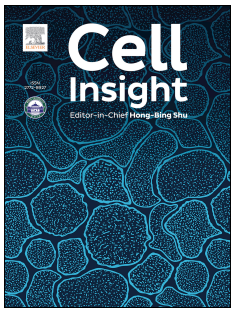
Journal Pre-proof



Rapid and direct discovery of functional tumor specific neoantigens by high resolution mass spectrometry and novel algorithm prediction

Huajian Tian, Guifei Li, Cookson K.C. Chiu, E. Li, Yuzong Chen, Ting Zhu, Min Hu, Yanjie Wang, Suping Wen, Jiajia Li, Shuangxue Luo, Zhicheng Chen, Huimei Zeng, Nan Zheng, Jinyong Wang, Weijun Shen, Xi Kang

PII: S2772-8927(25)00025-2

DOI: https://doi.org/10.1016/j.cellin.2025.100251

Reference: CELLIN 100251

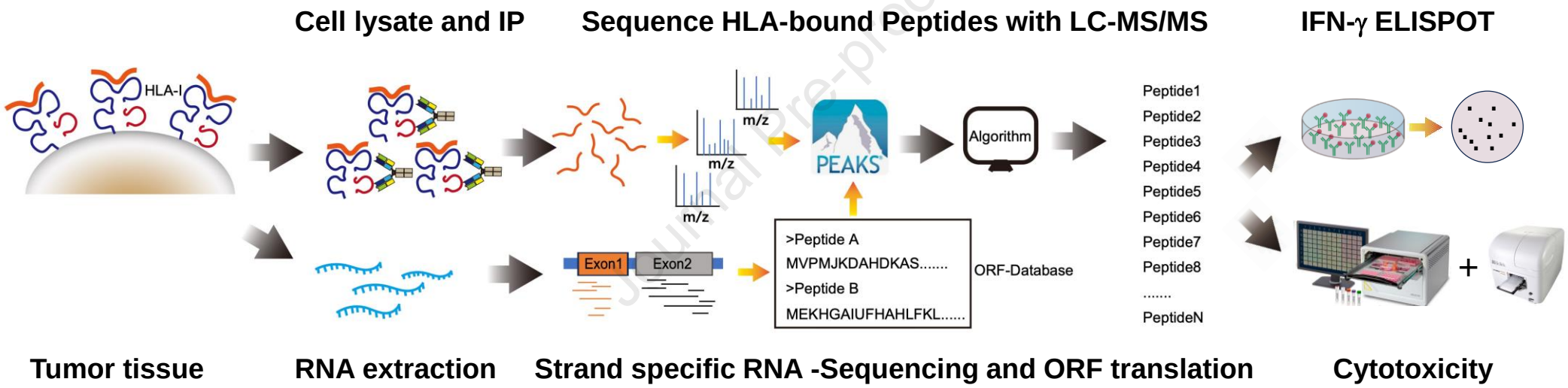
To appear in: *Cell Insight*

Received Date: 6 November 2024

Revised Date: 1 May 2025

Accepted Date: 1 May 2025

Please cite this article as: Tian, H., Li, G., Chiu, C.K.C., Li, E, Chen, Y., Zhu, T., Hu, M., Wang, Y., Wen, S., Li, J., Luo, S., Chen, Z., Zeng, H., Zheng, N., Wang, J., Shen, W., Kang, X., Rapid and direct discovery of functional tumor specific neoantigens by high resolution mass spectrometry and novel algorithm prediction, *Cell Insight*, https://doi.org/10.1016/j.cellin.2025.100251.



Rapid and direct discovery of functional tumor specific neoantigens by high resolution mass spectrometry and novel algorithm prediction

Huajian Tian[1,†], Guifei Li[1,†], Cookson K.C. Chiu [2], E Li[2], Yuzong Chen[4,5], Ting Zhu[5], Min Hu[3], Yanjie Wang[3], Suping Wen[1], Jiajia Li[1], Shuangxue Luo[1], Zhicheng Chen[1], Huimei Zeng[1], Nan Zheng[1], Jinyong Wang[6,7], Weijun Shen[1]*, Xi Kang[1]*

1.Translation innovation center, Shenzhen Bay Laboratory, Shenzhen, P. R. China
2.Multi-omics Mass Spectrometry Core, Biomedical Research Core Facilities, Shenzhen Bay Laboratory, Shenzhen, P. R. China
3. Genomics core, Biomedical Research Core Facilities, Shenzhen Bay Laboratory, Shenzhen, P. R. China
4. Institute of Biomedical Health Technology and Engineering, Shenzhen Bay Laboratory, Shenzhen, P. R. China
5.The State Key Laboratory of Chemical Oncogenomics, Key Laboratory of Chemical Biology, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, P. R. China
6.Institute of Infectious Diseases, Shenzhen Bay Laboratory, Shenzhen, P. R. China
7.Shenzhen International Institute for Biomedical Research, Shenzhen, P. R. China.

†. These authors contribute equally
*Corresponding：**kangxi@szbl.ac.cn**
wshen@szbl.ac.cn

Abstract

While immune cell therapies have transformed cancer treatment, achieving comparable success in solid tumors remains a significant challenge compared to hematologic malignancies like non-Hodgkin lymphoma (NHL) and multiple myeloma (MM). Over the past four decades, various immunotherapeutic strategies, including tumor vaccines, tumor-infiltrating lymphocyte (TIL) therapies, and T-cell receptor (TCR) therapies, have demonstrated clinical efficacy in select solid tumors, suggesting potential advantages over CAR-T and CAR-NK cell therapies in specific contexts. The dynamic nature of the cancer-immunity cycle, characterized by the continuous evolution of tumor-specific neoantigens, enables tumors to evade immune surveillance. This highlights the urgent need for rapid and accurate identification of functional tumor neoantigens to inform the design of personalized tumor vaccines. These vaccines can be based on mRNA, dendritic cells (DCs), or synthetic peptides. In this study, we established a novel platform integrating immunoprecipitation with mass spectrometry (IP-MS) for efficient and direct identification of tumor-specific neoantigen peptides. By combining this approach with our proprietary AI-based prediction algorithm and high-throughput *in vitro* functional validation, we can generate patient-specific neoantigen candidates within six weeks, accelerating personalized tumor vaccine development.

Introduction

Cancer remains a leading cause of mortality worldwide. Despite advancements such as CAR-T cell therapy and therapeutic antibodies, these treatments exhibit efficacy only in specific cancer types(Miao et al., 2024, Vera et al., 2024). These therapies typically target Tumor-Associated Antigens (TAAs) expressed on the cell membrane surface, activating cytotoxic effector cells either directly or via antibody-dependent cellular cytotoxicity (ADCC)(Guan et al., 2024). However, TAAs often show the antigen density, presents a quantitative rather than qualitative difference between tumor and normal cells, leading to "on-target, off-tumor" side effects, which can be toxic(Dharani et al., 2024). In contrast, Tumor-Specific Antigens (TSAs), derived from mutated tumor proteins and processed into 8-12 amino acid peptides (neoantigens), offer a more selective approach. These neoantigens are presented on the cell surface by the Major Histocompatibility Complex (MHC), known as HLA in humans (Abdollahi et al., 2024, Janelle et al., 2020), forming a peptide-MHC complex (pMHC) that can trigger specific T cell immune responses through T Cell Receptor (TCR) engagement(Smith et al., 2015). This pMHC-TCR interaction forms the basis for tumor vaccines, a promising therapeutic modality(Buonaguro and Tagliamonte, 2023, Nguyen et al., 2016).

Tumor vaccines, with a history spanning over 25 years, have generally progressed through three main stages of development: pure or modified peptides, Dendritic Cell (DC) vaccines, and mRNA vaccines. Early clinical trials utilized synthesized tumor neoantigen peptides to treat patients with metastatic melanoma(Rosenberg et al., 1998). Over the past decade, advances in tumor vaccines have accelerated, paralleling progress in other immunotherapies. Loading neoantigens onto DCs, or engineering DCs to present neoantigens, has become an increasingly adopted strategy(Ding et al., 2021, Lesterhuis et al., 2011). In 2010, Sipuleucel-T (Provenge®), the first DC vaccine, was approved for the treatment of prostate cancer (Borno et al., 2020, Sartor et al., 2020), leading to the registration of hundreds of clinical studies. Concurrently, the widespread adoption of Next Generation Sequencing (NGS) has facilitated

the expansion of neoantigen identification for diverse indications. For instance, researchers reported the discovery of 26 significant gene mutations through Whole-Exome Sequencing (WES), subsequently screening the KRAS (G12D) mutation and observing tumor regression following re-infusion of TIL cells stimulated by these neoantigens(Tran et al., 2016, Tran et al., 2014).

mRNA tumor vaccines also utilize WES to sequence patient tumor samples and predict the interaction between MHC-I epitopes and potential neoantigens. Subsequently, mRNA loaded with these predicted neoantigens is administered to patients, resulting in observed therapeutic efficacies(Lang et al., 2022, Sahin et al., 2017). This technology was the foundation for BioNTech, which gained global recognition for its COVID-19 mRNA vaccine. Beyond viral mRNA vaccines, BioNTech has advanced three tumor vaccine pipelines with personalized neoantigens into phase II trials, while Moderna has three clinical phase III pipelines(Kopetz et al., 2022, Weber et al., 2024).

Next-Generation Sequencing (NGS) currently remains the predominant technique in tumor vaccine development. By comparing transcriptome sequences from patient derived normal and tumor tissues, researchers identify mutated gene sequences within tumor tissue and perform comprehensive analyses to predict potential tumor neoantigens(Karasaki et al., 2017, Peng et al., 2019). However, these candidates are predicted from DNA/RNA sequences, rather than directly from the antigenic peptides (8-12mers) presented by MHC molecules. Consequently, regardless of whether mutations are point mutations or frameshift mutations, the precise start and end points of these neoantigens cannot be accurately predicted. Although AI is increasingly being leveraged to enhance NGS-based prediction, direct discovery of *bona fide* neoantigens is considered more reliable. Furthermore, the same TSA cannot guarantee the generation of identical neoantigens in the same patient across different time points(Ford et al., 2018, Sharpnack et al., 2022). Therefore, a significant need exists for the rapid and accurate identification of authentic tumor-specific neoantigens.

Recently, Immunoprecipitation-Mass Spectrometry (IP-MS) has emerged as a method for

identifying tumor-specific neoantigen sequences, representing authentic peptides eluted from the pMHC complex(Huang et al., 2024, Pak et al., 2021, Sturm et al., 2021). In this study, we enhanced the IP-MS method by integrating RNA-seq to calibrate potential errors arising from MS *de novo* sequencing. Furthermore, our collaborator developed a novel algorithm that yielded divergent prediction results compared to netMHCpan4.1, and these results were subsequently validated through functional experiments. This collaborative approach enriches the research process and leads to more innovative outcomes.

Result

Advancing Tumor-Specific Neoantigen Discovery

Neoantigen-based tumor vaccines have shown promise over the past two decades. The conventional approach for personalized tumor neoantigen screening uses DNA/RNA next-generation sequencing (NGS), which translates nucleotide sequences into amino acid sequences to predict potential neoantigen combinations(Ding et al., 2021, Rojas et al., 2023). Although artificial intelligence improves these predictions, it still falls short of directly identifying authentic neoantigen peptides. Immunoprecipitation-mass spectrometry (IP-MS) has recently emerged as a method for identifying tumor neoantigens(Huang et al., 2024, Pak et al., 2021, Sturm et al., 2021), often used with algorithms like netMHCpan4.1 for neoantigen scoring (Fig. 1A). Given the co-evolution of tumors and the immune system, driven by the cancer-immunity cycle, the most effective neoantigens must be personalized for each patient and can change over time(Mellman et al., 2023). Even within the same patient. Rapid and accurate identification of functional tumor neoantigens is, therefore, crucial for optimizing therapeutic windows and maximizing efficacy in treating relapsed and refractory (R/R) cancers.

Because NGS methods do not directly provide peptide information displayed in pMHC complexes, we chose the IP-MS method. We improved the IP-MS approach by integrating RNA-seq data to correct potential errors from MS *de novo* sequencing. Furthermore, our collaborator Dr. Yuzong Chen developed a novel algorithm (details to be published

152 separately), which provides different prediction results compared to netMHCpan4.1 (Fig.
153 1B). This platform has been biochemically validated, tested with tumor cell lines, and
154 analyzed with clinical samples (data not shown). In this study, we describe the technical
155 aspects of this platform using a tumor cell line and its xenograft model.

156 **Immunoprecipitation of pMHC complexes for LC-MS/MS analysis**

157 To identify presented peptides, we performed pMHC-immunoprecipitation followed by mass
158 spectrometry (MS) on a xenograft tumor sample. To simulate *in vivo* tumor growth, the MIA
159 PaCa-2 cell line was subcutaneously injected into SCID-beige mice. Once tumors reached a
160 volume of 1000 mm^3, they were harvested and lysed to extract proteins. These proteins were
161 then immunoprecipitated using the W6/32 antibody, which was validated for quality control
162 before immunoprecipitation (Fig. 2A-B). To confirm the antibody's ability to enrich the
163 pMHC complex under defined biochemical conditions, we expressed the HLA heavy chain
164 A2H (HLA-A2) and the HLA light chain β2M in *E. coli*. The purified A2H, β2M, and a
165 synthesized peptide were then refolded into the pMHC complex (Fig. S1) following
166 established protocols. The results demonstrated that W6/32 successfully immunoprecipitated
167 the pMHC complex, as evidenced by a sharp, clear band for A2H, while the β2M band
168 appeared condensed with nonspecific bands at the bottom (Fig. 2C).

169 Furthermore, W6/32 was used to enrich the pMHC complex from MIA PaCa-2 xenograft
170 tumor lysate, and the results indicated successful precipitation of the pMHC complex (Fig.
171 2D). The eluted peptides were subsequently ultrafiltered and analyzed by LC-MS/MS.
172 PEAKS Online was used to analyze the MS raw data, with the principles of the PEAKS
173 Online workflow and Data-Dependent Acquisition (DDA) analysis detailed in Fig. 3A-B.

174 **Calibrating peptide sequences with RNA-seq and ORF library**

175 Although IP-MS can directly identify tumor neoantigens, rigorous validation is essential.
176 Peptide *de novo* sequencing relies on tandem mass spectrometry to calculate the mass of
177 amino acid residues on the peptide backbone based on the mass difference between two
178 fragment ions. During this process, precursor ions of neoantigens are isolated in the

179 quadrupole and fragmented via collision-induced dissociation with nitrogen gas, producing b-
180 and y- ions. Sufficient fragmentation, indicated by high cleavage coverage along the
181 neoantigen peptide sequence, provides critical information for the *de novo* sequencing
182 algorithm. In our analysis, incomplete peptide fragmentation during mass spectrometry (MS)
183 can lead to ambiguous sequence assignments (Fig. 3C). For example, a fragmentation pattern
184 with 87.5% sequence coverage (7 out of 8 possible cleavage sites identified) suggested that a
185 neoantigen sequence could be either "PLVDTHKSR" or "PLVDTHSKR." This uncertainty,
186 arising from incomplete fragmentation, can introduce errors in *de novo* assembly. To address
187 this, we created a peptide database of open reading frame (ORF) translations based on RNA-
188 seq data, termed ORF-DB. This peptide library was generated through ORF translation (Fig.
189 S2), using ORF-DB as a reference to assemble RNA-seq reads into amino acid sequences for
190 MS analysis. To ensure accuracy, RNA-seq results underwent stringent quality control,
191 achieving an overall alignment rate of 94.69%. This process yielded 14,390 genes, 33,676
192 transcripts, and 22,875 that matched the reference genome (Table S1-1#). In total, we
193 identified 18,211 single-nucleotide polymorphisms (SNPs) and 1,567 insertions/deletions
194 (indels) in the sample (Table S1-2#). On average, there were 3,228 synonymous and 2,070
195 non-synonymous mutations within the coding region (Table S1-3#). Overall, 1,480 mutated
196 genes were identified among the 2,070 non-synonymous mutations (Table S2).

197 **Neoantigen identification in MIA PaCa-2 xenograft tumors**

198 The software module PEAKS Online was used to analyze the MS raw data for neoantigen
199 identification via database searching and *de novo* sequencing. For database searching, we
200 used the widely accepted proteomics reference library, UniProt (human, Swiss-Prot).
201 Matching the MS results with the UniProt database identified 1,869 peptides, designated
202 IP(UniProt) (Table S3). Following our established workflow, we also aligned the MS results
203 with our ORF-DB, yielding 1,526 peptides (designated IP(ORF-DB), Table S4). Analysis
204 indicated that most identified peptides were between 8 and 12 amino acids in length, with 9-
205 mer peptides being the most abundant. Prediction results showed that peptides binding to
206 HLA-A*24:02 were predominantly 9-mer and 10-mer peptides (Fig. 3D). To validate the

accuracy of IP(ORF-DB), we compared the overlap between IP(UniProt) and IP(ORF-DB). Using a CAA threshold (CAA% > 0), we obtained 614 peptides of 8-12 amino acids from IP(UniProt) and 748 peptides of the same length from IP(ORF-DB). This comparison revealed 442 overlapping peptides (Fig. 3E left, Table S6), indicating high feasibility of the results.

Next, we aimed to validate the accuracy of the peptides identified through *de novo* sequencing (Table S5). Comparing the *de novo* sequencing dataset with the database search results, we found 284 identical peptides between IP (*de novo*) and IP (UniProt), accounting for 62% of the IP (*de novo*) peptides (Fig. 3E middle, Table S7). Additionally, there were 277 overlapping peptides when compared with the dataset from ORF-DB, constituting 60% of the IP (*de novo*) peptides (Fig. 3E right, Table S8). These findings suggest that errors may have occurred during peptide sequence assembly in the *de novo* process.

Given the significant interest in HLA-A in immunotherapies, we identified the HLA genotype of MIA PaCa-2 as HLA-A*24:02 (Fig. S3). Through comprehensive data analysis, we focused on the most frequent peptides, specifically those with lengths of 8-12 amino acids. Using both netMHCpan4.1 and our new algorithm, we predicted their binding affinity to HLA-A*24:02, with 9-mer peptides constituting the largest proportion (Table 1).

Moreover, the W6/32 antibody binds to all MHC Class I molecules (i.e., HLA-A, HLA-B, and HLA-C), which are crucial for T cell activation. Each individual inherits one set of HLA genes from each parent, leading to a unique combination of HLA alleles. Theoretically, an individual can possess up to 12 different HLA subtypes. Following predictions using netMHCpan4.1 and our new algorithm, we screened out 98 and 108 peptides, respectively (Table S9). In summary, despite bioinformatic predictions, the potential candidate pool remains extensive, highlighting the necessity for high-throughput functional testing.

Comparison of neoantigen identification by IP-MS and MAE

Given that the mild acid elution (MAE) method has been previously used for neoantigen identification(Abdollahi et al., 2024, Sturm et al., 2021), we incorporated this technique

234 alongside our current IP-MS strategy for comparison. Aligning the MAE MS results with the
235 UniProt database, we identified 1980 peptides with lengths of 8-12 amino acids (CAA% > 0)
236 (Table S10-1#). However, only 44 peptides overlapped between the MAE(UniProt) and
237 IP(UniProt) datasets (Fig. S4, Table. S10-2#). Among these, 10 peptides were predicted to bind
238 to HLA-A*24:02 (Fig. S4A), suggesting that the MAE method can offer some insights into
239 specific neoantigens. However, because the peptides identified by the MAE method did not
240 undergo pMHC enrichment, they were more susceptible to contamination by non-target
241 proteins, resulting in a total of 4575 peptides in MAE(UniProt) (Table. S10-3#,4#). This
242 increased complexity in screening for true neoantigens ultimately leads to lower confidence in
243 the peptides obtained via the MAE method compared to our IP-MS method.
244

245 ***In vitro* validation of neoantigen functionality**

246 To validate the functionality of identified neoantigens, we selected 19 peptide sequences
247 (Table 1, Fig. S5) compatible with the HLA-A*24:02 allele for *in vitro* assays. These
248 candidates were derived from the IP(ORF-DB) dataset, and their binding affinity was
249 predicted using both netMHCpan4.1 and our new algorithm. Two peptides were classified as
250 non-binders (NB) by both algorithms. The remaining 17 peptides, originating from the 442
251 peptides common to IP(UniProt) and IP(ORF-DB), were predicted to bind to HLA-A*24:02.
252 The predicted binding levels of peptides No. 1-15 were similar between the two algorithms;
253 however, the predicted scores for peptides No. 16-19 showed discrepancies.

254 Through this process, we screened candidate antigenic peptides with high predicted affinity
255 for HLA-A*24:02. To determine whether these peptides genuinely possess immunogenicity,
256 *in vitro* validation was performed (Fig. 4A). Given that the natural TCR-neoantigen
257 interaction elicits a mild T cell activation signal, we fine-tuned the *in vitro* experimental
258 design to mimic physiological conditions and demonstrate specific cytotoxicity. To mimic
259 natural T cell signaling, we stimulated peripheral blood mononuclear cells (PBMCs),
260 containing both antigen-presenting cells (APCs) and T cells, with individual peptides. After 5
261 days of stimulation, the PBMCs were tested using IFN-γ ELISPOT and T cell degranulation

assays. To demonstrate specific cytotoxicity *in vitro* following peptide stimulation, anti-CD3/CD28 antibodies were added to activate T cells. Cytotoxicity was assessed using real-time Incucyte live imaging or bioluminescence assays, depending on the experimental needs.

To test our candidates *in vitro*, we first conducted a primary screening of all candidate peptides. Synthesized peptides were used to stimulate allogeneic donor's HLA-A*24:02 PBMCs, with or without anti-CD3/CD28 antibody. IFN-γ ELISPOT and T cell cytotoxicity assays were then performed. Initially, the peptides were combined into 4-5 peptide pools for rough screening. Subsequently, reactive pools were selected for individual fine screening. Preliminary results indicated that POOL-2 (peptides No.6-10), POOL-3 (peptides No.11-15), and POOL-4 (peptides No.16-19) induced IFN-γ secretion. Unexpectedly, POOL-1 (peptides No.1-5) failed to induce IFN-γ secretion or PBMC cytotoxicity (Fig. 4B-4D), even though both algorithms provided strong binding predictions (Table 1). These results underscore the importance of experimental validation to confirm prediction results before developing tumor vaccines. Furthermore, the killing assay confirmed that POOL-2, POOL-3, and POOL-4 induced PBMCs-mediated cytotoxicity (Fig. 4E). The target cell-specific lysis was inhibited by an HLA blocking antibody, consistent with the known ability of the W6/32 antibody to inhibit pMHC-TCR interaction(Matsushita et al., 2016, Yao et al., 2023). Next, to accurately identify the most immunogenic peptide, we stimulated PBMCs with each peptide individually, repeating the IFN-γ ELISPOT assay and evaluating cytotoxicity. Each peptide from POOL-2, 3, and 4 was subjected to individual experiments. The results showed that peptides No. 6-15 induced IFN-γ secretion independently (Fig. 5A-5B) and stimulated PBMCs to lyse cancer cells (Fig. 5C-5D). The T cell degranulation assay also indicated that peptides No. 6-15 activated T cells to varying degrees (Fig. 5E-5G, Fig. S6). These results generally aligned with both algorithms' predictions, although some discrepancies were noted.

Comparison of prediction accuracy between algorithms

This study employed netMHCpan4.1, an AI algorithm based on a neural network, alongside our innovative new algorithm developed by Dr. Yuzong Chen (publication forthcoming). This new algorithm is a deep learning-based predictive model integrating biological knowledge and

trained on quantitative measurements of peptide-MHC binding affinity(Shen et al., 2022, Shen et al., 2021). Given that both neural networks and deep learning are major algorithms for AI development, we compared their differences in Table S11. Our experimental findings suggest that the predictions made by the two algorithms show a considerable degree of agreement. However, we observed inconsistencies, particularly in the peptide predictions No. 16-19 (Fig. 6A). To ascertain which algorithm yielded more accurate results, we conducted further *in vitro* evaluation experiments. The results showed that peptides predicted as Non-Binders (NB) by netMHCpan4.1, specifically No. 16, No. 17, and No. 19, stimulated immune cells to produce IFN-γ and triggered immune cells to eliminate cancer cells. Conversely, peptide No. 18, predicted as Weak Binding (WB) by netMHCpan4.1, did not exhibit these functionalities (Fig. 6B-6G). These findings indicate that our new algorithm's predictions surpass those of netMHCpan4.1 in terms of accuracy.

TCR activity towards the validated neoantigen

Among the neoantigens screened, an ideal candidate would be functional for developing tumor vaccines or TCR-based therapeutics. We identified peptide No. 19, derived from Survivin, a promising target with an existing clinical-phase TCR drug from AbbVie(Chervin, 2021). ABBV-184 is a TCR-CD3 engager protein drug targeting the same peptide as No. 19 in our list(Chervin et al., 2023, Peterlin et al., 2024). To evaluate the binding between the MHC complex loaded with peptide No. 19 and the TCR, as well as the consequent T cell signaling, we performed an *in vitro* T cell-specific activation assay (Fig. 7A). First, T2 cells, which express MHC molecules without any pre-loaded peptides, were loaded with FITC-conjugated peptide No. 19. Peptide loading was verified by flow cytometry (Fig. 7B). Second, we established a Jurkat NFAT-GFP reporter cell line (Fig. S7). Furthermore, we cloned the single-chain TCR sequence (scTCR) from ABBV-184 into a lentiviral-based 2nd generation CAR backbone and transduced this scTCR into the Jurkat NFAT-GFP cell line to establish a stable scTCR reporter cell line (Fig. 7C). Co-culturing peptide No. 19-loaded T2 cells with either wild-type or scTCR-stable Jurkat NFAT-GFP cells, we observed that scTCR Jurkat cells significantly boosted NFAT-GFP reporter expression upon contact with peptide No. 19-loaded T2 cells, consistent with previous reports (Fig. 7D). Finally, we harvested the co-cultured cells,

320 stained them with a T cell marker (CD3) and a T cell activation marker (CD69), and analyzed
321 them via flow cytometry alongside the NFAT-GFP reporter (Fig. 7E-7H). These results
322 demonstrate that peptide No. 19, identified using our platform, can react with its targeted TCR
323 and induce a robust T cell activation signal.
324

325 Discussion

326 Neoantigen discovery and identification, particularly of tumor-specific antigens resulting
327 from cancer cell mutations, offers significant potential for improving cancer treatment and
328 advancing tumor immunotherapy. Given the critical role of individual genetic backgrounds in
329 determining neoantigen effectiveness, personalized approaches to tumor vaccines are
330 essential.

331 Detecting neoantigen expression in tumor tissues enables the screening of suitable antigens
332 for vaccine design. This enhances therapeutic efficacy by ensuring vaccines elicit a robust
333 immune response against the unique mutations present in an individual's tumor. Rapid
334 neoantigen identification can improve treatment outcomes, reduce recurrence rates, and
335 propel tumor immunotherapy forward.

336 Despite advancements in next-generation sequencing (NGS) technology, directly identifying
337 neoantigen peptide sequences remains challenging due to the complexities of peptide splicing
338 and post-translational modifications that NGS alone cannot resolve. To address these
339 limitations, a novel approach has been introduced: constructing a tumor tissue-specific
340 peptide database by translating open reading frames (ORFs) derived from RNA-seq results.
341 This ORF-DB more accurately reflects the tumor's characteristics compared to traditional
342 databases like UniProt, enhancing neoantigen sequence identification precision. Integrating
343 mass spectrometry (MS) results can further aid in obtaining neoantigen sequence information
344 and predicting their binding affinity to HLA alleles. Enrichment-based approaches
345 significantly improve neoantigen identification accuracy, as traditional methods like mild
346 acid direct elution (MAE) may yield lower authenticity due to potential contamination with

347 non-target proteins. A collaborative approach enriches the research process and leads to more
348 innovative and inclusive outcomes.

349 This research aims to develop a method for the rapid and precise identification of effective
350 neoantigens from tumor tissues. Initially, a peptide database specific to the tumor tissue was
351 constructed by translating open reading frames (ORFs) derived from RNA-seq results on the
352 tumor tissue. Although the traditional UniProt database (human, Swiss-Prot) contains a more
353 significant number of proteins and can provide more peptide information, the ORF-DB
354 developed in this study more accurately reflects the characteristics of the tumor tissue than
355 the UniProt database. By comparing these two database search methods, information about
356 bona fide neoantigens can be accurately obtained. *De novo* sequencing serves as a
357 fundamental analytical function of PEAKS. However, during the experimental process, it was
358 found that errors can occur during the assembly process due to incomplete fragmentation of
359 peptide segments. Such errors can be rectified through a database search with ORF-DB. This
360 approach has the potential to enhance the precision of neoantigen sequence identification.

361 The RNA-seq result indicates the presence of 1,480 mutated genes in the MIA PaCa-2 cell
362 line. However, the neoantigens produced by these mutations were not identified, indicating
363 that the proteins encoded by the mutated genes may not necessarily be presented by HLA.
364 Previous studies have also indicated that the mutation burden in pancreatic cancer is
365 relatively low, which aligns with these findings(Rojas et al., 2023).

366 After analyzing the mass spectrometry results to obtain neoantigen sequence information,
367 these results were combined with the sequencing outcomes of HLA typing. The binding
368 affinities of the neoantigens for HLA-A*24:02 were predicted and scored using netMHCpan
369 4.1 and a proprietary algorithm. The scores produced by the developed algorithm were
370 mostly consistent with those from netMHCpan 4.1.

371 Additional studies are necessary to verify the potential applications of these neoantigens. It is
372 important to understand their biological functions in detail. For example, *in vitro* assays have
373 demonstrated that the representative neoantigens are immunogenic. Even some peptides are

predicted to have strong binding to HLA-A*24:02 but are unable to induce an immune response in PBMCs.

The findings suggested that the two algorithms' predictions are almost consistent. Nevertheless, discrepancies were noted in some neoantigens. To ascertain the reliability of each algorithm's predictions, supplementary *in vitro* evaluations were carried out. These experiments showed that *in vitro* assays are more consistent with the prediction from the new algorithm. These results underscore the superior accuracy of the new algorithm compared to netMHCpan4.1 in predicting peptide binding affinity.

In summary, the rapid and precise identification of effective neoantigens from tumor tissues is crucial for enhancing cancer treatment and advancing tumor immunotherapy. The development of tailored tumor vaccines based on individual genetic profiles, coupled with innovative methodologies for neoantigen discovery, represents a significant stride toward personalized cancer therapies. This research not only contributes to the understanding of neoantigen biology but also lays the groundwork for future advancements in cancer immunotherapy, ultimately aiming to improve patient outcomes and reduce the burden of cancer recurrence.

By establishing a robust workflow for neoantigen discovery and validation, researchers can expedite the transition from identification to pre-clinical functional evaluation, marking a significant milestone in the development of effective cancer vaccines. In business research, incorporating psychological concepts like Theory of Mind (ToM) into management theories exemplifies interdisciplinary synergy, fostering more robust and versatile research frameworks. ToM enables researchers to construct arguments that are attuned to the mental states and motivations of their audience, ensuring that arguments are both persuasive and empathetic.

Methods

Cells

401 MIA PaCa-2: The MIA PaCa-2 cell line (American Type Culture Collection [ATCC], Manassas,
402 VA, USA) was cultured in DMEM supplemented with 10% FBS and 1% antibiotics at 37°C in
403 a 5% CO2 atmosphere. Cells were passaged every 2-3 days at a sub-cultivation ratio of 1:3
404 using trypsinization and were assumed to have undergone 20 passages.

405 293T: 293T cells, provided by cell stock of Shenzhen Bay Laboratory, were cultured in
406 DMEM supplemented with 10% FBS and 1% antibiotics at 37°C in a 5% CO2 atmosphere.

407 W6/32 Hybridoma: The W6/32 hybridoma cell line (ATCC, Manassas, VA, USA) was
408 cultured in DMEM supplemented with 10% FBS and 1% antibiotics at 37°C in a 5% CO2
409 atmosphere. Cells were maintained in triangular culture bottles and passaged every 4-5 days
410 at a sub-cultivation ratio of 1:3. Each cell line was assumed to have undergone 5 passages.

411 T2: The T2 cell line (ATCC, Manassas, VA, USA) was cultured in RPMI 1640 supplemented
412 with 10% FBS and 1% antibiotics at 37°C in a 5% CO2 atmosphere. Cells were passaged
413 every 2-3 days at a sub-cultivation ratio of 1:3.

414 CHO-S: CHO-S cells, provided by cell stock of Shenzhen Bay Laboratory, were cultured in
415 DMEM supplemented with 10% FBS and 1% antibiotics at 37°C in a 5% CO2 atmosphere.

416 Jurkat: Jurkat cells, provided by cell stock of Shenzhen Bay Laboratory, were cultured in
417 RPMI 1640 supplemented with 10% FBS and 1% antibiotics at 37°C in a 5% CO2
418 atmosphere.

419 All cell lines were preserved in liquid nitrogen using a mixture of 95% FBS and 5% DMSO.

420 hPBMC: Human peripheral blood mononuclear cells (hPBMC) were obtained from human
421 peripheral blood provided by Taicang First People's Hospital and hPBMC isolation service
422 provided by Milestone Biotechnologies. hPBMC were preserved in liquid nitrogen using
423 STEMCELL Technologies' CryoStor® CS10 freezing buffer.

424 **Lentiviral Packaging**

425 HEK293T cells were utilized as packaging cells for lentivirus production via transfection.
426 Briefly, 1×10^7 293T cells were seeded onto 10 cm plate and cultured to 80% confluence. A
427 mixture of 60 μg of PEI (Polysciences, 24765-1) and 24 μg of plasmid DNA (pSpAX2:
428 pMD2.G: target vector at a ratio of 5:7:12) was prepared in 0.5 mL Opti-MEM
429 (ThermoFisher, 11058021) by incubating each component separately for 5 minutes, followed
430 by combining them for 20 minutes before transfection. Eight hours post-transfection, the
431 medium was replaced with fresh, antibiotic-free medium supplemented with 2% FBS.
432 Supernatants containing lentivirus were collected at 48 and 72 hours post-transfection by
433 low-speed centrifugation and filtered through a 0.45 μm syringe filter (Millipore,
434 SLHVR33RB). The resulting lentivirus supernatant was either used immediately,
435 concentrated by centrifugation at 20,000g for 2 hours, or snap-frozen in liquid nitrogen and
436 stored at -80°C.

437 **Generation of MIA PaCa-2 stable cell line**

438 MIA PaCa-2 cells were seeded 24 hours before transduction. Lentivirus expressing either
439 luciferase or mKate was diluted in DMEM supplemented with 1 μg/mL polybrene. The existing
440 cell culture medium was removed, and the lentivirus/polybrene mixture was directly applied to
441 the cells, followed by incubation for 24 hours at 37°C. Subsequently, the medium was replaced
442 with fresh culture medium. At 48 hours post-transduction, puromycin (1 μg/mL) was added to
443 the medium for selection of successfully transduced cells. Luciferase substrate assays were
444 used to quantify luciferase expression in the MIA PaCa-2-luciferase cell line. The expression
445 of red fluorescent protein (mKate) in the MIA PaCa-2-mKate cell line was validated using
446 IncuCyte imaging.

447

448 **Generation of Jurkat NFAT-GFP reporter cell line**

449 NFAT-GFP lentiviral plasmid was generously provided by Dr. Chen Yu from Shenzhen Bay
450 Laboratory. To generate a stable reporter cell line, Jurkat cells were infected with lentivirus
451 encoding NFAT-GFP. Following lentiviral transduction, single-cell clones were sorted into U-
452 bottom 96-well plates and cultured in RPMI 1640 supplemented with 20% FBS, 1 μg/mL
453 puromycin, and 1% antibiotics. After clonal expansion, each clone was passaged at a 1:2 split

454 ratio. One aliquot of each clone was either untreated or treated with PMA-Ionomycin
455 (ThermoFisher, 00-4975-03). GFP reporter signals were then evaluated via FACS analysis.
456 Positive clones were identified, recorded, and the corresponding aliquot was selected for
457 further experiments.
458

459 **Generation of Jurkat-scTCR stable cell line**

460 The T-cell receptor (TCR) sequence of ABBV-184 that targets No.19 peptide (Chervin, 2021)
461 was synthesized and cloned into a lentiviral vector as a single-chain TCR (scTCR) fused with
462 a V5 tag and the transmembrane and intracellular signaling domains of a second-generation
463 chimeric antigen receptor (CAR). Jurkat NFAT-GFP reporter cells were then transduced with
464 the packaged lentivirus. At 48 hours post-transduction, transduction efficiency was assessed by
465 flow cytometry (FACS) via V5 tag staining (ThermoFisher, 12-6796-42). V5-positive
466 polyclonal cells were subsequently sorted to establish a stable cell line.
467

468 **Jurkat-NFAT-GFP Activation Assay**

469 Wild-type Jurkat NFAT-GFP cells or Jurkat NFAT-GFP cells stably expressing scTCR were co-
470 cultured with peptide-loaded T2 cells at a 1:1 ratio and incubated for 24 hours. NFAT-GFP
471 reporter activity was then quantified using a High-Content Imaging System (Perkin Elmer,
472 Opera Phenix Plus) and analyzed with Harmony software. Following co-culture, cells were
473 harvested and stained with a viability dye (Tonbo Bio, 13-0870), anti-CD3 (ThermoFisher, 25-
474 0037-42), and anti-CD69 (Biolegend, 310922) antibodies to assess T cell activation. CD69
475 expression (a marker of T cell activation) and GFP expression (NFAT reporter activity) were
476 quantitatively analyzed by flow cytometry.
477

478 **Xenograft model**

479 Female CB17.B6-*Prkdc*^{scid}*Lyst*^{bg}/Crl mice (4-6 weeks old, from Beijing Vital River Laboratory
480 Animal Technology Co., Ltd) were used for *in vivo* tumorigenicity studies. Each mouse
481 received a single subcutaneous injection (200 μL) into the right flank, consisting of 1×10^7
482 MIA PaCa-2 cells suspended in PBS and Matrigel. Injection sites were monitored until palpable
483 tumors were established. Once tumors reached the predetermined size, they were excised for

484 subsequent analysis. All animal studies were conducted following the guidelines and
485 regulations of the Regional Ethics Committee for Animal Experiments at Shenzhen Bay
486 Laboratory. Mice were maintained under standard housing conditions with a 12-hour light/dark
487 cycle and ad libitum access to food and water.
488

489 **Production of W6/32 antibody**

490 8-10 weeks old Balb/c mice from the same consistent lineage were utilized for monoclonal
491 antibody production. To enhance hybridoma cell engraftment, mice were pre-sensitized with
492 an intraperitoneal injection of mineral oil one week before hybridoma cell inoculation.
493 Subsequently, 1×10^6 hybridoma cells were injected into the peritoneal cavity of each mouse.
494 Approximately 7 days post-injection, ascites fluid was harvested. The W6/32 antibody was
495 then purified from the ascites fluid using Protein A Sepharose affinity chromatography.
496 Following elution, the buffer was exchanged for PBS via ultrafiltration.
497

498 **W6/32 and Cell binding**

499 T2 and CHO-S cells were cultured in their respective optimized media. Upon reaching the
500 desired confluency, cells were harvested and centrifuged at 300g for 5 minutes to remove the
501 supernatant. To eliminate residual culture medium components, cells were resuspended in PBS
502 and washed twice. The W6/32 antibody was diluted in PBS to the appropriate working
503 concentration. The diluted W6/32 antibody was added to the cell suspension, with an isotype-
504 matched mouse IgG (mIgG) as a negative control. The mixture was incubated at room
505 temperature for 30 minutes, unbound primary antibody was rinsed, and secondary antibody
506 (CST, 4410S) was then applied to incubate at room temperature in dark for an additional 30
507 minutes. Cells were washed 2-3 times with PBS and were resuspended in PBS at a
508 concentration of 1×10^6 cells/mL and transferred to sample tubes for flow cytometry analysis.
509

510 **Western Blot**

511 Equal amounts of protein were separated by SDS-PAGE using polyacrylamide gels with a
512 concentration gradient ranging from 10-15%. Following electrophoresis, proteins were
513 transferred to a PVDF membrane, which was then blocked with 5% non-fat dry milk. The

514 membrane was subsequently incubated with the following primary antibody: anti-HLA (HC10,
515 ThermoFisher, MUB2037P) at a 1:1000 dilution. After washing, the membrane was incubated
516 with anti-HRP-conjugated secondary antibodies (CST, 7074 or 7076) at a 1:5000 dilution.
517 Protein bands were visualized using enhanced chemiluminescence (ECL) assays (Beyotime,
518 P0018S). Images were cropped to display the molecular weight marker proteins (kDa).
519

520 **HLA Genotyping**

521 HLA genotyping of the MIA PaCa-2 cell line was performed by BGI. Genomic DNA was
522 extracted, and HLA genes were amplified via PCR using sequence-specific primers. The
523 amplified products were then purified and prepared for sequencing. Sequencing was conducted
524 using the 3730xl DNA Analyzer (Applied Biosystems). The resulting sequencing data were
525 analyzed using specialized software to determine the HLA genotype.
526

527 **RNA-seq and ORF translation**

528 Total RNA was extracted using RNAiso (Takara Bio, 9019). Poly(A) enrichment or ribosomal
529 RNA depletion using specific probes was employed to remove ribosomal RNA. Strand-specific
530 RNA-seq libraries were constructed, selecting for insert fragment sizes between 350 and 450
531 base pairs. Sequencing was performed on the Illumina NovaSeq™ X Plus platform using
532 paired-end 150 bp reads to generate deep sequencing data. Reads were aligned to the hg38
533 reference genome (Gencode V38) using Hisat2 V2.2.1 with the following parameters: -rna-
534 strandness RF -dta -no-mixed -no-unal. Transcript assembly was performed using Stringtie
535 2.2.2 (parameter -rf) with reference annotation. Open reading frames (ORFs) and coding
536 regions were predicted from the assembled GTF output using TransDecoder v5.7.1, and these
537 predicted coding regions were subsequently mapped back to the genome.
538

539 **pMHC enrichment by immunoprecipitation**

540 For pMHC isolation, MIA PaCa-2 xenograft tumor tissue (100 mg wet weight) was minced
541 and lysed in 500 μL of CHAPS buffer supplemented with protease inhibitors (ThermoFisher,
542 87785). pMHC complexes were immunoprecipitated using the W6/32 antibody conjugated to
543 Protein A/G magnetic beads (MCE, HY-K0202). Peptides were eluted with 10% acetic acid

544 and subsequently filtered using a 3 kDa molecular weight cut-off ultrafiltration device
545 (Millipore, UFC5003) to retain MHC-I heavy chain (A2H), β2M light chain, and the W6/32
546 antibody. Peptide extracts were then desalted using C18-ZipTips (Millipore, ZTC18S) and
547 concentrated by vacuum centrifugation. Before mass spectrometry analysis, the peptides were
548 resuspended in 0.1% formic acid (ThermoFisher, A117-50).
549

550 **Mild Acid Elution (MAE)**

551 MIA PaCa-2 Xenograft tumor sample (wet weight 50 mg) was cut into small pieces and washed
552 three times with PBS at 200g for 10 mins centrifugation. Tumor sample was put into MAE
553 buffer (10% acetic acid) for 1 min. The MAE buffer was filtered on a 3 kDa molecular weight
554 cut-off ultrafiltration tube (Millipore, UFC5003) to retain MHC-I A2H heavy chain, β2M light
555 chain, and W6/32 antibody. Peptide extracts were desalted by C18-ZipTips and dried using a
556 Speed-Vac. Before MS analysis, the peptides were resuspended in 0.1 % formic acid (FA).
557

558 **LC-MS/MS analysis**

559 LC-MS/MS analysis was performed on an Easy nLC 1200 (ThermoFisher, Bremen, Germany)
560 coupled to an Orbitrap Fusion Lumos equipped with a nanospray flex ion source
561 (ThermoFisher, Bremen, Germany). The peptides were dissolved in water with 0.1% formic
562 acid and separated on a commercial RP-HPLC pre-column (75 μm×2 cm) (ThermoFisher,
563 164535) and RP-HPLC analytical column (75 μm×25 cm) (ThermoFisher, 164941), both
564 packed with 2 μm C18 beads. The peptides were eluted over a 90 mins segmented gradient.
565 The Orbitrap Fusion Lumos acquired data in a data-dependent manner, alternating between
566 full-scan MS and MS2 scans. Isolated precursor ions were sequentially fragmented in a 3 secs
567 cycle. Dynamic exclusion was set to 30 secs, and precursors with charge states were isolated
568 for MS/MS experiments.
569

570 **MS data processing and database searching**

571 Qualitative and quantitative analysis of mass spectrometry raw data was performed using a
572 multi-stage workflow in PEAKS Online 10 (Bioinformatics Solutions Inc., Waterloo, Canada).
573 Initially, the data were searched against the human UniProt database. This was followed by a

574 de novo peptidome sequencing search. Precursor and fragment mass error tolerances were set
575 to 10 ppm and 0.02 Da, respectively. Methionine oxidation (+15.9949 Da) was included as a
576 variable modification. All results were filtered to a 1% false discovery rate (FDR). For *de novo*
577 sequencing, only peptides with a de novo score of 70% or higher were retained for subsequent
578 analysis.
579

580 **Synthetic peptides**

581 Peptides were synthesized using solid-phase peptide synthesis (SPPS) with Fmoc chemistry,
582 proceeding from the C-terminus to the N-terminus. The synthesis employed a resin support,
583 Fmoc-protected amino acids, and appropriate peptide condensation reagents. Following
584 synthesis, peptides were cleaved from the resin and purified by reversed-phase chromatography
585 on a C18 column, using a gradient elution of acetonitrile and water containing 0.05%
586 trifluoroacetic acid. Peptide synthesis was performed using a Symphony X Polypeptide
587 Synthesizer at the Translation Innovation Center of Shenzhen Bay Laboratory or by Genscript.
588 Synthesized peptides were subsequently dissolved in DMSO for downstream assays.
589

590 **T-cell cytotoxicity**

591 Untreated 24-well plates (NEST, 702001) were coated with anti-CD3/CD28 antibodies (1
592 μg/mL, Biolegend, 317347/302943). Peripheral blood mononuclear cells (PBMCs), cultured
593 in a medium consisting of 45% Click's medium, 45% RPMI 1640, 10% FBS, 1% antibiotics,
594 and 20 Units/mL IL-2, were seeded at a density of 1 x 10^6 cells per well. PBMCs were
595 stimulated with the indicated peptide(s) or DMSO. Single peptide treatments were performed
596 at a concentration of 10 μg/mL. For mixed peptides treatments, a pool of 4 or 5 peptides was
597 used at a concentration of 2 μg/mL per peptide. An equivalent volume of DMSO was used as
598 a negative control. Five days post-stimulation, PBMCs were harvested and co-cultured
599 overnight with MIA PaCa-2 cells stably expressing either firefly luciferase (*fluc*) or mKate at
600 an effector-to-target (E:T) ratio of 20:1. Target cells without any treatment served as the
601 spontaneous death control, while target cells lysed with NP-40 lysis buffer represented the
602 maximal killing control. Bioluminescence-based cytotoxicity was quantified using a BioTek
603 Synergy Neo2 plate reader, and specific lysis was calculated using the following equation:

604 Specific lysis (%) = 100 × (Spontaneous death RLU - Test RLU) / (Spontaneous death RLU -
605 Maximal killing RLU). In parallel, IncuCyte-based cytotoxicity was monitored in real-time by
606 quantitative measurement of live, mKate+ tumor cells. Values were normalized to the t=0-hour
607 measurement.
608
609 **ELISPOT**
610 IFN-γ ELISPOT assays were. Peripheral blood mononuclear cells (PBMCs) were seeded at a
611 density of 1 × 10^6 cells per well. PBMCs were stimulated with the indicated peptide(s), PMA-
612 Ionomycin (positive control), or DMSO (negative control). Five days post-stimulation,
613 stimulated PBMCs were. IFN-γ spots were quantified using a CTL S6 Ultra ELISPOT reader.
614 IFN-γ ELISPOT assays were performed using an ELISPOT kit (BD Biosciences, 551849).
615 1×10^6 PBMCs were seeded per well. PBMCs were treated with the indicated peptide(s), PMA-
616 Ionomycin, or DMSO. 5 days post-treatment, stimulated PBMCs were performed using an
617 ELISPOT kit (BD Biosciences, 551849). IFN-γ spots were counted using CTL S6 Ultra.
618
619 **T cell degranulation**
620 Peripheral blood mononuclear cells (PBMCs) were seeded at a density of 1×10^6 cells per well.
621 PBMCs were stimulated with the indicated peptide(s), PMA-Ionomycin (positive control), or
622 DMSO (negative control). Seven days post-stimulation, PBMCs and target cells were co-
623 cultured. Concurrently, APC-conjugated anti-CD107a antibody (Biolegend, 301102) was
624 added to each well at 2.5 μL per test. After 1 hour, Monensin (ThermoFisher, 00-4505-51) was
625 added, and co-culture was continued for an additional 5.5 hours. CD8-FITC antibody
626 (Biolegend, 301505) was added for the last 30 minutes to facilitate T cell gating. CD107a
627 staining was then quantified by flow cytometry.
628

629 ## Authors' contribution

630 **Huajian Tian** and **Guifei Li** conducted experiments, prepared figures, wrote the manuscript;
631 **Cookson K.C. Chiu** and **E Li** provided the Mass Spectrometry service and data analysis,
632 integrated the manuscript preparation; **Yuzong Chen** and **Ting Zhu** provided the novel

algorithm for neoantigen-MHC affinity and prediction; **Min Hu** and **Yanjie Wang** conducted the RNA-seq and data analysis; **Suping Wen** established the Jurkat NFAT-GFP reporter cell line; **Jiajia Li** and **Shuangxue Luo** provided the MIA PaCa-2 xenograft mice model; **Zhicheng Chen** and **Huimei Zeng** produced the W6/32 antibody from hybridoma cell line; **Nan Zheng** provided part of the peptide synthesis; **Jinyong Wang** provided the biosafety level II support and related data generation; **Weijun Shen** supervised the research project and administrate the operation, **Xi Kang** designed the experimental framework, directed the project, conducted result analysis, prepared figures and reviewed the manuscript.

Acknowledgments

We thank Dr. Stephen H.K. Wong (currently in Innoventbio) proposed this project and joined the brainstorming. We thank Dr. Haidi Yin from Biotech Center in Shenzhen Medical Academy of Research and Translation (SMART) joined the brainstorming. We thank Dr. Xuefen Huang from USC reviewed the manuscripts. We thank Dr. Aimin Hui from EnCureGen, and Dr. Junhui Chen from Peking University Shenzhen Hospital for discussing the clinical collaboration of this project. We thank Dr. Nieng Yan, Dr. Huan Tu, and Ms. Nicole KY Zhao for caring about the clinical and commercial translation of this project. This work was financially supported by the startup funding for the Center of Translational Medicine, Shenzhen Bay Laboratory, and Natural Science Foundation of Guangdong Province (2021A1515010478).

Conflict of interest

The authors state that the manuscript was carried out without any commercial relationships that could bring any conflict of interest. Shenzhen Bay Laboratory may still keep the potential for the clinical translation of this study.

Figure Legend

Fig.1 Comparison between traditional neoantigen identification methods and the technical approach in this study.

(A) Current methods for discovering tumor neoantigens.
(B) Schematic diagram of technical methods used in this research.

Fig.2 Immunoprecipitation pMHC complex of Xenograft tumor tissue.
(A)Identify W6/32 antibody purity by SDS-PAGE with non-reducing or reducing loading buffer. N, non-reducing; R, reducing; FL, Full Length; HC, heavy chain; LC, light chain.
(B)Quality Controlling the binding affinity of W6/32 with MHC complex using T2 cell line or negative controlled CHO-S cell line.
(C)Immunoprecipitant the refold pMHC complex with W6/32. All protein samples were loaded with reduced buffer. The result was presented by Coomassie staining, right panel. HC, heavy chain; LC, light chain. Schematic of pMHC refold in biochemical condition, left panel.
(D) Immunoprecipitant the MIA PaCa-2 Xenograft tumor lysate with W6/32. The result was shown by Western Blot.

Fig.3 The analysis of MS result in PEAKS Online.
(A) Overview of data acquisition strategies, analysis approaches, and core algorithms of PEAKS Online.
(B) An example of integrated workflow for DDA data analysis. DDA： Data-dependent acquisition.
(C) An example of ORF-DB calibrated the error of *de novo* sequencing. The *de novo* sequencing mistake was marked in red while the accurate peptide sequence in ORF-DB was marked in green. ORF-DB: a peptide database of open reading frame (ORF) translations based on RNA-seq data.
(D) Peptides length distribution and Sequence logos of IP(UniProt) and IP(ORF-DB). (E) The peptides overlapping of IP(UniProt) vs. IP(ORF-DB), IP(UniProt) vs. IP (*de novo)* and IP(ORF-DB) vs. IP (*de novo*); CT(UniProt)/ CT(ORF-DB)/ CT (*de novo*) was a control group to filter the non-specific binding peptides; IP: immunoprecipitated with W6/32; CT: immunoprecipitated with mIgG.

Fig.4 Neoantigens POOLs activated the Immune responses and cytotoxicity. (A)Schematic diagram of *in vitro* functional assays of neoantigens, HLA-A*24:02 PBMCs were used for IFN-γ ELISPOT and cytotoxic assays. w/wo=with or without.
(B) PBMCs stimulated by neoantigen pools were used for IFN-γ ELISPOT detection; anti-HLA(W6/32) was used to block T-cell recognition;
(C) Number of ELISPOTs was calculated. Student two-tailed test. Data represent mean ± SEM from 3 independent replicates (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). αHLA: anti-HLA antibody(W6/32).
(D) Detection of neoantigen pools stimulated cytotoxic effects on target cells by using bio-luminance assay. Student two-tailed test. Data represent mean ± SEM from 3 independent replicates (*$P < 0.05$, **$P < 0.01$; ns, not significant). fLuc: Firefly luciferase
(E) IncuCyte analysis the inhibitory effects of PBMCs against target cells. Anti-HLA(W6/32) was used to block T-cell recognition. Student two-tailed test. Data represent mean ± SEM from 3 independent replicates (****$P < 0.0001$). αHLA: anti-HLA antibody(W6/32). Red area (%) indicates the live cells percentage.

Fig.5 Screening neoantigens by measuring immune responses and tumor killing effects.
(A) PBMCs induced by No.1-15 peptides were used for IFN-γ ELISPOT detection.
(B) Number of ELISPOTs was calculated. Student two-tailed test. Data represent mean ± SEM from 3 independent replicates (*P < 0.05, **P < 0.01).
(C) Detection of cytotoxic effects on target cells induced by No.1-15 peptides via using bio-luminance assay. Student two-tailed test. Data represent mean ± SEM from 3 independent replicates (*P < 0.05). fLuc: Firefly luciferase
(D) IncuCyte analysis of the inhibitory effects of PBMCs against target cells. Student two-tailed test. Data represent mean ± SEM from 3 independent replicates (****P < 0.0001).; the CT group was only without PBMCs. Peptide No.2 was a negative control. The red area (%) indicated live cells percentage.
(E-G) CD8+T cells are activated by neoantigens *in vitro*. CD8+T cells were stimulated with peptides or control (DMSO) and the efficacy was evaluated by T cell degranulation. CD107a positive percentage and mean fluorescent intensity were summarized. Data represent mean ± SEM from 3 independent replicates (*P < 0.05, **P < 0.01; ns, not significant). PMA-Iono: PMA-Ionomycin; MFI: Mean fluorescence intensity.

Fig.6 Measurement of immune responses and tumor killing effects induced by No.16-19 peptides.
(A) Comparison of the differences in affinity prediction results for peptides between the new algorithm and netMHCpan4.1.
(B) PBMCs induced by No.16-19 peptides were used for IFN-γ ELISPOT assay. PMA-Iono: PMA-Ionomycin;
(C) Number of ELISPOTs was calculated. Student two-tailed test. Data represent mean ± SEM from 3 independent replicates (*P < 0.05, **P < 0.01; ns, not significant).
(D) IncuCyte analysis of the inhibitory effects of PBMCs against target cells. Student two-tailed test. Data represent mean ± SEM from 3 independent replicates (****P < 0.0001).; the CT group was only MIA PaCa-2 without PBMCs. Peptide No.2 was a negative control.
(E-G) CD8+T cells activated by neoantigens *in vitro*. $CD8^{+}$T cells were stimulated with peptides or control (DMSO) and the efficacy was evaluated by T cell degranulation. CD107a positive percentage and mean fluorescent intensity were summarized. MFI: Mean fluorescence intensity. Data represent mean ± SEM from 3 independent replicates (*P < 0.05, **P < 0.01; ns, not significant).

Fig.7 No.19 peptide loaded pMHC triggers T cell activation via the specific TCR interaction
(A) Schematic of scTCR Jurkat-NFAT-GFP activated upon binding with T2 loaded with No.19 peptide.
(B) 10 μg/mL FITC conjugated No.19 Peptide was loaded onto empty T2 cells for 2 hours at 37°C, T2 cells were then rinsed twice and checked via FACS and controlled by T2 cells without any peptide loading.
(C) Jurkat NFAT-GFP were stably transduced with the lentivirus containing scTCR-V5-BBZ. The positive transduced cells were polyclonal sorted by Flow cytometer. The sorted cells were maintained and the scTCR expression was checked before co-culture.

(D) 24 hours post of the co-culture of No.19 peptide loaded T2 cells and the Jurkat NFAT-GFP cells. High-content imaging analysis at 10× magnification was applied to evaluate the NFAT-GFP reporter signals (Left). Besides, the Mean fluorescence intensity (MFI) of GFP (right) was quantified by the imaging software. Data represent mean ± SEM from 3 independent replicates (*P < 0.05).
(E-H) After High-content imaging analysis, the co-culture cells were harvested and stained with viability dye, CD3, and CD69 antibodies. Flow cytometry was applied to compare the NFAT-GFP reporter signals (E-F) as well as the CD69 expression stands for the T cell activation(G-H). Data represent mean ± SEM from 3 independent replicates (**P < 0.01, ***P < 0.001).

Supplementary figure legend

Supplementary Fig.1 A2H and β2M refold to pMHC.
(A) Schematic of pMHC Tetramer produced.
(B)Identify A2H and β2M expression by SDS-PAGE. All protein samples were loaded with reduced buffer. The arrow was directed at the band of A2H and β2M
(C) Purification of pMHC by AKTA. The arrow was directed at the peak of pMHC
(D)Peaks collected from AKTA purification were detected by SDS-page
(E)Purification of Peaks-2 detected by AKTA
(F) Detection of pMHC-tetramer binding to scTCR Jurkat by flow cytometry. pMHC contained a No. 19 peptide.

Supplementary Fig.2 Neoantigens-related genes expression level and ORF-DB establishment.
(A) Heatmap showing the expression levels (log10(TPM+1)) of genes associated with neoantigens.
(B) The frequency of various assembled transcript lengths.
(C) The distribution of different types of assembled transcripts.
(D) The distribution of transcript structures, categorized into 5'uORFs, 3'dORFs, lncRNAs, and CDS, based on their relative abundance in the dataset. ORF: Open-reading frame; CDS: Coding sequence.
Supplementary Fig.3 MIA PaCa-2 HLA typing sequencing result.

Supplementary Fig.4 LC-MS/MS result of neoantigens discovery by MAE method.
(A) Comparison of peptide sequences between MAE and IP. MAE: Mild acid elution, IP: Immunoprecipitation.
(B) False discovery rate (FDR) curve. X axis was the number of peptide-spectrum matches (PSM) being kept. Y axis was the corresponding FDR.
(C) Distribution of precursor mass error of filtered PSM in ppm (Parts per million).
(D) Histogram of peptide ΔRT. RT: Retention time.
(E) Scatterplot of peptide RT versus Predicted RT.

794 **Supplementary Fig.5 Mass-spec of 19 peptides selected from IP (ORF-DB)**
795
796 **Supplementary Fig.6 Gating Strategy of T cell degranulation assay**
797
798 **Supplementary Fig.7 Jurkat NFAT-GFP single cell clones screening**
799
800

Abbreviation	Full name
NHL	Non-Hodgkin lymphoma
MM	Multiple myeloma
TIL	Tumor-infiltrating lymphocyte
TCR	T-cell receptor
DCs	Dendritic cells
HLA	Human leukocyte antigen
IP-MS	Immunoprecipitation followed by mass spectrometry
TAAs	Tumor associated antigens
ADCC	Antibody dependent cellular cytotoxicity
TSAs	Tumor specific antigens
MHC	Major histocompatibility complex
pMHC	The peptide and MHC complex
NGS	Next generation sequencing
WES	Whole-exome sequencing
ATCC	American tissue cell culture
DMEM	Dulbecco's Modified Eagle's medium
DMSO	Dimethyl sulfoxide
FA	Formic acid
RLU	Relative light unit
FDR	False discovery rate
PBMCs	Peripheral blood mononuclear cells
ORF	Open reading frame
CDS	Coding sequence
DDA	Data-dependent acquisition
MAE	Mild acid elution
CAA	Confident amino acid
fLuc	Firefly luciferase
MFI	Mean fluorescence intensity
PMA	Phorbol myristate acetate
SA-PE	Streptavidin-PE
PSM	Peptide-spectrum matches
PPM	Parts per million

RT	Retention Time

801

802 # References

803 ABDOLLAHI, P., NORSETH, H. M. & SCHJESVOLD, F. 2024. Advances and challenges in anti-cancer vaccines for
804 multiple myeloma. *Front Immunol,* 15**,** 1411352.
805 BORNO, H., MOSES, K. A., MONK, P., FITCH, K., HARMON, M., FLANDERS, S. & MCKAY, R. R. 2020. Sipuleucel-T
806 (sip-T) and oral androgen axis inhibitors in black men being treated for metastatic castration-resistance
807 prostate cancer (mCRPC): Observations from the Medicare Fee for Service (FFS) population. *Journal of*
808 *Clinical Oncology,* 38.
809 BUONAGURO, L. & TAGLIAMONTE, M. 2023. Peptide-based vaccine for cancer therapies. *Front Immunol,* 14**,**
810 1210044.
811 CHERVIN, A. S., ET AL., . 2021. *Bispecific binding molecules. WO2021163366A1*.
812 CHERVIN, A. S., STONE, J. D., KONIECZNA, I., CALABRESE, K. M., WANG, N., HARIBHAI, D., DONG, F., WHITE, M. K.,
813 RODRIGUEZ, L. E., BUKOFZER, G. T., ELLIS, P. A., COSGROVE, C., HECQUET, C., CLARIN, J. D., PALMA, J. P.
814 & REILLY, E. B. 2023. ABBV-184: A Novel Survivin-specific TCR/CD3 Bispecific T-cell Engager is Active
815 against Both Solid Tumor and Hematologic Malignancies. *Mol Cancer Ther,* 22**,** 903-912.
816 DHARANI, S., CHO, H., FERNANDEZ, J. P., JUILLERAT, A., VALTON, J., DUCHATEAU, P., POIROT, L. & DAS, S. 2024.
817 TALEN-edited allogeneic inducible dual CAR T cells enable effective targeting of solid tumors while
818 mitigating off-tumor toxicity. *Mol Ther*.
819 DING, Z. Y., LI, Q., ZHANG, R., XIE, L., SHU, Y., GAO, S., WANG, P. P., SU, X. Q., QIN, Y., WANG, Y. L., FANG, J. M.,
820 ZHU, Z. Z., XIA, X. Y., WEI, G. C., WANG, H., QIAN, H., GUO, X. L., GAO, Z. B., WANG, Y., WEI, Y. Q., XU, Q.,
821 XU, H. & YANG, L. 2021. Personalized neoantigen pulsed dendritic cell vaccine for advanced lung cancer.
822 *Signal Transduction and Targeted Therapy,* 6.
823 FORD, M. J., JONES, R., ALLEN, D., AMUNUGAMA, R., PISANO, M., MOBLEY, J., DOMANSKI, P., HO, B., BOCHAR, D.
824 & HOLT, M. 2018. Mass spectrometry as a tool for MHC class I and II neoantigen discovery. *Cancer*
825 *Research,* 78.
826 GUAN, J., ZHOU, S. X., WU, W. W., ZHU, T. Y., CAO, L., WU, M., LI, N. H., FU, F. G., LIAO, Z. G., LIN, S. M., LIU, Z. Y.,
827 ZHU, M. J., LUO, N. N., ZHANG, Y., CHIA, T. S., CHEN, B. L. & HE, K. J. 2024. IBI334, a novel ADCC-enhanced
828 B7-H3/EGFR bispecific antibody, demonstrated potent pre-clinical efficacy in solid tumors. *Cancer*
829 *Research,* 84.
830 HUANG, D., ZHU, X., YE, S., ZHANG, J., LIAO, J., ZHANG, N., ZENG, X., WANG, J., YANG, B., ZHANG, Y., LAO, L., CHEN,
831 J., XIN, M., NIE, Y., SAW, P. E., SU, S. & SONG, E. 2024. Tumour circular RNAs elicit anti-tumour immunity
832 by encoding cryptic peptides. *Nature,* 625**,** 593-602.
833 JANELLE, V., RULLEAU, C., DEL TESTA, S., CARLI, C. & DELISLE, J. S. 2020. T-Cell Immunotherapies Targeting
834 Histocompatibility and Tumor Antigens in Hematological Malignancies. *Front Immunol,* 11**,** 276.
835 KARASAKI, T., NAGAYAMA, K., KUWANO, H., NITADORI, J. I., SATO, M., ANRAKU, M., HOSOI, A., MATSUSHITA, H.,
836 TAKAZAWA, M., OHARA, O., NAKAJIMA, J. & KAKIMI, K. 2017. Prediction and prioritization of
837 neoantigens: integration of RNA sequencing data with whole-exome sequencing. *Cancer Sci,* 108**,** 170-
838 177.
839 KOPETZ, S., MORRIS, V. K., ALONSO-ORDUNA, V., GARCIA-ALFONSO, P., REBOREDO, M., MONTES, A., PAEZ, J. D.,
840 REINACHER-SCHICK, A. C., HÖHLER, T., CARRASCO, J., GALLIGAN, B. M., MANNING, L., PREUSSNER, L.,
841 TURECI, O. & SAHIN, U. 2022. A phase 2 multicenter, open-label, randomized, controlled trial in patients
842 with stage II/III colorectal cancer who are ctDNA positive following resection to compare efficacy of

843 autogene cevumeran versus watchful waiting. *Journal of Clinical Oncology,* 40.
844 LANG, F., SCHRÖRS, B., LÖWER, M., TÜRECI, Ö. & SAHIN, U. 2022. Identification of neoantigens for individualized
845 therapeutic cancer vaccines. *Nature Reviews Drug Discovery,* 21**,** 261-282.
846 LESTERHUIS, W. J., SCHREIBELT, G., SCHARENBORG, N. M., BROUWER, H. M., GERRITSEN, M. J., CROOCKEWIT, S.,
847 COULIE, P. G., TORENSMA, R., ADEMA, G. J., FIGDOR, C. G., DE VRIES, I. J. & PUNT, C. J. 2011. Wild-type
848 and modified gp100 peptide-pulsed dendritic cell vaccination of advanced melanoma patients can lead
849 to long-term clinical responses independent of the peptide used. *Cancer Immunol Immunother,* 60**,** 249-
850 60.
851 MATSUSHITA, M., OTSUKA, Y., TSUTSUMIDA, N., TANAKA, C., UCHIUMI, A., OZAWA, K., SUZUKI, T., ICHIKAWA, D.,
852 ABURATANI, H., OKAMOTO, S., KAWAKAMI, Y. & HATTORI, Y. 2016. Identification of Novel HLA-A*24:02-
853 Restricted Epitope Derived from a Homeobox Protein Expressed in Hematological Malignancies. *PLoS*
854 *One,* 11**,** e0146371.
855 MELLMAN, I., CHEN, D. S., POWLES, T. & TURLEY, S. J. 2023. The cancer-immunity cycle: Indication, genotype, and
856 immunotype. *Immunity,* 56**,** 2188-2205.
857 MIAO, L. L., ZHANG, J., XU, W., QIAN, Q., ZHANG, G. C., YUAN, Q., LV, Y. T., ZHANG, H. G., SHEN, C. Y. & WANG, W.
858 2024. Global research trends in CAR-T cell therapy for solid tumors: A comprehensive visualization and
859 bibliometric study (2012-2023). *Human Vaccines & Immunotherapeutics,* 20.
860 NGUYEN, A., SANBORN, J. Z., VASKE, C. J., RABIZADEH, S., NIAZI, K., SOON-SHIONG, P. & BENZ, S. C. 2016.
861 Identifying patient-specific neoepitopes for cell-based and vaccine immunotherapy targets in breast
862 cancer patients by HLA typing and predicting MHC presentation from whole genome and RNA
863 sequencing data. *Journal of Clinical Oncology,* 34.
864 PAK, H., MICHAUX, J., HUBER, F., CHONG, C., STEVENSON, B. J., MULLER, M., COUKOS, G. & BASSANI-STERNBERG,
865 M. 2021. Sensitive Immunopeptidomics by Leveraging Available Large-Scale Multi-HLA Spectral Libraries,
866 Data-Independent Acquisition, and MS/MS Prediction. *Mol Cell Proteomics,* 20**,** 100080.
867 PENG, M., MO, Y. Z., WANG, Y., WU, P., ZHANG, Y. J., XIONG, F., GUO, C., WU, X., LI, Y., LI, X. L., LI, G. Y., XIONG, W.
868 & ZENG, Z. Y. 2019. Neoantigen vaccine: an emerging tumor immunotherapy. *Molecular Cancer,* 18.
869 PETERLIN, P., SAADA-BOUZID, E., MOSKOVITZ, M., PIGNEUX, A., YUDA, J., SINNOLLAREDDY, M., HENNER, W. R.,
870 CHEN, D., FREISE, K. J., LEIBMAN, R. S., AVIGDOR, A. & SHIMIZU, T. 2024. First-in-human clinical trial
871 results with ABBV-184, a first-in-class T-cell receptor/anti-CD3 bispecific protein, in adults with
872 previously treated AML or NSCLC. *Expert Rev Anticancer Ther,* 24**,** 893-904.
873 ROJAS, L. A., SETHNA, Z., SOARES, K. C., OLCESE, C., PANG, N., PATTERSON, E., LIHM, J., CEGLIA, N., GUASP, P.,
874 CHU, A., YU, R., CHANDRA, A. K., WATERS, T., RUAN, J., AMISAKI, M., ZEBBOUDJ, A., ODGEREL, Z., PAYNE,
875 G., DERHOVANESSIAN, E., MULLER, F., RHEE, I., YADAV, M., DOBRIN, A., SADELAIN, M., LUKSZA, M.,
876 COHEN, N., TANG, L., BASTURK, O., GONEN, M., KATZ, S., DO, R. K., EPSTEIN, A. S., MOMTAZ, P., PARK,
877 W., SUGARMAN, R., VARGHESE, A. M., WON, E., DESAI, A., WEI, A. C., D'ANGELICA, M. I., KINGHAM, T.
878 P., MELLMAN, I., MERGHOUB, T., WOLCHOK, J. D., SAHIN, U., TURECI, O., GREENBAUM, B. D., JARNAGIN,
879 W. R., DREBIN, J., O'REILLY, E. M. & BALACHANDRAN, V. P. 2023. Personalized RNA neoantigen vaccines
880 stimulate T cells in pancreatic cancer. *Nature,* 618**,** 144-150.
881 ROSENBERG, S. A., YANG, J. C., SCHWARTZENTRUBER, D. J., HWU, P., MARINCOLA, F. M., TOPALIAN, S. L., RESTIFO,
882 N. P., DUDLEY, M. E., SCHWARZ, S. L., SPIESS, P. J., WUNDERLICH, J. R., PARKHURST, M. R., KAWAKAMI,
883 Y., SEIPP, C. A., EINHORN, J. H. & WHITE, D. E. 1998. Immunologic and therapeutic evaluation of a
884 synthetic peptide vaccine for the treatment of patients with metastatic melanoma. *Nat Med,* 4**,** 321-7.
885 SAHIN, U., DERHOVANESSIAN, E., MILLER, M., KLOKE, B. P., SIMON, P., LÖWER, M., BUKUR, V., TADMOR, A. D.,
886 LUXEMBURGER, U., SCHRÖRS, B., OMOKOKO, T., VORMEHR, M., ALBRECHT, C., PARUZYNSKI, A., KUHN,

887 A. N., BUCK, J., HEESCH, S., KATHARINA, H., MÜLLER, F., ORTSEIFER, I., VOGLER, I., GODEHARDT, E., ATTIG,
888 S., RAE, R., BREITKREUZ, A., TOLLIVER, C., SUCHAN, M., MARTIC, G., HOHBERGER, A., SORN, P.,
889 DIEKMANN, J., CIESLA, J., WAKSMANN, O., BÜRCK, A. K., WITT, M., ZILLGEN, M., ROTHERMEL, A.,
890 KASEMANN, B., LANGER, D., BOLTE, S., DIKEN, M., KREITER, S., NEMECEK, R., GEBHARDT, C., GRABBE,
891 S., HÖLLER, C., UTIKAL, J., HUBER, C., LOQUAI, C. & TÜRECI, Ö. 2017. Personalized RNA mutanome
892 vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature,* 547**,** 222-+.
893 SARTOR, O., ARMSTRONG, A. J., AHAGHOTU, C., MCLEOD, D. G., COOPERBERG, M. R., PENSON, D. F., KANTOFF,
894 P. W., VOGELZANG, N. J., HUSSAIN, A., PIECZONKA, C. M., SHORE, N. D., QUINN, D. I., HEATH, E. I.,
895 TUTRONE, R. F., SCHELLHAMMER, P. F., HARMON, M., CHANG, N. N., FREEDLAND, S. J. & HIGANO, C. S.
896 2020. Overall survival (OS) of African-American (AA) and Caucasian (CAU) men who received sipuleucel-
897 T for metastatic castration-resistant prostate cancer (mCRPC)-final PROCEED analysis. *Cancer*
898 *Epidemiology Biomarkers & Prevention,* 29.
899 SHARPNACK, M. F., JOHNSON, T. S., CHALKLEY, R., HAN, Z., CARBONE, D., HUANG, K. & HE, K. 2022. TSAFinder:
900 exhaustive tumor-specific antigen detection with RNAseq. *Bioinformatics,* 38**,** 2422-2427.
901 SHEN, W. X., LIU, Y., CHEN, Y., ZENG, X., TAN, Y., JIANG, Y. Y. & CHEN, YU Z. 2022. AggMapNet: enhanced and
902 explainable low-sample omics deep learning with feature-aggregated multi-channel networks. *Nucleic*
903 *Acids Research,* 50**,** e45-e45.
904 SHEN, W. X., ZENG, X., ZHU, F., WANG, Y. L., QIN, C., TAN, Y., JIANG, Y. Y. & CHEN, Y. Z. 2021. Out-of-the-box deep
905 learning prediction of pharmaceutical properties by broadly learned knowledge-based molecular
906 representations. *Nature Machine Intelligence,* 3**,** 334-343.
907 SMITH, S. N., HARRIS, D. T. & KRANZ, D. M. 2015. T Cell Receptor Engineering and Analysis Using the Yeast Display
908 Platform. *Yeast Surface Display.*
909 STURM, T., SAUTTER, B., WORNER, T. P., STEVANOVIC, S., RAMMENSEE, H. G., PLANZ, O., HECK, A. J. R. &
910 AEBERSOLD, R. 2021. Mild Acid Elution and MHC Immunoaffinity Chromatography Reveal Similar Albeit
911 Not Identical Profiles of the HLA Class I Immunopeptidome. *J Proteome Res,* 20**,** 289-304.
912 TRAN, E., ROBBINS, P. F., LU, Y. C., PRICKETT, T. D., GARTNER, J. J., JIA, L., PASETTO, A., ZHENG, Z. L., RAY, S., GROH,
913 E. M., KRILEY, I. R. & ROSENBERG, S. A. 2016. T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer.
914 *New England Journal of Medicine,* 375**,** 2255-2262.
915 TRAN, E., TURCOTTE, S., GROS, A., ROBBINS, P. F., LU, Y. C., DUDLEY, M. E., WUNDERLICH, J. R., SOMERVILLE, R.
916 P., HOGAN, K., HINRICHS, C. S., PARKHURST, M. R., YANG, J. C. & ROSENBERG, S. A. 2014. Cancer
917 Immunotherapy Based on Mutation-Specific CD4+T Cells in a Patient with Epithelial Cancer. *Science,* 344**,**
918 641-645.
919 VERA, D. G., WAGHELA, H., NUH, M., PAN, J. & LULLA, P. 2024. Approved CAR-T therapies have reproducible
920 efficacy and safety in clinical practice. *Human Vaccines & Immunotherapeutics,* 20.
921 WEBER, J. S., CARLINO, M. S., KHATTAK, A., MENIAWY, T., ANSSTAS, G., TAYLOR, M. H., KIM, K. B., MCKEAN, M.,
922 LONG, G., SULLIVAN, R. J., FARIES, M., TRAN, T. T., COWEY, C. L., PECORA, A., SHAHEEN, M., SEGAR, J.,
923 MEDINA, T., ATKINSON, V., GIBNEY, G. T., LUKE, J. J., THOMAS, S., BUCHBINDER, E., HEALY, J. A., HUANG,
924 M., MORRISSEY, M., FELDMAN, I., SEHGAL, V., ROBERT-TISSOT, C., HOU, P. J., ZHU, L. L., BROWN, M.,
925 AANUR, P., MEEHAN, R. S. & ZAKS, T. 2024. Individualised neoantigen therapy mRNA-4157 (V940) plus
926 pembrolizumab versus pembrolizumab monotherapy in resected melanoma (KEYNOTE-942): a
927 randomised, phase 2b study. *Lancet,* 403**,** 632-644.
928 YAO, T., ZHANG, Z., LI, Q., HUANG, R., HONG, Y., LI, C., ZHANG, F., HUANG, Y., FANG, Y., CAO, Q., JIN, X., LI, C.,
929 WANG, Z., LIN, X. J., LI, L., WEI, W., WANG, Z. & SHEN, J. 2023. Long-Read Sequencing Reveals Alternative
930 Splicing-Driven, Shared Immunogenic Neoepitopes Regardless of SF3B1 Status in Uveal Melanoma.

931 *Cancer Immunol Res,* 11**,** 1671-1687.

932

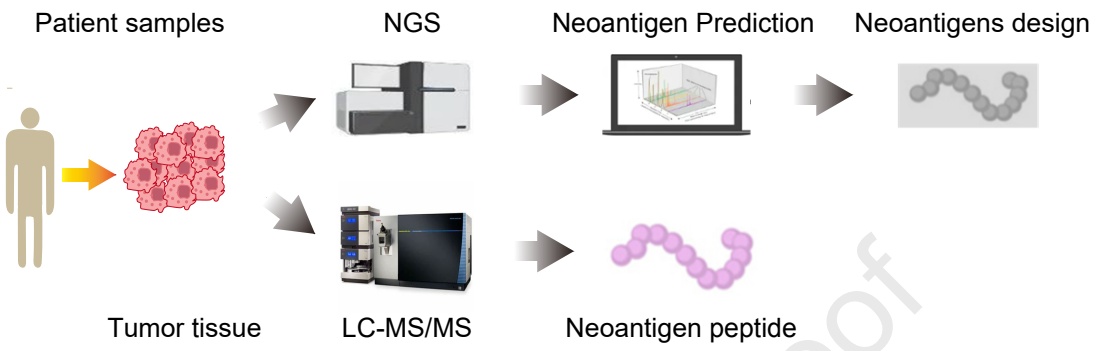
Table 1

Table.1 Affinity prediction analysis of netMHCpan4.1 and New algorithm

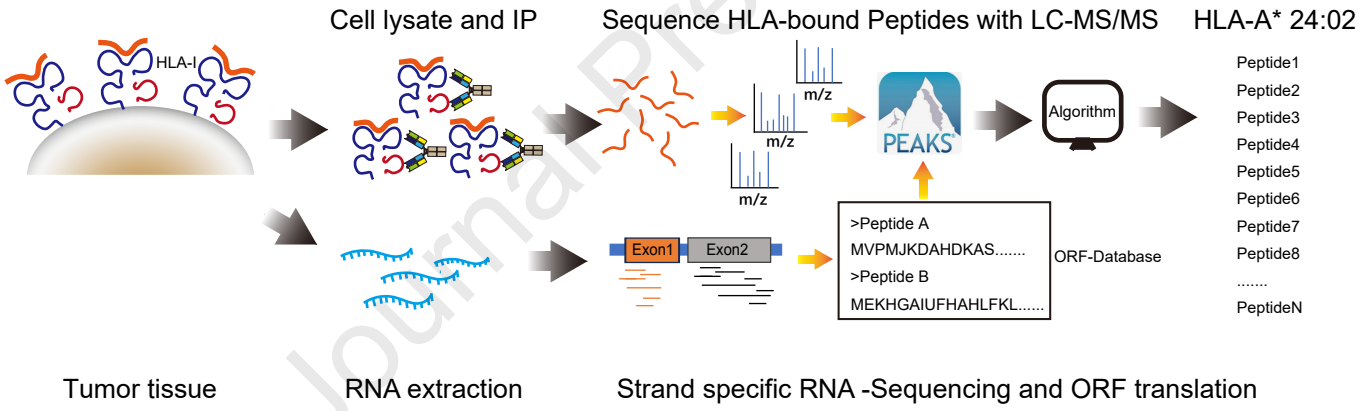
Peptide No.	Peptide	CAA (%)	Length	HLA	netMHCpan4.1 Predicted affinity (nM)	Bind Level	New algorithm Predicted affinity (nM)	Bind Level
1	EYPDRIMNTF	50	10	A*24:02	106.24	SB	157.59	SB
2	FEGFPDKQPR	80	10	A*24:02	42011.50	NB	33321.40	NB
3	LYADVGGKQF	100	10	A*24:02	241.58	SB	315.21	SB
4	RYFDPANGKF	80	10	A*24:02	52.65	SB	307.32	SB
5	YDESGPSIVH	70	10	A*24:02	45240.08	NB	31383.19	NB
6	AYVHMVTHF	44.4	9	A*24:02	24.67	SB	37.85	SB
7	EYNSDLHQF	77.8	9	A*24:02	119.28	SB	331.30	SB
8	KFIDTTSKF	100	9	A*24:02	70.39	SB	1263.09	WB
9	KYISGPHEL	88.9	9	A*24:02	51.21	SB	489.38	SB
10	RYIDTHNRV	66.7	9	A*24:02	102.93	SB	1985.71	WB
11	TYGEIFEKF	77.8	9	A*24:02	6.46	SB	84.16	SB
12	VYIKHPVSL	100	9	A*24:02	45.15	SB	107.29	SB
13	VYISEHEHF	77.8	9	A*24:02	10.33	SB	102.06	SB
14	VYPDGIRHI	55.6	9	A*24:02	131.16	SB	391.65	SB
15	KYTPPPHHI	66.7	9	A*24:02	87.68	SB	865.07	SB
16	KYTPDAMLH	77.8	9	A*24:02	20306.50	NB	3011.98	WB
17	SGPERILSI	66.7	9	A*24:02	7074.60	NB	4970.61	WB
18	SFVDTRTLL	66.7	9	A*24:02	3106.86	WB	14600.04	NB
19	LTLGEFLKL	88.9	9	A*24:02	7160.54	NB	5263.05	WB

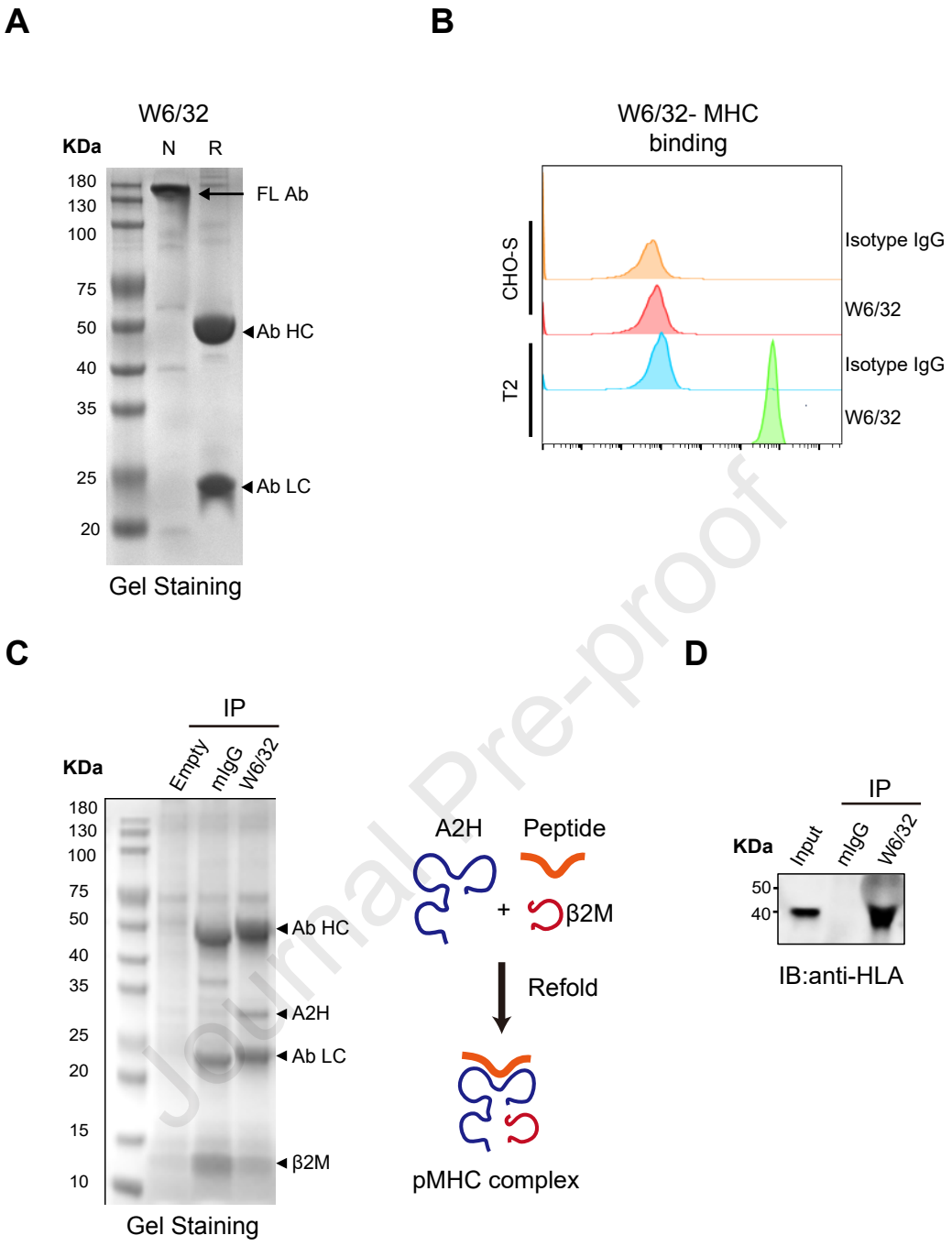
WB=weak binding; SB-strong binding; NB=no binding

A



B





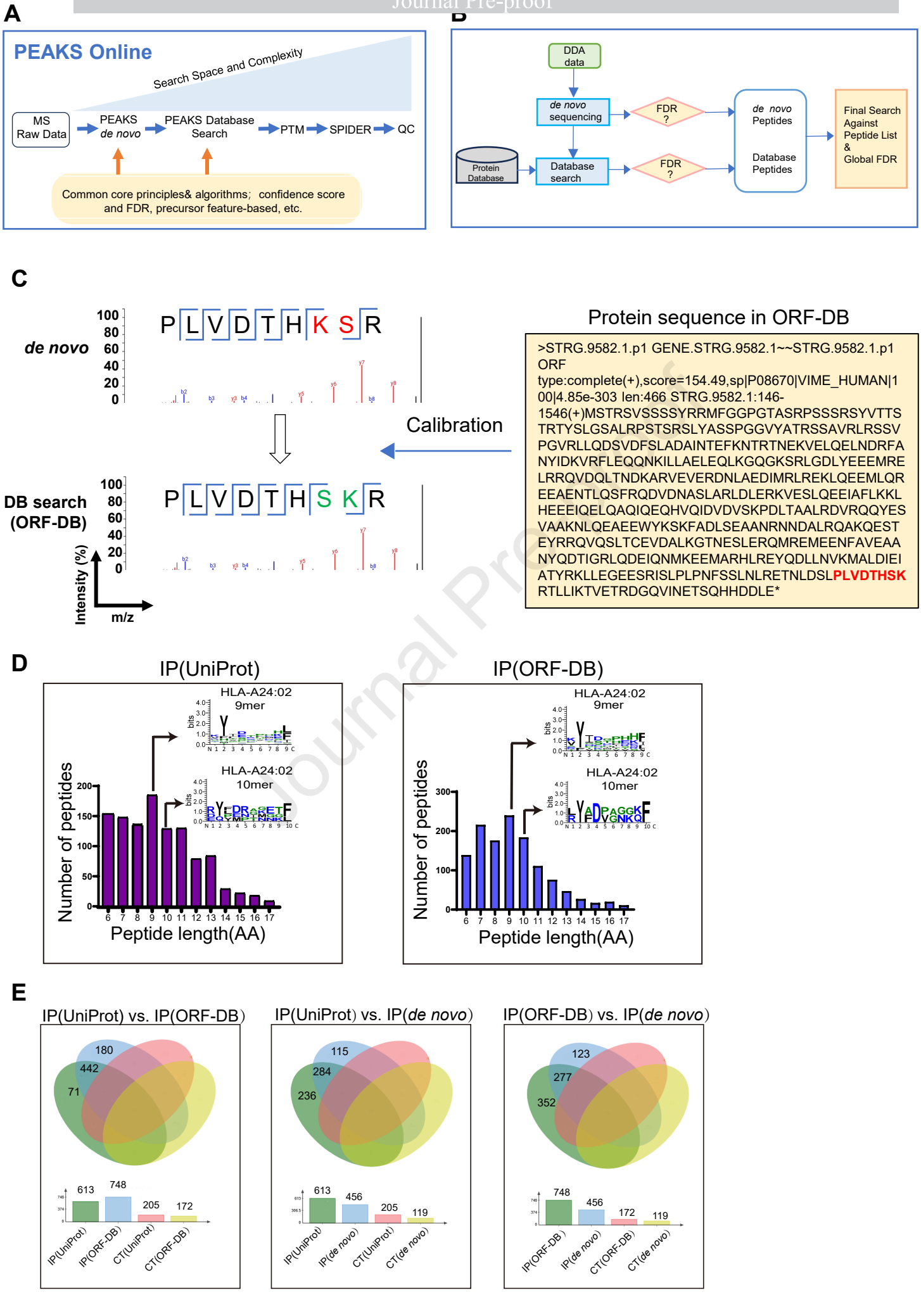


Fig. 4

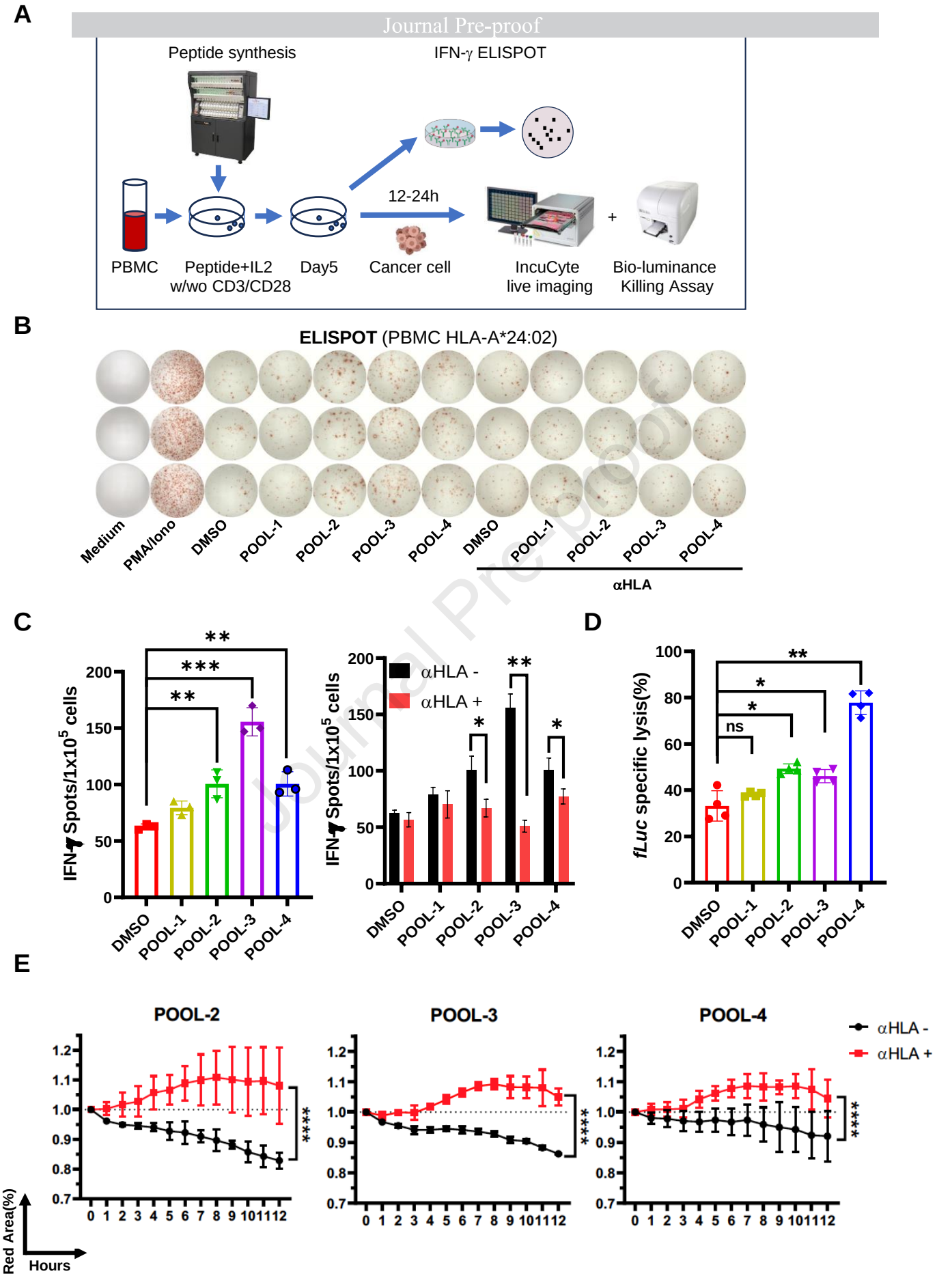


Fig. 5

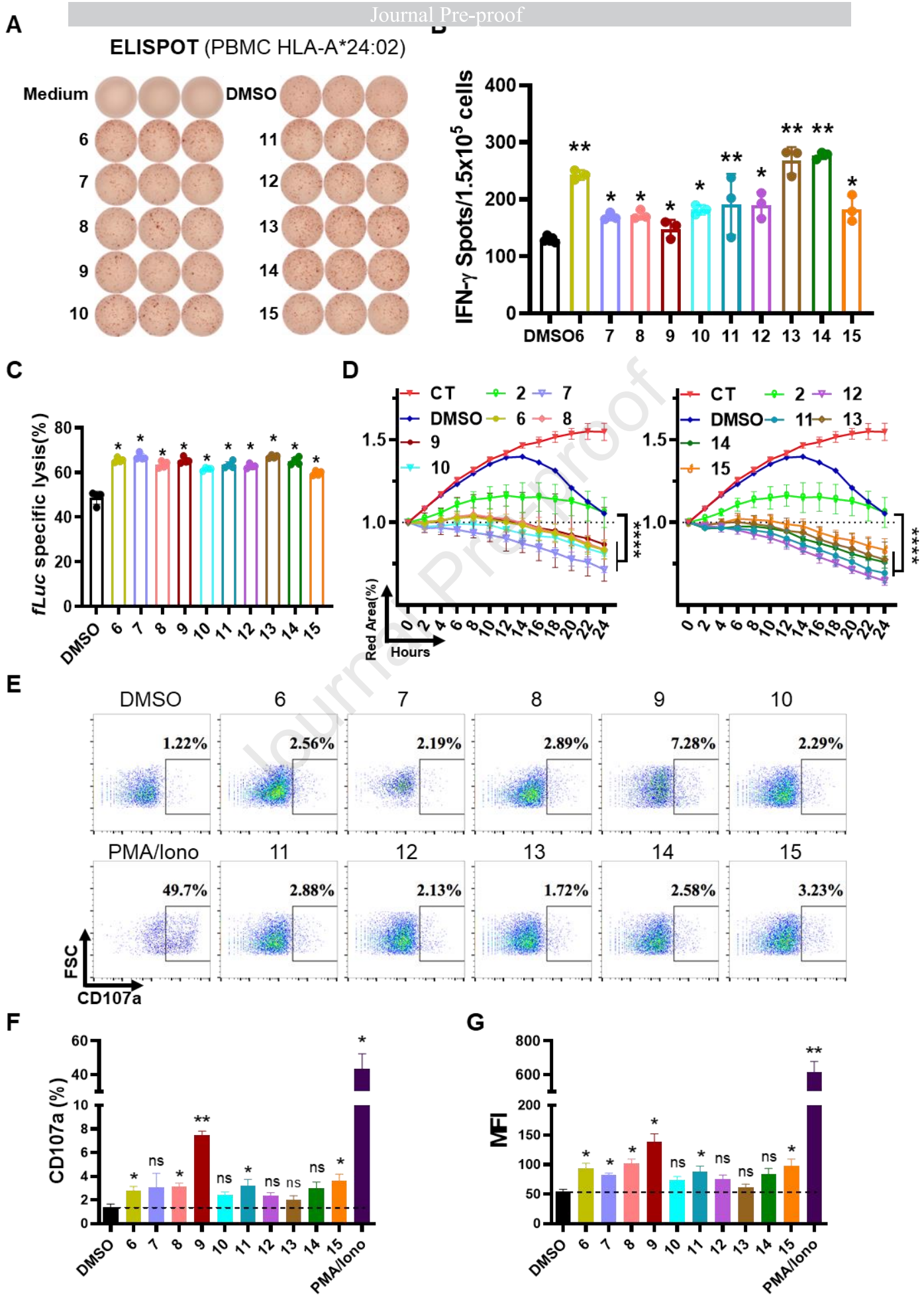
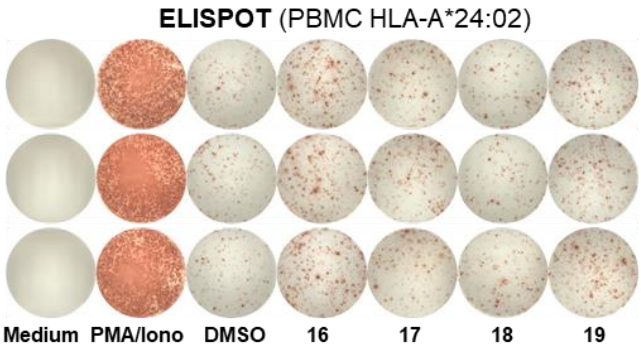


Fig. 6

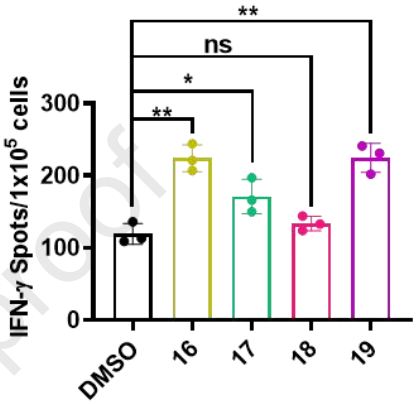
A

Peptide No.	Peptide	CAA (%)	Length	HLA	netMHCpan4.1 Predicted affinity (nM)	Bind Level	New algorithm Predicted affinity (nM)	Bind Level	*In vitro* Efficacy
16	KYTPDAMLH	77.8	9	A*24:02	20306.50	NB	3011.98	WB	Yes
17	SGPERILSI	66.7	9	A*24:02	7074.60	NB	4970.61	WB	Yes
18	SFVDTRTLL	66.7	9	A*24:02	3106.86	WB	14600.04	NB	No
19	LTLGEFLKL	88.9	9	A*24:02	7160.54	NB	5263.06	WB	Yes

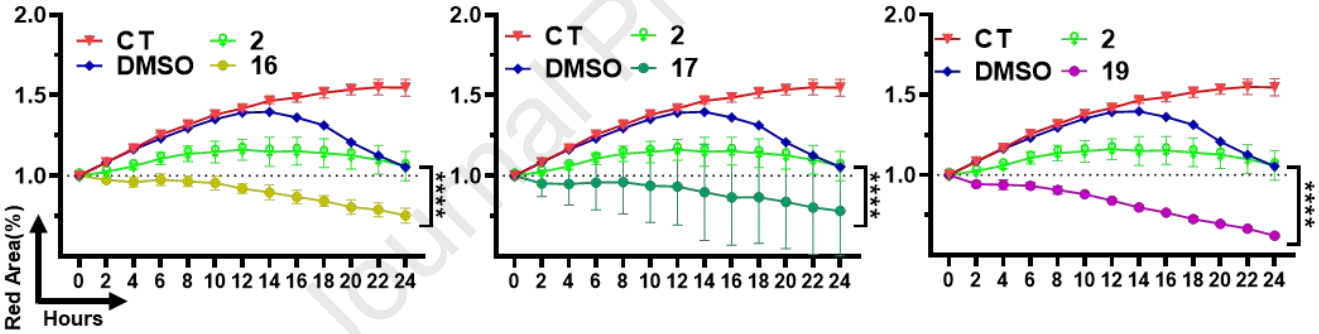
B



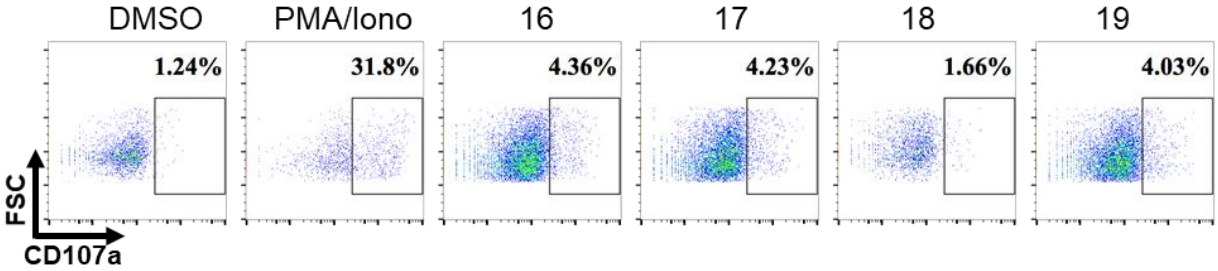
C



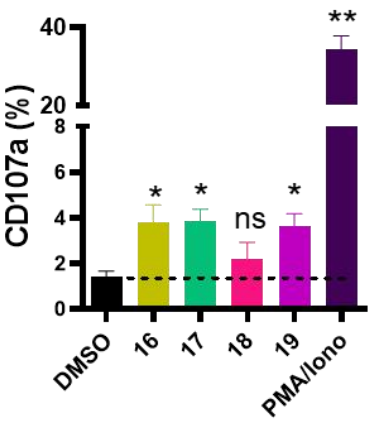
D



E



F



G

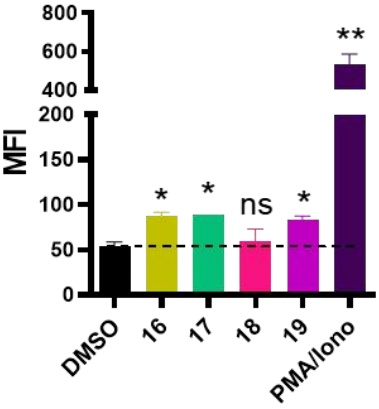
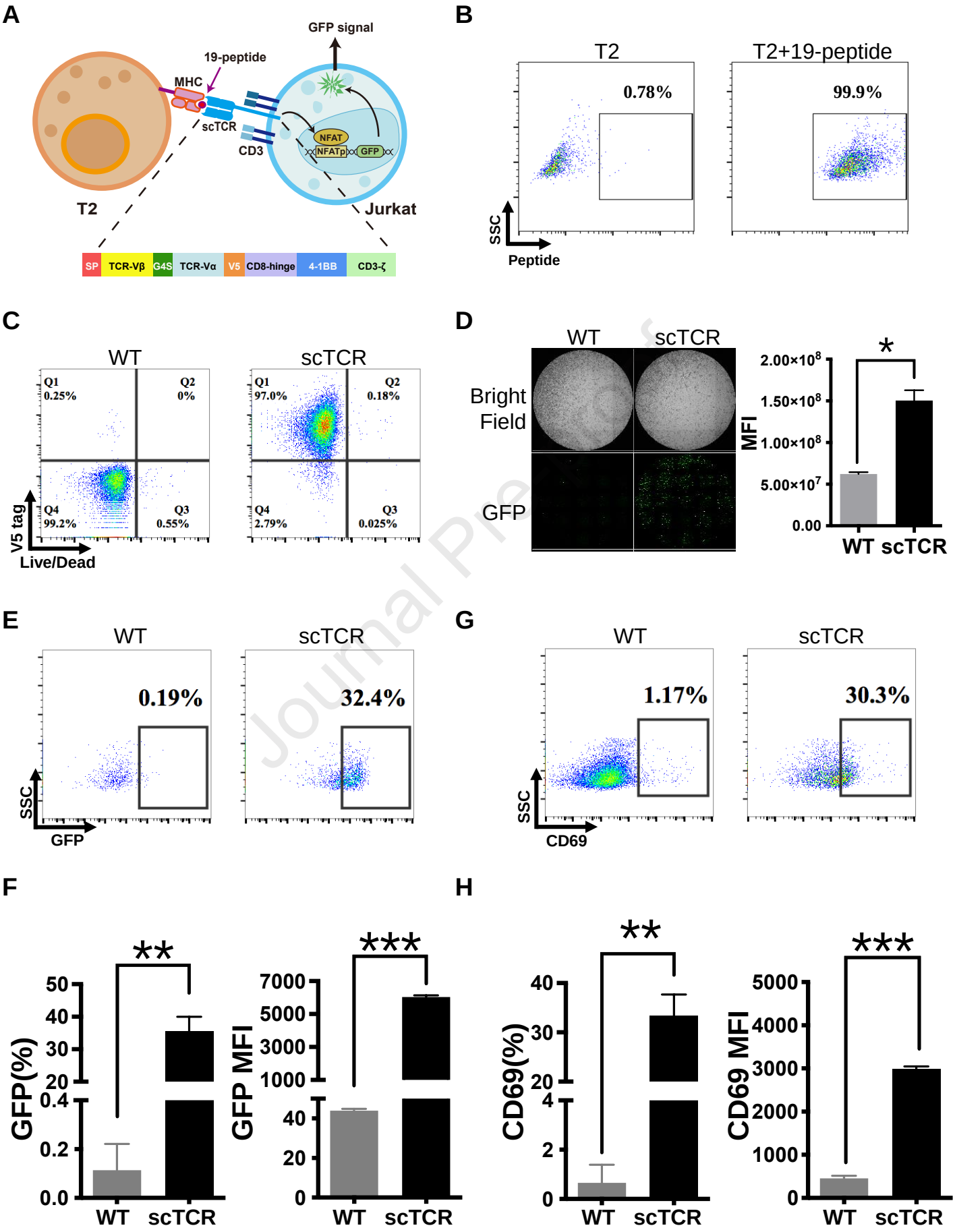


Fig. 7



- IP-MS directly identifies *bona fide* personalized neoantigen peptides

- RNA-seq calibrates the MS data to avoid errors during amino acid de novo sequencing

- A new algorithm prediction is better than netMHCpan4.1

- Rapid in vitro validation ensures neoantigen candidates are functional for tumor vaccine development

Conflict of Interest Statement

Manuscript Title: **Rapid and direct discovery of functional tumor specific neoantigens by high resolution mass spectrometry and novel algorithm prediction**

The authors declare that there are no conflicts of interest related to this manuscript.

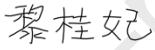
Xi Kang
kangxi@szbl.ac.cn; kangxi@whu.edu.cn

11.06.2024

Huajian Tian
tianhj@szbl.ac.cn; 13250710561@163.com

11.06.2024

Guifei Li
ligf@szbl.ac.cn; lgf2012@foxmail.com



11.06.2024