Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

Cancer Cell

CellPress

Review

Challenges in neoantigen-directed therapeutics

Lien Lybaert,[1,10] Steve Lefever,[1,10] Bruno Fant,[1] Evelien Smits,[2] Bruno De Geest,[3] Karine Breckpot,[4] Luc Dirix,[5] Steven A. Feldman,[6] Wim van Criekinge,[7] Kris Thielemans,[4] Sjoerd H. van der Burg,[8] Patrick A. Ott,[9] and Cedric Bogaert[1,*]
[1]myNEO NV, 9000 Ghent, Belgium
[2]Center for Oncological Research, University of Antwerp, 2610 Wilrijk, Belgium
[3]Department of Pharmaceutics, Ghent University, 9000 Ghent, Belgium
[4]Laboratory of Molecular and Cellular Therapy, Department of Biomedical Sciences, Vrije Universiteit Brussel, Brussels, Belgium
[5]Translational Cancer Research Unit, Center for Oncological Research, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium
[6]Center for Cancer Cell Therapy, Stanford University School of Medicine, Stanford, CA, USA
[7]Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium
[8]Medical Oncology, Oncode Institute, Leiden University Medical Center, Leiden, the Netherlands
[9]Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA
[10]These authors contributed equally
*Correspondence: cedric@myneo.me
https://doi.org/10.1016/j.ccell.2022.10.013

SUMMARY

A fundamental prerequisite for the efficacy of cancer immunotherapy is the presence of functional, antigen-specific T cells within the tumor. Neoantigen-directed therapy is a promising strategy that aims at targeting the host's immune response against tumor-specific antigens, thereby eradicating cancer cells. Initial forays have been made in clinical environments utilizing vaccines and adoptive cell therapy; however, many challenges lie ahead. We provide an in-depth overview of the current state of the field with an emphasis on *in silico* neoantigen discovery and the clinical aspects that need to be addressed to unlock the full potential of this therapy.

INTRODUCTION

The promise of utilizing the body's immune system to treat cancer is in part linked to the discovery of "cancer immunoediting," which states that the immune system can play a vital role in the protection of the host against tumorigenesis but can also shape and even promote tumor growth.[1] The understanding that tumors develop upon immune evasion paved the way for new treatment options to shift the balance from a pro-tumoral environment toward the development of an unfavorable setting for cancer cells by (re)boosting the immune system.[2,3]

So far, immune checkpoint inhibitors (ICIs) are considered the most promising class of immunotherapeutics for the treatment of solid tumors, leading to FDA approvals for many cancer indications.[4] ICIs target inhibitory receptors expressed on tumor cells and on immuno-suppressive cells residing in the lymphatic system, peripheral blood, and the tumor microenvironment (TME). Blockage of these receptors (checkpoints) enables the patient's immune system to once again identify and attack cancer cells. The most common of these checkpoints currently targeted are the programmed-death (ligand)-1 (PD-1/PD-L1) transmembrane proteins. The emergence of ICIs has led to a changed standard of care for multiple tumor types resulting in improved outcomes for many cancer patients. In BRAF wild-type metastatic melanoma, for example, the overall survival (OS) is significantly higher when treated with ICIs compared with chemotherapy, i.e., up to a 1-year survival rate of 72.9% versus 42.1%.[5] Furthermore, durable responses have been achieved in cancer types that lacked treatments capable of inducing long-term tumor control, such as advanced head and neck cancer and non-small cell lung cancer.[6] This has led to the approvals of pembrolizumab (Keytruda) in 2016 and 2017 as the first cancer immuno-therapy for first-line treatment of different solid tumors. In addition, it was also approved for the first tissue/site-agnostic indication in microsatellite instability-high or mismatch repair deficient solid tumors. The striking results of ICIs in advanced disease triggered the testing of ICI in early-stage patients.[7]

Nevertheless, many patients still do not benefit from ICI treatment. For example, Leger et al. showed that the majority of NSCLC patients do not respond to nivolumab[8] and others report similar numbers across a variety of cancers. An analysis by Haslam et al. estimated the percentage of US cancer patients eligible for ICI drugs to be 43%, whereas in actuality only 12% of patients respond to the therapy.[9] Although expression of PD-L1 is still considered to be the most widely used biomarker to predict response to most classes of ICI therapy, a cohort study performed by Yarchoan et al. revealed that, of 10,000 samples across all major tumor types, only 15.2% of patients had PD-L1-positive tumors.[10] Taken together, these studies show that the initial hope that ICI monotherapy might be the key in the fight against cancer for all cancer patients was unrealistic. Many challenges still need to be overcome to increase response rates to ICIs as well as to find other immunotherapeutic approaches for patients who do not benefit from the current standard of care.

The observation of a correlation between tumor mutational burden (TMB), linked to the amount of tumor-unique mutated peptides, and response rate in patients who are treated with PD-1 inhibition[11,12] provided initial evidence for the importance

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

BOX 1. Tumor-specific antigens (TSAs) versus tumor-associated antigens (TAAs)

The difference between TSAs and TAAs lies in their expression pattern. TSAs are antigens arising from features uniquely present in tumor cells, such as somatic mutations, or tumor-specific gene expression and splicing patterns. TAAs on the other hand, are neoantigens generated by features that are enriched in tumor cells, although these are also present in normal cells. An example of TAAs are the cancer-testis antigens, such as the MAGE-A antigens that are expressed in both normal and tumor cells, but show a significant elevated expression in the latter.[32] As TAAs are self antigens it is a lot harder to induce a potent immune reaction against these antigens due to immune tolerance. In case the immune system can be directed toward these types of antigens, there is the risk for off-target toxicity where non-diseased tissues and cells can also be affected, resulting in potentially dire effects on the patients' well-being.

of neoantigens as the targets of immune responses mediating cancer control. In addition, several studies in mice and humans have shown that neoantigen-specific T cells can induce tumor regression in patients who are receiving both tumor-infiltrating lymphocyte (TIL) therapy[13–17] and ICI therapy.[18–21] In this respect, personalized immunotherapies—aiming to increase the number of neoantigen-specific T cells—are being explored. These neoantigen-directed therapies involve cancer vaccines or adoptive cell therapies (ACT), i.e., neoantigen-specific T cell receptor (TCR) or chimeric antigen receptor (CAR) T cell therapy.

In 2015,[22] 2017,[23,24] and 2019,[25] the first clinical phase I trials testing personalized therapies were successful in high and low TMB tumors. Hu et al. showed persistent neoantigen-specific memory T cells 2–4.5 years after neoantigen vaccination, indicating the potential promise of neoantigen-driven immunotherapy in providing long-term control of and protection against metastases.[26] Furthermore, it is expected that improvements in DNA and RNA sequencing (RNA-seq) technology and reduced costs will further boost the potential of neoantigen-based immunotherapy. It should also be noted that, usually, only minor off-target effects develop because neoantigens are tumor specific.[27] Altogether, these data shows the promise of personalized immunotherapy to (re-)activate the immune system against neoantigens expressed by the patient's cancer cells.

THE CLINICAL CHALLENGE

Despite the theoretical promise of neoantigen-based vaccines, there is considerable skepticism about neoantigen-driven immunotherapy, so far no neoantigen-directed therapy has yet received regulatory approval for use in cancer patients. This lack of enthusiasm may be explained by the relatively modest anti-tumor activity seen in neoantigen-directed vaccine trials to date. Promising efficacy in epithelial cancers has been reported when neoantigen-specific T cells are administered to the patients directly, albeit with limited data available. The thus far modest results may be attributed to suboptimal vaccine platforms, poor choice of antigen, substandard clinical trial, and therapy design, as well as unconsidered tumoral heterogeneity, among other reasons.[28] ACT, on the other hand, in particular CAR T cells, have shown remarkably effective results already, resulting in the FDA approval of tisagenlecleucel and axicabtagene ciloleucel in several hematological cancers. Aside from hematological cancers, ACT therapy is now also being explored in solid cancers, which account for the majority of all tumor-related casualties. Although anti-tumor responses have been shown with ACT therapy in select solid tumor cases, the bulk of the efforts have failed to be efficient enough.[29–31] The relatively modest anti-tumor activity of ACT therapies for solid tumors may be attributed to the use of tumor-associated tumor-specific antigens (TAAs) rather than tumor-specific antigens (TSAs) (see Box 1).[15] Another important reason why ACT therapy has shown relatively low efficacy in solid tumors compared with hematologic malignancies is likely related to the tumor itself. The presence of a suppressed immune system in patients with advanced cancers may play a role. Moreover, solid tumors exhibit a vastly complex environment of immunosuppressive circuits that can all lead to their evasion from immune attack.[33,34] This tumoral heterogeneity may in fact be one of the main reasons why cancer immunotherapy, in general, fails for many types of cancers.

Tumoral heterogeneity

In recent years, the importance of the immunological tumor profile has been elucidated by the discovery of the significant impact the TME has on cancer development and treatment response. Where disappointing clinical results were poorly understood before, nowadays these can often be explained by the heterogeneity of cancers on a genetic[35] and immunological level. The genetic heterogeneous nature of a tumor is discussed in detail below (Section "Genetic variability (clonality): genetic heterogeneity").

The concept of tumoral immune heterogeneity became undeniable when checkpoint inhibition therapy did not live up to the expectations after several disappointing clinical results in different cancer types.[36] Notably, it was observed that some patients benefited significantly, whereas others—having the apparent same type of cancer—did not respond at all. These results revealed specific immune evasion mechanisms that tumors use to evade and survive the immune system, which can differ between distinct tumor types (inter-tumoral) and also within the same tumor type (intra-tumoral). This heterogeneity introduces complexity because every tumor develops its own unique immune profile to evade the anti-tumor immune response on multiple levels. Three "immune types" have been defined, each with its own therapy-response profile related to a different status of immune activity: immune inflamed, immune excluded, and immune deserted.[37]

Combination therapy

Since the anti-tumor activity of ICI monotherapy is relatively low in most tumor types, different combinatorial strategies are being investigated, of which neoantigen-driven therapies are a prime candidate to both focus and activate the immune system of patients. A single immunotherapy modality is unlikely to adequately address the heterogeneity and adaptability of tumors in which resistance mechanisms can be multi-faceted and acquired

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

CellPress

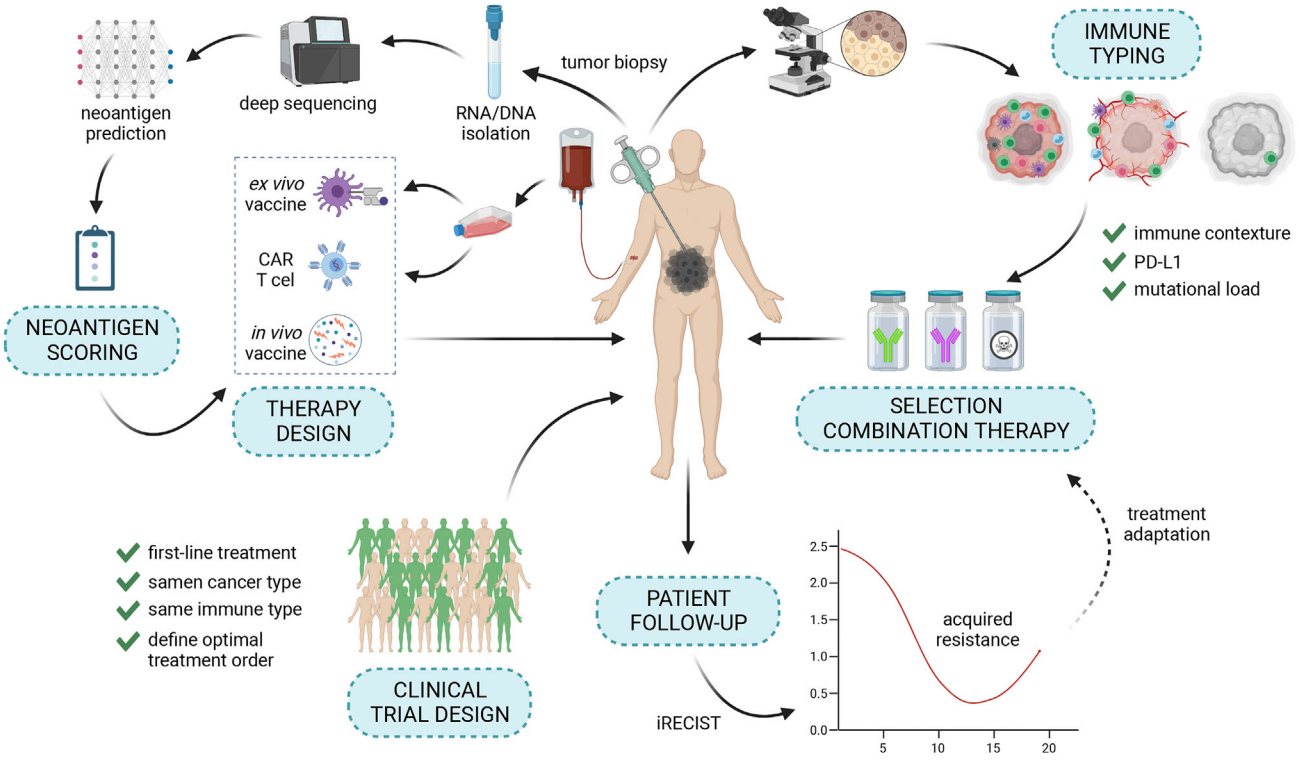

Figure 1. Overview of the different factors that aid to unlock the full potential of neoantigen-directed therapeutics
The general workflow starts with a tumor biopsy, followed by immune typing, therapy selection, and/or neoantigen-based therapy design. During and after therapy, patients' tumoral status is checked regularly to identify a relapse and/or resistance.

during treatment. Hence, treating the patient with double or triple therapy could lead to more efficient tumor elimination by attacking the cancer cells on multiple levels. In addition, longitudinal evaluation during treatment is needed to unveil the potential development of certain resistance mechanisms and to adjust the therapy accordingly[38] (Figure 1). It has been observed that tumors relapsing under ICI treatment show a different mutational landscape with a different spectrum of neo-epitope variants.[39] This could be explained by the selective pressure imposed by the ICI-activated immune system on the selected set of spontaneously recognized neoantigens. The timing of the different treatments can be critical too. For instance, it was demonstrated that an anti-OX40 antibody followed by a PD-1-inhibitor provided a beneficial effect, while the opposite was true when treatments were reversed.[40,41] But not only does the sequence of combination therapies matter, but the optimal dose and optimal duration is also not always clear. Furthermore, adverse event management will become more important and dose regimens should be optimized in combination therapy to achieve the highest clinical benefit with acceptable toxicity. Since neoantigen-driven cancer vaccination, and ACT in principle, can both be used to treat every tumor immune type, provided it is combined with the right TME-modulating complementary therapy, these therapies may offer benefit even for patients who have less-inflamed tumors. Given that combination immunotherapy often significantly increases treatment-related adverse events,[42] the favorable toxicity profile associated with cancer vaccination offers an unprecedented combination advantage.[28] One could also expect that the dose of the therapy could even be lowered when combined with cancer vaccines, due to more optimal immune system (re-)activation.

Inflamed tumors: Neoantigen-driven therapeutics and ICIs

Inflamed tumors carry a highly immunogenic tumor profile, with large quantities of immune cells infiltrating into the tumor. The activity of these immune cells, however, is often limited due to the evasion techniques that the tumors use. Considering the central role that neoantigens play as targets of effective anti-tumor immune responses, coupled with the recognition that the TME is often adversarial and counteracts these immune responses, combining neoantigen-driven therapeutics with ICIs may have significant clinical value. Whereas vaccines can induce, broaden, and boost the neoantigen-specific T cell population, ACT can in principle deliver large quantities of neoantigen-specific T cells without the need for *in vivo* expansion. The combination with ICI therapy will ultimately lead to an expanded tumor-reactive T cell repertoire that can recognize neoantigen-presenting cancer cells.[43] In addition, widening the breadth of neoantigen reactivity may delay the development of acquired resistance to immune checkpoint blockade related to selective pressure.[44] Neoantigen-specific, tumor-infiltrating T cells generated by neoantigen-directed cancer vaccination, or ACT, can also stimulate IFN-ɣ production in the TME. This is often correlated with upregulated PD-L1 expression on tumor cells, preventing a long-lasting effector T cell response by inducing an inactive state.[44]

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

The potential of this approach has recently been shown in clinical studies.[13,23,24,44] Sahin et al. and Ott et al. both showed the potential benefit of combining cancer vaccination with ICI therapy in melanoma patients, whereas a neoantigen-targeting TIL-ACT approach combined with pembrolizumab was reported by Zacharakis et al. in a breast cancer patient. In these cases, checkpoint inhibition therapy has indeed been shown to enable the effector function of neoantigen-directed T cells once they home in on the immuno-suppressive TME.

Excluded tumors: Neoantigen-driven therapeutics and anti-angiogenic therapy

Immune-excluded cancers develop a dense stroma that prevents infiltration of T cells and other anti-tumorigenic immune cells. Consequently, the neoantigen-directed T cells induced or reactivated by neoantigen-driven therapeutics or ICIs are not able to reach the tumor. Combining neoantigen-driven therapeutics or ICIs with anti-angiogenic therapy, directed against vascularization could, in these cases, facilitate tumor T cell infiltration.

Following encouraging preclinical studies, the combination of anti-angiogenic therapy and therapeutic vaccines is currently being investigated in clinical environments. However, so far, no benefit has been reported in clinical phase III studies.[45] One might expect that checkpoint inhibition could be needed in these tumors to reactivate the exhausted T cells. This approach is supported by multiple studies testing ICI therapy combined with anti-angiogenic therapy showing significant clinical benefit in immune-excluded tumors.[46] So far, the most promising results have been achieved in renal cell[47,48] and hepatocellular carcinoma[49]; however, final outcomes of large phase III clinical tests for other indications are still to be observed.[50,51]

Deserted tumors: Neoantigen-driven therapeutics and ICD-induction therapy

Identifying efficient treatment options for immune-deserted tumors will be critical, as immunotherapy options are limited for these patients. Immunogenic cell death (ICD)-induction therapy is one of the strategies being explored. Radio/chemotherapy and oncolytic virotherapy both cause ICD, resulting in the sudden release of different molecules. These involve cancer antigens, damaged proteins, damage-associated molecular patterns, and pro-inflammatory cytokines that attract and can therefore trigger the immune system against cancer cells.[52] Furthermore, neoantigen-driven therapeutics have been shown to have promise as well, and, consequently, combination therapies could add further benefit to those patients.

Cytoreductive therapies reduce the tumor mass or eliminate remaining cancer cells and metastases before or after surgery, respectively. However, it was found to also create an *in situ* vaccine. It can also reduce the amount of suppressive immune cells and it is described that this could increase vaccine efficacy.[53] Recently, Melief et al. reported prolonged survival rates in cervical cancer patients following combined treatment with cytoreductive chemotherapy and therapeutic vaccination in an open-label single-arm phase I/II study.[54] The median OS of the treated patients was 16.1 months, whereas 10–13.3 months had been reported in previous studies treating patients with chemotherapy only. A randomized trial should be performed to compare this combination treatment—or triple therapy when adding a checkpoint inhibitor—versus SoC to further investigate the modes of action and the true potential of adding vaccination to the combination therapy.

Neoantigen-specific response identification

The neoantigen-directed nature of personalized immunotherapies provides a unique opportunity to improve our understanding of therapeutically induced vaccine- and tumor-specific immunity. Novel technologies and the advancement of standard assays have vastly enriched the toolbox to assess vaccine-induced immune responses—both in the peripheral blood and the tumor—which are illustrated in Figure 2.

Immune assessments in the peripheral blood

Enzyme-linked immune absorbent spot (ELISPOT), multiplexed intracellular staining, and antigen-specific major histocompatibility complex (MHC) multimers are now widely available to assess frequencies, phenotypes, and function of vaccine-induced T cell responses over time (*ex vivo* or after *in vitro* stimulation). Bulk frequencies of T cells specific for individual vaccine epitopes can be reliably measured by IFN-γ ELISPOT, while expression of activation markers, such as 41BB and the degranulation marker CD107a, allow the assessment of activation status and cytotoxic potential of vaccine-induced T cells. Bulk TCR sequencing can also assess quantitative changes in T cell clones. These tools can therefore provide valuable information on the dynamics of circulating T cell populations, including neoepitope-specific T cells over time. Of note, comprehensive and in-depth analysis of peripheral immune responses induced by multi-neoepitope vaccines typically requires large numbers of PBMC necessitating cell collection by leukapheresis.

Immunogenomic tools, including single-cell TCR sequencing (scTCR-seq) and single-cell RNA-seq (scRNA-seq) of MHC-multimer or IFN-γ catch-sorted T cells, enable the dissection of vaccine-specific T cell responses on a clonal level and can provide high-resolution information on the phenotypes and function of these cells.[26,55] However, confirming the specificity of individual TCRs to vaccine neoepitopes is currently still cumbersome and resource intense. High-dimensional single-cell transcriptomics of $CD45^+$ sorted immune cells in the peripheral blood over time also allows in-depth characterization of immune cell populations and subsets in the context of vaccination.

Immune assessments in the tumor

While the analysis of circulating T cell responses is highly informative, interrogating TILs and other immune cell populations in the tumor after vaccination and linking this information to the clinical context likely provides the most relevant information with regard to the efficacy of a personalized neoantigen vaccine. Tumor sample quantity and quality are often limiting; however, new technologies allow the extraction of high-dimensional information from small amounts of tumor material, such as core biopsies. Multiplexed immunohistochemistry and immunofluorescence approaches enable thorough characterization of tumor immune infiltrates, including their spatial resolution from FFPE samples. Changes in TCR clonotype frequencies of TILs between pre- and post-vaccination time points can be assessed using paired scTCR-seq on the fresh or fresh-frozen tumor core biopsies allowing the tracking of individual clonotypes over time. scRNA-seq provides information on the cellular states and function of T cells bearing the individual TCR clonotypes. scRNA-seq also allows the dissection of the TME more broadly by providing

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

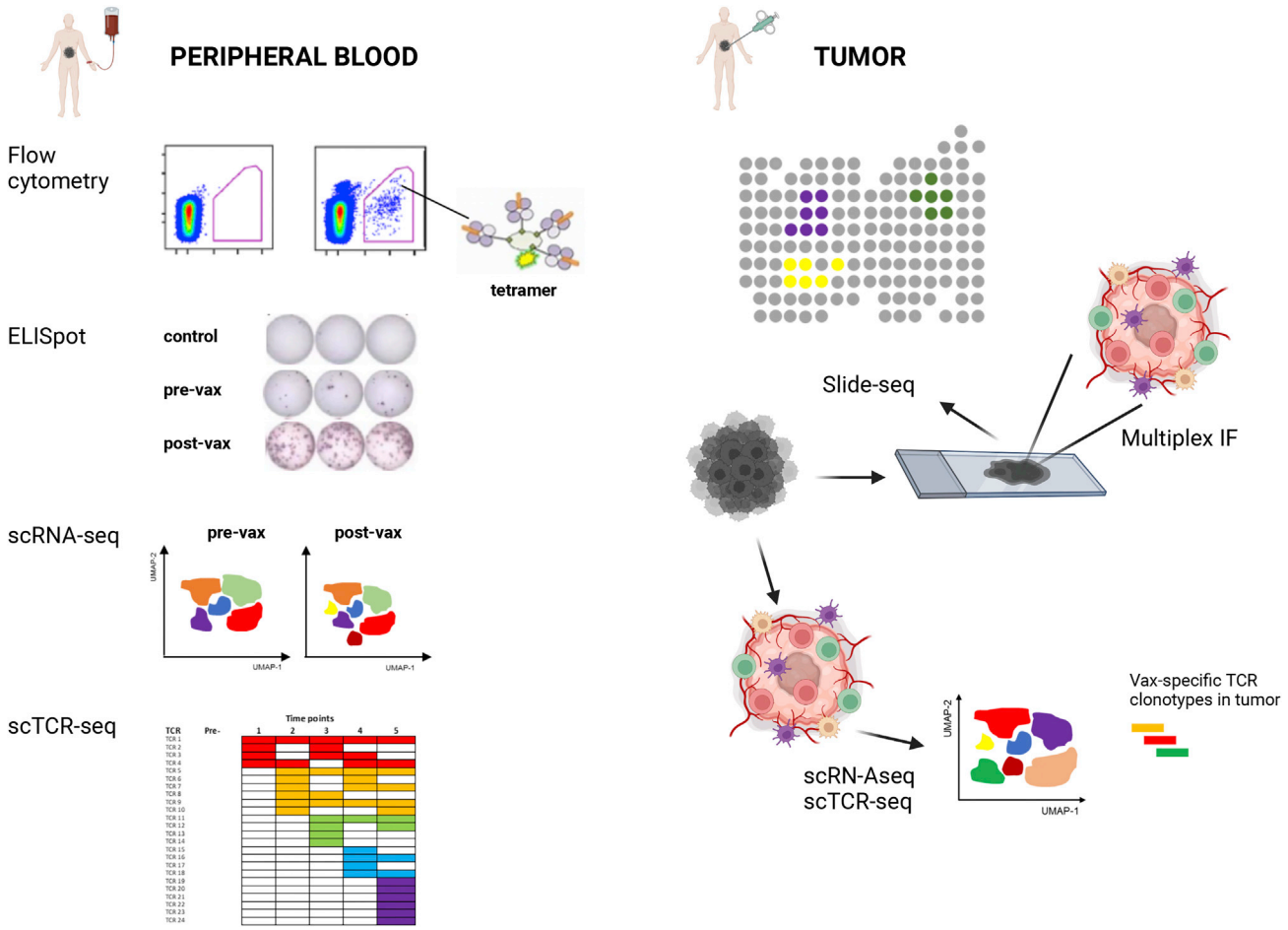

Figure 2. Conventional assays and immunogenomics tools available for the analyses of neoantigen vaccine-induced immune responses in the peripheral blood and tumor

information on the tumor and stromal cells, as well as individual immune cell subsets in the context of vaccination. Information on the geographic distribution of these different cell populations within the tumor at single-cell resolution can be generated through spatial transcriptomics approaches, such as Slide-seq or high-definition spatial transcriptomics, which capture mRNA from tissue sections onto a surface with DNA-barcoded microparticles with defined localization.[56]

THE BIOINFORMATICS CHALLENGE

Although neoantigen-based cancer therapies are a promising innovative approach to attacking and killing tumor cells, their effectiveness is highly dependent on the identity and nature of the selected neoantigens (Figure 1). Such neoantigens need to abide by several conditions to be applicable. Firstly, they have to be tumor specific to prevent off-target effects impacting a patient's healthy tissues. Secondly, they need to be presented on a tumor's cell surface so that immune cells can recognize and kill the associated tumor cell; and, finally, potent neoantigens must be able to trigger an immune response—i.e., they need to be immunogenic—when they are presented by MHC complexes to TCRs associated with immune cells. Separating such neoantigens from the complete set of tumor peptides in the patient is no simple task in which multiple factors, such as the patient's human leukocyte antigen (HLA)-type, tumor cell heterogeneity, and intrinsic features associated with peptide processing, need to be taken into account. In the next sections, all steps and corresponding tools required to elucidate and prioritize the most effective neoantigens are described in depth.

Current-standard neoantigen prediction wireframe

The standard neoantigen prediction setup consists of four steps: raw sequencing data pre-processing, followed by HLA typing; read mapping and variant calling; prediction of MHC binding and presentation; and neoantigen prioritization and selection. With the advent of next-generation sequencing and the ever-decreasing sequencing cost, whole-exome sequencing (WES) data for matched normal and tumor DNA is becoming routinely available for cancer patients. Such data are the primary input for any prediction pipeline. Whole-genome sequencing (WGS) can be used instead of WES; it results in more sensitive and robust variant calls but comes at a higher cost. WGS allows for high-confidence screening of a larger potential source of neoantigens, such as gene fusion events and chromosomal

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

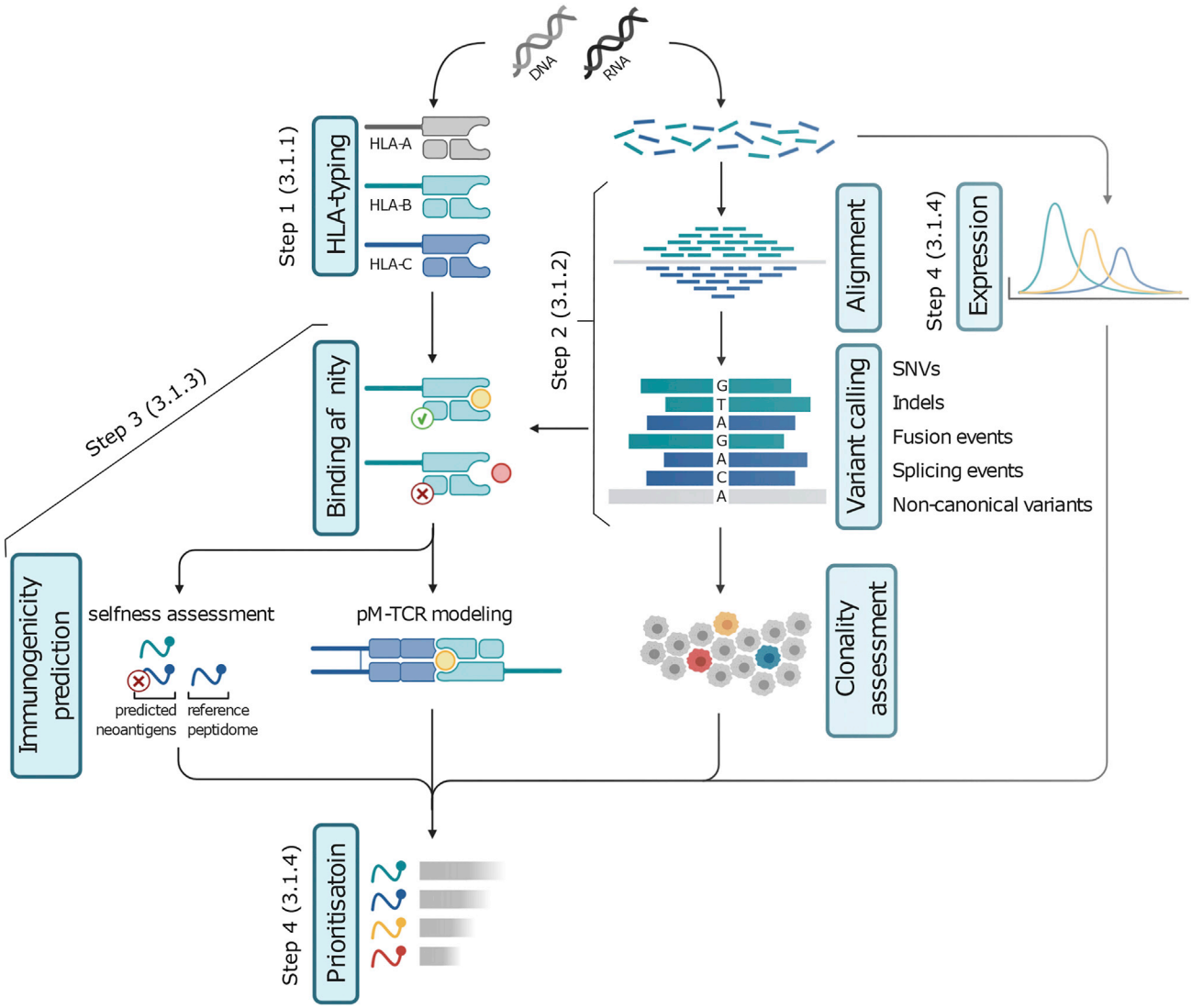

Figure 3. Illustration of all necessary steps for optimal neoantigen prediction and selection

The four major steps include HLA-typing, read mapping and variant calling, MHC binding/immunogenicity prediction, neoantigen prioritization—using host-gene expression and variant clonality assessment among others—and selection.

rearrangements, and it yields more reliable copy-number estimates.[57] Figure 3 depicts an overview of the steps required for optimal neoantigen prediction and selection, while an overview of all tools and software packages described in this section can be found in Table 1 and Box 2. Other steps that can strongly increase the breadth and precision of effective neoantigen discovery are described in the five next sections below.

HLA typing

Neoantigens must bind to and be presented by MHC molecules to elicit a T cell-driven immune response. The HLA genes, coding for these complexes, are highly polymorphic and patient specific. Actionable neoantigen prediction is therefore largely dependent on the correct inference of the HLA haplotype. Most of the pipelines focus on MHC class I (MHC-I) prediction using DNA-derived sequencing data as input. Despite its importance, MHC class II (MHC-II) prediction is less common in current prediction pipelines (Section "MHC-II epitopes"). Although the highly polymorphic nature of the HLA gene region across the population complicates HLA read mapping, some of these tools can reach up to 99% prediction accuracy in comparison with HLA-specific typing assays used in the clinic.[134]

Overall, most software tools for HLA-I typing show good performance. The most commonly used packages include OptiType[135] and PolySolver,[65] but accuracies exceeding 95% have also been reported for other packages, including HISAT-genotype, HLA*PRG, HLA-HD, HLA-LA, and xHLA (Table 1).[61,64,136–138] RNA-seq data can also be used for HLA-typing, with well-performing tools, such as ArcasHLA, Opti-Type, PHLAT, and Seq2HLA (Table 1),[58,135,139,140] although studies have shown a lower performance compared with DNA sequencing.[135] In contrast, an ensemble approach, where combining RNA- and DNA-based HLA typing using several algorithms, further increases precision and is recommended for automated pipelines. A comprehensive overview of all HLA-I and HLA-II typing tools,

Table 1. Overview of the HLA-typing, and MHC-I and -II binding predictors

Overview of the HLA-typing predictors[a]

	Input requirement, application (MHC-I or -II) and resolution of the tool			Accuracy values in the corresponding study (%), when using a specific (RNA/WGS/WES) input type for MHC-I or -II typing						Study info	
				MHC I			MHC II				
	Input type	Class	Resolution*	RNA	WGS	WES	RNA	WGS	WES	Year	Reference
ArcasHLA	RNA	I & II	4	97.73–100			94.13			2020	Orenbuch et al.[58]
						98.30				2021	Lee et al.[59]
				99.40			98.10			2022	Claeys et al.[60]
Athlates	DNA	I & II	4			98.50			100	2016	Xie et al.[61]
HISAT-genotype						99.33			94.55	2020	Orenbuch et al.[58]
					~100			~100		2021	Liu et al.[62]
HLA*PRG	DNA	I & II	6		99.50	83.30		98.07	93.00	2016	Xie et al.[61]
					100			98.33		2018	Lee et al.[63]
					100	83.33		98.67	94.67	2019	Dilthey et al.[64]
HLAforest	RNA	I & II	4			42.10					Shukla et al.[65]
					~70.00			~70.00		2021	Liu et al.[62]
				84.20		61.30				2014	Bai et al.[66]
HLA-HD	DNA/RNA	I & II	6			~98.00			90.04	2021	Li et al.[67]
						94.10				2021	Lee et al.[59]
					~100			~100		2021	Liu et al.[62]
				98.00			99.40		96.20	2022	Claeys et al.[60]
HLA-LA	DNA	I & II	6			96.00				2021	Lee et al.[59]
						94.40			96.70	2022	Claeys et al.[60]
					100	91.00		98.67	95.33	2019	Dilthey et al.[64]
HLAminer	DNA/RNA	I & II	4			43.80					Shukla et al.[65]
					~69.00			~69.00		2021	Liu et al.[62]
									52.80	2022	Claeys et al.[60]
				39.80		54.15				2014	Bai et al.[66]
HLAProfiler	RNA	I & II	4	99.50						2020	Orenbuch et al.[58]
HLAscan	DNA	I & II	4						74.20	2022	Claeys et al.[60]
HLA-VBSeq	DNA	I & II	6		97.50			38.00		2016	Xie et al.[61]
						~63.00			~40.00	2021	Li et al.[67]
					~35.00			~35.00		2021	Liu et al.[62]
									60.20	2022	Claeys et al.[60]
Kourami	DNA	I & II	6			~82.00			~60.00	2021	Li et al.[67]
						72.10				2021	Lee et al.[59]
					100			100		2018	Lee et al.[63]
					98.67	91.67		100	96.67	2019	Dilthey et al.[64]

(*Continued on next page*)

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

Table 1. *Continued*

Overview of the HLA-typing predictors[a]

	Input requirement, application (MHC-I or -II) and resolution of the tool			Accuracy values in the corresponding study (%), when using a specific (RNA/WGS/WES) input type for MHC-I or -II typing						Study info	
				MHC I			MHC II				
	Input type	Class	Resolution*	RNA	WGS	WES	RNA	WGS	WES	Year	Reference
Opti-Type	DNA/RNA	I	4			97.70					Shukla et al.[65]
				96.42		99.11				2021	Yi et al.[68]
						97.20				2017	Nam et al.[57]
					97.00					2016	Xie et al.[61]
						95.40				2021	Li et al.[67]
				95.70–99.67		97.87				2020	Orenbuch et al.[58]
						89.40				2021	Lee et al.[59]
				99.20		98.00				2022	Claeys et al.[60]
PHLAT	DNA/RNA	I & II	6			84.00					Shukla et al.[65]
				84.52		93.80				2021	Yi et al.[68]
						85.60				2017	Nam et al.[57]
					81.85	87.70		85.45	87.30	2016	Xie et al.[61]
				95.70						2020	Orenbuch et al.[58]
						93.00				2021	Lee et al.[59]
				95.40			98.90			2022	Claeys et al.[60]
				92.30		94.15				2014	Bai et al.[66]
					66.67			86.67		2018	Lee et al.[63]
PolySolver	DNA	I	4			97.00					Shukla et al.[65]
						95.80				2021	Yi et al.[68]
						94.00				2017	Nam et al.[57]
						94.90				2021	Li et al.[67]
						94.90				2022	Claeys et al.[60]
Seq2HLA	RNA	I & II	4			62.10					Shukla et al.[65]
				91.07						2021	Yi et al.[68]
				96.17						2020	Orenbuch et al.[58]
						97.30				2021	Lee et al.[59]
					~80.00			~80.00		2021	Liu et al.[62]
				96.70			87.90			2022	Claeys et al.[60]
				~32.00						2014	Bai et al.[66]
SOAPHLA	DNA	I & II			91.00			97.00		2016	Xie et al.[61]
						96.40				2013	Cao et al.[69]
						~80.00			~84.00	2021	Li et al.[67]

(*Continued on next page*)

Table 1. ***Continued***

Overview of the HLA-typing predictors[a]											
	Input requirement, application (MHC-I or -II) and resolution of the tool			Accuracy values in the corresponding study (%), when using a specific (RNA/WGS/WES) input type for MHC-I or -II typing						Study info	
				MHC I			MHC II				
	Input type	Class	Resolution*	RNA	WGS	WES	RNA	WGS	WES	Year	Reference
xHLA	DNA	I & II	4		99.90	99.27		98.07	98.70	2016	Xie et al.[61]
						~93.00			70.80	2021	Li et al.[67]
					100	94.00		100	96.50	2019	Dilthey et al.[64]

Overview of the MHC-I and -II binding predictors[b]									
				Performance (%) (AUC/PPV) when applying the tool on a specific benchmarking set				Study info	
	Contents of the training set	Modeling approach	MHC class compatibility	AUC^{ROC}	AUC^{PR}	PPV		Year	Reference
MHC binding affinity predictors (trained on binding affinity data)									
ARB	Allele	Scoring matrix	II	0.91				2020	Paul et al.[231]
				0.70				2009	Nielsen et al.[232]
				0.72				2019	Liu et al.[233]
				0.72				2020	Heng et al.[234]
				~0.80				2020	Saxena et al.[235]
				0.66				2017	Hu et al.[236]
DeepMHC	Pan	DCNN	I	0.73				2017	Hu et al.[236]
DeepMHCII	Pan	DCNN	II	0.86				2021	You et al.[210]
DeepSeqPan	Pan	DCNN	I	0.74				2019	Liu et al.[233]
				0.74				2020	Heng et al.[234]
DeepSeqPanII	Pan	DCNN	II	0.78				2021	Liu et al.[230]
				0.76				2021	You et al.[210]
Heng et al.	Pan	ANN	I	0.72				2020	Heng et al.[234]
KISS	Pan	SVM	I	~0.90				2009	Zhang et al.[145]
				0.73				2012	Zhang et al.[145]
MHC2pred	Pan	SVM	II	0.59				2009	Nielsen et al.[232]
MHCAttnNet	Pan	ANN	I & II	0.89	0.94	0.97		2020	Venkatesh et al.[207]
NetMHC	Allele	ANN	I	0.93				2017	Bhattacharya et al.[237]
				0.76		0.62		2020	Shao et al.[238]
				0.98				2020	Paul et al.[231]
				0.76				2019	Liu et al.[233]
				0.76				2020	Heng et al.[234]
				0.82	0.90	0.95		2020	Venkatesh et al.[207]
					0.51			2019	Boehm et al.[239]
				0.88				2018	Zhao et al.[150]

(*Continued on next page*)

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

Table 1. *Continued*

Overview of the MHC-I and -II binding predictors[b]

	Contents of the training set	Modeling approach	MHC class compatibility	Performance (%) (AUC/PPV) when applying the tool on a specific benchmarking set			Study info	
				AUC^{ROC}	AUC^{PR}	PPV	Year	Reference
NMER	Pan	RF	I	0.91			2021	Gartner et al.[154]
NN-align	Pan	ANN	II	0.62			2019	Chen et al.[151]
				~0.80			2021	Liu et al.[230]
				0.82			2009	Nielsen et al.[232]
				0.91			2019	Zhao et al.[150]
OnionMHC	Allele	ANN	I	0.83			2020	Saxena et al.[235]
PickPocket	Pan	scoring matrix	I	~0.90			2009	Zhang et al.[145]
				0.80			2020	Heng et al.[234]
				0.80			2019	Liu et al.[233]
				0.80			2018	Zhao et al.[150]
				~0.82			2020	Saxena et al.[235]
RANKPEP		scoring matrix	I & II	0.88			2020	Paul et al.[231]
				0.64			2009	Nielsen et al.[232]
SMM-align	Allele	scoring matrix	II	0.64			2019	Chen et al.[151]
				~0.71			2021	Liu et al.[230]
				0.73			2009	Nielsen et al.[232]
				0.77			2019	Liu et al.[233]
				0.77			2020	Heng et al.[234]
				0.87			2018	Zhao et al.[150]
				0.71			2017	Hu et al.[236]
Smmpmbed	Allele	SMM	I	0.92			2017	[237]
				0.88			2020	Paul et al.[231]
				0.84			2019	Liu et al.[233]
				0.84			2020	Heng et al.[234]
				0.86			2018	Zhao et al.[150]
				~0.81			2020	Saxena et al.[235]
TEPITOPE	Allele	scoring matrix	II	~0.38			2021	Liu et al.[230]
				0.59			2019	Chen et al.[151]
MHC presentation predictors (trained on ligandomic/MS data)								
EDGE		ANN	I			0.37	2022	Mill et al.[157]
EPIP	Pan	scoring matrix	I			0.52	2020	Weipeng et al.[240]
FIONA-P	Pan	CNN	II	0.91	0.89	0.22	2022	Shi et al.[114]
ForestMHC	Pan	RF	I		0.73		2019	Boehm et al.[239]

(*Continued on next page*)

Table 1. ***Continued***

Overview of the MHC-I and -II binding predictors[b]

	Contents of the training set	Modeling approach	MHC class compatibility	Performance (%) (AUC/PPV) when applying the tool on a specific benchmarking set			Study info	
				AUC^{ROC}	AUC^{PR}	PPV	Year	Reference
HLAthena	Pan	ANN	I	0.77			2020	Saxena et al.[235]
MARIA	Pan	ANN	II	0.89			2019	Chen et al.[151]
				0.87	0.84	0.33	2020	Shi et al.[114]
MHCflurry	Pan	ANN	I	0.93		0.73	2020	Reynisson et al.[153]
				0.79		0.62	2020	Shao et al.[238]
				0.93			2017	Bhattacharya et al.[241]
				0.98			2020	Paul et al.[231]
						~0.40	2021	Pyke et al.[152]
						0.24	2020	Weipeng et al.[240]
				~0.96			2019	Phloyphisut et al.[242]
				0.83			2019	Liu et al.[233]
				0.83			2020	Heng et al.[234]
				0.82	0.91	0.97	2020	Venkatesh et al.[207]
				0.91			2018	Zhao et al.[150]
						0.18	2022	Mill et al.[157]
				0.76			2021	Gartner et al.[154]
MHCnuggets	Pan	ANN	II	0.74		0.63	2020	Shao et al.[238]
				0.93			2017	Bhattacharya et al.[237]
MHCSeqNet	Pan	ANN	I	~0.96–0.98			2019	Phloyphisut et al.[242]
MixMHC2pred	Allele	ANN	II	0.81			2019	Racle et al.[209]
				0.89	0.88	0.28	2022	Shi et al.[114]
					0.74		2019	Boehm et al.[239]
				0.84			2018	Zhao et al.[150]
				0.93		0.77	2020	Reynisson et al.[153]
neoMS	Pan	ANN	I			0.61	2022	Mill et al.[157]
NetMHCIIpan	Pan	ANN	II	0.73			2019	Racle et al.[209]
				0.61			2019	Chen et al.[151]
				0.76	0.74	0.13	2022	Shi et al.[114]
				~0.82			2021	Liu et al.[230]
				0.89			2018	Zhao et al.[150]
				0.85			2021	You et al.[210]

(*Continued on next page*)

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

Table 1. *Continued*

Overview of the MHC-I and -II binding predictors[b]

	Contents of the training set	Modeling approach	MHC class compatibility	Performance (%) (AUC/PPV) when applying the tool on a specific benchmarking set			Study info	
				AUC^{ROC}	AUC^{PR}	PPV	Year	Reference
NetMHCpan	Pan	ANN	I	~0.92			2009	Zhang et al.[145]
				0.82			2012	Zhang et al.[145]
				0.95		0.79	2020	Reynisson et al.[153]
				0.78		0.63	2020	Shao et al.[238]
				0.93			2017	Bhattacharya et al.[237]
				0.98			2020	Paul et al.[231]
						~0.47	2021	Pyke et al.[152]
						0.37	2020	Weipeng et al.[240]
				~0.86			2019	Phloyphisut et al.[242]
				0.76			2019	Liu et al.[233]
				0.76			2020	Heng et al.[234]
					0.65		2019	Boehm et al.[239]
				0.87			2018	Zhao et al.[150]
				0.81			2020	Saxena et al.[235]
				0.72			2017	Hu et al.[236]
				0.78			2021	Gartner et al.[154]
SHERPA	pan	GBDT	I			~0.62	2021	Pyke et al.[152]

AUC^{ROC}, AUC under the ROC curve; AUC^{PR}, AUC under the precision-recall curve; PPV, positive predictive value; DCNN, deep convolutional neural network; ANN, artificial neural network; CNN, convolutional neural network; RF, random forest; SVM, support vector machines; GBDT, gradient boost decision tree.

*The number of digits with which a tool is able to predict an HLA type determines its resolution. Four-digit resolution means that HLA types are predicted up to the level of a synonymous DNA mutation in the HLA coding region (e.g., HLA-A*02:101:01), while at six-digit resolution differences in non-coding HLA regions and changes in expression are reported as well (e.g., HLA-A*02:101:01:02N).

[a]HLA-typing tools are annotated by input type and type of HLA molecule for which they can be used. Where available performance/accuracy have been included for the specific input and HLA types.

[b]MHC-I and -II binding predictors are annotated by approach, i.e., the prediction model and the content of the training set—either a single (allele) or multiple (pan) HLA-alleles—and the type of MHC molecule for which they can be used. Where available performance values (AUC^{ROC}, AUC^{PR}, and PPV) have been included. In the panel they are grouped based on the data they have been trained on: binding affinity predictors were trained on binding affinity data, while presentation predictors were trained on ligandomics and/or MS data.

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

BOX 2. Overview of the tools mentioned in section "The bioinformatics challenge"

Tools are grouped by application (e.g., expression analysis, variant calling, etc.), and where possible a short description of the applicability, the approach and prediction model used, and the performance values (AUC/accuracy) are included. HMM, hidden-Markov model; BMM, Bayesian mixture model; MLMDP, maximum likelihood mixture decomposition problem; BIC, Bayesian information criterion; DPMM, Dirichlet process mixture model; LM, likelihood model; MCMC, Markov chain Monte Carlo; ANN, artificial neural network; SVM, support vector machine; LSTM, long- to short-term memory; CNN, convolutional neural network.

- Variant calling
  1. Mutect2,[70] SNV & small indel caling
  2. Strelka2,[71,72] SNV & small indel calling
  3. Varscan2,[73] SNV, small indel, & CNV calling
  4. SomaticSniper,[74] SNV calling
  5. Lancet,[75] small indel calling
  6. EBCall,[76] small indel calling
  7. Pindel,[77] large indel & structural variant calling
  8. GRIDSS,[78] large indel & structural variant calling
- Peptide proteolysis prediction
  1. NetChop20S,[79] MHC-I, trained on *in vitro* digestion data
  2. NetchomCterm,[79] trained on MHC-I ligand data
  3. ProteaSMM,[80] MHC-I, trained on *in vitro* digestion data
- TAP-associated transport prediction
  1. TAPREG,[81] SVM modeling
  2. PREDTAP,[82] MHC-I, ANN, & HMM modeling
- Expression estimation
  1. Kallisto[83]
  2. Sailfish[84]
- Gene fusion identification
  1. Pizzly,[85] requires kallisto output
  2. FusionCatcher,[86] Bowtie-based read-alignment followed by ensemble-based fusion identification
  3. STAR-fusion,[87] identifies chimeric junctions using STAR-based split reads
  4. GFusion,[88] TopHat-based read-alignment
  5. INTEGRATE,[89] graph-based identification following read-alignment
  6. CICERO,[90] read-alignment followed by two rounds of BLAT mapping
  7. JAFFA,[91] Bowtie-based read-alignment followed by BLAT & Oases transcriptome mapping
  8. Arriba,[92] Based on the STAR aligner, can be used for other structural rearrangements as wel
- Alternative splicing detection
  1. DEXSeq,[93] exon-based
  2. EdgeR,[94] exon-based
  3. SplAdder,[95] event-based
  4. Limma,[96] exon-based
  5. KeepMeAround,[97] event-based
  6. MAJIQ,[98] event-based
  7. AIDE,[99] event-based
  8. SUPPA2,[100] event-based
- Clonality assessment
  1. PyClone,[101] BMM & VAF-based modeling
  2. PhyloWGS,[102] MCMC/BIC & VAF-based modeling
  3. Meltos,[103] Bayesian DPMM & VAF-based modeling
  4. SVclone,[104] Dirichlet process & VAF-based modeling
  5. TUSV,[105] LM & VAF-based modeling
  6. WEAVER,[106] Dirichlet process & VAF-based modeling
  7. PhyloSub,[107] DPMM & VAF-based modeling
  8. CloneHD,[108] HMM/BMM & VAF-based modeling
  9. SciClone,[109] HMM/BMM & VAF-based modeling
  10. CaSpER,[110] HMM & CNV-based modeling
  11. ExPANdS,[111] CNV-based modeling
  12. THeTa,[112] MLMDP/BIC & CNV-based modeling
  13. TITAN,[113] HMM & CNV-based modeling

(Continued on next page)

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

BOX 2. *Continued*

- pMHC-TCR/immunogenicity binding prediction
  1. FIONA-I,[114] MHC-II, sequence-based, CNN modeling
  2. PRIME,[115] combines MHC-I presentation & peptide-intrinsic TCR recognition
  3. neoIM,[116] HC-I, sequence-based, random forest modeling (AUC = 0.88)
  4. Repitope,[117] MHC-I & -II, sequence-based, linear contact modeling
  5. pTuneos,[118] MHC-I, sequence-based random forest modeling
  6. Neopepsee,[119] MHC-I, sequence-based, locally weighted naïve Bayes modeling (AUC = 0.98)
  7. NetTCR,[120] sequence-based (peptide, TCRβ), CNN modeling (AUC = 0.73)
  8. TCRex,[121] sequence-based (TCRβ), random forest modeling
  9. ATMTCR,[122] sequence-based (peptide, TCRα/β)
  10. TCRdist,[123] HLA-2 specific, sequence- and structure-based (peptide, TCRα/β) (AUC = 0.72–0.82)
  11. SwarmTCR,[124] sequence-based (peptide, TCRα/β) (AUC = ∼0.83)
  12. SETE,[125] sequence-based (peptide, TCRβ), gradient-boosting decision tree modeling (AUC = ∼0.8)
  13. ERGO,[126] sequence-based (peptide, TCRα/β), LSTM modeling (AUC = 0.70)
  14. ERGO-II,[126] sequence-based (T cell and MHC type, peptide, TCRα/β), LSTM modeling (AUC = 0.70)
  15. TCRFlexDock,[127] structure-based, applies Rosetta
  16. GLIPH,[128] sequence-based (TCRα/β), TCR clustering modeling
  17. DynaDom,[129] structure-based (pMHC, Vα/β)
  18. TCRpMHCmodels,[130] sequence- and structure-based (MHC, peptide, TCRα/β)
  19. Dhusia et al.,[131] sequence- and structure-based, random forest modeling (accurace = 0.62)
  20. Milighetti et al.,[132] sequence- and structure-based (peptide, TCRα/β) (AUC = ∼0.70)
  21. Riley et al.,[132] HLA-A2 specific, applies Rosetta and Monte-Carlo-based conformational sample and energy minimization (AUC = 0.73)
  22. NetTCR2,[133] HLA-2 specific, sequence-based, CNN modeling (peptide, TCRα/β) (AUC = 0.89)

their required input type, and accuracy values in corresponding comparison studies can be found in Table 1.

Next to allelic loss of HLA, aberrations in or reduced expression at HLA gene loci strongly impact neoantigen presentation and often cause tumor immune evasion.[65,141] Thus, while prioritizing the predicted neoantigens, expression levels and mutational events of HLA genes are often considered. PolySolver can identify genetic changes at these loci, while the R package LOHHLA elucidates HLA copy-number variations resulting in loss of heterozygosity, as often observed in, e.g., lung cancer.[141]

Read mapping and variant calling

Following HLA typing, the sequencing reads are mapped on a human reference genome, with an appropriate aligner. For DNA- and RNA-seq, Burrows-Wheeler and STAR are frequently used aligners.[142,143] This is followed by the characterization of tumor-specific variants, which involves the screening of exome- or genome-wide tumoral mutations, extended with all other events known to affect the antigen repertoire of a tumor cell. Most commonly, however, the focus is limited to single nucleotide variants (SNVs), insertions, and deletions (indels). Oftentimes it is crucial to also consider other types of tumor variants that have been shown to result in neoantigens, as described further in section "Challenging low-mutation tumors: digging deeper".

Indels frequently result in a shifted reading frame which heavily impacts the ensuing peptides and thus deliver potentially higher and more immunogenic epitopes than simple SNVs. Examples of variant callers are Mutect2,[70] Strelka(2),[71] Varscan(2),[73] and SomaticSniper.[74] Of these, the former two are best suited for detecting SNVs at low allele fractions and provide the most reliable calls as validated by Sanger sequencing (MuTect2, Strelka, Varscan2, and SomaticSniper showed validation percentages of 89.2%, 72.3%, 35.4%, and 32.3%, respectively).[144] Furthermore, by supplying a set of germline variants and false-positive mutations identified in normal samples, Mutect2 can filter out non-somatic variants and sequencing and alignment artifacts.[71]

Neoantigen binding affinity and presentation prediction

Presentation of the neoantigens on the patient's MHC molecules is crucial for efficient T cell recognition. All presentation prediction models start by splitting the mutant protein into k-mer peptides using a sliding window with peptide lengths commonly ranging from 8 to 11 amino acids (AAs) for MHC-I. These are fed into the prediction models that can be subdivided based on the machine learning approach and the type of training data. Linear regression-based models, such as PickPocket and smmpmbec, assume that each AA residue contributes linearly to the binding affinity, while more recent methods rely on more sophisticated neural networks.[145–149] Although the latter requires more training data, they model the non-linear binding affinities between the peptide sequence and the MHC complex better than linear regression.[150]

Two main strategies have emerged to address HLA diversity in the context of presentation prediction. Allele-specific models have been trained on data corresponding to a single HLA type. While they usually perform very well on data matching the individual HLA types they were trained on, they underperform or even cannot be used on HLA types absent from their training data, such as rare or novel HLA types. In contrast, pan-HLA models are trained on data obtained from multiple HLA types. As such, they are able to learn cross-HLA type features important for neoantigen binding and are thus better equipped for predictions involving previously unseen HLA types. Most of the commonly

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

used and more recently implemented models, such as MARIA, MHCflurry, NetMHCpan, NMER, neoMS, and SHERPA, have followed the latter approach, as can be seen in Table 1.[148,151–154]

Earlier algorithms were mostly trained on data obtained through *in vitro* peptide-MHC binding affinity assays and did not account for a number of factors potentially affecting peptide presentation. These include peptide/MHC (pMHC) stability, peptide degradation, intracellular processing, and transport. Overall, this results in an increased level of false-positive peptide presentation predictions as strong MHC binders can still result in low presentation due to inadequate proteasomal degradation and TAP-mediated transport. This was demonstrated by Bassani-Sternberg et al.,[155] who showed that only 10% of peptides predicted as presented by this first generation of algorithms can be found by mass spectrometry (MS) on the surface of the cell.[22,24,155] For this reason, when using binding affinity-trained models in prediction workflows, it is important to include tools trained on *in vitro* proteasome digestion data, among others. Examples include NetChop20S, NetChomCterm, and Protea SMM for MHC-I-compatible peptide proteolysis.[79] NetChop Cterm—which is trained on *in vivo* MHC-I ligand data—generally results in the most accurate predictions,[79] while PRED^{TAP} and TAPREG can be used for TAP-associated transport simulations.[81,82]

More recent methods include *in vivo* ligandome data, generated from pMHC-I immunoprecipitation followed by MS. As such, the overall peptide presentation process is considered at once, giving rise to more accurate presentation estimates.[147–149] NetMHCpan4, MHCflurry, EPIP, EDGE, and MHCnuggets are commonly used neoantigen presentation predictors.[147–149,151,156] The former two show the highest accuracy in presentation versus non-presentation classifications, with area under the curve (AUC) values commonly exceeding 0.9 (Table 1). However, initial results of novel methods, such as neoMS, NMER, and SHERPA, seem to perform similarly well or even better, but still need more comparative studies to better benchmark them against the more established tools.[152,154,157] In general, consensus between methods is quite low, making decisions based on consensus estimates from multiple predictors difficult.[158] A general summary of the available MHC-I and MHC-II binding prediction tools is available in Table 1, showing the type of data they are trained on (pan- versus allele-specific, MHC-I versus MHC-II), the approach used, and the AUC or positive predictive values in the respective studies.

Neoantigen prioritization

The final step in any prediction pipeline is neoantigen prioritization. Current assumptions indicate that MHC-I binding of the neoantigen might be the strongest contributor to neoantigen immunogenicity.[159] Hence, most prediction workflows assign the highest weight to this parameter during prioritization. The copy-number status, as well as the expression level of the HLA alleles, to which the neoantigen is predicted to bind, can also be considered for prioritization because lowly expressed HLA genes will result in a limited presentation of the associated neoantigen.[160] In this context, neoantigens predicted to bind to multiple HLA alleles should receive a higher ranking. Other features, such as proteasomal degradation and TAP-based transportation, have shown a marginal impact when using MS-trained prediction models.[161]

A second aspect to consider during prioritization is the expression of the host gene harboring the mutation. This can be achieved by performing an RNA-seq-based gene expression analysis using commonly known tools, such as Kallisto, Salmon, and Sailfish, among others.[83,84,162] Gene expression-based filtering is important since, even if the mutation is present in the tumoral DNA, the absence of or limited active transcription of the host gene will inhibit the generation of mutated peptides to be presented by the cell's MHC molecules. In this context, protein quantification is equally important but more challenging and less commonly established in current neoantigen prediction pipelines.

Recently, the Tumor Neoantigen Selection Alliance published a set of features and corresponding thresholds most significantly contributing to neoantigen presentation and immunogenicity of MHC-I presented peptides.[163] In addition to binding affinity (<34 nM) and gene expression level (>33 TPM), it was found that hydrophobicity and pMHC stability (>1.4 h) are correlated with neoantigen presentation potential. For immunogenicity, foreignness related to immunogenic tolerance and agretopicity were identified as parameters with the highest impact, since ~50% of the elucidated immunogenic tumor epitopes in the study were associated with either low agretopicity or high foreignness.

Overview of publicly available prediction pipelines

Among the currently available neoantigen prediction pipelines, several trends and core components can be observed (Figure 4). The majority of the pipelines do not inherently perform WES/WGS pre-processing and variant calling. Instead, they rely on user-generated variant lists. Only pTuneos, TSNAD, OpenVax, nextNEOpi, and TruNeo can process raw sequencing input.[118,164–167] In contrast, a large number of workflows allow RNA-seq pre-processing. Most often this relates to Kallisto- or Sailfish-based analysis to prioritize neoantigens associated with high expression of the host gene, with only a few pipelines performing RNA-seq-based read alignment and variant calling.[83,84]

Most of the pipelines also implement HLA typing, often using Opti-Type with individual implementations of PolySolver, HLA-miner, SOAPHLA, and HLALOH.[65,135,141,168,169] Only Neoepiscope, neoPredPipe, and MuPeXI expect a user-supplied list of HLA alleles.[170–172] With some exceptions, all pipelines focus on SNVs and indels. NeoFuse, pVACtools, nextNEOpi, INTEGRATE-NEO, and TruNeo can handle fusion events, while ASNEO is compatible with splicing and intron retention events.[165,167,173–175] Variant phasing is only integrated into Neoepiscope and pVACtools, while approximately half of the pipelines take foreignness into account.[170,174] Finally, all pipelines implement MHC-binding analysis, underscoring the importance of this feature. All but one of the workflows apply NetMHCpan for this; only NeoFuse uses MHCflurry exclusively, while nextNEOpi applies both.[147,148,167,173] MHC-II compatibility, however, is limited to neoANT-HILL, nextNEOpi, and pVACtools.[167,174,176] The inclusion of proteomic data and peptide-related features is only performed by a selection of the pipelines.[174,176] In addition to the publicly available pipelines shown in Figure 4, other commercial and proprietary pipelines implementing a significant portion of the discussed features exist.

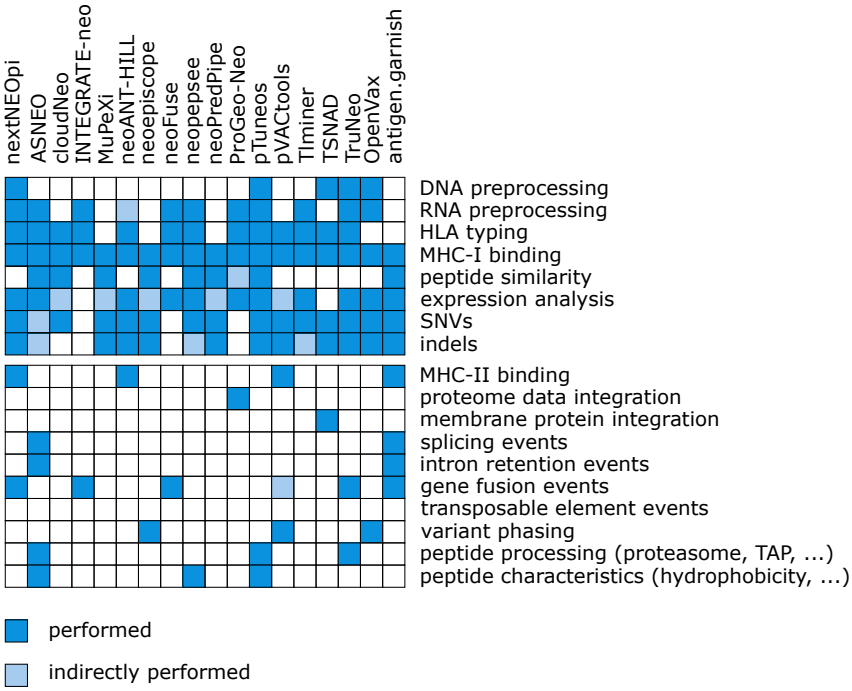

Figure 4. Overview of the currently available neoantigen prediction pipelines (vertical) and their focus points (horizontal)

Dark-shaded squares indicate that the corresponding analysis has been implemented directly in the pipeline (e.g., the pipeline directly calls Kallisto for expression analysis), while lighter-shaded squares indicate features where the pipeline relies on results generated independently from the pipeline that needs to be supplied by the user when running the pipeline (e.g., a list of gene expression values).

Challenging low-mutation tumors: Digging deeper

Finding immunogenic neoantigens in TMB tumors is challenging. It is therefore critical to consider the complete mutational repertoire and identify as many disparate peptides as possible. Indeed, Laumont et al. postulated that, although there are clear indications that the immune system can drive successful tumor elimination in tumors characterized by a low TMB, different and more thorough neoantigen discovery methods are required.[177] In this context, the inclusion of non-canonical variants (see Box 3) is important and the need for tools that allow the detection of an antigen landscape as broad as possible is high. This means exploring the domain of gene fusion events,[178] transposable element activity, and neoisoforms.[179,180] Tumor-specific gene fusion events or transposable element insertion into the coding sequence of a host gene will strongly perturb the nucleotide and peptide sequence of the corresponding gene, giving rise to a potentially large number of tumor-specific neoantigens. The same holds true for the generation of tumor-specific neoisoforms due to, for example, altered splicing patterns or the use of alternative translation start codons. Altered proteasomal degradation in tumor cells, on the other hand, can potentially generate protein degradation products absent in normal cells due to differential splicing patterns. In hematologic malignancies, taking into account immunoglobulin diversity and somatic hypermutation in B cells might also increase the pool of potential neoantigens in tumor types characterized by a low degree of mutations.[181]

Indels, gene fusions, and splicing variants

The best-performing callers for small indels are Lancet, Strelka2, and EBCall, the latter demonstrating the least sensitivity to coverage variability.[75,76,182] For the identification of larger indels (0.5–10 kb) and structural variants, Pindel and GRIDSS are ideally suited.[77,78] Gene fusions are a potent new neoantigen source and have been reported to drive the development of approximately 16% of all cancers.[183] A large set of tools is available for the elucidation of gene fusions based on WGS or RNA-seq data, including pizzly, FusionCatcher, STAR-fusion, GFusion, INTEGRATE, arriba, CICERO, and JAFFA.[85–92] Unfortunately, such tools still result in high false-positive rates, which could be addressed by only retaining predictions supported by multiple data types or methods.

Alternative splicing events have been shown to be 30% more common in tumors than in healthy samples and thus constitute another pool of potential new neoantigen candidates.[179] Splicing aberrations can be detected using RNA-seq by means of DEXSeq-, Limma-, or EdgeR-based differential expression analysis or through specialized event-based tools, such as SplAdder, KeepMeAround, MAJIQ, rMATS, AIDE, and SUPPA2.[94–100,184,185] Mehmood et al. showed that, in general, all methods scored well, with overall the exon-based methods, MAJIQ and rMATS, performing best in the context of both precision and qPCR-based validation of identified events.[186]

Non-canonical variants

In addition to mutation-driven events discussed above, tumoral alterations caused by aberrant expression should be examined as well. This requires a further extension of the variant search domain with non-canonical variants (see Box 3), such as transcripts with alternative start codons, uncanonical open reading frames (ORFs), and small ORFs (smORF), in for example long non-coding RNAs and 5′ untranslated regions.[187–189] These variants are potentially more valuable targets than classical mutations because they rather depend on the abnormal translation of a normal, unmutated part of the genome. As such, the same event is more likely to happen across several patients. Therefore, the resulting aberrantly expressed tumor-specific antigens (aeTSAs) are more likely to be shared, thus requiring a less individualized therapeutic approach. Moreover, it has been shown that non-mutated aeTSA events greatly outnumber mutated TSA events.[177]

Approaches integrating sequencing and MS ligandome datasets are essential for the inclusion of non-exonic variants in neoantigen prediction pipelines.[180,190,191] However, MS is currently still biased toward the elucidation of highly expressed peptides, resulting in an underestimation of the number of MHC-presented peptides.[104] Fortunately, approaches are being studied to

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

BOX 3. Canonical versus non-canonical variants

In biology, canonical refers to established or standard mechanisms or signaling pathways in normal cells, such as the known splicing pattern or translation start site of a gene. In tumor cells or cells under stress, these features can sometimes be activated or perturbed, resulting in aberrant transcript or protein structures, or mechanisms deviating from normal execution, most of the time resulting in disease. Such events are referred to as non-canonical events. For example, translation initiation canonically starts from an AUG start codon but, in some cases, non-AUG codons have also been shown to function as translation start sites. Translation starting from such non-canonical start sites will typically result in alternative peptide sequences. If this happens in tumor cells, the resulting peptides are likely to be specific to these cells and could be used as targets for cancer therapies.

remedy this.[192,193] Other newer technological developments, such as Ribo-sequencing, allow us to verify non-canonical gene expression to evaluate the expression of intronic regions, and to discover alternative ORFs within known genes. MS, on the other hand, could help reveal tumor-specific post-translational modifications and proteasome-generated spliced peptides presented *in vivo* by MHC complexes.[194,195] The inclusion of variants originating from non-canonical sources has the potential to significantly increase the pool of TSAs. Unfortunately, workflows to identify such variants are still less well established and result in high degrees of false-positive hits. Thorough validation of the identified variants and more robust identification algorithms are needed to increase the success rate of non-canonical- derived neoantigens before they can be routinely used in clinical applications.

Genetic variability (clonality): Genetic heterogeneity

It is essential to select antigens representative of the whole tumor, taking sub-clonality within the tumor and its metastases into account. In this respect, multi-region sequencing of representative biopsies is necessary to estimate if a given neoantigen is expressed by all tumor cells. Gerlinger et al. reported that intratumoral heterogeneity may foster tumor evolution, and indicated the inadequacies associated with sequencing a single tumor specimen as well as personalized therapy addressing only one neoantigen.[196] In addition, it is essential to implement *in silico* filtering to generate functionally important tumor alterations (driver mutations) with high allelic fractions to increase the probability of selecting highly clonal epitopes.

To aid the selection of clonal mutations, while discarding subclonal variants, various bioinformatics methods have been developed. In general, clonal evolution assessment can be performed using either copy-number changes or the variant allele fractions (VAFs). VAF-based methods rely on the fact that subclonal variants have lower fractions than clonal mutations in bulk sequencing datasets. However, VAFs can be affected by tumor purities and copy-number alterations (CNAs). As such, VAF-based clonal evolution assessments largely depend on the accuracy of CNA detection methods, such as PyClone, PhyloWGS, Meltos, SVclone, TUSV, WEAVER, and PhyloSub.[101–107] These methods apply the Dirichlet process, while CloneHD and SciClone use hidden Markov and Bayesian mixture models, respectively.[108,109] Copy-number-based modeling of sub-clonal populations uses the fraction of reads covering the alternative allele, which deviates from 0.5 in the case of gains or losses, corrected for tumor purity and ploidy. Methods applying this approach include CaSpER, ExPAndS, THeTa, and TITAN.[110–113] Although benchmark studies are still lacking, recent studies have shown THeTa and ExPANdS to perform well.[112,197] CNAs and (sub)clonal mutations can also be inferred from single-cell sequencing data; however, it should be noted that, with single-cell data, one must consider potential allelic dropout and lowly covered regions.[198]

In the clinic, multi-region biopsy sampling is often not feasible due to the size or inaccessibility of the tumor. In these cases, WES on liquid biopsies, complemented with sequencing of a matching normal sample, may prove a promising alternative.[199] Such an approach would enable the assessment of a patient's complete somatic tumor load in a simpler, less-invasive way, bypassing the need for sub-clonality analysis.

MHC-II epitopes

The presentation of MHC-I-restricted tumor antigens by dendritic cells is necessary to activate $CD8^+$ T cells.[200] The latter are highly efficient in recognizing and eradicating tumor cells that present the tumor antigens for which they have been primed.[201] This explains why there has been a long-standing belief in the field that $CD4^+$ T cells are not important for tumor control given that the majority of solid tumors lack MHC-II expression.[202] However, it has been shown that $CD4^+$ T cells have both indirect and direct effector functions and that targeting only MHC-I-restricted neoantigens could result in a suboptimal anti-tumor immune response.

The observation that MHC-II-mediated pathways could have a significant impact on current cancer treatments has triggered the development of workflows capable of predicting MHC-II-binding epitopes.[203] Unfortunately, these tools still show a lower performance when compared with MHC-I-driven models. This is largely because MHC-II molecules are heterodimeric glycoproteins consisting of a highly polymorphic alpha and beta chains. The higher number of HLA genes observed for MHC-II (6 versus 3 for MHC-I) makes them much more variable than MHC-I molecules. In addition, MHC-II complexes have open-ended binding grooves, allowing the binding of peptides to be more variable in both length (13–25 AA for MHC-II, compared with 8–15 AA for MHC-I) and sequence composition.[147,204]

As mentioned before, a crucial part of a neoantigen identification pipelines includes the HLA-typing and MHC-affinity binding prediction steps. Multiple tools are capable of typing both HLA-I and HLA-II, including ArcasHLA, seq2HLA, HLA*PRG, HLAforest, HLA-HD, HLA-LA, Kourami, xHLA, Athlates, PHLAT, and SOAPTyping.[58,61,63,64,137–140,169,205,206] Among these, xHLA, Kourami, HLA-LA, HLA-HD, and HLA*PRG perform best for DNA input types according to multiple studies (accuracy >90%–95%), while PHLAT, HLA-HD, and ArcasHLA show

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

good accuracies on RNA (although here the number of comparison studies is limited) (Table 1).[58,61,63,64,137–139]

Despite the large amount of experimental data, MHC-II-binding affinity predictors still fail in reaching adequate accuracy levels. The overall impact of incorporating MS data in these models is yet to be determined, but it is likely that models trained on such data could result in improved accuracy.[207] As the secondary structure of peptides was reported to play an important role in pre-processing prior to MHC-II-associated presentation, structure-based models and better delineation of binding footprints could further help refine presentation and binding predictors.[132,208] Currently established MHC-II-binding prediction tools apply either neural network-based approaches (including NetMHCIIpan, MixMHC2pred, FIONA-P, MHCAttnNet, DeepMHCII, MHCnuggets, MoDec, MARIA, and NN-align)[114,151,153,207,209–212] or utilize scoring matrices to assess binding (e.g., TEPITOPE, MHC2pred, SMM-align, ARB, and RANKPEP).[213–217] Overall, the former types of model are preferred, with FIONA-P, MARIA, MHCAttnNet, and MixMHC2pred all having AUC values well above 0.85 (Table 1).[114,151,207,209]

Immunogenicity prediction

Immunogenicity assessment can be done *in vitro* through ELISPOT assays (see section "Immune assessments in the peripheral blood") in which cytokine production by antigen-specific $CD8^+$ T cells is measured or by intracellular cytokine staining (peptide-HLA multimer staining).[218] *In vitro* immunogenicity assessment is an important tool, as over 90% of predicted neoantigens are estimated to be non-immunogenic.[163] However, although novel and optimized screening approaches are being explored,[219] these assays are generally time-consuming and expensive, further lengthening the manufacturing time and increasing the total production cost of the therapy. This highlights the need for novel, automated prediction approaches.

Research on *in silico* prediction of immunogenicity is still largely lacking. Since only a limited number of epitopes have been shown to elicit an immunogenic response, prediction accuracy could be improved by prioritizing dominant epitopes.[220] Most prediction workflows try to enrich for truly immunogenic neoantigen candidates by prioritizing predicted neoantigens based on dissimilarity to self (Figure 4). This is based on the observation that neoantigens, similar in sequence to native peptides, are likely to be subjected to central tolerance, while neoantigens without a native counterpart are more easily identified as foreign by the immune system.

In addition to peptide pre-processing and MHC-binding affinity, the peptide AA sequence is believed to significantly affect pMHC-TCR recognition and interaction as well as the degree of immunogenicity.[159,221] In addition, Capietto et al. showed that the position of a mutation within the peptide is important and that integrating this feature in neoantigen prediction pipelines resulted in a better ranking of the immunogenic neoantigen candidates.[222] Other studies have reported immunogenic peptides to be enriched in hydrophobic AA at TCR interaction sites and the importance of AA weight, size, and charge in establishing pMHC-TCR complexes.[119,163,223] This shows that modeling pMHC-TCR interactions should be possible. Standalone tools implementing such pMHC-TCR prediction models using known pMHC-TCR complexes include NetTCR(2), TCRex, ATMTCR, TCRdist, SwarmTCR, SETE, and ERGO(-II).[120–124,126,133,224] Other tools and pipelines model these interactions through machine learning by using confirmed immunogenic and non-immunogenic peptides as positive and negative training sets.[117,120,121,126] Examples are PRIME, neoIM, Ineo-Epp, Repitope, FIONA-I, pTuneos, and Neopepsee.[114–119,225] Unfortunately, because of limited validated pMHC-TCR interaction data, most models are still insufficiently trained and often produce inaccurate predictions.

The aforementioned methods attempt to model pMHC-TCR interactions based on the primary AA sequence of the peptide (and the individual AA characteristics), often combined with HLA typing and sometimes TCR sequence information. A better approach might be to extend these features with the impact of post-translational modifications on peptide structure and to integrate these as a whole. Indeed, since peptide sequence and individual AA characteristics largely contribute to secondary and tertiary peptide structures, structure-based models could help elucidate their impact on neoantigen immunogenicity potential.[127,132] Up to now, few attempts have been pursued, e.g., TCRFlexDock, GLIPH, DynaDom, and TCRpMHCmodels, and the models created by Dhusia et al., Milighetti et al., and Riley et al.[127–132,226] They generally rely on TCR, MHC, and peptide features located in close vicinity to each other, such as the TCR CDR3α, and CDR3β that have been shown to interact with the MHC-presented peptide in TCR-pMHC complexes. The downside of these approaches is that they highly rely on elucidated TCR-pMHC complexes, such as crystallographic structures to model the interaction. Such structures are technically difficult and time-consuming to obtain, therefore high-quality data to train the algorithms are still lacking, resulting in limited robustness and underperformance of structure-based binding predictors. Structure-based predictors are currently still outperformed by sequence-trained models, with the latter achieving—and depending on the test set sometimes exceeding—AUC values of 0.9, while the former commonly reach AUC values around 0.7 (Box 2). As such, more research and data are needed to assess the accuracy of structure-trained binding prediction workflows and improve adoption in neoantigen discovery pipelines.[127]

Neoantigen selection

Neoantigen prioritization is the final step of the neoantigen discovery process. It is crucial to filter out the unsuitable neoantigens and retain only highly confident and actionable candidates. For example, hundreds of high-affinity neoantigens can be predicted in a tumor sample, but peptide screens routinely detect T cell reactivity against only a limited set of neoantigens per tumor.[14]

In addition, it is expected that inclusion of subdominant neoantigens, i.e., antigens that require immunization as they do not naturally induce an immune response, into any therapeutic strategy will lead to a decreased selection of antigen loss variants.[227] This is of particular interest in the context of clonal tumor evolution where targeting single dominant oncogenes results in the development of multiple resistant clones,[22] while targeting multiple neoantigens may result in the downregulation of particular epitopes while others persist.[16] Also, in the context of tumor

Please cite this article in press as: Lybaert et al., Challenges in neoantigen-directed therapeutics, Cancer Cell (2022), https://doi.org/10.1016/j.ccell.2022.10.013

heterogeneity and sub-clonality (Section "Genetic variability (clonality): genetic heterogeneity"), it is essential to select a set of well-represented alterations that cover most, if not all, patient's cancer cells to prevent immune escape.

The size of the final neoantigen pool is another critical question to address. Targeting multiple epitopes at once will lead to a broader and potentially stronger immune response while limiting the chance of outgrowth of tumor cells that reverted the targeted alterations.[228] In addition, improper target selection and inclusion of falsely identified epitopes in therapy, as well as cross-reactivity of the neoantigen-specific T cells with the non-mutated version of the antigen, might not correctly redirect the immune system against the tumor. This could potentially bring the immune system toward imbalance.[229]

To weigh both factors, up to 20 neoantigens are commonly selected for neoantigen-driven immunotherapy, taking a cautious approach to maximize the true positive fraction of immunogenic neoantigens for each patient. Nevertheless, the optimal number of neoepitopes needed is still unknown.[230]

Conclusions

This review addresses bioinformatics and clinical challenges that lie ahead for neoantigen-directed immunotherapy to unlock its full potential. An overview of the necessary steps is illustrated in Figure 1.

Neoantigen identification is therein a challenging task and different hurdles need to be taken to select immunogenic epitopes. Correct identification seems to be essential either as a target for cancer vaccination or to selectively expand reactive T cells for ACT, or to predict responses toward ICIs. Neoantigen prediction has become more accurate over recent years and has shifted to include new types of alterations, such as gene fusions, splicing events, and non-canonical ORFs. Unfortunately, most neoantigen prediction toolsets are still compromised by a large number of false positives, unable to elicit an immunogenic response, which diminishes the chances of therapy success or makes the validation process a time-consuming and costly process. An improved investigation and centralization of the different translational datasets associated with historic and future immunotherapies will help to shed more light on parameters associated with highly effective neoantigens and successful neoantigen-based treatments. More specifically, a better understanding of the mechanisms underlying neoantigen immunogenicity will have a significant impact on therapy success rate. For this, more extensive and accurately elucidated pMHC and pMHC-TCR complexes are needed. Unfortunately, achieving this is still expensive and labor-intensive since it requires crystallographic analysis of every individual pMHC and pMHC-TCR complex. In this context, more high-throughput methods applying crosslinking on the complexes of interest followed by proteolytic digestion and MS could already help elucidate important contact regions in a faster and more cost-efficient way. Incorporation of such interaction data into structure-based pMHC and pMHC-TCR-binding affinity prediction models could prove a major leap forward in improving neoantigen immunogenicity modeling. Neoantigen prioritization could further be improved by a more profound knowledge of a cell's protein and RNA degradation processes, which could help better delineate the proteins that are likely to be expressed and presented at the cell's surface using sequencing data alone. Finally, a more sensitive identification of low(er) expressed peptides by MS will likely enhance the elucidation of peptides arising from more exotic artifacts, such as tumor-specific smORFs and post-translational modifications, thus further increasing the pool of potential neoantigens to be used for cancer therapy.

Aside from correct neoantigen identification, neoantigen-directed immunotherapy faces several clinical challenges most importantly tumoral heterogeneity. Three different immune types exist, each exhibiting a distinct immune activity and response profile. As a result, there are large differences in how tumors behave, not only between cancer types but also within the same tumor type and in individual patients with a particular tumor type. The latter may in part explain why some cancer patients respond very well to immunotherapy, whereas others—with the same tumor type—do not. Overall, it is imperative to define the immune type of a patient's tumor and to design the treatment regimen accordingly. Combining immune profiling with mutational profiling should become standard-of-care as the idea of "one size fits all" has become outdated, prompting a more personalized approach. By analyzing different immunological and mutational (bio)markers, the most appropriate therapy can be selected for each patient. This also implies that we should rethink how clinical trials are designed, as multiple clinical trial failures can be attributed to suboptimal patient stratification and selection. Also, not until recently, only patients having completed their standard-of-care treatment could be enrolled. This has significantly undermined clinical efforts since these patients have highly advanced disease profiles and are often close to being out of treatment options. Since the approval of pembrolizumab in 2016 for first-line treatment of advanced lung cancers, other indications have followed for ICIs, changing the mindset toward testing other immunotherapies in first-line advanced cancers as well. Another interesting approach concerns a novel set of study designs that are used to increase trial efficiency by accommodating the cancer subpopulation heterogeneity. Basket trials, for instance, select patients across tumor types based on a common biomarker for a single therapy, whereas umbrella trials differentiate patients based on molecular subtypes of a single cancer type and treat each of these subgroups with a matched targeted treatment. Another design option, adaptive platform trials, evaluates different therapy arms versus a common control arm, wherein new experimental arms can be added, or therapy arms can be dropped for which (in)effectivity has been demonstrated. Thus, these new trial platforms can increase efficiency and offer a more effective approach as they can be tailored to match the research objectives of different cancer (sub)types.

Taken together, neoantigen-directed immunotherapy remains promising. Although significant challenges remain to fully unlock the potential of neoantigen-directed ACT and cancer vaccination, the knowledge obtained in recent years on tumor heterogeneity and its impact on treatment response should help to overcome these challenges. This will allow the selection of the most appropriate (combination) therapy as well as a more optimal design of clinical trials. Better pMHC/TCR-binding models and an increased understanding of neoantigen immunogenicity will likely increase the robustness and efficacy of neoantigen-based treatments further. Neoantigen-directed immunotherapy is thus

expected to play a significant part in cancer treatment due to the central role that neoantigens play in tumor development and elimination, creating the potential to aid in making immunotherapy available for every patient, even for patients with difficult-to-treat solid tumors.

DECLARATION OF INTERESTS

L.L., S.L., and B.F. are employees and shareholders at myNEO NV. E.S. is a member of the scientific advisory board of Intra-Vacc. W.v.C. is co-founder and shareholder at myNEO NV. K.T. is a scientific founder of and holds equity in eTheRNA Immunotherapies, and is member of the scientific advisory board of myNEO NV. S.H.v.d.B. is named as inventor on the patent for the use of synthetic long peptides as vaccine as well as on a patent for TEIPP neoantigen-specific treatments. He also serves as a paid member of the strategy board of the vaccine companies: ISA Pharmaceuticals, PCI Biotech, DCprime, and AGLAIA, and also performs contract research on neoantigens for Enara BIO and Frame Pharmaceuticals. P.A.O. reports research funding from and has advised Neon Therapeutics, Bristol-Meyers Squibb, Medrck, CytomX, Pfizer, Novartis, Celldex, Amgen, Array, AstraZeneca/MedImmune, Armo BioSciences, Xencor, Oncorus, and Roche/Genentech. C.B. is co-founder and shareholder at myNEO NV.

REFERENCES

1. Dunn, G.P., Bruce, A.T., Ikeda, H., Old, L.J., and Schreiber, R.D. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. Nat. Immunol. *3*, 991–998.

2. Kim, R., Emi, M., and Tanabe, K. (2007). Cancer immunoediting from immune surveillance to immune escape. Immunology *121*, 1–14.

3. Croci, D.O., Zacarías Fluck, M.F., Rico, M.J., Matar, P., Rabinovich, G.A., and Scharovsky, O.G. (2007). Dynamic cross-talk between tumor and immune cells in orchestrating the immunosuppressive network at the tumor microenvironment. Cancer Immunol. Immunother. *56*, 1687–1700.

4. Vaddepally, R.K., Kharel, P., Pandey, R., Garje, R., and Chandra, A.B. (2020). Review of indications of FDA-approved immune checkpoint inhibitors per NCCN guidelines with the level of evidence. Cancers *12*, E738.

5. Atkinson, V. (2017). Recent advances in malignant melanoma. Intern. Med. J. *47*, 1114–1121.

6. Ferris, R.L., Blumenschein, G., Jr., Fayette, J., Guigay, J., Colevas, A.D., Licitra, L., Harrington, K., Kasper, S., Vokes, E.E., Even, C., et al. (2016). Nivolumab for recurrent squamous-cell carcinoma of the head and neck. N. Engl. J. Med. *375*, 1856–1867.

7. Wakelee, H.A., Altorki, N.K., Zhou, C., Csőszi, T., Vynnychenko, I.O., Goloborodko, O., Luft, A., Akopov, A., Martinez-Marti, A., Kenmotsu, H., et al. (2021). IMpower010: primary results of a phase III global study of atezolizumab versus best supportive care after adjuvant chemotherapy in resected stage IB-IIIA non-small cell lung cancer (NSCLC). J. Clin. Oncol. *39*, 8500.

8. Leger, P.D., Rothschild, S., Castellanos, E., Pillai, R.N., York, S.J., and Horn, L. (2017). Response to salvage chemotherapy following exposure to immune checkpoint inhibitors in patients with non-small cell lung cancer. J. Clin. Oncol. *35*, 9084.

9. Haslam, A., and Prasad, V. (2019). Estimation of the percentage of US patients with cancer who are eligible for and respond to checkpoint inhibitor immunotherapy drugs. JAMA Netw. Open *2*, e192535.

10. Yarchoan, M., Albacker, L.A., Hopkins, A.C., Montesion, M., Murugesan, K., Vithayathil, T.T., Zaidi, N., Azad, N.S., Laheru, D.A., Frampton, G.M., and Jaffee, E.M. (2019). PD-L1 expression and tumor mutational burden are independent biomarkers in most cancers. JCI Insight *4*.

11. Yarchoan, M., Hopkins, A., and Jaffee, E.M. (2017). Tumor mutational burden and response rate to PD-1 inhibition. N. Engl. J. Med. *377*, 2500–2501.

12. Marabelle, A., Fakih, M., Lopez, J., Shah, M., Shapira-Frommer, R., Nakagawa, K., Chung, H.C., Kindler, H.L., Lopez-Martin, J.A., Miller, W.H., Jr., et al. (2020). Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. Lancet Oncol. *21*, 1353–1365.

13. Zacharakis, N., Chinnasamy, H., Black, M., Xu, H., Lu, Y.C., Zheng, Z., Pasetto, A., Langhan, M., Shelton, T., Prickett, T., et al. (2018). Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. Nat. Med. *24*, 724–730.

14. Robbins, P.F., Lu, Y.C., El-Gamil, M., Li, Y.F., Gross, C., Gartner, J., Lin, J.C., Teer, J.K., Cliften, P., Tycksen, E., et al. (2013). Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. Nat. Med. *19*, 747–752.

15. Verdegaal, E., van der Kooij, M.K., Visser, M., van der Minne, C., de Bruin, L., Meij, P., Terwisscha van Scheltinga, A., Welters, M.J., Santegoets, S., de Miranda, N., et al. (2020). Low-dose interferon-alpha preconditioning and adoptive cell therapy in patients with metastatic melanoma refractory to standard (immune) therapies: a phase I/II study. J. Immunother. Cancer *8*, e000166.

16. Verdegaal, E.M.E., de Miranda, N.F.C.C., Visser, M., Harryvan, T., van Buuren, M.M., Andersen, R.S., Hadrup, S.R., van der Minne, C.E., Schotte, R., Spits, H., et al. (2016). Neoantigen landscape dynamics during human melanoma-T cell interactions. Nature *536*, 91–95.

17. Linnemann, C., van Buuren, M.M., Bies, L., Verdegaal, E.M.E., Schotte, R., Calis, J.J.A., Behjati, S., Velds, A., Hilkmann, H., Atmioui, D.E., et al. (2015). High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. Nat. Med. *21*, 81–85.

18. Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Luber, B.S., Azad, N.S., Laheru, D., et al. (2015). PD-1 blockade in tumors with mismatch-repair deficiency. N. Engl. J. Med. *372*, 2509–2520.

19. Gubin, M.M., Zhang, X., Schuster, H., Caron, E., Ward, J.P., Noguchi, T., Ivanova, Y., Hundal, J., Arthur, C.D., Krebber, W.J., et al. (2014). Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. Nature *515*, 577–581.

20. Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P., Ho, T.S., et al. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. Science *348*, 124–128.

21. Van Allen, E.M., Miao, D., Schilling, B., Shukla, S.A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Foppen, M.H.G., Goldinger, S.M., et al. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science *350*, 207–211.

22. Carreno, B.M., Magrini, V., Becker-Hapak, M., Kaabinejadian, S., Hundal, J., Petti, A.A., Ly, A., Lie, W.R., Hildebrand, W.H., Mardis, E.R., and Linette, G.P. (2015). Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. Science *348*, 803–808.

23. Ott, P.A., Hu, Z., Keskin, D.B., Shukla, S.A., Sun, J., Bozym, D.J., Zhang, W., Luoma, A., Giobbie-Hurder, A., Peter, L., et al. (2017). An immunogenic personal neoantigen vaccine for patients with melanoma. Nature *547*, 217–221.

24. Sahin, U., Derhovanessian, E., Miller, M., Kloke, B.P., Simon, P., Löwer, M., Bukur, V., Tadmor, A.D., Luxemburger, U., Schrörs, B., et al. (2017). Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature *547*, 222–226.

25. Keskin, D.B., Anandappa, A.J., Sun, J., Tirosh, I., Mathewson, N.D., Li, S., Oliveira, G., Giobbie-Hurder, A., Felt, K., Gjini, E., et al. (2019). Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. Nature *565*, 234–239.

26. Hu, Z., Leet, D.E., Allesøe, R.L., Oliveira, G., Li, S., Luoma, A.M., Liu, J., Forman, J., Huang, T., Iorgulescu, J.B., et al. (2021). Personal neoantigen vaccines induce persistent memory T cell responses and epitope spreading in patients with melanoma. Nat. Med. *27*, 515–525. https://doi.org/10.1038/s41591-020-01206-4.

27. Türeci, Ö., Vormehr, M., Diken, M., Kreiter, S., Huber, C., and Sahin, U. (2016). Targeting the heterogeneity of cancer with individualized neoepitope vaccines. Clin. Cancer Res. *22*, 1885–1896.

28. Saxena, M., van der Burg, S.H., Melief, C.J.M., and Bhardwaj, N. (2021). Therapeutic cancer vaccines. Nat. Rev. Cancer *21*, 360–378.

29. Alcantara, M., Du Rusquec, P., and Romano, E. (2020). Current clinical evidence and potential solutions to increase benefit of CAR T-cell therapy for patients with solid tumors. OncoImmunology *9*, 1777064.

30. Fousek, K., and Ahmed, N. (2015). The evolution of T-cell therapies for solid malignancies. Clin. Cancer Res. *21*, 3384–3392.

31. Zhao, L., and Cao, Y.J. (2019). Engineered T cell therapy for cancer in the clinic. Front. Immunol. *10*, 2250.

32. Weon, J.L., and Potts, P.R. (2015). The MAGE protein family and cancer. Curr. Opin. Cell Biol. *37*, 1–8. https://doi.org/10.1016/j.ceb.2015.08.002.

33. Wagner, J., Wickman, E., DeRenzo, C., and Gottschalk, S. (2020). CAR T cell therapy for solid tumors: bright future or dark reality? Mol. Ther. *28*, 2320–2339.

34. Titov, A., Valiullina, A., Zmievskaya, E., Zaikova, E., Petukhov, A., Miftakhova, R., Bulatov, E., and Rizvanov, A. (2020). Advancing CAR T-cell therapy for solid tumors: lessons learned from lymphoma treatment. Cancers *12*, 125.

35. Dagogo-Jack, I., and Shaw, A.T. (2018). Tumour heterogeneity and resistance to cancer therapies. Nat. Rev. Clin. Oncol. *15*, 81–94.

36. Sharma, P., Hu-Lieskovan, S., Wargo, J.A., and Ribas, A. (2017). Primary, adaptive, and acquired resistance to cancer immunotherapy. Cell *168*, 707–723.

37. Chen, D.S., and Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. Nature *541*, 321–330.

38. Aurisicchio, L., Pallocca, M., Ciliberto, G., and Palombo, F. (2018). The perfect personalized cancer therapy: cancer vaccines against neoantigens. J. Exp. Clin. Cancer Res. *37*, 86.

39. Beyranvand Nejad, E., Welters, M.J.P., Arens, R., and van der Burg, S.H. (2017). The importance of correctly timing cancer immunotherapy. Expert Opin. Biol. Ther. *17*, 87–103.

40. Messenheimer, D.J., Jensen, S.M., Afentoulis, M.E., Wegmann, K.W., Feng, Z., Friedman, D.J., Gough, M.J., Urba, W.J., and Fox, B.A. (2017). Timing of PD-1 blockade is critical to effective combination immunotherapy with anti-OX40. Clin. Cancer Res. *23*, 6165–6177.

41. Verma, V., Shrimali, R.K., Ahmad, S., Dai, W., Wang, H., Lu, S., Nandre, R., Gaur, P., Lopez, J., Sade-Feldman, M., et al. (2019). PD-1 blockade in subprimed CD8 cells induces dysfunctional PD-1+CD38hi cells and anti-PD-1 resistance. Nat. Immunol. *20*, 1231–1243.

42. Park, R., Lopes, L., Cristancho, C.R., Riano, I.M., and Saeed, A. (2020). Treatment-related adverse events of combination immune checkpoint inhibitors: systematic review and meta-analysis. Front. Oncol. *10*, 258. https://www.frontiersin.org/article/10.3389/fonc.2020.00258.

43. Schumacher, T.N., Scheper, W., and Kvistborg, P. (2019). Cancer neoantigens. Annu. Rev. Immunol. *37*, 173–200.

44. Jiang, T., Shi, T., Zhang, H., Hu, J., Song, Y., Wei, J., Ren, S., and Zhou, C. (2019). Tumor neoantigens: from basic research to clinical applications. J. Hematol. Oncol. *12*, 93.

45. Mougel, A., Terme, M., and Tanchot, C. (2019). Therapeutic cancer vaccine and combinations with antiangiogenic therapies and immune checkpoint blockade. Front. Immunol. *10*, 467.

46. Lee, W.S., Yang, H., Chon, H.J., and Kim, C. (2020). Combination of anti-angiogenic therapy and immune checkpoint blockade normalizes vascular-immune crosstalk to potentiate cancer immunity. Exp. Mol. Med. *52*, 1475–1485.

47. Rini, B.I., Plimack, E.R., Stus, V., Gafanov, R., Hawkins, R., Nosov, D., Pouliot, F., Alekseev, B., Soulières, D., Melichar, B., et al. (2019). Pembrolizumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. N. Engl. J. Med. *380*, 1116–1127.

48. Motzer, R.J., Penkov, K., Haanen, J., Rini, B., Albiges, L., Campbell, M.T., Venugopal, B., Kollmannsberger, C., Negrier, S., Uemura, M., et al. (2019). Avelumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. N. Engl. J. Med. *380*, 1103–1115.

49. Finn, R.S., Qin, S., Ikeda, M., Galle, P.R., Ducreux, M., Kim, T.Y., Kudo, M., Breder, V., Merle, P., Kaseb, A.O., et al. (2020). Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma. N. Engl. J. Med. *382*, 1894–1905.

50. Yi, M., Jiao, D., Qin, S., Chu, Q., Wu, K., and Li, A. (2019). Synergistic effect of immune checkpoint blockade and anti-angiogenesis in cancer treatment. Mol. Cancer *18*, 60.

51. Ciciola, P., Cascetta, P., Bianco, C., Formisano, L., and Bianco, R. (2020). Combining immune checkpoint inhibitors with anti-angiogenic agents. J. Clin. Med. *9*, 675.

52. Fucikova, J., Kepp, O., Kasikova, L., Petroni, G., Yamazaki, T., Liu, P., Zhao, L., Spisek, R., Kroemer, G., and Galluzzi, L. (2020). Detection of immunogenic cell death and its relevance for cancer therapy. Cell Death Dis. *11*, 1013.

53. Fumet, J.-D., Limagne, E., Thibaudin, M., and Ghiringhelli, F. (2020). Immunogenic cell death and elimination of immunosuppressive cells: a double-edged sword of chemotherapy. Cancers *12*, E2637.

54. Melief, C.J.M., Welters, M.J.P., Vergote, I., Kroep, J.R., Kenter, G.G., Ottevanger, P.B., Tjalma, W.A.A., Denys, H., van Poelgeest, M.I.E., Nijman, H.W., et al. (2020). Strong vaccine responses during chemotherapy are associated with prolonged cancer survival. Sci. Transl. Med. *12*, eaaz8235.

55. Quek, C., Bai, X., Long, G.v., Scolyer, R.A., and Wilmott, J.S. (2021). High-dimensional single-cell transcriptomics in melanoma and cancer immunotherapy. Genes *12*, 1629. https://doi.org/10.3390/genes12101629.

56. Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. Science *363*, 1463–1467.

57. Nam, J.-Y., Kim, N.K.D., Kim, S.C., Joung, J.G., Xi, R., Lee, S., Park, P.J., and Park, W.Y. (2016). Evaluation of somatic copy number estimation tools for whole-exome sequencing data. Brief. Bioinform. *17*, 185–192.

58. Orenbuch, R., Filip, I., Comito, D., Shaman, J., Pe'er, I., and Rabadan, R. (2020). arcasHLA: high-resolution HLA typing from RNAseq. Bioinformatics *36*, 33–40.

59. Lee, M., Seo, J.H., Song, S., Song, I.H., Kim, S.Y., Kim, Y.A., Gong, G., Kim, J.E., and Lee, H.J. (2021). A new human leukocyte antigen typing algorithm combined with currently available genotyping tools based on next-generation sequencing data and guidelines to select the most likely human leukocyte antigen genotype. Front. Immunol. *12*, 688183.

60. Claeys, A., Staut, J., Merseburger, P., Marchal, K., and van den Eynden, J. (2022). Benchmark of tools for in silico prediction of MHC class I and class II genotypes from NGS data. Preprint at bioRxiv. https://doi.org/10.1101/2022.04.28.489842.

61. Xie, C., Yeo, Z.X., Wong, M., Piper, J., Long, T., Kirkness, E.F., Biggs, W.H., Bloom, K., Spellman, S., Vierra-Green, C., et al. (2017). Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. Proc. Natl. Acad. Sci. USA *114*, 8059–8064.

62. Liu, P., Yao, M., Gong, Y., Song, Y., Chen, Y., Ye, Y., Liu, X., Li, F., Dong, H., Meng, R., et al. (2021). Benchmarking the human leukocyte antigen typing performance of three assays and seven next-generation sequencing-based algorithms. Front. Immunol. *12*, 652258.

63. Lee, H., and Kingsford, C. (2018). Graph-guided assembly for novel human leukocyte antigen allele discovery. Genome Biol. *19*, 16.

64. Dilthey, A.T., Gourraud, P.A., Mentzer, A.J., Cereb, N., Iqbal, Z., and McVean, G. (2016). High-accuracy HLA type inference from whole-genome sequencing data using population reference graphs. PLoS Comput. Biol. *12*, e1005151.

65. Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S.,

et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. *33*, 1152–1158.

66. Bai, Y., Ni, M., Cooper, B., Wei, Y., and Fury, W. (2014). Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. BMC Genom. *15*, 325.

67. Li, X., Zhou, C., Chen, K., Huang, B., Liu, Q., and Ye, H. (2021). Benchmarking HLA genotyping and clarifying HLA impact on survival in tumor immunotherapy. Mol. Oncol. *15*, 1764–1782.

68. Yi, J., Chen, L., Xiao, Y., Zhao, Z., and Su, X. (2021). Investigations of sequencing data and sample type on HLA class Ia typing with different computational tools. Brief. Bioinform. *22*, bbaa143.

69. Cao, H., Wu, J., Wang, Y., Jiang, H., Zhang, T., Liu, X., Xu, Y., Liang, D., Gao, P., Sun, Y., et al. (2013). An integrated tool to study MHC region: accurate SNV detection and HLA genes typing in human MHC region using targeted high-throughput sequencing. PLoS One *8*, e69388.

70. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

71. Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., and Saunders, C.T. (2018). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods *15*, 591–594.

72. Saunders, C.T., Wong, W.S.W., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics *28*, 1811–1817.

73. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576.

74. Larson, D.E., Harris, C.C., Chen, K., Koboldt, D.C., Abbott, T.E., Dooling, D.J., Ley, T.J., Mardis, E.R., Wilson, R.K., and Ding, L. (2012). SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics *28*, 311–317.

75. Narzisi, G., Corvelo, A., Arora, K., Bergmann, E.A., Shah, M., Musunuri, R., Emde, A.K., Robine, N., Vacic, V., and Zody, M.C. (2018). Genome-wide somatic variant calling using localized colored de Bruijn graphs. Commun. Biol. *1*, 20.

76. Shiraishi, Y., Sato, Y., Chiba, K., Okuno, Y., Nagata, Y., Yoshida, K., Shiba, N., Hayashi, Y., Kume, H., Homma, Y., et al. (2013). An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. Nucleic Acids Res. *41*, e89.

77. Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865–2871.

78. Cameron, D.L., Schröder, J., Penington, J.S., Do, H., Molania, R., Dobrovic, A., Speed, T.P., and Papenfuss, A.T. (2017). GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Res. *27*, 2050–2060.

79. Calis, J.J.A., Reinink, P., Keller, C., Kloetzel, P.M., and Keşmir, C. (2015). Role of peptide processing predictions in T cell epitope identification: contribution of different prediction programs. Immunogenetics *67*, 85–93.

80. Nielsen, M., Lundegaard, C., Lund, O., and Keşmir, C. (2005). The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. Immunogenetics *57*, 33–41.

81. Diez-Rivero, C.M., Chenlo, B., Zuluaga, P., and Reche, P.A. (2010). Quantitative modeling of peptide binding to TAP using support vector machine. Proteins *78*, 63–72.

82. Zhang, G.L., Petrovsky, N., Kwoh, C.K., August, J.T., and Brusic, V. (2006). PRED(TAP): a system for prediction of peptide binding to the human transporter associated with antigen processing. Immunome Res. *2*, 3.

83. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

84. Patro, R., Mount, S.M., and Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat. Biotechnol. *32*, 462–464.

85. Melsted, P., Hateley, S., Joseph, I.C., Pimentel, H., Bray, N., and Pachter, L. (2017). Fusion detection and quantification by pseudoalignment. Preprint at bioRxiv. https://doi.org/10.1101/166322.

86. Nicorici, D., Şatalan, M., Edgren, H., Kangaspeska, S., Murumägi, A., Kallioniemi, O., Virtanen, S., and Kilkku, O. (2014). FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data. Preprint at bioRxiv. https://doi.org/10.1101/011650.

87. Haas, B.J., Dobin, A., Stransky, N., Li, B., Yang, X., Tickle, T., Bankapur, A., Ganote, C., Doak, T.G., Pochet, N., et al. (2017). STAR-fusion: fast and accurate fusion transcript detection from RNA-seq. Preprint at bioRxiv. https://doi.org/10.1101/120295.

88. Zhao, J., Chen, Q., Wu, J., Han, P., and Song, X. (2017). GFusion: an effective algorithm to identify fusion genes from cancer RNA-seq data. Sci. Rep. *7*, 6880.

89. Zhang, J., White, N.M., Schmidt, H.K., Fulton, R.S., Tomlinson, C., Warren, W.C., Wilson, R.K., and Maher, C.A. (2016). INTEGRATE: gene fusion discovery using whole genome and transcriptome data. Genome Res. *26*, 108–118.

90. Tian, L., Li, Y., Edmonson, M.N., Zhou, X., Newman, S., McLeod, C., Thrasher, A., Liu, Y., Tang, B., Rusch, M.C., et al. (2020). CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. Genome Biol. *21*, 126.

91. Davidson, N.M., Majewski, I.J., and Oshlack, A. (2015). JAFFA: high sensitivity transcriptome-focused fusion gene detection. Genome Med. *7*, 43.

92. Uhrig, S., Ellermann, J., Walther, T., Burkhardt, P., Fröhlich, M., Hutter, B., Toprak, U.H., Neumann, O., Stenzinger, A., Scholl, C., et al. (2021). Accurate and efficient detection of gene fusions from RNA sequencing data. Genome Res. *31*, 448–460.

93. Li, Y., Rao, X., Mattox, W.W., Amos, C.I., and Liu, B. (2015). RNA-seq analysis of differential splice junction usage and intron retentions by DEXSeq. PLoS One *10*, e0136653.

94. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

95. Kahles, A., Ong, C.S., Zhong, Y., and Rätsch, G. (2016). SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. Bioinformatics *32*, 1840–1847.

96. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

97. Pimentel, H., Conboy, J., and Pachter, L. (2016). Keep me around: intron retention detection and analysis. Preprint at arXiv. https://doi.org/10.48550/arXiv.1510.00696.

98. Vaquero-Garcia, J., Barrera, A., Gazzara, M.R., González-Vallinas, J., Lahens, N.F., Hogenesch, J.B., Lynch, K.W., and Barash, Y. (2016). A new view of transcriptome complexity and regulation through the lens of local splicing variations. Elife *5*, e11752.

99. Li, W.V., Li, S., Tong, X., Deng, L., Shi, H., and Li, J.J. (2019). AIDE: annotation-assisted isoform discovery with high precision. Genome Res. *29*, 2056–2072.

100. Trincado, J.L., Entizne, J.C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D.J., and Eyras, E. (2018). SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biol. *19*, 40.

101. Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., Ha, G., Aparicio, S., Bouchard-Côté, A., and Shah, S.P. (2014). PyClone: statistical inference of clonal population structure in cancer. Nat. Methods *11*, 396–398.

102. Deshwar, A.G., Vembu, S., Yung, C.K., Jang, G.H., Stein, L., and Morris, Q. (2015). PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. Genome Biol. *16*, 35.

103. Ricketts, C., Seidman, D., Popic, V., Hormozdiari, F., Batzoglou, S., and Hajirasouliha, I. (2020). Meltos: multi-sample tumor phylogeny reconstruction for structural variants. Bioinformatics *36*, 1082–1090.

104. Cmero, M., Yuan, K., Ong, C.S., Schröder, J., PCAWG Evolution and Heterogeneity Working Group, Corcoran, N.M., Papenfuss, T., Hovens, C.M., Markowetz, F., and Macintyre, G.; PCAWG Consortium (2020). Inferring structural variant cancer cell fraction. Nat. Commun. *11*, 730.

105. Eaton, J., Wang, J., and Schwartz, R. (2018). Deconvolution and phylogeny inference of structural variations in tumor genomic samples. Bioinformatics *34*, i357–i365.

106. Li, Y., Zhou, S., Schwartz, D.C., and Ma, J. (2016). J. Allele-specific quantification of structural variations in cancer genomes. Cell Syst. *3*, 21–34.

107. Jiao, W., Vembu, S., Deshwar, A.G., Stein, L., and Morris, Q. (2014). Inferring clonal evolution of tumors from single nucleotide somatic mutations. BMC Bioinf. *15*, 35.

108. Fischer, A., Vázquez-García, I., Illingworth, C.J.R., and Mustonen, V. (2014). High-definition reconstruction of clonal composition in cancer. Cell Rep. *7*, 1740–1752.

109. Miller, C.A., White, B.S., Dees, N.D., Griffith, M., Welch, J.S., Griffith, O.L., Vij, R., Tomasson, M.H., Graubert, T.A., Walter, M.J., et al. (2014). SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. PLoS Comput. Biol. *10*, e1003665.

110. Serin Harmanci, A., Harmanci, A.O., and Zhou, X. (2020). CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. Nat. Commun. *11*, 89.

111. Andor, N., Harness, J.V., Müller, S., Mewes, H.W., and Petritsch, C. (2014). EXPANDS: expanding ploidy and allele frequency on nested subpopulations. Bioinformatics *30*, 50–60.

112. Oesper, L., Mahmoody, A., and Raphael, B.J. (2013). Inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. Genome Biol. *14*, R80.

113. Ha, G., Roth, A., Khattra, J., Ho, J., Yap, D., Prentice, L.M., Melnyk, N., McPherson, A., Bashashati, A., Laks, E., et al. (2014). TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. Genome Res. *24*, 1881–1893.

114. Shi, X., Xiaohua, W., and Caiyi, F. (2022). A highly effective system for predicting MHC-II epitopes with immunogenicity. Front. Oncol. *12*, 888556.

115. Schmidt, J., Smith, A.R., Magnin, M., Racle, J., Devlin, J.R., Bobisse, S., Cesbron, J., Bonnet, V., Carmona, S.J., Huber, F., et al. (2021). Prediction of neo-epitope immunogenicity reveals TCR recognition determinants and provides insight into immunoediting. Cell Rep. Med. *2*, 100194.

116. Pfitzer, L., Lybaert, L., Bogaert, C., and Fant, B. (2022). Improving T-cell mediated immunogenic epitope identification via machine learning: the neoIM model. Preprint at bioRxiv. https://doi.org/10.1101/2022.06.03.494687.

117. Ogishi, M., and Yotsuyanagi, H. (2019). Quantitative prediction of the landscape of T cell epitope immunogenicity in sequence space. Front. Immunol. *10*, 827.

118. Zhou, C., Wei, Z., Zhang, Z., Zhang, B., Zhu, C., Chen, K., Chuai, G., Qu, S., Xie, L., Gao, Y., and Liu, Q. (2019). pTuneos: prioritizing tumor neoantigens from next-generation sequencing data. Genome Med. *11*, 67.

119. Kim, S., Kim, H.S., Kim, E., Lee, M.G., Shin, E.C., Paik, S., and Kim, S. (2018). Neopepsee: accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information. Ann. Oncol. *29*, 1030–1036.

120. Jurtz, V.I., Jessen, L.E., Bentzen, A.K., Jespersen, M.C., Mahajan, S., Vita, R., Jensen, K.K., Marcatili, P., Hadrup, S.R., Peters, B., and Nielsen, M. (2018). NetTCR: sequence-based prediction of TCR binding to peptide-MHC complexes using convolutional neural networks. Preprint at bioRxiv. https://doi.org/10.1101/433706.

121. Gielis, S., Moris, P., De Neuter, N., Bittremieux, W., Ogunjimi, B., Laukens, K., and Maysman, P. (2018). TCRex: a webtool for the prediction of T-cell receptor sequence epitope specificity. Preprint at bioRxiv. https://doi.org/10.1101/373472.

122. Michael, C., Seojin, B., Pengfei, Z., and Heewook, L. (2022). ATM-TCR: TCR-epitope binding affinity prediction using a multi-head self-attention model. Front. Immunol. *13*, 893247.

123. Dash, P., Fiore-Gartland, A.J., Hertz, T., Wang, G.C., Sharma, S., Souquette, A., Crawford, J.C., Clemens, E.B., Nguyen, T.H.O., Kedzierska, K., et al. (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. Nature *547*, 89–93.

124. Ehrlich, R., Kamga, L., Gil, A., Luzuriaga, K., Selin, L.K., and Ghersi, D. (2021). SwarmTCR: a computational approach to predict the specificity of T cell receptors. BMC Bioinf. *22*, 422.

125. Tong, Y., Wang, J., Zheng, T., Zhang, X., Xiao, X., Zhu, X., Lai, X., and Liu, X. (2020). SETE: sequence-based Ensemble learning approach for TCR Epitope binding prediction. Comput. Biol. Chem. *87*, 107281.

126. Springer, I., Besser, H., Tickotsky-Moskovitz, N., Dvorkin, S., and Louzoun, Y. (2020). Prediction of specific TCR-peptide binding from large dictionaries of TCR-peptide pairs. Front. Immunol. *11*, 1803.

127. Pierce, B.G., and Weng, Z. (2013). A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes. Protein Sci. *22*, 35–46.

128. Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L.E., Rubelt, F., Ji, X., Han, A., Krams, S.M., Pettus, C., et al. (2017). Identifying specificity groups in the T cell receptor repertoire. Nature *547*, 94–98.

129. Hoffmann, T., Marion, A., and Antes, I. (2017). DynaDom: structure-based prediction of T cell receptor inter-domain and T cell receptor-peptide-MHC (class I) association angles. BMC Struct. Biol. *17*, 2.

130. Jensen, K.K., Rantos, V., Jappe, E.C., Olsen, T.H., Jespersen, M.C., Jurtz, V., Jessen, L.E., Lanzarotti, E., Mahajan, S., Peters, B., et al. (2019). TCRpMHCmodels: structural modelling of TCR-pMHC class I complexes. Sci. Rep. *9*, 14530.

131. Dhusia, K., Su, Z., and Wu, Y. (2021). A structural-based machine learning method to classify binding affinities between TCR and peptide-MHC complexes. Mol. Immunol. *139*, 76–86.

132. Riley, T.P., Keller, G.L.J., Smith, A.R., Davancaze, L.M., Arbuiso, A.G., Devlin, J.R., and Baker, B.M. (2019). Structure based prediction of neoantigen immunogenicity. Front. Immunol. *10*, 2047.

133. Montemurro, A., Schuster, V., Povlsen, H.R., Bentzen, A.K., Jurtz, V., Chronister, W.D., Crinklaw, A., Hadrup, S.R., Winther, O., Peters, B., et al. (2021). NetTCR-2.0 enables accurate prediction of TCR-peptide binding by using paired TCRα and β sequence data. Commun. Biol. *4*, 1060.

134. Kiyotani, K., Mai, T.H., and Nakamura, Y. (2017). Comparison of exome-based HLA class I genotyping tools: identification of platform-specific genotyping errors. J. Hum. Genet. *62*, 397–405.

135. Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics *30*, 3310–3316.

136. Kim, D., Paggi, J., and Salzberg, S.L. (2018). HISAT-Genotype: next generation genomic analysis platform on a personal computer. Preprint at bioRxiv. https://doi.org/10.1101/266197.

137. Dilthey, A.T., Mentzer, A.J., Carapito, R., Cutland, C., Cereb, N., Madhi, S.A., Rhie, A., Koren, S., Bahram, S., McVean, G., and Phillippy, A.M. (2019). HLA-LA - HLA typing from linearly projected graph alignments. Bioinformatics *35*, 4394–4396.

138. Kawaguchi, S., Higasa, K., Shimizu, M., Yamada, R., and Matsuda, F. (2017). HLA-HD: an accurate HLA typing algorithm for next-generation sequencing data. Hum. Mutat. *38*, 788–797.

139. Bai, Y., Wang, D., and Fury, W. (2018). PHLAT: inference of high-resolution HLA types from RNA and whole exome sequencing. Methods Mol. Biol. *1802*, 193–201.

140. Boegel, S., Löwer, M., Schäfer, M., Bukur, T., de Graaf, J., Boisguérin, V., Türeci, O., Diken, M., Castle, J.C., and Sahin, U. (2012). HLA typing from RNA-Seq sequence reads. Genome Med. *4*, 102.

141. McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., et al. (2017). Allele-specific HLA loss and immune escape in lung cancer evolution. Cell *171*, 1259–1271.e11.

142. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

143. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

144. Cai, L., Yuan, W., Zhang, Z., He, L., and Chou, K.-C. (2016). In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. Sci. Rep. *6*, 36540.

145. Zhang, H., Lund, O., and Nielsen, M. (2009). The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. Bioinformatics *25*, 1293–1299.

146. Kim, Y., Sidney, J., Pinilla, C., Sette, A., and Peters, B. (2009). Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. BMC Bioinf. *10*, 394.

147. Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. J. Immunol. *199*, 3360–3368.

148. O'Donnell, T.J., Rubinsteyn, A., Bonsack, M., Riemer, A.B., Laserson, U., and Hammerbacher, J. (2018). MHCflurry: open-source class I MHC binding affinity prediction. Cell Systems *7*, 129–132.e4.

149. Bulik-Sullivan, B., Busby, J., Palmer, C.D., Davis, M.J., Murphy, T., Clark, A., Busby, M., Duke, F., Yang, A., Young, L., et al. (2019). Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. Nat. Biotechnol. *37*, 55–63.

150. Zhao, W., and Sher, X. (2018). Systematically benchmarking peptide-MHC binding predictors: from synthetic to naturally processed epitopes. PLoS Comput. Biol. *14*, e1006457.

151. Chen, B., Khodadoust, M.S., Olsson, N., Wagar, L.E., Fast, E., Liu, C.L., Muftuoglu, Y., Sworder, B.J., Diehn, M., Levy, R., et al. (2019). Predicting HLA class II antigen presentation through integrated deep learning. Nat. Biotechnol. *37*, 1332–1343.

152. Pyke, R.M., Mellacheruvu, D., Dea, S., Abbott, C.W., Zhang, S.V., Phillips, N.A., Harris, J., Bartha, G., Desai, S., McClory, R., et al. (2021). Precision neoantigen discovery using large-scale immunopeptidomes and composite modeling of MHC peptide presentation. Mol. Cell. Proteomics *20*, 100111.

153. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). 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*, W449–W454.

154. Gartner, J.J., Parkhurst, M.R., Gros, A., Tran, E., Jafferji, M.S., Copeland, A., Hanada, K.I., Zacharakis, N., Lalani, A., Krishna, S., et al. (2021). A machine learning model for ranking candidate HLA class I neoantigens based on known neoepitopes from multiple human tumor types. Nat. Cancer *2*, 563–574.

155. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L.J., and Mann, M. (2015). Mass spectrometry of human leukocyte antigen class i peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol. Cell. Proteomics *14*, 658–673.

156. Hu, W., Qiu, S., Li, Y., Zhang, L., Xiang, H., Han, X., Chen, L., Li, S., Ren, Z., Hou, G., et al. (2019). EPIC: MHC-I epitope prediction integrating mass spectrometry derived motifs and tissue-specific expression profiles. Preprint at bioRxiv. https://doi.org/10.1101/567081.

157. Mill, N.A., Bogaert, C., van Criekinge, W., and Fant, B. (2022). neoMS: attention-based prediction of MHC-I epitope presentation. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.13.491845.

158. Bonsack, M., Hoppe, S., Winter, J., Tichy, D., Zeller, C., Küpper, M.D., Schitter, E.C., Blatnik, R., and Riemer, A.B. (2019). Performance evaluation of MHC class-I binding prediction tools based on an experimentally validated MHC–peptide binding data set. Cancer Immunol. Res. *7*, 719–736.

159. Wang, S., Li, J., Chen, X., Wang, L., Liu, W., and Wu, Y. (2016). Analyzing the effect of peptide-HLA-binding ability on the immunogenicity of potential CD8+ and CD4+ T cell epitopes in a large dataset. Immunol. Res. *64*, 908–918.

160. Paulson, K.G., Voillet, V., McAfee, M.S., Hunter, D.S., Wagener, F.D., Perdicchio, M., Valente, W.J., Koelle, S.J., Church, C.D., Vandeven, N., et al. (2018). Acquired cancer resistance to combination immunotherapy from transcriptional loss of class I HLA. Nat. Commun. *9*, 3868.

161. Hackl, H., Charoentong, P., Finotello, F., and Trajanoski, Z. (2016). Computational genomics tools for dissecting tumour–immune cell interactions. Nat. Rev. Genet. *17*, 441–458.

162. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

163. Wells, D.K., van Buuren, M.M., Dang, K.K., Hubbard-Lucey, V.M., Sheehan, K.C.F., Campbell, K.M., Lamb, A., Ward, J.P., Sidney, J., Blazquez, A.B., et al. (2020). Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. Cell *183*, 818–834.e13.

164. Zhou, Z., Lyu, X., Wu, J., Yang, X., Wu, S., Zhou, J., Gu, X., Su, Z., and Chen, S. (2017). TSNAD: an integrated software for cancer somatic mutation and tumour-specific neoantigen detection. R. Soc. Open Sci. *4*, 170050.

165. Tang, Y., Wang, Y., Wang, J., Li, M., Peng, L., Wei, G., Zhang, Y., Li, J., and Gao, Z. (2020). TruNeo: an integrated pipeline improves personalized true tumor neoantigen identification. BMC Bioinf. *21*, 532.

166. Kodysh, J., and Rubinsteyn, A. (2020). OpenVax: an open-source computational pipeline for cancer neoantigen prediction BT - bioinformatics for cancer immunotherapy: methods and protocols, S. Boegel, ed. (Springer US), pp. 147–160. https://doi.org/10.1007/978-1-0716-0327-7_10.

167. Rieder, D., Fotakis, G., Ausserhofer, M., René, G., Paster, W., Trajanoski, Z., and Finotello, F. (2022). nextNEOpi: a comprehensive pipeline for computational neoantigen prediction. Bioinformatics *38*, 1131–1132.

168. Warren, R.L., Choe, G., Freeman, D.J., Castellarin, M., Munro, S., Moore, R., and Holt, R.A. (2012). Derivation of HLA types from shotgun sequence datasets. Genome Med. *4*, 95.

169. Cao, H., Wu, J., Wang, Y., Jiang, H., Zhang, T., Liu, X., Xu, Y., Liang, D., Gao, P., Sun, Y., et al. (2013). An integrated tool to study MHC region: accurate SNV detection and HLA genes typing in human MHC region using targeted high-throughput sequencing. PLoS One *8*, e69388.

170. Wood, M.A., Nguyen, A., Struck, A.J., Ellrott, K., Nellore, A., and Thompson, R.F. (2020). Neoepiscope improves neoepitope prediction with multivariant phasing. Bioinformatics *36*, 713–720.

171. Schenck, R.O., Lakatos, E., Gatenbee, C., Graham, T.A., and Anderson, A.R.A. (2019). NeoPredPipe: high-throughput neoantigen prediction and recognition potential pipeline. BMC Bioinf. *20*, 264.

172. Bjerregaard, A.-M., Nielsen, M., Hadrup, S.R., Szallasi, Z., and Eklund, A.C. (2017). MuPeXI: prediction of neo-epitopes from tumor sequencing data. Cancer Immunol. Immunother. *66*, 1123–1130.

173. Fotakis, G., Rieder, D., Haider, M., Trajanoski, Z., and Finotello, F. (2020). NeoFuse: predicting fusion neoantigens from RNA sequencing data. Bioinformatics *36*, 2260–2261.

174. Hundal, J., Kiwala, S., McMichael, J., Miller, C.A., Xia, H., Wollam, A.T., Liu, C.J., Zhao, S., Feng, Y.Y., Graubert, A.P., et al. (2020). pVACtools: a computational toolkit to identify and visualize cancer neoantigens. Cancer Immunol. Res. *8*, 409–420.

175. Zhang, J., Mardis, E.R., and Maher, C.A. (2017). INTEGRATE-neo: a pipeline for personalized gene fusion neoantigen discovery. Bioinformatics *33*, 555–557.

176. Coelho, A.C.M.F., Fonseca, A.L., Martins, D.L., Lins, P.B.R., da Cunha, L.M., and de Souza, S.J. (2020). An integrated tool for identification of potential neoantigens. BMC Med. Genomics *13*, 30.

177. Laumont, C.M., Vincent, K., Hesnard, L., Audemard, É., Bonneil, É., Laverdure, J.P., Gendron, P., Courcelles, M., Hardy, M.P., Côté, C., et al. (2018). Noncoding regions are the main source of targetable tumor-specific antigens. Sci. Transl. Med. *10*, eaau5516.

178. Yang, W., Lee, K.W., Srivastava, R.M., Kuo, F., Krishna, C., Chowell, D., Makarov, V., Hoen, D., Dalin, M.G., Wexler, L., et al. (2019). Immunogenic neoantigens derived from gene fusions stimulate T cell responses. Nat. Med. *25*, 767–775.

179. Kahles, A., Lehmann, K.V., Toussaint, N.C., Hüser, M., Stark, S.G., Sachsenberg, T., Stegle, O., Kohlbacher, O., Sander, C.; Cancer Genome Atlas Research Network, and Rätsch, G. (2018). Comprehensive analysis of alternative splicing across tumors from 8, 705 patients. Cancer Cell *34*, 211–224.e6.

180. Smart, A.C., Margolis, C.A., Pimentel, H., He, M.X., Miao, D., Adeegbe, D., Fugmann, T., Wong, K.K., and Van Allen, E.M. (2018). Intron retention is a source of neoepitopes in cancer. Nat. Biotechnol. *36*, 1056–1058.

181. Khodadoust, M.S., Olsson, N., Wagar, L.E., Haabeth, O.A.W., Chen, B., Swaminathan, K., Rawson, K., Liu, C.L., Steiner, D., Lund, P., et al. (2017). Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. Nature *543*, 723–727.

182. Krøigård, A.B., Thomassen, M., Lænkholm, A.V., Kruse, T.A., and Larsen, M.J. (2016). Evaluation of nine somatic variant callers for detection of somatic mutations in exome and targeted deep sequencing data. PLoS One *11*, e0151664.

183. Gao, Q., Liang, W.W., Foltz, S.M., Mutharasu, G., Jayasinghe, R.G., Cao, S., Liao, W.W., Reynolds, S.M., Wyczalkowski, M.A., Yao, L., et al. (2018). Driver fusions and their implications in the development and treatment of human cancers. Cell Rep. *23*, 227–238.e3.

184. Shen, S., Park, J.W., Lu, Z.x., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. USA *111*, E5593–E5601.

185. Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Res. *22*, 2008–2017.

186. Mehmood, A., Laiho, A., Venäläinen, M.S., McGlinchey, A.J., Wang, N., and Elo, L.L. (2020). Systematic evaluation of differential splicing tools for RNA-seq studies. Brief. Bioinform. *21*, 2052–2065.

187. Koster, J., and Plasterk, R.H.A. (2019). A library of Neo Open Reading Frame peptides (NOPs) as a sustainable resource of common neoantigens in up to 50% of cancer patients. Sci. Rep. *9*, 6577.

188. Ouspenskaia, T., Law, T., Clauser, K.R., Klaeger, S., Sarkizova, S., Aguet, F., Li, B., Christian, E., Knisbacher, B.A., Le, P.M., et al. (2020). Thousands of novel unannotated proteins expand the MHC I immunopeptidome in cancer. Preprint at bioRxiv. https://doi.org/10.1101/2020.02.12.945840.

189. Kochetov, A.V. (2006). [Alternative translation start sites and their significance for eukaryotic proteome]. Mol. Biol. *40*, 788–795.

190. Laumont, C.M., Daouda, T., Laverdure, J.P., Bonneil, É., Caron-Lizotte, O., Hardy, M.P., Granados, D.P., Durette, C., Lemieux, S., Thibault, P., and Perreault, C. (2016). Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. Nat. Commun. *7*, 10238.

191. Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B.J., et al. (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. Nat. Commun. *11*, 1293.

192. Peck Justice, S.A., McCracken, N.A., Victorino, J.F., Qi, G.D., Wijeratne, A.B., and Mosley, A.L. (2021). Boosting detection of low-abundance proteins in thermal proteome profiling experiments by addition of an isobaric trigger channel to TMT multiplexes. Anal. Chem. *93*, 7000–7010.

193. Lee, H.Y., Kim, E.G., Jung, H.R., Jung, J.W., Kim, H.B., Cho, J.W., Kim, K.M., and Yi, E.C. (2019). Refinements of LC-MS/MS spectral counting statistics improve quantification of low abundance proteins. Sci. Rep. *9*, 13653.

194. Cobbold, M., De La Peña, H., Norris, A., Polefrone, J.M., Qian, J., English, A.M., Cummings, K.L., Penny, S., Turner, J.E., Cottine, J., et al. (2013). MHC class I-associated phosphopeptides are the targets of memory-like immunity in leukemia. Sci. Transl. Med. *5*, 203ra125.

195. Malaker, S.A., Penny, S.A., Steadman, L.G., Myers, P.T., Loke, J.C., Raghavan, M., Bai, D.L., Shabanowitz, J., Hunt, D.F., and Cobbold, M. (2017). Identification of glycopeptides as posttranslationally modified neoantigens in leukemia. Cancer Immunol. Res. *5*, 376–384.

196. Gerlinger, M., Rowan, A.J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N. Engl. J. Med. *366*, 883–892.

197. Yadav, V.K., and De, S. (2015). An assessment of computational methods for estimating purity and clonality using genomic data derived from heterogeneous tumor tissue samples. Brief. Bioinform. *16*, 232–241.

198. Kim, K.I., and Simon, R. (2014). Using single cell sequencing data to model the evolutionary history of a tumor. BMC Bioinf. *15*, 27.

199. Tailor, T.D., Rao, X., Campa, M.J., Wang, J., Gregory, S.G., and Patz, E.F., Jr. (2019). Whole exome sequencing of cell-free DNA for early lung cancer: a pilot study to differentiate benign from malignant CT-detected pulmonary lesions. Front. Oncol. *9*, 317.

200. Akira, S., Uematsu, S., and Takeuchi, O. (2006). Pathogen recognition and innate immunity. Cell *124*, 783–801.

201. Andersen, M.H., Schrama, D., Thor Straten, P., and Becker, J.C. (2006). Cytotoxic T cells. J. Invest. Dermatol. *126*, 32–41.

202. Kim, H.-J., and Cantor, H. (2014). CD4 T-cell subsets and tumor immunity: the helpful and the not-so-helpful. Cancer Immunol. Res. *2*, 91–98.

203. Ferris, S.T., Durai, V., Wu, R., Theisen, D.J., Ward, J.P., Bern, M.D., Davidson, J.T., 4th, Bagadia, P., Liu, T., Briseño, C.G., et al. (2020). cDC1 prime and are licensed by CD4+ T cells to induce anti-tumour immunity. Nature *584*, 624–629.

204. Chicz, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Stern, L.J., Vignali, D.A., and Strominger, J.L. (1992). Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. Nature *358*, 764–768.

205. Kim, H.J., and Pourmand, N. (2013). HLA haplotyping from RNA-seq data using hierarchical read weighting. PLoS One *8*, e67885.

206. Liu, C., Yang, X., Duffy, B., Mohanakumar, T., Mitra, R.D., Zody, M.C., and Pfeifer, J.D. (2013). ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. Nucleic Acids Res. *41*, e142.

207. Venkatesh, G., Grover, A., Srinivasaraghavan, G., and Rao, S. (2020). MHCAttnNet: predicting MHC-peptide bindings for MHC alleles classes I and II using an attention-based deep neural model. Bioinformatics *36*, i399–i406.

208. Bui, H.-H., Schiewe, A.J., von Grafenstein, H., and Haworth, I.S. (2006). Structural prediction of peptides binding to MHC class I molecules. Proteins *63*, 43–52.

209. Racle, J., Michaux, J., Rockinger, G.A., Arnaud, M., Bobisse, S., Chong, C., Guillaume, P., Coukos, G., Harari, A., Jandus, C., et al. (2019). Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. Nat. Biotechnol. *37*, 1283–1286.

210. You, R., Qu, W., Mamitsuka, H., and Zhu, S. (2022). DeepMHCII: a novel binding core-aware deep interaction model for accurate MHC-II peptide binding affinity prediction. Bioinformatics *38*, i220–i228.

211. Shao, X.M., Bhattacharya, R., Huang, J., Sivakumar, I.K.A., Tokheim, C., Zheng, L., Hirsch, D., Kaminow, B., Omdahl, A., Bonsack, M., et al. (2020). High-throughput prediction of MHC class I and II neoantigens with MHCnuggets. Cancer Immunol. Res. *8*, 396–408. https://doi.org/10.1158/2326-6066.CIR-19-0464.

212. Nielsen, M., and Andreatta, M. (2017). NNAlign: a platform to construct and evaluate artificial neural network models of receptor-ligand interactions. Nucleic Acids Res. *45*, W344–W349.

213. Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L., Tuereci, O., Sahin, U., Braxenthaler, M., Gallazzi, F., Protti, M.P., Sinigaglia, F., and Hammer, J. (1999). Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat. Biotechnol. *17*, 555–561.

214. Bui, H.-H., Sidney, J., Peters, B., Sathiamurthy, M., Sinichi, A., Purton, K.A., Mothé, B.R., Chisari, F.V., Watkins, D.I., and Sette, A. (2005). Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. Immunogenetics *57*, 304–314.

215. Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., and Stevanović, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics *50*, 213–219.

216. Reche, P.A., Glutting, J.-P., and Reinherz, E.L. (2002). Prediction of MHC class I binding peptides using profile motifs. Hum. Immunol. *63*, 701–709.

217. Nielsen, M., Lundegaard, C., and Lund, O. (2007). Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinf. *8*, 238.

218. Ali, M., Foldvari, Z., Giannakopoulou, E., Böschen, M.L., Strønen, E., Yang, W., Toebes, M., Schubert, B., Kohlbacher, O., Schumacher, T.N., and Olweus, J. (2019). Induction of neoantigen-reactive T cells from healthy donors. Nat. Protoc. *14*, 1926–1943.

219. Marijt, K.A., Blijleven, L., Verdegaal, E.M.E., Kester, M.G., Kowalewski, D.J., Rammensee, H.G., Stevanović, S., Heemskerk, M.H.M., van der Burg, S.H., and van Hall, T. (2018). Identification of non-mutated neoantigens presented by TAP-deficient tumors. J. Exp. Med. *215*, 2325–2337.

220. Moutaftsi, M., Peters, B., Pasquetto, V., Tscharke, D.C., Sidney, J., Bui, H.H., Grey, H., and Sette, A. (2006). A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. Nat. Biotechnol. *24*, 817–819.

221. Feltkamp, M.C., Vierboom, M.P., Kast, W.M., and Melief, C.J. (1994). Efficient MHC class I-peptide binding is required but does not ensure MHC class I-restricted immunogenicity. Mol. Immunol. *31*, 1391–1401.

222. Capietto, A.-H., Jhunjhunwala, S., Pollock, S.B., Lupardus, P., Wong, J., Hänsch, L., Cevallos, J., Chestnut, Y., Fernandez, A., Lounsbury, N., et al. (2020). Mutation position is an important determinant for predicting cancer neoantigens. J. Exp. Med. *217*, e20190179.

223. Calis, J.J.A., Maybeno, M., Greenbaum, J.A., Weiskopf, D., De Silva, A.D., Sette, A., Keşmir, C., and Peters, B. (2013). Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Comput. Biol. *9*, e1003266.

224. Springer, I., Tickotsky, N., and Louzoun, Y. (2021). Contribution of T Cell receptor alpha and beta CDR3, MHC typing, V and J genes to peptide binding prediction. Front. Immunol. *12*, 664514.

225. Wang, G., Wan, H., Jian, X., Li, Y., Ouyang, J., Tan, X., Zhao, Y., Lin, Y., and Xie, L. (2020). INeo-epp: a novel T-cell HLA class-I immunogenicity or neoantigenic epitope prediction method based on sequence-related amino acid features. BioMed Res. Int. *2020*, 5798356.

226. Milighetti, M., Shawe-Taylor, J., and Chain, B. (2021). Predicting T cell receptor antigen specificity from structural features derived from homology models of receptor-peptide-major histocompatibility complexes. Front. Physiol. *12*, 790998.

227. Yarchoan, M., Johnson, B.A., 3rd, Lutz, E.R., Laheru, D.A., and Jaffee, E.M. (2017). Targeting neoantigens to augment antitumour immunity. Nat. Rev. Cancer *17*, 569–622.

228. D'Alise, A.M., Leoni, G., Cotugno, G., Troise, F., Langone, F., Fichera, I., De Lucia, M., Avalle, L., Vitale, R., Leuzzi, A., et al. (2019). Adenoviral vaccine targeting multiple neoantigens as strategy to eradicate large tumors combined with checkpoint blockade. Nat. Commun. *10*, 2688.

229. Joseph, C.G., Darrah, E., Shah, A.A., Skora, A.D., Casciola-Rosen, L.A., Wigley, F.M., Boin, F., Fava, A., Thoburn, C., Kinde, I., et al. (2014). Association of the autoimmune disease scleroderma with an immunologic response to cancer. Science *343*, 152–157.

230. Liu, Z., Jin, J., Cui, Y., Xiong, Z., Nasiri, A., Zhao, Y., and Hu, J. (2022). DeepSeqPanII: an interpretable recurrent neural network model with attention mechanism for peptide-HLA class II binding prediction. IEEE/ACM Trans. Comput. Biol. Bioinform. *19*, 2188–2196. https://doi.org/10.1109/TCBB.2021.3074927.

231. Paul, S., Croft, N.P., Purcell, A.W., Tscharke, D.C., Sette, A., Nielsen, M., and Peters, B. (2020). Benchmarking predictions of MHC class I restricted T cell epitopes in a comprehensively studied model system. PLoS Comput. Biol. *16*, e1007757.

232. Nielsen, M., and Lund, O. (2009). An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinf. *10*, 296.

233. Liu, Z., Song, L., Zhang, P., Cao, Z., Hao, J., Tian, Y., Luo, A., Zhang, P., and Ma, J. (2019). DeepSeqPan, a novel deep convolutional neural network model for pan-specific class I HLA-peptide binding affinity prediction. Sci. Rep. *9*, 20425.

234. Heng, Y., Kuang, Z., Huang, S., Chen, L., Shi, T., Xu, L., and Mei, H. (2020). A pan-specific GRU-based recurrent neural network for predicting HLA-I-binding peptides. ACS Omega *5*, 18321–18330.

235. Saxena, S., Animesh, S., Fullwood, M.J., and Mu, Y. (2021). A deep learning model for peptide - HLA-A*02: 01 binding predictions using both structure and sequence feature sets. J. Micromech. Mol. Phys. *05*, 2050009. https://doi.org/10.1142/S2424913020500095.

236. Hu, J., and Liu, Z. (2017). DeepMHC: deep convolutional neural networks for high-performance peptide-MHC binding affinity prediction. Preprint at bioRxiv. https://doi.org/10.1101/239236.

237. Bhattacharya, R., Tokheim, C., Sivakumar, A., Guthrie, V.B., Anagnostou, V., Velculescu, V.E., and Karchin, R. (2017). Prediction of peptide binding to MHC Class I proteins in the age of deep learning. Preprint at bioRxiv. https://doi.org/10.1101/154757.

238. Shao, X.M., Bhattacharya, R., Huang, J., Sivakumar, I.K.A., Tokheim, C., Zheng, L., Hirsch, D., Kaminow, B., Omdahl, A., Bonsack, M., et al. (2020). High-throughput prediction of MHC Class i and II neoantigens with MH cnuggets. Cancer Immunol. Res. *8*, 396–408.

239. Boehm, K.M., Bhinder, B., Raja, V.J., Dephoure, N., and Elemento, O. (2019). Predicting peptide presentation by major histocompatibility complex class I: an improved machine learning approach to the immunopeptidome. BMC Bioinf. *20*, 7.

240. Weipeng, H., Qiu, S., Li, Y., Lin, X., Zhang, L., Xiang, H., Han, X., Zhu, S., Chen, L., Li, S., et al. (2020). EPIP: MHC-I epitope prediction integrating mass spectrometry derived motifs and tissue-specific expression profiles. Preprint at bioRxiv. https://doi.org/10.1101/567081.

241. Bhattacharya, R., Sivakumar, A., Tokheim, C., Guthrie, V.B., Anagnostou, V., Velculescu, V.E., and Karchin, R. (2017). Evaluation of machine learning methods to predict peptide binding to MHC Class I proteins. Preprint at bioRxiv. https://doi.org/10.1101/154757.

242. Phloyphisut, P., Pornputtapong, N., Sriswasdi, S., and Chuangsuwanich, E. (2019). A deep neural network model for universal MHC binding prediction. BMC Bioinf. *20*, 270.