

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7668742号  
(P7668742)

(45)発行日 令和7年4月25日(2025.4.25)

(24)登録日 令和7年4月17日(2025.4.17)

|                         |                             |          |                         |         |
|-------------------------|-----------------------------|----------|-------------------------|---------|
| (51)国際特許分類              |                             | F I      |                         |         |
| C 1 2 Q                 | 1/6869(2018.01)             | C 1 2 Q  | 1/6869                  | Z Z N A |
| C 1 2 N                 | 5/10 (2006.01)              | C 1 2 N  | 5/10                    |         |
| C 1 2 N                 | 5/0783(2010.01)             | C 1 2 N  | 5/0783                  |         |
| C 1 2 N                 | 15/12 (2006.01)             | C 1 2 N  | 15/12                   |         |
| C 0 7 K                 | 14/725(2006.01)             | C 0 7 K  | 14/725                  |         |
| 請求項の数 13 (全146頁) 最終頁に続く |                             |          |                         |         |
| (21)出願番号                | 特願2021-552549(P2021-552549) | (73)特許権者 | 518213307               |         |
| (86)(22)出願日             | 令和2年3月6日(2020.3.6)          |          | グリットストーン バイオ インコーポレ     |         |
| (65)公表番号                | 特表2022-524328(P2022-524328  |          | イテッド                    |         |
|                         | A)                          |          | アメリカ合衆国 9 4 6 0 8 カリフォル |         |
| (43)公表日                 | 令和4年5月2日(2022.5.2)          |          | ニア州 エメリービル ホートン ストリ     |         |
| (86)国際出願番号              | PCT/US2020/021508           |          | ート 5 9 5 9 スイート 3 0 0   |         |
| (87)国際公開番号              | WO2020/181240               | (74)代理人  | 100102978               |         |
| (87)国際公開日               | 令和2年9月10日(2020.9.10)        |          | 弁理士 清水 初志               |         |
| 審査請求日                   | 令和5年3月3日(2023.3.3)          | (74)代理人  | 100160923               |         |
| (31)優先権主張番号             | 62/814,801                  |          | 弁理士 山口 裕孝               |         |
| (32)優先日                 | 平成31年3月6日(2019.3.6)         | (74)代理人  | 100119507               |         |
| (33)優先権主張国・地域又は機関       |                             |          | 弁理士 刑部 俊                |         |
|                         | 米国(US)                      | (74)代理人  | 100142929               |         |
| (31)優先権主張番号             | 62/826,822                  |          | 弁理士 井上 隆一               |         |
| (32)優先日                 | 平成31年3月29日(2019.3.29)       | (74)代理人  | 100148699               |         |
| 最終頁に続く                  |                             | 最終頁に続く   |                         |         |

(54)【発明の名称】 M H C クラス I I モデルによる新生抗原の特定

(57)【特許請求の範囲】

【請求項1】

対象の1つ以上の細胞に由来する、1つ以上のクラスII MHCアレルによって前記細胞の表面上に提示される可能性の高い少なくとも1つの新生抗原に対して抗原特異的な1つ以上のT細胞を特定するための方法であって、

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ステップ；

前記新生抗原のそれぞれの前記ペプチド配列を、対応する数値ベクトルにコード化するステップであって、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、前記ペプチド配列における前記アミノ酸の位置のセットとに関する情報を含む、ステップ；

前記新生抗原のセットについての提示尤度のセットを生成するために、コンピュータのプロセッサを使用して、機械学習させた提示モデルに前記数値ベクトルを入力するステップであって、前記セットの中の各提示尤度が、対応する新生抗原が前記1つ以上のクラスII MHCアレルによって前記対象の前記細胞の前記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、

複数の試料の各試料についての、前記試料中に存在するものとして特定されたクラスII MHCアレルのセットの中の少なくとも1つのクラスII 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、及び

前記試料のそれぞれについての、前記ペプチドを構成する複数のアミノ酸と、前記ペプチドにおける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

を含む訓練データセットに少なくとも基づいて特定される、複数のパラメータと、  
入力として受け取られる前記数値ベクトルと、前記数値ベクトル及び前記パラメータに  
基づいて出力として生成される前記提示尤度との間の関係を表す、関数と  
を含み、ここで、前記機械学習させた提示モデルは、ペプチド配列の結合コアまたは結合  
アンカーのアミノ酸の位置を特定するように設計される、ステップ；

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

前記サブセットの中の少なくとも1つの前記新生抗原に対して抗原特異的である1つ以  
上のT細胞を特定するステップ

を含む、前記方法。

【請求項2】

(a) 前記少なくとも1つの新生抗原が、対象の1つ以上の腫瘍細胞に由来し、1つ以  
上のクラスII MHCアレルによって前記腫瘍細胞の表面上に提示される可能性の高い、  
新生抗原であり、かつ、

(b) 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前記ステッ  
プが、

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つを取得するステ  
ップであって、前記ヌクレオチドシーケンシングデータが、前記腫瘍細胞由来の前記ヌ  
クレオチドシーケンシングデータと前記正常細胞由来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た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、ステップ

を含む、

請求項1に記載の方法。

【請求項3】

前記機械学習させた提示モデルに前記数値ベクトルを入力する前記ステップが、

前記ペプチド配列の前記特定の位置の前記特定のアミノ酸に基づいて、クラスII MHC  
アレルが前記新生抗原を提示するかどうかを示す、前記1つ以上のクラスII MHCア  
レルの各クラスII MHCアレル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新生抗原の前記ペプチド配列に適用すること  
を含み、任意で、

(a) 前記機械学習させた提示モデルに前記数値ベクトルを入力する前記ステップが、  
対応するクラスII MHCアレルが前記対応する新生抗原を提示する尤度を示す、各ク  
ラスII MHCアレルについての対応するアレルごとの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；及び

前記新生抗原の前記提示尤度を生成するために、前記アレルごとの尤度を組み合わせる  
こと

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つ以上のク  
ラスII 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) 前記機械学習させた提示モデルに前記数値ベクトルを入力する前記ステップが、  
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であ  
って、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  
以上のクラスII MHCアレル間で干渉するものとしてモデル化する、前記変換すること  
をさらに含む、

請求項1または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

10

20

30

40

50

特定され、前記方法が、

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ステップをさらに含む、

(a) 前記1つ以上のクラスII MHCアレルの各クラスII 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ステップ；

対応するクラスII MHCアレルが前記対応する新生抗原を提示する尤度を示す、各クラスII MHCアレルについてのアレルごとの尤度を生成するために、各クラスII 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ステップ；及び

前記提示尤度を生成するために、前記アレルごとの尤度を組み合わせるステップをさらに含む、または

(b) 前記クラスII 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ステップ；及び

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ステップを任意でさらに含む、

請求項3に記載の方法。

【請求項5】

(a) 前記1つ以上のクラスII MHCアレルが、2つ以上の異なるクラスII MHCアレルを含む、および／または

(b) 前記少なくとも1つのクラスII MHCアレルが、2つ以上の異なるタイプのクラスII MHCアレルを含む、および／または

(c) 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、および／または

(d) 前記ペプチド配列をコード化する前記ステップが、ワンホットコード化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、および／または

(e) 前記複数の試料が、

(i) 単一のクラスII 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、

(ii) 複数のクラスII 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、

(iii) 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、

(iv) 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

(v) 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、および／または

(f) 前記訓練データセットが、

(i) 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

(ii) 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、および／または

(g) 前記提示尤度のセットが、少なくとも、前記対象における前記1つ以上のクラスII MHCアレルの発現レベルによって、さらに特定され、前記発現レベルがRNA-seqまたは質量分析により測定される、および／または

(h) 前記提示尤度のセットが、

(i) 前記新生抗原のセットの中の新生抗原と前記1つ以上のクラスII MHCアレルとの間の予測される親和性、及び

(ii) 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

10

20

30

40

50

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、および／または

(i) 前記提示尤度のセットが、

(i) 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

(ii) 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、および／または

(j) 前記選択された新生抗原のセットを選択する前記ステップが、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に比べて前記腫瘍細胞表面上に提示される尤度が高い新生抗原を選択することを含む、および／または

(k) 前記選択された新生抗原のセットを選択する前記ステップが、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に比べ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高い新生抗原を選択することを含む、および／または

(l) 前記選択された新生抗原のセットを選択する前記ステップが、前記提示モデルに基づいて、選択されない新生抗原に比べ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細胞に提示されることができる尤度が高い新生抗原を選択することを含み、任意で、前記APCは樹状細胞(DC)である、および／または

(m) 前記選択された新生抗原のセットを選択する前記ステップが、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に比べ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低い新生抗原を選択することを含む、および／または

(n) 前記選択された新生抗原のセットを選択する前記ステップが、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に比べ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低い新生抗原を選択することを含む、および／または

(o) 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、脾臓癌、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より選択される、および／または

(p) 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ステップをさらに含み、任意で、

前記個別化がんワクチンのための出力が、前記選択された新生抗原のセットをコードする少なくとも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1～4のいずれか一項に記載の方法。

#### 【請求項6】

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1～5のいずれか一項に記載の方法。

#### 【請求項7】

前記ニューラルネットワークモデルが、前記クラスII MHCアレルについての複数のネットワークモデルを含み、各ネットワークモデルが、前記クラスII MHCアレルのうちの対応するクラスII 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、

(a) 各ネットワークモデルが、1つ以上の畳み込みニューラルネットワークをさらに含み、前記1つ以上の畳み込みニューラルネットワークのそれぞれが、1つ以上の層に配置されかつ異なるサイズのフィルターを有する一連のノードを含み、前記1つ以上の畳み込みニューラルネットワークのそれぞれの前記フィルターが、前記ペプチド配列の結合コアまたは結合アンカーを含む各新生抗原のペプチド配列におけるアミノ酸の位置を特定するようなサイズである、および／または

(b) 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回の訓練反復において一緒に更新される、および／または

10

20

30

40

50

(c) 前記機械学習させた提示モデルが、1つ以上のノードの層を含むディープラーニングモデルである、

請求項6に記載の方法。

【請求項8】

(a) 前記1つ以上のT細胞を特定する前記ステップが、前記1つ以上のT細胞を拡大増殖させる条件下で前記1つ以上のT細胞を前記サブセットの中の前記新生抗原のうちの1つ以上と共培養することを含む、および／または

(b) 前記1つ以上のT細胞を特定する前記ステップが、前記1つ以上のT細胞を、前記サブセットの中の前記新生抗原のうちの1つ以上を含むMHCマルチマーと、前記T細胞と前記MHCマルチマーとの結合を可能とする条件下で接触させることを含む、  
請求項1～7のいずれか一項に記載の方法。

10

【請求項9】

前記1つ以上の特定されたT細胞の1つ以上のT細胞受容体(TCR)を特定するステップをさらに含む、請求項1～8のいずれか一項に記載の方法。

【請求項10】

前記1つ以上のT細胞受容体を特定する前記ステップが、前記1つ以上の特定されたT細胞の前記T細胞受容体配列をシーケンシングすることを含む、請求項9に記載の方法。

【請求項11】

前記1つ以上の特定されたT細胞受容体のうちの少なくとも1つを発現するように複数のT細胞を遺伝子操作するステップ；及び

20

前記複数のT細胞を拡大増殖させる条件下で前記複数のT細胞を培養するステップをさらに含む、請求項10に記載の方法。

【請求項12】

前記1つ以上の特定されたT細胞受容体のうちの少なくとも1つを発現するように前記複数のT細胞を遺伝子操作する前記ステップが、

前記1つ以上の特定されたT細胞の前記T細胞受容体配列を発現ベクターにクローニングすること；及び

前記複数のT細胞のそれぞれに前記発現ベクターをトランスフェクトすることを含む、請求項11に記載の方法。

【請求項13】

30

前記1つ以上の特定されたT細胞を拡大増殖させる条件下で前記1つ以上の特定されたT細胞を培養するステップをさらに含む、請求項1～12のいずれか一項に記載の方法。

【発明の詳細な説明】

【背景技術】

【0001】

腫瘍特異的な新生抗原に基づいた治療用ワクチン及びT細胞療法は、次世代の個別化がん免疫療法として極めて有望である<sup>1～3</sup>。非小細胞肺癌(NSCLC)及びメラノーマなどの高い遺伝子変異量を有するがんは、新生抗原を生じる可能性が比較的高いことから、かかる治療法の特に有望な標的である<sup>4, 5</sup>。初期の証拠により、新生抗原に基づいたワクチン接種がT細胞応答を誘発し<sup>6</sup>、新生抗原を標的としたT細胞療法が、選択された患者において腫瘍退縮を引き起こしうる<sup>7</sup>ことが示されている。

40

【0002】

詳細には、新生抗原ベースのワクチン接種及び新生抗原を標的としたT細胞療法で使用するためのMHCクラスIIにより提示される新生抗原の特定は、新生抗原応答性のTILの最大50%が、MHCクラスIIアレルによって提示される新生抗原に応答するCD4細胞を含むことから、有望な治療である。これらのCD4細胞は、抗腫瘍応答においてCD8細胞を助けることが示されており、場合によっては腫瘍細胞を直接攻撃することが示されている。がん治療におけるMHCクラスII提示新生抗原のこのような有望性にもかかわらず、MHCクラスII提示新生抗原の陽性適中率(PPV)は、CD8細胞によって認識されるMHCクラスI提示新生抗原のPPVよりも低い。

50

## 【0003】

MHCクラスII提示新生抗原のこれらの比較的低い提示予測結果は、MHCクラスII分子と比較したMHCクラスII分子の構造に一部拠るものと考えられる。具体的には、MHCクラスII分子は、MHCクラスII分子と比較してより開放したペプチド結合溝を有する傾向がある。この構造の違いにより、MHCクラスII分子がアミノ酸8～10個の長さのペプチドと結合する傾向があるのに対して、MHCクラスII分子はより様々な長さのペプチドと結合する（図14F）。MHCクラスII分子によって提示されるペプチドの長さのばらつきのため、MHCクラスII分子によって提示されるペプチドは、MHCクラスII分子によって提示されるペプチドと比較して予測がより難しくなり得る。

## 【0004】

したがって、MHCクラスII提示新生抗原及び新生抗原認識T細胞の特定は、腫瘍応答を評価し<sup>77, 110</sup>、腫瘍進化を調べ<sup>111</sup>、次世代の個別化療法を設計する<sup>112</sup>うえで中心的な課題となっている。現在の新生抗原特定法は、時間と手間がかかるか<sup>84, 96</sup>、または精度が充分とはいえない<sup>87, 91-93</sup>。さらに、新生抗原認識T細胞は、TILの主成分であり<sup>84, 96, 113, 114</sup>、がん患者の末梢血中を循環している<sup>107</sup>ことが近年、示されたが、新生抗原反応性T細胞を特定するための現在の方法は、（1）TIL<sup>97, 98</sup>または白血球除去療法<sup>107</sup>などの取得困難な臨床検体に頼ったものである、（2）非実用的に大きなペプチドのライブラリーをスクリーニングする必要がある、または（3）少数のMHCアレルについてのみ実用上、利用可能であり得るMHCアレルに頼ったものである、という制約条件のいくつかの組み合わせを有している。

## 【0005】

さらに、次世代のシーケンシング、RNA遺伝子発現、及び新生抗原ペプチドのMHC結合親和性の予測を用いた、変異に基づいた分析を取り入れた初期の方法が提案されている<sup>8</sup>。しかしながら、これらの提案されている方法では、遺伝子発現及びMHC結合以外の多くの段階（例えば、TAP輸送、プロテアソーム切断、MHC結合、ペプチド-MHC複合体の細胞表面への輸送、及び／またはTCRによるMHCの認識；エンドサイトーシスまたはオートファジー、細胞外またはリソソームプロテアーゼ（例えばカテプシン）による切断、及び／またはHLA-DMにより触媒されるHLA結合に対するCLIPペプチドとの競合）を含む<sup>9</sup>エピトープ生成プロセスの全体をモデル化することはできない。したがって、既存の方法は、陽性適中率（PPV）が低くなるという問題を有する傾向にある（図1A）。

## 【0006】

実際、複数のグループによって実施された、腫瘍細胞により提示されるペプチドの分析は、遺伝子発現及びMHC結合親和性を用いて提示されることが予測されたペプチドのうち5%未満しか腫瘍表面のMHC上に見られないことを示している<sup>10, 11</sup>（図1B）。結合予測とMHC提示との間のこのような低い相関は、変異の数単独に対してチェックポイント阻害剤反応について結合に制限された新生抗原の予測精度の向上が認められないことによってさらに指示されている<sup>12</sup>。これらの提示予測の失敗は、MHCクラスIIアレルによって提示される新生抗原の場合に、特に当てはまる。

## 【0007】

提示を予測するための既存の方法のこのような低い陽性適中率（PPV）は、新生抗原に基づいたワクチンの設計における、また、新生抗原に基づいたT細胞療法における問題を提示する。PPVの低い予測を用いてワクチンが設計される場合、大部分の患者で治療用新生抗原が投与される可能性は低くなり、複数の新生抗原が投与される患者はさらに少なくなるものと考えられる（提示されるペプチドのすべてが免疫原性であると仮定したとしても）。同様に、治療用のT細胞がPPVの低い予測に基づいて設計される場合、大部分の患者で腫瘍新生抗原に対する反応性を有するT細胞が投与される可能性が低くなり、予測後に下流の検査方法を用いて予測新生抗原を特定する時間及び物理的リソースのコストが不要に高くなり得る。したがって、現行の方法による新生抗原ワクチン接種及びT細胞療法は、腫瘍を有する対象の相当数において奏功する可能性は低い（図1C）。

## 【0008】

さらに、これまでのアプローチは、シス作用性の変異のみを用いて候補新生抗原を生成するものであり、複数の腫瘍タイプで生じ、多くの遺伝子で異常スプライシングにつながるスプライシング因子の変異<sup>13</sup>、及びプロテアーゼ切断部位を生じるかまたは除去する変異を含む、新生ORFのさらなるソースをほとんどの場合で考慮していなかった。

## 【0009】

最後に、腫瘍ゲノム及びトランスクリプトーム解析に対する標準的アプローチは、ライブラリ構築、エクソーム及びトランスクリプトームの捕捉、シーケンシング、またはデータ分析における最適に満たない条件のために、候補新生抗原を生ずる体細胞突然変異を見逃す可能性がある。同様に、標準的な腫瘍分析のアプローチでは、配列アーチファクトまたは生殖系列多型を新生抗原として誤って助長してしまう場合があり、それぞれワクチン容量の非効率的な利用または自己免疫のリスクにつながりうる。

## 【発明の概要】

## 【0010】

本明細書では、個別化がんワクチン用の、T細胞療法用の、またはその両方のためのMHCクラスIIアレルにより提示される新生抗原を特定及び選択するための最適化されたアプローチが開示される。第1に、次世代シーケンシング(NGS)を用いて新生抗原候補を特定するための最適化された腫瘍エクソーム及びトランスクリプトーム解析アプローチに対する取り組みを行う。これらの方法は、すべてのクラスのゲノム変化にわたって最も感度及び特異性の高い新生抗原候補が開発されるように、NGSによる腫瘍解析の標準的アプローチに立脚したものである。第2に、特異性の問題を克服し、ワクチン添加用に、及び／またはT細胞療法の標的として開発されるMHCクラスIIアレル提示新生抗原が抗腫瘍免疫をより誘発しやすくするために高PPVのMHCクラスIIアレル提示新生抗原選択に対する新規なアプローチが提供される。これらのアプローチには、実施形態に応じて、ペプチド-MHCクラスIIアレルマッピングを共にモデル化する訓練された統計学的回帰または非線形ディープラーニングMHCクラスIIモデル、ならびに異なる長さのペプチドにわたって統計学的効力を共有する、複数の長さのペプチドについてのMHCクラスIIアレルごとのモチーフが含まれる。特に非線形MHCクラスIIディープラーニングモデルは同じ細胞内の異なるMHCアレルを独立したものとして扱うように設計及び訓練することができるため、線形モデル同士が互いに干渉する線形モデルに伴う問題が解決される。最後に、MHCクラスIIアレル提示新生抗原に基づいた個別化ワクチンの設計及び製造に関する、また、T細胞療法用の個別化されたMHCクラスIIアレル提示新生抗原特異的なT細胞の製造におけるさらなる懸案事項が解決される。

## 【0011】

本明細書で開示されるモデルは、結合親和性で訓練された最新の予測ツール及びMSペプチドデータに基づいた初期の予測ツールの性能を最大で一桁上回る。MHCクラスIIアレルによるペプチドの提示をより高い信頼性で予測することにより、本モデルは、限られた量の患者の末梢血を使用し、患者当たりにつきスクリーニングするペプチドの数が少なく、MHCマルチマーに必ずしも頼らない、臨床的実用性のあるプロセスを用いた個別化療法のためのMHCクラスIIアレル提示新生抗原特異的なまたは腫瘍抗原特異的なT細胞のより時間効率かつ費用効率の高い特定を可能とする。しかしながら、別の実施形態では、本明細書に開示されるモデルを用いることで、MHCクラスIIアレル提示新生抗原または腫瘍抗原特異的なT細胞を特定するためにスクリーニングする必要がある、MHCマルチマーに結合したペプチドの数が少なくなることによって、MHCマルチマーを用いたMHCクラスIIアレル提示腫瘍抗原特異的なT細胞のより時間効率及び費用効率の高い特定が可能となる。

## 【0012】

TIL新生エピトープデータセット及び予測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MHCクラスIIモデルの予測性能は、MHCクラスIIアレル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MHCクラス

10

20

30

40

50

II アレル提示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、MHCクラスII アレル提示抗原標的化免疫療法のための実用的なインシリコのMHCクラスII アレル提示抗原の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]

対象の1つ以上の腫瘍細胞に由来する、1つ以上のクラスII MHC アレルによって前記腫瘍細胞の表面上に提示される可能性の高い少なくとも1つの新生抗原に対して抗原特異的な1つ以上のT細胞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つを取得するステップであって、前記ヌクレオチドシーケンシングデータが、前記腫瘍細胞由来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由来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各新生抗原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た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、ステップ；

前記新生抗原のそれぞれの前記ペプチド配列を、対応する数値ベクトルにコード化するステップであって、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、前記ペプチド配列における前記アミノ酸の位置のセットとに関する情報を含む、ステップ；

前記新生抗原のセットについての提示尤度のセットを生成するために、コンピュータのプロセッサを使用して、機械学習させた提示モデルに前記数値ベクトルを入力するステップであって、前記セットの中の各提示尤度が、対応する新生抗原が前記1つ以上のクラスII MHC アレルによって前記対象の前記腫瘍細胞の前記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、

複数の試料の各試料についての、前記試料中に存在するものとして特定されたクラスII MHC アレルのセットの中の少なくとも1つのクラスII 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、及び

前記試料のそれぞれについての、前記ペプチドを構成する複数のアミノ酸と、前記ペプチドにおける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

を含む訓練データセットに少なくとも基づいて特定される、複数のパラメータと、入力として受け取られる前記数値ベクトルと、前記数値ベクトル及び前記パラメータに基づいて出力として生成される前記提示尤度との間の関係を表す、関数とを含む、ステップ；

選択された新生抗原のセットを生成するために、前記提示尤度のセットに基づいて前記新生抗原のセットのサブセットを選択するステップ；

前記サブセットの中の少なくとも1つの前記新生抗原に対して抗原特異的である1つ以上のT細胞を特定するステップ；ならびに

前記1つ以上の特定されたT細胞を返すステップを含む、前記方法。

[本発明1002]

前記機械学習させた提示モデルに前記数値ベクトルを入力する前記ステップが、前記ペプチド配列の前記特定の位置の前記特定のアミノ酸に基づいて、クラスII MHC アレルが前記新生抗原を提示するかどうかを示す、前記1つ以上のクラスII MHC アレルの各クラスII MHC アレル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新生抗原の前記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]

前記機械学習させた提示モデルに前記数値ベクトルを入力する前記ステップが、対応するクラスII MHC アレルが前記対応する新生抗原を提示する尤度を示す、各ク

10

20

30

40

50

ラスII MHCアレルについての対応するアレルごとの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；及び

前記新生抗原の前記提示尤度を生成するために、前記アレルごとの尤度を組み合わせること

をさらに含む、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つ以上のクラスII 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]

前記機械学習させた提示モデルに前記数値ベクトルを入力する前記ステップが、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であって、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つ以上のクラスII MHCアレル間で干渉するものとしてモデル化する、前記変換することをさらに含む、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、前記方法が、

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ステップをさらに含む、本発明1002～1005のいずれかの方法。

[本発明1007]

前記1つ以上のクラスII MHCアレルの各クラスII 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ステップ；

対応するクラスII MHCアレルが前記対応する新生抗原を提示する尤度を示す、各クラスII MHCアレルについてのアレルごとの尤度を生成するために、各クラスII 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ステップ；及び

前記提示尤度を生成するために、前記アレルごとの尤度を組み合わせるステップをさらに含む、本発明1006の方法。

[本発明1008]

前記クラスII 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ステップ；及び

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ステップをさらに含む、本発明1006の方法。

[本発明1009]

前記1つ以上のクラスII MHCアレルが、2つ以上の異なるクラスII MHCアレルを含む、本発明1001～1008のいずれかの方法。

[本発明1010]

前記少なくとも1つのクラスII MHCアレルが、2つ以上の異なるタイプのクラスII 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前記ステップが、ワンホットコード化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) 単一のクラスII 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、

10

20

30

40

50

(b) 複数のクラスⅡ 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-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、本発明1001~1013のいずれかの方法。

10

[本発明1015]

前記提示尤度のセットが、少なくとも、前記対象における前記1つ以上のクラスⅡ M  
H C アレルの発現レベルによって、さらに特定され、前記発現レベルがRNA-seq また  
は質量分析により測定される、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の中の新生抗原と前記1つ以上のクラスⅡ 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、本発明1001~1015のい  
ずれかの方法。

20

[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。

30

[本発明1018]

前記選択された新生抗原のセットを選択する前記ステップが、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に比べて前記腫瘍細胞表面上に提示される尤度が  
高い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前記ステップが、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に比べ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高い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。

40

[本発明1020]

前記選択された新生抗原のセットを選択する前記ステップが、前記提示モデルに基づい  
て、選択されない新生抗原に比べ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細胞に提示されることができる尤度が高い新生抗原を選択することを含み、任  
意で、前記APCは樹状細胞 (DC) 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前記ステップが、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に比べて中枢性寛容または末梢性寛容によって阻  
害される尤度が低い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。

50

[本発明1022]

前記選択された新生抗原のセットを選択する前記ステップが、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に比べ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低い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より選択される、本発明1001～1022のいずれかの方法。

10

[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ステップをさらに含む、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のための出力が、前記選択された新生抗原のセットをコードする少なくとも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。

20

[本発明1027]

前記ニューラルネットワークモデルが、前記クラスII MHCアレルについての複数のネットワークモデルを含み、各ネットワークモデルが、前記クラスII MHCアレルのうちの対応するクラスII 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]

各ネットワークモデルが、1つ以上の畳み込みニューラルネットワークをさらに含み、前記1つ以上の畳み込みニューラルネットワークのそれぞれが、1つ以上の層に配置されかつ異なるサイズのフィルターを有する一連のノードを含み、前記1つ以上の畳み込みニューラルネットワークのそれぞれの前記フィルターが、前記ペプチド配列の結合コアまたは結合アンカーを含む各新生抗原のペプチド配列におけるアミノ酸の位置を特定するようなサイズである、本発明1027の方法。

30

[本発明1029]

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回の訓練反復において一緒に更新される、本発明1027または1028の方法。

[本発明1030]

前記機械学習させた提示モデルが、1つ以上のノードの層を含むディープラーニングモデルである、本発明1026～1029のいずれかの方法。

40

[本発明1031]

前記1つ以上のT細胞を特定する前記ステップが、前記1つ以上のT細胞を拡大増殖させる条件下で前記1つ以上のT細胞を前記サブセットの中の前記新生抗原のうちの1つ以上と共培養することを含む、本発明1001～1030のいずれかの方法。

[本発明1032]

前記1つ以上のT細胞を特定する前記ステップが、前記1つ以上のT細胞を、前記サブセットの中の前記新生抗原のうちの1つ以上を含むMHCマルチマーと、前記T細胞と前記MHCマルチマーとの結合を可能とする条件下で接触させることを含む、本発明1001～1031のいずれかの方法。

[本発明1033]

50

前記1つ以上の特定されたT細胞の1つ以上のT細胞受容体（TCR）を特定するステップをさらに含む、本発明1001～1032のいずれかの方法。

[本発明1034]

前記1つ以上のT細胞受容体を特定する前記ステップが、前記1つ以上の特定されたT細胞の前記T細胞受容体配列をシーケンシングすることを含む、本発明1033の方法。

[本発明1035]

本発明1001～1034のいずれかの前記サブセットの中の少なくとも1つの選択された新生抗原に対して抗原特異的な、単離されたT細胞。

[本発明1036]

前記1つ以上の特定されたT細胞受容体のうちの少なくとも1つを発現するように複数のT細胞を遺伝子操作するステップ；

前記複数のT細胞を拡大増殖させる条件下で前記複数のT細胞を培養するステップ；及び  
前記拡大増殖させたT細胞を前記対象に注入するステップ

をさらに含む、本発明1034の方法。

[本発明1037]

前記1つ以上の特定されたT細胞受容体のうちの少なくとも1つを発現するように前記複数のT細胞を遺伝子操作する前記ステップが、

前記1つ以上の特定されたT細胞の前記T細胞受容体配列を発現ベクターにクローニングすること；及び

前記複数のT細胞のそれぞれに前記発現ベクターをトランスフェクトすること  
を含む、本発明1036の方法。

[本発明1038]

前記1つ以上の特定されたT細胞を拡大増殖させる条件下で前記1つ以上の特定されたT細胞を培養するステップ；及び

前記拡大増殖させたT細胞を前記対象に注入するステップ  
をさらに含む、本発明1001～1037のいずれかの方法。

【図面の簡単な説明】

【0013】

本発明のこれらの特徴、態様、及び側面、ならびに他の特徴、態様、及び側面は、以下の説明文及び添付の図面に関してより深い理解が得られるであろう。

【0014】

【図1A】 新生抗原の特定に対する現在の臨床的アプローチを示す。

【0015】

【図1B】 予測された結合ペプチドのうち、腫瘍細胞上に提示されるものは5%未満であることを示す。

【0016】

【図1C】 新生抗原予測の特異性の問題の影響を示す。

【0017】

【図1D】 結合予測が、新生抗原の特定に充分ではないことを示す。

【0018】

【図1E】 ペプチド長の関数としてのMHC-I提示の確率を示す。

【0019】

【図1F】 Promega社のダイナミックレンジ標準から生成された、例示的なペプチドスペクトルを示す。配列番号1を開示する。

【0020】

【図1G】 特性の追加が、いかにモデルの陽性適中率を増大させるかを示す。

【0021】

【図2A】 一実施形態による、患者におけるペプチド提示の尤度を特定するための環境の概略である。

【0022】

10

20

30

40

50

【図 2 B】一実施形態による、提示情報を取得する方法を説明する。配列番号 2 0 を開示する。

【図 2 C】一実施形態による、提示情報を取得する方法を説明する。配列番号 3 ～ 8 を示される順序で開示する。

【0 0 2 3】

【図 3】一実施形態による、提示特定システムのコンピュータ論理構成要素を説明する、ハイレベルブロック図である。

【0 0 2 4】

【図 4】一実施形態による、訓練データの例示的なセットを説明する。「ペプチド配列」を配列番号 1 0 ～ 1 3 として、また、「C 側フランキング配列」を配列番号 1 5、2 1 ～ 2 2、及び 2 2 として、示される順序でそれぞれ開示する。

【0 0 2 5】

【図 5】M H C アレルに関連した例示的なネットワークモデルを説明する。

【0 0 2 6】

【図 6 A】一実施形態による、M H C アレルによって共有される例示的なネットワークモデル  $NN_H(\cdot)$  を示す。

【0 0 2 7】

【図 6 B】別の実施形態による、M H C アレルによって共有される例示的なネットワークモデル  $NN_H(\cdot)$  を示す。

【0 0 2 8】

【図 7】例示的なネットワークモデルを用いた、M H C アレルに関連したペプチドの提示尤度の生成を説明する。

【0 0 2 9】

【図 8】例示的なネットワークモデルを用いた、M H C アレルに関連したペプチドの提示尤度の生成を説明する。

【0 0 3 0】

【図 9】例示的なネットワークモデルを用いた、M H C アレルに関連したペプチドの提示尤度の生成を説明する。

【0 0 3 1】

【図 1 0】例示的なネットワークモデルを用いた、M H C アレルに関連したペプチドの提示尤度の生成を説明する。

【0 0 3 2】

【図 1 1】例示的なネットワークモデルを用いた、M H C アレルに関連したペプチドの提示尤度の生成を説明する。

【0 0 3 3】

【図 1 2】例示的なネットワークモデルを用いた、M H C アレルに関連したペプチドの提示尤度の生成を説明する。

【0 0 3 4】

【図 1 3 A】N S C L C 患者における変異負荷の標本度数分布を示す。

【0 0 3 5】

【図 1 3 B】一実施形態による、患者が最小変異負荷を満たすかどうかの選択基準に基づいて選択された患者に対してシミュレートしたワクチン中の提示新生抗原の数を示す。

【0 0 3 6】

【図 1 3 C】一実施形態による、提示モデルに基づいて特定された治療サブセットを含むワクチンに関連付けられた選択された患者と、従来技術のモデルによって特定された治療サブセットを含むワクチンに関連付けられた選択された患者との間のシミュレートしたワクチン中の提示新生抗原の数を比較したものである。

【0 0 3 7】

【図 1 3 D】H L A - A \* 0 2 : 0 1 についての単一のアレルごとの提示モデルに基づいて特定された治療サブセットを含むワクチンに関連付けられた選択された患者と、H L A

10

20

30

40

50

－A＊02：01についてのアレルごとの提示モデル及びHLA－B＊07：02についてのアレルごとの提示モデルの両方に基づいて特定された治療サブセットを含むワクチンに関連付けられた選択された患者との間のシミュレートしたワクチン中の提示新生抗原の数を比較したものである。一実施形態によれば、ワクチン容量は、 $v = 20$ 個のエピトープに設定される。

【0038】

【図13E】一実施形態による、変異負荷に基づいて選択された患者と、期待効用値スコアにより選択された患者との間でシミュレートしたワクチン中の提示新生抗原の数を比較したものである。

【0039】

【図14A】質量分析を用いたヒト腫瘍細胞及び腫瘍浸潤リンパ球（TIL）上のクラスII MHCアレルから溶出されたペプチドの長さのヒストグラムである。

【0040】

【図14B】2つの例示的なデータセットについてmRNA定量化と残基当たりの提示ペプチドとの依存関係を示す。

【0041】

【図14C】2つの例示的なデータセットを用いて訓練及び試験した例示的な提示モデルの性能結果を比較したものである。

【0042】

【図14D】HLAクラスII分子を含む合計73種の試料の各試料について質量分析を用いてシーケンシングしたペプチドの量を示すヒストグラムである。

【0043】

【図14E】特定のMHCクラスII分子のアレルが同定された試料の量を示すヒストグラムである。

【0044】

【図14F】合計で73種の試料中でMHCクラスII分子によって提示されるペプチドの割合を、一定範囲のペプチド長の各ペプチド長について示したヒストグラムである。

【0045】

【図14G】73種の試料中に存在する遺伝子について、遺伝子発現とMHCクラスII分子による遺伝子発現産物の提示の発生率との間の関係を示す線グラフである。

【0046】

【図14H】ペプチドの試験データセット内のペプチドがMHCクラスII分子によって提示される尤度を予想するうえでの入力の変化する同じモデルの性能を比較した線グラフである。

【0047】

【図14I】ペプチドの試験データセット内のペプチドがMHCクラスII分子によって提示される尤度を予測するうえでの3つの異なるモデルの性能を比較した線グラフである。

【0048】

【図14J】HLA－DRB（MHCクラスII遺伝子）によるペプチド提示を予測するように構成された、図14IのBi-LSTMモデルの例示的な実施形態を示す。

【0049】

【図14K】図14IのBi-LSTM、MLP、RNN、及び結合親和性モデルの完全適合率－再現率曲線を示す線グラフである。

【0050】

【図14L】2つの異なる基準を用いたベスト・イン・クラスの従来モデルと、2つの異なる入力を有する本明細書に開示される提示モデルの、ペプチドの試験データセット内のペプチドがMHCクラスII分子によって提示される尤度を予想するうえでの性能を比較した線グラフである。

【0051】

【図14M】HLAクラスII分子を含むヒト腫瘍（NSCLC、リンパ腫、及び卵巣癌

10

20

30

40

50

）及び細胞株（EBV）からなる合計230種の試料の各試料について0.1未満のq値で質量分析を用いてシーケンシングしたペプチドの量を示すヒストグラムである。

【0052】

【図14N】特定のMHCクラスII分子のアレルが同定された試料の量を示すヒストグラムである。

【0053】

【図14O】MHCクラスII分子に結合したペプチド及びMHCクラスII分子に結合したペプチドを示す。

【0054】

【図14P】MHCクラスII分子によるペプチド提示を予測するように構成された、図14Qのインセプションモデルのインセプションニューラルネットワークの例示的な実施形態を示す。

10

【0055】

【図14Q】ペプチドの試験データセット内のペプチドが、試験データセット内に存在するMHCクラスII分子の少なくとも1つによって提示される尤度を予測するうえでの「Bi-LSTM」及び「インセプション」提示モデルの性能を比較した線グラフである。

【0056】

【図15A】図15は、「MSモデル」、HLA-DRB1\*15:01及びHLA-DRB5\*01:01にわたって最も低いNetMHCIIpanパーセンタイルランクを採用する、「NetMHCIIpan rank」: NetMHCIIpan 3.1<sup>77</sup>、及びHLA-DRB1\*15:01及びHLA-DRB5\*01:01にわたってnM単位の最も強い親和性を採用する「NetMHCIIpan nM」: NetMHCIIpan 3.1の、HLA-DRB1\*15:01／HLA-DRB5\*01:01試験データセット内でのペプチドのランク付けにおける予測性能を比較したものである。

20

【図15B】図15Aの説明を参照。

【0057】

【図16】TCRをレシピエント細胞に導入するためのTCRコンストラクトの例示的な実施形態を示す。

【0058】

【図17-1】図17は、治療法の開発のためにTCRを発現系にクローニングするための例示的なP526コンストラクト骨格のヌクレオチド配列を示す。図17は、配列番号23を開示する。

30

【図17-2】図17-1の続きを示す。

【0059】

【図18-1】図18は、治療法の開発のために患者の新生抗原特異的TCRのクロナタイプ1を発現系にクローニングするための例示的なコンストラクト配列を示す。図18は、配列番号24を開示する。

【図18-2】図18-1の続きを示す。

【0060】

【図19-1】図19は、治療法の開発のために患者の新生抗原特異的TCRのクロナタイプ3を発現系にクローニングするための例示的なコンストラクト配列を示す。図19は、配列番号25を開示する。

40

【図19-2】図19-1の続きを示す。

【0061】

【図20】一実施形態により、患者に個別化された新生抗原特異的治療を行うための方法のフローチャートである。

【0062】

【図21】図1及び3に示した実体を実施するための例示的なコンピュータを説明する。

【発明を実施するための形態】

【0063】

50

## 詳細な説明

### Ⅰ. 定義

一般に、特許請求の範囲及び明細書において使用される用語は、当業者により理解される通常の意味を有するものとして解釈されるものとする。特定の用語を、さらなる明確性を与えるために下記に定義する。通常の意味と与えられる定義との間に矛盾が存在する場合、与えられる定義が用いられるものとする。

#### 【0064】

本明細書で使用するところの「抗原」という用語は、免疫反応を誘導する物質のことである。

#### 【0065】

本明細書で使用するところの「新生抗原」という用語は、例えば、腫瘍細胞の変異、または腫瘍細胞に特異的な翻訳後修飾によって、抗原に対応する野生型の親抗原とは異なるものとする少なくとも1つの変化を有する抗原のことである。新生抗原は、ポリペプチド配列またはヌクレオチド配列を含んでよい。変異は、フレームシフトもしくは非フレームシフト挿入欠失 (indel)、ミスセンスもしくはナンセンス置換、スプライス部位変化、ゲノム再編成もしくは遺伝子融合、または、新生ORFを生じる任意のゲノム変化もしくは発現変化を含むことができる。変異はまた、スプライス変異体も含むことができる。腫瘍細胞に特異的な翻訳後修飾は、異常リン酸化を含むことができる。腫瘍細胞に特異的な翻訳後修飾はまた、プロテアソームによって生成されるスプライス抗原も含むことができる。Liepe et al., A large fraction of HLA class I ligands are proteasome-generated spliced peptides; Science. 2016 Oct 21; 354 (6310): 354-358を参照されたい。

#### 【0066】

本明細書で使用するところの「腫瘍新生抗原」という用語は、対象の腫瘍細胞または組織中に存在するが、対象の対応する正常細胞または組織中には存在しない新生抗原のことである。

#### 【0067】

本明細書において使用される場合、「新生抗原ベースのワクチン」という用語は、1つ以上の新生抗原、例えば複数の新生抗原に基づいたワクチンコンストラクトのことである。

#### 【0068】

本明細書において使用される場合、「候補新生抗原」という用語は、新生抗原を表しうる新たな配列を生じる変異、または他の異常のことである。

#### 【0069】

本明細書において使用される場合、「コード領域」という用語は、遺伝子における、タンパク質をコードする部分のことである。

#### 【0070】

本明細書において使用される場合、「コード変異」という用語は、コード領域で生じる変異のことである。

#### 【0071】

本明細書において使用される場合、「ORF」という用語は、オープンリーディングフレームを意味する。

#### 【0072】

本明細書において使用される場合、「新生ORF」という用語は、変異またはスプライシングなどの他の異常により生じる腫瘍特異的なORFのことである。

#### 【0073】

本明細書において使用される場合、「ミスセンス変異」という用語は、1つのアミノ酸から別のアミノ酸への置換を引き起こす変異である。

#### 【0074】

本明細書において使用される場合、「ナンセンス変異」という用語は、アミノ酸から終

10

20

30

40

50

止コドンへの置換を引き起こす変異である。

【0075】

本明細書において使用される場合、「フレームシフト変異」という用語は、タンパク質のフレームに変更を引き起こす変異である。

【0076】

本明細書において使用される場合、「挿入欠失」という用語は、1つ以上の核酸の挿入または欠失である。

【0077】

本明細書において使用される場合、2つ以上の核酸またはポリペプチドの配列との関連での「同一性」(%)という用語は、下記の配列比較アルゴリズム（例えば、BLASTP及びBLASTN、または当業者が利用可能な他のアルゴリズム）のうちの1つを用いて、または目視検査により測定される、最大の一致について比較し、整列させた場合に、ヌクレオチドまたはアミノ酸残基の特定の比率(%)が同じである2つ以上の配列または部分配列のことを指す。用途に応じて、「同一性」(%)は、比較される配列の領域にわたって、例えば、機能ドメインにわたって存在するか、あるいは、比較される2つの配列の完全長にわたって存在することができる。

【0078】

配列比較では、一般的に、1つの配列が、試験配列が比較される参照配列として機能する。配列比較アルゴリズムを用いる場合、試験配列及び参照配列をコンピュータに入力し、必要な場合には部分配列座標を指定し、配列アルゴリズムプログラムのパラメータを指定する。次いで、配列比較アルゴリズムが、指定されたプログラムパラメータに基づいて、参照配列に対する試験配列の配列同一性(%)を算出する。あるいは、配列の類似性または相違性は、選択された配列位置（例えば、配列モチーフ）における特定のヌクレオチドの、または翻訳後の配列ではアミノ酸の有無の組み合わせによって確立することもできる。

【0079】

比較を行うための配列の最適なアラインメントは、例えば、Smith & Waterman, Adv. Appl. Math. 2:482 (1981)の局所相同性アルゴリズムによって、Needleman & Wunsch, J. Mol. Biol. 48:443 (1970)の相同性アラインメントアルゴリズムによって、Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988)の類似性の探索法によって、これらのアルゴリズムのコンピュータ処理による実行(Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.におけるGAP、BESTFIT、FASTA、及びTFASTA)によって、または目視検査によって実施することができる（一般的には、下記のAusubel et al.を参照）。

【0080】

配列同一性(%)及び配列類似性(%)を決定するのに適したアルゴリズムの1つの例として、Altschul et al., J. Mol. Biol. 215:403-410 (1990)に記載されるBLASTアルゴリズムがある。BLAST解析を行うためのソフトウェアは、National Center for Biotechnology Informationを通して公に入手可能である。

【0081】

本明細書において使用される場合、「ノンストップまたはリードスルー」という用語は、天然の終止コドンの除去を引き起こす変異のことである。

【0082】

本明細書において使用される場合、「エпитープ」という用語は、抗体またはT細胞受容体が一般的に結合する、抗原の特異的な部分のことである。

【0083】

10

20

30

40

50

本明細書において使用される場合、「免疫原性」という用語は、例えば、T細胞、B細胞、またはその両方を介して免疫応答を誘発する能力のことである。

【0084】

本明細書において使用される場合、「HLA結合親和性」、「MHC結合親和性」という用語は、特異的な抗原と特異的なMHCアレルとの結合の親和性を意味する。

【0085】

本明細書において使用される場合、「ベイト」という用語は、DNAまたはRNAの特異的な配列を試料から濃縮するために使用される核酸プローブのことである。

【0086】

本明細書において使用される場合、「変異」という用語は、対象の核酸と、コントロールとして使用される参照ヒトゲノムとの差異である。

10

【0087】

本明細書において使用される場合、「変異コール」という用語は、典型的にはシーケンシングからの、変異の存在のアルゴリズム的決定である。

【0088】

本明細書において使用される場合、「多型」という用語は、生殖細胞系列変異、すなわち、個体のすべてのDNA保有細胞において見出される変異である。

【0089】

本明細書において使用される場合、「体細胞変異」という用語は、個体の非生殖細胞において生じる変異である。

20

【0090】

本明細書において使用される場合、「アレル」という用語は、遺伝子の1つのバージョンまたは遺伝子配列の1つのバージョンまたはタンパク質の1つのバージョンのことである。

【0091】

本明細書において使用される場合、「HLA型」という用語は、HLA遺伝子アレルの相補体のことである。

【0092】

本明細書において使用される場合、「ナンセンス変異依存分解機構」または「NMD」という用語は、未成熟な終止コドンに起因する細胞によるmRNAの分解のことである。

30

【0093】

本明細書において使用される場合、「トランカル変異 (truncal mutation)」という用語は、腫瘍の発生の初期に生じ、腫瘍の細胞の大部分に存在する変異である。

【0094】

本明細書において使用される場合、「サブクローナル変異」という用語は、腫瘍の発生において後期に生じ、腫瘍の細胞の一部のみに存在する変異である。

【0095】

本明細書において使用される場合、「エクソーム」という用語は、タンパク質をコードするゲノムのサブセットである。エクソームは、ゲノムの集合的なエクソンでありうる。

40

【0096】

本明細書において使用される場合、「ロジスティック回帰」という用語は、従属変数が1に等しい確率のロジットが従属変数の線形関数としてモデル化される、統計からのバイナリデータ用の回帰モデルである。

【0097】

本明細書において使用される場合、「ニューラルネットワーク」という用語は、多層の線形変換に続いて一般的に確率的勾配降下法及び逆伝搬により訓練された要素ごとの非線形変換を行うことからなる分類または回帰のための機械学習モデルである。

【0098】

本明細書において使用される場合、「プロテオーム」という用語は、細胞、細胞の群、

50

または個体によって発現される、及び／または翻訳されるすべてのタンパク質のセットのことである。

【0099】

本明細書において使用される場合、「ペプチドーム」という用語は、細胞表面上のMHC-IまたはMHC-IIによって提示されるすべてのペプチドのセットのことである。ペプチドームは、細胞または細胞の集合の性質を指す場合もある（例えば、腫瘍ペプチドームは、腫瘍を含むすべての細胞のペプチドームの和集合を意味する）。

【0100】

本明細書において使用される場合、「ELISPOT」という用語は、ヒト及び動物において免疫応答を観察するための一般的な方法である、酵素結合免疫吸着スポットアッセイを意味する。

10

【0101】

本明細書において使用される場合、「デキストラマー」という用語は、フローサイトメトリーにおいて抗原特異的T細胞染色に使用される、デキストランベースのペプチド-MHCマルチマーである。

【0102】

本明細書において使用される場合、「MHCマルチマー」という用語は、複数のペプチド-MHCモノマー単位からなるペプチド-MHC複合体である。

【0103】

本明細書において使用される場合、「MHCテトラマー」という用語は、4つのペプチド-MHCモノマー単位からなるペプチド-MHC複合体である。

20

【0104】

本明細書において使用される場合、「寛容または免疫寛容」という用語は、1つ以上の抗原、例えば、自己抗原に対する免疫不応答の状態のことである。

【0105】

本明細書において使用される場合、「中枢性寛容」という用語は、自己反応性T細胞クローンを欠失させること、または自己反応性T細胞クローンの免疫抑制性制御性T細胞（Treg）への分化を促進することのいずれかにより、胸腺において与えられる寛容である。

【0106】

本明細書において使用される場合、「末梢性寛容」という用語は、中枢性寛容を生き延びた自己反応性T細胞を下方制御もしくはアネルギー化すること、またはこれらのT細胞のTregへの分化を促進することにより、末梢系において与えられる寛容である。

30

【0107】

「試料」という用語は、静脈穿刺、排泄、射精、マッサージ、生検、針吸引、洗浄試料、擦過、外科的切開、もしくは介入、または当技術分野において公知の他の手段を含む手段によって対象から採取された、単一細胞、または複数の細胞、または細胞の断片、または体液のアリコートを含むことができる。

【0108】

「対象」という用語は、インビボ、エクスピボ、またはインビトロ、雄または雌のいずれかの、細胞、組織、または生物体、ヒトまたは非ヒトを包含する。対象という用語は、ヒトを含む哺乳動物を含める。

40

【0109】

「哺乳動物」という用語は、ヒト及び非ヒトの両方を包含し、ヒト、非ヒト霊長類、イヌ、ネコ、マウス、ウシ、ウマ、及びブタを含むが、それらに限定されない。

【0110】

「臨床的因子」という用語は、対象の状態、例えば、疾患の活性または重症度の測定を指す。「臨床的因子」は、非試料マーカを含む、対象の健康状態のすべてのマーカ、ならびに／または、非限定的に年齢及び性別などの、対象の他の特徴を包含する。臨床的因子は、対象または所定の条件下の対象由来の試料（または試料の集団）の評定から取得

50

され得るスコア、値、または値のセットであることができる。臨床的因子はまた、マーカー、及び／または遺伝子発現代替物などの他のパラメータによっても予測することができる。臨床的因子は、腫瘍タイプ、腫瘍サブタイプ、及び喫煙歴を含むことができる。

【0111】

略語：MHC：主要組織適合性複合体；HLA：ヒト白血球抗原、またはヒトMHC遺伝子座；NGS：次世代シーケンシング；PPV：陽性適中率；TSNA：腫瘍特異的新生抗原；FFPE：ホルマリン固定パラフィン包埋；NMD：ナンセンス変異依存分解機構；NSCLC：非小細胞肺癌；DC：樹状細胞。

【0112】

本明細書及び添付の特許請求の範囲において使用される場合、単数形「a」、「an」、及び「the」は、文脈によってそうでない旨が明示されない限り、複数の指示物を含む点に留意されたい。

【0113】

本明細書において直接定義されていない用語は、本発明の技術分野の範囲内で理解されるような、一般的にそれらに付随する意味を有するものとして理解されるべきである。本発明の態様の組成物、装置、方法など、ならびにそれらの製造または使用法を説明するうえで実施者にさらなる手引きを与える目的で特定の用語が本明細書で検討される。同じものについて複数の言い方がなされうる点は認識されるであろう。したがって、代替的な語及び同義語が、本明細書で検討される用語の任意の1つ以上について用いられる場合がある。本明細書においてある用語が詳述または検討されているか否かに重きが置かれるべきではない。いくつかの同義語または代用可能な方法、材料などが提供される。1つまたは数個の同義語または均等物の記載は、明確に述べられない限り、他の同義語または均等物の使用を除外しない。用語の例を含む例の使用は、あくまで説明を目的としたものにすぎず、本明細書における発明の態様の範囲及び意味を限定しない。

【0114】

本明細書の本文において引用されるすべての参考文献、発行特許、及び特許出願は、あらゆる目的でそれらの全容を参照により本明細書に援用するものである。

【0115】

II. 新生抗原を特定する方法

本明細書では、腫瘍細胞の表面上にクラスII MHCアレルによって提示される確率の高い、対象の腫瘍細胞由来の新生抗原に対して抗原特異的であるT細胞を特定するための方法を開示する。本方法は、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、及び／または全ゲノムのヌクレオチドシーケンシングデータを取得することを含む。このヌクレオチドシーケンシングデータを用いて新生抗原のセットの中の各新生抗原のペプチド配列が取得される。新生抗原のセットは、腫瘍細胞からのヌクレオチドシーケンシングデータと、正常細胞からのヌクレオチドシーケンシングデータと比較することによって特定される。具体的には、新生抗原のセットの中の各新生抗原のペプチド配列は、ペプチド配列を対象の正常細胞から特定された対応する野生型の親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。本方法は、新生抗原のセットの中の各新生抗原のペプチド配列を、対応する数値ベクトルにコード化することをさらに含む。各数値ベクトルは、ペプチド配列を構成するアミノ酸及びペプチド配列におけるアミノ酸の位置を記述する情報を含む。本方法は、機械学習させた提示モデルに数値ベクトルを入力することにより、新生抗原のセットの中の各新生抗原についての提示尤度を生成することをさらに含む。各提示尤度は、対応する新生抗原が対象の腫瘍細胞の表面上のクラスII MHCアレルによって提示される尤度を表す。機械学習させた提示モデルは、複数のパラメータ及び関数を含む。複数のパラメータは、訓練データセットに基づいて特定される。訓練データセットは、複数の試料の中のそれぞれの試料についての、その試料中に存在するものとして特定されたクラスII MHCアレルのセットの中の少なくとも1つのクラスII 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と、ペプチドを構成するアミノ酸及び／またはペプチドにおけるアミノ酸の位置を記述

10

20

30

40

50

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とを含む。関数は、機械学習させた提示モデルによって入力として受け取られる数値ベクトルと、機械学習させた提示モデルによって数値ベクトル及び複数のパラメータに基づいて出力として生成される提示尤度との間の関係を表す。本方法は、新生抗原のセットのサブセットを、提示尤度に基づいて選択することにより、選択された新生抗原のセットを生成することをさらに含む。本方法は、サブセットの中の新生抗原のうちの少なくとも1つに対して抗原特異的なT細胞を特定することと、これらの特定されたT細胞を返すこととをさらに含む。

#### 【0116】

いくつかの実施形態では、機械学習させた提示モデルに数値ベクトルを入力することは、機械学習させた提示モデルを新生抗原のペプチド配列に適用してクラスII MHCアレルのそれぞれについて依存性スコアを生成することを含む。あるクラスII MHCアレルについての依存性スコアは、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そのクラスII MHCアレルが新生抗原を提示するかどうかを示す。さらなる実施形態では、機械学習させた提示モデルに数値ベクトルを入力することは、依存性スコアを変換することによって、各クラスII MHCアレルについて、対応するクラスII MHCアレルが対応する新生抗原を提示する尤度を示す、対応するアレルごとの尤度を生成することと、アレルごとの尤度を組み合わせることによって新生抗原の提示尤度を生成することと、をさらに含む。いくつかの実施形態では、依存性スコアを変換することは、新生抗原の提示をクラスII 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。代替的な実施形態では、機械学習させた提示モデルに数値ベクトルを入力することは、依存性スコアの組み合わせを変換して提示尤度を生成することをさらに含む。かかる実施形態では、依存性スコアの組み合わせを変換することは、新生抗原の提示をクラスII MHCアレル間の干渉としてモデル化する。

#### 【0117】

いくつかの実施形態では、提示尤度のセットは、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る。かかる実施形態では、本方法は、機械学習させた提示モデルをアレル非相互作用特性に適用することによってアレル非相互作用特性に対する依存性スコアを生成することをさらに含む。依存性スコアは、対応する新生抗原のペプチド配列が、アレル非相互作用特性に基づいて提示されるかどうかを示す。いくつかの実施形態では、本方法は、各クラスII MHCアレルについての依存性スコアをアレル非相互作用特性についての依存性スコアと組み合わせることと、各クラスII MHCアレルについての組み合わせられた依存性スコアを変換して各クラスII MHCアレルについてアレルごとの尤度を生成することと、アレルごとの尤度を組み合わせることで提示尤度を生成することと、をさらに含む。あるクラスII MHCアレルについてのアレルごとの尤度は、そのクラスII MHCアレルが対応する新生抗原を提示する尤度を示す。代替的な実施形態では、本方法は、クラスII MHCアレルについての依存性スコアとアレル非相互作用特性についての依存性スコアとを組み合わせることと、組み合わせられた依存性スコアを変換して提示尤度を生成することと、をさらに含む。

#### 【0118】

いくつかの実施形態では、クラスII MHCアレルは2つ以上の異なるクラスII MHCアレルを含む。

#### 【0119】

いくつかの実施形態では、訓練データセットの試料中に存在するものとして特定されたクラスII MHCアレルのセットの中の少なくとも1つのクラスII MHCアレルは、2つ以上の異なるタイプのクラスII MHCアレルを含む。

#### 【0120】

いくつかの実施形態では、ペプチド配列は、アミノ酸9個以外の長さを有するペプチド配列を含む。

#### 【0121】

いくつかの実施形態では、ペプチド配列をコード化することは、ワンホットコード化ス

10

20

30

40

50

キームを用いてペプチド配列をコード化することを含む。

【0122】

特定の実施形態では、複数の試料は、単一のクラスII MHCアレルを発現するように操作された細胞株、複数のクラスII MHCアレルを発現するように操作された細胞株、複数の患者から得られたまたは由来するヒト細胞株、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。

【0123】

いくつかの実施形態では、訓練データセットは、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、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。

10

【0124】

いくつかの実施形態では、提示尤度のセットは、RNA-seqまたは質量分析により測定される、対象内のクラスII MHCアレルの発現レベルによってさらに特定される。

【0125】

いくつかの実施形態では、提示尤度のセットは、新生抗原のセットの中の新生抗原とクラスII MHCアレルとの間の予測される親和性、及び、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少なくとも一方を含む特性によってさらに特定される。

【0126】

いくつかの実施形態では、数値的尤度のセットは、そのソースタンパク質配列内の新生抗原コード化ペプチド配列に隣接するC末端配列、及びそのソースタンパク質配列内の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。

20

【0127】

いくつかの実施形態では、選択された新生抗原のセットを選択することは、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に比べて、腫瘍細胞表面上に提示される尤度が高い新生抗原を選択することを含む。

【0128】

いくつかの実施形態では、選択された新生抗原のセットを選択することは、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に比べて、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高い新生抗原を選択することを含む。

30

【0129】

いくつかの実施形態では、選択された新生抗原のセットを選択することは、提示モデルに基づいて、選択されない新生抗原に比べて、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細胞に提示される尤度が高い新生抗原を選択することを含む。かかる実施形態では、APCは場合により樹状細胞（DC）である。

【0130】

いくつかの実施形態では、選択された新生抗原のセットを選択することは、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に比べて、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低い新生抗原を選択することを含む。

40

【0131】

いくつかの実施形態では、選択された新生抗原のセットを選択することは、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に比べて、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低い新生抗原を選択することを含む。

【0132】

いくつかの実施形態では、1つ以上の腫瘍細胞は、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、及び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より選択される。

【0133】

50

いくつかの実施形態では、本方法は、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。かかる実施形態では、個別化がんワクチン用の出力は、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ことができる。

【0134】

いくつかの実施形態では、機械学習させた提示モデルは、ニューラルネットワークモデルである。かかる実施形態では、ニューラルネットワークモデルは、それぞれのネットワークモデルが、クラスII MHCアレルのうちの対応するクラスII MHCアレルに割り当てられ、1つ以上の層に配置された一連のノードを含む、クラスII MHCアレルについての複数のネットワークモデルを含み得る。かかる実施形態では、ニューラルネットワークモデルは、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することができ、少なくとも2つのネットワークモデルのパラメータが少なくとも1回の訓練反復において一緒に更新される。

10

【0135】

かかる実施形態では、各ネットワークモデルは1つ以上の畳み込みニューラルネットワークをさらに含むことができ、1つ以上の畳み込みニューラルネットワークのそれぞれは、1つ以上の層に配置され、異なるサイズのフィルターを有する一連のノードを含む。1つ以上の畳み込みニューラルネットワークのそれぞれのフィルターは、ペプチド配列の結合コアまたは結合アンカーを含む各新生抗原のペプチド配列におけるアミノ酸の位置を特定するようなサイズとすることができる。

20

【0136】

いくつかの実施形態では、機械学習させた提示モデルは、1層以上のノードの層を含むディープラーニングモデルとすることができる。

【0137】

いくつかの実施形態では、T細胞を特定することは、T細胞を増殖させる条件下でT細胞をサブセットの中の新生抗原のうちの1つ以上と共培養することを含む。

【0138】

いくつかの実施形態では、T細胞を特定することは、T細胞を、サブセットの中の新生抗原のうちの1つ以上を含むMHCマルチマーと、T細胞とMHCマルチマーとの結合が可能な条件下で接触させることを含む。

30

【0139】

いくつかの実施形態では、本方法は、特定されたT細胞のT細胞受容体(TCR)を特定することをさらに含む。かかる実施形態では、T細胞受容体を特定することは、特定されたT細胞のT細胞受容体配列をシークエンシングすることを含む。かかる実施形態では、本方法は、1つ以上の特定されたT細胞受容体のうちの少なくとも1つを発現するようにT細胞を遺伝子操作することと、T細胞が増殖する条件下でT細胞を培養することと、増殖させたT細胞を対象に注入することと、をさらに含むことができる。かかる実施形態では、特定されたT細胞受容体のうちの少なくとも1つを発現するようにT細胞を遺伝子操作することは、特定されたT細胞のT細胞受容体配列を発現ベクターにクローニングすることと、T細胞のそれぞれに発現ベクターをトランスフェクトすることと、を含むことができる。

40

【0140】

いくつかの実施形態では、本方法は、特定されたT細胞を、特定されたT細胞が増殖する条件下で培養することと、増殖させたT細胞を対象に注入することと、をさらに含むことができる。

【0141】

本明細書では、上記に述べた新生抗原のサブセットの中の少なくとも1つの選択された新生抗原に対して抗原特異的である単離されたT細胞も開示される。

【0142】

国際特許出願公開第WO2018/195357号及び国際特許出願公開第WO201

50

9 / 0 5 0 9 9 4 号を、それらの全容にわたって参照により本明細書に援用する。国際特許出願公開第WO 2 0 1 8 / 1 9 5 3 5 7 号は、MHCクラスII分子による抗原提示を予測するための方法について記載している。国際特許出願公開第WO 2 0 1 9 / 0 5 0 9 9 4 号は、MHC分子によって提示される抗原に抗原特異的なT細胞を特定するための方法について記載している。これらの公開公報について本明細書のこのセクションで言及しているが、国際特許出願公開第WO 2 0 1 8 / 1 9 5 3 5 7 号及び同第WO 2 0 1 9 / 0 5 0 9 9 4 号に示される開示は、それらの全容を本出願のすべてのセクションに参照により援用する。

【0143】

#### III. 新生抗原における腫瘍特異的変異の特定

また、ある特定の変異（例えば、がん細胞中に存在する変異またはアレル）の特定のための方法も、本明細書に開示する。特に、これらの変異は、がんを有する対象のがん細胞のゲノム、トランスクリプトーム、プロテオーム、またはエクソーム中に存在し得るが、対象由来の正常組織には存在し得ない。

【0144】

腫瘍における遺伝子変異は、それらが腫瘍において排他的にタンパク質のアミノ酸配列における変更をもたらす場合、腫瘍の免疫学的ターゲティングに有用と考えることができる。有用な変異は、以下を含む：（１）タンパク質において異なるアミノ酸をもたらす非同義変異；（２）C末端に新規の腫瘍特異的配列を有する、より長いタンパク質の翻訳をもたらす、終止コドンが修飾されているかまたは欠失しているリードスルー変異；（３）成熟mRNAにおけるイントロンの包含、したがって固有の腫瘍特異的タンパク質配列をもたらす、スプライス部位変異；（４）２種類のタンパク質の接合部に腫瘍特異的配列を有するキメラタンパク質を生じる、染色体再編成（すなわち、遺伝子融合）；（５）新規の腫瘍特異的タンパク質配列を有する新たなオープンリーディングフレームをもたらす、フレームシフト変異または欠失。変異はまた、非フレームシフト挿入欠失、ミスセンスもしくはナンセンス置換、スプライス部位変化、ゲノム再編成もしくは遺伝子融合、または、新生ORFを生じる任意のゲノム変化もしくは発現変化のうちの１つ以上も含むことができる。

【0145】

例えば、腫瘍細胞におけるスプライス部位、フレームシフト、リードスルー、または遺伝子融合の変異から生じた、変異を有するペプチドまたは変異したポリペプチドは、腫瘍対正常細胞において、DNA、RNA、またはタンパク質をシーケンシングすることによって特定することができる。

【0146】

また、変異は、以前に特定された腫瘍特異的変異を含むことができる。公知の腫瘍変異は、Catalogue of Somatic Mutations in Cancer (COSMIC) データベースで見出すことができる。

【0147】

様々な方法を、個体のDNAまたはRNAにおいて特定の変異またはアレルの存在を検出するために利用可能である。この分野における進歩は、正確で、容易な、かつ安価な大規模SNP遺伝子型判定を提供している。例えば、動的アレル特異的ハイブリダイゼーション (DASH)、マイクロプレートアレイ対角線ゲル電気泳動 (MADGE)、パイロシーケンシング、オリゴヌクレオチド特異的ライゲーション、TaqManシステム、及びAffymetrix SNPチップなどの種々のDNA「チップ」技術を含むいくつかの技法が、記載されている。これらの方法は、典型的にはPCRによる、標的遺伝子領域の増幅を利用する。さらに他の方法は、侵襲性切断による小さなシグナル分子の生成及びその後の質量分析、または、固定化されたパッドロックプローブ及びローリングサークル増幅に基づく。特異的な変異を検出するための、当技術分野において公知の方法のいくつかを、下記に要約する。

【0148】

10

20

30

40

50

PCRベースの検出手段は、多数のマーカの多重増幅を同時に含むことができる。例えば、サイズがオーバーラップせず、同時に解析することができるPCR産物を生成するようにPCRプライマーを選択することが、当技術分野において周知である。あるいは、差次的にラベル化され、したがって、各々を差次的に検出することができるプライマーで異なるマーカを増幅することが可能である。当然、ハイブリダイゼーションベースの検出手段により、試料における複数のPCR産物の差次的な検出が可能になる。複数のマーカの多重解析を可能にする他の技法が、当技術分野において公知である。

【0149】

いくつかの方法が、ゲノムDNAまたは細胞RNAにおける単一ヌクレオチド多型の解析を容易にするために開発されている。例えば、一塩基多型は、例えば、Mundy, C. R. (米国特許第4,656,127号)において開示されているような、特化されたエキソヌクレアーゼ抵抗性ヌクレオチドを用いることによって検出することができる。この方法にしたがって、多型部位のすぐ3'のアレル配列に対して相補的なプライマーを、特定の動物またはヒトから取得された標的分子に対してハイブリダイズさせる。標的分子上の多型部位が、存在する特定のエキソヌクレアーゼ抵抗性ヌクレオチド誘導体に対して相補的であるヌクレオチドを含有する場合、その誘導体は、ハイブリダイズされたプライマーの末端上に組み込まれる。そのような組み込みのために、プライマーはエキソヌクレアーゼに対して抵抗性になり、それによりその検出が可能になる。試料のエキソヌクレアーゼ抵抗性誘導体の同一性は既知であるため、プライマーがエキソヌクレアーゼに対して抵抗性になったという知見により、標的分子の多型部位に存在するヌクレオチドが、反応において使用されたヌクレオチド誘導体のものに対して相補的であることが明らかになる。この方法は、多量の外来性配列データの決定を必要としないという利点を有する。

【0150】

多型部位のヌクレオチドの同一性を決定するために、溶液ベースの方法を使用することができる(Cohen, D. et al. (フランス国特許第2,650,840号;PCT出願第WO91/02087号)。米国特許第4,656,127号のMundyの方法におけるように、多型部位のすぐ3'のアレル配列に対して相補的であるプライマーを使用する。この方法は、多型部位のヌクレオチドに対して相補的である場合は、プライマーの末端上に組み込まれるようになる、ラベル化ジデオキシヌクレオチド誘導体を用いて、その部位のヌクレオチドの同一性を決定する。

【0151】

Genetic Bit AnalysisまたはGBAとして公知である代替的な方法が、Golet, P. et al. (PCT出願第92/15712号)により記載されている。Golet, P. et al.の方法は、ラベル化ターミネーターと、多型部位の3'の配列に対して相補的であるプライマーとの混合物を使用する。Golet, P. et al.の方法は、ラベル化ターミネーターと、多型部位の3'の配列に対して相補的であるプライマーとの混合物を使用する。Cohen et al. (フランス国特許第2,650,840号;PCT出願第WO91/02087号)の方法とは対照的に、Golet, P. et al.の方法は、プライマーまたは標的分子が固相に固定化される、不均一相アッセイであることができる。

【0152】

DNAにおいて多型部位をアッセイするための、いくつかのプライマーガイドヌクレオチド組み込み手順が、記載されている(Komher, J. S. et al., Nucl. Acids. Res. 17:7779-7784 (1989); Sokolov, B. P., Nucl. Acids Res. 18:3671 (1990); Syvanen, A.-C., et al., Genomics 8:684-692 (1990); Kuppuswamy, M. N. et al., Proc. Natl. Acad. Sci. (U.S.A.) 88:1143-1147 (1991); Prezant, T. R. et al., Hum. Mutat. 1:159-164 (1992); Ugozzoli, L. et al., GATA 9:107-112 (1992); Nyren, P. et a

10

20

30

40

50

l., Anal. Biochem. 208:171-175 (1993))。これらの方法は、それらが、多型部位で塩基間を識別するためにラベル化デオキシヌクレオチドの組み込みを利用する点で、GBAとは異なる。そのような形式において、シグナルは、組み込まれたデオキシヌクレオチドの数に比例するため、同じヌクレオチドのランにおいて起こる多型は、ランの長さに比例するシグナルを結果としてもたらしることができる (Syvanen, A.-C., et al., Amer. J. Hum. Genet. 52:46-59 (1993))。

#### 【0153】

数多くのイニシアティブは、DNAまたはRNAの何百万もの個々の分子から並行して直接、配列情報を取得する。リアルタイムの単一分子の合成によるシーケンシング技術は、シーケンシングされる鋳型に対して相補的であるDNAの新生鎖の中に組み込まれる際の、蛍光ヌクレオチドの検出に依拠する。1つの方法において、長さが30～50塩基のオリゴヌクレオチドを、ガラスのカバーガラスに、5'端で共有結合性に固着させる。これらの固着した鎖は、2つの機能を果たす。第1に、それらは、鋳型が、表面結合オリゴヌクレオチドに対して相補的な捕捉尾部を有して構成されている場合に、標的鋳型鎖の捕捉部位として作用する。それらはまた、配列読み取りの基礎を形成する、鋳型指向性プライマー伸長のためのプライマーとしても作用する。捕捉プライマーは、複数サイクルの合成、検出、及び、色素を除去するための色素-リンカーの化学的切断を用いた、配列決定のための、固定された位置部位として機能する。各サイクルは、ポリメラーゼ/ラベル化ヌクレオチド混合物の添加、リンス、画像化、及び色素の切断からなる。代替的な方法において、ポリメラーゼは、蛍光ドナー分子で修飾されてスライドガラス上に固定化され、他方、各ヌクレオチドは、γ-ホスファートに付着したアクセプター蛍光部分で色分けされている。ヌクレオチドが、新規の鎖の中に組み込まれるようになる際に、システムが、蛍光タグ付加されたポリメラーゼと蛍光修飾されたヌクレオチドとの間の相互作用を検出する。他の合成によるシーケンシング技術もまた、存在する。

#### 【0154】

任意の適している合成によるシーケンシングプラットフォームを、変異を特定するために使用することができる。上記のように、4種類の主要な合成によるシーケンシングプラットフォームを、現在利用可能である: Roche/454 Life Sciencesより販売されるGenome Sequencer、Illumina/Solexaより販売される1G Analyzer、Applied Biosystemsより販売されるSOLiDシステム、及びHelicos Bioscienceより販売されるHeliscopeシステム。合成によるシーケンシングプラットフォームはまた、Pacific Biosciences及びVisiGen Biotechnologiesによっても記載されている。いくつかの実施形態において、シーケンシングされる多数の核酸分子は、支持体（例えば、固体支持体）に結合している。核酸を支持体上に固定化するために、捕捉配列/万能プライミング部位を、鋳型の3'端及び/または5'端に付加することができる。核酸は、支持体に共有結合性に付着した相補的配列に対して捕捉配列をハイブリダイズすることによって、支持体に結合させることができる。捕捉配列（万能捕捉配列とも呼ばれる）は、万能プライマーとして二重に働き得る、支持体に付着した配列に対して相補的な核酸配列である。

#### 【0155】

捕捉配列に対する代替物として、カップリングペア（例えば、抗体/抗原、受容体/リガンド、または、例えば米国特許出願第2006/0252077号に記載されているようなアビジン-ビオチンペアなど）のメンバーを、各断片に連結させて、そのカップリングペアのそれぞれの第2のメンバーでコーティングされた表面上に捕捉させることができる。

#### 【0156】

捕捉に続いて、配列を、例えば、鋳型依存的な合成によるシーケンシングを含む、例えば、実施例及び米国特許第7,283,337号に記載されているような、単一分子検

10

20

30

40

50

出／シーケンシングによって解析することができる。合成によるシーケンシングにおいて、表面に結合した分子は、ポリメラーゼの存在下で、多数のラベル化ヌクレオチド三リン酸に曝露される。鋳型の配列は、成長する鎖の3'端の中に組み込まれるラベル化ヌクレオチドの順序によって決定される。これは、リアルタイムで行うことができ、ステップ・アンド・リピートモードで行うことができる。リアルタイム解析のために、各ヌクレオチドに対して異なる光ラベルを組み込むことができ、複数のレーザーを、組み込まれたヌクレオチドの刺激のために利用することができる。

【0157】

シーケンシングはまた、他の大規模並列処理シーケンシング、または次世代シーケンシング（NGS）技法及びプラットフォームも含むことができる。大規模並列処理シーケンシング技法及びプラットフォームの追加的な例は、Illumina HiSeqまたはMiSeq、ThermoPGMまたはProton、PacBio RS IまたはSequel、QiagenのGene Reader、及びOxford Nanopore MinIONである。追加的な類似した現在の大規模並列処理シーケンシング技術、及びこれらの技術の将来世代を、使用することができる。

【0158】

任意の細胞タイプまたは組織を利用して、本明細書に記載した方法における使用のための核酸試料を取得することができる。例えば、DNAまたはRNA試料を、腫瘍または体液、例えば、公知の技法（例えば、静脈穿刺）によって取得された血液、もしくは唾液から取得することができる。あるいは、核酸試験を、乾燥試料（例えば、髪または皮膚）に対して行うことができる。加えて、試料を、シーケンシングのために腫瘍から取得することができ、別の試料を、正常組織が腫瘍と同じ組織タイプのものである場合に、シーケンシングのために正常組織から取得することができる。試料を、シーケンシングのために腫瘍から取得することができ、別の試料を、正常試料が腫瘍とは別個の組織タイプのものである場合に、シーケンシングのために正常組織から取得することができる。

【0159】

腫瘍は、肺癌、黒色腫、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、及びT細胞リンパ球性白血病、非小細胞肺癌、及び小細胞肺癌のうちの1つ以上を含むことができる。

【0160】

あるいは、タンパク質質量分析を使用して、腫瘍細胞上のMHCタンパク質に結合した変異したペプチドの存在を特定または実証することができる。ペプチドは、腫瘍細胞から、または腫瘍から免疫沈降させたHLA分子から酸溶出することができ、次いで、質量分析を用いて特定することができる。

【0161】

#### IV. 新生抗原

新生抗原は、ヌクレオチドまたはポリヌクレオチドを含むことができる。例えば、新生抗原は、ポリペプチド配列をコードするRNA配列であることができる。ワクチンにおいて有用な新生抗原は、したがって、ヌクレオチド配列またはポリペプチド配列を含むことができる。

【0162】

本明細書に開示する方法によって特定された腫瘍特異的変異を含む単離されたペプチド、公知の腫瘍特異的変異を含むペプチド、および、本明細書に開示する方法によって特定された変異ポリペプチドまたはその断片を、本明細書に開示する。新生抗原ペプチドは、新生抗原に関連するポリペプチド配列をコードするヌクレオチド配列（例えば、DNAまたはRNA）を含む場合に、それらのコード配列の文脈において記載することができる。

【0163】

新生抗原ヌクレオチド配列によってコードされる1つ以上のポリペプチドは、以下のうちの少なくとも1つを含むことができる：1000nm未満のIC50値でのMHCとの

10

20

30

40

50

結合親和性、MHCクラスIペプチドについてはアミノ酸8～15個、8、9、10、11、12、13、14、または15個の長さ、プロテアソーム切断を促進するペプチド内またはその近くの配列モチーフの存在、及び、TAP輸送を促進する配列モチーフの存在。MHCクラスIIのポリペプチドではアミノ酸6～30、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、または30個の長さ、細胞外またはリソソームプロテアーゼ（例えば、カテプシン類）による切断またはHLA-DMにより触媒されるHLA結合を促進するペプチド内またはその近くの配列モチーフの存在。

【0164】

1つ以上の新生抗原は、腫瘍の表面上に存在することができる。

10

【0165】

1つ以上の新生抗原は、腫瘍を有する対象において免疫原性であることができ、例えば、対象においてT細胞応答またはB細胞応答を惹起することができ得る。

【0166】

対象において自己免疫応答を誘導する1つ以上の新生抗原は、腫瘍を有する対象のためのワクチン生成の文脈において、考察から排除することができる。

【0167】

少なくとも1つの新生抗原ペプチド分子のサイズは、約5個、約6個、約7個、約8個、約9個、約10個、約11個、約12個、約13個、約14個、約15個、約16個、約17個、約18個、約19個、約20個、約21個、約22個、約23個、約24個、約25個、約26個、約27個、約28個、約29個、約30個、約31個、約32個、約33個、約34個、約35個、約36個、約37個、約38個、約39個、約40個、約41個、約42個、約43個、約44個、約45個、約46個、約47個、約48個、約49個、約50個、約60個、約70個、約80個、約90個、約100個、約110個、約120個、またはそれよりも多いアミノ分子残基、及びこれらの範囲から導出される任意の範囲を含むことができるが、それらに限定されない。具体的な実施形態において、新生抗原ペプチド分子は、アミノ酸50個以下である。

20

【0168】

新生抗原ペプチド及びポリペプチドは、MHCクラスIについては長さが15残基以下で、通常約8～約11残基の間からなり、特に9または10残基であることができ；MHCクラスIIについては、6～30残基であることができる。

30

【0169】

望ましい場合、より長いペプチドを、いくつかのやり方において設計することができる。1つの例において、HLAアレル上のペプチドの提示尤度が予測されるかまたは公知である場合、より長いペプチドは、（1）各々の対応する遺伝子産物のN末端及びC末端に向かって2～5アミノ酸の伸長を有する個々の提示されるペプチド；（2）各々について伸長した配列を有する、提示されるペプチドのいくつかまたはすべての連鎖のいずれかからなることができる。別の例において、シーケンシングにより、腫瘍中に存在する長い（10残基より長い）新生エпитープ配列（例えば、新規のペプチド配列をもたらすフレームシフト、リードスルー、またはイントロンの包含による）が明らかになる場合、より長いペプチドは、（3）新規の腫瘍特異的アミノ酸のストレッチ全体からなることになり、したがって、最強のHLAに提示されるより短いペプチドの計算的なまたはインビトロ試験ベースの選択の必要を回避する。いずれの例においても、より長いペプチドの使用によって、患者細胞による内因性のプロセッシングが可能になり、より有効な抗原提示及びT細胞応答の誘導がもたらされ得る。

40

【0170】

新生抗原ペプチド及びポリペプチドは、HLAタンパク質上に提示されることができる。いくつかの態様において、新生抗原ペプチド及びポリペプチドは、野生型ペプチドよりも強い親和性でHLAタンパク質上に提示される。いくつかの態様において、新生抗原ペプチドまたはポリペプチドは、少なくとも5000nM未満、少なくとも1000nM未

50

満、少なくとも500 nM未満、少なくとも250 nM未満、少なくとも200 nM未満、少なくとも150 nM未満、少なくとも100 nM未満、少なくとも50 nM未満、またはそれよりも小さいIC50を有することができる。

【0171】

いくつかの態様において、新生抗原ペプチド及びポリペプチドは、対象に投与された場合に、自己免疫応答を誘導せず、及び／または免疫寛容を引き起こさない。

【0172】

また、少なくとも2種類以上の新生抗原ペプチドを含む組成物も提供する。いくつかの実施形態において、組成物は、少なくとも2種類の異なるペプチドを含有する。少なくとも2種類の異なるペプチドは、同じポリペプチドに由来することができる。異なるポリペプチドとは、ペプチドが、長さ、アミノ酸配列、またはその両方において異なることを意味する。ペプチドは、腫瘍特異的変異を含有することが知られているか、または見出されている任意のポリペプチドに由来する。新生抗原ペプチドが由来することができる、適しているポリペプチドは、例えば、COSMICデータベースにおいて見出すことができる。COSMICは、ヒトがんにおける体細胞性変異についての総合的な情報の管理を行う。ペプチドは、腫瘍特異的変異を含有する。いくつかの態様において、腫瘍特異的変異は、特定のがんタイプについてのドライバー変異である。

【0173】

望ましい活性または性質を有する新生抗原ペプチド及びポリペプチドは、望ましいMHC分子に結合して適切なT細胞を活性化する非改変ペプチドの生物学的活性を増大させるかまたは実質的にそのすべてを少なくとも保持しつつ、特定の望ましい属性、例えば、改善された薬理学的特徴を与えるように改変することができる。例として、新生抗原ペプチド及びポリペプチドを、保存的または非保存的のいずれかの置換などの、種々の改変にさらに供することができ、そのような改変は、改善されたMHC結合、安定性、または提示などの、それらの使用におけるある特定の利点を提供し得る。保存的置換とは、アミノ酸残基を、生物学的及び／または化学的に類似している別のもので、例えば、1つの疎水性残基を別の疎水性残基、または1つの極性残基を別の極性残基で置き換えることを意味する。置換は、Gly、Ala；Val、Ile、Leu、Met；Asp、Glu；Asn、Gln；Ser、Thr；Lys、Arg；及びPhe、Tyrなどの組み合わせを含む。単一アミノ酸置換の効果はまた、D-アミノ酸を用いて探査してもよい。そのような改変は、例えば、Merrifield, Science 232:341-347 (1986), Barany & Merrifield, The Peptides, Gross & Meienhofer, eds. (N.Y., Academic Press), pp. 1-284 (1979)；及びStewart & Young, Solid Phase Peptide Synthesis, (Rockford, Ill., Pierce), 2d Ed. (1984)に記載されているように、周知のペプチド合成手順を用いて行うことができる。

【0174】

種々のアミノ酸模倣物または非天然アミノ酸でのペプチド及びポリペプチドの改変は、インビボでのペプチド及びポリペプチドの安定性の増大に特に有用である場合がある。安定性は多くの方法でアッセイすることができる。例として、ペプチダーゼ、ならびに、ヒト血漿及び血清などの種々の生物学的媒質が、安定性を試験するために使用されている。例えば、Verhoef et al., Eur. J. Drug Metab Pharmacokin. 11:291-302 (1986)を参照されたい。ペプチドの半減期は、25%ヒト血清(v/v)アッセイを用いて好都合に決定することができる。プロトコールは、概して以下のようなものである。プールしたヒト血清(AB型、非熱不活性化)を、使用前に遠心分離によって脱脂する。次いで、血清を、RPMI組織培養培地で25%に希釈し、ペプチド安定性を試験するために使用する。あらかじめ決定された時間間隔で、少量の反応溶液を取り出して、6%水性トリクロロ酢酸またはエタノールのいずれかに添加する。濁った反応試料を15分間冷却(4℃)し、次いで、スピンドして沈降血清タ

10

20

30

40

50

ンパク質を沈殿させる。次いで、ペプチドの存在を、安定性特異的クロマトグラフィー条件を用いた逆相HPLCによって決定する。

#### 【0175】

ペプチド及びポリペプチドを、改善された血清半減期以外の望ましい属性を提供するために修飾することができる。例として、CTL活性を誘導するペプチドの能力を、Tヘルパー細胞応答を誘導することができる少なくとも1つのエピトープを含有する配列への連結によって増強することができる。免疫原性ペプチド／Tヘルパーコンジュゲートは、スペーサー分子によって連結することができる。スペーサーは、典型的には、生理学的条件下で実質的に無電荷である、アミノ酸またはアミノ酸模倣物などの相対的に小さな中性分子から構成される。スペーサーは、典型的には、例えば、Ala、Gly、または、非極性アミノ酸もしくは中性極性アミノ酸の他の中性スペーサーから選択される。任意で存在するスペーサーは、同じ残基から構成される必要はなく、したがって、ヘテロオリゴマーまたはホモオリゴマーであり得ることが、理解されるであろう。存在する場合、スペーサーは、通常、少なくとも1または2残基、より通常は、3～6残基であろう。あるいは、ペプチドを、スペーサーなしでTヘルパーペプチドに連結することができる。

#### 【0176】

新生抗原ペプチドは、ペプチドのアミノ末端またはカルボキシ末端のいずれかで、直接またはスペーサーを介してのいずれかでTヘルパーペプチドに連結することができる。新生抗原ペプチドまたはTヘルパーペプチドのいずれかのアミノ末端を、アシル化することができる。例示的なTヘルパーペプチドは、破傷風毒素の830～843、インフルエンザの307～319、マラリアスポロゾイトの周囲382～398及び378～389を含む。

#### 【0177】

タンパク質またはペプチドは、標準的な分子生物学的技法を通したタンパク質、ポリペプチド、もしくはペプチドの発現、天然由来源からのタンパク質もしくはペプチドの単離、またはタンパク質もしくはペプチドの化学合成を含む、当業者に公知の任意の技法によって作製することができる。種々の遺伝子に対応する、ヌクレオチドならびにタンパク質、ポリペプチド及びペプチドの配列は、以前に開示されており、当業者に公知のコンピュータ処理されたデータベースで見出すことができる。1つのそのようなデータベースは、National Institutes of Healthのウェブサイト位置する、National Center for Biotechnology InformationのGenbank及びGenPeptデータベースである。公知の遺伝子のコード領域は、本明細書に開示する技法を用いて、または当業者に公知であるように、増幅及び／または発現させることができる。あるいは、タンパク質、ポリペプチド、及びペプチドの種々の商業的調製物が、当業者に公知である。

#### 【0178】

さらなる態様において、新生抗原は、新生抗原ペプチドまたはその一部をコードする核酸（例えば、ポリヌクレオチド）を含む。ポリヌクレオチドは、例えば、DNA、cDNA、PNA、CNA、RNA（例えば、mRNA）、例えば、ホスホロチオアートバックボーンを有するポリヌクレオチドなどの、ポリヌクレオチドの一本鎖及び／もしくは二本鎖、または天然形態もしくは安定化形態のいずれか、または、それらの組み合わせであることができ、イントロンを含有してもよく、または含有しなくてもよい。またさらなる態様は、ポリペプチドまたはその一部を発現することができる発現ベクターを提供する。様々な細胞タイプ用の発現ベクターが、当技術分野において周知であり、過度の実験なしで選択することができる。概して、DNAを、プラスミドなどの発現ベクター中に、発現のための適正な方向及び正確なリーディングフレームで挿入する。必要な場合は、DNAを、望ましい宿主によって認識される適切な転写及び翻訳調節性制御ヌクレオチド配列に連結することができるが、そのような制御は、概して発現ベクターにおいて利用可能である。次いで、ベクターを、標準的な技法を通して宿主中に導入する。手引きは、例えば、Sambrook et al. (1989) Molecular Cloning, A L

10

20

30

40

50

aboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y. において見出すことができる。

【0179】

#### IV. ワクチン組成物

また、特異的な免疫応答、例えば、腫瘍特異的な免疫応答を生じることができる免疫原性組成物、例えば、ワクチン組成物も、本明細書に開示する。ワクチン組成物は、典型的に、例えば、本明細書に記載した方法を用いて選択された多数の新生抗原を含む。ワクチン組成物はまた、ワクチンと呼ぶこともできる。

【0180】

ワクチンは、1～30種類のペプチド、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、もしくは30種類の異なるペプチド、6、7、8、9、10、11、12、13、もしくは14種類の異なるペプチド、または12、13、もしくは14種類の異なるペプチドを含有することができる。ペプチドは、翻訳後修飾を含むことができる。ワクチンは、1～100種類もしくはそれよりも多いヌクレオチド配列、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100種類もしくはそれよりも多い異なるヌクレオチド配列、6、7、8、9、10、11、12、13、もしくは14種類の異なるヌクレオチド配列、または12、13、もしくは14種類の異なるヌクレオチド配列を含有することができる。ワクチンは、1～30種類の新生抗原配列、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100種類もしくはそれよりも多い異なる新生抗原配列、6、7、8、9、10、11、12、13、もしくは14種類の異なる新生抗原配列、または12、13、もしくは14種類の異なる新生抗原配列を含有することができる。

【0181】

一実施形態では、異なるペプチド及び／もしくはポリペプチド、またはそれらをコードするヌクレオチド配列は、ペプチド及び／またはポリペプチドが、異なるMHCクラスI分子及び／または異なるMHCクラスII分子などの異なるMHC分子と結合することができるように選択される。いくつかの態様において、1つのワクチン組成物は、最も頻繁に存在するMHCクラスI分子及び／またはMHCクラスII分子と結合することができるペプチド及び／またはポリペプチドのコード配列を含む。したがって、ワクチン組成物は、少なくとも2種類の好ましい、少なくとも3種類の好ましい、または少なくとも4種類の好ましいMHCクラスI分子及び／またはMHCクラスII分子と結合することができる異なる断片を含むことができる。

【0182】

ワクチン組成物は、特異的な細胞傷害性T細胞応答、及び／または特異的なヘルパーT細胞応答を生じることができる。

10

20

30

40

50

## 【0183】

ワクチン組成物は、アジュバント及び／または担体をさらに含むことができる。有用なアジュバント及び担体の例を、本明細書の下記に示す。組成物は、例えば、タンパク質などの担体、または、例えば、T細胞に対してペプチドを提示することができる樹状細胞（DC）などの抗原提示細胞と結合することができる。

## 【0184】

アジュバントは、ワクチン組成物中へのその混合が、新生抗原に対する免疫応答を増大させるか、または別の方法で修飾する任意の物質である。担体は、新生抗原がそれに結合することができる足場構造、例えば、ポリペプチドまたは多糖であることができる。任意で、アジュバントは、共有結合性または非共有結合性にコンジュゲートされる。

10

## 【0185】

抗原に対する免疫応答を増大させるアジュバントの能力は、典型的に、免疫媒介性反応の有意なもしくは実質的な増大、または疾患症候の低減によって明示される。例えば、体液性免疫の増大は、典型的に、抗原に対して生じた抗体の力価の有意な増大によって明示され、T細胞活性の増大は、典型的に、細胞増殖、または細胞性細胞傷害、またはサイトカイン分泌の増大において明示される。アジュバントはまた、例えば、主として体液性またはTh応答を、主として細胞性またはTh応答へと変更することによって、免疫応答を変化させ得る。

## 【0186】

適しているアジュバントは、1018 ISS、アラム、アルミニウム塩、Amplivax、AS15、BCG、CP-870, 893、CpG7909、CyaA、dSLIM、GM-CSF、IC30、IC31、イミキモド、ImuFact IMP321、IS Patch、ISS、ISCOMATRIX、JuvImmune、LipoVac、MF59、モノホスホリル脂質A、Montanide IMS 1312、Montanide ISA206、Montanide ISA 50V、Montanide ISA-51、OK-432、OM-174、OM-197-MP-EC、ONTAK、PepTelベクターシステム、PLGマイクロ粒子、レシキモド、SRL172、ピロソーム及び他のウイルス様粒子、YF-17D、VEGFトラップ、R848、β-グルカン、Pam3Cys、サポニンに由来するAquila's QS21 stimulon (Aquila Biotech, Worcester, Mass., USA)、マイコバクテリア抽出物及び合成細菌細胞壁模倣物、及びRibix Detox. QuilまたはSuperfosなどの他の専売アジュバントを含むが、それらに限定されない。不完全フロインドまたはGM-CSFなどのアジュバントが、有用である。樹状細胞に特異的ないくつかの免疫学的アジュバント（例えば、MF59）及びそれらの調製物が、以前に記載されている（Dupuis M, et al., Cell Immunol. 1998; 186(1): 18-27; Allison AC; Dev Biol Stand. 1998; 92: 3-11）。また、サイトカインを使用することもできる。いくつかのサイトカインは、リンパ組織に対する樹状細胞の遊走への影響（例えば、TNF-α）、Tリンパ球に対する効率的な抗原提示細胞への樹状細胞の成熟の加速化（例えば、GM-CSF、IL-1、及びIL-4）（具体的にその全体が参照により本明細書に組み入れられる、米国特許第5,849,589号）、及び免疫アジュバントとしての作用（例えば、IL-12）に直接結び付けられている（Gabrilovich DI, et al., J Immunother Emphasis Tumor Immunol. 1996(6): 414-418）。

20

30

40

## 【0187】

CpG免疫刺激性オリゴヌクレオチドもまた、ワクチン設定においてアジュバントの効果を増強することが報告されている。TLR 7、TLR 8、及び／またはTLR 9に結合するRNAなどの他のTLR結合分子がまた、使用されてもよい。

## 【0188】

有用なアジュバントの他の例は、化学的に修飾されたCpG（例えば、CpR、Ide

50

ra)、Poly(I:C)(例えば、polyi:C12U)、非CpG細菌DNAまたはRNA、ならびに、治療的に及び／またはアジュバントとして作用し得る、シクロホスファミド、スニチニブ、ベバシズマブ、セブレックス、NCX-4016、シルデナフィル、タダラフィル、バルデナフィル、ソラフィニブ、XL-999、CP-547632、パゾパニブ、ZD2171、AZD2171、イピリムマブ、トレメリムマブ、及びSC58175などの免疫活性小分子及び抗体を含むが、それらに限定されない。アジュバント及び添加物の量及び濃度は、当業者が過度の実験なしで容易に決定することができる。追加的なアジュバントは、顆粒球マクロファージコロニー刺激因子(GM-CSF、サルグラモスチム)などのコロニー刺激因子を含む。

#### 【0189】

ワクチン組成物は、1種類よりも多い異なるアジュバントを含むことができる。さらに、治療用組成物は、上記の任意またはそれらの組み合わせを含む、任意のアジュバント物質を含むことができる。ワクチン及びアジュバントを、任意の適切な配列において、一緒にまたは別々に投与することもまた、企図される。

#### 【0190】

担体(または賦形剤)は、アジュバントから独立して存在することができる。担体の機能は、例えば、活性または免疫原性を増大させるため、安定性を与えるため、生物学的活性を増大させるため、または血清半減期を増大させるために、特に変異体の分子量を増大させることであり得る。さらに、担体は、T細胞に対してペプチドを提示するのを助けることができる。担体は、当業者に公知の任意の適している担体、例えば、タンパク質または抗原提示細胞であることができる。担体タンパク質は、キーホールリンペットヘモシアニン、トランスフェリンなどの血清タンパク質、ウシ血清アルブミン、ヒト血清アルブミン、サイログロブリンもしくはオボアルブミン、免疫グロブリン、またはインスリンなどのホルモン、またはパルミチン酸であることができるが、それらに限定されない。ヒトの免疫化のためには、担体は概して、ヒトに許容されかつ安全な、生理学的に許容される担体である。しかし、破傷風トキソイド及び／またはジフテリアトキソイドは、適している担体である。あるいは、担体は、デキストラン、例えばセファロースであることができる。

#### 【0191】

細胞傷害性T細胞(CTL)は、無傷の外来抗原自体よりも、MHC分子に結合したペプチドの形態において抗原を認識する。MHC分子自体は、抗原提示細胞の細胞表面に位置する。したがって、CTLの活性化は、ペプチド抗原、MHC分子、及びAPCの三量体複合体が存在する場合に可能である。対応して、ペプチドがCTLの活性化のために使用される場合だけではなく、追加的にそれぞれのMHC分子を有するAPCが添加される場合に、それは免疫応答を増強し得る。したがって、いくつかの実施形態において、ワクチン組成物は、追加的に、少なくとも1つの抗原提示細胞を含有する。

#### 【0192】

新生抗原はまた、ワクシニア、鶏痘、自己複製アルファウイルス、マラバウイルス、アデノウイルス(例えば、Tatsis et al., *Adenoviruses, Molecular Therapy* (2004) 10, 616-629を参照されたい)、または、第2、第3、もしくはハイブリッド第2／第3世代のレンチウイルス、及び特異的な細胞タイプもしくは受容体を標的とするように設計された任意の世代の組換えレンチウイルスを含むがそれらに限定されないレンチウイルス(例えば、Hu et al., *Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases*, *Immunol Rev.* (2011) 239 (1): 45-61、Sakuma et al., *Lentiviral vectors: basic to translational*, *Biochem J.* (2012) 443 (3): 603-18、Cooper et al., *Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promo*

10

20

30

40

50

ter, Nucl. Acids Res. (2015) 43 (1) : 682-690、Zufferey et al., Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, J. Virol. (1998) 72 (12) : 9873-9880を参照されたい)などの、ウイルスベクターベースのワクチンプラットフォームに含めることもできる。上述のウイルスベクターベースのワクチンプラットフォームのパッケージング能力に依存して、このアプローチは、1つ以上の新生抗原ペプチドをコードする1つ以上のヌクレオチド配列を送達することができる。配列は、非変異配列が隣接していてもよく、リンカーによって分離されていてもよく、または、細胞内区画を標的とする1つもしくは複数の配列が先行していてもよい(例えば、Gros et al., Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, Nat Med. (2016) 22 (4) : 433-8、Stronen et al., Targeting of cancer neoantigens with donor-derived T cell receptor repertoires, Science. (2016) 352 (6291) : 1337-41、Lu et al., Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions, Clin Cancer Res. (2014) 20 (13) : 3401-10を参照されたい)。宿主中への導入時に、感染した細胞は、新生抗原を発現し、それにより、ペプチドに対する宿主免疫(例えば、CTL)応答を惹起する。免疫化プロトコールにおいて有用なワクチンベクター及び方法は、例えば、米国特許第4,722,848号に記載されている。別のベクターは、BCG(カルメット・ゲラン桿菌)である。BCGベクターは、Stover et al. (Nature 351 : 456-460 (1991))に記載されている。新生抗原の治療的投与または免疫化に有用な、多種多様の他のワクチンベクター、例えば、チフス菌(*Salmonella typhi*)ベクターなどが、本明細書における記載から当業者に明らかであろう。

【0193】

#### IV. A. ワクチン設計及び製造のさらなる考慮事項

##### IV. A. 1. すべての腫瘍サブクローンをカバーするペプチドのセットの決定

すべての、または大部分の腫瘍サブクローンによって提示されるものを意味するトランカルペプチド(truncal peptide)が、ワクチン中への包含について優先される<sup>53</sup>。任意で、高い確率で提示されかつ免疫原性であることが予測されるトランカルペプチドがない場合、または、高い確率で提示されかつ免疫原性であることが予測されるトランカルペプチドの数が、追加的な非トランカルペプチドをワクチンに含めることができるほど少ない場合には、腫瘍サブクローンの数及び同一性を推定すること、及びワクチンによってカバーされる腫瘍サブクローンの数を最大化するようにペプチドを選ぶことによって、さらなるペプチドを優先順位付けすることができる<sup>54</sup>。

【0194】

##### IV. A. 2. 新生抗原の優先順位決定

上記の新生抗原フィルターのすべてを適用した後、ワクチン技術が対応できるよりも多くの候補新生抗原が、依然としてワクチン包含に利用可能である可能性がある。追加的に、新生抗原解析の種々の態様についての不確定度が残っている可能性があり、候補ワクチン新生抗原の様々な性状の間にトレードオフが存在する可能性がある。したがって、選択プロセスの各段階でのあらかじめ決定されたフィルターの代わりに、少なくとも以下の軸を有する空間に候補新生抗原を置き、積分アプローチを用いて選択を最適化する、積分多次元モデルを考えることができる。

1. 自己免疫または寛容のリスク(生殖細胞系列のリスク)(より低い自己免疫のリスクが、典型的に好ましい)

2. シークエンシングアーチファクトの確率（より低いアーチファクトの確率が、典型的に好ましい）
3. 免疫原性の確率（より高い免疫原性の確率が、典型的に好ましい）
4. 提示の確率（より高い提示の確率が、典型的に好ましい）
5. 遺伝子発現（より高い発現が、典型的に好ましい）
6. HLA遺伝子のカバレッジ（新生抗原のセットの提示に関与する、より多い数のHLA分子は、腫瘍が、HLA分子の下方制御または変異を介して免疫攻撃を回避する確率を低くする可能性がある）
7. HLAクラスのカバレッジ（HLA-I及びHLA-IIの両方をカバーすることで、治療応答の確率が高まり、腫瘍の免疫回避の確率が低くなる可能性がある）。

10

【0195】

さらに、場合によっては、新生抗原が患者の腫瘍のすべてまたは一部において喪失するかまたは不活性化されたHLAアレルによって提示されることが予測される場合には、これらの新生抗原のワクチン接種における優先順位を下げる（例えば除外）することができる。HLAアレルの喪失は、体細胞変異、ヘテロ接合性の喪失、または遺伝子座のホモ接合欠失のいずれかによって生じうる。HLAアレルの体細胞変異の検出方法は当該技術分野では周知のものである（例えば、Shukla et al., 2015）。体細胞LOH及びホモ接合欠失（HLA遺伝子座を含む）の検出方法についても同様に述べられている（Carter et al., 2012; McGranahan et al., 2017; Van Loo et al., 2010）。

20

【0196】

#### V. 治療及び製造方法

本明細書に開示する方法を用いて特定された複数の新生抗原などの1つ以上の新生抗原を対象に投与することにより、対象において腫瘍特異的な免疫応答を誘導し、腫瘍に対するワクチン接種を行い、対象のがんの症状を治療及び／または緩和する方法も提供される。

【0197】

いくつかの態様において、対象は、がんと診断されているか、またはがんを発症するリスクにある。対象は、ヒト、イヌ、ネコ、ウマ、または、腫瘍特異的な免疫応答が望ましい任意の動物であることができる。腫瘍は、乳、卵巣、前立腺、肺、腎臓、胃、結腸、精巣、頭頸部、脾臓、脳、黒色腫、及び他の組織器官の腫瘍などの、任意の固形腫瘍、ならびに、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、及びB細胞リンパ腫を含むリンパ腫及び白血病などの、血液腫瘍であることができる。

30

【0198】

新生抗原は、CTL応答を誘導するのに十分な量で投与することができる。

【0199】

新生抗原は、単独で、または他の治療用物質との組み合わせで投与することができる。治療用物質は、例えば、化学療法剤、放射線、または免疫療法である。特定のがんのための任意の適している治療的処置を、施すことができる。

【0200】

加えて、対象に、チェックポイント阻害因子などの抗免疫抑制性／免疫刺激性物質をさらに投与することができる。例えば、対象に、抗CTLA抗体または抗PD-1または抗PD-L1をさらに投与することができる。抗体によるCTLA-4またはPD-L1の遮断は、患者においてがん性細胞に対する免疫応答を増強することができる。特に、CTLA-4遮断は、ワクチン接種プロトコルを採用した場合に有効であることが示されている。

40

【0201】

ワクチン組成物に含まれるべき各新生抗原の最適量、及び最適投薬レジメンを、決定することができる。例えば、新生抗原またはその変異体は、静脈内（i.v.）注射、皮下（s.c.）注射、皮内（i.d.）注射、腹腔内（i.p.）注射、筋肉内（i.m.）

50

）注射のために調製することができる。注射の方法は、s . c . 、 i . d . 、 i . p . 、 i . m . 、及び i . v . を含む。DNAまたはRNA注射の方法は、i . d . 、 i . m . 、 s . c . 、 i . p . 、及び i . v . を含む。ワクチン組成物の投与の他の方法は、当業者に公知である。

#### 【0202】

ワクチンは、組成物中に存在する新生抗原の選択、数、及び／または量が、組織、がん、及び／または患者に特異的であるように編集することができる。例として、ペプチドの厳密な選択は、所定の組織における親タンパク質の発現パターンによって手引きされ得る。選択は、がんの特異的なタイプ、疾患の状態、より早期の処置レジメン、患者の免疫状態、及び当然、患者のHLAハロタイプに依存し得る。さらに、ワクチンは、特定の患者の個人的な必要にしたがって、個別化された構成要素を含有することができる。例は、特定の患者における新生抗原の発現にしたがって新生抗原の選択を変えること、または、処置の第1のラウンドまたはスキームの後の二次的処置についての調整を含む。

#### 【0203】

がんのためのワクチンとして使用されるべき組成物について、正常組織において多量に発現している類似した正常な自己ペプチドを有する新生抗原は、本明細書に記載した組成物において、避けられるか、または少量で存在することができる。他方で、患者の腫瘍が、多量のある特定の新生抗原を発現することが公知である場合、このがんの処置のためのそれぞれの薬学的組成物は、多量に存在することができ、及び／または、この特定の新生抗原もしくはこの新生抗原の経路に特異的な1種類よりも多い新生抗原を含めることができる。

#### 【0204】

新生抗原を含む組成物を、既にかんを患っている個体に投与することができる。治療的適用において、組成物は、腫瘍抗原に対する有効なCTL応答を惹起し、かつ、症候及び／または合併症を治療するかまたは少なくとも部分的に停止するのに十分な量で、患者に投与される。これを達成するのに妥当な量を、「治療的有效用量」として定義する。この用途のために有効な量は、例えば、組成物、投与の様式、処置される疾患の病期及び重症度、患者の体重及び健康の全身状態、ならびに処方医の判断に依存するであろう。組成物は、概して、重篤な疾患状態、すなわち、命に関わるか、または潜在的に命に関わる状況、特にがんが転移している場合に使用できることを、心に留めるべきである。そのような例において、外来性物質の最小化、及び新生抗原の相対的な非毒性の性質を考慮して、実質的過剰量のこれらの組成物を投与することが、可能であり、かつ処置する医師が望ましいとすることができる。

#### 【0205】

治療的用途のために、投与は、腫瘍の検出または外科的除去時に始めることができる。これに、少なくとも症候が実質的に減ずるまで、及びその後ある期間にわたって、ブースト用量が続く。

#### 【0206】

治療的処置のための薬学的組成物（例えば、ワクチン組成物）は、非経口、局部、経鼻、経口、または局所投与について意図される。薬学的組成物は、非経口的に、例えば、静脈内、皮下、皮内、または筋肉内に投与することができる。組成物は、腫瘍に対する局所免疫応答を誘導するために、外科的切除の部位に投与することができる。新生抗原の溶液を含む非経口投与用の組成物を、本明細書に開示し、ワクチン組成物は、許容される担体、例えば、水性担体に溶解または懸濁される。様々な水性担体、例えば、水、緩衝水、0.9%食塩水、0.3%グリシン、ヒアルロン酸などを使用することができる。これらの組成物は、従来の周知の滅菌技法によって滅菌することができ、または滅菌濾過することができる。結果として生じた水溶液を、そのまま使用のためにパッケージングするか、または凍結乾燥することができ、凍結乾燥調製物は、投与前に滅菌溶液と組み合わせられる。組成物は、pH調整剤及び緩衝剤、等張化剤、湿潤剤など、例えば、酢酸ナトリウム、乳酸ナトリウム、塩化ナトリウム、塩化カリウム、塩化カルシウム、ソルビタンモノラウ

ラート、トリエタノールアミンオレアートなどのような、生理学的条件に近づけるために必要とされる、薬学的に許容される補助物質を含有してもよい。

【0207】

新生抗原はまた、それらをリンパ組織などの特定の細胞組織にターゲティングする、リポソームを介して投与することもできる。リポソームはまた、半減期を増大させるのにも有用である。リポソームは、エマルジョン、フォーム、ミセル、不溶性単層、液晶、リン脂質分散物、ラメラ層などを含む。これらの調製物において、送達されるべき新生抗原は、単独で、または、CD45抗原に結合するモノクローナル抗体などの、例えば、リンパ系細胞の間で優性な受容体に結合する分子、または他の治療用組成物もしくは免疫原性組成物と共に、リポソームの一部として組み込まれる。したがって、所望の新生抗原で満たされたリポソームは、リンパ系細胞の部位へ方向付けられることができ、そこで、リポソームは次いで、選択された治療用／免疫原性組成物を送達する。リポソームは、概して、中性及び負電荷を有するリン脂質、及びコレステロールなどのステロールを含む、標準的な小胞形成脂質から形成され得る。脂質の選択は、概して、例えば、リポソームサイズ、酸不安定性、及び血流におけるリポソームの安定性の考慮により手引きされる。例えば、Szoka et al., Ann. Rev. Biophys. Bioeng. 9; 467 (1980)、米国特許第4,235,871号、第4,501,728号、第4,501,728号、第4,837,028号、及び第5,019,369号に記載されているように、様々な方法を、リポソームを調製するために利用可能である。

10

【0208】

免疫細胞へのターゲティングのために、リポソーム中に組み込まれるべきリガンドは、例えば、所望の免疫系細胞の細胞表面決定基に特異的な抗体またはその断片を含むことができる。リポソーム懸濁液は、とりわけ、投与の様式、送達されるペプチド、及び処置される疾患の病期にしたがって変動する用量で、静脈内、局所、局部などに投与することができる。

20

【0209】

治療目的または免疫化目的で、本明細書に記載したペプチド、及び任意でペプチドの1つ以上をコードする核酸をまた、患者に投与することもできる。数多くの方法が、核酸を患者に送達するために好都合に使用される。例として、核酸を、「裸のDNA」として直接送達することができる。このアプローチは、例として、Wolff et al., Science 247:1465-1468 (1990)、ならびに米国特許第5,580,859号及び第5,589,466号に記載されている。核酸はまた、例として、米国特許第5,204,253号に記載されているような弾道送達を用いて投与することもできる。単にDNAからなる粒子を、投与することができる。あるいは、DNAを、金粒子などの粒子に接着させることができる。核酸配列を送達するためのアプローチは、エレクトロポレーションを伴うかまたは伴わない、ウイルスベクター、mRNAベクター、及びDNAベクターを含むことができる。

30

【0210】

核酸はまた、カチオン性脂質などのカチオン性化合物に複合体化させて送達することもできる。脂質媒介性遺伝子送達法は、例として、9618372WOAWO 96/18372; 9324640WOAWO 93/24640; Mannino & Gould-Fogerite, BioTechniques 6(7): 682-691 (1988); 米国特許第5,279,833号 Rose、米国特許第5,279,833号; 9106309WOAWO 91/06309; 及び Felgner et al., Proc. Natl. Acad. Sci. USA 84: 7413-7414 (1987) に記載されている。

40

【0211】

新生抗原はまた、ワクシニア、鶏痘、自己複製アルファウイルス、マラバウイルス、アデノウイルス（例えば、Tatsis et al., Adenoviruses, Molecular Therapy (2004) 10, 616-629を参照された

50

い)、または、第2、第3、もしくはハイブリッド第2／第3世代のレンチウイルス、及び特異的な細胞タイプもしくは受容体を標的とするように設計された任意の世代の組換えレンチウイルスを含むがそれらに限定されないレンチウイルス(例えば、Hu et al., Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, Immunol Rev. (2011) 239 (1): 45-61、Sakuma et al., Lentiviral vectors: basic to translational, Biochem J. (2012) 443 (3): 603-18、Cooper et al., Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter, Nucl. Acids Res. (2015) 43 (1): 682-690、Zufferey et al., Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, J. Virol. (1998) 72 (12): 9873-9880を参照されたい)などの、ウイルスベクターベースのワクチンプラットフォームに含めることもできる。上述のウイルスベクターベースのワクチンプラットフォームのパッケージング能力に依存して、このアプローチは、1つ以上の新生抗原ペプチドをコードする1つ以上のヌクレオチド配列を送達することができる。配列は、非変異配列が隣接していてもよく、リンカーによって分離されていてもよく、または、細胞内区画を標的とする1つもしくは複数の配列が先行していてもよい(例えば、Gros et al., Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, Nat Med. (2016) 22 (4): 433-8、Stronen et al., Targeting of cancer neoantigens with donor-derived T cell receptor repertoires, Science. (2016) 352 (6291): 1337-41、Lu et al., Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions, Clin Cancer Res. (2014) 20 (13): 3401-10を参照されたい)。宿主中への導入時に、感染した細胞は、新生抗原を発現し、それにより、ペプチドに対する宿主免疫(例えば、CTL)応答を惹起する。免疫化プロトコールにおいて有用なワクシニアベクター及び方法は、例えば、米国特許第4,722,848号に記載されている。別のベクターは、BCG(カルメット・ゲラン桿菌)である。BCGベクターは、Stover et al. (Nature 351: 456-460 (1991))に記載されている。新生抗原の治療的投与または免疫化に有用な、多種多様の他のワクチンベクター、例えば、チフス菌ベクターなどが、本明細書における記載から当業者に明らかであろう。

#### 【0212】

核酸を投与する手段は、1つ以上のエピトープをコードするミニ遺伝子構築物を使用する。ヒト細胞における発現のための、選択されたCTLエピトープをコードするDNA配列(ミニ遺伝子)を作製するために、エピトープのアミノ酸配列を逆翻訳する。各アミノ酸に対するコドン選択を手引きするために、ヒトコドン使用頻度表を使用する。これらのエピトープをコードするDNA配列を、直接隣り合わせて、連続的なポリペプチド配列を作製する。発現及び／または免疫原性を最適化するために、追加の要素を、ミニ遺伝子設計中に組み入れることができる。逆翻訳して、ミニ遺伝子配列に含めることができるアミノ酸配列の例は、ヘルパーTリンパ球エピトープ、リーダー(シグナル)配列、及び小胞体保持シグナルを含む。加えて、CTLエピトープのMHC提示は、CTLエピトープに近接した合成の(例えば、ポリアラニン)または天然に存在するフランキング配列を含む

ことによって、改善することができる。ミニ遺伝子配列は、ミニ遺伝子のプラス鎖及びマイナス鎖をコードするオリゴヌクレオチドをアセンブルすることによって、DNAに変換される。オーバーラップするオリゴヌクレオチド（30～100塩基長）を、周知の技法を用いて適切な条件下で、合成し、リン酸化し、精製し、アニーリングする。オリゴヌクレオチドの端は、T4DNAリガーゼを用いて連結する。CTLエピトープポリペプチドをコードするこの合成ミニ遺伝子を、次いで、望ましい発現ベクター中にクローニングすることができる。

#### 【0213】

精製プラスミドDNAは、様々な製剤を用いて、注射のために調製することができる。これらのうちでもっとも単純なものは、滅菌リン酸緩衝食塩水（PBS）における凍結乾燥DNAの再構成である。様々な方法が記載されており、新たな技法が利用可能になり得る。上記で言及したように、核酸は、カチオン性脂質で好都合に製剤化される。加えて、糖脂質、融合性リポソーム、ペプチド、及び保護的、相互作用の、非縮合性（PINC）と集合的に呼ばれる化合物もまた、精製プラスミドDNAと複合体化させて、安定性、筋肉内分散、または特異的な器官もしくは細胞タイプへの輸送などの変数に影響を及ぼすことができる。

#### 【0214】

また、本明細書に開示する方法のステップを行うこと；及び、多数の新生抗原または多数の新生抗原のサブセットを含む腫瘍ワクチンを生産するステップを含む、腫瘍ワクチンを製造する方法も、本明細書に開示する。

#### 【0215】

本明細書に開示する新生抗原は、当技術分野において公知の方法を用いて製造することができる。例えば、本明細書に開示する新生抗原またはベクター（例えば、1つ以上の新生抗原をコードする少なくとも1つの配列を含むベクター）を生産する方法は、新生抗原またはベクターを発現するのに適している条件下で宿主細胞を培養するステップであって、宿主細胞が、新生抗原またはベクターをコードする少なくとも1つのポリヌクレオチドを含むステップ、及び、新生抗原またはベクターを精製するステップを含むことができる。標準的な精製法は、クロマトグラフィー技法、電気泳動技法、免疫学的技法、沈降技法、透析技法、濾過技法、濃縮技法、及びクロマトフォーカシング技法を含む。

#### 【0216】

宿主細胞は、チャイニーズハムスター卵巣（CHO）細胞、NS0細胞、酵母、またはHEK293細胞を含むことができる。宿主細胞は、本明細書に開示する新生抗原またはベクターをコードする少なくとも1つの核酸配列を含む、1つ以上のポリヌクレオチドで形質転換することができ、任意で、単離されたポリヌクレオチドは、新生抗原またはベクターをコードする少なくとも1つの核酸配列に機能的に連結されたプロモーター配列をさらに含む。ある特定の実施形態において、単離されたポリヌクレオチドは、cDNAであることができる。

#### 【0217】

### VI. 新生抗原の特定

#### VI. A. 新生抗原候補の特定

腫瘍及び正常のエクソーム及びトランスクリプトームのNGS解析のための研究法を、新生抗原の特定のスペースに記載し、適用している<sup>6, 14, 15</sup>。下記の例は、臨床設定における新生抗原の特定について、より大きな感度及び特異性のためのある特定の最適化を考慮している。これらの最適化は、実験室プロセスに関連するもの及びNGSデータ解析に関連するものの、2つの区域にグループ化することができる。

#### 【0218】

#### VI. A. 1. 実験室プロセスの最適化

本明細書に提示したプロセスの改善は、標的とされるがんパネルにおける信頼できるがんドライバー遺伝子の評価について開発された概念<sup>16</sup>を、新生抗原の特定のために必要な全エクソーム設定及び全トランスクリプトーム設定に拡大することによって、低い腫瘍

10

20

30

40

50

含量及び少ない体積の臨床標本からの高精度の新生抗原の発見における難題に対処する。  
具体的には、これらの改善は、以下を含む：

1. 低い腫瘍含量またはサブクローン状態のいずれかにより、低い変異体アレル頻度で存在する変異を検出するための、腫瘍エクソームにわたる深い（500xよりも大きい）固有の平均カバレッジのターゲティング。

2. 可能性のある新生抗原の見逃しが最も少ないように、100x未満でカバーされる塩基が5%未満である、例として、

a. 個々のプローブQCを有するDNAベースの捕捉プローブの使用<sup>17</sup>

b. 十分にカバーされていない領域についての追加的なベイトの包含

3. 可能性のある新生抗原が体細胞性／生殖細胞系列ステータスについて分類されていないままである（したがってTSNAとして使用可能ではない）ことが最も少ないように、20x未満でカバーされる塩基が5%未満である、正常エクソームにわたる均一カバレッジのターゲティング。

4. 必要とされるシーケンシングの総量を最小化するために、配列捕捉プローブは、非コードRNAは新生抗原を生じることができないことから、遺伝子のコード領域のみについて設計される。追加的な最適化は、以下を含む：

a. GCリッチであり、標準的なエクソームシーケンシングでは十分に捕捉されないHLA遺伝子についての補充的プローブ<sup>18</sup>。

b. 不十分な発現、プロテアソームによる最適に満たない消化、または異例の配列特性などの要因により、候補新生抗原を少ししかまたは全く生成しないと予測される遺伝子の排除。

5. 変異検出、遺伝子及びスプライス変異体（「アイソフォーム」）発現の定量、ならびに融合物検出を可能にするために、腫瘍RNAが同様に、高深度（100Mリードよりも大きい）でシーケンシングされる。FFPE試料由来のRNAは、DNAにおいてエクソームを捕捉するために使用されるのと同じまたは類似したプローブで、プローブベース濃縮<sup>19</sup>を用いて抽出される。

【0219】

#### V I . A . 2 . N G S データ解析の最適化

解析法の改善は、一般的な研究変異コーリングアプローチの最適に満たない感度及び特異性に対処し、具体的には、臨床設定における新生抗原の特定のために関連するカスタマイズ化を考慮する。これらは、以下を含む：

1. アラインメントのための、HG38参照ヒトゲノムまたはより後のバージョンの使用（それが、以前のゲノムリリースとは対照的に、集団多型をより良好に反映する複数のMHC領域アセンブリーを含有するため）。

2. 様々なプログラム<sup>5</sup>からの結果をマージすることによる、単一変異コーラー20の限界の克服。

a. 単一ヌクレオチド変異及び挿入欠失は、以下を含む一連のツールで、腫瘍DNA、腫瘍RNA、及び正常DNAから検出される：Strelka<sup>21</sup>及びMutect<sup>22</sup>などの、腫瘍及び正常DNAの比較に基づくプログラム；ならびに、低純度の試料において特に有利である<sup>23</sup>、UNCeqRなどの、腫瘍DNA、腫瘍RNA、及び正常DNAを組み入れるプログラム。

b. 挿入欠失は、Strelka及びABRA<sup>24</sup>などの、局所リアセンブリーを行うプログラムで決定される。

c. 構造的再編成は、Pindel<sup>25</sup>またはBreakseq<sup>26</sup>などの専用のツールを用いて決定される。

3. 試料スワップを検出して阻止するために、同じ患者についての試料由来の変異コールが、選ばれた数の多型部位で比較される。

4. 例として、以下による、人工的コールの広範囲のフィルタリングが行われる：

a. 潜在的に、低いカバレッジの例においては緩やかな検出パラメータで、及び挿入欠失の例においては許容的な近接基準での、正常DNAにおいて見出される変異の除去。

b. 低いマッピング品質または低い塩基品質による変異の除去<sup>27</sup>。

c. たとえ対応する正常において観察されないとしても、再出現するシーケンシングアーチファクトから生じる変異の除去<sup>27</sup>。例は、主として1本の鎖上に検出される変異を含む。

d. 無関連のコントロールのセットにおいて検出される変異の除去<sup>27</sup>。

5. seq2HLA<sup>28</sup>、ATHLATES<sup>29</sup>、またはOptiType<sup>30</sup>のうちの1つを使用する、かつまた、エクソーム及びRNAシーケンシングデータを組み合わせる<sup>28</sup>、正常エクソームからの正確なHLAコーリング。追加的な潜在的最適化は、ロングリードDNAシーケンシングなどの、HLAタイピングのための専用アッセイの採用<sup>30</sup>、または、RNA断片を連結して連続性を保持するための方法の適応<sup>31</sup>を含む。

6. 腫瘍特異的のスプライス変異体から生じた新生ORFの堅牢な検出は、CLASS<sup>32</sup>、Bayesemblem<sup>33</sup>、StringTie<sup>34</sup>、またはそのリファレンスガイドモードにおける類似したプログラム（すなわち、各実験からそれらの全体の転写産物を再作製するように試みるよりもむしろ、公知の転写産物構造を用いる）を用いて、RNA-seqデータから転写産物をアセンブルすることによって、行われる。Cufflinks<sup>35</sup>が、この目的で一般的に使用されるが、それは頻繁に、信じ難いほど多数のスプライス変異体を産生し、それらの多くは、完全長遺伝子よりもはるかに短く、単純なポジティブコントロールをリカバーすることができない場合がある。コード配列及び潜在的なナンセンス変異依存分解機構は、変異体配列を再導入した、SpliceR<sup>36</sup>及びMAMBA<sup>37</sup>などのツールで決定される。遺伝子発現は、Cufflinks<sup>35</sup>またはExpress（Roberts and Pachter, 2013）などのツールで決定される。野生型及び変異体特異的な発現カウント及び／または相対レベルは、ASE<sup>38</sup>またはHTSeq<sup>39</sup>などの、これらの目的で開発されたツールで決定される。潜在的なフィルタリング段階は、以下を含む：

a. 不十分に発現されていると考えられる候補新生ORFの除去。

b. ナンセンス変異依存分解機構（NMD）を引き起こすと予測される候補新生ORFの除去。

7. 腫瘍特異的と直接検証することができない、RNAにおいてのみ観察される候補新生抗原（例えば、新生ORF）は、例として以下を考慮することにより、追加的なパラメータにしたがって、腫瘍特異的である可能性が高いとして分類される：

a. 腫瘍DNAのみのシス作用性フレームシフトまたはスプライス部位変異の支持の存在。

b. スプライシング因子における腫瘍DNAのみのトランス作用性変異の確証の存在。例として、R625変異体SF3B1での3つの独立して公開された実験において、最も差次的にスプライシングを呈する遺伝子は、1つの実験がブドウ膜黒色腫患者を検討し<sup>40</sup>、第2の実験がブドウ膜黒色腫細胞株を検討し<sup>41</sup>、及び第3の実験が乳がん患者を検討した<sup>42</sup>にもかかわらず、一致していた。

c. 新規のスプライシングアイソフォームについては、RNASeqデータにおける「新規の」スプライス-ジャンクションリードの確証の存在。

d. 新規の再編成については、正常DNAには存在しない腫瘍DNAにおけるエクソン近傍リードの確証の存在。

e. GTEx<sup>43</sup>などの遺伝子発現大要からの欠如（すなわち、生殖細胞系列起源の可能性をより低くする）。

8. アラインメント及びアノテーションベースのエラー及びアーチファクトを直接避けるために、アセンブルされたDNAの腫瘍及び正常リード（またはそのようなリード由来のkマー）を比較することによる、参照ゲノムアラインメントベースの解析の補完（例えば、生殖細胞系列変異またはリピートコンテキスト挿入欠失の近くに生じる体細胞性変異について）。

【0220】

ポリアデニル化RNAを有する試料において、RNA-seqデータにおけるウイルス

10

20

30

40

50

RNA及び微生物RNAの存在は、患者の応答を予測し得る追加的因子の特定に向かって、RNA COMPASS 4.4または類似した方法を用いて評価される。

#### 【0221】

#### V I . B . H L A ペプチドの単離及び検出

H L A ペプチド分子の単離は、組織試料の溶解及び可溶化後に、古典的な免疫沈降（I P）法を用いて行った<sup>55~58</sup>。清澄化した溶解物を、H L A 特異的 I P に使用した。

#### 【0222】

免疫沈降は、抗体がH L A 分子に特異的である、ビーズにカップリングした抗体を用いて行った。汎クラス I H L A 免疫沈降のためには、汎クラス I C R 抗体を使用し、クラス I I H L A - D R のためには、H L A - D R 抗体を使用する。抗体を、一晚インキュベーション中に、N H S - セファロースビーズに共有結合で付着させる。共有結合性の付着後、ビーズを洗浄して、I P のために等分した<sup>59、60</sup>。ビーズに共有結合されていない抗体を用いて免疫沈降を行うこともできる。一般的に、これは、抗体をカラムに保持するために P r o t e i n A 及び／または P r o t e i n G でコーティングしたセファロースまたは磁気ビーズを使用して行われる。M H C / ペプチド複合体を選択的に濃縮するために使用することができるいくつかの抗体を下記に示す。

| 抗体名   | 特異性                     |
|-------|-------------------------|
| W6/32 | クラス I HLA-A, B, C       |
| L243  | クラス II - HLA-DR         |
| Tu36  | クラス II - HLA-DR         |
| LN3   | クラス II - HLA-DR         |
| Tu39  | クラス II - HLA-DR, DP, DQ |

#### 【0223】

清澄化した組織溶解物を、免疫沈降のために抗体ビーズに添加する。免疫沈降後、ビーズを溶解物から除去し、追加的な I P を含む追加的な実験のために、溶解物を保存する。標準的な技法を用いて、I P ビーズを洗浄して非特異的結合を除去し、H L A / ペプチド複合体をビーズから溶出する。分子量スピンカラムまたは C 1 8 分画を用いて、タンパク質構成要素をペプチドから除去する。結果として生じたペプチドを、S p e e d V a c 蒸発によって乾燥させ、いくつかの場合には、M S 解析の前に - 2 0 ° C で保存する。

#### 【0224】

乾燥したペプチドを、逆相クロマトグラフィーに適している H P L C 緩衝液において再構成し、F u s i o n L u m o s 質量分析計（T h e r m o）における勾配溶出のために、C - 1 8 マイクロキャピラリー H P L C カラム上にロードする。ペプチド質量／電荷（m / z）の M S 1 スペクトルを、O r b i t r a p 検出器において高解像度で収集し、その後、M S 2 低解像度スキャンを、選択イオンの H C D フラグメンテーション後にイオントラップ検出器において収集した。追加的に、M S 2 スペクトルは、C I D もしくは E T D フラグメンテーション法、または、ペプチドのより大きなアミノ酸カバレッジを獲得するための3つの技法の任意の組み合わせのいずれかを用いて、取得することができる。M S 2 スペクトルはまた、O r b i t r a p 検出器において高解像度質量精度で測定することもできる。

#### 【0225】

各解析由来の M S 2 スペクトルを、C o m e t <sup>61、62</sup>を用いてタンパク質データベースに対して検索し、ペプチド特定を、P e r c o l a t o r <sup>63~65</sup>を用いてスコア化する。P E A K S s t u d i o（B i o i n f o r m a t i c s S o l u t i o n s I n c .）及び他のサーチエンジンを用いてさらなるシークエンシングを行うか、またはスペクトルマッチング及びデノボシークエンシング<sup>75</sup>を含むシークエンシング法を用いる

ことができる。

【0226】

#### V I . B . 1 . 総合的 H L A ペプチドシーケンシングのための M S 検出限界の研究

ペプチド Y V Y V A D V A A K (配列番号 1) を用いて、何が検出の限界かを、L C カラム上にロードした様々な量のペプチドを用いて決定した。試験したペプチドの量は、1 p m o l、100 f m o l、10 f m o l、1 f m o l、及び 100 a m o l であった。(表 1) 結果を図 1 F に示す。これらの結果は、検出の最低限界 (L o D) がアトモルの範囲 ( $10^{-18}$ ) にあること、ダイナミックレンジが 5 桁に及ぶこと、及び、シグナル対ノイズが、低いフェムトモル範囲 ( $10^{-15}$ ) でシーケンシングに十分であるように見えることを示す。

| ペプチド m/z | カラムへのロード    | 1e9個の細胞中のコピー／細胞 |
|----------|-------------|-----------------|
| 566.830  | 1 p m o l   | 600             |
| 562.823  | 100 f m o l | 60              |
| 559.816  | 10 f m o l  | 6               |
| 556.810  | 1 f m o l   | 0.6             |
| 553.802  | 100 a m o l | 0.06            |

【0227】

#### V I I . 提示モデル

##### V I I . A . システムの概要

図 2 A は、1つの実施形態にしたがう、患者におけるペプチド提示の尤度を特定するための環境 100 の概要である。環境 100 は、それ自体が提示情報記憶装置 165 を含む提示特定システム 160 を導入するコンテキストを提供する。

【0228】

提示特定システム 160 は、図 30 に関して下記で議論されるようなコンピュータ計算システムにおいて具現化された、1つまたはコンピュータモデルであり、M H C アレルのセットに関連するペプチド配列を受け取り、ペプチド配列が、関連する M H C アレルのセットの 1 つ以上によって提示されるであろう尤度を決定する。提示特定システム 160 はクラス I 及びクラス I I M H C アレルの両方に適用することができる。これは、様々なコンテキストにおいて有用である。提示特定システム 160 の 1 つの具体的な用途の例は、患者 110 の腫瘍細胞由来の M H C アレルのセットに関連する候補新生抗原のヌクレオチド配列を受け取り、候補新生抗原が、腫瘍に関連する M H C アレルの 1 つ以上によって提示され、及び／または患者 110 の免疫系において免疫原性応答を誘導するであろう尤度を決定することができることである。システム 160 によって決定された際に高い尤度を有するそれらの候補新生抗原を、ワクチン 118 における包含のために選択することができ、そのような抗腫瘍免疫応答が、腫瘍細胞を提供する患者 110 の免疫系から誘発され得る。さらに、高い提示尤度を有する候補新生抗原に対する反応性を有する T C R を有する T 細胞を T 細胞療法で使用するために作製することが可能であり、これにより、患者 110 の免疫系から抗腫瘍免疫応答も誘発される。

【0229】

提示特定システム 160 は、1つ以上の提示モデルを通して提示尤度を決定する。具体的には、提示モデルは、所定のペプチド配列が、関連する M H C アレルのセットについて提示されるかどうかの尤度を生成し、尤度は、記憶装置 165 に保存された提示情報に基づいて生成される。例えば、提示モデルは、ペプチド配列「Y V Y V A D V A A K (配列番号 1)」が、試料の細胞表面上のアレルのセット H L A - A \* 0 2 : 0 1、H L A - A

\*03:01、HLA-B\*07:02、HLA-B\*08:03、HLA-C\*01:04について提示されるかどうかの尤度を生成し得る。提示情報165は、MHCアレルによってペプチドが提示されるようにこれらのペプチドが様々なタイプのMHCアレルに結合するかどうかについての情報を含み、これは、モデルにおいて、ペプチド配列におけるアミノ酸の位置に応じて決定される。提示モデルは、提示情報165に基づいて、認識されていないペプチド配列が、MHCアレルの関連するセットと結合して提示されるかどうかを予測することができる。上記に述べたように、提示モデルはクラスI及びクラスII MHCアレルの両方に適用することができる。

【0230】

#### VII.B. 提示情報

図2は、1つの実施形態にしたがう、提示情報を取得する方法を説明する。提示情報165は、2つの一般的部類の情報：アレル相互作用情報及びアレル非相互作用情報を含む。アレル相互作用情報は、MHCアレルのタイプに依存する、ペプチド配列の提示に影響を及ぼす情報を含む。アレル非相互作用情報は、MHCアレルのタイプに非依存的な、ペプチド配列の提示に影響を及ぼす情報を含む。

【0231】

#### VII.B.1. アレル相互作用情報

アレル相互作用情報は、主として、ヒト、マウスなど由来の1つ以上の特定されたMHC分子によって提示されていることが公知である、特定されたペプチド配列を含む。注目すべきことに、これは、腫瘍試料から取得されたデータを含んでもよく、または含まなくてもよい。提示されたペプチド配列は、単一のMHCアレルを発現する細胞から特定されてもよい。この例において、提示されたペプチド配列は、概して、あらかじめ決定されたMHCアレルを発現するように操作されてその後合成タンパク質に曝露された単一アレル細胞株から収集される。MHCアレル上に提示されたペプチドは、酸溶出などの技法によって単離され、質量分析により特定される。図2Bは、あらかじめ決定されたMHCアレルHLA-DRB1\*12:01上に提示された例示的なペプチドYEMFNDKSKRAPPDDKMF（配列番号2）が単離され、質量分析により特定される、この例を示す。この状況においては、ペプチドが、単一のあらかじめ決定されたMHCタンパク質を発現するように操作された細胞を通して特定されるため、提示されたペプチドとそれが結合したMHCタンパク質との間の直接の関連が、決定的に既知である。

【0232】

提示されたペプチド配列はまた、複数のMHCアレルを発現する細胞から収集されてもよい。典型的にヒトにおいては、6種類の異なるタイプのMHC I分子及び最大で12種類の異なるタイプのMHC II分子が細胞で発現している。そのような提示されたペプチド配列は、複数のあらかじめ決定されたMHCアレルを発現するように操作されている複数アレル細胞株から特定されてもよい。そのような提示されたペプチド配列はまた、正常組織試料または腫瘍組織試料のいずれかの、組織試料から特定されてもよい。この例において特に、MHC分子は、正常組織または腫瘍組織から免疫沈降させることができる。複数のMHCアレル上に提示されたペプチドは、同様に、酸溶出などの技法によって単離され、質量分析により特定されることができる。図2Cは、6種類の例示的なペプチドYEMFNDKSF（配列番号3）、HROEIFSHDFJ（配列番号4）、FJIEJFOESS（配列番号5）、NEIOREIREI（配列番号6）、JFKSIFEMMSJDSUIFLKSJFIEIFJ（配列番号7）、及びKNFLENFIESOFI（配列番号8）が、特定されたクラスI MHCアレルHLA-A\*01:01、HLA-A\*02:01、HLA-B\*07:02、HLA-B\*08:01、及びクラスII MHCアレルHLA-DRB1\*10:01、HLA-DRB1:11:01上に提示されており、単離され、質量分析により特定される、この例を示す。単一アレル細胞株とは対照的に、結合したペプチドが、特定される前のMHC分子から単離されるため、提示されたペプチドとそれが結合したMHCタンパク質との間の直接の関連は、未知である可能性がある。

10

20

30

40

50

## 【0233】

アレル相互作用情報はまた、ペプチド-MHC分子複合体の濃度、及びペプチドのイオン化効率の両方に依存する、質量分析イオン電流も含むことができる。イオン化効率は、配列依存性様式で、ペプチドごとに変動する。概して、イオン効率は、およそ2桁にわたってペプチドごとに変動し、他方、ペプチド-MHC複合体の濃度は、それよりも大きい範囲にわたって変動する。

## 【0234】

アレル相互作用情報はまた、所定のMHCアレルと所定のペプチドとの間の結合親和性の測定値または予測値も含むことができる。1つ以上の親和性モデルが、そのような予測値を生成することができる(72, 73, 74)。例えば、図1Dに示した例に戻ると、提示情報165は、ペプチドYEMFNDKSF(配列番号3)とクラスIIアレルHLA-A\*01:01との間の1000nMの結合親和性予測値を含み得る。IC50>1000nmであるペプチドはわずかし、MHCによって提示されず、より低いIC50値が、提示の確率を増大させる。提示情報165は、ペプチドKNFLENFIESOFI(配列番号8)とクラスIIアレルHLA-DRB1:11:01との間の結合親和性予測値を含み得る。

10

## 【0235】

アレル相互作用情報はまた、MHC複合体の安定性の測定値または予測値も含むことができる。1つ以上の安定性モデルが、そのような予測値を生成することができる。より安定なペプチド-MHC複合体(すなわち、より長い半減期を有する複合体)は、腫瘍細胞上、及びワクチン抗原に遭遇する抗原提示細胞上に高コピー数で提示される可能性がより高い。例えば、図2Cに示した例に戻ると、提示情報165は、クラスI分子HLA-A\*01:01について1時間の半減期の安定性予測値を含み得る。提示情報165はクラスII分子HLA-DRB1:11:01の半減期の安定性予測値も含み得る。

20

## 【0236】

アレル相互作用情報はまた、ペプチド-MHC複合体の形成反応の、測定されたかまたは予測された速度も含むことができる。より速い速度で形成する複合体は、高濃度で細胞表面上に提示される可能性がより高い。

## 【0237】

アレル相互作用情報はまた、ペプチドの配列及び長さも含むことができる。MHCクラスI分子は典型的に、8~15ペプチドの長さを有するペプチドを提示することを好む。提示されたペプチドの60~80%は、長さ9を有する。MHCクラスII分子は一般的にペプチド6~30個の長さを有するペプチドを提示する傾向にある。

30

## 【0238】

アレル相互作用情報はまた、新生抗原コード化ペプチド上のキナーゼ配列モチーフの存在、及び新生抗原コード化ペプチド上の特異的な翻訳後修飾の有無も含むことができる。キナーゼモチーフの存在は、MHC結合を増強または干渉し得る、翻訳後修飾の確率に影響を及ぼす。

## 【0239】

アレル相互作用情報はまた、(RNA seq、質量分析、または他の方法によって測定されたかまたは予測された際の)翻訳後修飾のプロセスに関与するタンパク質、例えば、キナーゼの発現または活性レベルも含むことができる。

40

## 【0240】

アレル相互作用情報はまた、質量分析プロテオミクスまたは他の手段によって評価された際の、特定のMHCアレルを発現する他の個体由来の細胞における、類似した配列を有するペプチドの提示の確率も含むことができる。

## 【0241】

アレル相互作用情報はまた、(例えば、RNA-seqまたは質量分析によって測定された際の)問題の個体における特定のMHCアレルの発現レベルも含むことができる。高レベルで発現しているMHCアレルに最も強く結合するペプチドは、低レベルで発現して

50

いるMHCアレルに最も強く結合するペプチドよりも、提示される可能性がより高い。

【0242】

アレル相互作用情報はまた、特定のMHCアレルを発現する他の個体における、特定のMHCアレルによる提示の、全体的な新生抗原コード化ペプチド配列非依存的確率も含むことができる。

【0243】

アレル相互作用情報はまた、他の個体における同じファミリーの分子（例えば、HLA-A、HLA-B、HLA-C、HLA-DQ、HLA-DR、HLA-DP）のMHCアレルによる提示の、全体的なペプチド配列に非依存的な確率も含むことができる。例えば、HLA-C分子は典型的に、HLA-AまたはHLA-B分子よりも低いレベルで発現しており、したがって、HLA-Cによるペプチドの提示は、HLA-AまたはHLA-B I Iによる提示よりも先験的に確率が低い。別の例として、HLA-DPは一般的にHLA-DRまたはHLA-DQよりも低いレベルで発現されることから、HLA-DPによるペプチドの提示はHLA-DRまたはHLA-DQによる提示よりもより確率が低いものと推測される。

【0244】

アレル相互作用情報はまた、特定のMHCアレルのタンパク質配列も含むことができる。

【0245】

下記のセクションに列挙される任意のMHCアレル非相互作用情報もまた、MHCアレル相互作用情報としてモデル化することができる。

【0246】

#### V I I . B . 2 . アレル非相互作用情報

アレル非相互作用情報は、そのソースタンパク質配列内の、新生抗原コード化ペプチドに隣接するC末端配列を含むことができる。MHC-Iでは、C末端フランキング配列（隣接配列）は、ペプチドのプロテアソームプロセッシングに影響を及ぼし得る。しかし、C末端フランキング配列は、ペプチドが小胞体に輸送され、細胞の表面上のMHCアレルと遭遇する前に、プロテアソームによってペプチドから切断される。その結果、MHC分子は、C末端フランキング配列についてのいかなる情報も受け取らず、したがって、C末端フランキング配列の効果は、MHCアレルタイプに応じて変動することができない。例えば、図2Cに示した例に戻ると、提示情報165は、ペプチドのソースタンパク質から特定された、提示されたペプチドF J I E J F O E S S（配列番号5）のC末端フランキング配列F O E I F N D K S L D K F J I（配列番号9）を含み得る。

【0247】

アレル非相互作用情報はまた、mRNA定量測定値も含むことができる。例えば、mRNA定量データは、質量分析訓練データを提供する同じ試料について取得することができる。図13Hに関して後に記載するように、RNA発現は、ペプチド提示の強い予測因子であると特定された。一実施形態では、mRNA定量測定値は、ソフトウェアツールRSEMから特定される。RSEMソフトウェアツールの詳細な実行は、Bo Li and Colin N. Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 12:323, August 2011で見出すことができる。一実施形態では、mRNA定量は、100万個のマッピングされたリードあたりの転写産物のキログラムあたりの断片の単位（FPKM）で測定される。

【0248】

アレル非相互作用情報はまた、そのソースタンパク質配列内の、ペプチドに隣接するN末端配列も含むことができる。

【0249】

アレル非相互作用情報はペプチド配列のソース遺伝子も含むことができる。ソース遺伝子はペプチド配列のEnsemblタンパク質ファミリーとして定義することができる。

他の例では、ソース遺伝子はペプチド配列のソースDNAまたはソースRNAとして定義することができる。ソース遺伝子は、例えば、タンパク質をコードするヌクレオチドのストリングとして表すか、またはその代わりに、特定のタンパク質をコードしていることが知られている既知のDNAまたはRNA配列の命名されたセットに基づいてよりカテゴリー化された形で表すことができる。別の例では、アレル非相互作用情報は、EnsemblまたはRefSeqのようなデータベースから抽出されたペプチド配列のソース転写産物もしくはアイソフォームまたは潜在的なソース転写産物もしくはアイソフォームのセットも含むことができる。

#### 【0250】

アレル非相互作用情報はまた、ペプチド配列が由来する細胞の組織タイプ、細胞タイプ、または腫瘍タイプも含むことができる。

10

#### 【0251】

アレル非相互作用情報はまた、(RNA-seqまたは質量分析によって測定された際の)任意で、腫瘍細胞における対応するプロテアーゼの発現にしたがって重み付けされる、ペプチドにおけるプロテアーゼ切断モチーフの存在も含むことができる。プロテアーゼ切断モチーフを含有するペプチドは、プロテアーゼによってより容易に分解され、したがって細胞内で安定性がより低いことになるため、提示される可能性がより低い。

#### 【0252】

アレル非相互作用情報はまた、適切な細胞タイプにおいて測定された際の、ソースタンパク質の代謝回転速度も含むことができる。より速い代謝回転速度(すなわち、より低い半減期)は提示の確率を増大させるが、類似していない細胞タイプにおいて測定された場合、この特性の予測力は低い。

20

#### 【0253】

アレル非相互作用情報はまた、RNA-seqもしくはプロテオーム質量分析によって測定された際、または、DNAもしくはRNA配列データにおいて検出される生殖細胞系列もしくは体細胞性スプライシング変異のアノテーションから予測された際の、任意で、腫瘍細胞において最も高発現している特異的なスプライス変異体(「アイソフォーム」)を考慮する、ソースタンパク質の長さも含むことができる。

#### 【0254】

アレル非相互作用情報はまた、(RNA-seq、プロテオーム質量分析、または免疫組織化学によって測定され得る)腫瘍細胞におけるプロテアソーム、イムノプロテアソーム、胸腺プロテアソーム、または他のプロテアーゼの発現のレベルも含むことができる。異なるプロテアソームは、異なる切断部位の好みを有する。より大きい重みが、その発現レベルに比例して、プロテアソームの各タイプの切断の好みに与えられる。

30

#### 【0255】

アレル非相互作用情報はまた、(例えば、RNA-seqまたは質量分析によって測定された際の)ペプチドのソース遺伝子の発現も含むことができる。可能な最適化は、腫瘍試料内の間質細胞及び腫瘍浸潤リンパ球の存在を説明する、測定された発現を調整することを含む。より高発現している遺伝子由来のペプチドは、提示される可能性がより高い。検出不可能なレベルの発現を有する遺伝子由来のペプチドは、考察から排除することができる。

40

#### 【0256】

アレル非相互作用情報はまた、新生抗原コード化ペプチドのソースmRNAが、ナンセンス変異依存分解機構のモデル、例えば、Rivas et al, Science 2015からのモデルによって予測されるようなナンセンス変異依存分解機構に供されるであろう確率も含むことができる。

#### 【0257】

アレル非相互作用情報はまた、細胞周期の種々の段階の最中の、ペプチドのソース遺伝子の典型的な組織特異的発現も含むことができる。(RNA-seqまたは試料分析プロテオミクスによって測定された際に)全体的に低いレベルで発現しているが、細胞周期の

50

特異的な段階の最中に高レベルで発現していることが公知である遺伝子は、非常に低いレベルで安定に発現している遺伝子よりも、より提示されるペプチドを産生する可能性が高い。

【0258】

アレル非相互作用情報はまた、例えば、UniProtまたはPDB <http://www.rcsb.org/pdb/home/home.do>において与えられるような、ソースタンパク質の特性の総合的なカタログも含むことができる。これらの特性は、とりわけ、タンパク質の二次構造及び三次構造、細胞内局在化11、遺伝子オントロジー（GO）用語を含み得る。具体的には、この情報は、タンパク質のレベルで作用するアノテーション、例えば、5' UTR長、及び特異的残基のレベルで作用するアノテーション、例えば、残基300～310のヘリックスモチーフを含有し得る。これらの特性はまた、ターンモチーフ、シートモチーフ、及び無秩序残基も含むことができる。

10

【0259】

アレル非相互作用情報はまた、ペプチドを含有するソースタンパク質のドメインの性状を説明する特性、例えば、二次構造または三次構造（例えば、 $\alpha$ ヘリックス対 $\beta$ シート）；選択的スプライシングも含むことができる。

【0260】

アレル非相互作用情報はまた、ペプチドのソースタンパク質におけるペプチドの位置での提示ホットスポットの有無を説明する特性も含むことができる。

【0261】

アレル非相互作用情報はまた、他の個体における問題のペプチドのソースタンパク質由来のペプチドの提示の確率（それらの個体におけるソースタンパク質の発現レベル、及びそれらの個体の様々なHLAタイプの影響を調整した後）も含むことができる。

20

【0262】

アレル非相互作用情報はまた、ペプチドが、技術的バイアスのために質量分析によって検出されないか、または過剰に表されるであろう確率も含むことができる。

【0263】

腫瘍細胞、間質、または腫瘍浸潤リンパ球（TIL）の状態について情報を与える、RNAseq、マイクロアレイ、Nanostringなどの標的化パネルなどの、遺伝子発現アッセイ、または、RT-PCRなどのアッセイによって測定される遺伝子モジュールを代表する単一／複数遺伝子によって測定された際の、種々の遺伝子モジュール／経路の発現（ペプチドのソースタンパク質を含有する必要はない）。

30

【0264】

アレル非相互作用情報はまた、腫瘍細胞におけるペプチドのソース遺伝子のコピー数も含むことができる。例えば、腫瘍細胞においてホモ接合性欠失に供される遺伝子由来のペプチドは、ゼロの提示確率を割り当てることができる。

【0265】

アレル非相互作用情報はまた、ペプチドがTAPに結合する確率、または、測定されたかもしくは予測された、TAPに対するペプチドの結合親和性も含むことができる。TAPに結合する可能性がより高いペプチド、またはより高い親和性でTAPに結合するペプチドは、MHC-Iによって提示される可能性がより高い。

40

【0266】

アレル非相互作用情報はまた、（RNA-seq、プロテオーム質量分析、免疫組織化学によって測定され得る）腫瘍細胞におけるTAPの発現レベルも含むことができる。MHC-Iでは、より高いTAP発現レベルは、すべてのペプチドの提示の確率を増大させる。

【0267】

アレル非相互作用情報はまた、以下を含むがそれらに限定されない、腫瘍変異の有無も含むことができる：

i. EGFR、KRAS、ALK、RET、ROS1、TP53、CDKN2A、CD

50

KN2B、NTRK1、NTRK2、NTRK3などの公知のがんドライバー遺伝子におけるドライバー変異。

i i . 抗原提示機構に関与するタンパク質をコードする遺伝子（例えば、B2M、HLA-A、HLA-B、HLA-C、TAP-1、TAP-2、TAPBP、CALR、CNX、ERP57、HLA-DM、HLA-DMA、HLA-DMB、HLA-DO、HLA-DOA、HLA-DOBHLA-DP、HLA-DPA1、HLA-DPB1、HLA-DQ、HLA-DQA1、HLA-DQA2、HLA-DQB1、HLA-DQB2、HLA-DR、HLA-DRA、HLA-DRB1、HLA-DRB3、HLA-DRB4、HLA-DRB5、または、プロテアソームもしくはイムノプロテアソームの構成要素をコードする遺伝子のいずれか）におけるもの。その提示が、腫瘍において機能喪失変異の影響下にある抗原提示機構の構成要素に依拠するペプチドは、提示の確率が低減している。

【0268】

以下を含むがそれらに限定されない、機能的生殖細胞系列多型の有無：

i . 抗原提示機構に関与するタンパク質をコードする遺伝子（例えば、B2M、HLA-A、HLA-B、HLA-C、TAP-1、TAP-2、TAPBP、CALR、CNX、ERP57、HLA-DM、HLA-DMA、HLA-DMB、HLA-DO、HLA-DOA、HLA-DOBHLA-DP、HLA-DPA1、HLA-DPB1、HLA-DQ、HLA-DQA1、HLA-DQA2、HLA-DQB1、HLA-DQB2、HLA-DR、HLA-DRA、HLA-DRB1、HLA-DRB3、HLA-DRB4、HLA-DRB5、または、プロテアソームもしくはイムノプロテアソームの構成要素をコードする遺伝子のいずれか）におけるもの。

【0269】

アレル非相互作用情報はまた、腫瘍タイプ（例えば、NSCLC、黒色腫）も含むことができる。

【0270】

アレル非相互作用情報はまた、例としてHLAアレル接尾辞によって反映されるような、HLAアレルの公知の機能性も含むことができる。例えば、アレル名HLA-A\*24:09NにおけるNの接尾辞は、発現せず、したがってエピトープを提示する可能性が低いヌルアレルを示し；完全なHLAアレル接尾辞の命名法は、<https://www.ebi.ac.uk/ipd/imgt/hla/nomenclature/suffixes.html>に記載されている。

【0271】

アレル非相互作用情報はまた、臨床的腫瘍サブタイプ（例えば、扁平上皮肺癌対非扁平上皮）も含むことができる。

【0272】

アレル非相互作用情報はまた、喫煙歴も含むことができる。

【0273】

アレル非相互作用情報はまた、日焼け、日光曝露、または他の変異原に対する曝露の経歴も含むことができる。

【0274】

アレル非相互作用情報はまた、任意でドライバー変異によって層別化される、関連性のある腫瘍タイプまたは臨床的サブタイプにおけるペプチドのソース遺伝子の局所的発現も含むことができる。関連性のある腫瘍タイプにおいて典型的に高レベルで発現している遺伝子は、提示される可能性がより高い。

【0275】

アレル非相互作用情報はまた、すべての腫瘍における、または同じタイプの腫瘍における、または少なくとも1つの共有されたMHCアレルを有する個体由来の腫瘍における、または少なくとも1つの共有されたMHCアレルを有する個体中の同じタイプの腫瘍における、変異の頻度も含むことができる。

## 【0276】

変異した腫瘍特異的ペプチドの例において、提示の確率を予測するために使用される特性の一覧はまた、変異のアノテーション（例えば、ミスセンス、リードスルー、フレームシフト、融合など）、または、変異がナンセンス変異依存分解機構（NMD）を結果としてもたらすと予測されるかどうかも含み得る。例えば、ホモ接合性早期終止変異のために腫瘍細胞において翻訳されないタンパク質セグメント由来のペプチドは、ゼロの提示確率を割り当てることができる。NMDは、提示の確率を減少させる、mRNA翻訳の減少を結果としてもたらす。

## 【0277】

## V I I . C . 提示特定システム

図3は、1つの実施形態による、提示特定システム160のコンピュータ論理構成要素を説明する、ハイレベルブロック図である。この例示的实施形態において、提示特定システム160は、データ管理モジュール312、コード化モジュール314、訓練モジュール316、及び予測モジュール320を含む。提示特定システム160はまた、訓練データ記憶装置170及び提示モデル記憶装置175も含む。モデル管理システム160のいくつかの実施形態は、本明細書に記載したものとは異なるモジュールを有する。同様に、機能は、本明細書に記載したものは異なる様式で、モジュールの間に分配され得る。

## 【0278】

## V I I . C . 1 . データ管理モジュール

データ管理モジュール312は、提示情報165から訓練データ170のセットを生成する。各々の訓練データのセットは、多数のデータ例を含有し、各データ例 $i$ は、少なくとも、提示されるかまたは提示されないペプチド配列 $p^i$ と、ペプチド配列 $p^i$ と結合した1つ以上の関連するMHCアレル $a^i$ と、提示特定システム160が、独立変数の新たな値を予測することに関心があるという情報を表す従属変数 $y^i$ とを含む、独立変数 $z^i$ のセットを含有する。

## 【0279】

本明細書の残りの部分を通じて言及される1つの特定の実現形態において、従属変数 $y^i$ は、ペプチド $p^i$ が、1つ以上の関連するMHCアレル $a^i$ によって提示されたかどうかを示す、バイナリーラベル（標識）である。しかし、他の実現形態において、従属変数 $y^i$ は、提示特定システム160が、独立変数 $z^i$ に依存して予測することに関心があるという任意の他の種類の情報を表し得ることが、認識される。例えば、別の実現形態において、従属変数 $y^i$ はまた、データ例について特定された質量分析イオン電流を示す数値であってよい。

## 【0280】

データ例 $i$ についてのペプチド配列 $p^i$ は、 $k_i$ 個のアミノ酸の配列であり、 $k_i$ は、データ例 $i$ の間で、ある範囲内で変動し得る。例えば、その範囲は、MHCクラスIについては8～15、またはMHCクラスIIについては6～30であり得る。システム160の1つの具体的な実現形態において、訓練データセット中のすべてのペプチド配列 $p^i$ は、同じ長さ、例えば9を有し得る。ペプチド配列中のアミノ酸の数は、MHCアレルのタイプ（例えば、ヒトにおけるMHCアレルなど）に応じて変動し得る。データ例 $i$ についてのMHCアレル $a^i$ は、どのMHCアレルが対応するペプチド配列 $p^i$ と結合して存在したかを示す。

## 【0281】

データ管理モジュール312はまた、訓練データ170に含有されるペプチド配列 $p^i$ 及び結合したMHCアレル $a^i$ と共に、結合親和性 $b^i$ 及び安定性 $s^i$ の予測値などの追加的なアレル相互作用変数も含み得る。例えば、訓練データ170は、ペプチド $p^i$ と、 $a^i$ において示される結合したMHC分子の各々との間の結合親和性予測値 $b^i$ を含有し得る。別の例として、訓練データ170は、 $a^i$ において示されるMHCアレルの各々についての安定性予測値 $s^i$ を含有し得る。

## 【0282】

10

20

30

40

50

データ管理モジュール 312 はまた、ペプチド配列  $p^i$  と共に、C 末端フランキング配列及び mRNA 定量測定値などのアレル非相互作用変数  $w^i$  も含み得る。

【0283】

データ管理モジュール 312 はまた、訓練データ 170 を生成するために、MHC アレルによって提示されないペプチド配列も特定する。概して、これは、提示の前に、提示されるペプチド配列を含むソースタンパク質の「より長い」配列を特定することを含む。提示情報が、操作された細胞株を含有する場合、データ管理モジュール 312 は、細胞に曝露した合成タンパク質における、細胞の MHC アレル上に提示されなかった一連のペプチド配列を特定する。提示情報が組織試料を含有する場合、データ管理モジュール 312 は、提示されたペプチド配列の起源であるソースタンパク質を特定して、組織試料細胞の MHC アレル上に提示されなかった、ソースタンパク質における一連のペプチド配列を特定する。

【0284】

データ管理モジュール 312 はまた、ランダム配列のアミノ酸を有するペプチドを人工的に生成し、生成された配列を、MHC アレル上に提示されないペプチドとして特定する。これは、ペプチド配列をランダムに生成することによって達成することができ、MHC アレル上に提示されないペプチドについての多量の合成データをデータ管理モジュール 312 が容易に生成することを可能にする。実際には、小さなパーセンテージのペプチド配列が MHC アレルによって提示されるため、合成で生成されたペプチド配列は、たとえそれらが細胞によってプロセッシングされたタンパク質に含まれたとしても、MHC アレルによって提示されていない可能性が非常に高い。

【0285】

図 4 は、1 つの実施形態による、訓練データ 170 A の例示的なセットを説明する。具体的には、訓練データ 170 A 中の最初の 3 つのデータ例は、アレル HLA-C\*01:03 と、3 種類のペプチド配列 QCEIOWAREFLKEIGJ (配列番号 10)、FIEUHFWEI (配列番号 11)、及び FEWRHRJTRUJR (配列番号 12) とを含む単一アレル細胞株からのペプチド提示情報を示す。訓練データ 170 A 中の 4 番目のデータ例は、アレル HLA-B\*07:02、HLA-C\*01:03、HLA-A\*01:01 と、ペプチド配列 QIEJOEIJ E (配列番号 13) とを含む複数アレル細胞株からのペプチド情報を示す。最初のデータ例は、ペプチド配列 QCEIOWARE (配列番号 14) が、アレル HLA-DRB3:01:01 によって提示されなかったことを示す。前の 2 つの段落において議論したように、ネガティブ標識されたペプチド配列は、データ管理モジュール 312 によってランダムに生成されてもよいし、提示されるペプチドのソースタンパク質から特定されてもよい。訓練データ 170 A はまた、ペプチド配列-アレルペアについて、1000 nM の結合親和性予測値及び 1 時間の半減期の安定性予測値も含む。訓練データ 170 A はまた、ペプチド FJELFISBOSJFIE (配列番号 15) の C 末端フランキング配列、及び  $10^2$  TPM の mRNA 定量測定値などの、アレル非相互作用変数も含む。4 番目のデータ例は、ペプチド配列 QIEJOEIJ E (配列番号 13) が、アレル HLA-B\*07:02、HLA-C\*01:03、または HLA-A\*01:01 のうちの 1 つによって提示されたことを示す。訓練データ 170 A はまた、アレルの各々についての結合親和性予測値及び安定性予測値、ならびに、ペプチドの C 末端フランキング配列及びペプチドについての mRNA 定量測定値も含む。

【0286】

#### VII. C. 2. コード化モジュール

コード化モジュール 314 は、訓練データ 170 に含有される情報を、1 つ以上の提示モデルを生成するために使用することができる数値的表示へとコード化する。一実現形態では、コード化モジュール 314 は、配列 (例えば、ペプチド配列または C 末端フランキング配列) を、あらかじめ決定された 20 文字のアミノ酸アルファベットについて、ワンホットでコード化する。具体的には、 $k_i$  個のアミノ酸を有するペプチド配列  $p^i$  は、 $20 \cdot k_i$  要素の行ベクトルとして表され、ペプチド配列の  $j$  番目の位置のアミノ酸のアル

10

20

30

40

50

コード化モジュール 314 はまた、各データ例  $i$  についてのラベル（標識） $y_i$  を、 $\{$

0, 1} のセットからの値を有するバイナリー変数としてコード化し、1の値は、ペプチド  $x^i$  が、関連するMHCアレル  $a^i$  のうちの1つによって提示されたことを示し、0の値は、ペプチド  $x^i$  が、関連するMHCアレル  $a^i$  のいずれによっても提示されなかったことを示す。従属変数  $y_i$  が質量分析イオン電流を表す場合、コード化モジュール314は、 $[0, \infty]$  の間のイオン電流値について  $[-\infty, \infty]$  の範囲を有する  $\log$  関数などの種々の関数を用いて、値を追加的にスケール調整し得る。

【0291】

コード化モジュール314は、ペプチド  $p_i$  及び関連するMHCアレル  $h$  についてのアレル相互作用変数  $x_{h^i}$  のペアを、アレル相互作用変数の数値的表示が次々に連結されている行ベクトルとして表し得る。例えば、コード化モジュール314は、 $x_{h^i}$  を、 $[p^i]$ 、 $[p^i b_{h^i}]$ 、 $[p^i s_{h^i}]$ 、または  $[p^i b_{h^i} s_{h^i}]$  と同等の行ベクトルとして表し得、 $b_{h^i}$  は、ペプチド  $p_i$  及び関連するMHCアレル  $h$  についての結合親和性予測値であり、同様に、 $s_{h^i}$  は、安定性についてのものである。あるいは、アレル相互作用変数の1つ以上の組み合わせは、個々に（例えば、個々のベクトルまたは行列として）保存されてもよい。

10

【0292】

1つの例において、コード化モジュール314は、結合親和性について測定されたかまたは予測された値をアレル相互作用変数  $x_{h^i}$  に組み入れることによって、結合親和性情報を表す。

【0293】

1つの例において、コード化モジュール314は、結合安定性について測定されたかまたは予測された値をアレル相互作用変数  $x_{h^i}$  に組み入れることによって、結合安定性情報を表す。

20

【0294】

1つの例において、コード化モジュール314は、結合オンレートについて測定されたかまたは予測された値をアレル相互作用変数  $x_{h^i}$  に組み入れることによって、結合オンレート情報

【0295】

1つの例において、クラスI MHC分子によって提示されるペプチドについて、コード化モジュール314は、ペプチド長を、ベクトル

30

$T_k = [\mathbb{1}(L_k=8) \mathbb{1}(L_k=9) \mathbb{1}(L_k=10) \mathbb{1}(L_k=11) \mathbb{1}(L_k=12) \mathbb{1}(L_k=13) \mathbb{1}(L_k=14) \mathbb{1}(L_k=15)]$

（ただし、

$\mathbb{1}$

は指標関数であり、 $L_k$  はペプチド  $p_k$  の長さを意味する）として表す。ベクトル  $T_k$  を、アレル相互作用変数  $x_{h^i}$  に含めることができる。別の例では、クラスIIのMHC分子によって提示されるペプチドについて、コード化モジュール314はペプチド長をベクトル

$T_k = [\mathbb{1}(L_k=6) \mathbb{1}(L_k=7) \mathbb{1}(L_k=8) \mathbb{1}(L_k=9) \mathbb{1}(L_k=10) \mathbb{1}(L_k=11)$

$\mathbb{1}(L_k=12) \mathbb{1}(L_k=13) \mathbb{1}(L_k=14) \mathbb{1}(L_k=15) \mathbb{1}(L_k=16) \mathbb{1}(L_k=17) \mathbb{1}(L_k=18) \mathbb{1}(L_k=19) \mathbb{1}(L_k=20) \mathbb{1}(L_k=21)$

40

$\mathbb{1}(L_k=22) \mathbb{1}(L_k=23) \mathbb{1}(L_k=24) \mathbb{1}(L_k=25) \mathbb{1}(L_k=26) \mathbb{1}(L_k=27) \mathbb{1}(L_k=28) \mathbb{1}(L_k=29) \mathbb{1}(L_k=30)]$

（ただし、

$\mathbb{1}$

は指標関数であり、 $L_k$  はペプチド  $p_k$  の長さを意味する）として表す。ベクトル  $T_k$  を、アレル相互作用変数  $x_{h^i}$  に含めることができる。

【0296】

1つの例において、コード化モジュール314は、MHCアレルのRNA-seqベ

50

スの発現レベルをアレル相互作用変数  $x_{hi}$  に組み入れることによって、MHCアレルのRNA発現情報を表す。

【0297】

同様に、コード化モジュール314は、アレル非相互作用変数  $w^i$  を、アレル非相互作用変数の数値的表示が次々に連鎖している行ベクトルとして表し得る。例えば、 $w^i$  は、 $[c^i]$  または  $[c^i m^i w^i]$  と同等の行ベクトルであってもよく、 $w_i$  は、ペプチド  $p^i$  のC末端フランキング配列及びペプチドに関連するmRNA定量測定値  $m^i$  に加えて任意の他のアレル非相互作用変数を表す、行ベクトルである。あるいは、アレル非相互作用変数の1つ以上の組み合わせは、個々に（例えば、個々のベクトルまたは行列として）保存されてもよい。

【0298】

1つの例において、コード化モジュール314は、代謝回転速度または半減期をアレル非相互作用変数  $w^i$  に組み入れることによって、ペプチド配列についてのソースタンパク質の代謝回転速度を表す。

【0299】

1つの例において、コード化モジュール314は、タンパク質長をアレル非相互作用変数  $w^i$  に組み入れることによって、ソースタンパク質またはアイソフォームの長さを表す。

【0300】

1つの例において、コード化モジュール314は、 $\beta 1_i$ 、 $\beta 2_i$ 、 $\beta 5_i$  サブユニットを含むイムノプロテアソーム特異的プロテアソームサブユニットの平均発現を、アレル非相互作用変数  $w^i$  に組み入れることによって、イムノプロテアソームの活性化を表す。

【0301】

1つの例において、コード化モジュール314は、（RSEMなどの技法によってFPKM、TPMの単位で定量された）ペプチド、またはペプチドの遺伝子もしくは転写産物のソースタンパク質のRNA-seq存在量を、ソースタンパク質の存在量をアレル非相互作用変数  $w^i$  に組み入れることによって表す。

【0302】

1つの例において、コード化モジュール314は、例えば、Rivas et. al. Science, 2015におけるモデルによって推定されるような、ペプチドの起源の転写産物がナンセンス変異依存分解機構（NMD）を受けるであろう確率を、この確率をアレル非相互作用変数  $w^i$  に組み入れることによって表す。

【0303】

1つの例において、コード化モジュール314は、RNA-seqを介して評価された遺伝子モジュールまたは経路の活性化状況を、例えば、経路における遺伝子の各々について、例えばRSEMを用いてTPMの単位で、経路における遺伝子の発現を定量すること、次いで、経路における遺伝子にわたる要約統計量、例えば平均値をコンピュータ計算することによって表す。平均を、アレル非相互作用変数  $w^i$  に組み入れることができる。

【0304】

1つの例において、コード化モジュール314は、ソース遺伝子のコピー数を、コピー数をアレル非相互作用変数  $w^i$  に組み入れることによって表す。

【0305】

1つの例において、コード化モジュール314は、（例えば、ナノモル単位での）測定されたかまたは予測されたTAP結合親和性をアレル非相互作用変数  $w^i$  に含むことによって、TAP結合親和性を表す。

【0306】

1つの例において、コード化モジュール314は、RNA-seqによって測定され（かつ、例えばRSEMによってTPMの単位で定量された）TAP発現レベルをアレル非相互作用変数  $w^i$  に含むことによって、TAP発現レベルを表す。

【0307】

1つの例において、コード化モジュール314は、腫瘍変異を、アレル非相互作用変数

10

20

30

40

50

$w^i$ における指標変数 (indicator variable) のベクトル (すなわち、ペプチド  $p^k$  が K R A S G 1 2 D 変異を有する試料に由来するならば  $d^k = 1$ 、それ以外は 0) として表す。

【0308】

1つの例において、コード化モジュール314は、抗原提示遺伝子における生殖細胞系列多型を、指標変数のベクトル (すなわち、ペプチド  $p^k$  が T A P において特異的な生殖細胞系列多型を有する試料に由来するならば  $d^k = 1$ ) として表す。これらの指標変数を、アレル非相互作用変数  $w^i$  に含めることができる。

【0309】

1つの例において、コード化モジュール314は、腫瘍タイプを、腫瘍タイプ (例えば、N S C L C、黒色腫、大腸癌など) のアルファベットについての長さ1のワンホットコード化ベクトルとして表す。これらのワンホットコード化変数を、アレル非相互作用変数  $w^i$  に含めることができる。

【0310】

1つの例において、コード化モジュール314は、M H C アレル接尾辞を、4桁の H L A アレルを様々な接尾辞で処理することによって表す。例えば、H L A - A \* 2 4 : 0 9 N は、モデルの目的で、H L A - A \* 2 4 : 0 9 とは異なるアレルと考えられる。あるいは、N 接尾辞で終わる H L A アレルは発現しないため、N 接尾辞の M H C アレルによる提示の確率は、すべてのペプチドについてゼロに設定することができる。

【0311】

1つの例において、コード化モジュール314は、腫瘍サブタイプを、腫瘍サブタイプ (例えば、肺腺癌、肺扁平上皮細胞癌など) のアルファベットについての長さ1のワンホットコード化ベクトルとして表す。これらのワンホットコード化変数を、アレル非相互作用変数  $w^i$  に含めることができる。

【0312】

1つの例において、コード化モジュール314は、喫煙歴を、アレル非相互作用変数  $w^i$  に含めることができる、バイナリー指標変数 (患者が喫煙歴を有するならば  $d^k = 1$ 、それ以外は 0) として表す。あるいは、喫煙歴を、喫煙の重症度のアルファベットについての長さ1のワンホットコード化変数としてコード化することができる。例えば、喫煙状況を、1が非喫煙者を示し、5が現在の大量喫煙者を示す、1~5のスケールに査定することができる。喫煙歴は、主として肺腫瘍と関連性があるため、複数の腫瘍タイプに対するモデルを訓練する場合、この変数は、患者が喫煙の経歴を有し、かつ腫瘍タイプが肺腫瘍であるならば1と同等であり、それ以外はゼロであると定義することもできる。

【0313】

1つの例において、コード化モジュール314は、日焼け歴を、アレル非相互作用変数  $w^i$  に含めることができる、バイナリー指標変数 (患者が重症の日焼けの経歴を有するならば  $d^k = 1$ 、それ以外は 0) として表す。重症の日焼けは、主として黒色腫と関連性があるため、複数の腫瘍タイプに対するモデルを訓練する場合、この変数は、患者が重症の日焼けの経歴を有し、かつ腫瘍タイプが黒色腫であるならば1と同等であり、それ以外はゼロであると定義することもできる。

【0314】

1つの例において、コード化モジュール314は、ヒトゲノムにおける各遺伝子または転写産物についての特定の遺伝子または転写産物の発現レベルの分布を、T C G A などの参照データベースを用いることによって、発現レベルの分布の要約統計量 (例えば、平均値、中央値) として表す。具体的には、腫瘍タイプ黒色腫を有する試料におけるペプチド  $p^k$  について、ペプチド  $p^k$  の起源の遺伝子または転写産物の、測定された遺伝子または転写産物の発現レベルをアレル非相互作用変数  $w^i$  に含むことができるだけでなく、T C G A によって測定された際の、黒色腫におけるペプチド  $p^k$  の起源の遺伝子または転写産物の、平均値及び／または中央値の遺伝子または転写産物発現も含めることができる。

【0315】

1つの例において、コード化モジュール314は、変異タイプを、変異タイプ（例えば、ミスセンス、フレームシフト、NMD誘導性など）のアルファベットについての長さ1のワンホットコード化変数として表す。これらのワンホットコード化変数を、アレル非相互作用変数 $w^i$ に含めることができる。

【0316】

1つの例において、コード化モジュール314は、タンパク質のタンパク質レベルの特性を、ソースタンパク質のアノテーション（例えば、5'UTR長）の値として、アレル非相互作用変数 $w^i$ において表す。別の例において、コード化モジュール314は、ペプチド $p^i$ についてのソースタンパク質の残基レベルのアノテーションを、ペプチド $p^i$ がヘリックスモチーフとオーバーラップするならば1と同等であり、それ以外は0であるか、または、ペプチド $p^i$ がヘリックスモチーフ内に完全に含有されるならば1と同等である指標変数を、アレル非相互作用変数 $w^i$ に含むことによって表す。別の例において、ヘリックスモチーフアノテーション内に含有されるペプチド $p^i$ における残基の割合を表す特性を、アレル非相互作用変数 $w^i$ に含めることができる。

【0317】

1つの例において、コード化モジュール314は、ヒトプロテオームにおけるタンパク質またはアイソフォームのタイプを、ヒトプロテオームにおけるタンパク質またはアイソフォームの数と同等の長さを有する指標ベクトル $o^k$ として表し、対応する要素 $o^k_i$ は、ペプチド $p^k$ がタンパク質 $i$ に由来するならば1であり、それ以外は0である。

【0318】

1つの例において、コード化モジュール314は、ペプチド $p^i$ のソース遺伝子 $G = gene(p^i)$ を $L$ 個の可能なカテゴリーを有するカテゴリー変数として表す（ただし、 $L$ は添え字を付したソース遺伝子の数の上限1、2、…、 $L$ を示す）。

【0319】

1つの例において、コード化モジュール314は、ペプチド $p^i$ の組織タイプ、細胞タイプ、腫瘍タイプ、または腫瘍組織学タイプ $T = 組織(p^i)$ を $M$ 個の可能なカテゴリーを有するカテゴリー変数として表す（ただし、 $M$ は添え字を付したタイプ1、2、…、 $M$ の数の上限を示す）。組織のタイプとしては、例えば、肺組織、心組織、腸組織、神経組織などを挙げることができる。細胞のタイプとしては、樹状細胞、マクロファージ、CD4 T細胞などを挙げることができる。肺腺癌、肺扁平上皮癌、メラノーマ、非ホジキンリンパ腫などを挙げることができる。

【0320】

コード化モジュール314はまた、ペプチド $p^i$ 及び関連するMHCアレル $h$ についての変数 $z^i$ の全体的なセットを、アレル相互作用変数 $x^i$ 及びアレル非相互作用変数 $w^i$ の数値的表示が次々に連鎖している行ベクトルとしても表し得る。例えば、コード化モジュール314は、 $z^i_h$ を、 $[x^i_h w^i]$ または $[w^i x^i_h]$ と同等の行ベクトルとして表し得る。

【0321】

#### VIII. 訓練モジュール

訓練モジュール316は、ペプチド配列に関連するMHCアレルによってペプチド配列が提示されるかどうかの尤度を生成する、1つ以上の提示モデルを構築する。具体的には、ペプチド配列 $p^k$ 及びペプチド配列 $p^k$ に関連するMHCアレル $a^k$ のセットを与えられ、各提示モデルは、ペプチド配列 $p^k$ が、関連するMHCアレル $a^k$ のうちの1つ以上によって提示されるであろう尤度を示す、推定値 $u_k$ を生成する。

【0322】

#### VIII. A. 概要

訓練モジュール316は、165に保存された提示情報から生成された、記憶装置170に保存された訓練データセットに基づいて、1つ以上の提示モデルを構築する。概して、提示モデルの具体的なタイプに関わらず、提示モデルのすべては、損失関数が最小化されるように、訓練データ170における独立変数と従属変数との間の依存性を捕捉する。

具体的には、損失関数  $(y_{i \in S}, u_{i \in S}; \theta)$  は、訓練データ 170 における 1 つ以上のデータ例  $S$  についての従属変数  $y_{i \in S}$  の値と、提示モデルによって生成されたデータ例  $S$  についての推定された尤度  $u_{i \in S}$  との間の矛盾を表す。本明細書の残りの部分を通じて言及される 1 つの特定の実現形態において、損失関数  $(y_{i \in S}, u_{i \in S}; \theta)$  は、以下のような等式 (1a) によって与えられる負の  $\log$  尤度関数である。

$$\ell(y_{i \in S}, u_{i \in S}; \theta) = \sum_{i \in S} (y_i \log u_i + (1 - y_i) \log(1 - u_i)) \quad (1a)$$

しかし、実際には、別の損失関数が使用されてもよい。例えば、質量分析イオン電流について予測がなされる場合、損失関数は、以下のような等式 1b によって与えられる平均二乗損失である。

$$\ell(y_{i \in S}, u_{i \in S}; \theta) = \sum_{i \in S} (\|y_i - u_i\|_2^2) \quad (1b)$$

### 【0323】

提示モデルは、1 つ以上のパラメータ  $\theta$  が、独立変数と従属変数との間の依存性を数学的に明記する、パラメトリックモデルであり得る。典型的に、損失関数  $(y_{i \in S}, u_{i \in S}; \theta)$  を最小化するパラメトリックタイプの提示モデルの種々のパラメータは、例えば、バッチ勾配アルゴリズム、確率的勾配アルゴリズムなどの、勾配ベースの数値的最適化アルゴリズムを通して決定される。あるいは、提示モデルは、モデル構造が、訓練データ 170 から決定され、固定されたパラメータのセットに厳密には基づかない、ノンパラメトリックモデルであり得る。

### 【0324】

#### VIII.B. アレルごとのモデル

訓練モジュール 316 は、アレルごとベースでペプチドの提示尤度を予測するための提示モデルを構築し得る。この例において、訓練モジュール 316 は、単一の MHC アレルを発現する細胞から生成された訓練データ 170 におけるデータ例  $S$  に基づいて、提示モデルを訓練し得る。

### 【0325】

一実現形態では、訓練モジュール 316 は、

$$u_k^h = \Pr(p^k \text{ 提示}; \text{MHC アレル } h) = f(g_h(x_h^k; \theta_h)) \quad (2)$$

によって、特定のアレル  $h$  についてのペプチド  $p^k$  の推定提示尤度  $u^k$  をモデル化し、ただし、ペプチド配列  $x_h^k$  は、ペプチド  $p^k$  及び対応する MHC アレル  $h$  についてのコード化されたアレル相互作用変数を意味し、 $f(\cdot)$  は、任意の関数であり、記載の便宜上、本明細書中を通して変換関数と呼ばれる。さらに、 $g_h(\cdot)$  は、任意の関数であり、記載の便宜上、本明細書中を通して依存性関数と呼ばれ、MHC アレル  $h$  について決定されたパラメータ  $\theta_h$  のセットに基づいて、アレル相互作用変数  $x_h^k$  についての依存性スコアを生成する。各 MHC アレル  $h$  についてのパラメータ  $\theta_h$  のセットの値は、 $\theta_h$  に関する損失関数を最小化することによって決定することができ、ここで  $i$  は、単一の MHC アレル  $h$  を発現する細胞から生成された訓練データ 170 のサブセット  $S$  における各例である。

### 【0326】

依存性関数  $g_h(x_h^k; \theta_h)$  の出力は、MHC アレル  $h$  が、少なくともアレル相互作用特性  $x_h^k$  に基づいて、及び特に、ペプチド  $p^k$  のペプチド配列のアミノ酸の位置に基づいて、対応する新生抗原を提示するかどうかを示す、MHC アレル  $h$  についての依存性スコアを表す。例えば、MHC アレル  $h$  についての依存性スコアは、MHC アレル  $h$  が、ペプチド  $p^k$  を提示する可能性が高い場合に、高い値を有し得、提示の可能性が高くない場合に、低い値を有し得る。変換関数  $f(\cdot)$  は、入力を変換し、より具体的には、この例において  $g_h(x_h^k; \theta_h)$  によって生成された依存性スコアを、ペプチド  $p^k$  が MHC

アレルによって提示されるであろう尤度を示す適切な値に変換する。

【0327】

本明細書の残りの部分を通じて言及される1つの特定の実現形態において、 $f(\cdot)$ は、適切なドメイン範囲について $[0, 1]$ 内の範囲を有する関数である。1つの例において、 $f(\cdot)$ は、

$$f(z) = \frac{\exp(z)}{1 + \exp(z)} \quad (4)$$

によって与えられる $\exp$ 関数である。

別の例として、 $f(\cdot)$ はまた、ドメイン $z$ の値が0以上である場合、

$$f(z) = \tanh(z) \quad (5)$$

によって与えられる双曲線正接関数であることもできる。あるいは、予測が、範囲 $[0, 1]$ の外側の値を有する質量分析イオン電流についてなされる場合、 $f(\cdot)$ は、例えば、恒等関数、指数関数、 $\log$ 関数などの任意の関数であることができる。

【0328】

したがって、ペプチド配列 $p^k$ がMHCアレル $h$ によって提示されるであろうアレルごとの尤度は、MHCアレル $h$ についての依存性関数 $g_h(\cdot)$ をペプチド配列 $p^k$ のコード化されたバージョンに適用して、対応する依存性スコアを生成することによって、生成することができる。依存性スコアは、ペプチド配列 $p^k$ がMHCアレル $h$ によって提示されるであろうアレルごとの尤度を生成するように、変換関数 $f(\cdot)$ によって変換されてもよい。

【0329】

#### VIII.B.1 アレル相互作用変数についての依存性関数

本明細書を通して言及される1つの特定の実現形態において、依存性関数 $g_h(\cdot)$ は、 $x_h^i$ における各アレル相互作用変数を、関連するMHCアレル $h$ について決定されたパラメータ $\theta_h$ のセットにおける対応するパラメータと線形結合する、

$$g_h(x_h^i; \theta_h) = x_h^i \cdot \theta_h \quad (6)$$

によって与えられるアフィン関数である。

【0330】

本明細書を通して言及される別の特定の実現形態において、依存性関数 $g_h(\cdot)$ は、1つ以上の層において配置された一連のノードを有するネットワークモデル $NN_h(\cdot)$ によって表される、

$$g_h(x_h^i; \theta_h) = NN_h(x_h^i; \theta_h) \quad (7)$$

によって与えられるネットワーク関数である。ノードは、パラメータ $\theta_h$ のセットにおける関連するパラメータを各々有する接続を通して、他のノードに接続され得る。1つの特定のノードでの値は、特定のノードに関連する活性化関数によってマッピングされた関連するパラメータによって重み付けられた、特定のノードに接続されたノードの値の和として表され得る。アフィン関数と対照的に、ネットワークモデルは、提示モデルが非線形性、及び異なる長さのアミノ酸配列を有するプロセスデータを組み入れることができるため、有利である。具体的には、非線形モデリングを通して、ネットワークモデルは、ペプチド配列中の異なる位置のアミノ酸間の相互作用、及びこの相互作用がペプチド提示にいか

【0331】

概して、ネットワークモデル $NN_h(\cdot)$ は、人工ニューラルネットワーク(ANN)、畳み込みニューラルネットワーク(CNN)、深層ニューラルネットワーク(DNN)などのフィードフォワードネットワーク、及び／または、長・短期記憶ネットワーク(LSTM)、双方向LSTMネットワーク、双方向再帰型ネットワーク、深層双方向再帰型

ネットワーク、多層知覚ネットワーク（N L P）などの再帰型ニューラルネットワーク（R N N）などとして、構造化され得る。

【0332】

本明細書の残りの部分を通じて言及される1つの例において、 $h = 1, 2, \dots, m$ における各MHCアレルは、別々のネットワークモデルに関連し、 $NN_h(\cdot)$ は、MHCアレル $h$ に関連するネットワークモデルからの出力を意味する。

【0333】

図5は、任意のMHCアレル $h = 3$ に関連した例示的なネットワークモデル $NN_3(\cdot)$ を説明する。図5に示すように、MHCアレル $h = 3$ についてのネットワークモデル $NN_3(\cdot)$ は、層 $l = 1$ での3種類の入力ノード、層 $l = 2$ での4種類のノード、層 $l = 3$ での2種類のノード、及び層 $l = 4$ での1種類の出力ノードを含む。ネットワークモデル $NN_3(\cdot)$ は、10種類のパラメータ $\theta_3(1), \theta_3(2), \dots, \theta_3(10)$ のセットに関連している。ネットワークモデル $NN_3(\cdot)$ は、MHCアレル $h = 3$ についての3種類のアレル相互作用変数 $x_3^k(1), x_3^k(2)$ 、及び $x_3^k(3)$ についての入力値（コード化されたポリペプチド配列データ及び使用される任意の他の訓練データを含む、個々のデータ例）を受け取り、値 $NN_3(x_3^k)$ を出力する。ネットワーク関数は、異なるアレル相互作用変数をそれぞれが入力として取る1つ以上のネットワークモデルを含んでもよい。

【0334】

別の例において、特定されたMHCアレル $h = 1, 2, \dots, m$ は、単一ネットワークモデル $NN_H(\cdot)$ に関連しており、 $NN_h(\cdot)$ は、MHCアレル $h$ に関連する単一ネットワークモデルの1つ以上の出力を意味する。そのような例において、パラメータ $\theta_h$ のセットは、単一ネットワークモデルについてのパラメータのセットに対応し得、したがって、パラメータ $\theta_h$ のセットは、すべてのMHCアレルによって共有され得る。

【0335】

図6Aは、MHCアレル $h = 1, 2, \dots, m$ によって共有される例示的なネットワークモデル $NN_H(\cdot)$ を説明する。図6Aに示すように、ネットワークモデル $NN_H(\cdot)$ は、MHCアレルに各々対応する、 $m$ 個の出力ノードを含む。ネットワークモデル $NN_3(\cdot)$ は、MHCアレル $h = 3$ についてのアレル相互作用変数 $x_3^k$ を受け取り、MHCアレル $h = 3$ に対応する値 $NN_3(x_3^k)$ を含む、 $m$ 個の値を出力する。

【0336】

さらに別の例において、単一ネットワークモデル $NN_H(\cdot)$ は、MHCアレル $h$ のアレル相互作用変数 $x_h^k$ 及びコード化されたタンパク質配列 $d_h$ を与えられて依存性スコアを出力する、ネットワークモデルであり得る。そのような例において、パラメータ $\theta_h$ のセットは、再び、単一ネットワークモデルについてのパラメータのセットに対応し得、したがって、パラメータ $\theta_h$ のセットは、すべてのMHCアレルによって共有され得る。したがって、そのような例において、 $NN_h(\cdot)$ は、単一ネットワークモデルに対して入力 $[x_h^k d_h]$ を与えられた、単一ネットワークモデル $NN_H(\cdot)$ の出力を意味する。そのようなネットワークモデルは、訓練データにおいて未知であったMHCアレルについてのペプチド提示確率を、単にそれらのタンパク質配列を特定することによって正しく予測することができるため、有利である。

【0337】

図6Bは、MHCアレルによって共有される例示的なネットワークモデル $NN_H(\cdot)$ を説明する。図6Bに示すように、ネットワークモデル $NN_H(\cdot)$ は、MHCアレル $h = 3$ のアレル相互作用変数及びタンパク質配列を入力として受け取り、MHCアレル $h = 3$ に対応する依存性スコア $NN_3(x_3^k)$ を出力する。

【0338】

さらに別の例において、依存性関数 $g_h(\cdot)$ は、

$$g_h(x_h^k, \theta_h) = g'_h(x_h^k, \theta'_h) + \theta_h^0$$

10

20

30

40

50

として表すことができ、式中、 $g_h(x_h^k; \theta_h)$  は、パラメータ  $\theta_h$  のセットを伴うアフィン関数、ネットワーク関数などであり、MHC アレル  $h$  についての提示のベースライン確率を表す、MHC アレルのアレル相互作用変数についてのパラメータのセットにおけるバイアスパラメータ  $\theta_h^0$  を伴う。

【0339】

別の実現形態において、バイアスパラメータ  $\theta_h^0$  は、MHC アレル  $h$  の遺伝子ファミリーにしたがって共有されてもよい。すなわち、MHC アレル  $h$  についてのバイアスパラメータ  $\theta_h^0$  は  $\theta_{\text{遺伝子}(h)}^0$  と同等であり得、遺伝子  $(h)$  は、MHC アレル  $h$  の遺伝子ファミリーである。例えば、クラス I MHC アレル HLA-A\*02:01、HLA-A\*02:02、及び HLA-A\*02:03 は、「HLA-A」の遺伝子ファミリーに割り当てられてもよく、これらの MHC アレルの各々についてのバイアスパラメータ  $\theta_h^0$  が共有されてもよい。別の例として、クラス II MHC アレル HLA-DRB1:10:01、HLA-DRB1:11:01、及び HLA-DRB3:01:01 を「HLA-DRB」の遺伝子ファミリーに割り当て、これらの MHC アレルのそれぞれのバイアスパラメータ  $\theta_h^0$  を共有することができる。

【0340】

例として、等式 (2) に戻ると、アフィン依存性関数  $g_h(\cdot)$  を用いた、 $m=4$  の異なる特定された MHC アレルの中で MHC アレル  $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k^3 = f(x_3^k; \theta_3)$$

によって生成することができ、式中、 $x_3^k$  は、MHC アレル  $h=3$  について特定されたアレル相互作用変数であり、 $\theta_3$  は、損失関数最小化を通して MHC アレル  $h=3$  について決定されたパラメータのセットである。

【0341】

別の例として、別々のネットワーク変換関数  $g_h(\cdot)$  を用いた、 $m=4$  の異なる特定された MHC アレルの中で MHC アレル  $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k^3 = f(NN_3(x_3^k; \theta_3))$$

によって生成することができ、式中、 $x_3^k$  は、MHC アレル  $h=3$  について特定されたアレル相互作用変数であり、 $\theta_3$  は、MHC アレル  $h=3$  に関連するネットワークモデル  $NN_3(\cdot)$  について決定されたパラメータのセットである。

【0342】

図 7 は、例示的なネットワークモデル  $NN_3(\cdot)$  を用いた、MHC アレル  $h=3$  に関連したペプチド  $p^k$  の提示尤度の生成を説明する。図 7 に示すように、ネットワークモデル  $NN_3(\cdot)$  は、MHC アレル  $h=3$  についてのアレル相互作用変数  $x_3^k$  を受け取り、出力  $NN_3(x_3^k)$  を生成する。出力は、関数  $f(\cdot)$  によってマッピングされて、推定提示尤度  $u_k$  を生成する。

【0343】

#### VIII.B.2. アレル非相互作用変数を伴うアレルごと

一実現形態では、訓練モジュール 316 は、アレル非相互作用変数を組み入れて、

$$u_k^h = \Pr(p^k \text{ 提示}) = f(g_w(w^k; \theta_w) + g_h(x_h^k; \theta_h)) \quad (8)$$

によって、ペプチド  $p^k$  の推定提示尤度  $u_k$  をモデル化し、式中、 $w^k$  は、ペプチド  $p^k$  についてのコード化されたアレル非相互作用変数を意味し、 $g_w(\cdot)$  は、アレル非相互作用変数について決定されたパラメータ  $\theta_w$  のセットに基づく、アレル非相互作用変数  $w^k$  についての関数である。具体的には、各 MHC アレル  $h$  についてのパラメータ  $\theta_h$  のセット及びアレル非相互作用変数についてのパラメータ  $\theta_w$  のセットの値を、 $\theta_h$  及び  $\theta_w$  に関

する損失関数を最小化することによって決定することができ、 $i$  は、単一の MHC アレルを発現する細胞から生成された訓練データ 170 のサブセット  $S$  における各例である。

【0344】

依存性関数  $g_w(w^k; \theta_w)$  の出力は、アレル非相互作用変数の影響に基づいて、1つ以上の MHC アレルによってペプチド  $p^k$  が提示されるかどうかを示す、アレル非相互作用変数についての依存性スコアを表す。例えば、アレル非相互作用変数についての依存性スコアは、ペプチド  $p^k$  の提示に正の影響を及ぼすことが公知である C 末端フランキング配列とペプチド  $p^k$  が結合している場合は、高い値を有し得、ペプチド  $p^k$  の提示に負の影響を及ぼすことが公知である C 末端フランキング配列とペプチド  $p^k$  が結合している場合は、低い値を有し得る。

【0345】

等式 (8) によると、ペプチド配列  $p^k$  が MHC アレル  $h$  によって提示されるであろうアレルごとの尤度は、MHC アレル  $h$  についての関数  $g_h(\cdot)$  を、ペプチド配列  $p^k$  のコード化されたバージョンに適用して、アレル相互作用変数について対応する依存性スコアを生成することによって、生成することができる。アレル非相互作用変数についての関数  $g_w(\cdot)$  もまた、アレル非相互作用変数についての依存性スコアを生成するように、アレル非相互作用変数のコード化されたバージョンに適用される。両方のスコアが組み合わされ、組み合わされたスコアが、MHC アレル  $h$  によってペプチド配列  $p^k$  が提示されるであろうアレルごとの尤度を生成するように、変換関数  $f(\cdot)$  によって変換される。

【0346】

あるいは、訓練モジュール 316 は、等式 (2) においてアレル非相互作用変数  $w^k$  をアレル相互作用変数  $x_h^k$  に付加することにより、予測におけるアレル非相互作用変数  $w^k$  を含んでもよい。したがって、提示尤度は、

$$u_k^h = \Pr(p^k \text{ 提示}; \text{アレル } h) = f(g_h([x_h^k w^k]; \theta_h)) \quad (9)$$

によって与えられ得る。

【0347】

#### VIII.B.3 アレル非相互作用変数についての依存性関数

アレル相互作用変数についての依存性関数  $g_h(\cdot)$  と同様に、アレル非相互作用変数についての依存性関数  $g_w(\cdot)$  は、アフィン関数、または別々のネットワークモデルがアレル非相互作用変数  $w^k$  に関連しているネットワーク関数であり得る。

【0348】

具体的には、依存性関数  $g_w(\cdot)$  は、 $w^k$  におけるアレル非相互作用変数を、パラメータ  $\theta_w$  のセットにおける対応するパラメータと線形結合する、

$$g_w(w^k; \theta_w) = w^k \cdot \theta_w.$$

によって与えられるアフィン関数である。

【0349】

依存性関数  $g_w(\cdot)$  はまた、パラメータ  $\theta_w$  のセットにおける関連するパラメータを有するネットワークモデル  $NN_w(\cdot)$  によって表される、

$$g_w(w^k; \theta_w) = NN_w(w^k; \theta_w)$$

によって与えられるネットワーク関数である。ネットワーク関数は、異なるアレル非相互作用変数をそれぞれが入力として取る 1 つ以上のネットワークモデルを含んでもよい。

【0350】

別の例において、アレル非相互作用変数についての依存性関数  $g_w(\cdot)$  は、

$$g_w(w^k; \theta_w) = g'_w(w^k; \theta'_w) + h(m^k; \theta_w^m) \quad (10)$$

によって与えられ得、式中、 $g'_w(w^k; \theta'_w)$  は、アレル非相互作用パラメータ  $\theta'_w$  の

セットを伴うアフィン関数、ネットワーク関数などであり、 $m^k$ は、ペプチド $p^k$ についてのmRNA定量測定値であり、 $h(\cdot)$ は、定量測定値を変換する関数であり、かつ $\theta_w^m$ は、mRNA定量測定値についての依存性スコアを生成するようにmRNA定量測定値と組み合わせられる、アレル非相互作用変数についてのパラメータのセットにおけるパラメータである。本明細書の残りの部分を通じて言及される1つの特定の実施形態において、 $h(\cdot)$ はlog関数であるが、実際には、 $h(\cdot)$ は、様々な異なる関数のうちのいずれか1つであり得る。

【0351】

さらに別の例において、アレル非相互作用変数についての依存性関数 $g_w(\cdot)$ は、

$$g_w(w^k; \theta_w) = g'_w(w^k; \theta'_w) + \theta_w^o \cdot o^k \quad (11)$$

10

によって与えられ、式中、 $g'_w(w^k; \theta'_w)$ は、アレル非相互作用パラメータ $\theta'_w$ のセットを伴うアフィン関数、ネットワーク関数などであり、 $o^k$ は、ペプチド $p^k$ についてヒトプロテオームにおけるタンパク質及びアイソフォームを表す、セクションV E E . C . 2で述べた指標ベクトルであり、かつ $\theta_w^o$ は、指標ベクトルと組み合わせられるアレル非相互作用変数についてのパラメータのセットにおける、パラメータのセットである。1つのバリエーションにおいて、 $o^k$ 及びパラメータ $\theta_w^o$ のセットの次元が有意に高い場合

、 $\lambda \cdot ||\theta^o||$  ( $||\cdot||$ は、L1ノルム、L2ノルム、組み合わせなどを表す)などのパラメータ正則化項を、パラメータの値を決定する時に損失関数に加えることができる。ハイパーパラメータ $\lambda$ の最適値を、適切な方法を通して決定することができる。

20

【0352】

さらに別の例において、アレル非相互作用変数に対する依存性関数 $g_w(\cdot)$ は下式により与えられる。すなわち、

$$g_w(w^k; \theta_w) = g'_w(w^k; \theta'_w) + \sum_{l=1}^L \mathbb{1}(\text{遺伝子}(p^k = l)) \cdot \theta_w^l \quad (12)$$

ただし、 $g'_w(w^k; \theta'_w)$ は、アレル非相互作用パラメータ $\theta'_w$ のセットを伴うアフィン関数、ネットワーク関数などであり、

30

$\mathbb{1}(\text{遺伝子}(p^k = l))$

は、ペプチド $p^k$ がアレル非相互作用変数に関して上記に述べたソース遺伝子 $l$ に由来するものである場合に1に等しいインジケータ関数であり、 $\theta_w^l$ はソース遺伝子 $l$ の「抗原性」を示すパラメータである。1つのバリエーションにおいて、 $L$ が十分に大きく、したがって、パラメータの数 $\theta_w^l = 1, 2, \dots, L$ が十分に大きい場合、 $\lambda \cdot ||\theta^l||$ のようなパラメータ正則化項（ただし、 $||\cdot||$ は、L1ノルム、L2ノルム、組み合わせなど）をパラメータの値を決定する際に損失関数に加えることができる。ハイパーパラメータ $\lambda$ の最適値は適当な方法によって決定することができる。

40

【0353】

さらに別の例において、アレル非相互作用変数に対する依存性関数 $g_w(\cdot)$ は下式により与えられる。すなわち、

$$g_w(w^k; \theta_w) = g'_w(w^k; \theta'_w) + \sum_{m=1}^M \sum_{l=1}^L \mathbb{1}(\text{遺伝子}(p^k) = l, \text{組織}(p^k) = m) \cdot \theta_w^{lm} \quad (12b)$$

ただし、 $g'_w(w^k; \theta'_w)$ は、アレル非相互作用パラメータ $\theta'_w$ のセットを伴うアフィン関数、ネットワーク関数などであり、

$\mathbb{1}(\text{遺伝子}(p^k) = l, \text{組織}(p^k) = m)$

50

は、アレル非相互作用変数に関して上記に述べたようにペプチド  $p^k$  がソース遺伝子  $l$  に由来するものである場合、かつペプチド  $p^k$  が組織タイプ  $m$  に由来するものである場合に 1 に等しいインジケータ関数であり、 $\theta_w^{lm}$  はソース遺伝子  $l$  と組織タイプ  $m$  との組み合わせの抗原性を示すパラメータである。詳細には、組織タイプ  $m$  の遺伝子  $l$  の抗原性は、組織タイプ  $m$  の細胞が、RNA 発現及びペプチド配列コンテキストについての調節後に遺伝子  $l$  由来のペプチドを提示する残留傾向を示し得る。

【0354】

1つのバリエーションにおいて、 $L$  または  $M$  が十分に大きく、したがって、パラメータの数  $\theta_w^{lm} = 1, 2, \dots, LM$  が十分に大きい場合、 $\lambda \cdot \|\theta_w^{lm}\|$  のようなパラメータ正則化項（ただし、 $\|\cdot\|$  は、 $L_1$  ノルム、 $L_2$  ノルム、組み合わせなど）をパラメータの値を決定する際に損失関数に加えることができる。ハイパーパラメータ  $\lambda$  の最適値は適当な方法によって決定することができる。別のバリエーションにおいて、同じソース遺伝子に対する係数が組織タイプ間で大きく異ならないように、パラメータの値を決定する際にパラメータ正則化項を損失関数に加えることができる。例えば、以下のようなペナルティ項：

$$\lambda \cdot \sum_{l=1}^L \sqrt{\sum_{m=1}^M (\theta_w^{lm} - \bar{\theta}_w^l)^2}$$

（ただし、

$$\bar{\theta}_w^l$$

はソース遺伝子  $l$  の組織タイプにわたった平均の抗原性である）は、損失関数中の異なる組織タイプにわたった抗原性の標準偏差にペナルティを付加することができる。

【0355】

実際には、式 (10)、(11)、(12a) 及び (12b) のいずれかの追加項を組み合わせることによってアレル非相互作用変数に関する依存性関数  $g_w(\cdot)$  を生成することができる。例えば、式 (10) の mRNA 定量測定値を示す項  $h(\cdot)$  と式 (12) のソース遺伝子の抗原性を示す項とを他の任意のアフィン関数またはネットワーク関数とともに互いに加え合わせることで、アレル非相互作用変数に関する依存性関数を生成することができる。

【0356】

例として、等式 (8) に戻ると、アフィン変換関数  $g_h(\cdot)$ 、 $g_w(\cdot)$  を用いた、 $m=4$  の異なる特定された MHC アレルの中で MHC アレル  $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k^3 = f(w^k \cdot \theta_w + x_3^k \cdot \theta_3)$$

によって生成することができ、式中、 $w^k$  は、ペプチド  $p^k$  について特定されたアレル非相互作用変数であり、 $\theta_w$  は、アレル非相互作用変数について決定されたパラメータのセットである。

【0357】

別の例として、ネットワーク変換関数  $g_h(\cdot)$ 、 $g_w(\cdot)$  を用いた、 $m=4$  の異なる特定された MHC アレルの中で MHC アレル  $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k^3 = f(NN_w(w^k; \theta_w) + NN_3(x_3^k; \theta_3))$$

によって生成することができ、式中、 $w^k$  は、ペプチド  $p^k$  について特定されたアレル相互作用変数であり、 $\theta_w$  は、アレル非相互作用変数について決定されたパラメータのセットである。

【0358】

図8は、例示的なネットワークモデル $NN_3(\cdot)$ 及び $NN_w(\cdot)$ を用いた、MHCアレル $h=3$ に関連したペプチド $p^k$ の提示尤度の生成を説明する。図8に示すように、ネットワークモデル $NN_3(\cdot)$ は、MHCアレル $h=3$ についてのアレル相互作用変数 $x_3^k$ を受け取り、出力 $NN_3(x_3^k)$ を生成する。ネットワークモデル $NN_w(\cdot)$ は、ペプチド $p^k$ についてのアレル非相互作用変数 $w^k$ を受け取り、出力 $NN_w(w^k)$ を生成する。出力は、組み合わせられ、関数 $f(\cdot)$ によってマッピングされて、推定提示尤度 $u_k$ を生成する。

【0359】

#### VIII. C. 複数アレルモデル

訓練モジュール316はまた、2つ以上のMHCアレルが存在する複数アレル設定においてペプチドの提示尤度を予測するための提示モデルを構築し得る。この例において、訓練モジュール316は、単一のMHCアレルを発現する細胞、複数のMHCアレルを発現する細胞、またはそれらの組み合わせから生成された訓練データ170におけるデータ例Sに基づいて、提示モデルを訓練し得る。

【0360】

##### VIII. C. 1. 実施例1：アレルごとのモデルの最大値

一実現形態では、訓練モジュール316は、複数のMHCアレル $H$ のセットに関連したペプチド $p^k$ の推定提示尤度 $u_k$ を、等式(2)～(11)と共に上記で説明したような、単一アレルを発現する細胞に基づいて決定されたセット $H$ におけるMHCアレル $h$ の各々について決定された提示尤度 $u_k^{h \in H}$ の関数としてモデル化する。具体的には、提示尤度 $u_k$ は、 $u_k^{h \in H}$ の任意の関数であることができる。一実現形態では、等式(12)に示すように、関数は最大値関数であり、提示尤度 $u_k$ は、セット $H$ における各MHCアレル $h$ についての提示尤度の最大値として決定することができる。

$$u_k = \Pr(p^k \text{ 提示}; \text{アレル } H) = \max(u_k^{h \in H})$$

【0361】

##### VIII. C. 2. 実施例2. 1：和の関数モデル

一実現形態では、訓練モジュール316は、ペプチド $p^k$ の推定提示尤度 $u_k$ を、

$$u_k = \Pr(p^k \text{ 提示}) = f\left(\sum_{h=1}^m a_h^k \cdot g_h(x_h^k; \theta_h)\right) \quad (13)$$

によってモデル化し、式中、要素 $a_h^k$ は、ペプチド配列 $p^k$ に関連する複数のMHCアレル $H$ について1であり、 $x_h^k$ は、ペプチド $p^k$ 及び対応するMHCアレルについてのコード化されたアレル相互作用変数を意味する。各MHCアレル $h$ についてのパラメータ $\theta_h$ のセットの値は、 $\theta_h$ に関する損失関数を最小化することによって決定することができる。iは、単一のMHCアレルを発現する細胞及び／または複数のMHCアレルを発現する細胞から生成された訓練データ170のサブセットSにおける各例である。依存性関数 $g_h$ は、セクションVIII. B. 1において上記で導入された依存性関数 $g_h$ のいずれかの形態であり得る。

【0362】

等式(13)によると、ペプチド配列 $p^k$ が1つ以上のMHCアレル $h$ によって提示されるであろう提示尤度は、依存性関数 $g_h(\cdot)$ を、MHCアレル $H$ の各々についてペプチド配列 $p^k$ のコード化されたバージョンに適用して、アレル相互作用変数についての対応するスコアを生成することによって、生成することができる。各MHCアレル $h$ についてのスコアが組み合わせられて、ペプチド配列 $p^k$ がMHCアレル $H$ のセットによって提示されるであろう提示尤度を生成するように変換関数 $f(\cdot)$ によって変換される。

【0363】

等式(13)の提示モデルは、各ペプチド $p^k$ についての関連するアレルの数が1より

10

20

30

40

50

も大きいことができる点で、等式（２）のアレルごとのモデルとは異なる。換言すると、 $a_h^k$ における１つよりも多い要素が、ペプチド配列  $p^k$  に関連する複数の MHC アレル H について１の値を有することができる。

【０３６４】

例として、アフィン変換関数  $g_h(\cdot)$  を用いた、 $m=4$  の異なる特定された MHC アレルの中で MHC アレル  $h=2$ 、 $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k = f(x_2^k \cdot \theta_2 + x_3^k \cdot \theta_3)$$

によって生成することができ、式中、 $x_2^k$ 、 $x_3^k$  は、MHC アレル  $h=2$ 、 $h=3$  について特定されたアレル相互作用変数であり、 $\theta_2$ 、 $\theta_3$  は、MHC アレル  $h=2$ 、 $h=3$  について決定されたパラメータのセットである。

【０３６５】

別の例として、ネットワーク変換関数  $g_h(\cdot)$ 、 $g_w(\cdot)$  を用いた、 $m=4$  の異なる特定された MHC アレルの中で MHC アレル  $h=2$ 、 $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k = f(NN_2(x_2^k; \theta_2) + NN_3(x_3^k; \theta_3))$$

によって生成することができ、式中、 $NN_2(\cdot)$ 、 $NN_3(\cdot)$  は、MHC アレル  $h=2$ 、 $h=3$  について特定されたネットワークモデルであり、 $\theta_2$ 、 $\theta_3$  は、MHC アレル  $h=2$ 、 $h=3$  について決定されたパラメータのセットである。

【０３６６】

図 9 は、例示的なネットワークモデル  $NN_2(\cdot)$  及び  $NN_3(\cdot)$  を用いた、MHC アレル  $h=2$ 、 $h=3$  に関連したペプチド  $p^k$  の提示尤度の生成を説明する。図 9 に示すように、ネットワークモデル  $NN_2(\cdot)$  は、MHC アレル  $h=2$  についてのアレル相互作用変数  $x_2^k$  を受け取り、出力  $NN_2(x_2^k)$  を生成し、ネットワークモデル  $NN_3(\cdot)$  は、MHC アレル  $h=3$  についてのアレル相互作用変数  $x_3^k$  を受け取り、出力  $NN_3(x_3^k)$  を生成する。出力は、組み合わせられ、関数  $f(\cdot)$  によってマッピングされて、推定提示尤度  $u_k$  を生成する。

【０３６７】

#### V I I I . C . ３ . 実施例 ２ . ２ : アレル非相互作用変数を伴う和の関数モデル

一実現形態では、訓練モジュール 316 は、アレル非相互作用変数を組み入れて、

$$u_k = \Pr(p^k \text{ 提示}) = f\left(g_w(w^k; \theta_w) + \sum_{h=1}^m a_h^k \cdot g_h(x_h^k; \theta_h)\right) \quad (14)$$

によって、ペプチド  $p^k$  の推定提示尤度  $u_k$  をモデル化し、式中、 $w^k$  は、ペプチド  $p^k$  についてのコード化されたアレル非相互作用変数を意味する。具体的には、各 MHC アレル  $h$  についてのパラメータ  $\theta_h$  のセット及びアレル非相互作用変数についてのパラメータ  $\theta_w$  のセットの値を、 $\theta_h$  及び  $\theta_w$  に関する損失関数を最小化することによって決定することができ、 $i$  は、単一の MHC アレルを発現する細胞及び／または複数の MHC アレルを発現する細胞から生成された訓練データ 170 のサブセット  $S$  における各例である。依存性関数  $g_w$  は、セクション V I I I . B . 3 において上記で導入された依存性関数  $g_w$  のいずれかの形態であり得る。

【０３６８】

したがって、等式（１４）によると、１つ以上の MHC アレル H によってペプチド配列  $p^k$  が提示されるであろう提示尤度は、関数  $g_h(\cdot)$  を、MHC アレル H の各々についてペプチド配列  $p^k$  のコード化されたバージョンに適用して、各 MHC アレル  $h$  のアレル相互作用変数について対応する依存性スコアを生成することによって、生成することができる。アレル非相互作用変数についての関数  $g_w(\cdot)$  もまた、アレル非相互作用変数に

ついでに依存性スコアを生成するように、アレル非相互作用変数のコード化されたバージョンに適用される。スコアが組み合わされ、組み合わされたスコアが、MHCアレルHによってペプチド配列  $p^k$  が提示されるであろう提示尤度を生成するように、変換関数  $f(\cdot)$  によって変換される。

【0369】

等式(14)の提示モデルにおいて、各ペプチド  $p^k$  についての関連するアレルの数は、1よりも大きいことができる。換言すると、 $a_h^k$  における1つよりも多い要素が、ペプチド配列  $p^k$  に関連する複数のMHCアレルHについて1の値を有することができる。

【0370】

例として、アフィン変換関数  $g_h(\cdot)$ 、 $g_w(\cdot)$  を用いた、 $m=4$  の異なる特定されたMHCアレルの中でMHCアレル  $h=2$ 、 $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k = f(w^k \cdot \theta_w + x_2^k \cdot \theta_2 + x_3^k \cdot \theta_3)$$

によって生成することができ、式中、 $w^k$  は、ペプチド  $p^k$  について特定されたアレル非相互作用変数であり、 $\theta_w$  は、アレル非相互作用変数について決定されたパラメータのセットである。

【0371】

別の例として、ネットワーク変換関数  $g_h(\cdot)$ 、 $g_w(\cdot)$  を用いた、 $m=4$  の異なる特定されたMHCアレルの中でMHCアレル  $h=2$ 、 $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k = f(NN_w(w^k; \theta_w) + NN_2(x_2^k; \theta_2) + NN_3(x_3^k; \theta_3))$$

によって生成することができ、式中、 $w^k$  は、ペプチド  $p^k$  について特定されたアレル相互作用変数であり、 $\theta_w$  は、アレル非相互作用変数について決定されたパラメータのセットである。

【0372】

図10は、例示的なネットワークモデル  $NN_2(\cdot)$ 、 $NN_3(\cdot)$ 、及び  $NN_w(\cdot)$  を用いた、図10に示すように、ネットワークモデル  $NN_2(\cdot)$  は、MHCアレル  $h=2$  についてのアレル相互作用変数  $x_2^k$  を受け取り、出力  $NN_2(x_2^k)$  を生成する。ネットワークモデル  $NN_3(\cdot)$  は、MHCアレル  $h=3$  についてのアレル相互作用変数  $x_3^k$  を受け取り、出力  $NN_3(x_3^k)$  を生成する。ネットワークモデル  $NN_w(\cdot)$  は、ペプチド  $p^k$  についてのアレル非相互作用変数  $w^k$  を受け取り、出力  $NN_w(w^k)$  を生成する。出力は、組み合わされ、関数  $f(\cdot)$  によってマッピングされて、推定提示尤度  $u_k$  を生成する。

【0373】

あるいは、訓練モジュール316は、等式(15)においてアレル非相互作用変数  $w^k$  をアレル相互作用変数  $x_h^k$  に付加することにより、予測におけるアレル非相互作用変数  $w^k$  を含んでもよい。したがって、提示尤度は、

$$u_k = \Pr(p^k \text{ 提示}) = f\left(\sum_{h=1}^m a_h^k \cdot g_h([x_h^k w^k]; \theta_h)\right) \quad (15)$$

によって与えられ得る。

【0374】

#### VIII.C.4. 実施例3.1：潜在的なアレルごとの尤度を用いたモデル

別の実現形態において、訓練モジュール316は、ペプチド  $p^k$  の推定提示尤度  $u_k$  を、

$$u_k = \Pr(p^k \text{ 提示}) = r\left(s(v = [a_1^k \cdot u_k^{r1}(\theta) \dots a_m^k \cdot u_k^{rm}(\theta)])\right) \quad (16)$$

10

20

30

40

50

によってモデル化し、式中、要素  $a_{h^k}$  は、ペプチド配列  $p^k$  に関連する複数の MHC アレル  $h \in H$  について 1 であり、 $u_k^h$  は、MHC アレル  $h$  についての潜在的なアレルごとの提示尤度であり、ベクトル  $v$  は、要素  $v_h$  が、 $a_{h^k} \cdot u_k^h$  に対応するベクトルであり、 $s(\cdot)$  は、 $v$  の要素をマッピングする関数であり、かつ  $r(\cdot)$  は、入力の値を所定の範囲中にクリップするクリッピング関数である。より詳細に下記に記載するように、 $s(\cdot)$  は、総和関数または二次関数であってもよいが、他の実施形態において、 $s(\cdot)$  は、最大値関数などの任意の関数であり得ることが認識される。潜在的なアレルごとの尤度についてのパラメータ  $\theta$  のセットの値は、 $\theta$  に関する損失関数を最小化することによって決定することができ、 $i$  は、単一の MHC アレルを発現する細胞及び／または複数の MHC アレルを発現する細胞から生成された訓練データ 170 のサブセット  $S$  における各例である。

10

【0375】

等式 (17) の提示モデルにおける提示尤度は、各々が、個々の MHC アレル  $h$  によってペプチド  $p^k$  が提示されるであろう尤度に対応する、潜在的なアレルごとの提示尤度  $u_k^h$  の関数としてモデル化される。潜在的なアレルごとの尤度は、潜在的なアレルごとの尤度についてのパラメータが、単一アレル設定に加えて、提示されるペプチドと対応する MHC アレルとの間の直接の関連が未知である複数アレル設定から学習され得る点で、セクション V I I I . B のアレルごとの提示尤度とは異なる。したがって、複数アレル設定において、提示モデルは、ペプチド  $p^k$  が全体として MHC アレル  $H$  のセットによって提示されるかどうかを推定できるだけでなく、どの MHC アレル  $h$  がペプチド  $p^k$  を提示した可能性が最も高いかを示す個々の尤度  $u_k^h \in H$  も提供することもできる。この利点は、提示モデルが、単一の MHC アレルを発現する細胞についての訓練データを伴わずに潜在的な尤度を生成できることである。

20

【0376】

本明細書の残りの部分を通じて言及される 1 つの特定の実現形態において、 $r(\cdot)$  は、範囲  $[0, 1]$  を有する関数である。例えば、 $r(\cdot)$  は、クリップ関数：

$$r(z) = \min(\max(z, 0), 1)$$

であってもよく、 $z$  と 1 の間の最小値が、提示尤度  $u_k$  として選ばれる。別の実現形態において、 $r(\cdot)$  は、

$$r(z) = \tanh(z)$$

30

として与えられる双曲線正接関数であり、ドメイン  $z$  の値は、0 以上である。

【0377】

#### V I I I . C . 5 . 実施例 3 . 2 : 関数の和モデル

1 つの特定の実現形態において、 $s(\cdot)$  は、総和関数であり、提示尤度は、潜在的なアレルごとの提示尤度を総和することによって与えられる。

$$u_k = \Pr(p^k \text{ 提示}) = r\left(\sum_{h=1}^m a_h^k \cdot u_k^h(\theta)\right) \quad (17)$$

【0378】

40

1 つの実現形態では、MHC アレル  $h$  についての潜在的なアレルごとの提示尤度を、

$$u_k^h = f(g_h(x_h^k, \theta_h)) \quad (18)$$

によって生成して、提示尤度が、

$$u_k = \Pr(p^k \text{ 提示}) = r\left(\sum_{h=1}^m a_h^k \cdot f(g_h(x_h^k, \theta_h))\right) \quad (19)$$

によって推定されるようにする。

50

## 【0379】

等式(19)によると、1つ以上のMHCアレルHによってペプチド配列 $p^k$ が提示されるであろう提示尤度は、関数 $g_h(\cdot)$ を、MHCアレルHの各々についてペプチド配列 $p^k$ のコード化されたバージョンに適用して、アレル相互作用変数についての対応する依存性スコアを生成することによって、生成することができる。各依存性スコアは、最初に、潜在的なアレルごとの提示尤度 $u_k^h$ を生成するように、関数 $f(\cdot)$ によって変換される。アレルごとの尤度 $u_k^h$ が組み合わせられ、組み合わせられた尤度にクリッピング関数が、値を範囲 $[0, 1]$ 中にクリップするために適用されて、ペプチド配列 $p^k$ がMHCアレルHのセットによって提示されるであろう提示尤度が生成され得る。依存性関数 $g_h$ は、セクションVII I . B . 1において上記で導入された依存性関数 $g_h$ のいずれかの形態であり得る。

10

## 【0380】

例として、アフィン変換関数 $g_h(\cdot)$ を用いた、 $m=4$ の異なる特定されたMHCアレルの中でMHCアレル $h=2$ 、 $h=3$ によってペプチド $p^k$ が提示されるであろう尤度は、

$$u_k = r\left(f(x_2^k \cdot \theta_2) + f(x_3^k \cdot \theta_3)\right)$$

によって生成することができ、式中、 $x_2^k$ 、 $x_3^k$ は、MHCアレル $h=2$ 、 $h=3$ について特定されたアレル相互作用変数であり、 $\theta_2$ 、 $\theta_3$ は、MHCアレル $h=2$ 、 $h=3$ について決定されたパラメータのセットである。

20

## 【0381】

別の例として、ネットワーク変換関数 $g_h(\cdot)$ 、 $g_w(\cdot)$ を用いた、 $m=4$ の異なる特定されたMHCアレルの中でMHCアレル $h=2$ 、 $h=3$ によってペプチド $p^k$ が提示されるであろう尤度は、

$$u_k = r\left(f(NN_2(x_2^k; \theta_2)) + f(NN_3(x_3^k; \theta_3))\right)$$

によって生成することができ、式中、 $NN_2(\cdot)$ 、 $NN_3(\cdot)$ は、MHCアレル $h=2$ 、 $h=3$ について特定されたネットワークモデルであり、 $\theta_2$ 、 $\theta_3$ は、MHCアレル $h=2$ 、 $h=3$ について決定されたパラメータのセットである。

30

## 【0382】

図11は、例示的なネットワークモデル $NN_2(\cdot)$ 及び $NN_3(\cdot)$ を用いた、MHCアレル $h=2$ 、 $h=3$ に関連したペプチド $p^k$ の提示尤度の生成を説明する。図9に示すように、ネットワークモデル $NN_2(\cdot)$ は、MHCアレル $h=2$ についてのアレル相互作用変数 $x_2^k$ を受け取り、出力 $NN_2(x_2^k)$ を生成し、ネットワークモデル $NN_3(\cdot)$ は、MHCアレル $h=3$ についてのアレル相互作用変数 $x_3^k$ を受け取り、出力 $NN_3(x_3^k)$ を生成する。各出力は、関数 $f(\cdot)$ によってマッピングされ、組み合わせられて、推定提示尤度 $u_k$ を生成する。

## 【0383】

別の実現形態において、予測が、質量分析イオン電流の $\log$ についてなされる場合、 $r(\cdot)$ は $\log$ 関数であり、 $f(\cdot)$ は指数関数である。

40

## 【0384】

V I I I . C . 6 . 実施例3.3：アレル非相互作用変数を伴う関数の和モデル

1つの実現形態では、MHCアレル $h$ についての潜在的なアレルごとの提示尤度を、

$$u_k^h = f\left(g_h(x_h^k; \theta_h) + g_w(w^k; \theta_w)\right) \quad (20)$$

によって生成して、提示尤度が、

50

$$u_k = \Pr(p^k \text{ 提示}) = r \left( \sum_{h=1}^m a_h^k \cdot f \left( g_w(w^k; \theta_w) + g_h(x_h^k; \theta_h) \right) \right) \quad (21)$$

によって生成されるようにして、ペプチド提示に、アレル非相互作用変数の影響を組み入れる。

【0385】

等式(21)によると、1つ以上のMHCアレルHによってペプチド配列  $p^k$  が提示されるであろう提示尤度は、関数  $g_h(\cdot)$  を、MHCアレルHの各々についてペプチド配列  $p^k$  のコード化されたバージョンに適用して、各MHCアレルhのアレル相互作用変数について対応する依存性スコアを生成することによって、生成することができる。アレル非相互作用変数についての関数  $g_w(\cdot)$  もまた、アレル非相互作用変数についての依存性スコアを生成するように、アレル非相互作用変数のコード化されたバージョンに適用される。アレル非相互作用変数のスコアが、アレル相互作用変数の依存性スコアの各々に組み合わせられる。組み合わせられたスコアの各々が、潜在的なアレルごとの提示尤度を生成するように、関数  $f(\cdot)$  によって変換される。潜在的な尤度が組み合わせられ、組み合わせられた出力にクリッピング関数が、値を範囲[0, 1]中にクリップするために適用されて、MHCアレルHによってペプチド配列  $p^k$  が提示されるであろう提示尤度が生成され得る。依存性関数  $g_w$  は、セクションV.I.II.B.3において上記で導入された依存性関数  $g_w$  のいずれかの形態であり得る。

【0386】

例として、アフィン変換関数  $g_h(\cdot)$ 、 $g_w(\cdot)$  を用いた、 $m=4$  の異なる特定されたMHCアレルの中でMHCアレル  $h=2$ 、 $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k = r \left( f(w^k \cdot \theta_w + x_2^k \cdot \theta_2) + f(w^k \cdot \theta_w + x_3^k \cdot \theta_3) \right)$$

によって生成することができ、式中、 $w^k$  は、ペプチド  $p^k$  について特定されたアレル非相互作用変数であり、 $\theta_w$  は、アレル非相互作用変数について決定されたパラメータのセットである。

【0387】

別の例として、ネットワーク変換関数  $g_h(\cdot)$ 、 $g_w(\cdot)$  を用いた、 $m=4$  の異なる特定されたMHCアレルの中でMHCアレル  $h=2$ 、 $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k = r \left( f(NN_w(w^k; \theta_w) + NN_2(x_2^k; \theta_2)) + f(NN_w(w^k; \theta_w) + NN_3(x_3^k; \theta_3)) \right)$$

によって生成することができ、式中、 $w^k$  は、ペプチド  $p^k$  について特定されたアレル相互作用変数であり、 $\theta_w$  は、アレル非相互作用変数について決定されたパラメータのセットである。

【0388】

図12は、例示的なネットワークモデル  $NN_2(\cdot)$ 、 $NN_3(\cdot)$ 、及び  $NN_w(\cdot)$  を用いた、MHCアレル  $h=2$ 、 $h=3$  に関連したペプチド  $p^k$  の提示尤度の生成を説明する。図12に示すように、ネットワークモデル  $NN_2(\cdot)$  は、MHCアレル  $h=2$  についてのアレル相互作用変数  $x_2^k$  を受け取り、出力  $NN_2(x_2^k)$  を生成する。ネットワークモデル  $NN_w(\cdot)$  は、ペプチド  $p^k$  についてのアレル非相互作用変数  $w^k$  を受け取り、出力  $NN_w(w^k)$  を生成する。出力は、組み合わせられ、関数  $f(\cdot)$  によってマッピングされる。ネットワークモデル  $NN_3(\cdot)$  は、MHCアレル  $h=3$  についてのアレル相互作用変数  $x_3^k$  を受け取り、出力  $NN_3(x_3^k)$  を生成し、これも、同じネットワークモデル  $NN_w(\cdot)$  の出力  $NN_w(w^k)$  と組み合わせられ、関数  $f(\cdot)$  によってマッ

ピングされる。両方の出力が組み合わされて、推定提示尤度  $u_k$  を生成する。

【0389】

別の実現形態では、MHCアレル  $h$  についての潜在的なアレルごとの提示尤度を、

$$u_k^h = f(g_h([x_h^k w^k]; \theta_h)) \quad (22)$$

によって生成して、提示尤度が、

$$u_k = \Pr(p^k \text{ 提示}) = r\left(\sum_{h=1}^m a_h^k \cdot f(g_h([x_h^k w^k]; \theta_h))\right)$$

10

によって生成されるようにする。

【0390】

#### VIII. C. 7. 実施例4：二次モデル

一実現形態では、 $s(\cdot)$  は、二次関数であり、ペプチド  $p^k$  の推定提示尤度  $u_k$  は、

$$u_k = \Pr(p^k \text{ 提示}) = \sum_{h=1}^m a_h^k \cdot u_k^{th}(\theta) - \sum_{h=1}^m \sum_{j < h} a_h^k \cdot a_j^k \cdot u_k^{th}(\theta) \cdot u_k^{tj}(\theta) \quad (23)$$

によって与えられ、式中、要素  $u_k^h$  は、MHCアレル  $h$  についての潜在的なアレルごとの提示尤度である。潜在的なアレルごとの尤度についてのパラメータ  $\theta$  のセットの値は、 $\theta$  に関する損失関数を最小化することによって決定することができ、 $i$  は、単一のMHCアレルを発現する細胞及び／または複数のMHCアレルを発現する細胞から生成された訓練データ170のサブセット  $S$  における各例である。潜在的なアレルごとの提示尤度は、上記の等式(18)、(20)、及び(22)において示すいずれかの形態であり得る。

20

【0391】

一態様において、等式(23)のモデルは、ペプチド配列  $p^k$  が、2つのMHCアレルによって同時に提示されるであろう可能性が存在し、2つのHLAアレルによる提示は統計学的に独立していることを含意し得る。

【0392】

等式(23)によると、1つ以上のMHCアレル  $H$  によってペプチド配列  $p^k$  が提示されるであろう提示尤度は、潜在的なアレルごとの提示尤度を組み合わせること、及び、MHCアレル  $H$  によってペプチド配列  $p^k$  が提示されるであろう提示尤度を生成するように、MHCアレルの各ペアがペプチド  $p^k$  を同時に提示するであろう尤度を総和から差し引くことによって、生成することができる。

30

【0393】

例として、アフィン変換関数  $g_h(\cdot)$  を用いた、 $m=4$  の異なる特定されたHLAアレルの中でHLAアレル  $h=2$ 、 $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k = f(x_2^k \cdot \theta_2) + f(x_3^k \cdot \theta_3) - f(x_2^k \cdot \theta_2) \cdot f(x_3^k \cdot \theta_3)$$

40

によって生成することができ、式中、 $x_2^k$ 、 $x_3^k$  は、HLAアレル  $h=2$ 、 $h=3$  について特定されたアレル相互作用変数であり、 $\theta_2$ 、 $\theta_3$  は、HLAアレル  $h=2$ 、 $h=3$  について決定されたパラメータのセットである。

【0394】

別の例として、ネットワーク変換関数  $g_h(\cdot)$ 、 $g_w(\cdot)$  を用いた、 $m=4$  の異なる特定されたHLAアレルの中でHLAアレル  $h=2$ 、 $h=3$  によってペプチド  $p^k$  が提示されるであろう尤度は、

$$u_k = f(NN_2(x_2^k; \theta_2)) + f(NN_3(x_3^k; \theta_3)) - f(NN_2(x_2^k; \theta_2)) \cdot f(NN_3(x_3^k; \theta_3))$$

50

によって生成することができ、式中、 $NN_2(\cdot)$ 、 $NN_3(\cdot)$ は、HLAアレル $h=2$ 、 $h=3$ について特定されたネットワークモデルであり、 $\theta_2$ 、 $\theta_3$ は、HLAアレル $h=2$ 、 $h=3$ について決定されたパラメータのセットである。

【0395】

#### IX 実施例5：予測モジュール

予測モジュール320は、配列データを受け取って、提示モデルを用いて配列データ中の候補新生抗原を選択する。具体的には、配列データは、患者の腫瘍組織細胞から抽出されたDNA配列、RNA配列、及び／またはタンパク質配列であってよい。予測モジュール320は、配列データを、MHC-Iについては8～15個のアミノ酸を有する、またはMHC-IIについては6～30個のアミノ酸を有する複数のペプチド配列 $p^k$ に処理する。例えば、予測モジュール320は、所定の配列「IEFROEIEFJEF（配列番号16）」を、9個のアミノ酸を有する3種類のペプチド配列「IEFROEIEFJ（配列番号17）」、「EFROEIEFJE（配列番号18）」、及び「FROEIEFJEF（配列番号19）」に処理することができる。一実施形態では、予測モジュール320は、患者の正常組織細胞から抽出された配列データをその患者の腫瘍組織細胞から抽出された配列データと比較して1つ以上の変異を有する部分を特定することによって、変異したペプチド配列である候補新生抗原を特定することができる。

【0396】

予測モジュール320は、提示モデルの1つ以上を処理されたペプチド配列に適用してペプチド配列の提示尤度を推定する。具体的には、予測モジュール320は、提示モデルを候補新生抗原に適用することによって、腫瘍HLA分子上に提示される可能性が高い1つ以上の候補新生抗原ペプチド配列を選択することができる。一実現形態では、予測モジュール320は、あらかじめ決定された閾値を上回る推定提示尤度を有する候補新生抗原配列を選択する。別の実現形態では、提示モデルは、最も高い推定提示尤度を有する $v$ 個の候補新生抗原配列を選択する（ $v$ は、一般的に、ワクチン中で送達することができるエピトープの最大数である）。所定の患者について選択された候補新生抗原を含むワクチンを患者に注射して免疫応答を誘導することができる。

【0397】

#### X. 実施例6：患者選択モジュール

患者選択モジュール324は、患者が選択基準を満たすかどうかに基づいてワクチン治療及び／またはT細胞療法に対する患者のサブセットを選択する。一実施形態では、選択基準は、提示モデルによって生成される患者の新生抗原候補の提示尤度に基づいて決定される。選択基準を調整することにより、患者選択モジュール324は、患者の新生抗原候補の提示尤度に基づいてワクチン投与及び／またはT細胞療法を受ける患者数を調整することができる。具体的には、厳密な選択基準では、ワクチン及び／またはT細胞療法によって治療される患者の数はより少なくなるが、有効な治療（例えば、1つ以上の腫瘍特異的新生抗原（TSNA）及び／または1つ以上の新生抗原応答性T細胞）を受けるワクチン及び／またはT細胞療法による治療患者の比率は高くなり得る。これに対して、緩い選択基準では、ワクチン及び／またはT細胞療法で治療される患者の数はより多くなるが、有効な治療を受けるワクチン及び／またはT細胞療法による治療患者の比率は低くなり得る。患者選択モジュール324は、治療を受ける患者の目標比率と、有効な治療を受ける患者の比率との間の所望のバランスに基づいて選択基準を変更する。

【0398】

いくつかの実施形態では、ワクチン治療を受ける患者を選択するための選択基準は、T細胞療法を受ける患者を選択するための選択基準と同じである。しかしながら、代替的な実施形態では、ワクチン治療を受ける患者を選択するための選択基準は、T細胞療法を受ける患者を選択するための選択基準と異なり得る。以下のセクションX.A及びX.Bでは、ワクチン治療を受ける患者を選択するための選択基準、及びT細胞療法を受ける患者を選択するための選択基準についてそれぞれ検討する。

【0399】

### X. A. ワクチン治療を行う患者の選択

一実施形態では、患者に、ワクチン容量  $v$  を有するその患者に対する個別化ワクチンに潜在的に含ませることが可能な  $v$  個の新生抗原候補の対応する治療サブセットが関連付けられる。一実施形態では、ある患者に対する治療サブセットは、提示モデルによって決定される最も高い提示尤度を有する新生抗原候補である。例えば、ワクチンが  $v = 20$  個のエピトープを含み得る場合、ワクチンは、提示モデルによって決定される最も高い提示尤度を有する各患者の治療サブセットを含み得る。しかしながら、他の実施形態では、ある患者に対する治療サブセットは、他の方法に基づいて決定することもできる点は認識される。例えば、ある患者に対する治療サブセットは、その患者に対する新生抗原候補のセットからランダムに選択することができ、または、ペプチド配列の結合親和性もしくは安定性をモデル化する従来技術のモデル、または提示モデルから得られる提示尤度及びこれらのペプチド配列に関する親和性または安定性情報を含む特定の因子の組み合わせに一部基づいて決定することができる。

【0400】

一実施形態では、患者選択モジュール 324 は、患者の腫瘍変異負荷が最小の変異負荷に等しいかまたはそれよりも高い場合に患者が選択基準を満たすものと判定する。ある患者の腫瘍変異負荷 (TMB) は、腫瘍エクソームの非同義変異の総数を示す。一実施形態では、患者選択モジュール 324 は、患者の TMB の絶対数が所定の閾値に等しいかまたはそれよりも高い場合にワクチン治療を行う患者を選択する。別の実現形態では、患者選択モジュール 324 は、患者の TMB が患者のセットについて決定された TMB 間の閾値パーセンタイル内にある場合にワクチン治療を行う患者を選択する。

【0401】

別の実施形態では、患者選択モジュール 324 は、患者の治療サブセットに基づく患者の効用値スコアが最小の効用値スコアに等しいかまたはそれよりも高い場合に患者が選択基準を満たすものと判定する。一実施形態では、効用値スコアは、治療サブセットからの提示抗原の推定数の尺度である。

【0402】

提示抗原の推定数は、新生抗原の提示を 1 つ以上の確率分布のランダム変数としてモデル化することによって予測することができる。一実現形態では、患者  $i$  の効用値スコアは、治療サブセットからの提示新生抗原候補の期待数、またはその特定の関数である。例として、各新生抗原の提示は、提示 (成功) の確率が新生抗原候補の提示尤度によって与えられるベルヌーイのランダム変数としてモデル化することができる。詳細には、それぞれが最も高い提示尤度  $u_{i1}$ 、 $u_{i2}$ 、 $\dots$ 、 $u_{iv}$  を有する  $v$  個の新生抗原候補  $p_{i1}$ 、 $p_{i2}$ 、 $\dots$ 、 $p_{iv}$  の治療サブセット  $S_i$  について、新生抗原候補  $p_{ij}$  の提示は、ランダム変数  $A_{ij}$  によって与えられ、ここで、

$$P(A_{ij} = 1) = u_{ij}, \quad P(A_{ij} = 0) = 1 - u_{ij}. \quad (24)$$

である。

提示される新生抗原の期待数は、各新生抗原候補の提示尤度の総和により与えられる。換言すれば、患者  $i$  の効用値スコアは、下式として表される。

$$\text{util}_i(S_i) = E \left[ \sum_{j=1}^v A_{ij} \right] = \sum_{j=1}^v u_{ij} \quad (25)$$

患者選択モジュール 324 は、ワクチン治療について最小効用値に等しいかまたはそれよりも高い効用値スコアを有する患者のサブセットを選択する。

【0403】

別の実現形態では、患者  $i$  の効用値スコアは、少なくとも閾値数の新生抗原  $k$  が提示される確率である。1 つの例では、新生抗原候補の治療サブセット  $S_i$  内の提示抗原の数は、提示 (成功) の確率がエピトープのそれぞれの提示尤度によって与えられるポアソン二

項ランダム変数としてモデル化される。詳細には、患者  $i$  の提示抗原の数は、ランダム変数  $N_i$  によって与えることができる。

$$N_i = \sum_{j=1}^v A_{ij} \sim \text{PBD}(u_{i1}, u_{i2}, \dots, u_{iv}) \quad (26)$$

式中、 $\text{PBD}(\cdot)$  は、ポアソン二項分布を示す。少なくとも閾値数の新生抗原  $k$  が提示される確率は、提示抗原の数  $N_i$  が  $k$  に等しいかまたはそれよりも大きい確率の操作によって与えられる。換言すれば、患者  $i$  の効用値スコアは、下式として表される。

$$\text{util}_i(S_i) = \mathbb{P}[N_i \geq k] = \sum_{m=k}^{\infty} \mathbb{P}[N_i = m] \quad (27)$$

10

患者選択モジュール 324 は、ワクチン治療について最小効用値に等しいかまたはそれよりも高い効用値スコアを有する患者のサブセットを選択する。

【0404】

別の実現形態では、患者  $i$  の効用値スコアは、1つ以上の患者の HLA アレルに対して固定閾値（例えば 500 nM）よりも低い結合親和性または予測される結合親和性を有する新生抗原候補の治療サブセット  $S_i$  内の新生抗原の数である。1つの例では、固定閾値は、1000 nM ~ 10 nM の範囲である。場合により、効用値スコアは、RNA-seq によって発現されたものとして検出された新生抗原のみをカウントしてもよい。

20

【0405】

別の実現形態では、患者  $i$  の効用値スコアは、その患者の 1つ以上の HLA アレルに対する結合親和性とその HLA アレルに対するランダムなペプチドの結合親和性の閾値パーセンタイル以下である、新生抗原候補の治療サブセット  $S_i$  内の新生抗原の数である。1つの例では、閾値パーセンタイルは、10 パーセンタイル ~ 0.1 パーセンタイルの範囲である。場合により、効用値スコアは、RNA-seq によって発現されたものとして検出された新生抗原のみをカウントしてもよい。

【0406】

式 (25) 及び (27) に関して説明した効用値スコアの例はあくまで例示的なものに過ぎず、患者選択モジュール 324 は他の統計学または確率分布を用いて効用値スコアを生成することもできる点は認識されよう。

30

【0407】

#### X. B. T細胞療法を行う患者の選択

別の実施形態では、ワクチン治療を受けることに代えて、またはそれに加えて、患者は T細胞療法を受けることができる。ワクチン治療と同様、患者が T細胞療法を受ける実施形態では、患者を上記に述べたような  $v$  個の新生抗原候補の対応する治療サブセットと関連付けることができる。この  $v$  個の新生抗原候補の治療サブセットを、 $v$  個の新生抗原候補のうちの 1つ以上に対する反応性を有する、患者由来の T細胞のインビトロでの特定に用いることができる。次に、これらの特定された T細胞を増殖させて個別化 T細胞療法において患者に注入することができる。

40

【0408】

2つの異なる時点において T細胞療法を受ける患者を選択することができる。第1の時点は、患者に対する  $v$  個の新生抗原候補の治療サブセットがモデルを用いて予測された後であるが、予測された  $v$  個の新生抗原候補の治療サブセットに対して特異的である T細胞のインビトロスクリーニングを行う前である。第2の時点は、予測された  $v$  個の新生抗原候補の治療サブセットに対して特異的である T細胞のインビトロスクリーニングを行った後である。

【0409】

最初に、その患者に対する  $v$  個の新生抗原候補の治療サブセットが予測された後で、か

50

つ予測された $v$ 個の新生抗原候補の治療サブセットに対して特異的である患者由来のT細胞のインビトロでの特定を行う前に、T細胞療法を受ける患者を選択することができる。詳細には、患者由来の新生抗原特異的なT細胞のインビトロスクリーニングはコストが嵩み得るため、患者が新生抗原特異的なT細胞を有する可能性が高い場合にのみ、新生抗原特異的なT細胞についてスクリーニングする患者を選択することが望ましいと考えられる。インビトロT細胞スクリーニングステップの前に患者を選択するには、ワクチン治療を行う患者を選択するために用いられるものと同じ基準を用いることができる。詳細には、いくつかの実施形態において、患者選択モジュール324は、患者の腫瘍変異負荷が上記に述べたような最小の変異負荷に等しいかまたはそれよりも高い場合にT細胞療法を受ける患者を選択することができる。別の実施形態では、患者選択モジュール324は、その患者

10

#### 【0410】

第2に、予測された $v$ 個の新生抗原候補の治療サブセットに対して特異的である患者由来のT細胞のインビトロ特定を行う前にT細胞療法を受ける患者を選択することに加えて、またはそれに代えて、予測された $v$ 個の新生抗原候補の治療サブセットに対して特異的であるT細胞のインビトロ特定を行った後でT細胞療法を受ける患者を選択することもできる。詳細には、ある患者は、新生抗原認識についての患者のT細胞のインビトロスクリーニングにおいてその患者について新生抗原特異的なTCRの少なくとも閾値量が特定された場合にT細胞療法を受けるものとして選択することができる。例えば、ある患者は、少なくとも2つの新生抗原特異的なTCRがその患者について特定された場合にのみ、または2つの異なる新生抗原に対して新生抗原特異的なTCRが特定された場合にのみ、T細胞療法を受けるものとして選択することができる。

20

#### 【0411】

別の実施形態では、ある患者は、その患者に対する $v$ 個の新生抗原候補の治療サブセットの閾値量の新生抗原が患者のTCRによって認識された場合にのみ、T細胞療法を受けるものとして選択することができる。例えば、ある患者は、その患者に対する $v$ 個の新生抗原候補の治療サブセットのうちの少なくとも1つの新生抗原が患者のTCRによって認識された場合にのみ、T細胞療法を受けるものとして選択することができる。さらなる実施形態では、ある患者は、その患者に対するTCRの少なくとも閾値量が、特定のHLA制限クラスの新生抗原ペプチドに対して新生抗原特異的なものとして特定された場合にのみ、T細胞療法を受けるものとして選択することができる。例えば、ある患者は、その患者に対する少なくとも1つのTCRが、新生抗原特異的なHLAクラスI制限新生抗原ペプチドとして特定された場合にのみ、T細胞療法を受けるものとして選択することができる。

30

#### 【0412】

いっそうさらなる実施形態では、ある患者は、特定のHLA制限クラスの新生抗原ペプチドの少なくとも閾値量が患者のTCRによって認識された場合にのみ、T細胞療法を受けるものとして選択することができる。例えば、ある患者は、少なくとも1つのHLAクラスI制限新生抗原ペプチドがその患者のTCRによって認識された場合にのみ、T細胞療法を受けるものとして選択することができる。例えば、ある患者は、少なくとも2つのHLAクラスII制限新生抗原ペプチドがその患者のTCRによって認識された場合にのみ、T細胞療法を受けるものとして選択することができる。上記の基準の任意の組み合わせを、その患者について予測された $v$ 個の新生抗原候補の治療サブセットに対して特異的であるT細胞をインビトロで特定した後でT細胞療法を受ける患者を選択するために用いることもできる。

40

#### 【0413】

### XI. 実施例7：例示的な患者選択性能を示す実験結果

セクションXで述べた患者選択の妥当性を、質量分析データにおいてシミュレートした

50

新生抗原のサブセットが提示されていることが分かっている、シミュレートした新生抗原候補の試験セットがそれぞれに関連付けられたシミュレートした患者のセットで患者の選択を行うことにより検証する。詳細には、試験セット内のそれぞれのシミュレートした新生抗原候補に、その新生抗原がバッサニー・スターンバーグデータセット（データセット「D1」）（データは、[www.ebi.ac.uk/pride/archive/projects/PXD0000394](http://www.ebi.ac.uk/pride/archive/projects/PXD0000394)にみることができる）からの複数アレルJY細胞株HLA-A\*02:01及びHLA-B\*07:02の質量分析データセットにおいて提示されているかどうかを示すラベル（標識）に関連付ける。図13Aとともに下記に詳細に述べるように、シミュレートした患者について多数の新生抗原候補を、非小細胞肺癌（NSCLC）患者における変異負荷の既知の度数分布に基づいてヒトプロテオームからサンプリングする。

10

#### 【0414】

同じHLAアレルについてのアレルごとの提示モデルを、IEDBデータセット（データセット「D2」）（データは、[http://www.iedb.org/doc/mhc\\_ligand\\_full.zip](http://www.iedb.org/doc/mhc_ligand_full.zip)にみることができる）からの単一アレルHLA-A\*02:01及びHLA-B\*07:02の質量分析データのサブセットである訓練セットを用いて訓練する。詳細には、各アレルの提示モデルを、N末端側及びC末端側のランキング配列をアレル非相互作用変数として、ネットワーク依存性関数 $g_h(\cdot)$ 及び $g_w(\cdot)$ ならびに $\exp$ it関数 $f(\cdot)$ とともに組み入れた式（8）に示されるアレルごとのモデルとした。アレルHLA-A\*02:01の提示モデルは、アレル相互作用変数としてペプチド配列が、アレル非相互作用変数としてN末端側及びC末端側のランキング配列が与えられるものとして、特定のペプチドがアレルHLA-A\*02:01上に提示される提示尤度を生成する。アレルHLA-B\*07:02の提示モデルは、アレル相互作用変数としてペプチド配列が、アレル非相互作用変数としてN末端側及びC末端側のランキング配列が与えられるものとして、特定のペプチドがアレルHLA-B\*07:02上に提示される提示尤度を生成する。

20

#### 【0415】

以下の例に図13A～13Eを参照して開示するように、ペプチド結合の予測について訓練された提示モデル及び従来技術のモデルなどの異なるモデルを、それぞれのシミュレートした患者に対する新生抗原候補の試験セットに適用することによって予測に基づき患者に対する異なる治療サブセットを特定する。ワクチン治療について選択基準を満たす患者を選択し、患者の治療サブセットにエピトープを含む個別化ワクチンに関連付ける。治療サブセットのサイズは、異なるワクチン容量に応じて異なる。提示モデルを訓練するために用いられる訓練セットとシミュレートした新生抗原候補の試験セットとの間の重複は導入されない。

30

#### 【0416】

以下の例では、ワクチンに含まれるエピトープ間で少なくとも特定の数の提示新生抗原を有する選択された患者の比率を分析する。この統計は、患者に免疫応答を誘発する潜在的な新生抗原を送達するうえでのシミュレートしたワクチンの有効性を示すものである。詳細には、ある試験セット内のシミュレートした新生抗原は、その新生抗原が質量分析データセットD2において提示されている場合に提示される。提示された新生抗原を有する患者の高い比率は、免疫応答を誘導することによって新生抗原ワクチンによる治療の奏功の可能性を示す。

40

#### 【0417】

### XI. A. 実施例7A: NSCLCがん患者における変異負荷の度数分布

図13Aは、NSCLC患者における変異負荷の標本度数分布を示す。NSCLCを含む異なる腫瘍タイプにおける変異負荷及び変異は、例えば、がんゲノムアトラス（the cancer genome atlas）（TCGA）（<https://cancergenome.nih.gov>）にみることができる。X軸は各患者の非同義変異の数を表し、Y軸は特定の数の非同義変異を有する標本患者の比率を表す。図13Aの標本度数

50

分布は、3～1786個の変異の範囲を示し、患者の30%は100個よりも少ない変異を有している。図13Aには示されていないが、変異負荷は非喫煙者と比較して喫煙者でより高く、変異負荷が患者における新生抗原負荷の強力な指標となり得ることが研究によって示されている。

【0418】

上記のセクションXIの冒頭で導入したように、シミュレートした患者の数のそれぞれに、新生抗原候補の試験セットが関連付けられる。各患者の試験セットは、各患者について図13Aに示される度数分布から変異負荷 $m_i$ をサンプリングすることによって生成される。各変異について、ヒトプロテオーム由来の21マーのペプチド配列を、シミュレートする変異配列を表すようにランダムに選択する。新生抗原候補配列の試験セットを、21マー内の変異にわたった各(8、9、10、11)マーのペプチド配列を特定することにより患者 $i$ について生成する。各新生抗原候補に、新生抗原候補配列が質量分析D1データセット内に存在するかどうかを示すラベル(標識)を関連付ける。例えば、データセットD1内に存在する新生抗原候補配列にはラベル「1」を関連付け、データセットD1内に存在しない配列にはラベル「0」を関連付けることができる。以下でより詳細に述べるように、図13B～13Eは、試験セット内の患者の提示新生抗原に基づいた患者選択の実験結果を示している。

【0419】

XI. B. 実施例7B: 変異負荷の選択基準に基づく新生抗原提示を有する選択された患者の比率

図13Bは、患者が最小変異負荷を満たすかどうかの選択基準に基づいて選択された患者に対してシミュレートしたワクチン中の提示新生抗原の数を示す。対応する試験において少なくとも特定の数の提示新生抗原を有する選択された患者の比率を特定する。

【0420】

図13Bにおいて、x軸は、「最小数の変異」のラベルで示される、腫瘍変異負荷に基づいたワクチン治療から除外される患者の比率を示す。例えば、「最小数の変異」200におけるデータポイントは、患者選択モジュール324が、少なくとも変異が200個の変異負荷を有するシミュレートした患者のサブセットのみを選択したことを示す。別の例として、「最小数の変異」300におけるデータポイントは、患者選択モジュール324が、少なくとも300個の変異を有するより低い比率のシミュレートした患者を選択したことを示す。y軸は、ワクチン容量 $v$ を有さない試験セット内の少なくとも特定の数の提示された新生抗原が関連付けられた選択された患者の比率を示す。詳細には、上のプロットは、少なくとも1個の新生抗原を提示する選択された患者の比率を示し、中間のプロットは、少なくとも2個の抗原を提示する選択された患者の比率を示し、下のプロットは、少なくとも3個の抗原を提示する選択された患者の比率を示す。

【0421】

図13Bに示されるように、提示された新生抗原を有する患者の比率は、変異負荷が高くなるほど顕著に増大している。これは、選択基準としての変異負荷が、新生抗原ワクチンが効果的な免疫反応を誘導する可能性の高い患者を選択するうえで効果的であり得ることを示している。

【0422】

XI. C. 実施例7C: 提示モデルと従来技術のモデルとによって特定されたワクチンにおける新生抗原提示の比較

図13Cは、提示モデルに基づいて特定された治療サブセットを含むワクチンに関連付けられた選択された患者と、従来技術のモデルによって特定された治療サブセットを含むワクチンに関連付けられた選択された患者との間のシミュレートしたワクチン中の提示新生抗原の数を比較したものである。左側のプロットは、限定的なワクチン容量として $v=10$ を仮定しており、右側のプロットは限定的なワクチン容量として $v=20$ を仮定している。患者は、提示された新生抗原の期待数を示す効用値スコアに基づいて選択される。

【0423】

図13Cにおいて、実線は、アレルHLA-A\*02:01及びHLA-B\*07:02に対する提示モデルに基づいて特定された治療サブセットを含むワクチンに関連付けられた患者を示す。各患者に対する治療サブセットは、試験セット内の配列に提示モデルのそれぞれを適用し、最も高い提示尤度を有する $v$ 個の新生抗原候補を特定することによって特定される。点線は、単一アレルHLA-A\*02:01に対する従来技術のモデルNETMHCpanに基づいて特定された治療サブセットを含むワクチンに関連付けられた患者を示す。NETMHCpanについての実施の詳細は、<http://www.cbs.dtu.dk/services/NetMHCpan>に示されている。各患者に対する治療サブセットは、試験セット内の配列にNETMHCpanモデルを適用し、最も高い推定される結合親和性を有する $v$ 個の新生抗原候補を特定することによって特定される。両グラフのx軸は、提示モデルに基づいて特定された治療サブセット内の提示新生抗原の期待数を示す期待効用値スコアに基づいてワクチン治療から除外された患者の比率を示す。期待効用値スコアは、セクションXにおいて式(25)に関連して述べたようにして決定される。y軸は、ワクチンに含まれる少なくとも特定の数の新生抗原(1、2、または3個の新生抗原)を提示する選択された患者の比率を示す。

【0424】

図13Cに示されるように、提示モデルに基づいた治療サブセットを含むワクチンに関連付けられた患者には、従来技術のモデルに基づいた治療サブセットを含むワクチンに関連付けられた患者よりも有意に高い割合で提示新生抗原を含むワクチンが投与される。例えば、右側のグラフに示されるように、従来技術のモデルに基づいたワクチンに関連付けられた選択された患者のわずか40%と比較して、提示モデルに基づいたワクチンに関連付けられた選択された患者の80%に、ワクチン中で少なくとも1つの提示新生抗原が投与される。これらの結果は、本明細書に述べられる提示モデルは、腫瘍を治療するための免疫反応を誘発する可能性の高いワクチンの新生抗原候補を選択するうえで効果的であることを示している。

【0425】

XI. D. 実施例7D: 提示モデルにより特定されたワクチンの新生抗原提示に対するHLAカバレッジの影響

図13Dは、HLA-A\*02:01についての単一のアレルごとの提示モデルに基づいて特定された治療サブセットを含むワクチンに関連付けられた選択された患者と、HLA-A\*02:01についてのアレルごとの提示モデル及びHLA-B\*07:02についてのアレルごとの提示モデルの両方に基づいて特定された治療サブセットを含むワクチンに関連付けられた選択された患者との間のシミュレートしたワクチン中の提示新生抗原の数を比較したものである。ワクチン容量は、 $v = 20$ 個のエピトープに設定する。各実験について、異なる治療サブセットに基づいて決定された期待効用値スコアに基づいて患者を選択する。

【0426】

図13Dにおいて、実線は、HLAアレルHLA-A\*02:01及びHLA-B\*07:02に対する提示モデルの両方に基づいた治療サブセットを含むワクチンに関連付けられた患者を示す。各患者に対する治療サブセットは、試験セット内の配列に提示モデルのそれぞれを適用し、最も高い提示尤度を有する $v$ 個の新生抗原候補を特定することによって特定される。点線は、HLAアレルHLA-A\*02:01に対する単一提示モデルに基づいた治療サブセットを含むワクチンに関連付けられた患者を示す。各患者に対する治療サブセットは、試験セット内の配列に単一HLAアレルのみについての提示モデルを適用し、最も高い提示尤度を有する $v$ 個の新生抗原候補を特定することによって特定される。実線のプロットでは、x軸は、両方の提示モデルにより特定された治療サブセットに対する期待効用値スコアに基づいてワクチン治療から除外された患者の比率を示す。点線のプロットでは、x軸は、単一の提示モデルにより特定された治療サブセットに対する期待効用値スコアに基づいてワクチン治療から除外された患者の比率を示す。y軸は、少なくとも特定の数の新生抗原(1、2、または3個の新生抗原)を提示する選択された患者

の比率を示す。

【0427】

図13Dに示されるように、両方のHLAアレルに対する提示モデルより特定された治療サブセットを含むワクチンに関連付けられた患者は、単一の提示モデルにより特定された治療サブセットを含むワクチンに関連付けられた患者よりも有意に高い割合で新生抗原を提示する。これらの結果は、高いHLAカバレッジを有する提示モデルを確立することの重要性を示すものである。

【0428】

XI.E. 実施例7E：変異負荷と提示新生抗原の期待数とによって選択された患者における新生抗原提示の比較

図13Eは、変異負荷に基づいて選択された患者と、期待効用値スコアにより選択された患者との間でシミュレートしたワクチン中の提示新生抗原の数を比較したものである。期待効用値スコアは、 $v = 20$ 個のエピトープのサイズを有する提示モデルにより特定された治療サブセットに基づいて決定する。

【0429】

図13Eにおいて、実線は、提示モデルにより特定された治療サブセットを含むワクチンに関連付けられた期待効用値スコアに基づいて選択された患者を示す。各患者に対する治療サブセットは、試験セット内の配列に提示モデルのそれぞれを適用し、最も高い提示尤度を有する $v = 20$ 個の新生抗原候補を特定することによって特定される。治療効用値スコアは、セクションXで式(25)に基づいて特定された治療サブセットの提示尤度に基づいて決定される。点線は、提示モデルにより特定された治療サブセットを含むワクチンに関連付けられた変異負荷に基づいて選択された患者を示す。x軸は、実線のプロットの期待効用値スコアに基づいてワクチン治療から除外された患者の比率、及び点線のプロットの変異負荷に基づいて除外された患者の比率を示す。y軸は、少なくとも特定の数の提示新生抗原(1、2、または3個の新生抗原)を含むワクチンが投与される選択された患者の比率を示す。図13Eに示されるように、期待効用値スコアに基づいて選択された患者には、変異負荷に基づいて選択された患者よりも高い割合で提示新生抗原を含むワクチンが投与される。しかしながら、変異負荷に基づいて選択された患者には、選択されない患者よりも高い割合で提示新生抗原を含むワクチンが投与される。したがって、変異負荷は効果的な新生抗原ワクチン遅漏における効果的な患者選択基準であるが、期待効用値スコアはより効果的である。

【0430】

XII. 実施例8：除外されたMHCクラスII質量分析データに対する質量分析訓練されたMHCクラスII提示モデルの評価

上記に述べた様々な提示モデルの有効性を、提示モデルを訓練するために使用されなかった訓練データ170のサブセットである、または訓練データ170と同様の変数及びデータ構造を有する訓練データ170からの別のデータセットである試験データT、試験した。

【0431】

提示モデルの性能を示す関連指標として、

$$\text{陽性適中率 (PPV)} = P(y_{i \in T} = 1 \mid u_{i \in T} \geq t) = \frac{\sum_{i \in T} \mathbb{1}(y_i = 1, u_i \geq t)}{\sum_{i \in T} \mathbb{1}(u_i \geq t)}$$

があり、これは関連するHLAアレル上に提示されることが正しく予測されたペプチドインスタンスの数の、そのHLAアレル上に提示されることが予測されたペプチドインスタンスの数に対する比を示す。一実現形態では、試験データTの中のペプチド $p^i$ は、対応する尤度推定値 $u_i$ が、所定の閾値 $t$ 以上である場合に、1つ以上の関連するHLAアレル上に提示されるものと予測された。提示モデルの性能を示す別の関連指標としては、

$$\text{再現率} = P(u_{i \in T} \geq t \mid y_{i \in T} = 1) = \frac{\sum_{i \in T} \mathbb{1}(y_i = 1, u_i \geq t)}{\sum_{i \in T} \mathbb{1}(y_i = 1)}$$

10

20

30

40

50

があり、これは、関連するHLAアレル上に提示されることが正しく予測されたペプチドインスタンスの数の、そのHLAアレル上に提示されることが知られていたペプチドインスタンスの数に対する比を示す。提示モデルの性能を示す別の関連指標として、受信者操作特性（ROC）の曲線下面積（AUC）がある。ROCは、下式により与えられる、偽陽性率（FPR）に対する再現率をプロットする。

$$\text{FPR} = P(u_{i \in T} \geq t | y_{i \in T} = 0) = \frac{\sum_{i \in T} \mathbb{1}(y_i = 0, u_i \geq t)}{\sum_{i \in T} \mathbb{1}(y_i = 0)}$$

#### 【0432】

#### XII.A. 質量分析データに対する提示モデルの性能

10

##### XII.A.1. 実施例1

図14Aは、質量分析を用いたヒト腫瘍細胞及び腫瘍浸潤リンパ球（TIL）上のクラスII MHCアレルから溶出されたペプチドの長さのヒストグラムである。具体的には、HLA-DRB1\*12:01ホモ接合体アレル（「データセット1」）及びHLA-DRB1\*12:01、HLA-DRB1\*10:01の複数アレル試料（「データセット2」）に対して質量分析ペプチドミクスを行った。結果は、クラスII MHCアレルから溶出されたペプチドの長さは、アミノ酸6～30個の範囲であることを示している。図14Aに示される頻度分布は、参考文献69の図1Cに示される、最新の質量分析法を用いてクラスII MHCから溶出されたペプチドの長さと同様である。

#### 【0433】

20

図14Bは、データセット1及びデータセット2についてmRNA定量化と残基当たりの提示ペプチドとの依存関係を示す。結果は、mRNAの発現とクラスII MHCアレルのペプチド提示との間には強い依存関係があることを示している。

#### 【0434】

具体的には、図14Bの横軸は、 $\log_{10}$ （ビン100万個当たりの転写物（TPM））で表したmRNAの発現を示す。図14Bの縦軸は、残基当たりのペプチド提示量を、 $10^{-2} < \log_{10} \text{TPM} < 10^{-1}$ の間のmRNA発現量に対応した最も低いビンの倍数として示したものである。1つの実線は、データセット1のmRNA定量化及びペプチド提示に関連したプロットであり、もう1つの実線はデータセット2のものである。図14Bに示されるように、mRNA発現量と対応する遺伝子内の残基当たりのペプチド提示量との間には強い相関がある。具体的には、RNA発現量が $10^1 < \log_{10} \text{TPM} < 10^2$ の範囲である遺伝子からのペプチドは、最低のビンと比較して提示される可能性が5倍以上高い。

30

#### 【0435】

これらの結果は、mRNA定量の測定値がペプチド提示を強く予測することから、これらの測定値を取り入れることにより、提示モデルの性能が大幅に改善されうること示す。

#### 【0436】

図14Cは、データセット1及びデータセット2を用いて訓練及び試験した例示的な提示モデルの性能結果を比較したものである。例示的な提示モデルのモデル特性の各セットについて、図14Cは、そのモデル特性のセットの中の特性がアレル相互作用特性として分類される場合、あるいはそのモデル特性のセットの中の特性がアレル非相互作用特性変数として分類される場合の再現率10%でのPPV値を示している。図14Cにみられるように、例示的な提示モデルのモデル特性の各セットについて、そのモデル特性のセットの中の特性がアレル相互作用特性として分類される場合に特定された再現率10%でのPPV値が左側に示され、そのモデル特性のセットの中の特性がアレル非相互作用特性として分類される場合に特定された再現率10%でのPPV値が右側に示されている。ペプチド配列の特性は、図14Cの目的では常にアレル相互作用特性として分類された点に注目されたい。結果は、この提示モデルによって14%～29%の範囲の10%リコールでのPPV値が実現され、この値はランダム予測におけるPPVよりも有意に高い（およそ500倍）ことを示している。

40

50

## 【0437】

この実験では、9～20個の長さのペプチド配列を検討した。データを、訓練、検証、及び試験セットに分けた。データセット1及びデータセット2の両方からの50個の残基のペプチドのブロックを訓練セット及び試験セットに割り当てた。プロテオーム内のどこかで重複したペプチドを除去し、訓練セットと試験セットの両方に現れるペプチド配列がないようにした。訓練セット及び試験セットにおけるペプチド提示の発生率は、非提示ペプチドを除去することにより50倍増加した。これは、データセット1及びデータセット2が、細胞の一部のみがクラスII HLAアレルであるヒト腫瘍試料由来のものであり、クラスII HLAアレルの純粋な試料よりもおよそ10倍低いペプチド収率となることによる（これは依然として、不完全な質量分析の感度による過小評価である）。訓練セットは、1064個の提示ペプチド及び3,810,070個の非提示ペプチドを含んでいた。試験セットは、314個の提示ペプチド及び807,400個の非提示ペプチドを含んでいた。

10

## 【0438】

例示モデル1は、ネットワーク依存性関数  $g_h(\cdot)$ 、 $\expit$  関数  $f(\cdot)$ 、及びアイデンティティ関数  $r(\cdot)$  を用いた式(22)の関数の総和モデルとした。ネットワーク依存性関数  $g_h(\cdot)$  は、256個の隠れノード及び正規化線形ユニット (ReLU) 活性化を伴う多層パーセプトロン (MLP) として構造化した。ペプチド配列以外に、アレル相互作用変数  $w$  は、ワンホットコード化されたC末端及びN末端フランキンク配列、ペプチド  $p^i$  のソース遺伝子の添え字  $G = gene(p^i)$  を示すカテゴリー変数、及びmRNA定量測定値を示す変数を含んでいた。例示モデル2は、C末端及びN末端フランキンク配列をアレル相互作用変数から省略した点以外は、例示モデル1と同じものとした。例示モデル3は、ソース遺伝子の添え字をアレル相互作用変数から省略した点以外は、例示モデル1と同じものとした。例示モデル4は、mRNA測定値をアレル相互作用変数から省略した点以外は、例示モデル1と同じものとした。

20

## 【0439】

例示モデル5は、ネットワーク依存性関数  $g_h(\cdot)$ 、 $\expit$  関数  $f(\cdot)$ 、アイデンティティ関数  $r(\cdot)$ 、及び式(12)の依存性関数  $g_w(\cdot)$  を用いた式(20)の関数の総和モデルとした。依存性関数  $g_w(\cdot)$  は、mRNA定量測定値を入力とし、16個の隠れノード及び正規化線形ユニット (ReLU) 活性化を伴うMLPとして構造化したネットワークモデル、及びC末端フランキンク配列を入力とし、32個の隠れノード及び正規化線形ユニット (ReLU) 活性化を伴うMLPとして構造化したネットワークモデルも含んでいた。ネットワーク依存性関数  $g_h(\cdot)$  は、256個の隠れノード及び正規化線形ユニット (ReLU) 活性化を伴う多層パーセプトロンとして構造化した。例示モデル6は、C末端及びN末端フランキンク配列のネットワークモデルを省略した点以外は、例示モデル5と同じものとした。例示モデル7は、ソース遺伝子の添え字をアレル非相互作用変数から省略した点以外は、例示モデル5と同じものとした。例示モデル8は、mRNA定量測定値のネットワークモデルを省略した点以外は、例示モデル5と同じものとした。

30

## 【0440】

試験セット内の提示ペプチドの発生率はおよそ  $1/2400$  であり、したがって、ランダム予測のPPVもおよそ  $1/2400 = 0.00042$  となるであろう。図14Cに示されるように、最も精度のよい提示モデルではおよそ29%のPPV値が実現され、この値はランダム予測のPPV値よりもおよそ500倍良好である。

40

## 【0441】

XII. A. 2. 実施例2

図14Dは、HLAクラスII分子を含むヒト腫瘍 (NSCLC、リンパ腫、及び卵巣癌) 及び細胞株 (EBV) からなる合計73種の試料の各試料について質量分析を用いてシーケンシングしたペプチドの量を示すヒストグラムである。図14Dに示されるように、各試料について平均で900種のペプチドをシーケンシングした。さらに、複数の

50

試料の各試料について、図14Dに示されるヒストグラムは、異なるq値の閾値で質量分析を用いてシーケンシングしたペプチドの量を示している。具体的には、複数の試料の各試料について、図14Dは、0.01未満のq値、0.05未満のq値、及び0.2未満のq値で質量分析を用いてシーケンシングしたペプチドの量を示している。

#### 【0442】

上記に述べたように、図14Dの73種の試料の各試料は、HLAクラスII分子を含んでいた。より具体的には、図14Dの73種の試料の各試料はHLA-DR分子を含んでいた。HLA-DR分子は、HLAクラスII分子の一種である。さらにより具体的には、図14Dの73種の試料の各試料は、HLA-DRB1分子、HLA-DRB3分子、HLA-DRB4分子、及び／またはHLA-DRB5分子を含んでいた。HLA-DRB1分子、HLA-DRB3分子、HLA-DRB4分子、及びHLA-DRB5分子は、HLA-DR分子の種類である。

#### 【0443】

この特定の実験は、HLA-DR分子、特にHLA-DRB1分子、HLA-DRB3分子、HLA-DRB4分子、及びHLA-DRB5分子を含む試料を用いて行ったが、代替的な実施形態では、この実験はHLAクラスII分子の任意の種類（複数可）のうちの1つ以上を含む試料を用いて行うことができる。例えば、代替的な実施形態では、HLA-DP及び／またはHLA-DQ分子を含む試料を用いて同じ実験を行うことができる。同じ方法を用いて任意の種類（複数可）のMHCクラスII分子をモデル化し、なおも信頼性の高い結果を得ることが可能であることは、当業者には周知である。例えば、Jensen、Kamilla Kjaergaard et al.<sup>76</sup>は、HLA-DR分子に対する、ならびにHLA-DP及びHLA-DQ分子に対する結合親和性をモデル化するために同じ方法を用いている最近の科学論文の一例である。したがって、当業者であれば、本明細書に記載される実験及びモデルを用いることで、HLA-DR分子のみでなく、他の任意のMHCクラスII分子を別々または同時にモデル化し、なおも信頼性の高い結果を得ることができる点は理解されよう。

#### 【0444】

合計73種の試料の各試料のペプチドをシーケンシングするため、各試料について質量分析を行った。試料について得られた質量スペクトルをCometで検索し、Percolatorでスコアリングしてペプチドをシーケンシングした。次いで、試料中でシーケンシングされたペプチドの量を複数の異なるPercolatorのq値の閾値について特定した。具体的には、その試料について、0.01未満のPercolatorのq値、0.05未満のPercolatorのq値、及び0.2未満のPercolatorのq値を用いてシーケンシングしたペプチドの量を決定した。

#### 【0445】

73種の試料の各試料について、異なるPercolatorのq値の閾値のそれぞれでシーケンシングしたペプチドの量を図14Dに示す。例えば、図14Dにみられるように、第1の試料では、0.2未満のq値で、質量分析を用いておよそ4700種のペプチドがシーケンシングされ、0.05未満のq値で、質量分析を用いておよそ3600種のペプチドがシーケンシングされ、0.01未満のq値で、質量分析を用いておよそ3200種のペプチドがシーケンシングされた。

#### 【0446】

全体として、図14Dは、質量分析を用いてMHCクラスII分子を含有する試料から大量のペプチドを、低いq値でシーケンシングすることができることを示している。換言すると、図14Dに示されるデータは、質量分析を用いて、MHCクラスII分子によって提示されうるペプチドを高い信頼度でシーケンシングすることができることを実証するものである。

#### 【0447】

図14Eは、特定のMHCクラスII分子のアレルが同定された試料の量を示すヒストグラムである。より具体的には、図14Eは、HLAクラスII分子を含む合計で73種

10

20

30

40

50

の試料について、特定のMHCクラスII分子が同定された試料の量を示している。

【0448】

図14Dに関して上記に述べたように、図14Dの73種の試料の各試料は、HLA-DRB1分子、HLA-DRB3分子、HLA-DRB4分子、及び／またはHLA-DRB5分子を含んでいた。したがって、図14Eは、HLA-DRB1分子、HLA-DRB3分子、HLA-DRB4分子、及びHLA-DRB5分子について、ある特定のアレルが同定された試料の量を示している。試料中に存在するHLAアレルを同定するには、その試料についてHLAクラスII DRのタイピングを行う。次いで、特定のHLAアレルが同定された試料の量を特定するには、HLAクラスII DRのタイピングを用いてHLAアレルが同定された試料の数を単純に合計する。例えば、図14Eに示されるように、合計で73種の試料のうちの17種の試料が、HLAクラスII分子のアレルHLA-DRB3\*01:01を含んでいた。換言すると、合計で73種の試料のうちの17種の試料がHLA-DRB3分子についてアレルHLA-DRB3\*01:01を含んでいた。全体として、図14Eは、HLAクラスII分子を含む73種の試料から広範囲のHLAクラスII分子のアレルを同定することができることを示している。

10

【0449】

図14Fは、合計で73種の試料中でMHCクラスII分子によって提示されるペプチドの割合を、一定範囲のペプチド長の各ペプチド長について示したヒストグラムである。合計で73種の試料の各試料中の各ペプチドの長さを決定するため、各ペプチドを図14Dに関して上記に述べたような質量分析を用いてシーケンシングし、次いでシーケンシングしたペプチド中の残基の数を単純に定量した。

20

【0450】

上記に述べたように、MHCクラスII分子は、一般的にはアミノ酸9～20個の長さを有するペプチドを提示する。したがって、図14Fは、アミノ酸9～20個の各ペプチド長について73種の試料中のMHCクラスII分子によって提示されるペプチドの割合を示している。例えば、図14Fに示されるように、73種の試料中のMHCクラスII分子によって提示されるペプチドのおよそ23%がアミノ酸14個の長さを有している。

【0451】

図14Fに示されるデータに基づけば、73種の試料中のMHCクラスII分子によって提示されるペプチドの最頻長は、アミノ酸14個及び15個の長さで特定された。73種の試料中のMHCクラスII分子によって提示されるペプチドについて特定されたこれらの最頻長は、MHCクラスII分子によって提示されるペプチドの最頻長についての以前の報告と一致する。さらに、やはり以前の報告と一致するように、図14Fのデータは、73種の試料に由来するMHCクラスII分子によって提示されるペプチドの60%超がアミノ酸14個及び15個以外の長さを有することを示している。換言すれば、図14Fは、MHCクラスII分子によって提示されるペプチドは最も高頻度でアミノ酸14個または15個の長さであるが、MHCクラスII分子によって提示されるペプチドの大きな割合がアミノ酸14個でも15個でもない長さであることを示している。したがって、すべての長さのペプチドが、MHCクラスII分子によって提示される等しい確率を有する、またはアミノ酸14個または15個の長さを有するペプチドのみがMHCIIクラスII分子によって提示されると仮定することは正確でない仮定である。図14Lに関して下記に詳細に述べるように、これらの誤った仮定は、MHCクラスII分子による提示を予測するための多くの最新のモデルで現在用いられており、したがって、これらのモデルにより予測される提示尤度はしばしば信頼性が低い。

30

40

【0452】

図14Gは、73種の試料中に存在する遺伝子について、遺伝子発現とMHCクラスII分子による遺伝子発現産物の提示の発生率との間の関係を示す線グラフである。より具体的には、図14Gは、遺伝子発現と、MHCクラスII分子によって提示されるペプチドのN末端を形成する、その遺伝子発現より生ずる残基の割合との間の関係を示している。合計で73種の試料の各試料中の遺伝子発現を定量化するため、各試料中に含まれるR

50

NAにRNAシーケンシングを行う。図14Gでは、遺伝子発現は、100万当たりの転写物 (transcripts per million) (TPM) の単位でRNAシーケンシングによって測定されている。73種の試料の各試料について遺伝子発現の提示の発生率を特定するため、HLAクラスII DRペプチドームデータの同定を各試料について行った。

【0453】

図14Gに示されるように、73種の試料について、遺伝子発現レベルと、MHCクラスII分子による発現遺伝子産物の残基の提示との間には強い相関が認められる。具体的には、図14Gに示されるように、発現レベルの最も低い遺伝子の発現から生じるペプチドは、発現レベルの最も高い遺伝子の発現から生じるペプチドと比較して、100倍以上、MHCクラスII分子によって提示されにくい。より簡単に言えば、より高度に発現される遺伝子の産物はMHCクラスII分子によってより高頻度で提示される。

【0454】

図14H~I及び図14K~Lは、ペプチドの試験データセット内のペプチドが、試験データセット内に存在するMHCクラスII分子の少なくとも1つによって提示される尤度を予測するうえでの異なる提示モデルの性能を比較した線グラフである。図14H~I及び図14K~Lに示されるように、あるペプチドが試験データセット内に存在するMHCクラスII分子の少なくとも1つによって提示される尤度を予測するうえでのモデルの性能は、そのモデルによって生成される各予測について、真の陽性率の偽陽性率に対する比を特定することにより決定される。所定のモデルについて特定されたこれらの比は、ROC (receiver operator characteristic) (受信者動作特性) 曲線として、x軸が偽陽性率を定量化し、y軸が真の陽性率を定量化した線グラフにおいて可視化することが可能である。曲線下面積 (AUC) を用いてモデルの性能を定量化する。具体的には、AUCが大きいモデルほど、AUCが小さいモデルと比較してより高い性能 (すなわちより高い精度) を有する。図14H、14I、及び14Lにおいて、傾き1の黒い破線 (真の陽性率の偽陽性率に対する比が1) は、ランダムに推定されたペプチド提示の尤度の予測曲線を示す。破線のAUCは0.5である。ROC曲線及びAUCの計測については、上記のセクションXIIの前半部分に関して詳細な考察がなされている。

【0455】

図14Hは、アレル相互作用及びアレル非相互作用変数の異なるセットが与えられた場合に、ペプチドの試験データセット内のペプチドがMHCクラスII分子によって提示される尤度を予測するうえでの5つの例示モデルの性能を比較した線グラフである。換言すれば、図14Hは、ペプチドがMHCクラスII分子によって提示される尤度を予測するうえでの異なるアレル相互作用及びアレル非相互作用変数の相対的な重要度を定量化するものである。

【0456】

図14Hの線グラフのROC曲線を作成するために用いた5つの例示モデルの各例示提示モデルのモデルアーキテクチャは、5つのシグモイド総和モデルの集合を含んだ。集合内の各シグモイド総和モデルは、試料当たり最大で4つの固有のHLA-DRアレルについてペプチド提示をモデル化するように構成した。さらに、集合内の各シグモイド総和モデルは、以下のアレル相互作用及びアレル非相互作用変数、すなわち、ペプチド配列、ランキング配列、TPM単位のRNA発現、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を行うように構成した。集合内の各シグモイド総和モデルのアレル相互作用成分は、256個の隠れユニットとしてReLU活性化を伴う1隠れ層のMLPであった。

【0457】

ペプチドの試験データセット内のペプチドがMHCクラスII分子によって提示される尤度を例示モデルを用いて予測するのに先立って、例示モデルを訓練及び検証した。例示モデルを訓練、検証、及び最終的に試験するため、73種の試料について上記に述べたデ

10

20

30

40

50

ータを訓練、検証、及び試験データセットに分けた。

【0458】

訓練、検証、及び試験データセットのうちの複数のデータセットにペプチドが現れないようにするため、以下の手順を行った。最初に、プロテオーム内の複数の位置に現れた、合計73種の試料に由来するすべてのペプチドを除去した。次いで、合計73種の試料に由来するペプチドを10個の隣接するペプチドのブロックに分割した。合計73種の試料に由来するペプチドの各ブロックを、訓練データセット、検証データセット、または試験データセットに個々に割り当てた。これにより、訓練、検証、及び試験データセットの複数のデータセットに現れるペプチドはなくなった。

【0459】

合計73種の試料中の38,035,453種のペプチドのうち、訓練データセットは、合計73種の試料のうちの69種に由来するMHCクラスII分子によって提示された33,570種のペプチドを含んだ。訓練データセットに含まれる33,570種のペプチドは、アミノ酸9~20個の長さであった。図14HのROCカーブを作成するために用いた例示モデルを、ADAM最適化アルゴリズム及び早期停止を用いて訓練データセット上で訓練した。

【0460】

検証データセットは、訓練データセットで用いられた同じ69種の試料に由来するMHCクラスII分子によって提示された3,925種のペプチドで構成されていた。検証データセットは、早期停止のみに用いた。

【0461】

試験データセットは、質量分析を用いて腫瘍試料から同定されたMHCクラスII分子によって提示されたペプチドを含んだ。具体的には、試験データセットは、4つの腫瘍試料から特定された232種のペプチドを含んだ。試験データセットに含まれるペプチドは、上記に述べた訓練データセットからは除外されていた。

【0462】

上記に述べたように、図14Hは、ペプチドがMHCクラスII分子によって提示される尤度を予測するうえでの異なるアレル相互作用変数及びアレル非相互作用変数の相対的な重要度を定量化するものである。やはり上記に述べたように、図14Hの線グラフのROC曲線を作成するために用いた例示モデルは、以下のアレル相互作用及びアレル非相互作用変数、すなわち、ペプチド配列、フランキング配列、TPM単位のRNA発現、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を行うように構成した。ペプチドがMHCクラスII分子によって提示される尤度を予測するために、これら5つの変数（ペプチド配列、フランキング配列、RNA発現、及び遺伝子識別子）のうちの4つの変数の相対的な重要度を定量化するため、上記に述べた5つの例示モデルの各例示モデルを、4つの変数の異なる組み合わせを用いて、試験データセットからのデータを用いて試験した。具体的には、試験データセットの各ペプチドについて、例示モデル1は、RNA発現を除いた、ペプチド配列、フランキング配列、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を生成した。同様に、試験データセットの各ペプチドについて、例示モデル2は、フランキング配列を除いた、ペプチド配列、RNA発現、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を生成した。同様に、試験データセットの各ペプチドについて、例示モデル3は、ペプチド配列を除いた、フランキング配列、RNA発現、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を生成した。同様に、試験データセットの各ペプチドについて、例示モデル4は、遺伝子識別子を除いた、フランキング配列、RNA発現、ペプチド配列、及び試料識別子に基づいてペプチド提示尤度の予測を生成した。最後に、試験データセットの各ペプチドについて、例示モデル5は、フランキング配列、RNA発現、ペプチド配列、遺伝子識別子、及び試料識別子の5つの変数のすべてに基づいてペプチド提示尤度の予測を生成した。

【0463】

これら5つの例示モデルのそれぞれの性能を図14Hの線グラフに示す。具体的には、

10

20

30

40

50

5つの例示モデルのそれぞれは、モデルによって生成された各予測について、真の陽性率の偽陽性率に対する比を示すROC曲線に関連付けられる。例えば、図14Hは、RNA発現を除いた、ペプチド配列、フランキング配列、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を生成した例示モデル1の曲線を示している。図14Hは、フランキング配列を除いた、ペプチド配列、RNA発現、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を生成した例示モデル2の曲線を示している。図14Hはまた、ペプチド配列を除いた、フランキング配列、RNA発現、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を生成した例示モデル3の曲線も示している。図14Hはまた、遺伝子識別子を除いた、フランキング配列、RNA発現、ペプチド配列、及び試料識別子に基づいてペプチド提示尤度の予測を生成した例示モデル4の曲線も示している。そして最後に、図14Hは、フランキング配列、RNA発現、ペプチド配列、試料識別子、及び遺伝子識別子の5つの変数のすべてに基づいてペプチド提示尤度の予測を生成した例示モデル5の曲線を示している。

#### 【0464】

上記に述べたように、あるペプチドがMHCクラスII分子によって提示される尤度を予測するうえでのモデルの性能は、そのモデルによって生成された各予測について、真の陽性率の偽陽性率に対する比を示すROC曲線のAUCを特定することによって定量化される。AUCが大きいモデルほど、AUCがより小さいモデルと比較してより高い性能（すなわちより高い精度）を有する。図14Hに示されるように、フランキング配列、RNA発現、ペプチド配列、試料識別子、及び遺伝子識別子の5つの変数のすべてに基づいてペプチド提示尤度の予測を生成した例示モデル5の曲線は0.98と最も高いAUCを実現した。したがって、ペプチド提示の予測を生成するために5つの変数すべてを用いた例示モデル5は、最良の性能を実現した。フランキング配列を除いた、ペプチド配列、RNA発現、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を生成した例示モデル2の曲線は、0.97と2番目に高いAUCを実現した。したがって、フランキング配列は、ペプチドがMHCクラスII分子によって提示される尤度を予測するうえで最も重要度の低い変数として特定することができる。遺伝子識別子を除いた、フランキング配列、RNA発現、ペプチド配列、及び試料識別子に基づいてペプチド提示尤度の予測を生成した例示モデル4の曲線は、0.96と3番目に高いAUCを実現した。したがって、遺伝子識別子は、ペプチドがMHCクラスII分子によって提示される尤度を予測するうえで2番目に重要度の低い変数として特定することができる。ペプチド配列を除いた、フランキング配列、RNA発現、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を生成した例示モデル3の曲線は、0.88と最も低いAUCを実現した。したがって、ペプチド配列は、ペプチドがMHCクラスII分子によって提示される尤度を予測するうえで最も重要な変数として特定することができる。RNA発現を除いた、ペプチド配列、フランキング配列、遺伝子識別子、及び試料識別子に基づいてペプチド提示尤度の予測を生成した例示モデル1の曲線は、0.95と2番目に低いAUCを実現した。したがって、RNA発現は、ペプチドがMHCクラスII分子によって提示される尤度を予測するうえで2番目に重要な変数として特定することができる。

#### 【0465】

図14Iは、ペプチドの試験データセット内のペプチドがMHCクラスII分子によって提示される尤度を予測するうえでの4つの異なる提示モデルの性能を比較した線グラフである。

#### 【0466】

図14Iで試験される第1のモデルは、本明細書では「結合親和性モデル」と呼ばれる。図14Iの結合親和性モデルは、予測を生成するための基準として最小のNetMHCII2.3で予測された結合親和性を用いた、ベスト・イン・クラスモデルである、NetMHCII2.3モデルである。具体的には、NetMHCII2.3モデルは、MHCクラスII分子のタイプ及びペプチド配列に基づいてペプチド提示尤度の予測を生成する。NetMHCII2.3モデルは、NetMHCII2.3のウェブサイト（www

. c b s . d t u . d k / s e r v i c e s / N e t M H C I I / , P M I D 2 9 3 1 5 5 9 8 ) <sup>7 6</sup>を用いて試験した。

【0467】

図14Iで試験される第2のモデルは、本明細書では「MLP」と呼ばれる。MLP（多層パーセプトロン）モデルは、アレル非相互作用変数 $w^k$ 及びアレル相互作用変数 $x_h^k$ が例えばニューラルネットワークのような別々の依存性関数に入力された後、これらの別々の依存性関数の出力が加え合わされる、上記に述べた提示モデルの一実施形態である。具体的には、完全非相互作用モデルは、アレル非相互作用変数 $w^k$ が依存性関数 $g_w$ に入力され、アレル相互作用変数 $x_h^k$ が別の依存性関数 $g_h$ に入力され、依存性関数 $g_w$ と依存性関数 $g_h$ の出力が互いに加え合わされる、上記に述べた提示モデルの一実施形態である。したがって、いくつかの実施形態では、完全非相互作用モデルは、上記に示した式8を用いてペプチド提示の尤度を決定する。さらに、アレル非相互作用変数 $w^k$ が依存性関数 $g_w$ に入力され、アレル相互作用変数 $x_h^k$ が別の依存性関数 $g_h$ に入力され、依存性関数 $g_w$ と依存性関数 $g_h$ の出力が加え合わされる完全非相互作用モデルの実施形態は、セクションVIIII. B. 2の前半部分、セクションVIIII. B. 3の後半部分、セクションVIIII. C. 3の前半部分、及びセクションVIIII. C. 6の前半部分に関して上記に詳細に述べられている。

【0468】

図14Iで試験される第3のモデルは、本明細書では「RNN」と呼ばれる。RNNモデルは、回帰型ニューラルネットワークを含み、上記に述べた完全非相互作用モデルと類似している。しかしながら、RNNモデルの回帰型ニューラルネットワークの層は、MLPモデルのニューラルネットワークの層とは異なっている。具体的には、RNNモデルの回帰型ニューラルネットワークの入力層は、一度に1つずつのペプチドをモデル化する可変長のペプチドストリングを受け入れる。このペプチドは、一度にアミノ酸1個ずつ、ニューラルネットワークノードに与えられ、その出力は、配列全体がモデル化されるまで配列中の次のアミノ酸とともにノードの入力にパイプ入力される。回帰層は、（1）データの連続性がモデルにキャプチャされる、及び（2）人工的にパディングを行うことなく、ペプチドの長さが変わりうる、という2つの理由でMHCクラスIIペプチドのモデル化に特に適用できる。回帰型ニューラルネットワークの次の層は、 $p=0.2$ のドロップアウト層と、最後にReLU活性化を用いる全結合64ノード層である。

【0469】

図14Iで試験される第4のモデルは、本明細書では「Bi-LSTM」と呼ばれる。Bi-LSTMモデルは、双方向長期短期記憶ニューラルネットワークで構成される。Bi-LSTMモデルは、ペプチド入力層以外は非相互作用モデルと同じである。Bi-LSTMモデルの入力層は、20マーペプチドストリングを受け入れ、続いてこの20マーペプチドストリングを $(n, 20, 21)$ テンソルとして埋め込む。Bi-LSTMモデルの双方向長期短期記憶ニューラルネットワークの次の各層は、128個のノードを有する回帰型長期短期記憶層、 $p=0.2$ のドロップアウト層、及び最後にReLU活性化を用いる全結合64ノード層からなる。従来のLSTMモデルでは、連続データの順序は、方向性があるものと仮定される（例えば、左から右へ、または右から左へ読まれる）。双方向LSTMでは、連続データは、左から右へ、及び右から左へ進み、両方向に処理される。ペプチドの結合は本質的には無方向性のタスクであるため、両方向での配列のモデル化によって、配列の両端からの情報がモデルの予測において同程度の重みを有することになる。

【0470】

図14Jを簡単に参照すると、図14Jは、HLA-DRB（MHCクラスII遺伝子）によるペプチド提示を予測するように構成された、図14IのBi-LSTMモデルの例示的な実施形態を示している。図14Jに示されるように、Bi-LSTMモデルは、アレル非相互作用特性（例えば、RNA配列、試料番号、タンパク質番号、及びフランキンク配列）を受け入れる共有ニューラルネットワークと、それぞれが異なるHLA-DR

B アレルに関連付けられ、コード化されたペプチド配列（アレル相互作用特性）を受け入れるように構成された別個のニューラルネットワークのセットを含む。ニューラルネットワークのセットのそれぞれの異なるニューラルネットワークは、B i - L S T Mニューラルネットワークを含む。図 1 4 J の B i - L S T Mモデルの例示的な実施形態では、異なるアレルに関連付けられた別個のニューラルネットワークのセットは、患者の試料 1 つ当たり H L A - D R B 遺伝子に最大で 4 つの異なるアレルが関連付けられているため、4 つの別個のニューラルネットワークを含んでいる。しかしながら、B i - L S T Mモデルが別の H L A 遺伝子によるペプチド提示を予測するように構成されている別の実施形態では、別個のニューラルネットワークのセットは、所定の H L A 遺伝子について患者試料中の各アレルの最大可能量に等しい量の別個のニューラルネットワークを含む。ニューラルネットワークのセットのうちのそれぞれの別個のニューラルネットワークは、モデルに入力されるペプチドが、所定のニューラルネットワークに関連付けられた H L A - D R B アレルによって提示される尤度を決定する。次いで、これらの尤度のそれぞれは、共有ニューラルネットワークからの出力と組み合わせられる。最後に、組み合わせられた尤度を合計してペプチドが H L A - D R B 遺伝子によって提示される全体の尤度を生成する。

#### 【 0 4 7 1 】

再び図 1 4 I を参照し、ペプチドの試験データセット内のペプチドが M H C クラス I I 分子によって提示される尤度を図 1 4 I の 4 つのモデルのそれぞれを用いて予測するのに先立って、各モデルを訓練及び検証した。結合親和性モデルは、免疫エピトープデータベース（I E D B , [www.iedb.org](http://www.iedb.org)）に寄託された H L A ペプチド結合親和性アッセイに基づいてそれ自体の訓練及び検証データセットを用いて訓練及び検証した。他の 3 つのモデルは、上記に述べた 6 9 種の試料の訓練データセットを用いて訓練し、上記に述べた検証データセットを用いて検証した。各モデルのこの訓練及び検証に続き、4 つのモデルのそれぞれを、上記に述べた試験データセットからの 4 つの除外された腫瘍試料を用いて試験した。具体的には、4 つのモデルのそれぞれについて、試験データセットからの 4 つの除外された腫瘍試料の各ペプチドをモデルに入力し、続いてモデルがそのペプチドの提示尤度を出力した。

#### 【 0 4 7 2 】

これら 4 つのモデルのそれぞれの性能を図 1 4 I の線グラフに示す。具体的には、4 つのモデルのそれぞれは、モデルによって生成された各予想について、真の陽性率の偽陽性率に対する比を示す R O C 曲線に関連付けられる。例えば、図 1 4 I は、結合親和性モデルの R O C 曲線、R N Nモデルの R O C 曲線、M L Pモデルの R O C 曲線、及び B i - L S T Mモデルの R O C 曲線を示している。

#### 【 0 4 7 3 】

上記に述べたように、あるペプチドが M H C クラス I I 分子によって提示される尤度を予測するうえでのモデルの性能は、そのモデルによって生成された各予測について、真の陽性率偽陽性率に対するの比を示す R O C 曲線の A U C を特定することによって定量化される。A U C が大きいモデルほど、A U C がより小さいモデルと比較してより高い性能（すなわちより高い精度）を有する。図 1 4 I に示されるように、B i - L S T Mモデルの曲線は 0 . 9 8 と最も高い A U C を実現した。したがって、B i - L S T Mモデルは最も良好な性能を実現した。B i - L S T Mモデルのこのようなピーク性能は、B i - L S T Mが、可変長さのペプチド、比較的長い長さのペプチド、及び繰り返すアミノ酸を有するペプチドを正確に予測する能力が最も高いことに一部よるものである。M L P及び R N Nモデルの曲線は、0 . 9 7 と 2 番目に高い A U C を実現した。したがって、M L P及び R N Nモデルは 2 番目に良好な性能を実現した。結合親和性モデルの曲線は、0 . 7 9 と最も低い A U C を実現した。したがって、結合親和性モデルは最も悪い性能となった。図 1 4 I で試験した B i - L S T M、M L P、及び R N Nモデルのそれぞれは、0 . 9 よりも大きい A U C を有している点に留意されたい。したがって、これらのモデル間のペプチド入力層のアーキテクチャの差異にもよらず、これらのモデルは、大幅に低い A U C を有する結合親和性モデルと異なり、比較的正確なペプチド提示の予測を得ることができる。

## 【0474】

図14Kは、図14Iに関して上記に述べた「Bi-LSTM」モデル、「MLP」モデル、「RNN」モデル、及び「結合親和性」モデルの完全適合率－再現率曲線を示す線グラフである。図14Kに示されるように、また、図14Iに基づいて予想されるように、「Bi-LSTM」モデルはAUC=0.23の最良の性能を実現し、「RNN」モデルはAUC=0.16の2番目に良好な性能を実現し、「MLP」モデルはAUC=0.11の3番目に良好な性能を実現し、「結合親和性」モデルはAUC=0.01の最も悪い性能となった。詳細には、質量分析データで訓練したBi-LSTMモデルは、結合親和性モデルの性能を大きく上回り、AUCが20倍を超えて増加した。

## 【0475】

図14Lは、2つの異なる基準が与えられた2つの例示的なベスト・イン・クラスの従来モデル、ならびにアレル相互作用変数及びアレル非相互作用変数の2つの異なるセットが与えられた2つの実施例提示モデルの、ペプチドの試験データセット内のペプチドがMHCクラスII分子によって提示される尤度を予想するうえでの性能を比較した線グラフである。具体的には、図14Lは、予想を生成するうえで最小NetMHCII2.3で予想された結合親和性を基準として用いる実施例のベスト・イン・クラス従来モデル（例示モデル1）、予想を生成するうえで最小NetMHCII2.3で予想された結合ランクを基準として用いる実施例のベスト・イン・クラス従来モデル（例示モデル2）、MHCクラスII分子のタイプ及びペプチド配列に基づいてペプチド提示尤度の予想を生成する例示モデル（例示モデル4）、ならびに、MHCクラスII分子のタイプ、ペプチド配列、RNA発現、遺伝子識別子、及びフランキンゲ配列に基づいてペプチド提示尤度の予想を生成する例示モデル（例示モデル3）の性能を比較した線グラフである。

## 【0476】

図14Lにおいて例示モデル1及び例示モデル2として用いられているベスト・イン・クラス従来モデルは、NetMHCII2.3モデルである。NetMHCII2.3モデルは、MHCクラスII分子のタイプ及びペプチド配列に基づいてペプチド提示尤度の予想を生成する。NetMHCII2.3モデルは、NetMHCII2.3のウェブサイト（[www.cbs.dtu.dk/services/NetMHCII/](http://www.cbs.dtu.dk/services/NetMHCII/)、PMID 29315598）<sup>76</sup>を用いて試験した。

## 【0477】

上記に述べたように、NetMHCII2.3モデルは2つの基準にしたがって試験した。具体的には、例示モデル1は、最小NetMHCII2.3で予想された結合親和性にしたがってペプチド提示尤度の予想を生成し、例示モデル2は、最小NetMHCII2.3で予想された結合ランクにしたがってペプチド提示尤度の予想を生成した。

## 【0478】

例示モデル3及び例示モデル4として用いられている提示モデルは、質量分析によって得られたデータを用いて訓練される、本明細書に開示される提示モデルの実施形態である。上記に述べたように、提示モデルは、アレル相互作用及びアレル非相互作用変数の2つの異なるセットに基づいてペプチド提示尤度の予想を生成した。具体的には、例示モデル4は、MHCクラスII分子のタイプ及びペプチド配列（NetMHCII2.3モデルで用いたものと同じ変数）に基づいてペプチド提示尤度の予想を生成し、例示モデル3は、MHCクラスII分子のタイプ、ペプチド配列、RNA発現、遺伝子識別子、及びフランキンゲ配列に基づいてペプチド提示尤度の予想を生成した。

## 【0479】

ペプチドの試験データセット内のペプチドがMHCクラスII分子によって提示される尤度を図14Lの例示モデルを用いて予想するのに先立って、各モデルを訓練及び検証した。NetMHCII2.3モデル（例示モデル1及び例示モデル2）は、免疫エピトープデータベース（IEDB, [www.iedb.org](http://www.iedb.org)）に寄託されたHLAペプチド結合親和性アッセイに基づいてそれ自体の訓練及び検証データセットを用いて訓練及び検証した。NetMHCII2.3モデルを訓練するために使用される訓練データセットは、

ほとんど15マーペプチドのみを含むことが知られている。これに対して、例示モデル3及び4は、図14Hに関して上記に述べた訓練データセットを用いて訓練し、図14Hに関して上記に述べた検証データセットを用いて検証した。

#### 【0480】

各モデルの訓練及び検証に続き、モデルのそれぞれを試験データセットを用いて試験した。上記に述べたように、NetMHCII2.3モデルはほとんど15マーペプチドのみを含むデータセットで訓練されるが、これはNetMHCII2.3が異なる重さのペプチドに異なる優先順位を与えることができないことを意味し、これにより、すべての長さのペプチドを含むHLAクラスIIの提示の質量分析データに対するNetMHCII2.3の予想性能が低下する。したがって、変動するペプチド長によって影響されないモデル間の適正な比較を与えるため、試験データセットは15マーペプチドのみを含むものとした。具体的には、試験データセットは、933種の15マーペプチドを含むものとした。試験データセット内の933種のペプチドのうちの40種が、MHCクラスII分子により、具体的には、HLA-DRB1\*07:01、HLA-DRB1\*15:01、HLA-DRB4\*01:03、及びHLA-DRB5\*01:01分子により提示された。試験データセットに含まれるペプチドは、上記に述べた訓練データセットからは除外されていた。

#### 【0481】

試験データセットを用いて各例示モデルを試験するため、例示モデルのそれぞれについて、試験データセット内の933種のペプチドの各ペプチドについて、モデルによってペプチドの提示尤度の予想を生成した。具体的には、試験データセット内の各ペプチドについて、例示1のモデルは、MHCクラスII分子によるそのペプチドの提示スコアを、MHCクラスII分子のタイプ及びペプチド配列を用い、試験データセット内の4つのHLAクラスII DRアレルにわたって最小NetMHCII2.3により予想された結合親和性によってペプチドをランク付けすることにより生成した。同様に、試験データセット内の各ペプチドについて、例示2のモデルは、MHCクラスII分子によるそのペプチドの提示スコアを、MHCクラスII分子のタイプ及びペプチド配列を用い、試験データセット内の4つのHLAクラスII DRアレルにわたって最小NetMHCII2.3により予想された結合ランク（すなわち、分位正規化した結合親和性）によってペプチドをランキングすることにより生成した。試験データセット内の各ペプチドについて、例示4のモデルは、MHCクラスII分子のタイプ及びペプチド配列に基づいてMHCクラスII分子によるそのペプチドの提示尤度を生成した。同様に、試験データセット内の各ペプチドについて、例示3のモデルは、MHCクラスII分子のタイプ、ペプチド配列、RNA発現、遺伝子識別子、及びフランキンク配列に基づいてMHCクラスII分子によるそのペプチドの提示尤度を生成した。

#### 【0482】

これら4つの例示モデルのそれぞれの性能を図14Lの線グラフに示す。具体的には、4つの例示モデルのそれぞれは、モデルによって生成された各予想について、真の陽性率の偽陽性率に対する比を示すROC曲線に関連付けられる。例えば、図14Lは、予想を生成するために最小NetMHCII2.3により予想された結合親和性を用いた例示1のモデルのROC曲線、予想を生成するために最小NetMHCII2.3により予想された結合ランクを用いた例示2のモデルのROC曲線、MHCクラスII分子のタイプ及びペプチド配列に基づいてペプチド提示尤度を生成した例示4のモデルのROC曲線、ならびに、MHCクラスII分子のタイプ、ペプチド配列、RNA発現、遺伝子識別子、及びフランキンク配列に基づいてペプチド提示尤度を生成した例示3のモデルのROC曲線を示している。

#### 【0483】

上記に述べたように、あるペプチドがMHCクラスII分子によって提示される尤度を予測するうえでのモデルの性能は、そのモデルによって生成された各予測について、真の陽性率の偽陽性率に対する比を示すROC曲線のAUCを特定することによって定量化さ

れる。AUCが大きいモデルほど、AUCが小さいモデルと比較してより高い性能（すなわちより高い精度）を有する。図14Lに示されるように、MHCクラスII分子のタイプ、ペプチド配列、RNA発現、遺伝子識別子、及びフランキング配列に基づいてペプチド提示尤度を生成した例示3のモデルの曲線は、0.95と最も高いAUCを実現した。したがって、MHCクラスII分子のタイプ、ペプチド配列、RNA発現、遺伝子識別子、及びフランキング配列に基づいてペプチド提示尤度を生成した例示3のモデルは、最も良好な性能を実現した。MHCクラスII分子のタイプ及びペプチド配列に基づいてペプチド提示尤度を生成した例示4のモデルの曲線は、0.91と2番目に高いAUCを実現した。したがって、MHCクラスII分子のタイプ及びペプチド配列に基づいてペプチド提示尤度を生成した例示4のモデルは、2番目に良好な性能を実現した。予想を生成するために最小NetMHCII2.3により予想した結合親和性を用いた例示1のモデルの曲線は、0.75と最も低いAUCとなった。したがって、予想を生成するために最小NetMHCII2.3により予想した結合親和性を用いた例示1のモデルの曲線は、最も悪い性能となった。予想を生成するために最小NetMHCII2.3により予想した結合ランクを用いた例示2のモデルの曲線は、0.76と2番目に低いAUCとなった。したがって、予想を生成するために最小NetMHCII2.3により予想した結合ランクを用いた例示2のモデルの曲線は、2番目に悪い性能となった。

【0484】

図14Lに示されるように、例示モデル1及び2と例示モデル3及び4との間の性能の隔たりは大きい。具体的には、NetMHCII2.3モデル（最小NetMHCII2.3により予想した結合親和性、または最小NetMHCII2.3により予想した結合ランクのいずれかの基準を用いるもの）の性能は、本明細書に開示される提示モデル（MHCクラスII分子の種類及びペプチド配列、またはMHCクラスII分子の種類、ペプチド配列、RNA発現、遺伝子識別子、及びフランキング配列に基づいてペプチド提示尤度を生成するもの）の性能よりもほぼ25%低い。したがって、図14Lは、本明細書に開示される提示モデルは、現在のベスト・イン・クラスの従来モデルであるNetMHCII2.3モデルよりも大幅に正確な提示予想を実現することが可能であることを示すものである。

【0485】

いっそうさらには、上記で述べたように、NetMHCII2.3モデルは、ほとんど15マーペプチドのみを含む訓練データセットで訓練される。その結果、NetMHCII2.3モデルは、どのペプチド長がMHCクラスII分子によってより提示されやすいかを学習するには訓練されない。したがって、NetMHCII2.3モデルは、MHCクラスII分子によるペプチド提示の尤度のその予想を、ペプチドの長さにしたがって重み付けしない。換言すれば、NetMHCII2.3モデルは、MHCクラスII分子によるペプチド提示の尤度のその予想をアミノ酸15個の最頻ペプチド長の外側の長さを有するペプチドについては変更しない。その結果、NetMHCII2.3モデルは、アミノ酸15個よりも長いまたは短い長さを有するペプチドの提示の尤度を過剰に予想する。

【0486】

これに対して、本明細書に開示される提示モデルは、質量分析によって得られたペプチドデータを用いて訓練されることから、すべての異なる長さのペプチドを含む訓練データセットで訓練することができる。その結果、本明細書に開示される開示モデルは、どのペプチド長がMHCクラスII分子によってより提示されやすいかを学習することができる。したがって、本明細書に開示される提示モデルは、MHCクラスII分子によるペプチド提示の尤度のその予測を、ペプチドの長さにしたがって重み付けすることができる。換言すれば、本明細書に開示される提示モデルは、MHCクラスII分子によるペプチド提示の尤度のそれらの予想をアミノ酸15個の最頻ペプチド長の外側の長さを有するペプチドについて変更することができる。その結果、本明細書に開示される提示モデルは、現在のベスト・イン・クラスの従来モデルであるNetMHCII2.3モデルと比較して、

アミノ酸15個よりも長いまたは短い長さを有するペプチドについて、大幅に正確な提示予想を実現することが可能である。これは、MHCクラスII分子によるペプチド提示の尤度を予想するために本明細書に開示される提示モデルを用いることの利点の1つである。

【0487】

### XII. A. 3. 実施例3

図14Mは、HLAクラスII分子を含むヒト腫瘍（NSCLC、リンパ腫、及び卵巣癌）及び細胞株（EBV）を含む合計230種の試料の各試料について0.1未満のq値で質量分析を用いてシーケンシングしたペプチドの量を示すヒストグラムである。図14Mに示されるように、0.1未満のq値で各試料について平均で1300種のペプチドをシーケンシングした。

10

【0488】

図14Dに関して上記に述べたように、図14Mの230種の試料の各試料は、HLAクラスII分子を含んでいた。より具体的には、図14Mの230種の試料の各試料はHLA-DR分子を含んでいた。HLA-DR分子は、HLAクラスII分子の一種である。さらにより具体的には、図14Mの230種の試料の各試料は、HLA-DRB1分子、HLA-DRB3分子、HLA-DRB4分子、及び／またはHLA-DRB5分子を含んでいた。HLA-DRB1分子、HLA-DRB3分子、HLA-DRB4分子、及びHLA-DRB5分子は、HLA-DR分子の種類である。

【0489】

この特定の実験は、HLA-DR分子、特にHLA-DRB1分子、HLA-DRB3分子、HLA-DRB4分子、及びHLA-DRB5分子を含む試料を用いて行ったが、代替的な実施形態では、この実験はHLAクラスII分子の任意の種類（複数可）のうちの1つ以上を含む試料を用いて行うことができる。例えば、代替的な実施形態では、HLA-DP及び／またはHLA-DQ分子を含む試料を用いて同一の実験を行うことができる。同じ方法を用いて任意の種類（複数可）のMHCクラスII分子をモデル化し、なおも信頼性の高い結果を得ることが可能であることは、当業者には周知である。例えば、Jensen、Kamilla Kjaergaard et al.<sup>76</sup>は、HLA-DR分子に対する、ならびにHLA-DP及びHLA-DQ分子に対する結合親和性をモデル化するために同一の方法を用いている最近の科学論文の一例である。したがって、当業者であれば、本明細書に記載される実験及びモデルを用いることで、HLA-DR分子のみでなく、他の任意のMHCクラスII分子を別々または同時にモデル化し、なおも信頼性の高い結果を得ることができる点は理解されよう。

20

30

【0490】

合計230種の試料の各試料のペプチドをシーケンシングするため、各試料について質量分析を行った。試料について得られた質量スペクトルをCometで検索し、Percolatorでスコアリングしてペプチドをシーケンシングした。次いで、試料中のシーケンシングされたペプチドの量を複数の異なるPercolatorのq値の閾値について特定した。具体的には、その試料について、0.01未満のPercolatorのq値、0.05未満のPercolatorのq値、及び0.2未満のPercolatorのq値を用いてシーケンシングしたペプチドの量を決定した。

40

【0491】

203種の試料の各試料について、異なるPercolatorのq値の閾値のそれぞれでシーケンシングしたペプチドの量を図14Mに示す。例えば、図14Mにみられるように、最初の試料については、0.1未満のq値を有するおよそ8000種のペプチドを質量分析を用いてシーケンシングした。

【0492】

全体として、図14Mは、質量分析を用いてMHCクラスII分子を含有する試料から大量のペプチドを、低いq値でシーケンシングすることができることを示している。換言すると、図14Mに示されるデータは、質量分析を用いて、MHCクラスII分子によ

50

って提示されうるペプチドを高い信頼度でシーケンシングすることができることを実証するものである。

【0493】

図14Nは、特定のMHCクラスII分子のアレルが同定された試料の量を示すヒストグラムである。より具体的には、図14Nは、HLAクラスII分子を含む合計で230種の試料について、ある特定のMHCクラスII分子が同定された試料の量を示している。

【0494】

図14Mに関して上記に述べたように、図14Mの230種の試料の各試料は、HLA-DRB1分子、HLA-DRB3分子、HLA-DRB4分子、及び／またはHLA-DRB5分子を含んでいた。したがって、図14Nは、HLA-DRB1分子、HLA-DRB3分子、HLA-DRB4分子、及びHLA-DRB5分子について、ある特定の

10

【0495】

試料中にどのHLA-DRB1、HLA-DRB3、HLA-DRB4、及びHLA-DRB5アレルが存在するかを特定するため、その試料についてHLAクラスII DRのタイピングを行った。次いで、特定のHLAアレルが同定された試料の量を特定するために、HLAクラスII DRのタイピングを用いてHLAアレルが同定された試料の数を単純に合計した。例えば、図14Nに示されるように、合計で230種の試料のうちの28種の試料が、HLAクラスII分子のアレルHLA-DRB3\*03:01を含んでいた。換言すると、合計で230種の試料のうちの28種の試料がHLA-DRB3分子についてアレルHLA-DRB3\*03:01を含んでいた。全体として、図14Nは、HLAクラスII分子を含む230種の試料から広範囲のHLAクラスII分子のアレルを同定することができることを示している。ヒト集団について言えば、米国特許出願第の白人集団におけるHLA-DRB1アレルのアレル頻度を、Maier s, M, et al. にみることができる<sup>161</sup>。

20

【0496】

図14Oは、MHCクラスI分子に結合したペプチドと、MHCクラスII分子に結合したペプチドを示している<sup>162</sup>。図14Oに示されるように、各ペプチドは、ペプチド骨格及び複数のアミノ酸を含んでいる。各MHC分子は、結合溝を有している。しかしながら、後述するように、MHCクラスI及びMHCクラスII分子の結合溝内へのペプチドの結合の仕方は異なる。

30

【0497】

本開示の全体を通じて述べるように、MHC分子によって提示されるペプチドの長さは異なり得る。具体的には、MHC分子によって提示されるペプチドはアミノ酸9〜20個の長さであり得る。ペプチドがMHC分子に結合してMHC分子によって提示される際、ペプチドの「結合コア」がMHC分子の結合溝内に位置する。具体的には、ペプチドの結合コアとは、ペプチドがMHC分子に結合してMHC分子によって提示される際にMHC分子の結合溝内に位置するペプチドのアミノ酸の配列である。さらに、ペプチドがMHC分子に結合してMHC分子によって提示される際、ペプチドの結合コアの「結合アンカー」が、MHC分子の結合溝内に物理的に結合する。具体的には、ペプチドの結合コアの結合アンカーとは、ペプチドがMHC分子に結合してMHC分子によって提示される際にMHC分子の結合溝に結合する結合コアの特定のアミノ酸である。

40

【0498】

図14Oに示されるように、MHCクラスI分子によって提示されるペプチドの結合コアは、ペプチドの全長を含んでいる。具体的には、図14Oに示されるように、MHCクラスI分子によって提示されるペプチドの全体がMHCクラスI分子の結合溝内に位置している。これに対して、MHCクラスII分子によって提示される分子では、ペプチドのアミノ酸の部分配列のみがペプチドの結合コアに含まれ得る。具体的には、図14Oに示されるように、MHCクラスII分子によって提示されるペプチドの両端は、MHCクラスII分子の結合溝内に位置していない。MHCクラスIIによって提示されるペプチド

50

の結合コアを構成するアミノ酸の部分配列は不明である場合もある。しかしながら、文献に述べられているように、MHCクラスII提示ペプチドの結合コアの最も一般的な長さは、アミノ酸9個である。

【0499】

さらに、MHCクラスII提示ペプチドの結合コアが不明であることに加えて、ペプチドの結合コアの結合アンカーを構成するアミノ酸の量及び位置も知られている場合もある。しかしながら、文献に述べられているように、MHCクラスII提示ペプチドの結合コアは通常、3～4個の結合アンカーを含み、結合アンカーは通常、結合コアの両端に位置するアミノ酸を含む。

【0500】

MHCクラスIとMHCクラスII分子へのペプチドの結合の相違のため、最適なペプチド提示予測性能を確保するには、ペプチド提示予測モデルが、MHCクラスII分子によるペプチド提示を詳細に予測するように構成されなければならない。具体的には、結合コアを含むアミノ酸の部分配列及びMHCクラスII分子により提示されるペプチドの結合コアの結合アンカーは不明であり得ることから、MHCクラスIIペプチド提示予測モデルはこの不確定性をモデル化するように構成されなければならない。詳細には、MHCクラスII分子によって提示されるペプチドの結合コア及び結合アンカーの不確定性をモデル化するためのインセプションモデルが開発された。

【0501】

図14PはMHCクラスII分子によるペプチド提示を予測するように構成された、図14Qのインセプションモデルのインセプションニューラルネットワークの例示的な実施形態を示す。インセプションモデルは、MHCクラスII分子によって提示されるペプチドの結合コア及び結合アンカーを特定し、これらの特定された結合コア及び結合アンカーを用いてMHCクラスII分子によるペプチドの提示を予測するように構成された提示モデルである。インセプションモデルは、アレル非相互作用特性（例えば、RNA配列、試料番号、タンパク質番号、及びフランキング配列）を受け入れる共有ニューラルネットワークと、アレル相互作用特性（例えば、ペプチド配列）を受け入れる別個のインセプションニューラルネットワークのセットとを含む。詳細には、別個のインセプションニューラルネットワークのセットの中の各別個のインセプションニューラルネットワークは、異なるMHCクラスIIアレル（例えば、HLA-DRBアレル）に関連付けられ、コード化されたペプチド配列を受け入れるように構成される。上記に述べたように、図14Pは、インセプションモデルのインセプションニューラルネットワークの例示的な実施形態を示す。

【0502】

最初に、MHCクラスII分子によって提示されるペプチドの長さは可変である（例えば、アミノ酸9～20個）ため、アミノ酸20個の最大長さよりも短いペプチドをアミノ酸20個の長さを有するようにパディングする。詳細には、ペプチドの長さがアミノ酸20個未満である場合、特定のアミノ酸Zをペプチドの左側に付加し、次いで、ペプチドの右側に付加する。このようなペプチドのパディングのパターンをペプチドがアミノ酸20個の長さを有するようになるまで繰り返す。ペプチドの両側にパディングすることにより、ペプチド長がすべてのペプチドにわたって一定とされる一方でペプチドの結合コアは不変に維持される。

【0503】

インセプションニューラルネットワークの入力層は、パディングされたペプチド配列を受け入れる。次いでパディングされたペプチドはワンホットでコード化される。図14Pに示されるように、各インセプションニューラルネットワークは、3つの一次元CNN層を含んでいる。3つのCNN層のうちの1つは、サイズ8の16個のフィルターを有している。3つのCNN層のうちの1つは、サイズ10の16個のフィルターを有している。3つのCNN層のうちの1つは、サイズ12の16個のフィルターを有している。これらのフィルターサイズは、上記に述べたように、文献においてMHCクラスIIにより提示

10

20

30

40

50

されるペプチドの最も一般的な結合コアの長さとして示されている約9個のアミノ酸からなる結合コアをインセプションニューラルネットワークに特定させるために意図的に選択されたものである。

【0504】

3つのCNN層のそれぞれの出力は、2つの一次元CNN層に入力される。2つのCNN層のうちの一方は、サイズ1の32個のフィルターを有している。2つのCNN層のうちの一方は、サイズ2の32個のフィルターを有している。これらのフィルターサイズは、MHCクラスIIによって提示されるペプチドの結合コア内の結合アンカーの位置を特定するために意図的に選択されたものである。

【0505】

これら2つのCNN層の出力は連結される。次いで、各連結された出力はbi-LSTM層に与えられる。各bi-LSTM層の出力は連結され、この連結結果は多層パーセプトロンに送信される。多層パーセプトロンの出力は、別個のインセプションニューラルネットワークの出力を含む。換言すれば、多層パーセプトロンの出力は、別個のインセプションニューラルネットワークに入力されるペプチドが、別個のインセプションニューラルネットワークに関連付けられたMHCクラスIIアレルによって提示される尤度を含む。各別個のインセプションニューラルネットワークからの提示尤度は、共有ニューラルネットワークからの出力と組み合わせられる。最後に、組み合わせられた尤度を合計してペプチドがMHCクラスIIアレルのうちの1つ以上によって提示される全体の尤度を生成する。

【0506】

図14Qは、ペプチドの試験データセット内のペプチドが、試験データセット内に存在するMHCクラスII分子の少なくとも1つによって提示される尤度を予測するうえでの「bi-LSTM」及び「インセプション」提示モデルの性能を比較した線グラフである。具体的には、図14Qは、「bi-LSTM」モデル及び「インセプション」モデルの完全適合率-再現率曲線を示す線グラフである。AUCを用いて各モデルの性能を定量化する。

【0507】

図14Qで試験される第1のモデルは、「bi-LSTM」モデルである。bi-LSTMモデルは、図14I及び14Jに関して上記に詳細に述べたモデルである。

【0508】

図14Qで試験される第2のモデルは、「インセプション」モデルである。インセプションモデルは、図14Pに関して上記に詳細に述べたモデルである。

【0509】

ペプチドの試験データセット内のペプチドがMHCクラスII分子によって提示される尤度をこれらのモデルを用いて予測するのに先立って、各例示モデルを訓練及び検証した。例示モデルを訓練、検証、及び最終的に試験するため、230種の試料について上記に述べたデータを訓練、検証、及び試験データセットに分けた。

【0510】

訓練、検証、及び試験データセットのうちの複数のデータセットにペプチドが現れないようにするため、以下の手順を行った。最初に、プロテオーム内の複数の位置に現れた、230種の試料に由来するすべてのペプチドを除去した。次いで、230種の試料に由来する残りのペプチドを10個の隣接するペプチドのブロックに分割した。隣接するペプチドの各ブロックを、訓練データセット、検証データセット、または試験データセットに一意に割り当てた。これにより、訓練、検証、及び試験データセットの複数のデータセットに現れるペプチドはなくなった。

【0511】

訓練データセットは、合計230種の試料のうちの226種に由来するMHCクラスII分子によって提示された188、210種のペプチドを含んでいた。訓練データセットに含まれる188、210種のペプチドは、アミノ酸9~20個の長さであった。bi-LSTMモデル及びインセプションモデルは、ADAM最適化アルゴリズム及び早期停止

10

20

30

40

50

を用いた訓練データセットでそれぞれ訓練した。

【0512】

検証データセットは、訓練データセットで用いられた同じ226種の試料に由来するMHCクラスII分子によって提示された21, 764種のペプチドを含んでいた。検証データセットは、早期停止のみに用いた。

【0513】

試験データセットは、質量分析を用いて腫瘍試料から特定されたMHCクラスII分子によって提示されたペプチドを含んでいた。具体的には、4つの腫瘍試料から特定された232種のペプチドで構成されていた。試験データセットに含まれるペプチドは、上記に述べた訓練データセットからは除外されていた。

10

【0514】

訓練データセット及び検証データセットをそれぞれ用いたBi-LSTM及びインセプションモデルの訓練及び検証後に、各モデルを試験データセットを用いて試験した。試験データセットに対するBi-LSTM及びインセプションモデルの性能を、完全適合率-再現率曲線及びAUCスコアとして図14Qに示す。図14Qに示されるように、Bi-LSTMモデルの性能はインセプションモデルを上回り、 $AUC = 0.347$ を実現した。Bi-LSTMは $AUC = 0.238$ を実現した。

【0515】

#### XII. A. 4. 実施例4

本明細書に開示される予測モデルをクラスIIのHLAペプチドに適用できるかどうかをさらに評価するため、公開されているクラスIIの質量分析データを、それぞれが単一のHLAクラスIIアレルを発現する2つの細胞株について得た。一方の細胞株はHLA-DRB1\*15:01を発現し、他方はHLA-DRB5\*01:01を発現した<sup>150</sup>。これら2つの細胞株を訓練データ用に使用した。試験データ用には、クラスIIの質量分析データをHLA-DRB1\*15:01及びHLA-DRB5\*01:01の両方を発現する別の細胞株から取得した<sup>151</sup>。RNAシーケンシングデータは、訓練用及び試験用細胞株のいずれでも入手できなかったため、異なるB細胞株B721.221からのRNAシーケンシングデータ<sup>92</sup>で代用した。

20

【0516】

クラスIIのデータでは9~20個の長さを有するペプチドを含めた点を除いて、HLAクラスIIデータの場合と同じ手順を用いて、ペプチドセットを訓練、検証、及び試験セットに分割した。訓練データは、HLA-DRB1\*15:01により提示された330種のペプチド、及びHLA-DRB5\*01:01により提示された103種のペプチドを含むものとした。検査データセットは、HLA-DRB1\*15:01またはHLA-DRB5\*01:01により提示された223種のペプチド及び4708種の非提示ペプチドを含むものとした。

30

【0517】

本発明者らは、10個のモデルのアンサンブルを訓練データセットで訓練してHLAクラスIIのペプチド提示を予測した。これらのモデルのアーキテクチャ及び訓練手順は、クラスIIのモデルでは入力ペプチドとして長さ11個ではなく長さ20個にワンホットでコード化され、パディングが0の配列を採用した点を除き、クラスIの提示を予測するために用いたものと同じとした。

40

【0518】

図15は、「MSモデル」、HLA-DRB1\*15:01及びHLA-DRB5\*01:01にわたって最も低いNetMHCIIpanパーセンタイルランクを採用する、「NetMHCIIpan rank」: NetMHCIIpan 3.1<sup>152</sup>、及びHLA-DRB1\*15:01及びHLA-DRB5\*01:01にわたってnM単位の最も強い親和性を採用する「NetMHCIIpan nM」: NetMHCIIpan 3.1の、HLA-DRB1\*15:01 / HLA-DRB5\*01:01試験データセット内でのペプチドのランク付けにおける予測性能を比較したものである。「MSモデル」は

50

、本明細書に開示されるMHCクラスIIの提示予測モデルである。

【0519】

詳細には、図15は、受信者操作特性(ROC)曲線、ならびにROC曲線下面積AUC(パネルA)及びAUC<sub>0.1</sub>(パネルB)統計をこれらのランク付け法について示したものである。AUC<sub>0.1</sub>は、エピトープ予測フィールドで一般的に考慮される、0~0.1FPR\*10のAUCである<sup>19</sup>。NetMHCIIpan nM法とNetMHCIIpan rank法の性能は同様であった。MSモデルの性能が最も高く、特にROC曲線の臨界高特異性領域において、比較した方法の性能を大幅に上回った(AUC<sub>0.1</sub>が0.27に対して0.41)。

【0520】

XII.B. MHCクラスIIアレルについて決定された提示モデルパラメータの例

以下は、クラスII MHCアレルであるHLA-DRB1\*12:01及びHLA-DRB1\*10:01の潜在的なアレルごとの提示尤度を生成する複数アレル提示モデル(式(16))のバリエーションについて決定されたパラメータのセットを示す。

$$u = \expit(\text{relu}(X \cdot W^1 + b^1) \cdot W^2 + b^2)$$

式中、 $\text{relu}(\cdot)$ は、正規化線形ユニット(ReLU)関数、 $W^1$ 、 $b^1$ 、 $W^2$ 、及び $b^2$ は、モデルについて決定されたパラメータ $\theta$ のセットである。アレル相互作用変数 $X$ は、入力ペプチド当たり1行のワンホットコード化され、中間パディングされた(middle-padded)ペプチド配列からなる $1 \times 399$ 行列に含まれる。 $W^1$ の大きさは $(399 \times 256)$ 、 $b^1$ の大きさは $(1 \times 256)$ 、 $W^2$ の大きさは $(256 \times 2)$ 、 $b^2$ の大きさは $(1 \times 2)$ である。出力の第1の行は、アレルHLA-DRB1\*12:01によるそのペプチド配列の潜在的なアレルごとの提示の確率を示し、出力の第2の行は、アレルHLA-DRB1\*10:01によるそのペプチド配列の潜在的なアレルごとの提示の確率を示す。実証目的で、 $b^1$ 、 $b^2$ 、 $W^1$ 、及び $W^2$ を補遺Aに示す。

【0521】

XIII. 実施例9: T細胞データのMHCクラスII提示モデルの評価

MHCクラスIIアレルによるペプチド提示の正確な予測が、ヒト腫瘍CD4エピトープ(すなわち、免疫療法のターゲット)を特定する能力につながるかどうかを評価するため、公開されているCD4+T細胞マルチマー/テトラマーアッセイデータをImmunepitope Database (IEDB)<sup>88</sup>よりダウンロードした。これらのデータは、14個のHLA-DRB1アレル、2個のHLA-DRB3アレル、1個のHLA-DRB4アレル、及び1個のHLA-DRB5アレルを含む18個の別個のHLA-DRBアレルを有するヒト試料からの長さ9~20残基の3,470種のペプチドからなるものであった。平均で、各アレルはそのアレルを含む33種の試料を有した。完全MHCクラスII MSモデル(セクションXII.A.2で上記に述べたものと同じモデル)を結合親和性予測ツールNetMHCII2.3に対して比較した。18個のアレルにわたって、完全MHCクラスII MSモデルの平均ROC曲線下面積(ROC AUC)は0.81であり、標準偏差が0.08であったのに対して、NetMHCII2.3のROC AUCは0.65に過ぎず、標準偏差は0.13とより大きかった。これらの結果は、CD4 T細胞エピトープを予測するうえでの完全MHCクラスII MSモデルのより優れた能力を証明するものである。アレルごとベースでは、HLA-DRB1\*01:01のようなより一般的なアレルのいくつかについて、ROC AUCは2つのモデル間で大幅に似通っていた。例えば、HLA-DRB1\*01:01では、完全MHCクラスII MSモデルのROC AUCは0.83であり、NetMHCII2.3モデルのROC AUCは0.81であった。しかしながら、大部分のアレルで、2つのモデル間の性能にはより幅広い違いがみられた。18個のアレルごとの試験のうち、完全MHCクラスII MSモデルは、17個のアレルでNetMHCII2.3モデルの性能を上回った。HLA-DRB1\*15:02の1個のアレルのみにおいて、NetMHCII2.3が完全MHCクラスII MSモデルの性能を上回った。しかしながら、このアレルは、そのアレルを含む試料が1つしか含まれていなかった完全MHCクラスII

M Sモデルの訓練データではよく表現されていなかった。

【0522】

#### X I V . 実施例 10 : 遡及的新生抗原 T 細胞データの M H C クラス I I 提示モデルによる評価

本実施例では、M H C クラス I I 分子によるペプチド提示の正確な予測が、ヒト腫瘍 C D 4 T 細胞エピトープを特定する能力につながるかどうかをさらに評価する。この評価を行うため、M H C クラス I I 提示モデルにより予測されたペプチドの C D 4 + 免疫原性をランク付けした。

【0523】

この評価に適切な試験データセットには、腫瘍細胞表面上の M H C クラス I I 分子によって提示され、T 細胞によって認識されるペプチドが含まれる。さらに、正式な性能評価には、陽性として示される（すなわち、T 細胞により認識された）ペプチドだけでなく、十分な数の陰性標識された（すなわち、試験したが、T 細胞により認識されなかった）ペプチドも必要とされた。質量分析データセットは、腫瘍提示に対応するが T 細胞認識には対応しない。逆に、ワクチン接種後のプライミングまたは T 細胞アッセイは T 細胞前駆細胞の存在に対応するが腫瘍提示には対応しない。例えば、ソース遺伝子の腫瘍内での発現が低レベルである強力な H L A 結合ペプチドは、こうしたペプチドは腫瘍によって提示されないため、治療上の有用性のない強い免疫後 C D 4 T 細胞応答を生じ得る。

【0524】

この評価に適した試験データセットを得るため、公開されているデータを最近の研究から収集した<sup>163</sup>。収集した試験データセットには、45人の患者にわたった、T I L に対して C D 4 + 応答性である69個の陽性標識された単一ヌクレオチド変異体 (S N V) 変異を含めた。上記に述べたように、収集した試験データセットには、陰性標識された S N V 変異も含めた。具体的には、患者当たり、平均で104個、中央値で106個の陰性標識 S N V 変異があった。

【0525】

試験データセットの各 S N V 変異を、配列の中央の13位のアミノ酸に S N V 変異を有するアミノ酸25個の配列として表した。次に、25個のアミノ酸の各配列について、S N V 変異を含むアミノ酸9個~20個の長さのすべての可能なペプチドを生成した。25個のアミノ酸の各配列は、118種の可能なペプチドを生じた。それぞれの可能なペプチドについて、アミノ酸5個のフランキング配列をペプチドの左側と右側に付加した。

【0526】

個別化免疫療法用の抗原の選択をシミュレートするため、試験データセット内の各患者の S N V 変異を、本明細書に開示されるインセプションモデル及び N e t M H C I I P a n 3 . 2 結合親和性モデルを T P M = 1 の遺伝子発現量閾値で用いて、患者の M H C クラス I I アレルによる提示の尤度の順番にランク付けした。使用したインセプションモデルは、試験データセット内の患者に存在する30種の M H C クラス I I アレルのうちの25種をカバーする、32種の異なる M H C クラス I I アレルによるペプチド提示を予測するように訓練した。

【0527】

インセプションモデルでは、各患者のそれぞれの S N V 変異の提示の尤度を計算するため、患者の特定された M H C クラス I I アレルのそれぞれについて、S N V 変異の118種の可能なペプチドのそれぞれの提示スコアをインセプションモデルを用いて決定した。次に、患者の M H C クラス I I アレルのそれぞれについてインセプションモデルによって決定された最も高い提示スコアを特定した。最後に、患者の M H C クラス I I アレルのそれぞれについてのこれらの最も高い提示スコアを合計して、患者の S N V 変異の全体の提示の尤度を求めた。

【0528】

N e t M H C I I P a n 3 . 2 モデルでは、各患者のそれぞれの S N V 変異の提示の尤度を計算するため、患者の特定された M H C クラス I I アレルのそれぞれについて、S N

10

20

30

40

50

V変異の118種の可能なペプチドのそれぞれの結合親和性をNetMHCIIPan3.2モデルを用いて決定した。次に、患者のMHCクラスIIアレルのそれぞれについてNetMHCIIPan3.2モデルによって決定された最も高い逆結合親和性を特定した。より低い結合親和性はより高い提示の尤度を示すため、最も高い逆結合親和性を特定した点に注意されたい。最後に、患者のMHCクラスIIアレルのそれぞれについてのこれらの最も高い逆結合親和性を合計して、患者のSNV変異の全体の提示の尤度を求めた。

【0529】

次に、各患者のSNV変異を、インセプションモデル及びNetMHCIIPan3.2モデルの両方で求めた、患者のMHCクラスIIによる提示の尤度の順番にランク付けした。抗原特異的免疫療法は、標的とするMHCクラスIIの特異性の数が技術的に制限されることから（例えば、現在の個別化ワクチンはおおよそ10～20個の体細胞変異をコードしており<sup>80～82</sup>、そのうちの約10個がMHCクラスII特異的なものであり得る）、各患者について上位1、2、3、4、5、及び10位のSNV変異をランク付けした。

【0530】

さらに、コントロールとして、各患者について、TPMが1以上である遺伝子に由来する患者のSNV変異のそれぞれをランダムにランク付けした。具体的には、各患者について、TPMが1以上である遺伝子に由来する患者のSNV変異のそれぞれを100回のトライアルでランダムにランク付けして、各患者の各SNV変異の全体のランク付けを決定した。

【0531】

SNV変異のランク付けを行った後、予測モデルを、少なくとも1つの既存のT細胞応答を有する各患者において上位1、2、3、4、5、及び10位にランクされたSNV変異における既存のT細胞応答の数をカウントすることによって比較した。次に、少なくとも1つの既存のT細胞応答を有する各患者において異なるモデルによって特定された上位1、2、3、4、5、及び10位にランクされたSNV変異について、T細胞（例えば、既存のT細胞応答）により認識されたSNV変異の割合を比較した。具体的には、下記表2に、合計69種の陽性標識SNV変異のうち、特定のモデルによって上位1、2、3、4、5、及び10位の予測として予測された陽性標識SNV変異の割合（%）を示す。表2に示されるように、インセプションモデルは、NetMHCIIPan3.2モデル及びランダムな予測と比較して、CD4+免疫原性MHCクラスII提示ペプチドを正確に予測する可能性が高い。

（表2）

| モデル             | 1位 | 2位  | 3位  | 4位  | 5位  | 10位 |
|-----------------|----|-----|-----|-----|-----|-----|
| インセプション         | 9% | 17% | 17% | 19% | 20% | 32% |
| NetMHCIIPan 3.2 | 9% | 12% | 16% | 18% | 19% | 29% |
| ランダム            | 1% | 3%  | 5%  | 6%  | 8%  | 16% |

【0532】

したがって、この評価は、従来の文献<sup>81, 82, 97</sup>に記載されるように、T細胞をブライミングできる新生抗原だけでなく、より厳密に、腫瘍によってT細胞に提示される新生抗原を特定することができるインセプションモデルのより優れた性能を確立するものである。

【0533】

#### XV. 実施例11：がん患者における新生抗原応答性T細胞の予期される特定

予期される本実施例では、改善された予測が通常の実験室からの新生抗原の特定を可能とすることを実証する。これを行うため、抗PD(L)1療法を行っている転移性NSCLCを有する患者からアーカイブされたFFPE腫瘍生検及び5～30mLの末梢血を分析する。体細胞変異（SNV及び短いindel）を、腫瘍全エクソームシーケンシング、腫瘍トランスクリプトームシーケンシング、及びマッチトノーマルエクソームシーケンシングを用いて各患者について特定する。MHCクラスII完全MSモデルを適

用して既存の抗腫瘍T細胞応答に対して試験を行うために患者当たり20個の新生エピトープを優先順位付けする。可能性の高いCD4応答に集中して分析を行うため、優先順位付けしたペプチドを8~11マーの最小エピトープとして合成し（「方法」）、次いで末梢血単核細胞（PBMC）を合成したペプチドと短期インビトロ刺激（IVS）培養中で培養して新生抗原応答性T細胞を増殖させる。2週間後、抗原特異的T細胞の存在を優先順位付けした新生エピトープに対するIFN- $\gamma$  ELISpotを用いて評価する。十分なPBMCが得られる患者で別々の実験をさらに行って認識された特異的抗原の完全または部分的な量み込みを行う。

#### 【0534】

最初に、患者特異的な新生抗原ペプチドプールに対するT細胞応答を、患者のそれぞれで検出する。それぞれの患者について、予測された新生抗原を、モデルランク付け及び任意の配列ホモロジーに従ってペプチドのそれぞれ2つのプールに組み合わせる（相同ペプチドは異なるプールに分割する）。次いで、それぞれの患者について、その患者についてインビトロ増殖させたPBMCをIFN- $\gamma$  ELISpotで2つの患者特異的な新生抗原ペプチドプールにより刺激する。DMSOのネガティブコントロール及びPHAのポジティブコントロールも行っており、バックグラウンド及びT細胞生存率をそれぞれ検出する。バックグラウンドの2倍よりも高い値を有する試料は陽性の応答性患者とみなす。さらに、インビトロ培養条件が、インビトロでのデノボプライミングを可能とするのではなく、既存のインビボプライミングされたメモリーT細胞のみを増殖させることを確認するため、HLAを一致させた健康なドナーで新生抗原を用いて一連のコントロール実験を行う。既存の新生抗原応答性T細胞が、IFN- $\gamma$  ELISpotを用いて患者特異的なペプチドプールで試験した患者の大部分で特定されることが予想される。さらに、大半の患者が、試験した新生抗原ペプチドのうちの少なくとも1つに応答することが予想される。

#### 【0535】

#### XV. A. ペプチド

カスタムメイドの組換え凍結乾燥ペプチドを購入し、滅菌DMSO中で10~50mMに戻し、一定分量に分けて-80℃で保存する。

#### 【0536】

#### XV. B. ヒト末梢血単核細胞（PBMC）

凍結保存した健康なドナーからのHLAタイピングされたPBMC（HIV、HCV、及びHBVについて結成陰性であることを確認済みのもの）を購入し、使用まで液体窒素中で保存する。新鮮な血液試料及びロイコパックも購入し、Ficoll-Paque密度勾配によりPBMCを単離した後、凍結保存する。患者のPBMCを地域の臨床処理センターで地域臨床標準業務手順書（SOP）及びIRBにより承認されたプロトコールに従って処理する。承認IRBは、Quorum Review IRB、Comitato Etico Interaziendale A.O.U. San Luigi Gonzaga di Orbassano、及びComite Etico de la Investigacion del Grupo Hospitalario Quiron en Barcelonaを含む。

#### 【0537】

PBMCを密度勾配遠心分離によって単離し、洗浄、カウントし、CryoStor CS10中、 $5 \times 10^6$ 細胞/mlで凍結保存する。凍結保存した細胞をcryoportで発送し、到着後にLN<sub>2</sub>中に移して保存する。凍結保存した細胞を解凍し、ベンゾナーゼを加えたOptmizer T-cell Expansion Basal Medium中で2回、ベンゾナーゼを加えない培地中で1回、洗浄する。細胞カウント及び生存率をGuava ViaCount試薬及びGuava easyCyte HTサイトメーター（EMD Millipore）上のモジュールを用いて評価する。次いで、細胞をその後のアッセイに適した濃度でアッセイに適した培地中に再懸濁する（次のセクションを参照）。

#### 【0538】

#### XV. C. インビトロ刺激 (IVS) 培養

健康なドナーまたは患者試料から得た既存のT細胞を、Ott et alにより適用されたアプローチ<sup>81</sup>と同様のアプローチで同族ペプチド及びIL-2の存在下で増殖させる。簡単に述べると、解凍したPBMCを一晩休ませ、24ウェル組織培養プレート中で10IU/mlのrhIL-2を添加したImmunoCult (商標) -XF T-cell Expansion Medium中、ペプチドプール (ペプチド当たり10μM) の存在下で14日間、刺激する。細胞を2×10<sup>6</sup>細胞/ウェルで播種し、培地の2/3を2〜3日ごとに交換することによって培養する。

【0539】

#### XV. D. IFNγ 酵素結合免疫スポット (ELISpot) アッセイ

IFNγ 産生T細胞の検出をELISpotアッセイ<sup>142</sup>により行う。簡単に述べると、PBMC (エキスピボまたはインビトロ増殖後のもの) を回収し、無血清RPMI中で洗浄し、抗ヒトIFNγ 捕捉抗体をコーティングしたELISpot Multiscreenプレート中、Optmizer T-cell Expansion Basal Medium (エキスピボ) またはImmunoCult (商標) -XF T-cell Expansion Medium (増殖させた培養物) 中でコントロールまたは同族ペプチドの存在下で培養する。5%CO<sub>2</sub>、37℃の加湿インキュベーター内で18時間インキュベートした後、細胞をプレートから除去し、膜に結合したIFNγ を、抗ヒトIFNγ 検出抗体、Vectastain Avidinペルオキシダーゼ複合体、及びAEC基質を用いて検出する。ELISpotプレートを乾燥させ、光から保護して保管し、標準化評価を行うために発送する<sup>143</sup>。

【0540】

#### XV. E. グランザイムB ELISA及びMSDマルチプレックスアッセイ

ELISpot上清中に分泌されたIL-2、IL-5及びTNF-αの検出をトリプレックスアッセイであるMSD U-PLEX Biomarkerアッセイ (カタログ番号K15067L-2) を使用して行う。アッセイは製造者の指示に従って行う。被検物質濃度 (pg/ml) を、各サイトカインについて既知の標準物質の連続希釈を用いて計算する。ELISpot上清中のグランザイムBの検出を、GranzymeB Duo Set (登録商標) ELISAを製造者の指示に従って使用して行う。簡単に述べると、ELISpot上清を試料希釈剤中、1:4に希釈し、グランザイムB標準の連続希釈液に並べて流して濃度 (pg/ml) を計算する。

【0541】

#### XV. F. IVSアッセイのネガティブコントロール実験—健康なドナーで試験した腫瘍細胞株由来の新生抗原

健康なドナーで試験した腫瘍細胞株由来の新生抗原についてIVSアッセイのネガティブコントロール実験を行う。かかる実験では、健康なドナーのPBMCを、IVS培養中、ポジティブコントロールペプチド (感染症に予め曝露したもの)、腫瘍細胞株由来のHLAを一致させた新生抗原 (曝露しないもの)、及びドナーが血清反応陰性である病原体に由来するペプチドを含むペプチドプールで刺激する。次いで、増殖させた細胞を、DMSO (ネガティブコントロール)、PHA及び一般的な感染症ペプチド (ポジティブコントロール)、新生抗原 (非曝露)、またはHIV及びHCVペプチド (ドナーが血清反応陰性であることを確認する) で刺激した後、IFNγ ELISpot (10<sup>5</sup>細胞/ウェル) により分析する。

【0542】

#### XV. G. IVSアッセイのネガティブコントロール実験—健康なドナーで試験される患者由来の新生抗原

健康なドナーにおける応答性について試験した患者由来の新生抗原についてIVSアッセイのネガティブコントロール実験を行う。具体的には、HLAを一致させた新生抗原ペプチドプールに対する健康なドナーにおけるT細胞応答の評価を行う。健康なドナーのPBMCを、エキスピボのIFNγ ELISpotにおいてコントロール (DMSO、CE

10

20

30

40

50

F、及びPHA) またはHLA一致させた患者由来新生抗原ペプチドで刺激する。さらに、新生抗原プールまたはCEFプールの存在下で増殖させたIVS培養後の健康なドナーのPBMCを、IFN $\gamma$ ELISpotにおいてコントロール(DMSO、CEF、及びPHA) またはHLAを一致させた患者由来の新生抗原ペプチドプールで刺激する。

【0543】

#### XVI. 実施例8～11の方法

以下の実施例8～11の方法は、将来的な予期される実施例10～11を実施するうえで用いられるので、未来形で説明される。しかしながら、以下の方法を説明するために未来形を用いてはいるが、これらの方法は実施例8及び9の実施において過去にも用いられているものである。

【0544】

#### XVI. A. 質量分析

##### XVI. A. 1. 試料

質量分析用にアーカイブされた凍結組織を販売元から得る。標本のサブセットも患者からプロスペクティブに採取される。

【0545】

##### XVI. A. 2. HLA免疫沈降法

HLAペプチド分子の単離を、組織試料の溶解及び可溶化後に確立された免疫沈降(IP)法を用いて行う<sup>87, 124-126</sup>。新鮮な凍結組織を粉碎し、溶解バッファー(1%CHAPS、20mM Tris-HCl、150mM NaCl、プロテアーゼ及びホスファターゼ阻害剤、pH=8)を加えて組織を可溶化し、得られた溶液を4℃で2時間遠心して破片をペレット化する。清澄化したライセートをHLA特異的IPに使用する。免疫沈降法はこれまでに述べられているようにして抗体W6/32を使用して行う<sup>127</sup>。ライセートを抗体ビーズに加え、4℃で一晩回転し、免疫沈降を行う。免疫沈降後、ビーズをライセートから除去する。IPビーズを洗浄して非特異的結合を除去し、HLA/ペプチド複合体を2N酢酸でビーズから溶出させる。タンパク質成分を分子量スピンカラムを使用してペプチドから除去する。得られたペプチドをSpeed Vac蒸発により乾燥状態とし、MS分析を行うまで-20℃で保存する。

【0546】

##### XVI. A. 3 ペプチドシーケンシング

乾燥させたペプチドをHPLCバッファーA中で戻し、C-18マイクロキャピラリーHPLCカラムにロードして質量分析計に勾配溶出させる。180分の0～40%B(溶媒A:0.1%ギ酸、溶媒B:80%アセトニトリル中0.1%ギ酸)の勾配を用いてペプチドをFusion Lumos質量分析計中に溶出させる。ペプチドの質量/電荷(m/z)のMS1スペクトルを分解能120,000でOrbitrap検出器中に収集し、続いて選択したイオンのHCD断片化後に20回のMS2の低分解能スキャンをOrbitrapまたはイオントラップ検出器中に収集する。MS2イオンの選択は、データ依存性取得モード及びイオンのMS2選択後に30秒間の動的排除を用いて行う。MS1スキャンの自動利得制御(AGC)を $4 \times 10^5$ に設定し、MS2スキャンについては $1 \times 10^4$ に設定する。HLAペプチドのシーケンシングには、MS2断片化を行うために+1、+2及び+3の荷電状態を選択することができる。

【0547】

各分析からのMS2スペクトルをComet<sup>128, 129</sup>を用いてタンパク質データベースに対して検索し、ペプチド特定をPercolator<sup>130～132</sup>を使用してスコア付けする。

【0548】

#### XVI. B. 機械学習

##### XVI. B. 1. データのコード化

各試料について、訓練データポイントは、試料中で発現された正確に1つの遺伝子にマッピングされる参照プロテオームからのすべての8～11マー(両端の値を含む)のペプ

10

20

30

40

50

チドとする。全体の訓練データセットは、各訓練試料からの訓練データセットを連結することにより生成する。8～11個の範囲はすべてのHLAクラスI提示ペプチドの約95%を捕捉するため、この長さの範囲を選択するが、12～15個の長さを加えることも、計算負荷がある程度増大する代わりに同じ方法を用いて行うことが可能である。ペプチド及びランキング配列を、ワンホットコード化スキームを用いてベクトル化する。複数の長さ(8～11個)のペプチドを、アミノ酸のアルファベットをパッド文字で増やし、すべてのペプチドを最大長さ11までパディングすることにより固定長さのベクターとして表す。訓練ペプチドのソースタンパク質のRNA存在量を、RSEM<sup>133</sup>から得られたイソフォームレベルのTPM(transcripts per million)推定値の対数として表す。各ペプチドについて、ペプチドごとのTPMを、ペプチドを含むイソフォームのそれぞれについてイソフォームごとTPM推定値の総和として計算する。OTPMで発現された遺伝子からのペプチドは訓練データから除外し、試験時に、非発現遺伝子からのペプチドに提示確率0を割り当てる。最後に、各ペプチドに、Ensemblタンパク質ファミリーIDを割り当て、各固有のEnsemblタンパク質ファミリーIDは遺伝子ごと提示傾向切片に対応する(次のセクションを参照)。

【0549】

#### XV.B.2. モデルアーキテクチャの諸元

完全提示モデルは以下の関数形を有する。

$$(式1) Pr(\text{ペプチド}i\text{提示}) = \sum_{k=1}^m a_k^i \cdot Pr(\text{アレル}a\text{によるペプチド}i\text{提示}),$$

式中、kは、1～mまでのデータセット内のHLAアレルの添え字であり、 $a_k^i$ は、標識変数(indicator variable)であり、アレルkがペプチドiが由来する試料中に存在する場合にその値は1であり、そうでない場合には0である。特定のペプチドiについて、 $a_k^i$ のうちの最大6個(ペプチドiの由来する試料のHLA型に対応する6個)以外のすべては、0である。確率の総和は、例えば $1 - \epsilon$ において、 $\epsilon = 10^{-6}$ でクリップする。

【0550】

アレルごとの提示確率を下記のようにモデル化する。

$$Pr(\text{アレル}a\text{によるペプチド}i\text{提示}) = \text{sigmoid}\{NN_a(\text{ペプチド}_i) + NN_{\text{ランキング}}(\text{ランキング}_i) + NN_{\text{RNA}}(\log(\text{TPM}_i)) + \alpha_{\text{試料}(i)} + \beta_{\text{タンパク質}(i)}\}$$

式中、変数は以下の意味を有する。sigmoidは、シグモイド( $\expit$ としても知られる)関数であり、ペプチド $_i$ は、ペプチドiのワンホットでコード化された中間パディングされたアミノ酸配列であり、 $NN_a$ は、提示の確率に対するペプチド配列の寄与をモデル化する線形最終層活性化によるニューラルネットワークであり、ランキング $_i$ は、ソースタンパク質中のペプチドiのワンホットでコード化されたランキング配列であり、 $NN_{\text{ランキング}}$ は、提示の確率に対するランキング配列の寄与をモデル化する線形最終層活性化によるニューラルネットワークであり、 $TPM_i$ は、TPM単位内のペプチドiのmRNAのソースmRNAの発現であり、試料(i)は、ペプチドiの由来する試料(すなわち患者)であり、 $\alpha_{\text{試料}(i)}$ は、試料ごと切片であり、タンパク質(i)は、ペプチドiのソースタンパク質であり、 $\beta_{\text{タンパク質}(i)}$ は、タンパク質ごと切片である(提示の遺伝子ごと傾向としても知られる)。

【0551】

各モデルのコンポーネントニューラルネットワークは以下のアーキテクチャを有する。

・ $NN_a$ のそれぞれは、入力次元数231(11残基×21個の可能な文字(パッド文字を含む)/残基)、幅256、隠れ層の整流線形単位(ReLU)活性化、出力層の線形活性化、及び訓練データセットのHLAアレルaごとに1個の出力ノードを有する単一隠れ層の多層パーセプトロン(MLP)の1個の出力ノードである。

・ $NN_{\text{ランキング}}$ は、入力次元数210(N末端ランキング配列の5残基+C末端フ

ランキング配列の5残基×残基ごとに21個の可能な文字（パッド文字を含む）、幅32、隠れ層の整流線形単位（ReLU）活性化、及び出力層の線形活性化を有する単一隠れ層のMLPである。

・NNRNAは、入力次元数1、幅16、隠れ層の整流線形単位（ReLU）活性化、及び出力層の線形活性化を有する単一隠れ層のMLPである。

【0552】

モデルの一部のコンポーネント（例えばNN<sub>a</sub>）は、特定のHLAに依存するが、多くのコンポーネント（NNフランキング、NNRNA、α試料（i）、βタンパク質（i））は依存しない。前者は「アレル相互作用性」と呼ばれ、後者は「アレル非相互作用性」と呼ばれる。アレル相互作用性及びアレル非相互作用性としてモデル化するための特性は、従来の生物学的知見に基づいて選択される。すなわち、HLAアレルはペプチドを見るため、ペプチド配列はアレル相互作用性としてモデル化されるが、ソースタンパク質、RNA発現またはフランキング配列に関する情報はHLAアレルに伝達されず（ペプチドは小胞体内でHLAと出会うまでにそのソースタンパク質から分離されているため）、したがってこれらの特性はアレル非相互作用性としてモデル化される。このモデルを、Keras v2.0.4<sup>134</sup>及びTheano v0.9.0<sup>135</sup>で実施する。

【0553】

ペプチドMSモデルは、完全MSモデル（式1）と同じ逆畳み込み手順を用いるが、提示のアレルごとの確率は、ペプチド配列及びHLAアレルのみを考慮した縮小されたアレルごとのモデルを用いて生成する。

$$\text{Pr}(\text{アレル}a \text{によるペプチド}i \text{提示}) = \text{sigmoid}\{\text{NN}_a(\text{ペプチド}i)\}$$

【0554】

ペプチドMSモデルは結合親和性予測と同じ特性を用いるが、モデルの重みは異なるデータタイプ（すなわち、HLAペプチド結合親和性データに対する質量分析データ）で訓練される。したがって、ペプチドMSモデルの予測性能を完全MSモデルと比較することで、全体の予測性能に対する非ペプチド特性（すなわち、RNA存在量、フランキング配列、遺伝子ID）の寄与が明らかとなり、ペプチドMSモデルの予測性能を結合親和性モデルと比較することで、全体の予測性能に対するペプチド配列の改善されたモデル化の重要度が明らかとなるであろう。

【0555】

#### XV B. 3. 訓練／検証／試験の分割

以下の手順を用いて訓練／検証／試験セットのうちの2つ以上に現れるペプチドがないようにする。すなわち、最初に、2つ以上のタンパク質に現れるすべてのペプチドを参照プロテオームから除去し、次に、プロテオームを10個の隣接したペプチドに分配する。各ブロックを訓練、検証、及び試験セットに固有に割り当てる。これにより、訓練、検証、または試験データセットのうちの2つ以上のデータセットに現れるペプチドはなくなる。検証セットは、早期停止のみに用いる。単一アレル試料からのペプチドが訓練データには含まれるが、訓練及び検証セットに組み込まれたペプチドのセット（提示及び非提示の両方）は、試験データとして使用されるペプチドのセットとは互いに素となる。

【0556】

#### XV B. 4. モデルの訓練

モデルを訓練するため、ペプチドごと損失が負のベルヌーイ対数尤度損失関数である（対数損失としても知られる）ものとして、すべてのペプチドを独立したものとしてモデル化する。正式には、ペプチドiの全体の損失に対する寄与は、

$$\text{損失}(i) = -\log(\text{Bernoulli}(y_i \mid \text{Pr}(\text{ペプチド}i \text{提示})))$$

であり、

式中、 $y_i$ は、ペプチドiの標示である（すなわち、ペプチドiが0である場合、 $y_i = 1$ 、そうでない場合には0）、 $\text{Bernoulli}(y \mid p)$ は、独立同分布に従う2値観測ベクトルyとして、パラメータ $p \in [0, 1]$ のベルヌーイ尤度を示す。このモデルを損失関数を最小化することにより訓練する。

## 【0557】

訓練時間を短縮するため、陰性標識訓練データの90%をアトランダムに除去することによりクラスバランスを調整する。モデル重みを、Glorot均一手順61を用いて初期化し、標準的なパラメータを用いたADAM62確率論的最適化アルゴリズムを用いてNvidia Maxwell TITAN X GPUで訓練する。全データの10%からなる検証セットを早期停止に用いる。このモデルを1/4エポックごとに検証セットで評価し、モデル訓練を検証損失（すなわち、検証セットでの負のベルヌーイ対数尤度）が低下しなくなった最初の1/4エポック後に終了する。

## 【0558】

完全提示モデルは、アンサンブル内のすべてのモデルについてモデル重さの異なるランダム初期化を行った同じ訓練データのシャッフルしたコピーで各レプリケートを独立して訓練した10個のモデルレプリケートのアンサンブルとする。試験時に、モデルレプリケートによって出力された確率の平均を取ることによって予測を生成する。

## 【0559】

XV I. B. 5. モチーフロゴ

モチーフロゴを、weblogolib Python API v3.5.0<sup>138</sup>を用いて生成する。結合親和性ロゴを生成するため、mhc\_ligand\_full.csv ファイルをImmune Epitope Database (IEDB<sup>88</sup>)からダウンロードし、以下の基準を満たすペプチドを保持する。すなわち、ナノモル(nM)単位の測定値、参照日が2000年以降であり、オブジェクトタイプが「直鎖状ペプチド」に等しく、ペプチド内のすべての残基が標準的な20文字のアミノ酸アルファベットから引用されるもの。測定された結合親和性が従来の結合閾値である500 nMよりも低いフィルタリングされたペプチドのサブセットを使用してロゴを生成する。IEDBの結合物質の数が少なすぎるアレルペアについてはロゴは生成されない。学習された提示モデルを表すロゴを生成するため、2,000,000種のランダムなペプチドについてモデル予測を各アレル及び各ペプチド長で予測する。各アレル及び各長さについて、学習された提示モデルにより上位1%（すなわち、上位の20,000種）にランク付けされたペプチドを用いてロゴを生成する。重要な点として、IEDBからのこの結合親和性データは、モデルの訓練または試験では使用せず、学習されたモチーフの比較にのみ使用する。

## 【0560】

XV I. B. 6. 結合親和性の予測

本発明者らは、オープンソースのGPU互換性HLAクラスI結合親和性予測ツールであるNetMHC II 2.3から結合親和性のみの予測ツールを使用してペプチド-MHCの結合親和性を予測する。複数のHLAアレルにわたって単一のペプチドについての結合親和性予測を組み合わせるため、最小の結合親和性を選択する。複数のペプチドにわたって結合親和性を組み合わせるため(すなわち、複数の変異ペプチドがまたがった変異をランク付けするため)、ペプチドにわたった最小の結合親和性を選択する。T細胞データセットに対するRNA発現の閾値を決定するため、TPM>0の閾値までのTCGAからの腫瘍型が一致したRNA-seqデータを使用する。初期のT細胞データセットのすべてを初期の公表中のTPM>0でフィルタリングするため、TPM>0でフィルタリングするためのTCGAのRNA-seqデータは使用されない。

## 【0561】

XV I. B. 7. 提示の予測

複数のHLAアレルにわたった単一のペプチドについて提示の確率を加え合わせるため、確率の総和を式1に示されるように特定する。複数のペプチドにわたって提示の確率を加え合わせるため(すなわち、複数の変異ペプチドがまたがった変異をランク付けするため)、提示の確率の総和を特定する。確率的には、ペプチドの提示がi.i.d.のベルヌーイランダム変数とみなされる場合、確率の総和は、提示される変異ペプチドの予想数に対応する。すなわち、

10

20

30

40

50

$$E[\text{変異}i\text{にまたがる提示される新生抗原の数}] = \sum_{j=1}^{n_i} \text{Pr}[\text{エピトープ}j\text{提示}]$$

ただし、 $\text{Pr}[\text{エピトープ}j\text{提示}]$ は、訓練された提示モデルをエピトープ  $j$  に適用することによって得られ、 $n_i$  は、変異  $i$  にまたがる変異エピトープの数を示す。例えば、そのソース遺伝子の末端から遠い  $\text{SNV}_i$  について、8 個のまたがった 8 マー、9 個のまたがった 9 マー、10 個のまたがった 10 マー、及び 11 個のまたがった 11 マーが、全部で  $n_i = 28$  のまたがった変異エピトープについて存在する。

【0562】

## XVI. C. 次世代シーケンシング

### XVI. C. 1. 試料

凍結切除された腫瘍のトランスクリプトーム分析を行うため、MS 分析に使用したのと同じ組織試料（腫瘍または隣接する正常組織）から RNA を得る。抗 PD1 療法を行っている患者の新生抗原エクソーム及びトランスクリプトーム分析を行うため、DNA 及び RNA をアーカイブされた FFPE 腫瘍生検から得る。隣接正常組織、一致血液、または PBMC を用いて正常エクソーム及び HLA タイピングを行うための正常 DNA を得る。

【0563】

### XVI. C. 2. 核酸抽出及びライブラリーの構築

血液由来の正常／生殖系 DNA を、Qiagen DNeasy カラムを製造者の推奨する手順に従って使用して単離する。組織試料からの DNA 及び RNA を Qiagen Allprep DNA/RNA 単離キットを製造者の推奨する手順に従って使用して単離する。これらの DNA 及び RNA を Picogreen 及び Ribogreen Fluorescence (Molecular Probes) によりそれぞれ定量し、収量が 50 ng 超の試料をライブラリーの構築に進める。DNA シーケンシングライブラリーを超音波せん断に続き、DNA Ultra II ライブラリー調製キットを製造者の推奨するプロトコールに従って使用することにより作製する。RNA Ultra II による熱断片化及びライブラリー構築により腫瘍 RNA シーケンシングライブラリーを作製する。得られたライブラリーを Picogreen (Molecular Probes) により定量する。

【0564】

### XVI. C. 3. 全エクソーム捕捉

DNA 及び RNA シーケンシングライブラリーのエクソン濃縮を、xGEN Whole Exome Panel を使用して行う。1~1.5 µg の正常 DNA または腫瘍 DNA もしくは RNA 由来ライブラリーを入力として用い、12 時間よりも長い時間にわたってハイブリダイズさせた後、ストレプトアビジン精製を行う。捕捉されたライブラリーを PCR により最小限増幅し、NEBNext Library Quant Kit により定量する。捕捉された各ライブラリーを等モル濃度でプールし、c-bot を用いてクラスター化し、75 塩基対の端部において HiSeq 4000 で、500x 超の腫瘍エクソーム、100x 超の正常エクソーム、及び 100M 超のリードの腫瘍トランスクリプトームのターゲットユニークな平均カバレッジにまでシーケンシングする。

【0565】

### XVI. C. 4. 分析

エクソームリード（FFPE 腫瘍エクソーム及び一致させた正常エクソーム）を、BWA-MEM<sup>144</sup> (v. 0.7.13-r1126) を使用して参照ヒトゲノム (hg38) とアラインする。RNA-seq リード（FFPE 及び凍結腫瘍組織試料）を、STAR (v. 2.5.1b) を使用してゲノム及び GENCODE 転写産物 (v. 25) とアラインする。RNA 発現を、同じ参照転写産物で RSEM<sup>133</sup> (v. 1.2.31) を使用して定量する。Picard (v. 2.7.1) を使用してデュプリケートアラインメントをマークし、アラインメントメトリックを計算する。GATK145 (v. 3.5

10

20

30

40

50

ー0)による塩基のクオリティー補正 (base quality score recalibration) 後のFFPE腫瘍試料について、FreeBayes<sup>146</sup> (1.0.2)によりペアリングした腫瘍ー正常エクソームを用いて置換及び短いindel変異を検出する。フィルターには、アレル頻度>4%;塩基クオリティーの中央値>25、支持リードの最小マッピングクオリティー=30、及び十分なカバレッジが得られたとして正常の代替的リードカウント≤2が含まれる。変異体はまた、両方の鎖でも検出されるであろう。反復領域に生じる体細胞変異体は除外する。翻訳及びアノテーションは、RefSeq転写産物を用いてsnpeff<sup>147</sup> (v.4.2)により行う。腫瘍RNAアライメントで確認された非同義、ノンストップ変異体を新生抗原予測に進める。Optitype<sup>148</sup> 1.3.1を使用してHLAタイプを生成する。

10

【0566】

#### XV I. C. 5. IVSコントロール実験用の腫瘍細胞株及び一致した正常細胞株

腫瘍細胞株及びその正常ドナーを一致させたコントロール細胞株をすべて購入し、販売元の指示に従って10<sup>8</sup>3~10<sup>8</sup>4細胞にまで増殖させた後、核酸抽出及びシーケンシングを行うために瞬間凍結させる。NGSを、Mutect<sup>149</sup> (3.1-0)を置換変異の検出にのみ用いた点以外は、概ね上記に述べたのと同様にして行う。

【0567】

#### XV I I. 実施例12: NSCLC患者の末梢血由来の新生抗原特異的メモリーT細胞のTCRの予期されるシーケンシング

次に、新生抗原特異的メモリーT細胞のTCRをNSCLC患者の末梢血からシーケンシングする。NSCLC患者からの末梢血単核球(PBMC)をELISpotインキュベーション後に回収する。具体的には、インビトロ増殖させた患者からのPBMCを、IFNγELISpotで、患者特異的な個別の新生抗原ペプチドにより、患者特異的な新生抗原ペプチドプールにより、また、DMSOネガティブコントロールにより刺激する。インキュベーション後、検出抗体の添加に先立って、PBMCを新しい培養プレートに移し、ELISpotアッセイが完了する間、インキュベーター内で維持する。陽性(応答性)ウェルをELISpotの結果に基づいて特定する。陽性ウェル及びネガティブコントロール(DMSO)ウェルからの細胞を加え合わせて、磁気標識抗体でCD137について染色し、Miltenyi磁気単離カラムを使用して濃縮を行う。

20

【0568】

上記に述べたように単離して増殖させたCD137濃縮及び枯渇T細胞画分を10x Genomicsシングルセル・レゾリューション・ペアド免疫TCRプロファイリングアプローチを用いてシーケンシングする。詳細には、生T細胞を、続くシングルセルcDNA作製及び完全長TCRのプロファイリング(5'UTR~定常領域、αとβを対合させる)を行うためにシングルセルエマルジョン中に分配する。1つのアプローチでは、転写産物の5'末端の分子バーコード付けしたテンプレート交換オリゴを利用し、第2のアプローチでは、3'末端の分子バーコード付けした定常領域オリゴを利用し、第3のアプローチでは、RNAポリメラーゼプロモーターをTCRの5'末端または3'末端に連結する。これらのアプローチはいずれも、シングルセルレベルでのα及びβTCRペアの特定及び逆量み込みを可能とするものである。得られたバーコード付けしたcDNA転写産物に最適化した酵素及びライブラリー構築ワークフローを行ってバイアスを低減し、細胞のプール内のクロノタイプの正確な表現を確実とする。各ライブラリーをIllumina社のMiSeqまたはHiSeq4000装置(対合末端150サイクル)で細胞当たり約5千~5万リードのターゲットシーケンシング深さに対してシーケンシングする。TCRα及びTCRβ鎖の存在を、オーソゴナルなアンカーPCRベースTCRシーケンシングアプローチ(Archerr)によって確認する。この特定のアプローチは、10x GenomicsベースのTCRシーケンシングと比較して限定された数の細胞を入力として使用し、用いられる酵素的操作の数が少ないという利点を有する。

30

40

【0569】

シーケンシングの出力を10xソフトウェア及びカスタムバイオインフォマティクス

50

パイプラインを使用して分析してT細胞受容体（TCR）α及びβ鎖を特定する。クロナタイプは、固有のCDR3アミノ酸配列のα、β鎖のペアとして定義する。クロナタイプを、2細胞よりも高い頻度で存在する1個のα鎖と1個のβ鎖のペアについてフィルタリングして患者のターゲットペプチド当たりのクロナタイプA最終リストを得た。

【0570】

要約すると、上記に述べた方法を用いることで、セクションXVの実施例11に関して上記に述べたようにして特定された患者の腫瘍新生抗原に対して新生抗原特異的である、患者の末梢血由来のメモリーCD4+T細胞が特定される。これらの特定された新生抗原特異的T細胞のTCRをシーケンシングする。またさらに、上記の提示モデルにより特定される患者の腫瘍新生抗原に対して新生抗原特異的であるシーケンシングしたTCRが特定される。

【0571】

#### XVII. 実施例13：T細胞療法の新生抗原特異的メモリーT細胞の使用

患者の腫瘍によって提示される新生抗原に対して新生抗原特異的であるT細胞及び／またはTCRが特定された後、これらの特定された新生抗原特異的T細胞及び／またはTCRを患者のT細胞療法に使用することができる。詳細には、これらの特定された新生抗原特異的T細胞及び／またはTCRを使用してT細胞療法において患者に注入するための治療量の新生抗原特異的T細胞を作製することができる。患者のT細胞療法で使用するための治療量の新生抗原特異的T細胞を作製するための2つの方法を、本明細書のセクションXVII.A.及びXVII.B.で考察する。第1の方法は、患者試料から特定された新生抗原特異的T細胞を増殖させることを含む（セクションXVII.A.）。第2の方法は、特定された新生抗原特異的T細胞をシーケンシングすることと、シーケンシングされたTCRを新たなT細胞にクローニングすることを含む（セクションXVII.B.）。本明細書では明示的に触れていないT細胞療法で使用するための新生抗原特異的T細胞を作製するための代替的な方法を用いて、T細胞療法で使用するための治療量の新生抗原特異的T細胞を作製することもできる。これらの方法の1つ以上によって新生抗原特異的T細胞が得られた後、これらの新生抗原特異的T細胞をT細胞療法を行うために患者に注入することができる。

【0572】

#### XVII.A. T細胞療法用の患者試料からの新生抗原特異的メモリーT細胞の特定及び増殖

患者のT細胞療法に使用するための治療量の新生抗原特異的T細胞を作製するための第1の方法は、患者試料から特定された新生抗原特異的T細胞を増殖させることを含む。

【0573】

詳細には、新生抗原特異的T細胞を患者のT細胞療法に使用するための治療量にまで増殖させるため、患者のがん細胞によって提示される可能性が最も高い新生抗原ペプチドのセットを、上記に述べた提示モデルを使用して特定する。さらに、T細胞を含有する患者試料を患者から得る。患者試料は、患者の末梢血、腫瘍浸潤リンパ球（TIL）、またはリンパ節細胞を含み得る。

【0574】

患者試料が患者の末梢血を含む実施形態では、以下の方法を用いて新生抗原特異的T細胞を治療量にまで増殖させることができる。一実施形態では、プライミングを行うことができる。別の実施形態では、既に活性化されているT細胞を上記に述べた方法のうちの1つ以上を用いて特定することができる。別の実施形態では、プライミング及び既に活性化されているT細胞の特定の両方を行うことができる。プライミング及び既に活性化されているT細胞の特定の両方を行うことの利点は、表現される特異性の数が最大となることである。プライミング及び既に活性化されているT細胞の特定の両方を行うことの難点は、このアプローチは困難で時間がかかることである。別の実施形態では、必ずしも活性化されていない新生抗原特異的細胞を単離することができる。かかる実施形態では、これらの新生抗原特異的細胞の抗原特異的または非特異的増殖を行うこともできる。これらのプラ

イミグされたT細胞を回収した後、プライミングされたT細胞を迅速増殖プロトコールに供することができる。例えば、いくつかの実施形態では、プライミングされたT細胞をRosenberg迅速増殖プロトコール (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2978753/>、<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2305721/>)<sup>153, 154</sup>に供することができる。

#### 【0575】

患者試料が患者のTILを含む実施形態では、以下の方法を用いて新生抗原特異的T細胞を治療量にまで増殖させることができる。一実施形態では、新生抗原特異的TILは、エキスピボで選別されたテトラマー／マルチマーとし、次いで選別されたTILを上記に述べたような迅速増殖プロトコールに供することができる。別の実施形態では、TILの新生抗原非特異的増殖を行った後、新生抗原特異的TILをテトラマー選別し、その後、選別されたTILを上記に述べたような迅速増殖プロトコールに供することができる。別の実施形態では、TILを迅速増殖プロトコールに供する前に抗原特異的培養を行うことができる (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4607110/>, <https://onlinelibrary.wiley.com/doi/pdf/10.1002/eji.201545849>)<sup>155, 156</sup>。

#### 【0576】

いくつかの実施形態では、Rosenberg迅速増殖プロトコールを改変することができる。例えば、抗PD1及び／または抗41BBをTIL培養に加えてより迅速な増殖を刺激することができる (<https://jitr.biomedcentral.com/articles/10.1186/s40425-016-0164-7>)<sup>157</sup>。

#### 【0577】

XVIIII. B. 新生抗原特異的T細胞の特定、特定された新生抗原特異的T細胞のシークエンシング、及びシークエンシングされたTCRの新たなT細胞へのクローニング

患者にT細胞療法を使用するための治療量の新生抗原特異的T細胞を作製するための第2の方法は、患者試料から新生抗原特異的T細胞を特定することと、特定された新生抗原特異的T細胞のTCRをシークエンシングすることと、シークエンシングされたTCRを新たなT細胞をクローニングすることと、を含む。

#### 【0578】

最初に、新生抗原特異的T細胞を患者試料から特定し、特定された新生抗原特異的T細胞のTCRをシークエンシングする。T細胞を単離することができる患者試料は、血液、リンパ節、または腫瘍のうちの1つ以上を含み得る。より詳細には、T細胞を単離することができる患者試料は、末梢血単核細胞(PBMC)、腫瘍浸潤細胞(TIL)、解離腫瘍細胞(DTC)、インビトロでプライミングされたT細胞、及び／またはリンパ球から単離された細胞のうちの1つ以上を含み得る。これらの細胞は、新鮮なもの及び／または凍結されたものであってよい。PBMC及びインビトロでプライミングされたT細胞は、がん患者及び／または健康な対象から得ることができる。

#### 【0579】

患者試料が得られた後、試料を増殖させ、かつ／またはプライミングすることができる。さまざまな方法を実施して患者試料を増殖させ、プライミングすることができる。一実施形態では、新鮮な及び／または凍結したPBMCをペプチドまたはタンデムミニ遺伝子の存在下で刺激することができる。別の実施形態では、新鮮な及び／または凍結したPBMCをペプチドまたはタンデムミニ遺伝子の存在下で抗原提示細胞(APC)により刺激及びプライミングすることができる。APCの例としては、B細胞、単球、樹状細胞、マクロファージまたは人工抗原提示細胞(関連するHLA及び／または共刺激分子を提示する細胞またはビーズなど、<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2929753/>に概説されている)。別の実施形態

では、P B M C、T I L、及び／または単離されたT細胞をサイトカインの存在下で刺激することができる（例えば、I L - 2、I L - 7、及び／またはI L - 1 5）。別の実施形態では、T I L及び／または単離されたT細胞を、最大刺激、サイトカイン（複数可）、及び／またはフィーダー細胞の存在下で刺激することができる。そのような実施形態では、T細胞は、活性化マーカー及び／またはマルチマー（例えば、テトラマー）によって単離することができる。別の実施形態では、T I L及び／または単離されたT細胞を、刺激性及び／または共刺激性マーカー（例えば、C D 3 抗体、C D 2 8 抗体）及び／またはビーズ（例えば、D y n a B e a d s）で刺激することができる。別の実施形態では、D T Cを、リッチ培地中、高用量のI L - 2でフィーダー細胞上で急速増殖プロトコルを用いて増殖させることができる。

10

#### 【0580】

次に、新生抗原特異的T細胞を特定して単離する。いくつかの実施形態では、T細胞は、予め増殖させることなく患者試料からエキスピボで単離される。一実施形態では、セクションX V I Iに関連して上記に述べた方法を用いて患者試料から新生抗原特異的T細胞を特定することができる。代替的な実施形態では、単離は、ポジティブセレクションによって特定の細胞集団について濃縮するか、またはネガティブセレクションにより特定の細胞集団を枯渇させることによって行われる。いくつかの実施形態では、ポジティブまたはネガティブセレクションは、それぞれポジティブまたはネガティブに選択された細胞上で発現されるか（マーカー<sup>+</sup>）または比較的高いレベルで発現される（マーカー<sup>高</sup>）1つ以上の表面マーカーに特異的に結合する1つ以上の抗体または他の結合物質と細胞を培養することによって行われる。

20

#### 【0581】

いくつかの実施形態では、T細胞は、C D 1 4などの、B細胞、単球、または他の白血球細胞などの非T細胞上で発現されるマーカーのネガティブセレクションによってP B M C試料から分離される。いくつかの態様では、C D 4 +またはC D 8 +の選択ステップを用いてC D 4 +ヘルパー及びC D 8 +細胞傷害性T細胞を分離する。そのようなC D 4 +及びC D 8 +の集団は、1つ以上のナイーブ、メモリー、及び／またはエフェクターT細胞の亜集団上で発現される、または比較的高い程度で発現されるマーカーについてポジティブまたはネガティブセレクションによって亜集団にさらに分別することができる。

#### 【0582】

いくつかの実施形態では、C D 4 +及びC D 8 +細胞は、それぞれの亜集団に関連付けられた表面抗原に基づいたポジティブまたはネガティブセレクションなどによって、ナイーブ、中枢メモリー、エフェクターメモリー、及び／または中枢メモリー幹細胞についてさらに濃縮されるか、または枯渇される。いくつかの実施形態では、中枢メモリーT細胞（T C M）細胞の濃縮を行うことで、例えば投与後の長期の生存、増殖、及び／または生着を改善させるなど、有効性が高められ、これはいくつかの態様で、かかる亜集団において特に安定的である（T e r a k u r a e t a l . ( 2 0 1 2 ) B l o o d . 1 : 7 2 - 8 2 ; W a n g e t a l . ( 2 0 1 2 ) J I m m u n o t h e r . 3 5 ( 9 ) : 6 8 9 - 7 0 1 を参照）。いくつかの実施形態では、T C Mを濃縮したC D 8 +T細胞とC D 4 +T細胞を組み合わせることでさらに有効性が高められる。

30

40

#### 【0583】

複数の実施形態では、メモリーT細胞は、C D 8 +末梢血リンパ球のC D 6 2 L +及びC D 6 2 L - サブセットの両方に存在する。P B M Cは、例えば抗C D 8 及び抗C D 6 2 L 抗体を用いることなどにより、C D 6 2 L - C D 8 +及び／またはC D 6 2 L + C D 8 +画分を濃縮または枯渇させることができる。

#### 【0584】

いくつかの実施形態では、中枢メモリーT（T C M）細胞の濃縮は、C D 4 5 R O、C D 6 2 L、C C R 7、C D 2 8、C D 3、及び／またはC D 1 2 7 の陽性または高い表面発現に基づいたものであり、いくつかの態様では、C D 4 5 R A 及び／またはグランザイムBを発現するか、または高度に発現する細胞のネガティブセレクションに基づいたもの

50

である。いくつかの態様では、TCM細胞について濃縮されたCD8+集団の単離は、CD4、CD14、CD45RAを発現する細胞の枯渇、及びCD62Lを発現する細胞のポジティブセレクションまたは濃縮によって行われる。一態様では、中枢メモリーT（TCM）細胞の濃縮は、CD4発現に基づいて選択された陰性画分から出発し、これにCD14及びCD45RAの発現に基づいてネガティブセレクションを、また、CD62Lに基づいてポジティブセレクションを行うことによって行われる。これらの選択はいくつかの態様では同時に行われ、他の態様では、いずれかの順序で順次行われる。いくつかの態様では、CD8+細胞集団または亜集団を調製するのに用いたのと同じCD4発現に基づく選択ステップをさらに用いて、CD4+細胞集団または亜集団を生成することにより、CD4に基づいた選択からの陽性及び陰性画分が保持され、必要に応じて1つ以上のさらなるポジティブまたはネガティブセレクションステップを行った後、方法の次のステップで用いられる。

#### 【0585】

特定の例では、PBMCの試料または他の白血球試料にCD4+細胞の選択を行う（陰性及び陽性画分の両方が保持される）。次に、陰性画分に、CD14及びCD45RAまたはROR1の発現に基づいたネガティブセレクションを、また、CD62LまたはCCR7などの中枢メモリーT細胞のマーカー特性に基づいたポジティブセレクションを行う（ポジティブセレクション及びネガティブセレクションはいずれかの順序で行う）。

#### 【0586】

CD4+Tヘルパー細胞は、細胞表面抗原を有する細胞集団を特定することにより、ナイーブ、中枢メモリー、及びエフェクター細胞に分別される。CD4+リンパ球は標準的方法によって得ることができる。いくつかの実施形態では、ナイーブCD4+Tリンパ球は、CD45RO-、CD45RA+、CD62L+、CD4+T細胞である。いくつかの実施形態では、中枢メモリーCD4+細胞は、CD62L+かつCD45RO+である。いくつかの実施形態では、エフェクターCD4+細胞は、CD62L-かつCD45RO-である。

#### 【0587】

一実施形態では、ネガティブセレクションによってCD4+細胞について濃縮するため、モノクローナル抗体カクテルは通常、CD14、CD20、CD11b、CD16、HLA-DR、及びCD8に対する抗体を含む。いくつかの実施形態では、抗体または結合パートナーは、磁性ビーズまたは常磁性ビーズのような固体支持体またはマトリックスに結合されることでポジティブ及び／またはネガティブセレクション用の細胞の分離を可能とする。例えば、いくつかの実施形態では、細胞及び細胞集団は免疫磁気（または親和性磁気）分離法を用いて分離または単離される（Methods in Molecular Medicine, vol. 58: Metastasis Research Protocols, Vol. 2: Cell Behavior In Vitro and In Vivo, p 17-25 Edited by: S. A. Brooks and U. Schumacher Humana Press Inc., Totowa, N. J. に概説されている）。

#### 【0588】

いくつかの態様では、分離する細胞の試料または組成を、常磁性ビーズ（例えば、DynabeadsまたはMACSビーズなど）などの磁気応答性粒子または微粒子などの微小な磁化可能または磁気応答性材料とインキュベートする。磁気応答性材料、例えば粒子は、材料が分離することが望ましい、例えば材料がネガティブまたはポジティブに選択することが望ましい細胞、複数の細胞、または細胞の集団に存在する分子、例えば表面マーカーに特異的に結合する結合パートナー、例えば抗体に直接的または間接的に付着される。

#### 【0589】

いくつかの実施形態では、磁性粒子またはビーズは、抗体または他の結合パートナーのような特定の結合要素に結合された磁気応答性材料を含む。磁気分離法に使用される多くの周知の磁気応答性材料が存在する。適当な磁性粒子としては、本明細書に参照により援

10

20

30

40

50

用するところのMoldayによる米国特許第4,452,773号、及び欧州特許明細書EP 452342 B号に記載されるものがある。他の例としては、Owenによる米国特許第4,795,698号、及びLibertiらによる米国特許第5,200,084号に記載されるものなどのコロイド粒径の粒子がある。

#### 【0590】

インキュベーションは、磁性粒子またはビーズに付着された、抗体または結合パートナーと特異的に結合する抗体もしくは結合パートナー、または分子（二次抗体または他の試薬など）が、試料中の細胞上に存在する場合に細胞表面分子に特異的に結合するような条件下で一般的に行われる。

#### 【0591】

いくつかの態様では、試料は磁場の中に置かれ、磁気応答性または磁化可能な粒子が付着された細胞が磁石に引き寄せられて非標識細胞から分離される。ポジティブセクションでは、磁石に引き寄せられた細胞が保持され、ネガティブセクションでは引き寄せられなかった細胞（非標識細胞）が保持される。いくつかの態様では、ポジティブセクションとネガティブセクションとの組み合わせが同じ選択ステップにおいて行われ、陽性及び陰性画分が保持され、さらに処理されるかまたはさらなる分離ステップに供される。

#### 【0592】

特定の実施形態では、磁気応答性粒子は、一次抗体もしくは他の結合パートナー、二次抗体、レクチン、酵素、またはストレプトアビジンでコーティングされる。特定の実施形態では、磁性粒子は1つ以上のマーカーに特異的な一次抗体のコーティングによって細胞に付着される。特定の実施形態では、ビーズではなく、細胞を一次抗体または結合パートナーで標識し、次いで、細胞型に特異的な二次抗体または他の結合パートナー（例えばストレプトアビジン）でコーティングした磁性ビーズを加える。特定の実施形態では、ストレプトアビジンでコーティングされた磁性ビーズがビオチン化された一次または二次抗体と組み合わせて使用される。

#### 【0593】

いくつかの態様では、磁気応答性粒子は、後でインキュベート、培養、及び／または操作しようとする細胞に付着されたままとされ、いくつかの態様では、粒子は患者に投与される細胞に付着されたままとされる。いくつかの実施形態では、磁化可能な、または磁気応答性粒子は細胞から除去される。磁化可能な粒子を細胞から除去するための方法は周知のものであり、例えば、非標識抗体、磁化可能な粒子、または切断可能なリンカーと結合させた抗体などの使用が含まれる。いくつかの実施形態では、磁化可能な粒子は生分解性である。

#### 【0594】

いくつかの実施形態では、親和性に基づく選択は、磁気活性化細胞選別（MACS）（Miltenyi Biotech, Auburn, Calif.）により行われる。磁気活性化細胞選別（MACS）システムは、磁化粒子が付着された細胞を高純度で選択することが可能である。いくつかの実施形態では、MACSは、外部磁場の印可後に非標的種と標的種が順次溶出されるモードで動作する。すなわち、磁化粒子に付着された細胞が保持される一方で付着していない種は溶出する。次いで、この第1の溶出ステップが完了した後、磁場に捕捉され、溶出が防止された種が、溶出して回収できるような何らかの形で遊離される。特定の実施形態では、大型T細胞以外のT細胞が標識され、細胞の不均質集団から枯渇される。

#### 【0595】

特定の実施形態では、単離または分離は、本方法の単離、細胞調製、分離、プロセッシング、インキュベーション、培養、及び／または配合ステップのうちの1つ以上を行うシステム、機器、または装置を用いて行われる。いくつかの態様では、例えば、エラー、ユーザによる取り扱い、及び／または汚染を最小限に抑制するため、これらのステップのそれぞれを閉鎖または滅菌環境内で行うシステムが用いられる。1つの例では、システムは、国際公開第WO2009/072003号、または米国特許出願公開第US201100

10

20

30

40

50

03380A1号である。

【0596】

いくつかの態様では、システムまたは装置は、単離、プロセッシング、及び配合ステップのうちの1つ以上、例えばすべてを、統合された、または自給式システムで、及び／または自動化若しくはプログラム可能な形で行う。いくつかの態様では、システムまたは装置は、ユーザがプロセッシング、単離、操作、及び配合ステップをプログラムし、制御し、その結果を評価し、及び／またはそのさまざまな側面を調節することを可能とする、システムまたは装置と通信するコンピュータ及び／またはコンピュータプログラムを含む。

【0597】

いくつかの態様では、分離及び／または他のステップは、例えば、閉鎖及び滅菌システム内で臨床規模レベルの細胞の自動化分離を行うためのCliniMACSシステム(Miltenyi Biotec)を用いて行われる。構成要素としては、統合マイクロコンピュータ、磁気分離ユニット、蠕動ポンプ、及びさまざまなピンチバルブが含まれ得る。統合コンピュータは、いくつかの態様では、機器のすべての構成要素を制御し、システムに、繰り返し手順を標準化された順序で実行するように指示する。磁気分離ユニットは、いくつかの態様では、可動式永久磁石及び選択カラム用のホルダーを含む。蠕動ポンプはピンチバルブとともに、チューブセット全体にわたった流量を制御し、システムを通るバッファの制御された流れ及び細胞の継続的な懸濁を確実にする。

【0598】

CliniMACSシステムは、いくつかの態様では、滅菌された非発熱性溶液中で供給される、抗体と結合された磁化可能な粒子を使用する。いくつかの実施形態では、磁性粒子で細胞を標識した後、細胞を洗浄して余分な粒子を除去する。次いで、細胞調製バッグをチューブセットに接続し、チューブセットをバッファの入ったバッグと細胞回収バッグに接続する。チューブセットは、予備カラム及び分離カラムを含む予め組み立てられた滅菌チューブで構成され、1回使用のみである。分離プログラムの開始後、システムは細胞試料を分離カラムに自動的に適用する。標識細胞がカラムに保持されるのに対して、非標識細胞は一連の洗浄ステップにより除去される。いくつかの実施形態では、本明細書に記載される方法とともに使用される細胞集団は非標識細胞であり、カラムに保持されない。いくつかの実施形態では、本明細書に記載される方法とともに使用される細胞集団は標識細胞であり、カラムに保持される。いくつかの実施形態では、本明細書に記載される方法とともに使用される細胞集団は、磁場の除去後にカラムから溶出され、細胞回収バッグ内に回収される。

【0599】

いくつかの実施形態では、分離及び／または他のステップは、CliniMACS Prodigyシステム(Miltenyi Biotec)を使用して行われる。CliniMACS Prodigyシステムは、いくつかの実施形態では、遠心分離による細胞の自動洗浄及び分画化を可能とする細胞プロセッシングユニットを備えている。CliniMACS Prodigyシステムは、オンボードカメラ及びソース細胞産物のマクロスコピック層を判別することによって最適な細胞分画化エンドポイントを決定する画像認識ソフトウェアも含むことができる。例えば、末梢血を、赤血球、白血球、及び血漿層に自動的に分離することができる。CliniMACS Prodigyシステムは、例えば、細胞分化及び増殖、抗原ローディング、及び長期の細胞培養などの細胞培養プロトコルを行う統合型細胞培養チャンバも含むことができる。入力ポートによって培地の無菌的な除去及び補充が可能であり、統合型顕微鏡を使用して細胞を監視することができる(例えば、Klebanoff et al. (2012) J Immunother. 35 (9): 651-660, Terakura et al. (2012) Blood. 1: 72-82, and Wang et al. (2012) J Immunother. 35 (9): 689-701を参照)。

【0600】

いくつかの実施形態では、本明細書に記載される細胞集団は、複数の細胞表面マーカー

について染色された細胞が流体流で搬送されるフローサイトメトリーによって回収及び濃縮（または枯渇）される。いくつかの実施形態では、本明細書に記載される細胞集団は、分取スケール（FACS）ソーティングにより回収及び濃縮（または枯渇）される。特定の実施形態では、本明細書に記載される細胞集団は、FACSに基づいた検出システムと組み合わせた微小電気機械的システム（MEMS）チップの使用により回収及び濃縮（または枯渇）される（例えば、WO2010/033140, Cho et al. (2010) Lab Chip 10, 1567-1573; 及び Godin et al. (2008) J Biophoton. 1 (5): 355-376を参照）。いずれの場合も、細胞を複数のマーカーで標識することができ、明確に定義された高純度のT細胞サブセットの単離を可能とする。

10

#### 【0601】

いくつかの実施形態では、ポジティブ及び／またはネガティブセクションによる分離を容易とするため、抗体または結合パートナーは1つ以上の検出可能なマーカーで標識される。例えば、分離は、蛍光標識された抗体との結合に基づいたものとして行うことができる。いくつかの例では、1つ以上の細胞表面マーカーに対して特異的な抗体または他の結合パートナーの結合に基づいた細胞の分離は、例えばフローサイトメトリー検出システムと組み合わせた、分取スケール（FACS）及び／または微小電気機械的（MEMS）チップを含む、蛍光活性化セルソーティング（FACS）などにより、流体流で搬送される。かかる方法は、同時に複数のマーカーに基づいたポジティブ及びネガティブセクションを可能とするものである。

20

#### 【0602】

いくつかの実施形態では、調製法は、単離、インキュベーション、及び／または操作の前または後で細胞を凍結、例えば凍結保存するステップを含む。いくつかの実施形態では、凍結ステップ及びそれに続く解凍ステップにより、細胞集団中の顆粒球及びある程度の単球が除去される。いくつかの実施形態では、細胞は、血漿及び血小板を除去するために、例えば洗浄ステップ後に凍結溶液中に懸濁される。いくつかの態様では、各種の公知の凍結溶液及びパラメータのいずれをも使用することができる。1つの例では、20% DMSO及び8% ヒト血清アルブミン（HSA）を含有するPBSまたは他の適当な細胞凍結媒体を使用する。次いで、これを、媒体で1:1に希釈することにより、それぞれ、DMSO及びHSAの最終濃度を10%及び4%とする。他の例としては、Cryostor（登録商標）、CTL-Cryo（商標）ABC凍結媒などが挙げられる。その後、細胞を毎分1℃の速度で-80℃にまで凍結し、液体窒素貯蔵タンクの蒸気相中で保存する。

30

#### 【0603】

いくつかの実施形態では、提供される方法には、培養（cultivation）、インキュベーション、培養（culture）、及び／または遺伝子操作のステップが含まれる。例えば、いくつかの実施形態では、枯渇させた細胞集団をインキュベート及び／または操作する方法、ならびに培養開始組成物が提供される。

#### 【0604】

したがって、いくつかの実施形態では、細胞集団は培養開始組成物中でインキュベートされる。インキュベーション及び／または操作は、ユニット、チャンバー、ウェル、カラム、チューブ、チューブセット、バルブ、バイアル、培養皿、バッグ、または他の培養または細胞培養用の容器などの培養容器中で行われる。

40

#### 【0605】

いくつかの実施形態では、細胞は、遺伝子操作に先立って、または遺伝子操作に関連してインキュベートかつ／または培養される。インキュベーションステップは、培養（culture）、培養（cultivation）、刺激、活性化、及び／または増殖を含み得る。いくつかの実施形態では、組成物または細胞は、刺激条件または刺激剤の存在下でインキュベートされる。かかる条件としては、例えば、組換え抗原受容体を導入する目的などで、抗原曝露を模倣し、及び／または遺伝子操作を行うために細胞をプライミングするために、集団内の細胞の分裂、増殖、活性化、及び／または生存を誘導するように設

50

計されたものが挙げられる。

#### 【0606】

こうした条件には、特定の培地、温度、酸素含有量、二酸化炭素含有量、時間、作用物質、例えば、栄養素、アミノ酸、抗生物質、イオン、及び／または刺激因子、例えば、サイトカイン、ケモカイン、抗原、結合パートナー、融合タンパク質、組換え可溶性受容体、ならびに細胞を活性化するように設計された他の任意の作用物質のうちの1つ以上が含まれ得る。

#### 【0607】

いくつかの実施形態では、刺激条件または刺激剤には、TCR複合体の細胞内シグナル伝達ドメインを活性化することができる1つ以上の作用物質、例えばリガンドが含まれる。いくつかの態様では、作用物質は、T細胞におけるTCR/CD3細胞内シグナル伝達カスケードをオンするか、または開始する。このような作用物質には、例えば、ビーズなどの固体支持体に結合した、例えば、抗CD3、抗CD28などの、TCR成分及び／または共刺激受容体に対して特異的な抗体などの抗体、及び／または1つ以上のサイトカインが含まれ得る。場合により、増殖方法は、抗CD3及び／または抗CD28抗体を（例えば、少なくとも約0.5 ng/mLの濃度で）培地に添加するステップをさらに含んでもよい。一部の態様において、刺激剤は、IL-2及び／またはIL-15、例えば、少なくとも約10単位/mLの濃度のIL-2が含まれる。

#### 【0608】

いくつかの態様では、インキュベーションは、Riddellらへの米国特許第6,040,177号、Klebanoff et al. (2012) J Immunother. 35 (9): 651-660、Terakura et al. (2012) Blood. 1: 72-82、及び／またはWang et al. (2012) J Immunother. 35 (9): 689-701に記載されるものなどの方法に従って行われる。

#### 【0609】

いくつかの実施形態では、T細胞は、（例えば、結果として生じる細胞集団が、増殖させようとする初期集団中で各Tリンパ球について少なくとも約5個、10個、20個、または40個、またはそれよりも多いPBMCフィーダー細胞を含有するように）、例えば、非分裂末梢血単核球（PBMC）などのフィーダー細胞を培養開始組成物に添加し、培養物を（例えば、T細胞の数を増やすのに十分な時間にわたって）インキュベートすることによって増殖される。いくつかの態様では、非分裂フィーダー細胞は、γ線を照射したPBMCフィーダー細胞を含んでもよい。いくつかの実施形態では、PBMCに約3000～3600ラドの範囲のγ線を照射して細胞分裂を阻止する。いくつかの実施形態では、PBMCフィーダー細胞をマイトマイシンCで不活化する。いくつかの態様では、T細胞の集団を加える前にフィーダー細胞を培地に加える。

#### 【0610】

いくつかの実施形態では、刺激条件は、ヒトTリンパ球の増殖に適した温度、例えば、少なくとも約25℃、一般的には少なくとも約30度、及び一般的には37℃または約37℃を含む。必要に応じて、インキュベーションは、フィーダー細胞として非分裂性のEBVで形質転換したリンパ芽球様細胞（LCL）を添加することをさらに含んでもよい。LCLに、約6000～10,000ラドの範囲のγ線を照射することができる。LCLフィーダー細胞はいくつかの態様では、LCLフィーダー細胞と初期Tリンパ球との比が少なくとも約10:1など、任意の適切な量で提供される。

#### 【0611】

各実施形態において、抗原特異的T細胞、例えば、抗原特異的CD4+細胞は、ナイーブまたは抗原特異的なTリンパ球を抗原で刺激することによって得られる。例えば、サイトメガロウイルス抗原に対する抗原特異的T細胞株またはクローンを、感染させた対象からT細胞を単離し、この細胞をインビトロで同じ抗原で刺激することによって作製することができる。

#### 【0612】

いくつかの実施形態では、新生抗原特異的T細胞は、機能性アッセイ（例えば、E L I S p o t）による刺激後に特定及び／または単離される。いくつかの実施形態では、新生抗原特異的T細胞は、細胞内サイトカイン染色によって多機能性細胞によって単離される。いくつかの実施形態では、新生抗原特異的T細胞は、活性化マーカー（例えば、C D 1 3 7、C D 3 8、C D 3 8 / H L A - D R二重陽性、及び／またはC D 6 9）を用いて特定及び／または単離される。いくつかの実施形態では、新生抗原特異的C D 4 +、ナチュラルキラーT細胞、及び／またはメモリーT細胞は、クラスIIマルチマー及び／または活性化マーカーを用いて特定及び／または単離される。いくつかの実施形態では、新生抗原特異的C D 4 + T細胞はメモリーマーカー（例えば、C D 4 5 R A、C D 4 5 R O、C C R 7、C D 2 7、及び／またはC D 6 2 L）を用いて特定及び／または単離される。いくつかの実施形態では、増殖中の細胞が特定及び／または単離される。いくつかの実施形態では、活性化されたT細胞が特定及び／または単離される。

10

#### 【0613】

患者試料からの新生抗原特異的T細胞の単離後、特定された新生抗原特異的T細胞の新生抗原特異的T C Rをシーケンシングする。新生抗原特異的T C Rをシーケンシングするには、最初にT C Rを特定する必要がある。T細胞の新生抗原特異的T C Rを特定する1つの方法は、T細胞を少なくとも1つの新生抗原を含むH L Aマルチマー（例えばテトラマー）と接触させ、H L AマルチマーとT C Rとの間の結合によりT C Rを特定することを含むことができる。新生抗原特異的T C Rを特定する別の方法は、T C Rを有する1つ以上のT細胞を得ることと、1つ以上のT細胞を少なくとも1つの抗原提示細胞（A P C）上に提示された少なくとも1つの新生抗原により活性化することと、少なくとも1つの新生抗原との相互作用により活性化された1つ以上の細胞を選択することによりT C Rを特定することと、を含むことができる。

20

#### 【0614】

新生抗原特異的T C Rの特定後、T C Rをシーケンシングすることができる。一実施形態では、セクションX V I Iに関連して上記に述べた方法を用いてT C Rをシーケンシングすることができる。別の実施形態では、T C RのT C R a及びT C R bをバルクシーケンシングした後、頻度に基づいてペアリングすることができる。別の実施形態では、T C Rは、Howie et al., Science Translational Medicine 2015 (doi:10.1126/scitranslmed.aac5624)の方法を用いてシーケンシングし、ペアリングすることができる。別の実施形態では、T C Rは、Han et al., Nat Biotech 2014 (PMID 24952902, doi 10.1038/nbt.2938)の方法を用いてシーケンシングし、ペアリングすることができる。別の実施形態では、ペアリングされたT C R配列を、<https://www.biorxiv.org/content/early/2017/05/05/134841>及び／または<https://patents.google.com/patent/US20160244825A1>により記載される方法を用いて得ることができる<sup>158, 159</sup>。

30

#### 【0615】

別の実施形態では、T細胞のクローン集団を、希釈を限定し、次いでT細胞のクローン集団のT C R a及びT C R bをシーケンシングすることによって得ることができる。さらに別の実施形態では、T細胞をウェルを有するプレート上でウェル1つ当たり1個のT細胞となるように選別し、次いで各T細胞のT C R a及びT C R bをシーケンシング及びペアリングすることができる。

40

#### 【0616】

次に、新生抗原特異的T細胞を患者試料から特定し、特定された新生抗原特異的T細胞のT C Rをシーケンシングした後、シーケンシングされたT C Rを新たなT細胞にクローニングする。これらのクローニングされたT細胞は、新生抗原特異的受容体を含む（例えば、T C Rを含む細胞外ドメインを含む）。そのような細胞の集団、及びそのような細胞を含む組成物も提供される。いくつかの実施形態では、組成物または集団は、そのよ

50

うな細胞について濃縮される（例えば、TCRを発現する細胞が、組成物中の全細胞、またはT細胞またはCD4+細胞などの特定の細胞の少なくとも1、5、10、20、30、40、50、60、70、80、90、91、92、93、94、95、96、97、98、99、または99%超を構成するように）。いくつかの実施形態では、組成物は、本明細書に開示されるTCRを有する少なくとも1つの細胞を含む。かかる組成物には、養子細胞療法などで投与するための医薬組成物及び製剤がある。細胞及び組成物を対象、例えば患者に投与するための治療方法も提供される。

#### 【0617】

したがって、TCR（複数可）を発現する遺伝子操作された細胞も提供される。細胞は、一般的に哺乳動物細胞などの真核細胞であり、通常はヒト細胞である。いくつかの実施形態では、細胞は、血液、骨髄、リンパ、またはリンパ器官に由来するものであり、自然免疫または獲得免疫などの免疫系の細胞である（例えば、リンパ球を含む骨髄細胞またはリンパ系細胞、典型的にはT細胞及びNK細胞）。他の例示的な細胞としては、人工多能性幹細胞（iPSC）を含む、多能性及び万能幹細胞などの幹細胞が含まれる。細胞は通常、対象から直接単離されるか、かつ／または対象から単離されて凍結されたものなどの初代細胞である。いくつかの実施形態では、細胞には、1つ以上のT細胞または全T細胞集団、CD4+細胞、ならびに機能、活性化状態、成熟度、分化能、増殖、再循環、局在化、及び／または持続能力、抗原特異性、抗原受容体の種類、特定の臓器またはコンパートメントにおける存在、マーカーまたはサイトカイン分泌プロファイル、及び／または分化の程度によって規定されるものなどのこれらの亜集団などの他の細胞タイプが含まれる。治療される対象に関連して、細胞は、同種及び／または自家であってよい。方法には、オフザシェルフ法が含まれる。オフザシェルフ法用などのいくつかの態様では、細胞は、例えば幹細胞、例えば人工多能性幹細胞（iPSC）である。いくつかの態様では、方法は、本明細書に記載されるようにして、対象から細胞を単離することと、それらの細胞を調製、プロセッシング、培養、及び／または操作することと、凍結保存の前または後でそれらの細胞を同じ患者に再導入することと、を含む。

#### 【0618】

T細胞及び／またはCD8+細胞のサブタイプ及び亜集団としては、ナイーブT（TN）細胞、エフェクターT細胞（TEFF）、メモリーT細胞そのサブタイプ、例えば、幹細胞メモリーT（TSCM）、中枢メモリーT（TCM）、エフェクターメモリーT（TEM）、または最終分化したエフェクターメモリーT細胞、腫瘍浸潤リンパ球（TIL）、未成熟T細胞、成熟T細胞、ヘルパーT細胞、細胞傷害性T細胞、粘膜関連インバリアントT（MAIT）細胞、天然および適応制御性T（Treg）細胞、ヘルパーT細胞、例えばTH1細胞、TH2細胞、TH3細胞、TH17細胞、TH9細胞、TH22細胞、濾胞ヘルパーT細胞、 $\alpha/\beta$ T細胞、及び $\delta/\gamma$ T細胞がある。

#### 【0619】

いくつかの態様において、細胞は、ナチュラルキラー（NK）細胞である。いくつかの態様において、細胞は、単球または顆粒球、例えば骨髄性細胞、マクロファージ、好中球、樹状細胞、マスト細胞、好酸球、及び／または好塩基球である。

#### 【0620】

細胞は、内因性TCRの発現を低減するかまたはノックアウトするように遺伝子改変することができる。かかる改変は、本明細書に参照により援用するところのMol Ther Nucleic Acids. 2012 Dec; 1(12): e63; Blood. 2011 Aug 11; 118(6): 1495-503; Blood. 2012 Jun 14; 119(24): 5697-5705; Torikai, Hiroki et al "HLA and TCR Knockout by Zinc Finger Nucleases: Toward "off-the-shelf" Allogeneic T-Cell Therapy for CD19+ Malignancies." Blood 116. 21 (2010): 3766; Blood. 2018 Jan 18; 131(3): 311-322. doi: 10.1182/blood-2017-05-787

598；及びWO2016069283に記載されている。

【0621】

細胞は、サイトカイン分泌を促進するために遺伝子改変することができる。かかる改変は、Hsu C、Hughes MS、Zheng Z、Bray RB、Rosenberg SA、Morgan RA、Primary human T lymphocytes engineered with a codon-optimized IL-15 gene resist cytokine withdrawal-induced apoptosis and persist long-term in the absence of exogenous cytokine. J Immunol. 2005；175：7226-34；Quintarelli C、Vera JF、Savoldo B、Giordano Attianese GM、Pule M、Foster AE、Co-expression of cytokine and suicide genes to enhance the activity and safety of tumor-specific cytotoxic T lymphocytes. Blood. 2007；110：2793-802；及び、Hsu C、Jones SA、Cohen CJ、Zheng Z、Kerstann K、Zhou J、Cytokine-independent growth and clonal expansion of a primary human CD8+ T-cell clone following retroviral transduction with the IL-15 gene. Blood. 2007；109：5168-77に記載されている。

【0622】

T細胞上のケモカイン受容体と腫瘍分泌ケモカインのミスマッチが、腫瘍微小環境内へのT細胞の不適当なトラフィックの主要因であることが示されている。治療効果を高めるため、腫瘍微小環境中でのケモカインの認識を高めるように遺伝子改変することができる。そのような改変の例は、Moon、EKCarpénito、CSun、JWang、LCKapoor、VPredina、J Expression of a functional CCR2 receptor enhances tumor localization and tumor eradication by retargeted human T-cells expressing a mesothelin-specific chimeric antibody receptor. Clin Cancer Res. 2011；17：4719-4730；及びCraddock、JALu、ABear、APule、MBrenner、MKRooney、CM et al.、Enhanced tumor trafficking of GD2 chimeric antigen receptor T-cells by expression of the chemokine receptor CCR2b. J Immunother. 2010；33：780-788に記載されている。

【0623】

細胞を、CD28及び41BBなどの共刺激／増強受容体の発現を増大させるために遺伝子改変することができる。

【0624】

T細胞療法の副作用には、サイトカイン放出症候群及び長期のB細胞枯渇が含まれ得る。レシピエント細胞への自殺／安全スイッチの導入によって細胞療法の安全性プロファイルを改善することができる。したがって、細胞を自殺／安全スイッチを含むように遺伝子改変することができる。自殺／安全スイッチは、遺伝子が発現される細胞に作用物質、例えば薬物に対する感度を与え、細胞が作用物質と接触する際に死なせる遺伝子であってよい。例示的な自殺／安全スイッチについては、Protein Cell. 2017 Aug；8（8）：573-589に記載されている。自殺／安全スイッチはHSV-TKとすることができる。自殺／安全スイッチはシトシンデアミナーゼ、プリンヌクレオシドホスホリラーゼ、またはニトロレダクターゼであってもよい。自殺／安全スイッチは、米国特許出願公開第US20170166877A1号に記載されるRapaCIDE（商標

10

20

30

40

50

）であってもよい。自殺／安全スイッチシステムは、*Haematologica*. 2009 Sep; 94 (9) : 1316-1320に記載されるCD20／リツキシマブであってもよい。これらの参考文献は、それらの全容を参照により援用するものである。

【0625】

TCRは、ヘテロ二量体化小分子の存在下でのみアセンブルする分割受容体としてレシピエント細胞に導入することができる。かかるシステムについては、本明細書に参照により援用するところの*Science*. 2015 Oct 16; 350 (6258) : aab4077、及び米国特許第9,587,020号に記載されている。

【0626】

いくつかの実施形態では、細胞は1つ以上の核酸、例えば、本明細書に開示されるTCRをコードしたポリヌクレオチドを含み、かかるポリヌクレオチドは遺伝子操作により導入され、それにより本明細書に開示される組換えまたは遺伝子操作されたTCRを発現する。いくつかの実施形態では、核酸は、異種、すなわち、別の生物または細胞から得られたものなどの、細胞または細胞から得られる試料中に通常は存在していないものであって、例えば、操作される細胞及び／またはかかる細胞が由来する生物には通常みられないものである。いくつかの実施形態では、核酸は、複数の異なる細胞タイプからの異なるドメインをコードした核酸のキメラ複合体を含むものなど、自然界にはみられない核酸など、天然に存在しないものである。

【0627】

核酸は、コドン最適化されたヌクレオチド配列を含むことができる。特定の理論または機序に束縛されるものではないが、ヌクレオチド配列のコドン最適化はmRNA転写産物の翻訳効率を高めるものと考えられる。ヌクレオチド配列のコドン最適化では、天然のコドンを、同じアミノ酸をコードしているが細胞内でより利用されやすいtRNAによって翻訳される別のコドンに置換することにより、翻訳効率を高めることができる。ヌクレオチド配列の最適化は、翻訳を妨げる二次mRNA構造を低減させることで翻訳効率を高めることもできる。

【0628】

コンストラクトまたはベクターを用いてレシピエント細胞にTCRを導入することができる。例示的なコンストラクトを本明細書に記載する。TCRのα鎖及びβ鎖をコードするポリヌクレオチドは、単一のコンストラクト内であってもよく、別々のコンストラクト内であってもよい。α鎖及びβ鎖をコードするポリヌクレオチドは、プロモーター、例えば異種プロモーターと機能的に連結することができる。異種プロモーターは、例えば、EF1α、CMV、PGK1、Ubc、βアクチン、CAGプロモーターのような強力なプロモーターとすることができる。異種プロモーターは、弱いプロモーターであってもよい。異種プロモーターは誘導性プロモーターであってもよい。例示的な誘導性プロモーターとしては、これらに限定されるものではないが、TRE、NFAT、GAL4、LACなどが挙げられる。他の例示的な誘導性発現系は、本明細書にその全容を援用するところの米国特許第5,514,578号、同第6,245,531号、同第7,091,038号、及び欧州特許第0517805号に記載されている。

【0629】

TCRをレシピエント細胞に導入するためのコンストラクトは、シグナルペプチドをコードしたポリヌクレオチド（シグナルペプチドエレメント）を含んでもよい。シグナルペプチドは、導入されたTCRの表面トラフィッキングを促進することができる。例示的なシグナルペプチドとしては、これらに限定されるものではないが、CD4シグナルペプチド、免疫グロブリンシグナルペプチドが挙げられ、具体例としてはGM-CSF及びIgGκが挙げられる。かかるシグナルペプチドは、本明細書に参照により援用するところの*Trends Biochem Sci*. 2006 Oct; 31 (10) : 563-71. Epub 2006 Aug 21; 及びAn, et al. "Construction of a New Anti-CD19 Chimeric Antigen Receptor and the Anti-Leukemia Function Study of

10

20

30

40

50

the Transduced T-cells.” *Oncotarget* 7.9 (2016): 10638–10649. PMC. Web. 16 Aug. 2018に記載されている。

#### 【0630】

場合により、例えば、 $\alpha$ 鎖及び $\beta$ 鎖が単一のコンストラクトまたはオープンリーディングフレームから発現される場合、またはコンストラクトにマーカー遺伝子が含まれるような場合には、コンストラクトはリボソームスキップ配列を含むことができる。リボソームスキップ配列は、2Aペプチド、例えば、P2AまたはT2Aペプチドであってよい。例示的なP2AまたはT2Aペプチドは、本明細書に参照によりその全容を援用するところの *Scientific Reports volume 7, Article number: 2193 (2017)* に記載されている。場合により、FURIN/PACE切断部位が2Aエレメントの上流に導入される。FURIN/PACE切断部位は、例えば、<http://www.nuolan.net/substrates.html> に記載されている。切断ペプチドは、第Xa因子の切断部位であってもよい。 $\alpha$ 鎖及び $\beta$ 鎖が単一のコンストラクトまたはオープンリーディングフレームから発現される場合、コンストラクトは配列内リボソーム進入部位 (IRES) を含むことができる。

#### 【0631】

コンストラクトは、1つ以上のマーカー遺伝子をさらに含むことができる。例示的なマーカー遺伝子としては、これらに限定されるものではないが、GFP、ルシフェラーゼ、HA、lacZが挙げられる。マーカーは、当業者には周知の抗生物質耐性マーカー、重金属耐性マーカー、または殺生物剤耐性マーカーなどの選択可能なマーカーであってよい。マーカーは、栄養要求性宿主で使用するための補完的マーカーとすることができる。例示的な補完的マーカー及び栄養要求性宿主は、*Gene*. 2001 Jan 24; 263 (1–2): 159–69に記載されている。かかるマーカーは、IRES、フレームシフト配列、2Aペプチドリンカー、TCRとの融合体によって発現させるか、または別のプロモーターから別に発現させることができる。

#### 【0632】

TCRをレシピエント細胞に導入するための例示的なベクターとしては、これらに限定されるものではないが、アデノ随伴ウイルス、アデノウイルス、アデノウイルス+改変ワクシニア、アンカラウイルス (MVA)、アデノウイルス+レトロウイルス、アデノウイルス+センダイウイルス、アデノウイルス+ワクシニアウイルス、アルファウイルス (VEE) レプリコンワクチン、アンチセンスオリゴヌクレオチド、ビフィドバクテリウム・ロンガム、CRISPR-Cas9、E. coli、フラビウイルス、遺伝子銃、ヘルペスウイルス、単純ヘルペスウイルス、ラクトコッカス・ラクティス、電気穿孔法、レンチウイルス、リポフェクション、リステリア・モノサイトゲネス、麻疹ウイルス、改変ワクシニアアンカラウイルス (MVA)、mRNA電気穿孔法、ネイキッド/プラスミドDNA、ネイキッド/プラスミドDNA+アデノウイルス、ネイキッド/プラスミドDNA+改変ワクシニアアンカラウイルス (MVA)、ネイキッド/プラスミドDNA+RNAトランスファー、ネイキッド/プラスミドDNA+ワクシニアウイルス、ネイキッド/プラスミドDNA+水泡口炎ウイルス、ニューカッスル病ウイルス、非ウイルス性PiggyBac (商標) (PB) トランスポゾン、ナノ粒子ベースのシステム、ポリオウイルス、ポックスウイルス、ポックスウイルス+ワクシニアウイルス、レトルウイルス、RNAトランスファー、RNAトランスファー+ネイキッド/プラスミドDNA、RNAウイルス、サッカロマイセス・セレビシエ、サルモネラ・ティフィムリウム、セムリキ森林ウイルス、センダイウイルス、志賀赤痢菌、シミアンウイルス、siRNA、スリーピング・ビューティートランスポゾン、ストレプトコッカス・ミュータンス、ワクシニアウイルス、ベネズエラウマ脳炎ウイルスレプリコン、水泡口炎ウイルス、及びコレラ菌が挙げられる。

#### 【0633】

好ましい実施形態では、TCRは、アデノ随伴ウイルス (AAV)、アデノウイルス、CRISPR-CAS9、ヘルペスウイルス、レンチウイルス、リポフェクション、mR

NA電気穿孔法、PiggyBac（商標）（PB）トランスポゾン、レトロウイルス、RNAトランスファー、またはスリーピング・ビューティートランスポゾンによってレシピエント細胞に導入される。

【0634】

いくつかの実施形態では、TCRをレシピエント細胞に導入するためのベクターは、ウイルスベクターである。例示的なウイルスベクターとしては、アデノウイルスベクター、アデノ随伴ウイルス（AAV）ベクター、レンチウイルスベクター、ヘルペスウイルスベクター、レトロウイルスベクターなどが挙げられる。かかるベクターは本明細書に記載される。

【0635】

TCRをレシピエント細胞に導入するためのTCRコンストラクトの例示的な実施形態を図16に示す。いくつかの実施形態では、TCRコンストラクトは、5'～3'の方向に以下のポリヌクレオチド配列、すなわち、プロモーター配列、シグナルペプチド配列、TCRβ可変（TCRβv）配列、TCRβ定常（TCRβc）配列、切断ペプチド（例えば、P2A）、シグナルペプチド配列、TCRα可変（TCRαv）配列、及びTCRα定常（TCRαc）配列を含む。いくつかの実施形態では、コンストラクトのTCRβc及びTCRαc配列は、1つ以上のマウス領域、例えば、本明細書に記載される完全マウス定常配列またはヒト→マウスへのアミノ酸交換を含む。いくつかの実施形態では、コンストラクトはさらに、TCRαc配列の3'末端、切断ペプチド配列（例えば、T2A）に続くレポーター遺伝子を含む。一実施形態では、コンストラクトは、5'～3'の方向に以下のポリヌクレオチド配列、すなわち、プロモーター配列、シグナルペプチド配列、TCRβ可変（TCRβv）配列、1つ以上のマウス領域を含むTCRβ定常（TCRβc）配列、切断ペプチド（例えば、P2A）、シグナルペプチド配列、TCRα可変（TCRαv）配列、及び1つ以上のマウス領域を含むTCRα定常（TCRαc）配列、切断ペプチド配列（例えば、T2A）、及びレポーター遺伝子を含む。

【0636】

図17は、治療法の開発のためにTCRを発現系にクローニングするための例示的なP526コンストラクト骨格のヌクレオチド配列を示す。

【0637】

図18は、治療法の開発のために患者の新生抗原特異的TCRのクローノタイプ1を発現系にクローニングするための例示的なコンストラクト配列を示す。

【0638】

図19は、治療法の開発のために患者の新生抗原特異的TCRのクローノタイプ3を発現系にクローニングするための例示的なコンストラクト配列を示す。

【0639】

TCRをコードする単離された核酸、かかる核酸を含むベクター、及びかかるベクターを含む宿主細胞、ならびにTCRを作製するための組換え法も提供される。

【0640】

核酸は組換え体であってよい。組換え核酸は、天然または合成核酸セグメントを、生細胞内で複製可能な核酸分子またはその複製産物に連結することにより生細胞の外部で構築することができる。本明細書の目的では、複製はインビトロ複製またはインビボ複製であってよい。

【0641】

TCRの組換え作製を行うには、TCRをコードする核酸（複数可）を単離してさらなるクローニング（すなわち、DNAの増幅）または発現を行うための複製可能なベクターに挿入することができる。いくつかの態様では、核酸は、例えば本明細書に参照により援用するところの米国特許第5,204,244号に記載されるような相同組み換えによって作製することができる。

【0642】

多くの異なるベクターが当該技術分野では周知である。ベクターの構成要素は一般的に

10

20

30

40

50

以下のもの、すなわち、例えば、本明細書にその全容を参照により援用するところの米国特許第5,534,615号に記載されるような、シグナル配列、複製起点、1つ以上のマーカー遺伝子、エンハンサーエレメント、プロモーター、及び転写終結配列のうちの1つ以上を含む。

【0643】

T C R、抗体、またはその抗原結合フラグメントを発現させるのに適した例示的なベクターまたはコンストラクトとしては、例えば、p U C シリーズ ( F e r m e n t a s L i f e S c i e n c e s )、p B l u e s c r i p t シリーズ ( S t r a t a g e n e , L a J o l l a , C A )、p E T シリーズ ( N o v a g e n , M a d i s o n , W I )、p G E X シリーズ ( P h a r m a c i a B i o t e c h , U p p s a l a , S w e d e n )、及び p E X シリーズ ( C l o n t e c h , P a l o A l t o , C A ) が挙げられる。A G T l o、A G T l 1、A Z a p I I ( S t r a t a g e n e )、A E M B L 4、及び A N M l 1 4 9 などのバクテリオファージも本明細書に開示される T C R を発現させるうえで適当である。

【0644】

#### X I X . 治療の概要のフローチャート

図20は、一実施形態による、患者に個別化された新生抗原特異的治療を行うための方法のフローチャートである。他の実施形態では、本方法は、図29に示されるもの以外の異なる及び／またはさらなるステップを含んでもよい。さらに、本方法の各ステップは、異なる実施形態において、図20に関連して述べられる順序とは異なる順序で行うことができる。

【0645】

提示モデルを、上記に述べた質量分析データを用いて訓練する(2001)。患者試料を取得する(2002)。いくつかの実施形態では、患者試料は、腫瘍生検及び／または患者の末梢血を含む。ステップ2002で得られた患者試料をシーケンシングして、患者試料からの腫瘍抗原ペプチドが提示される尤度を予測するための提示モデルに入力するデータを特定する。ステップ2002で得られた患者試料からの腫瘍抗原ペプチドの提示尤度を、訓練された提示モデルを用いて予測する(2003)。予測された提示尤度に基づいて患者に対する治療抗原を特定する(2004)。次いで、別の患者試料を取得する(2005)。患者試料は、患者の末梢血、腫瘍浸潤リンパ球(T I L)、リンパ、リンパ節細胞、及び／または他のT細胞源を含み得る。ステップ2005で得られた患者試料を新生抗原特異的T細胞についてインビボでスクリーニングする(2006)。

【0646】

治療プロセスのこの時点において、患者は、T細胞療法及び／またはワクチン治療を受けることができる。ワクチン治療を受けるには、患者のT細胞が特異的である新生抗原を特定する(2014)。次いで、特定された新生抗原を含むワクチンを作製する(2015)。最後に、ワクチンを患者に投与する(2016)。

【0647】

T細胞治療を受けるには、新生抗原特異的T細胞を増殖させるか、及び／または新たな新生抗原特異的T細胞を遺伝子操作する。T細胞療法で使用するための新生抗原特異的T細胞を増殖させるには、細胞を単純に増殖させ(2007)、患者に注入する(2008)。

【0648】

T細胞療法用の新たな新生抗原特異的T細胞を遺伝子操作するには、インビボで特定された新生抗原特異的T細胞のT C Rをシーケンシングする(2009)。次いで、これらのT C R配列を発現ベクターにクローニングする(2010)。次にこの発現ベクター2910を新たなT細胞にトランスフェクトする(2011)。トランスフェクトしたT細胞を増殖させる(2012)。そして最後に、増殖させたT細胞を患者に注入する(2013)。

【0649】

10

20

30

40

50

患者はT細胞療法及びワクチン療法の療法を受けることができる。一実施形態では、患者は最初にワクチン療法を受け、その後、T細胞療法を受ける。このアプローチの利点の1つは、ワクチン療法により、検出可能なレベルのT細胞によって認識される腫瘍特異的T細胞の数及び新生抗原の数を増やすことができる点である。

【0650】

別の実施形態では、患者はT細胞療法を受けた後、ワクチン療法を受けることができ、その場合、ワクチンに含まれるエピトープのセットは、T細胞療法のターゲットとなるエピトープのうちの1つ以上を含む。このアプローチの利点の1つは、ワクチンの投与によって治療T細胞の増殖及び持続を促進することができる点である。

【0651】

## XX. 例示的なコンピュータ

図21は、図1及び図3に示した実体を実施するための例示的なコンピュータ2100を説明する。コンピュータ2100は、チップセット2104に連結された少なくとも1つのプロセッサ2102を含む。チップセット2104は、メモリコントローラハブ2120及び入力／出力(I/O)コントローラハブ2122を含む。メモリ2106及びグラフィックスアダプタ2112は、メモリコントローラハブ2120に連結されており、ディスプレイ2118は、グラフィックスアダプタ2112に連結されている。記憶デバイス2108、入力装置2114、及びネットワークアダプタ2116は、I/Oコントローラハブ2122に連結されている。コンピュータ2100の他の実施形態は、異なるアーキテクチャを有する。

【0652】

記憶デバイス2108は、ハードドライブ、コンパクトディスク読み出し専用メモリ(CD-ROM)、DVD、またはソリッドステートメモリ装置などの、非一時的なコンピュータ可読の記憶媒体である。メモリ2106は、プロセッサ2102によって使用される命令及びデータを保持する。入力インターフェイス2114は、タッチスクリーンインターフェイス、マウス、トラックボール、もしくは他のタイプのポインティングデバイス、キーボード、またはそれらのいくつかの組み合わせであり、データをコンピュータ2100中に入力するために使用される。いくつかの実施形態において、コンピュータ2100は、ユーザーからのジェスチャーを介して、入力インターフェイス2114からの入力(例えば、コマンド)を受け取るように構成されていてもよい。グラフィックスアダプタ2112は、ディスプレイ2118上に画像及び他の情報を表示する。ネットワークアダプタ2116は、コンピュータ2100を、1つ以上のコンピュータネットワークに連結する。

【0653】

コンピュータ2100は、本明細書に記載した機能性を提供するためのコンピュータプログラムモジュールを遂行するように適合している。本明細書において使用される場合、「モジュール」という用語は、特定の機能性を提供するために使用されるコンピュータプログラム論理を指す。したがって、モジュールは、ハードウェア、ファームウェア、及び／またはソフトウェアにおいて実行されることができる。一実施形態では、プログラムモジュールは、記憶デバイス2108に保存され、メモリ2106中にロードされ、プロセッサ2102によって遂行される。

【0654】

図1の実体によって使用されるコンピュータ2100のタイプは、実体によって必要とされる実施形態及びプロセッシングパワーに応じて変動することができる。例えば、提示特定システム160は、単一のコンピュータ2100、または、例えばサーバーファームにおいてネットワークを通して互いに通信する複数のコンピュータ2100において、起動することができる。コンピュータ2100は、グラフィックスアダプタ2112及びディスプレイ2118などの、上記の構成要素のうちのいくつかを欠いてもよい。

【0655】

参考文献

10

20

30

40

50

1. Desrichard, A., Snyder, A. & Chan, T. A. Cancer Neoantigens and Applications for Immunotherapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* (2015). doi:10.1158/1078-0432.CCR-14-3175
2. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* 348, 69–74 (2015).
3. Gubin, M. M., Artyomov, M. N., Mardis, E. R. & Schreiber, R. D. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J. Clin. Invest.* 125, 3413–3421 (2015).
4. Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348, 124–128 (2015).
5. Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189–2199 (2014). 10
6. Carreno, B. M. et al. Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T-cells. *Science* 348, 803–808 (2015).
7. Tran, E. et al. Cancer immunotherapy based on mutation-specific CD4+ T-cells in a patient with epithelial cancer. *Science* 344, 641–645 (2014).
8. Hacothen, N. & Wu, C. J.-Y. United States Patent Application: 0110293637 - COMPOSITIONS AND METHODS OF IDENTIFYING TUMOR SPECIFIC NEOANTIGENS. (A1). at <http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20110293637.PGNR.>>
9. Lundegaard, C., Hoof, I., Lund, O. & Nielsen, M. State of the art and challenges in sequence based T-cell epitope prediction. *Immunome Res.* 6 Suppl 2, S3 (2010). 20
10. Yadav, M. et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* 515, 572–576 (2014).
11. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J. & Mann, M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics MCP* 14, 658–673 (2015).
12. Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 350, 207–211 (2015).
13. Yoshida, K. & Ogawa, S. Splicing factor mutations and cancer. *Wiley Interdiscip. Rev. RNA* 5, 445–459 (2014).
14. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550 (2014).
15. Rajasagi, M. et al. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* 124, 453–462 (2014). 30
16. Downing, S. R. et al. United States Patent Application: 0120208706 - OPTIMIZATION OF MULTIGENE ANALYSIS OF TUMOR SAMPLES. (A1). at <http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20120208706.PGNR.>>
17. Target Capture for NextGen Sequencing - IDT. at <http://www.idtdna.com/pages/products/nextgen/target-capture>
18. Shukla, S. A. et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* 33, 1152–1158 (2015).
19. Cieslik, M. et al. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* 25, 1372–1381 (2015).
20. Bodini, M. et al. The hidden genomic landscape of acute myeloid leukemia: subclonal structure revealed by undetected mutations. *Blood* 125, 600–605 (2015). 40

21. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinforma. Oxf. Engl.* 28, 1811–1817 (2012).
22. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219 (2013).
23. Wilkerson, M. D. et al. Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. *Nucleic Acids Res.* 42, e107 (2014).
24. Mose, L. E., Wilkerson, M. D., Hayes, D. N., Perou, C. M. & Parker, J. S. ABRA: improved coding indel detection via assembly-based realignment. *Bioinforma. Oxf. Engl.* 30, 2813–2815 (2014).
25. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinforma. Oxf. Engl.* 25, 2865–2871 (2009).
26. Lam, H. Y. K. et al. Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nat. Biotechnol.* 28, 47–55 (2010).
27. Frampton, G. M. et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* 31, 1023–1031 (2013).
28. Boegel, S. et al. HLA typing from RNA-Seq sequence reads. *Genome Med.* 4, 102 (2012).
29. Liu, C. et al. ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res.* 41, e142 (2013).
30. Mayor, N. P. et al. HLA Typing for the Next Generation. *PloS One* 10, e0127153 (2015).
31. Roy, C. K., Olson, S., Graveley, B. R., Zamore, P. D. & Moore, M. J. Assessing long-distance RNA sequence connectivity via RNA-templated DNA-DNA ligation. *eLife* 4, (2015).
32. Song, L. & Florea, L. CLASS: constrained transcript assembly of RNA-seq reads. *BMC Bioinformatics* 14 Suppl 5, S14 (2013).
33. Maretty, L., Sibbesen, J. A. & Krogh, A. Bayesian transcriptome assembly. *Genome Biol.* 15, 501 (2014).
34. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295 (2015).
35. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinforma. Oxf. Engl.* (2011). doi:10.1093/bioinformatics/btr355
36. Vitting-Seerup, K., Porse, B. T., Sandelin, A. & Waage, J. spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* 15, 81 (2014).
37. Rivas, M. A. et al. Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* 348, 666–669 (2015).
38. Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J. & Akey, J. M. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res.* 21, 1728–1737 (2011).
39. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* 31, 166–169 (2015).
40. Furney, S. J. et al. SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov.* (2013). doi:10.1158/2159-8290.CD-13-0330
41. Zhou, Q. et al. A chemical genetics approach for the functional assessment of novel cancer genes. *Cancer Res.* (2015). doi:10.1158/0008-5472.CAN-14-2930
42. Maguire, S. L. et al. SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* 235, 571–580 (2015).
43. Carithers, L. J. et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation Biobanking* 13, 311–319 (2015).
44. Xu, G. et al. RNA CoMPASS: a dual approach for pathogen and host transcriptome analysis of RNA-seq datasets. *PloS One* 9, e89445 (2014).

10

20

30

40

50

45. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinforma. Oxf. Engl.* (2015). doi:10.1093/bioinformatics/btv639
46. Jorgensen, K. W., Rasmussen, M., Buus, S. & Nielsen, M. NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* 141, 18–26 (2014).
47. Larsen, M. V. et al. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur. J. Immunol.* 35, 2295–2303 (2005).
48. cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* 57, 33–41 (2005). 10
49. Boisvert, F.-M. et al. A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. *Mol. Cell. Proteomics* 11, M111.011429–M111.011429 (2012).
50. Duan, F. et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* 211, 2231–2248 (2014).
51. Janeway's Immunobiology: 9780815345312: Medicine & Health Science Books @ Amazon.com. at <<http://www.amazon.com/Janeways-Immunobiology-Kenneth-Murphy/dp/0815345313>>
52. Calis, J. J. A. et al. Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput. Biol.* 9, e1003266 (2013).
53. Zhang, J. et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 346, 256–259 (2014).
54. Walter, M. J. et al. Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* 366, 1090–1098 (2012). 20
55. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* 1992. 255: 1261-1263.
56. Zarling AL, Polefrone JM, Evans AM, Mikesh LM, Shabanowitz J, Lewis ST, Engelhard VH, Hunt DF. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc Natl Acad Sci U S A.* 2006 Oct 3;103(40):14889-94.
57. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics.* 2015 Mar;14(3):658-73. doi: 10.1074/mcp.M114.042812.
58. Abelin JG, Trantham PD, Penny SA, Patterson AM, Ward ST, Hildebrand WH, Cobbold M, Bai DL, Shabanowitz J, Hunt DF. Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nat Protoc.* 2015 Sep;10(9):1308-18. doi: 10.1038/nprot.2015.086. Epub 2015 Aug 6 30
59. Barnstable CJ, Bodmer WF, Brown G, Galfre G, Milstein C, Williams AF, Ziegler A. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell.* 1978 May;14(1):9-20.
60. Goldman JM, Hibbin J, Kearney L, Orchard K, Th'ng KH. HLA-DR monoclonal antibodies inhibit the proliferation of normal and chronic granulocytic leukaemia myeloid progenitor cells. *Br J Haematol.* 1982 Nov;52(3):411-20.
61. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. *Proteomics.* 2013 Jan;13(1):22-4. doi: 10.1002/pmic.201200439. Epub 2012 Dec 4.
62. Eng JK, Hoopmann MR, Jahan TA, Egertson JD, Noble WS, MacCoss MJ. A deeper look into Comet--implementation and features. *J Am Soc Mass Spectrom.* 2015 Nov;26(11):1865-74. doi: 10.1007/s13361-015-1179-x. Epub 2015 Jun 27. 40

63. Lukas Käll, Jesse Canterbury, Jason Weston, William Stafford Noble and Michael J. MacCoss. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods* 4:923 – 925, November 2007
64. Lukas Käll, John D. Storey, Michael J. MacCoss and William Stafford Noble. Assigning confidence measures to peptides identified by tandem mass spectrometry. *Journal of Proteome Research*, 7(1):29-34, January 2008
65. Lukas Käll, John D. Storey and William Stafford Noble. Nonparametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics*, 24(16):i42-i48, August 2008
66. Bo Li and C. olin N. Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a referenfe genome. *BMC Bioinformatics*, 12:323, August 2011
67. Hillary Pearson, Tariq Daouda, Diana Paola Granados, Chantal Durette, Eric Bonneil, Mathieu Courcelles, Anja Rodenbrock, Jean-Philippe Laverdure, Caroline Côté, Sylvie Mader, Sébastien Lemieux, Pierre Thibault, and Claude Perreault. MHC class I-associated peptides derive from selective regions of the human genome. *The Journal of Clinical Investigation*, 2016,
68. Juliane Liepe, Fabio Marino, John Sidney, Anita Jeko, Daniel E. Bunting, Alessandro Sette, Peter M. Klotzel, Michael P. H. Stumpf, Albert J. R. Heck, Michele Mishito. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science*, 21, October 2016.
69. Mommen GP., Marino, F., Meiring HD., Poelen, MC., van Gaans-van den Brink, JA., Mohammed S., Heck AJ., and van Els CA. Sampling From the Proteome to the Human Leukocyte Antigen-DR (HLA-DR) Ligandome Proceeds Via High Specificity. *Mol Cell Proteomics* 15(4): 1412-1423, April 2016.
70. Sebastian Kreiter, Mathias Vormehr, Niels van de Roemer, Mustafa Diken, Martin Löwer, Jan Diekmann, Sebastian Boegel, Barbara Schrörs, Fulvia Vascotto, John C. Castle, Arbel D. Tadmor, Stephen P. Schoenberger, Christoph Huber, Özlem Türeci, and Ugur Sahin. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* 520, 692-696, April 2015.
71. Tran E., Turcotte S., Gros A., Robbins P.F., Lu Y.C., Dudley M.E., Wunderlich J.R., Somerville R.P., Hogan K., Hinrichs C.S., Parkhurst M.R., Yang J.C., Rosenberg S.A. Cancer immunotherapy based on mutation-specific CD4+ T-cells in a patient with epithelial cancer. *Science* 344(6184) 641-645, May 2014.
72. Andreatta M., Karosiene E., Rasmussen M., Stryhn A., Buus S., Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* 67(11-12) 641-650, November 2015.
73. Nielsen, M., Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 10:296, September 2009.
74. Nielsen, M., Lundegaard, C., Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 8:238, July 2007.
75. Zhang, J., et al. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics*. 11(4):1-8. 1/2/2012.
76. Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189–2199 (2014).
77. Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348, 124–128 (2015).
78. Gubin, M. M., Artyomov, M. N., Mardis, E. R. & Schreiber, R. D. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J. Clin. Invest.* 125, 3413–3421 (2015).
79. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* 348, 69–74 (2015).
80. Carreno, B. M. et al. Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T-cells. *Science* 348, 803–808 (2015).

10

20

30

40

50

81. Ott, P. A. et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 547, 217–221 (2017).
82. Sahin, U. et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* 547, 222–226 (2017).
83. Tran, E. et al. T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N. Engl. J. Med.* 375, 2255–2262 (2016).
84. Gros, A. et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat. Med.* 22, 433–438 (2016).
85. The problem with neoantigen prediction. *Nat. Biotechnol.* 35, 97–97 (2017).
86. Vitiello, A. & Zanetti, M. Neoantigen prediction and the need for validation. *Nat. Biotechnol.* 35, 815–817 (2017).
87. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J. & Mann, M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics MCP* 14, 658–673 (2015).
88. Vita, R. et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* 43, D405–412 (2015).
89. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinforma. Oxf. Engl.* 32, 511–517 (2016).
90. O'Donnell, T. J. et al. MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. *Cell Syst.* (2018). doi:10.1016/j.cels.2018.05.014
91. Bassani-Sternberg, M. et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* 7, 13404 (2016).
92. Abelin, J. G. et al. Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity* 46, 315–326 (2017).
93. Yadav, M. et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* 515, 572–576 (2014).
94. Stranzl, T., Larsen, M. V., Lundegaard, C. & Nielsen, M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics* 62, 357–368 (2010).
95. Bentzen, A. K. et al. Large-scale detection of antigen-specific T-cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat. Biotechnol.* 34, 1037–1045 (2016).
96. Tran, E. et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* 350, 1387–1390 (2015).
97. Stronen, E. et al. Targeting of cancer neoantigens with donor-derived T-cell receptor repertoires. *Science* 352, 1337–1341 (2016).
98. Trolle, T. et al. The Length Distribution of Class I-Restricted T-cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference. *J. Immunol. Baltim. Md* 1950 196, 1480–1487 (2016).
99. Di Marco, M. et al. Unveiling the Peptide Motifs of HLA-C and HLA-G from Naturally Presented Peptides and Generation of Binding Prediction Matrices. *J. Immunol. Baltim. Md* 1950 199, 2639–2651 (2017).
100. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning*. (MIT Press, 2016).
101. Sette, A. et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T-cell epitopes. *J. Immunol. Baltim. Md* 1950 153, 5586–5592 (1994).
102. Fortier, M.-H. et al. The MHC class I peptide repertoire is molded by the transcriptome. *J. Exp. Med.* 205, 595–610 (2008).
103. Pearson, H. et al. MHC class I-associated peptides derive from selective regions of the human genome. *J. Clin. Invest.* 126, 4690–4701 (2016).
104. Bassani-Sternberg, M. et al. Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allosteric regulating HLA specificity. *PLoS Comput. Biol.* 13, e1005725 (2017).

10

20

30

40

50

105. Andreatta, M., Lund, O. & Nielsen, M. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinforma. Oxf. Engl.* 29, 8–14 (2013).
106. Andreatta, M., Alvarez, B. & Nielsen, M. GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx248
107. Gros, A. et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat. Med.* 22, 433–438 (2016).
108. Zacharakis, N. et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nat. Med.* 24, 724–730 (2018).
109. Chudley, L. et al. Harmonisation of short-term in vitro culture for the expansion of antigen-specific CD8+ T-cells with detection by ELISPOT and HLA-multimer staining. *Cancer Immunol. Immunother.* 63, 1199–1211 (2014). 10
110. Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 350, 207–211 (2015).
111. Anagnostou, V. et al. Evolution of Neoantigen Landscape during Immune Checkpoint Blockade in Non-Small Cell Lung Cancer. *Cancer Discov.* 7, 264–276 (2017).
112. Carreno, B. M. et al. Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T-cells. *Science* 348, 803–808 (2015).
113. Stevanović, S. et al. Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. *Science* 356, 200–205 (2017).
114. Pasetto, A. et al. Tumor- and Neoantigen-Reactive T-cell Receptors Can Be Identified Based on Their Frequency in Fresh Tumor. *Cancer Immunol. Res.* 4, 734–743 (2016).
115. Gillette, M. A. & Carr, S. A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat. Methods* 10, 28–34 (2013). 20
116. Boegel, S., Löwer, M., Bukur, T., Sahin, U. & Castle, J. C. A catalog of HLA type, HLA expression, and neo-epitope candidates in human cancer cell lines. *Oncoimmunology* 3, e954893 (2014).
117. Johnson, D. B. et al. Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nat. Commun.* 7, 10582 (2016).
118. Robbins, P. F. et al. A Pilot Trial Using Lymphocytes Genetically Engineered with an NY-ESO-1-Reactive T-cell Receptor: Long-term Follow-up and Correlates with Response. *Clin. Cancer Res.* 21, 1019–1027 (2015).
119. Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189–2199 (2014).
120. Calis, J. J. A. et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput. Biol.* 9, e1003266 (2013). 30
121. Duan, F. et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* 211, 2231–2248 (2014).
122. Glanville, J. et al. Identifying specificity groups in the T-cell receptor repertoire. *Nature* 547, 94–98 (2017).
123. Dash, P. et al. Quantifiable predictive features define epitope-specific T-cell receptor repertoires. *Nature* 547, 89–93 (2017).
124. Hunt, D. F. et al. Pillars article: Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* 1992, 255: 1261–1263. *J. Immunol. Baltim. Md 1950* 179, 2669–2671 (2007).
125. Zarling, A. L. et al. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14889–14894 (2006).
126. Abelin, J. G. et al. Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nat. Protoc.* 10, 1308–1318 (2015). 40

127. Barnstable, C. J. et al. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell* 14, 9–20 (1978).
128. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* 13, 22–24 (2013).
129. Eng, J. K. et al. A deeper look into Comet--implementation and features. *J. Am. Soc. Mass Spectrom.* 26, 1865–1874 (2015).
130. Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J. Proteome Res.* 7, 29–34 (2008).
131. Käll, L., Storey, J. D. & Noble, W. S. Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinforma. Oxf. Engl.* 24, i42–48 (2008).
132. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* 4, 923–925 (2007).
133. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323 (2011).
134. Chollet, F. & others. Keras. (2015).
135. Bastien, F. et al. Understanding the difficulty of training deep feedforward neural networks. *Proc. Thirteen. Int. Conf. Artif. Intell. Stat.* 249–256 (2010).
136. Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* 249–256 (2010).
137. Kingma, D. & Ba, J. Adam: A method for stochastic optimization. *ArXiv Prepr. ArXiv14126980* (2014).
138. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18, 6097–6100 (1990).
139. Rubinsteyn, A., O'Donnell, T., Damaraju, N. & Hammerbacher, J. Predicting Peptide-MHC Binding Affinities With Imputed Training Data. *bioRxiv* (2016). doi:<https://doi.org/10.1101/054775>
140. Tran, E. et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* 350, 1387–1390 (2015).
141. Stronen, E. et al. Targeting of cancer neoantigens with donor-derived T-cell receptor repertoires. *Science* 352, 1337–1341 (2016).
142. Janetzki, S., Cox, J. H., Oden, N. & Ferrari, G. Standardization and validation issues of the ELISPOT assay. *Methods Mol. Biol. Clifton NJ* 302, 51–86 (2005).
143. Janetzki, S. et al. Guidelines for the automated evaluation of Elispot assays. *Nat. Protoc.* 10, 1098–1115 (2015).
144. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* 25, 1754–1760 (2009).
145. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498 (2011).
146. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* (2012).
147. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w<sup>1118</sup>; iso-2; iso-3. *Fly (Austin)* 6, 80–92 (2012).
148. Szolek, A. et al. OptiType: precision HLA typing from next-generation sequencing data. *Bioinforma. Oxf. Engl.* 30, 3310–3316 (2014).
149. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219 (2013).

10

20

30

40

50

150. Scholz, E. M. et al. Human Leukocyte Antigen (HLA)-DRB1\*15:01 and HLA-DRB5\*01:01 Present Complementary Peptide Repertoires. *Front. Immunol.* 8, 984 (2017).
151. Ooi, J. D. et al. Dominant protection from HLA-linked autoimmunity by antigen-specific regulatory T-cells. *Nature* 545, 243–247 (2017).
152. Karosiene, E. et al. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics* 65, 711–724 (2013).
153. Dudley ME, Gross CA, Langan MM, et al. CD8+ enriched “young” tumor infiltrating lymphocytes can mediate regression of metastatic melanoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2010;16(24):6122-6131. doi:10.1158/1078-0432.CCR-10-1297.
154. Dudley ME, Wunderlich JR, Shelton TE, Even J, Rosenberg SA. Generation of Tumor-Infiltrating Lymphocyte Cultures for Use in Adoptive Transfer Therapy for Melanoma Patients. *Journal of immunotherapy (Hagerstown, Md : 1997)*. 2003;26(4):332-342.
155. Cohen CJ, Gartner JJ, Horovitz-Fried M, et al. Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. *The Journal of Clinical Investigation*. 2015;125(10):3981-3991. doi:10.1172/JCI82416.
156. Kelderman, S. , Heemskerk, B. , Fanchi, L. , Philips, D. , Toebes, M. , Kvistborg, P. , Bauren, M. M., Rooij, N. , Michels, S. , Germeroth, L. , Haanen, J. B. and Schumacher, N. M. (2016), Antigen-specific TIL therapy for melanoma: A flexible platform for personalized cancer immunotherapy. *Eur. J. Immunol.*, 46: 1351-1360. doi:10.1002/eji.201545849.
157. Hall M, Liu H, Malafa M, et al. Expansion of tumor-infiltrating lymphocytes (TIL) from human pancreatic tumors. *Journal for Immunotherapy of Cancer*. 2016;4:61. doi:10.1186/s40425-016-0164-7.
158. Briggs A, Goldfless S, Timberlake S, et al. Tumor-infiltrating immune repertoires captured by single-cell barcoding in emulsion. *bioRxiv*. 2017. doi.org/10.1101/134841.
159. US Patent Application No. 20160244825A1.
160. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Arbor, A. (2014). Going Deeper with Convolutions. 1–9. ArXiv 1409.4842. doi.org/10.1109/CVPR.2015.7298594, P., Reed, S., Anguelov, D., ... Arbor, A. (2014). *Going Deeper with Convolutions*. 1–9.
161. Maiers, M, et al. "High-resolution HLA alleles and haplotypes in the United States population." *Human Immunology*, vol. 68, no. 9, 24 May 2007, pp. 779-88, doi:https://www.sciencedirect.com/science/article/abs/pii/S019885907000948?via%3Dihub.
162. Mak, Tak W., et al. *Primer to the Immune Response*. 2nd ed., Elsevier Inc., 2014, pp. 143-59, [www.sciencedirect.com/science/article/pii/B9780123852458000066?via%3Dihub](https://www.sciencedirect.com/science/article/pii/B9780123852458000066?via%3Dihub).
163. Parkhurst, M. R., Robbins, P. F., Tran, E., Prickett, T. D., Gartner, J. J., Jia, L., ... Rosenberg, S. A. (2019). Unique Neoantigens Arise from Somatic Mutations in Patients with Gastrointestinal Cancers. *Cancer Discovery*, 9(8), 1022–1035. <https://doi.org/10.1158/2159-8290.cd-18-1494>

10

20

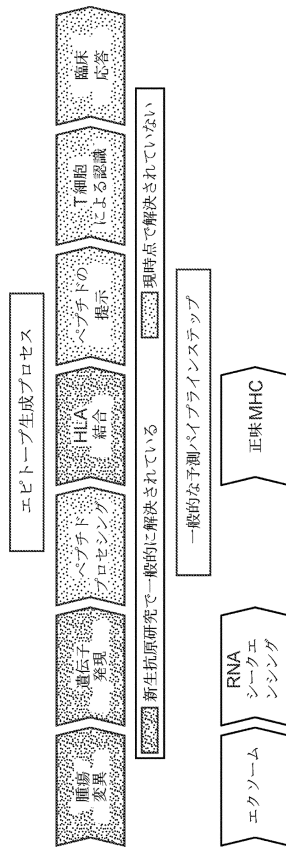
30

40

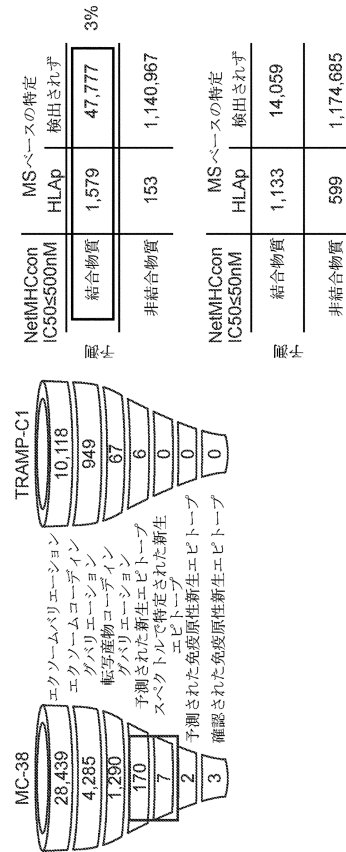
50

## 【図面】

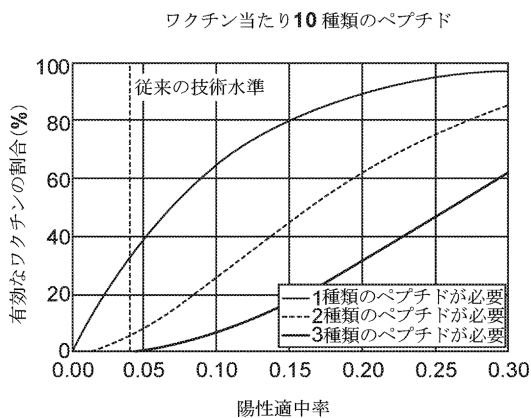
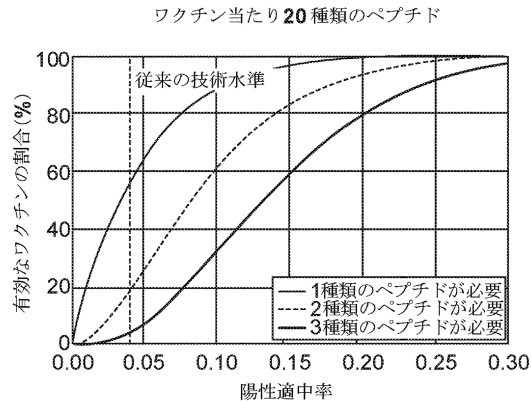
【図 1 A】



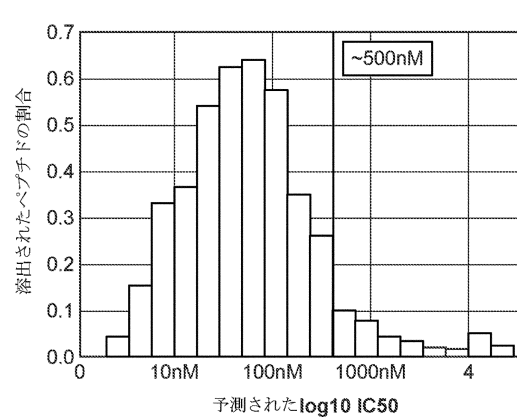
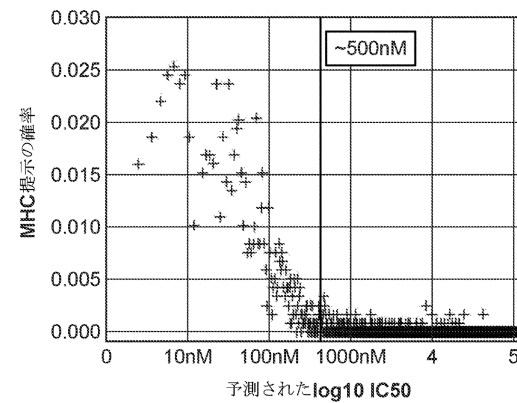
【図 1 B】



【図 1 C】



【図 1 D】



10

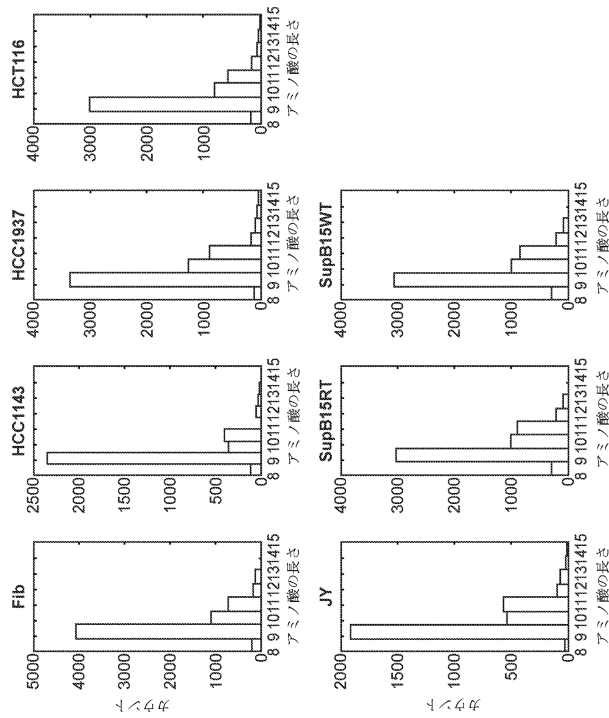
20

30

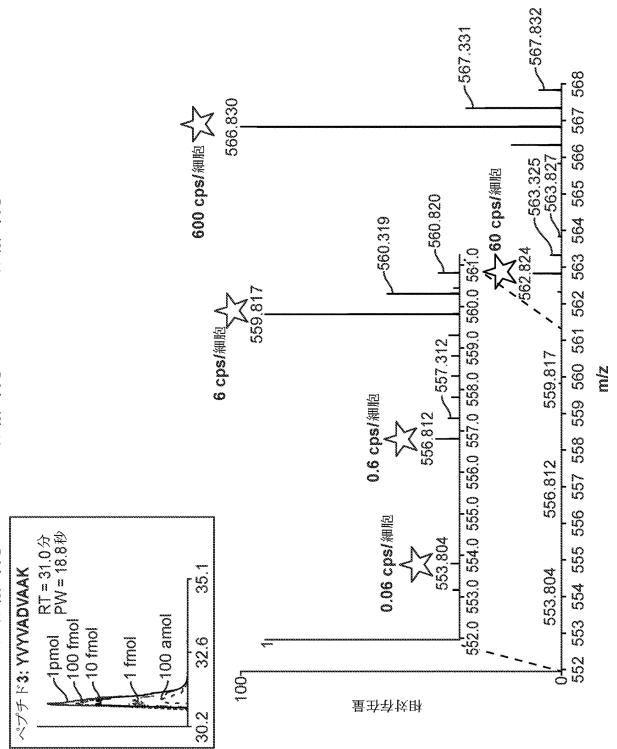
40

50

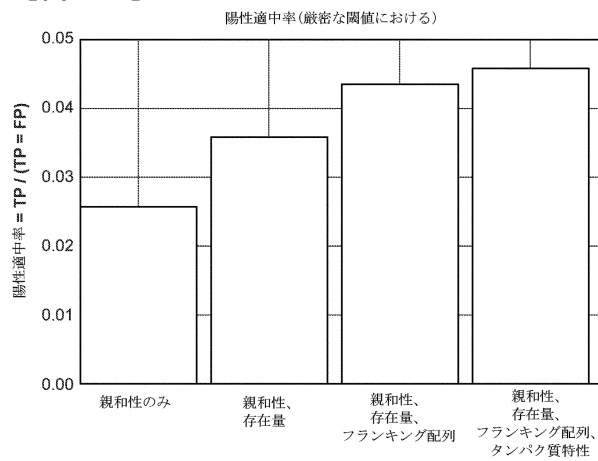
【図 1 E】



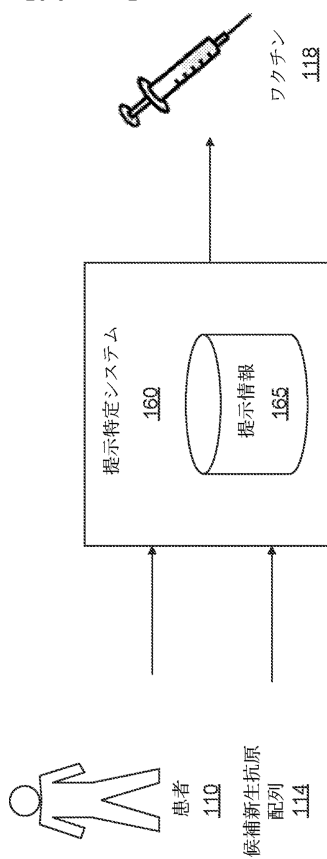
【図 1 F】



【図 1 G】



【図 2 A】



10

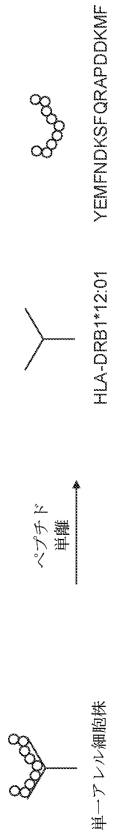
20

30

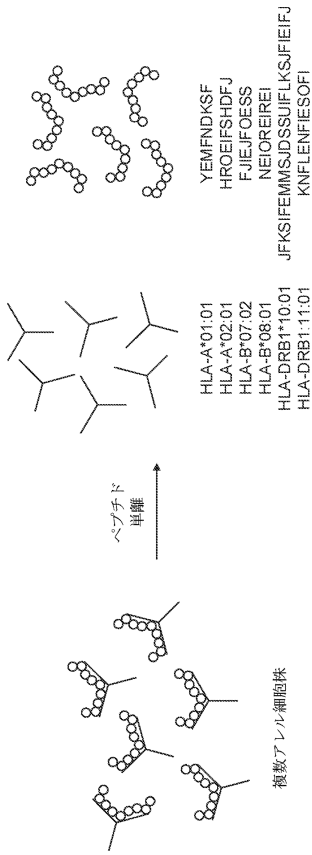
40

50

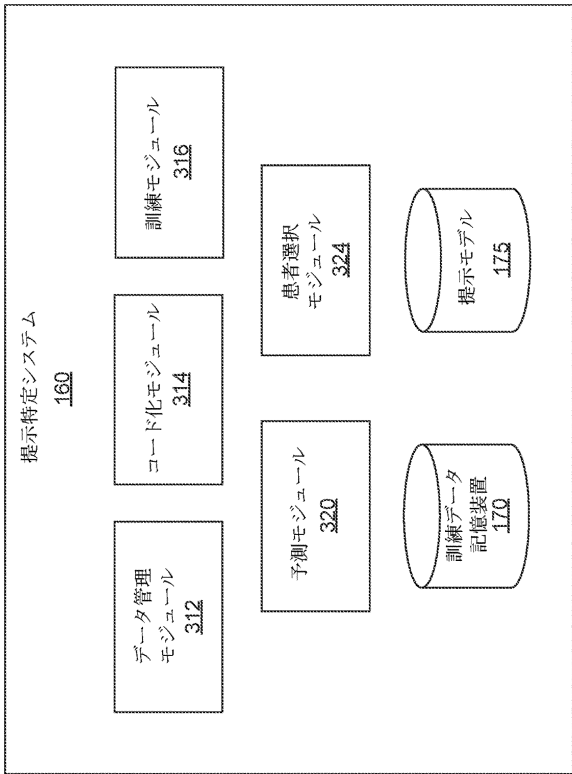
【図 2 B】



【図 2 C】



【図 3】



【図 4】

| 訓練データ<br>170A   |            |           |                |                  |                 |         |
|-----------------|------------|-----------|----------------|------------------|-----------------|---------|
| アレル依存 (x)       |            |           |                | アレル非依存 (w)       |                 |         |
| ペプチド配列 (p)      | 親和性 (b·nm) | 安定性 (s·h) | アレル (a)        | C末端フランキンゲン配列 (c) | mRNA Q. (m·TPM) | ラベル (v) |
| QCEIWAREFLKEIGJ | 1000       | 1         | HLA-DRB3*01:01 | FJELFISBOSJFIE   | 10 <sup>2</sup> | 提示されない  |
| FIUHFWI         | 1500       | 15        | HLA-C*01:03    | FEGRKUOOI        | 10 <sup>3</sup> | 提示される   |
| FEWRHRJTRJUR    | 650        | 20        | HLA-C*01:03    | PJFIOEJOJGEIO    | 10 <sup>1</sup> | 提示される   |
| QIEJOEIE        | 500        | 1         | HLA-B*07:02    | PJFIOEJOJGEIO    | 1               | 提示される   |
|                 | 600        | 14        | HLA-C*01:03    |                  |                 |         |
|                 | 1200       | 7         | HLA-A*01:01    |                  |                 |         |

10

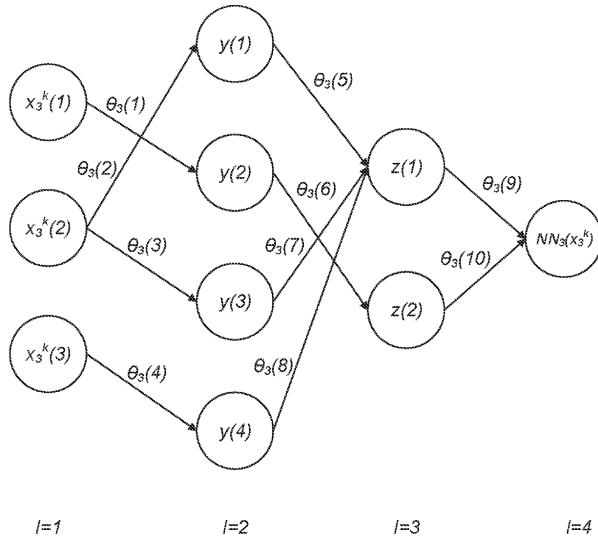
20

30

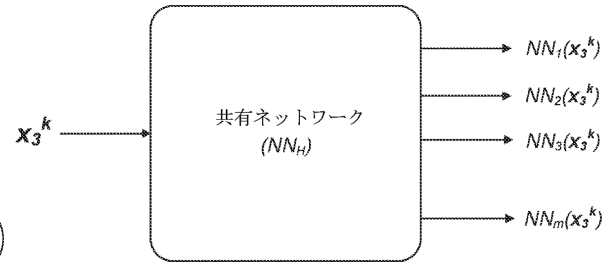
40

50

【図 5】

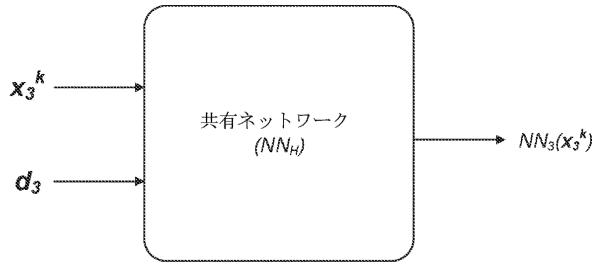


【図 6 A】

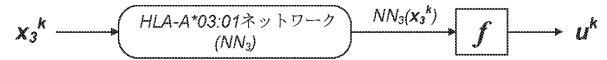


10

【図 6 B】

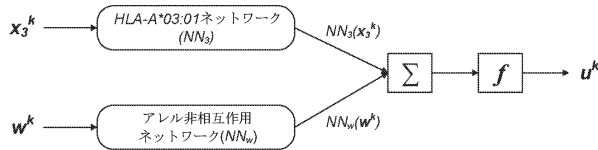


【図 7】

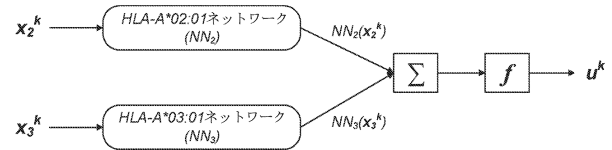


20

【図 8】



【図 9】

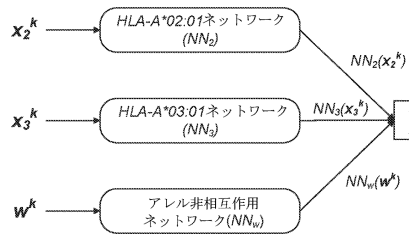


30

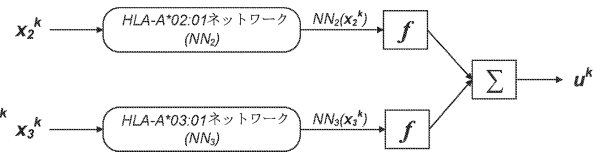
40

50

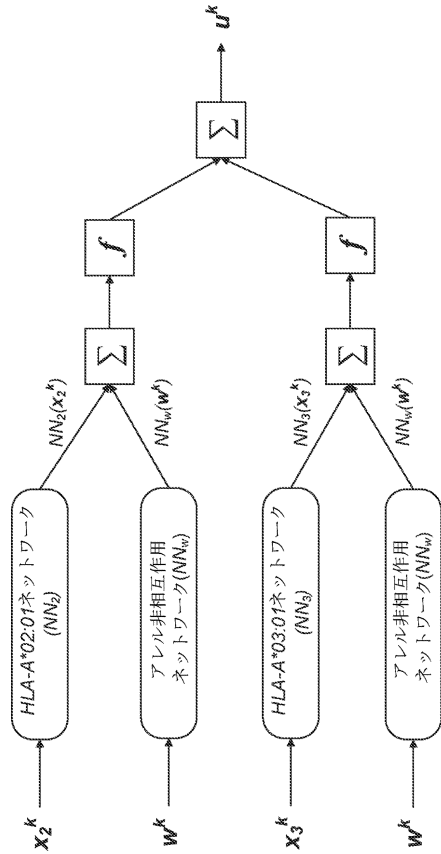
【図 10】



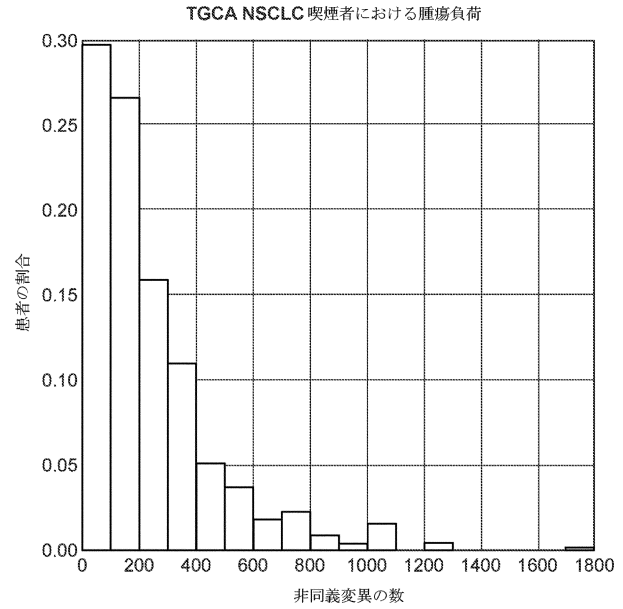
【図 11】



【図 12】



【図 13 A】



10

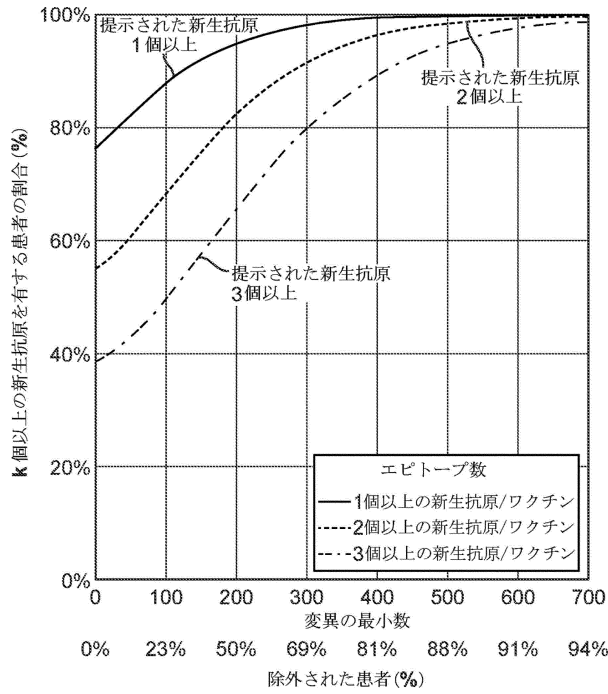
20

30

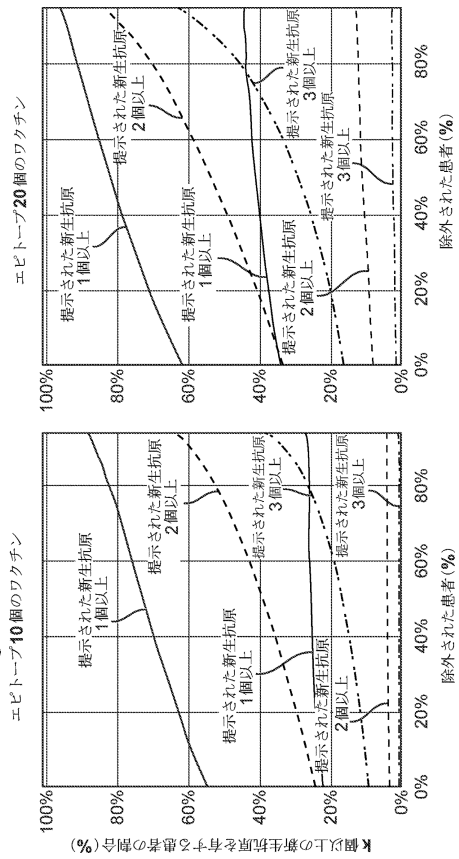
40

50

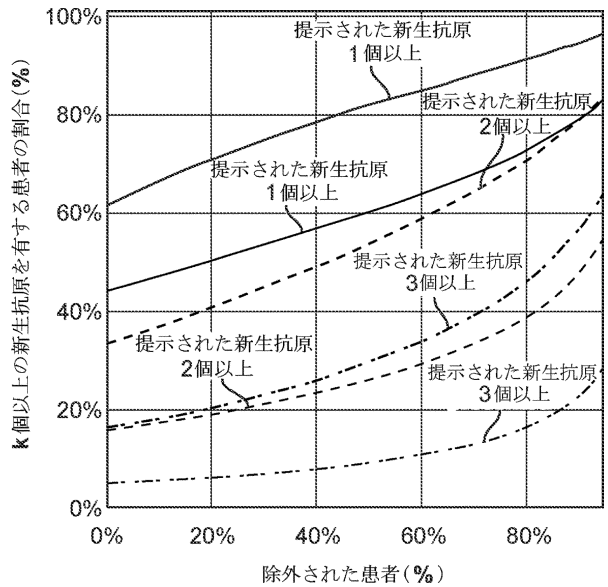
【図 1 3 B】



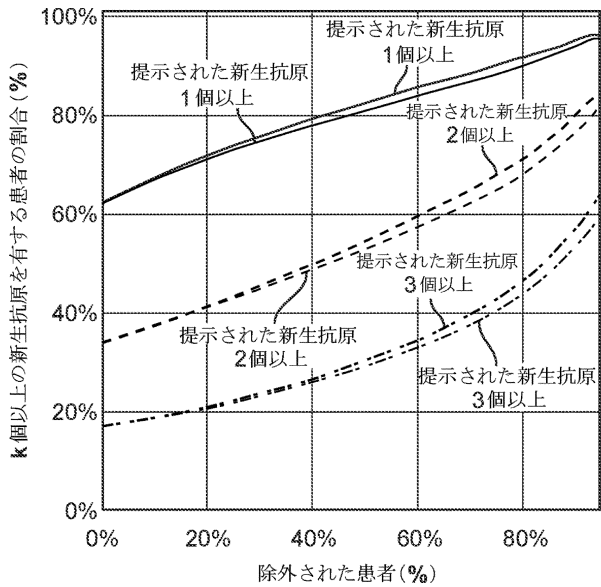
【図 1 3 C】



【図 1 3 D】



【図 1 3 E】



10

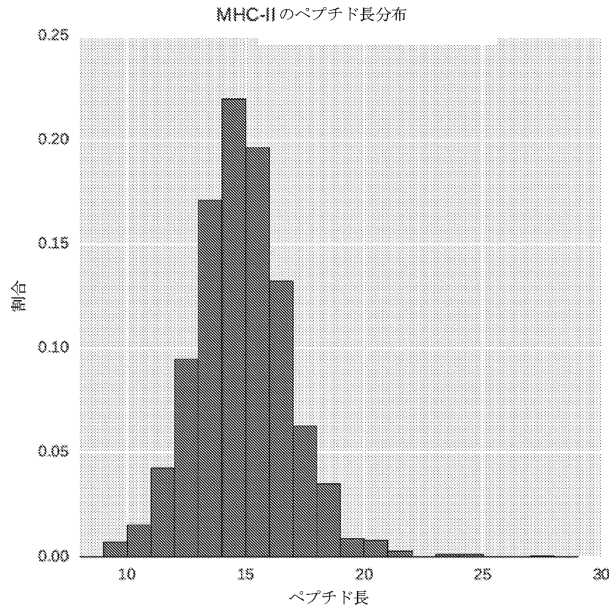
20

30

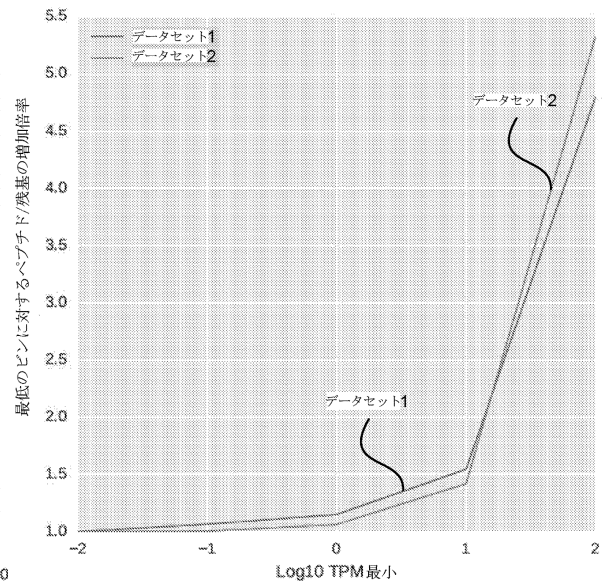
40

50

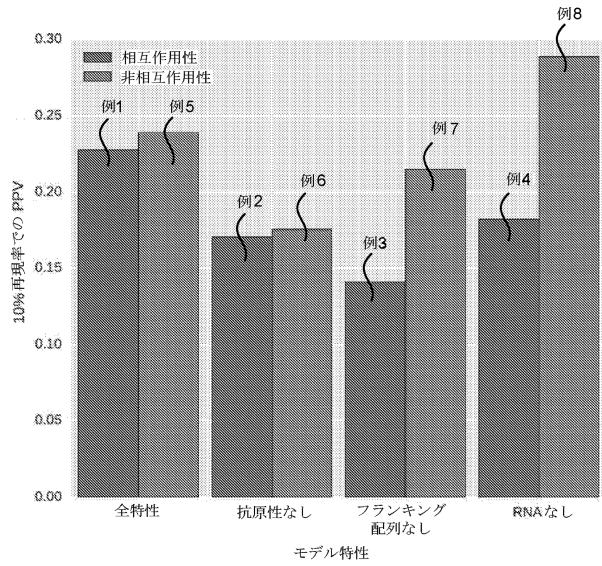
【図 1 4 A】



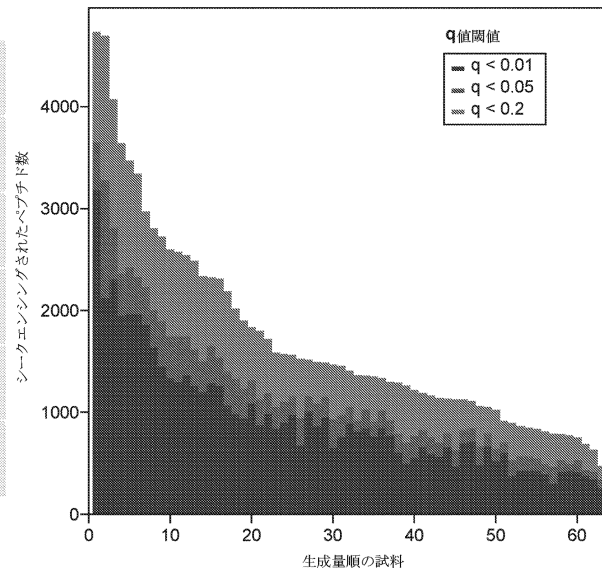
【図 1 4 B】



【図 1 4 C】



【図 1 4 D】



10

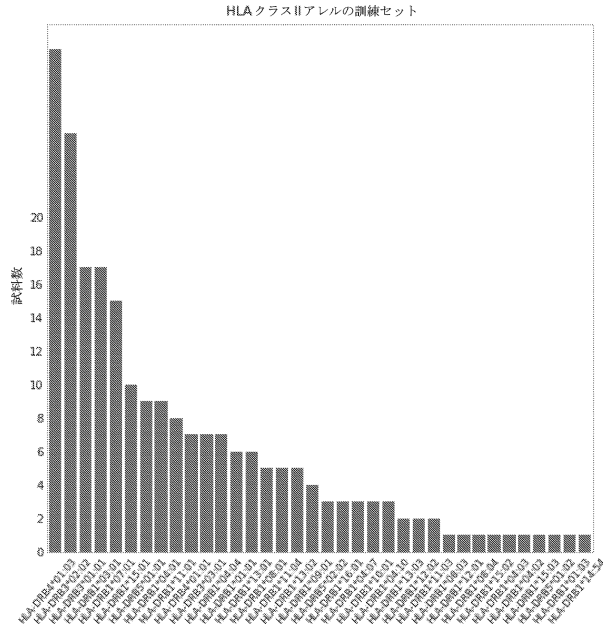
20

30

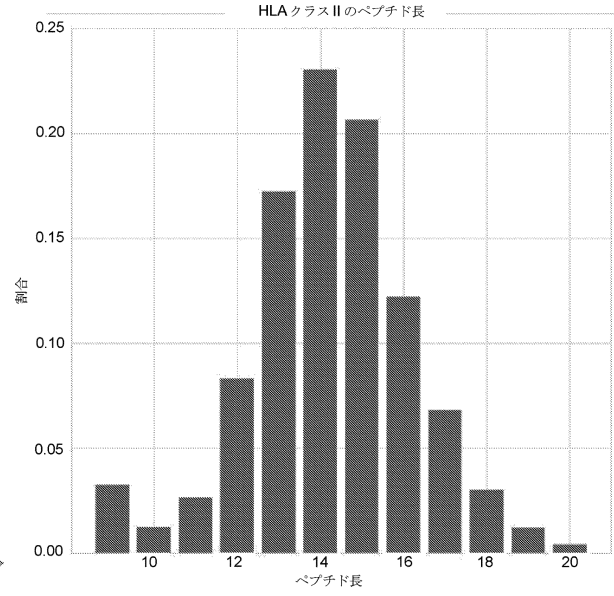
40

50

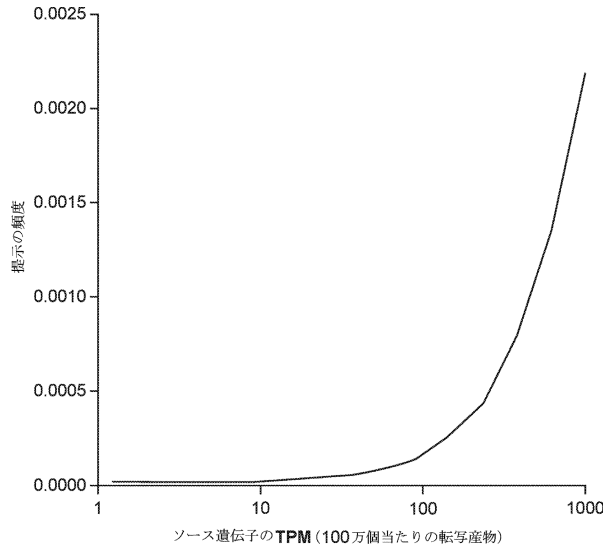
【図 1 4 E】



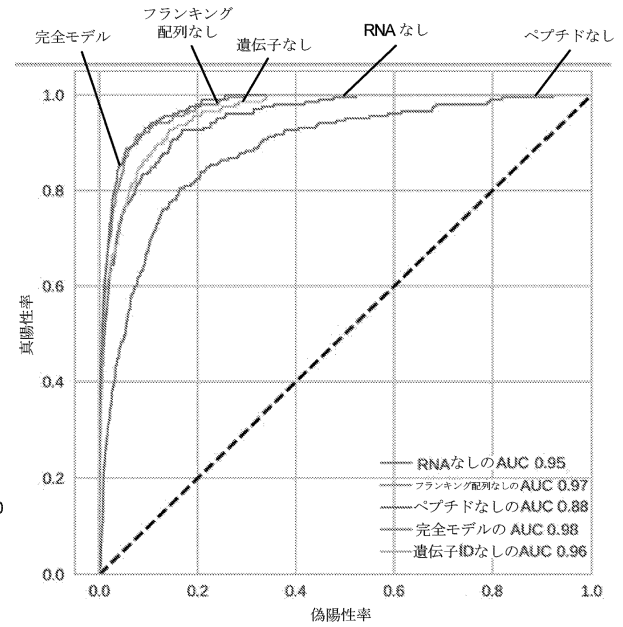
【図 1 4 F】



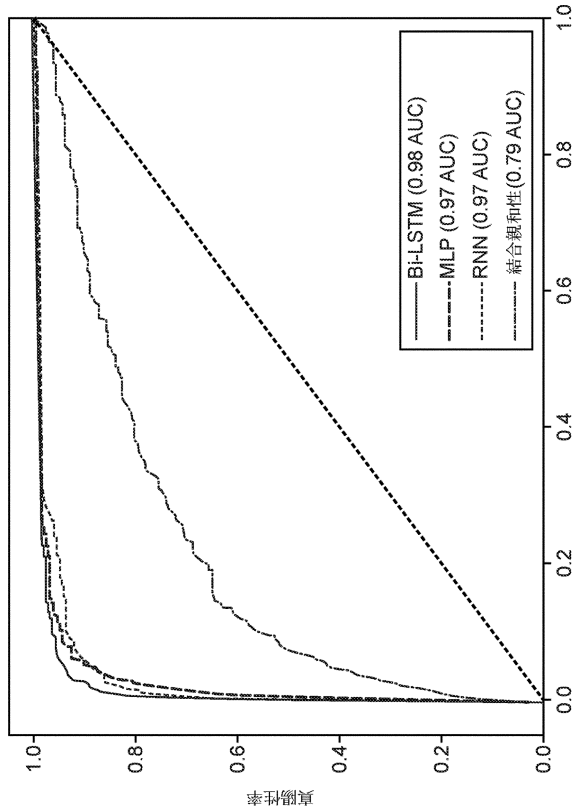
【図 1 4 G】



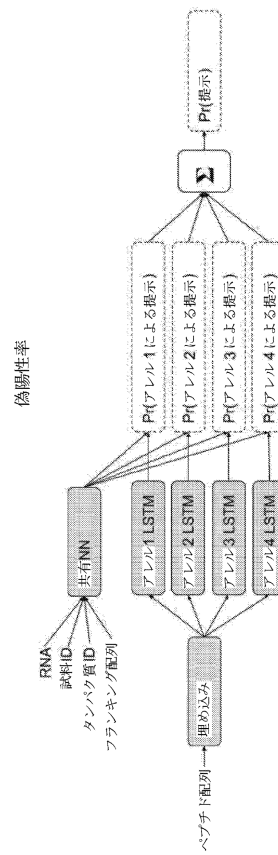
【図 1 4 H】



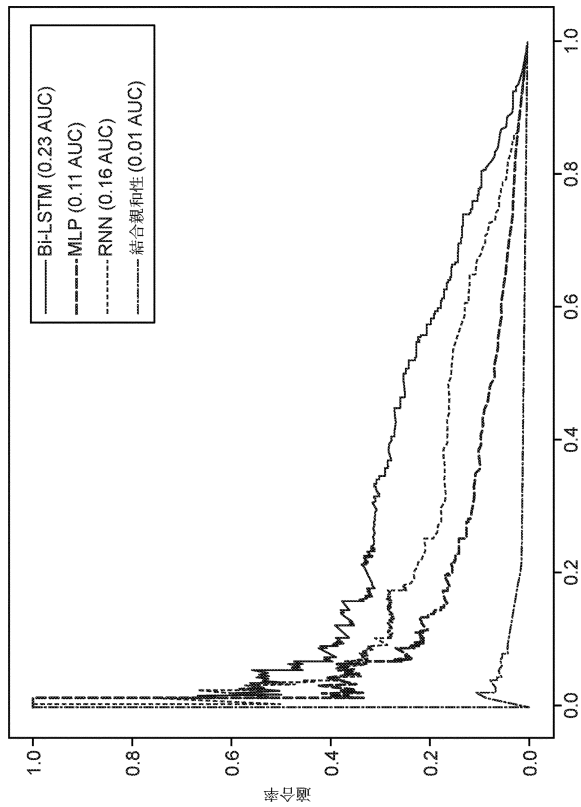
【図 14 I】



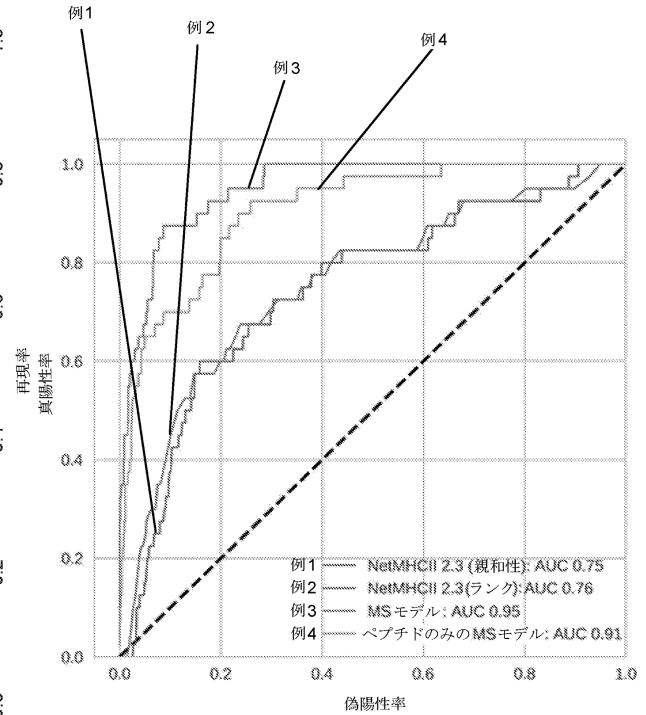
【図 14 J】



【図 14 K】



【図 14 L】



10

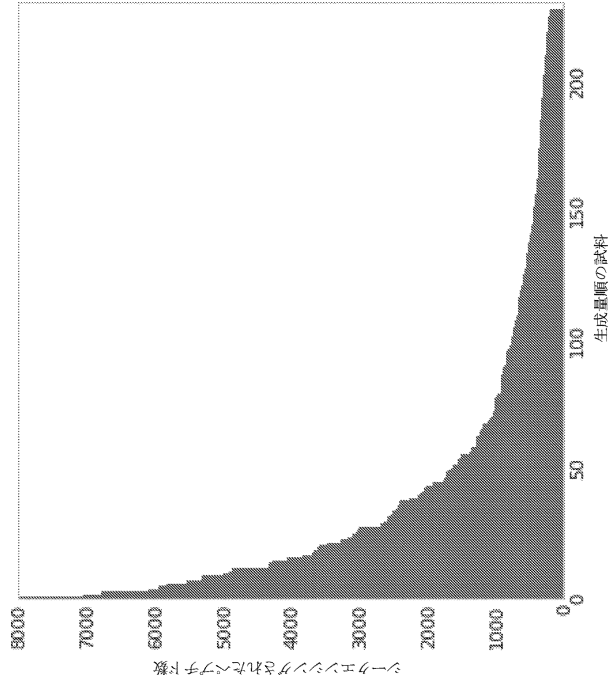
20

30

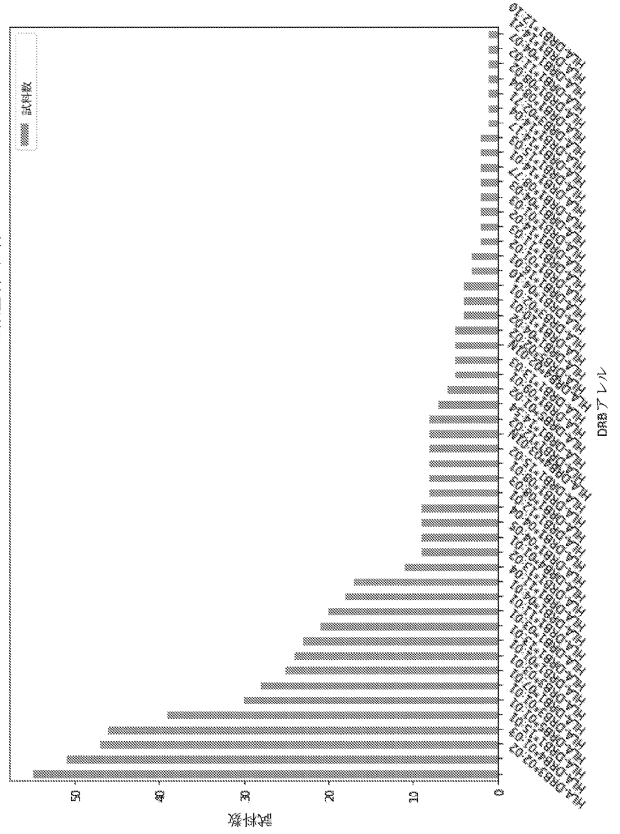
40

50

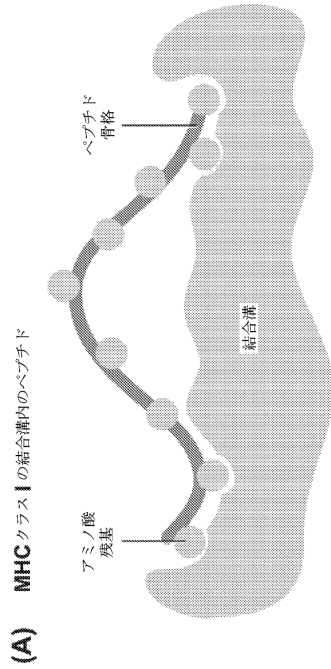
【図 1 4 M】



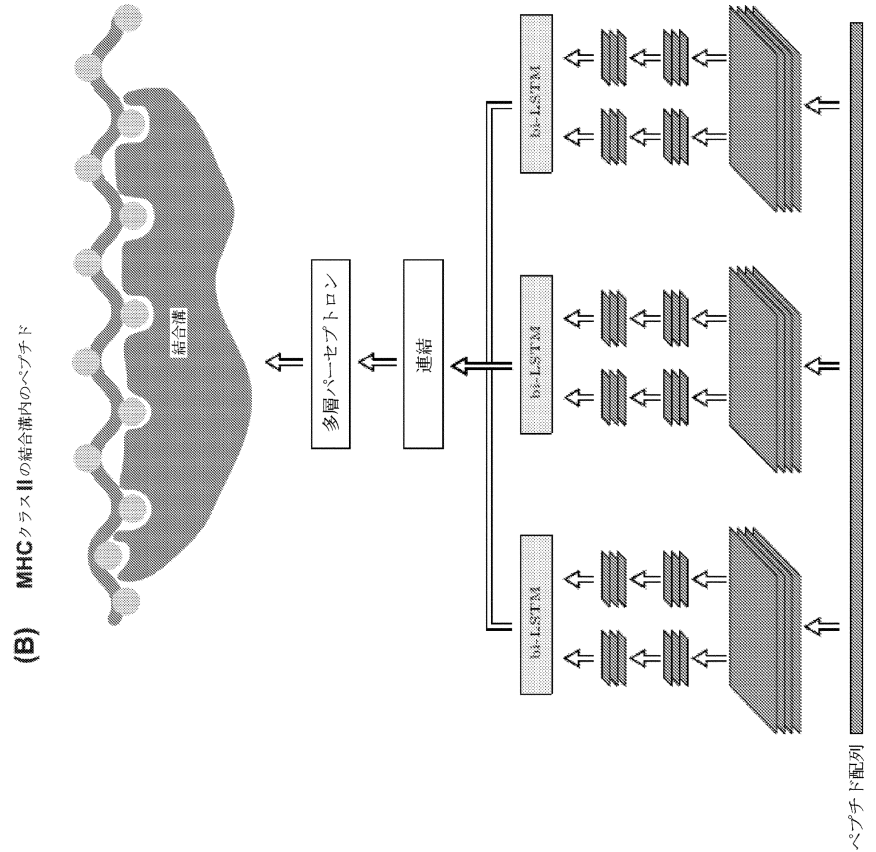
【図 1 4 N】



【図 1 4 O】



【図 1 4 P】



10

20

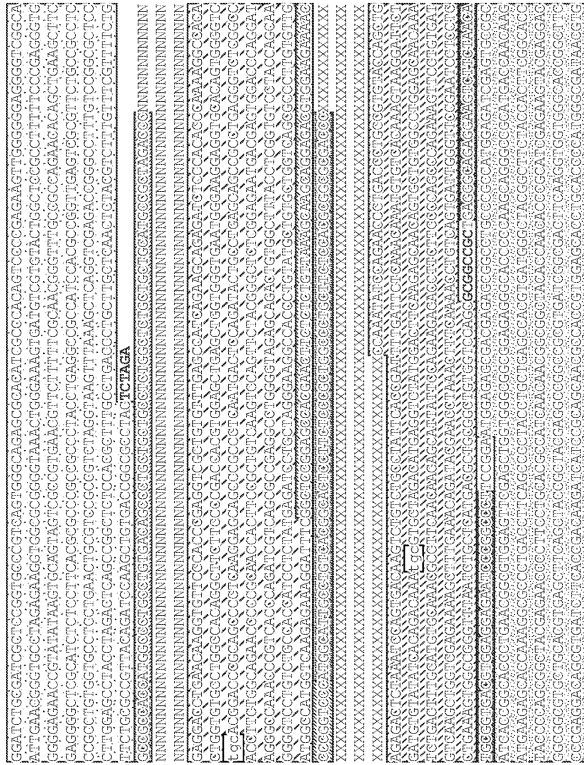
30

40

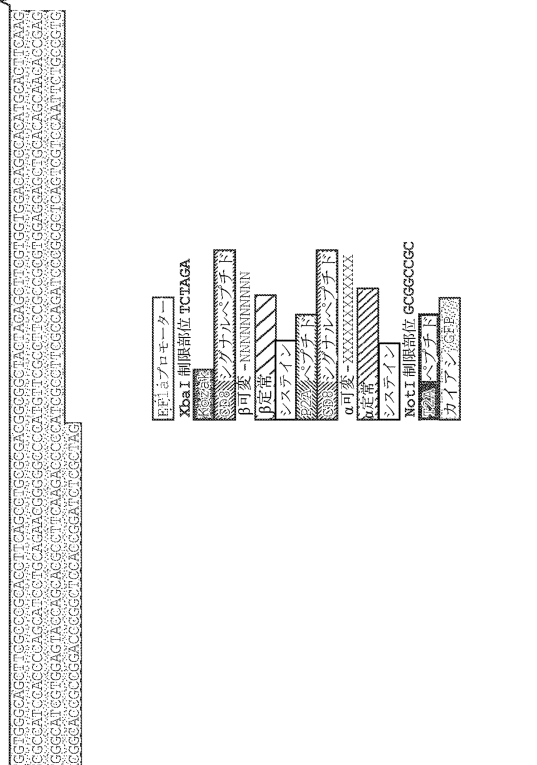
50



【図 17-1】



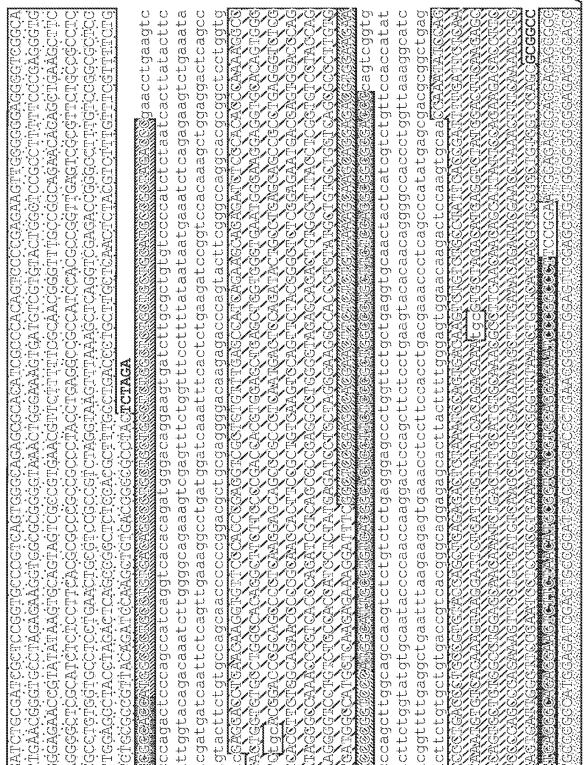
【図 17-2】



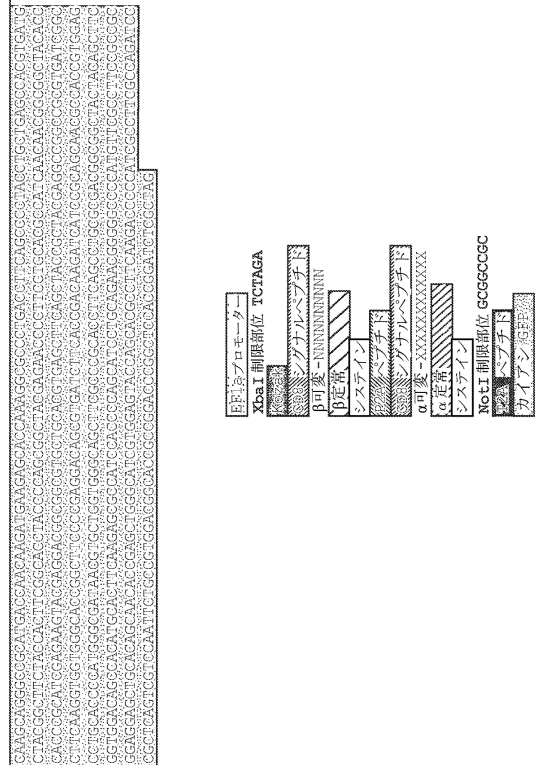
10

20

【図 18-1】



【図 18-2】

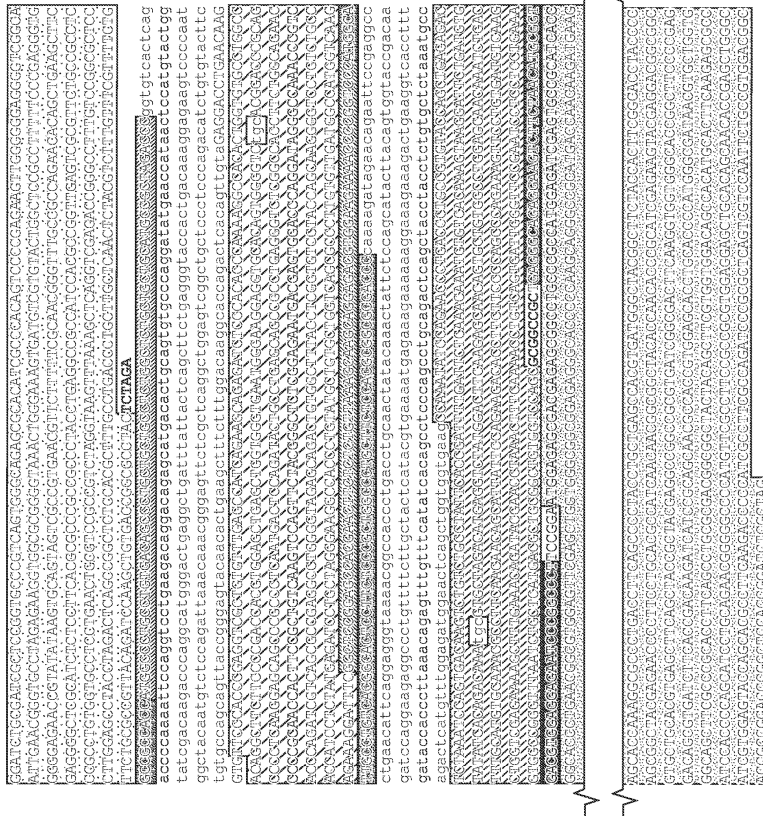


30

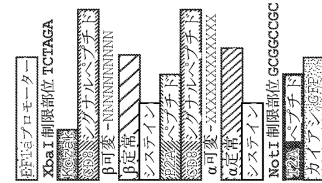
40

50

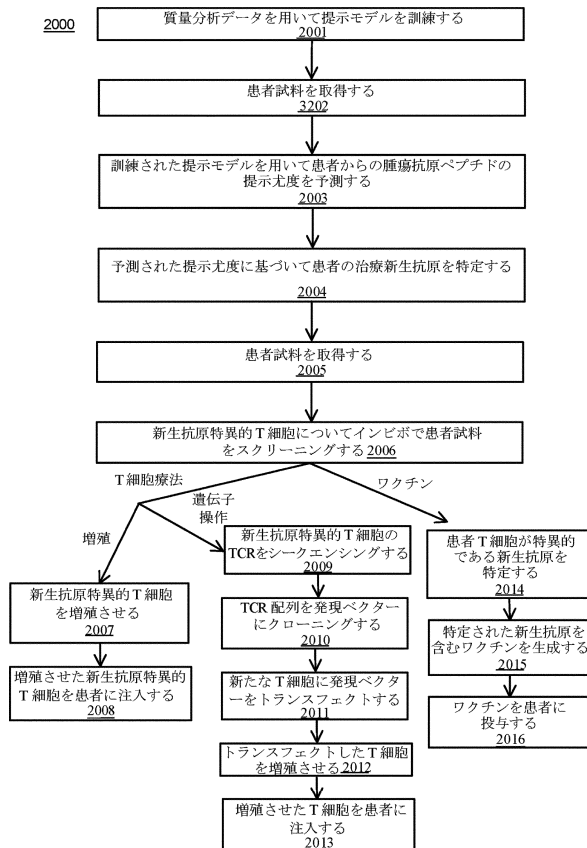
【図 19 - 1】



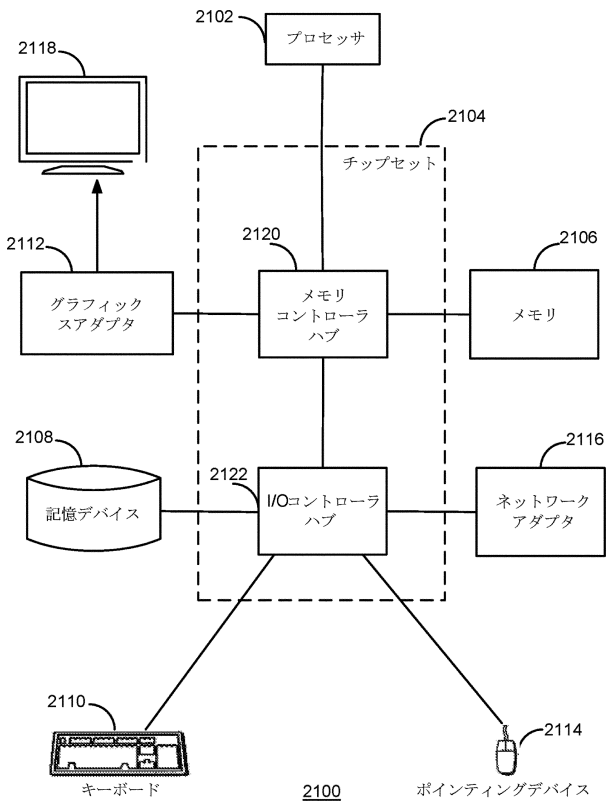
【図 19 - 2】



【図 20】



【図 21】



10

20

30

40

50

【配列表】  
0007668742000001.app

10

20

30

40

50

## フロントページの続き

## (51)国際特許分類

F I

C 0 7 K 14/74 (2006.01)

C 0 7 K 14/74

C 0 7 K 14/82 (2006.01)

C 0 7 K 14/82

G 1 6 B 30/00 (2019.01)

G 1 6 B 30/00

## (33)優先権主張国・地域又は機関

米国(US)

弁理士 佐藤 利光

(74)代理人 100128048

弁理士 新見 浩一

(74)代理人 100129506

弁理士 小林 智彦

(74)代理人 100205707

弁理士 小寺 秀紀

(74)代理人 100114340

弁理士 大関 雅人

(74)代理人 100121072

弁理士 川本 和弥

(72)発明者 バウチャー トーマス フランシス

アメリカ合衆国 9 4 6 0 8 カリフォルニア州 エメリービル ホートン ストリート 5 9 5 9 ス  
イト 3 0 0 グリットストーン バイオ インコーポレイテッド内

(72)発明者 ブリック-サリバン ブレンダン

アメリカ合衆国 9 4 6 0 8 カリフォルニア州 エメリービル ホートン ストリート 5 9 5 9 ス  
イト 3 0 0 グリットストーン バイオ インコーポレイテッド内

(72)発明者 バズビー ジェニファー

アメリカ合衆国 9 4 6 0 8 カリフォルニア州 エメリービル ホートン ストリート 5 9 5 9 ス  
イト 3 0 0 グリットストーン バイオ インコーポレイテッド内

(72)発明者 カオ ミン デュク

アメリカ合衆国 9 4 6 0 8 カリフォルニア州 エメリービル ホートン ストリート 5 9 5 9 ス  
イト 3 0 0 グリットストーン バイオ インコーポレイテッド内

(72)発明者 デイビス マシュー ジョセフ

アメリカ合衆国 9 4 6 0 8 カリフォルニア州 エメリービル ホートン ストリート 5 9 5 9 ス  
イト 3 0 0 グリットストーン バイオ インコーポレイテッド内

(72)発明者 ヤング ローレン エリザベス

アメリカ合衆国 9 4 6 0 8 カリフォルニア州 エメリービル ホートン ストリート 5 9 5 9 ス  
イト 3 0 0 グリットストーン バイオ インコーポレイテッド内

(72)発明者 フランシス ジョシュア マイケル

アメリカ合衆国 9 4 6 0 8 カリフォルニア州 エメリービル ホートン ストリート 5 9 5 9 ス  
イト 3 0 0 グリットストーン バイオ インコーポレイテッド内

(72)発明者 パーマー クリスティーヌ

アメリカ合衆国 9 4 6 0 8 カリフォルニア州 エメリービル ホートン ストリート 5 9 5 9 ス  
イト 3 0 0 グリットストーン バイオ インコーポレイテッド内

(72)発明者 スコベルネ モイツァ

アメリカ合衆国 9 4 6 0 8 カリフォルニア州 エメリービル ホートン ストリート 5 9 5 9 ス  
イト 3 0 0 グリットストーン バイオ インコーポレイテッド内

(72)発明者 イェレンスキー ロマン

アメリカ合衆国 9 4 6 0 8 カリフォルニア州 エメリービル ホートン ストリート 5 9 5 9 ス  
イト 3 0 0 グリットストーン バイオ インコーポレイテッド内

審査官 白井 美香保

(56)参考文献 国際公開第 2 0 1 8 / 1 9 5 3 5 7 (WO, A 1)

特表 2016-521128 (JP, A)

Isolation and characterization of an HLA-DPB1\*04:01-restricted MAGE-A3 T cell receptor for cancer immunotherapy, J. Immunotherapy, 2016年, Vol.39, No.5, p.191-201

Immunogenetics, 2015年, Vol.67, p.641-650

Machine learning methods for predicting HLA-peptide binding activity., Bioinformatics and Biology Insights, 2015年, Vol.9 (S3), p.21-29

Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification, Nature Biotechnology, 2018年12月17日, Vol.37, No.1, p.55-63

(58)調査した分野 (Int.Cl., DB名)

C12Q 1/00-3/00

C12N 1/00-7/08

CAPLUS/BIOSIS/MEDLINE/EMBASE (STN)