Universidad Nacional de San Agustín

Detección *in Silico* de Neoantígenos Utilizando Transformers y Transfer Learning en el Marco de Desarrollo de Vacunas Personalizadas para Tratar el Cáncer

MSc. Vicente Machaca Arceda

2023

Contenido

Contenido

Contexto y Motivación

An la actualidad, el cáncer representa el mayor problema de salud mundial [1].

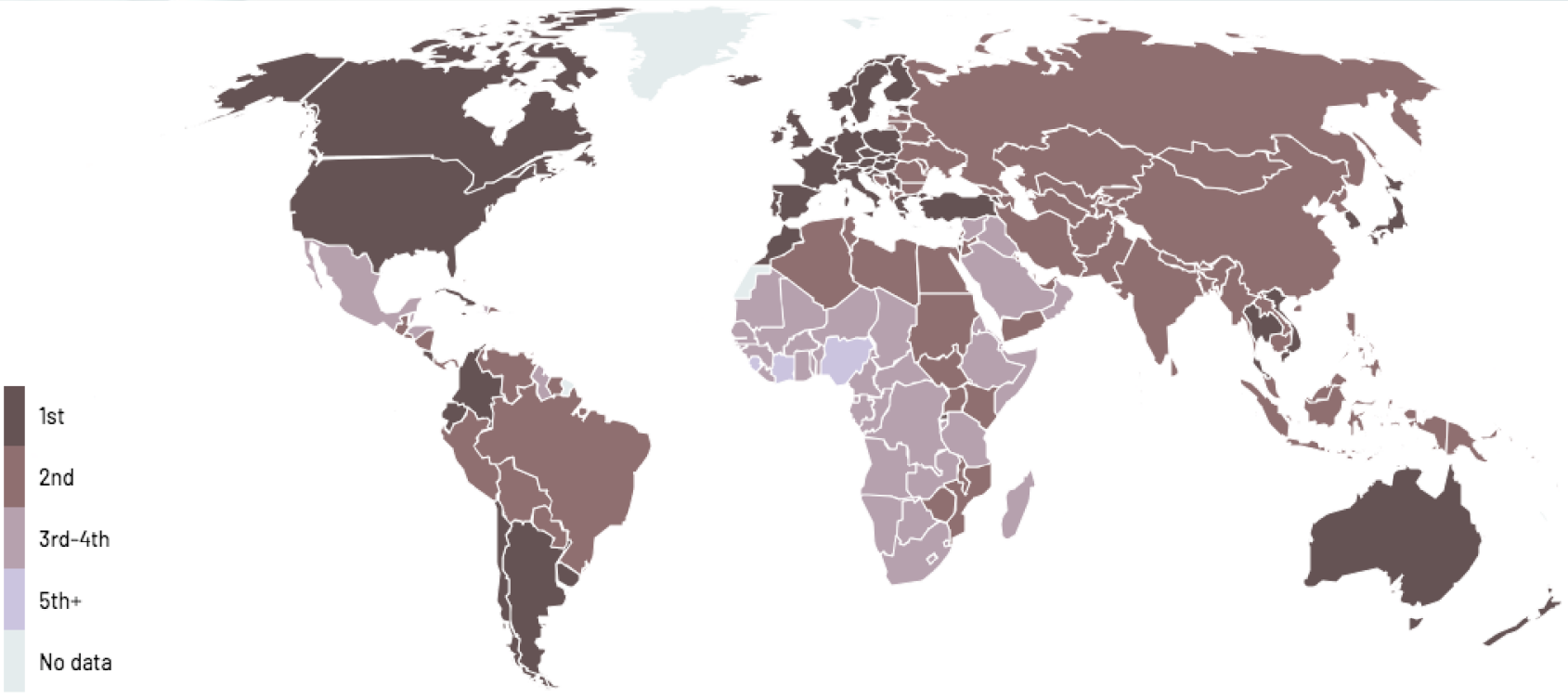


Figure: Ranking de las muertes por cáncer entre 30 y 69 años. **Fuente**: The Atlas Cancer [2].

Contexto y Motivación
Muertes por tipos de cáncer

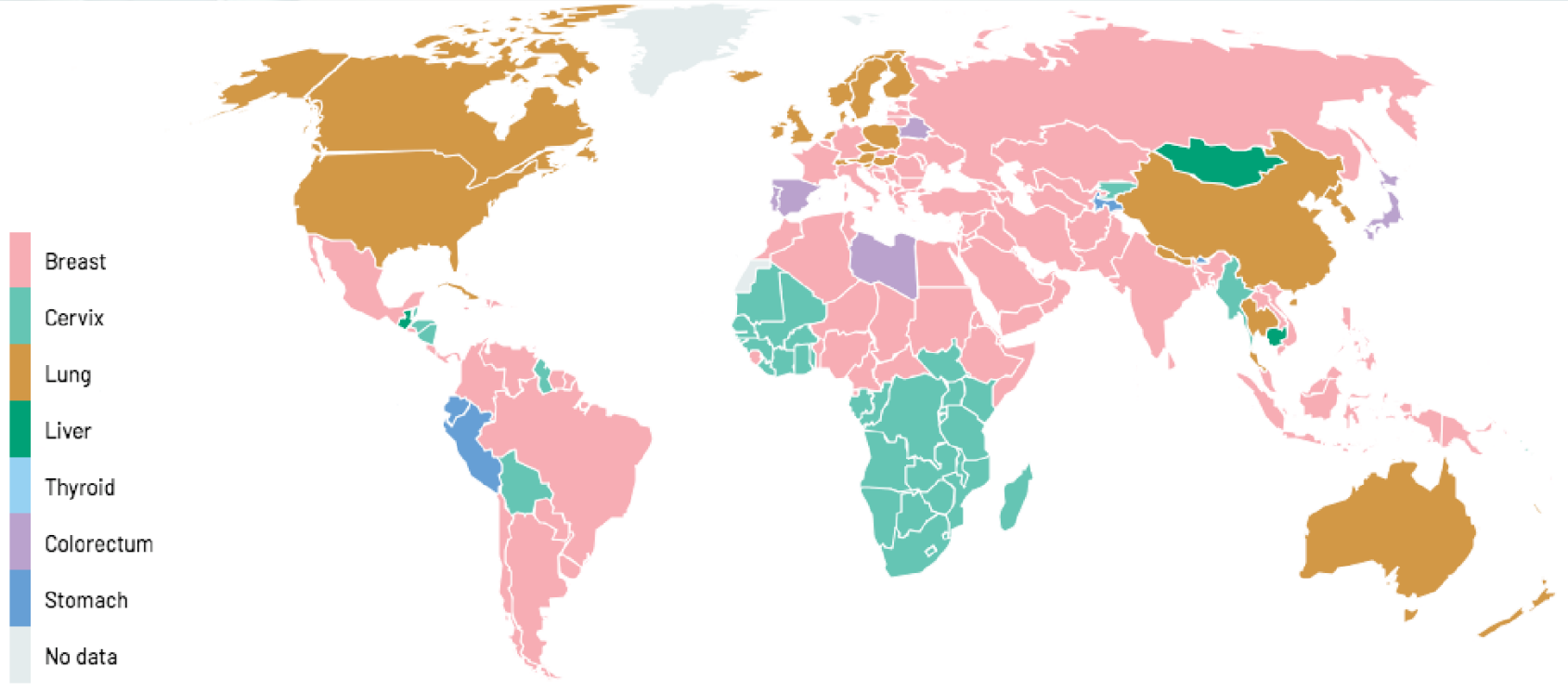


Figure: Ranking de las muertes por tipo de cáncer en mujeres. **Fuente**: The Atlas Cancer [2].

Contexto y Motivación

Muertes por tipos de cáncer

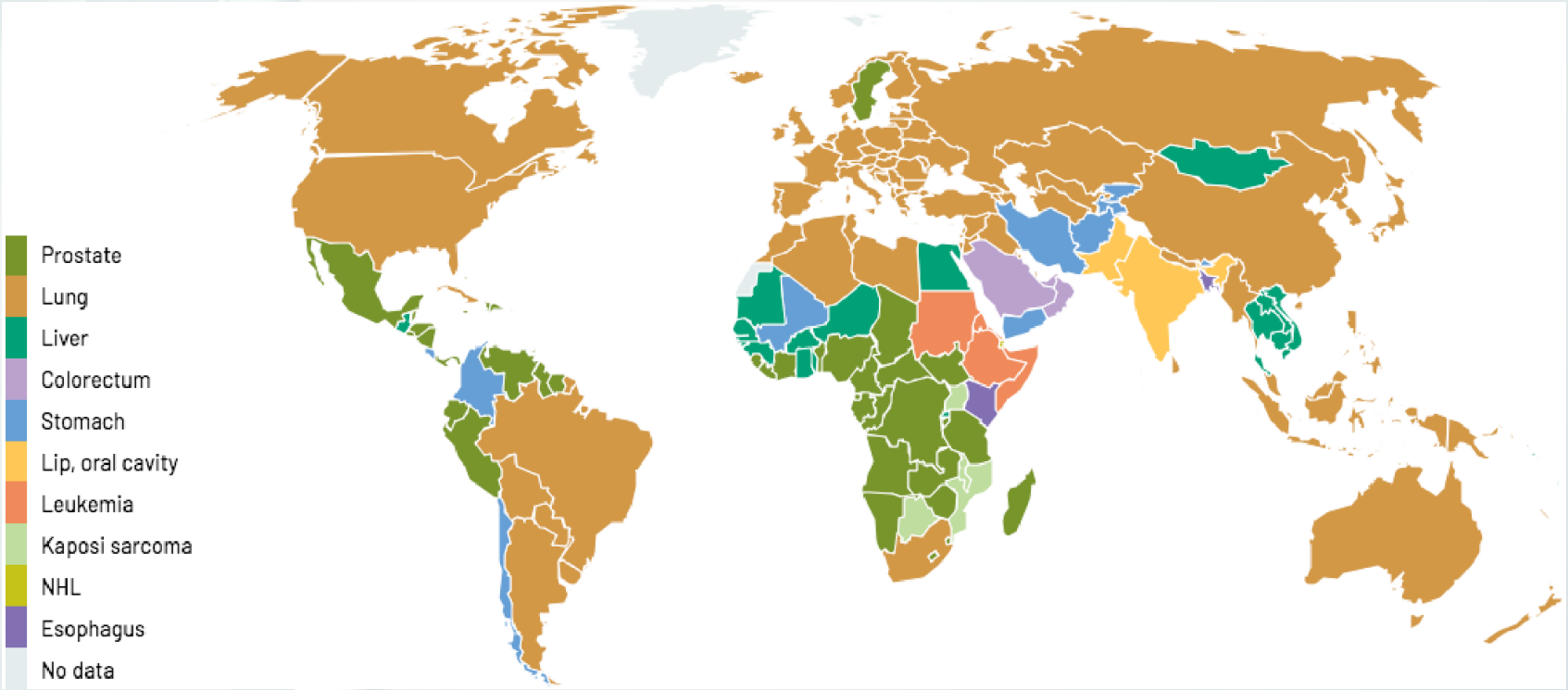


Figure: Ranking de las muertes por tipo de cáncer en hombres. **Fuente**: The Atlas Cancer [2].

Contexto y Motivación

Porcentaje de casos y muertes

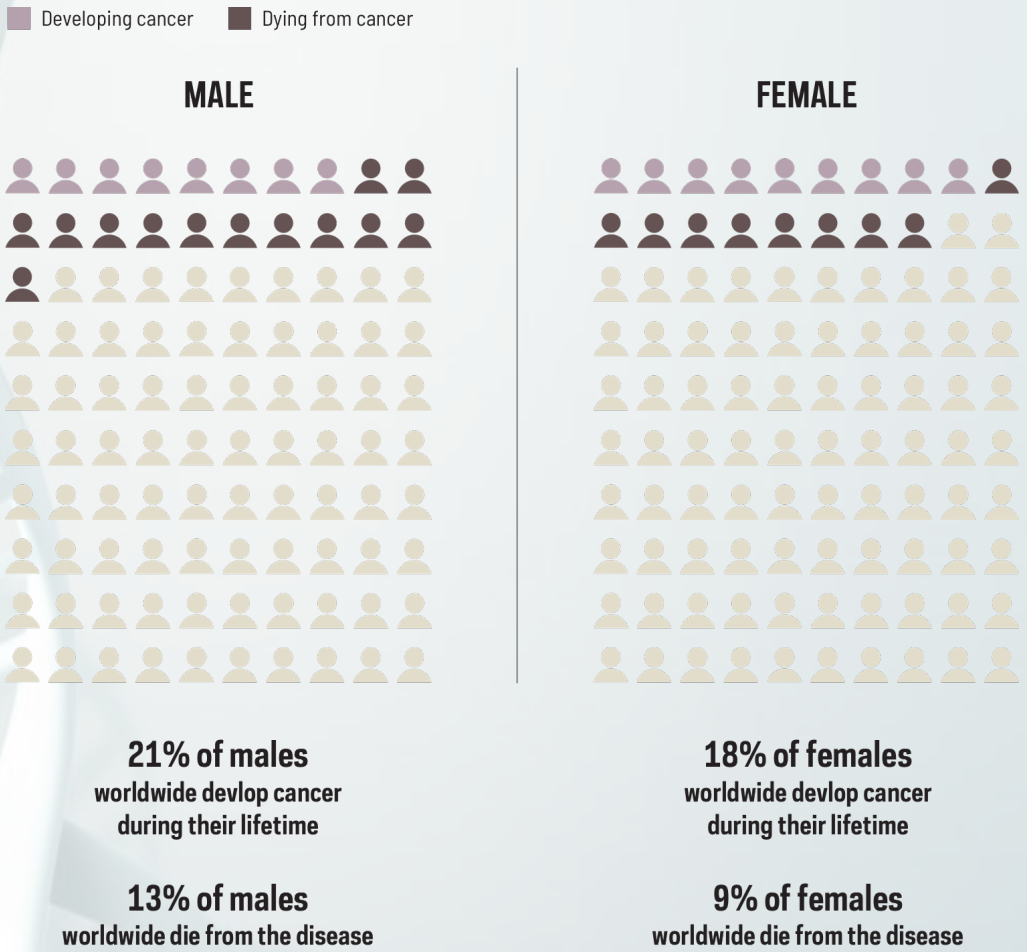


Figure: Porcentaje de casos y muertes por sexo. **Fuente** The Atlas Cancer [2].

Contexto y Motivación

Predicción de nuevos casos

Figure: Predicción de nuevos casos para el 2040. **Fuente** The Atlas Cancer [2].

Contenido

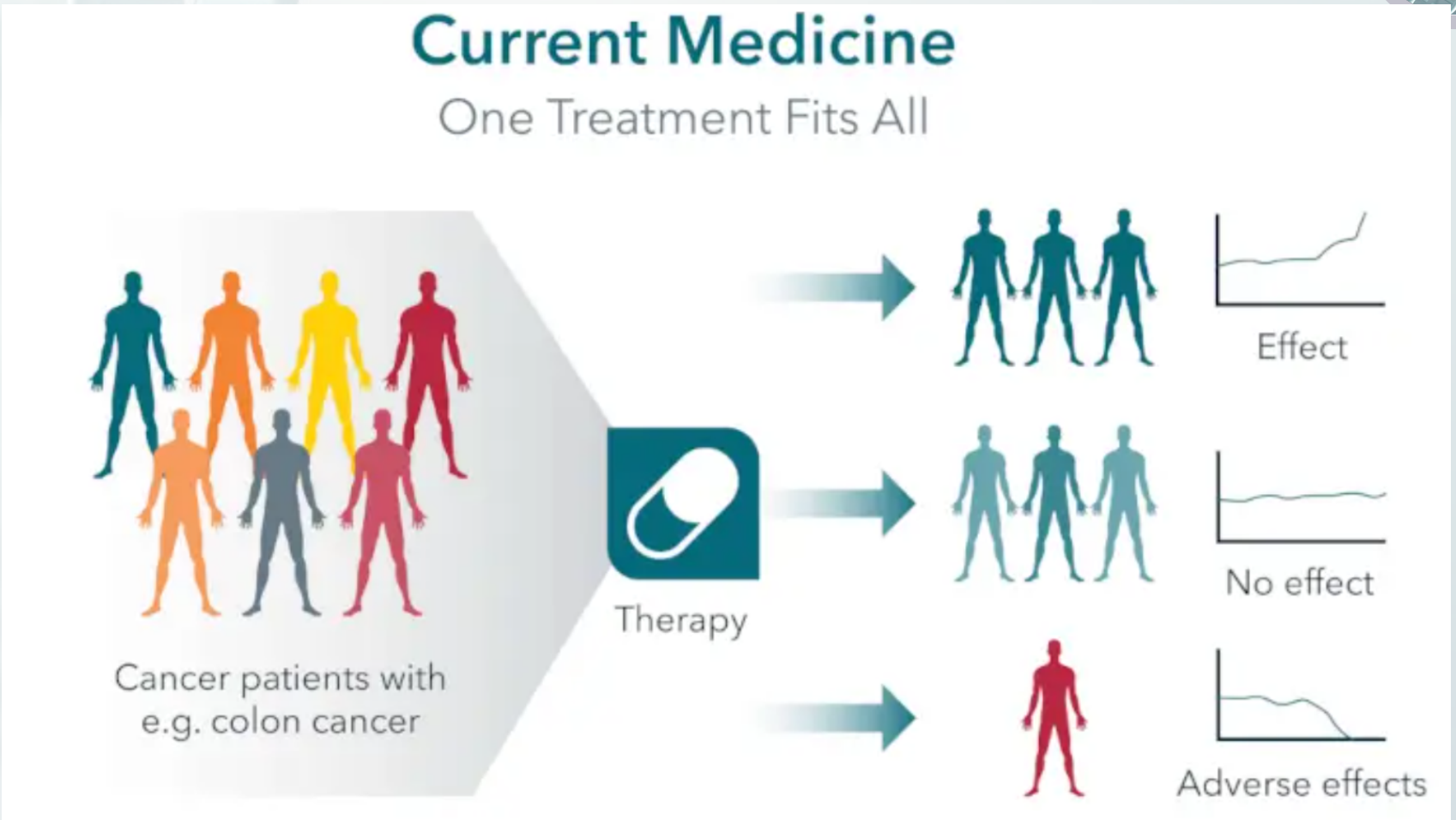


Figure: Pacientes con el mismo tipo de cáncer pueden reaccionar de forma disitinta a los mismos tratamientos. **Fuente** The Atlas Cancer [3].

Contexto y Motivación

Reacciones distintas para cada paciente

Figure: Cada paciente necesita un tratamiento personalizado. **Fuente** The Atlas Cancer [3].

Inmunoterapia del Cáncer

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

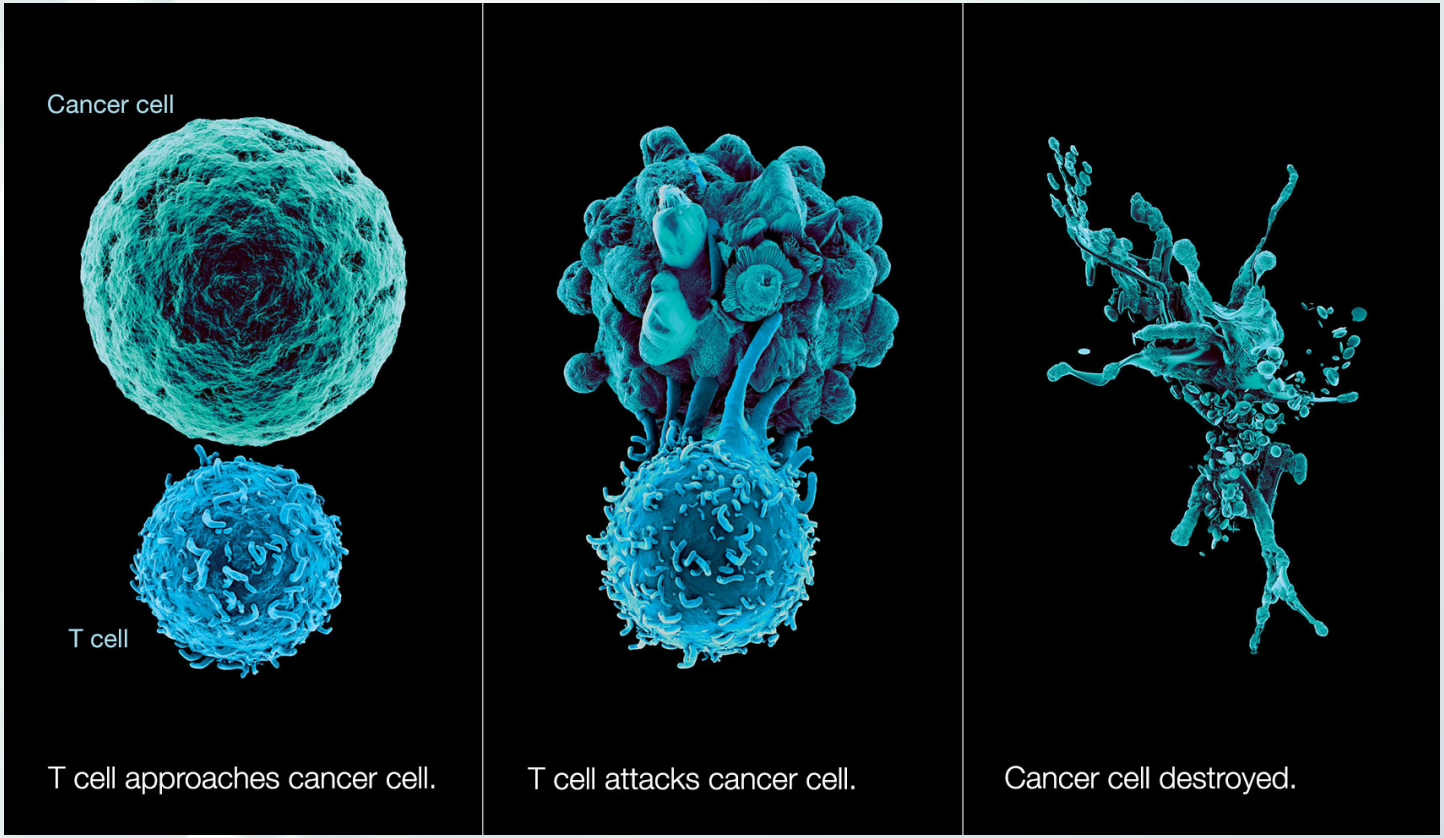


Figure: Ejemplo de como una célula T destruye células del cancer [5].

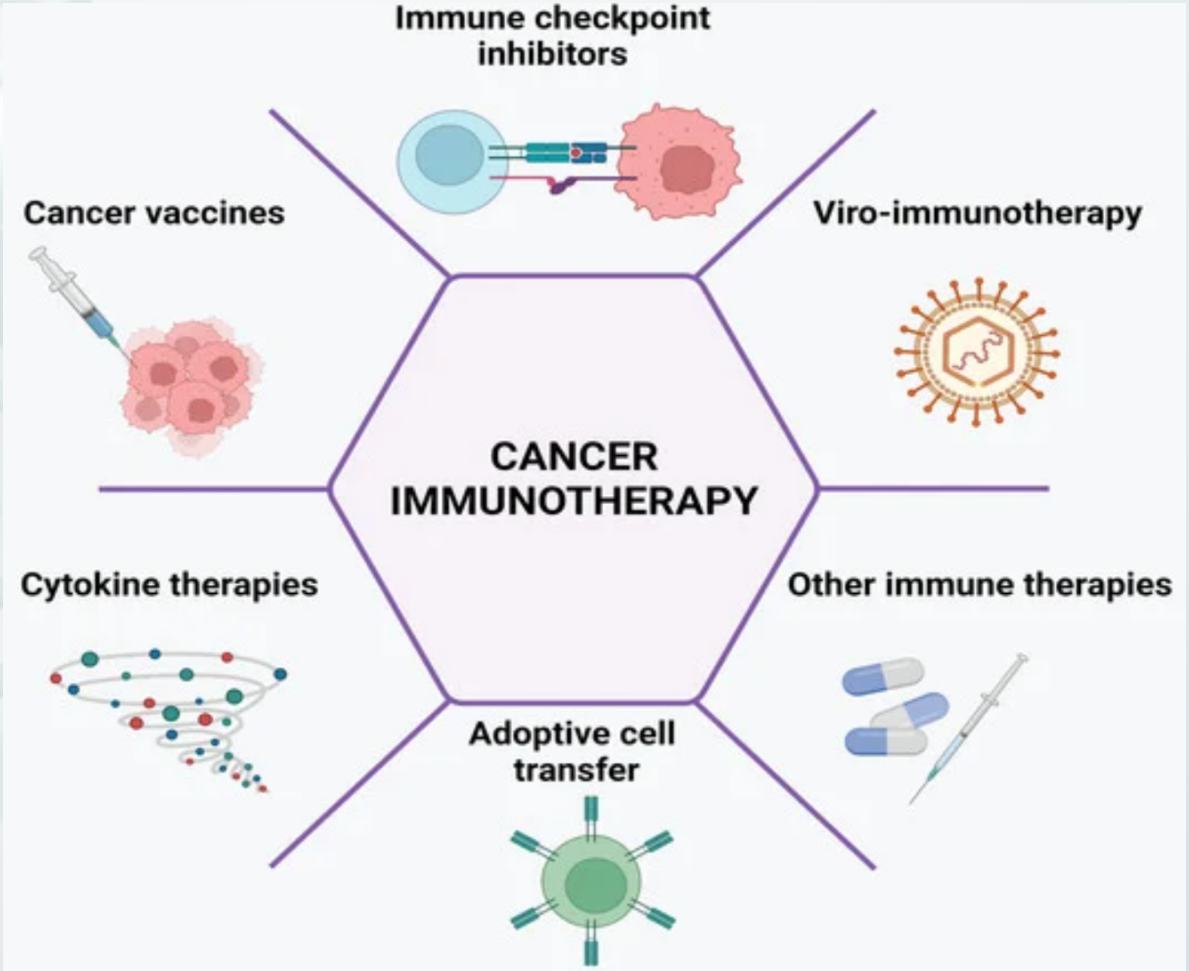


Figure: Tipos de tratamientos para la inmunoterapia del cáncer. **Fuente**: [6].

Contenido

Es una **proteína** que se forma en las células de Cáncer cuando ocurre mutaciones en el DNA [7, 8].

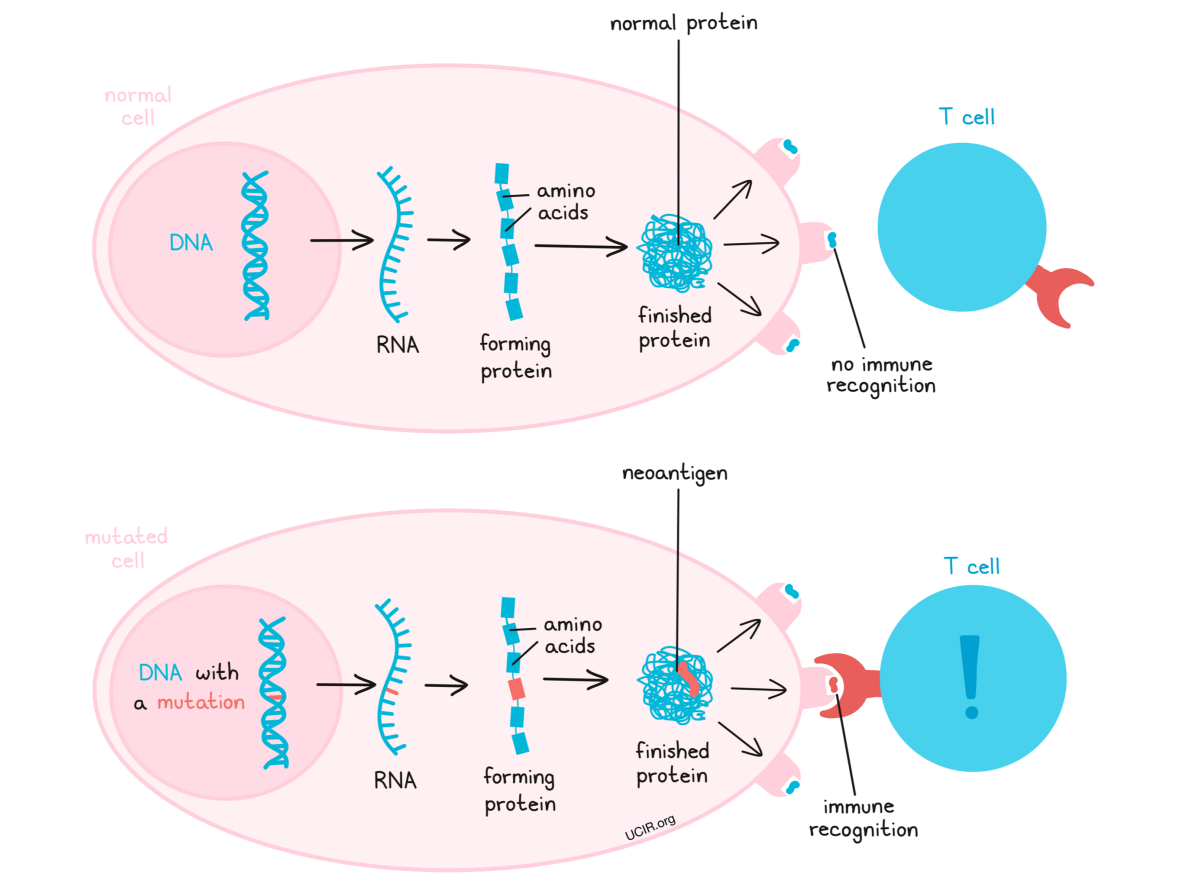


Figure: Neoantígenos y células T. **Fuente**: [9].

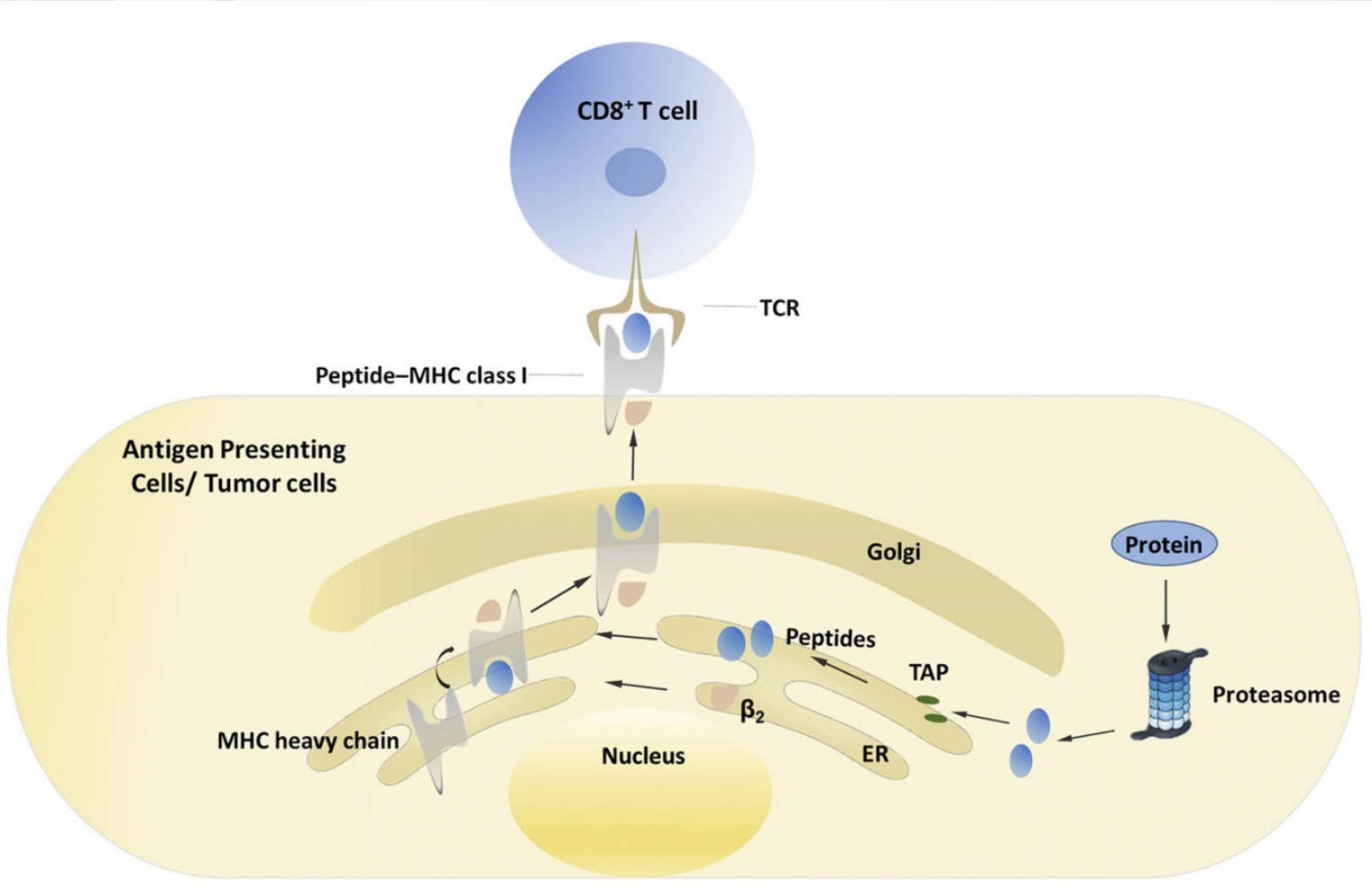


Figure: Presentación de antígenos por MHC-I. Fuente: [10]

Contexto y Motivación
Vacunas personalizadas

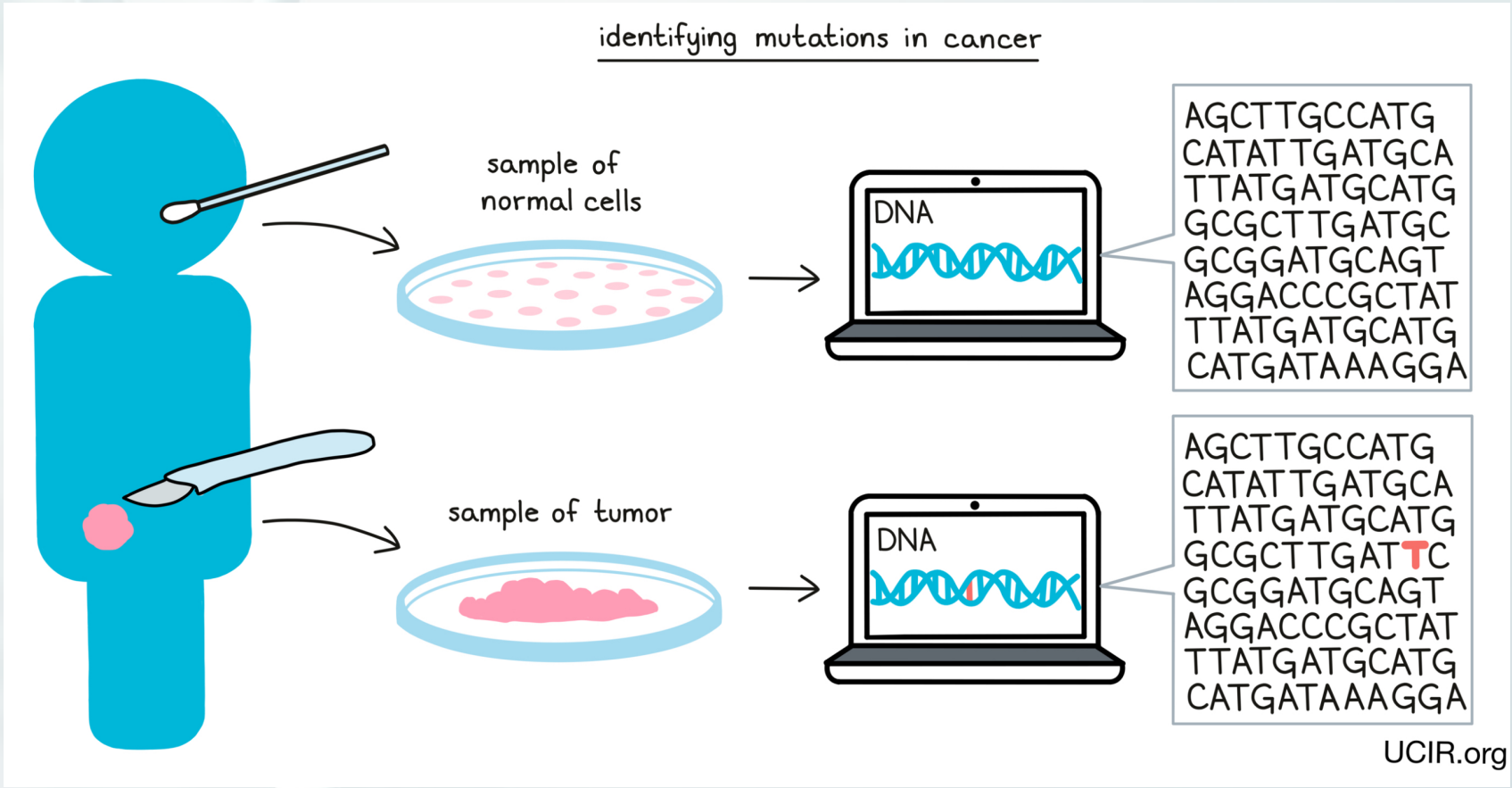


Figure: Proceso para la generación de vacunas contra el cáncer. **Fuente**: [9].

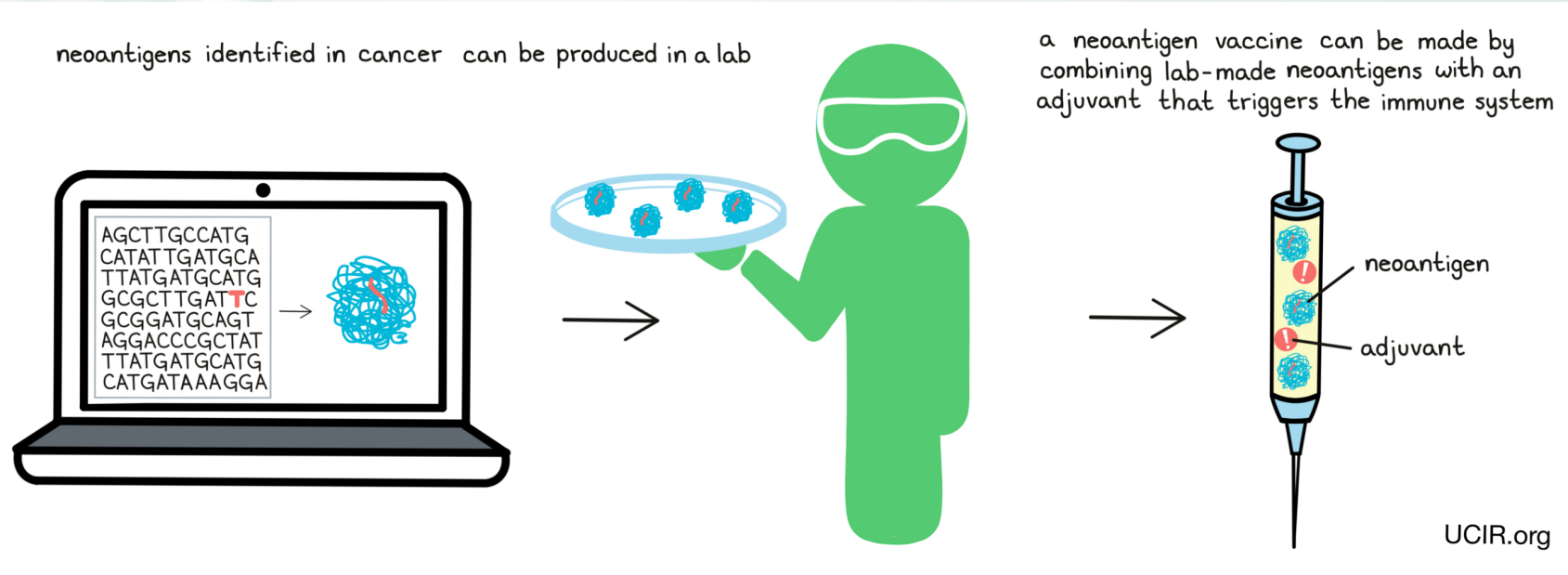


Figure: Proceso para la generación de vacunas contra el cáncer. **Fuente**: [9].

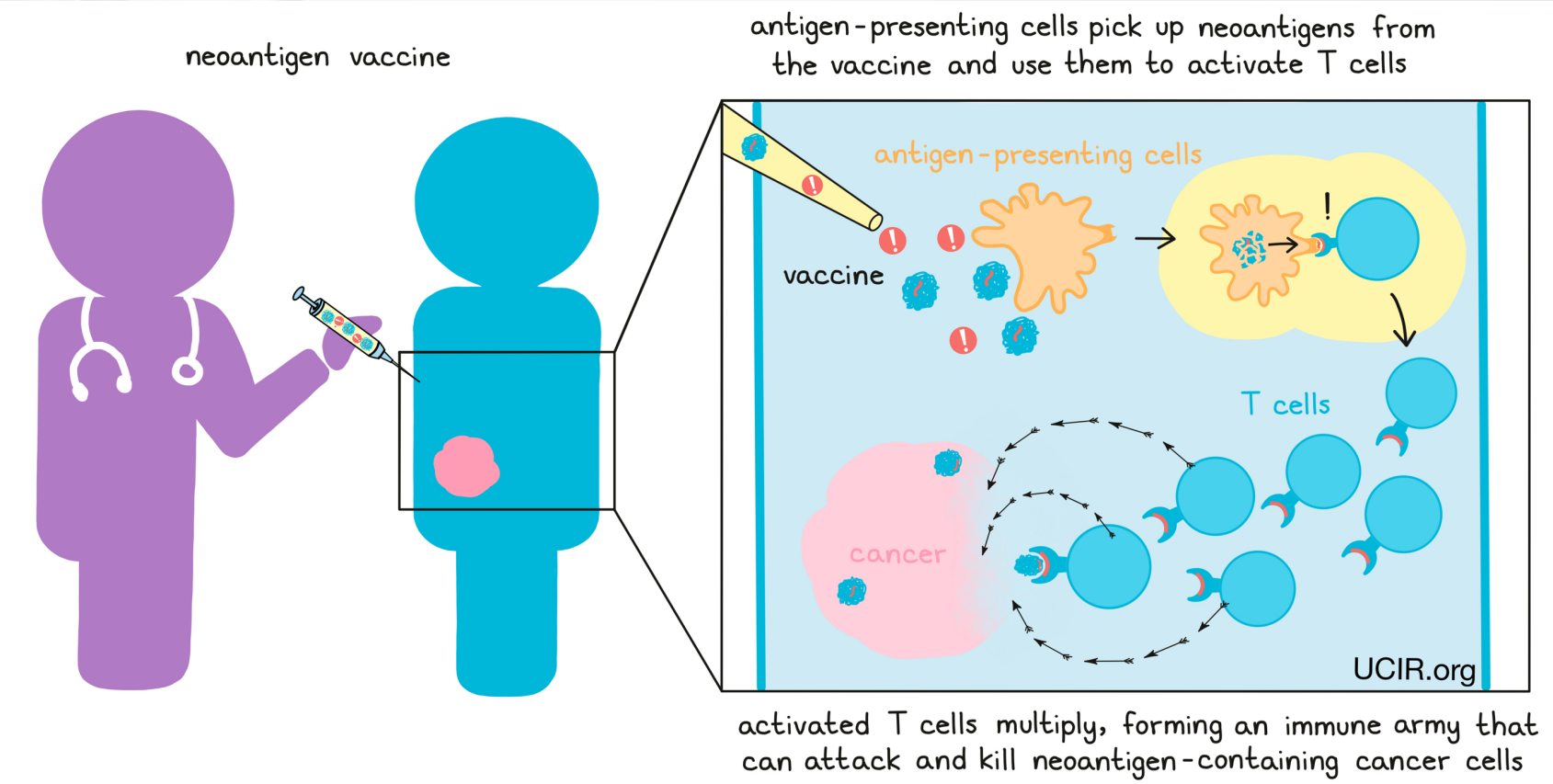


Figure: Proceso para la generación de vacunas contra el cáncer. **Fuente**: [9].

Contexto y Motivación
Vacunas personalizadas

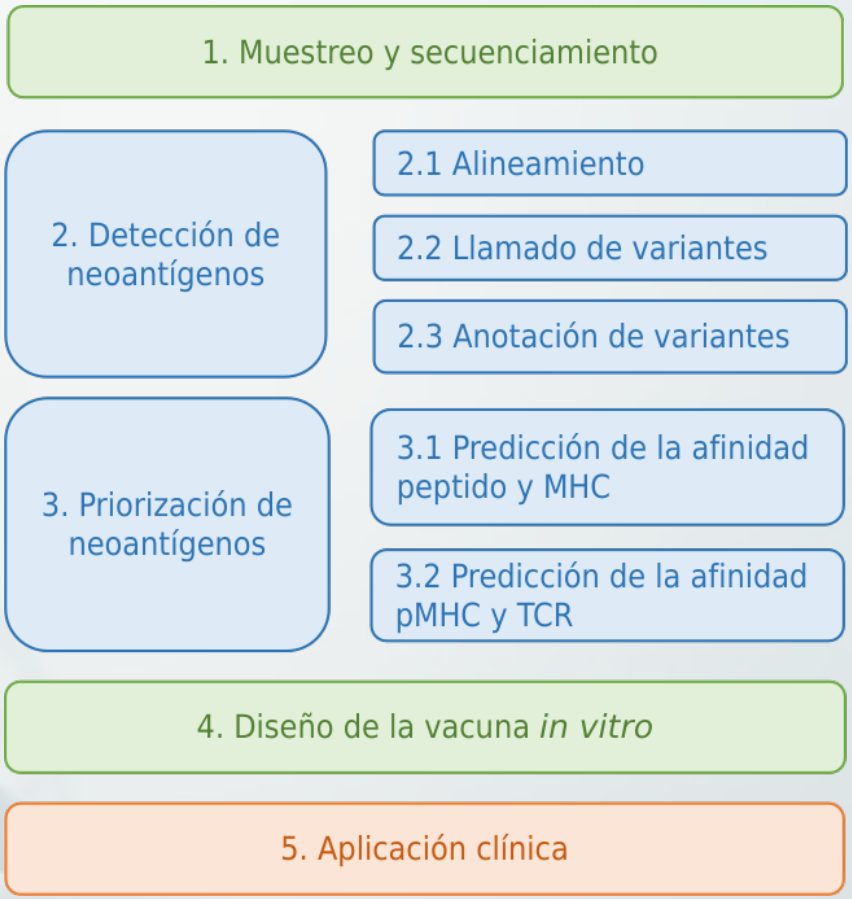


Figure: Resumen del proceso de generación de vacunas contra el cáncer.

Contenido

Problema

Less than 5% of detected neoantigens (peptides binded to MHC) succeed in activating the immune system [12].

This is a **binary classification problem**. A peptide could be represented like: $p = \{A, ..., Q\}$ and a MHC like: $q = \{A, N, ..., Q, E\}$. Finally, we need to know the probability of affinity between p and q (pMHC)

Problema

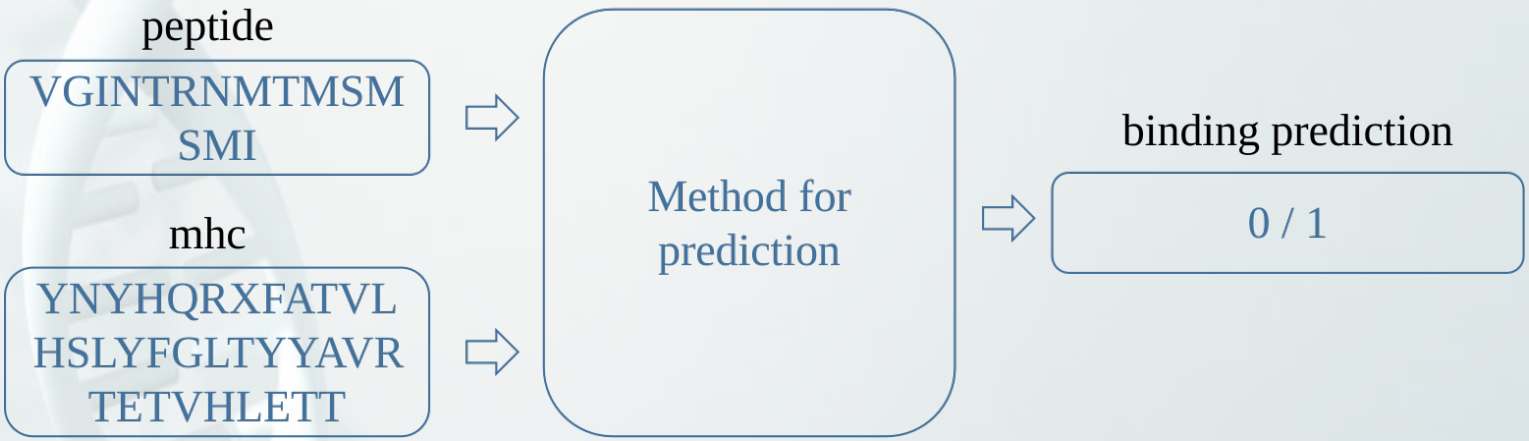


Figure: pMHC binding prediction problem.

Contenido

Proposal

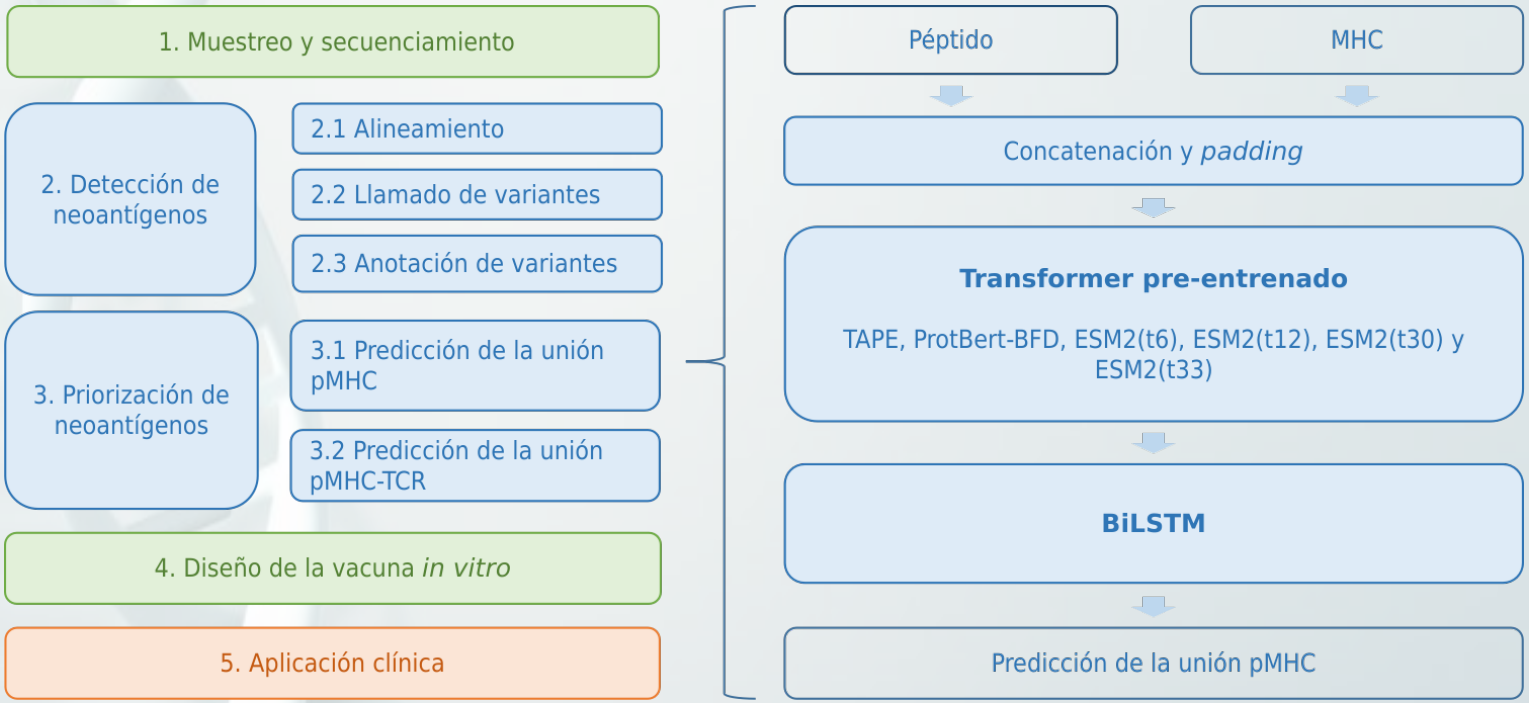


Figure: Proposal for pMHC binding prediction.

Contenido

Databases

Training: 539,019; Validation: 179,673; and Testing: 172,580.

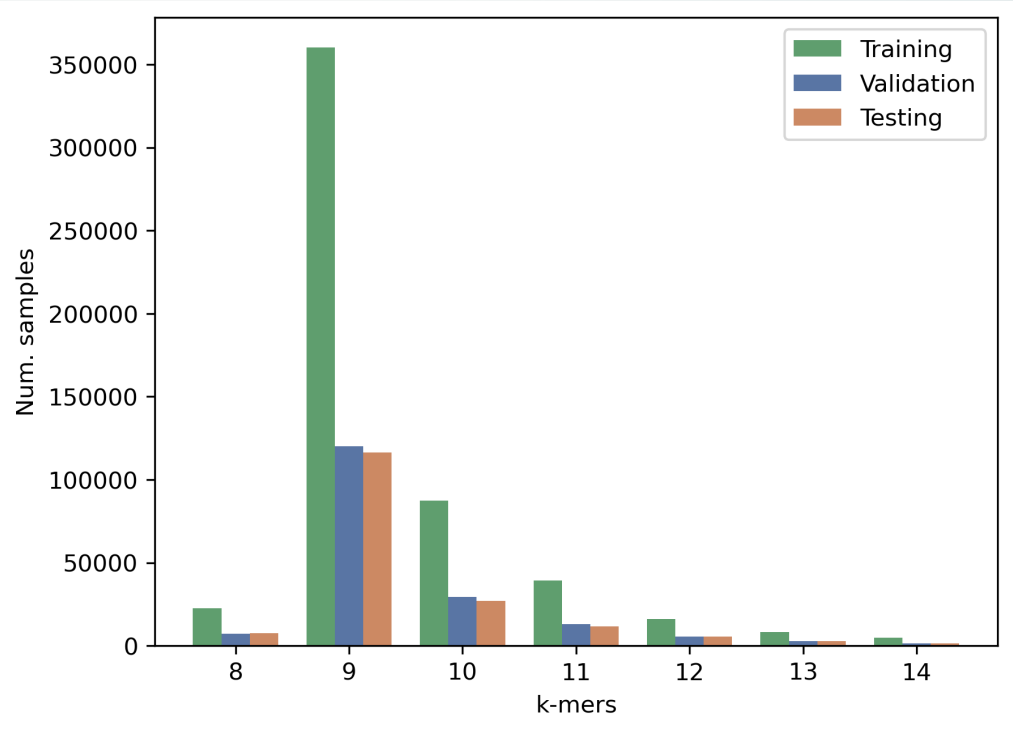


Figure: Number of samples per k-mer.

Contenido

Pre-trained models

Table: Differences between TAPE, ProtBert-DFB, and ESM2. HS: *Hidden size*; AH: *Attention heads*.

Model	BD	Samples	Layers	HS	AH	Params.
TAPE	Pfam	30M	12	768	12	92M
ProtBert-BFD	BFD	2122M	30	1024	16	420M
ESM2(t6)	Uniref50	60M	6	320	20	8M
ESM2(t12)	Uniref50	60M	12	480	20	35M
ESM2(t30)	Uniref50	60M	30	640	20	150M
ESM2(t33)	Uniref50	60M	33	1280	20	650M

Contenido

Results
(Training for 3 epochs)

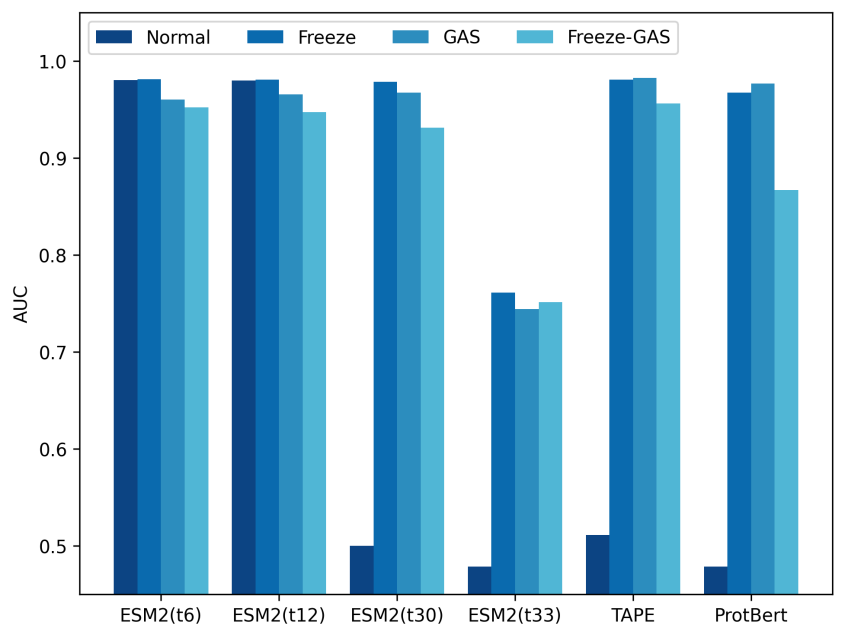
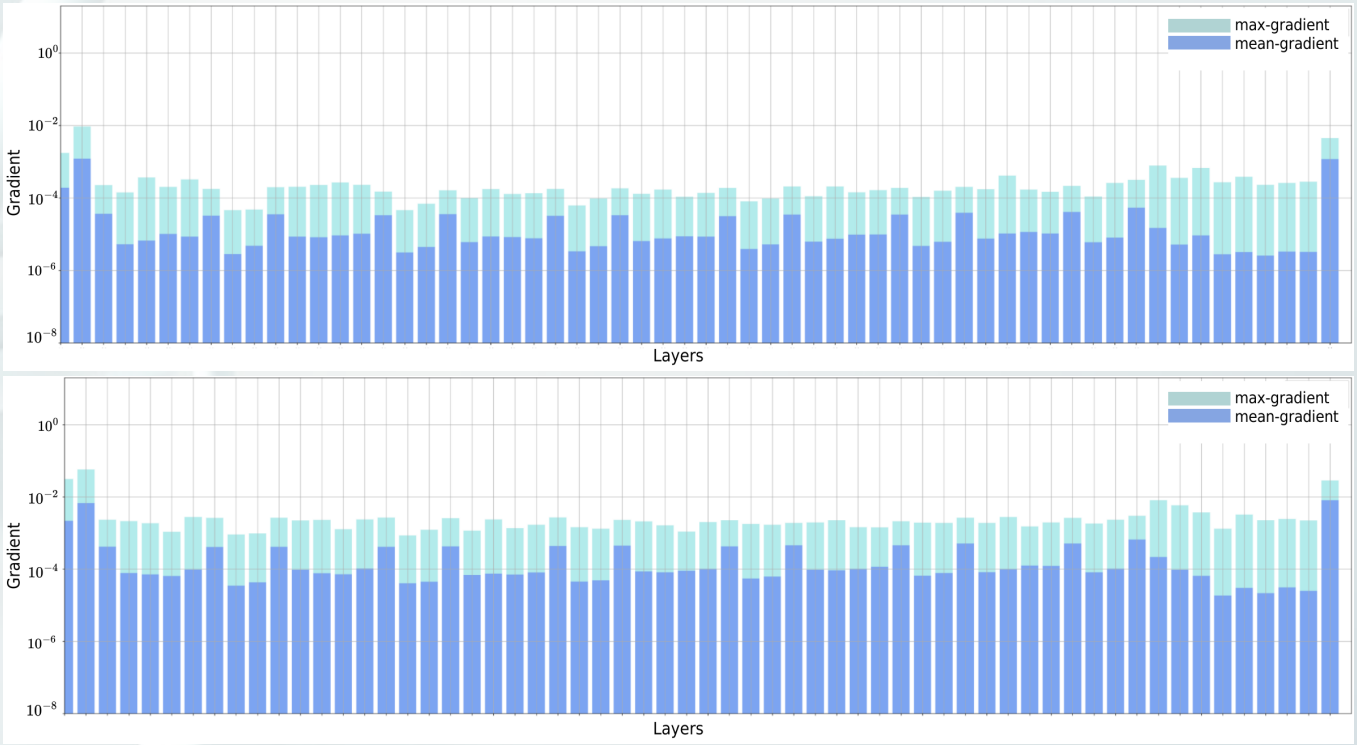


Figure: Comparative analysis of Area Under the Curve (AUC) in Transformer model architectures using various training methodologies.

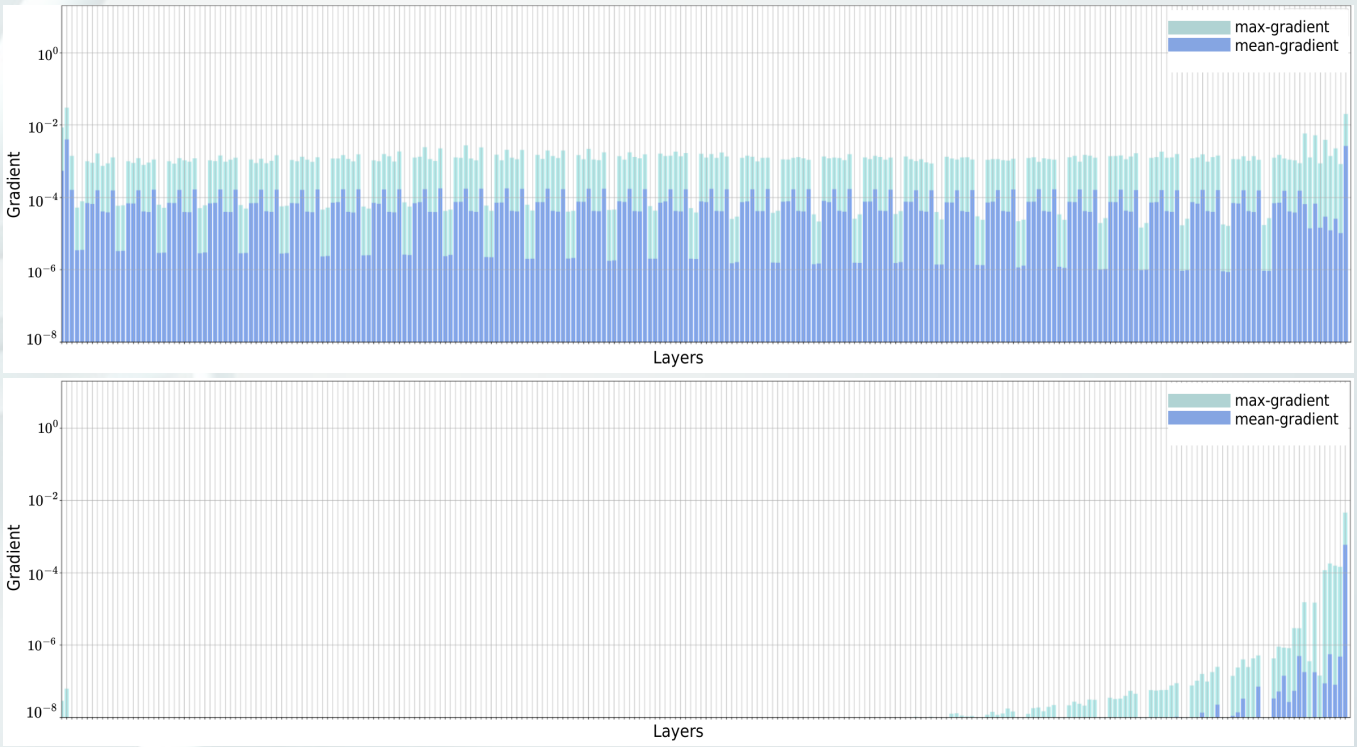
Results
Vanish gradient problem

Gradients for ESM2(t6)



Results
Vanish gradient problem

Gradients for ESM2(t30)



Results

Comparison (Training for 30 epochs)

	Accuracy	Precision	Recall	F1-score	AUC	MCC
ESM2(t6)-Normal	0.9390	0.9333	**0.9453**	0.9392	0.9797	0.8780
ESM2(t6)-Freeze	**0.9401**	**0.9398**	0.9402	**0.9400**	**0.9830**	**0.8802**
ESM2(t6)-GAS	0.9366	0.9322	0.9413	0.9368	0.9818	0.8732
ESM2(t6)-Freeze-GAS	0.9354	0.9326	0.9383	0.9355	0.9813	0.8708
ESM2(t30)-Normal	-	-	-	-	-	-
ESM2(t30)-Freeze	**0.9393**	0.9304	**0.9493**	**0.9397**	0.9787	**0.8787**
ESM2(t30)-GAS	0.9346	**0.9337**	0.9352	0.9345	0.9808	0.8691
ESM2(t30)-Freeze-GAS	0.9363	0.9319	0.9411	0.9365	**0.9818**	0.8726
TAPE-Normal	-	-	-	-	-	-
TAPE-Freeze	0.9395	**0.9404**	0.9382	0.9393	0.9815	0.8790
TAPE-GAS	**0.9415**	0.9352	**0.9484**	**0.9418**	**0.9841**	**0.8831**
TAPE-Freeze-GAS	0.9359	0.9297	0.9428	0.9362	0.9820	0.8719

Results
Comparison with state-of-art tools

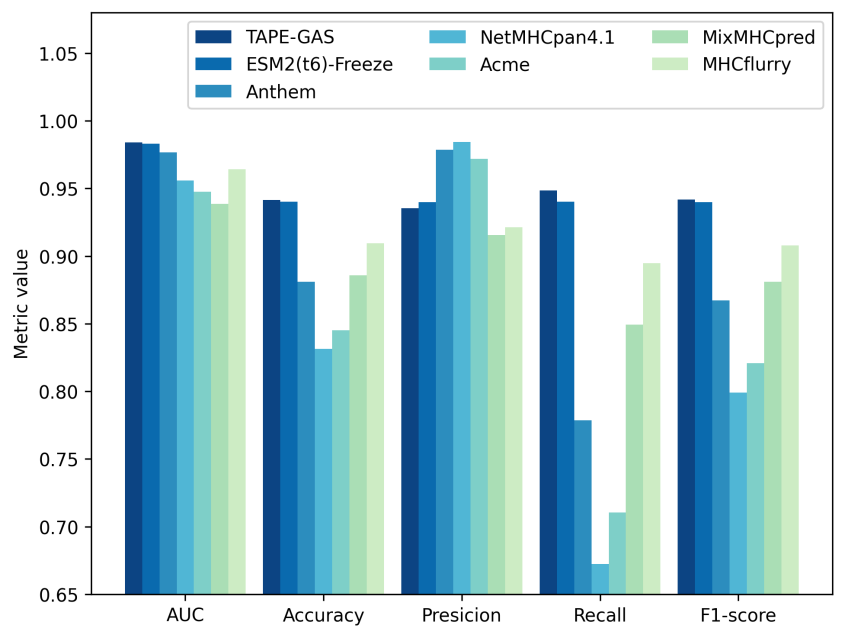


Figure: The AUC values for TAPE-GAS and ESM2(t6) trained for 30 epochs, in comparison to state-of-the-art methods.

Results
Comparison with state-of-art tools

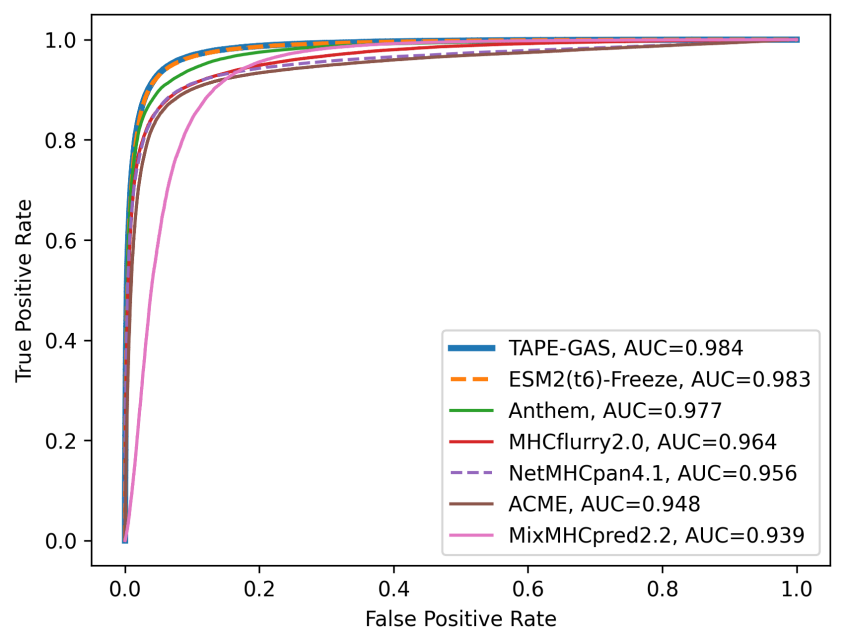


Figure: ROC curves for TAPE-GAS and ESM2(t6) trained for 30 epochs, in comparison to state-of-the-art methods.

Results
Comparison with state-of-art tools

Table: Performance evaluation of Transformer models TAPE-GAS and ESM2(t6)-Freeze, trained for 30 epochs, against Anthem, NetMHCpan4.1, ACME, MixMHCpred2.2, and MhcFlurry2.0.

	Accuracy	Precision	Recall	F1-score	AUC	MCC
TAPE-GAS	**0.9415**	0.9352	**0.9484**	**0.9418**	**0.9841**	**0.8831**
ESM2(t6)-Freeze	**0.9401**	0.9398	**0.9402**	**0.9400**	**0.9830**	**0.8802**
Anthem	0.8811	**0.9786**	0.7787	0.8673	0.9768	0.7785
NetMHCpan4.1	0.8312	**0.9844**	0.6724	0.7991	0.9557	0.6982
ACME	0.8452	0.9717	0.7105	0.8208	0.9476	0.7165
MixMHCpred2.2	0.8857	0.9155	0.8493	0.8811	0.9386	0.7733
MhcFlurry2.0	0.9093	0.9211	0.8948	0.9078	0.9642	0.8189

Contenido

Discussion

Fine-tuning ESM2 models

The most favorable results were obtained with the smallest model, **ESM2(t6)**. we believe is not sufficiently large for ESM2(t33), a model boasting 650 million parameters.

Another potential reason could be attributed to the use of **Rotary Position Embedding (RoPE)** used instead of absolute positional encoding.

Discussion

Layer Freezing and GAS

This approach involves locking the Transformer model while updating only the BiLSTM parameters. This method is generally well-suited to accelerate the training process, even though it may lead to a slight sacrifice in performance.

Surprisingly, **for ESM2 models, this methodology yielded the best results, while for TAPE and ProtBert-BFD, it yielded the expected outcomes**.

Discussion

TAPE, ProtBert-BFD and ESM2

ProtBert-BFD got the worst result despite this model were pre-trained with the largest dataset BFD with 2122M samples. We believe, this result is caused by the noisy information and sequence mistakes in BFD dataset.

TAPE achieved the best results. TAPE models were pre-trained using the Pfam dataset, it is derived from UniProtKB and **selectively includes sequences belonging to Reference Proteomes rather than the entire UniProtKB**

ESM2(t6) achieved results that closely rival TAPE. ESM2(t6) comprises only 8 million parameters, compared to 92 million parameters of TAPE.

References I

[1] Rebecca L Siegel, Kimberly D Miller, Nikita Sandeep Wagle, and Ahmedin Jemal,
"Cancer statistics, 2023,"
Ca Cancer J Clin, vol. 73, no. 1, pp. 17–48, 2023.

[2] Cancer Atlas,
"Cancer atlas - the burden," 2023.

[3] Personalized Medicine,
"Pdx and personalized medicine," 2023.

[4] Cancer.net,
"Qué es la inmunoterapia," 2022.

[5] NortShore,
"Immunotherapy," 2022.

References II

[6] Mateusz Kciuk, Esam Bashir Yahya, Montaha Mohamed Ibrahim Mohamed, Summya Rashid, Muhammad Omer Iqbal, Renata Kontek, Muhanad A Abdulsamad, and Abdulmutalib A Allaq,
"Recent advances in molecular mechanisms of cancer immunotherapy,"
Cancers, vol. 15, no. 10, pp. 2721, 2023.

[7] NCI,
"National cancer institute dictionary," 2022.

[8] Elizabeth S Borden, Kenneth H Buetow, Melissa A Wilson, and Karen Taraszka Hastings,
"Cancer neoantigens: Challenges and future directions for prediction, prioritization, and validation,"
Frontiers in Oncology, vol. 12, 2022.

References III

[9] UCIR,
"Neoantige-based therapy," 2023.

[10] Xiaomei Zhang, Yue Qi, Qi Zhang, and Wei Liu,
"Application of mass spectrometry-based mhc immunopeptidome profiling in neoantigen identification for tumor immunotherapy,"
Biomedicine & Pharmacotherapy, vol. 120, pp. 109542, 2019.

[11] Miao Peng, Yongzhen Mo, Yian Wang, Pan Wu, Yijie Zhang, Fang Xiong, Can Guo, Xu Wu, Yong Li, Xiaoling Li, et al.,
"Neoantigen vaccine: an emerging tumor immunotherapy,"
Molecular cancer, vol. 18, no. 1, pp. 1–14, 2019.

References IV

[12] L Mattos, M Vazquez, F Finotello, R Lepore, E Porta, J Hundal, P Amengual-Rigo, CKY 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.

[13] Nil Adell Mill, Cedric Bogaert, Wim van Criekinge, and Bruno Fant, "neoms: Attention-based prediction of mhc-i epitope presentation," *bioRxiv*, 2022.

[14] Ronghui You, Wei Qu, Hiroshi Mamitsuka, and Shanfeng Zhu, "Deepmhcii: a novel binding core-aware deep interaction model for accurate mhc-ii peptide binding affinity prediction," *Bioinformatics*, vol. 38, no. Supplement_1, pp. i220–i228, 2022.

References V

[15] Guangyuan Li, Balaji Iyer, VB Surya Prasath, Yizhao Ni, and Nathan Salomonis,
"Deepimmuno: deep learning-empowered prediction and generation of immunogenic peptides for t-cell immunity,"
Briefings in bioinformatics, vol. 22, no. 6, pp. bbab160, 2021.

[16] Franziska Lang, Pablo Riesgo-Ferreiro, Martin L¨ower, Ugur Sahin, and Barbara Schr¨ors,
"Neofox: annotating neoantigen candidates with neoantigen features,"
Bioinformatics, vol. 37, no. 22, pp. 4246–4247, 2021.

[17] Ko-Han Lee, Yu-Chuan Chang, Ting-Fu Chen, Hsueh-Fen Juan, Huai-Kuang Tsai, and Chien-Yu Chen,
"Connecting mhc-i-binding motifs with hla alleles via deep learning,"
Communications Biology, vol. 4, no. 1, pp. 1–12, 2021.

References VI

[18] Valentin Junet and Xavier Daura,
"Cnn-peppred: an open-source tool to create convolutional nn models for the discovery of patterns in peptide sets—application to peptide–mhc class ii binding prediction,"
Bioinformatics, vol. 37, no. 23, pp. 4567–4568, 2021.

[19] Baikang Pei and Yi-Hsiang Hsu,
"Iconmhc: a deep learning convolutional neural network model to predict peptide and mhc-i binding affinity,"
Immunogenetics, vol. 72, no. 5, pp. 295–304, 2020.

[20] Shikhar Saxena, Sambhavi Animesh, Melissa J Fullwood, and Yuguang Mu,
"Onionmhc: A deep learning model for peptide—hla-a* 02: 01 binding predictions using both structure and sequence feature sets,"

References VII

Journal of Micromechanics and Molecular Physics, vol. 5, no. 03, pp. 2050009, 2020.

[21] Felicia SL Ng, Michel Vandenberghe, Guillem Portella, Corinne Cayatte, Xiaotao Qu, Shino Hanabuchi, Aimee Landry, Raghothama Chaerkady, Wen Yu, Rosana Collepardo-Guevara, et al.,
"Minerva: Learning the rules of hla class i peptide presentation in tumors with convolutional neural networks and transfer learning,"
Available at SSRN 3704016, 2020.

[22] Tianyi Zhao, Liang Cheng, Tianyi Zang, and Yang Hu,
"Peptide-major histocompatibility complex class i binding prediction based on deep learning with novel feature,"
Frontiers in Genetics, vol. 10, pp. 1191, 2019.

References VIII

[23] Zhonghao Liu, Yuxin Cui, Zheng Xiong, Alierza Nasiri, Ansi Zhang, and Jianjun Hu,
"Deepseqpan, a novel deep convolutional neural network model for pan-specific class i hla-peptide binding affinity prediction,"
Scientific reports, vol. 9, no. 1, pp. 1–10, 2019.

[24] Youngmahn Han,
"Deep convolutional neural networks for peptide-mhc binding predictions," 2018.

[25] Xiaoyun Yang, Liyuan Zhao, Fang Wei, and Jing Li,
"Deepnetbim: deep learning model for predicting hla-epitope interactions based on network analysis by harnessing binding and immunogenicity information,"
BMC bioinformatics, vol. 22, no. 1, pp. 1–16, 2021.

References IX

[26] Jing Jin, Zhonghao Liu, Alireza Nasiri, Yuxin Cui, Stephen-Yves Louis, Ansi Zhang, Yong Zhao, and Jianjun Hu,
"Deep learning pan-specific model for interpretable mhc-i peptide binding prediction with improved attention mechanism,"
Proteins: Structure, Function, and Bioinformatics, vol. 89, no. 7, pp. 866–883, 2021.

[27] Yan Hu, Ziqiang Wang, Hailin Hu, Fangping Wan, Lin Chen, Yuanpeng Xiong, Xiaoxia Wang, Dan Zhao, Weiren Huang, and Jianyang Zeng,
"Acme: pan-specific peptide–mhc class i binding prediction through attention-based deep neural networks,"
Bioinformatics, vol. 35, no. 23, pp. 4946–4954, 2019.

References X

[28] Xuezhi Xie, Yuanyuan Han, and Kaizhong Zhang,
"Mhcherrypan: a novel pan-specific model for binding affinity prediction of class i hla-peptide,"
International Journal of Data Mining and Bioinformatics, vol. 24, no. 3, pp. 201–219, 2020.

[29] Yilin Ye, Jian Wang, Yunwan Xu, Yi Wang, Youdong Pan, Qi Song, Xing Liu, and Ji Wan,
"Mathla: a robust framework for hla-peptide binding prediction integrating bidirectional lstm and multiple head attention mechanism,"
BMC bioinformatics, vol. 22, no. 1, pp. 1–12, 2021.

References XI

[30] Zhonghao Liu, Jing Jin, Yuxin Cui, Zheng Xiong, Alireza Nasiri, Yong Zhao, and Jianjun Hu,
"Deepseqpanii: an interpretable recurrent neural network model with attention mechanism for peptide-hla class ii binding prediction,"
IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2021.

[31] Yu Heng, Zuyin Kuang, Wenzhao Xie, Haoqi Lan, Shuheng Huang, Linxin Chen, Tingting Shi, Lei Xu, Xianchao Pan, and Hu Mei,
"A simple pan-specific rnn model for predicting hla-ii binding peptides,"
Molecular Immunology, vol. 139, pp. 177–183, 2021.

References XII

[32] Limin Jiang, Hui Yu, Jiawei Li, Jijun Tang, Yan Guo, and Fei Guo, "Predicting mhc class i binder: existing approaches and a novel recurrent neural network solution," *Briefings in Bioinformatics*, vol. 22, no. 6, pp. bbab216, 2021.

[33] Xiaoshan M Shao, Rohit Bhattacharya, Justin Huang, IK Sivakumar, Collin Tokheim, Lily Zheng, Dylan Hirsch, Benjamin Kaminow, Ashton Omdahl, Maria 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 XIII

[34] Jingcheng Wu, Wenzhe Wang, Jiucheng Zhang, Binbin Zhou, Wenyi Zhao, Zhixi Su, Xun Gu, Jian Wu, Zhan Zhou, and Shuqing Chen,
"Deephlapan: a deep learning approach for neoantigen prediction considering both hla-peptide binding and immunogenicity,"
Frontiers in Immunology, p. 2559, 2019.

[35] Fuxu Wang, Haoyan Wang, Lizhuang Wang, Haoyu Lu, Shizheng Qiu, Tianyi Zang, Xinjun Zhang, and Yang Hu,
"Mhcroberta: pan-specific peptide–mhc class i binding prediction through transfer learning with label-agnostic protein sequences,"
Briefings in Bioinformatics, vol. 23, no. 3, pp. bbab595, 2022.

References XIV

[36] Yanyi Chu, Yan Zhang, Qiankun Wang, Lingfeng Zhang, Xuhong Wang, Yanjing Wang, Dennis Russell Salahub, Qin Xu, Jianmin Wang, Xue 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.

[37] Jun Cheng, Kaïdre Bendjama, Karola Rittner, and Brandon 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 XV

[38] Hans-Christof Gasser, Georges Bedran, Bo Ren, David Goodlett, Javier Alfaro, and Ajitha Rajan,
"Interpreting bert architecture predictions for peptide presentation by mhc class i proteins,"
arXiv preprint arXiv:2111.07137, 2021.

[39] Mikhail Shugay, Dmitriy V Bagaev, Ivan V Zvyagin, Renske M Vroomans, Jeremy Chase Crawford, Garry Dolton, Ekaterina A Komech, Anastasiya L Sycheva, Anna E Koneva, Evgeniy 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.

References XVI

[40] Randi Vita, Swapnil Mahajan, James A Overton, Sandeep Kumar Dhanda, Sheridan Martini, Jason R Cantrell, Daniel K Wheeler, Alessandro Sette, and Bjoern Peters, "The immune epitope database (iedb): 2018 update," *Nucleic acids research*, vol. 47, no. D1, pp. D339–D343, 2018.

[41] Jingcheng Wu, Wenyi Zhao, Binbin Zhou, Zhixi Su, Xun Gu, Zhan Zhou, and Shuqing Chen, "Tsnadb: a database for tumor-specific neoantigens from immunogenomics data analysis," *Genomics, proteomics & bioinformatics*, vol. 16, no. 4, pp. 276–282, 2018.

References XVII

[42] Wei-Jun Zhou, Zhi Qu, Chao-Yang Song, Yang Sun, An-Li Lai, Ma-Yao Luo, Yu-Zhe Ying, Hu Meng, Zhao Liang, Yan-Jie He, et al.,
"Neopeptide: an immunoinformatic database of t-cell-defined neoantigens,"
Database, vol. 2019, 2019.

[43] Deylane Menezes Teles Oliveira, Rafael Melo Santos de Serpa Brandão, Luiz Claudio Demes da Mata Sousa, Francisco das Chagas Alves Lima, Semiramis Jamil Hadad do Monte, Mário Sérgio Coelho Marroquim, Antonio Vanildo de Sousa Lima, Antonio Gilberto Borges Coelho, Jhonatan Matheus Sousa Costa, Ricardo Martins 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.

References XVIII

[44] Xiaoxiu Tan, Daixi Li, Pengjie Huang, Xingxing Jian, Huihui Wan, Guangzhi Wang, Yuyu Li, Jian Ouyang, Yong Lin, and Lu Xie, "dbpepneo: a manually curated database for human tumor neoantigen peptides," *Database*, vol. 2020, 2020.

[45] Manman Lu, Linfeng Xu, Xingxing Jian, Xiaoxiu Tan, Jingjing Zhao, Zhenhao Liu, Yu Zhang, Chunyu Liu, Lanming Chen, Yong Lin, et al., "dbpepneo2. 0: A database for human tumor neoantigen peptides from mass spectrometry and tcr recognition," *Frontiers in immunology*, p. 1583, 2022.

References XIX

[46] Le Zhang, Geng Liu, Guixue Hou, Haitao Xiang, Xi Zhang, Ying Huang, Xiuqing Zhang, Bo Li, and Leo 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, pp. 579, 2022.

[47] James Robinson, Dominic J Barker, Xenia Georgiou, Michael A Cooper, Paul Flicek, and Steven GE Marsh,
"Ipd-imgt/hla database,"
Nucleic acids research, vol. 48, no. D1, pp. D948–D955, 2020.

References XX

[48] Sora Kim, Han Sang Kim, Eunyoung Kim, MG 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.

[49] Alex Rubinsteyn, Julia Kodysh, Isaac Hodes, Sebastien Mondet, Bulent Arman Aksoy, John P Finnigan, Nina Bhardwaj, and Jeffrey Hammerbacher,
"Computational pipeline for the pgv-001 neoantigen vaccine trial,"
Frontiers in immunology, vol. 8, pp. 1807, 2018.

References XXI

[50] Ting-You Wang, Li Wang, Sk Kayum Alam, Luke H Hoeppner, and Rendong Yang,
"Scanneo: identifying indel-derived neoantigens using rna-seq data,"
Bioinformatics, vol. 35, no. 20, pp. 4159–4161, 2019.

[51] Ryan O Schenck, Eszter Lakatos, Chandler Gatenbee, Trevor A Graham, and Alexander RA @miscNCIdictionary2022, author = NCI, title = National Cancer Institute Dictionary, year = 2022, url = https://www.cancer.gov/publications/dictionaries/genetics-dictionary, urldate = 2022-03-20 Anderson,
"Neopredpipe: high-throughput neoantigen prediction and recognition potential pipeline,"
BMC bioinformatics, vol. 20, no. 1, pp. 1–6, 2019.

References XXII

[52] Jasreet Hundal, Susanna Kiwala, Joshua McMichael, Christopher A Miller, Huiming Xia, Alexander T Wollam, Connor J Liu, Sidi Zhao, Yang-Yang Feng, Aaron 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.

[53] Yuyu Li, Guangzhi Wang, Xiaoxiu Tan, Jian Ouyang, Menghuan Zhang, Xiaofeng Song, Qi Liu, Qibin Leng, Lanming Chen, and Lu Xie,
"Progeo-neo: a customized proteogenomic workflow for neoantigen prediction and selection,"
BMC medical genomics, vol. 13, no. 5, pp. 1–11, 2020.

References XXIII

[54] Mary A Wood, Austin Nguyen, Adam J Struck, Kyle Ellrott, Abhinav Nellore, and Reid F Thompson,
"Neoepiscope improves neoepitope prediction with multivariant phasing,"
Bioinformatics, vol. 36, no. 3, pp. 713–720, 2020.

[55] Ana Carolina MF Coelho, André L Fonseca, Danilo L Martins, Paulo BR Lins, Lucas M da Cunha, and Sandro J de Souza,
"neoant-hill: an integrated tool for identification of potential neoantigens,"
BMC Medical Genomics, vol. 13, no. 1, pp. 1–8, 2020.

References XXIV

[56] Carlos Wert-Carvajal, Rubén Sánchez-García, José R Macías, Rebeca Sanz-Pamplona, Almudena Méndez Pérez, Ramon Alemany, Esteban Veiga, Carlos Óscar S Sorzano, and Arrate 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.

[57] Laura Y Zhou, Fei Zou, and Wei Sun,
"Prioritizing candidate peptides for cancer vaccines by pepprmint: a statistical model to predict peptide presentation by hla-i proteins,"
bioRxiv, 2021.

References XXV

[58] Yuri Laguna Terai, Chun Huang, Baoli Wang, Xiaonan Kang, Jing Han, Jacqueline Douglass, Emily Han-Chung Hsiue, Ming Zhang, Raj Purohit, Taylor deSilva, et al.,
"Valid-neo: A multi-omics platform for neoantigen detection and quantification from limited clinical samples,"
Cancers, vol. 14, no. 5, pp. 1243, 2022.

[59] Qing Hao, Ping Wei, Yang Shu, Yi-Guan Zhang, Heng Xu, and Jun-Ning Zhao,
"Improvement of neoantigen identification through convolution neural network,"
Frontiers in immunology, vol. 12, 2021.

References XXVI

[60] Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C Lawrence Zitnick, Jerry Ma, et al.,
"Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences,"
Proceedings of the National Academy of Sciences, vol. 118, no. 15, 2021.

[61] Alexander V Gopanenko, Ekaterina N Kosobokova, and Vyacheslav S Kosorukov,
"Main strategies for the identification of neoantigens,"
Cancers, vol. 12, no. 10, pp. 2879, 2020.

References XXVII

[62] Nasser Hashemi, Boran Hao, Mikhail Ignatov, Ioannis Paschalidis, Pirooz Vakili, Sandor Vajda, and Dima Kozakov, "Improved predictions of mhc-peptide binding using protein language models," *bioRxiv*, 2022.