[0008] The situation is even more complicated for the HLA-II system. While there are also three types of HLA-II genes, referred to as HLA-DR, HLA-DP and HLA-DQ, they are each heterodimer complexes formed by two polymorphic genes (referred to as the respective alpha and beta chains). Again, each person inherits two alleles (one from the mother and one from the father) of each of the genes. Thus, in total, each person has (up to) twelve different HLA-II complexes. As with HLA-I, different people have different alleles, and thousands of different combinations have been observed.

SUMMARY

[0009] In an ~~embodiment~~ example, the present ~~invention~~ disclosure provides a method of ranking epitopes derived from neoantigens as targets for personalized immunotherapy. Candidate epitopes are collected based on patient data of a cancer patient. A set of scores are calculated for each of the candidate epitopes, each of the scores in a respective one of the sets for a respective one of the candidate epitopes representing an independent measure of a likelihood of the respective one of candidate epitopes to elicit an immune response in the cancer patient. The scores in each of the sets of scores are combined into a single score for each of the candidate epitopes. The single scores for the candidate epitopes in each case reflect an overall likelihood of eliciting the immune response in the patient. The candidate epitopes are ranked using the single scores for the immunotherapy.

[0009a] In one aspect, the present disclosure provides a method of ranking epitopes derived from neoantigens as targets for personalized immunotherapy, the method comprising: collecting candidate epitopes based on patient data of a cancer patient; calculating a set of scores for each of the candidate epitopes, each of the scores in a respective one of the sets for a respective one of the candidate epitopes representing an independent measure of a likelihood of the respective one of candidate epitopes to elicit an immune response in the cancer patient, wherein each of the sets of scores includes at least a first score indicating a likelihood of human leukocyte antigen (HLA) binding which is determined using HLA alleles which are specific to the cancer patient, a second score indicating a T-cell response which is predicted using a T-cell receptor (TCR) repertoire which is identified using healthy ribonucleic acid (RNA)-sequence data specific to the cancer patient and a third score based on tumor RNA-sequence data specific to the cancer patient; combining the scores in each of the sets of scores into a single score for each of the candidate epitopes, the single scores for the candidate epitopes in each case reflecting an overall likelihood of eliciting the immune

response in the patient; extracting experimentally-verified properties of the epitopes and domain knowledge about the epitopes, and embedding each of the epitopes in a vector space based on the experimentally-verified properties and the domain knowledge, and ranking the candidate epitopes using the single scores for the immunotherapy, wherein the candidate epitopes are ranked based on the single scores and the embeddings, and the ranking is performed in an order of largest weighted distances in the vector space, the weighted distances in each case being determined based on Euclidean distances in the vector space multiplied by the single score for each respective one of the candidate epitopes such that one of the candidate epitopes with a largest weighted distance from an origin of the vector space is ranked first followed by one of the candidate epitopes having a largest weighted difference from the top-ranked epitope.

[0009b] In another aspect, the present disclosure provides a computer system for ranking epitopes derived from neoantigens as targets for personalized immunotherapy, the computer system comprising memory and one or more processors which, alone or in combination, are configured to provide for execution of a method comprising: collecting candidate epitopes based on patient data of a cancer patient; calculating a set of scores for each of the candidate epitopes, each of the scores in a respective one of the sets for a respective one of the candidate epitopes representing an independent measure of a likelihood of the respective one of candidate epitopes to elicit an immune response in the cancer patient, wherein each of the sets of scores includes at least a first score indicating a likelihood of human leukocyte antigen (HLA) binding which is determined using HLA alleles which are specific to the cancer patient, a second score indicating a T-cell response which is predicted using a T-cell receptor (TCR) repertoire which is identified using healthy ribonucleic acid (RNA)-sequence data specific to the cancer patient and a third score based on tumor RNA-sequence data specific to the cancer patient; combining the scores in each of the sets of scores into a single score for each of the candidate epitopes, the single scores for the candidate epitopes in each case reflecting an overall likelihood of eliciting the immune response in the patient; and extracting experimentally-verified properties of the epitopes and domain knowledge about the epitopes, and embedding each of the epitopes in a vector space based on the experimentally-verified properties and the domain knowledge, ranking the candidate epitopes using the single scores for the immunotherapy, wherein the candidate epitopes are ranked based on the single scores and the embeddings, and the ranking is performed in an order of largest weighted distances in the vector space, the weighted distances in each case being determined based on Euclidean

distances in the vector space multiplied by the single score for each respective one of the candidate epitopes such that one of the candidate epitopes with a largest weighted distance from an origin of the vector space is ranked first followed by one of the candidate epitopes having a largest weighted difference from the top-ranked epitope.

[0009c] In another aspect, the present disclosure provides a non-transitory computer-readable medium having instructions thereon, which, upon execution by one or more processors, alone or in combination, and using memory, provides for execution of a method comprising: collecting candidate epitopes based on patient data of a cancer patient; calculating a set of scores for each of the candidate epitopes, each of the scores in a respective one of the sets for a respective one of the candidate epitopes representing an independent measure of a likelihood of the respective one of candidate epitopes to elicit an immune response in the cancer patient, wherein each of the sets of scores includes at least a first score indicating a likelihood of human leukocyte antigen (HLA) binding which is determined using HLA alleles which are specific to the cancer patient, a second score indicating a T-cell response which is predicted using a T-cell receptor (TCR) repertoire which is identified using healthy ribonucleic acid (RNA)-sequence data specific to the cancer patient and a third score based on tumor RNA-sequence data specific to the cancer patient; combining the scores in each of the sets of scores into a single score for each of the candidate epitopes, the single scores for the candidate epitopes in each case reflecting an overall likelihood of eliciting the immune response in the patient; extracting experimentally-verified properties of the epitopes and domain knowledge about the epitopes, and embedding each of the epitopes in a vector space based on the experimentally-verified properties and the domain knowledge and ranking the candidate epitopes using the single scores for the immunotherapy, wherein the candidate epitopes are ranked based on the single scores and the embeddings, and the ranking is performed in an order of largest weighted distances in the vector space, the weighted distances in each case being determined based on Euclidean distances in the vector space multiplied by the single score for each respective one of the candidate epitopes such that one of the candidate epitopes with a largest weighted distance from an origin of the vector space is ranked first followed by one of the candidate epitopes having a largest weighted difference from the top-ranked epitope.

BRIEF DESCRIPTION OF THE DRAWING

[0010] The present invention will be described in even greater detail below based on the exemplary figures. The invention is not limited to the exemplary embodiments. All features described and/or illustrated herein can be used alone or combined in different combinations in embodiments of the invention. The features and advantages of various embodiments of the present invention will become apparent by reading the following detailed description with reference to the attached drawings which illustrate the following:

[0011] FIG. 1 is a schematic overview of a computer system and method for ranking and selecting target epitopes for immunotherapy;

[0012] FIG. 2 is a flowchart of a method for preparation of plasmid carrying sequences encoding for patient-specific neopeptides; and

CLAIMS

What is claimed is:

1. A method of ranking epitopes derived from neoantigens as targets for personalized immunotherapy, the method comprising:

collecting candidate epitopes based on patient data of a cancer patient;

calculating a set of scores for each of the candidate epitopes, each of the scores in a respective one of the sets for a respective one of the candidate epitopes representing an independent measure of a likelihood of the respective one of candidate epitopes to elicit an immune response in the cancer patient,

wherein each of the sets of scores includes at least a first score indicating a likelihood of human leukocyte antigen (HLA) binding which is determined using HLA alleles which are specific to the cancer patient, a second score indicating a T-cell response which is predicted using a T-cell receptor (TCR) repertoire which is identified using healthy ribonucleic acid (RNA)-sequence data specific to the cancer patient and a third score based on tumor RNA-sequence data specific to the cancer patient;

combining the scores in each of the sets of scores into a single score for each of the candidate epitopes, the single scores for the candidate epitopes in each case reflecting an overall likelihood of eliciting the immune response in the patient;

extracting experimentally-verified properties of the epitopes and domain knowledge about the epitopes, and

embedding each of the epitopes in a vector space based on the experimentally-verified properties and the domain knowledge, and

ranking the candidate epitopes using the single scores for the immunotherapy,

wherein the candidate epitopes are ranked based on the single scores and the embeddings, and the ranking is performed in an order of largest weighted distances in the vector space, the weighted distances in each case being determined based on Euclidean distances in the vector space multiplied by the single score for each respective one of the

candidate epitopes such that one of the candidate epitopes with a largest weighted distance from an origin of the vector space is ranked first followed by one of the candidate epitopes having a largest weighted difference from the top-ranked epitope.

~~2. The method according to claim 1, wherein each of the sets of scores includes at least a first score indicating a likelihood of human leukocyte antigen (HLA) binding which is determined using HLA alleles which are specific to the cancer patient, and a second score indicating a T-cell response which is predicted using a T-cell receptor (TCR) repertoire which is identified using healthy ribonucleic acid (RNA) sequence data specific to the cancer patient.~~

~~3. The method according to claim 2, wherein each of the sets of scores further comprises a third score based on tumor RNA-sequence data specific to the cancer patient.~~

~~4. The method according to claim 1, further comprising:~~

~~extracting experimentally-verified properties of the epitopes and domain knowledge about the epitopes, and~~

~~embedding each of the epitopes in a vector space based on the experimentally-verified properties and the domain knowledge.~~

~~5. The method according to claim 4, wherein the candidate epitopes are ranked based on the single scores and the embeddings.~~

~~6. The method according to claim 5, wherein the ranking is performed in an order of largest weighted distances in the vector space, the weighted distances in each case being determined based on Euclidean distances in the vector space multiplied by the single score for each respective one of the candidate epitopes such that one of the candidate epitopes with a largest weighted distance from an origin of the vector space is ranked first followed by one of the candidate epitopes having a largest weighted difference from the top-ranked epitope.~~

~~7.~~2. The method according to claim 4~~1~~, wherein the embedding is performed using a representation learning embedding framework which uses an affinity graph in which nodes represent the epitopes and edges connect the epitopes which have a similarity measure above a predefined threshold, wherein attributes of the nodes include at least the experimentally-

derived properties and the domain knowledge, and wherein an embedding function is learned for each of the attributes to map the attributes to numeric vectors.

~~8.~~3. The method according to claim ~~4~~1, wherein the embedding is performed by direct embedding in which at least the experimentally-derived properties and the domain knowledge are each embedded using numeric vectors which are concatenated together.

~~9.~~4. The method according to claim ~~4~~1, wherein the embeddings include vector representations of biochemical properties of the epitopes.

~~10.~~5. The method according to claim ~~4~~1, wherein the embeddings include vector representations of amino acid sequences of the epitopes.

~~11.~~6. A computer system for ranking epitopes derived from neoantigens as targets for personalized immunotherapy, the computer system comprising memory and one or more processors which, alone or in combination, are configured to provide for execution of a method comprising:

collecting candidate epitopes based on patient data of a cancer patient;

calculating a set of scores for each of the candidate epitopes, each of the scores in a respective one of the sets for a respective one of the candidate epitopes representing an independent measure of a likelihood of the respective one of candidate epitopes to elicit an immune response in the cancer patient,

wherein each of the sets of scores includes at least a first score indicating a likelihood of human leukocyte antigen (HLA) binding which is determined using HLA alleles which are specific to the cancer patient, a second score indicating a T-cell response which is predicted using a T-cell receptor (TCR) repertoire which is identified using healthy ribonucleic acid (RNA)-sequence data specific to the cancer patient and a third score based on tumor RNA-sequence data specific to the cancer patient;

combining the scores in each of the sets of scores into a single score for each of the candidate epitopes, the single scores for the candidate epitopes in each case reflecting an overall likelihood of eliciting the immune response in the patient; and

extracting experimentally-verified properties of the epitopes and domain knowledge about the epitopes, and

embedding each of the epitopes in a vector space based on the experimentally-verified properties and the domain knowledge,

ranking the candidate epitopes using the single scores for the immunotherapy,

wherein the candidate epitopes are ranked based on the single scores and the embeddings, and the ranking is performed in an order of largest weighted distances in the vector space, the weighted distances in each case being determined based on Euclidean distances in the vector space multiplied by the single score for each respective one of the candidate epitopes such that one of the candidate epitopes with a largest weighted distance from an origin of the vector space is ranked first followed by one of the candidate epitopes having a largest weighted difference from the top-ranked epitope.

~~12. The computer system according to claim 11, wherein each of the sets of scores includes at least a first score indicating a likelihood of human leukocyte antigen (HLA) binding which is determined using HLA alleles which are specific to the cancer patient, and a second score indicating a T-cell response which is predicted using a T-cell receptor (TCR) repertoire which is identified using healthy ribonucleic acid (RNA) sequence data specific to the cancer patient.~~

~~13. The computer system according to claim 11, being further configured to provide for execution of the following steps:~~

~~extracting experimentally-verified properties of the epitopes and domain knowledge about the epitopes, and~~

~~embedding each of the epitopes in a vector space based on the experimentally-verified properties and the domain knowledge,~~

~~wherein the candidate epitopes are ranked based on the single scores and the embeddings.~~

~~14. The computer system according to claim 13, wherein the ranking is performed in an order of largest weighted distances in the vector space, the weighted distances in each case being determined based on Euclidean distances in the vector space multiplied by the single~~

~~score for each respective one of the candidate epitopes such that one of the candidate epitopes with a largest weighted distance from an origin of the vector space is ranked first followed by one of the candidate epitopes having a largest weighted difference from the top-ranked epitope.~~

~~15.~~7. A non-transitory computer-readable medium having instructions thereon, which, upon execution by one or more processors, alone or in combination, and using memory, provides for execution of a method comprising:

collecting candidate epitopes based on patient data of a cancer patient;

calculating a set of scores for each of the candidate epitopes, each of the scores in a respective one of the sets for a respective one of the candidate epitopes representing an independent measure of a likelihood of the respective one of candidate epitopes to elicit an immune response in the cancer patient,

wherein each of the sets of scores includes at least a first score indicating a likelihood of human leukocyte antigen (HLA) binding which is determined using HLA alleles which are specific to the cancer patient, a second score indicating a T-cell response which is predicted using a T-cell receptor (TCR) repertoire which is identified using healthy ribonucleic acid (RNA)-sequence data specific to the cancer patient and a third score based on tumor RNA-sequence data specific to the cancer patient;

combining the scores in each of the sets of scores into a single score for each of the candidate epitopes, the single scores for the candidate epitopes in each case reflecting an overall likelihood of eliciting the immune response in the patient; extracting experimentally-verified properties of the epitopes and domain knowledge about the epitopes, and

embedding each of the epitopes in a vector space based on the experimentally-verified properties and the domain knowledge and

ranking the candidate epitopes using the single scores for the immunotherapy,

wherein the candidate epitopes are ranked based on the single scores and the embeddings, and the ranking is performed in an order of largest weighted distances in the vector space, the weighted distances in each case being determined based on Euclidean distances in the vector space multiplied by the single score for each respective one of the

candidate epitopes such that one of the candidate epitopes with a largest weighted distance from an origin of the vector space is ranked first followed by one of the candidate epitopes having a largest weighted difference from the top-ranked epitope.