Revista Colombiana de
Hematología y Oncología
Julio - Diciembre de 2023 • Volumen 10 - Número 2

EDITORIAL

Machine learning and Deep learning applications in CAR-T research and development.

Aplicación de *Machine Learning* y *Deep Learning* en investigación y desarrollo de terapias CAR-T

» **Bonell Patiño-Escobar, MD.**[1,2]

[1] UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, CA, USA
[2]Department of Laboratory Medicine, University of California, San Francisco, CA, USA

Doi: https://doi.org/10.51643/22562915.671

Chimeric antigen receptors (CARs) stand out as prominent synthetic biology tools designed to enhance the immune response against tumors. CARs have demonstrated impressive efficacy, inducing remission or even providing a cure for individuals with hematological neoplasms. However, research regarding their effectiveness in treating solid tumors is somewhat limited. These limitations reveal ongoing challenges in CARs development that are currently being addressed. To optimize the translational potential of these cellular therapies, the approach should be firmly grounded in fundamental immunology. Thus, needing improved predictive pre-clinical models (particularly in vivo) that closely resemble human diseases. This approach aims to enhance efficacy while maintaining a critical focus on safety considerations and the ability to overcome various inhibitory mechanisms[1].

When it comes to predictive models, each in vivo model has their pros and cons since those should incorporate humanized components and undergo validation confirming that there is faithful replication of human diseases, facilitating the rapid translation of findings from preclinical to clinical testing[1]. Moreover, there are other data-driven predictive models that propose in silico advancements. Although these models require experimental validation in the laboratory, they play a pivotal role in bridging the gap between dry and wet lab research, expediting discoveries that would otherwise take years and resulting in substantial time and cost savings across various screening processes.

Machine learning and deep learning (ML/DL) constitute a domain that combines automation, experiential learning, and essential statistical and computational principles[2,3]. Over the past two decades, ML/DL has found widespread applications across diverse fields, achieving remarkable success. In the realm of artificial intelligence (AI), ML/DL has focused on developing practical software, with diverse applications such as computer vision, speech recognition, natural language processing, and robot control, among others. Similarly, the impact of machine learning extends to fields like biology and social

science, where it enables the novel analysis of high-throughput experimental data, bringing about positive advancements. Notable recent achievements in deep learning encompass various milestones, such as playing Go2, natural language processing3 (including the famous GPT-34), detecting and identifying objects in images5, and predicting protein structures[6].

Diverse ML/DL approaches have enhanced cancer biology research, particularly in identifying target candidates for immunotherapy predictions. Choudhry et al, demonstrated that an epigenetic analysis, relying on publicly available ATAC-seq and bulk RNAseq datasets, employed a machine learning approach based on random forest and XGBoost ensemble techniques to successfully predict CD38 transcriptional regulation in myeloma cells from patient samples to sensitize myeloma cells to current approved immunotherapies such as Daratumumab or Isatuximab, both monoclonal antibodies directed to CD38 antigen in the tumor cell surface[7].

In a more recent study conducted by Hie et al., a general protein language model showcased the remarkable capacity of deep learning models to efficiently guide the evolution of human antibodies. This was achieved through the suggestion of evolutionarily plausible mutations, an analysis encompassing a staggering 98 million protein sequences, of which only a few thousand were antibody-related. Intriguingly, this feat was accomplished without providing the model with any prior knowledge about the target antigen, binding specificity, or protein structure. To test their hypothesis, the research team conducted evolutionary campaigns guided by the likelihood predictions from the language model, to enhance the affinity of seven antibodies. These antibodies represented a diverse range of antigens and degrees of maturity, including but not limited to influenza A, ebolavirus, SARS-CoV-1, and SARS-CoV-2. The outcome was a remarkable improvement in the binding affinities of all clinically relevant antibodies tested, even in cases where these antibodies were already highly evolved, ultimately resulting in enhanced affinity and naturalization[8].

Antibody maturation is a field that numerous research groups have been actively pursuing, aiming to enhance binding properties. This pursuit involves improving monoclonal antibodies for various indications and enhancing binders for cell therapies as well. Therefore, it comes as no surprise that the application of ML/DL and AI has permeated the realms of immunotherapy, cell engineering, and cell therapy. Particularly in the field of CAR-T and CAR-NK cells, ML/DL can aid in addressing various challenges related to activation, exhaustion, persistence, tonic signaling, on-target off-tumor effects, and synapse interactions with tumor cells.[9,10]

One of the primary challenges in CAR-T design lies in ensuring the CAR-T's precise discrimination between normal and cancerous tissues. An in-silico screening approach aimed at identifying multiple antigen signatures capable of enhancing CAR-T's tumor recognition abilities through the integration of various antigen inputs using Boolean logic (AND / NOT). Dannenfelser and colleagues analyzed and screened over 2.5 million dual antigens and 60 million triple antigens across 33 different cancer cell types and 34 normal tissues. Their findings revealed that dual antigens significantly outperformed the best single clinically investigated CAR targets, while triple antigens were predicted to be the most effective discriminators between tumor and normal tissue in several tumor types. This underscores the importance of utilizing Boolean logic gates, employing

2- to 3 antigens, to enhance CAR-Ts' tumor recognition capabilities based on bioinformatic analyses that predict the effect of cellular immunotherapy.[11]

As overall survival rates in cancer patients have improved, resistance to cornerstone therapies has become more prevalent. This is particularly evident in cases of multiple myeloma. To address this challenge, there has been a proposal to explore pharmacological strategies that can enhance vulnerabilities to immunotherapy in resistant diseases. Such an approach may offer a unified perspective for overcoming resistance to Proteasome Inhibitors in the new era of myeloma treatment. The characterization of surface proteomic changes in these contexts holds the potential to unveil novel strategies for diagnosing and specifically treating drug-resistant diseases.[12]

In a study conducted by Ferguson et al., a bioinformatic approach was employed to analyze a vast array of genes/proteins, totaling more than 33,654. They devised a five-criteria scoring system that integrated surface proteomics data with publicly available mRNA transcriptome data. This innovative approach led to the identification of new potential surface targets for antigen-specific immunotherapies in myeloma, including targets such as CCR10, TXNDC11, LILR4B, and SEMA4A. The expression of these targets was validated in both cell lines and primary samples.[13] This predictive model facilitated the design of a novel anti-CCR10 CAR-T therapy, leveraging its natural ligand (CCL27), and demonstrating activity against cell lines in vitro.[13,14]

Beyond the binders expressed on the surface of CAR-T cells, the costimulatory domains typically reside intracellularly, responsible for initiating downstream signals to activate and co-stimulate ultimately enhancing the killing of tumors. This costimulatory domain stands as a pivotal component within the CAR construct, regardless of the specific binder employed. Its role extends beyond bolstering the function and efficacy of CAR-T cell therapy, encompassing the determination of T-cell fate, including aspects related to memory phenotype, exhaustion, and persistence. These factors are closely linked to improved antitumor activity, enhanced persistence, and efficient cytotoxicity against tumor cells, both in vitro and in vivo.[15,18]

Daniels et al., embarked on the construction of a library of CARs, encompassing approximately 2300 synthetic costimulatory domains, created through combinations of 13 signaling motifs. These CARs facilitated diverse outcomes in human T cells, with outcomes sensitive to variations in motif combinations and configurations. The utilization of neural networks, trained to decipher the combinatorial language of CAR signaling motifs, allowed for the extraction of key design principles. For instance, non-native combinations of motifs, binding tumor necrosis factor receptor–associated factors (TRAFs) and phospholipase C gamma 1 (PLCg1), were found to enhance cytotoxicity and stemness, factors associated with effective tumor eradication. Consequently, libraries constructed from fundamental signaling building blocks, complemented by machine learning, can effectively guide the engineering of receptors tailored to exhibit desired phenotypes[19].

Indeed, as mentioned, it's clear that synthetic biology is progressively relying on computational predictions through ML/DL and AI to enhance living therapies. Rather than undergoing painstaking and time-consuming laboratory screenings, bioinformatic analyses empower the identification of promising candidates based on predictive models, which can then be validated in vitro or in

vivo. This methodology undeniably expedites the discovery of novel targets and pathways, thereby facilitating the progression and refinement of the next generation of cellular therapies. Moreover, it preserves both time and resources, a particularly valuable asset in low to middle income countries.

Referencias

1. Garcia JM, Burnett CE, Roybal KT. Toward the clinical development of synthetic immunity to cancer [Internet]. Immunological Reviews. John Wiley and Sons Inc; 2023. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/imr.13245
2. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, et al. Mastering the game of Go without human knowledge. Nature [Internet]. 2017 Oct 18;550(7676):354–9. Available from: https://www.nature.com/articles/nature24270
3. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language Models are Unsupervised Multitask Learners [Internet]. Available from: https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf
4. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language Models are Few-Shot Learners. 2020 May 28 [cited 2023 Nov 7]; Available from: https://arxiv.org/abs/2005.14165v4
5. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. 2017 Mar 20; Available from: http://arxiv.org/abs/1703.06870
6. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature [Internet]. 2021 Aug 26;596(7873):583–9. Available from: https://www.nature.com/articles/s41586-021-03819-2#citeas
7. Choudhry P, Gugliemini O, Geng H, Sarin V, Sarah L, Paranjape N, et al. Functional multi-omics reveals genetic and pharmacologic regulation of surface CD38 in multiple myeloma. Available from: https://doi.org/10.1101/2021.08.04.455165
8. Hie BL, Shanker VR, Xu D, Bruun TUJ, Weidenbacher PA, Tang S, et al. Efficient evolution of human antibodies from general protein language models. Nat Biotechnol [Internet]. 2023; Available from: https://www.nature.com/articles/s41587-023-01763-2
9. Naghizadeh A, Tsao WC, Cho JH, Xu H, Mohamed M, Li D, et al. In vitro machine learning-based CAR T immunological synapse quality measurements correlate with patient clinical outcomes. PLoS Comput Biol [Internet]. 2022 Mar 1;18(3). Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009883
10. Lee M, Lee YH, Song J, Kim G, Jo YJ, Min HS, et al. Deep-learning based three-dimensional 1 label-free tracking and analysis of immunological synapses of car-t cells. Elife [Internet]. 2020 Dec 1;9:1–53. Available from: https://elifesciences.org/articles/49023
11. Dannenfelser R, Allen GM, VanderSluis B, Koegel AK, Levinson S, Stark SR, et al. Discriminatory Power of Combinatorial Antigen Recognition in Cancer T Cell Therapies. Cell Syst [Internet]. 2020 Sep 23;11(3):215-228.e5. Available from: https://www.sciencedirect.com/science/article/pii/S2405471220302866
12. Patiño-Escobar B, Talbot A, Wiita AP. Overcoming proteasome inhibitor resistance in the immunotherapy era. Trends Pharmacol Sci [Internet]. 2023 Aug 1;44(8):507–18. Available from: https://www.cell.com/trends/pharmacological-sciences/fulltext/S0165-6147(23)00111-6
13. Ferguson ID, Patiño-Escobar B, Tuomivaara ST, Lin YHT, Nix MA, Leung KK, et al. The surfaceome of multiple myeloma cells

suggests potential immunotherapeutic strategies and protein markers of drug resistance. Nat Commun [Internet]. 2022 Dec 1;13(1). Available from: https://www.nature.com/articles/s41467-022-31810-6

14. Patino-Escobar B, Kasap C, Ferguson I, Hale M, Wiita A. P-103: Profiling the myeloma cell surface proteome reveals CCR10 as a potential immunotherapeutic target. Clin Lymphoma Myeloma Leuk [Internet]. 2021 Oct;21:S94–5. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2152265021022370
15. Long AH, Haso WM, Shern JF, Wanhainen KM, Murgai M, Ingaramo M, et al. 4-1BB costimulation ameliorates T cell exhaustion induced by tonic signaling of chimeric antigen receptors. Nat Med [Internet]. 2015 Jun 9;21(6):581–90. Available from: https://www.nature.com/articles/nm.3838
16. Boucher JC, Li G, Kotani H, Cabral ML, Morrissey D, Lee SB, et al. CD28 costimulatory domain-targeted mutations enhance chimeric antigen receptor T-cell function. Cancer Immunol Res [Internet]. 2021 Jan 1;9(1):62–74. Available from: https://aacrjournals.org/cancerimmunolres/article/9/1/62/470226/CD28-Costimulatory-Domain-Targeted-Mutations
17. Gattinoni L, Speiser DE, Lichterfeld M, Bonini C. T memory stem cells in health and disease [Internet]. Vol. 23, Nature Medicine. Nature Publishing Group; 2017. p. 18–27. Available from: https://www.nature.com/articles/nm.4241
18. Singh N, Perazzelli J, Grupp SA, Barrett DM. Early memory phenotypes drive T cell proliferation in patients with pediatric malignancies [Internet]. Available from: https://www.science.org/doi/10.1126/scitranslmed.aad5222
19. Daniels KG, Wang S, Simic MS, Bhargava HK, Capponi S, Tonai Y, et al. Decoding CAR T cell phenotype using combinatorial signaling motif libraries and machine learning [Internet]. Available from: https://www.science.org/doi/abs/10.1126/science.abq0225?af=R&utm_source=sfmc&utm_medium=email&utm_campaign=1stRelease&utm_content=alert&et_rid=719334783&et_cid=4522378