A Deep Learning–Driven Framework Integrating Organoid-Based Functional Validation Identifies Universal Neoantigens from Recurrent Glioma Mutations

Chen Wang[1], Ting Sun[2], Yufei He[2], Mingchen Yu[3], Chang-Qing Pan[1], Huimin Hu[3], Yishuo Sun[1], Di Wang[1], Zhongliang Cui[1], Jiazheng Zhang[1], You Zhai[1], Zhongfang Shi[3], Ziwei Li[1], Menghui Xu[3], Young Taek Oh[4], Tao Jiang[1,3,5,6,7,8], Zhiyuan Xu[2], Guanzhang Li[1,3,8], Jing Zhang[2], and Wei Zhang[1,3,5,6,7,8]

ABSTRACT

Glioblastoma (GBM) is the most common malignant intracranial tumor in adults, with a median survival of only 16 to 20 months. Neoantigen therapy has shown advantages in the treatment of GBM, as it improves the immunosuppressive microenvironment within the tumor. However, the identification of truly immunogenic neoantigens remains a major challenge. Current computational prediction tools primarily focus on antigen presentation, whereas algorithms that incorporate T-cell immunogenicity features remain limited. Furthermore, standard validation methods, such as enzyme-linked immunospot (ELISpot) assays, lack physiologic relevance and do not fully recapitulate the tumor microenvironment. In this study, we developed a neoantigen prediction algorithm, TCRscore, based on publicly available datasets by integrating human leukocyte antigen binding and T-cell receptor (TCR) recognition features. Twenty-one patient-derived GBM organoid models were established from isocitrate dehydrogenase wild-type tumors to validate the performance of the algorithm. Predicted neoantigens were evaluated using ELISpot assays, flow cytometry, and *in vitro* killing assays based on organoid–T cell coculture systems. TCRscore outperformed six existing tools in predicting immunogenic neoepitopes. The organoid models retained the key histologic and transcriptomic features of parental tumors and provided an effective platform for functional validation. Coculture assays confirmed that neoantigen-specific T cells could induce targeted killing in GBM organoids. In particular, the analysis identified that the recurrent $PIK3R1^{G376R}$ mutation contributed to a potential shared neoantigen in GBM. Overall, by integrating TCRscore with organoid-based validation, this study provides a high-fidelity, high-quality GBM neoantigen database with significantly enhanced prediction accuracy.

Significance: A clinically impactful framework that integrates a TCR-aware AI algorithm with glioblastoma organoids enables accurate neoantigen prediction and validation, advancing both personalized and population-level immunotherapy strategies.

Introduction

GBM, as defined by the 2021 World Health Organization (WHO) Central Nervous System (CNS) Tumor Classification, refers specifically to glioblastoma, isocitrate dehydrogenase (IDH) wild type, and is one of the most aggressive and deadly forms of brain cancer, characterized by rapid growth, high heterogeneity, and resistance to conventional therapies such as surgery, radiation, and chemotherapy (1). Despite significant advances in cancer treatment, the prognosis for patients with GBM remains poor, with median survival times <2 years from diagnosis (2). This stark reality highlights the urgent need for novel therapeutic approaches that effectively address the unique challenges presented by GBM, including its infiltrative growth pattern, ability to adapt and evolve, and its highly immunosuppressive microenvironment (3).

Neoantigens, tumor-specific peptides arising from somatic mutations, offer a promising opportunity to apply personalized cancer immunotherapies (4). Unlike conventional antigens, neoantigens are unique to each patient's tumor, making them ideal targets for precision medicine. By leveraging the ability of the immune system to recognize these unique antigens, neoantigen-based treatments elicit robust and targeted immune responses against GBM cells while minimizing damage to healthy tissues (5). Notably, two

[1]Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, PR China. [2]Laboratory for Biomechanics and Mechanobiology of Ministry of Education, Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, School of Engineering Medicine, Beihang University, Beijing, PR China. [3]Beijing Neurosurgical Institute, Capital Medical University, Beijing, PR China. [4]Department of Biological Sciences, Kangwon National University, Chuncheon-Si, Republic of Korea. [5]Brain Tumor Center, Beijing Institute of Brain Disorders, Capital Medical University, Beijing, PR China. [6]Beijing Engineering Research Center of Targeted Drugs and Cell Therapy for CNS Tumors, Beijing, PR China. [7]China National Clinical Research Center for Neurological Diseases, Beijing, PR China. [8]Chinese Glioma Genome Atlas Network (CGGA) and Asian Glioma Genome Atlas Network (AGGA), Beijing, PR China.

C. Wang and T. Sun contributed equally to this article.

Corresponding Authors: Wei Zhang, Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, No. 119 South Fourth Ring Road, West Fengtai District, Beijing 100070, PR China. E-mail: zhangwei_vincent@mail.ccmu.edu.cn; Guanzhang Li, liguanzhang122@163.com; Jing Zhang, Laboratory for Biomechanics and Mechanobiology of Ministry of Education, Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, No. 37 Xueyuan Road, Haidian District, Beijing 100191, PR China. E-mail: jz2716@buaa.edu.cn; and Zhiyuan Xu, 1213495110@qq.com

Cancer Res 2026;86:3074–91

doi: 10.1158/0008-5472.CAN-25-2679

landmark clinical trials by Hilf and colleagues and Keskin and colleagues (6, 7) demonstrated the feasibility and therapeutic potential of personalized neoantigen vaccines in patients with GBM, providing critical proof of concept for neoantigen-based immunotherapy. Neoantigen-based therapies provide a means to specifically train the patient's immune system to recognize and attack tumor cells without affecting normal cells, potentially offering a safer and more effective treatment option (8).

Recent advances in sequencing technologies and computational biology have facilitated the identification and validation of neoantigens, bringing personalized immunotherapy closer to clinical reality for patients with GBM (9). However, one of the most fundamental and still-unanswered questions in neoantigen and antigen biology is the lack of understanding of why certain neoantigens fail to elicit T-cell responses (10). Recently, substantial efforts have been directed toward ascertaining which peptides are presented on major histocompatibility complex class I (MHC-I) molecules (11), facilitating the development of highly accurate predictors for peptide–MHC-I binding affinities (12). Despite precisely predicting the peptide–MHCs (pMHC) generated, determining what differentiates epitopes from nonepitopes remains an unresolved issue. The capacity to establish a connection between pMHCs and T-cell receptor (TCR) sequences is fundamental for tracking the complex interactions between the immune system and tumors, along with optimizing the design and execution of diverse immunotherapeutic strategies.

To bridge this knowledge gap and improve the precision of existing pMHC-binding prediction, recent *in silico* algorithms have begun to incorporate T cell–related features into their models (13), including TCR-facing residues, peptide hydrophobicity, and structural compatibility with known TCR motifs (14). Deep learning models can extract features from noisy raw data by learning high-level representations. They demonstrate great potential in discerning intricate, nonlinear relationships among multiple immunologic variables (15). Nevertheless, current models face substantial constraints in incorporating amino acid descriptors—encompassing the quantitative values of topologic, physicochemical, and three-dimensional structural properties—when analyzing the interactions among human leukocyte antigens (HLA), neoantigens, and TCRs. More specifically, they fail to fully leverage the advantages of *Z* descriptors in quantifying intrinsic amino acid characteristics and predicting peptide–protein interactions (16). These limitations lead to their poor capacity to generalize across diverse heterogeneous tumors and patient-specific immune repertoires, necessitating the development of cutting-edge bioinformatics algorithms for neoantigen prediction.

In parallel with computational developments, experimental validation of neoantigen immunogenicity is crucial. *In vitro* models, primarily using enzyme-linked immunospot (ELISpot) technology, are commonly used to evaluate the ability of candidate neoantigens to elicit immune responses. ELISpot assays measure the release of cytokines, such as interferon gamma (IFNγ), by immune cells in response to stimulation with neoantigen peptides (17). Despite their utility, ELISpot models have limitations, such as an inability to fully replicate the complex tumor microenvironment, which includes a diverse array of immune and stromal cells that influence immune responses (18). This may affect the accuracy (ACC) of predicting immune responses in patients, underscoring the need for more advanced preclinical models.

To address this challenge, tumor organoid models have emerged as a promising tool. Derived from patient tumors, organoids maintain the architectural and cellular diversity of the original tissue, including the immune microenvironment (19). By providing a more physiologically relevant environment that preserves the parental tumor's immune context, GBM organoids offer a platform to more accurately assess neoantigen immunogenicity (20). The integration of tumor organoid models with neoantigen peptide screening allows for a more precise evaluation of how these peptides interact with the immune system, including their ability to activate T cells in a context that closely resembles the patient's tumor (21). These models advance our understanding of immunogenicity in a patient-specific context and facilitate the development of more effective neoantigen-based therapies for GBM.

In this study, we present an integrated framework that combines a *Z* descriptor–based T cell–aware deep learning model, TCRscore, with patient-derived GBM organoids for neoantigen discovery and validation. TCRscore was developed based on publicly available datasets and incorporates conventional features, such as MHC binding affinity and gene expression, as well as TCR recognition potential, aiming to better capture the true immunogenicity of candidate peptides. We validated the model using a unique cohort of 21 patients with GBM using patient-derived organoids. Using TCRscore, we prioritized neoantigen candidates for each patient and conducted functional validation. Previous studies have demonstrated that peptides of 9 amino acids in length (9-mer) are the most commonly presented neoantigens by HLA class I molecules (22), which we also confirmed in our study. We therefore focused primarily on 9-mer peptides. Our results showed that TCRscore outperformed several widely used neoantigen prediction tools in both predictive ACC and experimental validation rate. By integrating TCRscore with organoid-based models, we have constructed a high-fidelity, high-quality GBM neoantigen database. In addition, our analysis also identified *PIK3R1*^{G376R} as a recurrent shared mutation in glioma that showed high prevalence across multiple datasets, and this mutation-derived neoantigens strongly predicted binding to common HLA alleles, highlighting its potential value for population-targeted immunotherapies such as shared neoantigen vaccines or TCR-based therapies.

Altogether, our study establishes a robust and generalizable strategy for personalized neoantigen discovery in GBM by integrating computational modeling and organoid validation. This approach advances the precision of neoantigen selection and provides a practical preclinical framework for evaluating immunotherapeutic targets in solid tumors.

Materials and Methods

pMHC and TCR sequence data collection

HLA class I allele sequences, neoantigen peptides, and TCR β-chain CDR3 regions were retrieved from publicly accessible immune repertoire and epitope databases. Precisely, 256; 46,224; 48,210; 33,710; 3,944; and 6,903 TCR sequences (corresponding to TCR β-chain CDR3 regions) were obtained from the TCR3d (tcr3d.ibbr.umd.edu; ref. 23), PIRD Dataset (db.cngb.org/pird/tbadb; ref. 24), VDJdb Dataset (vdjdb.cdr3.net; ref. 25), and McPAS (friedmanlab.weizmann.ac.il/McPAS-TCR/; ref. 26) datasets, as well as those of Joglekar and colleagues (27) and Chen and colleagues (28), respectively. Following the elimination of duplicate sequences, 122,912 unique TCR sequence information records remained. With respect to the antigen-HLA sequence pairs, the NetMHCpan-4.1 training dataset was employed. This dataset comprises 160,581 entries, each annotated with binding affinity values.

Protein sequence coding module

For numerical encoding, we used amino acid *Z* descriptors derived from the principal components of physicochemical properties through partial least squares projection. These descriptors have been employed in predicting cell-penetrating peptides (29) and immunogenic proteins (30), exhibiting considerable potential in predicting immunogenicity. Unlike the fourth and fifth descriptors, which were not clearly derivable, the first three *Z* descriptors corresponded to lipophilicity (z1), volume (z2), and polarity (z3; refs. 31–33). Therefore, we employed the first three *Z* descriptors to represent each amino acid sequence (which may be HLA, neoantigen, or TCR sequences) followed by a *Z* score transformation, resulting in a *Z* score–converted ($3 \times N$) numeric matrix for each AA sequence. For HLA class I, alleles were converted into the canonical NetMHCpan 34-residue pseudosequence, which comprises polymorphic peptide-contacting positions. Peptides were encoded in a 15-mer frame, a choice that fully encompasses common MHC-I ligands (8–11-mers) while accommodating infrequent bulged ligands (12–15-mers) and preventing information loss from clipping. Shorter peptides were zero-padded to 15 residues. TCR CDR3 sequences were zero-padded to a fixed length of 50 residues. This length covered essentially all natural CDR3s and provided headroom for potential extensions (e.g., adding CDR3α) with minimal or no changes to the autoencoder architecture. Therefore, each biological input was transformed into a fixed-size matrix: 34×3 for HLA, 15×3 for peptides, and 50×3 for TCRs.

TCR representation module

To capture latent patterns within TCR sequences, an autoencoder architecture was implemented. The encoder functions to compress the CDR3 matrix into a low-dimensional embedding, and conversely, the decoder undertakes the task of reconstructing the input, thereby guaranteeing the fidelity of the features. Specifically, the TCR CDR3 sequences were transformed into a two-dimensional matrix (50×3) using the *Z* descriptor; it is fed into an autoencoder neural network for further dimensionality reduction and feature extraction. This process aims to obtain more concise features for the TCR CDR3 sequence. The reduced dimensionality features also alleviate the computational load of the neural network model.

The autoencoder neural network is a powerful, unsupervised deep learning model comprising three main components: an encoding (input) part for compressing data, a processing (bottleneck) component for handling the compressed data, and a decoder (output). When data are introduced into the autoencoder, they are first encoded and then compressed to a smaller size. Subsequently, the network is trained on this encoded/compressed data, and a reconstructed copy of the original data is produced as output.

Specifically, the data are initially sent through an encoder comprising 4 one-dimensional convolutional layers. In each convolutional layer, the number of channels is set to 32, 16, 8, and 4, and the size of the one-dimensional convolutional kernels is 3, 3, 2, and 2. Considering that the input data have been standardized using *Z* score normalization, the tanh function is employed as the activation function. A fully connected layer, which serves as the bottleneck layer of the autoencoder, follows the convolutional layers.

To construct the decoder part, the layers preceding the bottleneck layer are reversed. The input of the encoder and the output of the decoder have the same dimensions in the feature matrix. The training effectiveness of the network is evaluated by comparing the similarity between input TCRs and reconstructed TCRs. The most central layer of the network is extracted as the final encoding result, which is stored in matrix format.

pMHC feature extraction module

A neural network model featuring two neural network branch structures was utilized to extract the features of pMHC. The HLA and antigen sequences, both preprocessed by the *Z* descriptor, were input into different branches of the neural network.

The two branches of the neural network share an identical structure, each consisting of 4 one-dimensional convolutional layers and a fully connected layer. For the one-dimensional convolutional layers, the number of channels was set to 32, 16, 4, and 1 successively, with a convolution kernel size of 2 in each layer. The output of the last convolutional layer was flattened and then fed into the fully connected layer. The fully connected layer outputs of the two branches were combined and input into a fully connected layer with 16 neurons. This layer served as the coding layer for feature extraction. Subsequently, the architecture included a fully connected layer with four neurons and an output layer with a single dimension. Tanh is employed as the activation function for all layers. This model is a supervised extended network model that uses the binding affinity between HLA and the antigen as a label, adopts the mean squared error (MSE) as the loss function, and utilizes the Adam optimizer for training purposes.

TCR–pMHC module

To learn the binding specificity between TCR and pMHC, we constructed a supervised neural network as a model. The TCR and pMHC features were used as inputs to the model. The combination of these two, in which negative and positive cases were denoted by 0 and 1, respectively, served as labels for training. After feature extraction, both the TCR and pMHC feature vectors were one-dimensional.

To avoid overfitting in complex networks, a network with a simple structure was selected. The TCR and pMHC features were separately input into two branch structures. Each branch structure comprised two fully connected layers, with the number of neurons set to 32 and 16. Subsequently, the outputs of the two branch structures were concatenated and sequentially fed into three fully connected layers. The number of neurons in these layers was 8, 4, and 1, and the output of the last fully connected layer represented the final output of the model, corresponding to the binding specificity of TCR and pMHC. The sigmoid function was used as the activation function for all layers. The Adam optimizer and MSE were employed as the optimizer and loss function of the model, respectively.

Generating artificial negative TCR–pMHC sequence pairs

Two approaches were used to address the imbalance between negative and positive samples in the dataset. First, the synthetic minority oversampling technique (SMOTE) method was utilized, as it randomly selects neighboring samples based on the existing minority samples to synthesize new minority samples (34). Second, binding-negative TCR–pMHC pairs were fabricated using artificial negative TCR sequences. Specifically, this method involves randomly replacing one or more amino acids in the TCR sequences from the database of the present study. If the resulting sequence does not exist in the database of authentic TCR sequences, it is added to the negative TCR sequence library as an artificial negative TCR sequence. By pairing any TCR sequence from the negative

TCR sequence library with any known pMHC sequence, binding-negative TCR–pMHC pairs are generated.

HLA evolutionary divergence module construction

To predict the broadness of the immunopeptidome presented by the encoded HLA molecules, we constructed an HLA evolutionary divergence (HED) module to quantify the genetic disparity (or divergence) among alleles at each individual HLA locus by employing Grantham distance metrics to compute the differences in the composition, polarity, and volume of polymorphic amino acid side chains across distinct HLA alleles (35).

Generating TCRscore model

We finally assembled the TCRscore model by integrating the protein sequence coding, TCR representation, pMHC feature extraction, TCR–pMHC, and HED modules.

Evaluation of TCR representation module

In the TCR representation module, the TCR β-chain CDR3 sequences that had been converted using the *Z* descriptor were encoded. Subsequently, a comparison was made between the heatmap of the descriptor matrix corresponding to the original TCR β-chain CDR3 sequences and that of the matrix output by the decoder. The striking congruence between these two heatmaps, quantified by the Pearson correlation between the original and reconstructed descriptor matrices (each 50 × 3 *Z* descriptor matrix is flattened into a length-150 vector), served as compelling evidence that the encoder–decoder model could accurately reconstruct TCR β-chain CDR3 sequences.

Evaluation of pMHC feature extraction module

To assess the performance of the pMHC feature extraction module, the entire dataset was randomly partitioned into training and independent test sets at an 8:2 ratio. Subsequently, the training set was further randomly split into subtraining and internal validation sets at a 9:1 ratio. Moreover, we utilized both the internal and external validation datasets to compute the Pearson correlation coefficients between their input and output matrices as the number of training epochs progressed. The results clearly demonstrated that as the number of training epochs increased, the model's performance gradually improved and ultimately reached a stable state.

Evaluation of the pMHC–TCR module

To conduct a more comprehensive evaluation of the performance of the pMHC-TCR module, the entire dataset was randomly divided into training and independent test sets at a 9:1 ratio. Thereafter, the training set underwent further random partitioning into subtraining and internal validation sets at various ratios: 9:1, 8:2, 7:3, and 6:4. This partitioning and subsequent model training procedure was iterated 50 times, resulting in 50 trained models. Then, the ACC of these 50 models was computed. Finally, among all the models, the model that achieved the highest ACC on the internal validation set, when the subtraining set/internal validation set ratio was 9:1, was selected as our ultimate model.

Determination of the cutoff value for the TCRscore model and performance comparison

We retrieved 126 experimentally validated neoantigen datasets from the public domain (36–42). Only samples with fully resolved HLA allelic information were included. For each neoantigen, restricted HLA-related information and experimental immunogenicity results corresponding to the patient were available. For each HLA–peptide pair in the samples, we computed the TCRscore and ranked them according to the TCRscore values. Subsequently, considering the number of peptides retained at different thresholds, we conducted 1,000 random samplings. For each random sampling, we calculated and fitted the ACC to ascertain the optimal threshold using positive predictive value (PPV) and ACC, which were calculated as follows:

$$PPV = \frac{N_{correct\ predicted\ positive\ antigens}}{N_{All\ predicted\ positive\ antigens}}$$

$$ACC = \frac{N_{correct\ predicted\ positive\ antigens} + N_{correct\ predicted\ negative\ antigens}}{N_{All\ predicted\ positive\ antigens} + N_{all\ predicted\ negative\ antigens}}$$

where *N* represents the total predicted results. PPV highlighted the predictor's ability to prioritize true neoantigens.

This subset of data were then used to compare the performance of our method with that of the six other widely used neoantigen prediction tools including ImmuneApp (43), PRIME 2.0 (44), NeoPred (45), MixMHCpred (44), MHCflurry 2.0 (46), and HLAthena (47).

Brain glioma tissue samples and database

Tumor tissues and blood samples were obtained from patients undergoing surgery in the Department of Neurosurgery at Beijing Tiantan Hospital, affiliated with Capital Medical University. Tumor tissues were independently diagnosed as GBM by at least two senior neuropathologists in accordance with the 2021 WHO Classification of Tumors of the CNS. The samples collected for this study have also been archived in the Chinese Glioma Genome Atlas (CGGA) database. Sample collection was approved by the Medical Ethics Committee of Beijing Tiantan Hospital, Capital Medical University (Ethics Approval No: KY2014-002-02), and written informed consent was obtained from all patients before sample collection. Studies were conducted in accordance with the Declaration of Helsinki.

Peripheral blood mononuclear cell isolation and cryopreservation

Peripheral blood mononuclear cells (PBMC) were isolated and cryopreserved following a modified protocol (48). Briefly, heparinized blood (20 mL) was diluted with an equal volume of phosphate-buffered saline (PBS) and carefully layered over 20 mL of lymphocyte separation medium in a 50 mL separation tube to create a distinct interface. The samples were centrifuged at 2,500 rpm for 20 minutes, and the mononuclear cell layer was collected and diluted 1:1 with PBS. Following a centrifugation step at 1,500 rpm for 15 minutes, the cell pellet was resuspended in 2 mL of red blood cell lysis buffer, gently mixed, and left to stand for 3 minutes. The suspension was then diluted to 45 mL with PBS and centrifuged again to obtain a purified PBMC pellet. The cells were resuspended in 5 mL of PBMC culture medium [RPMI 1640 with 10% fetal bovine serum (FBS), 100 U/mL IL2, and 1% antibiotics], counted, and plated at 1×10^6 cells/mL. After 6 hours of incubation, the PBMCs were observed under a light microscope to assess morphology. The cells were subsequently centrifuged, and the resulting pellet was resuspended in 1 mL of PBMC freezing medium (90% RPMI 1640, 10% FBS). The suspension was transferred to cryovials, labeled, and stored in liquid nitrogen for long-term preservation.

Construction, culture, and passaging of patient-derived organoid models

Sterile culture dishes and microscissors, tweezers, PBS, Normocin, and an organoid culture system (DMEM, 50% neurobasal, 1%

GlutaMax, 1% non-essential amino acids, (NEAAs), 1% penicillin-streptomycin, 1% N2, 1× B27) were prepared. All reagents were obtained from Gibco and Thermo Fisher Scientific. Six-well plates and 50 mL centrifuge tubes were exposed to UV light in a laminar flow hood for 30 minutes before use. Fresh tumor tissue was obtained from the operating room, placed in PBS, and transported on ice for preservation. Thereafter, the tissue was washed thoroughly in a sterile culture dish with 15 mL PBS and 150 μL Normocin, followed by three additional washes with 15 mL PBS each. The tissue was then immersed in 5 to 10 mL PBS and minced into pieces approximately 1 mm in diameter using sterile microscissors. Next, the minced tissue was transferred to a centrifuge tube, to which 15 mL of PBS was added, and the suspension was centrifuged at 800 rpm for 5 minutes. Subsequently, the supernatant was discarded, and the tissue pellet was resuspended in culture medium and transferred to a six-well plate. Organoids were cultured in a sterile incubator at 37°C under 5% CO_2 and 90% humidity on a shaker (120 rpm), with daily observation and medium changes. After 2 to 3 weeks, tumor fragments formed spherical organoids. When organoids reached ≥3 mm, they were transferred to a laminar flow hood and minced again using sterile micro scissors into pieces <1 mm in diameter. The centrifugation and culture steps were repeated for organoid passaging.

After 2 to 3 weeks of culture, mature organoids were harvested and cryopreserved following a modified protocol from Jacob and colleagues (49). Briefly, organoids were cut into small fragments, pretreated with Y-27632, and frozen in medium containing 10% DMSO using a controlled-rate freezing method before transfer to liquid nitrogen. Upon recovery, thawed organoids were gradually rehydrated in Y-27632–supplemented medium and typically resumed normal growth within 1 to 2 weeks.

IHC validation of organoid model (Ki-67)

Organoids were fixed with 1 mL of 4% paraformaldehyde (PFA) at room temperature for 20 minutes, followed by overnight fixation with PFA. The organoids were then dehydrated using a sucrose gradient (10%, 20%, and 30%)—each step lasting approximately 4 hours—until they settled at the bottom. The organoids were embedded in optimal cutting temperature (OCT) compound, with five organoids placed per mold, and surgical tumor tissue was covered with OCT compound and frozen at −80°C for 30 minutes. Sections were cut to 5 μm thickness after trimming to 20 μm. Before staining, sections were air-dried for 1 hour at room temperature and stored at −20°C if not used immediately. Afterward, the sections were circled with a hydrophobic pen, fixed with ice-cold methanol for 15 minutes, and washed 3 times with PBS. Peroxidase-blocking solution was added for 15 minutes, followed by three PBS washes. Next, goat serum–blocking solution was added for at least 30 minutes, followed by three PBS washes. The primary antibody (Ki-67, 1:100; Zhongshan Golden Bridge; RRID:AB_2890067) was diluted and added to the sections, which were incubated overnight at 4°C. After washing off the primary antibody, the secondary antibody was added and incubated at room temperature for 1 hour. Subsequently, sections were washed 3 times with PBS, and DAB chromogenic solution was added in the dark, with the reaction stopped after approximately 1 minute under light microscope observation. The sections were then stained with hematoxylin for 12 minutes, rinsed with tap water for 10 minutes, and sequentially immersed in 50%, 70%, and 100% ethanol for 5 minutes each, followed by xylene for 5 minutes. Finally, the sections were mounted, air-dried, and scanned using a slide scanner.

Immunofluorescence staining on frozen sections

Frozen sections were air-dried in a 37°C oven for 10 to 20 minutes and fixed in 4% PFA for 30 minutes. After washing with PBS (pH 7.4), antigen retrieval was performed using EDTA buffer (pH 9.0) in a microwave oven (medium power for 5 minutes, rest for 5 minutes, then low power for 10 minutes), followed by natural cooling and PBS washes. Endogenous peroxidase activity was blocked using 3% hydrogen peroxide for 15 minutes at room temperature. The sections were then incubated with 3% BSA in PBS for 30 minutes to block nonspecific binding.

Primary antibodies were applied overnight at 4°C in a humidified chamber. The following antibodies were used: anti-SOX2 (23064; Cell Signaling Technology, RRID:AB_2714146), anti-OLIG2 (YT7967; Immunoway Biotechnology, RRID:AB_3712855), anti-DCX (ab207175; Abcam, RRID:AB_2894710), anti-GFAP (ab4648; Abcam, RRID:AB_449329), and anti–cleaved caspase 3 (MAB835; R&D Systems, RRID:AB_2243951). After PBS washes, the corresponding horseradish peroxidase (HRP)–conjugated secondary antibodies were applied for 50 minutes at room temperature, followed by tyramide signal amplification with different fluorophores. Multiple rounds of staining were performed sequentially using different tyramide dyes, with antigen retrieval and antibody stripping steps between rounds as needed. The four primary antibodies were detected using tyramide-conjugated fluorophores at distinct emission channels: GFAP (480 nm), doublecortin (520 nm), SOX2 (570 nm), and OLIG2 (620 nm), allowing for multiplexed visualization without spectral overlap. Finally, nuclei were counterstained with DAPI, and slides were mounted using an antifade mounting medium.

Whole-exome sequencing data–somatic mutation calling

Raw sequencing data underwent quality control using FastQC (RRID:SCR_014583). Subsequently, Trimmomatic (RRID:SCR_011848; ref. 50) was used for trimming to eliminate low-quality reads and remove adapters. The quality-controlled sequencing reads were aligned to the human reference genome (hg19) using Burrows–Wheeler Aligner, with blood sample sequencing data serving as a reference. Somatic mutations were detected using MuTect2 (RRID:SCR_026692; ref. 51), VarScan (RRID:SCR_006849; ref. 52), MuSE (RRID: SCR_020252; ref. 53), SomaticSniper (54), and Strelka (RRID:SCR_005109; ref. 55). ANNOVAR (RRID: SCR_012821; ref. 56) was used to filter and annotate the detected mutations. To obtain high-quality somatic mutations, the following criteria were imposed: (i) The number of mutated reads in the tumor sample must exceed one, and the variant allele frequency should be ≥ 0.05; (ii) sequencing depth in the tumor tissue should be at least 15 (depth of coverage ≥ 15); and (iii) the mutation must be supported by a minimum of two tools. The filtered somatic mutations were then annotated and utilized for subsequent analyses. Copy-number variation (CNV) profiles were generated from whole-exome sequencing (WES) data using CNVkit (RRID:SCR_021917; ref. 57).

WES data–HLA typing

HLA haplotypes were predicted using OptiType (58) software based on WES sequencing data from blood samples.

WES data–neoantigen prediction

Using the annotated VCF file, peptides were generated by sliding a window across the positions of affected mutations. This process produced all possible 9-mer peptides corresponding to non-synonymous somatic mutations. The binding affinity between these peptides and the patient's HLA-I alleles was predicted using NetMHCpan-4.0. High-affinity binders were defined as those peptides

with an IC_{50} value of ≤500 nmol/L. Subsequently, peptides with a mutated-form binding affinity of <500 nmol/L and corresponding wild-type peptide affinity >500 nmol/L were selected as candidate antigens for further analysis.

We screened these candidate antigens according to two criteria. First, they had to be present in both tumor tissues and organoids. Second, the expression levels (transcripts per million value, nonlog) of the mutant genes should be >1. Next, we used the TCRscore tool to calculate the immunogenicity scores of these 9-mer candidate antigens. Peptides with a TCRscore exceeding 0.6 were classified as neoantigens.

The Cancer Genome Atlas and CGGA cohorts

Mutation data for The Cancer Genome Atlas (TCGA)-GBM and TCGA-lower grade glioma (LGG) were downloaded from the Genomic Data Commons (GDC) in MAF format. Primary tumor samples (01A) were selected. Following this selection, hypermutated samples were identified and excluded from the analysis (mutation count >500), ultimately resulting in 824 samples for subsequent analysis. From the CGGA database, primary samples were chosen for mutation calling, producing 180 samples for further study. The mutation-calling approach was identical to that described above. MuTect2, VarScan, MuSE, SomaticSniper, and Strelka were utilized to detect somatic mutations, which were then filtered and annotated with ANNOVAR. The filtering criteria were the same as previously detailed. The R package maftools (version 2.22.0) was employed for visualizing sample mutations and calculating the tumor mutational burden (TMB).

RNA expression and HLA frequency analysis

RNA sequencing (RNA-seq) was performed on tumor tissue and organoids from all 21 samples in the cohort. Total RNA was extracted using the TRIzol reagent (Invitrogen) following the manufacturer's protocol. RNA purity and quantification were evaluated using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific), and RNA integrity was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies). Then, the libraries were constructed using the VAHTS Universal V6 RNA-seq Library Prep Kit according to the manufacturer's instructions. Transcriptome sequencing was performed on the Illumina NovaSeq 6000 platform to generate 150 bp paired-end reads. The transcriptome sequencing was conducted by OE Biotech Co., Ltd. FastQC was used to conduct quality control analysis on the raw sequencing data. Subsequently, Trimmomatic was applied for trimming, effectively filtering out low-quality reads and removing adapters. The quality-controlled sequencing reads were then aligned to the human reference genome (hg19) using HISAT2(RRID:SCR_015530; ref. 59). Read counts aligned to protein-coding genes for each sample were acquired using htseq-count.

HLA allele frequency data were sourced from the Allele Frequency Net Database (AFND, http://allelefrequencies.net) and HLA allele distribution in TCGA healthy samples (HLA-A, HLA-B, HLA-C). HLA alleles with a frequency exceeding 3% were designated as common HLAs for subsequent analysis (60).

RNA data–somatic mutation

Based on the mutations detected by WES data as reference, we conducted mutation detection on RNA data. Two distinct tools were employed for this purpose: Bamreadcount (arXiv:2107.12817v1, RRID:SCR_023653) and MutScan (61). Bamreadcount is capable of detecting mutations at corresponding sites from BAM files, whereas MutScan can directly perform mutation detection from raw FASTQ files. For single-nucleotide variants (SNV), mutations supported by both tools are utilized for downstream analysis. In contrast, for insertions and deletions, only the results from MutScan are considered the final version.

Peptide synthesis

The peptides used in this experiment were synthesized by Beijing SciLight Biotechnology Ltd. Each peptide was synthesized with a purity of 95%. The synthesis process involved several key steps: First, the Fmoc-protected monomers had their amino-protecting groups removed using a basic solvent (piperidine), and the carboxyl group of the subsequent amino acid was activated using an activating agent, allowing the activated monomer to react and form a peptide bond with the free amino group. These two steps were repeated in cycles until the desired peptide chain was fully synthesized. Once synthesis was complete, the peptide was cleaved from the resin, and the protecting groups were removed using trifluoroacetic acid. Finally, the synthesized peptide was analyzed and purified using reverse-phase high-performance liquid chromatography.

Peptide preparation and solubilization

Peptides were reconstituted at 2 mg/mL based on net charge: Peptides with a net positive or negative charge were first dissolved in sterile, nuclease-free water; neutral or poorly soluble peptides were dissolved in DMSO.

PBMC–peptide coculture

Patient PBMCs were thawed by placing the cryovial in a 37°C water bath for 5 minutes. Next, PBMC culture medium was prewarmed, and the cell suspension was transferred to a 50 mL centrifuge tube, followed by the addition of 15 mL of PBMC culture medium and centrifugation at 1,200 rpm for 5 minutes. The supernatant was discarded, and 1 mL of culture medium was added to thoroughly resuspend and count the cells. Thereafter, 5×10^5 PBMCs were added to a 24-well plate, followed by 500 μL of culture medium to resuspend and culture the cells. The peptide solution was thawed at room temperature, and 5 μL (20 μg/mL) of both wild-type and mutant peptide solutions were added to the culture system. Labels were marked on the surface of the plate, and the 24-well plate was placed in a sterile incubator at 37°C under 5% CO_2 and 90% humidity. The stimulation was repeated every 3 days by adding 5 μL of peptide solution, with additional culture medium as needed, for 2 to 3 cycles. After stimulation, cell samples were collected for the ELISpot assay.

ELISpot assay

Materials such as RPMI 1640 and PBMC media, phytohemagglutinin (PHA), IFNγ ELISpot kits (3420-2A, Mabtech, RRID:AB_3712871), ELISpot polyvinylidene difluoride (PVDF) 96-well plates (3654-TP-10, Mabtech), AEC substrate solution, 35% alcohol, PBS, FBS (Gibco, Thermo Fisher Scientific), multichannel pipettes, and reagent reservoirs were prepared. Plate preparation involved adding 20 μL of 35% alcohol per well to the 96-well plate, incubating for 50 seconds, discarding the alcohol, washing the plate 5 times with PBS, adding 100 μL of diluted capture mAb per well, and incubating at 4°C overnight. For cell incubation, the wells were washed, blocked with medium, and incubated at room temperature for 30 minutes. Stimulated PBMCs were resuspended and added to the plate, with three replicates for each group. Positive control wells received 5 μL of PHA, whereas negative control wells received only cell

suspension. The plate was incubated at 37°C under 5% CO_2 and 90% humidity for 24 to 48 hours. ELISpot detection involved washing the plate, adding the detection antibody, incubating, washing again, adding streptavidin-HRP, incubating, washing, adding AEC substrate solution, and observing until spots appeared. The reaction was stopped by washing with PBS. Finally, the PVDF plate was dried in a dark environment for 24 hours before imaging using an AID multifunctional microplate reader (AID Autoimmun Diagnostika GmbH, RRID:SCR_025287). For the HLA-blocking experiments, PBMCs were preincubated with Anti-human MHC Class I (HLA-A, HLA-B, HLA-C)-InVivo (1 μg/ml, A2172, Selleck, RRID:AB_3718661) prior to peptide stimulation.

Organoid coculture and immunofluorescence staining (organoid frozen sections)

Organoids derived from patient tissue were cocultured with the PBMC–peptide system, as described above. The coculture was maintained in RPMI 1640 medium supplemented with 10% FBS. After 3 days of coculture, organoid samples were collected, embedded in OCT compound, and sectioned for subsequent analysis.

A hydrophobic barrier was drawn around the sample using a hydrophobic pen, followed by fixation with PFA for 10 minutes, washing with PBS, antigen retrieval with ice-cold antigen retrieval solution at room temperature for 5 minutes, and washing again with PBS. Blocking was performed using QuickBlock immunostaining blocking buffer (Beyotime) for 15 to 20 minutes, followed by PBS washing. An autofluorescence quencher (Biotium) was applied for 5 minutes, followed by three washes with PBS. The primary antibody was diluted at a ratio of 1:200 to 1:400 and incubated overnight at 4°C. After three washes with PBS, the fluorescent secondary antibody was diluted at a ratio of 1:400 and incubated for 1 hour at room temperature. The slides were then washed with PBS 3 times and dried with absorbent paper, and the coverslip was cleaned using lint-free tissue. One drop of mounting medium containing DAPI was added to each slide (ensuring no bubbles), which was then covered with a coverslip and sealed tightly, expelling any air bubbles. Finally, images were captured using a fluorescence microscope with a 10× objective.

Flow cytometry analysis

The flow cytometry analysis involved collecting and washing the cell pellet (approximately 1×10^5 cells or a visible pellet in an EP tube), followed by incubation with directly conjugated flow antibodies at 4°C for 30 minutes and washing with PBS. The flow antibodies used in this experiment included CD107a/CD69 antibody (Elabscience, RRID:AB_3712858, RRID:AB_3712859) labeled with APC and CD3 antibody (Elabscience, RRID:AB_3712857) labeled with PE. Thereafter, the pellet was resuspended in 400 μL PBS and subjected to flow cytometry analysis. Each sample was run in triplicate, with 10,000 cells collected per run, and the main *in vitro* experiments were repeated 2 to 3 times.

Statistical analysis

Statistical analyses were performed using R (version 4.4.4; R Foundation for Statistical Computing, RRID:SCR_001905) and GraphPad Prism (version 9.3; GraphPad Software, RRID:SCR_002798). Continuous data are described as means ± SDs. For comparisons between two groups, an unpaired *t* test was used, whereas comparisons among more than two groups were performed using one-way ANOVA. Statistical significance was defined as follows: ****, $P < 0.0001$; ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$; ns, $P > 0.05$.

Results

Deep learning framework for predicting TCR–neoantigen–HLA interactions

We constructed a deep learning–based model, named TCRscore, to predict the immunogenicity of neoantigen peptides by integrating the intrinsic features of TCRs, neoantigen peptides, HLA class I molecules, and the HED metric. The TCRscore model comprises five modules: the protein sequence coding (**Fig. 1A**), TCR representation (**Fig. 1B** and **C**), pMHC feature extraction (**Fig. 1D** and **E**), TCR–pMHC (**Fig. 1F**), and HED (**Fig. 1G**) modules. The protein sequence coding module numerically encoded HLA sequences, peptide sequences, and TCR CDR3 regions using amino acid *Z* descriptors. Considering that the fourth and fifth descriptors are not clearly derivable, we utilized the first three *Z* descriptors, which correspond to lipophilicity (z1 or A), volume (z2 or B), and polarity (z3 or C; refs. 31–33).

In the TCR representation module, TCR sequences are first transformed via an autoencoder, which effectively compresses the high-dimensional sequence data into low-dimensional latent features. This approach preserves the key structural characteristics while reducing noise. Meanwhile, in the pMHC feature extraction module, the pMHC complex is processed using a one-dimensional convolutional neural network, enabling the automatic extraction of spatial and functional binding features.

These encoded TCR and pMHC features are subsequently concatenated and fed into the TCR–pMHC module. This module has been trained to predict a binding score, which reflects the likelihood of effective TCR–pMHC recognition. To address class imbalance, we employed SMOTE, which is effective in enhancing model generalizability (34). The HED module was then used to compute the differences in the composition, polarity, and volume of polymorphic amino acid side chains across distinct HLA alleles. Ultimately, we assembled the TCRscore model by integrating the aforementioned modules. This integrated model was then used to prioritize neoantigen peptides according to their predicted immunogenicity levels.

The predictive performance of the model has been thoroughly explored and verified, both for the individual modules and the entire model. The comparison between the input TCR and reconstructed TCRs showed that CDR3s were faithfully reconstructed, demonstrating the success of autoencoder training (**Fig. 1C**; Supplementary Fig. S1). The predicted and true binding affinities between the MHC molecule and antigens in independent validation datasets were highly correlated (**Fig. 1E**). TCR–pMHC prediction achieved low loss (**Fig. 1H**), high ACC (**Fig. 1I**), and high area under the curve (AUC) values (**Fig. 1J**).

To compare the performance of our TCRscore model with other state-of-the-art methods, we started with a set of 126 experimentally validated neoantigens retrieved from the public domain (Supplementary Table S1; refs. 36–42). Each of these neoantigens had all HLA-related patient-specific information available. The TCRscore model demonstrated robust predictive capabilities for neoantigens, as evidenced by its PPV across various prediction thresholds (**Fig. 1K**), which markedly surpassed that of random sampling. This effectively validated its utility in neoantigen selection. From this set, we evaluated the performance of TCRscore against these other methods: ImmuneApp (43), PRIME 2.0 (44), NeoPred (45), MixMHCpred (44), MHCflurry 2.0 (46), and HLAthena (47). As

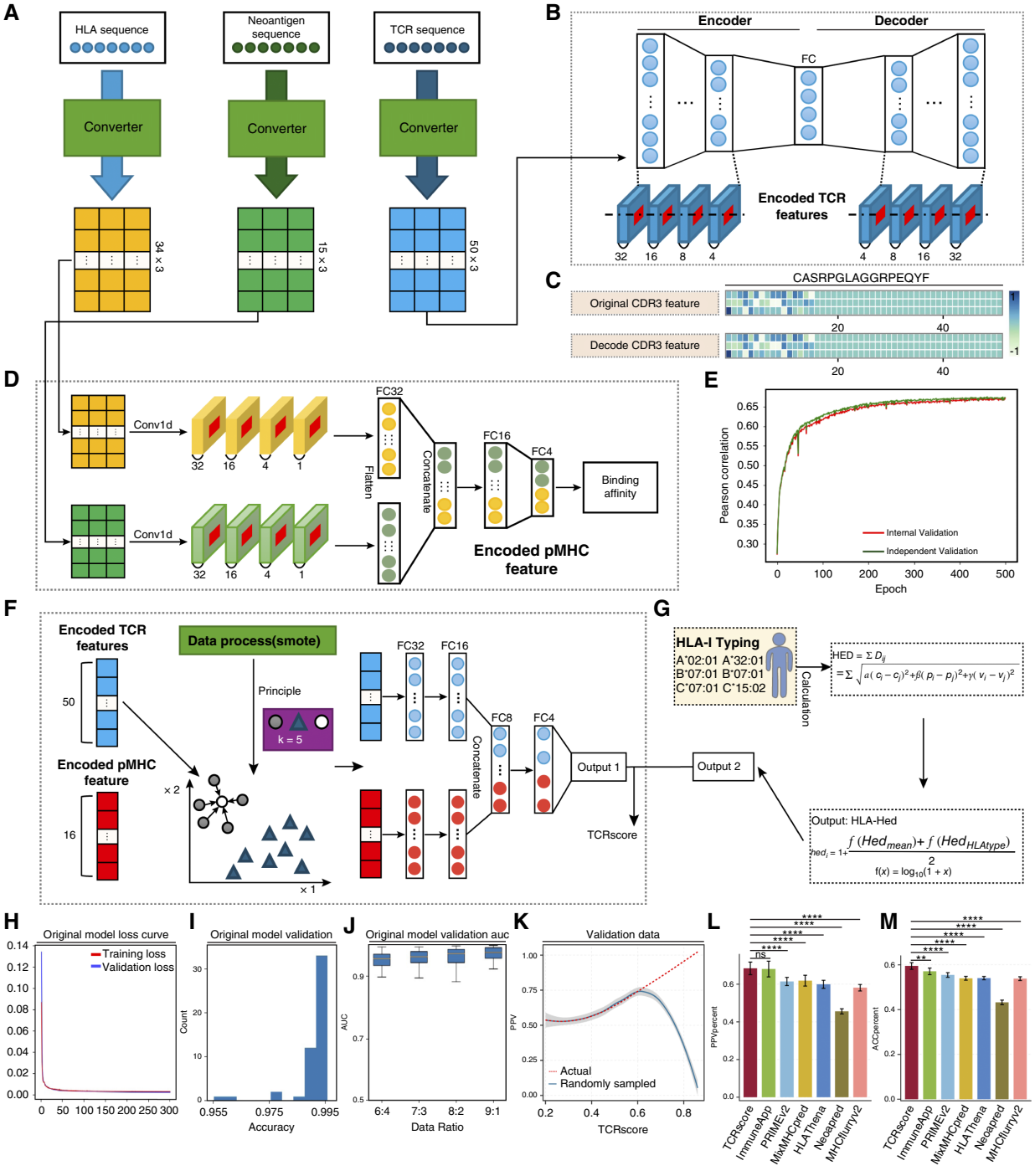

Figure 1.

Construction of a deep learning–based algorithm for neoantigen prediction and screening workflow. **A,** Input processing of HLA, neoantigen, and TCR sequences using amino acid *Z* descriptors. **B,** TCR sequence encoding via autoencoder to extract latent CDR3 features. **C,** Comparison of original and reconstructed CDR3 features to validate encoding performance. **D,** Convolutional neural network (CNN)-based model for pMHC feature extraction and binding affinity prediction. **E,** Internal and independent validation of the predicted binding between neoantigens and MHC proteins generated by the pMHC feature extraction module. **F,** Integration of TCR and pMHC features, with SMOTE for data balancing and final interaction scoring. **G,** HLA typing and calculation of HED to refine immunogenicity scoring. **H,** Training and validation loss curves demonstrating model convergence. **I,** ACC distribution on the validation dataset. **J,** AUC performance under different training/validation data splits. **K,** PPV curve for the validation set compared with random sampling. **L** and **M,** Summary of mean PPVpercent (**L**) and ACCpercent (**M**) with 95% confidence intervals (CI) across all valid *n* values for each method ($n = 126$). PPVpercent: PPV within the top *n*% predictions. ACCpercent: ACC within the top *n*% predictions. Statistical comparisons were performed using paired Wilcoxon tests; **, $P < 0.01$; ****, $P < 0.0001$.

demonstrated by the PPV (**Fig. 1L**) and ACC (**Fig. 1M**), TCRscore exhibited significantly better performance than all the other methods.

Patient-derived GBM organoids recapitulate histologic and molecular features of primary tumors

We have successfully established 21 patient-derived GBM organoid models from IDH wild-type tumor samples. According to the 2021 WHO classification, IDH wild-type glioblastoma represents the majority of adult GBM cases and is associated with more aggressive clinical behavior and poorer prognosis compared with IDH-mutant gliomas (62). The clinical and molecular characteristics of the cohort are summarized in Supplementary Table 2. The median age was 55 years (range, 28–69 years), with a male-to-female ratio of approximately 1.6:1. All patients were histologically confirmed as GBM (WHO CNS 4), and none harbored mutations in IDH1 or IDH2, consistent with the IDH wild-type classification.

To evaluate the proliferative capacity of these organoids, we performed Ki-67 IHC staining on representative samples. As shown in **Fig. 2A**, high Ki-67 expression was consistently observed across organoids from different patients (P06, P09, P12, P17, and P20 are shown), indicating robust proliferative activity. To assess histologic fidelity, hematoxylin and eosin staining was performed. The architectural characteristics of the original tumors were preserved by the organoids, including cellular density and pleomorphism (**Fig. 2B**). Furthermore, immunofluorescence staining revealed similar expression patterns of glioma lineage markers—SOX2, OLIG2, DCX, and GFAP—between organoids and their matched parental tumors (**Fig. 2C**), supporting the retention of cell type heterogeneity within the model.

WES analysis revealed that the mutational profiles of organoids closely resembled those of their corresponding tumors. As shown in **Fig. 2D**, key GBM-related mutations, including those in *TP53*, *EGFR*, and *NF1*, were preserved across organoid–tumor pairs. We further analyzed the proportion of shared versus unique mutations using RNA-seq and WES data. A considerable proportion of mutations were determined to be shared, particularly at the RNA level, with a minority being organoid- or tumor-specific. Fifty-four percent of nonsynonymous SNVs (neoantigen-associated mutations, range: 10%–80%) were shared by organoids and matching tissues on average (**Fig. 2E**).

Correlation analysis of WES and RNA variant data confirmed high genomic concordance between the tumor and its corresponding organoid samples but rarely with other tissues (**Fig. 2F**). Moreover, we evaluated genome-wide structural consistency between organoids and primary tumors by analyzing CNVs. As shown in **Fig. 2G**, CNV heatmaps were generated for all samples across chromosomes 1 to 22 and X. The organoids retained the major CNV patterns observed in their matched tumors, including common amplifications and deletions across genomic regions. We calculated the mean CNVs across chromosomes and evaluated the correlation between organoid and matched tumor tissue samples. The correlation coefficients ranged from 0.39 to 0.96, with a median of 0.70. A global comparison based on WES data revealed a high degree of similarity, further supporting the structural genomic fidelity of the organoid models.

TMB was quantified and compared between tumors and their derived organoids. No significant difference was observed in TMB (**Fig. 2H** and **I**). Notably, tumor purity, inferred from transcriptomic data, was slightly but significantly higher in the organoid models compared with that in their parental tumors (**Fig. 2J** and **K**). This was likely due to reduced stromal or immune cell content during *in vitro* culture, implying that the organoids efficiently concentrated cancer cells, making them a reliable model for studying tumor biology.

Additionally, we assessed TMB distribution across large public glioma cohorts, including CGGA and TCGA, stratifying patients based on key clinical and molecular features. TMB levels varied across subgroups but remained within expected ranges for IDH wild-type GBM (Supplementary Fig. S2A and S2B). TMB was calculated using our own 21-patient cohort and showed a comparable distribution, which is consistent with that of publicly available datasets (Supplementary Fig. S2C), thus providing a robust foundation for downstream neoantigen analysis.

Considering that mutation expression levels are a key determinant of the immunogenicity of neoantigens generated by nonsynonymous mutations, we further evaluated the neoantigen-associated mutations landscape retained in GBM organoid models at the transcription level by conducting a paired comparison of the predicted neoantigen-related mutations between the organoids and their matched parental tumors. Venn diagrams illustrated the number of mutations shared and specific to each patient at the RNA level (Supplementary Fig. S3A and S3B). In most cases, a considerable number of neoantigen-associated mutations were shared between the organoid and tumor, with high concordance observed in patients such as P05, P15, and P17. Therefore, organoid models may largely preserve the immunogenic mutational landscape of the original tumor. However, two patients (P19 and P21) were not included in the figure due to a complete lack of overlapping neoantigen-associated mutations, possibly reflecting extreme transcriptional differences or minimal mutational overlap detectable by our pipeline. Furthermore, we conducted transcriptome-wide correlation analyses based on RNA-seq gene expression data to quantify the degree of transcriptional similarity between tumors and their corresponding organoids. This layer demonstrated strong concordance (Supplementary Fig. S3C), further confirming the preservation of gene expression patterns during organoid derivation. Overall, our data support the use of organoids as a reliable platform for neoantigen-level studies.

Collectively, GBM organoids faithfully recapitulate the key histologic and molecular features of their matched parental tumors, supporting their utility as preclinical models for GBM research.

TCRscore accurately predicts immunogenic neoantigens and outperforms existing tools in our GBM cohort

To systematically evaluate the predictive power of our model, we first established a comprehensive neoantigen screening workflow (**Fig. 3A**). Initially, we prescreened for candidate peptides by integrating whole-exome and whole-transcriptome sequencing data, HLA typing, and *in silico* neoepitope prediction (NetMHCpan-4.0). Only peptides derived from mutated genes with an expression level greater than 1 and present in both tumor tissues and organoids were selected for further analysis. Before selecting candidate neoantigens, we further analyzed the distribution of peptide lengths (8–12-mer) using the IEDB database, which revealed that 9-mer peptides were the most prevalent (Supplementary Fig. S4). Therefore, we focused primarily on 9-mer peptides in the subsequent experimental validations. Overall, 127 candidate peptides were screened from the GBM cohort comprising 21 IDH wild-type GBMs, each of which had associated patient-derived GBM organoid models established (**Fig. 3B**; Supplementary Table S3). Subsequently, we ranked all 127 candidates according to their TCRscore values to assess the model's capacity to prioritize immunogenic peptides. The peptides were then partitioned into two groups:

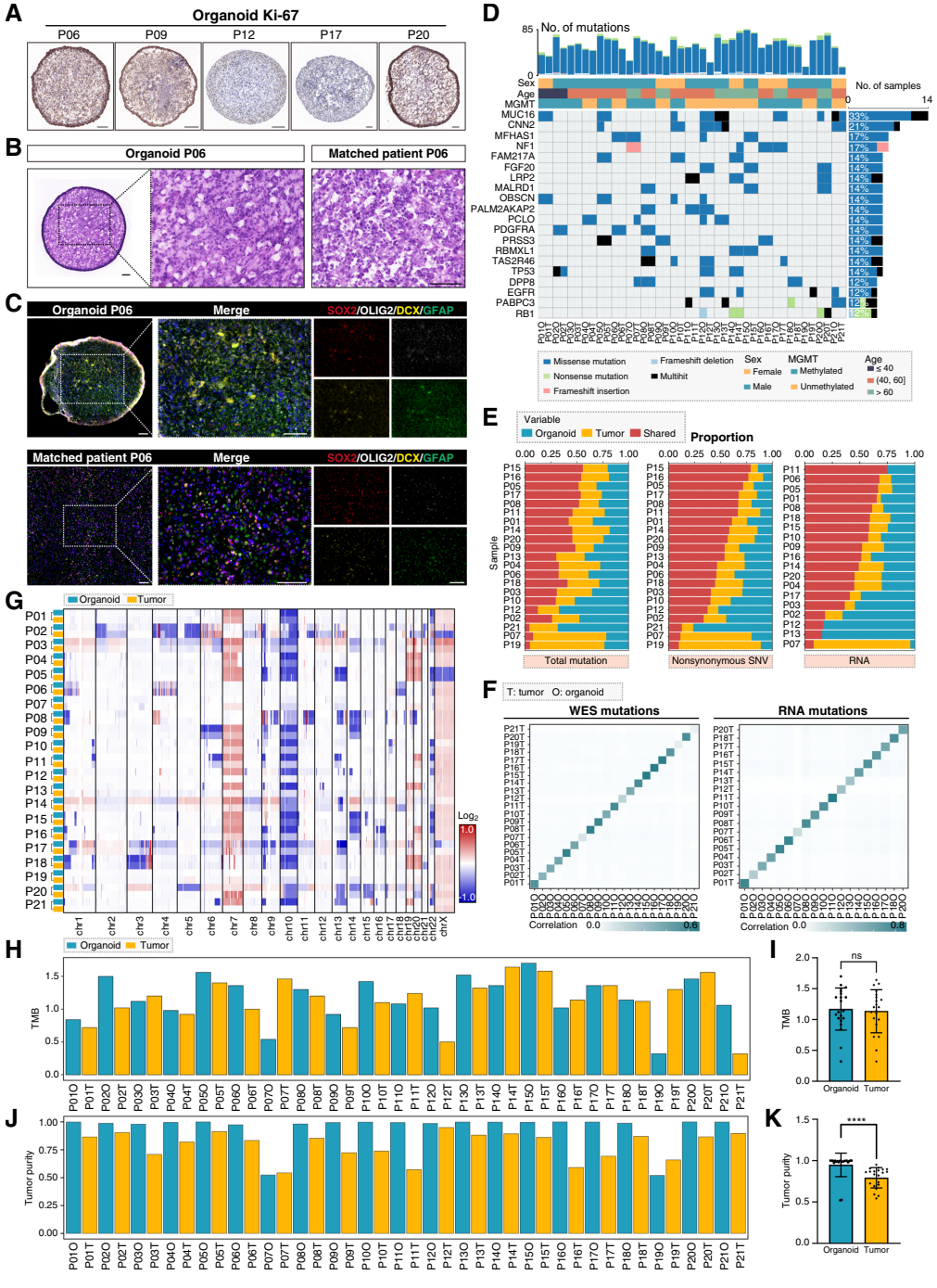

Figure 2.

Histologic and molecular characterization of patient-derived IDH wild-type GBM organoid models. A total of 21 IDH wild-type GBM organoid models were successfully established and compared with their matched parental tumors to assess histologic and transcriptomic fidelity. **A,** Representative IHC staining of Ki-67 in selected GBM organoid models (P06, P09, P12, P17, P20; P, patient), demonstrating high proliferative activity across models. **B,** Hematoxylin and eosin staining of organoid P06 and its matched parental tumor, revealing preserved histologic features. **C,** Immunofluorescence staining for SOX2, OLIG2, DCX, and GFAP in organoid P06 and matched tumor tissue, showing comparable expression patterns of glioma lineage markers. **D,** Mutation landscape of the organoid–tumor pairs based on WES, including mutation types, gene alterations, and patient metadata. **E,** Proportion of shared and unique mutations in total mutations, nonsynonymous SNVs, and RNA-seq (RNA) across organoid and tumor samples. **F,** Correlation heatmaps of WES and RNA-based mutation profiles between tumor and organoid samples, confirming high genetic similarity. **G,** Heatmap of CNV profiles for all organoid–tumor pairs across the genome. Each column represents a chromosome, and each row corresponds to a sample (organoid or tumor). The color scale indicates $\log_2$ copy-number changes. **H** and **I,** TMB in organoid and tumor samples, with no significant difference observed (ns, not significant). **J** and **K,** Tumor purity comparison showing slightly higher purity in GBM organoids compared with tumors (****, $P < 0.0001$).

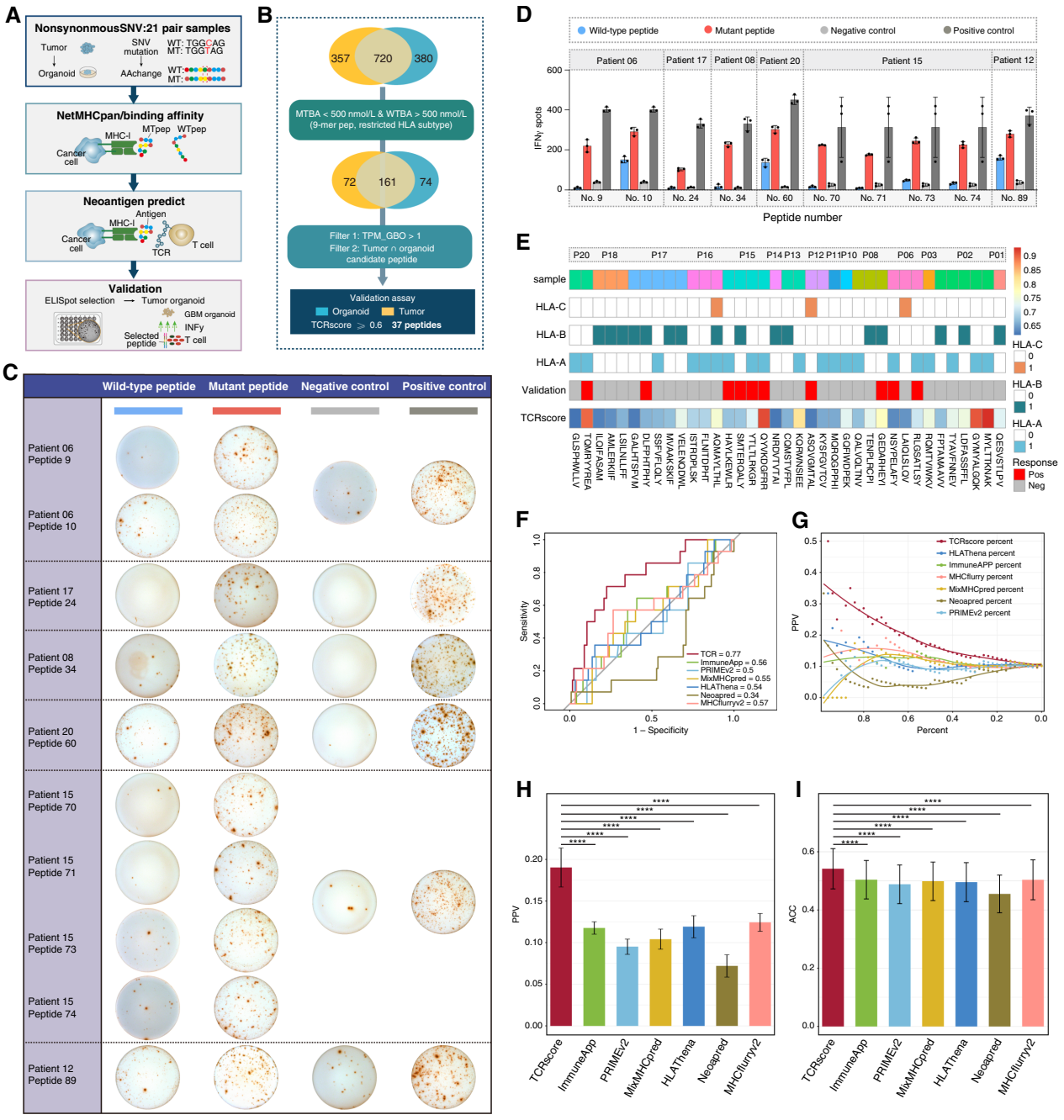

Figure 3.

Experimental validation and benchmarking of neoantigen prediction methods in our GBM cohort. **A,** Workflow for neoantigen prediction and screening, integrating genomic, transcriptomic, and HLA-binding information. **B,** Inclusion criteria and summary of our GBM cohort selection from public datasets for downstream neoantigen evaluation. **C** and **D,** Representative ELISpot assay results demonstrating T-cell reactivity to predicted neoantigen peptides. **E,** Distribution of immunogenic peptides (ELISpot-positive) across different HLA alleles and their presence in individual patient samples. **F,** ROC curves and AUC values for TCRscore and six existing neoantigen prediction tools: ImmuneApp, PRIME 2.0, NeoPred, MixMHCpred, MHCflurry 2.0, and HLAthena, evaluated on 127 peptides. **G,** PPV (PPVpercent) calculated for all methods as the fraction of validated immunogenic peptides ranked within the top *n*% of predicted peptides ($n = 127$). **H,** Summary of mean PPVpercent with 95% confidence intervals (CI) across all valid *n* values for each method. Statistical comparisons were performed using paired Wilcoxon tests; ****, $P < 0.0001$. **I,** Summary of mean ACCpercent (ACC within the top *n*% predictions) with 95% CI for all methods. Statistical comparisons were performed using paired Wilcoxon tests; ****, $P < 0.0001$. WT, wild type.

high-ranking peptides (the top-scoring candidates predicted to be immunogenic, TCRscore ≥0.6, **Fig. 1K**) and low-ranking peptides (the bottom-scoring candidates anticipated to be non-immunogenic, TCRscore <0.6). This categorization enabled the evaluation of the true positive rate and potential false negatives, thereby offering a more comprehensive appraisal of the model's performance.

For the high-ranking peptides, we conducted IFNγ ELISpot assays using autologous T cells to assess their immunogenicity. Several peptides elicited strong IFNγ responses, as determined by a significant increase in spot counts upon stimulation with mutant peptides compared with corresponding wild-type peptides (**Fig. 3C** and **D**). This result confirms the ability of these neoepitopes to effectively activate T cells. To further verify T-cell specificity, we selected five positive peptides and performed HLA blocking experiments as well as additional assays using PBMCs predicted not to bind the corresponding HLA alleles, which further confirmed HLA-restricted T-cell recognition (Supplementary Fig. S5). The validated peptides, including their sequences and source genes, predicted using HLA class I restriction, are summarized in Supplementary Table S4.

ELISpot assays were primarily performed using autologous PBMCs collected from patients with GBM at the time of surgery. Considering that GBM-associated immune dysfunction may affect assay sensitivity, additional validation was conducted using PBMCs from healthy donors. These donors were selected based on the presence of HLA class I alleles predicted to bind the corresponding neoantigen peptides, allowing for relevant pMHC presentation. This approach helped control potential variability in T-cell function and further validated the immunogenicity of selected peptides. The detailed HLA typing of the 21 patients with GBM and the healthy donor PBMCs is listed in Supplementary Tables S5 and S6.

We conducted an in-depth investigation into the distribution of immunogenic (ELISpot-positive) peptides among the various HLA alleles and their occurrence in individual patient samples. Our analysis revealed that these peptides were not limited to a single allele or an individual, and the peptides with immunogenicity that bind to HLA-A are more numerous (**Fig. 3E**). Then, we benchmarked the TCRscore against six widely used neoantigen prediction tools: ImmuneApp, PRIME 2.0, NeoPred, MixMHCpred, MHCflurry 2.0, and HLAthena (Supplementary Table S7). Using the ELISpot results as ground truth, receiver operating characteristic analysis was performed. As shown in **Fig. 3F**, TCRscore achieved the highest AUC (AUC = 0.77), indicating a superior ability to discriminate immunogenic from nonimmunogenic peptides.

To further assess practical predictive value, we calculated the PPVpercent, defined as the proportion of true immunogenic peptides ranked within the top *n*% of predictions. Across all *n* thresholds, TCRscore consistently outperformed the other methods (**Fig. 3G**). A summary of the mean PPVpercent with 95% confidence intervals is shown in **Fig. 3H**. Statistical analysis using paired Wilcoxon tests revealed a significant TCRscore advantage (****, $P < 0.0001$). Similarly, ACCpercent (ACC among top-ranked peptides) was calculated for each model. The TCRscore achieved the highest mean ACCpercent with narrow confidence intervals and significantly outperformed other tools (**Fig. 3I**; ****, $P < 0.0001$; Supplementary Figure S6A). These results demonstrate that the TCRscore ranks immunogenic peptides more accurately and also provides higher practical enrichment of validated neoantigens. In summary, the TCRscore achieved more than a 35% improvement in mean PPV and more than a 7% increase in mean ACC compared with existing neoantigen prediction methods.

To evaluate the limitations of the model and identify potential false negatives, we investigated the peptides that had low TCRscore rankings yet triggered positive immune responses in ELISpot assays. As presented in Supplementary Fig. S6B, three such peptides induced detectable IFNγ secretion, notwithstanding their low prioritization by the model. Quantitative analysis affirmed that the spot counts for these peptides were significantly higher than those of the wild-type and negative controls (Supplementary Fig. S6C). Supplementary Table S8 presents detailed information on the ELISpot-positive peptides with low TCRscore rankings. These peptides exemplify false-negative instances, emphasizing that although the TCRscore is generally highly effective, certain immunogenic peptides may not achieve high rankings.

Experimental confirmation of neoantigen immunogenicity using flow cytometry and organoid models

To functionally validate the immunogenicity of predicted neoantigens, we conducted flow cytometry assays to assess the activation and cytotoxic potential of neoantigen-specific $CD8^+$ T cells. CD69 was used as an early activation marker of T cells, whereas CD107a served as an indicator of degranulation and cytotoxic function. Compared with the wild-type controls, the peptide-stimulated T cells exhibited markedly increased expressions of both CD69 and CD107a, indicating enhanced activation and effector function (**Fig. 4A**).

Quantitative analysis across six patient-derived peptides was used to verify the significantly elevated percentages of $CD69^+$ and $CD107a^+$ $CD8^+$ T cells in response to mutant peptides (**Fig. 4B**). These results suggest that the predicted neoepitopes robustly activated the T cells and promoted cytotoxic granule release, consistent with functional immunogenicity.

To further assess the antitumor effects of neoantigen-specific T cells, we cocultured them with autologous GBM organoids. Immunofluorescence staining for GFAP (tumor marker) and Ki-67 (proliferation marker) revealed substantial reductions in $Ki\text{-}67^+$ tumor cells following exposure to mutant peptide-reactive T cells (**Fig. 4C**). In contrast, wild-type peptide-treated organoids retained elevated levels of Ki-67 expression, indicating limited immune-mediated cytotoxicity.

Quantification of $Ki\text{-}67^+$ cells among $GFAP^+$ tumor populations showed a consistent and significant decrease in the mutant peptide groups across all tested samples (**Fig. 4D**; ****, $P < 0.0001$).

In parallel, staining for cleaved caspase 3 (CC3) revealed that mutant peptides robustly triggered apoptosis in GBM organoids, as evidenced by a significantly higher proportion of $CC3^+$ cells compared with wild-type peptide treatment (Supplementary Fig. S7; ****, $P < 0.0001$).

These findings demonstrate that neoantigen-specific T cells not only become activated but also mediate effective antitumor responses in glioblastoma organoids.

PIK3R1^{G376R} identified as a shared neoantigen in multiple independent glioma cohorts

To identify recurrent driver mutations with neoantigen potential, we analyzed somatic mutation profiles from primary glioma samples in both TCGA and CGGA datasets. The top 25 high-frequency mutations were summarized along with clinical characteristics such as IDH mutation status, grade, and survival outcome (**Fig. 5A** and **B**). Among the shared alterations, *PIK3R1* emerged as a commonly mutated gene across both cohorts (TCGA: 6%, CGGA: 7%).

We used MutSigCV to identify the significantly mutated driver genes in each dataset. As shown in **Fig. 5C**, multiple genes overlapped between TCGA and CGGA, including *PIK3R1*, indicating its recurrent role in glioma genesis. Mutation mapping across the PIK3R1 protein revealed a clustering of alterations within the iSH2 domain, with *p.G376R* being the most frequent hotspot in both

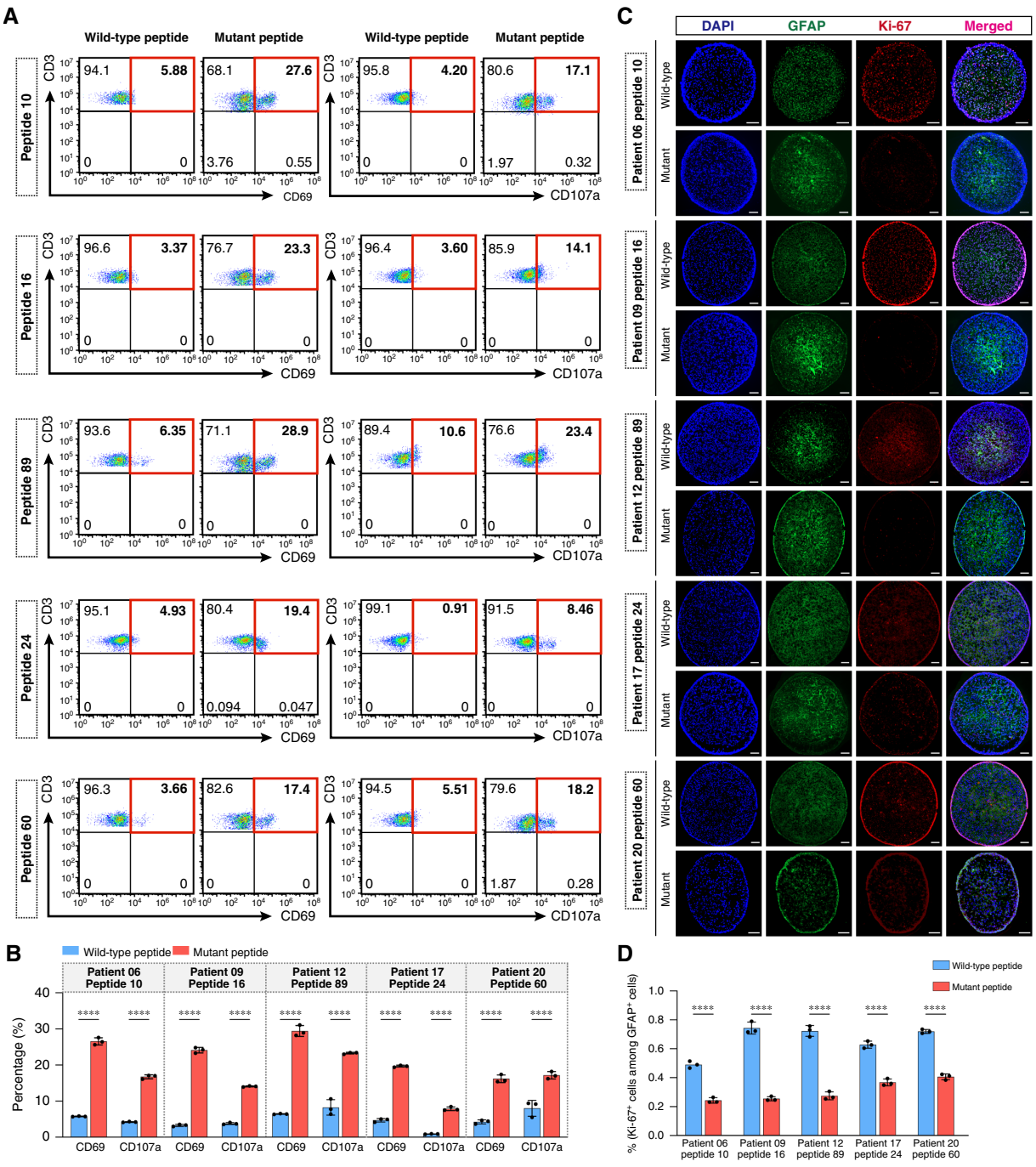

Figure 4.

Flow cytometric and organoid-based validation of neoantigen immunogenicity and T cell–mediated cytotoxicity. **A,** Representative flow cytometry plots showing CD69 and CD107a expression in $CD8^+$ T cells stimulated with wild-type or mutant peptides. Mutant peptides induced stronger T-cell activation and degranulation. **B,** Quantification of $CD69^+$ and $CD107a^+$ $CD8^+$ T cells in response to each peptide. Data are mean ± SEM; ****, $P < 0.0001$ (paired *t* test). **C,** Immunofluorescence images of glioblastoma organoids cocultured with peptide-specific T cells. Ki-67 staining (red) indicates proliferating tumor cells. Mutant peptide–reactive T cells reduced $Ki\text{-}67^+$ cells more effectively than wild-type controls. **D,** Quantification of $Ki\text{-}67^+$ cells among $GFAP^+$ tumor cells in organoids. Data are shown as mean ± SEM; ****, $P < 0.0001$.

datasets (**Fig. 5D**; Supplementary Fig. S8). To further quantify its recurrence, we examined the proportion of glioma samples harboring *PIK3R1* mutations. In both TCGA and CGGA cohorts, a substantial fraction of the *PIK3R1*-mutant samples carried the *p.G376R* substitution (**Fig. 5E**, 14% in TCGA; 25% in CGGA), supporting its classification as a shared mutation present across patients.

Considering its high recurrence and driver nature, we evaluated the potential of *p.G376R* to generate shared neoantigens. We predicted all possible 8- to 11-mer peptides spanning the mutant site and assessed their binding to common HLA class I alleles. Several peptides exhibited high TCRratio scores across multiple HLA types (**Fig. 5F**), indicating strong presentation and T-cell recognition potential. Notably, these HLA alleles are frequently represented in the TCGA glioma cohort, suggesting that *PIK3R1*^{G376R}-derived neoepitopes could be broadly applied across the patient population.

To functionally validate the immunogenicity of the recurrent *PIK3R1*^{G376R}-derived neoepitope, we further performed coculture assays. Flow cytometry analysis revealed that stimulation with the mutant peptide led to a significant increase in CD69$^+$ and CD107a$^+$ CD8$^+$ T cells compared with the wild-type counterpart (**Fig. 5G** and **H**), indicating effective T-cell activation and degranulation. To assess cytotoxic activity in a physiologically relevant context, we employed glioblastoma organoids derived from patient 15. Immunofluorescence staining demonstrated that T cells specific to the mutant peptide induced a stronger reduction in tumor cell proliferation, as evidenced by a lower proportion of Ki-67$^+$ cells among GFAP$^+$ tumor cells (**Fig. 5I** and **J**). Consistently, CC3 staining revealed a significant increase in tumor cell apoptosis following mutant peptide-specific T-cell treatment (Supplementary Fig. S9A and S9B). These results validate the *p.G376R* mutant peptide as a shared, functional neoantigen capable of eliciting cytotoxic T-cell responses both phenotypically and in an organoid-based model.

These results support the *PIK3R1*^{G376R}-derived neoepitope as a recurrent shared neoantigen in glioma, with potential utility for population-targeted immunotherapy approaches.

Discussion

Neoantigen-based personalized immunotherapy represents a promising approach to precision oncology. The binding of neoantigen peptides to MHC molecules is the first step of an immune response, in which the MHC molecules present the neoantigen peptides to T cells, subsequently activating the immune system and leading to tumor destruction (20). The predictive screening of neoantigen peptides primarily depends on two aspects: the binding ability between neoantigen peptides and MHC molecules, along with the immunogenicity of the neoantigen peptides. Increasing attention has been paid to the immunogenicity of neoantigen peptides, specifically their ability to induce T-cell responses (63). Previous studies have shown that only a small proportion of neoantigen peptides are immunogenic, and accurately screening for these immunogenic neoantigens is necessary for successful treatment using this therapy (64). NetMHCpan is a widely recognized algorithm for predicting the binding ability between neoantigen peptides and MHC. However, consensus on algorithms for immunogenicity prediction is currently lacking, which remains an important challenge for researchers in this field. For *in vitro* validation, IFNγ ELISpot is currently the most used technology. Meanwhile, *in vivo* models predominantly use murine tumor cell line xenograft models, as the development of humanized mouse models is challenging, costly, and time-consuming and has not yet been successfully applied to neoantigen research (65). Therefore, two major issues currently exist in the field of neoantigen therapy research: (i) the lack of a recognized algorithm based on peptide immunogenicity and (ii) the lack of a reliable preclinical model that accurately reflects the real tumor environment (66).

In the present study, we first developed TCRscore, a deep learning–based algorithm trained on publicly available datasets, which incorporates pMHC binding, gene expression, and TCR recognition features to improve neoantigen prioritization. We compared the performance of the TCRscore model with other state-of-the-art methods in publicly validated neoantigens, and TCRscore exhibited significantly better performance. To further evaluate its performance in an actual clinical context, we applied TCRscore to our cohort comprising 21 patients diagnosed with IDH wild-type GBM. TCRscore was used to prioritize candidate neoepitopes for experimental validation. Employing ELISpot, flow cytometry, and coculture assays, we showed that TCRscore attained a higher validation rate for immunogenic neoantigens in comparison with six widely used prediction tools.

The complex immune microenvironment within GBM presents obstacles to immunotherapy. Immune checkpoint inhibitors, which represent a milestone in cancer immunotherapy, have shown promising results in solid tumors such as melanoma (67). However, their efficacy in GBM is limited, likely due to the immunosuppressive microenvironment and tumor heterogeneity of GBM (68). Similarly, CAR T-cell therapy, an active immunotherapy, has achieved remarkable success in hematologic malignancies but faces substantial challenges in GBM, including target loss and off-target effects (69). In 2018, two independent clinical studies demonstrated that neoantigen therapy could improve the immune microenvironment of GBM, bringing new immunotherapeutic opportunities for GBM (6, 7). The efficacy of immunotherapy, including neoantigen-based approaches, is closely related to the immune microenvironment within GBM. Previously, ELISpot has been widely used as an important preclinical model for validating neoantigen peptides. However, as an *in vitro* model, ELISpot cannot replicate the real tumor microenvironment, limiting its ability to predict the true tumor-killing effect elicited by activated T cells. In the current study, we propose that GBM organoid models could serve as ideal preclinical models for neoantigen peptide validation, as we have previously reported in our earlier work (20).

GBM organoid models were initially developed by Hongjun Song's team at the University of Pennsylvania, demonstrating their ability to retain the heterogeneity, mutational characteristics, and immune microenvironment of the parental tumor and successfully applying them to drug screening studies (49). Certain immune components in organoid models may gradually diminish or be lost during *in vitro* culture, but the key components, such as macrophages, T cells, and DC cells, are partially preserved to recapitulate the dynamic immune interactions. Moreover, ongoing researches are actively exploring ways to maintain all possible immune components, including the optimization of media formulations and tissue preservation techniques. Notably, a recent study demonstrated that glioblastoma organoids could faithfully reproduce CAR T-cell responses observed in clinical settings, including cytokine patterns in cerebrospinal fluid (70). This underscores that despite the limitations of immune components, current organoid models can still offer potential and predictive insights into immunotherapy efficacy.

In our research, we utilized multiomics approaches to demonstrate that organoid models retain the mutational characteristics of their parental tumors. Building on the initial validation of peptide immunogenicity using ELISpot and flow cytometry, we further sorted neoantigen-reactive T cells and cocultured them with matched patient organoids. Most reactive T cells achieved substantial cytotoxic effects, further confirming the ACC of immunogenicity algorithm predictions and the sensitivity of organoid models as preclinical validation tools. Notably, in organoid-killing assays, not all of the 13 immunogenic peptides identified through ELISpot elicited strong cytotoxic responses, potentially due to the suppressive tumor immune microenvironment, which also underscores the necessity for preclinical validation models.

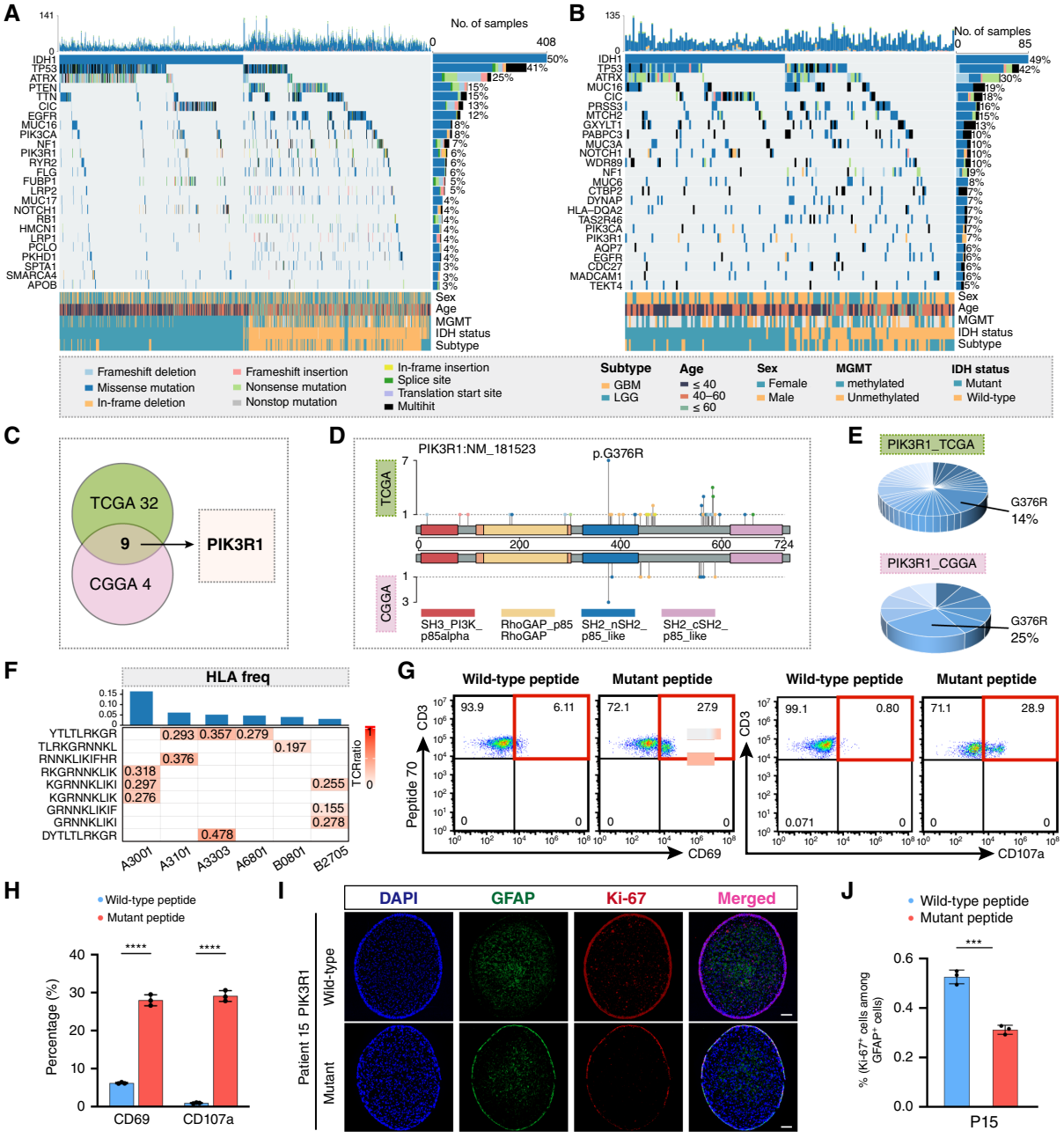

Figure 5.

PIK3R1^{G376R} as a recurrent shared neoantigen candidate in glioblastoma. **A** to **B,** Top 25 high-frequency somatic mutations and corresponding clinical features in TCGA (A) and CGGA (B) primary glioma cohorts. **C,** Venn diagram showing overlapping significantly mutated driver genes identified by MutSigCV in TCGA and CGGA datasets. **D,** Mapping of mutation sites and types across the *PIK3R1* protein in glioma samples from TCGA and CGGA, highlighting the *p.G376R* hotspot mutation. **E,** Distribution of *PIK3R1* mutations in glioma samples, with pie charts indicating the frequency of the recurrent *p.G376R* variant in both cohorts. **F,** Predicted binding potential of 8–11-mer peptides spanning the *PIK3R1*^{G376R} mutation to common HLA class I alleles. Heatmap values represent the TCRratio, reflecting the likelihood of T-cell recognition. The bar plot above shows the HLA allele frequency within the TCGA glioma cohort. These findings support *PIK3R1*^{G376R} as a shared neoantigen candidate with potential population coverage. **G,** Flow cytometry plots showing CD69 and CD107a expression in CD8$^+$ T cells stimulated with wild-type or mutant peptides. **H,** Quantification of CD69$^+$ and CD107a$^+$ CD8$^+$ T cells in response to each peptide. Data are mean ± SEM; ****, $P < 0.0001$ (paired *t* test). **I,** Immunofluorescence images of glioblastoma organoids cocultured with peptide-specific T cells. Ki-67 staining (red) indicates proliferating tumor cells. Mutant peptide–reactive T cells reduced Ki-67$^+$ cells more effectively than wild-type controls. **J,** Quantification of Ki-67$^+$ cells among GFAP$^+$ tumor cells in organoids. Data are shown as mean ± SEM; ***, $P < 0.001$.

In the present study, our work provides a comprehensive framework that combines deep learning–based prediction with advanced *in vitro* validation, addressing two major bottlenecks in neoantigen research: prediction ACC and the physiologic relevance of validation systems. In parallel, the integration of matched multiomics and functional validation data across 21 organoid models allowed us to establish a high-fidelity, high-ACC neoantigen dataset, which may serve as a valuable resource for future prediction model development. These findings may inform future efforts in personalized immunotherapy, particularly in tumors such as glioblastoma, in which robust immune targets and modeling systems are urgently needed.

Our findings also highlight the value of identifying shared neoantigens with therapeutic relevance across patients. Through integrated genomic analysis of large-scale glioma cohorts, we identified *PIK3R1*^{G376R} as a recurrent hotspot mutation and potential shared neoepitope. This mutation exhibited high frequency across independent datasets and produced peptides with strong predicted binding and T-cell recognition potential across multiple common HLA alleles. Considering its high recurrence and immunogenicity, *PIK3R1*^{G376R} represents a compelling candidate for population-targeted neoantigen-based immunotherapies, including T-cell receptor-engineered T cell or shared vaccine development strategies. *TP53* has been reported as one of the most classic and well-known cancer markers; however, its HLA-A*02:01–restricted recognition of mutated *p53 p.R175H* was identified, and the minimal peptide epitope HMTEVVRHC accounts for only about 5% of *TP53* mutations (71). Therefore, future studies could further evaluate the prevalence of *PIK3R1*^{G376R} in pan-cancer contexts and its potential for clinical application.

This study had two main limitations: (i) Our focus was on 9-mer neoantigen peptides, as the literature suggests that 9-mer neoantigen peptides exhibit advantages in MHC binding in GBM; however, we did not verify this in our data, nor did we perform experimental validation, and (ii) we did not validate candidate neoantigen peptides at the clinical level. Although organoid models largely reflect the intratumoral environment of the parental tumor, *in vitro* culture models still differ from the real human environment.

In our future work, we will advance this research. The progress of clinical trials for neoantigen dendritic cell vaccines is highly meaningful. To the best of our knowledge, leading international institutions have advanced this therapy to phase III trials (NCT06077760, NCT05685004, NCT05933577), and mRNA neoantigen vaccines have shown initial success in various cancers and seem to be highly promising in treating GBM (72, 73). Further clinical trials are needed to explore the underlying mechanisms of GBM neoantigens to offer more treatment options to patients with GBM.

Data Availability

Raw WES and RNA-seq data of tumor tissues, blood, and organoids of 21 patients with glioblastoma are deposited at the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa-human/browse/HRA012176). All data requests will be granted by the Data Access Committee. The data analyzed in this study were obtained from the TCGA-GDC portal for the TCGA-GBM and TCGA-LGG cohorts. The data that support the findings of this study are also available from the CGGA data portal, specifically the WES dataset WESeq_286. Correspondence and requests for materials should be addressed to Chen Wang or Ting Sun. The algorithms and code from the article are available at https://github.com/zhangjbig/TCRscore and https://codeocean.com/capsule/7595656/tree/v1.

Authors' Disclosures

No disclosures were reported.

Authors' Contributions

C. Wang: Data curation, validation, investigation, visualization, methodology, writing–original draft. **T. Sun:** Data curation, software, formal analysis, validation, investigation, visualization, methodology, writing–original draft, writing–review and editing. **Y. He:** Validation, investigation. **M. Yu:** Validation, investigation. **C.-Q. Pan:** Validation, investigation. **H. Hu:** Validation, investigation. **Y. Sun:** Data curation, validation. **D. Wang:** Validation, investigation. **Z. Cui:** Investigation. **J. Zhang:** Investigation. **Y. Zhai:** Investigation. **Z. Shi:** Investigation. **Z. Li:** Investigation. **M. Xu:** Investigation. **Y.T. Oh:** Investigation. **T. Jiang:** Supervision. **Z. Xu:** Data curation, software, supervision, validation. **G. Li:** Supervision, validation, investigation, visualization. **J. Zhang:** Conceptualization, resources, data curation, software, formal analysis, supervision, validation, writing–original draft, writing–review and editing. **W. Zhang:** Supervision, funding acquisition, validation, investigation, writing–original draft, project administration, writing–review and editing.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (82525053 for W. Zhang), the Key Joint Project of the Beijing Natural Science Foundation (Z250008 for W. Zhang), the Youth Beijing Scholar Program (No. 055 for W. Zhang), the Beijing Outstanding Young Scientist Program (JWZQ20240101026 for W. Zhang), the Fundamental Research Funds for General Universities (for J. Zhang), the Youth Thousand Scholar Program of China (for J. Zhang), the Research Ward Excellence Clinical Research Program of the Beijing Municipal Health Commission (BRWEP2024W032040200 for W. Zhang), the Dengfeng Talents Program of the Beijing Hospitals Authority (DFL20240503 for W. Zhang), the Beijing Natural Science Foundation (JQ24057 for G. Li), the Yangfan Program of the Beijing Hospitals Authority (ZLRK202314 for W. Zhang), the Beijing Hospital Authority Clinical Medicine Development special funding support, code ZLRK202333 (for J. Zhang), the Excellent Youth Talents Program of Capital Medical University (A2205 for W. Zhang), and the Beijing Municipal Health Commission (JYY2023-2 for W. Zhang). We sincerely thank the healthy donors who contributed PBMCs for this study, including Hao Yan, Jingchen Yang, Bohan Liu, Yiyun Yin, Yankun Chen, Cheng Cheng, Yin Ren, and others. C. Wang, M. Yu, C.-Q. Pan, and Y. Sun conducted the major biological experiments and data acquisition. C. Wang, H. Hu, D. Wang, and others were responsible for sample collection, statistical analysis, and additional material preparation. W. Zhang, J. Zhang, G. Li, and Z. Xu supervised the overall project, designed the study, guided the experimental and analytic procedures, and were responsible for manuscript revision and correspondence.

Note

Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Received June 24, 2025; revised October 22, 2025; accepted March 16, 2026; posted first March 26, 2026.

References

1. Weller M, Wen PY, Chang SM, Dirven L, Lim M, Monje M, et al. Glioma. Nat Rev Dis Primers 2024;10:33.
2. Wilson L, Atallah A. Veliparib plus temozolomide for MGMT-methylated glioblastoma. JAMA Oncol 2025;11:569.
3. Zhang C, Neha, Zhang J, Dhaha P, Li X, Mishra SK, et al. Aging and senescence: key players in brain tumor progression and drug resistance. Drug Resist Updat 2025;81:101228.
4. Dunn GP, Sherpa N, Manyanga J, Johanns TM. Considerations for personalized neoantigen vaccination in Malignant glioma. Adv Drug Deliv Rev 2022;186:114312.
5. Woroniecka K, Fecci PE. Immuno-synergy? Neoantigen vaccines and checkpoint blockade in glioblastoma. Neuro Oncol 2020;22:1233–4.
6. Keskin DB, Anandappa AJ, Sun J, Tirosh I, Mathewson ND, Li S, et al. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. Nature 2019;565:234–9.

7. Hilf N, Kuttruff-Coqui S, Frenzel K, Bukur V, Stevanović S, Gouttefangeas C, et al. Actively personalized vaccination trial for newly diagnosed glioblastoma. Nature 2019;565:240–5.
8. Dunn GP, Cloughesy TF, Maus MV, Prins RM, Reardon DA, Sonabend AM. Emerging immunotherapies for malignant glioma: from immunogenomics to cell therapy. Neuro Oncol 2020;22:1425–38.
9. Huber F, Arnaud M, Stevenson BJ, Michaux J, Benedetti F, Thevenet J, et al. A comprehensive proteogenomic pipeline for neoantigen discovery to advance personalized cancer immunotherapy. Nat Biotechnol 2025;43:1360–72.
10. Shemesh CS, Hsu JC, Hosseini I, Shen BQ, Rotte A, Twomey P, et al. Personalized cancer vaccines: clinical landscape, challenges, and opportunities. Mol Ther 2021;29:555–70.
11. Li B, Jing P, Zheng G, Pi C, Zhang L, Yin Z, et al. Neo-intline: integrated pipeline enables neoantigen design through the in-silico presentation of T-cell epitope. Signal Transduct Target Ther 2023;8:397.
12. Paul S, Croft NP, Purcell AW, Tscharke DC, Sette A, Nielsen M, et al. Benchmarking predictions of MHC class I restricted T cell epitopes in a comprehensively studied model system. PLoS Comput Biol 2020;16:e1007757.
13. Miller AM, Koşaloğlu-Yalçın Z, Westernberg L, Montero L, Bahmanof M, Frentzen A, et al. A functional identification platform reveals frequent, spontaneous neoantigen-specific T cell responses in patients with cancer. Sci Transl Med 2024;16:eabj9905.
14. Zhou W, Xiang W, Yu J, Ruan Z, Pan Y, Wang K, et al. NeoTCR: an immunoinformatic database of experimentally-supported functional neoantigen-specific TCR sequences. Genomics Proteomics Bioinformatics 2025;23:qzae010.
15. Qu W, You R, Mamitsuka H, Zhu S. DeepMHCI: an anchor position-aware deep interaction model for accurate MHC-I peptide binding affinity prediction. Bioinformatics 2023;39:btad551.
16. Dimitrov I, Garnev P, Flower DR, Doytchinova I. Peptide binding to the HLA-DRB1 supertype: a proteochemometrics analysis. Eur J Med Chem 2010;45:236–43.
17. Church C, Pulliam T, Longino N, Park SY, Smythe KS, Makarov V, et al. Transcriptional and functional analyses of neoantigen-specific CD4 T cells during a profound response to anti-PD-L1 in metastatic Merkel cell carcinoma. J Immunother Cancer 2022;10:e005328.
18. Danisch S, Slabik C, Cornelius A, Albanese M, Tagawa T, Chen Y-FA, et al. Spatiotemporally skewed activation of programmed cell death receptor 1-Positive T cells after Epstein-Barr virus infection and tumor development in long-term fully humanized mice. Am J Pathol 2019;189:521–39.
19. Neal JT, Li X, Zhu J, Giangarra V, Grzeskowiak CL, Ju J, et al. Organoid modeling of the tumor immune microenvironment. Cell 2018;175:1972–88.e16.
20. Wang C, Yu M, Zhang W. Neoantigen discovery and applications in glioblastoma: an immunotherapy perspective. Cancer Lett 2022;550:215945.
21. Wang W, Yuan T, Ma L, Zhu Y, Bao J, Zhao X, et al. Hepatobiliary tumor organoids reveal HLA class I neoantigen landscape and antitumoral activity of neoantigen peptide enhanced with immune checkpoint inhibitors. Adv Sci (Weinh) 2022;9:e2105810.
22. Mamrosh JL, Sherman DJ, Cohen JR, Johnston JA, Joubert MK, Li J, et al. Quantitative measurement of the requirement of diverse protein degradation pathways in MHC class I peptide presentation. Sci Adv 2023;9:eade7890.
23. Gowthaman R, Pierce BG. TCR3d: the T cell receptor structural repertoire database. Bioinformatics 2019;35:5323–5.
24. Zhang W, Wang L, Liu K, Wei X, Yang K, Du W, et al. PIRD: Pan immune repertoire database. Bioinformatics 2020;36:897–903.
25. Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Res 2018;46:D419–27.
26. Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. Bioinformatics 2017;33:2924–9.
27. Joglekar AV, Leonard MT, Jeppson JD, Swift M, Li G, Wong S, et al. T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. Nat Methods 2019;16:191–8.
28. Chen G, Yang X, Ko A, Sun X, Gao M, Zhang Y, et al. Sequence and structural analyses reveal distinct and highly diverse human $CD8^+$ TCR repertoires to immunodominant viral antigens. Cell Rep 2017;19:569–83.
29. Hansen M, Kilk K, Langel U. Predicting cell-penetrating peptides. Adv Drug Deliv Rev 2008;60:572–9.
30. Doytchinova IA, Flower DR. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine 2007;25:856–66.
31. Hellberg S, Sjöström M, Skagerberg B, Wold S. Peptide quantitative structure-activity relationships, a multivariate approach. J Med Chem 1987;30:1126–35.
32. Spänig S, Heider D. Encodings and models for antimicrobial peptide classification for multi-resistant pathogens. BioData Min 2019;12:7.
33. Sandberg M, Eriksson L, Jonsson J, Sjöström M, Wold S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. J Med Chem 1998;41:2481–91.
34. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intelligence Res 2002;16:321–57.
35. Chowell D, Krishna C, Pierini F, Makarov V, Rizvi NA, Kuo F, et al. Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. Nat Med 2019;25:1715–20.
36. Nielsen JS, Sedgwick CG, Shahid A, Zong Z, Brumme ZL, Yu S, et al. Toward personalized lymphoma immunotherapy: identification of common driver mutations recognized by patient CD8+ T cells. Clin Cancer Res 2016;22:2226–36.
37. Zhang X, Kim S, Hundal J, Herndon JM, Li S, Petti AA, et al. Breast cancer neoantigens can induce $CD8^+$ T-Cell responses and antitumor immunity. Cancer Immunol Res 2017;5:516–23.
38. Tran E, Ahmadzadeh M, Lu YC, Gros A, Turcotte S, Robbins PF, et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. Science 2015;350:1387–90.
39. Martin SD, Wick DA, Nielsen JS, Little N, Holt RA, Nelson BH. A library-based screening method identifies neoantigen-reactive T cells in peripheral blood prior to relapse of ovarian cancer. Oncoimmunology 2017;7:e1371895.
40. Chen F, Zou Z, Du J, Su S, Shao J, Meng F, et al. Neoantigen identification strategies enable personalized immunotherapy in refractory solid tumors. J Clin Invest 2019;129:2056–70.
41. Yao P, Gao M, Hu W, Wang J, Wang Y, Wang Q, et al. Proteogenomic analysis identifies neoantigens and bacterial peptides as immunotherapy targets in colorectal cancer. Pharmacol Res 2024;204:107209.
42. Garcia-Marquez MA, Thelen M, Bauer E, Maas L, Wennhold K, Lehmann J, et al. Germline homozygosity and allelic imbalance of HLA-I are common in esophagogastric adenocarcinoma and impair the repertoire of immunogenic peptides. J Immunother Cancer 2024;12:e007268.
43. Xu H, Hu R, Dong X, Kuang L, Zhang W, Tu C, et al. ImmuneApp for HLA-I epitope prediction and immunopeptidome analysis. Nat Commun 2024;15:8926.
44. Gfeller D, Schmidt J, Croce G, Guillaume P, Bobisse S, Genolet R, et al. Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 $CD8^+$ T-cell epitopes. Cell Syst 2023;14:72–83.e5.
45. Jiang D, Xi B, Tan W, Chen Z, Wei J, Hu M, et al. NeoaPred: a deep-learning framework for predicting immunogenic neoantigen based on surface and structural features of peptide-human leukocyte antigen complexes. Bioinformatics 2024;40:btae547.
46. O'Donnell TJ, Rubinsteyn A, Laserson U. MHCflurry 2.0: improved pan-allele prediction of MHC class I-Presented peptides by incorporating antigen processing. Cell Syst 2020;11:42–8.e7.
47. Sarkizova S, Klaeger S, Le PM, Li LW, Oliveira G, Keshishian H, et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. Nat Biotechnol 2020;38:199–209.
48. Corkum CP, Ings DP, Burgess C, Karwowska S, Kroll W, Michalak TI. Immune cell subsets and their gene expression profiles from human PBMC isolated by Vacutainer Cell Preparation Tube (CPT) and standard density gradient. BMC Immunol 2015;16:48.
49. Jacob F, Salinas RD, Zhang DY, Nguyen PTT, Schnoll JG, Wong SZH, et al. A patient-derived glioblastoma organoid model and biobank recapitulates Inter- and intra-tumoral heterogeneity. Cell 2020;180:188–204.e22.
50. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20.
51. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213–9.
52. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568–76.

53. Fan Y, Xi L, Hughes DS, Zhang J, Zhang J, Futreal PA, et al. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. Genome Biol 2016;17:178.
54. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics 2012;28:311–7.
55. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012;28:1811–7.
56. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
57. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol 2016;12:e1004873.
58. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics 2014;30:3310–6.
59. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019;37:907–15.
60. Gurung HR, Heidersbach AJ, Darwish M, Chan PPF, Li J, Beresini M, et al. Systematic discovery of neoepitope-HLA pairs for neoantigens shared among patients and tumor types. Nat Biotechnol 2024;42:1107–17.
61. Chen S, Huang T, Wen T, Li H, Xu M, Gu J. MutScan: fast detection and visualization of target mutations by scanning FASTQ data. BMC Bioinformatics 2018;19:16.
62. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro Oncol 2021;23:1231–51.
63. O'Brien H, Salm M, Morton LT, Szukszto M, O'Farrell F, Boulton C, et al. Breaking the performance ceiling for neoantigen immunogenicity prediction. Nat Cancer 2023;4:1618–21.
64. Rosenberg SA, Parkhurst MR, Robbins PF. Adoptive cell transfer immunotherapy for patients with solid epithelial cancers. Cancer Cell 2023;41:646–8.
65. Yang SF, Weng MT, Liang JD, Chiou LL, Hsu YC, Lee YT, et al. Neoantigen vaccination augments antitumor effects of anti-PD-1 on mouse hepatocellular carcinoma. Cancer Lett 2023;563:216192.
66. Wang Z, Gu Y, Sun X, Huang H. Computation strategies and clinical applications in neoantigen discovery towards precision cancer immunotherapy. Biomark Res 2025;13:96.
67. Bagchi S, Yuan R, Engleman EG. Immune checkpoint inhibitors for the treatment of cancer: clinical impact and mechanisms of response and resistance. Annu Rev Pathol 2021;16:223–49.
68. Arrieta VA, Dmello C, McGrail DJ, Brat DJ, Lee-Chang C, Heimberger AB, et al. Immune checkpoint blockade in glioblastoma: from tumor heterogeneity to personalized treatment. J Clin Invest 2023;133:e163447.
69. Albelda SM. CAR T cell therapy for patients with solid tumours: key lessons to learn and unlearn. Nat Rev Clin Oncol 2024;21:47–66.
70. Logun M, Wang X, Sun Y, Bagley SJ, Li N, Desai A, et al. Patient-derived glioblastoma organoids as real-time avatars for assessing responses to clinical CAR-T cell therapy. Cell Stem Cell 2025;32:181–90.e4.
71. Lo W, Parkhurst M, Robbins PF, Tran E, Lu YC, Jia L, et al. Immunologic recognition of a shared p53 mutated neoantigen in a patient with metastatic colorectal cancer. Cancer Immunol Res 2019;7:534–43.
72. Rojas LA, Sethna Z, Soares KC, Olcese C, Pang N, Patterson E, et al. Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer. Nature 2023;618:144–50.
73. Liau LM, Ashkan K, Brem S, Campian JL, Trusheim JE, Iwamoto FM, et al. Association of autologous tumor lysate-loaded dendritic cell vaccination with extension of survival among patients with newly diagnosed and recurrent glioblastoma: a phase 3 prospective externally controlled cohort trial. JAMA Oncol 2023;9:112–21.