SYSTEMS BIOLOGY

Immunopeptidomics-guided discovery and characterization of neoantigens for personalized cancer immunotherapy

Yangyang Cai[1]†, Manyu Gong[2]†, Mengqian Zeng[1]†, Feng Leng[1]†, Dezhong Lv[1], Jiyu Guo[3], Hao Wang[2], Yapeng Li[4], Quan Lin[5], Jing Jing[5], Ying Zhang[2]*, Juan Xu[1]*, Yongsheng Li[3,5]*

Neoantigens have emerged as ideal targets for personalized cancer immunotherapy. We depict the pan-cancer peptide atlas by comprehensively collecting immunopeptidomics from 531 samples across 14 cancer and 29 normal tissues, and identify 389,165 canonical and 70,270 noncanonical peptides. We reveal that noncanonical peptides exhibit comparable presentation levels as canonical peptides across cancer types. Tumor-specific peptides exhibit significantly distinct biochemical characteristics compared with those observed in normal tissues. We further propose an immunopeptidomic-guided machine learning-based neoantigen screening pipeline (MaNeo) to prioritize neo-peptides as immunotherapy targets. Benchmark analysis reveals MaNeo results in the accurate identification of shared and tumor-specific canonical and noncanonical neo-peptides. Last, we use MaNeo to detect and validate three neo-peptides in cancer cell lines, which can effectively induce increased proliferation of active T cells and T cell responses to kill cancer cells but not damage healthy cells. The pan-cancer peptide atlas and proposed MaNeo pipeline hold great promise for the discovery of canonical and noncanonical neoantigens for cancer immunotherapies.

INTRODUCTION

Neoantigen-mediated cancer immunotherapies, such as T cell therapy and vaccines, induce $CD8^+$ cytotoxic T cell responses to kill tumor cells by recognizing human lymphocyte antigen (HLA)–presented antigens (*1*, *2*). Now, neoepitopes predominantly originate from nonsynonymous mutations within the cancer genome. However, "cold" tumors are subject to a relatively low number of mutations and a sparse occurrence of neoepitopes in patients with cancer, often resulting in a lack of suitable therapeutic targets (*3*, *4*). Accumulating evidence suggests that immunogenic epitopes can also arise from putative translation products of genomic noncoding regions and from noncanonical translation, including untranslated regions (UTRs) and alternative translations of protein-coding genes and noncoding genes, which are called noncanonical epitopes (*5*). Noncanonical epitopes have the potential to drive antitumor immunity without the risk of on-target off-tumor adverse effects, leading to a more precise and safer treatment strategy for patients with various types of cancer (*6*, *7*). Thus, noncanonical epitopes have emerged as promising immunotherapeutic targets.

Traditional proteomic techniques can identify peptides that specifically occur in tumor tissues, accelerating the recognition of tumor neoantigens. However, determining which peptides are bound by major histocompatibility complex (MHC) molecules is not possible (*8*). Another commonly used approach relies on the identification of HLA-bound peptides by immunoprecipitation from the surface of cells and elution the bound peptides before mass spectrometry (MS) (named the immunopeptidome) (*9*). In contrast to traditional proteomes, the immunopeptidome provides the real-time landscape of epitopes. In addition, current proteomics reduce the potential search space to only peptides with either lysine or arginine terminal residues (*8*), whereas immunopeptidome may have any terminal residue. The immunopeptidome coupled with diverse customized theoretical reference libraries has revealed the existence of various MHC-bound epitopes derived from different alterations, such as genomic variants, aberrantly expressed canonical tumor-specific antigens, noncanonical translation, alternative splicing, and defective ribosomal products (*10*–*12*). An immunopeptidome study of mismatch repair-deficient colorectal cancer patients demonstrated that neoantigens originating from noncoding regions were capable of eliciting adaptive immune responses (*13*). We also characterized the immunopeptidome landscape by integrating multi-omics data in acute myeloid leukemia (AML) and identified both canonical and noncanonical MHC-associated peptides (*4*). Immunopeptidome analysis will help researchers accurately identify distinct classes of tumor neoantigens with higher resolution and lower noise, which is crucial for the development of personalized cancer immunotherapy (*6*).

Although several studies have been performed on nonmutated neoantigens, the majority of methods that rank neoantigens rely only on the expression levels of peptides or source genes on the basis of RNA sequencing (RNA-seq) data from tumor samples and normal samples from the Genotype-Tissue Expression (GTEx) project (*14*, *15*). However, overwhelming evidence has demonstrated that RNA-seq data cannot directly reveal translating open reading frames (ORFs) and mimic the degradation of peptides in vivo (*16*, *17*). In addition, the antigenic repertoires from diverse samples exhibit high heterogeneity (*18*). Heterogeneity is strongly influenced by various factors, including the type of tumor, the genetic background

[1]State Key Laboratory of Frigid Zone Cardiovascular Diseases (SKLFZCD), College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China. [2]Department of Pharmacology (Key Laboratory of Cardiovascular Medicine Research, Ministry of Education), College of Pharmacy, Harbin Medical University, Harbin, Heilongjiang 150081, PR China. [3]State Key Laboratory of Frigid Zone Cardiovascular Diseases (SKLFZCD), School of Interdisciplinary Medicine and Engineering, Harbin Medical University, Harbin, Heilongjiang 150081, China. [4]The Second Affiliated Hospital of Harbin Medical University, Harbin 150081, China. [5]Department of Radiation Oncology, Harbin Medical University Cancer Hospital, Harbin, Heilongjiang 150040, China.
*Corresponding author. Email: liyongsheng@ems.hrbmu.edu.cn (Yongsheng Li); xujuanbiocc@ems.hrbmu.edu.cn (J.X.); jennying223@hrbmu.edu.cn (Y.Z.)
†These authors contributed equally to this work.

of the host, and the tumor microenvironment. Consequently, the identification and characterization of neoantigens require the integration of a comprehensive immunopeptidome and proteome from patients with different tumors and healthy tissues.

To fill these gaps, we collected and reanalyzed the publicly available immunopeptidomic datasets encompassing 14 tumors and 29 normal tissues, against the reviewed protein sequences and customized libraries, which were constructed on the basis of nonredundant canonical and noncanonical ORFs with "NTG" start codons from the Ribosome profiling sequencing (Ribo-seq) data. We detected 389,165 canonical and 70,270 noncanonical peptides across cancer types and normal tissues. The presentation levels of the canonical and noncanonical peptides were comprehensively characterized. Furthermore, we classified peptides into tumor and normal peptides and examined their differences in various physical and chemical properties. We next proposed MaNeo, a machine learning-based pipeline for prioritizing neo-peptides as candidate immunotherapy targets on the basis of various features of cancer. In total, 14 neo-peptides were prioritized with MaNeo and the immunogenicities of three neo-peptides were functionally validated in several cancer cell lines. Orthogonal integration of available immunopeptidomics data will help find and further prioritize neoantigens for personalized cancer immunotherapy.

RESULTS

Immunopeptidomics mapping reveals noncanonical peptides in cancer

To construct a comprehensive peptide atlas of cancer and normal samples, we developed a pipeline to integrate publicly available immunopeptidome data from cancer, adjacent, and healthy normal samples (fig. S1). In total, we constructed the tumor-normal peptide atlas encompassed 308,703 peptides from 14 different cancer types and 151,258 peptides from 29 normal tissues across 531 samples from 43 datasets (Fig. 1A; fig. S2, A to C; and table S1). We found that skin cutaneous melanoma (SKCM) was the cancer type with the most abundant peptides, followed by meningioma (MG), ovarian cancer (OV), and breast cancer (BRCA). There was no evident association between the number of peptides and the number of samples (fig. S2D; $R = 0.34$, $P = 0.24$). Next, the constructed tumor-normal peptide atlas was exploited to prioritize neo-peptides in cancer. We annotated the types of peptides according to the genome regions, characterized differential features across diverse cancers and tissue types, and further proposed a machine-learning model to screen neo-peptides (Fig. 1B). Last, we developed an online webserver (Immunogenic Search Engine, ISE) to help researchers prioritize neo-peptides in cancer.

The constructed atlas spanned 107 HLA alleles, including 31 HLA-A, 49 HLA-B, and 27 HLA-C alleles, which collectively covered 98.67, 92.73, and 98.29% of the worldwide population, respectively (fig. S2C). We found that different allele sets were highly coincident between tumor and normal samples (Fig. 1C). In addition, the majority of alleles were found to be associated with diverse cancer types (fig. S2E). On the basis of pan-cancer mass spectrometry–based immunopeptidome, we found that noncanonical peptides originating from noncoding regions are processed and presented by HLA-I on the cell surface. In total, we identified 70,720 unique noncanonical peptides, which were derived mainly from the translation of long noncoding RNAs (lncRNAs), 5′ upstream open reading frames (uORFs) and 3′ dORFs (Fig. 1D), accounting for approximately 15% of all HLA-presented peptides. Notably, 14% of the peptides were derived from out-of-frame regions of protein-coding genes, suggesting that a canonical protein may not be the unique translation product of an annotated coding sequence (CDS). This result was consistent with previous observations that tryptophan-associated accumulation of ribosomes could lead to ribosomal frame shifting and generate out-of-frame peptides (*19*). Although the majority of genes derived more canonical peptides, 683 genes were identified that can produce a greater number of noncanonical peptides (Fig. 1E and table S2), including TPM1 and its paralog TPM3, which stabilize the cytoskeleton and inhibit tumor cell migration and invasion (*20*). We next selected the top 10 encoded-peptide genes from tumor and normal tissues and found that each gene generated noncanonical peptides despite fewer than canonical peptides. In addition, genes had a high degree of overlap across different tumor and normal tissues (fig. S3, A and B, and table S3). These results suggest that numerous antigens in tumor cells may be by-products of normal cellular proliferation, limiting their potential as therapeutic cancer vaccines due to the risk of side effects (*7*). Notably, among these genes, Protein Kinase, DNA-Activated, Catalytic Polypeptide (PRKDC) is notably prevalent across tumors and serves not only as a predictive biomarker but also as a drug target for immune checkpoint inhibitors (*21*).

The expression levels of overlapping genes between tumor and normal tissues were further characterized via the The Cancer Genome Atlas (TCGA) and GTEx datasets. Our analysis revealed that these genes were expressed in distinct tissue types. Notably, the expression of source genes could partially reflect the presence of peptides in tissues because several genes are capable of encoding a large number of peptides despite exhibiting low expression levels in those tissues (fig. S3C). For example, titin gene (TTN) was highly expressed in the GTEx RNA-seq data for heart and muscle tissues (*22*); however, immunopeptidome analysis indicated that TTN not only generated plenty of peptides in heart and muscle tissues but also generated hundreds of peptides in the tongue and esophagus (fig. S3D). Similarly, apolipoprotein B (APOB) was predominantly expressed in the liver and rarely expressed in the lung (*23*); however, it derived numerous peptides in both tissues (fig. S3D). Furthermore, we expanded the findings to include entire peptide-associated genes and retained those with transcripts per million ≤ 1 across all GTEx tissues. We found that most genes could still derive peptides in normal tissues (fig. S4). These results indicated that the immunopeptidome could effectively detect noncanonical peptides that the RNA-seq data inadequately captured the peptide presentation landscape across cancer and normal tissues. In contrast, immunopeptidome analysis can directly identify HLA-presented peptides with low RNA expression levels, particularly noncanonical peptides.

Comparable presentation levels of canonical and noncanonical peptides

Cancer antigens derived from noncoding genomic regions, which have attracted increasing interest, represent the majority of targetable tumor-specific antigens capable of eliciting antitumor immune responses (*24*, *25*). Similar tissue distributions were observed between noncanonical and canonical peptides across various tissues (Fig. 2A). Each tissue contributes a unique set of peptides,

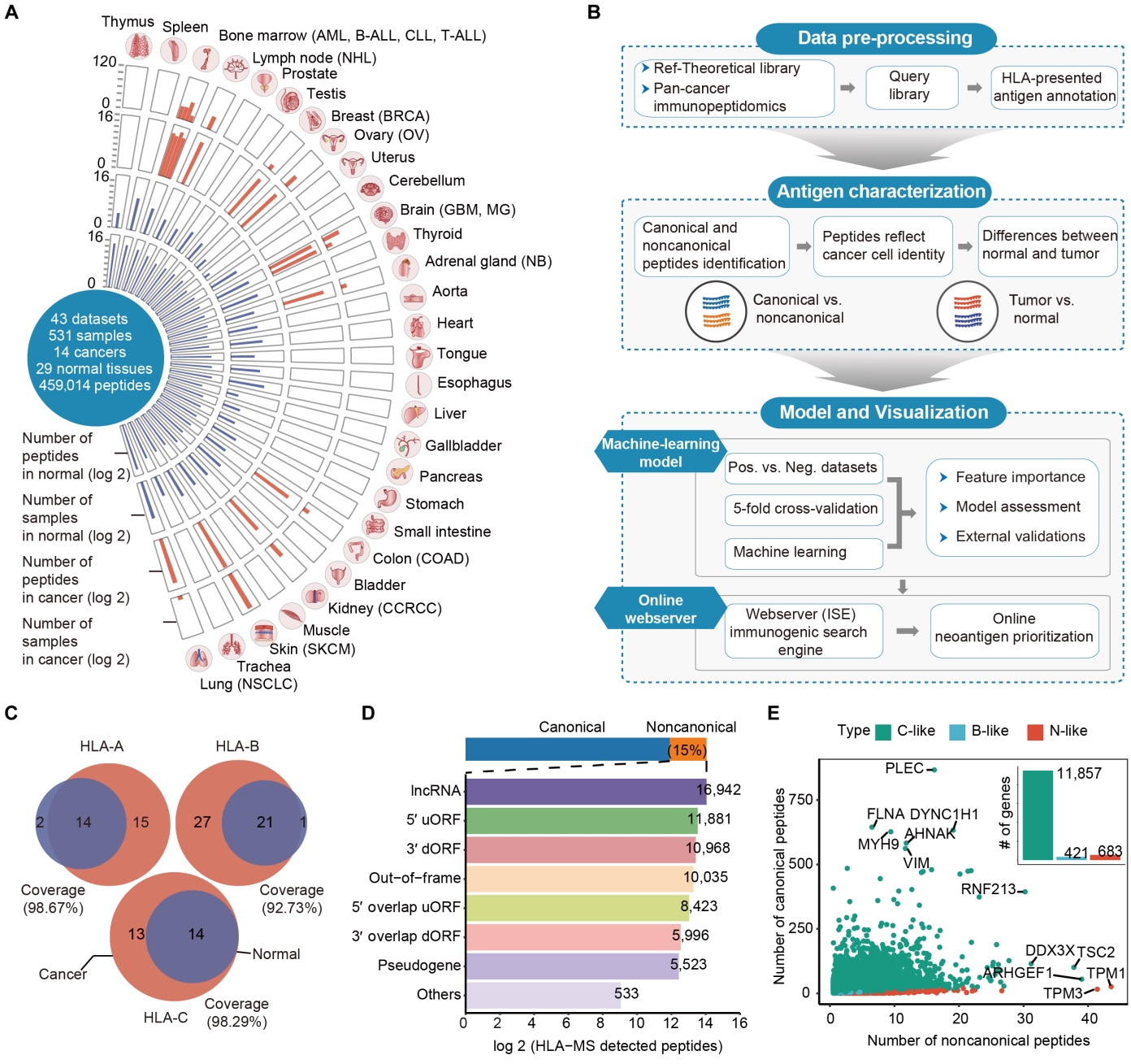

Fig. 1. Overview of the HLA-presented tumor-normal immunopeptidome atlas. (**A**) Overview of the immunopeptidome cohort across the 14 cancer types and 29 normal tissues. AML: Acute myeloid leukemia, B-ALL: B cell acute lymphoblastic leukemia, T-ALL: T cell acute lymphoblastic leukemia, CLL: Chronic lymphocytic leukemia, NHL: Non-Hodgkin's lymphoma, BRCA: Breast cancer, OV: Ovarian cancer, GBM: Glioblastoma, M.G.: Meningioma, NB: Neuroblastoma, COAD: Colon adenocarcinoma, CCRCC: Clear cell renal carcinoma, SKCM: Skin cutaneous melanoma, NSCLC: Non–small cell lung cancer. (**B**) Workflow for the construction of tumor-normal immunopeptidome atlas. (**C**) Coverage of HLA-A, B, and C alleles among the worldwide population, respectively. (**D**) Number of peptides among diverse ORF types. (**E**) Number of gene-encoded canonical and non-canonical peptides. C-like represents the gene deriving canonical peptides more than noncanonical ones. B-like represents the gene deriving canonical peptides equal to noncanonical ones. N-like represents the gene deriving canonical peptides less than noncanonical ones. lncRNA, long noncoding RNA.

with sex organs being the largest contributor to unique sequences. Furthermore, the brain, as the most complex and heterogeneous tissue, exhibited more heterogeneity at the peptide level. A previous study investigated the different spatial regions of gliomas in the same sample and reported marked heterogeneity of tumor neoantigens (*26*).

Consistent with previous studies (*27*), all the peptides presented similar length distributions ranging from 8 to 12, with a modal length of 9 (Fig. 2B). Biochemical features provide valuable insights into the molecular properties of peptides, allowing for a deeper understanding of their stable binding with HLA-I molecules and recognition by T cell receptors (TCRs). To evaluate the biochemical

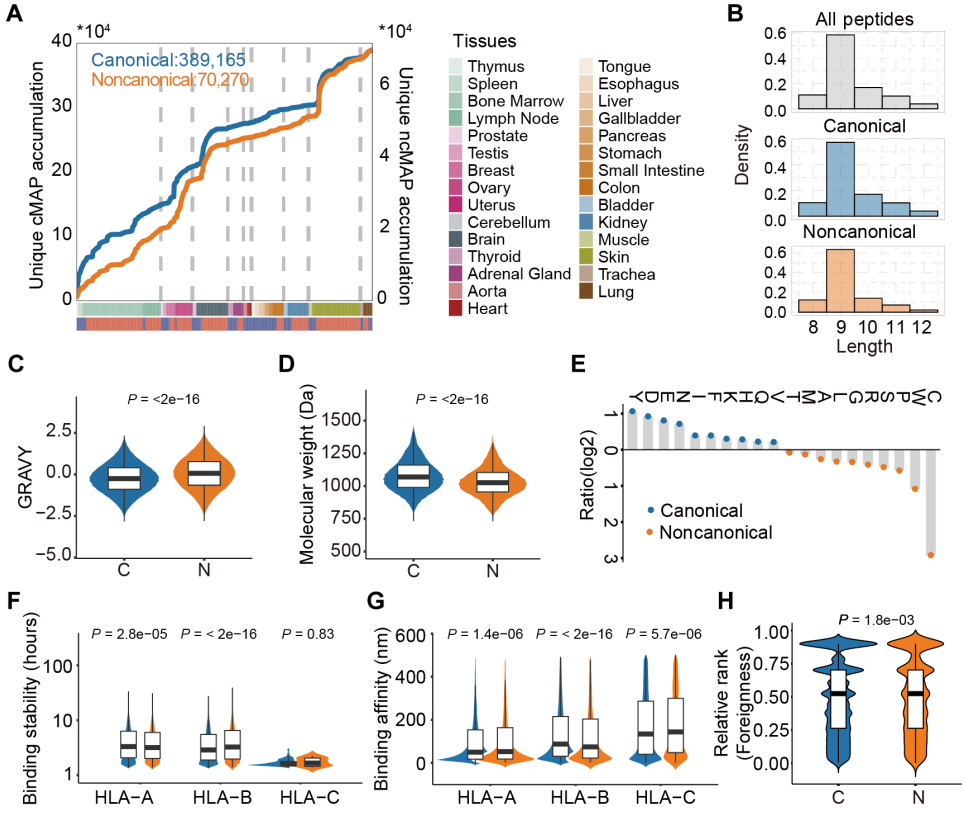

Fig. 2. Comparison of distinct aspects of canonical and noncanonical peptides. (**A**) Accumulated unique canonical and non-canonical peptide distributions from each tissue. (**B**) Comparison of the length distributions of the canonical and noncanonical peptides. (**C**) GRAVY is shown for the canonical (n = 389,165) and noncanonical peptides (n = 70,270) with the Wilcoxon rank sum test. (**D**) Molecular weights of the peptides were compared via the Wilcoxon rank sum test (389,165 canonical versus 70,270 noncanonical peptides). (**E**) Usage tendency of different amino acids in canonical and noncanonical peptides. The orange color indicates that the amino acid is enriched in noncanonical peptides, whereas the blue color indicates that the amino acid is enriched in canonical peptides. (**F** to **H**) Presentation levels between canonical and noncanonical peptides are from (F) binding stability (HLA-A: 101,200 canonical versus 10,455 noncanonical, HLA-B: canonical: 102,232 versus 7235 noncanonical, HLA-C: 163 canonical versus 7 noncanonical), (G) binding affinity (HLA-A: 160,290 canonical versus 17,459 noncanonical, HLA-B: canonical: 166,184 versus 12,116 noncanonical, HLA-C: 71,012 canonical versus 5742 noncanonical) and (H) the relative rank of foreignness (TCR recognition probability homologous to known pathogen-derived peptides in the IEDB) (149,130 canonical versus 24,852 noncanonical). Wilcoxon rank sum test was used to compare the differences between groups.

features of the canonical and noncanonical peptides, we next analyzed the amino acid composition (AAC) and known determinants of immunogenicity. The noncanonical peptides presented significantly higher GRAVY index scores (Fig. 2C; $P < 2 \times 10^{-16}$), indicating a preference for noncanonical peptides for antigen presentation and recognition by T cells (*28*). A notable bias toward the usage of small-molecular weight amino acids was observed within noncanonical peptides (Fig. 2D; $P < 2 \times 10^{-16}$).

In addition, compared with canonical peptides, noncanonical peptides generally contain more cysteine and tryptophan residues at each position (Fig. 2E). The incorporation of tryptophan residues within epitopes plays an important role in modulating the immunogenicity of antigens. Owing to its highly hydrophobic nature and distinctive aromatic binuclear ring structure, tryptophan serves as a critical residue in the design of vaccines and diagnostic reagents. Its unique physicochemical properties not only enhance antigen-antibody interactions but also contribute to the stabilization of epitope structures, thereby optimizing immune recognition and response (*29*). In contrast, the canonical peptides presented acidic residues (that is, aspartate and glutamate) at a higher frequency than did the noncanonical peptides (Fig. 2E and fig. S5, A and B). We also examined the presentation levels of canonical and noncanonical peptides from various perspectives. The binding stability was next inferred with NetMHCStabpan (v.1.0) with default parameters (*30*). The binding of noncanonical peptides to HLA-B alleles was more stable than that of canonical peptides (Fig. 2F), suggesting that HLA alleles are capable of presenting noncanonical peptides on the cell membrane through a series of transport processes. Binding affinity was also assessed via NetMHCpan (v.4.1) (*31*), MHCflurry (v2.0) (*32*), and MixMHCpred (v2.1) (*33*). We found that noncanonical peptides that bind to HLA-B alleles were relatively close (Fig. 2G and fig. S5, C and D). We also evaluated the probability of peptides being recognized by TCRs with the multistate thermodynamic model proposed recently (*34*). Our findings revealed a similar distribution of probabilities between canonical and noncanonical peptides (Fig. 2H), indicating the presence of noncanonical peptides that can trigger a cytotoxic T cell response. Together, these findings suggest that noncanonical peptides have comparable presentation levels, enrich the diversity, and extend the repertoire of tumor antigens.

HLA-presented peptides reflect cancer cell identity

To improve our understanding of the specificity of HLA-presented peptides, we conducted a comprehensive dissection of sequence

characteristics for peptides detected in cancerous and normal tissues. First, we categorized peptides exclusive to tumor samples as tumor peptides, whereas the remaining peptides were classified as normal (Fig. 3A). Notably, approximately 73 and 58% of the peptides were presented in only one cancer or normal tissue, respectively. However, many peptides were present in several tumors and normal tissues, accounting for 23.0 to 73.2% (median: 56.7%) and 53.5 to 95.8% (median: 90.5%), respectively (Fig. 3B and fig. S6, A and B). We next performed unsupervised clustering based on position frequency vectors of peptides across all tumor and normal tissues. We observed a clear separation between tumor and normal tissues, with greater similarity observed among normal tissues than among tumors (Fig. 3C, figs. S7A and S8A, and taqble S4). Furthermore, principal components analysis (PCA) confirmed this observation, revealing obvious differences in the HLA-presented peptides across various tissue types (Fig. 3D and figs. S7B and S8B). We also inspected motif shifts between tumor and normal peptides and found that the motif "xExxxxxxx" was shared across normal tissues, whereas the motif "xxxxxxxxK/R" was shared across tumors (fig. S9).

We next explored the usage of amino acids under different conditions. As expected, tumor peptides presented a greater content of positively charged residues (histidine, lysine, and arginine) than did normal peptides. This observation may be attributed to the distinct metabolic reprogramming of tumor cells, which often results in an acidic tumor microenvironment, in contrast to the physiological conditions of noncancerous cells (*35*, *36*). Moreover, cysteine was also enriched in eight types of cancer, highlighting the importance of incorporating noncanonical antigens (Figs. 2E and 3E, and figs. S7C and S8C). The amino acids exhibited notable differences at particular positions across various cancer types. For example, enrichment of arginine was frequently observed at positions 2 and 9 in AML, MG, and NHL (fig. S6C), indicating the presence of primary anchors. The anchoring sites of peptides are critical for binding to HLAs. Compared with normal peptides, tumor peptides tend to use combinations of basic-basic amino acids (Fig. 3F). Conversely, normal peptides exhibited the opposite bias, favoring acidic-acidic dipeptide composition (DPC) (Fig. 3F and figs. S7D and S8D). In particular, we found that cysteine-X and X-cysteine junctions were prevalent across five cancer types, including CLL, NHL, OV, MG, and CCRCC (fig. S6D). In addition, the biochemical properties of HLA-presented peptides—such as GRAVY, net charge, and isoelectric point (pI)—are crucial for their transport from the intracellular environment to the cell surface (*37*). The tumor peptides presented significant differences in their distributions of biochemical properties, characterized by low hydrophobicity, high net charge, and high pI (Fig. 3G and figs. S7E and S8E; $P < 2.2 \times 10^{-16}$). These results suggest that we can use the sequence features of peptides to reflect cell identity, as these peptides display obvious differences in different microenvironments.

Machine learning–based prioritization of immunogenic neo-peptides

To accurately identify neo-peptides exclusively derived from tumor cells, we developed a machine learning–based neoantigen screening pipeline (MaNeo) guided by the collected immunopeptidomes (Fig. 4A). For each cancer type, we constructed machine learning models against 29 normal tissues. In total, we generated 406 cancer- and tissue-specific models for each encoding mode and two machine learning–based classifiers [Random forest (RF) and XGBoost], which were designed to prioritize the neo-peptides. To assess the performance of MaNeo, we used multiple metrics, including the area under the receiver operating characteristic curve (AUROC), the area under the precision–recall curve (AUPRC), sensitivity, specificity, precision, weighted F1 score, and balanced accuracy. The performance of different models based on diverse features varied greatly, with the multicode model demonstrating superior results compared with the one-hot encoding approach, which in turn outperformed the peptide to MHC binding energy covariance matrices (PMBEC) model (Fig. 4B). All of these encodings were found to be more effective than the Blosum62 method (Fig. 4B). Notably, the multicode models exhibited high precision while maintaining high recall, indicating their ability to identify true positives accurately. We also calculated the standard deviations of model metrics and demonstrated exceptional stability, with the maximum SD remaining below 0.1 (fig. S10). Furthermore, the multicode models consistently presented high and robust area under the curve (AUC) values, with >99% of the models achieving AUROC values > 0.8 and averages of 0.880 for RF and 0.877 for XGBoost (Fig. 4B, fig. S11, and table S5). Among the cancer types analyzed, BRCA exhibited the best performance. Overall, we selected diverse combinations of embedding-multicode to construct our MaNeo pipeline.

To better interpret the importance of features in the RF and XGBoost classifiers, we quantified the contributions of various feature modalities in determining the final score. We conducted a screening process to identify the top 20 features with the highest contributions for each multicode model. We subsequently aggregated the frequencies of these features across all the models and selected the 20 most prevalent feature vectors (fig. S12, A and B). We found that the most important feature groups for RF and XGBoost stemmed from PMBEC embedding, biochemical properties, and AAC (fig. S12, C and D). The amino acids at positions 2 and 9, as anchor residues of HLA molecules, are determinants of the binding process of HLAs with peptides (*38*). A high weight was attributed to biochemical properties, indicating that subtle changes in the cellular microenvironment might affect the process of antigen processing to peptides in the proteasome (*39*). Because RF and XGBoost share many similar high-contributing features, they differ in some aspects. In particular, XGBoost often assigns greater importance to one-hot encoding than does RF. These results suggested that MaNeo enables the capture of multiple-dimensional information from diverse feature groups to help classification.

We demonstrated that MaNeo can screen neo-peptides when trained with fivefold cross-validation against specific tumors. However, experimentally determining the neo-peptides derived from various sources for patients involve extensive time, effort, and cost expenses. To determine the minimum number of peptides required for training to achieve good generalization performance, we compared the performance of models trained on diverse subsets of peptides. A given proportion of all peptides were randomly selected as training data for each model, while the remaining peptides were used for testing. Compared with models using all peptides for RF and XGBoost, the average performance decreased only slightly by 0.009 and 0.008, respectively, in AUPRC when the model was trained on 60% of the available peptides (Fig. 4C), respectively. We further explored the generalization capabilities of MaNeo for novel cancer types by screening all neo-peptides on an unseen tumor that was reserved during training (leave-tumor-out). To gain insight into peptide presentation between tumor and normal cells, we mixed all

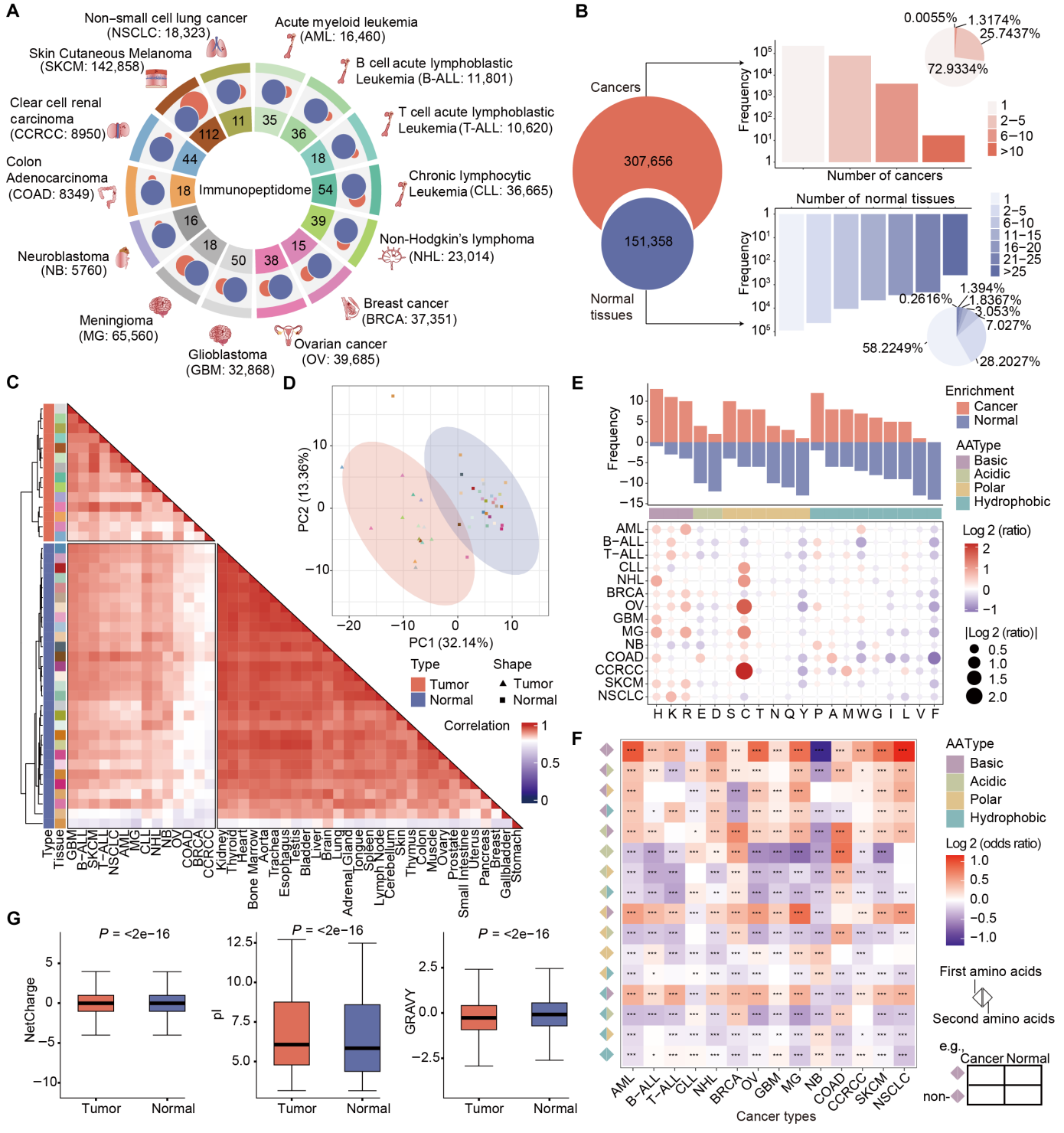

Fig. 3. Differences in the features of tumor and normal peptides. (**A**) Peptides were classified exclusively into tumor samples as tumor peptides, whereas the rest were classified as normal peptides. The size of the circle indicates the number of peptides. The red color represents the tumor peptides, whereas the blue color represents the normal peptides. (**B**) The Venn diagram displaying the number of all the tumor and normal peptides. The bar and pie plots show the frequency and proportion of peptides found in several tumors or normal tissues. (**C**) Heatmap with hierarchically clustered correlations of position frequency vectors across tumors and normal tissues. (**D**) PCA based on the matrix of the AAC at each position to distinguish tumors from normal tissues. (**E**) Stacked bar plot at the top illustrates the amino acid enrichment across various types of tumors. The dot plot at the bottom indicates enriched amino acids in each tumor. (**F**) Connection of amino acids displays different degrees of enrichment among tumors, $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$. (**G**) Comparisons of biochemical features—such as net charge, pI, and GRAVY—between tumor ($n = 307{,}656$) and normal peptides ($n = 151{,}358$) with the Wilcoxon rank sum test.

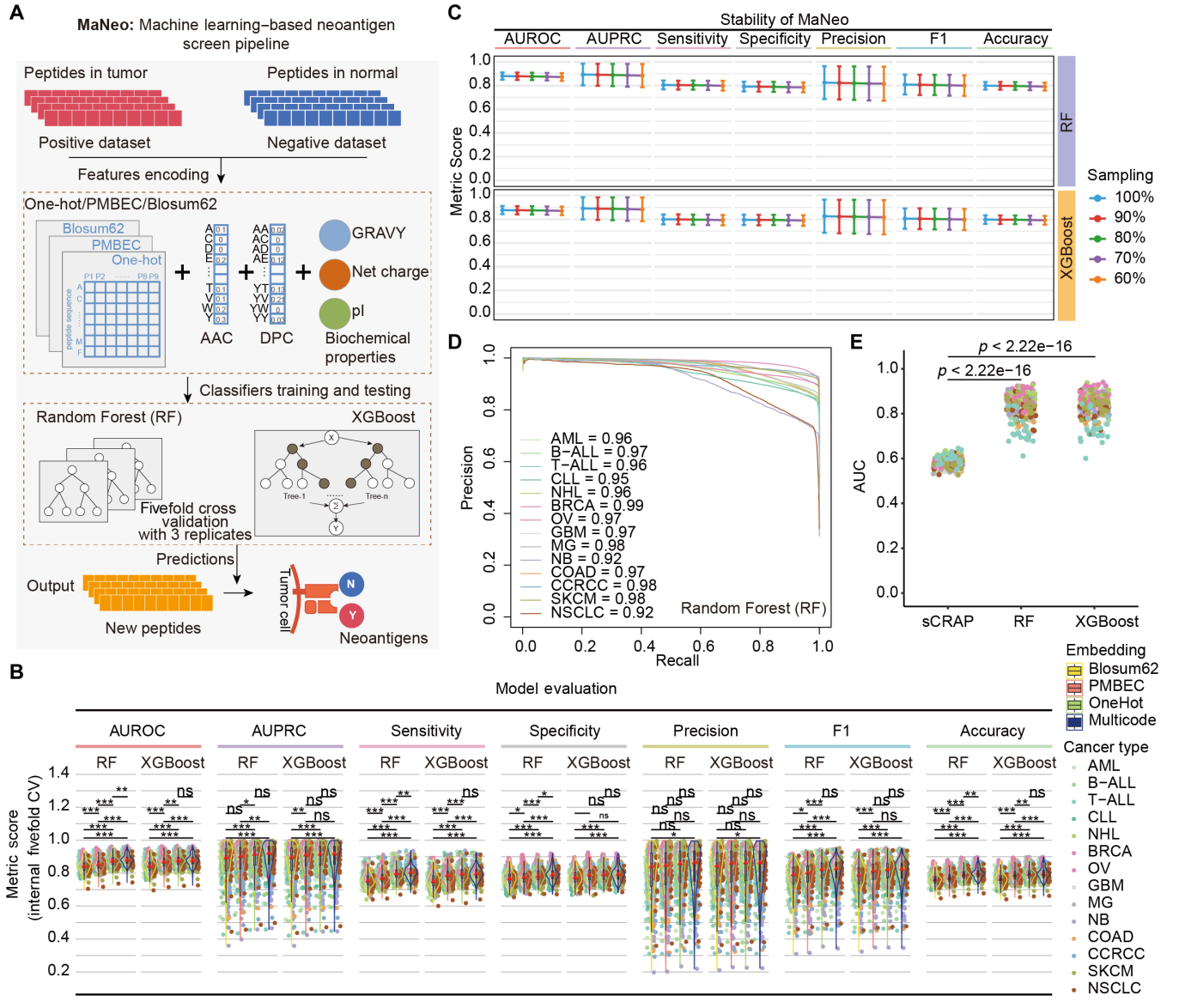

Fig. 4. Predicting neo-peptides within the HLA-presented peptides. (**A**) Schematic representation of MaNeo for the training and testing procedures used to develop the predictive models. For each machine learning model, the hyperparameter was tuned with 5-fold cross-validation with three replicates. (**B**) Model performance was assessed via various metrics ($n = 406$, Wilcoxon rank sum test), $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$. (**C**) Performance of the models when a smaller amount of training data is used. (**D**) AUPRC curves for the different groups of tumors generated via the RF classifier. (**E**) Comparison of performance between the MaNeo and sCRAP algorithms ($n = 377$, Wilcoxon rank sum test). ns, not significant.

the tumor and normal peptides to construct the pan–tumor-normal model. For the pan–tumor-normal model, the performance for RF was considerably better than that of XGBoost, with an average AUPRC of 0.963 ± 0.021 during classification (Fig. 4D and fig. S13). These results suggest that MaNeo could be extended to other tumor types by obtaining experimental data covering a broader range of neo-peptides and learning their common sequence features.

Next, the model performance was compared against that of selective cross-reactive antigen presentation (sCRAP), which was proposed to recognize unmutated neo-peptides as peptide-centric chimeric antigen receptor T-Cell (CAR-T) targets (*40*). Our models significantly outperformed sCRAP ($P < 0.0001$) with 1.4-, 1.2-, 1.4-, 1.3-, 1.1-, 1.3-, and 1.4-fold improvements in the AUC, AUPRC, sensitivity, specificity, precision, weighted F1 score, and balanced accuracy, respectively (Fig. 4E and fig. S14). This may be attributed to the fact that MaNeo takes into account not only the sequence composition similarity but also the biochemical characteristics. Together, the expansive, high-quality data coupled with model training resulted in accurate prediction of immunogenic neo-peptides for all profiled cancer types.

Validation of immunogenic neo-peptides in independent cancer cohorts

Having demonstrated that our models can effectively identify neo-peptides that are absent in normal tissues, we next explored the

application of these models for prioritizing neo-peptides. In this process, we identified immunogenic neo-peptides that bind to HLA molecules and elicit adaptive immune responses. We hypothesized that if peptides can be predicted as tumor peptides against more normal tissues, they are more likely to be neo-peptides. To assess the generalization ability of our models to prioritize immunogenic neo-peptides within an independent dataset, we ranked and classified each peptide on the basis of the predicted score and threshold of the corresponding models (Fig. 5A). We eventually identified 2523 immunogenic neo-peptides, accounting for 0.08 to 0.32% of all peptides per tumor, of which 1476 were canonical, and 1047 were noncanonical (Fig. 5A and table S6).

We found that the source genes of immunogenic neo-peptides were significantly enriched in oncogenic pathways, including the mitotic spindle, G_2-M checkpoint, and notch signaling pathways (Fig. 5B and table S7). Although the majority of immunogenic neo-peptides were patient specific, ~12.6 and ~5.9% of the canonical and noncanonical immunogenic neo-peptides, respectively, were observed in ≥2 samples (Fig. 5C). The immunogenic neo-peptides present in at least five samples were identified and ranked according to their immunogenic and tumor specificity scores. Last, 14 recurrent neo-peptides were identified across cancer types (Fig. 5D). Considering the number of samples in which the immunogenic neo-peptides were present and their respective ranks, we prioritized three strong candidates for functional validations, including "KLNIRPLLR," "KLFSVTRNR," and "RLPQKPLHR" (Fig. 5D).

A potential concern about prioritized neo-peptides is whether they are absent in nonmalignant cells. We thus examined the RNA and protein expression of genes encoding three immunogenic neo-peptides in solid tumors and healthy tissues. We found that EFTUD2, NUP160, and AKNA presented variable but ubiquitous RNA expression patterns across various tissues (fig. S15). These observations suggest that RNA transcript levels are not sufficient to represent both translation and peptide levels. We next evaluated the abundance of proteins that derived neo-peptides on the basis of proteome from tumor cell lines and in-house heart tissues. We found that both EFTUD2 and NUP160 were significantly up-regulated in tumor cell lines, demonstrating their tumor-specific protein expression (fig. S16). In addition, we used Western blot to detect the protein expression levels of the parental proteins EFTUD2, NUP160, and AKNA. The results demonstrated that the protein expression levels of EFTUD2, NUP160, and AKNA were greater in tumor cells (A375 and A549) than in normal cells (AC16, THLE-2, Beas-2B, and HK-2) (fig. S17). Last, we explored the immunopeptidome data used to identify canonical and noncanonical neo-peptides and demonstrated that these peptides were tumor specific and absent in normal cells. All these results indicate that the combination of immunomodome data is very necessary for the identification of tumor neoantigens.

Functional validation of immunogenic neo-peptides

To demonstrate the immune reactivity of candidate neo-peptides, we next evaluated the antitumor effects induced by neo-peptide–loaded dendritic cells (DCs) through the activation of $CD8^+$ T cells. Healthy volunteers were recruited to collect peripheral blood, and flow cytometry sorting technology was used to isolate DCs and $CD8^+$ T cells from the peripheral blood. The purity of the isolated DCs was greater than 90% and that of the isolated $CD8^+$ T cells was greater than 80% (fig. S18). Phenotypically, we observed that subsequent to neoantigen presentation, there was substantial stimulation of $CD8^+$ T cell activation and up-regulation of the levels of two cytokines

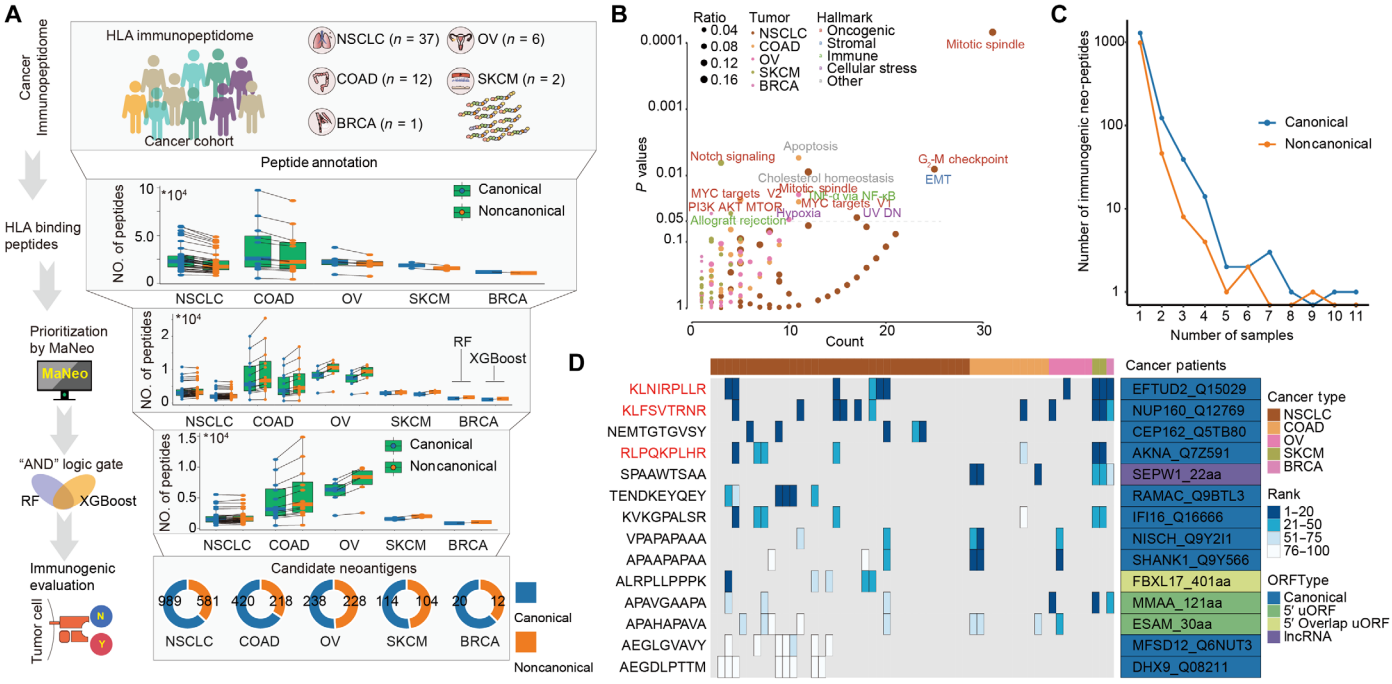

Fig. 5. Validation of MaNeo in independent cancer cohorts. (**A**) Schematic of the computational platform for the identification of immunogenic neo-peptides. (**B**) Gene set enrichment with source genes of immunogenic neo-peptides for each tumor. (**C**) Number of samples in which the immunogenic neo-peptides were detected (1 to 11). (**D**) Immunogenic neo-peptides that were presented in at least five samples. The colors of the heatmap represent the rank of the peptides in the samples. PI3K, phosphatidylinositol 3-kinase; UV, ultraviolet; MTOR, mammalian target of rapamycin; NF-κB, nuclear factor κB.

[interferon-γ (IFN-γ) and tumor necrosis factor–α (TNF-α)] with dose dependent (Fig. 6A). Moreover, the proliferation of $CD8^+$ T cells stimulated by neo-peptide–loaded DCs (NP/DC-$CD8^+$ T cells) was significantly increased (Fig. 6B).

In addition, we also evaluated neo-peptide–specific T cell induction via flow cytometry using KLNIRPLLR, RLPQKPLHR, and KLFSVTRNR tetramers along with $CD8^+$ T cell markers. Compared with the control group, the KLNIRPLLR tetramer-binding $CD8^+$ T cell population expanded from 0.64 to 14.0% of the total $CD8^+$ T cells, the RLPQKPLHR tetramer-binding $CD8^+$ T cell population expanded from 0.64% to 13.1% of the total $CD8^+$ T cells, and the KLFSVTRNR tetramer-binding $CD8^+$ T cell population expanded from 0.64 to 12.4% of the total of $CD8^+$ T cells (Fig. 6C). We further confirmed that, compared with those in the control group of $CD8^+$ T cells stimulated with DCs without neo-peptides (no peptide) loading, the neo-peptides enhanced the immune function of NP/DC-$CD8^+$ T cells against tumor cells (Fig. 6D; A375 and A549). In contrast, these neo-peptides did not significantly affect human cardiomyocytes (AC16), human hepatocytes (THLE-2), human bronchial epithelial cells (Beas-2B), or human renal epithelial cells (HK-2) (Fig. 6D). These findings indicate that the candidate neo-peptides can trigger tumor immune response to eliminate tumors but has no cytotoxic effect on normal cells. These results further demonstrate that MaNeo, trained on large-scale immunopeptidome datasets, exhibited robust generalization capabilities and yielded highly accurate results when it was applied to independent cancer cohorts.

ISE: An immunogenic search engine for personalized immunotherapy

To help researchers apply the MaNeo pipeline described here to tumors of interest, we developed a user-friendly immunogenic search engine, ISE (www.bio-bigdata.com.cn/ISE/index/), for neoantigen prioritization. Users can select the cancer and model of interest and then upload a list of candidate amino acid sequences. After successfully submitting the required input files, users can directly obtain real-time analysis results on the right side of the page. The time depends on the input file size and internet speed. ISE can predict whether the peptide of interest is immunogenic or not in particular cancer. All the data generated can be downloaded for further analysis, and the query results are shown in a user-friendly way. ISE should serve as a guide for biologists interested in identifying the immunogenicity of candidate peptides.

DISCUSSION

The identification of neoantigens is becoming increasingly important for the development of novel immunotherapies for cancer. Most of the existing neoantigen prediction algorithms—such as NMER model (*27*), personalized Variant Antigens by Cancer Sequencing (pVACseq) (*41*), and pTuneos (*42*)—focus primarily on mutation-derived peptides. These algorithms rely on genomic and transcriptomic data to identify tumor-specific mutations and generate candidate peptides. Although these algorithms have been instrumental in advancing the field, they inherently limit the scope of neoantigen discovery, particularly for antigens derived from noncoding regions of protein-coding genes and noncoding RNAs. Recent efforts have shifted toward the exploration of aberrantly expressed neoantigens. For example, several methods—such as sCARP (*40*), k-depl. (*14*), and BamQuery (*6*)—have been developed to screen neoantigens on the basis of expression levels of the RNA transcripts encoding them. However, a notable limitation of transcriptome-based approaches is that they are insufficient for directly identifying translated ORFs or accurately reflecting the expression levels of peptides presented by HLA alleles in vivo. In this study, we generated a comprehensive atlas of neoantigens by combining a coordinated large-scale analysis multiple types of omics data from different tumor and normal tissues. The integration of the immunopeptidome can help to identify ORFs that are likely to be translated into peptides. Furthermore, mass spectrometry–based immunopeptidomics allows for the direct identification of peptide sequences presented on the cell surface by MHC molecules. By combining multiple omics datasets and computational pipelines, we will provide a more accurate representation of the actual peptide repertoire available for immune surveillance.

Characterization of presentation levels is crucial for understanding the diversity of peptides that can be recognized by the immune system (*43*). We discovered that noncanonical peptides outperformed canonical peptides in some aspects. For example, HLA-B alleles tend to bind noncanonical peptides in a closer and more stable manner. The noncanonical peptides exhibited a preference for recognition by T cells. In addition, cysteine was enriched in the noncanonical peptides, but it was underrepresented in canonical peptides (*44*). We believe that noncanonical peptides have comparable potential as targets for cancer immunotherapy. In contrast to that of noncancerous cells, the metabolism of tumor cells leads to a tumor microenvironment that is commonly acidic (*35*, *36*). We hypothesized that difference in the microenvironment result in abnormalities in HLA-presented peptides. We classified the peptides into distinct groups: tumor peptides and normal peptides. Our aim was to scrutinize and identify any significant differences between two sets of peptides. To achieve this goal, we harnessed a series of distinctive features and characteristics inherent to each peptide. PCA revealed notable differences in HLA-presented peptides across various tissue types. Compared with normal tissues, tumor peptides had a higher content of positively charged residues, with cysteine enriched in eight cancer types. Specific amino acid positions showed significant differences, with arginine enrichment at positions 2 and 9 in particular cancers, suggesting primary anchors. A previous study revealed that HLA-presented peptides exhibited strong preferences for basic residues at anchor positions (*45*). Tumor peptides often use basic-basic amino acid combinations, whereas normal peptides prefer acidic-acidic combinations. Cysteine-X and X-cysteine junctions were common in five cancer types. The biochemical properties of HLA-presented peptides—including GRAVY, net charge, and pI—differ significantly between tumor and normal peptides, with tumors characterized by low hydrophobicity, high net charge, and high pI. These results indicate that subtle changes in the cellular microenvironment influence differences in HLA-presented antigens, which in turn reflect cell identity (*26*).

Functionally validation of all the peptides by experimental methods is often expensive, labor-intensive and time-consuming. Leveraging these identified features, we proceeded to develop and construct a sophisticated computational pipeline named MaNeo to facilitate the prioritization of neo-peptides, which have the potential to serve as highly effective targets for immunotherapy treatments. MaNeo exhibited high precision and recall, indicating accurate identification of true positives. We demonstrated the stability of MaNeo by extracting different proportions of training set samples and

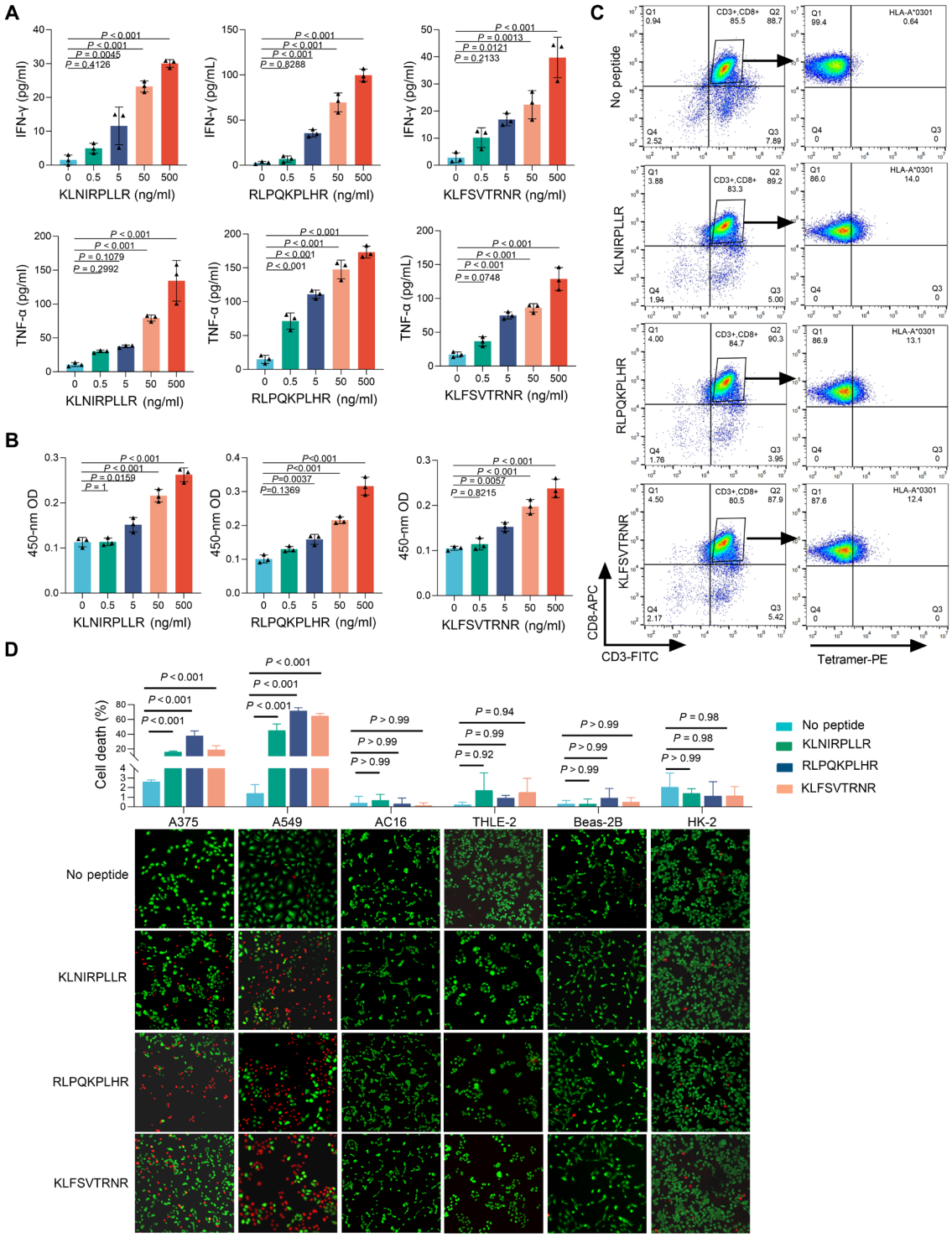

Fig. 6. In vitro validation of neoantigens for tumors. (**A**) Activation assay (IFN-γ and TNF-α enzyme-linked immunosorbent assay) for NP/DC-CD8$^+$ T cells and no peptide. $n = 3$ [means ± SD; ordinary one-way analysis of variance (ANOVA) followed by Dunnett's multiple comparisons test]. (**B**) CCK-8 assay for NP/DC-CD8$^+$ T, $n = 3$ (means ± SD; ordinary one-way ANOVA followed by Dunnett's multiple comparisons test). (**C**) CD8$^+$ T cells from healthy individuals stained with KLNIRPLLR, RLPQKPLHR, and KLFSVTRNR tetramers. (**D**) Cytotoxicity assay for tumor cells (A375: human malignant melanoma line, A549: human NSCLC cancer cell line) and other tissue cells (AC16, THLE-2, Beas-2B, and HK-2). $n = 3$ (means ± SD; ordinary two-way ANOVA followed by Tukey's multiple comparisons test).

leave-one-out method. Even when trained on 60% of the peptides, the average performance of MaNeo slightly decreased by 0.009 and 0.008 in AUPRC for RF and XGBoost, respectively. MaNeo was then tested on an independent tumor to assess its ability to generalize to new cancer types, showing that it could be applied to other tumors by learning common sequence features. The pan–tumor-normal model performed better with RF than XGBoost did, achieving an average AUPRC of 0.963 ± 0.021. The performance of MaNeo was also significantly better than that of the recently developed sCRAP (*40*), with improvements in various metrics. These results suggest that MaNeo considers both sequence similarity and biochemical characteristics, which leads to accurate predictions of immunogenic neo-peptides across different cancer types.

In independent valid cancer cohorts, we identified 2523 immunogenic neo-peptides, which accounted for 0.08 to 0.32% of all peptides per tumor, with 1476 being canonical and 1047 noncanonical. These neo-peptides are enriched in oncogenic pathways. While most neo-peptides were unique to the samples, several were observed in multiple samples. We identified 14 recurrent peptides and validated three of them in cancer cell lines. Notably, they are capable of triggering T cell responses and proliferation. Tetramer binding assay further confirmed that all three neo-peptides were specifically recognized by $CD8^+$ T cell surface TCRs. In particular, the tetramer bound to the KLNIRPLLR peptide displayed the optimal specific binding affinity toward $CD8^+$ T cells. In addition, T cells activated by neo-peptides can eliminate cancer cell lines without damaging healthy cells from the heart, liver, lung, or kidney. Overall, these results provide further evidence that MaNeo can facilitate the identification of immunogenic neo-peptides in various tumors. Last, we created a user-friendly webserver, ISE, to assist clinical researchers in applying MaNeo to screen neo-peptides as potential immunotherapeutic targets.

Some limitations still need to be addressed in future studies. The discovery of personalized antigens, which depend on peptide presentation on cancer cells by HLA-I and HLA-II, is essential for the development of cancer vaccines and T cell therapies (*45*, *46*). However, the present study only effectively integrates data from the HLA-I immunopeptidome and constructs a pipeline to accurately identify and prioritize HLA-I–presented neoantigens, lacking the exploration of HLA-II–presented neoantigens. Furthermore, with the increase in the number of matched Ribo-seq data with immunopeptidome cohorts, we can use additional large-scale Ribo-seq data to uncover translated neoantigens, which can increase our capacity to identify translation in transcripts that are extremely sample specific. In conclusion, our study demonstrated the immunopeptidomics-guided discovery and characterization of neoantigens for personalized cancer immunotherapy.

MATERIALS AND METHODS

MS-based immunopeptidome data collection and analyses

We collected immunopeptidomic datasets from available public databases to find and quantify all HLA-presented peptides on the cell surface, encompassing 14 cancer types and 29 healthy tissues (table S1). The information on the HLA alleles of the samples was retrieved from the corresponding studies. In these datasets, samples were lysed from human tissue or cells, HLA-I–peptide complexes were immunoaffinity purified via pan-HLA-I antibodies (w6/32), and peptides were directly eluted from purified HLA-molecules with acid treatment. The eluted peptides were desalted and concentrated to remove acidic buffers and other impurities. Last, the isolated peptides were subjected to MS analysis.

The canonical human proteome reviewed with Swiss-Prot was obtained from the UniProt database (downloaded in February 2022). To expand the repertoire of HLA-I peptides derived from noncanonical sources, we acquired an in-house customized noncanonical theory library from IEAtlas (*47*). In addition to the constructed proteome reference, the corresponding raw files were searched via MaxQuant (v.2.1.0.0) with a maximum allowed precursor mass tolerance of 20 parts per million. Peptide length was restricted to between 8 and 12 amino acids. No fixed modifications were applied, whereas methionine oxidation and N-terminal acetylation were used as variable modifications. The enzyme specificity was set to unspecific. To reduce false-positive identifications, a peptide spectrum match false discovery rate of 0.05 was set. The "match between runs" option was enabled with default parameters. For the annotation of identified peptides, if a peptide matched multiple hits, then we mapped the peptides in the following priority order: canonical, 5′ overlap uORF, 5′ uORF, out-of-frame, 3′ overlap dORF, 3′ dORF, lncRNA, pseudogene, and "other" ORFs (*48*, *49*). We detected 70,270 and 389,165 unique noncanonical and canonical peptides, respectively. Peptides that were exclusive to tumor tissues were classified as tumor peptides, whereas the remaining peptides, which arose as by-products of normal proliferative programs, were regarded as normal peptides, potentially triggering unintended off-target effects.

MaNeo: Machine learning–based models for neoantigen prioritization

Machine learning models have been established to predict the likelihood that a given peptide will be tumor specific or presented on the surface of normal cells. To construct models that are tumor and normal tissue specific, positive and negative training consisted of MS-identified tumor and normal peptides in specific cancer and normal tissues, respectively. This process was performed for each combination of the 14 cancer types and 29 normal tissues. Because our datasets are unbalanced, we used stratified cross-validation to ensure that we had a similar proportion of positive and negative peptides in both the training and test sets. Model training was conducted in a three times fivefold cross-validation setting with different model initialization random seeds; during each cross-validation iteration, the training data were split into five equal parts, four parts were used to train the model, and the remaining one was used for evaluation. We exploited the average of the three initializations as the final prediction score for a given peptide. The RF and XGBoost models were trained via the "RandomForestClassifier" and "XGBClassifier" functions of the Sklearn Python library.

Four different models were trained with four different embeds of the peptide sequence: (i) one-hot binary encoding, (ii) similarity encoding via the PMBEC matrix, (iii) similarity encoding via the blosum62 matrix, and (iv) combining the above three encoding methods into multicode encoding. For encoding features of non–9-nucleotide oligomer peptides, all peptides were collapsed/padded to length of 9 to correct the discrepancies in length, where the 8-nucleotide oligomer peptides duplicated the seventh position and the other skipped positions 8 until the C -1 residue (*8*, *50*). In addition to encoding features of the peptide sequences, the MaNeo models incorporated the following features: (i) amino acid properties, including AAC and

DPC for 20 standard amino acids present in the peptide sequence, which formed a 420-dimensional feature vector, and (ii) peptide-level characteristics, the three-dimensional feature vector of biochemical properties [pI, net charge, and average hydrophobicity (GRAVY)], were included to reflect the microenvironment changes outside the tumor and normal cells (*35*, *36*).

Model performance evaluation

To evaluate model performance, multiple metrics—including the AUROC, AUPRC, balanced accuracy, weighted F1-score, sensitivity, specificity, and precision—were employed to quantify the predictive power of the models. We applied the function "coords" of the R package "pROC" (v.1.18.4) to determine the optimal threshold for the tumor-specific prediction score calculated by the model against each combination of the cancer type and normal tissue.

The MSV000082648 dataset was obtained from the MassIVE database as a validation set, including 37 patients with non–small cell lung cancer (NSCLC), 12 patients with colon adenocarcinoma (COAD), 6 patients with OV, 2 patients with SKCM, and 1 patient with BRCA. We exploited Comet (*51*) to search each spectrum against a customized protein database and further annotated the types of identified peptides. The peptide predicted to be a tumor peptide in all models aimed at the cancer where the peptide was identified was retained. Next, an "AND" logic gate for the RF and XGBoost classifiers was used to screen neo-peptides. Last, we evaluated the immunogenicity of the neo-peptides via NetMHCpan (v.4.1), NetMHCStabPan (v.1.0), and foreignness. The neo-peptide with a binding affinity < 500 nM, binding stability > 1.4 hours, and foreignness > 10^{-16} was reserved as an immunogenic neo-peptide. We ranked the immunogenic neo-peptides according to their model predicted scores and immunogenic metrics.

Feature importance estimation

To explain the XGBoost model predictions, we measured the contribution of each feature to the prediction results via the function "Explainer" of the Shapley Additive explanation (SHAP) Python library. We selected the top 20 features by mean SHAP absolute values and ordered them on the basis of contribution. For the RF classifier, feature importance was calculated via the feature_importances output of the RandomForestClassifier function. The top 20 features were filtered by feature importance.

Sequence similarity estimation between cancers and normal tissues

To measure the similarity among cancer types and normal tissues, we computed position frequency matrices for HLA-presented peptides and converted them into 180-dimensional vectors. The pairwise correlations of these frequency vectors were used to calculate the Euclidean distance followed by hierarchical clustering. In addition, we performed PCA for the frequency vectors to distinguish tumors from normal tissues.

RNA expression analyses of antigen-related genes

To facilitate comparisons among TCGA datasets and distinct tissues in GTEx, expression quantification across transcript isoforms was uniformed transcripts per million units. For GTEx, we averaged expression for each tissue and retained the genes whose transcripts per million ≤ 1 (*52*). For the TCGA analysis, we acquired 151 AML samples, 1,104 BRCA samples with 113 matched healthy breast tissue samples, 379 OV samples, 471 COAD samples with 41 matched healthy colon tissue samples, 168 GBM samples with five matched healthy brain tissue samples, 1,027 NSCLC samples with 172 matched healthy lung tissue samples, and 535 CCRCC samples with 72 matched healthy kidney tissue samples and SKCM samples. The genes were further filtered on the basis of the antigen parent transcript expression across different cancer types in the TCGA, where the mean expression of TCGA samples was required to be at least twice or greater than the average level detected in any GTEx tissue.

Peptides

Peptides (KLNIRPLLR, RLPQKPLHR, and KLFSVTRNR) with purities greater than 85% were purchased from Go Top Peptide Biotech Co., Ltd. The peptide was synthesized with a custom-built rapid-flow automated peptide synthesizer as previously described (*53*). The lyophilized peptides were diluted to 500 ng/ml in deionized water and stored at −20°C. Purified DCs and $CD8^+$ T cells were used in the peptide loading assay. For the test group, peptide was added at final concentrations of 0.5, 5, 50, and 500 ng/ml, while for the control group, the same volume of deionized water was added.

Cell lines

The cell lines employed in this study were provided by the Department of Pharmacology of Harbin Medical University and have all undergone STR identification. The A375, Beas-2B, and AC16 cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM) (Gibco, catalog no. 11995065, USA) media, supplemented with 10% fetal bovine serum (FBS) (VivaCell, catalog no. C04001-500, China) and 1% penicillin-streptomycin (Meilunbio, catalog no. MA0110, China). The THLE-2 and HK-2 cell lines were cultured in DMEM/F-12K (Gibco, catalog no. 11320033, USA) supplemented with 10% FBS and 1% penicillin-streptomycin. The A549 cell line was cultured in Ham's F-12 K (Gibco, catalog no. 21127022, USA) media supplemented with 10% FBS and 1% penicillin-streptomycin. All the cultures were maintained in humidified cell incubators at 37°C with 5% CO_2 and were routinely tested for mycoplasma contamination.

Western blot

Protein extraction from A549, A375, AC16, Beas-2B, THLE-2, and HK-2 cells was performed using a cell lysate prepared with radioimmunoprecipitation assay lysis buffer, phosphatase inhibitor, and protease inhibitor at a 100:1:1 ratio. The protein concentration was quantified via a BCA kit (Beyotime, catalog no. P0011, China) according to the manufacturer's protocol. Total cellular proteins were electrophoresed on 8% SDS–polyacrylamide gel electrophoresis at 110 V and then transferred to a nitrocellulose membrane (NC) at 300 mA. The NC membrane was blocked with rapid-blocking solution (Beyotime, catalog no. P0252, China) and incubated overnight at 4°C with primary antibodies against EFTUD2 (Proteintech, catalog no. 10208-1-AP, USA), NUP160 (Proteintech, catalog no. 16084-1-AP, USA), AKNA (Proteintech, catalog no. 23785-1-AP, USA), and β-actin (Proteintech, catalog no. 20536-1-AP, USA), followed by a 50-min incubation with secondary antibodies at room temperature. The process was finalized by chemiluminescence imaging.

T cell and DC isolation

Human peripheral blood mononuclear cells (PBMCs) were obtained from the whole blood of consenting healthy volunteers. This study was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of Harbin Medical University (ethics no. IRB5099724). All the volunteers agreed to participate in this study with written informed consent. T cells were isolated with modifications as previously described (*54*), with modifications. In brief, after the blood was mixed 1:1 with phosphate-buffered saline (PBS), the diluted whole blood was overlaid on top of Ficoll-Paque PLUS (Cytiva, catalog no.17144002, USA) before density gradient centrifugation (centrifugation at 400*g* for 30 min). The isolated PBMCs were washed with PBS-buffered twice (centrifugation at 400*g* for 15 min). PBMCs were cultured for 5 hours in RPMI 1640 medium (Gibco, catalog no.11875093, USA) supplemented with 10% FBS (VivaCell, catalog no.C04001-500, China) and 1% penicillin-streptomycin (Meilunbio, catalog no.MA0110, China) followed by rinsing with PBS to remove non-adherent cells, thereby obtaining monocytes. The monocytes were cultured in RPMI 1640 complete medium supplemented with recombinant human interleukin-4 (IL-4) (50 ng/ml; PerproTech, catalog no. 200-04-20UG, USA) and granulocyte-macrophage colony-stimulating factor (100 ng/ml; PerproTech, catalog no. 300-03-20UG, USA) to induce the maturation of DCs. Similarly, monocytes were cultured in RPMI 1640 complete medium supplemented with recombinant human IL-2 (50 ng/ml; PerproTech, catalog no. 200-02-50UG, USA) and lipopolysaccharide (10 ng/ml; Solarbio, catalog no. L8880, China) to induce the maturation of T cells.

Purification of $CD8^+$ T cells and DCs

$CD8^+$ T cells and DCs purified populations were ascertained by flow cytometry. The following anti-human monoclonal antibodies were used for T cells surface staining: CD3–fluorescein isothiocyanate (FITC) (BioLegend, catalog no. 300406, USA) and CD8a–allophycocyanin (BioLegend, catalog no. 300912, USA). The following anti-human monoclonal antibodies were used for DCs surface staining: HLA-DR–phycoerythrin (BioLegend, catalog no. 307606, USA) and CD11c-APC (BioLegend, catalog no. 301614, USA). The cell samples were sorted via fluorescence-activated cell sorting (FACS) (Beckman Coulter, MoFlo XDP, USA), and the data were analyzed with FlowJo software. The sorted cells were used for subsequent experiments.

ELISA assays

Purified DCs were seeded in 24-well flat bottom plates (5×10^5 cells in 500 μl of RPMI 1640 complete medium) and incubated with neo-peptides (KLNIRPLLR, RLPQKPLHR, and KLFSVTRNR) for 3 days to obtain mature DCs loaded with the neo-peptides. The neo-peptide–loaded DCs were cocultured with activated T cells at a ratio of 1:10 for 2 days, followed by restimulation with neo-peptides. These restimulated $CD8^+$ T cells were used for subsequent enzyme-linked immunosorbent assay (ELISA) assays. IFN-γ and TNF-α were measured using by Elabscience ELISA kit performed according to the manufacturer's instructions (IFN-γ ELISA Kit, Elabscience, catalog no. E-EL-H0108, China) (TNF-α ELISA Kit, Elabscience, catalog no. E-EL-H0109, China).

CCK-8 assay

The restimulated $CD8^+$ T cells were added to a 96-well plate at a concentration of 1×10^4 cells per 100 μl. Then, 10 μl of Cell Counting Kit-8 (CCK-8) solution (New Cell & Molecular Biotech, C6005, China) was added to each well. The plate was incubated in a cell culture incubator for 3 hours, and the plate was read at 450 nm with a multifunctional microplate reader (Aglient, BioTek Synergy HTX, China).

Cytotoxicity assay

The cytotoxicity assay was performed with modifications as previously described (*55*). In brief, purified DCs were seeded in 24-well flat bottom plates (5×10^5 cells). For the neo-peptides (KLNIRPLLR, RLPQKPLHR, and KLFSVTRNR) groups, the DCs were incubated with neo-peptides (KLNIRPLLR, RLPQKPLHR, KLFSVTRNR) for 3 days to obtain mature DCs loaded with the neo-peptides. The neo-peptide-loaded DCs were co-cultured with activated $CD8^+$ T cells at a ratio of 1:10 for 2 days, followed by restimulation with neo-peptides. For the control group, purified DCs were directly co-cultured with $CD8^+$ T cells at a ratio of 1:10 for 2 days. Resuspend the $CD8^+$ T cells in the above-mentioned groups with tumor cells (A549 and A375) or other somatic cells (Beas-2B, AC16, THLE-2, and HK-2) at a ratio of 5:1 (*56*). The mixed cells were subsequently seeded into a 24-well plate and cultured for 3 days. After 3 days, the culture medium and suspended T cells were removed, and the adherent tumor cells were subsequently stained with the Calcein AM/PI Double Staining Kit (Elabscience, E-CK-A354, China). Confocal laser microscopy (Olympus, catalog no. 0972818, Japan) was used to obtain images. ImageJ software was used to count and calculate the proportion of dead cells.

Tetramer assembly and staining

The tetramer assembly and staining experiment was performed as previously described with appropriate modifications (*57*). According to the manufacturer's instructions, the LEGEND MAX Flex-T Human Class I Peptide Exchange ELISA Kit (BioLegend, catalog no.447207, USA) was used to detect the affinity between the HLA class I α-chain/β2-microglobulin and immunogenic neo-peptides (KLNIRPLLR, RLPQKPLHR, and KLFSVTRNR). Peptide-MHC tetramers were produced via the ultraviolet-irradiation–mediated peptide exchange method. $CD8^+$ T cell sorting was performed as previously described. The cells were collected, washed, resuspended in FACS buffer, and stained with fluorescent dye conjugated antibodies for 15 min at 4°C. CD3-FITC (BioLegend, catalog no.300406, USA), CD8a-APC (BioLegend, catalog no.300912, USA), and Flex-T HLA-A*03:01 Monomer UVX (BioLegend, catalog no. 280005; 0.05 μg/1 $\times 10^6$ cells) were used. After being washed twice in FACS buffer, the cells were analyzed via a FACS (Apogee, Micro Plus, UK), and the data were analyzed with FlowJo software.

Webserver implementation

The webserver-backed server was constructed with Django (version 4.2.14). The web frontend was implemented with Hyper Text Markup Language and Cascading Style Sheets, incorporating jQuery (version 3.7.1) and Datatable (v.2.1.8) plugins for the visualization of result tables. Statistical analyses were conducted via Python version 3.6.8. The ISE underwent testing across several popular web browsers, with a preference for Google Chrome, followed by Firefox and Apple Safari.

Statistical analyses

All the statistical analyses were conducted with R software (v.4.1.2). *P* values for the comparisons were calculated between two groups

with Wilcoxon signed-rank tests. Spearman correlation was used to investigate the correlation coefficients and raw *P* values between features measured on a continuous scale. We used the R package "ggseqlogo" to identify motifs under the background of diverse tumor and normal tissues. Enrichment analysis was performed via the R packages "clusterProfiler" (v.4.2.2), and reference gene sets were obtained from the HALLMARK pathways in MSigDB.

Supplementary Materials

The PDF file includes:
Figs. S1 to S18
Legends for tables S1 to S7

Other Supplementary Material for this manuscript includes the following:
Tables S1 to S7

REFERENCES AND NOTES

1. F. Lang, B. Schrörs, M. Löwer, Ö. Türeci, U. Sahin, Identification of neoantigens for individualized therapeutic cancer vaccines. *Nat. Rev. Drug Discov.* **21**, 261–282 (2022).
2. P. D. Katsikis, K. J. Ishii, C. Schliehe, Challenges in developing personalized neoantigen cancer vaccines. *Nat. Rev. Immunol.* **24**, 213–227 (2024).
3. M. E. Camarena, P. Theunissen, M. Ruiz, J. Ruiz-Orera, B. Calvo-Serra, R. Castelo, C. Castro, P. Sarobe, P. Fortes, J. Perera-Bel, M. M. Albà, Microproteins encoded by noncanonical ORFs are a major source of tumor-specific antigens in a liver cancer patient meta-cohort. *Sci. Adv.* **10**, eadn3628 (2024).
4. Y. Cai, D. Li, D. Lv, J. Yu, Y. Ma, T. Jiang, N. Ding, Z. Liu, Y. Li, J. Xu, MHC-I-presented non-canonical antigens expand the cancer immunotherapy targets in acute myeloid leukemia. *Sci. Data* **11**, 831 (2024).
5. C. Chong, G. Coukos, M. Bassani-Sternberg, Identification of tumor antigens with immunopeptidomics. *Nat. Biotechnol.* **40**, 175–188 (2022).
6. M. V. R. Cuevas, M. P. Hardy, J.-D. Larouche, A. Apavaloaei, E. Kina, K. Vincent, P. Gendron, J.-P. Laverdure, C. Durette, P. Thibault, S. Lemieux, C. Perreault, G. Ehx, BamQuery: A proteogenomic tool to explore the immunopeptidome and prioritize actionable tumor antigens. *Genome Biol.* **24**, 188 (2023).
7. G. Li, S. Mahajan, S. Ma, E. D. Jeffery, X. Zhang, A. Bhattacharjee, M. Venkatasubramanian, M. T. Weirauch, E. R. Miraldi, H. L. Grimes, G. M. Sheynkman, T. Tilburgs, N. Salomonis, Splicing neoantigen discovery with SNAF reveals shared targets for cancer immunotherapy. *Sci. Transl. Med.* **16**, eade2886 (2024).
8. A. Kacen, A. Javitt, M. P. Kramer, D. Morgenstern, T. Tsaban, M. D. Shmueli, G. C. Teo, F. da Veiga Leprevost, E. Barnea, F. Yu, A. Admon, L. Eisenbach, Y. Samuels, O. Schueler-Furman, Y. Levin, A. I. Nesvizhskii, Y. Merbl, Post-translational modifications reshape the antigenic landscape of the MHC I immunopeptidome in tumors. *Nat. Biotechnol.* **41**, 239–251 (2023).
9. M. Arnaud, J. Chiffelle, R. Genolet, B. Navarro Rodrigo, M. A. S. Perez, F. Huber, M. Magnin, T. Nguyen-Ngoc, P. Guillaume, P. Baumgaertner, C. Chong, B. J. Stevenson, D. Gfeller, M. Irving, D. E. Speiser, J. Schmidt, V. Zoete, L. E. Kandalaft, M. Bassani-Sternberg, S. Bobisse, G. Coukos, A. Harari, Sensitive identification of neoantigens and cognate TCRs in human solid tumors. *Nat. Biotechnol.* **40**, 656–660 (2022).
10. F. Huber, M. Arnaud, B. J. Stevenson, J. Michaux, F. Benedetti, J. Thevenet, S. Bobisse, J. Chiffelle, T. Gehert, M. Müller, H. Pak, A. I. Krämer, E. R. Altimiras, J. Racle, M. Taillandier-Coindard, K. Muehlethaler, A. Auger, D. Saugy, B. Murgues, A. Benyagoub, D. Gfeller, D. D. Laniti, L. Kandalaft, B. N. Rodrigo, H. Bouchaab, S. Tissot, G. Coukos, A. Harari, M. Bassani-Sternberg, A comprehensive proteogenomic pipeline for neoantigen discovery to advance personalized cancer immunotherapy. *Nat. Biotechnol.*, 10.1038/s41587-024-02420-y , (2024).
11. Capturing the diversity of protein modifications on presented tumor antigens. *Nat. Biotechnol.* **41**, 195–196 (2023).
12. Y. Li, Y. Zhang, T. Pan, P. Zhou, W. Zhou, Y. Gao, S. Zheng, J. Xu, Shedding light on the hidden human proteome expands immunopeptidome in cancer. *Brief. Bioinform.* **23**, bbac034 (2022).
13. G. Rospo, R. Chila, V. Matafora, V. Basso, S. Lamba, A. Bartolini, A. Bachi, F. Di Nicolantonio, A. Mondino, G. Germano, A. Bardelli, Non-canonical antigens are the largest fraction of peptides presented by MHC class I in mismatch repair deficient murine colorectal cancer. *Genome Med.* **16**, 15 (2024).
14. G. Ehx, J.-D. Larouche, C. Durette, J.-P. Laverdure, L. Hesnard, K. Vincent, M.-P. Hardy, C. Thériault, C. Rulleau, J. Lanoix, E. Bonneil, A. Feghaly, A. Apavaloaei, N. Noronha, C. M. Laumont, J.-S. Delisle, L. Vago, J. Hébert, G. Sauvageau, S. Lemieux, P. Thibault, C. Perreault, Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. *Immunity* **54**, 737–752.e10 (2021).
15. A. M. Jaeger, L. E. Stopfer, R. Ahn, E. A. Sanders, D. A. Sandel, W. A. Freed-Pastor, W. M. Rideout III, S. Naranjo, T. Fessenden, K. B. Nguyen, P. S. Winter, R. E. Kohn, P. M. K. Westcott, J. M. Schenkel, S.-L. Shanahan, A. K. Shalek, S. Spranger, F. M. White, T. Jacks, Deciphering the immunopeptidome in vivo reveals new tumour antigens. *Nature* **607**, 149–155 (2022).
16. C. Huang, L. Chen, S. R. Savage, R. V. Eguez, Y. Dou, Y. Li, F. da Veiga Leprevost, E. J. Jaehnig, J. T. Lei, B. Wen, M. Schnaubelt, K. Krug, X. Song, M. Cieślik, H.-Y. Chang, M. A. Wyczalkowski, K. Li, A. Colaprico, Q. K. Li, D. J. Clark, Y. Hu, L. Cao, J. Pan, Y. Wang, K.-C. Cho, Z. Shi, Y. Liao, W. Jiang, M. Anurag, J. Ji, S. Yoo, D. C. Zhou, W.-W. Liang, M. Wendl, P. Vats, S. A. Carr, D. R. Mani, Z. Zhang, J. Qian, X. S. Chen, A. R. Pico, P. Wang, A. M. Chinnaiyan, K. A. Ketchum, C. R. Kinsinger, A. I. Robles, E. An, T. Hiltke, M. Mesri, M. Thiagarajan, A. M. Weaver, A. G. Sikora, J. Lubinski, M. Wierzbicka, M. Wiznerowicz, S. Satpathy, M. A. Gillette, G. Miles, M. J. Ellis, G. S. Omenn, H. Rodriguez, E. S. Boja, S. M. Dhanasekaran, L. Ding, A. I. Nesvizhskii, A. K. El-Naggar, D. W. Chan, H. Zhang, B. Zhang, Clinical Proteomic Tumor Analysis Consortium, Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell* **39**, 361–379.e16 (2021).
17. S. R. Savage, X. Yi, J. T. Lei, B. Wen, H. Zhao, Y. Liao, E. J. Jaehnig, L. K. Somes, P. W. Shafer, T. D. Lee, Z. Fu, Y. Dou, Z. Shi, D. Gao, V. Hoyos, Q. Gao, B. Zhang, Pan-cancer proteogenomics expands the landscape of therapeutic targets. *Cell* **187**, 4389–4407.e15 (2024).
18. S. Pounraj, S. Chen, L. Ma, R. Mazzieri, R. Dolcetti, B. H. A. Rehm, Targeting tumor heterogeneity with neoantigen-based cancer vaccines. *Cancer Res.* **84**, 353–363 (2024).
19. O. Bartok, A. Pataskar, R. Nagel, M. Laos, E. Goldfarb, D. Hayoun, R. Levy, P.-R. Körner, I. Z. M. Kreuger, J. Champagne, E. A. Zaal, O. B. Bleijerveld, X. Huang, J. Kenski, J. Wargo, A. Brandis, Y. Levin, O. Mizrahi, M. Alon, S. Lebon, W. Yang, M. M. Nielsen, N. Stern-Ginossar, M. Altelaar, C. R. Berkers, T. Geiger, D. S. Peeper, J. Olweus, Y. Samuels, R. Agami, Anti-tumour immunity induces aberrant peptide presentation in melanoma. *Nature* **590**, 332–337 (2021).
20. K. Peng, X. Chen, A. Lin, Z. Tong, W. Lin, PolyC-RNA-binding protein 1 (PCBP1) enhances tropomyosin 3 (TPM3) mRNA stability to promote the progression of esophageal squamous cell carcinoma. *Bioengineered* **13**, 8581–8592 (2022).
21. K. T. Tan, C.-N. Yeh, Y.-C. Chang, J.-H. Cheng, W.-L. Fang, Y.-C. Yeh, Y.-C. Wang, D.-S. Hsu, C.-E. Wu, J.-I. Lai, P. M.-H. Chang, M.-H. Chen, M.-L. Lu, S.-J. Chen, Y. Chao, M. Hsiao, M.-H. Chen, *PRKDC*: New biomarker and drug target for checkpoint blockade immunotherapy. *J. Immunother. Cancer* **8**, e000485 (2020).
22. A. J. Tijsen, L. Cócera Ortega, Y. J. Reckman, X. Zhang, I. van der Made, S. Aufiero, J. Li, S. C. Kamps, A. van den Bout, H. D. Devalla, K. Y. van Spaendonck-Zwarts, S. Engelhardt, L. Gepstein, J. S. Ware, Y. M. Pinto, Titin circular RNAs create a back-splice motif essential for SRSF10 splicing. *Circulation* **143**, 1502–1512 (2021).
23. Y. Huang, S. Zheng, Z. Guo, X. de Mollerat du Jeu, X.-J. Liang, Z. Yang, H.-Y. Zhang, S. Gao, Z. Liang, Ionizable liposomal siRNA therapeutics enables potent and persistent treatment of Hepatitis B. *Signal Transduct. Target. Ther.* **7**, 38 (2022).
24. R. G. Gupta, F. Li, J. Roszik, G. Lizée, Exploiting tumor neoantigens to target cancer evolution: Current challenges and promising therapeutic approaches. *Cancer Discov.* **11**, 1024–1039 (2021).
25. C. M. Laumont, K. Vincent, L. Hesnard, É. Audemard, É. Bonneil, J.-P. Laverdure, P. Gendron, M. Courcelles, M.-P. Hardy, C. Côté, C. Durette, C. St-Pierre, M. Benhammadi, J. Lanoix, S. Vobecky, E. Haddad, S. Lemieux, P. Thibault, C. Perreault, Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Transl. Med.* **10**, eaau5516 (2018).
26. M. O. Schaettler, M. M. Richters, A. Z. Wang, Z. L. Skidmore, B. Fisk, K. E. Miller, T. L. Vickery, A. H. Kim, M. R. Chicoine, J. W. Osbun, E. C. Leuthardt, J. L. Dowling, G. J. Zipfel, R. G. Dacey, H.-C. Lu, T. M. Johanns, O. L. Griffith, E. R. Mardis, M. Griffith, G. P. Dunn, Characterization of the genomic and immunologic diversity of malignant brain tumors through multisector analysis. *Cancer Discov.* **12**, 154–171 (2022).
27. J. J. Gartner, M. R. Parkhurst, A. Gros, E. Tran, M. S. Jafferji, A. Copeland, K.-I. Hanada, N. Zacharakis, A. Lalani, S. Krishna, A. Sachs, T. D. Prickett, Y. F. Li, M. Florentin, S. Kivitz, S. C. Chatmon, S. A. Rosenberg, P. F. Robbins, A machine learning model for ranking candidate HLA class I neoantigens based on known neoepitopes from multiple human tumor types. *Nat. Cancer* **2**, 563–574 (2021).
28. S. Kalaora, A. Nagler, D. Nejman, M. Alon, C. Barbolin, E. Barnea, S. L. C. Ketelaars, K. Cheng, K. Vervier, N. Shental, Y. Bussi, R. Rotkopf, R. Levy, G. Benedek, S. Trabish, T. Dadosh, S. Levin-Zaidman, L. T. Geller, K. Wang, P. Greenberg, G. Yagel, A. Peri, G. Fuks, N. Bhardwaj, A. Reuben, L. Hermida, S. B. Johnson, J. R. Galloway-Peña, W. C. Shropshire, C. Bernatchez, C. Haymaker, R. Arora, L. Roitman, R. Eilam, A. Weinberger, M. Lotan-Pompan, M. Lotem, A. Admon, Y. Levin, T. D. Lawley, D. J. Adams, M. P. Levesque, M. J. Besser, J. Schachter, O. Golani, E. Segal, N. Geva-Zatorsky, E. Ruppin, P. Kvistborg, S. N. Peterson, J. A. Wargo, R. Straussman, Y. Samuels, Identification of bacteria-derived HLA-bound peptides in melanoma. *Nature* **592**, 138–143 (2021).

29. S. Barik, The uniqueness of tryptophan in biology: Properties, metabolism, interactions and localization in proteins. *Int. J. Mol. Sci.* **21**, 8776 (2020).
30. M. Rasmussen, E. Fenoy, M. Harndahl, A. B. Kristensen, I. K. Nielsen, M. Nielsen, S. Buus, Pan-specific prediction of peptide-MHC class I complex stability, a correlate of T cell immunogenicity. *J. Immunol.* **197**, 1517–1524 (2016).
31. B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
32. T. J. O'Donnell, A. Rubinsteyn, U. Laserson, MHCflurry 2.0: Improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Syst.* **11**, 42–48.e7 (2020).
33. K. M. Boehm, B. Bhinder, V. J. Raja, N. Dephoure, O. Elemento, Predicting peptide presentation by major histocompatibility complex class I: An improved machine learning approach to the immunopeptidome. *BMC Bioinformatics* **20**, 7 (2019).
34. M. Łuksza, N. Riaz, V. Makarov, V. P. Balachandran, M. D. Hellmann, A. Solovyov, N. A. Rizvi, T. Merghoub, A. J. Levine, T. A. Chan, J. D. Wolchok, B. D. Greenbaum, A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* **551**, 517–520 (2017).
35. R. D. Leone, J. D. Powell, Metabolism of immune cells in cancer. *Nat. Rev. Cancer* **20**, 516–531 (2020).
36. D. Rekow, Computer-aided design and manufacturing in dentistry: A review of the state of the art. *J. Prosthet. Dent.* **58**, 512–516 (1987).
37. Z. Liu, C. Zhou, Y. Qin, Z. Wang, L. Wang, X. Wei, Y. Zhou, Q. Li, H. Zhou, W. Wang, Y.-X. Fu, M. Zhu, W. Liang, Coordinating antigen cytosolic delivery and danger signaling to program potent cross-priming by micelle-based nanovaccine. *Cell Discov.* **3**, 17007 (2017).
38. R. H. McCusker, D. R. Campion, A. L. Cartwright, Effect of growth hormone-secreting tumors on adipose tissue cellularity in young and mature rats. *Growth* **50**, 128–137 (1986).
39. N. Pishesha, T. J. Harmand, H. L. Ploegh, A guide to antigen processing and presentation. *Nat. Rev. Immunol.* **22**, 751–764 (2022).
40. M. Yarmarkovich, Q. F. Marshall, J. M. Warrington, R. Premaratne, A. Farrel, D. Groff, W. Li, M. di Marco, E. Runbeck, H. Truong, J. S. Toor, S. Tripathi, S. Nguyen, H. Shen, T. Noel, N. L. Church, A. Weiner, N. Kendsersky, D. Martinez, R. Weisberg, M. Christie, L. Eisenlohr, K. R. Bosse, D. S. Dimitrov, S. Stevanovic, N. G. Sgourakis, B. R. Kiefel, J. M. Maris, Targeting of intracellular oncoproteins with peptide-centric CARs. *Nature* **623**, 820–827 (2023).
41. J. Hundal, B. M. Carreno, A. A. Petti, G. P. Linette, O. L. Griffith, E. R. Mardis, M. Griffith, pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. *Genome Med.* **8**, 11 (2016).
42. C. Zhou, Z. Wei, Z. Zhang, B. Zhang, C. Zhu, K. Chen, G. Chuai, S. Qu, L. Xie, Y. Gao, Q. Liu, pTuneos: Prioritizing tumor neoantigens from next-generation sequencing data. *Genome Med.* **11**, 67 (2019).
43. C. Tretter, N. de Andrade Krätzig, M. Pecoraro, S. Lange, P. Seifert, C. von Frankenberg, J. Untch, G. Zuleger, M. Wilhelm, D. P. Zolg, F. S. Dreyer, E. Bräunlein, T. Engleitner, S. Uhrig, M. Boxberg, K. Steiger, J. Slotta-Huspenina, S. Ochsenreither, N. von Bubnoff, S. Bauer, M. Boerries, P. J. Jost, K. Schenck, I. Dresing, F. Bassermann, H. Friess, D. Reim, K. Grützmann, K. Pfütze, B. Klink, E. Schrock, B. Haller, B. Kuster, M. Mann, W. Weichert, S. Fröhling, R. Rad, M. Hiltensperger, A. M. Krackhardt, Proteogenomic analysis reveals RNA as a source for tumor-agnostic neoantigen identification. *Nat. Commun.* **14**, 4632 (2023).
44. M. Wilhelm, D. P. Zolg, M. Graber, S. Gessulat, T. Schmidt, K. Schnatbaum, C. Schwencke-Westphal, P. Seifert, N. de Andrade Kratzig, J. Zerweck, T. Knaute, E. Bräunlein, P. Samaras, L. Lautenbacher, S. Klaeger, H. Wenschuh, R. Rad, B. Delanghe, A. Huhmer, S. A. Carr, K. R. Clauser, A. M. Krackhardt, U. Reimer, B. Kuster, Deep learning boosts sensitivity of mass spectrometry-based immunopeptidomics. *Nat. Commun.* **12**, 3346 (2021).
45. M. Stražar, J. Park, J. G. Abelin, H. B. Taylor, T. K. Pedersen, D. R. Plichta, E. M. Brown, B. Eraslan, Y.-M. Hung, K. Ortiz, K. R. Clauser, S. A. Carr, R. J. Xavier, D. B. Graham, HLA-II immunopeptidome profiling and deep learning reveal features of antigenicity to inform antigen discovery. *Immunity* **56**, 1681–1698.e13 (2023).
46. M. H. Dezfulian, T. Kula, T. Pranzatelli, N. Kamitaki, Q. Meng, B. Khatri, P. Perez, Q. Xu, A. Chang, A. C. Kohlgruber, Y. Leng, A. A. Jupudi, M. L. Joachims, J. A. Chiorini, C. J. Lessard, A. D. Farris, S. K. Muthuswamy, B. M. Warner, S. J. Elledge, TScan-II: A genome-scale platform for the de novo identification of $CD4^+$ T cell epitopes. *Cell* **186**, 5569–5586.e21 (2023).
47. Y. Cai, D. Lv, D. Li, J. Yin, Y. Ma, Y. Luo, L. Fu, N. Ding, Y. Li, Z. Pan, X. Li, J. Xu, IEAtlas: An atlas of HLA-presented immune epitopes derived from non-coding regions. *Nucleic Acids Res.* **51**, D409–D417 (2023).
48. A. I. Kraemer, C. Chong, F. Huber, H. Pak, B. J. Stevenson, M. Müller, J. Michaux, E. R. Altimiras, S. Rusakiewicz, L. Simó-Riudalbas, E. Planet, M. Wiznerowicz, J. Dagher, D. Trono, G. Coukos, S. Tissot, M. Bassani-Sternberg, The immunopeptidome landscape associated with T cell infiltration, inflammation and immune editing in lung cancer. *Nat. Cancer* **4**, 608–628 (2023).
49. R. Jackson, L. Kroehling, A. Khitun, W. Bailis, A. Jarret, A. G. York, O. M. Khan, J. R. Brewer, M. H. Skadow, C. Duizer, C. C. D. Harman, L. Chang, P. Bielecki, A. G. Solis, H. R. Steach, S. Slavoff, R. A. Flavell, The translation of non-canonical open reading frames controls mucosal immunity. *Nature* **564**, 434–438 (2018).
50. S. Sarkizova, S. Klaeger, P. M. Le, L. W. Li, G. Oliveira, H. Keshishian, C. R. Hartigan, W. Zhang, D. A. Braun, K. L. Ligon, P. Bachireddy, I. K. Zervantonakis, J. M. Rosenbluth, T. Ouspenskaia, T. Law, S. Justesen, J. Stevens, W. J. Lane, T. Eisenhaure, G. Lan Zhang, K. R. Clauser, N. Hacohen, S. A. Carr, C. J. Wu, D. B. Keskin, A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat. Biotechnol.* **38**, 199–209 (2020).
51. J. K. Eng, T. A. Jahan, M. R. Hoopmann, Comet: An open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24 (2013).
52. D. Hebenstreit, M. Fang, M. Gu, V. Charoensawan, A. van Oudenaarden, S. A. Teichmann, RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol. Syst. Biol.* **7**, 497 (2011).
53. A. J. Mijalis, D. A. Thomas III, M. D. Simon, A. Adamo, R. Beaumont, K. F. Jensen, B. L. Pentelute, A fully automated flow-based approach for accelerated peptide synthesis. *Nat. Chem. Biol.* **13**, 464–466 (2017).
54. M. Ali, Z. Foldvari, E. Giannakopoulou, M.-L. Böschen, E. Strønen, W. Yang, M. Toebes, B. Schubert, O. Kohlbacher, T. N. Schumacher, J. Olweus, Induction of neoantigen-reactive T cells from healthy donors. *Nat. Protoc.* **14**, 1926–1943 (2019).
55. R. Naghavian, W. Faigle, P. Oldrati, J. Wang, N. C. Toussaint, Y. Qiu, G. Medici, M. Wacker, L. K. Freudenmann, P.-E. Bonté, M. Weller, L. Regli, S. Amigorena, H.-G. Rammensee, J. S. Walz, S. D. Brugger, M. Mohme, Y. Zhao, M. Sospedra, M. C. Neidert, R. Martin, Microbial peptides activate tumour-infiltrating lymphocytes in glioblastoma. *Nature* **617**, 807–817 (2023).
56. Z. Yu, W. Liu, Y. He, M. Sun, J. Yu, X. Jiao, Q. Han, H. Tang, B. Zhang, Y. Xian, J. Qi, J. Gong, W. Xin, G. Shi, F. Shan, R. Zhang, J. Li, M. Wei, HLA-A2.1-restricted ECM1-derived epitope LA through DC cross-activation priming $CD8^+$ T and NK cells: A novel therapeutic tumour vaccine. *J. Hematol. Oncol.* **14**, 71 (2021).
57. D. Li, C. Chen, J. Li, J. Yue, Y. Ding, H. Wang, Z. Liang, L. Zhang, S. Qiu, G. Liu, Y. Gao, Y. Huang, D. Li, R. Zhang, W. Liu, X. Wen, B. Li, X. Zhang, X. Zhang, R.-H. Xu, A pilot study of lymphodepletion intensity for peripheral blood mononuclear cell-derived neoantigen-specific CD8 + T cell therapy in patients with advanced solid tumors. *Nat. Commun.* **14**, 3447 (2023).

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (32322020, 32170676, and 32060152) and the Natural Science Foundation of Heilongjiang Province (Key Program; ZD2023C007). **Author contributions:** Conceptualization: Yongsheng Li, J.X., and Y.Z. Methodology: Y.C., M.G., M.Z., F.L., D.L., J.G., H.W., Yapeng Li, and J.J. Software: Y.C., M.G., M.Z., F.L., D.L., and J.J. Validation: Y.C., M.G., M.Z., F.L., D.L., J.G., and J.J. Formal analysis: Y.C., M.G., M.Z., F.L., D.L., J.G., H.W., and J.J. Investigation: Y.C., M.G., M.Z., F.L., D.L., J.G., H.W., Yapeng Li, Q.L., and J.J. Resources: Y.C., M.G., M.Z., F.L., D.L., J.G., H.W., Yapeng Li, Q.L., Yongsheng Li, J.X., and Y.Z. Data curation: Y.C.. Visualization: Y.C., M.G., M.Z., and F.L. Supervision: Yongsheng Li, J.X., and Y.Z. Writing—original draft: Y.C., Yongsheng Li, J.X., and J.J. Writing—review and editing: Y.C., Yongsheng Li, J.X., Y.Z., and J.J. Project administration: Yongsheng Li, J.X., and Y.Z. Funding acquisition: Yongsheng Li, J.X., and Y.Z. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 30 December 2024
Accepted 16 April 2025
Published 21 May 2025
10.1126/sciadv.adv6445