

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B1)

(11) 特許番号

特許第6500144号  
(P6500144)

(45) 発行日 平成31年4月10日(2019.4.10)

(24) 登録日 平成31年3月22日(2019.3.22)

(51) Int.Cl.		F I			
<b>G 1 6 B</b>	<b>40/00</b>	<b>(2019.01)</b>	G O 6 F	19/24	Z N A
<b>C 1 2 Q</b>	<b>1/02</b>	<b>(2006.01)</b>	C 1 2 Q	1/02	
<b>G O 1 N</b>	<b>33/53</b>	<b>(2006.01)</b>	G O 1 N	33/53	
<b>A 6 1 K</b>	<b>39/00</b>	<b>(2006.01)</b>	A 6 1 K	39/00	G

請求項の数 21 (全 52 頁)

(21) 出願番号	特願2018-61901 (P2018-61901)	(73) 特許権者	518092159
(22) 出願日	平成30年3月28日(2018.3.28)		K O T A I バイオテクノロジーズ株式会社
(62) 分割の表示	特願2018-49440 (P2018-49440) の分割		大阪府大阪市中央区東高麗橋2番5-90 5号 グラン・コープ東高麗橋
原出願日	平成30年3月16日(2018.3.16)	(74) 代理人	100078282
審査請求日	平成30年3月28日(2018.3.28)		弁理士 山本 秀策
早期審査対象出願		(74) 代理人	100113413
			弁理士 森下 夏樹
		(74) 代理人	100118371
			弁理士 ▲胸▼谷 剛志
		(74) 代理人	100181674
			弁理士 飯田 貴敏
		(74) 代理人	100181641
			弁理士 石川 大輔

最終頁に続く

(54) 【発明の名称】 免疫実体の効率的クラスタリング

(57) 【特許請求の範囲】

【請求項1】

(i) 少なくとも2つの免疫実体 (immunological entity) の特徴量を提供するステップであって、該ステップは該少なくとも2つの免疫実体の三次元構造モデルから特徴量を計算することを除く、ステップと、

(ii) 該特徴量に基づいて、抗原特異性または結合モードを特定せずに該免疫実体の抗原特異性または結合モードの分析を機械学習させるステップと、

(iii) 該抗原特異性または結合モードの分類または異同の決定を行うステップとを含む、免疫実体の集合を解析する方法。

【請求項2】

免疫実体の集合を解析する方法であって、該方法は：

(a) 該免疫実体の集合のメンバーの少なくとも1つの対について特徴量を抽出するステップであって、該ステップは該少なくとも1つの対の三次元構造モデルから特徴量を計算することを除く、ステップと、

(b) 該特徴量を用いた機械学習により該対について抗原特異性または結合モードの間の距離を算出し、または該抗原特異性または結合モードが一致するかどうかを判定するステップと、

(c) 該距離に基づいて該免疫実体の集合をクラスタリングするステップと、

(d) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを含む、方法。

## 【請求項 3】

免疫実体の集合を解析する方法であって、該方法は：

( a a ) 該免疫実体の集合のメンバーの少なくとも 1 つの対をなす配列それぞれについて特徴量を抽出するステップであって、該ステップは該少なくとも 1 つの対をなす配列の免疫実体の三次元構造モデルから特徴量を計算することを除く、ステップと、

( b b ) 該特徴量を高次元ベクトル空間に射影し、ここで、該メンバーの空間上の距離は該メンバーの機能類似性を反映する、ステップと、

( c c ) 該距離に基づいて該免疫実体の集合をクラスタリングするステップと、

( d d ) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを包含する、方法。

10

## 【請求項 4】

前記特徴量は配列情報、CDR 1 - 3 配列の長さ、配列一致度、フレームワーク領域の配列一致度、分子の全電荷 / 親水性 / 疎水性 / 芳香族アミノ酸の数、各 CDR、フレームワーク領域の電荷 / 親水性 / 疎水性 / 芳香族アミノ酸の数、各アミノ酸の数、重鎖 - 軽鎖の組み合わせ、体細胞変異数、変異の位置、アミノ酸モチーフの存在 / 一致度、参照配列セットに対する希少度、および参照配列による結合 HLA のオッズ比からなる群より選択される少なくとも 1 つを含む、請求項 1 または 2 に記載の方法。

## 【請求項 5】

前記免疫実体は抗体、抗体の抗原結合断片、B 細胞受容体、B 細胞受容体の断片、T 細胞受容体、T 細胞受容体の断片、キメラ抗原受容体 (CAR)、またはこれらのいずれかまたは複数を含む細胞である、請求項 1、2 または 4 のいずれか一項に記載の方法。

20

## 【請求項 6】

前記機械学習による計算は前記特徴量を入力とし、ランダムフォレストまたはブースティングで行い、

前記クラスタリングは結合距離に基づく単純な閾値に基づくもの、階層的クラスタリング、あるいは非階層的クラスタリング法で行う、

請求項 2 に記載の方法。

## 【請求項 7】

前記解析は

バイオマーカーの同定、あるいは治療ターゲットとなる免疫実体または該免疫実体を含む細胞の同定のいずれか 1 つまたは複数を包含、請求項 2 または 6 に記載の方法。

30

## 【請求項 8】

前記機械学習は、回帰的な手法、ニューラルネットワーク法、サポートベクトルマシン、およびランダムフォレスト等の機械学習アルゴリズムからなる群より選択される、請求項 1 ~ 2、4 ~ 7 のいずれか一項に記載の方法。

## 【請求項 9】

前記特徴量は配列情報、CDR 1 - 3 配列の長さ、配列一致度、フレームワーク領域の配列一致度、分子の全電荷 / 親水性 / 疎水性 / 芳香族アミノ酸の数、各 CDR、フレームワーク領域の電荷 / 親水性 / 疎水性 / 芳香族アミノ酸の数、各アミノ酸の数、重鎖 - 軽鎖の組み合わせ、体細胞変異数、変異の位置、アミノ酸モチーフの存在 / 一致度、参照配列セットに対する希少度、および参照配列による結合 HLA のオッズ比からなる群より選択される少なくとも 1 つを含む、請求項 3 に記載の方法。

40

## 【請求項 10】

前記免疫実体は抗体、抗体の抗原結合断片、B 細胞受容体、B 細胞受容体の断片、T 細胞受容体、T 細胞受容体の断片、キメラ抗原受容体 (CAR)、またはこれらのいずれかまたは複数を含む細胞である、請求項 3 または 9 に記載の方法。

## 【請求項 11】

前記高次元ベクトル空間に射影するステップ ( b b ) は教師あり、半教師あり ( Siamese network )、または教師なし ( Auto-encoder ) のいずれかの方法で行い、前記クラスタリングするステップ ( c c ) は高次元空間上の距離に基づく

50

単純な閾値に基づくもの、階層的クラスタリング、あるいは非階層的クラスタリング法で行う、

で行う、請求項 3、9 または 10 のいずれかに記載の方法。

【請求項 12】

前記解析は

バイオマーカーの同定、あるいは治療ターゲットとなる免疫実体または該免疫実体を含む細胞の同定のいずれか 1 つまたは複数を含む、請求項 3 または 9 ~ 11 のいずれかに記載の方法。

【請求項 13】

請求項 1 ~ 12 のいずれか一項に記載の方法をコンピュータに実行させるプログラム。 10

【請求項 14】

請求項 1 ~ 12 のいずれか一項に記載の方法をコンピュータに実行させるプログラムを格納した記録媒体。

【請求項 15】

請求項 1 ~ 12 のいずれか一項に記載の方法をコンピュータに実行させるプログラムを含むシステム。

【請求項 16】

前記抗原特異性または結合モードについて、生体情報と関連付ける工程を包含するステップを包含する、請求項 1 ~ 2、または 4 ~ 7 のいずれか一項に記載の方法。

【請求項 17】

請求項 1 ~ 2、または 4 ~ 7 のいずれか一項に記載の方法を用いて、抗原特異性または結合モードが同一である免疫実体を同一のクラスターに分類する工程を包含する、抗原特異性または結合モードのクラスターを生成する方法。

【請求項 18】

請求項 17 に記載の方法で生成されたクラスターに基づき、前記免疫実体の保有者を既知の疾患または障害あるいは生体の状態と関連付ける工程を包含する、疾患または障害あるいは生体の状態を同定する方法。

【請求項 19】

前記機能類似性について、生体情報と関連付ける工程を包含するステップを包含する、請求項 3、または 8 ~ 11 のいずれか一項に記載の方法。

【請求項 20】

請求項 3、または 8 ~ 11 のいずれか一項に記載の方法を用いて、前記機能類似性が同一である免疫実体を同一のクラスターに分類する工程を包含する、機能類似性のクラスターを生成する方法。

【請求項 21】

請求項 20 に記載の方法で生成されたクラスターに基づき、前記免疫実体の保有者を既知の疾患または障害あるいは生体の状態と関連付ける工程を包含する、疾患または障害あるいは生体の状態を同定する方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、抗体等の免疫実体をエピトープに基づいて分類する方法、エピトープクラスターの作成およびその応用に関する。

【背景技術】

【0002】

抗体は抗原に特異的にかつ高親和性で結合するタンパク質である。ヒト抗体は重鎖、軽鎖と呼ばれる 2 つの高分子配列からなる。重鎖、軽鎖はそれぞれさらに可変領域と定常領域という 2 つの領域に分けられる。そして、この可変領域は、抗体の生理活性に重要な多様性をもたらすことがわかっている。この可変領域はさらにフレームワーク領域と相補性決定領域 (CDR) に分けられる。抗体はターゲットとして結合する分子を抗原という。 50

抗体は一般的にCDRが抗原と物理的に相互作用することにより、抗原を特異的にまた高親和性で結合する。抗原において抗体と物理的に相互作用する領域を「エピトープ」と呼ぶ。

#### 【0003】

抗体は非常に多様性に富んでいる。各個人は $10^{11}$ ものアミノ酸配列の異なる抗体を作り出すことができる。この多様性によってB細胞レパトアは多様な抗原と、さらに同じ抗原の異なるエピトープと異なる親和性で結合することができる。CDR領域のアミノ酸配列が多様性の源泉である。CDRの中でも重鎖の3番目のループ(CDR-H3)が最も多様性に富む。複数のアミノ酸配列の非常に異なる抗体が同一のまたは非常に似たエピトープと結合することがある。この「配列の縮退」によって抗体、とりわけ別々の個体によって作られた抗体を抗原やエピトープによって比較することは非常に難しい。

10

#### 【0004】

抗体は商業的に非常に価値のある分子で、現在最も商業的に成功している薬の多くが抗体医薬である。さらに、抗体医薬は製薬業界において最も急速に成長している分野である。抗体は高親和性と特異性という特長を生かし、医療用だけでなく、基礎研究や製薬以外の産業界においても広く利用されている。

#### 【0005】

T細胞もまた、B細胞と構造的によく似た受容体(TCR)を発現する。重要な違いは、TCRは可溶性ではなく、常にT細胞に結合している点である(B細胞は可溶性の受容体である抗体と、細胞膜に結合したBCRとを産生する。)。BCRほどの多様性はないものの、T細胞もこれまで非常によく研究されてきた。とりわけ、悪性腫瘍に対する作用では細胞傷害性T細胞による細胞破壊が重要である。

20

#### 【0006】

近年、抗体やTCRのアミノ酸配列を次世代シーケンシング技術によって大規模に同定することが可能になった。他方、それらの抗体、TCRと結合する抗原やエピトープの同定が課題であり、商業的にも大きな需要が期待される。

#### 【0007】

既存の抗原同定方法は、抗体やTCRを1つまたは複数の抗原候補と相互作用させ、実験的に相互作用を同定する方法である(例えば、表面プラズモン共鳴)。これに変わる技術としては、プロテインチップや、各種のライブラリ法がある。これらは比較的安価で高速であるが、関節リウマチ等幾つかの疾患で重要な翻訳後修飾を受けたタンパク質やペプチドに対しては適用することができない。また、構造エピトープの同定は困難である。

30

#### 【0008】

これらの実験的なスクリーニング技術は、抗原が同定されている必要がある。言い換えると、抗体、TCRの発見の前に抗原が同定されていなくてはならない。

#### 【0009】

非特許文献1は、残基ペアリング優先度およびクロスブロッキング法を用いる抗体特異的B細胞エピトープを予測する計算法を開示する。

#### 【先行技術文献】

#### 【非特許文献】

40

#### 【0010】

【非特許文献1】Sela-Culang I. et al., Structure 22, 646 - 657, 2014

#### 【発明の概要】

#### 【課題を解決するための手段】

#### 【0011】

本発明者らは、機能をあらかじめ規定しないで、通常、免疫実体結合物(抗原やエピトープ)を個々に別の「機能(=例えば、抗原Aに対する特異性を持つかどうか)」として取り扱われる抗原特異性または結合モードというものに一般性があると仮定し、その類似度を評価することで、免疫実体を分類することができることを見出した。これにより、従来知られていない機能(例えば、抗原特異性または結合モード)に対しても適用可能である

50

。従って、通常抗体抗原反応などの免疫実体の反応にいう機能(例えば、特定の抗原特異性または結合モード)を予め特定しないことにより、本発明は一般化することができる。好ましい実施形態では、「機能」は特定の抗原特異性または結合モード(抗原制御能)である。本発明において、機能をあらかじめ特定しないことは、学習セットに様々な抗原に対するものを含ませることができ、それらを各機能ごとに類似度を推測することに反映させることができる。

【0012】

本発明は、以下を提供する

(1) (i) 少なくとも2つの免疫実体 (immunological entity) の特徴量を提供するステップと、

10

(ii) 該特徴量に基づいて、抗原特異性または結合モードを特定せずに該免疫実体の抗原特異性または結合モードの分析を機械学習させるステップと、

(iii) 該抗原特異性または結合モードの分類または異同の決定を行うステップとを含む、免疫実体の集合を解析する方法。

(2) 免疫実体の集合を解析する方法であって、該方法は：

(a) 該免疫実体の集合のメンバーの少なくとも1つの対について特徴量を抽出するステップと、

(b) 該特徴量を用いた機械学習により該対について抗原特異性または結合モードの間の距離を算出し、または該抗原特異性または結合モードが一致するかどうかを判定するステップと、

20

(c) 該距離に基づいて該免疫実体の集合をクラスタリングするステップと、

(d) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを含む、方法。

(3) 免疫実体の集合を解析する方法であって、該方法は：

(aa) 該免疫実体の集合のメンバーの少なくとも1つの対をなす配列それぞれについて特徴量を抽出するステップと、

(bb) 該特徴量を高次元ベクトル空間に射影し、ここで、該メンバーの空間上の距離は該メンバーの機能類似性を反映する、ステップと、

(cc) 該距離に基づいて該免疫実体の集合をクラスタリングするステップと、

(dd) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを含む、方法。

30

(4) 前記特徴量は配列情報、CDR1-3配列の長さ、配列一致度、フレームワーク領域の配列一致度、分子の全電荷/親水性/疎水性/芳香族アミノ酸の数、各CDR、フレームワーク領域の電荷/親水性/疎水性/芳香族アミノ酸の数、各アミノ酸の数、重鎖-軽鎖の組み合わせ、体細胞変異数、変異の位置、アミノ酸モチーフの存在/一致度、参照配列セットに対する希少度、および参照配列による結合HLAのオッズ比からなる群より選択される少なくとも1つを含む、前記項目のいずれか一項に記載の方法。

(5) 前記免疫実体は抗体、抗体の抗原結合断片、B細胞受容体、B細胞受容体の断片、T細胞受容体、T細胞受容体の断片、キメラ抗原受容体(CAR)、またはこれらのいずれかまたは複数を含む細胞である、前記項目のいずれか一項に記載の方法。

40

(6) 前記機械学習による計算は前記特徴量を入力とし、ランダムフォレストまたはブースティングで行い、前記クラスタリングは結合距離に基づく単純な閾値に基づくもの、階層的クラスタリング、あるいは非階層的クラスタリング法で行う、前記項目のいずれか一項に記載の方法。

(7) 前記解析は、バイオマーカーの同定、あるいは治療ターゲットとなる免疫実体または該免疫実体を含む細胞の同定のいずれか1つまたは複数を含む、前記項目のいずれか一項に記載の方法。

(8) 前記高次元ベクトル空間計算(bb)は教師あり、半教師あり(Siamese network)、または教師なし(Auto-encoder)のいずれかの方法で行い、

50

前記クラスタリング ( c c ) は高次元空間上の距離に基づく単純な閾値に基づくもの、階層的クラスタリング、あるいは非階層的クラスタリング法で行う、前記項目のいずれか一項に記載の方法。

( 9 ) 前記解析 ( d d ) はバイオマーカーの同定、あるいは治療ターゲットとなる免疫実體または該免疫実體を含む細胞の同定のいずれか1つまたは複数を含む、前記項目のいずれか一項に記載の方法。

( 10 ) 前記機械学習は、回帰的な手法、ニューラルネットワーク法、サポートベクトルマシン、およびランダムフォレスト等の機械学習アルゴリズムからなる群より選択される、前記項目のいずれか一項に記載の方法。

( 11 ) 前記項目のいずれか一項に記載の方法をコンピュータに実行させるプログラム。

( 12 ) 前記項目のいずれか一項に記載の方法をコンピュータに実行させるプログラムを格納した記録媒体。

( 13 ) 前記項目のいずれか一項に記載の方法をコンピュータに実行させるプログラムを含むシステム

( 14 ) 前記抗原特異性または結合モードについて、生体情報と関連付ける工程を包含するステップを包含する、前記項目のいずれか一項に記載の方法。

( 15 ) 前記項目のいずれか一項に記載の方法を用いて、抗原特異性または結合モードが同一である免疫実體を同一のクラスターに分類する工程を包含する、抗原特異性または結合モードのクラスターを生成する方法。

( 16 ) 前記項目のいずれか一項に記載の方法で生成されたクラスターに基づき、前記免疫実體の保有者を既知の疾患または障害あるいは生体の状態と関連付ける工程を包含する、疾患または障害あるいは生体の状態を同定する方法。

( 17 ) 前記項目のいずれか一項に基づいて同定された抗原特異性または結合モードを有する免疫実體を含む、前記生体情報の同定のための組成物。

( 18 ) 前記項目のいずれか一項に記載の方法に基づいて同定された抗原特異性または結合モードを有する免疫実體を含む、疾患または障害あるいは生体の状態を診断するための組成物。

( 19 ) 前記項目のいずれか一項に記載の方法に基づいて同定されたエピトープに対する免疫実體を含む、疾患または障害あるいは生体の状態を治療または予防するための組成物。

( 20 ) 前記組成物はワクチンを含む、前記項目のいずれか一項に記載の組成物。

( 21 ) 前記項目のいずれか一項に記載の方法で同定された抗原特異性または結合モードを有する構造を有する免疫実體 (例えば、抗体)、エピトープまたは免疫実體結合物 (例えば、抗原)。

( 22 ) 前記免疫実體、エピトープまたは免疫実體結合物について、生体情報と関連付ける工程を包含するステップを包含する、前記項目のいずれかに記載の方法。

( 23 ) 前記クラスタリング、分類または解析した免疫実體、エピトープまたは免疫実體結合物を同定する工程をさらに包含する、前記項目のいずれかに記載の方法。

( 24 ) 前記同定は、アミノ酸配列の決定、三次元構造の同定、三次元構造以外の構造上の同定、および生物学的機能の同定からなる群より選択される少なくとも1つを含む、前記項目のいずれか一項に記載の方法。

( 25 ) 前記同定は、前記免疫実體、エピトープまたは免疫実體結合物の構造を決定することを含む、前記項目のいずれか一項に記載の方法。

( 26 ) 前記項目のいずれか一項に記載の分類方法を用いて、抗原特異性または結合モードが同一である免疫実體、エピトープまたは免疫実體結合物を同一のクラスターに分類する工程を包含する、免疫実體、エピトープまたは免疫実體結合物のクラスターを生成する方法。

( 27 ) 前記免疫実體、エピトープまたは免疫実體結合物を、その特性および既知の免疫実體、エピトープまたは免疫実體結合物との類似性からなる群より選択される少なくとも1つの評価項目を評価し、所定の基準を満たした免疫実體を対象に前記クラスター分類を

10

20

30

40

50

行うことを特徴とする、前記項目のいずれか一項に記載の方法。

(28) 前記項目のいずれか一項に記載の方法で生成クラスターに基づき同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物の保有者を既知の疾患または障害あるいは生体の状態と関連付ける工程を包含する、疾患または障害あるいは生体の状態の同定法。

(29) 前記項目のいずれか一項に記載の方法で生成されたクラスターを一つまたは複数用いて、該クラスターの保有者の疾患または障害あるいは生体の状態を評価する工程を含む、疾患または障害あるいは生体の状態の同定法。

(30) 前記評価は、前記複数のクラスターの存在量の順位および/または存在比に基づく分析、または一定数のB細胞を調べ、その中に興味あるBCRと類似のもの/クラスターがあるかどうかという定量による分析からなる群より選択される少なくとも1つの指標を用いてなされる、前記項目のいずれか一項に記載の方法。

10

(31) 前記評価は、前記クラスター以外の指標も用いてなされる、前記項目のいずれか一項に記載の方法。

(32) 前記クラスター以外の指標は、疾患関連遺伝子、疾患関連遺伝子の多型、疾患関連遺伝子の発現プロファイル、エピジェネティクス解析、TCRおよびBCRのクラスターの組合せから選択される少なくとも1つを含む、前記項目のいずれか一項に記載の方法。

(33) 前記疾患または障害あるいは生体の状態の同定は、前記疾患または障害あるいは生体の状態の診断、予後、薬力学、予測、代替法の決定、患者層の特定、安全性の評価、毒性の評価、およびこれらのモニタリングからなる群より選択される少なくとも1つを含む、前記項目のいずれか一項のいずれかに記載の方法。

20

(34) 前記項目のいずれか一項に記載の方法で同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物、および/または前記項目のいずれか一項に記載の方法で生成されたクラスターを一つまたは複数用いて、疾患または障害あるいは生体の状態の指標となるバイオマーカーの評価を行う工程を含む、該バイオマーカーの評価のための方法。

(35) 前記項目のいずれか一項に記載の方法で同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物、および/または前記項目のいずれか一項に記載の方法で生成されたクラスターを一つまたは複数用いて、疾患または障害あるいは生体の状態との関連付け、バイオマーカーを決定する工程を含む、該バイオマーカーの同定のための方法。

30

(36) 前記項目のいずれか一項に基づいて同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物に対する免疫実体を含む、前記生体情報の同定のための組成物。

(37) 前記項目のいずれか一項に基づいて同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物またはそれを含む免疫実体結合物(例えば、抗原)を含む、前記生体情報の同定のための組成物。

(38) 前記項目のいずれか一項に基づいて同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物を含む、前記項目のいずれか一項に記載の疾患または障害あるいは生体の状態を診断するための組成物。

40

(39) 前記項目のいずれか一項に基づいて同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物を標的とする物質を含む、前記項目のいずれか一項に記載の疾患または障害あるいは生体の状態を診断するための組成物。

(40) 前記項目のいずれか一項に基づいて同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物を含む、前記項目のいずれか一項に記載の疾患または障害あるいは生体の状態を診断するための組成物。

(41) 前記項目のいずれか一項のいずれかに記載の方法に基づいて同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物を含む、前記項目のいずれか一項に記載の疾患または障害あるいは生体の状態を治療または予防するため

50

の組成物。

(42) 前記免疫実体は、抗体、抗体の抗原結合断片、T細胞受容体、T細胞受容体の断片、B細胞受容体、B細胞受容体の断片、キメラ抗原受容体(CAR)、これらのいずれかまたは複数を含む細胞(例えば、キメラ抗原受容体(CAR)を含むT細胞)からなる群より選択される、前記項目のいずれか一項に記載の組成物。

(43) 前記項目のいずれか一項に基づいて同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物を標的とする物質を含む、前記項目のいずれか一項に記載の疾患または障害あるいは生体の状態を予防または治療するための組成物。

(44) 前記項目のいずれか一項に基づいて同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物を有する、前記項目のいずれか一項に記載の疾患または障害あるいは生体の状態を治療または予防するための組成物。

(45) 前記組成物はワクチンを含む、前記項目のいずれか一項に記載の組成物。

(46) 前記項目のいずれか一項に基づいて同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物を有する、疾患または障害あるいは生体の状態を予防または治療するためのワクチンを評価するための組成物。

(47) 免疫実体の集合を解析する方法をコンピュータに実行させるコンピュータプログラムであって、該方法は、

(i) 少なくとも2つの免疫実体(immunological entity)の特徴量を提供するステップと、

(ii) 該特徴量に基づいて、抗原特異性または結合モードを特定せずに該免疫実体の抗原特異性または結合モードの分析を機械学習させるステップと、

(iii) 該抗原特異性または結合モードの分類または異同の決定を行うステップとを包含する、プログラム。

(48) 免疫実体の集合を解析する方法をコンピュータに実行させるコンピュータプログラムであって、該方法は、

(a) 該免疫実体の集合のメンバーの少なくとも1つの対について特徴量を抽出するステップと、

(b) 該特徴量を用いた機械学習により該対について抗原特異性または結合モードの間の距離を算出し、または該抗原特異性または結合モードが一致するかどうかを判定するステップと、

(c) 該距離に基づいて該免疫実体の集合をクラスタリングするステップと、

(d) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを包含する、プログラム。

(49) 免疫実体の集合を解析する方法をコンピュータに実行させるコンピュータプログラムであって、該方法は、

(aa) 該免疫実体の集合のメンバーの少なくとも1つの対をなす配列それぞれについて特徴量を抽出するステップと、

(bb) 該特徴量を高次元ベクトル空間に射影し、ここで、該メンバーの空間上の距離は該メンバーの機能類似性を反映する、ステップと、

(cc) 該距離に基づいて該免疫実体の集合をクラスタリングするステップと、

(dd) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを包含する、プログラム。

(50) 前記項目に記載される1つまたは複数の特徴をさらに含む、前記項目のいずれか一項に記載のプログラム。

(51) 免疫実体の集合を解析する方法をコンピュータに実行させるコンピュータプログラムを格納した記録媒体であって、該方法は、

(i) 少なくとも2つの免疫実体(immunological entity)の特徴量を提供するステップと、

(ii) 該特徴量に基づいて、抗原特異性または結合モードを特定せずに該免疫実体の抗

10

20

30

40

50

原特異性または結合モードの分析を機械学習させるステップと、

( i i i ) 該抗原特異性または結合モードの分類または異同の決定を行うステップとを包含する、記録媒体。

( 5 2 ) 免疫実体の集合を解析する方法をコンピュータに実行させるコンピュータプログラムを格納した記録媒体であって、該方法は、

( a ) 該免疫実体の集合のメンバーの少なくとも1つの対について特徴量を抽出するステップと、

( b ) 該特徴量を用いた機械学習により該対について抗原特異性または結合モードの間の距離を算出し、または該抗原特異性または結合モードが一致するかどうかを判定するステップと、

10

( c ) 該距離に基づいて該免疫実体の集合をクラスタリングするステップと、

( d ) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを包含する、記録媒体。

( 5 3 ) 免疫実体の集合を解析する方法をコンピュータに実行させるコンピュータプログラムを格納した記録媒体であって、該方法は、

( a a ) 該免疫実体の集合のメンバーの少なくとも1つの対をなす配列それぞれについて特徴量を抽出するステップと、

( b b ) 該特徴量を高次元ベクトル空間に射影し、ここで、該メンバーの空間上の距離は該メンバーの機能類似性を反映する、ステップと、

( c c ) 該距離に基づいて該免疫実体の集合をクラスタリングするステップと、

20

( d d ) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを包含する、記録媒体。

( 5 4 ) 前記項目に記載される1つまたは複数の特徴をさらに含む、前記項目のいずれか一項に記載の記録媒体。

( 5 5 ) 免疫実体の集合を解析するシステムであって、該システムは、

( I ) 少なくとも2つの免疫実体 ( i m m u n o l o g i c a l e n t i t y ) の特徴量を提供する特徴量提供部と、

( I I ) 該特徴量に基づいて、抗原特異性または結合モードを特定せずに該免疫実体の抗原特異性または結合モードの分析を機械学習させる機械学習部と、

( I I I ) 該抗原特異性または結合モードの分類または異同の決定を行う分類部とを包含する、システム。

30

( 5 6 ) 免疫実体の集合を解析するシステムであって、該システムは、

( A ) 該免疫実体の集合のメンバーの少なくとも1つの対について特徴量を抽出する特徴量提供部と、

( B ) 該特徴量を用いた機械学習により該対について抗原特異性または結合モードの間の距離を算出し、または該抗原特異性または結合モードが一致するかどうかを判定する判定部と、

( C ) 該距離に基づいて該免疫実体の集合をクラスタリングするクラスタリング部と、

( D ) 必要に応じて該クラスタリングによる分類に基づいて解析する解析部とを包含する、システム。

40

( 5 7 ) 免疫実体の集合を解析するシステムであって、該システムは、

( A ) 該免疫実体の集合のメンバーの少なくとも1つの対をなす配列それぞれについて特徴量を抽出する特徴量提供部と、

( B ' ) 該特徴量を高次元ベクトル空間に射影し、ここで、該メンバーの空間上の距離は該メンバーの機能類似性を反映する、ステップと、

( C ) 該距離に基づいて該免疫実体の集合をクラスタリングするクラスタリング部と、

( D ) 必要に応じて該クラスタリングによる分類に基づいて解析する解析部と

を包含する、システム。

( 5 8 ) 前記項目に記載される1つまたは複数の特徴をさらに含む、前記項目のいずれか一項に記載のシステム。

50

## 【 0 0 1 3 】

抗体、T C Rの発見の前に抗原などの免疫実体結合物が同定される必要がないことは本発明のクラスタリングアルゴリズムの重要なアドバンテージである。本発明の技術は抗原などの免疫実体結合物に対する事前の知識を必要としない。本発明の技術の魅力的な応用の一つとしては、抗体、T C Rクラスターを病気のバイオマーカー、創薬ターゲット候補の同定、抗体医薬、キメラ抗原受容体として遺伝子改変T細胞治療に利用することである。例えば、ある種の白血病やリンパ腫ではB C RおよびT C Rが典型的な配列パターンを示すことが知られており、抗原などの免疫実体結合物がわかっていなくても、それを同定することで病気の診断に用いることができる。

## 【 0 0 1 4 】

本発明において、上記の1つまたは複数の特徴は、明示された組み合わせに加え、さらに組み合わせて提供され得ることが意図される。本発明のなおさらなる実施形態および利点は、必要に応じて以下の詳細な説明を読んで理解すれば、当業者に認識される。

## 【発明の効果】

## 【 0 0 1 5 】

抗体やT C Rをエピトープごとにクラスタリングすることは実際的に大きな効果を生む。とりわけ、免疫実体結合物（例えば、抗原）、抗原特異性、結合モードまたはエピトープごとに分けられたクラスターそれ自体が、免疫実体結合物（例えば、抗原）が同定されていなくても、価値のあるものである。このようなクラスタリングは幾つかの直接的な利益がある。例えば、別々の個体からの抗体、T C Rレパトアの比較が可能になる（例：ドナーXはドナーYと比較して、クラスターZの発現が多い。）。また、疾患特異的、新規免疫実体結合物（例えば、抗原）やエピトープの発見の可能性もある。新規免疫実体結合物（例えば、抗原）の発見は創薬において極めて価値がある。加えて、興味あるエピトープに対する抗体の定量的評価ができる。既存のプロテインチップと組み合わせることで、より定量的かつ、高解像度・高精度な情報が得られる。さらに言えば、下流の解析を容易化、低コスト化できる。例えば、N個のB C RまたはT C Rをスクリーニングするのではなく、N個がM個（ $N > M$ ）のクラスターに含まれているのであれば、M個のスクリーニングで済ませることができる。さらにまた、免疫実体結合物（例えば、抗原）または抗原特異性、結合モードもしくはエピトープ既知のB C R、T C Rを用いたバーチャルスクリーニング（類似性探索による、免疫実体結合物（例えば、抗原）、エピトープの推定）ができる。実験的なスクリーニングと相補的な技術になることも特徴であるといえる。

## 【 0 0 1 6 】

異なるアミノ酸配列を持つ抗体が同一のエピトープを認識し得、あるいは同一の抗原特異性もしくは結合モードを有し得るので、既存のバイオインフォマティクスツール、例えば配列アライメントは、エピトープごとの抗体のクラスタリングには妥当な手法とはいえない。また、構造バイオインフォマティクスにおいてはいわゆるタンパク質複合体構造を予測するドッキングや既知のタンパク質複合体の界面との類似性に基づいて複合体構造を予測する手法があるが、これらもエピトープごとの抗体のクラスタリングには妥当な手法とはいえない。T C Rも同様の問題があるが、さらに免疫実体結合物（例えば、抗原）が1次元的なペプチドとそれを提示する分子であるM H Cとの複合体であり、M H Cそれ自体も多様であることが問題を複雑にしている。それゆえ、抗体やT C Rを頑強な手法でエピトープ、抗原特異性もしくは結合モードごとにクラスタリングできる手法はこれまでの手法では不可能であった重要な発明である。

## 【図面の簡単な説明】

## 【 0 0 1 7 】

【図1A】図1Aは、本発明の実施形態を例示するフローチャートである。左は対ごとに評価する場合を示し、右は全体から評価する場合を示す。データセットの種類に応じた射影は、あらかじめ各配列間の距離がわかっている場合には、(学習)あらかじめ、(例えば)左の方法で各配列間の(抗原特異性の意味での)距離が分かっている場合、(e.g. ニューラルネットワークを用いて)その配列間の距離を再現するような多次元空間上のベクトルと

10

20

30

40

50

して配列を射影する。特徴量は各配列から任意のものを抽出し、ニューラルネットワークの入力とする。(予測)上記で学習されたモデルに配列から特徴量を抽出したものを入力すると、予測結果が得られる。また、あらかじめ各配列の抗原特異性の正誤が分かっている場合には、(e.g.ニューラルネットワークを用いて)抗原特異性の正の配列対は近く、誤の配列対は近くなるような高次元空間への射影を行う。この時、ニューラルネットワークの入力は各配列から抽出した任意の特徴量ベクトルであり、各配列間の高次元空間における距離に応じて学習を行い最適なモデルを構成する。(予測)上記で学習されたモデルに配列から特徴量を抽出したものを入力すると、予測結果が得られる。

【図1B】図1Bは、テストセットに対するBCRクラスタリングの結果を示す。ノードが各PDB構造を表し、予測の結果、エッジが同じ抗原特異性を持つと判定されたものである。

10

【図2】図2は、20のエピトープを認識するTCRのクラスタリング結果を示す。

【図3】図3は、EBV由来エピトープ(FLRGGRAYGL(配列番号1))特異的TCRのクラスタリング結果(右)と対応する結晶構造(左:PDBより得た構造を重ね書きしたもの)を示す。

【図4】図4は、2種類のHIV由来ペプチド特異的TCRとデータベース上のTCRとのクラスタリング結果を示す。

【図5】図5は、本発明のシステムの模式図である。

【図6】図6は、本発明を実施するフローチャート例の模式図である。左は対ごとに評価する場合を示し、右は全体から評価する場合を示す。

20

【発明を実施するための形態】

【0018】

以下、本発明を最良の形態を示しながら説明する。本明細書の全体にわたり、単数形の表現は、特に言及しない限り、その複数形概念をも含むことが理解されるべきである。従って、単数形の冠詞(例えば、英語の場合は「a」、「an」、「the」など)は、特に言及しない限り、その複数形概念をも含むことが理解されるべきである。また、本明細書において使用される用語は、特に言及しない限り、当該分野で通常用いられる意味で用いられることが理解されるべきである。したがって、他に定義されない限り、本明細書中で使用される全ての専門用語および科学技術用語は、本発明の属する分野の当業者によって一般的に理解されるのと同じ意味を有する。矛盾する場合、本明細書(定義を含めて)が優先する。

30

【0019】

(定義)

以下に本明細書において特に使用される用語の定義および/または基本的技術内容を適宜説明する。

【0020】

本明細書において「免疫実体(immunological entity)」とは、免疫反応を担う任意の物質をいう。免疫実体には、抗体、抗体の抗原結合断片、T細胞受容体、T細胞受容体の断片、B細胞受容体、B細胞受容体の断片、キメラ抗原受容体(CAR)、これらのいずれかまたは複数を含む細胞(例えば、キメラ抗原受容体(CAR)を含むT細胞(CAR-T))等が含まれる。免疫実体は広く考えることができ、アルパカ等の動物が産生するナノボディ(nanobody)や人工的に多様性(diversity)を持たせたファージディスプレイ等(これにはscFvやナノボディを含む)の解析で使用される免疫学的に関連する実体(entity)も同様に包含される。本明細書において特に断らない限り「第一」および「第二」等(「第三」...等)の記載は、相互に異なる実体であることを示す。

40

【0021】

本明細書において「抗体」とは、当該分野で通常使用されるのと同様の意義で用いられ、抗原が生体の免疫系と接触する(抗原刺激)ときに免疫系でつくられる、抗原と高度に特異的な反応をするタンパク質をいう。本発明で用いられるエピトープに対する抗体は、

50

それぞれ、特定のエピトープに結合すればよく、その由来、種類、形状などは問われない。本明細書に記載される抗体はフレームワーク領域と抗原結合領域（CDR）とに分割することができる。

#### 【0022】

本明細書において「T細胞レセプター（TCR）」とは、T細胞受容体、T細胞抗原レセプター、T細胞抗原受容体ともいい、免疫系をつかさどるT細胞の細胞膜に発現する受容体（レセプター）をいい、抗原を認識する。鎖、鎖、鎖および鎖が存在し、

または の二量体を構成する。前者の組み合わせからなるTCRを TCR、後者の組み合わせからなるTCRを TCRと呼び、それぞれのTCRを持つT細胞は T細胞、 T細胞と呼ばれる。構造的にB細胞の産生する抗体のFabフラグメントと非常に類似しており、MHC分子に結合した抗原分子を認識する。成熟T細胞の持つTCR遺伝子は遺伝子再編成を経ているため、一個体は多様性に富んだTCRを持ち、様々な抗原を認識することができる。TCRはさらに細胞膜に存在する不可変なCD3分子と結合し複合体を形成する。CD3は細胞内領域にITAM（immunoreceptor tyrosine-based activation motif）と呼ばれるアミノ酸配列を持ち、このモチーフが細胞内のシグナル伝達に関与するとされている。それぞれのTCR鎖は可変部（V）と定常部（C）から構成され、定常部は細胞膜を貫通して短い細胞質部分を持つ。可変部は細胞外に存在して、抗原-MHC複合体と結合する。可変部には超可変部、あるいは相補性決定領域（CDR）と呼ばれる領域が3つ存在し、この領域が抗原-MHC複合体と結合する。3つのCDRはそれぞれCDR1、CDR2、CDR3と呼ばれる。TCRの遺伝子再構成は免疫グロブリンとして知られるB細胞受容体の過程と同様である。TCRの遺伝子再編成ではまず、鎖のVDJ再編成が行われ、続いて鎖のVJ再編成が行われる。鎖の再編成が行われる際に鎖の遺伝子は染色体上から欠失するため、TCRを持つT細胞が TCRを同時に持つことはない。逆に TCRを持つT細胞ではこのTCRを介したシグナルが鎖の発現を抑制するため、TCRを持つT細胞が TCRを同時に持つこともない。

#### 【0023】

本明細書において「B細胞レセプター（BCR）」とは、B細胞受容体、B細胞抗原レセプター、B細胞抗原受容体とも呼ばれ、膜結合型免疫グロブリン（mIg）分子と会合したIg / Ig（CD79a / CD79b）ヘテロ二量体（ / ）から構成されるものをいう。mIgサブユニットは抗原に結合し、受容体の凝集を起こすが、一方、 / サブユニットは細胞内に向かってシグナルを伝達する。BCRが凝集すると、チロシンキナーゼのSyk及びBtkと同様に、SrcファミリーキナーゼのLyn、Blk、及びFynを速やかに活性化するといわれる。BCRシグナル伝達の複雑さによって多くの異なる結果が生じるが、その中には、生存、耐性（アレルギー；抗原に対する過敏反応の欠如）またはアポトーシス、細胞分裂、抗体産生細胞または記憶B細胞への分化などが含まれる。TCRの可変領域の配列が異なるT細胞が何億種類も生成され、またBCR（または抗体）の可変領域の配列が異なるB細胞が何億種類も生成される。TCRとBCRの個々の配列はゲノム配列の再構成や変異導入により異なるので、T細胞やB細胞の抗原特異性については、TCR・BCRのゲノム配列またはmRNA（cDNA）の配列を決定

#### 【0024】

本明細書において「キメラ抗原受容体（CAR）」とは、腫瘍抗原に特異的なモノクローナル抗体可変領域の軽鎖（VL）と重鎖（VH）を直列に結合させた単鎖抗体（scFv）をN末端側に、T細胞受容体（TCR）鎖をC末端側に持つキメラ蛋白の総称であり、腫瘍免疫回避機構に打ち勝つための遺伝子操作を加えた人工T細胞受容体を患者T細胞に遺伝子導入し、そのT細胞を体外で増幅培養した後に患者に輸注するという遺伝子・細胞治療法において使用される人工T細胞受容体である（Dotto G, et al. Hum Gene Ther 20: 1229-1239, 2009）。本発明により同定またはクラスター化されたエピトープを用いて、このようなCARを生産するこ

とができ、生産されたCARまたはそれを含む遺伝子改変T細胞を用いて遺伝子細胞治療法を実現することができる(Credit: Brentjens R, et al. "Driving CAR T cells forward." Nat Rev Clin Oncol. 2016 13, 370-383等を参照)。

【0025】

本明細書において「V領域」とは、抗体、TCRまたはBCR等の免疫実体の可変領域の可変部(V)領域をいう。

【0026】

本明細書において「D領域」とは、抗体、TCRまたはBCR等の免疫実体の可変領域のD領域をいう。

【0027】

本明細書において「J領域」とは、抗体、TCRまたはBCR等の免疫実体の可変領域のJ領域をいう。

【0028】

本明細書において「C領域」とは、抗体、TCRまたはBCR等の免疫実体の定常部(C)領域をいう。

【0029】

本明細書において「可変領域のレパトア(repertoire)」とは、TCRまたはBCRで遺伝子再構成により任意に作り出されたV(D)J領域の集合をいう。TCRレパトア、BCRレパトア等と熟語で使用されるが、これらは例えば、T細胞レパトア、B細胞レパトアなどと称されることもある。例えば、「T細胞レパトア」とは、抗原認識または免疫実体結合物の認識において重要な役割を果たすT細胞レセプター(TCR)の発現によって特徴づけられるリンパ球の集合をいう。T細胞レパトアの変化は、生理的状態および疾患状態における免疫状態の有意な指標をもたらすため、T細胞レパトア解析は、疾患の発症に関与する抗原特異性T細胞の同定およびTリンパ球の異常の診断のために行われてきた。TCRやBCRはゲノム上に存在する複数のV領域、D領域、J領域、C領域の遺伝子断片の遺伝子再構成によって、多様な遺伝子配列を創出している。

【0030】

本明細書において「アイソタイプ」とは、IgM、IgA、IgG、IgEおよびIgD等において、同じタイプに属するが、相互に配列が異なるタイプを言う。アイソタイプは種々の遺伝子の略称や記号を用いて表示される。

【0031】

本明細書において「サブタイプ」とは、BCRの場合IgAおよびIgGにおいて存在するタイプ内のタイプであって、IgGについては、IgG1、IgG2、IgG3もしくはIgG4、IgAについてはIgA1もしくはIgA2が存在する。TCRについても、鎖および鎖において存在することが知られており、それぞれTRBC1、TRBC2、あるいはTRGC1、TRGC2が存在する。

【0032】

本明細書において「免疫実体結合体」とは、抗体、TCRまたはBCR等の免疫実体によって特異的に結合され得る任意の基質をいう。本明細書において「抗原」と称した場合、広義には「免疫実体結合物」を指すことがあるが、当該分野において「抗原」とは狭義には抗体と対で用いられることがあり、狭義には「抗原」は「抗体」に特異的に結合され得る任意の基質をいう。

【0033】

本明細書において「エピトープ」とは、抗体またはリンパ球レセプター(TCR、BCR等)等の免疫実体が結合する免疫実体結合物(例えば、抗原)分子中の部位をいう。アミノ酸の直鎖がエピトープを構成することもあるが(直鎖状エピトープ)、タンパク質の離れた部分が立体構造を構成しエピトープとして機能することもある(コンフォメーションエピトープ)。本発明が対象とするエピトープはこのようなエピトープの詳細な分類は問わない。ある抗体等の免疫実体に関してエピトープが同じであれば、他の配列を有す

10

20

30

40

50

る抗体等の免疫実体であっても同様に利用することができることが理解される。

【0034】

本明細書において「抗原特異性」とは、免疫実体についていうとき、その結合パートナー（例えば、抗原）との結合特異性を言い、ある特定の結合パートナーには結合するが、その他の結合パートナーとは結合しないか低い親和性で結合する性質を言う。

【0035】

本明細書において「結合モード」とは、免疫実体とその結合パートナーとの間の3次元的な結合の様式（モード）をいい、物理的な概念を表すものである。理論に束縛されることを望まないが、概して、結合モードが複数集合すると抗原特異性を形成すると考えられるが、これに限定されるものではない。

【0036】

本明細書において免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードが「同一」か「異なる」かは、本発明に基づく分類に従って、類似度（アミノ酸配列、三次元構造、抗原特異性または結合モード等）によって判断することができる。「同一」は、化学式やアミノ酸配列等が完全に同一であることに限られるというものではなく、機能または立体構造が実質的に同質であることをいい、本発明では、代表的に抗原特異性または結合モードによって決定され得、同一の免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードのクラスターに属する免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードは本発明では「同一」と判断される。従って、「異なる」免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードとは、「同一」のクラスターに属するものではない免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードを指す。一つの実施形態では、免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードが「同一」か「異なる」かによって、同一のクラスターに属するかどうかを決定することができる。クラスター分析を行った場合、ある免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードは、別の免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードと比較して同じクラスターに属する場合は同一と判断し、別のクラスターに属する場合は異なると判断する。したがって、結合する免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードが同一である免疫実体、エピトープ、免疫実体結合物を同一のクラスターに分類し、クラスターを生成することもできる。また、免疫実体、エピトープまたは免疫実体結合物を、その特性および既知の免疫実体との類似性からなる群より選択される少なくとも1つの評価項目を評価し、所定の基準を満たした免疫実体、エピトープまたは免疫実体結合物を対象に前記クラスター分類を行うことができる。したがって、1つの実施形態では、免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードが同一である場合、該免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードの三次元構造が少なくとも一部重複あるいはすべて重複することがあり、あるいは、該免疫実体またはエピトープまたは免疫実体結合物（または、抗原特異性もしくは結合モードを担う）のアミノ酸配列または他の化学物質の部分構造が少なくとも一部あるいはすべて重複することがある。重要な指標としては、確実に確認できる構造データ等とよく合うように閾値を決めるのが適切であるが、統計学的有意性を重視する場合、他の閾値を採用することもあり得、当業者は状況に応じて本明細書の記載を参考に適宜閾値を設定することができる。例えば、階層的クラスタリング手法（例えば、群平均法(average linkage clustering)、最短距離法(NN法)、K-NN法、Ward法、最長距離法、重心法)を用いてクラスタリング分析をした場合に求められる最大距離が特定の値未満のものを同一クラスターとみなすことができる。このような値としては、1未満、0.95未満、0.9未満、0.85未満、0.8未満、0.75未満、0.7未満、0.65未満、0.6未満、0.55未満、0.5未満、0.45未満、0.4未満、0.35未満、0.3未満、0.25未満、0.2未満、0.15未満、0.1未満、0.05未満などを挙げることができるがこれらに限定されない。クラスタリング手法としては階層的手法に限られず、非階層的手法を用いてもよい。

10

20

30

40

50

## 【0037】

本明細書において免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードの「クラスター」とは、一般に、ある集団の要素（この場合免疫実体またはエピトープ、免疫実体結合物、あるいは抗原特異性、または結合モード）を、外的基準や群の数の指定なしに、多次元空間における要素の分布から、類似したものを集めたものをいい、本明細書では、多数の免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードのうち少なくとも1つが類似したものを集めたものをいう。同一のクラスターに属する免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードには、同様の抗体が結合する。多変量解析によって分類することができ、種々のクラスター分析手法を用いてクラスターを構成することができる。本発明が提供する免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードのクラスターは、そのクラスターへ属していることを示すことにより、生体内の状態（例えば、疾患、障害や薬効、特に免疫状態等）を反映することが示された。

10

## 【0038】

本発明において、クラスタリングによる分類に基づく解析としては、例えば、クラスタリング結果からそれぞれのクラスターを遺伝子のように見なし、遺伝子発現解析のように用いることが想定される。具体的には、例えば、1. 継時変化を追う場合、特定の、あるいは複数のクラスターに属する配列の増減を見る。増減しているクラスターの数を見る。それぞれのクラスターに特徴的な量（V/D/J遺伝子、CDRの長さ、親水性、疎水性、保存されている残基等）を見出す。2. 複数検体の特定の層に興味がある場合、特定の層に優位に存在、増減しているクラスターを同定する。増減しているクラスターの数を見る。それぞれのクラスターに特徴的な量を見出す。3. 機能に興味がある場合（機能＝抗原特異性または結合モード）、興味ある抗原に特異的な配列（ELISPOTアッセイ、pMHCテトラマーによるsorting等、別の実験により得ているとする）を含むクラスターに注目し、その増減を見る。（機能＝細胞の機能）別々にソーティングして配列決定した異なるサブタイプの細胞から得られたクラスタリング結果を比較する。4. 別の実験ソースと比較する場合、遺伝子発現解析や、オミックス解析、細菌叢、サイトカイン、細胞種の数の大小等との相関、あるいはそれらと組み合わせた1-3までの解析を適宜使用することができることが想定され得る。

20

## 【0039】

本明細書において、「機械学習」とは、当該分野で使用される最も広義の意味で解され、機械（コンピュータ）が学習することをいい、人間が自然に行っている学習能力と同様の機能をコンピュータで実現しようとする技術・手法のことである。機械が学習するためには、学習の元となるデータを入力値として用いる。入力値を「機械学習アルゴリズム」と呼ばれる処理を通して、データを分類したり、認識したりする処理を見つけ出す。そして、この学習した処理を使うことで、学習後に入力された未だ学習していないデータに対しても、分類したり識別したりすることができるようになる。機械学習により、分類や認識、識別、あるいは回帰（予測）を行うことができる。機械学習には教師あり学習と教師なし学習があり、強化学習という手法もある。ディープラーニングは機械学習の一部であり、機械学習は人工知能（AI）の一部といえる。機械学習は、開発者が全ての動作をプログラムするのではなく、データをAI自身が解析し、法則性やルールを見つけ出す、つまり、トレーニングによって特定のタスクを実行できるようになる人工知能をさすが、ディープラーニングは機械学習の中の一手法で、ニューラルネットワークとその関連技術の発展形であり、従来の機械学習と異なり、人間の神経を参考にしたニューラルネットワークを何層も重ねることにより、データの分析と学習を強化した人工知能をさす。ディープラーニングで用いられるオートエンコーダ（auto-encoder）によって全体評価を行う場合には、免疫実体の配列自身を入力として高次元ベクトル空間に射影することになる。すなわち、オートエンコーダ自身が特徴量を抽出して高次元ベクトル空間に射影する。特徴量はそのまま高次元ベクトル空間要素となる。射影は恒等写像を含むと解釈することができる。

30

40

50

## 【0040】

本明細書において「分類」とは、抗原特異性または結合モードについて言及するとき、ある一定の基準に基づいてその抗原特異性または結合モードが同様の性質を有するグループごとに分けることをいう。本発明では、クラスタリングによって分類することができる。

## 【0041】

本明細書において「異同」とは、抗原特異性または結合モードについて言及するとき、その抗原特異性または結合モードが同様の性質や構造を有しているかどうかをいう。

## 【0042】

本明細書において「抗原特異性または結合モードを特定する」とは特定の興味ある抗原、あるいは抗原に属する結合モードにのみ注目することを言う。解析の対象を特定することとすることができる。

10

## 【0043】

本明細書において「抗原特異性または結合モードを特定しない」とは、特定の興味ある抗原、あるいは抗原に属する結合モードではなく、多様な抗原に対する抗原特異性や結合モードを（好ましくは、等しく）一般的に扱うことを言う。

## 【0044】

本明細書において「類似度」とは、免疫実体結合物（例えば、抗原）、免疫実体、エピトープ、抗原特異性、または結合モード等の分子またはその一部あるいはそれが形成する空間的配置等について、分子が類似している度合いをいう。類似度は、長さの違い、配列類似度などに基づいて決定することができる。理論に束縛されることを望まないが、本発明の一部の実施形態では、この類似度に基づいて免疫実体、エピトープ、免疫実体結合物、抗原特異性、または結合モードを分類したところ、同一のクラスターに属するエピトープに結合する抗体、TCR、BCR等は、同一のカテゴリーに入る疾患、障害、症状や生理現象等に割り当てられ得ることが理解される。従って、本発明の手法を用いて同じ免疫実体、エピトープ、抗原特異性、または結合モードのクラスターに反応する抗体、TCR、BCR等を有するかどうかを調べることによって、各種の診断（がんの罹患、投与薬の適合性等）を行うことができる。本発明の解析において、類似度を用いることができる。

20

## 【0045】

本明細書において「類似性スコア」とは、類似性を示す具体的な数値をいい、「類似度」ともいう。構造類似度を計算した場合に使用される技法に応じて、適宜適切なスコアが採用されうる。類似性スコアは、例えば、回帰的な手法、ニューラルネットワーク法や、サポートベクトルマシン、ランダムフォレストといった機械学習アルゴリズムなどを用いて算出することができる。本発明の解析において類似性スコアを用いることができる。

30

## 【0046】

本明細書において「特徴量」とは、機械学習などの解析や計算をする場合に、結果に影響すると考えられる要素をいう。免疫実体の解析において有用な特徴量としては、例えば、配列情報、CDR1-3配列の長さ、配列一致度、フレームワーク領域の配列一致度、分子の全電荷/親水性/疎水性/芳香族アミノ酸の数、各CDR、フレームワーク領域の電荷/親水性/疎水性/芳香族アミノ酸の数、各アミノ酸の数、重鎖-軽鎖の組み合わせ、体細胞変異数、変異の位置、アミノ酸モチーフの存在/一致度、参照配列セットに対する希少度、参照配列による結合HLAのオッズ比等を挙げることができ、これらは1つまたは複数種類用いることができるが、これらに限定されない。特徴量は特徴ベクトルとして機械学習アルゴリズムの入力として用いられる。

40

## 【0047】

本明細書において「距離」とは、抗原特異性の距離は抗原特異性が一致するかどうかの判定を意味する。「距離」は任意の数値を設定することができ、具体的には、「距離」を0か1かで予測する設定にした場合、クラスタリングは単なる1をまとめる作業となる。他方で、距離を[0-1]で表現する場合、クラスタリングのメリットは単なる距離（対の関係）ではなく、周辺にある対の密度など他のパラメータも考慮することができる。本

50

発明では、どちらも可能である。距離に関する情報は、本発明では、特徴量を提供するために用いることができる情報である。

【0048】

本明細書において「相補性決定領域(CDR)」とは、抗体等の免疫実体において、実際に免疫実体結合物(例えば、抗原)に接触して結合部位を形成している領域である。CDRに関する情報は、本発明では、特徴量を提供するために用いることができる情報である。一般的にCDRは、抗体および抗体に相当する分子(免疫実体)のFv(重鎖可変領域(VH)および軽鎖可変領域(VL)を含む)上に位置している。また一般的にCDRは、5~30アミノ酸残基程度からなるCDR1、CDR2、CDR3が存在する。そして、抗原抗体反応では、特に重鎖のCDRが抗体の抗原への結合に寄与していることが知られている。またCDRの中でも、CDR3、特にCDR-H3が抗体の抗原への結合における寄与が最も高いことが知られている。例えば、"Willy et al., Biochemical and Biophysical Research Communications Volume 356, Issue 1, 27 April 2007, Pages 124-128"には、重鎖CDR3を改変させることで抗体の結合能を上昇させたことが記載されている。CDRの定義およびその位置を決定する方法は複数報告されている。例えば、Kabatの定義(Sequences of Proteins of Immunological Interest, 5th ed., Public Health Service, National Institutes of Health, Bethesda, MD. (1991))、またはChothiaの定義(Chothia et al., J. Mol. Biol., 1987; 196:901-917)を採用してもよい。本発明の一実施形態においては、Kabatの定義を好適な例として採用するが、必ずしもこれに限定されない。また、場合によっては、Kabatの定義とChothiaの定義の両方を考慮して決定しても良く(改変Chothia法)、例えば、各々の定義によるCDRの重複部分を、または各々の定義によるCDRの両方を含んだ部分をCDRとすることもでき、あるいはIMGTまたはHoneggerに従って決定することもできる。そのような方法の具体例としては、Kabatの定義とChothiaの定義の折衷案である、Oxford Molecular's AbM antibody modeling softwareを用いたMartinらの方法(Proc. Natl. Acad. Sci. USA, 1989; 86:9268-9272)がある。このようなCDRの情報を用いて、本発明を実施することができる。本明細書において「CDR3」とは、3つめの相補性決定領域(complementarity-determining region: CDR)をいい、ここで、CDRとは、可変領域のうち、直接免疫実体結合物(例えば、抗原)と接触する領域であり、特に変化が大きく、この超可変領域のことをいう。軽鎖と重鎖の可変領域に、それぞれ3つのCDR(CDR1~CDR3)と、3つのCDRを取り囲む4つのFR(FR1~FR4)が存在する。CDR3領域は、V領域、D領域、J領域にまたがって存在するとされているため、可変領域の鍵を握るといわれており、分析対象として用いられる。

【0049】

本明細書において「フレームワーク領域」とは、CDR以外のFv領域の領域をいい、通常FR1、FR2、FR3およびFR4からなり、抗体間で比較的よく保存されているとされる(Kabat et al., "Sequence of Proteins of Immunological Interest" US Dept. Health and Human Services, 1983.)。それゆえ、本発明において、各配列の比較の際にフレームワーク領域を固定する手法を採用しうる。フレームワーク領域に関する情報は、本発明では、特徴量を提供するために用いることができる情報である。

【0050】

本明細書において「遺伝子領域」とは、フレームワーク領域および抗原結合領域(CDR)や、V領域、D領域、J領域およびC領域等の各領域をさす。このような遺伝子領域は、当該分野で公知であり、データベース等を参酌して適宜決定することができる。本明細書において遺伝子の「相同性」とは、2以上の遺伝子配列の、互いに対する同一性の程度をいい、一般に「相同性」を有するとは、同一性または類似性の程度が高いことをいう。従って、ある2つの遺伝子の相同性が高いほど、それらの配列の同一性または類似性は高い。2種類の遺伝子が相同性を有するか否かは、配列の直接の比較、または核酸の場合ストリンジェントな条件下でのハイブリダイゼーション法によって調べられ得る。本明細

10

20

30

40

50

書において「相同性検索」とは、相同性の検索をいう。好ましくは、コンピュータを用いてインシリコで行うことができる。遺伝子領域に関する情報は、本発明では、特徴量を提供するために用いることができる情報である。

#### 【0051】

本明細書においてアミノ酸配列等の領域の「同定」とは、アミノ酸配列をある観点で特徴づけることをいい、一つの性質を有する特徴で特定される領域を定めることをいう。同定には、具体的にアミノ酸番号を含む領域を特定すること、それらの領域に関する特徴をリンクさせることなどが含まれるがこれらに限定されない。本明細書においてアミノ酸配列等の領域の「分割」とは、アミノ酸配列を特徴づけたのち、一つの性質を有する特徴で定められる領域ごとに区別し別々の領域にすることをいう。このような同定および分割は、バイオインフォマティクス分野で使用される任意の技術、例えばKabat、Chotia、改変Chotia、IMGT、Honegger等を用いて実施することができる。本明細書において、アミノ酸配列等の領域の処理の際、フレームワーク等に例示される保存領域を同定することが一つの重要な特徴であり、同定の結果、保存領域と非保存領域（例えば、CDR等）とに分割されることも想定される。2つ以上の免疫実体の保存領域または非保存領域の一部を同定して重ね合わせを行う場合、それぞれの免疫実体の一部は実質的に対応関係にあることが好ましい。本明細書において「対応関係」にあるとは、保存領域についていう場合は、第一の免疫実体の一部と、第二の免疫実体の一部とについて、三次元構造の位置を考慮したときに互いに重ね合わせられ得る関係にある。非保存領域の場合は、本明細書において説明される同一残基の定義を行うことで、三次元構造の位置を考慮したときに互に対応するアミノ酸残基が存在することになる。したがって、「対応関係」は、配列等のアラインメントまたは同一残基の同定などを行うことによって確認することができる。

#### 【0052】

本明細書において「アラインメント」（英語では、alignment（名詞）またはalign（動詞））とは、アラインメント、整列とも言い、バイオインフォマティクスにおいて、DNAやRNA、タンパク質の一次構造の類似した領域を特定できるように並べたものをいう。機能的、構造的、あるいは進化的な配列の関係性を知るヒントを与えることが多い。アラインメントされたアミノ酸残基等の配列は、典型的には行列の行として表現され、同一あるいは類似性質の配列が同じ列に並ぶようギャップが挿入される。2つの配列を比較する場合は、ペアワイズ配列アラインメントと称され、2配列間でのアラインメントで、部分的、あるいは全体の類似性を詳しく調べるときに用いる。アラインメントには、代表的には動的計画法を用いることができ、代表的な手法として、グローバルアラインメントについてはNeedleman-Wunsch法（ニードルマン＝ウンシュ法）、ローカルアラインメントについてはSmith-Waterman法（スミス＝ウォーターマン法）を利用することができる。ここで、グローバルアラインメントとは配列中の全残基がアラインメントされるようにしたもので、ほぼ同じ長さの配列間での比較に有効である。ローカルアラインメントは、配列が全体としては似ておらず、部分的類似を見つけた場合に有効である。本明細書において「ミスマッチ」とは、核酸配列、アミノ酸配列等をアラインメントしたときに、互いに同一ではない塩基またはアミノ酸が存在することをいう。「ギャップ」は、アラインメントにおいて、一方には存在するが他方には存在しない塩基またはアミノ酸が存在することをいう。アラインメントに関する情報は、本発明では、特徴量を提供するために用いることができる情報である。

#### 【0053】

本明細書において「アサイン」とは、ある配列（例えば、核酸配列、タンパク質配列等）に、特定の遺伝子名、機能、特徴領域（例えば、V領域、J領域など）等情報を割り当てることをいう。具体的には、ある配列に特定の情報を入力またはリンクさせる等により達成することができる。

#### 【0054】

本明細書において「特異的」とは、対象となる配列に結合するが、少なくとも対象とな

10

20

30

40

50

る抗体、TCRまたはBCRのプールにおいて、好ましくは存在する抗体、TCRまたはBCRの配列すべてにおいて、他の配列とは結合性が低い、好ましくは結合しないことをいう。特異的な配列は好ましくは対象となる配列に対して完全に相補的であることが有利であるが、必ずしも限定されない。

【0055】

本明細書において「タンパク質」、「ポリペプチド」、「オリゴペプチド」および「ペプチド」は、本明細書において同じ意味で使用され、任意の長さのアミノ酸のポリマーをいう。このポリマーは、直鎖であっても分岐していてもよく、環状であってもよい。アミノ酸は、天然のものであっても非天然のものであってもよく、改変されたアミノ酸であってもよい。この用語はまた、複数のポリペプチド鎖の複合体へとアセンブルされたものを包含し得る。この用語はまた、天然または人工的に改変されたアミノ酸ポリマーも包含する。そのような改変としては、例えば、ジスルフィド結合形成、グリコシル化、脂質化、アセチル化、リン酸化または任意の他の操作もしくは改変（例えば、標識成分との結合体化）が包含される。この定義にはまた、例えば、アミノ酸の1または2以上のアナログを含むポリペプチド（例えば、非天然アミノ酸などを含む）、ペプチド様化合物（例えば、ペプチド）および当該分野において公知の他の改変が包含される。

10

【0056】

本明細書において、「アミノ酸」は、本発明の目的を満たす限り、天然のものでも非天然のものでもよい。

【0057】

本明細書において「ポリヌクレオチド」、「オリゴヌクレオチド」および「核酸」は、本明細書において同じ意味で使用され、任意の長さのヌクレオチドのポリマーをいう。この用語はまた、「オリゴヌクレオチド誘導体」または「ポリヌクレオチド誘導体」を含む。「オリゴヌクレオチド誘導体」または「ポリヌクレオチド誘導体」とは、ヌクレオチドの誘導体を含むか、またはヌクレオチド間の結合が通常とは異なるオリゴヌクレオチドまたはポリヌクレオチドをいい、互換的に使用される。そのようなオリゴヌクレオチドとして具体的には、例えば、2'-O-メチル-リボヌクレオチド、オリゴヌクレオチド中のリン酸ジエステル結合がホスホロチオエート結合に変換されたオリゴヌクレオチド誘導体、オリゴヌクレオチド中のリン酸ジエステル結合がN3'-P5'ホスホロアミデート結合に変換されたオリゴヌクレオチド誘導体、オリゴヌクレオチド中のリボースとリン酸ジエステル結合とがペプチド核酸結合に変換されたオリゴヌクレオチド誘導体、オリゴヌクレオチド中のウラシルがC-5プロピニルウラシルで置換されたオリゴヌクレオチド誘導体、オリゴヌクレオチド中のウラシルがC-5チアゾールウラシルで置換されたオリゴヌクレオチド誘導体、オリゴヌクレオチド中のシトシンがC-5プロピニルシトシンで置換されたオリゴヌクレオチド誘導体、オリゴヌクレオチド中のシトシンがフェノキサジン修飾シトシン（phenoxazine-modified cytosine）で置換されたオリゴヌクレオチド誘導体、DNA中のリボースが2'-O-プロピルリボースで置換されたオリゴヌクレオチド誘導体およびオリゴヌクレオチド中のリボースが2'-メトキシエトキシリボースで置換されたオリゴヌクレオチド誘導体などが例示される。他にそうではないと示されなければ、特定の核酸配列はまた、明示的に示された配列と同様に、その保存的に改変された改変体（例えば、縮重コドン置換体）および相補配列を包含することが企図される。具体的には、縮重コドン置換体は、1またはそれ以上の選択された（または、すべての）コドンの3番目の位置が混合塩基および/またはデオキシイノシン残基で置換された配列を作製することにより達成され得る（Batzer et al., *Nucleic Acid Res.*19:5081(1991); Ohtsuka et al., *J. Biol. Chem.* 260: 2605-2608(1985); Rossolini et al., *Mol. Cell. Probes* 8:91-98(1994)）。本明細書において「核酸」はまた、遺伝子、cDNA、mRNA、オリゴヌクレオチド、およびポリヌクレオチドと互換可能に使用される。本明細書において「ヌクレオチド」は、天然のものでも非天然のものでもよい。

20

30

40

【0058】

本明細書において「遺伝子」とは、遺伝形質を規定する因子をいう。通常染色体上に一

50

定の順序に配列している。タンパク質の一次構造を規定する遺伝子を構造遺伝子といい、その発現を左右する遺伝子を調節遺伝子という。本明細書では、「遺伝子」は、「ポリヌクレオチド」、「オリゴヌクレオチド」および「核酸」をさすことがある。「遺伝子産物」とは、遺伝子に基づいて産生された物質でありタンパク質、mRNAなどをさす。

#### 【0059】

アミノ酸は、その一般に公知の3文字記号か、またはIUPAC - IUB Biochemical Nomenclature Commissionにより推奨される1文字記号のいずれかにより、本明細書中で言及され得る。ヌクレオチドも同様に、一般に認知された1文字コードにより言及され得る。本明細書では、アミノ酸配列および塩基配列の類似性、同一性および相同性の比較は、配列分析用ツールであるBLASTを用いてデフォルトパラメータを用いて算出される。同一性の検索は例えば、NCBIのBLAST 2.2.28 (2013.4.2発行)を用いて行うことができる。本明細書における同一性の値は通常は上記BLASTを用い、デフォルトの条件でアラインした際の値をいう。ただし、パラメータの変更により、より高い値が出る場合は、最も高い値を同一性の値とする。複数の領域で同一性が評価される場合はそのうちの最も高い値を同一性の値とする。類似性は、同一性に加え、類似のアミノ酸についても計算に入れた数値である。

#### 【0060】

本明細書において遺伝子の「相同性」とは、2以上の遺伝子配列の、互いに対する同一性の程度をいい、一般に「相同性」を有するとは、同一性または類似性の程度が高いことをいう。従って、ある2つの遺伝子の相同性が高いほど、それらの配列の同一性または類似性は高い。2種類の遺伝子が相同性を有するか否かは、配列の直接の比較、または核酸の場合ストリンジェントな条件下でのハイブリダイゼーション法によって調べられ得る。2つの遺伝子配列を直接比較する場合、その遺伝子配列間でDNA配列が、代表的には少なくとも50%同一である場合、好ましくは少なくとも70%同一である場合、より好ましくは少なくとも80%、90%、95%、96%、97%、98%または99%同一である場合、それらの遺伝子は相同性を有する。従って本明細書において「相同体」または「相同遺伝子産物」は、本明細書にさらに記載する複合体のタンパク質構成要素と同じ生物学的機能を発揮する、別の種、好ましくは哺乳動物におけるタンパク質を意味する。

#### 【0061】

本明細書において「精製された」物質または生物学的因子（例えば、核酸またはタンパク質など）とは、その生物学的因子に天然に随伴する因子の少なくとも一部が除去されたものをいう。従って、通常、精製された生物学的因子におけるその生物学的因子の純度は、その生物学的因子が通常存在する状態よりも高い（すなわち濃縮されている）。本明細書中で使用される用語「精製された」は、好ましくは少なくとも75重量%、より好ましくは少なくとも85重量%、よりさらに好ましくは少なくとも95重量%、そして最も好ましくは少なくとも98重量%の、同型の生物学的因子が存在することを意味する。本発明で用いられる物質は、好ましくは「精製された」物質である。本明細書において「単離」されたとは、天然に存在する状態で付随する任意のものを少なくとも1つ除去したものをいい、例えば、ゲノム配列からその特定の遺伝子配列を取り出した場合も単離とという。

#### 【0062】

本明細書において「マーカー（物質、タンパク質または遺伝子（核酸）」とは、ある状態（例えば、正常細胞状態、形質転換状態、疾患状態、障害状態、あるいは増殖能、分化状態のレベル、有無等）にあるかまたはその危険性があるかどうかを追跡する示標となる物質をいう。このようなマーカーとしては、遺伝子（核酸 = DNAレベル）、遺伝子産物（mRNA、タンパク質など）、代謝物質、酵素などを挙げることができる。本発明において、ある状態（例えば、分化障害などの疾患）についての検出、診断、予備的検出、予測または事前診断は、その状態に関連するマーカーに特異的な薬剤、剤、因子または手段、あるいはそれらを含む組成物、キットまたはシステム等を用いて実現することができる。本明細書において、「遺伝子産物」とは、遺伝子によってコードされるタンパク質ま

10

20

30

40

50

たは mRNA をいう。

【0063】

本明細書において「被験体」とは、本発明の診断または検出等の対象となる対象（例えば、ヒト等の生物または生物から取り出した器官あるいは細胞等）をいう。

【0064】

本明細書において「試料」とは、被験体等から得られた任意の物質をいい、例えば、細胞等が含まれる。当業者は本明細書の記載をもとに適宜好ましい試料を選択することができる。

【0065】

本明細書において「薬剤」、「剤」または「因子」（いずれも英語では agent に相当する）は、広義には、交換可能に使用され、意図する目的を達成することができる限りどのような物質または他の要素（例えば、光、放射能、熱、電気などのエネルギー）でもあってもよい。そのような物質としては、例えば、タンパク質、ポリペプチド、オリゴペプチド、ペプチド、ポリヌクレオチド、オリゴヌクレオチド、ヌクレオチド、核酸（例えば、cDNA、ゲノムDNAのようなDNA、mRNAのようなRNAを含む）、ポリサッカリド、オリゴサッカリド、脂質、有機低分子（例えば、ホルモン、リガンド、情報伝達物質、有機低分子、コンビナトリアルケミストリで合成された分子、医薬品として利用され得る低分子（例えば、低分子リガンドなど）など）、これらの複合分子が挙げられるがそれらに限定されない。ポリヌクレオチドに対して特異的な因子としては、代表的には、そのポリヌクレオチドの配列に対して一定の配列相同性を（例えば、70%以上の配列同一性）もって相補性を有するポリヌクレオチド、プロモーター領域に結合する転写因子のようなポリペプチドなどが挙げられるがそれらに限定されない。ポリペプチドに対して特異的な因子としては、代表的には、そのポリペプチドに対して特異的に指向された抗体またはその誘導体あるいはその類似物（例えば、単鎖抗体）、そのポリペプチドがレセプターまたはリガンドである場合の特異的なリガンドまたはレセプター、そのポリペプチドが酵素である場合、その基質などが挙げられるがそれらに限定されない。

【0066】

本明細書において「検出剤」とは、広義には、目的の対象を検出することができるあらゆる薬剤をいう。

【0067】

本明細書において「診断剤」とは、広義には、目的の状態（例えば、疾患など）を診断することができるあらゆる薬剤をいう。

【0068】

本発明の検出剤は、検出可能とする部分（例えば、抗体等）に他の物質（例えば、標識等）を結合させた複合体または複合分子であってもよい。本明細書において使用される場合、「複合体」または「複合分子」とは、2以上の部分を含む任意の構成体を意味する。例えば、一方の部分がポリペプチドである場合は、他方の部分は、ポリペプチドであってもよく、それ以外の物質（例えば、糖、脂質、核酸、他の炭化水素等）であってもよい。本明細書において複合体を構成する2以上の部分は、共有結合で結合されていてもよくそれ以外の結合（例えば、水素結合、イオン結合、疎水性相互作用、ファンデルワールス力等）で結合されていてもよい。2以上の部分がポリペプチドの場合は、キメラポリペプチドとも称しうる。従って、本明細書において「複合体」は、ポリペプチド、ポリヌクレオチド、脂質、糖、低分子などの分子が複数種連結してできた分子を含む。

【0069】

本明細書において「相互作用」とは、2つの物質についていうとき、一方の物質と他方の物質との間で力（例えば、分子間力（ファンデルワールス力）、水素結合、疎水性相互作用など）を及ぼしあうこという。通常、相互作用をした2つの物質は、会合または結合している状態にある。

【0070】

本明細書中で使用される用語「結合」は、2つの物質の間、あるいはそれらの組み合わせ

10

20

30

40

50

せの間での、物理的相互作用または化学的相互作用を意味する。結合には、イオン結合、非イオン結合、水素結合、ファンデルワールス結合、疎水性相互作用などが含まれる。物理的相互作用（結合）は、直接的または間接的であり得、間接的なものは、別のタンパク質または化合物の効果を介するかまたは起因する。直接的な結合とは、別のタンパク質または化合物の効果を介してもまたはそれらに起因しても起こらず、他の実質的な化学中間体を伴わない、相互作用をいう。結合または相互作用を測定することによって、本発明のマーカの発現の度合い等を測定することができる。

#### 【0071】

従って、本明細書においてポリヌクレオチドまたはポリペプチドなどの生物学的因子に対して「特異的に」相互作用する（または結合する）「因子」（または、薬剤、検出剤等）とは、そのポリヌクレオチドまたはそのポリペプチドなどの生物学的因子に対する親和性が、他の無関連の（特に、同一性が30%未満の）ポリヌクレオチドまたはポリペプチドに対する親和性よりも、代表的には同等またはより高いか、好ましくは有意に（例えば、統計学的に有意に）高いものを包含する。そのような親和性は、例えば、ハイブリダイゼーションアッセイ、結合アッセイなどによって測定することができる。

#### 【0072】

本明細書において第一の物質または因子が第二の物質または因子に「特異的に」相互作用する（または結合する）とは、第一の物質または因子が、第二の物質または因子に対して、第二の物質または因子以外の物質または因子（特に、第二の物質または因子を含む試料中に存在する他の物質または因子）に対するよりも高い親和性で相互作用する（または結合する）ことをいう。物質または因子について特異的な相互作用（または結合）としては、例えば、リガンド-レセプター反応、核酸におけるハイブリダイゼーション、タンパク質における抗原抗体反応、酵素-基質反応など、核酸およびタンパク質の両方が関係する場合、転写因子とその転写因子の結合部位との反応など、タンパク質-脂質相互作用、核酸-脂質相互作用などが挙げられるがそれらに限定されない。従って、物質または因子がともに核酸である場合、第一の物質または因子が第二の物質または因子に「特異的に相互作用する」ことには、第一の物質または因子が、第二の物質または因子に対して少なくとも一部に相補性を有することが包含される。また例えば、物質または因子がともにタンパク質である場合、第一の物質または因子が第二の物質または因子に「特異的に」相互作用する（または結合する）こととしては、例えば、抗原抗体反応による相互作用、レセプター-リガンド反応による相互作用、酵素-基質相互作用などが挙げられるがそれらに限定されない。2種類の物質または因子がタンパク質および核酸を含む場合、第一の物質または因子が第二の物質または因子に「特異的に」相互作用する（または結合する）ことには、転写因子と、その転写因子が対象とする核酸分子の結合領域との間の相互作用（または結合）が包含される。

#### 【0073】

本明細書においてポリヌクレオチドまたはポリペプチド発現の「検出」または「定量」は、例えば、マーカ-検出剤への結合または相互作用を含む、mRNAの測定および免疫学的測定方法を含む適切な方法を用いて達成され得るが、本発明では、PCR産物の量をもって測定することができる。分子生物学的測定方法としては、例えば、ノーザンプロット法、ドットプロット法またはPCR法などが例示される。免疫学的測定方法としては、例えば、方法としては、マイクロタイタープレートを用いるELISA法、RIA法、蛍光抗体法、発光イムノアッセイ(LIA)、免疫沈降法(IP)、免疫拡散法(SRID)、免疫比濁法(TIA)、ウェスタンプロット法、免疫組織染色法などが例示される。また、定量方法としては、ELISA法またはRIA法などが例示される。アレイ(例えば、DNAアレイ、プロテインアレイ)を用いた遺伝子解析方法によっても行われ得る。DNAアレイについては、(秀潤社編、細胞工学別冊「DNAマイクロアレイと最新PCR法」)に広く概説されている。プロテインアレイについては、Nat Genet. 2002 Dec; 32 Suppl: 526-32に詳述されている。遺伝子発現の分析法としては、上述に加えて、RT-PCR、RACE法、SSCP法、免疫沈降法、two-hybridシステム、in vitro翻訳などが挙げら

10

20

30

40

50

れるがそれらに限定されない。そのようなさらなる分析方法は、例えば、ゲノム解析実験法・中村祐輔ラボ・マニュアル、編集・中村祐輔羊土社（2002）などに記載されており、本明細書においてそれらの記載はすべて参考として援用される。

【0074】

本明細書において「手段」とは、ある目的（例えば、検出、診断、治療）を達成する任意の道具となり得るものをいい、特に、本明細書では、「選択的に認識（検出）する手段」とは、ある対象を他のものとは異なって認識（検出）することができる手段をいう。

【0075】

本発明により検出された結果は、免疫系の状態の指標として有用である。従って、本発明によって、免疫系の状態の指標を識別し、疾患の状態を知るために用いることができる。

10

【0076】

本明細書において「診断」とは、被験体における疾患、障害、状態などに関連する種々のパラメータを同定し、そのような疾患、障害、状態の現状または未来を判定することをいう。本発明の方法、装置、システムを用いることによって、体内の状態を調べることができ、そのような情報を用いて、被験体における疾患、障害、状態、投与すべき処置または予防のための処方物または方法などの種々のパラメータを選定することができる。本明細書において、狭義には、「診断」は、現状を診断することをいうが、広義には「早期診断」、「予測診断」、「事前診断」等を含む。本発明の診断方法は、原則として、身体から出たものを利用することができ、医師などの医療従事者の手を離れて実施することができることから、産業上有用である。本明細書において、医師などの医療従事者の手を離れて実施することができることを明確にするために、特に「予測診断、事前診断もしくは診断」を「支援」として称することがある。

20

【0077】

本発明の診断薬等の医薬等としての処方手順は、当該分野において公知であり、例えば、日本薬局方、米国薬局方、他の国の薬局方などに記載されている。従って、当業者は、本明細書の記載があれば、過度な実験を行うことなく、使用すべき量を決定することができる。

【0078】

（好ましい実施形態の説明）

30

以下に本発明の好ましい実施形態を説明する。以下に提供される実施形態は、本発明のよりよい理解のために提供されるものであり、本発明の範囲は以下の記載に限定されるべきでないことが理解される。従って、当業者は、本明細書中の記載を参酌して、本発明の範囲内で適宜改変を行うことができることは明らかである。これらの実施形態について、当業者は適宜、任意の実施形態を組み合わせ得る。

【0079】

（結合モードクラスター化技術）

1つの局面において、本発明は、(i)少なくとも2つの免疫実体 (immunological entity) の特徴量（例えば、配列情報）を提供するステップと、(ii)該特徴量に基づいて、該免疫実体の抗原特異性または結合モードの分析を機械学習させるステップと、(iii)該抗原特異性または結合モードの分類または異同の決定を行うステップとを含む、免疫実体の抗原特異性または結合モードを分析する方法を提供する。

40

【0080】

1つの実施形態では、本発明は、(i)少なくとも2つの免疫実体 (immunological entity) の特徴量（例えば、配列情報）を提供するステップと、(ii)該特徴量に基づいて、該免疫実体の抗原特異性または結合モード（例えば、「エピトープ」）の分析を、抗原特異性または結合モードを特定せずに機械学習させるステップと、(iii)該抗原特異性または結合モードの分類または異同の決定を行うステップとを含む、免疫実体の抗原特異性または結合モードを分析する方法を提供する。

50

## 【0081】

1つの実施形態では、本発明は、対ごとに免疫実体を評価する方法に関する。この実施形態では、本発明は、免疫実体の集合を解析する方法であって、該方法は：(a)該免疫実体の集合のメンバーの少なくとも1つの対について特徴量を抽出するステップと(b)該特徴量を用いた機械学習により該対について抗原特異性または結合モードの間の距離を算出し、または該抗原特異性または結合モードが一致するかどうかを判定するステップと、(c)該距離に基づいて該免疫実体の集合をクラスタリングするステップと、(d)必要に応じて該クラスタリングによる分類に基づいて解析するステップとを包含する、方法を提供する。

## 【0082】

本発明において「距離」の算出方法は代表的に以下のとおりである。まず既知の実験データから学習データを構築する。学習データにおいては、代表的に、免疫実体の対のアミノ酸配列情報とラベル情報(その対が同一のエピトープ/結合モードを持つかどうか、または同一の抗原分子に結合するかどうかの情報)を含む。学習データを得た実験手法によって得られるラベル情報が異なり得る。例えば、X線結晶構造解析では分子の結合情報が原子レベルで得られるため、結合モードの情報が得られる。次にこの実験データを用いて、代表的に、機械学習により同一エピトープ/結合モード/抗原に対するものを1、異なるものを0として学習させる。学習の結果、機械学習は与えられた免疫実体対が同一結合モード/エピトープ/抗原に結合する確率を返す。この確率が距離となる。本発明では、例示した手法と同様の他の手法を用いて算出することもできる。

## 【0083】

ここで、この実施形態では、機械学習による予測の際の「特徴量」の取り扱いについてみると、特徴量は特徴ベクトルとして機械学習アルゴリズムの入力として用いられる。本発明は、抗原特異性および結合モードのいずれでも分析することができる。抗原特異性は生物学的な定義で、結合モードは物理学的な定義であり、実質的に同一の対象を指すといえる。本発明の予測方法において、グループとしてまとめられるのは、物理学的な結合モードであり、結合モードが一義的に解析し得るが、複数の結合モードを含みうる抗原特異性も、結果的に解析することができる。

## 【0084】

本発明の実施形態において、クラスタリングは距離に基づいて算出することができる。例えば、対ごとの評価では、全体評価と異なり、ステップb)で距離を算出する。他方で、抗原特異性または結合モードが一致するかどうかの判定において、抗原特異性または結合モードの距離とは、抗原特異性または結合モードが一致するかどうかの判定を意味する。具体的には、1つの実施形態では、「距離」を0か1かで予測してしまうとクラスタリングは単なる1をまとめる作業となる。他方、別の実施形態では、距離を[0-1]で表現することで、クラスタリングのメリットは単なる距離(対の関係)ではなく、周辺にある対の密度など他のパラメータも考慮することができる。

## 【0085】

本発明の実施形態で利用され得る解析手法としては、クラスタリング結果からそれぞれのクラスターを遺伝子のように見なし、遺伝子発現解析のように用いることが挙げられる。具体的には、例えば、1.継時変化を追う場合、特定の、あるいは複数のクラスターに属する配列の増減を見る。増減しているクラスターの数を見る。それぞれのクラスターに特徴的な量(V/D/J遺伝子、CDRの長さ、親水性、疎水性、保存されている残基等)を見出す。2.複数検体の特定の層に興味がある場合、特定の層に優位に存在、増減しているクラスターを同定する。増減しているクラスターの数を見る。それぞれのクラスターに特徴的な量を見出す。3.機能に興味がある場合(機能=抗原特異性または結合モード)、興味ある抗原に特異的な配列(ELISPOTアッセイ、pMHCテトラマーによるsorting等、別の実験により得ているとする)を含むクラスターに注目し、その増減を見る。(機能=細胞の機能)別々にソーティングして配列決定した異なるサブタイプの細胞から得られたクラスタリング結果を比較する。4.別の実験ソースと比較する場合、遺伝子発現解析や、オミックス解析、細菌

10

20

30

40

50

叢、サイトカイン、細胞種の数の大小等との相関、あるいはそれらと組み合わせた1-3までの解析などを挙げることができる。

【0086】

「対ごと」の実施形態においては、対で1つの特徴量ベクトル、本明細書の別の箇所で説明する「全体」の実施形態においては、配列1つに対し1つの特徴量ベクトルを算出する。

【0087】

従って、対ごとで行う態様の1つの実施形態では、「免疫実体の集合のメンバーの少なくとも1つの対について特徴量を抽出するステップ」は代表的には以下のとおり行う。すなわち、まず、各配列の遺伝子情報や領域の情報を得；次に配列をCDR、フレームワーク等の領域に分け；全体、あるいは領域ごとの各配列の特徴量を得；対の特徴量として、それぞれの配列の特徴量の一致、一致度や差を求め；最後に一連の操作で得られた特徴量を一つの特徴量ベクトルとしてまとめることによって、特徴量を抽出し、対ごとの場合は、対で1つの特徴量ベクトルを算出することができる。

10

【0088】

全体で行う態様の1つの実施形態では、「免疫実体の集合のメンバーの少なくとも1つの対をなす配列それぞれについて特徴量を抽出するステップ」は代表的には以下のとおり行う。まず、各配列の遺伝子情報や領域の情報を得；次に配列をCDR、フレームワーク等の領域に分け；全体、あるいは領域ごとの各配列の特徴量を得；最後に一連の操作で得られた特徴量を一つの特徴量ベクトルとしてまとめることによって、特徴量を抽出し、全体的場合は、配列1つに対し1つの特徴量ベクトルを算出し総合として全体の特徴量ベクトルを抽出することができる。

20

【0089】

別の実施形態では、本発明は、対ごとに解析する方法において前記機械学習による計算を提供する。

【0090】

対ごとで行う態様の1つの実施形態では、「特徴量を用いた機械学習により該対について抗原特異性または結合モードの間の距離を算出する」ステップは代表的に以下のように行う。例えば、ランダムフォレスト、ブースティング等の手法を用いて特徴量（例えば、(a)で抽出された数値）から対の距離を算出する。

30

【0091】

対ごとで行う態様の1つの実施形態では、「抗原特異性または結合モードが一致するかどうか」の「判定」は、ある閾値（例えば、0.5、0.6等の適宜の数値）に基づく判定などの任意の手法で行うことができる。

【0092】

対ごとで行う態様とは異なり、全体で行う態様でおこなわれる実施形態では、「該特徴量を高次元ベクトル空間に射影するステップ」は、代表的には以下のとおり行う。すなわち、埋め込み(embedding)という技術を用い得る。埋め込みでは、代表的に、各配列からなる高次元ベクトルは機械学習により学習データから同一結合モード/エピトープ/抗原を認識するものを近くに、そうでないものを遠くに配置するように学習される。高次元ベクトル空間は機械学習によりこの配置が可能となるように選ばれる。

40

【0093】

対ごとで行う態様とは異なり、全体で行う態様で行われる実施形態では、「該メンバーの空間上の距離は該メンバーの機能類似性を反映する」とは以下を意味する。すなわち、機械学習により学習データから同一結合モード/エピトープ/抗原を認識するものを近くに、そうでないものを遠くに配置されている。従って、機能類似性を反映とは、より近い距離にある配列は類似機能を持つと期待されることを意味する。このステップは結合距離に基づく単純な閾値に基づくもの、階層的クラスタリング、非階層的クラスタリング、あるいはその組み合わせで行うことができる。

【0094】

50

クラスタリングは、対ごとで行う態様でも全体で行う態様でも同様のものを用いることができ、「距離に基づいて該免疫実体の集合をクラスタリングするステップ」では、代表的には、例えば、距離に基づく単純な閾値に基づくもの、階層的クラスタリング、非階層的クラスタリング法、あるいはその組み合わせを用いる。クラスタリング結果が学習セットの偽陽性を最小化する、ランド指数/マシューの相関係数(Matthews correlation coefficient: MCC)を最大化する、偽陽性を一定の割合未満にしながらランド指数/MCCを最大化する等、目的と学習セットの正解ラベルの種類(結合モード/エピトープ/抗原)に応じ望ましい結果を得るように最適なクラスタリングのパラメータを用いることができる。

#### 【0095】

全体または対ごとでのいずれの態様でも実施することができる1つの実施形態では、具体的には、クラスタリングによる分類に基づいて解析するステップは、例えば、クラスタリング結果からそれぞれのクラスターを遺伝子のように見なし、遺伝子発現解析のように用いることが挙げられる。具体的には、1. 継時変化を追う場合、特定の、あるいは複数のクラスターに属する配列の増減を見る。増減しているクラスターの数を見る。それぞれのクラスターに特徴的な量(V/D/J遺伝子、CDRの長さ、親水性、疎水性、保存されている残基等)を見出す。2. 複数検体の特定の層に興味がある場合、特定の層に優位に存在、増減しているクラスターを同定する。増減しているクラスターの数を見る。それぞれのクラスターに特徴的な量を見出す。3. 機能に興味がある場合、(機能=抗原特異性または結合モード)興味ある抗原に特異的な配列(ELISPOTアッセイ、pMHCテトラマーによるsorting等、別の実験により得ているとする)を含むクラスターに注目し、その増減を見る。(機能=細胞の機能)別々にソーティングして配列決定した異なるサブタイプの細胞から得られたクラスタリング結果を比較する。4. 別の実験ソースと比較する場合、遺伝子発現解析や、オミックス解析、細菌叢、サイトカイン、細胞種の数の大小等との相関、あるいはそれらと組み合わせた1-3までの解析で行うことができる。そして、具体的な実施形態では、「該クラスタリングによる分類に基づいて」行われる「解析」には、代表的には、バイオマーカーの同定、あるいは治療ターゲットとなる免疫実体または該免疫実体を含む細胞の同定のいずれか1つまたは複数を含む解析が挙げられるがこれらに限定されない。例えば、単一あるいは複数のクラスターを用いて、患者由来等興味ある検体、検体群に特異的なクラスター、あるいは一群のクラスターに属する免疫実体発現の有無、発現量、あるいは発現パターンの差異を統計的に評価することにより、疾患の有無や診断、予後、再燃可能性、重症度、ワクチンの有効性等を予測するバイオマーカーとして、あるいは自己免疫疾患等治療ターゲットとなる病原性免疫実体及びそれを発現する細胞の探索、細胞療法やワクチンの開発ターゲットとすべき免疫実体の同定を行うことができる。

#### 【0096】

本発明の実施形態では、例えば、ディープラーニング(Deep Learning)で用いられるオートエンコーダ(auto-encoder)によって全体評価を行う場合には、免疫実体の配列自身を入力として高次元ベクトル空間に射影する。オートエンコーダ自身が特徴量を抽出して高次元ベクトル空間に射影する。この場合、抽出された特徴量はそのまま高次元ベクトル空間要素となる。

#### 【0097】

1つの実施形態では、前記免疫実体は抗体、抗体の抗原結合断片、B細胞受容体、B細胞受容体の断片、T細胞受容体、T細胞受容体の断片、キメラ抗原受容体(CAR)、またはこれらのいずれかまたは複数を含む細胞である。

#### 【0098】

1つの実施形態では、本発明は、全体評価での免疫実体の集合の解析を提供する。その実施形態では、本発明は、免疫実体の集合を解析する方法であって、該方法は:(aa)該免疫実体の集合のメンバーの少なくとも1つの対をなす配列それぞれについて特徴量を抽出するステップと、(bb)該特徴量を高次元ベクトル空間に射影し、ここで、該メンバーの空間上の距離は該メンバーの機能類似性を反映する、ステップと、(cc)該距離

10

20

30

40

50

に基づいて該免疫実体の集合をクラスタリングするステップと、(d d) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを包含する、方法を提供する。

【0099】

全体で行う場合の1つの実施形態では、免疫実体の集合のメンバーの少なくとも1つの対をなす配列それぞれについて特徴量を抽出するステップは、対ごとに行うものと同様に実施することができ、例えば、オートエンコーダにより提供することなども例示することができる。

【0100】

全体で行う場合の1つの実施形態では、特徴量を高次元ベクトル空間に射影し、ここで、該メンバーの空間上の距離は該メンバーの機能類似性を反映する、ステップは、対ごとに行うものと同様に実施することができる。このように、空間上の距離を得た後のクラスタリング以降は同様に、対ごとであっても、全体であっても同様に実施することができる。

10

【0101】

例えば、1つの実施形態において、全体で行う場合、前記高次元ベクトル空間計算(b)は、例えば、教師あり、半教師あり(Siamese network)、教師なし(Auto-encoder)のいずれかの方法で行うことができるが、これに限定されない。

【0102】

1つの実施形態では、距離に基づいて免疫実体の集合をクラスタリングするステップは、具体的には以下のようにして行うことができる。例えば、高次元空間上の距離に基づく単純な閾値に基づくもの、階層的クラスタリング、あるいは非階層的クラスタリング法、あるいはその組み合わせを用いることができる。この時、クラスタリング結果が学習セットの偽陽性を最小化する、ランド指数/マシューの相関係数(Matthews correlation coefficient: MCC)を最大化する、偽陽性を一定の割合未満にしながらランド指数/MCCを最大化する等、目的と学習セットの正解ラベルの種類(結合モード/エピトープ/抗原)に応じ望ましい結果を得るように最適なクラスタリングのパラメータを用いるなどの、種々の手順で処理することで、実施することができる。

20

【0103】

1つの実施形態では、クラスタリングによる分類に基づいて解析するステップは、例えば、クラスタリング結果からそれぞれのクラスターを遺伝子のように見なし、遺伝子発現解析のように用いることが挙げられる。具体的には、1. 継時変化を追う場合、特定の、あるいは複数のクラスターに属する配列の増減を見る。増減しているクラスターの数をみる。それぞれのクラスターに特徴的な量(V/D/J 遺伝子、CDRの長さ、親水性、疎水性、保存されている残基等)を見出す。2. 複数検体の特定の層に興味がある場合、特定の層に優位に存在、増減しているクラスターを同定する。増減しているクラスターの数をみる。それぞれのクラスターに特徴的な量を見出す。3. 機能に興味がある場合、(機能=抗原特異性または結合モード)興味ある抗原に特異的な配列(ELISPOTアッセイ、pMHCテトラマーによるsorting等、別の実験により得ているとする)を含むクラスターに注目し、その増減を見る。(機能=細胞の機能)別々にソーティングして配列決定した異なるサブタイプの細胞から得られたクラスタリング結果を比較する。4. 別の実験ソースと比較する場合、遺伝子発現解析や、オミックス解析、細菌叢、サイトカイン、細胞種の数的大小等との相関、あるいはそれらと組み合わせた1-3までの解析で行うことができる。

30

40

【0104】

上記はあくまで一例であり、本発明を実施するために、より多くの項を含んだより複雑な関数型を用いることもできる。

【0105】

1つの実施形態では、前記機械学習は、回帰的な手法、ニューラルネットワーク法や、サポートベクトルマシン、ランダムフォレストといった機械学習アルゴリズムからなる群より選択される。

50

## 【0106】

本発明の評価ステップは免疫実体結合物（例えば、抗原）が既知という特別なケースや、一部の抗体ターゲットを知っている場合には応用として、これら既知のケースをクラスタリングに含むことができる。すなわち、免疫実体結合物（例えば、抗原）/エピトープ（抗原特異性や結合モード）既知の免疫実体（例えば、抗体）を用いることで、免疫実体（例えば、抗体）の免疫実体結合物（例えば、抗原）/エピトープ（抗原特異性や結合モード）を予測することができる。

## 【0107】

本明細書で記載されるクラスター分類したエピトープは、生体情報と関連付けることができる。例えば、本発明の分類方法に基づいて同定されたエピトープの一つまたは複数のクラスターに基づき、前記抗体の保有者を既知の疾患または障害あるいは生体の状態と関連付けることができる。

10

## 【0108】

本発明が関与し得る疾患または障害あるいは生体の状態には、例えば、異物（例えば、細菌やウイルス等）の感染状態のほか、非自己と認識される自己由来の実体（例えば、新生成物（がん、腫瘍）や自己免疫疾患に関連する実体）がある。免疫系は、生物にとって内因性の分子（「自己」分子）を、生物に対する外因性または外来性の物質（「非自己分子」と識別するように機能する。免疫系は、応答を媒介する構成成分に基づいて異物に対して2つのタイプの適応応答（体液性応答および細胞性応答）を有する。体液性応答は抗体により媒介され、他方で、細胞性免疫はリンパ球として類別される細胞に関与する。最近の抗がんおよび抗ウイルス戦略において、抗がんもしくは抗ウイルス治療または療法的手段として、宿主免疫系を利用することが一つの重要な戦略となっている。本発明の分類およびクラスター化技術は、体液性応答および細胞性応答のいずれの戦略でも応用することができる。

20

## 【0109】

免疫系は、宿主の異物からの防御において、3つの段階（認識、活性化、およびエフェクター）を経て機能する。認識段階では、免疫系は、身体中の外来抗原または侵入物の存在を認識し、それを認知させる。外来抗原は、例えば異物（ウイルスタンパク質由来の細胞表面マーカー等）のほか、非自己と認識される細胞（がん細胞）の細胞表面マーカー等であり得る。免疫系が侵入物を認識すると、免疫系の抗原特異的細胞は、侵入物誘発性シグナルに応答して増殖および分化する（活性化段階）。最終的に、免疫系のエフェクター細胞が検出された侵入物に応答して、それを中和するエフェクター段階である。エフェクター細胞は免疫応答を実行する役割を担う。エフェクター細胞としては、B細胞、T細胞やナチュラルキラー（NK）細胞等が挙げられる。B細胞は、侵入物に対する抗体を生成し、抗体は補体系と組み合わせて、特定の標的である免疫実体、エピトープ、抗原特異性または結合モード（抗原等の免疫実体結合物なども含む）を含むかあるいは関連する細胞ないし生物を破壊へと導く。T細胞は、ヘルパーT細胞、制御性T細胞、細胞傷害性T細胞（CTL細胞）等の種類があり、ヘルパーT細胞はサイトカインを分泌し、他の細胞の増殖等を刺激し免疫応答の有効性を強化する。制御性T細胞は免疫応答を下方制御する。CTL細胞は、表面上に外来抗原を提示する細胞を直接溶解・融解することで破壊する。NK細胞は、ウイルス感染細胞や悪性腫瘍細胞等を認識し破壊するとされる。したがって、これらのエフェクター細胞が対象とするあるいは関連性の高い、免疫実体、エピトープ、抗原特異性または結合モードを分類し、これを疾患または障害あるいは生体の状態と結びつけることは、治療や診断の有効性に非常の重要な役割を果たすといえる。

30

40

## 【0110】

このように、T細胞は、特定の抗原シグナルに応答して機能する抗原特異的免疫細胞である。Bリンパ球およびそれらが産生する抗体は、また抗原特異的物体である。本発明は、これらの特定の免疫実体結合物（例えば、抗原）について、免疫実体、エピトープ、抗原特異性または結合モードのクラスターを用いて分類し、最終的な機能（特定の疾患または障害あるいは生体の状態との関連）ごとに分類し、クラスター化することができること

50

を提供する。

【0111】

上述のようにB細胞は遊離型または可溶性の抗原に应答するが、T細胞はこれらに应答しない。T細胞が抗原に应答するためには、抗原がペプチドにプロセシングされ、腫瘍組織適合性複合体(MHC)でコードされる提示構造に結合されることが必要である(「MHC拘束」と呼ばれる)。T細胞はこのメカニズムにより自己細胞と非自己細胞とを識別する。抗原が認識可能なMHC分子により提示されない場合、T細胞は抗原シグナルを認識しない。認識可能なMHC分子に結合したペプチドに特異的なT細胞はMHCペプチド複合体に結合し免疫応答が進行する。MHCには2つのクラスがあり(クラスI MHC、クラスII MHC)、CD4<sup>+</sup>T細胞はクラスII MHCタンパク質を優先的に相互作用し、他方で、細胞傷害性T細胞(CD8<sup>+</sup>)はクラスI MHCと優先的に相互作用するとされる。これらのMHCタンパク質はいずれのクラスのものも、細胞の外部表面上にその大部分の構造が含まれる膜貫通タンパク質であり、その外部にペプチド結合間隙がある。この間隙には内因性、外来性のいずれのタンパク質の断片も細胞外環境に結合および提示される。この際、プロフェッショナル抗原提示細胞(pAPC)と呼ばれる細胞が、MHCタンパク質を用いてT細胞に対する抗原を提示し、種々の特定の今日刺激分子を用いてT細胞がとる分化、活性化の経路を誘導し、免疫系の効果を実現する。本発明の免疫実体、エピトープ、抗原特異性または結合モードの分類およびクラスター化技術は、これらのMHCが関与する治療や診断に関しても従来提供できない応用法を提供する。

10

【0112】

非自己実体については、従来の免疫系を十分活用することで、治療や診断に関する応用法を提供することができるが、自己についてはさらなる工夫が必要でありうる。がん細胞等は正常細胞と由来を同じくし、遺伝子レベルでは正常細胞と実質的に同一であるからである。ただし、がん細胞は腫瘍関連抗原(TuAA)を提示することが知られており、この抗原または他の免疫実体結合物を活用することで、被験者の免疫系を活用しがん細胞を攻撃することができる。このような腫瘍関連抗原もまた、本発明の技術により免疫実体、エピトープ、抗原特異性または結合モードを指標に分類し、クラスター化することができる。例えば、腫瘍関連抗原を応用し抗がんワクチン等に応用することができる。従来例えば、活性化腫瘍細胞全体を使用する技術が米国特許第5,993,828号で開示されている。あるいは単離された腫瘍抗原を含有する組成物を応用する技術も試みられている(例えば、Krishnadas DK et al., Cancer Immunol Immunother. 2015 Oct;64(10):1251-60)。同定されたエピトープを認識するキメラ抗原受容体(CAR)を用いた遺伝子改変T細胞(CAR-Tとも称する。)を用いることもできる。また、PD-1やPD-L1などの免疫チェックポイントに関する作用に基づく免疫チェックポイント阻害剤等を利用した免疫療法も最近注目を浴びている。PD-1は抗原提示細胞に発現するPD-1リガンド(PD-L1及びPD-L2)と結合し、リンパ球に抑制性シグナルを伝達してリンパ球の活性化状態を負に調節している。PD-1リガンドは抗原提示細胞以外にヒトの様々な腫瘍組織に発現しており、悪性黒色腫においても切除した腫瘍組織におけるPD-L1の発現と術後の生存期間との間に負の相関関係があるとされている。PD-1抗体やPD-L1抗体でPD-1とPD-L1との結合を阻害するとその細胞傷害活性が回復するとされており、抗原特異的なT細胞の活性化及びがん細胞に対する細胞傷害活性を増強することで持続的な抗腫瘍効果を示すことができる(例えば、ニボルマブ等)。このような免疫活性の負の調節機構をもとに戻すメカニズムについても、本発明のエピトープの分類、クラスター化法を応用することができる。

20

30

40

【0113】

本発明の1つの実施形態では、ワクチンについては、ウイルス疾患についても本発明の免疫実体、エピトープ、免疫実体結合物、抗原特異性または結合モードの分類、クラスター化法を応用することができる。ウイルスに対するワクチンは、弱毒化生ウイルスのほか、不活化ワクチンのほか、サブユニットワクチン等が利用されている。サブユニットワクチンの成功率は高くないが、エンベロープタンパク質に基づいた組換えB型肝炎ワクチン

50

等での成功例が報告されている。本発明の免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードの分類、クラスター化法を用いると、適切に生体の状態を関連付けることができるため、サブユニットワクチン等での有効性も上昇すると考えられる。また、適切なクラスターの定量的な評価によって、ワクチンの有効性評価にもつながると考えられる。またあるワクチンが有効な症例との比較により、層別化も可能である。結果として有効性が上がる、あるいは上市の可能性が高まることも考えられる。実際に本発明の手法を用いて、ワクチンに反応するクラスターをインシリコで同定した結果が示されている。

#### 【0114】

1つの実施形態では、本発明の免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードの分類、クラスター化法で使用されうる免疫実體として抗体、抗体の抗原結合断片、B細胞受容体、B細胞受容体の断片、T細胞受容体、T細胞受容体の断片、キメラ抗原受容体(CAR)、これらのいずれかまたは複数を含む細胞(例えば、キメラ抗原受容体(CAR)を含むT細胞(CAR-T))等を挙げることができる。

#### 【0115】

ここで、本発明は、本発明の手法に基づいて分類された免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードのクラスターを生成する方法を提供し、ここで、この方法は、結合するエピトープが同一である免疫実體を同一のクラスターに分類する工程を包含する。また、1つの実施形態では、免疫実體、エピトープまたは免疫実體結合物を、その特性および既知の免疫実體、エピトープまたは免疫実體結合物との類似性からなる群より選択される少なくとも1つの評価項目を評価し、所定の基準を満たした免疫実體、エピトープまたは免疫実體結合物を対象に前記クラスター分類を行うことができる。複数の前記免疫実體、エピトープまたは免疫実體結合物が同一である場合、該免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードの三次元構造が少なくとも一部または全部重複することがあり、複数の前記エピトープが同一である場合、該免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードに関連するアミノ酸配列が少なくとも一部または全部重複することがある。

#### 【0116】

##### <抗原特異性・結合モードと抗原分類>

さらに別の局面では、本発明は、本発明の方法で同定された抗原特異性または結合モードを有するかあるいはそれらに基づく構造を有する免疫実體、エピトープ、免疫実體結合物、抗原特異性、結合モード、抗原(またはそれに対応する免疫実體結合物)、あるいはそれらのクラスターを提供する。ここで定義される免疫実體、エピトープ、免疫実體結合物、抗原特異性、結合モード、抗原等は、本明細書の<(結合モードクラスター化技術)>に記載される任意の特徴を有し得、あるいはそれらの技術で同定、分類またはクラスター化されたものでありうる。ここで、クラスターを生成する方法としては、結合するエピトープ、免疫実體結合物、抗原特異性または結合モードが同一である免疫実體を同一のクラスターに分類する工程、あるいは、結合する免疫実體、抗原特異性または結合モードが同一であるエピトープまたは免疫実體結合物を同一のクラスターに分類する工程を包含することを挙げることができる。好ましい実施形態では、免疫実體、エピトープまたは免疫実體結合物を、その特性および既知の免疫実體、エピトープまたは免疫実體結合物との類似性からなる群より選択される少なくとも1つの評価項目を評価し、所定の基準を満たした免疫実體を対象にクラスター分類を行うことができる。ここで採用され得る基準としては、例えば、複数の前記免疫実體、エピトープ、免疫実體結合物抗原特異性または結合モードが同一である場合、該免疫実體、エピトープ、免疫実體結合物抗原特異性または結合モードの三次元構造が少なくとも一部重複することがあり得、あるいは、複数の前記免疫実體、エピトープ、または免疫実體結合物の抗原特異性または結合モードが同一である場合、該エピトープまたは免疫実體結合物のアミノ酸配列または化学構造の少なくとも一部が重複してもよい。

#### 【0117】

本発明の1つの実施形態は、分類された免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モード、またはクラスター化された免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードおよび上記免疫実體、エピトープ、抗原特異性または結合モードを含むかまたは関連する免疫実體結合物(例えば、抗原)またはポリペプチドに関する。

【0118】

ここで、分類された免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードまたはクラスター化された免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードの記述(同定)方法として以下を挙げるができる。すなわち、本発明の手法で同定された免疫実體(例えば、抗体)、エピトープまたは免疫実體結合物のクラスターは高い精度で同一の免疫実體、エピトープ、免疫実體結合物などのパートナーを認識し、または抗原特異性または結合モードを有するものと考えられるため、クラスターが認識する免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードの同定には、エピトープまたは免疫実體結合物(例えば、抗原)が既知の免疫実體(例えば、抗原既知抗体)に対する類似性評価や、実験的な抗原スクリーニング(または、他の免疫実體結合物のスクリーニング)、さらに望ましくは抗原-抗体ペア(あるいは、他の免疫実體-免疫実體結合物)の変異体実験、NMR化学シフト、結晶構造解析、相互作用に関わる免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードの同定、またはインビトロもしくはインビボ実験により機能評価を行い同定することができる。従って、既存の免疫実體、エピトープ、免疫実體結合物、抗原特異性または結合モードおよびそれに基づく免疫実體等が提供されるとしても、本発明のようにクラスター化または分類されたものは、特定の情報を持ったものであり、特定の用途に使用することができ、特定の効果および機能を有するものということができ、その点で従来のエピトープまたは免疫実體結合物(例えば、抗原)およびそれに基づく免疫実體にない新たな特徴を付与するものであり、新規かつ顕著な特徴のある技術的事項を提供するといえる。

【0119】

<プログラム、媒体、システム構成>

1つの局面では、本発明は、本発明の方法を実行させるプログラムを提供する。ここで採用され得る任意の特徴は本明細書の<結合モードクラスター化技術>に記載される任意の特徴またはその組み合わせでありうる。

【0120】

1つの局面では、たとえば、(i)少なくとも2つの免疫実體(immunological entity)の特徴量を提供するステップと、(ii)該特徴量に基づいて、抗原特異性または結合モードを特定せずに該免疫実體の抗原特異性または結合モードの分析を機械学習させるステップと、(iii)該抗原特異性または結合モードの分類または異同の決定を行うステップとを含む、免疫実體の集合を解析する方法を実行させるプログラムを提供する。

【0121】

あるいは、1つの局面では、たとえば、免疫実體の集合を解析する方法であって、該方法は：

- (a) 該免疫実體の集合のメンバーの少なくとも1つの対について特徴量を抽出するステップと、
- (b) 該特徴量を用いた機械学習により該対について抗原特異性または結合モードの間の距離を算出し、または該抗原特異性または結合モードが一致するかどうかを判定するステップと、
- (c) 該距離に基づいて該免疫実體の集合をクラスタリングするステップと、
- (d) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを包含する、方法を実行させるプログラムを提供する。

【0122】

あるいは、1つの局面では、たとえば、免疫実體の集合を解析する方法であって、該方

10

20

30

40

50

法は：

( a a ) 該免疫実体の集合のメンバーの少なくとも1つの対をなす配列それぞれについて特徴量を抽出するステップと、

( b b ) 該特徴量を高次元ベクトル空間に射影し、ここで、該メンバーの空間上の距離は該メンバーの機能類似性を反映する、ステップと、

( c c ) 該距離に基づいて該免疫実体の集合をクラスタリングするステップと、

( d d ) 必要に応じて該クラスタリングによる分類に基づいて解析するステップとを包含する、方法を実行させるプログラムを提供する。

【 0 1 2 3 】

上述のプログラムにおいて、ここで採用され得る任意の特徴は本明細書の<結合モードクラスタ化技術>に記載される任意の特徴またはその組み合わせでありうる。

10

【 0 1 2 4 】

別の局面では、本発明は、本発明の方法を実行させるプログラムを格納した記録媒体を提供する。1つの実施形態では、記録媒体は、内部に格納され得るROMやHDD、磁気ディスク、USBメモリ等のフラッシュメモリなどの外部記憶装置でありうる。ここで採用され得る任意の特徴は本明細書の<結合モードクラスタ化技術>に記載される任意の特徴またはその組み合わせでありうる。本発明の記録媒体は、上記本発明の上記プログラムを格納するものであり得る。

【 0 1 2 5 】

別の局面では、本発明は、本発明の方法を実行させるプログラムを含むシステムを提供する。ここで採用され得る任意の特徴は本明細書の<結合モードクラスタ化技術>に記載される任意の特徴またはその組み合わせでありうる。1つの実施形態では、本発明のシステムは、( I ) 少なくとも2つの免疫実体 ( immunological entity ) の特徴量を提供する特徴量提供部と、( I I ) 該特徴量に基づいて、抗原特異性または結合モードを特定せずに該免疫実体の抗原特異性または結合モードの分析を機械学習させる機械学習部と、( I I I ) 該抗原特異性または結合モードの分類または異同の決定を行う分類部とを備える免疫実体の集合を解析するシステムを提供する。ここで採用され得る任意の特徴は本明細書の<結合モードクラスタ化技術>に記載される任意の特徴またはその組み合わせでありうる。これらの各部分は別々の構成要素で実現されてもよく、これら2つ以上が、1つの構成要素によって実現されていてもよい。

20

30

【 0 1 2 6 】

他の実施形態では、本発明は、( A ) 該免疫実体の集合のメンバーの少なくとも1つの対について特徴量を抽出する特徴量抽出部または特徴量提供部と、( B ) 該特徴量を用いた機械学習により該対について抗原特異性または結合モードの間の距離を算出し、または該抗原特異性または結合モードが一致するかどうかを判定する判定部と、( C ) 該距離に基づいて該免疫実体の集合をクラスタリングするクラスタリング部と、( D ) 必要に応じて該クラスタリングによる分類に基づいて解析する解析部とを備える免疫実体の集合を解析システムを提供する。ここで採用され得る任意の特徴は本明細書の<結合モードクラスタ化技術>に記載される任意の特徴またはその組み合わせでありうる。これらの各部分は別々の構成要素で実現されてもよく、これら2つ以上が、1つの構成要素によって実現されていてもよい。

40

【 0 1 2 7 】

他の局面では、本発明は、免疫実体の集合を解析するシステムを提供し、このシステムは、( A ) 該免疫実体の集合のメンバーの少なくとも1つの対をなす配列それぞれについて特徴量を抽出する特徴量抽出部または特徴量提供部と、( B ' ) 該特徴量を高次元ベクトル空間に射影し、ここで、該メンバーの空間上の距離は該メンバーの機能類似性を反映する、射影部と、( C ) 該距離に基づいて該免疫実体の集合をクラスタリングするクラスタリング部と、( D ) 必要に応じて該クラスタリングによる分類に基づいて解析する解析部とを備える。ここで採用され得る任意の特徴は本明細書の<結合モードクラスタ化技術>に記載される任意の特徴またはその組み合わせでありうる。これらの各部分は別々の構

50

成要素で実現されてもよく、これら2つ以上が、1つの構成要素によって実現されていてもよい。

【0128】

次に、図5の機能ブロック図を参照して、本発明のシステム1000の構成を説明する。なお、本図においては、単一のシステムで実現した場合を示しているが、複数のシステムで実現される場合も本発明の範囲に包含されることが理解される。

【0129】

本発明のシステム1000は、コンピュータシステムに内蔵されたCPU1001にシステムバス1020を介してRAM1003、ROMやHDD、磁気ディスク、USBメモリ等のフラッシュメモリなどの外部記憶装置1005及び入出力インターフェース(I/F)1025が接続されて構成される。入出力I/F1025には、キーボードやマウスなどの入力装置1009、ディスプレイなどの出力装置1007、及びモデムなどの通信デバイス1011がそれぞれ接続されている。外部記憶装置1005は、情報データベース格納部1030とプログラム格納部1040とを備えている。何れも、外部記憶装置1005内に確保された一定の記憶領域である。

【0130】

このようなハードウェア構成において、入力装置1009を介して各種の指令(コマンド)が入力されることで、又は通信I/Fや通信デバイス1011等を介してコマンドを受信することで、この記憶装置1005にインストールされたソフトウェアプログラムがCPU1001によってRAM1003上に呼び出されて展開され実行されることで、OS(オペレーションシステム)と協働して本発明の機能を奏するようになっている。もちろん、このような協働する場合以外の仕組みでも本発明を実装することは可能である。

【0131】

本発明の実装において、免疫実体(これらは、抗体、B細胞受容体またはT細胞受容体等でありうる)のアミノ酸配列またはこれと同等の情報(例えば、これをコードする核酸配列等)、その他の特徴量は、入力装置1009を介して入力され、あるいは、通信I/Fや通信デバイス1011等を介して入力されるか、あるいは、データベース格納部1030に格納されたものであってもよい。特徴量に基づいて抗原特異性または結合モードを特定せずに該免疫実体の抗原特異性または結合モードの分析を機械学習させるステップ; 該抗原特異性または結合モードの分類または異同の決定を行うステップ; 該特徴量を用いた機械学習により該対について抗原特異性または結合モードの間の距離を算出し、または該抗原特異性または結合モードが一致するかどうかを判定するステップ; 該距離に基づいて該免疫実体の集合をクラスタリングするステップ; 必要に応じて該クラスタリングによる分類に基づいて解析するステップ; 特徴量を高次元ベクトル空間に射影し、ここで、該メンバーの空間上の距離は該メンバーの機能類似性を反映する、ステップ; 該距離に基づいて該免疫実体の集合をクラスタリングするステップ; 必要に応じて該クラスタリングによる分類に基づいて解析するステップなどは、プログラム格納部1040に格納されたプログラム、または、入力装置1009を介して各種の指令(コマンド)が入力されることで、又は通信I/Fや通信デバイス1011等を介してコマンドを受信することで、この外部記憶装置1005にインストールされたソフトウェアプログラムによって実行することができる。取得されたデータや分割されたデータは、出力装置1007を通じて出力されるかまたは情報データベース格納部1030等の外部記憶装置1005に格納されてもよい。データは、出力装置1007を通じて出力されるかまたは情報データベース格納部1030等の外部記憶装置1005に格納されてもよい。

【0132】

データベース格納部1030には、これらのデータや計算結果、もしくは通信デバイス1011等を介して取得した情報が随時書き込まれ、更新される。各入力配列セット中の各々の配列、参照データベースの各遺伝子情報ID等の情報を各マスタテーブルで管理することにより、蓄積対象となるサンプルに帰属する情報を、各マスタテーブルにおいて定義されたIDにより管理することが可能となる。

10

20

30

40

50

【 0 1 3 3 】

データベース格納部 1 0 3 0 には、上記計算結果は、疾患、障害、生体情報等の既知の情報と関連付けて格納されてもよい。このような関連付けは、ネットワーク（インターネット、イントラネット等）を通じて入手可能なデータをそのまままたはネットワークのリンクとしてなされてもよい。

【 0 1 3 4 】

また、プログラム格納部 1 0 4 0 に格納されるコンピュータプログラムは、コンピュータを、上記した処理システム、機械学習、解析、射影、距離計算、分類、分割などの計算または処理等を行う処理を実施するシステムとして構成するものである。これらの各機能は、それぞれが独立したコンピュータプログラムやそのモジュール、ルーチンなどであり、上記 CPU 1 0 0 1 によって実行されることでコンピュータを各システムや装置として構成させるものである。なお、以下においては、それぞれのシステムにおける各機能が協働してそれぞれのシステムを構成しているものとする。

10

【 0 1 3 5 】

1 つの局面において、本発明は、データベースを用いて被験体のエピトープまたはそのクラスターを解析し、および/または診断もしくは診断結果に基づき治療する方法を提供する。この方法および本明細書で説明される 1 つまたは複数の異なる特徴を含む方法を、本明細書において「本発明の免疫実体の効率的クラスタリング」とも呼ぶ。そして本発明のレパトア解析法を実現するシステムを「本発明の免疫実体の効率的クラスタリング解析システム」ともいう。

20

【 0 1 3 6 】

本発明の免疫実体の効率的クラスタリングシステムを図 5 に示し、その具体的なアルゴリズムである本発明の免疫実体の効率的クラスタリング解析システムは、図 6 に例示されている。

【 0 1 3 7 】

図 6 では、S 1 0 0（ステップ（1））において、特徴量の提供又は抽出が実施される。対ごとにする場合、データセット上の全ての対について、特徴量を抽出する。全体の場合、データセット上の全ての配列を高次元ベクトル空間（空間上での距離が、配列間の機能類似性を反映する）に射影する。

30

【 0 1 3 8 】

S 1 5 0（ステップ（1A））では、対ごとで実施する場合、機械学習による予測がなされる。ここでは、データセット上の全ての対について、抗原特異性（結合モード）が一致するかどうか判定される。

【 0 1 3 9 】

S 2 0 0（ステップ（2））では、クラスタリングを行う。対ごとの評価の場合、データセット上の全ての対について、予測された配列対の間の距離に応じてクラスタを作成する。全体の場合、クラスタリングは、データセット上の全ての対について、抗原特異性（結合モード）が一致するかどうかを判定する。

【 0 1 4 0 】

S 3 0 0（ステップ（3））では、解析を行う。

40

【 0 1 4 1 】

提供されるデータは、外部記憶装置 1 0 0 5 に保存されたものであってもよいが、通常は、通信デバイス 1 0 1 1 を通じて、公共で提供されるデータベースとして取得することができる。あるいは入力装置 1 0 0 9 を用いて入力し、必要に応じて RAM 1 0 0 3 または外部記憶装置 1 0 0 5 に記録してもよい。ここでは、免疫実体の配列情報や他の特徴量を含むデータベースが提供される。配列情報や他の特徴量はまた、実際に得られた試料の配列を決定することによって入手され得る。RNA または DNA を腫瘍および健常組織から、各組織からポリ A + RNA を単離して cDNA を調製し、標準プライマーを用いて cDNA の配列決定を行うことによって、単離され、配列情報を入手し得る。かかる技術は当分野において周知である。また、患者のゲノムの全てまたは一部の配列決定が、当該分

50

野において周知である。ハイスループットDNA配列決定法は当分野において公知であり、例えば、イルミナ（登録商標）配列決定技術によるMiSeq（商標）シリーズのシステムを含む。これは、大規模パラレルSBS手法を用いて、数十億の塩基の高品質のDNA配列を1処理あたりに生成する。あるいは、抗体のアミノ酸配列を質量分析によって決定することもできる。本発明のシステムにおいてS100を実現する部分は特徴量提供部とも呼ばれる。

#### 【0142】

<組成物、治療、診断、医薬等>

本発明はまた、実施形態としては、上述の分類またはクラスター化された免疫実體、エピトープ、ポリペプチド、免疫実體結合物（例えば、抗原；抗原としては、エピトープを含むペプチド等の他、糖鎖等翻訳後修飾を含むもの、DNA/RNAといった核酸、低分子も含まれる）、免疫実體または免疫実體結合物またはクラスターに対して実質的類似性を有するか、同一クラスターに属する抗原特異性または結合モードに関連するポリペプチドを含む。他の好ましい実施形態としては、上記のいずれかに対して機能的類似性を有するポリペプチドを含む。さらなる実施形態は、本発明は、上述の分類またはクラスター化されたエピトープ、ポリペプチド、免疫実體結合物（例えば、抗原）、またはクラスター、ならびにそれらに対して実質的類似性を有するポリペプチド、同一クラスターに属する抗原特異性または結合モードに関連するポリペプチドをコードする核酸を含む。ここで採用され得る任意の特徴は本明細書の<結合モードクラスター化技術>に記載される任意の特徴またはその組み合わせ、あるいはそれらの技術で同定、分類またはクラスター化されたものでありうる。

#### 【0143】

1つの実施形態では、本発明の免疫実體、エピトープまたは免疫実體結合体であるポリペプチド、あるいは抗原特異性または結合モードを含む免疫実體、エピトープまたは免疫実體結合体、クラスターまたはそれらを含むポリペプチドは、HLA-A2分子に対して親和性を有することができる。親和性は、結合アッセイ、エピトープ認識の制限アッセイ、予測アルゴリズム等により決定することができる。エピトープ、クラスターまたはそれらを含むポリペプチドは、HLA-B7、HLA-B51分子等に対して親和性を有することができる。

#### 【0144】

本発明の他の実施形態では、本発明は、本発明で分類またはクラスター化された免疫実體、エピトープまたは免疫実體結合体、あるいは抗原特異性または結合モードを有する免疫実體、エピトープ、または免疫実體結合体、それらを含むまたは関連するクラスターまたはポリペプチド、を包含するポリペプチド、および薬学的に許容可能なアジュバント、キャリア、希釈剤、賦形剤等を含む薬学的組成物を提供する。アジュバントはポリヌクレオチドであり得る。ポリヌクレオチドはジヌレオチドを含むことができる。アジュバントは、ポリヌクレオチドによりコードされ得る。アジュバントはサイトカインであり得る。

#### 【0145】

さらなる実施形態では、本発明は、本発明で分類またはクラスター化された免疫実體、エピトープ、抗原特異性、結合モード、または免疫実體結合物（例えば、抗原）を含むポリペプチドをコードする核酸を含む本明細書中に記載する核酸のいずれかを含む薬学的組成物に関する。かかる組成物は、薬学的に許容可能なアジュバント、キャリア、希釈剤、賦形剤等を含むことができる。

#### 【0146】

さらなる実施形態では、本発明は、本発明で分類またはクラスター化された免疫実體、エピトープ、または免疫実體結合体の少なくとも1つに特異的に結合するか、同一のクラスターに属する抗原特異性、結合モードを有する、単離されおよび/または精製された抗体、抗原結合断片または他の免疫実體（例えば、B細胞受容体、B細胞受容体の断片、T細胞受容体、T細胞受容体の断片、キメラ抗原受容体（CAR）、またはこれらのいずれ

10

20

30

40

50

かまたは複数を含む細胞)に関する。他の実施形態では、本発明は、本発明で分類またはクラスター化された免疫実体、エピトープ、の少なくとも1つに特異的に結合するか、同一のクラスターに属する抗原特異性、結合モードを有する、または任意の他の適切なエピトープを含むペプチド-MHCタンパク質複合体に特異的に結合する単離されおよび/または精製された抗体または他の免疫実体に関する。いずれかの実施形態からの抗体はモノクローナル抗体またはポリクローナル抗体であり得る。これらの組成物は、薬学的に許容可能なアジュバント、キャリア、希釈剤、賦形剤等を含むことができる。

#### 【0147】

さらなる実施形態では、本発明は、本発明で分類またはクラスター化された免疫実体、エピトープ、または免疫実体結合体の少なくとも1つに特異的に相互作用するか、同一のクラスターに属する抗原特異性、結合モードを有するT細胞受容体(TCR)および/またはB細胞受容体(BCR)、それらの断片、またはその結合ドメインを含む単離されたタンパク質分子、またはTCRおよび/またはBCRのレパトア、キメラ抗原受容体(CAR)、またはこれらのいずれかまたは複数を含む細胞(例えば、キメラ抗原受容体(CAR)を含む遺伝子改変T細胞(CAR-T細胞ともいう)等)または他の免疫実体に関する。他の実施形態では、本発明は、本発明で分類またはクラスター化されたエピトープまたは任意の他の適切なエピトープを含むペプチド-MHCタンパク質複合体に特異的に結合する単離されおよび/または精製された抗体または他の免疫実体に関する。これらの組成物は、薬学的に許容可能なアジュバント、キャリア、希釈剤、賦形剤等を含むことができる。

#### 【0148】

さらなる局面では、本発明は、本発明の方法で生成されたクラスターに基づき、前記免疫実体の保有者を既知の疾患または障害あるいは生体の状態と関連付ける工程を包含する、疾患または障害あるいは生体の状態の同定法を提供する。あるいは別の局面では、本発明は、本発明の方法で生成されたクラスターを一つまたは複数用いて、該クラスターの保有者の疾患または障害あるいは生体の状態を評価する工程を含む疾患または障害あるいは生体の状態の同定法を提供する。ここで採用され得る任意の特徴は本明細書の<結合モードクラスター化技術>に記載される任意の特徴またはその組み合わせ、あるいはそれらの技術で同定、分類またはクラスター化されたものでありうる。ここで、上記評価は、評価は、前記複数のクラスターの存在量の順位、複数のクラスターの存在比に基づく分析、一定数のB細胞を調べ、その中に興味あるBCRと類似のもの/クラスターがあるかどうかという定量による分析などから選択される少なくとも1つの指標を用いてなされうるが、これらに限定されない。さらに別の実施形態では、上記評価は、前記クラスター以外の指標(例えば、疾患関連遺伝子、疾患関連遺伝子の多型、疾患関連遺伝子の発現プロファイル、エピジェネティクス解析、TCRおよびBCRのクラスターの組合せなどを挙げることができる。)も用いてなされる。本発明を用いることで、例えば、具体的には免疫系で重要な疾患特異的遺伝子(HLA allele等)、疾患関連遺伝子多型や遺伝子発現プロファイル(RNA-seq等)、エピジェネティクス解析(メチル化解析等)と組み合わせることができる。

#### 【0149】

1つの実施形態では、本発明が同定し得る疾患または障害あるいは生体の状態の同定は、前記疾患または障害あるいは生体の状態の診断、予後、薬力学、予測、代替法の決定、患者層の特定、安全性の評価、毒性の評価、およびこれらのモニタリングなどでありうる。

#### 【0150】

別の局面において、本発明は本発明で同定または分類されたエピトープ、免疫実体結合体あるいは精製したクラスターを一つまたは複数用いて、疾患または障害あるいは生体の状態の指標となるバイオマーカーの評価を行う工程を含む、該バイオマーカーの評価のための方法を提供する。あるいは本発明は本発明で同定または分類されたエピトープ、あるいは精製したクラスターを一つまたは複数用いて、疾患または障害あるいは生体の状態と

10

20

30

40

50

関連付け、バイオマーカーを決定する工程を含む、該バイオマーカーの同定のための方法を提供する。ここで、バイオマーカーの同定法については以下のような手法を用いることができる。例えば、シーケンサーで読んだB細胞レパトアの興味あるクラスターの存在、大きさ、占有率等をマーカーとして同定し、またそれらを利用することができる。

【0151】

さらなる実施形態では、本発明は、本発明で分類またはクラスター化された免疫実體、エピトープ、または免疫実體結合体の少なくとも1つに特異的に相互作用するか、同一のクラスターに属する抗原特異性、結合モードを有するポリペプチドをコードする構築物を含む本明細書中に記載する組換え構築物を発現する宿主細胞に関する。宿主細胞は樹状細胞、マクロファージ、腫瘍細胞、腫瘍由来細胞、細菌、真菌、原生動物等であり得る。この実施形態はまたこのような宿主細胞、および薬学的に許容可能なアジュバント、キャリア、希釈剤、賦形剤等を含む薬学的組成物を提供する。

10

【0152】

別の局面では、本発明は、本発明に基づいて同定された免疫実體、エピトープ、免疫実體結合体またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有する抗原または免疫実體結合物を含む、前記生体情報の同定のための組成物を提供する。あるいは、本発明は、本発明に基づいて同定された免疫実體、エピトープ、免疫実體結合体またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有する免疫実體、エピトープ、抗原などの免疫実體結合物などを含む、疾患または障害あるいは生体の状態を診断するための組成物を提供する。ここで採用され得る任意の特徴は本明細書の<結合モードクラスター化技術>に記載される任意の特徴またはその組み合わせ、あるいはそれらの技術で同定、分類またはクラスター化されたものでありうる。

20

【0153】

別の局面では、本発明は、本発明に基づいて同定されたエピトープまたは免疫実體結合体に対する免疫実體を標的とする物質を含む、疾患または障害あるいは生体の状態を診断するための組成物を提供する。あるいは、本発明は、本発明に基づいて同定された免疫実體、エピトープ、免疫実體結合体またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有する免疫実體、エピトープ、抗原などの免疫実體結合物を含む、疾患または障害あるいは生体の状態を診断するための組成物を提供する。ここで採用され得る任意の特徴は本明細書の<結合モードクラスター化技術>に記載される任意の特徴またはその組み合わせ、あるいはそれらの技術で同定、分類またはクラスター化されたものでありうる。したがって、免疫実體としては、例えば、抗体、抗体の抗原結合断片、T細胞受容体、T細胞受容体の断片、B細胞受容体、B細胞受容体の断片、キメラ抗原受容体(CAR)、これらのいずれかまたは複数を含む細胞(例えば、キメラ抗原受容体(CAR)を含むT細胞)などを挙げるすることができる。

30

【0154】

さらに別の局面では、本発明は、本発明に基づいて同定された免疫実體、エピトープ、免疫実體結合体またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有する免疫実體を含む疾患または障害あるいは生体の状態を治療または予防するための組成物を提供する。ここで採用され得る任意の特徴は本明細書の<結合モードクラスター化技術>に記載される任意の特徴またはその組み合わせ、あるいはそれらの技術で同定、分類またはクラスター化されたものでありうる。また、使用され得る免疫実體は、抗体、抗原結合断片、キメラ抗原受容体(CAR)、キメラ抗原受容体(CAR)を含むT細胞などを挙げるができるがこれらに限定されない。

40

【0155】

別の局面において、本発明は、本発明に基づいて同定された免疫実體、エピトープ、免疫実體結合体またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有する免疫実體を標的とする物質を含む、疾患または障害あるいは生体の状態を予防または治療するための組成物を提供する。ここで採用され得る任意の特徴は本明細書の<結合モードクラスター化技術>に記載される任意の特徴またはその組み合わせ、ある

50

いはそれらの技術で同定、分類またはクラスター化されたものでありうる。使用され得る物質としては、ペプチド、ポリペプチド、タンパク質、核酸、糖、低分子、高分子、金属イオンこれらの複合体を挙げることができるがこれらに限定されない。

【0156】

別の局面では、本発明は、本発明に基づいて同定された免疫実体、エピトープ、免疫実体結合体またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有する免疫実体結合物（例えば、抗原）を含む、疾患または障害あるいは生体の状態を治療または予防するための組成物を提供する。ここで採用され得る任意の特徴は本明細書の〈結合モードクラスター化技術〉に記載される任意の特徴またはその組み合わせ、あるいはそれらの技術で同定、分類またはクラスター化されたものでありうる。

10

【0157】

さらなる実施形態では、本発明は、本発明に基づいて同定された免疫実体、エピトープ、免疫実体結合体またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有する免疫実体結合物（例えば、抗原）またはポリペプチド、上記および本明細書中に記載の組成物、上記および本明細書中に記載のT細胞または宿主細胞のような少なくとも1つの構成成分を含むワクチンまたは免疫治療用組成物に関する。

【0158】

本発明はまた、診断方法または治療方法に関する。この方法は、本明細書中に開示するものを含むワクチンまたは免疫治療用組成物のような薬学的組成物を動物に投与するステップを含むことができる。投与は、例えば、経皮、結節内、結節周囲、経口、静脈内、皮内、筋内、腹腔内、粘膜、エアロゾル吸入、滴注等のよう送達様式を含むことができる。この方法は、標的細胞の状態を示す特徴を決定するためにアッセイする工程をさらに含むことができる。上記方法は、第1のアッセイ工程、および第2のアッセイ工程をさらに含むことができ、ここで第1のアッセイ工程は治療薬等の投与工程前に行われ、該第2のアッセイ工程は上記治療薬等の投与工程後に行われる。この場合、第1のアッセイ工程で決定される特徴を、第2のアッセイ工程で決定される特徴と比較する工程であって、それにより結果を得る、比較する工程をさらに含むことができる。結果は、例えば、免疫応答の徴候、標的細胞数の減少、標的細胞を含む腫瘍の質量またはサイズの低下、細胞内寄生生物感染標的細胞の数または濃度の低減等であり得、本発明に基づいて同定された免疫実体、エピトープまたはそれを含む、あるいは抗原特異性または結合モードに基づいて、判定を行うことができる。

20

30

【0159】

本発明は、本発明に基づいて同定された免疫実体、エピトープ、免疫実体結合体またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有するクラスター、このエピトープを含む免疫実体結合物（例えば、抗原）またはポリペプチドを、受動/養子免疫治療薬を作製する方法に関する。この方法は、本明細書中の他の箇所に記載するもののようなT細胞または宿主細胞を、薬学的に許容可能なアジュバント、キャリア、希釈剤、賦形剤等と組み合わせることを含むことができる。賦形剤としては、緩衝剤、結合剤、爆破剤、希釈剤、香味料、潤滑剤などを含むことができる。

【0160】

1つの局面において、本発明は、本発明に基づいて同定された免疫実体、エピトープ、免疫実体結合体またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有する免疫実体結合物（例えば、抗原）またはポリペプチド等を用いて、障害、疾患または生体状態を診断する方法に関する。上記方法は、被験体組織を、上記および本明細書中の他の箇所に記載するもののいずれかを含む、例えばT細胞、宿主細胞、抗体、タンパク質を含む少なくとも1つの構成成分と接触させること、および上記組織または該構成成分の特徴に基づいて疾患を診断することを含むことができる。接触工程は、例えば *in vivo* または *in vitro* で行われ得る。本発明は、分類したエピトープを同定する工程をさらに包含する。このような同定する工程には、その構造の決定が含まれ、このほか、例えば、アミノ酸配列の決定、三次元構造の同定、その他構造上の同定、

40

50

生物学的機能の同定などを含むがこれらに限定されない。

【0161】

さらなる実施形態では、本発明はワクチンを作製する方法に関する。この方法は、本明細書中の他の箇所に記載するもののいずれかを含むエピトープ、免疫実体結合体、組成物、構築物、T細胞、宿主細胞を含めた少なくとも1つの構成成分を、薬学的に許容可能なアジュバント、キャリア、希釈剤、賦形剤等と組み合わせることを含むことができる。別の実施形態では、本発明は、本発明のクラスタリングおよび分類法ならびにそれにより同定されたエピトープ、免疫実体または免疫実体結合物、あるいは同定された抗原特異性、結合モードを有するエピトープ、免疫実体または免疫実体結合物を用いて、ワクチンの評価または改善を行うことができ、同定されたエピトープまたは免疫実体結合物、同定された結合モードまたは抗原特異性またはそれを含む免疫実体、エピトープまたは免疫実体結合物、あるいはクラスターそれ自体を用いてバイオマーカーの評価および/または作成あるいは改善をおこなうことができる。ここで、「改善」は、クラスタリングにより、抗体価を上げたいクラスターを同定するなどして、ワクチン接種時の中和抗体産生をより適切に評価することができ、通常の実験と並行して行うことで、ワクチン性能改善のための手法を提供することを意味する。バイオマーカーの「評価」としては、例えば、まずそれ自身バイオマーカーとなり得るようなクラスター（例えば、疾患の状態と相関するクラスター）を同定し、より単純な実験（例えば、ELISA結合アッセイ等）を用いて実施することができる。）が適切に期待するクラスターの変化をフォローできているか調べる方法を例として挙げるができる。この場合、クラスターそれ自身がマーカーとしての機能を果たしているということが前提であるが、同様に（クラスターの情報を反映するように）作製も行うこともできる。

10

20

【0162】

本発明はまた、本発明に基づいて同定された免疫実体、エピトープ、免疫実体結合物またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有するまたは対する免疫実体を含む、疾患または障害あるいは生体の状態を予防または治療するためのワクチンを評価するための組成物を提供する。これらの評価は、例えば、インフルエンザウイルスの例があり、これらを応用することができる。別の局面において、本発明は、本発明に基づいて同定された免疫実体、エピトープ、免疫実体結合物またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有する免疫実体結合物（例えば、抗原）またはポリペプチド等を用いて、疾患を治療または予防する方法に関する。この方法は、本明細書中の他の箇所に記載するようなワクチンまたは免疫治療用組成物を動物に投与することを含む動物の治療方法を、例えば放射線療法、化学療法、生化学療法、手術を含む少なくとも1つの治療様式と組み合わせることを含むことができる。

30

【0163】

本発明はまた、本発明に基づいて同定された免疫実体、エピトープ、免疫実体結合物またはそれを含む、あるいは同一クラスターに属する抗原特異性または結合モードを有する、本発明で分類またはクラスター化されたエピトープ、このエピトープを含むクラスター、このエピトープを含む、同一クラスターに属する抗原特異性または結合モードを有する免疫実体結合物（例えば、抗原）またはポリペプチド等を含むワクチンまたは免疫治療用生成物に関する。さらに他の実施形態は、本明細書中の他の箇所に記載するポリペプチドをコードする単離されたポリヌクレオチドに関する。他の実施形態は、これらのポリヌクレオチドを含むワクチンまたは免疫治療用生成物に関する。ポリヌクレオチドは、DNA、RNA等であり得る。

40

【0164】

1つの実施形態では、本発明はまた、送達（デリバリー）デバイス、および本明細書中の他の箇所に記述した実施形態のいずれかを含むキットに関する。送達デバイスは、カテーテル、シリンジ、内部または外部ポンプ、リザーバ、吸入器、マイクロインジェクター、パッチ、および送達の任意の経路に適した任意の他の同様のデバイスであり得る。上述のように、送達デバイスに加えて、キットはまた、本明細書中に開示する実施形態のいず

50

れかを含むことができる。例えば、キットは、単離されたエピトープ、ポリペプチド、クラスター、核酸、免疫実体結合物（例えば、抗原）、上述のいずれかを含む薬学的組成物、抗体、T細胞、T細胞受容体、エピトープ-MHC複合体、ワクチン、免疫治療薬等を含むことができるが、これらに限定されない。キットはまた、使用のための詳細な説明書および任意の他の同様の品目のような構成物を含むことができる。

【0165】

結合モードや抗原特異性が同一である、免疫実体、エピトープ、免疫実体結合物および/またはそれらのクラスターをワクチンまたは薬学的組成物に含めるための特に望ましい戦略は、2000年4月28日に出願された「EPITOPE SYNCHRONIZATION IN ANTIGEN PRESENTING CELLS」という表題の米国特許出願第09/560,465号に開示されている。

10

【0166】

本発明で使用され得るワクチンは、本発明で分類、同定またはクラスター化したエピトープ、免疫実体結合物、あるいは同定された抗原特異性または結合モードを有するエピトープまたは免疫実体結合物を提示させるのに有効な濃度で当該エピトープまたは免疫実体結合物（例えば、抗原）を含有する。好ましくは、本発明のワクチンは、任意に1つまたは複数の免疫性エピトープと組み合わせて、複数の本発明のエピトープあるいはそのクラスターを含むことができる。本発明のワクチン製剤は、標的に対してエピトープを提示させるのに十分な濃度でペプチドおよび/または核酸を含有する。本発明の製剤は好ましくは、約1 $\mu$ g~1mg/(ワクチン調製物100 $\mu$ l)の総濃度でエピトープまたはそれを含有する。ペプチドワクチンおよび/または核酸ワクチンに関する従来の投与量および投薬を本発明とともに使用することができ、かかる投薬レジメンは当該分野で十分に理解されている。一実施形態では、成人に関する一回の投与量はかかる組成物約1~約5000 $\mu$ lであることが好適であり、一回または複数回で、例えば1週間、2週間、1ヶ月、またはそれ以上に分けた2回、3回、4回またはそれ以上の投与量で投与される。本発明のワクチンは、遺伝的に宿主中でエピトープを発現するように操作されたウイルス、細菌または原生動物のような組換え生物を含むことができる。

20

【0167】

本発明のワクチン、組成物、方法は、ワクチンの性能を増強するために、製剤にアジュバントを配合することができる。具体的には、エピトープの送達および取り込みを増強するように設計することができる。本発明により意図されるアジュバントは、当業者に既知であり、例えばGM-CSF、G-CSF、IL-2、IL-12、BCG、破傷風トキソイド、オステオポンチン、およびETA-1が挙げられる。

30

【0168】

本発明のワクチン等は、任意の適切な手法で投与することができる。本発明のワクチンは、当該技術分野で公知の標準的なワクチン送達プロトコルと一致した様式で患者に投与される。エピトープ送達方法としては、注射、滴注、または吸入による送達を含む、経皮、結節内、結節周辺、経口、静脈内、皮内、筋内、腹腔内、および粘膜投与が挙げられるが、これらに限定されない。CTL応答を誘発するためのワクチン送達の特有用な方法は、2002年1月17日に発行されたオーストラリア特許第739189号、1999年9月1日に提出された米国特許出願第09/380,534号、および2001年2月2日に提出されたその一部同時継続の米国特許出願第09/776,232号に開示されており、これらは本明細書において参考として援用される。

40

【0169】

1つの実施形態では、本発明はまた、本発明で分類、同定またはクラスター化した免疫実体、エピトープまたは免疫実体結合物、あるいは同定された抗原特異性または結合モードを有する免疫実体、エピトープまたは免疫実体結合物を提示させるのに有効な濃度で、免疫実体、エピトープまたはそれを含有する免疫実体結合物（例えば、抗原）に対して特異的に結合するタンパク質、抗体、これらが発現し得る細胞、特異的なB細胞およびT細胞等を含みうる。これらの試薬は、免疫グロブリン、すなわちその生成方法が当該分野で周知

50

であるポリクローナル血清またはモノクローナル抗体の形態をとる。ペプチド - MHC 分子複合体に関する特異性を有する mAb の生成は当該技術分野で公知である (Aharoni et al. Nature 351:147-150, 1991等)。一般的な構築および使用は、「T CELL RECEPTORS AND THEIR USE IN THERAPEUTIC AND DIAGNOSTIC METHODS」という表題の米国特許第 5, 830, 755号でも取り扱われている。

【0170】

1つの実施形態では、本発明で分類、同定またはクラスター化した結合モード、抗原特異性、免疫実体、エピトープまたは免疫実体結合物を提示させるのに有効な濃度で結合モードまたは抗原特異性を有する免疫実体、エピトープまたは免疫実体結合物 (例えば、抗原) のいずれかを、結合モード、抗原特異性、免疫実体エピトープまたは免疫実体結合物に関連した病原状態の診断 (イメージング、または他の検出)、モニタリング、および治療で使用するために、酵素、放射化学物質、蛍光タグ、および毒素と結合させることができる。したがって、毒素結合体は、腫瘍細胞を死滅させるのに投与することができ、放射標識は、結合モード、抗原特異性、免疫実体、エピトープまたは免疫実体結合物に関する陽性腫瘍のイメージングを容易にすることができ、酵素結合体は、癌を診断し生検組織におけるエピトープ発現を確認するために、ELISA 様アッセイで使用することができる。さらなる実施形態では、上述に記載するような T 細胞は、結合モード、抗原特異性またはエピトープおよび/またはサイトカインによる刺激により達成される増殖後に、養子免疫療法として患者に投与することができる。

【0171】

別の実施形態では、本発明は、本発明で分類、同定またはクラスター化した結合モードまたは抗原特異性を有するエピトープまたは免疫実体結合物と MHC との複合体または、結合モードもしくは抗原特異性を有するエピトープまたは免疫実体結合物としてのペプチド - MHC 複合体を提供する。特に好適な実施形態では、複合体は、米国特許第 5, 635, 363号 (四量体)、または米国特許第 6, 015, 884号 (Ig - 二量体) に記載されるもののような可溶性の多量体タンパク質であり得る。かかる試薬は、特定の T 細胞応答を検出およびモニタリングする際に、ならびにかかる T 細胞を精製する際に有用である。

【0172】

別の実施形態では、本発明で分類、同定またはクラスター化した結合モードまたは抗原特異性を有する免疫実体、エピトープまたは免疫実体結合物を用いて、機能アッセイを行い、免疫性の内因性レベル、免疫学的刺激 (例えば、ワクチン) に対する応答を評価し、疾患と治療の進路による免疫状態をモニタリングすることができる。免疫性の内因性レベルを測定する場合を除いて、これらのアッセイのいずれも、対処される問題の性質に応じて、*in vivo* でも *in vitro* でも、免疫の予備工程を前提とし得る。かかる免疫は、本発明の様々な実施形態を用いて、あるいは同様の免疫性を誘起することができる他の形態の免疫原を用いて実施することができる。同族 TCR の発現を検出することができる PCR および四量体 / Ig - 二量体型解析を除いて、これらのアッセイは概して、特定の機能活性を検出するために、上述のような本発明の様々な実施形態を好適に使用することができる *in vitro* 抗原性刺激の工程から利益を得る (高細胞溶解性応答はときには直接検出され得る)。最終的に、細胞溶解活性の検出は、同一のクラスターに属する結合モードもしくは抗原特異性を有する物質またはエピトープ提示標的細胞を必要とし、それは本発明の様々な実施形態を用いて生成することができる。任意の特定の工程に関して選択される特定の実施形態は、対処されるべき問題、使用の容易性、コスト等に依存するが、任意の特定組の状況に関する別の実施形態を上回る一実施形態の利点は当業者に明らかである。

【0173】

このような機能アッセイでは、本発明の結合モードもしくは抗原特異性に関連するか、または免疫実体、エピトープ、免疫実体結合物、または MHC 分子とのその複合体の形態で、活性化工程もしくは読み取り工程、またはその両方で用いることができる。当該分野

で公知のT細胞機能の多くのアッセイ(詳細な手順は、Current Protocols in Immunology 1999 John Wiley & Sons Inc., N.Yのような標準的な免疫学的参照文献に見出すことができる)のうち、細胞プールの応答を測定するアッセイと個々の細胞の応答を測定するアッセイの2つのカテゴリーを実施することができる。前者は、応答強度の全体的な測定が可能なのに対して、後者は、応答細胞の相対的頻度を決定し得る。全体的な応答を測定するアッセイの例は、細胞傷害性アッセイ、ELISA、およびサイトカイン分泌を検出する増殖アッセイである。個々の細胞(またはそれらに由来する小クローン)の応答を測定するアッセイとしては、限界希釈解析(LDA)、ELISPOT、未分泌サイトカインのフローサイトメトリー検出(米国特許第5,445,939号、米国特許第5,656,446号および同第5,843,689号に記載されており、それらのための試薬は、商品名「FAST IMMUNE」でBecton, Dickinson & Companyで販売されている)、および上述のように、かつ上記に引用するように四量体またはIg-二量体により特異的TCRの検出が挙げられる(Yee, C. et al. Current Opinion in Immunology, 13:141-146, 2001にも参照)。

10

**【0174】**

本発明はキットとして提供されることができる。本明細書において「キット」とは、通常2つ以上の区画に分けて、提供されるべき部分(例えば、検査薬、診断薬、治療薬、抗体、標識、説明書など)が提供されるユニットをいう。安定性等のため、混合されて提供されるべきでなく、使用直前に混合して使用することが好ましいような組成物の提供を目的とするときに、このキットの形態は好ましい。そのようなキットは、好ましくは、提供される部分(例えば、検査薬、診断薬、治療薬)をどのように使用するか、あるいは、試薬をどのように処理すべきかを記載する指示書または説明書を備えていることが有利である。本明細書においてキットが試薬キットとして使用される場合、キットには、通常、検査薬、診断薬、治療薬、抗体等の使い方などを記載した指示書などが含まれる。

20

**【0175】**

このように、本発明のさらなる局面では、本発明はキットに関するものであり、このキットは、(a)本発明の医薬組成物を溶液形状または凍結乾燥形状で包含する容器と、(b)選択的に、該凍結乾燥製剤用の希釈剤または再構成液を包含する第2の容器と、(c)選択的に、(i)該溶液の使用または(ii)該凍結乾燥製剤の再構成および/または使用に関する説明書とを有する。該キットは、1もしくはそれ以上の(iii)緩衝剤、(iv)希釈剤、(v)フィルター、(vi)針、または(v)シリンジをさらに有する。該容器は、好ましくは瓶、バイアル瓶、シリンジ、または試験管であり、多用途容器でよい。該医薬組成物は、好ましくは乾燥凍結される。

30

**【0176】**

本発明のキットは、好ましくは、本発明の乾燥凍結製剤およびその再構成および/または使用に関する説明書を、適切な容器内に有する。適切な容器として含まれるのは、例えば、瓶、バイアル瓶(例えばデュアルチャンババイアル)、シリンジ(デュアルチャンバシリンジなど)、および試験管である。該容器は、ガラスまたはプラスチックのような様々な材料から形成することができる。好ましくは、該キットおよび/または容器は、該容器上にある、あるいは該容器に伴う、再構成および/または使用の方法を示す説明書を包含する。例えば、そのラベルは、該乾燥凍結製剤を再構成して上記のペプチド濃度にするという説明を示すことができる。該ラベルは、さらに、該製剤が皮下注射に有用であるもしくは皮下注射のためのものであるという説明を示すことができる。

40

**【0177】**

該製剤の容器は、繰り返し投与(例えば2~6回の投与)に使うことができる多用途バイアル瓶でもよい。該キットは、さらに、適切な希釈剤(例えば重曹溶液)を有する第2の容器を有することができる。

**【0178】**

該希釈剤と該凍結乾燥製剤を混合して作られる再構成された製剤の最終ペプチド濃度は、好ましくは少なくとも0.15mg/mL/ペプチド(=75µg、0.5mlの場合)であり、好ましくは3

50

mg/mL/ペプチド(=1500 µg、0.5mlの場合)以下であるがこれらに限定されない。該キットは、さらに、商業的観点およびユーザーの観点から見て望ましいその他の材料(その他の緩衝剤、希釈剤、フィルター、針、シリンジ、およびパッケージに挿入される使用説明書を含む)を含むことができる。

【0179】

本発明のキットは、他の構成要素(例えば他の化合物またはこれら他の化合物の医薬組成物)と共に、もしくはそれらなしに、本発明の医薬組成物の製剤を包含する単一の容器を有すること、または各構成要素によって別の容器を有することができる。

【0180】

好ましくは、本発明のキットは、第2の化合物(アジュバント(例えばGM-CSF)、化学療法薬剤、天然生成物、ホルモンまたは拮抗薬、他の医薬など)またはその医薬組成物の併投与との組み合わせとして使用するためにパッケージされた本発明の処方を含む。該キットの構成要素は、予め複合体として作られたもの、もしくは、患者に投与するまで各構成要素が異なる別々の容器に入ったものが可能である。該キットの構成要素は、1もしくはそれ以上の液体溶液として提供することができ、好ましくは水溶液であり、より好ましくは滅菌水溶液である。該キットの構成要素は、固体として提供することも可能であり、好ましくは別の異なる容器にて提供される適切な溶剤をそれに加えて液体に変換することができる。

10

【0181】

療法キットの容器としては、バイアル、試験管、フラスコ、瓶、シリンジ、もしくは固体または液体を密封する他の任意の手段が可能である。通常、複数の構成要素がある場合、別々に投薬できるように、該キットは第2のバイアルまたはその他の容器を包含する。該キットは、薬学的に許容される液体用の別の容器も包含することができる。好ましくは、治療キットは、該キットの構成要素である本発明の薬剤を投与することを可能にする器具(例えば1もしくはそれ以上の針、シリンジ、点眼器、ピペットなど)を包含する。

20

【0182】

本発明の医薬組成物は、経口(経腸)、経鼻腔、経眼、皮下、皮内、筋内、静脈内、または経皮のような任意の許容される経路によって該ペプチドを投与するのに適したものである。好ましくは、該投与は皮下投与であり、最も好ましくは皮内投与である。投与は注入ポンプによって行うことができる。

30

【0183】

本明細書において「指示書」は、本発明を使用する方法を医師または他の使用者に対する説明を記載したものである。この指示書は、本発明の検出方法、診断薬の使い方、または医薬などを投与することを指示する文言が記載されている。また、指示書には、投与部位として、経口、食道への投与(例えば、注射などによる)することを指示する文言が記載されていてもよい。この指示書は、本発明が実施される国の監督官庁(例えば、日本であれば厚生労働省、米国であれば食品医薬品局(FDA)など)が規定した様式に従って作成され、その監督官庁により承認を受けた旨が明記される。指示書は、いわゆる添付文書(package insert)であり、通常は紙媒体で提供されるが、それに限定されず、例えば、電子媒体(例えば、インターネットで提供されるホームページ、電子メール)のような形態でも提供され得る。

40

【0184】

本明細書において「または」は、文章中に列挙されている事項の「少なくとも1つ以上」を採用できるときに使用される。「もしくは」も同様である。本明細書において「2つの値」の「範囲内」と明記した場合、その範囲には2つの値自体も含む。

【0185】

(一般技術)

本明細書において用いられる分子生物学的手法、生化学的手法、微生物学的手法、バイオインフォマティクスは、当該分野において公知であり、周知でありまたは慣用される任意のものが使用され得る。

50

## 【0186】

本明細書において引用された、科学文献、特許、特許出願などの参考文献は、その全体が、各々具体的に記載されたのと同じ程度に本明細書において参考として援用される。

## 【0187】

以上、本発明を、理解の容易のために好ましい実施形態を示して説明してきた。以下に、実施例に基づいて本発明を説明するが、上述の説明および以下の実施例は、例示の目的のみに提供され、本発明を限定する目的で提供したのではない。従って、本発明の範囲は、本明細書に具体的に記載された実施形態にも実施例にも限定されず、特許請求の範囲によってのみ限定される。

## 【実施例】

10

## 【0188】

以下に実施例を記載する。必要な場合、以下の実施例において、全ての実験は、大阪大学倫理委員会で承認されたガイドラインに従って実施した。試薬類は具体的には実施例中に記載した製品を使用した。他メーカー (Sigma-Aldrich、和光純薬、ナカライ、R & D Systems、USCN Life Science INC等) の同等品でも代用可能である。

## 【0189】

(実施例1: 抗体の抗原特異的クラスタリング)

抗体-抗原複合体の結晶構造から抗原エピトープ特異性に基づいて抗体配列をクラスタリングした。

20

## 【0190】

(データセット)

S A b D a b (<http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/Welcome.php>, 2017年3月16日版) から抗原抗体複合体結晶構造リストをダウンロードした。閾値を3.5として抗体と接触している抗原の重原子を探した。抗原の残基長さが3以上のものを残し、さらに、抗原抗体の配列の重複をCD-HITを用いて除いた。FASTA (デフォルト設定を使用) を用いて抗原配列のローカルアライメントを行い、一致部分を抜き出し、各抗原配列上の抗体配列との接触残基が65%以上保存されていて、5残基以上同一抗原残基に接触しているもので、かつ接触残基のRMSDが5.0Å未満のものを同一エピトープを認識するとした。最後に、抗体の重鎖と軽鎖配列を繋げてCD-HITを用いて90%以上の配列相同性があるものは削除した。全体として23,220の対が得られ、そのうち465が正、残りが誤のデータセットとなった。ここから、ランダムに80%を学習セットに使い、残りの20%をテストセットに用いた。(表1、表2)。

30

## 【0191】

【表 1 - 1】

1ahwBA	2x71AB	4fp8IM	4xvsHL
1bj1HL	2xraHL	4fqjHL	4xwgHL
1bogBA	2yc1AB	4g6fHL	4y5yAB
1bvkBA	3b2uCD	4g6jHL	4yb1HL
1deeFE	3baeHL	4hixHL	4yd1BC
1e4xHL	3bkyHL	4hxxAB	4ydvHL
1fd1HL	3bn9DC	4hs6HL	4zffHL
1fnsHL	3eyfBA	4hs8HL	4zptAB
1frgHL	3g5vBA	4i9wED	5acoGJ
1ggiHL	3gbmHL	4ij3CB	5c7xMN
1himLH	3gi8HL	4jb9HL	5cbaCD
1hysDC	3gjfHL	4jdtHL	5cd5CD
1jpsHL	3grwHL	4jg0HL	5d1qBA
1kc5HL	3h42HL	4jo1IM	5d1qCD
1n64HL	3hb3CD	4jpvHL	5d1xBA
1nmaHL	3ixtAB	4jznIP	5eocHL
1orsBA	3j5mDC	4jzoCF	5f3bAB
1otsCD	3jabHL	4k4mHL	5f3hAB
1qljIM	3jwoHL	4ki5CD	5f9oHL
1s5hBA	3l5wHL	4kv5HL	5fecAB
1tziBA	3lexHL	4lspHL	5fgcEB
1uj3BA	3ma9HL	4lsuHL	5gjsHL
1v7mHL	3mlrHL	4lvhHI	5h37KL
1xgpBA	3mlwHL	4m48HL	5hhvHL
1yy9DC	3mnwBA	4mhhJK	5i8cAB
2aepHL	3o41AB	4py8IJ	5i9qBC
2b0sHL	3pjsBA	4r2gJI	5igxHL
2b2xIM	3q3gDC	4ravAB	5kaqQR
2dtgAB	3s88HL	4rdqIH	5ke1CD
2dtgCD	3sm5HL	4rfoHL	5kveHL
2eizBA	3so3CB	4slqHL	5sx4JI
2i91HG	3u7yHL	4slrHL	5te4HL
2ipuHL	3ujiHL	4slsHL	5tljDC
2j4wHL	3ujjHL	4tvpDE	5tlkHG

10

20

30

【 0 1 9 2 】

【表 1 - 2】

2nyyDC	3ulvDC	4u6hAB	5tpwHL
2os1HL	3wxvHL	4ut6HL	5u3kHL
2otuBA	4d9qED	4utaHL	5u3oHL
2r4rHL	4edwHL	4utbHL	5wt9HL
2uziHL	4edxBA	4xmphHL	
2vdrHL	4ffvDC	4xnmHL	
2virBA	4ffyhHL	4xnyHL	

40

表 1. 学習セット

【 0 1 9 3 】

50

【表 2】

1g9mHL	3mlyHL	4k9eHL	4yf1HL
1kcsHL	3rvvDC	4m62IM	5b8cBA
1mlcBA	3se8HL	4o58HL	5c7kEF
1ncaHL	3uyrHL	4onfHL	5drzBA
1tjgHL	3ztjGH	4pljDC	5esvHL
1tpxBC	4a18HL	4r8wHL	5eszAB
1yqvHL	4dgiHL	4rfnHL	5f96HL
2fd6HL	4fqihHL	4tsaHL	5f9wBC
2q8aHL	4hhaBA	4tvpHL	5iq9HL
3be1HL	4hlzGH	4u0rBC	5kvfHL
3ifoAB	4hpoHL	4uu9AB	5u3jHL
3ifpAB	4iofCD	4xnzHL	5u3mHL
3l5xHL	4j6rHL	4y5vGH	
3lh2IM	4jpwHL	4ydiHL	

10

表 2. テストセット

【 0 1 9 4 】

(特徴量抽出)

それぞれに抗体の対に対して、重鎖と軽鎖それぞれ3つのCDRと4つのFR(フレームワーク)領域を同定した。上記の特徴量を各領域毎に得た。

- ・ B L O S U M 6 2 に基づく配列相同性スコア
- ・ アミノ酸配列の長さの差
- ・ アライメントされた残基の個数。

20

【 0 1 9 5 】

(機械学習とハイパーパラメータ最適化)

pythonの機械学習用ライブラリである、sklearnのGridSearchCVを用いて、ランダムフォレストのtreeの数と各treeのleafの数を、5回交差検証の結果平均MCC(Matthews correlation coefficient)が最高となるようグリッド探索を行った。最高のMCCを与えるハイパーパラメータは(treeの数、treeのleafの数)=(9, 60)となった。

30

【 0 1 9 6 】

最適なハイパーパラメータを用いて、再度学習セット全体を用いて学習を行い、その結果をテストセットに適用した。その結果、MCCは0.85となった。

【 0 1 9 7 】

(結論)

抗体-抗原複合体の結晶構造から抗原エピトープ特異性に基づいて抗体配列をクラスタリングすることができることが判明した。

【 0 1 9 8 】

(実施例 2 : TCRの抗原特異的クラスタリング)

本実施例では、TCR-pMHC結合情報のみからTCRのクラスタリングを行い、クラスターが異なる結合特異性(モード)を反映していることを示す。

40

【 0 1 9 9 】

(データセット)

TCR配列データを下記の3つのデータベースから取得した(2017年10月2日データ取得)。

- ・ ATLAS: <https://zlab.umassmed.edu/atlas/web/help.php>
- ・ VDJdb: <https://vdjdb.cdr3.net/>
- ・ McPAS-TCR: <http://friedmanlab.weizmann.ac.il/McPAS-TCR/>

50

これらのうち、ヒトとマウスに由来するTCRのみを抽出し、重複したエントリ(V遺伝子、J遺伝子、CDR3配列が同じもの)を削除し、結果として10727のユニークなTCRベータ鎖のデータセット(それぞれpMHCの情報があるもの)を作成した。

#### 【0200】

(特徴量抽出)

機械学習に用いる特徴量は以下のものを用いた。

(1) V-、J-遺伝子に基づく特徴量

ヒトおよびマウスのTRAV、TRBV、TRAJ、TRBJ遺伝子のアミノ酸配列情報をIMGT(<http://www.imgt.org/vquest/refseqh.html>)より取得し、各遺伝子ファミリーのグローバルアライメントを行い、多重配列アライメントを得た。IMGTの定義に基づくCDR1、CDR2、FR1、FR2、FR3、FR4を抜き出した。81番目から86番目(IMGTの定義に基づく)のアミノ酸で定義される、CDR2.5領域(Dash, P., Fiore-Gartland, A. J., Hertz, T., Wang, G. C., Sharma, S., Souquette, A., ... Thomas, P. G. (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature*. <https://doi.org/10.1038/nature22383>)も抜き出した。

10

#### 【0201】

(2) CDR3に基づく特徴量

CDR3領域(IMGTの定義に基づいて105番目から117番目のアミノ酸)の配列を抜き出した。データベースに記載してあるものは全長からではなくそのまま使用した。さらに削られたCDR3(CDR3の最初の3つのアミノ酸と最後の2つのアミノ酸を削ったもの)を得た。

20

#### 【0202】

(3) 物理量の特徴量

・全電荷

個々の領域(CDR1, CDR2, CDR2.5, CDR3, FR1, FR2, FR3, FR4)ごとに、各領域に含まれる側鎖のpH7.5における電荷を足し合わせた。

・CDR3領域の疎水性

Kyte & Doolittleの疎水性指数(index of hydrophobicity)を計算した。

30

(4) 対の比較に基づく特徴量

上記のTCRごとの特徴量に加え、全てのTCRの対ごとの特徴量も計算した。

#### 【0203】

配列間距離: TCR-AとTCR-Bに対して、上記で生成した多重配列アライメントを元に、BLOSUM62代入行列(Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences*, 89(22), 10915-10919. <https://doi.org/10.1073/pnas.89.22.10915>)に基づき、個々の領域(CDR1, CDR2, CDR2.5, CDR3, FR1, FR2, FR3, FR4)ごとの配列間距離を計算した。

#### 【0204】

非配列特徴量: プーリアン型特徴量として各領域(CDR1, CDR2, CDR2.5, CDR3, FR1, FR2, FR3, FR4)の電荷が同じ符号(+または-)を持つかどうか。および、CDR3領域の疎水性の差の絶対値を考慮した。

40

#### 【0205】

(機械学習アルゴリズムとハイパーパラメータ最適化)

(1) 機械学習予測モデル

オープンソースのLightGBM gradient boostingフレームワーク(<https://github.com/Microsoft/LightGBM>)を用いて、対のTCRが同じエピトープに結合するかどうかを学習させた。この時、以下のハイパーパラメータを最適化した: treeの数、treeごとのleafの数、学習レート、正誤の相対的重み。

50

## 【0206】

## (2) クラスタリングアルゴリズム

予測結果に基づいて階層的クラスタリング法によってクラスタリングを行う。この時、固定された予測値の閾値を設定するが閾値もハイパーパラメータの最適化の際に最適化される。

## 【0207】

## (3) 学習・テストセット分割と評価

データセットから生成した対の情報は、結合するエピトープに基づき、80%のエピトープが学習セットに、20%がテストセットに割り振る。この割り振りを10回繰り返す。

10

## 【0208】

スコアリング：様々なハイパーパラメータセットに基づいた学習によって作成された予測モデルをテストセットに適用し、評価を行う。評価はMCCスコア、変形ランド指数、一様性スコア(homogeneity score)を用いて行なった。10回の学習・予測・クラスタリング・評価を繰り返した。一様性スコアが0.9より大きいものうち、平均MCCスコアが最高のものを選択した。

## 【0209】

## (結果)

ハイパーパラメータは、(treeの数、treeごとのleafの数、学習レート、正誤の相対的重み) = (50, 30, 0.1, 1.6) が最適であった。また階層的クラスタリングの閾値は0.6に設定された。(図2)

20

最適化されたモデルをTCR-pMHC結晶構造が知られているEBV(Epstein-Barr Virus)由来のエピトープを認識するTCRに対して適用した。その結果、同じpMHCでも異なる位置を認識するTCRは別々のクラスターに分かれており、クラスタリング結果は結合モードを反映していることがわかった(図3)。

## 【0210】

## (実施例3: HIV由来抗原特異的TCRの抗原ペプチドと提示MHCの予測)

本実施例では、抗原未知TCRと抗原既知TCR配列とのクラスタリングを行い、抗原既知TCR配列の情報から、抗原未知TCR配列の抗原を予測できることを示す。

## 【0211】

## (データセット)

国立感染症研究所で得られた14例のヒト検体由来のHIV由来ペプチドA特異的TCR配列115本と、7例のヒト検体由来のHIV由来ペプチドB特異的TCR配列82本と、実施例2で用いたデータセットに含まれる236本の重複のないHIV抗原(7種類)特異的TCRを用いた。

30

## 【0212】

## (予測)

データセットに実施例2の最適なハイパーパラメータを用いて得られた機械学習モデルを適用した。階層的クラスタリングの閾値も同様(0.6)である。クラスタリング結果を図4に示す。ペプチドA特異的配列とB特異的配列が分離していることがわかる。また、実施例2で用いたデータセットが含まれるクラスターはクラスター内のTCR配列が認識するpMHC情報から、認識する抗原が予測された。

40

## 【0213】

## (備考)

## 略語(Abbreviations)

TCR: T cell receptor

ML: Machine learning

CDR: Complementarity-determining region(s)

MCC: Matthews correlation coefficient

BLOSUM: BLOcksSubstitution Matrix a.a. amino acid

50

(注記)

以上のように、本発明の好ましい実施形態を用いて本発明を例示してきたが、本発明は、特許請求の範囲によってのみその範囲が解釈されるべきであることが理解される。本明細書において引用した特許、特許出願および文献は、その内容自体が具体的に本明細書に記載されているのと同様にその内容が本明細書に対する参考として援用されるべきであることが理解される。

【産業上の利用可能性】

【0214】

免疫関連の疾患について、精確度が高い、臨床応用が可能である。

【配列表フリーテキスト】

【0215】

配列番号1：EBV由来エピトープ(FLRGRAYGL)

【要約】 (修正有)

【課題】免疫関連の疾患について、精確度が高い、臨床応用可能な免疫実体を分類する手法を提供する。

【解決手段】機能をあらかじめ規定しないで、通常、免疫実体結合物(抗原やエピトープ)を個々に別の「機能(=例えば、抗原Aに対する特異性を持つかどうか)」として取り扱われる抗原特異性というものに一般性があると仮定し、その類似度を評価することで、免疫実体を分類する。

【選択図】図1A

10

20

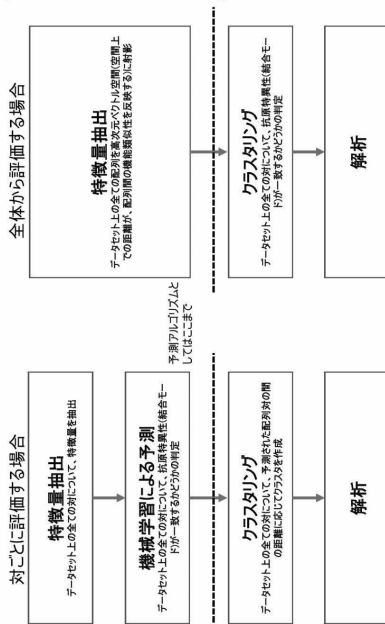
【図1A】

データセットの種類に応じた射影

1. あらかじめ各配列間の距離的かついては、各配列間の抗原特異性の意味での距離が分かっている場合、(例えば)左の方法で各配列間の抗原特異性の意味での距離が分かっている場合、(e.g. ニューラルネットワークを用いて)その配列間の距離を再現する。多次元空間上のベクトルとして配列を射影する。特徴量は各配列から任意のもの抽出し、ニューラルネットワークの入力とする。(予測)上記で学習されたモデルに配列から特徴量を抽出したものを入力すると、予測結果が得られる。

2. あらかじめ各配列の抗原特異性の正誤が分かっている場合、(e.g. ニューラルネットワークを用いて)抗原特異性の正の配列対は近く、負の配列対は遠くなるような多次元空間への射影を行う。この時、ニューラルネットワークの入力は各配列から抽出した任意の特徴量ベクトルであり、各配列間の多次元空間における距離に応じて学習を行い、最適なモデルを構成する。(予測)上記で学習されたモデルに配列から特徴量を抽出したものを入力すると、予測結果が得られる。

Fig.1A



【図1B】

テストセットに対するBCRクラスティングの結果。ノードが各PDB構造を表し、エッジが予測の結果同じ抗原特異性を持つと判定されたもの。

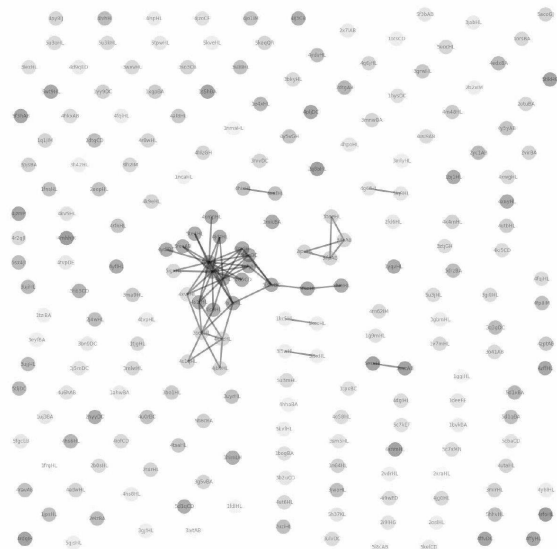


Fig.1B

【 図 2 】

20のエピトープを認識するTCRのクラスタリング結果

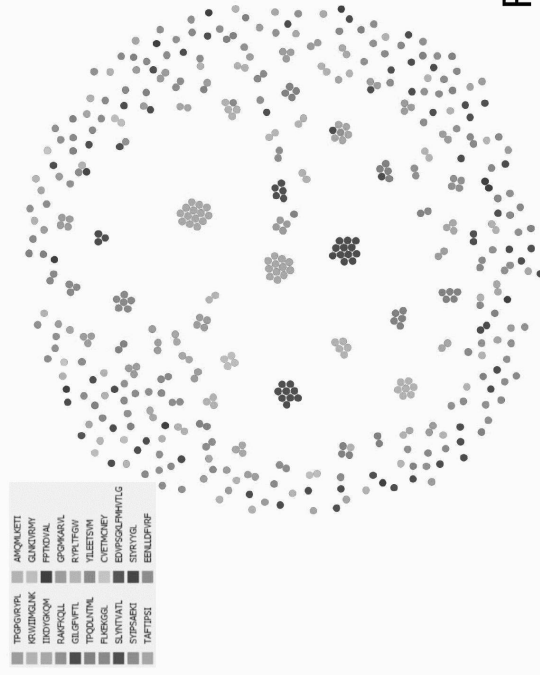


Fig.2

【 図 4 】

2種類のHIV由来ペプチド特異的TCRとデータベース上のTCRとのクラスタリング結果

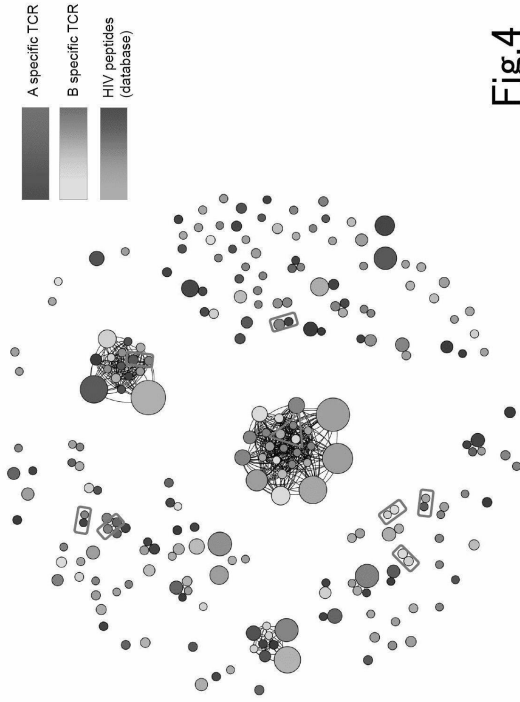


Fig.4

【 図 3 】

EBV由来エピトープ(FLRGRAYGL)特異的TCRのクラスタリング結果(右)と対応する結晶構造(左:PDBより得た構造を重ね書きしたもの)

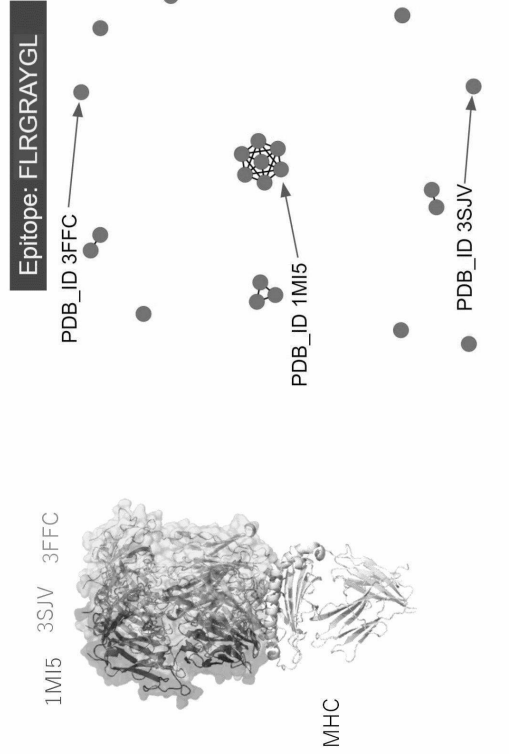


Fig.3

【 図 5 】

システム1000

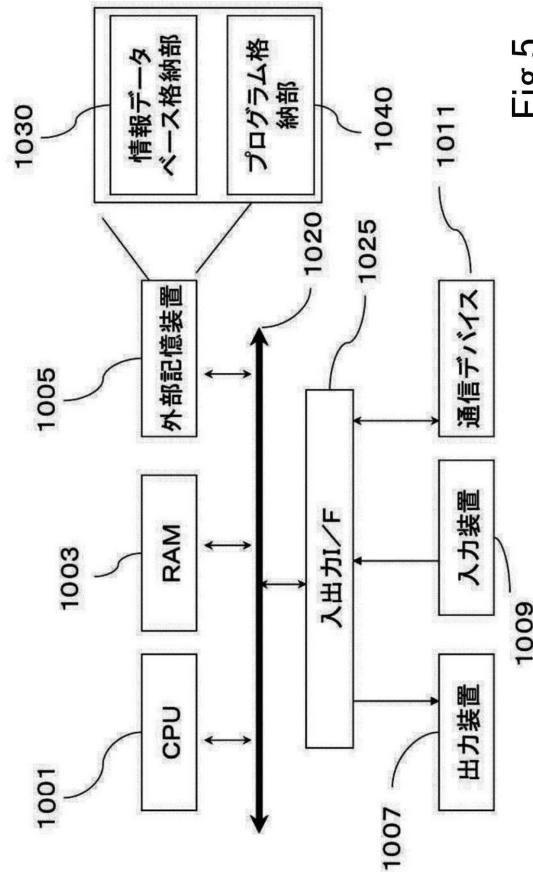


Fig.5



## フロントページの続き

- (74)代理人 230113332  
弁護士 山本 健策
- (72)発明者 山下 和男  
大阪府大阪市中央区東高麗橋2番5-905号 グラン・コープ東高麗橋 KOTAIバイオテク  
ノロジーズ株式会社内
- (72)発明者 ビヨー ジョエル  
大阪府大阪市中央区東高麗橋2番5-905号 グラン・コープ東高麗橋 KOTAIバイオテク  
ノロジーズ株式会社内
- (72)発明者 サックス ニコラ クロード ポール  
大阪府大阪市中央区東高麗橋2番5-905号 グラン・コープ東高麗橋 KOTAIバイオテク  
ノロジーズ株式会社内

審査官 渡邊 加寿磨

- (56)参考文献 国際公開第2018/052131(WO, A1)  
特表2005-512044(JP, A)  
米国特許出願公開第2014/0032186(US, A1)  
米国特許出願公開第2004/0014069(US, A1)  
米国特許出願公開第2017/0076036(US, A1)  
特表2002-502064(JP, A)  
特開2006-236151(JP, A)  
特開2007-272627(JP, A)  
米国特許出願公開第2017/0228523(US, A1)  
神崎洋治, " ロボットをビジネスに生かすAI技術(2):人工能はどうやって「学ぶ」のか  
教師あり学習、教師なし学習、強化学習(1/2)", [online], アイティメディア株式会社,  
2016年11月30日, [検索日:2018.7.5], URL, <https://web.archive.org/web/20161130100425/http://www.atmarkit.co.jp/ait/articles/1610/01/news002.html>

## (58)調査した分野(Int.Cl., DB名)

G16B	5/00	-	99/00
G16H	10/00	-	80/00
A61K	39/00		
C12Q	1/02		
G01N	33/53		

专利名称(译)	有效聚类免疫实体		
公开(公告)号	<a href="#">JP6500144B1</a>	公开(公告)日	2019-04-10
申请号	JP2018061901	申请日	2018-03-28
[标]发明人	山下和男		
发明人	山下 和男 ビヨー ジョエル サククス ニコラ クロード ポール		
IPC分类号	G16B40/00 C12Q1/02 G01N33/53 A61K39/00		
FI分类号	G06F19/24.ZNA C12Q1/02 G01N33/53 A61K39/00.G G16B40/00		
F-TERM分类号	4B063/QA05 4B063/QQ79 4B063/QR48 4B063/QR71 4C085/AA03 4C085/BA01 4C085/EE01		
代理人(译)	夏木森下 饭田TakashiSatoshi 石川大介 山本健作		
审查员(译)	渡邊 加寿磨		
其他公开文献	JP2019160261A		
外部链接	<a href="#">Espacenet</a>		

摘要(译)

提供一种对免疫学实体进行分组的方法，该方法可以在免疫相关疾病领域中在临床上高精度应用。例如，通常假定每种免疫学实体偶联物（抗原和表位）是否存在对抗原A）的特异性，并对它们的相似性进行评估以对免疫学实体进行分组。图1A

(19) 日本国特許庁(JP)	(12) 特許公報(B1)	(11) 特許番号 特許第6500144号 (P6500144)
(45) 発行日 平成31年4月10日(2019.4.10)	(24) 登録日 平成31年3月22日(2019.3.22)	
(51) Int. Cl.	F I	
G 1 6 B 40/00 (2019.01)	G O 6 F 19/24 Z N A	
C 1 2 Q 1/02 (2006.01)	C 1 2 Q 1/02	
G O 1 N 33/53 (2006.01)	G O 1 N 33/53	
A 6 1 K 39/00 (2006.01)	A 6 1 K 39/00 G	
請求項の数 21 (全 52 頁)		
(21) 出願番号 特願2018-61901(P2018-61901)	(73) 特許権者 518092159	
(22) 出願日 平成30年3月28日(2018.3.28)	K O T A I バイオテクノロジー株式会社	
(62) 分割の表示 特願2018-49440(P2018-49440)の分割	大阪府大阪市中央区東高麗橋2番5-905号 グラン・コープ東高麗橋	
原出願日 平成30年3月16日(2018.3.16)	(74) 代理人 100078282	
審査請求日 平成30年3月28日(2018.3.28)	弁理士 山本 秀策	
早期審査対象出願	(74) 代理人 100113413	
	弁理士 森下 夏樹	
	(74) 代理人 100118371	
	弁理士 ▲駒▼谷 剛志	
	(74) 代理人 100181674	
	弁理士 飯田 貴敏	
	(74) 代理人 100181641	
	弁理士 石川 大輔	
(54) 【発明の名称】 免疫実体の効率的クラスタリング	最終頁に続く	