Seminars in Cancer Biology 91 (2023) 50–69

Contents lists available at ScienceDirect

Seminars in Cancer Biology

journal homepage: www.elsevier.com/locate/semcancer

Artificial intelligence in cancer immunotherapy: Applications in neoantigen recognition, antibody design and immunotherapy response prediction

Tong Li [a,1], Yupeng Li [a,1], Xiaoyi Zhu [b,d,1], Yao He [a], Yanling Wu [b,d], Tianlei Ying [b,d,*,1], Zhi Xie [a,c,**]

[a] State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China
[b] MOE/NHC Key Laboratory of Medical Molecular Virology, Shanghai Institute of Infectious Disease and Biosecurity, School of Basic Medical Sciences, Shanghai Medical College, Fudan University, Shanghai, China
[c] Center for Precision Medicine, Sun Yat-sen University, Guangzhou, China
[d] Shanghai Engineering Research Center for Synthetic Immunology, Shanghai, China

ARTICLE INFO

Keywords:
Artificial intelligence
Cancer immunotherapy
Deep learning
Machine learning

ABSTRACT

Cancer immunotherapy is a method of controlling and eliminating tumors by reactivating the body's cancer-immunity cycle and restoring its antitumor immune response. The increased availability of data, combined with advancements in high-performance computing and innovative artificial intelligence (AI) technology, has resulted in a rise in the use of AI in oncology research. State-of-the-art AI models for functional classification and prediction in immunotherapy research are increasingly used to support laboratory-based experiments. This review offers a glimpse of the current AI applications in immunotherapy, including neoantigen recognition, antibody design, and prediction of immunotherapy response. Advancing in this direction will result in more robust predictive models for developing better targets, drugs, and treatments, and these advancements will eventually make their way into the clinical setting, pushing AI forward in the field of precision oncology.

1. Introduction

Immunotherapy has long been a key treatment for cancer. In addition to standard chemotherapy, radiotherapy, and surgery, the continuous development of biotechnology and the characterization of tumor molecular mechanisms has led to immunotherapy playing an increasingly important role in cancer treatment and becoming one of the leading treatments for cancer [1]. Different from traditional cancer treatment methods, immunotherapy is a treatment approach that uses the body's own immune system to fight cancer by targeting multiple targets and dynamically modulating the immune system [2]. The two primary cancer immunotherapies are immune checkpoint blockade (ICB) and adoptive cell therapy (ACT) [3]. The immune system's response to cancer cells, which has been suppressed, is strengthened using ICB. In the ICB strategy, immune checkpoint inhibitors, such as antibodies neutralized against programmed cell death protein 1 (PD-1) or cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), are used to reactivate tumor-specific T cells. The use of therapeutic antibodies can

Abbreviations: AI, artificial intelligence; ICB, immune checkpoint blockade; ACT, adoptive cell therapy; PD-1, programmed cell death protein 1; CTLA-4, cytotoxic T-lymphocyte-associated protein 4; TAs, tumor-associated antigens; NGS, next-generation sequencing; ADCs, antibody—drug conjugates; mAbs, monoclonal antibodies; ADAs, anti-drug antibodies; PD-L1, programmed cell death ligand-1; TMB, tumor mutational burden; MSI, microsatellite instability; dMMR, deficient mismatch repair; irAEs, immune-related adverse events; DNNs, deep neural networks; MHC, major histocompatibility complex; ML, machine learning; NNs, neural networks; ANNs, artificial neural networks; MS, mass spectrometry; BA, binding affinity; AP, antigen processing; DL, deep learning; PWMs, position weight matrices; NLP, natural language processing; GRU, gate recurrent unit; CNNs, convolutional neural networks; pMHCs, peptide-MHC complexes; TCRs, T cell surface receptors; TILs, tumor infiltrating lymphocytes; IHC, immunohistochemistry; TME, tumor microenvironment; FFPE, formalin-fixed, paraffin-embedded; NSCLC, non-small cell lung cancer; HCDR3, heavy-chain complementarity-determining region 3; DMS, deep mutational scanning; LDA, linear discriminant analysis.

* Corresponding author at: MOE/NHC Key Laboratory of Medical Molecular Virology, Shanghai Institute of Infectious Disease and Biosecurity, School of Basic Medical Sciences, Shanghai Medical College, Fudan University, Shanghai, China.
** Corresponding author at: State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China.
E-mail addresses: tlying@fudan.edu.cn (T. Ying), xiezhi@gmail.com (Z. Xie).
[1] Equally contributed authors.

https://doi.org/10.1016/j.semcancer.2023.02.007
Received 10 November 2022; Received in revised form 13 February 2023; Accepted 28 February 2023
Available online 3 March 2023

help to improve antitumor immunity by directing or engineering immune cells. Therapeutic antibodies (such as anti-CD20 rituximab) bind directly to tumor-associated antigens (TAs) to direct immune responses. For the ACT strategy, the patient's own immune cells are harvested and modified in vitro to enhance their antitumor ability [4]. Engineered ACT therapies, such as CAR-T and TCR therapies, involve the modification of a patient's T cells to enable them to target known TAs [5]. Both types of cancer immunotherapy rely on individual patient tumor-specific mutations that promote T-cell-mediated immune responses specific to the cancer [6]. In this review, we will discuss challenges in three important topics in cancer immunotherapy, including neoantigen recognition, antibody design, and immunotherapy response prediction (Fig. 1).

Neoantigens are mutations that encode immunologically active proteins that can cause the immune system to recognize the affected cell as foreign. Neoantigens are the key to the development of personalized cancer immunotherapies, such as personalized cancer vaccines, ICB and ACT [7]. The prediction of neoantigens, also known as neoepitopes, is a major challenge for computational approaches to immunotherapy and a necessity for narrowing down mutations for inclusion in vaccines or for high-throughput methods assessing T-cell recognition in vitro, as only a small number of mutations are actually immunogenic [8]. Many neoantigens predicted by next-generation sequencing (NGS) may not be able to be efficiently translated into proteins or peptides; this may explain why NGS predictions of certain tumors are not consistent with actual treatment effects [9]. A key challenge for immunotherapies that involve the transfer of TCRs into recipient patient T cells is the identification of safe target antigens. If engineered TCR-T cells cross-react with self-antigens in healthy tissue, the side effects can be devastating [10]. Not all tumors contain a sufficient number of immunogenic mutations; therefore, identification of a wider variety of shared TAs (gene fusions, alternative splicing, mutational frame shifts, endogenous retroviruses) could potentially broaden the scope and number of therapeutic cancer vaccines and immunotherapy efficacy assessments [11,12].

The design of antibodies is a key for therapeutic antibodies or antibody—drug conjugates (ADCs) that are used in cancer treatment. There are eleven ADCs and two bispecific antibodies, blinatumomab and amivantamab, that are alternative forms of approved anticancer antibodies. The success of these antibodies in treating cancer, particularly in the treatment of hematological malignancies, has spurred intensified efforts to develop next-generation anticancer antibodies with improved response rates or durations [13]. The prediction of antibody structure has a wide range of applications in the engineering of antibodies. Another challenge in computational antibody development is the ability to predict 'developability'. The term 'developability' of an antibody already encompasses a range of desirable drug-like qualities, including its manufacturability, storage stability, ease of administration, and favorable pharmacological behavior in patients [14]. This approach is similar to the Lipinski rule of five, which has proven to be very valuable in small molecule drug development [15]. The antibody sequence space is estimated to be composed of up to 10^{18} unique molecules [16]. One traditional way to find new antibodies is to create phage display libraries. These libraries provide access to up to 10^{11} possible molecules, which is just a small sample of the full possible space [16,17]. The challenge for antibody engineers is to develop a computational method to explore the antibody sequence space and identify new functional antibodies. In addition, all protein-based therapeutics, including monoclonal antibodies (mAbs), may potentially be immunogenic and elicit immune responses in humans. The immunogenicity of mAbs and anti-drug antibodies (ADAs) to cause adverse reactions is also a major concern [13,18,19].

Significant improvements in the cancer immunotherapy of different types of tumors has been achieved. Nevertheless, positive results from the treatment are only observed in a portion of patients [20]. The mechanisms of immune resistance have become a point of focus. Understanding the mechanisms of immunotherapy resistance and identifying potential predictors are essential for the development of effective treatments. The currently FDA-approved biomarkers for predicting the response to immunotherapy are programmed cell death ligand-1 (PD-L1), tumor mutational burden (TMB), and microsatellite instability/deficient mismatch repair (MSI/dMMR) [20]. Although some researchers believe that the expression of PD-L1 can be used as a biomarker for predicting immunotherapy, good efficacy may still be achieved for some cancer patients with low PD-L1 expression [21]. The complexity of the tumor microenvironment and the heterogeneity of tumor patients mean that a single biomarker cannot be used to predict the efficacy of immunotherapy. It is necessary to develop a predictive model that includes multiple parameters related to tumor-host interactions [22–24]. On the other hand, immune-related adverse events (irAEs) are a continuing complication of checkpoint blockade. Adverse events of grade 3 or higher occurred for $\geq 2/3$ of patients who received anti-PD-1 or anti-PD-L1 drugs [25,26]. It is becoming increasingly crucial to comprehend what causes these irAEs and how to address them.

AI technology encompasses multiple technologies with the common goal of computationally simulating human intelligence. In the past decade, unprecedented success has been achieved in processing natural data forms such as images, text, and speech using deep learning (DL) algorithms, such as deep neural networks (DNNs) [27,28]. AI has many applications in cancer research and precision medicine, such as cancer diagnosis, molecular characterization, tumor microenvironment characterization, pharmacogenomics discovery, and clinical outcome prediction [29,30]. AI is excellent for complex pattern recognition in large quantities of data and can provide quantitative assessments in an automated fashion. AI approaches allow algorithms to become better at performing specific tasks and generating decision support systems by

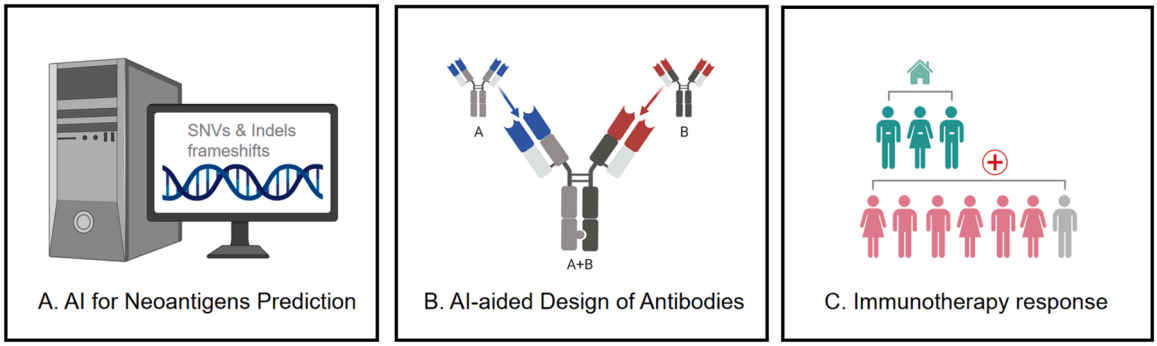

Fig. 1. Applications of AI in the cancer immunotherapy. AI is being applied to immunotherapy, comprising of three steps: **A.** identifying novel neoantigens, **B.** designing antibodies, and **C.** predicting immunotherapy effects. This order of tasks is in line with the process of drug research and development, comprising target discovery, drug design, and efficacy measures.

accumulating medical data, such as omic, radiology, pathology, and clinical data, and associated outcomes.

Here, we will review the application of AI in immunotherapy, focusing on three main themes (Fig. 1). First, tumor neoantigens are the basis of immunotherapy, and the use of AI to predict immunogenic tumor antigens rapidly and accurately, reducing experimental screening and validation, represents a major unsolved challenge. Second, despite the significant success of tumor therapeutic antibodies, there is still much room for improvement that is inspiring much innovation in antibody design. AI-enhanced antibodies hold great potential for further success in cancer treatment. Third, we discuss the challenges of predicting immunotherapy response, i.e., the identification of patients most likely to respond to immunotherapy using multimodal, multiscale biomarkers and the characterization of the tumor immune microenvironment. In the end of this review, we will summarize the challenges and current limitations of AI in tumor immunotherapy applications and provide our thoughts on how to improve the applicability of AI in the future.

2. AI for neoantigen prediction

Cancer immunotherapy is based on the body's antitumor immunity, which begins with the activation of specific T cells for tumor elimination [31]. Neoantigens are tumor-specific peptides produced by somatic mutations that are able to elicit T cell immune responses and therefore are potential targets for cancer immunotherapies [32,33]. However, the proportion of tumor peptides recognized by T cells is very low [34,35]. In this case, it is necessary to precisely identifying neoantigens. There are two strategies for neoantigen detection: identifying TAs with T cell response experimentally, which is expansive with low efficiency, or predicting neoepitopes with computational approaches, which have high a number of false-positive results but have been continually improved. In this section, we will review applications of AI for neoantigen prediction and focus on the type of algorithms, datasets, and categories that are cores for these methods.

2.1. AI for peptide-MHC binding prediction

The binding of peptides to the major histocompatibility complex (MHC) is a key step for immune recognition [36,37]. Although identifying epitopes that activate immune response with this single process has been questioned [38], peptide-MHC binding is one of the most critical steps for neoantigen prediction in many current approaches [39], and many prediction methods are still emerging today.

MHC molecules are located on the cell surface. They are classified into type I and type II molecules, presenting peptides to CD8+ and CD4+ T cells, respectively. Tumor mutant peptides presented by MHCs could be recognized by T cells and activate a specific response to eliminate tumor cells. For decades, methods for peptide-MHC binding prediction have been developed, and MHC I binding methods still represent most approaches because the MHC I groove is closed and accommodates only short peptides. AI algorithms, especially machine learning (ML) algorithms, have become mainstream with the accumulation of experimental data (Fig. 2); thus, accuracy is continuously improving. ML algorithms are a category of algorithms that are used to identify and learn patterns in data through the use of statistical methods. Then, classifications or predictions are made on new data. In ML, no rigid assumptions must be made, and thus, it becomes possible to find complex nonlinear dependencies [40] similar to peptide-MHC interactions. There are many ML models, such as neural networks (NNs), decision trees, and support vector machine (SVM).

Data are the core of ML, and initially, algorithms learned from a large amount of peptide-MHC binding affinity data. NetMHC [41], which is a well-known predictor for peptide-MHC I binding, has been utilized in several benchmark analyses [42–44]. It uses binding affinity measurements from IEDB [45] (Table 2) to train an NN model, which acquired the nonlinear relationship between peptide sequences and the binding affinity of homologous MHC molecules [46]. Using the latest version of NetMHC 4.0 [47], more than 90 % of naturally presented MHC ligands have been successfully identified with 98 % specificity [48]. However, there is a long distance between the identified neoantigens and therapy requirements for a high false-positive prediction rate [49]. One potential cause for this deviation lies in the methods trained from affinity measurements only modeling a single event for peptide-MHC binding.

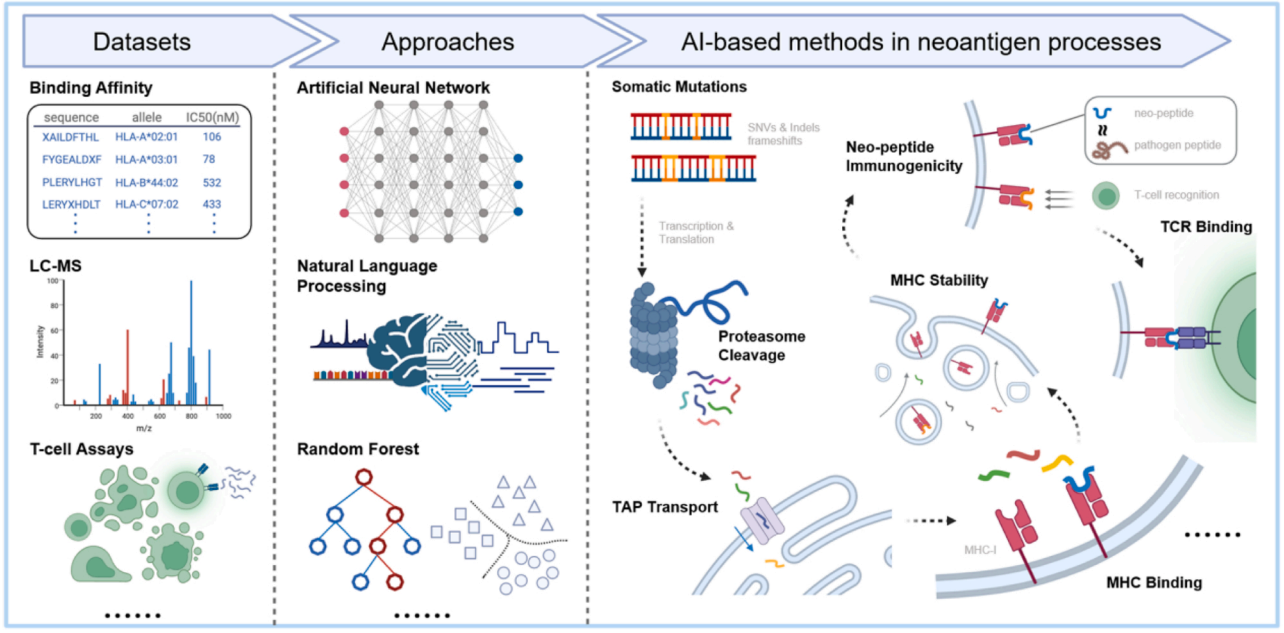

Fig. 2. AI-based approaches applied on neoantigen prediction. The neoantigen processes is showed in the MHC-I pathway. SNV: single nucleotide variants, Indel: insertion-deletion, MHC: major histocompatibility complex, pMHC: peptide-MHC complex, TCR: T-cell receptor, TAP: transporter associated with antigen processing, LC-MS: Liquid Chromatograph Mass Spectrometer.

Realizing this problem, another NN-based method, NetMHCpan 4.0, was developed and trained on both affinity measurements and 85,217 MHC ligands eluted by mass spectrometry (MS), providing more antigen processing signals [36]. The most recent versions are NetMHCpan-4.1 and NetMHCIIpan-4.0, for MHC I and MHC II ligand predictions, respectively. Similar to NetMHC 4.0 and NetMHCpan 4.0, MHCflurry [50] was also trained on both affinity measurements and MHC binding peptides with MS-identified ligands in training sets expanded to 226,684 ligands and performed well [46]. The updated version MHCflurry2.0 was used to train two models to predict binding affinity (BA) and antigen processing (AP), separately, on larger datasets. The BA model with deeper hidden layers (2–3) outperforms the shallow version, but most of the improvement in accuracy may be primarily due to the incorporation of new datasets [51].

The improvement of data quality, both in data size and relevance to the peptide presentation process, and the refinement of AI algorithms continue to improve the prediction tools. EDGE is an approach that model epitope presentation from two types of MHCs [38]. The training sets contain 142,844 MHC-presented peptides from 101 individuals with different cancers, which could help to capture more information on antitumor immunity [34,52,53], and were used to construct a model with DL. In DL, hierarchical layers are used to discover complex patterns from high-dimensional data after transforming input in a nonlinear fashion. In EDGE, additional allele-specific models and two features, transcript abundance and flanking sequence [38,54], are also integrated to perfect the pipeline. Furthermore, the performance of EDGE was validated with the identification of neoantigen-reactive T cells in cancer patients. Different from the methods above, MixMHCped2.0 was trained on motifs extracted from MHC I binding peptides with position weight matrices (PWMs), which mathematically describe each allele's binding motif by computing the frequency of each amino acid at each position [55]. A PWM is a linear-based approach; even so, it has been widely used to represent peptide binding to MHCs by deriving the corresponding motifs [56–58]. The model seems to perform well [44] and has aroused the thought that the MHC binding process is not as complex as it seems [59]. In addition, in MixMHCped 2.0, the length distribution and multiple specificity of peptides were integrated, enabling more features related to neoantigen presentation to be addressed. Some other tools employed multiple AI approaches in model construction. MHCSeqNet is a DL model trained on both MHC binding affinity and MHC ligand peptidome datasets [60]. The architecture of the model comprised five layers with different NN approaches. For example, Natural Language Processing (NLP), which can capture the semantic and syntactic relationships between words in a sentence, was applied in peptide embedding layer, and Gate Recurrent Unit (GRU), a subset of RNN and is a very useful for fixing the vanishing gradient problem, was chosen as the peptide processing layer.

Although many tools have been developed for MHC I peptide processing, predicting neoantigens presented by MHC II is still challenging. Firstly, the open groove results in a variant ligand length and flexible core position. Secondly, the MHC II pathways are more complex and poorly studied compared to those of MHC I. Thirdly, there is limited data available. Even so, MHC II target tools are necessary in some cases in which neoantigens recognized by T cells are MHC II restricted [35]. NeonMHC2 is a specific predictor of peptide-MHC II presentation [61]. To overcome the lack of high-quality data required for peptide-MHC II binding prediction, a technology that enables efficient isolation of MHC II peptides binding a single allele for MS-based identification with 6 distinct cancer and antigen-presenting cells (APCs) cell lines were developed. Then, MHC II binding motifs were resolved from prepared datasets, and a model with convolutional neural networks (CNNs) was constructed and successful used in the field of computer vision because of its proficiency in translationally invariant pattern recognition [62]. MARIA is another state-of-art tools for MHC II antigen prediction [63]. To comprehensively capture the features in antigen presentation, the DL approach integrated with separate models on peptide binding, presentation, and cleavage was made.

The improvement of peptide-MHC binding prediction with AI algorithms relies on datasets meeting requirements related to both quantity and relevance to biological processes and ingenious categories for model construction. Here, we focus on tools using ML and summarize some typical methods with model training-associated information, including the approaches, datasets, and assessments in independent studies (Table 1). A recent review provided a comprehensive overview of the effectiveness and usage of existing tools and pipelines [64]. In addition, many benchmarks have been published [44,46,64], but assessments on more recent tools are needed. In general, epitope-MHC interactions are critical in the process of neoantigen-elicited immune therapy and are involved. Accurate prediction of MHC binding or presentation could undoubtedly contribute to neoantigen identification, but defining true neoantigens requires the use of AI to build analytical tools for other critical steps, such as peptide posttranslational modification and immunogenicity [59].

2.2. AI methods for peptide immunogenicity

Immunogenicity is the ability to induce adaptive immune responses [65], which are necessary for neoantigens to be used in tumor immunotherapy through T cell response activation. The fraction of somatic mutations to be recognized by a spontaneously occurring T cell response in a given cancer is very small [34,35,66], and the reason why some mutant peptides are immunogenic and others are not is not clear [67]. In this case, approaches for distinguishing immunogenic peptides are needed. As mentioned above, the processes of peptide binding and presentation by MHCs are crucial for T cell recognition, and great improvements have been driven by the increase in MS data and ML models. However, the information directly related to peptide immunogenicity cannot be captured. Therefore, several tools trained from immunogenicity-related data or features with AI categories have been developed.

It was suggested that peptide-MHC binding stability could be correlated with immunogenicity in the early days [36]. Peptides that elicit immune responses need to have stable binding to MHCs to form peptide-MHC complexes (pMHCs), which are then transported to the cell surface and maintained for a long enough time to be recognized and bound by T cell surface receptors (TCRs). NetMHCstabpan is an NN-based method trained on the half-life of MHC I [67]. The combination of predicted pMHC affinity and stability can significantly improve the ability to screen epitopes recognized by T cells [67,68]. Furthermore, the performance of affinity and stability predictions in the case of the differences in the amounts of available data was compared using balanced networks. The process of neoantigen calling T cell immune response is diverse and involves multiple steps, which make it challenging to predict with a single feature. In Neopepsee, collected epitopes exhibit T cell response, and on the basis of 14 selected immunogenicity features, ML-based classifiers for four models were constructed: Gaussian naïve Bayes (GNB), locally weighted naïve Bayes (LWNB), RF, and SVM [38]. Although limitations may exist on training sets lacking negative candidates corresponding to verified neoantigens, integrating multiple features for model learning could be a useful strategy as described in another study [69]. To better utilize immunogenicity for neoantigen prediction, methods based on more complicated algorithms have gradually been developed. In DeepHLApan, two models, the binding model and immunogenicity model, were constructed to separately predict the probabilities of presenting peptides to the cell surface and eliciting T cell activation [70]. To achieve these goals, datasets of peptides with MHC binding and immunogenicity measurements were collected. Then, RNN, which is a DL approach that handles variable-length sequences with a recurrent hidden state whose activation at each time is dependent on that of the previous time, was applied for training. Using this method, an immunogenic score for querying peptides can be output and applied to CD8+ T cell epitopes.

Table 1
Application of different AI algorithms in immunotherapy.

1. AI-based methods for neoantigen prediction

Name	Year	MHC	Approach	Neoantigen process	Type of learning data	Training sets	Independent benchmark
EDGE [38]	2019	I/II	DL	MHC binding	MHC ligands	142,844 peptides from 101 samples	
DeepHLApan [225]	2019	I	RNN	MHC binding & TCR binding	MHC ligands & T-cell recognized neoepitopes	437,077 peptides for binding model & 32,785 peptides for immunogenic	
NMER [63]	2021	I	RF	TCR binding	T-cell recognized neoepitopes	185 neoepitopes identified by TILs from 96 samples	
NetMHC-4.0 [47]	2016	I	NN	MHC binding	Affinity measurements	Affinity measurements from IEDB	[44,46]
NetMHCpan-4.0 [36]	2017	I	NN	MHC binding	Affinity measurements & MHC ligands	Affinity measurements from IEDB & 85,217 peptides	[44,46]
NetMHCstabpan [67]	2016	I	NN	pMHC stability	Half-life of pMHC in vitro	28,166 binding measurements	[44]
MHCflurry [50]	2018	I	NN	MHC binding	Affinity measurements & MHC ligands	230,735 affinity measurements & 226,684 peptides	[44,46]
MHCflurry-2.0 [51]	2020	I	NN	binding affinity & MHC ligands	Affinity measurements & MHC ligands	219,596 affinity measurements & 493,473 peptides for binding model; 399,392 peptides for presentation model	
Neonmhc2 [61]	2019	II	CNN	MHC binding	MHC ligands	> 50,000 peptides	
Neopesee [226]	2018	I	ML	Immunogenicity	Epitopes with T-cell response	311 neoepitopes with T-cell response	
pMTnet [227]	2021	I	DL	Immunogenicity	Affinity measurements & pMHC-TCR pairs	172,422 affinity measurements & 243,747 human TCRβ CDR3 sequences & 32,607 pMHC-TCR pairs	
ForestMHC [228]	2019	I	RF	MHC binding	9-mer peptides from MS	> 160,000 peptides	
PRIME [96]	2021	I	LR	MHC binding & TCR binding	Peptides with immunogenicity or not	4958 peptides	[229]
MARIA [54]	2019	II	RNN	MHC binding & protease cleavage & expression	MHC ligands	8374 peptides for presentation model & 33,909 peptides for binding model & 12,150 for cleavage model	
MHCSeqNet [60]	2019	I	NLP, GRU	MHC binding & ligand prediction	MHC ligands	228,348 peptides	
HLAthena [230]	2019	I	NN	MHC binding	MHC ligands	> 185,000 peptides	
NetTCR-2.0 [231]	2021	I/II	CNN	Immunogenicity	TCR-peptide pairs	9204 CDR3β-peptide pairs & 4598 CDR3α-/β-peptide pairs	
NetMHCPan-4.1 [232]	2020	I	NN	MHC binding	Affinity measurements & MHC ligands	208,093 affinity measurements & 665,492 peptides	
NetMHCIIpan-4.0 [232]	2020	II	NN	MHC binding	Affinity measurements & MHC ligands	108,959 affinity measurements & 381,066 peptides	

2. AI-based methods for antibody design

Role	Name	Model	Training input	Description	Ref.
Target binding prediction and optimization	Ens-Grad	Ensemble of CNN	51,130 CDRH3 sequences and their R2-to-R3 enrichment in phage-display panning	Designing antibody CDRH3 regions with target affinities and improved specificity	[107]
	-	CNN	CDRH3 sequences of 11,300 binding and 27,539 non-binding trastuzumab variants	Optimization of trastuzumab in affinity and pharmaceutical properties	[108]
	-	CNN and generative adversarial network	6003 non-binder and 1345 binder sequences for CTLA-4; 6052 non-binder and 1719 binder sequences for PD-1	Classifying and generating sequences of CTLA-4 and PD-1 binding antibodies	[109]
	-	LSTM	959 VH sequences	Generating antibody sequences with improved affinity to kynurenine	[110]
	-	linear discriminant analysis model	4000 sequences (2000/2000 with high and low specificity; 1516/2484 with high and low antigen binding)	Co-optimizing the affinity and specificity of emibetuzumab	[111]
Antibody structure prediction	DeepH3	Series of CNN	1388 structures from SAbDab	Predicting inter-residue distances and orientations of antibody CDRH3 region	[116]
	DeepAb	ReNet and bi-LSTM encoder	118,386 paired sequences from the OAS database and 1692 structures from SAbDab	Predicting relative distances and orientations of antibody Fv region and realizing structure by Rosetta	[121]
	DeepSCAb	ResNet for the inter-residue module and a transformer encoder model for the rotamer module	1433 structures from SAbDab	Predicting full Fv structures including side-chain geometries	[126]
	ABlooper	five E(n)-equivariant graph neural networks	3438 structures from SAbDab	Predicting structures of antibody CDR loops	[127]
	NanoNet	Two 2D CNN	~ 2000 heavy chains of mAbs and nanobody structures	Modeling nanobodies and VH domains of antibodies	[130]
Pharmaceutical property	AbLSTM	Bi-LSTM network	25,000 antibody sequences from BCR repertoire sequencing	Evaluating the nativeness of antibody candidates	[135]
	BioPhi	RoBERTa transformer encoder model for Sapiens	Human antibody repertoires in the OAS database	A platform composed of OASis which is for antibody humanness evaluation and Sapiens for antibody humanization	[136]
	solPredict		Protein solubility data of 220 antibodies		[138]

(*continued on next page*)

Table 1 (*continued*)

2. AI-based methods for antibody design

Role	Name	Model	Training input	Description	Ref.
		Support vector machine and random forest models with language-based transfer learning		Predicting the apparent solubility of antibodies in histidine (pH 6.0) buffer	
	DeepSCM	1D CNN	6596 antibody Fv sequences and their SCM scores	A surrogate model for the SCM to predict antibody viscosity	[140]
	-	Random forest classifier	64 clinical-stage antibodies and their PK data	Identifying biophysical assays and in silico properties correlated with antibody PK	[144]

3. AI-based methods for immunotherapy response prediction

	Target	Summary	Data type	Tumor type	Technology	Validation	Ref.
Prediction of associated biomarker	MSI	Predict MSI directly from H&E histology	H&E histology	Gastrointestinal cancer	CNN with deep residual learning (resnet18)	AUC = 0.84	[158]
	MSI or dMMR	Identified dMMR or MSI from H&E-stained slides	H&E-stained slides	Colorectal cancer	Modified ShuffleNet/ MobileNetV2 architecture	AUPRC = 0.92–0.931, 66.6–67 % SPE and 76.0–95 % SEN	[161, 163]
	PD-L1 score	Quantitatively analyze the PD-L1 score of tumor cells	PD-L1-stained digital images	Cutaneous melanoma	ML algorithm (random forest classifier)	Correlation between label and prediction (r = 0.97, P < 0.0001).	[164]
	PD-L1 expression	Predict the TPS of PD-L1 expression	WSIs of the 22c3 assay	Non-small cell lung cancer	U-Net structure with residual blocks	ACC = 0.9326 and SPE = 0.9641	[166]
	TMB	Predicting TMB from histopathological images	H&E-stained histopathological slides	Lung adenocarcinoma	Inception-v3 and the random forest architecture	AUPRC = 0.92, precision = 0.89	[168]
	TMB status	Predicting TMB from images and clinical information	Histopathological images and clinical data	Colorectal cancer	Multi-modal deep learning model based on ResNet	AUPRC = 0.817	[172]
Deciphering TME	TME reconstruction	Prediction of 51 unique cell subpopulations	Bulk RNA-seq	Pan-cancer from TCGA	Decision tree ML deconvolution algorithm	Cytometric, immunohistochemical, or scRNA-seq	[182]
	TILs maps	Predict spatial patterns of TILs	H&E images	13 TCGA tumor types	Semi-supervised CNN and unsupervised CAE	Prediction correlate with pathologist and molecular estimates	[189]
	Tumor Cellularity	Predict tumor purity from H&E Image	Whole Slide H&E Image	Breast cancer	Weakly-supervised segmentation model with Resnet-34	Cohen's kappa coefficient = 0.69	[193]
	TME	Predict spatial mapping of multiple cell types	H&E images and spatial transcriptomic data	Lung adenocarcinoma	CNN	Correlated with bulk RNA-seq	[202]
Directly prediction	Immunotherapy response	Predict ICI treatment responses in three different cancer types	Clinical outcome and transcriptomic data	Melanoma, gastric cancer, and bladder cancer	Protein—protein interaction network (PPI)-based ML	AUCs > 0.7	[207]
	Immunotherapy response	A predictive model of immunotherapy using genomics data	Genomic data and RNA-seq	29 cancer types from TCGA	MultiModal Network	Compared to benchmark markers (p = 0.009)	[208]
	ICB efficacy	Predict ICB response	Genomic, molecular, demographic and clinical data	16 different cancer types	Ensemble learning random forest ML model	Pan-cancer AUC = 0.85	[209]
	Immunotherapy response	Prediction of response to PD-(L)1 blockade	Medical imaging, histopathological, and genomic features	Non-small cell lung cancer	Multiple-instance LR and multimodal dynamic attention	AUC = 0.80	[216]

The indirectly strategy means that AI is used to predict TMB, MSI status, PD-L1 expression or TME cellular composition. The directly strategy means that AI is used to directly predict treatment response. A selection of representative studies have been referenced, and it should be noted that the list may not encompass all categories completely.

Some other tools make use of data closer to the "real scene". NMER, a model for MHC I neoepitopes, was trained on 185 neoantigens recognized by tumor infiltrating lymphocytes (TILs) and corresponding negative controls from 96 individuals with metastatic cancers [61]. The high-quality data and RF model were used to provide an estimate of the relative influence of multiple features of MHC I epitopes and facilitate neoantigen identification.

In addition, pMHC-TCR binding prediction is a way in which to measure the ability of epitopes to elicit T cell immune responses directly. The TCR binds to the pMHC by interacting with the side chains of peptides exposed on the outside of the MHC groove [71,72]. The same neoepitope-MH complex can bind to T cells with different TCRs, which may be composed of molecularly distinct α- and β-chains [73–77]. pMTnet can be used to predict TCR binding specificities of neoantigens presented by MHC I. It utilizes the TCR sequence, pMHC binding affinity measurements, and TCR-pMHC binding pairs to train multilayer networks with transfer learning, a branch of DL that applies additional data or existing models to relevant new tasks. To lower the difficulty level of the prediction task, three steps were performed. The protein sequences of epitopes and MHC and the TCR sequence (CDR3β) were modeled into numeric embedding in the first two steps. Then, these two embeddings were combined into the final model, which was used to the predict TCR epitope binding specificity given the sequence of the TCR and epitope and the type of MHC I.

Another approach is to identify TCRs specific to available neoantigens to promote the understanding of pMHC-TCR binding mechanisms and measure peptide immunogenicity. Some clustering-based methods were developed. iSMART is used to group TCRs by their CDRβ3 sequence to identify tumor antigen-specific TCRs through pan-cancer multi-omics analysis [78]. In addition, novel tumor-associated

antigens were also found using this model. DeepTCR is a platform able to model complex TCR sequencing data by learning a joint representation of a TCR by its CDR3 sequences and V/D/J gene usage [79]. The model applied both unsupervised and supervised DL on TCR sequences to learn patterns for descriptive or predictive purposes, and improved antigen-specific TCR classification.

Although T cell acts a pivotal part in immunotherapy with neoantigen, the role of B cells is also noteworthy. B cells can act as APCs to assist in neoantigen spreading [80] or be activated to differentiate and produce tumor-specific antibodies [81]. A study of murine lung adenocarcinoma (LUAD) found that B cell-recognized neoantigens drive tumor-specific B cell and TFH cell responses which is important for anti-tumor effector CD8 T cell responses [82]. Therefore, the ability to elicit B cell response as epitopes can be used to evaluate neoantigen immunogenicity. B cell epitopes are divided into two categories: linear and conformational epitopes [83]. ML methods for B cell epitope prediction have been used, such as neural network, SVM or random forest, which are trained on features of known B cell epitopes [84]. For example, EpiDope trained a DNN with almost 25,000 experimentally verified epitope and non-epitope regions for linear B cell epitope prediction [85]. Epitope3D collected almost 1600 unbound antigen structures from experimental antibody-antigen complexes to predict conformational B cell epitope with Adaboost classifier [86]. Till now, optimization on neoantigen selection based on B cell reorganization faces great challenges because of the low performances of B cell epitope prediction tools [87] and the lack of B cell response to neoantigen-targeting vaccines [88].

Neoantigen prediction based on immunogenicity with multiple indicators integrated with AI modeling has gradually become mainstream, and problems have emerged, including the definition and integration of standard data and the evaluation of different methods [86] (Table 1). Furthermore, although sequence-based methods have developed rapidly, insights into the structure-dependent mechanisms of epitope-specific T cell recognition is also important [87]. Approaches for structural modeling [88], such as AlphaFold [89], would help.

2.3. Neoantigen prediction with multiple dimensions

In addition to MHC presentation and TCR recognition, studying other processes can also help with neoantigen identification. Next, we will present some AI-based methods focusing on these processes (Fig. 2).

In the approach of computational neoantigen prediction, aberrant peptide detection is the first step, and somatic mutations, including SNVs, small insertions, deletions (indels), frameshift mutations, and gene fusions, are the source of the most characterized neoantigens [9,90,91]. Among these, SNVs and small indels are commonly studied. In MuTect, a Bayesian classifier is applied to detect somatic point mutations [92]. A Bayesian classifier is a type of ML classifier based on Bayes' theorem, which is simple and effective for predicting a class of datasets. The method takes tumor and corresponding normal samples as input data and is sensitive to low-allelic fraction events. When translated into the cytoplasm, the neopeptides need to undergo proteasome cleavage and transport before binding to the MHC molecule. NetChop was trained on MHC I ligand data using an NN to predict the combined specificity of cleavage by immunoproteasomes and constitutive proteasomes [93]. Furthermore, TAP and MHC affinity data were integrated in NetChop for better effect. Another method that was used to perform predictions for all MHC I molecules is NetCTLpan [94]. AutoRT has been used to predict peptide retention with high accuracy [95]. Peptide retention time is an intrinsic feature that refers to the time points when peptides elute from a liquid chromatography (LC) column, as recorded by the instrument. The method was trained on more than 100,000 peptides using DL and can be used to classify tumor-specific peptides.

Furthermore, some other properties are lacking in AI-based predictors but show impressive performance for neoantigen ranking, such as peptide foreignness and agretopicity [39], which means neoepitopes with similar sequences to pathogen peptides or dissimilarity to self are more likely to elicit T cell responses [73]. Recently, the two features were demonstrated in PRIME, a predictor of immunogenic epitopes, but were not combined in the model [96]. AI approaches may be effective in modeling them through training on peptidome datasets.

As AI-based tools for neoantigen prediction cover more processes, it will be a challenge to integrate multiple processes. pTuneos [97] is a computational strategy that is used to determine neoantigens with multiple dimensions, such as MHC I presentation, T cell recognition, proteasome cleavage, and TAP transport. In this study, a novel RF model was used to score the epitope based on the related features. pVACtools is a cancer immunotherapy tools suite for neoantigen identification and visualization [98]. The framework produces an end-to-end solution for neoantigen characterization and DNA-based cancer vaccine design. Several features were considered for neoantigen ranking, and the MHC presentation model supports many current methods, of which eight are for MHC I and four are for MHC II.

The size and quality of training data are at the heart of all ML and DL models and largely determine their performance, robustness, and accuracy [59]. Thus, training and integrating models from datasets from multiple types can be useful. There are basically three types of data used in current neoantigen prediction methods: binding affinity data, eluted peptide MS data, and T cell assay data. IEDB is a free resource that allows the user to easily search for experimental data characterizing antibodies and epitopes studied in humans, nonhuman primates, and other animal species [99]. The database was established in 2004 and now contains more than 1,500,000 epitopes primarily gathered from more than 23,000 studies, and the data are continually expanded. Epitope information can be obtained from assays of B cells, T cells, and MHC ligands. Databases of validated neoantigens collected from a large number of studies could be useful without considering data imbalance [67]. dbPepNeo is a database of human neoantigens that have been experimentally validated and compiled from immunology literature and existing databases (Table 2). In its updated version, more than 800,00 neoantigens were provided with additional novel data types, such as MHC II neoantigens and fusion or noncoding region-sourced neoantigens. Immune cells recognize heterogenous molecular structures through immune receptors (BCRs and TCRs). Neoantigen-TCR interactions can be learned from the immune repertoire, which has been reviewed in detail in other articles [59].

3. AI-aided design of therapeutic monoclonal antibodies

Since the first monoclonal antibody (mAb), rituximab, was approved in the USA for the treatment of patients with non-Hodgkin's lymphoma in 1997 [100], mAbs have become a major component in cancer therapy, topping the list of best-selling drugs worldwide, largely owing to their high degrees of specificity and efficacy [101]. However, most of the antibodies on the market were developed using conventional discovery techniques, namely, display library screening and animal immunization, which are time-consuming and labor-intensive. For continued exploitation of therapeutic antibodies, it is critical to identify novel methods to mine these molecules more efficiently. The development of computational methodologies and ML approaches in the past decade opened a new era of in silico antibody engineering, as summarized by other teams [102–105]. In this part, we mainly focus on new possibilities that DL brings to therapeutic antibody design in terms of structure prediction, target binding screening and affinity maturation as well as pharmaceutical property prediction (Fig. 3).

3.1. AI for target binding antibody screening and affinity maturation

The advent of high-throughput sequencing (HTS) has enabled much easier and inexpensive access to large-scale antibody sequence information [106], offering a powerful method for screening hybridoma, phage, or yeast display libraries. However, even at the hundreds of

Table 2
Datasets for different applications of immunotherapy.

1. Datasets for neoantigen prediction

Name	Year	Last updated	Overview of the data
IEDB [99]	2004	2023/1/30	> 1,500,000 epitopes from > 23,000 studies, including B cell, T cell, and MHC ligand assays
dbPepNeo2 [233]	2022	No record	801 HC neoantigens and 864,884 low LC peptidomes, validated neoantigen peptides, TCRs, and HLA peptidomes

2. Datasets for antibody design

Database name	Description	Link	Ref.
Observed Antibody Space (OAS)	Collecting immune repertoires	https://opig.stats.ox.ac.uk/webapps/oas/	[122]
Structural Antibody Database (SAbDab)	Database of antibody structures	https://opig.stats.ox.ac.uk/webapps/newsabdab/sabdab/	[117]

3. Datasets for immunotherapy response prediction

Dataset name	Description	Link	Type(s) of cancer (s)	Ref.
Predict MSI in gastrointestinal cancer	Histological images for MSI vs. MSS classification in gastrointestinal cancer	https://doi.org/10.5281/zenodo.2530789 https://doi.org/10.5281/zenodo.2530835 https://doi.org/10.5281/zenodo.2532612	Gastrointestinal cancer	[158]
The Cancer Genome Atlas (TCGA)	TCGA is a pool of molecular data sets publicly accessible and freely available to cancer researchers	https://portal.gdc.cancer.gov/	Pan-cancer	[161]
NCT-CRC-HE-100K and CRC-VAL-HE-7K datasets	This is a set of 100,000 image patches from H&E stained histological images of colorectal cancer (CRC) and normal tissue.	https://doi.org/10.5281/zenodo.1214455	Colorectal cancer	[163]
MMDL-colorectal cancer	Predicting colorectal cancer TMB from histopathological images and clinical information	https://github.com/hkmgeneis/MMDL/tree/master	Colorectal cancer	[172]
Kassandra	Precise reconstruction of the TME using bulk RNA-seq and a ML algorithm trained on artificial transcriptomes	https://science.bostongene.com/kassandra/	Pan-cancer	[182]
CIBERSORTx	Determining cell type abundance and expression from bulk tissues with digital cytometry	http://cibersortx.stanford.edu/	Pan-cancer	[183]
Scaden	DL–based cell composition analysis from tissue expression profiles	https://www.science.org/doi/10.1126/sciadv.aba2619#sec-4	Pan-cancer	[184]
MethylCIBERSORT	Pan-cancer deconvolution of tumor composition using DNA methylation	https://www.nature.com/articles/s41467-018-05570-1#Sec12	Pan-cancer	[185]
MethylNet	An automated and modular deep learning approach for DNA methylation analysis	https://doi.org/10.24433/CO.6373790.v1	Pan-cancer	[186]
Human-interpretable image features (HIFs)	HIFs derived from densely mapped cancer pathology slides predict diverse molecular phenotypes	https://www.nature.com/articles/s41467-021-21896-9#data-availability	Pan-cancer	[194]
NetBio	Network-based ML approach to predict immunotherapy response in cancer patients	https://zenodo.org/record/4661265 http://research-pub.gene.com/IMvigor210CoreBiologies/https://string-db.org/	Pan-cancer	[207]
DeepTCR	DL reveals predictive sequence concepts within immune repertoires to immunotherapy	https://zenodo.org/record/6590069	Pan-cancer	[210]
PD-(L)1 blockade in patients with NSCLC	Multimodal integration of radiology, pathology and genomics for prediction of response	https://www.synapse.org/#!Synapse:syn26642505 https://www.cbioportal.org/study/summary?id=lung_msk_mind_2020	Non-small cell lung cancer	[216]

millions scale, there still exists a large, unexplored latent sequence space, where the unique advantages of DL can be integrated to make rational predictions. In this aspect, Liu's team provided a successful example combining a phage display library, HTS and NNs to design the heavy-chain complementarity-determining region 3 (HCDR3) of antibodies for desired affinity and specificity [107]. Instead of using accurate antibody affinity data, they applied round-to-round sequence enrichment during library panning as a measurement of antibody affinity and designed an ensemble of 18 NNs with 6 different architectures to predict the enrichment of antibodies outside the training space, which was demonstrated to outperform any single network in all metrics considered. To propose novel sequences with improved affinity, the authors further employed a gradient-based optimization framework (Ens-Grad) to provide modification strategies maximizing the enrichment of output sequences. As validated by experimental methods, the optimized sequences produced by Ens-Grad possessed on average and at the extremes better affinities than those observed in the training set, thus demonstrating the great potential of DL to empower conventional antibody affinity engineering schemes.

In another work, Mason et al. combined a mammalian cell display library with DL to optimize the classic therapeutic antibody trastuzumab targeting human epidermal growth factor receptor 2 (HER2) [108]. Although only the variable HCDR3 region was considered, the theoretical protein sequence space was still too large due to combinatorial explosion and contained sparse potential binders. To address this problem, the authors performed single-site deep mutational scanning (DMS) in the HCDR3 of trastuzumab to guide the rational design of the combinatorial mutagenesis library, whose theoretical size was sharply reduced from 10^{13} to 7.17×10^{8}. By sorting a miniscule library of 10^{4} variants, they identified 11,300 and 27,539 unique HER2 binders and nonbinders and used their sequence information as input to train multiple ML models, and the best performance was achieved using CNN. To experimentally validate the precision of their model, the authors randomly selected 30 predicted binders and 12 predicted nonbinders and found only one false negative variant. One highlight of this work is the inclusion of several in silico filtering steps to search for potential binders with better developability parameters than the parent trastuzumab.

In contrast to optimizing a specific antibody, Lim et al. isolated B cells from CTLA-4- or PD-1-immunized mice, encapsulated them into microfluidic droplets, performed PCR amplification of natively paired heavy and light chain sequences and generated a yeast library to display single-chain variable fragments (scFvs) [109]. By deep sequencing the

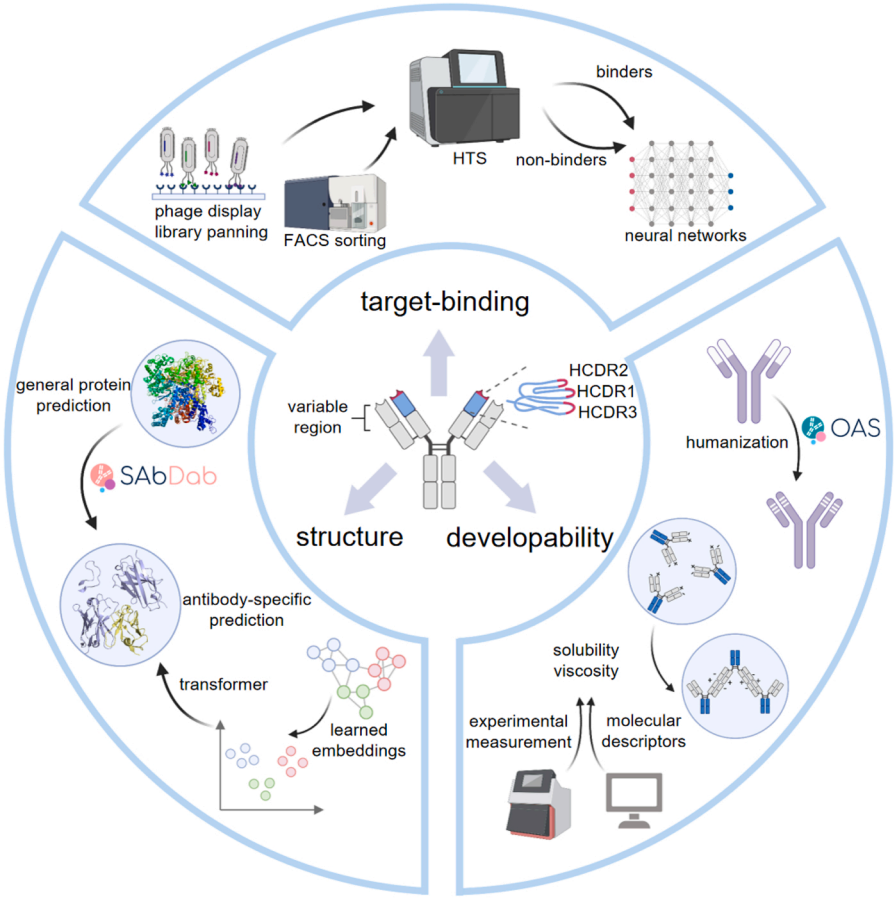

Fig. 3. Applications of DL in the discovery of therapeutic antibodies. The diversity of antibodies is enriched in their variable regions, especially the heavy chain complementarity determining region 3 (HCDR3). Here we highlight three key components in antibody development: (i) **target-binding prediction:** combining phage-display library panning or FACS sorting with high-throughput sequencing (HTS), we could obtain sequences with labeled target binding abilities to take as input to train neural networks to predict binding abilities of novel sequences; (ii) **antibody structure prediction:** based on some models for general protein structure prediction (e.g., AlphaFold2, RoseTTAFold), antibody-specific models, predicting only the HCDR3 region, the variable fragment, or nanobodies, can be trained using data from the structural antibody database (SAbDab). Moreover, pretrained protein models can learn meaningful representations of sequences thus improve the performance of structure prediction; (iii) **antibody developability prediction:** the large amount of human B cell receptor repertoire data from the Observed Antibody Space (OAS) database have be employed to train neural networks to evaluate the humanness of antibody sequences. Besides, DL can be applied to predict antibody solubility and viscosity, using experimental measurements, or calculated molecular descriptors (e.g., spatial charge map) as training data.

library before and after FACS enrichment, the authors identified the binding ability of a panel of antibody sequences, which were further used as input to train a CNN model for binder classification. To produce synthetic antibodies against CTLA-4 and PD-1, a generative adversarial network was also built. However, due to the largely clonally related training data, the sequences generated by this model were highly similar. Thus, sequences with novel properties beyond those found through experimental work were unable to be produced. How to break through the 'copy problem' of deep generative models to produce out-of-distribution sequences remains a challenge in this field.

As they did not provide continuous inference for affinity maturation, previous approaches to predict antibody-antigen binding were restricted to classifying binders and nonbinders. A breakthrough addressing this issue was recently made. In Saka's work, a long short-term memory (LSTM) network was employed to generate antibody sequences with improved affinity to kynurenine [110]. Impressively, the authors confirmed that the likelihood of generated sequences from this model correlated with the binding affinity well. Based on this characteristic, this model could generate antibodies with over 1800-fold higher affinity than the parental clone, outperforming the conventional frequency-based screening method. Makowski et al. trained a linear discriminant analysis (LDA) model to project binary affinity and specificity labels of variants of a clinical-stage antibody, emibetuzumab, into a continuous one-dimensional space [111]. Encouragingly, the authors found that the resulting LDA weights correlated with continuous values of antibody affinity and specificity well (Spearman's correlation coefficients of 0.87 for affinity and 0.67 for nonspecific binding), thus enabling the identification of Pareto optimal clones in a novel mutational space. In summary, with this model, the authors managed to optimize affinity and specificity simultaneously without making trade-offs between them.

3.2. AI for antibody structure prediction

Knowledge of the structure is of particular significance for the rational design of antibodies. Since experimental structure determination methods such as crystallography, NMR and cryo-EM are costly and laborious, researchers have continued to explore alternative strategies, such as computational modeling, to gain insights into antibody structure in a high-throughput manner. However, the strategies lack sufficient accuracy and reliability. Notably, recent breakthroughs in DL-based general protein structure prediction have ushered a new era, where accurately predicting the structure of most proteins is possible [112, 113]. However, the performance in the subfield problem, namely, antibody structure prediction, is less satisfactory due to the highly variable nature of antibody structures. Thus, the development of antibody-specific models is urgently needed.

Unlike five other loops adopting canonical folds, the CDR H3 loop has contributed the most to antibody diversity and presents a long-standing challenge for antibody structure prediction. To address this, Grey's team was inspired by RaptorX [114], an architecture that performed well on general protein structure prediction in CASP13, and proposed a deep residual NN [115], DeepH3 [116]. This model was trained on 1433 antibody structures selected from the SAbDab database [117] (Table 2) and took one-hot encoded, heavy and light chain variable chain concatenated sequences as input. In addition to the parameter of interresidue distances utilized by RaptorX, DeepH3 was also trained to predict dihedral and planar angles, which were demonstrated to be more effective for scoring CDR H3 loop structures than distances. The output of DeepH3 was then fed into Rosetta to *de novo* predict CDR H3 loops [118,119]. The average lower root-mean-squared deviation (RMSD) of the best-scoring structures generated with DeepH3 on the RosettaAntibody Benchmark [120] was 2.2 ± 1.1 Å, while that of the set

of five best-scoring structures for each target decreased to 1.9 ± 0.9 Å.

Encouraged by the outstanding performance of DeepH3, the authors constructed another architecture, DeepAb, to predict the whole antibody variable region from a sequence [121]. Despite the similarity in the two-stage modeling strategy, DeepAb includes two additions over the previous work. Instead of directly passing sequence information to the network, the authors pretrained an unsupervised representation learning model with 118,386 paired heavy and light chain sequences from the Observed Antibody Space (OAS) database [122] to embed general immunoglobin sequence patterns, such as evolutionary and structure information (Table 2). Another improvement is the involvement of an interpretable attention-based mechanism [123], which enabled the identification of physically important residue pairs, thus providing interpretable insights into its predictions. With both enhancements, when using DeepAb on two benchmark datasets, more accurate structure predictions were achieved [120,124] than when using previous grafting-based methods, as illustrated by the structures of two therapeutic antibodies, rituximab and sonepcizumab.

To achieve more accurate structure prediction, Grey's team investigated whether the inclusion of side-chain conformations, which are also important contributors to antigen binding [125], could enhance the performance of their prediction models. DeepSCAb was first pretrained to predict pairwise geometries, and then the output was fed into the rotamer module to predict side-chain distributions [126]. After updating the interresidue module to obtain the final interresidue outputs, the predicted structure was finally realized using Rosetta. Compared with DeepH3, DeepSCAb achieved lower cross-entropy loss on both the training and validation datasets. Regarding the decoy discrimination task, DeepSCAb outperformed DeepH3 for both the top-1 and top-5 scoring decoys, suggesting that side-chain geometry training can improve interresidue predictions. The authors also observed that the performance of DeepSCAb depends on the quality of interresidue geometry prediction. Although this method complemented existing methods for antibody structures, it was expected that it would be less important to predict side chains separately when the backbone prediction becomes more accurate.

However, it is notable that the interresidue distances and orientations used by DeepH3, DeepAb and DeepSCAb are invariant features that are unable to commute with the group action, resulting in networks without the equivariant property. Moreover, the requirement of dividing prediction into two separate processes, that is, obtaining interresidue geometries and then feeding them to Rosetta to produce final structures, largely restricted the generative speed. ABlooper is a DL architecture developed by Abanades et al. to predict antibody CDR loop structures [127]. Specifically, this network is composed of five simultaneously trained E(n)-equivariant graph NNs (E(n)-EGNNs) [128], and their output predictions were averaged to obtain a final prediction. Due to its end-to-end manner, ABlooper can be used to predict CDR backbone atoms for 100 structures in under 5 s with comparable accuracy to DeepAb. The speed of this approach makes it an ideal tool for high-throughput structure prediction, especially considering the rapidly increasing antibody repertoire sequencing data available. In addition, ABlooper can provide a quality estimate of each generated loop structure, encouraging physically plausible final predictions.

Compared with mAbs, nanobodies are a subclass of therapeutic antibodies that are advantageous in small size and cost-effective manufacturing systems [129]. However, due to longer CDR3 loops and the lack of light chains, nanobodies are more structurally flexible and require specifically designed models for accurate modeling. One such method is NanoNet, a residual convolutional network trained on structures of antibody variable heavy chains as well as nanobodies [130]. NanoNet is characterized by the direct prediction of 3D coordinates of the backbone and Cβ atoms without dividing the modeling process into framework and CDR regions. Thus, this model could make thousands of predictions in a matter of seconds with an accuracy comparable to other state-of-the-art antibody structure prediction models. Notably, the speed and accuracy of NanoNet will largely accelerate antibody-Nb docking for epitope mapping.

In addition, antibody modeling can be used in understanding the structure of antibody-antigen complexes. Davila et al. developed a pipeline, AbAdapt, that models antibody and antigen structures, predicts epitope and paratope, and combines them with rigid docking [131]. To optimize each step of the pipeline, the authors developed and trained ML models, employing 622 antibody-antigen pairs with known structures as input for cross validation. By integrating AbAdapt with more accurate protein modeling method AlphaFold2, it is encouraging to observe significant improvement in docking, paratope prediction and antibody-specific epitope prediction [132].

3.3. AI for prediction and optimization of antibody pharmaceutical properties

Another key component of therapeutic antibodies that contributes to clinical efficacy is desirable pharmaceutical properties, which include multiple aspects, such as specificity, immunogenicity, aggregation propensity, viscosity, solubility, and pharmacokinetics [133,134]. Despite the widely recognized significance of these developability parameters, it is difficult to experimentally assess them in a high-throughput manner during the early discovery stage of antibody candidates due to limitation in terms of material quality, including low concentration and low purity. Therefore, the development of alternative predictive tools to filter out candidates with unfavorable characteristics is highly needed to accelerate antibody design procedures and improve the success rate. Computational tools have been employed for therapeutic antibody design for decades, as recently systematically reviewed elsewhere [104], but their algorithms usually only take several known factors into consideration, limiting the accuracy and precision. In this section, we focus on some explorations using machine (deep) learning strategies to engineer therapeutic antibodies (Table 1).

Synthetically designed libraries are a major source of therapeutic antibodies, from which antibodies commonly require an engineering process, namely, humanization, to reduce the immunogenicity concerns of sequences and improve the safety profile. Indeed, even antibodies derived from natural repertoires often undergo further mutations to achieve a balance between affinity and biophysical properties at the risk of disrupting their 'nativeness', which refers to the similarity to naturally occurring antibodies. Therefore, to evaluate the nativeness of antibody candidates, Wollacott et al. developed a bidirectional LSTM (ABLSTM) network harnessing 25,000 antibody sequences from B-cell receptor repertoire sequencing as training data [135]. Crucially, favored by its ability to capture long-range dependencies between positions, the model surpassed other state-of-the-art methods in discriminating human antibody sequences from those originating from other species. In addition, without the process of sequence alignment, which can be time-consuming and ambiguous, the LSTM model can rapidly assess a library of 10000 antibody sequences within a few minutes.

However, one limitation of ABLSTM is that it only provides a single score of sequence humanness but without exact amino acid positions for improvement. Based on the OAS database, which is a collection of more than five hundred million human sequences, Prihoda et al. built an open-source platform BioPhi [136]. This tool is composed of two methods: OASis, which is designed for antibody humanness evaluation, and Sapiens, which is designed for antibody humanization. Specifically, OASis is used to evaluate the prevalence of each overlapping 9-mer peptide within a given antibody sequence in a peptide reference library, which was constructed from antibody repertoires in the OAS database. Based on this design, OASis is advantageous over ABLSTM in its interpretability and granularity, providing a visual report that highlights regions with the largest risk. Sapiens is a transformer NN trained based on masked language modeling (MLM) to recognize and repair masked amino acids using 20 million heavy chain sequences and 19 million light chain sequences as the training set. On the test sets of 177 humanized

antibodies, it was demonstrated that Sapiens could provide the same quality of humanization solutions as those designed by experts.

Aimed at convenient and patient-centric dosing schemes, therapeutic antibodies are often formulated at high concentrations to reduce dosage volume and relieve administration pain. However, such high concentrations, commonly exceeding 100 mg/ML, require antibodies with superior solubility and viscosity [137]. solPredict is a protein language model capable of quantitatively predicting the apparent solubility of antibodies in histidine (pH 6.0) buffer conditions [138]. This model was trained on 220 antibodies with extrapolated protein solubility data obtained from the PEG-induced precipitation method for its advantage of high throughput and minimal material needed. Specifically, this model employed embeddings extracted from a pretrained protein language model (ESM-1b transformer) as input [139] to enrich biological property information from the sequence. Such transfer learning empowered efficient learning with limited training data and a simple architecture with two fully connected hidden layers. Using 40 antibodies independent from training data, solPredict was demonstrated to correlate with experimental solubility data well (Spearman correlation coefficient = 0.86, Pearson correlation coefficient = 0.84, R^2 = 0.69, and RMSE = 4.40).

Unlike solPredict, which uses a relatively high-throughput experimental dataset and transfer learning to train its model, a CNN model aimed at predicting antibody viscosity, DeepSCM, provides novel insight into the attainment of training data [140]. DeepSCM is a surrogate model for the spatial charge map (SCM), a well-developed model accounting for the surface-exposed negative charges on the antibody variable fragment (Fv) region to predict antibody viscosity at high concentrations [141]. Despite the reliable performance of the SCM score, its application is largely hindered by difficulty in model construction and high computational cost. Therefore, by collecting 6596 antibody Fv sequences and building a homology model to run molecular dynamics simulations followed by SCM score calculation, a 1D CNN architecture was constructed to predict the viscosity of antibodies at high concentrations. A linear correlation coefficient with an SCM score of 0.9 was achieved on the test set (N = 1320), and the viscosities of 37/38 therapeutic antibodies were successfully classified at a fast speed. Further work combining structural descriptors such as SCM with DL models to predict other properties can be envisioned.

Pharmacokinetic (PK) is another key property during early discovery phases of antibody candidates, which can influence both clinical efficacy and toxicity. Despite the well-documented associations with neonatal Fc receptor (FcRn) affinity and target-mediated clearance, recent observations of unexpectedly rapid clearance of therapeutic mAbs have suggested the urgence to clarify other mechanisms influencing antibody transport and processing [142,143]. From a series of 28 in silico properties calculated from antibody structure and 12 measurements from in vitro assays, Grinshpun et al. applied a random forest classifier to identify isoelectric point and poly-specificity as the best factors to discriminate between normal and fast clearing antibodies [144]. In the future, the additional available high-quality PK data for clinical-stage antibodies would greatly improve the predictive power of ML approaches and enable the characterization of complex combinations of multiple factors, thereby promoting the development of more efficacious anti-tumor therapies.

4. AI for predicting immunotherapy response

Immunotherapy that utilizes ICIs has been shown to be an effective treatment against various types of tumors. Major issues that limit the use of cancer immunotherapy include patient selection and the prediction of treatment response. The ability to prospectively identify patients who would benefit most from a particular treatment plan can help reduce the risk of adverse clinical outcomes and high treatment costs. It is crucial to establish early on whether patients are unresponsive to ongoing treatments and whether clinicians need to modify the treatment plan accordingly. This is especially important in the case of ICI immunotherapy, where disease progression may not follow the typical pattern [145].

Predicting immunotherapy response hinges on the existence of reliable biomarkers. A more thorough understanding of the mechanism and cell types are necessary for the response and resistance [24]. Immunohistochemistry (IHC) staining for PD-L1 expression has been used in many hospital pathology departments as a predictive marker. The lack of consistency and reproducibility in the interpretation of PD-L1 by clinicians is because PD-L1 can be expressed in both tumor cells and various immune cells [146,147]. TMB and MSI are important biomarkers of immunotherapy response in different cancer types [148,149]. Digital pathology analysis and AI methods can be used to accurately obtain the staining score of PD-L1, TMB status, and MSI status, and advantages in reproducibility and diagnostic efficiency were noted.

Cancer is a complex, multifaceted disease with numerous microscopic, macroscopic, and molecular features that can influence treatment responses and patient prognosis individually or collectively. The tumor microenvironment (TME) is a critical factor in cancer progression, metastasis and response to therapy [150]. However, the detailed molecular and cellular interactions in the TME largely remain unknown. The spatial heterogeneity of the TME [151], tumor cellularity [152], and infiltrating immune cell populations, such as CD4+ and CD8+ T cells [153], contributes to how tumors evolve, metastasize, or respond to immunotherapy.

In this section, we focus on AI-based methods for predicting biomarkers and resolving the complex heterogeneity of the TME using big data such as omics and imaging data. AI has the potential to greatly improve patient selection and outcome prediction by providing detailed insights into tumor evolution and the microenvironment in a noninvasive manner. The goal of precision medicine and immunotherapy for cancer is to use patient medical data to optimize patient management and treatment and improve survival. These studies are still in the preliminary stage and require further research and validation before they can be applied clinically (Table 1).

4.1. Prediction of biomarkers associated with immunotherapy

Detecting the expression status of tumor molecular pathological markers is necessary before clinical immunotherapy. For example, the FDA approved Keytruda (pembrolizumab) as a first-line treatment for patients with MSI-high (MSI-H) metastatic colorectal tumors [154]. Higher-order genomic features, such as TMB and endogenous mutational processes (e.g., MSI and homologous recombination deficiency (HRD)), as well as large-scale features (e.g., whole-genome duplication), are clinically relevant [149,155–157]. There is a growing need for expeditious and low-cost methods of biomarker detection. Such as prediction directly from H&E-stained histopathology images, qPCR, IHC or NGS, which are methods that are already available and do not require additional tissue.

MSI testing is not always conducted on patients in clinical practice, as this requires additional genetic or immunohistochemical testing. Deep residual learning can be used to predict the MSI status of patients with gastrointestinal tumors from H&E-stained histology slides, which are widely available. Interestingly, the analysis of formalin-fixed, paraffin-embedded (FFPE) slides showed better predictive accuracy (AUC = 0.84) than snap-frozen slides (AUC = 0.77) [158]. MSI status can be predicted from ctDNA of endometrial cancer patients to inform immunotherapy-based therapy [159,160].

Based on a multicentre study of colorectal cancer, H&E WSIs contain predictive information on MSI status as a biomarker of response to ICB. In this study, the researchers collected H&E-stained slides and molecular profiling results of 8836 colorectal tumors from Germany, the Netherlands, the United Kingdom and the United States to evaluate the performance of the DL model by cross-validation and external cohort validation [161]. The DL MSI and dMMR detectors showed similar

characteristics to criterion standard tests [162], reaching clinical-grade performance. MSINet, a transfer learning model based on the MobileNetV2 architecture, was used to classify tissues, and subsequently classify MSI status in H&E-stained WSIs (40 × magnification) from a colorectal cancer cohort of 100 primary tumors (50 with microsatellite stability and 50 with MSI) from Stanford Medical Center. The authors used the H&E-stained WSIs dataset from The Cancer Genome Atlas (TCGA) to validate their model on an external cohort and compared its performance with that of five gastrointestinal pathologists (Table 2). The DL models were more accurate than the experienced gastrointestinal pathologists in predicting MSI. The authors observed a performance gap between the Stanford internal test set and the external TCGA dataset. This gap may be due to the heterogeneity of the TCGA dataset, which can be attributed to the original institution, slide preparation, and scanning procedure. The generalization performance of the model is affected by datasets from multiple institutions [163].

Some studies have employed AI to predict PD-L1 expression. A supervised ML algorithm (random forest classifier) was used to quantitatively analyze the PD-L1 score of tumor cells in PD-L1-stained digital images of melanoma. The results were highly consistent with pathologists' scoring of PD-L1 under a microscope ($r = 0.97$, $P < 0.0001$). Digital image analysis can reduce the difference between manual analysis of PD-L1 [164]. A semi supervised generative AC-GAN architecture was used to construct a model for the quantitative assessment of PD-L1 tumor cell expression scores by integrating the manual annotation and evaluation results of multiple pathologists. The model was then used to score PD-L1 expression on 270 needle biopsy specimens of advanced non-small cell lung cancer. Automated scoring of expressions showed that the model is in good agreement with human scoring [165]. Wu et al. established a tumor cell automatic identification model through a fully CNN model based on the U-ResNet structure. The tumor proportion score (TPS) of specific PD-L1 expression was output. In this study, good agreement between the TPS score results of the AI system and the trained pathologists on the PD-L1 antibody 22c3 and SP263 test sets was found (22c3: $r = 0.9429–0.9458$; SP263: $r = 0.9787$). Meanwhile, AI-assisted diagnostic tests showed significant improvements in the intragroup consistency and diagnostic efficiency of untrained general pathologists [166].

TMB is another important biomarker of the response to checkpoint immunotherapy [167]. Image2TMB, a DL model based on Inception-v3 and the random forest architecture, was used to determine TMB status from frozen H&E slides of a lung adenocarcinoma (LUAD) TCGA cohort in an attempt to replace the use of whole-exome sequencing (WES), which has high cost, operational complexity and long turnover times, for TMB gold standard determination [168]. FFPE slides, CT scans, and MR images have also been used to predict TMB in lung adenocarcinoma, non-small cell lung cancer, and low-grade glioma cohorts [169–171]. Multimodal DL involves building models that can analyze and connect information from multiple modalities. Multimodal DL that incorporated histopathological images and clinical information was used to predict TMB in colorectal cancer. Many clinical features are significantly associated with TMB status [172]. The top five clinical features are tumor stage, pathologic T, pathologic N, pathologic M and age.

4.2. Deciphering the tumor immune microenvironment

The prognostic and predictive roles of the TME in solid and hematologic tumors have been explored through the use of checkpoint inhibitors and CAR-T-cell therapy [173–177]. Higher levels of certain cytokines and chemokines, activated cytotoxic T cells, and tumor-infiltrating lymphocytes (TILs) are associated with superior clinical outcomes in large B-cell lymphoma pretreatment TME. In addition to the efficacy of CAR-T-cell therapy, pretreatment TME was also used to predict the safety of CAR-T cells. Higher intratumoral regulatory T (Treg) cell levels were associated with a lower incidence of neurotoxicity [177]. The density of CD20+ B cells and tertiary lymphoid structures in tumor tissue, as well as the ratio of tertiary lymphoid structures to tumor area, are novel predictors of the efficacy of tumor immunotherapy, particularly in its early stages [178–180].

AI can be used to predict the TME cellular composition and spatial distribution. Neural-based models provide an accurate description of the tumor immune microenvironment of solid tumors of the colon, breast, lung, and pancreas by integrating RNA-Seq and imaging data in a clinical setting. This was then evaluated model's predictions against expert pathology review [181]. ML and DL methods can be used to accurately estimate TME cellular composition using bulk transcriptome [182–184] or methylation data [185,186]. Using a decision tree ML deconvolution algorithm trained on an extensive collection of > 9400 tissue and blood-sorted cellular RNA profiles, Kassandra was able to accurately reconstruct the TME. The performance of Kassandra was validated on 4000 H&E slides and 1000 tissues by comparative cytometry, IHC, or single-cell RNA-seq. Digital TME reconstruction revealed that the presence of PD-1-positive CD8+ T cells is correlated with immunotherapy response and increases the predictive potential of established biomarkers [182]. However, using these algorithms for clinical workflows requires more rigorous benchmarks on real clinical samples. A consensus must be reached on how each cell type is defined in the field [187].

Fassler et al. employed a DL model consisting of an autoencoder (ColorAE) and a U-Net CNN to detect and classify six cell classes from histopathological images obtained from IHC of pancreatic ductal adenocarcinoma tissues. These methods can be used to quantitatively describe the spatial distribution of immune cells in the TME [188]. The spatial organization and density of TILs differed between different tumor types, tumor molecular subtypes, and immune subtypes and was associated with the immunotherapy response. The spatial patterns of TILs identified using a deep CNN model [189] are strongly correlated with the cellular composition of the tumor as assessed by CIBERSORT [190] (support vector regression model).

Tumor cells promote TME formation by creating physical barriers, inhibiting immune cells, and recruiting immunosuppressive cells [191]. Tumor cellularity (tumor purity) is an important indicator of residual disease (pathological response) after treatment. A DNN (InceptionNet architecture) was used to quantify tumor cellularity from H&E-stained WSIs (20 × magnification) of 62 breast cancer patients. The authors trained two DNN models: one to distinguish tumors from healthy tissue and the other to output a regression score of tumor cellularity (between 0 % and 100 %). The predictive scores from the study showed a high degree of agreement with the tumor cellularity reported by two independent pathologists [192]. Rakhlin et al. evaluated three powerful new DL-based methods for the automated assessment of tumor cellularity in posttreatment breast surgical specimens stained with haematoxylin and eosin [193].

Reliable predictions are produced for digital pathology using many AI approaches, but their inner workings are often opaque, leading to them being referred to as "black boxes". The lack of interpretability is a significant barrier to clinical integration. A pan-cancer analysis found that annotation-guided interpretable signatures can be used to predict the expression of four immune checkpoint proteins and homologous recombination defects. This DL approach provides an interpretable window into the composition and spatial structure of the TME. These interpretable features include simple cell (e.g., lymphocyte density in tumor tissue) and tissue quantities (e.g., area of necrotic tissue) to complex spatial features capturing tissue architecture, tissue morphology, and cell—cell proximity [194].

Single-cell transcriptomics and spatial transcriptomics offer new ways to study the TME. The use of single-cell transcriptomics provides insight into the mechanisms of cell—cell communication in the TME [195,196], identifying distinct cell types [197,198]. The application of spatial transcriptomic platforms has led to the identification of novel roles for specific types of infiltrating immune cells in cancer progression and response to therapy [199–201]. Choi et al. developed and validated

a DL model that integrates H&E images and spatial transcriptome data of lung adenocarcinoma to decipher the spatial mapping of multiple cell types in the TME. The cell types identified by the model using H&E image patches were significantly correlated with those generated by spatial transcriptomic data [202]. Given the complexity integration of H&E image and spatial transcriptomic data, DL approaches are well-suited for its analysis and interpretation. By utilizing GIST (Guiding-Image Spatial Transcriptomics), which is a conceptually novel methodology, the combination of spatial transcriptomics data with cell-type-informative paired tissue H&E stain images, such as from the reverse side of the same tissue section, enables improved inferences of tissue cell type composition in spatial transcriptomics data. Utilizing the output of AI annotated H&E tissue images, it is possible to significantly enhance the recognition of clinically meaningful immune cell infiltration in breast cancer tissue [203]. The integration of H&E images and spatial transcriptomics data can not only identify the cell types in the TME, but also enhance the accuracy of gene expression prediction. ST-Net, a DL algorithm, has been developed to predict local gene expression from H&E images, using a new dataset of 30,612 spatially resolved gene expression data matched to histopathology images from 23 patients with breast cancer [204]. Bergenstråhle et al. presented a method that combines spatial gene expression data with histological image data from the same tissue section, in order to infer higher-resolution expression maps [205].

4.3. *Predicting immunotherapy response*

The use of ICIs has greatly improved clinical care for cancer patients and has expanded to a growing list of cancer types, including melanoma and bladder and gastroesophageal cancers [206]. The accumulation of a large amount of omics data related to immunotherapy in the past decade has been facilitated by advances in sequencing technology. DL technology is especially effective in analyzing such high-dimensional data.

AI can be used to predict immunotherapy response by analyzing omics data in combination with other data (Fig. 4). Kong et al. collected clinical outcome and transcriptomic data from more than 700 ICI-treated patients. They employed protein—protein interaction network (PPI)-based ML (NetBio) to predict the response to ICI treatment in three different cancer types. The NetBio model was more accurate than predictions based on other traditional ICI therapy biomarkers, including PD-1, PD-L1, or CTLA-4, and markers associated with the TME, including CD8 T cell, T-cell exhaustion, cancer-associated fibroblast (CAF), and tumor-associated macrophage (TAM) markers [207]. Tumor genomic features have attracted interest for their potential to impact the response to immunotherapy. By using deep belief networks, Xie et al. were able to build a predictive model of immunotherapy using genomics data from multiple types of cancer. The model integrates genomic data from multiple perspectives, such as TMB, MSI, and somatic copy number variation, to classify different types of tumors into different genomic clusters. The model was applied to tumors from two melanoma immunotherapy clinical cohorts, and it was shown that different benefits from immunotherapy were observed for melanoma patients with different genomic classes [208]. The authors provided proof of principle that DL modeling may have the potential to discover intrinsic statistical correlations across modalities in multifactorial input data, which could help to understand the molecular mechanisms of primary resistance to immunotherapy.

An ensemble learning random forest ML model with 16 input features obtained through the integration of the genomic, molecular, demographic and clinical data of 1479 patients with 16 different cancer types treated with ICB from a comprehensively curated cohort (MSK-IMPACT) was used to predict ICB response. Of the patients studied, approximately 37 % had non-small cell lung cancer (NSCLC), 13 % had melanoma, and the remaining 50 % had other types of cancer, such as renal cell carcinoma and bladder, head and neck and colorectal cancer. These patients were treated with PD-1/PD-L1 inhibitors, CTLA-4 blockade or a combination of both immunotherapy agents. According to a retrospective analysis, the model had high sensitivity and specificity in predicting clinical response to immunotherapy. Furthermore, the model was also used to predict overall survival and progression-free survival on test data across different cancer types [209]. T-cell receptor (TCR) sequencing can be used to characterize cancer-specific immune responses. Sidhom et al. utilized a sequence classifier from DeepTCR, a set of previously described DL algorithms, to search for

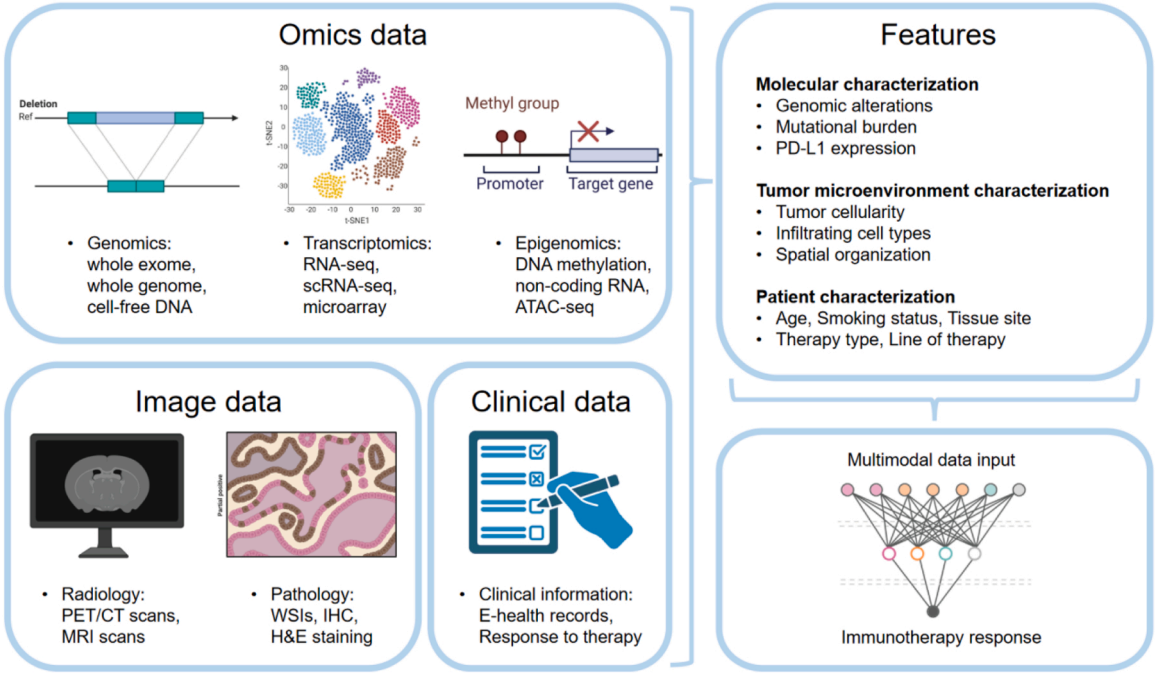

Fig. 4. Application of DL in prediction of immunotherapy response. The integration of datasets from multiple disparate modalities, including omics, images, and clinical datasets, can increase the relevant feature space of AI models, thereby enabling end-to-end immunotherapy response prediction. PET/CT, positron emission tomography/computed tomography; MRI, magnetic resonance imaging; WSIs, whole-slide images; IHC, immunohistochemistry; H&E, haematoxylin and eosin.

sequence concepts (e.g., motifs) that could predict immunotherapy response [210].

Noninvasive imaging data can also be used to predict the response to immunotherapy. A CT-derived radiological biomarker was developed and validated to differentiate immunotherapy responders from non-responders in NSCLC and melanoma patients by generating AI-based feature descriptions on pretreatment contrast-enhanced CT imaging data. In this study, it was found that morphologically more heterogeneously distributed lesions with compact borders and nonuniform density patterns were more likely to respond to immunotherapy [211]. The automatic extraction of deterministic, quantitative features is tractable as radiology data are digitized. These features are associated with clinical outcomes such as responses to ICB in pan-cancer analyses. In a retrospective multicohort study, radiomic features were found to predict clinical outcome in patients with advanced solid tumors following anti-PD-L1 or anti-PD-1 immunotherapy [212]. Johannet et al. reported an AI method that uses CNNs trained on treatment-naïve histopathology slides combined with patient clinical characteristics to predict the response to checkpoint immunotherapy in patients with advanced melanoma [213].

The application of AI to data collected during routine clinical workup of immunotherapy, particularly data from radiology and histopathology, is becoming more widespread in a substantial effort to better identify patients who are likely to respond positively to treatment [214,215]. Although various data modalities exist, they are often studied independently due to a lack of multimodal datasets and integrative algorithms. Vanguri et al. demonstrated the predictive capacity of combined medical imaging, histopathological, and genomic features to predict response to immunotherapy in a cohort of 247 patients with advanced NSCLC with multimodal baseline data obtained during diagnostic clinical workup, including CT scan images, digitized PD-L1 IHC slides and known outcomes to immunotherapy (Table 2). The authors demonstrated that a ML approach (dynamic attention with masking (DyAM)) to automatically extract discriminative features from disparate modalities has complementary and combinatorial capabilities in identifying high- and low-risk NSCLC patients undergoing immunotherapy. Multiple data modalities obtained through routine clinical diagnostic workup can be integrated to improve predictions of immunotherapy response [216] (Fig. 4).

As data scale increases and model interpretability research advances, the clinical design of immunotherapy can benefit from AI model-based recommendation systems [217]. With regard to immunotherapy, recommendation systems could learn from retrospective data to assist in future clinical decision-making for new patients on the basis of multiple patient measurements, such as a pretreatment CT scan and H&E-stained biopsy sample. All research with the potential to influence patient treatment should undergo careful evaluation sequences and be driven by protocols with a predefined statistical analysis plan that includes the utilization of AI models to direct the clinical design of immunotherapy. Evaluating DL systems for medical applications involves two primary steps: DL study and clinical trials [218]. After the AI models have been tested and verified in multi-center and external multiple cohorts, a prospective evaluation of their medical utility in randomized phase III clinical trials is necessary.

5. Conclusion

This review highlights the advances that have been made in utilizing AI to identify immunogenic neoantigens, design antibodies, and predict immunotherapy responses. These growing fields have been spurred by the exponential growth of omic-, radiological, pathological, experimental, and clinical data. This makes it possible to encode these data and use them for training AI models. AI has indisputable potential to enhance cancer immunotherapy and more broadly affect the field of cancer. With AI's success in research, the question becomes whether and when AI can be fully integrated into the clinic and become a routine practice for cancer physicians and patients. However, the application of AI in cancer immunotherapy still has some limitations, such as the insufficient amount of available data, lack of data sharing, data biases, lack of code sharing, and model interpretability. Currently, there is a trend of data bias in public datasets, with more data available from European sources than from Asian sources [219] (Table 2). Cell lines are essential tools in preclinical antibody development, and data obtained from these sources will complement experimental data obtained from patient-derived organoids [220]. In addition to data bias, there is also a gap between the ease of obtaining data from various platforms and the ease of access by external agencies for independent use, especially for private or controlled access datasets. The lack of data sharing prevents effective validation of the AI model across multiple centers. These problems can be mitigated through future clinical research, related databases, and the development of AI algorithms such as weakly supervised learning, semi-supervised learning, and active learning, which can reduce the burden of annotation [221]. Data variability is a significant challenge in implementing DL for immunotherapy, and the inconsistency of data batches and quality issues often result in unsuccessful external validation. For example, the immunohistochemical staining intensity and quality between laboratories may differ. Therefore, it is essential to establish a standardized system including inclusion and exclusion criteria, such as the use of databases with standardized disease codes and unified vocabulary in electronic health records.

AI systems need to go through multiple stages of development and evaluation in clinical trials to be used in clinics [218]. Implementing and adopting AI models to aid or enhance clinical workflows in the clinic requires addressing model uncertainty and interpretability. The AI model is an inscrutable "black box", leading clinical experts to distrust them. Model interpretability has received extensive attention, and many excellent reviews specifically focus on model interpretability [222–224]. The incorporation of prior knowledge into a transparent AI model and feature visualization are effective solutions. Technology forecasts have inherent inaccuracies. AI differs from human intelligence in several ways and being proficient in one area does not guarantee success in others. The uncertainty of the model may come from the data selection, data accuracy and completeness, inherent biases in the data, artefacts, and model specification error. AI, biologists, and clinicians should cooperate and learn from each other to better serve patients' immunotherapy treatments.

Technological advancements have allowed for a variety of methods to collect data at the individual patient level, and an ideal AI-based model for immunotherapy should include all data relevant to clinical information and biomarkers. Multimodal AI algorithms will play an important role in applications such as immunotherapy. A thorough understanding of learned models from both biological and clinical perspectives is essential for investigators who want to implement multimodal AI methods rationally. Depending on the goals of the study, understanding a model can arguably be as important as increasing its predictive power. The long-term prospects for this field of research are very promising, as immunotherapy, like other fields of biology, is undergoing a transformation that is merging with computational and data science. Therefore, it is expected that ML and DL will make new progress in applications such as neoantigen recognition, antibody design, and immunotherapy response prediction.

CRediT authorship contribution statement

TL, YP.L and XY.Z wrote the manuscript draft and made figures and tables. All authors wrote, read, and approved the manuscript.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

Data Availability

No data was used for the research described in the article.

Acknowledgements

We thank the Center for Precision Medicine at Sun Yat-sen University for the long-term support.

Funding

This project was supported by National Key R&D Program of China (2019YFA0904400, Z.X. and T.Y.) and the National Natural Science Foundation of China (32270984, Y.W.).

References

[1] M. Dougan, G. Dranoff, S.K. Dougan, Cancer immunotherapy: beyond checkpoint blockade, Annu. Rev. Cancer Biol. 3 (2019) 55–75, https://doi.org/10.1146/annurev-cancerbio-030518-055552.
[2] S. Tan, D. Li, X. Zhu, Cancer immunotherapy: pros, cons and beyond, Biomed. Pharmacother. 124 (2020), 109821, https://doi.org/10.1016/j.biopha.2020.109821.
[3] A.D. Waldman, J.M. Fritz, M.J. Lenardo, A guide to cancer immunotherapy: from T cell basic science to clinical practice, Nat. Rev. Immunol. 20 (2020) 651–668, https://doi.org/10.1038/s41577-020-0306-5.
[4] S.A. Rosenberg, N.P. Restifo, Adoptive cell transfer as personalized immunotherapy for human cancer, Science 348 (2015) 62–68, https://doi.org/10.1126/science.aaa4967.
[5] M. Sadelain, R. Brentjens, I. Rivière, The basic principles of chimeric antigen receptor designmaking better chimeric antigen receptors, Cancer Discov. 3 (2013) 388–398, https://doi.org/10.1158/2159-8290.cd-12-0548.
[6] M. Luksza, N. Riaz, V. Makarov, V.P. Balachandran, M.D. Hellmann, A. Solovyov, N.A. Rizvi, T. Merghoub, A.J. Levine, T.A. Chan, et al., A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy, Nature 551 (2017) 517–520, https://doi.org/10.1038/nature24473.
[7] M. Yarchoan, B.A. Johnson, E.R. Lutz, D.A. Laheru, E.M. Jaffee, Targeting neoantigens to augment antitumour immunity, Nat. Rev. Cancer 17 (2017) 209–222, https://doi.org/10.1038/nrc.2017.74.
[8] A.-M. Bjerregaard, M. Nielsen, S.R. Hadrup, Z. Szallasi, A.C. Eklund, MuPeXI: prediction of neo-epitopes from tumor sequencing data, Cancer Immunol. Immunother. 66 (2017) 1123–1130, https://doi.org/10.1007/s00262-017-2001-3.
[9] C.C. Smith, S.R. Selitsky, S. Chai, P.M. Armistead, B.G. Vincent, J.S. Serody, Alternative tumour-specific antigens, Nat. Rev. Cancer 19 (2019) 465–478, https://doi.org/10.1038/s41568-019-0162-4.
[10] C.H. Lee, M. Salio, G. Napolitani, G. Ogg, A. Simmons, H. Koohy, Predicting cross-reactivity and antigen specificity of T cell receptors, Front. Immunol. 11 (2020), 565096, https://doi.org/10.3389/fimmu.2020.565096.
[11] Y. Pan, K.E. Kadash-Edmondson, R. Wang, J. Phillips, S. Liu, A. Ribas, R. Aplenc, O.N. Witte, Y. Xing, RNA dysregulation: an expanding source of cancer immunotherapy targets, Trends Pharm. Sci. 42 (2021) 268–282, https://doi.org/10.1016/j.tips.2021.01.006.
[12] L. Chen, Y. Zhang, Y. Yang, Y. Yang, H. Li, X. Dong, H. Wang, Z. Xie, Q. Zhao, An integrated approach for discovering noncanonical MHC-I peptides encoded by small open reading frames, J. Am. Soc. Mass Spectrom. 32 (2021) 2346–2357, https://doi.org/10.1021/jasms.1c00076.
[13] P.J. Carter, A. Rajpal, Designing antibodies as therapeutics, Cell 185 (2022) 2789–2805, https://doi.org/10.1016/j.cell.2022.05.029.
[14] T. Jain, T. Sun, S. Durand, A. Hall, N.R. Houston, J.H. Nett, B. Sharkey, B. Bobrowicz, I. Caffry, Y. Yu, et al., Biophysical properties of the clinical-stage antibody landscape, Proc. Natl. Acad. Sci. USA 114 (2017) 944–949, https://doi.org/10.1073/pnas.1616408114.
[15] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 23 (1997) 3–25, https://doi.org/10.1016/s0169-409x(00)00129-0.
[16] W. Wilman, S. Wróbel, W. Bielska, P. Deszynski, P. Dudzic, I. Jaszczyszyn, J. Kaniewski, J. M\lokosiewicz, A. Rouyan, T. Sat\lawa, et al., Machine-designed biotherapeutics: opportunities, feasibility and advantages of deep learning in computational antibody discovery, Brief. Bioinform. 23 (2022), bbac267, https://doi.org/10.1093/bib/bbac267.
[17] M.A. Alfaleh, H.O. Alsaab, A.B. Mahmoud, A.A. Alkayyal, M.L. Jones, S. M. Mahler, A.M. Hashem, Phage display derived monoclonal antibodies: from bench to bedside, Front. Immunol. 11 (2020), 1986, https://doi.org/10.3389/fimmu.2020.01986.
[18] R. Akbar, H. Bashour, P. Rawat, P.A. Robert, E. Smorodina, T.-S. Cotet, K. Flem-Karlsen, R. Frank, B.B. Mehta, M.H. Vu, et al., Progress and challenges for the machine learning-based design of fit-for-purpose monoclonal antibodies, in: Mabs, Taylor & Francis, 2022, p. 2008790. ⟨https://doi.org/10.1080/19420862.2021.2008790⟩).
[19] N. Doneva, I. Doytchinova, I. Dimitrov, Predicting immunogenicity risk in biopharmaceuticals, Symmetry 13 (2021) 388, https://doi.org/10.3390/sym13030388.
[20] C.Y. Kang, S.E. Duarte, H.S. Kim, E. Kim, J. Park, A.D. Lee, Y. Kim, L. Kim, S. Cho, Y. Oh, et al., Artificial intelligence-based radiomics in the era of immuno-oncology, Oncologist 27 (2022) e471–e483, https://doi.org/10.1093/oncolo/oyac036.
[21] S.L. Topalian, J.M. Taube, R.A. Anders, D.M. Pardoll, Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy, Nat. Rev. Cancer 16 (2016) 275–287, https://doi.org/10.1038/nrc.2016.36.
[22] N.A. Saunders, F. Simpson, E.W. Thompson, M.M. Hill, L. Endo-Munoz, G. Leggatt, R.F. Minchin, A. Guminski, Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives, EMBO Mol. Med. 4 (2012) 675–684, https://doi.org/10.1002/emmm.201101131.
[23] A. Kalbasi, A. Ribas, Tumour-intrinsic resistance to immune checkpoint blockade, Nat. Rev. Immunol. 20 (2020) 25–39, https://doi.org/10.1038/s41577-019-0218-4.
[24] J.J. Havel, D. Chowell, T.A. Chan, The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy, Nat. Rev. Cancer 19 (2019) 133–150, https://doi.org/10.1038/s41568-019-0116-x.
[25] Y. Wang, S. Zhou, F. Yang, X. Qi, X. Wang, X. Guan, C. Shen, N. Duma, J. V. Aguilera, A. Chintakuntlawar, et al., Treatment-related adverse events of PD-1 and PD-L1 inhibitors in clinical trials: a systematic review and meta-analysis, JAMA Oncol. 5 (2019) 1008–1019, https://doi.org/10.1001/jamaoncol.2019.0393.
[26] Q. Wang, R. Xu, Immunotherapy-related adverse events (irAEs): extraction from FDA drug labels and comparative analysis, JAMIA Open 2 (2019) 173–178, https://doi.org/10.1093/jamiaopen/ooy045.
[27] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444, https://doi.org/10.1038/nature14539.
[28] C. Cao, F. Liu, H. Tan, D. Song, W. Shu, W. Li, Y. Zhou, X. Bo, Z. Xie, Deep learning and its applications in biomedicine, Genom. Proteom. Bioinform. 16 (2018) 17–32, https://doi.org/10.1016/j.gpb.2017.07.003.
[29] B. Bhinder, C. Gilvary, N.S. Madhukar, O. Elemento, Artificial intelligence in cancer research and precision medicine, Cancer Discov. 11 (2021) 900–915, https://doi.org/10.1158/2159-8290.cd-21-0090.
[30] Y. Deng, M. Feng, Y. Jiang, Y. Zhou, H. Qin, F. Xiang, Y. Wang, H. Bu, J. Bao, Development of pathological reconstructed high-resolution images using artificial intelligence based on whole slide image, MedComm 1 (2020) 410–417, https://doi.org/10.1002/mco2.39.
[31] W.J. Lesterhuis, J.B.A.G. Haanen, C.J.A. Punt, Cancer immunotherapy–revisited, Nat. Rev. Drug Discov. 10 (2011) 591–600, https://doi.org/10.1038/nrd3500.
[32] F. Chen, Z. Zou, J. Du, S. Su, J. Shao, F. Meng, J. Yang, Q. Xu, N. Ding, Y. Yang, Q. Liu, Q. Wang, Z. Sun, S. Zhou, S. Du, J. Wei, B. Liu, Neoantigen identification strategies enable personalized immunotherapy in refractory solid tumors, J. Clin. Invest. 129 (2019) 2056–2070, https://doi.org/10.1172/JCI99538.
[33] M.M. Richters, H. Xia, K.M. Campbell, W.E. Gillanders, O.L. Griffith, M. Griffith, Best practices for bioinformatic characterization of neoantigens for clinical utility, Genome Med. 11 (2019) 56, https://doi.org/10.1186/s13073-019-0666-2.
[34] E. Tran, M. Ahmadzadeh, Y.-C. Lu, A. Gros, S. Turcotte, P.F. Robbins, J.J. Gartner, Z. Zheng, Y.F. Li, S. Ray, J.R. Wunderlich, R.P. Somerville, S.A. Rosenberg, Immunogenicity of somatic mutations in human gastrointestinal cancers, Science 350 (2015) 1387–1390, https://doi.org/10.1126/science.aad1253.
[35] M.R. Parkhurst, P.F. Robbins, E. Tran, T.D. Prickett, J.J. Gartner, L. Jia, G. Ivey, Y.F. Li, M. El-Gamil, A. Lalani, J.S. Crystal, A. Sachs, E. Groh, S. Ray, L.T. Ngo, S. Kivitz, A. Pasetto, R. Yossef, F.J. Lowery, S.L. Goff, W. Lo, G. Cafri, D. C. Deniger, P. Malekzadeh, M. Ahmadzadeh, J.R. Wunderlich, R.P.T. Somerville, S.A. Rosenberg, Unique neoantigens arise from somatic mutations in patients with gastrointestinal cancers, Cancer Discov. 9 (2019) 1022–1035, https://doi.org/10.1158/2159-8290.CD-18-1494.
[36] V. Jurtz, S. Paul, M. Andreatta, P. Marcatili, B. Peters, M. Nielsen, NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data, J. Immunol. 199 (2017) 3360–3368, https://doi.org/10.4049/jimmunol.1700893.
[37] D. Gfeller, P. Guillaume, J. Michaux, H.-S. Pak, R.T. Daniel, J. Racle, G. Coukos, M. Bassani-Sternberg, The length distribution and multiple specificity of naturally presented HLA-I ligands, J. Immunol. 201 (2018) 3705–3716, https://doi.org/10.4049/jimmunol.1800914.
[38] B. Bulik-Sullivan, J. Busby, C.D. Palmer, M.J. Davis, T. Murphy, A. Clark, M. Busby, F. Duke, A. Yang, L. Young, N.C. Ojo, K. Caldwell, J. Abhyankar, T. Boucher, M.G. Hart, V. Makarov, V.T.D. Montpreville, O. Mercier, T.A. Chan, G. Scagliotti, P. Bironzo, S. Novello, N. Karachaliou, R. Rosell, I. Anderson, N. Gabrail, J. Hrom, C. Limvarapuss, K. Choquette, A. Spira, R. Rousseau, C. Voong, N.A. Rizvi, E. Fadel, M. Frattini, K. Jooss, M. Skoberne, J. Francis, R. Yelensky, Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification, Nat. Biotechnol. (2018), https://doi.org/10.1038/nbt.4313.
[39] D.K. Wells, M.M. van Buuren, K.K. Dang, V.M. Hubbard-Lucey, K.C.F. Sheehan, K. M. Campbell, A. Lamb, J.P. Ward, J. Sidney, A.B. Blazquez, A.J. Rech, J. M. Zaretsky, B. Comin-Anduix, A.H.C. Ng, W. Chour, T.V. Yu, H. Rizvi, J.M. Chen, P. Manning, G.M. Steiner, X.C. Doan, , Tumor Neoantigen Selection Alliance, T. Merghoub, J. Guinney, A. Kolom, C. Selinsky, A. Ribas, M.D. Hellmann, N. Hacohen, A. Sette, J.R. Heath, N. Bhardwaj, F. Ramsdell, R.D. Schreiber, T. N. Schumacher, P. Kvistborg, N.A. Defranoux, Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen

prediction, Cell 183 (2020) 818–834.e13, https://doi.org/10.1016/j.cell.2020.09.015.
[40] D. Bzdok, N. Altman, M. Krzywinski, Statistics versus machine learning, Nat. Methods 15 (2018) 233–234, https://doi.org/10.1038/nmeth.4642.
[41] M. Nielsen, C. Lundegaard, P. Worning, S.L. Lauemøller, K. Lamberth, S. Buus, S. Brunak, O. Lund, Reliable prediction of T-cell epitopes using neural networks with novel sequence representations, Protein Sci. 12 (2003) 1007–1017, https://doi.org/10.1110/ps.0239403.
[42] C. Lundegaard, O. Lund, M. Nielsen, Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers, Bioinformatics 24 (2008) 1397–1398, https://doi.org/10.1093/bioinformatics/btn128.
[43] B. Peters, H.-H. Bui, S. Frankild, M. Nielson, C. Lundegaard, E. Kostem, D. Basch, K. Lamberth, M. Harndahl, W. Fleri, S.S. Wilson, J. Sidney, O. Lund, S. Buus, A. Sette, A community resource benchmarking predictions of peptide binding to MHC-I molecules, PLoS Comput. Biol. 2 (2006), e65, https://doi.org/10.1371/journal.pcbi.0020065.
[44] S. Mei, F. Li, A. Leier, T.T. Marquez-Lago, K. Giam, N.P. Croft, T. Akutsu, A. I. Smith, J. Li, J. Rossjohn, A.W. Purcell, J. Song, A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction, Brief. Bioinform. 21 (2020) 1119–1135, https://doi.org/10.1093/bib/bbz051.
[45] R. Vita, J.A. Overton, J.A. Greenbaum, J. Ponomarenko, J.D. Clark, J.R. Cantrell, D.K. Wheeler, J.L. Gabbard, D. Hix, A. Sette, B. Peters, The immune epitope database (IEDB) 3.0, Nucleic Acids Res. 43 (2015) D405–412, https://doi.org/10.1093/nar/gku938.
[46] W. Zhao, X. Sher, Systematically benchmarking peptide-MHC binding predictors: from synthetic to naturally processed epitopes, PLoS Comput. Biol. 14 (2018), e1006457, https://doi.org/10.1371/journal.pcbi.1006457.
[47] M. Andreatta, M. Nielsen, Gapped sequence alignment using artificial neural networks: application to the MHC class I system, Bioinformatics 32 (2016) 511–517, https://doi.org/10.1093/bioinformatics/btv639.
[48] M. Nielsen, M. Andreatta, NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets, Genome Med. 8 (2016) 33, https://doi.org/10.1186/s13073-016-0288-x.
[49] The problem with neoantigen prediction, Nat. Biotechnol. 35 (2017) 97, https://doi.org/10.1038/nbt.3800.
[50] T.J. O'Donnell, A. Rubinsteyn, M. Bonsack, A.B. Riemer, U. Laserson, J. Hammerbacher, MHCflurry: open-source class I MHC binding affinity prediction, Cell Syst. 7 (2018) 129–132.e4, https://doi.org/10.1016/j.cels.2018.05.014.
[51] T.J. O'Donnell, A. Rubinsteyn, U. Laserson, MHCflurry 2.0: improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing, Cell Syst. 11 (2020) 42–48.e7, https://doi.org/10.1016/j.cels.2020.06.010.
[52] E. Strønen, M. Toebes, S. Kelderman, M.M. van Buuren, W. Yang, N. van Rooij, M. Donia, M.-L. Böschen, F. Lund-Johansen, J. Olweus, T.N. Schumacher, Targeting of cancer neoantigens with donor-derived T cell receptor repertoires, Science 352 (2016) 1337–1341, https://doi.org/10.1126/science.aaf2288.
[53] M. Bassani-Sternberg, E. Bräunlein, R. Klar, T. Engleitner, P. Sinitcyn, S. Audehm, M. Straub, J. Weber, J. Slotta-Huspenina, K. Specht, Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry, Nat. Commun. 7 (1) (2016) 16.
[54] B. Chen, M.S. Khodadoust, N. Olsson, L.E. Wagar, E. Fast, C.L. Liu, Y. Muftuoglu, B.J. Sworder, M. Diehn, R. Levy, M.M. Davis, J.E. Elias, R.B. Altman, A. A. Alizadeh, Predicting HLA class II antigen presentation through integrated deep learning, Nat. Biotechnol. 37 (2019) 1332–1343, https://doi.org/10.1038/s41587-019-0280-2.
[55] M. Bassani-Sternberg, D. Gfeller, Unsupervised HLA peptidome deconvolution improves ligand prediction accuracy and predicts cooperative effects in peptide-HLA interactions, J. Immunol. 197 (2016) 2492–2499, https://doi.org/10.4049/jimmunol.1600808.
[56] M. Rasmussen, M. Harndahl, A. Stryhn, R. Boucherma, L.L. Nielsen, F. A. Lemonnier, M. Nielsen, S. Buus, Uncovering the peptide-binding specificities of HLA-C: a general strategy to determine the specificity of any MHC class I molecule, J. Immunol. 193 (2014) 4790–4802, https://doi.org/10.4049/jimmunol.1401689.
[57] H. Rammensee, J. Bachmann, N.P. Emmerich, O.A. Bachor, S. Stevanović, SYFPEITHI: database for MHC ligands and peptide motifs, Immunogenetics 50 (1999) 213–219, https://doi.org/10.1007/s002510050595.
[58] P.A. Reche, E.L. Reinherz, Prediction of peptide-MHC binding using profiles, Methods Mol. Biol. 409 (2007) 185–200, https://doi.org/10.1007/978-1-60327-118-9_13.
[59] M. Pertseva, B. Gao, D. Neumeier, A. Yermanos, S.T. Reddy, Applications of machine and deep learning in adaptive immunity, Annu. Rev. Chem. Biomol. Eng. 12 (2021) 39–62, https://doi.org/10.1146/annurev-chembioeng-101420-125021.
[60] P. Phloyphisut, N. Pornputtapong, S. Sriswasdi, E. Chuangsuwanich, MHCSeqNet: a deep neural network model for universal MHC binding prediction, BMC Bioinform. 20 (2019) 270, https://doi.org/10.1186/s12859-019-2892-4.
[61] J.G. Abelin, D. Harjanto, M. Malloy, P. Suri, T. Colson, S.P. Goulding, A.L. Creech, L.R. Serrano, G. Nasir, Y. Nasrullah, C.D. McGann, D. Velez, Y.S. Ting, A. Poran, D.A. Rothenberg, S. Chhangawala, A. Rubinsteyn, J. Hammerbacher, R. B. Gaynor, E.F. Fritsch, J. Greshock, R.C. Oslund, D. Barthelme, T.A. Addona, C. M. Arieta, M.S. Rooney, Defining HLA-II ligand processing and binding rules with mass spectrometry enhances cancer epitope prediction, Immunity 51 (2019) 766–779.e17, https://doi.org/10.1016/j.immuni.2019.08.012.
[62] Y. Lecun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L. D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Comput. 1 (1989) 541–551.
[63] J.J. Gartner, M.R. Parkhurst, A. Gros, E. Tran, M.S. Jafferji, A. Copeland, K.-I. Hanada, N. Zacharakis, A. Lalani, S. Krishna, A. Sachs, T.D. Prickett, Y.F. Li, M. Florentin, S. Kivitz, S.C. Chatmon, S.A. Rosenberg, P.F. Robbins, A machine learning model for ranking candidate HLA class I neoantigens based on known neoepitopes from multiple human tumor types, Nat. Cancer 2 (2021) 563–574, https://doi.org/10.1038/s43018-021-00197-6.
[64] L. Lybaert, S. Lefever, B. Fant, E. Smits, B. De Geest, K. Breckpot, L. Dirix, S. A. Feldman, W. van Criekinge, K. Thielemans, S.H. van der Burg, P.A. Ott, C. Bogaert, Challenges in neoantigen-directed therapeutics, Cancer Cell 41 (2023) 15–40, https://doi.org/10.1016/j.ccell.2022.10.013.
[65] T. Blankenstein, P.G. Coulie, E. Gilboa, E.M. Jaffee, The determinants of tumour immunogenicity, Nat. Rev. Cancer 12 (2012) 307–313, https://doi.org/10.1038/nrc3246.
[66] C. Linnemann, M.M. van Buuren, L. Bies, E.M.E. Verdegaal, R. Schotte, J.J. A. Calis, S. Behjati, A. Velds, H. Hilkmann, D.E. Atmioui, M. Visser, M.R. Stratton, J.B.A.G. Haanen, H. Spits, S.H. van der Burg, T.N.M. Schumacher, High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma, Nat. Med. 21 (2015) 81–85, https://doi.org/10.1038/nm.3773.
[67] M. Rasmussen, E. Fenoy, M. Harndahl, A.B. Kristensen, I.K. Nielsen, M. Nielsen, S. Buus, Pan-specific prediction of peptide-MHC class I complex stability, a correlate of T cell immunogenicity, J. Immunol. 197 (2016) 1517–1524, https://doi.org/10.4049/jimmunol.1600582.
[68] K.W. Jørgensen, M. Rasmussen, S. Buus, M. Nielsen, NetMHCstab – predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery, Immunology 141 (2014) 18–26, https://doi.org/10.1111/imm.12160.
[69] J. Hennecke, D.C. Wiley, T cell receptor-MHC interactions up close, Cell 104 (2001) 1–4, https://doi.org/10.1016/s0092-8674(01)00185-4.
[70] C. Szeto, C.A. Lobos, A.T. Nguyen, S. Gras, TCR recognition of peptide-MHC-I: rule makers and breakers, Int. J. Mol. Sci. 22 (2020) 68, https://doi.org/10.3390/ijms22010068.
[71] S. Bobisse, R. Genolet, A. Roberti, J.L. Tanyi, J. Racle, B.J. Stevenson, C. Iseli, A. Michel, M.-A. Le Bitoux, P. Guillaume, J. Schmidt, V. Bianchi, D. Dangaj, C. Fenwick, L. Derré, I. Xenarios, O. Michielin, P. Romero, D.S. Monos, V. Zoete, D. Gfeller, L.E. Kandalaft, G. Coukos, A. Harari, Sensitive and frequent identification of high avidity neo-epitope specific CD8 + T cells in immunotherapy-naive ovarian cancer, Nat. Commun. 9 (2018) 1092, https://doi.org/10.1038/s41467-018-03301-0.
[72] G. Cafri, R. Yossef, A. Pasetto, D.C. Deniger, Y.-C. Lu, M. Parkhurst, J.J. Gartner, L. Jia, S. Ray, L.T. Ngo, M. Jafferji, A. Sachs, T. Prickett, P.F. Robbins, S. A. Rosenberg, Memory T cells targeting oncogenic mutations detected in peripheral blood of epithelial cancer patients, Nat. Commun. 10 (2019) 449, https://doi.org/10.1038/s41467-019-08304-z.
[73] Q. Leng, M. Tarbe, Q. Long, F. Wang, Pre-existing heterologous T-cell immunity and neoantigen immunogenicity, Clin. Transl. Immunol. 9 (2020), e01111, https://doi.org/10.1002/cti2.1111.
[74] Z. Hu, D.E. Leet, R.L. Allesøe, G. Oliveira, S. Li, A.M. Luoma, J. Liu, J. Forman, T. Huang, J.B. Iorgulescu, R. Holden, S. Sarkizova, S.H. Gohil, R.A. Redd, J. Sun, L. Elagina, A. Giobbie-Hurder, W. Zhang, L. Peter, Z. Ciantra, S. Rodig, O. Olive, K. Shetty, J. Pyrdol, M. Uduman, P.C. Lee, P. Bachireddy, E.I. Buchbinder, C. H. Yoon, D. Neuberg, B.L. Pentelute, N. Hacohen, K.J. Livak, S.A. Shukla, L. R. Olsen, D.H. Barouch, K.W. Wucherpfennig, E.F. Fritsch, D.B. Keskin, C.J. Wu, P.A. Ott, Personal neoantigen vaccines induce persistent memory T cell responses and epitope spreading in patients with melanoma, Nat. Med. 27 (2021) 515–525, https://doi.org/10.1038/s41591-020-01206-4.
[75] H. Zhang, L. Liu, J. Zhang, J. Chen, J. Ye, S. Shukla, J. Qiao, X. Zhan, H. Chen, C. J. Wu, Y.-X. Fu, B. Li, Investigation of antigen-specific T-cell receptor clusters in human cancers, Clin. Cancer Res. 26 (2020) 1359–1371, https://doi.org/10.1158/1078-0432.CCR-19-3249.
[76] J.-W. Sidhom, H.B. Larman, D.M. Pardoll, A.S. Baras, DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires, Nat. Commun. 12 (2021) 1605, https://doi.org/10.1038/s41467-021-21879-w.
[77] C.M. Laumont, A.C. Banville, M. Gilardi, D.P. Hollern, B.H. Nelson, Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities, Nat. Rev. Cancer 22 (2022) 414–430, https://doi.org/10.1038/s41568-022-00466-1.
[78] P. Zaenker, E.S. Gray, M.R. Ziman, Autoantibody production in cancer–the humoral immune response toward autologous antigens in cancer patients, Autoimmun. Rev. 15 (2016) 477–483, https://doi.org/10.1016/j.autrev.2016.01.017.
[79] C. Cui, J. Wang, E. Fagerberg, P.-M. Chen, K.A. Connolly, M. Damo, J.F. Cheung, T. Mao, A.S. Askari, S. Chen, B. Fitzgerald, G.G. Foster, S.C. Eisenbarth, H. Zhao, J. Craft, N.S. Joshi, Neoantigen-driven B cell and CD4 T follicular helper cell collaboration promotes anti-tumor CD8 T cell responses, Cell 184 (2021) 6101–6118.e13, https://doi.org/10.1016/j.cell.2021.11.007.
[80] M.C. Jespersen, S. Mahajan, B. Peters, M. Nielsen, P. Marcatili, Antibody specific B-cell epitope predictions: leveraging information from antibody-antigen protein complexes, Front. Immunol. 10 (2019), 298, https://doi.org/10.3389/fimmu.2019.00298.

[81] A. Ras-Carmona, A.A. Lehmann, P.V. Lehmann, P.A. Reche, Prediction of B cell epitopes in proteins using a novel sequence similarity-based method, Sci. Rep. 12 (2022) 13739, https://doi.org/10.1038/s41598-022-18021-1.
[82] M. Collatz, F. Mock, E. Barth, M. Hölzer, K. Sachse, M. Marz, EpiDope: a deep neural network for linear B-cell epitope prediction, Bioinformatics 37 (2021) 448–455, https://doi.org/10.1093/bioinformatics/btaa773.
[83] B.M. da Silva, Y. Myung, D.B. Ascher, D.E.V. Pires, epitope3D: a machine learning method for conformational B-cell epitope prediction, Brief. Bioinform. 23 (2022), bbab423, https://doi.org/10.1093/bib/bbab423.
[84] G. Cia, F. Pucci, M. Rooman, Critical review of conformational B-cell epitope prediction methods, Brief. Bioinform. 24 (2023), bbac567, https://doi.org/10.1093/bib/bbac567.
[85] E. Blass, P.A. Ott, Advances in the development of personalized neoantigen-based therapeutic cancer vaccines, Nat. Rev. Clin. Oncol. 18 (2021) 215–229, https://doi.org/10.1038/s41571-020-00460-2.
[86] F. Lang, B. Schrörs, M. Löwer, Ö. Türeci, U. Sahin, Identification of neoantigens for individualized therapeutic cancer vaccines, Nat. Rev. Drug Discov. 21 (2022) 261–282, https://doi.org/10.1038/s41573-021-00387-y.
[87] J.R. Devlin, J.A. Alonso, C.M. Ayres, G.L.J. Keller, S. Bobisse, C.W. Vander Kooi, G. Coukos, D. Gfeller, A. Harari, B.M. Baker, Structural dissimilarity from self drives neoepitope escape from immune tolerance, Nat. Chem. Biol. 16 (2020) 1269–1276, https://doi.org/10.1038/s41589-020-0610-1.
[88] J. Graves, J. Byerly, E. Priego, N. Makkapati, S.V. Parish, B. Medellin, M. Berrondo, A review of deep learning methods for antibodies, Antibodies 9 (2020) 12, https://doi.org/10.3390/antib9020012.
[89] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.
[90] S. Turajlic, K. Litchfield, H. Xu, R. Rosenthal, N. McGranahan, J.L. Reading, Y.N. S. Wong, A. Rowan, N. Kanu, M. Al Bakir, T. Chambers, R. Salgado, P. Savas, S. Loi, N.J. Birkbak, L. Sansregret, M. Gore, J. Larkin, S.A. Quezada, C. Swanton, Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis, Lancet Oncol. 18 (2017) 1009–1021, https://doi.org/10.1016/S1470-2045(17)30516-8.
[91] W. Yang, K.-W. Lee, R.M. Srivastava, F. Kuo, C. Krishna, D. Chowell, V. Makarov, D. Hoen, M.G. Dalin, L. Wexler, R. Ghossein, N. Katabi, Z. Nadeem, M.A. Cohen, S.K. Tian, N. Robine, K. Arora, H. Geiger, P. Agius, N. Bouvier, K. Huberman, K. Vanness, J.J. Havel, J.S. Sims, R.M. Samstein, R. Mandal, J. Tepe, I. Ganly, A. L. Ho, N. Riaz, R.J. Wong, N. Shukla, T.A. Chan, L.G.T. Morris, Immunogenic neoantigens derived from gene fusions stimulate T cell responses, Nat. Med. 25 (2019) 767–775, https://doi.org/10.1038/s41591-019-0434-2.
[92] K. Cibulskis, M.S. Lawrence, S.L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E.S. Lander, G. Getz, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples, Nat. Biotechnol. 31 (2013) 213–219, https://doi.org/10.1038/nbt.2514.
[93] C. Keşmir, A.K. Nussbaum, H. Schild, V. Detours, S. Brunak, Prediction of proteasome cleavage motifs by neural networks, Protein Eng. 15 (2002) 287–296, https://doi.org/10.1093/protein/15.4.287.
[94] T. Stranzl, M.V. Larsen, C. Lundegaard, M. Nielsen, NetCTLpan: pan-specific MHC class I pathway epitope predictions, Immunogenetics 62 (2010) 357–368, https://doi.org/10.1007/s00251-010-0441-4.
[95] B. Wen, K. Li, Y. Zhang, B. Zhang, Cancer neoantigen prioritization through sensitive and reliable proteogenomics analysis, Nat. Commun. 11 (2020) 1759, https://doi.org/10.1038/s41467-020-15456-w.
[96] J. Schmidt, A.R. Smith, M. Magnin, J. Racle, J.R. Devlin, S. Bobisse, J. Cesbron, V. Bonnet, S.J. Carmona, F. Huber, G. Ciriello, D.E. Speiser, M. Bassani-Sternberg, G. Coukos, B.M. Baker, A. Harari, D. Gfeller, Prediction of neo-epitope immunogenicity reveals TCR recognition determinants and provides insight into immunoediting, Cell Rep. Med. 2 (2021), 100194, https://doi.org/10.1016/j.xcrm.2021.100194.
[97] C. Zhou, Z. Wei, Z. Zhang, B. Zhang, C. Zhu, K. Chen, G. Chuai, S. Qu, L. Xie, Y. Gao, Q. Liu, pTuneos: prioritizing tumor neoantigens from next-generation sequencing data, Genome Med. 11 (2019) 67, https://doi.org/10.1186/s13073-019-0679-x.
[98] J. Hundal, S. Kiwala, J. McMichael, C.A. Miller, H. Xia, A.T. Wollam, C.J. Liu, S. Zhao, Y.-Y. Feng, A.P. Graubert, A.Z. Wollam, J. Neichin, M. Neveau, J. Walker, W.E. Gillanders, E.R. Mardis, O.L. Griffith, M. Griffith, pVACtools: a computational toolkit to identify and visualize cancer neoantigens, Cancer Immunol. Res. 8 (2020) 409–420, https://doi.org/10.1158/2326-6066.CIR-19-0401.
[99] R. Vita, S. Mahajan, J.A. Overton, S.K. Dhanda, S. Martini, J.R. Cantrell, D. K. Wheeler, A. Sette, B. Peters, The immune epitope database (IEDB): 2018 update, Nucleic Acids Res. 47 (2019) D339–D343, https://doi.org/10.1093/nar/gky1006.
[100] A.J. Grillo-López, Rituximab (Rituxan®/MabThera®): the first decade (1993–2003), Expert Rev. Anticancer Ther. 3 (2003) 767–779, https://doi.org/10.1586/14737140.3.6.767.
[101] H. Kaplon, A. Chenoweth, S. Crescioli, J.M. Reichert, Antibodies to watch in 2022, MAbs 14 (2022), 2014296, https://doi.org/10.1080/19420862.2021.2014296.
[102] A.M. Hummer, B. Abanades, C.M. Deane, Advances in computational structure-based antibody design, Curr. Opin. Struct. Biol. 74 (2022), 102379, https://doi.org/10.1016/j.sbi.2022.102379.
[103] R. Akbar, H. Bashour, P. Rawat, P.A. Robert, E. Smorodina, T.-S. Cotet, K. Flem-Karlsen, R. Frank, B.B. Mehta, M.H. Vu, T. Zengin, J. Gutierrez-Marcos, F. Lund-Johansen, J.T. Andersen, V. Greiff, Progress and challenges for the machine learning-based design of fit-for-purpose monoclonal antibodies, MAbs 14 (2022), 2008790, https://doi.org/10.1080/19420862.2021.2008790.
[104] R.A. Norman, F. Ambrosetti, A.M.J.J. Bonvin, L.J. Colwell, S. Kelm, S. Kumar, K. Krawczyk, Computational approaches to therapeutic antibody design: established methods and emerging trends, Brief. Bioinform. 21 (2020) 1549–1567, https://doi.org/10.1093/bib/bbz095.
[105] E.K. Makowski, L. Wu, P. Gupta, P.M. Tessier, Discovery-stage identification of drug-like antibodies using emerging experimental and computational methods, MAbs 13 (2021), 1895540, https://doi.org/10.1080/19420862.2021.1895540.
[106] G. Georgiou, G.C. Ippolito, J. Beausang, C.E. Busse, H. Wardemann, S.R. Quake, The promise and challenge of high-throughput sequencing of the antibody repertoire, Nat. Biotechnol. 32 (2014) 158–168, https://doi.org/10.1038/nbt.2782.
[107] G. Liu, H. Zeng, J. Mueller, B. Carter, Z. Wang, J. Schilz, G. Horny, M. E. Birnbaum, S. Ewert, D.K. Gifford, Antibody complementarity determining region design using high-capacity machine learning, Bioinformatics 36 (2020) 2126–2133, https://doi.org/10.1093/bioinformatics/btz895.
[108] D.M. Mason, S. Friedensohn, C.R. Weber, C. Jordi, B. Wagner, S.M. Meng, R. A. Ehling, L. Bonati, J. Dahinden, P. Gainza, B.E. Correia, S.T. Reddy, Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning, Nat. Biomed. Eng. (2021), https://doi.org/10.1038/s41551-021-00699-9.
[109] Y.W. Lim, A.S. Adler, D.S. Johnson, Predicting antibody binders and generating synthetic antibodies using deep learning, MAbs 14 (2022), 2069075, https://doi.org/10.1080/19420862.2022.2069075.
[110] K. Saka, T. Kakuzaki, S. Metsugi, D. Kashiwagi, K. Yoshida, M. Wada, H. Tsunoda, R. Teramoto, Antibody design using LSTM based deep generative model from phage display library for affinity maturation, Sci. Rep. 11 (2021) 5852, https://doi.org/10.1038/s41598-021-85274-7.
[111] E.K. Makowski, P.C. Kinnunen, J. Huang, L. Wu, M.D. Smith, T. Wang, A.A. Desai, C.N. Streu, Y. Zhang, J.M. Zupancic, J.S. Schardt, J.J. Linderman, P.M. Tessier, Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space, Nat. Commun. 13 (2022) 3788, https://doi.org/10.1038/s41467-022-31457-3.
[112] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.
[113] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G.R. Lee, J. Wang, Q. Cong, L.N. Kinch, R.D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J.H. Pereira, A.V. Rodrigues, A.A. van Dijk, A. C. Ebrecht, D.J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C.K. Yip, J.E. Burke, K.C. Garcia, N.V. Grishin, P. D. Adams, R.J. Read, D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network, Science (2021) 7.
[114] J. Xu, Distance-based protein folding powered by deep learning, Proc. Natl. Acad. Sci. USA 116 (2019) 16856–16865, https://doi.org/10.1073/pnas.1821309116.
[115] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, 2016, pp. 770–8. ⟨https://doi.org/10.1109/CVPR.2016.90⟩.
[116] J.A. Ruffolo, C. Guerra, S.P. Mahajan, J. Sulam, J.J. Gray, Geometric potentials from deep learning improve prediction of CDR H3 loop structures, Bioinformatics 36 (2020) i268–i275, https://doi.org/10.1093/bioinformatics/btaa457.
[117] J. Dunbar, K. Krawczyk, J. Leem, T. Baker, A. Fuchs, G. Georges, J. Shi, C. M. Deane, SAbDab: the structural antibody database, Nucl. Acids Res. 42 (2014) D1140–D1146, https://doi.org/10.1093/nar/gkt1043.
[118] A. Leaver-Fay, M. Tyka, S.M. Lewis, O.F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P.D. Renfrew, C.A. Smith, W. Sheffler, I.W. Davis, S. Cooper, A. Treuille, D.J. Mandell, F. Richter, Y.-E.A. Ban, S.J. Fleishman, J.E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J.J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J.J. Gray, B. Kuhlman, D. Baker, P. Bradley, Rosetta3: an object-oriented software suite for the simulation and design of macromolecules, in: Methods in Enzymology, Elsevier, 2011, pp. 545–574, https://doi.org/10.1016/B978-0-12-381270-4.00019-6.
[119] J.K. Leman, B.D. Weitzner, S.M. Lewis, J. Adolf-Bryfogle, N. Alam, R.F. Alford, M. Aprahamian, D. Baker, K.A. Barlow, P. Barth, B. Basanta, B.J. Bender, K. Blacklock, J. Bonet, S.E. Boyken, P. Bradley, C. Bystroff, P. Conway, S. Cooper, B.E. Correia, B. Coventry, R. Das, R.M. De Jong, F. DiMaio, L. Dsilva, R. Dunbrack, A.S. Ford, B. Frenz, D.Y. Fu, C. Geniesse, L. Goldschmidt, R. Gowthaman, J. J. Gray, D. Gront, S. Guffy, S. Horowitz, P.-S. Huang, T. Huber, T.M. Jacobs, J. R. Jeliazkov, D.K. Johnson, K. Kappel, J. Karanicolas, H. Khakzad, K.R. Khar, S. D. Khare, F. Khatib, A. Khramushin, I.C. King, R. Kleffner, B. Koepnick, T. Kortemme, G. Kuenze, B. Kuhlman, D. Kuroda, J.W. Labonte, J.K. Lai, G. Lapidoth, A. Leaver-Fay, S. Lindert, T. Linsky, N. London, J.H. Lubin, S. Lyskov, J. Maguire, L. Malmström, E. Marcos, O. Marcu, N.A. Marze, J. Meiler,

R. Moretti, V.K. Mulligan, S. Nerli, C. Norn, S. Ó'Conchúir, N. Ollikainen, S. Ovchinnikov, M.S. Pacella, X. Pan, H. Park, R.E. Pavlovicz, M. Pethe, B. G. Pierce, K.B. Pilla, B. Raveh, P.D. Renfrew, S.S.R. Burman, A. Rubenstein, M. F. Sauer, A. Scheck, W. Schief, O. Schueler-Furman, Y. Sedan, A.M. Sevy, N. G. Sgourakis, L. Shi, J.B. Siegel, D.-A. Silva, S. Smith, Y. Song, A. Stein, M. Szegedy, F.D. Teets, S.B. Thyme, R.Y.-R. Wang, A. Watkins, L. Zimmerman, R. Bonneau, Macromolecular modeling and design in Rosetta: recent methods and frameworks, Nat. Methods 17 (2020) 665–680, https://doi.org/10.1038/s41592-020-0848-2.

[120] B.D. Weitzner, J.R. Jeliazkov, S. Lyskov, N. Marze, D. Kuroda, R. Frick, J. Adolf-Bryfogle, N. Biswas, R.L. Dunbrack, J.J. Gray, Modeling and docking of antibody structures with Rosetta, Nat. Protoc. 12 (2017) 401–416, https://doi.org/10.1038/nprot.2016.180.

[121] J.A. Ruffolo, J. Sulam, J.J. Gray, Antibody structure prediction using interpretable deep learning, Patterns 3 (2022), 100406, https://doi.org/10.1016/j.patter.2021.100406.

[122] A. Kovaltsuk, J. Leem, S. Kelm, J. Snowden, C.M. Deane, K. Krawczyk, Observed antibody space: a resource for data mining next-generation sequencing of antibody repertoires, J.I 201 (2018) 2502–2509, https://doi.org/10.4049/jimmunol.1800708.

[123] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is All you Need, n.d., 11.

[124] M.I.J. Raybould, C. Marks, K. Krawczyk, B. Taddese, J. Nowak, A.P. Lewis, A. Bujotzek, J. Shi, C.M. Deane, Five computational developability guidelines for therapeutic antibody profiling, Proc. Natl. Acad. Sci. USA 116 (2019) 4025–4030, https://doi.org/10.1073/pnas.1810576116.

[125] M.L. Chiu, D.R. Goulet, A. Teplyakov, G.L. Gilliland, Antibody structure and function: the basis for engineering therapeutics, Antibodies 8 (2019) 55, https://doi.org/10.3390/antib8040055.

[126] D. Akpinaroglu, J.A. Ruffolo, S.P. Mahajan, J.J. Gray, Simultaneous prediction of antibody backbone and side-chain conformations with deep learning, PLoS One 17 (2022), e0258173, https://doi.org/10.1371/journal.pone.0258173.

[127] B. Abanades, G. Georges, A. Bujotzek, C.M. Deane, ABlooper: fast accurate antibody CDR loop structure prediction with accuracy estimation, Bioinformatics 38 (2022) 1877–1880, https://doi.org/10.1093/bioinformatics/btac016.

[128] V.G. Satorras, E. Hoogeboom, M. Welling, E(n) Equivariant Graph Neural Networks, 2022. ⟨http://arxiv.org/abs/2102.09844⟩, (Accessed 9 October 2022).

[129] S. Muyldermans, Nanobodies: natural single-domain antibodies, Annu. Rev. Biochem. 82 (2013) 775–797, https://doi.org/10.1146/annurev-biochem-063011-092449.

[130] T. Cohen, M. Halfon, D. Schneidman-Duhovny, NanoNet: rapid and accurate end-to-end nanobody modeling by deep learning, Front. Immunol. 13 (2022), 958584, https://doi.org/10.3389/fimmu.2022.958584.

[131] A. Davila, Z. Xu, S. Li, J. Rozewicki, J. Wilamowski, S. Kotelnikov, D. Kozakov, S. Teraguchi, D.M. Standley, AbAdapt: an adaptive approach to predicting antibody–antigen complex structures from sequence, Bioinform. Adv. 2 (2022), vbac015, https://doi.org/10.1093/bioadv/vbac015.

[132] Z. Xu, A. Davila, J. Wilamowski, S. Teraguchi, D.M. Standley, Improved antibody-specific epitope prediction using AlphaFold and AbAdapt**, ChemBioChem 23 (2022), https://doi.org/10.1002/cbic.202200303.

[133] A. Jarasch, H. Koll, J.T. Regula, M. Bader, A. Papadimitriou, H. Kettenberger, Developability assessment during the selection of novel therapeutic antibodies, J. Pharmaceut. Sci. 104 (2015) 1885–1898, https://doi.org/10.1002/jps.24430.

[134] Y. Xu, D. Wang, B. Mason, T. Rossomando, N. Li, D. Liu, J.K. Cheung, W. Xu, S. Raghava, A. Katiyar, C. Nowak, T. Xiang, D.D. Dong, J. Sun, A. Beck, H. Liu, Structure, heterogeneity and developability assessment of therapeutic antibodies, MAbs 11 (2019) 239–264, https://doi.org/10.1080/19420862.2018.1553476.

[135] A.M. Wollacott, C. Xue, Q. Qin, J. Hua, T. Bohnuud, K. Viswanathan, V. B. Kolachalama, Quantifying the nativeness of antibody sequences using long short-term memory networks, Protein Eng. Des. Sel. 32 (2019) 347–354, https://doi.org/10.1093/protein/gzz031.

[136] D. Prihoda, J. Maamary, A. Waight, V. Juan, L. Fayadat-Dilman, D. Svozil, D. A. Bitton, BioPhi: a platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning, MAbs 14 (2022), 2020203, https://doi.org/10.1080/19420862.2021.2020203.

[137] S. Mitragotri, P.A. Burke, R. Langer, Overcoming the challenges in administering biopharmaceuticals: formulation and delivery strategies, Nat. Rev. Drug Discov. 13 (2014) 655–672, https://doi.org/10.1038/nrd4363.

[138] J. Feng, solPredict: Antibody apparent solubility prediction from sequence by transfer learning, n.d., 27.

[139] A. Rives, J. Meier, T. Sercu, S. Goyal, Z. Lin, J. Liu, D. Guo, M. Ott, C.L. Zitnick, J. Ma, R. Fergus, Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences, Proc. Natl. Acad. Sci. USA 118 (2021), e2016239118, https://doi.org/10.1073/pnas.2016239118.

[140] P.-K. Lai, DeepSCM: an efficient convolutional neural network surrogate model for the screening of therapeutic antibody viscosity, Comput. Struct. Biotechnol. J. 20 (2022) 2143–2152, https://doi.org/10.1016/j.csbj.2022.04.035.

[141] N.J. Agrawal, B. Helk, S. Kumar, N. Mody, H.A. Sathish, H.S. Samra, P.M. Buck, L. Li, B.L. Trout, Computational tool for the early screening of monoclonal antibodies for their viscosities, MAbs 8 (2016) 43–48, https://doi.org/10.1080/19420862.2015.1099773.

[142] H. Wu, D.S. Pfarr, S. Johnson, Y.A. Brewah, R.M. Woods, N.K. Patel, W.I. White, J. F. Young, P.A. Kiener, Development of motavizumab, an ultra-potent antibody for the prevention of respiratory syncytial virus infection in the upper and lower respiratory tract, J. Mol. Biol. 368 (2007) 652–665, https://doi.org/10.1016/j.jmb.2007.02.024.

[143] R.L. Kelly, Y. Yu, T. Sun, I. Caffry, H. Lynaugh, M. Brown, T. Jain, Y. Xu, K. D. Wittrup, Target-independent variable region mediated effects on antibody clearance can be FcRn independent, MAbs 8 (2016) 1269–1275, https://doi.org/10.1080/19420862.2016.1208330.

[144] B. Grinshpun, N. Thorsteinson, J.N. Pereira, F. Rippmann, D. Nannemann, V. D. Sood, Y. Fomekong Nanfack, Identifying biophysical assays and *in silico* properties that enrich for slow clearance in clinical-stage therapeutic antibodies, MAbs 13 (2021), 1932230, https://doi.org/10.1080/19420862.2021.1932230.

[145] R. Ferrara, I. Matos, Atypical patterns of response and progression in the era of immunotherapy combinations, Future Oncol. 16 (2020) 1707–1713, https://doi.org/10.2217/fon-2020-0186.

[146] W.A. Cooper, P.A. Russell, M. Cherian, E.E. Duhig, D. Godbolt, P.J. Jessup, C. Khoo, C. Leslie, A. Mahar, D.F. Moffat, et al., , Intra-and interobserver reproducibility assessment of PD-L1 biomarker in non–small cell lung cancerreproducibility of PD-L1 biomarker assessment in NSCLC, Clin. Cancer Res. 23 (2017) 4569–4577, https://doi.org/10.1158/1078-0432.ccr-17-0151.

[147] H. Brunnström, A. Johansson, S. Westbom-Fremer, M. Backman, D. Djureinovic, A. Patthey, M. Isaksson-Mettävainio, M. Gulyas, P. Micke, PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: inter-pathologist variability is higher than assay variability, Mod. Pathol. 30 (2017) 1411–1421, https://doi.org/10.1038/modpathol.2017.59.

[148] X. Bao, H. Zhang, W. Wu, S. Cheng, X. Dai, X. Zhu, Q. Fu, Z. Tong, L. Liu, Y. Zheng, et al., Analysis of the molecular nature associated with microsatellite status in colon cancer identifies clinical implications for immunotherapy, J. Immunother. Cancer 8 (2020), https://doi.org/10.1136/jitc-2020-001437.

[149] R.M. Samstein, C.-H. Lee, A.N. Shoushtari, M.D. Hellmann, R. Shen, Y. Y. Janjigian, D.A. Barron, A. Zehir, E.J. Jordan, A. Omuro, et al., Tumor mutational load predicts survival after immunotherapy across multiple cancer types, Nat. Genet. 51 (2019) 202–206, https://doi.org/10.1038/s41588-018-0312-8.

[150] F. Runa, S. Hamalian, K. Meade, P. Shisgal, P. Gray, J. Kelber, Tumor microenvironment heterogeneity: challenges and opportunities, Curr. Mol. Biol. Rep. 3 (2017) 218–229, https://doi.org/10.1007/s40610-017-0073-7.

[151] J.Y. Lee, B. Kannan, B.Y. Lim, Z. Li, A.H. Lim, J.W. Loh, T.K. Ko, C.C.-Y. Ng, J. Y. Chan, The multi-dimensional biomarker landscape in cancer immunotherapy, Int. J. Mol. Sci. 23 (2022) 7839, https://doi.org/10.3390/ijms23147839.

[152] S. Haider, S. Tyekucheva, D. Prandi, N.S. Fox, J. Ahn, A.W. Xu, A. Pantazi, P. J. Park, P.W. Laird, C. Sander, et al., Systematic assessment of tumor purity and its clinical implications, JCO Precis. Oncol. 4 (2020) 995–1005, https://doi.org/10.1200/po.20.00016.

[153] J. Borst, T. Ahrends, N. Bąbała, C.J. Melief, W. Kastenmüller, CD4+ T cell help in cancer immunology and immunotherapy, Nat. Rev. Immunol. 18 (2018) 635–647, https://doi.org/10.1038/s41577-018-0044-0.

[154] S.J. Casak, L. Marcus, L. Fashoyin-Aje, S.L. Mushti, J. Cheng, Y.-L. Shen, W. F. Pierce, L. Her, K.B. Goldberg, M.R. Theoret, et al., FDA approval summary: pembrolizumab for the first-line treatment of patients with MSI-H/dMMR advanced unresectable or metastatic colorectal carcinomafda approval summary: pembrolizumab, Clin. Cancer Res. 27 (2021) 4680–4684, https://doi.org/10.1158/1078-0432.ccr-21-0557.

[155] D.T. Le, J.N. Durham, K.N. Smith, H. Wang, B.R. Bartlett, L.K. Aulakh, S. Lu, H. Kemberling, C. Wilt, B.S. Luber, et al., Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade, Science 357 (2017) 409–413, https://doi.org/10.1126/science.aan6733.

[156] P. Priestley, J. Baber, M.P. Lolkema, N. Steeghs, E. de Bruijn, C. Shale, K. Duyvesteyn, S. Haidari, A. van Hoeck, W. Onstenk, et al., Pan-cancer whole-genome analyses of metastatic solid tumours, Nature 575 (2019) 210–216, https://doi.org/10.1038/s41586-019-1689-y.

[157] S. Pei, T. Liu, X. Ren, W. Li, C. Chen, Z. Xie, Benchmarking variant callers in next-generation and third-generation sequencing analysis, Brief. Bioinform. 22 (2021), bbaa148, https://doi.org/10.1093/bib/bbaa148.

[158] J.N. Kather, A.T. Pearson, N. Halama, D. Jäger, J. Krause, S.H. Loosen, A. Marx, P. Boor, F. Tacke, U.P. Neumann, et al., Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer, Nat. Med. 25 (2019) 1054–1056, https://doi.org/10.1038/s41591-019-0462-y.

[159] E.L. Moss, D.N. Gorsia, A. Collins, P. Sandhu, N. Foreman, A. Gore, J. Wood, C. Kent, L. Silcock, D.S. Guttery, Utility of circulating tumor DNA for detection and monitoring of endometrial cancer recurrence and progression, Cancers 12 (2020) 2231, https://doi.org/10.3390/cancers12082231.

[160] C. Lu, Y.-C. Zhang, Z.-H. Chen, Q. Zhou, Y.-L. Wu, Applications of circulating tumor DNA in immune checkpoint inhibition: emerging roles and future perspectives, Front. Oncol. 12 (2022), https://doi.org/10.3389/fonc.2022.836891 (836891–836891).

[161] A. Echle, H.I. Grabsch, P. Quirke, P.A. van den Brandt, N.P. West, G.G. Hutchins, L.R. Heij, X. Tan, S.D. Richman, J. Krause, et al., Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning, Gastroenterology 159 (2020) 1406–1416, https://doi.org/10.1053/j.gastro.2020.06.021.

[162] R.A. Barnetson, A. Tenesa, S.M. Farrington, I.D. Nicholl, R. Cetnarskyj, M. E. Porteous, H. Campbell, M.G. Dunlop, Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer, N. Engl. J. Med. 354 (2006) 2751–2763, https://doi.org/10.1056/nejmoa053493.

[163] R. Yamashita, J. Long, T. Longacre, L. Peng, G. Berry, B. Martin, J. Higgins, D. L. Rubin, J. Shen, Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study, Lancet Oncol. 22 (2021) 132–141, https://doi.org/10.1016/s1470-2045(20)30535-0.

[164] V.H. Koelzer, A. Gisler, J.C. Hanhart, J. Griss, S.N. Wagner, N. Willi, G. Cathomas, M. Sachs, W. Kempf, D.S. Thommen, et al., Digital image analysis improves

precision of PD-L1 scoring in cutaneous melanoma, Histopathology 73 (2018) 397–406, https://doi.org/10.1111/his.13528.
[165] A. Kapil, A. Meier, A. Zuraw, K.E. Steele, M.C. Rebelatto, G. Schmidt, N. Brieu, Deep semi supervised generative learning for automated tumor proportion scoring on NSCLC tissue needle biopsies, Sci. Rep. 8 (1) (2018) 10, https://doi.org/10.1038/s41598-018-35501-5.
[166] J. Wu, C. Liu, X. Liu, W. Sun, L. Li, N. Gao, Y. Zhang, X. Yang, J. Zhang, H. Wang, et al., Artificial intelligence-assisted system for precision diagnosis of PD-L1 expression in non-small cell lung cancer, Mod. Pathol. 35 (2022) 403–411, https://doi.org/10.1038/s41379-021-00904-9.
[167] T.A. Chan, M. Yarchoan, E. Jaffee, C. Swanton, S.A. Quezada, A. Stenzinger, S. Peters, Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic, Ann. Oncol. 30 (2019) 44–56, https://doi.org/10.1093/annonc/mdy495.
[168] M.S. Jain, T.F. Massoud, Predicting tumour mutational burden from histopathological images using multiscale deep learning, Nat. Mach. Intell. 2 (2020) 356–362, https://doi.org/10.1038/s42256-020-0190-5.
[169] Y. Niu, L. Wang, X. Zhang, Y. Han, C. Yang, H. Bai, K. Huang, C. Ren, G. Tian, S. Yin, et al., Predicting tumor mutational burden from lung adenocarcinoma histopathological images using deep learning, Front. Oncol. 12 (2022), https://doi.org/10.3389/fonc.2022.927426.
[170] B. He, D. Dong, Y. She, C. Zhou, M. Fang, Y. Zhu, H. Zhang, Z. Huang, T. Jiang, J. Tian, et al., Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker, J. Immunother. Cancer 8 (2020), https://doi.org/10.1136/jitc-2020-000550.
[171] L.H.T. Lam, N.T. Chu, T.-O. Tran, D.T. Do, N.Q.K. Le, A radiomics-based machine learning model for prediction of tumor mutational burden in lower-grade gliomas, Cancers 14 (2022) 3492, https://doi.org/10.3390/cancers14143492.
[172] K. Huang, B. Lin, J. Liu, Y. Liu, J. Li, G. Tian, J. Yang, Predicting colorectal cancer tumor mutational burden from histopathological images and clinical information using multi-modal deep learning, Bioinformatics (2022), https://doi.org/10.1093/bioinformatics/btac641.
[173] J. Galon, D. Bruni, Tumor immunology and tumor evolution: intertwined histories, Immunity 52 (2020) 55–81, https://doi.org/10.1016/j.immuni.2019.12.018.
[174] F. Pagès, B. Mlecnik, F. Marliot, G. Bindea, F.-S. Ou, C. Bifulco, A. Lugli, I. Zlobec, T.T. Rau, M.D. Berger, et al., International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study, Lancet 391 (2018) 2128–2139, https://doi.org/10.1016/s0140-6736(18)30789-x.
[175] M. Van den Eynde, B. Mlecnik, G. Bindea, T. Fredriksen, S.E. Church, L. Lafontaine, N. Haicheur, F. Marliot, M. Angelova, A. Vasaturo, et al., The link between the multiverse of immune microenvironments in metastases and the survival of colorectal cancer patients, Cancer Cell 34 (2018) 1012–1026, https://doi.org/10.1016/j.ccell.2018.11.003.
[176] D. Bruni, H.K. Angell, J. Galon, The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy, Nat. Rev. Cancer 20 (2020) 662–680, https://doi.org/10.1038/s41568-020-0285-7.
[177] N. Scholler, R. Perbost, F.L. Locke, M.D. Jain, S. Turcan, C. Danan, E.C. Chang, S. S. Neelapu, D.B. Miklos, C.A. Jacobson, et al., Tumor immune contexture is a determinant of anti-CD19 CAR T cell efficacy in large B cell lymphoma, Nat. Med. 28 (2022) 1872–1882, https://doi.org/10.1038/s41591-022-01916-x.
[178] B.A. Helmink, S.M. Reddy, J. Gao, S. Zhang, R. Basar, R. Thakur, K. Yizhak, M. Sade-Feldman, J. Blando, G. Han, et al., B cells and tertiary lymphoid structures promote immunotherapy response, Nature 577 (2020) 549–555, https://doi.org/10.1038/s41586-019-1922-8.
[179] R. Cabrita, M. Lauss, A. Sanna, M. Donia, M. Skaarup Larsen, S. Mitra, I. Johansson, B. Phung, K. Harbst, J. Vallon-Christersson, et al., Tertiary lymphoid structures improve immunotherapy and survival in melanoma, Nature 577 (2020) 561–565, https://doi.org/10.1038/s41586-019-1914-8.
[180] F. Petitprez, A. de Reyniès, E.Z. Keung, T.W.-W. Chen, C.-M. Sun, J. Calderaro, Y.-M. Jeng, L.-P. Hsiao, L. Lacroix, A. Bougouïin, et al., B cells are associated with survival and immunotherapy response in sarcoma, Nature 577 (2020) 556–560, https://doi.org/10.1038/s41586-019-1906-8.
[181] D. Reiman, L. Sha, I. Ho, T. Tan, D. Lau, A.A. Khan, Integrating RNA expression and visual features for immune infiltrate prediction, in: BIOCOMPUTING 2019: Proceedings of the Pacific Symposium, World Scientific, 2019, pp. 284–295, https://doi.org/10.1142/9789813279827_0026.
[182] A. Zaitsev, M. Chelushkin, D. Dyikanov, I. Cheremushkin, B. Shpak, K. Nomie, V. Zyrin, E. Nuzhdina, Y. Lozinsky, A. Zotova, et al., Precise reconstruction of the TME using bulk RNA-seq and a machine learning algorithm trained on artificial transcriptomes, Cancer Cell 40 (2022) 879–894, https://doi.org/10.1016/j.ccell.2022.07.006.
[183] A.M. Newman, C.B. Steen, C.L. Liu, A.J. Gentles, A.A. Chaudhuri, F. Scherer, M. S. Khodadoust, M.S. Esfahani, B.A. Luca, D. Steiner, et al., Determining cell type abundance and expression from bulk tissues with digital cytometry, Nat. Biotechnol. 37 (2019) 773–782, https://doi.org/10.1038/s41587-019-0114-2.
[184] K. Menden, M. Marouf, S. Oller, A. Dalmia, D.S. Magruder, K. Kloiber, P. Heutink, S. Bonn, Deep learning–based cell composition analysis from tissue expression profiles, Sci. Adv. 6 (2020), eaba2619, https://doi.org/10.1126/sciadv.aba2619.
[185] A. Chakravarthy, A. Furness, K. Joshi, E. Ghorani, K. Ford, M.J. Ward, E.V. King, M. Lechner, T. Marafioti, S.A. Quezada, et al., Pan-cancer deconvolution of tumour composition using DNA methylation, Nat. Commun. 9 (1) (2018) 13, https://doi.org/10.1038/s41467-018-05570-1.
[186] J.J. Levy, A.J. Titus, C.L. Petersen, Y. Chen, L.A. Salas, B.C. Christensen, MethylNet: an automated and modular deep learning approach for DNA methylation analysis, BMC Bioinform. 21 (1) (2020) 15, https://doi.org/10.1186/s12859-020-3443-8.
[187] D. Lau, A.M. Bobe, A.A. Khan, RNA sequencing of the tumor microenvironment in precision cancer immunotherapy, Trends Cancer 5 (2019) 149–156, https://doi.org/10.1016/j.trecan.2019.02.006.
[188] D.J. Fassler, S. Abousamra, R. Gupta, C. Chen, M. Zhao, D. Paredes, S.A. Batool, B. S. Knudsen, L. Escobar-Hoyos, K.R. Shroyer, et al., Deep learning-based image analysis methods for brightfield-acquired multiplex immunohistochemistry images, Diagn. Pathol. 15 (1) (2020) 11, https://doi.org/10.1186/s13000-020-01003-0.
[189] J. Saltz, R. Gupta, L. Hou, T. Kurc, P. Singh, V. Nguyen, D. Samaras, K.R. Shroyer, T. Zhao, R. Batiste, et al., Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images, Cell Rep. 23 (2018) 181–193, https://doi.org/10.1016/j.celrep.2018.03.086.
[190] B. Chen, M.S. Khodadoust, C.L. Liu, A.M. Newman, A.A. Alizadeh, Profiling tumor infiltrating immune cells with CIBERSORT, in: Cancer Systems Biology, Springer, 2018, pp. 243–259.
[191] C. Peterson, N. Denlinger, Y. Yang, Recent advances and challenges in cancer immunotherapy, Cancers 14 (2022) 3972, https://doi.org/10.3390/cancers14163972.
[192] S. Akbar, M. Peikari, S. Salama, A.Y. Panah, S. Nofech-Mozes, A.L. Martel, Automated and manual quantification of tumour cellularity in digital slides for tumour burden assessment, Sci. Rep. 9 (2019) 1–9, https://doi.org/10.1038/s41598-019-50568-4.
[193] A. Rakhlin, A. Tiulpin, A.A. Shvets, A.A. Kalinin, V.I. Iglovikov, S. Nikolenko, Breast tumor cellularity assessment using deep neural networks, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019. ⟨https://doi.org/10.1109/ICCVW.2019.00048⟩.
[194] J.A. Diao, J.K. Wang, W.F. Chui, V. Mountain, S.C. Gullapally, R. Srinivasan, R. N. Mitchell, B. Glass, S. Hoffman, S.K. Rao, et al., Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes, Nat. Commun. 12 (1) (2021) 15, https://doi.org/10.1038/s41467-021-21896-9.
[195] L. Peng, F. Wang, Z. Wang, J. Tan, L. Huang, X. Tian, G. Liu, L. Zhou, Cell–cell communication inference and analysis in the tumour microenvironments from single-cell transcriptomics: data resources and computational strategies, Brief. Bioinform. 23 (2022), bbac234, https://doi.org/10.1093/bib/bbac234.
[196] C. Wu, J. Yang, W. Xiao, Z. Jiang, S. Chen, D. Guo, P. Zhang, C. Liu, H. Yang, Z. Xie, Single-cell characterization of malignant phenotypes and microenvironment alteration in retinoblastoma, Cell Death Dis. 13 (2022) 1–12, https://doi.org/10.1038/s41419-022-04904-8.
[197] J. Dohmen, A. Baranovskii, J. Ronen, B. Uyar, V. Franke, A. Akalin, Identifying tumor cells at the single-cell level using machine learning, Genome Biol. 23 (1) (2022) 23, https://doi.org/10.1186/s13059-022-02683-1.
[198] Y. He, H. Yuan, C. Wu, Z. Xie, DISC: a highly scalable and accurate inference of gene expression and structure for single-cell transcriptomes using semi-supervised deep learning, Genome Biol. 21 (1) (2020) 28, https://doi.org/10.1186/s13059-020-02083-3.
[199] S.N. Nerurkar, D. Goh, C.C.L. Cheung, P.Q.Y. Nga, J.C.T. Lim, J.P.S. Yeong, Transcriptional spatial profiling of cancer tissues in the era of immunotherapy: the potential and promise, Cancers 12 (2020) 2572, https://doi.org/10.3390/cancers12092572.
[200] M. Larroquette, J.-P. Guegan, B. Besse, S. Cousin, M. Brunet, S. Le Moulec, F. Le Loarer, C. Rey, J.-C. Soria, F. Barlesi, et al., Spatial transcriptomics of macrophage infiltration in non-small cell lung cancer reveals determinants of sensitivity and resistance to anti-PD1/PD-L1 antibodies, J. Immunother. Cancer 10 (2022), e003890, https://doi.org/10.1136/jitc-2021-003890.
[201] S.M. Lewis, M.-L. Asselin-Labat, Q. Nguyen, J. Berthelet, X. Tan, V.C. Wimmer, D. Merino, K.L. Rogers, S.H. Naik, Spatial omics and multiplexed imaging to explore cancer biology, Nat. Methods 18 (2021) 997–1012, https://doi.org/10.1038/s41592-021-01203-6.
[202] H. Choi, K.J. Na, J. Koh, Y.T. Kim, Deep learning-based tumor microenvironment cell types mapping from H&E images of lung adenocarcinoma using spatial transcriptomic data, Cancer Res. 82 (2022), https://doi.org/10.1158/1538-7445.AM2022-5131 (5131–5131).
[203] A. Zubair, R.H. Chapple, S. Natarajan, W.C. Wright, M. Pan, H.-M. Lee, H. Tillman, J. Easton, P. Geeleher, Cell type identification in spatial transcriptomics data can be improved by leveraging cell-type-informative paired tissue images using a Bayesian probabilistic model, Nucleic Acids Res. 50 (2022), https://doi.org/10.1093/nar/gkac320 (e80–e80).
[204] B. He, L. Bergenstråhle, L. Stenbeck, A. Abid, A. Andersson, Å. Borg, J. Maaskola, J. Lundeberg, J. Zou, Integrating spatial gene expression and breast tumour morphology via deep learning, Nat. Biomed. Eng. 4 (2020) 827–834, https://doi.org/10.1038/s41551-020-0578-x.
[205] L. Bergenstråhle, B. He, J. Bergenstråhle, X. Abalo, R. Mirzazadeh, K. Thrane, A. L. Ji, A. Andersson, L. Larsson, N. Stakenborg, et al., Super-resolved spatial transcriptomics by deep data fusion, Nat. Biotechnol. 40 (2022) 476–479, https://doi.org/10.1038/s41587-021-01075-3.
[206] T.N. Gide, J.S. Wilmott, R.A. Scolyer, G.V. Long, Primary and acquired resistance to immune checkpoint inhibitors in metastatic melanomaresistance to immunotherapy in melanoma, Clin. Cancer Res. 24 (2018) 1260–1270, https://doi.org/10.1158/1078-0432.ccr-17-2267.
[207] J. Kong, D. Ha, J. Lee, I. Kim, M. Park, S.-H. Im, K. Shin, S. Kim, Network-based machine learning approach to predict immunotherapy response in cancer patients, Nat. Commun. 13 (2022) 1–15, https://doi.org/10.1038/s41467-022-31535-6.

[208] F. Xie, J. Zhang, J. Wang, A. Reuben, W. Xu, X. Yi, F.S. Varn, Y. Ye, J. Cheng, M. Yu, et al., Multifactorial deep learning reveals pan-cancer genomic tumor clusters with distinct immunogenomic landscape and response to immunotherapydeep learning modeling tumor immune landscape, Clin. Cancer Res. 26 (2020) 2908–2920, https://doi.org/10.1158/1078-0432.ccr-19-1744.
[209] D. Chowell, S.-K. Yoo, C. Valero, A. Pastore, C. Krishna, M. Lee, D. Hoen, H. Shi, D.W. Kelly, N. Patel, et al., Improved prediction of immune checkpoint blockade efficacy across multiple cancer types, Nat. Biotechnol. 40 (2022) 499–506, https://doi.org/10.1038/s41587-021-01070-8.
[210] J.-W. Sidhom, G. Oliveira, P. Ross-MacDonald, M. Wind-Rotolo, C.J. Wu, D. M. Pardoll, A.S. Baras, Deep learning reveals predictive sequence concepts within immune repertoires to immunotherapy, Sci. Adv. 8 (2022), eabq5089, https://doi.org/10.1126/sciadv.abq5089.
[211] S. Trebeschi, S. Drago, N. Birkbak, I. Kurilova, A. Călin, A.D. Pizzi, F. Lalezari, D. Lambregts, M. Rohaan, C. Parmar, et al., Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers, Ann. Oncol. 30 (2019) 998–1004, https://doi.org/10.1093/annonc/mdz108.
[212] R. Sun, E.J. Limkin, M. Vakalopoulou, L. Dercle, S. Champiat, S.R. Han, L. Verlingue, D. Brandao, A. Lancia, S. Ammari, et al., A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study, Lancet Oncol. 19 (2018) 1180–1191, https://doi.org/10.1016/s1470-2045(18)30413-3.
[213] P. Johannet, N. Coudray, D.M. Donnelly, G. Jour, I. Illa-Bochaca, Y. Xia, D. B. Johnson, L. Wheless, J.R. Patrinely, S. Nomikou, et al., Using machine learning algorithms to predict immunotherapy response in patients with advanced melanomapredicting immunotherapy response in advanced melanoma, Clin. Cancer Res. 27 (2021) 131–140, https://doi.org/10.1158/1078-0432.ccr-20-2415.
[214] K. Bera, N. Braman, A. Gupta, V. Velcheti, A. Madabhushi, Predicting cancer outcomes with radiomics and artificial intelligence in radiology, Nat. Rev. Clin. Oncol. 19 (2022) 132–146, https://doi.org/10.1038/s41571-021-00560-7.
[215] A. Echle, N.T. Rindtorff, T.J. Brinker, T. Luedde, A.T. Pearson, J.N. Kather, Deep learning in cancer pathology: a new generation of clinical biomarkers, Br. J. Cancer 124 (2021) 686–696, https://doi.org/10.1038/s41416-020-01122-x.
[216] R.S. Vanguri, J. Luo, A.T. Aukerman, J.V. Egger, C.J. Fong, N. Horvat, A. Pagano, J. de A.B. Araujo-Filho, L. Geneslaw, H. Rizvi, et al., Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L) 1 blockade in patients with non-small cell lung cancer, Nat. Cancer (2022) 1–14, https://doi.org/10.1038/s43018-022-00416-8.
[217] K.M. Boehm, P. Khosravi, R. Vanguri, J. Gao, S.P. Shah, Harnessing multimodal data integration to advance precision oncology, Nat. Rev. Cancer 22 (2022) 114–126, https://doi.org/10.1038/s41568-021-00408-3.
[218] A. Kleppe, O.-J. Skrede, S. De Raedt, K. Liestøl, D.J. Kerr, H.E. Danielsen, Designing deep learning studies in cancer diagnostics, Nat. Rev. Cancer 21 (2021) 199–211, https://doi.org/10.1038/s41568-020-00327-9.
[219] J. Yuan, Z. Hu, B.A. Mahal, S.D. Zhao, K.H. Kensler, J. Pi, X. Hu, Y. Zhang, Y. Wang, J. Jiang, et al., Integrated analysis of genetic ancestry and genomic alterations across cancers, Cancer Cell 34 (2018) 549–560, https://doi.org/10.1016/j.ccell.2018.08.019.
[220] B. Herpers, B. Eppink, M.I. James, C. Cortina, A. Cañellas-Socias, S.F. Boj, X. Hernando-Momblona, D. Glodzik, R.C. Roovers, M. van de Wetering, et al., Functional patient-derived organoid screenings identify MCLA-158 as a therapeutic EGFR$\times$ LGR5 bispecific antibody with efficacy in epithelial tumors, Nat. Cancer 3 (2022) 418–436, https://doi.org/10.1038/s43018-022-00359-0.
[221] V. Cabannes, From Weakly Supervised Learning to Active Learning, ArXiv Preprint ArXiv:2209.11629, 2022. ⟨https://doi.org/10.48550/arXiv.2209.11629⟩.
[222] G. Novakovsky, N. Dexter, M.W. Libbrecht, W.W. Wasserman, S. Mostafavi, Obtaining genetics insights from deep learning via explainable artificial intelligence, Nat. Rev. Genet. (2022) 1–13, https://doi.org/10.1038/s41576-022-00532-2.
[223] W.J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, B. Yu, Definitions, methods, and applications in interpretable machine learning, Proc. Natl. Acad. Sci. USA 116 (2019) 22071–22080, https://doi.org/10.1073/pnas.1900654116.
[224] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115, https://doi.org/10.1016/j.inffus.2019.12.012.
[225] J. Wu, W. Wang, J. Zhang, B. Zhou, W. Zhao, Z. Su, X. Gu, J. Wu, Z. Zhou, S. Chen, DeepHLApan: a deep learning approach for neoantigen prediction considering both HLA-peptide binding and immunogenicity, Front. Immunol. 10 (2019), 2559, https://doi.org/10.3389/fimmu.2019.02559.
[226] S. Kim, H.S. Kim, E. Kim, M.G. Lee, E.-C. Shin, S. Paik, S. Kim, Neopepsee: accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information, Ann. Oncol. 29 (2018) 1030–1036, https://doi.org/10.1093/annonc/mdy022.
[227] T. Lu, Z. Zhang, J. Zhu, Y. Wang, P. Jiang, X. Xiao, C. Bernatchez, J.V. Heymach, D.L. Gibbons, J. Wang, L. Xu, A. Reuben, T. Wang, Deep learning-based prediction of the T cell receptor-antigen binding specificity, Nat. Mach. Intell. 3 (2021) 864–875, https://doi.org/10.1038/s42256-021-00383-2.
[228] K.M. Boehm, B. Bhinder, V.J. Raja, N. Dephoure, O. Elemento, Predicting peptide presentation by major histocompatibility complex class I: an improved machine learning approach to the immunopeptidome, BMC Bioinform. 20 (2019) 7, https://doi.org/10.1186/s12859-018-2561-z.
[229] P.R. Buckley, C.H. Lee, R. Ma, I. Woodhouse, J. Woo, V.O. Tsvetkov, D. S. Shcherbinin, A. Antanaviciute, M. Shughay, M. Rei, A. Simmons, H. Koohy, Evaluating performance of existing computational models in predicting CD8+ T cell pathogenic epitopes and cancer neoantigens, Brief. Bioinform. 23 (2022), bbac141, https://doi.org/10.1093/bib/bbac141.
[230] S. Sarkizova, S. Klaeger, P.M. Le, L.W. Li, G. Oliveira, H. Keshishian, C. R. Hartigan, W. Zhang, D.A. Braun, K.L. Ligon, P. Bachireddy, I.K. Zervantonakis, J.M. Rosenbluth, T. Ouspenskaia, T. Law, S. Justesen, J. Stevens, W.J. Lane, T. Eisenhaure, G. Lan Zhang, K.R. Clauser, N. Hacohen, S.A. Carr, C.J. Wu, D. B. Keskin, A large peptidome dataset improves HLA class I epitope prediction across most of the human population, Nat. Biotechnol. 38 (2020) 199–209, https://doi.org/10.1038/s41587-019-0322-9.
[231] A. Montemurro, V. Schuster, H.R. Povlsen, A.K. Bentzen, V. Jurtz, W. D. Chronister, A. Crinklaw, S.R. Hadrup, O. Winther, B. Peters, L.E. Jessen, M. Nielsen, NetTCR-2.0 enables accurate prediction of TCR-peptide binding by using paired TCRα and β sequence data, Commun. Biol. 4 (2021) 1060, https://doi.org/10.1038/s42003-021-02610-3.
[232] B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, 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. 48 (2020) W449–W454, https://doi.org/10.1093/nar/gkaa379.
[233] M. Lu, L. Xu, X. Jian, X. Tan, J. Zhao, Z. Liu, Y. Zhang, C. Liu, L. Chen, Y. Lin, L. Xie, dbPepNeo2.0: a database for human tumor neoantigen peptides from mass spectrometry and TCR recognition, Front. Immunol. 13 (2022), 855976, https://doi.org/10.3389/fimmu.2022.855976.