

# UC San Diego

## UC San Diego Previously Published Works

### Title

MHC class I loaded ligands from breast cancer cell lines: A potential HLA-I-typed antigen collection

### Permalink

<https://escholarship.org/uc/item/64f378ww>

### Authors

Rozanov, Dmitri V

Rozanov, Nikita D

Chiotti, Kami E

et al.

### Publication Date

2018-03-01

### DOI

10.1016/j.jprot.2018.01.004

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



# HHS Public Access

Author manuscript

*J Proteomics*. Author manuscript; available in PMC 2018 June 13.

Published in final edited form as:

*J Proteomics*. 2018 March 30; 176: 13–23. doi:10.1016/j.jpro.2018.01.004.

## MHC class I loaded ligands from breast cancer cell lines: A potential HLA-I-typed antigen collection

Dmitri V. Rozanov<sup>1</sup>, Nikita D. Rozanov<sup>2</sup>, Kami Chiotti<sup>1</sup>, Ashok Reddy<sup>3</sup>, Phillip A. Wilmarth<sup>3</sup>, Larry L. David<sup>3</sup>, Seung W. Cha<sup>4</sup>, Sunghee Woo<sup>5</sup>, Pavel Pevzner<sup>6</sup>, Vineet Bafna<sup>7</sup>, Gregory G. Burrows<sup>8</sup>, Juha K. Rantala<sup>9</sup>, Trevor Levin<sup>10</sup>, Pavana Anur<sup>1</sup>, Katie Johnson-Camacho<sup>1</sup>, Shaadi Tabatabaei<sup>1</sup>, Daniel J. Munson<sup>11</sup>, Tullia C. Bruno<sup>12</sup>, Jill E. Slansky<sup>11</sup>, John W. Kappler<sup>13</sup>, Naoto Hirano<sup>14</sup>, Sebastian Boegel<sup>15</sup>, Bernard A. Fox<sup>16</sup>, Colt Egelston<sup>17</sup>, Diana L. Simons<sup>17</sup>, Grecia Jimenez<sup>17</sup>, Peter P. Lee<sup>17</sup>, Joe W. Gray<sup>10,18</sup>, and Paul T. Spellman<sup>1</sup>

<sup>1</sup>Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, Oregon, United States of America <sup>2</sup>Oregon State University, Corvallis, Oregon, United States of America <sup>3</sup>Proteomics Shared Resource, Oregon Health and Science University, Portland, Oregon, United States of America <sup>4</sup>Electrical and Computer Engineering, University of California, San Diego, California, United States of America <sup>5</sup>School of Medicine, Johns Hopkins University, Baltimore, Maryland, United States of America <sup>6</sup>The NIH Center for Computational Mass Spectrometry, University of California, San Diego, San Diego, California, United States of America <sup>7</sup>Computer Science & Engineering, University of California, San Diego, California, United States of America <sup>8</sup>Neurology and Biochemistry & Molecular Biology, Oregon Health and Science University, Portland, Oregon, United States of America <sup>9</sup>Misvik biology Ltd, Turku, Finland <sup>10</sup>Department of Biomedical Engineering, Oregon Health & Science University, Portland, Oregon, United States of America <sup>11</sup>Department of Immunology & Microbiology, University of Colorado, Denver, Colorado, United States of America <sup>12</sup>Department of Immunology, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America <sup>13</sup>National Jewish Medical and Research Center, Denver, Colorado, United States of America <sup>14</sup>Margaret Cancer Centre, Toronto, Ontario, Canada <sup>15</sup>University Medical Center, Johannes Gutenberg-University, Mainz, Germany <sup>16</sup>Laboratory of Molecular and Tumor Immunology, Chiles Research Institute Providence PDX Medical Center, Portland, Oregon, United States of America <sup>17</sup>City of Hope National Medical Center, Duarte, California, United States of America <sup>18</sup>Center for Health & Healing, Oregon Health and Science University, Portland, Oregon, United States of America

### Abstract

Breast cancer therapy based on amplifying a patient's antitumor immune response depends on the availability of appropriate MHC class I-restricted, breast cancer-specific epitopes. To build a catalog of peptides presented by breast cancer cells, we undertook systematic MHC class I immunoprecipitation followed by elution of MHC class I-loaded peptides in breast cancer cell lines. We determined the sequence of 3,196 MHC class I-bound peptides representing 1,921 proteins from a panel of 20 breast cancer cell lines including basal, luminal, and claudin-low subtypes. The data has been deposited to the ProteomeXchange with identifier PXD006406. After removing duplicate peptides, i.e., the same peptide eluted from more than one cell line, the total number of unique peptides was 2,740. Of the unique peptides eluted, more than 1,750 had been

previously identified, and of these, sixteen have been shown to be immunogenic. Importantly, only 3 of these immunogenic peptides have been identified in breast cancer cells in earlier studies. MHC class I binding probability of eluted peptides was used to plot the distribution of MHC class I allele-specific peptides in accordance with the binding score for each breast cancer cell line. We also determined that the tested breast cancer cells presented 89 mutation-containing peptides and peptides derived from aberrantly translated genes, 7 of which were shared between four or two different cell lines. Overall, the high throughput identification of MHC class I-loaded peptides is an effective strategy for systematic characterization of cancer peptides, and could be employed for design of multi-peptide anticancer vaccines.

---

## Introduction

Breast cancer is the most frequently occurring cancer in women in all racial and ethnic groups [1]. Despite the positive outcomes for most breast cancer patients, the side effects of current treatment are substantial [2–4], and lower toxicity treatments are needed. Further, for a significant minority of breast cancer patients, current treatments are inadequate. Substantial advances in the field of immunotherapeutics have resulted in approvals of both vaccines [5] and immune checkpoint inhibitors [6].

It is now recognized that invasive ductal carcinoma of the breast is a heterogeneous disease consisting of several major molecularly defined subtypes, including Luminal A, Luminal B, HER2, Basal, triple-negative, and the claudin-low subset [7, 8]. These subtypes have distinct clinical, genomic, and proteomic features; and it is becoming clear that there are differences between breast cancer subtype and response to specific therapeutic agents [9, 10]. Luminal tumor cells consisting of luminal A and B cells look like cells that start in the inner (luminal) cells lining the mammary ducts. Luminal A tumors have the best prognosis with high survival rate and low recurrence rate. Luminal B cells have poorer prognosis than luminal A cells and luminal B cells also tend to be estrogen receptor (ER) positive. Triple negative breast cancer consists of several subsets. One subset is basal-like. Basal-like tumor cells look like outer (basal) cells surrounding the mammary ducts. Most triple negative tumor cells are also basal-like and vice versa most basal-like tumor cells are also triple negative cells. Triple negative/basal-like tumors are often aggressive and have a poorer prognosis as compared to luminal A and B tumors. Claudin-low cells represent less common molecular subtypes of breast cancer. Claudin-low cells have more enriched cancer-stem cell-like features and higher activities of ER and progesterone receptor (PR) pathways than basal tumor cells. Claudin-low cells are also least differentiated among all subtypes of breast cancer cells and preferentially display a triple-negative phenotype. Claudin-low tumors are associated with poor survival. Human epidermal growth factor receptor 2 (HER2) subtype breast tumors can be HER2-positive (70%) and HER2-negative (30%). HER2 tumors tend to be both ER- and PR-negative and have poor prognosis [11].

A powerful and safe approach to breast cancer therapy is to amplify the patient's anti-tumor immune response by stimulating autologous T cells with antigen-presenting cells (APCs) loaded with breast cancer-specific antigens. T-cell based immunotherapeutic approaches require knowledge of cancer-specific antigens, and the capacity to determine which of these

cancer-specific antigens are presented by the cancer cells of the individual patient. During the past two decades, antigens from various tumor cells have been identified, and their MHC class I-restricted epitopes have been predicted and confirmed in T cell-based assays [12–15]. Current strategies to identify immunogenic tumor specific MHC class I-restricted epitopes rely on either expression analysis of tumor associated antigens (TAA) followed by synthesis of predicted peptides and T cell activation assays [16, 17], or elution of MHC class I-loaded peptides directly from tumors followed by mass spectrometry analysis, gene expression profiling, and *in vitro* T cell assays [18]. The first strategy largely depends on the accuracy of the predicting tool which is often compromised by the inability to take into account all of many factors such as proteasomal cleavage specificity, additional trimming at the N-terminus of peptide, and transport of peptides into the endoplasmic reticulum [19]. Proteasomal cleavage is especially important in peptide prediction due to differences in regulation of gene expression and cleavage preferences of proteasomes in tumor cells compared to normal cells [20, 21]. The second strategy, while less subject to error, is limited because tumor cells extracted from cancer patients are often unavailable or only available in small quantities. To address these shortcomings, we used well established breast cancer cell lines to elute MHC class I-loaded peptides followed by identification of their sequences by mass spectrometry.

We describe systematic MHC class I immunoprecipitation followed by elution of MHC class I-loaded peptides from 20 breast cancer cell lines belonging to the four major subtypes. We determined the sequence of more than 2,700 unique non-mutated MHC class I-loaded peptides from 1,921 proteins as well as the sequence of 85 mutation-containing peptides and 4 peptides derived from aberrantly translated genes. MHC class I binding probability of eluted peptides was also used to plot the distribution of non-mutated MHC class I allele-specific peptides in accordance with the binding score. Finally, by analyzing the available peptide data set, we determined that more than 1,750 peptides have been identified in previous studies, and of these, sixteen peptides have been shown to be immunogenic.

## Materials and Methods

### Cell lines, antibodies, and reagents

Unless otherwise indicated, breast cancer and non-transformed cell lines used in this study were from American Type Culture Collection (Manassas, VA) (Table 1). Breast cancer T4 cell line and non-transformed S1 cell line were from Dr. Bissell [22]. Unless special medium was required, cells were cultured in DMEM or RPMI supplemented with 10% FBS (DMEM/FBS or RPMI/FBS) and 10 µg/mL gentamicin. Unless otherwise indicated, all reagents were from Sigma (St. Louis, MO). Mouse MHC class I (HLA-A, B, C) antibody (clone W6/32) was from Cedarlane (Burlington, NC). Mouse HLA-A\*02 antibody (clone BB7.2), *N*-octyl-β-D-glucopyranoside, SuperSignal West Pico chemiluminescent substrate, interferon (IFN)-α, IFN-γ, and protein G plus agarose were from Pierce (Rockford, IL). HRP-conjugated donkey anti-mouse IgG and Alexa Fluor 488-conjugated donkey anti-mouse IgG were from Jackson ImmunoResearch (West Grove, PA).

### Western blotting and immunocytochemistry (ICC)

For Western blotting, samples were analyzed by staining with MHC class I (HLA-A, B, C) antibodies (clone W6/32) (1 µg/ml) overnight followed by HRP-conjugated donkey anti-mouse IgG and Pico chemiluminescent substrate. For ICC analysis, cells were washed with PBS and stained for 2 h on ice using MHC class I (HLA-A, B, C) antibodies (clone W6/32) and HLA-A\*02 antibodies (clone BB7.2) (5 µg/mL each) in DMEM/FBS. Cells were then washed with PBS, fixed for 15 min with 4% paraformaldehyde, blocked for 1 h in PBS supplemented with 1% casein, 0.05% Tween 20, and stained with Alexa Fluor 488-conjugated donkey anti-mouse IgG for 2 h. Images were acquired on a fluorescence microscope (Olympus) and quantitated by scanR analysis software (Olympus). Nuclear DNA was stained with 4',6-diamidino-2-phenylindole (DAPI).

### Immunoprecipitation and elution of MHC class I-bound peptides and mass spectrometry (MS) analysis of eluted peptides

MHC class I-bound peptides were isolated according to published protocols [23–25] using mouse anti-human HLA (clone W6/32)-specific antibodies followed by acid treatment and concentration by vacuum centrifugation. Briefly,  $10^8$  cells were pelleted and lysed in 40 ml of PBS containing 50 mM n-octyl-β-d-glucopyranoside, 1 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub>, 1 mM phenylmethylsulfonyl fluoride, 1 µg/ml leupeptin, 1 µg/ml pepstatin, and 1 µg/ml aprotinin for 1 h at 4°C. Insoluble material was removed by centrifugation. Next, supernatants were incubated overnight at 4°C with 200 µg of mouse anti-human HLA antibodies (clone W6/32) and 100 µl of a 50% Protein G plus agarose slurry (Pierce). Following washes, the beads were treated with 0.1% trifluoroacetic acid (TFA) for 30 min and eluted peptides were passed through 10 kDa cut-off filter units (Millipore). Flow through solutions were dried by vacuum centrifugation and dissolved in 5% formic acid. Peptides were separated using liquid chromatography with a nanoAcquity UPLC system (Waters), then delivered to an LTQ Velos linear ion trap mass spectrometer (Thermo Fisher Scientific) using electrospray ionization with a Captive Spray Source (Michrom Biosciences). Samples were applied at 15 µl/min to a Symmetry C18 trap cartridge (Waters) for 10 min, then switched onto a 75 µm × 250 mm nanoAcquity BEH 130 C18 column with 1.7 µm particles (Waters) using mobile phases water (A) and acetonitrile (B) containing 0.1% formic acid, 7–30% acetonitrile gradient over 106 min, and 300 nl/min flow rate. Data-dependent collection of MS/MS spectra used the dynamic exclusion feature of the instrument's control software (repeat count equal to 1, exclusion list size of 500, exclusion duration of 30 sec, and exclusion mass width of -1.0 to +4.0) to obtain MS/MS spectra of the ten most abundant parent ions (minimum signal of 5,000) following each survey scan from m/z 400 to 1,400.

### Mass spectrometry (MS) data analysis

Several releases of the stable human Swiss-Prot canonical protein databases were used (versions 2011.06, 2011.08, 2011.11, 2012.03, 2012.06) with about 20,200 sequences during the course of the project. The database releases were compared and all proteins associated with identified peptides were present in all databases so the results could be safely combined. We used sequence-reversed databases to estimate error thresholds [26]. After adding 179 common contaminant sequences, reversed sequences were constructed and

concatenated for final protein databases of roughly 40,800 sequences. Peptide identification was performed with SEQUEST/PAWS [27], using no enzyme specificity, average parent mass tolerance of 2.5 Da, monoisotopic fragment ion mass tolerance of 1.0 Da, and variable modification of +16 Da on methionine residues with a maximum of 2 modifications per peptide. The data from each cell line corresponded to a single liquid chromatography (LC) MS run. It is difficult to apply target/decoy statistical error control methods when dealing with small datasets. We combined set search score thresholds to exclude the main decoy score distributions similarly across all 22 datasets from all 20 cell lines into combined target and decoy score histograms to establish score thresholds to achieve an overall desired false discovery rate (FDR) of 4.65%. Score thresholds were set independently for peptides of different charge states (1+, 2+, and 3+). The individual sample FDRs varied from the overall FDR depending on the number of peptides present in each sample preparation and ranged from 2.8% to 7.8%. The MS proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRoteomics IDentifications (PRIDE) database partner repository [28] with the dataset identifier PXD006406.

For the identification of recently predicted mutation-containing HLA peptides [29], the Thermo LTQ RAW files were converted to compressed text files using MSConvert of the ProteoWizard (version 3.0.11383) toolkit [30]. Comet (version 2017.01 rev. 2) [31] was used to search a custom FASTA database constructed from predicted HLA peptides [29] to assign peptide spectrum matches (PSMs). The Comet parameters were an average parent ion mass tolerance of 2.5 Da, a monoisotopic fragment ion mass tolerance of 1.0005 Da, no static amino acid modifications, variable oxidation of methionine (M+15.9949 Da), and no enzyme for the protein digestion. The FASTA database was created from the peptides listed in Supplemental Table 4 of [29]. The FASTA accession (key) was constructed by concatenating the cell line name, the protein identifier, and the peptide sequence. The protein sequence was the predicted HLA peptide sequence. There were 40,813 total predicted peptides. Common contaminants (179 sequences) were included, and decoy sequences added by sequence reversal. The final FASTA file had 81,984 sequences.

The Comet search scores were processed with the PAW pipeline [27]. PSMs were separated by cell line (a single mass spectrometer RAW file) and by charge state (1+, 2+, or 3+), and filtered to allow only matches to the full-length predicted HLA peptides (shortened peptide forms are scored during the no enzyme search). The separate PSM lists were then sorted from highest discriminant function score to lowest. Matches to the HLA peptides were accepted until the first decoy match score. The ranking by discriminant score and acceptance filtering were performed with Excel.

To identify altered peptides including peptides with insertions and deletions, we also used the Enosi pipeline [32]. The tool combines sample specific RNA-seq data with RNA-seq data from the TCGA project to create a custom database encoding mutations, splice variants, and other genomic lesions, and uses a two stage FDR controlled pipeline to identify altered peptides that are not part of the global reference.

## HLA typing

**HLA typing**—RNA-seq data were obtained from the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48213>). The HLA type of each cell line was identified by using paired end reads in fastq format as input to the seq2HLA tool [33].

## MHC class I binding predictions and distribution of MHC class I allele-specific peptides

MHC class I binding predictions were made using IEDB analysis resource tool, which combines predictions from the Consensus method [34] consisting of Artificial Neural Network (ANN) [35, 36], Stabilized Matrix Method (SMM) [37], and Scoring Matrices derived from Combinatorial Peptide Libraries (Comlib) [38], and NetMHCpan [39] method. To predict peptide:MHC-I binding affinity, we used the Consensus method consisting of ANN, SMM, and Comlib if any corresponding predictor is available for the molecule. Otherwise, NetMHCpan is used. Based on availability of predictors and previously observed predictive performance, this selection tries to use the best possible method for a given MHC molecule. Thus, multiple methods were used to predict peptide binding affinity. If two methods predicted the binding affinity of the same peptide, the consensus value was taken.

## Results

### ICC and RNA-seq analyses of breast cancer cell lines

In addition to MHC class I-positive breast cancer cells, we sought to identify HLA-A\*02-positive cell lines because the HLA-A\*02 allele occurs frequently in all ethnic groups. HLA-A\*02 has been identified in 35% of African-Americans and in 50% of Caucasians [40]. Each cell line was stained with the mouse anti-human HLA (clone W6/32) and HLA-A\*02-specific (clone BB7.2) antibody followed by Alexa Fluor 488-conjugated donkey anti-mouse IgG. Control staining was performed with non-specific mouse IgG antibodies. Representative images of the stained MCF7 and MDA-MB-231 cells are shown in Figure 1A. As expected, staining with control mouse IgG showed no signal. Typical quantitative data of HLA-A\*02 for MCF7 and MDA-MB-231 is shown in Figure 1B. Quantitative levels of total MHC class I and HLA-A\*02 expression in all analyzed breast cancer cell lines are shown in Supporting Information Figure S1.

The level of MHC class I and HLA-A\*02 expression in MDA-MB-231 cells was arbitrarily set to 100%, and the expression level of MHC class I and HLA-A\*02 in other lines was calculated as a percentage of MDA-MB-231 staining (Table 1). Cell lines were grouped by subtypes and by their measured HLA-A\*02 protein throughout the manuscript. MHC class I status in selected cell lines was also determined using RNA-seq analysis. The data showed that, in most cases, MHC class I staining (protein expression) did not correlate with the level of MHC class I mRNA. In addition, MHC class I staining was below the limit of detection in MDA-MB-157, JIMT1, CAL-51, ZR75B, SKBR3, and BT474 cells despite the fact that MHC class I mRNA was present in these cells.

We identified MHC class I alleles in the breast cancer cell lines by RNA-seq using seq2HLA [33] to map RNA-seq reads against a reference database of HLA alleles [29] (Table 1). Our HLA typing agrees with the data published by Boegel et al [29]; however, there is a discrepancy between our data and the data used by the TRON Cell Line Portal (<http://cellines.tron-mainz.de/>) with regard to MCF-7 cell line. This discrepancy could arise if there were multiple, genetically unique subpopulations within the MCF-7 breast cancer cell line as described by Tan et al. [42]. The validity of our HLA typing was also confirmed by the observation of the same HLA-typing profile in LY2 cells, as both MCF7 and LY2 cell lines were derived from the same patient [43]. With regard to HLA-A\*02 allele, there was general agreement between protein expression and genotype; however, some cell lines were phenotypically HLA-A\*02-negative even though HLA-A\*02 mRNA was present (Table 1).

### MHC class I immunoprecipitation and elution of MHC class I-loaded peptides

To identify MHC class I-restricted peptides expressed on the cell surface, MHC molecules immunoprecipitated as described in Methods were analyzed by Western blotting using anti-MHC class I antibodies (Supporting Information Figure S2). The major purpose of this Western blotting analysis was to confirm the robustness of our protocol to elute MHC class I ligands from the cell surface of breast cancer cells. The obtained results with MDA-MB-231 cells showed that the selected antibodies were able to immunoprecipitate MHC class I molecules. In addition, we found that the selected antibodies were not able to deplete MHC class I molecules from the lysate of approximately  $10^8$  cells. Because the total level of MHC class I expression determined by immunocytochemistry is comparable in most analyzed cell lines (Table 1), we concluded that  $10^8$  cells expressed enough MHC class I molecules to saturate the used amount of MHC class I antibodies.

Flow through solutions of eluates from control IgG and anti-MHC class I samples were subjected to MS analysis. IgG control sample from HCC1187 cells contained few peptides (Supporting Information Figure S3). In contrast, anti-MHC class I sample from HCC187 cells contained several hundred peptides. Immunoprecipitation and elution of peptides were repeated for all selected cell lines (Table 1) in independent experiments.

We used the PAW processing pipeline [27] to identify cell surface peptides and control peptide false discovery rates. Peptide summary reports are presented in Supporting Information Table S1. Peptides that map to a single protein are indicated as unique (TRUE in “Unique” column), while redundant peptides are indicated as non-unique (FALSE in “Unique” column). Common contaminating peptides and peptides eluted from both the control IgG agarose and empty agarose were removed from the Table S1. In addition, we found that some peptides were derived from MHC class I proteins themselves. We removed these peptides from our dataset because we could not exclude the possibility that these peptides were non-specifically captured due to the degradation of MHC class I proteins during immunoprecipitation and acid treatment. Indeed, we identified a few mouse IgG-derived peptides in the eluted samples that were non-specifically captured during the treatment. Peptides and corresponding proteins were arranged according to the frequency that these peptides were observed (MS/MS total spectral counts). The total numbers of identified peptide spectral counts from target and decoy databases matches are also tallied in



a tab of Supporting Information Table S1. Peptide FDR were estimated from decoy (reversed) protein entries [26].

The data in the Table 1 suggest that high levels of cell surface MHC class I protein expression do not guarantee high levels of peptide loading. For example, in three independent experiments, only small number of peptides was recovered from MDA-MB-231 cells, which expressed high levels of MHC class I. We did not identify any shared peptides between these three peptide identifications most likely because there was a large antigen excess as compared to the ability of antibodies to precipitate it (Figure S2). In addition, we assume that MDA-MB-231 cells have impaired MHC class I function. In support of our conclusion, the available RNAseq data showed that MDA-MB-231 cells harbor mutation-containing genes *bach1*, *cpsf1*, and *ier5*, which are directly involved in antigen processing and presentation [44]. Mutations in these genes most likely result in defects in the expression of MHC class I-loaded peptides on the cell surface. It implies that MDA-MB-231 cells mostly express MHC class I proteins without any loaded peptides like T2 cells that have defects in transporter associated with antigen processing (TAP) proteins. Thus, it would be hard to identify shared peptides that are presented *via* defected MHC class I presentation machinery. A total of 3,196 peptides were eluted from MHC class I molecules immunoprecipitated from the surface of breast cancer cells belonging to the 20 cell lines. After removal of duplicate peptides, i.e., the same peptide eluted from more than one cell line, the total number of unique peptides and corresponding proteins was 2,740 and 1,921, respectively. The identity and frequency of shared peptides in all cell lines are presented in the Supporting Information Table S2. We also included in the Supporting information Table S2 gene expression data of all corresponding proteins from which eluted peptides were derived.

Next, we asked if the eluted peptides contain antigenic mutations, which are primary targets for T cell response. For this purpose, our raw MS data was screened against a catalog of predicted mutation-containing HLA peptides (Supporting Information Table S3) [29]. We determined that among eluted peptides there are 85 peptides that contain mutation in their sequence. In addition, we used the Enosi pipeline, which can detect peptides derived from mutated proteins and from either proteins with deletions and insertions or aberrantly translated genes using RNA-seq data from the TCGA project [32]. We identified two peptides derived from reverse translated genes *tap1* (ATAPGLGGGPEPLGR) and *ikbkap* (EIISDPGVQGYSR) in HCC1806 cells as well as one peptide derived from translations of a +1 frame-shifted gene *cp4* (AVASINSSEALR) in UACC812 cells and one peptide derived +2 frame-shifted gene *clipr1* (TAFESITSSDQR) in HCC1806 cells (Supporting Information Table S4).

### Data mining and validation

Analysis of our data showed that some peptides were frequently presented on the cell surface of different breast cancer cells (Supporting Information Table S2). As can be expected, peptides derived from proteins that are expressed at high levels in cancer and normal cells such as elongation factor 2 (EEF2), fructose-bisphosphate aldolase A (ALDOA), E3 ubiquitin-protein ligase RNF213 (RNF213), cytoplasmic dynein 1 heavy

chain 1 (DYHC1), helicase with zinc finger domain 2 (HELZ2), and eukaryotic translation initiation factor 3 subunit D (EIF3D) were most frequently presented in the context of MHC class I. However, many of the peptides were derived from low abundance proteins, furthermore, a subset of those had corresponding mRNA levels below the threshold of detection. Even mutation-containing peptides were derived from genes that have both high and low level of expression in the analyzed breast cancer cells. The presence of peptides from low-abundance genes implies some selectivity process. It is noteworthy that identical peptides were eluted from the cell surface of different cell lines confirming the reproducibility of the obtained results. To further elaborate on this issue, we identified peptides either specific or shared between different breast cancer subtypes, which could be of interest to design therapies targeted to specific breast tumor types (Supporting Information Table S5).

To determine if the eluted MHC class I-bound peptides had been identified in previous studies, we searched IEDB, which consists of more than 167,000 human unique peptides. IEDB is sponsored by the National Institute for Allergy and Infectious Diseases (NIAID) who launched a large-scale antibody and T cell epitope discovery program. Data sources to be integrated into the IEDB include publications in peer-reviewed journals, published patents, and direct submissions from institutions or companies. Each publication can also be classified by its general topic (e.g. infectious diseases, autoimmunity, allergy, transplantation, HIV or cancer). Thus, IEDB includes epitopes from a variety of sources and constitutes the most comprehensive database/resource for HLA ligands [45]. We determined that, of the 2,740 eluted unique peptides, 1,751 peptides have previously been shown to bind MHC class I proteins. Importantly, 16 of the peptides in our data set were active in T cell activation assays published earlier (Table 2) and half of these immunogenic peptides were shared between different breast cancer cell lines. Surprisingly, we found that these immunogenic epitopes were seldom related to breast cancer, and one of them was previously identified as a target of T cells in chronic lymphocytic leukemia (CLL) patients [57].

To find new breast cancer specific MHC class I-loaded epitopes that could have the ability to activate T cell response, we used gene expression profiling to determine the MHC class I-presented genes that have alterations or elevated expression levels in breast tumors compared to normal cells. First, we determined genes whose expression is altered in invasive breast cancers by copy number amplification, homozygous deletion, mRNA upregulation or downregulation, and mutation using the cBioPortal for Cancer Genomics that contains large-scale cancer genomics data sets (Supporting Information Table S6). We arranged all identified genes in accordance with the frequency of alterations in breast cancer samples. For further analysis we selected genes that have alterations in at least 20% of breast cancers.

Second, we used gene expression data from of 708 breast tumors and 329 normal tissues from TCGA (20), EBI (21), and GEO (22) to identify among MHC I-presented genes those genes that have preferential expression in breast cancer samples over normal samples. We averaged expression among all tumor and normal samples for each gene and arranged all genes by the level of differential expression in tumor and normal samples (Supporting Information Table S6). In this analysis we selected genes that have at least 2 times higher expression in breast tumors than in normal tissues. We also attempted to select breast cancer

specific candidate genes using gene expression data sets based on RNA-seq data for 62 breast carcinoma cell lines and 6 non-transformed cell lines (Supporting Information Table S6). We averaged expression data for each gene across all breast carcinoma cell lines and non-transformed cells and for further analysis we selected genes that have 2 times higher expression in transformed than in non-transformed cells. Interestingly, most genes that were overexpressed in breast tumors as compared to normal tissues do not show a similar overexpression pattern in breast carcinoma cell lines as compared to non-transformed cells.

Next, we calculated the immunogenicity score for this set of proteins that were associated with cancer and compared it with that of the known 1217 immunogenic epitopes present in the IEDB database. To determine the immunogenicity score, we employed a tool that was used to predict immunogenicity of viral MHC class I ligands [59]. Previous studies demonstrated that this tool can be successfully used to predict immunogenicity of cancer-related peptides [60]. We find that our cancer-related dataset contains a number of peptides that have score higher than 0.20, which is close to the top 20% most immunogenic known epitopes (Supporting Information Table S6). These peptides are good candidates for further testing in T cell-based immunogenicity assays.

The availability of comprehensive HLA-A, B, and C typing also allowed us to predict the binding probability of each peptide to the HLA molecules present in the corresponding cells including mutation-containing peptides and peptides derived from aberrantly translated genes (Supporting Information Table S7). For this purpose, we used a Consensus method [34] recommended by IEDB, which consists of ANN [35, 36], SMM [37], Comblib [38], and NetMHCpan [39]. It has been shown that the consensus method is superior to any single prediction technique tested [61]. The predicted affinity of each peptide was expressed as a percentile rank, generated by comparing the peptide IC<sub>50</sub> value in relation to the corresponding HLA allele against those of a set of random peptides from the SWISSPROT database. Lower rank values indicate higher predicted peptide affinities. Next, we plotted the number of peptides (counts) in relation to these percentile ranks for each HLA allele (Figure 2). For strong binders, we selected peptides with rank ranging from 0.1 to 1. Our data showed that UACC812, CAMA-1, SUM185PE, HCC1806, and T47D cells use only one HLA allele to present almost all high affinity peptides (rank 0.1–1), while other cells use two or more HLA alleles to load high affinity peptides. To distinguish binders from non-binders, we used rank 10. Based on these selection criteria, we determined the number and ratio of peptides that are not predicted to bind any HLA allele. The ratio of non-binders was in the range of FDR (Supporting Information Table 1) thereby confirming the validity of rank 10 to distinguish between binders and non-binders. We also determined the quantity of shared high affinity peptides between HLA alleles in corresponding cells and ratio of these shared peptides in relation to all eluted peptides (Figure 2). Our data showed that some high affinity peptides are shared between different HLA alleles while other peptides are not.

## Discussions

We selected breast cancer for these studies because breast cancer is a global problem. Worldwide, approximately a quarter of all cancer cases in women are breast cancers. In addition, as a model of breast cancer we have approximately 70 breast cancer cell lines,

which are the focus of intensive molecular and phenotypic characterization [22, 43, 62]. Efforts to develop a breast cancer vaccine have largely focused on eliciting immune responses to HER2 [63–65]. It is unclear, however, whether HER2-derived epitopes are the most potent antigens for HER2-overexpressing breast cancer cells. Identification of new and potent T cell-activating epitopes expressed on the surface of cancer cells remains critical for development of successful immunotherapies. In this paper, we have demonstrated that established cancer cell lines can be used as a source to identify MHC class I-restricted and immunogenic epitopes. The previously identified immunogenic peptides that were present in our dataset do not harbor any mutations in their sequences, thereby their ability to activate T cells would be a result of alterations in either the expression or processing pattern as compared to normal cells.

We identified 2,740 unique peptides eluted from breast cancer cells and 1,921 corresponding proteins. Actually, few papers have described surveys of tumor-associated immunopeptidomes with identifications of more than 100 HLA-bound peptides [60, 66, 67]. For example, Bassani-Sternberg et al. [60] developed a high-throughput MS-based workflow that yielded more than 22,000 unique HLA peptides across seven cancer cell lines and primary cells. Authors used a similar approach, however, with 5 times more cells than we used in the immunoprecipitation experiments. Another distinction from our approach is the covalent binding of W6/32 antibody to Protein-A Sepharose beads that could improve the yield and purity of the eluted HLA-I complexes. We believe that the increased number of cells and use W6/32 antibody covalently bound to the beads resulted in approximately 10-fold improvement in the yield of the eluted peptides as compared to our protocol. On the other hand, the dataset published by Bassani-Sternberg et al. [60] contained only four previously identified immunogenic epitopes while our dataset contains sixteen immunogenic epitopes identified in previous studies. This 4-fold difference in the number of immunogenic epitopes identified in our studies and in the work of Bassani-Sternberg et al. [60] can be result of either type of cancer cells used in the MHC class I immunoprecipitation and epitope elution experiments or difference in the experimental techniques. Most other related works identified far fewer MHC class I-bound ligands. One of the explanations for the inability to elute a large number of peptides from the cell surface can be related to downregulation of HLA ligand presentation machinery in cancer cells. There are many different ways used by cancer cells to escape immune attack. One of them is to reduce the cell's ability to present HLA ligands on the cell surface. For example, our data and results published by others showed that approximately 40% of breast cancers lost MHC class I RNA expression [68]. MHC class I expression can also be downregulated at the post transcriptional level. In agreement, Table 1 showed that expression of MHC class I RNA does not always correlate with that of MHC class I protein abundance. In addition, HLA ligand presentation can be downregulated by reduction/loss of TAP expression that occur in 29% of primary and 42% of metastatic breast cancers [68]. The available RNA-seq and immunohistochemistry data allowed us to relate HLA expression with the quantity of eluted peptides. RNA-seq and HLA expression data do not always correlate with the number of eluted peptides. Thus, we conclude that RNA-seq and protein expression data cannot be used to estimate the level of HLA-loaded peptides on the cell surface and additional factors should be taken into account.

Our data suggest that in most cases the identified peptides were expressed at high or moderate levels on the cell surface. This observation is in agreement with the recently published results of Bassani-Sternberg et al. [60], who demonstrated a strong correlation between protein abundance and HLA-presentation. We have not seen any depletion of antigen after immunoprecipitation, and more frequent MHC class I/peptide complexes could occupy all antigen binding sites of the antibodies used for immunoprecipitations. The fact that only stable MHC class I/peptide complexes are efficiently transported to the extracellular membrane and have an extended half-life on the cell surface indicates that the level of peptide expression is closely linked to peptide affinity for the presenting MHC class I molecule [69]. In agreement, plotting the number of peptides in relation to percentile ranks demonstrated that the ratio of high affinity peptides (rank 0.1–1) is higher than that of lower affinity peptides for most HLA alleles which are preferentially used to present peptides. It is important; however, to identify low affinity peptides, which can be immunogenic [70]. There are some approaches to address this issue. The first approach is to increase the antibody/antigen ratio and couple the antibodies to a solid support [60]. A second approach, far less specific, is to use a lectin-coated resin to capture all glycosylated MHC class I molecules [71]. A third, non-specific, approach is to elute peptides directly from cells as has been described previously [67, 72]. This technique is especially useful for clinical samples where tumor material is often available in only very small quantities.

Recently, Boegel et al. [29] showed that, on average, breast carcinoma cells harbor approximately 300 mutation-containing genes per genome, and the authors predicted that 60–70% of the corresponding mutation-containing peptides can be efficiently presented by MHC class I molecules present in the same cells. To confirm that MHC class I molecules are loaded with mutation-containing peptides, we searched our raw MS data against database of the predicted antigenic mutations [29]. Interestingly, we determined that our eluted peptide pool includes 85 mutation-containing peptides, and 7 of them were shared between different breast cancer cell lines. These shared mutation-containing peptides could provide a basis to design a multi-peptide cancer vaccine. To identify peptides that are not present in normal cells, we also employed an algorithm specifically designed to identify altered peptides [32]. We identified four non-normal peptides that were derived from alternative translations of genes, including *tap1* (reverse frame translation), *c4a* (+1 frame shifted translation), *ikbkap* (reverse frame translation), and *glpr1* (+2 frame shifted translation). Altogether, our data indicated that altered or mutation-containing peptides are presented at the level of 2–3% on the cell surface of breast cancer cells. Our data also suggest that it is feasible to identify immunogenic epitopes derived from wild type genes expressed in breast cancer cells. Importantly, our data also showed that either non-mutated immunogenic HLA ligands or mutation-containing neo-epitopes can be shared between different cancer cells that do not have identical HLA alleles. The ideal candidate epitopes for vaccine development will be those that are commonly presented by tumor cells. Our data also confirmed that immunogenic epitopes can be shared between different cancer types including leukemia and epithelial cancers.

The knowledge of HLA genotype and probability of binding of each peptide to MHC class I present in the corresponding cells allowed us to plot the distribution of ranked MHC class I allele-specific peptides for each breast cancer cell line. Our data showed that breast cancer

cells use different quantity of MHC class I alleles to present high affinity peptides (rank 0.1–1.0). For example, HCC1806, CAMA-1, SUM185PE, T47D, and UACC812 cells use only one MHC class I allele to present almost all high affinity peptides, while other cells use two or more MHC class I alleles to present high affinity peptides. Based on the inability of many MHC class I alleles to efficiently bind and present high affinity peptides, we propose that during breast cancer progression malignant cells lose MHC class I functionality to interfere with peptide presentation machinery and, as a consequence, avoid immune system attack. Our conclusion that interfering with MHC class I functionality can be used by cancer cells to escape the immune attack is supported by the published data showing that approximately 40% of breast cancers lost MHC class I RNA expression [68].

The current work was driven mostly by our optimized technique to elute MHC class I-loaded peptides from different breast cancer cells with an aim to build a potential HLA-I-typed antigen collection. We understand that a limitation of our work is the absence of immunogenicity analysis of the eluted peptides. We proposed that the gap between the basic research and clinical utility of the project may be effectively closed by developing a high throughput protocol to identify immunogenic epitopes among eluted peptides. To accomplish this task, we developed a technique to attach a patient's own PBMC-derived antigen-presenting dendritic cells (DCs) to the array matrix. The attached DCs can be matured to present pre-printed peptides on the cell surface (Fig. 3). Subsequent ELISPOT assays using a patient's own T cells will identify peptides that can induce the T cell-based immune response.

We envision that future design of personalized vaccines will be based on the HLA allele subtypes of the individual patient and the ability of these HLA allele subtypes to present MHC class I-restricted peptides. Distribution profiles of presentation of known immunogenic epitopes by each MHC class I allele type will likely inform the design of multi-peptide-based vaccines. In support of these conclusions, in a recent report, nine peptides were used as the basis for a multi-peptide vaccine against renal cell carcinoma cells [18].

## Conclusions

In summary and in agreement with previous studies [60], our results demonstrate that established cancer cell lines can be used to identify tumor-specific and immunogenic MHC class I-restricted peptides. Because it is extremely difficult in most cases to obtain enough fresh tumor for efficient immunoprecipitation and MS analysis of eluted peptides, we used a more feasible cell line-based approach to elute peptides from the cell surface followed by sequence determination using MS.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This study was supported by the U.S. Department of Defense Award W81XWH-11-1-0549. Mass spectrometric analysis was performed by the OHSU Proteomics Shared Resource with partial support from NIH core grants P30EY010572 and P30CA069533, and equipment grants S10RR025571 and R01DC002368-15S1. We thank

consumer advocates Joan Venticinque, Susie Brain, and Carolyn Charkey for assistance in manuscript preparation and for valuable comments. We thank consumer advocates Joan Venticinque, Susie Brain, and Carolyn Charkey for assistance in manuscript preparation and for valuable comments.

## References

1. Ehemann C, Henley SJ, Ballard-Barbash R, Jacobs EJ, Schymura MJ, Noone AM, et al. Annual Report to the Nation on the status of cancer, 1975–2008, featuring cancers associated with excess weight and lack of sufficient physical activity. *Cancer*. 2012; 118:2338–2366. [PubMed: 22460733]
2. Hanna YM, Baglan KL, Stromberg JS, Vicini FA, Decker DA. Acute and subacute toxicity associated with concurrent adjuvant radiation therapy and paclitaxel in primary breast cancer therapy. *Breast J*. 2002; 8:149–153. [PubMed: 12047471]
3. Henry NL, Skaar TC, Dantzer J, Li L, Kidwell K, Gersch C, et al. Genetic associations with toxicity-related discontinuation of aromatase inhibitor therapy for breast cancer. *Breast Cancer Res Treat*. 2013; 138:807–816. [PubMed: 23546553]
4. Darby SC, McGale P, Taylor CW, Peto R. Long-term mortality from heart disease and lung cancer after radiotherapy for early breast cancer: prospective cohort study of about 300,000 women in US SEER cancer registries. *Lancet Oncol*. 2005; 6:557–565. [PubMed: 16054566]
5. Jähnisch H, Füssel S, Kiessling A, Wehner R, Zastrow S, Bachmann M, et al. Dendritic cell-based immunotherapy for prostate cancer. *Clin Dev Immunol*. 2010; 2010:517493. [PubMed: 21076523]
6. Emens LA. Breast cancer immunobiology driving immunotherapy: vaccines and immune checkpoint blockade. *Expert Rev Anticancer Ther*. 2012; 12:1597–611. [PubMed: 23253225]
7. Sørli T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001; 98:10869–10874. [PubMed: 11553815]
8. Weigelt B, Mackay A, A'hern R, Natrajan R, Tan DS, Dowsett M, et al. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol*. 2010; 11:339–349. [PubMed: 20181526]
9. Wainberg ZA, Anghel A, Desai AJ, Ayala R, Luo T, Safran B, et al. Lapatinib, a dual EGFR and HER2 kinase inhibitor, selectively inhibits HER2-amplified human gastric cancer cells and is synergistic with trastuzumab in vitro and *in vivo*. *Clin Cancer Res*. 2010; 16:1509–1519. [PubMed: 20179222]
10. Colleoni M, Cole BF, Viale G, Regan MM, Price KN, Maiorano E, et al. Classical cyclophosphamide, methotrexate, and fluorouracil chemotherapy is more effective in triple-negative, node-negative breast cancer: results from two randomized trials of adjuvant chemoendocrine therapy for node-negative breast cancer. *J Clin Oncol*. 2010; 28:2966–2973. [PubMed: 20458051]
11. Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, et al. Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res Treat*. 2013; 142:237–55. [PubMed: 24162158]
12. Peoples GE, Goedegebuure PS, Smith R, Linehan DC, Yoshino I, Eberlein TJ. Breast and ovarian cancer-specific cytotoxic T lymphocytes recognize the same HER2/neu-derived peptide. *Proc Natl Acad Sci U S A*. 1995; 92:432–6. [PubMed: 7831305]
13. Jaramillo A, Majumder K, Manna PP, Fleming TP, Doherty G, Dipersio JF, et al. Identification of HLA-A3-restricted CD8+ T cell epitopes derived from mammaglobin-A, a tumor-associated antigen of human breast cancer. *Int J Cancer*. 2002; 102:499–506. [PubMed: 12432553]
14. Müller MR, Grünebach F, Nencioni A, Brossart P. Transfection of dendritic cells with RNA induces CD4- and CD8-mediated T cell immunity against breast carcinomas and reveals the immunodominance of presented T cell epitopes. *J Immunol*. 2003; 170:5892–6. [PubMed: 12794115]
15. Wang W, Epler J, Salazar LG, Riddell SR. Recognition of breast cancer cells by CD8+ cytotoxic T-cell clones specific for NY-BR-1. *Cancer Res*. 2006; 66:6826–33. [PubMed: 16818660]
16. Schultze JL, Vonderheide RH. From cancer genomics to cancer immunotherapy: toward second-generation tumor antigens. *Trends Immunol*. 2001; 22:516–23. [PubMed: 11525943]

17. Weinschenk T, Gouttefangeas C, Schirle M, Obermayr F, Walter S, Schoor O, et al. Integrated functional genomics approach for the design of patient-individual antitumor vaccines. *Cancer Res.* 2002; 62:5818–27. [PubMed: 12384544]
18. Walter S, Weinschenk T, Stenzl A, Zdrojowy R, Pluzanska A, Szczylik C, et al. Multi-peptide immune response to cancer vaccine IMA901 after single-dose cyclophosphamide associates with longer patient survival. *Nat Med.* 2012; 18:1254–61. [PubMed: 22842478]
19. Backert L, Kohlbacher O. Immunoinformatics and epitope prediction in the age of genomic medicine. *Genome Med.* 2015; 7:119. [PubMed: 26589500]
20. Fortier M, Caron E, Hardy MP, Voisin G, Lemieux S, Perreault C, et al. The MHC class I peptide repertoire is molded by the transcriptome. *J Exp Med.* 2008; 205:595–610. [PubMed: 18299400]
21. Comber JD, Philip R. MHC class I antigen presentation and implications for developing a new generation of therapeutic vaccines. *Ther Adv Vaccines.* 2014; 2:77–89. [PubMed: 24790732]
22. Kenny PA, Lee GY, Myers CA, Neve RM, Semeiks JR, Spellman PT, et al. The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. *Mol Oncol.* 2007; 1:84–96. [PubMed: 18516279]
23. Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature.* 1991; 351:290–6. [PubMed: 1709722]
24. Seeger FH, Schirle M, Gatfield J, Arnold D, Keilholz W, Nickolaus P, et al. The HLA-A\*6601 peptide motif: prediction by pocket structure and verification by peptide analysis. *Immunogenetics.* 1999; 49:571–6. [PubMed: 10380706]
25. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, et al. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science.* 1992; 255:1261–3. [PubMed: 1546328]
26. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods.* 2007; 4:207–14. [PubMed: 17327847]
27. Wilmarth PA, Riviere MA, David LL. Techniques for accurate protein identification in shotgun proteomic studies of human, mouse, bovine, and chicken lenses. *J Ocul Biol Dis Infor.* 2009; 2:223–34. [PubMed: 20157357]
28. Vizcaíno JA, Côté RG, Csordas A, Dienes JA, Fabregat A, Foster JM, et al. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 2013; 41(D1):D1063–9. [PubMed: 23203882]
29. Boegel S, Löwer M, Bukur T, Sahin U, Castle JC. A catalog of HLA type, HLA expression, and neo-epitope candidates in human cancer cell lines. *Oncoimmunology.* 2014; 3:e954893. [PubMed: 25960936]
30. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K. A cross-platform toolkit for mass spectrometry and proteomics. *Nature biotechnology.* 2012; 30(10):918–920.
31. Eng JK, Jahan TA, Hoopmann MR. Comet: An open-source MS/MS sequence database search tool. *Proteomics.* 2013; 13(1):22–24.30. [PubMed: 23148064]
32. Woo S, Cha SW, Na S, Guest C, Liu T, Smith RD, et al. Proteogenomic strategies for identification of aberrant cancer peptides using large-scale next-generation sequencing data. *Proteomics.* 2014; 14:2719–30. [PubMed: 25263569]
33. Boegel S, Löwer M, Schäfer M, Bukur T, de Graaf J, Boisguérin V, et al. HLA typing from RNA-Seq sequence reads. *Genome Med.* 2012; 4:102. [PubMed: 23259685]
34. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, et al. Immune epitope database analysis resource. *Nucleic Acids Res.* 2012; 40(Web Server issue):W525–30. [PubMed: 22610854]
35. Nielsen M, Lundegaard C, Wornig P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* 2003; 12:1007–17. [PubMed: 12717023]
36. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res.* 2008; 36(Web Server issue):W509–12. [PubMed: 18463140]

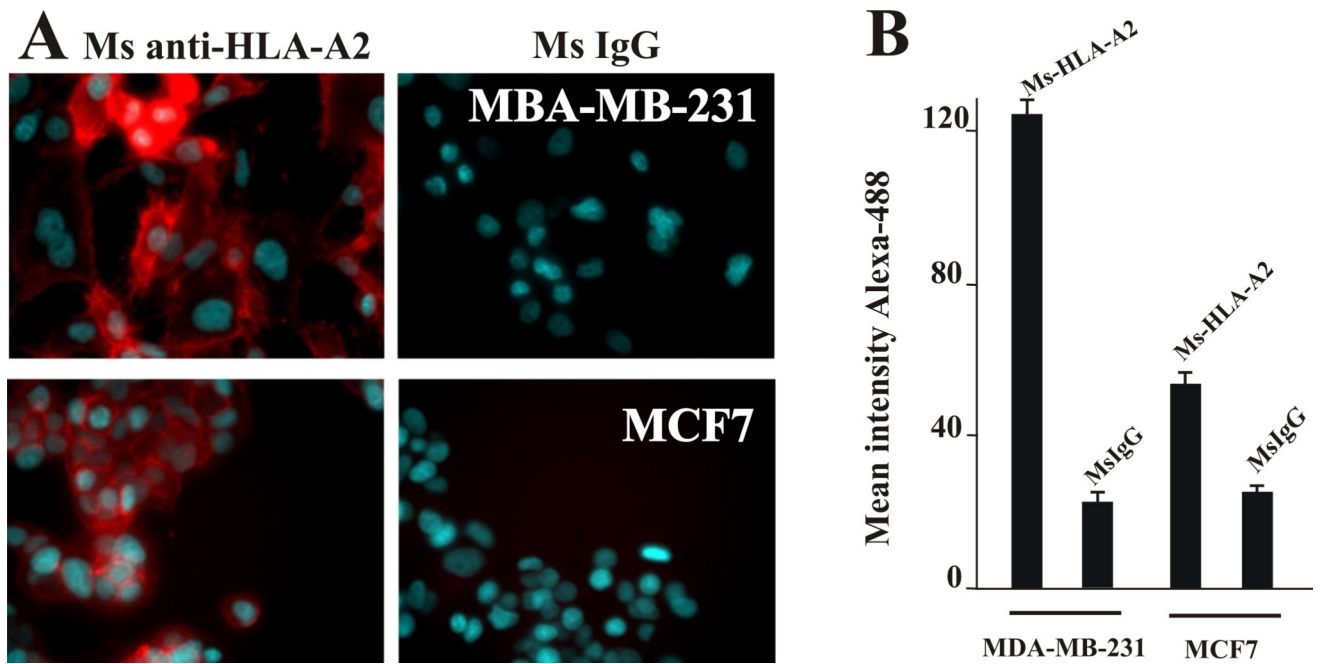


37. Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*. 2005; 6:132. [PubMed: 15927070]
38. Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, et al. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Res*. 2008; 4:2. [PubMed: 18221540]
39. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*. 2009; 61:1–13. [PubMed: 19002680]
40. Ellis JM, Henson V, Slack R, Ng J, Hartzman RJ, Katovich Hurley C. Frequencies of HLA-A2 alleles in five U.S. population groups. Predominance of A\*02011 and identification of HLA-A\*0231. *Hum Immunol*. 2000; 61:334–40. [PubMed: 10689125]
41. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010; 28:511–15. [PubMed: 20436464]
42. Tan Z, Nie S, McDermott SP, Wicha MS, Lubman DM. Single Amino Acid Variant Profiles of Subpopulations in the MCF-7 Breast Cancer Cell Line. *J Proteome Res*. 2017; 16:842–51. [PubMed: 28076950]
43. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell*. 2006; 10:515–27. [PubMed: 17157791]
44. <http://immunet.princeton.edu/>
45. Salimi N, Fleri W, Peters B, Sette A. The immune epitope database: a historical retrospective of the first decade. *Immunology*. 2012; 137:117–23. [PubMed: 22681406]
46. Barnea E, Beer I, Patoka R, Ziv T, Kessler O, Tzehoval E, et al. Analysis of endogenous peptides bound by soluble MHC class I molecules: a novel approach for identifying tumor-specific antigens. *Eur J Immunol*. 2002; 32:213–22. [PubMed: 11782012]
47. Flad T, Spengler B, Kalbacher H, Brossart P, Baier D, Kaufmann R, et al. Direct identification of major histocompatibility complex class I-bound tumor-associated peptide antigens of a renal carcinoma cell line by a novel mass spectrometric method. *Cancer Res*. 1998; 58:5803–11. [PubMed: 9865739]
48. Wei CY, Chung WH, Huang HW, Chen YT, Hung SI. Direct interaction between HLA-B and carbamazepine activates T cells in patients with Stevens-Johnson syndrome. *J Allergy Clin Immunol*. 2012; 129:1562–9.e5. [PubMed: 22322005]
49. Krausa P, Münz C, Keilholz W, Stevanovic S, Jones EY, Browning M, et al. Definition of peptide binding motifs amongst the HLA-A\*30 allelic group. *Tissue Antigens*. 2000; 56:10–18. [PubMed: 10958351]
50. Barber LD, Percival L, Valiante NM, Chen L, Lee C, Gumperz JE, et al. The inter-locus recombinant HLA-B\*4601 has high selectivity in peptide binding and functions characteristic of HLA-C. *J Exp Med*. 1996; 184:735–40. [PubMed: 8760827]
51. Barber LD, Percival L, Arnett KL, Gumperz JE, Chen L, Parham P. Polymorphism in the alpha 1 helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity. *J Immunol*. 1997; 158:1660–9. [PubMed: 9029102]
52. Mahler M, Mierau R, Schlumberger W, Blüthner M. A population of autoantibodies against a centromere-associated protein A major epitope motif cross-reacts with related cryptic epitopes on other nuclear autoantigens and on the Epstein-Barr nuclear antigen 1. *J Mol Med (Berl)*. 2001; 79:722–31. [PubMed: 11862315]
53. Whitelegg AM, Oosten LE, Jordan S, Kester M, van Halteren AG, Madrigal JA, et al. Investigation of peptide involvement in T cell allorecognition using recombinant HLA class I multimers. *J Immunol*. 2005; 175:1706–14. [PubMed: 16034111]
54. Wei ML, Cresswell P. HLA-A2 molecules in an antigen-processing mutant cell contain signal sequence-derived peptides. *Nature*. 1992; 356:443–6. [PubMed: 1557127]
55. Henderson RA, Michel H, Sakaguchi K, Shabanowitz J, Appella E, Hunt DF, et al. HLA-A2.1-associated peptides from a mutant cell line: a second pathway of antigen presentation. *Science*. 1992; 255:1264–6. [PubMed: 1546329]

56. Alvarez I, Carrascal M, Canals F, Muixí L, Abián J, Jaraquemada D. Analysis of the HLA class I associated peptide repertoire in a hepatocellular carcinoma cell line reveals tumor-specific peptides as putative targets for immunotherapy. *Proteomics Clin Appl*. 2007; 1:286–98. [PubMed: 21136679]
57. Kowalewski DJ, Schuster H, Backert L, Berlin C, Kahn S, Kanz L, et al. HLA ligandome analysis identifies the underlying specificities of spontaneous antileukemia immune responses in chronic lymphocytic leukemia (CLL). *Proc Natl Acad Sci U S A*. 2015; 112:E166–75. [PubMed: 25548167]
58. Jarmalavicius S, Welte Y, Walden P. High immunogenicity of the human leukocyte antigen peptidomes of melanoma tumor cells. *J Biol Chem*. 2012; 287:33401–11. [PubMed: 22869377]
59. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol*. 2013; 9:e1003266. [PubMed: 24204222]
60. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics*. 2015; 14:658–73. [PubMed: 25576301]
61. Moutaftsi M, Peters B, Pasquetto V, Tschärke DC, Sidney J, Bui HH, et al. A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat Biotechnol*. 2006; 24:817–9. [PubMed: 16767078]
62. Heiser LM, Sadanandam A, Kuo WL, Benz SC, Goldstein TC, Ng S, et al. Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc Natl Acad Sci U S A*. 2012; 109:2724–9. [PubMed: 22003129]
63. Disis ML, Wallace DR, Gooley TA, Dang Y, Slota M, Lu H, et al. Concurrent trastuzumab and HER2/neu-specific vaccination in patients with metastatic breast cancer. *J Clin Oncol*. 2009; 27:4685–92. [PubMed: 19720923]
64. Patil R, Clifton GT, Holmes JP, Amin A, Carmichael MG, Gates JD, et al. Clinical and immunologic responses of HLA-A3+ breast cancer patients vaccinated with the HER2/neu-derived peptide vaccine E75 in a phase I/II clinical trial. *J Am Coll Surg*. 2010; 210:140–7. [PubMed: 20113933]
65. Wiedermann U, Wiltschke C, Jasinska J, Kundi M, Zurbriggen R, Garner-Spitzer E, et al. A virosomal formulated Her-2/neu multi-peptide vaccine induces Her-2/neu-specific immune responses in patients with metastatic breast cancer: a phase I study. *Breast Cancer Res Treat*. 2010; 119:673–83. [PubMed: 20092022]
66. Kirner A, Mayer-Mokler A, Reinhardt C. IMA901: a multi-peptide cancer vaccine for treatment of renal cell cancer. *Hum Vaccin Immunother*. 2014; 10:3179–89. [PubMed: 25625928]
67. Bassani-Sternberg M, Bräunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun*. 2016; 7:13404. [PubMed: 27869121]
68. Marincola FM, Jaffee EM, Hicklin DJ, Ferrone S. Escape of human solid tumors from T-cell recognition: molecular mechanisms and functional significance. *Adv Immunol*. 2000; 74:181–273. [PubMed: 10605607]
69. Croft NP, Smith SA, Wong YC, Tan CT, Dudek NL, Flesch IE, et al. Kinetics of antigen expression and epitope presentation during virus infection. *PLoS Pathog*. 2013; 9:e1003129. [PubMed: 23382674]
70. Duan F, Duitama J, Al Seesi S, Ayres CM, Corcelli SA, Pawashe AP, et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J Exp Med*. 2014; 211:2231–48. [PubMed: 25245761]
71. Bagriaciuk EU, Kirkpatrick A, Miller KS. Glycosylation of native MHC class Ia molecules is required for recognition by allogeneic cytotoxic T lymphocytes. *Glycobiology*. 1996; 6:413–21. [PubMed: 8842705]
72. Fortier MH, Caron E, Hardy MP, Voisin G, Lemieux S, Perreault C, et al. The MHC class I peptide repertoire is molded by the transcriptome. *J Exp Med*. 2008; 205:595–610. [PubMed: 18299400]

### Significance

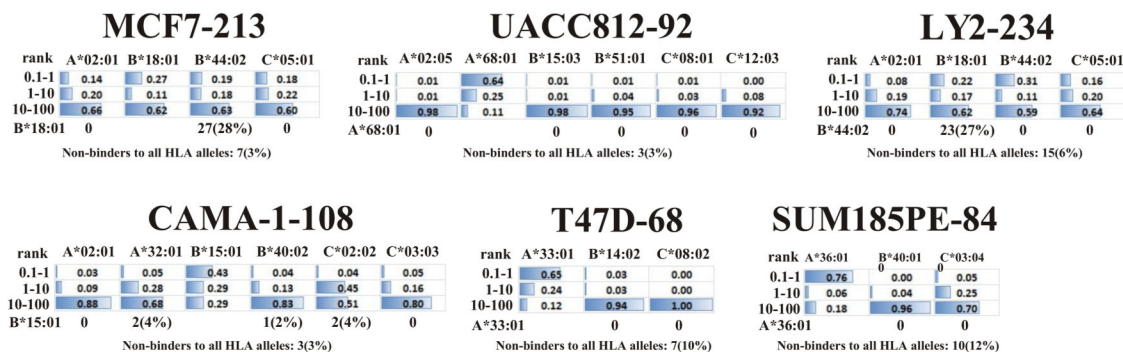
By employing proteomic analyses of eluted peptides from breast cancer cells, the current study has built an initial HLA-I-typed antigen collection for breast cancer research. It was also determined that immunogenic epitopes can be identified using established cell lines and that shared immunogenic peptides can be found in different cancer types such as breast cancer and leukemia. Importantly, out of 3,196 eluted peptides that included duplicate peptides in different cells 89 peptides either contained mutation in their sequence or were derived from aberrant translation suggesting that mutation-containing epitopes are on the order of 2–3 % in breast cancer cells. Finally, our results suggest that interfering with MHC class I function is one of the mechanisms of how tumor cells escape immune system attack.



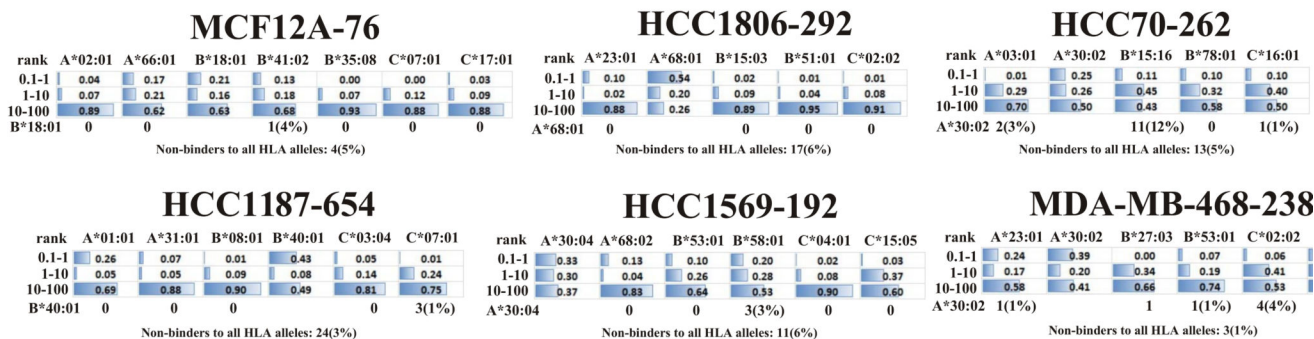
**Figure 1.**  
**A.** Immunostaining of HLA-A\*02. MCF7 and MDA-MB-231 cells were stained using HLA-A\*02-specific antibody (clone BB7.2). As a control, cells were stained using non-specific mouse IgG. Nuclei were visualized by DAPI. Original magnification, 400×. **B.** Quantitative representation of HLA-A\*02 expression in breast cancer cell lines. Stained images were scanned and analyzed using ScanR Acquire and ScanR Analysis software, respectively. The bars represent the mean of intensities between several thousands of scanned cells. Error bars indicate the standard error of the mean.

# Peptide binding probability

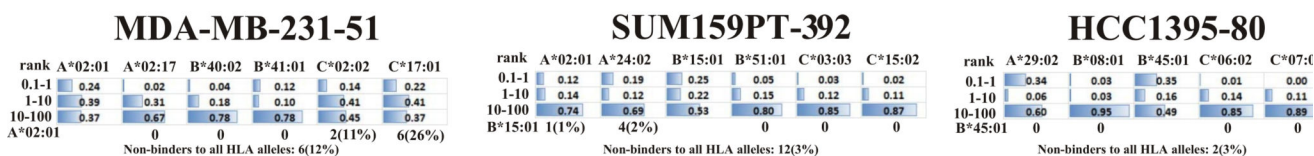
## Luminal cells



## Basal cells

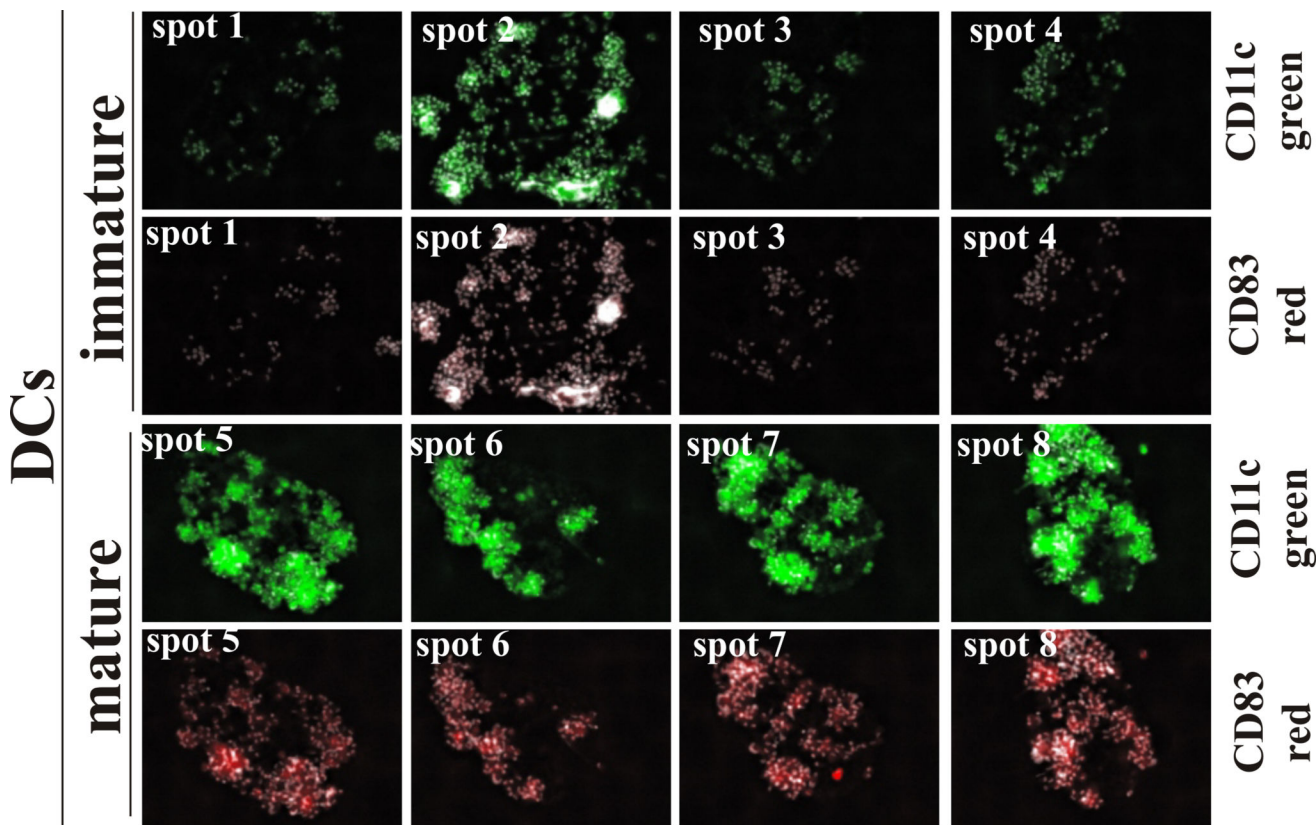


## Claudin-low cells

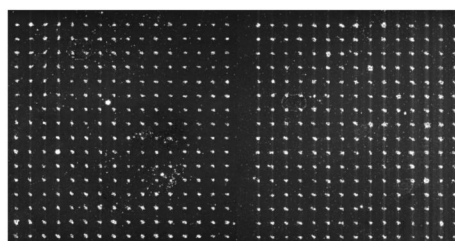


**Figure 2.**

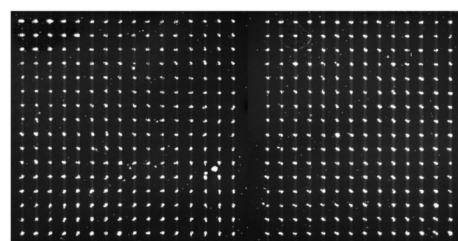
Binding probabilities of each peptide to the HLA allele present in the corresponding cells. Only 8–14-mer peptides were used for this analysis. Percentile rank for each peptide was generated by comparing the peptide's binding EC50 against those of a set of random peptides from SWISSPROT database. Numbers to the right of cell line name represent the number of analyzed peptides. Numbers below the MHC class I allele type indicate the ratio of ranked peptides in relation to all peptides. High affinity peptides (rank 0.1–1) are presented by two, three or four, and one MHC class I alleles. Numbers below plots and numbers in brackets represent quantity of shared high affinity peptides between different alleles and ratio of these shared peptides in relation to all eluted peptides, respectively. The number of peptides that are not predicted to bind any HLA allele (non-binders) and their ratio to all peptides are also indicated below the plots.



**CD11c-DC marker, CD83-DC maturation marker**



**immature Dcs**



**matured Dcs**

**Figure 3.** Attachment and maturation of DCs on the array matrix. *Upper panel.* Immunocytochemistry analysis of attached DCs. *Lower panel.* Image of immature and mature DCs on the array.

Table 1

HLA phenotype and genotype in breast cancer cell lines.

| cell line  | Subtype     | Relative to MDA-MB-231, % |             |                         |                         |                         |                 | HLA-A <sup>a</sup>      | HLA-B <sup>a</sup> | HLA-C <sup>a</sup> |
|------------|-------------|---------------------------|-------------|-------------------------|-------------------------|-------------------------|-----------------|-------------------------|--------------------|--------------------|
|            |             | protein                   |             |                         | mRNA                    |                         |                 |                         |                    |                    |
|            |             | HLA-A*02                  | MHC class I | HLA-A                   | HLA-B                   | HLA-C                   |                 |                         |                    |                    |
| MCF7       | Luminal     | 29                        | 18          | 15                      | 19                      | 46                      | A*02:01/A*02:01 | B*44:02/B*18:01         | C*05:01/C*05:01    |                    |
| UACC812    | Luminal     | 25                        | 29          | 55                      | 89                      | 245                     | A*68:01/A*02:05 | B*15:03/B*51:01         | C*08:01/C*12:03    |                    |
| LY2        | Luminal     | 17                        | 31          | 13                      | 28                      | 61                      | A*02:01/A*02:01 | B*44:02/B*18:01         | C*05:01/C*05:01    |                    |
| HCC1428    | Luminal     | 5                         | 6           | 38                      | 59                      | 319                     | A*01:01/A*02:01 | B*07:02/B*07:02         | C*07:02/C*07:02    |                    |
| CAMA-1     | Luminal     | 4                         | 15          | 14                      | 44                      | 86                      | A*02:01/A*32:01 | B*40:02/B*15:01         | C*02:02/C*03:03    |                    |
| HCC1419    | Luminal     | 2                         | 7           | 15                      | 36                      | 126                     | A*24:02/A*02:01 | B*46:01/B*52:01         | C*03:03/C*01:02    |                    |
| T47D       | Luminal     | 0                         | 46          | 35                      | 587                     | 392                     | A*33:01/A*33:01 | B*14:02/B*14:02         | C*08:02/C*08:02    |                    |
| SUM185PE   | Luminal     | 0                         | 100         | 0                       | 58                      | 184                     | A*36:01/A*36:01 | B*40:01/B*40:01         | C*03:04/C*03:04    |                    |
| SKBR3      | Luminal     | 0                         | 0           | 1                       | 0                       | 3                       | A*02:01/A*11:01 | B*35:21/B*14:02         | C*03:04/C*03:04    |                    |
| ZR75B      | Luminal     | 0                         | 0           | 2                       | 24                      | 84                      | A*02:24/A*02:24 | B*35:01/B*35:01         | C*04:01/C*04:01    |                    |
| BT474      | Luminal     | 0                         | 0           | 0                       | 14                      | 115                     | A*29:02/A*02:46 | B*07:02/B*35:08         | C*07:02/C*16:01    |                    |
| HCC1500    | Basal       | 39                        | 25          | 93                      | 139                     | 215                     | A*02:02/A*02:02 | B*15:80/B*37:19         | C*03:04/C*04:01    |                    |
| MCF12A     | Basal       | 21                        | 26          | 50                      | 43                      | 149                     | A*66:01/A*02:01 | B*41:02/B*18:01/B*35:08 | C*17:01/C*07:01    |                    |
| HCC1806    | Basal       | 1                         | 103         | 112                     | 144                     | 206                     | A*68:01/A*23:01 | B*51:01/B*15:03         | C*02:02/C*02:02    |                    |
| HCC70      | Basal       | 0                         | 54          | 406                     | 2302                    | 1394                    | A*30:02/A*03:01 | B*78:01/B*15:16         | C*16:01/C*16:01    |                    |
| HCC1187    | Basal       | 0                         | 104         | 408                     | 1708                    | 1540                    | A*31:01/A*01:01 | B*08:01/B*40:01         | C*07:01/C*03:04    |                    |
| HCC1569    | Basal       | 0                         | 38          | 259                     | 581                     | 533                     | A*30:04/A*68:02 | B*58:01/B*53:01         | C*04:01/C*15:05    |                    |
| MDA-MB-468 | Basal       | 0                         | 57          | 70                      | 111                     | 246                     | A*30:02/A*23:01 | B*27:03/B*53:01         | C*02:02/C*04:01    |                    |
| MDA-MB-157 | Basal       | 0                         | 0           | 83                      | 232                     | 310                     | A*30:01/A*02:01 | B*53:01/B*53:01         | C*06:02/C*04:01    |                    |
| CAL-120    | Basal       | 0                         | 38          | 18                      | 5                       | 19                      | A*03:01/A*03:01 | B*56:01/B*44:02         | C*01:02/C*05:01    |                    |
| CAL-51     | Basal       | 0                         | 0           | 10                      | 4                       | 11                      | A*24:02/A*29:02 | B*44:26/B*51:01         | C*16:01/C*16:01    |                    |
| MDA-MB-231 | Claudin-low | 100                       | 100         | 100 (287 <sup>b</sup> ) | 100 (146 <sup>b</sup> ) | 100 (109 <sup>b</sup> ) | A*02:17/A*02:01 | B*41:01/B*40:02         | C*17:01/C*02:02    |                    |
| BT549      | Claudin-low | 35                        | 78          | 39                      | 10                      | 92                      | A*01:01/A*02:01 | B*15:17/B*56:01         | C*07:01/C*03:03    |                    |
| SUM159PT   | Claudin-low | 26                        | 21          | 98                      | 68                      | 67                      | A*24:02/A*02:01 | B*51:01/B*15:01         | C*15:02/C*03:03    |                    |

| cell line | Subtype     | Relative to MDA-MB-231, % |             |       |       |       |                 | HLA-A <sup>a</sup> | HLA-B <sup>a</sup> | HLA-C <sup>a</sup> |
|-----------|-------------|---------------------------|-------------|-------|-------|-------|-----------------|--------------------|--------------------|--------------------|
|           |             | protein                   |             | mRNA  |       |       |                 |                    |                    |                    |
|           |             | HLA-A*02                  | MHC class I | HLA-A | HLA-B | HLA-C |                 |                    |                    |                    |
| HCC1395   | Claudin-low | 1                         | 171         | 24    | 272   | 281   | A*29:02/A*29:02 | B*08:01/B*45:01    | C*07:01/C*06:02    |                    |
| JIMT1     | HER2        | 0                         | 0           | 29    | 14    | 164   | A*68:01/A*03:01 | B*08:01/B*40:01    | C*07:06/C*03:04    |                    |

<sup>a</sup>  $p < 0.05$ . The p-value generated by the HLA prediction Seq2HLA software [33].

<sup>b</sup> Expected fragments per kilobase of transcript per million fragments sequenced (FPKM) [41].



**Table 2**

Immunogenic peptides that have been eluted from the cell surface of breast cancer cells.

| Peptide sequence | Assay  | Peptide source   | MHC class I restriction      | tumor type  | Number of identifications in the current work | Number of identifications in previous studies | Reference  |
|------------------|--|--|------------------------------|---|---|---|------------|
| GLLGTLVQL        | <sup>51</sup> Cr release                               | Catenin β-1 [CTNB1] (400–408)  | HLA-A*02:01                  | breast, ovarian, prostate   | 1   | 6   | 44         |
| ALSDHHIYL        | <sup>51</sup> Cr release                               | Fructose-bisphosphate aldolase A [ALDOA] (216–224)                       | HLA-A*02:01                  | breast, ovarian, prostate, renal cell carcinoma                         | 3   | 7   | 44, 45     |
| SLFVSNHAY        | <sup>51</sup> Cr release                               | Fructose-bisphosphate aldolase A [ALDOA] (356–364)                       | HLA-B*15:02, HLA-A*3003      | in vitro, B cell line JS  | 4   | 1   | 46, 47     |
| SQFGGSQY         | <sup>51</sup> Cr release                               | Eukaryotic translation initiation factor 3 subunit D [EIF3D] (61–69)     | HLA-B*15:02; HLA-B*15:01;    | in vitro; B lymphoblastoid cell line 721.221                            | 2   | 2   | 46, 48, 49 |
| NVIRDAVTY        | <sup>51</sup> Cr release                               | Histone H4 [H4] (65–73)  | HLA-B*15:02                  | in vitro  | 1   | 0   | 46         |
| ISDPSKVTL        | Immunoblot detection of antibody/antigen binding       | Coilin [COIL] (257–266)  |                              | in vitro  | 2   | 0   | 50         |
| LLDVPTAAV        | Positive MHC: epitope complex binding to TCR           | Gamma-interferon-inducible lysosomal thiol reductase [GILT] (27–35)      | HLA-A*02:01                  | in vitro; T2 cells; breast, ovarian, prostate; hepatocellular carcinoma | 4   | 4   | 44, 51–54  |
| YGYDNYKEY        | ELISPOT IFN-γ release                                  | CDC47L protein [CDA7L] (422–430)   | HLA-A*03, HLA-C*03, HLA-C*12 | chronic lymphocytic leukemia (CLL)                                      | 2   | 7 positive CLLs (23%)                         | 55         |
| RTAPIGR          | Response of PBMC CD8 T cells of melanoma               | Cell division cycle 5-like protein [CDC5L] (83–91)                       | HLA-A*11                     | melanoma tumor cells  | 2   | 1   | 56         |
| SVQGHYR          | Response of PBMC CD8 T cells of melanoma               | ADP/ATP translocase 2 [SLC25A5] (180–188)                                | HLA-A*11                     | melanoma tumor cells  | 2   | 1   | 56         |
| AVQEFGLARFK      | Response of tumor-infiltrating CD8 T cells of melanoma | protein of relevant evolutionary and lymphoid interest [PRELI] (142–152) | HLA-A*11                     | melanoma tumor cells  | 1   | 1   | 56         |
| VYLDKFIRL        | Response of tumor-infiltrating CD8 T cells of melanoma | guanine nucleotide-binding protein-like 3-like protein [GNL3L] (293–301) | HLA-A*24                     | melanoma tumor cells  | 1   | 1   | 56         |
| ATAGDGLIELRK     | Response of PBMC CD8 T cells of melanoma               | Prohibitin [PHB], partial (229–240)                                      | HLA-A*11                     | melanoma tumor cells  | 1   | 1   | 56         |
| ATAGIIVNR        | Response of PBMC CD8 T cells of melanoma               | Clathrin heavy chain 1 [CLTC] (311–320)                                  | HLA-A*11                     | melanoma tumor cells  | 1   | 1   | 56         |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

| Peptide sequence | Assay  | Peptide source                                      | MHC class I restriction | tumor type           | Number of identifications in the current work | Number of identifications in previous studies | Reference |
|------------------|--|---|-------------------------|----------------------|---|---|-----------|
| YPFKPPKV         | Response of tumor-infiltrating CD8 T cells of melanoma | Ubiquitin-conjugating enzyme E2 E2 [UBE2E2] (59–66) | HLA-B*51                | melanoma tumor cells | 1   | 1   | 56        |
| SSVPGVRL         | Response of tumor-infiltrating CD8 T cells of melanoma | vimentin [VIM] (72–80)                              | HLA-B*35                | melanoma tumor cells | 1   | 1   | 56        |