

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7155470号  
(P7155470)

(45)発行日 令和4年10月19日(2022.10.19)

(24)登録日 令和4年10月11日(2022.10.11)

(51)国際特許分類	F I
G 1 6 B 20/00 (2019.01)	G 1 6 B 20/00
C 1 2 Q 1/686(2018.01)	C 1 2 Q 1/686 Z
C 1 2 Q 1/6869(2018.01)	C 1 2 Q 1/6869 Z Z N A
C 1 2 N 15/12 (2006.01)	C 1 2 N 15/12

請求項の数 18 (全38頁)

(21)出願番号	特願2020-502531(P2020-502531)	(73)特許権者	520178700 エーシーティー ジェノミックス (アイピー) カンパニー リミテッド 香港 シャーティン ホンコン サイエンスパーク サイエンス パーク ウェスト アベニュー ビルディング 15ダブリュ 8階 803
(86)(22)出願日	平成30年3月31日(2018.3.31)	(74)代理人	100079049 弁理士 中島 淳
(65)公表番号	特表2020-518083(P2020-518083 A)	(74)代理人	100084995 弁理士 加藤 和詳
(43)公表日	令和2年6月18日(2020.6.18)	(72)発明者	ヤン、ベイ - ジア 台湾 114 タイペイ シティ ネイフーディストリクト シンフー セカンド ロード ナンバー 208 4エフ 最終頁に続く
(86)国際出願番号	PCT/US2018/025597		
(87)国際公開番号	WO2018/183980		
(87)国際公開日	平成30年10月4日(2018.10.4)		
審査請求日	令和1年9月30日(2019.9.30)		
(31)優先権主張番号	62/479,320		
(32)優先日	平成29年3月31日(2017.3.31)		
(33)優先権主張国・地域又は機関	米国(US)		

(54)【発明の名称】 免疫原性がん特異的エピトープのためのランク付けシステム

(57)【特許請求の範囲】

【請求項1】

腫瘍サブクローン群における少なくとも1種の免疫原性変異ペプチドを選択するための方法であって、

(a) 複数の変異配列を得ること、

(b) 疾患関連変異に由来する少なくとも1つのエピトープを同定すること、

(c) 前記少なくとも1つのエピトープの免疫原性に関連する複数の因子を取り込むこと、ここで前記複数の因子は、ペプチドレベルスコアを計算するために用いられる第1の因子と、試料レベルスコアを計算するために用いられる第2の因子とを含み、前記第2の因子は下記式により計算されるサブクローン純度 (s)、及び腫瘍純度を含む；

【数1】

$$S = \frac{AF_{\text{観察}}}{M} (pC_t + C_n(1 - p))$$

〔式中、AF<sub>観察</sub>は観察アレル頻度であり、pは腫瘍純度であり、C<sub>t</sub>は腫瘍変異(t)のコピー数(C)であり、C<sub>n</sub>は正常条件(n)のコピー数(C)であり、Mは変異アレル数である。〕、

- ( d ) 前記複数の因子を重み付けること、
- ( e ) 前記複数の因子の重みに基づいて前記少なくとも 1 つのエピトープに免疫原性スコアを割り当てること、
- ( f ) 前記少なくとも 1 つのエピトープをランク付けすること、及び
- ( g ) 前記工程 ( f ) におけるランク付け結果に基づいて少なくとも 1 つの免疫原性変異ペプチドを選択すること、

を含み、

前記免疫原性変異ペプチドが、T細胞応答を引き起こし得る少なくとも 1 つのエピトープを含む、方法。

【請求項 2】

100 個以下のエピトープが選択される、請求項 1 に記載の方法。

【請求項 3】

50 個以下のエピトープが選択される、請求項 1 に記載の方法。

【請求項 4】

30 個以下のエピトープが選択される、請求項 1 に記載の方法。

【請求項 5】

10 個以下のエピトープが選択される、請求項 1 に記載の方法。

【請求項 6】

前記第 1 の因子が、MHC クラス I 及び MHC クラス II による前記少なくとも 1 つのエピトープの提示を対象を含む、請求項 1 に記載の方法。

【請求項 7】

前記選択された少なくとも 1 つのエピトープの MHC クラス I との結合親和性が、1500 nM 未満の IC50 値である、請求項 6 に記載の方法。

【請求項 8】

前記第 1 の因子が MHC クラス I 結合安定性を含む、請求項 6 に記載の方法。

【請求項 9】

前記第 1 の因子が、タンパク質存在量、遺伝子発現、又はそれらの組み合わせを含む、請求項 6 に記載の方法。

【請求項 10】

前記第 1 の因子が、細胞傷害性 T 細胞において免疫応答を引き起こす前記少なくとも 1 つのエピトープの能力を対象を含む、請求項 1 に記載の方法。

【請求項 11】

前記第 1 の因子が、ヘルパー T 細胞における免疫応答を引き起こす前記少なくとも 1 つのエピトープの能力を対象を含む、請求項 1 に記載の方法。

【請求項 12】

前記第 1 の因子が、自己ペプチドに対する前記少なくとも 1 つのエピトープの類似性を含む、請求項 1 に記載の方法。

【請求項 13】

前記第 1 の因子が、既知の抗原に対する前記少なくとも 1 つのエピトープの相同性を含む、請求項 1 に記載の方法。

【請求項 14】

前記変異の変異頻度が少なくとも 10% である、請求項 1 に記載の方法。

【請求項 15】

前記変異の変異頻度が少なくとも 30% である、請求項 1 に記載の方法。

【請求項 16】

前記第 2 の因子は、さらに MHC アレル本数を含む、請求項 1 に記載の方法。

【請求項 17】

前記免疫原性スコアが、MHC クラス I 及び MHC クラス II による前記少なくとも 1 つのエピトープの提示と、ヘルパー T 細胞及び細胞傷害性 T 細胞の両方において免疫応答を引き起こす前記少なくとも 1 つのエピトープの能力と、を含む前記第 1 の因子、並びに

10

20

30

40

50

前記第 2 の因子、が統合されてなる、請求項 1 に記載の方法。

【請求項 1 8】

腫瘍サブクローン群における少なくとも 1 種の免疫原性変異ペプチドを選択するためのシステムであって、

コンピュータにより実行可能であり下記 ( a ) ~ ( g ) の工程を含むプログラムを記憶するハードウェアメモリを含み、

前記免疫原性変異ペプチドが、T 細胞応答を引き起こし得る少なくとも 1 つのエピトープを含む、システム：

( a ) 複数の変異配列を得ること、

( b ) 疾患関連変異に由来する少なくとも 1 つのエピトープを同定すること、

( c ) 前記少なくとも 1 つのエピトープの免疫原性に関連する複数の因子を取り込むこと、ここで前記複数の因子は、ペプチドレベルスコアを計算するために用いられる第 1 の因子と、試料レベルスコアを計算するために用いられる第 2 の因子とを含み、前記第 2 の因子は下記式により計算されるサブクローン純度 (  $s$  )、及び腫瘍純度を含む：

【数 2】

$$S = \frac{AF_{\text{観察}}}{M} (pC_t + C_n(1 - p))$$

〔式中、 $AF_{\text{観察}}$ は観察アレル頻度であり、 $p$ は腫瘍純度であり、 $C_t$ は腫瘍変異 (  $t$  ) のコピー数 (  $C$  ) であり、 $C_n$ は正常条件 (  $n$  ) のコピー数 (  $C$  ) であり、 $M$ は変異アレル数である。〕、

( d ) 前記複数の因子を重み付けること、

( e ) 前記複数の因子の重みに基づいて前記少なくとも 1 つのエピトープに免疫原性スコアを割り当てること、

( f ) 前記少なくとも 1 つのエピトープをランク付けすること、及び

( g ) 前記工程 ( f ) におけるランク付け結果に基づいて少なくとも 1 つの免疫原性変異ペプチドを選択すること。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、免疫原性 T 細胞エピトープを同定し、予測し、ランク付けるシステム及び方法に関する。本開示は、ゲノミクス、次世代配列決定、免疫腫瘍学及び精密医療の分野に関連付けられる。

【背景技術】

【0002】

腫瘍特異抗原を使用することによって腫瘍細胞に対する免疫応答を刺激することががんと闘う際に顕著になってきた。これらの抗原は、腫瘍ゲノミクスと免疫治療における臨床的有益性との間を結びつけるものとして決定されてきた。簡潔に言えば、がん変異を有する遺伝子は、前記変異を有するペプチドを生じる。これらのペプチドは、続いて主要組織適合性複合体 ( MHC ) クラス I 及び主要組織適合性複合体 ( MHC ) II に結合し、抗原として腫瘍細胞表面上に提示される。免疫系、特に細胞傷害性 T 細胞及びヘルパー T 細胞は、これらの抗原を非自己と認識し得、免疫応答を誘発しうる。これらの抗原の内の多くは、腫瘍に限定的なものであり、前もって前記免疫系によって認められるものではない。ゆえに、それらは、処置が正常細胞を損傷することなく前記腫瘍細胞を目標とする免疫治療の適切な標的の役割を果たす。

【0003】

腫瘍特異抗原を用いて T 細胞応答を引き起こすことによって、様々な結果が生じてきた

。これらの抗原を利用することは、以下の2つの障害に直面した。第1の障害は、免疫細胞が、これらの抗原を非自己と認識し、正常細胞を攻撃することなく免疫応答を引き起こすにちがいないということであり、第2の障害は、前記T細胞が前記抗原を非自己と認識する場合であっても、変異腫瘍細胞を含む我々の身体内の細胞には免疫チェックポイントとして知られている安全チェックがあり、それによってT細胞が長期で高振幅攻撃することが防止されるということである。免疫治療の成功ためには、これらの免疫チェックポイントを機能不全にする必要がある。近年、前記第2の障害は注目を浴びている。これらの免疫チェックポイントに対する阻害剤、例えば抗PD1抗体、抗PDL1抗体及び抗CTLA4抗体が開発された後、ますます多くの薬物、臨床試験及び標的がん型が続いた。しかし、免疫チェックポイント阻害剤処置の応答速度がおおよそ20%以下であることから、依然として実質的な改善の余地がある。しかるに、処置前に免疫治療に感受性がある患者を選択することが理想的である。多量の又は質の高い腫瘍特異抗原を有することは、処置応答及び生存率に対する関連性が高いことが示唆されてきた。免疫チェックポイント阻害剤の投与における患者選択を容易にするために、これらの抗原を正確に同定することが非常に要求される。

10

#### 【0004】

対照的に、T細胞認識のための抗原の選択を直接活用することの進捗状況は、比較的停滞したままである。T細胞は、外来の抗原を認識した後の攻撃を引き起こす。しかし、現在のところ、腫瘍特異抗原のT細胞認識のための有効な手法はない。治療手法は、総じて、がんワクチン接種と養子細胞移入とに分類され得る。治療的がんワクチン接種は、ナイーブレパートリから前記T細胞を大きくすること及び存在するT細胞を再活性化することによって、腫瘍進行を遅延させ、腫瘍退縮を促進することを目指す。前記ワクチンは、前記腫瘍特異抗原を含むものであって、これらの抗原が、免疫応答を刺激するそれらの能力について選択されるものである。しかし、ワクチンの進歩は、前記ワクチンを無効なままにする免疫原性抗原の準最適選択によって妨げられてきた。養子細胞移入は、前記免疫細胞を、腫瘍細胞を攻撃するように養成することに直接焦点を当てる。前記免疫細胞、典型的にはT細胞又は樹状細胞は、前記患者から回収され、実験室で養成される。腫瘍特異抗原を認識することによって腫瘍細胞を首尾よく排除する前記T細胞は前記患者に再注入されて戻され、前記T細胞は前記患者の中の腫瘍細胞を攻撃する。しかし、この手法は、低成功率につながる無効な抗原の選択による問題も生じる。両手法において示されるように、免疫原性腫瘍特異抗原の最適選択は、免疫治療における臨床的有益性を達成するために必要である。

20

30

#### 【0005】

免疫原性腫瘍特異抗原を正確に同定するための信頼性のある方法は、各種免疫治療戦略に、広く適用可能であり、且つ決定的に有益である。腫瘍特異抗原の現行の同定は、概して、変異を確認することと、エピトープ、すなわち前記抗原の抗原決定基部分に対するMHCの結合親和性を予測することと、からなる。複数のエピトープ予測ツールが矛盾した結果を伴って存在するが、その後の実験によって、前記予測されたエピトープで検証されたのはおおよそ55%がだけである(Rajasagi Mら、Blood. 2014 Jul 17; 124(3): 453-62)。典型的な方法は、ペプチド配列に基づくものであって、MHC及びそれらの対応する免疫細胞の両クラスを同時に考慮するものではない。さらに、各患者又は各試料は、予測に影響を与える特異的特性を有するが、これらの試料特異的特性は、現在の新生抗原ランク付け方法に組み込まれてこなかった。これらの特性は、アレル本数(dosage of alleles)と総称され得る。多量の変異アレル及びMHCアレルは、免疫系が腫瘍特異抗原を認識する可能性を識別可能に高くするものがあるので、前記エピトープ予測に影響を及ぼす。本開示には、ペプチドレベル情報及び試料レベル情報の両方を用いて正確に免疫原性エピトープを同定し、予測し、ランク付けるシステム及び方法が記載される。ペプチドレベル情報は、MHCクラスI及びII提示、CD4活性化及びCD8活性化を同時に統合し、試料レベル情報は、アレル本数、すなわち変異アレルのクローナリティとMHCアレルの数とを含む。さらに、前記システム及び前記方法

40

50

は、それぞれ、細胞プロセス (cellular processes)、腫瘍特異的特性、抗原提示プロセス及び免疫活性化プロセスにおける重要な構成要素に基づいた因子の包括的リストを取り込む。各因子における重みを伴うエピトープの最適選択が開示される。本開示には、がんワクチン、養子細胞移入又は免疫チェックポイント阻害剤の使用などの、個々に合わせた処置 (personalized treatment) におけるさらなる開発のためのエピトープランク付け方法も記載される。

#### 【発明の概要】

##### 【0006】

本開示には、個体の疾患組織に由来するエピトープを同定し、予測し、ランク付けるシステム及び方法であって、前記エピトープが前記個体内の前記疾患に対する免疫応答を引き起こす、システム及び方法が記載される。前記システムは、ペプチドレベル情報及び試料レベル情報を同時に考慮する。ペプチドレベル情報は、MHCクラスI、MHCクラスII、ヘルパーT細胞活性化及び細胞傷害性T細胞活性化に関連するエピトープ配列を含む各種構成要素に係る計算を含む。試料レベル情報は、変異アレルのクローナリティとMHCアレルの数とを含む腫瘍特異的因子である。前記同定システム及び前記同定方法は、前記因子を統合し、且つ前記因子に重みを指定するものであって、前記重みは、免疫応答を引き起こす際の前記因子の寄与の大きさを表すものである。前記エピトープには免疫原性スコアが付与される。本開示は、前記エピトープをランク付けるためのシステム及び方法をも提供する。前記ランク付けシステム及び前記ランク付け方法は、前記個体に特有のエピトープの優先順位を付けるものであって、個々に合わせた前記疾患の処置のために前記エピトープを利用することができるものである。

##### 【0007】

前記システムは、変異及びコピー数変異の一覧としての変異体情報を有するものであって、前記変異体が次世代配列決定プラットフォームを用いて同定されるものである。前記変異体と共に、配列決定結果の未処理の読み取りを含む前記関連した配列決定情報も提供される。前記システムは、前記変異に関連付けられる主要組織適合性複合体 (MHC) の型をも有する。いくつかの実施形態において、MHCの前記型は、前記変異を呈していたか又は呈していなかった単一の個体のものである。前記システムは、a) 前記変異を含む前記ペプチド配列、b) 各エピトープについてのペプチドレベルスコアであって、前記スコアが、提示され且つ免疫応答を活性化する前記ペプチドの能力を表す、ペプチドレベルスコア、c) 各エピトープについての試料レベルスコアであって、前記スコアが不均一腫瘍 (heterogeneous tumors) 内の変異クローナリティを表す、試料レベルスコア、及びd) 各エピトープのランクであって、前記ランクが、免疫治療処置に有効であると予測されるエピトープに優先的な順位を付けるものである、ランク、を含む一組の変異関連エピトープを出力する。

##### 【0008】

本開示の前記システム及び前記方法は、以下の工程の内のいくつか、又は全てを含む。  
 1) 変異体呼出し (variant calling)、アノテーション、コピー数検出、ヘテロ接合性の消失及び腫瘍純度を含む次世代配列決定データに基づいて前記変異の特性を決定すること、  
 2) 前記変異を有する前記遺伝子の特性を決定すること、  
 3) 前記遺伝子の発現を決定することであって、発現が、公開されている利用可能なりポジトリにおける組織特異的データ及び疾患特異的データに基づく、こと、  
 4) 前記遺伝子のタンパク質存在量を決定することであって、存在量が、公開されている利用可能なりポジトリにおける組織特異的データ及び疾患特異的データに基づく、こと、  
 5) 前記変異を含む前記ペプチド配列を得ること。MHCクラスIについて、前記ペプチドは、長さが8~15のアミノ酸、選好的には長さが8~11のアミノ酸である。MHCクラスIIについて、前記ペプチドは、長さが9~23のアミノ酸である。  
 6) MHCクラスI及びMHCクラスIIへの前記ペプチドの結合を予測すること、  
 7) CD4+T細胞及びCD8+T細胞に対する前記ペプチドの免疫応答の活性化を予測すること、  
 8) 前記ペプチドが適切な抗原提示プロセスを経るかを予測すること、  
 9) 前記ペプチドをそれらの未変異対応物と比較すること、  
 10) 前

10

20

30

40

50

記ペプチドを既知の抗原と比較すること、11) MHCクラスIアレル本数を決定し、MHCクラスI計算に取り込むこと、12) 前記ペプチドレベルにおける重み割り当て及び免疫原性予測のために工程1~11における前記因子を取り込み、統合すること、13) 試料レベルスコアのための前記クローン変異頻度 (clonal mutation frequency for sample-level score) を計算すること、14) ペプチドレベルスコア及び試料レベルスコアを免疫原性スコアとして集合化すること、15) 前記遺伝子のコピー数消失を決定することであって、前記遺伝子の全コピーの消失によって免疫原性スコアが0という低さになること、16) 前記免疫原性に基づいて前記ペプチドをランク付けること。

#### 【0009】

前記エピトープの免疫原性は、以下の因子の内の1つ、1超又は任意の複数の組み合わせに基づく。前記因子としては、i) 変異頻度、ii) コピー数変化、iii) ヘテロ接合性の消失、iv) 腫瘍純度、v) 変異アレルのクローナリティ、vi) 既知の抗原配列との相同性 (抗原相同性)、vii) MHCクラスIについての野生型との類似性 (自己類似性)、viii) MHCクラスIIについての野生型との類似性 (自己類似性)、ix) 遺伝子発現、x) タンパク質存在量、xi) プロテアソーム切断、xii) TAP輸送、xiii) MHCクラスI結合親和性、xiv) MHCクラスII結合親和性、xv) MHCクラスI結合安定性、xvi) MHCクラスIアレル本数、及びxvii) 免疫原性T細胞エピトープのコンセンサス配列マトリックスに対するペプチド類似性、が挙げられるが、それらに限定されるものではない。

#### 【0010】

いくつかの実施形態において、前記個体は、より多量のMHCアレルを含むものであって、ここでより多量とは、前記個体がMHCアレルのホモ接合性対、すなわち各親からのものを有することを意味するものである。より多量のMHCアレルによって、抗原提示における付加的効果がある可能性がある。アレル本数 (allele dosage) の効果は、前記モデルのMHCクラスI計算内に付加される。

#### 【0011】

ペプチドレベルスコアについて、viからxvまでの因子を用いて4つの機械学習モデルが構築される。モデルIは、遺伝子発現、タンパク質存在量、プロテアソーム切断、TAP輸送、MHCクラスI結合親和性、MHCクラスI結合安定性、及びMHCアレル本数を取り込んでMHCクラスI提示を予測する。モデルIIは、MHCクラスII結合親和性を取り込んでMHCクラスII提示を予測する。両モデルは、質量分析法によってMHC-ペプチド結合を測定したデータを用いて訓練される。モデルIIIは、自己類似性及び抗原相同性を取り込んでヘルパーT細胞活性化を予測する。モデルIVは、自己類似性、抗原相同性及びMHCクラスI免疫原性を取り込んで細胞傷害性T細胞活性化を予測する。モデルIII及びモデルIVは、インビトロT細胞免疫応答アッセイ結果を用いて訓練される。これらの4つのモデル及びそれらの組み合わせのいずれかを統合するために、機械学習リグレッサ及び解析的方法が用いられる。前記最後のモデル内の因子は、最適化のための反復モデルチューニングにより特徴選択及び機械学習モデルのシステム内で重み付けされ、前記モデルは、既知の免疫原性エピトープにより検証される。

#### 【0012】

試料レベルスコアについては、因子i~vが考慮される。これらの因子を解析的に算出して、前記変異がクローン変異であるかどうかを決定する。腫瘍は、各サブクローンが別個の遺伝子構造のものである複数のサブクローンを含む可能性がある。クローン変異は、前記クローンの大部分に出現する変異であると定義される。すなわち、これらの変異は、がん変異進化の「幹」において初期に生じる。クローン変異に由来する腫瘍特異抗原は、前記腫瘍細胞の大部分において提示されるので、免疫攻撃の標的となる可能性がある。逆に、小さいサブクローンの変異に由来する腫瘍特異性エピトープは、「葉」変異であり、前記腫瘍細胞の一部に提示されるのみである。これらの腫瘍細胞が攻撃される場合であっても、他のクローンは影響を受けないままである。クローン変異は、期待されるアレル頻度の最大尤度に基づいて変異アレルの数を算出し、次いで変異アレルの数字を用いてサブ

10

20

30

40

50

クローン純度を推定することによって決定される。前記試料レベルスコアは、前記サブクローン純度及び腫瘍純度から算出される。

【0013】

前記免疫原性スコアは、前記ペプチドレベルスコア及び前記試料レベルスコアの両方を取り込む。次いで、各ペプチドについての前記免疫原性スコアは、前記統合同定システムの最終出力としてランク付けられる。

【図面の簡単な説明】

【0014】

【図1】前記システムのフローチャート。前記システムにおける全体的プロセス及び各主要工程を記載する。

10

【0015】

【図2】入力及び出力。システムは、入力として次世代配列決定プラットフォームから生成される変異体情報及び試料情報を受け取る。システムは、MHC情報も受け取る。システムは、アミノ酸配列と、ペプチドレベルスコアと、試料レベルスコアと、免疫原性エピトープのランクとを出力する。

【0016】

【図3】T細胞特徴による免疫応答の比較。免疫応答するペプチドと無応答のペプチドとの、ペプチドレベルスコアの分布。A)実施例5は、CD4+細胞についての自己類似性及び抗原相同性を考慮する。B)実施例6は、CD8+細胞についてのT細胞免疫原性を考慮する。p値は、ウィルコクソン順位和検定によって算出される。

20

【0017】

【図4】抗原提示特徴及びT細胞特徴による免疫応答の比較。免疫応答するペプチドと無応答のペプチドとの、ペプチドレベルスコアの分布。A)実施例7は、CD4+についての実施例4における自己類似性、抗原相同性及び特徴を考慮する。B)実施例8は、CD8+についての実施例1におけるT細胞免疫原性及び特徴を考慮する。p値は、ウィルコクソン順位和検定によって算出される。

【0018】

【図5】抗原予測スコア及びT細胞特徴による免疫応答の比較。免疫応答するペプチドと無応答のペプチドとの、ペプチドレベルスコアの分布。A)実施例9は、CD4+についての実施例4における前記モデルの自己類似性、抗原相同性及び予測スコアを考慮する。B)実施例10は、CD8+についての実施例1における前記モデルのT細胞免疫原性及び予測スコアを考慮する。p値は、ウィルコクソン順位和検定によって算出される。

30

【0019】

【図6】抗原提示情報及び両T細胞特徴による免疫応答の比較。A)実施例5及び実施例6を考慮する実施例11。B)実施例7及び実施例8を考慮する実施例12。C)実施例9及び実施例10を考慮する実施例13。

【0020】

【図7】トップ50のペプチドレベルスコアにおける応答ペプチド。棒は、CD8+に対して応答することが実験的に確認され、且つ各組についてのトップ50のペプチドレベルスコアにランク付けられたペプチドの数を示す。各プロットは、特定の患者を表す。破線は、前記患者における応答ペプチドの総数を表す。

40

【発明を実施するための形態】

【0021】

ある実施形態において、本開示には、疾患特異的エピトープを同定し、前記エピトープの免疫原性を予測し、精密医療の壮大なスキームにおける個体のさらなる個々に合わせた処置のために前記エピトープをランク付ける、統合されたシステム及び方法が記載される。前記システム及び前記方法は、免疫原性エピトープの正確で実用的な同定のための、配列に基づいた変異体呼出し、配列に基づいたコピー数の決定、配列アラインメント、類似性マトリックス、機械学習、最適化及び数学的モデル化を統合する(図1)。前記システム及び前記方法は、細胞プロセス、腫瘍特異的特性、抗原提示プロセス及び免疫活性化プ

50

プロセスを構築する各構成要素を考慮する。各プロセスにおける前記構成要素は、前記細胞における前記構成要素の実際の機能に従い因子として計算される。各因子は、前記エピトープの免疫原性に対する寄与の大きさによって重み付けされる。重み付けされた因子によって、免疫原性となる理由に対する検査が可能になり、その点で臨床現場及び研究現場の両方における検査が容易になる。前記因子は、前記システムによって全て考慮されるペプチドレベル情報及び試料レベル情報を表すものであって、前記システムがこれらの因子を利用して前記エピトープの免疫原性のスコアを付けるものである。本開示は、前記同定されたエピトープをランク付けるためのスキームをも提供する。前記ランク付けは、前記エピトープの予測された免疫原性スコアに基づく。

【0022】

本開示において用いられる用語は、実施形態及び請求項を記載する目的のために理解されるべきである。前記用語の時制及び語幹のいかなる変化も本開示を限定するものではないことを理解すべきである。一般に理解される前記用語のいかなる同義語も、本開示を限定するために用いられるものではないことが理解されるべきである。

【0023】

本開示は、本明細書に記載される特定の метод論、プロトコル又は手法に限定されるものではないが、その理由は、これらが変化する可能性があるためである。本明細書に記載される特定の実施形態は、単に例であって、本開示の範囲を限定するものと解釈されるべきではない。

【0024】

本開示において用いられる場合、単数形の「1つの(a)」、「1つの(an)」及び「前記(the)」は、内容が明確に示されない限り、具体的には、それらが指す用語の複数形をも包含する。

【0025】

「構成要素」という用語は、前記変異における特定の特性、前記遺伝子における特定の特性、前記細胞プロセスにおける特定の工程、又は前記試料の特定の特性を指す。

【0026】

「因子」という用語は、因子が、式によって算出され得るか、計算ツールによって予測され得るか、又はカテゴリとして層別化され得る、因子の計算表現を指す。

【0027】

「ペプチド」という用語は、様々な長さのアミノ酸配列を指し、免疫原性であっても免疫原性でなくてもよく、腫瘍関連であっても腫瘍関連でなくてもよい。「抗原」という用語は、免疫原性であり、且つ免疫系によって認識され得るペプチドを指す。「エピトープ」という用語は、細胞の表面に提示され得る短い抗原を指す。前記エピトープは、より長い抗原のプロテアソーム切断によって生成され得る。

【0028】

「がんワクチン」という用語は、がんを、前記がんに対する身体の免疫系を高めることによって処置することを目指す治療ワクチンを指す。前記治療ワクチンは、一般に投与される予防ワクチンと混同してはならず、一般に投与される前記予防ワクチンは、予防のために疾患の前に投与されるものである。

【0029】

「主要組織適合性複合体」という用語及びその略語である「MHC」は、前記MHCの任意の変形物及び名称を指し、前記変形物及び前記名称としては、前記MHCのクラス、代替的名称、例えばヒト白血球抗原(HLA)、型、例えばA、B、C、DRB1、DPA1、DPB1、DQA1、DQB1などが挙げられるが、それらに限定されるものではない。

【0030】

「変異」という用語は、特に明記しない限り、ミスセンス変異、フレームシフト変異及びプライス部位変異を含む非同義体細胞変異を指す。「変異体」という用語は、変異を含むが、さらに、コピー数変異、染色体再配置、融合、転座及び逆位を含む構造変異を含

10

20

30

40

50

む。体細胞変異体は、生殖系列には存在せず、後の生存において、特にがん発生の間に生じた変異体と定義される。これらの変異体は、腫瘍化につながる可能性があるか、又はがんと共に進行するパッセンジャである。

【 0 0 3 1 】

「総深さ」という用語は、遺伝子内の特定の位置で配列決定される読み取りの総数である。

【 0 0 3 2 】

いくつかの実施形態において、前記システム又は前記方法は、入力として N G S 配列決定データを受け取る ( 図 2 )。N G S データは、変異体呼び出しフォーマット ( V C F )、配列アラインメントマップ ( S A M )、バイナリーアラインメントマップ ( B A M )、F A S T Q 又は他の未処理の若しくは処理されたファイルフォーマットの中にあることができる。前記 V C F ファイルは、ゲノム内の位置に関する情報を含む。前記情報としては、変異アレル、参照アレル、染色体、染色体位置、アレル頻度及び総深さが挙げられるが、それらに限定されるものではない。いくつかの実施形態において、ユーザは、コピー数変異、腫瘍純度及びヘテロ接合性の消失を含む、アレルの大きなセグメントの変化に由来する情報をも提供することになる。いくつかの実施形態において、前記システムは、S A M / B A M ファイルを受け取るものであって、上述のデータが全て前記 S A M / B A M ファイルから計算され得るものである。いくつかの実施形態において、前記システムは、F A S T Q ファイルを受け入れるものであって、上述のデータが参照ゲノムに対する配列アラインメントを行った後で計算され得るものである。

【 0 0 3 3 】

前記システムは、入力として M H C の型をも受け取る ( 図 2 )。いくつかの実施形態において、前記 M H C クラス I は、A、B 及び C が挙げられるがそれらに限定されるものではない上位型であり、少なくとも 4 桁の分解能で型決めされる。いくつかの実施形態において、前記 M H C クラス II は、D R B 1、D P A 1、D P B 1、D Q A 1、D Q B 1 が挙げられるがそれらに限定されるものではない上位型であり、少なくとも 4 桁の分解能で型決めされる。いくつかの実施形態において、前記 M H C 型は、本明細書に記載される前記 N G S データから得られ得る。

【 0 0 3 4 】

本開示には、前記エピトープの同定と、その免疫原性の予測とが記載されており、本開示は、以下の因子の内の 1 つ、1 超又は任意の複数の組み合わせに基づく。i ) 変異体呼び出しによって決定される変異の変異頻度、i i ) コピー数変化、i i i ) 変異についてのヘテロ接合性の消失 ( L O H )、i v ) 腫瘍純度、v ) 変異アレルのクローナリティ、v i ) 配列アラインメントによって決定される既知の抗原との相同性、v i i ) M H C クラス I に対する変異ペプチドの M H C 結合親和性と野生型ペプチドの M H C 結合親和性との比によって決定される、野生型ペプチドとの類似性、v i i i ) M H C クラス II に対する変異ペプチドの M H C 結合親和性と野生型ペプチドの M H C 結合親和性との比によって決定される、野生型ペプチドとの類似性、i x ) 公開されているデータベースから得うる、組織特異的実験及び疾患特異的実験から決定される遺伝子発現、x ) 公開されているデータベースから得うる、組織特異的実験及び疾患特異的実験から決定されるタンパク質存在量、x i ) プロテアソーム分解データによって決定されるプロテアソーム切断、x i i ) 輸送速度データによって決定される T A P 輸送、x i i i ) インビトロ結合アッセイによって決定される M H C クラス I 結合親和性、x i v ) インビトロ結合アッセイによって決定される M H C クラス II 結合親和性、x v ) M H C クラス I 結合安定性、x v i ) M H C クラス I アレル本数、x v i i ) インビトロ T 細胞増殖アッセイ又はエクスピボ T 細胞増殖アッセイによって決定される M H C クラス I 免疫原性。

【 0 0 3 5 】

エピトープは、抗原提示プロセスによって細胞の表面に出現する。がん特異的エピトープの場合、遺伝子変異によって変異ペプチドが得られ、次いで、前記変異ペプチドは、プロテアソームによって短いエピトープに切断され、次いで、T A P タンパク質によって小

10

20

30

40

50

胞体の中に輸送される。小胞体の内部において、前記エピトープはMHC複合体に結合する。次いで、前記MHC複合体と共に、前記エピトープは、免疫細胞認識のために前記細胞表面上に提示される。この抗原提示プロセスにおける各工程は、前記エピトープの免疫原性に寄与する。

#### 【0036】

腫瘍における変異は、ひとつひとつの腫瘍細胞内で生じなくてもよい。免疫原性エピトープを生じさせる変異が腫瘍細胞のより大きい部分で生じる場合、免疫細胞は、前記腫瘍細胞のほとんどを標的として認識するので、前記腫瘍を消滅させる可能性が高い。ゆえに、0~100%の変異頻度によって表される腫瘍内で検出される変異の割合は、エピトープの免疫原性の決定における重要な一面である。変異頻度がより高いことは、変異が、腫瘍のより大部分の中にある可能性があり、したがって免疫細胞攻撃の有効性に影響を与える可能性があることを表す。同様に、コピー数変化、ヘテロ接合性の消失(LOH)、腫瘍純度及び変異アレルのクローナリティを含む変異の他の特性は、免疫細胞攻撃につながる可能性がある変異エピトープを生成する腫瘍細胞の比率を反映する。

10

#### 【0037】

エピトープが由来する遺伝子は、発現する必要がある。腫瘍試料における遺伝子発現は、NGS(例:RNA-seq)、マイクロアレイ、定量的リアルタイムPCR又はノーザンブロットを用いて測定され得る。組織特異的遺伝子発現及びがん特異的遺伝子発現は、公開されている利用可能なデータベースからも得られ得る。公開されている利用可能なデータベースからデータを利用することによって、実際に発現する遺伝子の決定が可能になる。遺伝子発現についての現在の理解は、ゲノムが全面的に転写されるということである。転写調節プロセスが、複雑で、且つ入り組んで制御される場合であっても、少量の発現がなお検出され、あまり調節されない。ゆえに、個別的な遺伝子発現データには多くのノイズが存在する。同一疾患を有する様々な個体にわたって同じ組織の中である遺伝子が発現する場合、それは、前記遺伝子が概してその細胞型において転写されることを示す。エピトープ形成のために翻訳が起こるように前記遺伝子が発現する必要がある。いくつかのデータセットにおいて、遺伝子発現は、質的に、低、中、高又はなしと決定される。これらのデータセットにおいて、遺伝子発現値は、0、1、2及び3などの数値に変換され得る。他のデータセットにおいて、遺伝子発現は、比又は任意の単位などの各種の単位を有する数値であることができる。いくつかの実施形態において、前記モデルは、数値又は変換された数値を統合機械学習モデルに取り込む。発現しない遺伝子は、低、0又はなしと表される。別の実施形態において、前記システムは、発現しない遺伝子をフィルタリングで除く。逆に、各発現検出方法によって高度と考えられる遺伝子発現は、エピトープの量に寄与する。高量のエピトープは、MHC複合体に遭遇する可能性がより高いので、提示される可能性が高い。

20

30

#### 【0038】

同様に、タンパク質存在量情報は、質量分析法、免疫蛍光法、免疫組織化学又はウェスタンブロットによって測定され得る。タンパク質存在量データは、パブリックドメインからも得うる。前記腫瘍特異性エピトープを有する前記タンパク質の量は、エピトープがMHCに結合する可能性を決定することに役立つ。エピトープは、免疫原性が非常に高い可能性があるが、量が少ない可能性がある。この状況下で、かかるエピトープは、免疫応答を引き起こすにおいて有効ではない。いくつかのデータセットにおいて、タンパク質存在量は、質的に、低、中、高又はなしと決定される。これらのデータセットにおいて、タンパク質存在量値は、0、1、2及び3などの数値に変換され得る。他のデータセットにおいて、タンパク質存在量は、比又は任意の単位などの各種の単位を有する数値であることができる。いくつかの実施形態において、前記モデルは、数値又は変換された数値を統合機械学習モデルに取り込む。発現しないタンパク質は、低、0又はなしと表される。別の実施形態において、前記システムは、タンパク質存在量に関するスコアにゼロを割り当てるか、又は発現しないタンパク質をフィルタリングで除く。逆に、各存在量検出方法によって高度と考えられるタンパク質存在量は、エピトープの量に寄与する。高量のエピトープ

40

50

プは、前記MHC複合体に遭遇する可能性がより高いので、提示される可能性が高い。

【0039】

変異していない野生型ペプチドに対する変異ペプチドの類似性が決定される。変異ペプチドが野生型ペプチドと類似する場合、それは自己と認識される可能性があり、T細胞によって許容される可能性がある。変異ペプチドの、変異していない野生型配列に対しての類似性は、MHCに対するそれらの結合親和性の差を用いて算出され得る。前記差は、前記野生型ペプチドの結合親和性に対しての前記変異ペプチドの結合親和性の比が、MHCクラスI及びMHCクラスIIの両方について算出される、差の数値尺度に翻訳される。

【0040】

既知の抗原に対する変異ペプチドの相同性が決定される。既知の抗原は、通常はT細胞が免疫応答を引き起こす細菌、ウイルス及び他の病原体に由来する。ゆえに、既知の抗原と非常に類似している変異ペプチドは、免疫原性を引き起こす可能性が高い。抗原相同性は、全ての既知の抗原に対しての、変異ペプチドの配列アラインメントによって決定されるものであって、配列の同一性と同一性の長さとが考慮されるものである。相同性は、抗原と同じ配列を含む変異ペプチドの割合として決定され得る。

10

【0041】

エピトープは、小胞体の中でMHC複合体に遭遇する。小胞体の内部にあるためには、変異ペプチドは、プロテアソームによって適切な長さのエピトープに切断される必要がある。プロテアソームによって切断され得るペプチド内の部位は、0～1のスコアであると予測できる。切断される可能性があるエピトープが提示前に分解するかもしれないので、エピトープの中の切断可能部位は最小又は0であることが好ましい。次いで、前記エピトープは、TAPタンパク質によって小胞体の中に輸送される必要がある。TAPによる前記エピトープの輸送効率性は、IC50値として表され、ここでIC50がより小さいと輸送がより良好であることを示す。容易に輸送されるエピトープは、MHC複合体に遭遇する可能性がより高い。

20

【0042】

エピトープは、提示のためにMHC複合体に結合することが必要である。MHCクラスIは、8～15のアミノ酸又はそれより長い、好ましくは8～11のアミノ酸のエピトープに結合してもよい。MHCクラスIIは、9～23のアミノ酸又はそれより長い、好ましくは15及び16のアミノ酸のエピトープに結合してもよい。エピトープは、MHC上の、各型のMHCに特異的な位置に固定される。これらの固定位置上の特異的アミノ酸は、結合の親和性を示すので提示にとって重要であるが、他の位置にあるアミノ酸も結合親和性に影響を及ぼす。1500nM未満又は1000nM未満、好ましくは500nM未満のIC50によって示されるクラスI及びクラスIIの両方に対する強固なMHC結合親和性は、エピトープがMHC複合体に結合する可能性があり、細胞の表面上に提示されることを表す。

30

【0043】

結合親和性に加えて、MHCに対するエピトープ結合の安定性も、抗原提示において重要である。エピトープは、MHC複合体との非常に強固な親和性を有する可能性があるが、エピトープがMHCに結合したままとなる時間は、提示にとって十分に長いものではない。エピトープが提示前にMHC複合体から解離する場合、前記エピトープは細胞表面上に現れない。0～1の半減期スコアによって示されるMHC結合安定性は、エピトープがMHCに結合したままとなる時間を表す。より長い期間結合していることによって、エピトープが細胞の前記表面上に提示される可能性が増加する。

40

【0044】

MHCクラスI免疫原性は、免疫応答の引き起こしにおけるエピトープ配列の特異的組成(specific composition)の能力を表す。特異的エピトープ配列は、細胞障害性T細胞の活性化T細胞受容体(TCR)に対して生化学的な影響を及ぼす可能性がある。TCRを誘発することは、T細胞攻撃における最初の工程である。-1～1の範囲のスコアによって表される高MHCクラスI免疫原性は、T細胞がエピトープに接触すると増殖する

50

可能性があることを示す。

【 0 0 4 5 】

いくつかの個体は、両親に由来する同じ型のMHCアレル、すなわちMHCのホモ接合性対を有する。ホモ接合性MHCアレルは、アレル本数によって付加された効果をもたらす可能性がある。ホモ接合性MHCアレルは、可能なエピトープが結合するアレル本数を増大させるので、細胞表面に前記エピトープを提示する可能性を識別可能に増大させる。さらに、細胞表面上の高量のMHCも、T細胞がエピトープを認識する可能性を増大させる。ゆえに、MHCクラスIの計算において、付加された効果が考慮される。

【 0 0 4 6 】

免疫治療処置の下で個体から免疫応答情報を用いることを含む上記の方法の内のいずれかによれば、免疫応答の欠如は、抗原提示機構に關与するものに欠陥があることによる可能性がある。これらの欠陥によって、抗原提示経路は行き詰ったものになり、その結果、エピトープは、前記エピトープが免疫原性であるにもかかわらず細胞の表面に提示されなくなる。これらの状況は、免疫治療に対する応答の理由を決定する際の因子を混乱させるものとして提示される。ゆえに、抗原提示機構の欠陥を有する個体は考慮されない。

10

【 0 0 4 7 】

上記の方法の内のいずれか1つによれば、各因子は、前記同定システムに基づいて重み付けされるものであって、前記同定システムは、特徴選択、機械学習、検証、及び最適化のための反復モデルチューニングを含むものである(図1)。特徴選択の後で用いられる前記特徴は、以下のものとして列挙される。

20

【 0 0 4 8 】

ペプチドレベル特徴

【 0 0 4 9 】

MHC I 提示：遺伝子発現、タンパク質存在量、プロテアソーム切断、TAP輸送、MHCクラスI結合親和性、MHCクラスI結合安定性

【 0 0 5 0 】

MHC II 提示：MHCクラスII結合親和性

【 0 0 5 1 】

ヘルパーT細胞活性：自己類似性、抗原相同性

【 0 0 5 2 】

細胞傷害性T細胞活性：自己類似性、抗原相同性、MHCクラスI免疫原性

30

【 0 0 5 3 】

試料レベル特徴

【 0 0 5 4 】

変異アレルのクローナリティ、MHCクラスIアレル本数

【 0 0 5 5 】

ペプチドレベルスコアについては、前記それぞれの特徴を用いた4つのモデル、すなわちMHCクラスI、MHCクラスII、ヘルパーT細胞活性化、及び細胞傷害性T細胞活性化が構築される。さらに、MHCクラスIは、そのモデルへ試料レベル特徴、MHCクラスIアレル本数を取り込む。これらの4つのモデル及びそれらの組み合わせのいずれかを統合するために、機械学習リグレッサの組み合わせを用いる。最適化のための反復モデルチューニングでペプチドレベルモデルを訓練し、既知の免疫原性エピトープで前記モデルを検証する。次いで、乗算などの数学的方法及び解析的方法で任意の2つ以上の機械学習リグレッサを統合して、最終的なペプチドレベルスコアを得る。

40

【 0 0 5 6 】

試料レベルスコアについては、変異アレルのクローナリティを算出する。クローン変異又は「幹変異」は、腫瘍発生の早期の段階の間に生じた変異と定義されるので、枝で生じる変異とは対照的に、がん変異進化の樹の幹に属する。クローン変異は、概して、腫瘍のクローンのほとんどの中に存在する。クローン変異の決定は、期待されるアレル頻度と観察されたアレル頻度との統計的有意性に基づいて変異アレルの期待数を割り当て、次いで

50

変異アレルの数を用いてサブクローン純度を推定することに基づく。サブクローン純度及び腫瘍純度の部分は、試料レベルスコアである。

【0057】

免疫原性スコアは、ペプチドレベルスコア及び試料レベルスコアが統合されたものである。該モデルは、パラメーターチューニングの複数の反復によってチューニングされ、複数回、再構築され、訓練されたものであり、前記反復のそれぞれにおいて、選択される特徴又は重みは変わる。最良の性能を有するモデルが最終モデルである。各エピトープに免疫原性スコアが与えられる。免疫原性スコアの大きさの順序は、前記統合システムによって計算された各腫瘍特異性エピトープのランクを表す。

【0058】

MHCクラスI提示のためのモデル、MHCクラスII提示のためのモデル、ヘルパーT細胞活性化のためのモデル、細胞傷害性T細胞活性化のためのモデル、及びこれらのモデルを試料レベルスコアと統合して最終免疫原性スコアに達すること含む、上記のモデルにおいて、機械学習方法が用いられる。これらのモデルを訓練するために適切なくつかの機械学習方法、例えば、回帰ベースモデル、ツリーベースモデル、ベイズモデル、サポートベクターマシン、ブースティングモデル、及びニューラルネットワークベースモデルがある。

【0059】

開示された前記システム及び前記方法は、免疫腫瘍学の分野で有益である。前記システム及び前記方法は、個体のための疾患の処置を容易にするための手法を提供する。前記システム及び前記方法から同定された前記免疫原性エピトープは、個体ごとに決定され、個々に合わせた医療又は個別的医療のための手法を提供する。前記システムは、各種の免疫治療戦略において使用され得る一組の免疫原性エピトープを提供する。エピトープ同定、免疫原性予測及びエピトープランク付けの前記統合システムは、チェックポイント阻害剤、がんワクチン又は養子細胞移入などの免疫治療を考慮している患者のために有用である。がんワクチン及び養子細胞移入において、ランク付けられた前記エピトープは、ワクチン合成又は免疫細胞養成のための選択された一組の非常に有望な候補の役割を果たす。免疫チェックポイント阻害剤治療において、免疫原性エピトープの数は、前記薬物を投与することについての応答予測の信頼性のある源を供する。本システムは、前記疾患の個体又は一般集団において精密医療を実施するために適切である。

【実施例】

【0060】

(実施例1. MHCクラスI複合体に対する前記ペプチド結合親和性及び前記ペプチド結合安定性を考慮することによるMHCクラスI複合体によるペプチド提示の予測)

【0061】

ペプチドは、抗原であるためには、前記MHC複合体によって細胞の表面に提示されて免疫細胞によって認識されることが可能であることを必要とする。このプロセスは、抗原提示細胞におけるMHCクラスII複合体によるペプチド提示であって、CD4+T細胞に前記ペプチドを提示するペプチド提示と、抗原提示細胞におけるMHCクラスI複合体によるペプチド提示であって、CD8+T細胞に前記ペプチドを提示するペプチド提示と、腫瘍細胞におけるMHCクラスI複合体によるCD8+T細胞へのペプチド提示とを含む。本実施例において、MHCクラスI複合体によってペプチド提示を予測するための選択された特徴を有するモデルを構築した。

【0062】

MHCクラスI複合体へのペプチド結合、ペプチドの結合親和性及び結合安定性に影響を及ぼす2つの特性を考慮することによってMHCクラスI提示を予測するためのモデルを構築した。NetMHC4.0 (Andreatta M and Nielsen M, Bioinformatics (2016) Feb 15; 32(4): 511-7; Nielsen Mら、Protein Sci., (2003) 12: 1007-17) を用いて、MHCクラスI複合体に対するペプチドの結合親和性(IC50)を算出した

10

20

30

40

50

。NetMHC4.0において利用可能でなかったHLA複合体については、NetMHCpan3.0(Nielsen M and Andreatta M, Genome Medicine (2016): 8: 33; Hoof Iら、Immunogenetics 61.1(2009): 1-13)を用いた。NetMHCstabpan1.0(Rasmussen Mら、J Immunol. 2016 Aug 15; 197(4): 1517-24)を用いてMHCクラスI複合体に対するペプチド結合についての安定性を算出した。訓練データとして、Bassani-Sternbergら、Molecular & Cellular Proteomics, 2015及びBassani-Sternbergら、Nature Communications, 2016から回収されたデータを用いて、機械学習モデルを訓練した。未変性配列として標識されない1超の遺伝子から生成されるペプチド配列であって、長さが9~11でないものを前記訓練データから除く。提示されたペプチド-HLA複合体として同定される前記ペプチド及び対応するHLA型を正の訓練データとして用いた。他のHLA型と対になる同一ペプチド配列は、負の訓練データであると考えられた。各ペプチドの対応するHLA型に対する結合親和性及び結合安定性を上記のように算出した。特徴としてのMHCクラスI複合体への前記ペプチド結合親和性及び前記ペプチド結合安定性によりロジスティック回帰モデルを構築し、Scikit-learnにおけるLogistic Regression(Fabian Pedregosaら、JMLR(2011)Oct 12: 2825-2830)を用いてMHCクラスIによるペプチド提示を予測した。10倍の交差検証を行った。表1に前記訓練データの試験データセットについての精度及び受信者操作曲線下面積(ROCのAUC)を示す。このモデルにおいて、前記結合親和性は結合安定性よりも強く寄与することが分かった。

10

20

## 【0063】

(実施例2. 前記ペプチド結合親和性及び前記遺伝子発現レベルを考慮することによるMHCクラスI複合体によるペプチド提示の予測)

## 【0064】

MHCクラスI複合体に結合するペプチドの能力の他に、ペプチドが提示され得るようにはペプチドが発現することも重要である。本実施例において、遺伝子発現とMHCクラスI複合体に対するペプチド結合能力とを対象に含むこと(accounting for)によってMHCクラスI複合体によるペプチド提示を予測するためのモデルを構築した。

30

## 【0065】

実施例1に記載されたようにMHCクラスI複合体へのペプチドの結合親和性を算出した。ペプチドについての遺伝子発現レベルは、前記ペプチドを生成する遺伝子のRNA発現レベルとして算出される。Illumina Body Map(Petryszak Rら、Nucleic Acids Res. 2016 Jan 4; 44(D1): D746-52)を用いて、ペプチドについての前記遺伝子発現レベルを得た。実施例1に記載されたようにBassani-Sternbergらから回収された前記データを、同じ方法によってフィルタリングされる訓練データとして用いて、機械学習モデルを訓練した。提示されたペプチド-HLA複合体として同定される前記ペプチド及び対応するHLA型を正の訓練データとして用いた。他のHLA型と対になる同一ペプチド配列は、負の訓練データであると考えられた。各ペプチドの対応するHLA型に対する結合親和性と各ペプチドについての前記遺伝子発現レベルとを上記のように得た。特徴としてのMHCクラスI複合体への前記ペプチド結合親和性とペプチドについての前記遺伝子発現レベルとによりロジスティック回帰モデルを構築し、Scikit-learnにおけるLogistic Regressionを用いてMHCクラスIによるペプチド提示を予測した。10倍の交差検証を行った。表1に前記訓練データの試験データセットについての精度及びROCのAUCを示す。このモデルにおいて、前記結合親和性は遺伝子発現よりも強く寄与する。

40

## 【0066】

(実施例3. 前記ペプチド結合親和性及び前記タンパク質存在量を考慮することによる

50

MHCクラスI複合体によるペプチド提示の予測)

【0067】

MHCクラスI複合体に結合するペプチドの能力の他に、ペプチドの存在量も前記MHC複合体によって提示されるペプチドの量に影響を及ぼす可能性がある。本実施例において、ペプチド存在量も考慮される。MHCクラスI複合体に対する前記ペプチド結合能力とペプチドの存在量とを対象に含むこと (accounting for) によってMHCクラスI複合体によるペプチド提示を予測するための、選択された特徴を有するモデルを構築した。

【0068】

MHCクラスI複合体への前記ペプチド結合能力及びペプチド結合確率に影響を及ぼす2つの特性、すなわちペプチドの結合親和性及びペプチドの存在量をそれぞれ考慮することによってMHCクラスI提示を予測するためのモデルを構築した。実施例1に記載されたようにMHCクラスI複合体へのペプチドの結合親和性を算出した。本実施例におけるペプチドの存在量は、前記ペプチドを含み、且つ遺伝子から生成される前記タンパク質の最大存在量と定義される、前記ペプチドについての前記タンパク質存在量によって表される。PaxDb Protein Abundance DatabaseのH.サピエンス-全生物(統合)データベース(Wang, M.ら、Proteomics 2015, 10.1002/pmic.201400441)を用いてペプチドについてのタンパク質存在量を得た。実施例1に記載されたようにBassani-Sternbergらから回収された前記データを、同じようにフィルタリングされる訓練データとして用いて、機械学習モデルを訓練した。提示されたペプチド-HLA複合体として同定される前記ペプチド及び対応するHLA型を正の訓練データとして用いた。他のHLA型と対になる同一ペプチド配列は、負の訓練データであると考えられた。各ペプチドの対応するHLA型に対する結合親和性と各ペプチドについての前記タンパク質存在量とを上記のように算出した。特徴としてのMHCクラスI複合体への前記ペプチド結合親和性とペプチドについての前記タンパク質存在量とによりロジスティック回帰モデルを構築し、Scikit-learnにおけるLogistic Regressionを用いてMHCクラスIによるペプチド提示を予測した。10倍の交差検証を行った。表1に前記訓練データの試験データセットについての精度及びROCのAUCを示す。この作業において、前記結合親和性はタンパク質存在量よりも前記モデルに強く寄与することが分かった。

【0069】

(実施例4. MHCクラスII複合体に対する前記ペプチド結合親和性を考慮することによるMHCクラスII複合体によるペプチド提示の予測)

【0070】

ペプチドは、抗原であるためには、前記MHC複合体によって細胞の表面に提示され免疫細胞によって認識されることが可能であることを必要とする。このプロセスは、抗原提示細胞におけるMHCクラスII複合体によるペプチド提示であって、CD4+T細胞に前記ペプチドを提示するペプチド提示と、抗原提示細胞におけるMHCクラスI複合体によるペプチド提示であって、CD8+T細胞に前記ペプチドを提示するペプチド提示と、腫瘍細胞におけるMHCクラスI複合体によるCD8+T細胞へのペプチド提示とを含む。本実施例において、MHCクラスII複合体によってペプチド提示を予測するためのモデルを構築した。

【0071】

MHCクラスII複合体へのペプチドの結合親和性を考慮することによってMHCクラスII提示を予測するためのモデルを構築した。NetMHCII2.2(Nielson Mら、BMC Bioinformatics. 2007 Jul 4; 8:238)を用いてMHCクラスII複合体に対するペプチド結合親和性を算出した。利用可能でなかった前記HLA型について、NetMHCIIpan3.1(Andreatta Mら、Immunogenetics. 2015 Nov; 67(11-12):641-50)を用いて前記ペプチド結合親和性を算出した。訓練データとして、Chongら、Molecular & Cellular Proteomics, 2017から回収さ

10

20

30

40

50

れた前記データを用いて機械学習モデルを訓練した。長さが9未満の前記ペプチド配列を前記訓練データから除く。提示されたペプチド - H L A複合体として同定される前記ペプチド及び対応するH L A型を正の訓練データとして用いた。他のH L A型と対になる同一ペプチド配列は、負の訓練データであると考えられた。前記特徴としてM H CクラスI I複合体への前記ペプチド結合親和性を用いてロジスティック回帰モデルを構築し、S c i k i t - l e a r nにおけるL o g i s t i c R e g r e s s i o nを用いてM H CクラスI Iによるペプチド提示を予測した。10倍の交差検証を行った。表1に前記訓練データの試験データセットについての精度及びR O CのA U Cを示す。

【0072】

【表1】

実施例1～実施例4のモデル性能

	実施例1 (MHCクラス I)	実施例2 (MHCクラス I)	実施例3 (MHCクラス I)	実施例4 (MHCクラスII)
ROCのAUC	0.88	0.88	0.88	0.78
精度	0.82	0.81	0.81	0.71

10

【0073】

(実施例5 . M H CクラスI Iについての前記自己類似性と、既知の抗原との前記ペプチド相同性とを考慮することによるC D 4 + T細胞の免疫応答を刺激するペプチドの能力の予測)

20

【0074】

ペプチドが免疫原性を有する抗原であるためには、M H CクラスI複合体及びM H CクラスI I複合体による提示の他に、C D 4 + T細胞及びC D 8 + T細胞の免疫応答を刺激するペプチドの能力も重要である。本実施例において、C D 4 + T細胞免疫応答を刺激するペプチドの能力を予測するための選択された特徴を有するモデルを構築した。

【0075】

C D 4 + T細胞による認識に影響を及ぼすペプチドの2つの特性を考慮することによってC D 4 + T細胞免疫応答を刺激するペプチドの能力を予測するためのモデルを構築した。前記特性は、M H CクラスI Iについての前記ペプチドとヒトタンパク質配列との間の類似性 ( M H CクラスI Iについての自己類似性 ( self-similarity for MHC Class II ) と称される ) 及び既知の抗原とのペプチド相同性である。3つの工程におけるペプチドのM H CクラスI Iについての自己類似性を算出した。第1に、E N S E M B L G R c h 3 7 ( www . e n s e m b l . o r g / ) から全タンパク質配列を検索し、9 ~ 23のアミノ酸の全ての可能な長さに切り取った。全てのペプチド配列が変異ペプチドであるというわけではないので、全ペプチド上において野生型ペプチドに対する変異ペプチドの関係を模倣した。すなわち、前記切り取られたヒトタンパク質配列に各ペプチドをアラインメントし、等しい長さを有し、且つ1つのミスマッチのみを有するものを選択した。これらの選択され切り取られたヒトタンパク質配列は、自己ペプチド ( self-peptides ) であると考えられる。第2に、実施例4に記載されたように前記M H CクラスI I複合体に対するペプチド及びその対応する自己ペプチドの結合親和性を算出した。第3に、M H CクラスI Iについての自己類似性を、より大きい結合親和性で除算されたより小さい結合親和性であると定義した。ヒト配列に対して2つ以上のミスマッチを有する前記ペプチドについて、それらの自己類似性として0を割り当てた。B L A S T ( h t t p s : / / b l a s t . n c b i . n l m . n i h . g o v / B l a s t . c g i ) を用いて、既知の抗原を有する前記ペプチド配列をアラインメントすることによって、前記ペプチド相同性を算出した。I E D Bデータベース ( www . i e d b . o r g ) の抗原データセットから、既知の抗原配列を検索して、ウイルス抗原又は細菌抗原で標識されたものを選択した。既知の抗原によってアラインメントされた、より高い比率の配列を有する前記ペプチドは、相同的であると考えられる。いかなる既知の抗原ともマッチしなかったペプチドについて、それらのペプチド相同性には0が割り当てられる。I E D Bデータベース ( h t t p

30

40

50

(<http://www.iedb.org>) から回収された標識 CD4 + T細胞免疫応答転帰を伴う、ファイル名が「tcell\_full\_v3.csv」であり、且つ MHC クラスが「II」と標識されたデータを訓練データとして用いて、CD4 + T細胞の免疫応答を予測するための機械学習モデルを訓練した。ペプチド長が 9 ~ 30 ではなく、細胞型が正常 T細胞ではなく、アッセイ群が免疫シグナル放出、T細胞活性化及び T細胞 - APC 結合と標識されない前記データを前記訓練データから除く。特徴としての MHC クラス II についての自己類似性と、既知の抗原とのペプチド相同性によりロジスティック回帰モデルを構築し、Scikit-learn における Logistic Regression を用いて、CD4 + T細胞免疫応答を刺激するペプチドの能力を予測した。前記モデルで 10 倍の交差検証を行った。表 2 に前記訓練データの試験データセットについての精度及び ROC の AUC を示す。

10

## 【0076】

モデル訓練の後、試験データとして、Ott Pら、Nature, 2017 から回収された標識 CD4 + T細胞免疫応答転帰を伴う、T細胞免疫応答について実験的に試験された前記データを用いて、前記モデルを試験した。上記のように特徴算出を行い、前記試験データにおける各ペプチドについての前記算出特徴及び訓練パラメータを有する前記モデルについての前記予測スコアを算出した。正の応答データ及び負の応答データについての前記予測スコアのボックスプロットと、ウィルコクソン順位和検定から算出される p 値とが図 3 A において示される。

## 【0077】

(実施例 6 . 前記ペプチド免疫原性を考慮することによって CD8 + T細胞の免疫応答をペプチドが刺激する能力の予測)

20

## 【0078】

ペプチドが免疫原性を有する抗原であるためには、MHC クラス I 複合体及び MHC クラス II 複合体による提示の他に、CD4 + T細胞及び CD8 + T細胞の免疫応答を刺激するペプチドの能力も重要である。本実施例において、CD8 + T細胞免疫応答を刺激するペプチドの能力を予測するための選択された特徴を有するモデルを構築した。

## 【0079】

CD8 + T細胞による前記認識に影響を及ぼすペプチドの特性を考慮することによって CD8 + T細胞免疫応答を刺激するペプチドの能力を予測するためのモデルを構築した。前記特性は、前記ペプチドの免疫原性である。IEDB 免疫原性予測因子 (Calis JJ, PLoS Comput Biol. (2013) Oct 9 (10): e1003266) を用いてペプチドの免疫原性を算出した。IEDB データベース (<http://www.iedb.org>) から回収された標識 CD8 + T細胞免疫応答転帰を伴う、ファイル名が「tcell\_full\_v3.csv」であり、且つ MHC クラスが「I」と標識されたデータを訓練データとして用いて、CD8 + T細胞の免疫応答を予測するための機械学習モデルを訓練した。ペプチド長が 8 ~ 11 ではなく、細胞型が正常 T細胞ではなく、アッセイ群が免疫シグナル放出、T細胞活性化及び T細胞 - APC 結合と標識されない前記データを前記訓練データから除く。前記特徴としての前記 IEDB 免疫原性予測因子によって予測される前記ペプチド免疫原性によりロジスティック回帰モデルを構築し、Scikit-learn における Logistic Regression を用いて、CD8 + T細胞免疫応答を刺激するペプチドの能力を予測した。前記モデルで 10 倍の交差検証を行ったが、表 2 に前記訓練データの試験データセットについての精度及び ROC の AUC を示す。モデル訓練の後、試験データとして、Ott Pら、Nature, 2017 から回収された標識 CD8 + T細胞免疫応答転帰を伴う、T細胞免疫応答について実験的に試験された前記データを用いて、前記モデルを試験した。上記のように特徴算出を行い、前記試験データにおける各ペプチドについての前記算出特徴及び訓練パラメータを有する前記モデルについての前記予測スコアを算出した。正の応答データ及び負の応答データについての前記予測スコアのボックスプロットと、ウィルコクソン順位和検定から算出される p 値とが図 3 B において示される。

30

40

50

## 【0080】

(実施例7. MHCクラスII についての前記自己類似性と、既知の抗原との前記ペプチド相同性と、実施例4における特徴とを考慮することによるCD4 + T細胞の免疫応答を刺激するペプチドの能力の予測)

## 【0081】

CD4 + T細胞の免疫応答は、前記エピトープが抗原提示細胞によって前記CD4 + T細胞に提示されることを必要とする。本実施例において、免疫応答を刺激するペプチドの能力と、さらに、前記ペプチドが抗原提示細胞上のMHCクラスII によって提示される能力との両方を明らかにすることによって、CD4 + T細胞免疫応答を刺激するペプチドの能力を予測するための選択された特徴を有するモデルを構築した。

10

## 【0082】

CD4 + T細胞による前記認識とMHCクラスII 複合体による前記提示とに影響を及ぼすペプチドの特性を考慮することによってCD4 + T細胞免疫応答を刺激するペプチドの能力を予測するためのモデルを構築した。前記特性は、MHCクラスII についての前記自己類似性、既知の抗原とのペプチド相同性、及びMHCクラスII 複合体に対する前記ペプチドの結合親和性である。実施例5に記載されたようにMHCクラスII についての前記自己類似性と、前記既知の抗原との相同性とを算出し、実施例4に記載されたようにMHCクラスII 複合体に対する前記結合親和性を算出した。実施例5に記載されたように、前記IEDBデータベースから回収される標識CD4 + T細胞免疫応答転帰を伴う前記データを訓練データとして用いて、CD4 + T細胞の免疫応答を予測するための機械学習モデルを訓練した。データフィルタリングプロセスは、実施例5に記載されたものと同じである。特徴としてのMHCクラスII についての前記自己類似性と、既知の抗原との前記ペプチド相同性と、MHCクラスII 複合体に対する前記ペプチド結合親和性によりロジスティック回帰モデルを構築し、Scikit-learnにおけるLogistic Regressionを用いてCD4 + T細胞免疫応答を刺激するペプチドの能力を予測した。前記モデルで10倍の交差検証を行った。表2に前記訓練データの試験データセットについての精度及びROCのAUCを示す。本実施例において、MHCクラスII 複合体に対するペプチド結合親和性は、MHCクラスII についての自己類似性及び既知の抗原とのペプチド相同性よりも強い予測因子であることが分かった。

20

## 【0083】

モデル訓練の後、実施例5に記載されたように、試験データとして、OttPらから回収された標識CD4 + T細胞免疫応答転帰を伴う前記データを用いて、前記モデルを試験した。上記のように特徴算出を行い、前記試験データにおける各ペプチドについての前記算出特徴及び訓練パラメータを有する前記モデルについての前記予測スコアを算出した。正の応答データ及び負の応答データについての前記予測スコアのボックスプロットと、ウィルコクソン順位和検定から算出されるp値とを図4Aに示す。

30

## 【0084】

(実施例8. 前記ペプチド免疫原性と実施例1における前記特徴とを考慮することによるCD8 + T細胞の免疫応答を刺激するペプチドの能力の予測)

## 【0085】

CD8 + T細胞の免疫応答は、前記エピトープがMHCクラスI によって前記CD8 + T細胞に提示されることを必要とする。本実施例において、免疫応答を刺激するペプチドの能力と、さらに、前記ペプチドがMHCクラスI によって提示される能力との両方を対象に含むこと(accounting for)によって、CD8 + T細胞免疫応答を刺激するペプチドの能力を予測するための選択された特徴を有するモデルを構築した。

40

## 【0086】

CD8 + T細胞による前記認識とMHCクラスI 複合体による前記提示とに影響を及ぼすペプチドの特性を考慮することによってCD8 + T細胞免疫応答を刺激するペプチドの能力を予測するためのモデルを構築した。前記特性は、前記ペプチドの免疫原性、MHCクラスI 複合体に対する前記ペプチドの結合親和性及び結合安定性である。実施例6に記

50

載されたように前記ペプチド免疫原性を算出し、実施例 1 に記載されたように MHC クラス I 複合体に対する前記結合親和性及び前記結合安定性を算出した。実施例 6 に記載されたように、前記 I E D B データベースから回収される標識 C D 8 + T 細胞免疫応答転帰を伴う前記データを訓練データとして用いて、C D 8 + T 細胞の免疫応答を予測するための機械学習モデルを訓練した。データフィルタリングプロセスは、実施例 6 に記載されたものと同じである。特徴としての前記ペプチド免疫原性と、MHC クラス I 複合体に対する前記ペプチド結合親和性及び前記ペプチド結合安定性とによりロジスティック回帰モデルを構築し、Scikit-learnにおける Logistic Regression を用いて C D 8 + T 細胞免疫応答を刺激するペプチドの能力を予測した。前記モデルで 10 倍の交差検証を行った。表 2 に前記訓練データの試験データセットについての精度及び R O C の A U C を示す。このモデルにおいてペプチド結合安定性が最も寄与し、次にペプチド結合親和性が寄与し、最後にペプチド免疫原性が寄与したことが分かった。

10

【 0 0 8 7 】

モデル訓練の後、実施例 6 に記載されたように、試験データとして、O t t P から回収された標識 C D 8 + T 細胞免疫応答転帰を伴う前記データを用いて、前記モデルを試験した。上記のように特徴算出を行い、前記試験データにおける各ペプチドについての前記算出特徴及び訓練パラメータを有する前記モデルについての前記予測スコアを算出した。正の応答データ及び負の応答データについての前記予測スコアのボックスプロットと、ウィルコクソン順位検定から算出される p 値とを図 4 B に示す。

【 0 0 8 8 】

( 実施例 9 . MHC クラス I I についての前記自己類似性と、既知の抗原との前記ペプチド相同性と、実施例 4 における前記モデルの予測スコアとを考慮することによる C D 4 + T 細胞の免疫応答を刺激するペプチドの能力の予測 )

20

【 0 0 8 9 】

C D 4 + T 細胞の免疫応答は、前記エピトープが抗原提示細胞によって前記 C D 4 + T 細胞に提示されることを必要とする。本実施例において、実施例 4 において構築されたモデルを考慮することによって、免疫応答を刺激するペプチドの能力と、さらに、前記ペプチドが抗原提示細胞上の MHC クラス I I によって提示される能力との両方を対象に含むこと ( accounting for ) によって、C D 4 + T 細胞免疫応答を刺激するペプチドの能力を予測するための選択された特徴を有するモデルを構築した。

30

【 0 0 9 0 】

C D 4 + T 細胞による前記認識と MHC クラス I I 複合体による前記提示とに影響を及ぼすペプチドの特性を考慮することによって C D 4 + T 細胞免疫応答を刺激するペプチドの能力を予測するためのモデルを構築した。前記特性は、MHC クラス I I についての前記自己類似性、既知の抗原とのペプチド相同性、及び実施例 4 に記載された前記 MHC クラス I I 提示モデルの予測スコアである。実施例 5 に記載されたように MHC クラス I I についての前記自己類似性と、前記既知の抗原との相同性とを算出した。前記 MHC クラス I I 提示モデルの前記予測スコアを算出するために、実施例 4 に記載されたように MHC クラス I I 複合体に対する前記結合親和性を算出した。実施例 4 から得られた上記記載の算出特徴と訓練パラメータとを有する前記 MHC クラス I I 提示モデルの前記予測スコアを算出した。実施例 5 に記載されたように、前記 I E D B データベースから回収される標識 C D 4 + T 細胞免疫応答転帰を伴う前記データを訓練データとして用いて、C D 4 + T 細胞の免疫応答を予測するための機械学習モデルを訓練した。データフィルタリングプロセスは、実施例 5 に記載されたものと同じである。特徴としての MHC クラス I I についての前記自己類似性と、既知の抗原との前記ペプチド相同性と、前記 MHC クラス I I 提示モデルの前記予測スコアとによりロジスティック回帰モデルを構築し、Scikit-learnにおける Logistic Regression を用いて C D 4 + T 細胞免疫応答を刺激するペプチドの能力を予測した。前記モデルで 10 倍の交差検証を行ったが、表 2 に前記訓練データの試験データセットについての精度及び R O C の A U C を示す。

40

【 0 0 9 1 】

50

モデル訓練の後、実施例 5 に記載されたように、O t t P から回収された標識 C D 4 + T 細胞免疫応答転帰を伴う前記データを用いて、前記モデルを試験した。上記のように特徴算出を行い、前記試験データにおける各ペプチドについての前記算出特徴及び訓練パラメータを有する前記モデルについての前記予測スコアを算出した。正の応答データ及び負の応答データについての前記予測スコアのボックスプロットと、ウィルコクソン順位和検定から算出される p 値とが図 5 A において示される。

【 0 0 9 2 】

( 実施例 1 0 . 前記ペプチド免疫原性と実施例 1 における前記モデルの予測スコアとを考慮することによる C D 8 + T 細胞の免疫応答を刺激するペプチドの能力の予測 )

【 0 0 9 3 】

C D 8 + T 細胞の免疫応答は、前記エピトープが M H C クラス I によって前記 C D 8 + T 細胞に提示されることを必要とする。本実施例において、実施例 1 においてモデル化されたように、免疫応答を刺激するペプチドの能力と、さらに、前記ペプチドが M H C クラス I によって提示される能力との両方を対象に含むこと ( accounting for ) によって、C D 8 + T 細胞免疫応答を刺激するペプチドの能力を予測するための選択された特徴を有するモデルを構築した。

【 0 0 9 4 】

C D 8 + T 細胞による前記認識と M H C クラス I 複合体による前記提示とに影響を及ぼすペプチドの特性を考慮することによって C D 8 + T 細胞免疫応答を刺激するペプチドの能力を予測するためのモデルを構築した。前記特性は、前記ペプチドの免疫原性、及び実施例 1 に記載された前記 M H C クラス I 提示モデルの予測スコアである。実施例 6 に記載されたように前記ペプチド免疫原性を算出した。前記 M H C クラス I 提示モデルの前記予測スコアを算出するために、実施例 1 に記載されたように M H C クラス I 複合体に対する前記結合親和性及び前記結合安定性を算出した。実施例 1 から得られた上記記載の算出特徴と訓練パラメータとを有する前記 M H C クラス I 提示モデルの前記予測スコアを算出した。実施例 6 に記載されたように、前記 I E D B データベースから回収される標識 C D 8 + T 細胞免疫応答転帰を伴う前記データを訓練データとして用いて、C D 8 + T 細胞の免疫応答を予測するための機械学習モデルを訓練した。データフィルタリングプロセスは、実施例 6 に記載されたものと同じである。特徴としての前記ペプチド免疫原性と、前記 M H C クラス I 提示モデルの前記予測スコアとによりロジスティック回帰モデルを構築し、S c i k i t - l e a r n における L o g i s t i c R e g r e s s i o n を用いて C D 8 + T 細胞免疫応答を刺激するペプチドの能力を予測した。前記モデルで 1 0 倍の交差検証を行った。表 2 に前記訓練データの試験データセットについての精度及び R O C の A U C を示す。このモデルにおいて、前記 M H C クラス I 提示モデルの前記予測スコアは、前記ペプチド免疫原性特徴よりも寄与する。

【 0 0 9 5 】

モデル訓練の後、実施例 6 に記載されたように、試験データとして、O t t P から回収された標識 C D 8 + T 細胞免疫応答転帰を伴う前記データを用いて、前記モデルを試験した。上記のように特徴算出を行い、前記試験データにおける各ペプチドについての前記算出特徴及び訓練パラメータを有する前記モデルについての前記予測スコアを算出した。正の応答データ及び負の応答データについての前記予測スコアのボックスプロットと、ウィルコクソン順位和検定から算出される p 値とが図 5 B において示される。

【 0 0 9 6 】

【 表 2 】

実施例 5 ~ 実施例 1 0 のモデル性能

	実施例 5 (CD4+ T細胞)	実施例 6 (CD8+ T細胞)	実施例 7 (CD4+ T細胞)	実施例 8 (CD8+ T細胞)	実施例 9 (CD4+ T細胞)	実施例 10 (CD8+ T細胞)
ROCのAUC	0.59	0.56	0.58	0.64	0.59	0.61
精度	0.59	0.55	0.6	0.62	0.63	0.59

10

20

30

40

50

## 【 0 0 9 7 】

( 実施例 1 1 . 実施例 1、実施例 4 ~ 実施例 6 からの抗原提示及び免疫応答の両方を考慮する、前記ペプチドが免疫原であることの予測 )

## 【 0 0 9 8 】

ペプチドが、免疫応答を、すなわち免疫原として引き起こすことが可能であるためには、前記ペプチドが M H C クラス I 複合体及び M H C クラス I I 複合体によって提示され、さらに C D 4 + T 細胞応答及び C D 8 + T 細胞応答を刺激する能力が必要である。そのため、本実施例においては、前記ペプチドレベルスコアを算出することによって免疫原としての前記ペプチドを予測するために実施例 1 及び実施例 4 ~ 実施例 6 を統合した。

## 【 0 0 9 9 】

実施例 1 及び 4 からの抗原提示情報並びに実施例 5 及び実施例 6 からの C D 4 + 情報及び C D 8 + 情報を統合して前記ペプチドレベルスコアを算出することによって、統合された一組のモデルを構築した。前記ペプチドレベルスコアは、M H C クラス I 複合体及び M H C クラス I I 複合体によって提示され、さらに C D 4 + T 細胞応答及び C D 8 + T 細胞応答を刺激するペプチドの能力を表す。実施例 1、実施例 4、実施例 5 及び実施例 6 から算出されるあらゆる予測スコアを乗算することによって前記ペプチドレベルスコアを算出した。なお、M H C クラス I I 提示及び C D 4 + T 細胞免疫応答予測のペプチドは、M H C クラス I 提示及び C D 8 + T 細胞免疫応答のペプチドよりも長い。これを解決するために、C D 8 + ペプチドを含む、あらゆる、可能な、より長いペプチドから、M H C クラス I I 提示及び C D 4 + T 細胞免疫応答についての特徴を算出した。前記 C D 4 + 情報及び前記 C D 8 + 情報を統合するために、C D 4 + T 細胞免疫応答を刺激する最も強い結合能力を有するものを選択した。P a t r i c k A . O t t らから検索された C D 8 + T 細胞免疫応答転帰で標識されたデータを試験データとして用いて、前記ペプチドレベルスコアを試験した。前記試験データにおける各ペプチドについての前記ペプチドレベルスコアを算出した。正の応答データ及び負の応答データについての前記予測スコアのボックスプロットと、ウィルコクソン順位和検定によって算出される p 値とを図 6 A に示す。

## 【 0 1 0 0 】

( 実施例 1 2 . 実施例 7 ~ 実施例 8 からの抗原提示及び免疫応答の両方を考慮する、前記ペプチドが免疫原であることの予測 )

## 【 0 1 0 1 】

実施例 1 1 で論じられたように、免疫原としてのペプチドを予測することは、前記ペプチドが M H C クラス I 複合体及び M H C クラス I I 複合体によって提示され、さらに C D 4 + T 細胞応答及び C D 8 + T 細胞応答を刺激することを必要とする。本実施例においては、前記ペプチドレベルスコアを算出することによって免疫原としての前記ペプチドを予測するために、実施例 7 ~ 実施例 8 を統合した。

## 【 0 1 0 2 】

実施例 7 ~ 実施例 8 からの抗原提示情報及び免疫応答情報を統合して前記ペプチドレベルスコアを算出することによって、統合された一組のモデルを構築した。前記ペプチドレベルスコアは、M H C クラス I 複合体及び M H C クラス I I 複合体によって提示され、さらに C D 4 + T 細胞応答及び C D 8 + T 細胞応答を刺激するペプチドの能力を表す。実施例 7 ~ 実施例 8 から算出されるあらゆる予測スコアを乗算することによって前記ペプチドレベルスコアを算出した。なお、M H C クラス I I 提示及び C D 4 + T 細胞免疫応答予測のペプチドは、M H C クラス I 提示及び C D 8 + T 細胞免疫応答のペプチドよりも長い。これを解決するために、C D 8 + ペプチドを含む、あらゆる、可能な、より長いペプチドから、M H C クラス I I 提示及び C D 4 + T 細胞免疫応答についての特徴を算出した。前記 C D 4 + 情報及び前記 C D 8 + 情報を統合するために、C D 4 + T 細胞免疫応答を刺激する最も強い結合能力を有するものを選択した。P a t r i c k A . O t t らから検索された C D 8 + T 細胞免疫応答転帰で標識されたデータを試験データとして用いて、前記ペプチドレベルスコアを試験した。前記試験データにおける各ペプチドについての前記ペプチドレベルスコアを算出した。正の応答データ及び負の応答データについての前記予測

10

20

30

40

50

スコアのボックスプロットと、ウィルコクソン順位和検定によって算出される p 値とを図 6 B に示す。

【 0 1 0 3 】

( 実施例 1 3 . 実施例 9 ~ 実施例 1 0 からの抗原提示及び免疫応答の両方を考慮する、前記ペプチドが免疫原であることの予測 )

【 0 1 0 4 】

実施例 1 1 で論じられるように、免疫原としてのペプチドを予測することは、前記ペプチドが M H C クラス I 複合体及び M H C クラス I I 複合体によって提示され、さらに C D 4 + T 細胞応答及び C D 8 + T 細胞応答を刺激することを必要とする。本実施例においては、前記ペプチドレベルスコアを算出することによって免疫原としての前記ペプチドを予測するために、実施例 9 ~ 実施例 1 0 を統合した。

10

【 0 1 0 5 】

実施例 9 ~ 実施例 1 0 からの抗原提示情報及び免疫応答情報を統合して前記ペプチドレベルスコアを算出することによって、統合された一組のモデルを構築した。前記ペプチドレベルスコアは、M H C クラス I 複合体及び M H C クラス I I 複合体によって提示され、さらに C D 4 + T 細胞応答及び C D 8 + T 細胞応答を刺激するペプチドの能力を表す。実施例 9 ~ 実施例 1 0 から算出されるあらゆる予測スコアを乗算することによって前記ペプチドレベルスコアを算出した。なお、M H C クラス I I 提示及び C D 4 + T 細胞免疫応答予測のペプチドは、M H C クラス I 提示及び C D 8 + T 細胞免疫応答のペプチドよりも長い。これを解決するために、C D 8 + ペプチドを含む、あらゆる、可能な、より長いペプチドから、M H C クラス I I 提示及び C D 4 + T 細胞免疫応答についての特徴を算出した。前記 C D 4 + 情報及び前記 C D 8 + 情報を統合するために、C D 4 + T 細胞免疫応答を刺激する最も強い結合能力を有するものを選択した。P a t r i c k A . O t t らから検索された C D 8 + T 細胞免疫応答転帰で標識されたデータを試験データとして用いて、前記ペプチドレベルスコアを試験した。前記試験データにおける各ペプチドについての前記ペプチドレベルスコアを算出した。正の応答データ及び負の応答データについての前記予測スコアのボックスプロットと、ウィルコクソン順位和検定によって算出される p 値とを図 6 C において示す。

20

【 0 1 0 6 】

( 実施例 1 4 . ペプチドレベルスコアを用いて免疫原としての患者の試料からのペプチドを同定し、スコア付け、ランク付け、評価すること )

30

【 0 1 0 7 】

患者について免疫原としてのペプチドを同定し、スコア付け、ランク付け、評価するために、O t t P ら ( N a t u r e . 2 0 1 7 J u l 1 3 ; 5 4 7 ( 7 6 6 2 ) : 2 1 7 - 2 2 1 ) からの前記ペプチドにおける我々のスコア付け方法を検索し、比較した。O t t P らは、6 人の黒色腫患者から免疫性長鎖ペプチド ( I M P ) を合成し、免疫細胞における前記ペプチドの免疫原性を実験的に試験した。これらの免疫性長鎖ペプチドは、長さが 1 5 ~ 3 0 のアミノ酸の範囲であった。試験された前記ペプチドは、C D 8 + 応答について長さが 9 ~ 1 0 のアミノ酸の、より短いものであり、前記応答ペプチドが同定された。よって、我々は、前記免疫性長鎖ペプチドから、長さが 9 ~ 1 0 のアミノ酸であり、前記変異アミノ酸を有する、より短い重複ペプチドを検索した変異。次いで、実施例 1 1 ~ 実施例 1 3 に記載されたように、各患者の、より短い前記ペプチドについて、ペプチドレベルスコアを算出した。5 0 はワクチン選択において一般的な数なので、トップ 5 0 のスコアを有する前記ペプチドを選択した。実施例 1 2 を用いた患者 1 についてのデータを表 3 に示す。次いで、図 7 に、実施例 1 1 ~ 実施例 1 3 を用いて算出された、各患者についてのトップ 5 0 の前記ペプチドの中で C D 8 + 応答を引き起こすことが可能であった応答ペプチドの数を示す。

40

【 0 1 0 8 】

50

【表 3】

実施例 1 2 からの算出を用いた、患者 1 のトップ 50 のペプチドレベルスコア

ランク	患者の ID	IMP の ID	配列	ペプチドレベルスコア	CD8 応答
1	1	1-IMP04	KLKFTLVF	0.345	1
2	1	1-IMP07	RFLEYLPLRF	0.336	1
3	1	1-IMP02	VQKVASKIPF	0.328	0
4	1	1-IMP06	TLFHFTYEL	0.321	0
5	1	1-IMP06	TLFHFTYELL	0.317	0
6	1	1-IMP11	KFGDLTNNF	0.303	0
7	1	1-IMP26	PREEFLRLC	0.302	0
8	1	1-IMP08	KLFESKAEL	0.299	0
9	1	1-IMP17	LCPREEFLR	0.298	0
10	1	1-IMP02	PPDRITEES	0.297	0
11	1	1-IMP04	VLAKKLVF	0.295	0
12	1	1-IMP03	KKKWFLFQD	0.292	0
13	1	1-IMP02	PPDRITEE	0.291	0
14	1	1-IMP07	HTELERFLE	0.288	0
15	1	1-IMP08	KLFESKAELA	0.287	0
16	1	1-IMP07	TELERFLEY	0.286	0
17	1	1-IMP07	LLHTELERF	0.286	0
18	1	1-IMP02	FPDRITEES	0.285	0
19	1	1-IMP19	VSVGDFSQEF	0.283	0
20	1	1-IMP02	IPFPDRITEE	0.280	0
21	1	1-IMP17	CPREEFLRLC	0.280	0
22	1	1-IMP12	ALFASRPRF	0.279	0
23	1	1-IMP03	FLFQDSKKI	0.277	0
24	1	1-IMP08	DKLFESKAE	0.277	0
25	1	1-IMP03	SKKKWFLFQD	0.276	0
26	1	1-IMP02	IPFPDRITE	0.276	0
27	1	1-IMP12	GGALFASRP	0.276	0
28	1	1-IMP07	LHTELERFLE	0.275	0

10

20

30

【 0 1 0 9 】

40

50

表3 続き

ランク	患者の ID	IMP の ID	配列	ペプチドレベルスコア	CD8 応答
29	1	1-IMP26	LSPREEFLR	0.275	0
30	1	1-IMP26	SPREEFLRLC	0.274	0
31	1	1-IMP19	VGDFSQEFS	0.272	0
32	1	1-IMP03	KKKWFLFQDS	0.271	0
33	1	1-IMP10	DSGIPENFVN	0.271	0
34	1	1-IMP19	VGDFSQEFSP	0.271	0
35	1	1-IMP12	RGGALFASRP	0.270	0
36	1	1-IMP19	SVGDFSQEF	0.269	0
37	1	1-IMP10	LADSGIPEN	0.267	0
38	1	1-IMP04	KKLKFVTLV	0.266	0
39	1	1-IMP02	SKIPFPDRIT	0.265	0
40	1	1-IMP10	GIPENSFNV	0.265	0
41	1	1-IMP02	KIPFPDRIT	0.264	0
42	1	1-IMP11	GKFGDLTNN	0.264	0
43	1	1-IMP08	DSDKLFESK	0.263	0
44	1	1-IMP08	EDSDKLFES	0.263	0
45	1	1-IMP07	LLHTELERFL	0.262	0
46	1	1-IMP12	RRGGALFASR	0.262	0
47	1	1-IMP10	IPENSFNVS	0.262	0
48	1	1-IMP02	ASKIPFPDR	0.261	0
49	1	1-IMP06	LFHTFYELLI	0.261	0
50	1	1-IMP07	ILLHTELERF	0.259	0

10

20

## 【 0 1 1 0 】

( 実施例 15 . 変異アレルのクローナリティから試料レベルスコアを決定すること )

## 【 0 1 1 1 】

試料特異的データから変異アレルのクローナリティを算出した。ホルマリン固定パラフィン包埋 ( F F P E ) 試料を得て、各患者について末梢血単核細胞試料を対にした。Q I A a m p ( 登録商標 ) D N A F F P E T i s s u e K i t ( Q I A G E N ( 登録商標 ) 、ドイツ、ヒルデン) を用いて、ゲノムDNAを抽出する。18136対の単位複製配列を標的とする多重化PCRを用いて、DNAを増幅した。製造業者の推奨プロトコルに従って、I o n P I C h i p ( T h e r m o F i s h e r S c i e n t i f i c 、米国マサチューセッツ州、ウォルサム) を有するI o n P r o t o n ( 商標 ) ( T h e r m o F i s h e r S c i e n t i f i c 、米国マサチューセッツ州、ウォルサム) システムを用いて、前記試料のエキソーム ( e x o m e ) の配列決定を行った。製造業者から提供されたソフトウェアT o r r e n t V a r i a n t C a l l e r ( T V C ) v . 4 . 4 によって未処理配列の読み取りを処理し、. b a mファイル及び. v c fファイルを生じた。T V C は、各変異体についての変異頻度も算出した。V a r i a n t E f f e c t P r e d i c t o r v . 7 4 によって前記変異体をアノテートした。次いで、d b S N P 1 3 8 、 1 0 0 0 G e n o m e 及び正常対血液によって一塩基変異多型 ( S N P ) 及び生殖細胞変異のフィルタリングを行った。残りの変異体を手作業でチェックした。コピー数及び腫瘍純度を、それぞれO N C O C N V 及びA D T E x によって. b a mファイルから決定した。正常対血液に対するF F P E のS N P アレル頻度の偏差8%超によってヘテロ接合性の消失 ( L O H ) を決定した。

30

40

50

【0112】

変異アレルのクローナリティは、期待アレル頻度及び観察アレル頻度の統計的有意性に基づいて変異アレルの期待数を割り当て、次いでサブクローン純度を推定することに基づく。前記期待アレル頻度 (McGrananら、Science (2016) Mar 25 ; 351 (6280) : 1463 - 9) は、以下のように算出される。

【0113】

【数1】

$$AF_{期待} = \frac{pM}{pC_t + C_n(1-p)}$$

10

【0114】

ここで、AF<sub>期待</sub>は期待アレル頻度を表し、tは腫瘍変異を表し、pは腫瘍純度を表し、Cはコピー数を表し、nは正常条件を表し、Mは変異アレル数を表す。Mを割り当てることは、<sup>2</sup>統計的有意性を有するAF<sub>期待</sub>と最も近い観察アレル頻度に基づき、ここでAF<sub>期待</sub>は、表4 (Sunら、Cancer Res (2014) 74 (19S) : 1893 から修正) からのものである。

【0115】

【表4】

コピー数、LOH及び腫瘍純度に対する変異アレル数及び期待アレル頻度

コピー数*	LOH状態	腫瘍純度**												
		M	0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1	
1	LOH	1	0	0.05	0.11	0.18	0.25	0.33	0.43	0.54	0.67	0.82	1	
2	LOH	2	0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1	
2	het	1	0	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	
3	LOH	3	0	0.14	0.27	0.39	0.50	0.60	0.69	0.78	0.86	0.93	1	
3	het	1	0	0.05	0.09	0.13	0.17	0.20	0.23	0.26	0.29	0.31	0.33	
3	het	2	0	0.10	0.18	0.26	0.33	0.40	0.46	0.52	0.57	0.62	0.67	
4	LOH	4	0	0.18	0.33	0.46	0.57	0.67	0.75	0.82	0.89	0.95	1	
4	Het	3	0	0.14	0.25	0.35	0.43	0.50	0.56	0.62	0.67	0.71	0.75	
4	Het	1	0	0.05	0.08	0.12	0.14	0.17	0.19	0.21	0.22	0.24	0.25	
4	Het	2	0	0.09	0.17	0.23	0.29	0.33	0.38	0.41	0.44	0.47	0.50	
...														

20

\* 1~4の範囲のコピー数のみが示されているが、いかなるコピー数においても算出を行うことができる

\*\* 増分が0.1の腫瘍純度のみが示されるが、いかなる腫瘍純度においても算出を行うことができる

30

【0116】

割り当てられたMを用いて、以下のものによってサブクローン純度sを算出することができる。

【0117】

【数2】

$$s = \frac{AF_{観察}}{M} (pC_t + C_n(1-p))$$

40

【0118】

次いで、サブクローン純度sを腫瘍純度pで割ることによって前記試料レベルスコアを決定したが、前記試料レベルスコアは、身体内の変異アレルの実際の量を表す。

【0119】

50

【数3】

$$\text{試料レベルスコア} = \frac{s}{p}$$

【0120】

なお、前記観察頻度が期待頻度よりも大きい場合、前記変異はクローン変異であるとみなし、<sup>2</sup>試験を行わなかった。さらに、この状況下で、前記サブクローン純度は腫瘍純度よりも大きく、1の試料レベルスコアを割り当てる。表5に試料についての前記試料レベルスコアを示す。

【0121】

【表5】

試料についての試料レベルスコア

変異体	変異頻度	腫瘍純度	LOH	コピー数	クローナリティ	p-値	期待頻度	サブクローン純度	試料レベルスコア
ANK1 chr8_41552157_T>C	0.723	66.8	LOH	2	クローナリティ	観察>期待	0.668	0.723	1
CHD5 chr1_6194848_G>A	0.684	66.8	LOH	2	クローナリティ	観察>期待	0.668	0.684	1
SLC3A2 chr11_62652688_G>T	0.408	66.8	HET	2	クローナリティ	観察>期待	0.334	0.816	1
MRC1 chr10_18122699_G>A	0.407	66.8	HET	2	クローナリティ	観察>期待	0.334	0.814	1
TGS1 chr8_56698861_A>T	0.392	66.8	HET	2	クローナリティ	観察>期待	0.334	0.784	1
SPACA1 chr6_88769199_C>A	0.391	66.8	HET	2	クローナリティ	観察>期待	0.334	0.782	1
ARHGAP21 chr10_24896718_G>T	0.38	66.8	HET	2	クローナリティ	観察>期待	0.334	0.760	1
NSRP1 chr17_28511815_G>T	0.375	66.8	HET	2	クローナリティ	観察>期待	0.334	0.750	1
SCN3A chr2_165984466_C>T	0.373	66.8	HET	2	クローナリティ	観察>期待	0.334	0.746	1
TET3 chr2_74328527_G>A	0.373	66.8	HET	2	クローナリティ	観察>期待	0.334	0.746	1
ZDHH1 chr16_67440152_C>T	0.366	66.8	HET	2	クローナリティ	観察>期待	0.334	0.732	1
GOS2 chr1_209849154_T>G	0.363	66.8	HET	2	クローナリティ	観察>期待	0.334	0.726	1
FANCM chr14_45645307_A>T	0.354	66.8	HET	2	クローナリティ	観察>期待	0.334	0.708	1
NFASC chr1_204957844_G>A	0.326	66.8	HET	2	クローナリティ	0.637	0.334	0.652	0.977
SON chr21_34921835_A>G	0.636	66.8	LOH	2	クローナリティ	0.585	0.668	0.636	0.953
SSTR4 chr20_23017227_A>C	0.636	66.8	LOH	2	クローナリティ	0.209	0.668	0.636	0.953
TMEM104 chr17_72832365_C>T	0.317	66.8	HET	2	クローナリティ	0.555	0.334	0.634	0.950

【0122】

10

20

30

40

50

表5続き

変異体	変異頻度	腫瘍純度	LOH	コピー数	クローナリティ	p-値	期待頻度	サブクローン純度	試料レベルスコア
ZC3H3 chr8_144557605_C>A	0.315	66.8	HET	2	クローナリティ	0.432	0.334	0.630	0.944
KDELC2 chr11_108361810_T>C	0.315	66.8	HET	2	クローナリティ	0.765	0.334	0.630	0.944
OSBPL5 chr11_3125480_G>A	0.621	66.8	LOH	2	クローナリティ	0.197	0.668	0.621	0.930
TP53 chr17_7578445_A>T	0.584	66.8	LOH	2	サブクローナリティ	0.028	0.668	0.584	0.875
SP5 chr2_171572922_C>G	0.291	66.8	HET	2	クローナリティ	0.262	0.334	0.582	0.872
COL6A5 chr3_130150786_C>A	0.57	66.8	LOH	2	サブクローナリティ	0.006	0.668	0.570	0.854
LRRC8E chr19_7964551_G>A	0.564	66.8	LOH	2	サブクローナリティ	0.000	0.668	0.564	0.845
NUP160 chr11_47809795_T>C	0.274	66.8	HET	2	クローナリティ	0.213	0.334	0.548	0.821
ANO2 chr12_5685076_A>C	0.41	66.8	HET	3	サブクローナリティ	0.010	0.5	0.547	0.819
MUC16 chr19_9086656_C>T	0.539	66.8	LOH	2	サブクローナリティ	0	0.668	0.539	0.807
ARFGEF1 chr8_68208826_G>T	0.265	66.8	HET	2	クローナリティ	0.207	0.334	0.530	0.794
APC2 chr19_1456374_GA>TT	0.524	66.8	LOH	2	サブクローナリティ	0.000	0.668	0.524	0.785
CSE1L chr20_47675019_A>T	0.611	66.8	LOH	4	サブクローナリティ	0	0.801	0.509	0.763
KCNAS chr12_5154521_G>A	0.382	66.8	HET	3	サブクローナリティ	0.003	0.5	0.510	0.763
TRIM34 chr11_5664563_G>T	0.454	66.8	LOH	2	サブクローナリティ	0	0.668	0.454	0.680

10

20

30

【 0 1 2 3 】

40

50

表5続き

変異体	変異頻度	腫瘍純度	LOH	コピー数	クローナリティ	p-値	期待頻度	サブクローン純度	試料レベルスコア
CCDC141 chr2_179730534_C>T	0.399	66.8	LOH	2	サブクローナリティ	0	0.668	0.399	0.598
KIAA1429 chr8_95500959_C>A	0.395	66.8	LOH	2	サブクローナリティ	0	0.668	0.395	0.592
RNF17 chr13_25338358_C>T	0.39	66.8	LOH	2	サブクローナリティ	0	0.668	0.390	0.584
GBF1 chr10_104126933_A>G	0.379	66.8	LOH	2	サブクローナリティ	0	0.668	0.379	0.568
PRDX6 chr1_173454623_G>A	0.367	66.8	LOH	2	サブクローナリティ	0	0.668	0.367	0.550
GPR115 chr6_47681616_C>T	0.359	66.8	LOH	2	サブクローナリティ	0	0.668	0.359	0.538
DCHS1 chr11_6644247_C>A	0.354	66.8	LOH	2	サブクローナリティ	0	0.668	0.354	0.530
VPS13A chr9_79865083_A>T	0.35	66.8	LOH	2	サブクローナリティ	0	0.668	0.350	0.524
COL12A1 chr6_75851835_G>T	0.345	66.8	LOH	2	サブクローナリティ	0	0.668	0.345	0.517
EPN3 chr17_48616588_A>G	0.341	66.8	LOH	2	サブクローナリティ	0	0.668	0.341	0.511
SF3B1 chr2_198272747_T>C	0.33	66.8	LOH	2	サブクローナリティ	0	0.668	0.330	0.494
FAM65B chr6_24861205_C>A	0.326	66.8	LOH	2	サブクローナリティ	0	0.668	0.326	0.488
NF1 chr17_29509534_G>A	0.323	66.8	LOH	2	サブクローナリティ	0	0.668	0.323	0.484

10

20

30

【 0 1 2 4 】

40

50

表5続き

変異体	変異頻度	腫瘍純度	LOH	コピー数	クローナリティ	p-値	期待頻度	サブクローン純度	試料レベルスコア
C4BPA chr1_207317200_G>T	0.317	66.8	LOH	2	サブクローナリティ	0	0.668	0.317	0.475
DST chr6_56494089_T>A	0.303	66.8	LOH	2	サブクローナリティ	0	0.668	0.303	0.454
NALCN chr13_101944438_A>C	0.284	66.8	LOH	2	サブクローナリティ	0	0.668	0.284	0.425
NEK2 chr1_211836839_G>A	0.261	66.8	LOH	2	サブクローナリティ	0	0.668	0.261	0.391
HES1 chr3_193854752_C>G	0.293	66.8	LOH	3	サブクローナリティ	0	0.751	0.261	0.390
DOPEY1 chr6_83839042_C>T	0.245	66.8	LOH	2	サブクローナリティ	0	0.668	0.245	0.367
TATDN3 chr1_212981098_C>T	0.111	66.8	HET	2	サブクローナリティ	0.001	0.334	0.222	0.333
CCDC34 chr11_27360421_G>A	0.148	66.8	LOH	2	サブクローナリティ	0	0.668	0.148	0.222
SNX19 chr11_130785825_C>G	0.145	66.8	LOH	2	サブクローナリティ	0	0.668	0.145	0.217
CPS1 chr2_211441138_G>A	0.12	66.8	LOH	2	サブクローナリティ	0	0.668	0.120	0.180
PLOD1 chr1_12025537_G>A	0.118	66.8	LOH	2	サブクローナリティ	0	0.668	0.118	0.177
RBBP6 chr16_24579145_C>T	0.117	66.8	LOH	2	サブクローナリティ	0	0.668	0.117	0.175
DUT chr15_48634228_G>A	0.114	66.8	LOH	2	サブクローナリティ	0	0.668	0.114	0.171

【 0 1 2 5 】

10

20

30

40

50

表5続き

変異体	変異頻度	腫瘍純度	LOH	コピー数	クローナリティ	p-値	期待頻度	サブクローン純度	試料レベルスコア
LAMC1 chr1_183103819_G>A	0.109	66.8	LOH	2	サブクローナリティ	0	0.668	0.109	0.163
PELP1 chr17_4575375_C>T	0.109	66.8	LOH	2	サブクローナリティ	0	0.668	0.109	0.163
HAUS2 chr15_42858894_A>C	0.103	66.8	LOH	2	サブクローナリティ	0	0.668	0.103	0.154
RFC1 chr4_39322052_C>T	0.103	66.8	LOH	2	サブクローナリティ	0	0.668	0.103	0.154
TSSK2 chr22_19119814_C>T	0.102	66.8	LOH	2	サブクローナリティ	0	0.668	0.102	0.153
KBTBD12 chr3_127642629_G>A	0.1	66.8	LOH	2	サブクローナリティ	0	0.668	0.100	0.150
DCTN2 chr12_57939846_C>T	0.109	66.8	LOH	3	サブクローナリティ	0	0.751	0.097	0.145

10

20

【 0 1 2 6 】

( 実施例 1 6 . ペプチドレベルスコア及び試料レベルスコアを用いて免疫原としてのがん試料からのペプチドを同定し、スコア付け、ランク付けること )

【 0 1 2 7 】

がん試料についての免疫原としてのペプチドを同定し、スコア付け、ランク付けるために、前記がん試料からの前記ペプチドにおいて我々のスコア付け方法を適用した。前記試料についての前記エキソーム配列決定プロセスは、実施例 1 5 に記載されている。前記試料の体細胞変異を確認した後、8 ~ 23 のアミノ酸の長さを有する前記変異アミノ酸を有する前記ペプチドを検索した。次いで、実施例 1 2 に記載されたように、前記ペプチドレベルスコアを算出し、検索された前記ペプチドについて、実施例 1 5 に記載されたように、前記試料レベルスコアを算出した。前記ペプチド関連情報及び前記試料関連情報の両方を統合して免疫原として各ペプチドをランク付けるために、各ペプチドの前記ペプチドレベルスコア及び前記試料レベルスコアを乗算することによって、各ペプチドについての前記免疫原性スコアを算出した。トップ 5 0 の免疫原性スコアを有する胃がん試料からの前記ペプチドを選択し、表 6 に示した。

30

【 0 1 2 8 】

40

50

【表 6】

ランク	変異 ID	配列	ペプチドレベル スコア	試料レベル スコア	免疫原性 スコア
1	chr16_67440152_C>T	LYLFFAVIGFEILVPLPHHW	0.393	1.000	0.393
2	chr16_67440152_C>T	LYLFFAVIGFEILVPLL	0.390	1.000	0.390
3	chr16_67440152_C>T	HPLQIVAWLLYFFAVIGFEILV	0.381	1.000	0.381
4	chr17_28511815_G>T	RVNFRREKVIETPENDFKHHR	0.379	1.000	0.379
5	chr16_67440152_C>T	LYLFFAVIGFEI	0.379	1.000	0.379
6	chr10_18122699_G>A	STLTWHQAR	0.369	1.000	0.369
7	chr16_67440152_C>T	LYLFFAVIGFEIL	0.364	1.000	0.364
8	chr16_67440152_C>T	LYLFFAVIGFEILVPLPHHWV	0.363	1.000	0.363
9	chr2_165984466_C>T	KMQECFQKAFFR	0.359	1.000	0.359
10	chr16_67440152_C>T	LYLFFAVIGFEILV	0.358	1.000	0.358
11	chr17_28511815_G>T	RVNFRREKVIETPENDFK	0.352	1.000	0.352
12	chr16_67440152_C>T	AWLLYFFAVIGFEILVPLL	0.350	1.000	0.350
13	chr16_67440152_C>T	FFAVIGFEILVPLPHHW	0.350	1.000	0.350
14	chr17_28511815_G>T	RVNFRREK	0.349	1.000	0.349
15	chr17_72832365_C>T	MYTLNFACCDVVGLAAVRFFL	0.366	0.950	0.348
16	chr10_18122699_G>A	TLTWHQAR	0.345	1.000	0.345
17	chr8_41552157_T>C	KLVPLVQAAPFENAATKR	0.344	1.000	0.344
18	chr10_18122699_G>A	KSTLTWHQAR	0.343	1.000	0.343
19	chr16_67440152_C>T	AWLLYFFAVIGFEI	0.343	1.000	0.343
20	chr8_41552157_T>C	KSKLVPLVQAAPFENAATKR	0.340	1.000	0.340

10

20

30

【 0 1 2 9 】

40

50

表 6 続き

ランク	変異 ID	配列	ペプチドレベル スコア	試料レベル スコア	免疫原性 スコア
21	chr1_209849154_T>G	ALFGVVLGRMETVCSPTAAR	0.338	1.000	0.338
22	chr1_209849154_T>G	VVLGRMETVCSPTAAR	0.337	1.000	0.337
23	chr1_209849154_T>G	KLYVLGSLALFGVVLGR	0.337	1.000	0.337
24	chr10_18122699_G>A	STLTWHQARK	0.336	1.000	0.336
25	chr1_209849154_T>G	VVLGRMETVCSPTAARRLR	0.336	1.000	0.336
26	chr10_24896718_G>T	KTSAPLIRR	0.336	1.000	0.336
27	chr17_72832365_C>T	MYTLNFACCDVVGLAAVRFF	0.351	0.950	0.333
28	chr8_41552157_T>C	VPLVQAAFPENAV	0.333	1.000	0.333
29	chr2_165984466_C>T	KMQECFQKAFFRK	0.331	1.000	0.331
30	chr17_72832365_C>T	MYTLNFACCDVVGLAAVRF	0.348	0.950	0.331
31	chr1_209849154_T>G	ALFGVVLGR	0.330	1.000	0.330
32	chr2_74328527_G>A	SLTEKPWALR	0.329	1.000	0.329
33	chr17_28511815_G>T	KSSADDEIEETRVNFR	0.329	1.000	0.329
34	chr8_56698861_A>T	KYLVEIVQESW	0.329	1.000	0.329
35	chr1_209849154_T>G	KMKLYVLGSLALFGVVLGR	0.328	1.000	0.328
36	chr1_209849154_T>G	ALFGVVLGRMETVCSPTAARR	0.328	1.000	0.328
37	chr17_28511815_G>T	DEIEETRVNFRREKVIETPEND	0.327	1.000	0.327
38	chr17_72832365_C>T	MYTLNFACCDVVGLAAVRFFLGL	0.344	0.950	0.327
39	chr2_74328527_G>A	STSALAGPSLTEKPWALR	0.326	1.000	0.326
40	chr6_88769199_C>A	KYMWKLLRQDQSQSIILVNDSEI	0.326	1.000	0.326

【 0 1 3 0 】

10

20

30

40

50

表 6 続き

ランク	変異 ID	配列	ペプチドレベル スコア	試料レベル スコア	免疫原性 スコア
41	chr17_28511815_G>T	DDEIEETRVNFRREKVIETPEND	0.326	1.000	0.326
42	chr1_209849154_T>G	RMETVCSPFTAAR	0.326	1.000	0.326
43	chr2_165984466_C>T	KMQECFQKAFFRKP	0.324	1.000	0.324
44	chr10_24896718_G>T	KTSAPLIR	0.324	1.000	0.324
45	chr1_209849154_T>G	MVKLYVLGSLALFGVVLGR	0.323	1.000	0.323
46	chr1_209849154_T>G	VLALFGVVLGRMETVCSPFTAAR	0.323	1.000	0.323
47	chr17_28511815_G>T	VNFRREKVIETPEND	0.322	1.000	0.322
48	chr17_28511815_G>T	ADDEIEETRVNFRREKVIETPEN	0.320	1.000	0.320
49	chr2_74328527_G>A	ALAGPSLTKPWALR	0.320	1.000	0.320
50	chr1_209849154_T>G	SVLALFGVVLGR	0.320	1.000	0.320

10

20

## (付記)

本開示は以下の態様を含む。

< 1 > 少なくとも 1 種の免疫原性変異ペプチドを選択するための方法であって、

( a ) 複数の変異配列を得ること、

( b ) 疾患関連変異に由来する少なくとも 1 つのエピトープを同定すること、

( c ) 前記少なくとも 1 つのエピトープの免疫原性に関連する複数の因子を取り込むこと、

( d ) 前記複数の因子を重み付けること、

( e ) 前記複数の因子の重みに基づいて前記少なくとも 1 つのエピトープに免疫原性スコアを割り当てること、

( f ) 前記少なくとも 1 つのエピトープをランク付けすること、及び

( g ) 工程 ( f ) におけるランク付け結果に基づいて前記免疫原性変異ペプチドを選択すること、

を含み、

前記免疫原性変異ペプチドが、T 細胞応答を引き起こし得る少なくとも 1 つのエピトープを含む、方法。

< 2 > 工程 ( c ) ~ 工程 ( e ) の内の 1 つが、機械学習モデルを利用して達成される、  
< 1 > に記載の方法。

< 3 > 1 0 0 個以下のエピトープが選択される、< 1 > 又は < 2 > に記載の方法。

< 4 > 5 0 個以下のエピトープが選択される、< 3 > に記載の方法。

< 5 > 3 0 個以下のエピトープが選択される、< 4 > に記載の方法。

< 6 > 1 0 個以下のエピトープが選択される、< 5 > に記載の方法。

< 7 > 1 0 個 ~ 3 0 個のエピトープが選択される、< 5 > に記載の方法。

< 8 > 前記複数の因子が、MHC クラス I 及び MHC クラス II による前記エピトープの提示を対象を含む、< 1 > ~ < 7 > のいずれか一項に記載の方法。

< 9 > MHC クラス I を有する前記選択されたエピトープの結合親和性が、1 5 0 0 n M 未満の IC 5 0 値である、< 8 > に記載の方法。

30

40

50

< 1 0 > 前記複数の因子がMHCクラスI結合安定性を含む、< 8 >又は< 9 >に記載の方法。

< 1 1 > 前記複数の因子が、タンパク質存在量、遺伝子発現、又はそれらの組み合わせを含む、< 8 > ~ < 1 0 >のいずれか一項に記載の方法。

< 1 2 > 前記複数の因子が、細胞傷害性T細胞において免疫応答を引き起こす前記エピトープの能力を対象に含む、< 1 > ~ < 1 1 >のいずれか一項に記載の方法。

< 1 3 > 前記複数の因子が、ヘルパーT細胞における免疫応答を引き起こす前記エピトープの能力を対象に含む、< 1 > ~ < 1 2 >のいずれか一項に記載の方法。

< 1 4 > 前記複数の因子が、自己ペプチドに対する前記エピトープの類似性を含む、< 1 2 >又は< 1 3 >に記載の方法。

10

< 1 5 > 前記複数の因子が、既知の抗原に対する前記エピトープの相同性を含む、< 1 2 > ~ < 1 4 >のいずれか一項に記載の方法。

< 1 6 > 前記変異の変異頻度が少なくとも10%である、< 1 > ~ < 1 5 >のいずれか一項に記載の方法。

< 1 7 > 前記変異の前記変異頻度が少なくとも30%である、< 1 6 >に記載の方法。

< 1 8 > 前記変異が、2つ以上のコピーと共に存在する、< 1 > ~ < 1 7 >のいずれか一項に記載の方法。

< 1 9 > 前記複数の因子の内の1つがヘテロ接合性の消失である、< 1 > ~ < 1 8 >のいずれか一項に記載の方法。

< 2 0 > 前記複数の因子の内の1つがアレル本数である、< 1 > ~ < 1 9 >のいずれか一項に記載の方法。

20

< 2 1 > 前記複数の因子の内の1つが、前記疾患関連変異のクローナリティである、< 1 > ~ < 2 0 >のいずれか一項に記載の方法。

< 2 2 > 前記免疫原性スコアが、ペプチドレベルスコアを算出するために用いられ得る因子と、試料レベルスコアを算出するために用いられ得る因子と、を含む前記複数の因子が統合されてなる、< 1 > ~ < 2 1 >のいずれか一項に記載の方法。

< 2 3 > 前記免疫原性スコアが、MHCクラスI及びMHCクラスIIによる前記エピトープの提示と、ヘルパーT細胞及び細胞傷害性T細胞の両方において免疫応答を引き起こす前記エピトープの能力と、前記疾患関連変異のクローナリティと、を含む複数の因子が統合されてなる、< 1 >に記載の方法。

30

< 2 4 > 少なくとも1種の免疫原性変異ペプチドを選択するためのシステムであって、コンピュータにより実行可能な手段である

( a ) 複数の変異配列を得ること、

( b ) 疾患関連変異に由来する少なくとも1つのエピトープを同定すること、

( c ) 前記少なくとも1つのエピトープの免疫原性に関連する複数の因子を取り込むこと、

( d ) 前記複数の因子を重み付けること、

( e ) 前記複数の因子の重みに基づいて前記少なくとも1つのエピトープに免疫原性スコアを割り当てること、

( f ) 前記少なくとも1つのエピトープをランク付けすること、及び

40

( g ) 工程 ( f ) におけるランク付け結果に基づいて前記免疫原性変異ペプチドを選択すること

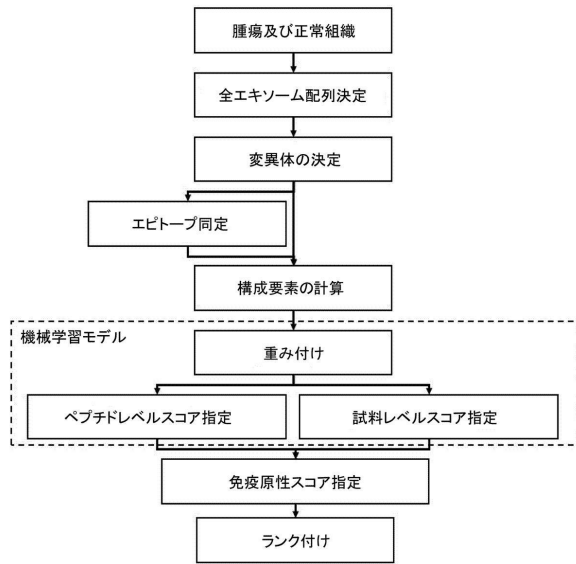
を記憶するハードウェアメモリを含み、

前記免疫原性変異ペプチドが、T細胞応答を引き起こし得る少なくとも1つのエピトープを含む、システム。

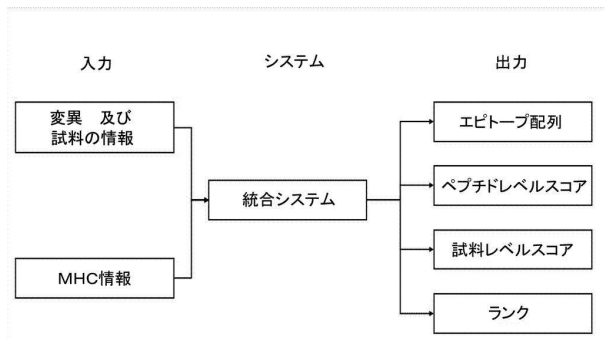
50

【 図 面 】

【 図 1 】

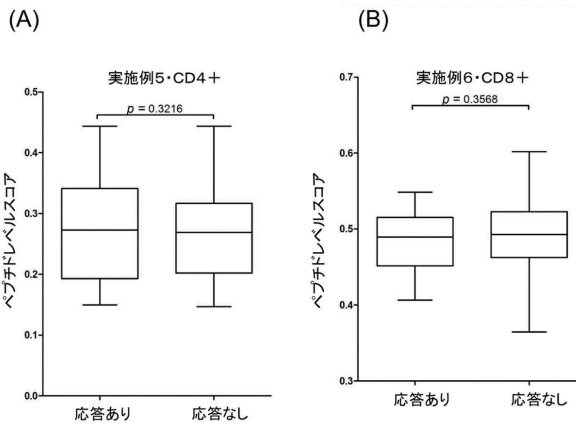


【 図 2 】

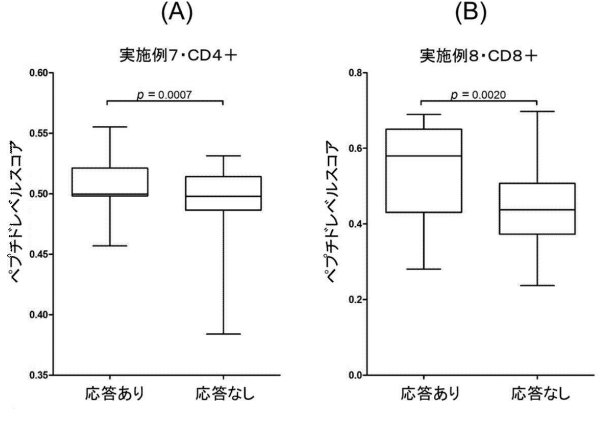


10

【 図 3 】



【 図 4 】



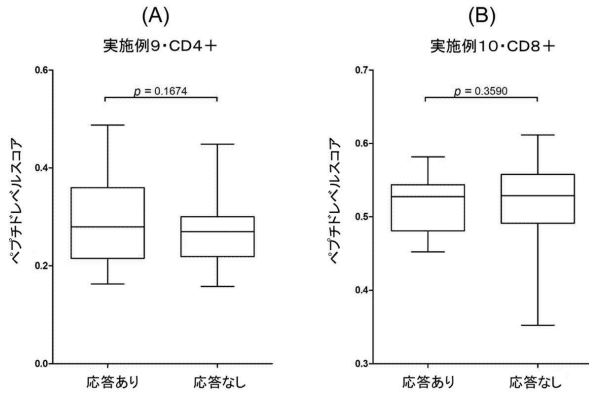
20

30

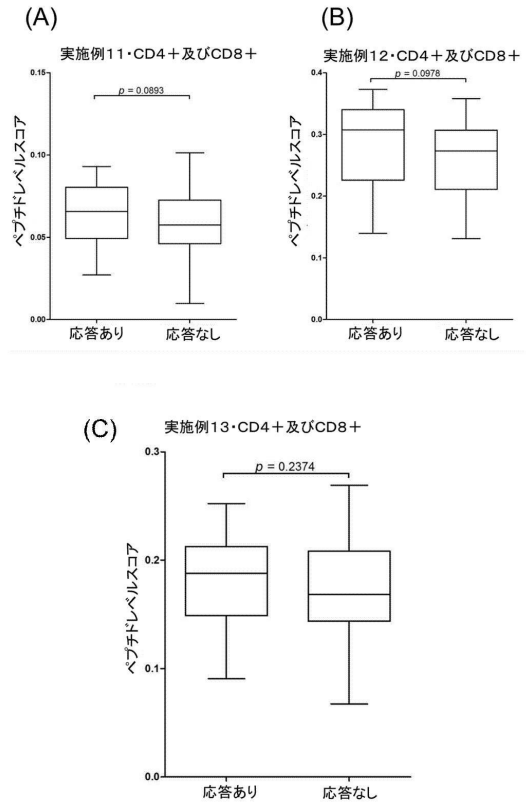
40

50

【 図 5 】



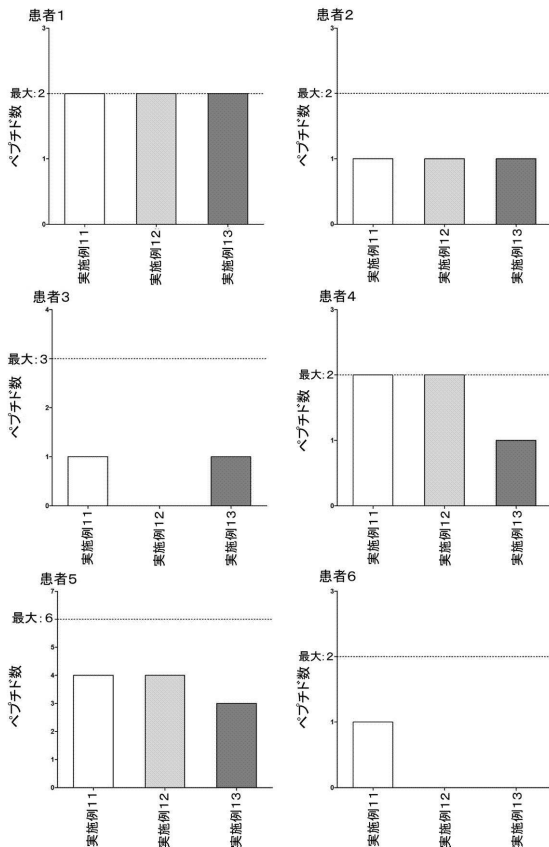
【 図 6 】



10

20

【 図 7 】



30

40

50

【配列表】

0007155470000001.app

10

20

30

40

50

## フロントページの続き

- (72)発明者 チェン、ジェン - ハオ  
台湾 114 タイペイ シティ ネイフー ディストリクト シンフー セカンド ロード ナンバー 2  
08 4エフ
- (72)発明者 チェン、イン - ジャ  
台湾 114 タイペイ シティ ネイフー ディストリクト シンフー セカンド ロード ナンバー 2  
08 4エフ
- (72)発明者 チェン、シュ - ジェン  
台湾 104 タイペイ シティ チョンサン ディストリクト ホージャン ストリート ナンバー 6  
9 - 1 2エフ - 1
- (72)発明者 チェン、フア - チェン  
台湾 106 タイペイ シティ ダーニ ディストリクト トウンホワ エス . ロード セクション 2  
レーン 63 ナンバー 17 4エフ

審査官 藤澤 美穂

- (56)参考文献 国際公開第2017/011660 (WO, A1)  
国際公開第2016/174085 (WO, A1)  
特開2005-301523 (JP, A)  
特表2016-506907 (JP, A)  
国際公開第2016/040682 (WO, A1)  
国際公開第2016/128376 (WO, A1)  
国際公開第2016/172722 (WO, A1)  
米国特許出願公開第2016/0101170 (US, A1)  
国際公開第2017/042394 (WO, A1)
- (58)調査した分野 (Int.Cl., DB名)  
G16B 5/00 - 99/00  
C12Q 1/686  
C12Q 1/6869  
C12N 15/12