

# Desarrollo de una Aplicación Web para la Detección de Neoantígenos en el Marco de Desarrollo de Vacunas Personalizadas para Tratar el Cáncer

PhD(c). Vicente Machaca Arceda y Ms. Richard Escobedo

17 de agosto de 2023

## 1. Breve estado de la cuestión

El cáncer representa el mayor problema de salud mundial y es la principal causa de muerte, con alrededor de un millón de fallecimientos reportados en 2020. Además, los métodos tradicionales basados en cirugías, radioterapias y quimioterapias tienen baja efectividad (Peng et al., 2019). En este contexto, surge el desarrollo de la inmunoterapia de cáncer, que tiene como objetivo estimular el sistema inmunológico de un paciente (Borden et al., 2022). En esta área, ha emergido la investigación basada en la detección de neoantígenos, hay tres tratamientos: vacunas personalizadas, terapias de células T adoptivas e inhibidores de puntos de control inmunológico. De los métodos mencionados anteriormente, se considera que el desarrollo de vacunas personalizadas tiene la mayor probabilidad de éxito (Borden et al., 2022).

Los neoantígenos son péptidos mutados específicos del tumor y se consideran las principales causas de una respuesta inmune (Borden et al., 2022; Chen et al., 2021; Gopanenko et al., 2020). El objetivo es entrenar los linfocitos (células T) de un paciente para que reconozcan los neoantígenos y activen el sistema inmunológico (Mattos et al., 2020; Peng et al., 2019). El ciclo de vida de un neoantígeno para las células con núcleo se puede resumir de la siguiente manera. Primero, una proteína se degrada en péptidos (posibles neoantígenos) en el citoplasma. A continuación, los péptidos se unen al Complejo Mayor de Histocompatibilidad (MHC), conocido como unión péptido-MHC (pMHC *binding*). Luego, este compuesto sigue una vía hasta llegar a la membrana celular (pMHC *presentation*). Finalmente, el pMHC es reconocido por el *T-cell Receptor* (TCR), lo que desencadena el sistema inmunológico. En este contexto, este proyecto se centra en la predicción de la unión del MHC con un péptido (pMHC). Si la unión ocurre, entonces el péptido en cuestión podría ser considerado un neoantígeno.

Referente a la predicción de la unión pMHC, existen dos enfoques: *allele-specific* y *pan-specific*; el primero desarrolla modelos por cada tipo de MHC, si consideramos que existen miles de variantes de MHC, este enfoque no es viable; el segundo enfoque *allele-specific*, desarrolla un solo modelo para predecir la unión pMHC. En los siguientes párrafos, explicaremos brevemente los principales trabajos desarrollados hasta la actualidad.

NetMHCpan4.1 (Reynisson et al., 2020) es un método *pan-specific* considerado como una línea base para la predicción de pMHC-I. Este método utiliza Redes Neuronales Artificiales (ANN). Mejoró sus versiones anteriores al aumentar el conjunto de datos de entrenamiento con 13245212 puntos de datos que cubren 250 moléculas distintas de MHC-I; además, el modelo se actualizó de NN\_align a NN\_alignMA (Alvarez et al., 2019). Además, MHCflurry2.0 (Donnell et al., 2020) es otro método de vanguardia; utiliza un predictor de afinidad de unión pan-allelic, un predictor de presentación de antígeno independiente del alelo y utiliza datos de Espectrometría de Masas (MS); después de experimentos, MHCflurry2.0 superó a NetMHCpan4.0. En cuanto a la predicción *pan-specific* de pMHC-II, NetMHCIIpan4.0 (Reynisson et al., 2020) utilizó desconvolución de motif y datos de *eluted ligands* por MS con 4086230 puntos de datos que cubren un total de 116 MHC-II distintos. Por otro lado, NetMHC4.0 (Andreatta and Nielsen, 2016) es *allele-specific*; actualizó sus versiones anteriores, agregando *padding* a los aminoácidos y utilizó ANNs.

Los *transformers* se consideran una revolución en la inteligencia artificial y se han aplicado con éxito en varias tareas de procesamiento del lenguaje natural (NLP, por sus siglas en inglés) (Patwardhan et al., 2023). Además, estos modelos se han utilizado en la detección de neoantígenos, centrándose en la predicción del enlace pMHC. Por ejemplo, BERTMHC (Cheng et al., 2021) es un método *pan-specific* para predecir el enlace pMHC-II; utiliza una arquitectura BERT y transfer learning de *Tasks Assessing Protein Embeddings* (TAPE) (Rao et al., 2019). Los autores aplican una *mean pooling* seguida de una capa *Fully Connected* (FC) después del modelo TAPE. En los experimentos, BERTMHC superó a NetMHCIIpan3.2 y PUFFIN. Además, ImmunoBERT (Gasser et al., 2021) también utilizó transfer learning de TAPE; sin embargo, los autores se enfocaron en la predicción de pMHC-I.

Además, MHCroBERTa (Wang et al., 2022) y HLAB (Zhang et al., 2022) también utilizaron *transfer learning*. MHCroBERTa utilizó entrenamiento auto-supervisado a partir de las bases de datos UniProtKB y Swiss-prot; luego, aplicaron *fine-tuning* con datos de IEDB (Vita et al., 2018). MHCroBERTa superó a NetMHCpan4.0 y MHCflurry2.0 en SRCC. Por otro lado, HLAB (Zhang et al., 2022) utilizó *transfer learning* de ProtBert-BFD (Elnaggar et al., 2021); utilizó un modelo BiLSTM en cascada. Además, en el *allele* HLA-A\*01:01, HLAB superó ligeramente a los métodos de vanguardia, incluido NetMHCpan4.1, en al menos 0.0230 en AUC y 0.0560 en precisión. Luego, nosotros hemos validado el uso de BERT y *transfer learning* (Machaca, 2023).

Luego, TransPHLA (Chu et al., 2022) es un método *allele-specific* que aplica *self-attention* a los péptidos. Los autores desarrollaron AOMP, que toma la unión de pMHC como entrada y devuelve péptidos mutantes con mayor afinidad hacia el *allele* MHC. Además, TransPHLA superó a los métodos de vanguardia, incluido NetMHCpan4.1, y es efectivo para cualquier longitud de péptido y MHC, y es más rápido para hacer predicciones. Además, el método DapNet-HLA *allele-specific* (Jing et al., 2023) obtuvo resultados interesantes, utilizó un conjunto de datos adicional (Swiss-Prot) para muestras negativas y combinó las ventajas de CNN, SENet (para agrupamiento) y LSTM. La propuesta obtuvo puntuaciones altas; sin embargo, el método no se comparó con métodos de vanguardia. Finalmente, nosotros en un trabajo anterior hemos desarrollado un *review* donde se detalla las principales ventajas y limitaciones de este problema (Machaca et al., 2023).

Finalmente, debido a la complejidad del proceso y la gran cantidad de métodos desarrollados, se ha desarrollado software y *pipelines* que pretenden facilitar el uso de estas herramientas. Entre las más recientes tenemos: Somaticseq (Fang et al., 2015), NeoPredPipe (Schenck et al., 2019), Cloud-Neo (Bais et al., 2017), MuPeXI (Bjerregaard et al., 2017), NeoepitopePred (Tran et al., 2015), Neoepiscopes (Yossef et al., 2018), pVACtools (Hundal et al., 2020) y NeoFuse (Gros et al., 2016). Estas herramientas en su mayoría toman como entrada archivos VCF y archivos de alineamiento Bam, para la detección de mutaciones (inserciones, eliminaciones y fusión de genes) y posibles neo antígenos.

## 2. Planteamiento del problema

Menos del 5 % de neoantígenos detectados llegan a la membrana y activan el sistema immune (Mattos et al., 2020; Mill et al., 2022; Bulik-Sullivan et al., 2019; Bassani-Sternberg et al., 2015; Yadav et al., 2014). Además, existen herramientas con buen desempeño en el problema de pMHC *binding*, pero con resultados pobres en pMHC *presentation* (Bulik-Sullivan et al., 2019). En este contexto, esta investigación se enfoca en la predicción de la unión pMHC. Este problema se puede representar como un problema de clasificación binaria, tomando un péptido y el MHC como entrada. Estos son secuencias de aminoácidos, el péptido se pueden representar como:  $p = \{A, \dots, Q\}$  y el MHC como:  $q = \{A, N, \dots, Q, E, G\}$ . Luego, tenemos que predecir si  $p$  y  $q$  pueden enlazarse.

## 3. Objetivos de la investigación

### 3.1. Objetivo general

Desarrollar una aplicación Web para la detección de neoantígenos en el marco del desarrollo de vacunas personalizadas para tratar el Cáncer.

### 3.2. Objetivos específicos

1. Implementar un método basado en *transformers* y *transfer learning* para predecir la unión pMHC.
2. Comparar los resultados del método propuesto con NetMHCpan4.1.
3. Implementar la aplicación Web para la detección de neoantígenos.

## 4. Importancia de la investigación

El cáncer es el mayor problema de salud mundial; sin embargo, los métodos tradicionales basados en cirugías, radioterapias y quimioterapias tienen baja efectividad (Peng et al., 2019). En este contexto, los neoantígenos son factores clave en el desarrollo de vacunas contra el cáncer (Borden et al., 2022; Chen et al., 2021; Gopanenko et al., 2020). Si se logra desarrollar un método con un buen desempeño, la inmunoterapia del cáncer basada en el desarrollo de vacunas personalizadas, podría utilizarse como alternativa a otros métodos como radioterapias y quimioterapias.

En el área de la inteligencia artificial y específicamente en el tópico de *deep learning*, el desarrollo de un método basado en *transformers* y *transfer learning*, representa una contribución importante y demuestra que este tipo de modelos no solo pueden utilizarse en campos del procesamiento natural del lenguaje (como lo hace chatGPT) sino también en otros ámbitos como la Immunoinformática.

Finalmente, tener una aplicación Web desarrollada por la universidad La Salle, realza el nombre de la universidad y la iguala a otras universidades que también desarrollan aplicaciones basadas en inteligencia artificial para solucionar problemas interdisciplinarios.

## 5. Posibles soluciones y consecuencias de la investigación

Para el desarrollo del proyecto, se propone el uso de *transformers* y *deep learning* al ser una de las técnicas con mayor investigación en la actualidad. Sin embargo, existen otras posibles soluciones basadas en redes neuronales totalmente conectadas y otros modelos básicos de *machine learning* como Support Vector Machines, Random Forest, etc; lamentablemente, estas últimas propuestas tienen resultados deficientes.

Entre las consecuencias de la investigación tenemos: (1) una publicación con el método propuesto para la predicción de la unión pMHC y (2) una aplicación Web para la predicción de la unión pMHC. Luego, esta investigación abre las puertas para futuros proyectos que involucren el desarrollo de un *pipeline* para la detección de neoantígenos, pero esta vez tomando como entrada una secuencia de ADN. Otro futuro proyecto de mayor envergadura involucra un trabajo interdisciplinario con biotecnólogos, biólogos y oncólogos; con el objetivo de desarrollar vacunas personalizadas para tratar el Cáncer.

## 6. Metodología de la Investigación

Hemos dividido la propuesta en dos partes: la primera se basa en el desarrollo de un modelo de *deep learning* basado en *transformers* y *transfer learning* para la predicción de la unión pMHC. Actualmente, ya tenemos resultados preliminares con un desempeño comparable a NetMHCpan4.1 (*state-of-art method*) (Machaca, 2023) y adicionalmente, hemos desarrollado un *review* sobre los principales métodos del estado del arte (Machaca et al., 2023). En la segunda parte del proyecto, se propone desarrollar la aplicación Web para la detección de neoantígenos, enfocados en la predicción de la unión pMHC.

### 6.1. Desarrollo del modelo de inteligencia artificial

En este proyecto, se propone utilizar una arquitectura BERT con *transfer learning*. Analizaremos alternativas como TAPE (Rao et al., 2019), ProtBERT-BFD (Elnaggar et al., 2021), ESM-1b

(Rives et al., 2021) y ESM2 (Lin et al., 2023) cada una con 92 millones, 420 millones, 650 millones y 15 billones de parámetros respectivamente. TAPE fue entrenado con 30 millones de proteínas, ProtBERT-BFD con 2122 millones de proteínas y 250 millones de proteínas para ESM-1b. Además, ESM-1b obtuvo mejores resultados en precisión de contacto que TAPE y ProtBERT-BFD (Rives et al., 2021); sin embargo, ya contamos la versión nueva ESM2 (Lin et al., 2023).

Además, HLAB (Zhang et al., 2022) propuso el uso de ProtBERT-BFD (Elnaggar et al., 2021) con un modelo BiLSTM en cascada y superó a NetMHCpan4.1 (método de vanguardia) en el *allele* HLA-A\*01:01. Por lo tanto, en este proyecto, proponemos utilizar el modelo preentrenado ESM2 (Lin et al., 2023) con un capa BiLSTM similar a HLAB (Zhang et al., 2022). Para el *fine-tuning*, utilizaremos conjuntos de datos de NetMHCpan4.1 y NetMHCIIpan4.0.

En resumen, en la Figura. 1, presentamos la propuesta: primero, se concatenan y se aplica *padding* al péptido y la pseudo secuencia MHC; en segundo lugar, utilizamos el modelo *transformer* ESM2 para obtener una nueva representación de los aminoácidos; finalmente, utilizamos un BiLSTM para predecir el enlace pMHC.

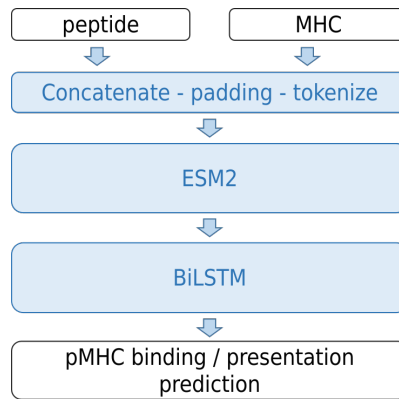


Figura 1: Propuesta: Utilizamos el modelo de transformer ESM2 seguido de BiLSTM para predecir el enlace pMHC.

## 6.2. Desarrollo de la aplicación Web

La aplicación Web se desarrollará para que un usuario pueda realizar predicciones de la unión pMHC. La página Web tomará como entrada un conjunto de péptidos y un tipo de MHC. Luego, retornará y mostrará que péptidos pueden unirse a dicho MHC y por lo tanto se les considera posibles neoantígenos.

Referente a las tecnologías a utilizar: para el *backend*, se utilizará el *framework* Express.js de Javascript y el gestor de base de datos Mongo. Para el *frontend*, se propone utilizar Next.js. Se ha seleccionado estas tecnologías debido a su versatilidad y la enorme documentación que presentan.

## 7. Presupuesto y cronograma

En la Tabla 1, presentamos el presupuesto para el trabajo de investigación. Este asciende a la suma de 4000 mil soles.

Tabla 1: Presupuesto. Abreviaciones, PC: *Personal Computer*

Insumo o material	Unidades	Precio	Total
Incentivo para el investigador principal	1	700	700
Incentivo para el co-investigador	1	700	700
Hosting y dominio	1	600	600
Workshops y cursos	1	1000	1000
Servicios de <i>cloud computing</i> para entrenar los modelos	1	1000	1000
<b>Total</b>			<b>4000</b>

En la Tabla 2, presentamos el cronograma de actividades por mes.

Tabla 2: Cronograma de actividades por mes.

Actividades	I	II	III	IV	V	VI	VII
<b>HITO I</b>							
Revisión de la literatura	x	x	x	x			
Implementación del modelo de inteligencia artificial	x	x	x	x			
Evaluación y comparación del modelo			x	x	x		
<b>HITO II</b>							
Implementación de la página Web				x	x	x	
Redacción del artículo de investigación					x	x	x
Difusión de resultados							x

## 8. Propuesta de la conferencia o revista

El artículo de investigación será presentado a la conferencia: “18th International Conference on Practical Applications of Computational Biology & Bioinformatics” ([enlace](#)). Esta conferencia indexa los artículos en Scopus a también en la revista Lecture Notes in Networks and Systems que es H-index 27 según Scimago ([enlace](#)). A su vez, esta conferencia invita a los mejores artículos a una publicación extendida en revistas cuartil Q1.

## Referencias

- Alvarez, B., Reynisson, B., Barra, C., Buus, S., Ternet, N., Connelley, T., Andreatta, M., and Nielsen, M. (2019). Nalign\_ma: mhc peptidome deconvolution for accurate mhc binding motif characterization and improved t-cell epitope predictions. *Molecular & Cellular Proteomics*, 18(12):2459–2477.
- Andreatta, M. and Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: application to the mhc class i system. *Bioinformatics*, 32(4):511–517.
- Bais, P., Namburi, S., Gatti, D. M., Zhang, X., and Chuang, J. H. (2017). Cloudneo: a cloud pipeline for identifying patient-specific tumor neoantigens. *Bioinformatics*, 33(19):3110–3112.
- Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J., and Mann, M. (2015). Mass spectrometry of human leukocyte antigen class i peptidomes reveals strong effects of protein abundance and turnover on antigen presentation\*[s]. *Molecular & Cellular Proteomics*, 14(3):658–673.
- Bjerregaard, A.-M., Nielsen, M., Hadrup, S. R., Szallasi, Z., and Eklund, A. C. (2017). Mupexi: prediction of neo-epitopes from tumor sequencing data. *Cancer Immunology, Immunotherapy*, 66(9):1123–1130.
- Borden, E. S., Buetow, K. H., Wilson, M. A., and Hastings, K. T. (2022). Cancer neoantigens: Challenges and future directions for prediction, prioritization, and validation. *Frontiers in Oncology*, 12.
- Bulik-Sullivan, B., Busby, J., Palmer, C. D., Davis, M. J., Murphy, T., Clark, A., Busby, M., Duke, F., Yang, A., Young, L., et al. (2019). Deep learning using tumor hla peptide mass spectrometry datasets improves neoantigen identification. *Nature biotechnology*, 37(1):55–63.
- Chen, I., Chen, M., Goedegebuure, P., and Gillanders, W. (2021). Challenges targeting cancer neoantigens in 2021: a systematic literature review. *Expert Review of Vaccines*, 20(7):827–837.
- Cheng, J., Bendjama, K., Rittner, K., and Malone, B. (2021). Bertmhc: improved mhc-peptide class ii interaction prediction with transformer and multiple instance learning. *Bioinformatics*, 37(22):4172–4179.
- Chu, Y., Zhang, Y., Wang, Q., Zhang, L., Wang, X., Wang, Y., Salahub, D. R., Xu, Q., Wang, J., Jiang, X., et al. (2022). A transformer-based model to predict peptide-hla class i binding and optimize mutated peptides for vaccine design. *Nature Machine Intelligence*, 4(3):300–311.
- Donnell, T. J., Rubinsteyn, A., and Laserson, U. (2020). Mhcflurry 2.0: improved pan-allele prediction of mhc class i-presented peptides by incorporating antigen processing. *Cell systems*, 11(1):42–48.
- Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., et al. (2021). Prottrans: Toward understanding the language of life through self-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(10):7112–7127.
- Fang, L. T., Afshar, P. T., Chhibber, A., Mohiyuddin, M., Fan, Y., Mu, J. C., Gibeling, G., Barr, S., Asadi, N. B., Gerstein, M. B., et al. (2015). An ensemble approach to accurately detect somatic mutations using somaticseq. *Genome biology*, 16(1):1–13.
- Gasser, H.-C., Bedran, G., Ren, B., Goodlett, D., Alfaro, J., and Rajan, A. (2021). Interpreting bert architecture predictions for peptide presentation by mhc class i proteins. *arXiv preprint arXiv:2111.07137*.
- Gopanenko, A. V., Kosobokova, E. N., and Kosorukov, V. S. (2020). Main strategies for the identification of neoantigens. *Cancers*, 12(10):2879.
- Gros, A., Parkhurst, M. R., Tran, E., Pasetto, A., Robbins, P. F., Ilyas, S., Prickett, T. D., Gartner, J. J., Crystal, J. S., Roberts, I. M., et al. (2016). Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nature medicine*, 22(4):433–438.

- Hundal, J., Kiwala, S., McMichael, J., Miller, C. A., Xia, H., Wollam, A. T., Liu, C. J., Zhao, S., Feng, Y.-Y., Graubert, A. P., et al. (2020). pvactools: a computational toolkit to identify and visualize cancer neoantigens. *Cancer immunology research*, 8(3):409–420.
- Jing, Y., Zhang, S., and Wang, H. (2023). Dapnet-hla: Adaptive dual-attention mechanism network based on deep learning to predict non-classical hla binding sites. *Analytical Biochemistry*, 666:115075.
- Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130.
- Machaca, V. E., Goyzueta, V., Cruz, M., and Tupac, Y. (2023). Deep learning and transformers in mhc-peptide binding and presentation towards personalized vaccines in cancer immunology: A brief review. In *International Conference on Practical Applications of Computational Biology & Bioinformatics*, pages 14–23. Springer.
- Machaca, V. M. A. (2023). Neoantigen detection using transformers and transfer learning in the cancer immunology context. In *International Conference on Practical Applications of Computational Biology & Bioinformatics*, pages 97–102. Springer.
- Mattos, L., Vazquez, M., Finotello, F., Lepore, R., Porta, E., Hundal, J., Amengual-Rigo, P., Ng, C., Valencia, A., Carrillo, J., et al. (2020). Neoantigen prediction and computational perspectives towards clinical benefit: recommendations from the esmo precision medicine working group. *Annals of oncology*, 31(8):978–990.
- Mill, N. A., Bogaert, C., van Crielinge, W., and Fant, B. (2022). neoms: Attention-based prediction of mhc-i epitope presentation. *bioRxiv*.
- Patwardhan, N., Marrone, S., and Sansone, C. (2023). Transformers in the real world: A survey on nlp applications. *Information*, 14(4):242.
- Peng, M., Mo, Y., Wang, Y., Wu, P., Zhang, Y., Xiong, F., Guo, C., Wu, X., Li, Y., Li, X., et al. (2019). Neoantigen vaccine: an emerging tumor immunotherapy. *Molecular cancer*, 18(1):1–14.
- Rao, R., Bhattacharya, N., Thomas, N., Duan, Y., Chen, P., Canny, J., Abbeel, P., and Song, Y. (2019). Evaluating protein transfer learning with tape. *Advances in neural information processing systems*, 32.
- Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). Netmhcpa-4.1 and netmhciipa-4.0: improved predictions of mhc antigen presentation by concurrent motif deconvolution and integration of ms mhc eluted ligand data. *Nucleic acids research*, 48(W1):W449–W454.
- Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., et al. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15).
- Schenck, R. O., Lakatos, E., Gatenbee, C., Graham, T. A., and @miscNCIdictionary2022, author = NCI, title = National Cancer Institute Dictionary, year = 2022, url = <https://www.cancer.gov/publications/dictionaries/genetics-dictionary>, urldate = 2022-03-20
- Anderson, A. R. (2019). Neopredpipe: high-throughput neoantigen prediction and recognition potential pipeline. *BMC bioinformatics*, 20(1):1–6.
- Tran, E., Ahmadzadeh, M., Lu, Y.-C., Gros, A., Turcotte, S., Robbins, P. F., Gartner, J. J., Zheng, Z., Li, Y. F., Ray, S., et al. (2015). Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science*, 350(6266):1387–1390.
- Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A., and Peters, B. (2018). The immune epitope database (iedb): 2018 update. *Nucleic acids research*, 47(D1):D339–D343.
- Wang, F., Wang, H., Wang, L., Lu, H., Qiu, S., Zang, T., Zhang, X., and Hu, Y. (2022). Mhcroberta: pan-specific peptide-mhc class i binding prediction through transfer learning with label-agnostic protein sequences. *Briefings in Bioinformatics*, 23(3):bbab595.

- Yadav, M., Jhunjhunwala, S., Phung, Q. T., Lupardus, P., Tanguay, J., Bumbaca, S., Franci, C., Cheung, T. K., Fritsche, J., Weinschenk, T., et al. (2014). Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature*, 515(7528):572–576.
- Yossef, R., Tran, E., Deniger, D. C., Gros, A., Pasetto, A., Parkhurst, M. R., Gartner, J. J., Prickett, T. D., Cafri, G., Robbins, P. F., et al. (2018). Enhanced detection of neoantigen-reactive t cells targeting unique and shared oncogenes for personalized cancer immunotherapy. *JCI insight*, 3(19).
- Zhang, Y., Zhu, G., Li, K., Li, F., Huang, L., Duan, M., and Zhou, F. (2022). Hlab: learning the bilstm features from the protbert-encoded proteins for the class i hla-peptide binding prediction. *Briefings in Bioinformatics*.