Can AI Design Cancer Vaccines? Evaluating Neural Networks for Epitope Prediction

Caitlin Roach
Queen's University
caitlin.roach@queensu.ca

Jennifer Qiu
Queen's University
20jqq@queensu.ca

Jaeson Wang
Queen's University
21jw145@queensu.ca

Salma Elsayed
Queen's University
salma.elsayed@queensu.ca

Adeel Haq
Queen's University
21ah33@queensu.ca

***Abstract*—Immunotherapy as a form of cancer treatment can be effective, but often causes the immune system to attack healthy tissues, leading to significant side effects. Therapeutic cancer vaccines offer a safer, tumour-specific alternative, but their efficiency relies on accurate epitope prediction, which is used to identify regions of a protein that can trigger an immune response in a patient. This study evaluates MHCflurry on a clinically relevant melanoma-associated antigen to assess the real-world applications of computational epitope prediction to therapeutic melanoma vaccines. We assessed predicted epitopes based on binding affinity, presentation, and processing scores, identifying the peptide sequence AQAPATEEQEA as the strongest candidate. We visualized results and key findings for a quantitative analysis of the peptide sequences. Our findings suggest that while computational tools like MHCflurry show promise in the design of cancer vaccines, they require experimental validation before implementation or clinical application.**

I. Introduction

Melanoma is a type of skin cancer in which malignant (cancer) cells rapidly multiply in the cells that colour the skin (melanocytes). This is the most dangerous type of skin cancer due to its aggressive growth and ability to spread to any organ in the body (National Cancer Institute, 2025). Melanoma is diagnosed by biopsy and analysis of skin lesions. Treatment plans include radiation therapy, chemotherapy, and, notably, immunotherapy.

Immunotherapy trains the body's immune system to recognize, target, and attack malignant cells, stopping or slowing cancer growth while preventing it from spreading to other organs. Types of immunotherapy include immune checkpoint inhibitors (ICIs), CAR T cell therapy, antibody-drug conjugates, and therapeutic vaccines. This treatment can be effective, but often causes the immune system to attack healthy cells and tissues, leading to what are called immune-related adverse events (irAEs). irAEs can affect many organ systems, including the skin, liver, and gastrointestinal system (Vaddepally et al., 2022). These effects are often serious. In a 2023 study, combination therapy with ICIs resulted in a 25% to 30% incidence of grade 4 (life-threatening) hepatitis and a grade 3 (severe) toxicity rate of approximately 15% (Yin et al., 2023).

Given this need for safer immunotherapy options, therapeutic vaccines present an opportunity for safer alternatives to traditional immunotherapy. Therapeutic cancer vaccines aim to provide tumor-specific treatment with fewer side effects, resulting in far less severe irAE rates than other forms of treatment. A 2025 study on a personalized therapeutic vaccine for advanced kidney cancer resulted in only mild flu-like symptoms, with no serious side effects reported (Braun et al., 2025), highlighting the potential that vaccines have to reduce irAE rates.

These vaccines are designed using epitope prediction, a computational method used to identify regions of a protein that can trigger an immune response in a patient. Machine learning shows great potential in the prediction of clinically relevant epitopes, improving vaccine efficacy while minimizing adverse effects on patients such as irAEs.

A. Motivation

The development of safe, widely accessible cancer treatment with minimal side effects is crucial to optimize quality of life for patients and their families, improve overall survival rates, and address toxicities like irAEs.

An optimal treatment would mitigate the broader societal impacts of cancer, including economic burden due to reduced labour force participation and productivity (OECD, 2024), strain on healthcare systems (Prager et al., 2018), and psychological impact of traumatic treatment plans (van Roij et al., 2019). Therapeutic vaccines offer a promising candidate for this treatment, and accurate epitope prediction is crucial for their development.

B. Problem Definition

Although immunotherapy can be effective in cancer treatment, high irAE rates pose serious risks to patient safety, limiting widespread use. Severe irAEs often lead to significant health complications, longer hospitalizations, and treatment discontinuation, resulting in lower overall survival rates (Liang et al., 2024).

Therapeutic cancer vaccines offer a promising alternative to traditional immunotherapy by inducing a targeted immune response in the patient, lowering the risk of off-target effects.

However, their development and implementation is limited by challenges in epitope prediction.

Recent advances in machine learning have introduced new approaches to epitope prediction. This study aims to evaluate machine learning epitope prediction tools like MHCflurry, determining their feasibility in identifying strong binding epitopes to inform melanoma vaccine design for improved efficiency and patient safety.

II. Related Work

In response to the growing demand for safer and more effective cancer treatments, many studies have explored computational approaches for epitope prediction, leveraging machine learning models for efficient vaccine design.

For instance, Tarek et al. (2018) applied computational epitope prediction tools to evaluate peptide sequences for non-small cell lung cancer vaccine design, identifying several promising candidates, but found that binding affinity scores did not always correlate with real-world immunogenicity. Similarly, Roudko et al. (2020) utilized computational tools to predict peptide interactions, but noted biases in training data.

Our study builds on this research by integrating MHCflurry, a machine learning-based predictor, with a real-time web application to improve accessibility for researchers. Unlike previous studies, we emphasize interpretability, open-source data, and scalability for broader vaccine design applications. In this study, we evaluate MHCflurry on a clinically verified melanoma-associated antigen (MAGE-A3) and analyze its performance in epitope prediction for immunotherapy.

III. Methodology

Due to computational limitations, we opted to conduct this study using MHCflurry, a well-documented, pre-trained machine learning-based epitope prediction tool. While other predictors exist (e.g., NetMHCpan), MHCflurry was selected for its open-source accessibility and previous validation in literature.

MHCflurry predicts peptide binding affinities using a system of artificial neural networks (ANNs). These ANNs use both classification and regression to identify potential candidates for vaccine design. Accuracy is determined through a mean square error loss function.

We evaluated this system on a clinically verified melanoma-associated antigen (MAGE-A3), sourced from the Immune Epitope Database (IEDB) to study its accuracy at predicting epitope candidates for melanoma vaccine design. We ran MHCflurry on 9-mer peptide sequences from MAGE-A3, predicting binding affinities for the HLA-A*02:01 allele (due to high clinical relevance). Candidates were evaluated and ranked based on their predicted binding affinity scores, with lower values indicating stronger binding potential.

We used Python 3 to create a web application, which was deployed using Gradio for a smooth user experience. We leveraged matplotlib and seaborn to visualize results from MHCflurry quickly and efficiently from within the user's browser.

A. Evaluation Methods

We evaluated the model on several key metrics, including half-maximal inhibitory concentration (IC50) scores as a measurement of the binding affinity of predicted epitope candidates. A very low IC50 score suggests an exceptional candidate for vaccine development, as strong binding increases the chances of recognition by the immune system.

Interpreting the results graphically, we were able to evaluate MHCflurry's predictions on the basis of interactions within the body's immune system, exploring real-world applications to vaccine design.

IV. Results and Discussion

While developing this project, we aimed to assess the feasibility of computational tools like MHCflurry with the goal of improving the efficiency, accuracy, and safety of cancer immunotherapy.

This computational approach identified strong epitopes for melanoma vaccine design that are not yet documented in existing immunogenicity databases. This suggests potential novel epitope candidates for melanoma vaccine design while highlighting the need for further statistical analysis and future experimental validation.

A. Results

For readability, we interpreted the results of this study both graphically and in terms of raw data.

TABLE I
MHCflurry MAGE-A3 Results

peptide	affinity (IC50 in nM)	presentation	processing
AQAPATEEQEA	2162.32	0.058	0.100
ALGLVGAQAPA	2912.19	0.032	0.008
AASSSSTLVEV	4185.49	0.023	0.007
GLEARGEALGL	4205.67	0.027	0.052
TLGEVPAAESP	11775.71	0.010	0.051
SNQEEEGPSTF	13409.89	0.007	0.004
TLVEVTLGEVP	13981.99	0.014	0.206
STLVEVTLGEV	14313.88	0.008	0.040
GLVGAQAPATE	16062.32	0.006	0.004
SPDPPQSPQGA	16979.16	0.006	0.006

This table shows the raw output of MHCflurry on MAGE-A3, which predicted peptide sequences (epitopes) as potential candidates for vaccine design. In this study, peptides are evaluated on key metrics such as IC50 score, presentation on the surface of a cell, and processing within the cell. During this prediction run, MHCflurry identified the AQAPATEEQEA sequence the strongest candidate among the predicted epitopes due to its low IC50 score, coupled with high presentation and processing metrics. This suggests that this specific sequence has a strong likelihood of being recognized by immune cells, and would therefore be an ideal candidate for vaccine design.

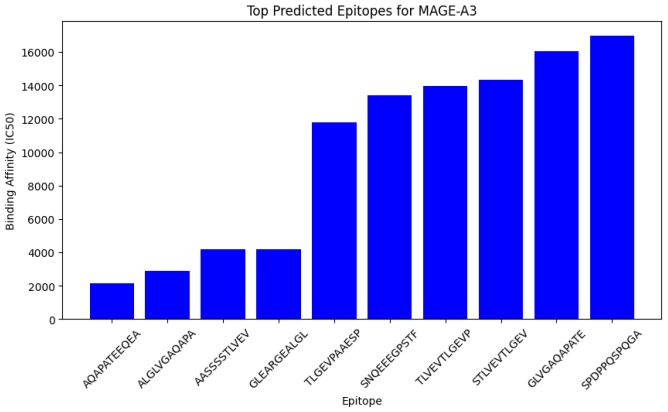

Fig. 1. Visualization of epitope candidates for MAGE-A3

MHCflurry identified multiple epitope candidates with varying binding affinities. The top candidate, AQAPATEEQEA, is shown here with the strongest predicted binding, suggesting a promising target for melanoma vaccine design.

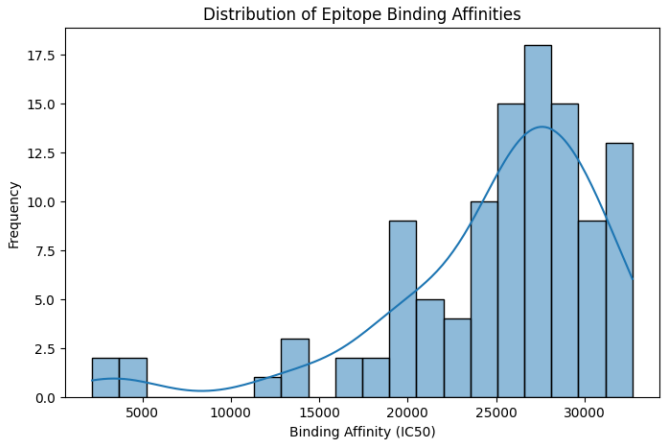

Fig. 2. Distribution of binding affinity (IC50) scores in epitope predictions from MHCflurry

The majority of predicted epitopes have high IC50 values, indicating weak binding affinity. However, the peptide AQAPATEEQEA falls in the low IC50 range, reinforcing its potential.

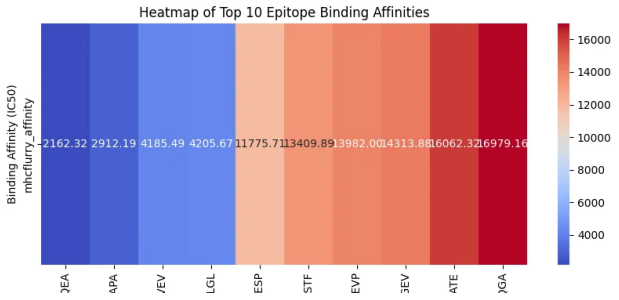

Fig. 3. Heatmap visualization of top predicted MAGE-A3 epitopes

This heatmap ranks the top 10 predicted epitopes for MAGE-A3 based on their IC50 values. Lower IC50 values (blue) indicate stronger binding, while higher IC50 values (red) indicate weak binding. AQAPATEEQEA shows the strongest binding affinity, making it a promising target.

B. Ethical Considerations

Although therapeutic cancer vaccines offer a promising safe alternative to traditional cancer treatments, their development, evaluation, and integration with existing medical frameworks present significant ethical challenges, particularly regarding access to treatment, accountability, and data privacy.

Therapeutic cancer vaccines raise ethical concerns about who benefits from this type of specialized cancer treatment, calling into question how factors like socioeconomic status, geographical location, insurance coverage, and income inequality will influence cancer treatment. When developing new treatments for diseases as prevalent as cancer, it is also important to consider the social determinants of health - the non-medical living conditions that influence well-being - and how they may affect treatment. Factors like income, education, employment, and access to healthcare services all affect the impact of therapeutic cancer vaccines, and must be considered.

Due to the complex nature of biological training data, computational tools like MHCflurry raise ethical concerns around algorithmic bias, transparency, and data privacy. Such models must be evaluated to identify and mitigate biased outputs, ensuring that these tools are well-equipped to serve diverse populations.

Furthermore, as machine learning continues to advance, it remains crucial that all computational results are rigorously tested and examined in both lab work and clinical settings to ensure integrity, accuracy, and efficiency.

V. Conclusion

This study explored computational epitope prediction with machine learning tool MHCflurry, revealing novel epitope candidates for melanoma vaccine design. Further analysis and clinical review are needed to verify clinical significance and real-world feasibility of these results. We demonstrated the potential of machine learning-based epitope prediction in the process of designing therapeutic melanoma vaccines, highlighting the growing role of machine learning in bioinformatics and computational biology.

As machine learning continues to advance, its integration with laboratory research could significantly enhance the precision and accessibility of therapeutic cancer vaccines, improving quality of life and overall survival rates for patients.

VI. Future Work

In the future, we plan to develop this tool further, potentially integrating computational epitope prediction with deep learning models to create an entirely open-source vaccine design tool for cancer researchers. By developing an open-source web application, we plan to make epitope prediction for cancer therapeutics accurate and widely available, enabling

researchers to quickly test vaccine candidates with minimal computational cost. We will need to conduct further statistical analysis and comparison to real-world benchmarks to confirm the real-world applications and immunogenicity of epitopes predicted using this tool.

Furthermore, integrating Major Histocompatibility Complex (MHC) class I and II prediction could potentially improve the effectiveness of this vaccine development tool. Class I epitopes recognize and destroy harmful cells, while class II epitopes enhance immunological memory, preventing the recurrence of cancer (Wang et al., 2021).

VII. Limitations

During the development of this project, we experienced several challenges that limited the scope of our research.

Due to system compatibility issues, we were unable to run epitope-predict natively on local hardware. We adapted the implementation to run on Google Colab, which introduced limitations in computational resources and runtime constraints.

Additionally, the results of this study are based solely on computational predictions. While MHCflurry provides key insights into potential epitope candidates, experimental validation is necessary to confirm the real-world applications of the identified epitope candidates.

References

[Braun et al., 2025] Braun, D., Moranzoni, G., Chea, V., et al. (2025). A neoantigen vaccine generates antitumour immunity in renal cell carcinoma. *Nature*.

[Cook et al., 2024] Cook, S., Samuel, V., Meyers, D. E., and et al. (2024). Immune-related adverse events and survival among patients with metastatic nsclc treated with immune checkpoint inhibitors. *JAMA Network Open*.

[Farrell, 2021] Farrell, D. (2021). epitopepredict: a tool for integrated mhc binding prediction. *GigaByte*.

[Institute, 2025] Institute, N. C. (2025). Melanoma treatment (pdq®) – health professional version.

[Mohammad M. Tarek, 2018] Mohammad M. Tarek, Ayman E. Shafei, M. A. A. M. M. M. (2018). Computational prediction of vaccine potential epitopes and 3-dimensional structure of xage-1b for non-small cell lung cancer immunotherapy. *Biomedical Journal*.

[O'Donnell et al., 2020] O'Donnell, T., Rubinsteyn, A., and Laserson, U. (2020). Mhcflurry 2.0: Improved pan-allele prediction of mhc i-presented peptides by incorporating antigen processing. *Cell Systems*.

[OECD, 2024] OECD (2024). *Tackling the Impact of Cancer on Health, the Economy and Society*. OECD Publishing, Paris.

[O'Donnell et al., 2018] O'Donnell, T., Rubinsteyn, A., Bonsack, M., Riemer, A. B., Laserson, U., and Hammerbacher, J. (2018). Mhcflurry: Open-source class i mhc binding affinity prediction. *Cell Systems*.

[Prager et al., 2018] Prager, G. W., Braga, S., Bystricky, B., Qvortrup, C., Criscitiello, C., Esin, E., Sonke, G. S., Martínez, G. A., Frenel, J. S., Karamouzis, M., Strijbos, M., Yazici, O., Bossi, P., Banerjee, S., Troiani, T., Eniu, A., Ciardiello, F., Tabernero, J., Zielinski, C. C., Casali, P. G., and Ilbawi, A. (2018). Global cancer control: responding to the growing burden, rising costs and inequalities in access. *ESMO Open*, 3(2):e000285.

[Vaddepally et al., 2022] Vaddepally, R., Doddamani, R., Sodavarapu, S., Madam, N. R., Katkar, R., Kutadi, A. P., Mathew, N., Garje, R., and Chandra, A. B. (2022). Review of immune-related adverse events (iraes) in non-small-cell lung cancer (nsclc)-their incidence, management, multi-organ iraes, and rechallenge. *Biomedicines*, 10(4):790.

[van Roij et al., 2019] van Roij, J., Brom, L., Youssef-El Soud, M., van de Poll-Franse, L., and Raijmakers, N. J. H. (2019). Social consequences of advanced cancer in patients and their informal caregivers: a qualitative study. *Supportive Care in Cancer*, 27(4):1187–1195.

[Wang et al., 2021] Wang, X., Yu, Z., Liu, W., Tang, H., Yi, D., and Wei, M. (2021). Recent progress on MHC-I epitope prediction in tumor immunotherapy. *American Journal of Cancer Research*, 11(6):2401.

[Yin et al., 2023] Yin, Q., Wu, L., Han, L., Zheng, X., Tong, R., Li, L., Bai, L., and Bian, Y. (2023). Immune-related adverse events of immune checkpoint inhibitors: a review. *Frontiers in Immunology*, 14:1167975.

[O'Donnell et al., 2020] [O'Donnell et al., 2018] [Yin et al., 2023] [Cook et al., 2024] [Vaddepally et al., 2022] [OECD, 2024] [van Roij et al., 2019] [Prager et al., 2018] [Braun et al., 2025] [Mohammad M. Tarek, 2018] [Farrell, 2021] [Institute, 2025] [Wang et al., 2021]