

CLAIMS

1. A method of identifying one of more antigenic peptides comprising:
 - (a) obtaining one or more tissue samples,
 - (b) acquiring mass spectrometry spectra data for one or more antigenic peptides, wherein the data comprises retention time;
 - (c) receiving, on at least one processor, test data comprising peptide spectrum data,
 - (d) evaluating, using the at least one processor, the test data using a classifier which is an electronic representation of a classification system, each said classifier trained using an electronically stored set of training data vectors, each training data vector representing an individual peptide and comprising a peptide spectrum data for the peptide, each training data vector further comprising a classification with respect to whether or not the peptide is antigenic,
wherein the classification system comprises Deep Learning algorithm selected from the group consisting of pDeep, DeepMass, or PROSIT, or an ensemble thereof,
 - (e) outputting, using the at least one processor, a classification of the sample from the peptide spectrum data concerning the likelihood of whether or not the peptide is antigenic based on the evaluating step;
 - (f) identifying one or more antigenic peptides.
2. The method of claim 1, wherein the mass spectrometry comprises tandem mass spectrometry (MS/MS).
3. The method of any one of claim 1 or 2, wherein the library or database comprises over 70%, over 80%, over 85%, over 90%, over 95%, or 100% antigenic peptide data.
4. The method of any one of claims 1–3, wherein the library or database comprises less than 30%, less than 25%, less than 20%, less than 15%, less than 10%, or less than 5% tryptic peptides data.
5. The method of any one of claims 1–4, wherein the prediction peptide performance score is greater than 0.95.

6. The method of any one of claims 1–4, wherein the prediction peptide performance score is from 0.92 to 0.98.
7. The method of any one of claims 1–6, wherein the peptide spectrum match (PSM) have a false discovery rate (FDR) of less than 0.05.
8. The method of any one of claims 1–7, wherein antigenic peptides are identified with greater accuracy than tryptic peptides.
9. The method of any one of claims 1–8, wherein the antigenic peptides are 8 to 11 amino acid or 8 to 9 amino acids in length.
10. The method of any one of claims 1–9, wherein the one or more antigenic peptides identified are overexpressed or presented in one or more specific cancer tissues.
11. The method of any one of claims 1–9, wherein the tissue is a cancer tissue and is selected from the group consisting hepatocellular carcinoma (HCC), colorectal carcinoma (CRC), glioblastoma (GB), gastric cancer (GC), esophageal cancer, non-small cell lung cancer (NSCLC), pancreatic cancer (PC), renal cell carcinoma (RCC), benign prostate hyperplasia (BPH), prostate cancer (PCA), ovarian cancer (OC), melanoma, breast cancer (BRCA), chronic lymphocytic leukemia (CLL), Merkel cell carcinoma (MCC), small cell lung cancer (SCLC), Non-Hodgkin lymphoma (NHL), acute myeloid leukemia (AML), gallbladder cancer and cholangiocarcinoma (GBC, CCC), urinary bladder cancer (UBC), uterine cancer (UEC), and combination thereof.
12. The method of any one of claims 1–11, wherein the one or more tissue samples comprise tumor tissue and corresponding healthy tissue and are obtained from an individual for testing.
13. The method of any one of claims 1–12, wherein prior to (a) obtaining one or more tumor samples from an individual, the HLA-subtypes of the individual are determined.
14. The method of claim 1, wherein the Deep Learning algorithm is pDeep.

- 15.** The method of any one of claims 1–14, further comprising a step of utilizing the identified one or more peptides in adoptive cellular therapy.
- 16.** The method of any one of claims 1–15, wherein the one or more antigenic peptides bind to an MHC molecule.
- 17.** The method of claim 1, wherein the training data vectors are stored and/or accessed from a public database, non-public database, or a combination thereof.
- 18.** The method of any one of claims 1-17, wherein the peptides are HLA-associated peptides.
- 19.** The method of any one of claims 1-18, wherein the peptide spectra data is experimental spectrum data, predicted spectrum data, or a combination thereof.
- 20.** The method of claim 19, wherein the peptide spectra data is experimental spectrum data.
- 21.** The method of claim 1, wherein the test data is obtained from mass spectroscopy, is predicated spectra data, or a combination thereof.
- 22.** The method of claim 1, wherein the test data is obtained from a peptide obtained from a physiological sample from a patient, preferably a tumor biopsy.
- 23.** The method of claim 1, wherein the tissue sample is a tumor biopsy.
- 24.** The method of claim 1, wherein the tissue sample is normal tissue.
- 25.** The method of claim 1, wherein the training data comprises peptide data obtained from at least 1–30 different types of cancer.

- 26.** The method of claim 1, wherein the training data comprises peptide data obtained from cancer samples selected from the group consisting of hepatocellular carcinoma (HCC), colorectal carcinoma (CRC), glioblastoma (GB), gastric cancer (GC), esophageal cancer, non-small cell lung cancer (NSCLC), pancreatic cancer (PC), renal cell carcinoma (RCC), benign prostate hyperplasia (BPH), prostate cancer (PCA), ovarian cancer (OC), melanoma, breast cancer (BRCA), chronic lymphocytic leukemia (CLL), Merkel cell carcinoma (MCC), small cell lung cancer (SCLC), Non-Hodgkin lymphoma (NHL), acute myeloid leukemia (AML), gallbladder cancer and cholangiocarcinoma (GBC, CCC), urinary bladder cancer (UBC), uterine cancer (UEC), or a combination thereof.
- 27.** The method of claim 1, wherein the training data comprises peptide data obtained from at least 1–50 different types of normal tissue.
- 28.** The method of claim 1, wherein the training data comprises peptide data obtained from normal tissues samples selected from the group consisting of adipose tissue, ovary, adrenal gland, pancreas, artery, peripheral nerve tissue, bladder, pituitary gland, blood cells, placenta, bone marrow, pleura, central nervous system tissue, preferably brain or spinal cord, prostate, breast, rectum, cartilage, salivary gland, cervix, ski, colon, small intestine, preferably duodenum, jejunum, ileum, appendix, stomach, esophagus, spleen, eye, gall bladder, testis, heart, thymus, kidney, thyroid, liver, trachea, lung, ureter, lymph node, uterus, muscle, vein, or a combination thereof.
- 29.** The method of claim 1, wherein the training data comprises peptide data evaluated from over 1500, over 2000, over 2500, or over 3000 tissue samples.
- 30.** The method of claim 1, wherein the training data comprises over 100 million, over 150 million, over 180 million, or over 200 million MS/MS spectra.
- 31.** The method of claim 16, wherein the MHC molecule is MHC Class I or MHC Class II.