

Principales Estrategias y Métodos Basados en Deep Learning Para la Detección de Neo Antígenos en el Marco del Desarrollo de Vacunas Personalizadas en la Inmunoterapia del Cáncer

Proyecto en colaboración Universidad La Salle y UCSP

PhD(c). Vicente Machaca Arceda

Introducción

Proyecto

Neo antígenos

Problema y Objetivos

Revisión Sistemática

Metodología

Discusión

Introducción

Proyecto

Neo antígenos

Problema y Objetivos

Revisión Sistemática

Metodología

Discusión



Universidad Católica
San Pablo

- ▶ Seis (06) meses.
- ▶ Equipo:
 - ▶ Vicente Machaca Arceda (ULaSalle).
 - ▶ Valeria Goyzueta (ULaSalle).
 - ▶ Yván Tupac (UCSP).
 - ▶ Maria Cruz (UCSP).

Introducción

Proyecto

Neo antígenos

Problema y Objetivos

Revisión Sistemática

Metodología

Discusión

Inmunoterapia del Cáncer

Es un tipo de tratamiento contra el Cáncer que estimula las defensas naturales del cuerpo para combatir el Cáncer [1].

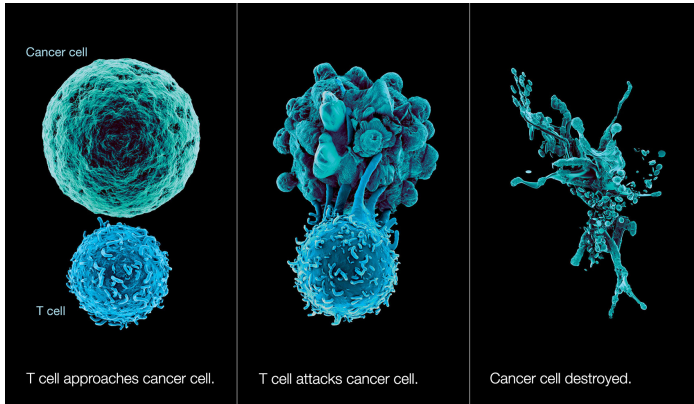


Figure: Ejemplo de como una célula T destruye células del cancer [2].

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** [3, 4].

En la actualidad hay varios métodos para detectar a predecir neo antígenos, pero **solo una pequeña cantidad de ellos** logran estimular al sistema inmune [5, 6].

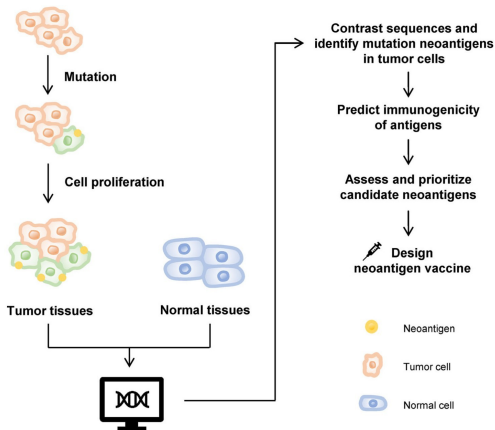


Figure: Proceso para la generación de vacunas personalizadas [7].

Introducción

Proyecto

Neo antígenos

Problema y Objetivos

Revisión Sistemática

Metodología

Discusión

Problema

Menos del 5% de los neoantígenos detectados logran activar a las células T (sistema inmune) [8].

Objetivo general

Desarrollar una revisión sistemática de métodos que utilizan *deep learning* para la detección de neo antígenos.

Introducción

Proyecto

Neo antígenos

Problema y Objetivos

Revisión Sistemática

Metodología

Discusión

Table: Research questions used in SLR.

Preguntas de investigación

- Q1.** Como se utilizan las técnicas de *deep learning* para la detección de neo antígenos?
- Q2.** Que tipos de datos y pre procesamiento es utilizado en la detección de ne antígenos?
- Q3.** Que bases de datos son utilizadas en la detección de neo antígenos?
- Q4.** Que método basado en *deep learning* tiene los mejores resultados?

Table: Cadenas de busqueda utilizadas en la RSL.

Cadena de busqueda

neoantigen AND (detection OR pipeline) AND deep learning

(MHC OR HLA) AND binding AND deep learning

(MHC-I OR MHC-II OR MHC OR HLA) AND (peptide OR epitope) AND (binding OR affinity OR prediction OR detection OR presentation)

TCR interaction prediction

Table: Bases de datos utilizadas en la RSL.

Bases de datos

IEEE Xplore

Science Direct

Springer

ACM Digital Library

PubMed

BioRxiv

Table: Crterios de inclusi3n y exclusi3n de artculos utilizados en la RSL.

Crterios de inclusi3n	Crterios de exclusi3n
Artculos con categora ERA (A, B o C) si son conferencias y Journals Q1, Q2 o Q3.	Trabajos de baja calidad, que no esten rankeados.
Se basan en uso de deep learning	Preprints

Table: Cantidad de artículos encontrados y seleccionados según los criterios de inclusión y exclusión en la RSL.

Año	Artículos encontrados	Artículos seleccionados
2018	57	21
2019	72	31
2020	86	29
2021	61	34
2022	58	19
Total	334	134

Stage 1.

Prediction of
peptides from
mutations

Stage 2.

Prediction of
pMHC binding
and presentation

Stage 3.

Prediction of
pMHC interaction
with TCR

Figure: Fases para la detección de neo antígenos.

Discusión

¿Como se utilizan las técnicas de *deep learning*?

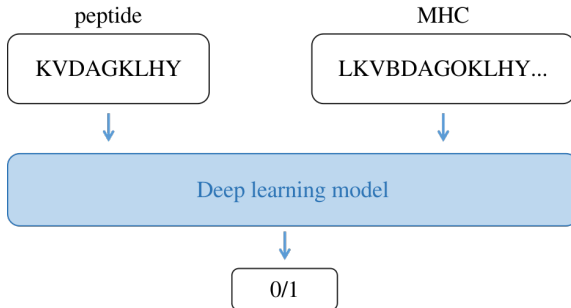


Name	Year ref.	Input	Output
Neopepsee	2018 [9]	RNA-seq, somatic mutations (VCF), HLA type (optional)	Neoantigens and gene expression levels
PGV Pipeline	2018 [10]	DNA-seq	Neoantigens
ScanNeo	2019 [11]	RNA-seq	Neoantigens
NeoPredPipe	2019 [12]	Mutations (VCF) y HLA type	Neoantigens and variant annotation
pVACtools	2020 [13]	Mutations (VCF)	Neoantigens
ProGeo-neo	2020 [14]	RNA-seq y somatic mutations (VCF)	Neoantigens
Neoepiscope	2020 [15]	Somatic mutations (VCF) and BAM files	Neoantigens and mutations
NeoANT-HILL	2020 [16]	RNA-seq y somatic mutations (VCF)	Neoantigens and gene expression levels
NAP-CNB	2021 [17]	RNA-seq	Neoantigens
PEPPRMINT	2021 [18]	DNA-seq	Neoantigens
Valid-NEO	2022 [19]	Somatic mutations (VCF), HLA type (optional)	Neoantigens

Discusión

¿Como se utilizan las técnicas de *deep learning*?

La detección de neo antígenos es visto como un problema de clasificación binaria.



Discusión

¿Que tipos de datos y pre procesamiento es utilizado?

Se consideran las cadenas de aminoacidos, como pre procesamiento se utiliza *one-hot encoding* y *BLOSUM*.

A	R	N	V
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	0
0	0	0	0
0	0	0	0
.	.	.	.
.	.	.	.
.	.	.	.
0	0	0	1

	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W
C	0																			
S	-1	4																		
T	-1	1	5																	
P	-3	-1	-1	7																
A	0	1	0	-1	4															
G	-3	0	-2	-2	0	6														
N	-3	-1	0	-2	-2	0	6													
D	-3	1	-1	-1	-2	-1	1	6												
E	-4	0	-1	-1	-1	-2	0	2	5											
Q	-3	0	-1	-1	-1	-2	0	0	2	5										
H	-3	0	-2	-2	-2	-2	1	-1	0	0	8									
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5								
K	-3	-1	-1	-1	-1	-2	0	-1	1	1	-1	2	5							
M	-1	0	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5						
I	-1	-2	-1	-3	-1	-1	-4	-3	-3	-3	-3	-3	1	4						
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	2	2	4					
V	-1	-2	0	-2	0	-3	-3	-3	-2	-3	-3	-2	1	3	1	4				
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6			
Y	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	-2	-2	-1	-1	-1	-1	3	7		
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11

Discusión

¿Que bases de datos son utilizadas?



Name	Year ref.	Description
VDJdb	2018 [20]	TCR binding to pMHC, contains 5491 samples.
IEDB	2018 [21]	The bigger database, contains information <i>T-cell epitopes</i>
TSNAdb	2018 [22]	It contains 7748 samples of mutations and HLA of 16 types of cancer.
NeoPeptide	2019 [23]	It contains samples of neoantigens resulting from somatic mutations and related items. 1818137 epitopes of more than 36000 neoantigens.
pHLA3D	2019 [24]	Presents 106 3D structures of the alpha, <i>beta2M</i> chains, and peptides of HLA-I molecules
dbPepNeo	2020 [25]	It has validated samples of the <i>peptide-MHC</i> bond, from MS. It contains 407794 low-quality samples, 247 medium-quality, and 295 high-quality samples.
dbPepNeo2.0	2022 [26]	It gathers a list of neoantigens and HLA molecules. It presents 801 high-quality and 842,289 poor-quality HLAs. Also, 55 class II neoantigens and 630 TCR-bound neo antigens.
IntroSpect	2022 [27]	Tool for building databases on <i>peptide-MHC binding</i> . It uses data from <i>Mass Spectrometry</i> .
IPD-IMGT/HLA	2022 [28]	With 25000 MHC molecules and 45 alleles.

Discusión

Que métodos basados en *deep learning* tiene los mejores resultados?



NetMHCpan4.1 es considerado el método *state-of-art*.

Recientemente, **HLAB** [29] ha propuesto el uso de transfer learning de un modelo basado en BERT (ProtTrans) seguido de una red BiLSTM. Con esta propuesta, supero ligeramente a NetMHCpan4.1.

TransPHLA [30] es *allele-specific* y según sus experimentos, supero ligeramente a NetMHCpan4.1 en algunos tipos de MHC.

- [1] 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>
- [2] NortShore, “Immunotherapy,” 2022. [Online]. Available: <https://www.northshore.org/kellogg-cancer-center/our-services/immunotherapy/>
- [3] NCI, “National cancer institute dictionary,” 2022. [Online]. Available: <https://www.cancer.gov/publications/dictionaries/genetics-dictionary>
- [4] 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.

- [5] 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.
- [6] 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.
- [7] 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.

- [8] 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.
- [9] S. Kim, H. S. Kim, E. Kim, M. Lee, E.-C. Shin, and S. Paik, “Neopepsee: accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information,” *Annals of Oncology*, vol. 29, no. 4, pp. 1030–1036, 2018.
- [10] A. Rubinsteyn, J. Kodysh, I. Hodes, S. Mondet, B. A. Aksoy, J. P. Finnigan, N. Bhardwaj, and J. Hammerbacher, “Computational pipeline for the pgv-001 neoantigen vaccine trial,” *Frontiers in immunology*, vol. 8, p. 1807, 2018.

- [11] 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.
- [12] 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.
- [13] 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.

- [14] 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.
- [15] M. A. Wood, A. Nguyen, A. J. Struck, K. Ellrott, A. Nellore, and R. F. Thompson, "Neoepiscope improves neoepitope prediction with multivariant phasing," *Bioinformatics*, vol. 36, no. 3, pp. 713–720, 2020.
- [16] A. C. M. Coelho, A. L. Fonseca, D. L. Martins, P. B. Lins, L. M. da Cunha, and S. J. de Souza, "neoant-hill: an integrated tool for identification of potential neoantigens," *BMC Medical Genomics*, vol. 13, no. 1, pp. 1–8, 2020.

- [17] C. Wert-Carvajal, R. Sánchez-García, J. R. Macías, R. Sanz-Pamplona, A. M. Pérez, R. Alemany, E. Veiga, C. Ó. S. Sorzano, and A. Muñoz-Barrutia, "Predicting mhc i restricted t cell epitopes in mice with nap-cnb, a novel online tool," *Scientific reports*, vol. 11, no. 1, pp. 1–10, 2021.
- [18] L. Y. Zhou, F. Zou, and W. Sun, "Prioritizing candidate peptides for cancer vaccines by peppermint: a statistical model to predict peptide presentation by hla-i proteins," *bioRxiv*, 2021.
- [19] Y. L. Terai, C. Huang, B. Wang, X. Kang, J. Han, J. Douglass, E. H.-C. Hsiue, M. Zhang, R. Purohit, T. deSilva *et al.*, "Valid-neo: A multi-omics platform for neoantigen detection and quantification from limited clinical samples," *Cancers*, vol. 14, no. 5, p. 1243, 2022.

- [20] M. Shugay, D. V. Bagaev, I. V. Zvyagin, R. M. Vroomans, J. C. Crawford, G. Dolton, E. A. Komech, A. L. Sycheva, A. E. Koneva, E. S. Egorov *et al.*, “Vdjdb: a curated database of t-cell receptor sequences with known antigen specificity,” *Nucleic acids research*, vol. 46, no. D1, pp. D419–D427, 2018.
- [21] R. Vita, S. Mahajan, J. A. Overton, S. K. Dhanda, S. Martini, J. R. Cantrell, D. K. Wheeler, A. Sette, and B. Peters, “The immune epitope database (iedb): 2018 update,” *Nucleic acids research*, vol. 47, no. D1, pp. D339–D343, 2018.
- [22] J. Wu, W. Zhao, B. Zhou, Z. Su, X. Gu, Z. Zhou, and S. Chen, “Tsnadb: a database for tumor-specific neoantigens from immunogenomics data analysis,” *Genomics, proteomics & bioinformatics*, vol. 16, no. 4, pp. 276–282, 2018.

- [23] W.-J. Zhou, Z. Qu, C.-Y. Song, Y. Sun, A.-L. Lai, M.-Y. Luo, Y.-Z. Ying, H. Meng, Z. Liang, Y.-J. He *et al.*, “Neopeptide: an immunoinformatic database of t-cell-defined neoantigens,” *Database*, vol. 2019, 2019.
- [24] 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.
- [25] X. Tan, D. Li, P. Huang, X. Jian, H. Wan, G. Wang, Y. Li, J. Ouyang, Y. Lin, and L. Xie, “dbpepneo: a manually curated database for human tumor neoantigen peptides,” *Database*, vol. 2020, 2020.

- [26] M. Lu, L. Xu, X. Jian, X. Tan, J. Zhao, Z. Liu, Y. Zhang, C. Liu, L. Chen, Y. Lin *et al.*, “dbpepneo2. 0: A database for human tumor neoantigen peptides from mass spectrometry and tcr recognition,” *Frontiers in immunology*, p. 1583, 2022.
- [27] L. Zhang, G. Liu, G. Hou, H. Xiang, X. Zhang, Y. Huang, X. Zhang, B. Li, and L. J. Lee, “Introspect: Motif-guided immunopeptidome database building tool to improve the sensitivity of hla i binding peptide identification by mass spectrometry,” *Biomolecules*, vol. 12, no. 4, p. 579, 2022.
- [28] J. Robinson, D. J. Barker, X. Georgiou, M. A. Cooper, P. Flicek, and S. G. Marsh, “Ipd-imgt/hla database,” *Nucleic acids research*, vol. 48, no. D1, pp. D948–D955, 2020.

- [29] Y. Zhang, G. Zhu, K. Li, F. Li, L. Huang, M. Duan, and F. Zhou, "Hlab: learning the bilstm features from the protbert-encoded proteins for the class i hla-peptide binding prediction," *Briefings in Bioinformatics*, 2022.
- [30] Y. Chu, Y. Zhang, Q. Wang, L. Zhang, X. Wang, Y. Wang, D. R. Salahub, Q. Xu, J. Wang, X. Jiang *et al.*, "A transformer-based model to predict peptide–hla class i binding and optimize mutated peptides for vaccine design," *Nature Machine Intelligence*, vol. 4, no. 3, pp. 300–311, 2022.

Questions?

