cancers

MDPI

Review

Main Strategies for the Identification of Neoantigens

Alexander V. Gopanenko, Ekaterina N. Kosobokova and Vyacheslav S. Kosorukov *

N.N. Blokhin National Medical Research Center of Oncology, Ministry of Health of the Russian Federation, 115478 Moscow, Russia; alexandr.gopanenko@yandex.ru (A.V.G.); ekkos@mail.ru (E.N.K.)
* Correspondence: kosorukov@ronc.ru; Tel.: +7-499-324-2274

Received: 4 September 2020; Accepted: 5 October 2020; Published: 7 October 2020

Simple Summary: This review provides an overview of currently available approaches applied for neoantigens discovery—tumor-specific peptides that appeared due to the mutation process and distinguish tumors from normal tissues. Focusing on genomics-based approaches and computational pipelines, we cover all steps required for selecting appropriate candidate peptides starting from NGS-derived data. Moreover, additional approaches such as mass-spectrometry-based and structure-based methods are discussed highlighting their advantages and disadvantages. This review also provides a description of available complex bioinformatics pipelines ensuring automated data processing resulting in a list of neoantigens. We propose the possible ideal pipeline that could be implemented in the neoantigens identification process. We discuss the integration of results generated by different approaches to improve the accuracy of neoantigens selection.

Abstract: Genetic instability of tumors leads to the appearance of numerous tumor-specific somatic mutations that could potentially result in the production of mutated peptides that are presented on the cell surface by the MHC molecules. Peptides of this kind are commonly called neoantigens. Their presence on the cell surface specifically distinguishes tumors from healthy tissues. This feature makes neoantigens a promising target for immunotherapy. The rapid evolution of high-throughput genomics and proteomics makes it possible to implement these techniques in clinical practice. In particular, they provide useful tools for the investigation of neoantigens. The most valuable genomic approach to this problem is whole-exome sequencing coupled with RNA-seq. High-throughput mass-spectrometry is another option for direct identification of MHC-bound peptides, which is capable of revealing the entire MHC-bound peptidome. Finally, structure-based predictions could significantly improve the understanding of physicochemical and structural features that affect the immunogenicity of peptides. The development of pipelines combining such tools could improve the accuracy of the peptide selection process and decrease the required time. Here we present a review of the main existing approaches to investigating the neoantigens and suggest a possible ideal pipeline that takes into account all modern trends in the context of neoantigen discovery.

Keywords: neoantigens; whole-exome and RNA sequencing; peptide-MHC binding predictions; pMHC-TCR binding predictions

1. Introduction

Cancer burden significantly impacts human health and quality of life. It is one of the leading causes of death worldwide (9.6 million deaths in 2018, according to WHO) [1]. Nowadays, there are different options to treat each type of cancer. These options include surgery, radiotherapy, chemotherapy, immunotherapy [2–4] and are often applied in combination to improve the outcomes. Recently developed immune checkpoint inhibitors (ICIs), such as anti-CTLA-4 and anti-PD-1/PD-L1, allowed us to achieve significant progress in the treatment of several cancer types [5,6]. Now they are an

established treatment for melanoma [7,8], NSCLC (non-small-cell lung carcinoma) [9,10], renal cell carcinoma [11,12], and are believed to be effective for some other cancer types [5,6,13–16]. Notably, it was observed that the efficacy of such therapy depends on the individual patient [13,17,18], which makes it necessary to investigate biomarkers of improved response. It is currently believed that a better response to ICIs is associated with PD-L1 expression, the number of tumor infiltrated lymphocytes, etc. [13,17–19]. It was also noticed that the efficacy of treatment with ICIs correlates with the tumor mutation burden (TMB) [20,21]. However, this trend is only observed in particular cancer types (e.g., melanoma, NSCLC, etc.) [21]. Taking into account that TMB can apparently reflect the so-called neoantigen burden [22–24], one could conclude that these observations highlight the importance of neoantigens as a triggering factor for tumor immune response related to T-cell activation [25–27]. It is not surprising that TMB is now justifiably considered to be a potential biomarker of response to immune checkpoint blockade therapy on par with other, well-established biomarkers [13,17–19,28].

In general, the set of somatic mutations is specific for each individual tumor specimen, with only a small number being common across samples [27,29]. Thus, it could be expected that the set of neoantigens has the same specificity. Being presented only on the surface of tumor cells by the major histocompatibility complex (MHC), which ensures their interaction with the TCR (T-cell receptor) [30], neoantigens have high immunogenic potential because they are distinguished from peptides generated by the degradation of normal proteins and thus can be recognized by the host's immune system as non-self, which would prevent them from inducing central and peripheral immune tolerance mechanisms [31]. It is important that in contrast to TAAs (tumor-associated antigens) such as HER2, MART-1, MUC1, and CGAs (cancer germline antigens or CTAs—cancer/testis antigens) such as MAGE (melanoma-associated antigen), NY-ESO-1 (cancer/testis antigen 1) [32–34], which can be expressed at low levels in a variety of healthy tissues, neoantigens are specific to tumors only. Thus targeting of neoantigens (e.g., by peptide vaccination) probably leads to less side effects associated with the targeting of TAAs and CGAs, including autoimmune toxicity related to immune activation in non-target tissues [35], cytokine release syndrome [36], and others. These properties make neoantigens the perfect target for immunotherapy. It seems logical that the facilitation of neoantigen presentation to the immune system could increase neoantigen-mediated T-cell response. One possible way to achieve this aim is by utilizing the neoantigen vaccines that were already developed and investigated in a limited number of trials for several cancer types such as high-risk melanoma [37,38], glioblastoma [39], etc., yielding promising results.

Different variants of somatic mutations that lead to the production of new proteins and, thus, to mutated peptide variants were established as the primary sources of neoantigens. Neoantigens can be a result of mutations leading to amino acid changes (SNV, single nucleotide variants), frameshift mutations such as insertions or deletions, jointly known as INDELs, that modify the protein sequence downstream of their position. SNV can also break stop codons, leading to passing through translation, which yields new protein sequences corresponding to normally non-coding regions. Another source of neoantigens is the translation of non-coding regions of the genome, such as long-noncoding RNAs [40]. Nowadays, the widespread use of so-called high-throughput techniques, mainly NGS, combined with a significant reduction in their cost, helps identify somatic mutation spectra and convert the resulting data into predictions of neoantigens. The most suitable NGS-based approaches applied for these purposes are whole-genome sequencing (WGS) and whole-exome sequencing (WES) coupled with RNA-seq. In contrast with whole-genome sequencing, WES offers a balance between costs and benefits. WES targets about 2–3% of the whole genome that is related to protein-coding genes [41] and is believed to be the main source of somatic mutations leading to the appearance of mutated proteins. Simultaneous use of RNA-seq provides information on gene expression levels, allowing to select the genes that are more likely to produce real proteins and also, in some cases, could serve as a starting point for somatic mutation calling [42,43]. Moreover, RNA-seq is capable of revealing other types of neoantigen sources, such as gene fusion [44,45], alternative splicing isoforms and RNA editing events [46]. In the absence of clinical HLA (human leukocyte antigen) typing data, RNA-seq can be an

additional source for in silico typing to improve the confidence of WES-based HLA determination [47]. High-confidence somatic mutation data that were identified based on WES and RNA-seq are commonly used in the isolation and ranking (or prioritization) of mutated peptide sequences aiming to detect peptides with the highest probability of being bound to the MHC and presented on the cell surface for TCR recognition [48,49]. Most algorithms typically apply HLA-peptide binding affinity as the primary ranking parameter [50]. As mentioned above, WES and RNA-seq also provide a foundation for HLA typing [48–50]. The individual set of HLA alleles (HLA allotype) is also required for peptide ranking because MHC-peptide affinity strongly depends on HLA variants [51]. At present, the prediction power of HLA typing tools that extract data from WES and RNA-seq is almost equal to the gold standards of HLA typing, such as PCR [52,53]. Nevertheless, if clinical data on the HLA type are available, it is more advisable to use them [50]. Perhaps the main weakness of the NGS-based approach is the absence of direct experimental evidence of the existence of predicted neoantigens. Today, the rapid progress in high-throughput mass-spectrometry approaches and the ensuing significant increase in sensitivity and accuracy [54–56] allows for direct identification of HLA binding epitopes. Modern LC-MS/MS makes it possible to identify thousands of MHC-bound peptides in a single experiment compared to tens in early studies [56]. Moreover, the MS data containing information on MHC-peptide interactions could serve as a reliable source for the creation of training datasets that play a critical role for machine learning-based approaches to MHC-peptide predictions [56,57]. This is especially important since prediction algorithms are mainly trained on biochemical data that could miss the entire picture of HLA-peptide binding features. Additionally, MS allows us to reveal tumor-specific post-translational modifications [58,59] and proteasome-generated spliced peptides [60] that also could contribute to the tumor-specific antigenic landscape. Finally, structure-based approaches could also improve the accuracy of the peptide selection process [61] by helping identify the effects of the structure and physicochemical properties of peptides on their immunogenic potential. A serious limitation of such approaches is due to the need for high-resolution crystallographic models and significant computing power to implement this analysis [32]. As most neoantigens have no immunogenic properties, elaborate pipelines are required to overcome all the obstacles described above to select the best candidates. The ideal algorithm for peptide selection should be able to answer the following general questions: (1) Is the gene containing a somatic mutation transcribed and, more importantly, is its mRNA translated into a protein that can be processed by the immunoproteasome into corresponding peptides? (2) Which of the peptides could result from proteasome-mediated degradation of proteins, and which of them could reach the MHC and interact with the MHC? (3) Which of these peptides have the highest affinity to the MHC and are more likely to be presented on the cell surface? (4) Can TCRs recognize the MHC-bound neoantigen complex?

It is important to note that the approaches mentioned above require downstream data analysis tools and algorithms for obtaining reliable results that could be used in clinical practice. Significant progress was made in this field during the last decade. The accuracy of neoantigen prediction is a top priority in the process of neoantigen vaccine development. Here we attempt to give a brief review of existing methods that are used to investigate neoantigens, including genomics-based in silico predictions, MS- and structure-based approaches, and describe their possible interactions and cross-validation potential. This review does not aim to give a detailed description of the available approaches and tools that are described in numerous reviews (see, e.g., [49,50,52,61–63]). It is meant to provide a bird's eye view of the main trends in the context of neoantigen identification, present interactions between different approaches and propose possible improvements.

2. Genomics-Based Approaches and Current Bioinformatics Pipelines

Currently, genomics-based strategies are some of the most promising in the field of neoantigen development. The widespread use of NGS-based techniques stimulates the development of bioinformatics tools, including those that are implemented in clinical practice. In the context of neoantigen discovery, it is important to note that the accuracy of peptide selection significantly depends

on bioinformatics pipelines that are applied to the processing of the data obtained by WES and RNA-seq. A limited number of complex pipelines for these purposes were developed and described during the last several years [64–68]. For detailed information of selected currently available genomics-centered pipelines, see Table 1. Most of them combine principally the same set of tools (in terms of tool class) intended to carry out the main steps of analysis including raw data pre-processing to remove low-quality data, mapping to the reference genome, somatic mutation calling, mutated peptide sequences isolation and peptide ranking according to their predicted capacity to be presented on the cell surface by the MHC and recognized by the TCR. A general overview of these steps is shown in Figure 1. Current comprehensive best practices in bioinformatics in the context of neoantigen identification are presented in [50]. In this report, authors not only describe the appropriate tools for each analysis step but also provide fundamental guidelines that could serve as a basis for creating standardized consensus rules for neoantigen research.

In general, the preliminary data that can be extracted during WES/WGS and RNA-seq analysis consist of bam-files containing sequenced reads aligned to the reference genome, a set of germline/somatic mutations, the patient's HLA allotypes, estimates of gene expression levels as well as information regarding the abundance of transcript isoforms. Somatic mutation data and RNA-seq alignments are also used to determine mutant protein sequences.

Somatic variant identification is one of the most important and, at the same time, one of the most delicate parts of all pipelines. It is now firmly established that not only neoantigens arising from single nucleotide variants (SNV) could be candidates for vaccines. Other mutation types are also considered to be sources of neoantigens. Among them are INDELs (short insertions and deletions) [69], gene fusions [70], exon-exon junctions [71], intron retentions [72] and some other alternative splicing events [46]. RNA transcription and splicing errors [73], as well as RNA editing examples [46], could also be recognized as neoantigen sources. Non-coding genome regions such as non-coding exons, UTRs, non-coding RNAs, and others could also be neoantigen sources [74]. This list could potentially be extended by V(D)J recombination and somatic hypermutation events [75,76] that are important for blood malignancies, and sequences of viruses that are associated with some tumors [77,78]. Proteasome-generated spliced peptides, as well as peptides bearing tumor-specific post-translational modifications, could also be a source of neoantigens [60,79], but they are out of the scope of NGS-based approaches. It was reported that neoantigens resulting from non-SNV variants could make up to 15% of all neoantigens [80]. Some authors state that non-coding regions could be the main source of neoepitopes [74]. Moreover, recent proteogenomics studies of ovarian cancer revealed that the composition of tumor-specific antigens resulting from non-mutated non-exonic regions includes 29% of intronic and 22% of intergenic sequences, and most importantly, many of them are shared across tumors [81]. Thus the variety of mutation types makes it necessary to select the right tool for the identification of each type if this tool is available. Tools for the identification of some of the mutation types listed above are discussed in [50]; additionally, comprehensive comparisons can be found in [63,82,83]. Mutect2 and Strelka2 are the most reliable somatic variant callers for SNV identification [63]. It is advisable to run several somatic callers simultaneously, which could potentially improve calling accuracy [84]. It is also good practice to conduct manual verification of somatic mutation caller results by viewing them in genomic browsers and to carry out additional validation utilizing targeted sequencing approaches [50]. Identification of other neoantigen sources is also possible due to tools such as Strelka [85] and EBCall [86], which are designed for INDELs calling, and Pindel [87], which is a specialized tool for calling large INDELs. A variety of tools for gene fusion identification were also developed, such as INTEGRATE [44] (and INTEGRATE-neo pipeline [88]), STAR-fusion [45], etc. There is now a clear demand for the development of tools that can provide proper identification of all the neoantigen sources listed above.

Furthermore, after detecting all the variants of interest, one wants to know whether they could, in principle, yield a neoantigen that has a chance to bind to the MHC molecule. Firstly, it is well known that the immunoproteasome has a limited specificity, which means that not every possible mutated

peptide will be produced during protein degradation [89]. Secondly, not all peptides produced by the proteasome would reach the required compartment of cells and could, in principle, interact with the MHC. It is known that before being presented by MHC class I, peptides are at first transported into the EPR (endoplasmic reticulum) by special transporters known as TAP (Transporter associated with antigen processing) and then trimmed by ER-related aminopeptidases (ERAP) [90]. There are several tools assessing TAP transport efficiency for peptides [91–94] and a number of tools that allow us to take proteasome cleavage specificity into account, such as NetChop20S, ProteaSMM [93,95] for MHC class I pathway and PepCleaveCD4, MHC II NP [96,97] for MHC class II pathway. It should also be taken into consideration that genes that code transporters of antigen-presenting machinery such as *TAP1*, *TAP2*, *B2M*, etc., can have mutations influencing their activity, and that these genes can have different expression levels in various tumor types, which has an additional impact on peptide presentation [98,99]. Thus, taking proteasome cleavage specificity and TAP transport limitations into account, the final list of peptides based on identified somatic variants should be created and subjected to subsequent prioritization procedures.

As mentioned above, currently available epitope prediction algorithms are based on the idea that the affinity of the peptide to a given MHC class molecule is the dominant contributor to neoantigen immunogenicity, and thus this parameter is considered to be the primary factor for peptide prioritization. It relies on the observation that only about 1 of 10,000 peptides resulting from protein degradation will be presented by the MHC [100]. It is also well-known that different MHC allotypes differ in specificity with respect to peptide binding. Therefore, it is crucial to know the HLA type before ranking peptides. The gold standard for HLA allotype determination is clinical HLA typing by sequence-specific PCR [101,102]. However, currently available HLA typers based on WES/RNA-seq data provide a high enough accuracy rate and can also be used for HLA allotype identification when a clinical HLA type is unavailable. Although HLA class I typing algorithms can reach an accuracy of up to 99% [103,104], HLA class II typers remain less effective and require additional development. It is no less important to estimate HLA locus gene expression as well as to determine somatic mutation patterns in this locus, as they both can be a cause of neoantigen presentation loss leading to resistance to immunotherapy [105–107].

Prediction of peptide-MHC binding affinity is the most critical step of the neoantigen discovery process. Many tools for such analysis exist [57,108–110]. These tools utilize large-scale peptide-MHC binding affinity data derived from biochemical measurements and eluted ligands data obtained by high-throughput mass-spectrometry analysis of MHC ligandome [57,111] to train machine learning-based classifiers that can identify binders and non-binders and calculate affinity scores. The machine learning approaches include linear regression (LR) and artificial neural networks (ANN). Depending on the experimental data that are used to train these algorithms, they can be classified on binding affinity (BA) trained methods, eluted ligands (EL) trained methods, and mixed trained methods utilizing both BA- and EL-datasets. Since the performance of different algorithms varies, a number of comprehensive benchmarking studies were carried out to compare the accuracy of these tools [48,49,112,113]. For instance, according to [49], where a dataset for 32 HLA class I and 24 HLA class II was used, ANN-based approaches showed better performance than LR-based, and among 19 predictors that were benchmarked, MHCflurry (AUC = 0.911 ± 0.010) and ann_align (AUC = 0.911 ± 0.004) showed the highest accuracy in terms of the AUC (Area Under ROC Curve) for MHC class I 9-mer and MHC class II 15-mer, respectively, in binding versus non-binding classification. In another benchmarking study [48], using an experimentally validated dataset with binding affinity data for 743 peptides (8- to 11-mers), derived from the HPV16 E6 and E7 proteins, none of the algorithms outperformed the others. However, different algorithms showed better performance for particular HLA types and peptide lengths [48]. In one of the most recent benchmarking studies [114], the performance of 15 algorithms was tested on a dataset described in [115], which contains 220 naturally processed vaccinia virus (VACV) peptides that were eluted from VACV-infected cells and tested for T cell immune response in infected C57Bl/6 mice. ANN-based NetMHCpan 4.0-L (AUC = 0.977),

NetMHCpan 4.0-B (AUC = 0.975) and MHCflurry-L (AUC = 0.973) were reported to achieve the best performance which was in general agreement with the results previously reported in [49]. More recently, improved versions of NetMHCpan (v.4.1) and NetMHCIIpan (v.4.0) as well as MHCflurry (v.2.0) were presented [57,109]. In [57] NNAlign_MA was used to update NetMHCpan and NetMHCIIpan which outperformed the current state-of-the-art methods including NetMHCpan 4.0 and MHCflurry. O'Donnell et al. incorporated an antigen processing predictor that uses data on MHC ligands, identified by mass-spectrometry, into MHCflurry 2.0 [109], allowing it to achieve better accuracy than the currently available tools. It seems logical that the simultaneous use of several MHC-binding predictors could improve peptide prioritization. It should be noted that currently available MHC-binding predictors suffer from inadequate support for rare MHC alleles and poor performance for MHC class II molecules. Another significant inherited weakness of this approach is the failure to consider the effect of post-translational modification on binding affinity. Despite these weaknesses, this approach is the gold standard in the prediction of MHC-peptide interactions.

It is well-known that not all peptides presented by the MHC (pMHC complexes) trigger T cell activation [116,117]. For instance, in [117], the authors summarized data on candidate neoantigens predicted to be MHC-binders from 13 suitable published works, which included information about assessing the peptides' immunogenic potential. It turned out that only 53 of 1948 neopeptide-MHC combinations elicited T cell response. In [118] it was reported that among 50 long peptides (MHC-binding prediction was performed using NetMHC 3.0) that were selected based on non-synonymous 563 somatic mutations in genes that are expressed in B16F10 murine melanoma, only one-third were immunogenic, and 60% of them elicited immune response directed against the mutated sequences. According to [119], only 25 of 66 27-mer peptides selected by predicted binding affinity to MHC I and MHC II and expression level were immunogenic according to IFNg ELISpot assay. Remarkably, in mouse models, the majority of immunogenic neoantigens (up to 90%) were associated with $CD4^+$ T cell response [118–120]. Since the primary goal of neoantigen identification (in the context of cancer vaccines development) is to select those that would trigger or boost T-cell-mediated immune response (preferably $CD8^+$ T cell response), it is essential to know which of the peptides with a high MHC binding affinity will be recognized by T-cells. This brings about the challenge of determining the specificity of MHC-epitope-TCR interactions, which could be an additional layer of the neoantigens ranking process. It is an established fact that T cells recognize pMHC complexes predominantly by the complementarity determining region 3 (CDR3) loops of the TCR [121]. Based on the fact that different individuals having different TCR repertoires can recognize the same epitopes arising from the same agents (e.g., immunodominant viral epitopes [122–124]), one may suggest that such epitopes have intrinsic patterns that make them more recognizable by the TCR. On the other hand, it was observed that TCR repertoires that are specific to the same epitope have similarities in their core sequences [125]. Such reasoning allows us to suggest that it is possible to perform a simulation based on sequences of peptides and TCR repertoires. Several approaches to predicting epitope-TCR binding were developed (e.g., TCRex [126], NetTCR [127], Repitope [128], ERGO [129], Deepwalk approach [130]). For instance, TCRex is based on the principle that similar TCR sequences often target the same epitope [126], Repitope is based on the idea that sequences of epitopes contain some intrinsic hidden pattern that is prone to activating T cell response [128]. Unfortunately, this class of tools is at the initial stage of development, and their prediction power suffers from insufficient training data on TCR–epitope interactions. Meanwhile, in the present time, other strategies are being successfully implemented to improve the immunogenicity of neoantigens [131,132]. Thus, in [131] the weak B16F10 neoantigens described in [118] were fused to the transmembrane domain of diphtheria toxin (DTT), significantly enhancing their ability to elicit $CD8^+$ T cell response and inhibit tumor growth. A bi-adjuvant vaccine containing a neoantigen supplemented with two adjuvants such as the Toll-like receptor (TLR) 7/8 agonist R848 and the TLR9 agonist CpG, boosted the immunogenicity of the neoantigen due to efficient co-delivery and synergism of adjuvants [132].

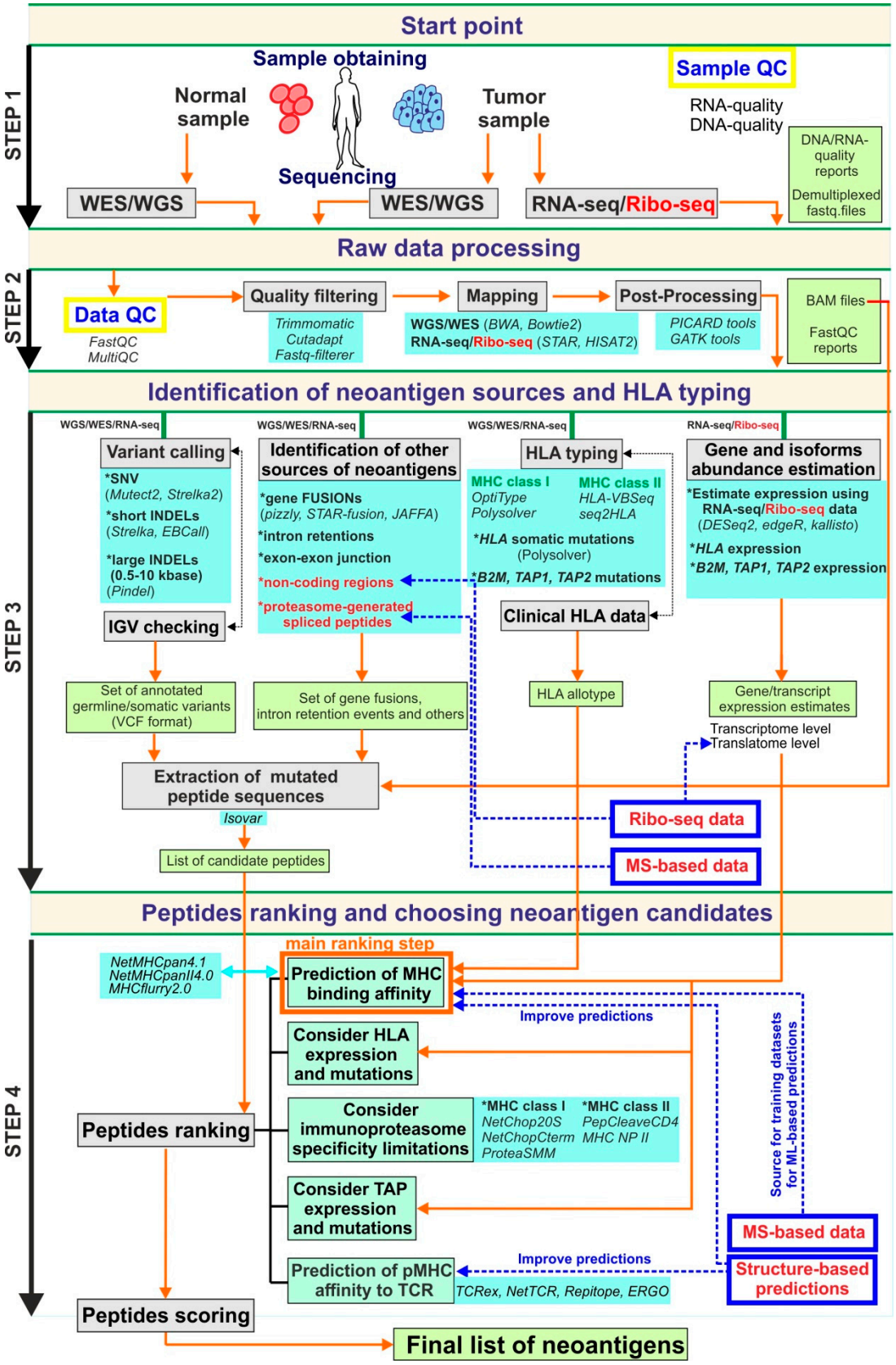

Figure 1. The schematic description of the possible ideal genomics-centric pipeline for neoantigen identification. In this scheme, the pipeline is formally split into four steps. Step 1 is related to sample obtaining, DNA/RNA-extraction, libraries preparation and high-throughput sequencing. Step 2 is

associated with raw NGS data processing, quality-filtering and obtaining aligned to reference reads in an appropriate format suitable for downstream analysis. The aim of Step 3 is to obtain all possible information from processed NGS including all variants set, HLA allotype, expression estimations, as well as, candidate peptide sequences that further to prioritization procedures of Step 4. At the final Step 4 candidate peptide list obtaining based on identified variants used for peptide ranking using mainly MHC binding estimators. Additional options such as TCR-pMHC binding affinity scoring, TAP-transport, etc., should be considered during the prioritization step. Grey boxes are reflected main procedures that should be done in each step; light green boxes contain data that should be generated on each step; light blue boxes present possible tools that could be applied for the corresponding procedure; blue boxes with red text (as well as red text alone) are related to additional approaches that could add value to this pipeline; the orange arrows show the information flow through pipeline workflow; the blue dashed arrows highlight the steps that could be improved by additional approaches.

At present, NGS techniques seem to play the primary role in the identification of therapeutic variants of neoantigens. However, the current implementation of genomics-based approaches with somatic mutations discovery as the initial step cannot answer several crucial questions. One of them is whether these mutations lead to the production of mutated proteins that can act as sources of neoantigens. It is believed that the level of mRNA does not always correlate with the protein level because not all mRNAs are translated with the same efficiency [133]. Moreover, some mRNAs are not translated at all due to being sequestered from the actively translated pool, for example, by deposition in P bodies [134]. A possible improvement that could solve this issue is based on utilizing a relatively new high-throughput technique called ribosome profiling developed in 2009 [135]. This approach involves high-throughput sequencing of ribosome-protected mRNA fragments and allows us to identify all translated mRNA, providing a snapshot of the total cellular translatome. Thus, the problem of protein production from mRNAs can be almost solved. It could also potentially reveal the translation of previously mentioned non-coding genome regions, revealing additional sources of tumor-specific neoantigens [136]. Nevertheless, genomics-based approaches are unable to solve all challenges related to the effects of post-translation modification of peptides on peptide stability and the ability to be bound by the MHC and cannot reveal proteasome-generated spliced peptide isoforms. Additionally, a lot of problems arise from highly polymorphic MHC molecules. Rare allotypes, especially MHC class II, are not supported by a sufficient volume of experimental data regarding the possibility of these types of MHC to bind peptides, making a precise ranking of neoantigens by their affinity to these MHC molecules impossible. However, integration of the above-mentioned steps into the "ideal" pipeline could significantly improve the accuracy of neoantigen prediction. Only by selecting the appropriate neoantigens can specific immune response against tumors be facilitated in clinical practice, as shown in Figure 1.

Table 1. The list of currently available computational pipelines for neoantigen prediction *.

Pipeline	Source, Required Input Data and Otput:	Workflow and Features:	Refs.
EpiToolkit 2015	Source: http://www.epitoolkit.de (not available) Description: Web-based pipeline focused on vaccine design. It includes simplified interfaces allowing to combine tools into a workflow. Input: Not described. Output: Interactive presentation of the results as HTML and Internal representation (List of predicted peptides with scores).	1. MHC genotyping (OptyType) 2. Epitope and neoepitope prediction (SYFPEITHI, BIMAS, SVMHC, NetMHC family, UniTope, TEPTITOPEpzn) 3. Epitope selection for vaccine design 4. Epitope assembly	[137]
FRED2 (FRamework for Epitope Detection) 2016	Source: https://github.com/FRED-2/Fred2 Description: Computational pipeline for T-cell epitope detection and vaccine design implemented in Python. Can be extended by additional tools. Input: Sequencing reads (FASTA format). Output: Not described.	1. HLA typing (OptiType, Polysolver, seq2HLA, ATHLATES) 2. T-cell epitope prediction • Epitope prediction (NetMHC 3.0) • TAPPrediction • CleavagePrediction (NetChop) 3. Epitope selection (OptiTope) 4. Epitope assembly (String-of-Beads, Spacer Design)	[138]
TepiTool 2016	Source: http://tools.iedb.org/tepitool/ Description: Web-based user-friendly computational pipeline for T cell epitope prediction hosted by IEDB. It is applicable to human, chimpanzee, cow, gorilla, macaque, mouse and pig. The web-tool associated article contains a step-by-step protocol of analysis with a comprehensive description of each step, recommendations to do, and a description of anticipated results. Input: Protein sequences in single-letter amino acid code (FASTA format), the list of HLA alleles. Output: Tables with peptide sequences with predicted features.	1. Provide sequence data 2. Select the host species and MHC allele class 3. Select the alleles for binding prediction 4. Select peptides to be included in the prediction 5. Select preferred methods for binding prediction and peptide selection and cutoff values (for MHC class I—Consensus (IEDB recommended 2006), NetMHCpan 2.8, NetMHC 3.4, etc; for MHC class II - Consensus (IEDB recommended 2006), NetMHCIIpan 3.0, NetMHCII 2.2, etc.) 6. Review selection, enter job details and submit data	[139]
Vaxrank 2017	Source: https://github.com/openvax/vaxrank Description: Computational framework for selecting neoantigens for vaccine peptides based on tumor mutations, tumor RNA sequencing and HLA type data. It was designed and used in the Personalized Genomic Vaccine Phase I trial (NCT02721043). Input: Tumor mutations (VCF format), tumor RNA-seq (BAM format), patient HLA alleles. Output: Set of vaccine peptides.	1. Determination of RNA abundance and extraction of mutated protein sequences 2. Predicting MHC binding (MHCtools) 3. Ranking mutant sequences 4. Optimizing sequences for peptide synthesis	[66,67]

Table 1. *Cont.*

Pipeline	Source, Required Input Data and Otput:	Workflow and Features:	Refs.
neoantigeneR 2017	Source: https://rdrr.io/github/tangshao2016/neoantigenR/ Description: R-based pipelines for neoantigen prediction using raw NGS data. Input: DNA-Seq, RNA-Seq, ExomeSeq (tumor and/or normal) short or long sequence reads (FASTA format), GFF annotation. Output: The list of high-affinity HLA class I binding neoantigen candidates.	1. Sequence alignment and isoform calling (Bowtie2, Cufflinks) 2. Epitope prediction: extracting putative novel peptide sequences 3. Candidate scoring by MHC binding prediction (NetMHC 3.4)	[140]
CloudNeo 2017	Source: https://github.com/TheJacksonLaboratory/CloudNeo Description: Cloud-based (implemented on CWL) workflow for neoantigen identification using NGS data. Input: VCF format (list of non-synonymous mutations), BAM format (for HLA typing). Output: HLA binding affinity predictions for all mutated peptides.	1. VCF processing and extraction of mutated peptide sequences (Protein_Translator) 2. HLA typing (Polysolver, HLAminer) 3. Peptide-MHC affinity prediction (NetMHCpan 3.0)	[141]
MuPexi (Mutant peptide extractor and informer) 2017	Source: http://www.cbs.dtu.dk/services/MuPeXI/ Description: Web-based tool for neo-epitope identification using somatic mutation calls (SNV, INDELs) and obtaining information about HLA binding affinity, expression level, similarities to self-peptides and mutant allele frequency for each mutated peptide. Supplemented by brief instructions and output format description. Input: Somatic mutation calls (VCF format), list of HLA types, gene expression profile (optional). Output: Table with all tumor-specific peptides derived from substitutions, insertions and deletions with annotation (HLA binding affinity and similarity to normal peptides).	1. Effect prediction (The Ensembl Variant Effect Predictor)—selecting of non-synonymous mutations 2. Neo-peptide extraction 3. The similarity to normal peptide estimation: removing mutated peptides similar to peptides in the human proteome from prioritization 4. Prediction of HLA binding (NetMHCpan 3.0) 5. Gene expression profiling 6. Annotation 7. Prioritization	[142]
TIminer (Tumor Immunology miner) 2017	Source: https://icbi.imed.ac.at/software/timiner/timiner.shtml (not available) Description: Computational framework that provides complex immunogenomic analysis including HLA typing, neoantigens prediction, characterization of immune infiltrates and quantification of tumor immunogenicity. Input: RNA-seq reads (FASTQ format), somatic DNA mutations (VCF format). Output: Not described.	1. HLA genotyping (Optitype) 2. Prediction of tumor neoantigens (NetMHCpan 3.0) 3. Characterization of tumor-infiltrating immune cells from bulk RNA-seq data (kallisto) 4. Quantification of tumor immunogenicity from expression data	[143]

Table 1. *Cont.*

Pipeline	Source, Required Input Data and Otput:	Workflow and Features:	Refs.
TSNAD (Tumor-specific neoantigen detector) 2017	Source: https://github.com/jiujiezz/tsnad Description: Pipeline with GUI allowing to identify tumor-specific mutant proteins according to GATK best practices. It provides two strategies: 1.Extraction of extracellular mutations from membrane proteins; 2. MHC affinity prediction for class I MHC. Allows us to start from raw NGS data. Input: Pair-ended sequencing data (FASTQ format) from WES. Output: List of somatic mutations with annotations, extracellular mutations of the membrane proteins and the MHC-binding information (TXT format).	1. Detection of cancer somatic mutations according to GATK best practices (Trimmomatic, BWA, samtools, Picard tools, GATK tools, ANNOVAR) 2. Prediction of neoantigens (TMHMM—for extracellular mutations, NetMHCpan 2.8—for MHC-binding affinity prediction for class I MHC).	[144]
INTEGRATE-neo 2017	Source: https://github.com/ChrisMaherLab/INTEGRATE-Neo Description: The pipeline is focused on the discovery of neoantigens derived from gene fusions. Input: Reads in FASTQ format, the human reference genome in FASTA format, gene models in GenePred format, genes fusion in BEDPE format predicted by INTEGRATE. Output: BEDPE format file.	1. Gene fusion peptide prediction 2. HLA allele prediction (HLAminer) 3. Gene fusion neoantigen discovery (NetMHC 4.0)	[88]
NeoepitopePred 2017	Source: https://github.com/stjude/NeoepitopePred Description: Workflow for identification of putative neoepitopes derived from SNV and gene fusions based on WGS data. Input: FASTQ format (PE or SE) or BAM format files, Output: Not described.	1. HLA typing—stjude-hlatype applet (OptiType) 2. Predict affinity of peptides to HLA—stjude-epitope applet (NetMHCcons 1.1) 3. Identification of Fusion junctions (CICERO)	[145]
Neopepsee 2018	Source: https://sourceforge.net/projects/neopepsee/ Description: Machine learning-based neoantigen prediction tool for NGS data. Input: Raw RNA-seq data (FASTQ format) and list of somatic mutations (VCF format), clinical HLA typing (if available) Output: mutated peptide sequences and gene expression levels, determination of immunogenic neoantigens.	1. Transcript isoform prediction 2. HLA type prediction (HLAminer) 3. MHC binding affinity prediction (IEDB-Peptide binding to MHC class I molecules) 4. Feature calculation 5. Immunogenicity classification (IEDB-T cell class I pMHC immunogenicity predictor)	[65]

Table 1. *Cont.*

Pipeline	Source, Required Input Data and Otput:	Workflow and Features:	Refs.
ScanNeo 2019	Source: https://github.com/ylab-hi/ScanNeo Description: Computational pipeline for the identification of short and large indels-derived neoantigens utilizing RNA-seq data. ScanNeo consists of independent modules implementing three analysis steps. Input: RNA-seq data in BAM format. Output: Ranked set of neoantigens.	1. Indels discovery: • duplicated reads removal (Picard tools) • spliced reads removal (sambamba) • realignment (BWA-MEM) • indels calling (transIndel) 2. Annotation and filtering • Putative PCR slippage derived indels removal • Indel annotation (Variant Effect Predictor) • Germline indels removal 3. Neoantigen prediction • Indel-derived peptide sequences generation (pVac-seq) • High-affinity peptides prediction (NetMHC 3.0 and NetMHCpan 3.0) • Prediction results merging and filtering Note: HLA typing carries out using yara aligner and OptiType tool or HLA type provides by the user.	[146]
DeepHLApan 2019	Source: http://biopharm.zju.edu.cn/deephlapan/ Description: Deep learning approach for neoantigen prediction considering both HLA-peptide binding (binding model) and immunogenicity (immunogenicity model) of peptide-HLA complex. Input: CSV format files with head of "Annotation,HLA,peptide". Only HLA-A,B,C alleles. Output: Binding score (ranges from 0 to 1, the probability that peptide binds with HLA), Immunogenicity score (ranges from 0 to 1; 0.5 is the threshold to select the predicted immunogenic pHLA).	1. The binding model for predicting the probability of the peptide being presented to the tumor cell membrane by HLA 2. Immunogenicity model for predicting the potential of pHLA eliciting T-cell activation.	[147]

Table 1. *Cont.*

Pipeline	Source, Required Input Data and Otput:	Workflow and Features:	Refs.
pTuneous (prioritizing tumor neoantigens from next-generation sequencing data) 2019	Source: https://github.com/bm2-lab/pTuneos Description: In silico tool to predict the immunogenicity of SNV-derived neoepitopes that consider MHC presentation and T-cell recognition ability. It is based on experimentally validated neoantigens. It contains Pre&RecNeo module—learning-based framework allowing to predict and prioritize neoepitopes recognized by T cells and RefinedNeo module—neoepitope scoring schema allowing to evaluate the naturally processed and presented neoepitope immunogenicity Input: PairMatchDNA (WES) mode accept WES and RNA-seq sequencing data (FASTQ format), VCF mode accepts VCF format file with mutation set, expression profile (e.g., obtained by *kallisto*), copy number profile (e.g., obtained by *sequenza*). Output: TSV files (snv_neo_model.tsv and indel_neo_models.tsv) containing extracted mutated peptides derived from non-synonimous SNV and INDELs and corresponding immunity score measures.	WES mode: • Sequencing quality control (Trimmomatic) • Mutation calling (Strelka) • HLA typing (Optitype) • Expression profiling (kallisto) • Neoantigen prediction, filtering and annotation (NetMHCpan 4.0) VCF mode: • Neoantigen prediction, filtering and annotation	[148]
NeoPredPipe 2019	Source: https://github.com/MathOnco/NeoPredPipe Description: Pipeline that provides predictions on multi-region sequence data and assessing intra-tumor heterogeneity (IHC) of the antigenic landscape of tumors. Input: Multi- or single region VCF files (with a set of somatic mutations), Patient HLA Types (optional) Output: Annotated variants, predicted neoantigens, predicted recognition potential, a summary of IHC statistics	1. Variant annotation (ANNOVAR) 2. Neoantigen prediction (NetMHCpan 4.0) 3. Peptide matching 4. Neoantigen recognition potential	[68]
pVACtools 2020	Source: https://pvactools.readthedocs.io/en/latest/ Description: Computational toolkit allowing identification of altered peptides derived from SNV, INDELs, gene fusions and providing prediction of peptide-MHC binding for MHC class I and class II. Input: VCF format files, FASTA with peptides Output: A set of files containing information about predicted epitopes before and after the filtering process supplying information about binding affinity scores and other parameters.	1. Prediction of neoantigens from somatic alterations (pVACseq and pVACfuse (gene fusions)) 2. Prediction of neoantigens for peptides in a FASTA file 3. Prioritization and selection (pVACviz with graphics-based interface) 4. Design of DNA and synthetic long peptide-based vaccines (pVACvector)	[64,149]

Table 1. *Cont.*

Pipeline	Source, Required Input Data and Otput:	Workflow and Features:	Refs.
ProGeo-neo 2020	Source: https://github.com/kbvstmd/ProGeo-neo Description: Neoantigen prediction workflow that integrates genomic and mass spectrometry data. It consists of three modules: construction of customized protein sequence database, HLA allele prediction, neoantigen prediction and filtration. Input: RNA-seq data (FASTQ format), Genomic variants (VCF format), LC-MS/MS data (Raw format). Output: List of candidate peptides	1. HLA typing (OptiType) 2. Identification tumor-specific antigens for NGS data (WES/RNA-seq) (BWA, GATK tools) 3. MHC binding prediction (NetMHCpan 4.0) 4. Verifying MHC-peptides using mass spectrometry data (MaxQuant) 5. Checking potential immunogenicity of T-cell-recognition	[150]
Neoepiscope 2020	Source: https://github.com/pdxgx/neoepiscope Description: Neoepitope identification pipeline that incorporates germline context and considers variant phasing for SNV and indels. Requires DNA-sequencing data. Input: Set of somatic and germline mutations (VCF format), BAM files. Output: TSV file with the information of mutations and neoepitopes	1. VCF files preprocessing (merging somatic and germline variants) 2. Haplotype phasing (HapCUT2) 3. Neoepitope prediction (MHCflurry, MHCnuggets, etc.)	[80]
neoANT-HILL 2020	Source: https://github.com/neoanthill/neoANT-HILL Description: User-friendly python-based toolkit that combines several pipelines that ensure fully-automated identification of potential neoantigens with a graphical interface. It allows starting from raw NGS data as well as ready-to-use variant calls. Input: Somatic variants (VCF format) and/or RNA-seq data (raw or aligned) Output: User-defined generic directory that contains variant calling data, FASTA with WT and MT sequences, predicted HLA types, gene expression estimates, tumor-infiltrating immune cells quantifications.	1. Expression estimation (kallisto) 2. Variant discovery (GATK tools) 3. HLA typing (OptiType) 4. Tumor-infiltrating immune-cell estimation (quanTIseq) 5. Variant annotation (snpEff) 6. MHC binding affinity prediction (IEDB tools, MHCflurry)	[151]
INeo-Epp 2020	Source: http://www.biostatistics.online/INeo-Epp/antigen.php Description: User-friendly web-tool implementing T-cell HLA class I immunogenicity prediction method based on sequence-related amino acid features utilizing the random forest algorithm. Input: Candidate peptide sequences (8-12 aa recommended), HLA allotype Output: Table containing peptides sequences annotated with score, %rank and prediction.	1. Providing peptide sequences and HLA types. 2. Annotation of peptides with score metrics. 3. Selecting immunogenic peptides with a score > 0.5 as recommended.	[152]

* The descriptions of the pipelines presented in the table are based on information provided in associated articles and obtained from the web-based source descriptions that are available on source websites. It is limited by highlighting the main features that distinguished the pipelines from each other. The date of the pipeline appearance is based on the publishing date of the supported article if other information is not provided. The source link is cited as "not available" if the website was not available at the time of writing. The output and input descriptions are presented as described in supporting articles or web-based sources (if available). In cases where a clear description was lacking, these fields were cited as "Not described". "Workflow and features" field contains information on the main steps that are available within the workflow. The main tools utilized as a part of the described workflows are also provided if they are described in supporting articles or in web-based sources.

3. Mass Spectrometry-Based Approaches

Genomics-based approaches represent the gold standard that is applied for neoantigen vaccine development, including in silico peptide prediction. Neoantigen candidate selection relies on the spectra of somatic mutations identified by WES/RNA-seq. This approach suffers from a lack of direct experimental evidence of the real presence of predicted epitopes on the cell surface as a complex with MHC molecules [153,154]. Lacking data could be obtained using high-throughput mass spectrometry techniques [153,154] that at present allow us to analyze large amounts of peptides or whole proteins simultaneously. This review does not aim to give a detailed characterization of MS-based approaches; for a comprehensive review on this topic, the reader could refer to [56,153,154].

A typical MS workflow (IP-based) starts with immunoprecipitation of MHC-peptide complexes using beads conjugated with MHC-specific antibodies or beads bound with dummy antibodies as negative controls. Subsequent washing steps ensure the removal of unbound and non-specifically bound peptides, whereupon the eluted material is subjected to MS analysis. Another strategy is mild acid elution (MAE) of MHC-bound peptides from the cell surface by treatment under mildly acidic conditions [155], followed by MS analysis. This method has a significant false-positive rate and low specificity due to contamination with a large quantity of non-specific peptides. Detailed comparison of IP- and MAE-based approaches are presented in [156]. To find information on MHC peptidome identification by MS approach, the reader could refer to Zhang et al. [154] where authors provide a summary of 40 studies that were carried out from 1990 to 2019.

Unlike the genomics-based approach, which only provides for neoantigen prediction, mass-spectrometry allows us to take a real snapshot of the total MHC-bound protein interactome. Additionally, it could reveal not only neoantigens that originate from somatic mutation variants but also those which arise due to proteasome-mediated peptide splicing [157,158]. Using mass-spectrometry, it was shown that the proportion of spliced peptides relative to peptides displayed by HLA class I varies from 2-6% reported in [159] to 30% reported in [60]. Moreover, MS allows us to identify the post-translational modifications (PTM) of peptides bound to the MHC, thus shedding light on the importance of PTM for binding affinity [59]. Mass-spectrometry derived data served for the development of the first tool allowing to predict the interaction between HLA class I molecules and phosphorylated peptides [160]. In addition, MS-based profiling of HLA peptidome could generate high-quality training data that could potentially significantly improve current prediction models [57,111,161], and could also be used for benchmarking available tools.

Nevertheless, MS also has some limitations. They include low sensitivity and reproducibility. These problems are especially acute for low-abundance peptides, including tumor-specific neoantigens. Moreover, the washing stages of MHC-peptide complexes during IP could result in a loss of bound peptides. These issues impose a limitation on the initial quantities of biological material. For typical experiments, 1 g of tumor tissue or anywhere from a hundred million to billions of cells are required [156]. It should also be noted that cancer cells and tumor tissues have different HLA molecules; thus, peptides that were identified from this type of material are relevant for different HLA molecules, adding the problem of specificity of the HLA ligandome.

In summary, by combining genomics-based predictions with high-throughput HLA-ligandome mass-spectrometry data, the performance of neoantigen discovery procedures could be significantly enhanced. For instance, the currently available ProGeo-neo pipeline [150] utilizes LC-MS/MS data to verify NGS-based derived neoantigen candidates.

4. Structure-Based Approaches

Structure-based predictions are another option that can improve the state of the art in the context of neoantigen discovery [61,162]. While the genomics-based approach utilizes sequence-based methods, the structure-based prediction is additionally capable of uncovering the significance of peptide structure and physicochemical properties, as well as the importance of post-translational modifications, such as phosphorylation [163], citrullination [164], and glycosylation [165], for peptide binding to the MHC

and the TCR. Moreover, structure-based approaches could yield predictions that will be applicable to all types of MHC and TCR receptors, mitigating the limitations of small training datasets for rare MHC alleles, which are required for machine learning-based predictions.

Despite the slow progress in the development of structure-based approaches due to the need for serious computational resources and high-resolution models, some attempts in this direction were made. In 2000 Schueler-Furman et al. [166] developed an approach utilizing a pairwise potential matrix that can be applied to a wide range of MHC I molecules for predicting peptide binding. In the following years, new algorithms for the prediction of peptide-MHC complexes binding were developed. PePSSI (peptide-MHC prediction of structure through solvated interfaces) [167] is an approach that allows predicting the structure of peptides bound to HLA-A2. It includes a sampling of peptide backbone conformations and flexible movement of MHC side chains and can explicitly take water molecules at the pMHC interface into account. Initially, PePSSI was tested to predict the conformation of eight peptides bound to HLA-A2, for which crystallography data are available. Analysis of predicted structures in comparison with structures derived from X-ray models showed them to be in good agreement. In [168] a method based on molecular dynamics simulations and estimation of free energy of binding between peptides and HLA molecules was proposed. Another approach, HLAffy, is based on the strength of a mechanistic model of peptide-HLA recognition [169]. It can predict epitopes for any class I HLA by assessing the binging affinity of peptide-HLA complexes by learning pair potentials that are important for peptide binding. Notably, this list of methods and descriptions of structure-based approaches is not exhaustive. For a more comprehensive review of this topic, please refer to [162].

As was mentioned above, some neoantigens that have a high binding affinity to MHC will not be effectively recognized by the TCR [170,171], which makes them unable to trigger T cell-mediated immune response. This fact allows us to suggest the existence of some peptide features that determine their recognition by the TCR independently of MHC binding. In recently published works, it was reported that immunogenic peptides are enriched in hydrophobic and aromatic amino acids at positions interacting with the TCR [172,173]. Other parameters that are believed to influence TCR binding are amino acid charge and bulkiness, WT and mutant sequence divergence and sequence entropy [65,174]. Currently, available tools attempt to solve these challenges by considering these features in the context of the peptide sequence [65,173–175]. However, it is evident that the impact of properties such as amino acid charge and size and the composition of hydrophobic residues should be taken into account in the conformation of the peptide bound to the MHC. In this connection, structure-based predictions could be one of the possible ways to determine the impact of physicochemical features of peptides on their immunogenic potential [61,176,177]. In [177], the authors developed a flexible backbone docking protocol called TCRFlexDock utilizing RosettaDock and ZRANK and benchmarked it using 20 structures of TCR/pMHC (17 for MHC class I and 3 for MHC class II) complexes, for which resolved structures of unbound components are available. Testing revealed that protein–protein docking algorithms are able to produce accurate structural models of TCR/pMHC based on unbound component structures [177]. In [176], the authors used a force-field approach utilizing refined versions of FoldX and Rosetta force fields to perform prediction of related targets of the TCR. TCR:p:MHCII complex-based benchmark containing epitope and non-epitope containing pMHC complexes was developed, and immunogenicity was estimated by calculating interaction energies between the TCR and each of the p:MHCII complexes. It was found that the predictive power of this approach depends on the ability to predict protein-MHC complex binding and model the structure of the TCR:p:MHC complex [176]. Riley et al. [61] developed a procedure for accurate and rapid modeling of the structure of nonameric peptides bound with a common class I MHC type HLA-A2 and applied it for analyzing a dataset containing thousands of immunogenic, non-immunogenic and non-HLA2-A2 binding peptides. After that, they trained a neural network (NN) on structural features that affect TCR and peptide binding energies. It was shown that structurally-parameterized NN outperformed other methods that do not include explicit structural or energetic properties in the assessment of $CD8^+$ T cell response of HLA-A2 bound nonameric peptides [61]. Thus, a combination of MHC-binding prediction based on NGS-data

with a structure-based approach could significantly improve the accuracy of immunogenic peptide selection that is of special importance in the context of peptide-based cancer vaccine development.

5. Neoantigen Peptide Databases

A growing number of pan-cancer analyses aimed at neoantigen evaluation make it necessary to consolidate the resulting data into appropriate databases. One of the most potent sources is the IEDB database, that contains data related to immunoepitopes and their identification [65,178,179]. This database is compiled from preliminary data from the literature [178,179]. It has been maintained since 2004 and is currently being actively developed [179]. It also includes an IEDB-AR, a specialized collection of tools for the prediction and evaluation of T and B cell epitopes [180]. In recent years new databases specializing on neoantigen-related data were established (e.g., TANTIGEN, TSNAdb, NeoPeptide, dbPepNeo) [181–184]. A detailed description of these databases is provided in Table 2.

Table 2. Selected databases containing information related to tumor neoantigens *.

Database, Year of Appearance	Source and Description	Refs.
IEDB (The Immune Epitope Database) 2004	Source: https://www.iedb.org/ Description: IEDB is one of the most powerful sources of experimental data concerning immune epitope discovery. It contains information regarding T cell epitopes of human and other organisms. It also provides tools that could be useful for neoantigen prediction. They include MHC class I and II binding predictors, including proteasome cleavage and TAP transport processing steps, as well as tools for immunogenicity predictions. For a review of this topic see [180]	[178–180]
TANTIGEN and TANTIGEN 2.0 (Tumor T-cell Antigen Database) 2009	Source: http://projects.met-hilab.org/tadb/ Description: This database contains information about more than 1000 tumor peptides stemming from 292 different proteins. According to the description presented in [181], all peptides in the database are marked as belonging to one of four categories: (1) peptides measured in vitro to bind the HLA, but not reported to elicit either in vivo or in vitro T cell response, (2) peptides found to bind the HLA and to elicit an in vitro T cell response, (3) peptides shown to elicit in vivo tumor rejection, and (4) peptides processed and naturally presented as defined by physical detection. Moreover, peptides are annotated that are naturally processed HLA binders, e.g., peptides eluted from HLA in mass-spectrometry studies. The database also contains predicted binding peptides of 15 HLA class I and Class II.	[181]
TSNAdb 2018	Source: http://biopharm.zju.edu.cn/tsnadb/ Description: TSNAdb is a freely available database developed by Wu et al. [182]. It contains results of somatic mutation identification and HLA typing analysis of 7748 tumor samples of 16 different cancer types obtained from The Cancer Genome Atlas (TCGA) and The Cancer Immunome Atlas (TCIA). Based on this data, the author predicted binding affinity between mutant/wild-type peptides and HLA class I molecules using netMHCpan v2.8/v4.0. Thus, the database contains information about 3707562/1146961 potential antigens.	[182]
NeoPeptide 2019	Source: http://www.neopeptide.cn/ and https://github.com/lyotvincent/NeoPeptide Description: NeoPeptide contains information about neoantigens resulting from somatic mutations gleaned from published literature and immunological resources. As described in [183], it contains 1,818,137 epitopes obtained from more than 36,000 neoantigens that were found in different cancer types (NSCLC, breast cancer, melanoma, etc.) and specifies characteristics such as mutation site, subunit sequence, and MHC complex restriction. The database includes data concerning experimentally characterized epitopes, which are also derived from MHC binding and MHC ligand elution experiments. Information on neoantigens is cited with references to the sources.	[183]

Table 2. *Cont.*

Database, Year of Appearance	Source and Description	Refs.
dbPepNeo 2020	Source: http://www.biostatistics.online/dbPepNeo/ Description: dbPepNeo is a manually curated database of experimentally confirmed human tumor antigens that bind specifically to HLA class I, which contains information extracted from peer-reviewed articles and the publicly available data sources. The database relies on mass spectrometry (MS) validation and specific T-cell immunoassays. The peptides were classified according to validation methods: 1. Low confidence (407794): validated by MS only; 2. Medium confidence (247): contain a somatic mutation and are validated by MS and WES/WGS; 3. High confidence (295): immunogenicity was validated directly by utilizing specific T-cell response experiments. dbPepNeo also includes the following tools: ProGeo-neo (see Table 1) and INeo-Epp, a machine learning algorithm for neoepitope immunogenicity prediction using neoantigen peptide features.	[184]

* The information presented here is based on the introductions to these databases provided in the respective articles as well as on details specified on source websites. The database creation date is based on the publication date of the supporting article unless specified otherwise.

6. Conclusions

The importance of neoantigens as potential immunotherapeutic agents and prognostic biomarkers is difficult to overstate. The need to further our understanding of neoantigen involvement in the development of tumor-related T cell immune response drives research in this field. It is considered an established fact that tumor mutation burden and, most notably, neoantigen burden exhibit a significant correlation with response to immune checkpoint therapy for certain cancer types (e.g., melanoma, NSCLC, etc.), once again highlighting the effect of neoantigens on the activation of the immune response to tumors. It seems promising to implement these properties of neoantigens in clinical practice. Firstly, estimates of neoantigen prevalence could be utilized as a potential biomarker of response to immune checkpoint inhibitors therapy that has a tremendous relevance, since the number of therapeutic options grows rapidly, and well-defined criteria for developing decision-making algorithms are required. Secondly, it seems logical to apply artificially synthesized tumor-specific neoantigens (e.g., in the form of RNA/DNA or peptide-based vaccines), which are able to activate specific T-cell-mediated immune response leading to tumor degradation. Both applications need an appropriate algorithm allowing accurate and rapid identification of neoantigens specific to individual tumors. To the best of our knowledge, WES and RNA-seq are the starting point for neoantigen discovery approaches that were already used in several trials on model organisms as well as in humans. These methods were implemented in clinical practice owing to the associated significant cost reduction and the development of suitable tools allowing to analyze the generated data. In the context of neoantigen identification, genomic-based approaches potentially allow us to discover almost all types of possible neoantigen sources (with the possible exception of proteasome-generated spliced peptides). However, its application is limited by a lack of evidence about the actual existence of predicted peptides and their ability to be bound and presented by MHC molecules and recognized by the TCR. The development of best practices for the bioinformatic prediction of neoantigens [50] is an important initial point on the way to the unification and standardization of tools and methods that are intended to achieve these goals. Revealing weak points of analysis is useful to equip the current pipeline with new tools that could improve its accuracy by solving the challenges described. It is equally important to note that MS-based methods now are essential because they not only lead to a better understanding of MHC-bound proteome but also help accumulate of data that can be used for improving training datasets and extend the knowledge about the nature of new peptide isoforms (e.g., proteasome-spliced isoforms) and the effect of post-translation modifications on the affinity to the MHC as well as the ability to be recognized by TCRs. Finally, the structure-based prediction could help to overcome limitations of sequence-based approaches, e.g., training datasets for each MHC allele, by adding information

regarding the general effects of intrinsic structural and physicochemical properties of peptides on their ability to bind to the MHC and be recognized by the TCR in pMHC complexes.

Authors of this review have a clear understanding that so far the implementation of all the above-mentioned techniques simultaneously for diagnostics purposes (e.g., tumor mutation and neoantigen burden estimation) as well as for the development of personalized neoantigen-based vaccines does not seem realistic for a number of reasons, including the fact that high expenditures and a large highly-qualified team are required to implement all these methods, and a long time is required for collecting and analyzing the generated data and to producing a ready-to-use vaccine. However, at the same time, considering that each approach has its gaps, it is obvious that they cannot predict peptides with absolute certainty and do not guarantee that the predicted peptides are indeed immunogenic. To overcome emerging challenges, it is necessary to integrate efforts of all investigators in this field and create clear and standardized guidelines for the main approaches available (genomics-based, MS-based and structure-based) to make it possible to combine and accumulate generated data, which could potentially improve existing models used in predictions. By including additional high-throughput techniques, such as ribosome profiling approach, and developing tools for solving currently unresolved issues, the current state of the art could be significantly improved.

Author Contributions: A.V.G. wrote the manuscript; E.N.K. and V.S.K. critically reviewed the manuscript; E.N.K. contributed to the preparation of figures. All authors have read and agreed to the published version of the manuscript.

Funding: This work was carried out with the financial support of the Russian Ministry of Health within the framework of experimental scientific development program № AAAA-A18-118032290146-5.

Acknowledgments: The authors thank O. Kossinova (Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk) for critical reading of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
2. Falzone, L.; Salomone, S.; Libra, M. Evolution of cancer pharmacological treatments at the turn of the third millennium. *Front. Pharmacol.* **2018**, *9*. [CrossRef] [PubMed]
3. Schirrmacher, V. From chemotherapy to biological therapy: A review of novel concepts to reduce the side effects of systemic cancer treatment (Review). *Int. J. Oncol.* **2019**, *54*, 407–419. [CrossRef]
4. Urruticoechea, A.; Alemany, R.; Balart, J.; Villanueva, A.; Vinals, F.; Capella, G. Recent advances in cancer therapy: An overview. *Curr. Pharm. Des.* **2010**, *16*, 3–10. [CrossRef] [PubMed]
5. Li, B.; Chan, H.L.; Chen, P. Immune checkpoint inhibitors: Basics and challenges. *Curr. Med. Chem.* **2019**, *26*, 3009–3025. [CrossRef] [PubMed]
6. Qin, S.; Xu, L.; Yi, M.; Yu, S.; Wu, K.; Luo, S. Novel immune checkpoint targets: Moving beyond PD-1 and CTLA-4. *Mol. Cancer* **2019**, *18*, 155. [CrossRef]
7. Queirolo, P.; Boutros, A.; Tanda, E.; Spagnolo, F.; Quaglino, P. Immune-checkpoint inhibitors for the treatment of metastatic melanoma: A model of cancer immunotherapy. *Semin. Cancer Biol.* **2019**, *59*, 290–297. [CrossRef]
8. Dobry, A.S.; Zogg, C.K.; Hodi, F.S.; Smith, T.R.; Ott, P.A.; Iorgulescu, J.B. Management of metastatic melanoma: Improved survival in a national cohort following the approvals of checkpoint blockade immunotherapies and targeted therapies. *Cancer Immunol. Immunother.* **2018**, *67*, 1833–1844. [CrossRef]
9. Qiu, Z.; Chen, Z.; Zhang, C.; Zhong, W. Achievements and futures of immune checkpoint inhibitors in non-small cell lung cancer. *Exp. Hematol. Oncol.* **2019**, *8*, 19. [CrossRef]
10. Yan, Y.F.; Zheng, Y.F.; Ming, P.P.; Deng, X.X.; Ge, W.; Wu, Y.G. Immune checkpoint inhibitors in non-small-cell lung cancer: Current status and future directions. *Brief. Funct. Genom.* **2019**, *18*, 147–156. [CrossRef]
11. Flippot, R.; Escudier, B.; Albiges, L. Immune checkpoint inhibitors: Toward new paradigms in renal cell carcinoma. *Drugs* **2018**, *78*, 1443–1457. [CrossRef] [PubMed]

12. Stuhler, V.; Maas, J.M.; Rausch, S.; Stenzl, A.; Bedke, J. Immune checkpoint inhibition for the treatment of renal cell carcinoma. *Expert Opin. Biol. Ther.* **2020**, *20*, 83–94. [CrossRef] [PubMed]
13. Darvin, P.; Toor, S.M.; Sasidharan Nair, V.; Elkord, E. Immune checkpoint inhibitors: Recent progress and potential biomarkers. *Exp. Mol. Med.* **2018**, *50*, 1–11. [CrossRef] [PubMed]
14. Tolba, M.F. Revolutionizing the landscape of colorectal cancer treatment: The Potential role of immune checkpoint inhibitors. *Int. J. Cancer* **2020**. [CrossRef]
15. Park, J.C.; Faquin, W.C.; Durbeck, J.; Faden, D.L. Immune checkpoint inhibitors in sinonasal squamous cell carcinoma. *Oral Oncol.* **2020**, 104776. [CrossRef]
16. Kandalaft, L.E.; Odunsi, K.; Coukos, G. Immune therapy opportunities in ovarian cancer. *Am. Soc. Clin. Oncol. Educ. Book* **2020**, *40*, 1–13. [CrossRef]
17. Nakamura, Y. Biomarkers for immune checkpoint inhibitor-mediated tumor response and adverse events. *Front. Med.* **2019**, *6*. [CrossRef]
18. Longo, V.; Brunetti, O.; Azzariti, A.; Galetta, D.; Nardulli, P.; Leonetti, F.; Silvestris, N. Strategies to improve cancer immune checkpoint inhibitors efficacy, other than abscopal effect: A systematic review. *Cancers* **2019**, *11*, 539. [CrossRef]
19. Havel, J.J.; Chowell, D.; Chan, T.A. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **2019**, *19*, 133–150. [CrossRef]
20. Lyu, G.Y.; Yeh, Y.H.; Yeh, Y.C.; Wang, Y.C. Mutation load estimation model as a predictor of the response to cancer immunotherapy. *NPJ Genom. Med.* **2018**, *3*, 12. [CrossRef]
21. Wu, Y.; Xu, J.; Du, C.; Wu, Y.; Xia, D.; Lv, W.; Hu, J. The predictive value of tumor mutation burden on efficacy of immune checkpoint inhibitors in cancers: A systematic review and meta-analysis. *Front. Oncol.* **2019**, *9*. [CrossRef] [PubMed]
22. 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. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science (NY)* **2015**, *348*, 124–128. [CrossRef] [PubMed]
23. Yi, M.; Qin, S.; Zhao, W.; Yu, S.; Chu, Q.; Wu, K. The role of neoantigen in immune checkpoint blockade therapy. *Exp. Hematol. Oncol.* **2018**, *7*, 28. [CrossRef] [PubMed]
24. Li, L.; Rao, X.; Wen, Z.; Ding, X.; Wang, X.; Xu, W.; Meng, C.; Yi, Y.; Guan, Y.; Chen, Y.; et al. Implications of driver genes associated with a high tumor mutation burden identified using next-generation sequencing on immunotherapy in hepatocellular carcinoma. *Oncol. Lett.* **2020**, *19*, 2739–2748. [CrossRef]
25. Schumacher, T.N.; Scheper, W.; Kvistborg, P. Cancer neoantigens. *Annu. Rev. Immunol.* **2019**, *37*, 173–200. [CrossRef]
26. Chu, Y.; Liu, Q.; Wei, J.; Liu, B. Personalized cancer neoantigen vaccines come of age. *Theranostics* **2018**, *8*, 4238–4246. [CrossRef]
27. Schumacher, T.N.; Schreiber, R.D. Neoantigens in cancer immunotherapy. *Science (NY)* **2015**, *348*, 69–74. [CrossRef]
28. Chan, T.A.; Yarchoan, M.; Jaffee, E.; Swanton, C.; Quezada, S.A.; Stenzinger, A.; Peters, S. Development of tumor mutation burden as an immunotherapy biomarker: Utility for the oncology clinic. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2019**, *30*, 44–56. [CrossRef]
29. Zhou, J.; Zhao, W.; Wu, J.; Lu, J.; Ding, Y.; Wu, S.; Wang, H.; Ding, D.; Mo, F.; Zhou, Z.; et al. Neoantigens derived from recurrently mutated genes as potential immunotherapy targets for gastric cancer. *Biomed. Res. Int.* **2019**, *2019*, 8103142. [CrossRef]
30. Neefjes, J.; Jongsma, M.L.M.; Paul, P.; Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **2011**, *11*, 823–836. [CrossRef]
31. Coulie, P.G.; Van den Eynde, B.J.; van der Bruggen, P.; Boon, T. Tumour antigens recognized by T lymphocytes: At the core of cancer immunotherapy. *Nat. Rev. Cancer* **2014**, *14*, 135–146. [CrossRef] [PubMed]
32. Li, L.; Goedegebuure, S.P.; Gillanders, W.E. Preclinical and clinical development of neoantigen vaccines. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2017**, *28*, xii11–xii17. [CrossRef] [PubMed]
33. Zhang, J.Y.; Looi, K.S.; Tan, E.M. Identification of tumor-associated antigens as diagnostic and predictive biomarkers in cancer. *Methods Mol. Biol.* **2009**, *520*, 1–10. [CrossRef] [PubMed]
34. Criscitiello, C. Tumor-associated antigens in breast cancer. *Breast Care (Basel)* **2012**, *7*, 262–266. [CrossRef] [PubMed]

35. Morgan, R.A.; Chinnasamy, N.; Abate-Daga, D.; Gros, A.; Robbins, P.F.; Zheng, Z.; Dudley, M.E.; Feldman, S.A.; Yang, J.C.; Sherry, R.M.; et al. Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy. *J. Immunother. (Hagerstown, Md. 1997)* **2013**, *36*, 133–151. [CrossRef]
36. Lee, D.W.; Gardner, R.; Porter, D.L.; Louis, C.U.; Ahmed, N.; Jensen, M.; Grupp, S.A.; Mackall, C.L. Current concepts in the diagnosis and management of cytokine release syndrome. *Blood* **2014**, *124*, 188–195. [CrossRef]
37. 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. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **2017**, *547*, 217–221. [CrossRef]
38. Sahin, U.; Derhovanessian, E.; Miller, M.; Kloke, B.P.; Simon, P.; Lower, M.; Bukur, V.; Tadmor, A.D.; Luxemburger, U.; Schrors, B.; et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **2017**, *547*, 222–226. [CrossRef]
39. 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. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* **2019**, *565*, 234–239. [CrossRef]
40. Sahu, A.; Singhal, U.; Chinnaiyan, A.M. Long noncoding RNAs in cancer: From function to translation. *Trends Cancer* **2015**, *1*, 93–109. [CrossRef]
41. Suwinski, P.; Ong, C.; Ling, M.H.T.; Poh, Y.M.; Khan, A.M.; Ong, H.S. Advancing personalized medicine through the application of whole exome sequencing and big data analytics. *Front. Genet.* **2019**, *10*. [CrossRef] [PubMed]
42. Coudray, A.; Battenhouse, A.M.; Bucher, P.; Iyer, V.R. Detection and benchmarking of somatic mutations in cancer genomes using RNA-seq data. *Peer J.* **2018**, *6*, e5362. [CrossRef] [PubMed]
43. Sheng, Q.; Zhao, S.; Li, C.I.; Shyr, Y.; Guo, Y. Practicability of detecting somatic point mutation from RNA high throughput sequencing data. *Genomics* **2016**, *107*, 163–169. [CrossRef] [PubMed]
44. Zhang, J.; White, N.M.; Schmidt, H.K.; Fulton, R.S.; Tomlinson, C.; Warren, W.C.; Wilson, R.K.; Maher, C.A. Integrate: Gene fusion discovery using whole genome and transcriptome data. *Genome Res.* **2016**, *26*, 108–118. [CrossRef]
45. Haas, B.J.; Dobin, A.; Stransky, N.; Li, B.; Yang, X.; Tickle, T.; Bankapur, A.; Ganote, C.; Doak, T.G.; Pochet, N.; et al. STAR-Fusion: Fast and accurate fusion transcript detection from RNA-Seq. *bioRxiv* **2017**. [CrossRef]
46. Park, J.; Chung, Y.J. Identification of neoantigens derived from alternative splicing and RNA modification. *Genom. Inform.* **2019**, *17*, e23. [CrossRef]
47. Orenbuch, R.; Filip, I.; Comito, D.; Shaman, J.; Pe'er, I.; Rabadan, R. arcasHLA: High-resolution HLA typing from RNAseq. *Bioinformatics (Oxf. Engl.)* **2019**, *36*, 33–40. [CrossRef]
48. Bonsack, M.; Hoppe, S.; Winter, J.; Tichy, D.; Zeller, C.; Küpper, M.; Schitter, E.C.; Blatnik, R.; Riemer, A.B. Performance evaluation of MHC class-I binding prediction tools based on an experimentally validated MHC-peptide binding dataset. *Cancer Immunol. Res.* **2019**. [CrossRef]
49. Zhao, W.; Sher, X. Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes. *PLoS Comput. Biol.* **2018**, *14*, e1006457. [CrossRef]
50. Richters, M.M.; Xia, H.; Campbell, K.M.; Gillanders, W.E.; Griffith, O.L.; Griffith, M. Best practices for bioinformatic characterization of neoantigens for clinical utility. *Genome Med.* **2019**, *11*, 56. [CrossRef]
51. Gfeller, D.; Bassani-Sternberg, M. Predicting antigen presentation—What could we learn from a million peptides? *Front. Immunol.* **2018**, *9*. [CrossRef] [PubMed]
52. Matey-Hernandez, M.L.; Maretty, L.; Jensen, J.M.; Petersen, B.; Sibbesen, J.A.; Liu, S.; Villesen, P.; Skov, L.; Belling, K.; Have, C.T.; et al. Benchmarking the HLA typing performance of Polysolver and Optitype in 50 Danish parental trios. *BMC Bioinf.* **2018**, *19*, 239. [CrossRef] [PubMed]
53. Bunce, M.; Passey, B. HLA typing by sequence-specific primers. *Methods Mol. Biol.* **2013**, *1034*, 147–159. [CrossRef] [PubMed]
54. Bassani-Sternberg, M.; Braunlein, E.; Klar, R.; Engleitner, T.; Sinitcyn, P.; Audehm, S.; Straub, M.; Weber, J.; Slotta-Huspenina, J.; Specht, K.; et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* **2016**, *7*, 13404. [CrossRef]
55. Abelin, J.G.; Keskin, D.B.; Sarkizova, S.; Hartigan, C.R.; Zhang, W.; Sidney, J.; Stevens, J.; Lane, W.; Zhang, G.L.; Eisenhaure, T.M.; et al. Mass spectrometry profiling of HLA-Associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity* **2017**, *46*, 315–326. [CrossRef]

56. Creech, A.L.; Ting, Y.S.; Goulding, S.P.; Sauld, J.F.K.; Barthelme, D.; Rooney, M.S.; Addona, T.A.; Abelin, J.G. The role of mass spectrometry and proteogenomics in the advancement of HLA epitope prediction. *Proteomics* **2018**, *18*, e1700259. [CrossRef]
57. Reynisson, B.; Alvarez, B.; Paul, S.; Peters, B.; Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **2020**, *48*, W449–W454. [CrossRef]
58. Doyle, H.A.; Mamula, M.J. Post-translational protein modifications in antigen recognition and autoimmunity. *Trends Immunol.* **2001**, *22*, 443–449. [CrossRef]
59. Engelhard, V.H.; Altrich-Vanlith, M.; Ostankovitch, M.; Zarling, A.L. Post-translational modifications of naturally processed MHC-binding epitopes. *Curr. Opin. Immunol.* **2006**, *18*, 92–97. [CrossRef]
60. Liepe, J.; Marino, F.; Sidney, J.; Jeko, A.; Bunting, D.E.; Sette, A.; Kloetzel, P.M.; Stumpf, M.P.; Heck, A.J.; Mishto, M. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science (NY)* **2016**, *354*, 354–358. [CrossRef]
61. Riley, T.P.; Keller, G.L.J.; Smith, A.R.; Davancaze, L.M.; Arbuiso, A.G.; Devlin, J.R.; Baker, B.M. Structure based prediction of neoantigen immunogenicity. *Front. Immunol.* **2019**, *10*. [CrossRef] [PubMed]
62. Bassani-Sternberg, M. Mass Spectrometry Based Immunopeptidomics for the Discovery of Cancer Neoantigens. *Methods Mol. Biol. (Clifton, N.J.)* **2018**, *1719*, 209–221. [CrossRef]
63. Chen, Z.; Yuan, Y.; Chen, X.; Chen, J.; Lin, S.; Li, X.; Du, H. Systematic comparison of somatic variant calling performance among different sequencing depth and mutation frequency. *Sci. Rep.* **2020**, *10*, 3501. [CrossRef] [PubMed]
64. 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. pVACtools: A computational toolkit to identify and visualize cancer neoantigens. *Cancer Immunol. Res.* **2020**, *8*, 409. [CrossRef]
65. Kim, S.; Kim, H.S.; Kim, E.; Lee, M.G.; Shin, E.C.; Paik, S.; Kim, S. Neopepsee: Accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2018**, *29*, 1030–1036. [CrossRef]
66. Rubinsteyn, A.; Hodes, I.; Kodysh, J.; Hammerbacher, J. Vaxrank: A computational tool for designing personalized cancer vaccines. *bioRxiv* **2017**, 142919. [CrossRef]
67. Rubinsteyn, A.; Kodysh, J.; Hodes, I.; Mondet, S.; Aksoy, B.A.; Finnigan, J.P.; Bhardwaj, N.; Hammerbacher, J. Computational pipeline for the PGV-001 neoantigen vaccine trial. *Front. Immunol.* **2018**, *8*. [CrossRef]
68. Schenck, R.O.; Lakatos, E.; Gatenbee, C.; Graham, T.A.; Anderson, A.R.A. Neopredpipe: High-throughput neoantigen prediction and recognition potential pipeline. *BMC Bioinf.* **2019**, *20*, 264. [CrossRef]
69. Turajlic, S.; Litchfield, K.; Xu, H.; Rosenthal, R.; McGranahan, N.; Reading, J.L.; Wong, Y.N.S.; Rowan, A.; Kanu, N.; Al Bakir, M.; et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: A pan-cancer analysis. *Lancet Oncol.* **2017**, *18*, 1009–1021. [CrossRef]
70. 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. Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat. Med.* **2019**, *25*, 767–775. [CrossRef]
71. David, J.K.; Maden, S.K.; Weeder, B.R.; Thompson, R.F.; Nellore, A. Putatively cancer-specific exon–exon junctions are shared across patients and present in developmental and other non-cancer cells. *NAR Cancer* **2020**, *2*. [CrossRef]
72. Smart, A.C.; Margolis, C.A.; Pimentel, H.; He, M.X.; Miao, D.; Adeegbe, D.; Fugmann, T.; Wong, K.K.; Van Allen, E.M. Intron retention is a source of neoepitopes in cancer. *Nat. Biotechnol.* **2018**, *36*, 1056–1058. [CrossRef] [PubMed]
73. Shen, L.; Zhang, J.; Lee, H.; Batista, M.T.; Johnston, S.A. RNA Transcription and splicing errors as a source of cancer frameshift neoantigens for vaccines. *Sci. Rep.* **2019**, *9*, 14184. [CrossRef] [PubMed]
74. Laumont, C.M.; Vincent, K.; Hesnard, L.; Audemard, E.; Bonneil, E.; Laverdure, J.P.; Gendron, P.; Courcelles, M.; Hardy, M.P.; Cote, C.; et al. Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Trans. Med.* **2018**, *10*. [CrossRef]
75. Khodadoust, M.S.; Olsson, N.; Wagar, L.E.; Haabeth, O.A.; Chen, B.; Swaminathan, K.; Rawson, K.; Liu, C.L.; Steiner, D.; Lund, P.; et al. Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature* **2017**, *543*, 723–727. [CrossRef]

76. Khodadoust, M.S.; Olsson, N.; Chen, B.; Sworder, B.; Shree, T.; Liu, C.L.; Zhang, L.; Czerwinski, D.K.; Davis, M.M.; Levy, R.; et al. B-cell lymphomas present immunoglobulin neoantigens. *Blood* **2019**, *133*, 878–881. [CrossRef]
77. Walboomers, J.M.; Jacobs, M.V.; Manos, M.M.; Bosch, F.X.; Kummer, J.A.; Shah, K.V.; Snijders, P.J.; Peto, J.; Meijer, C.J.; Muñoz, N. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J. Pathol.* **1999**, *189*, 12–19. [CrossRef]
78. Gillison, M.L.; Koch, W.M.; Capone, R.B.; Spafford, M.; Westra, W.H.; Wu, L.; Zahurak, M.L.; Daniel, R.W.; Viglione, M.; Symer, D.E.; et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *JNCI J. Natl. Cancer Inst.* **2000**, *92*, 709–720. [CrossRef]
79. Kumai, T.; Ishibashi, K.; Oikawa, K.; Matsuda, Y.; Aoki, N.; Kimura, S.; Hayashi, S.; Kitada, M.; Harabuchi, Y.; Celis, E.; et al. Induction of tumor-reactive T helper responses by a posttranslational modified epitope from tumor protein p53. *Cancer Immunol. Immunother. Cii* **2014**, *63*, 469–478. [CrossRef]
80. Wood, M.A.; Nguyen, A.; Struck, A.J.; Ellrott, K.; Nellore, A.; Thompson, R.F. Neoepiscope improves neoepitope prediction with multivariant phasing. *Bioinformatics (Oxf. Engl.)* **2019**, *36*, 713–720. [CrossRef]
81. Zhao, Q.; Laverdure, J.P.; Lanoix, J.; Durette, C.; Coté, C.; Bonneil, E.; Laumont, C.M.; Gendron, P.; Vincent, K.; Courcelles, M.; et al. Proteogenomics uncovers a vast repertoire of shared tumor-specific antigens in ovarian cancer. *Cancer Immunol. Res.* **2020**. [CrossRef] [PubMed]
82. Warden, C.D.; Adamson, A.W.; Neuhausen, S.L.; Wu, X. Detailed comparison of two popular variant calling packages for exome and targeted exon studies. *PeerJ* **2014**, *2*, e600. [CrossRef] [PubMed]
83. Alioto, T.S.; Buchhalter, I.; Derdak, S.; Hutter, B.; Eldridge, M.D.; Hovig, E.; Heisler, L.E.; Beck, T.A.; Simpson, J.T.; Tonon, L.; et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nat. Commun.* **2015**, *6*, 10001. [CrossRef] [PubMed]
84. Callari, M.; Sammut, S.J.; De Mattos-Arruda, L.; Bruna, A.; Rueda, O.M.; Chin, S.F.; Caldas, C. Intersect-then-combine approach: Improving the performance of somatic variant calling in whole exome sequencing data using multiple aligners and callers. *Genome Med.* **2017**, *9*, 35. [CrossRef] [PubMed]
85. Saunders, C.T.; Wong, W.S.; Swamy, S.; Becq, J.; Murray, L.J.; Cheetham, R.K. Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics (Oxf. Engl.)* **2012**, *28*, 1811–1817. [CrossRef] [PubMed]
86. Shiraishi, Y.; Sato, Y.; Chiba, K.; Okuno, Y.; Nagata, Y.; Yoshida, K.; Shiba, N.; Hayashi, Y.; Kume, H.; Homma, Y.; et al. An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic Acids Res.* **2013**, *41*, e89. [CrossRef] [PubMed]
87. Ye, K.; Schulz, M.H.; Long, Q.; Apweiler, R.; Ning, Z. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics (Oxf. Engl.)* **2009**, *25*, 2865–2871. [CrossRef]
88. Zhang, J.; Mardis, E.R.; Maher, C.A. INTEGRATE-neo: A pipeline for personalized gene fusion neoantigen discovery. *Bioinformatics (Oxf. Engl.)* **2017**, *33*, 555–557. [CrossRef]
89. Sijts, E.J.A.M.; Kloetzel, P.M. The role of the proteasome in the generation of MHC class I ligands and immune responses. *Cell. Mol. Life Sci. CMLS* **2011**, *68*, 1491–1502. [CrossRef]
90. Rock, K.L.; Reits, E.; Neefjes, J. Present yourself! By MHC Class I and MHC Class II molecules. *Trends Immunol.* **2016**, *37*, 724–737. [CrossRef]
91. Peters, B.; Bulik, S.; Tampe, R.; Van Endert, P.M.; Holzhütter, H.G. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J. Immunol. (Baltimore, Md. 1950)* **2003**, *171*, 1741–1749. [CrossRef] [PubMed]
92. Bhasin, M.; Raghava, G.P.S. Analysis and prediction of affinity of TAP binding peptides using cascade SVM. *Protein Sci.* **2004**, *13*, 596–607. [CrossRef] [PubMed]
93. Tenzer, S.; Peters, B.; Bulik, S.; Schoor, O.; Lemmel, C.; Schatz, M.M.; Kloetzel, P.M.; Rammensee, H.G.; Schild, H.; Holzhütter, H.G. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell. Mol. Life Sci. CMLS* **2005**, *62*, 1025–1037. [CrossRef] [PubMed]
94. Bhasin, M.; Lata, S.; Raghava, G.P. TAPPred prediction of TAP-binding peptides in antigens. *Methods Mol. Biol.* **2007**, *409*, 381–386. [CrossRef] [PubMed]

95. Nielsen, M.; Lundegaard, C.; Lund, O.; Keşmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: Insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **2005**, *57*, 33–41. [CrossRef] [PubMed]
96. Hoze, E.; Tsaban, L.; Maman, Y.; Louzoun, Y. Predictor for the effect of amino acid composition on CD4 + T cell epitopes preprocessing. *J. Immunol. Methods* **2013**, *391*, 163–173. [CrossRef] [PubMed]
97. Paul, S.; Karosiene, E.; Dhanda, S.K.; Jurtz, V.; Edwards, L.; Nielsen, M.; Sette, A.; Peters, B. Determination of a predictive cleavage motif for eluted major histocompatibility complex class II ligands. *Front. Immunol.* **2018**, *9*, 1795. [CrossRef]
98. Romero, J.M.; Jiménez, P.; Cabrera, T.; Cózar, J.M.; Pedrinaci, S.; Tallada, M.; Garrido, F.; Ruiz-Cabello, F. Coordinated downregulation of the antigen presentation machinery and HLA class I/beta2-microglobulin complex is responsible for HLA-ABC loss in bladder cancer. *Int. J. Cancer* **2005**, *113*, 605–610. [CrossRef]
99. Leone, P.; Shin, E.C.; Perosa, F.; Vacca, A.; Dammacco, F.; Racanelli, V. MHC class I antigen processing and presenting machinery: Organization, function, and defects in tumor cells. *J. Natl. Cancer Inst.* **2013**, *105*, 1172–1187. [CrossRef]
100. Yewdell, J.W.; Reits, E.; Neefjes, J. Making sense of mass destruction: Quantitating MHC class I antigen presentation. *Nat. Rev. Immunol.* **2003**, *3*, 952–961. [CrossRef]
101. Melista, E.; Rigo, K.; Pasztor, A.; Christiansen, M.; Bertinetto, F.E.; Meintjes, P.; Hague, T. Towards a new gold standard—NGS corrections to sanger SBT genotyping results. *Hum. Immunol.* **2015**, *76*, 148. [CrossRef]
102. Choo, S.Y. The HLA system: Genetics, immunology, clinical testing, and clinical implications. *Yonsei Med. J.* **2007**, *48*, 11–23. [CrossRef] [PubMed]
103. Bauer, D.C.; Zadoorian, A.; Wilson, L.O.W.; Thorne, N.P. Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. *Brief. Bioinf.* **2018**, *19*, 179–187. [CrossRef] [PubMed]
104. Kiyotani, K.; Mai, T.H.; Nakamura, Y. Comparison of exome-based HLA class I genotyping tools: Identification of platform-specific genotyping errors. *J. Hum. Genet.* **2017**, *62*, 397–405. [CrossRef]
105. 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. Acquired cancer resistance to combination immunotherapy from transcriptional loss of class I HLA. *Nat. Commun.* **2018**, *9*, 3868. [CrossRef]
106. Paulson, K.G.; Tegeder, A.; Willmes, C.; Iyer, J.G.; Afanasiev, O.K.; Schrama, D.; Koba, S.; Thibodeau, R.; Nagase, K.; Simonson, W.T.; et al. Downregulation of MHC-I expression is prevalent but reversible in Merkel cell carcinoma. *Cancer Immunol. Res.* **2014**, *2*, 1071–1079. [CrossRef]
107. 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. Allele-Specific HLA Loss and immune escape in lung cancer evolution. *Cell* **2017**, *171*, 1259–1271.e1211. [CrossRef]
108. Jurtz, V.; Paul, S.; Andreatta, M.; Marcatili, P.; Peters, B.; Nielsen, M. NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **2017**, *199*, 3360–3368. [CrossRef]
109. O'Donnell, T.J.; Rubinsteyn, A.; Laserson, U. MHCflurry 2.0: Improved pan-allele prediction of MHC Class I-presented peptides by incorporating antigen processing. *Cell Syst.* **2020**, *11*, P42–P48.e7. [CrossRef]
110. O'Donnell, T.J.; Rubinsteyn, A.; Bonsack, M.; Riemer, A.B.; Laserson, U.; Hammerbacher, J. MHCflurry: Open-source class I MHC binding affinity prediction. *Cell Syst.* **2018**, *7*, 129–132.e124. [CrossRef]
111. Reynisson, B.; Barra, C.; Kaabinejadian, S.; Hildebrand, W.H.; Peters, B.; Nielsen, M. Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *J. Proteome Res.* **2020**, *19*, 2304–2315. [CrossRef] [PubMed]
112. Paul, S.; Croft, N.P.; Purcell, A.W.; Tscharke, D.C.; Sette, A.; Nielsen, M.; Peters, B. Benchmarking predictions of MHC class I restricted T cell epitopes. *bioRxiv* **2019**, 694539. [CrossRef]
113. Trolle, T.; Metushi, I.G.; Greenbaum, J.A.; Kim, Y.; Sidney, J.; Lund, O.; Sette, A.; Peters, B.; Nielsen, M. Automated benchmarking of peptide-MHC class I binding predictions. *Bioinformaticsatics (Oxf. Engl.)* **2015**, *31*, 2174–2181. [CrossRef] [PubMed]
114. Paul, S.; Croft, N.P.; Purcell, A.W.; Tscharke, D.C.; Sette, A.; Nielsen, M.; Peters, B. Benchmarking predictions of MHC class I restricted T cell epitopes in a comprehensively studied model system. *PLoS Comput. Biol.* **2020**, *16*, e1007757. [CrossRef]

115. Croft, N.P.; Smith, S.A.; Pickering, J.; Sidney, J.; Peters, B.; Faridi, P.; Witney, M.J.; Sebastian, P.; Flesch, I.E.A.; Heading, S.L.; et al. Most viral peptides displayed by class I MHC on infected cells are immunogenic. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 3112–3117. [CrossRef]
116. Sette, A.; Vitiello, A.; Reherman, B.; Fowler, P.; Nayersina, R.; Kast, W.M.; Melief, C.J.; Oseroff, C.; Yuan, L.; Ruppert, J.; et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J. Immunol. (Baltimore, Md. 1950)* **1994**, *153*, 5586–5592.
117. Bjerregaard, A.M.; Nielsen, M.; Jurtz, V.; Barra, C.M.; Hadrup, S.R.; Szallasi, Z.; Eklund, A.C. An Analysis of natural T cell responses to predicted tumor neoepitopes. *Front. Immunol.* **2017**, *8*, 1566. [CrossRef]
118. Castle, J.C.; Kreiter, S.; Diekmann, J.; Löwer, M.; van de Roemer, N.; de Graaf, J.; Selmi, A.; Diken, M.; Boegel, S.; Paret, C.; et al. Exploiting the mutanome for tumor vaccination. *Cancer Res.* **2012**, *72*, 1081. [CrossRef]
119. Bekri, S.; Uduman, M.; Gruenstein, D.; Mei, A.H.C.; Tung, K.; Rodney-Sandy, R.; Bogen, B.; Buell, J.; Stein, R.; Doherty, K.; et al. Neoantigen synthetic peptide vaccine for multiple myeloma elicits T cell immunity in a pre-clinical model. *Blood* **2017**, *130*, 1868. [CrossRef]
120. Kreiter, S.; Vormehr, M.; van de Roemer, N.; Diken, M.; Löwer, M.; Diekmann, J.; Boegel, S.; Schrörs, B.; Vascotto, F.; Castle, J.C.; et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* **2015**, *520*, 692–696. [CrossRef]
121. Borg, N.A.; Ely, L.K.; Beddoe, T.; Macdonald, W.A.; Reid, H.H.; Clements, C.S.; Purcell, A.W.; Kjer-Nielsen, L.; Miles, J.J.; Burrows, S.R.; et al. The CDR3 regions of an immunodominant T cell receptor dictate the 'energetic landscape' of peptide-MHC recognition. *Nat. Immunol.* **2005**, *6*, 171–180. [CrossRef] [PubMed]
122. Gras, S.; Saulquin, X.; Reiser, J.B.; Debeaupuis, E.; Echasserieau, K.; Kissenpfennig, A.; Legoux, F.; Chouquet, A.; Le Gorrec, M.; Machillot, P.; et al. Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. *J. Immunol. (Baltimore, Md. 1950)* **2009**, *183*, 430–437. [CrossRef] [PubMed]
123. Chen, G.; Yang, X.; Ko, A.; Sun, X.; Gao, M.; Zhang, Y.; Shi, A.; Mariuzza, R.A.; Weng, N.P. Sequence and structural analyses reveal distinct and highly diverse human cd8(+) tcr repertoires to immunodominant viral antigens. *Cell Rep.* **2017**, *19*, 569–583. [CrossRef] [PubMed]
124. Nivarthi, U.K.; Gras, S.; Kjer-Nielsen, L.; Berry, R.; Lucet, I.S.; Miles, J.J.; Tracy, S.L.; Purcell, A.W.; Bowden, D.S.; Hellard, M.; et al. An extensive antigenic footprint underpins immunodominant TCR adaptability against a hypervariable viral determinant. *J. Immunol. (Baltimore, Md. 1950)* **2014**, *193*, 5402–5413. [CrossRef]
125. 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. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **2017**, *547*, 89–93. [CrossRef]
126. Gielis, S.; Moris, P.; Neuter, N.D.; Bittremieux, W.; Ogunjimi, B.; Laukens, K.; Meysman, P. TCRex: A webtool for the prediction of T-cell receptor sequence epitope specificity. *bioRxiv* **2018**, 373472. [CrossRef]
127. 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.; et al. NetTCR: Sequence-based prediction of TCR binding to peptide-MHC complexes using convolutional neural networks. *bioRxiv* **2018**, 433706. [CrossRef]
128. Ogishi, M.; Yotsuyanagi, H. Quantitative prediction of the Landscape of T cell epitope immunogenicity in sequence space. *Front. Immunol.* **2019**, *10*. [CrossRef]
129. Springer, I.; Besser, H.; Tickotsky-Moskovitz, N.; Dvorkin, S.; Louzoun, Y. Prediction of specific TCR-peptide binding from large dictionaries of TCR-peptide pairs. *bioRxiv* **2020**, 650861. [CrossRef]
130. Bi, J.; Zheng, Y.; Yan, F.; Hou, S.; Li, C. Prediction of epitope-associated TCR by using network topological similarity based on deepwalk. *IEEE Access* **2019**, *7*, 151273–151281. [CrossRef]
131. Zhang, Y.; Lin, Z.; Wan, Y.; Cai, H.; Deng, L.; Li, R. The Immunogenicity and anti-tumor efficacy of a rationally designed neoantigen vaccine for B16F10 mouse melanoma. *Front. Immunol.* **2019**, *10*, 2472. [CrossRef] [PubMed]
132. Ni, Q.; Zhang, F.; Liu, Y.; Wang, Z.; Yu, G.; Liang, B.; Niu, G.; Su, T.; Zhu, G.; Lu, G.; et al. A bi-adjuvant nanovaccine that potentiates immunogenicity of neoantigen for combination immunotherapy of colorectal cancer. *Sci. Adv.* **2020**, *6*, eaaw6071. [CrossRef] [PubMed]
133. Gingold, H.; Pilpel, Y. Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* **2011**, *7*, 481. [CrossRef] [PubMed]

134. Wang, C.; Schmich, F.; Srivatsa, S.; Weidner, J.; Beerenwinkel, N.; Spang, A. Context-dependent deposition and regulation of mRNAs in P-bodies. *Elife* **2018**, *7*, e29815. [CrossRef] [PubMed]
135. Ingolia, N.T.; Ghaemmaghami, S.; Newman, J.R.S.; Weissman, J.S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (NY)* **2009**, *324*, 218. [CrossRef] [PubMed]
136. Zeng, C.; Fukunaga, T.; Hamada, M. Identification and analysis of ribosome-associated lncRNAs using ribosome profiling data. *BMC Genom.* **2018**, *19*, 414. [CrossRef] [PubMed]
137. Schubert, B.; Brachvogel, H.P.; Jürges, C.; Kohlbacher, O. EpiToolKit—A web-based workbench for vaccine design. *Bioinformatics (Oxf. Engl.)* **2015**, *31*, 2211–2213. [CrossRef]
138. Schubert, B.; Walzer, M.; Brachvogel, H.P.; Szolek, A.; Mohr, C.; Kohlbacher, O. FRED 2: An immunoinformatics framework for Python. *Bioinformatics (Oxf. Engl.)* **2016**, *32*, 2044–2046. [CrossRef]
139. Paul, S.; Sidney, J.; Sette, A.; Peters, B. TepiTool: A Pipeline for Computational Prediction of T Cell Epitope Candidates. *Curr. Protoc. Immunol.* **2016**, *114*, 18.19.11–18.19.24. [CrossRef]
140. Tang, S.; Madhavan, S. neoantigenR: An annotation based pipeline for tumor neoantigen identification from sequencing data. *bioRxiv* **2017**, 171843. [CrossRef]
141. Bais, P.; Namburi, S.; Gatti, D.M.; Zhang, X.; Chuang, J.H. CloudNeo: A cloud pipeline for identifying patient-specific tumor neoantigens. *Bioinformatics (Oxf. Engl.)* **2017**, *33*, 3110–3112. [CrossRef] [PubMed]
142. Bjerregaard, A.M.; Nielsen, M.; Hadrup, S.R.; Szallasi, Z.; Eklund, A.C. MuPeXI: Prediction of neo-epitopes from tumor sequencing data. *Cancer Immunol. Immunother.* **2017**, *66*, 1123–1130. [CrossRef] [PubMed]
143. Tappeiner, E.; Finotello, F.; Charoentong, P.; Mayer, C.; Rieder, D.; Trajanoski, Z. TIminer: NGS data mining pipeline for cancer immunology and immunotherapy. *Bioinformatics (Oxf. Engl.)* **2017**, *33*, 3140–3141. [CrossRef] [PubMed]
144. Zhou, Z.; Lyu, X.; Wu, J.; Yang, X.; Wu, S.; Zhou, J.; Gu, X.; Su, Z.; Chen, S. TSNAD: An integrated software for cancer somatic mutation and tumour-specific neoantigen detection. *R Soc. Open Sci.* **2017**, *4*, 170050. [CrossRef] [PubMed]
145. Chang, T.C.; Carter, R.A.; Li, Y.; Li, Y.; Wang, H.; Edmonson, M.N.; Chen, X.; Arnold, P.; Geiger, T.L.; Wu, G.; et al. The neoepitope landscape in pediatric cancers. *Genome Med.* **2017**, *9*, 78. [CrossRef]
146. Wang, T.Y.; Wang, L.; Alam, S.K.; Hoeppner, L.H.; Yang, R. ScanNeo: Identifying indel-derived neoantigens using RNA-Seq data. *Bioinformatics (Oxf. Engl.)* **2019**, *35*, 4159–4161. [CrossRef]
147. Wu, J.; Wang, W.; Zhang, J.; Zhou, B.; Zhao, W.; Su, Z.; Gu, X.; Wu, J.; Zhou, Z.; Chen, S. DeepHLApan: A deep learning approach for neoantigen prediction considering both HLA-peptide binding and immunogenicity. *Front. Immunol.* **2019**, *10*. [CrossRef]
148. Zhou, C.; Wei, Z.; Zhang, Z.; Zhang, B.; Zhu, C.; Chen, K.; Chuai, G.; Qu, S.; Xie, L.; Gao, Y.; et al. pTuneos: Prioritizing tumor neoantigens from next-generation sequencing data. *Genome Med.* **2019**, *11*, 67. [CrossRef]
149. Hundal, J.; Carreno, B.M.; Petti, A.A.; Linette, G.P.; Griffith, O.L.; Mardis, E.R.; Griffith, M. pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med.* **2016**, *8*, 11. [CrossRef]
150. Li, Y.; Wang, G.; Tan, X.; Ouyang, J.; Zhang, M.; Song, X.; Liu, Q.; Leng, Q.; Chen, L.; Xie, L. ProGeo-neo: A customized proteogenomic workflow for neoantigen prediction and selection. *BMC Med. Genom.* **2020**, *13*, 52. [CrossRef]
151. Coelho, A.C.M.F.; Fonseca, A.L.; Martins, D.L.; Lins, P.B.R.; da Cunha, L.M.; de Souza, S.J. neoANT-HILL: An integrated tool for identification of potential neoantigens. *BMC Med. Genom.* **2020**, *13*, 30. [CrossRef] [PubMed]
152. Wang, G.; Wan, H.; Jian, X.; Li, Y.; Ouyang, J.; Tan, X.; Zhao, Y.; Lin, Y.; Xie, L. 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**, *2020*, 5798356. [CrossRef] [PubMed]
153. Chen, R.; Fulton, K.M.; Twine, S.M.; Li, J. Identification of MHC Peptides Using Mass Spectrometry For Neoantigen Discovery And Cancer Vaccine Development. *Mass. Spectrum. Rev.* **2019**. [CrossRef] [PubMed]
154. Zhang, X.; Qi, Y.; Zhang, Q.; Liu, W. Application of mass spectrometry-based MHC immunopeptidome profiling in neoantigen identification for tumor immunotherapy. *Biomed. Pharmacother.* **2019**, *120*, 109542. [CrossRef]
155. Storkus, W.J.; Zeh, H.J., 3rd; Salter, R.D.; Lotze, M.T. Identification of T-cell epitopes: Rapid isolation of class I-presented peptides from viable cells by mild acid elution. *J. Immunother. Emphas. Tumor Immunol. Off. J. Soc. Biol. Ther.* **1993**, *14*, 94–103. [CrossRef]

156. Kote, S.; Pirog, A.; Bedran, G.; Alfaro, J.; Dapic, I. Mass Spectrometry-based identification of MHC-associated peptides. *Cancers* **2020**, *12*, 535. [CrossRef]
157. Vigneron, N.; Stroobant, V.; Ferrari, V.; Abi Habib, J.; Van den Eynde, B.J. Production of spliced peptides by the proteasome. *Mol. Immunol.* **2019**, *113*, 93–102. [CrossRef]
158. Liepe, J.; Sidney, J.; Lorenz, F.K.M.; Sette, A.; Mishto, M. Mapping the MHC class I–spliced immunopeptidome of cancer cells. *Cancer Immunol. Res.* **2019**, *7*, 62. [CrossRef]
159. Mylonas, R.; Beer, I.; Iseli, C.; Chong, C.; Pak, H.S.; Gfeller, D.; Coukos, G.; Xenarios, I.; Müller, M.; Bassani-Sternberg, M. Estimating the contribution of proteasomal spliced peptides to the HLA-I ligandome. *Mol. Cell Proteom.* **2018**, *17*, 2347. [CrossRef]
160. Solleder, M.; Guillaume, P.; Racle, J.; Michaux, J.; Pak, H.; Müller, M.; Coukos, G.; Bassani-Sternberg, M.; Gfeller, D. Mass spectrometry based immunopeptidomics leads to robust predictions of phosphorylated HLA class I ligands. *bioRxiv* **2019**, 836189. [CrossRef]
161. 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. Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nat. Biotechnol.* **2019**, *37*, 55–63. [CrossRef] [PubMed]
162. Antunes, D.A.; Abella, J.R.; Devaurs, D.; Rigo, M.M.; Kavraki, L.E. Structure-based methods for binding mode and binding affinity prediction for peptide-MHC complexes. *Curr. Top. Med. Chem.* **2018**, *18*, 2239–2255. [CrossRef] [PubMed]
163. Mohammed, F.; Stones, D.H.; Zarling, A.L.; Willcox, C.R.; Shabanowitz, J.; Cummings, K.L.; Hunt, D.F.; Cobbold, M.; Engelhard, V.H.; Willcox, B.E. The antigenic identity of human class I MHC phosphopeptides is critically dependent upon phosphorylation status. *Oncotarget* **2017**, *8*, 54160–54172. [CrossRef] [PubMed]
164. Durrant, L.G.; Metheringham, R.L.; Brentville, V.A. Autophagy, citrullination and cancer. *Autophagy* **2016**, *12*, 1055–1056. [CrossRef]
165. Galli-Stampino, L.; Meinjohanns, E.; Frische, K.; Meldal, M.; Jensen, T.; Werdelin, O.; Mouritsen, S. T-cell recognition of tumor-associated carbohydrates: The nature of the glycan moiety plays a decisive role in determining glycopeptide immunogenicity. *Cancer Res.* **1997**, *57*, 3214–3222.
166. Schueler-Furman, O.; Altuvia, Y.; Sette, A.; Margalit, H. Structure-based prediction of binding peptides to MHC class I molecules: Application to a broad range of MHC alleles. *Protein Sci.* **2000**, *9*, 1838–1846. [CrossRef]
167. Bui, H.H.; Schiewe, A.J.; von Grafenstein, H.; Haworth, I.S. Structural prediction of peptides binding to MHC class I molecules. *Proteins* **2006**, *63*, 43–52. [CrossRef]
168. Yanover, C.; Bradley, P. Large-scale characterization of peptide-MHC binding landscapes with structural simulations. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6981. [CrossRef]
169. Mukherjee, S.; Bhattacharyya, C.; Chandra, N. HLaffy: Estimating peptide affinities for Class-1 HLA molecules by learning position-specific pair potentials. *Bioinformatics (Oxf. Engl.)* **2016**, *32*, 2297–2305. [CrossRef]
170. Ochoa-Garay, J.; McKinney, D.M.; Kochounian, H.H.; McMillan, M. The ability of peptides to induce cytotoxic T cells in vitro does not strongly correlate with their affinity for the H-2Ld molecule: Implications for vaccine design and immunotherapy. *Mol. Immunol.* **1997**, *34*, 273–281. [CrossRef]
171. Feltkamp, M.C.; Vierboom, M.P.; Kast, W.M.; Melief, C.J. Efficient MHC class I-peptide binding is required but does not ensure MHC class I-restricted immunogenicity. *Mol. Immunol.* **1994**, *31*, 1391–1401. [CrossRef]
172. Calis, J.J.; Maybeno, M.; Greenbaum, J.A.; Weiskopf, D.; De Silva, A.D.; Sette, A.; Kesmir, C.; Peters, B. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput. Biol.* **2013**, *9*, e1003266. [CrossRef] [PubMed]
173. Chowell, D.; Krishna, S.; Becker, P.D.; Cocita, C.; Shu, J.; Tan, X.; Greenberg, P.D.; Klavinskis, L.S.; Blattman, J.N.; Anderson, K.S. TCR contact residue hydrophobicity is a hallmark of immunogenic CD8+ T cell epitopes. *Proc. Nat. Acad. Sci. USA* **2015**, *112*, E1754–E1762. [CrossRef] [PubMed]
174. Tung, C.W.; Ziehm, M.; Kämper, A.; Kohlbacher, O.; Ho, S.Y. POPISK: T-cell reactivity prediction using support vector machines and string kernels. *BMC Bioinf.* **2011**, *12*, 446. [CrossRef]
175. Trolle, T.; Nielsen, M. NetTepi: An integrated method for the prediction of T cell epitopes. *Immunogenetics* **2014**, *66*, 449–456. [CrossRef]
176. Lanzarotti, E.; Marcatili, P.; Nielsen, M. Identification of the cognate peptide-MHC target of T cell receptors using molecular modeling and force field scoring. *Mol. Immunol.* **2018**, *94*, 91–97. [CrossRef]

177. Pierce, B.G.; Weng, Z. A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes. *Protein Sci.* **2013**, *22*, 35–46. [CrossRef]
178. Vita, R.; Overton, J.A.; Greenbaum, J.A.; Ponomarenko, J.; Clark, J.D.; Cantrell, J.R.; Wheeler, D.K.; Gabbard, J.L.; Hix, D.; Sette, A.; et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **2014**, *43*, D405–D412. [CrossRef]
179. Vita, R.; Mahajan, S.; Overton, J.A.; Dhanda, S.K.; Martini, S.; Cantrell, J.R.; Wheeler, D.K.; Sette, A.; Peters, B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **2019**, *47*, D339–D343. [CrossRef]
180. Dhanda, S.K.; Mahajan, S.; Paul, S.; Yan, Z.; Kim, H.; Jespersen, M.C.; Jurtz, V.; Andreatta, M.; Greenbaum, J.A.; Marcatili, P.; et al. IEDB-AR: Immune epitope database—Analysis resource in 2019. *Nucleic Acids Res.* **2019**, *47*, W502–W506. [CrossRef]
181. Olsen, L.R.; Tongchusak, S.; Lin, H.; Reinherz, E.L.; Brusic, V.; Zhang, G.L. TANTIGEN: A comprehensive database of tumor T cell antigens. *Cancer Immunol. Immunother.* **2017**, *66*, 731–735. [CrossRef] [PubMed]
182. Wu, J.; Zhao, W.; Zhou, B.; Su, Z.; Gu, X.; Zhou, Z.; Chen, S. TSNAdb: A Database for Tumor-specific neoantigens from immunogenomics data analysis. *Genom. Proteom. Bioinform.* **2018**, *16*, 276–282. [CrossRef] [PubMed]
183. Zhou, W.J.; Qu, Z.; Song, C.Y.; Sun, Y.; Lai, A.L.; Luo, M.Y.; Ying, Y.Z.; Meng, H.; Liang, Z.; He, Y.J.; et al. NeoPeptide: An immunoinformatic database of T-cell-defined neoantigens. *Database* **2019**, *2019*. [CrossRef] [PubMed]
184. Tan, X.; Li, D.; Huang, P.; Jian, X.; Wan, H.; Wang, G.; Li, Y.; Ouyang, J.; Lin, Y.; Xie, L. dbPepNeo: A manually curated database for human tumor neoantigen peptides. *Database* **2020**, *2020*. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).