

OPEN

Streamlined selection of cancer antigens for vaccine development through integrative multi-omics and high-content cell imaging

Ki-Cheol Han^{1,8*}, Daechan Park^{2,8}, Shinyeong Ju¹, Young Eun Lee^{1,3}, Sun-Hee Heo^{4,5}, Young-Ae Kim⁴, Ji Eun Lee¹, Yuna Lee³, Kyong Hwa Park⁶, Se-Ho Park³, Hee Jin Lee^{4,5}, Cheolju Lee^{1,7*} & Mihue Jang^{1,7*}

Identification of tumor antigens that induce cytotoxic T lymphocytes (CTLs) is crucial for cancer-vaccine development. Despite their predictive ability, current algorithmic approaches and human leukocyte antigen (HLA)-peptidomic analysis allow limited selectivity. Here, we optimized a method to rapidly screen and identify highly immunogenic epitopes that trigger CTL responses. We used a combined application of this method involving immune-specific signature analysis and HLA-associated peptidomics using samples from six patients with triple-negative breast cancer (TNBC) in order to select immunogenic HLA epitopes for *in vitro* testing. Additionally, we applied high-throughput imaging at the single-cell level in order to confirm the immunoreactivity of the selected peptides. The results indicated that this method enabled identification of promising CTL peptides capable of inducing antitumor immunity. This platform combining high-resolution computational analysis, HLA-peptidomics, and high-throughput immunogenicity testing allowed rapid and robust identification of highly immunogenic epitopes and represents a powerful technique for cancer-vaccine development.

Cancer immunotherapy to boost T cell-mediated immune response in order to target and eliminate cancer cells has proven therapeutically efficacious in a variety of human malignancies¹. In particular, therapeutic cancer vaccines against tumor-related epitopes that directly stimulate T cells have been clinically effective and are currently available². Cancer cells express several antigens, including self-antigens derived from tumor tissues, as well as mutation-derived antigens (i.e., neoantigens), that can be recognized as foreign antigens by the host immune system^{3–5}. Recently, cancer vaccines targeting individual neoantigens have become an attractive form of cancer therapeutics by virtue of their ability to elicit a robust T cell immune response^{6–8}. Three independent clinical studies demonstrated the efficacy of this approach in eliciting neoantigen-specific T cell responses in patients with late-stage melanoma^{9–11}. Nevertheless, several challenges to personalized cancer treatment targeting distinct neoantigens remain as obstacles to wide clinical application of these vaccines. First, the prediction of highly immunogenic neoantigens is difficult. Recent advances in next-generation sequencing technology and MHC-epitope databases allow the accurate prediction of neoantigens; however, existing tools are inadequate for accurate prediction of immunogenicity due to the multiple associated factors, such as proteasomal-processing ability, intracellular routing, specific binding to highly polymorphic HLA molecules, and peptide/MHC (p/MHC)-complex

¹Center for Theragnosis, Biomedical Research Institute, Korea Institute of Science and Technology, Seongbuk-Gu, Seoul, 02792, Republic of Korea. ²Department of Biological Sciences, College of Natural Sciences, Ajou University, Suwon, 16499, Republic of Korea. ³College of Life Science and Biotechnology, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Republic of Korea. ⁴Department of Pathology, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul, 05505, Republic of Korea. ⁵Asan Center for Cancer Genome Discovery, Asan Institute for Life Sciences, University of Ulsan College of Medicine, Asan Medical Center, Seoul, 05505, Republic of Korea. ⁶Oncology/Hematology, Department of Internal medicine, Korea University College of medicine, Seongbuk-Gu, Seoul, 02841, Republic of Korea. ⁷Department of Converging Science and Technology, KHU-KIST, Kyung Hee University, Seoul, 02447, Republic of Korea. ⁸These authors contributed equally: Ki-Cheol Han and Daechan Park. *email: biolord@kist.re.kr; clee270@kist.re.kr; mihue@kist.re.kr

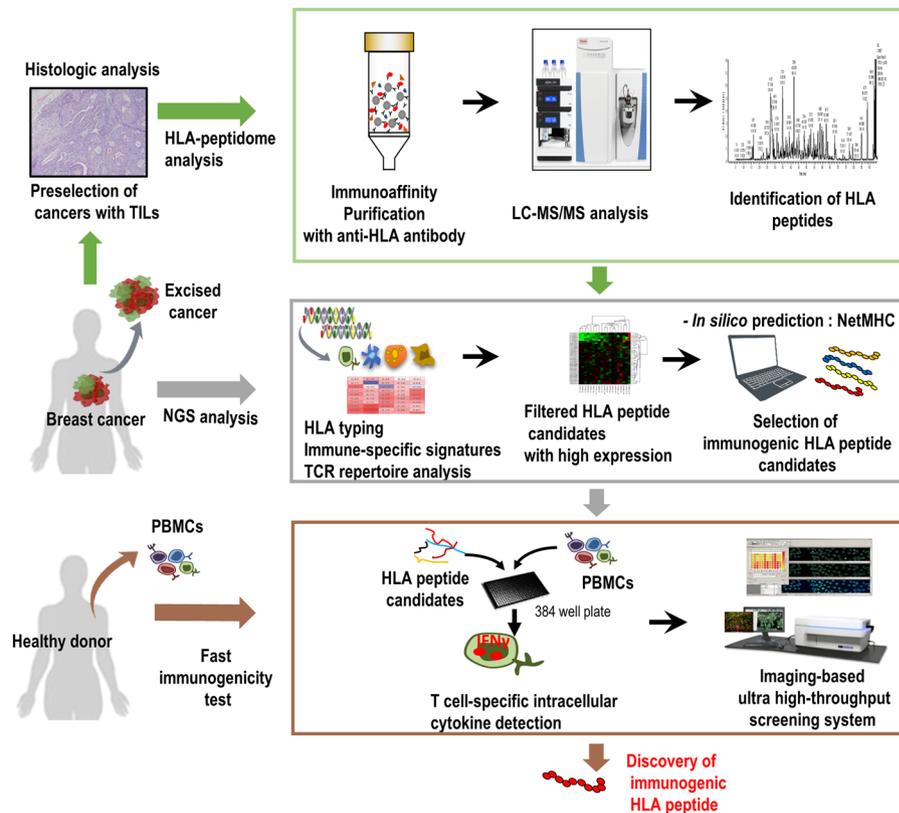


Figure 1. Scheme of the rapid high-throughput approach for discovering natural CTL epitopes. Preselected TIL-resident TNBC tumors underwent HLA-peptidomic analysis to identify HLA-bound peptides. Integrated WTS data revealed a higher priority to select promising HLA-peptides *via* high-resolution bioinformatics analysis, showing immune-cell-specific signatures and TCR-repertoire diversity in tumors. Combined NGS analysis and the use of predictive algorithms for MHC-binding affinity enabled selection of highly immunogenic HLA-peptide candidates. Analysis of IFN γ -producing CD8 $^{+}$ T cell response using the high-content imaging system in a 384-well format at the single-cell level for discovery of immunogenic HLA epitopes eliciting a CTL response.

stability^{6,12}. Therefore, it is necessary to optimize the predictive accuracy of MHC-binding affinity and selection of immunogenic neoantigens in order to improve currently available algorithms and bioinformatics tools¹³.

Traditional cancer vaccines target different types of self-antigens, including overexpressed tumor-specific proteins and oncogenes that might also represent universal antigens among patients with the same malignancy^{14–16}. Self-antigens derived from tumor-related proteins are grouped into three categories^{17,18}: tumor-associated antigens (TAAs), tumor-specific antigens (TSAs), and cancer/testis antigens (CTAs). TAAs are typically overexpressed in tumor tissues relative to their levels in other tissues, whereas TSAs are exclusively expressed in cancer cells, and CTAs are expressed only in germline tissues and trophoblastic cells in addition to being expressed in cancer cells. Using genetic and immunological approaches, several self-antigens have been identified as adequate targets for cancer immunotherapy¹⁶; however, vaccination against self-antigens can still potentially induce central- and peripheral-tolerance responses resulting in low therapeutic efficacy^{5,19} or an autoimmune response against normal tissues⁶. Therefore, identification of appropriate self-antigens is important for successful cancer-vaccine immunotherapy.

Cancer vaccines targeting cancer antigens have been developed using multiple approaches, including peptides, proteins, nucleic acids, and viral vectors²⁰. Among these strategies, peptides used in cancer vaccines are 20–30 amino acids, showing lower antigen complexity and low costs for manufacturing²¹. In particular, MHC class-I-restricted peptides (i.e., CTL epitopes) recognized by CD8 $^{+}$ CTLs play a key role in attacking cancer cells by promoting the activation and proliferation of antigen-specific CTLs²². To develop effective CTL-associated peptide vaccines, many methods use bioinformatics algorithms and proteomics approaches to identify multiple HLA-class-bound peptides^{23,24}. Recently, proteomics approaches associated with mass spectrometry (MS)²⁵ and specifically LC-MS/MS, were used for HLA-peptidomic analysis to directly facilitate the discovery of numerous natural peptides in complex with specific HLA molecules expressed on cell surfaces^{26,27}. Additionally, recent advances in MS-based technologies led to the development of a robust approach to quantify the dynamics of epitope presentation^{28,29}. However, the high numbers of HLA-restricted peptides available cannot be comprehensively tested to facilitate effective vaccine development. Therefore, we established a rapid and robust screening system combining immune-signature investigation, HLA-peptidomic analysis, and high-throughput immunogenicity testing to evaluate antigen-specific CTL responses (Fig. 1).

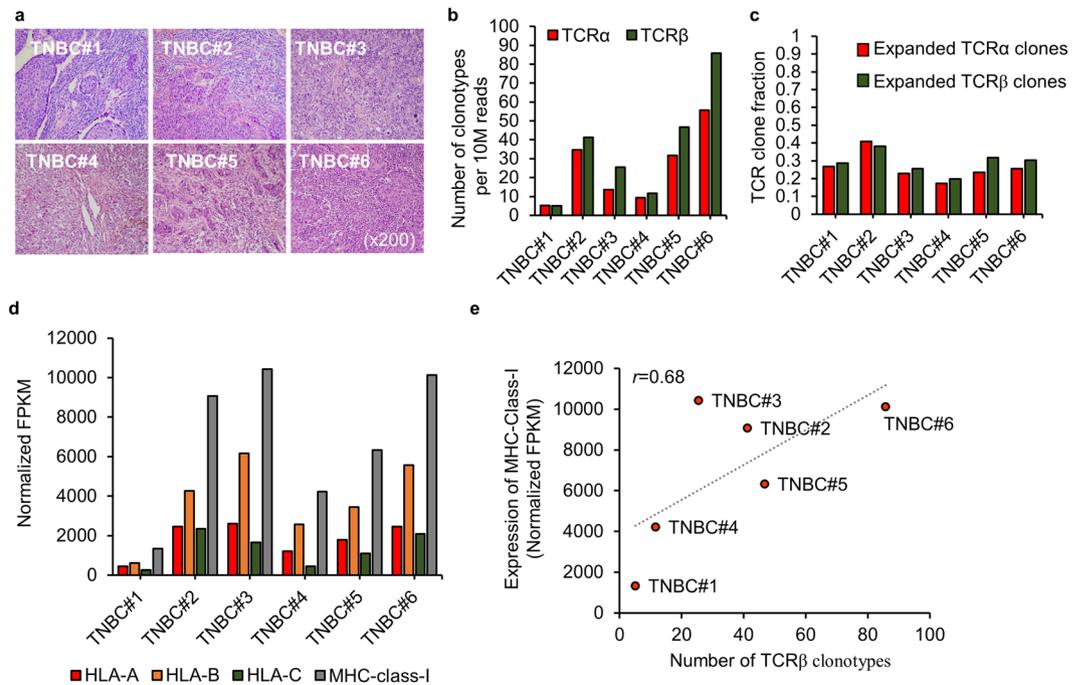


Figure 2. Positive linear correlation between TCR diversity and MHC gene expression. **(a)** Histologic analysis showing TIL-resident tumors from six patients with TNBC. **(b,c)** TCR-repertoire analysis showing the number of TCR clonotypes per 10 million reads **(b)** and the clonality of expanded TCR clones defined as the frequency of the top 10% of TCR clones **(c)**. **(d)** Comparison of MHC class I expression in tissue samples from each patient. MHC class I expression indicates the total expression of HLA-A, HLA-B, and HLA-C molecules. **(e)** The relationship between the number of TCRβ clonotypes and the expression of MHC class I genes. Pearson's correlation was calculated between two groups.

Results

Relationship between T cell receptor (TCR) diversity and MHC gene expression. TILs (tumor-infiltrating lymphocytes) play a significant role in tumor-sites. Additionally, large amounts of TILs correlate with improved tumor survival³⁰; therefore, we preselected TIL-resident TNBC tissues for histologic analysis to identify potentially promising cancer epitopes (Fig. 2a and Supplementary Fig. S1) and scored TIL density by measuring the proportion of the stromal area infiltrated by lymphocytes, as previously described³¹. To select highly immunogenic HLA epitopes, we analyzed the intratumoral heterogeneity of the TCR repertoires in TIL-resident cancers from six patients with TNBC (Fig. 2b,c). The TCR repertoires comprise somatic recombination of the TCRα and β chains, allowing the specificity of each T cell clone to be determined by rearrangement of the V, D, and J segments of the TCRβ chain during generation of the highly variable complementary determining region^{32,33}. To evaluate TCR diversity of TILs, we assembled CDR3 sequences using the sequence reads of RNA-seq data. A unique CDR3 sequence of TCRα and TCRβ, respectively, was defined as a clone, and the number of clonotypes represents the number of unique clones per sample after normalization with the corresponding RNA-seq depth (Fig. 2b). The T cell clonal fraction was defined as the frequency of the top 10% of TCRα or TCRβ clones among total TCR clones (Fig. 2c). The top 10 most abundant TCRα and TCRβ sequences in each patient are shown in Supplementary Fig. S2, and expression of the three HLA-class I genes (HLA-A/B/C) is shown in Fig. 2d. We found a linear positive correlation between the number of TCRβ clonotypes and the expression of MHC-class I genes according to Pearson's correlation coefficient ($r = 0.68$) (Fig. 2e).

TIL immunoprofiles and immune-specific signatures. Several recent studies demonstrated the strong relationship between significant overall survival (OS) and cancer patients harboring a high number of CD8⁺ T cells and a low number of FoxP3⁺ T cells³⁴. In particular, the abundance of regulatory T (Treg) cells and macrophages correlated with worse outcomes, whereas the abundance of intratumoral CD8⁺ T cells and CD4⁺ T-helper (Th)1 cells correlated with better prognosis³⁵. Additionally, immune-specific signatures in TILs are of potential clinical significance; therefore, we estimated the distribution of TIL types in each patient according to WTS data using CIBERSORT computational analysis³⁶ (Fig. 3 and Supplementary Fig. S3). The relative proportion of each infiltrated immune cell was evaluated in each patient by quantifying immune composition from bulk-tissue gene-expression profiles, as enrichment of CD8/CD45RO and Th1 cells are considered positive prognostic factors³⁷. The results indicated that CD8⁺ T cells were highly infiltrated in both patients TNBC#2 and TNBC#6, whereas a large proportion of Treg cells was observed in patients TNBC#5 and TNBC#6 (Fig. 3a–c). Patient TNBC#6 displayed a highly enriched frequency of CD8⁺ T cells, as well as Treg cells, suggesting increased accumulation of TILs. Interestingly, we found a significantly increased proportion of CD8⁺ T cells relative to Treg cells in patient TNBC#2 as compared with that observed in other patients (Fig. 3d), suggesting the emergence

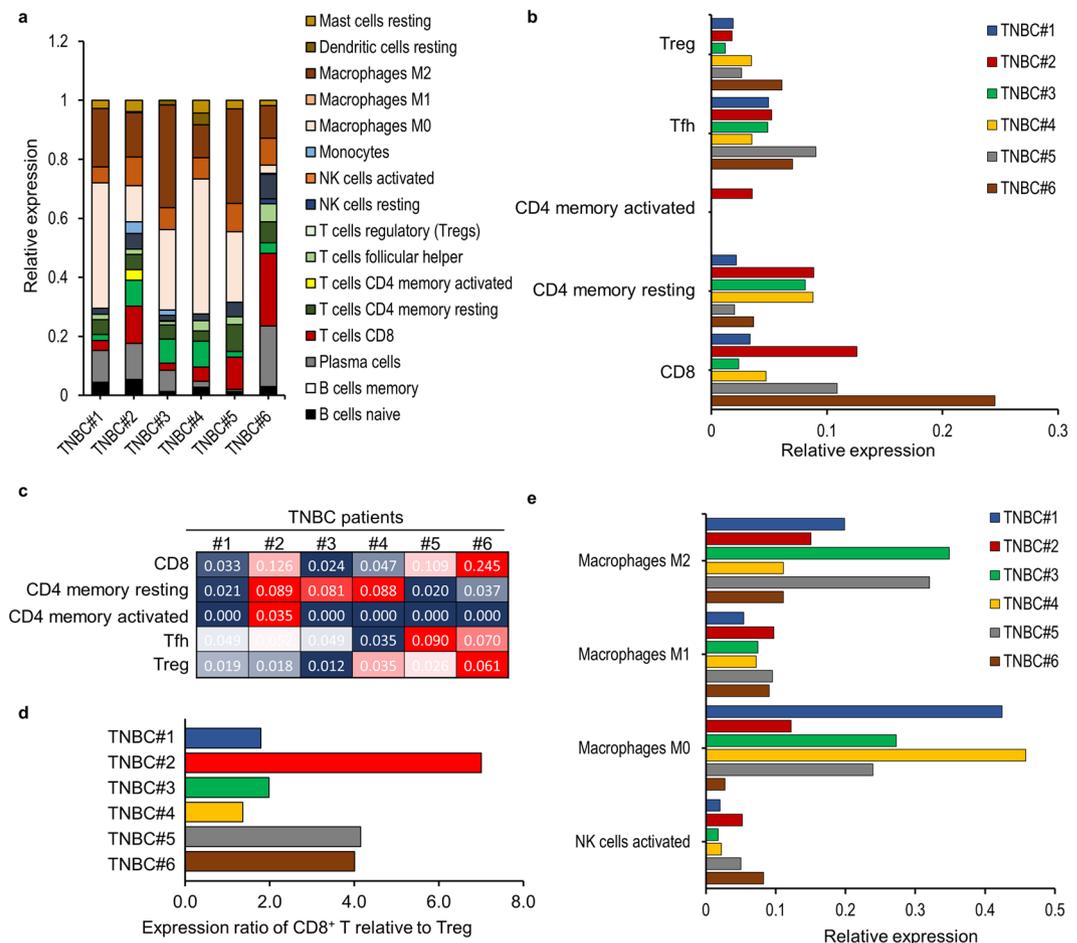


Figure 3. Immune-specific signatures in TILs. Characterization of specific immune-cell types estimated using CIBERSORT. Immune-cell population was evaluated by whole-transcription expression profiles from bulk tissue lysates. **(a)** Overview of immune-cell signatures among TILs. **(b,c)** The relative proportion of adoptive T cell-specific signatures among TILs. **(c)** The proportion of CD8⁺ T cells relative to Treg cells. **(d)** The relative proportion of innate immune-cell-specific signatures among TILs.

of promising tumor-associated antigens. On the other hand, we found elevated levels of immune-suppressive macrophages in patients TNBC#1, TNBC#3, and TNBC#5, which is predictive of a negative outcome (Fig. 3e).

HLA-peptidome and LC-MS/MS results. To identify naturally existing MHC class I-restricted ligands, we used an immunoproteomic approach involving tissue from six patients with TNBCs and an MHC-I antibody specific for the HLA-A, B, and C molecules (Fig. 4 and Supplementary Fig. S4). After immunoprecipitation, high-resolution LC-MS/MS analysis identified and quantified the HLA peptide sequences with a 1% false discovery rate (FDR) (Supplementary Fig. S5). Notably, the number of eluted peptides from each of the six patients were substantially different (Supplementary Fig. S6), although >96% of the HLA peptides analyzed by LC-MS/MS showed typical properties associated with epitope length. As expected, most of the peptides were nine amino acids long, with only a few having 13 to 15 amino acids, suggesting a high level of consistency (Fig. 4a and Supplementary Fig. S6). Clustering of the 9-mer HLA peptides showed predominant enrichment of residues at peptide positions 2 and 9 and consistent with the anchor motifs required by the binding groove of each HLA molecule²⁷ (Fig. 4b and Supplementary Fig. S7). Additionally, we found high numbers of CD8⁺ T and CD4⁺ Th1 cells infiltrating into the tumor sites of patient TNBC#2 and relative to Treg cells, with patient TNBC#2 showing a 4-fold higher number of CD8⁺ T cells as compared with that in patient TNBC#1 and accompanied by the lowest expression of MHC class I genes, suggesting a higher accumulation of antigen-specific CTLs in patient TNBC#2 (Supplementary Figs. S3,S8). A total of 594 peptides were identified from the tissue of patient TNBC#2 along with elevated expression of HLA genes, whereas only five peptides were received from tissue from patient TNBC#1 and all showing the lowest expression of MHC class I genes (Fig. 4c and Supplementary Fig. S8). Moreover, we observed a positive correlation between the number of eluted peptides relative to input lysate and the expression of MHC class I genes (Fig. 4c).

We then performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis to investigate the genes associated with the 594 HLA-binding peptides in patient TNBC#2 and the homotypic HLA-A*11:01 allele (Fig. 4d). Interestingly, the high-count genes ($N > 10$) were significantly enriched in KEGG pathways related to cancer, protein processing in the endoplasmic reticulum (ER), viral carcinogenesis, and

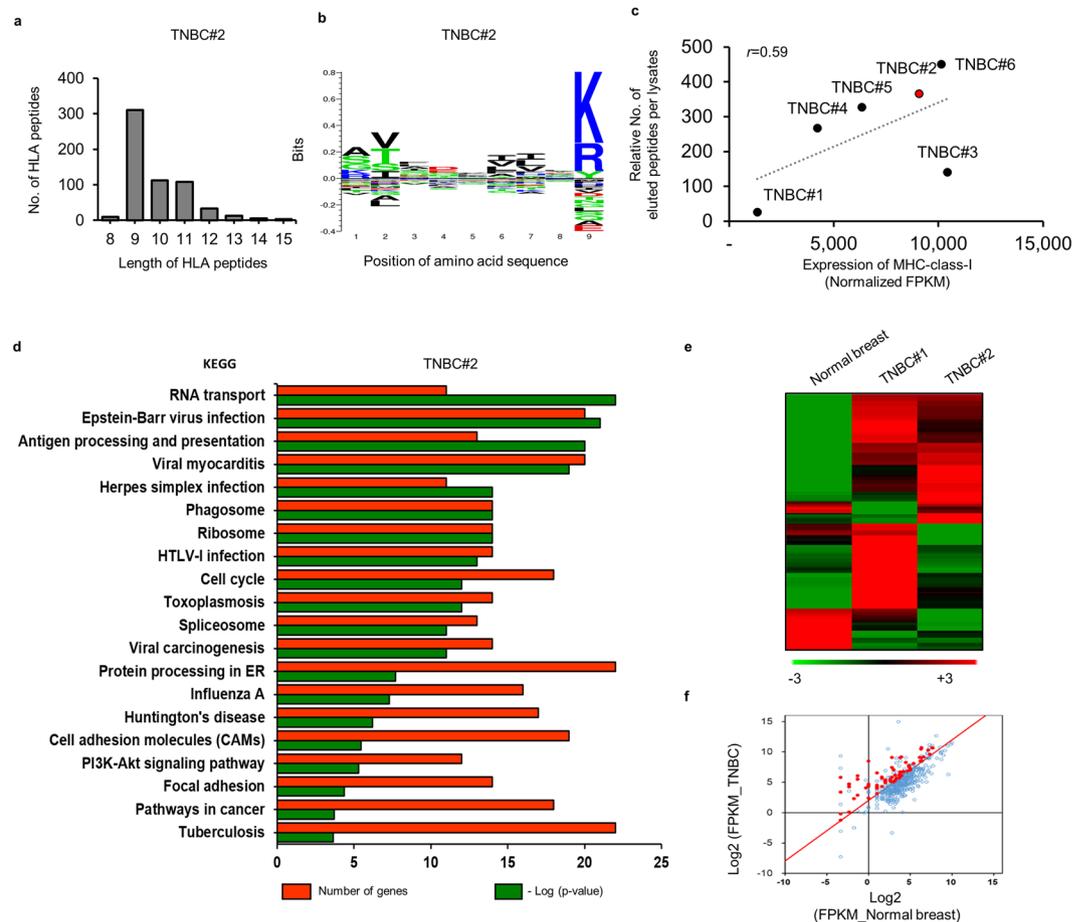


Figure 4. Identification of TNBC-associated HLA peptides by HLA-peptidomics. **(a)** The number and typical length distribution of HLA peptides. **(b)** Logo plot showing sequence analysis of all 9-mer HLA peptides. **(c)** The relationship between the relative number of eluted peptides per lysates and MHC gene expression. Pearson's correlation was calculated between two groups. **(d)** KEGG enrichment pathway analysis was performed to determine the function of genes derived from correlated HLA peptides. **(e,f)** Comparison of the transcription profiles of correlated genes in normal and TNBC tissue in order to remove genes exhibiting low expression (≤ 2 \log_2 fold change). **(e)** Heatmap created by MultiExperiment Viewer software. **(f)** Red dots in the scatterplot indicate genes exhibiting a fold change of ≥ 4 \log_2 according to mRNA levels relative to levels observed in normal breast tissue.

antigen processing and presentation. Numerous cancer-related genes overexpressed in cancer tissues contribute to cancer-specific or associated epitopes³⁸, and HLA epitopes require proteasomal digestion and translocation into the ER to bind MHC class-I molecules³⁹. It would be expected that the expression of genes encoding machinery responsible for antigen processing would be elevated under these circumstances. These results suggested that the eluted HLA peptides identified were naturally presented by HLA molecules. We further investigated the levels of the eluted peptides based on RNA-seq analysis of corresponding mRNA from the same sample (Fig. 4e,f). Compared with normal breast tissues, 174 of 594 peptides corresponding to proteins from the same sample showed elevated abundances in cancer tissues accompanied by significant differences in mRNA expression (≥ 2 \log_2 fold change). Subsequent *in silico* prediction of the HLA-binding affinities to the 174 HLA peptides and calculation of their respective binding affinity to specific alleles (predicted IC_{50})⁴⁰ yielded a list of the top 20 highest ranking peptides derived from patient TNBC#2 (Table 1).

A rapid imaging-based screening method to determine antigen-specific T cell response at the single-cell level.

To determine whether the experimentally identified peptides can functionally elicit an immune response, we evaluated cytokine production by the CD8⁺ T cells. Currently, intracellular cytokine staining (ICS)-based detection methods for monitoring *ex vivo* IFN γ response show low throughput relative to the number of candidate antigens being tested. Moreover, an individual antigen test requires large amounts of immune cells⁴¹. Therefore, we developed an efficient and comprehensive screening system to test CTL response based on a high-content, high-throughput imaging approach (Supplementary Fig. S9). This fluorescence-imaging-based screening system allows the use of lower numbers of viable cells up-scaled performance^{42,43}. Development of a 384-well format capable of screening mixed populations of T cells for their response against large number of

Ranking	Gene	Eluted Peptide	IC50 (nM)	Immuno-genicity score	HLA type	Ranking	Gene	Eluted Peptide	IC50 (nM)	Immuno-genicity score	HLA type
1	DYNLRB1	SLMHSFILK	5	0.030	A11:01	11	LDHA	GSLFLRTPK	12	0.178	A11:01
2	PRKDC	STFDTQITK	6	0.124	A11:01	12	PRKDC	STFDTQITKK	12	0.017	A11:01
3	NDUFC2	KTYGEIFEK	8	0.413	A11:01	13	PSMD14	AAMLDTVVFK	12	0.133	A11:01
4	EIF4A1	GIYAYGFEK	8	0.224	A11:01	14	LDHB	GSLFLQTPK	14	0.021	A11:01
5	PSMD14	AMLDTVVFK	8	0.199	A11:01	15	AURKB	KSHFIVALK	14	0.323	A11:01
6	DHCR7	AVSTFAMVK	9	0.012	A11:01	16	BCAP31	GAMEHFHMK	14	0.110	A11:01
7	TOMM5	RVTFFILKK	9	0.105	A11:01	17	RPL24	ASLADIMAK	15	0.057	A11:01
8	ROPN1B	SALGVITIK	10	0.242	A11:01	18	EIF3H	STYYGSFVTR	19	0.035	A11:01
9	TUBB	STAIQELFK	11	0.187	A11:01	19	NDUFA11	GTFLEGVAK	19	0.214	A11:01
10	TCPI	GVFEPTIVK	11	0.301	A11:01	20	C3	ATFGTQVVEK	19	0.129	A11:01

Table 1. Top 20 immunogenic CTL peptide candidates with scores were generated from a prediction tool, NetMHC.

peptides enables a cost-effective approach to phenotype analysis. Notably, cancer-associated antigens are highly attractive targets for determining their efficacy in triggering a T cell response; however, numerous clinical trials targeting TAAs for vaccine development have failed to demonstrate clinical efficacy due to immune self-tolerance. To identify highly immunogenic peptides incapable of eliciting self-tolerance, we tested the antigen-specific T cell response in PBMCs from a healthy donor. Monitoring *ex vivo* IFN γ -producing PBMC reactivity using our fluorescence-labelled cell-based screening system (Fig. 5a–c, Supplementary Figs. S10,S11) revealed significant IFN γ responses to two individual epitopes (eIF4A1-P and TCP1-P) in CD8⁺ T cells labelled with a FITC conjugated anti-human CD8 antibody (Fig. 5b). Additionally, treatment with the HLA-A*11:01-specific epitopes allowed detection of APC-conjugated IFN γ released from CD8⁺ T cells (Fig. 5a,b), with PMA and ionomycin co-treatment used to trigger T cell activation as a positive control. Similarly, PBMCs from two of the three healthy donors were reactive against same two epitopes (Fig. 5c).

To further analyze the peptide-induced CD8⁺ T cells, we generated HLA-A*11:01 tetramers targeting specific peptide-reactive CD8⁺ T cells. FACS analysis revealed that 9.99% and 7.50% of T cells were targeted by eIF4A1-P and TCP1-P, respectively, and detectable on day 12 of *ex vivo* T cell expansion (Fig. 5d). These findings suggested the efficacy of our method to screen highly immunogenic CTL epitopes using an imaging system on the detection of intracellular IFN γ levels following peptide stimulation. The two genes associated with the peptide epitopes, the translation initiation factor *eIF4A1* and *TCPI*, a member of the chaperonin-containing complex TCP1-containing ring complex (TRiC), are involved in tumor proliferation and survival. *eIF4A1* controls translation initiation and is a critical checkpoint protein involved in cell proliferation and tumorigenesis⁴⁴. Additionally, *TCPI*, as a TRiC member, is involved in tumor survival and growth and an oncogene driver⁴⁵. Moreover, these two genes are significantly overexpressed in tumor tissues according to WTS data and data from the Genotype-Tissue Expression Project public proteomic database, suggesting their potential as therapeutic targets. These results identified two epitopes derived from *eIF4A1* and *TCPI* as potential promising immunogenic antigens to boost T cell response.

Discussion

Patients with TNBC, which is defined by the absence of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2, have a higher tendency for recurrence at ~3 years after diagnosis^{46,47}. There are currently no known therapeutic targets for TNBC patients due to the molecular heterogeneity of the disease⁴⁸. Recent accumulation of massive and comprehensive bioinformatics data allowed identification of potential therapeutic targets associated with clinical survival⁴⁹. Interestingly, a TNBC subtype characterized by high levels of immune genes involved in T cell function, immune response, and antigen processing, was found to be associated with favorable prognosis, suggesting a close correlation between immune-gene signatures and better clinical outcomes⁵⁰. Therefore, it is possible that TILs controlling clinical cancer progression represent key factors for preselection of tumors prior to HLA-immunopeptidomics. In this study, TIL-resident tissues were pre-selected comprehensively to investigate a diversity of TCR repertoires and immune profile as predictors of clinical outcome. We then found a positive correlation between TCR diversity, reflecting clonal composition, and the expression of MHC- class I molecules, suggesting that active tumor-antigen presentation promotes the generation of antigen-specific TILs. Additionally, immune-specific signature analysis can discriminate specific immune-cell types in each patient and thus enhance the efficiency of selective HLA-peptidomic approaches. Notably, the expressional comparison of CD8⁺ T cells relative to immune-suppressive Treg cells is extremely crucial to select the high antigenicity of antigens reflecting the therapeutic efficacy. Thus, the extreme increase in the number of CD8⁺ T cells relative to that of Treg cells in patient TNBC#2 presented a major analytical reason for further *in vitro* T cell response testing. Furthermore, our results indicated a positive correlation between the number of peptides identified via HLA-peptidomics and the amount of HLA molecules expressed on the surface of cancer cells. It is also possible that the observed difference in the number of eluted peptides might have been influenced by HLA-expression levels resulting from active induction of antigen presentation on MHC molecules, which subsequently elicited a strong immune response. Thus, these findings suggested that several factors should be considered for successful HLA-peptidomic approaches influenced by TCR diversity and elevated expression of HLA genes.

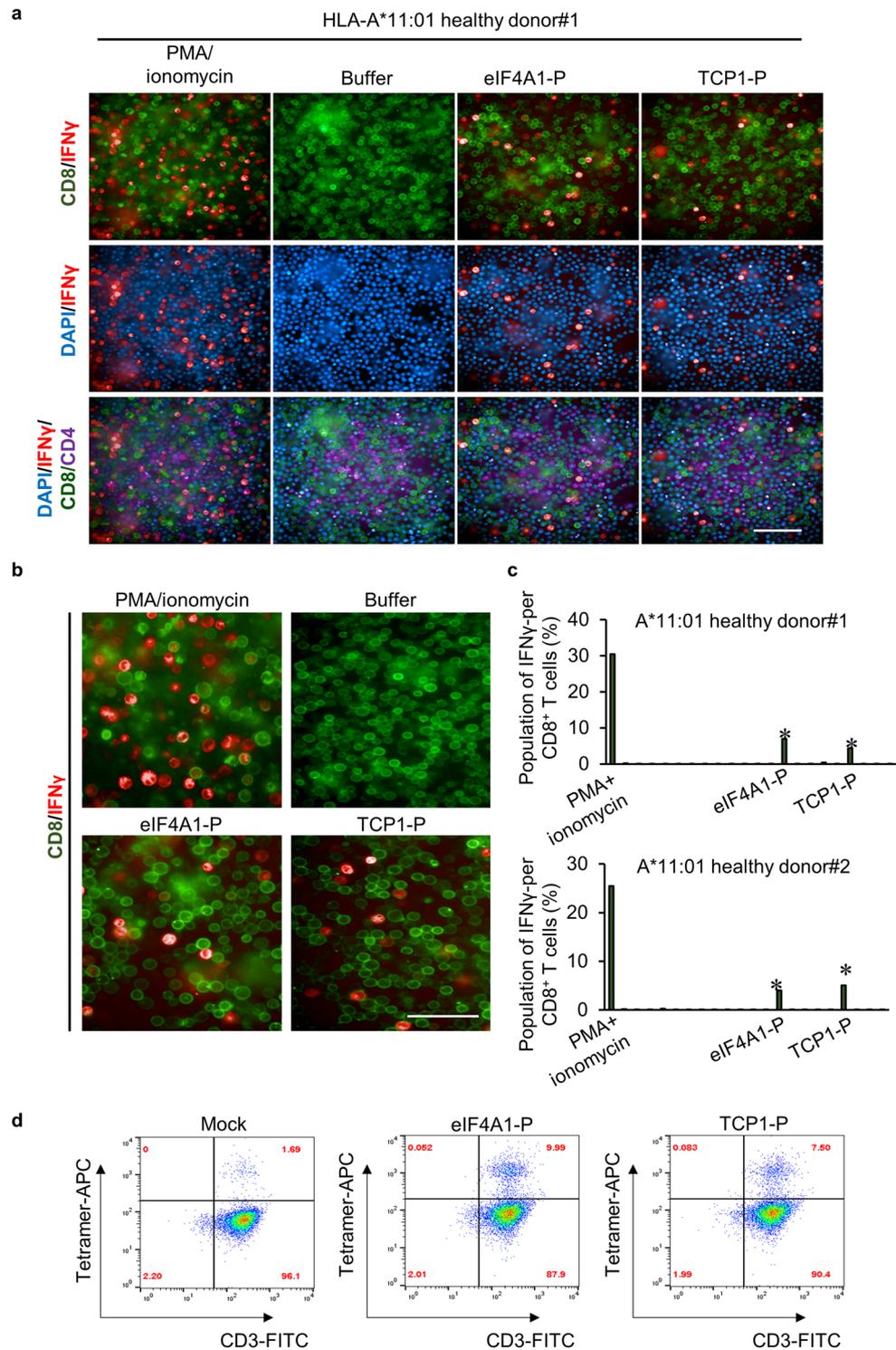


Figure 5. Identification of IFN γ response against peptide-reactive T cells using high-throughput imaging. (a) The 384-well screening system used to detect IFN γ -producing CD8⁺ T cells. (b) Detection of intracellular IFN γ levels in CD8⁺ T cells following peptide administration. PMA and ionomycin were used for non-specific T cell stimulation. Representative images are shown. (c) Bar graphs depicting the frequency of antigen-reactive CD8⁺ T cells. * $P < 0.005$, one-way analysis of variance, followed by Tukey's multiple comparison test. (d) Peptide/A*11:01 MHC-tetramer staining for antigen-specific T cells analyzed by flow cytometry.

Additionally, we showed that integrating HLA-peptidomics with imaging-based immunogenicity screening is applicable for the discovery of highly immunogenic CTL epitopes. Characterization of antigen-specific cellular immune response is essential to confirm vaccine-related effects specific to a cancer antigen. Currently, there are

only few assays (e.g., the enzyme-linked immunosorbent spot assay) capable of quantifying T cell responses^{51,52}, and the choice of which assay to use depends on the experimental scale, cost, equipment, reproducibility, and required detection sensitivity. A high-throughput imaging system provides an optimal platform for highly sensitive and quantitative analysis of individual T cells at the single-cell level. Moreover, ICS-based cytokine detection allows identification of specific cell subpopulations, even when using a small number of cells.

The immunopeptidome approach serves massive HLA-associated peptides as the collection of cancer epitopes; however, there are obstacles to rapidly determining the optimal set of promising epitopes by testing an enormous pool of peptide candidates in a single measurement. In this study, the HLA-peptidomics approach combined with comprehensive analysis of immune-specific signatures and TCR repertoires showed high selectivity to determine the immunogenic T-cell epitopes. Sequentially, the high-content imaging system allowed high-resolution analysis for T cell reactivity.

Despite the need for discovery of tumor-derived antigens for effective cancer vaccine development, selection of antigens that elicit robust immune response remains challenging. Here, we report a smart strategy for streamlined selection of cancer antigens in vaccine development. Through integrative multi-omics and high-content cell imaging, we identified highly immunogenic epitopes from patients with TNBC. Identification of potential vaccine epitopes coupled with immune-specific signature analysis, HLA-peptidomics, and single-cell-based immunogenicity testing offers a discriminative and powerful tool for cancer-vaccine development.

Methods

Patients and ethics statement. Tumor samples were collected from six different patients with TNBC who were diagnosed at stage I/II at the Korea University Anam Hospital (Seoul, Korea) and Asan Medical Center (Seoul, Korea). For patients TNBC#1 and TNBC#2, the biospecimens and data used in this study were provided by the Biobank of Korea University Anam Hospital (BBB2017-03-01). The ethics approval of this study was granted by the IRB committee at the Korea University Anam Hospital (BBB2017-03-01). Patients provided written informed consent for the collection and analysis of their tumor samples, and this project was reviewed and approved by the Institutional Review Board of the Korea University Anam Hospital (2017AN0399). Hematoxylin and eosin-stained samples were reviewed and selected by a breast pathologist based on the tumor-infiltrating lymphocyte (TIL) score as the proportion of the stromal area infiltrated by lymphocytes and according to the recommendations of the International TIL Breast Cancer Working Group³¹. Four patients-derived tissues (TNBC#3–#6) and peripheral blood mononuclear cells (PBMCs) from three healthy donors were obtained from Asan Medical Center, with requirements for informed donor consent for the use of data followed and approved by the Institutional Review Board (2017-0784 and 2016-0935) of Asan Medical Center. Healthy Donors also provided written informed consent for the collection and analysis of their PMBCs. All methods were carried out in accordance with relevant guidelines and regulations, approved by the institutional review board of Korea University Anam Hospital and Asan Medical center.

Analysis of high-throughput sequencing. DNA and RNA were simultaneously extracted from cryo-pulverized TNBC tissue powder using an AllPrep kit (Qiagen, Hilden, Germany), and libraries for whole-exome sequencing and total RNA sequencing, respectively, were prepared using TruSeq library prep kits (Illumina, San Diego, CA, USA). The libraries were sequenced using the HiSeq platform (Illumina), and raw data were mapped onto hg38 using the bwa mem algorithm (<http://bio-bwa.sourceforge.net/>). Variant calling was performed using the Genome Analysis Toolkit (<https://software.broadinstitute.org/gatk/>) as previously described⁵³. A custom proteogenomic search database was generated for the variants using the proteomics tool QUILTS (<http://www.fenyolab.org/tools/tools.html>). For WTS data, contaminating adapters and low-quality bases were removed with Trimmomatic⁵⁴, and the trimmed data were mapped onto hg38 using STAR (version 2.5.3a)⁵⁵. Gene expression, including that of HLA genes, was calculated by RSEM (version 1.3.0)⁵⁶, and HLA presentation on multiple cancer cells was evaluated using the Expression Atlas (<https://www.proteinatlas.org/humanproteome/tissue/cancer>)⁵⁷. For identification and quantification of TILs from WTS data, we used MiXCR (v.2.1.3; <https://mixcr.readthedocs.io/en/master/>) with the alignment parameter *-p rna-seq*⁵⁸. HLA typing at 4-digit resolution was performed using the HLAScan method and WES data (Synthekabio, Korea).

Purification of HLA-class I peptides. The LC-MS/MS analysis was performed as previously described⁵⁹. HLA-class I peptides were purified from TNBC tissues, and frozen tissues were pulverized with a CP02 cryo-OPREP automated dry pulverizer (Covaris, Woburn, MA, USA), followed by incubation at 4 °C for 1 h with lysis buffer containing 0.25% sodium deoxycholate, 0.2 mM iodoacetamide, 1 mM EDTA, 1 mM PMSE, 1% octyl- β -D-glucopyranoside (Sigma-Aldrich, St. Louis, MO, USA), and a protease-inhibitor cocktail (Roche, Mannheim, Germany) in phosphate-buffered saline (PBS). The lysates were cleared by centrifugation for 20 min at 15,000 g and 4 °C. HLA-class I molecules were purified using the W6/32 monoclonal antibody bound to Amino-Link beads (Thermo Scientific, Waltham, MA, USA), as previously described⁶⁰. The anti-HLA-class I antibody was purchased from Abcam (Cambridge, UK). HLA-peptide complexes were eluted from the affinity column using five column volumes of 0.1 N acetic acid. The eluted HLA-class I proteins and the released peptides were loaded on Sep-Pak tC18 columns (Waters, Milford, MA, USA), and the peptide fraction was eluted with 30% acetonitrile in 0.1% trifluoroacetic acid, followed by drying by vacuum centrifugation.

LC-MS/MS analysis of HLA-class I peptides. The LC-MS/MS analysis was performed as previously described⁵⁹. Dried peptide samples were reconstituted in 10 μ L of 0.1% formic acid, and an aliquot containing ~4 μ L was injected from a cooled (10 °C) autosampler into a reversed-phase Magic C18aq column (15 cm \times 75 μ m (packed in-house); Michrom BioResources, Auburn, CA, USA) on an Eksigent nanoLC 2D system at a flow rate of 300 nL/min. Prior to use, the column was equilibrated with 95% buffer A (0.1% formic acid in water) and 5%

buffer B (0.1% formic acid in acetonitrile). The peptides were eluted with a linear gradient from 5% to 30% buffer B over 70 min and 30% to 70% buffer B over 5 min, followed by an organic wash and aqueous re-equilibration at a flow rate of 300 nL/min, with a total run time of 95 min. The high-performance liquid chromatography system was coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) operated in data-dependent acquisition mode. Full scans (m/z 300–1800) were acquired at a resolution of 60,000 using an automatic gain-control (AGC) target value of $1e6$ and a maximum ion-injection time of 10 ms. Tandem mass spectra were generated for up to 5 precursors by collision-induced dissociation in the ion-trap using a normalized collision energy of 35%. The dynamic exclusion was set to 60 s, and fragment ions were detected at a normal scan mode using an AGC target value of $1e5$ and a maximum ion-injection time of 500 ms. Source ionization parameters were as follows: spray voltage, 1.9 kV; and capillary temperature, 275 °C.

Data analysis of the HLA-class I peptidome. MS data were analyzed using MaxQuant software (v.1.5.8.3)⁶¹ against the UniProt database (April 40, 2016; <https://www.uniprot.org/>), and personalized human database from patient-derived WES data. N-terminal acetylation and methionine oxidation were set as variable modifications, enzyme specificity was set as “unspecific,” and FDRs for peptides and proteins were set at 0.01 and 1, respectively. Possible sequence matches were restricted to eight to 15 amino acids, a maximum peptide mass of 2000 Da, and a maximum charge state of three. Main search peptide tolerance was set at 10, and the box for “Use MS2 centroids” was checked. Hits to the reverse database and contaminants were removed from the “peptide.txt” output file produced by MaxQuant. Peptides were subjected to HLA-class I binding and immunogenicity prediction analyses using NetMHC (<http://www.cbs.dtu.dk/services/NetMHC/>) and the MHC I immunogenicity portion of the IEDB Analysis Resource (<http://tools.iedb.org/immunogenicity/>), respectively. The logo-plots were constructed using a Seq2Logo method⁶².

Rapid single cell-based imaging analysis of T cell reactivity. The IFN γ response was evaluated in PBMCs isolated from three HLA-A*11:01-positive healthy donors. T cell responses to treatment with individual peptides were monitored after 12 days of *in vitro* culture, as previously described¹¹. Briefly, PBMCs were cultured in RPMI-1640 supplemented with L-glutamine, non-essential amino acids, HEPES, β -mercaptoethanol, sodium pyruvate, penicillin/streptomycin (Gibco; Thermo Fisher Scientific), and 10% human AB serum (Gibco), with a total of 5×10^4 PBMCs used in each well. For antigen-specific T cell expansion, individual peptides (2 μ g/mL) were incubated with PBMCs in the presence of interleukin (IL)-7 (20 ng/mL; Peprotech, Rocky Hill, NJ, USA) for 3 days. Each peptide with >95% purity was synthesized by Synpeptide (Shanghai, China). For non-specific T cell expansion, T cells were stimulated using a T cell activation/expansion kit (Miltenyi Biotec, Bergisch Gladbach, Germany), and cells were cultured every 3 days by replacing the medium with fresh half-medium containing IL-7 (5 ng/mL) and IL-15 (5 ng/mL; Peprotech). On day 12, for *ex vivo* intracellular IFN γ detection, each peptide (10 μ g/mL) or phytohemagglutinin (PMA) (50 ng/mL; Sigma-Aldrich) plus ionomycin (1 μ g/mL; Sigma-Aldrich) for the positive control was administered in the presence of Brefeldin A (1:1000; BioLegend, San Diego, CA, USA) for overnight incubation according to the ICS protocol. After intracellular fixation using BD Cytofix/Cytoperm fixation/permeabilization kit (BD Biosciences, San Jose, CA, USA), cells were stained with antibodies against surface markers and IFN γ at 4 °C for either 1 h or overnight. For visualization of IFN γ -producing T cells, a fluorescein isothiocyanate (FITC)-conjugated anti-human CD8 α antibody (R&D Systems, Minneapolis, MN, USA), an Alexa594-conjugated anti-human CD4 antibody (BioLegend), and an allophycocyanin (APC)-conjugated anti-human IFN γ antibody (BioLegend) were specifically labeled for CD8 $^+$ T, CD4 $^+$ T, and intracellular IFN γ capture, respectively. Cells were then washed with PBS buffer, and high-throughput imaging was performed using the Operetta CLS high-content analysis system equipped with Harmony software (PerkinElmer, Waltham, MA, USA).

Detection of antigen-specific T cells. To generate the p/MHC tetramer, a biotinylated HLA-A*11:01 monomer complexed with each peptide was obtained from ImmunoMAX Co., Ltd. (Seoul, Korea). Peptides with >95% purity were synthesized by Synpeptide (China), and their sequences are provided in Table 1. To generate an APC-labeled p/MHC complex tetramer, p/MHC complex monomers were tetramerized in the presence of APC-conjugated streptavidin (BD Biosciences). T cells were then stained with 1 ng/ μ L of the p/MHC tetramer in Fluorescence-activated cell sorting (FACS) buffer (BioLegend) for 10 min at room temperature, and after washing, immunofluorescence was detected by FACS. T cells were gated according to the populations of cells labeled with the FITC-conjugated anti-human CD3 antibody (BioLegend).

Received: 4 November 2019; Accepted: 9 March 2020;

Published online: 03 April 2020

References

- Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74, <https://doi.org/10.1126/science.aaa4971> (2015).
- van der Burg, S. H., Arens, R., Ossendorp, F., van Hall, T. & Melief, C. J. Vaccines for established cancer: overcoming the challenges posed by immune evasion. *Nat Rev Cancer* **16**, 219–233, <https://doi.org/10.1038/nrc.2016.16> (2016).
- Buonaguro, L., Petrizzo, A., Tornesello, M. L. & Buonaguro, F. M. Translating tumor antigens into cancer vaccines. *Clin Vaccine Immunol* **18**, 23–34, <https://doi.org/10.1128/CVI.00286-10> (2011).
- Tagliamonte, M., Petrizzo, A., Tornesello, M. L., Buonaguro, F. M. & Buonaguro, L. Antigen-specific vaccines for cancer treatment. *Hum Vaccin Immunother* **10**, 3332–3346, <https://doi.org/10.4161/21645515.2014.973317> (2014).
- Yarchoan, M., Johnson, B. A. III, Lutz, E. R., Laheru, D. A. & Jaffee, E. M. Targeting neoantigens to augment antitumour immunity. *Nat Rev Cancer* **17**, 569, <https://doi.org/10.1038/nrc.2017.74> (2017).
- Efremova, M., Finotello, F., Rieder, D. & Trajanoski, Z. Neoantigens Generated by Individual Mutations and Their Role in Cancer Immunity and Immunotherapy. *Front Immunol* **8**, 1679, <https://doi.org/10.3389/fimmu.2017.01679> (2017).

7. Monach, P. A., Meredith, S. C., Siegel, C. T. & Schreiber, H. A unique tumor antigen produced by a single amino acid substitution. *Immunity* **2**, 45–59 (1995).
8. Lennerz, V. *et al.* The response of autologous T cells to a human melanoma is dominated by mutated neoantigens. *Proc Natl Acad Sci USA* **102**, 16013–16018, <https://doi.org/10.1073/pnas.0500090102> (2005).
9. Carreno, B. M. *et al.* Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803–808, <https://doi.org/10.1126/science.aaa3828> (2015).
10. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222–226, <https://doi.org/10.1038/nature23003> (2017).
11. Ott, P. A. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221, <https://doi.org/10.1038/nature22991> (2017).
12. Li, L., Goedegebuure, S. P. & Gillanders, W. E. Preclinical and clinical development of neoantigen vaccines. *Ann Oncol* **28**, xii11–xii17, <https://doi.org/10.1093/annonc/mdx681> (2017).
13. Capietto, A. H., Jhunjhunwala, S. & Delamarre, L. Characterizing neoantigens for personalized cancer immunotherapy. *Curr Opin Immunol* **46**, 58–65, <https://doi.org/10.1016/j.coi.2017.04.007> (2017).
14. Guo, C. *et al.* Therapeutic cancer vaccines: past, present, and future. *Adv Cancer Res* **119**, 421–475, [10.1016/B978-0-12-407190-2.00007-1](https://doi.org/10.1016/B978-0-12-407190-2.00007-1) (2013).
15. Melief, C. J., van Hall, T., Arens, R., Ossendorp, F. & van der Burg, S. H. Therapeutic cancer vaccines. *J Clin Invest* **125**, 3401–3412, <https://doi.org/10.1172/JCI80009> (2015).
16. Melero, I. *et al.* Therapeutic vaccines for cancer: an overview of clinical trials. *Nat Rev Clin Oncol* **11**, 509–524, <https://doi.org/10.1038/nrclinonc.2014.111> (2014).
17. Aurisicchio, L., Pallocca, M., Ciliberto, G. & Palombo, F. The perfect personalized cancer therapy: cancer vaccines against neoantigens. *J Exp Clin Cancer Res* **37**, 86, <https://doi.org/10.1186/s13046-018-0751-1> (2018).
18. Hu, Z., Ott, P. A. & Wu, C. J. Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nat Rev Immunol* **18**, 168–182, <https://doi.org/10.1038/nri.2017.131> (2018).
19. Guo, Y., Lei, K. & Tang, L. Neoantigen Vaccine Delivery for Personalized Anticancer Immunotherapy. *Front Immunol* **9**, 1499, <https://doi.org/10.3389/fimmu.2018.01499> (2018).
20. Ahmed, M. S. & Bae, Y. S. Dendritic cell-based therapeutic cancer vaccines: past, present and future. *Clin Exp Vaccine Res* **3**, 113–116, [10.7774/cevr.2014.3.2.113](https://doi.org/10.7774/cevr.2014.3.2.113) (2014).
21. Combadiere, B., Beaujean, M., Chaudesaigues, C. & Vieillard, V. Peptide-Based Vaccination for Antibody Responses Against HIV. *Vaccines (Basel)* **7**, <https://doi.org/10.3390/vaccines7030105> (2019).
22. Valdes-Zayas, A. *et al.* Immunologic Response Elicited in Breast Cancer Patients Receiving a NeuGcGM3-based Vaccine as Adjuvant Therapy. *J Immunother* **40**, 289–301, <https://doi.org/10.1097/CJI.0000000000000175> (2017).
23. Kalaora, S. *et al.* Combined Analysis of Antigen Presentation and T-cell Recognition Reveals Restricted Immune Responses in Melanoma. *Cancer Discov* **8**, 1366–1375, <https://doi.org/10.1158/2159-8290.CD-17-1418> (2018).
24. Bassani-Sternberg, M. *et al.* Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* **7**, 13404, <https://doi.org/10.1038/ncomms13404> (2016).
25. Melief, C. J. M. & Kessler, J. H. Novel insights into the HLA class I immunopeptidome and T-cell immunosurveillance. *Genome Med* **9**, 44, <https://doi.org/10.1186/s13073-017-0439-8> (2017).
26. Abelin, J. G. *et al.* Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity* **46**, 315–326, <https://doi.org/10.1016/j.immuni.2017.02.007> (2017).
27. Khodadoust, M. S. *et al.* Antigen presentation profiling reveals recognition of lymphoma immunoglobulin neoantigens. *Nature* **543**, 723–727, <https://doi.org/10.1038/nature21433> (2017).
28. Croft, N. P. *et al.* Kinetics of antigen expression and epitope presentation during virus infection. *Plos Pathog* **9**, e1003129, <https://doi.org/10.1371/journal.ppat.1003129> (2013).
29. Caron, E. *et al.* Analysis of Major Histocompatibility Complex (MHC) Immunopeptidomes Using Mass Spectrometry. *Mol Cell Proteomics* **14**, 3105–3117, [10.1074/mcp.M115.052431](https://doi.org/10.1074/mcp.M115.052431), [10.1074/mcp.O115.052431](https://doi.org/10.1074/mcp.O115.052431) (2015).
30. Gooden, M. J., de Bock, G. H., Leffers, N., Daemen, T. & Nijman, H. W. The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. *Br J Cancer* **105**, 93–103, <https://doi.org/10.1038/bjc.2011.189> (2011).
31. Adams, S. *et al.* Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J Clin Oncol* **32**, 2959–2966, <https://doi.org/10.1200/JCO.2013.55.0491> (2014).
32. Madi, A. *et al.* T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *Elife* **6**, <https://doi.org/10.7554/eLife.22057> (2017).
33. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402, <https://doi.org/10.1038/334395a0> (1988).
34. Diana, A. *et al.* Prognostic value, localization and correlation of PD-1/PD-L1, CD8 and FOXP3 with the desmoplastic stroma in pancreatic ductal adenocarcinoma. *Oncotarget* **7**, 40992–41004, <https://doi.org/10.18632/oncotarget.10038> (2016).
35. DeNardo, D. G. *et al.* Leukocyte complexity predicts breast cancer survival and functionally regulates response to chemotherapy. *Cancer Discov* **1**, 54–67, <https://doi.org/10.1158/2159-8274.CD-10-0028> (2011).
36. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453–457, <https://doi.org/10.1038/nmeth.3337> (2015).
37. Steele, K. E. *et al.* Measuring multiple parameters of CD8+ tumor-infiltrating lymphocytes in human cancers by image analysis. *J Immunother Cancer* **6**, 20, <https://doi.org/10.1186/s40425-018-0326-x> (2018).
38. Lang, J. M., Andrei, A. C. & McNeel, D. G. Prioritization of cancer antigens: keeping the target in sight. *Expert Rev Vaccines* **8**, 1657–1661, <https://doi.org/10.1586/erv.09.134> (2009).
39. Rock, K. L., York, I. A. & Goldberg, A. L. Post-proteasomal antigen processing for major histocompatibility complex class I presentation. *Nat Immunol* **5**, 670–677, <https://doi.org/10.1038/ni1089> (2004).
40. Hoof, I. *et al.* NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* **61**, 1–13, <https://doi.org/10.1007/s00251-008-0341-z> (2009).
41. Busch, D. H. & Pamer, E. G. MHC class I/peptide stability: implications for immunodominance, *in vitro* proliferation, and diversity of responding CTL. *J Immunol* **160**, 4441–4448 (1998).
42. Li, L., Zhou, Q., Voss, T. C., Quick, K. L. & LaBarbera, D. V. High-throughput imaging: Focusing in on drug discovery in 3D. *Methods* **96**, 97–102, <https://doi.org/10.1016/j.ymeth.2015.11.013> (2016).
43. Knight, S. *et al.* Enabling 1536-Well High-Throughput Cell-Based Screening through the Application of Novel Centrifugal Plate Washing. *SLAS Discov* **22**, 732–742, <https://doi.org/10.1177/2472555216683650> (2017).
44. Eberle, J., Fecker, L. F., Bittner, J. U., Orfanos, C. E. & Geilen, C. C. Decreased proliferation of human melanoma cell lines caused by antisense RNA against translation factor eIF-4A1. *Br J Cancer* **86**, 1957–1962, <https://doi.org/10.1038/sj.bjc.6600351> (2002).
45. Guest, S. T., Kratche, Z. R., Bollig-Fischer, A., Haddad, R. & Ethier, S. P. Two members of the TRiC chaperonin complex, CCT2 and TCP1 are essential for survival of breast cancer cells and are linked to driving oncogenes. *Exp Cell Res* **332**, 223–235, <https://doi.org/10.1016/j.yexcr.2015.02.005> (2015).

46. Kwon, J. *et al.* A Prognostic Model for Patients with Triple-Negative Breast Cancer: Importance of the Modified Nottingham Prognostic Index and Age. *J Breast Cancer* **20**, 65–73, <https://doi.org/10.4048/jbc.2017.20.1.65> (2017).
47. Jitariu, A. A., Cimpean, A. M., Ribatti, D. & Raica, M. Triple negative breast cancer: the kiss of death. *Oncotarget* **8**, 46652–46662, <https://doi.org/10.18632/oncotarget.16938> (2017).
48. Stagg, J. & Allard, B. Immunotherapeutic approaches in triple-negative breast cancer: latest research and clinical prospects. *Ther Adv Med Oncol* **5**, 169–181, <https://doi.org/10.1177/1758834012475152> (2013).
49. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70, <https://doi.org/10.1038/nature11412> (2012).
50. Fridman, W. H., Pages, F., Sautes-Fridman, C. & Galon, J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer* **12**, 298–306, <https://doi.org/10.1038/nrc3245> (2012).
51. Desbien, A. L. *et al.* Squalene emulsion potentiates the adjuvant activity of the TLR4 agonist, GLA, via inflammatory caspases, IL-18, and IFN-gamma. *Eur J Immunol* **45**, 407–417, <https://doi.org/10.1002/eji.201444543> (2015).
52. Beebe, E. A. & Orr, M. T. Assessment of Antigen-Specific Cellular Immunogenicity Using Intracellular Cytokine Staining, ELISpot, and Culture Supernatants. *Methods Mol Biol* **1494**, 313–320, https://doi.org/10.1007/978-1-4939-6445-1_24 (2017).
53. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303, <https://doi.org/10.1101/gr.107524.110> (2010).
54. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
55. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, <https://doi.org/10.1093/bioinformatics/bts635> (2013).
56. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, <https://doi.org/10.1186/1471-2105-12-323> (2011).
57. Petryszak, R. *et al.* Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res* **44**, D746–752, <https://doi.org/10.1093/nar/gkv1045> (2016).
58. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* **12**, 380–381, <https://doi.org/10.1038/nmeth.3364> (2015).
59. Yeom, J., Ju, S., Choi, Y., Paek, E. & Lee, C. Comprehensive analysis of human protein N-termini enables assessment of various protein forms. *Sci Rep* **7**, 6599, <https://doi.org/10.1038/s41598-017-06314-9> (2017).
60. Bassani-Sternberg, M. *et al.* Soluble plasma HLA peptidome as a potential source for cancer biomarkers. *Proc Natl Acad Sci USA* **107**, 18769–18776, <https://doi.org/10.1073/pnas.1008501107> (2010).
61. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372, <https://doi.org/10.1038/nbt.1511> (2008).
62. Thomsen, M. C. & Nielsen, M. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res* **40**, W281–287, <https://doi.org/10.1093/nar/gks469> (2012).

Acknowledgements

We would like to express my special appreciation and thank Dr. Gil-Je Lee (Perkin Elmer) for her constant help quantitatively in analysing immune cell imaging. This research was supported by the intramural research program (2E30350) of KIST, and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1C1C1008181 and No. 2017M3C9A5031595).

Author contributions

C.L. conceived the study. K.H., C.L. and M.J. designed the experiments. K.H., D.P. and M.J. wrote the manuscript. K.H., S.J. and J.L. carried out HLA-peptidomic experiments. D.P. performed NGS analysis. S.H., H.L., Y.A.K. and K.P. harvested clinical samples and analyse H&E samples. S.P., Y.L., Y.E.L. and M.J. performed the experiments regarding to T cell immunogenicity.

Competing interests

KIST has filed a patent on which M.J., C.L., K.H., J.L. and S.J. are listed as inventors. The other authors declare no competing financial interest.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-62244-z>.

Correspondence and requests for materials should be addressed to K.-C.H., C.L. or M.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020