

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第1部門第1区分

【発行日】令和3年11月11日(2021.11.11)

【公表番号】特表2020-536553(P2020-536553A)

【公表日】令和2年12月17日(2020.12.17)

【年通号数】公開・登録公報2020-051

【出願番号】特願2020-520270(P2020-520270)

【国際特許分類】

C 1 2 Q	1/6869	(2018.01)
A 6 1 K	39/00	(2006.01)
A 6 1 P	35/00	(2006.01)
A 6 1 P	11/00	(2006.01)
A 6 1 P	17/00	(2006.01)
A 6 1 P	15/00	(2006.01)
A 6 1 P	13/12	(2006.01)
A 6 1 P	13/08	(2006.01)
A 6 1 P	1/04	(2006.01)
A 6 1 P	11/02	(2006.01)
A 6 1 P	11/04	(2006.01)
A 6 1 P	27/16	(2006.01)
A 6 1 P	1/02	(2006.01)
A 6 1 P	1/18	(2006.01)
A 6 1 P	25/00	(2006.01)
A 6 1 P	35/02	(2006.01)
A 6 1 P	43/00	(2006.01)
G 0 1 N	33/53	(2006.01)
G 0 1 N	33/574	(2006.01)
C 1 2 N	15/12	(2006.01)
C 4 0 B	40/08	(2006.01)

【 F I 】

C 1 2 Q	1/6869	Z N A Z
A 6 1 K	39/00	H
A 6 1 P	35/00	
A 6 1 P	11/00	
A 6 1 P	17/00	
A 6 1 P	15/00	
A 6 1 P	13/12	
A 6 1 P	13/08	
A 6 1 P	1/04	
A 6 1 P	11/02	
A 6 1 P	11/04	
A 6 1 P	27/16	
A 6 1 P	1/02	
A 6 1 P	1/18	
A 6 1 P	25/00	
A 6 1 P	35/02	
A 6 1 P	43/00	1 0 5
G 0 1 N	33/53	M
G 0 1 N	33/574	D

C 1 2 N 15/12
C 4 0 B 40/08

【手続補正書】

【提出日】令和3年9月28日(2021.9.28)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

腫瘍細胞の表面上に提示される可能性の高い、対象の1つ以上の腫瘍細胞に由来する1つ以上の新生抗原を特定するための方法であって、

前記対象の前記腫瘍細胞及び正常細胞からエクソーム、トランスクリプトーム、または全ゲノムのヌクレオチドシーケンシングデータのうちの少なくとも1つを取得する工程であって、前記ヌクレオチドシーケンシングデータが、前記腫瘍細胞からの前記ヌクレオチドシーケンシングデータと前記正常細胞からの前記ヌクレオチドシーケンシングデータを比較することにより特定された新生抗原のセットのそれぞれのペプチド配列を表すデータを取得するために用いられ、各新生抗原のペプチド配列が、前記ペプチド配列を前記対象の前記正常細胞から特定される対応する野生型のペプチド配列とは異なるものとする少なくとも1つの変化を含む、前記取得する工程；

前記新生抗原のそれぞれのペプチド配列を、対応する数値ベクトルにコード化する工程であって、各数値ベクトルが、前記ペプチド配列を構成する複数のアミノ酸と前記ペプチド配列内の前記アミノ酸の位置のセットとに関する情報を含む、前記コード化する工程；

前記新生抗原のそれぞれのペプチド配列を、前記対象の前記ヌクレオチドシーケンシングデータの複数のkマーブブロックのうちの1つ以上のkマーブブロックと関連付ける工程；

前記新生抗原のセットについて提示尤度のセットを生成するために、コンピュータプロセッサを使用して、前記数値ベクトル及び前記1つ以上の関連付けられたkマーブブロックを機械学習させた提示モデルに入力する工程であって、前記セット内の各提示尤度が、対応する新生抗原が1つ以上のMHCアレルによって前記対象の前記腫瘍細胞の表面上に提示される尤度を表し、前記機械学習させた提示モデルが、

訓練データセットに少なくとも基づいて特定される複数のパラメータであって、前記訓練データセットが、

複数の試料の各試料について、前記試料中に存在するものとして特定されたMHCアレルのセット内の少なくとも1つのMHCアレルに結合したペプチドの存在を測定する質量分析によって得られた標識、

前記試料のそれぞれについて、前記ペプチドを構成する複数のアミノ酸と前記ペプチド内の前記アミノ酸の位置のセットとに関する情報を含む数値ベクトルとしてコード化された訓練ペプチド配列、及び

前記試料のそれぞれについて、前記試料の前記訓練ペプチド配列のそれぞれについて、前記訓練ペプチド配列と、前記訓練ペプチド配列の前記ヌクレオチドシーケンシングデータの複数のkマーブブロックのうちの1つ以上のkマーブブロックとの間の関連付け

を含み、

前記複数のパラメータのサブセットが、前記1つ以上のkマーブブロックにおける提示ホットスポットの有無を表す、前記複数のパラメータと、

入力として受け取られた前記数値ベクトル及び前記1つ以上のkマーブブロックと、前記数値ベクトル、前記1つ以上のkマーブブロック、及び前記パラメータに基づいた出力

として生成された前記提示尤度との間の関係を表す関数と
を含む、

前記入力する工程；

選択された新生抗原のセットを生成するために、前記新生抗原のセットのサブセットを、前記提示尤度のセットに基づいて選択する工程；ならびに

前記選択された新生抗原のセットを返す工程

を含む、前記方法。

【請求項2】

前記数値ベクトルを前記機械学習させた提示モデルに入力する工程が、

前記ペプチド配列の特定の位置の特定のアミノ酸に基づいてMHCアレルが前記新生抗原を提示するかどうかを示す依存性スコアを前記1つ以上のMHCアレルのそれぞれについて生成するために、前記機械学習させた提示モデルを前記新生抗原のペプチド配列に適用すること

を含む、請求項1に記載の方法。

【請求項3】

(A) 前記数値ベクトルを前記機械学習させた提示モデルに入力する工程が、各MHCアレルについて、対応するMHCアレルが対応する新生抗原を提示するであろう尤度を示す対応するアレル毎尤度を生成するために、前記依存性スコアを変換することと、前記アレル毎尤度を組み合わせて前記新生抗原の提示尤度を生成することとをさらに含み、任意で、前記依存性スコアを変換することが、前記新生抗原の提示を前記1つ以上のMHCアレルにわたって相互排他的なものとしてモデル化する；または

(B) 前記数値ベクトルを前記機械学習させた提示モデルに入力する工程が、前記提示尤度を生成するために前記依存性スコアの組み合わせを変換することとをさらに含み、前記依存性スコアの組み合わせを変換することが、前記新生抗原の提示を前記1つ以上のMHCアレル間で干渉するものとしてモデル化する、

請求項2に記載の方法。

【請求項4】

前記提示尤度のセットが、少なくとも1つ以上のアレル非相互作用特性によってさらに特定され、

前記アレル非相互作用特性に基づいて前記対応する新生抗原のペプチド配列が提示されるかどうかを示す前記アレル非相互作用特性についての依存性スコアを生成するために、前記機械学習させた提示モデルを前記アレル非相互作用特性に適用すること

をさらに含む、請求項2または3に記載の方法。

【請求項5】

(A) 前記1つ以上のMHCアレルの各MHCアレルについての前記依存性スコアを、前記アレル非相互作用特性についての前記依存性スコアと組み合わせることと、

対応するMHCアレルが対応する新生抗原を提示する尤度を示す、各MHCアレルについてのアレル毎尤度を生成するために、各MHCアレルについての前記組み合わせられた依存性スコアを変換することと、

前記提示尤度を生成するために、前記アレル毎尤度を組み合わせることと

をさらに含む；または

(B) 前記MHCアレルのそれぞれについての前記依存性スコアと、前記アレル非相互作用特性についての前記依存性スコアとを組み合わせることと、

前記提示尤度を生成するために、前記組み合わせられた依存性スコアを変換することと

をさらに含む、

請求項4に記載の方法。

【請求項6】

前記少なくとも1つ以上のアレル非相互作用特性が、前記新生抗原のペプチド配列と、前記新生抗原の前記ヌクレオチドシーケンシングデータの複数のkマーブブロックのうち

の1つ以上のkマーブロックとの間の関連付けを含む、請求項4または5に記載の方法。

【請求項7】

- (A) 前記1つ以上のMHCアレルが2つ以上の異なるMHCアレルを含む；
 (B) 前記ペプチド配列が、アミノ酸9個以外の長さを有するペプチド配列を含む；
 (C) 前記ペプチド配列をコード化することが、ワン・ホットエンコーディングスキームを用いて前記ペプチド配列をコード化することを含む；
 (D) 前記複数の試料が、
 (a) 単一のMHCアレルを発現するように操作された1つ以上の細胞株、
 (b) 複数のMHCアレルを発現するように操作された1つ以上の細胞株、
 (c) 複数の患者から得られた、または複数の患者に由来する1つ以上のヒト細胞株、
 (d) 複数の患者から得られた新鮮なまたは凍結された腫瘍試料、及び
 (e) 複数の患者から得られた新鮮なまたは凍結された組織試料
のうちの少なくとも1つを含む；かつ/または
 (E) 前記訓練データセットが、
 (a) 前記ペプチドのうちの少なくとも1つについてのペプチド-MHC結合親和性の測定値に関連するデータ、及び
 (b) 前記ペプチドのうちの少なくとも1つについてのペプチド-MHC結合安定性の測定値に関連するデータ
のうちの少なくとも1つをさらに含む、
 請求項1~6のいずれか1項に記載の方法。

【請求項8】

- (A) 前記提示尤度のセットが、RNA-seqまたは質量分析により測定される、前記対象における前記1つ以上のMHCアレルの少なくとも発現レベルによってさらに特定される；
 (B) 前記提示尤度のセットが、
 (a) 前記新生抗原のセット内の新生抗原と前記1つ以上のMHCアレルとの間の予測される親和性、及び
 (b) 前記新生抗原コード化ペプチド-MHC複合体の予測される安定性のうちの少なくとも1つを含む特性によってさらに特定される；かつ/または
 (C) 前記数値的尤度のセットが、
 (a) そのソースタンパク質配列内の、前記新生抗原コード化ペプチド配列に隣接するC末端配列、及び
 (b) そのソースタンパク質配列内の、前記新生抗原コード化ペプチド配列に隣接するN末端配列
のうちの少なくとも1つを含む特性によってさらに特定される、
 請求項1~7のいずれか1項に記載の方法。

【請求項9】

- 前記選択された新生抗原のセットを選択することが、
 (A) 前記機械学習させた提示モデルに基づいて、選択されない新生抗原と比較して前記腫瘍細胞表面上に提示される尤度が増大している新生抗原；
 (B) 前記機械学習させた提示モデルに基づいて、選択されない新生抗原と比較して前記対象において腫瘍特異的な免疫応答を誘導することができる尤度が増大している新生抗原；
 (C) 前記提示モデルに基づいて、選択されない新生抗原と比較してプロフェッショナル抗原提示細胞(APC)によってナイーブT細胞に対して提示されることができ尤度が増大している新生抗原であって、任意で、前記APCが樹状細胞(DC)である新生抗原；
 (D) 前記機械学習させた提示モデルに基づいて、選択されない新生抗原と比較して中枢性寛容または末梢性寛容によって阻害される尤度が減少している新生抗原；及び/または

(E) 前記機械学習させた提示モデルに基づいて、選択されない新生抗原と比較して前記対象において正常組織に対する自己免疫応答を誘導することができる尤度が減少している新生抗原

を選択することを含む、請求項1~8のいずれか1項に記載の方法。

【請求項10】

(A) 前記1つ以上の腫瘍細胞が、肺癌、メラノーマ、乳癌、卵巣癌、前立腺癌、腎臓癌、胃癌、結腸癌、精巣癌、頭頸部癌、膵臓癌、脳癌、B細胞リンパ腫、急性骨髄性白血病、慢性骨髄性白血病、慢性リンパ球性白血病、T細胞リンパ球性白血病、非小細胞肺癌、及び小細胞肺癌からなる群から選択される；かつ/または

(B) 前記方法が、前記選択された新生抗原のセットから個別化がんワクチンを構築するための出力を生成することをさらに含み、任意で、前記個別化がんワクチン用の出力が、前記選択された新生抗原のセットをコードした少なくとも1つのペプチド配列または少なくとも1つのヌクレオチド配列を含む、

請求項1~9のいずれか1項に記載の方法。

【請求項11】

前記機械学習させた提示モデルが、ニューラルネットワークモデルである、請求項1~10のいずれか1項に記載の方法。

【請求項12】

前記ニューラルネットワークモデルがMHCアレルについての複数のネットワークモデルを含み、それぞれのネットワークモデルが複数のMHCアレルのうちの対応するMHCアレルに割り当てられ、かつ1つ以上の層に配置された一連のノードを含み、任意で、前記ニューラルネットワークモデルが、前記ニューラルネットワークモデルのパラメータを更新することによって訓練され、少なくとも2つのネットワークモデルの前記パラメータが、少なくとも1回の訓練イテレーションについて一緒に更新される、請求項11に記載の方法。

【請求項13】

前記機械学習させた提示モデルが、ノードの1つ以上の層を含むディープラーニングモデルである、請求項11または12に記載の方法。

【請求項14】

前記1つ以上のMHCアレルが、クラスI MHCアレルである、請求項1~13のいずれか1項に記載の方法。

【請求項15】

コンピュータプロセッサと、

前記コンピュータプロセッサにより実行されると、前記コンピュータプロセッサに、
対象の腫瘍細胞及び正常細胞からエクソーム、トランスクリプトーム、または全ゲノムのヌクレオチドシーケンシングデータのうちの少なくとも1つを取得することであって、前記ヌクレオチドシーケンシングデータが、前記腫瘍細胞からの前記ヌクレオチドシーケンシングデータと前記正常細胞からの前記ヌクレオチドシーケンシングデータとを比較することにより特定された新生抗原のセットのそれぞれのペプチド配列を表すデータを取得するために用いられ、各新生抗原のペプチド配列が、前記ペプチド配列を前記対象の前記正常細胞から特定される対応する野生型のペプチド配列とは異なるものとする少なくとも1つの変化を含む、前記取得すること；

前記新生抗原のそれぞれのペプチド配列を、対応する数値ベクトルにコード化することであって、各数値ベクトルが、前記ペプチド配列を構成する複数のアミノ酸と前記ペプチド配列内の前記アミノ酸の位置のセットとに関する情報を含む、前記コード化すること；

前記新生抗原のそれぞれのペプチド配列を、前記対象の前記ヌクレオチドシーケンシングデータの複数のkマーブブロックのうちの1つ以上のkマーブブロックと関連付けること；

前記新生抗原のセットについて提示尤度のセットを生成するために、前記数値ベク

トル及び前記1つ以上の関連付けられたkマーブブロックを機械学習させた提示モデルに入力することによって、前記セット内の各提示尤度が、対応する新生抗原が1つ以上のMHCアレルによって前記対象の前記腫瘍細胞の表面上に提示される尤度を表し、前記機械学習させた提示モデルが、

訓練データセットに少なくとも基づいて特定される複数のパラメータであって、前記訓練データセットが、

複数の試料の各試料について、前記試料中に存在するものとして特定されたMHCアレルのセット内の少なくとも1つのMHCアレルに結合したペプチドの存在を測定する質量分析によって得られた標識、

前記試料のそれぞれについて、前記ペプチドを構成する複数のアミノ酸と前記ペプチド内の前記アミノ酸の位置のセットとに関する情報を含む数値ベクトルとしてコード化された訓練ペプチド配列、及び

前記試料のそれぞれについて、前記試料の前記訓練ペプチド配列のそれぞれについて、前記訓練ペプチド配列と、前記訓練ペプチド配列の前記ヌクレオチドシーケンシングデータの前記kマーブブロックのうちの1つ以上のkマーブブロックとの間の関連付け

を含み、

前記複数のパラメータのサブセットが、前記1つ以上のkマーブブロックにおける提示ホットスポットの有無を表す、前記複数のパラメータと、

入力として受け取られた前記数値ベクトル及び前記1つ以上のkマーブブロックと、前記数値ベクトル、前記1つ以上のkマーブブロック、及び前記パラメータに基づいた出力として生成された前記提示尤度との間の関係を表す関数と

を含む、

前記入力すること；

選択された新生抗原のセットを生成するために、前記新生抗原のセットのサブセットを、前記提示尤度のセットに基づいて選択すること；ならびに

前記選択された新生抗原のセットを返すこと

を行わせるコンピュータプログラム命令を格納したメモリとを含む、コンピュータシステム。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0010

【補正方法】変更

【補正の内容】

【0010】

TIL新生エピトープデータセット及び予想される新生抗原反応性T細胞の特定タスクにおける本明細書に開示されるモデルの予測性能は、HLAのプロセッシング及び提示をモデル化することによって、治療に有用な新生エピトープの予測を得ることが今や可能であることを示す。要約すると、この作業は、抗原標的化免疫療法のための実用的なインシリコの抗原特定を可能とすることにより、患者の治療に向かう進行を加速するものである。

[本発明1001]

腫瘍細胞の表面上に提示される可能性の高い、対象の1つ以上の腫瘍細胞に由来する1つ以上の新生抗原を特定するための方法であって、

前記対象の前記腫瘍細胞及び正常細胞からエクソーム、トランスクリプトーム、または全ゲノムのヌクレオチドシーケンシングデータのうちの少なくとも1つを取得する工程であって、前記ヌクレオチドシーケンシングデータが、前記腫瘍細胞からの前記ヌクレオチドシーケンシングデータと前記正常細胞からの前記ヌクレオチドシーケンシングデータとを比較することにより特定された新生抗原のセットのそれぞれのペプチド配列を表すデータを取得するために用いられ、各新生抗原のペプチド配列が、前記ペプチド配列を前記対象の前記正常細胞から特定される対応する野生型のペプチド配列とは異なるもの

とする少なくとも1つの変化を含む、前記取得する工程；

前記新生抗原のそれぞれのペプチド配列を、対応する数値ベクトルにコード化する工程であって、各数値ベクトルが、前記ペプチド配列を構成する複数のアミノ酸と前記ペプチド配列内の前記アミノ酸の位置のセットとに関する情報を含む、前記コード化する工程；

前記新生抗原のそれぞれのペプチド配列を、前記対象の前記ヌクレオチドシーケンシングデータの複数のkマーブブロックのうちの1つ以上のkマーブブロックと関連付ける工程；

前記新生抗原のセットについて提示尤度のセットを生成するために、コンピュータプロセッサを使用して、前記数値ベクトル及び前記1つ以上の関連付けられたkマーブブロックを機械学習させた提示モデルに入力する工程であって、前記セット内の各提示尤度が、対応する新生抗原が1つ以上のMHCアレルによって前記対象の前記腫瘍細胞の表面上に提示される尤度を表し、前記機械学習させた提示モデルが、

訓練データセットに少なくとも基づいて特定される複数のパラメータであって、前記訓練データセットが、

複数の試料の各試料について、前記試料中に存在するものとして特定されたMHCアレルのセット内の少なくとも1つのMHCアレルに結合したペプチドの存在を測定する質量分析によって得られた標識、

前記試料のそれぞれについて、前記ペプチドを構成する複数のアミノ酸と前記ペプチド内の前記アミノ酸の位置のセットとに関する情報を含む数値ベクトルとしてコード化された訓練ペプチド配列、及び

前記試料のそれぞれについて、前記試料の前記訓練ペプチド配列のそれぞれについて、前記訓練ペプチド配列と、前記訓練ペプチド配列の前記ヌクレオチドシーケンシングデータの複数のkマーブブロックのうちの1つ以上のkマーブブロックとの間の関連付け

を含み、

前記複数のパラメータのサブセットが、前記1つ以上のkマーブブロックにおける提示ホットスポットの有無を表す、前記複数のパラメータと、

入力として受け取られた前記数値ベクトル及び前記1つ以上のkマーブブロックと、前記数値ベクトル、前記1つ以上のkマーブブロック、及び前記パラメータに基づいた出力として生成された前記提示尤度との間の関係を表す関数と

を含む、

前記入力する工程；

選択された新生抗原のセットを生成するために、前記新生抗原のセットのサブセットを、前記提示尤度のセットに基づいて選択する工程；ならびに

前記選択された新生抗原のセットを返す工程を含む、前記方法。

[本発明1002]

前記数値ベクトルを前記機械学習させた提示モデルに入力する工程が、

前記ペプチド配列の特定の位置の特定のアミノ酸に基づいてMHCアレルが前記新生抗原を提示するかどうかを示す依存性スコアを前記1つ以上のMHCアレルのそれぞれについて生成するために、前記機械学習させた提示モデルを前記新生抗原のペプチド配列に適用すること

を含む、本発明1001の方法。

[本発明1003]

前記数値ベクトルを前記機械学習させた提示モデルに入力する工程が、

各MHCアレルについて、対応するMHCアレルが対応する新生抗原を提示するであろう尤度を示す対応するアレル毎尤度を生成するために、前記依存性スコアを変換することと、

前記アレル毎尤度を組み合わせて前記新生抗原の提示尤度を生成することとを含む、本発明1002の方法。

[本発明1004]

前記依存性スコアを変換することが、前記新生抗原の提示を前記1つ以上のMHCアレルにわたって相互排他的なものとしてモデル化する、本発明1003の方法。

[本発明1005]

前記数値ベクトルを前記機械学習させた提示モデルに入力する工程が、前記提示尤度を生成するために前記依存性スコアの組み合わせを変換することをさらに含み、前記依存性スコアの組み合わせを変換することが、前記新生抗原の提示を前記1つ以上のMHCアレル間で干渉するものとしてモデル化する、本発明1002の方法。

[本発明1006]

前記提示尤度のセットが、少なくとも1つ以上のアレル非相互作用特性によってさらに特定され、

前記アレル非相互作用特性に基づいて前記対応する新生抗原のペプチド配列が提示されるかどうかを示す前記アレル非相互作用特性についての依存性スコアを生成するために、前記機械学習させた提示モデルを前記アレル非相互作用特性に適用することをさらに含む、本発明1002～1005のいずれかの方法。

[本発明1007]

前記1つ以上のMHCアレルの各MHCアレルについての前記依存性スコアを、前記アレル非相互作用特性についての前記依存性スコアと組み合わせることと、

対応するMHCアレルが対応する新生抗原を提示する尤度を示す、各MHCアレルについてのアレル毎尤度を生成するために、各MHCアレルについての前記組み合わせられた依存性スコアを変換することと、

前記提示尤度を生成するために、前記アレル毎尤度を組み合わせることとをさらに含む、本発明1006の方法。

[本発明1008]

前記MHCアレルのそれぞれについての前記依存性スコアと、前記アレル非相互作用特性についての前記依存性スコアとを組み合わせることと、

前記提示尤度を生成するために、前記組み合わせられた依存性スコアを変換することとをさらに含む、本発明1006の方法。

[本発明1009]

前記少なくとも1つ以上のアレル非相互作用特性が、前記新生抗原のペプチド配列と、前記新生抗原の前記ヌクレオチドシーケンシングデータの複数のkマーブブロックのうち1つ以上のkマーブブロックとの間の関連付けを含む、本発明1006～1008のいずれかの方法。

[本発明1010]

前記1つ以上のMHCアレルが2つ以上の異なるMHCアレルを含む、本発明1001～1009のいずれかの方法。

[本発明1011]

前記ペプチド配列が、アミノ酸9個以外の長さを有するペプチド配列を含む、本発明1001～1010のいずれかの方法。

[本発明1012]

前記ペプチド配列をコード化することが、ワン・ホットエンコーディングスキームを用いて前記ペプチド配列をコード化することを含む、本発明1001～1011のいずれかの方法。

[本発明1013]

前記複数の試料が、

- (a) 単一のMHCアレルを発現するように操作された1つ以上の細胞株、
 - (b) 複数のMHCアレルを発現するように操作された1つ以上の細胞株、
 - (c) 複数の患者から得られた、または複数の患者に由来する1つ以上のヒト細胞株、
 - (d) 複数の患者から得られた新鮮なまたは凍結された腫瘍試料、及び
 - (e) 複数の患者から得られた新鮮なまたは凍結された組織試料
- のうちの少なくとも1つを含む、本発明1001～1012のいずれかの方法。

[本発明1014]

前記訓練データセットが、

(a) 前記ペプチドのうちの少なくとも1つについてのペプチド - M H C 結合親和性の測定値に関連するデータ、及び

(b) 前記ペプチドのうちの少なくとも1つについてのペプチド - M H C 結合安定性の測定値に関連するデータ

のうちの少なくとも1つをさらに含む、本発明1001 ~ 1013のいずれかの方法。

[本発明1015]

前記提示尤度のセットが、RNA - s e qまたは質量分析により測定される、前記対象における前記1つ以上のM H C アレルの少なくとも発現レベルによってさらに特定される、本発明1001 ~ 1014のいずれかの方法。

[本発明1016]

前記提示尤度のセットが、

(a) 前記新生抗原のセット内の新生抗原と前記1つ以上のM H C アレルとの間の予測される親和性、及び

(b) 前記新生抗原コード化ペプチド - M H C 複合体の予測される安定性

のうちの少なくとも1つを含む特性によってさらに特定される、本発明1001 ~ 1015のいずれかの方法。

[本発明1017]

前記数値的尤度のセットが、

(a) そのソースタンパク質配列内の、前記新生抗原コード化ペプチド配列に隣接するC末端配列、及び

(b) そのソースタンパク質配列内の、前記新生抗原コード化ペプチド配列に隣接するN末端配列

のうちの少なくとも1つを含む特性によってさらに特定される、本発明1001 ~ 1016のいずれかの方法。

[本発明1018]

前記選択された新生抗原のセットを選択することが、前記機械学習させた提示モデルに基づいて、選択されない新生抗原と比較して前記腫瘍細胞表面上に提示される尤度が増大している新生抗原を選択することを含む、本発明1001 ~ 1017のいずれかの方法。

[本発明1019]

前記選択された新生抗原のセットを選択することが、前記機械学習させた提示モデルに基づいて、選択されない新生抗原と比較して前記対象において腫瘍特異的な免疫応答を誘導することができる尤度が増大している新生抗原を選択することを含む、本発明1001 ~ 1018のいずれかの方法。

[本発明1020]

前記選択された新生抗原のセットを選択することが、前記提示モデルに基づいて、選択されない新生抗原と比較してプロフェッショナル抗原提示細胞 (A P C) によってナイーブT細胞に対して提示されることができる尤度が増大している新生抗原を選択することを含み、任意で、前記A P C が樹状細胞 (D C) である、本発明1001 ~ 1019のいずれかの方法。

[本発明1021]

前記選択された新生抗原のセットを選択することが、前記機械学習させた提示モデルに基づいて、選択されない新生抗原と比較して中枢性寛容または末梢性寛容によって阻害される尤度が減少している新生抗原を選択することを含む、本発明1001 ~ 1020のいずれかの方法。

[本発明1022]

前記選択された新生抗原のセットを選択することが、前記機械学習させた提示モデルに基づいて、選択されない新生抗原と比較して前記対象において正常組織に対する自己免疫応答を誘導することができる尤度が減少している新生抗原を選択することを含む、本発明

1001～1021のいずれかの方法。

[本発明1023]

前記1つ以上の腫瘍細胞が、肺癌、メラノーマ、乳癌、卵巣癌、前立腺癌、腎臓癌、胃癌、結腸癌、精巣癌、頭頸部癌、膵臓癌、脳癌、B細胞リンパ腫、急性骨髄性白血病、慢性骨髄性白血病、慢性リンパ球性白血病、T細胞リンパ球性白血病、非小細胞肺癌、及び小細胞肺癌からなる群から選択される、本発明1001～1022のいずれかの方法。

[本発明1024]

前記選択された新生抗原のセットから個別化がんワクチンを構築するための出力を生成することをさらに含む、本発明1001～1023のいずれかの方法。

[本発明1025]

前記個別化がんワクチン用の出力が、前記選択された新生抗原のセットをコードした少なくとも1つのペプチド配列または少なくとも1つのヌクレオチド配列を含む、本発明1024の方法。

[本発明1026]

前記機械学習させた提示モデルが、ニューラルネットワークモデルである、本発明1001～1025のいずれかの方法。

[本発明1027]

前記ニューラルネットワークモデルがMHCアレルについての複数のネットワークモデルを含み、それぞれのネットワークモデルが複数のMHCアレルのうちの対応するMHCアレルに割り当てられ、かつ1つ以上の層に配置された一連のノードを含む、本発明1026の方法。

[本発明1028]

前記ニューラルネットワークモデルが、前記ニューラルネットワークモデルのパラメータを更新することによって訓練され、少なくとも2つのネットワークモデルの前記パラメータが、少なくとも1回の訓練イテレーションについて一緒に更新される、本発明1027の方法。

[本発明1029]

前記機械学習させた提示モデルが、ノードの1つ以上の層を含むディープラーニングモデルである、本発明1026～1028のいずれかの方法。

[本発明1030]

前記1つ以上のMHCアレルが、クラスI MHCアレルである、本発明1001～1029のいずれかの方法。

[本発明1031]

コンピュータプロセッサと、

前記コンピュータプロセッサにより実行されると、前記コンピュータプロセッサに、対象の腫瘍細胞及び正常細胞からエクソーム、トランスクリプトーム、または全ゲノムのヌクレオチドシーケンシングデータのうちの少なくとも1つを取得することであって、前記ヌクレオチドシーケンシングデータが、前記腫瘍細胞からの前記ヌクレオチドシーケンシングデータと前記正常細胞からの前記ヌクレオチドシーケンシングデータとを比較することにより特定された新生抗原のセットのそれぞれのペプチド配列を表すデータを取得するために用いられ、各新生抗原のペプチド配列が、前記ペプチド配列を前記対象の前記正常細胞から特定される対応する野生型のペプチド配列とは異なるものとする少なくとも1つの変化を含む、前記取得すること；

前記新生抗原のそれぞれのペプチド配列を、対応する数値ベクトルにコード化することであって、各数値ベクトルが、前記ペプチド配列を構成する複数のアミノ酸と前記ペプチド配列内の前記アミノ酸の位置のセットとに関する情報を含む、前記コード化すること；

前記新生抗原のそれぞれのペプチド配列を、前記対象の前記ヌクレオチドシーケンシングデータの複数のkマーブブロックのうちの1つ以上のkマーブブロックと関連付けること；

前記新生抗原のセットについて提示尤度のセットを生成するために、前記数値ベクトル及び前記1つ以上の関連付けられたkマープロックを機械学習させた提示モデルに入力することであって、前記セット内の各提示尤度が、対応する新生抗原が1つ以上のMHCアレルによって前記対象の前記腫瘍細胞の表面上に提示される尤度を表し、前記機械学習させた提示モデルが、

訓練データセットに少なくとも基づいて特定される複数のパラメータであって、前記訓練データセットが、

複数の試料の各試料について、前記試料中に存在するものとして特定されたMHCアレルのセット内の少なくとも1つのMHCアレルに結合したペプチドの存在を測定する質量分析によって得られた標識、

前記試料のそれぞれについて、前記ペプチドを構成する複数のアミノ酸と前記ペプチド内の前記アミノ酸の位置のセットとに関する情報を含む数値ベクトルとしてコード化された訓練ペプチド配列、及び

前記試料のそれぞれについて、前記試料の前記訓練ペプチド配列のそれぞれについて、前記訓練ペプチド配列と、前記訓練ペプチド配列の前記ヌクレオチドシーケンシングデータの前記kマープロックのうちの1つ以上のkマープロックとの間の関連付け

を含み、

前記複数のパラメータのサブセットが、前記1つ以上のkマープロックにおける提示ホットスポットの有無を表す、前記複数のパラメータと、

入力として受け取られた前記数値ベクトル及び前記1つ以上のkマープロックと、前記数値ベクトル、前記1つ以上のkマープロック、及び前記パラメータに基づいた出力として生成された前記提示尤度との間の関係を表す関数と

を含む、

前記入力すること；

選択された新生抗原のセットを生成するために、前記新生抗原のセットのサブセットを、前記提示尤度のセットに基づいて選択すること；ならびに

前記選択された新生抗原のセットを返すこと

を行わせるコンピュータプログラム命令を格納したメモリとを含む、コンピュータシステム。