BioMedInformatics

MDPI

Review

Understanding and Therapeutic Application of Immune Response in Major Histocompatibility Complex (MHC) Diversity Using Multimodal Artificial Intelligence

Yasunari Matsuzaka [1,2,*] and Ryu Yashiro [2,3]

[1] Division of Molecular and Medical Genetics, Center for Gene and Cell Therapy, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan
[2] Administrative Section of Radiation Protection, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira 187-8551, Japan; ryuy@niid.go.jp
[3] Department of Mycobacteriology, Leprosy Research Center, National Institute of Infectious Diseases, Tokyo 162-8640, Japan
* Correspondence: yasunari80808@ims.u-tokyo.ac.jp; Tel.: +81-3-5449-5372

Abstract: Human Leukocyte Antigen (HLA) is like a device that monitors the internal environment of the body. T lymphocytes immediately recognize the HLA molecules that are expressed on the surface of the cells of the different individual, attacking it defeats microorganisms that is one of the causes of rejection in organ transplants performed between people with unmatched HLA types. Over 2850 and 3580 different polymorphisms have been reported for HLA-A and HLA-B respectively, around the world. HLA genes are associated with the risk of developing a variety of diseases, including autoimmune diseases, and play an important role in pathological conditions. By using a deep learning method called multi-task learning to simultaneously predict the gene sequences of multiple HLA genes, it is possible to improve accuracy and shorten execution time. Some new systems use a model called convolutional neural network (CNNs) in deep learning, which uses neural networks consisting of many layers and can learn complex correlations between SNP information and HLA gene sequences based on reference data for HLA imputation, which serves as training data. The learned model can output predicted values of HLA gene sequences with high accuracy using SNP information as input. To investigate which part of the input information surrounding the HLA gene is used to make learning predictions, predictions were made using not only a small number of nearby SNP information but also many SNP information distributed over a wider area by visualizing the learning information of the model. While conventional methods are strong at learning using nearly SNP information and not good at learning using SNP information located at distant locations, some new systems are thought that prediction accuracy may have improved because this problem was overcome. HLA genes are involved in the onset of a variety of diseases and are attracting attention. As an important area from the perspective of elucidating pathological conditions and realizing personalized medicine. The applied multi-task learning to two different HLA imputation reference panels—a Japanese panel (n = 1118) and type I diabetes genetics consortium panel (n = 5122). Through 10-fold cross-validation on these panels, the multi-task learning achieved higher imputation accuracy than conventional methods, especially for imputing low-frequency and rare HLA alleles. The increased prediction accuracy of HLA gene sequences is expected to increase the reliability of HLA analysis, including integrated analysis between different racial populations, and is expected to greatly contribute to the identification of HLA gene sequences associated with diseases and further elucidation of pathological conditions.

Keywords: convolutional neural network (CNNs); cytotoxic T cells (CTLs); human leukocyte antigen (HLA); long short-term memory (LSTM); major histocompatibility complex (MHC); masked language modeling (MLM); natural language processing (NLP); tumor associated antigens (TAAs)

Citation: Matsuzaka, Y.; Yashiro, R. Understanding and Therapeutic Application of Immune Response in Major Histocompatibility Complex (MHC) Diversity Using Multimodal Artificial Intelligence. *BioMedInformatics* **2024**, *4*, 1835–1864. https://doi.org/10.3390/biomedinformatics4030101

Academic Editor: Rosalba Giugno

Received: 21 December 2023
Revised: 29 May 2024
Accepted: 1 August 2024
Published: 5 August 2024

1. Introduction

Human Leukocyte Antigen (HLA) are proteins expressed on the surface of cells. When foreign substances such as microorganisms invade inside or outside the cell, it has the function of notifying T lymphocytes, which are one of the components of white blood cells (WBCs) contained in the blood, of the presence of foreign substances and instructing them to eliminate them by expressing on the cell surface to combine the fragments created by the decomposition of foreign substances [1]. Thus, HLA is like a device that monitors the internal environment of the body, and T lymphocytes are constantly monitoring whether there are any strange foreign particles bound to HLA molecules, by which protects our bodies from attack by foreign substances such as microorganisms. When cells from a different individual invade our body, T lymphocytes immediately recognize the HLA molecules that are expressed on the surface of the cells of the different individual, attacking in the same way it defeats microorganisms that is one of the causes of rejection in organ transplants performed between people with unmatched HLA types [2].

Regarding HLA, which is human major histocompatibility complex (MHC), HLA antigens, which are human MHC antigens, were discovered in 1952 by Dausset in France, who discovered antibodies that reacted with leukocyte agglutination tests in the serum of patients with frequent blood transfusions, that began with the naming of the corresponding antigen Mac antigen, currently HLA-A2 antigen. However, more than half a century after its discovery, HLA is only found in WBCs, but is distributed in almost all cells and body fluids. Then, it has been revealed that it functions as a histocompatibility antigen, an important molecular marker involved in human immunity that the immune system to distinguish between "self" and "non-self". The more HLA mismatches between the patients undergoing allogeneic hematopoietic stem cell transplantation and the donor, the higher the incidence of adverse complications, such as engraftment failure and graft-versus-host disease (GVHD) [3]. There are six types of HLA proteins, HLA-A, HLA-B, and HLA-C, which are class I molecules presented as transmembrane glycoproteins on the surface of all nucleated cells and HLA-DR, HLA-DQ, and HLA-DP, which are class II molecules usually expressed only by professional antigen-presenting cells, including B cells, macrophages, dendritic cells, Langerhans cells, thymic epithelium, and activated but not quiescent T cells due to their different functions and structures (Figure 1). Additionally, each of these molecules has structural variations called "polymorphisms". For example, over 2850 and 3580 different polymorphisms have been reported for HLA-A and HLA-B respectively, around the world [4]. A test that examines the types of polymorphisms and determines the combination of HLA molecules possessed by an individual is called "HLA typing". There are various methods for HLA typing, but recently DNA typing is commonly used to estimate the gene sequence that determines polymorphisms in HLA molecules [5]. Donor selection for hematopoietic cell transplantation is based on the degree of concordance of HLA-A, -B, -C, -DR molecule polymorphisms inferred from HLA typing results [6,7].

The original class I molecules consists of an alpha heavy chain attached to a beta2 microglobulin molecules, which as two peptide-binding domains, one immunoglobulin (Ig)-like domain, and one transmembrane region with a cytoplasmic tail. The heavy chains of class I molecules are encoded by genes at the HLA-A, HLA-B, and HLA-C loci. T cells that express CD8 molecules respond to MHC class I molecules. These lymphocytes often have cytoplasmic functions and must be able to recognize any infected cells. Because all nucleated cells express MHC class I molecules, all infected cells act as antigen-presenting cells to CD8-positive T cells that CD8 binds to the non-polymorphic portion of the heavy chains of class I molecules. Some MHC class I genes directs peptides to specific receptors on HLA-G, which play a role in protecting the fetus from maternal immune responses, and HLA-E, which presents peptides to specific receptors on NK cells, encodes a non-classical MHC molecule [8,9]. MHC class II molecules, on the other hand, consist of wo polypeptide (alpha and beta) chains; each chains have a peptide-binding domain, an Ig-like domain, and a transmembrane region with a cytoplasmic tail. Both polypeptides chains are encoded by genes in the HLA-DP, HLA-DQ, or HLA-DR regions of chromosome 6 (Figure 1). Most

nucleated cells can be induced to express MHC class II molecules by interferon (IFN)-γ [10]. T cells that respond to class II molecules express CD4 and are often helper cells. Additionally, the MHC class III region of the genome encodes several molecules important in inflammation [11]. These include the complement (C2, C4, and factor B), tumor necrosis factor (TNF) alpha, lymphotoxin, and three heat shock proteins.

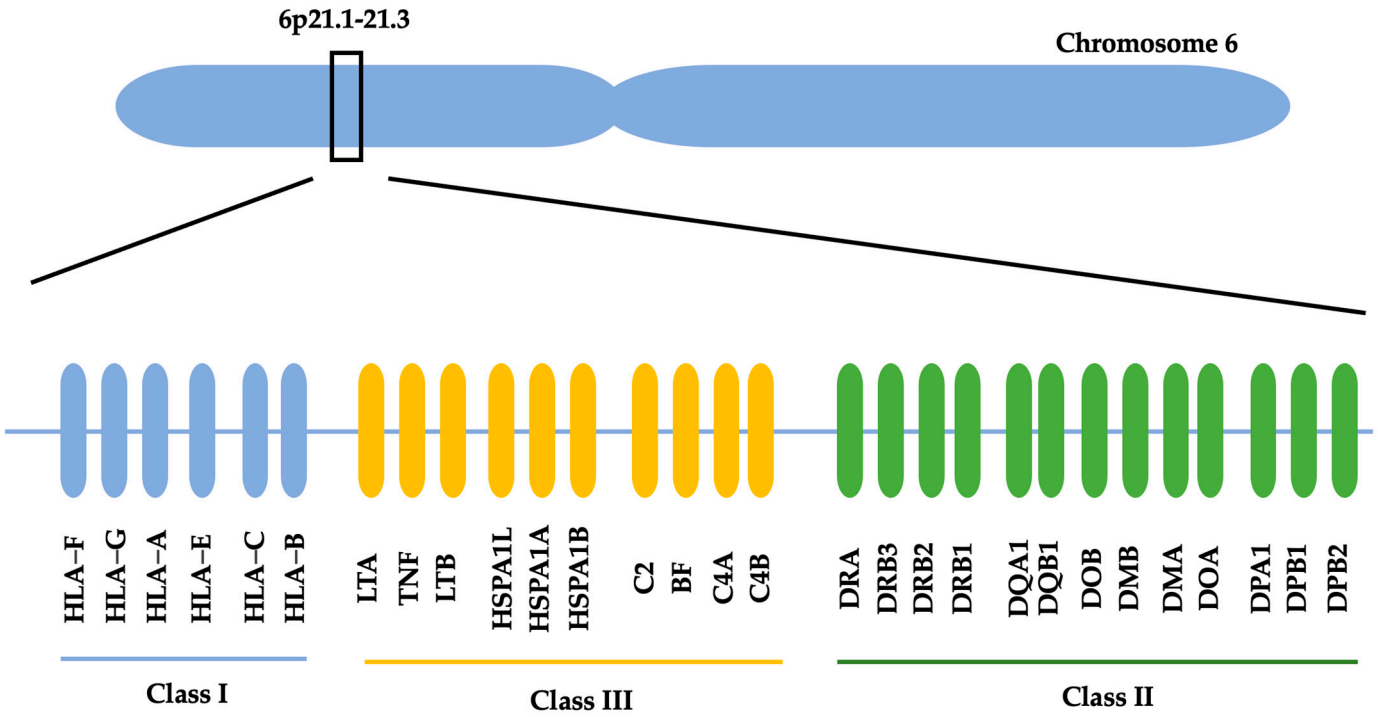

Figure 1. Genomic map of the human major histocompatibility complex.

Additionally, by changing specific parts of the foreign body, the immune response of T lymphocytes to the foreign body can be significantly enhanced or suppressed, suggesting that this can be applied to further enhance the T lymphocyte response, immune response to a specific foreign substance when it is useful for the body [12,13]. For example, it could be applied to the development of powerful vaccines to increase resistance, immunity against pathogenic bacteria and viruses. Some diseases, such as rheumatism, develop when T lymphocytes react to components of the body that T lymphocytes should not normally react to [14]. In this case, it is possible to suppress the immune response of T lymphocytes that react to self by using a self-imitation product that has slightly changed the structure of self-components that T lymphocytes mistake for foreign substances [15]. In this review, we summarized the advance in prediction models of sequence validation of HLA using deep learning and their clinical application.

2. Major Histocompatibility Complex (MHC)

Research on the MHC began in 1936 when Gorer in England and Snell in the US discovered it as an antigen related to the blood type of mice. They found that the genetic region governing this antigen strongly influences the rejection response to skin grafts and name its Histocompatibility-2 antigen (H-2 antigen). Subsequently, analysis using inbred mice revealed that the gene region controlling the H-2 antigen is a gene complex consisting of multiple regions located on chromosome 17. In the early 1960s, Benacerraf, McDevitt and colleagues revealed that the MHC region contains a gene region that controls the mouse immune response and named this region the I region. T cells contacts MHC molecules with the T cell receptor (TCR) as well as co-receptors CD8 and CD4. Cytotoxic T cells (CTLs) express CD8 on their surface and bind to MHC class I molecules. CTLs release perforin and

granzyme by recognizing target cells such as virus-infected cells, directly induce apoptosis and damage them [16,17]. Helper T cells (Th cells) express CD4 on their surface and bind to cells that express MHC class II molecules on their surface. Th1 cells produce IL-2 and IFN-γ, which activate CTLs, macrophages, and NK cells. Th2 cells produce cytokines such as IL-4 and activate B cells. Th17 cells produce IL-17 and are involved in inflammation [18]. Regulatory T cells (Treg) produce TGF-beta and IL-10 and have the function of actively terminating an immune response once it has started.

2.1. Vaccine and MHC

Thus, relationship between MHC region and immune response are critical. In cancer cells, cancer-specific proteins are broken down into peptides and presented as cancer antigen peptides along with MHC class I molecules on the cancer cell surface [19]. Cancer antigen specific CTLs that recognize this are activated and attack cancer cells [20–22]. However, if the number and strength of CTLs, immunity are insufficient, cancer will continue to proliferate and metastasize, becoming malignant [23]. Cancer vaccine therapy is a method of treating cancer by administering antigenic peptides into the patient's blood vessels to strongly induce cancer specific CTLs [24–26]. To further enhance the effectiveness of cancer vaccine therapy, research is underway on ways to use antigen-presenting cells such as dendritic cells that present cancer antigen peptides, and on combinations with gene therapy. To further enhance the effectiveness of cancer vaccine therapy, research is underway on ways to use antigen-presenting cells (APCs) such as dendritic cells that present cancer antigen peptides, and on combinations with gene therapy. Among these, cancer vaccine therapy is a treatment method that is expected to exhibit cancer cell-specific antitumor effects by directly inoculating peptides and patients derived from tumor associated antigens (TAAs), which are highly expressed only in cancer cells and rarely expressed in normal cells, or mRNAs and genes derived from TAAs, and DCs loaded with or introduced with these to cancer patients [27]. In applying tumor-specific cancer immune responses to cancer immunotherapy, it is important to first identify the target TAA [28]. Various TAAs, such as cancer testis antigens and carcinoembryonic antigens, have been identified through comprehensive cDNA microarray analysis of cancer and normal tissues, and have been applied to cancer immunotherapy. These TAAs are released into the tissue fluid when some cancer cells die due to apoptosis, etc. [22].

A long peptide (LP) consisting of 10 to 20 amino acids that is taken up and degraded by DC endosome binds to HLA class II molecules and enters the cell [29,30]. It is expressed on the surface and is recognized by the T cell receptor (TCR) of $CD4^+$ Th cells, which activates the Th cells [31,32]. At the same time, DCs transport TAA to the cytoplasm, where it is degrades by proteasomes, resulting in a short peptide (SP) consisting of 9 to 11 amino acids, which is bound to HLA class I molecules and attached to the cell surface, which is recognized by the TCR of $CD8^+$ CTLs, cross-presentation pathway, and the CTLs are activated [33]. DC express the costimulatory molecules CD80/86, and when this binds to CD28 of naïve T cells, T cells are strongly activated and differentiate into effector T cells. Effector T cells do not require costimulatory molecules; they recognize HLA-TAA peptide complexes expressed on the surface of cancer cells and exhibit an immune response against cancer cells. $CD4^+$ Th1 cells produce Th1 cytokines, which promote CTL induction and activation. Thus, cancer antigen vaccine therapy is a treatment method that actively immunizes cancer patients with TAAs from the outside to induce and enhance antitumor immunity, with the aim of inducing T cell immune responses against TAAs [25,26].

2.2. SP Vaccine Therapy

Thus, Th1 cytokines promote CTL induction and activation. CTLs are very important effector cells that kill cancer cells [21,34]. When TAA-derived SP is administered to cancer patients, it binds to HLA-I on DCs and is presented on the cell surface [35–37]. By inducing and activating peptide-specific CTLs and attacking cancer cells that express the same TAA-SP recognized by these CTLs, it is expected that a cancer-specific antitumor effect will

occur [38,39]. Methods for identifying new TAAs include the Serex method and cDNA microarray analysis [40]. Furthermore, using algorithm analysis and HLA-expressing transgenic mice based on the amino acid sequence of the newly identified TAA, many SPs that induce HLA-I-restricted CTLs have been identified and have been applied to cancer vaccine therapy [25,41,42]. Furthermore, using proteome analysis techniques such as mass spectrometry, attempts have been made to identify TAAs and TAA-derived T cell epitopes by determining the amino acid sequences of peptides bound to HLA molecules on the surfaces of cancer cells and antigen-presenting cells [43,44]. The advantage of identifying TAA and T cell epitopes using such proteome analysis technology is that TAA peptides presented by HLA expressed on the surface of cancer tissues can be directly identified. Using methods such as proteosome analysis such as mass spectrometry and genetic analysis, multiple TAA-SPs that bind to HLA-I expressed in renal cell carcinoma have been identified and are being applied to cancer peptide vaccine therapy [45–48]. Clinical research on TAA peptide vaccine therapy using SP has been conducted for a variety of cancers, including esophageal cancer, head and neck cancer, lung cancer, and bladder cancer [25,41,49–51]. Furthermore, no serious adverse events were observed in either case, and tumor shrinkage and survival effects have been observed in some patients with advanced cancer [52,53]. SPs recognized by HLA-A24-restricted CTLs for novel cancer testis antigens LY6K, CDCA1, and IMP3, which are frequently expressed in esophageal, oral cavity, and lung cancers, identified by genome-wide cDNA. Furthermore, in vaccinated patients, an increase in overall survival was observed in direct proportion to the number of cancer antigen peptides to which CTLs responded [50]. However, monotherapy with cancer antigen SP peptide vaccines for patients with advanced cancer has a low response rate of less than 5% in most cases, and satisfactory has not yet been demonstrated [25,49,51].

2.3. LP Vaccine Therapy

Thus, monotherapy with cancer antigen SP peptide vaccines for patients with advanced cancer has a low response rate. To induce a stranger antitumor immune response, the presence of TAA-specific Th1 cells as well as is important [54,55]. Furthermore, the TAA-SP mentioned above induces unresponsiveness in CTLs by binding to HLA-I in somatic cells other than antigen-presenting cells that do not express costimulatory molecules such as CD80/86. In addition, in immunization with TAA-SP and incomplete Freund's adjuvant (IFA), CTLs accumulate at the immunized area for a long time and do not migrate to the tumor area [25,56]. On the other hand, LP cannot directly bind to HLA-I, and after being taken up by antigen-presenting cells and processed within the cells, it is presented by HLA-II and not only induces Th cells, but also induce Th cells [57,58]. SP is produced through the cross-presentation pathway, which can be presented by HLA-I and induce TAA-specific CTL [51,59]. Therefore, LP is useful not only in inducing stronger antitumor effects but also in preventing the induction of unresponsiveness [60,61]. Clinical trials using LP are currently being conducted on a variety of cancers, and some have even reported survival benefits. Furthermore, it was confirmed that the identified LP-specific Th cells exist in the peripheral blood of cancer patients. Furthermore, because of cDNA analysis of TCR genes expressed in LP-specific Th cell clones in DEPDC1 and MPHOSPH1, which are cancer antigens that are frequently and highly expressed in bladder cancer, both bulk Th cell lines and clones expressed a single TCR alpha and beta chain gene, in which this pair was responsible for cancer antigen specificity and HLA restriction. By administering such LPs as a vaccine to cancer patients, it is expected that TAA peptide vaccine therapy with even stronger antitumor effects will be developed [25,26,41]. Furthermore, the TCR gene can be used for monitoring Th cells induced by LP vaccine therapy and for adoptive immunotherapy for cancer using peripheral blood T cells that have been forced to express the TCR gene [62–66]. When applying such peptide vaccine therapy clinically, there are some points to be considered regarding HLA-I, which presents SP contained in LP, that is a loss of heterozygosity (LOH) of the HLA-I gene in tumor [67–69]. Tumor cells expressing HLA-I and TAA are eliminated by CTL induced by peptide vaccine administration [70].

However, tumor cells that have undergone LOH escape attack by CTL and proliferate and metastasize [71,72]. To avoid this phenomenon, first, select an LP containing an SP that can induce CTLs restricted to HLA, such as HLA-A24, which has Bw4, a ligand for KIR, an inhibitory receptor of NK cells. Thus, even if HLA-A24 is deleted due to LOH, tumor cells can be eliminated by NK cells [73,74]. Furthermore, due to this mechanism, LOH of HLA-A24 in tumor cells is less likely to occur. Based on the same idea, the target patients are those who are HLA-A24 homozygote or whose opposite haplotype is Bw4 negative, or HLA-A2 homozygote; patients with B*40:02 are excluded. Furthermore, to prevent tumor escape from CTL due to LOH, LP containing multiple HLA-restricted CTL epitopes SP is utilized [75].

2.4. Cancer Antigen Vaccine Therapy Using DC and Neoantigen

Thus, tumor cells that have undergone LOH escape attack by CTL. Also, DC vaccine therapy is a treatment method that utilizes the powerful antigen-presenting ability of DCs [76,77]. There are methods such as loading DC with TAA peptide or cancer cell lysate and introducing mRNA or DNA encoding TAA into DC [78–80]. By inoculating cancer patients with these DCs as a vaccine, cancer cell-specific T cells are induced and activated. Immature DCs are induced by culturing monocytes collected from peripheral blood in the presence of GM-CSF and IL-4, and these immature DCs are loaded with TAA-derived peptides and mRNA or DNA encoding TAA [81,82]. A common method is to further stimulate DCs to mature and activate them, inducing them to have strong antigen-presenting ability, and then using them as a vaccine [81,83]. Cancer develops through the accumulation of genetic mutations in multiple stages, and therefore has many somatic mutations. As a result, when a peptide containing an amino acid mutation is presented by HLA-I, it is expected to be recognized as a "non-self-antigen". [57,84,85] Cancer antigen vaccine therapy using mutated peptides derived from antigens, neoantigens encoded by somatic missense mutant genes that occur in cancer cells has attracted attention [69,86,87]. An advantage of applying neoantigens to cancer vaccine therapy is that some T cells that exhibit high affinity for TAA may be eliminated from the body through negative selection in the thymus [26,69,88]. Because these T cells have not undergone selection in the thymus, they are less susceptible to immune tolerance, and specific T cells are more likely to be efficiently induced [89]. Comprehensive genetic analysis of cancer cells, mainly malignant melanoma, has shown that mutated peptides caused by somatic mutations in the genes of cancer cells have an immunogenic effect on CTL and Th1 cells within the tumor [69,90–92]. This immune response plays an important role in antitumor immunity. Mutations that produce neoantigens are divided into driver mutations and passenger mutations [93,94]. Driver mutations are genetic mutations that promote the transformation of malignant tumor cells, and passenger mutations are random mutations that are not directly involved in cancer formation [90,93,95]. Most genetic mutations that supply neoantigens originate from passenger mutations rather than driver mutations. In addition, neoantigens generated by driver mutations are expressed from the early stage of tumor development, and T cells undergo epigenetic changes due to long-term exposure to neoantigens, making them unresponsive [88]. Passenger mutations differ among individual patients even when tumors arise from the same organ, and they are numerous and highly immunogenic, making them suitable targets for cancer vaccine therapy using neoantigens peptides. Molecular targeted therapy is the best cancer treatment that targets driver mutations [96,97]. Comprehensive exon and RNA sequence analysis of relatively early-stage lung cancer tissues revealed that LOH is frequently detected in patients with many neoantigen peptides, and immune escape is more likely to occur [69,98–100]. HLA-I allele of cancer patients constrain the types of mutations in driver genes that play a direct and important role in the development and progression of cancer cells [101–104]. In patients with specific HLA alleles, neoantigen peptides generated by these driver mutations are presented to T cells by the HLA and inducer tumor immunity, thereby eliminating tumor cells [105–109].

Antibody therapy that lifts the immunosuppressive state of T cells characteristic of cancer patients has shown remarkable results by inhibiting immune checkpoint molecules such as anti-CTLA-4, CTL antigen 4 antibodies and anti-PD-1, programmed cell death-1/PD-1 antibodies [110]. Neoantigen-specific T cells already exist in the tumor tissue before treatment, and these T cells are reactivated by anti-CTLA-4 antibody therapy or anti-PD-1 antibody therapy and play an important role in antitumor effects [111–114]. In addition, in patients with a good response to immune checkpoint inhibition therapy, more nonsynonymous substitution mutant genes with amino acid substitutions are observed, and neoantigen-specific T cells are detected in tumor tissue before treatment [69,111,114]. These results suggest that enhancement of the immune response to neoantigens may be the mechanism by which immune checkpoint blockade therapy exerts its effectiveness [69,115,116]. It is expected that the combination of cancer antigen vaccines containing neoantigens and immune checkpoint inhibition point inhibition therapy will produce strong antitumor effects [26,117]. An analysis of many patients with advanced cancer treated with anti-PD-1 antibodies or anti-CTLA-4 antibodies showed that HLA-B44 supertype-positive, a group of HLA antigens that have common motifs in the amino acids sequence of their binding peptides, based on the common attributes of the pocket in the HLA peptide-accommodating groove and HLA heterozygous patients show a good response, while HLA-B62 supertype-positive patients, HLA homozygous patients, and HLA LOH patients show a poor response [118]. Thus, there is a correlation between the effectiveness of immune checkpoint inhibitor therapy and the type and number of HLA alleles [119,120]. This phenomenon also indicates that T cells that recognize neoantigen peptides presented by HLA are responsible for the immune response that destroys tumor cells [69,106,121].

As clinical trials progress, cancer antigen vaccine therapy has few side effects than conventional cancer treatments and is effective in prolonging survival while maintain patients' quality of life. However, the clinical effect of cancer antigen vaccine therapy alone on advanced cancer remains weak, but there are great expectations for vaccine therapy that targets multiple neoantigens [42,69,122,123]. Furthermore, the development of more potent cancer antigen will require more complex cancer immunotherapy that combine treatments that control immunosuppressive mechanisms in the tumor microenvironment [124]. Additionally, advances in genetic analysis technology have made it possible to evaluate the cancer and immune status of individual cancer patients using biomarkers and other methods [125–127]. Future efforts will also be focused on developing personalized cancer immunotherapy based on this information.

3. Computer-Based Inference of HLA Gene Sequences Related to Immune Function

Thus, there are great expectations for vaccine therapy that targets multiple neoantigens. WBC, one of the blood cells contained in human blood, are cells responsible for the immune response that recognize self and non-self and attacks foreign antigens. Like blood types for red blood cells (RBCs), blood types for WBCs also exist, and are determined by individual differences in the genomic sequence of the HLA gene. HLA genes are associated with the risk of developing a variety of diseases, including autoimmune diseases, and play an important role in pathological conditions [128]. Since HLA gene sequences are complex and require high costs to decipher, the HLA gene sequence is comprehensively predicted and analyzed on a computer from information on multiple single nucleotide polymorphisms (SNPs) located around the genome sequence [129]. HLA gene sequences associated with the risk of developing various diseases have been elucidated, but the reported risk sequences often differ between races [130]. For example, in type 1 diabetes in Westerners, amino acid other than aspartic acid at position 57 of the HLA-DQB1 molecule was known to be the most strongly associated risk sequence, but a similar association was not observed in Japanese. It is expected that further knowledge will be obtained by performing integrated analysis among different ethnic groups, but the prediction accuracy for low-frequency HLA gene sequences decreases significantly with conventional HLA imputation methods [130,131].

There was a problem in that it was difficult to obtain highly reliable results in integrated analysis between groups that required sequences.

HLA gene complex plays a crucial role in the development of autoimmune diseases. Certain HLA alleles have been identified as risk factors for various autoimmune conditions. Some of the key HLA gene sequences associated with an increased risk of developing autoimmune diseases are as follows; in rheumatoid arthritis (RA), HLA-DRB1*04:01, HLA-DRB1*04:04, and HLA-DRB1*04:05 (shared epitope alleles) [132], in type 1 diabetes mellitus (DM),HLA-DRB1*03:01, HLA-DRB1*04:01, and HLA-DRB1*04:05 and HLA-DQB1*02:01 and HLA-DQB1*03:02 [133], in multiple sclerosis, HLA-DRB1*15:01 and HLA-A*02:01 (protective allele) [134], in systemic lupus erythematosus (SLE), HLA-DRB1*03:01 and HLA-DRB1*15:01 [135,136], in ankylosing spondylitis, HLA-B*27 [137], in celiac disease, HLA-DQA1*05 and HLA-DQB1*02 (encoding for DQ2.5 heterodimer) and HLA-DQA1*03 and HLA-DQB1*03:02 (encoding for DQ8 heterodimer) [138], in psoriasis, HLA-C*06:02 [139], and in inflammatory bowel diseases, HLA-DRB1*01:03 and HLA-DRB1*07:01 (Crohn's disease) and HLA-DRB1*01:02 and HLA-DRB1*15:01 (ulcerative disease) [140–142]. It's important to note that the presence of these HLA alleles does not necessarily mean an individual will develop an autoimmune disease, as other genetic and environmental factors also contribute to the risk.

Additionally, the specific HLA associations may vary across different populations and ethnic groups. For instance, in Caucasian populations, certain HLA alleles like HLA-DRB1*03:01, HLA-DRB1*04:01, and HLA-DRB1*04:04 are strongly linked to an increased risk of RA [143,144]. In contrast, studies in Arab populations have reported inconsistent associations, with HLA-DRB1*04:01 and HLA-DRB1*10 being implicated in some cases [145–147]. Similarly, for type 1 DM, HLA-DRB1*03:01 and HLA-DRB1*04:01 alleles confer high risk in Caucasians, while HLA-DRB1*04:05 and HLA-DRB1*04:06 are associated with increased risk in Arab populations [148]. In SLE, HLA-DRB1*03:01 and HLA-DRB1*15:01 are risk factors in Caucasians, but HLA-DRB1*16:02 and HLA-DRB1*15:01 are implicated in Arab SLE patients [149]. These differences likely arise from the unique genetic backgrounds and evolutionary histories of different ethnic groups, leading to variations in the HLA allele frequencies and linkage disequilibrium patterns. Additionally, environmental factors and gene-environment interactions may also contribute to the observed ethnic disparities in HLA associations with autoimmune diseases.

In recently, artificial intelligence (AI) technologies, including machine learning methods, have been attracting attention to solve problems that are difficult to solve with conventional analysis methods. Deep learning is a machine learning method that uses multilayer neural networks. This technology has been attracting attention since the mid-2010s, mainly because it shows predictive performance that far exceeds existing methods in fields such as image recognition and natural language processing [150,151]. In deep learning models used in protein engineering, the input is often the amino acid sequence of the protein, and the output is the functional value doe prediction [152–154]. It has the characteristic of being able to learn complex features that cannot be captured by general statistical and machine learning analysis methods. Some new systems use a model called convolutional neural network (CNNs) in deep learning, which uses neural networks consisting of many layers and can learn complex correlations between SNP information and HLA gene sequences based on reference data for HLA imputation, which serves as training data [155]. The learned model can output predicted values of HLA gene sequences with high accuracy using SNP information as input [156].

By using a deep learning method called multi-task learning to simultaneously predict the gene sequences of multiple HLA genes, it is possible to improve accuracy and shorten execution time [130]. To investigate which part of the input information surrounding the HLA gene is used to make learning predictions, predictions were made using not only a small number of nearby SNP information but also many SNP information distributed over a wider area by visualizing the learning information of the model [29,129,130]. While conventional methods are strong at learning using nearly SNP information and not good at

learning using SNP information located at distant locations, some new systems are thought that prediction accuracy may have improved because this problem was overcome. HLA genes are involved in the onset of a variety of diseases and are attracting attention [128,157]. As an important area from the perspective of elucidating pathological conditions and realizing personalized medicine. The increased prediction accuracy of HLA gene sequences is expected to increase the reliability of HLA analysis, including integrated analysis between different racial populations, and is expected to greatly contribute to the identification of HLA gene sequences associated with diseases and further elucidation of pathological conditions [129,158,159].

On the multi-task learning approach used for predicting HLA gene sequences, model architecture is a multitask convolutional neural network with a shared part of two convolutional layers and a fully connected layer, followed by individual fully connected layers for each HLA gene group [130]. The HLA genes are grouped into 4 groups: (1) Class I, (2) Class II, (3) Class III, and (4) Other genes like TAP. The shared convolutional layers extract common features across HLA genes in a group, while the individual fully connected layers predict allelic dosages for each specific HLA gene. SoftMax activation is used to output imputation dosages ranging from 0.0 to 1.0 for each allele. Dropout and batch normalization are applied for regularization. As for training data, Two HLA reference panels are used for training: Japanese (n = 1118) and Type 1 Diabetes Genetics Consortium (T1DGC) (n = 5122). The input is pre-phased SNV genotypes encoded as one-hot vectors based on reference/alternate alleles. Hierarchical fine-tuning is used by transferring parameters from higher (2-digit) to lower (4/6-digit) allele resolution models during training. As for validation, 10-fold cross-validation is performed on the reference panels to evaluate imputation accuracy. An independent Japanese HLA dataset (n = 908) is also used for validation. Multi-ethnic validation is done using 1000 Genomes Project data. Accuracy metrics used are sensitivity, positive predictive value (PPV), r2 for dosage, and concordance rate for best-guess genotypes. Thus, the multi-task deep learning approach allows leveraging shared features across HLA genes while still predicting individual gene sequences, outperforming conventional methods, especially for low-frequency alleles.

As for limitations of methodology of fixed prediction targets, unlike genotype imputation methods that can handle missing genotypes, the multi-task learning approach treats HLA allele imputation as a fixed classification problem, which may limit its flexibility [130]. For dependence on linkage disequilibrium (LD), while the multi-task learning is less dependent on distance-based LD decay patterns compared to conventional methods, its performance may still be influenced by complex LD structures in the MHC region [130]. For hierarchical fine-tuning assumptions, the multi-task learning approach assumes a hierarchical structure of HLA alleles (2-digit, 4-digit, 6-digit) and transfers parameters across these levels. However, some alleles in the reference panel may not follow this hierarchy, potentially affecting the model's performance [130]. As for broader limitations of the multi-task learning approach for HLA imputation in limited applications in population genetics, the application of deep neural networks, particularly multi-task learning, to population genetics data has been limited so far, highlighting the need for further exploration and validation [130]. For computational complexity, while multi-task learning aims to reduce processing time by grouping tasks, the computational complexity of deep learning models can still be a limitation, especially for large-scale biobank data [130]. As for interpretability, deep learning models, including multi-task architectures, are often criticized for their lack of interpretability, making it challenging to understand the underlying patterns learned by the model. As for data requirements, multi-task learning approaches may require large and diverse reference panels to effectively capture the complex patterns and dependencies across HLA genes, which can be challenging to obtain [130]. It's important to note that the authors acknowledge the need for further research and validation to address these limitations and explore the full potential of multi-task learning for HLA imputation and other applications in population genetics.

The studies provide several empirical examples demonstrating the effectiveness of the proposed method for multi-task learning to predict HLA gene sequences. The applied multi-task learning to two different HLA imputation reference panels—a Japanese panel (n = 1118) and type I diabetes genetics consortium panel (n = 5122). Through 10-fold cross-validation on these panels, the multi-task learning achieved higher imputation accuracy than conventional methods, especially for imputing low-frequency and rare HLA alleles [130]. They further evaluated multi-task learning on an independent HLA dataset (n = 908) and found it outperformed other methods [130]. Simulation studies on multi-ethnic individuals from the 1000 Genomes Project demonstrated the multi-task learning's ability to accurately impute HLA alleles across diverse populations [130]. The authors applied multi-task learning to impute HLA alleles in large genome-wide association studies from the BioBank Japan (BBJ) and UK Biobank (UKB) cohorts. They then performed trans-ethnic fine-mapping analysis in the MHC region using the imputed HLA data, highlighting a practical application [130]. In a preprint, Naito et al. reported results from applying their similar multi-task CNN architecture to the type I diabetes genetics consortium dataset, achieving 97.6% imputation accuracy, superior to SNP2HLA and comparable to HIBAG. The studies provide concrete empirical evidence across multiple reference panels, independent datasets, simulations, and real data applications demonstrating the effectiveness of the multi-task deep learning approach for accurate HLA gene imputation.

Multi-task learning can contribute to improving accuracy and reducing execution time compared to training separate models for each task by leveraging shared representations and features across related tasks [160,161]. As for improved accuracy, multi-task learning allows the model to learn more robust and generalizable feature representations by sharing knowledge across related tasks. This helps capture underlying patterns that are useful for multiple tasks, leading to better generalization. By jointly optimizing for multiple objectives, multi-task learning can regularize the model and prevent overfitting to any single task, resulting in improved overall accuracy. Multi-task learning enables transfer learning, where knowledge gained from one task can be transferred to related tasks, boosting their performance. As for reduced execution time, with multi-task learning, a single model is trained to perform multiple tasks simultaneously, eliminating the need to train and deploy separate models for each task. This reduces the overall training and inference time. Shared representations and parameters across tasks in multi-task learning architectures like Hard Parameter Sharing (HPS) and Cross-Stitch Networks lead to fewer parameters compared to training individual models, resulting in faster execution. Efficient multi-task learning methods like FastCAR optimize the network architecture and loss function weighting schemes, allowing for faster convergence and reduced training time while maintaining accuracy [162]. Techniques like dynamic task weighting and gradient modulation in multi-task learning L can help balance the learning of different tasks, leading to faster convergence and lower execution times. Thus, multi-task learning improves accuracy by leveraging shared representations and knowledge transfer across related tasks, while reducing execution time by eliminating redundant computations and optimizing architectures and loss functions for efficient multi-task training and inference.

Leveraging insights into HLA diversity holds significant potential for personalized medicine and disease treatment in several key areas. In organ transplantation, HLA typing plays a crucial role in matching donors and recipients for organ transplantation. By utilizing AI and machine learning to analyze complex HLA data, we can optimize donor-recipient matching, reducing the risk of organ rejection and increasing the chances of successful transplantation. This personalized approach improves transplant outcomes. In disease susceptibility, certain HLA alleles are linked to an increased risk of developing complex diseases like type 1 diabetes, RA, and some cancers. AI and machine learning can identify patterns in massive HLA datasets, pinpointing specific disease-associated alleles. This enables early detection, risk stratification, and personalized treatment plans tailored to an individual's HLA profile. In pharmacogenomics, HLA typing is valuable in pharmacogenomics, the study of how genes influence drug responses. By leveraging AI to analyze HLA

data, we can predict an individual's potential reaction to specific drugs, avoiding adverse effects and optimizing therapeutic efficacy. This personalized approach enhances drug safety and treatment outcomes. In vaccine development, understanding HLA diversity across populations can inform vaccine design and deployment strategies. Tailoring vaccines to account for HLA variation within target populations may maximize their effectiveness and enable a more precise, personalized approach to vaccination programs [163]. Thus, harnessing HLA diversity through AI and machine learning offers opportunities for personalized medicine across organ transplantation, disease risk assessment, pharmacogenomics, and vaccine development. This approach has the potential to revolutions healthcare by enabling early intervention, optimized treatments, and improved patient outcomes tailored to an individual's unique genetic profile.

Multi-task learning for predicting HLA gene sequences has several potential real-world applications in the context of HLA diversity and immune response. In disease association studies, HLA genes play a crucial role in the immune system and are associated with various autoimmune diseases, infectious diseases, and adverse drug reactions. Accurate imputation of HLA alleles can facilitate large-scale association studies, leading to a better understanding of disease mechanisms and personalized medicine approaches [164]. In transplantation medicine, HLA matching is critical for successful organ and stem cell transplantation. Precise HLA typing of donors and recipients can improve transplant outcomes by reducing the risk of rejection and graft-versus-host disease [164]. HLA diversity influences the immune response to pathogens and vaccines in vaccine development. Accurate HLA imputation can help identify population-specific HLA alleles associated with vaccine efficacy or adverse reactions, informing the design of more effective and safer vaccines [163]. HLA molecules present antigenic peptides to T cells in immunotherapy, playing a crucial role in cancer immunotherapy. Precise HLA typing can aid in developing personalized cancer immunotherapies, such as neoantigen-based vaccines and adoptive T-cell therapies [163]. In population genetics studies, HLA genes are among the most polymorphic in the human genome, making them valuable markers for studying human evolution, migration patterns, and population structure [163]. By leveraging multi-task learning and deep neural networks, the proposed methodology can accurately impute HLA alleles, including low-frequency and rare variants, from SNP data across diverse populations. This can facilitate large-scale studies and applications in precision medicine, transplantation, vaccine development, and population genetics, ultimately leading to a better understanding and management of immune-related diseases and conditions.

4. Multimodal AI

Thus, harnessing HLA diversity through AI and machine learning offers opportunities for personalized medicine across organ transplantation, disease risk assessment, pharmacogenomics, and vaccine development. Transformer brought about such a dramatic change in the field of AI that have "changed everything about deep learning." Natural language models using Transformer, BERT (bidirectional encoder representations from Transformer) (Figure 2), GPT-3, PaLM are rapidly improving their performance and are becoming common knowledge. Transformer, which is a neural network consisting only of attention mechanisms, is used not only in the field of natural language but also in a variety of other fields, including Vision Transformer for image recognition, Alphafold2 for protein 3-dimension (3D) structure analysis, CLIP that combines language and images, and Gato, which performs over 600 tasks, is developed using Transformer. With Gato, data such as images and language are essentially the same regardless of modality, making it possible to create multiple AI.Large language model (LLM) is capable of multimodal learning, which means learning data with different properties, modes at the same time [164]. In other words, multimodal AI is a deep learning method that inputs multiple types of data and processes it in an integrated manner. CNNs create decision machines and process data from a single type of information input.

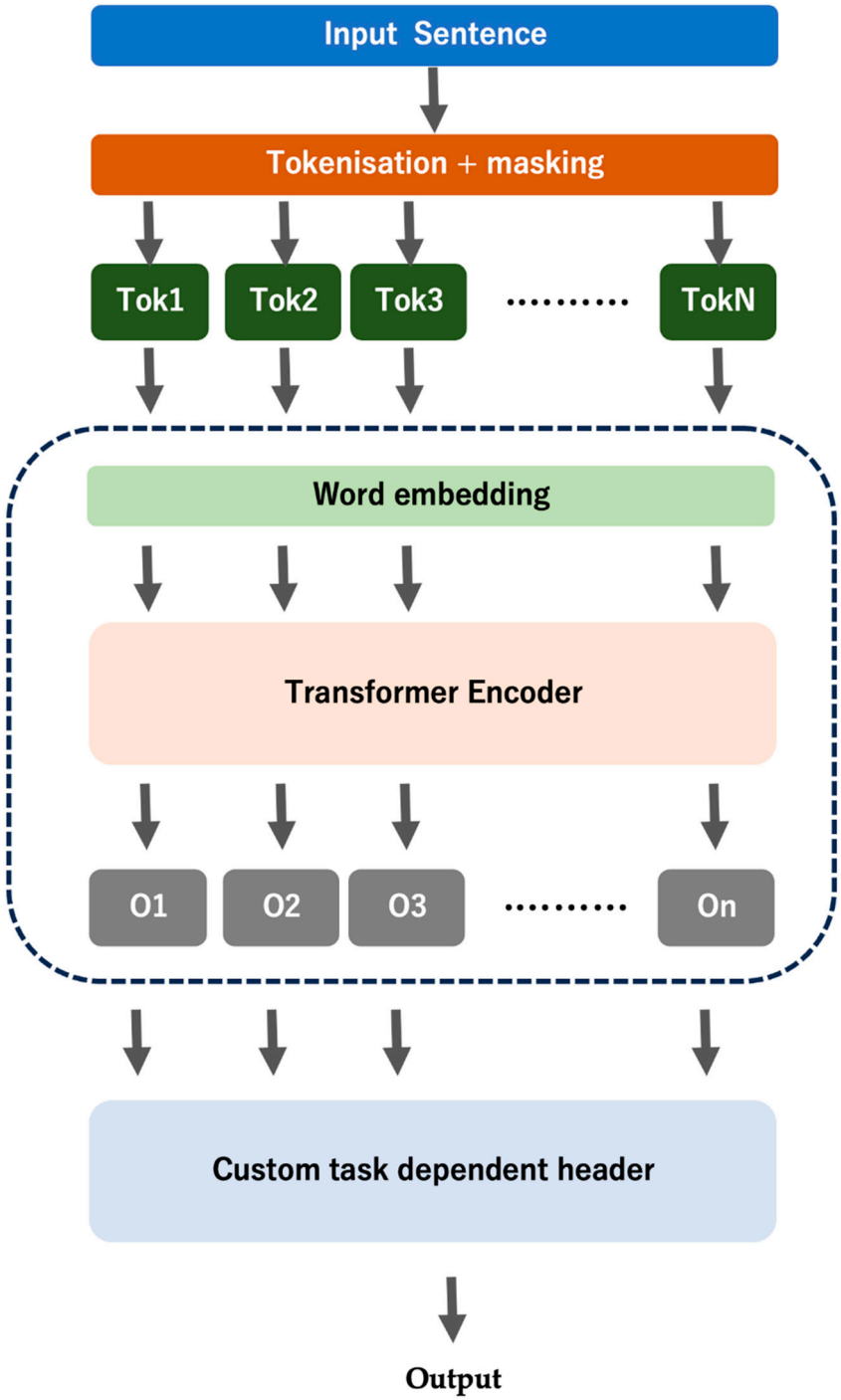

Figure 2. Schematic representation of Bidirectional Encoder Representations from Transformers (BERT). Input sentence is split into multiple tokens (Tok N) and fed to a BERT model, which outputs embedded output feature vectors, ON, for each token. By attaching different head layers on top, it transforms BERT into a task-oriented model.

On the other hand, multimodal AI refers to a deep learning method that takes multiple types of data as input and processes them in an integrated manner. When humans process information, they combine multiple sensory information that comes in from the outside, such as visual information, smell, touch, taste, and hearing, represented by the five senses. Multimodal AI uses a method of creating a deep learning model from multiple data like the information processing performed by the human brain and creating a judger. For example, when determining an object in a certain image, in most cases an AI model is

created by learning only the image, visual information. This type of learning and making decisions based on only one piece of information is called "single modal". On the other hand, in the case of multimodal AI, it becomes possible to judge objects based on multiple modals of the five human senses, such as visual information, auditory information, and olfactory information. However, this is a future vision of multimodal AI that may be realize in the future, and the multimodal learning that is used has not reached that level. One possibility is to use the meta information of the image to improve the accuracy of discrimination. As multimodal learning progresses, a single AI model will be able to determine multiple factors [165]. This increases the possibility of being able to deal with unpredictable abnormal patterns.

Advances in the field of machine learning, represented by natural language processing (NLP), and the increasing functionality of computers have established a foundation for efficiently extracting information from the vast amount of sequence data that has been accumulated through the development of sequencing technology. Language mode-related research is rapidly developing in protein research using machine learning. A language model is a representation learning technology that learns the characteristics of the data itself from large-scale data [166]. It was originally proposed as a statistical model that learns the feature representation of each word that makes up a natural language sentence generation probability function. Protein language models are language models that are trained by treating the primary sequence of a protein as a "sentence" and each amino acid as a "word" [167–170]. Broadly divided by architecture, there are word2vec, CNNs, recurrent neural networks (RNNs), and those based on Transformer, which is a neural network proposed in connection with machine translation. Unlike conventional CNNs and RNNs, Transformer is unique in that it can incorporate dependencies between distant tokens through its self-attention mechanism. Here, the words (processing units) processed by the language model are called tokens, which in this paper correspond to each amino acid included in a protein sequence [171]. When using Transformer, information is processed using a three-layered structure: (i) vector representation (token embedding) specific to each amino acid, (ii) incorporation of context information via self-attention, and (iii) calculation of amino acid appearance probability for each residue position. The amino acid representation finally obtained by Transformer is called an embedding and corresponds to the output of (ii). An architecture has been proposed that combines an encoder to obtain embeddings and a decoder to generate arrays.

To treat biological sequences with machine learning, it is first necessary to obtain their feature representation [172,173]. For proteins, classically, position-specific vectors expressions based on so-called descriptors, including AAindex, ST-scale, Z-scale, T-scale, FASGAI, MS-WHIM, ProtFP, VHSE etc., or MSA that reflect the physicochemical and structural properties of amino acids in a sequence scoring matrix (PSSM), etc. have been used [174,175]. A language model in natural language processing applied to proteins, amino acid sequences is called a protein language model [176–179]. Since there are many unknowns about how proteins function in living organisms, it is hoped that new knowledge will be obtained by using large amounts of data and pre-learning methods such as BERT [171,180] (Figure 2). The Transformer used in BERT is an architecture widely used in natural language processing (Figure 2). There are increasing attempts to improve the interpretability of models by analyzing the attention mechanism in this Transformer. In particular, the one that interprets BERT is called BERTology. There are two main analysis methods for Transformer-based pre-trained models, which are effective for the target task by performing learning on a different task before the target task to be solved for acquiring features. To make robust predictions possible even when there is little training data, it is possible to solve this problem by pre-training BERT on a large-scale protein database [171,180,181]. Analysis of attention mechanism: among the amino acid residues whose attention weights exceed a threshold,

it checks how many combinations have biological characteristics [182]. Specifically, it is calculated using the following formula, which is an evaluation index close to precision.

$$\mathrm{fa}(x) = \frac{\sum_{x \in X} \sum_{i=1}^{|x|} \sum_{j=1}^{|x|} \mathbf{f}(i,j) \cdot \mathbf{Lai}, \mathbf{j} > \theta}{\sum_{x \in X} \sum_{i=1}^{|x|} \sum_{j=1}^{|x|} \mathbf{Lai}, \mathbf{j} > \theta} \quad (1)$$

A probe task is an auxiliary task used to improve the interpretability of pre-trained models in natural language processing, which checks what kind of information the internal representation acquired in the pre-trained model contains. Specially, it input the representation obtained from the model into a classifier, have it solved some tasks, and check whether useful information is encoded for that task. There are two types of probe tasks for Transformer-based models: embedded probes and attention probes. Embedded probes target the output of each layer, and attention probes target attention weight [183,184]. In the analysis of attention mechanisms, contact maps are important features in protein secondary structures [176,185,186]. A contact map is a combination of spatially close to amino acids in protein folding. Using this contact map as a feature for amino acid pairs and applying the attention analysis method described above, fa(x) for each pre-trained model ranged from 44.5% to 63.2%. In addition, the attention heads that most reflected the contact map were in deeper layers. Considering that the background probability in the contact map is 1.3%, the pre-learning in the protein language model has acquired a high-order representation that reflects the contact map [173,178,187]. Binding sites are the sites where proteins interact with other molecules and are very important features for protein function [188]. The proportion of attentions that refer to this binding site, fa(x) is between 45.6% and 50.7%, which is a very high value considering that the background probability of the binding site is 4.8%. Additionally, most attention heads refer to binding sites at a high rate [189]. The reason why the binding site is given such importance in the model even though it is a feature that reflects interaction with external molecules is that it has been suggested that structural motifs may be highly conserved because they are features directly related to protein function [188,190,191]. Post-translational modifications are changes made to proteins after they are translated from mRNA, and phosphorylation is a typical post-translational modification, which play a major role in protein structure and function [192]. In post-translational modification, fa(x) is 64%, which is a very high value considering that the background probability of post-translational modification is 0.8% [193]. However, the number of attention heads that refer to post-translational modification site is small [194]. As a result of performing embedding probes and attention probes in each layer of the pre-learning model, the F1 score is used for secondary structure prediction, and the precision is used for binding site prediction and contact prediction. Prediction of secondary structures such as helices, turns/bends, and strands show prediction accuracy even using outputs at relatively low layers [195].

One of the advantages of using language models is that their pre-training can be performed unsupervised, which is also called self-supervised learning because it learns using the input array itself as a teacher. The specific method differs depending on the architecture of the language model. Herein, we will focus on masked language modeling (MLM), which was introduced in BERT and is widely used. With this method, stochastically mask some tokens in the input array. By inputting this masked sequence into Transformer, the likelihood of each amino acid at the masked position can be calculated. Therefore, in MLM, the negative log likelihood L (2) for the original amino acid at all mask positions M is minimized as a loss function. Thus, by inferring the excluded residue from information on surrounding residues (context), it is possible to learn the appearance pattern of amino acids in the sequence [196,197].

$$\mathrm{L} = -\sum_{i \in M} \log(\mathrm{pi}, \mathrm{a}) \quad (2)$$

In a method using a BERT model with a self-attention mechanism pre-trained using a large-scale protein database, it is possible to predict B cell epitopes and peptide binding to MHC class II [171,180,198]. A vaccine is a substance, such as a weakened pathogen, that is artificially administered to create immunity against an infectious disease. Historically, direct ingestion of live, attenuated pathogens has triggered an immune response within the body, but later dead pathogens, their antigens, and toxins alone are used as vaccines. Lymphocytes, especially B cells and T cells, are involved in acquiring immunity against infectious diseases. B cells that induce antigens-specific immune response in vivo express B cell receptors (BCRs) on their cell membranes, which directly bind to antigen proteins and recognize their epitope regions, which allows the production of large amounts of antigen-specific antibodies. On the other hand, the T antigen peptide produced when the pathogen protein is phagocytosed by antigen-presenting cells as an immunogen must be used as an immunogen for cells to recognize pathogen proteins, needs to be presented on the cell surface by binding to MHC molecules. The T cell then recognizes the presented peptide and responds and activates B cells that produce antibodies or directly destroys infected cells [199]. Based on these properties, research has recently begun on peptide vaccines, which attempt to acquire immunity by administering small peptide fragments that have a structure accepted by B cells and T cells [200].

By administering a vaccine that mimics the structure of a specific antigen epitope region, it is possible to induce the production of monoclonal antibodies specific to that antigen [201]. To search for peptide vaccine candidates that act against B cells and T cells, it is necessary to predict the peptides that bind to BCRs and the peptides that bind to MHC molecules. It has traditionally been thought that this prediction requires precise clarification of the steric bounds between proteins [202]. Experimentally determining the 3D structure of an antibody-antigen complex is extremely time-consuming and labor-intensive, making it unsuitable for assaying large numbers of candidate peptides, such as in vaccine development [203]. When predicting bonds using computer simulations, there are still issues in which bonds can only be investigated when proteins are considered as rigid bodies, and issues that require enormous amounts of time and computational resources even when using molecular dynamics-based methods. On the other hand, in recent years, research has been progressing on binding prediction using machine learning that does not explicitly treat binding between proteins. However, high accuracy has not yet been achieved due to challenges in B cell epitope prediction and prediction of peptide binding to MHC molecules, especially MHC class II molecules, using machine learning [204,205]. The first problem is that it is difficult to learn long-distance dependencies. In recent years, long-distance dependency learning has been attempted using LSTM (Long Short-Term Memory), a type of deep learning, for sequence data of amino acid sequences. LSTM is a model that deals with the problem of long-distance dependencies using storage parts called cells [206,207]. However, it still needs to go through a long-distance network between amino acids, and information may be lost [208,209].

Furthermore, when applied to proteins, understanding the dependence relationships caused by the higher-order structure of the protein remains an issue [210–212]. In the task of predicating peptides that bind to MHC class II, it is necessary to capture the complex interactions between amino acids in both MHC class II and the peptide, which is difficult to capture such interactions with LSTM [207,213,214]. In predicting peptides that bind to MHC class I molecules, one of the two major classes of MHC molecules, there are fewer problems learning long-distance dependencies, and predictions using machine learning can achieve higher accuracy, because the amino acid sequence is shorter than MHC class II [214–216]. The second issue is that sufficient generalization performance cannot be obtained when supervised learning data is small [217]. The simplest approach to improving generalization performance is to prepare more supervised learning data, but this is not easy because biological experiments to increase training data require a large amount of cost. To solve these problems, BERT, which is based on a large-scale protein database, does not store sequence information as an intermediate vector like LSTM, but can

model long sequences with variable lengths by using only an attention mechanism [218]. In LSTM, the farther the distance between amino acids, the more networks it must pass through, and as a result, there is a risk of losing information about amino acids that are far apart. The shorter this distance, the easier it is to learn long-distance dependencies, but the attention mechanism, which also allows to capture complex interactions between two series data, has the advantage of being able to directly model relationships between amino acids, regardless of their positions in the sequence [217,219]. Another great feature of BERT is that it can use large amounts of unsupervised data for pre-training [220]. By pre-training BERT based on a large amount of protein data, generalization performance can be improved even when training data is small [172,182]. One of the main features of BERT is that it uses a bidirectional Transformer in its architecture, which allows it to learn long-distance dependencies better than LSTM [221]. Transformer is a neural network consisting only of attention mechanisms, and Transformer can capture long-distance dependencies better than LSTM. Model with multiple layers of unidirectional Transformers perform better than LSTM in various NLP tasks. BERT furthermore makes the Transformer bidirectional so that sequence information from the opposite direction is also used to improve performance. In this task, it is important to understand the long-range dependence of each amino acid.

When predicting the binding of peptides to MHC class II, it is important to learn the complex interactions between MHC class II and peptides [222,223]. BERT can capture such long-distance dependencies and can be expected to learn complex interactions by adding multiple layers of attention mechanisms. Another main feature of BERT is that generalizations performance can be improved by pretraining with MLM on large datasets, which in MLM is a task in which some words are deleted from an input sentence and the deleted words are predicted from the surrounding context. This task allows to learn the features of the word itself, considering the context. MLM can use large amounts of unsupervised text data because the words themselves functions as a type of supervised data [224]. Through this pre-training, BERT has further improved its performance in various NLP tasks. In addition, pre-training can learn more efficiently than other methods even when there is little data.

In the early stages of research, B cell epitope prediction used only the physicochemical properties of the amino acids that make up proteins as features [225,226]. Since then, a method based on machine learning that incorporates information on the amino acid sequence itself has achieved relatively high accuracy, and many methods are used, including those using support vector machines, Random Forests, forward propagation neural networks, RNNs, and LSTM with attention mechanisms [227]. Although Lbtope and BepiPred-2.0 use features of dipeptide composition, long-range features of antigen proteins are not incorporated into the model [228]. Because DLBepitope uses a simple forward propagation neural network, it is difficult to learn complex dependencies between amino acids compared to models with sequence information and attention mechanisms. ABCpred attempts to learn sequence information using RNN, but RNN cannot capture long-distance dependencies. The method using LSTM with attention mechanism achieves higher accuracy than BepiPred-2.0 by combining features that take such long-distance dependencies into account and structural and chemical features of the entire antigen protein [229–231]. However, since LSTM is used, information between amino acids that are far apart may be lost. Many machine learning methods have been proposed for predicting peptide binding to MHC class II, but the following two issues need to be appropriately addressed [227,232,233]. The first issue is the problem of vanishing gradients and long-range dependence problems that occur in algorithms for long sequence data such as amino acid sequences [234]. DeepSeqPanll deals with this by using a model that combines LSTM with an attention mechanism and CNN, but this problem remains because it uses LSTM in the model [235]. The second issue is the need to consider the interaction between peptides and MHC class II [222]. As input a graph representation and amino acid sequence of the protein, graph representation of compound and amino acid sequence project onto a low-dimensional vector using a graph neural network (GNNs) and CNNs, respectively, and the interaction between

these vectors is captured using an attention mechanism, showing higher performance than existing methods [236–238]. However, since GNNs require structural information as input, they cannot be trained in advance using a large-scale protein database, which will be described later, and CNNs cannot capture long-distance dependencies [239,240]. By using only attention mechanism in the model, simultaneously solve the problem of long-range dependence and the problem of considering the interaction between peptide and MHC class II, while improving generalization performance by performing pre-learning on a large-scale protein database [171,178,241,242]. Furthermore, machine learning in general has the problem of not being able to sufficiently learn features from the training data when the training data is small [243]. Simply increasing the training data is difficult because it requires a large amount of cost, and in recent years, methods such as data augmentation and pre-learning have been used. Data augmentation is a method of extending learning data by making predictions using a machine learning model for similar data and adding temporary labels. In predicting peptide binding to MHC class II, the performance of the model is improved by expanding the training data for peptide binding prediction using ligand data obtained from mass spectrometry [57,244].

In addition, there are two issues: it is difficult to learn complex dependencies between distant amino acids, and accuracy is low when there is insufficient training data [206,208,245]. To solve these issues, there is a method for predicting peptide binding using BERT, in which the advantages are (1) the attention mechanism allows learning long-range dependencies and interactions between amino acids in MHC class II and peptide bounds; (2) the generalization performance of the model can be improved by using a large-scale unsupervised database for pre-training [246–249]. One of the main features of BERT is that it can learn long-distance dependencies better than LSTM by adopting a bidirectional Transformer in its architecture, which can capture long-distance dependencies better than LSTM. Models with multiple layers of unidirectional Transformer perform better than LSTM in various NLP tasks. BERT further improves performance by making the Transformer bidirectional and using sequence information from the opposite direction. In this task, it is important to capture the long-range dependencies of each amino acid [250]. In this task, it is important to capture the long-range dependencies of each amino acid [206,208,209]. When predicting the binding of peptides to MHC class II, it is important to learn the complex interactions between MHC class II and peptides. BERT can capture such long-distance dependencies and is expected to learn complex interactions by adding multiple layers of attention mechanisms. Another main feature of BERT is that it can improve generalization performance by pre-training large-scale datasets using MLM in NLP, in which some words are deleted from an input sentence and the deleted words are predicted from the surrounding context and the words themselves function as a type of supervised data, so a large amount of unsupervised text data can be used. Through this pre-training, BERT can further improve its performance in various NLP tasks [251,252]. In addition, pre-training can learn more efficiently than other methods even when there is little data [253]. Therefore, BERT is applied to peptide bond prediction to capture long-distance dependencies between amino acids.

B-cell epitope prediction requires only a single peptide sequence as input, so BERT can be used directly by inputting the peptide sequence directly to BERT [254,255]. In predicting peptide binding of MHC class II molecules, it is not obvious how to apply BERT because it handles as input the triple of a peptide sequence, the amino acid sequence of the alpha chain of the MHC class II molecules, and the amino acid sequence of the beta chain of the MHC class II molecule. Traditionally, models before BERT project multiple inputs into vectors independently and then apply a bidirectional attention mechanism [256]. It is possible to effectively model relationships between multiple inputs by concatenating inputs with a special token called a separate token and applying a bidirectional self-attention mechanism. As a result, it is expected that the model can be modeled in the same way regardless of whether the input is a single amino acid sequence or multiple amino acid sequence. However, when BERT is trained only on supervised learning data, that is, when no prior training is performed, BERT cannot learn well. This is because there is little

training data used to predict B cell epitopes and peptide binding to MHC class II, making it difficult to adjust the huge parameters of BERT. Therefore, when pre-training BERT using the large-scale protein database Pfam, the amount of supervised learning data used to predict B cell epitopes is B cell epitopes is approximately 200,000, and the amount of supervised learning data used to predict peptide binding to MHC class II is approximately 50,000, Pfam is a database of approximately 31 million proteins [171,180,181]. Therefore, by pre-training Pfam using MLM, it is possible to adjust the huge parameters of BERT [257]. In addition, BERT pretrained with Pfam can improve generalization performance for tasks such as learning protein structure and function, secondary structure prediction, and contact prediction [171]. Regarding the architecture of the model, after receiving as input a sequence of N characters of amino acids x = (x1, x2, x3,. . .. .xN), the output is a sequence of H-dimensional embedding vector Z = (z1, x2, x3,. . ... zN) in the l layer. When the average vector S of the output Z = (z1, x2, x3,. . ... zN) of the final layer of BERT is an aggregate representation of the input sequence, the vector S∈RH takes the inner product with the weight W∈RH of the classification layer. As a result, it is projected into one dimension, and the final output o is activated by the sigmoid function. Cross entropy error is used for optimization.

$$S = 1/n\sum_{i}^{n} Zi \quad (3)$$

$$O = \text{sigmoid}\ (SW^{T}) \quad (4)$$

In addition, a comparison of AUC and SECC on test data for BERT with pre-training and BERT without pre-training shows that pre-training contributes to improving generalization performance, but this is because the parameters of BERT, which has a high degree of freedom, cannot be learned sufficiently unless pre-training is performed using a large-scale database. One way to improve model performance is to perform data augmentation separately from pre-learning. In addition, by visualizing attention, BERT can see where the learned model is paying attention when making predictions. BERT focuses on amino acids in close positions in the 3D structure and on binding sites [256].

Med-UniC is a novel framework designed to integrate multimodal medical data from English and Spanish for cross-lingual medical vision-language pre-training. It addresses the challenge of community bias caused by different languages by proposing Cross-lingual Text Alignment Regularization (CTR). CTR explicitly unifies cross-lingual semantic representations of medical reports from diverse language communities through latent language disentanglement, mitigating bias from determining positive-negative sample pairs within analogous reports [258]. In cross-lingual medical language model, constructs a cross-lingual vocabulary by combining English and Spanish medical terms. Pre-trains on mixed English and Spanish medical reports using masked language modeling. Selectively updates high layers to alleviate catastrophic forgetting during vision- language pre-training. In vision-language alignment, incorporates contrastive learning to predict matched image-text pairs while mapping negative pairs far apart. Uses non-linear visual and linguistic projectors to map images and texts to a joint embedding space. In Cross-Lingual Text Alignment Regularization (CTR), Optimizes through latent language disentanglement, not depending on negative samples. Mitigates bias from determining positive-negative sample pairs within analogous reports. Ensures cross-lingual representation is not biased toward any specific language community. Med-UniC achieves superior performance across 5 medical image tasks and 10 datasets covering over 30 diseases, offering a versatile framework for unifying multimodal medical data from diverse linguistic communities. Experimental results highlight the presence of community bias in cross-lingual vision-language pre-training and show that reducing this bias enhances performance in both vision-language and uni-modal visual task [258].

ETP (ECG-Text Pre-training) is an innovative approach that leverages both electrocardiogram (ECG) data and corresponding textual reports within a cross-modal learning paradigm to learn transferable ECG representations [258]. The key idea is to use a trainable

ECG encoder and a frozen language model to embed paired ECG signals and automatically generated clinical reports separately [259]. The self-supervised pre- training objective aims to maximize the similarity between paired ECG and report embeddings while minimizing the similarity between ECG and other mismatched reports [259]. This multimodal ECG-text pre-training approach enables zero-shot learning for ECG classification tasks, achieving around 10% performance improvement without using any annotated data compared to supervised and self-supervised baselines that rely on annotated data [259]. Notably, ETP demonstrates high generalizability, effectiveness, and efficiency, attaining the highest recall and F1 scores on the MIT-BIH dataset despite containing different ECG classes from the pre-training dataset [259]. The frozen language model component plays a crucial role in helping ECG zero-shot learning by providing rich semantic representations from the textual reports, which guide the ECG encoder to learn transferable representations that can generalize to unseen classes during inference [259].

The integration of genomic data, AI technologies, and clinical outcomes significantly contributes to a better understanding of HLA diversity and its role in various pathological conditions. HLA genes are highly polymorphic and play a crucial role in immune response, making them important for understanding disease susceptibility and outcomes. Advanced genomic technologies like next-generation sequencing and imputation methods enable high-resolution characterization of HLA allelic variations [260]. AI and computational approaches facilitate analysis of large-scale HLA genomic data and identification of disease associations [261]. Genome-wide association studies consistently show strong associations between HLA variations and immune-mediated diseases like autoimmune disorders, but conventional methods are limited in fine-mapping these highly polymorphic regions. High-resolution HLA sequencing and AI-driven analysis enable unambiguous identification of disease-associated polymorphisms. Integrating HLA genomics with clinical data has revealed associations between specific HLA-KIR combinations and transplantation outcomes like graft rejection and graft-versus-host disease [261]. This aids in donor selection and personalized transplantation strategies. Studies combining HLA sequencing, AI algorithms, and phenotypic data from large biobanks like UK Biobank have uncovered novel HLA associations with cancer risk and heterogeneity across subtypes [260,262]. This highlights HLA diversity's role in tumor immunology and immunotherapy response. AI-driven analysis of HLA diversity along with KIR genomic regions provides insights into innate immunity mechanisms relevant for infectious diseases and cancer immunotherapies [261]. Therefore, the synergy between high-throughput genomics, AI-powered bioinformatics, and deep phenotyping through biobanks enables comprehensive decoding of HLA polymorphisms and their intricate connections to various pathological conditions, paving the way for improved diagnostics and therapeutics [260–262].

Recent developments in multimodal AI have significant implications for cancer immunotherapy. Multimodal AI models can integrate diverse data sources like radiology, histology, genomics, and electronic health records to provide a comprehensive view of the patient's immune response and tumor microenvironment. This holistic assessment enables better identification of biomarkers and predictors for immunotherapy response and resistance mechanisms [263,264]. As for improved patient stratification, Multimodal AI can capture the complexity of antitumor immune responses by combining multiple modalities like tumor genomics, immune cell profiling, and imaging data. This allows for more accurate stratification of patients who are likely to respond to immunotherapies like immune checkpoint blockade (ICB) [264]. In identification of multimodal biomarkers, Single biomarkers often have limited predictive power for immunotherapy response. Multimodal AI can discover novel multimodal biomarkers by integrating complementary information across modalities, providing a more robust predictor of treatment outcomes [262,264]. In precision immunotherapy selection, by profiling the patient's immune status holistically, multimodal AI can identify the key rate-limiting factors hindering antitumor immunity. This enables selection of optimal immunotherapy strategies or combinations tailored to overcome the specific resistance mechanisms in each patient [264]. As for monitoring

treatment response, longitudinal multimodal data can be leveraged by AI to monitor dynamic changes in the tumor microenvironment during immunotherapy. This allows early prediction of treatment response or resistance, enabling timely adaptation of therapeutic strategies [262,264]. In novel biomarker discovery, AI-driven exploration of multimodal data associations can uncover previously unknown patterns and relationships between different modalities. These insights can guide the discovery of novel biomarkers and therapeutic targets for cancer immunotherapy [262,264]. Therefore, multimodal AI approaches provide a powerful framework to dissect the complexities of antitumor immunity, enabling more precise and personalized cancer immunotherapy strategies.

5. Conclusions

Cancer antigen vaccine therapy is a treatment method that actively immunizes cancer patients with TAAs from the outside to induce and enhance antitumor immunity, with the aim of inducing T cell immune responses against TAAs. As clinical trials progress, cancer antigen vaccine therapy has few side effects than conventional cancer treatments and is effective in prolonging survival while maintain patients' quality of life. Furthermore, the development of more potent cancer antigen will require more complex cancer immunotherapy that combine treatments that control immunosuppressive mechanisms in the tumor microenvironment. In patients with specific HLA alleles, neoantigen peptides generated by these driver mutations are presented to T cells by the HLA and inducer tumor immunity, thereby eliminating tumor cells. Additionally, advances in genetic analysis technology have made it possible to evaluate the cancer and immune status of individual cancer patients using biomarkers and other methods. When predicting the binding of peptides to MHC class II, it is important to learn the complex interactions between MHC class II and peptides. BERT can capture such long-distance dependencies and can be expected to learn complex interactions by adding multiple layers of attention mechanisms. BERT is applied to peptide bond prediction to capture long-distance dependencies between amino acids. Another great feature of BERT is that it can use large amounts of unsupervised data for pre-training. BERT pretrained with Pfam can improve generalization performance for tasks such as learning protein structure and function, secondary structure prediction, and contact prediction. In addition, by visualizing attention, BERT can see where the learned model is paying attention when making predictions. The prediction tools will be useful for the novel therapeutic approaches for immune diseases via HLA immune responses.

Author Contributions: Writing—review and editing, Y.M.; supervision, R.Y.; funding acquisition, Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Armony, G.; Heck, A.J.R.; Wu, W. Extracellular crosslinking mass spectrometry reveals HLA class I—HLA class II interactions on the cell surface. *Mol. Immunol.* **2021**, *136*, 16–25. [CrossRef] [PubMed]
2. Liu, S.; Bos, N.A.; Verschuuren, E.A.M.; van Baarle, D.; Westra, J. Biological Characteristics of HLA-G and Its Role in Solid Organ Transplantation. *Front. Immunol.* **2022**, *13*, 902093. [CrossRef] [PubMed]
3. Neuchel, C.; Gowdavally, S.; Tsamadou, C.; Platzbecker, U.; Sala, E.; Wagner-Drouet, E.; Valerius, T.; Kröger, N.; Wulf, G.; Einsele, H.; et al. Higher risk for chronic graft-versus-host disease (GvHD) in HLA-G mismatched transplants following allogeneic hematopoietic stem cell transplantation: A retrospective study. *HLA* **2022**, *100*, 349–360. [CrossRef] [PubMed]
4. Arrieta-Bolaños, E.; Hernández-Zaragoza, D.I.; Barquera, R. An HLA map of the world: A comparison of HLA frequencies in 200 worldwide populations reveals diverse patterns for class I and class II. *Front. Genet.* **2023**, *14*, 866407. [CrossRef] [PubMed]
5. Do, M.D.; Le, L.G.H.; Nguyen, V.T.; Dang, T.N.; Nguyen, N.H.; Vu, H.A.; Mai, T.P. High-Resolution HLA Typing of HLA-A, -B, -C, -DRB1, and -DQB1 in Kinh Vietnamese by Using Next-Generation Sequencing. *Front. Genet.* **2020**, *11*, 383. [CrossRef] [PubMed]
6. Crocchiolo, R.; Rombolà, G. Human Leucocyte Antigen System and Selection of Unrelated Hematopoietic Stem Cell Donors: Impact of Patient-Donor (Mis)matching and New Challenges with the Current Technologies. *J. Clin. Med.* **2023**, *12*, 646. [CrossRef] [PubMed]

7. Little, A.M.; Akbarzad-Yousefi, A.; Anand, A.; Diaz Burlinson, N.; Dunn, P.P.J.; Evseeva, I.; Latham, K.; Poulton, K.; Railton, D.; Vivers, S.; et al. BSHI guideline: HLA matching and donor selection for haematopoietic progenitor cell transplantation. *Int. J. Immunogenet.* **2021**, *48*, 75–109. [CrossRef] [PubMed]
8. Bastidas-Legarda, L.Y.; Khakoo, S.I. Conserved and variable natural killer cell receptors: Diverse approaches to viral infections. *Immunology* **2019**, *156*, 319–328. [CrossRef] [PubMed]
9. Pende, D.; Falco, M.; Vitale, M.; Cantoni, C.; Vitale, C.; Munari, E.; Bertaina, A.; Moretta, F.; Del Zotto, G.; Pietra, G.; et al. Killer Ig-Like Receptors (KIRs): Their Role in NK Cell Modulation and Developments Leading to Their Clinical Exploitation. *Front. Immunol.* **2019**, *10*, 1179. [CrossRef] [PubMed]
10. Tichauer, J.E.; Arellano, G.; Acuña, E.; González, L.F.; Kannaiyan, N.R.; Murgas, P.; Panadero-Medianero, C.; Ibañez-Vega, J.; Burgos, P.I.; Loda, E.; et al. Interferon-gamma ameliorates experimental autoimmune encephalomyelitis by inducing homeostatic adaptation of microglia. *Front. Immunol.* **2023**, *14*, 1191838. [CrossRef] [PubMed]
11. Saline, M.; Rödström, K.E.; Fischer, G.; Orekhov, V.Y.; Karlsson, B.G.; Lindkvist-Petersson, K. The structure of superantigen complexed with TCR and MHC reveals novel insights into superantigenic T cell activation. *Nat. Commun.* **2010**, *1*, 119. [CrossRef] [PubMed]
12. Mariani, E.; Lisignoli, G.; Borzì, R.M.; Pulsatelli, L. Biomaterials: Foreign Bodies or Tuners for the Immune Response? *Int. J. Mol. Sci.* **2019**, *20*, 636. [CrossRef] [PubMed]
13. Carnicer-Lombarte, A.; Chen, S.T.; Malliaras, G.G.; Barone, D.G. Foreign Body Reaction to Implanted Biomaterials and Its Impact in Nerve Neuroprosthetics. *Front. Bioeng. Biotechnol.* **2021**, *9*, 622524. [CrossRef] [PubMed]
14. Carlé, C.; Degboe, Y.; Ruyssen-Witrand, A.; Arleevskaya, M.I.; Clavel, C.; Renaudineau, Y. Characteristics of the (Auto)Reactive T Cells in Rheumatoid Arthritis According to the Immune Epitope Database. *Int. J. Mol. Sci.* **2023**, *24*, 4296. [CrossRef] [PubMed]
15. Sun, L.; Su, Y.; Jiao, A.; Wang, X.; Zhang, B. T cells in health and disease. *Signal Transduct. Target. Ther.* **2023**, *8*, 235. [CrossRef] [PubMed]
16. Sanchez, E.E.; Tello-Lafoz, M.; Guo, A.J.; de Jesus, M.; Elbanna, Y.A.; Winer, B.Y.; Budhu, S.; Chan, E.; Rosiek, E.; Kondo, T.; et al. Apoptotic contraction drives target cell release by cytotoxic T cells. *Nat. Immunol.* **2023**, *24*, 1434–1442. [CrossRef] [PubMed]
17. Sanchez, E.; Huse, M. How cytotoxic T cells release their dying targets. *Nat. Immunol.* **2023**, *24*, 1413–1414. [CrossRef] [PubMed]
18. Milovanovic, J.; Arsenijevic, A.; Stojanovic, B.; Kanjevac, T.; Arsenijevic, D.; Radosavljevic, G.; Milovanovic, M.; Arsenijevic, N. Interleukin-17 in Chronic Inflammatory Neurological Diseases. *Front. Immunol.* **2020**, *11*, 947. [CrossRef] [PubMed]
19. Zhou, Y.; Lih, T.M.; Pan, J.; Höti, N.; Dong, M.; Cao, L.; Hu, Y.; Cho, K.C.; Chen, S.Y.; Eguez, R.V.; et al. Proteomic signatures of 16 major types of human cancer reveal universal and cancer-type-specific proteins for the identification of potential therapeutic targets. *J. Hematol. Oncol.* **2020**, *13*, 170. [CrossRef] [PubMed]
20. Abd Hamid, M.; Peng, Y.; Dong, T. Human cancer germline antigen-specific cytotoxic T cell-what can we learn from patient. *Cell. Mol. Immunol.* **2020**, *17*, 684–692. [CrossRef] [PubMed]
21. Weigelin, B.; den Boer, A.T.; Wagena, E.; Broen, K.; Dolstra, H.; de Boer, R.J.; Figdor, C.G.; Textor, J.; Friedl, P. Cytotoxic T cells are able to efficiently eliminate cancer cells by additive cytotoxicity. *Nat. Commun.* **2021**, *12*, 5217. [CrossRef] [PubMed]
22. Leko, V.; Rosenberg, S.A. Identifying and Targeting Human Tumor Antigens for T Cell-Based Immunotherapy of Solid Tumors. *Cancer Cell* **2020**, *38*, 454–472. [CrossRef] [PubMed]
23. Raskov, H.; Orhan, A.; Christensen, J.P.; Gögenur, I. Cytotoxic $CD8^+$ T cells in cancer and cancer immunotherapy. *Br. J. Cancer* **2021**, *124*, 359–367. [CrossRef] [PubMed]
24. Kartikasari, A.E.R.; Prakash, M.D.; Cox, M.; Wilson, K.; Boer, J.C.; Cauchi, J.A.; Plebanski, M. Therapeutic Cancer Vaccines-T Cell Responses and Epigenetic Modulation. *Front. Immunol.* **2019**, *9*, 3109. [CrossRef] [PubMed]
25. Liu, J.; Fu, M.; Wang, M.; Wan, D.; Wei, Y.; Wei, X. Cancer vaccines as promising immuno-therapeutics: Platforms and current progress. *J. Hematol. Oncol.* **2022**, *15*, 28. [CrossRef] [PubMed]
26. Fan, T.; Zhang, M.; Yang, J.; Zhu, Z.; Cao, W.; Dong, C. Therapeutic cancer vaccines: Advancements, challenges, and prospects. *Signal Transduct. Target. Ther.* **2023**, *8*, 450. [CrossRef] [PubMed]
27. Zhang, X.; Cui, H.; Zhang, W.; Li, Z.; Gao, J. Engineered tumor cell-derived vaccines against cancer: The art of combating poison with poison. *Bioact. Mater.* **2022**, *22*, 491–517. [CrossRef] [PubMed]
28. Wang, D.R.; Wu, X.L.; Sun, Y.L. Therapeutic targets and biomarkers of tumor immunotherapy: Response versus non-response. *Signal Transduct. Target. Ther.* **2022**, *7*, 331. [CrossRef] [PubMed]
29. Chen, B.; Khodadoust, M.S.; Olsson, N.; Wagar, L.E.; Fast, E.; Liu, C.L.; Muftuoglu, Y.; Sworder, B.J.; Diehn, M.; Levy, R.; et al. Predicting HLA class II antigen presentation through integrated deep learning. *Nat. Biotechnol.* **2019**, *37*, 1332–1343. [CrossRef] [PubMed]
30. Wendorff, M.; Garcia Alvarez, H.M.; Østerbye, T.; ElAbd, H.; Rosati, E.; Degenhardt, F.; Buus, S.; Franke, A.; Nielsen, M. Unbiased Characterization of Peptide-HLA Class II Interactions Based on Large-Scale Peptide Microarrays; Assessment of the Impact on HLA Class II Ligand and Epitope Prediction. *Front. Immunol.* **2020**, *11*, 1705. [CrossRef] [PubMed]
31. Shah, K.; Al-Haidari, A.; Sun, J.; Kazi, J.U. T cell receptor (TCR) signaling in health and disease. *Signal Transduct. Target. Ther.* **2021**, *6*, 412. [CrossRef] [PubMed]
32. Mørch, A.M.; Bálint, Š.; Santos, A.M.; Davis, S.J.; Dustin, M.L. Coreceptors and TCR Signaling—The Strong and the Weak of It. *Front. Cell Dev. Biol.* **2020**, *8*, 597627. [CrossRef] [PubMed]

33. Colbert, J.D.; Cruz, F.M.; Rock, K.L. Cross-presentation of exogenous antigens on MHC I molecules. *Curr. Opin. Immunol.* **2020**, *64*, 1–8. [CrossRef] [PubMed]
34. Weigelin, B.; Friedl, P. T cell-mediated additive cytotoxicity—Death by multiple bullets. *Trends Cancer* **2022**, *8*, 980–987. [CrossRef] [PubMed]
35. Del Prete, A.; Salvi, V.; Soriani, A.; Laffranchi, M.; Sozio, F.; Bosisio, D.; Sozzani, S. Dendritic cell subsets in cancer immunity and tumor antigen sensing. *Cell. Mol. Immunol.* **2023**, *20*, 432–447. [CrossRef] [PubMed]
36. Marciscano, A.E.; Anandasabapathy, N. The role of dendritic cells in cancer and anti-tumor immunity. *Semin. Immunol.* **2021**, *52*, 101481. [CrossRef] [PubMed]
37. Wylie, B.; Macri, C.; Mintern, J.D.; Waithman, J. Dendritic Cells and Cancer: From Biology to Therapeutic Intervention. *Cancers* **2019**, *11*, 521. [CrossRef] [PubMed]
38. Zhao, X.; Zhang, Z.; Wen, C.; Huang, J.; Yang, S.; Liu, J.; Geng, H.; Peng, B.; Li, Z.; Zhang, Y. The safety and anti-tumor effect of multiple peptides-pulsed dendritic cells combined with induced specific cytotoxic T lymphocytes for patients with solid tumors. *Front. Immunol.* **2023**, *14*, 1284334. [CrossRef] [PubMed]
39. Kang, T.H.; Yang, A.; Tsai, Y.; Ferrall, L.; Hung, C.F. Targeted tumor coating with antigenic, CTL-recognizable peptides via Annexin A5 chimeric constructs following chemotherapy redirects adaptive $CD8^+$ T cell immunity for tumor clearance. *Cell. Mol. Immunol.* **2021**, *18*, 1578–1580. [CrossRef] [PubMed]
40. Rao, M.S.; Van Vleet, T.R.; Ciurlionis, R.; Buck, W.R.; Mittelstadt, S.W.; Blomme, E.A.G.; Liguori, M.J. Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. *Front. Genet.* **2019**, *9*, 636. [CrossRef] [PubMed]
41. Buonaguro, L.; Tagliamonte, M. Peptide-based vaccine for cancer therapies. *Front. Immunol.* **2023**, *14*, 1210044. [CrossRef] [PubMed]
42. Biswas, N.; Chakrabarti, S.; Padul, V.; Jones, L.D.; Ashili, S. Designing neoantigen cancer vaccines, trials, and outcomes. *Front. Immunol.* **2023**, *14*, 1105420. [CrossRef] [PubMed]
43. Zhang, B.; Bassani-Sternberg, M. Current perspectives on mass spectrometry-based immunopeptidomics: The computational angle to tumor antigen discovery. *J. Immunother. Cancer* **2023**, *11*, e007073. [CrossRef] [PubMed]
44. Becker, J.P.; Riemer, A.B. The Importance of Being Presented: Target Validation by Immunopeptidomics for Epitope-Specific Immunotherapies. *Front. Immunol.* **2022**, *13*, 883989. [CrossRef] [PubMed]
45. Sönmez, M.G.; Sönmez, L.Ö. New treatment modalities with vaccine therapy in renal cell carcinoma. *Urol. Ann.* **2019**, *11*, 119–125. [CrossRef] [PubMed]
46. Xu, H.; Zheng, X.; Zhang, S.; Yi, X.; Zhang, T.; Wei, Q.; Li, H.; Ai, J. Tumor antigens and immune subtypes guided mRNA vaccine development for kidney renal clear cell carcinoma. *Mol. Cancer* **2021**, *20*, 159. [CrossRef] [PubMed]
47. Hu, J.; Yuan, Z.; Jiang, Y.; Mo, Z. Identification of Five Tumor Antigens for Development and Two Immune Subtypes for Personalized Medicine of mRNA Vaccines in Papillary Renal Cell Carcinoma. *J. Pers. Med.* **2023**, *13*, 359. [CrossRef] [PubMed]
48. Xu, Y.; Miller, C.P.; Warren, E.H.; Tykodi, S.S. Current status of antigen-specific T-cell immunotherapy for advanced renal-cell carcinoma. *Hum. Vaccin. Immunother.* **2021**, *17*, 1882–1896. [CrossRef] [PubMed]
49. Abd-Aziz, N.; Poh, C.L. Development of Peptide-Based Vaccines for Cancer. *J. Oncol.* **2022**, *2022*, 9749363. [CrossRef] [PubMed]
50. Mizukoshi, E.; Nakagawa, H.; Tamai, T.; Kitahara, M.; Fushimi, K.; Nio, K.; Terashima, T.; Iida, N.; Arai, K.; Yamashita, T.; et al. Peptide vaccine-treated, long-term surviving cancer patients harbor self-renewing tumor-specific $CD8^+$ T cells. *Nat. Commun.* **2022**, *13*, 3123. [CrossRef] [PubMed]
51. Stephens, A.J.; Burgess-Brown, N.A.; Jiang, S. Beyond Just Peptide Antigens: The Complex World of Peptide-Based Cancer Vaccines. *Front. Immunol.* **2021**, *12*, 696791. [CrossRef] [PubMed]
52. Hopkins, A.M.; Kichenadasse, G.; Karapetis, C.S.; Rowland, A.; Sorich, M.J. Early tumor shrinkage identifies long-term disease control and survival in patients with lung cancer treated with atezolizumab. *J. Immunother. Cancer* **2020**, *8*, e000500. [CrossRef] [PubMed]
53. Yang, X.; Xian, X.; Wang, Y.; Qiu, M. Assessing prognostic value of early tumor shrinkage and depth of response in first-line therapy for patients with advanced unresectable pancreatic cancer. *BMC Gastroenterol.* **2021**, *21*, 294. [CrossRef] [PubMed]
54. Basu, A.; Ramamoorthi, G.; Albert, G.; Gallen, C.; Beyer, A.; Snyder, C.; Koski, G.; Disis, M.L.; Czerniecki, B.J.; Kodumudi, K. Differentiation and Regulation of T_H Cells: A Balancing Act for Cancer Immunotherapy. *Front. Immunol.* **2021**, *12*, 669474. [CrossRef] [PubMed]
55. Lee, J.; Lozano-Ruiz, B.; Yang, F.M.; Fan, D.D.; Shen, L.; González-Navajas, J.M. The Multifaceted Role of Th1, Th9, and Th17 Cells in Immune Checkpoint Inhibition Therapy. *Front. Immunol.* **2021**, *12*, 625667. [CrossRef] [PubMed]
56. Kumai, T.; Yamaki, H.; Kono, M.; Hayashi, R.; Wakisaka, R.; Komatsuda, H. Antitumor Peptide-Based Vaccine in the Limelight. *Vaccines* **2022**, *10*, 70. [CrossRef] [PubMed]
57. Chu, Y.; Zhang, Y.; Wang, Q.; Zhang, L.; Wang, X.; Wang, Y.; Salahub, D.R.; Xu, Q.; Wamg, J.; Jiang, X.; et al. A transformer-based model to predict peptide–HLA class I binding and optimize mutated peptides for vaccine design. *Nat. Mach. Intell.* **2022**, *4*, 300–311. [CrossRef]
58. Karnaukhov, V.; Paes, W.; Woodhouse, I.B.; Partridge, T.; Nicastri, A.; Brackenridge, S.; Scherbinin, D.; Chudakov, D.M.; Zvyagin, I.V.; Ternette, N.; et al. HLA binding of self-peptides is biased towards proteins with specific molecular functions. *bioRxiv* **2021**, Preprint. [CrossRef] [PubMed]

59. Embgenbroich, M.; Burgdorf, S. Current Concepts of Antigen Cross-Presentation. *Front. Immunol.* **2018**, *9*, 1643. [CrossRef] [PubMed]
60. Xie, M.; Liu, D.; Yang, Y. Anti-cancer peptides: Classification, mechanism of action, reconstruction and modification. *Open Biol.* **2020**, *10*, 200004. [CrossRef] [PubMed]
61. Karami Fath, M.; Babakhaniyan, K.; Zokaei, M.; Yaghoubian, A.; Akbari, S.; Khorsandi, M.; Soofi, A.; Nabi-Afjadi, M.; Zalpoor, H.; Jalalifar, F.; et al. Anti-cancer peptide-based therapeutic strategies in solid tumors. *Cell. Mol. Biol. Lett.* **2022**, *27*, 33. [CrossRef] [PubMed]
62. Liu, Y.; Yan, X.; Zhang, F.; Zhang, X.; Tang, F.; Han, Z.; Li, Y. TCR-T Immunotherapy: The Challenges and Solutions. *Front. Oncol.* **2022**, *11*, 794183. [CrossRef] [PubMed]
63. Chandran, S.S.; Klebanoff, C.A. T cell receptor-based cancer immunotherapy: Emerging efficacy and pathways of resistance. *Immunol. Rev.* **2019**, *290*, 127–147. [CrossRef] [PubMed]
64. Zhang, Y.; Liu, Z.; Wei, W.; Li, Y. TCR engineered T cells for solid tumor immunotherapy. *Exp. Hematol. Oncol.* **2022**, *11*, 38. [CrossRef] [PubMed]
65. Shafer, P.; Kelly, L.M.; Hoyos, V. Cancer Therapy With TCR-Engineered T Cells: Current Strategies, Challenges, and Prospects. *Front. Immunol.* **2022**, *13*, 835762. [CrossRef] [PubMed]
66. Li, D.; Li, X.; Zhou, W.L.; Huang, Y.; Liang, X.; Jiang, L.; Yang, X.; Sun, J.; Li, Z.; Han, W.D.; et al. Genetically engineered T cells for cancer immunotherapy. *Signal Transduct. Target. Ther.* **2019**, *4*, 35. [CrossRef] [PubMed]
67. Hazini, A.; Fisher, K.; Seymour, L. Deregulation of HLA-I in cancer and its central importance for immunotherapy. *J. Immunother. Cancer* **2021**, *9*, e002899. [CrossRef] [PubMed]
68. Liu, W.; Tang, H.; Li, L.; Wang, X.; Yu, Z.; Li, J. Peptide-based therapeutic cancer vaccine: Current trends in clinical application. *Cell Prolif.* **2021**, *54*, e13025. [CrossRef] [PubMed]
69. Xie, N.; Shen, G.; Gao, W.; Huang, Z.; Huang, C.; Fu, L. Neoantigens: Promising targets for cancer therapy. *Signal Transduct. Target. Ther.* **2023**, *8*, 9. [CrossRef] [PubMed]
70. Dhatchinamoorthy, K.; Colbert, J.D.; Rock, K.L. Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. *Front. Immunol.* **2021**, *12*, 636568. [CrossRef] [PubMed]
71. Wang, M.; Jiang, H.; Liu, X.; Wang, X. Biophysics involved in the process of tumor immune escape. *iScience* **2022**, *25*, 104124. [CrossRef] [PubMed]
72. Dutta, S.; Ganguly, A.; Chatterjee, K.; Spada, S.; Mukherjee, S. Targets of Immune Escape Mechanisms in Cancer: Basis for Development and Evolution of Cancer Immune Checkpoint Inhibitors. *Biology* **2023**, *12*, 218. [CrossRef] [PubMed]
73. Garrido, F.; Aptsiauri, N. Cancer immune escape: MHC expression in primary tumours versus metastases. *Immunology* **2019**, *158*, 255–266. [CrossRef] [PubMed]
74. Sabbatino, F.; Liguori, L.; Polcaro, G.; Salvato, I.; Caramori, G.; Salzano, F.A.; Casolaro, V.; Stellato, C.; Col, J.D.; Pepe, S. Role of Human Leukocyte Antigen System as A Predictive Biomarker for Checkpoint-Based Immunotherapy in Cancer Patients. *Int. J. Mol. Sci.* **2020**, *21*, 7295. [CrossRef] [PubMed]
75. Jin, X.; Ding, Y.; Sun, S.; Wang, X.; Zhou, Z.; Liu, X.; Li, M.; Chen, X.; Shen, A.; Wu, Y.; et al. Screening HLA-A-restricted T cell epitopes of SARS-CoV-2 and the induction of $CD8^+$ T cell responses in HLA-A transgenic mice. *Cell. Mol. Immunol.* **2021**, *18*, 2588–2608. [CrossRef] [PubMed]
76. Calmeiro, J.; Carrascal, M.A.; Tavares, A.R.; Ferreira, D.A.; Gomes, C.; Falcão, A.; Cruz, M.T.; Neves, B.M. Dendritic Cell Vaccines for Cancer Immunotherapy: The Role of Human Conventional Type 1 Dendritic Cells. *Pharmaceutics* **2020**, *12*, 158. [CrossRef] [PubMed]
77. Yu, J.; Sun, H.; Cao, W.; Song, Y.; Jiang, Z. Research progress on dendritic cell vaccines in cancer immunotherapy. *Exp. Hematol. Oncol.* **2022**, *11*, 3. [CrossRef] [PubMed]
78. Firdessa-Fite, R.; Creusot, R.J. Nanoparticles versus Dendritic Cells as Vehicles to Deliver mRNA Encoding Multiple Epitopes for Immunotherapy. *Mol. Ther. Methods Clin. Dev.* **2019**, *16*, 50–62. [CrossRef] [PubMed]
79. de Mey, W.; Locy, H.; De Ridder, K.; De Schrijver, P.; Autaers, D.; Lakdimi, A.; Esprit, A.; Franceschini, L.; Thielemans, K.; Verdonck, M.; et al. An mRNA mix redirects dendritic cells towards an antiviral program, inducing anticancer cytotoxic stem cell and central memory $CD8^+$ T cells. *Front. Immunol.* **2023**, *14*, 1111523. [CrossRef] [PubMed]
80. Ahmed, R.; Sayegh, N.; Graciotti, M.; Kandalaft, L.E. Electroporation as a method of choice to generate genetically modified dendritic cell cancer vaccines. *Curr. Opin. Biotechnol.* **2020**, *65*, 142–155. [CrossRef] [PubMed]
81. Perez, C.R.; De Palma, M. Engineering dendritic cell vaccines to improve cancer immunotherapy. *Nat. Commun.* **2019**, *10*, 5408. [CrossRef] [PubMed]
82. Najafi, S.; Mortezaee, K. Advances in dendritic cell vaccination therapy of cancer. *Biomed. Pharmacother.* **2023**, *164*, 114954. [CrossRef] [PubMed]
83. Kumbhari, A.; Egelston, C.A.; Lee, P.P.; Kim, P.S. Mature Dendritic Cells May Promote High-Avidity Tuning of Vaccine T Cell Responses. *Front. Immunol.* **2020**, *11*, 584680. [CrossRef] [PubMed]
84. Nguyen, A.T.; Szeto, C.; Gras, S. The pockets guide to HLA class I molecules. *Biochem. Soc. Trans.* **2021**, *49*, 2319–2331. [CrossRef] [PubMed]

85. Nagler, A.; Kalaora, S.; Barbolin, C.; Gangaev, A.; Ketelaars, S.L.C.; Alon, M.; Pai, J.; Benedek, G.; Yahalom-Ronen, Y.; Erez, N.; et al. Identification of presented SARS-CoV-2 HLA class I and HLA class II peptides using HLA peptidomics. *Cell Rep.* **2021**, *35*, 109305. [CrossRef] [PubMed]
86. Ruangapirom, L.; Sutivijit, N.; Teerapakpinyo, C.; Mutirangura, A.; Doungkamchan, C. Identification of Shared Neoantigens in *BRCA1*-Related Breast Cancer. *Vaccines* **2022**, *10*, 1597. [CrossRef] [PubMed]
87. Okada, M.; Shimizu, K.; Fujii, S.I. Identification of Neoantigens in Cancer Cells as Targets for Immunotherapy. *Int. J. Mol. Sci.* **2022**, *23*, 2594. [CrossRef] [PubMed]
88. Ho, S.Y.; Chang, C.M.; Liao, H.N.; Chou, W.H.; Guo, C.L.; Yen, Y.; Nakamura, Y.; Chang, W.C. Current Trends in Neoantigen-Based Cancer Vaccines. *Pharmaceuticals* **2023**, *16*, 392. [CrossRef] [PubMed]
89. Benlaribi, R.; Gou, Q.; Takaba, H. Thymic self-antigen expression for immune tolerance and surveillance. *Inflamm. Regen.* **2022**, *42*, 28. [CrossRef] [PubMed]
90. Chandran, S.S.; Ma, J.; Klatt, M.G.; Dündar, F.; Bandlamudi, C.; Razavi, P.; Wen, H.Y.; Weigelt, B.; Zumbo, P.; Fu, S.N.; et al. Immunogenicity and therapeutic targeting of a public neoantigen derived from mutated PIK3CA. *Nat. Med.* **2022**, *28*, 946–957. [CrossRef] [PubMed]
91. Claeys, A.; Luijts, T.; Marchal, K.; Van den Eynden, J. Low immunogenicity of common cancer hot spot mutations resulting in false immunogenic selection signals. *PLoS Genet.* **2021**, *17*, e1009368. [CrossRef] [PubMed]
92. Capietto, A.H.; Jhunjhunwala, S.; Delamarre, L. Cancer neoantigens and immunogenicity: Mutation position matters. *Mol. Cell. Oncol.* **2020**, *7*, 1740071. [CrossRef] [PubMed]
93. Lang, F.; Schrörs, B.; Löwer, M.; Türeci, Ö.; Sahin, U. Identification of neoantigens for individualized therapeutic cancer vaccines. *Nat. Rev. Drug Discov.* **2022**, *21*, 261–282. [CrossRef] [PubMed]
94. Yu, G.; He, X.; Li, X.; Wu, Y. Driving neoantigen-based cancer vaccines for personalized immunotherapy into clinic: A burdensome journey to promising land. *Biomed. Pharmacother.* **2022**, *153*, 113464. [CrossRef] [PubMed]
95. Kumar, S.; Warrell, J.; Li, S.; McGillivray, P.D.; Meyerson, W.; Salichos, L.; Harmanci, A.; Martinez-Fundichely, A.; Chan, C.W.Y.; Nielsen, M.M.; et al. Passenger Mutations in More Than 2500 Cancer Genomes: Overall Molecular Functional Impact and Consequences. *Cell* **2020**, *180*, 915–927.e16. [CrossRef] [PubMed]
96. Grodzka, A.; Knopik-Skrocka, A.; Kowalska, K.; Kurzawa, P.; Krzyzaniak, M.; Stencel, K.; Bryl, M. Molecular alterations of driver genes in non-small cell lung cancer: From diagnostics to targeted therapy. *EXCLI J.* **2023**, *22*, 415–432. [CrossRef] [PubMed]
97. Min, H.Y.; Lee, H.Y. Molecular targeted therapy for anticancer treatment. *Exp. Mol. Med.* **2022**, *54*, 1670–1694. [CrossRef] [PubMed]
98. Ye, L.; Creaney, J.; Redwood, A.; Robinson, B. The Current Lung Cancer Neoantigen Landscape and Implications for Therapy. *J. Thorac. Oncol.* **2021**, *16*, 922–932. [CrossRef] [PubMed]
99. Rosenthal, R.; Cadieux, E.L.; Salgado, R.; Bakir, M.A.; Moore, D.A.; Hiley, C.T.; Lund, T.; Tanić, M.; Reading, J.L.; Joshi, K.; et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature* **2019**, *567*, 479–485. [CrossRef] [PubMed]
100. Anichini, A.; Perotti, V.E.; Sgambelluri, F.; Mortarini, R. Immune Escape Mechanisms in Non Small Cell Lung Cancer. *Cancers* **2020**, *12*, 3605. [CrossRef] [PubMed]
101. Kherreh, N.; Cleary, S.; Seoighe, C. No evidence that HLA genotype influences the driver mutations that occur in cancer patients. *Cancer Immunol. Immunother.* **2022**, *71*, 819–827. [CrossRef] [PubMed]
102. Filip, I.; Wang, A.; Kravets, O.; Orenbuch, R.; Zhao, J.; Perea-Chamblee, T.E.; Manji, G.A.; López de Maturana, E.; Malats, N.; Olive, K.P.; et al. Pervasiveness of HLA allele-specific expression loss across tumor types. *Genome Med.* **2023**, *15*, 8. [CrossRef] [PubMed]
103. Fangazio, M.; Ladewig, E.; Gomez, K.; Garcia-Ibanez, L.; Kumar, R.; Teruya-Feldstein, J.; Rossi, D.; Filip, I.; Pan-Hammarström, Q.; Inghirami, G.; et al. Genetic mechanisms of HLA-I loss and immune escape in diffuse large B cell lymphoma. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2104504118. [CrossRef] [PubMed]
104. Martínez-Jiménez, F.; Priestley, P.; Shale, C.; Baber, J.; Rozemuller, E.; Cuppen, E. Genetic immune escape landscape in primary and metastatic cancer. *Nat. Genet.* **2023**, *55*, 820–831. [CrossRef] [PubMed]
105. 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] [PubMed]
106. Sim, M.J.W.; Sun, P.D. T Cell Recognition of Tumor Neoantigens and Insights Into T Cell Immunotherapy. *Front. Immunol.* **2022**, *13*, 833017. [CrossRef] [PubMed]
107. Zhou, W.; Yu, J.; Li, Y.; Wang, K. Neoantigen-specific TCR-T cell-based immunotherapy for acute myeloid leukemia. *Exp. Hematol. Oncol.* **2022**, *11*, 100. [CrossRef] [PubMed]
108. Gurung, H.R.; Heidersbach, A.J.; Darwish, M.; Chan, P.P.F.; Li, J.; Beresini, M.; Zill, O.A.; Wallace, A.; Tong, A.J.; Hascall, D.; et al. Systematic discovery of neoepitope-HLA pairs for neoantigens shared among patients and tumor types. *Nat. Biotechnol.* 2023, *in press*. [CrossRef] [PubMed]
109. Cattaneo, C.M.; Battaglia, T.; Urbanus, J.; Moravec, Z.; Voogd, R.; de Groot, R.; Hartemink, K.J.; Haanen, J.B.A.G.; Voest, E.E.; Schumacher, T.N.; et al. Identification of patient-specific $CD4^+$ and $CD8^+$ T cell neoantigens through HLA-unbiased genetic screens. *Nat. Biotechnol.* **2023**, *41*, 783–787. [CrossRef] [PubMed]
110. Zahavi, D.; Weiner, L. Monoclonal Antibodies in Cancer Therapy. *Antibodies* **2020**, *9*, 34. [CrossRef] [PubMed]

111. Li, S.; Simoni, Y.; Zhuang, S.; Gabel, A.; Ma, S.; Chee, J.; Islas, L.; Cessna, A.; Creaney, J.; Bradley, R.K.; et al. Characterization of neoantigen-specific T cells in cancer resistant to immune checkpoint therapies. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2025570118. [CrossRef] [PubMed]
112. Salvatori, E.; Lione, L.; Compagnone, M.; Pinto, E.; Conforti, A.; Ciliberto, G.; Aurisicchio, L.; Palombo, F. Neoantigen cancer vaccine augments anti-CTLA-4 efficacy. *NPJ Vaccines* **2022**, *7*, 15. [CrossRef] [PubMed]
113. Brightman, S.E.; Becker, A.; Thota, R.R.; Naradikian, M.S.; Chihab, L.; Zavala, K.S.; Ramamoorthy Premlal, A.L.; Griswold, R.Q.; Dolina, J.S.; Cohen, E.E.W.; et al. Neoantigen-specific stem cell memory-like $CD4^+$ T cells mediate $CD8^+$ T cell-dependent immunotherapy of MHC class II-negative solid tumors. *Nat. Immunol.* **2023**, *24*, 1345–1357. [CrossRef] [PubMed]
114. Li, J.; Xiao, Z.; Wang, D.; Jia, L.; Nie, S.; Zeng, X.; Hu, W. The screening, identification, design and clinical application of tumor-specific neoantigens for TCR-T cells. *Mol. Cancer* **2023**, *22*, 141. [CrossRef] [PubMed]
115. Liu, L.; Chen, J.; Zhang, H.; Ye, J.; Moore, C.; Lu, C.; Fang, Y.; Fu, Y.X.; Li, B. Concurrent delivery of immune checkpoint blockade modulates T cell dynamics to enhance neoantigen vaccine-generated antitumor immunity. *Nat. Cancer* **2022**, *3*, 437–452. [CrossRef] [PubMed]
116. Dolina, J.S.; Lee, J.; Brightman, S.E.; McArdle, S.; Hall, S.M.; Thota, R.R.; Zavala, K.S.; Lanka, M.; Ramamoorthy Premlal, A.L.; Greenbaum, J.A.; et al. Linked $CD4^+$/$CD8^+$ T cell neoantigen vaccination overcomes immune checkpoint blockade resistance and enables tumor regression. *J. Clin. Investig.* **2023**, *133*, e164258. [CrossRef] [PubMed]
117. Nasti, T.H.; Eberhardt, C.S. Vaccination against Cancer or Infectious Agents during Checkpoint Inhibitor Therapy. *Vaccines* **2021**, *9*, 1396. [CrossRef] [PubMed]
118. Rotte, A. Combination of CTLA-4 and PD-1 blockers for treatment of cancer. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 255. [CrossRef] [PubMed]
119. Akturk, H.K.; Couts, K.L.; Baschal, E.E.; Karakus, K.E.; Van Gulick, R.J.; Turner, J.A.; Pyle, L.; Robinson, W.A.; Michels, A.W. Analysis of Human Leukocyte Antigen DR Alleles, Immune-Related Adverse Events, and Survival Associated with Immune Checkpoint Inhibitor Use among Patients with Advanced Malignant Melanoma. *JAMA Netw. Open* **2022**, *5*, e2246400. [CrossRef] [PubMed]
120. Correale, P.; Saladino, R.E.; Giannarelli, D.; Sergi, A.; Mazzei, M.A.; Bianco, G.; Giannicola, R.; Iuliano, E.; Forte, I.M.; Calandruccio, N.D.; et al. HLA Expression Correlates to the Risk of Immune Checkpoint Inhibitor-Induced Pneumonitis. *Cells* **2020**, *9*, 1964. [CrossRef] [PubMed]
121. Puig-Saus, C.; Sennino, B.; Peng, S.; Wang, C.L.; Pan, Z.; Yuen, B.; Purandare, B.; An, D.; Quach, B.B.; Nguyen, D.; et al. Neoantigen-targeted $CD8^+$ T cell responses with PD-1 blockade therapy. *Nature* **2023**, *615*, 697–704. [CrossRef] [PubMed]
122. Niemi, J.V.L.; Sokolov, A.V.; Schiöth, H.B. Neoantigen Vaccines; Clinical Trials, Classes, Indications, Adjuvants and Combinatorial Treatments. *Cancers* **2022**, *14*, 5163. [CrossRef] [PubMed]
123. Peng, M.; Mo, Y.; Wang, Y.; Wu, P.; Zhang, Y.; Xiong, F.; Guo, C.; Wu, X.; Li, Y.; Li, X.; et al. Neoantigen vaccine: An emerging tumor immunotherapy. *Mol. Cancer* **2019**, *18*, 128. [CrossRef] [PubMed]
124. Gupta, S.L.; Basu, S.; Soni, V.; Jaiswal, R.K. Immunotherapy: An alternative promising therapeutic approach against cancers. *Mol. Biol. Rep.* **2022**, *49*, 9903–9913. [CrossRef] [PubMed]
125. Posner, A.; Sivakumaran, T.; Pattison, A.; Etemadmoghadam, D.; Thio, N.; Wood, C.; Fisher, K.; Webb, S.; DeFazio, A.; Wilcken, N.; et al. Immune and genomic biomarkers of immunotherapy response in cancer of unknown primary. *J. Immunother. Cancer* **2023**, *11*, e005809. [CrossRef] [PubMed]
126. Pender, A.; Titmuss, E.; Pleasance, E.D.; Fan, K.Y.; Pearson, H.; Brown, S.D.; Grisdale, C.J.; Topham, J.T.; Shen, Y.; Bonakdar, M.; et al. Genome and Transcriptome Biomarkers of Response to Immune Checkpoint Inhibitors in Advanced Solid Tumors. *Clin. Cancer Res.* **2021**, *27*, 202–212. [CrossRef] [PubMed]
127. Bai, R.; Lv, Z.; Xu, D.; Cui, J. Predictive biomarkers for cancer immunotherapy with immune checkpoint inhibitors. *Biomark. Res.* **2020**, *8*, 34. [CrossRef] [PubMed]
128. Muñiz-Castrillo, S.; Vogrig, A.; Honnorat, J. Associations between HLA and autoimmune neurological diseases with autoantibodies. *Auto. Immun. Highlights* **2020**, *11*, 2. [CrossRef] [PubMed]
129. Solomon, B.D.; Zheng, H.; Dillon, L.W.; Goldman, J.D.; Hourigan, C.S.; Heath, J.R.; Khatri, P. Prediction of HLA genotypes from single-cell transcriptome data. *Front. Immunol.* **2023**, *14*, 1146826. [CrossRef] [PubMed]
130. Naito, T.; Suzuki, K.; Hirata, J.; Kamatani, Y.; Matsuda, K.; Toda, T.; Okada, Y. A deep learning method for HLA imputation and trans-ethnic MHC fine-mapping of type 1 diabetes. *Nat. Commun.* **2021**, *12*, 1639. [CrossRef] [PubMed]
131. Geffard, E.; Limou, S.; Walencik, A.; Daya, M.; Watson, H.; Torgerson, D.; Barnes, K.C.; Cesbron Gautier, A.; Gourraud, P.A.; Vince, N. Easy-HLA: A validated web application suite to reveal the full details of HLA typing. *Bioinformatics* **2020**, *36*, 2157–2164. [CrossRef] [PubMed]
132. Valdés-Corona, L.F.; Hernández-Doño, S.; Rodríguez-Reyna, T.S.; García-Silva, R.; Jakez, J.; Escamilla-Tilch, M.; Lima, G.; Llorente, L.; Pineda, C.; Yunis, E.; et al. Aspartic $acid^{70}$ in the HLA-DRB1 chain and shared epitope alleles partially explain the high prevalence of autoimmunity in Mexicans. *J. Transl. Autoimmun.* **2020**, *3*, 100057. [CrossRef] [PubMed]
133. Noble, J.A.; Valdes, A.M. Genetics of the HLA region in the prediction of type 1 diabetes. *Curr. Diab Rep.* **2011**, *11*, 533–542. [CrossRef] [PubMed]
134. Hedström, A.K.; Olsson, T.; Alfredsson, L. The increased risk of multiple sclerosis associated with HLA-DRB1*15:01 and smoking is modified by alcohol consumption. *Sci. Rep.* **2021**, *11*, 21237. [CrossRef] [PubMed]

135. Kawasaki, A.; Kusumawati, P.A.; Kawamura, Y.; Kondo, Y.; Kusaoi, M.; Amano, H.; Kusanagi, Y.; Itoh, K.; Fujimoto, T.; Tamura, N.; et al. Genetic dissection of HLA-DRB1*15:01 and XL9 region variants in Japanese patients with systemic lupus erythematosus: Primary role for HLA-DRB1*15:01. *RMD Open* **2023**, *9*, e003214. [CrossRef] [PubMed]
136. Miglioranza Scavuzzi, B.; van Drongelen, V.; Kaur, B.; Fox, J.C.; Liu, J.; Mesquita-Ferrari, R.A.; Kahlenberg, J.M.; Farkash, E.A.; Benavides, F.; Miller, F.W.; et al. The lupus susceptibility allele DRB1*03:01 encodes a disease-driving epitope. *Commun. Biol.* **2022**, *5*, 751. [CrossRef] [PubMed]
137. Dashti, N.; Mahmoudi, M.; Aslani, S.; Jamshidi, A. HLA-B*27 subtypes and their implications in the pathogenesis of ankylosing spondylitis. *Gene* **2018**, *670*, 15–21. [CrossRef] [PubMed]
138. Karell, K.; Louka, A.S.; Moodie, S.J.; Ascher, H.; Clot, F.; Greco, L.; Ciclitira, P.J.; Sollid, L.M.; Partanen, J.; European Genetics Cluster on Celiac Disease. HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: Results from the European Genetics Cluster on Celiac Disease. *Hum. Immunol.* **2003**, *64*, 469–477. [CrossRef] [PubMed]
139. Dand, N.; Duckworth, M.; Baudry, D.; Russell, A.; Curtis, C.J.; Lee, S.H.; Evans, I.; Mason, K.J.; Alsharqi, A.; Becher, G.; et al. HLA-C*06:02 genotype is a predictive biomarker of biologic treatment response in psoriasis. *J. Allergy Clin. Immunol.* **2019**, *143*, 2120–2130. [CrossRef] [PubMed]
140. Goyette, P.; Boucher, G.; Mallon, D.; Ellinghaus, E.; Jostins, L.; Huang, H.; Ripke, S.; Gusareva, E.S.; Annese, V.; Hauser, S.L.; et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat. Genet.* **2015**, *47*, 172–179. [CrossRef] [PubMed]
141. Mahdi, B.M. Role of HLA typing on Crohn's disease pathogenesis. *Ann. Med. Surg.* **2015**, *4*, 248–253. [CrossRef] [PubMed]
142. Degenhardt, F.; Mayr, G.; Wendorff, M.; Boucher, G.; Ellinghaus, E.; Ellinghaus, D.; ElAbd, H.; Rosati, E.; Hübenthal, M.; Juzenas, S.; et al. Transethnic analysis of the human leukocyte antigen region for ulcerative colitis reveals not only shared but also ethnicity-specific disease associations. *Hum. Mol. Genet.* **2021**, *30*, 356–369. [CrossRef] [PubMed]
143. Asquith, M.; Sternes, P.R.; Costello, M.E.; Karstens, L.; Diamond, S.; Martin, T.M.; Li, Z.; Marshall, M.S.; Spector, T.D.; le Cao, K.A.; et al. HLA Alleles Associated with Risk of Ankylosing Spondylitis and Rheumatoid Arthritis Influence the Gut Microbiome. *Arthritis Rheumatol.* **2019**, *71*, 1642–1650. [CrossRef] [PubMed]
144. Inoue, M.; Nagafuchi, Y.; Ota, M.; Tsuchiya, H.; Tateishi, S.; Kanda, H.; Fujio, K. Carriers of HLA-DRB1*04:05 have a better clinical response to abatacept in rheumatoid arthritis. *Sci. Rep.* **2023**, *13*, 15250. [CrossRef] [PubMed]
145. Al Naqbi, H.; Mawart, A.; Alshamsi, J.; Al Safar, H.; Tay, G.K. Major histocompatibility complex (MHC) associations with diseases in ethnic groups of the Arabian Peninsula. *Immunogenetics* **2021**, *73*, 131–152. [CrossRef] [PubMed]
146. Tay, G.K.; Al Naqbi, H.; Mawart, A.; Baalfaqih, Z.; Almaazmi, A.; Deeb, A.; Alsafar, H. Segregation Analysis of Genotyped and Family-Phased, Long Range MHC Classical Class I and Class II Haplotypes in 5 Families With Type 1 Diabetes Proband in the United Arab Emirates. *Front. Genet.* **2021**, *12*, 670844. [CrossRef] [PubMed]
147. Ali, A.A.; Khalid, K.E.; Mohammed, S.E.; Akhtar, M.S.; Saeed, O.K. Association of Human Leukocyte Antigen (HLA) class II (DRB1 and DQB1) alleles and haplotypes with Rheumatoid Arthritis in Sudanese patients. *Front. Immunol.* **2023**, *14*, 1178546. [CrossRef] [PubMed]
148. van Gerven, N.M.; de Boer, Y.S.; Zwiers, A.; Verwer, B.J.; Drenth, J.P.; van Hoek, B.; van Erpecum, K.J.; Beuers, U.; van Buuren, H.R.; den Ouden, J.W.; et al. HLA-DRB1*03:01 and HLA-DRB1*04:01 modify the presentation and outcome in autoimmune hepatitis type-1. *Genes Immun.* **2015**, *16*, 247–252. [CrossRef] [PubMed]
149. Barber, M.R.W.; Drenkard, C.; Falasinnu, T.; Hoi, A.; Mak, A.; Kow, N.Y.; Svenungsson, E.; Peterson, J.; Clarke, A.E.; Ramsey-Goldman, R. Global epidemiology of systemic lupus erythematosus. *Nat. Rev. Rheumatol.* **2021**, *17*, 515–532. [CrossRef] [PubMed]
150. Kigo, S.N.; Omondi, E.O.; Omolo, B.O. Assessing predictive performance of supervised machine learning algorithms for a diamond pricing model. *Sci. Rep.* **2023**, *13*, 17315. [CrossRef] [PubMed]
151. Chatterjee, S.; Chaudhuri, R.; Vrontis, D.; Papadopoulos, T. Examining the impact of deep learning technology capability on manufacturing firms: Moderating roles of technology turbulence and top management support. *Ann. Oper. Res.* **2022**, *30*, 1–21. [CrossRef] [PubMed]
152. Sanderson, T.; Bileschi, M.L.; Belanger, D.; Colwell, L.J. ProteInfer, deep neural networks for protein functional inference. *Elife.* **2023**, *12*, e80942. [CrossRef] [PubMed]
153. Luo, Y.; Jiang, G.; Yu, T.; Liu, Y.; Vo, L.; Ding, H.; Su, Y.; Qian, W.W.; Zhao, H.; Peng, J. ECNet is an evolutionary context-integrated deep learning framework for protein engineering. *Nat. Commun.* **2021**, *12*, 5743. [CrossRef] [PubMed]
154. Ko, C.W.; Huh, J.; Park, J.W. Deep learning program to predict protein functions based on sequence information. *MethodsX* **2022**, *9*, 101622. [CrossRef] [PubMed]
155. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [CrossRef] [PubMed]
156. Yang, X.; Zhao, L.; Wei, F.; Li, J. DeepNetBim: Deep learning model for predicting HLA-epitope interactions based on network analysis by harnessing binding and immunogenicity information. *BMC Bioinform.* **2021**, *22*, 231. [CrossRef] [PubMed]
157. Debebe, B.J.; Boelen, L.; Lee, J.C.; IAVI Protocol C Investigators; Thio, C.L.; Astemborski, J.; Kirk, G.; Khakoo, S.I.; Donfield, S.M.; Goedert, J.J.; et al. Identifying the immune interactions underlying HLA class I disease associations. *Elife* **2020**, *9*, e54558. [CrossRef] [PubMed]

158. Yi, J.; Chen, L.; Xiao, Y.; Zhao, Z.; Su, X. Investigations of sequencing data and sample type on HLA class Ia typing with different computational tools. *Brief. Bioinform.* **2021**, *22*, bbaa143. [CrossRef] [PubMed]
159. Wang, Y.Y.; Mimori, T.; Khor, S.S.; Gervais, O.; Kawai, Y.; Hitomi, Y.; Tokunaga, K.; Nagasaki, M. HLA-VBSeq v2: Improved HLA calling accuracy with full-length Japanese class-I panel. *Hum. Genome Var.* **2019**, *6*, 29. [CrossRef] [PubMed]
160. Qu, H.; Zheng, C.; Ji, H.; Huang, R.; Wei, D.; Annis, S.; Drummond, F. A deep multi-task learning approach to identifying mummy berry infection sites, the disease stage, and severity. *Front. Plant Sci.* **2024**, *15*, 1340884. [CrossRef] [PubMed]
161. Sobotka, D.; Herold, A.; Perkonigg, M.; Beer, L.; Bastati, N.; Sablatnig, A.; Ba-Ssalamah, A.; Langs, G. Improving Vessel Segmentation with Multi-Task Learning and Auxiliary Data Available Only during Model Training. *Comput. Med. Imaging Graph.* **2024**, *114*, 102369. [CrossRef] [PubMed]
162. Berg, M.; Petoukhov, I.; van den Ende, I.; Meyer, K.B.; Guryev, V.; Vonk, J.M.; Carpaij, O.; Banchero, M.; Hendriks, R.W.; van den Berge, M.; et al. FastCAR: Fast correction for ambient RNA to facilitate differential gene expression analysis in single-cell RNA-sequencing datasets. *BMC Genom.* **2023**, *24*, 722. [CrossRef] [PubMed]
163. Kwok, A.J.; Mentzer, A.; Knight, J.C. Host genetics and infectious disease: New tools, insights and translational opportunities. *Nat. Rev. Genet.* **2021**, *22*, 137–153. [CrossRef] [PubMed]
164. Yin, S.; Fu, C.; Zhao, S.; Li, K.; Sun, X.; Xu, T.; Chen, E. A Survey on Multimodal Large Language Models. *arXiv* **2023**, arXiv:2306.13549. [CrossRef]
165. Qureshi, M.N.I.; Oh, J.; Cho, D.; Jo, H.J.; Lee, B. Multimodal Discrimination of Schizophrenia Using Hybrid Weighted Feature Concatenation of Brain Functional Connectivity and Anatomical Features with an Extreme Learning Machine. *Front. Neuroinform.* **2017**, *11*, 59. [CrossRef] [PubMed]
166. Steinberg, E.; Jung, K.; Fries, J.A.; Corbin, C.K.; Pfohl, S.R.; Shah, N.H. Language models are an effective representation learning technique for electronic health record data. *J. Biomed. Inform.* **2021**, *113*, 103637. [CrossRef] [PubMed]
167. Ofer, D.; Brandes, N.; Linial, M. The language of proteins: NLP, machine learning & protein sequences. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1750–1758. [CrossRef] [PubMed]
168. Ferruz, N.; Schmidt, S.; Höcker, B. ProtGPT2 is a deep unsupervised language model for protein design. *Nat. Commun.* **2022**, *13*, 4348. [CrossRef] [PubMed]
169. Valentini, G.; Malchiodi, D.; Gliozzo, J.; Mesiti, M.; Soto-Gomez, M.; Cabri, A.; Reese, J.; Casiraghi, E.; Robinson, P.N. The promises of large language models for protein design and modeling. *Front. Bioinform.* **2023**, *3*, 1304099. [CrossRef] [PubMed]
170. Jha, K.; Karmakar, S.; Saha, S. Graph-BERT and language model-based framework for protein-protein interaction identification. *Sci. Rep.* **2023**, *13*, 5663. [CrossRef] [PubMed]
171. Brandes, N.; Ofer, D.; Peleg, Y.; Rappoport, N.; Linial, M. ProteinBERT: A universal deep-learning model of protein sequence and function. *Bioinformatics* **2022**, *38*, 2102–2110. [CrossRef] [PubMed]
172. Lv, Z.; Li, M.; Wang, Y.; Zou, Q. Editorial: Machine learning for biological sequence analysis. *Front. Genet.* **2023**, *14*, 1150688. [CrossRef] [PubMed]
173. Detlefsen, N.S.; Hauberg, S.; Boomsma, W. Learning meaningful representations of protein sequences. *Nat. Commun.* **2022**, *13*, 1914. [CrossRef] [PubMed]
174. Ruan, X.; Zhou, D.; Nie, R.; Guo, Y. Predictions of Apoptosis Proteins by Integrating Different Features Based on Improving Pseudo-Position-Specific Scoring Matrix. *Biomed. Res. Int.* **2020**, *2020*, 4071508. [CrossRef] [PubMed]
175. Li, Z.; Nie, R.; You, Z.; Cao, C.; Li, J. Using discriminative vector machine model with 2DPCA to predict interactions among proteins. *BMC Bioinform.* **2019**, *20*, 694. [CrossRef] [PubMed]
176. Sgarbossa, D.; Lupo, U.; Bitbol, A.F. Generative power of a protein language model trained on multiple sequence alignments. *Elife.* **2023**, *12*, e79854. [CrossRef] [PubMed]
177. Madani, A.; Krause, B.; Greene, E.R.; Subramanian, S.; Mohr, B.P.; Holton, J.M.; Olmos, J.L., Jr.; Xiong, C.; Sun, Z.Z.; Socher, R.; et al. Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **2023**, *41*, 1099–1106. [CrossRef] [PubMed]
178. Wu, F.; Wu, L.; Radev, D.; Xu, J.; Li, S.Z. Integration of pre-trained protein language models into geometric deep learning networks. *Commun. Biol.* **2023**, *6*, 876. [CrossRef] [PubMed]
179. McWhite, C.D.; Armour-Garb, I.; Singh, M. Leveraging protein language models for accurate multiple sequence alignments. *Genome Res.* **2023**, *33*, 1145–1153. [CrossRef] [PubMed]
180. Kang, H.; Goo, S.; Lee, H.; Chae, J.W.; Yun, H.Y.; Jung, S. Fine-tuning of BERT Model to Accurately Predict Drug-Target Interactions. *Pharmaceutics* **2022**, *14*, 1710. [CrossRef] [PubMed]
181. Jin, Y.; Yang, Y. ProtPlat: An efficient pre-training platform for protein classification based on FastText. *BMC Bioinform.* **2022**, *23*, 66. [CrossRef] [PubMed]
182. Uddin, M.R.; Mahbub, S.; Rahman, M.S.; Bayzid, M.S. SAINT: Self-attention augmented inception-inside-inception network improves protein secondary structure prediction. *Bioinformatics* **2020**, *36*, 4599–4608. [CrossRef] [PubMed]
183. Wiemers, E.A.; Redick, T.S. The influence of thought probes on performance: Does the mind wander more if you ask it? *Psychon. Bull. Rev.* **2019**, *26*, 367–373. [CrossRef] [PubMed]
184. Robison, M.K.; Miller, A.L.; Unsworth, L. Examining the effects of probe frequency, response options, and framing within the thought-probe method. *Behav. Res. Methods.* **2019**, *51*, 398–408. [CrossRef] [PubMed]

185. Gu, J.; Zhang, T.; Wu, C.; Liang, Y.; Shi, X. Refined Contact Map Prediction of Peptides Based on GCN and ResNet. *Front. Genet.* **2022**, *1*, 859626. [CrossRef] [PubMed]
186. Chen, C.; Wu, T.; Guo, Z.; Cheng, J. Combination of deep neural network with attention mechanism enhances the explainability of protein contact prediction. *Proteins* **2021**, *89*, 697–707. [CrossRef] [PubMed]
187. Chen, C.S.; Zhou, J.; Wang, F.; Liu, X.; Dou, D. Structure-aware protein self-supervised learning. *Bioinformatics* **2023**, *39*, btad189. [CrossRef] [PubMed]
188. Konc, J.; Janežič, D. Protein binding sites for drug design. *Biophys. Rev.* **2022**, *14*, 1413–1421. [CrossRef] [PubMed]
189. Chen, C.; Hou, J.; Shi, X.; Yang, H.; Birchler, J.A.; Cheng, J. DeepGRN: Prediction of transcription factor binding site across cell-types using attention-based deep neural networks. *BMC Bioinform.* **2021**, *22*, 38. [CrossRef] [PubMed]
190. Kuzmanic, A.; Bowman, G.R.; Juarez-Jimenez, J.; Michel, J.; Gervasio, F.L. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. *Acc. Chem. Res.* **2020**, *53*, 654–661. [CrossRef] [PubMed]
191. Guterres, H.; Lee, H.S.; Im, W. Ligand-Binding-Site Structure Refinement Using Molecular Dynamics with Restraints Derived from Predicted Binding Site Templates. *J. Chem. Theory Comput.* **2019**, *15*, 6524–6535. [CrossRef] [PubMed]
192. Ramazi, S.; Zahiri, J. Posttranslational modifications in proteins: Resources, tools and prediction methods. *Database* **2021**, *2021*, baab012. [CrossRef] [PubMed]
193. Pascovici, D.; Wu, J.X.; McKay, M.J.; Joseph, C.; Noor, Z.; Kamath, K.; Wu, Y.; Ranganathan, S.; Gupta, V.; Mirzaei, M. Clinically Relevant Post-Translational Modification Analyses-Maturing Workflows and Bioinformatics Tools. *Int. J. Mol. Sci.* **2018**, *20*, 16. [CrossRef] [PubMed]
194. Yan, Y.; Wang, D.; Xin, R.; Soriano, R.A.; Ng, D.C.M.; Wang, W.; Ping, P. Protocol for the prediction, interpretation, and mutation evaluation of post-translational modification using MIND-S. *STAR Protoc.* **2023**, *4*, 102682. [CrossRef] [PubMed]
195. Lyu, Z.; Wang, Z.; Luo, F.; Shuai, J.; Huang, Y. Protein Secondary Structure Prediction With a Reductive Deep Learning Method. *Front. Bioeng. Biotechnol.* **2021**, *9*, 687426. [CrossRef] [PubMed]
196. Braghetto, A.; Orlandini, E.; Baiesi, M. Interpretable Machine Learning of Amino Acid Patterns in Proteins: A Statistical Ensemble Approach. *J. Chem. Theory Comput.* **2023**, *19*, 6011–6022. [CrossRef] [PubMed]
197. ElAbd, H.; Bromberg, Y.; Hoarfrost, A.; Lenz, T.; Franke, A.; Wendorff, M. Amino acid encoding for deep learning applications. *BMC Bioinform.* **2020**, *21*, 235. [CrossRef] [PubMed]
198. Chandra, A.; Tünnermann, L.; Löfstedt, T.; Gratz, R. Transformer-based deep learning for predicting protein properties in the life sciences. *Elife* **2023**, *12*, e82819. [CrossRef] [PubMed]
199. Szeto, C.; Lobos, C.A.; Nguyen, A.T.; Gras, S. TCR Recognition of Peptide-MHC-I: Rule Makers and Breakers. *Int. J. Mol. Sci.* **2020**, *22*, 68. [CrossRef] [PubMed]
200. Malonis, R.J.; Lai, J.R.; Vergnolle, O. Peptide-Based Vaccines: Current Progress and Future Challenges. *Chem. Rev.* **2020**, *120*, 3210–3229. [CrossRef] [PubMed]
201. Rouet, R.; Henry, J.Y.; Johansen, M.D.; Sobti, M.; Balachandran, H.; Langley, D.B.; Walker, G.J.; Lenthall, H.; Jackson, J.; Ubiparipovic, S.; et al. Broadly neutralizing SARS-CoV-2 antibodies through epitope-based selection from convalescent patients. *Nat. Commun.* **2023**, *14*, 687. [CrossRef] [PubMed]
202. Røgen, P. Quantifying steric hindrance and topological obstruction to protein structure superposition. *Algorithms Mol. Biol.* **2021**, *16*, 1. [CrossRef] [PubMed]
203. Gaudreault, F.; Corbeil, C.R.; Sulea, T. Enhanced antibody-antigen structure prediction from molecular docking using AlphaFold2. *Sci. Rep.* **2023**, *13*, 15107. [CrossRef] [PubMed]
204. Bukhari, S.N.H.; Jain, A.; Haq, E.; Mehbodniya, A.; Webber, J. Machine Learning Techniques for the Prediction of B-Cell and T-Cell Epitopes as Potential Vaccine Targets with a Specific Focus on SARS-CoV-2 Pathogen: A Review. *Pathogens* **2022**, *11*, 146. [CrossRef] [PubMed]
205. Liu, T.; Shi, K.; Li, W. Deep learning methods improve linear B-cell epitope prediction. *BioData Min.* **2020**, *13*, 1. [CrossRef] [PubMed]
206. Guo, Y.; Li, W.; Wang, B.; Liu, H.; Zhou, D. DeepACLSTM: Deep asymmetric convolutional long short-term memory neural models for protein secondary structure prediction. *BMC Bioinform.* **2019**, *20*, 341. [CrossRef] [PubMed]
207. Yi, H.C.; You, Z.H.; Zhou, X.; Cheng, L.; Li, X.; Jiang, T.H.; Chen, Z.H. ACP-DL: A Deep Learning Long Short-Term Memory Model to Predict Anticancer Peptides Using High-Efficiency Feature Representation. *Mol. Ther. Nucleic Acids* **2019**, *17*, 1–9. [CrossRef] [PubMed]
208. Zhong, W.; He, C.; Xiao, C.; Liu, Y.; Qin, X.; Yu, Z. Long-distance dependency combined multi-hop graph neural networks for protein-protein interactions prediction. *BMC Bioinform.* **2022**, *23*, 521. [CrossRef] [PubMed]
209. Zhao, Y.; Liu, Y. OCLSTM: Optimized convolutional and long short-term memory neural network model for protein secondary structure prediction. *PLoS ONE* **2021**, *16*, e0245982. [CrossRef] [PubMed]
210. Liu, X.R.; Zhang, M.M.; Gross, M.L. Mass Spectrometry-Based Protein Footprinting for Higher-Order Structure Analysis: Fundamentals and Applications. *Chem. Rev.* **2020**, *120*, 4355–4454. [CrossRef] [PubMed]
211. Murgas, K.A.; Saucan, E.; Sandhu, R. Hypergraph geometry reflects higher-order dynamics in protein interaction networks. *Sci. Rep.* **2022**, *12*, 20879. [CrossRef] [PubMed]
212. Al Mughram, M.H.; Herrington, N.B.; Catalano, C.; Kellogg, G.E. Systematized analysis of secondary structure dependence of key structural features of residues in soluble and membrane-bound proteins. *J. Struct. Biol. X* **2021**, *5*, 100055. [CrossRef] [PubMed]

213. Li, Y.; Li, X.; Liu, Y.; Yao, Y.; Huang, G. MPMABP: A CNN and Bi-LSTM-Based Method for Predicting Multi-Activities of Bioactive Peptides. *Pharmaceuticals* **2022**, *15*, 707. [CrossRef] [PubMed]
214. Yin, S.; Mi, X.; Shukla, D. Leveraging Machine Learning Models for Peptide-Protein Interaction Prediction. *arXiv* **2023**, arXiv:2310.18249. [CrossRef] [PubMed]
215. Boehm, K.M.; Bhinder, B.; Raja, V.J.; Dephoure, N.; Elemento, O. Predicting peptide presentation by major histocompatibility complex class I: An improved machine learning approach to the immunopeptidome. *BMC Bioinform.* **2019**, *20*, 7. [CrossRef] [PubMed]
216. Wu, C.; Gao, R.; Zhang, Y.; De Marinis, Y. PTPD: Predicting therapeutic peptides by deep learning and word2vec. *BMC Bioinform.* **2019**, *20*, 456. [CrossRef] [PubMed]
217. Caro, M.C.; Huang, H.Y.; Cerezo, M.; Sharma, K.; Sornborger, A.; Cincio, L.; Coles, P.J. Generalization in quantum machine learning from few training data. *Nat. Commun.* **2022**, *13*, 4919. [CrossRef] [PubMed]
218. Zaheer, M.; Guruganesh, G.; Dubey, A.; Ainslie, J.; Alberti, C.; Ontanon, S.; Pham, P.; Ravula, A.; Wang, Q.; Yang, L.; et al. Big Bird: Transformers for Longer Sequences. *arXiv* **2021**, arXiv:2007.14062. [CrossRef]
219. Yang, Z.; Zeng, X.; Zhao, Y.; Chen, R. AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduct. Target. Ther.* **2023**, *8*, 115. [CrossRef] [PubMed]
220. Su, Y.; Liu, F.; Meng, Z.; Lan, T.; Shu, L.; Shareghi, E.; Collier, N. TaCL: Improving BERT Pre-training with Token-aware Contrastive Learning. *arXiv* **2022**, arXiv:2111.04198. [CrossRef]
221. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805. [CrossRef]
222. Tikhonova, I.G.; Gigoux, V.; Fourmy, D. Understanding Peptide Binding in Class A G Protein-Coupled Receptors. *Mol. Pharmacol.* **2019**, *96*, 550–561. [CrossRef] [PubMed]
223. Bashore, C.; Prakash, S.; Johnson, M.C.; Conrad, R.J.; Kekessie, I.A.; Scales, S.J.; Ishisoko, N.; Kleinheinz, T.; Liu, P.S.; Popovych, N.; et al. Targeted degradation via direct 26S proteasome recruitment. *Nat. Chem. Biol.* **2023**, *19*, 55–63. [CrossRef] [PubMed]
224. Meng, Y.; Huang, J.; Wang, G.; Wang, Z.; Zhang, C.; Han, J. Unsupervised Word Embedding Learning by Incorporating Local and Global Contexts. *Front. Big Data* **2020**, *3*, 9. [CrossRef] [PubMed]
225. Jespersen, M.C.; Mahajan, S.; Peters, B.; Nielsen, M.; Marcatili, P. Antibody Specific B-Cell Epitope Predictions: Leveraging Information From Antibody-Antigen Protein Complexes. *Front. Immunol.* **2019**, *10*, 298. [CrossRef] [PubMed]
226. Zeng, Y.; Wei, Z.; Yuan, Q.; Chen, S.; Yu, W.; Lu, Y.; Gao, J.; Yang, Y. Identifying B-cell epitopes using AlphaFold2 predicted structures and pretrained language model. *Bioinformatics* **2023**, *39*, btad187. [CrossRef] [PubMed]
227. Hou, Q.; Waury, K.; Gogishvili, D.; Feenstra, K.A. Ten quick tips for sequence-based prediction of protein properties using machine learning. *PLoS Comput. Biol.* **2022**, *18*, e1010669. [CrossRef] [PubMed]
228. Ras-Carmona, A.; Lehmann, A.A.; Lehmann, P.V.; Reche, P.A. Prediction of B cell epitopes in proteins using a novel sequence similarity-based method. *Sci. Rep.* **2022**, *12*, 13739. [CrossRef] [PubMed]
229. Lu, S.; Li, Y.; Ma, Q.; Nan, X.; Zhang, S. Structure-Based B-cell Epitope Prediction Model Through Combing Local and Global Features. *Front. Immunol.* **2022**, *13*, 890943. [CrossRef] [PubMed]
230. Noumi, T.; Inoue, S.; Fujita, H.; Sadamitsu, K.; Sakaguchi, M.; Tenma, A.; Nakagami, H. Epitope Prediction of Antigen Protein using Attention-Based LSTM Network. *J. Inf. Process.* **2020**, *29*, 321–327. [CrossRef]
231. Syrlybaeva, R.; Strauch, E.M. Deep learning of protein sequence design of protein-protein interactions. *Bioinformatics* **2023**, *39*, btac733. [CrossRef] [PubMed]
232. Mittal, S.; Manna, S.; Pathak, B. Machine Learning Prediction of the Transmission Function for Protein Sequencing with Graphene Nanoslit. *ACS Appl. Mater. Interfaces* **2022**, *14*, 51645–51655. [CrossRef] [PubMed]
233. Wu, T.-H.; Lin, P.-C.; Chou, H.-H.; Shen, M.-R.; Hsieh, S.-Y. Pathogenicity Prediction of Single Amino Acid Variants With Machine Learning Model Based on Protein Structural Energies. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2023**, *20*, 606–615. [CrossRef]
234. Grechishnikova, D. Transformer neural network for protein-specific de novo drug generation as a machine translation problem. *Sci. Rep.* **2021**, *11*, 321. [CrossRef] [PubMed]
235. Liu, Z.; Jin, J.; Cui, Y.; Xiong, Z.; Nariri, A.; Zhao, Y.; Hu, J. DeepSeqPanII: An Interpretable Recurrent Neural Network Model with Attention Mechanism for Peptide-HLA Class II Binding Prediction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 2188–2196. [CrossRef]
236. Elbasani, E.; Njimbouom, S.N.; Oh, T.J.; Kim, E.H.; Lee, H.; Kim, J.D. GCRNN: Graph convolutional recurrent neural network for compound-protein interaction prediction. *BMC Bioinform.* **2022**, *22*, 616. [CrossRef] [PubMed]
237. Xia, C.; Feng, S.H.; Xia, Y.; Pan, X.; Shen, H.B. Fast protein structure comparison through effective representation learning with contrastive graph neural networks. *PLoS Comput. Biol.* **2022**, *18*, e1009986. [CrossRef] [PubMed]
238. Jha, K.; Saha, S.; Singh, H. Prediction of protein-protein interaction using graph neural networks. *Sci. Rep.* **2022**, *12*, 8360. [CrossRef] [PubMed]
239. Réau, M.; Renaud, N.; Xue, L.C.; Bonvin, A.M.J.J. DeepRank-GNN: A graph neural network framework to learn patterns in protein-protein interfaces. *Bioinformatics* **2023**, *39*, btac759. [CrossRef] [PubMed]
240. McDonnell, K.; Abram, F.; Howley, E. Application of a Novel Hybrid CNN-GNN for Peptide Ion Encoding. *J. Proteome Res.* **2023**, *22*, 323–333. [CrossRef] [PubMed]

241. Pan, T.; Li, C.; Bi, Y.; Wang, Z.; Gasser, R.B.; Purcell, A.W.; Akutsu, T.; Webb, G.I.; Imoto, S.; Song, J. PFresGO: An attention mechanism-based deep-learning approach for protein annotation by integrating gene ontology inter-relationships. *Bioinformatics* **2023**, *39*, btad094. [CrossRef] [PubMed]
242. Bhattacharya, N.; Thomas, N.; Rao, R.; Dauparas, J.; Koo, P.K.; Baker, D.; Song, Y.S.; Ovchinnikov, S. Interpreting Potts and Transformer Protein Models through the Lens of Simplified Attention. *Pac. Symp. Biocomput.* **2022**, *27*, 34–45. [PubMed]
243. Xu, P.; Ji, X.; Li, M.; Lu, W. Small data machine learning in materials science. *NPJ Comput. Mater.* **2023**, *9*, 4. [CrossRef]
244. Nielsen, M.; Andreatta, M.; Peters, B.; Buus, S. Immunoinformatics: Predicting Peptide-MHC Binding. *Annu. Rev. Biomed. Data Sci.* **2020**, *3*, 191–215. [CrossRef] [PubMed]
245. Lee, M. Recent Advances in Deep Learning for Protein-Protein Interaction Analysis: A Comprehensive Review. *Molecules* **2023**, *28*, 5169. [CrossRef] [PubMed]
246. Myronov, A.; Mazzocco, G.; Król, P.; Plewczynski, D. BERTrand-peptide:TCR binding prediction using Bidirectional Encoder Representations from Transformers augmented with random TCR pairing. *Bioinformatics* **2023**, *39*, btad468. [CrossRef] [PubMed]
247. Cheng, J.; Bendjama, K.; Rittner, K.; Malone, B. BERTMHC: Improved MHC-peptide class II interaction prediction with transformer and multiple instance learning. *Bioinformatics* **2021**, *37*, 4172–4179. [CrossRef] [PubMed]
248. Guntuboina, C.; Das, A.; Mollaei, P.; Kim, S.; Barati Farimani, A. PeptideBERT: A Language Model Based on Transformers for Peptide Property Prediction. *J. Phys. Chem. Lett.* **2023**, *14*, 10427–10434. [CrossRef] [PubMed]
249. Wang, R.; Jin, J.; Zou, Q.; NAkai, K.; Wei, L. Predicting protein-peptide binding residues via interpretable deep learning. *Bioinformatics* **2022**, *38*, 3351–3360. [CrossRef]
250. Ji, Y.; Zhou, Z.; Liu, H.; Davuluri, R.V. DNABERT: Pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. *Bioinformatics* **2021**, *37*, 2112–2120. [CrossRef]
251. Sanchez, C.; Zhang, Z. The Effects of In-domain Corpus Size on pre-training BERT. *arXiv* **2022**, arXiv:2212.07914. [CrossRef]
252. Tejani, A.S.; Ng, Y.S.; Xi, Y.; Fielding, J.R.; Browning, T.G.; Rayan, J.C. Performance of Multiple Pretrained BERT Models to Automate and Accelerate Data Annotation for Large Datasets. *Radiol. Artif. Intell.* **2022**, *4*, e220007. [CrossRef] [PubMed]
253. Hou, L.; Pang, R.Y.; Zhou, T.; Wu, Y.; Song, X.; Song, X.; Zhou, D. Token Dropping for Efficient BERT Pretraining. *arXiv* **2022**, arXiv:2203.13240. [CrossRef]
254. Jiang, L.; Jiang, J.; Wang, X.; Zhang, Y.; Zheng, B.; Liu, S.; Zhang, Y.; Liu, C.; Wan, Y.; Xiang, D.; et al. IUP-BERT: Identification of Umami Peptides Based on BERT Features. *Foods* **2022**, *11*, 3742. [CrossRef] [PubMed]
255. Charoenkwan, P.; Nantasenamat, C.; Hasan, M.M.; Manavalan, B.; Shoombuatong, W. BERT4Bitter: A bidirectional encoder representations from transformers (BERT)-based model for improving the prediction of bitter peptides. *Bioinformatics* **2021**, *37*, 2556–2562. [CrossRef] [PubMed]
256. Luo, X.; Ding, H.; Tang, M.; Gandhi, P.; Zhang, Z.; He, Z. Attention Mechanism with BERT for Content Annotation and Categorization of Pregnancy-Related Questions on a Community Q&A Site. *Proc. IEEE Int. Conf. Bioinform. Biomed.* **2020**, *2020*, 1077–1081. [CrossRef] [PubMed]
257. Huang, S.; Wu, X.; Wu, X.; Wang, K. Sentiment analysis algorithm using contrastive learning and adversarial training for POI recommendation. *Soc. Netw. Anal. Min.* **2023**, *13*, 75. [CrossRef] [PubMed]
258. Wan, Z.; Liu, C.; Zhang, M.; Fu, J.; Wang, B.; Cheng, S.; Ma, L.; Quilodrán-Casas, C.; Arcucci, R. Med-UniC: Unifying Cross-Lingual Medical Vision-Language Pre-Training by Diminishing Bias. *arXiv* **2024**, arXiv:2305.19894. [CrossRef]
259. Li, J.; Liu, C.; Cheng, S.; Arcucci, R.; Hong, S. Frozen Language Model Helps ECG Zero-Shot Learning. *arXiv* **2023**, arXiv:2303.12311. [CrossRef]
260. Wang, Q.L.; Wang, T.M.; Deng, C.M.; Zhang, W.L.; He, Y.Q.; Xue, W.Q.; Liao, Y.; Yang, D.W.; Zheng, M.Q.; Jia, W.H. Association of HLA diversity with the risk of 25 cancers in the UK Biobank. *EBioMedicine* **2023**, *92*, 104588. [CrossRef] [PubMed]
261. Sakaue, S.; Hosomichi, K.; Hirata, J.; Nakaoka, H.; Yamazaki, K.; Yawata, M.; Yawata, N.; Naito, T.; Umeno, J.; Kawaguchi, T.; et al. Decoding the diversity of killer immunoglobulin-like receptors by deep sequencing and a high-resolution imputation method. *Cell Genom.* **2022**, *2*, 100101. [CrossRef] [PubMed]
262. Bycroft, C.; Freeman, C.; Petkova, D.; Band, G.; Elliott, L.T.; Sharp, K.; Motyer, A.; Vukcevic, D.; Delaneau, O.; O'Connell, J.; et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **2018**, *562*, 203–209. [CrossRef] [PubMed]
263. Lipkova, J.; Chen, R.J.; Chen, B.; Lu, M.Y.; Barbieri, M.; Shao, D.; Vaidya, A.J.; Chen, C.; Zhuang, L.; Williamson, D.F.K.; et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* **2022**, *40*, 1095–1110. [CrossRef] [PubMed]
264. Roelofsen, L.M.; Kaptein, P.; Thommen, D.S. Multimodal predictors for precision immunotherapy. *Immunooncol Technol.* **2022**, *14*, 100071. [CrossRef] [PubMed]