



Universidad de Ingeniería y Tecnología

Detección de neo antígenos utilizando *deep learning* en el marco del desarrollo de vacunas personalizadas en la inmunoterapia del Cáncer

MSc. Vicente Machaca Arceda

2022-I



Marco teórico

Bioinformática y DNA

Mutaciones

Neo antígenos

Problema y Objetivos

Motivación y Problema

Objetivo

Estado del arte

Propuesta

Resultados

Conclusiones

Trabajos futuros



Marco teórico

Bioinformática y DNA

Mutaciones

Neo antígenos

Problema y Objetivos

Motivación y Problema

Objetivo

Estado del arte

Propuesta

Resultados

Conclusiones

Trabajos futuros

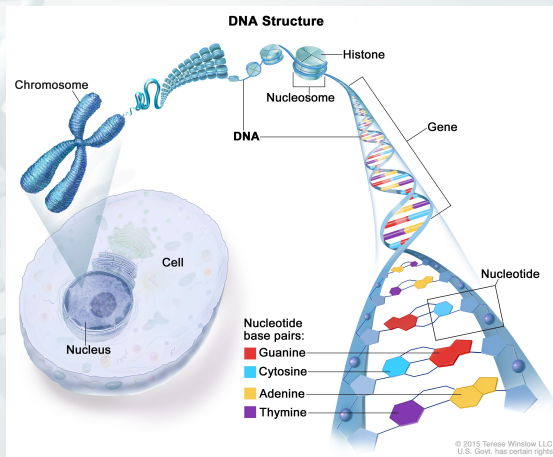


Figure: Where DNA is located [1].

DNA

De DNA a proteínas

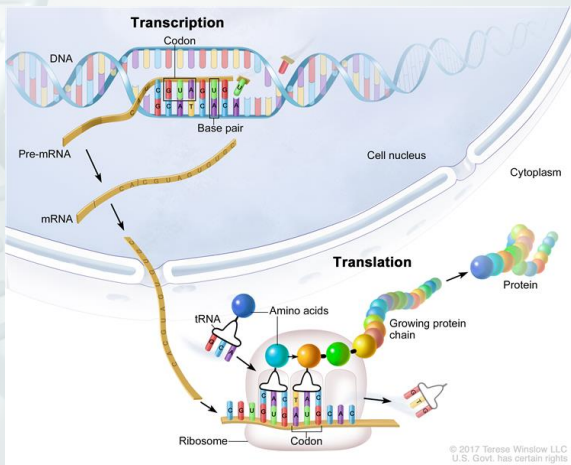


Figure: Transcription and translation [2].

Variantes y Mutaciones

Tipos



- ▶ **Single-Nucleotide Variant (SNV)**, cambios a menos de 10 bases.
- ▶ **Structural Variation (SV)**, cambios a mas de 10 bases, incluso pueden llegar a aumentar la cantidad de cromosomas.

Variantes y Mutaciones

Ejemplo

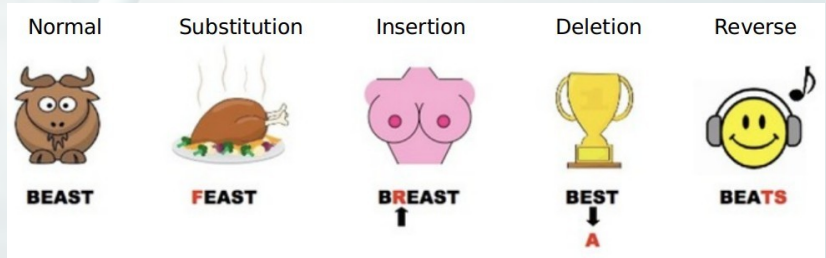


Figure: Overview of the Different Types of Point Mutations.

Variantes y Mutaciones



Single Nucleotide Variant



Deletion



Insertion



Tandem Duplication



Interspersed Duplication



Inversion



Translocation



Copy Number Variant



Types of Variants

Figure: Example of structural variants. Source: [3]

Variantes y Mutaciones

Frameshift

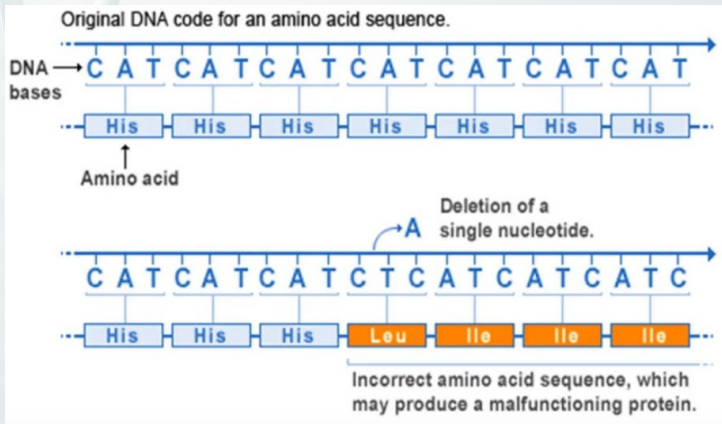


Figure: Ejemplo de una mutación INDELS causante de un *frameshift*.

Fusión de genes

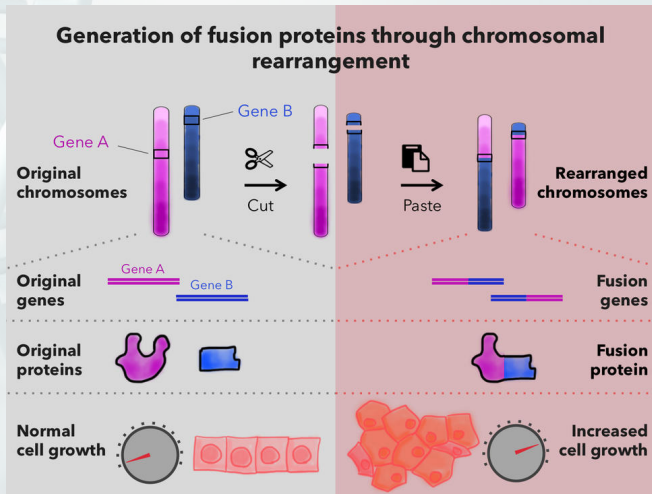


Figure: Ejemplo de una fusión de genes.

Variaciones a nivel de cromosomas



Figure: Los 46 cromosomas presentes en una célula.

Variaciones a nivel de cromosomas

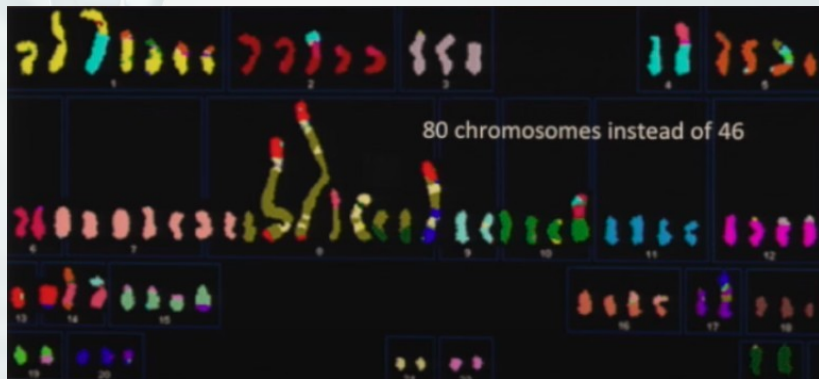
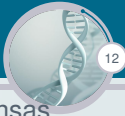


Figure: Cromosomas de una mujer con Cáncer de mama (1971).

Inmunoterapia del Cáncer



Es un tipo de tratamiento contra el Cáncer que estimula las defensas naturales del cuerpo para combatir el Cáncer [4].

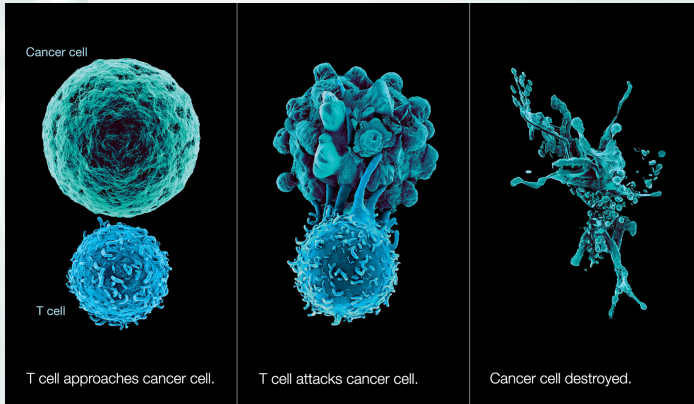


Figure: Ejemplo de como una célula T destruye células del cancer [5].

Inmunoterapia del Cáncer

Neo antígenos



Es una **proteína** que se forma en las células de Cáncer cuando ocurre mutaciones en el DNA, cumplen un rol importante al **estimular una respuesta inmune** [1, 6].

En la actualidad hay varios métodos para detectar a predecir neo antígenos, pero **solo una pequeña cantidad de ellos** logran estimular al sistema inmune [7, 8].

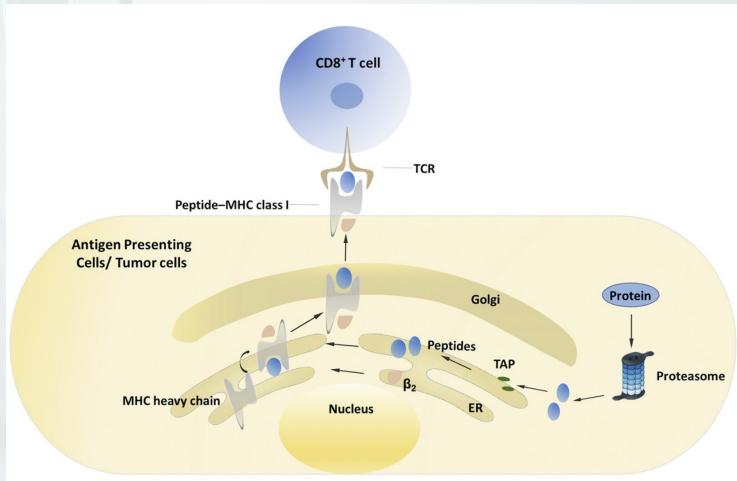


Figure: Presentación de antígenos por MHC-I. Fuente: [9]

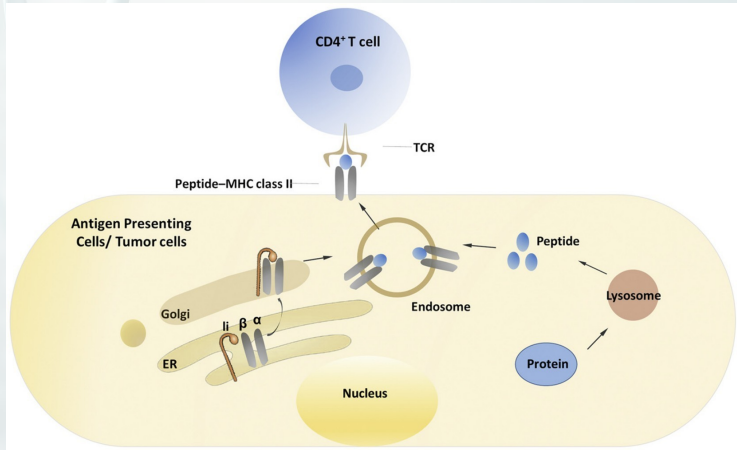


Figure: Presentación de antígenos por MHC-II. Fuente: [9]

Inmunoterapia del Cáncer

Generación de vacunas

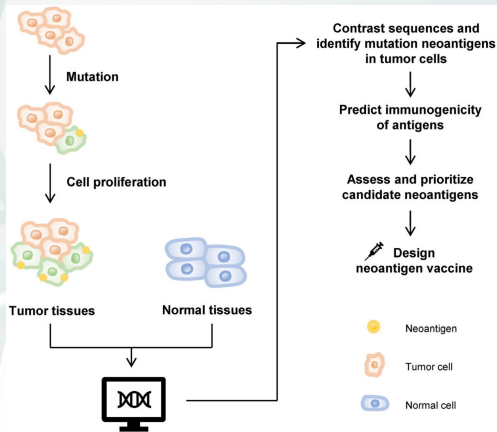


Figure: Proceso para la generación de vacunas personalizadas [10].



Marco teórico

Bioinformática y DNA

Mutaciones

Neo antígenos

Problema y Objetivos

Motivación y Problema

Objetivo

Estado del arte

Propuesta

Resultados

Conclusiones

Trabajos futuros



El cáncer representa el mayor problema de salud mundial, pero lamentablemente los métodos basados en cirugías, radioterapias, quimioterapias tienen baja efectividad [10].

La inmunoterapia del cáncer es una alternativa para el desarrollo de vacunas personalizadas, pero este proceso depende de una correcta detección de neo antígenos [11, 10].



Menos del 5% de péptidos detectados en *pMHC binding*, llegan a la membrana de la células, para que luego sean reconocidos por las células T (*pMHC presentation*) [11, 12, 13, 14, 15]. Además, propuestas recientes solo llegan aun 0.6 de *presicion* y 0.4 de *recall* [12].

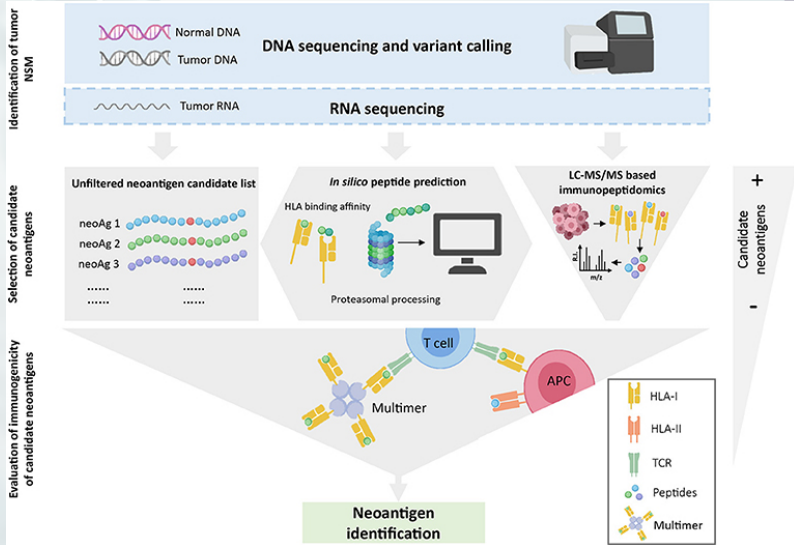
Dada como entrada dos secuencias de aminoacidos, representando a un péptido y a una molécula MHC, el objetivo es predecir su afinidad, como si fuera un problema de clasificación binaria.



Objetivo general

Proponer un método basado en *deep learning* para la detección de neo antígenos, enfocados en el problema de *pMHC presentation*.

Objetivos





Marco teórico

Bioinformática y DNA

Mutaciones

Neo antígenos

Problema y Objetivos

Motivación y Problema

Objetivo

Estado del arte

Propuesta

Resultados

Conclusiones

Trabajos futuros



Año	Nombre	Referencia
2020	ProGeo-neo	[16]
2020	INeo-Epp	[17]
2020	pVACtools	[18]
2019	NeoPredPipe	[19]
2019	DeepHLApan	[20]
2019	ScanNeo	[21]
2017	CloudNeo	[22]
...

Estado del arte

Peptide-MHC binding



Año	Nombre	Modelo	Referencia
2022	AEM	Transformer	[23]
2021	BERTMHC	Transformer	[24]
2021	APPM	3 CNN	[8]
2020	NetMHCpan4.1	ANN	[25]
2020	MHCflurry2.0	ANN	[26]
2020	MHCnuggets	ANN	[27]
2019	PUFFIN	ANN	[28]
..



Marco teórico

Bioinformática y DNA
Mutaciones
Neo antígenos

Problema y Objetivos

Motivación y Problema
Objetivo

Estado del arte

Propuesta

Resultados

Conclusiones

Trabajos futuros

La propuesta se basa en los modelos BERTMHC [24] y APPM [8].

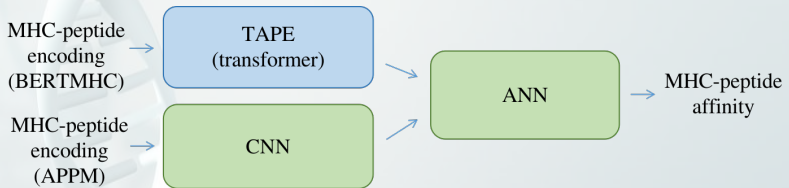


Figure: Modelo propuesto para la predicción del enlace péptido y MHC.

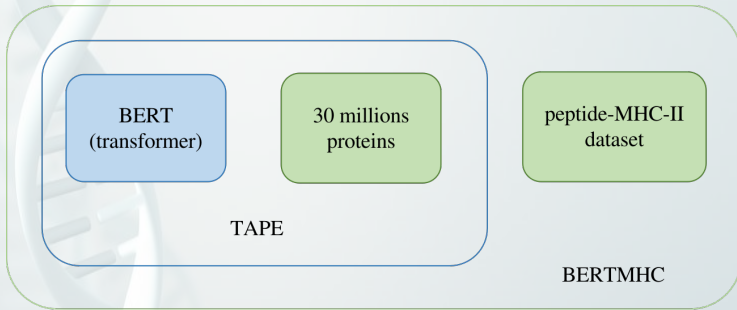


Figure: BERTMHC.

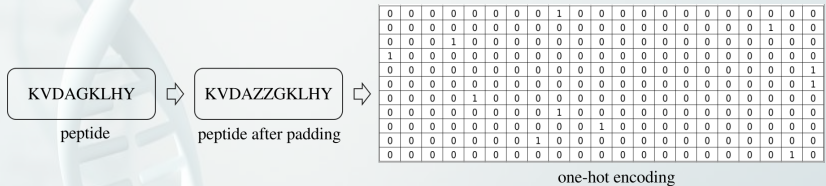


Figure: Proceso para obtener una matriz (imagen) a partir de un péptido (APPM).



Marco teórico

Bioinformática y DNA

Mutaciones

Neo antígenos

Problema y Objetivos

Motivación y Problema

Objetivo

Estado del arte

Propuesta

Resultados

Conclusiones

Trabajos futuros



Table: Cantidad de muestras por tipo de *allele*.

<i>Alleles</i>	Label = 1	Label = 0	Train	Test
A*01:01	3398	48700	45498	6600
A*02:01	6779	165342	160921	11200
A*02:03	1780	116299	107879	10200
A*31:01	1879	45918	41597	6200
B*44:02	1525	44760	40085	6200
B*44:03	1487	39482	34769	6200
MHC-II alleles	1917	496	1533	384



Table: Resultados obtenidos en cada base de datos.

<i>Allele</i>	<i>Accuracy</i>	<i>F1 score</i>	<i>Precision</i>	<i>Recall</i>
A*01:01	0.978	0.917	0.982	0.887
A*0201	0.962	0.956	0.965	0.948
A*02:03	0.992	0.979	0.994	0.969
A*31:01	0.980	0.968	0.989	0.951
B*44:02	0.991	0.981	0.968	0.997
B*44:03	0.992	0.987	0.995	0.980



Table: AUC entre la propuesta, BERTMHC, NetMHCpan3.2, PUFFIN y MHCnuggets.

Modelo	AUC
Propuesta	0.72
BERTMHC	0.72
NetMHCpan3	0.68
PUFFIN	0.69
MHCnuggets	0.58



Marco teórico

Bioinformática y DNA

Mutaciones

Neo antígenos

Problema y Objetivos

Motivación y Problema

Objetivo

Estado del arte

Propuesta

Resultados

Conclusiones

Trabajos futuros

Conclusiones



En esta investigación se propuso el uso de un modelo *transformer* ya entrenado con una base de datos de 30 millones de proteínas. Luego, esta red fue conectada de forma paralela con una red CNN.

El uso de *transfer learning* es una buena opción para suplir la falta de muestras en ciertos problemas y reducir el tiempo de entrenamiento.

La propuesta llegó a mejorar los mejores métodos de detección de afinidad entre un péptido y una proteína MHC-II. Como trabajo futuro, se planteará la misma propuesta para proteínas MHC-I.

Predecir la afinidad entre un péptido y una proteína MHC, es uno de los pasos más importantes para calificar al péptido como un neo antígeno, capaz de generar una respuesta inmunitaria.



Marco teórico

Bioinformática y DNA

Mutaciones

Neo antígenos

Problema y Objetivos

Motivación y Problema

Objetivo

Estado del arte

Propuesta

Resultados

Conclusiones

Trabajos futuros



Recientemente un trabajo [23] también propone el uso de *transfer learning* pero de un modelo pre-entrenado con 250 millones de proteínas. Entonces, se plantea utilizar la misma red, aumentar la cantidad de muestras y evaluar los resultados.

Actualmente se cuenta con una base de datos de proteínas MHC [29], entonces utilizando AlphaFold de Google, se plantea predecir la estructura de varios péptidos y analizar el enlace péptido-MHC desde un punto de vista de la computación gráfica.



- [1] NCI. (2022) National cancer institute dictionary. [Online]. Available: <https://www.cancer.gov/publications/dictionaries/genetics-dictionary>
- [2] —, “Nci dictionary of cancer terms,” <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/transcription>, 2020, accessed: 2020-03-20.
- [3] PacBio, “Two review articles assess structural variation in human genomes,” <https://www.pacb.com/blog/two-review-articles-assess-structural-variation-in-human-genomes/>, 2021, accessed: 2021-05-07. [Online]. Available: <https://www.pacb.com/blog/two-review-articles-assess-structural-variation-in-human-genomes/>



- [4] Cancer.net. (2022) Qué es la inmunoterapia. [Online]. Available: <https://www.cancer.net/es/desplazarse-por-atencion-del-cancer/como-se-trata-el-cancer/inmunoterapia/que-es-la-inmunoterapia>
- [5] NortShore. (2022) Immunotherapy. [Online]. Available: <https://www.northshore.org/kellogg-cancer-center/our-services/immunotherapy/>
- [6] E. S. Borden, K. H. Buetow, M. A. Wilson, and K. T. Hastings, “Cancer neoantigens: Challenges and future directions for prediction, prioritization, and validation,” *Frontiers in Oncology*, vol. 12, 2022.
- [7] I. Chen, M. Chen, P. Goedegebuure, and W. Gillanders, “Challenges targeting cancer neoantigens in 2021: a systematic literature review,” *Expert Review of Vaccines*, vol. 20, no. 7, pp. 827–837, 2021.



- [8] Q. Hao, P. Wei, Y. Shu, Y.-G. Zhang, H. Xu, and J.-N. Zhao, "Improvement of neoantigen identification through convolution neural network," *Frontiers in immunology*, vol. 12, 2021.
- [9] X. Zhang, Y. Qi, Q. Zhang, and W. Liu, "Application of mass spectrometry-based mhc immunopeptidome profiling in neoantigen identification for tumor immunotherapy," *Biomedicine & Pharmacotherapy*, vol. 120, p. 109542, 2019.
- [10] M. Peng, Y. Mo, Y. Wang, P. Wu, Y. Zhang, F. Xiong, C. Guo, X. Wu, Y. Li, X. Li *et al.*, "Neoantigen vaccine: an emerging tumor immunotherapy," *Molecular cancer*, vol. 18, no. 1, pp. 1–14, 2019.



- [11] L. Mattos, M. Vazquez, F. Finotello, R. Lepore, E. Porta, J. Hundal, P. Amengual-Rigo, C. Ng, A. Valencia, J. Carrillo *et al.*, “Neoantigen prediction and computational perspectives towards clinical benefit: recommendations from the esmo precision medicine working group,” *Annals of oncology*, vol. 31, no. 8, pp. 978–990, 2020.
- [12] N. A. Mill, C. Bogaert, W. van Crielinge, and B. Fant, “neoms: Attention-based prediction of mhc-i epitope presentation,” *bioRxiv*, 2022.
- [13] B. Bulik-Sullivan, J. Busby, C. D. Palmer, M. J. Davis, T. Murphy, A. Clark, M. Busby, F. Duke, A. Yang, L. Young *et al.*, “Deep learning using tumor hla peptide mass spectrometry datasets improves neoantigen identification,” *Nature biotechnology*, vol. 37, no. 1, pp. 55–63, 2019.



- [14] M. Bassani-Sternberg, S. Pletscher-Frankild, L. J. Jensen, and M. Mann, “Mass spectrometry of human leukocyte antigen class i peptidomes reveals strong effects of protein abundance and turnover on antigen presentation*[s],” *Molecular & Cellular Proteomics*, vol. 14, no. 3, pp. 658–673, 2015.
- [15] M. Yadav, S. Jhunjunwala, Q. T. Phung, P. Lupardus, J. Tanguay, S. Bumbaca, C. Franci, T. K. Cheung, J. Fritsche, T. Weinschenk *et al.*, “Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing,” *Nature*, vol. 515, no. 7528, pp. 572–576, 2014.
- [16] Y. Li, G. Wang, X. Tan, J. Ouyang, M. Zhang, X. Song, Q. Liu, Q. Leng, L. Chen, and L. Xie, “Progeo-neo: a customized proteogenomic workflow for neoantigen prediction and selection,” *BMC medical genomics*, vol. 13, no. 5, pp. 1–11, 2020.



- [17] G. Wang, H. Wan, X. Jian, Y. Li, J. Ouyang, X. Tan, Y. Zhao, Y. Lin, and L. Xie, “Ineo-epp: a novel t-cell hla class-i immunogenicity or neoantigenic epitope prediction method based on sequence-related amino acid features,” *BioMed research international*, vol. 2020, 2020.
- [18] J. Hundal, S. Kiwala, J. McMichael, C. A. Miller, H. Xia, A. T. Wollam, C. J. Liu, S. Zhao, Y.-Y. Feng, A. P. Graubert *et al.*, “pvactools: a computational toolkit to identify and visualize cancer neoantigens,” *Cancer immunology research*, vol. 8, no. 3, pp. 409–420, 2020.
- [19] R. O. Schenck, E. Lakatos, C. Gatenbee, T. A. Graham, and A. R. Anderson, “Neopredpipe: high-throughput neoantigen prediction and recognition potential pipeline,” *BMC bioinformatics*, vol. 20, no. 1, pp. 1–6, 2019.



- [20] J. Wu, W. Wang, J. Zhang, B. Zhou, W. Zhao, Z. Su, X. Gu, J. Wu, Z. Zhou, and S. Chen, “Deephlapan: a deep learning approach for neoantigen prediction considering both hla-peptide binding and immunogenicity,” *Frontiers in Immunology*, p. 2559, 2019.
- [21] T.-Y. Wang, L. Wang, S. K. Alam, L. H. Hoepfner, and R. Yang, “Scanneo: identifying indel-derived neoantigens using rna-seq data,” *Bioinformatics*, vol. 35, no. 20, pp. 4159–4161, 2019.
- [22] P. Bais, S. Namburi, D. M. Gatti, X. Zhang, and J. H. Chuang, “Cloudneo: a cloud pipeline for identifying patient-specific tumor neoantigens,” *Bioinformatics*, vol. 33, no. 19, pp. 3110–3112, 2017.
- [23] N. Hashemi, B. Hao, M. Ignatov, I. Paschalidis, P. Vakili, S. Vajda, and D. Kozakov, “Improved predictions of mhc-peptide binding using protein language models,” *bioRxiv*, 2022.



- [24] J. Cheng, K. Bendjama, K. Rittner, and B. Malone, “Bertmhc: improved mhc–peptide class ii interaction prediction with transformer and multiple instance learning,” *Bioinformatics*, vol. 37, no. 22, pp. 4172–4179, 2021.
- [25] B. Reynisson, B. Alvarez, S. Paul, B. Peters, and M. Nielsen, “Netmhciipan-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 research*, vol. 48, no. W1, pp. W449–W454, 2020.
- [26] T. J. O’Donnell, A. Rubinsteyn, and U. Laserson, “Mhcflurry 2.0: Improved pan-allele prediction of mhc class i-presented peptides by incorporating antigen processing,” *Cell systems*, vol. 11, no. 1, pp. 42–48, 2020.



- [27] X. M. Shao, R. Bhattacharya, J. Huang, I. A. Sivakumar, C. Tokheim, L. Zheng, D. Hirsch, B. Kaminow, A. Omdahl, M. Bonsack *et al.*, “High-throughput prediction of mhc class i and ii neoantigens with mhcnuggets,” *Cancer immunology research*, vol. 8, no. 3, pp. 396–408, 2020.
- [28] H. Zeng and D. K. Gifford, “Quantification of uncertainty in peptide-mhc binding prediction improves high-affinity peptide selection for therapeutic design,” *Cell systems*, vol. 9, no. 2, pp. 159–166, 2019.
- [29] D. M. T. Oliveira, R. M. S. de Serpa Brandão, L. C. D. da Mata Sousa, F. d. C. A. Lima, S. J. H. do Monte, M. S. C. Marroquim, A. V. de Sousa Lima, A. G. B. Coelho, J. M. S. Costa, R. M. Ramos *et al.*, “phla3d: An online database of predicted three-dimensional structures of hla molecules,” *Human Immunology*, vol. 80, no. 10, pp. 834–841, 2019.

