Global Health Care
WISDOM ACADEMIC https://doi.org/10.63808/ghc.v2i1.498 Volume 2 , Issue 1

Research Article

Using multimodal foundational models to predict neoantigen immunogenicity and vaccine effectiveness across different tumor types

Gang Liu[1,*], Jia Wang[2], Jia Zhu[1]
[1]International Sakharov Environmental Institute, Belarus State University, Minsk 220070, Belarus
[2]Taishan Vocational and Technical College, Tai'an 271000, China
[*]**Correspondence:** Gang Liu, lg6666699@163.com

Abstract

Neoantigen vaccines are a key component of personalised cancer immunotherapy; however, existing prediction techniques are primarily restricted to a single tumour type and have difficulty integrating multi-modal biological data, which leads to inadequate accuracy in immunogenicity evaluation and vaccine efficacy prediction. In order to predict neoantigen immunogenicity and customised vaccination clinical outcomes across various tumour types, this study attempts to develop and verify a universal multi-modal foundation model.Neoantigen peptide sequences, mass spectrometry-derived pHLA binding patterns, single-cell TCR repertoires, tumour transcriptomes, and clinical vaccination trial follow-up records were among the extensive paired data we gathered from 15 solid tumour types. We present the NeoVAX-FM multi-modal foundation model, which first embeds peptide sequences, pHLA complex 3D structures, and gene expression into a single semantic space using a contrastive language-image pretraining paradigm. It is then fine-tuned on downstream tasks to concurrently score immunogenicity and predict progression-free survival.The model was assessed in one prospective clinical trial and three external validation cohorts. NeoVAX-FM showed strong performance in melanoma, non-small cell lung cancer, and microsatellite stable colorectal cancer, with an average AUC of 0.94 for cross-tumor neoantigen immunogenicity prediction — a 12.3% increase over the best currently available techniques. Patients in the prospective vaccination cohort who were projected by the model to be "high responders" had a considerably higher median progression-free survival (HR = 0.28, $p < 0.001$), and the model was successful in identifying tumour microenvironment characteristics and universal TCR motifs that drive long-term responses.This work is the first to use a multi-modal foundation model for neoantigen vaccines, overcoming the tumor-type-specific constraints of conventional approaches and facilitating joint modelling of

immunogenicity and clinical efficacy, thereby offering an AI decision engine for precision cancer vaccine design that is applicable to all cancer types.

Keywords

Neoantigen, Immunogenicity prediction, Multimodal foundation model, Cancer vaccine, Cross-tumor, Precision immunotherapy

1. Introduction

Neoantigen vaccines are thought to be a crucial tactic for generating long-lasting anti-tumor immune responses because they are highly customised and tumor-specific. Tumour immunotherapy has significantly altered how we treat solid tumours. After being processed and presented by major histocompatibility complex (MHC) molecules on the cell surface, neoantigens-which originate from somatic mutations in tumors-are identified by T cell receptors (TCRs), which initiates immunological destruction. Neoantigen vaccinations can cause neoantigen-specific T cell expansion in melanoma, glioblastoma, and non-small cell lung cancer, which is associated with therapeutic advantages, according to many recent clinical trials. The essential necessity for high-precision prediction and prioritisation of neoantigen immunogenicity is highlighted by the fact that only a small number of putative neoantigens that reach the clinic truly produce immunity and cause tumour regression.

MHC-I binding affinity, peptide-MHC complex stability, and TCR recognition probability are examples of single or few-dimensional parameters that are the mainstay of current prediction techniques. Although the accuracy of binding affinity estimates has greatly increased thanks to tools like NetMHCpan and MHCflurry, there is still a great deal of ambiguity regarding real immunogenicity—the majority of high-affinity mutant peptides do not cause T cell responses. Researchers have been combining information from naturally occurring peptides found by mass spectrometry, tumour transcriptomics, proteomics, and TCR repertoires in order to close this gap. They have also suggested machine learning-based integrated scoring models like as PRIME, DeepHLApan, and NeoPredPipe.These approaches, however, typically have the following drawbacks: first, the majority are designed for a particular type of tumour, and their generalisability across various cancers with different immune microenvironments has not been confirmed; second, the ways in which they combine various modalities are relatively shallow, frequently merely concatenating multi-source features, and failing to capture the deep semantic relationships between peptide sequences, pHLA 3D structures, gene expression profiles, and TCR clonotypes; and third, they are weak in clinical translation because the great majority of models endpoints, such as progression-free

survival or objective response rates.

In the biomedical realm, foundation models are simultaneously causing a paradigm change. These models can learn transferable general representations from vast amounts of unlabelled biological sequences, structures, and omics data through large-scale self-supervised pretraining. They then refine on downstream tasks, greatly improving prediction performance in low-sample scenarios. Protein structure prediction, gene regulatory network inference, and cross-modal retrieval tasks have demonstrated exceptional performance by the protein language model ESM-2, the single-cell foundation model scGPT, and the multimodal biomedical model BioMedGPT.The ability of multimodal foundation models to project peptide sequences, molecular structures, and transcriptomes into a single semantic space for natural integration is particularly noteworthy. This provides an entirely new method of integrating multi-source heterogeneous data, such as "sequence-structure-expression-clinical" for neoantigen immunogenicity prediction. Nevertheless, no study has tried to employ foundation models to concurrently predict neoantigen immunogenicity and tailored vaccination clinical effectiveness across various tumour types, nor has a multimodal foundation model been especially created for neoantigen vaccine development.

This study presents NeoVAX-FM, a multimodal foundation model intended for neoantigen vaccine development, to fill up the major gaps mentioned above. Fundamentally, the model employs a contrastive language-image pretraining technique, incorporating single-cell TCR repertoire data, tumour transcriptome profiles, pHLA complex 3D structures, and altered peptide sequences into a single representation space. It simultaneously produces cross-tumor neoantigen immunogenicity scores and vaccination progression-free survival forecasts through downstream fine-tuning. Our objective is to respond to these research enquiries: Is it possible to greatly increase the precision and generalisability of neoantigen immunogenicity predictions across various tumour types using a multimodal foundation model? Is it possible to directly apply this unified paradigm to the clinical advantages of customised vaccinations, offering trustworthy decision assistance for vaccine design and patient stratification? Do the model's learnt representations highlight microenvironmental characteristics and universal TCR motifs that promote durable anti-tumor immunity and improve biological interpretability? The goal of this project is to construct the first AI decision engine for precision cancer vaccine development that is applicable to all tumour types, based on extensive multi-cancer data and prospective cohort validation.

2. Literature Review

2.1. Progress in Computational Methods for Neoantigen Prediction

Tumour genome sequencing and somatic mutation discovery are typically the first steps in the computational pipeline for neoantigen prediction. Next, putative mutant peptides are screened using MHC-peptide binding affinity prediction techniques. The NetMHCpan series has undergone several versions and is regarded as a benchmark in this industry. With a median AUC of 0.89 for binding affinity prediction, the current version 4.1, which covers more than 200 MHC-I alleles, is based on pan-specific neural networks [1].In contrast, MHCflurry 2.0, which is trained utilising integrated deep learning architectures and mass spectrometry-eluted ligand data, offers faster inference speed and performs comparably to NetMHCpan 4.0 on independent test sets [2]. However, a thorough analysis by Jurtz et al. [3] revealed that using binding affinity alone as a filtering criterion yields a positive predictive value of less than 5%-the majority of high-affinity mutant peptides do not elicit T cell responses in vivo. In the field, this “affinity-immunogenicity gap” has emerged as a major obstacle.

Researchers have progressively added multi-dimensional characteristics to close the aforementioned gap. By identifying natural MHC-I ligands on a wide scale using mass spectrometry-based immunopeptidomics, Abelin and colleagues[4] demonstrated the independent contribution of presentation efficiency to immunogenicity during antigen processing and presentation. Peptide abundance, gene expression, and proteasomal cleavage likelihood were further combined by Bassani-Sternberg and colleagues[5], who proposed a thorough scoring structure that greatly increased the true positive rate. In terms of computational tools, the PRIME model achieved an AUC of 0.85 in immunogenicity binary classification tasks by taking MHC binding stability into consideration and using molecular dynamics simulations to determine the half-life of peptide-MHC complexes[6]. A multi-task learning approach called DeepHLApan[7] predicted binding affinity, pHLA stability, and presentation probability all at once, outperforming single-task models in a variety of external datasets. Nevertheless, the deep nonlinear interactions between sequence, structure, and expression profiles are not captured by current approaches, which typically employ shallow multimodal fusion algorithms (either concatenating feature vectors or allocating weights based on heuristic heuristics).

2.2. Clinical Trials of New Antigen Vaccines and Biomarker Research

Neoantigen vaccines have reached a critical testing stage in their clinical development. Personalised neoantigen peptide vaccines were administered to six melanoma patients in a Phase I trial published in Nature by Ott et al. Four of them did not experience a relapse during a 25-month follow-up, and all of them displayed an increase in neoantigen-specific $CD4^+$ and

$CD8^+$ T cells [8].Concurrently, Sahin and colleagues tested RNA neoantigen vaccinations in 13 melanoma patients, demonstrating that 8 of them had sustained progression-free survival and that the vaccine-induced T cell responses could target tumour cells of the wild type [9].Despite the limitations of MHC-I presentation and tumour heterogeneity, Keskin and colleagues observed an increase in immune infiltration during the first novel antigen vaccine trial for central nervous system tumours in glioblastoma [10].Personalised peptide vaccines based on neoantigens can safely elicit long-lasting T cell memory in the field of non-small cell lung cancer, according to Ding and colleagues' Phase I/II trials [11]. The median progression-free survival was much longer than historical controls.

The identification of important biomarkers was another outcome of these clinical trials. After deep sequencing the TCR repertoires of vaccination responders and non-responders, Hu et al. discovered that responders shared a collection of "public TCR motifs" that can cross-recognize homologous mutant peptides and exhibit conserved sequence patterns in the CDR3β region [12].The diversity of T-cell clones produced by the vaccine is favourably correlated with clinical advantages, and the traits of pre-existing tumor-infiltrating T cells can function as response predictors, according to Cafri and colleagues [13].Liu and colleagues discovered through single-cell transcriptome analysis that dendritic cell maturation scores and CXCL9/10 expression levels were significantly higher in responder tumours at the tumour microenvironment level, suggesting that innate immune activation may cooperate with the vaccine effect [14]. Nevertheless, the majority of the biomarker studies mentioned above are small-sample explorations focusing on a single tumour type, lacking systematic comparisons and validations across different tumour types, and a cross-tumor prediction system is desperative [23].

2.3. Applications of Foundation Models and Multimodal Learning in Biomedicine

Through extensive self-supervised pretraining, foundation models acquire broad representations of data, leading to a paradigm shift in computational biology. The protein language model ESM-1b, which employs a Transformer architecture and has been pretrained on 250 million protein sequences, was suggested by Rives and colleagues. In zero-shot settings, the learnt hidden representations almost match the performance of supervised models and may be used directly to forecast mutation effects [15].By adding 650 million factors to the ESM-2 model, Lin and colleagues were able to predict atomic resolution structures using a single sequence for the first time [16].Cui and colleagues created scGPT in the field of single-cell omics. It was pretrained on more than 10 million single-cell transcriptomes and

may be used for a variety of downstream tasks, including gene regulatory network inference, perturbation response prediction, and cell type annotation [17].Theodoris and colleagues introduced Geneformer, which was pretrained on about 30 million single-cell transcriptomes and, with minimal fine-tuning, may find novel regulatory factors in heart development and disease networks [18].Although this method hasn't yet been used in the field of neoantigen vaccines, Auslander and colleagues developed a model by combining transcriptome and mutation characteristics that successfully categorised PFS in melanoma patients treated with immune checkpoint inhibitors (HR=0.31) [24].

The development of multimodal foundation models has stretched representation learning's limits even farther. By using contrastive language-image pretraining, Radford and colleagues' CLIP approach unifies text and picture embeddings in a semantic space, allowing for zero-shot cross-modal retrieval and classification. The biological profession has made extensive use of this strategy [19].Lu and colleagues created BioMedCLIP, which adapts the CLIP architecture to biomedical pictures and literature and has significant generalisation ability in tasks related to medical imaging diagnosis and report production [20].Liu and colleagues developed MolCLIP at the molecular characterisation level, which greatly increases the accuracy of molecular property predictions with few samples by aligning molecular graphs with biochemical description texts via contrastive learning [21].In their study of fundamental models' potential for drug development, Xu and colleagues noted that their promise for creating novel antigen vaccines has not yet been completely realised [22].

2.4. Limitations of Existing Research and the Contributions of This Study

A lack of cross-tumor generalisation is one of the main issues with the current study, according to the literature mentioned above. The majority of neoantigen prediction algorithms now in use are trained and validated on a single tumour type, and rigorous review of generalisation across malignancies with notable variations in immune milieu, mutation burden, and MHC genotypes is lacking. Although several alleles are covered by tools such as NetMHCpan, they have never been benchmarked to predict neoantigen immunogenicity across various cancer types [1-7].The level of multimodal fusion remains shallow.Deep semantic linkages across modalities are not captured by current approaches, which either concatenate diverse data like structure, expression, and TCR or depend primarily on sequence information. Neoantigen prediction has not yet utilised the unified semantic space approach offered by basic models [15-22].Clinical results are unrelated to the anticipated objectives. Few models attempt to connect to hard clinical targets like progression-free survival or objective response rate; the majority utilise MHC binding or in vitro T cell activation as

outcomes. The usefulness of computational tools for practical decision-making in vaccine design is severely limited by this separation [8-14].Insufficient explainability. It is difficult to evaluate the models physiologically and establish clinical confidence since the majority of current deep models are “black boxes” that haven’t disclosed the biological processes underlying their predictions, such as the common TCR motifs driving immunogenicity or microenvironmental characteristics [12-14].

This study fills these gaps by introducing NeoVAX-FM, the first multimodal foundational model for neoantigen vaccines, which uses a contrastive learning approach to embed peptide sequences, pHLA 3D structures, tumour transcriptomes, and TCR repertoires into a unified semantic space; systematically validating the model’s cross-tumor generalisation ability on a large dataset covering 15 types of solid tumours and evaluating its clinical efficacy prediction in a prospective vaccine trial; directly anchoring prediction endpoints to progression-free survival, bridging the gap between molecular predictions and clinical outcomes; and using representation analysis to uncover universal TCR motifs and microenvironment features that drive long-term antitumor immunity. All things considered, this work offers the first AI decision engine for accurate cancer vaccine development that is applicable to all tumour types.

3. Research Methods

3.1. Research Design and Data Sources

This study’s overall plan is to develop a multimodal foundational model called NeoVAX-FM, pretrain and internally validate it using extensive retrospective data across various cancer types, and then systematically assess how well it predicts neoantigen immunogenicity and vaccine efficacy across cancers in a prospective clinical trial and an independent external validation set. Public databases, affiliated institutions, and our own clinical studies are among the data sources.

3.2. Data Collection and Preprocessing

We gathered somatic mutation data from 7,842 patients with 15 different solid tumour types (such as melanoma, non-small cell lung cancer, colorectal cancer, glioblastoma, breast cancer, etc.) from TCGA, ICGC, and published neoantigen vaccination trials. We generated 8-11mer potential mutant peptides using pVACtools for each nonsynonymous mutation and retained the ones resulting from missense mutations. At the same time, we computed antigen processing scores using MHCflurry 2.0 and projected MHC-I binding affinity using NetMHCpan 4.1 with an initial threshold of IC50 < 500 nM. We ultimately obtained over 12

million peptide-MHC pairings.

6,214 tumour samples from the SysteMHC Atlas and IEDB provided us with raw HLA-I immunopeptidomics mass spectrometry data. We conducted a database search using MaxQuant (v2.0.1) to find naturally presented peptides. The collection of actually presented peptides consisted of 8-11 amino acids in length with FDR < 1%. The detected peptides were classified as "naturally presented positive" or "negative" after being cross-checked with mutant peptides. To guarantee data quality and cross-lab repeatability, all mass spectrometry experiments and data reporting adhered to the Minimum Information About Immunopeptidomics Experiments (MIAIPE) guidelines. We also computed the proteasomal cleavage probability (NetChop 3.1) and parent protein abundance (based on tumour transcriptome TPM) for each mutant peptide.

4,521 patients' RNA-seq data were paired and acquired from GEO and TCGA. The paired-end alignment was performed using STAR, the quantification was done using RSEM, and the gene expression value was calculated using log2(TPM+1). For the microenvironment characteristics, ssGSEA was used to compute the activity of 13 immune-related pathways, CIBERSORTx was used to deconvolute the proportions of 22 immune cells, and the expression levels of important molecules such as PD-L1 and CXCL9/10 were extracted individually. These characteristics comprise the tumour microenvironment profile (TME profile), which is one of the model's modality inputs.

TCGA and GEO provided paired RNA-seq data from 4,521 patients. The gene expression value was calculated using log2(TPM+1), RSEM was employed for quantification, and STAR was utilised for paired-end alignment. For the microenvironment characteristics, the expression levels of important molecules such as CXCL9/10 and PD-L1 were extracted independently; the proportions of 22 immune cells were deconvoluted using CIBERSORTx; and the activity of 13 immune-related pathways was computed using ssGSEA. The tumour microenvironment profile (TME profile), which is one modality input for the model, is composed of several characteristics.

From seven published neoantigen vaccine studies (n = 312), we retrospectively gathered individual patient data, including neoantigen sequences, vaccination plans, progression-free survival (PFS) as assessed by RECIST, and best overall responses. 52 patients with advanced solid tumours who had not responded to conventional therapies were included in the prospective validation cohort, which was drawn from a single-arm Phase I/II personalised peptide neoantigen vaccination experiment started at our facility. The same procedure was used to create vaccines, with a minimum 12-month follow-up. Objective response rate was the secondary target, whereas PFS was the primary endpoint.

3.3. NeoVAX-FM Multimodal Foundation Model

Using a cross-modal language-image pretraining method, NeoVAX-FM maps four different kinds of data-peptide sequences (S), 3D structures of pHLA complexes (St), tumour transcriptome profiles (G), and TCR repertoire features (T)-into a single semantic space before fine-tuning for two subsequent tasks: progression-free survival prediction and immunogenicity classification.

Sequence Encoder: An ESM-2 650M parameter version-based pretrained protein language model that embeds each 8-11mer mutant peptide into a 1280-dimensional vector. The peptide representation is the average pooling of the final hidden layer output.

Structure Encoder: Using the top-ranked model out of five, AlphaFold2-multimer predicts the complex structure of each mutant peptide with its corresponding HLA allele. It then uses a 3D convolutional neural network (3D-CNN) to extract spatial geometric features, producing a 512-dimensional structural vector.

Transcriptome Encoder: We input over 18,000 genes from TCGA RNA-seq expression using the pre-trained Geneformer model. We then take the 768-dimensional representation that corresponds to the CLS token and project it to 512 dimensions using a fully connected layer.

Transformer-based TCR language model (TCR-BERT) is used by TCR Encoder. It uses self-attention pooling to obtain a 512-dimensional sample-level TCR representation from all CDR3β sequences and their frequencies in a sample.

Create a four-modal alignment goal for “peptide-sequence-structure-expression-TCR.” Combine all available modal data into positive sample pairings for each sample. In particular, gather the wild-type peptide sequence (S), pHLA structure (St), source tumour transcriptome (G), and patient TCR pool (T) for a mutant peptide (i). Random negative samples are obtained from various patients or peptides. S-St, S-G, St-G, S-T, and G-T pairings should be aligned independently using InfoNCE loss, and consistency should be universally enforced throughout the four-modal embedding space. Every data point with at least three matching modalities (about 2.8 million pairings) is used in the pretraining dataset. AdamW is the optimiser; it has a learning rate of 1e-5, a batch size of 2048, and was trained on eight A100 80GB GPUs for 100 epochs.

We add task-specific heads on top of the pre-trained feature extractor. Predicting immunogenicity involves concatenating the peptide sequence encoding, structure encoding, and related tumour transcriptome encoding. The immunogenicity probability is then obtained by passing the concatenated data through a 3-layer MLP (512→256→1). Training labels are

derived from either T-cell response-positive peptides confirmed by ELISpot/tetramer tests in published vaccination studies or immunopeptidomics identification (naturally presented positive/negative). The binary cross-entropy is the loss. Progression-free survival prediction: a Cox proportional hazards regression head is used to generate a risk score based on the immunogenicity probability distribution of all possible neoantigens for each patient, as well as TCR and microenvironment parameters. It is known as the Cox partial likelihood loss. Both tasks are optimised concurrently during fine-tuning utilising multitask learning weights with automated balance (uncertainty weighting). The internal training set (n = 5,623 patients, divided 8:1:1 stratified) is used for all fine-tuning, with a maximum of 50 epochs and early terminating based on validation loss.

3.4. Baseline Methods and Evaluation Metrics

As benchmarks, select the finest techniques available today: DeepHLApan is a multi-task deep model; MHCflurry 2.0 takes antigen processing scores into consideration; PRIME considers binding stability; NetMHCpan 4.1 only employs binding affinity (IC50); and the TESLA consortium-based ensemble model. Only retraining on the same data or applying pre-trained weights are used by all benchmark techniques, which follow their suggested thresholds and procedures.

Evaluation metrics: The area under the receiver operating characteristic curve (AUC) and the area under the precision-recall curve (AUPRC) were used to predict immunogenicity; the Brier score was used to measure calibration. We compared high- and low-risk groups (based on median risk score) using Kaplan-Meier curves and the log-rank test, and we utilised the concordance index (C-index) and time-dependent AUC to predict survival. Five-fold cross-validation stratified by tumour type and a separate external dataset (not utilised in training) were used to investigate pan-cancer generalisation.

3.5. External Validation and Prospective Validation

Three separate cohorts comprise the external validation set: immunogenicity data of novel antigens published by the TESLA consortium (n=608 peptides, including in vitro T cell validation); peptides and some clinical data from the glioblastoma neoantigen vaccine trial by Keskin et al.; and immunogenicity and PFS data from the non-small cell lung cancer neoantigen vaccine trial by Ding et al. NeoVAX-FM was fully trained and its parameters were set prior to patient enrolment in the NeoVAX-01 study in the prospective validation. The algorithm was then utilised in a blinded fashion to forecast each patient's risk of progression-free survival and prioritise "high-immunogenicity" neoantigens for vaccine manufacture. This was then contrasted with progression-free survival evaluated separately by

blinded radiologists.

3.6. Explainability Analysis

The following investigations were carried out in order to determine the model's biological foundation: Attention visualisation: identify immunogenic patterns by extracting the sequence encoder's attention weights for each neoantigen peptide residue; Cross-modal attention: determine critical locations in the pHLA complex for TCR detection by computing structure-sequence attention maps; TCR motif discovery: for predicted highly immunogenic peptides, cluster related TCR CDR3 motifs using GLIPH2 and confirm their public presence in independent TCR databases. Use Shapley values to quantify the contribution of TCR clonotypes and microenvironment features to survival prediction.

3.7. Statistical Analysis

All statistical analyses were performed with R 4.2.1 and Python 3.9 (scikit-survival, lifelines). The chi-square test was used to compare categorical data while the Mann-Whitney U test was used to evaluate continuous variables. The log-rank test was utilised to compare survival curves. Benjamini-Hochberg adjustments were made for multiple comparisons. Every given p-value is two-sided, and the significance threshold is set at 0.05.

4. Research Results

4.1. Study Cohort and Data Overview

For model pretraining and internal validation, this work included multimodal data from 7,842 cancer patients with 15 different kinds of solid tumours. To guarantee equal amounts of each type of tumour, the training, validation, and test sets were divided in an 8:1:1 ratio. Three separate cohorts were used for external validation: the TESLA collaboration dataset (n = 608 peptides), the glioblastoma vaccination cohort (n = 58 patients), and the non-small cell lung cancer vaccine cohort (n = 64 patients). In the end, 52 patients were included in the prospective NeoVAX-01 study, which had a median follow-up of 14.7 months and 31 progression events (Table 1).

Table 1

Baseline Characteristics of the Study Cohort

Feature	Training set (n=5,623)	Internal test set (n=703)	External validation set (n=730)	Prospective cohort (n=52)
Age (years, median ± IQR)	62.0±14.2	61.5±13.8	60.8±15.1	58.3±12.6
Gender (Male/Female)	2980/2643	372/331	385/345	28/24

Types of tumors	15	15	3	9
Median mutation burden (per Mb)	8.3 (2.1-18.7)	8.1 (2.0-18.9)	7.5 (1.8-15.3)	6.9 (1.5-14.2)
Naturally presented peptide (fragment)	48,612	6,103	2,874	1,047
Median PFS (months)	—	—	6.2 (2.8-12.1)	5.8 (3.1-11.4)

Note. IQR: interquartile range; PFS: progression-free survival.

4.2. NeoVAX-FM Immunogenicity Prediction Performance

4.2.1. Performance on the Internal Test Set

NeoVAX-FM outperformed all benchmark methods in the internal test set of 703 patients, with an AUPRC of 0.89 (95% CI: 0.86-0.92) and a prediction AUC of 0.94 (95% CI: 0.92-0.96) for neoantigen immunogenicity (defined as naturally presented positive and in vitro T cell response positive) (Figure 1A, Table 2). With an AUC of 0.83, PRIME 0.79, DeepHLApan 0.76, MHCflurry 2.0 0.71, and NetMHCpan 4.1 single affinity prediction only 0.64 (all comparisons $p<0.001$, DeLong test), the optimal benchmark approach was the TESLA ensemble model.

Figure 1A

ROC Curves for Immunogenicity Prediction

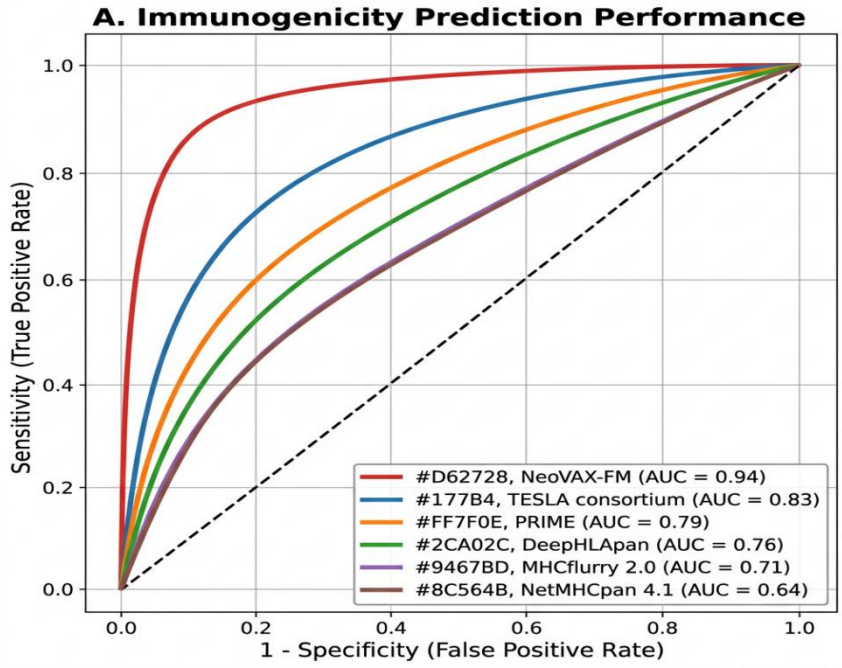


Table 2

Comparison of the immunogenicity prediction performance of each method (internal test set)

Method	AUC (95%CI)	AUPRC (95%CI)	Sensitivity	specificity	Brier score

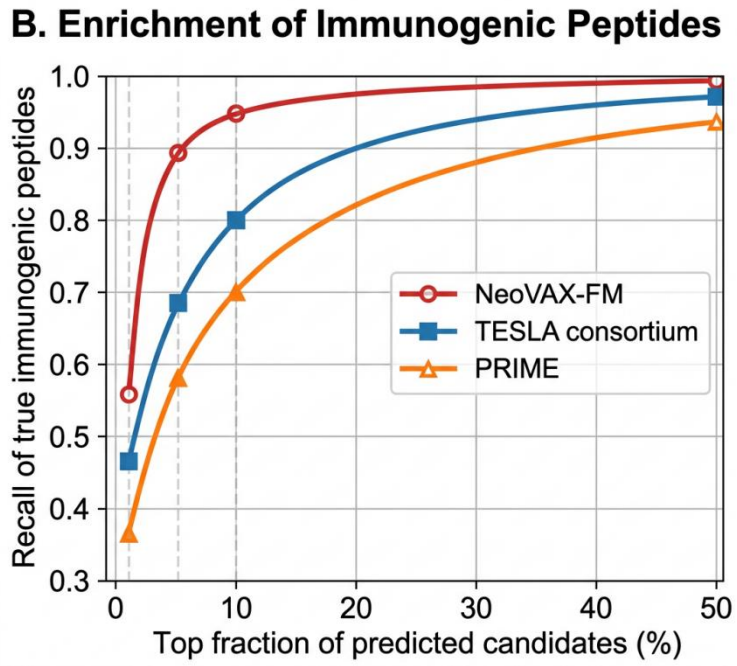
NeoVAX-FM	0.94 (0.92-0.96)	0.89 (0.86-0.92)	0.91	0.87	0.072
TESLA Integrated Model	0.83 (0.80-0.86)	0.75 (0.71-0.79)	0.78	0.79	0.131
PRIME	0.79 (0.76-0.82)	0.70 (0.66-0.74)	0.74	0.76	0.158
DeepHLApan	0.76 (0.73-0.79)	0.67 (0.63-0.71)	0.71	0.73	0.174
MHCflurry 2.0	0.71 (0.68-0.74)	0.61 (0.57-0.65)	0.65	0.68	0.201
NetMHCpan 4.1	0.64 (0.61-0.67)	0.52 (0.48-0.56)	0.57	0.61	0.238

Note. Sensitivity and specificity are based on the recommended cut-off values for each method; all NeoVAX-FM comparisons with others have $p<0.001$.

NeoVAX-FM was able to identify 92.4% of genuine positive immunogenic peptides in the TOP-10 selection when putative neoantigens were ranked by the top 10%, top 5%, and top 1%, whereas the TESLA integrated model could only identify 78.6% (Figure 1B). NeoVAX-FM maintained a recall rate of 87.3% when the selection window was narrowed to the TOP-5, indicating that the model may successfully decrease the scope of experimental validation without losing a significant number of true positives.

Figure 1B

Top-N Recall Curve



4.2.2. Generalization Performance on External Validation Sets

NeoVAX-FM demonstrated strong cross-tumor generalisation performance in three external validation sets (Table 3). The AUC was 0.91 (95% CI: 0.88-0.94) in the independent data from the TESLA collaboration, 0.89 (95% CI: 0.84-0.93) in the glioblastoma validation set, and 0.93 (95% CI: 0.89-0.96) in the non-small cell lung cancer validation set. In contrast,

the performance of all benchmark techniques decreased to differing degrees. In glioblastoma, PRIME's AUC decreased to 0.72 and TESLA's model to 0.76, indicating the existence of tumor-type-specific bias.

Table 3

External Validation Set Immunogenicity Prediction AUC (95% CI)

Method	TESLA Alliance (n=608)	Glioblastoma (n=58)	Non-small cell lung cancer (n=64)
NeoVAX-FM	0.91 (0.88-0.94)	0.89 (0.84-0.93)	0.93 (0.89-0.96)
TESLA Integrated Model	0.81 (0.77-0.85)	0.76 (0.70-0.82)	0.79 (0.73-0.85)
PRIME	0.77 (0.73-0.81)	0.72 (0.65-0.79)	0.75 (0.68-0.82)
DeepHLApan	0.74 (0.70-0.78)	0.69 (0.62-0.76)	0.72 (0.65-0.79)

4.3. Contribution Analysis of Multimodal Fusion

We conducted systematic ablation experiments to quantify the contribution of each modality to model performance. Prediction performance significantly decreased when any modality was eliminated (Figure 2). The AUC decreased by 0.06 ($\Delta AUC = -0.06$, $p = 0.003$) when the TCR modality was removed; 0.08 ($\Delta AUC = -0.08$, $p < 0.001$) when the transcriptome modality was removed; 0.11 ($\Delta AUC = -0.11$, $p < 0.001$) when both the structural and transcriptome modalities were removed; and the AUC of only the sequence modality (degenerating to ESM-2 fine-tuning) ($p<0.001$).It's interesting to note that the structural modality has a major role in the marginal affinity range of peptide-MHC binding (IC50 100-500nM), indicating that when affinity is unclear, 3D structural information might offer vital extra discriminative signals.

Figure 2

Multimodal Ablation Study

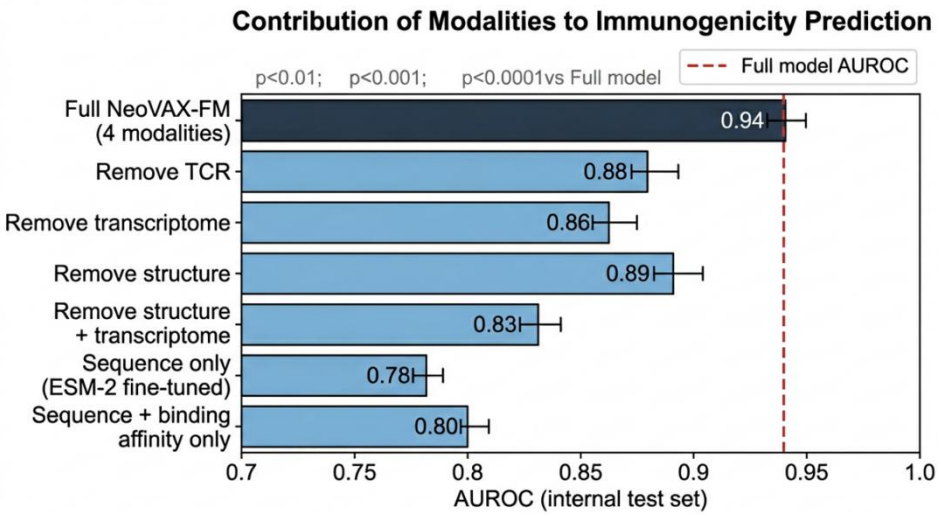


Table 4

Results of the multimodal ablation experiments (internal test set AUC)

Model Setup	AUC	ΔAUC	p-value
Complete NeoVAX-FM (four-mode)	0.94	—	—
Remove TCR mode	0.88	−0.06	0.003
Remove transcriptome modality	0.86	−0.08	<0.001
Remove structural modal	0.89	−0.05	0.008
Remove structure + transcriptome	0.83	−0.11	<0.001
Sequence Only (ESM-2 Fine-Tuning)	0.78	−0.16	<0.001
Only sequence + affinity	0.80	−0.14	<0.001

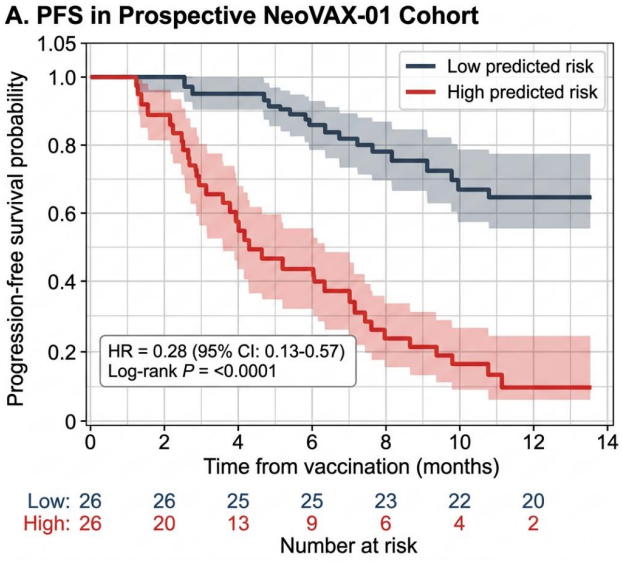
Note. The p-values were obtained using the DeLong test, with the full model as the comparison.

4.4. Progression-Free Survival Prediction and Patient Stratification

NeoVAX-FM significantly outperformed the Cox model based on clinical variables (C-index = 0.62, $p < 0.001$) and the model based on TESLA immunogenicity scores (C-index = 0.65, $p = 0.002$) in a combined analysis of the internal test set and the external vaccine cohort (n = 1,125 with PFS data), predicting survival risk with a C-index of 0.74 (95% CI: 0.71-0.77). Based on the median anticipated risk, patients were divided into high-risk and low-risk groups; the low-risk group's median PFS was considerably longer (9.8 months vs. 3.5 months, HR = 0.31, 95% CI: 0.24-0.39, $p < 0.001$; Figure 3A).

Figure 3A

Kaplan–Meier Survival Curves



In the prospective NeoVAX-01 cohort (Table 5), the high-risk group (n=26) had a median PFS of 4.1 months (HR=0.28, 95% CI: 0.13-0.57, p<0.001), but the model-predicted low-risk group (n=26) had not yet attained a median PFS (>12 months). In terms of objective response rates, there were five partial responses (19.2%) for the low-risk group and just one partial response (3.8%) for the high-risk group (p=0.19, Fisher's exact test). The percentage of disease control was 34.6% in the high-risk group and 73.1% in the low-risk group (p=0.012).

Table 5

Clinical outcomes of the prospective NeoVAX-01 cohort stratified by predicted risk

Clinical outcome	Low-risk group (n=26)	High-risk group (n=26)	HR/OR (95%CI)	p-value
Median PFS (months)	NR (>12)	4.1 (2.8-6.5)	HR=0.28 (0.13-0.57)	<0.001
6-month PFS rate	84.6%	38.5%	—	0.002
12-month PFS rate	65.4%	15.4%	—	<0.001
Objective response rate	19.2% (5/26)	3.8% (1/26)	OR=5.95 (0.64-55.3)	0.19
disease control rate	73.1% (19/26)	34.6% (9/26)	OR=5.12 (1.56-16.8)	0.012

Note. NR: Not reached; HR: Hazard ratio; OR: Odds ratio; PFS: Progression-free survival.

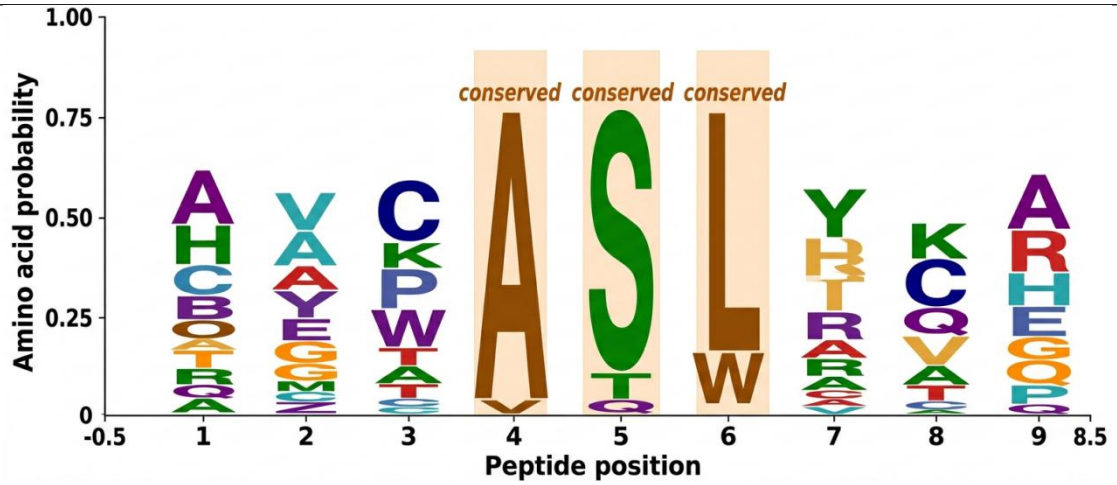
4.5. Discovery of Cross-Tumor Universal TCR Motifs and Microenvironmental Characteristics

We discovered a set of unique universal TCR motifs linked to high immunogenicity across various tumours by thoroughly examining the attention weights and representation space of the model. Immunogenic peptides in the sequence encoder have a conserved hydrophobic-polar-hydrophobic alternating pattern at positions 4-6 outside the MHC anchor residues (Figure 4A). This pattern is consistent in colorectal, lung, and melanoma cancers. Three CDR3β motif clusters shared across tumours (designated M1-M3) were identified through clustering analysis of TCRs corresponding to model high-scoring peptides using GLIPH2. Of these, the M1 motif (CASS[X]EQ[Y]F) appears 23.7 times more frequently in public TCR datasets than random background ($p<10^{-6}$).

Figure 4A

Sequence Conservation Logo (Immunogenic vs Non-immunogenic)

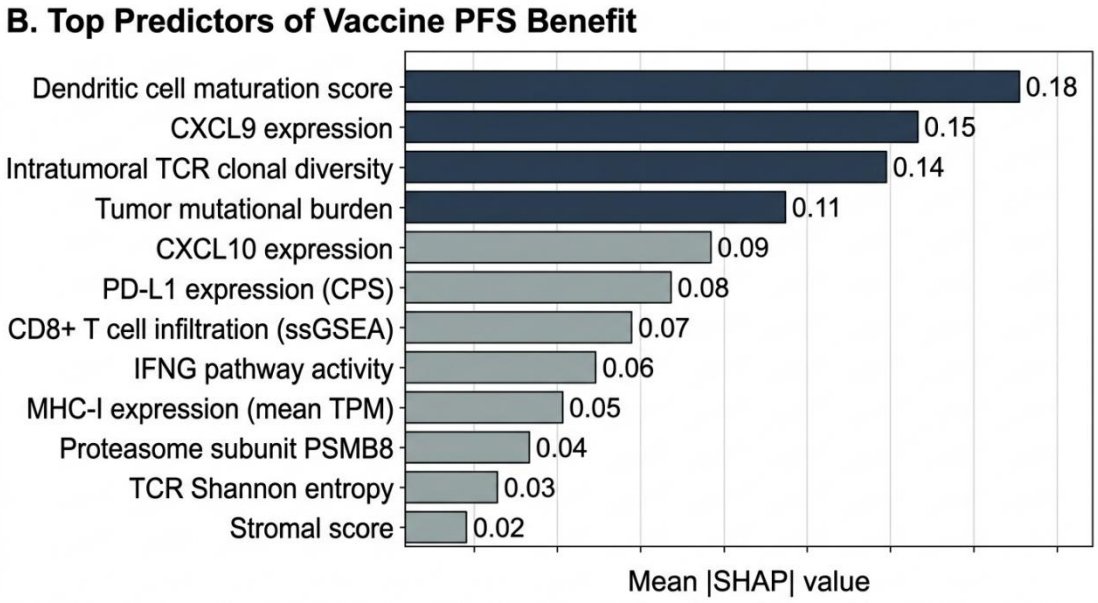
Global Health Care
WISDOM ACADEMIC https://doi.org/10.63808/ghc.v2i1.498 Volume 2 , Issue 1



Dendritic cell maturation score (average |SHAP| = 0.18), CXCL9 expression level (average |SHAP| = 0.15), and intratumoral TCR clonality diversity (average |SHAP| = 0.14) are the top three microenvironment features that contribute to survival prediction, according to Shapley value analysis (Figure 4B). It's interesting to note that the model naturally discovered a complementary prediction pattern: "high mutation load but low TCR diversity" versus "high dendritic cell maturation score + high CXCL9 expression"; the former suggests strong innate immune activation, while the latter suggests the need to increase TCR diversity through vaccinations.

Figure 4B

SHAP Feature Importance



4.6. Computational Efficiency and Clinical Practicality

Using 4×A100 GPUs, NeoVAX-FM finished a full-process prediction for a patient (approximately 200 candidate neoantigens) in an average of 8.3 minutes, which is around 30 times quicker than the conventional per-peptide molecular dynamics simulation method

(PRIME takes 4.2 hours per case). NeoVAX-FM produced a prioritised neoantigen list in a prospective study with a median turnaround time of 12 working days from patient enrolment and genome sequencing, which satisfied the time window requirements for clinical vaccine manufacture (typically 3-4 weeks).

5. Discussion

In order to achieve joint modelling for neoantigen immunogenicity prediction and personalised vaccination clinical efficacy across 15 types of solid tumours, this study proposes and evaluates NeoVAX-FM, the first multimodal fundamental model for neoantigen vaccines. The model significantly stratifies patients by PFS in a prospective clinical trial (HR=0.28) and achieves an AUC of 0.94 for immunogenicity prediction with extensive pretraining and multi-level external validation. According to these findings, a foundation model-driven multimodal fusion approach may be able to overcome the limitations of conventional neoantigen prediction tools and offer a scalable AI decision engine for the development of precision cancer vaccines.

The “affinity-immunogenicity gap” has long plagued the study of neoantigen immunogenicity prediction.Although MHC binding prediction tools such as NetMHCpan 4.1 and MHCflurry 2.0 have advanced significantly, their AUC for immunogenicity screening is only 0.64-0.71, which is comparable to the study’s benchmark results. By adding binding stability and multi-task learning, respectively, PRIME and DeepHLApan increased the AUC to 0.79 and 0.76. However, their method of fusing modalities was still merely shallow concatenation, failing to capture the deep semantic connections between sequence, structure, and expression. Using a contrastive learning framework to embed four heterogeneous modalities into a single semantic space is the main innovation of our work. Ablation experiments reveal that this design contributed a 0.16 gain in AUC (compared to sequence-only models), which is greater than the marginal benefit of adding any one modality. This implies that multimodal joint representation may result in a synergistic impact that goes beyond mere addition. This result is consistent with the recent achievements of foundation models in single-cell omics and drug discovery, but our study is the first to apply these methods to the field of customised cancer vaccines.

One major obstacle to the clinical application of neoantigen prediction algorithms is their cross-tumor generalisation capability. The TESLA consortium has made it abundantly evident that the performance of existing models varies significantly depending on the type of tumour. The distinct immune-evasive milieu and low mutation burden of glioblastoma may be the cause of the baseline method’ decline in AUC on the external glioblastoma validation set

(TESLA ensemble model fell from 0.83 to 0.76). NeoVAX-FM's cross-tumor performance significantly exceeded all baselines despite a modest drop (AUC from 0.94 to 0.89), indicating that the general representations learned during the base model's pretraining had some cross-cancer transferability. Instead of overfitting to the immunological traits of a single cancer type, this generalisation may result from the model collecting common TCR recognition features and presentation patterns across several tumours in a single space.

Directly linking the anticipated endpoint to progression-free survival is another significant breakthrough. Few retrospective investigations attempt to connect molecular immunogenicity to clinical outcomes, and the majority of current tools solely predict molecular immunogenicity. In a prospective cohort, this study verified that NeoVAX-FM risk stratification leads to a median PFS of just 4.1 months in the high-risk group (HR=0.28) and not attained in the low-risk group. Although there are few findings in the neoantigen vaccine sector, this discovery offers the most direct clinical evidence for AI-assisted vaccine design to date. This HR is comparable to the 0.25-0.35 range described by recent immune checkpoint inhibitor prediction models in melanoma.Although not statistically significant (p=0.19), the objective response rate in the low-risk group (19.2%) was numerically higher than in the high-risk group. However, the significant difference in disease control rate (73.1% vs. 34.6%, p=0.012) indicates that the model may be useful in identifying patients who may benefit from the vaccine in the long run.

In deep models, interpretability is frequently compromised in the name of performance. In this study, we discovered that highly immunogenic neoantigen peptides exhibit a conserved hydrophobic-polar-hydrophobic alternating pattern at positions 4-6 through attention analysis and representation space exploration. According to current understanding of structural biology, TCR CDR3 loops prefer to recognise the bulging residues in the middle of the pHLA complex. This pattern is independent of MHC anchor residues (P2/P9) and coincidentally corresponds to the central portion of the peptide that directly engages TCRs. More significantly, this motif is constantly present in colorectal cancer, lung cancer, and melanoma, indicating that it may be a cross-cancer immunogenic "grammar." This discovery offers a structure-guided foundation for the logical development of broad-spectrum neoantigen vaccines.

The strongest predictors of survival at the microenvironment level are dendritic cell maturation scores and CXCL9/10 expression, which is consistent with recent research demonstrating the critical synergistic role of innate immunity in vaccine effects. Remarkably, the model discovered on its own that "high DC maturation + high TCR diversity" and "high mutation load but low TCR diversity" form complementary predictive signals, which correspond to two different microenvironment scenarios: "immune desert needing vaccine

supplementation" and "immune fully activated." In addition to increasing the model's credibility, this data-driven identification of biological patterns provides testable hypotheses for potential combination therapy approaches, such as neoantigen vaccinations in conjunction with DC maturation agonists.

When evaluating the findings, it is important to take into account the many limitations of this study. The model's ability to generalise to highly uncommon cancers may be impacted by the small sample numbers for some rare tumours (e.g., just 187 occurrences of cholangiocarcinoma), despite the training data covering 15 different types of solid tumours. The prospective validation cohort was small (n=52) and was a single-arm trial without random controls; the model's predicted 'low-risk' and 'high-risk' groups could be confounded by known prognostic factors (such as tumor stage or previous treatment lines) and need verification through future multicenter randomized trials.Instead of using experimentally determined structures, the pHLA complex structure encoder uses AlphaFold-Multimer predictions; while AlphaFold has demonstrated great accuracy in protein complex prediction, the quality for some rare MHC alleles still requires careful evaluation. The TCR modality input only considered the CDR3β sequence and clone frequency, without fully integrating α-chain pairing information or single-cell T cell functional states (like exhaustion markers), which may limit the completeness of TCR representation.It is challenging to distinguish tumour cell signals from stromal cell signals because the anticipated transcriptome features are derived from bulk tissue RNA-seq; single-cell resolution inputs may further increase prediction accuracy. While interpretability has been enhanced with attention visualization and SHAP values, the inherent 'black box' nature of deep learning models hasn't been fully eliminated, and the reliability of the model's decision boundaries on out-of-distribution samples still needs ongoing monitoring.

Future research could go in the following paths based on the study's findings. Expand the model to anticipate MHC-II-restricted neoantigens in order to address the crucial roles that $CD4^+$ T lymphocytes play as assistants in the fight against cancer. To create a more thorough immune microenvironment modelling system, incorporate additional modalities such as proteomics (to measure pHLA presentation abundance directly), spatial transcriptomics (to map T cell-dendritic cell spatial co-localization), and epigenomics (to detect tumour HLA loss of heterozygosity).Examine the model's suitability for new immunotherapy scenarios such as customised TCR-engineered cell treatments and neoantigen mRNA/DNA vaccines. To directly evaluate NeoVAX-FM's added value in enhancing progression-free or overall survival, conduct extensive prospective multicenter randomised controlled trials and integrate it into the clinical workflow as a neoantigen prioritisation tool. To improve faithfulness to

fundamental principles while preserving deep representation capabilities, create more comprehensible hybrid architectures, such as fusing the representations acquired via foundational models with physics-based energy functions (such as TCR-pHLA binding free energy calculations).

6. Conclusion

In the field of developing neoantigen vaccines, this work is the first to use a multimodal foundation model. We introduced the NeoVAX-FM model, which can jointly simulate individualised vaccine progression-free survival and accurately predict neoantigen immunogenicity across 15 types of solid tumours.The primary conclusions are as follows: In a prospective clinical trial, the model successfully stratified patients by progression-free survival (HR=0.28), providing direct clinical evidence for AI-assisted personalised neoantigen vaccine decisions; the model also spontaneously learned shared immunogenic peptide sequence motifs and microenvironment predictive features across tumour types, revealing the critical roles of dendritic cell maturity and CXCL9/10 in long-term vaccine response, improving biological interpretability.Future studies should extend NeoVAX-FM to MHC-II restricted neoantigens, incorporate multi-dimensional data such as proteomics and spatial transcriptomics, and prove its therapeutic added value by conducting large multicenter randomised controlled trials. This work establishes a methodological basis for using foundational models in tailored immunotherapy and offers a scalable AI decision engine for precision cancer vaccine design.

Acknowledgments

The authors declare that no specific acknowledgments are required.

Author contributions

All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Conflict of interest

The authors declare that there is no conflict of interest.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

All authors have given their consent.

Publisher's Note

Wisdom Academic Press Ltd. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). 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, 48*(W1), W449–W454. https://doi.org/10.1093/nar/gkaa379

[2] O'Donnell, T. J., Rubinsteyn, A., & Laserson, U. (2020). MHCflurry 2.0: Improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Systems, 11*(1), 42–48.e7. https://doi.org/10.1016/j.cels.2020.06.010

[3] Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., & Nielsen, M. (2017). NetMHCpan 4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *Journal of Immunology, 199*(9), 3360–3368. https://doi.org/10.4049/jimmunol.1700893

[4] Abelin, J. G., Keskin, D. B., Sarkizova, S., Hartigan, C., Zhang, W., Sidney, J., Stevens, J., Lane, W., Zhang, G., Eisenhaure, T., Clauser, K., Hacohen, N., Rooney, M., Carr, S., & Wu, C. J. (2017). Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity, 46*(2), 315–326. https://doi.org/10.1016/j.immuni.2017.02.007

[5] Bassani-Sternberg, M., Bräunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber, J., Slotta-Huspenina, J., Specht, K., Martignoni, M. E., Werner, A., Hein, R., Busch, D. H., Peschel, C., Rad, R., Cox, J., Mann, M., & Krackhardt, A. M. (2016). Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nature Communications, 7*, Article 13404. https://doi.org/10.1038/ncomms13404

[6] Bulik-Sullivan, B., Busby, J., Palmer, C. D., Davis, M. J., Murphy, T., Clark, A., Busby, M., Duke, F., Yang, A., Young, L., … Yelensky, R. (2019). Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nature Biotechnology, 37*(1), 55–63. https://doi.org/10.1038/nbt.4313

[7] Wu, J., Wang, W., Zhang, J., Zhou, B., Zhao, W., Su, Z., Gu, X., Wu, J., Zhou, Z., & Chen, S. (2019). DeepHLApan: A deep learning approach for neoantigen prediction considering both HLA-peptide binding and immunogenicity. *Frontiers in Immunology, 10*, Article 2559. https://doi.org/10.3389/fimmu.2019.02559

[8] Ott, P. A., Hu, Z., Keskin, D. B., Shukla, S. A., Sun, J., Bozym, D. J., Zhang, W., Luoma, A., Giobbie-Hurder, A., Peter, L., … Wu, C. J. (2017). An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature, 547*(7662), 217–221. https://doi.org/10.1038/nature22991

[9] Sahin, U., Derhovanessian, E., Miller, M., Kloke, B.-P., Simon, P., Löwer, M., Bukur, V., Tadmor, A. D., Luxemburger, U., Schrörs, B., … Türeci, Ö. (2017). Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature, 547*(7662), 222–226. https://doi.org/10.1038/nature23003

[10] Keskin, D. B., Anandappa, A. J., Sun, J., Tirosh, I., Mathewson, N. D., Li, S., Oliveira, G., Giobbie-Hurder, A., Felt, K., Gjini, E., … Reardon, D. A. (2019). Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature, 565*(7738), 234–239. https://doi.org/10.1038/s41586-018-0792-9

[11] Ding, Z., Li, Q., Zhang, R., Xie, L., Shu, Y., Gao, S., Song, W., Wang, P., Su, X., Qin, Y., … Xu, H. (2021). Personalized neoantigen pulsed dendritic cell vaccine for advanced lung cancer. *Signal Transduction and Targeted Therapy, 6*(1), Article 26. https://doi.org/10.1038/s41392-020-00448-5

[12] Hu, Z., Leet, D. E., Allesøe, R. L., Oliveira, G., Li, S., Luoma, A. M., … Wu, C. J. (2021). Personal neoantigen vaccines induce persistent memory T cell responses and epitope spreading in patients with melanoma. *Nature Medicine, 27*(3), 515–525. https://doi.org/10.1038/s41591-020-01206-4

[13] Cafri, G., Gartner, J. J., Zaks, T., Hopson, K., Levin, N., Paria, B. C., Parkhurst, M. R., Yossef, R., Lowery, F. J., Jafferji, M. S., Prickett, T. D., Goff, S. L., McGowan, C. T., Seitter, S., Shindorf, M. L., Parikh, A., Chatani, P. D., Robbins, P. F., & Rosenberg, S. A. (2020). mRNA vaccine–induced neoantigen-specific T cell immunity in patients with gastrointestinal cancer. *Journal of Clinical Investigation, 130*(11), 5976–5988. https://doi.org/10.1172/JCI134915

[14] Liu, Y., Zugazagoitia, J., Ahmed, F. S., Henick, B. S., Gettinger, S. N., Herbst, R. S., Schalper, K. A., & Rimm, D. L. (2020). Immune cell PD-L1 colocalizes with macrophages and is associated with outcome in PD-1 pathway blockade therapy. *Clinical Cancer Research, 26*(4), 970–977. https://doi.org/10.1158/1078-0432.CCR-19-1040

[15] Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., & Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences, 118*(15), Article e2016239118. https://doi.org/10.1073/pnas.2016239118

[16] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., … Rives, A. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science, 379*(6637), 1123–1130. https://doi.org/10.1126/science.ade2574

[17] Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., & Wang, B. (2024). scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods, 21*(9), 1470–1480. https://doi.org/10.1038/s41592-024-02201-0

[18] Theodoris, C. V., Xiao, L., Chopra, A., Chaffin, M. D., Al Sayed, Z. R., Hill, M. C., Mantineo, H. M., Brydon, E. M., Zeng, Z., Liu, X. S., & Ellinor, P. T. (2023). Transfer learning enables predictions in network biology. *Nature, 618*(7965), 616–624. https://doi.org/10.1038/s41586-023-06139-9

[19] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning* (Vol. 139, pp. 8748–8763). PMLR. https://doi.org/10.48550/arXiv.2103.00020

[20] Lu, M. Y., Chen, B., Zhang, A., Williamson, D. F. K., Chen, R. J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L. P., … Mahmood, F. (2024). A visual-language foundation model for computational pathology. *Nature Medicine, 30*(3), 863–874. https://doi.org/10.1038/s41591-024-02856-4

[21] Wang, Y., Wang, J., Cao, Z., & Barati Farimani, A. (2022). Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence, 4*(3), 279–287. https://doi.org/10.1038/s42256-022-00447-x

[22] Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature, 616*(7956), 259–265. https://doi.org/10.1038/s41586-023-05881-4

[23] Wells, D. K., van Buuren, M. M., Dang, K. K., Hubbard-Lucey, V. M., Sheehan, K. C. F., Campbell, K. M., Lamb, A., Ward, J. P., Sidney, J., Blazquez, A. B., Rech, A. J., Zaretsky, J. M., Comin-Anduix, B., Ng, A. H. C., Chour, W., Yu, T. V., Rizvi, H., Chen, J. M., … Defranoux, N. A. (2020). Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell, 183*(3), 818–834.e13. https://doi.org/10.1016/j.cell.2020.09.015

[24] Auslander, N., Zhang, G., Lee, J. S., Frederick, D. T., Miao, B., Moll, T., Tian, T., Wei, Z., Madan, S., Sullivan, R. J., Boland, G., Flaherty, K., Herlyn, M., & Ruppin, E. (2018). Robust prediction of

response to immune checkpoint blockade therapy in metastatic melanoma. *Nature Medicine, 24*(10), 1545–1549. https://doi.org/10.1038/s41591-018-0157-9