



OPEN ACCESS

EDITED BY

Zlatko Trajanoski,
Medical University of Innsbruck, Austria

REVIEWED BY

Martin Löwer,
Translationale Onkologie an der
Universitätsmedizin der Johannes Gutenberg-
Universität Mainz, Germany
Cansu Cimen Bozkus,
Icahn School of Medicine at Mount Sinai,
United States

*CORRESPONDENCE

Nan Deng

✉ Ndeng1@mdanderson.org

Eduardo Vilar

✉ EVilar@mdanderson.org

RECEIVED 01 March 2024

ACCEPTED 03 May 2024

PUBLISHED 21 May 2024

CITATION

Deng N, Sinha KM and Vilar E (2024) MONET:
a database for prediction of neoantigens
derived from microsatellite loci.
Front. Immunol. 15:1394593.
doi: 10.3389/fimmu.2024.1394593

COPYRIGHT

© 2024 Deng, Sinha and Vilar. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

MONET: a database for prediction of neoantigens derived from microsatellite loci

Nan Deng^{1*}, Krishna M. Sinha¹ and Eduardo Vilar^{1,2,3*}

¹Department of Clinical Cancer Prevention, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, ²Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, ³Department of Clinical Cancer Genetics Program, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

Background: Microsatellite instability (MSI) secondary to mismatch repair (MMR) deficiency is characterized by insertions and deletions (indels) in short DNA sequences across the genome. These indels can generate neoantigens, which are ideal targets for precision immune interception. However, current neoantigen databases lack information on neoantigens arising from coding microsatellites. To address this gap, we introduce The Microsatellite Neoantigen Discovery Tool (MONET).

Method: MONET identifies potential mutated tumor-specific neoantigens (neoAgs) by predicting frameshift mutations in coding microsatellite sequences of the human genome. Then MONET annotates these neoAgs with key features such as binding affinity, stability, expression, frequency, and potential pathogenicity using established algorithms, tools, and public databases. A user-friendly web interface (<https://monet.mdanderson.org/>) facilitates access to these predictions.

Results: MONET predicts over 4 million and 15 million Class I and Class II potential frameshift neoAgs, respectively. Compared to existing databases, MONET demonstrates superior coverage (>85% vs. <25%) using a set of experimentally validated neoAgs.

Conclusion: MONET is a freely available, user-friendly web tool that leverages publicly available resources to identify neoAgs derived from microsatellite loci. This systems biology approach empowers researchers in the field of precision immune interception.

KEYWORDS

neoantigen, microsatellite, Lynch syndrome, mismatch repair, somatic mutation, indels

Introduction

Microsatellite instability (MSI) is caused by the accumulation of insertions and deletions (indels) in short-segment DNA sequences of mono-, di-, tri-nucleotide, and longer repeats known as microsatellites due to mismatch repair (MMR) deficiency. MMR deficiency is secondary to inactivating mutations in one of the four MMR genes (*MLH1*, *MSH2*, *MHS6*, and *PMS2*) or epigenetic silencing of *MLH1* (sporadic MMR deficiency) (1). These indels lead to frameshift mutations, thus resulting in the generation of mutated neoantigens (neoAgs) that are unique to tumor cells and highly unlikely to be found in normal cells. Unlike non-synonymous single-nucleotide variants (SNVs), which are mutations that typically generate neoAgs with only one altered amino acid(s), frameshift mutations typically generate completely different amino acid sequences. These frameshifted sequences have low probabilities of being tolerated by the host's immune system. These mutated frameshift proteins possess intrinsic immunogenicity and are, therefore, attractive targets for cancer interception and therapy (2, 3).

Currently, there are several epitope databases available to facilitate *in silico* vaccine design (Table 1). The Immune Epitope Database (IEDB) (4) is a globally accessible gateway to experimentally validated immune epitopes, while other databases, such as Antijen (5) and caped (6) focus on curated cancer epitopes from research manuscripts. In addition, the GNIFdb (7) and TSNAdb databases (8, 9) utilize computational approaches to predict putative neoAg based on high-frequency mutations detected in cancers. However, there is currently no antigen database dedicated to neoAgs derived from microsatellite tracts. Tumors displaying high levels of MSI (MSI-H) may harbor indels in up to 80% of microsatellite loci (10), thus suggesting that coding MSI could generate a significant number of potential neoAg candidates with high degree of immunogenicity. Leveraging this knowledge gap, we developed a new database named MicrOsatellite Neoantigen Discovery Tool (MONET) that focuses on the prediction of putative neoAgs derived from MSI in cancers.

Here, we introduce the MONET database, which includes all possible neoAg derived from microsatellite loci present in the human reference genome. To generate MONET, we used computational algorithms to predict neoAg-derived epitopes with high affinity for CD8⁺ and CD4⁺ T cells that are specific to a variety of MHC-I and MHC-II alleles, respectively. We also evaluated the binding stability and foreignness of predicted epitopes, which are crucial factors in assessing their immunogenicity. Furthermore, we integrated data on gene expression levels of neoAgs in different tumor types leveraging the TCGA database, and mutation allele frequencies from public databases in order to provide a comprehensive immunogenicity and population coverage. Our work demonstrates that MONET has excellent coverage and outperforms other available antigen databases when tested against a set of verified epitopes derived from microsatellites. In addition, a user-friendly web-interface has been implemented and housed at <https://monet.mdanderson.org/>,

Abbreviations: MONET, The Microsatellite Neoantigen Discovery Tool; neoAg, Neoantigen; MHC, Major Histocompatibility Complex; MMR, Mismatch Repair; MSI, Microsatellite Instability.

where users can query candidate target genes and obtain curated lists of potential neoAg epitopes based on their corresponding MHC alleles without the need for complex computational efforts.

Methods

Generation of potential frameshift neoantigens

To generate potential frameshift neoAg derived from microsatellite loci, we utilized MSIensor2 (<https://github.com/niu-lab/msisensor2>) (11) to scan the human reference genome (GRCh38) for the identification of all microsatellite loci. Short nucleotide repeats exceeding five units were identified as microsatellites. For larger repeated motifs (ranging from 2 to 5 base pairs), a minimum repeat number of three was used. Only microsatellites within protein-coding regions were retained to generate potential mutant proteins. Insertions/deletions differing by a multiple of 3 will share the same reading frame, resulting in identical downstream sequences, so that we can generate two types of downstream frameshift sequences for each identified microsatellite: $3n+1$ and $3n+2$ shifts of nucleotide bases, where n is an integer ($-3, -2, -1, 0, 1, 2, 3$, and so on). For example, in the $3n+1$ series, two nucleotide deletions ($n = -1$ and $3n+1 = -2$) in the sequence will share the same reading frame with one nucleotide insertion ($n = 0$, $3n+1 = 1$), and four nucleotide insertions ($n = 1$, $3n+1 = 4$) and so on. This applies similarly to the $3n+2$ series. Therefore, to efficiently represent these frameshift mutations, we introduce two *in silico* variants for each microsatellite: a two-nucleotide deletion ($n = -1$, $3n+1 = -2$) and a one-nucleotide deletion ($n = -1$, $3n+2 = -1$). These variants

TABLE 1 Available antigen databases.

Database	URL	Type	Target
GNIFdb	http://www.oncoimmunobank.cn/index.php	Computational	Glioma
IEDB	https://www.iedb.org/	Curated	General
TSNAdb v2.0	https://pgx.zju.edu.cn/tsnadb/	Computational	Pan cancer
dbPepNeo2.0	http://119.3.70.71/dbPepNeo2/home.html	Curated	Pan cancer
CAD v1.0	http://cad.bio-it.cn/	Curated	Pan Cancer
NeoPeptide	Not available	Curated	Pan Cancer
NEPdb	http://nep.whu.edu.cn/	Curated	Pan Cancer
TANTIGEN 2.0	http://projects.methilab.org/tadb	Curated	Pan Cancer
Antijen	http://www.ddg-pharmfac.net/antijen/Antijen/antijenhomepage.htm	Curated	General
caped	https://caped.icp.ucl.ac.be/	Curated	General

encompass the spectrum of frameshift mutations within each series and generate entirely distinct downstream amino acid sequences compared to the wild-type sequence, thus making them ideal neoAg candidates.

Putative neoantigen epitopes

It is important to note the significant complexity of potential mutant amino acid sequences generated at the junction region, where the wild-type and mutant sequences meet within the microsatellite region. The size and location of indels at this junction significantly impact the resulting mutant sequence. However, based on our previous experimental data (10), very few (<5%) verified neoAg epitopes originate from these junction regions. Therefore, we excluded the sequences around the junction at which the frameshifted amino acid occurred to prevent this complexity. Then, putative neoAg epitopes were predicted towards a panel of higher frequency MHC-Class I (12) and MHC-Class II (13) alleles (Table 2) providing coverage for 97% and 99% of the general population, respectively. Various algorithms such as MHCflurry (14), MHCnuggets (15), NetMHC (16), PickPocket (17), SMM-align (18), NNalign (19) that are implemented in pVACtools (20) were employed to predict neoAgs. Any epitope and allele pairs with a binding affinity of $IC_{50} < 50$ nM in any algorithm were considered potential epitopes for subsequent processes. We predicted the binding stability for Class I epitopes using NetMHCstab (21) and assessed foreignness to the human proteome using antigen.garnish (<https://github.com/andrewrech/antigen.garnish>) (22). Also, other characteristics that could play an important role in

epitope selection such as terminal amino acids and the Gravy score (average hydrophathy) were annotated.

Annotation of neoantigens

To gain a deeper understanding of the potential epitopes, we also annotated the corresponding mutations using the Ensembl Variant Effect Predictor (VEP) (23). Annotations include associated gene names, IDs, genomic coordinates, and among others. Another critical factor for the effectiveness of neoAgs being presented by MHC is their expression level (24–26). Higher expression levels will increase the probability of these epitopes being presented by MHC molecules. We integrated the expression levels of the corresponding genes generating the neoAgs from the Cancer Genome Atlas Program (TCGA, <https://www.cancer.gov/tcga>) project by using all datasets that contain pairs of normal and tumor. Additionally, for the dataset with MSI status information, we distinguished and listed the MSI-H and MSS groups separately. Differentially expressed genes (Benjamini-Hochberg adjusted p-value < 0.05) between normal and cancer tissues in each of the different cancer data set were labelled. Furthermore, we annotated the frequency of the mutation and associated phenotypes using the dbSNP (RRID: SCR_002338) (27) and ClinVar (RRID:SCR_006169) database (28). This annotation provided valuable insights into the frequency and clinical implications of the mutations from where the neoAgs are derived. Moreover, we included evidence of experimentally confirmed epitopes using IEDB (RRID:SCR_006604), which contributes to the reliability and validity of the identified epitopes.

TABLE 2 MHC Alleles used to predict putative neoantigens.

Class I Allele tested (n=27)	Class II Allele tested (n=27)
HLA-A*01:01	DRB1*01:01
HLA-A*02:01	DRB1*03:01
HLA-A*02:03	DRB1*04:01
HLA-A*02:06	DRB1*04:05
HLA-A*03:01	DRB1*07:01
HLA-A*11:01	DRB1*08:02
HLA-A*23:01	DRB1*09:01
HLA-A*24:02	DRB1*11:01
HLA-A*26:01	DRB1*12:01
HLA-A*30:01	DRB1*13:02
HLA-A*30:02	DRB1*15:01
HLA-A*31:01	DRB3*01:01
HLA-A*32:01	DRB3*02:02
HLA-A*33:01	DRB4*01:01
HLA-A*68:01	DRB5*01:01
HLA-A*68:02	DQA1*05:01-DQB1*02:01
HLA-B*07:02	DQA1*05:01-DQB1*03:01
HLA-B*08:01	DQA1*03:01-DQB1*03:02
HLA-B*15:01	DQA1*04:01-DQB1*04:02
HLA-B*35:01	DQA1*01:01-DQB1*05:01
HLA-B*40:01	DQA1*01:02-DQB1*06:02
HLA-B*44:02	DPA1*02:01-DPB1*01:01
HLA-B*44:03	DPA1*01:03-DPB1*02:01
HLA-B*51:01	DPA1*01:03-DPB1*04:01
HLA-B*53:01	DPA1*03:01-DPB1*04:02
HLA-B*57:01	DPA1*02:01-DPB1*05:01
HLA-B*58:01	DPA1*02:01-DPB1*14:01

MONET website infrastructure

We constructed the website using a microservices architecture with Docker. The backend database (MySQL) was employed to store and manage data. Express.js served as the middleware, responsible for translating HTTP requests into MySQL queries. Vue.js was utilized to develop the user interface. Nginx served as the HTTP server communicating between the containers hosting the frontend and backend components of the website.

Comparisons of the epitope coverage derived from microsatellite loci across various databases

We selected 100 epitopes derived from mutations in microsatellite loci among the top-ranked based on their predicted immunogenicity (Supplementary Table S2) in LS patients and showed that 65 out of 100 predicted neoAg candidates were validated for their immunogenicity using *in vitro* ELISPOT assays (10). These peptides were used to assess the coverage across different databases. Specifically, on April 9th, 2024, we individually searched the 65 validated immunogenic epitopes

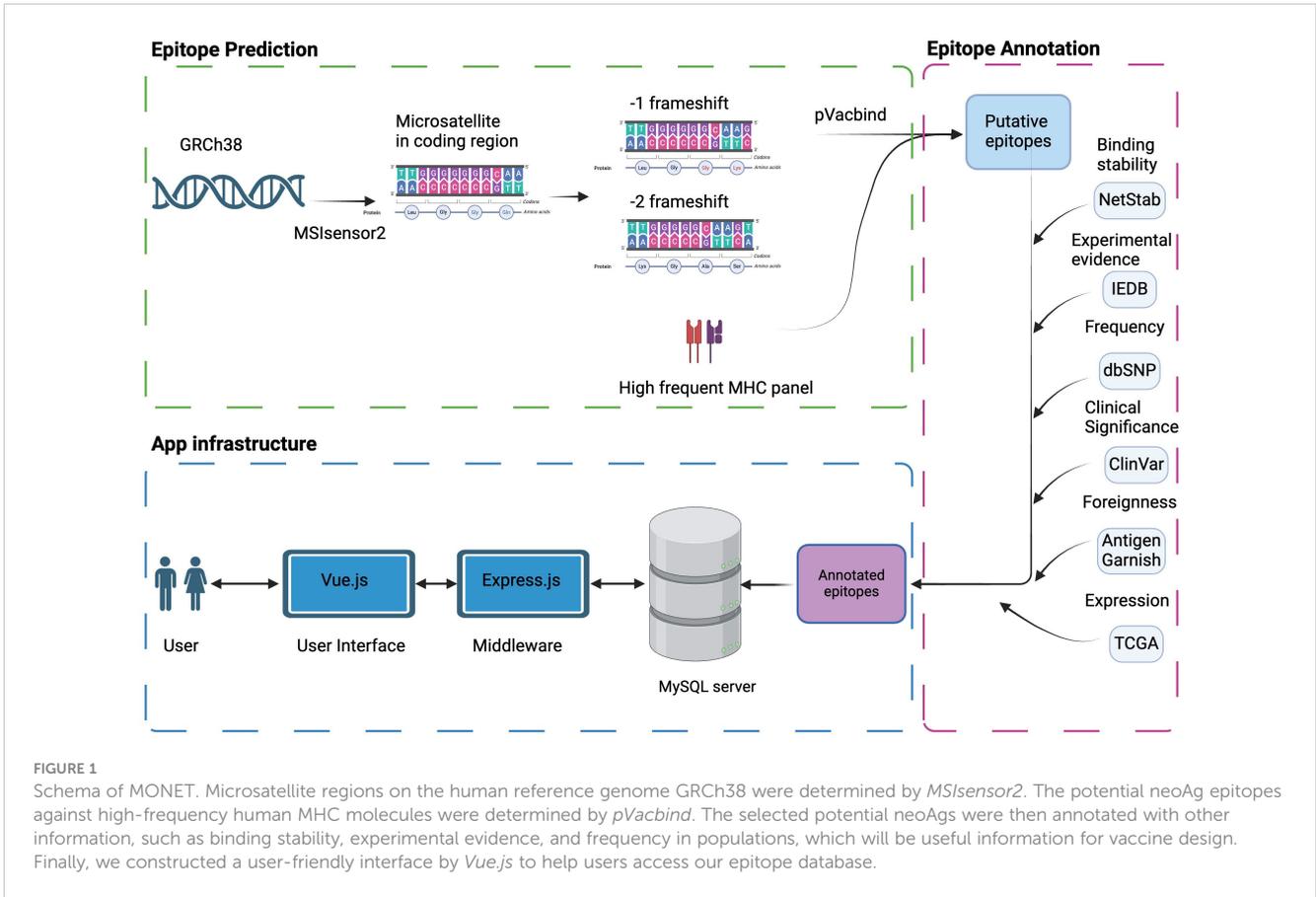


FIGURE 1

Schema of MONET. Microsatellite regions on the human reference genome GRCh38 were determined by *MSIsensor2*. The potential neoAg epitopes against high-frequency human MHC molecules were determined by *pVacbind*. The selected potential neoAgs were then annotated with other information, such as binding stability, experimental evidence, and frequency in populations, which will be useful information for vaccine design. Finally, we constructed a user-friendly interface by *Vue.js* to help users access our epitope database.

within MONET and the 10 databases listed in Table 1. Then, we evaluated the number of epitopes recorded in each database. The services of dbPepNeo, NeoPeptide, and TANTIGEN databases were unavailable at the time of our research. Therefore, we reported results from MONET and the other 7 databases.

Data Availability

Public data analyzed in this study was obtained from multiple sources including: <https://www.ncbi.nlm.nih.gov/genome/>, <https://www.ncbi.nlm.nih.gov/snp/>, <https://www.ncbi.nlm.nih.gov/clinvar/>,

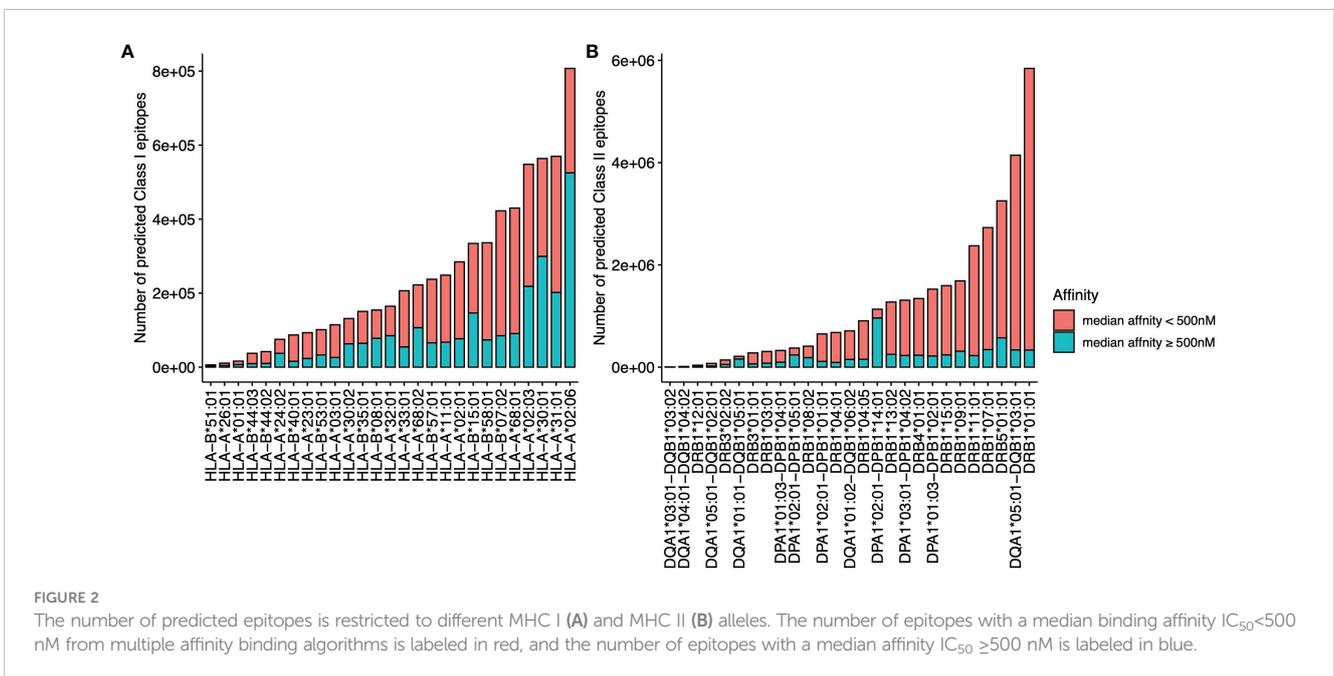


FIGURE 2

The number of predicted epitopes is restricted to different MHC I (A) and MHC II (B) alleles. The number of epitopes with a median binding affinity $IC_{50} < 500$ nM from multiple affinity binding algorithms is labeled in red, and the number of epitopes with a median affinity $IC_{50} \geq 500$ nM is labeled in blue.

<https://www.iedb.org/>, <https://www.cancer.gov/tcga>. The epitope prediction data in this study are available at <https://monet.mdanderson.org>.

Results

Prediction of frameshift neoantigens in microsatellites

The overall data processing is depicted in **Figure 1**. We identified a total of 34,067,744 microsatellite loci in the human genome GRCh38. Among them, only 3,934,634 microsatellites were located in protein-coding regions (**Table 3**). After eliminating duplicated sequences, we obtained 492,578 unique frameshift mutation neoAg sequences (**Table 3**). From these sequences, we identified 4,449,128 MHC Class I and 15,589,846 MHC Class II potential epitopes. The number of epitopes binding to MHC Class I alleles for each allele ranged from approximately 6,000 to 800,000 (**Figure 2A**), while binders to Class II alleles displayed a higher number of predicted epitopes ranging from approximately 5,000 to 6 million (**Figure 2B**).

Annotation of epitopes and mutations

After generation of predicted potential epitopes using our pipeline, the binding stability was annotated using antigen.garish (**Figure 1**). Then, the expression levels of the corresponding genes carrying the mutations were annotated using TCGA data. A total of 64,288 epitopes have been experimentally verified and recorded in the IEDB (**Table 3**). A total of 542,442 mutations in MONET were recorded in the dbSNP database, which contains human variants

including small indels original from germline or somatic mutations. Of these mutations, 13,289 are linked to the ClinVar database, which associates human variation with their potentially clinically relevant results. While most of the variant recorded in dbSNP and ClinVar are of germline origin, 198 somatic mutations in MONET have been recorded in ClinVar. The top ClinVar phenotypes (**Table S1**) include malignant tumor of prostate (Rank 1, n = 31), carcinoma of colon (Rank 2, n= 24), and colorectal cancer (Rank 5, n=16). In one hand, these clinical records verified our putative neoAgs that might present in dMMR/MSI-H cancers. In the other hand, the limited coverage of the current clinically available databases suggests that our computational results could be very valuable in bridging this gap.

Web interface

In MONET, the landing page can be accessed at <https://monet.mdanderson.org>. The key functions of MONET are: ‘Search Neoantigens’ and ‘Best Neoantigens’. Both can be accessed from the sidebar.

Search neoantigens.

In the ‘Search Neoantigens’ page, users have the option to search for neoAg using various parameters, including gene symbol and related IDs, epitope sequence, mutation HVGSp/HVGSc ID, and ClinVar phenotype. Moreover, users can narrow down their search by limiting it to specific MHC class I and II types. This functionality proves particularly useful when users have a specific target gene or disease in mind and would like to search into the details of a particular epitope (**Figure 3A**).

As an example, users can utilize the search function to find potential Class I neoAg epitopes resulting from mutations in *TGFBR2*. The search engine generates 520 summarized results, providing information such as peptide sequence, mutation location, related gene, and associated HLA (Human Leukocyte Antigen) alleles. To further refine the results, users can utilize the filter sidebar on the right-hand side. This allows filtering based on peptide sequence, affinity cutoff, and target HLA alleles. In cases where multiple genes are returned, users can also apply a filter based on the gene symbol. Once a specific epitope of interest is identified, users can select it to further access detailed information about the epitope (**Figure 3B**).

Detailed epitope information is offered in five tabs, each providing specific details (**Figure 4**): 1. *Prediction affinity* will display a table of predicted affinities generated by different algorithms for various HLA types (**Figure 4**) and other features of the epitope such as the Gravy score; 2. *The mutant gene* will show essential gene information related to the epitope, including the gene’s location, related IDs, and both wild-type and mutant sequences of the protein. The mutant sequence is highlighted in red (**Supplementary Figure S1**); 3. *ClinVar* and *dbSNP* will provide access to the gene ID and corresponding link to ClinVar

TABLE 3 Statistics of the MONET database.

	Count
Microsatellites on reference genome	34,067,744
Microsatellites on exon regions	3,934,634
Genes with microsatellites on exon regions	18,591
Potential mutations	11,803,902
No duplicated downstream sequence	492,578
Possible Class I epitopes (best affinity <50nM)	4,449,128
Possible Class II epitopes (best affinity <50nM)	15,589,846
Mutations in dbSNP	542,442
Mutations in ClinVar	13,289
Iedb experiment verified epitope	65,220
Iedb experiment verified TCR and MHC information	932
Iedb experiment verified MHC ligand	64,288

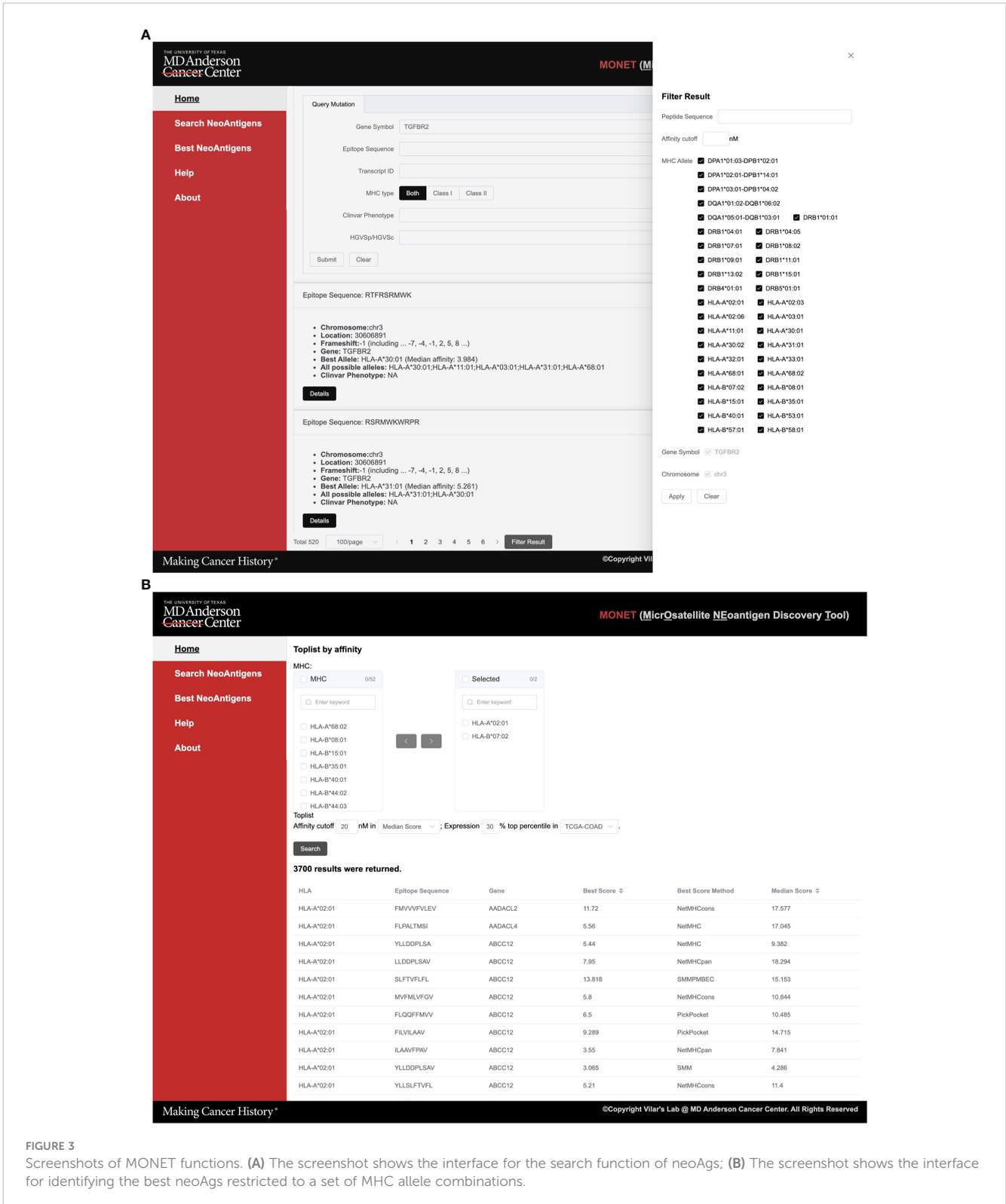


FIGURE 3 Screenshots of MONET functions. (A) The screenshot shows the interface for the search function of neoAgs; (B) The screenshot shows the interface for identifying the best neoAgs restricted to a set of MHC allele combinations.

and dbSNP databases. Additionally, if available, basic information about related diseases and frequency data will be presented (Supplementary Figure S2); 4. If the peptide has been experimentally tested in the IEDB (Immune Epitope Database), this tab will showcase

relevant information from IEDB; 5. Users can explore the expression of the target gene in solid normal tissue versus primary solid tumors across different TCGA databases. Thus, this tab displays a bar plot illustrating this information (Supplementary Figure S3).

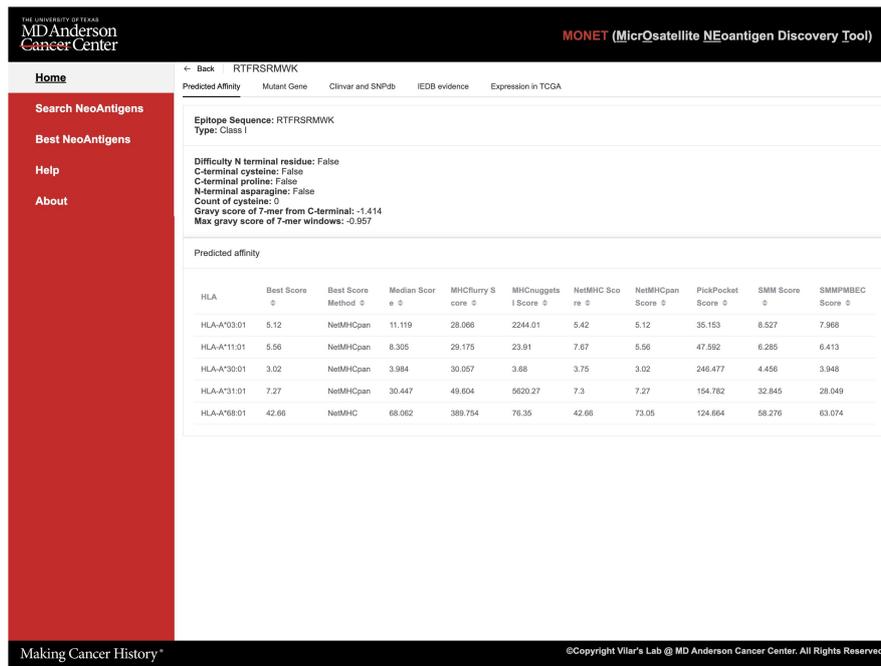


FIGURE 4 Screenshot of detailed reports of the epitope result page.

Best neoantigens

Within the best neoAg search engine function, users have the capability to search for all potential epitopes specific to one or multiple HLA alleles (Figure 3B). This feature allows for customized filtering options, such as an affinity and expression cutoff in a specific TCGA cancer type. This functionality proves particularly valuable when users intend to identify the top potential epitopes for a specific set of HLA alleles from either an individual patient or a group of patients.

As an example, users can select HLA-A*02:01 and HLA-B*07:02 alleles along with a median affinity cutoff of 30 nM across multiple algorithms. Furthermore, they can narrow down the search to include only the top 20% most highly expressed genes in the COAD dataset (Figure 3B). In total, 3,186 results for HLA-A*02:01 and 2,614 results for HLA-B*07:02 returned. These potential epitopes are generated from mutations in 1,483 genes. Users have the option to download the results directly from the page, and detailed information for each peptide can be accessed on the corresponding peptide details page.

Evaluating the coverage of neoantigens derived from microsatellite loci across various databases

The most distinctive feature of MONET is its focus on epitopes derived from microsatellite loci that are targets of MMRd. We compared the coverage of 65 of such epitopes, the immunogenicity of which has

been validated using ELISpot assays (10), across MONET and other popular databases. MONET covers 57 out of these 65 verified epitopes (Figure 5, Supplementary Table S2). In comparison to MONET, TSNAdb, only covers 15 of the 65 epitopes, and IEDB covers just one. GNIFdb, CAD, NEPdb, CAPAD, and Antijen do not contain any of these epitopes. Therefore, MONET demonstrates excellent coverage (>85%) of our target epitopes, which are derived from microsatellite loci due to MMRd, thus significantly outperforming other available databases in this field.

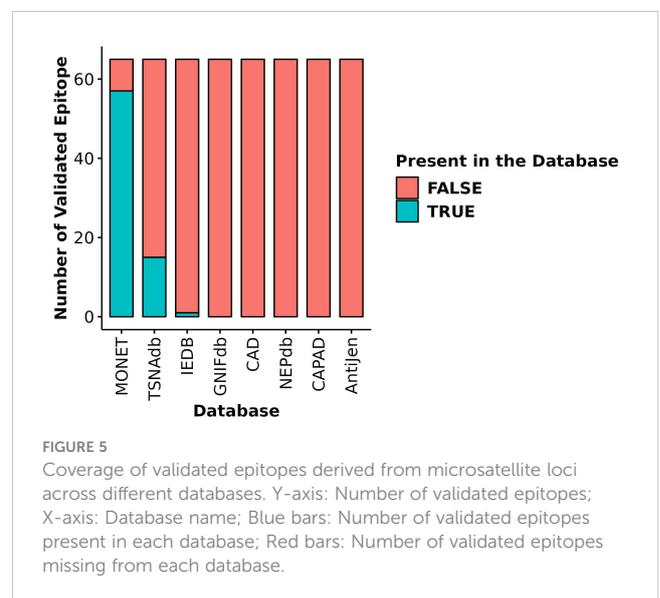


FIGURE 5 Coverage of validated epitopes derived from microsatellite loci across different databases. Y-axis: Number of validated epitopes; X-axis: Database name; Blue bars: Number of validated epitopes present in each database; Red bars: Number of validated epitopes missing from each database.

Discussion

The MONET database serves as a comprehensive resource for researchers investigating neoAg related to MSI cancers. This database covers possible neoAg derived from microsatellites within the human genome, which is specific to a panel of high-frequency MHC class I and class II alleles. Researchers can utilize the database to conduct searches based on target genes or epitope sequences, as well as target MHC alleles, to obtain comprehensive information on potential target epitopes. The user-friendly interface makes the database accessible to cancer researchers without acquiring any bioinformatic skills.

We acknowledge that MONET has several limitations. MONET solely focuses on target epitopes for humans, but we have ongoing efforts to broaden its scope and incorporate epitopes for other model systems such as mouse, rat, and rhesus, which will be particularly valuable for cancer vaccine studies using model organisms. MONET uses multiple MHC-peptide binding affinity algorithms to identify potential neoAgs. We currently treat all algorithms equally and calculate the minimum, median, or mean value of these algorithms to evaluate the peptides. However, these algorithms have large differences in performance (29, 30). In the future, we plan to exclude those algorithms with inferior performance and prioritize the weight of algorithms with the best performance based on data generated by our team or public data sets. MONET currently lacks a REST API (Representational State Transfer Application Programming Interface), which could streamline data access and analysis for bioinformaticians seeking customized or bulk queries. A REST API is planned for integration in the upcoming MONET release.

In summary, MONET is a systems biology tool that has the goal of facilitating the identification of mutated neoAgs derived from microsatellite loci by leveraging publicly available state-of-the-art tools and by providing a user-friendly online website that is freely available to the scientific community.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

Author contributions

ND: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. KS: Writing – original draft, Writing – review & editing. EV: Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by a gift from the Feinberg Family Foundation and grants R01CA260761, R01CA257375 and U01 CA233056 (US National Institutes of Health/National Cancer Institute) to EV; the generous philanthropic contributions to The University of Texas MD Anderson Cancer Center Moon Shots Program, The MD Anderson Cancer Center SPORE in Gastrointestinal Cancer P50 CA221707 (US National Institutes of Health/National Cancer Institute); and P30 CA016672 (US National Institutes of Health/National Cancer Institute) to the University of Texas MD Anderson Cancer Center Core Support Grant.

Acknowledgments

We thank Gaston Benavides Jr., Jenny Chen, and Rui Jiang for their IT support. We also thank Greg Holland and Ana Bolivar for their testing and suggestions on the webpage design.

Conflict of interest

EV has a consulting or advisory role with Janssen Research and Development, Recursion Pharma, Guardant Health, The Rising Tide Foundation, and Nouscom, s.r.l. EV has received research support from Janssen Research and Development.

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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2024.1394593/full#supplementary-material>

References

- Vilar E, Gruber SB. Microsatellite instability in colorectal cancer—the stable evidence. *Nat Rev Clin Oncol*. (2010) 7(3):153–62. doi: 10.1038/nrclinonc.2009.237
- Schumacher TN, Scheper W, Kvistborg P. Cancer neoantigens. *Annu Rev Immunol*. (2019) 37:173–200. doi: 10.1146/annurev-immunol-042617-053402
- Lang F, Schrors B, Lower M, Tureci O, Sahin U. Identification of neoantigens for individualized therapeutic cancer vaccines. *Nat Rev Drug Discov*. (2022) 21(4):261–82. doi: 10.1038/s41573-021-00387-y
- Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res*. (2019) 47(D1):D339–43. doi: 10.1093/nar/gky1006
- Toseland CP, Clayton DJ, McSparron H, Hemsley SL, Blythe MJ, Paine K, et al. Antigen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res*. (2005) 1(1):4. doi: 10.1186/1745-7580-1-4
- Hutchison S, Pritchard AL. Identifying neoantigens for use in immunotherapy. *Mamm Genome*. (2018) 29(11–12):714–30. doi: 10.1007/s00335-018-9771-6
- Li W, Sun T, Li M, He Y, Li L, Wang L, et al. GNIFdb: a neoantigen intrinsic feature database for glioma. *Database (Oxford)*. (2022) 2022. doi: 10.1093/database/baac004
- Wu J, Chen W, Zhou Y, Chi Y, Hua X, Wu J, et al. TSNAdb v2.0: the updated version of tumor-specific neoantigen database. *Genomics Proteomics Bioinf*. (2023) 21(2):259–66. doi: 10.1101/2022.07.28.501872
- Wu J, Zhao W, Zhou B, Su Z, Gu X, Zhou Z, et al. TSNAdb: A database for tumor-specific neoantigens from immunogenomics data analysis. *Genomics Proteomics Bioinf*. (2018) 16(4):276–82. doi: 10.1016/j.gpb.2018.06.003
- Bolivar AM, Duzagac F, Deng N, Reyes-Urbe L, Chang K, Wu W, et al. Genomic landscape of lynch syndrome colorectal neoplasia identifies shared mutated neoantigens for immunoprevention. *Gastroenterology*. (2024) 166(5):787–801.e11. doi: 10.1053/j.gastro.2024.01.016
- Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, et al. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics*. (2014) 30(7):1015–6. doi: 10.1093/bioinformatics/btt755
- Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics*. (2011) 63(6):325–35. doi: 10.1007/s00251-011-0513-0
- Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc Natl Acad Sci USA*. (2013) 110(22):E2046–53. doi: 10.1073/pnas.1305227110
- O'Donnell TJ, Rubinsteyn A, Laserson U. MHCflurry 2.0: improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Syst*. (2020) 11(4):418–9. doi: 10.1016/j.cels.2020.09.001
- Shao XM, Bhattacharya R, Huang J, Sivakumar IKA, Tokheim C, Zheng L, et al. High-throughput prediction of MHC class I and II neoantigens with MHCnuggets. *Cancer Immunol Res*. (2020) 8(3):396–408. doi: 10.1158/2326-6066.CIR-19-0464
- Lundegaard C, Lamberth C, Harndahl M, Buus S, Lund O, Nielsen M. 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*. (2008) 36:W509–12. doi: 10.1093/nar/gkn202
- Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics*. (2009) 25(10):1293–9. doi: 10.1093/bioinformatics/btp137
- Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinf*. (2007) 8:238. doi: 10.1186/1471-2105-8-238
- Nielsen M, Andreatta M. NNAlign: a platform to construct and evaluate artificial neural network models of receptor-ligand interactions. *Nucleic Acids Res*. (2017) 45(W1):W344–9. doi: 10.1093/nar/gkx276
- Hundal J, Kiwala S, McMichael J, Miller CA, Xia H, Wollam AT, et al. pVACtools: A computational toolkit to identify and visualize cancer neoantigens. *Cancer Immunol Res*. (2020) 8(3):409–20. doi: 10.1158/2326-6066.CIR-19-0401
- Jorgensen KW, Rasmussen M, Buus S, Nielsen M. NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology*. (2014) 141(1):18–26. doi: 10.1111/imm.12160
- Richman LP, Vonderheide RH, Rech AJ. Neoantigen dissimilarity to the self-proteome predicts immunogenicity and response to immune checkpoint blockade. *Cell Syst*. (2019) 9(4):375–382.e4. doi: 10.1016/j.cels.2019.08.009
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol*. (2016) 17(1):122. doi: 10.1186/s13059-016-0974-4
- Westcott PMK, Sacks NJ, Schenkel JM, Ely ZA, Smith O, Hauck H, et al. Low neoantigen expression and poor T-cell priming underlie early immune escape in colorectal cancer. *Nat Cancer*. (2021) 2(10):1071–85. doi: 10.1038/s43018-021-00247-z
- Borden ES, Ghafoor S, Buetow KH, LaFleur BJ, Wilson MA, Hastings KT. NeoScore integrates characteristics of the neoantigen:MHC class I interaction and expression to accurately prioritize immunogenic neoantigens. *J Immunol*. (2022) 208(7):1813–27. doi: 10.4049/jimmunol.2100700
- Wells DK, van Buuren MM, Dang KK, Hubbard-Lucey VM, Sheehan KCF, Campbell KM, et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell*. (2020) 183(3):818–34.e13. doi: 10.1016/j.cell.2020.09.015
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. (2001) 29(1):308–11. doi: 10.1093/nar/29.1.308
- Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. (2018) 46(D1):D1062–7. doi: 10.1093/nar/gkx1153
- Bonsack M, Hoppe S, Winter J, Tichy D, Zeller C, Kupper MD, et al. Performance evaluation of MHC class-I binding prediction tools based on an experimentally validated MHC-peptide binding data set. *Cancer Immunol Res*. (2019) 7(5):719–36. doi: 10.1158/2326-6066.CIR-18-0584
- Mei S, Li F, Leier A, Marquez-Lago TT, Giam K, Croft NP, et al. A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. *Brief Bioinform*. (2020) 21(4):1119–35. doi: 10.1093/bib/bbz051