



# CAD v1.0: Cancer Antigens Database Platform for Cancer Antigen Algorithm Development and Information Exploration

Jijun Yu<sup>1,2</sup>, Luoxuan Wang<sup>3</sup>, Xiangya Kong<sup>4</sup>, Yang Cao<sup>5</sup>, Mengmeng Zhang<sup>1,6</sup>, Zhaolin Sun<sup>6</sup>, Yang Liu<sup>5</sup>, Jing Wang<sup>1,2</sup>, Beifen Shen<sup>1,2</sup>, Xiaochen Bo<sup>7\*</sup> and Jiannan Feng<sup>1,2\*</sup>

<sup>1</sup>State Key Laboratory of Toxicology and Medical Countermeasures, Beijing Institute of Pharmacology and Toxicology, Beijing, China, <sup>2</sup>Beijing Key Laboratory of Therapeutic Gene Engineering Antibody, Beijing, China, <sup>3</sup>State Key Laboratory of Toxicology and Medical Countermeasures, Beijing Key Laboratory of Neuropsychopharmacology, Beijing Institute of Pharmacology and Toxicology, Beijing, China, <sup>4</sup>Beijing Geneworks Technology Co., Ltd., Beijing, China, <sup>5</sup>Center of Growth, Metabolism and Aging, Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China, <sup>6</sup>Beijing Capital Agribusiness Future Biotechnology Co, Beijing, China, <sup>7</sup>Department of Biotechnology, Beijing Institute of Radiation Medicine, Beijing, China

## OPEN ACCESS

### Edited by:

Venkata Yellapantula,  
Memorial Sloan Kettering Cancer  
Center, United States

### Reviewed by:

Guideng Li,  
Chinese Academy of Medical  
Sciences, China  
Muzamil Yaqub Want,  
Roswell Park Comprehensive Cancer  
Center, United States

### \*Correspondence:

Xiaochen Bo  
boxc@bmi.ac.cn  
Jiannan Feng  
fengjiannan1970@qq.com

### Specialty section:

This article was submitted to  
Nanobiotechnology,  
a section of the journal  
Frontiers in Bioengineering and  
Biotechnology

Received: 25 January 2022

Accepted: 06 April 2022

Published: 12 May 2022

### Citation:

Yu J, Wang L, Kong X, Cao Y,  
Zhang M, Sun Z, Liu Y, Wang J,  
Shen B, Bo X and Feng J (2022) CAD  
v1.0: Cancer Antigens Database  
Platform for Cancer Antigen Algorithm  
Development and  
Information Exploration.  
Front. Bioeng. Biotechnol. 10:819583.  
doi: 10.3389/fbioe.2022.819583

Cancer vaccines have gradually attracted attention for their tremendous preclinical and clinical performance. With the development of next-generation sequencing technologies and related algorithms, pipelines based on sequencing and machine learning methods have become mainstream in cancer antigen prediction; of particular focus are neoantigens, mutation peptides that only exist in tumor cells that lack central tolerance and have fewer side effects. The rapid prediction and filtering of neoantigen peptides are crucial to the development of neoantigen-based cancer vaccines. However, due to the lack of verified neoantigen datasets and insufficient research on the properties of neoantigens, neoantigen prediction algorithms still need to be improved. Here, we recruited verified cancer antigen peptides and collected as much relevant peptide information as possible. Then, we discussed the role of each dataset for algorithm improvement in cancer antigen research, especially neoantigen prediction. A platform, Cancer Antigens Database (CAD, <http://cad.bio-it.cn/>), was designed to facilitate users to perform a complete exploration of cancer antigens online.

**Keywords:** tumor-associated antigens (TAAs), tumor-specific antigens (TSAs), prediction model, cancer antigen, neoantigen

## INTRODUCTION

Colev (189) was the first to attempt to leverage patient immune systems to fight against cancer. Since then, immunotherapy has made great progress, especially in immune checkpoint blockades of cytotoxic T-lymphocyte-associated protein 4 and programmed cell death protein 1 (PD-1) in melanoma (Hodi et al., 2010; Robert et al., 2015) and other cancers (Zou et al., 2016). However, not all patients benefit from immune checkpoint inhibitors, which may cause immune-related adverse events. Therefore, additional immunotherapies need to be developed.

Recently, cancer vaccines have attracted increasing attention with promising results in preclinical studies (Castle et al., 2012; Gubin et al., 2014) and clinical outcomes (Li et al., 2017; Ott et al., 2017; Sahin et al., 2017) in individual or combination therapies. Cancer antigens are peptides in tumor cells

present in antigen-presenting cells, which then provoke T-cell activation to kill tumor cells. The process of immune tumor killing is displayed in **Supplementary Figure S1A**. Ehx and Perreault (2019) suggested classifying cancer antigens into the following three categories based on specificity and mutation situations: tumor-associated antigens (TAAs), aberrantly expressed tumor-specific antigens (aeTSAs), and mutated tumor-specific antigens (mTSAs). TAAs are typically proteins that are overexpressed in tumor cells but also exist in normal tissues. aeTSAs drive the majority of epigenetic changes in atypical translation events, which are widely expressed in various cancers and could be shared among patients (Probst et al., 2017; Schuster et al., 2017; Gee et al., 2018; Laumont et al., 2018). mTSAs, also known as neoantigens, are patient-specific mutation peptides that only exist in tumor cells. The classification method was also used to organize recruited cancer antigens in the current study.

The efficacies of many TAAs and aeTSAs have been investigated in clinical trials to date, but global clinical outcomes have not been encouraging. However, many neoantigen-based cancer vaccines have generated effective antitumor responses in multiple preclinical studies (Gubin et al., 2014; Li et al., 2017). For example, Castle et al. (2012) identified 16 candidate neoantigens confirmed to be immunogenic using IFN-gamma enzyme-linked immunospot assay. Peptide immunization has conferred *in vivo* tumor control in protective and therapeutic settings. Apart from allograft transplantation models, the establishment of immunodeficient nude and NOD.scid.gamma mice make it possible to study human cancer cells in mice. Zhang et al. (2017) found that two neoantigens, ROBO3 A1265V and PALB2 H198D, with adoptive transfer of autologous peripheral blood mononuclear cells stimulated *in vitro* with mutant peptides decreased tumor growth. In addition, many recent preclinical studies have achieved a broad immune response, such as increased cytotoxic T-lymphocyte response and decrease in the tumor growth rate in mouse models (Ahn et al., 2015; Chen et al., 2015; Rekoske et al., 2015; Lambricht et al., 2016; Lopes et al., 2017, 2018; Wu et al., 2017; Zhao et al., 2017; Gao et al., 2018).

In addition to the preclinical research mentioned earlier, cancer neoantigen vaccines have also made great progress in clinical research. Numerous clinical studies have confirmed the safety and efficacy of cancer neoantigen vaccines (Ott et al., 2017, 2020; Keskin et al., 2019; Platten et al., 2021), and the outcomes of clinical trials on melanoma were particularly encouraging. Ott et al. (2017) demonstrated the potential of neoantigen vaccines in the treatment of melanoma. Of six vaccinated patients after surgical resection, four had no recurrence at the 20- to 32-month follow-up; the remaining two with recurrence were treated with PD-1. After antibody therapy, complete remission was achieved. Keskin et al. (2019) (NC02287428) found that neoantigen-specific T cells from peripheral blood can migrate into intracranial glioblastoma, which provided evidence that cancer neoantigen vaccines could enhance the immune microenvironment of glioblastoma cells. These clinical studies have indicated that neoantigen vaccines

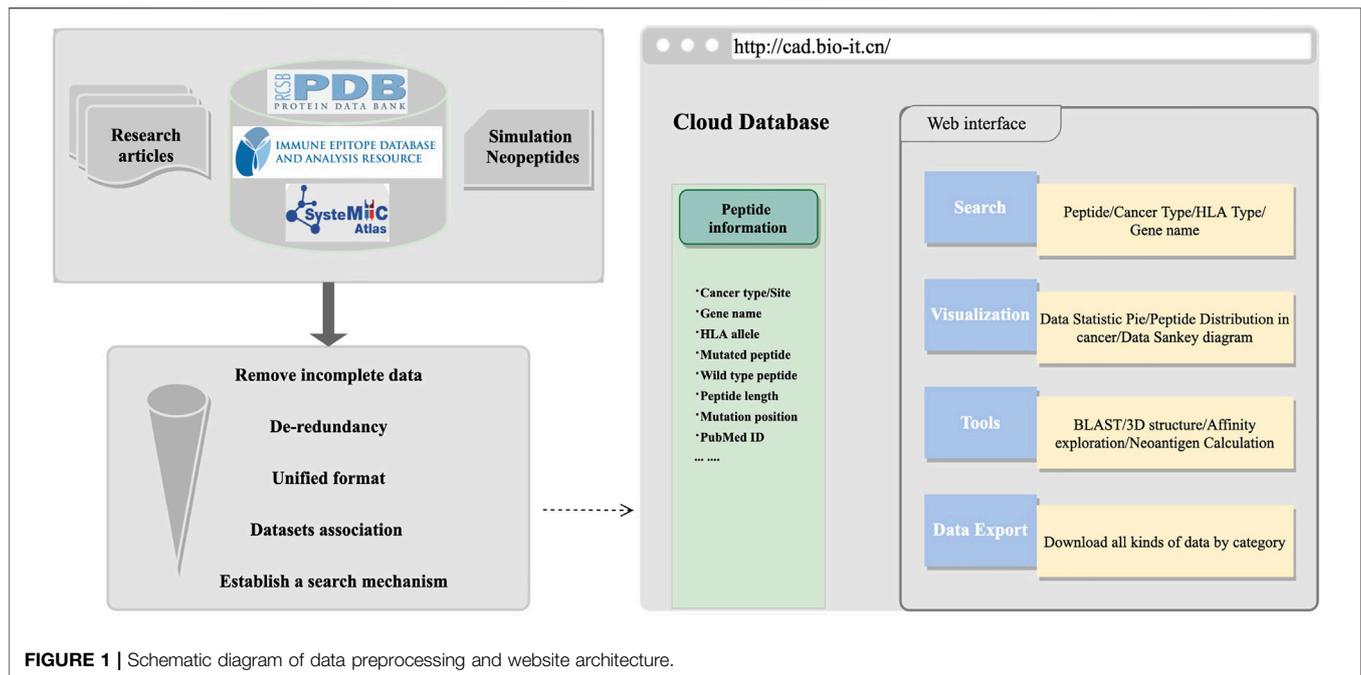
could play a meaningful role in different tumors. This evidence provides great confidence for further exploration of the clinical path of neoantigen vaccines.

In both preclinical and clinical studies, rapid and accurate neoantigen prediction plays a key role in the development of neoantigen vaccines. In this study, we explore the use of relevant datasets in neoantigen prediction.

The process of neoantigen prediction includes 1) sampling of patient tumor and adjacent tissues, 2) whole-genome sequencing or whole-exome sequencing and RNA-seq sequencing, 3) identifying somatic mutations and calculating the expression levels and HLA subtypes, 4) calculating the binding affinity of potential neoantigens to HLA subtypes, and 5) screening candidate neoantigen. A flowchart of neoantigen prediction is displayed in **Supplementary Figure S1B**.

Currently, multiple software programs and algorithms have been developed in neoantigen prediction pipelines (Soria-Guerra et al., 2015; Peng et al., 2019; Lancaster et al., 2020), and the most commonly used tools are NetMHC (Lundegaard et al., 2008a) and NetMHCpan (Reynisson et al., 2021), which use allele-specific epitope prediction and pan-specific machine learning methods, respectively. After affinity screening, researchers further screen candidate neoantigens by peptide immunogenicity prediction, such as neoantigen prediction workflows INeo-Epp (Wang et al., 2020) and pTuneos (Chi Zhou et al., 2019). These algorithms are based on the immunogenicity-related characteristics of T-cell epitope peptides combined with machine learning algorithms.

Insufficient and low-quality datasets affect the accuracy of prediction models. Presently, many cancer antigens or neoantigen peptides have been generated. It is necessary to recruit and organize these datasets to facilitate the improvement of existing algorithms or the development of new algorithms. Existing databases help accomplish this purpose. The Cancer-Testis Database (Almeida et al., 2009) is a knowledge base of high-throughput and curated data on cancer-testis antigen a. It provides an important resource for the exploration of cancer-testis antigens. NEPdb (Xia et al., 2021), dbPepNeo (Tan et al., 2020), and NeoPeptide (Wei-Jun Zhou et al., 2019) contain curated neoantigens but lack any other types of cancer antigens; TSNAdb v1.0 (Wu et al., 2018) predicts candidate neoantigens based on The Cancer Genome Atlas and The Cancer Immunome Atlas datasets, which are useful for comparing candidate neoantigens in different cancers. The Tumor T-cell Antigen Database, TANTIGEN 2.0 (Zhang et al., 2021), contains HLA ligands and T-cell epitopes and classifies cancer antigens. However, this database does not contain invalid peptides and lacks clinically relevant data. The blueprint of the Cancer Epitope Database and Analysis Resource (CEDAR) database, proposed by Koşaloğlu-Yalçın et al. (2021), is based on the Immune Epitope Database (IEDB) and tries to collect more specific information and classify it in detail. The aforementioned databases focus on online information exploration but not on algorithm development. To fill this gap, we built the Cancer Antigens Database (CAD), which recruited all cancer antigen peptides and relevant datasets, established neoantigen simulation datasets, carried out detailed



data preprocessing, and explained the scope of application and precautions of different datasets in algorithm development. A user-friendly platform was concurrently established to facilitate online exploration.

## DATA COLLECTION

We recruited cancer antigens verified from published articles (~900) and collected associated peptides from other resources, such as the IEDB (Vita et al., 2019) for peptide binding and T-cell epitope datasets, the SystemMHC atlas (Shao et al., 2018) for mass spectrometry (MS) datasets, the VDJdb database (Bagaev et al., 2020) for antigens with T-cell receptor (TCR) sequence datasets, and the Protein Data Bank for three-dimensional structure of peptide-major histocompatibility complex (p-MHC) or pMHC-TCR complex. After recruiting all verified cancer antigens, a small amount of neoantigen data were insufficient for the development of algorithms; therefore, we also generated simulation neopeptides, for which dbSNP datasets from the National Center for Biotechnology Information (Sherry et al., 2001), the reviewed SwissProt human proteins sequence from Uniprot (Bateman et al., 2021), and verified T-cell epitopes from IEDB were used.

Finally, more than 800 cancer antigens were recruited for our database, which includes cancers such as skin, brain, and kidney cancers, as well as a total of 267 verified neoantigens. Except for verified cancer antigens, information on associated peptides, including cleaned MHC-peptide binding (569953), T-cell epitopes (66151), pMHC MS (509536), antigens with TCR sequences (60267), and more than 6,000 simulated neopeptides, was included in our datasets (only HLA-A\*0201 was included; for more HLA alleles, users can generate their own

HLA alleles datasets. Refer to methods described in **Supplementary Figure S2** and code from <https://github.com/yujijun/NeoSimData>).

## USAGE ABOUT DATASETS

The majority of datasets in our database are suitable for algorithm development. For example, all curated verified cancer antigen datasets can be used as preliminary verification for predicting candidate antigens to test whether the same or similar cancer antigens have been studied or reported. The simulated neopeptide datasets have been used in feature-based neoantigen immunogenicity algorithms (Kim et al., 2018). A large number of neoantigen simulation datasets have greatly filled the gap of insufficient neoantigen datasets, providing a choice for the application of machine learning and even deep learning in the immunogenicity prediction of neoantigens. We provided an upgraded version to facilitate the generation of more flexible and widely applicable simulation datasets.

Part of the MHC-binding datasets in IEDB has been used in binding prediction algorithms; it should be noted that many peptides from IEDB belonged to bacteria or viruses but not humans and also were not obtained by standardized experimental methodologies in cancer models, which may reduce the accuracy of the algorithm prediction to a certain extent (Jiang et al., 2019). In addition, redundant information can also cause inaccurate model evaluation. Therefore, cleaned and selected human origin datasets have been generated and stored in our database. At present, tumor neoantigen prediction and filtering are mainly based on the binding affinity of MHC and peptides. However, peptide and MHC binding is a necessary but not sufficient condition for T-cell activation; therefore, many

peptides screened by peptide–MHC binding cannot activate T cells to induce immune responses. T-cell epitope datasets were a useful resource to predict peptide immunogenicity for cancer treatment. For example, T-cell epitope datasets could be used as prior knowledge to improve the accuracy of algorithms, just as Balachandran et al. (2017) and Luksza et al. (2017) developed a new approach for assessing whether a tumor is immunogenically based on estimated likelihood of TCR recognition for each predicted neoantigen. These estimates were computed from sequence similarities between the predicted neoantigens and a dataset of immunogenic epitopes. Wang et al. (2020) developed INeo-Epp, a random forest classifier for T-cell immunogenic HLA-I-presenting antigen epitopes and neoantigens based on sequence-related amino acid features. Riley et al. (2019) trained a neural network on structural features that influence TCR and peptide-binding energies.

Several groups are interested in systematically studying the binding of TCR to peptides/MHC. Gielis et al. (2020) developed a web tool TCRex for the prediction of T-cell receptor sequence epitope specificity, which allows users to upload TCR sequences and predict interaction with multiple known epitopes; Montemurro et al. (2021) developed NetTCR-2.0, which enables accurate prediction of TCR–peptide binding by the “shallow” convolutional neural network; Jokinen et al. (2021) developed TCRGP, a novel Gaussian process method that predicts recognition between T-cell receptors and epitopes, which has better performance in algorithm evaluation than existing state-of-the-art methods in epitope specificity predictions. Some databases have been built for curating such research; for example, the VDJdb database (Bagaev et al., 2020) curates TCR sequences with known antigen specificities. Peptide information of the dataset has also been integrated into our website.

Finally, we compiled a list of benchmark datasets in our database, which could be used for testing and verification of neoantigen pipelines. It included the entire process from original sequence datasets, predicted neoantigens, and experimentally verified immunogenic peptides, which have been used for the evaluation of several complete neoantigen prediction platforms such as pVACtools (Hundal et al., 2020), pTuneo (Chi Zhou et al., 2019), and neoepiscopy (Wood et al., 2020). All these datasets can also be used for cross-sectional evaluation and comparison between different neoantigen prediction pipelines (Supplementary Table S1). Detailed datasets statistics and usage are shown in Supplementary Table S2, and all the datasets mentioned before can be downloaded from our database (<http://cad.bio-it.cn/#/Download>).

## CONSTRUCTION OF CANCER ANTIGEN PLATFORM

These datasets can be used in algorithm development and exploration of cancer antigens online. After all datasets were collected or generated, we organized the information into a unified format, including tumor name, tissue site, gene name, peptides, MHC alleles, and mutation information, and then, a

user-friendly retrieval mechanism was established. The HOME page had detailed statistics about peptide information. On the SEARCH page, multiple retrieval methods were established. To facilitate users to perform comprehensive peptide exploration, we provided information such as hydrophobic and hydrophilic properties, which have been proved to have an important influence on the immunogenicity of antigens (Chowell et al., 2015). We also integrated the sequence alignment program of BLAST (Zeng et al., 2007) for sequence similarity exploration and constructed pMHC–TCR protein structure modeling tools for structure interaction exploration. To conduct a more in-depth analysis of cancer antigens, especially neoantigens, we introduced and linked some of the commonly used epitopes prediction (Buus et al., 2003; Nielsen et al., 2003, 2007a; Peters et al., 2003; Peters and Sette, 2005; Tenzer et al., 2005; Lundegaard et al., 2006, 2008b, 2008a), affinity prediction (Sturniolo et al., 1999; Nielsen et al., 2007b; Sidney et al., 2008; Wang et al., 2008, 2010; Nielsen and Lund, 2009; Andreatta et al., 2015; Jensen et al., 2018; Reynisson et al., 2021), and neoantigen prediction pipelines in the TOOLS page; all database or software mentioned is organized in Supplementary Table S3. The schematic diagram framework of the web construction processes is shown in Figure 1. Cancer antigen researchers are encouraged to use this platform to submit relevant information about cancer antigens on the SUBMIT page and feel free to download interesting datasets on the DOWNLOAD page. More information about this website can be found at <http://cad.bio-it.cn/#/FAQ>.

## USAGE OF CANCER ANTIGEN PLATFORM

A chosen peptide could be explored in this comprehensive platform. Specific processes can refer to Supplementary Figure S3. We also provide multiple useful tools, such as the sequence alignment tool, which provides the opportunity to explore and compare new cancer antigens with prior knowledge in our database. In addition, we also facilitated online MHC–peptide structure modeling, exploring the peptide–MHC binding and/or pMHC–TCR structure. For specific usage cases, refer to the description in Supplementary Figure S4. It can also be locally run by users with an algorithmic foundation. Users can refer to the source code at [https://github.com/yujijun/pMHC\\_TCR\\_binding](https://github.com/yujijun/pMHC_TCR_binding).

## DISCUSSION

Given the problems in the prediction of cancer antigens, especially neoantigens, we systematically organized and explained the usage of datasets. All these datasets and their usage proposals will greatly promote the accuracy of tumor neoantigen prediction, especially in machine learning or deep learning algorithm scenarios. However, in the process of algorithm development, developers must pay attention to the characteristics of each dataset. For example, pMHC MS datasets lack negative observations (peptides that do not bind), posing challenges in creating predictive models (Villani et al.,

2018). In the situation of immunogenicity study of antigens, the T-cell epitope datasets may be better than peptide-MHC binding or MS datasets because not all peptides presented by MHC can provoke T-cell activation. Multiple datasets and strategies can be integrated to improve overall results. For example, combining pMHC MS datasets into MHC-binding datasets might make prediction of peptide-MHC binding more accurate.

Except for considering the attributes of datasets, many methods of improving performance inherent in machine learning or deep learning can also be considered. De-redundancy of datasets may improve the scalability of the algorithm and also make a more accurate evaluation. In some of the datasets mentioned before, negative will be much larger than positive. Many down-sampling methods could be used to prevent model prediction bias; in the case of insufficient training data, users can generate simulated datasets as previously mentioned.

In addition to the detailed introduction of the data and algorithms mentioned previously, a one-stop interaction platform was established, which is convenient for all cancer antigen researchers to conduct online exploration of cancer antigen properties, such as the affinity and hydrophobicity, pMHC or PMHC-TCR docking characteristics, and key binding sites. This is very important for the re-excavation of existing cancer antigen information. Information is still not detail enough for readers to explore at a more specific. Therefore, we are looking forward to the blueprint mentioned in the CEDAR database (Koşaloğlu-Yalçın et al., 2021). CEDAR was built based on IEDB (Vita et al., 2019), including all cancer-specific epitope data from various T-cell and B-cell experiments, MHC-binding assays, and MHC ligandomics by MS. Simultaneously, the peptide information will be associated with biologically, immunologically, and clinically relevant information, and a fine-tuned classification and retrieval mechanism will be established. For researchers without programming experience, the information on relevant epitope terms can be accurately investigated online, which facilitates experiments based on prior knowledge. At the same time, it provides online calculation and objective evaluation between different epitope tools, which greatly reduces the difficulty of selecting epitope

prediction tools for those who have no programming experience and are unfamiliar with algorithms.

In this study, we curated the most comprehensive datasets of verified cancer antigens and systematically explained the usage of the various datasets in neoantigen algorithm development and then established the online exploration platform for cancer antigens and integrated useful tools to conveniently and comprehensively investigate. We believe all these efforts will support researchers in cancer antigens with or without programming experience. We will continue to improve our platform to make it more informative and convenient to use.

## DATA AVAILABILITY STATEMENT

The datasets for this study can be found at <http://cad.bio-it.cn/#/Download>, further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

JY, XB, and JF designed the whole subject; JY, XK, MZ, and ZS collected the data; JY, XK, YL, and JW developed the websites and tools; BS, XB, and JF curated the database; JY and LW created all figures and tables in the article; JY wrote the manuscript; and all authors read and approved the final manuscript.

## FUNDING

This work has been supported by the National Natural Science Foundation of China (No. 31771010).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2022.819583/full#supplementary-material>

## REFERENCES

- Ahn, E., Kim, H., Han, K. T., and Sin, J.-I. (2015). A Loss of Antitumor Therapeutic Activity of CEA DNA Vaccines Is Associated with the Lack of Tumor Cells' Antigen Presentation to Ag-specific CTLs in a colon Cancer Model. *Cancer Lett.* 356, 676–685. doi:10.1016/j.canlet.2014.10.019
- Almeida, L. G., Sakabe, N. J., deOliveira, A. R., Silva, M. C. C., Mundstein, A. S., Cohen, T., et al. (2009). CTdatabase: a Knowledge-Base of High-Throughput and Curated Data on Cancer-Testis Antigens. *Nucleic Acids Res.* 37, D816–D819. doi:10.1093/nar/gkn673
- Andreata, M., Karosiene, E., Rasmussen, M., Stryhn, A., Buus, S., and Nielsen, M. (2015). Accurate Pan-specific Prediction of Peptide-MHC Class II Binding Affinity with Improved Binding Core Identification. *Immunogenetics* 67, 641–650. doi:10.1007/s00251-015-0873-y
- Bagaev, D. V., Vroomans, R. M. A., Samir, J., Stervbo, U., Rius, C., Dolton, G., et al. (2020). VDjdb in 2019: Database Extension, New Analysis Infrastructure and a T-Cell Receptor Motif Compendium. *Nucleic Acids Res.* 48, D1057–D1062. doi:10.1093/nar/gkz874
- Balachandran, V. P., Łuksza, M., Łuksza, M., Zhao, J. N., Makarov, V., Moral, J. A., et al. (2017). Identification of Unique Neoantigen Qualities in Long-Term Survivors of Pancreatic Cancer. *Nature* 551, 512–516. doi:10.1038/nature24462
- Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., et al. (2021). UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi:10.1093/nar/gkaa1100
- Buus, S., Lauemøller, S. L., Worning, P., Kesmir, C., Frimurer, T., Corbet, S., et al. (2003). Sensitive Quantitative Predictions of Peptide-MHC Binding by a 'Query by Committee' Artificial Neural Network Approach. *Tissue Antigens* 62, 378–384. doi:10.1034/j.1399-0039.2003.00112.x
- Castle, J. C., Kreiter, S., Diekmann, J., Löwer, M., van de Roemer, N., de Graaf, J., et al. (2012). Exploiting the Mutanome for Tumor Vaccination. *Cancer Res.* 72, 1081–1091. doi:10.1158/0008-5472.CAN-11-3722
- Chen, R., Wang, S., Yao, Y., Zhou, Y., Zhang, C., Fang, J., et al. (2015). Antimetastatic Effects of DNA Vaccine Encoding Single-Chain Trimer Composed

- of MHC I and Vascular Endothelial Growth Factor Receptor 2 Peptide. *Oncol. Rep.* 33, 2269–2276. doi:10.3892/or.2015.3820
- Chowell, D., Krishna, S., Becker, P. D., Cocita, C., Shu, J., Tan, X., et al. (2015). TCR Contact Residue Hydrophobicity Is a Hallmark of Immunogenic CD8 + T Cell Epitopes. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1754–E1762. doi:10.1073/pnas.1500973112
- Colev, W. B. (1893). I. The Treatment of Malignant Tumors by Repeated Inoculations of Erysipelas. *Ann. Surg.* 18, 68. doi:10.1097/00000658-189307000-00009
- Ehx, G., and Perreault, C. (2019). Discovery and Characterization of Actionable Tumor Antigens. *Genome Med.* 11, 10–12. doi:10.1186/s13073-019-0642-x
- Gao, F.-S., Zhan, Y.-T., Wang, X.-D., and Zhang, C. (2018). Enhancement of Antitumor Effect of Plasmid DNA-Carrying MUC1 by the Adjuvanticity of FLT3L in Mouse Model. *Immunopharmacol. Immunotoxicol.* 40, 353–357. doi:10.1080/08923973.2018.1498099
- Gee, M. H., Han, A., Lofgren, S. M., Beausang, J. F., Mendoza, J. L., Birnbaum, M. E., et al. (2018). Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes. *Cell* 172, 549–563.e16. doi:10.1016/j.cell.2017.11.043
- Gielis, S., Moris, P., Bittremieux, W., De Neuter, N., Ogunjimi, B., Laukens, K., et al. (2020). Identification of Epitope-specific T Cells in T-Cell Receptor Repertoires. *Methods Mol. Biol.* 2120, 183–195. doi:10.1007/978-1-0716-0327-7\_13
- Gubin, M. M., Zhang, X., Schuster, H., Caron, E., Ward, J. P., Noguchi, T., et al. (2014). Checkpoint Blockade Cancer Immunotherapy Targets Tumour-Specific Mutant Antigens. *Nature* 515, 577–581. doi:10.1038/nature13988
- Hodi, F. S., O'Day, S. J., McDermott, D. F., Weber, R. W., Sosman, J. A., Haanen, J. B., et al. (2010). Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N. Engl. J. Med.* 363, 711–723. doi:10.1056/nejmoa1003466
- Hundal, J., Kiwala, S., McMichael, J., Miller, C. A., Xia, H., Wollam, A. T., et al. (2020). PVAtools: A Computational Toolkit to Identify and Visualize Cancer Neoantigens. *Cancer Immunol. Res.* 8, 409–420. doi:10.1158/2326-6066.CIR-19-0401
- Jensen, K. K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J. A., Yan, Z., et al. (2018). Improved Methods for Predicting Peptide Binding Affinity to MHC Class II Molecules. *Immunology* 154, 394–406. doi:10.1111/imm.12889
- Jiang, T., Shi, T., Zhang, H., Hu, J., Song, Y., Wei, J., et al. (2019). Tumor Neoantigens: From Basic Research to Clinical Applications. *J. Hematol. Oncol.* 12, 1–13. doi:10.1186/s13045-019-0787-5
- Jokinen, E., Huuhtanen, J., Mustjoki, S., Heinonen, M., and Lähdesmäki, H. (2021). Predicting Recognition between T Cell Receptors and Epitopes with TCRGP. *Plos Comput. Biol.* 17, e1008814. doi:10.1371/journal.pcbi.1008814
- Keskin, D. B., Anandappa, A. J., Sun, J., Tirosh, I., Mathewson, N. D., Li, S., et al. (2019). Neoantigen Vaccine Generates Intratumoral T Cell Responses in Phase Ib Glioblastoma Trial. *Nature* 565, 234–239. doi:10.1038/s41586-018-0792-9
- Kim, S., Kim, H. S., Kim, E., Lee, M. G., Shin, E.-C., Paik, S., et al. (2018). Neopepsee: Accurate Genome-Level Prediction of Neoantigens by Harnessing Sequence and Amino Acid Immunogenicity Information. *Ann. Oncol.* 29, 1030–1036. doi:10.1093/annonc/mdy022
- Koşaloğlu-Yalçın, Z., Blazeska, N., Carter, H., Nielsen, M., Cohen, E., Kufe, D., et al. (2021). The Cancer Epitope Database and Analysis Resource: A Blueprint for the Establishment of a New Bioinformatics Resource for Use by the Cancer Immunology Community. *Front. Immunol.* 12, 735609. doi:10.3389/fimmu.2021.735609
- Lambricht, L., Vanvarenberg, K., De Beuckelaer, A., Van Hoecke, L., Grooten, J., Ucakar, B., et al. (2016). Coadministration of a Plasmid Encoding HIV-1 Gag Enhances the Efficacy of Cancer DNA Vaccines. *Mol. Ther.* 24, 1686–1696. doi:10.1038/mt.2016.122
- Lancaster, E. M., Jablons, D., and Kratz, J. R. (2020). Applications of Next-Generation Sequencing in Neoantigen Prediction and Cancer Vaccine Development. *Genet. Test. Mol. Biomarkers* 24, 59–66. doi:10.1089/gtmb.2018.0211
- Laumont, C. M., Vincent, K., Hesnard, L., Audemard, É., Bonnel, É., Laverdure, J.-P., et al. (2018). Noncoding Regions Are the Main Source of Targetable Tumor-Specific Antigens. *Sci. Transl. Med.* 10, eaau5516. doi:10.1126/scitranslmed.aau5516
- Li, L., Goedegebuure, S. P., and Gillanders, W. E. (2017). Preclinical and Clinical Development of Neoantigen Vaccines. *Ann. Oncol.* 28, xii11–xii17. doi:10.1093/annonc/mdx681
- Lopes, A., Vanvarenberg, K., Prét, V., and Vandermeulen, G. (2017). Codon-Optimized P1A-Encoding DNA Vaccine: Toward a Therapeutic Vaccination against P815 Mastocytoma. *Mol. Ther. - Nucleic Acids* 8, 404–415. doi:10.1016/j.omtn.2017.07.011
- Lopes, A., Vanvarenberg, K., Kos, Š., Lucas, S., Colau, D., Van den Eynde, B., et al. (2018). Combination of Immune Checkpoint Blockade with DNA Cancer Vaccine Induces Potent Antitumor Immunity against P815 Mastocytoma. *Sci. Rep.* 8, 15732. doi:10.1038/s41598-018-33933-7
- Łuksza, M., Riaz, N., Makarov, V., Balachandran, V. P., Hellmann, M. D., Solovyyov, A., et al. (2017). A Neoantigen Fitness Model Predicts Tumour Response to Checkpoint Blockade Immunotherapy. *Nature* 551, 517–520. doi:10.1038/nature24473
- Lundegaard, C., Nielsen, M., and Lund, O. (2006). The Validity of Predicted T-Cell Epitopes. *Trends Biotechnol.* 24, 537–538. doi:10.1016/j.tibtech.2006.10.001
- Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., and Nielsen, M. (2008a). NetMHC-3.0: Accurate Web Accessible Predictions of Human, Mouse and Monkey MHC Class I Affinities for Peptides of Length 8–11. *Nucleic Acids Res.* 36, W509–W512. doi:10.1093/nar/gkn202
- Lundegaard, C., Lund, O., and Nielsen, M. (2008b). Accurate Approximation Method for Prediction of Class I MHC Affinities for Peptides of Length 8, 10 and 11 Using Prediction Tools Trained on 9mers. *Bioinformatics* 24, 1397–1398. doi:10.1093/bioinformatics/btn128
- Montemurro, A., Schuster, V., Povlsen, H. R., Bentzen, A. K., Jurtz, V., Chronister, W. D., et al. (2021). NetTCR-2.0 Enables Accurate Prediction of TCR-Peptide Binding by Using Paired TCR $\alpha$  and  $\beta$  Sequence Data. *Commun. Biol.* 4, 1–13. doi:10.1038/s42003-021-02610-3
- Nielsen, M., and Lund, O. (2009). NN-align. An Artificial Neural Network-Based Alignment Algorithm for MHC Class II Peptide Binding Prediction. *BMC Bioinformatics* 10, 296. doi:10.1186/1471-2105-10-296
- Nielsen, M., Lundegaard, C., Worning, P., Laumøller, S. L., Lamberth, K., Buus, S., et al. (2003). Reliable Prediction of T-Cell Epitopes Using Neural Networks with Novel Sequence Representations. *Protein Sci.* 12, 1007–1017. doi:10.1110/ps.0239403
- Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., et al. (2007a). NetMHCpan, a Method for Quantitative Predictions of Peptide Binding to Any HLA-A and -B Locus Protein of Known Sequence. *PLoS One* 2, e796. doi:10.1371/journal.pone.0000796
- Nielsen, M., Lundegaard, C., and Lund, O. (2007b). Prediction of MHC Class II Binding Affinity Using SMM-Align, a Novel Stabilization Matrix Alignment Method. *BMC Bioinformatics* 8, 1–12. doi:10.1186/1471-2105-8-238
- Ott, P. A., Hu, Z., Keskin, D. B., Shukla, S. A., Sun, J., Bozym, D. J., et al. (2017). An Immunogenic Personal Neoantigen Vaccine for Patients with Melanoma. *Nature* 547, 217–221. doi:10.1038/nature22991
- Ott, P. A., Hu-Lieskovan, S., Chmielowski, B., Govindan, R., Naing, A., Bhardwaj, N., et al. (2020). A Phase Ib Trial of Personalized Neoantigen Therapy Plus Anti-PD-1 in Patients with Advanced Melanoma, Non-small Cell Lung Cancer, or Bladder Cancer. *Cell* 183, 347–362.e24. doi:10.1016/j.cell.2020.08.053
- Peng, M., Mo, Y., Wang, Y., Wu, P., Zhang, Y., Xiong, F., et al. (2019). Neoantigen Vaccine: An Emerging Tumor Immunotherapy. *Mol. Cancer* 18, 1–14. doi:10.1186/s12943-019-1055-6
- Peters, B., and Sette, A. (2005). Generating Quantitative Models Describing the Sequence Specificity of Biological Processes with the Stabilized Matrix Method. *BMC Bioinformatics* 6, 1–9. doi:10.1186/1471-2105-6-132
- Peters, B., Bulik, S., Tampe, R., van Endert, P. M., and Holzhütter, H.-G. (2003). Identifying MHC Class I Epitopes by Predicting the TAP Transport Efficiency of Epitope Precursors. *J. Immunol.* 171, 1741–1749. doi:10.4049/jimmunol.171.4.1741
- Platten, M., Bunse, L., Wick, A., Bunse, T., Le Cornet, L., Harting, I., et al. (2021). A Vaccine Targeting Mutant IDH1 in Newly Diagnosed Glioma. *Nature* 592, 463–468. doi:10.1038/s41586-021-03363-z
- Probst, P., Kopp, J., Oxenius, A., Colombo, M. P., Ritz, D., Fugmann, T., et al. (2017). Sarcoma Eradication by Doxorubicin and Targeted TNF Relies upon CD8+ T-Cell Recognition of a Retroviral Antigen. *Cancer Res.* 77, 3644–3654. doi:10.1158/0008-5472.CAN-16-2946

- Rekoske, B. T., Smith, H. A., Olson, B. M., Maricque, B. B., and McNeel, D. G. (2015). PD-1 or PD-L1 Blockade Restores Antitumor Efficacy Following SSX2 Epitope-Modified DNA Vaccine Immunization. *Cancer Immunol. Res.* 3, 946–955. doi:10.1158/2326-6066.CIR-14-0206
- Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data. *Nucleic Acids Res.* 48, W449–W454. doi:10.1093/NAR/GKAA379
- Riley, T. P., Keller, G. L. J., Smith, A. R., Davancaze, L. M., Arbujo, A. G., Devlin, J. R., et al. (2019). Structure Based Prediction of Neoantigen Immunogenicity. *Front. Immunol.* 10, 2047. doi:10.3389/fimmu.2019.02047
- Robert, C., Schachter, J., Long, G. V., Arance, A., Grob, J. J., Mortier, L., et al. (2015). Pembrolizumab versus Ipilimumab in Advanced Melanoma. *N. Engl. J. Med.* 372, 2521–2532. doi:10.1056/nejmoa1503093
- Sahin, U., Derhovanessian, E., Miller, M., Kloke, B.-P., Simon, P., Löwer, M., et al. (2017). Personalized RNA Mutanome Vaccines Mobilize Poly-specific Therapeutic Immunity against Cancer. *Nature* 547, 222–226. doi:10.1038/nature23003
- Schuster, H., Peper, J. K., Bösmüller, H.-C., Röhle, K., Backert, L., Bilich, T., et al. (2017). The Immunopeptidomic Landscape of Ovarian Carcinomas. *Proc. Natl. Acad. Sci. U.S.A.* 114, E9942–E9951. doi:10.1073/pnas.1707658114
- Shao, W., Pedrioli, P. G. A., Wolski, W., Scurtescu, C., Schmid, E., Vizcaíno, J. A., et al. (2018). The Systemic MHC Atlas Project. *Nucleic Acids Res.* 46, D1237–D1247. doi:10.1093/nar/gkx664
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: The NCBI Database of Genetic Variation. *Nucleic Acids Res.* 29, 308–311. doi:10.1093/nar/29.1.308
- Sidney, J., Assarsson, E., Moore, C., Ngo, S., Pinilla, C., Sette, A., et al. (2008). Quantitative Peptide Binding Motifs for 19 Human and Mouse MHC Class I Molecules Derived Using Positional Scanning Combinatorial Peptide Libraries. *Immunome Res.* 4, 2–14. doi:10.1186/1745-7580-4-2
- Soria-Guerra, R. E., Nieto-Gomez, R., Govea-Alonso, D. O., and Rosales-Mendoza, S. (2015). An Overview of Bioinformatics Tools for Epitope Prediction: Implications on Vaccine Development. *J. Biomed. Inform.* 53, 405–414. doi:10.1016/j.jbi.2014.11.003
- Sturniolo, T., Bono, E., Ding, J., Radrizzani, L., Tuercio, O., Sahin, U., et al. (1999). Generation of Tissue-specific and Promiscuous HLA Ligand Databases Using DNA Microarrays and Virtual HLA Class II Matrices. *Nat. Biotechnol.* 17, 555–561. doi:10.1038/9858
- Tan, X., Li, D., Huang, P., Jian, X., Wan, H., Wang, G., et al. (2020). DbPepNeo: A Manually Curated Database for Human Tumor Neoantigen Peptides. *Database* 2020, 1–10. doi:10.1093/database/baaa004
- Tenzen, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M. M., et al. (2005). Modeling the MHC Class I Pathway by Combining Predictions of Proteasomal Cleavage, TAP Transport and MHC Class I Binding. *Cmls, Cel. Mol. Life Sci.* 62, 1025–1037. doi:10.1007/s00018-005-4528-2
- Villani, A.-C., Sarkizova, S., and Hacohen, N. (2018). Systems Immunology: Learning the Rules of the Immune System. *Annu. Rev. Immunol.* 36, 813–842. doi:10.1146/annurev-immunol-042617-053035
- Vita, R., Mahajan, S., Overton, J. A., Dhandra, S. K., Martini, S., Cantrell, J. R., et al. (2019). The Immune Epitope Database (IEDB): 2018 Update. *Nucleic Acids Res.* 47, D339–D343. doi:10.1093/nar/gky1006
- Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., and Peters, B. (2008). A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. *PLoS Comput. Biol.* 4, e1000048. doi:10.1371/journal.pcbi.1000048
- Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M., et al. (2010). Peptide Binding Predictions for HLA DR, DP and DQ Molecules. *BMC Bioinformatics* 11, 568. doi:10.1186/1471-2105-11-568
- Wang, G., Wan, H., Jian, X., Li, Y., Ouyang, J., Tan, X., et al. (2020). InEo-Epp: A Novel T-Cell HLA Class-I Immunogenicity or Neoantigenic Epitope Prediction Method Based on Sequence-Related Amino Acid Features. *Biomed. Res. Int.* 2020, 1–12. doi:10.1155/2020/5798356
- Wood, M. A., Nguyen, A., Struck, A. J., Ellrott, K., Nellore, A., and Thompson, R. F. (2020). Neoepitope Improves Neoepitope Prediction with Multivariant Phasing. *Bioinformatics* 36, 713–720. doi:10.1093/bioinformatics/btz653
- Wu, Y., Zhai, W., Sun, M., Zou, Z., Zhou, X., Li, G., et al. (2017). A Novel Recombinant Multi-Epitope Vaccine Could Induce Specific Cytotoxic T Lymphocyte Response *In Vitro* and *In Vivo*. *Ppl* 24, 573–580. doi:10.2174/0929866524666170419152700
- Wu, J., Zhao, W., Zhou, B., Su, Z., Gu, X., Zhou, Z., et al. (2018). TSNAdb: A Database for Tumor-specific Neoantigens from Immunogenomics Data Analysis. *Genomics, Proteomics Bioinformatics* 16, 276–282. doi:10.1016/j.gpb.2018.06.003
- Xia, J., Bai, P., Fan, W., Li, Q., Li, Y., Wang, D., et al. (2021). NEPdb: A Database of T-Cell Experimentally-Validated Neoantigens and Pan-Cancer Predicted Neoepitopes for Cancer Immunotherapy. *Front. Immunol.* 12, 644637. doi:10.3389/fimmu.2021.644637
- Zeng, J., Yang, L., Du, H., Xiao, L., Jiang, L., Wu, J., et al. (2007). Gapped BLAST and PSI-BLAST: a New Generation of Protein Database Search Programs. *World J. Microbiol. Biotechnol.* 25, 3389–3402. doi:10.1016/B978-1-4832-3211-9.50009-7
- Zhang, X., Kim, S., Hundal, J., Herndon, J. M., Li, S., Petti, A. A., et al. (2017). Breast Cancer Neoantigens Can Induce CD8+ T-Cell Responses and Antitumor Immunity. *Cancer Immunol. Res.* 5, 516–523. doi:10.1158/2326-6066.CIR-16-0264
- Zhang, G., Chitkushev, L., Olsen, L. R., Keskin, D. B., and Brusic, V. (2021). TANTIGEN 2.0: a Knowledge Base of Tumor T Cell Antigens and Epitopes. *BMC Bioinformatics* 22, 40. doi:10.1186/s12859-021-03962-7
- Zhao, Y., Wei, Z., Yang, H., Li, X., Wang, Q., Wang, L., et al. (2017). Enhance the Anti-renal Carcinoma Effect of a DNA Vaccine Targeting G250 Gene by Co-expression with Cytotoxic T-Lymphocyte Associated antigen-4(CTLA-4). *Biomed. Pharmacother.* 90, 147–152. doi:10.1016/j.biopha.2017.03.015
- Zhou, C., Wei, Z., Zhang, Z., Zhang, B., Zhu, C., Chen, K., et al. (2019). pTuneos: Prioritizing Tumor Neoantigens from Next-Generation Sequencing Data. *Genome Med.* 11, 67. doi:10.1186/s13073-019-0679-x
- Zhou, W.-J., Qu, Z., Song, C.-Y., Sun, Y., Lai, A.-L., Luo, M.-Y., et al. (2019). NeoPeptide: An Immunoinformatic Database of T-Cell-Defined Neoantigens. *Oxford. Database.* doi:10.1093/database/baz128
- Zou, W., Wolchok, J. D., Chen, L., Weiping, Z., Jedd, D., W., and Lieping, C. (2016). PD-L1 (B7-H1) and PD-1 Pathway Blockade for Cancer Therapy: Mechanisms, Response Biomarkers, and Combinations. *Sci. Transl. Med.* 8, 1–34. doi:10.1126/scitranslmed.aad7118.PD-L1

**Conflict of Interest:** Author XK is employed by Beijing Geneworks Technology Co. and Author MZ is employed Beijing Capital Agribusiness Future Biotechnology Co.

The remaining 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.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yu, Wang, Kong, Cao, Zhang, Sun, Liu, Wang, Shen, Bo and Feng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.