LaSalle Universidad

DIRECCIÓN DE INVESTIGACIÓN

Universidad Católica San Pablo

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 interno en colaboración con la UCSP

PhD(c). Vicente Machaca Arceda

Contenido

Presentación

- PhD(c). Vicente Enrique Machaca Arceda.
- Profesor UTEC (TP).
- Investigador en Bioinformática y Aprendizaje de Maquina.
- Index-h 5 .

Presentación
Publicaciones

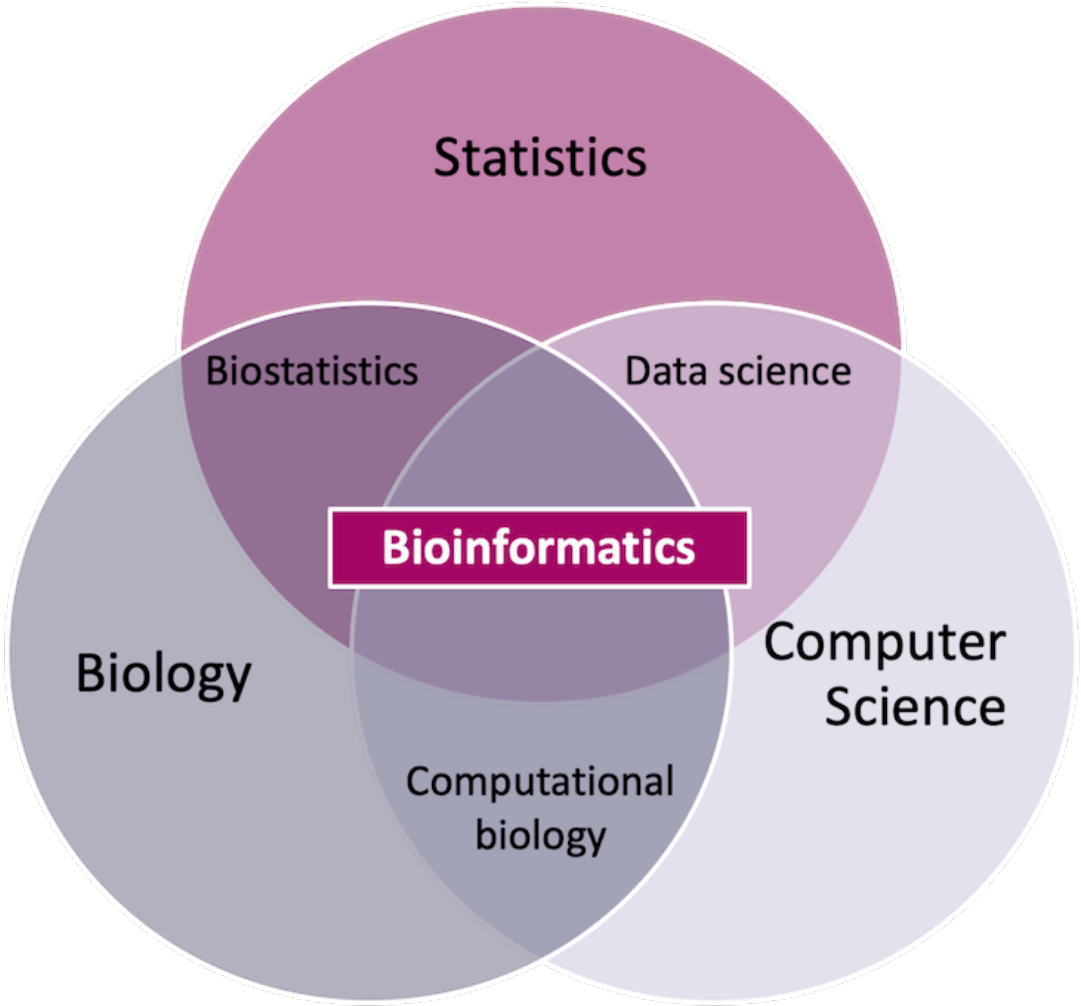
Year	Country	Title
2020	USA	Small Ship Detection on Optical Satellite Imagery with YOLO and YOLT
2018	Brasil	Fast Car Crash Detection in Video
2016	Chile	Fast Face Detection in Violent Video Scenes
2016	Costa Rica	Real Time Violence Detection in Video with ViF and Horn-Schunck
2016	Costa Rica	Optimization model for face detection in video sequences
2015	Chile	Real Time Violence Detection in Video

Presentation
Publications

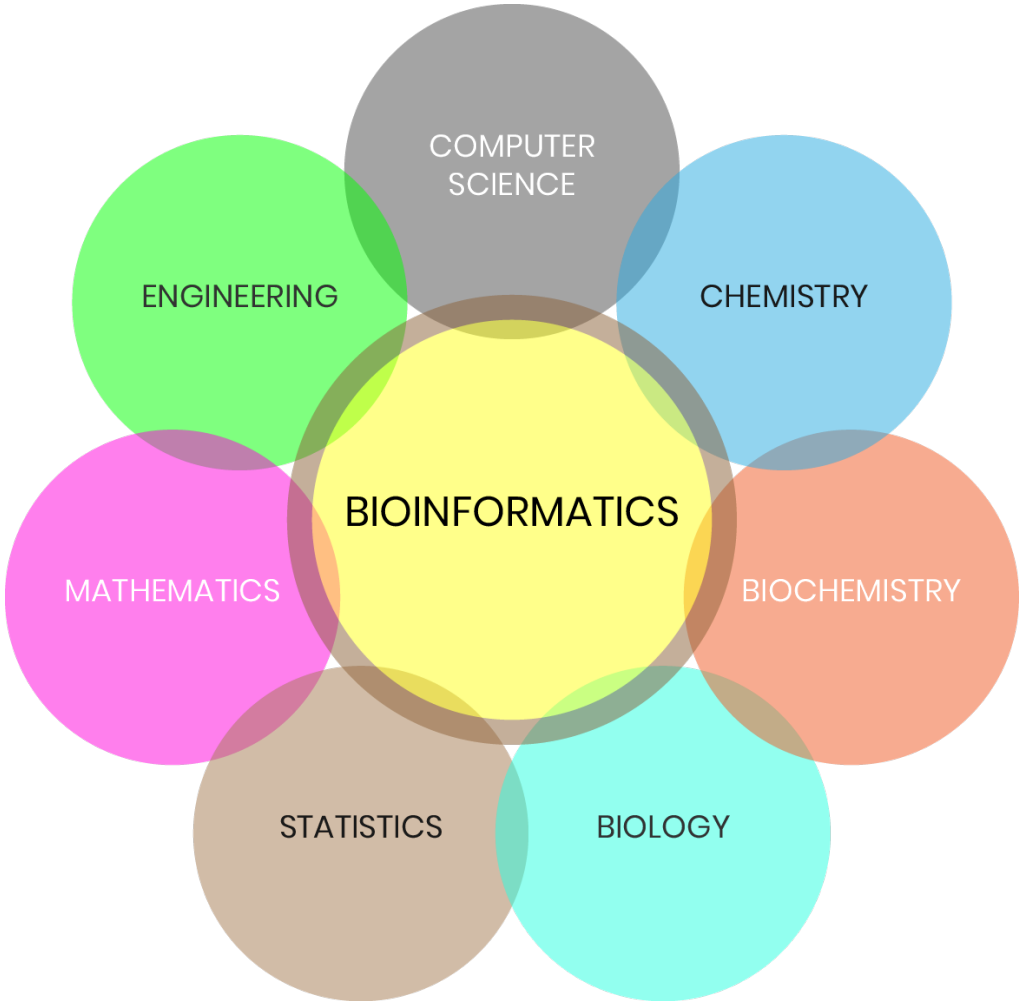
Year	Country	Title
2022	USA	ArgosMol: A Web Tool for Protein Structure Prediction and Visualization
2021	Chapter	COVID-19 Pandemic: Analysis and Statistics of Confirmed Cases
2020	Canada	An Analysis of k-Mer Frequency Features with Machine Learning Models for Viral Sub-typing of Polyomavirus and HIV-1 Genomes
2020	Canada	An analysis of k-mer frequency features with SVM and CNN for viral subtyping classification
2020	Canada	Forecasting time series with Multiplicative Trend Exponential Smoothing and LSTM: COVID-19 case study

Contenido

Bioinformática



Bioinformática



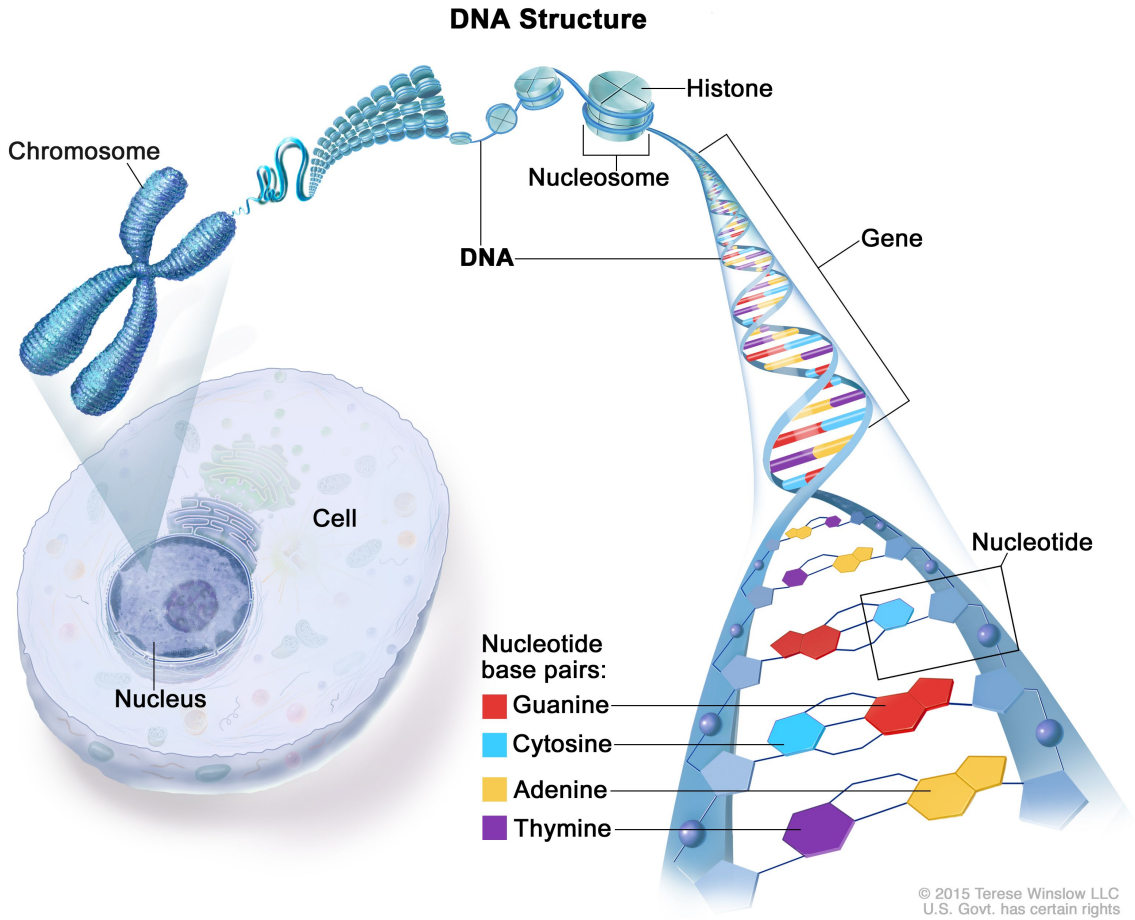


Figure: Where DNA is located [1].

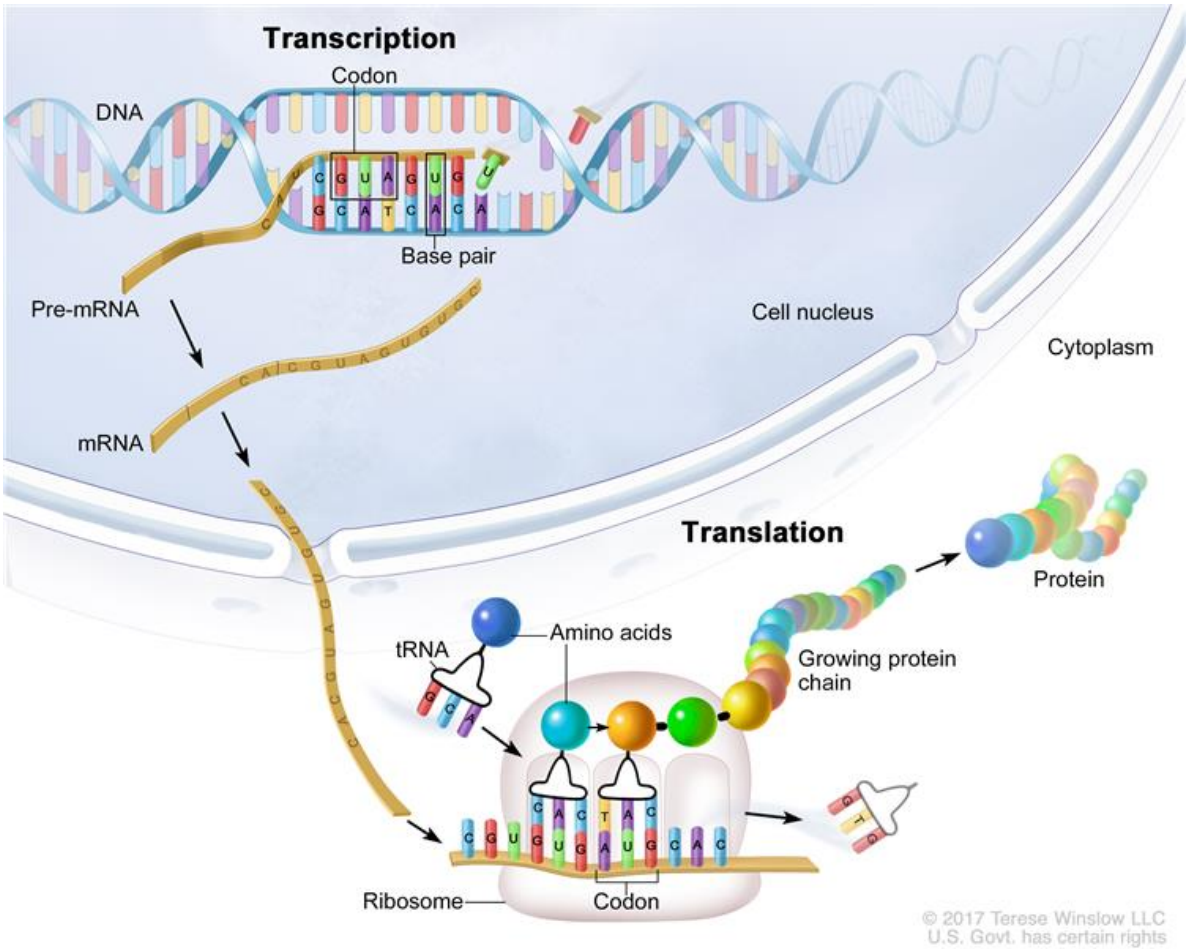


Figure: Transcription and translation [2].

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

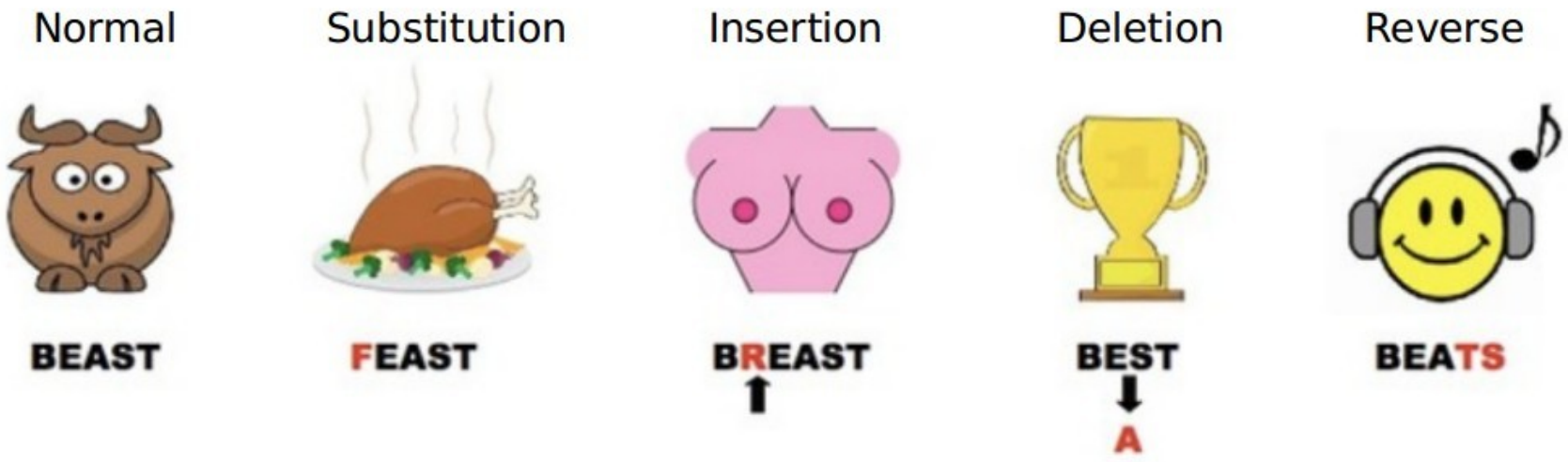


Figure: Overview of the Different Types of Point Mutations.

Variantes y Mutaciones

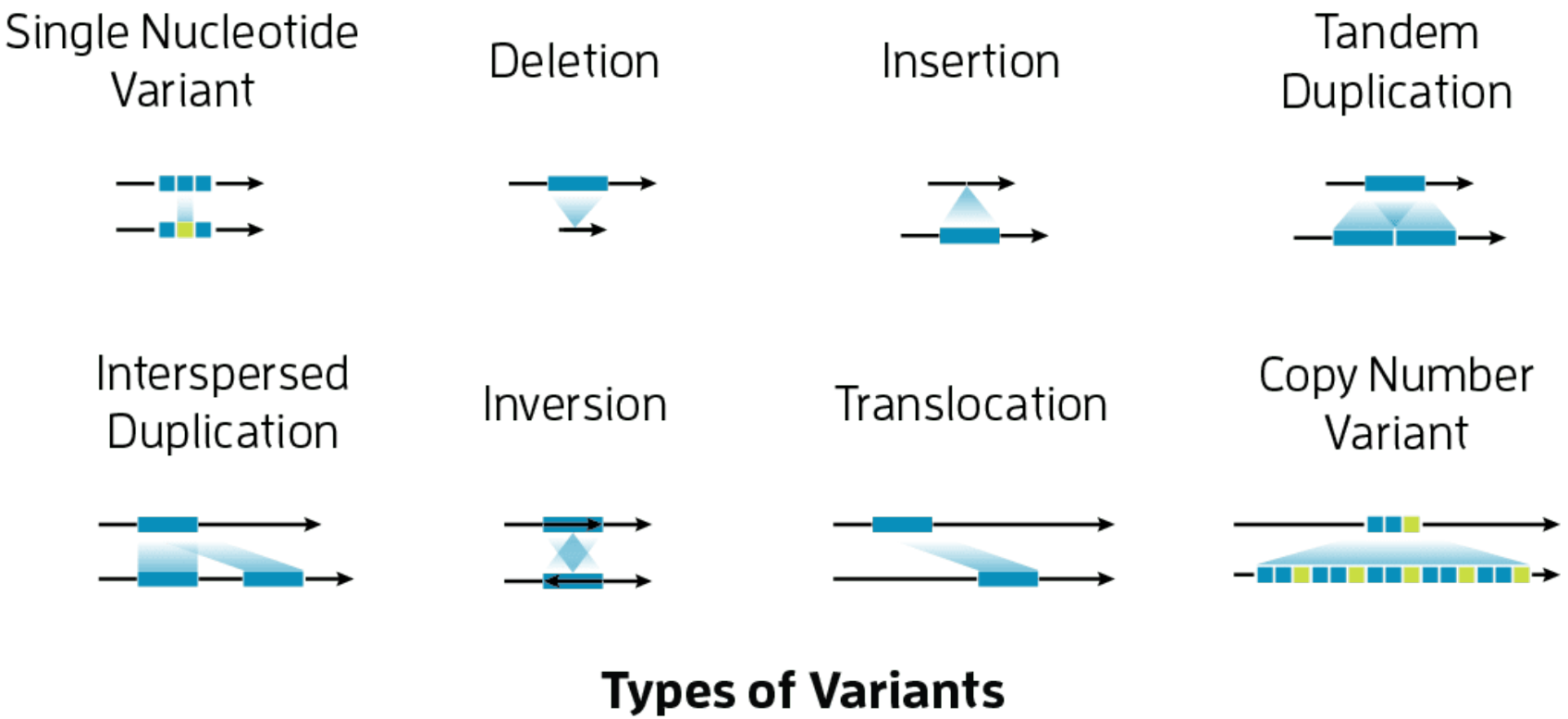


Figure: Example of structural variants. Source: [3]

Variantes y Mutaciones

Frameshift

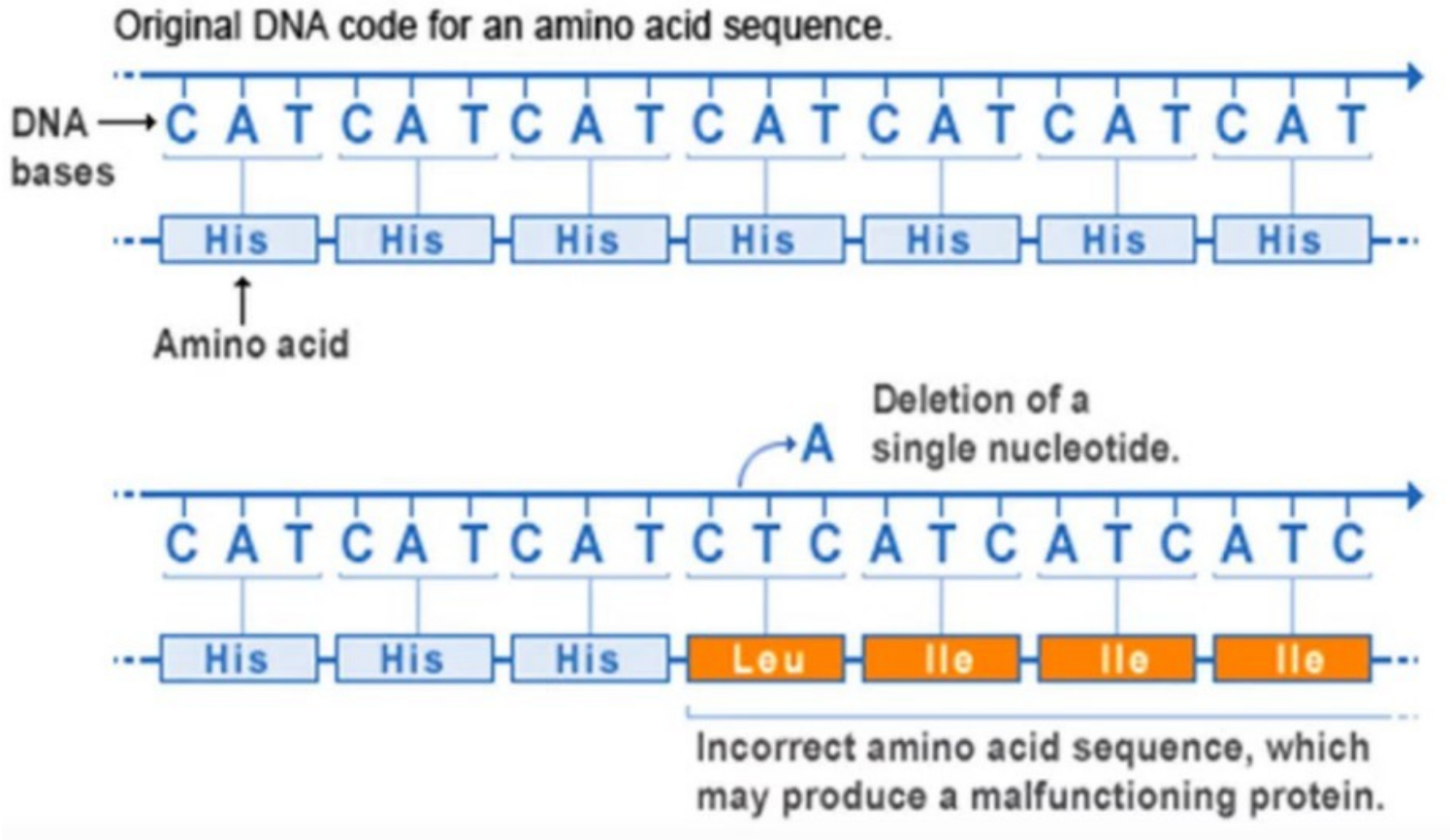


Figure: Ejemplo de una mutación INDELS causante de un *frameshift*.

Variantes y Mutaciones
Frameshift

Los *Frameshift variants* estan muy relacionados a la enfermedad Tay–Sachs (destrucción de células nerviosas). Tambien, incrementa la susceptibilidad a varios tipos de Cáncer [4, 5].

Fusión de genes

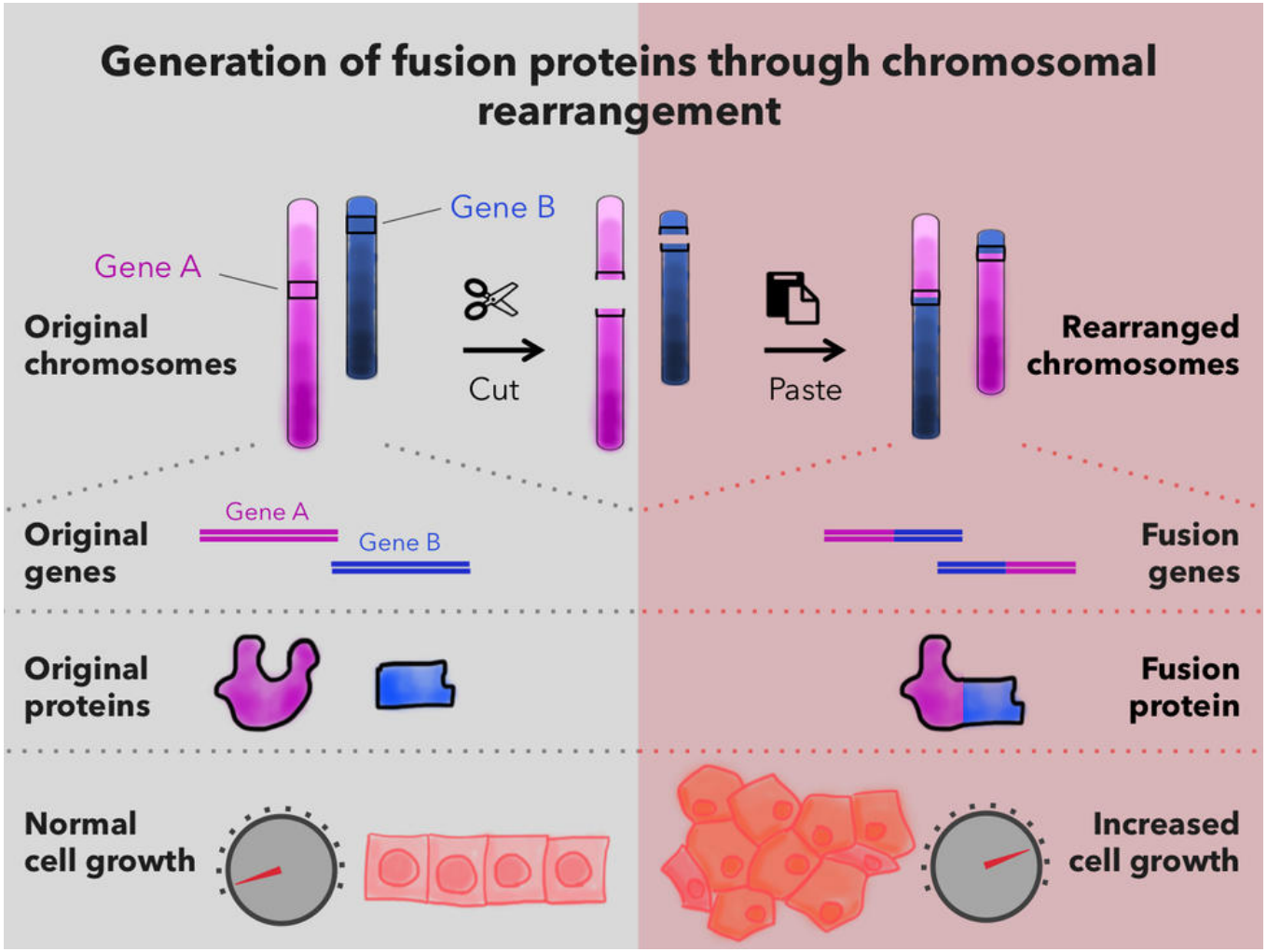


Figure: Ejemplo de una fución de genes.

Variaciones a nivel de cromosomas

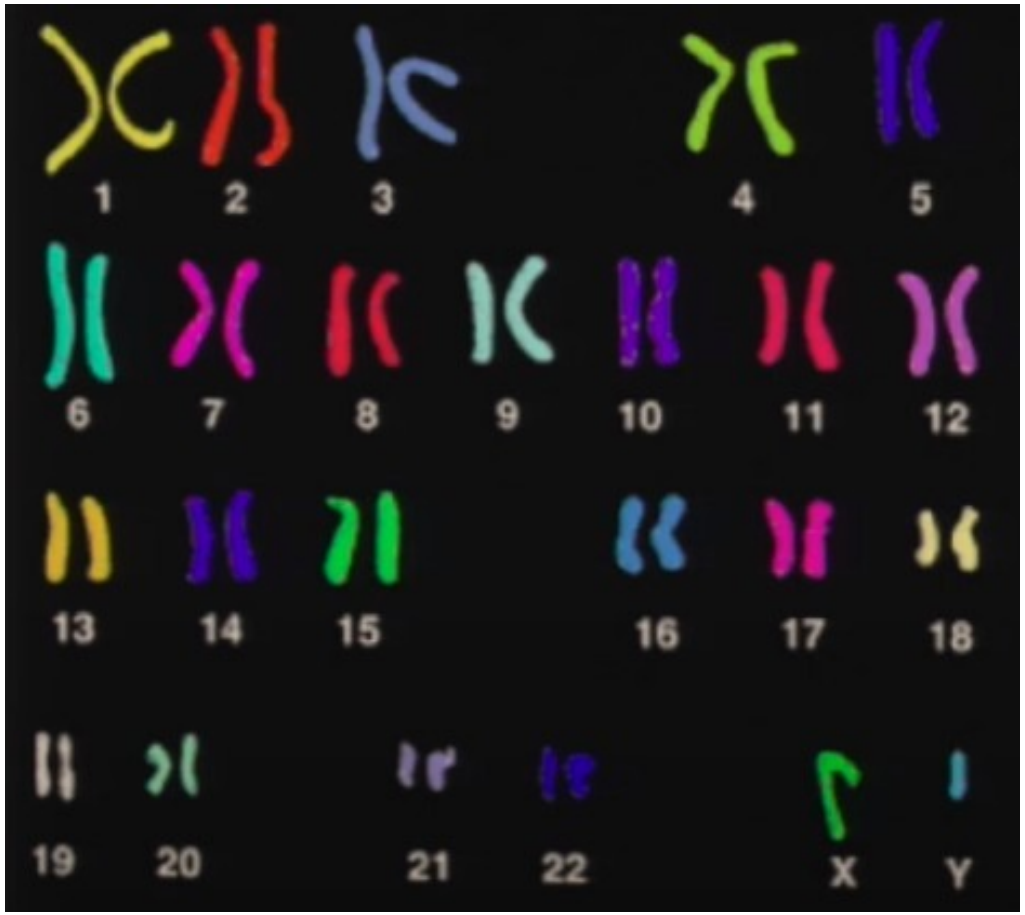


Figure: Los 46 cromosomas presentes en una célula.

Variaciones a nivel de cromosomas

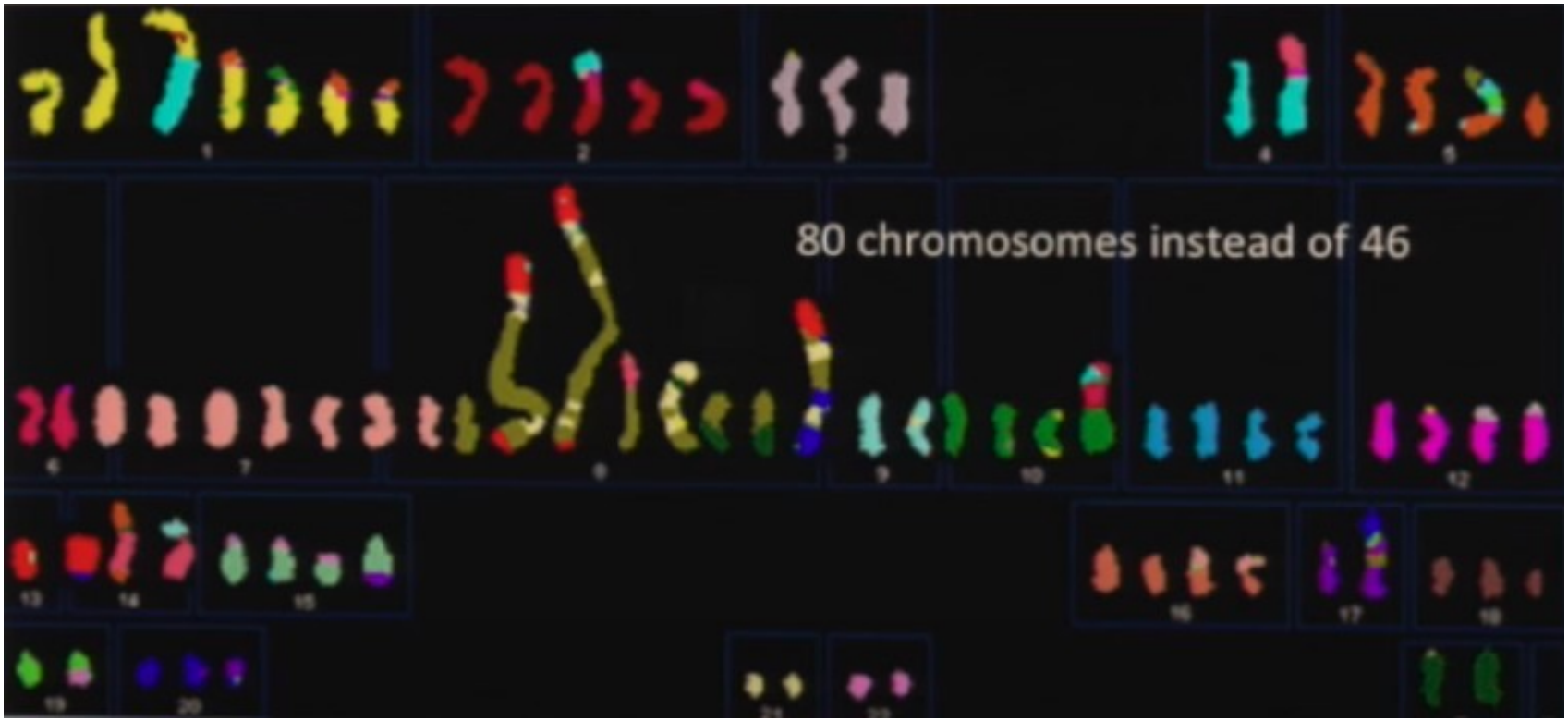


Figure: Cromosomas de una mujer con Cáncer de mama (1971).

Inmunoterapia del Cáncer

Es un tipo de tratamiento contra el Cáncer que estimula las defensas naturales del cuerpo para combatir el Cáncer [6].

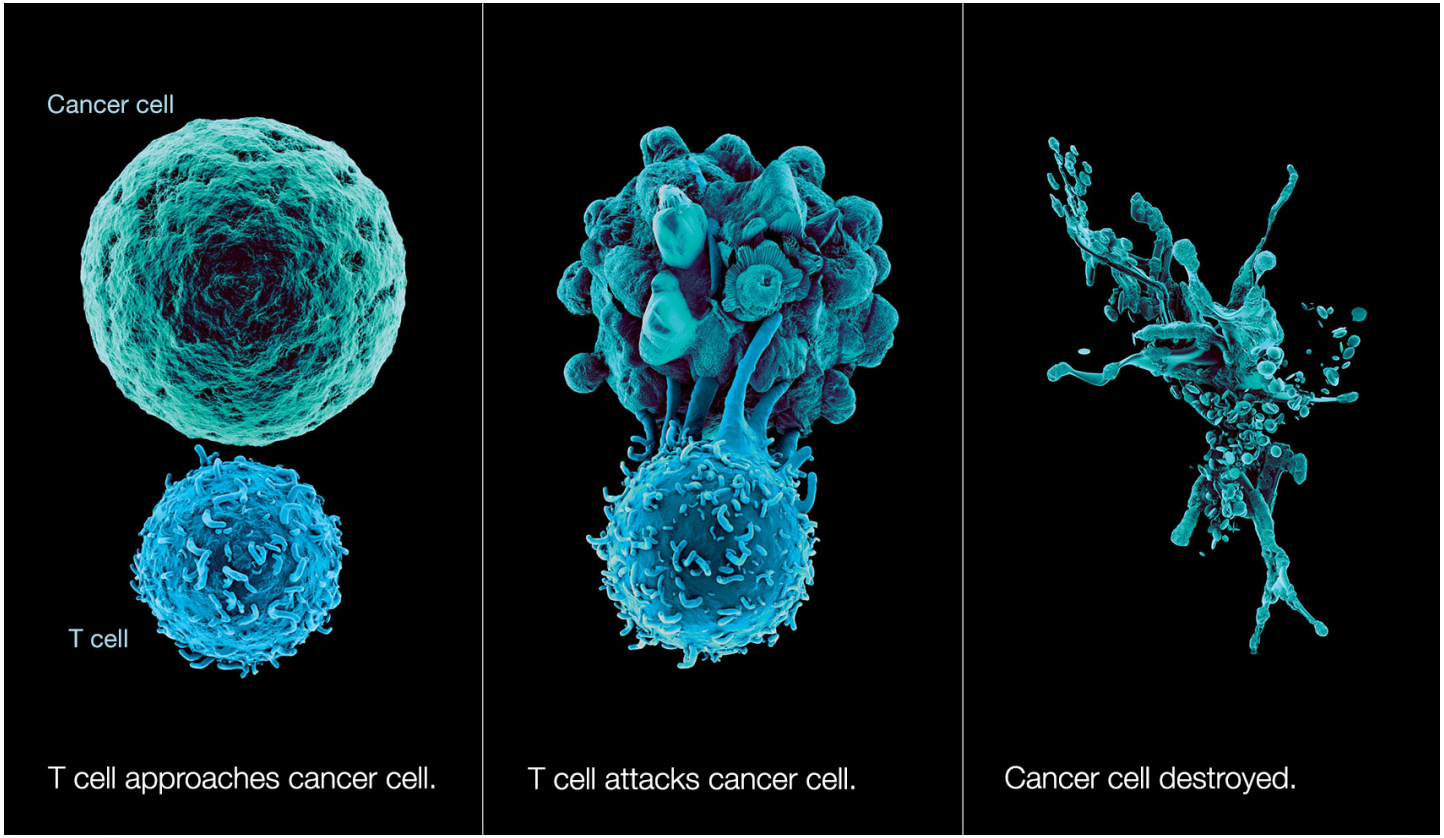


Figure: Ejemplo de como una célula T destruye células del cancer [7].

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, 8].

En la actualidad hay varios métodos para detectar a predecir neo antígenos, pero **solo una pequeña cantidad de ellos** logran estimular al sistema inmune [9, 10].

MHC-I

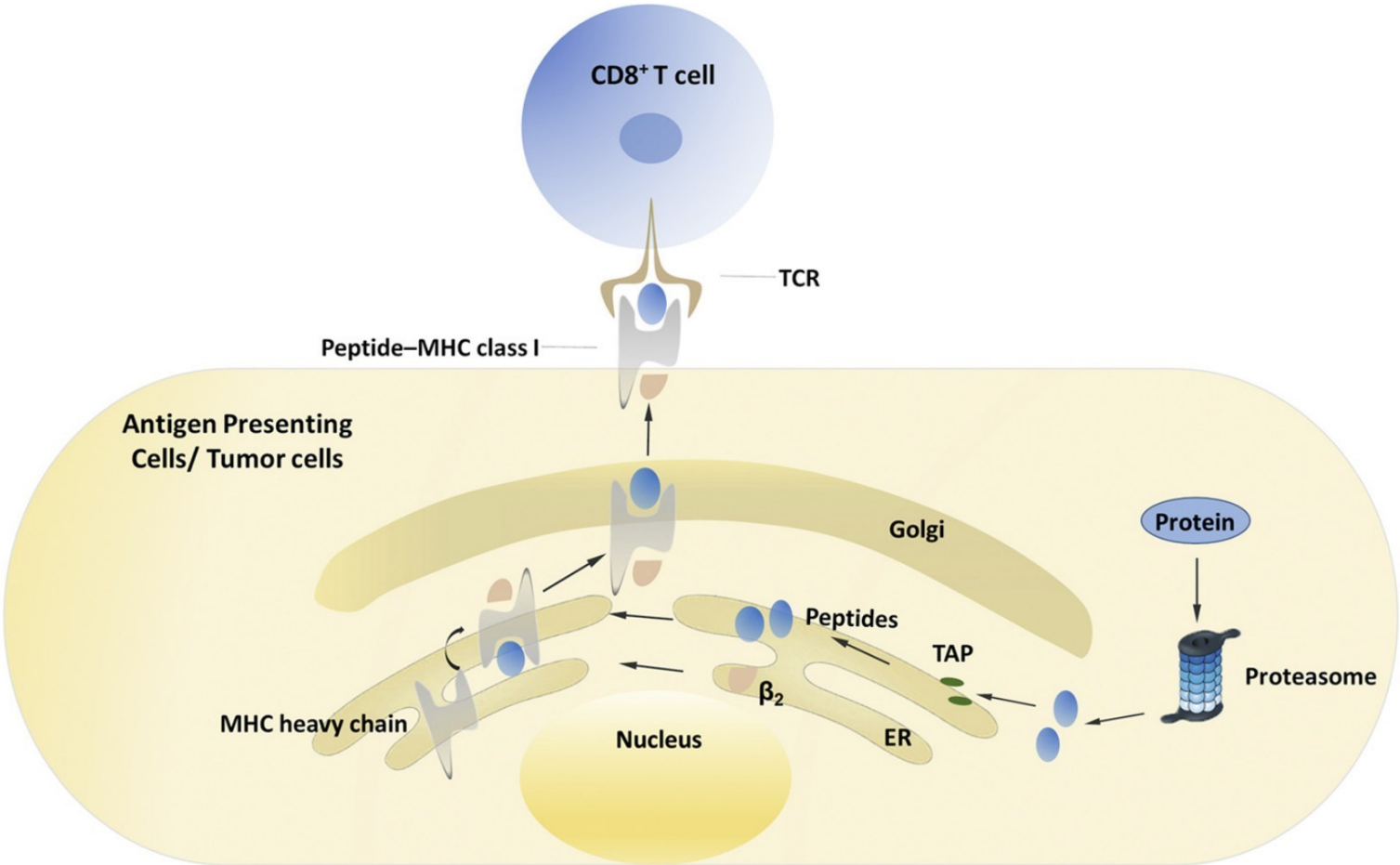


Figure: Presentación de antígenos por MHC-I. Fuente: [11]

MHC-II

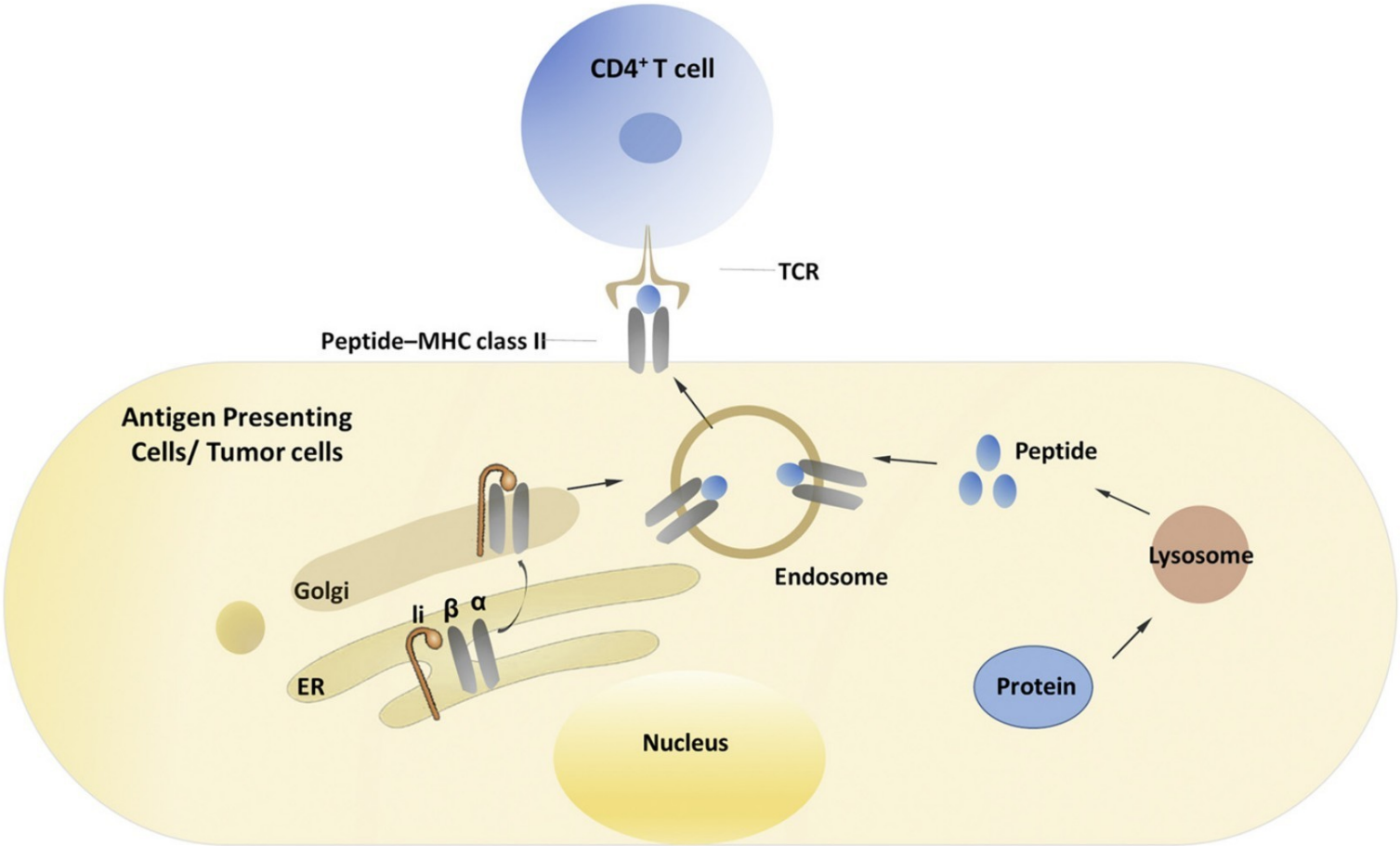


Figure: Presentación de antígenos por MHC-II. Fuente: [11]

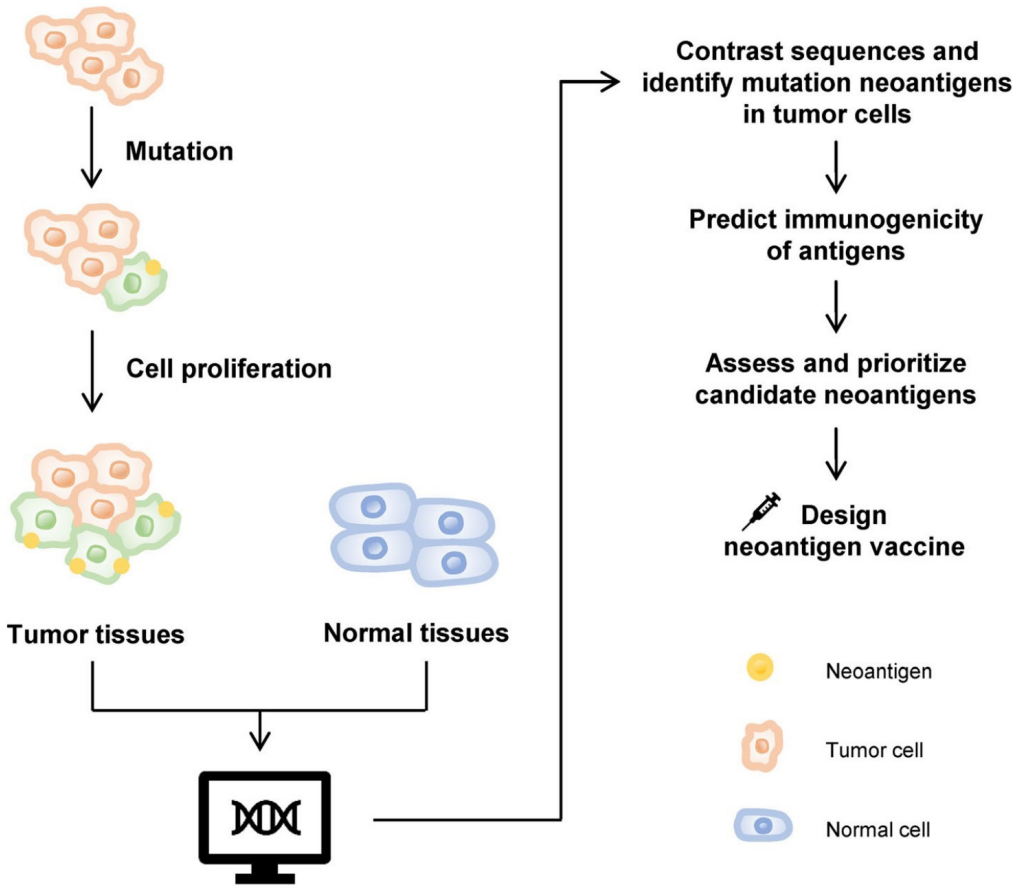


Figure: Proceso para la generación de vacunas personalizadas [12].

Contenido

Motivación

El cáncer representa el mayor problema de salud mundial, pero lamentablemente los métodos basados en cirugías, radioterapias, quimioterapias tienen baja efectividad [12].

La inmunoterapia del cáncer es una alternativa para el desarrollo de vacunas personalizadas, pero este proceso depende de una correcta detección de neo antígenos [13, 12].

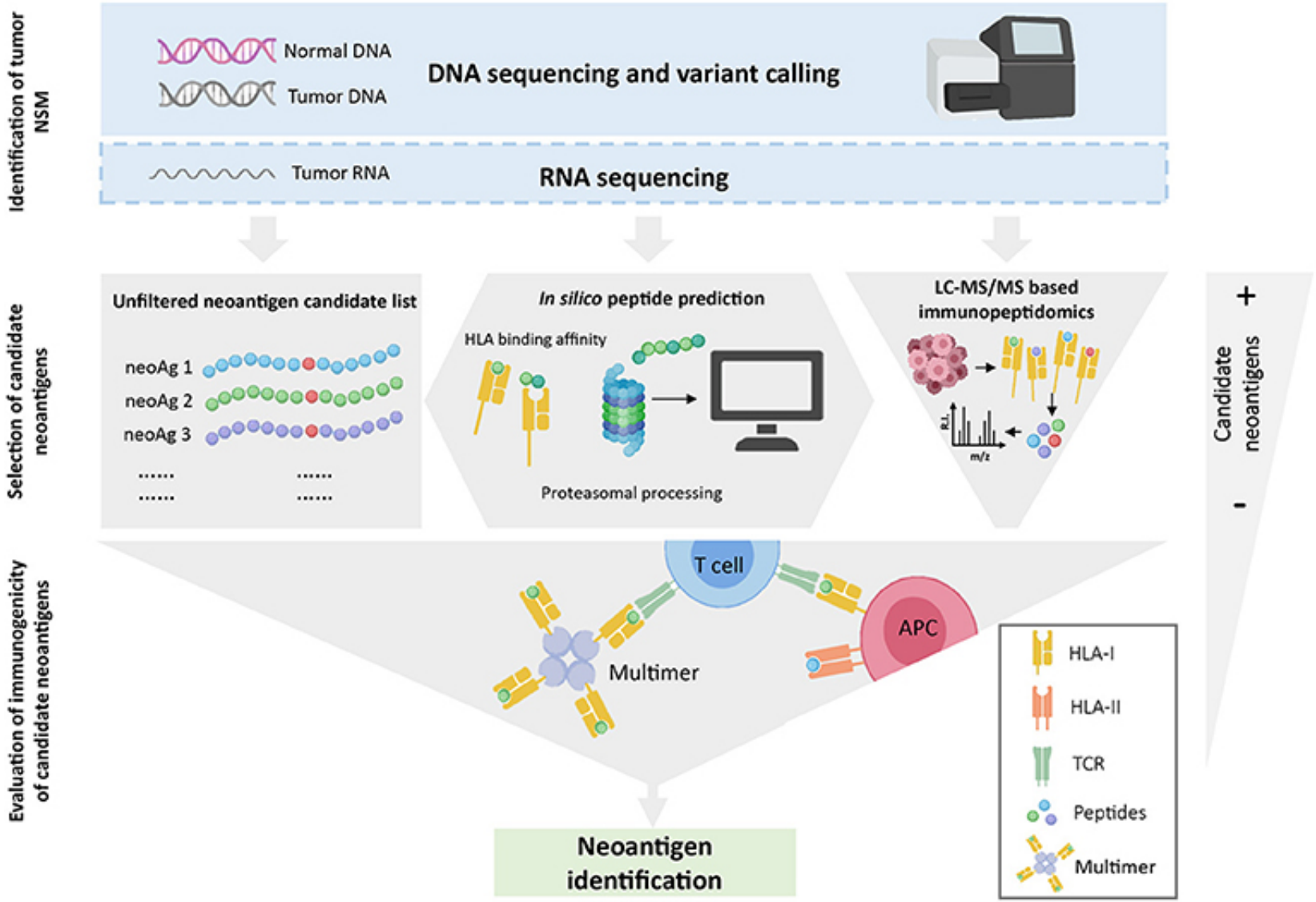
Problema

Menos del 3% de los neoantígenos detectados logran activar a las células T (sistema inmune) [13].

Objetivo general

Desarrollar un método basado en *deep learning* que mejore el acierto de la detección de neo antígenos.

Objetivos


Identification of tumor NSM
Normal DNA
Tumor DNA
DNA sequencing and variant calling
Tumor RNA
RNA sequencing
Selection of candidate neoantigens
Unfiltered neoantigen candidate list
neoAg 1
neoAg 2
neoAg 3
In silico peptide prediction
HLA binding affinity
Proteasomal processing
LC-MS/MS based immunopeptidomics
Candidate neoantigens
+
-
Evaluation of immunogenicity of candidate neoantigens
T cell
APC
Multimer
HLA-I
HLA-II
TCR
Peptides
Multimer
Neoantigen identification

Contenido

Estado del arte
Pipelines

Año	Nombre	Referencia
2020	ProGeo-neo	[14]
2020	INeo-Epp	[15]
2020	pVACtools	[16]
2019	NeoPredPipe	[17]
2019	DeepHLApan	[18]
2019	ScanNeo	[19]
2017	CloudNeo	[20]
...	...	...

Estado del arte

Peptide-MHC binding

Año	Nombre	Modelo	Referencia
2022	AEM	Transformer	[21]
2021	**BERTMHC**	Transformer	[22]
2021	**APPM**	3 CNN	[10]
2020	NetMHCpan4.1	ANN	[23]
2020	MHCflurry2.0	ANN	[24]
2020	MHCnuggets	ANN	[25]
2019	PUFFIN	ANN	[26]
..	...	...	...

Contenido

Propuesta

La propuesta se basa el los modelos BERTMHC [22] y APPM [10].

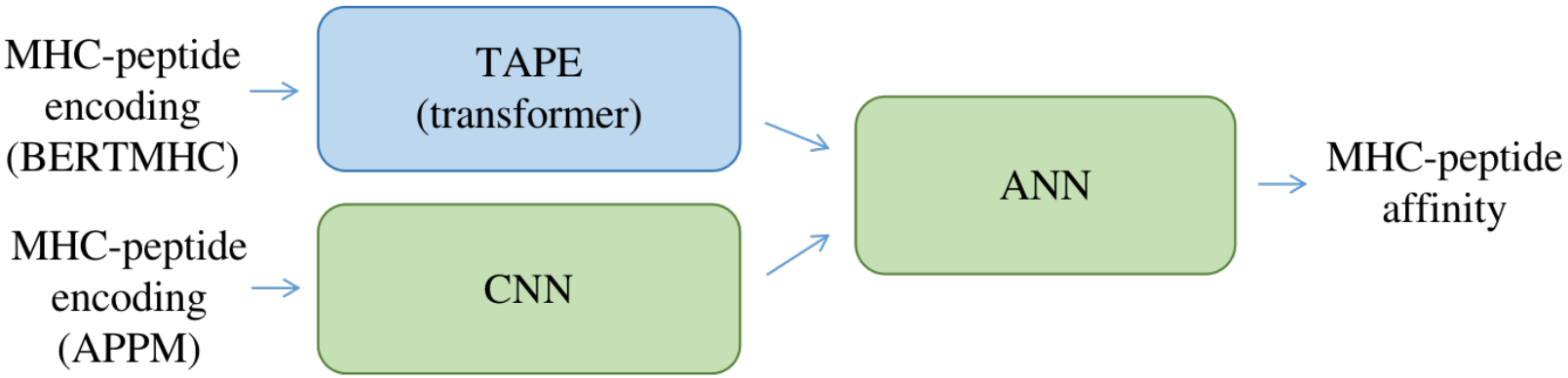


Figure: Modelo propuesto para la predicción del enlace péptido y MHC.

Propuesta
BERTMHC

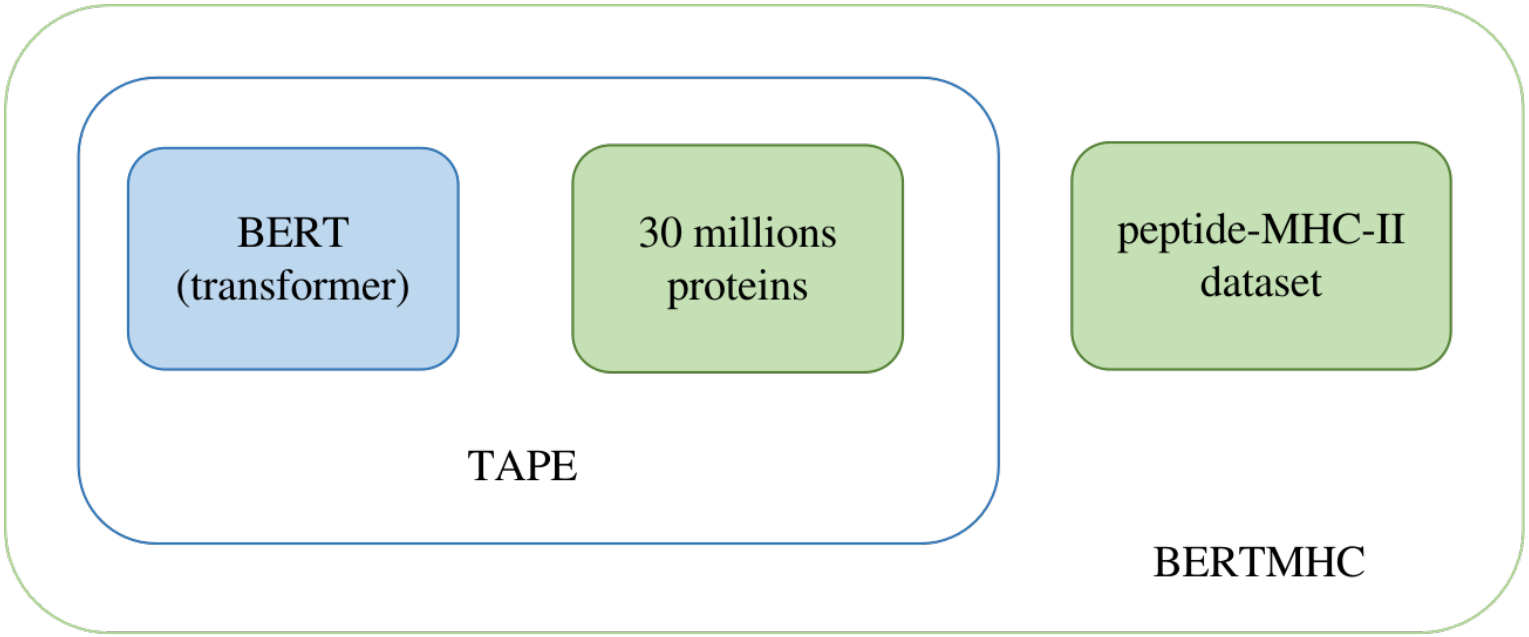


Figure: BERTMHC.

Propuesta
APPM

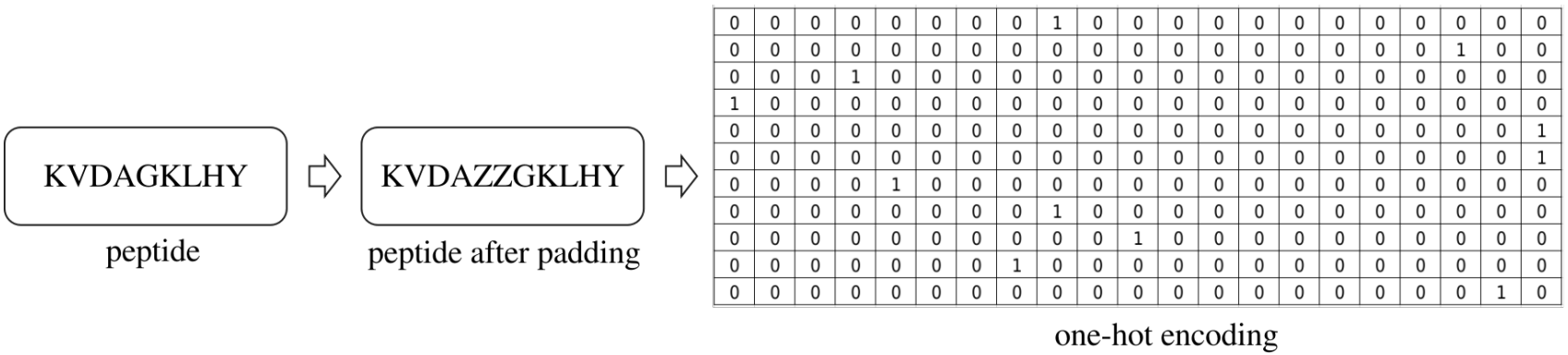


Figure: Proceso para obtener una matriz (imagen) a partir de un péptido (APPM).

Contenido

Base de datos

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

Alleles	**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

Resultados

Table: Resultados obtenidos en cada base de datos.

Allele	*Accuracy*	*F1 score*	*Precision*	*Recall*
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

Resultados

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

Modelo	**AUC**
Propuesta	0.73
BERTMHC	0.72
NetMHCpan3	0.68
PUFFIN	0.69
MHCnuggets	0.58

Contenido

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 llego 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 paso mas importantes par calificar al péptido como un neo antígeno, capaz de generar una respuesta inmunitaría.

Contenido

Trabajos futuros

Recientemente un trabajo [21] tambien 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 [27], 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.

References I

[1] NCI, “National cancer institute dictionary,” 2022. [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/

References II

[4] P. A. Zimmerman, A. Buckler-White, G. Alkhatib, T. Spalding, J. Kubofcik, C. Combadiere, D. Weissman, O. Cohen, A. Rubbert, G. Lam *et al.*, "Inherited resistance to hiv-1 conferred by an inactivating mutation in cc chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background, and quantified risk," *Molecular medicine*, vol. 3, no. 1, pp. 23–36, 1997.

[5] C. Xu, "A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data," *Computational and structural biotechnology journal*, vol. 16, pp. 15–24, 2018.

[6] Cancer.net. (2022) Qué es la inmunoterapia. [Online]. Available: https://www.cancer.net/es/desplazarse-por-atencion-del-cáncer/como-se-trata-el-cáncer/inmunoterapia/qué-es-la-inmunoterapia

References III

[7] NortShore, “Immunotherapy,” 2022. [Online]. Available: https://www.northshore.org/kellogg-cancer-center/our-services/immunotherapy/

[8] 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.

[9] 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.

[10] 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.

References IV

[11] 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.

[12] 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.

[13] 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.

References V

[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] 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.

[16] 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.

References VI

[17] 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.

[18] 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.

[19] T.-Y. Wang, L. Wang, S. K. Alam, L. H. Hoeppner, and R. Yang, "Scanneo: identifying indel-derived neoantigens using rna-seq data," *Bioinformatics*, vol. 35, no. 20, pp. 4159–4161, 2019.

References VII

[20] 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.

[21] 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.

[22] 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.

References VIII

[23] B. Reynisson, B. Alvarez, S. Paul, B. Peters, and M. Nielsen, "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 research*, vol. 48, no. W1, pp. W449–W454, 2020.

[24] 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.

[25] X. M. Shao, R. Bhattacharya, J. Huang, I. 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 mhcnuggetshigh-throughput prediction of neoantigens with mhcnuggets," *Cancer immunology research*, vol. 8, no. 3, pp. 396–408, 2020.

References IX

[26] 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.

[27] 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.

Questions?