(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7236543号
(P7236543)

(45)発行日　令和5年3月9日(2023.3.9)

(24)登録日　令和5年3月1日(2023.3.1)

(51)国際特許分類

Ｇ１６Ｂ	*40/20*	*(2019.01)*	Ｇ１６Ｂ	40/20
Ｃ１２Ｎ	*15/12*	*(2006.01)*	Ｃ１２Ｎ	15/12
Ｃ０７Ｋ	*14/47*	*(2006.01)*	Ｃ０７Ｋ	14/47
Ｃ１２Ｑ	*1/02*	*(2006.01)*	Ｃ１２Ｑ	1/02
Ｃ１２Ｑ	*1/68*	*(2018.01)*	Ｃ１２Ｑ	1/68

ＦＩ

ＺＮＡ

請求項の数　19　（全246頁）　最終頁に続く

(21)出願番号　特願2021-536205(P2021-536205)
(86)(22)出願日　令和1年12月20日(2019.12.20)
(65)公表番号　特表2022-518355(P2022-518355A)
(43)公表日　令和4年3月15日(2022.3.15)
(86)国際出願番号　PCT/US2019/068084
(87)国際公開番号　WO2020/132586
(87)国際公開日　令和2年6月25日(2020.6.25)
審査請求日　令和4年12月20日(2022.12.20)
(31)優先権主張番号　62/783,914
(32)優先日　平成30年12月21日(2018.12.21)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　62/826,827
(32)優先日　平成31年3月29日(2019.3.29)
最終頁に続く

(73)特許権者　518344612
ビオンテック ユーエス インコーポレイテッド
アメリカ合衆国 マサチューセッツ ０２１３９， ケンブリッジ， エリー ストリート ４０， スイート １１０
(74)代理人　100078282
弁理士　山本 秀策
(74)代理人　100113413
弁理士　森下 夏樹
(74)代理人　100181674
弁理士　飯田 貴敏
(74)代理人　100181641
弁理士　石川 大輔
(74)代理人　230113332
最終頁に続く

(54)【発明の名称】　ＨＬＡクラスＩＩ特異的エピトープの予測およびＣＤ４＋ Ｔ細胞の特徴付けのための方法およびシステム

(57)【特許請求の範囲】

【請求項１】

がんの処置を必要とするヒト対象においてがんを処置するための候補ペプチド配列のセットのペプチド配列のサブセットを選択する、コンピュータで実行される方法であって、

（ａ）コンピュータプロセッサを使用して、候補ペプチド配列のセットのアミノ酸情報を、訓練された機械学習クラスＩＩ　ＨＬＡ－ペプチド提示予測モデルに入力して、複数の提示予測を生成するステップであって、候補ペプチド配列の前記セットの各候補ペプチド配列が、対象のゲノム、トランスクリプトームもしくはエクソーム、または前記対象における病原体もしくはウイルスによってコードされ；前記複数の提示予測が、候補ペプチド配列の前記セットの各候補ペプチド配列についてのＨＬＡ提示予測を含み、前記複数の提示予測のうちの各提示予測により、候補ペプチド配列の前記セットのうちのペプチド配列が、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質により提示される提示可能性が示され、前記訓練された機械学習クラスＩＩ　ＨＬＡ－ペプチド提示予測モデルが、

（ｉ）以下：

（１）訓練用ペプチドの配列、

（２）前記訓練用ペプチド配列と関連するＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質がいかなるタンパク質であるか、および、

（３）前記訓練用ペプチドのうちの１つまたは複数が訓練用細胞中の前記ＨＬＡクラスＩＩ対立遺伝子によってコードされる前記タンパク質によって提示された、質量分析によ

る観察
を含む複数のパラメータ、を含む訓練用データに少なくとも基づいて同定される複数のパラメーター；および
（ｉｉ）入力として受け取られた前記アミノ酸配列情報と、前記アミノ酸配列情報および前記複数のパラメータに基づいて出力として生成された提示可能性との間の関係を提示する関数
を含む、ステップと、
（ｂ）少なくとも前記複数の提示予測に基づいて、候補ペプチド配列の前記セットのうちペプチド配列のサブセットを選択して、選択されたペプチド配列のセットを生成するステップであって、ペプチド配列の前記選択されたサブセットは、
（ｉ）対象のゲノム、トランスクリプトームもしくはエクソーム、または前記対象における病原体もしくはウイルスによってコードされる、および
（ｉｉ）前記対象の細胞の前記ＨＬＡクラスＩＩ対立遺伝子によってコードされる前記タンパク質によって提示されると予測される；
ここで、医薬組成物は、前記ヒト対象に投与されるべきであることが示され、前記医薬組成物は、（ｉ）ペプチド配列の前記選択されたサブセットのうちの１つまたは複数を有するポリペプチド、（ｉｉ）（ｉ）のポリペプチドをコードするポリヌクレオチド、（ｉｉｉ）（ｉ）もしくは（ｉｉ）を含むＡＰＣ、または（ｉｖ）（ｂ）において選択されたペプチド配列と複合体を形成した前記ヒト対象のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされるタンパク質に特異的なＴ細胞受容体（ＴＣＲ）を含むＴ細胞を含み；前記訓練された機械学習ＨＬＡ－ペプチド提示予測モデルの陽性適中率（ＰＰＶ）が、提示ＰＰＶ決定法に従うと１０％のリコール率で少なくとも０．２である、方法。
【請求項２】
訓練用細胞中で前記ＨＬＡクラスＩＩによってコードされる前記タンパク質によって提示されると質量分析によって観察された訓練用ペプチドの前記配列の各配列が、少なくとも１５アミノ酸の長さを有する、請求項１に記載の方法。
【請求項３】
訓練された機械学習クラスＩＩ　ＨＬＡ－ペプチド提示予測モデルの陽性適中率（ＰＰＶ）が、提示ＰＰＶ決定法に従うと少なくとも０．３である、請求項１に記載の方法。
【請求項４】
（ｉ）前記訓練用細胞が、単一ＭＨＣクラスＩＩ複合体またはＤＲ、ＤＰおよびＤＱからなる群から選択されるクラスＩＩ　ＨＬＡ遺伝子座の単一対立遺伝子バリアントによってコードされるタンパク質を発現する訓練用細胞を含み、前記単一ＭＨＣクラスＩＩ複合体またはクラスＩＩ　ＨＬＡ遺伝子座の前記単一対立遺伝子バリアントによってコードされるタンパク質が、前記対象の細胞によって発現される；または（ｉｉ）前記訓練用データが、逆畳み込みによって得られる訓練用データを含む、請求項３に記載の方法。
【請求項５】
前記訓練用細胞が、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされるタンパク質を発現し、クラスＩＩ　ＨＬＡ対立遺伝子によってコードされる前記タンパク質が、親和性タグを含む、請求項３に記載の方法。
【請求項６】
複数の試験ペプチド配列のアミノ酸情報を処理して、複数の試験提示予測を生成し、各試験提示予測により、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる前記１つまたは複数のタンパク質により前記複数の試験ペプチド配列のうちの所与の試験ペプチド配列が提示され得る可能性が示される場合に、前記機械学習ＨＬＡ－ペプチド提示予測モデルの陽性適中率（ＰＰＶ）が少なくとも０．２であり、前記複数の試験ペプチド配列が、（ｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された少なくとも１つのヒットペプチド配列と、（ｉｉ）生物体のゲノムによってコードされるタンパク質内に含有される少なくとも４９９のデコイペプチド配列とを含む少なくとも５００の試験ペプチド配列を含み、前記生物体と前記対象が

同じ種であり、前記複数の試験ペプチド配列が、前記少なくとも１つのヒットペプチド配列と前記少なくとも４９９のデコイペプチド配列を１：４９９の比で含み、前記機械学習ＨＬＡ－ペプチド提示予測モデルにより、前記複数の試験ペプチド配列の上位０．２％が、細胞において発現される前記ＨＬＡタンパク質によって提示されることが予測される、請求項３に記載の方法。

【請求項７】

（ｉ）前記少なくとも１つのヒットペプチド配列が、少なくとも１０ヒットペプチド配列を含み、（ｉｉ）前記少なくとも４９９のデコイペプチド配列が、少なくとも４９９０のデコイペプチド配列を含む、請求項６に記載の方法。

【請求項８】

前記少なくとも１つのヒットペプチドのうちのいずれかの任意の９個連続するアミノ酸サブ配列が、前記少なくとも４９９０のデコイペプチド配列のうちの任意の９個連続するアミノ酸サブ配列と重複しない、請求項７に記載の方法。

【請求項９】

候補ペプチド配列の前記セットのうちの各ペプチド配列が、がんと関連する、請求項１に記載の方法。

【請求項１０】

候補ペプチド配列の前記セットのうちの各ペプチド配列が、（ｉ）変異を含む、（ｉｉ）前記対象のがん細胞において発現される、および（ｉｉｉ）前記対象の非がん細胞のゲノムによってコードされない、請求項９に記載の方法。

【請求項１１】

クラスＩＩ　ＨＬＡ対立遺伝子によってコードされる前記タンパク質が、ＨＬＡ－ＤＰＢ１＊０１：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０２：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０３：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０２／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０６：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＲＢ１＊０１：０１、ＨＬＡ－ＤＲＢ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０３：０１、ＨＬＡ－ＤＲＢ１＊０３：０２、ＨＬＡ－ＤＲＢ１＊０４：０１、ＨＬＡ－ＤＲＢ１＊０４：０２、ＨＬＡ－ＤＲＢ１＊０４：０３、ＨＬＡ－ＤＲＢ１＊０４：０４、ＨＬＡ－ＤＲＢ１＊０４：０５、ＨＬＡ－ＤＲＢ１＊０４：０７、ＨＬＡ－ＤＲＢ１＊０７：０１、ＨＬＡ－ＤＲＢ１＊０８：０１、ＨＬＡ－ＤＲＢ１＊０８：０２、ＨＬＡ－ＤＲＢ１＊０８：０３、ＨＬＡ－ＤＲＢ１＊０８：０４、ＨＬＡ－ＤＲＢ１＊０９：０１、ＨＬＡ－ＤＲＢ１＊１０：０１、ＨＬＡ－ＤＲＢ１＊１１：０１、ＨＬＡ－ＤＲＢ１＊１１：０２、ＨＬＡ－ＤＲＢ１＊１１：０４、ＨＬＡ－ＤＲＢ１＊１２：０１、ＨＬＡ－ＤＲＢ１＊１２：０２、ＨＬＡ－ＤＲＢ１＊１３：０１、ＨＬＡ－ＤＲＢ１＊１３：０２、ＨＬＡ－ＤＲＢ１＊１３：０３、ＨＬＡ－ＤＲＢ１＊１４：０１、ＨＬＡ－ＤＲＢ１＊１５：０１、ＨＬＡ－ＤＲＢ１＊１５：０２、ＨＬＡ－ＤＲＢ１＊１５：０３、ＨＬＡ－ＤＲＢ１＊１６：０１、ＨＬＡ－ＤＲＢ３＊０１：０１、ＨＬＡ－ＤＲＢ３＊０２：０２、ＨＬＡ－ＤＲＢ３＊０３：０１、ＨＬＡ－ＤＲＢ４＊０１：０１、ＨＬＡ－ＤＲＢ５＊０１：０１、ＨＬＡ－ＤＲＢ１＊０１：０１、ＨＬＡ－ＤＲＢ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０３：０１、ＨＬＡ－ＤＲＢ１＊０４：０１、ＨＬＡ－ＤＲＢ１＊０４：０２、ＨＬＡ－ＤＲＢ１＊０４：０４、ＨＬＡ－ＤＲＢ１＊０４：０５、ＨＬＡ－ＤＲＢ１＊０７：０１、ＨＬＡ－ＤＲＢ１＊０８：０１、ＨＬＡ－ＤＲＢ１＊０８：０２、ＨＬＡ－ＤＲＢ１＊０８：０３、ＨＬＡ－ＤＲＢ１＊０９：０１、ＨＬＡ－ＤＲＢ１＊１１：０１、ＨＬＡ－ＤＲＢ１＊１１：０２、ＨＬＡ－ＤＲＢ１＊１１：０４、ＨＬＡ－ＤＲＢ１＊１２：０１、ＨＬＡ－ＤＲＢ１＊１３：０１、ＨＬＡ－ＤＲＢ１＊１３：０２、ＨＬＡ－ＤＲＢ１＊１３：０３、ＨＬＡ－ＤＲＢ１＊１４：０１、ＨＬＡ－ＤＲＢ１＊１５：０１、ＨＬＡ－ＤＲＢ１＊１５：０２、ＨＬＡ－ＤＲＢ１＊１５：０３、ＨＬＡ－ＤＲＢ１＊１６：０２、ＨＬＡ－ＤＲＢ３＊０１：０１、ＨＬＡ－ＤＲＢ３＊０２：０１、ＨＬＡ－ＤＲＢ３＊０２：０２、ＨＬＡ－ＤＲＢ３＊０３：０１、ＨＬＡ－ＤＲＢ４＊０１：０１、ＨＬＡ－ＤＲＢ４＊

０１：０３、ＨＬＡ－ＤＲＢ５＊０１：０１；ＨＬＡ－ＤＰＢ１＊０１：０１、ＨＬＡ－ＤＰＢ１＊０２：０１、ＨＬＡ－ＤＰＢ１＊０２：０２、ＨＬＡ－ＤＰＢ１＊０３：０１、ＨＬＡ－ＤＰＢ１＊０４：０１、ＨＬＡ－ＤＰＢ１＊０４：０２、ＨＬＡ－ＤＰＢ１＊０５：０１、ＨＬＡ－ＤＰＢ１＊０６：０１、ＨＬＡ－ＤＰＢ１＊１１：０１、ＨＬＡ－ＤＰＢ１＊１３：０１、ＨＬＡ－ＤＰＢ１＊１７：０１、ＨＬＡ－ＤＱＡ１＊０１：０１／ＨＬＡ－ＤＱＢ１＊０５：０１、ＨＬＡ－ＤＱＡ１＊０１：０２／ＨＬＡ－ＤＱＢ１＊０６：０２、ＨＬＡ－ＤＱＡ１＊０１：０２／ＨＬＡ－ＤＱＢ１＊０６：０４、ＨＬＡ－ＤＱＡ１＊０１：０３／ＨＬＡ－ＤＱＢ１＊０６：０３、ＨＬＡ－ＤＱＡ１＊０２：０１／ＨＬＡ－ＤＱＢ１＊０２：０２、ＨＬＡ－ＤＱＡ１＊０２：０１／ＨＬＡ－ＤＱＢ１＊０３：０３、ＨＬＡ－ＤＱＡ１＊０３：０１／ＨＬＡ－ＤＱＢ１＊０３：０２、ＨＬＡ－ＤＱＡ１＊０３：０３／ＨＬＡ－ＤＱＢ１＊０３：０１、ＨＬＡ－ＤＱＡ１＊０５：０１／ＨＬＡ－ＤＱＢ１＊０２：０１、ＨＬＡ－ＤＱＡ１＊０５：０５／ＨＬＡ－ＤＱＢ１＊０３：０１、ならびにそれらの任意の組み合わせからなる群から選択される、請求項１に記載の方法。

【請求項１２】

前記方法が、ゲノム、トランスクリプトーム、またはエキソーム配列決定によって前記対象の複数のポリヌクレオチド配列を得ることを含み、前記複数のポリヌクレオチド配列が、候補ペプチド配列の前記セットをコードする、請求項１に記載の方法。

【請求項１３】

前記ゲノム、トランスクリプトーム、またはエキソーム配列決定が、全ゲノム配列決定、全トランスクリプトーム配列決定、または全エクソーム配列決定である、請求項１２に記載の方法。

【請求項１４】

候補ペプチド配列の前記セットが、前記複数の提示予測に基づいて順位付けされる、請求項１に記載の方法。

【請求項１５】

選択されたペプチド配列の前記セットの前記ペプチド配列の各々が、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされるタンパク質に、５００ｎＭもしくはそれ未満のＩＣ５０または５００ｎＭもしくはそれ未満の推定ＩＣ５０で結合する、請求項１に記載の方法。

【請求項１６】

前記医薬組成物が、前記選択されたペプチド配列のうちの少なくとも２つを含む１もしくは複数のポリペプチド、または前記選択されたペプチド配列のうちの少なくとも２つをコードする１もしくは複数のポリヌクレオチドを含む、請求項１に記載の方法。

【請求項１７】

前記医薬組成物が、アジュバントをさらに含む、請求項１６に記載の方法。

【請求項１８】

前記医薬組成物が、前記対象においてＣＤ４＋　Ｔ細胞応答および／またはＣＤ８＋　Ｔ細胞応答を引き出す、請求項１６に記載の方法。

【請求項１９】

前記訓練された機械学習クラスＩＩ　ＨＬＡ－ペプチド提示予測モデルの陽性適中率（ＰＰＶ）が、提示ＰＰＶ決定法に従うと、２０％のリコール率で、少なくとも０．２である、請求項１に記載の方法。

【発明の詳細な説明】

【技術分野】

【０００１】

相互参照

本出願は、それぞれの全体が参照により本明細書に組み込まれる、２０１８年１２月２１日出願の米国仮出願第６２／７８３，９１４号；２０１９年３月２９日出願の米国仮出願第６２／８２６，８２７号；２０１９年５月３１日出願の米国仮出願第６２／８５５，３７９号；および２０１９年８月２３日出願の米国仮出願第６２／８９１，１０１号の利

益を主張する。
【背景技術】
【０００２】
背景
主要組織適合性遺伝子複合体（ＭＨＣ）は、ヒト白血球抗原（ＨＬＡ）遺伝子をコードする遺伝子複合体である。ＨＬＡ遺伝子は、循環Ｔ細胞に対してヒト細胞の表面上にディスプレイされるタンパク質ヘテロ二量体として発現される。ＨＬＡ遺伝子は、高度に多型的であり、それにより、ＨＬＡ遺伝子による適応免疫系の微調整が可能になる。適応免疫応答は、一部において、ヒト白血球抗原（ＨＬＡ）ヘテロ二量体と結合した疾患関連ペプチド抗原をディスプレイする細胞を同定し、排除するＴ細胞の能力に依拠する。
【０００３】
ヒトでは、内因性および外因性タンパク質は、プロテアソームによって、ならびに細胞質基質およびエンドソーム／リソソームのプロテアーゼおよびペプチダーゼによってプロセシングされてペプチドになり、ＭＨＣ遺伝子によってコードされる２つのクラスの細胞表面タンパク質によって提示される。これらの細胞表面タンパク質は、ヒト白血球抗原（ＨＬＡクラスＩおよびクラスＩＩ）と称され、それらに結合し、免疫応答を引き出すペプチドの群はＨＬＡエピトープと称される。ＨＬＡエピトープは、免疫系が病原体感染および自己がん化（ｔｒａｎｓｆｏｒｍａｔｉｏｎ ｏｆ ｓｅｌｆ）などの危険シグナルを検出することを可能にする重要な構成成分である。ＣＤ４＋ Ｔ細胞は、樹状細胞およびマクロファージなどの抗原提示細胞（ＡＰＣ）上にディスプレイされるクラスＩＩ ＭＨＣ（ＨＬＡ－ＤＲ、ＨＬＡ－ＤＱ、およびＨＬＡ－ＤＰ）エピトープを認識する。ＨＬＡクラスＩＩ－リガンドの内因性プロセシングおよび提示は複雑な過程であり、種々のシャペロン、および酵素のサブセットを必要とするが、それらの全てが十分に特徴付けられているわけではない。ＨＬＡクラスＩＩ－ペプチドの提示により、ヘルパーＴ細胞が活性化され、その後、Ｂ細胞分化および抗体産生ならびにＣＴＬ応答が促進される。活性化されたヘルパーＴ細胞により、他のＴ細胞を活性化し、その分化を誘導するサイトカインおよびケモカインの分泌もなされる。
【０００４】
ＨＬＡクラスＩＩヘテロ二量体毎のペプチド結合の優先性を理解することが、いずれのがんまたは腫瘍特異的抗原によりがんまたは腫瘍特異的Ｔ細胞応答が引き出される可能性が高いかを首尾よく予測するための鍵になる。特定のＨＬＡクラスＩＩ関連ペプチド（例えば、新抗原ペプチド）を同定し、単離する方法が必要とされている。そのような方法体系および単離された分子は、例えば、これだけに限定されないが、免疫に基づく治療薬を含めた治療薬の開発に有用である。
【発明の概要】
【課題を解決するための手段】
【０００５】
概要
本明細書に記載の方法および組成物は、広範囲の適用に使用される。例えば、本明細書に記載の方法および組成物を、免疫原性抗原ペプチドを同定するために使用することができ、また、個別化医薬品などの薬物の開発、ならびに抗原特異的Ｔ細胞の単離および特徴付けのために使用することができる。
【０００６】
ＣＤ４＋ Ｔ細胞応答は、抗腫瘍活性を有し得る。クラスＩＩ予測を使用しない場合、ＣＤ４＋ Ｔ細胞応答が高い率で示される可能性がある（例えば、ＮｅｏＶａｘによる試験においてＳＬＰエピトープの６０％（ＮＴ－００１において４９％、Ott et al., Nature, 2017 Jul 13; 547 (7662): 217-221を参照されたい）、およびＢｉｏｎｔｅｃｈによる試験においてｍＲＮＡエピトープの４８％、Sahin et al., Nature, 2017 Jul 13; 547 (7662): 222-226を参照されたい）。これらのエピトープが一般にはネイティブなまま（腫瘍によってまたは食作用性ＤＣによって）提示されるかどうかは

明らかでない場合がある。真に提示されるＨＬＡクラスＩＩ結合性エピトープの同定を改善することにより、高ＣＤ４＋　Ｔ応答率を治療有効性に変換することが望ましい可能性がある。

【０００７】

遺伝子発現、酵素による切断、および経路／局在化の偏りの役割は、頑強には数量化されていない可能性がある。より関連性のある経路がオートファジー（腫瘍細胞によるＨＬＡクラスＩＩ提示）であるかファゴサイトーシス（ＡＰＣによる腫瘍エピトープのＨＬＡクラスＩＩ提示）であるかは不明であり得るが、大多数の既存のＭＳデータはオートファジーに由来すると推定することができる。ＮｅｔＭＨＣＩＩｐａｎが現行の予測標準であり得るが、これは正確ではないと考えられる場合がある。３つのＨＬＡクラスＩＩ遺伝子座（ＤＲ、ＤＰ、およびＤＱ）のうち、ＨＬＡ－ＤＲのある特定の共通の対立遺伝子に関してのみデータが存在し得る。

【０００８】

ＨＬＡ　Ｃｌａｓｓ　ＩＩ提示の規則の学習に関しては、当該分野の標準および本提唱手法を含めた、種々のデータ生成手法が存在し得る。当該分野の標準は、ＮｅｔＭＨＣＩＩｐａｎ予測器の基礎であり得る親和性測定、ロースループットがもたらされることおよび放射性試薬の必要性を含み得、また、当該分野の標準では、プロセシングの役割が欠損している。本提唱手法は、質量分析を含み得、その場合、細胞株／組織／腫瘍からのデータがオートファジーに関するプロセシング規則の決定に役立ち得、また、単一対立遺伝子ＭＳにより、対立遺伝子特異的結合規則の決定が可能になり得る（複対立遺伝子ＭＳデータは、効率的な学習のためにはあまりにも複雑であることが推定される（Bassani-Sternberg. MCP. 2018））。

【０００９】

新しいＨＬＡクラスＩＩ予測器を検証するための以下のような種々のやり方があり得る：デフォルト設定であり得る、提出されたＭＳデータに対する検証；免疫モニタリングデータにより、腫瘍提示ではなくＡＰＣに対するワクチンペプチドローディングを評価することができ、データを多くの異なる対立遺伝子にわたって薄く広げることができる遡及的ワクチン試験（例えば、ＮＴ－００１）；不調和に予測されるペプチド（２～３の対立遺伝子に対してのみ）についての測定値を得るように構成することができる生化学的親和性測定；Ｎｅｏｎ優先エピトープおよびＮｅｔＭＨＣＩＩｐａｎ優先エピトープによりｅｘｖｉｖｏにおけるＴ細胞応答が誘導される率を試験するために構成することができるＴ細胞誘導。

【００１０】

Ｔ細胞誘導による検証に関しては、デフォルトの手法は、不調和に予測されるＴＣＧＡのｎｅｏＯＲＦを評価することを含み得、誘導材料は、健康なドナーＡＰＣおよびＴ細胞を含み得、誘導および読み取りは、ＳＬＰ（約１５ｍｅｒのペプチド）によるものであり得る。ランダムなペプチドでは高い応答率が生じる可能性があり、また、ＳＬＰではプロセシングへの対処が不十分である可能性がある。可能性のある解決法は、ｍＲＮＡによる誘導を含み得る。

【００１１】

本明細書に開示される方法は、エピトープ予測のための対立遺伝子特異的機械学習法の訓練のためのＬＣ－ＭＳ／ＭＳ単一対立遺伝子データを生成することを含み得る。そのような方法は、予測モデルの性能を向上させる偽陽性を厳密に取り除くために品質メトリクスのセットを利用してＬＣ－ＭＳ／ＭＳデータ品質を向上させること；ＨＬＡ－リガンドームＬＣ－ＭＳ／ＭＳデータセットから対立遺伝子特異的ＨＬＡクラスＩＩ結合性コアを同定すること；機械学習アルゴリズムを利用してＨＬＡクラスＩＩ－リガンドおよびエピトープ予測を改善すること；ならびに／または、ＨＬＡクラスＩＩ－リガンド提示に影響を及ぼし、ＨＬＡクラスＩＩエピトープ予測を改善する生物学的変数、例えば、遺伝子発現、切断性、遺伝子の偏り、細胞の局在化、および二次構造などを同定することを含み得る。

【００１２】
（ａ）複数の提示予測を生成するために、機械学習ＨＬＡペプチド提示予測モデルを使用して複数の候補ペプチド配列のアミノ酸情報を処理するステップであって、複数の候補ペプチド配列の各候補ペプチド配列は、対象のゲノムもしくはエクソームによってコードされ、複数の提示予測は、複数の候補ペプチド配列の各々についてのＨＬＡ提示予測を含み、各ＨＬＡ提示予測により、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数の候補ペプチド配列のうちの所与の候補ペプチド配列が提示され得る可能性が示され、機械学習ＨＬＡペプチド提示予測モデルは、訓練用細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された訓練用ペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップと、（ｂ）少なくとも複数の提示予測に基づいて、複数のペプチド配列のうち、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１つによって提示されるペプチド配列を同定するステップと、を含み、機械学習ＨＬＡペプチド提示予測モデルの陽性適中率（ＰＰＶ）は、提示ＰＰＶ決定法に従うと少なくとも０．０７である、方法が本明細書で提供される。
【００１３】
（ａ）複数の結合予測を生成するために、機械学習ＨＬＡペプチド結合予測モデルを使用して対象のゲノムまたはエクソームによってコードされる複数のペプチド配列のアミノ酸情報を処理するステップであって、複数の結合予測は、複数の候補ペプチド配列の各々についてのＨＬＡ結合予測を含み、各結合予測により、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質が複数の候補ペプチド配列のうちの所与の候補ペプチド配列に結合する可能性が示され、機械学習ＨＬＡペプチド結合予測モデルは、ＨＬＡクラスＩＩタンパク質またはＨＬＡクラスＩＩタンパク質類似体に結合することが同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップと、（ｂ）少なくとも複数の結合予測に基づいて、複数のペプチド配列のうち、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１つとの結合に関する閾値結合予測確率値よりも大きな確率を有するペプチド配列を同定するステップとを含み、機械学習ＨＬＡペプチド結合予測モデルの陽性適中率（ＰＰＶ）は、結合ＰＰＶ決定法に従うと少なくとも０．１である、方法が本明細書で提供される。
【００１４】
一部の実施形態では、機械学習ＨＬＡペプチド提示予測モデルは、訓練用細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された訓練用ペプチドの配列の配列情報を含む訓練用データを使用して訓練される。
【００１５】
一部の実施形態では、方法は、提示予測に基づいて、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１つによって提示されることが同定された少なくとも２つのペプチドに順位を付けることを含む。
【００１６】
一部の実施形態では、方法は、２つまたはそれよりも多くの順位付けされたペプチドのうちの１つまたは複数のペプチドを選択することを含む。
【００１７】
一部の実施形態では、方法は、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１つによって提示されることが同定された複数のうちの１つまたは複数のペプチドを選択することを含む。
【００１８】
一部の実施形態では、方法は、提示予測に基づいて順位付けされた２つまたはそれよりも多くのペプチドのうちの１つまたは複数のペプチドを選択することを含む。
【００１９】
一部の実施形態では、複数の試験ペプチド配列のアミノ酸情報を処理して、複数の試験

提示予測を生成し、各試験提示予測により、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数の試験ペプチド配列のうちの所与の試験ペプチド配列が提示され得る可能性が示される場合に、機械学習ＨＬＡペプチド提示予測モデルの陽性適中率（ＰＰＶ）が少なくとも０．０７であり、複数の試験ペプチド配列は、（ｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された少なくとも１つのヒットペプチド配列と、（ｉｉ）生物体のゲノムによってコードされるタンパク質内に含有される少なくとも４９９のデコイペプチド配列とを含む少なくとも５００の試験ペプチド配列を含み、生物体と対象が同じ種であり、複数の試験ペプチド配列は、少なくとも１つのヒットペプチド配列と少なくとも４９９のデコイペプチド配列を１：４９９の比で含み、機械学習ＨＬＡペプチド提示予測モデルにより、複数の試験ペプチド配列の上位パーセンテージは、細胞において発現されるＨＬＡタンパク質によって提示されることが予測される。

【００２０】

一部の実施形態では、複数の試験ペプチド配列のアミノ酸情報を処理して、複数の試験結合予測を生成し、各試験結合予測により、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質が複数の試験ペプチド配列のうちの所与の試験ペプチド配列に結合する可能性が示される場合に、機械学習ＨＬＡペプチド提示予測モデルの陽性適中率（ＰＰＶ）が少なくとも０．１であり、複数の試験ペプチド配列は、（ｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された少なくとも１つのヒットペプチド配列と、（ｉｉ）細胞において発現されるＨＬＡタンパク質、例えば、細胞において発現される単一ＨＬＡタンパク質（例えば、単一対立遺伝子細胞）によって提示されることが質量分析によって同定された少なくとも１つのペプチド配列を含むタンパク質内に含有される少なくとも１９のデコイペプチド配列とを含む少なくとも２０の試験ペプチド配列を含み、複数の試験ペプチド配列は、少なくとも１つのヒットペプチド配列と少なくとも１９のデコイペプチド配列を１：１９の比で含み、機械学習ＨＬＡペプチド提示予測モデルにより、複数の試験ペプチド配列の上位パーセンテージが細胞において発現されるＨＬＡタンパク質に結合することが予測される。

【００２１】

一部の実施形態では、少なくとも１つのヒットペプチド配列とデコイペプチド配列の間にアミノ酸配列の重複が存在しない。

【００２２】

一部の実施形態では、機械学習ＨＬＡペプチド提示予測モデルの陽性適中率（ＰＰＶ）は、少なくとも０．０８、０．０９、０．１、０．１１、０．１２、０．１３、０．１４、０．１５、０．１６、０．１７、０．１８、０．１９、０．２、０．２１、０．２２、０．２３、０．２４、０．２５、０．２６、０．２７、０．２８、０．２９、０．３、０．３１、０．３２、０．３３、０．３４、０．３５、０．３６、０．３７、０．３８、０．３９、０．４、０．４１、０．４２、０．４３、０．４４、０．４５、０．４６、０．４７、０．４８、０．４９、０．５、０．５１、０．５２、０．５３、０．５４、０．５５、０．５６、０．５７、０．５８、０．５９、０．６、０．６１、０．６２、０．６３、０．６４、０．６５、０．６６、０．６７、０．６８、０．６９、０．７、０．７１、０．７２、０．７３、０．７４、０．７５、０．７６、０．７７、０．７８、０．７９、０．８、０．８１、０．８２、０．８３、０．８４、０．８５、０．８６、０．８７、０．８８、０．８９、０．９、０．９１、０．９２、０．９３、０．９４、０．９５、０．９６、０．９７、０．９８または０．９９である。

【００２３】

一部の実施形態では、少なくとも１つのヒットペプチド配列は、少なくとも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、４

６、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または１００のヒットペプチド配列を含む。

【００２４】

一部の実施形態では、少なくとも４９９のデコイペプチド配列は、少なくとも５００、６００、７００、８００、９００、１０００、１１００、１２００、１３００、１４００、１５００、１６００、１７００、１８００、１９００、２０００、２１００、２２００、２３００、２４００、２５００、２６００、２７００、２８００、２９００、３０００、３１００、３２００、３３００、３４００、３５００、３６００、３７００、３８００、３９００、４０００、４１００、４２００、４３００、４４００、４５００、４６００、４７００、４８００、４９００、５０００、５１００、５２００、５３００、５４００、５５００、５６００、５７００、５８００、５９００、６０００、６１００、６２００、６３００、６４００、６５００、６６００、６７００、６８００、６９００、７０００、７１００、７２００、７３００、７４００、７５００、７６００、７７００、７８００、７９００、８０００、８１００、８２００、８３００、８４００、８５００、８６００、８７００、８８００、８９００、９０００、９１００、９２００、９３００、９４００、９５００、９６００、９７００、９８００、９９００、１００００、１１０００、１２０００、１３０００、１４０００、１５０００、１６０００、１７０００、１８０００、１９０００、２００００、２１０００、２２０００、２３０００、２４０００、２５０００、２６０００、２７０００、２８０００、２９０００、３００００、３１０００、３２０００、３３０００、３４０００、３５０００、３６０００、３７０００、３８０００、３９０００、４００００、４１０００、４２０００、４３０００、４４０００、４５０００、４６０００、４７０００、４８０００、４９０００、５００００、５２５００、５５０００、５７５００、６００００、６２５００、６５０００、６７５００、７００００、７２５００、７５０００、７７５００、８００００、８２５００、８５０００、８７５００、９００００、９２５００、９５０００、９７５００、１０００００、１２５０００、１５００００、１７５０００、２００００、２２５０００、２５００００、２７５００、３００００、３２５０００、３５００００、３７５０００、４００００、４２５０００、４５００００、４７５０００、５００００、６００００、７００００、８００００、９００００または１０００００のデコイペプチド配列を含む。当業者は、ヒット：デコイの比の変化によりＰＰＶが変化することを理解することができる。

【００２５】

一部の実施形態では、少なくとも500の試験ペプチド配列は、少なくとも６００、７００、８００、９００、１０００、１１００、１２００、１３００、１４００、１５００、１６００、１７００、１８００、１９００、２０００、２１００、２２００、２３００、２４００、２５００、２６００、２７００、２８００、２９００、３０００、３１００、３２００、３３００、３４００、３５００、３６００、３７００、３８００、３９００、４０００、４１００、４２００、４３００、４４００、４５００、４６００、４７００、４８００、４９００、５０００、５１００、５２００、５３００、５４００、５５００、５６００、５７００、５８００、５９００、６０００、６１００、６２００、６３００、６４００、６５００、６６００、６７００、６８００、６９００、７０００、７１００、７２００、７３００、７４００、７５００、７６００、７７００、７８００、７９００、８０００、８１００、８２００、８３００、８４００、８５００、８６００、８７００、８８００、８９００、９０００、９１００、９２００、９３００、９４００、９５００、９６００、９７００、９８００、９９００、１００００、１１０００、１２０００、１３０００、１４０００、１５０００、１６０００、１７０００、１８０００、１９０００、２００００、２１０００、２２０００、２３０００、２４０００、２５０００、２６０００、２７０００、２８０００、２９０００、３００００、３１０００、３２０００、３３

０００、３４０００、３５０００、３６０００、３７０００、３８０００、３９０００、４００００、４１０００、４２０００、４３０００、４４０００、４５０００、４６０００、４７０００、４８０００、４９０００、５００００、５２５００、５５０００、５７５００、６００００、６２５００、６５０００、６７５００、７００００、７２５００、７５０００、７７５００、８００００、８２５００、８５０００、８７５００、９００００、９２５００、９５０００、９７５００、１０００００、１２５０００、１５００００、１７５０００、２００００、２２５０００、２５００００、２７５０００、３００００、３２５０００、３５００００、３７５０００、４００００、４２５０００、４５００００、４７５０００、５００００、６００００、７００００、８００００、９００００または１０００００の試験ペプチド配列を含む。

【００２６】

一部の実施形態では、上位パーセンテージは、上位０．２０％、０．３０％、０．４０％、０．５０％、０．６０％、０．７０％、０．８０％、０．９０％、１．００％、１．１０％、１．２０％、１．３０％、１．４０％、１．５０％、１．６０％、１．７０％、１．８０％、１．９０％、２．００％、２．１０％、２．２０％、２．３０％、２．４０％、２．５０％、２．６０％、２．７０％、２．８０％、２．９０％、３．００％、３．１０％、３．２０％、３．３０％、３．４０％、３．５０％、３．６０％、３．７０％、３．８０％、３．９０％、４．００％、４．１０％、４．２０％、４．３０％、４．４０％、４．５０％、４．６０％、４．７０％、４．８０％、４．９０％、５．００％、５．１０％、５．２０％、５．３０％、５．４０％、５．５０％、５．６０％、５．７０％、５．８０％、５．９０％、６．００％、６．１０％、６．２０％、６．３０％、６．４０％、６．５０％、６．６０％、６．７０％、６．８０％、６．９０％、７．００％、７．１０％、７．２０％、７．３０％、７．４０％、７．５０％、７．６０％、７．７０％、７．８０％、７．９０％、８．００％、８．１０％、８．２０％、８．３０％、８．４０％、８．５０％、８．６０％、８．７０％、８．８０％、８．９０％、９．００％、９．１０％、９．２０％、９．３０％、９．４０％、９．５０％、９．６０％、９．７０％、９．８０％、９．９０％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％である。

【００２７】

一部の実施形態では、少なくとも１つのヒットペプチド配列は、少なくとも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または１００のヒットペプチド配列を含む。

【００２８】

一部の実施形態では、少なくとも１９のデコイペプチド配列は、少なくとも３０、４０、５０、６０、７０、８０、９０、１００、１１０、１２０、１３０、１４０、１５０、１６０、１７０、１８０、１９０、２００、２１０、２２０、２３０、２４０、２５０、２６０、２７０、２８０、２９０、３００、３１０、３２０、３３０、３４０、３５０、３６０、３７０、３８０、３９０、４００、４１０、４２０、４３０、４４０、４５０、４６０、４７０、４８０、４９０、５００　６００、７００、８００、９００、１０００、１１００、１２００、１３００、１４００、１５００、１６００、１７００、１８００、１９００、２０００、２１００、２２００、２３００、２４００、２５００、２６００、２７００、２８００、２９００、３０００、３１００、３２００、３３００、３４００、３５００、３６００、３７００、３８００、３９００、４０００、４１００、４２００、４３００、４４００、４５００、４６００、４７００、４８００、４９００、５０００

、５１００、５２００、５３００、５４００、５５００、５６００、５７００、５８００、５９００、６０００、６１００、６２００、６３００、６４００、６５００、６６００、６７００、６８００、６９００、７０００、７１００、７２００、７３００、７４００、７５００、７６００、７７００、７８００、７９００、８０００、８１００、８２００、８３００、８４００、８５００、８６００、８７００、８８００、８９００、９０００、９１００、９２００、９３００、９４００、９５００、９６００、９７００、９８００、９９００、１００００、１１０００、１２０００、１３０００、１４０００、１５０００、１６０００、１７０００、１８０００、１９０００、２００００、２１０００、２２０００、２３０００、２４０００、２５０００、２６０００、２７０００、２８０００、２９０００、３００００、３１０００、３２０００、３３０００、３４０００、３５０００、３６０００、３７０００、３８０００、３９０００、４００００、４１０００、４２０００、４３０００、４４０００、４５０００、４６０００、４７０００、４８０００、４９０００、５００００、５２５００、５５０００、５７５００、６００００、６２５００、６５０００、６７５００、７００００、７２５００、７５０００、７７５００、８００００、８２５００、８５０００、８７５００、９００００、９２５００、９５０００、９７５００、１０００００、１２５０００、１５００００、１７５０００、２０００００、２２５０００、２５００００、２７５０００、３０００００、３２５０００、３５００００、３７５０００、４０００００、４２５０００、４５００００、４７５０００、５０００００、６０００００、７０００００、８０００００、９０００００または１００００００のデコイペプチド配列を含む。

【００２９】

一部の実施形態では、少なくとも２０の試験ペプチド配列は、少なくともを含み、少なくとも５００の試験ペプチド配列は、少なくとも３０、４０、５０、６０、７０、８０、９０、１００、１１０、１２０、１３０、１４０、１５０、１６０、１７０、１８０、１９０、２００、２１０、２２０、２３０、２４０、２５０、２６０、２７０、２８０、２９０、３００、３１０、３２０、３３０、３４０、３５０、３６０、３７０、３８０、３９０、４００、４１０、４２０、４３０、４４０、４５０、４６０、４７０、４８０、４９０、５００　６００、７００、８００、９００、１０００、１１００、１２００、１３００、１４００、１５００、１６００、１７００、１８００、１９００、２０００、２１００、２２００、２３００、２４００、２５００、２６００、２７００、２８００、２９００、３０００、３１００、３２００、３３００、３４００、３５００、３６００、３７００、３８００、３９００、４０００、４１００、４２００、４３００、４４００、４５００、４６００、４７００、４８００、４９００、５０００、５１００、５２００、５３００、５４００、５５００、５６００、５７００、５８００、５９００、６０００、６１００、６２００、６３００、６４００、６５００、６６００、６７００、６８００、６９００、７０００、７１００、７２００、７３００、７４００、７５００、７６００、７７００、７８００、７９００、８０００、８１００、８２００、８３００、８４００、８５００、８６００、８７００、８８００、８９００、９０００、９１００、９２００、９３００、９４００、９５００、９６００、９７００、９８００、９９００、１００００、１１０００、１２０００、１３０００、１４０００、１５０００、１６０００、１７０００、１８０００、１９０００、２００００、２１０００、２２０００、２３０００、２４０００、２５０００、２６０００、２７０００、２８０００、２９０００、３００００、３１０００、３２０００、３３０００、３４０００、３５０００、３６０００、３７０００、３８０００、３９０００、４００００、４１０００、４２０００、４３０００、４４０００、４５０００、４６０００、４７０００、４８０００、４９０００、５００００、５２５００、５５０００、５７５００、６００００、６２５００、６５０００、６７５００、７００００、７２５００、７５０００、７７５００、８００００、８２５００、８５０００、８７５００、９００００、９２５００、９５０００、９７５００、１０００００、１２５０００、１５００００、１７５０００、２０００００、２２５０００、２５００００、２７５０００、３０００００、３２５０００、３５００００、３７５０００、４００

０００、４２５０００、４５００００、４７５０００、５０００００、６０００００、７００００、８００００、９００００または１０００００００の試験ペプチド配列試験ペプチド配列を含む。

【００３０】

一部の実施形態では、上位パーセンテージは、上位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％または４０％である。

【００３１】

一部の実施形態では、ＰＰＶは、表１１の対応するＨＬＡ対立遺伝子によってコードされるタンパク質についての表１１の２列目のそれぞれのＰＰＶよりも大きい。一部の実施形態では、ＰＰＶは、表１１の対応するＨＬＡ対立遺伝子によってコードされるタンパク質についての表１１の３列目のそれぞれのＰＰＶと少なくとも等しい。

【００３２】

一部の実施形態では、ＰＰＶは、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質についての表１２の２列目のそれぞれのＰＰＶと等しいまたはそれよりも大きい。

【００３３】

一部の実施形態では、ＰＰＶは、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質についての表１６の２列目のそれぞれのＰＰＶよりも大きい。

【００３４】

一部の実施形態では、対象は、単一の対象である。

【００３５】

一部の実施形態では、対象は、哺乳動物である。

【００３６】

一部の実施形態では、対象は、ヒトである。

【００３７】

一部の実施形態では、訓練用細胞は、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる単一のタンパク質を発現する細胞である。

【００３８】

一部の実施形態では、訓練用細胞は、単一対立遺伝子ＨＬＡ細胞、またはアフィニティータグを有するＨＬＡ対立遺伝子を発現する細胞である。

【００３９】

一部の実施形態では、対象の細胞は、がん細胞を含む。

【００４０】

一部の実施形態では、方法は、ペプチド配列を同定するためのものである。

【００４１】

一部の実施形態では、方法は、ペプチド配列を選択するためのものである。

【００４２】

一部の実施形態では、方法は、がん治療を調製するためのものである。

【００４３】

一部の実施形態では、方法は、対象特異的がん治療を調製するためのものである。

【００４４】

一部の実施形態では、方法は、がん細胞特異的がん治療を調製するためのものである。

【００４５】

一部の実施形態では、複数のペプチド配列の各ペプチド配列は、がんに関連付けられる。

【００４６】

一部の実施形態では、複数のペプチド配列のうちの少なくとも１つのペプチド配列は、対象のがん細胞によって過剰発現される。

【００４７】

一部の実施形態では、複数のペプチド配列の各ペプチド配列は、対象のがん細胞によって過剰発現される。

【００４８】

一部の実施形態では、複数のペプチド配列のうちの少なくとも１つのペプチド配列は、がん細胞特異的ペプチドである。

【００４９】

一部の実施形態では、複数のペプチド配列の各ペプチド配列は、がん細胞特異的ペプチドである。

【００５０】

一部の実施形態では、複数のペプチド配列の各ペプチド配列は、対象のがん細胞によって発現される。

【００５１】

一部の実施形態では、複数のペプチド配列のうちの少なくとも１つのペプチド配列は、対象の非がん細胞によってコードされない。

【００５２】

一部の実施形態では、複数のペプチド配列の各ペプチド配列は、対象の非がん細胞によってコードされない。

【００５３】

一部の実施形態では、複数のペプチド配列のうちの少なくとも１つのペプチド配列は、対象の非がん細胞では発現されない。

【００５４】

一部の実施形態では、複数のペプチド配列の各ペプチド配列は、対象の非がん細胞では発現されない。

【００５５】

一部の実施形態では、方法は、対象の複数のペプチド配列を得ることを含む。

【００５６】

一部の実施形態では、方法は、対象の複数のポリヌクレオチド配列を得ることを含む。

【００５７】

一部の実施形態では、方法は、対象のゲノムもしくはエクソームによってまたは対象における病原体もしくはウイルスによってコードされる複数のペプチド配列をコードする対象の複数のポリヌクレオチド配列を得ることを含む。

【００５８】

一部の実施形態では、方法は、対象のゲノムまたはエクソームによってコードされる複数のペプチド配列をコードする対象の複数のポリヌクレオチド配列をコンピュータプロセッサによって得ることを含む。

【００５９】

一部の実施形態では、方法は対象の複数のポリヌクレオチド配列をゲノム配列決定またはエクソーム配列決定によって得ることを含む、請求項のいずれか一項に記載の方法。

【００６０】

一部の実施形態では、方法は、対象の複数のポリヌクレオチド配列を全ゲノム配列決定または全エクソーム配列決定によって得ることを含む。

【００６１】

一部の実施形態では、処理するステップは、コンピュータプロセッサによって処理することを含む。

【００６２】

一部の実施形態では、処理するステップは、少なくとも複数のペプチド配列のアミノ酸情報に基づいて複数の予測変数を生成することを含む。

【００６３】

一部の実施形態では、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数の予測変数を処理すること。

【００６４】
一部の実施形態では、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質は、対象によって発現されるクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質である。
【００６５】
一部の実施形態では、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質は、対象のがん細胞によって発現されるクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質である。
【００６６】
一部の実施形態では、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質は、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる単一のタンパク質である。
【００６７】
一部の実施形態では、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質は、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる２種、３種、４種、５種または６種またはそれよりも多くのタンパク質である。
【００６８】
一部の実施形態では、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質は、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる各タンパク質である。
【００６９】
一部の実施形態では、方法は、対象に、ペプチド配列の選択されたサブセットの１つまたは複数を含む組成物を投与するステップをさらに含む。
【００７０】
一部の実施形態では、複数のペプチド配列を同定するステップは、対象のがん細胞由来のＤＮＡ、ＲＮＡ、またはタンパク質配列を対象の正常細胞由来のＤＮＡ、ＲＮＡ、またはタンパク質配列と比較することを含み、ここで、複数のペプチドの各々は、対象のがん細胞に存在し、対象の正常な細胞には存在しない少なくとも１つの変異を含む。
【００７１】
一部の実施形態では、機械学習ＨＬＡ－ペプチド提示予測モデルは、少なくとも訓練用データに基づいて同定された複数の予測変数を含み、ここで、訓練用データは、アミノ酸位置情報を含む訓練用ペプチド配列情報を含み、訓練用ペプチド配列情報は、細胞において発現されるＨＬＡタンパク質に関連付けられる、複数の予測変数；ならびに、アミノ酸位置情報と、アミノ酸位置情報および複数の予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む。
【００７２】
一部の実施形態では、同定するステップは、複数のペプチド配列のうち、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１つによって提示される確率が提示予測確率値の閾値よりも大きいペプチド配列を同定することを含む。
【００７３】
一部の実施形態では、機械学習ＨＬＡペプチド提示予測モデルによって提示されることが予測された複数の試験ペプチド配列の０．２％のうちの１つまたは複数の、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１つによって提示される確率が提示予測確率値の閾値よりも高い。
【００７４】
一部の実施形態では、機械学習ＨＬＡペプチド提示予測モデルによって提示されることが予測された複数の試験ペプチド配列の０．２％の各々の、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１

つによって提示される確率が提示予測確率値の閾値よりも高い。
【００７５】
一部の実施形態では、陽性の数は、ヒットの数と等しくなるように制約される。
【００７６】
一部の実施形態では、質量分析は、単一対立遺伝子質量分析である。
【００７７】
一部の実施形態では、ペプチドは、オートファジーを通じて、細胞において発現されるＨＬＡタンパク質によって提示される。
【００７８】
一部の実施形態では、ペプチドは、ファゴサイトーシスを通じて、細胞において発現されるＨＬＡタンパク質によって提示される。
【００７９】
一部の実施形態では、複数の予測変数は、ペプチドを含む供給源タンパク質の発現レベル予測器を含む。
【００８０】
一部の実施形態では、複数の予測変数は、ペプチドを含む供給源タンパク質の安定性予測器を含む。
【００８１】
一部の実施形態では、複数の予測変数は、ペプチドを含む供給源タンパク質の分解速度予測器を含む。
【００８２】
一部の実施形態では、複数の予測変数は、ペプチドを含む供給源タンパク質のタンパク質切断性予測器を含む。
【００８３】
一部の実施形態では、複数の予測変数は、ペプチドを含む供給源タンパク質の細胞または組織局在化予測器を含む。
【００８４】
一部の実施形態では、複数の予測変数は、ペプチドを含む供給源タンパク質の細胞内プロセシング方式の予測器を含み、供給源タンパク質のプロセシング方式は、供給源タンパク質が、とりわけ、オートファジー、ファゴサイトーシス、および細胞内輸送に供されるかどうかの予測器を含む。
【００８５】
一部の実施形態では、訓練用データの品質を、複数の品質メトリクスを使用して向上させる。
【００８６】
一部の実施形態では、複数の品質メトリクスは、一般的な夾雑ペプチド除去、高いスコア化ピーク強度、高いスコア、および高い質量精度を含む。
【００８７】
一部の実施形態では、スコア化ピーク強度は、少なくとも５０％である。
【００８８】
一部の実施形態では、スコア化ピーク強度は、少なくとも６０％である。
【００８９】
一部の実施形態では、スコアは、少なくとも７である。
【００９０】
一部の実施形態では、質量精度は、最大で５ｐｐｍである。
【００９１】
一部の実施形態では、細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の免疫沈降ＨＬＡタンパク質によって提示されるペプチドである。
【００９２】

一部の実施形態では、細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の外因性ＨＬＡタンパク質によって提示されるペプチドである。

【００９３】

一部の実施形態では、細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の組換えＨＬＡタンパク質によって提示されるペプチドである。

【００９４】

一部の実施形態では、複数の予測変数は、ペプチド－ＨＬＡ親和性予測変数を含む。

【００９５】

一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、酵素特異性なし修飾なしペプチドデータベースを検索することによって同定されるペプチドを含む。

【００９６】

一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、逆データベース検索戦略を使用してペプチドデータベースを検索することによって同定されるペプチドを含む。

【００９７】

一部の実施形態では、ＨＬＡタンパク質は、ＨＬＡ－ＤＲタンパク質、ＨＬＡ－ＤＱタンパク質、またはＨＬＡ－ＤＰタンパク質を含む。

【００９８】

一部の実施形態では、ＨＬＡタンパク質は、ＨＬＡ－ＤＰＢ１＊０１：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０２：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０３：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０２／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０６：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＱＢ１＊０２：０１／ＨＬＡ－ＤＱＡ１＊０５：０１、ＨＬＡ－ＤＱＢ１＊０２：０２／ＨＬＡ－ＤＱＡ１＊０２：０１、ＨＬＡ－ＤＱＢ１＊０６：０２／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＱＢ１＊０６：０４／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０１：０１、ＨＬＡ－ＤＲＢ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０３：０１、ＨＬＡ－ＤＲＢ１＊０３：０２、ＨＬＡ－ＤＲＢ１＊０４：０１、ＨＬＡ－ＤＲＢ１＊０４：０２、ＨＬＡ－ＤＲＢ１＊０４：０３、ＨＬＡ－ＤＲＢ１＊０４：０４、ＨＬＡ－ＤＲＢ１＊０４：０５、ＨＬＡ－ＤＲＢ１＊０４：０７、ＨＬＡ－ＤＲＢ１＊０７：０１、ＨＬＡ－ＤＲＢ１＊０８：０１、ＨＬＡ－ＤＲＢ１＊０８：０２、ＨＬＡ－ＤＲＢ１＊０８：０３、ＨＬＡ－ＤＲＢ１＊０８：０４、ＨＬＡ－ＤＲＢ１＊０９：０１、ＨＬＡ－ＤＲＢ１＊１０：０１、ＨＬＡ－ＤＲＢ１＊１１：０１、ＨＬＡ－ＤＲＢ１＊１１：０２、ＨＬＡ－ＤＲＢ１＊１１：０４、ＨＬＡ－ＤＲＢ１＊１２：０１、ＨＬＡ－ＤＲＢ１＊１２：０２、ＨＬＡ－ＤＲＢ１＊１３：０１、ＨＬＡ－ＤＲＢ１＊１３：０２、ＨＬＡ－ＤＲＢ１＊１３：０３、ＨＬＡ－ＤＲＢ１＊１４：０１、ＨＬＡ－ＤＲＢ１＊１５：０１、ＨＬＡ－ＤＲＢ１＊１５：０２、ＨＬＡ－ＤＲＢ１＊１５：０３、ＨＬＡ－ＤＲＢ１＊１６：０１、ＨＬＡ－ＤＲＢ３＊０１：０１、ＨＬＡ－ＤＲＢ３＊０２：０２、ＨＬＡ－ＤＲＢ３＊０３：０１、ＨＬＡ－ＤＲＢ４＊０１：０１、ＨＬＡ－ＤＲＢ５＊０１：０１からなる群から選択されるＨＬＡクラスＩＩタンパク質を含む。

【００９９】

一部の実施形態では、ＨＬＡ－ＤＲは、ＤＲＡ＊０１：０１と対形成する。

【０１００】

一部の実施形態では、ＨＬＡタンパク質は、ＤＰＡ＊０１：０３／ＤＰＢ＊０４：０１、ＤＲＢ１＊０１：０１、ＤＲＢ１＊０１：０２、ＤＲＢ１＊０３：０１、ＤＲＢ１＊０４：０１、ＤＲＢ１＊０４：０２、ＤＲＢ１＊０４：０４、ＤＲＢ１＊０４：０５、ＤＲＢ１＊０７：０１、ＤＲＢ１＊０８：０１、ＤＲＢ１＊０８：０２、ＤＲＢ１＊０８：０３、ＤＲＢ１＊０９：０１、ＤＲＢ１＊１１：０１、ＤＲＢ１＊１１：０２、ＤＲＢ１＊１１：０

４、ＤＲＢ１＊１２：０１、ＤＲＢ１＊１３：０１、ＤＲＢ１＊１３：０２、ＤＲＢ１＊１３：０３、ＤＲＢ１＊１４：０１、ＤＲＢ１＊１５：０１、ＤＲＢ１＊１５：０２、ＤＲＢ１＊１５：０３、ＤＲＢ１＊１６：０２、ＤＲＢ３＊０１：０１、ＤＲＢ３＊０２：０１、ＤＲＢ３＊０２：０２、ＤＲＢ３＊０３：０１、ＤＲＢ４＊０１：０１、ＤＲＢ４＊０１：０３およびＤＲＢ５＊０１：０１からなる群から選択されるＨＬＡクラスＩＩタンパク質である。

【０１０１】

一部の実施形態では、ＨＬＡ－ＤＲタンパク質は、ＤＲＡ＊０１：０１を二量体中に含む。

【０１０２】

一部の実施形態では、ＨＬＡタンパク質は、ＤＰＢ１＊０１：０１、ＤＰＢ１＊０２：０１、ＤＰＢ１＊０２：０２、ＤＰＢ１＊０３：０１、ＤＰＢ１＊０４：０１、ＤＰＢ１＊０４：０２、ＤＰＢ１＊０５：０１、ＤＰＢ１＊０６：０１、ＤＰＢ１＊１１：０１、ＤＰＢ１＊１３：０１、ＤＰＢ１＊１７：０１からなる群から選択されるＨＬＡ－ＤＰタンパク質を含む。

【０１０３】

一部の実施形態では、ＨＬＡ－ＤＰタンパク質は、ＤＰＡ１＊０１：０３を含んで対形成する。

【０１０４】

一部の実施形態では、ＨＬＡタンパク質は、Ａ１＊０１：０１＋Ｂ１＊０５：０１、Ａ１＊０１：０２＋Ｂ１＊０６：０２、Ａ１＊０１：０２＋Ｂ１＊０６：０４、Ａ１＊０１：０３＋Ｂ１＊０６：０３、Ａ１＊０２：０１＋Ｂ１＊０２：０２、Ａ１＊０２：０１＋Ｂ１＊０３：０３、Ａ１＊０３：０１＋Ｂ１＊０３：０２、Ａ１＊０３：０３＋Ｂ１＊０３：０１、Ａ１＊０５：０１＋Ｂ１＊０２：０１およびＡ１＊０５：０５＋Ｂ１＊０３：０１からなる群から選択されるＨＬＡ－ＤＱタンパク質複合体を含む。

【０１０５】

一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、ＨＬＡ－ペプチドのＭＳ／ＭＳスペクトルをペプチドまたはタンパク質データベース内の１つまたは複数のペプチドまたはタンパク質のＭＳ／ＭＳスペクトルと比較することによって同定されるペプチドを含む。

【０１０６】

一部の実施形態では、変異は、点変異、スプライス部位変異、フレームシフト変異、リードスルー変異、および遺伝子融合変異からなる群から選択される。

【０１０７】

一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドの長さは、１５～４０アミノ酸である。

【０１０８】

一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、ＨＬＡタンパク質によって提示されるペプチドを、ＨＬＡ－ペプチドのＭＳ／ＭＳスペクトルとペプチドまたはタンパク質データベース内の１つまたは複数のペプチドまたはタンパク質のＭＳ／ＭＳスペクトルと比較することによって同定することにより同定されるペプチドを含む。

【０１０９】

一部の実施形態では、個別化がん治療は、アジュバントをさらに含む。

【０１１０】

一部の実施形態では、個別化がん治療は、免疫チェックポイント阻害剤をさらに含む。

【０１１１】

一部の実施形態では、訓練用データは、構造化されたデータ、時系列データ、構造化されていないデータ、リレーショナルデータ、またはこれらの任意の組合せを含む。

【０１１２】

一部の実施形態では、構造化されていないデータは、画像データを含む。

【０１１３】
一部の実施形態では、リレーショナルデータは、顧客システム、企業システム、運用システム、ウェブサイト、ウェブアクセス可能なアプリケーションプログラムインターフェース（ＡＰＩ）、またはこれらの任意の組合せからのデータを含む。
【０１１４】
一部の実施形態では、訓練用データを、クラウドに基づくデータベースにアップロードする。
【０１１５】
一部の実施形態では、訓練を、畳み込みニューラルネットワークを使用して実施する。
【０１１６】
一部の実施形態では、畳み込みニューラルネットワークは、少なくとも２つの畳み込み層を含む。
【０１１７】
一部の実施形態では、畳み込みニューラルネットワークは、少なくとも１つのバッチ正規化ステップを含む。
【０１１８】
一部の実施形態では、畳み込みニューラルネットワークは、少なくとも１つの空間ドロップアウトステップを含む。
【０１１９】
一部の実施形態では、畳み込みニューラルネットワークは、少なくとも１つグローバルマックスプーリングステップを含む。
【０１２０】
一部の実施形態では、畳み込みニューラルネットワークは、少なくとも１つの高密度層を含む。
【０１２１】
一部の実施形態では、ペプチド配列を同定するステップは、対象のがん細胞において発現される変異を用いてペプチド配列を同定することを含む。
【０１２２】
一部の実施形態では、ペプチド配列を同定するステップは、対象の正常細胞では発現されないペプチド配列を同定することを含む。
【０１２３】
一部の実施形態では、ペプチド配列を同定するステップは、ウイルス性ペプチド配列を同定することを含む。
【０１２４】
一部の実施形態では、ペプチド配列を同定するステップは、過剰発現されたペプチド配列を同定することを含む。
【０１２５】
対象に対する免疫治療のためのＨＬＡクラスＩＩ特異的ペプチドを同定する方法であって、コンピュータプロセッサにより、エピトープを含む候補ペプチド、およびそれぞれがエピトープを含む複数のペプチド配列を得るステップと、免疫細胞に対する複数のペプチド配列の各々についての提示予測を生成するために、コンピュータプロセッサにより、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、ＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルは、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップと、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質から、候補ペプチドに結合することが機械学習ＨＬＡ－ペプチド提示予測モデルによって予測されたタンパク質を選択するステップであって、タンパク質は、免疫細胞に対する候

補ペプチドの提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップと、候補ペプチドを選択されたタンパク質と接触させ、したがって、候補ペプチドを、選択されたタンパク質と会合したプレースホルダーペプチドと競合させるステップと、候補ペプチドがプレースホルダーに置き換わるかどうかに基づいて、候補ペプチドを選択されたタンパク質に特異的な免疫治療のためのペプチドとして同定するステップとを含む方法が本明細書で提供される。

【０１２６】

一部の実施形態では、得るステップは、候補ペプチドを同定することを含み、候補ペプチドを同定することは、対象のがん細胞由来のＤＮＡ、ＲＮＡ、またはタンパク質配列を対象の正常細胞由来のＤＮＡ、ＲＮＡ、またはタンパク質配列と比較することを含む。

【０１２７】

一部の実施形態では、処理するステップは、少なくとも複数のペプチド配列のアミノ酸情報に基づいて複数の予測変数を同定すること、および機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数の予測変数を処理することを含む。

【０１２８】

一部の実施形態では、機械学習ＨＬＡ－ペプチド提示予測モデルは、少なくとも訓練用データに基づいて同定された複数の予測変数を含み、ここで、訓練用データは、アミノ酸位置情報を含む訓練用ペプチド配列情報であって、細胞において発現されるＨＬＡタンパク質に関連付けられる訓練用ペプチド配列情報；ならびに、アミノ酸位置情報と、アミノ酸位置情報および複数の予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む。

【０１２９】

一部の実施形態では、陽性の数は、ヒットの数と等しくなるように制約される。

【０１３０】

一部の実施形態では、質量分析は、単一対立遺伝子質量分析である。

【０１３１】

一部の実施形態では、複数の予測変数は、ペプチドを含む供給源タンパク質の発現レベル予測器、安定性予測器、分解速度予測器、切断性予測器、細胞または組織局在化予測器、ならびにオートファジー、ファゴサイトーシス、および細胞内輸送を含む細胞内プロセシング方式予測器のうちの任意の１つまたは複数を含む。

【０１３２】

一部の実施形態では、訓練用データの品質を、複数の品質メトリクスを使用して向上させる。

【０１３３】

一部の実施形態では、複数の品質メトリクスは、一般的な夾雑ペプチド除去、高いスコア化ピーク強度、高いスコア、および高い質量精度を含む。

【０１３４】

一部の実施形態では、スコア化ピーク強度は、少なくとも５０％である。

【０１３５】

一部の実施形態では、スコア化ピーク強度は、少なくとも６０％である。

【０１３６】

一部の実施形態では、プレースホルダーペプチドは、ＣＬＩＰペプチドである。

【０１３７】

一部の実施形態では、プレースホルダーペプチドは、ＣＭＶペプチドである。

【０１３８】

一部の実施形態では、３方法は、標的ペプチドによるプレースホルダーペプチドの置換えのＩＣ５０を測定するステップをさらに含む。

【０１３９】

一部の実施形態では、標的ペプチドによるプレースホルダーペプチドの置換えのＩＣ５０は、５００ｎＭ未満である。

【０１４０】
一部の実施形態では、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質からの少なくとも１つのタンパク質は、ＨＬＡクラスＩＩ四量体または多量体である。
【０１４１】
一部の実施形態では、標的ペプチドを質量分析によってさらに同定する。
【０１４２】
一部の実施形態では、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる少なくとも１つのタンパク質は、組換えタンパク質である。
【０１４３】
一部の実施形態では、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる少なくとも１つのタンパク質は、真核細胞において発現するものである。
【０１４４】
一部の実施形態では、ペプチドは、オートファジーを通じて、細胞において発現されるＨＬＡタンパク質によって提示される。
【０１４５】
一部の実施形態では、ペプチドは、ファゴサイトーシスを通じて、細胞において発現されるＨＬＡタンパク質によって提示される。
【０１４６】
一部の実施形態では、細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の免疫沈降ＨＬＡタンパク質によって提示されるペプチドである。
【０１４７】
一部の実施形態では、細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の外因性ＨＬＡタンパク質によって提示されるペプチドである。
【０１４８】
一部の実施形態では、細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の組換えＨＬＡタンパク質によって提示されるペプチドである。
【０１４９】
一部の実施形態では、複数の予測変数は、ペプチド－ＨＬＡ親和性予測変数を含む。
【０１５０】
一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、酵素特異性なし修飾なしペプチドデータベースを検索することによって同定されるペプチドを含む。
【０１５１】
一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、逆データベース検索戦略を使用してペプチドデータベースを検索することによって同定されるペプチドを含む。
【０１５２】
一部の実施形態では、ＨＬＡタンパク質は、ＨＬＡ－ＤＲタンパク質、ＨＬＡ－ＤＱタンパク質、またはＨＬＡ－ＤＰタンパク質を含む。
【０１５３】
一部の実施形態では、免疫治療は、がん免疫治療である。
【０１５４】
一部の実施形態では、エピトープは、がん特異的エピトープである。
【０１５５】
一部の実施形態では、ＨＬＡクラスＩＩ対立遺伝子によってコードされる少なくとも１つのタンパク質は、二量体形態で存在するＨＬＡタンパク質の少なくともアルファ１サブユニットおよびベータ１サブユニットを含む。

【０１５６】
一部の実施形態では、ペプチドがいかなるペプチドかは既知である。
【０１５７】
一部の実施形態では、ペプチドがいかなるペプチドかは未知である。
【０１５８】
一部の実施形態では、ペプチドがいかなるペプチドかを質量分析によって決定する。
【０１５９】
一部の実施形態では、ペプチド交換アッセイは、ペプチド蛍光プローブまたはタグの検出を含む。
【０１６０】
一部の実施形態では、プレースホルダーペプチドは、ＣＬＩＰペプチドである。一部の実施形態では、プレースホルダーペプチドは、ＰＶＳＫＭＲＭＡＴＰＬＬＭＱＡ（配列番号１）というアミノ酸配列を有する。
【０１６１】
一部の実施形態では、ポリ核酸構築物は、プロモーター、分泌シグナル、二量体形成因子、リボソームスキッピング配列、精製および／または検出用の１つまたは複数のタグのうちの１つまたは複数をさらに含む発現ベクターを含む。
【０１６２】
一部の実施形態では、プレースホルダーペプチド配列は、ベクター内の核酸配列によってコードされる。
【０１６３】
一部の実施形態では、切断可能ドメインをコードする配列を、プレースホルダーペプチドをコードする配列とＨＬＡベータ１ペプチドをコードする配列の間に配置する。
【０１６４】
ＭＨＣクラスＩＩ結合性ペプチドの免疫原性をアッセイするための方法であって、ＭＨＣクラスＩＩ結合性ペプチドに結合することが機械学習ＨＬＡ－ペプチド提示予測モデルによって予測された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質を選択するステップであって、機械学習ＨＬＡ－ペプチド提示予測モデルは、所与のペプチド配列についての提示予測を生成するように構成されており、提示予測により、ＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質により所与のペプチド配列が提示され得る可能性が示され、タンパク質は、ＭＨＣクラスＩＩ結合性ペプチドの提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップと、ペプチドを選択されたタンパク質と接触させ、したがって、ペプチドを、選択されたタンパク質と会合したプレースホルダーペプチドと競合させ、プレースホルダーペプチドを置き換え、それにより、ＨＬＡクラスＩＩタンパク質とＭＨＣクラスＩＩ結合性ペプチドとを含む複合体を形成するステップと、複合体をＣＤ４＋　Ｔ細胞と接触させるステップと、サイトカインの誘導、ケモカインの誘導、および細胞表面マーカーの発現からなる群から選択されるＣＤ４＋　Ｔ細胞の活性化パラメータの１つまたは複数についてアッセイするステップとを含む方法が本明細書で提供される。
【０１６５】
一部の実施形態では、ＨＬＡクラスＩＩ対立遺伝子は、四量体または多量体である。
【０１６６】
一部の実施形態では、サイトカインはＩＬ－２である。
【０１６７】
がん免疫治療のために対象におけるＣＤ４＋　Ｔ細胞活性化を誘導するための方法であって、がんに関連付けられ、がん変異を含むペプチド配列を同定するステップであって、ペプチド配列の同定は、対象のがん細胞由来のＤＮＡ、ＲＮＡ、またはタンパク質配列を対象の正常細胞由来のＤＮＡ、ＲＮＡ、またはタンパク質配列と比較することを含む、ステップと、対象の細胞によって通常発現され、ペプチドに結合することが機械学習ＨＬＡ－ペプチド提示予測モデルによって予測された、ＨＬＡクラスＩＩ対立遺伝子によってコ

ードされるタンパク質を選択するステップであって、予測モデルの陽性適中率は、少なくとも０．１％、０．１％～５０％または最大で５０％の再現率で少なくとも０．１であり、タンパク質は、同定されたペプチド配列の提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップと、同定されたペプチドを、選択された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質と接触させて、同定されたペプチドは、５００ｎＭ未満のＩＣ５０値で、選択された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質と会合したプレースホルダーペプチドと競合してプレースホルダーペプチドに置き換わるかどうかを検証するステップと、必要に応じて、同定されたペプチドを精製するステップと、同定されたペプチドの配列を含むポリペプチドまたはポリペプチドをコードするポリヌクレオチドを対象に有効量で投与するステップとを含む方法が本明細書で提供される。

【０１６８】

ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、コンピュータプロセッサにより、ポリペプチド配列の複数のペプチド配列を得るステップと、複数のペプチド配列の各々についての提示予測を生成するために、コンピュータプロセッサにより、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列のエピトープ配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルは、細胞において発現されるＨＬＡタンパク質に関連する配列情報を含む訓練用データを使用して訓練される、ステップと、複数の提示予測に基づいて、ポリペプチド配列の複数のペプチド配列の各々が対象に対して免疫原性ではないことを決定または予測するステップと、対象に、薬物を含む組成物を投与するステップとを含む方法が本明細書で提供される。

【０１６９】

４つの個々のＨＬＡタンパク質アルファ１およびベータ１ヘテロ二量体をコンジュゲートすることによってＨＬＡクラスＩＩ四量体または多量体を製造するための方法であって、真核細胞において、ＨＬＡタンパク質のアルファ鎖およびベータ鎖をコードする核酸配列、分泌シグナル、ビオチン化モチーフおよび同定用または精製用の少なくとも１つのタグを含むベクターを発現させ、したがって、各ＨＬＡタンパク質アルファ１およびベータ１ヘテロ二量体を、二量体を形成した状態で分泌させるステップであって、ヘテロ二量体が、プレースホルダーペプチドと会合している、ステップと、分泌されたヘテロ二量体を細胞培地から精製するステップと、ペプチド交換アッセイを使用してペプチド結合活性を検証するステップと、ストレプトアビジンを添加し、それにより、ヘテロ二量体をコンジュゲートして四量体にするステップと、四量体を精製し、１ｍｇ／Ｌを超える収量を得るステップとを含む方法が本明細書で提供される。多量体、例えば、五量体、六量体または八量体も同様に生成することができ、本明細書において同等に意図されている。

【０１７０】

一部の実施形態では、ベクターは、ＣＭＶプロモーターを含む。

【０１７１】

一部の実施形態では、ベクターは、ベータ１鎖と切断可能部位を介して連結したプレースホルダーペプチドをコードする配列を含む。

【０１７２】

一部の実施形態では、ペプチド交換アッセイは、プレースホルダーペプチドをベータ鎖から予め切断することを伴う。

【０１７３】

一部の実施形態では、切断可能部位は、トロンビン切断部位である。

【０１７４】

一部の実施形態では、ペプチド交換アッセイは、ＦＲＥＴアッセイである。

【０１７５】

一部の実施形態では、精製は、カラムクロマトグラフィー、イオン交換クロマトグラフィー、分子ふるいクロマトグラフィー、アフィニティークロマトグラフィー、またはＬＣ－ＭＳのうちのいずれか１つによるものである。

【０１７６】

ＨＬＡ－ＤＲヘテロ二量体、またはＨＬＡ－ＤＰヘテロ二量体、またはＨＬＡ－ＤＱヘテロ二量体のいずれかを含むＨＬＡクラスＩＩ四量体または多量体であって、各ヘテロ二量体がアルファ鎖とベータ鎖とを含み、ヘテロ二量体が精製されたものであり、１ｍｇ／Ｌを超える濃度で存在する、ＨＬＡクラスＩＩ四量体または多量体が本明細書で提供される。

【０１７７】

一部の実施形態では、ＨＬＡクラスＩＩ四量体は、表８Ａ～８Ｃから選択される。

【０１７８】

一部の実施形態では、ＨＬＡクラスＩＩ四量体は、ＨＬＡ－ＤＲタンパク質、ＨＬＡ－ＤＰタンパク質、およびＨＬＡ－ＤＱタンパク質からなる群から選択されるヘテロ二量体対を含む。

【０１７９】

一部の実施形態では、ＨＬＡタンパク質は、ＨＬＡ－ＤＰＢ１＊０１：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０２：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０３：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０２／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０６：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＱＢ１＊０２：０１／ＨＬＡ－ＤＱＡ１＊０５：０１、ＨＬＡ－ＤＱＢ１＊０２：０２／ＨＬＡ－ＤＱＡ１＊０２：０１、ＨＬＡ－ＤＱＢ１＊０６：０２／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＱＢ１＊０６：０４／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０１：０１、ＨＬＡ－ＤＲＢ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０３：０１、ＨＬＡ－ＤＲＢ１＊０３：０２、ＨＬＡ－ＤＲＢ１＊０４：０１、ＨＬＡ－ＤＲＢ１＊０４：０２、ＨＬＡ－ＤＲＢ１＊０４：０３、ＨＬＡ－ＤＲＢ１＊０４：０４、ＨＬＡ－ＤＲＢ１＊０４：０５、ＨＬＡ－ＤＲＢ１＊０４：０７、ＨＬＡ－ＤＲＢ１＊０７：０１、ＨＬＡ－ＤＲＢ１＊０８：０１、ＨＬＡ－ＤＲＢ１＊０８：０２、ＨＬＡ－ＤＲＢ１＊０８：０３、ＨＬＡ－ＤＲＢ１＊０８：０４、ＨＬＡ－ＤＲＢ１＊０９：０１、ＨＬＡ－ＤＲＢ１＊１０：０１、ＨＬＡ－ＤＲＢ１＊１１：０１、ＨＬＡ－ＤＲＢ１＊１１：０２、ＨＬＡ－ＤＲＢ１＊１１：０４、ＨＬＡ－ＤＲＢ１＊１２：０１、ＨＬＡ－ＤＲＢ１＊１２：０２、ＨＬＡ－ＤＲＢ１＊１３：０１、ＨＬＡ－ＤＲＢ１＊１３：０２、ＨＬＡ－ＤＲＢ１＊１３：０３、ＨＬＡ－ＤＲＢ１＊１４：０１、ＨＬＡ－ＤＲＢ１＊１５：０１、ＨＬＡ－ＤＲＢ１＊１５：０２、ＨＬＡ－ＤＲＢ１＊１５：０３、ＨＬＡ－ＤＲＢ１＊１６：０１、ＨＬＡ－ＤＲＢ３＊０１：０１、ＨＬＡ－ＤＲＢ３＊０２：０２、ＨＬＡ－ＤＲＢ３＊０３：０１、ＨＬＡ－ＤＲＢ４＊０１：０１およびＨＬＡ－ＤＲＢ５＊０１：０１からなる群から選択されるＨＬＡクラスＩＩタンパク質である。

【０１８０】

一部の実施形態では、ヘテロ二量体対は、真核細胞において発現される。

【０１８１】

一部の実施形態では、ヘテロ二量体対は、ベクターによってコードされる。

【０１８２】

本明細書に記載のＨＬＡタンパク質のアルファ鎖およびベータ鎖をコードする核酸配列、分泌シグナル、ビオチン化モチーフおよび同定用または精製用の少なくとも１つのタグを含み、したがって、各ＨＬＡタンパク質アルファ１およびベータ１ヘテロ二量体を、二量体を形成した状態で分泌させるベクターであって、分泌されたヘテロ二量体が、必要に応じてプレースホルダーペプチドと会合している、ベクターが本明細書で提供される。

【０１８３】

本明細書に記載のベクターを含む細胞が本明細書で提供される。

【０１８４】
一部の実施形態では、ＨＬＡクラスＩＩヘテロ二量体を真核細胞から細胞培養培地中に分泌させ、それを、カラムクロマトグラフィー、イオン交換クロマトグラフィー、分子ふるいクロマトグラフィー、アフィニティークロマトグラフィーまたはＬＣ－ＭＳのうちのいずれか１つによってさらに精製する。
【０１８５】
ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、コンピュータプロセッサにより、ポリペプチド配列の複数のペプチド配列を得るステップと、複数のペプチド配列の各々についての提示予測を生成するために、コンピュータプロセッサにより、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列のエピトープ配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルは、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練されるステップと、複数の提示予測に基づいて、ポリペプチド配列の複数のペプチド配列のうちの少なくとも１つが対象に対して免疫原性であることを決定または予測するステップとを含む方法が本明細書で提供される。
【０１８６】
ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、コンピュータプロセッサを使用してポリペプチド配列のペプチド配列のアミノ酸情報を機械学習ＨＬＡ－ペプチド提示予測モデルに入力して、ペプチド配列についての提示予測のセットを生成するステップであって、各提示予測により、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により所与のペプチド配列のエピトープ配列が提示される確率が示され、機械学習ＨＬＡ－ペプチド提示予測モデルは、少なくとも訓練用データに基づいて同定された複数の予測変数を含み、訓練用データは、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；アミノ酸位置情報を含む訓練用ペプチド配列情報であって、細胞において発現されるＨＬＡタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られたアミノ酸位置情報とアミノ酸位置情報および予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む、ステップと、提示予測のセットに基づいて、ポリペプチド配列のペプチド配列の各々が対象に対して免疫原性ではないことを決定または予測するステップと、対象に、薬物を含む組成物を投与するステップとを含む方法が本明細書で提供される。
【０１８７】
ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、コンピュータプロセッサを使用してポリペプチド配列のペプチド配列のアミノ酸情報を機械学習ＨＬＡ－ペプチド提示予測モデルに入力して、ペプチド配列についての提示予測のセットを生成するステップであって、各提示予測により、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により所与のペプチド配列のエピトープ配列が提示される確率が示され、機械学習ＨＬＡ－ペプチド提示予測モデルは、少なくとも訓練用データに基づいて同定された複数の予測変数を含み、訓練用データは、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；アミノ酸位置情報を含む訓練用ペプチド配列情報であって、細胞において発現されるＨＬＡタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られたアミノ酸位置情報とアミノ酸位置情報および予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む、ステップと、提示予測のセットに基づいて、ポリペプチド配列のペプチド配列のうちの少なくとも１つが対象に対して免疫原性であることを決定または予測す

るステップとを含む方法が本明細書で提供される。
【０１８８】
ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、コンピュータプロセッサにより、ポリペプチド配列の複数のペプチド配列を得るステップと、複数のペプチド配列の各々についての提示予測を生成するために、コンピュータプロセッサにより、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列のエピトープ配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルは、細胞において発現されるＨＬＡタンパク質に関連する配列情報を含む訓練用データを使用して訓練されるステップと、複数の提示予測に基づいて、ポリペプチド配列の複数のペプチド配列の各々が対象に対して免疫原性ではないことを決定または予測するステップと、対象に、薬物を含む組成物を投与するステップとを含む方法が本明細書で提供される。
【０１８９】
一部の実施形態では、方法は、薬物を対象に投与しないことを決定することをさらに含む。
【０１９０】
一部の実施形態では、薬物は、抗体またはその結合性断片を含む。
【０１９１】
一部の実施形態では、ポリペプチド配列のペプチド配列は、８、９、１０、１１、または１２アミノ酸の長さを有し、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされるタンパク質は、対象の細胞のクラスＩ　ＭＨＣ対立遺伝子によってコードされるタンパク質である。
【０１９２】
一部の実施形態では、ポリペプチド配列のペプチド配列は、１５、１６、１７、１８、１９、２０、２１、２２、２３、２４、または２５アミノ酸の長さを有し、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされるタンパク質は、対象の細胞のクラスＩＩ　ＭＨＣ対立遺伝子によってコードされるタンパク質である。
【０１９３】
自己免疫疾患または状態を有する対象を処置する方法であって、（ａ）対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣによって提示される発現されたタンパク質のエピトープを同定または予測するステップであって、同定または予測されたエピトープとクラスＩまたはクラスＩＩ　ＭＨＣとを含む複合体が、対象のＣＤ８　Ｔ細胞またはＣＤ４　Ｔ細胞の標的とされる、ステップと、（ｂ）複合体に結合するＴ細胞受容体（ＴＣＲ）を同定するステップと、（ｃ）ＴＣＲを対象由来の調節性Ｔ細胞または同種異系の調節性Ｔ細胞において発現させるステップと、（ｄ）ＴＣＲを発現する調節性Ｔ細胞を対象に投与するステップとを含む方法が本明細書で提供される。
【０１９４】
一部の実施形態では、自己免疫疾患または状態は、糖尿病である。
【０１９５】
一部の実施形態では、細胞は、膵島細胞である。
【０１９６】
自己免疫疾患または状態を有する対象を処置する方法であって、対象に、（ｉ）対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣによって提示されることが同定または予測された発現されたタンパク質のエピトープと（ｉｉ）クラスＩまたはクラスＩＩ　ＭＨＣとを含む複合体であって、対象のＣＤ８　Ｔ細胞またはＣＤ４　Ｔ細胞の標的とされる、複合体に結合するＴ細胞受容体（ＴＣＲ）を発現する調節性Ｔ細胞を投与するステップを含む方法が本明細書で提供される。
【０１９７】

対象の個別化がん治療のためのペプチド配列を同定するためのコンピュータシステムであって、対象の複数のペプチド配列が保存されるように構成されたデータベースと、前記データベースと作動可能にカップリングした１つまたは複数のコンピュータプロセッサであって、複数のペプチド配列の各々についての提示予測を生成するために、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、対象の細胞のクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップ；および対象の個別化がん治療のための複数のペプチド配列のサブセットを少なくとも複数の提示予測に基づいて選択するステップを行うように個別に集合的にプログラムされた１つまたは複数のコンピュータプロセッサとを含むコンピュータシステムが本明細書で提供される。

【０１９８】

対象に対する免疫治療のためのＨＬＡクラスＩＩ特異的ペプチドを同定するためのコンピュータシステムであって、エピトープおよびそれぞれがエピトープを含む複数のペプチド配列を含む候補ペプチドが保存されるように構成されたデータベースと、前記データベースと作動可能にカップリングした１つまたは複数のコンピュータプロセッサであって、免疫細胞に対する複数のペプチド配列の各々についての提示予測を生成するために、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、ＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップ；対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質から、候補ペプチドに結合することが機械学習ＨＬＡ－ペプチド提示予測モデルによって予測されたタンパク質を選択するステップであって、当該タンパク質が、免疫細胞に対する候補ペプチドの提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップ；および、候補ペプチドを選択されたタンパク質と接触させ、したがって、候補ペプチドを選択されたタンパク質と会合したプレースホルダーペプチドと競合させた際に、候補ペプチドによってプレースホルダーペプチドが置き換えられるかどうかに基づいて、候補ペプチドを、選択されたタンパク質に特異的な免疫治療のためのペプチドとして同定するステップを行うように個別に集合的にプログラムされた１つまたは複数のコンピュータプロセッサとを含む、コンピュータシステムが本明細書で提供される。

【０１９９】

ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングするためのコンピュータシステムであって、ポリペプチド配列の複数のペプチド配列が保存されるように構成されたデータベースと、前記データベースと作動可能にカップリングした１つまたは複数のコンピュータプロセッサであって、複数のペプチド配列の各々についての提示予測を生成するために、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列のエピトープ配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質に関連する配列情報を含む訓練用データを使用して訓練される、ステップ；および、複数の提示予測に基づいて、ポリペプチド配列の複数のペプチド配列の各々が対象に対して免疫原性ではないことを決定または予測するステップであって、薬物を含む組成物を対象に投与する、ステップを行うように個別に集合的にプログラム

された、１つまたは複数のコンピュータプロセッサとを含む、コンピュータシステムが本明細書で提供される。

【０２００】

ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングするためのコンピュータシステムであって、ポリペプチド配列の複数のペプチド配列が保存されるように構成されたデータベースと、前記データベースと作動可能にカップリングした１つまたは複数のコンピュータプロセッサであって、複数のペプチド配列の各々についての提示予測を生成するために、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列のエピトープ配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップ；および、複数の提示予測に基づいて、ポリペプチド配列の複数のペプチド配列のうちの少なくとも１つが対象に対して免疫原性であることを決定または予測するステップを行うように個別に集合的にプログラムされた、１つまたは複数のコンピュータプロセッサとを含む、コンピュータシステムが本明細書で提供される。

【０２０１】

１つまたは複数のコンピュータプロセッサによって実行されると、対象の個別化がん治療のためのペプチド配列を同定するための方法であって、対象の複数のペプチド配列を得るステップと、複数のペプチド配列の各々についての提示予測を生成するために、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、対象の細胞のクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップと、少なくとも複数の提示予測に基づいて、対象の個別化がん治療のための複数のペプチド配列のサブセットを選択するステップとを含む方法をインプリメントする、機械により実行可能なコードを含む非一時的コンピュータ可読媒体が本明細書で提供される。

【０２０２】

１つまたは複数のコンピュータプロセッサによって実行されると、対象に対する免疫治療のためのＨＬＡクラスＩＩ特異的ペプチドを同定する方法であって、エピトープおよびそれぞれがエピトープを含む複数のペプチド配列を含む候補ペプチドを得るステップと、免疫細胞に対する複数のペプチド配列の各々についての提示予測を生成するために、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、ＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップと、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質から、候補ペプチドに結合することが機械学習ＨＬＡ－ペプチド提示予測モデルによって予測されたタンパク質を選択するステップであって、当該タンパク質が、免疫細胞に対する候補ペプチドの提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップと、候補ペプチドを選択されたタンパク質と接触させ、したがって、候補ペプチドをプレースホルダーペプチドと競合させた際に、候補ペプチドによってプレースホルダーペプチドが置き換えられるかどうかに基づいて、候補ペプチドを、選択されたタンパク質に特異的な免疫治療のためのペプチドとして同定するステップとを含む方法をインプリメント

する、機械により実行可能なコードを含む非一時的コンピュータ可読媒体が本明細書で提供される。

【０２０３】

１つまたは複数のコンピュータプロセッサによって実行されると、ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、ポリペプチド配列の複数のペプチド配列を得るステップと、複数のペプチド配列の各々についての提示予測を生成するために、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列のエピトープ配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質に関連する配列情報を含む訓練用データを使用して訓練される、ステップと、複数の提示予測に基づいて、ポリペプチド配列の複数のペプチド配列の各々が対象に対して免疫原性ではないことを決定または予測するステップであって、薬物を含む組成物を対象に投与する、ステップを含む方法をインプリメントする、機械により実行可能なコードを含む非一時的コンピュータ可読媒体が本明細書で提供される。

【０２０４】

１つまたは複数のコンピュータプロセッサによって実行されると、ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、ポリペプチド配列の複数のペプチド配列を得るステップと、複数のペプチド配列の各々についての提示予測を生成するために、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のペプチド配列のうちの所与のペプチド配列のエピトープ配列が提示され得る可能性が示され、機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップと、複数の提示予測に基づいて、ポリペプチド配列の複数のペプチド配列のうちの少なくとも１つが対象に対して免疫原性であることを決定または予測するステップとを含む方法をインプリメントする、機械により実行可能なコードを含む非一時的コンピュータ可読媒体が本明細書で提供される。

【０２０５】

複数の提示予測を生成するために、機械学習ＨＬＡペプチド提示予測モデルを使用して複数の候補ペプチド配列のアミノ酸情報を処理するステップであって、複数のうちの各候補ペプチド配列が、対象のゲノムまたはエクソームによってコードされ、複数の提示予測が、複数の候補ペプチド配列の各々についてのＨＬＡ提示予測を含み、各提示予測により、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数のうちの所与の候補ペプチド配列が提示され得る可能性が示され、機械学習ＨＬＡペプチド提示予測モデルが、訓練用細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された訓練用ペプチドの配列に関する配列情報を含む訓練用データを使用して訓練される、ステップと、少なくとも複数の提示予測に基づいて、複数のペプチド配列のうち、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１つによって提示される確率が提示予測確率値の閾値よりも大きいペプチド配列を同定するステップとを含み、複数の試験ペプチド配列のアミノ酸情報を処理して、複数の試験提示予測を生成し、各試験提示予測により、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数の試験ペプチド配列のうちの所与の試験ペプチド配列が提示され得る可能性が示される場合に、機械学習ＨＬＡペプチド提示予測モデルの陽性適中率（ＰＰＶ）が少なくとも０．０７であり、複数の試験ペプチド配列が、（ｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によっ

て同定された少なくとも１つのヒットペプチド配列と、（ｉｉ）生物体のゲノムによってコードされるタンパク質内に含有される少なくとも４９９のデコイペプチド配列とを含む少なくとも５００の試験ペプチド配列を含み、生物体と対象が同じ種であり、複数の試験ペプチド配列が、少なくとも１つのヒットペプチド配列と少なくとも４９９のデコイペプチド配列を１：４９９の比で含み、機械学習ＨＬＡペプチド提示予測モデルにより、複数の試験ペプチド配列の０．２％が細胞において発現されるＨＬＡタンパク質によって提示されることが予測される、方法が本明細書で提供される。

【０２０６】

複数の結合予測を生成するために、機械学習ＨＬＡ－ペプチド結合予測モデルを使用して対象のゲノムまたはエクソームによってコードされる複数のペプチド配列のアミノ酸情報を処理するステップであって、複数の結合予測が、複数の候補ペプチド配列の各々についてのＨＬＡ結合予測を含み、各結合予測により、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質が複数の候補ペプチド配列のうちの所与の候補ペプチド配列に結合する可能性が示され、機械学習ＨＬＡペプチド結合予測モデルが、ＨＬＡクラスＩＩタンパク質またはＨＬＡクラスＩＩタンパク質類似体に結合することが同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップと、少なくとも複数の結合予測に基づいて、複数のペプチド配列のうち、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１つとの結合に関する閾値結合予測確率値よりも大きな確率を有するペプチド配列を同定するステップとを含み、複数の試験ペプチド配列のアミノ酸情報を処理して、複数の試験結合予測を生成し、各試験結合予測により、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質が複数の試験ペプチド配列のうちの所与の試験ペプチド配列に結合する可能性が示される場合に、機械学習ＨＬＡペプチド結合予測モデルの陽性適中率（ＰＰＶ）が少なくとも０．１であり、複数の試験ペプチド配列が、（ｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された少なくとも１つのヒットペプチド配列および（ｉｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定されたペプチド配列を含むタンパク質に含有される少なくとも１９のデコイペプチド配列を含む試験ペプチド配列を少なくとも５０含み、生物体と対象が同じ種であり、複数の試験ペプチド配列が、少なくとも１つのヒットペプチド配列と少なくとも１９のデコイペプチド配列を１：１９の比で含み、機械学習ＨＬＡペプチド提示予測モデルにより、複数の試験ペプチド配列の５％が細胞において発現されるＨＬＡタンパク質に結合することが予測される、方法が本明細書で提供される。

【０２０７】

一部の実施形態では、機械学習ＨＬＡペプチド提示予測モデルは、訓練用細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された訓練用ペプチドの配列に関する配列情報を含む訓練用データを使用して訓練される

【０２０８】

一部の実施形態では、機械学習ＨＬＡペプチド提示予測モデルによって提示されることが予測された複数の試験ペプチド配列の０．２％のうちの１つまたは複数の、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１つによって提示される確率が提示予測確率値の閾値よりも高い。

【０２０９】

一部の実施形態では、機械学習ＨＬＡペプチド提示予測モデルによって提示されることが予測された複数の試験ペプチド配列の０．２％の各々の、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質のうちの少なくとも１つによって提示される確率が提示予測確率値の閾値よりも高い。

【０２１０】

一部の実施形態では、ＰＰＶは、表１３の対応するＨＬＡ対立遺伝子によってコードされるタンパク質についての表１１の２列目のそれぞれのＰＰＶよりも大きい。一部の実施

形態では、ＰＰＶは、表１１の対応するＨＬＡ対立遺伝子によってコードされるタンパク質についての表１１の３列目のそれぞれのＰＰＶと少なくとも等しい。

【０２１１】

一部の実施形態では、ＰＰＶは、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質についての表１２の２列目のそれぞれのＰＰＶよりも大きい。

【０２１２】

一部の実施形態では、ＰＰＶは、表１６の対応するＨＬＡ対立遺伝子によってコードされるタンパク質についての表１６の２列目のそれぞれのＰＰＶと少なくとも等しい。

【０２１３】

個別化がん治療を調製するための方法であって、がんに関連付けられるペプチド配列を同定するステップであって、対象のがん細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列を対象の正常細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列と比較することを含む、ステップと、同定されたペプチド配列についての提示予測のセットを生成するために、コンピュータプロセッサを使用して同定されたペプチド配列のアミノ酸位置情報を機械学習ＨＬＡ－ペプチド提示予測モデルに入力するステップであって、各提示予測が、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質が同定されたペプチド配列のうちの所与の配列を提示する確率を表し、機械学習ＨＬＡ－ペプチド提示予測モデルが、少なくとも訓練用データに基づいて同定された複数の予測変数を含み、訓練用データが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；アミノ酸位置情報を含む訓練用ペプチド配列情報であって、細胞において発現されるＨＬＡタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られたアミノ酸位置情報とアミノ酸位置情報および予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む、ステップと、提示予測のセットに基づいて同定された、個別化がん治療を調製するためのペプチド配列のサブセットを選択するステップであって、予測モデルの陽性適中率が、少なくとも０．１％、０．１％～５０％または最大で５０％の再現率で少なくとも０．１である、ステップとを含む方法が本明細書で提供される。

【０２１４】

機械学習ＨＬＡ－ペプチド提示予測モデルを訓練するステップであって、コンピュータプロセッサを使用して、ＨＬＡクラスＩＩ対立遺伝子を発現する細胞由来の１つまたは複数のＨＬＡ－ペプチド複合体から単離されたＨＬＡ－ペプチドのアミノ酸位置情報配列をＨＬＡ－ペプチド提示予測モデルに入力することを含むステップを含み、機械学習ＨＬＡ－ペプチド提示予測モデルが、少なくとも訓練用データに基づいて同定された複数の予測変数が、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；訓練用ペプチドのアミノ酸位置情報を含む訓練用ペプチド配列情報であって、細胞において発現されるＨＬＡタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られたアミノ酸位置情報とアミノ酸位置情報および予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む、方法が本明細書で提供される。

【０２１５】

一部の実施形態では、提示モデルの陽性適中率は、少なくとも０．１％、０．１％～５０％または最大で５０％の再現率で、少なくとも０．２５である。

【０２１６】

一部の実施形態では、提示モデルの陽性適中率は、少なくとも０．１％、０．１％～５０％または最大で５０％の再現率で、少なくとも０．４である。

【０２１７】

一部の実施形態では、提示モデルの陽性適中率は、少なくとも０．１％、０．１％～５０％または最大で５０％の再現率で、少なくとも０．６である。

【０２１８】

一部の実施形態では、質量分析は、単一対立遺伝子質量分析である。

【０２１９】
一部の実施形態では、ペプチドは、オートファジーを通じて、細胞において発現されるＨＬＡタンパク質によって提示される。
【０２２０】
一部の実施形態では、ペプチドは、ファゴサイトーシスを通じて、細胞において発現されるＨＬＡタンパク質によって提示される。
【０２２１】
一部の実施形態では、訓練用データの品質を、複数の品質メトリクスを使用して向上させる。
【０２２２】
一部の実施形態では、複数の品質メトリクスは、一般的な夾雑ペプチド除去、高いスコア化ピーク強度、高いスコア、および高い質量精度を含む。
【０２２３】
一部の実施形態では、スコア化ピーク強度は、少なくとも５０％である。
【０２２４】
一部の実施形態では、スコア化ピーク強度は、少なくとも６０％である。
【０２２５】
一部の実施形態では、スコアは、少なくとも７である。
【０２２６】
一部の実施形態では、質量精度は、最大で５ｐｐｍである。
【０２２７】
一部の実施形態では、質量精度は、最大で２ｐｐｍである。
【０２２８】
一部の実施形態では、骨格切断スコアは、少なくとも５である。
【０２２９】
一部の実施形態では、骨格切断スコアは、少なくとも８である。
【０２３０】
一部の実施形態では、細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の免疫沈降ＨＬＡタンパク質によって提示されるペプチドである。
【０２３１】
一部の実施形態では、細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の外因性ＨＬＡタンパク質によって提示されるペプチドである。
【０２３２】
一部の実施形態では、細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の組換えＨＬＡタンパク質によって提示されるペプチドである。
【０２３３】
一部の実施形態では、複数の予測変数は、ペプチド－ＨＬＡ親和性予測変数を含む。
【０２３４】
一部の実施形態では、複数の予測変数は、供給源タンパク質発現レベル予測変数を含む。
【０２３５】
一部の実施形態では、複数の予測変数は、ペプチド切断性予測変数を含む。
【０２３６】
一部の実施形態では、訓練用ペプチド配列情報は、ペプチドデータベースに対して酵素特異性なし修飾なしを検索することによって同定されるペプチドを含む、ＨＬＡタンパク質によって提示されるペプチド由来の配列を含む。一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、新規のペプチド配列決定ツールを検索することによって同定されるペプチドを含む。

【０２３７】
一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、逆データベース検索戦略を使用してペプチドデータベースを検索することによって同定されるペプチドを含む。
【０２３８】
一部の実施形態では、ＨＬＡタンパク質は、ＨＬＡ－ＤＲタンパク質、およびＨＬＡ－ＤＰタンパク質またはＨＬＡ－ＤＱタンパク質を含む。一部の実施形態では、ＨＬＡタンパク質は、ＨＬＡ－ＤＲタンパク質、およびＨＬＡ－ＤＰタンパク質またはＨＬＡ－ＤＱタンパク質からなる群から選択されるＨＬＡ－ＤＲタンパク質を含む。一部の実施形態では、ＨＬＡタンパク質は、ＨＬＡ－ＤＰＢ１＊０１：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０２：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０３：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０２／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０６：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＱＢ１＊０２：０１／ＨＬＡ－ＤＱＡ１＊０５：０１、ＨＬＡ－ＤＱＢ１＊０２：０２／ＨＬＡ－ＤＱＡ１＊０２：０１、ＨＬＡ－ＤＱＢ１＊０６：０２／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＱＢ１＊０６：０４／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０１：０１、ＨＬＡ－ＤＲＢ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０３：０１、ＨＬＡ－ＤＲＢ１＊０３：０２、ＨＬＡ－ＤＲＢ１＊０４：０１、ＨＬＡ－ＤＲＢ１＊０４：０２、ＨＬＡ－ＤＲＢ１＊０４：０３、ＨＬＡ－ＤＲＢ１＊０４：０４、ＨＬＡ－ＤＲＢ１＊０４：０５、ＨＬＡ－ＤＲＢ１＊０４：０７、ＨＬＡ－ＤＲＢ１＊０７：０１、ＨＬＡ－ＤＲＢ１＊０８：０１、ＨＬＡ－ＤＲＢ１＊０８：０２、ＨＬＡ－ＤＲＢ１＊０８：０３、ＨＬＡ－ＤＲＢ１＊０８：０４、ＨＬＡ－ＤＲＢ１＊０９：０１、ＨＬＡ－ＤＲＢ１＊１０：０１、ＨＬＡ－ＤＲＢ１＊１１：０１、ＨＬＡ－ＤＲＢ１＊１１：０２、ＨＬＡ－ＤＲＢ１＊１１：０４、ＨＬＡ－ＤＲＢ１＊１２：０１、ＨＬＡ－ＤＲＢ１＊１２：０２、ＨＬＡ－ＤＲＢ１＊１３：０１、ＨＬＡ－ＤＲＢ１＊１３：０２、ＨＬＡ－ＤＲＢ１＊１３：０３、ＨＬＡ－ＤＲＢ１＊１４：０１、ＨＬＡ－ＤＲＢ１＊１５：０１、ＨＬＡ－ＤＲＢ１＊１５：０２、ＨＬＡ－ＤＲＢ１＊１５：０３、ＨＬＡ－ＤＲＢ１＊１６：０１、ＨＬＡ－ＤＲＢ３＊０１：０１、ＨＬＡ－ＤＲＢ３＊０２：０２、ＨＬＡ－ＤＲＢ３＊０３：０１、ＨＬＡ－ＤＲＢ４＊０１：０１およびＨＬＡ－ＤＲＢ５＊０１：０１からなる群から選択されるＨＬＡ－ＤＲタンパク質を含む。
【０２３９】
一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、ＨＬＡ－ペプチドのＭＳ／ＭＳスペクトルとペプチドデータベース内の１つまたは複数のＨＬＡ－ペプチドのＭＳ／ＭＳスペクトルと比較することによって同定されるペプチドを含む。
【０２４０】
一部の実施形態では、変異は、点変異、スプライス部位変異、フレームシフト変異、リードスルー変異、および遺伝子融合変異からなる群から選択される。
【０２４１】
一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、１５～４０アミノ酸の長さを有する。
【０２４２】
一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドは、（ａ）単一のＨＬＡクラスＩＩ対立遺伝子を発現する細胞株から１つまたは複数のＨＬＡ複合体を単離するステップと、（ｂ）１つまたは複数の単離されたＨＬＡ複合体から１つまたは複数のＨＬＡ－ペプチドを単離するステップと、（ｃ）１つまたは複数の単離されたＨＬＡ－ペプチドについてのＭＳ／ＭＳスペクトルを得るステップと、（ｄ）ペプチドデータベースから１つまたは複数の単離されたＨＬＡ－ペプチドのＭＳ／ＭＳスペクトルに対応するペプチド配列を得るステップであって、ステップ（ｄ）で得られた１つまたは複数の配列により、１つまたは複数の単離されたＨＬＡ－ペプチドの配列が同定される、ステップによ

って同定されるペプチドを含む。

【０２４３】

一部の実施形態では、個別化がん治療は、アジュバントをさらに含む。

【０２４４】

一部の実施形態では、個別化がん治療は、免疫チェックポイント阻害剤をさらに含む。

【０２４５】

一部の実施形態では、訓練用データは、構造化されたデータ、時系列データ、構造化されていないデータ、リレーショナルデータ、またはこれらの任意の組合せを含む。

【０２４６】

一部の実施形態では、構造化されていないデータは、画像データを含む。

【０２４７】

一部の実施形態では、リレーショナルデータは、顧客システム、企業システム、運用システム、ウェブサイト、ウェブアクセス可能なアプリケーションプログラムインターフェース（ＡＰＩ）、またはこれらの任意の組合せからのデータを含む。

【０２４８】

一部の実施形態では、訓練用データを、クラウドに基づくデータベースにアップロードする。

【０２４９】

一部の実施形態では、訓練するステップを、畳み込みニューラルネットワークを使用して実施する。

【０２５０】

一部の実施形態では、畳み込みニューラルネットワークは、少なくとも２つの畳み込み層を含む。

【０２５１】

一部の実施形態では、畳み込みニューラルネットワーク（ＣＮＮ）は、少なくとも１つのバッチ正規化ステップを含む。

【０２５２】

一部の実施形態では、畳み込みニューラルネットワークは、少なくとも１つの空間ドロップアウトステップを含む。

【０２５３】

一部の実施形態では、畳み込みニューラルネットワークは、少なくとも１つグローバルマックスプーリングステップを含む。

【０２５４】

一部の実施形態では、畳み込みニューラルネットワークは、少なくとも１つの高密度層を含む。

【０２５５】

一部の実施形態では、ペプチド配列を同定するステップは、対象のがん細胞において発現される変異を用いてペプチド配列を同定することを含む。

【０２５６】

一部の実施形態では、ペプチド配列を同定するステップは、対象の正常細胞では発現されないペプチド配列を同定することを含む。

【０２５７】

一部の実施形態では、ペプチド配列を同定するステップは、過剰発現されたペプチド配列を同定することを含む。

【０２５８】

一部の実施形態では、ペプチド配列を同定するステップは、ウイルス性ペプチド配列を同定することを含む。一態様では、対象に特異的な免疫治療のためのＨＬＡクラスＩＩ特異的ペプチドを同定するための方法であって、エピトープを含む候補ペプチドを同定するステップと、コンピュータプロセッサを使用して、各々がエピトープを含む複数のペプチド配列のアミノ酸情報を機械学習ＨＬＡ－ペプチド提示予測モデルに入力して、免疫細胞

に対するペプチド配列のＨＬＡ提示予測のセットを生成するステップであって、各提示予測が、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質がエピトープを含む所与のペプチド配列を提示する確率を表し、予測モデルの陽性適中率が、少なくとも０．１％、０．１％～５０％または最大で５０％の再現率で少なくとも０．１である、ステップと、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質から候補ペプチドに結合することが予測モデルによって予測されたタンパク質を選択するステップであって、当該タンパク質が、免疫細胞に対する候補ペプチドの提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップと、候補ペプチドをＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質と接触させ、したがって、候補ペプチドをＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質と会合したプレースホルダーペプチドと競合させるステップと、候補ペプチドによってプレースホルダーペプチドが置き換えられるかどうかに基づいて、候補ペプチドをＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質に特異的な免疫治療のためのペプチドとして同定するステップとを含む方法が本明細書で提供される。

【０２５９】

一部の実施形態では、免疫治療は、がん免疫治療である。

【０２６０】

一部の実施形態では、同定するステップは、対象のがん細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列を対象の正常細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列と比較することを含む。一部の実施形態では、エピトープは、がん特異的エピトープである。

【０２６１】

一部の実施形態では、ＨＬＡクラスＩＩ対立遺伝子によってコードされる少なくとも１つのタンパク質は、二量体形態で存在するＨＬＡタンパク質の少なくともアルファ１サブユニットおよびベータ１サブユニットまたはその断片を含む。一部の実施形態では、プレースホルダーペプチドは、ＣＬＩＰペプチドである。一部の実施形態では、プレースホルダーペプチドは、ＣＭＶペプチドである。一部の実施形態では、方法は、標的ペプチドによるプレースホルダーペプチドの置換えのＩＣ５０を測定するステップをさらに含む。一部の実施形態では、標的ペプチドによるプレースホルダーペプチドの置換えのＩＣ５０は、５００ｎＭ未満である。一部の実施形態では、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質からの少なくとも１つのタンパク質は、ＨＬＡクラスＩＩ四量体または多量体である。一部の実施形態では、標的ペプチドを質量分析によってさらに同定する。一部の実施形態では、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる少なくとも１つのタンパク質は、組換えタンパク質である。一部の実施形態では、対象の細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる少なくとも１つのタンパク質は、真核細胞において発現される。

【０２６２】

一態様では、ＨＬＡクラスＩＩタンパク質との結合に関する候補ペプチドの特異性を検証するためのアッセイ方法であって、真核細胞において、ＭＨＣ－ＩＩ結合性エピトープを含むペプチドに結合することができるアルファ鎖およびベータ鎖またはその一部を含むＨＬＡクラスＩＩタンパク質をコードする核酸配列を含むポリ核酸構築物を発現させるステップであって、発現されたＨＬＡクラスＩＩタンパク質またはその一部が、プレースホルダーペプチドと会合したままである、ステップと、真核細胞において発現されたＨＬＡクラスＩＩタンパク質またはその一部を単離するステップと、（ａ）漸増量の候補ペプチドを添加して、候補ペプチドによってＨＬＡクラスＩＩタンパク質またはその一部と会合したプレースホルダーペプチドが置き換えられるかどうかを決定し、（ｂ）置換え反応のＩＣ５０を算出してプレースホルダーペプチドと比べた候補ペプチドのＨＬＡクラスＩＩタンパク質またはその一部に対する親和性を決定し、それにより、候補ペプチドのＨＬＡクラスＩＩタンパク質との結合に関する特異性を検証することによってペプチド交換アッセイを実施するステップとを含む方法が本明細書で提供される。

【０２６３】

一部の実施形態では、ペプチドがいかなるペプチドかは既知である。一部の実施形態では、ペプチドがいかなるペプチドかは未知である。一部の実施形態では、ペプチドがいかなるペプチドかを質量分析によって決定する。

【０２６４】

一部の実施形態では、ペプチド交換アッセイは、ペプチド蛍光プローブまたはタグの検出を含む。一部の実施形態では、プレースホルダーペプチドは、ＣＬＩＰペプチドである。

【０２６５】

一部の実施形態では、ポリ核酸構築物は、プロモーター、リンカー、１つまたは複数のプロテアーゼ切断部位、分泌シグナル、二量体形成因子、リボソームスキッピング配列、精製および／または検出のための１つまたは複数のタグのうちの１つまたは複数をさらに含む発現ベクターを含む。

【０２６６】

一態様では、ＭＨＣクラスＩＩ結合性ペプチドの免疫原性をアッセイするための方法であって、ペプチドに結合することが機械学習ＨＬＡ－ペプチド提示予測モデルによって予測された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質を選択するステップであって、予測モデルの陽性適中率が、少なくとも０．１％、０．１％～５０％または最大で５０％の再現率で少なくとも０．１であり、タンパク質が、同定されたペプチド配列の提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップと、ペプチドを、選択された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質と接触させ、したがって、ペプチドを、選択された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質と会合したプレースホルダーペプチドと競合させ、プレースホルダーペプチドを置き換え、それにより、ＨＬＡクラスＩＩタンパク質と同定されたペプチドとを含む複合体を形成するステップと、ＨＬＡクラスＩＩタンパク質と同定されたペプチドの複合体をＣＤ４＋　Ｔ細胞と接触させるステップと、サイトカインの誘導、ケモカインの誘導および細胞表面マーカーの発現から選択されるＣＤ４＋　Ｔ細胞の活性化パラメータの１つまたは複数についてアッセイするステップとを含む方法が本明細書で提供される。

【０２６７】

一部の実施形態では、ＨＬＡクラスＩＩ対立遺伝子は、四量体または多量体である。一部の実施形態では、サイトカインはＩＬ－２である。一部の実施形態では、サイトカインはＩＦＮ－ガンマである。

【０２６８】

一態様では、がん免疫治療のために対象におけるＣＤ４＋　Ｔ細胞活性化を誘導するための方法であって、がんに関連付けられ、がん変異を含むペプチド配列を同定するステップであって、対象のがん細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列を対象の正常細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列と比較することを含む、ステップと、対象の細胞によって通常発現され、ペプチドに結合することが機械学習ＨＬＡ－ペプチド提示予測モデルによって予測された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質を選択するステップであって、予測モデルの陽性適中率が、少なくとも０．１％、０．１％～５０％または最大で５０％の再現率で少なくとも０．１であり、タンパク質が、同定されたペプチド配列の提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップと、同定されたペプチドを、選択された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質と接触させて、同定されたペプチドが、５００ｎＭ未満のＩＣ５０値で、選択された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質と会合したプレースホルダーペプチドと競合してプレースホルダーペプチドに置き換わるかどうかを検証するステップと、同定されたペプチドを精製するステップと、同定されたペプチドを対象に有効量で投与するステップとを含む方法が本明細書で提供される。

【０２６９】

一態様では、ＨＬＡクラスＩＩ四量体または多量体を製造する方法であって、真核細胞において、ＨＬＡタンパク質のアルファ鎖およびベータ鎖をコードする核酸配列、リンカ

ー、１つまたは複数のプロテアーゼ切断部位、分泌シグナル、ビオチン化モチーフおよび同定用または精製用の少なくとも１つのタグを含むベクターを発現させ、したがって、各ＨＬＡタンパク質アルファ１およびベータ１ヘテロ二量体を、二量体を形成した状態で分泌させるステップであって、ヘテロ二量体がプレースホルダーペプチドと会合している、ステップと、分泌されたヘテロ二量体を細胞培地から精製するステップと、ペプチド交換アッセイを使用してペプチド結合活性を検証するステップと、ストレプトアビジンを添加し、それにより、ヘテロ二量体をコンジュゲートして四量体にするステップと、四量体を精製するステップと、１ｍｇ／Ｌを超える収量を得るステップとを含む方法が本明細書で提供される。

【０２７０】

一部の実施形態では、ベクターは、ＣＭＶプロモーターを含む。一部の実施形態では、ベクターは、ベータ１鎖と切断可能部位を介して連結したプレースホルダーペプチドをコードする配列を含む。一部の実施形態では、ペプチド交換アッセイは、プレースホルダーペプチドをベータ鎖から予め切断することを伴う。一部の実施形態では、切断可能部位は、トロンビン切断部位である。一部の実施形態では、ペプチド交換アッセイは、ＦＲＥＴアッセイである。一部の実施形態では、精製は、カラムクロマトグラフィー、バッチクロマトグラフィー、イオン交換クロマトグラフィー、分子ふるいクロマトグラフィー、アフィニティークロマトグラフィーまたはＬＣ－ＭＳのうちのいずれか１つによる。

【０２７１】

一態様では、ＨＬＡ－ＤＲヘテロ二量体、またはＨＬＡ－ＤＰヘテロ二量体、またはＨＬＡ－ＤＱヘテロ二量体のいずれかを含むＨＬＡクラスＩＩ四量体を含む組成物であって、各ヘテロ二量体が、アルファ鎖およびベータ鎖を含み、精製されており、０．２５ｍｇ／Ｌよりも高い濃度で存在する、組成物が本明細書で提供される。一部の実施形態では、ＨＬＡクラスＩＩ四量体は、ＨＬＡ－ＤＲタンパク質、およびＨＬＡ－ＤＰタンパク質またはＨＬＡ－ＤＱタンパク質からなる群から選択することができるタンパク質からなる群から選択されるヘテロ二量体対を含む。一部の実施形態では、ＨＬＡタンパク質は、ＨＬＡ－ＤＰＢ１＊０１：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０２：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０３：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０２／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０６：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＱＢ１＊０２：０１／ＨＬＡ－ＤＱＡ１＊０５：０１、ＨＬＡ－ＤＱＢ１＊０２：０２／ＨＬＡ－ＤＱＡ１＊０２：０１、ＨＬＡ－ＤＱＢ１＊０６：０２／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＱＢ１＊０６：０４／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０１：０１、ＨＬＡ－ＤＲＢ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０３：０１、ＨＬＡ－ＤＲＢ１＊０３：０２、ＨＬＡ－ＤＲＢ１＊０４：０１、ＨＬＡ－ＤＲＢ１＊０４：０２、ＨＬＡ－ＤＲＢ１＊０４：０３、ＨＬＡ－ＤＲＢ１＊０４：０４、ＨＬＡ－ＤＲＢ１＊０４：０５、ＨＬＡ－ＤＲＢ１＊０４：０７、ＨＬＡ－ＤＲＢ１＊０７：０１、ＨＬＡ－ＤＲＢ１＊０８：０１、ＨＬＡ－ＤＲＢ１＊０８：０２、ＨＬＡ－ＤＲＢ１＊０８：０３、ＨＬＡ－ＤＲＢ１＊０８：０４、ＨＬＡ－ＤＲＢ１＊０９：０１、ＨＬＡ－ＤＲＢ１＊１０：０１、ＨＬＡ－ＤＲＢ１＊１１：０１、ＨＬＡ－ＤＲＢ１＊１１：０２、ＨＬＡ－ＤＲＢ１＊１１：０４、ＨＬＡ－ＤＲＢ１＊１２：０１、ＨＬＡ－ＤＲＢ１＊１２：０２、ＨＬＡ－ＤＲＢ１＊１３：０１、ＨＬＡ－ＤＲＢ１＊１３：０２、ＨＬＡ－ＤＲＢ１＊１３：０３、ＨＬＡ－ＤＲＢ１＊１４：０１、ＨＬＡ－ＤＲＢ１＊１５：０１、ＨＬＡ－ＤＲＢ１＊１５：０２、ＨＬＡ－ＤＲＢ１＊１５：０３、ＨＬＡ－ＤＲＢ１＊１６：０１、ＨＬＡ－ＤＲＢ３＊０１：０１、ＨＬＡ－ＤＲＢ３＊０２：０２、ＨＬＡ－ＤＲＢ３＊０３：０１、ＨＬＡ－ＤＲＢ４＊０１：０１、ＨＬＡ－ＤＲＢ５＊０１：０１）からなる群から選択される。

【０２７２】

一部の実施形態では、ヘテロ二量体対を真核細胞において発現させる。一部の実施形態では、ヘテロ二量体対は、ベクターによってコードされる。一部の実施形態では、ベクタ

ーは、ＨＬＡタンパク質のアルファ鎖およびベータ鎖をコードする核酸配列、分泌シグナル、ビオチン化モチーフおよび同定用または精製用の少なくとも１つのタグを含み、したがって、各ＨＬＡタンパク質アルファ１およびベータ１ヘテロ二量体は二量体を形成した状態で分泌され、分泌されたヘテロ二量体はプレースホルダーペプチドと会合している。一部の実施形態では、ベクターは、ＨＬＡタンパク質のアルファ鎖およびベータ鎖をコードする核酸配列、分泌シグナル、ビオチン化モチーフおよび同定用または精製用の少なくとも１つのタグを含み、したがって、各ＨＬＡタンパク質アルファ１およびベータ１ヘテロ二量体は二量体を形成した状態で分泌され、分泌されたヘテロ二量体はプレースホルダーペプチドと会合している。

【０２７３】

一部の実施形態では、ＨＬＡクラスＩＩヘテロ二量体を真核細胞から細胞培養培地中に分泌させ、カラムまたはバッチクロマトグラフィー、イオン交換クロマトグラフィー、分子ふるいクロマトグラフィー、アフィニティークロマトグラフィーまたはＬＣ－ＭＳのうちのいずれか１つによって精製する。

【０２７４】

一態様では、ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、コンピュータプロセッサを使用してポリペプチド配列のペプチド配列のアミノ酸情報を機械学習ＨＬＡ－ペプチド提示予測モデルに入力して、ペプチド配列についての提示予測のセットを生成するステップであって、各提示予測が、対象の細胞のＨＬＡクラスＩまたはクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質により所与のペプチド配列のエピトープ配列が提示される確率を表し、機械学習ＨＬＡ－ペプチド提示予測モデルが、少なくとも訓練用データに基づいて同定された複数の予測変数を含み、訓練用データが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；アミノ酸位置情報を含む訓練用ペプチド配列情報であって、細胞において発現されるＨＬＡタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られたアミノ酸位置情報とアミノ酸位置情報および予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む、ステップと、（ｂ）提示予測のセットに基づいて、ポリペプチド配列のペプチド配列の各々が対象に対して免疫原性ではないことを決定または予測するステップと、（ｃ）対象に、薬物を含む組成物を投与するステップとを含む方法が本明細書で提供される。

【０２７５】

一態様では、ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、（ａ）コンピュータプロセッサを使用してポリペプチド配列のペプチド配列のアミノ酸情報を機械学習ＨＬＡ－ペプチド提示予測モデルに入力して、ペプチド配列についての提示予測のセットを生成するステップであって、各提示予測が、対象の細胞のＨＬＡクラスＩまたはクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質により所与のペプチド配列のエピトープ配列が提示される確率を表し、機械学習ＨＬＡ－ペプチド提示予測モデルが、少なくとも訓練用データに基づいて同定された複数の予測変数を含み、訓練用データが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；アミノ酸位置情報を含む訓練用ペプチド配列情報であって、細胞において発現されるＨＬＡタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られたアミノ酸位置情報とアミノ酸位置情報および予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む、ステップと、（ｂ）提示予測のセットに基づいて、ポリペプチド配列のペプチド配列のうちの少なくとも１つが対象に対して免疫原性であることを決定または予測するステップとを含む方法が本明細書で提供される。

【０２７６】

一実施形態では、方法は、薬物を対象に投与しないことを決定することをさらに含む。

【０２７７】

一実施形態では、薬物は、抗体またはその結合性断片を含む。

【０２７８】

一実施形態では、ポリペプチド配列のペプチド配列は、８、９、１０、１１または１２アミノ酸の長さを有するポリペプチド配列の各連続したペプチド配列を含み、対象の細胞のＨＬＡクラスＩまたはクラスＩＩ対立遺伝子によってコードされるタンパク質は、対象の細胞のＨＬＡクラスＩ対立遺伝子によってコードされるタンパク質である。

【０２７９】

一実施形態では、ポリペプチド配列のペプチド配列は、１５、１６、１７、１８、１９、２０、２１、２２、２３、２４、または２５アミノ酸の長さを有するポリペプチド配列の各連続したペプチド配列を含み、対象の細胞のＨＬＡクラスＩまたはクラスＩＩ対立遺伝子によってコードされるタンパク質は、対象の細胞のクラスＩＩ　ＭＨＣ対立遺伝子によってコードされるタンパク質である。

【０２８０】

一態様では、自己免疫疾患または状態を有する対象を処置する方法であって、（ａ）対象の細胞のＨＬＡクラスＩまたはクラスＩＩによって提示される、発現されたタンパク質のエピトープを同定または予測するステップであって、同定または予測されたエピトープとＨＬＡクラスＩまたはクラスＩＩとを含む複合体が、対象のＣＤ８　Ｔ細胞またはＣＤ４Ｔ細胞の標的とされる、ステップと、（ｂ）複合体に結合するＴ細胞受容体（ＴＣＲ）を同定するステップと、（ｃ）ＴＣＲを対象由来の調節性Ｔ細胞または同種異系の調節性Ｔ細胞において発現させるステップと、（ｄ）ＴＣＲを発現する調節性Ｔ細胞を対象に投与するステップとを含む方法が本明細書で提供される。

【０２８１】

一実施形態では、自己免疫疾患または状態は、糖尿病である。

【０２８２】

一実施形態では、細胞は、膵島細胞である。

【０２８３】

一態様では、自己免疫疾患または状態を有する対象を処置する方法であって、対象に、（ｉ）対象の細胞のＨＬＡクラスＩまたはクラスＩＩによって提示されることが同定または予測された、発現されたタンパク質のエピトープと（ｉｉ）ＨＬＡクラスＩまたはクラスＩＩとを含む複合体に結合するＴ細胞受容体（ＴＣＲ）を発現する調節性Ｔ細胞を投与するステップを含み、複合体が、対象のＣＤ８　Ｔ細胞またはＣＤ４　Ｔ細胞の標的とされる、方法が本明細書で提供される。

【０２８４】

本開示の単に例示的な実施形態が示され、記載されている以下の詳細な説明から、本開示のさらなる態様および利点が当業者には容易に明らかになろう。理解される通り、本開示は、他のおよび異なる実施形態も可能であり、全て本開示から逸脱することなく、そのいくつかの詳細を種々の明白な観点で改変することができる。したがって、図および説明は、事実上例示的なものであり、制限するものではないとみなされるべきである。

【０２８５】

ＭＡＰＴＡＣ（商標）をハイスループットペプチド結合アッセイのために使用することができ、その場合、ＬＣ－ＭＳ／ＭＳを使用し、異なる時点において異なる条件下で、例えば３７℃での加熱下で、ＭＡＰＴＡＣ（商標）構築物を用いた単離後にＨＬＡクラスＩＩに結合したペプチドを測定して、異なる安定性を有するペプチドの集団の配列を得る。

【０２８６】

一態様では、対象におけるがんを処置するための方法であって、がんに関連付けられるペプチド配列を同定するステップであって、対象のがん細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列を対象の正常細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列と比較することを含むステップと、コンピュータプロセッサを使用して、同定されたペプチド配列のアミノ酸情報を機械学習ＨＬＡ－ペプチド提示予測モデルに入力して、同定されたペプチド配列についての提示予測のセットを生成するステップであって、各提示予測が、対象の

細胞のＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質が同定されたペプチド配列のうちの所与の配列を提示する確率を表し、機械学習ＨＬＡ－ペプチド提示予測モデルが、少なくとも訓練用データに基づいて同定された複数の予測変数を含み、訓練用データが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；アミノ酸位置情報を含む訓練用ペプチド配列情報であって、細胞において発現されるＨＬＡタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られたアミノ酸位置情報とアミノ酸位置情報および予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む、ステップと、提示予測のセットに基づいて同定された、個別化がん治療を調製するためのペプチド配列のサブセットを選択するステップと、対象にペプチドの１つまたは複数を含む組成物を投与するステップであって、予測モデルの陽性適中率が、少なくとも０．１％、０．１％～５０％または最大で５０％の再現率で少なくとも０．１である、ステップとを含む方法が本明細書で提供される。

【０２８７】

一部の実施形態では、機械学習ＨＬＡ－ペプチド提示予測モデルは、細胞において発現されるＨＬＡタンパク質によって提示され、逆相オフライン分画を実施した後に質量分析によって同定されたペプチドの配列の配列情報を含む。

【０２８８】

一部の実施形態では、予測モデルは、ＮｅｔＭＨＣＩＩｐａｎと比較して１．１倍～１００倍の改善を示す。一部の実施形態では、予測モデルは、ＮｅｔＭＨＣＩＩｐａｎと比較して１．１倍、２倍、３倍、４倍、５倍、６倍、７倍、７．４倍、８倍、９倍、１０倍、１１倍、１２倍、１３倍、１４倍、１５倍、１６倍、１７倍、１８倍、１８倍、２０倍、２１倍、２２倍、２３倍、２４倍、２５倍、２６倍、２７倍、２８倍、２９倍、３０倍、３１倍、３２倍、３３倍、３４倍、３５倍、３６倍、３７倍、３８倍、３９倍、４０倍、４１倍、４２倍、４３倍、４４倍、４５倍、５０倍、５５倍、６０倍、６１倍、６２倍、６３倍、６４倍、６５倍、６６倍、６７倍、６８倍、６９倍、７０倍、７１倍、７２倍、７３倍、７４倍、７５倍、７６倍、７７倍、７８倍、７９倍、８０倍、８１倍、８倍、８３倍、８４倍、８５倍、８６倍、８７倍、８８倍、８９倍、９０倍、９１倍、９２倍、９３倍、９４倍、９５倍、９６倍、９７倍、９８倍、９９倍、１００倍またはそれよりも大きな改善を示す。

【０２８９】

参照による組込み

本明細書において言及されている全ての刊行物、特許および特許出願は、個々の刊行物、特許、または特許出願が、具体的にかつ個別に参照により組み込まれることが示されたものと同じく参照により本明細書に組み込まれる。参照により組み込まれている刊行物および特許または特許出願が本明細書に含有される開示と矛盾する範囲では、本明細書がそのような矛盾する材料のいずれにも取って代わり、かつ／または優先するものとする。

【０２９０】

本発明の新規の特徴は、添付の特許請求の範囲において詳細に記載されている。本発明の原理が利用されている例示的な実施形態が記載されている以下の詳細な説明、および以下の付属図（本明細書では「図」とも）を参照することにより、本発明の特徴および利点のよりよい理解が得られよう。

【図面の簡単な説明】

【０２９１】

【図１Ａ】図１Ａは、ＭＨＣクラスＩタンパク質とドッキングしたペプチドを表す図である。図は、配列番号３６を開示する。

【０２９２】

【図１Ｂ】図１Ｂは、ＭＨＣクラスＩＩタンパク質とドッキングしたペプチドを表す例示的な図を示す。図は、配列番号３７を開示する。

【０２９３】

【図２】図２は、単一対立遺伝子ＨＬＡクラスＩＩ結合性ペプチドデータを生成するための例示的な実験手法を示す。ＨＬＡクラスＩＩペプチドを、ＨＬＡクラスＩＩを発現しない細胞を含めた任意の細胞に導入し、その結果、特定のＨＬＡクラスＩＩ対立遺伝子（複数可）を細胞において発現させる。遺伝子操作されたＨＬＡ発現細胞の集団を回収し、溶解させ、それらのＨＬＡ－ペプチド複合体にタグ付けし（例えば、ビオチン化）、免疫精製する（例えば、ビオチン－ストレプトアビジン相互作用を使用して）。単一のＨＬＡに特異的なＨＬＡ付随ペプチドをそれらのタグ付けされた（例えば、ビオチン化された）複合体から溶出させ、評価する（例えば、高分解能ＬＣ－ＭＳ／ＭＳを使用して配列決定を行う）ことができる。

【０２９４】

【図３】図３は、Ｎｅｏｎ　ＢＡＰ、Ｅｘｐｉ２９３細胞株；Ｎｅｏｎ　ＢＡＰ、Ａ３７５細胞株；ＩＥＤＢ、親和性＜５０ｎＭ；および汎ＨＬＡ　Ｃｌａｓｓ　ＩＩ　Ａｂ、ホモ接合性ＬＣＬにわたるＨＬＡクラスＩＩ－ＤＲＢ１＊１１：０１関連ペプチドの例示的な配列ロゴ表示を示す。図３は、ＭＳ由来モチーフの例が既知パターンとマッチし、トランスフェクトされた細胞株にわたって一貫性を示すことを示す。

【０２９５】

【図４】図４は、ＨＬＡクラスＩＩ結合予測器の性能の例示的な描写である。図４は、ヒットペプチドをランダムにシャッフリングすることによって１：１９（ヒット：デコイ）の比で生成された観察された質量分析ペプチドおよびデコイペプチドからなる検証用データセットに適用した結合予測器（ｎｅｏｎｍｈｃ２）およびＮｅｔＭＨＣＩＩｐａｎの性能を示す棒プロットである。ＮＥＯＮ結合予測器ｎｅｏｎｍｈｃ２に関しては、示されている各ＭＨＣ　ＩＩ対立遺伝子に対して別々のモデルを築く。棒の高さは、検証セット内の実際にヒットペプチドであった予測された結合体の分率と定義される陽性適中率（ＰＰＶ）を示す。対立遺伝子は、その対立遺伝子について予測した場合のモデルの性能によってソートされている。

【０２９６】

【図５】図５は、結合予測器の検証に対するスコア化ピーク強度（ＳＰＩ）閾値の例示的な影響を示す。図５は、種々のスコア化ピーク強度（ＳＰＩ）カットオフを有するペプチドのセットで訓練／検証を行った場合のＨＬＡクラスＩＩ結合予測器の性能を示す。訓練された各対立遺伝子特異的モデルについて、以下の３つの設定でのモデルの性能が示されている：７０　ＳＰＩよりも大きいまたはそれと等しい観察されたＭＳヒットペプチドを使用してデータセットに対して訓練および評価を行う、５０　ＳＰＩよりも大きいまたはそれと等しいペプチドで訓練を行い、７０　ＳＰＩよりも大きいまたはそれと等しいペプチドで検証を行う、ならびに、５０　ＳＰＩよりも大きいまたはそれと等しいペプチドで訓練および検証を行う。

【０２９７】

【図６】図６は、ＬＣ－ＭＳ／ＭＳによる対立遺伝子プロファイリングによって観察された、７０スコア化ピーク強度（ＳＰＩ）カットオフよりも大きいまたはそれと等しいペプチドの数からの代表的データを示す例示的な棒プロットを示す。各棒は、対立遺伝子の観察されたペプチドの総数を表す。３５種のＨＬＡ－ＤＲ対立遺伝子についての収集データが存在する。３５種のＨＬＡ－ＤＲ対立遺伝子についての収集データは、ＨＬＡ－ＤＲに関して＞９５％の集団カバレッジを有する（ＵＳＡ対立遺伝子頻度）。

【０２９８】

【図７Ａ】図７Ａは、示されているＨＬＡクラスＩＩ対立遺伝子についてデータの試験区分に適用した場合のモデルのＰＰＶを示す。使用したデコイペプチドは、ヒット対デコイ比１：１９で陽性（ヒット）ペプチド配列のスクランブル配列であった。ＰＰＶを、試験区分におけるペプチドの上位スコア化５％を同定し、それぞれのＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質への結合について陽性であるそれらの分率を決定することによって決定した。

【０２９９】

【図７Ｂ－１】図７Ｂは、訓練セットのサイズに応じた例示的な予測性能（訓練セットを人為的にダウンサンプリングすることによって得られた曲線）を示す。図７Ｂは、一般に、収集した３５種のＨＬＡ－ＤＲ対立遺伝子について、訓練セットのサイズを拡大すると、ＰＰＶの値が増大することを示す。
【図７Ｂ－２】図７Ｂは、訓練セットのサイズに応じた例示的な予測性能（訓練セットを人為的にダウンサンプリングすることによって得られた曲線）を示す。図７Ｂは、一般に、収集した３５種のＨＬＡ－ＤＲ対立遺伝子について、訓練セットのサイズを拡大すると、ＰＰＶの値が増大することを示す。
【図７Ｂ－３】図７Ｂは、訓練セットのサイズに応じた例示的な予測性能（訓練セットを人為的にダウンサンプリングすることによって得られた曲線）を示す。図７Ｂは、一般に、収集した３５種のＨＬＡ－ＤＲ対立遺伝子について、訓練セットのサイズを拡大すると、ＰＰＶの値が増大することを示す。
【０３００】
【図８】図８は、プロセシングに関連する変数により予測をさらに改善することができることを実証する例示的なグラフを示す。タンパク質をコードするエクソームから選択されたＭＳにより観察されたペプチドランダム配列を区別することができる。訓練用データ区分に対して、結合強度（ＮｅｔＭＨＣＩＩｐａｎまたはＮｅｏｎの予測器）およびプロセシング特色（ＲＮＡ－Ｓｅｑ発現および派生遺伝子レベルの偏り項）を使用してロジスティック回帰を当てはめて、ＨＬＡクラスＩＩ提示を予測することができる。別々の評価区分に対して、ＭＳにより観察されたＭＨＣ　ＩＩペプチド（「ヒット」）と重複するエクソンの位置を、ＭＳで観察されなかったランダムなエクソンの位置と並行してスコア化することができる（１：４９９の比）。上位０．２％（１／５００）を陽性とコールすることができ、陽性適中率を、この閾値を用いて評価することができる。
【０３０１】
【図９】図９は、例示的なニューラルネットワークアーキテクチャを示す。入力ペプチドが２０ｍｅｒとして示され、それよりも短いペプチドには「欠損」文字が記入される。各ペプチドは３１次元の埋め込みを有し、したがって、ニューラルネットワークへの入力は２０×３１行列である。ニューラルネットワークによって処理する前に、２０×３１行列に対する特色の正規化を訓練セットにおける特色値の平均および標準偏差に基づいて実施する。第１の畳み込み層は、９アミノ酸および５０フィルター（チャネルとも称される）のカーネルを有し、Ｒｅｃｔｉｆｉｅｄ　Ｌｉｎｅａｒ　Ｕｎｉｔ（ＲｅＬＵ）活性化関数を用いる。この後に、バッチ正規化、次いでドロップアウト率２０％での空間ドロップアウトが続く。この後に、３アミノ酸および２０フィルターのカーネルを有し、ＲｅＬＵ活性化関数を用いる別の畳み込み層、次いで、再度バッチ正規化およびドロップアウト率２０％での空間ドロップアウトが続く。次いで、２０フィルターの各々において最大限に活性化されたニューロンを取ってグローバル最大プーリングを適用し、次いで、これらの２０個の値を、シグモイド活性化関数を使用して単一のニューロンを有する全結合（高密度）層を通過させる。この層の出力を結合／非結合予測として扱う。Ｌ２正則化をそれぞれ０．０５、０．１、および０．０１の重みを有する第１の畳み込み層、第２の畳み込み層、および高密度層の重みに適用する。使用した追加的なモデルでは畳み込み層の数および各層のカーネルサイズを変動させた。
【０３０２】
【図１０】図１０は、本明細書で提供される方法をインプリメントするようにプログラムされたまたは他のやり方で構成された例示的なコンピュータ制御システムを示す。
【０３０３】
【図１１Ａ】図１１Ａは、ＭＡＰＴＡＣ（商標）実験ワークフローの例示的な概要を示す。図は、配列番号３８を開示する。
【０３０４】
【図１１Ｂ】図１１Ｂは、反復実験にわたってマージした例示的な対立遺伝子当たりのペプチド計数を示す。

【０３０５】
【図１１Ｃ】図１１Ｃは、ＭＡＰＴＡＣ（商標）によってプロファイリングされたＨＬＡクラスＩおよびＨＬＡクラスＩＩ対立遺伝子についての例示的なペプチド長分布を示す。
【０３０６】
【図１１Ｄ】図１１Ｄは、ＭＡＰＴＡＣ（商標）およびＩＥＤＢ（対立遺伝子ＤＲＢ１＊０１：０１、ＤＲＢ１＊０３：０１、ＤＲＢ１＊０９：０１、およびＤＲＢ１＊１１：０１）、ヒトプロテオーム、ならびに以前の刊行物からの複対立遺伝子ＭＳデータについて観察された例示的な残基当たりのシステイン頻度を示す。
【０３０７】
【図１２Ａ】図１２Ａは、個体の＞１％に存在するＨＬＡ－ＤＲ、ＨＬＡ－ＤＰ、およびＨＬＡ－ＤＱ対立遺伝子についてのコーカサス人での頻度、および、強力な結合体（＜５０ｎＭ）として測定された、示されている供給源由来のペプチドの計数を示す。
【０３０８】
【図１２Ｂ】図１２Ｂは、ＩＥＤＢペプチドの例示的な長さ分布を関連するＨＬＡクラスＩＩ親和性測定値と共に示す。
【０３０９】
【図１２Ｃ】図１２Ｃは、（１）Ｅｘｐｉ２９３、（２）ＨｅＬａ、および（３）２種のＨＬＡクラスＩ対立遺伝子および２種のＨＬＡクラスＩＩ対立遺伝子：ＨＬＡ－Ａ＊０２：０１、ＨＬＡ－Ｂ＊４５：０１、ＨＬＡ－ＤＲＢ１＊０１：０１、およびＨＬＡ－ＤＲＢ１＊１１：０１を個別にトランスフェクトしたＡ３７５細胞株の例示的なウエスタンブロットを示す。膜に抗ビオチンリガーゼエピトープタグをブロットして、ビオチンアクセプターペプチド（ＢＡＰ）、およびローディング対照として抗ベーター－チューブリンを可視化した。レーンはＭＡＰＴＡＣ（商標）プロトコールの間に収集された以下の画分に対応した：レーン１、入力、レーン２、ビオチン化入力、およびレーン３、プルダウン後の入力。
【０３１０】
【図１２Ｄ】図１２Ｄは、ＭＡＰＴＡＣ（商標）およびＩＥＤＢ（対立遺伝子ＤＲＢ１＊０１：０１、ＤＲＢ１＊０３：０１、ＤＲＢ１＊０９：０１、およびＤＲＢ１＊１１：０１）、ヒトプロテオーム、ならびに以前の刊行物からの複対立遺伝子ＭＳデータについて観察された例示的な残基当たりのアミノ酸頻度を示す。
【０３１１】
【図１２Ｅ】図１２Ｅは、個体の＞１％に存在するＨＬＡ－ＤＲ、ＨＬＡ－ＤＰ、およびＨＬＡ－ＤＱ対立遺伝子についてのコーカサス人での頻度、および、強力な結合体（＜５０ｎＭ）として測定された、示されている供給源由来のペプチドの計数を示す。この図は、図１２Ａに対する追加的なデータを含む。追加的なデータはｔｏｏｌｓ．ｉｅｄｂ．ｏｒｇ／ｍａｉｎ／ｄａｔａｓｅｔｓ／から取得した。
【０３１２】
【図１２Ｆ】図１２Ｆは、ＭＡＰＴＡＣ（商標）（還元およびアルキル化されたもの）、ＭＡＰＴＡＣ（商標）（無処理）およびＩＥＤＢ（対立遺伝子ＤＲＢ１＊０１：０１、ＤＲＢ１＊０３：０１、ＤＲＢ１＊０９：０１、およびＤＲＢ１＊１１：０１）、ヒトプロテオーム、ならびに以前の刊行物からの複対立遺伝子ＭＳデータについて観察された例示的な残基当たりのアミノ酸頻度を示す。
【０３１３】
【図１３】図１３は、ＭＡＰＴＡＣ（商標）およびＩＥＤＢ毎のＭＨＣ　ＩＩ対立遺伝子についてのコア結合性配列ロゴの例示的な表示を示す。配列ロゴは図表示であり、各アミノ酸の高さは、対立遺伝子によってコードされるＭＨＣタンパク質に結合するペプチドにおけるそのアミノ酸の発生頻度に比例する。エントロピーが最も低い位置が色で示されており、その色がアミノ酸特性に対応する。ペプチドは示されているデータセットに由来し、ＣＮＮに基づく予測器（方法）に従ってアラインメントされている。ロゴは、全体的なモチーフと密接にはマッチしないものを含めた全てのペプチドを表す（例えば、「トラッ

シュ」クラスターに隔離されるペプチドがない）。
【０３１４】
【図１４Ａ】図１４Ａは、２つの異なる細胞株（Ａ３７５＆ｅｘｐｉ２９３）において、結合アッセイ、安定性アッセイ、可溶性ＨＬＡ（ｓＨＬＡ）質量分析、単一対立遺伝子質量分析、およびＭＡＰＴＡＣ（商標）を含めた異なるＨＬＡ－リガンドプロファイリング技術を使用して解析されたＨＬＡ－Ａ＊０２：０１結合性ペプチド（リガンド）についての例示的な配列ロゴを示す。
【０３１５】
【図１４Ｂ】図１４Ｂは、ヒューリスティックに定義されたアンカーを０個、１個、２個、３個、および４個示すＭＡＰＴＡＣ（商標）ペプチドの例示的な分率を示す。
【０３１６】
【図１４Ｃ】図１４Ｃは、ＭＡＰＴＡＣ（商標）により観察されたペプチド（対立遺伝子当たり２０種のペプチド、各々がＳＰＩ＞７０およびサイズが２以上の入れ子セットを有する）ならびにプロテオームからサンプリングした、長さを釣り合わせたデコイについてのＮｅｔＭＨＣＩＩｐａｎにより予測された結合親和性の例示的な分布を示す。
【０３１７】
【図１５Ａ】図１５Ａは、単一対立遺伝子ＭＨＣペプチドとスクランブルされた長さを釣り合わせたデコイを区別するように訓練された畳み込みニューラルネットワーク（ＣＮＮ）の例示的なアーキテクチャを示す。概略図は、アミノ酸特色の埋め込み、フィルターサイズが異なる２つの畳み込み層の使用、ならびにグローバル最大プーリングの、最終的なロジスティック出力ノードへの入力としての使用を示す。
【０３１８】
【図１５Ｂ】図１５Ｂは、ｎｅｏｎｍｈｃ２またはＮｅｔＭＨＣＩＩｐａｎのいずれかからの結合予測を用いて測定されたＩＥＤＢ親和性の相関についてのケンドールのタウ統計値を示す例示的な結果である。評価されたペプチドは、ＮｅｔＭＨＣＩＩｐａｎが発表された翌年にＩＥＤＢにポストされたもののみを含む。
【０３１９】
【図１６】図１６は、訓練用データセットのサイズに応じたｎｅｏｎｍｈｃ２の性能の例示的な描写である。
【０３２０】
【図１７Ａ】図１７Ａは、汎ＤＲおよび汎クラスＩＩ　ＭＨＣ　ＭＳデータセットにスパイクされたＭＡＰＴＡＣ（商標）ペプチド（対立遺伝子当たり２０種）についての例示的なクラスター割り当てを示す。データセットに対して、ＧｉｂｂｓＣｌｕｓｔｅｒを使用して逆畳み込みを行った。各四角は１つのＭＡＰＴＡＣ（商標）ペプチドを表す。四角の色は、それが割り当てられるクラスターを示し、灰色の棒は、ペプチドが実際に由来する対立遺伝子を示す。ギブスクラスター解法におけるクラスターの総数（右側）を、相互情報（ＭＩ）メトリクスを使用して選択した。ＭＩスコアにより、試料がどのように選別されるかも決定される；高ＭＩ解法の試料が上位に現れる。
【０３２１】
【図１７Ｂ】図１７Ｂは、ＧｉｂｂｓＣｌｕｓｔｅｒによって逆畳み込みされた複対立遺伝子ＭＳデータについての例示的なコア結合性配列ロゴを示す。ペプチドの各セットは、ＭＡＰＴＡＣ（商標）スパイクインに最良にアラインメントされるクラスターに対応する。
【０３２２】
【図１７Ｃ】図１７Ｃは、ＭＡＰＴＡＣ（商標）データまたは逆畳み込みされた複対立遺伝子データのいずれかを使用してホールドアウトＭＡＰＴＡＣ（商標）ペプチドを予測するモデルの代表的な性能を示す。各対立遺伝子について、２つのデータ供給源のうちの大きい方（通常ＭＡＰＴＡＣ（商標））をダウンサンプリングし、したがって、予測器は等しい数の訓練例に基づく。ＮｅｔＭＨＣＩＩｐａｎ性能がさらなる比較として示されている。
【０３２３】

【図１７Ｄ】図１７Ｄは、示されている供給源からの複対立遺伝子ＭＳデータに由来する例示的なコア結合性配列ロゴを示す。

【０３２４】

【図１８Ａ】図１８Ａは、ペプチド対供給源遺伝子発現の分率（ＭＳにより観察されたペプチドおよびランダムなプロテオームデコイについての百万当たりの転写物数（ＴＰＭ））の例示的なグラフを示す（Schuster et al. 2017から再プロットしたデータ）。

【０３２５】

【図１８Ｂ】図１８Ｂは、結腸直腸がん、黒色腫、および卵巣がんデータセットの共同解析によって決定される、遺伝子当たりのクラスＩＩペプチドの例示的な観察数と予測数とを示す（Loeffler et al., 2018、およびSchuster et al., 2017）。予測計数は、遺伝子の長さに発現レベルを掛けることによって導き出される。予測計数および観察計数を関連性のある試料にわたって合計した。血漿中に存在することが公知である遺伝子にそれらの濃度に従って印をつけた（挿入図）。

【０３２６】

【図１８Ｃ】図１８Ｃは、オートファジーに関連する遺伝子についての富化スコア（観察された観察対予測された観察の比、図１８Ｂと同様）の例示的な分布を示す。

【０３２７】

【図１８Ｄ】図１８Ｄは、各供給源遺伝子の局在化に従った富化スコアの例示的な分布を示す。供給源遺伝子の局在化を、Ｕｎｉｐｒｏｔ（ｕｎｉｐｒｏｔ＿ｓｐｒｏｔ．ｄａｔ）を使用して決定した。

【０３２８】

【図１８Ｅ】図１８Ｅは、細胞内局在化特性に基づいて隔離されたＭＨＣ－ＩＩ結合親和性を有するペプチドの総数の分率の予測された頻度と観察された頻度の比較を表す例示的なデータを示す。

【０３２９】

【図１８Ｆ】図１８Ｆは、２つの異なる遺伝子発現プロファイルに関する観察におけるペプチドの相対的一致の例示的な代表的データを示す。各試料について、遺伝子レベルペプチド計数をバルク腫瘍遺伝子発現およびプロフェッショナルＡＰＣ（マクロファージ）遺伝子発現プロファイルの線形結合としてモデル化した。係数の比により、各発現プロファイルとペプチドレパートリーの相対的一致が決定される。エラーバーはブートストラップリサンプリングによってコンピュータ計算された９５％信頼区間に対応した。

【０３３０】

【図１９Ａ】図１９Ａは、５つの試験例におけるＨＬＡ－ＤＲＢ１の発現レベルの例示的な代表的データを示す。各ドットは、細胞にわたって平均した、個々の患者の個々の細胞型における発現を表す。

【０３３１】

【図１９Ｂ】図１９Ｂは、ＴＣＧＡ患者のＲＮＡ－Ｓｅｑから入力された腫瘍および間質由来ＨＬＡ－ＤＲＢ１発現の例示的な代表的データを示す。横棒は個々の患者に対応し、腫瘍型によって群分けされている。患者が、ＤＮＡに基づく変異コールによって決定されるＨＬＡクラスＩＩ経路遺伝子（ＣＩＩＴＡ、ＣＤ７４またはＣＴＳＳＳ）の変異を有する場合、その患者を含めた。各患者について、腫瘍に起因するＨＬＡ－ＤＲＢ１発現の分率をｍｉｎ（１，２ｆ）として推定し、ここで、ｆは、変異を示すＣＩＩＴＡ、ＣＤ７４、またはＣＴＳＳにおけるＲＮＡ－Ｓｅｑリードの分率である。

【０３３２】

【図１９Ｃ】図１９Ｃは、チェックポイント遮断免疫治療前およびチェックポイント遮断免疫治療後の生検を含む追加的な単一細胞ＲＮＡ－Ｓｅｑ試験の例示的な代表的データを示す。

【０３３３】

【図２０】図２０は、天然のドナー組織に対する予測の全体的な性能を評価する例示的な代表的な実験データを示す。

【０３３４】
【図２１Ａ】図２１Ａは、細胞ＨＬＡクラスＩＩリガンドームを予測する統合された提示モデルを示す例示的な代表的データを示す。これは、汎ＤＲデータセットについてのヒット対デコイ比１：４９９でのＰＰＶを表す（図３０Ｂおよび図３２Ｅにおいても解析されている）。予測器は、結合予測（ＮｅｔＭＨＣＩＩｐａｎまたはｎｅｏｎｍｈｃ２）を使用し、必要に応じて遺伝子発現、遺伝子の偏り（図３２Ａの通り）、および以前に観察されたＨＬＡ－ＤＱペプチドとの重複を使用するものである。各候補ペプチドについて、結合スコアを、試料遺伝子型におけるＨＬＡ－ＤＲ対立遺伝子にわたる最大値として算出した。
【０３３５】
【図２１Ｂ】図２１Ｂは、樹状細胞（細胞溶解物から解析）によって提示されるＳＩＬＡＣを使用して同定された腫瘍由来ペプチドについての、図２１Ａと同じヒット：デコイ比および性能メトリクスを使用し、プロセシング特色の使用を伴うまたは伴わない予測性能を示す例示的な代表的データを示す。
【０３３６】
【図２１Ｃ】図２１Ｃは、ＵＶ処理実験において観察された、重標識されたペプチドについての例示的な発現および遺伝子の偏りのスコア（赤色のドット、Ｋ５６２発現に従ってプロットした）を、軽標識されたペプチド（灰色のドット、ＤＣ発現に従ってプロットした）と比較して示す。
【０３３７】
【図２１Ｄ】図２１Ｄは、溶解物およびＵＶ処理実験による重標識されたペプチド供給源遺伝子の重複を表す例示的な図を示す。遺伝子名に機能的クラス毎に色が付いている。
【０３３８】
【図２２Ａ】図２２Ａは、ＨＬＡクラスＩＩにより駆動されるＣＤ４＋　Ｔ細胞およびＴ細胞応答を検証するための、本明細書に開示されるアッセイプロトコールを表す例示的なフロー図を示す。
【０３３９】
【図２２Ｂ】図２２Ｂは、ペプチド交換アッセイのための例示的なＨＬＡタンパク質二量体構築物の設計（上のパネル）および例示的なアッセイワークフローの図表示（下のパネル）を示す。図は、配列番号２０を「１０×Ｈｉｓ」として開示する。
【０３４０】
【図２３】図２３は、新しい結合性ペプチドをスクリーニングするためのＭＨＣ－ＩＩ発現のための例示的なベクター設計の例示的な図解、および発現されたタンパク質産物の表示を示す。図は、配列番号３９を開示し、配列番号２０を「１０×Ｈｉｓ」として開示する。
【０３４１】
【図２４】図２４は、プレースホルダーペプチドのトランスフェクション、精製およびベータ鎖からの切断の例示的な流れ図を示す。
【０３４２】
【図２５Ａ】図２５Ａは、発現されたＭＨＣ－ＩＩペプチドの分泌の増加に関連するＣＬＩＰペプチドをコードするベクターを示す例示的な図解を示す。図は、配列番号２１を開示する。
【０３４３】
【図２５Ｂ】図２５Ｂは、それぞれＣＬＩＰ０をコードする核酸およびＣＬＩＰ１をコードする核酸のより短い形態およびより長い形態の例示的な図表示を示す。図は、出現する順に、それぞれ、配列番号１および２１を開示する。
【０３４４】
【図２５Ｃ】図２５Ｃは、より長いクリップを伴うまたは伴わないアルファ鎖およびベータ鎖のクーマシーゲル分析の例示的な代表的結果を示す。
【０３４５】

【図２６Ａ】図２６Ａは、ＴＲ－ＦＲＥＴアッセイの例示的な図解を示す。
【０３４６】
【図２６Ｂ】図２６Ｂは、特定のペプチドを使用した蛍光共鳴エネルギー移動（ＦＲＥＴ）アッセイを使用したＨＬＡクラスＩＩペプチド結合アッセイからの例示的な代表的な極性化データを示す。
【０３４７】
【図２６Ｃ】図２６Ｃは、特定のペプチドを使用した蛍光共鳴エネルギー移動（ＦＲＥＴ）アッセイを使用したＨＬＡクラスＩＩペプチド結合アッセイからの例示的な代表的な極性化データを示す。
【０３４８】
【図２６Ｄ】図２６Ｄは、蛍光の増大から算出したＭＨＣ－構築物結合ペプチドの例示的なパーセント置換えを示す。
【０３４９】
【図２６Ｅ】図２６Ｅは、蛍光の増大から算出したＭＨＣ－構築物結合ペプチドの例示的なパーセント置換えを示す。
【０３５０】
【図２６Ｆ】図２６Ｆは、示差走査蛍光光度法（ＤＳＦ）を使用したアッセイを使用した例示的なペプチド交換を示す。フルオロフォアの結合および高い蛍光がもたらされる、ＭＨＣクラスＩＩヘテロ二量体も同様に解離させる熱を用いたＭＨＣクラスＩＩからのペプチド解離の検出の例示的な機構を示す図表示が示されている。エピトープペプチドによるプレースホルダーペプチドの移動の例示的な概略図も示されている。温度に対してプロットした例示的な融解曲線も示されている。
【０３５１】
【図２６Ｇ】図２６Ｇは、例示的な可溶性ＨＬＡ－ＤＭ構築物およびＭＨＣクラスＩＩペプチド交換を実施するためのその使用を示す。示されている構築物は、ＣＭＶプロモーター、分泌配列（リーダー）の下流のＨＬＡ－ＤＭベータ鎖のコード配列およびＨＬＡ－ＤＭアルファ鎖のコード配列、ならびにベータ鎖コード配列の３’末端のＢＡＰ配列；アルファ鎖コード配列の３’末端におけるＨｉｓタグを含有する。２つの鎖は介在するリボソームスキッピング配列によって分離されている。構築物をＥｘｐｉ－ＣＨＯ細胞において発現させ、培養培地中に分泌されたタンパク質を精製した。図は、配列番号２０を「１０×Ｈｉｓ」として開示する。
【０３５２】
【図２６Ｈ】図２６Ｈは、ペプチド交換を実施するためのＨＬＡ－ｓＤＭを使用した例示的な分子ふるいクロマトグラフィーデータを示す。
【０３５３】
【図２７Ａ】図２７Ａは、例示的なＤＲＢ四量体レパートリーを築く例示的な図解を示す。
【０３５４】
【図２７Ｂ】図２７Ｂは、例示的なクラスＩＩ四量体レパートリーを築く例示的な図解を示す。
【０３５５】
【図２７Ｃ】図２７Ｃは、ペプチド交換のためのＤＲＢ１対立遺伝子についてのＤＲＢ四量体レパートリーカバレッジの要約の例示的な図解を示す。
【０３５６】
【図２７Ｄ】図２７Ｄは、ヒトＭＨＣクラスＩＩ対立遺伝子産生の例示的なカバレッジを示す。
【０３５７】
【図２７Ｅ】図２７Ｅは、Ｆｌｕエピトープ（記憶応答）またはＨＩＶエピトープ（ナイーブな応答）を用いて誘導された試料の四量体染色からの例示的な結果を示す。
【０３５８】
【図２８Ａ】図２８Ａは、ペプチドを、ＨＬＡクラスＩＩ制限について、エピトープペプ

チドについての対立遺伝子制限を迅速に同定するためのスクリーニング方法を可能にする蛍光偏光アッセイによって評価する方法の例示的な図表示を示す。図２８Ａのアッセイ原理により、親和性測定、およびペプチド交換の明白な測定が可能になる。

【０３５９】

【図２８Ｂ】図２８Ｂは、ＤＲＢ１＊０１：０１を用いた蛍光偏光アッセイにおいて探究された多数のアッセイ条件の例示的な要約を示す（上のパネル）。可溶性ＭＨＣクラスＩＩ対立遺伝子および界面活性剤ミセル内に膜貫通ドメインを有する全長ＭＨＣクラスＩＩ対立遺伝子の図表も示されており（下のパネル）、これらはどちらも、アッセイに使用するための切断可能なリンカーを有するプレースホルダーペプチドを用いて構築されたものである。

【０３６０】

【図２８Ｃ】図２８Ｃは、図２８Ｂの下のパネルにおいて前に示された全長対立遺伝子および可溶性対立遺伝子を調査するためのアッセイの例示的な図表示を示す。手短に言えば、全長対立遺伝子および可溶性対立遺伝子の両方を細胞において発現させる。膜に結合した全長対立遺伝子の形態は膜を透過処理することによって回収し、一方、分泌形態は細胞の上清から回収する。回収されたクラスＩＩ　ＨＬＡ対立遺伝子タンパク質を、ニッケル（Ni^{2+}）カラムを通過させることによって精製する。

【０３６１】

【図２８Ｄ】図２８Ｄは、精製方法がペプチド効力に影響を及ぼさないことを示す例示的なデータを示す。左側にはＬ２４３により精製された全長ＨＬＡ－ＤＲ１およびNi^{2+}により精製された全長ＨＬＡ－ＤＲ１を使用した実験からの平均ＩＣ５０値が示されている。

【０３６２】

【図２８Ｅ】図２８Ｅは、可溶性形態（ｓＤＲ１）または全長形態（ｆＤＲ１）の選択がペプチド効力に影響を及ぼさないことを示す例示的なデータを示す。左側にはｓＤＲ１形態またはｆＤＲ１を使用した実験からの平均ＩＣ５０値が示されている。ＦＰ、蛍光偏光。

【０３６３】

【図２８Ｆ】図２８Ｆは、ｎｅｏｎｍｈｃ２により予測されたペプチドおよびＮｅｔＭＨＣＩＩｐａｎにより予測されたペプチドの結合アッセイおよび不調和ペプチドの同定における例示的な評価の例示的な図解を示す。

【０３６４】

【図２８Ｇ】図２８Ｇは、ｎｅｏｎｍｈｃ２により予測されたペプチドを評価するための例示的な蛍光偏光結合スクリーニングデータを示す；ヒートマップとして、また使用したペプチドの各濃度について示されているプローブ結合のパーセント阻害としても示されている。緑色は、色の強度に比例して良好な結合を示す。黄色は中間の結合を示し、赤色は不十分な結合を示し、また、対応するパーセント阻害値としても示されている。

【０３６５】

【図２８Ｈ】図２８Ｈは、ｎｅｏｎｍｈｃ２により予測されたペプチドの評価の例示的な結合アッセイにおける要約を示す。

【０３６６】

【図２９】図２９は、各ＨＬＡ対立遺伝子毎の平均ＭＡＰＴＡＣ（商標）実験的反復（細胞５０００万個）からのペプチドの例示的な平均計数を示す。

【０３６７】

【図３０Ａ－Ｂ】図３０Ａ～３０Ｃは、ＨＬＡクラスＩＩ　ＭＡＰＴＡＣ（商標）対立遺伝子＋／－ＨＬＡ－ＤＭについての例示的な結合性コア解析および複対立遺伝子逆畳み込みの忠実度を示す。図３０Ａは、ＨＬＡ－ＤＭ同時トランスフェクション（ｅｘｐｉ２９３細胞株）を伴うまたは伴わないＭＡＰＴＡＣ（商標）およびＩＥＤＢに従った、ＨＬＡ－ＤＲ、ＨＬＡ－ＤＱ、およびＨＬＡ－ＤＰ対立遺伝子についての代表的な例示的な配列ロゴの１つを示し、各アミノ酸の高さはその頻度に比例する。頻度が１０％よりも高いアミノ酸は化学的特性に応じて色で示されている；他のアミノ酸は全て灰色で示されている

。ペプチドをＧｉｂｂｓＣｌｕｓｔｅｒツール（補足的方法）に従ってアラインメントし、ロゴは、モチーフ全体と密接にはマッチしないものを含めた全てのペプチドを表す（例えば、「トラッシュ」クラスターに隔離されるペプチドがない）。図３０Ｂは、汎ＤＲＭＳデータセットにスパイクインされたＭＡＰＴＡＣ（商標）ペプチド（対立遺伝子当たり２０種）についてのクラスター割り当ての例示的な記述を示す。データセットに対して、ＧｉｂｂｓＣｌｕｓｔｅｒを使用して逆畳み込みを行った。色付きの四角は各々が１つのＭＡＰＴＡＣ（商標）ペプチドを表す。四角の色は、それが割り当てられるクラスターを示し、灰色の棒は、ペプチドが由来する対立遺伝子を示す。図３０Ｃは、図３０Ｂに示されている対立遺伝子についての、アンカー位置において予測された残基を０個、１個、２個、３個、または４個を示すペプチドの共有を示す例示的なグラフを示す。アンカー位置をエントロピーが最も低い４つの位置と定義し、「予測される」残基をそれらの位置における頻度が１０％以上である残基と定義した。

【図３０Ｃ】図３０Ａ～３０Ｃは、ＨＬＡクラスＩＩ　ＭＡＰＴＡＣ（商標）対立遺伝子＋／－ＨＬＡ－ＤＭについての例示的な結合性コア解析および複対立遺伝子逆畳み込みの忠実度を示す。図３０Ａは、ＨＬＡ－ＤＭ同時トランスフェクション（ｅｘｐｉ２９３細胞株）を伴うまたは伴わないＭＡＰＴＡＣ（商標）およびＩＥＤＢに従った、ＨＬＡ－ＤＲ、ＨＬＡ－ＤＱ、およびＨＬＡ－ＤＰ対立遺伝子についての代表的な例示的な配列ロゴの１つを示し、各アミノ酸の高さはその頻度に比例する。頻度が１０％よりも高いアミノ酸は化学的特性に応じて色で示されている；他のアミノ酸は全て灰色で示されている。ペプチドをＧｉｂｂｓＣｌｕｓｔｅｒツール（補足的方法）に従ってアラインメントし、ロゴは、モチーフ全体と密接にはマッチしないものを含めた全てのペプチドを表す（例えば、「トラッシュ」クラスターに隔離されるペプチドがない）。図３０Ｂは、汎ＤＲ　ＭＳデータセットにスパイクインされたＭＡＰＴＡＣ（商標）ペプチド（対立遺伝子当たり２０種）についてのクラスター割り当ての例示的な記述を示す。データセットに対して、ＧｉｂｂｓＣｌｕｓｔｅｒを使用して逆畳み込みを行った。色付きの四角は各々が１つのＭＡＰＴＡＣ（商標）ペプチドを表す。四角の色は、それが割り当てられるクラスターを示し、灰色の棒は、ペプチドが由来する対立遺伝子を示す。図３０Ｃは、図３０Ｂに示されている対立遺伝子についての、アンカー位置において予測された残基を０個、１個、２個、３個、または４個を示すペプチドの共有を示す例示的なグラフを示す。アンカー位置をエントロピーが最も低い４つの位置と定義し、「予測される」残基をそれらの位置における頻度が１０％以上である残基と定義した。

【０３６８】

【図３１Ａ】図３１Ａ～３１Ｆは、ｎｅｏｎｍｈｃ２結合予測アルゴリズムの例示的なアーキテクチャおよびベンチマーキングを示す。図３１Ａは、単一対立遺伝子ＨＬＡクラスＩＩペプチドと、スクランブルされた長さを釣り合わせたデコイとを区別するように訓練された畳み込みニューラルネットワーク（ＣＮＮ）の例示的なアーキテクチャを示す。概略図は、アミノ酸特色の埋め込み層、幅６の２つの畳み込み層、最後までスキップ接続の存在、および最終的なロジスティック出力ノードへの入力として平均プーリング操作と最大プーリング操作の組合せの使用を示す。図３１Ｂは、訓練またはハイパーパラメータ最適化に使用していないＭＡＰＴＡＣ（商標）データの区分に関して評価した、ＮｅｔＭＨＣＩＩｐａｎおよびｎｅｏｎｍｈｃ２についての例示的な陽性適中率（ＰＰＶ）を示す。各対立遺伝子について、ＭＳにより観察されたペプチドｎ個を、供給源遺伝子の同じセットからサンプリングした長さを釣り合わせたデコイ１９ｎ個と併せてスコア化し、各予測器の上位に順位付けされたペプチド（例えば、上位５％）ｎ個を陽性とコールした。この評価プロトコールによると、偽陽性の数と偽陰性の数が必ず等しくなるので、ＰＰＶは再現率と同一である。図３１Ｃは、ＴＧＥＭデータセットに対する例示的なＮｅｔＭＨＣＩＩｐａｎのＰＰＶおよびｎｅｏｎｍｈｃ２のＰＰＶを示す。各対立遺伝子について、上位に順位付けされたペプチドｎ個を陽性とコールし、ここで、ｎは、評価したセット内の確認された免疫原性エピトープの数である。図３１Ｄは、新抗原ペプチドについての例示的なｅｘ　ｖｉｖｏにおけるＴ細胞誘導の結果を示す。ペプチドを、ＨＬＡ－ＤＲＢ１＊１

１：０１についての高ｎｅｏｎｍｈｃ２スコアおよび弱いＮｅｔＭＨＣＩＩｐａｎスコアに基づいて選択した。図は、出現する順に、それぞれ、配列番号８７～８９、９１、９０、２、９２～９４、３および９５～９６を開示する。図３１Ｅは、単一対立遺伝子ＭＡＰＴＡＣデータに対して訓練したモデルと、ホールドアウト単一対立遺伝子データに対して評価される逆畳み込みされた複対立遺伝子データとの比較を示す。値は、訓練用データセットをダウンサンプリングして逆畳み込み訓練セットのサイズをマッチさせた場合のｎｅｏｎｍｈｃ２について示されている通りである。図３１Ｆは、ＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１、逆畳み込みにより訓練された予測器、およびｎｅｏｎｍｈｃ２のＴＧＥＭデータセットに対するＰＰＶを示す（ダウンサンプリングを伴う、および伴わない）。各対立遺伝子について、上位に順位付けされたペプチドｎ個を陽性とコールし、ここで、ｎは、評価されたセットにおいて確認された免疫原性エピトープの数である。

【図３１Ｂ】図３１Ａ～３１Ｆは、ｎｅｏｎｍｈｃ２結合予測アルゴリズムの例示的なアーキテクチャおよびベンチマーキングを示す。図３１Ａは、単一対立遺伝子ＨＬＡクラスＩＩペプチドと、スクランブルされた長さを釣り合わせたデコイとを区別するように訓練された畳み込みニューラルネットワーク（ＣＮＮ）の例示的なアーキテクチャを示す。概略図は、アミノ酸特色の埋め込み層、幅６の２つの畳み込み層、最後までスキップ接続の存在、および最終的なロジスティック出力ノードへの入力として平均プーリング操作と最大プーリング操作の組合せの使用を示す。図３１Ｂは、訓練またはハイパーパラメータ最適化に使用していないＭＡＰＴＡＣ（商標）データの区分に関して評価した、ＮｅｔＭＨＣＩＩｐａｎおよびｎｅｏｎｍｈｃ２についての例示的な陽性適中率（ＰＰＶ）を示す。各対立遺伝子について、ＭＳにより観察されたペプチドｎ個を、供給源遺伝子の同じセットからサンプリングした長さを釣り合わせたデコイ１９ｎ個と併せてスコア化し、各予測器の上位に順位付けされたペプチド（例えば、上位５％）ｎ個を陽性とコールした。この評価プロトコールによると、偽陽性の数と偽陰性の数が必ず等しくなるので、ＰＰＶは再現率と同一である。図３１Ｃは、ＴＧＥＭデータセットに対する例示的なＮｅｔＭＨＣＩＩｐａｎのＰＰＶおよびｎｅｏｎｍｈｃ２のＰＰＶを示す。各対立遺伝子について、上位に順位付けされたペプチドｎ個を陽性とコールし、ここで、ｎは、評価したセット内の確認された免疫原性エピトープの数である。図３１Ｄは、新抗原ペプチドについての例示的なｅｘ　ｖｉｖｏにおけるＴ細胞誘導の結果を示す。ペプチドを、ＨＬＡ－ＤＲＢ１＊１１：０１についての高ｎｅｏｎｍｈｃ２スコアおよび弱いＮｅｔＭＨＣＩＩｐａｎスコアに基づいて選択した。図は、出現する順に、それぞれ、配列番号８７～８９、９１、９０、２、９２～９４、３および９５～９６を開示する。図３１Ｅは、単一対立遺伝子ＭＡＰＴＡＣデータに対して訓練したモデルと、ホールドアウト単一対立遺伝子データに対して評価される逆畳み込みされた複対立遺伝子データとの比較を示す。値は、訓練用データセットをダウンサンプリングして逆畳み込み訓練セットのサイズをマッチさせた場合のｎｅｏｎｍｈｃ２について示されている通りである。図３１Ｆは、ＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１、逆畳み込みにより訓練された予測器、およびｎｅｏｎｍｈｃ２のＴＧＥＭデータセットに対するＰＰＶを示す（ダウンサンプリングを伴う、および伴わない）。各対立遺伝子について、上位に順位付けされたペプチドｎ個を陽性とコールし、ここで、ｎは、評価されたセットにおいて確認された免疫原性エピトープの数である。

【図３１Ｃ】図３１Ａ～３１Ｆは、ｎｅｏｎｍｈｃ２結合予測アルゴリズムの例示的なアーキテクチャおよびベンチマーキングを示す。図３１Ａは、単一対立遺伝子ＨＬＡクラスＩＩペプチドと、スクランブルされた長さを釣り合わせたデコイとを区別するように訓練された畳み込みニューラルネットワーク（ＣＮＮ）の例示的なアーキテクチャを示す。概略図は、アミノ酸特色の埋め込み層、幅６の２つの畳み込み層、最後までスキップ接続の存在、および最終的なロジスティック出力ノードへの入力として平均プーリング操作と最大プーリング操作の組合せの使用を示す。図３１Ｂは、訓練またはハイパーパラメータ最適化に使用していないＭＡＰＴＡＣ（商標）データの区分に関して評価した、ＮｅｔＭＨＣＩＩｐａｎおよびｎｅｏｎｍｈｃ２についての例示的な陽性適中率（ＰＰＶ）を示す。各対立遺伝子について、ＭＳにより観察されたペプチドｎ個を、供給源遺伝子の同じセッ

トからサンプリングした長さを釣り合わせたデコイ１９ｎ個と併せてスコア化し、各予測器の上位に順位付けされたペプチド（例えば、上位５％）ｎ個を陽性とコールした。この評価プロトコールによると、偽陽性の数と偽陰性の数が必ず等しくなるので、ＰＰＶは再現率と同一である。図３１Ｃは、ＴＧＥＭデータセットに対する例示的なＮｅｔＭＨＣＩＩｐａｎのＰＰＶおよびｎｅｏｎｍｈｃ２のＰＰＶを示す。各対立遺伝子について、上位に順位付けされたペプチドｎ個を陽性とコールし、ここで、ｎは、評価したセット内の確認された免疫原性エピトープの数である。図３１Ｄは、新抗原ペプチドについての例示的なｅｘ　ｖｉｖｏにおけるＴ細胞誘導の結果を示す。ペプチドを、ＨＬＡ－ＤＲＢ１＊１１：０１についての高ｎｅｏｎｍｈｃ２スコアおよび弱いＮｅｔＭＨＣＩＩｐａｎスコアに基づいて選択した。図は、出現する順に、それぞれ、配列番号８７～８９、９１、９０、２、９２～９４、３および９５～９６を開示する。図３１Ｅは、単一対立遺伝子ＭＡＰＴＡＣデータに対して訓練したモデルと、ホールドアウト単一対立遺伝子データに対して評価される逆畳み込みされた複対立遺伝子データとの比較を示す。値は、訓練用データセットをダウンサンプリングして逆畳み込み訓練セットのサイズをマッチさせた場合のｎｅｏｎｍｈｃ２について示されている通りである。図３１Ｆは、ＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１、逆畳み込みにより訓練された予測器、およびｎｅｏｎｍｈｃ２のＴＧＥＭデータセットに対するＰＰＶを示す（ダウンサンプリングを伴う、および伴わない）。各対立遺伝子について、上位に順位付けされたペプチドｎ個を陽性とコールし、ここで、ｎは、評価されたセットにおいて確認された免疫原性エピトープの数である。

【図３１Ｄ】図３１Ａ～３１Ｆは、ｎｅｏｎｍｈｃ２結合予測アルゴリズムの例示的なアーキテクチャおよびベンチマーキングを示す。図３１Ａは、単一対立遺伝子ＨＬＡクラスＩＩペプチドと、スクランブルされた長さを釣り合わせたデコイとを区別するように訓練された畳み込みニューラルネットワーク（ＣＮＮ）の例示的なアーキテクチャを示す。概略図は、アミノ酸特色の埋め込み層、幅６の２つの畳み込み層、最後までスキップ接続の存在、および最終的なロジスティック出力ノードへの入力として平均プーリング操作と最大プーリング操作の組合せの使用を示す。図３１Ｂは、訓練またはハイパーパラメータ最適化に使用していないＭＡＰＴＡＣ（商標）データの区分に関して評価した、ＮｅｔＭＨＣＩＩｐａｎおよびｎｅｏｎｍｈｃ２についての例示的な陽性適中率（ＰＰＶ）を示す。各対立遺伝子について、ＭＳにより観察されたペプチドｎ個を、供給源遺伝子の同じセットからサンプリングした長さを釣り合わせたデコイ１９ｎ個と併せてスコア化し、各予測器の上位に順位付けされたペプチド（例えば、上位５％）ｎ個を陽性とコールした。この評価プロトコールによると、偽陽性の数と偽陰性の数が必ず等しくなるので、ＰＰＶは再現率と同一である。図３１Ｃは、ＴＧＥＭデータセットに対する例示的なＮｅｔＭＨＣＩＩｐａｎのＰＰＶおよびｎｅｏｎｍｈｃ２のＰＰＶを示す。各対立遺伝子について、上位に順位付けされたペプチドｎ個を陽性とコールし、ここで、ｎは、評価したセット内の確認された免疫原性エピトープの数である。図３１Ｄは、新抗原ペプチドについての例示的なｅｘ　ｖｉｖｏにおけるＴ細胞誘導の結果を示す。ペプチドを、ＨＬＡ－ＤＲＢ１＊１１：０１についての高ｎｅｏｎｍｈｃ２スコアおよび弱いＮｅｔＭＨＣＩＩｐａｎスコアに基づいて選択した。図は、出現する順に、それぞれ、配列番号８７～８９、９１、９０、２、９２～９４、３および９５～９６を開示する。図３１Ｅは、単一対立遺伝子ＭＡＰＴＡＣデータに対して訓練したモデルと、ホールドアウト単一対立遺伝子データに対して評価される逆畳み込みされた複対立遺伝子データとの比較を示す。値は、訓練用データセットをダウンサンプリングして逆畳み込み訓練セットのサイズをマッチさせた場合のｎｅｏｎｍｈｃ２について示されている通りである。図３１Ｆは、ＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１、逆畳み込みにより訓練された予測器、およびｎｅｏｎｍｈｃ２のＴＧＥＭデータセットに対するＰＰＶを示す（ダウンサンプリングを伴う、および伴わない）。各対立遺伝子について、上位に順位付けされたペプチドｎ個を陽性とコールし、ここで、ｎは、評価されたセットにおいて確認された免疫原性エピトープの数である。

【図３１Ｅ－Ｆ】図３１Ａ～３１Ｆは、ｎｅｏｎｍｈｃ２結合予測アルゴリズムの例示的なアーキテクチャおよびベンチマーキングを示す。図３１Ａは、単一対立遺伝子ＨＬＡク

ラスＩＩペプチドと、スクランブルされた長さを釣り合わせたデコイとを区別するように訓練された畳み込みニューラルネットワーク（ＣＮＮ）の例示的なアーキテクチャを示す。概略図は、アミノ酸特色の埋め込み層、幅６の２つの畳み込み層、最後までスキップ接続の存在、および最終的なロジスティック出力ノードへの入力として平均プーリング操作と最大プーリング操作の組合せの使用を示す。図３１Ｂは、訓練またはハイパーパラメータ最適化に使用していないＭＡＰＴＡＣ（商標）データの区分に関して評価した、ＮｅｔＭＨＣＩＩｐａｎおよびｎｅｏｎｍｈｃ２についての例示的な陽性適中率（ＰＰＶ）を示す。各対立遺伝子について、ＭＳにより観察されたペプチドｎ個を、供給源遺伝子の同じセットからサンプリングした長さを釣り合わせたデコイ１９ｎ個と併せてスコア化し、各予測器の上位に順位付けされたペプチド（例えば、上位５％）ｎ個を陽性とコールした。この評価プロトコールによると、偽陽性の数と偽陰性の数が必ず等しくなるので、ＰＰＶは再現率と同一である。図３１Ｃは、ＴＧＥＭデータセットに対する例示的なＮｅｔＭＨＣＩＩｐａｎのＰＰＶおよびｎｅｏｎｍｈｃ２のＰＰＶを示す。各対立遺伝子について、上位に順位付けされたペプチドｎ個を陽性とコールし、ここで、ｎは、評価したセット内の確認された免疫原性エピトープの数である。図３１Ｄは、新抗原ペプチドについての例示的なｅｘ　ｖｉｖｏにおけるＴ細胞誘導の結果を示す。ペプチドを、ＨＬＡ－ＤＲＢ１＊１１：０１についての高ｎｅｏｎｍｈｃ２スコアおよび弱いＮｅｔＭＨＣＩＩｐａｎスコアに基づいて選択した。図は、出現する順に、それぞれ、配列番号８７～８９、９１、９０、２、９２～９４、３および９５～９６を開示する。図３１Ｅは、単一対立遺伝子ＭＡＰＴＡＣデータに対して訓練したモデルと、ホールドアウト単一対立遺伝子データに対して評価される逆畳み込みされた複対立遺伝子データとの比較を示す。値は、訓練用データセットをダウンサンプリングして逆畳み込み訓練セットのサイズをマッチさせた場合のｎｅｏｎｍｈｃ２について示されている通りである。図３１Ｆは、ＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１、逆畳み込みにより訓練された予測器、およびｎｅｏｎｍｈｃ２のＴＧＥＭデータセットに対するＰＰＶを示す（ダウンサンプリングを伴う、および伴わない）。各対立遺伝子について、上位に順位付けされたペプチドｎ個を陽性とコールし、ここで、ｎは、評価されたセットにおいて確認された免疫原性エピトープの数である。

【０３６９】

【図３２Ａ－Ｅ】図３２Ａ～３２Ｅは、ＨＬＡクラスＩＩ腫瘍ペプチドームにおける例示的な遺伝子表示およびタンパク質プロセシングを示す。図３２Ａは、結腸直腸がん、黒色腫、および卵巣がんデータセットの共同解析によって決定された遺伝子当たりのＨＬＡクラスＩＩペプチドの観察数と予測数の例示的な結果を示す。予測計数は、遺伝子の長さに発現レベルを掛けることによって導き出される。予測計数および観察計数を関連性のある試料にわたって合計した。血漿中に存在することが公知である遺伝子にそれらの濃度に従って印をつけた。図３２Ｂは、細胞内局在当たりのペプチドの予測された頻度と観察された頻度の例示的な結果を示す。図３２Ｃは、プロテアソームによって調節される遺伝子についての富化スコアの分布の例示的な結果を示す（観察された観察対予測された観察の比、一部図３２Ｂの通り）。遺伝子セットは、既知のユビキチン化部位を含むもの、およびプロテアソーム阻害剤を適用すると存在量が増加するものを含んだ。図３２Ｄは、ＨＬＡクラスＩＩペプチドが、ｉ）カテプシンおよび他の酵素によりタンパク質を破壊切断してペプチド断片にし、その後、それらにＨＬＡが結合する、ｉｉ）タンパク質またはアンフォールディングされたポリペプチドがＨＬＡに結合し、その後、切断されてペプチド長になる、ｉｉｉ）タンパク質が部分的に消化された後、結合し、結合後にさらにトリミングされる、のいずれに従ってプロセシングされるかに関する３つの例示的なワーキングモデルを示す図を示す。各モデルは異なる予測手法に対応する。図３２Ｅは、プロセシングに関連する変数およびｎｅｏｎｍｈｃ２結合予測を含むロジスティック回帰モデルについて観察されたＰＰＶの、結合予測のみを使用したモデルと比較した絶対的な上昇を示す。ＨＬＡ－ＤＲ抗体によってプロファイリングされた１１の試料に対して評価を行った（図３０Ｂにおける解析と同じ試料）；各点は１つの試料に対応する。アスタリスクは、両側の対応のあるｔ検定に従った有意な改善を示す（＊：ｐ＜０．０１、＊＊：ｐ＜０．００１、

＊＊＊：ｐ＜０．０００１）。同じ解析が図４０Ｂに示されているが、代わりにＮｅｔＭＨＣＩＩｐａｎを基本の予測器として使用する。デコイ選択およびＰＰＶ算出のための方法は図３１Ｂにおいて使用されているものと同一である。

【０３７０】

【図３３Ａ－Ｃ】図３３Ａ～３３Ｇは、樹状細胞によって提示される腫瘍抗原の同定および予測の例示的な結果を示す。図３３Ａは、がん細胞（Ｋ５６２）に由来するＤＣ提示ＨＬＡ－ＩＩリガンドを同定するための実験的なワークフローの例示的な図表示を示す。がん細胞を、完全な組み入れのためにＳＩＬＡＣ培地中で成長させ、溶解させたかまたは照射処理し、次いで、単球由来樹状細胞と一緒にプレーティングした。提示されたペプチドを汎ＤＲ抗体によって単離し、ＬＣ－ＭＳ／ＭＳによって配列決定した。図３３Ｂは、樹状細胞によって提示される腫瘍由来ペプチドについての、図２１Ａと同じヒット対デコイ比および性能メトリクスを使用した予測性能を表す例示的なデータを示す。プロセシング特色の使用を伴うまたは伴わない、ＮｅｔＭＨＣＩＩｐａｎに基づくモデルおよびｎｅｏｎｍｈｃ２に基づくモデルについての性能が示されている。図３３Ｃは、ＵＶ処理実験において観察された、重標識されたペプチドの供給源遺伝子についての例示的な遺伝子発現分布（赤色の曲線、Ｋ５６２発現に従ってプロットした）を軽標識されたペプチドの供給源遺伝子（灰色の曲線、ＤＣ発現に従ってプロットした）と比較して示す。図３３Ｄは、ＮｅｔＭＨＣＩＩｐａｎに基づくモデルおよびｎｅｏｎｍｈｃ２に基づくモデルを使用し、プロセシング特色を用いて、および用いずに、提示された腫瘍抗原を予測することについてのヒット対デコイ比１：４９９でのＰＰＶの例示的なグラフを示す。データ点は、左から右に、試料：ドナー１　ＨＯＣｌで処理した細胞：ＮｅｔＭＨＣＩＩｐａｎ連続した発現；ＮｅｔＭＨＣＩＩｐａｎ連続した発現＋遺伝子の偏り；ＮｅｔＭＨＣＩＩｐａｎ連続した発現＋遺伝子の偏り＋ＤＱ重複、完全プロセシング方式；ドナー１、ＵＶ処理：ｎｅｏｎｍｈｃ２；ｎｅｏｎｍｈｃ２＋閾値発現；ｎｅｏｎｍｈｃ２＋連続した発現；ｎｅｏｎｍｈｃ２＋連続した発現＋遺伝子の偏り；ｎｅｏｎｍｈｃ２＋連続した発現＋遺伝子の偏り＋ＤＱ重複を表す。図３３Ｅは、それぞれ重（Ｋ５６２由来）ペプチドおよび軽（ＤＣ由来）ペプチドの予測における種々の遺伝子の局在および機能的クラスの有意性を示す。Ｐ値を、ｎｅｏｎｍｈｃ２結合スコアおよび供給源遺伝子発現を制御するロジスティック回帰に従って算出する。棒の色は、回帰において係数に関連するサインを示す。図３３Ｆは、ＵＶ処理実験およびＨＯＣｌ処理実験における腫瘍細胞由来ペプチド供給源遺伝子の重複を示す結果の例示的な図表示を示す（機能的クラス毎に色付けされている）。図３３Ｇは、第１のドナーにおいて観察された重標識されたペプチドに当てはめたロジスティックモデルを使用して第２のドナーにおいて提示される腫瘍抗原を予測することについてのＰＰＶを示す例示的なデータを示す。モデルを、ｎｅｏｎｍｈｃ２結合単独で；結合および発現；または結合、発現、およびペプチドがミトコンドリア遺伝子に由来するものであるかどうかを示すバイナリ変数を使用して当てはめた。

【図３３Ｄ－Ｅ】図３３Ａ～３３Ｇは、樹状細胞によって提示される腫瘍抗原の同定および予測の例示的な結果を示す。図３３Ａは、がん細胞（Ｋ５６２）に由来するＤＣ提示ＨＬＡ－ＩＩリガンドを同定するための実験的なワークフローの例示的な図表示を示す。がん細胞を、完全な組み入れのためにＳＩＬＡＣ培地中で成長させ、溶解させたかまたは照射処理し、次いで、単球由来樹状細胞と一緒にプレーティングした。提示されたペプチドを汎ＤＲ抗体によって単離し、ＬＣ－ＭＳ／ＭＳによって配列決定した。図３３Ｂは、樹状細胞によって提示される腫瘍由来ペプチドについての、図２１Ａと同じヒット対デコイ比および性能メトリクスを使用した予測性能を表す例示的なデータを示す。プロセシング特色の使用を伴うまたは伴わない、ＮｅｔＭＨＣＩＩｐａｎに基づくモデルおよびｎｅｏｎｍｈｃ２に基づくモデルについての性能が示されている。図３３Ｃは、ＵＶ処理実験において観察された、重標識されたペプチドの供給源遺伝子についての例示的な遺伝子発現分布（赤色の曲線、Ｋ５６２発現に従ってプロットした）を軽標識されたペプチドの供給源遺伝子（灰色の曲線、ＤＣ発現に従ってプロットした）と比較して示す。図３３Ｄは、ＮｅｔＭＨＣＩＩｐａｎに基づくモデルおよびｎｅｏｎｍｈｃ２に基づくモデルを使用し

、プロセシング特色を用いて、および用いずに、提示された腫瘍抗原を予測することについてのヒット対デコイ比１：４９９でのＰＰＶの例示的なグラフを示す。データ点は、左から右に、試料：ドナー１　ＨＯＣｌで処理した細胞：ＮｅｔＭＨＣＩＩｐａｎ連続した発現；ＮｅｔＭＨＣＩＩｐａｎ連続した発現＋遺伝子の偏り；ＮｅｔＭＨＣＩＩｐａｎ連続した発現＋遺伝子の偏り＋ＤＱ重複、完全プロセシング方式；ドナー１、ＵＶ処理：ｎｅｏｎｍｈｃ２；ｎｅｏｎｍｈｃ２＋閾値発現；ｎｅｏｎｍｈｃ２＋連続した発現；ｎｅｏｎｍｈｃ２＋連続した発現＋遺伝子の偏り；ｎｅｏｎｍｈｃ２＋連続した発現＋遺伝子の偏り＋ＤＱ重複を表す。図３３Ｅは、それぞれ重（Ｋ５６２由来）ペプチドおよび軽（ＤＣ由来）ペプチドの予測における種々の遺伝子の局在および機能的クラスの有意性を示す。Ｐ値を、ｎｅｏｎｍｈｃ２結合スコアおよび供給源遺伝子発現を制御するロジスティック回帰に従って算出する。棒の色は、回帰において係数に関連するサインを示す。図３３Ｆは、ＵＶ処理実験およびＨＯＣｌ処理実験における腫瘍細胞由来ペプチド供給源遺伝子の重複を示す結果の例示的な図表示を示す（機能的クラス毎に色付けされている）。図３３Ｇは、第１のドナーにおいて観察された重標識されたペプチドに当てはめたロジスティックモデルを使用して第２のドナーにおいて提示される腫瘍抗原を予測することについてのＰＰＶを示す例示的なデータを示す。モデルを、ｎｅｏｎｍｈｃ２結合単独で；結合および発現；または結合、発現、およびペプチドがミトコンドリア遺伝子に由来するものであるかどうかを示すバイナリ変数を使用して当てはめた。

【図３３Ｆ－Ｇ】図３３Ａ～３３Ｇは、樹状細胞によって提示される腫瘍抗原の同定および予測の例示的な結果を示す。図３３Ａは、がん細胞（Ｋ５６２）に由来するＤＣ提示ＨＬＡ－ＩＩリガンドを同定するための実験的なワークフローの例示的な図表示を示す。がん細胞を、完全な組み入れのためにＳＩＬＡＣ培地中で成長させ、溶解させたかまたは照射処理し、次いで、単球由来樹状細胞と一緒にプレーティングした。提示されたペプチドを汎ＤＲ抗体によって単離し、ＬＣ－ＭＳ／ＭＳによって配列決定した。図３３Ｂは、樹状細胞によって提示される腫瘍由来ペプチドについての、図２１Ａと同じヒット対デコイ比および性能メトリクスを使用した予測性能を表す例示的なデータを示す。プロセシング特色の使用を伴うまたは伴わない、ＮｅｔＭＨＣＩＩｐａｎに基づくモデルおよびｎｅｏｎｍｈｃ２に基づくモデルについての性能が示されている。図３３Ｃは、ＵＶ処理実験において観察された、重標識されたペプチドの供給源遺伝子についての例示的な遺伝子発現分布（赤色の曲線、Ｋ５６２発現に従ってプロットした）を軽標識されたペプチドの供給源遺伝子（灰色の曲線、ＤＣ発現に従ってプロットした）と比較して示す。図３３Ｄは、ＮｅｔＭＨＣＩＩｐａｎに基づくモデルおよびｎｅｏｎｍｈｃ２に基づくモデルを使用し、プロセシング特色を用いて、および用いずに、提示された腫瘍抗原を予測することについてのヒット対デコイ比１：４９９でのＰＰＶの例示的なグラフを示す。データ点は、左から右に、試料：ドナー１　ＨＯＣｌで処理した細胞：ＮｅｔＭＨＣＩＩｐａｎ連続した発現；ＮｅｔＭＨＣＩＩｐａｎ連続した発現＋遺伝子の偏り；ＮｅｔＭＨＣＩＩｐａｎ連続した発現＋遺伝子の偏り＋ＤＱ重複、完全プロセシング方式；ドナー１、ＵＶ処理：ｎｅｏｎｍｈｃ２；ｎｅｏｎｍｈｃ２＋閾値発現；ｎｅｏｎｍｈｃ２＋連続した発現；ｎｅｏｎｍｈｃ２＋連続した発現＋遺伝子の偏り；ｎｅｏｎｍｈｃ２＋連続した発現＋遺伝子の偏り＋ＤＱ重複を表す。図３３Ｅは、それぞれ重（Ｋ５６２由来）ペプチドおよび軽（ＤＣ由来）ペプチドの予測における種々の遺伝子の局在および機能的クラスの有意性を示す。Ｐ値を、ｎｅｏｎｍｈｃ２結合スコアおよび供給源遺伝子発現を制御するロジスティック回帰に従って算出する。棒の色は、回帰において係数に関連するサインを示す。図３３Ｆは、ＵＶ処理実験およびＨＯＣｌ処理実験における腫瘍細胞由来ペプチド供給源遺伝子の重複を示す結果の例示的な図表示を示す（機能的クラス毎に色付けされている）。図３３Ｇは、第１のドナーにおいて観察された重標識されたペプチドに当てはめたロジスティックモデルを使用して第２のドナーにおいて提示される腫瘍抗原を予測することについてのＰＰＶを示す例示的なデータを示す。モデルを、ｎｅｏｎｍｈｃ２結合単独で；結合および発現；または結合、発現、およびペプチドがミトコンドリア遺伝子に由来するものであるかどうかを示すバイナリ変数を使用して当てはめた。

【０３７１】
【図３４Ａ－Ｂ】図３４Ａ～３４Ｂは、図２９に関連するＭＡＰＴＡＣ（商標）データの例示的な特徴付けを示す。図３４Ａは、親和性タグ付けされたＨＬＡ－Ａ＊０２：０１－ＢＡＰをコードするＭＡＰＴＡＣ（商標）構築物をトランスフェクトしたＥｘｐｉ２９３細胞株のＦＡＣＳによる例示的なＨＬＡ細胞表面解析を示す。図３４Ｂは、親和性タグ付けされたＨＬＡ－ＤＲＢ１＊１１：０１－ＢＡＰをコードするＭＡＰＴＡＣ（商標）構築物をトランスフェクトしたＥｘｐｉ２９３細胞株のＦＡＣＳによる例示的なＨＬＡ細胞表面解析を示す（下）。トランスフェクトされたＥｘｐｉ２９３細胞（オレンジ色）のＨＬＡ細胞表面発現を染色されたトランスフェクトされていないＥｘｐｉ２９３（青色）、無染色のトランスフェクトされていないＥｘｐｉ２９３（赤色）、染色したＰＢＭＣ（暗い緑色）、および無染色のＰＢＭＣ（明るい緑色）と比較した。全てのＨＬＡクラスＩ染色にＷ６／３２（汎ＨＬＡクラスＩ）を利用し、一方、ＨＬＡクラスＩＩ染色にはＲＥＡ３３２（汎ＨＬＡクラスＩＩ）を利用した。
【０３７２】
【図３５】図３５は、図３０Ａに関連する、ＭＡＰＴＡＣ（商標）とＩＥＤＢロゴの例示的な比較を示す。良好なＮｅｔＭＨＣＩＩｐａｎスコアを示さなかったが、ＭＳではよく支持された（スコア化ピーク強度＞７０および入れ子セットサイズ≧１）、ＭＳにより観察されたペプチドに対する測定された親和性およびＮｅｔＭＨＣＩＩｐａｎで予測された親和性。
【０３７３】
【図３６Ａ】図３６Ａ～３６Ｃは、図３０Ａ～３０Ｃに関連する、ＨＬＡ－ＤＲ１　ＭＡＰＴＡＣ（商標）データ忠実度の例示的な解析を示す。図３６Ａは、一般的な対立遺伝子についての、ＨＬＡ－ＤＲ１　ＭＡＰＴＡＣ（商標）ペプチド（緑色）（長さ１２～２３）の、ランダムにプロテオームからサンプリングした５０，０００の長さを釣り合わせたデコイペプチド（青色）と比較した、例示的なＮｅｔＭＨＣＩＩｐａｎ３．１スコアを示す。図３６Ｂは、良好なＮｅｔＭＨＣＩＩｐａｎスコアを示さないが、ＭＳではよく支持される（スコア化ピーク強度＞７０および入れ子セットサイズ≧１）例示的なＭＳにより観察されたペプチドに対する例示的な測定された親和性およびＮｅｔＭＨＣＩＩｐａｎにより予測された親和性を示す。図は、出現する順に、上から下、左から右に、それぞれ、配列番号４０～８６を開示する。図３６Ｃは、異なる細胞型においてＭＡＰＴＡＣ（商標）によって決定されるＨＬＡ－ＤＲＢ１対立遺伝子についての例示的なＨＬＡクラスＩＩ配列ロゴを示す。
【図３６Ｂ】図３６Ａ～３６Ｃは、図３０Ａ～３０Ｃに関連する、ＨＬＡ－ＤＲ１　ＭＡＰＴＡＣ（商標）データ忠実度の例示的な解析を示す。図３６Ａは、一般的な対立遺伝子についての、ＨＬＡ－ＤＲ１　ＭＡＰＴＡＣ（商標）ペプチド（緑色）（長さ１２～２３）の、ランダムにプロテオームからサンプリングした５０，０００の長さを釣り合わせたデコイペプチド（青色）と比較した、例示的なＮｅｔＭＨＣＩＩｐａｎ３．１スコアを示す。図３６Ｂは、良好なＮｅｔＭＨＣＩＩｐａｎスコアを示さないが、ＭＳではよく支持される（スコア化ピーク強度＞７０および入れ子セットサイズ≧１）例示的なＭＳにより観察されたペプチドに対する例示的な測定された親和性およびＮｅｔＭＨＣＩＩｐａｎにより予測された親和性を示す。図は、出現する順に、上から下、左から右に、それぞれ、配列番号４０～８６を開示する。図３６Ｃは、異なる細胞型においてＭＡＰＴＡＣ（商標）によって決定されるＨＬＡ－ＤＲＢ１対立遺伝子についての例示的なＨＬＡクラスＩＩ配列ロゴを示す。
【図３６Ｃ】図３６Ａ～３６Ｃは、図３０Ａ～３０Ｃに関連する、ＨＬＡ－ＤＲ１　ＭＡＰＴＡＣ（商標）データ忠実度の例示的な解析を示す。図３６Ａは、一般的な対立遺伝子についての、ＨＬＡ－ＤＲ１　ＭＡＰＴＡＣ（商標）ペプチド（緑色）（長さ１２～２３）の、ランダムにプロテオームからサンプリングした５０，０００の長さを釣り合わせたデコイペプチド（青色）と比較した、例示的なＮｅｔＭＨＣＩＩｐａｎ３．１スコアを示す。図３６Ｂは、良好なＮｅｔＭＨＣＩＩｐａｎスコアを示さないが、ＭＳではよく支持

される（スコア化ピーク強度＞７０および入れ子セットサイズ≧１）例示的なＭＳにより観察されたペプチドに対する例示的な測定された親和性およびＮｅｔＭＨＣＩＩｐａｎにより予測された親和性を示す。図は、出現する順に、上から下、左から右に、それぞれ、配列番号４０～８６を開示する。図３６Ｃは、異なる細胞型においてＭＡＰＴＡＣ（商標）によって決定されるＨＬＡ－ＤＲＢ１対立遺伝子についての例示的なＨＬＡクラスＩＩ配列ロゴを示す。

【０３７４】

【図３７Ａ－Ｂ】図３７Ａ～３７Ｃは、図３０Ａ～３０Ｃに関連する、ＭＡＰＴＡＣ（商標）モチーフの追加的な例示的な解析を示す。図３７Ａは、ＨＬＡ－ＤＭ同時トランスフェクション（ｅｘｐｉ２９３細胞株）を伴う、および伴わない実験についてのＭＡＰＴＡＣ（商標）由来の配列ロゴを示す。図３７Ｂは、ＭＡＰＴＡＣ（商標）およびＩＥＤＢに従ったいくつかのＨＬＡクラスＩ対立遺伝子についての配列ロゴを示す。Ａ＊３２：０１はＰ２において高頻度Ｑを示さず、Ｃ＊０３：０３はＰ９において高頻度Ｙを示さず、複対立遺伝子逆畳み込みを使用した以前の試験とは異なることに留意されたい；Ｂ＊５２：０１のロゴについては以前に発表されていない。図３７Ｃは、ＭＡＰＡＣ（商標）により観察されたペプチドのＣＤ７４の遺伝子配列に対する例示的なアラインメントを示す。

【図３７Ｃ】図３７Ａ～３７Ｃは、図３０Ａ～３０Ｃに関連する、ＭＡＰＴＡＣ（商標）モチーフの追加的な例示的な解析を示す。図３７Ａは、ＨＬＡ－ＤＭ同時トランスフェクション（ｅｘｐｉ２９３細胞株）を伴う、および伴わない実験についてのＭＡＰＴＡＣ（商標）由来の配列ロゴを示す。図３７Ｂは、ＭＡＰＴＡＣ（商標）およびＩＥＤＢに従ったいくつかのＨＬＡクラスＩ対立遺伝子についての配列ロゴを示す。Ａ＊３２：０１はＰ２において高頻度Ｑを示さず、Ｃ＊０３：０３はＰ９において高頻度Ｙを示さず、複対立遺伝子逆畳み込みを使用した以前の試験とは異なることに留意されたい；Ｂ＊５２：０１のロゴについては以前に発表されていない。図３７Ｃは、ＭＡＰＡＣ（商標）により観察されたペプチドのＣＤ７４の遺伝子配列に対する例示的なアラインメントを示す。

【０３７５】

【図３８Ａｉ】図３８Ａｉ～３８Ｄは、図３１Ａ～３１Ｄに関連する例示的なｎｅｏｎｍｈｃ２性能統計値およびＴ細胞フロー染色を示す。図３８Ａｉは、例示的な訓練用データセットのサイズに応じたｎｅｏｎｍｈｃ２の性能を示す。ＰＰＶを、図３１Ｂにおけるものと同じ様式で、同じ評価ペプチドを使用して評価した；しかし、より小さな訓練用データセットを模倣するために訓練用データをランダムにダウンサンプリングした。図３８Ａｉｉは、ＧｉｂｂｓＣｌｕｓｔｅｒ（デフォルト設定；「トラッシュクラスター許容）を使用した複対立遺伝子ＨＬＡ－ＤＲリガンドームに由来するペプチドクラスターについての例示的な配列ロゴを示す。図３８Ｂは、ｎｅｏｎｍｈｃ２を用いた新抗原ペプチド予測で再コールされた誘導試料由来のＣＤ４＋細胞によるＩＦＮ－γ発現の例示的な代表的なフローサイトメトリープロットを示す。デルタ値を、新抗原を伴って再コールされた場合のＩＦＮ－γを発現するＣＤ４＋細胞のパーセント（＋ペプチド）を、新抗原の非存在下で再コールされたＩＦＮ－γを発現するＣＤ４＋のパーセント（ペプチドなし）から差し引くことによって算出した。左の２つのフロープロットは、ＣＤ４＋　Ｔ細胞によるＴ細胞応答を誘導した新抗原（ＰＥＡＳＬＹＧＡＬＳＫＧＳＧＧ（配列番号２））およびＴ細胞応答を誘導しなかった新抗原（ＰＡＴＹＩＬＩＬＫＥＦＣＬＶＧ（配列番号３））を表す。図３８Ｃは、単一のｎｅｏｎｍｈｃ２新抗原ペプチドを伴って再コールされたウェルからの例示的なデルタ値を示す。ペプチドを、陽性応答（３％を超えるデルタ応答、強調表示されている）を有する場合には誘導ヒットとみなした。図は、出現する順に、それぞれ、配列番号８７～９１、２、９２～９４、３、および９５～９６を開示する。図３８Ｄは、ＧｉｂｂｓＣｌｕｓｔｅｒ（デフォルト設定；「トラッシュ」クラスター許容）を使用して複対立遺伝子ＨＬＡ－ＤＲリガンドームについて導き出されたペプチドクラスターについての例示的な配列ロゴを示す。

【図３８Ａｉｉ】図３８Ａｉ～３８Ｄは、図３１Ａ～３１Ｄに関連する例示的なｎｅｏｎｍｈｃ２性能統計値およびＴ細胞フロー染色を示す。図３８Ａｉは、例示的な訓練用デー

タセットのサイズに応じたｎｅｏｎｍｈｃ２の性能を示す。ＰＰＶを、図３１Ｂにおけるものと同じ様式で、同じ評価ペプチドを使用して評価した；しかし、より小さな訓練用データセットを模倣するために訓練用データをランダムにダウンサンプリングした。図３８Ａｉｉは、ＧｉｂｂｓＣｌｕｓｔｅｒ（デフォルト設定；「トラッシュクラスター許容）を使用した複対立遺伝子ＨＬＡ－ＤＲリガンドームに由来するペプチドクラスターについての例示的な配列ロゴを示す。図３８Ｂは、ｎｅｏｎｍｈｃ２を用いた新抗原ペプチド予測で再コールされた誘導試料由来のＣＤ４＋細胞によるＩＦＮ－γ発現の例示的な代表的なフローサイトメトリープロットを示す。デルタ値を、新抗原を伴って再コールされた場合のＩＦＮ－γを発現するＣＤ４＋細胞のパーセント（＋ペプチド）を、新抗原の非存在下で再コールされたＩＦＮ－γを発現するＣＤ４＋のパーセント（ペプチドなし）から差し引くことによって算出した。左の２つのフロープロットは、ＣＤ４＋　Ｔ細胞によるＴ細胞応答を誘導した新抗原（ＰＥＡＳＬＹＧＡＬＳＫＧＳＧＧ（配列番号２））およびＴ細胞応答を誘導しなかった新抗原（ＰＡＴＹＩＬＩＬＫＥＦＣＬＶＧ（配列番号３））を表す。図３８Ｃは、単一のｎｅｏｎｍｈｃ２新抗原ペプチドを伴って再コールされたウェルからの例示的なデルタ値を示す。ペプチドを、陽性応答（３％を超えるデルタ応答、強調表示されている）を有する場合には誘導ヒットとみなした。図は、出現する順に、それぞれ、配列番号８７～９１、２、９２～９４、３、および９５～９６を開示する。図３８Ｄは、ＧｉｂｂｓＣｌｕｓｔｅｒ（デフォルト設定；「トラッシュ」クラスター許容）を使用して複対立遺伝子ＨＬＡ－ＤＲリガンドームについて導き出されたペプチドクラスターについての例示的な配列ロゴを示す。

【図３８Ｂ－Ｃ】図３８Ａｉ～３８Ｄは、図３１Ａ～３１Ｄに関連する例示的なｎｅｏｎｍｈｃ２性能統計値およびＴ細胞フロー染色を示す。図３８Ａｉは、例示的な訓練用データセットのサイズに応じたｎｅｏｎｍｈｃ２の性能を示す。ＰＰＶを、図３１Ｂにおけるものと同じ様式で、同じ評価ペプチドを使用して評価した；しかし、より小さな訓練用データセットを模倣するために訓練用データをランダムにダウンサンプリングした。図３８Ａｉｉは、ＧｉｂｂｓＣｌｕｓｔｅｒ（デフォルト設定；「トラッシュクラスター許容）を使用した複対立遺伝子ＨＬＡ－ＤＲリガンドームに由来するペプチドクラスターについての例示的な配列ロゴを示す。図３８Ｂは、ｎｅｏｎｍｈｃ２を用いた新抗原ペプチド予測で再コールされた誘導試料由来のＣＤ４＋細胞によるＩＦＮ－γ発現の例示的な代表的なフローサイトメトリープロットを示す。デルタ値を、新抗原を伴って再コールされた場合のＩＦＮ－γを発現するＣＤ４＋細胞のパーセント（＋ペプチド）を、新抗原の非存在下で再コールされたＩＦＮ－γを発現するＣＤ４＋のパーセント（ペプチドなし）から差し引くことによって算出した。左の２つのフロープロットは、ＣＤ４＋　Ｔ細胞によるＴ細胞応答を誘導した新抗原（ＰＥＡＳＬＹＧＡＬＳＫＧＳＧＧ（配列番号２））およびＴ細胞応答を誘導しなかった新抗原（ＰＡＴＹＩＬＩＬＫＥＦＣＬＶＧ（配列番号３））を表す。図３８Ｃは、単一のｎｅｏｎｍｈｃ２新抗原ペプチドを伴って再コールされたウェルからの例示的なデルタ値を示す。ペプチドを、陽性応答（３％を超えるデルタ応答、強調表示されている）を有する場合には誘導ヒットとみなした。図は、出現する順に、それぞれ、配列番号８７～９１、２、９２～９４、３、および９５～９６を開示する。図３８Ｄは、ＧｉｂｂｓＣｌｕｓｔｅｒ（デフォルト設定；「トラッシュ」クラスター許容）を使用して複対立遺伝子ＨＬＡ－ＤＲリガンドームについて導き出されたペプチドクラスターについての例示的な配列ロゴを示す。

【図３８Ｄ】図３８Ａｉ～３８Ｄは、図３１Ａ～３１Ｄに関連する例示的なｎｅｏｎｍｈｃ２性能統計値およびＴ細胞フロー染色を示す。図３８Ａｉは、例示的な訓練用データセットのサイズに応じたｎｅｏｎｍｈｃ２の性能を示す。ＰＰＶを、図３１Ｂにおけるものと同じ様式で、同じ評価ペプチドを使用して評価した；しかし、より小さな訓練用データセットを模倣するために訓練用データをランダムにダウンサンプリングした。図３８Ａｉｉは、ＧｉｂｂｓＣｌｕｓｔｅｒ（デフォルト設定；「トラッシュクラスター許容）を使用した複対立遺伝子ＨＬＡ－ＤＲリガンドームに由来するペプチドクラスターについての例示的な配列ロゴを示す。図３８Ｂは、ｎｅｏｎｍｈｃ２を用いた新抗原ペプチド予測で

再コールされた誘導試料由来のＣＤ４＋細胞によるＩＦＮ－γ発現の例示的な代表的なフローサイトメトリープロットを示す。デルタ値を、新抗原を伴って再コールされた場合のＩＦＮ－γを発現するＣＤ４＋細胞のパーセント（＋ペプチド）を、新抗原の非存在下で再コールされたＩＦＮ－γを発現するＣＤ４＋のパーセント（ペプチドなし）から差し引くことによって算出した。左の２つのフロープロットは、ＣＤ４＋　Ｔ細胞によるＴ細胞応答を誘導した新抗原（ＰＥＡＳＬＹＧＡＬＳＫＧＳＧＧ（配列番号２））およびＴ細胞応答を誘導しなかった新抗原（ＰＡＴＹＩＬＩＬＫＥＦＣＬＶＧ（配列番号３））を表す。図３８Ｃは、単一のｎｅｏｎｍｈｃ２新抗原ペプチドを伴って再コールされたウェルからの例示的なデルタ値を示す。ペプチドを、陽性応答（３％を超えるデルタ応答、強調表示されている）を有する場合には誘導ヒットとみなした。図は、出現する順に、それぞれ、配列番号８７～９１、２、９２～９４、３、および９５～９６を開示する。図３８Ｄは、ＧｉｂｂｓＣｌｕｓｔｅｒ（デフォルト設定；「トラッシュ」クラスター許容）を使用して複対立遺伝子ＨＬＡ－ＤＲリガンドームについて導き出されたペプチドクラスターについての例示的な配列ロゴを示す。

【０３７６】

【図３９Ａ－Ｃ】図３９Ａ～３９Ｃは、図３２Ａ～３２Ｅに関連する、ＨＬＡクラスＩＩについての追加的な例示的な細胞起源解析を示す。図３９Ａは、汎ＤＲ抗体によってプロファイリングされた４つのＰＢＭＣ試料（図３０ＢのＲＧ１２４８、ＲＧ１１０４、ＲＧ１０９５、およびＨＤＳＣ）において観察されたＨＬＡクラスＩＩペプチドについての、ペプチド供給源遺伝子がヒト血漿中に存在するかどうかに応じた例示的なパーセント順位ｎｅｏｎｍｈｃ２スコアを示す。各ペプチドについて、ドナーに存在する対立遺伝子にわたって最良（最低）のパーセント順位を使用した。ランダムな長さを釣り合わせたプロテオームデコイについてのスコアが比較のために示されている。箱型図は５パーセンタイル、２５パーセンタイル、５０パーセンタイル、７５パーセンタイル、および９５パーセンタイルを示す。図３９Ｂは、図３２Ａにおけるものと同じ方法体系を使用してＨＬＡクラスＩについて遺伝子毎の観察されたペプチドと予測されたペプチドの例示的な計数を示す。データは同じ腫瘍型（結腸直腸、卵巣、および黒色腫）に対応する。ヒト血漿中に存在する遺伝子が青色で強調表示されており、それらの濃度に応じたサイズになっている。図３９Ｃは、２つの異なる遺伝子発現プロファイルに関する例示的なペプチド観察の相対的一致を示す。各試料について、遺伝子レベルペプチド計数をバルク腫瘍遺伝子発現およびプロフェッショナルＡＰＣ遺伝子発現プロファイルの線形結合としてモデル化した。係数の比により、各発現プロファイルのペプチドレパートリーとの相対的一致が決定される。エラーバーはブートストラップリサンプリングによってコンピュータ計算された９５％信頼区間に対応する。

【０３７７】

【図４０Ａ－Ｂ】図４０Ａ～４０Ｂは、図３２Ａ～３２Ｅに関連するプロセシングモチーフの追加的な例示的な解析を示す。図４０Ａは、ドナーＰＢＭＣ、単球由来樹状細胞、結腸直腸がん、黒色腫、卵巣がん、およびｅｘｐｉ２９３細胞株（大多数のＭＡＰＴＡＣ（商標）データ生成に使用される）において観察される、平均プロテオーム頻度（上流位置Ｕ３～Ｕ１および下流位置Ｄ１～Ｄ３を適用する）に対するまたは平均ペプチド頻度（内部の位置Ｎ１～Ｃ１を適用する）に対するＮ末端およびＣ末端ペプチドカット部分付近の例示的なアミノ酸頻度を示す。図４０Ｂは、図３２Ｅと同じであるが、ＮｅｔＭＨＣＩＩｐａｎを基本の予測器として使用した解析を示す。ＨＬＡ－ＤＲ抗体によってプロファイリングされた８つの試料（図３１Ｂにおいて解析したものと同じ試料）について、ＮｅｔＭＨＣＩＩｐａｎによる予測に加えてプロセシングに関連する変数を含めたロジスティック回帰モデルに関してはＰＰＶの絶対的な上昇が観察された（ＮｅｔＭＨＣＩＩｐａｎのみのモデルと比較して）。アスタリスクは有意な改善を示す（＊：ｐ＜０．０１、＊＊：ｐ＜０．００１、＊＊＊：ｐ＜０．０００１）両側の対応のあるｔ検定に従った。

【０３７８】

【図４１】図４１は、ペプチドの上流、ペプチド内部、およびペプチドの下流の位置を指

すために使用した例示的な命名システムを示す。
【０３７９】
【図４２Ａ】図４２Ａは、ｎＬＣ－ＭＳ／ＭＳによる、内因性にプロセシングされ、ＨＬＡ－１およびＨＬＡクラスＩＩによって提示されるペプチドの解析の例示的なワークフローを表す図を示す。
【０３８０】
【図４２Ｂ】図４２Ｂは、トリプシンペプチドの、ＦＡＩＭＳを伴うまたは伴わないｎＬＣ－ＭＳ／ＭＳ分析からの例示的な実験結果を示すグラフを示す。示されている分析規模でのＦＡＩＭＳを伴うまたは伴わないｎＬＣ－ＭＳ／ＭＳ分析によるＨＬＡ－１およびＨＬＡクラスＩＩペプチドの検出における代表的な重複も示されている。
【０３８１】
【図４３Ａ】図４３Ａは、ＦＡＩＭＳを伴うまたは伴わない例示的なＨＬＡクラスＩ酸性および塩基性逆相分画ペプチド検出を示す。
【０３８２】
【図４３Ｂ】図４３Ｂは、保持時間に対してプロットした、ＨＬＡクラスＩと結合した独特のペプチドの検出を示す例示的な実験結果を示す。
【０３８３】
【図４４Ａ】図４４Ａは、ＦＡＩＭＳを伴うまたは伴わない例示的なＨＬＡクラスＩＩ酸および塩基性逆相分画ペプチド検出を示す。
【０３８４】
【図４４Ｂ】図４４Ｂは、保持時間に対してプロットした、ＨＬＡクラスＩＩと結合した独特のペプチドの検出を示す例示的な実験結果を示す。
【０３８５】
【図４５－１】図４５は、示されている方法を使用して検出したＨＬＡクラスＩ結合性ペプチドの交差サイズの例示的なグラフ（左）ならびに、ＨＬＡクラスＩ結合性ペプチドのＬＣ－ＭＳ／ＭＳによる検出のための例示的な標準のワークフローおよび最適化されたワークフローのベン図を示す（右）。
【図４５－２】図４５は、示されている方法を使用して検出したＨＬＡクラスＩ結合性ペプチドの交差サイズの例示的なグラフ（左）ならびに、ＨＬＡクラスＩ結合性ペプチドのＬＣ－ＭＳ／ＭＳによる検出のための例示的な標準のワークフローおよび最適化されたワークフローのベン図を示す（右）。
【０３８６】
【図４６－１】図４６は、示されている方法を使用して検出したＨＬＡクラスＩＩ結合性ペプチドの交差サイズの例示的なグラフ（左）、ならびにＨＬＡクラスＩＩ結合性ペプチドのＬＣ－ＭＳ／ＭＳによる検出の例示的な標準のワークフローおよび最適化されたワークフローのベン図を示す（右）。
【図４６－２】図４６は、示されている方法を使用して検出したＨＬＡクラスＩＩ結合性ペプチドの交差サイズの例示的なグラフ（左）、ならびにＨＬＡクラスＩＩ結合性ペプチドのＬＣ－ＭＳ／ＭＳによる検出の例示的な標準のワークフローおよび最適化されたワークフローのベン図を示す（右）。
【発明を実施するための形態】
【０３８７】
詳細な説明
用語は全て、当業者によって理解される通り理解されることが意図されている。別段の定義のない限り、本明細書において使用される全ての科学技術用語は、本開示が関係する分野の当業者に一般に理解されるものと同じ意味を有する。
【０３８８】
本明細書で使用される節の見出しは、単に構成目的のものであり、記載の主題を限定するものとは解釈されない。
【０３８９】

本開示の種々の特色は単一の実施形態に関して記載され得るが、特色を別々にまたは任意の適切な組合せで提供することもできる。逆に、本開示は本明細書では明確にするために別々の実施形態に関して記載されている場合があるが、本開示を単一の実施形態で実施することもできる。

【０３９０】

本開示は、改善された免疫治療のためのＨＬＡクラスＩＩ特異的ペプチドの使用を可能にする新しいコンピュータに基づく機械学習ＨＬＡ－ペプチド提示予測モデルを使用して、特定のＨＬＡクラスＩＩアルファ鎖およびベータ鎖対による抗原、具体的にはがん抗原の提示を高度な信頼度で予測することができるという重要な所見に基づく。

【０３９１】

一態様では、本開示は、ペプチドのＨＬＡクラスＩＩタンパク質（アルファ鎖とベータ鎖のヘテロ二量体で構成される）への高忠実度の結合により、特定のペプチドがＴリンパ球に提示され、それにより、特異的な免疫応答が引き出され、あらゆる交差反応性または無差別免疫が回避されることが確実になるような、特定のＨＬＡクラスＩＩアルファ鎖およびベータ鎖ヘテロ二量体と正確に対形成または結合し得るペプチドを予測するための方法を提供する。いくつかの最近の試験により、ＣＤ４＋　Ｔ細胞がＨＬＡクラスＩＩにより提示されたリガンドを認識し、腫瘍の制御に寄与することができることも示されている。がんワクチンおよび他の免疫治療はＣＤ４＋　Ｔ細胞応答を方向付けることを利用することが理想的であるが、現行の試みでは、現行の予測ツールの正確度が不十分なことから、ＨＬＡクラスＩＩ抗原予測が完全に無視されている。

【０３９２】

一態様では、本開示は、特定の同類ＨＬＡクラスＩＩタンパク質を発現する対象にペプチドを治療的に投与した場合に、ＣＤ４＋　Ｔ細胞を活性化し、免疫記憶を刺激するＨＬＡクラスＩＩタンパク質の能力によって、ペプチドを用いてより持続的かつ頑強な免疫応答を活性化することができるような、特定のＨＬＡクラスＩＩタンパク質に正確に結合し得るペプチドを予測するための方法を提供する。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して改善が示される。一部の実施形態では、本発明で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約１．１倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約２倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約３倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約４倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約５倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約６倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約７倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約８倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約９倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約１０倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約１５倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩ

タンパク質予測に関して、現在利用可能な予測器に対して少なくとも約２０倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約３０倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約４０倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約５０倍の改善が示される。一部の実施形態では、本明細書で提供される方法では、特定のＨＬＡクラスＩＩタンパク質予測に関して、現在利用可能な予測器に対して少なくとも約６０倍の改善が示される。

【０３９３】

一態様では、特定の対象に対して調整または個別化された免疫治療の方法が本明細書で提示される。あらゆる対象または患者がＨＬＡクラスＩおよびＨＬＡクラスＩＩタンパク質の特定のアレイを発現する。ＨＬＡタイピングは、対象によって発現されるＨＬＡタンパク質の特定のレパートリーを決定することを可能にする周知の技法である。特定の対象によって発現されるＨＬＡヘテロ二量体が分かったら、特定のＨＬＡクラスＩＩアルファ鎖およびベータ鎖ヘテロ二量体に高忠実度で結合することができるペプチドを予測するための、本明細書に記載の、改善され洗練された信頼できる方法により、対象に対して特異的に調整された特異的な免疫応答を生じさせることができることを確実にする。

【０３９４】

本出願では、他に特に指定がなければ、単数形の使用は複数形を含む。本明細書において使用される場合、単数形「１つの（ａ）」、「１つの（ａｎ）」および「その（ｔｈｅ）」は、文脈により明確に別段の規定がなされない限り、複数の指示対象を包含することに留意しなければならない。本出願では、「または（ｏｒ）」の使用は、別段の指定のない限り、「および／または（ａｎｄ／ｏｒ）」を意味する。さらに、「含む（ｉｎｃｌｕｄｉｎｇ）」という用語ならび「含む（ｉｎｃｌｕｄｅ）」、「含む（ｉｎｃｌｕｄｅｓ）」、および「含まれる（ｉｎｃｌｕｄｅｄ）」などの他の形態の使用は限定されるものではない。「１つまたは複数の（ｏｎｅ　ｏｒ　ｍｏｒｅ）」または「少なくとも１つの（ａｔ　ｌｅａｓｔ　ｏｎｅ）」という用語、例えば、メンバーの群のうちの１つまたは複数の（ｏｎｅ　ｏｒ　ｍｏｒｅ）または少なくとも１つの（ａｔ　ｌｅａｓｔ　ｏｎｅ）メンバー（複数可）などは、それ自体は明らかであり、さらなる例証により、この用語は、とりわけ、前記メンバーのうちのいずれか１つへの言及、または前記メンバーのうちの任意の２つまたはそれよりも多く、例えば、前記メンバーのうちのいずれか≧３、≧４、≧５、≧６または≧７など、および最大前記メンバー全てなどへの言及を包含する。

【０３９５】

本明細書における「一部の実施形態」、「ある実施形態」、「一実施形態」または「他の実施形態」への言及は、実施形態に関連して記載されている特色、構造、または特徴が本開示の少なくとも一部の実施形態に含まれるが、必ずしも全ての実施形態に含まれるわけではない。

【０３９６】

本明細書および請求項（複数可）において使用される場合、「含む（ｃｏｍｐｒｉｓｉｎｇ）」（ならびに「含む（ｃｏｍｐｒｉｓｅ）」および「含む（ｃｏｍｐｒｉｓｅｓ）」などの、含む（ｃｏｍｐｒｉｓｉｎｇ）の任意の形態）、「有する（ｈａｖｉｎｇ）」（ならびに「有する（ｈａｖｅ）」および「有する（ｈａｓ）」などの、有する（ｈａｖｉｎｇ）の任意の形態）、「含む（ｉｎｃｌｕｄｉｎｇ）」（ならびに「含む（ｉｎｃｌｕｄｅｓ）」および「含む（ｉｎｃｌｕｄｅ）」などの、含む（ｉｎｃｌｕｄｉｎｇ）の任意の形態）または「含有する（ｃｏｎｔａｉｎｉｎｇ）」（ならびに、「含有する（ｃｏｎｔａｉｎｓ）」および「含有する（ｃｏｎｔａｉｎ）」などの、含有する（ｃｏｎｔａｉｎｉｎｇ）の任意の形態）という単語は、包括的またはオープンエンドであり、追加的な、列挙されていない要素または方法のステップは排除されない。本明細書において考

察されている任意の実施形態を本開示の任意の方法または組成物に関して実施することができ、逆もまた同じであることが意図されている。さらに、本開示の組成物を使用して本開示の方法を実現することができる。

【０３９７】

「約」または「およそ」という用語は、パラメータ、量、持続時間などの測定可能な値について言及する際に本明細書で使用される場合、指定の値から＋／－２０％またはそれ未満、＋／－１０％またはそれ未満、＋／－５％またはそれ未満、または＋／－１％またはそれ未満の変動を、そのような変動が本開示の実施に適する限りは包含するものとする。「約」または「およそ」という修飾語が指す値は、それ自体も具体的に開示されることが理解されるべきである。

【０３９８】

「免疫応答」という用語は、Ｔ細胞共刺激のモジュレーションによる影響を受けるＴ細胞媒介性および／またはＢ細胞媒介性免疫応答を含む。例示的な免疫応答としては、Ｔ細胞応答、例えば、サイトカイン産生、および細胞性細胞傷害性が挙げられる。さらに、免疫応答という用語は、Ｔ細胞活性化の影響を間接的に受ける免疫応答、例えば、抗体産生（体液性応答）およびサイトカイン応答性細胞、例えばマクロファージの活性化を含む。

【０３９９】

「受容体」は、リガンドに結合することができる生体分子または分子群を意味すると理解されるべきである。受容体は、細胞、細胞形成または生物体における情報の伝達に役立ち得る。受容体は、少なくとも１つの受容体単位を含み、２つまたはそれよりも多くの受容体単位を含有し得、その場合、各受容体単位は、タンパク質分子、例えば、糖タンパク質分子からなり得る。受容体は、リガンドの構造を補完する構造を有し、そのリガンドを結合パートナーとして複合体を形成し得る。細胞の表面上のリガンドに結合した後の受容体のコンフォメーションの変化によってシグナル伝達情報が伝達され得る。本開示によると、受容体は、リガンド、例えば、適切な長さのペプチドまたはペプチド断片と受容体／リガンド複合体を形成することができるＭＨＣクラスＩおよびクラスＩＩのタンパク質を指し得る。ＨＬＡクラスＩおよびクラスＩＩ対立遺伝子によってコードされるクラスＩおよびクラスＩＩ　ＭＨＣペプチドは、本明細書では、当業者の考察の状況の範囲内でよく理解される通り、それぞれＨＬＡクラスＩペプチドおよびＨＬＡクラスＩＩペプチド、またはＨＬＡクラスＩおよびＨＬＡクラスＩＩペプチド、またはＨＬＡクラスＩクラスＩＩタンパク質、またはＨＬＡクラスＩおよびＨＬＡクラスＩＩタンパク質、またはＨＬＡクラスＩおよびクラスＩＩ分子、またはそのようなそれらの一般的なバリアントと称されることも多い。

【０４００】

「リガンド」は、受容体と複合体を形成することができる分子である。本開示によると、リガンドは、例えば、適切な長さを有し、そのアミノ酸配列内に適切な結合性モチーフを有するペプチドまたはペプチド断片を意味し、したがって、ペプチドまたはペプチド断片は、ＭＨＣクラスＩまたはＭＨＣクラスＩＩのタンパク質（すなわち、ＨＬＡクラスＩおよびＨＬＡクラスＩＩタンパク質）に結合し、それと複合体を形成することができると理解されるべきである。

【０４０１】

「抗原」は、免疫応答を刺激することができる分子であり、がん細胞または感染因子または自己免疫疾患によって産生され得る。ヘルパーＴリンパ球（Ｔヘルパー（ＴＨ）細胞）であれ細胞傷害性Ｔリンパ球（ＣＴＬ）であれ、Ｔ細胞によって認識される抗原は、インタクトなタンパク質として認識されるのではなく、細胞の表面上でＨＬＡクラスＩまたはクラスＩＩタンパク質と会合した小さなペプチドとして認識される。天然に存在する免疫応答の過程中、抗原提示細胞（ＡＰＣ）上でＨＬＡクラスＩＩ分子と会合して認識される抗原は、細胞の外側から取得され、内部移行し、ＨＬＡクラスＩＩ分子と会合する小さなペプチドにプロセシングされる。ＡＰＣはまた、外因性抗原をプロセシングし、プロセシングされた抗原をＨＬＡクラスＩ分子上に提示することによってペプチド抗原を交差提

示し得る。ＨＬＡクラスＩ　ＭＨＣ分子と会合して認識されるペプチドを生じる抗原は、一般に、細胞内で産生されるペプチドであり、これらの抗原は、プロセシングされ、クラスＩ　ＭＨＣ分子と会合する。所与のＨＬＡクラスＩまたはクラスＩＩ分子と会合するペプチドは共通の結合性モチーフを有することを特徴とすることが現在理解されており、多数の異なるＨＬＡクラスＩおよびクラスＩＩ分子の結合性モチーフが決定されている。所与の抗原のアミノ酸配列に対応し、所与のＨＬＡクラスＩまたはクラスＩＩ分子の結合性モチーフを含有する合成ペプチドを合成することもできる。次いで、これらのペプチドを妥当なＡＰＣに付加することができ、ＡＰＣを使用してｉｎ　ｖｉｔｒｏまたはｉｎ　ｖｉｖｏにおけるＴヘルパー細胞またはＣＴＬ応答を刺激することができる。結合性モチーフ、ペプチドを合成するための方法、およびＴヘルパー細胞またはＣＴＬ応答を刺激するための方法は全て当業者には公知であり、容易に利用可能である。

【０４０２】

「ペプチド」という用語は、本明細書では「変異体ペプチド」および「新抗原性ペプチド」と互換的に使用される。同様に、「ポリペプチド」という用語は、本明細書では「変異体ポリペプチド」および「新抗原性ポリペプチド」と互換的に使用される。「新抗原」または「ネオエピトープ」とは、発現されたタンパク質において腫瘍特異的変異から生じる腫瘍抗原または腫瘍エピトープのクラスを意味する。本開示は、腫瘍特異的変異を含むペプチド、既知の腫瘍特異的変異、および本開示の方法によって同定された変異体ポリペプチドまたはその断片を含むペプチドをさらに含む。これらのペプチドおよびポリペプチドは、本明細書では「新抗原性ペプチド」または「新抗原性ポリペプチド」と称される。ポリペプチドまたはペプチドは、種々の長さであり得、中性（非荷電）形態または塩の形態のいずれかであり得、また、グリコシル化、側鎖の酸化、リン酸化、または任意の翻訳後修飾などの修飾を有さないかまたはこれらの修飾を含有し得、修飾により本明細書に記載のポリペプチドの生物活性が破壊されない条件に供され得る。一部の実施形態では、本開示の新抗原性ペプチドは、ＨＬＡクラスＩに関しては、２２残基またはそれ未満の長さ、例えば、約８残基から約２２残基まで、約８残基から約１５残基まで、または９残基または１０残基を含み得、ＨＬＡ　Ｃｌａｓｓ　ＩＩに関しては、４０残基またはそれ未満の長さ、例えば、約８残基から約４０残基までの長さ、約８残基から約２４残基までの長さ、約１２残基から約１９残基まで、または約１４残基から約１８残基までを含み得る。一部の実施形態では、新抗原性ペプチドまたは新抗原性ポリペプチドは、ネオエピトープを含む。

【０４０３】

「エピトープ」という用語は、本明細書で定義される抗体、抗体ペプチド、および／または抗体様分子（これだけに限定されないが、Ｔ細胞受容体を含む）に特異的に結合することができる任意のタンパク質決定因子を含む。エピトープ決定因子は、一般には、アミノ酸または糖側鎖などの分子の化学的に活性な表面基からなり、一般に、特定の３次元構造的特徴ならびに特定の電荷特徴を有する。

【０４０４】

「Ｔ細胞エピトープ」は、クラスＩまたはクラスＩＩのＭＨＣ分子がペプチド提示ＭＨＣ分子またはＭＨＣ複合体の形態に結合し得、次いで、これらの形態で、それぞれ細胞傷害性Ｔリンパ球またはヘルパーＴ細胞によって認識され、それが結合する、ペプチド配列である。

【０４０５】

「抗体」という用語は、本明細書で使用される場合、ＩｇＧ（ＩｇＧ１、ＩｇＧ２、ＩｇＧ３、およびＩｇＧ４を含む）、ＩｇＡ（ＩｇＡ１およびＩｇＡ２を含む）、ＩｇＤ、ＩｇＥ、ＩｇＭ、ならびにＩｇＹを含み、また、単鎖全抗体を含めた全抗体、およびその抗原結合性（Ｆａｂ）断片を包含するものとする。抗原結合性抗体断片としては、これだけに限定されないが、Ｆａｂ、Ｆａｂ’およびＦ（ａｂ’）２、Ｆｄ（ＶＨおよびＣＨ１からなる）、単鎖可変断片（ｓｃＦｖ）、単鎖抗体、ジスルフィド連結した可変断片（ｄｓＦｖ）およびＶＬドメインまたはＶＨドメインのいずれかを含む断片が挙げられる。抗体

は、任意の起源動物に由来するものであり得る。単鎖抗体を含めた抗原結合性抗体断片は、可変領域（複数可）を単独で、または以下の全体もしくは部分と組み合わせて含み得る：ヒンジ領域、ＣＨ１ドメイン、ＣＨ２ドメイン、およびＣＨ３ドメイン。可変領域（複数可）およびヒンジ領域、ＣＨ１ドメイン、ＣＨ２ドメイン、およびＣＨ３ドメインの任意の組合せも含まれる。抗体は、例えば、ＨＬＡ付随ペプチドまたはＨＬＡ－ＨＬＡ結合性ペプチド（ＨＬＡ－ペプチド）複合体に特異的に結合するモノクローナル抗体、ポリクローナル抗体、キメラ抗体、ヒト化抗体、およびヒトモノクローナル抗体およびポリクローナル抗体であり得る。可溶性ＨＬＡ－ペプチド複合体または膜結合ＨＬＡ付随ペプチド、例えば、膜からタンパク質分解により切断されたものなどの可溶性タンパク質を富化するために種々の免疫親和性技法が適することが当業者には理解されよう。これらとしては、（１）可溶性タンパク質に特異的に結合することができる１種または複数種の抗体を固定されたまたは可動性の基材（例えば、プラスチックウェルまたは樹脂、ラテックスまたは常磁性ビーズ）に固定化し、（２）生体試料由来の可溶性タンパク質を含有する溶液を、抗体をコーティングした基材を通過させ、それにより、可溶性タンパク質を抗体に結合させる技法が挙げられる。抗体および結合した可溶性タンパク質を伴う基材を溶液から分離し、必要に応じて、例えば抗体を浸す溶液のｐＨおよび／またはイオン強度および／またはイオン組成を変動させることによって抗体および可溶性タンパク質を引き離す。あるいは、抗体と可溶性タンパク質を合わせ、高分子凝集体を形成させる免疫沈降技法を使用することができる。高分子凝集体を分子ふるい技法によってまたは遠心分離によって溶液から分離することができる。

【０４０６】

「免疫精製（ＩＰ）」（または免疫アフィニティー精製または免疫沈降）という用語は、所望の抗原を試料から単離するための、当技術分野で周知のおよび広く使用されているプロセスである。一般に、このプロセスは、所望の抗原を含有する試料を、抗原に対する抗体を固相に共有結合により付着させた親和性マトリックスと接触させることを伴う。試料中の抗原は親和性マトリックスに免疫化学的結合を通じて結合するようになる。次いで、親和性マトリックスを洗浄して、あらゆる結合しなかった種を除去する。親和性マトリックスと接触している溶液の化学組成を変更することによって抗原を親和性マトリックスから取り出す。親和性マトリックスを含有するカラムで免疫精製を行うことができ、その場合、溶液は溶離液である。あるいは、免疫精製はバッチプロセスにおけるものであってよく、その場合、親和性マトリックスを溶液中で懸濁液として維持する。このプロセスの重要なステップは、マトリックスから抗原を取り出すステップである。これは、一般に、親和性マトリックスと接触している溶液のイオン強度を、例えば無機塩を添加することによって増大させることによって実現される。ｐＨを変更することも、抗原と親和性マトリックスの免疫化学的結合を解離させるために有効であり得る。

【０４０７】

「作用物質」は、任意の小分子化学化合物、抗体、核酸分子、またはポリペプチド、またはその断片である。

【０４０８】

「変更」または「変化」は、増大または低減である。変更は、わずか１％、２％、３％、４％、５％、１０％、２０％、３０％、もしくは４０％、５０％、６０％、またはさらには７０％、７５％、８０％、９０％、もしくは１００％もの大きさであり得る。

【０４０９】

「生物学的試料」は、生物体に由来する任意の組織、細胞、流体、または他の材料である。本明細書で使用される場合、「試料」という用語は、生物体に由来する任意の組織、細胞、流体、または他の材料などの生物学的試料を含む。「特異的に結合する」は、化合物（例えば、ペプチド）が、試料、例えば生体試料中のある分子（例えば、ポリペプチド）を認識し、それに結合するが、他の分子には実質的に認識せず、それに結合しないことを指す。

【０４１０】

「捕捉用試薬」は、分子（例えば、核酸分子またはポリペプチド）を選択または単離するための、分子（例えば、核酸分子またはポリペプチド）に特異的に結合する試薬を指す。

【０４１１】

本明細書で使用される場合、「決定すること」、「評価すること」、「アッセイすること」、「測定すること」、「検出すること」という用語およびそれらの文法上の等価物は、定量的決定および定性的決定の両方を指し、したがって、「決定すること」という用語は、本明細書では「アッセイすること」、「測定すること」などと互換的に使用される。定量的決定が意図されている場合、分析物の「量を決定すること」という句などが使用される。定性的決定および／または定量的決定が意図されている場合、分析物の「レベルを決定すること」または分析物を「検出すること」という句が使用される。

【０４１２】

「断片」は、参照タンパク質または核酸と実質的に同一のタンパク質または核酸の部分である。一部の実施形態では、部分は、本明細書に記載の参照タンパク質または核酸の生物活性の少なくとも５０％、７５％、または８０％、または９０％、９５％、またはさらには９９％を保持する。

【０４１３】

「単離された」、「精製された」、「生物学的に純粋な」という用語およびそれらの文法上の等価物は、材料が、そのネイティブな状態で見いだされる、通常はそれに付随する構成成分を種々の程度で含まないことを指す。「単離する」は、元々の供給源または周囲のものからのある程度の分離を示す。「精製する」は、単離よりも高い程度の分離を示す。「精製された」または「生物学的に純粋な」タンパク質は、他の材料を十分に含まず、したがって、いかなる不純物もタンパク質の生物学的特性に実質的な影響を及ぼすこともなく、他の有害な結果を引き起こすこともない。すなわち、本開示の核酸またはペプチドは、組換えＤＮＡ技法によって作製された場合には細胞性材料、ウイルス性材料、もしくは培養培地、または、化学的に合成された場合には化学的前駆体もしくは他の化学物質を実質的に含まなければ、精製されている。純度および均一性は、一般には、分析化学技法、例えば、ポリアクリルアミドゲル電気泳動または高速液体クロマトグラフィーを使用して決定される。「精製された」という用語は、電気泳動ゲルにおいて核酸またはタンパク質から基本的に１つのバンドが生じることを示し得る。修飾、例えばリン酸化またはグリコシル化に供することができるタンパク質に関しては、異なる修飾により異なる単離されたタンパク質を生じさせることができ、それを別々に精製することができる。

【０４１４】

「単離された」ポリペプチド（例えば、ＨＬＡ－ペプチド複合体由来のペプチド）またはポリペプチド複合体（例えば、ＨＬＡ－ペプチド複合体）は、天然に付随する構成成分から分離された本開示のポリペプチドまたはポリペプチド複合体である。一般には、ポリペプチドまたはポリペプチド複合体は、少なくとも６０重量％が天然に付随するタンパク質および天然に存在する有機分子を含まなければ、単離されている。調製物は、少なくとも７５重量％、少なくとも９０重量％、または少なくとも９９重量％が本開示のポリペプチドまたはポリペプチド複合体であり得る。本開示の単離されたポリペプチドまたはポリペプチド複合体は、例えば、天然のソースから抽出することによって、そのようなポリペプチドまたはポリペプチド複合体の１つもしくは複数の構成成分をコードする組換え核酸を発現させることによって、またはポリペプチドまたはポリペプチド複合体の１つもしくは複数の構成成分を化学的に合成することによって、得ることができる。純度は、任意の妥当な方法によって、例えば、カラムクロマトグラフィー、ポリアクリルアミドゲル電気泳動、またはＨＰＬＣ分析によって測定することができる。一部の場合では、ＨＬＡ対立遺伝子にコードされるＭＨＣクラスＩＩタンパク質（すなわち、ＭＨＣクラスＩＩペプチド）は、本文書内では互換的にＨＬＡクラスＩＩタンパク質（またはＨＬＡクラスＩＩペプチド）と称される。

【０４１５】

「ベクター」という用語は、異種核酸を輸送することまたは異種核酸の発現を媒介する

ことができる核酸分子を指す。プラスミドは、「ベクター」という用語に包含される属のうちの１種である。ベクターは、一般には、複製開始点ならびに宿主細胞における複製および／または維持のために必要な他の実体を含有する核酸配列を指す。作動可能に連結した遺伝子および／または核酸配列の発現を方向付けることができるベクターは、本明細書では、「発現ベクター」と称される。一般に、有用な発現ベクターは、多くの場合、「プラスミド」の形態であり、これは、それらのベクター形態では染色体に結合せず、一般にはコードされるＤＮＡの安定発現または一過性発現のために必要な実体を含む環状二本鎖ＤＮＡ分子を指す。本明細書に開示される方法に使用することができる他の発現ベクターとしては、これだけに限定されないが、プラスミド、エピソーム、細菌人工染色体、酵母人工染色体、バクテリオファージまたはウイルスベクターが挙げられ、そのようなベクターは、宿主のゲノムに組み込まれるものまたは細胞において自律的に複製することができるものであり得る。ベクターは、ＤＮＡベクターまたはＲＮＡベクターであり得る。当業者に公知の、同等の機能を果たす他の形態の発現ベクター、例えば、自己複製する染色体外ベクターまたは宿主ゲノム内に組込むことができるベクターも使用することができる。例示的なベクターは、自律複製することができ、かつ／またはそれらが連結した核酸を発現させることができるものである。

【０４１６】

「スペーサー」または「リンカー」という用語は、融合タンパク質に関して使用される場合、融合タンパク質を構成するタンパク質を接合するペプチドを指す。一般に、スペーサーは、タンパク質またはＲＮＡ配列間を接合することまたはタンパク質またはＲＮＡ配列間のいくらかの最小距離または他の空間的関係を保存すること以外の特定の生物活性を有さない。しかし、一部の実施形態では、スペーサーの構成物であるアミノ酸を、分子のフォールディング、正味の電荷、または疎水性などの分子のいくつかの特性に影響を及ぼすように選択することができる。本開示のある実施形態における使用に適したリンカーは当業者には周知であり、それらとしては、これだけに限定されないが、直鎖状または分枝鎖炭素リンカー、複素環式炭素リンカー、またはペプチドリンカーが挙げられる。リンカーは、２つの抗原性ペプチドを、一部の実施形態では各抗原性ペプチドが適当にフォールディングすることを確実にするために十分な距離をあけて分離するために使用される。例示的なペプチドリンカー配列は柔軟な伸長したコンフォメーションをとり、順序づけられた二次構造を生じさせる傾向を示さない。柔軟なタンパク質領域内の典型的なアミノ酸は、Ｇｌｙ、ＡｓｎおよびＳｅｒを含む。Ｇｌｙ、ＡｓｎおよびＳｅｒを含有するアミノ酸配列の事実上あらゆる順列が上記のリンカー配列の基準を満たすと予測される。他のほぼ中性のアミノ酸、例えば、ＴｈｒおよびＡｌａなどもリンカー配列に使用することができる。リンカーとして使用することができるさらに他のアミノ酸配列は、Maratea et al. (1985), Gene 40: 39-46；Murphy et al. (1986) Proc. Nat'l. Acad. Sci. USA 83: 8258-62；米国特許第４，９３５，２３３号；および米国特許第４，７５１，１８０号に開示されている。

【０４１７】

「新形成」という用語は、不適切に高レベルの細胞分裂、不適切に低レベルのアポトーシス、またはその両方によって引き起こされるまたはそれをもたらす任意の疾患を指す。神経膠芽腫は新形成またはがんの１つの非限定的な例である。「がん」または「腫瘍」または「過剰増殖障害」という用語は、制御されない増殖、不死性、転移能、急速な成長および増殖率、ならびにある特定の特徴的な形態学的特色などの、がんを引き起こす細胞に典型的な特徴を有する細胞が存在することを指す。がん細胞は、多くの場合、腫瘍の形態であるが、そのような細胞は、動物内に単独で存在し得る、または、白血病細胞などの非腫瘍形成性がん細胞であり得る。がんとしては、これだけに限定されないが、Ｂ細胞がん（例えば、多発性骨髄腫、ワルデンストレームマクログロブリン血症）、重鎖病（例えば、アルファ鎖病、ガンマ鎖病、およびミュー鎖病など）、良性単クローン性免疫グロブリン血症、および免疫細胞アミロイドーシス、黒色腫、乳がん、肺がん、気管支がん、結腸直腸がん、前立腺がん（例えば、転移性前立腺がん、ホルモン難治性前立腺がん）、膵が

ん、胃がん、卵巣がん、膀胱がん（ｕｒｉｎａｒｙ ｂｌａｄｄｅｒ ｃａｎｃｅｒ）、脳または中枢神経系がん、末梢神経系がん、食道がん（ｅｓｏｐｈａｇｅａｌ ｃａｎｃｅｒ）、子宮頸がん、子宮または子宮体がん、口腔または咽頭のがん、肝がん、腎がん（ｋｉｄｎｅｙ ｃａｎｃｅｒ）、精巣がん、胆道がん、小腸または虫垂がん、唾液腺がん、甲状腺がん、副腎がん、骨肉腫、軟骨肉腫、血液学的組織のがんなどが挙げられる。本開示に包含される方法に適用可能ながんの型の他の非限定的な例としては、ヒト肉腫および癌腫、例えば、線維肉腫、粘液肉腫、脂肪肉腫、軟骨肉腫、骨原性肉腫、脊索腫、血管肉腫、内皮肉腫、リンパ管肉腫、リンパ管内皮肉腫、滑膜腫、中皮腫、ユーイング腫瘍、平滑筋肉腫、横紋筋肉腫、結腸癌、結腸直腸がん、膵がん、乳がん、卵巣がん、扁平上皮癌、基底細胞癌、腺癌、汗腺癌、脂腺癌、乳頭癌、乳頭状腺癌、嚢胞腺癌、髄様癌、気管支原性肺癌、腎細胞癌、ヘパトーマ、胆管癌、肝がん、絨毛癌、セミノーマ、胎児性癌、ウィルムス腫瘍、子宮頸がん、骨がん、脳腫瘍、精巣がん、肺癌、小細胞肺癌、膀胱癌、上皮癌、神経膠腫、星状細胞腫、髄芽腫、頭蓋咽頭腫、上衣腫、松果体腫、血管芽細胞腫、聴神経腫、乏枝神経膠腫、髄膜腫、黒色腫、神経芽細胞腫、網膜芽細胞腫；白血病、例えば、急性リンパ性白血病および急性骨髄球性白血病（骨髄芽球性白血病、前骨芽球性白血病、骨髄単球性白血病、単球性白血病および赤白血病）；慢性白血病（慢性骨髄球性（顆粒球性）白血病および慢性リンパ性白血病）；ならびに真性赤血球増加症、リンパ腫（ホジキン病および非ホジキン病）、多発性骨髄腫、ワルデンストレームマクログロブリン血症、および重鎖病が挙げられる。一部の実施形態では、がんは、これだけに限定されないが、膀胱がん（ｂｌａｄｄｅｒ ｃａｎｃｅｒ）、乳がん、子宮頸がん、結腸がん、婦人科のがん、腎がん（ｒｅｎａｌ ｃａｎｃｅｒ）、喉頭がん、肺がん、口腔がん、頭頸部がん、卵巣がん、膵がん、前立腺がん、または皮膚がんなどの上皮がんである。他の実施形態では、がんは、乳がん、前立腺がん、肺がん、または結腸がんである。さらに他の実施形態では、上皮がんは、非小細胞肺がん、非乳頭状腎細胞癌、子宮頸癌、卵巣癌（例えば、漿液性卵巣癌）、または乳癌である。上皮がんは、これだけに限定されないが、漿液性、類内膜、粘液性、明細胞、ブレンナー、または未分化を含めた種々の他のやり方で特徴付けることができる。一部の実施形態では、本開示を、これだけに限定されないが、マントル細胞リンパ腫を含めたリンパ腫またはそのサブタイプの処置、診断、および／または予後判定に使用する。リンパ球増殖性障害も増殖性疾患とみなされる。

【０４１８】

「ワクチン」という用語は、疾患（例えば、新形成／腫瘍／感染因子／自己免疫疾患）の予防および／または処置のために免疫を生じさせるための組成物を意味すると理解されるべきである。したがって、ワクチンは、抗原を含む医薬であり、ヒトまたは動物においてワクチン接種によって特定の防御および保護物質を生成するために使用されることが意図されたものである。「ワクチン組成物」は、薬学的に許容される賦形剤、担体または希釈剤を含み得る。本開示の態様は、抗原に基づくワクチンの調製における当該技術の使用に関する。これらの実施形態では、ワクチンは、１つまたは複数の疾患特異的抗原性ペプチド（またはそれらをコードする対応する核酸）を指すものとする。一部の実施形態では、抗原に基づくワクチンは、少なくとも２種、少なくとも３種、少なくとも４種、少なくとも５種、少なくとも６種、少なくとも７種、少なくとも８種、少なくとも９種、少なくとも１０種、少なくとも１１種、少なくとも１２種、少なくとも１３種、少なくとも１４種、少なくとも１５種、少なくとも１６種、少なくとも１７種、少なくとも１８種、少なくとも１９種、少なくとも２０種、少なくとも２１種、少なくとも２２種、少なくとも２３種、少なくとも２４種、少なくとも２５種、少なくとも２６種、少なくとも２７種、少なくとも２８種、少なくとも２９種、少なくとも３０種、またはそれよりも多くの抗原性ペプチドを含有する。一部の実施形態では、抗原に基づくワクチンは、２～１００種、２～７５種、２～５０種、２～２５種、２～２０種、２～１９種、２～１８種、２～１７種、２～１６種、２～１５種、２～１４種、２～１３種、２～１２種、２～１０種、２～９種、２～８種、２～７種、２～６種、２～５種、２～４種、３～１００種、３～７５種、３～５０種、３～２５種、３～２０種、３～１９種、３～１８種、３～１７種、３～１６

種、３～１５種、３～１４種、３～１３種、３～１２種、３～１０種、３～９種、３～８種、３～７種、３～６種、３～５種、４～１００種、４～７５種、４～５０種、４～２５種、４～２０種、４～１９種、４～１８種、４～１７種、４～１６種、４～１５種、４～１４種、４～１３種、４～１２種、４～１０種、４～９種、４～８種、４～７種、４～６種、５～１００種、５～７５種、５～５０種、５～２５種、５～２０種、５～１９種、５～１８種、５～１７種、５～１６種、５～１５種、５～１４種、５～１３種、５～１２種、５～１０種、５～９種、５～８種、または５～７種の抗原性ペプチドを含有する。一部の実施形態では、抗原に基づくワクチンは、２種、３種、４種、５種、６種、７種、８種、９種、１０種、１１種、１２種、１３種、１４種、１５種、１６種、１７種、１８種、１９種、または２０種の抗原性ペプチドを含有する。一部の場合では、抗原性ペプチドは、新抗原性ペプチドである。一部の場合では、抗原性ペプチドは、１つまたは複数のネオエピトープを含む。

【０４１９】

「薬学的に許容される」という用語は、連邦または州政府の規制当局によって承認されたもしくは承認可能であることまたは米国薬局方または他のヒトを含めた動物における使用に関して一般に認識されている薬局方に収載されていることを指す。「薬学的に許容される賦形剤、担体または希釈剤」は、作用物質と一緒に対象に投与することができ、作用物質の薬理活性を破壊せず、かつ、治療量の作用物質を送達するために十分な用量で投与された場合に非毒性である、賦形剤、担体または希釈剤を指す。本明細書に記載のプールされた疾患特異的抗原の「薬学的に許容される塩」は、過剰な毒性、刺激、アレルギー反応、または他の問題もしくは併発症を伴わずにヒトまたは動物の組織と接触させて使用するのに適すると当技術分野において一般にみなされる酸または塩基塩であり得る。そのような塩としては、アミンなどの塩基性残基の無機酸塩および有機酸塩、ならびにカルボン酸などの酸性残基のアルカリまたは有機塩が挙げられる。特定の医薬塩としては、これだけに限定されないが、塩酸、リン酸、臭化水素酸、リンゴ酸、グリコール酸、フマル酸、硫酸、スルファミン酸、スルファニル酸、ギ酸、トルエンスルホン酸、メタンスルホン酸、ベンゼンスルホン酸、エタン二スルホン酸、２－ヒドロキシエチルスルホン酸、硝酸、安息香酸、２－アセトキシ安息香酸、クエン酸、酒石酸、乳酸、ステアリン酸、サリチル酸、グルタミン酸、アスコルビン酸、パモン酸、コハク酸、フマル酸、マレイン酸、プロピオン酸、ヒドロキシマレイン酸、ヨウ化水素酸、フェニル酢酸、アルカン酸、例えば、酢酸、ＨＯＯＣ－（ＣＨ２）ｎ－ＣＯＯＨ（式中、ｎは０～４である）などの酸の塩などが挙げられる。同様に、薬学的に許容されるカチオンとしては、これだけに限定されないが、ナトリウム、カリウム、カルシウム、アルミニウム、リチウムおよびアンモニウムが挙げられる。当業者は、本開示および当技術分野における知見から、Remington's Pharmaceutical Sciences, 17th ed., Mack Publishing Company, Easton, PA, p.1418 (1985)によって列挙されているものを含めた、本明細書で提供されるプールされた疾患特異的抗原のためのさらなる薬学的に許容される塩を理解されよう。一般に、薬学的に許容される酸または塩基塩は、塩基性または酸性部分を含有する親化合物から、任意の従来の化学的方法によって合成することができる。簡単に述べると、そのような塩は、遊離の酸または塩基の形態のこれらの化合物を化学量論量の妥当な塩基または酸と妥当な溶媒中で反応させることによって調製することができる。

【０４２０】

本開示の方法に有用な核酸分子としては、本開示のポリペプチドまたはその断片をコードする任意の核酸分子が挙げられる。そのような核酸分子は、内因性核酸配列と１００％同一である必要はないが、一般には実質的な同一性を示す。内因性配列に対して実質的な同一性を有するポリヌクレオチドは、一般には、二本鎖核酸分子の少なくとも一方の鎖とハイブリダイズすることができる。「ハイブリダイズする」は、種々のストリンジェンシーの条件下で相補的なポリヌクレオチド配列またはその一部の間で核酸分子が対形成して二本鎖分子を形成する場合を指す（例えば、Wahl, G. M. and S. L. Berger (1987) Methods Enzymol. 152: 399；Kimmel, A. R. (1987) Methods Enzymol.

152: 507を参照されたい）。例えば、ストリンジェントな塩濃度は、普通は、約７５０ｍＭ未満のＮａＣｌおよび７５ｍＭのクエン酸三ナトリウム、約５００ｍＭ未満のＮａＣｌおよび５０ｍＭのクエン酸三ナトリウム、または約２５０ｍＭ未満のＮａＣｌおよび２５ｍＭのクエン酸三ナトリウムであり得る。低ストリンジェンシーのハイブリダイゼーションは有機溶媒、例えばホルムアミドを用いずに得ることができるが、一方、高ストリンジェンシーのハイブリダイゼーションは少なくとも約３５％ホルムアミド、または少なくとも約５０％ホルムアミドの存在下で得ることができる。ストリンジェントな温度条件は、普通は、少なくとも約３０℃、少なくとも約３７℃、または少なくとも約４２℃の温度を含み得る。ハイブリダイゼーション時間、界面活性剤、例えばドデシル硫酸ナトリウム（ＳＤＳ）の濃度、およびキャリアＤＮＡを含めるか除外するかなどの変動する追加的なパラメータは当業者には周知である。これらの種々の条件を必要に応じて組み合わせることにより、種々のレベルのストリンジェンシーが実現される。例示的な実施形態では、ハイブリダイゼーションを３０℃、７５０ｍＭのＮａＣｌ、７５ｍＭのクエン酸三ナトリウム、および１％ＳＤＳで行うことができる。別の例示的な実施形態では、ハイブリダイゼーションを３７℃、５００ｍＭのＮａＣｌ、５０ｍＭのクエン酸三ナトリウム、１％ＳＤＳ、３５％ホルムアミド、および１００μｇ／ｍｌの変性サケ精子ＤＮＡ（ｓｓＤＮＡ）で行うことができる。別の例示的な実施形態では、ハイブリダイゼーションを４２℃、２５０ｍＭのＮａＣｌ、２５ｍＭのクエン酸三ナトリウム、１％ＳＤＳ、５０％ホルムアミド、および２００μｇ／ｍｌのｓｓＤＮＡで行うことができる。これらの条件の有用な変動は当業者には容易に明らかになるであろう。大多数の適用に関しては、ハイブリダイゼーション後の洗浄ステップに関してもストリンジェンシーを変動させることができる。洗浄ストリンジェンシー条件は、塩濃度によって、および温度によって定義することができる。上記の通り、洗浄ストリンジェンシーは、塩濃度を低下させることによって、または温度を上昇させることによって高めることができる。例えば、洗浄ステップのストリンジェントな塩濃度は、約３０ｍＭ未満のＮａＣｌおよび３ｍＭのクエン酸三ナトリウム、または約１５ｍＭ未満のＮａＣｌおよび１．５ｍＭのクエン酸三ナトリウムであり得る。洗浄ステップのストリンジェントな温度条件は、少なくとも約２５℃、少なくとも約４２℃、または少なくとも約６８℃の温度を含み得る。例示的な実施形態では、洗浄ステップを２５℃、３０ｍＭのＮａＣｌ、３ｍＭのクエン酸三ナトリウム、および０．１％ＳＤＳで行うことができる。他の例示的な実施形態では、洗浄ステップを４２℃、１５ｍＭのＮａＣｌ、１．５ｍＭのクエン酸三ナトリウム、および０．１％ＳＤＳで行うことができる。別の例示的な実施形態では、洗浄ステップを６８℃、１５ｍＭのＮａＣｌ、１．５ｍＭのクエン酸三ナトリウム、および０．１％ＳＤＳで行うことができる。これらの条件の追加的な変動は当業者には容易に明らかになるであろう。ハイブリダイゼーション技法は当業者には周知であり、例えば、Benton and Davis (Science 196: 180, 1977)；Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72: 3961, 1975）；Ausubel et al. (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001)；Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York)；およびSambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New Yorkに記載されている。

【０４２１】

「実質的に同一」とは、ポリペプチドまたは核酸分子が参照アミノ酸配列（例えば、本明細書に記載のアミノ酸配列のうちのいずれか１つ）または参照核酸配列（例えば、本明細書に記載の核酸配列のうちのいずれか１つ）に対して少なくとも５０％の同一性を示すことを指す。そのような配列は、アミノ酸または核酸レベルで、比較のために使用される配列に対して少なくとも６０％、８０％または８５％、９０％、９５％、９６％、９７％、９８％、またはさらには９９％またはそれよりも大きく同一であり得る。配列同一性は、一般には、配列解析ソフトウェア（例えば、Ｓｅｑｕｅｎｃｅ Ａｎａｌｙｓｉｓ Ｓｏｆｔｗａｒｅ Ｐａｃｋａｇｅ ｏｆ ｔｈｅ Ｇｅｎｅｔｉｃｓ Ｃｏｍｐｕｔｅｒ Ｇｒ

ｏｕｐ，Ｕｎｉｖｅｒｓｉｔｙ ｏｆ Ｗｉｓｃｏｎｓｉｎ Ｂｉｏｔｅｃｈｎｏｌｏｇｙ Ｃｅｎｔｅｒ，１７１０ Ｕｎｉｖｅｒｓｉｔｙ Ａｖｅｎｕｅ，Ｍａｄｉｓｏｎ，Ｗｉｓ．５３７０５、ＢＬＡＳＴ、ＢＥＳＴＦＩＴ、ＧＡＰ、またはＰＩＬＥＵＰ／ＰＲＥＴＴＹＢＯＸプログラム）を使用して測定される。そのようなソフトウェアでは、種々の置換、欠失、および／または他の修飾に対して相同性の程度を割り当てることにより、同一または同様の配列を適合させる。保存的置換は、一般には、以下の群内での置換を含む：グリシン、アラニン；バリン、イソロイシン、ロイシン；アスパラギン酸、グルタミン酸、アスパラギン、グルタミン；セリン、トレオニン；リシン、アルギニン；およびフェニルアラニン、チロシン。同一性の程度を決定するための例示的な手法では、ＢＬＡＳＴプログラムを、密接に関連する配列を示すｅ－３からｅ－ｍ°の間の確率スコアで使用することができる。「参照」は比較標準である。

【０４２２】

「対象」または「患者」という用語は、処置、観察、または実験の対象となる動物を指す。単に例として、対象としては、これだけに限定されないが、これだけに限定されないが、ヒトまたは非ヒト哺乳動物、例えば、非ヒト霊長類、マウス、ウシ、ウマ科の動物、イヌ科の動物、ヒツジ、またはネコ科の動物などを含めた哺乳動物が挙げられる。

【０４２３】

「処置する（ｔｒｅａｔ）」、「処置された（ｔｒｅａｔｅｄ）」、「処置すること（ｔｒｅａｔｉｎｇ）」、「処置（ｔｒｅａｔｍｅｎｔ）」などという用語は、障害および／またはそれに伴う症状（例えば、新形成または腫瘍または感染因子または自己免疫疾患）を低減すること、防止すること、または好転させることを指すものとする。「処置すること（ｔｒｅａｔｉｎｇ）」は、疾患（例えば、がんまたは感染因子による感染症または自己免疫疾患）の発症後、または発症の疑い後に対象に対して治療を施行することを指し得る。「処置すること（ｔｒｅａｔｉｎｇ）」は、「緩和すること（ａｌｌｅｖｉａｔｉｎｇ）」という概念を含み、これは、疾患に関連する任意の症状もしくは他の病的影響の発生もしくは再発の頻度、もしくは重症度、および／または治療に付随する副作用を少なくすることを指す。「処置すること（ｔｒｅａｔｉｎｇ）」という用語は、「管理すること（ｍａｎａｇｉｎｇ）」という概念も包含し、これは、患者における疾患または障害の重症度を低下させること、例えば、疾患を有する患者の寿命を延ばすこともしくは生存を延長すること、またはその再発を遅延させること、例えば、疾患に罹患している患者における寛解の期間を長くすることを指す。除外するものではないが、障害または状態を処置することとは、障害、状態、またはそれに伴う症状を完全に排除することを求めるものではないことが理解される。

【０４２４】

「防止する（ｐｒｅｖｅｎｔ）」、「防止すること（ｐｒｅｖｅｎｔｉｎｇ）」、「防止（ｐｒｅｖｅｎｔｉｏｎ）」という用語およびそれらの文法上の等価物は、本明細書で使用される場合、作用物質または化合物の投与の開始時点では疾患または状態に付随する症状が発生していない対象におけるそのような症状の発症を回避するまたは遅延させることを意味する。

【０４２５】

「治療効果」という用語は、障害（例えば、新形成、腫瘍、または感染因子による感染症または自己免疫疾患）またはそれに付随する病的逸脱の症状の１つまたは複数のいくらかの程度の軽減を指す。「治療有効量」は、本明細書で使用される場合、単回用量または複数回用量を細胞または対象に投与した際に、そのような処置の非存在下で予測されるものを超えた、そのような障害を有する患者の生存の延長、障害の１つまたは複数の徴候または症状の低減、防止または遅延などに有効な作用物質の量を指す。「治療有効量」は、治療効果を実現するために必要な量を限定するものと意図されている。当技術分野における通常の技術を有する医師または獣医師は、必要な医薬組成物の「治療有効量」（例えば、ＥＤ５０）を容易に決定し、処方することができる。例えば、医師または獣医師は、医薬組成物に使用する本開示の化合物の用量を所望の治療効果を実現するために必要な用量

よりも低いレベルで開始し、所望の効果が実現されるまで投薬量を徐々に増加することができる。疾患、状態、および障害は、本明細書では互換的に使用される。

【０４２６】

当業者は、「ペプチドタグ」、「アフィニティータグ」、「エピトープタグ」、または「親和性アクセプタータグ」という用語は、本明細書では互換的に使用されることを理解されよう。本明細書で使用される場合、「親和性アクセプタータグ」という用語は、タグ付けされたタンパク質を例えばアフィニティー精製によって容易に検出または精製することを可能にするアミノ酸配列を指す。親和性アクセプタータグは、一般に（必ずしもではないが）ＨＬＡ対立遺伝子のＮ末端またはＣ末端またはその付近に配置される。種々のペプチドタグが当技術分野で周知である。非限定的な例として、ポリ－ヒスチジンタグ（例えば、４～１５個の連続したＨｉｓ残基（配列番号４）、例えば、８個の連続したＨｉｓ残基（配列番号５）など）；ポリ－ヒスチジン－グリシンタグ；ＨＡタグ（例えば、Field et al., Mol. Cell. Biol., 8: 2159, 1988）；ｃ－ｍｙｃタグ（例えば、Evans et al., Mol. Cell. Biol., 5: 3610, 1985）；単純ヘルペスウイルス糖タンパク質Ｄ（ｇＤ）タグ（例えば、Paborsky et al., Protein Engineering, 3: 547, 1990）；ＦＬＡＧタグ（例えば、Hopp et al., BioTechnology, 6: 1204, 1988；米国特許第４，７０３，００４号および４，８５１，３４１号）；ＫＴ３エピトープタグ（例えば、Martine et al., Science, 255: 192, 1992）；チューブリンエピトープタグ（例えば、Skinner, Biol. Chem., 266: 15173, 1991）；Ｔ７遺伝子１０タンパク質ペプチドタグ（例えば、Lutz-Freyemuth et al., Proc. Natl. Acad. Sci. USA, 87: 6393, 1990）；ストレプトアビジンタグ（ＳｔｒｅｐＴａｇ（商標）またはＳｔｒｅｐＴａｇＩＩ（商標）；例えば、Schmidt et al., J. Mol. Biol., 255 (5):753-766, 1996または米国特許第５，５０６，１２１号を参照されたい；Ｓｉｇｍａ－Ｇｅｎｏｓｙｓからも市販されている）；またはＶｅｓｉｃｕｌａｒ Ｓｔｏｍａｔｉｓウイルス糖タンパク質に由来するＶＳＶ－Ｇエピトープタグ；またはサルウイルス５（ＳＶ５）パラミクソウイルスのＰタンパク質およびＶタンパク質に見いだされる小さなエピトープ（Ｐｋ）に由来するＶ５タグが挙げられる。一部の実施形態では、親和性アクセプタータグは「エピトープタグ」であり、これは、認識可能なエピトープ（抗体結合性部位）をＨＬＡ－タンパク質に付加して、対応する抗体の結合をもたらし、それにより、タグ付けされたタンパク質の同定またはアフィニティー精製を可能にするペプチドタグの一種である。エピトープタグの非限定的な例は、ＩｇＧに結合するプロテインＡまたはプロテインＧである。一部の実施形態では、ＩｇＧ Ｓｅｐｈａｒｏｓｅ ６ Ｆａｓｔ Ｆｌｏｗクロマトグラフィー樹脂のマトリックスをヒトＩｇＧに共有結合によりカップリングする。この樹脂は、プロテインＡでタグ付けされたタンパク質を迅速かつ好都合に精製するための高い流速を可能にするものである。多数の他のタグ部分が当業者には公知であり、当業者が構想することができ、また、本明細書で意図されている。親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体のエレメントとして発現させることができる限りは、任意のペプチドタグを使用することができる。

【０４２７】

本明細書で使用される場合、「親和性分子」という用語は、親和性アクセプターペプチドに化学的特異性で結合する分子またはリガンドを指す。化学的特異性は、タンパク質の結合性部位の、特異的なリガンドに結合する能力である。タンパク質が結合することができるリガンドが少ないほど、その特異性は大きくなる。特異性は、所与のタンパク質とリガンドの結合の強度を記述するものである。この関係は、タンパク質－リガンド系に関して結合した状態と結合していない状態の平衡を特徴付けるものである解離定数（ＫＤ）によって説明することができる。

【０４２８】

「親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体」という用語は、ＨＬＡクラスＩまたはクラスＩＩ関連ペプチドまたはその一部と、それと特異的に結合した、親和性アクセプターペプチドを含む単一の対立遺伝子の組換えＨＬＡクラスＩまたはクラ

スＩＩペプチドとを含む複合体を指す。

【０４２９】

「特異的結合（ｓｐｅｃｉｆｉｃ　ｂｉｎｄｉｎｇ）」または「特異的に結合（ｓｐｅｃｉｆｉｃａｌｌｙ　ｂｉｎｄｉｎｇ）」という用語は、親和性分子および親和性アクセプタータグまたはエピトープおよびＨＬＡペプチドの相互作用に関して使用される場合、相互作用がタンパク質上の特定の構造（例えば、抗原性決定因子またはエピトープ）の存在に依存することを意味する；言い換えれば、親和性分子は、タンパク質全般にではなく、特異的な親和性アクセプターペプチド構造を認識し、それに結合する。

【０４３０】

本明細書で使用される場合、「親和性」という用語は、結合対の２つのメンバー、例えば、「親和性アクセプタータグ」と「親和性分子」およびＨＬＡ結合性ペプチドとＨＬＡクラスＩまたはクラスＩＩ分子の間の結合の強度の評価基準を指す。ＫＤは解離定数であり、容量モル濃度の単位を有する。親和定数は解離定数の逆である。親和定数は、時にはこの化学的実体を記載するための一般用語として使用される。親和定数は結合のエネルギーの直接評価基準である。親和性は、例えば、市販のＢｉａｃｏｒｅ　ＳＰＲ単位を使用して表面プラズモン共鳴（ＳＰＲ）によって実験的に決定することができる。親和性はまた、阻害性濃度５０（ＩＣ５０）として表すこともでき、これは、ペプチドの５０％が置き換えられる濃度である。同様に、ｌｎＩＣ５０はＩＣ５０の自然対数を指す。Ｋｏｆｆは、例えば、親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体からの親和性分子の解離に関する解離速度定数を指す。

【０４３１】

一部の実施形態では、親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体は、ビオチンアクセプターペプチド（ＢＡＰ）を含み、ストレプトアビジン／ＮｅｕｔｒＡｖｉｄｉｎビーズを使用して複合体細胞混合物から免疫精製される。ビオチン－アビジン／ストレプトアビジン結合は、天然で知られている最も強い非共有結合性相互作用である。この特性は、ビオチンが共有結合により付着したタンパク質の免疫精製などの広範囲の適用のための生物学的ツールとして活用される。例示的な実施形態では、ＨＬＡ対立遺伝子をコードする核酸配列に免疫精製のための親和性アクセプタータグとしてビオチンアクセプターペプチド（ＢＡＰ）を組み込む。ＢＡＰを、ｉｎ　ｖｉｖｏまたはｉｎ　ｖｉｔｒｏにおいて、タグ内の単一のリシン残基を特異的にビオチン化することができる（例えば、米国特許第５，７２３，５８４号；５，８７４，２３９号；および５，９３２，４３３号；および英国特許第ＧＢ２３７００３９号）。ＢＡＰは、一般には１５アミノ酸長であり、ビオチンアクセプター残基として単一のリシンを含有する。一部の実施形態では、ＢＡＰを単一の対立遺伝子ＨＬＡペプチドのＮ末端またはＣ末端またはその付近に配置する。一部の実施形態では、ＢＡＰをＨＬＡクラスＩペプチドの重鎖ドメインとβ２ミクログロブリンドメインの間に配置する。一部の実施形態では、ＢＡＰをＨＬＡクラスＩＩペプチドのβ鎖ドメインとα鎖ドメインの間に配置する。一部の実施形態では、ＢＡＰをＨＬＡクラスＩの重鎖のα１ドメイン、α２ドメイン、およびα３ドメインの間のループ領域内、またはＨＬＡクラスＩＩのそれぞれα鎖およびβ鎖のα１ドメインとα２ドメインの間およびβ１ドメインとβ２ドメインの間に配置する。ビオチン化のためにＢＡＰが組み込まれたＨＬＡクラスＩおよびクラスＩＩ発現のために設計された例示的な構築物および免疫精製は図２に記載されている。

【０４３２】

本明細書で使用される場合、「ビオチン」という用語は、化合物ビオチン自体ならびにその類似体、誘導体およびバリアントを指す。したがって、「ビオチン」という用語は、ビオチン様化合物を含めたビオチン（シス－ヘキサヒドロ－２－オキソ－１Ｈ－チエノ［３，４］イミダゾール－４－ペンタン酸）および任意の誘導体およびその類似体を含む。そのような化合物としては、例えば、ビオチン－ｅ－Ｎ－リシン、ビオシチンヒドラジド、２－イミノビオチンのアミノまたはスルフヒドリル誘導体およびビオチニル－Ｅ－アミノカプロン酸－Ｎ－ヒドロキシスクシンイミドエステル、スルホスクシンイミドイミノビ

オチン、ビオチンブロモアセチルヒドラジド、ｐ－ジアゾベンゾイルビオシチン、３－（Ｎ－マレイミドプロピオニル）ビオシチン、デスチオビオチンなどが挙げられる。「ビオチン」という用語は、リザビジン、アビジン、ストレプトアビジン、タマビジン部分、または他のアビジン様ペプチドのうちの１つまたは複数に特異的に結合し得るビオチンバリアントも含む。

【０４３３】

本明細書で使用される場合、「ＰＰＶ決定法」は、提示ＰＰＶ決定法を指し得る。例えば、「ＰＰＶ決定法」は、（ａ）複数の試験提示予測を生成するために、機械学習ＨＬＡペプチド提示予測モデルなどのＨＬＡペプチド提示予測モデルを使用して複数の試験ペプチド配列のアミノ酸情報を処理するステップであって、各試験提示予測により、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子などの細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質により複数の試験ペプチド配列のうちの所与の試験ペプチド配列が提示され得る可能性が示され、複数の試験ペプチド配列が、（ｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された少なくとも１つのヒットペプチド配列および（ｉｉ）対象と同じ種の生物体などの生物体のゲノムによってコードされるタンパク質内に含有される少なくとも４９９のデコイペプチド配列を含む少なくとも５００の試験ペプチド配列を含み、複数の試験ペプチド配列が、デコイペプチド配列の数に対するヒットペプチド配列の数が１つ未満という比、例えば、少なくとも１つのヒットペプチド配列の少なくとも４９９のデコイペプチド配列に対する１：４９９の比などを含む、ステップと、（ｂ）複数の試験ペプチド配列の上位パーセンテージ、例えば、上位複数の試験ペプチド配列の０．２％などを、細胞のクラスＩＩ　ＨＬＡ対立遺伝子によって提示されるものとして同定またはコールするステップと、（ｃ）ＨＬＡペプチド提示予測モデルのＰＰＶを算出するステップであって、ＰＰＶが、複数のうちの、質量分析によって細胞のクラスＩＩ　ＨＬＡ対立遺伝子によって提示されるものとして観察されたペプチドである、細胞のクラスＩＩ　ＨＬＡ対立遺伝子によって提示されるものとして同定またはコールされた試験ペプチド配列の分率である、ステップとを含む方法を指し得る。一部の実施形態では、デコイペプチドは、同じ長さのものである、すなわち、ヒットペプチドと同じ数のアミノ酸を含む。一部の実施形態では、デコイペプチドは、ヒットペプチドと比較してアミノ酸を１つ多くまたは１つ少なく含み得る。一部の実施形態では、デコイペプチドは、内因性ペプチドであるペプチドである。一部の実施形態では、デコイペプチドは合成ペプチドである。一部の実施形態では、デコイペプチドは、第１のＭＨＣクラスＩまたはクラスＩＩタンパク質に結合することが質量分析によって同定された内因性ペプチドであり、第１のＭＨＣクラスＩまたはクラスＩＩタンパク質は、ヒットペプチドに結合する第２のＭＨＣクラスＩまたはクラスＩＩタンパク質とは別個である。一部の実施形態では、デコイペプチドは、スクランブルペプチドであり得、例えば、デコイペプチドは、ペプチドの長さの範囲内でアミノ酸位がヒットペプチドのアミノ酸位に対して再配列されたアミノ酸配列を含み得る。一部の実施形態では、ＰＰＶ決定法は、提示ＰＰＶ決定法であり得る。一部の実施形態では、ヒットペプチド配列の数とデコイペプチド配列の数の比は、約１：１０、１：２０、１：５０、１：１００、１：２５０、１：５００、１：１０００、１：１５００、１：２０００、１：２５００、１：５０００、１：７５００、１：１００００、１：２５０００、１：５００００または１：１０００００である。一部の実施形態では、少なくとも１つのヒットペプチド配列は、少なくとも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または１００のヒットペプチド配列を含む。一部の実施形態では、少

なくとも４９９のデコイペプチド配列は、少なくとも５００　６００、７００、８００、９００、１０００、１１００、１２００、１３００、１４００、１５００、１６００、１７００、１８００、１９００、２０００、２１００、２２００、２３００、２４００、２５００、２６００、２７００、２８００、２９００、３０００、３１００、３２００、３３００、３４００、３５００、３６００、３７００、３８００、３９００、４０００、４１００、４２００、４３００、４４００、４５００、４６００、４７００、４８００、４９００、５０００、５１００、５２００、５３００、５４００、５５００、５６００、５７００、５８００、５９００、６０００、６１００、６２００、６３００、６４００、６５００、６６００、６７００、６８００、６９００、７０００、７１００、７２００、７３００、７４００、７５００、７６００、７７００、７８００、７９００、８０００、８１００、８２００、８３００、８４００、８５００、８６００、８７００、８８００、８９００、９０００、９１００、９２００、９３００、９４００、９５００、９６００、９７００、９８００、９９００、１００００、１１０００、１２０００、１３０００、１４０００、１５０００、１６０００、１７０００、１８０００、１９０００、２００００、２１０００、２２０００、２３０００、２４０００、２５０００、２６０００、２７０００、２８０００、２９０００、３００００、３１０００、３２０００、３３０００、３４０００、３５０００、３６０００、３７０００、３８０００、３９０００、４００００、４１０００、４２０００、４３０００、４４０００、４５０００、４６０００、４７０００、４８０００、４９０００、５００００、５２５００、５５０００、５７５００、６００００、６２５００、６５０００、６７５００、７００００、７２５００、７５０００、７７５００、８００００、８２５００、８５０００、８７５００、９００００、９２５００、９５０００、９７５００、１０００００、１２５０００、１５００００、１７５０００、２００００、２２５０００、２５００００、２７５０００、３００００、３２５０００、３５００００、３７５０００、４００００、４２５０００、４５００００、４７５０００、５０００００、６００００、７００００、８００００、９００００または１０００００のデコイペプチド配列を含む。一部の実施形態では、少なくとも５００の試験ペプチド配列は、少なくとも６００、７００、８００、９００、１０００、１１００、１２００、１３００、１４００、１５００、１６００、１７００、１８００、１９００、２０００、２１００、２２００、２３００、２４００、２５００、２６００、２７００、２８００、２９００、３０００、３１００、３２００、３３００、３４００、３５００、３６００、３７００、３８００、３９００、４０００、４１００、４２００、４３００、４４００、４５００、４６００、４７００、４８００、４９００、５０００、５１００、５２００、５３００、５４００、５５００、５６００、５７００、５８００、５９００、６０００、６１００、６２００、６３００、６４００、６５００、６６００、６７００、６８００、６９００、７０００、７１００、７２００、７３００、７４００、７５００、７６００、７７００、７８００、７９００、８０００、８１００、８２００、８３００、８４００、８５００、８６００、８７００、８８００、８９００、９０００、９１００、９２００、９３００、９４００、９５００、９６００、９７００、９８００、９９００、１００００、１１０００、１２０００、１３０００、１４０００、１５０００、１６０００、１７０００、１８０００、１９０００、２００００、２１０００、２２０００、２３０００、２４０００、２５０００、２６０００、２７０００、２８０００、２９０００、３００００、３１０００、３２０００、３３０００、３４０００、３５０００、３６０００、３７０００、３８０００、３９０００、４００００、４１０００、４２０００、４３０００、４４０００、４５０００、４６０００、４７０００、４８０００、４９０００、５００００、５２５００、５５０００、５７５００、６００００、６２５００、６５０００、６７５００、７００００、７２５００、７５０００、７７５００、８００００、８２５００、８５０００、８７５００、９００００、９２５００、９５０００、９７５００、１０００００、１２５０００、１５００００、１７５０００、２００００、２２５０００、２５００００、２７５０００、３００００、３２５０００、３５００００、３７５０００、４００００、４２５０００、４５００００、４７５０００、５０００

００、６０００００、７０００００、８０００００、９０００００または１０００００の試験ペプチド配列を含む。一部の実施形態では、複数の試験ペプチド配列の上位パーセンテージを細胞のクラスＩＩ　ＨＬＡ対立遺伝子によって提示されるものとして同定またはコールするステップは、上位０．２０％、０．３０％、０．４０％、０．５０％、０．６０％、０．７０％、０．８０％、０．９０％、１．００％、１．１０％、１．２０％、１．３０％、１．４０％、１．５０％、１．６０％、１．７０％、１．８０％、１．９０％、２．００％、２．１０％、２．２０％、２．３０％、２．４０％、２．５０％、２．６０％、２．７０％、２．８０％、２．９０％、３．００％、３．１０％、３．２０％、３．３０％、３．４０％、３．５０％、３．６０％、３．７０％、３．８０％、３．９０％、４．００％、４．１０％、４．２０％、４．３０％、４．４０％、４．５０％、４．６０％、４．７０％、４．８０％、４．９０％、５．００％、５．１０％、５．２０％、５．３０％、５．４０％、５．５０％、５．６０％、５．７０％、５．８０％、５．９０％、６．００％、６．１０％、６．２０％、６．３０％、６．４０％、６．５０％、６．６０％、６．７０％、６．８０％、６．９０％、７．００％、７．１０％、７．２０％、７．３０％、７．４０％、７．５０％、７．６０％、７．７０％、７．８０％、７．９０％、８．００％、８．１０％、８．２０％、８．３０％、８．４０％、８．５０％、８．６０％、８．７０％、８．８０％、８．９０％、９．００％、９．１０％、９．２０％、９．３０％、９．４０％、９．５０％、９．６０％、９．７０％、９．８０％、９．９０％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％を細胞のクラスＩＩ　ＨＬＡ対立遺伝子によって提示されるものとして同定またはコールすることを含む。一部の実施形態では、細胞は、単一対立遺伝子細胞である。

【０４３４】

本明細書で使用される場合、「ＰＰＶ決定法」は、結合ＰＰＶ決定法を指し得る。例えば、「ＰＰＶ決定法」は、（ａ）複数の試験結合予測を生成するために、機械学習ＨＬＡペプチド結合予測モデルなどのＨＬＡペプチド結合予測モデルを使用して複数の試験ペプチド配列のアミノ酸情報を処理するステップであって、各試験結合予測により、対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子などの細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質が複数の試験ペプチド配列のうちの所与の試験ペプチド配列に結合する可能性が示され、複数の試験ペプチド配列が、（ｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された少なくとも１つのヒットペプチド配列および（ｉｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された少なくとも１つのペプチド配列を含むタンパク質内に含有される少なくとも１９のデコイペプチド配列を含む少なくとも２０の試験ペプチド配列を含み、複数の試験ペプチド配列が、デコイペプチド配列の数に対するヒットペプチド配列の数が１つ未満という比、例えば、少なくとも１つのヒットペプチド配列と少なくとも１９のデコイペプチド配列である１：１９の比などを含み、（ｂ）複数の試験ペプチド配列の上位パーセンテージ、例えば、上位複数の試験ペプチド配列の５％などを、ＨＬＡタンパク質に結合するものとして同定またはコールするステップと、（ｃ）ＨＬＡペプチド結合予測モデルのＰＰＶを算出するステップであって、ＰＰＶが、複数のうちの、質量分析によって細胞のクラスＩＩ　ＨＬＡ対立遺伝子によって提示されるものとして観察されたペプチドである、細胞のクラスＩＩ　ＨＬＡ対立遺伝子に結合するものとして同定またはコールされた試験ペプチド配列の分率である方法を指し得る。一部の実施形態では、ヒットペプチド配列の数とデコイペプチド配列の数の比は、約１：２、１：３、１：４、１：５、１：１０、１：２０、１：２５、１：３０、１：４０、１：５０、１：７５、１：１００、１：２００、１：２５０、１：５００または１：１０００である。一部の実施形態では、少なくとも１つのヒットペプチド配列は、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または１００のヒットペプチド配列を含む。一部の実施形態では、少なくとも１９のデコイペプチド配列は、少なくとも３０、４０、５０、６０、７０、８０、９０、１００、１１０、１２０、１３０、１４０、１５０、１６０、１７０、１８０、１９０、２００、２１０、２２０、２３０、２４０、２５０、２６０、２７０、２８０、２９０、３００、３１０、３２０、３３０、３４０、３５０、３６０、３７０、３８０、３９０、４００、４１０、４２０、４３０、４４０、４５０、４６０、４７０、４８０、４９０、５００ ６００、７００、８００、９００、１０００、１１００、１２００、１３００、１４００、１５００、１６００、１７００、１８００、１９００、２０００、２１００、２２００、２３００、２４００、２５００、２６００、２７００、２８００、２９００、３０００、３１００、３２００、３３００、３４００、３５００、３６００、３７００、３８００、３９００、４０００、４１００、４２００、４３００、４４００、４５００、４６００、４７００、４８００、４９００、５０００、５１００、５２００、５３００、５４００、５５００、５６００、５７００、５８００、５９００、６０００、６１００、６２００、６３００、６４００、６５００、６６００、６７００、６８００、６９００、７０００、７１００、７２００、７３００、７４００、７５００、７６００、７７００、７８００、７９００、８０００、８１００、８２００、８３００、８４００、８５００、８６００、８７００、８８００、８９００、９０００、９１００、９２００、９３００、９４００、９５００、９６００、９７００、９８００、９９００、１００００、１１０００、１２０００、１３０００、１４０００、１５０００、１６０００、１７０００、１８０００、１９０００、２００００、２１０００、２２０００、２３０００、２４０００、２５０００、２６０００、２７０００、２８０００、２９０００、３００００、３１０００、３２０００、３３０００、３４０００、３５０００、３６０００、３７０００、３８０００、３９０００、４００００、４１０００、４２０００、４３０００、４４０００、４５０００、４６０００、４７０００、４８０００、４９０００、５００００、５２５００、５５０００、５７５００、６００００、６２５００、６５０００、６７５００、７００００、７２５００、７５０００、７７５００、８００００、８２５００、８５０００、８７５００、９００００、９２５００、９５０００、９７５００、１０００００、１２５０００、１５００００、１７５０００、２００００、２２５０００、２５００００、２７５０００、３００００、３２５０００、３５００００、３７５０００、４００００、４２５００００、４５００００、４７５０００、５００００、６００００、７００００、８００００、９００００または１０００００のデコイペプチド配列を含む。一部の実施形態では、少なくとも２０の試験ペプチド配列は、少なくとも３０、４０、５０、６０、７０、８０、９０、１００、１１０、１２０、１３０、１４０、１５０、１６０、１７０、１８０、１９０、２００、２１０、２２０、２３０、２４０、２５０、２６０、２７０、２８０、２９０、３００、３１０、３２０、３３０、３４０、３５０、３６０、３７０、３８０、３９０、４００、４１０、４２０、４３０、４４０、４５０、４６０、４７０、４８０、４９０、５００ ６００、７００、８００、９００、１０００、１１００、１２００、１３００、１４００、１５００、１６００、１７００、１８００、１９００、２０００、２１００、２２００、２３００、２４００、２５００、２６００、２７００、２８００、２９００、３０００、３１００、３２００、３３００、３４００、３５００、３６００、３７００、３８００、３９００、４０００、４１００、４２００、４３００、４４００、４５００、４６００、４７００、４８００、４９００、５０００、５１００、５２００、５３００、５４００、５５００、５６００、５７００、５８００、５９００、６０００、６１００、６２００、６３００、６４００、６５００、６６００、６７００、６８００、６９００、７０００、７１００、７２００、７３００、７４００、７５００、７６００、７７００、７８００、７９００、８０００、８１００、８２００、８３００、８４

００、８５００、８６００、８７００、８８００、８９００、９０００、９１００、９２００、９３００、９４００、９５００、９６００、９７００、９８００、９９００、１００００、１１０００、１２０００、１３０００、１４０００、１５０００、１６０００、１７０００、１８０００、１９０００、２００００、２１０００、２２０００、２３０００、２４０００、２５０００、２６０００、２７０００、２８０００、２９０００、３００００、３１０００、３２０００、３３０００、３４０００、３５０００、３６０００、３７０００、３８０００、３９０００、４００００、４１０００、４２０００、４３０００、４４０００、４５０００、４６０００、４７０００、４８０００、４９０００、５００００、５２５００、５５０００、５７５００、６００００、６２５００、６５０００、６７５００、７００００、７２５００、７５０００、７７５００、８００００、８２５００、８５０００、８７５００、９００００、９２５００、９５０００、９７５００、１０００００、１２５０００、１５００００、１７５０００、２００００、２２５０００、２５００００、２７５０００、３００００、３２５０００、３５００００、３７５０００、４００００、４２５０００、４５００００、４７５０００、５００００、６００００、７００００、８００００、９００００または１０００００の試験ペプチド配列を含む。一部の実施形態では、複数の試験ペプチド配列の上位パーセンテージを細胞のクラスＩＩ　ＨＬＡ対立遺伝子によって提示されるものとして同定またはコールするステップは、上位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％、または４０％を細胞のクラスＩＩ　ＨＬＡ対立遺伝子によって提示されるものとして同定またはコールすることを含む。一部の実施形態では、細胞は、単一対立遺伝子細胞である。

【０４３５】

ヒト白血球抗原（ＨＬＡ）系

免疫系は、２つの機能的下位系：自然免疫系および適応免疫系に分類することができる。自然免疫系は、感染に対する防御の第一選択であり、大多数の潜在的な病原体が、例えば注目すべき感染症を引き起こす前に自然免疫系によって迅速に中和される。適応免疫系は、侵入生物体の抗原と称される分子構造に反応する。自然免疫系とは異なり、適応免疫系は病原体に対して高度に特異的である。適応免疫では長続きする保護ももたらされ得る；例えば、麻疹から回復したある人はそれから生涯にわたって麻疹から保護される。適応免疫反応には２つの型が存在し、それらには体液性免疫反応および細胞媒介性免疫反応が含まれる。体液性免疫反応では、Ｂ細胞によって体液中に分泌された抗体が病原体由来の抗原に結合し、それにより、種々の機構、例えば、補体媒介性溶解を通じた病原体の消失が導かれる。細胞媒介性免疫反応では、他の細胞を破壊することができるＴ細胞が活性化される。例えば、疾患に関連するタンパク質が細胞に存在する場合、それらのタンパク質は細胞内でタンパク質分解により断片化されてペプチドになる。次いで、特定の細胞内タンパク質自体が抗原またはこのように形成されたペプチドに付着し、それらを細胞の表面に輸送し、そこで、それらを体のＴ細胞内の分子防御機構に対して提示する。細胞傷害性Ｔ細胞がこれらの抗原を認識し、その抗原を有する細胞を死滅させる。

【０４３６】

「主要組織適合性遺伝子複合体（ＭＨＣ）」、「ＭＨＣ分子」、または「ＭＨＣタンパク質」という用語は、タンパク質抗原のタンパク質分解による切断の結果生じたペプチドに結合することができる、潜在的なＴ細胞エピトープを表し、それらを細胞表面に輸送し、ペプチドを特定の細胞、例えば、細胞傷害性Ｔリンパ球またはヘルパーＴ細胞に提示するタンパク質を指す。ヒトＭＨＣはＨＬＡ複合体とも称される。したがって、「ヒト白血球抗原（ＨＬＡ）系」、「ＨＬＡ分子」または「ＨＬＡタンパク質」という用語は、ヒトにおけるＭＨＣタンパク質をコードする遺伝子複合体を指す。ＭＨＣという用語は、マウス種では「Ｈ－２」複合体と称される。「主要組織適合性遺伝子複合体（ＭＨＣ）」、「ＭＨＣ分子」、「ＭＨＣタンパク質」および「ヒト白血球抗原（ＨＬＡ）系」、「ＨＬＡ

分子」、「ＨＬＡタンパク質」という用語は本明細書では互換的に使用されることが当業者には理解されよう。

【０４３７】

ＨＬＡタンパク質は、ＨＬＡクラスＩおよびＨＬＡクラスＩＩと称される２つの型に分類される。２つのＨＬＡクラスのタンパク質の構造は非常に類似しているが、全く異なる機能を有する。ＨＬＡクラスＩタンパク質は、大多数の腫瘍細胞を含めた、体のほとんど全ての細胞の表面上に存在する。ＨＬＡクラスＩタンパク質には、通常、内因性タンパク質または細胞の内側に存在する病原体を起源とする抗原がローディングされ、次いで、ナイーブなまたは細胞傷害性Ｔリンパ球（ＣＴＬ）に提示される。ＨＬＡクラスＩＩタンパク質は、これだけに限定されないが、樹状細胞、Ｂ細胞、およびマクロファージを含めた抗原提示細胞（ＡＰＣ）上に存在する。ＨＬＡクラスＩＩタンパク質は主に、外部の抗原供給源、例えば、細胞の外側からプロセシングされたペプチドをヘルパーＴ細胞に提示する。ＨＬＡクラスＩタンパク質が結合するペプチドの大部分は、生物体自体の健康な宿主細胞において産生される細胞質タンパク質を起源とし、通常は免疫反応を刺激しない。

【０４３８】

ＨＬＡクラスＩ分子（図１）は、非共有結合により連結した２つのポリペプチド鎖、ＨＬＡによりコードされるα鎖（重鎖、４４～４７ｋＤ）およびβ２ミクログロブリン（またはβ２ｍ）と称されるＨＬＡによりコードされないサブユニット（１２ｋＤ）からなる。α鎖は、３つの細胞外ドメイン、α１、α２およびα３、ならびに膜貫通領域を有し、α１領域およびα２領域は、約７～１３アミノ酸（例えば、約８～１１アミノ酸、または９アミノ酸または１０アミノ酸）のペプチドに結合することができる。ＨＬＡクラス１分子は、適切な結合性モチーフを有するペプチドに結合し、それを細胞傷害性Ｔリンパ球に提示する。ＨＬＡクラス１重鎖は、ＨＬＡ－Ａ単量体とも称されるＨＬＡ－Ａ対立遺伝子のタンパク質産物、またはＨＬＡ－Ｂ対立遺伝子のタンパク質産物（同様に、ＨＬＡ－Ｂ単量体）またはＨＬＡ－Ｃ対立遺伝子のタンパク質産物（ＨＬＡ－Ｃ単量体）であり得、その各々がβ－２－ミクログロブリンと複合体を形成する。α１は、非ＨＬＡタンパク質β２ｍの上に載っている；β２ｍは、ヒト１５番染色体上に位置するベータ２－ミクログロブリン遺伝子によってコードされる。α３ドメインは膜貫通領域と接続しており、それにより、ＨＬＡクラスＩ分子が細胞膜とつながっている。提示されるペプチドは、α１／α２ヘテロ二量体（２つの同一でないサブユニットで構成される分子）の中心領域にあるペプチド結合性溝の底に保持される。ＨＬＡクラスＩ－Ａ、ＨＬＡクラスＩ－ＢまたはＨＬＡクラスＩ－Ｃは、高度に多型的である。ＨＬＡクラス１－Ａ遺伝子（ＨＬＡ－Ａ遺伝子と称される）、ＨＬＡクラス１－Ｂ遺伝子（ＨＬＡ－Ｂ遺伝子と称される）およびＨＬＡクラス１－Ｃ遺伝子（ＨＬＡ－Ｃ遺伝子と称される）の各々は、８つのエクソンを含有し、エクソン１はリーダーペプチドをコードし、エクソン２および３はα１およびα２ドメインをコードし、エクソン５は膜貫通領域をコードし、エクソン６および７は細胞質尾部をコードする。エクソン２およびエクソン３の多型が各クラス１分子のペプチド結合特異性を担う。ＨＬＡクラスＩ－Ｂ遺伝子（ＨＬＡ－Ｂ）は多くの可能性のあるバリエーション、発現パターンおよび提示される抗原を有する。この群は、ＨＬＡ遺伝子座内でコードされる群、例えば、ＨＬＡ－Ｅ、ＨＬＡ－Ｆ、ＨＬＡ－Ｇに細分され、また、例えば、ＵＬＢＰ、Ｒａｅ１およびＨ６０などのストレスリガンドではない。これらの分子の多くの抗原／リガンドは依然として不明であるが、ＣＤ８＋　Ｔ細胞、ＮＫＴ細胞、およびＮＫ細胞の各々と相互作用し得る。

【０４３９】

一部の実施形態では、本開示は、非古典的ＨＬＡクラスＩ－Ｅ対立遺伝子を利用する。ＨＬＡ－Ｅ分子は、ナチュラルキラー（ＮＫ）細胞およびＣＤ８＋　Ｔ細胞によって認識される。ＨＬＡ－Ｅは、肺、肝臓、皮膚および胎盤細胞を含めたほとんど全ての組織において発現される。ＨＬＡ－Ｅ発現は固形腫瘍（例えば、骨肉腫および黒色腫）においても検出される。ＨＬＡ－Ｅ分子は、ＣＤ８＋　Ｔ細胞に発現されるＴＣＲに結合し、その結果、Ｔ細胞活性化がもたらされる。ＨＬＡ－Ｅは、ＮＫ細胞およびＣＤ８＋　Ｔ細胞上に

発現されるＣＤ９４／ＮＫＧ２受容体に結合することも公知である。ＣＤ９４は、ＮＫＧ２のいくつかの異なるアイソフォームと対形成して、細胞の活性化を阻害する潜在性を有する受容体（ＮＫＧ２Ａ、ＮＫＧ２Ｂ）または細胞の活性化を促進する潜在性を有する受容体（ＮＫＧ２Ｃ）を形成し得る。ＨＬＡ－Ｅは、大多数のＨＬＡ－Ａ分子、ＨＬＡ－Ｂ分子、ＨＬＡ－Ｃ分子、およびＨＬＡ－Ｇ分子のリーダー配列のアミノ酸残基３～１１に由来するペプチドに結合することができるが、それ自体のリーダーペプチドには結合することができない。ＨＬＡ－Ｅは、ＨＬＡ－Ａ、ＨＬＡ－Ｂ、およびＨＬＡ－Ｃ対立遺伝子と同様に、内因性タンパク質に由来するペプチドを提示することも示されている。生理的条件下で、ＣＤ９４／ＮＫＧ２Ａと、ＨＬＡクラスＩリーダー配列由来のペプチドがローディングされたＨＬＡ－Ｅの会合により、通常、阻害シグナルが誘導される。サイトメガロウイルス（ＣＭＶ）は、ＨＬＡ－Ａリーダーを模倣するＵＬ４０糖タンパク質を発現することにより、ＮＫ細胞免疫サーベイランスからエスケープする機構を利用する。しかし、ＣＤ８＋　Ｔ細胞が、ＣＭＶ　Ｔｏｌｅｄｏ株に由来するＵＬ４０ペプチドがローディングされたＨＬＡ－Ｅを認識することができ、ＣＭＶに対する防御において役割を果たすことも報告されている。いくつかの試験により、感染性疾患およびがんにおけるＨＬＡ－Ｅのいくつかの重要な機能が明らかになっている。

【０４４０】

ペプチド抗原は、それ自体が小胞体内ＨＬＡクラスＩの分子に競合的な親和性結合によって付着した後、細胞表面上に提示される。ここで、個々のペプチド抗原の親和性は、そのアミノ酸配列およびアミノ酸配列内の規定された位置における特定の結合性モチーフの存在に直接結びつけられる。そのようなペプチドの配列が既知である場合、疾患細胞に対する免疫系を、例えばペプチドワクチンを使用して操作することが可能である。

【０４４１】

ＭＨＣ分子は、高度に多型的である、すなわち、多くのＭＨＣバリアントが存在する。各バリアントは、タンパク質をコードする遺伝子のバリエーションによってコードされ、そのようなバリアント遺伝子の各々は対立遺伝子と称される。ヒトに関しては、ＭＨＣはヒト白血球抗原（ＨＬＡ）として公知であり、３つの型のＨＬＡクラスＩＩ分子：ＤＰ、ＤＱおよびＤＲを伴う。ＨＬＡクラスＩＩペプチド（図１）は、２つの鎖、αおよびβを有し、各々が２つのドメイン－α１およびα２ならびにβ１およびβ２を有し、各鎖が、膜貫通ドメイン、それぞれα２およびβ２を有し、それによりＨＬＡクラスＩＩ分子と細胞膜が繋がっている。α１とβ１のヘテロ二量体からペプチド結合性溝が形成される。最も広く研究されているＨＬＡ－ＤＲ分子は、それぞれαドメインおよびβドメインに対応するＤＲＡおよびＤＲＢを有する。ＤＲＢは多様であり、ＤＲＡはほぼ同一である。したがって、ＤＲＢ対立遺伝子の結合特異性により、対応するＨＬＡ－ＤＲの結合特異性が示される。各ＭＨＣタンパク質はそれ自体の結合特異性を有する、つまり、ＭＨＣ分子に対するペプチド結合性のセットは、別のＭＨＣ分子に対するものとは異なり得る。古典的分子は、ペプチドをＣＤ４＋リンパ球に提示する。細胞内機能を有する付属因子である非古典的分子は、細胞膜上には露出せず、リソソームの膜内部にあり、通常、抗原性ペプチドを古典的ＨＬＡクラスＩＩ分子にローディングする。

【０４４２】

ＨＬＡクラスＩＩ系では、マクロファージおよび未成熟樹状細胞などの食細胞が実体をファゴサイトーシスによってファゴソームに取り込み（しかしＢ細胞は、エンドソームへのより一般的なエンドサイトーシスを示す）、それらはリソソームと融合し、リソソームの酸性酵素により取り込まれたタンパク質が切断されて多くの異なるペプチドになる。オートファジーは、ＨＬＡクラスＩＩペプチドの別の供給源である。宿主のゲノムにコードされる宿主で生じるＨＬＡクラスＩＩバリアントとの分子間相互作用における物理化学ダイナミクスにより、特定のペプチドは、免疫優性を示し、ＨＬＡクラスＩＩ分子にローディングされる。これらは、細胞表面に輸送され、外面化する。最も多く研究されているＨＬＡクラスＩＩ遺伝子のサブクラスは、ＨＬＡ－ＤＰＡ１、ＨＬＡ－ＤＰＢ１、ＨＬＡ－ＤＱＡ１、ＨＬＡ－ＤＱＢ１、ＨＬＡ－ＤＲＡ、およびＨＬＡ－ＤＲＢ１である。

【０４４３】
ＨＬＡクラスＩＩ分子によるＣＤ４＋ヘルパーＴ細胞へのペプチドの提示は、外来抗原に対する免疫応答に必要である（Roche and Furuta, 2015）。ＣＤ４＋ Ｔ細胞は、活性化されたら、Ｂ細胞分化および抗体産生、ならびにＣＤ８＋ Ｔ細胞（ＣＴＬ）応答を促進する。ＣＤ４＋ Ｔ細胞はまた、他の免疫細胞を活性化し、その分化を誘導するサイトカインおよびケモカインも分泌する。ＨＬＡクラスＩＩ分子は、相互作用してＨＬＡクラスＩペプチド結合性溝よりも大きく開いたペプチド結合性溝を形成するα鎖とβ鎖のヘテロ二量体である（Unanue et al., 2016）。ＨＬＡクラスＩＩに結合したペプチド分子は、Ｎ末端側またはＣ末端側のいずれかに、溝から突出した隣接残基を伴う９－アミノ酸結合性コアを有すると考えられている（Jardetzky et al., 1996；Stern et al., 1994）。これらのペプチドは、通常、１２～１６アミノ酸の長さであり、多くの場合、結合レジスターのＰ１位、Ｐ４位、Ｐ６／７位およびＰ９位にアンカー残基を３～４個含有する（Rossjohn et al., 2015）。

【０４４４】
ＨＬＡ対立遺伝子は、共優性様式で発現される、つまり、両親から遺伝した対立遺伝子（バリアント）が同等に発現される。例えば、人は各々が、３つのクラスＩ遺伝子（ＨＬＡ－Ａ、ＨＬＡ－ＢおよびＨＬＡ－Ｃ）の各々の対立遺伝子を２つ有し、したがって、６つの異なる型のＨＬＡクラスＩＩを発現し得る。ＨＬＡクラスＩＩ遺伝子座では、人は各々がＨＬＡ－ＤＰ遺伝子の対（ＤＰＡ１およびＤＰＢ１、α鎖およびβ鎖をコードする）、ＨＬＡ－ＤＱ（ＤＱＡ１およびＤＱＢ１、α鎖およびβ鎖について）、１つの遺伝子ＨＬＡ－ＤＲα（ＤＲＡ１）、および１つまたは複数の遺伝子ＨＬＡ－ＤＲβ（ＤＲＢ１およびＤＲＢ３、ＤＲＢ４またはＤＲＢ５）を受け継ぐ。例えば、ＨＬＡ－ＤＲＢ１は、ほぼ４００種よりも多くの公知の対立遺伝子を有する。これは、１つのヘテロ接合性個体が６種または８種の機能性ＨＬＡクラスＩＩ対立遺伝子：各親から３種またはそれよりも多くを受け継ぎ得ることを意味する。したがって、ＨＬＡ遺伝子は、高度に多型的である；多くの異なる対立遺伝子が異なる集団内の個体に存在する。ＨＬＡタンパク質をコードする遺伝子は、多くの可能性のあるバリエーションを有し、それにより、各人の免疫系が広範囲の外来インベーダーと反応することが可能になる。いくつかのＨＬＡ遺伝子は、数百の同定されたバージョン（対立遺伝子）を有し、その各々に特定の数字が与えられる。一部の実施形態では、ＨＬＡクラスＩ対立遺伝子は、ＨＬＡ－Ａ＊０２：０１、ＨＬＡ－Ｂ＊１４：０２、ＨＬＡ－Ａ＊２３：０１、ＨＬＡ－Ｅ＊０１：０１（非古典的）である。一部の実施形態では、ＨＬＡクラスＩＩ対立遺伝子は、ＨＬＡ－ＤＲＢ＊０１：０１、ＨＬＡ－ＤＲＢ＊０１：０２、ＨＬＡ－ＤＲＢ＊１１：０１、ＨＬＡ－ＤＲＢ＊１５：０１、およびＨＬＡ－ＤＲＢ＊０７：０１である。

【０４４５】
対象の対象特異的ＨＬＡ対立遺伝子またはＨＬＡ遺伝子型は、当技術分野で公知の任意の方法によって決定することができる。例示的な実施形態では、ＨＬＡ遺伝子型を、その全体が参照により本明細書に組み込まれる、２０１５年６月１１日に公開されたＷ０２０１５０８５１４７としての国際特許出願第ＰＣＴ／ＵＳ２０１４／０６８７４６号に記載の任意の方法によって決定する。簡単に述べると、方法は、多型遺伝子型を決定することを含み、これは、配列決定データセットから抽出されたリードの、多型遺伝子の対立遺伝子バリアントを含む遺伝子参照セットに対するアラインメントを生成するステップと、アラインメント内の各対立遺伝子バリアントについて第１の事後確率または事後確率由来スコアを決定するステップと、最大の第１の事後確率または事後確率由来スコアを有する対立遺伝子バリアントを第１の対立遺伝子バリアントとして同定するステップと、第１の対立遺伝子バリアントおよび１つまたは複数の他の対立遺伝子バリアントとアラインメントされる１つまたは複数の重複リードを同定するステップと、重み付け因子を使用して１つまたは複数の他の対立遺伝子バリアントについて第２の事後確率または事後確率由来スコアを決定するステップと、最大の第２の事後確率または事後確率由来スコアを有する対立遺伝子バリアントを選択することによって第２の対立遺伝子バリアントを同定するステッ

プであって、第１の対立遺伝子バリアントおよび第２の対立遺伝子バリアントにより多型遺伝子の遺伝子型が定義される、ステップと、第１の対立遺伝子バリアントおよび第２の対立遺伝子バリアントの出力をもたらすステップとを含み得る。

【０４４６】

一部の実施形態では、本明細書に記載のＭＨＣクラスＩＩペプチド：抗原性ペプチド結合および提示予測方法は、個々のＨＬＡ対立遺伝子によってコードされるＭＨＣクラスＩＩペプチドの大きなレパートリーから結合体を予測することができるものである。一部の実施形態では、ＭＡＰＴＡＣ技術を、質量分析により検証されたＨＬＡ適合ペプチドの大きなデータベースを用いて訓練する。一部の実施形態では、質量分析により検証されたＨＬＡ適合ペプチドの大きなデータベースは、１．２×１０^{6}種よりも多くのそのようなＨＬＡ適合ペプチドを含む。一部の実施形態では、質量分析により検証されたＨＬＡ適合ペプチドの大きなデータベースは、ＭＨＣクラスＩおよびクラスＩＩ対立遺伝子亜型の両方を含む１５０種よりも多くのＨＬＡ対立遺伝子を包含する。一部の実施形態では、データベースは、ＨＬＡ－ＩおよびＨＬＡ－ＩＩ（ＤＲ亜型）のＵＳ集団の少なくとも９５％を包含する。

【０４４７】

本明細書に記載の通り、変異エピトープが誘導免疫応答において有効であること、および自発性腫瘍退縮または長期生存の場合が、変異エピトープに対するＣＤ８＋　Ｔ細胞応答と相関すること、および「免疫編集」を追跡して、マウスおよび人間における優性の変異抗原の発現を変更することができることのエビデンスが動物およびヒトの両方において数多く存在する。

【０４４８】

配列決定技術により、各腫瘍が、遺伝子のタンパク質コード内容を変更する患者特異的変異を多数含有することが明らかになった。そのような変異により、単一のアミノ酸の変化（ミスセンス変異によって引き起こされる）から、フレームシフト、終結コドンのリードスルーまたはイントロン領域の翻訳（新規のオープンリーディングフレーム変異；ｎｅｏＯＲＦ）に起因した新規アミノ酸配列の長い領域の付加までにわたって変更されたタンパク質が創出される。これらの変異タンパク質は、ネイティブなタンパク質とは異なり、自己寛容の免疫を弱める効果を受けないので、腫瘍に対する宿主の免疫応答の有益な標的である。したがって、変異タンパク質は、患者の正常細胞と比較して、免疫原性である可能性がより高く、また、腫瘍細胞に対してより特異的である。本質的に、がん関連変異を含有する短いペプチド（８～２４アミノ酸長）は、がん免疫治療の候補である。

【０４４９】

一部の実施形態では、ペプチドに対する変異コールのために、予測方法を駆動するアルゴリズムをさらに利用することができる。一部の実施形態では、予測方法を、ペプチド内のドライバー変異の状態、および／またはＲＮＡ発現の状態、および／または切断予測を決定するために使用することができる。

【０４５０】

「Ｔ細胞」という用語は、ＣＤ４＋　Ｔ細胞およびＣＤ８＋　Ｔ細胞を含む。Ｔ細胞という用語はまた、Ｔヘルパー１型Ｔ細胞およびＴヘルパー２型Ｔ細胞の両方を含む。Ｔ細胞は、本明細書で使用される場合、一般に、機能および細胞表面抗原（クラスター分類抗原、またはＣＤ）によって分類され、Ｔ細胞はまた、抗原に対するＴ細胞受容体の結合を容易にし、２種の主要なクラス：ヘルパーＴ（ＴＨ）細胞および細胞傷害性Ｔリンパ球（ＣＴＬ）に分類される。

【０４５１】

成熟ヘルパーＴ（ＴＨ）細胞は、表面タンパク質ＣＤ４を発現し、ＣＤ４＋　Ｔ細胞と称される。Ｔ細胞の発生後、成熟したナイーブなＴ細胞は胸腺を離れ、リンパ節を含めた体の中に拡散し始める。ナイーブＴ細胞は、それらが応答するようにプログラムされた抗原に曝露したことがないＴ細胞である。全てのＴ細胞と同様に、ナイーブＴ細胞は、Ｔ細胞受容体－ＣＤ３複合体を発現する。Ｔ細胞受容体（ＴＣＲ）は、定常領域および可変領

域の両方からなる。可変領域により、いずれの抗原にＴ細胞が応答しうるかが決定される。ＣＤ４＋　Ｔ細胞は、ＭＨＣクラスＩＩに対する親和性を有するＴＣＲを有し、タンパク質およびＣＤ４が胸腺における成熟化の間のＭＨＣ親和性の決定に関与する。ＭＨＣクラスＩＩタンパク質は、一般に、特殊化された抗原提示細胞（ＡＰＣ）の表面にしか見いだされない。特殊化された抗原提示細胞（ＡＰＣ）は、主に、樹状細胞、マクロファージおよびＢ細胞であるが、樹状細胞は、ＭＨＣクラスＩＩを構成的に（いつでも）発現する唯一の細胞群である。濾胞樹状細胞などの一部のＡＰＣの表面にはネイティブな（またはプロセスされていない）抗原も結合するが、プロセシングされていない抗原はＴ細胞と相互作用せず、Ｔ細胞の活性化に関与しない。ＨＬＡクラスＩタンパク質に結合するペプチド抗原は、一般には、ＨＬＡクラスＩＩタンパク質に結合するペプチド抗原よりも短い。

【０４５２】

細胞傷害性Ｔリンパ球（ＣＴＬ）は、細胞傷害性Ｔ細胞、細胞溶解性Ｔ細胞、ＣＤ８＋Ｔ細胞、またはキラーＴ細胞としても公知であり、標的化された細胞においてアポトーシスを誘導するリンパ球を指す。ＣＴＬは、標的細胞と、ＴＣＲと標的細胞表面上のプロセシングされた抗原（Ａｇ）の相互作用を介して抗原特異的コンジュゲートを形成し、その結果、標的化された細胞のアポトーシスがもたらされる。アポトーシス小体はマクロファージによって排除される。「ＣＴＬ応答」という用語は、ＣＴＬ細胞によって媒介される一次免疫応答を指すために使用される。細胞傷害性Ｔリンパ球は、それらの表面上にＴ細胞受容体（ＴＣＲ）およびＣＤ８分子の両方を有する。Ｔ細胞受容体は、ＨＬＡクラスＩの分子と複合体を形成したペプチドを認識し、それに結合することができる。各細胞傷害性Ｔリンパ球は、特異的なＭＨＣ／ペプチド複合体に結合することができる独特のＴ細胞受容体を発現する。大多数の細胞傷害性Ｔ細胞は、特異的な抗原を認識することができるＴ細胞受容体（ＴＣＲ）を発現する。ＴＣＲがＨＬＡクラスＩ分子と結合するためには、ＨＬＡクラスＩ分子の定常部分に結合するＣＤ８と称される糖タンパク質がＴＣＲに付随していなければならない。したがって、これらのＴ細胞は、ＣＤ８＋　Ｔ細胞と称される。ＣＤ８とＭＨＣ分子の親和性により、抗原特異的活性化の間、Ｔ細胞と標的細胞が密接に結合した状態に保たれる。ＣＤ８＋　Ｔ細胞は、活性化されたらＴ細胞として認識される、一般に、免疫系内で所定の細胞傷害性役割を有するものと分類される。しかし、ＣＤ８＋　Ｔ細胞はまた、いくつかのサイトカインを作り出す能力も有する。

【０４５３】

「Ｔ細胞受容体（ＴＣＲ）」は、抗原の提示に応答したＴ細胞の活性化に関与する細胞表面受容体である。ＴＣＲは、一般に、２つの鎖、アルファおよびベータで構成され、これらがアセンブルしてヘテロ二量体を形成し、ＣＤ３形質導入サブユニットと会合して、細胞表面に存在するＴ細胞受容体複合体を形成する。ＴＣＲの各アルファ鎖およびベータ鎖は、免疫グロブリン様Ｎ末端可変（Ｖ）領域および定常（Ｃ）領域、疎水性膜貫通ドメイン、および短い細胞質内領域からなる。免疫グロブリン分子に関しては、アルファ鎖およびベータ鎖の可変領域はＶ（Ｄ）Ｊ組換えによって生成され、Ｔ細胞の集団内で抗原特異性の大きな多様性が創出される。しかし、インタクトな抗原を認識する免疫グロブリンとは対照的に、Ｔ細胞は、ＭＨＣ分子に付随するプロセシングされたペプチド断片によって活性化され、Ｔ細胞による抗原認識に対して、ＭＨＣ拘束性として公知の余分の要素が導入される。Ｔ細胞受容体を通じたドナーとレシピエントの間のＭＨＣ不一致の認識により、Ｔ細胞の増殖およびＧＶＨＤの潜在的な発生が導かれる。ＴＣＲの正常な表面発現は、複合体の７つの構成成分の全ての調和した合成およびアセンブリに依存することが示されている（Ashwell and Klusner 1990）。ＴＣＲαまたはＴＣＲβの不活化の結果、Ｔ細胞の表面からのＴＣＲの消失がもたらされ得、それにより、アロ抗原の認識、したがってＧＶＨＤが防止される。しかし、ＴＣＲ破壊の結果、一般に、ＣＤ３シグナル伝達構成成分の消失がもたらされ、さらなるＴ細胞増大の手段が変更される。

【０４５４】

「ＨＬＡペプチドーム」という用語は、特定のＨＬＡクラスと特異的に相互作用し、何千もの異なる配列を包含し得るペプチドのプールを指す。ＨＬＡペプチドームは、細胞に

おいて発現される正常なタンパク質および異常なタンパク質の両方に由来するペプチドの多様性を含む。したがって、ＨＬＡペプチドームを、がん特異的ペプチドを同定するため、腫瘍免疫治療薬の開発のため、ならびにがん細胞内のタンパク質合成および分解スキームに関する情報源として研究することができる。一部の実施形態では、ＨＬＡペプチドームは、可溶性ＨＬＡペプチド（ｓＨＬＡ）のプールである。一部の実施形態では、ＨＬＡペプチドームは、膜結合ＨＬＡ（ｍＨＬＡ）のプールである。

【０４５５】

「抗原提示細胞」または「ＡＰＣ」は、プロフェッショナル抗原提示細胞（例えば、Ｂリンパ球、マクロファージ、単球、樹状細胞、ランゲルハンス細胞）、ならびに他の抗原提示細胞（例えば、ケラチノサイト、内皮細胞、アストロサイト、線維芽細胞、オリゴデンドロサイト、胸腺上皮細胞、甲状腺上皮細胞、グリア細胞（脳）、膵ベータ細胞、および血管内皮細胞）を含む。「抗原提示細胞」または「ＡＰＣ」は、主要組織適合性遺伝子複合体（ＭＨＣ）分子を発現し、ＭＨＣと複合体を形成した外来抗原をその表面上に提示することができる細胞である。

【０４５６】

単一対立遺伝子ＨＬＡ細胞株

単一のＨＬＡクラスＩ対立遺伝子、ＨＬＡクラスＩＩ対立遺伝子の単一の対、または単一のＨＬＡクラスＩ対立遺伝子およびＨＬＡクラスＩＩ対立遺伝子の単一の対を発現する単一対立遺伝子細胞株を、適切な細胞集団に単一のＨＬＡ対立遺伝子をコードするポリ核酸、例えばベクターを形質導入またはトランスフェクトすることによって生成することができる（図２）。適切な細胞集団としては、例えば、単一のＨＬＡクラスＩ対立遺伝子を外因的に発現するＨＬＡクラスＩ欠損細胞株、ＨＬＡクラスＩＩ対立遺伝子の単一の外因性対を発現するＨＬＡクラスＩＩ欠損細胞株、または単一のＨＬＡクラスＩおよび／またはクラスＩＩ対立遺伝子の単一の対を外因的に発現するクラスＩおよびクラスＩＩ欠損細胞株が挙げられる。例示的な実施形態として、ＨＬＡクラスＩ欠損Ｂ細胞株は、Ｂ７２１．２２１である。しかし、ＨＬＡクラスＩおよび／またはＨＬＡクラスＩＩが欠損した他の細胞集団を生成することができることが当業者には明らかである。内因性ＨＬＡクラスＩまたはＨＬＡクラスＩＩ遺伝子を欠失させる／不活化するための典型的な方法としては、例えばＴＨＰ－１細胞におけるＣＲＩＳＰＲ－Ｃａｓ９媒介性ゲノム編集が挙げられる。一部の実施形態では、細胞の集団は、マクロファージ、Ｂ細胞、および樹状細胞などのプロフェッショナル抗原提示細胞である。細胞は、Ｂ細胞または樹状細胞であり得る。一部の実施形態では、細胞は、腫瘍細胞または腫瘍細胞株に由来する細胞である。一部の実施形態では、細胞は、患者から単離されたものである。一部の実施形態では、細胞は、感染因子またはその一部を含有する。一部の実施形態では、細胞の集団は、少なくとも１０７個の細胞を含む。一部の実施形態では、細胞の集団を、例えば、少なくとも１つの遺伝子の発現および／または活性を増大させるまたは低減することによってさらに改変する。一部の実施形態では、遺伝子は、免疫プロテアソームのメンバーをコードする。免疫プロテアソームは、ＨＬＡクラスＩ結合性ペプチドのプロセシングに関与し、ＬＭＰ２（β１ｉ）、ＭＥＣＬ－１（β２ｉ）、およびＬＭＰ７（β５ｉ）サブユニットを含むことが公知である。免疫プロテアソームはまた、インターフェロン－ガンマによって誘導することができる。したがって、一部の実施形態では、細胞の集団を、１つまたは複数のサイトカイン、増殖因子、または他のタンパク質と接触させることができる。細胞を、インターフェロン－ガンマ、ＩＬ－１０、ＩＬ－６、および／またはＴＮＦ－αなどの炎症性サイトカインで刺激することができる。細胞の集団を、ストレス（熱ストレス、酸素欠乏、グルコース飢餓性、ＤＮＡ損傷剤など）などの種々の環境条件に供することもできる。一部の実施形態では、細胞を、化学療法薬、放射線、標的化治療、または免疫治療のうちの１つまたは複数と接触させる。したがって、本明細書に開示される方法を使用して、ＨＬＡペプチドのプロセシングおよび提示に対する種々の遺伝子または条件の影響を試験することができる。一部の実施形態では、使用する条件を、ＨＬＡ－ペプチドの集団を同定する患者の状態に適合するように選択する。

【０４５７】
本開示の単一のＨＬＡ－対立遺伝子は、ウイルスに基づく系（例えば、アデノウイルス系、アデノ随伴ウイルス（ＡＡＶ）ベクター、ポックスウイルス、またはレンチウイルス）を使用してコードし、発現させることができる。アデノ随伴ウイルスによる送達、アデノウイルスによる送達、およびレンチウイルスによる送達に使用することができるプラスミドは以前に記載されている（例えば、参照により本明細書に組み込まれる米国特許第６，９５５，８０８号および同第６，９４３，０１９号、および米国特許出願第２００８０２５４００８号を参照されたい）。本開示の実施に使用することができるベクターの中で、細胞の宿主ゲノムへの組込みは、レトロウイルス遺伝子移入法を用いて可能であり、多くの場合、挿入された導入遺伝子の長期発現がもたらされる。例示的な実施形態では、レトロウイルスは、レンチウイルスである。さらに、多くの異なる細胞型および標的組織において高い形質導入効率が観察されている。レトロウイルスのトロピズムは、外来エンベロープタンパク質を組み入れ、標的細胞の潜在的な標的集団を増大させることによって変更することができる。レトロウイルスを工学的に操作して、レンチウイルスがある特定の細胞型のみに感染するように、挿入された導入遺伝子の条件付き発現を可能にすることができる。細胞型特異的プロモーターを使用して、特定の細胞型における発現を標的とすることができる。レンチウイルスベクターは、レトロウイルスベクターである（したがって、レンチウイルスベクターおよびレトロウイルスベクターのどちらも本開示の実施に使用することができる）。さらに、レンチウイルスベクターは、非分裂細胞に形質導入または感染することができ、また、一般には高いウイルス力価を生じる。

【０４５８】
レトロウイルスによる遺伝子移入系の選択は、標的組織に依存し得る。レトロウイルスベクターは、最大６～１０ｋｂの外来配列をパッケージングする能力を有するシス作用性の長い末端リピートで構成される。最小のシス作用性ＬＴＲは、複製およびベクターのパッケージングに十分なものであり、したがって、恒久的な発現をもたらすために所望の核酸を標的細胞に組み込むために使用する。本開示の実施に使用することができる、広く使用されているレトロウイルスベクターとしては、マウス白血病ウイルス（ＭｕＬＶ）に基づくもの、テナガザル白血病ウイルス（ＧａＬＶ）に基づくもの、サル免疫不全ウイルス（ＳＩＶ）に基づくもの、ヒト免疫不全ウイルス（ＨＩＶ）に基づくもの、およびこれらの組合せが挙げられる（例えば、Buchscher et al., (1992) J. Virol. 66: 2731-2739；Johann et al., (1992) J. Virol. 66: 1635-1640；Sommnerfelt et al., (1990) Virol. 176: 58-59；Wilson et al., (1998) J. Virol. 63: 2374-2378；Miller et al., (1991) J. Virol. 65: 2220-2224；ＰＣＴ／ＵＳ９４／０５７００を参照されたい）。また、最小の非霊長類レンチウイルスベクター、例えば、ウマ伝染性貧血ウイルス（ＥＩＡＶ）に基づくレンチウイルスベクターが本開示の実施に有用である（例えば、Balagaan, (2006) J Gene Med; 8: 275-285, Published online 21 November 2005 in Wiley InterScience DOI: 10.1002/jgm.845を参照されたい）。ベクターは、標的遺伝子の発現を駆動するサイトメガロウイルス（ＣＭＶ）プロモーターを有し得る。したがって、本開示は、本開示の実施に有用なベクターの中でも、レトロウイルスベクターおよびレンチウイルスベクターを含めたウイルスベクターを意図している。

【０４５９】
任意のＨＬＡ対立遺伝子を細胞集団において発現させることができる。例示的な実施形態では、ＨＬＡ対立遺伝子は、ＨＬＡクラスＩ対立遺伝子である。一部の実施形態では、ＨＬＡクラスＩ対立遺伝子は、ＨＬＡ－Ａ対立遺伝子またはＨＬＡ－Ｂ対立遺伝子である。一部の実施形態では、ＨＬＡ対立遺伝子は、ＨＬＡクラスＩＩ対立遺伝子である。ＨＬＡクラスＩおよびクラスＩＩ対立遺伝子の配列は、ＩＰＤ－ＩＭＧＴ／ＨＬＡデータベースにおいて見いだすことができる。例示的なＨＬＡ対立遺伝子としては、これだけに限定されないが、ＨＬＡ－Ａ＊０２：０１、ＨＬＡ－Ｂ＊１４：０２、ＨＬＡ－Ａ＊２３：０１、ＨＬＡ－Ｅ＊０１：０１、ＨＬＡ－ＤＲＢ＊０１：０１、ＨＬＡ－ＤＲＢ＊０１：０２、

ＨＬＡ－ＤＲＢ＊１１：０１、ＨＬＡ－ＤＲＢ＊１５：０１、およびＨＬＡ－ＤＲＢ＊０７：０１が挙げられる。

【０４６０】

一部の実施形態では、ＨＬＡ対立遺伝子を、目的の遺伝子型に対応するように選択する。一部の実施形態では、ＨＬＡ対立遺伝子は、変異ＨＬＡ対立遺伝子であり、罹患患者において天然に存在しない対立遺伝子または天然に存在する対立遺伝子であり得る。本明細書に開示される方法には、種々の障害に関連するＨＬＡ対立遺伝子ならびに低頻度で存在する対立遺伝子に対するＨＬＡ結合性ペプチドの同定という別の利点がある。したがって、一部の実施形態では、本明細書で提供される方法により、ＨＬＡ対立遺伝子を、それが集団内、例えばコーカサス人集団内で１％未満の頻度で存在する場合であっても同定することができる。

【０４６１】

一部の実施形態では、ＨＬＡ対立遺伝子をコードする核酸配列は、ＨＬＡ－タンパク質を免疫精製するために使用することができる親和性アクセプタータグをさらに含む。適切なタグは当技術分野で周知である。一部の実施形態では、親和性アクセプタータグは、ポリ－ヒスチジンタグ、ポリ－ヒスチジン－グリシンタグ、ポリ－アルギニンタグ、ポリ－アスパラギン酸タグ、ポリ－システインタグ、ポリ－フェニルアラニン、ｃ－ｍｙｃタグ、単純ヘルペスウイルス糖タンパク質Ｄ（ｇＤ）タグ、ＦＬＡＧタグ、ＫＴ３エピトープタグ、チューブリンエピトープタグ、Ｔ７遺伝子１０タンパク質ペプチドタグ、ストレプトアビジンタグ、ストレプトアビジン結合性ペプチド（ＳＰＢ）タグ、Ｓｔｒｅｐ－タグ、Ｓｔｒｅｐ－タグＩＩ、アルブミン結合性タンパク質（ＡＢＰ）タグ、アルカリホスファターゼ（ＡＰ）タグ、ブルータングウイルスタグ（Ｂ－タグ）、カルモジュリン結合性ペプチド（ＣＢＰ）タグ、クロラムフェニコール　アセチルトランスフェラーゼ（ＣＡＴ）タグ、コリン結合性ドメイン（ＣＢＤ）タグ、キチン結合性ドメイン（ＣＢＤ）タグ、セルロース結合性ドメイン（ＣＢＰ）タグ、ジヒドロ葉酸レダクターゼ（ＤＨＦＲ）タグ、ガラクトース結合性タンパク質（ＧＢＰ）タグ、マルトース結合性タンパク質（ＭＢＰ）、グルタチオン－Ｓ－トランスフェラーゼ（ＧＳＴ）、Ｇｌｕ－Ｇｌｕ（ＥＥ）タグ、ヒトインフルエンザ赤血球凝集素（ＨＡ）タグ、西洋ワサビペルオキシダーゼ（ＨＲＰ）タグ、ＮＥ－タグ、ＨＳＶタグ、ケトステロイドイソメラーゼ（ＫＳＩ）タグ、ＫＴ３タグ、ＬａｃＺタグ、ルシフェラーゼタグ、ＮｕｓＡタグ、ＰＤＺドメインタグ、ＡｖｉＴａｇ、カルモジュリン－タグ、Ｅ－タグ、Ｓ－タグ、ＳＢＰ－タグ、Ｓｏｆｔａｇ１、Ｓｏｆｔａｇ３、ＴＣタグ、ＶＳＶ－タグ、Ｘｐｒｅｓｓタグ、Ｉｓｏｐｅｐｔａｇ、ＳｐｙＴａｇ、ＳｎｏｏｐＴａｇ、Ｐｒｏｆｉｎｉｔｙ　ｅＸａｃｔタグ、プロテインＣタグ、Ｓ１－タグ、Ｓ－タグ、ビオチン－カルボキシ担体タンパク質（ＢＣＣＰ）タグ、緑色蛍光タンパク質（ＧＦＰ）タグ、小さなユビキチン様修飾因子（ＳＵＭＯ）タグ、タンデムアフィニティー精製（ＴＡＰ）タグ、ＨａｌｏＴａｇ、Ｎｕｓ－タグ、チオレドキシン－タグ、Ｆｃ－タグ、ＣＹＤタグ、ＨＰＣタグ、ＴｒｐＥタグ、ユビキチンタグ、Ｖｅｓｃｕｌａｒ　Ｓｔｏｍａｔｉｓウイルス糖タンパク質に由来するＶＳＶ－Ｇエピトープタグ、またはサルウイルス５（ＳＶ５）のパラミクソウイルスのＰタンパク質およびＶタンパク質において見いだされる小さなエピトープ（Ｐｋ）に由来するＶ５タグである。一部の実施形態では、親和性アクセプタータグは、「エピトープタグ」であり、これは、認識可能なエピトープ（抗体結合性部位）をＨＬＡ－タンパク質に付加して、対応する抗体への結合をもたらし、それにより、タグ付けされたタンパク質の同定またはアフィニティー精製を可能にする、ペプチドタグの一種である。エピトープタグの非限定的な例は、ＩｇＧに結合するプロテインＡまたはプロテインＧである。一部の実施形態では、親和性アクセプタータグとしては、ビオチンアクセプターペプチド（ＢＡＰ）またはヒトインフルエンザ赤血球凝集素（ＨＡ）ペプチド配列が挙げられる。多数の他のタグ部分が当業者には公知であり、当業者が構想することができ、また、本明細書で意図されている。親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体のエレメントとして発現させることができる限りは、任意のペプチドタグを使用することができる。

【０４６２】
本明細書で提供される方法は、ＨＬＡ構築物の親和性プルダウンをトランスフェクトまたは形質導入した細胞からＨＬＡ－ペプチド複合体を単離することを含む（図３）。一部の実施形態では、複合体を、市販の抗体を用い、当技術分野で公知の標準の免疫沈降技法を使用して単離することができる。細胞をまず溶解させることができる。ＨＬＡクラスＩ－ペプチド複合体は、Ｗ６／３２抗体などのＨＬＡクラスＩ特異的抗体を使用して単離することができ、一方、ＨＬＡクラスＩＩ－ペプチド複合体は、Ｍ５／１１４．１５．２モノクローナル抗体などのＨＬＡクラスＩＩ特異的抗体を使用して単離することができる。一部の実施形態では、単一の（またはその対）ＨＬＡ対立遺伝子をペプチドタグとの融合タンパク質として発現させ、ＨＬＡ－ペプチド複合体を、ペプチドタグを認識する結合性分子を使用して単離する。

【０４６３】
方法は、前記ＨＬＡ－ペプチド複合体からペプチドを単離し、ペプチドの配列決定を行うことをさらに含む。酸溶出などの当業者に公知の任意の方法によって複合体からペプチドを単離する。任意の配列決定方法を使用することができるが、一部の実施形態では、質量分析を使用する方法、例えば、液体クロマトグラフィー－質量分析（ＬＣ－ＭＳまたはＬＣ－ＭＳ／ＭＳ、またはその代わりにＨＰＬＣ－ＭＳまたはＨＰＬＣ－ＭＳ／ＭＳ）などが利用される。これらの配列決定方法は当業者に周知であり、Medzihradszky KF and Chalkley RJ. Mass Spectrom Rev. 2015 Jan-Feb; 34 (1): 43-63に概説されている。

【０４６４】
一部の実施形態では、細胞の集団は、１つまたは複数の内因性ＨＬＡ対立遺伝子を発現する。一部の実施形態では、細胞の集団は、工学的に操作された１つまたは複数の内因性ＨＬＡクラスＩ対立遺伝子を欠く細胞の集団である。一部の実施形態では、細胞の集団は、工学的に操作された内因性ＨＬＡクラスＩ対立遺伝子を欠く細胞の集団である。一部の実施形態では、細胞の集団は、工学的に操作された１つまたは複数の内因性ＨＬＡクラスＩＩ対立遺伝子を欠く細胞の集団である。一部の実施形態では、細胞の集団は、工学的に操作された内因性ＨＬＡクラスＩＩ対立遺伝子を欠く細胞の集団または工学的に操作された内因性ＨＬＡクラスＩ対立遺伝子および内因性ＨＬＡクラスＩＩ対立遺伝子を欠く細胞の集団である。一部の実施形態では、細胞の集団は、例えば、蛍光活性化細胞選別（ＦＡＣＳ）によって富化または選別された細胞を含む。一部の実施形態では、蛍光活性化細胞選別（ＦＡＣＳ）を使用して、細胞の集団を選別する。一部の実施形態では、細胞の集団を、ＨＬＡクラスＩまたはクラスＩＩのいずれかまたはＨＬＡクラスＩおよびクラスＩＩの両方の細胞表面発現について予めＦＡＣＳにより選別する。例えば、ＦＡＣＳを使用して、細胞の集団をＨＬＡクラスＩ対立遺伝子、ＨＬＡクラスＩＩ対立遺伝子、またはこれらの組合せの細胞表面発現について選別することができる。

【０４６５】
個別化がんワクチンを調製するための方法
がん特異的変異、したがって、同じヒト対象のがん細胞にはＤＮＡに変異が存在するが正常細胞にはその変異が存在しないこと、および変異によりＤＮＡによってコードされるタンパク質の１つまたは複数のアミノ酸の変化が導かれることが同定されたら、その変異を宿主免疫応答の標的とすることができる。天然の免疫応答を変異タンパク質に方向付けることができ、それにより、そのタンパク質を発現するがん細胞の破壊を導くことができる。がん性組織における天然の寛容性応答および免疫無防備状態の環境に起因して、免疫治療は、そのような免疫応答を強化して、体の寛容性および免疫抑制効果を無効にすることを試みる臨床経路である。したがって、上記の変異を含むタンパク質またはペプチドは免疫治療の適切な候補である。

【０４６６】
変異タンパク質は、抗原提示細胞（ＡＰＣ）として作用するプロフェッショナル食細胞によって摂取され、細かく切られ、Ｔ細胞活性化のための抗原として主要組織適合性遺伝

子複合体（ＭＨＣ）タンパク質を含む抗原提示複合体として細胞表面上にディスプレイされる。ヒトＭＨＣタンパク質は、ヒト白血球抗原、ＨＬＡと称される。ＭＨＣタンパク質は、ＭＨＣ－クラスＩまたはクラスＩＩタンパク質であり得、いくつかの機能的な違いはクラスＩまたはクラスＩＩ　ＭＨＣタンパク質（ＨＬＡクラスＩおよびＨＬＡクラスＩＩタンパク質）のいずれかによるペプチドの提示に起因するが、顕著な違いの１つは、ＨＬＡクラスＩ－ペプチド複合体は細胞傷害性ＣＤ８＋　Ｔ細胞に対して抗原を提示し、一方、ＨＬＡクラスＩＩペプチド複合体もＣＤ４＋　Ｔ細胞を活性化し持続的な免疫応答を導くことができるという事実にある。ＣＤ８＋　Ｔ細胞は、感染細胞または腫瘍細胞などの疾患細胞の細胞毎の消失という課題に不可欠である。ＣＤ４＋　Ｔ細胞は、活性化されるとより持続的な効果を有し、最も重要なのは免疫記憶の生成である。ＣＤ４サブセットは、免疫学的脅威の型に応じて示差的に動員され、重複するまたは異なる機能を有する多数のサブセットが共動員され得る。これは、病原性脅威に対する免疫学的応答の平衡を保つのに役立つ。これらに関して、ＨＬＡクラスＩＩペプチド媒介性抗原提示により、持続的な調整された免疫応答がもたらされる。他方では、ペプチドへのＨＬＡクラスＩＩの結合はプロミスキャスであり得、したがって、非特異的なペプチド結合および免疫系への提示により、自己免疫などの異所性免疫応答が導かれる。

【０４６７】

一態様では、本開示は、ペプチドのＨＬＡクラスＩＩタンパク質（アルファ鎖とベータ鎖のヘテロ二量体で構成される）への高忠実度の結合により、特定のペプチドがＴリンパ球に提示され、それにより、特異的な免疫応答が引き出され、あらゆる交差反応性または無差別免疫が回避されることが確実になるような、特定のＨＬＡクラスＩＩアルファ鎖およびベータ鎖ヘテロ二量体と正確に対形成または結合し得るペプチドを予測するための方法を提供する。

【０４６８】

一態様では、本開示は、特定の同類ＨＬＡクラスＩＩタンパク質を発現する対象にペプチドを治療的に投与した場合に、ＣＤ４＋　Ｔ細胞を活性化し、免疫記憶を刺激するＨＬＡクラスＩＩタンパク質の能力により、ペプチドを用いてより持続的かつ頑強な免疫応答を活性化することができるような、特定のＨＬＡクラスＩＩタンパク質に正確に結合し得るペプチドを予測するための方法を提供する。一部の実施形態では、ＨＬＡクラスＩＩタンパク質に高い特異性で結合することが予測される所与のペプチドは、変異を含むペプチドであり、変異は、対象のがんまたは腫瘍細胞において蔓延しているものである；一方、変異ペプチドに結合することが予測される同じＨＬＡクラスＩＩタンパク質は、対応する変異していない野生型ペプチドに（ａ）結合しないか、または（ｂ）対象の変異ペプチドとの結合に対する親和性と比較して別個に低親和性で結合する。ＨＬＡの変異ペプチドに対する優先的な結合は、野生型ペプチドを発現する細胞は、ＨＬＡにより提示されたペプチドに対して反応性のＴ細胞による免疫攻撃から容赦されるので、免疫治療薬の開発に有利である。一部の実施形態では、ＨＬＡクラスＩＩタンパク質に特異的に結合することが予測されるペプチドは、翻訳後修飾を有するペプチドである。例示的な翻訳後修飾としては、これだけに限定されないが、リン酸化、ユビキチン化、脱リン酸化、グリコシル化、メチル化、またはアセチル化が挙げられる。一部の実施形態では、予測されるペプチドを、免疫治療に使用する前に翻訳後修飾に供する。

【０４６９】

一部の実施形態では、本明細書に開示される免疫治療法および戦略は、自己免疫反応におけるものなどの望ましくない免疫活性化の抑制にも適用可能であり得る。具体的には、特定のＨＬＡ亜型に対する潜在的な結合体として同定されたペプチドを、特定のＨＬＡ分子に結合し、免疫原性応答を引き起こすのではなく、寛容性を誘導するように調整することができる。

【０４７０】

一態様では、特定の対象に対して調整または個別化された免疫治療の方法が本明細書で提示される。あらゆる対象または患者がＨＬＡクラスＩおよびＨＬＡクラスＩＩタンパク

質の特定のアレイを発現する。ＨＬＡタイピングは、対象によって発現されるＨＬＡタンパク質の特定のレパートリーを決定することを可能にする周知の技法である。特定の対象によって発現されるＨＬＡヘテロ二量体が分かったら、特定のＨＬＡクラスＩＩアルファ鎖およびベータ鎖ヘテロ二量体に高忠実度で結合することができるペプチドを予測するための、本明細書に記載の、改善され洗練された信頼できる方法により、対象に対して特異的に調整された特異的な免疫応答を生じさせることができることを確実にする。

【０４７１】

ＨＬＡヘテロ二量体をコードする遺伝子は高度に多型的であり、ヒト集団にわたって４，０００種よりも多くのＨＬＡクラスＩＩ対立遺伝子バリアントが同定されている。母系および父系ＨＬＡハプロタイプから、個体は、ＨＬＡクラスＩＩ遺伝子座の各々について異なる対立遺伝子を受け継ぐことができ、各ＨＬＡクラスＩＩヘテロ二量体はα鎖およびβ鎖で構成される。特にＨＬＡ－ＤＰおよびＨＬＡ－ＤＱ対立遺伝子について、多数のαおよびβ鎖対形成の組合せに起因して、可能性のあるＨＬＡヘテロ二量体の集団は高度に複雑である。ＨＬＡクラスＩＩヘテロ二量体は、小胞体（ＥＲ）に変換され、タンパク質ＣＤ７４に由来するインバリアント鎖（Ｉｉ）と共にアセンブリされて安定な複合体になる。Ｉｉにより、適当なタンパク質フォールディングが可能になることによってクラスＩＩ複合体が安定化され、また、ＨＬＡクラスＩＩヘテロ二量体のエンドソーム／リソソーム区画への移出が可能になる。これらのＨＬＡクラスＩＩローディング区画内で、Ｉｉはカテプシンによってタンパク質分解により切断されてＣＬＩＰと称されるプレースホルダーペプチドになる。次いで、低ｐＨ環境において、非古典的ＨＬＡクラスＩＩヘテロ二量体であるシャペロンＨＬＡ－ＤＭによってＣＬＩＰがより親和性が高いペプチドに交換される。次いで、高親和性ペプチドがローディングされたＨＬＡクラスＩＩ複合体がトランス－ゴルジに向かい、最終的にＣＤ４＋　Ｔ細胞に対するディスプレイのために細胞表面に向かう。

【０４７２】

各ＨＬＡヘテロ二量体は、対立遺伝子特異的結合優先性を有する何千ものペプチドと結合すると推定される。実際、各ＨＬＡ対立遺伝子は、約１，０００～１０，０００種の独特のペプチドと結合し、それらをＴ細胞に提示すると推定される。そのようなＨＬＡ結合の多様性を考慮すると、ペプチドが特定のＨＬＡ対立遺伝子に結合する可能性があるかどうかの正確な予測は非常に困難である。α鎖とβ鎖の対形成の不均一性、コア結合性エピトープを確信的に割り当てる能力を限定するデータの複雑さ、および高分解能生化学的分析に必要な免疫沈降グレードの対立遺伝子特異的抗体の欠如に起因して、ＨＬＡクラスＩＩ分子の対立遺伝子特異的ペプチド結合性特徴に関してはほとんど分かっていない。さらに、所与のＨＬＡ対立遺伝子に由来するペプチドエピトープの分析では、多数のＨＬＡ対立遺伝子が細胞表面に提示される場合に不明確性が生じる。

【０４７３】

候補新抗原の予測は、主にＨＬＡクラスＩエピトープについて行われる（クラスＩ予測アルゴリズムについての実験データが、クラスＩＩと比べて利用可能なので）が、前臨床的および臨床的個別化新抗原ワクチン接種試験のどちらにおいても、多くの場合、ＣＤ４＋　Ｔ細胞応答が観察される。これらの知見から、ＨＬＡクラスＩＩエピトーププロセシングおよび提示もがん処置において極めて重要な役割を果たし得ることが実証される。ＨＬＡクラスＩＩ予測アルゴリズムは存在するが、ＨＬＡクラスＩＩヘテロ二量体のオープンエンドのペプチド結合性溝により、より長いペプチド（一般に、１５～４０アミノ酸）が結合することが可能になり、それにより、エピトープ提示の不均一性および複雑さが増大することが原因で、不正確である。したがって、ＨＬＡクラスＩＩペプチド結合性コアおよびクラスＩＩエピトーププロセシングおよび提示に関与する細胞プロセス特徴をよりよく理解するためのさらなる研究が必要とされている。プロテオミクス分野は、現在、ＨＬＡクラスＩＩヘテロ二量体形成の複雑さおよびＨＬＡクラスＩＩ－ペプチド複合体を単離するための免疫沈降グレードの抗体の利用可能性により限定されている。これらの難題を克服するために、クラスＩＩエピトープ予測方法のための対立遺伝子特異的ＨＬＡクラ

スＩＩ－リガンドームの特徴付けに関してＬＣ－ＭＳ／ＭＳに依拠する単一対立遺伝子ＨＬＡプロファイリングワークフローを開発した。以下の定義は、当業者への補足であり、本出願を対象とし、いかなる関連するまたは関連しない場合、例えば、いかなる共同所有の特許または出願にも帰するものではない。本明細書に記載の方法および材料と類似した、またはそれと等しい任意の方法および材料を本開示を試験するための実施に使用することができるが、例示的な材料および方法が本明細書に記載されている。したがって、本明細書において使用される用語法は、特定の実施形態を説明するためだけのものであり、限定的なものではない。

【０４７４】

個別化がんワクチンを調製する方法が本明細書に開示される。個別化がんワクチンを調製するための方法は、対象のがん細胞において発現される変異を用いてペプチド配列を同定するステップと、同定されたペプチド配列についての提示予測のセットを生成するために、コンピュータプロセッサを使用して同定されたペプチド配列のアミノ酸位置情報を機械学習ＨＬＡ－ペプチド提示予測モデルに入力するステップであって、各提示予測が、対象のがん細胞のクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質が同定されたペプチド配列のうちの所与の配列を提示する確率を表す、ステップと、個別化がんワクチンを調製するために、提示予測のセットに基づいて同定されたペプチド配列のサブセットを選択するステップとを含み得る。

【０４７５】

一部の実施形態では、本明細書に記載の方法で得られた１つまたは複数の結果は、定量的な値または以下の１つまたは複数を示す値をもたらすものである：診断の正確度の可能性、対象における状態の存在の可能性、対象に状態が発生している可能性、特定の処置の成功の可能性、またはこれらの任意の組合せ。一部の実施形態では、本明細書に記載の方法により、状態が発生するリスクまたは可能性を予測することができる。一部の実施形態では、本明細書に記載の方法は、状態の発生に関する早期診断指標であり得る。一部の実施形態では、本明細書に記載の方法により、状態の診断または存在を確認することができる。一部の実施形態では、本明細書に記載の方法により、状態の増悪をモニタリングすることができる。一部の実施形態では、本明細書に記載の方法により、対象における状態に対する処置の有効性をモニタリングすることができる。

【０４７６】

ＭＨＣ－ＩＩペプチドを同定するための方法

一態様では、免疫活性化のために、ＭＨＣ－ＩＩタンパク質によって提示される１つまたは複数のペプチドを同定する方法が本明細書で提示される。一部の実施形態では、１つまたは複数のペプチドは、エピトープを含む。一部の実施形態では、方法は、特異的なエピトープがＭＨＣ－ＩＩタンパク質によって提示される可能性のコンピュータによる予測を伴う。一部の実施形態では、方法は、ＭＨＣ－ＩＩ提示に対するエピトープの特異性のコンピュータによる予測を伴う。一部の実施形態では、コンピュータによる予測方法は、ペプチド－ＭＨＣ相互作用の評価を伴う。一部の実施形態では、コンピュータによる予測方法は、抗原提示に対するペプチドの対立遺伝子特異性の予測を伴う。一部の実施形態では、コンピュータによる予測方法は、バイオインフォマティクス情報、例えば、ヌクレオチド配列、生体分子の構造モチーフ、タンパク質間相互作用の特色および免疫原性などの機能的力価の組込みを伴う。一部の実施形態では、コンピュータによる予測方法は、機械学習を伴う。ＭＨＣクラスＩおよびクラスＩＩの両方について、単純パターンモチーフ、サポートベクターマシン（ＳＶＭ）、隠れマルコフモデル（ＨＭＭ）、ニューラルネットワーク（ＮＮ）モデル、定量的構造と活性の関連性（ＱＳＡＲ）分析、構造に基づく方法、および生物物理学的方法などの機械学習手法に基づいてペプチド－ＭＨＣ相互作用を予測するためのイムノインフォマティクス方法が多数開発されている。これらの方法は、２つのカテゴリー、すなわち、対立遺伝子内（対立遺伝子特異的）方法およびトランス－対立遺伝子（汎特異的）方法に分けることができる。対立遺伝子内方法は、実験的なペプチド結合性データの限定されたセットに対して特定のＭＨＣ分子についての訓練を行い、ペ

プチド結合性の予測をその分子に適用する。ＭＨＣ分子が極度に多型的であり、何千もの対立遺伝子バリアントが存在することから、十分な実験的な結合データが欠如していることと相まって、各対立遺伝子に対して予測モデルを築くことは不可能である。したがって、多くの対立遺伝子または種間にわたって拡大するペプチド結合性データを使用して、トランス－対立遺伝子および汎用方法、例えば、ＮｅｔＭＨＣＩＩｐａｎ（Karosiene E etal., 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 HLADQ. Immunogenetics (2013) 65(10):711-24）、およびＴＥＰＩＴＯＰＥｐａｎ（Zhang L、et al., TEPITOPEpan: extending TEPITOPE for peptide binding prediction covering over 700 HLA-DR molecules. PLoS One (2012) 7(2):e30483）などが開発されている。ＮｅｔＭＨＣｐａｎおよびＫＩＳＳなどの、ＭＨＣ－Ｉに対する同様の方法も利用可能である。

【０４７７】

一部の実施形態では、ペプチド配列は、対象の正常細胞では発現されないものであり得る。一部の実施形態では、対象のありとあらゆる細胞ががん細胞ではない場合がある。がん細胞は、これだけに限定されないが、甲状腺がん、副腎皮質性がん、肛門がん、再生不良性貧血、胆管がん、膀胱がん（ｂｌａｄｄｅｒ　ｃａｎｃｅｒ）、骨がん、骨転移、中枢神経系（ＣＮＳ）がん、末梢神経系（ＰＮＳ）がん、乳がん、キャッスルマン病、子宮頸がん、小児期非ホジキンリンパ腫、リンパ腫、結腸がん・直腸がん、子宮体がん、食道がん（ｅｓｏｐｈａｇｕｓ　ｃａｎｃｅｒ）、ユーイングファミリーの腫瘍（例えば、ユーイング肉腫）、眼がん、胆嚢がん（ｇａｌｌｂｌａｄｄｅｒ　ｃａｎｃｅｒ）、消化管カルチノイド腫瘍、消化管間質腫瘍、妊娠性絨毛疾患、ヘアリー細胞白血病、ホジキン病、カポジ肉腫、腎がん（ｋｉｄｎｅｙ　ｃａｎｃｅｒ）、喉頭・下咽頭がん、急性リンパ性白血病、急性骨髄性白血病、小児白血病、慢性リンパ性白血病、慢性骨髄性白血病、肝がん、肺がん、肺カルチノイド腫瘍、非ホジキンリンパ腫、男性乳がん、悪性中皮腫、多発性骨髄腫、骨髄異形成症候群、骨髄増殖性疾患、鼻腔および副鼻腔がん、鼻咽頭がん、神経芽細胞腫、口腔・中咽頭がん、骨肉腫、卵巣がん、膵がん、陰茎がん、下垂体腫瘍、前立腺がん、網膜芽細胞腫、横紋筋肉腫、唾液腺がん、肉腫（成人軟部組織がん）、黒色腫皮膚がん、非黒色腫皮膚がん、胃がん、精巣がん、胸腺がん、子宮がん（例えば、子宮肉腫）、膣がん、外陰がん、またはワルデンストレームマクログロブリン血症を含めた異なるがんによって産生されたものであり得る。

【０４７８】

同定するステップは、対象のがん細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列を対象の正常細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列と比較することを含み得る。対象のがん細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列は、対象の正常細胞由来のＤＮＡ、ＲＮＡまたはタンパク質配列とは異なり得る。同定するステップにより、感受性が高い核酸バリアントを同定することができる。

【０４７９】

機械学習ＨＬＡ－ペプチド提示予測モデルは、少なくとも訓練用データに基づいて同定された複数の予測変数を含み得る。訓練用データは、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；アミノ酸位置情報を含む訓練用ペプチド配列情報であって、細胞において発現されるＨＬＡタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られたアミノ酸位置情報とアミノ酸位置情報および予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含み得る。

【０４８０】

一部の実施形態では、訓練用データは、構造化されたデータ、時系列データ、構造化されていないデータ、およびリレーショナルデータをさらに含み得る。構造化されていないデータは、ロボット工学またはシミュレーションの正確なシミュレートまたは訓練における使用のための、音声データ、画像データ、ビデオ、機械的データ、電気的データ、化学

的データ、およびこれらの任意の組合せを含み得る。時系列データは、スマートメーター、スマート電化製品、スマートデバイス、モニタリングシステム、遠隔測定デバイス、またはセンサーのうちの１つまたは複数からのデータを含み得る。リレーショナルデータは、顧客システム、企業システム、運用システム、ウェブサイト、ウェブアクセス可能なアプリケーションプログラムインターフェース（ＡＰＩ）、またはこれらの任意の組合せからのデータを含む。これは、ユーザーが、ファイルまたは他のデータフォーマットをソフトウェアまたはシステムに入力する任意の方法によって行うことができる。

【０４８１】

一部の実施形態では、訓練用データはデータベースに保存することができる。データベースは、コンピュータ可読フォーマットで保存することができる。コンピュータプロセッサを、コンピュータ可読メモリに保存されたデータにアクセスするように構成することができる。一部の実施形態では、結果を得るために、コンピュータシステムを使用してデータを解析することができる。結果を遠隔操作でまたは内部で記憶媒体に保存することができ、薬物適用専門家などの人員と通信することができる。一部の実施形態では、コンピュータシステムを、結果を伝達するための構成要素と作動可能にカップリングすることができる。伝達用の構成要素は、有線および無線構成要素を含み得る。有線伝達構成要素の例としては、Ｕｎｉｖｅｒｓａｌ Ｓｅｒｉａｌ Ｂｕｓ（ＵＳＢ）接続、同軸ケーブル接続、Ｃａｔ５またはＣａｔ６ケーブルなどのイーサネット（登録商標）ケーブル、光ファイバーケーブル、または電話線を挙げることができる。無線通信構成要素の例としては、Ｗｉ－Ｆｉレシーバー、３Ｇもしくは４Ｇ ＬＴＥデータシグナルなどの、移動データ標準にアクセスするための構成要素、またはＢｌｕｅｔｏｏｔｈ（登録商標）レシーバーを挙げることができる。一部の実施形態では、記憶媒体内のこれらのデータの全てを収集し、アーカイブに入れて、データウェアハウスを築く。

【０４８２】

一部の実施形態では、データベースは、外部データベースを含む。外部データベースは、医学的データベース、例えば、これだけに限定されないが、Ａｄｖｅｒｓｅ Ｄｒｕｇ Ｅｆｆｅｃｔｓ Ｄａｔａｂａｓｅ、AHFS Supplemental File、Ａｌｌｅｒｇｅｎ Ｐｉｃｋｌｉｓｔ Ｆｉｌｅ、Ａｖｅｒａｇｅ ＷＡＣ Ｐｒｉｃｉｎｇ Ｆｉｌｅ、Ｂｒａｎｄ Ｐｒｏｂａｂｉｌｉｔｙ Ｆｉｌｅ、Ｃａｎａｄｉａｎ Ｄｒｕｇ Ｆｉｌｅ ｖ２、Ｃｏｍｐｒｅｈｅｎｓｉｖｅ Ｐｒｉｃｅ Ｈｉｓｔｏｒｙ、Ｃｏｎｔｒｏｌｌｅｄ Ｓｕｂｓｔａｎｃｅｓ Ｆｉｌｅ、Ｄｒｕｇ Ａｌｌｅｒｇｙ Ｃｒｏｓｓ－Ｒｅｆｅｒｅｎｃｅ Ｆｉｌｅ、Ｄｒｕｇ Ａｐｐｌｉｃａｔｉｏｎ Ｆｉｌｅ、Ｄｒｕｇ Ｄｏｓｉｎｇ & Ａｄｍｉｎｉｓｔｒａｔｉｏｎ Ｄａｔａｂａｓｅ、Ｄｒｕｇ Ｉｍａｇｅ Ｄａｔａｂａｓｅ ｖ２．０／Ｄｒｕｇ Ｉｍｐｒｉｎｔ Ｄａｔａｂａｓｅ ｖ２．０、Ｄｒｕｇ Ｉｎａｃｔｉｖｅ Ｄａｔｅ Ｆｉｌｅ、Ｄｒｕｇ Ｉｎｄｉｃａｔｉｏｎｓ Ｄａｔａｂａｓｅ、Ｄｒｕｇ Ｌａｂ Ｃｏｎｆｌｉｃｔ Ｄａｔａｂａｓｅ、Ｄｒｕｇ Ｔｈｅｒａｐｙ Ｍｏｎｉｔｏｒｉｎｇ Ｓｙｓｔｅｍ（ＤＴＭＳ）ｖ２．２ ／ ＤＴＭＳ Ｃｏｎｓｕｍｅｒ Ｍｏｎｏｇｒａｐｈｓ、Ｄｕｐｌｉｃａｔｅ Ｔｈｅｒａｐｙ Ｄａｔａｂａｓｅ、Ｆｅｄｅｒａｌ Ｇｏｖｅｒｎｍｅｎｔ Ｐｒｉｃｉｎｇ Ｆｉｌｅ、Ｈｅａｌｔｈｃａｒｅ Ｃｏｍｍｏｎ Ｐｒｏｃｅｄｕｒｅ Ｃｏｄｉｎｇ Ｓｙｓｔｅｍ Ｃｏｄｅｓ（ＨＣＰＣＳ）Ｄａｔａｂａｓｅ、ＩＣＤ－１０ Ｍａｐｐｉｎｇ Ｆｉｌｅｓ、Ｉｍｍｕｎｉｚａｔｉｏｎ Ｃｒｏｓｓ－Ｒｅｆｅｒｅｎｃｅ Ｆｉｌｅ、Ｉｎｔｅｇｒａｔｅｄ Ａ ｔｏ Ｚ Ｄｒｕｇ Ｆａｃｔｓ Ｍｏｄｕｌｅ、Ｉｎｔｅｇｒａｔｅｄ Ｐａｔｉｅｎｔ Ｅｄｕｃａｔｉｏｎ、Ｍａｓｔｅｒ Ｐａｒａｍｅｔｅｒｓ Ｄａｔａｂａｓｅ、Ｍｅｄｉ－Ｓｐａｎ Ｅｌｅｃｔｒｏｎｉｃ Ｄｒｕｇ Ｆｉｌｅ（ＭＥＤ－Ｆｉｌｅ）ｖ２、Ｍｅｄｉｃａｉｄ Ｒｅｂａｔｅ Ｆｉｌｅ，Ｍｅｄｉｃａｒｅ Ｐｌａｎｓ Ｆｉｌｅ、Ｍｅｄｉｃａｌ Ｃｏｎｄｉｔｉｏｎ Ｐｉｃｋｌｉｓｔ Ｆｉｌｅ，Ｍｅｄｉｃａｌ Ｃｏｎｄｉｔｉｏｎｓ Ｍａｓｔｅｒ Ｄａｔａｂａｓｅ、Ｍｅｄｉｃａｔｉｏｎ Ｏｒｄｅｒ Ｍａｎａｇｅｍｅｎｔ Ｄａｔａｂａｓｅ（ＭＯＭＤ）、Ｐａｒａｍｅｔｅｒｓ ｔｏ Ｍｏｎｉｔｏｒ Ｄａｔａｂａｓｅ、Ｐａｔｉｅｎｔ Ｓａｆｅｔｙ

Ｐｒｏｇｒａｍｓ Ｆｉｌｅ、Ｐａｙｍｅｎｔ Ａｌｌｏｗａｎｃｅ Ｌｉｍｉｔ－Ｐａｒｔ Ｂ（ＰＡＬ－Ｂ）ｖ２．０、Ｐｒｅｃａｕｔｉｏｎｓ Ｄａｔａｂａｓｅ、ＲｘＮｏｒｍ Ｃｒｏｓｓ－Ｒｅｆｅｒｅｎｃｅ Ｆｉｌｅ、Ｓｔａｎｄａｒｄ Ｄｒｕｇ Ｉｄｅｎｔｉｆｉｅｒｓ Ｄａｔａｂａｓｅ，Ｓｕｂｓｔｉｔｕｔｉｏｎ Ｇｒｏｕｐｓ Ｆｉｌｅ、Ｓｕｐｐｌｅｍｅｎｔａｌ Ｎａｍｅｓ Ｆｉｌｅ、Ｕｎｉｆｏｒｍ Ｓｙｓｔｅｍ ｏｆ Ｃｌａｓｓｉｆｉｃａｔｉｏｎ Ｃｒｏｓｓ－Ｒｅｆｅｒｅｎｃｅ Ｆｉｌｅ、またはＷａｒｎｉｎｇ Ｌａｂｅｌ Ｄａｔａｂａｓｅであり得る。

【０４８３】

一部の実施形態では、訓練用データを、他のデータ供給源を通じて得ることもできる。データ供給源は、センサーまたはスマートデバイス、例えば、電化製品、スマートメーター、ウェアラブル、モニタリングシステム、データストア、顧客システム、課金システム、財務システム、クラウドソースデータ、気象データ、ソーシャルネットワーク、または任意の他のセンサー、企業システムまたはデータストアなどを含み得る。スマートメーターまたはセンサーの例としては、顧客サイトに位置するメーターもしくはセンサー、または顧客と、生成もしくは供給源場所の間に位置するメーターもしくはセンサーを挙げることができる。広範な供給源のアレイからのデータを組み入れることにより、システムを、複雑かつ詳細な解析を実施できるものにすることができる。一部の実施形態では、データ供給源は、限定することなく他の医学的プラットフォームのためのセンサーまたはデータベースを含み得る。

【０４８４】

ＨＬＡタイピングは、抗体を使用した血清学的方法、または、配列特異的オリゴヌクレオチドプローブハイブリダイゼーション（Ｓｅｑｕｅｎｃｅ Ｓｐｅｃｉｆｉｃ Ｏｌｉｇｏｎｕｃｌｅｏｔｉｄｅ Ｐｒｏｂｅ Ｈｙｂｒｉｄｉｚａｔｉｏｎ）（ＳＳＯＰ）、もしくは配列に基づくタイピング（Ｓｅｑｕｅｎｃｅ Ｂａｓｅｄ Ｔｙｐｉｎｇ）（ＳＢＴ）などのＰＣＲに基づく方法のいずれかによって慣習的に行われる。前者は潜在的に高度な交差反応性および限定された分解能によって妨害され、一方、後者では、多型位置が原因でプライマー位置決めの可能性が非常に限られることに起因する、ＰＣＲの効率に関連する困難が生じる。

【０４８５】

一部の実施形態では、配列情報を、配列決定方法または質量分析を使用する方法、例えば、液体クロマトグラフィー－質量分析（ＬＣ－ＭＳまたはＬＣ－ＭＳ／ＭＳ、またはその代わりにＨＰＬＣ－ＭＳまたはＨＰＬＣ－ＭＳ／ＭＳ）などのいずれかによって同定する。これらの配列決定方法は当業者には周知であり得、また、Medzihradszky KF and Chalkley RJ. Mass Spectrom Rev. 2015 Jan-Feb； 34 (1): 43-63に概説されている。一部の実施形態では、質量分析は、単一対立遺伝子質量分析である。一部の実施形態では、質量分析は、ＭＳ分析、ＭＳ／ＭＳ分析、ＬＣ－ＭＳ／ＭＳ分析、またはこれらの組合せであり得る。一部の実施形態では、ＭＳ分析を使用して、インタクトなペプチドの質量を決定することができる。例えば、決定することは、インタクトなペプチドの質量を決定すること（例えば、ＭＳ分析）を含み得る。一部の実施形態では、ＭＳ／ＭＳ分析を使用して、ペプチド断片の質量を決定することができる。例えば、決定することは、ペプチド断片の質量を決定することを含み得、それを使用して、ペプチドまたはその一部のアミノ酸配列を決定することができる（例えば、ＭＳ／ＭＳ分析）。一部の実施形態では、ペプチド断片の質量を使用して、ペプチド内のアミノ酸の配列を決定することができる。一部の実施形態では、ＬＣ－ＭＳ／ＭＳ分析を使用して、複雑なペプチド混合物を分離することができる。例えば、決定することは、液体クロマトグラフィーによってなどで複雑なペプチド混合物を分離すること、およびインタクトなペプチドの質量、ペプチド断片の質量、またはこれらの組合せを決定することを含み得る（例えば、ＬＣ－ＭＳ／ＭＳ分析）。このデータを、例えばペプチド配列決定のために使用することができる。

【０４８６】

一部の実施形態では、訓練用ペプチド配列情報は、訓練用ペプチドのアミノ酸位置情報

を含む。一部の実施形態では、訓練用ペプチド配列情報は、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報の最大で約９０％、８０％、７０％、６０％、５０％、４０％、３０％、２０％、１０％またはそれ未満を含む。一部の実施形態では、訓練用ペプチド配列情報は、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報の少なくとも約１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％またはそれよりも多くを含み得る。

【０４８７】

任意の情報およびデータを、情報およびデータの供給源である対象と対にすることができる。対象または医療専門家は、ストレージまたはサーバーから情報およびデータを対象のアイデンティティーによって検索することができる。対象のアイデンティティーは、患者の写真、名前、住所、社会保障番号、誕生日、電話番号、郵便番号、またはこれらの任意の組合せを含み得る。対象のアイデンティティーは、暗号化し、視覚的な図式コードに符号化することができる。視覚的な図式コードは、対象のアイデンティティーに一意的に関連付けることができるワンタイムバーコードであり得る。バーコードは、ＵＰＣバーコード、ＥＡＮバーコード、Ｃｏｄｅ　３９バーコード、Ｃｏｄｅ　１２８バーコード、ＩＴＦバーコード、ＣｏｄａＢａｒバーコード、ＧＳ１　ＤａｔａＢａｒバーコード、ＭＳＩ　Ｐｌｅｓｓｅｙバーコード、ＱＲバーコード、Ｄａｔａｍａｔｒｉｘコード、ＰＤＦ４１７コード、またはＡｚｔｅｃバーコードであり得る。視覚的な図式コードは、表示画面上に表示されるように構成することができる。バーコードは、機械によって光学的に捕捉し、読み取ることができるＱＲを含み得る。バーコードにより、バージョン、フォーマット、位置、アラインメント、またはバーコードの読み取りおよび復号が可能になるバーコードのタイミングなどの要素を規定することができる。バーコードは、種々の型の情報をバイナリまたは英数字情報などの任意の型の適切なフォーマットでコードし得る。ＱＲコード（登録商標）は、ＱＲコード（登録商標）を画像処理デバイスによって合理的な距離からスキャンすることができる限りは種々のシンボルサイズを有し得る。ＱＲコード（登録商標）は、任意の画像ファイルフォーマットのものであり得る（例えば、ＥＰＳまたはＳＶＧベクトルグラフ、ＰＮＧ、ＴＩＦ、ＧＩＦ、またはＪＰＥＧラスターグラフィックスフォーマット）。

【０４８８】

一部の実施形態では、入力として受け取られたアミノ酸位置情報とアミノ酸位置情報および予測変数に基づいて出力として生成された提示可能性との関係を表す関数は、線形関数または非線形関数を含む。関数は、例えば、正規化線形ユニット（ＲｅＬＵ）活性化関数、Ｌｅａｋｙ　ＲｅＬｕ活性化関数、または、飽和双曲線正接関数、同一性関数、バイナリステップ関数、ロジスティック関数、ａｒｃＴａｎ関数、ソフトサイン関数、パラメトリック正規化線形ユニット関数、指数関数線形ユニット関数、ｓｏｆｔＰｌｕｓ関数、ベントアイデンティティ（ｂｅｎｔ　ｉｄｅｎｔｉｔｙ）関数、ｓｏｆｔＥｘｐｏｎｅｎｔｉａｌ関数、Ｓｉｎｕｓｏｉｄ関数、Ｓｉｎｃ関数、Ｇａｕｓｓｉａｎ関数、またはシグモイド関数などの他の関数、またはこれらの任意の組合せであり得る。

【０４８９】

一部の実施形態では、線形関数を線形回帰によって得る。一部の実施形態では、線形回帰は、従属変数と独立変数の間の最良の直線関係を当てはめることによって標的変数を予測するための方法である。最良当てはめは、各点における形状と実際の観察値の間の距離の全ての合計が最小であることを意味する。線形回帰は、単純線形回帰または多重線形回帰を含み得る。単純線形回帰では、単一の独立変数を使用して従属変数を予測することができる。多重線形回帰では、１つよりも多くの独立変数を使用して、最良の直線関係を当てはめることによって、従属変数を予測することができる。非線形関数は非線形回帰によって得ることができる。非線形回帰は、観察に基づくデータが、モデルパラメータの非線形組合せであり、１つまたは複数の独立変数に依存する関数によってモデル化される、回帰分析の一形態であり得る。非線形回帰は、ステップ関数、区分的関数、スプライン、お

よび一般加法モデルを含み得る。

【０４９０】

一部の実施形態では、提示可能性は、一次元の値（例えば、確率）によって示される。一部の実施形態では、確率は、事象が起こり得る可能性が測定されるように構成される。一部の実施形態では、確率は、約０からおよび１まで、０．１～０．９、０．２～０．８、０．３～０．７、または０．４～０．６にわたる。確率が高い事象ほど、その事象が起こり得る可能性が大きい。一部の実施形態では、事象は、非限定的な例として、ＨＬＡ－ペプチドがある特定のアミノ酸位置情報を有するいくつかのペプチドを提示するかどうか、および人がアミノ酸位置情報に基づいて病気にかかるかどうかを含めた、任意の型の状況を含む。一部の実施形態では、可能性は、多次元の値で示すことができる。多次元の値は、多次元空間、ヒートマップ、またはスプレッドシートによって示すことができる。

【０４９１】

一実施形態では、提示予測のセットに基づいて同定されたペプチド配列のサブセットの選択を、個別化がんワクチンが調製されるように構成する。一部の実施形態では、サブセットは、提示予測のセットに基づいて同定された、最大で約９０％、８０％、７０％、６０％、５０％、４０％、３０％、２０％、１０％またはそれ未満のペプチド配列を含む。他の場合では、サブセットは、提示予測のセットに基づいて同定された、少なくとも約１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％またはそれよりも多くのペプチド配列を含み得る。がんワクチンは、既存のがんを処置するか、またはがんの発生を防止するワクチンであり得る。ワクチンは、患者から取得した試料から調製することができ、その患者に対して特定のものであり得る。

【０４９２】

一部の実施形態では、ポックスウイルスを疾患（例えば、がん）ワクチンまたは免疫原性組成物に使用する。これらとしては、オルソポックスウイルス、アビポックス、ワクシニア、ＭＶＡ、ＮＹＶＡＣ、カナリアポックス、ＡＬＶＡＣ、鶏痘、ＴＲＯＶＡＣなどが挙げられる。ベクターの利点としては、構築が簡単であること、大量の外来ＤＮＡを適応させることができること、および発現レベルが高いことを挙げることができる。Ｃｈｏｒｄｏｐｏｘｖｉｒｉｎａｅサブファミリーポックスウイルス（脊椎動物のポックスウイルス）、例えば、とりわけ、オルソポックスウイルスおよびアビポックスウイルス、例えば、ワクシニアウイルス（例えば、Ｗｙｅｔｈ株、ＷＲ株（例えば、ＡＴＣＣ（登録商標）ＶＲ－１３５４）、Ｃｏｐｅｎｈａｇｅｎ株、ＮＹＶＡＣ、ＮＹＶＡＣ．１、ＮＹＶＡＣ．２、ＭＶＡ、ＭＶＡ－ＢＮ）、カナリアポックスウイルス（例えば、Ｗｈｅａｔｌｅｙ Ｃ９３株、ＡＬＶＡＣ）、鶏痘ウイルス（例えば、ＦＰ９株、Ｗｅｂｓｔｅｒ株、ＴＲＯＶＡＣ）、鳩痘、鳩痘、ウズラ痘、およびアライグマ痘、それらの合成のまたは天然に存在しない組換えなどの、本開示の実施に使用することができるポックスウイルス、その使用、ならびにそのような組換えの作出および使用方法に関する情報は、科学文献および特許文献に見いだすことができる。

【０４９３】

一部の実施形態では、ワクシニアウイルスを、抗原を発現させるために疾患ワクチンまたは免疫原性組成物に使用する。組換えワクシニアウイルスは、感染した宿主細胞の細胞質内で複製することができるものであり得、したがって、目的のポリペプチドにより免疫応答を誘導することができる。

【０４９４】

一部の実施形態では、ＡＬＶＡＣを疾患ワクチンまたは免疫原性組成物にベクターとして使用する。ＡＬＶＡＣは、外来導入遺伝子を発現するように改変することができ、原核生物抗原および真核生物抗原の両方のワクチン接種の方法として使用されてきた、カナリアポックスウイルスであり得る。

【０４９５】

一部の実施形態では、改変ワクシニアアンカラ（Ｍｏｄｉｆｉｅｄ Ｖａｃｃｉｎｉａ Ａｎｋａｒａ）（ＭＶＡ）ウイルスを抗原ワクチンまたは免疫原性組成物のためのウイル

スベクターとして使用する。ＭＶＡは、オルソポックスウイルスファミリーのメンバーであり得、ワクシニアウイルス（ＣＶＡ）のＡｎｋａｒａ株のニワトリ胚線維芽細胞に対する約５７０回の段階的継代によって生成された。これらの継代の結果として、得られたＭＶＡウイルスは、ＣＶＡと比較して３１キロベース少ないゲノム情報を含み得、また、高度に宿主細胞により制限される。ＭＶＡは、極度の弱毒化、すなわち、毒力または感染能の減弱を特徴とし得るが、それでもなお優れた免疫原性を保持する。種々の動物モデルにおいて試験した場合、ＭＶＡは、免疫抑制された個体においてさえ、弱毒性であることが証明され得る。さらに、ＭＶＡ－ＢＮ（登録商標）－ＨＥＲ２は、ＨＥＲ－２陽性乳がんの処置のために設計された免疫治療の候補になり得、現在臨床試験中である。

【０４９６】

一部の実施形態では、陽性適中率（ＰＰＶ）を予測モデルの一部として使用する。ＰＰＶは、精度測定値としても公知であり、例えば検査またはモデルによって疾患または状態を有すると診断された個体が実際にその疾患または状態を有する確率である。ＰＰＶは、真陽性結果の数を陽性が返された結果の総数（偽陽性を含む結果）で割ることによって算出することができる。ＰＰＶ＝真陽性／（真陽性＋偽陽性）。例えば、患者１００名のセットにおいて、モデルにより５０名の患者で陽性結果が同定され、そのうちの２５名が真陽性であった場合、ＰＰＶは、２５／５０＝０．５になる。ＰＰＶが１に近いことにより、検査またはモデルなどの診断方法がより正確であることが表される。ＰＰＶを使用して、予測モデルの正確度を決定することができる。ＰＰＶを使用して、モデルによって生成される可能性がある偽陽性の結果に適応するように予測モデルを調整することができる。

【０４９７】

再現率を予測モデルの一部として使用することができる。再現率は、試料セットにおける陽性の総数のうちの真陽性結果のパーセンテージとみなすことができる。再現率＝真陽性／（真陽性＋偽陰性）。例えば、患者１００名のセットにおいて、モデルにより５０名の患者で陽性結果が同定され、そのうちの２５名が真陽性であり、患者のセットにおいて合計７５名の陽性が存在した場合、再現率は、｛２５／（２５＋２５）｝×１００＝５０％になる。再現率を使用して、予測モデルの正確度を決定することができる。再現率を使用して、予測モデルを、モデルによって生成される可能性がある偽陽性の結果または偽陰性の結果に適応するように調整することができる。

【０４９８】

一部の実施形態では、予測モデルの陽性適中率は、０．１％～１０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれよりも高いものである。一部の実施形態では、予測モデルの陽性適中率は、０．１％～１０％の再現率で、最大で０．９、０．８、０．７、０．６、０．５、０．４、０．３、０．２、０．１またはそれ未満であり得る。予測モデルの陽性適中率は、０．１％未満の再現率で、少なくとも０．０５、０．１、０．２、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれよりも高いものであり得る。一部の実施形態では、予測モデルの陽性適中率は、０．１％未満の再現率で、最大で０．９、０．８、０．７、０．６、０．５、０．４、０．３、０．２、０．１またはそれ未満であり得る。予測モデルの陽性適中率は、１０％よりも高い再現率で、少なくとも０．０５、０．１、０．２、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれよりも高いものであり得る。一部の実施形態では、予測モデルの陽性適中率は、１０％よりも高い再現率で、最大で０．９、０．８、０．７、０．６、０．５、０．４、０．３、０．２、０．１またはそれ未満であり得る。

【０４９９】

一部の実施形態では、予測モデルの陽性適中率は、０．１％～１０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれより高い。一部の実施形態では、予測モデルの陽性適中率は、０．１％～０．５％、０．１％～１％、０．１％～２％、０．１％～３％、０．１％～４％、０．１％～５％、０．１％～６％、０．１％～７％、０．１％～８％、０．１％～

９％、０．１％～１０％、０．５％～１％、０．５％～２％、０．５％～３％、０．５％～４％、０．５％～５％、０．５％～６％、０．５％～７％、０．５％～８％、０．５％～９％、０．５％～１０％、１％～２％、１％～３％、１％～４％、１％～５％、１％～６％、１％～７％、１％～８％、１％～９％、１％～１０％、２％～３％、２％～４％、２％～５％、２％～６％、２％～７％、２％～８％、２％～９％、２％～１０％、３％～４％、３％～５％、３％～６％、３％～７％、３％～８％、３％～９％、３％～１０％、４％～５％、４％～６％、４％～７％、４％～８％、４％～９％、４％～１０％、５％～６％、５％～７％、５％～８％、５％～９％、５％～１０％、６％～７％、６％～８％、６％～９％、６％～１０％、７％～８％、７％～９％、７％～１０％、８％～９％、８％～１０％または９％～１０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれより高い。一部の実施形態では、予測モデルの陽性適中率は、０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％または１０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれより高い。一部の実施形態では、予測モデルの陽性適中率は、少なくとも０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％または９％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれより高い。一部の実施形態では、予測モデルの陽性適中率は、最大で０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％または１０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれより高い。

【０５００】

一部の実施形態では、予測モデルの陽性適中率は、１０％～２０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれより高い。一部の実施形態では、予測モデルの陽性適中率は、１０％～１１％、１０％～１２％、１０％～１３％、１０％～１４％、１０％～１５％、１０％～１６％、１０％～１７％、１０％～１８％、１０％～１９％、１０％～２０％、１１％～１２％、１１％～１３％、１１％～１４％、１１％～１５％、１１％～１６％、１１％～１７％、１１％～１８％、１１％～１９％、１１％～２０％、１２％～１３％、１２％～１４％、１２％～１５％、１２％～１６％、１２％～１７％、１２％～１８％、１２％～１９％、１２％～２０％、１３％～１４％、１３％～１５％、１３％～１６％、１３％～１７％、１３％～１８％、１３％～１９％、１３％～２０％、１４％～１５％、１４％～１６％、１４％～１７％、１４％～１８％、１４％～１９％、１４％～２０％、１５％～１６％、１５％～１７％、１５％～１８％、１５％～１９％、１５％～２０％、１６％～１７％、１６％～１８％、１６％～１９％、１６％～２０％、１７％～１８％、１７％～１９％、１７％～２０％、１８％～１９％、１８％～２０％または１９％～２０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれより高い。一部の実施形態では、予測モデルの陽性適中率は、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれより高い。一部の実施形態では、予測モデルの陽性適中率は、少なくとも１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％または１９％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれより高い。一部の実施形態では、予測モデルの陽性適中率は、最大で１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれより高い。

【０５０１】

一部の実施形態では、予測モデルの陽性適中率は、少なくとも０．１％、０．５％、１

％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％、または２０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれよりも高いものである。例えば、予測モデルの陽性適中率は、少なくとも１０％の再現率で少なくとも０．１であり得る。例えば、予測モデルの陽性適中率は、少なくとも１０％の再現率で、少なくとも０．２であり得る。例えば、予測モデルの陽性適中率は、少なくとも１０％の再現率で、少なくとも０．３であり得る。例えば、予測モデルの陽性適中率は、少なくとも１０％の再現率で、少なくとも０．４であり得る。例えば、予測モデルの陽性適中率は、少なくとも１０％の再現率で、少なくとも０．５であり得る。例えば、予測モデルの陽性適中率は、少なくとも１０％の再現率で、少なくとも０．６であり得る。例えば、予測モデルの陽性適中率は、少なくとも１０％少なくとも０．７であり得る。例えば、予測モデルの陽性適中率は、少なくとも１０％の再現率で、少なくとも０．８であり得る。例えば、予測モデルの陽性適中率は、少なくとも１０％の再現率で、少なくとも０．９であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．１であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．２であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．３であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．４であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．５であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．６であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．７であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．８であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．９であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．１であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．２であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．３であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．４であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．５であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．６であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．７であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．８であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．９であり得る。

【０５０２】

一部の実施形態では、予測モデルの陽性適中率は、約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％、または２０％の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれよりも高いものである。例えば、予測モデルの陽性適中率は、約１０％の再現率で少なくとも０．１であり得る。例えば、予測モデルの陽性適中率は、約１０％の再現率で、少なくとも０．２であり得る。例えば、予測モデルの陽性適中率は、約１０％の再現率で、少なくとも０．３であり得る。例えば、予測モデルの陽性適中率は、約１０％の再現率で、少なくとも０．４であり得る。例えば、予測モデルの陽性適中率は、約１０％の再現率で、少なくとも０．５であり得る。例えば、予測モデルの陽性適中率は、約１０％の再現率で、少なくとも０．６であり得る。例えば、予測モデルの陽性適中率は、約１０％の再現率で、少なくとも０．７であり得る。例えば、予測モデルの陽性適中率は、約１０％の再現率で、少なくとも０．８であり得る。例えば、予測モデルの陽性適中率は、約１０％の再現率で、少なくとも０．９であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．１であり得る。例えば、予測モデ

ルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．２であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．３であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．４であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．５であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．６であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．７であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．８であり得る。例えば、予測モデルの陽性適中率は、少なくとも５％の再現率で、少なくとも０．９であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．１であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．２であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．３であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．４であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．５であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．６であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．７であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．８であり得る。例えば、予測モデルの陽性適中率は、少なくとも２０％の再現率で、少なくとも０．９であり得る。

【０５０３】

一部の実施形態では、予測モデルの陽性適中率は、０．１％未満、０．５％未満、１％未満、２％未満、３％未満、４％未満、５％未満、６％未満、７％未満、８％未満、９％未満、１０％未満、１１％未満、１２％未満、１３％未満、１４％未満、１５％未満、１６％未満、１７％未満、１８％未満、１９％未満、または２０％未満の再現率で、少なくとも０．０５、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７、０．８、０．９またはそれよりも高いものである。例えば、予測モデルの陽性適中率は、最大で１０％の再現率で少なくとも０．１であり得る。例えば、予測モデルの陽性適中率は、最大で１０％の再現率で、少なくとも０．２であり得る。例えば、予測モデルの陽性適中率は、最大で１０％の再現率で、少なくとも０．３であり得る。例えば、予測モデルの陽性適中率は、最大で１０％の再現率で、少なくとも０．４であり得る。例えば、予測モデルの陽性適中率は、最大で１０％の再現率で、少なくとも０．５であり得る。例えば、予測モデルの陽性適中率は、最大で１０％の再現率で、少なくとも０．６であり得る。例えば、予測モデルの陽性適中率は、最大で１０％の再現率で、少なくとも０．７であり得る。例えば、予測モデルの陽性適中率は、最大で１０％の再現率で、少なくとも０．８であり得る。例えば、予測モデルの陽性適中率は、最大で１０％の再現率で、少なくとも０．９であり得る。例えば、予測モデルの陽性適中率は、最大で５％の再現率で、少なくとも０．１であり得る。例えば、予測モデルの陽性適中率は、最大で５％の再現率で、少なくとも０．２であり得る。例えば、予測モデルの陽性適中率は、最大で５％の再現率で、少なくとも０．３であり得る。例えば、予測モデルの陽性適中率は、最大で５％の再現率で、少なくとも０．４であり得る。例えば、予測モデルの陽性適中率は、最大で５％の再現率で、少なくとも０．５であり得る。例えば、予測モデルの陽性適中率は、最大で５％の再現率で、少なくとも０．６であり得る。例えば、予測モデルの陽性適中率は、最大で５％の再現率で、少なくとも０．７であり得る。例えば、予測モデルの陽性適中率は、最大で５％の再現率で、少なくとも０．８であり得る。例えば、予測モデルの陽性適中率は、最大で５％の再現率で、少なくとも０．９であり得る。例えば、予測モデルの陽性適中率は、最大で２０％の再現率で、少なくとも０．１であり得る。例えば、予測モデルの陽性適中率は、最大で２０％の再現率で、少なくとも０．２であり得る。例えば、予測モデルの陽性適中率は、最大で２０％の再現率で、少なくとも０．３であり得る。例えば、予測モデルの陽性適中率は、最大で２０％の再現率で、少なくとも０．４であり得る。例えば、予測モデルの陽性適中率は、最大で２０％の再現率で、少なくとも０．５であり得

る。例えば、予測モデルの陽性適中率は、最大で２０％の再現率で、少なくとも０．６であり得る。例えば、予測モデルの陽性適中率は、最大で２０％の再現率で、少なくとも０．７であり得る。例えば、予測モデルの陽性適中率は、最大で２０％の再現率で、少なくとも０．８であり得る。例えば、予測モデルの陽性適中率は、最大で２０％の再現率で、少なくとも０．９であり得る。

【０５０４】

一部の実施形態では、約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％、または２０％の再現率で、予測モデルの陽性適中率は０．０５％～０．６％である。約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％の再現率で、予測モデルの陽性適中率は、０．０５％～０．１％、０．０５％～０．１５％、０．０５％～０．２％、０．０５％～０．２５％、０．０５％～０．３％、０．０５％～０．３５％、０．０５％～０．４％、０．０５％～０．４５％、０．０５％～０．５％、０．０５％～０．５５％、０．０５％～０．６％、０．１％～０．１５％、０．１％～０．２％、０．１％～０．２５％、０．１％～０．３％、０．１％～０．３５％、０．１％～０．４％、０．１％～０．４５％、０．１％～０．５％、０．１％～０．５５％、０．１％～０．６％、０．１５％～０．２％、０．１５％～０．２５％、０．１５％～０．３％、０．１５％～０．３５％、０．１５％～０．４％、０．１５％～０．４５％、０．１５％～０．５％、０．１５％～０．５５％、０．１５％～０．６％、０．２％～０．２５％、０．２％～０．３％、０．２％～０．３５％、０．２％～０．４％、０．２％～０．４５％、０．２％～０．５％、０．２％～０．５５％、０．２％～０．６％、０．２５％～０．３％、０．２５％～０．３５％、０．２５％～０．４％、０．２５％～０．４５％、０．２５％～０．５％、０．２５％～０．５５％、０．２５％～０．６％、０．３％～０．３５％、０．３％～０．４％、０．３％～０．４５％、０．３％～０．５％、０．３％～０．５５％、０．３％～０．６％、０．３５％～０．４％、０．３５％～０．４５％、０．３５％～０．５％、０．３５％～０．５５％、０．３５％～０．６％、０．４％～０．４５％、０．４％～０．５％、０．４％～０．５５％、０．４％～０．６％、０．４５％～０．５％、０．４５％～０．５５％、０．４５％～０．６％、０．５％～０．５５％、０．５％～０．６％または０．５５％～０．６％であり得る。約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％の再現率で、予測モデルの陽性適中率は、０．０５％、０．１％、０．１５％、０．２％、０．２５％、０．３％、０．３５％、０．４％、０．４５％、０．５％、０．５５％または０．６％であり得る。約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％の再現率で、予測モデルの陽性適中率は、少なくとも０．０５％、０．１％、０．１５％、０．２％、０．２５％、０．３％、０．３５％、０．４％、０．４５％、０．５％または０．５５％であり得る。約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％の再現率で、予測モデルの陽性適中率は、最大で０．１％、０．１５％、０．２％、０．２５％、０．３％、０．３５％、０．４％、０．４５％、０．５％、０．５５％または０．６％であり得る。

【０５０５】

約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％の再現率で、予測モデルの陽性適中率は、０．４５％～０．９８％であり得る。約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または

２０％の再現率で、予測モデルの陽性適中率は、０．４５％～０．５％、０．４５％～０．５５％、０．４５％～０．６％、０．４５％～０．６５％、０．４５％～０．７％、０．４５％～０．７５％、０．４５％～０．８％、０．４５％～０．８５％、０．４５％～０．９％、０．４５％～０．９６％、０．４５％～０．９８％、０．５％～０．５５％、０．５％～０．６％、０．５％～０．６５％、０．５％～０．７％、０．５％～０．７５％、０．５％～０．８％、０．５％～０．８５％、０．５％～０．９％、０．５％～０．９６％、０．５％～０．９８％、０．５５％～０．６％、０．５５％～０．６５％、０．５５％～０．７％、０．５５％～０．７５％、０．５５％～０．８％、０．５５％～０．８５％、０．５５％～０．９％、０．５５％～０．９６％、０．５５％～０．９８％、０．６％～０．６５％、０．６％～０．７％、０．６％～０．７５％、０．６％～０．８％、０．６％～０．８５％、０．６％～０．９％、０．６％～０．９６％、０．６％～０．９８％、０．６５％～０．７％、０．６５％～０．７５％、０．６５％～０．８％、０．６５％～０．８５％、０．６５％～０．９％、０．６５％～０．９６％、０．６５％～０．９８％、０．７％～０．７５％、０．７％～０．８％、０．７％～０．８５％、０．７％～０．９％、０．７％～０．９６％、０．７％～０．９８％、０．７５％～０．８％、０．７５％～０．８５％、０．７５％～０．９％、０．７５％～０．９６％、０．７５％～０．９８％、０．８％～０．８５％、０．８％～０．９％、０．８％～０．９６％、０．８％～０．９８％、０．８５％～０．９％、０．８５％～０．９６％、０．８５％～０．９８％、０．９％～０．９６％、０．９％～０．９８％または０．９６％～０．９８％であり得る。約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％の再現率で、予測モデルの陽性適中率は、０．４５％、０．５％、０．５５％、０．６％、０．６５％、０．７％、０．７５％、０．８％、０．８５％、０．９％、０．９６％または０．９８％であり得る。約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％の再現率で、予測モデルの陽性適中率は、少なくとも０．４５％、０．５％、０．５５％、０．６％、０．６５％、０．７％、０．７５％、０．８％、０．８５％、０．９％または０．９６％であり得る。約０．１％、０．５％、１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、１５％、１６％、１７％、１８％、１９％または２０％の再現率で、予測モデルの陽性適中率は、最大で０．５％、０．５５％、０．６％、０．６５％、０．７％、０．７５％、０．８％、０．８５％、０．９％、０．９６％または０．９８％であり得る。

【０５０６】

機械学習ＨＬＡ－ペプチド提示予測モデルを訓練する方法

ある態様では、機械学習ＨＬＡ－ペプチド提示予測モデルを訓練する方法は、コンピュータプロセッサを使用して、ＨＬＡクラスＩＩ対立遺伝子を発現する細胞由来の１つまたは複数のＨＬＡ－ペプチド複合体から単離されたＨＬＡ－ペプチドのアミノ酸位置情報配列をＨＬＡ－ペプチド提示予測モデルに入力するステップを含み得、機械学習ＨＬＡ－ペプチド提示予測モデルを訓練することは、ニューラルネットワークのノードに対する重み付けされた値を、提供された訓練用データに最良に適合するように調整することを含み得る。

【０５０７】

訓練用データは、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；訓練用ペプチドのアミノ酸位置情報を含む訓練用ペプチド配列情報であって、細胞において発現されるＨＬＡタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られたアミノ酸位置情報とアミノ酸位置情報および予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含み得る。訓練用データ、訓練用ペプチド配列情報、関数、および提示可能性は本明細書の他の箇所で開示されている。

【０５０８】
訓練されたアルゴリズムは、１つまたは複数のニューラルネットワークを含み得る。ニューラルネットワークは、一連の層におけるいくつかの接続したニューロン（またはノード）のグラフに基づくコンピューティングシステムの一種であり得る。ニューラルネットワークは、データが提示される入力層；１つまたは複数の内部層および／または「隠れ」層；ならびに結果が提示される出力層を含み得る。ニューラルネットワークは、入力データセットと標的データセットの関係を、一連の接続の重みを調整することによって学習することができる。ニューロンは、他の層内のニューロンと、接続の強度を制御するパラメータである重みを有する接続を介して接続することができる。各層内のニューロンの数は、解かれる問題の複雑さに関連付けることができる。層内の必要なニューロンの最小数は問題の複雑さによって決定することができ、最大数はニューラルネットワークの一般化能力によって限定され得る。入力ニューロンは、提示されたデータを受けとり、次いで、そのデータを第１の隠れ層内のノードに接続の重みを通じて伝達することができ、接続の重みは、訓練中に改変される。結果ノードは、入力とそれらの付随する重みの全ての対の積を合計することができる。重み付けされた合計をバイアスでオフセットして、結果ノードの値を調整することができる。ノードまたはニューロンの出力を、閾値または活性化関数を使用してゲーティングすることができる。活性化関数は線形または非線形関数であり得る。活性化関数は、例えば、正規化線形ユニット（ＲｅＬＵ）活性化関数、Ｌｅａｋｙ ＲｅＬｕ活性化関数、または飽和双曲線正接関数、同一性関数、バイナリステップ関数、ロジスティック関数、ａｒｃＴａｎ関数、ソフトサイン関数、パラメトリック正規化線形ユニット関数、指数関数線形ユニット関数、ｓｏｆｔＰｌｕｓ関数、ベントアイデンティティ関数、ｓｏｆｔＥｘｐｏｎｅｎｔｉａｌ関数、Ｓｉｎｕｓｏｉｄ関数、Ｓｉｎｃ関数、Ｇａｕｓｓｉａｎ関数、またはシグモイド関数などの他の関数、またはこれらの任意の組合せであり得る。

【０５０９】
ニューラルネットワーク内の隠れ層は、データを処理し、その結果を次の層に重み付けされた接続の第２のセットを通じて伝達することができる。その後の各層は、前の層からの結果を、より複雑な関係に「プール」することができる。ニューラルネットワークを、それら自体が訓練の間（およびその後）に、出力値などの、所与の入力のセットからの所望の出力がもたらされるように改変されることを可能にすることにより、訓練用データ（１つまたは複数のセンサーから収集されたデータ）の既知の試料セットを用いて訓練することができる。訓練されるアルゴリズムは、畳み込みニューラルネットワーク、再帰型ニューラルネットワーク、拡張畳み込みニューラルネットワーク、全結合ニューラルネットワーク、深層生成モデル、およびボルツマンマシンを含み得る。

【０５１０】
ニューラルネットワークの重み付け因子、バイアス値、および閾値、または他の計算パラメータは、訓練相において、訓練用データの１つまたは複数のセットを使用して「教えられる」または「学習する」。例えば、パラメータを、訓練用データセットからの入力データおよび最急降下法または逆伝播法を使用して訓練することができ、したがって、ニューラルネットワークからの出力値（複数可）は、訓練用データセットに含まれる例と一致する。

【０５１１】
ニューラルネットワークの入力層に使用されるノードの数は、少なくとも約１０、５０、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００、２０００、３０００、４０００、５０００、６０００、７０００、８０００、９０００、１０，０００、２０，０００、３０，０００、４０，０００、５０，０００、６０，０００、７０，０００、８０，０００、９０，０００、１００，０００またはそれよりも多くであり得る。他の場合では、入力層に使用されるノードの数は、最大で約１００，０００、９０，０００、８０，０００、７０，０００、６０，０００、５０，０００、４０，０００、３０，０００、２０，０００、１０，０００、９０００、８０００、７０００

、６０００、５０００、４０００、３０００、２０００、１０００、９００、８００、７００、６００、５００、４００、３００、２００、１００、５０、または１０またはそれ未満であり得る。一部の場合では、ニューラルネットワークに使用される層の総数（入力層および出力層を含める）は、少なくとも約３、４、５、１０、１５、２０、またはそれよりも多くであり得る。他の場合では、層の総数は最大で約２０、１５、１０、５、４、３またはそれ未満であり得る。

【０５１２】

一部の場合では、ニューラルネットワークに使用される学習可能または訓練可能パラメータ、例えば、重み付け因子、バイアス、または閾値の総数は少なくとも約１０、５０、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００、２０００、３０００、４０００、５０００、６０００、７０００、８０００、９０００、１０，０００、２０，０００、３０，０００、４０，０００、５０，０００、６０，０００、７０，０００、８０，０００、９０，０００、１００，０００またはそれよりも多くであり得る。他の場合では、学習可能パラメータの数は、最大で約１００，０００、９０，０００、８０，０００、７０，０００、６０，０００、５０，０００、４０，０００、３０，０００、２０，０００、１０，０００、９０００、８０００、７０００、６０００、５０００、４０００、３０００、２０００、１０００、９００、８００、７００、６００、５００、４００、３００、２００、１００、５０、または１０またはそれ未満であり得る。

【０５１３】

ニューラルネットワークは、畳み込みニューラルネットワークを含み得る。畳み込みニューラルネットワークは、１つまたは複数の畳み込み層、拡張型層または全結合層を含み得る。畳み込み層の数は１～１０の間であり得、拡張型層の数は０～１０の間であり得る。畳み込み層の総数（入力層および出力層を含める）は、少なくとも約１、２、３、４、５、１０、１５、２０、またはそれよりも多くであり得、拡張型層の総数は少なくとも約１、２、３、４、５、１０、１５、２０、またはそれよりも多くであり得る。畳み込み層の総数は最大で約２０、１５、１０、５、４、３またはそれ未満であり得、拡張型層の総数は最大で約２０、１５、１０、５、４、３またはそれ未満であり得る。一部の実施形態では、畳み込み層の数は１～１０の間であり、全結合層の数は０～１０の間である。畳み込み層の総数（入力層および出力層を含める）は少なくとも約１、２、３、４、５、１０、１５、２０、またはそれよりも多くであり得、全結合層の総数は少なくとも約１、２、３、４、５、１０、１５、２０、またはそれよりも多くであり得る。畳み込み層の総数は最大で約２０、１５、１０、５、４、３またはそれ未満であり得、全結合層の総数は最大で約２０、１５、１０、５、４、３またはそれ未満であり得る。

【０５１４】

畳み込みニューラルネットワーク（ＣＮＮ）は、ディープおよびフィードフォワード人工ニューラルネットワークであり得る。ＣＮＮは、視覚心像の解析に適用可能であり得る。ＣＮＮは、入力層、出力層、および多数の隠れ層を含み得る。ＣＮＮの隠れ層は、畳み込み層、プーリング層、全結合層および正規化層を含み得る。層は、３次元：幅、高さおよび深さで組織化することができる。

【０５１５】

畳み込み層では、入力に対して畳み込み操作を適用し、畳み込み操作の結果を次の層に渡すことができる。画像を処理するために、畳み込み操作により、フリーパラメータの数を減らし、ネットワークが少数のパラメータでより深くなることを可能にすることができる。畳み込み層では、ニューロンは、前の層の制限された下位領域のみから入力を受け取ることができる。畳み込み層のパラメータは、学習可能フィルター（またはカーネル）のセットを含み得る。学習可能フィルターは、小さな受容野を有し、入力体積の全深さを通じて広がる。フォワードパスの間、各フィルターを入力体積の幅および高さにわたって畳み込み、フィルターのエントリーと入力の間のドット積をコンピュータ計算し、そのフィルターの２次元活性化マップを作製することができる。結果として、ネットワークは、入

力のいくつかの空間的位置において特定の型の特色を検出した場合に活性化するフィルターを学習することができる。

【０５１６】

プーリング層は、グローバルプーリング層を含み得る。グローバルプーリング層では、１つの層におけるニューロンクラスターの出力を組み合わせて単一のニューロンにして次の層に入れることができる。例えば、最大プーリング層では、前の層におけるニューロンのクラスターの各々からの最大値を使用することができ、平均プーリング層では、前の層におけるニューロンのクラスターの各々からの平均値を使用することができる。全結合層では、１つの層内のあらゆるニューロンを別の層内のあらゆるニューロンと接続することができる。全結合層では、各ニューロンが前の層のあらゆる要素から入力を受け取ることができる。正規化層はバッチ正規化層であり得る。バッチ正規化層では、ニューラルネットワークの性能および安定性を改善することができる。バッチ正規化層では、ニューラルネットワーク内の任意の層にゼロ平均／単位分散である入力を提供することができる。バッチ正規化層を使用する利点としては、ネットワークの訓練がより速いこと、学習率がより高いこと、重みを初期化するのが容易であること、実行可能な活性化関数がより多いこと、およびディープネットワークを創出するプロセスがより単純であることを挙げることができる。

【０５１７】

ニューラルネットワークは、再帰型ニューラルネットワークを含み得る。再帰型ニューラルネットワークは、例えば連続したデータ入力など、逐次的データを入力として受け取るように構成することができ、再帰型ニューラルネットワークソフトウェアモジュールにより、あらゆる時間ステップにおいて内部状態を更新することができる。再帰型ニューラルネットワークでは、内部状態（メモリ）を使用して入力の順序をプロセシングできる。再帰型ニューラルネットワークは、手書き文字認識または音声認識、次の単語の予測、作曲、画像キャプション生成、時系列異常検知、機械翻訳、シーンラベリング、および株式市場予測などの課題に適用可能であり得る。再帰型ニューラルネットワークは、完全再帰型ニューラルネットワーク、独立型再帰型ニューラルネットワーク、エルマンネットワーク、ジョーダンネットワーク、エコー状態、ニューラルヒストリーコンプレッサ、長・短期記憶、ゲート付き再帰型ユニット、多重時間スケールモデル、ニューラルチューリングマシン、微分可能ニューラルコンピュータ、ニューラルネットワーク・プッシュダウン・オートマトン、またはこれらの任意の組合せを含み得る。

【０５１８】

訓練されたアルゴリズムは、例えば、ＳＶＭ、ランダムフォレスト、クラスタリングアルゴリズム（またはソフトウェアモジュール）、勾配ブースティング、ロジスティック回帰、および／または決定木などの教師ありまたは教師なし学習法を含み得る。教師あり学習アルゴリズムは、入力データと出力データの関係を推測するために、標識された対になった訓練用データ例のセットを使用することに依拠するアルゴリズムであり得る。教師なし学習アルゴリズムは、訓練用データセットから出力データに推測を引き出すために使用されるアルゴリズムであり得る。教師なし学習アルゴリズムは、処理データにおける隠れパターンまたは群分けを見いだすための探索的データ解析に使用することができるクラスター解析を含み得る。教師なし学習法の１つの例は、主成分分析を含み得る。主成分分析は、１つまたは複数の変数の次元を縮約することを含み得る。所与の変数の次元は、少なくとも1、５、１０、５０、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００、１１００、１２００ １３００、１４００、１５００、１６００、１７００、１８００、またはそれよりも大きいものであり得る。所与の変数の次元は、最大で１８００、１６００、１５００、１４００、１３００、１２００、１１００、１０００、９００、８００、７００、６００、５００、４００、３００、２００、１００、５０、１０またはそれ未満であり得る。

【０５１９】

訓練用アルゴリズムは、統計学的技法によって達成することができる。一部の実施形態

では、統計学的な技法は、線形回帰、分類、リサンプリング法、サブセット選択、縮小、次元縮約、非線形モデル、木に基づく方法、サポートベクターマシン、教師なし学習、またはこれらの任意の組合せを含み得る。

【０５２０】

線形回帰は、従属変数と独立変数の間に最良の直線関係を当てはめることによって標的変数を予測するための方法であり得る。最良当てはめは、各点における形状と実際の観察の間の全ての距離の合計が最小であることを意味し得る。線形回帰は、単純線形回帰および多重線形回帰を含み得る。単純線形回帰では、単一の独立変数を使用して従属変数を予測することができる。多重線形回帰では、１つよりも多くの独立変数を使用して、最良の直線関係を当てはめることによって、従属変数を予測することができる。

【０５２１】

分類は、正確な予測および解析を実現するために、データの集合にカテゴリーを割り当てるデータマイニング技法であり得る。分類技法は、ロジスティック回帰および判別分析を含み得る。ロジスティック回帰は、従属変数が二分法（バイナリ）である場合に使用することができる。ロジスティック回帰は、１つの従属バイナリ変数と１つまたは複数の名義、順序、間隔または比率－レベル独立変数の関係を発見し、記述するために使用することができる。リサンプリングは、元のデータサンプルから繰り返されたサンプルを引き出すことを含む方法であり得る。リサンプリングは、概算確率値をコンピュータ計算するために一般的な分布表の利用を伴わなくてよい。リサンプリングにより、実際のデータに基づいて独特のサンプリング分布を生成することができる。一部の実施形態では、リサンプリングには、独特のサンプリング分布を生成するために、分析的方法ではなく実験的方法を使用することができる。リサンプリング技法は、ブートストラッピングおよび交差検証を含み得る。ブートストラッピングは、元のデータからの、元に戻すことを伴うサンプリングによって実施することができ、「選択されなかった」データ点を試験事例として取る。訓練用データを複数の部分に分割することにより、交差検証を実施することができる。

【０５２２】

サブセット選択により、応答に関連する予測器のサブセットを同定することができる。サブセット選択は、最良のサブセット選択、フォワードステップワイズ選択、バックワードステップワイズ選択、ハイブリッド法、またはこれらの任意の組合せを含み得る。一部の実施形態では、縮小が、全ての予測器を伴うモデルに合うが、推定される係数は最小二乗推定値に対してゼロに向かって縮小する。この縮小により、分散が減少し得る。縮小は、リッジ回帰およびラッソを含み得る。次元縮約により、ｍ＜ｎとして、ｎ＋１係数を推定する問題がｍ＋１係数という、より単純な問題に縮小し得る。これは、変数のｎ個の異なる線形結合、または投影をコンピュータ計算することによって達成することができる。次いで、これらのｎ個の投影を予測器として使用して最小二乗によって線形回帰モデルに当てはめる。次元縮約は、主成分回帰および部分最小二乗を含み得る。主成分回帰を使用して、変数の大きなセットから特色の低次元セットを導き出すことができる。主成分回帰に使用される主成分により、引き続き直交方向にデータの線形結合を使用してデータの大多数の分散を捕捉することができる。部分最小二乗は主成分回帰に対する教師あり代替であり得、これは、部分最小二乗では、新しい特色を同定するために応答変数が使用され得るからである。

【０５２３】

非線形回帰は、観察に基づくデータが、モデルパラメータの非線形組合せであり、１つまたは複数の独立変数に依存する関数によってモデル化される、回帰分析の一形態であり得る。非線形回帰は、ステップ関数、区分的関数、スプライン、一般加法モデル、またはこれらの任意の組合せを含み得る。

【０５２４】

木に基づく方法は、回帰および分類問題のどちらにも使用することができる。回帰および分類問題は、予測器空間をいくつかの単純領域への層別化またはセグメント化することを伴い得る。木に基づく方法は、バギング、ブースティング、ランダムフォレスト、また

はこれらの任意の組合せを含み得る。バギングにより、反復との組合せを使用して元のデータと同じカーナリティ／サイズの多段階を生成して元のデータセットから訓練用の追加的なデータを生成することによって予測の分散を減少させることができる。ブースティングにより、いくつかの異なるモデルを使用して出力を算出し、次いで、加重平均手法を使用して結果を平均することができる。ランダムフォレストアルゴリズムにより、訓練セットのランダムなブートストラップサンプルを引き出すことができる。サポートベクターマシンは、分類技法であり得る。サポートベクターマシンは、２つのクラスの点を最大の余地で最良に分離する超平面を見いだすことを含み得る。サポートベクターマシンでは、最適化問題を制約することができ、したがって、データを完璧に分類するという制約に供される余地が最大化される。

【０５２５】

教師なし方法は、入力データを含むデータセットから、標識された応答を伴わずに推測を引き出す方法であり得る。教師なし方法は、クラスタリング、主成分分析、ｋ平均クラスタリング、階層クラスタリング、またはこれらの任意の組合せを含み得る。

【０５２６】

質量分析は、単一対立遺伝子質量分析であり得る。一部の実施形態では、質量分析は、ＭＳ分析、ＭＳ／ＭＳ分析、ＬＣ－ＭＳ／ＭＳ分析、またはこれらの組合せであり得る。一部の実施形態では、ＭＳ分析を使用して、インタクトなペプチドの質量を決定することができる。例えば、決定することは、インタクトなペプチドの質量を決定すること（例えば、ＭＳ分析）を含み得る。一部の実施形態では、ＭＳ／ＭＳ分析を使用して、ペプチド断片の質量を決定することができる。例えば、決定することは、ペプチド断片の質量を決定することを含み得、それを使用して、ペプチドまたはその一部のアミノ酸配列を決定することができる（例えば、ＭＳ／ＭＳ分析）。一部の実施形態では、ペプチド断片の質量を使用して、ペプチド内のアミノ酸の配列を決定することができる。一部の実施形態では、ＬＣ－ＭＳ／ＭＳ分析を使用して、複雑なペプチド混合物を分離することができる。例えば、決定することは、液体クロマトグラフィーによってなどで複雑なペプチド混合物を分離すること、およびインタクトなペプチドの質量、ペプチド断片の質量、またはこれらの組合せを決定することを含み得る（例えば、ＬＣ－ＭＳ／ＭＳ分析）。このデータを、例えばペプチド配列決定のために使用することができる。

【０５２７】

ペプチドは、オートファジーを通じて、細胞において発現されるＨＬＡタンパク質によって提示され得る。オートファジーにより、細胞構成成分の順序正しい分解および再利用が可能になり得る。オートファジーは、マクロオートファジー、ミクロオートファジーおよびシャペロン媒介性オートファジーを含み得る。ペプチドは、ファゴサイトーシスを通じて、細胞において発現されるＨＬＡタンパク質によって提示され得る。ファゴサイトーシスは、病原体および細胞片を除去するために使用される主要な機構であり得る。例えば、マクロファージが病原性微生物を摂取すると、病原体はファゴソーム内に捕捉され、次いで、それがリソソームと融合して、ファゴリソソームを形成する。ＨＬＡクラスＩＩでは、マクロファージおよび未成熟樹状細胞などの食細胞が実体をファゴサイトーシスによってファゴソームに取り込み得（しかし、Ｂ細胞はエンドソームへのより一般的なエンドサイトーシスを示すが）、それらはリソソームと融合し、リソソームの酸性酵素により取り込まれたタンパク質が切断されて多くの異なるペプチドになる。

【０５２８】

訓練用データの品質は、複数の品質メトリクスを使用して向上させることができる。複数の品質メトリクスは、一般的な夾雑ペプチド除去、高スコア化ピーク強度、高スコア、および高い質量精度を含み得る。スコア化ピーク強度を、スコア化を実施する前に使用することができる。ＭＳ／ＭＳ検索により、まず、単純なフィルターを使用してＭＳ／ＭＳスペクトルを候補配列に対してスクリーニングする。このフィルターは最小のスコア化ピーク強度であり得る。スコア化ピーク強度の使用により、十分な数のスペクトルピークが調査され、このフィルターによって確立された閾値を満たさないことが見いだされたら、

候補配列を迅速かつ即座にはねることが可能になることにより、検索スピードが増強され得る。スコア化ピーク強度は少なくとも５０％であり得る。スコア化ピーク強度は少なくとも７０％であり得る。スコア化ピーク強度は少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％またはそれよりも大きいものであり得る。一部の場合では、スコア化ピーク強度は最大で９０％、８０％、７０％、６０％、５０％、４０％、３０％、２０％、１０％またはそれ未満であり得る。スコアは少なくとも７であり得る。スコアは、少なくとも１、２、３、４、５、６、７、８、９、１０、１５、２０またはそれよりも大きいものであり得る。一部の場合では、スコアは、最大で約２０、１５、１０、９、８、７、６、５、４、３、２、１またはそれ未満であり得る。質量精度は最大で５ｐｐｍであり得る。質量精度は最大で１０ｐｐｍ、９ｐｐｍ、８ｐｐｍ、７ｐｐｍ、６ｐｐｍ、５ｐｐｍ、４ｐｐｍ、３ｐｐｍ、２ｐｐｍ、１ｐｐｍまたはそれ未満であり得る。質量精度は少なくとも１ｐｐｍ、２ｐｐｍ、３ｐｐｍ、４ｐｐｍ、５ｐｐｍ、６ｐｐｍ、７ｐｐｍ、８ｐｐｍ、９ｐｐｍ、１０ｐｐｍまたはそれよりも大きいものであり得る。

【０５２９】

一部の実施形態では、質量精度は、最大で２ｐｐｍである。一部の実施形態では、骨格切断スコアは、少なくとも５である。一部の実施形態では、骨格切断スコアは、少なくとも８である。

【０５３０】

細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の免疫沈降ＨＬＡタンパク質によって提示されるペプチドであり得る。免疫沈降（ＩＰ）は、タンパク質抗原を溶液から、その特定のタンパク質に特異的に結合する抗体を使用して沈殿させる技法であり得る。このプロセスを使用して、多くの何千もの異なるタンパク質を含有する試料から特定のタンパク質を単離し、濃縮することができる。免疫沈降では、手順のある時点で抗体を固体基質にカップリングすることが必要になり得る。

【０５３１】

細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の外因性ＨＬＡタンパク質によって提示されるペプチドであり得る。単一の外因性ＨＬＡタンパク質は、１つまたは複数の外因性ペプチドを細胞の集団に導入することによって創出することができる。一部の実施形態では、導入することは、細胞の集団を１つもしくは複数の外因性ペプチドと接触させること、または細胞の集団において１つもしくは複数の外因性ペプチドを発現させることを含む。一部の実施形態では、導入することは、細胞の集団を１つまたは複数の外因性ペプチドをコードする１つまたは複数の核酸と接触させることを含む。一部の実施形態では、１つまたは複数のペプチドをコードする１つまたは複数の核酸はＤＮＡである。一部の実施形態では、１つまたは複数のペプチドをコードする１つまたは複数の核酸はＲＮＡであり、必要に応じて、ＲＮＡはｍＲＮＡである。一部の実施形態では、富化することは、四量体（または多量体）試薬の使用を含まない。

【０５３２】

細胞において発現されるＨＬＡタンパク質によって提示されるペプチドは、細胞において発現される単一の組換えＨＬＡタンパク質によって提示されるペプチドであり得る。組換えＨＬＡタンパク質は、組換えＨＬＡクラスＩまたはＨＬＡクラスＩＩ対立遺伝子によってコードされるものであり得る。ＨＬＡクラスＩは、ＨＬＡ－Ａ、ＨＬＡ－Ｂ、ＨＬＡ－Ｃからなる群から選択することができる。ＨＬＡクラスＩは、非古典的クラス－Ｉ－ｂ群であり得る。ＨＬＡクラスＩは、ＨＬＡ－Ｅ、ＨＬＡ－Ｆ、およびＨＬＡ－Ｇからなる群から選択することができる。ＨＬＡクラスＩは、ＨＬＡ－Ｅ、ＨＬＡ－Ｆ、およびＨＬＡ－Ｇからなる群から選択される非古典的クラス－Ｉ－ｂ群であり得る。一部の実施形態では、ＨＬＡクラスＩＩは、ＨＬＡクラスＩＩ　α鎖、ＨＬＡクラスＩＩ　β鎖、またはこれらの組合せを含む。

【０５３３】
複数の予測変数は、ペプチド－ＨＬＡ親和性予測変数を含み得る。複数の予測変数は、供給源タンパク質発現レベル予測変数を含み得る。供給源タンパク質発現レベルは、細胞内のペプチドの供給源タンパク質の発現レベルであり得る。一部の実施形態では、発現レベルは、供給源タンパク質の量または供給源タンパク質をコードするＲＮＡの量を測定することによって決定することができる。複数の予測変数は、ペプチド配列、アミノ酸の物理的性質、ペプチドの物理的性質、細胞内のペプチドの供給源タンパク質の発現レベル、タンパク質の安定性、タンパク質の翻訳率、ユビキチン化部位、タンパク質分解速度、リボソームのプロファイリングからの翻訳効率、タンパク質切断性、タンパク質の局在化、ＴＡＰ輸送を容易にする宿主タンパク質のモチーフ、オートファジーに供される宿主タンパク質、リボソームの停止に有利なモチーフ（例えば、ポリプロリンまたはポリリシンのひと続き）、ＮＭＤに有利なタンパク質の特色（例えば、長い３’ＵＴＲ、最後のエクソンの＞５０ｎｔ上流の終止コドン：エクソン接合部およびペプチド切断性）を含み得る。
【０５３４】
複数の予測変数は、ペプチド切断性予測変数を含み得る。ペプチド切断性は、切断可能なリンカーまたは切断配列に関連し得る。一部の実施形態では、切断可能なリンカーは、リボソームスキッピング部位または配列内リボソーム進入部位（ＩＲＥＳ）エレメントである。一部の実施形態では、リボソームスキッピング部位またはＩＲＥＳは細胞において発現された時に切断される。一部の実施形態では、リボソームスキッピング部位は、Ｆ２Ａ、Ｔ２Ａ、Ｐ２Ａ、およびＥ２Ａからなる群から選択される。一部の実施形態では、ＩＲＥＳエレメントは、一般的な細胞ＩＲＥＳ配列またはウイルスＩＲＥＳ配列から選択される。Ｆ２Ａ、または配列内リボソーム進入部位（ＩＲＥＳ）などの切断配列を、α鎖とβ２－ミクログロブリン（ＨＬＡクラスＩ）の間、またはα鎖とβ鎖（ＨＬＡクラスＩＩ）の間に配置することができる。一部の実施形態では、単一のＨＬＡクラスＩ対立遺伝子は、ＨＬＡ－Ａ＊０２：０１、ＨＬＡ－Ａ＊２３：０１およびＨＬＡ－Ｂ＊１４：０２、またはＨＬＡ－Ｅ＊０１：０１であり、ＨＬＡクラスＩＩ対立遺伝子は、ＨＬＡ－ＤＲＢ＊０１：０１、ＨＬＡ－ＤＲＢ＊０１：０２およびＨＬＡ－ＤＲＢ＊１１：０１、ＨＬＡ－ＤＲＢ＊１５：０１、またはＨＬＡ－ＤＲＢ＊０７：０１である。一部の実施形態では、切断配列は、Ｔ２Ａ、Ｐ２Ａ、Ｅ２Ａ、またはＦ２Ａ配列である。例えば、切断配列は、ＥＧＲＧＳＬＴＣＧＤＶＥＮＰＧＰ（配列番号６）（Ｔ２Ａ）、ＡＴＮＦＳＬＫＱＡＧＤＶＥＮＰＧＰ（配列番号７）（Ｐ２Ａ）、ＱＣＴＮＹＡＬＫＬＡＧＤＶＥＳＮＰＧＰ（配列番号８）（Ｅ２Ａ）、またはＶＫＱＴＬＮＦＤＬＫＬＡＧＤＶＥＳＮＰＧＰ（配列番号９）（Ｆ２Ａ）であり得る。
【０５３５】
一部の実施形態では、切断配列は、トロンビン切断部位ＣＬＩＰであり得る。
【０５３６】
ＨＬＡタンパク質によって提示されるペプチドは、酵素特異性なし修飾なしペプチドデータベースを検索することによって同定されるペプチドを含み得る。ペプチドデータベースは、修飾なしデータベースまたは修飾あり（例えば、リン酸化またはシステイニル化）データベースなどの酵素特異性なしペプチドデータベースであり得る。一部の実施形態では、ペプチドデータベースは、ポリペプチドデータベースである。一部の実施形態では、ポリペプチドデータベースは、タンパク質データベースであり得る。一部の実施形態では、方法は、逆データベース検索戦略を使用してペプチドデータベースを検索することをさらに含む。一部の実施形態では、方法は、逆データベース検索戦略を使用してタンパク質データベースを検索することをさらに含む。一部の実施形態では、例えば、通常のペプチドまたはタンパク質データベースには含まれない新しいペプチドを発見するために、新規の検索を実施する。ペプチドデータベースは、親和性アクセプターでタグ付けされたＨＬＡを含む１つまたは複数の細胞をそれぞれが含む第１の細胞の集団および第２の細胞の集団を準備するステップであって、配列親和性アクセプターでタグ付けされたＨＬＡが、親

和性アクセプターペプチドに作動可能に連結した異なるＨＬＡ対立遺伝子によってコードされる異なる組換えポリペプチドを含む、ステップと、親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体を富化するステップと、富化するステップからの、親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体に結合したペプチドまたはその一部を特徴付けるステップと、ＨＬＡ－対立遺伝子特異的ペプチドデータベースを生成するステップとによって生成することができる。

【０５３７】

ＨＬＡタンパク質によって提示されるペプチドは、ＨＬＡ－ペプチドのＭＳ／ＭＳスペクトルとペプチドデータベース内の１つまたは複数のＨＬＡ－ペプチドのＭＳ／ＭＳスペクトルと比較することによって同定されるペプチドを含み得る。

【０５３８】

ペプチドまたはペプチドをコードする核酸のいずれかに変異が存在し得る。変異は、点変異、スプライス部位変異、フレームシフト変異、リードスルー変異、および遺伝子融合変異からなる群から選択することができる。点変異は、ＤＮＡまたはＲＮＡの配列から単一のヌクレオチド塩基が変化した、挿入されたまたは欠失した遺伝子変異であり得る。スプライス部位変異は、前駆体メッセンジャーＲＮＡの成熟メッセンジャーＲＮＡへのプロセシングの間にスプライシングが生じる特定の部位においていくつかのヌクレオチドが挿入された、欠失した、または変化した遺伝子変異であり得る。フレームシフト変異は、ＤＮＡ配列における３で割り切れないいくつかのヌクレオチドのインデル（挿入または欠失）によって引き起こされる遺伝子変異であり得る。変異はまた、挿入、欠失、置換変異、遺伝子重複、染色体転座、および染色体逆位も含み得る。

【０５３９】

一部の実施形態では、ＨＬＡクラスＩＩタンパク質は、ＨＬＡ－ＤＲタンパク質を含む。

【０５４０】

一部の実施形態では、ＨＬＡクラスＩＩタンパク質は、ＨＬＡ－ＤＰタンパク質を含む。

【０５４１】

一部の実施形態では、ＨＬＡクラスＩＩタンパク質は、ＨＬＡ－ＤＱタンパク質を含む。

【０５４２】

一部の実施形態では、ＨＬＡクラスＩＩタンパク質は、ＨＬＡ－ＤＲタンパク質、およびＨＬＡ－ＤＰタンパク質またはＨＬＡ－ＤＱタンパク質からなる群から選択することができる。一部の実施形態では、ＨＬＡタンパク質は、ＨＬＡ－ＤＰＢ１＊０１：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０２：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０３：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０２／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０６：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＱＢ１＊０２：０１／ＨＬＡ－ＤＱＡ１＊０５：０１、ＨＬＡ－ＤＱＢ１＊０２：０２／ＨＬＡ－ＤＱＡ１＊０２：０１、ＨＬＡ－ＤＱＢ１＊０６：０２／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＱＢ１＊０６：０４／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０１：０１、ＨＬＡ－ＤＲＢ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０３：０１、ＨＬＡ－ＤＲＢ１＊０３：０２、ＨＬＡ－ＤＲＢ１＊０４：０１、ＨＬＡ－ＤＲＢ１＊０４：０２、ＨＬＡ－ＤＲＢ１＊０４：０３、ＨＬＡ－ＤＲＢ１＊０４：０４、ＨＬＡ－ＤＲＢ１＊０４：０５、ＨＬＡ－ＤＲＢ１＊０４：０７、ＨＬＡ－ＤＲＢ１＊０７：０１、ＨＬＡ－ＤＲＢ１＊０８：０１、ＨＬＡ－ＤＲＢ１＊０８：０２、ＨＬＡ－ＤＲＢ１＊０８：０３、ＨＬＡ－ＤＲＢ１＊０８：０４、ＨＬＡ－ＤＲＢ１＊０９：０１、ＨＬＡ－ＤＲＢ１＊１０：０１、ＨＬＡ－ＤＲＢ１＊１１：０１、ＨＬＡ－ＤＲＢ１＊１１：０２、ＨＬＡ－ＤＲＢ１＊１１：０４、ＨＬＡ－ＤＲＢ１＊１２：０１、ＨＬＡ－ＤＲＢ１＊１２：０２、ＨＬＡ－ＤＲＢ１＊１３：０１、ＨＬＡ－ＤＲＢ１＊１３：０２、ＨＬＡ－ＤＲＢ１＊１３：０３、ＨＬＡ－ＤＲＢ１＊１４：０１、ＨＬＡ－ＤＲＢ１＊１５：０１、ＨＬＡ－ＤＲＢ１＊１５：０２、ＨＬＡ－ＤＲＢ１＊１５：０３、ＨＬＡ－ＤＲＢ１＊１６：０１、ＨＬＡ－ＤＲＢ３＊０１：０１、ＨＬＡ－ＤＲＢ３＊０２：０２、ＨＬＡ－ＤＲＢ３＊０３

：０１、ＨＬＡ－ＤＲＢ４＊０１：０１、ＨＬＡ－ＤＲＢ５＊０１：０１）からなる群から選択されるＨＬＡクラスＩＩタンパク質である。ＨＬＡタンパク質によって提示されるペプチドの長さは、１５～４０アミノ酸であり得る。ＨＬＡタンパク質によって提示されるペプチドの長さは、少なくとも１、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アミノ酸、またはそれよりも多くのアミノ酸であり得る。一部の実施形態では、ＨＬＡタンパク質によって提示されるペプチドの長さは、最大で３０、２９、２８、２７、２６、２５、２４、２３、２２、２１、２０、１９、１８、１７、１６、１５、１４、１３、１２、１１アミノ酸、またはそれ未満であり得る。

【０５４３】

ＨＬＡタンパク質によって提示されるペプチドは、（ａ）単一のＨＬＡクラスＩＩ対立遺伝子を発現する細胞株から１つまたは複数のＨＬＡ複合体を単離するステップと、（ｂ）１つまたは複数の単離されたＨＬＡ複合体から１つまたは複数のＨＬＡ－ペプチドを単離するステップと、（ｃ）１つまたは複数の単離されたＨＬＡ－ペプチドについてのＭＳ／ＭＳスペクトルを得るステップと、（ｄ）ペプチドデータベースから１つまたは複数の単離されたＨＬＡ－ペプチドのＭＳ／ＭＳスペクトルに対応するペプチド配列を得るステップであって、１つまたは複数の配列が、ステップ（ａ、ｂ、ｃ）から得られたものである、ステップと、（ｄ）１つまたは複数の単離されたＨＬＡ－ペプチドの配列を同定するステップとによって同定されるペプチドを含み得る。

【０５４４】

単離するステップは、アフィニティータグが付されたＨＬＡ構築物をトランスフェクトまたは形質導入した細胞からＨＬＡ－ペプチド複合体を単離することを含み得る。一部の実施形態では、複合体を、市販の抗体を用い、当技術分野で公知の標準の免疫沈降技法を使用して単離することができる。細胞をまず溶解させることができる。ＨＬＡクラスＩＩ－ペプチド複合体を、Ｍ５／１１４．１５．２モノクローナル抗体などのＨＬＡクラスＩＩ特異的抗体を使用して単離することができる。一部の実施形態では、単一の（またはその対）ＨＬＡ対立遺伝子をペプチドタグとの融合タンパク質として発現させ、ＨＬＡ－ペプチド複合体を、ペプチドタグを認識する結合性分子を使用して単離する。

【０５４５】

単離するステップは、ペプチドをＨＬＡ－ペプチド複合体から単離し、ペプチドの配列決定を行うことを含み得る。酸溶出などの当業者に公知の任意の方法によって複合体からペプチドを単離する。任意の配列決定方法を使用することができるが、一部の実施形態では、質量分析を使用する方法、例えば、液体クロマトグラフィー－質量分析（ＬＣ－ＭＳまたはＬＣ－ＭＳ／ＭＳ、またはその代わりにＨＰＬＣ－ＭＳまたはＨＰＬＣ－ＭＳ／ＭＳ）などが利用される。これらの配列決定方法は当業者には周知であり得、また、Medzihradszky KF and Chalkley RJ. Mass Spectrom Rev. 2015 Jan-Feb; 34 (1): 43-63に概説されている。

【０５４６】

単離または精製に適した追加的な候補構成成分および分子は、ビオチン（ビオチン－アビジン特異的結合対）、抗体、受容体、リガンド、レクチン、または、例えば、プラスチックまたはポリスチレンビーズ、プレートまたはビーズ、磁気ビーズ、テストストリップ、および膜を含めた固体支持体を含む分子などの結合性分子を含み得る。コンジュゲートがそれらの種々の分子量に有効に分離される電荷差異によってコンジュゲートを分離するために、陽イオン交換クロマトグラフィーなどの精製方法を使用することができる。陽イオン交換クロマトグラフィーによって得られた画分の内容物は、従来の方法、例えば、質量分析、ＳＤＳ－ＰＡＧＥ、または分子実体を分子量によって分離するための他の公知の方法を使用して、分子量によって同定することができる。

【０５４７】

一部の実施形態では、方法は、特徴付けるステップの前にペプチドを親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体から単離するステップをさらに含む。一部の

実施形態では、ＨＬＡ－ペプチド複合体を、抗ＨＬＡ抗体を使用して単離する。一部の場合では、アフィニティータグを伴うまたは伴わないＨＬＡ－ペプチド複合体を、抗ＨＬＡ抗体を使用して単離する。一部の場合では、アフィニティータグを伴うまたは伴わない可溶性ＨＬＡ（ｓＨＬＡ）を細胞培養の培地から単離する。一部の場合では、アフィニティータグを伴うまたは伴わない可溶性ＨＬＡ（ｓＨＬＡ）を、抗ＨＬＡ抗体を使用して単離する。例えば、アフィニティータグを伴うまたは伴わない可溶性ＨＬＡ（ｓＨＬＡ）などのＨＬＡを、抗ＨＬＡ抗体を含有するビーズまたはカラムを使用して単離することができる。一部の実施形態では、ペプチドを、抗ＨＬＡ抗体を使用して単離する。一部の場合では、アフィニティータグを伴うまたは伴わない可溶性ＨＬＡ（ｓＨＬＡ）を、抗ＨＬＡ抗体を使用して単離する。一部の場合では、アフィニティータグを伴うまたは伴わない可溶性ＨＬＡ（ｓＨＬＡ）を、抗ＨＬＡ抗体を含有するカラムを使用して単離する。一部の実施形態では、方法は、親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体に結合したペプチドの末端から１つまたは複数のアミノ酸を除去するステップをさらに含む。

【０５４８】

個別化がんワクチンは、アジュバントをさらに含み得る。例えば、ＴＬＲ３ならびにＭＤＡ５およびＲＩＧ３のＲＮＡヘリカーゼ－ドメインのアゴニストであるポリ－ＩＣＬＣは、ワクチンアジュバントとしてのいくつかの望ましい特性が示されている。これらの特性としては、ｉｎ　ｖｉｖｏにおける免疫細胞の局所的な活性化および全身性の活性化の誘導、刺激性ケモカインおよびサイトカインの産生、ならびにＤＣによる抗原提示の刺激を挙げることができる。さらに、ポリ－ＩＣＬＣにより、ヒトにおける長続きするＣＤ４＋およびＣＤ８＋応答を誘導することができる。重要なことに、ポリ－ＩＣＬＣを用いたワクチン接種を受けた対象と、高度に有効な複製コンピテント黄熱病ワクチンを受けた志願者において、転写およびシグナルトランスダクション経路の上方制御の著しい類似性を認めることができる。さらに、最近の第１相試験において、ポリ－ＩＣＬＣとＮＹＥＳＯ－１ペプチドワクチン（モンタニドに加えて）の組合せを用いて免疫化した卵巣癌患者の＞９０％でＣＤ４＋　Ｔ細胞およびＣＤ８＋　Ｔ細胞の誘導、ならびにペプチドに対する抗体応答が示された。

【０５４９】

個別化がんワクチンは、免疫チェックポイント阻害剤をさらに含み得る。免疫チェックポイント阻害剤は、Ｔ細胞などの一部の型の免疫系細胞、および一部のがん細胞によって作られるある特定のタンパク質を遮断する薬物の１つの型を含み得る。これらのタンパク質は、免疫応答を制止するのに役立ち、また、Ｔ細胞によるがん細胞の死滅を防ぐことができる。これらのタンパク質が遮断されると、免疫系に対する「ブレーキ」が放出され、Ｔ細胞がより良好にがん細胞を死滅させることができるようになる。Ｔ細胞またはがん細胞に見いだされるチェックポイントタンパク質の例としては、ＰＤ－１／ＰＤ－Ｌ１およびＣＴＬＡ－４／Ｂ７－１／Ｂ７－２が挙げられる。一部の免疫チェックポイント阻害剤は、がんを処置するために使用される。

【０５５０】

訓練用データは、構造化されたデータ、時系列データ、構造化されていないデータ、およびリレーショナルデータをさらに含み得る。構造化されていないデータは、ロボット工学またはシミュレーションの正確なシミュレートまたは訓練における使用のための、音声データ、画像データ、ビデオ、機械的データ、電気的データ、化学的データ、およびこれらの任意の組合せを含み得る。時系列データは、スマートメーター、スマート電化製品、スマートデバイス、モニタリングシステム、遠隔測定デバイス、またはセンサーのうちの１つまたは複数からのデータを含み得る。リレーショナルデータは、顧客システム、企業システム、運用システム、ウェブサイト、ウェブアクセス可能なアプリケーションプログラムインターフェース（ＡＰＩ）、またはこれらの任意の組合せからのデータを含む。これは、ユーザーが、ファイルまたは他のデータフォーマットをソフトウェアまたはシステムに入力する任意の方法によって行うことができる。

【０５５１】

訓練用データをクラウドに基づくデータベースにアップロードすることができる。クラウドに基づくデータベースは、機械学習に基づくセンサー信号処理アルゴリズムを実行するローカルおよび／またはリモートコンピュータシステムからアクセス可能なものであり得る。クラウドに基づくデータベースおよび関連するソフトウェアを、電子データをアーカイブに入れるため、電子データを共有するため、および電子データを解析するために使用することができる。ローカルで生成したデータまたはデータセットを、それがアクセスすることができるクラウドに基づくデータベースにアップロードし、同じサイトまたは異なるサイトにある他の機械学習に基づく検出システムを訓練するために使用することができる。センサーデバイスおよび検出システムの試験性能を連続して改善するために、ローカルで生成したセンサーデバイスおよびシステムによる試験結果をクラウドに基づくデータベースにアップロードし、訓練用データセットをリアルタイムで更新するために使用することができる。

【０５５２】

訓練は、畳み込みニューラルネットワークを使用して実施することができる。畳み込みニューラルネットワーク（ＣＮＮ）は本明細書の他の箇所に記載されている。畳み込みニューラルネットワークは、少なくとも２つの畳み込み層を含み得る。畳み込み層の数は１～１０の間であり、拡張型層の数は０～１０の間である。畳み込み層の総数（入力層および出力層を含める）は、少なくとも約１、２、３、４、５、１０、１５、２０、またはそれよりも多くであり得、拡張型層の総数は少なくとも約１、２、３、４、５、１０、１５、２０、またはそれよりも多くであり得る。畳み込み層の総数は最大で約２０、１５、１０、５、４、３またはそれ未満であり得、拡張型層の総数は最大で約２０、１５、１０、５、４、３またはそれ未満であり得る。一部の実施形態では、畳み込み層の数は１～１０の間であり、全結合層の数は０～１０の間である。畳み込み層の総数（入力層および出力層を含める）は、少なくとも約１、２、３、４、５、１０、１５、２０、またはそれよりも多くであり得、全結合層の総数は少なくとも約１、２、３、４、５、１０、１５、２０、またはそれよりも多くであり得る。畳み込み層の総数は最大で約２０、１５、１０、５、４、３またはそれ未満であり得、全結合層の総数は最大で約２０、１５、１０、５、４、３またはそれ未満であり得る。

【０５５３】

畳み込みニューラルネットワークは、少なくとも１つのバッチ正規化ステップを含み得る。バッチ正規化層により、ニューラルネットワークの性能および安定性を改善することができる。バッチ正規化層では、ニューラルネットワーク内の任意の層にゼロ平均／単位分散である入力をもたらすことができる。バッチ正規化層の総数は、少なくとも約３、４、５、１０、１５、２０またはそれよりも多くであり得る。バッチ正規化層の総数は、最大で約２０、１５、１０、５、４、３またはそれ未満であり得る。

【０５５４】

畳み込みニューラルネットワークは、少なくとも１つの空間ドロップアウトステップを含み得る。空間ドロップアウトステップの総数は、少なくとも約３、４、５、１０、１５、２０またはそれよりも多くであり得、また、空間ドロップアウトステップの総数は、最大で約２０、１５、１０、５、４、３またはそれ未満であり得る。

【０５５５】

畳み込みニューラルネットワークは、少なくとも１つグローバル最大プーリングステップを含み得る。グローバルプーリング層では、１つの層におけるニューロンクラスターの出力を組み合わせて単一のニューロンにして次の層に入れることができる。例えば、最大プーリング層では、前の層におけるニューロンのクラスターの各々からの最大値を使用することができ、平均プーリング層では、前の層におけるニューロンのクラスターの各々からの平均値を使用することができる。畳み込みニューラルネットワークは、少なくとも約１、２、３、４、５、１０、１５、２０、またはそれよりも多くのグローバル最大プーリングステップを含み得る。畳み込みニューラルネットワークは、最大で約２０、１５、１０、５、４、３またはそれ未満のグローバル最大プーリングステップを含み得る。

【０５５６】
畳み込みニューラルネットワークは、少なくとも１つの高密度層を含み得る。畳み込みニューラルネットワークは、少なくとも約１、２、３、４、５、１０、１５、２０、またはそれよりも多くの高密度層を含み得る。畳み込みニューラルネットワークは、最大で約２０、１５、１０、５、４、３またはそれ未満の高密度層を含み得る。
治療方法
【０５５７】
腫瘍特異的ペプチドを使用した個別化免疫治療が記載されている。悪性細胞内の遺伝子変化（例えば、逆位、転座、欠失、ミスセンス変異、スプライス部位変異など）の結果として生じる腫瘍新抗原は、最も腫瘍特異的なクラスの抗原を表す。新抗原は、同定、最適化された抗原の選択、およびワクチンまたは免疫原性組成物に使用するための新抗原の産生が技術的に難しいことから、がんワクチンまたは免疫原性組成にはめったに使用されていない。免疫原として利用する特定のペプチドの効率的な選択には、いずれの腫瘍特異的ペプチドが患者に存在するＨＬＡ対立遺伝子に効率的に結合するかおよび抗腫瘍免疫の誘導のために患者の免疫系に有効に提示されるかを予測できることが必要である。治癒的かつ腫瘍特異的免疫治療を開発するための極めて重要な関門の１つは、自己免疫を回避するために、高度に特異的かつ制限された腫瘍抗原を同定および選択することである。これは、免疫治療のための候補腫瘍特異的ペプチドがＭＨＣクラスＩＩ抗原によって提示されるものである場合に、ＭＨＣクラスＩＩ－ペプチド結合および免疫系への提示にある特定のレベルの柔軟なプロミスキャス性が存在するので、特に重要である。同時に、ＭＨＣクラスＩＩにより提示されるペプチドは、細胞傷害性細胞の活性化だけでなくＣＤ４＋ｖｅメモリーＴ細胞の活性化にも必要である。したがって、ＭＨＣクラスＩＩ媒介性免疫原性応答は、頑強な、腫瘍からの保護における大きな効果のための長期の免疫原性をもたらすために必要である。これらの問題には、信頼できるペプチド－ＭＨＣ予測アルゴリズムを有すること、ならびにペプチド－ＭＨＣ相互作用および免疫原性をアッセイおよび検証するための信頼できるシステムを有することによって対処することができる。したがって、一部の実施形態では、高度に効率的かつ免疫原性のがんワクチンを、ＤＮＡレベルで腫瘍には存在するが高い割合のがんを有する対象由来の対応する生殖細胞系列試料には存在しない、新形成／腫瘍における候補変異を同定し、同定された変異を１つまたは複数のペプチド－ＭＨＣ結合予測アルゴリズムを用いて解析して、いずれのＭＨＣ（ヒト白血球抗原またはヒトの場合ＨＬＡ）が高い割合の患者ＨＬＡ対立遺伝子に結合するかを同定し、高い割合のがんを有する対象を処置するために適したがんワクチンまたは免疫原性組成物への使用のための、全ての新抗原ペプチドおよび予測された結合性ペプチドのセットから選択された複数の新抗原性ペプチドを合成することによって作製することができる。
【０５５８】
例えば、ペプチド配列決定情報を治療ワクチンに変換することは、高い割合の個体のＨＬＡペプチドに結合し得る変異ペプチドの予測を含み得る。免疫原として利用する特定の変異の効率的な選択には、いずれの変異ペプチドが高い割合の患者のＨＬＡ対立遺伝子に効率的に結合するかを予測できることが必要である。最近、検証された結合性ペプチドおよび非結合性ペプチドを用いたニューラルネットワークに基づく学習手法により、主要なＨＬＡ－ＡおよびＨＬＡ－Ｂ対立遺伝子に関する予測アルゴリズムの正確度が進歩した。しかし、進歩したニューラルネットワークに基づくアルゴリズムの使用はＨＬＡ－ペプチド結合規則を符号化することには役立っているが、いくつかの因子により、ＨＬＡ対立遺伝子に提示されるペプチドの予測力が限定されている。
【０５５９】
例えば、ペプチド配列決定情報を治療ワクチンに変換することは、薬物を長いペプチドの多エピトープワクチンとして製剤化することを含み得る。事実上できる限り多くの変異エピトープを標的とすることにより、免疫系の巨大な能力が活用され、免疫標的化遺伝子産物のダウンモジュレーションによって免疫学的エスケープの機会が防止され、エピトープ予測手法の公知の不正確さが補償される。合成ペプチドにより、多数の免疫原を効率的

に調製するため、および変異体エピトープの同定を有効なワクチンに迅速に変換するための有用な手段がもたらされる。ペプチドは、細菌または動物性物質が夾雑していない試薬を利用して容易に化学合成することおよび容易に精製することができる。サイズが小さいことにより、タンパク質の変異領域に明確に焦点を当てることが可能になり、また、他の構成成分（変異していないタンパク質またはウイルスベクター抗原）に由来する無関連の抗原性競合も低減する。

【０５６０】

例えば、ペプチド配列決定情報を治療ワクチンに変換することは、強力なワクチンアジュバントとの組合せを含み得る。有効なワクチンには、免疫応答を開始するために強力なアジュバントが必要になり得る。例えば、ＴＬＲ３ならびにＭＤＡ５およびＲＩＧ３のＲＮＡヘリカーゼ－ドメインのアゴニストであるポリ－ＩＣＬＣは、ワクチンアジュバントとしてのいくつかの望ましい特性が示されている。これらの特性としては、ｉｎ　ｖｉｖｏにおける免疫細胞の局所的な活性化および全身性の活性化の誘導、刺激性ケモカインおよびサイトカインの産生、ならびにＤＣによる抗原提示の刺激が挙げられる。さらに、ポリ－ＩＣＬＣにより、ヒトにおける長続きするＣＤ４＋およびＣＤ８＋応答を誘導することができる。重要なことに、ポリ－ＩＣＬＣを用いたワクチン接種を受けた対象と高度に有効な複製コンピテント黄熱病ワクチンを受けた志願者で転写およびシグナルトランスダクション経路の上方制御の著しい類似性が認められた。さらに、最近の第１相試験において、ポリ－ＩＣＬＣとＮＹＥＳＯ－１ペプチドワクチン（モンタニドに加えて）の組合せを用いて免疫化した卵巣癌患者の＞９０％でＣＤ４＋　Ｔ細胞およびＣＤ８＋　Ｔ細胞の誘導、ならびにペプチドに対する抗体応答が示された。同時に、ポリ－ＩＣＬＣは現在までに２５件を超える臨床試験において広範囲にわたって試験されており、比較的良性の毒性プロファイルが示されている。

【０５６１】

一部の実施形態では、免疫原性ペプチドを疾患または状態を有する対象由来の細胞から同定することができる。一部の実施形態では、免疫原性ペプチドは、疾患または状態を有する対象に対して特異的なものであり得る。一部の実施形態では、免疫原性ペプチドは、疾患または状態を有する対象のＨＬＡハプロタイプと適合するＨＬＡに結合し得る。

【０５６２】

一部の実施形態では、ペプチドのライブラリーを細胞において発現させることができる。一部の実施形態では、細胞は、同定されるまたは特徴付けられるペプチドを含む。一部の実施形態では、同定されるまたは特徴付けられるペプチドは、内因性ペプチドである。一部の実施形態では、ペプチドは、外因性ペプチドである。例えば、同定されるまたは特徴付けられるペプチドを、ペプチドのライブラリーをコードする複数の配列から発現させることができる。

【０５６３】

本明細書の開示前に、ＨＬＡペプチドームに関する大多数のＬＣ－ＭＳ／ＭＳ試験で多数のＨＬＡペプチドを発現する細胞が使用されており、これには、既存のバイオインフォマティクス予測器または「逆畳み込み」を使用してペプチを最大６種のＨＬＡクラスＩ対立遺伝子のうちの１種に割り当てることが必要である（Bassani-Sternberg and Gfeller, 2016）。したがって、既知のモチーフに密接に適合しないペプチドは、所与のＨＬＡ対立遺伝子に対する結合体として確信的に報告することができない。

【０５６４】

個体のＨＬＡペプチドに結合し得る変異ペプチドなどのペプチドの予測方法が本明細書で提供される。一部の実施形態では、本出願は、抗原を含むペプチドの所与のセットから、対象に対する免疫原性組成物を調製するために最も適したペプチドを同定する方法であって、ペプチドの所与のセットから、対象のＨＬＡタンパク質に結合することができる複数のペプチドを選択することを含み、ＨＬＡタンパク質に結合する前記能力が、ペプチドの配列を、前記対象のＨＬＡ－対立遺伝子の各々に対して特異的なＨＬＡ結合性ペプチドに対応するペプチド配列データベースを用いて訓練された機械を用いて解析することによ

って決定される、方法を提供する。抗原を含むペプチドの所与のセットから、対象に対する免疫原性組成物を調製するために最も適したペプチドを同定する方法であって、ペプチドの所与のセットから、対象のＨＬＡタンパク質に結合することができるものと決定された複数のペプチドを選択することを含み、ＨＬＡタンパク質に結合する能力が、ペプチドの配列を、本明細書において上に記載されている方法を行うことによって得られたペプチド配列データベースを用いて訓練された機械を用いて解析することによって決定される、方法が本明細書で提供される。したがって、一部の実施形態では、本開示は、対象特異的免疫原性組成物を調製するための複数の対象特異的ペプチドを同定する方法であって、対象が腫瘍を有し、対象特異的ペプチドが、対象および対象の腫瘍に特異的であり、対象の腫瘍の試料および対象の非腫瘍試料の配列決定を行うステップと、核酸配列決定に基づいて、対象のがん細胞のゲノムには存在するが対象由来の正常な組織には存在しない非サイレント変異、および対象のＨＬＡ遺伝子型を決定するステップと、同定された非サイレント変異から複数の対象特異的ペプチドを選択するステップであって、複数の対象特異的ペプチドの各々が、対象の腫瘍に特異的な異なる腫瘍エピトープを有し、各々が、本明細書に記載のＨＬＡ結合を予測するための方法において非サイレント変異に由来するペプチドの配列を解析することによって決定して、対象のＨＬＡタンパク質に結合することができるものと同定される、ステップとを含む方法を提供する。

【０５６５】

一部の実施形態では、個体に特異的なＨＬＡ－ペプチド複合体を特徴付ける方法が本明細書に開示される。

【０５６６】

一部の実施形態では、個体に特異的なＨＬＡ－ペプチド複合体を特徴付ける方法を使用して、状態または疾患を有する対象などの、それを必要とする個体における免疫治療薬を開発する。

【０５６７】

哺乳動物において抗腫瘍免疫もたらす方法であって、哺乳動物に、記載の方法に従って同定されたペプチドをコードする配列を含むポリ核酸を投与することを含む方法が本明細書で提供される。哺乳動物において抗腫瘍免疫もたらす方法であって、哺乳動物に、本明細書に記載の方法に従って同定されたペプチドの配列を有するペプチドを有効量で投与することを含む方法が本明細書で提供される。哺乳動物において抗腫瘍免疫もたらす方法であって、哺乳動物に、本明細書に記載の方法に従って同定されたペプチドの配列を含むペプチドを含む細胞を投与することを含む方法が本明細書で提供される。哺乳動物において抗腫瘍免疫もたらす方法であって、哺乳動物に、本明細書に記載の方法に従って同定されたペプチドの配列を含むペプチドをコードする配列を含むポリ核酸を含む細胞を投与することを含む方法が本明細書で提供される。一部の実施形態では、細胞は、ペプチドをＨＬＡ－ペプチド複合体として提示する。

【０５６８】

対象における疾患または障害を処置する方法であって、対象に、本明細書に記載の方法に従って同定されたペプチドをコードする配列を含むポリ核酸を投与することを含む方法が本明細書で提供される。対象における疾患または障害を処置する方法であって、対象に、本明細書に記載の方法に従って同定されたペプチドの配列を含むペプチドを有効量で投与することを含む方法が本明細書で提供される。対象における疾患または障害を処置する方法であって、対象に、本明細書に記載の方法に従って同定されたペプチドの配列を含むペプチドを含む細胞を投与することを含む方法が本明細書で提供される。対象における疾患または障害を処置する方法であって、対象に、本明細書に記載の方法に従って同定されたペプチドの配列を含むペプチドをコードする配列を含むポリ核酸を含む細胞を投与することを含む方法が本明細書で提供される。一部の実施形態では、疾患または障害は、がんである。一部の実施形態では、方法は、免疫チェックポイント阻害剤を対象に投与することをさらに含む。

【０５６９】

一部の実施形態では、それを必要とする個体に対する免疫治療薬を、ＨＬＡ－ペプチド複合体を特徴付けることによって開発する方法であって、ａ）それを必要とする個体に由来する細胞の集団を準備するステップであって、細胞の集団のうちの１つまたは複数の細胞が、親和性アクセプターでタグ付けされたＨＬＡクラスＩまたはＨＬＡクラスＩＩ対立遺伝子をコードする配列を含むポリ核酸を含み、親和性アクセプターでタグ付けされたＨＬＡをコードする配列が、ｉ）組換えＨＬＡクラスＩまたはＨＬＡクラスＩＩ対立遺伝子をコードする配列と、それと作動可能に連結したｉｉ）親和性アクセプターペプチドをコードする配列とを含む、ステップと、ｂ）細胞の集団の１つまたは複数の細胞のうちの少なくとも１つの細胞において、親和性アクセプターでタグ付けされたＨＬＡを発現させ、それにより、少なくとも１つの細胞において親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体を形成するステップと、ｃ）親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体を富化するステップと、それを必要とする個体に特異的なＨＬＡ－ペプチド複合体を特徴付けるステップと、ｄ）それを必要とする個体に特異的なＨＬＡ－ペプチド複合体に基づく免疫治療薬を開発するステップとを含み、個体が、疾患または状態を有する、方法が本明細書に開示される。

【０５７０】

一部の実施形態では、免疫治療薬は、核酸またはペプチド治療薬である。

【０５７１】

一部の実施形態では、方法は、１つまたは複数のペプチドを細胞の集団に導入することを含む。一部の実施形態では、方法は、細胞の集団に１つもしくは複数のペプチドを接触させることまたは細胞の集団において１つもしくは複数のペプチドを発現させることを含む。一部の実施形態では、方法は、細胞の集団に１つまたは複数のペプチドをコードする１つまたは複数の核酸を接触させることを含む。

【０５７２】

一部の実施形態では、方法は、患者特異的ＨＬＡに関連して同定されたペプチドに基づく免疫治療薬を開発することを含む。一部の実施形態では、細胞の集団は、それを必要とする個体に由来するものである。

【０５７３】

一部の実施形態では、方法は、細胞の集団においてペプチドのライブラリーを発現させることを含む。一部の実施形態では、方法は、親和性アクセプターでタグ付けされたＨＬＡ－ペプチド複合体のライブラリーを発現させることを含む。一部の実施形態では、ライブラリーは、疾患または状態に関連するペプチドのライブラリーを含む。一部の実施形態では、疾患または状態は、がんまたは感染因子による感染症または自己免疫疾患である。一部の実施形態では、方法は、細胞の集団のうちの１つまたは複数の細胞に感染因子またはその一部を導入することを含む。一部の実施形態では、方法は、それを必要とする個体に特異的なＨＬＡ－ペプチド複合体由来の１つまたは複数のペプチドを特徴付けることを含み、必要に応じて、ペプチドは、感染因子または自己免疫疾患の１つまたは複数の標的タンパク質に由来するものである。一部の実施形態では、方法は、感染因子または自己免疫疾患の１つまたは複数の標的タンパク質由来のペプチドの１つまたは複数の領域を特徴付けることを含む。一部の実施形態では、方法は、感染因子または自己免疫疾患に由来するＨＬＡ－ペプチド複合体由来のペプチドを同定することを含む。

【０５７４】

一部の実施形態では、感染因子は、病原体である。一部の実施形態では、病原体は、ウイルス、細菌、または寄生生物である。

【０５７５】

一部の実施形態では、ウイルスは、ＢＫウイルス（ＢＫＶ）、デングウイルス（ＤＥＮＶ－１、ＤＥＮＶ－２、ＤＥＮＶ－３、ＤＥＮＶ－４、ＤＥＮＶ－５）、サイトメガロウイルス（ＣＭＶ）、Ｂ型肝炎ウイルス（ＨＢＶ）、Ｃ型肝炎ウイルス（ＨＣＶ）、エプスタイン・バーウイルス（ＥＢＶ）、アデノウイルス、ヒト免疫不全ウイルス（ＨＩＶ）、ヒトＴ細胞リンパ増殖性ウイルス（ＨＴＬＶ－１）、インフルエンザウイルス、ＲＳＶ、

ＨＰＶ、狂犬病、流行性耳下腺炎風疹ウイルス、ポリオウイルス、黄熱病、Ａ型肝炎、Ｂ型肝炎、ロタウイルス、水痘ウイルス、ヒトパピローマウイルス（ＨＰＶ）、天然痘、帯状疱疹、およびこれらの組合せからなる群から選択される。

【０５７６】

一部の実施形態では、細菌は、Ｋｌｅｂｓｉｅｌｌａ ｓｐｐ．、Ｔｒｏｐｈｅｒｙｍａ ｗｈｉｐｐｌｅｉ、Ｍｙｃｏｂａｃｔｅｒｉｕｍ ｌｅｐｒａｅ、Ｍｙｃｏｂａｃｔｅｒｉｕｍ ｌｅｐｒｏｍａｔｏｓｉｓ、およびＭｙｃｏｂａｃｔｅｒｉｕｍ ｔｕｂｅｒｃｕｌｏｓｉｓからなる群から選択される。一部の実施形態では、細菌は、腸チフス、肺炎球菌、髄膜炎菌、ヘモフィルスＢ、炭疽、破傷風トキソイド、血清群Ｂ髄膜炎菌、ｂｃｇ、コレラ、およびこれらの組合せからなる群から選択される。

【０５７７】

一部の実施形態では、寄生生物は、蠕虫または原生動物である。一部の実施形態では、寄生生物は、Ｌｅｉｓｈｍａｎｉａ ｓｐｐ．（例えば、Ｌ．ｍａｊｏｒ、Ｌ．ｉｎｆａｎｔｕｍ、Ｌ．ｂｒａｚｉｌｉｅｎｓｉｓ、Ｌ．ｄｏｎｏｖａｎｉ、Ｌ．ｃｈａｇａｓｉ、Ｌ．ｍｅｘｉｃａｎａ）、Ｐｌａｓｍｏｄｉｕｍ ｓｐｐ．（例えば、Ｐ．ｆａｌｃｉｐａｒｕｍ、Ｐ．ｖｉｖａｘ、Ｐ．ｏｖａｌｅ、Ｐ．ｍａｌａｒｉａｅ）、Ｔｒｙｐａｎｏｓｏｍａ ｃｒｕｚｉ、Ａｓｃａｒｉｓ ｌｕｍｂｒｉｃｏｉｄｅｓ、Ｔｒｉｃｈｕｒｉｓ ｔｒｉｃｈｉｕｒａ、Ｎｅｃａｔｏｒ ａｍｅｒｉｃａｎｕｓ、およびＳｃｈｉｓｔｏｓｏｍａ ｓｐｐ．（Ｓ．ｍａｎｓｏｎｉ、Ｓ．ｈａｅｍａｔｏｂｉｕｍ、Ｓ．ｊａｐｏｎｉｃｕｍ）からなる群から選択される。

【０５７８】

一部の実施形態では、免疫治療薬は、工学的に作製された受容体である。一部の実施形態では、工学的に作製された受容体は、キメラ抗原受容体（ＣＡＲ）、Ｔ細胞受容体（ＴＣＲ）、またはＢ細胞受容体（ＢＣＲ）、養子Ｔ細胞治療（ＡＣＴ）、またはそれらの誘導体である。他の態様では、工学的に作製された受容体は、キメラ抗原受容体（ＣＡＲ）である。一部の態様では、ＣＡＲは、第１世代ＣＡＲである。他の態様では、ＣＡＲは、第２世代ＣＡＲである。さらに他の態様では、ＣＡＲは、第３の世代ＣＡＲである。

【０５７９】

一部の態様では、ＣＡＲは、細胞外部分、膜貫通部分、および細胞内部分を含む。一部の態様では、細胞内部分は、少なくとも１つのＴ細胞共刺激ドメインを含む。一部の態様では、Ｔ細胞共刺激ドメインは、ＣＤ２７、ＣＤ２８、ＴＮＦＲＳ９（４－１ＢＢ）、ＴＮＦＲＳＦ４（ＯＸ４０）、ＴＮＦＲＳＦ８（ＣＤ３０）、ＣＤ４０ＬＧ（ＣＤ４０Ｌ）、ＩＣＯＳ、ＩＴＧＢ２（ＬＦＡ－１）、ＣＤ２、ＣＤ７、ＫＬＲＣ２（ＮＫＧ２Ｃ）、ＴＮＦＲＳ１８（ＧＩＴＲ）、ＴＮＦＲＳＦ１４（ＨＶＥＭ）、またはこれらの任意の組合せからなる群から選択される。

【０５８０】

一部の態様では、工学的に作製された受容体は、標的に結合する。一部の態様では、結合は、疾患または状態に罹患している個体に特異的なＨＬＡ－ペプチド複合体を特徴付ける方法で同定されたペプチドに特異的なものである。

【０５８１】

一部の態様では、免疫治療薬は、本明細書に詳細に記載されている細胞である。一部の態様では、免疫治療薬は、疾患または状態に罹患している個体に特異的なＨＬＡ－ペプチド複合体を特徴付ける方法で同定されたペプチドに特異的に結合する受容体を含む細胞である。一部の態様では、免疫治療薬は、本発明のペプチド／核酸と組み合わせて使用される細胞である。一部の実施形態では、細胞は、患者の細胞である。一部の実施形態では、細胞は、Ｔ細胞である。一部の実施形態では、細胞は、腫瘍浸潤性リンパ球である。

【０５８２】

一部の態様では、状態または疾患を有する対象を、対象のＴ細胞受容体レパートリーに基づいて処置する。一部の実施形態では、抗原ワクチンを対象のＴ細胞受容体レパートリーに基づいて選択する。一部の実施形態では、対象を、本明細書に記載の方法を使用して

同定された抗原またはペプチドに特異的なＴＣＲを発現するＴ細胞を用いて処置する。一部の実施形態では、対象を、ＴＣＲ、例えば対象特異的ＴＣＲに特異的な、本明細書に記載の方法を使用して同定された抗原またはペプチドを用いて処置する。一部の実施形態では、対象を、ＴＣＲ、例えば対象特異的ＴＣＲを発現するＴ細胞に特異的な、本明細書に記載の方法を使用して同定された抗原またはペプチドを用いて処置する。一部の実施形態では、対象を、対象特異的ＴＣＲに特異的な、本明細書に記載の方法を使用して同定された抗原またはペプチドを用いて処置する。

【０５８３】

一部の実施形態では、免疫原性抗原組成物またはワクチンを、対象において同定されたＴＣＲに基づいて選択する。一実施形態では、Ｔ細胞レパートリーを同定することおよびそれを機能アッセイにおいて試験することを使用して、状態または疾患を有する対象に投与する免疫原性組成物またはワクチンを決定する。一部の実施形態では、免疫原性組成物は、抗原ワクチンである。一部の実施形態では、抗原ワクチンは、対象特異的抗原ペプチドを含む。一部の実施形態では、抗原ワクチンに含める抗原ペプチドを、抗原に結合する対象特異的ＴＣＲの数量化に基づいて選択する。一部の実施形態では、抗原ペプチドを、ペプチドのＴＣＲに対する結合親和性に基づいて選択する。一部の実施形態では、選択は、数量および結合親和性の組合せに基づく。例えば、機能アッセイでは抗原に強力に結合するが、ＴＣＲレパートリーに高度には表されないＴＣＲは、ＴＣＲを発現するＴ細胞は有利に増幅されるので、抗原ワクチンの良好な候補でありうる。

【０５８４】

一部の実施形態では、抗原を、ＴＣＲへの結合に基づいて、対象への投与のために選択する。一部の実施形態では、疾患または状態を有する対象由来のＴ細胞などのＴ細胞を増大させることができる。本明細書に記載の方法を使用して同定された免疫原性抗原ペプチドに特異的なＴＣＲを発現する増大したＴ細胞を対象に投与し戻すことができる。一部の実施形態では、適切な細胞、例えば、ＰＢＭＣに、本明細書に記載の方法を使用して同定された免疫原性抗原ペプチドに特異的なＴＣＲを発現させるためのポリヌクレオチドを形質導入またはトランスフェクトし、それを対象に投与する。本明細書に記載の方法を使用して同定された免疫原性抗原ペプチドに特異的なＴＣＲを発現するＴ細胞を増大させ、対象に投与し戻すことができる。一部の実施形態では、自己患部組織と一緒にインキュベートした場合に細胞溶解活性をもたらす、本明細書に記載の方法を使用して同定された免疫原性抗原ペプチドに特異的なＴＣＲを発現するＴ細胞を増大させ、対象に投与することができる。一部の実施形態では、本明細書に記載の方法を使用して同定された免疫原性抗原ペプチドへの結合をもたらす、機能アッセイに使用されるＴ細胞を増大させ、対象に投与することができる。一部の実施形態では、本明細書に記載の方法を使用して同定された対象特異的免疫原性抗原ペプチドに結合することが決定されたＴＣＲをＴ細胞において発現させ、対象に投与することができる。

【０５８５】

本明細書に記載の方法は、腫瘍または病原体に関連する抗原などの選択された抗原に特異的なＴ細胞などの免疫系細胞の養子移入を伴い得る。例えば、本明細書に記載の方法を使用して同定された免疫原性抗原ペプチドに対する特異性を有する新しいＴＣＲ　αおよびβ鎖を導入することによってＴ細胞受容体（ＴＣＲ）の特異性を変更することによりＴ細胞を遺伝子改変するために種々の戦略を使用することができる（例えば、米国特許第８，６９７，８５４号；ＰＣＴ特許公報第ＷＯ２００３０２０７６３号、同第ＷＯ２００４０３３６８５号、同第ＷＯ２００４０４４００４号、同第ＷＯ２００５１１４２１５号、同第ＷＯ２００６０００８３０号、同第ＷＯ２００８０３８００２号、同第ＷＯ２００８０３９８１８号、同第ＷＯ２００４０７４３２２号、同第ＷＯ２００５１１３５９５号、同第ＷＯ２００６１２５９６２号、同第ＷＯ２０１３１６６３２１号、同第ＷＯ２０１３０３９８８９号、同第ＷＯ２０１４０１８８６３号、同第ＷＯ２０１４０８３１７３号；米国特許第８，０８８，３７９号を参照されたい）。

【０５８６】

キメラ抗原受容体（ＣＡＲ）を使用して、多種多様な受容体キメラ構築物を有する、本明細書に記載の方法を使用して同定された免疫原性抗原ペプチドのような選択された標的に特異的なＴ細胞などの免疫応答性細胞を生成することができる（例えば、米国特許第５，８４３，７２８号；同第５，８５１，８２８号；同第５，９１２、１７０号；同第６，００４，８１１号；同第６，２８４，２４０号；同第６，３９２，０１３号；同第６，４１０，０１４号；同第６，７５３，１６２号；同第８，２１１，４２２号；および、ＰＣＴ公開第Ｗ０９２１５３２２号を参照されたい）。代替ＣＡＲ構築物を次に続く世代に属するものとして特徴付けることができる。第１世代ＣＡＲは、一般には、抗原に特異的な抗体の単鎖可変断片からなり、例えば、柔軟なリンカーによって、例えばＣＤ８ａヒンジドメインおよびＣＤ８ａ膜貫通ドメインによって、ＣＤ３ζまたはＦｃＲｙまたはｓｃＦｖ－ＦｃＲｙのいずれかの膜貫通および細胞内シグナル伝達ドメインに連結した、特異的な抗体のＶＨと連結したＶＬを含む（例えば、米国特許第７，７４１，４６５号；米国特許第５，９１２，１７２号；米国特許第５，９０６，９３６号を参照されたい）。第２世代ＣＡＲでは、ＣＤ２８、ＯＸ４０（ＣＤ１３４）、または４－１ＢＢ（ＣＤ１３７）などの１つまたは複数の共刺激分子の細胞内ドメインをエンドドメイン内に組み入れる。例えば、ｓｃＦｖ－ＣＤ２８／ＯＸ４０／４－ｌＢＢ－ＣＤ３（例えば、米国特許第８，９１１，９９３号；同第８，９１６，３８１号；同第８，９７５，０７１号；同第９，１０１，５８４号；同第９，１０２，７６０号；同第９，１０２，７６１号を参照されたい）。第３世代ＣＡＲは、共刺激エンドドメイン、例えばＣＤ３Ｃ鎖、ＣＤ９７、ＧＤＩ　ｌａ－ＣＤ１８、ＣＤ２、ＩＣＯＳ、ＣＤ２７、ＣＤ１５４、ＣＤＳ、ＯＸ４０、４－１ＢＢ、またはＣＤ２８シグナル伝達ドメイン、例えば、ｓｃＦｖ－ＣＤ２８－４－ｌＢＢ－ＣＤ３ＣまたはｓｃＦｖ－ＣＤ２８－ＯＸ４０－ＣＤ３Ｑの組合せを含む（例えば、米国特許第８，９０６，６８２号；米国特許第８，３９９，６４５号；米国特許第５，６８６，２８１号；ＰＣＴ公開第ＷＯ２０１４１３４１６５号；ＰＣＴ公開第ＷＯ２０１２０７９０００号を参照されたい）。一部の実施形態では、共刺激を、例えば、プロフェッショナル抗原提示細胞上の抗原との相互作用後に、共刺激で活性化され、増大するように選択された抗原特異的Ｔ細胞においてＣＡＲを発現させることによって調和させることができる。例えば、Ｔ細胞攻撃の標的化を改善するため、および／または副作用を最小限にするために、追加的な工学的に作製された受容体を免疫応答性細胞に提供することができる。

【０５８７】

プロトプラスト融合、リポフェクション、トランスフェクションまたは電気穿孔などの代替技法を使用して、標的免疫応答性細胞を形質転換することができる。例えば、ＣＤ３ζおよびＣＤ２８またはＣＤ１３７のいずれかを通じた第２世代抗原特異的ＣＡＲシグナル伝達を使用してＣＡＲを導入するために、レトロウイルスベクター、レンチウイルスベクター、アデノウイルスベクター、アデノ随伴ウイルスベクター、プラスミドまたはＳｌｅｅｐｉｎｇ　Ｂｅａｕｔｙトランスポゾンなどのトランスポゾンなどの多種多様なベクターを使用することができる（米国特許第６，４８９，４５８号；同第７，１４８，２０３号；同第７，１６０，６８２号；同第７，９８５，７３９号；同第８，２２７，４３２号を参照されたい）。ウイルスベクターとしては、例えば、ＨＩＶ、ＳＶ４０、ＥＢＶ、ＨＳＶまたはＢＰＶに基づくベクターを挙げることができる。

【０５８８】

形質転換の標的とされる細胞としては、例えば、Ｔ細胞、ナチュラルキラー（ＮＫ）細胞、細胞傷害性Ｔリンパ球（ＣＴＬ）、調節性Ｔ細胞、ヒト胚性幹細胞、腫瘍浸潤リンパ球（ＴＩＬ）またはリンパ系細胞が分化し得る多能性幹細胞を挙げることができる。所望のＣＡＲを発現するＴ細胞は、例えば、がん抗原と共刺激分子を同時発現するγ線照射した活性化および増殖細胞（ＡＰＣ）と共培養することによって選択することができる。工学的に操作されたＣＡＲ　Ｔ細胞を、例えば、ＩＬ－２およびＩＬ－２１などの可溶性因子の存在下でＡＰＣと共培養することによって増大させることができる。増大は、例えば、メモリーＣＡＲ　Ｔ細胞がもたらされるように行うことができる（例えば、非酵素的デジタルアレイおよび／またはマルチパネルフローサイトメトリーによってアッセイするこ

とができる）。このように、抗原を有する腫瘍に対する特定の細胞傷害活性を有するＣＡＲ　Ｔ細胞をもたらすことができる（必要に応じてインターフェロン－γなどの所望のケモカインの産生と併せて）。この種類のＣＡＲ　Ｔ細胞は、例えば、動物モデルにおいて、例えば、腫瘍異種移植片に脅威を与えるために使用することができる。

【０５８９】

前述のものなどの手法を適合させて、例えば、選択された抗原に結合する受容体を認識する抗原を含む免疫応答性細胞を有効量で投与することによって新形成または病原性感染症などの疾患を有する対象を処置し、かつ／またはその生存を増大させる方法であって、結合により、免疫応答性細胞が活性化され、それにより、疾患（例えば、新形成、病原体感染症、自己免疫障害、または同種移植反応など）が処置または防止される、方法をもたらすことができる。ＣＡＲ　Ｔ細胞治療における投薬は、例えば、１ｋｇ当たり１０６個から１０９個までの細胞を、例えばシクロホスファミドを用いたリンパ球枯渇の過程を伴ってまたは伴わずに投与することを伴い得る。

【０５９０】

可能性のある有害反応から守るために、工学的に操作された免疫応答性細胞に、トランスジェニック安全スイッチを、細胞を特定のシグナルへの曝露を受けやすくする導入遺伝子の形態で備えることができる。例えば、単純ヘルペスウイルスチミジンキナーゼ（ＴＫ）遺伝子を、例えば、幹細胞移植後のドナーリンパ球輸注として使用される同種異系のＴリンパ球に導入することにより、このように使用することができる。そのような細胞では、ガンシクロビルまたはアシクロビルなどのヌクレオシドプロドラッグの投与により、細胞死が引き起こされる。代替安全スイッチ構築物としては、例えば２つの非機能性ｉｃａｓｐ９分子と一緒にして活性な酵素を形成する小分子二量体形成剤の投与によって誘発される誘導性カスパーゼ９が挙げられる。細胞増殖制御をインプリメントするための多種多様な代替手法が記載されている（例えば、米国特許出願公開第２０１３００７１４１４号；ＰＣＴ特許公報第ＷＯ２０１１１４６８６２号；ＰＣＴ特許公報第ＷＯ２０１４０１１９８７号；ＰＣＴ特許公報第ＷＯ２０１３０４０３７１号を参照されたい）。養子治療のさらなる精密化では、代替インプリメンテーション、例えば、編集されたＣＡＲ　Ｔ細胞の提供のために、ゲノム編集を使用して免疫応答性細胞を調整することができる。

【０５９１】

細胞治療法は、Ｔ細胞のｅｘ　ｖｉｖｏにおける活性化および増大も伴い得る。一部の実施形態では、Ｔ細胞を、それを必要とする対象に投与する前に活性化することができる。これらの型の処置の例としては、腫瘍浸潤性リンパ球（ＴＩＬ）細胞（米国特許第５，１２６，１３２号を参照されたい）、細胞傷害性Ｔ細胞（米国特許第６，２５５，０７３号；および米国特許第５，８４６，８２７号を参照されたい）、増大させた腫瘍流入領域リンパ節細胞（米国特許第６，２５１，３８５号を参照されたい）、および種々の他のリンパ球調製物（米国特許第６，１９４，２０７号；米国特許第５，４４３，９８３号；米国特許第６，０４０，１７７号；および米国特許第５，７６６，９２０号を参照されたい）の使用が挙げられる。

【０５９２】

ｅｘ　ｖｉｖｏ活性化Ｔ細胞集団は、がん、感染性疾患、または他の病的状態、例えば、自己免疫疾患の状態に対する免疫応答を最大限に調整する状態にあり得る。活性化のためには、少なくとも２つのシグナルがＴ細胞に送達され得る。第１のシグナルは、通常、Ｔ細胞表面上のＴ細胞受容体（ＴＣＲ）を通じて送達される。ＴＣＲによる第１のシグナルは、通常、ＴＣＲが抗原提示細胞（ＡＰＣ）の表面上にＭＨＣ複合体と併せて発現されたペプチド抗原と相互作用すると誘発される。第２のシグナルは、通常、Ｔ細胞の表面上の共刺激受容体を通じて送達される。共刺激受容体は、一般に、ＡＰＣの表面上に発現される対応するリガンドまたはサイトカインによって誘発される。

【０５９３】

本明細書に記載の方法を使用して同定された免疫原性抗原ペプチドに特異的なＴ細胞を得、疾患を処置または防止する方法に使用することができることが意図されている。この

点について、本開示は、対象における疾患または状態を処置または防止する方法であって、対象に、本明細書に記載の方法を使用して同定された免疫原性抗原ペプチドに特異的な細胞を含む細胞集団を、対象における疾患を処置または防止するために有効な量で投与することを含む方法を提供する。一部の実施形態では、対象における疾患を処置または防止する方法は、疾患反応性Ｔ細胞を富化させた細胞集団を対象に、哺乳動物におけるがんを処置または防止するために有効な量で投与することを含む。細胞は、対象にとって同種異系または自己の細胞であり得る。

【０５９４】

本開示は、対象に抗原性ペプチドまたはワクチンを投与することにより、対象における疾患特異的免疫応答を誘導する、疾患に対するワクチン接種を行う、対象における疾患の症状を処置および／または緩和する方法をさらに提供する。

【０５９５】

本開示のペプチドまたは組成物は、ＣＴＬ応答を誘導するために十分な量で投与することができる。抗原性ペプチドまたはワクチン組成物を、単独で、または他の治療剤と組み合わせて投与することができる。例示的な治療剤としては、これだけに限定されないが、化学療法薬または生物学的治療薬作用物質、放射線、または免疫治療が挙げられる。特定の疾患に対する任意の適切な治療的処置を施行することができる。化学療法薬および生物学的治療剤の例としては、これだけに限定されないが、アルデスロイキン、アルトレタミン、アミホスチン、アスパラギナーゼ、ブレオマイシン、カペシタビン、カルボプラチン、カルムスチン、クラドリビン、シサプリド、シスプラチン、シクロホスファミド、シタラビン、ダカルバジン（ＤＴＩＣ）、ダクチノマイシン、ドセタキセル、ドキソルビシン、ドロナビノール、エポエチンアルファ、エトポシド、フィルグラスチム、フルダラビン、フルオロウラシル、ゲムシタビン、グラニセトロン、ヒドロキシウレア、イダルビシン、イホスファミド、インターフェロンアルファ、イリノテカン、ランソプラゾール、レバミソール、ロイコボリン、メゲストロール、メスナ、メトトレキサート、メトクロプラミド、マイトマイシン、ミトタン、ミトキサントロン、オメプラゾール、オンダンセトロン、パクリタキセル（Ｔａｘｏｌ（登録商標））、ピロカルピン、プロクロルペラジン、リツキシマブ、タモキシフェン、タキソール、トポテカン塩酸塩、トラスツズマブ、ビンブラスチン、ビンクリスチンおよび酒石酸ビノレルビンが挙げられる。さらに、対象に、抗免疫抑制性剤または免疫賦活性剤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。

【０５９６】

ワクチン組成物に含められる各ペプチドの量および投薬レジメンは、当業者が決定することができる。例えば、ペプチドまたはそのバリアントを静脈内（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．）注射用に調製することができる。ペプチド注射の典型的な方法としては、ｓ．ｃ、ｉ．ｄ．、ｉ．ｐ．、ｉ．ｍ．、およびｉ．ｖ．が挙げられる。ＤＮＡ注射の典型的な方法としては、ｉ．ｄ．、ｉ．ｍ．、ｓ．ｃ、ｉ．ｐ．およびｉ．ｖ．が挙げられる。ワクチン組成物の他の投与方法は当業者に公知である。

【０５９７】

医薬組成物を、組成物中に存在するペプチドの選択、数および／または量が疾患および／または患者特異的になるようにコンパイルすることができる。例えば、ペプチドの正確な選択を、副作用を回避するために、所与の組織における親タンパク質の発現パターンによってガイドすることができる。選択は、疾患の特定の型、疾患の状態、以前の処置レジメン、患者の免疫の状態、および患者のＨＬＡ－ハプロタイプに依存的なものであり得る。さらに、本開示によるワクチンは、特定の患者の個人的な必要性に応じて個別化された構成成分を含有し得る。例としては、特定の患者における関連する抗原の発現に応じて変動するペプチドの量、個人的なアレルギーまたは他の処置に起因する望ましくない副作用、および処置の第１のラウンドまたはスキームの後の二次的処置のための調整が挙げられる。

コンピュータ制御システム

【０５９８】

本開示は、本開示の方法をインプリメントするようにプログラムされたコンピュータ制御システムを提供する。図１０は、機械学習ＨＬＡ－ペプチド提示モデルの訓練を行うようにプログラムされたまたは他のやり方で構成されたコンピュータシステム（１００１）を示す。コンピュータシステム（１００１）により、例えば、アミノ酸位置情報を入力すること、帰属情報をデータセットに移行すること、およびデータセットを用いて訓練されたアルゴリズムを生成することなどの本開示の種々の態様を調節することができる。コンピュータシステム（１００１）は、ユーザー電子デバイスまたは遠隔コンピュータシステムであり得る。電子デバイスは、モバイル電子デバイスであり得る。

【０５９９】

コンピュータシステム（１００１）は、中央処理装置（ＣＰＵ、本明細書では「プロセッサ」および「コンピュータプロセッサ」とも）（１００５）を含み、これは、逐次的処理または並行処理のいずれかによるシングルコアプロセッサまたはマルチコアプロセッサであり得る。コンピュータシステム（１００１）はまた、メモリユニットまたはデバイス（１０１０）（例えば、ランダムアクセスメモリー、リードオンリーメモリー、フラッシュメモリ）、ストレージユニット（１０１５）（例えば、ハードディスク）、１つまたは複数の他のシステムと通信するための通信インターフェース（１０２０）（例えば、ネットワークアダプター）、ならびに、プリンター、モニター、ＵＳＢドライブおよび／またはＣＤ－ＲＯＭドライブなどの、外付けまたは内蔵のいずれかまたはその両方の周辺機器（１０２５）も含む。メモリ（１０１０）、ストレージユニット（１０１５）、インターフェース（１０２０）および周辺機器（１０２５）は、ＣＰＵ（１００５）と、マザーボードなどの通信バス（実線）を通じて通信する。ストレージユニット（１０１５）は、データを記憶させるためのデータストレージユニット（またはデータリポジトリ）であり得る。コンピュータシステム（１００１）は、コンピュータネットワーク（「ネットワーク」）（１０３０）に通信インターフェース（１０２０）を活用して作動可能にカップリングすることができる。ネットワーク（１０３０）は、インターネット、インターネットおよび／もしくはエクストラネット、またはインターネットと通信したイントラネットおよび／もしくはエクストラネットであり得る。ネットワーク（１０３０）は、一部の場合では、電気通信および／またはデータネットワークである。ネットワーク（１０３０）は、分散コンピューティングを支持するピアツーピアネットワークを可能にすることができる１つまたは複数のコンピュータサーバーを含み得る。ネットワーク（１０３０）は、一部の場合では、コンピュータシステム（１００１）を活用して、コンピュータシステム（１００１）とカップリングしたデバイスをクライアントまたはサーバーとして動作させることを可能にすることができるクライアント－サーバー構造をインプリメントすることができる。

【０６００】

ＣＰＵ（１００５）は、プログラムまたはソフトウェアに具体化することができる一連の機械可読指示を実行することができるものである。指示は、メモリ（１０１０）に記憶させることができる。指示をＣＰＵ（１００５）に方向付けることができ、その後、それにより、ＣＰＵ（１００５）を、本開示の方法をインプリメントするようにプログラムするまたは他のやり方で構成することができる。ＣＰＵ（１００５）によって実施される操作の例としては、フェッチ、復号、実行およびライトバックを挙げることができる。

【０６０１】

ＣＰＵ（１００５）は、集積回路などの回路の一部であり得る。システム（１００１）の１つまたは複数の他の構成成分を回路に含めることができる。一部の場合では、回路は、特定用途向け集積回路（ＡＳＩＣ）でる。

【０６０２】

ストレージユニット（１０１５）は、ドライバー、ライブラリーおよび保存プログラムなどの、ファイルを記憶することができるものである。ストレージユニット（１０１５）

は、ユーザーデータ、例えば、ユーザー選択およびユーザープログラムを記憶することができるものである。コンピュータシステム（１００１）は、例えば、一部の場合では、コンピュータシステム（１００１）とイントラネットまたはインターネットを通じて通信する遠隔サーバー上に位置するものなどの、コンピュータシステム（１００１）の外部の１つまたは複数の追加的なデータストレージユニットを含み得る。

【０６０３】

コンピュータシステム（１００１）は、１つまたは複数の遠隔コンピュータシステムとネットワーク（１０３０）を通じて通信し得る。例えば、コンピュータシステム（１００１）は、遠隔コンピュータシステムまたはユーザーと通信し得る。遠隔コンピュータシステムの例としては、パーソナルコンピュータ（例えば、携帯型ＰＣ）、スレートまたはタブレットＰＣ（例えば、Ａｐｐｌｅ（登録商標）ｉＰａｄ（登録商標）、Ｓａｍｓｕｎｇ（登録商標）Ｇａｌａｘｙ　Ｔａｂ）、電話機、スマートフォン（例えば、Ａｐｐｌｅ（登録商標）ｉＰｈｏｎｅ（登録商標）、Ａｎｄｒｏｉｄ対応デバイス、Ｂｌａｃｋｂｅｒｒｙ（登録商標））、または携帯情報端末が挙げられる。ユーザーは、コンピュータシステム（１００１）にネットワーク（１０３０）を介してアクセスすることができる。

【０６０４】

本明細書に記載の方法は、例えばメモリ（１０１０）またはデータストレージユニット（１０１５）などのコンピュータシステム（１００１）の電子ストレージ位置に記憶された機械（例えば、コンピュータプロセッサ）により実行可能なコードによってインプリメントすることができる。機械により実行可能なまたは機械可読コードは、ソフトウェアの形態で提供することができる。使用中、コードは、プロセッサ（１００５）によって実行することができる。一部の場合では、プロセッサ（１００５）によりすぐにアクセスすることができるように、コードをストレージユニット（１０１５）から取り出し、メモリ（１０１０）に記憶させることができる。一部の状況では、ストレージユニット（１０１５）を除外することができ、機械により実行可能な命令をメモリ（１０１０）に記憶させる。

【０６０５】

コードを、コードを実行するように適合させたプロセッサを有する機械での使用のためにプリコンパイルし、構成することができる、または、コードを実行時間の間にコンパイルすることができる。コードは、コードがプリコンパイル様式でまたは都度コンパイル様式で実行されることが可能になるように選択することができるプログラミング言語で供給することができる。

【０６０６】

コンピュータシステム（１００１）などの、本明細書で提供されるシステムおよび方法の態様は、プログラミングに具体化することができる。当該技術の種々の態様は、一般には、機械可読媒体の一種に保持されるまたはそれに具体化される機械（またはプロセッサ）により実行可能なコードおよび／または関連データの形態の「製品」または「製造品」と考えられ得る。機械により実行可能なコードは、ハードディスクなどのストレージユニットに、またはメモリ（例えば、リードオンリーメモリー、ランダムアクセスメモリー、フラッシュメモリ）に記憶させることができる。「ストレージ」型媒体は、コンピュータの有形メモリ、プロセッサなど、またはその関連モジュール、例えば、ソフトウェアプログラミングのために任意の時点で非一時的ストレージを提供することができる種々の半導体メモリ、テープドライブ、ディスクドライブなど、のいずれかまたは全てを含み得る。ソフトウェアの全てまたは一部は、時々、インターネットまたは種々の他の電気通信ネットワークを通じて通信することができる。そのような通信により、例えば、１つのコンピュータまたはプロセッサから別のコンピュータまたはプロセッサへの、例えば、管理サーバーまたはホストコンピュータからアプリケーションサーバーのコンピュータプラットフォームへのソフトウェアのローディングが可能になる。したがって、ソフトウェア要素を担持することができる別の型の媒体としては、有線および光通信線ネットワークを通じて、ならびに種々のエアリンクによって、ローカルデバイス間の物理的インターフェースを渡って使用されるものなどの光波、電波および電磁波が挙げられる。そのような波を運ぶ

物理的要素、例えば、有線または無線リンク、光リンクなども、ソフトウェアを担持する媒体とみなすことができる。本明細書で使用される場合、非一時的な有形「記憶」媒体に制限される場合を除き、コンピュータまたは機械「可読媒体」などの用語は、プロセッサに実行の指示を提供することに関与する任意の媒体を指す。

【０６０７】

したがって、コンピュータで実行可能なコードなどの機械可読媒体は、これだけに限定されないが、有形記憶媒体、搬送波媒体または物理的伝送媒体を含めた多くの形態を取り得る。非揮発性記憶媒体としては、光学ディスクまたは磁気ディスク、例えば、任意のコンピュータ（複数可）におけるストレージデバイスのいずれかなど、例えば、図に示されているデータベースなどをインプリメントするために使用することができるものなどが挙げられる。揮発性記憶媒体としては、そのようなコンピュータプラットフォームのメインメモリなどのダイナミックメモリが挙げられる。有形伝達媒体としては、同軸ケーブル；コンピュータシステム内のバスを構成する電線を含めた銅線および光ファイバーが挙げられる。搬送波伝送媒体は、電気もしくは電磁気シグナル、または、高周波（ＲＦ）および赤外（ＩＲ）データ通信の間に生じるものなどの音波もしくは光波の形態を取り得る。したがって、コンピュータ可読媒体の一般形態としては、例えば、フロッピー（登録商標）ディスク、フレシキブルディスク、ハードディスク、磁気テープ、任意の他の磁気媒体、ＣＤ－ＲＯＭ、ＤＶＤもしくはＤＶＤ－ＲＯＭ、任意の他の光学媒体、パンチカード、紙テープ、穴のパターンを有する任意の他の物理的記憶媒体、ＲＡＭ、ＲＯＭ、ＰＲＯＭおよびＥＰＲＯＭ、ＦＬＡＳＨ（登録商標）－ＥＰＲＯＭ、任意の他のメモリチップもしくはカートリッジ、搬送波伝達データもしくは指示、そのような搬送波を輸送するケーブルもしくはリンク、またはコンピュータによりプログラミングコードおよび／もしくはデータを読み取ることができる任意の他の媒体が挙げられる。これらの形態のコンピュータ可読媒体の多くは、１つまたは複数の指示の１つまたは複数のシーケンスを実行のためにプロセッサに運ぶことに関与し得る。

【０６０８】

コンピュータシステム（１００１）は、例えば、対象のがん細胞のクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により、同定されたペプチド配列のうちの所与の配列が提示される確率を提供するためのユーザーインタフェース（ＵＩ）（１０４０）を含む電子ディスプレイ（１０３５）を含み得るまたはそれと通信し得る。ＵＩの例としては、限定することなく、グラフィカルユーザーインターフェース（ＧＵＩ）およびウェブに基づくユーザーインタフェースが挙げられる。

【０６０９】

本開示の方法およびシステムは、１つまたは複数のアルゴリズムによってインプリメントすることができる。アルゴリズムは、中央処理装置（１００５）によって実行された際に、ソフトウェアによってインプリメントすることができる。アルゴリズムは、例えば、アミノ酸位置情報を入力し、帰属情報をデータセットに移行し、データセットを用いて訓練されたアルゴリズムを生成することができるものである。

【実施例】

【０６１０】

以下に提示する実施例は、単に例示する目的のものであり、本明細書に提示される特許請求の範囲を限定しない。

【０６１１】

（実施例１）

ＨＬＡ　Ｃｌａｓｓ　ＩＩ結合予測器の性能

本実施例では、ヒットペプチドをランダムにシャッフリングすることによって１：１９（ヒット：デコイ）の比で生成された観察された質量分析ペプチドおよびデコイペプチドを含む検証用データセットを使用して、ｎｅｏｎｍｈｃ２（ＮＥＯＮ）およびＮｅｔＭＨＣＩＩｐａｎの結合予測器を解析した（図４）。ＮＥＯＮ結合予測器に関しては、示されている各ＭＨＣ　ＩＩ対立遺伝子について別々のモデルを築いた。棒の高さは陽性適中率

（ＰＰＶ）を示す。対立遺伝子は、その対立遺伝子について予測した場合のモデルの性能によってソートされている。ＮＥＯＮ結合予測器では、ＮｅｔＭＨＣＩＩｐａｎと比較した場合、全ての対立遺伝子にわたってより高いＰＰＶが示された。

【０６１２】

本実施例では、結合予測器の検証に対するＳＰＩ閾値の影響も試験した（図５）。種々のスコア化ピーク強度（ＳＰＩ）カットオフを有するペプチドのセットで訓練／検証を行った場合のＨＬＡクラスＩＩ結合予測器の性能を示した。異なるＳＰＩカットオフ条件を使用した：７０　ＳＰＩよりも大きいまたはそれと等しい観察されたＭＳヒットペプチドを使用してデータセットに対して訓練および評価を行う、５０　ＳＰＩよりも大きいまたはそれと等しいペプチドで訓練を行い、７０　ＳＰＩよりも大きいまたはそれと等しいペプチドで検証を行う、ならびに、５０　ＳＰＩよりも大きいまたはそれと等しいペプチドで訓練および検証を行う。

【０６１３】

本実施例では、ＨＬＡ－ＤＲ（ＵＳＡ対立遺伝子頻度）について＞９５％の集団カバレッジを有する３５種のＨＬＡ－ＤＲ対立遺伝子についてのデータを収集して、ＬＣ－ＭＳ／ＭＳによる対立遺伝子プロファイリングによって観察された、７０スコア化ピーク強度（ＳＰＩ）カットオフよりも大きいまたはそれと等しいペプチドの数を示した（図６）。

【０６１４】

１つの例示的なセットアップでは、モデルＰＰＶ解析を、Ｎｅｏｎｍｈｃ２プログラムについてこれまでのところ生成された各クラスＩＩ対立遺伝子についての試験区分データに適用した。試験区分データは、ＭＳにより観察されたクラスＩＩ結合体である陽性例（例えば、ヒットサンプルペプチド）と陽性例のスクランブルバージョンである陰性例（例えば、デコイサンプルペプチド）で構成されるものであった。ヒット：デコイ比を１：１９に保った、例えば、各陽性サンプルに対して、１９の陰性サンプルを含め（すなわち、５％陽性サンプル）、検証のために試験区分を実施した。試験区分内のペプチドの最良スコア化５％を選択し、どの程度の分率が陽性であるかを問い合わせることによってＰＰＶスコアを生成した。結果を図７Ａに示す。

【０６１５】

収集されたＨＬＡ－ＤＲ対立遺伝子に関して、訓練セットのサイズを増大させるとＰＰＶの値が増加することが観察された（図７Ｂ）。

【０６１６】

本実施例では、プロセシングに関連する変数により、予測がさらに改善された（図８）。訓練用データ区分に対して、ロジスティック回帰を当てはめて、結合強度（ＮｅｔＭＨＣＩＩｐａｎまたはＮｅｏｎの予測器）およびプロセシング特色（ＲＮＡ－Ｓｅｑ発現および派生遺伝子レベルの偏り項）を使用してＨＬＡクラスＩＩ提示を予測した。別々の評価区分に対して、ＭＳにより観察されたＭＨＣ　ＩＩペプチド（「ヒット」）と重複するエクソンの位置をＭＳで観察されなかったランダムなエクソンの位置と比べてスコア化した（１：４９９の比）。一般に、プロセシングに関連する変数を用いたＮｅｏｎでは、ＮｅｔＭＨＣＩＩｐａｎ、Ｎｅｏｎの予測器、およびプロセシングに関連する変数を用いたＮｅｔＭＨＣＩＩｐａｎよりも高いＰＰＶが示された。

【０６１７】

（実施例２）

ニューラルネットワークアーキテクチャ

本実施例では、ニューラルネットワークを使用して、訓練用アルゴリズムを得る（図９）。入力ペプチドは２０ｍｅｒで表され、それよりも短いペプチドは「欠損」文字が記入される。各ペプチドは３１次元の埋め込みを有し、したがって、ニューラルネットワークへの入力は２０×３１行列であった。ニューラルネットワークによって処理する前に、２０×３１行列に対する特色の正規化を訓練セットにおける特色値の平均および標準偏差に基づいて実施した。第１の畳み込み層は、ＲｅＬＵ活性化関数を用いる９アミノ酸および５０フィルター（チャネルとも称される）のカーネルを有した。この後にバッチ正規化、

次いで空間ドロップアウトをドロップアウト率２０％で行った。この後に、カーネル３および２０フィルターおよびＲｅＬＵ活性化関数を用いた別の畳み込み層、次いで、再度バッチ正規化およびドロップアウト率２０％での空間ドロップアウトを行った。次いで、２０フィルターの各々において最大限に活性化されたニューロンを取ってグローバル最大プーリングを適用し、次いで、これらの２０個の値を、シグモイド活性化関数を使用し、単一のニューロンを有する全結合（ｄｅｎｓｅ）層に渡した。この出力を結合／非結合予測として扱った。Ｌ２正則化をそれぞれ０．０５、０．１、および０．０１の重みを有する第１の畳み込み層、第２の畳み込み層、および高密度層の重みに適用した。

【０６１８】

（実施例３）

単一対立遺伝子ＭＨＣクラスＩＩリガンドプロファイリングのためのスケーラブルプロトコール

ＭＨＣクラスＩＩ結合性モチーフに関する現在の知見は、２つのｉｎ ｖｉｔｒｏ結合アッセイに基づくものであり得、２つのアッセイの一方は、細胞ＭＨＣを使用してＥＣ５０を算出するものであり、他方は、精製されたＭＨＣを使用してＩＣ５０を算出するものである。主要なＨＬＡクラスＩＩ予測アルゴリズムであるＮｅｔＭＨＣＩＩｐａｎは、これらのデータに排他的に基づいて訓練される。

【０６１９】

限られた数のヒトＨＬＡクラスＩＩ対立遺伝子が、ほぼ全てが１５ｍｅｒである確認された結合性ペプチドの２００より多くの例（親和性＜１００ｎＭ）（図１２Ｅ）によって現在支持されている。これらの実験は最も一般的なコーカサス人ＨＬＡ－ＤＲ対立遺伝子のみにわたり、非コーカサス人集団に特異的な対立遺伝子（例えば、ＨＬＡ－ＤＲＢ１＊１５：０２）に関してはカバレッジが限定されており、一般的なＨＬＡ－ＤＰおよびＨＬＡ－ＤＱ対立遺伝子に関してはカバレッジがほぼない。現行のＨＬＡクラスＩＩ予測性能は、一般的なコーカサス人対立遺伝子に対するものであっても、ＭＨＣクラスＩの正確度に後れをとっており、ＲＯＣ曲線はランダムと比べるとわずかに良好なものである。

【０６２０】

これらの限定を考慮して、本明細書では、ＭＳに基づく同定のために単一の対立遺伝子によってコードされるＭＨＣタンパク質に結合するＨＬＡクラスＩＩ結合性ペプチドの効率的な単離を可能にする、Ｍｏｎｏ－Ａｌｌｅｌｉｃ Ｃａｐｔｕｒｅ ｂｙ Ｔａｇｇｅｄ Ａｌｌｅｌｅ ｃａｐｔｕｒｅ（ＭＡＰＴＡＣ（商標））と称する新規のバイオテクノロジーを開発した（図１１Ａおよび１１Ｂ）；この手法は、ＨＬＡクラスＩにも同様に機能する。ＨＬＡクラスＩＩに適用する場合、選択された対立遺伝子のアルファ鎖およびベータ鎖が遺伝子構築物にコードされ、ビオチン－アクセプターペプチド（ＢＡＰ）配列がベータ鎖のＣ末端に配置されている。次いで、これらの細胞を溶解させ、ＢｉｒＡ酵素と一緒にインキュベートして、捕捉用対立遺伝子のベータ鎖のＣ末端をビオチン化する。ＮｅｕｔｒＡｖｉｄｉｎプルダウンによりＭＨＣと結合したペプチドの集団を精製し、それを分子ふるいによってさらに単離し、ベストインクラスＬＣ－ＭＳ／ＭＳプロトコールを用いて配列決定を行う。

【０６２１】

一部の実施形態では、高磁場非対称波形イオン移動度分光分析（ＦＡＩＭＳ）を使用してＬＣ－ＭＳ／ＭＳ分析を評価する。一部の実施形態では、ペプチドを、ｎＬＣ－ＭＳ／ＭＳによる分析の前に、酸性逆相（ａＲＰ）および塩基性逆相（ｂＲＰ）オフライン分画の両方に供する。

【０６２２】

２日間のトランスフェクションが構築物の頑強な発現を実現するために十分であり（図１２Ｂ）、４種の異なる対立遺伝子について３種の別個の細胞株（ｅｘｐｉ２９３、Ａ３７５、およびＢ７２１）における妥当な細胞表面局在が伴った（図１２Ｃ）。

【０６２３】

ＨＬＡ－ＤＲＡは機能的にインバリアントであるので、この手法では、捕捉用ベータ鎖

が内因性アルファ鎖と対形成したとしても、単一対立遺伝子分解能が実現される。これは、この手法を、所与の細胞株における既存のＨＬＡ遺伝子型および発現レベルにかかわらず、ＨＬＡ－ＤＲ対立遺伝子をプロファイルするために使用することができることを意味する。

【０６２４】

ＨＬＡ－ＤＰおよびＨＬＡ－ＤＱに関しては、アルファ鎖およびベータ鎖はどちらも変動し、どちらもペプチド結合に寄与し、したがって、単一対立遺伝子分解能は、ネイティブなアルファ鎖が発現されない場合またはネイティブな対立遺伝子がホモ接合性であり、捕捉用対立遺伝子とマッチする場合のみに予測される。あるいは、ネイティブなアルファ鎖に対応するペプチドのバックグラウンドを確立するために、ベータ鎖のみの捕捉を使用することができる。

【０６２５】

プロファイリングされた対立遺伝子には、５種のＨＬＡ－ＤＲ対立遺伝子（ＤＲＢ１＊０３：０１、ＤＲＢ１＊０９：０１、ＤＲＢ１＊１１：０１、ＤＲＢ３＊０１：０１、およびＤＲＢ３＊０２：０２）ならびに１種のＨＬＡ－ＤＰ対立遺伝子（ＤＰＢ１＊０１：０１／ＤＰＡ１＊０１：０３）、１種のＨＬＡ－ＤＱ対立遺伝子（ＤＱＢ１＊０６：０２／ＤＱＡ１＊０１：０２）、および２種のクラスＩ対立遺伝子（表１）が含まれた。全ての場合において、２～３回の反復実験が少なくとも１５００種の独特のペプチドを観察するのに十分であった（図１１Ｂ）。プロファイリングされた対立遺伝子の中で、ほんの小さなパーセンテージのヒットが既知の夾雑物または完全なトリプシン消化物に対応した；他方では、ニセトランスフェクションにより比較的少数のペプチドが返され、それらは既知の夾雑物または完全なトリプシン消化物として大部分が同定可能であった（図１１Ｂ）。

表１は、例示的な実験に使用したサンプルの要約を示す。

【表１】

表 1

ドナー	DR1	DR3/4/5
D001000763	DRB1*03:01, DRB1*11:01	DRB3*01:01, DRB3*02:02
HD84	DRB1*01:01	存在せず
RG1248	DRB1*03:01	DRB3*01:01
RG1095	DRB1*03:01, DRB1*11:01	DRB3*01:01, DRB3*02:02
RG1104	DRB1*01:01, DRB1*11:01	DRB3*02:02
2010113472	DRB1*01:01, DRB1*11:01	DRB3*02:02
2010113438	DRB1*03:01, DRB1*11:01	DRB3*01:01, DRB3*02:02

【０６２６】

ＭＨＣ　ＩＩ結合性ペプチドの末端がＭＨＣ結合性溝内にフィットする必要はないので、多数の別個のペプチド種が同じコア結合性配列を共有すれば、それらは同等に良好に結合し得る。重複する配列を有するペプチドを「入れ子セット」内にプールしたところ、ＨＬＡクラスＩＩ対立遺伝子当たり５００～７００種の独特の入れ子セットが観察され、これらは、一般には５００～６００種の独特の遺伝子に由来するものであった。ＨＬＡクラスＩおよびＨＬＡクラスＩＩ結合性ペプチドの長さ分布は抗体に基づくプルダウンを使用した以前のＭＳ試験において観察されたものとマッチする（図１１Ｃ）。

【０６２７】

推定上のＭＨＣ結合性ペプチドの中で、大多数のアミノ酸がそれらの供給源プロテオーム頻度と一致するレベルで表された。例外にはシステイン、メチオニン、およびトリプトファンが含まれ、これらは、ＭＨＣ　ＩＩペプチドに関する以前のＭＳに基づく試験と一致して枯渇した。ＩＥＤＢからの対立遺伝子を釣り合わせた高親和性ペプチド（＜５０ｎＭ）ではシステイン、メチオニン、およびトリプトファンの枯渇は観察されなかった；しかし、ＩＥＤＢペプチドでは、プロテオームに関して、ロイシンおよびメチオニンの富化ならびにプロリン、アスパラギン酸、およびグルタミン酸の枯渇が示された。

【０６２８】

（実施例４）

ＭＡＰＴＡＣ（商標）プロトコールにより既知のおよび新規のＭＨＣ　ＩＩ結合性モチーフが明らかになる

クラスＩＩペプチドのＭＨＣ結合性部分配列はＮ末端またはＣ末端に対して固定された位置にないので、正確なクラスＩＩモチーフの発見では、各結合体ペプチドについて異なる結合レジスターの可能性を動的に検討しなければならない。Ｇｉｂｂ’ｓ　Ｃｌｕｓｔｅｒツールでは、期待値最大化（ＥＭ）アルゴリズムによってこの難題に対処する。畳み込みニューラルネットワーク（ＣＮＮ）を使用した新規モチーフ発見手法の使用を探求した。ＣＮＮは、コンピュータビジョンの分野では成功しており、同様に、翻訳不変パターン認識を実現しようとしている。ＣＮＮを、ＭＨＣ結合性ペプチドとそれらのスクランブルバージョンを区別するように訓練し、次いで、最後から２番目のネットワーク層において最大のノード活性化を実現した部分配列に従って陽性例をアラインメントした。単一対立遺伝子ＭＳデータに適用する場合、この手法では、ギブスクラスタリングと一致したモチーフが得られ、また、総体的な１位、４位、６位、および９位にアンカーが示された（図１３）。これらのモチーフは、ＩＥＤＢからの高親和性結合体（親和性＜５０ｎＭ；図１３）について観察されたＣＮＮ由来のモチーフと高度に一致した。ＤＲＢ１＊１１：０１に関しては、モチーフが細胞株にわたって安定であり、以前に汎ＤＲ抗体を用いてプロファイリングされたＤＲＢ１＊１１：０１ホモ接合性細胞株と一致することをさらに検証した。同様に、ＭＨＣクラスＩ対立遺伝子に関して導き出されたモチーフは、親和性に基づく方法および以前のＭＳに基づく試験によるものと一致した（図１４Ａ）。

【０６２９】

ＭＨＣクラスＩＩ対立遺伝子の全てが識別可能なモチーフを示したが、アンカー位置のエントロピーはＭＨＣクラスＩ対立遺伝子について観察されたものよりも顕著に高かった。したがって、各ＭＨＣクラスＩＩ対立遺伝子について各アンカー位置の好ましいアミノ酸を定義し、ペプチドの１０～２０％だけが４つのアンカー位置全てにおいて理想的な残基を示し、６０％もの多くが２つまたはそれよりも少ない予測されたアンカーを示すことが観察された（図１４Ｂおよび図３０Ｃ）。１３～１７ｍｅｒを結合潜在性についてＮｅｔＭＨＣＩＩｐａｎを使用してスコア化し、一方、ＭＳにより観察されたペプチドは全ての場合において予測された結合潜在性に関して富化され、スコアに、長さを釣り合わせたランダムなペプチドとの有意な重複が存在した（図１４Ｃ、および図３６Ａ）。

【０６３０】

（実施例５）

単一対立遺伝子ＭＨＣ　ＩＩ　ＭＳデータに対して訓練されたアルゴリズムにより免疫原性が予測される

次に、単一対立遺伝子ＭＳプラットフォームからのデータにより改善されたＭＨＣクラスＩＩ結合予測器を生成することができるかどうかを検討した。ＣＮＮ手法を足場として、フィルターサイズ、スキップ接続、および総受容野を有する多層ネットワークを創出した（図３１Ａ）。ｎｅｏｎｍｈｃ２と称されるこのディープラーニングモデルを訓練し、評価するために、プロテオームを遺伝子の７５％、１２．５％、および１２．５％を表す３つの区分に区分した。第１の区分はＣＮＮを確率的勾配降下法によって訓練するために使用し、第２の区分はアーキテクチャおよびハイパーパラメータ最適化のために使用した。第３の区分は、解析の最後に１回だけ、性能を評価するために使用した。評価の完全性

を確実にするために、パラロガスな遺伝子群内の全ての遺伝子が同じ区分に入るように注意を払った。

【０６３１】

ＭＳでは特にシステインに対してある程度の残基の偏りが示されるので（図１２Ｄ）、この問題は、陽性例の配列をランダムに並べ替えることによって生成した陰性訓練例（デコイと称される）を使用することによって軽減される。この手法には、天然のタンパク質の配列特性を学習するというリスクがあるので、それにより予測性能が人為的に膨張する可能性がある、モデル評価では、デコイをペプチド供給源遺伝子の観察されていない部分配列からランダムにサンプリングする別個のデコイ生成戦略を使用する。ヒット対デコイ比１：１９での陽性適中率（ＰＰＶ）の算出により、評価区分におけるＭＳペプチドの予測に関してｎｅｏｎｍｈｃ２がＮｅｔＭＨＣＩＩｐａｎと比べて改善されたＰＰＶを有することが示された（図４および図３１Ｂ）。ｎｅｏｎｍｈｃ２の訓練用データセットのサイズを人為的にダウンサンプリングする実験により、その性能がデータにより限定され、より深いカバレッジのデータを用いることで改善されることが示唆される（図１６）。

【０６３２】

ＮｅｔＭＨＣＩＩｐａｎを訓練したデータ型である結合親和性を予測するためにｎｅｏｎｍｈｃ２の能力を探求した。ＮｅｔＭＨＣｐａｎから同じペプチド測定値に対して訓練および評価を行う利点を取り除くために、ＩＥＤＢに寄託されたわずかに古いバージョンのＮｅｔＭＨＣＩＩｐａｎによりスコア化されたペプチドを使用して評価を実行した。ケンドールのタウ統計値を使用して予測の正確度を評価すると、全ての場合において、ＮｅｔＭＨＣＩＩｐａｎスコアはＭＳに基づく予測器と同様であったかまたはわずかに良好であった（図１５Ｂ）。興味深いことに、性能は、実施した親和性アッセイの型に依存した。ｎｅｏｎｍｈｃ２は、Ｓｅｔｔｅおよび共同研究者からの親和性測定値に関して予測した場合にはＮｅｔＭＨＣＩＩｐａｎにわずかに後れをとったが、Ｂｕｕｓおよび共同研究者からの測定値に関して予測した場合にはＮｅｔＭＨＣＩＩｐａｎにより大きな後れをとった。これらの結果をまとめて考慮して、これらのプラットフォーム間に内在性差異が存在すると思われたが、いずれの手法がより間違いのないものであるかはすぐには明らかにならなかった。

【０６３３】

明瞭さの改善を実現するために、天然のＣＤ４　Ｔ細胞応答を予測する能力を評価した。ＩＥＤＢからのデータは、一般に、応答の対立遺伝子制限が大抵未定義であるかまたは帰属かいずれかであるので、この目的のためには不適当であった。したがって、四量体によりガイドされるエピトープマッピング（ＴＧＥＭ）データの大きなデータセットをアセンブルした。これらの試験の全てで、予測への優先順位付けではなく、包括的な重複ペプチドスクリーニングを使用し、それにより、ＮｅｔＭＨＣＩＩｐａｎに有利な観察の偏りが取り除かれた。一方で、対立遺伝子制限は明白である。評価のために十分なデータが存在した全ての対立遺伝子に関しては、ｎｅｏｎｍｈｃ２は、ランダムよりもほんのわずかだけ性能が良かったＮｅｔＭＨＣｐａｎよりも実質的に優れた。したがって、ＭＡＰＴＡＣ（商標）プラットフォームは、免疫原性ＭＨＣクラスＩＩエピトープを同定するモデルを訓練するためのベストインクラスであり得る。

【０６３４】

（実施例６）

複対立遺伝子ＭＳデータに対して訓練されたアルゴリズムは劣る

標準の汎ＤＲおよび汎ＩＩ抗体精製に基づく多数の複対立遺伝子クラスＩＩデータベースがパブリックドメインに存在することを考慮して、適切な予測器を複対立遺伝子データのみを使用して訓練することができるかどうかを試験した。いくつかのグループがＭＨＣクラスＩ対立遺伝子モチーフの複対立遺伝子クラスＩデータからの逆畳み込みに関して成功を示しているが、これらの試みはまだ公的に入手可能な予測器には変換されていない。クラスＩＩモチーフの逆畳み込みは、各ペプチドの結合レジスターとクラスターメンバーシップの両方を同時に分解する必要性によってさらに複雑になる。クラスＩＩ逆畳み込み

の可能性を探究するためにギブスクラスターツールが使用されているが、この手法の忠実度は広範囲には検証されていない。

【０６３５】

クラスＩＩ逆畳み込みの正確度を評価するために、遺伝子型が公知である公的に入手可能な汎ＤＲデータセットを選択した。各データセットについて、我々の単一対立遺伝子データのうち２０ペプチドをドナーの遺伝子型の各対立遺伝子に対してスパイクインした（ハプロタイプおよび接合性に応じて１～２のＤＲ１対立遺伝子プラス０～２のＤＲ３／４／５対立遺伝子）。ギブスクラスタリングツールを各データセットに対して実行し、スパイクインペプチドが適正に共クラスター化されたかどうかをそれらの既知の対立遺伝子起源に従って観察した。この解析の初期バージョンでは、対立遺伝子数に対するクラスター数を固定したか、またはギブスクラスターにより最適なクラスターの数を自動的に決定されるようにした。しかし、いずれの手法でもペプチドが正確には逆畳み込みされなかったと思われる）。アルゴリズムを補助するために、スパイクインペプチドの真の供給源対立遺伝子とそれらが割り当てられたクラスターの間の調整された相互情報を算出することによって最適なクラスター計数を選択した。それにもかかわらず、全てではないが一部の場合に、ペプチドが、それらの供給源対立遺伝子を考慮せずに多様なクラスターにわたって分布した（図１７Ａ）。これらの結果により、現行の逆畳み込みプロトコールがＭＨＣクラスＩＩに対しては確実に正確なものではない可能性があることが示唆される。

【０６３６】

この解析に対する警告の１つは、一部のペプチドは１つよりも多くの対立遺伝子に結合することができることである。これに沿って、次の疑問は、それにもかかわらず複対立遺伝子データに由来する結合性モチーフが単一対立遺伝子データから観察されたものと合理的にマッチし得るかどうかである。これを評価するために、単一の対立遺伝子の各々のペプチドを捕捉するために最良の対応を有するクラスターを選択し、これらの集団に基づいたモチーフを築いた（例えば、図１７Ｂを参照されたい）。既知のアンカーのいくつかは多くのモチーフにより明白に実証されたが、他の位置は単一対立遺伝子のモチーフとは不調和である、または供給源データセット間で不調和である。さらに、偽性アンカーが現れた明らかな場合があった。最後に、逆畳み込みされたデータを、ペプチドを我々の単一対立遺伝子データセットの評価区分に予測することができたＣＮＮを訓練するために使用することができるかどうかを評価した。逆畳み込みされた複対立遺伝子データに対して訓練されたモデルは、全ての場合においてＭＡＰＴＡＣ（商標）訓練モデルに達しなかった（図１７Ｃ）。

【０６３７】

（実施例７）

供給源タンパク質の特色が提示可能性に影響を及ぼす

ＭＨＣクラスＩに関しては、提示されるエピトープのレパートリーの決定、したがって、タンパク質からペプチドへのプロセシングが特徴付けられたクラスＩＩレパートリーをどのように形づくるかにおいてプロテアソームが重要な役割を果たす。

【０６３８】

まず、いくつかの組織に基づくペプチドプロファイリングデータセットにおいて観察されたＭＨＣクラスＩＩペプチドのＮ末端およびＣ末端の正確な位置に焦点を当てた。位置に基づくアミノ酸頻度をデコイペプチドに対して比較することにより、有意な富化および枯渇を観察した。このパターンは、最近の観察と一致する。興味深いことに、全体的なパターンは、最良の特徴付けられたクラスＩＩプロセシング酵素であるカテプシンＳの既知の切断優先性（［ＲＰＩ］［ＦＭＬＷ］［ＫＱＴＲ］［ＡＬＳ］）とマッチしない。

【０６３９】

このモチーフの予測潜在性を決定するために、Ｎ末端およびＣ末端に対するＮＮに基づく予測器を築き、２つの切断変数を予測された結合潜在性（ＭＳ訓練ＣＮＮ毎）と一緒に使用したロジスティック回帰を当てはめて、同じ供給源遺伝子からサンプリングした真のＭＳペプチドと長さを釣り合わせたデコイペプチドとを区別した。

【０６４０】
この予測器では、結合潜在性を単独で考慮するモデルに対してペプチド予測のわずかな改善がもたらされたが、ＭＨＣクラスＩＩ結合性エピトープ（互換的にクラスＩＩエピトープと称される）の免疫原性はペプチド切断の正確な位置には依存しない可能性があるので、疑問は、切断の正確な部位が未知の場合に、それでもこのモデルに価値が付加されるかどうかである。したがって、ヒットおよびデコイの正確な切断位置を差し控え、その代わりに帰属結合性コアの付近（＋／－１５ＡＡ）のタンパク質位置にわたる複合切断性スコアをスコア化して、二度目予測スキームを実行した。興味深いことに、結合性のみの予測器に対して性能に改善は認められなかった。これらの結果は、クラスＩＩ切断予測の追加により、ＭＳにより観察されたリガンドの予測を改善することはできるが、おそらく正確なペプチド末端に依存しないＴ細胞認識は改善することができないことが示された以前の研究と一致する。

【０６４１】
ＭＨＣ　ＩＩペプチドの有意な分率がＭＨＣとの結合後にそれらのＮ末端およびＣ末端から「チュードバック」されるモデルが提唱された。このモデルの下では、プロリンによりエキソペプチダーゼの進行が遮断されるので、最後から２番目のプロリンシグネチャーが生じる。このシナリオでは、ＭＨＣリガンド末端の直接解析に由来するモチーフは、ペプチド断片生成の最初のステップではなく下流の編集を反映するので、潜在的に誤った方向に導くものである。したがって、クラスＩＩペプチドの付近の、それらの生成を説明することができる可能性がある他の配列特色を決定した。まず、規準のカテプシンＳシグネチャーを検索したが、ＭＳにより観察されたクラスＩＩペプチド付近で、ペプチド供給源遺伝子からサンプリングした長さを釣り合わせたデコイペプチドに対してカテプシンＳ部位の富化は認められなかった。このプロセシングシグネチャーは酵素の複雑なアンサンブルを反映する可能性があるので、新規のＣＮＮを、観察されたペプチドおよびデコイの周囲の上流および下流のタンパク質コンテキスト（＋－２５ＡＡ）に基づいて訓練した。

【０６４２】
ペプチド利用可能性を、タンパク質の一次配列ではなく、フォールディングされたまたはセミアンフォールディングされた状態によって決定する第３のモデルを検討した。相同性に基づくＡＣＣＰＲＯを使用して二次構造および溶媒がアクセス可能な領域を予測し、予測器のアンサンブルを使用して内因的に不規則なドメインを同定した。

【０６４３】
プロセシング優先領域を本質的に予測することが難しい場合、以前に公開された複対立遺伝子クラスＩＩ　ＭＳデータの大きな集合内の少なくとも１つのペプチドが及ぶタンパク質領域全ての一覧表を単に築き、重複を予測特色として使用することが可能であり得る。明らかに、以前に公開されたデータで表される対立遺伝子は同じまたは同様の結合性モチーフを有する可能性があるので、重複特色は結合情報により夾雑される。それにもかかわらず、この特色であっても、提示されるペプチドの予測がわずかに改善され、これにより、ＭＨＣクラスＩＩペプチドが強力なプロセシングホットスポットに供されない可能性があることが示唆される。

【０６４４】
次の疑問は、どの遺伝子がクラスＩＩ結合性ペプチドレパートリーに最も大きく寄与するかであった。ＭＨＣクラスＩリガンドを予測する場合、発現レベルなどの遺伝子レベル特色により、大きなブーストがもたらされることがすでに分かっている。以前に公開された、ヒト組織のクラスＩＩ結合レパートリーをプロファイリングするＭＳデータセットを活用し、ＭＳにより観察されたペプチドが、ランダムなデコイペプチド（プロテオームからサンプリングしたもの）よりも１桁高度に発現されることが観察された（図１８Ａ）。それにもかかわらず、遺伝子にマッピングされたクラスＩＩペプチドの約５％が、代表的なＲＮＡ－Ｓｅｑデータによると表面上は発現されないことが認められた。このパターンに基づいて、各遺伝子がクラスＩＩペプチドレパートリーにおいて過剰にまたは過少に表される程度を、各遺伝子の観察の数がその長さと発現レベルの積に比例するはずであると

いうベースライン予想を提唱することによって数量化しようと試みた（図１８Ｂ）。過剰に表された遺伝子の中で、ヒト組織血清において発現される、多くのクラスＩＩ結合性ペプチドを産生するが、ネイティブな組織では表面上は発現されないタンパク質の明らかな富化が認められた。これは、細胞外環境から試料採取された抗原の提示におけるＭＨＣクラスＩＩの公知の役割と一致する。

【０６４５】

オートファジーは、別の十分に確立されたクラスＩＩプロセシング経路であるので、各遺伝子（観察されたペプチドが５つよりも少なく、予測されたペプチドが５つよりも少ないあらゆる遺伝子を除外する）についての観察されたペプチド対予測されたペプチドの比を決定し、公知のオートファジー遺伝子またはマウスにおけるＡｔｇ５ノックアウトによって安定化される遺伝子の物理的パートナーに対して富化が認められるかどうかを決定した（図１８Ｃ）。クラスＩＩデータにおけるいずれの遺伝子セットも富化されないと思われた。実際、オートファジー遺伝子の物理的パートナーは適度に過少に表されると思われた。

【０６４６】

全ての細胞内局在を見渡すと（図１８Ｄおよび図１８Ｅ）、少数の区画が決定的に過剰にまたは過少に表された。２つの最も富化された区画は細胞膜およびリソソームであり、各々からおよそ２倍のクラスＩＩペプチドの予測数が生じた。膜タンパク質の富化が、オートファゴソームへの膜再利用または膜タンパク質の直接オートファジー経路へのゴルジ経路設定に関するかどうかは明らかでない。リソソームタンパク質の富化と以前に観察されたオートファジー遺伝子の枯渇との間の見かけの矛盾により、これらの傾向が、検討されているオートファジー関連遺伝子の特異的なサブセットに対して高度に感受性であることが示された。図１８Ｆは、２つの異なる遺伝子発現プロファイル、バルク腫瘍およびプロフェッショナル抗原提示細胞に関するペプチド観察の相対的一致を示す。

【０６４７】

（実施例８）

正確なＭＨＣ　ＩＩ予測にはエンドサトーシス経路を理解することが必要である

クラスＩＩ遺伝子の供給源経路を理解することに加えて、どの細胞型が大多数のクラスＩＩ提示を担うかを理解することが極めて重要であり得る。がんの場合では、線維芽細胞を含めた非プロフェッショナルＡＰＣおよび腫瘍自体が、炎症を起こした腫瘍微小環境（ＴＭＥ）内でクラスＩＩを提示すると考えられている。さらなる洞察を得るために、肺がん、頭頸部がん、および黒色腫をプロファイリングしたものである３つの最近公開された単一細胞ＲＮＡ－Ｓｅｑデータセットにおいて ＨＬＡ－ＤＲＢ１発現を解析した。細胞にわたって患者－細胞型レベルを平均することにより、規準のＡＰＣ（マクロファージ、樹状細胞、およびＢ細胞）が、腫瘍および他の間質細胞型よりもはるかに高レベルのクラスＩＩを提示し、この傾向が多数の患者および腫瘍型にわたって一致することが明らかになった。

【０６４８】

免疫治療によりこの傾向が乱されるかどうかを探索するために、チェックポイント遮断応答性腫瘍型からの追加的な単一細胞ＲＮＡ－Ｓｅｑを解析し、処置の前後のＨＬＡ－ＤＲＢ１発現を評価した。１つの確認された応答者を含んだ黒色腫コホートでは、治療前生検材料および治療後生検材料のどちらにおいても腫瘍細胞による一様に低いＨＬＡ－ＤＲＢ１発現が示された（図１９Ｃ）。抗ＰＤ－１治療に対して５５％臨床応答率を示した基底細胞癌コホートは、同様に、時点にかかわらず、腫瘍細胞由来ＨＬＡ－ＤＲＢ１の低発現を示した（図１９Ｃ）。

【０６４９】

これらの結果により、大多数の腫瘍内ＨＬＡクラスＩＩ提示が主にプロフェッショナルＡＰＣによって駆動され、「ホット」ＴＭＥ条件では一般的なパターンとの相違が保証されないことが示唆された。

【０６５０】

腫瘍細胞は腫瘍微小環境ではＡＰＣに数で勝るので、それにもかかわらずＭＨＣクラスＩＩ発現のレベルが低いことには、免疫学的関連性がある可能性がある。いかほどの全体的なクラスＩＩ発現が腫瘍細胞によってもたらされるか、対していかほどが間質によってもたらされるかを評価するために、クラスＩＩ特異的遺伝子に変異を有するＴＣＧＡ患者（ＣＩＩＴＡ、ＣＤ７４、およびＣＴＳＳに焦点を当てる）を同定し、体細胞（腫瘍特異的）バリアントを示したＲＮＡ－Ｓｅｑリードの画分を決定した。この情報を使用して、ＨＬＡ－ＤＲＢ１発現のどの画分が腫瘍に由来するか、対してどの画分が間質に由来するかを帰属させた（図１９Ｂ）。１７種の別個の腫瘍型を表す１５３名の患者において同定された変異に基づいて、大多数のクラスＩＩ発現が非腫瘍細胞から生じると思われる優性パターンが観察された。最高レベルのＴ細胞浸潤を有する患者だけに焦点を当て（上位１０％、以前に公開された１８遺伝子シグネチャー（Ayers et al., 2017）を使用して同定されたもの）腫瘍のＨＬＡ－ＤＲ低発現がそれでも正常だと思われ、１６名の患者のうち３名のみが＞１０００ＴＰＭ（腫瘍増悪および転移）を発現した。

【０６５１】

免疫治療によりこの傾向が乱されるかどうかを探索するために、チェックポイント遮断応答性腫瘍型からの追加的な単一細胞ＲＮＡ－Ｓｅｑを解析し、処置の前後のＨＬＡ－ＤＲＢ１発現を評価した。１つの確認された応答者を含んだ黒色腫コホートでは、治療前生検材料および治療後生検材料のどちらにおいても腫瘍細胞による一様に低いＨＬＡ－ＤＲＢ１発現が示された（図１９Ｃ）。抗ＰＤ－１治療に対して５５％臨床応答率を示した基底細胞癌コホートは、同様に、時点にかかわらず、腫瘍細胞由来ＨＬＡ－ＤＲＢ１の低発現を示した（図１９Ｃ）。

【０６５２】

これらの結果により、大多数の腫瘍内ＨＬＡクラスＩＩ提示が主にプロフェッショナルＡＰＣによって駆動され、「ホット」ＴＭＥ条件では一般的なパターンとの相違が保証されないことが示唆された。

【０６５３】

（実施例９）

新しい予測概念により、免疫原性新抗原のより正確な同定が可能になる

ｎｅｏｎｍｈｃ２および関連するプロセシング規則の有用性を探求するために、いくつかの予測シナリオにおける性能を検討した。まず、ＭＳにより同定されたペプチドを予測する能力を、汎ＤＲ抗体を用いてプロファイリングされた、７名の健康なドナー由来のＰＭＢＣにおいて評価した。この解析により、ＭＡＰＴＡＣ（商標）システムまたは我々の産生細胞株に固有のあらゆる系統的な偏りを制御することができる。ヒットとデコイを１：４９９の比で使用し、タンパク質をコードするエクソームからデコイをランダムにサンプリングして、ｎｅｏｎｍｈｃ２およびＮｅｔＭＨＣＩＩｐａｎベースのモデルならびに追加的なプロセシング特色が組み入れられたモデルの陽性適中率を評価した（図１８Ｂによる、発現、遺伝子レベルの偏り、および以前のＭＨＣ　ＩＩペプチドとの重複）。これらのモデルにより、結合予測およびプロセシング予測のどちらに関しても実質的な改善が確認された（図２０）。

【０６５４】

図２１Ａは、示されているさらなるプロセシングパラメータまたは特色を用いたＮｅｔＭＨＣＩＩｐａｎとｎｅｏｎｍｈｃ２の比較を示す。ＨＬＡ－ＤＲ抗体によってプロファイリングされた８つのＭＳ試料についての予測性能（実施例６において分析されたものと同じ試料、図１７Ａ）。予測器は、ＨＬＡ結合予測（ＮｅｔＭＨＣＩＩｐａｎまたはｎｅｏｎｍｈｃ２のいずれか）を最小に使用し、必要に応じて追加的なプロセシング関連変数：遺伝子発現、遺伝子の偏り（例えば、図１８Ｂ、図１８Ｃ、図１９Ｂによる）、および以前に観察されたＨＬＡ－ＤＱペプチドとの重複を使用する。本実施例では、デコイを、入手可能なデコイ配列をほぼ飽和させるヒット対デコイの比１：４９９が実現されるようにプロテオームからランダムにサンプリングした（ＭＳにより観察されたペプチドを決して産生しない遺伝子を含む）。図４と類似した様式で陽性適中率を算出した。例えば、ペ

プチドの上位０．２％を陽性としてコールし、ＰＰＶは、真のＭＳにより観察されたペプチドである陽性の分率である。各試料中の各候補ペプチドについて、結合スコアを、試料遺伝子型に存在するＨＬＡ－ＤＲ対立遺伝子にわたる最大値として算出した。２つの方法によって見いだされたペプチドの傾向にはかなりの相関があるが、本明細書に記載のモデルではより頑強な転帰が示される。図２１Ｂ（図３３Ｂも参照されたい）は、図２１Ａにおけるものと同じヒット：デコイ比および性能メトリクスを使用した、樹状細胞（溶解物）によって提示される腫瘍由来ペプチドについての予測性能を表す。プロセシング特色の使用を伴うまたは伴わないＮｅｔＭＨＣＩＩｐａｎおよび本明細書に記載のモデルについての性能が示されている。図２１Ｃは、重く標識されたペプチド各々についての発現レベルおよび遺伝子の偏りのスコアを示す。図２１Ｄは、溶解物およびＵＶ処理実験による重く標識されたペプチド供給源遺伝子の重複を表す図である。

【０６５５】

（実施例１０）

細胞株におけるクラスＩＩ　ＨＬＡペプチドの発現およびＭＨＣ－ＩＩと結合したペプチドの単離

構築物の設計、細胞培養およびＨＬＡ－ペプチド免疫沈降

本例示的試験では、単一の親和性タグ付けされたＨＬＡ構築物を細胞株（Ａ３７５、ＨＥＫ２９３Ｔ、Ｅｘｐｉ２９３、ＨｅＬａ）にトランスフェクトし、親和性タグ付けされたＨＬＡ－ペプチド複合体を免疫沈降させることによって単一対立遺伝子細胞株を生成した。図１２Ａおよび１２Ｅでは、ＭＨＣクラスＩＩ対立遺伝子頻度は、特に断りのない限り、ａｌｌｅｌｅｆｒｅｑｕｅｎｃｉｅｓ．ｎｅｔ／から得られた対立遺伝子頻度である。米国集団に関する対立遺伝子頻度は、６２．３％ヨーロッパ人、１３．３％アフリカ人、６．８％アジア人、および１７．６％ヒスパニックの混合を仮定することによって帰属させた。

【０６５６】

図１２Ａおよび図１２Ｅに関しては、ｍｈｃ＿ｌｉｇａｎｄ＿ｆｕｌｌ．ｃｓｖデータセットを２０１８年９月２１日にＩＥＤＢデータからダウンロードした。「細胞ＭＨＣ／競合／蛍光」、「細胞ＭＨＣ／競合／放射活性」、「細胞ＭＨＣ／直接／蛍光」、「精製されたＭＨＣ／競合／蛍光」、「精製されたＭＨＣ／競合／放射活性」、または「精製されたＭＨＣ／直接／蛍光」と同等の「方法／技法」および「解離定数ＫＤ」、「解離定数ＫＤ（およそのＥＣ５０）」、「解離定数ＫＤ（およそのＩＣ５０）」、「最大半量の有効濃度（ＥＣ５０）」、または「最大半数阻害濃度（ＩＣ５０）」と同等の「アッセイ群」を有するためには有効な親和性測定が必要であった。「著者」欄に列「Ｂｕｕｓ」が表示されている場合、測定値はＳｏｒｅｎ　Ｂｕｕｓグループ（Ｕｎｉｖｅｒｓｉｔｙ　ｏｆ　Ｃｏｐｅｎｈａｇｅｎ、Ｄｅｎｍａｒｋ）によるものであった。他の場合では、著者欄に列「Ｓｅｔｔｅ」または「Ｓｉｄｎｅｙ」が含まれる場合、測定値はＡｌｅｓｓａｎｄｒｏ　Ｓｅｔｔｅ　ｇｒｏｕｐ（Ｌａ　Ｊｏｌｌａ　Ｉｎｓｔｉｔｕｔｅ　ｆｏｒ　Ｉｍｍｕｎｏｌｏｇｙ、Ｕ．Ｓ．Ａ）によるものであった。他の全ての測定値は「他」と標識した。強力な結合体を数え上げるために、測定された親和性が５０ｎＭよりも強力なペプチドのみを計数した（図１２Ａ）。図１２Ｅはｔｏｏｌｓ．ｉｅｄｂ．ｏｒｇ／ｍａｉｎ／ｄａｔａｓｅｔｓ／からの追加的なデータを含み、親和性が＜１００ｎＭである強力な結合体を数え上げる。

【０６５７】

ＤＮＡ構築物の設計

ＨＬＡクラスＩおよびＨＬＡクラスＩＩ対立遺伝子の遺伝子配列をＩＰＤ－ＩＭＧＴ／ＨＬＡウェブページ（ｅｂｉ．ａｃ．ｕｋ／ｉｐｄ／ｉｍｇｔ／ｈｌａ）によって同定し、組換え発現構築物の設計に使用した。ＨＬＡクラスＩに関しては、α鎖をＣ末端ＧＳＧＧＳＧＧＳＡＧＧリンカー（配列番号１０）と融合し、その後にビオチン－アクセプター－ペプチド（ＢＡＰ）タグ配列ＧＬＮＤＩＦＥＡＱＫＩＥＷＨＥ（配列番号１１）、終止コドン、および変動するＤＮＡバーコードを続け、ｐＳＦ　Ｌｅｎｔｉベクター（Ｏｘｆ

ｏｒｄ Ｇｅｎｅｔｉｃｓ、Ｏｘｆｏｒｄ、ＵＫ）にＮｃｏＩおよびＸｂａＩ制限部位を介してクローニングした。ＨＬＡクラスＩＩ構築物は、同様にｐＳＦ ＬｅｎｔｉにＮｃｏＩおよびＸｂａＩ制限部位を介してクローニングし、Ｃ末端においてクラスＩ構築物由来のリンカー－ＢＡＰ配列（ＳＧＧＳＧＧＳＡＧＧＧＬＮＤＩＦＥＡＱＫＩＥＷＨＥ（配列番号１２））、その後、別の短いＧＳＧリンカー、Ｆ２Ａリボソームスキッピング配列（ＶＫＱＴＬＮＦＤＬＬＫＬＡＧＤＶＥＳＮＰＧＰ（配列番号１３））、α鎖の配列、ＨＡタグ（ＧＳＹＰＹＤＶＰＤＹＡ（配列番号１４））、終止コドン、および変動するＤＮＡバーコードと融合したβ鎖配列からなるものであった。全てのＤＮＡ配列の同一性をサンガーシーケンシングによって検証した。

【０６５８】

細胞培養および一過性トランスフェクション

Ｅｘｐｉ２９３細胞（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）を、Ｅｘｐｉ２９３培地（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）中、８％CO_2、３７℃、１２５ｒｐｍで振とうしながら成長させた。Ｅｘｐｉ２９３細胞を、定期的な隔週での継代を伴って、０．５×10^6個／ｍＬから６×10^6個／ｍＬの間の細胞密度で維持した。Ｅｘｐｉ２９３細胞懸濁液３０ｍＬを、およそ３×10^6個／ｍＬの細胞密度および＞９０％生存能力での一過性トランスフェクションのために使用した。簡単に述べると、ＤＮＡ３０μｇ（細胞懸濁液１ｍＬ当たり１μｇ／ｍＬのＤＮＡ）を１つのチューブ中、Ｏｐｔｉ－ＭＥＭ培地（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）１．５ｍＬに希釈し、一方、ＥｘｐｉＦｅｃｔａｍｉｎｅ（商標）２９３トランスフェクション試薬（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）８０μＬをＯｐｔｉ－ＭＥＭ、１．５ｍＬを含有する第２のチューブ中に希釈した。これらの２つのチューブを室温で５分間インキュベートし、組み合わせ、穏やかに混合し、室温で３０分間インキュベートした。ＤＮＡとＥｘｐｉＦｅｃｔａｍｉｎｅの混合物をＥｘｐｉ２９３細胞に添加し、３７℃、８％CO_2、８０％相対湿度でインキュベートした。４８時間後、トランスフェクトされた細胞を４回の技術的反復実験においてチューブ当たり細胞５０×10^6個で回収し、遠心分離し、１×Ｇｉｂｃｏ ＤＰＢＳ（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）で１回洗浄し、質量分析のために液体窒素で急速冷凍した。細胞１×10^6個の一定分量を各トランスフェクションバッチから収集し、抗ＢＡＰ（Ｒｏｃｋｌａｎｄ Ｉｍｍｕｎｏｃｈｅｍｉｃａｌｓ Ｉｎｃ．、Ｌｉｍｅｒｉｃｋ、ＰＡ）または抗ＨＡ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）ウエスタンブロットによって分析して、親和性タグ付けされたＨＬＡタンパク質発現を検証した。

【０６５９】

Ａ３７５細胞（ＡＴＣＣ）を、１０％ＦＢＳを伴うＤＭＥＭ中で成長させ、定期的な継代を伴って、８０％以下の集密での培養で維持した。質量分析実験のために、Ａ３７５細胞を、５００cm^2のプレートにおいて、７０％集密の細胞数から算出された通り、１００ｍＬ中、細胞１８．５×10^6個／ｍＬの播種密度で培養した。２４時間後、総培養体積に関して調整したＴｒａｎｓＩＴシステムプロトコールに従って細胞にＴｒａｎｓＩＴ－Ｘ２（Ｍｉｒｕｓ Ｂｉｏ、Ｍａｄｉｓｏｎ、ＷＩ）をトランスフェクトした。４８時間後、細胞培地を吸引し、細胞を１×Ｇｉｂｃｏ ＤＰＢＳ（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）で洗浄した。回収のために、Ａ３７５細胞を非酵素的細胞解離溶液（Ｓｉｇｍａ－Ａｌｄｒｉｃｈ）３０ｍＬと一緒に３７℃で１０分間インキュベートし、遠心分離し、１×ＤＰＢＳで洗浄し、試料当たり５０×10^6個の細胞で分注した。２９３Ｔ細胞およびＨｅＬａ細胞をＡＴＣＣから購入し、それぞれＤＭＥＭ、１０％ＦＢＳ、２ｍＭのＬ－グルタミンまたはＤＭＥＭ＋１０％ＦＢＳ中、３７℃、５％CO^2で培養した。両方の細胞株に、ＨＬＡ構築物をＴｒａｎｓＩＴ ＬＴ１試薬（Ｍｉｒｕｓ Ｂｉｏ、Ｍａｄｉｓｏｎ、ＷＩ）を使用して製造者の指示に従ってトランスフェクトし、トランスフェクションの４８時間後に、Ａ３７５細胞について記載されているものと同様に処理した。全ての試料から、細胞１×10^6個の一定分量を各トランスフェクションで収集し、抗ＢＡＰ（Ｒｏｃｋｌａｎｄ Ｉｍｍｕｎｏｃｈｅｍｉｃａｌｓ Ｉｎｃ．、Ｌｉｍｅｒｉｃｋ、ＰＡ）または抗ＨＡ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）ウエスタンブロットによっ

て分析して、親和性タグ付けされたＨＬＡタンパク質発現を検証した。
【０６６０】
ＢｉｒＡタンパク質の発現および精製
Ｃ末端ヘキサ－ヒスチジンタグ（配列番号１５）と融合したＥ．ｃｏｌｉ ＢｉｒＡをコードするｐＥＴ１９ベクターを使用した。化学的コンピテントＥ．ｃｏｌｉ ＢＬ２１（ＤＥ３）細胞（Ｎｅｗ Ｅｎｇｌａｎｄ Ｂｉｏｌａｂｓ）を、ＢｉｒＡ発現プラスミドで形質転換し、ＬＢブロスプラス１００μｇ／ｍｌのアンピシリン中、３７℃でOD_{600}が０．６～０．８になるまで成長させ、３０℃まで冷却した後、０．４ｍＭのイソプロピル－β－Ｄ－チオガラクトピラノシドを添加することによって発現を誘導した。Ｅ．ｃｏｌｉ細胞の成長を３０℃で４時間継続した。４℃、８０００×ｇで３０分間遠心分離することによってＥ．ｃｏｌｉ細胞を回収し、使用するまで－８０℃で保存した。組換えＢｉｒＡを発現する凍結細胞ペレットを、５ｍＭのイミダゾールを伴うＩＭＡＣ緩衝剤（５０ｍＭのNaH_2PO_4、ｐＨ８．０、３００ｍＭのＮａＣｌ）に再懸濁し、１ｍｇ／ｍｌのリゾチームと一緒に氷上で２０分間インキュベートし、超音波処理によって溶解させた。４℃、１６，０００×ｇで３０分間遠心分離することによって細胞デブリおよび不溶性材料を除去した。清澄化した上清をその後、ＡＫＴＡ ｐｕｒｅ ｃｈｒｏｍａｔｏｇｒａｐｈｙ ｓｙｓｔｅｍ（ＧＥ Ｈｅａｌｔｈｃａｒｅ）を使用してＨｉｓＴｒａｐ ＨＰ ５ｍＬカラムにローディングし、ＩＭＡＣ緩衝剤プラス２５ｍＭおよび５０ｍＭのイミダゾールで洗浄した後、５００ｍＭのイミダゾールを用いて溶出させた。ＢｉｒＡを含有する画分をプールし、２０ｍＭのＴｒｉｓ－ＨＣｌ、ｐＨ８．０に対して２５ｍＭのＮａＣｌを用いて透析し、ＨｉＴｒａｐ Ｑ ＨＰ ５ｍＬカラム（ＧＥ Ｈｅａｌｔｈｃａｒｅ）にローディングし、２５ｍＭのＮａＣｌから６００ｍＭのＮａＣｌまでの直線勾配を適用することによって溶出した。高度に純粋なＢｉｒＡを含有する画分をプールし、緩衝剤を保存緩衝剤（２０ｍＭのＴｒｉｓ－ＨＣｌ、ｐＨ８．０、１００ｍＭのＮａＣｌ、５％グリセロール）に交換し、およそ５～１０ｍｇ／ｍＬに濃縮し、分注し、－８０℃での保存のために液体窒素で急速冷凍した。ＢｉｒＡタンパク質濃度を、ＵＶ分光法により、算出された吸光係数ε＝４７，４４０$M^{-1}cm^{-1}$を使用してOD_{280nm}において決定した。
【０６６１】
ウエスタンブロッティングプロトコール
試料をＸＴ Ｓａｍｐｌｅ ＢｕｆｆｅｒおよびＸＴ Ｒｅｄｕｃｉｎｇ Ａｇｅｎｔ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）に添加し、９５℃で５分間加熱し、次いで、約１００，０００個の細胞に対応する体積を１０％Ｃｒｉｔｅｒｉｏｎ ＸＴ Ｂｉｓ－Ｔｒｉｓゲル（Ｂｉｏ－Ｒａｄ）にローディングし、ＰｏｗｅｒＰａｃ Ｂａｓｉｃ Ｐｏｗｅｒ Ｓｕｐｐｌｙ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）を使用し、ＸＴ ＭＥＳ Ｒｕｎｎｉｎｇ Ｂｕｆｆｅｒ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）を用いて２００Ｖで３５分間電気泳動した。ゲルを水で簡単にすすぎ、次いで、タンパク質を、Ｉｎｖｉｔｒｏｇｅｎ ｉＢｌｏｔ２ Ｇｅｌ Ｔｒａｎｓｆｅｒ Ｄｅｖｉｃｅ（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）で設定Ｐ３を使用してＩｎｖｉｔｒｏｇｅｎ ｉＢｌｏｔ Ｔｒａｎｓｆｅｒ Ｓｔａｃｋｓ（Ｔｈｅｒｍｏ Ｆｉｓｈｅｒ Ｓｃｉｅｎｔｉｆｉｃ）内のＰＶＤＦ膜に移した。Ｐｒｅｃｉｓｉｏｎ Ｐｌｕｓ Ｐｒｏｔｅｉｎ Ａｌｌ Ｂｌｕｅ Ｓｔａｎｄａｒｄ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）を使用して分子量をモニタリングした。次に、膜をＰｉｅｒｃｅ ＴＢＳ Ｔｗｅｅｎ２０（ＴＢＳＴ）緩衝剤（２５ｍＭのＴｒｉｓ、０．１５ｍＭのＮａＣｌ、０．０５％（ｖ／ｖ）Ｔｗｅｅｎ２０、ｐＨ７．５）で３×５分間洗浄し、ＴＢＳＴ－Ｍ（５％（ｗ／ｖ）無脂肪インスタントドライミルクを含有するＴＢＳＴ）中、室温で１時間ブロッキングし、次いで、ＴＢＳＴ－Ｂ（５％（ｗ／ｖ）Ｂｏｖｉｎｅ Ｓｅｒｕｍ Ａｌｂｕｍｉｎ（Ｓｉｇｍａ Ａｌｄｒｉｃｈ）を含有するＴＢＳＴ）ならびにウサギ抗ベータチューブリン抗体（カタログ番号ａｂ６０４６、Ａｂｃａｍ）およびウサギ抗ビオチンリガーゼエピトープタグ抗体（カタログ番号１００－４０１－Ｂ２１、Ｒｏｃｋｌａｎｄ Ｉｍｍｕｎｏｃｈｅｍｉｃａｌｓ

）の両方の１：５，０００希釈物中、４℃で終夜インキュベートした。次に、膜をＴＢＳＴで３×５分間洗浄し、ヤギ抗ウサギＩｇＧ（Ｈ＋Ｌ－西洋ワサビペルオキシダーゼとコンジュゲートした抗体（カタログ番号１７０－６５１５、Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）の１：１０，０００希釈物を含有するＴＢＳＴ－Ｍ中、室温で１時間インキュベートし、次いで、ＴＢＳＴを用いて室温で３×５分間洗浄した。最後に、膜をＰｉｅｒｃｅ ＥＣＬ Ｗｅｓｔｅｒｎ Ｂｌｏｔｔｉｎｇ Ｓｕｂｓｔｒａｔｅ（ＴｈｅｒｍｏＦｉｓｈｅｒ Ｓｃｉｅｎｔｉｆｉｃ、Ｒｏｃｋｆｏｒｄ、ＩＬ）に浸し、ＣｈｅｍｉＤｏｃ ＸＲＳ＋Ｉｍａｇｅｒ（Ｂｉｏ－Ｒａｄ）を使用して展開し、Ｉｍａｇｅ Ｌａｂ ｓｏｆｔｗａｒｅ（Ｂｉｏ－Ｒａｄ）を使用して可視化した。

【０６６２】

親和性タグ付けされたＨＬＡ－ペプチド複合体の単離

ＢＡＰタグが付されたＨＬＡ対立遺伝子を発現する細胞およびＢＡＰタグを伴わない内因性ＨＬＡ－ペプチド複合体のみを発現する陰性対照細胞株からの親和性タグ付けされたＨＬＡ－ペプチド複合体の単離を実施した。ＮｅｕｔｒＡｖｉｄｉｎビーズを付したアガロース樹脂を低温ＰＢＳ １ｍＬで３回洗浄した後、ＨＬＡ－ペプチドアフィニティー精製に使用した。ＢＡＰタグが付されたＨＬＡペプチドを発現する細胞５０×１０^{6}個を含有する凍結ペレットを氷上で２０分間にわたって解凍し、低温溶解緩衝剤［２０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１００ｍＭのＮａＣｌ、６ｍＭの$MgCl_2$、１．５％（ｖ／ｖ）Ｔｒｉｔｏｎ Ｘ－１００、６０ｍＭのオクチルグルコシド、０．２ｍＭの２－ヨードアセトアミド、１ｍＭのＥＤＴＡ、ｐＨ８、１ｍＭのＰＭＳＦ、１×完全ＥＤＴＡフリープロテアーゼ阻害剤カクテル（Ｒｏｃｈｅ、Ｂａｓｅｌ、Ｓｗｉｔｚｅｒｌａｎｄ）］１．２ｍＬ中に手動でピペッティングすることによって穏やかに溶解させた。溶解物を、≧２５０単位のベンゾナーゼヌクレアーゼ（Ｓｉｇｍａ－Ａｌｄｒｉｃｈ）と一緒に４℃で１５分間、回転させながらインキュベートしてＤＮＡ／ＲＮＡを分解し、４℃、１５，０００×ｇで２０分間遠心分離して細胞デブリおよび不溶性材料を除去した。清澄化した上清を新しいチューブに移し、０．５６μＭのビオチン、１ｍＭのＡＴＰ、および３μＭのＢｉｒＡを伴う１．５ｍＬチューブ中、室温で１０分間、回転させながらインキュベートすることにより、ＢＡＰタグが付されたＨＬＡペプチドをビオチン化した。上清を、２００μＬのＰｉｅｒｃｅ ｈｉｇｈ－ｃａｐａｃｉｔｙ ＮｅｕｔｒＡｖｉｄｉｎ ｂｅａｄｅｄ ａｇａｒｏｓｅ ｒｅｓｉｎ（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）スラリーに対応する体積で、４℃で３０分間、回転させながらインキュベートして、ビオチン化ＨＬＡ－ペプチド複合体を親和性富化した。最後に、ＨＬＡが結合した樹脂を低温洗浄緩衝剤（２０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１００ｍＭのＮａＣｌ、６０ｍＭのオクチルグルコシド、０．２ｍＭの２－ヨードアセトアミド、１ｍＭのＥＤＴＡ、ｐＨ８）１ｍＬで４回洗浄し、次いで、低温の１０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１ｍＬで４回洗浄した。洗浄間に、ＨＬＡが結合した樹脂を手動で穏やかに混合し、次いで、４℃、１，５００×ｇで１分間遠心分離することによってペレット化した。洗浄済みのＨＬＡが結合した樹脂を－８０℃で保存したかまたはすぐにＨＬＡ－ペプチドの溶出および脱塩に供した。

【０６６３】

抗体に基づくＨＬＡ－ペプチド複合体の単離

ＨＬＡクラスＩＩ ＤＲ－ペプチド複合体を健康なドナー末梢血単核細胞（ＰＢＭＣ）から単離した。ＧａｍｍａＢｉｎｄ Ｐｌｕｓ Ｓｅｐｈａｒｏｓｅ樹脂７５μＬに対応する体積を低温ＰＢＳ １ｍＬで３回洗浄し、抗体１０μｇと一緒に４℃で終夜、回転させながらインキュベートし、次いで、低温ＰＢＳ１ｍＬで３回洗浄した後、ＨＬＡ－ペプチド免疫沈降に使用した。細胞５０×１０^{6}個を含有する凍結ＰＢＭＣペレットを氷上で２０分間にわたって解凍し、低温溶解緩衝剤［２０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１００ｍＭのＮａＣｌ、６ｍＭのＭｇＣｌ２、１．５％（ｖ／ｖ）Ｔｒｉｔｏｎ Ｘ－１００、６０ｍＭのオクチルグルコシド、０．２ｍＭの２－ヨードアセトアミド、１ｍＭのＥＤＴＡ、ｐＨ８、１ｍＭのＰＭＳＦ、１×完全ＥＤＴＡフリープロテアーゼ阻害剤カクテル（Ｒｏｃｈｅ、Ｂａｓｅｌ、Ｓｗｉｔｚｅｒｌａｎｄ）］１．２ｍＬ中にピペッティングす

ることによって穏やかに溶解させた。溶解物を、＞２５０単位のベンゾナーゼヌクレアーゼ（Ｓｉｇｍａ－Ａｌｄｒｉｃｈ）と一緒に４℃で１５分間、回転させながらインキュベートしてＤＮＡ／ＲＮＡを分解し、４℃、１５，０００×ｇで２０分間遠心分離して細胞デブリおよび不溶性材料を除去した。次いで、上清を、ＧａｍｍａＢｉｎｄ Ｐｌｕｓ Ｓｅｐｈａｒｏｓｅ樹脂（ＧＥ Ｌｉｆｅ Ｓｃｉｅｎｃｅｓ）に結合させた抗ＨＬＡ ＤＲ抗体（ＴＡＬ １Ｂ５、製品番号ｓｃ－５３３１９；Ｓａｎｔａ Ｃｒｕｚ Ｂｉｏｔｅｃｈｎｏｌｏｇｙ、Ｄａｌｌａｓ、ＴＸ）と一緒に４℃で３時間、回転させながらインキュベートして、ＨＬＡ ＤＲ－ペプチド複合体を免疫沈降させた。最後に、ＨＬＡが結合した樹脂を低温洗浄緩衝剤（２０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１００ｍＭのＮａＣｌ、６０ｍＭのオクチルグルコシド、０．２ｍＭの２－ヨードアセトアミド、１ｍＭのＥＤＴＡ、ｐＨ８）１ｍＬで４回洗浄し、次いで、低温の１０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１ｍＬで４回洗浄した。洗浄間に、ＨＬＡが結合した樹脂を穏やかに混合し、次いで、４℃、１，５００×ｇで１分間遠心分離することによってペレット化した。洗浄済みのＨＬＡが結合した樹脂を－８０℃で保存したかまたはすぐにＨＬＡ－ペプチドの溶出および脱塩に供した。

【０６６４】

ＨＬＡ－ペプチドの溶出および脱塩

ＨＬＡ－ペプチドを親和性タグ付けされた内因性ＨＬＡ複合体から溶出させ、同時に、Ｓｅｐ－Ｐａｋ（Ｗａｔｅｒｓ、Ｍｉｌｆｏｒｄ、ＭＡ）固相抽出系を使用して脱塩した。簡単に述べると、Ｓｅｐ－Ｐａｋ Ｖａｃ １ｃｃ（５０ｍｇ）３７～５５μｍ粒子サイズｔＣ１８カートリッジを２４位置抽出マニホールド（Ｒｅｓｔｅｋ）に付着させ、ＭｅＯＨ、２００μＬ、その後、５０％（ｖ／ｖ）ＡＣＮ／１％（ｖ／ｖ）ＦＡ、１００μＬを用いて２回活性化し、その後、１％（ｖ／ｖ）ＦＡ、５００μＬで４回洗浄した。ＨＬＡ－ペプチドを親和性タグ付けされたＨＬＡペプチドから解離させ、ｔＣ１８固相へのペプチド結合を容易にするために、３％（ｖ／ｖ）ＡＣＮ／５％（ｖ／ｖ）ＦＡ、４００μＬをＨＬＡが結合したビーズが付されたアガロース樹脂を含有するチューブに添加した。ピペッティングすることによってスラリーを混合し、次いで、Ｓｅｐ－Ｐａｋカートリッジに移した。チューブおよびピペットチップを１％（ｖ／ｖ）ＦＡ（２×２００μＬ）ですすぎ、すすぎ液をカートリッジに移した。ローディング対照としてＰｉｅｒｃｅ Ｐｅｐｔｉｄｅ Ｒｅｔｅｎｔｉｏｎ Ｔｉｍｅ Ｃａｌｉｂｒａｔｉｏｎ（ＰＲＴＣ）混合物（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）１００ｆｍｏｌをカートリッジに添加した。ビーズが付されたアガロース樹脂を、１０％（ｖ／ｖ）ＡｃＯＨ、２００μＬと一緒に２回、５分間にわたってインキュベートして、ＨＬＡ－ペプチドを親和性タグ付けされたＨＬＡペプチドからさらに解離し、次いで、１％（ｖ／ｖ）ＦＡ、５００μＬで４回洗浄した。ＨＬＡ－ペプチドを、１５％（ｖ／ｖ）ＡＣＮ／１％（ｖ／ｖ）ＦＡ、２５０μＬ、その後、３０％（ｖ／ｖ）ＡＣＮ／１％（ｖ／ｖ）ＦＡ、２×２５０μＬを用いたステップ分画によってｔＣ１８から溶出させて新しい１．５ｍＬマイクロチューブ（Ｓａｒｓｔｅｄｔ）に入れた。活性化、試料ローディング、洗浄、および溶出に使用した溶液は重力によって流出させたが、残りの溶出液をカートリッジから取り出すために真空（≦－２．５ＰＳＩ）を使用した。ＨＬＡ－ペプチドを含有する溶出液を凍結させ、真空遠心分離によって乾燥させ、－８０℃で保存した後、第２の脱塩ワークフローに供した。

【０６６５】

以前に記載されている通りＥｍｐｏｒｅ Ｃ１８固相抽出ディスク（３Ｍ、Ｓｔ．Ｐａｕｌ、ＭＮ）の１６ゲージの穿孔を２つ使用して充填した社内で築いたＳｔａｇｅＴｉｐｓを用いてＨＬＡ－ペプチド試料の二次脱塩を実施した。ＳｔａｇｅＴｉｐｓを、ＭｅＯＨ、１００μＬを用い、その後、５０％（ｖ／ｖ）ＡＣＮ／０．１％（ｖ／ｖ）ＦＡ、５０μＬを用いて、２回活性化し、次いで、１％（ｖ／ｖ）ＦＡ、１００μＬで３回洗浄した。乾燥ＨＬＡ－ペプチドを、３％（ｖ／ｖ）ＡＣＮ／５％（ｖ／ｖ）２００μＬを添加し、次いで、ＳｔａｇｅＴｉｐｓにローディングすることによって可溶化した。チューブおよびピペットチップを１％（ｖ／ｖ）ＦＡ（２×１００μＬ）ですすぎ、すすぎ体積を

ＳｔａｇｅＴｉｐｓに移し、次いで、ＳｔａｇｅＴｉｐｓを１％（ｖ／ｖ）ＦＡ、１００μＬで５回洗浄した。ペプチドを、１５％（ｖ／ｖ）ＡＣＮ／０．１％（ｖ／ｖ）ＦＡ、２０μＬ、その後、３０％（ｖ／ｖ）ＡＣＮ／０．１％（ｖ／ｖ）ＦＡの２つの２０μＬのカットの段階的勾配を使用して溶出させた。試料ローディング、洗浄、および溶出は、卓上遠心分離機において１，５００～３，０００×ｇの最大スピードで実施した。溶出液を凍結させ、真空遠心分離によって乾燥させ、－８０℃で保存した。

【０６６６】

タンデム質量分析によるＨＬＡ－ペプチド配列決定

全てのｎａｎｏＬＣ－ＥＳＩ－ＭＳ／ＭＳ分析に同じ下記のＬＣ分離条件を使用した。試料を、ＰｉｃｏＦｒｉｔ（Ｎｅｗ Ｏｂｊｅｃｔｉｖｅ，Ｉｎｃ．、Ｗｏｂｕｒｎ、ＭＡ）にフィットさせたＰｒｏｘｅｏｎ Ｅａｓｙ ＮａｎｏＬＣ １２００（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ、Ｓａｎ Ｊｏｓｅ、ＣＡ）を使用してクロマトグラフィーにより分離した。１０μｍのエミッターを有する７５μｍ内径キャピラリーに、１．９μｍ粒子サイズ／２００Å孔径のＣ１８ Ｒｅｐｒｏｓｉｌビーズを用い、１０００ｐｓｉの圧力でＨｅを約３０～４０ｃｍまで充填し、分離の間、６０℃で加熱した。カラムを、１０×ベッド体積の緩衝剤Ａ（０．１％（ｖ／ｖ）ＦＡおよび３％（ｖ／ｖ）ＡＣＮ）を用いて平衡化し、試料を３％（ｖ／ｖ）ＡＣＮ／５％（ｖ／ｖ）ＦＡ、４μＬにローディングし、７～３０％の緩衝剤Ｂ（０．１％（ｖ／ｖ）ＦＡおよび８０％（ｖ／ｖ）ＡＣＮ）を８２分間、３０～９０％緩衝剤Ｂを６分間の直線勾配を用いてペプチドを溶出させ、次いで、９０％緩衝剤Ｂで１５分間保持してカラムを洗浄した。試料のサブセットを、６～４０％緩衝剤Ｂを８４分間、４０～６０％緩衝剤Ｂを９分間の直線勾配を用いて溶出させ、次いで、９０％緩衝剤Ｂを５分間および５０％緩衝剤Ｂで９分間保持してカラムを洗浄した。試料溶出のための直線勾配は毎分２００ｎＬの速度で流し、約１３秒のピーク幅中央値がもたらされた。

【０６６７】

データ依存性取得の間、溶出したペプチドを、Ｎａｎｏｓｐｒａｙ Ｆｌｅｘ Ｉｏｎ ｓｏｕｒｃｅ（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）を備えたＯｒｂｉｔｒａｐ Ｆｕｓｉｏｎ Ｌｕｍｏｓ ｍａｓｓ ｓｐｅｃｔｒｏｍｅｔｅｒ（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）に２．２～２．５ｋＶで導入した。フルスキャンＭＳを分解能６０，０００、３００ｍ／ｚから１，７００ｍ／ｚまで（ＡＧＣ標的４ｅ５、５０ｍｓ ｍａｘ ＩＴ）で取得した。各フルスキャンの後に、２秒間のサイクル時間、または、単離幅１．０ｍ／ｚ、衝突エネルギー３４（ＨＬＡクラスＩデータ）および３８（ＨＬＡクラスＩＩデータ）、ＡＧＣ標的５ｅ４、および最大充填時間２５０ｍｓ最大イオン時間を使用した分解能１５，０００でのデータ依存性ＭＳ２スキャンの上位１０を続けた。単離幅１．０ｍ／ｚを使用した理由は、ＨＬＡクラスＩＩペプチドはより長い傾向があり（＞４０アミノ酸のペプチドのサブセットを用いた中央値１６アミノ酸）、したがって、モノアイソトピックピークが常に同位体クラスターにおいて最も高いピークであるわけではなく、質量分析計取得ソフトウェアにより、指定のオフセットの非存在下で最も高い同位体ピークが単離ウインドウの中央に置かれる。したがって、１．０ｍ／ｚ単離ウインドウにより、同位体クラスターにおける最も高いピークではない場合であっても、クラスＩＩペプチドの電荷の状態は多くの場合＋２またはそれよりも高いので、モノアイソトピックピークの共単離が可能になる。選択された前駆体当たり約３ＰＳＭを可能にするための繰り返し回数１および排除持続時間５秒間で動的排除が可能になった。ＨＬＡ－ペプチド同定はＰＳＭ品質に依拠し、したがって、電荷の状態が異なる多数のＰＳＭによりペプチド同定の信頼度がさらに増大するので、前駆体当たり単一の電荷の状態に対する依存的スキャンが無効の間は同位元素を排除した。ＨＬＡクラスＩＩデータ収集のための電荷の状態のスクリーニングを、電荷の状態が１である（塩基性アンカー残基を有する対立遺伝子についてのみ）、＞７である、または割り当てられていない前駆イオンに対するＭＳ／ＭＳの誘発を防止するためのＰｅｐｔｉｄｅ Ｍｏｄｅを使用したモノアイソトピック前駆体選択（ＭＩＰＳ）と一緒に可能にした。ＨＬＡクラスＩに関してはデータ収集、電荷の状態が１である前駆イ

オン（質量範囲８００～１７００ｍ／ｚ）および２～４である前駆イオンを選択し、一方、電荷の状態が＞４である前駆イオンおよび割り当てられていない前駆イオンは排除した。

【０６６８】

高磁場非対称波形イオン移動度分光分析（ＦＡＩＭＳ）を使用したペプチドの検出を、以下のプロトコールを使用して評価した。Ａ３７５細胞により内因性にプロセシングされ、提示されたＨＬＡクラスＩおよびＨＬＡクラスＩＩペプチドを酸性逆相（ａＲＰ）および塩基性逆相（ｂＲＰ）オフライン分画の両方に供した後、ＦＡＩＭＳインターフェースを備えていないまたは備えたｏｒｂｉｔｒａｐ　ｆｕｓｉｏｎ　ｌｕｍｏｓ　ｔｒｉｂｉｄ質量分析計を使用してｎＬＣ－ＭＳ／ＭＳによる分析を行った。図４２Ａはワークフローを実証する。図４２Ｂは、少量のトリプシン試料ジャーカット細胞、ＨｅＬａ細胞を用いたＦＡＩＭＳのベンチマーキングを示す。ＨＬＡクラスＩ結合性ペプチドおよびＨＬＡクラスＩＩ結合性ペプチドの両方をＦＡＩＭＳを使用して解析した（それぞれ図４３Ａおよび図４３Ｂ；ならびに図４４Ａおよび図４４Ｂ）。それぞれの場合において、ペプチド検出のわずかな改善、特にｂＲＰ分画ペプチド。図４５および図４６は交差サイズを示す。

【０６６９】

ＬＣ－ＭＳ／ＭＳデータの解釈

この節は、例えば、図２９に関連する。質量スペクトルを、Ｓｐｅｃｔｒｕｍ　Ｍｉｌｌ　ｓｏｆｔｗａｒｅ　ｐａｃｋａｇｅ　ｖ６．０　ｐｒｅ－Ｒｅｌｅａｓｅ（Ａｇｉｌｅｎｔ　Ｔｅｃｈｎｏｌｏｇｉｅｓ、Ｓａｎｔａ　Ｃｌａｒａ、ＣＡ）を使用して解釈した。ＭＳ／ＭＳスペクトルの前駆体ＭＨ＋が６００～２０００（クラスＩ）／６００～４０００（クラスＩＩ）の範囲内に入らない、前駆体電荷が＞５（クラスＩ）／＞７（クラスＩＩ）である、または検出ピークが最低＜５である場合にはそのＭＳ／ＭＳスペクトルを検索から除外した。同じクロマトグラフィーピークで取得された同じ前駆体ｍ／ｚを有する同様のスペクトルのマージは無効にした。ＭＳ／ＭＳスペクトルを、２６４種の一般的な夾雑物と組み合わせた、ゲノムのｈｇ１９アノテーションおよびそのタンパク質コード転写物を伴う全てのＵＣＳＣ　Ｇｅｎｏｍｅ　Ｂｒｏｗｓｅｒ遺伝子を含有するデータベース（６３，６９１種のエントリー；１０，９１７，８６７種の独特の９ｍｅｒのペプチド）に対して検索した。データベース検索の前に、全てのＭＳ／ＭＳを、配列タグの長さが＞２である、例えば、最小で３つの質量がアミノ酸の鎖内質量から分離されるスペクトル品質フィルターを通過させなければならなかった。最小の骨格切断スコア（ＢＣＳ）を５に設定し、ＥＳＩ　ＱＥｘａｃｔｉｖｅ　ＨＬＡｖ２スコア化スキームを使用した。還元もアルキル化もされていないネイティブなＨＬＡ－ペプチド試料からのスペクトルの全てを、酵素特異性なし、システイニル化として固定されたシステインの修飾を使用し、以下の可変の修飾：酸化メチオニン（ｍ）、ピログルタミン酸（Ｎ末端ｑ）、カルバミドメチル化（ｃ）を用いて検索した。還元およびアルキル化されたＨＬＡ－ペプチド試料を、酵素特異性なし、カルバミドメチル化として固定されたシステインの修飾を使用し、以下の可変の修飾：酸化メチオニン（ｍ）、ピログルタミン酸（Ｎ末端ｑ）、システイニル化（ｃ）を用いて検索した。ネイティブなＨＬＡ－ペプチドデータセットならびに還元およびアルキル化されたＨＬＡ－ペプチドデータセットの両方に対して前駆体質量許容誤差±１０ｐｐｍ、生成物質量許容誤差±１０ｐｐｍ、および最小のマッチしたピーク強度３０％を使用した。個々のスペクトルについてのペプチドスペクトルマッチ（ＰＳＭ）を、Ｓｐｅｃｔｒｕｍ　Ｍｉｌｌ自動検証モジュールを使用して標的－デコイのＰＳＭ順位でのＦＤＲ推定に基づいた利用に確信的に割り当てられたものと自動的に指定して、スコア化閾値基準を設定した。ＨＬＡ対立遺伝子に対して最小の配列長７、自動可変範囲前駆体質量フィルタリング、ならびに全てのＬＣ－ＭＳ／ＭＳ実行にわたって最適化されたスコアおよびデルタ順位１－順位２スコア閾値を使用した自動閾値戦略により、各前駆体電荷の状態についてＰＳＭ　ＦＤＲ推定値＜１．０％がもたらされた。

【０６７０】

ＰＳＭ　ＦＤＲ推定値＜１．０％を通過した同定されたペプチドを、参照データベース内の２６４種の一般的な夾雑物タンパク質に割り当てられた全てのペプチドを除去するこ

とによって、および陰性対照ＭＡＰＴＡＣ（商標）親和性プルダウンにおいて同定されたペプチドを除去することによって、夾雑物に対してさらにフィルタリングした。さらに、参照データベースのｉｎ　ｓｉｌｉｃｏトリプシン消化物にマッピングされたペプチドはｕＰＬＣカラムの試料キャリーオーバーからのトリプシン夾雑物として除外することができないので、これらのペプチド同定の全てを除去した。

【０６７１】

ＭＡＰＴＡＣ（商標）プロトコールを使用したＨＬＡ－ＤＲ、ＨＬＡ－ＤＱ、ＨＬＡ－ＤＰヘテロ二量体の単一対立遺伝子割り当て

ＬＣ－ＭＳ／ＭＳにより同定されたペプチドへの単一対立遺伝子ＨＬＡの割り当ては、２つの手法に従った。ＨＬＡ－ＤＲＡ１の対立遺伝子バリエーションは限られており、ペプチド結合に影響を及ぼすとは考えられないので、ＤＲ実験（ＤＲＢ１、３、４および５のプロファイリング）からの全てのデータを、単一対立遺伝子性である、つまり、ペプチドが捕捉用アルファ鎖と対形成した捕捉用ベータ鎖を含むＨＬＡクラスＩＩヘテロ二量体に結合する可能性が最も高いとみなした。しかし、一部のペプチドが、別個の内因性に発現されたアルファ鎖と対形成したノックインベータ鎖を含むＨＬＡ　ＩＩヘテロ二量体に結合し得る可能性が残る。

【０６７２】

逆に、ＨＬＡ－ＤＰおよびＨＬＡ－ＤＱ遺伝子座に関しては、アルファ鎖は、重要な対立遺伝子バリアントを示し、したがって、ノックインおよび内因性アルファ鎖対立遺伝子の両方が存在することにより、多数のヘテロ二量体の潜在性が生じる。例えば、別個のＨＬＡ－ＤＰおよびＨＬＡ－ＤＱヘテロ二量体をコードするノックインアルファ鎖およびベータ鎖は、それぞれ内因性に発現されたアルファ鎖およびベータ鎖と対形成し得、それにより、各ＨＬＡ－ＤＰおよびＨＬＡ－ＤＱ　ＭＡＰＴＡＣ（商標）構築物について４つの独特のヘテロ二量体が作り上げられる。したがって、精製されたＭＡＰＴＡＣ（商標）ペプチド集団の中での結合特異性は単一対立遺伝子性ではない。この内因性対形成の問題を軽減するために、アルファ鎖およびＭＡＰＴＡＣ（商標）ベータ鎖を内因性に含むＨＬＡヘテロ二量体に結合する可能性が高いペプチドの集団を同定することを可能にした、アルファ鎖を欠く構築物を使用した（ｓａｎｓ－アルファノックイン）。これらのペプチドを対応するアルファ＋ベータ鎖ＭＡＰＴＡＣ（商標）実験からコンピュータにより差し引いて、単一対立遺伝子ＭＡＰＴＡＣ（商標）アルファ＋ベータ組合せに対する特異的なペプチドの集団を見積もった。

【０６７３】

各ペプチドを、ＵＣＳＣ　ｈｇ１９遺伝子アノテーション内の１つまたは複数のタンパク質をコードする転写物に割り当てた。多くのペプチド同定は他と重複し、したがって、大部分が重複する情報を構成するので、図１１Ｃに示されている通り、ペプチドを、各々が約１つの独特の結合事象に対応することが意図されている「入れ子セット」に群分けした。例えば、ペプチドＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤＶ（配列番号１６）、ＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１７）、およびＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤＶ（配列番号１８）は全て、保存された配列ＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１９）を含有し、おそらく全てが同じレジスター内のＭＨＣに結合する。所与のデータセットのペプチドを入れ子状にするために、各ノードが独特のペプチドに対応するグラフを築き、少なくとも１つの９ｍｅｒを共有し、少なくとも１つの共通の転写物にマッピング可能なペプチドの任意の対間にエッジを創出した。Ｒパッケージｉｇｒａｐｈ内のクラスターコマンドを使用して、接続したノードのクラスターを同定し、各クラスターを入れ子セットと定義した。この手順により、エッジ基準（共通の９ｍｅｒが１つ以上、共通の転写物が１つ以上）を満たす任意の２つのペプチドが同じ入れ子セット内に入ることが保証される。

【０６７４】

以前に公開されたＭＳデータの解析

以下の節は、少なくとも図１２Ａ～１２Ｆ、図３５、図３６Ａ～３６Ｂ、図３８Ｄ、図

３９Ａ～３９Ｃ、図４０Ａ～４０Ｂに関する。．ｒａｗファイルで提供された公開されたＬＣ－ＭＳ／ＭＳデータセットを、Ｓｐｅｃｔｒｕｍ Ｍｉｌｌ ｓｏｆｔｗａｒｅ ｐａｃｋａｇｅ ｖ６．０ ｐｒｅ－Ｒｅｌｅａｓｅ（Ａｇｉｌｅｎｔ Ｔｅｃｈｎｏｌｏｇｉｅｓ、Ｓａｎｔａ Ｃｌａｒａ、ＣＡ）を使用して再処理した。ＨＣＤ断片化ならびにオービトラップ（高分解能）へのＭＳおよびＭＳ／ＭＳデータ収集を利用したＴｈｅｒｍｏ Ｏｒｂｉｔｒａｐ機器（例えば、Ｖｅｌｏｓ、ＱＥｘａｃｔｉｖｅ、Ｆｕｓｉｏｎ、Ｌｕｍｏｓ）に収集されたデータセットを、上記の節「ＬＣ－ＭＳ／ＭＳデータの解釈」に記載されているパラメータを使用して解析した。ＣＩＤ断片化を利用したＭＳおよびＭＳ／ＭＳ高分解能データセットに対して、ＥＳＩオービトラップスコア化スキームを用い、上記と同じパラメータを使用した。オービトラップへのＭＳデータ収集およびイオントラップへのＭＳ／ＭＳデータ収集を用いたデータセットに対しても同様に上記と同じ以下のパラメータを使用し、以下の偏差を用いた。ＨＣＤデータに対してはＥＳＩ ＱＥｘａｃｔｉｖｅ ＨＬＡｖ２スコア化スキームを使用し、一方ＣＩＤデータに対してはＥＳＩオービトラップスコア化スキームを使用した。前駆体質量許容誤差±１０ｐｐｍ、生成物質量許容誤差±０．５Ｄａを使用した。高分解能ＭＳ／ＭＳデータセットおよび低分解能ＭＳ／ＭＳデータセットの両方に対して、個々のスペクトルについてのペプチドスペクトルマッチ（ＰＳＭ）を、Ｓｐｅｃｔｒｕｍ Ｍｉｌｌ自動検証モジュールを使用して標的－デコイのＰＳＭ順位でのＦＤＲ推定に基づいた利用に確信的に割り当てられたものと自動的に指定して、スコア化閾値基準を設定した。ＨＬＡ対立遺伝子に対して最小の配列長７、自動可変範囲前駆体質量フィルタリング、ならびに全てのＬＣ－ＭＳ／ＭＳ実行にわたって最適化されたスコアおよびデルタ順位１－順位２スコア閾値を使用した自動閾値戦略により、各前駆体電荷の状態についてＰＳＭ ＦＤＲ推定値＜１．０％がもたらされた。

【０６７５】

図１１Ｄに示されている通り、ヒトプロテオームにおけるアミノ酸頻度を、ＵＣＳＣ ｈｇ１９アノテーション内のタンパク質をコードする遺伝子全ての配列に基づいて算出した（多数の転写物アイソフォームによって表される遺伝子に対して１つの転写物をランダムに選択する）。ＩＥＤＢ頻度を、≦５０ｎＭの親和性が少なくとも１回観察されたペプチドの独特のセットを同定することによって決定した（Ｃ末端に六価ポリヒスチジンを有する一部のペプチドは除外する）。ＭＡＰＴＡＣ（商標）頻度をまず、入れ子セット当たりただ１つのペプチド（最長のもの）を使用し、５種のＤＲＢ１対立遺伝子（ＤＲＢ１＊０１：０１、ＤＲＢ１＊０３：０１、ＤＲＢ１＊０９：０１、およびＤＲＢ１＊１１：０１）にわたる標準の順相プロトコールに関して検討した。さらに、いくつかの対立遺伝子にわたってＭＡＰＴＡＣ（商標）頻度を塩基性逆相プロトコールについて別々に算出した。外部データセットからのＭＳデータを、潜在的対立遺伝子起源を考慮せずに、同様に入れ子セット当たり最長ペプチドを使用して解析した。

【０６７６】

クラスＩ（ＨＬＡクラスＩ結合性ペプチド）配列ロゴを築くこと

各クラスＩ対立遺伝子（図１４Ａに示されている）について、対応するペプチドの最初の５つの位置（ロゴ位置１～５にマッピングされる）および最後の４つの位置（ロゴ位置６～９にマッピングされる）におけるアミノ酸頻度をプロファイリングすることによって長さ９の配列ロゴを創出した。このように、ペプチドは、長さにかかわらず配列ロゴに寄与した。クラスＩＩロゴと同様に、文字の高さは各位置における各アミノ酸の頻度に比例し、また、エントロピーが低い位置に色分けを使用する。

【０６７７】

ＭＳにより観察されたペプチドの予測される親和性

この節は、少なくとも図１４Ｃに関連する。各クラスＩＩ対立遺伝子について、１４から１７までの独特のペプチド長を全て同定し、結合潜在性についてＮｅｔＭＨＣＩＩｐａｎを使用してスコア化した。比較のために、ランダムな長さを釣り合わせたペプチドをヒトプロテオームからサンプリングした。密度分布（図１４Ｃに示されている）を対数変換された値に基づいて決定した。一部の対立遺伝子は、ＮｅｔＭＨＣＩＩｐａｎにより予測

が支持されなかったので、解析から除外された。
【０６７８】
ＭＳにより観察されたペプチドについての測定された親和性
ペプチドを、予測されるＮｅｔＭＨＣＩＩｐａｎ結合親和性が不十分なものである場合（ＤＲＢ１＊０１：０１については＞１００ｎＭまたはＤＲＢ１＊０９：０１およびＤＲＢ１＊１１：０１については＞５００ｎＭ）または以前に公開された生化学的ＭＨＣ－ペプチド親和性アッセイにおいて試験されるヒューリスティックに定義されたアンカーが≦２である場合、親和性測定値について選択した。
【０６７９】
プロテオーム区分の訓練、チューニング、および試験の確立
この節は、少なくとも図１５Ａに関連する（図３１Ａも参照されたい）。各ノードがタンパク質をコードする転写物を表し、少なくとも５つの独特の９ｍｅｒのアミノ配列内容を共有する転写物の全ての対の間にエッジが存在する（ＵＣＳＣ　ｈｇ１９遺伝子アノテーション）グラフを創出した。Ｒパッケージｉｇｒａｐｈ内のクラスターコマンド（ｃｒａｎ．ｒ－ｐｒｏｊｅｃｔ．ｏｒｇ／ｗｅｂ／ｐａｃｋａｇｅｓ／ｉｇｒａｐｈ／ｃｉｔａｔｉｏｎ．ｈｔｍｌ）を使用して、接続したノードのクラスターを同定し、各クラスターを「転写物群」と定義した。このように、２つの転写物がエッジを共有する場合（共有される９ｍｅｒが５つ以上）、同じ転写物群に入ることが保証された。転写物群をランダムにサンプリングし、７５％を訓練区分に入れ、１２．５％をチューニング区分に入れ、１２．５％を試験区分に入れた。全ての解析において、ＭＳにより観察されたペプチド（および観察されていないデコイペプチド）をそれらの供給源転写物の区分に従って区分（訓練、チューニング、または試験）に入れた。供給源転写物が多数の異なる区分にマッピングされる非常に稀な場合には、割り当て優先順序を訓練し（優先性が最も高い）、チューニングし、次いで、試験した（優先性が最も低い）。プロテオームの同じ区分化を区分に基づく分析の全てに使用した。プロテオームの区分化のグラフに基づく手法を使用して、訓練および評価中に同様のペプチド配列が現れ、それにより予測性能が人為的に膨張し得る可能性を最小限にした。
【０６８０】
ＣＮＮに基づくクラスＩＩ結合予測器のアーキテクチャおよび訓練
少なくとも図１５Ａとの関連で、アミノ酸は「ワンホット」符号化によって表すことができるが、他方ではＰＭＢＥＣ行列およびＢＬＯＳＵＭ行列を使用してアミノ酸を符号化する選択をし、その場合、同様のアミノ酸は同様の特色プロファイルを有する。我々のペプチド特色付けの目的のために、解かれたタンパク質構造においてアミノ酸近接性に基づいた独特の行列を生成した。この手法の概念は、アミノ酸の典型的な近隣アミノ酸に化学的特性が反映されるはずであるというものである。約１００，０００種のＤＳＳＰタンパク質構造の各アミノ酸について（ｃｄｎ．ｒｃｓｂ．ｏｒｇ／ｅｔｌ／ｋａｂｓｃｈＳａｎｄｅｒ／ｓｓ．ｔｘｔ．ｇｚ，）、３Ｄ空間では最も近いが、一次配列では少なくとも１０アミノ酸離れている残基を決定した。このデータを使用して、アラニンの最も近くの近隣アミノ酸がアラニンである回数、アラニンの最も近くの近隣アミノ酸がシステインなどである回数を決定して、近接計数の２０×２０行列を創出した。行列の各要素をその対応する列および行の合計の積で割り、行列全体を対数変換した。最後に、行列全体の平均値を各要素から引いた。３つの追加的な物理的特色である疎水性、電荷、およびサイズを追加的な列として加え、したがって、各アミノ酸を２３の入力特色によって表した。
【０６８１】
ＭＡＰＴＡＣ（商標）により観察されたペプチドに関する予測性能のベンチマーキング、図２１Ａ～Ｂに関連する
所与の対立遺伝子についての予測性能を評価するために、観察することができたが（プロテオームに存在することが理由で）、ＭＳデータでは観察されなかったペプチドのセットを定義することが必要であった。これらの陰性例を「天然のデコイ」と称した（上記の「スクランブルデコイ」とは対照的に）。指導原理の通り、以下を決定した：

１．天然のデコイの長さ分布はＭＳにより観察されたヒットの長さ分布とマッチすべきである。
２．天然のデコイは、他の天然のデコイと重複する配列を含有すべきではない。
３．天然のデコイはヒットと重複するべきではない。
４．天然のデコイは、少なくとも１つのヒットから生じた遺伝子に由来すべきである。

【０６８２】

以下の疑似コードは、これらの原理を満たす評価を創出したプロセスを表す：

【表１９】

ヒットの空リスト2つ、$H_{minimal}$および$H_{exhaustive}$を初期化する
MSにより観察されたペプチドの各入れ子セットSについて:
S内に訓練またはチューニング区分内の転写物にマッピングすることができるペプチドがない場合:
S内の最も短いペプチドを$H_{minimal}$に加える
S内の全てのペプチドを$H_{exhaustive}$に加える
デコイペプチドの空リスト、Dを初期化する
試験区分内の各タンパク質をコードする転写物(最長が最初、最短が最後)について:
$H_{exhaustive}$内に転写物にマッピングされるペプチドがない場合:
飛ばして次の転写物に進む
転写物のタンパク質配列に重複ペプチドPのセットをかぶせる。ペプチド長は$H_{minimal}$の長さ分布からランダムにサンプリングする。重複は8アミノ酸である(P内の最後のペプチドが一般にはタンパク質の終わりからぶら下がる)
P内の最後のペプチドがまだぶら下がっている間に:
P内の最長ペプチドの長さから1アミノ酸を差し引く
P内の各ペプチドについて:
$H_{exhaustive}$内のペプチドと9merを共有せず、D内のペプチドのいずれにも観察されない9merを含有する場合:
当該ペプチドをDに加える
他の場合には:
当該ペプチドをはねる
$H_{minimal}$およびDは評価データセットを構成する

【０６８３】

このセットに対する性能を評価するために、ｎ個のヒットペプチド全てを、１９ｎ個のデコイのセット（デコイの完全なセットから、元に戻さずにランダムにサンプリングしたもの）と一緒に予測器（ｎｅｏｎｍｈｃ２またはＮｅｔＭＨＣＩＩｐａｎ）によって評価し、スコア化した。組み合わせたセット内のペプチドの上位５％を陽性コールと標識し、陽性適中率（ＰＰＶ）を、ヒットであった陽性コールの分率として算出した。陽性の数は

ヒットの数と等しくなるように制約されるので、この評価シナリオでは再現率がＰＰＶと等しいことに留意されたい。対立遺伝子にわたって一貫した１：１９の比を適用することは、そうでなければ各対立遺伝子について観察されるヒットの数の影響を高度に受ける性能値の安定化に役立つ。ヒットの数が対立遺伝子の内在性特性よりも実験条件および反復実験数に大きく関連することが仮定されたので、これは妥当であると思われた。１：１９の比は、ダウンサンプリングがインプリメントされなかった場合に使用されるものとさほど離れていない。

【０６８４】

ＩＥＤＢ親和性測定値に関する予測性能のベンチマーキング

図１５Ｂに関する通り、ＮｅｔＭＨＣＩＩｐａｎをＩＥＤＢ親和性データに対して訓練したので、そのサンプル外性能をわずかに旧式のバージョンおよびＩＥＤＢ測定値を使用して評価した。予測された親和性と測定された親和性の対応をケンドールのタウ係数によって決定した。同じ統計値を使用して、同じ測定値のセットに対するｎｅｏｎｍｈｃ２の性能を評価した。評価は、対立遺伝子によって、および公開群（ＳｅｔｔｅまたはＢｕｕｓ）によって別々に行った。

【０６８５】

天然のＣＤ４＋　Ｔ細胞応答の予測性能のベンチマーキング

ＩＥＤＢにおいて実証されたＣＤ４＋　Ｔ細胞応答の大部分は未知のまたはコンピュータにより帰属されたクラスＩＩ対立遺伝子制限を有するので、クラスＩＩ四量体によって実験的に確認された記録のサブセットに焦点を当てた。ほぼ全てのそのような記録はＷｉｌｌｉａｍ　Ｋｗｏｋ　Ｌａｂｏｒａｔｏｒｙ（Ｂｅｎａｒｏｙａ　Ｒｅｓｅａｒｃｈ　Ｉｎｓｔｉｔｕｔｅ、Ｓｅａｔｔｌｅ、ＷＡ）によって寄託されており、免疫反応性個体の血液を使用して、多様な病原体およびアレルゲンの四量体によりガイドされるエピトープマッピング（ＴＧＥＭ）を実施している。陰性ペプチドは一部の試験に関してはポストされているが、他の試験に関してはポストされていないので、供給源である刊行物を精査して、陽性および陰性ペプチド反応性の完全なセットを再構築した。２０ｍｅｒのペプチド全てをｎｅｏｎｍｈｃ２によって、およびＮｅｔＭＨＣＩＩｐａｎによってスコア化した。対立遺伝子にわたる陽性適中率（ＰＰＶ）を同等のやり方で対立遺伝子にわたって算出するために、各対立遺伝子についての陰性例を陽性対陰性の比が１：１９になるまでランダムにダウンサンプリングした。ＰＰＶを、ペプチドの上位のスコア化された５％にわたって実験的に確認された陽性の分率として算出した。性能を受信者操作者曲線によっても評価した。

【０６８６】

ＭＨＣ　ＩＩペプチドの逆畳み込みの性能の評価

ＧｉｂｂｓＣｌｕｓｔｅｒ（ｖ２．０）ツールの、複対立遺伝子ＭＨＣクラスＩＩペプチドデータを対立遺伝子起源によってクラスター化する能力を評価するために、ＤＲ遺伝子型が分かっている対象に対する公開されたＤＲ特異的実験の多様なセットからのペプチド（表２）をまずキュレートした。一部の場合では、元の刊行物ではＨＬＡ－ＤＲＢ１タイピングは提示されているが、ＨＬＡ－ＤＲＢ３／４／５についてのタイピングは省略されている。これらの場合に対処するために、ＩＭＧＴにより提供されたＤＲ１：ＤＲ３／４／５連結を仮定し、それが４桁タイピングを分解するのに不十分な場合は、集団「ＵＳＡＳａｎＦｒａｎｃｉｓｃｏＣａｕｃａｓｉａｎ」において観察された連結（ａｌｌｅｌｅｆｒｅｑｕｅｎｃｉｅｓ．ｎｅｔ，、集団ＩＤ　３０９８：表２を使用した。

【表２】

表２

DRB1	DRB3/4/5	ハプロタイプ頻度(%)
DRB1*03:01	DRB3*01:01	23.2
DRB1*04:01	DRB4*01:01	9
DRB1*04:02	DRB4*01:01	2.2
DRB1*04:04	DRB4*01:01	3.2
DRB1*07:01	DRB4*01:01	13.2
DRB1*09:01	DRB4*01:01	2.8
DRB1*11:01	DRB3*02:02	7.2
DRB1*11:04	DRB3*02:02	2.8
DRB1*13:01	DRB3*02:02	2.6
DRB1*13:02	DRB3*03:01	2.6
DRB1*14:01	DRB3*02:02	2.2
DRB1*15:01	DRB5*01:01	28.6
DRB4*01:01 は DRB4*01:03 と同一であることが示されていることに留意されたい (ebi.ac.uk/cgi-bin/ipd/imgt/hla/get_allele.cgi?DRB4*0101)		

【０６８７】

各（帰属）遺伝子型に存在する各ＤＲＢ１／３／４／５対立遺伝子について、我々の単一対立遺伝子ＭＡＰＴＡＣ（商標）データからの２０個のペプチドをスパイクインした。次いで、これらの強化されたデータセットをＧｉｂｂｓＣｌｕｓｔｅｒ－ｖ２．０にサブミットした。

【０６８８】

ＭＨＣ　ＩＩペプチドの観察された切断部位の特徴付け

免疫精製を使用してヒト組織をプロファイリングしたいくつかの試験にわたるペプチド同定をマージすることによる、天然にプロセシングされ、提示されるＭＨＣ　ＩＩペプチドの大きなデータセットが本明細書に開示される（表２）。多くのペプチドが同じＮ末端（例えば、ＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤＶ（配列番号１６）およびＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１７））または同じＣ末端（例えば、ＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１７）およびＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１９））を共有するので、Ｎ末端について１つ、およびＣ末端について１つの、２セットの重複していないカット部分をキュレートした。次いで、同等数の独特の観察されていないＮ末端およびＣ末端カット部分を少なくとも１つのＭＨＣ　ＩＩペプチドを産生した遺伝子のセットからランダムにサンプリングした。これらの４つのデータセットをＮ末端ヒット、Ｃ末端ヒット、Ｎ末端デコイ、およびＣ末端デコイと称した。さらに、図４１に示されている、ペプチドの上流、ペプチド内部、およびペプチドの下流の位置を指すための命名システムを使用した。

【０６８９】

Ｎ末端ヒットについてＵ１０位からＮ３位までの各アミノ酸の頻度を決定し、これらの頻度をＮ末端デコイについて観察されたものと比較した。所与のアミノ酸が所与の位置にある率にヒットとデコイで有意差があるかどうかを決定するために、２×２表（例えば、Ｕ１がリシンであるヒットの計数、Ｕ１がリシンであるデコイの計数；Ｕ１がリシンでないヒットの計数、およびＵ１がリシンでないデコイの計数）を創出し、カイ二乗検定によってスコア化した。Ｃ末端ヒットおよびデコイのＣ３位からＤ１０位までのアミノ酸頻度を解析するために類似の手法を使用した。

【０６９０】

第２の解析では、切断事象の直前の残基と切断事象の後の残基の統計学的結び付きを検討した。まず、Ｕ１：Ｎ１対（Ａ：Ａ、Ａ：Ｃ、Ａ：Ｄ、…、Ｙ：Ｖ、Ｙ：Ｗ、Ｙ：Ｙ）の計数はＮ末端ヒットとＮ末端デコイを比較し、各対についての富化／枯渇の有意性を、

２×２分割表のカイ二乗検定によって決定した（例えば、Ｐ：Ｋを有するヒットの計数、Ｐ：Ｋを有するデコイの計数；Ｐ：Ｋを有さないヒットの計数、Ｐ：Ｋを有さないデコイの計数）。Ｃ末端ヒットおよびデコイのＣ１：Ｄ１対の頻度を解析するために類似の手法を使用した。

【０６９１】

種々のクラスＩＩ切断予測器の性能のベンチマーキング

健康なドナー由来の末梢血をＤＲ結合性ペプチドについてプロファイリングした。これらの試料を使用して、切断関連変数／予測器の、提示されるクラスＩＩエピトープの同定を増強する能力をベンチマークした。

【０６９２】

結合潜在性および切断潜在性の両方を使用してペプチドの提示を予測する統合予測器を築くために、まず、データセットを、図４に関して記載されているものと同じ手法を使用するが、「試験」区分ではなく「チューニング」区分を使用して構築した。手短に言えば、これは、ヒットとデコイを１：２０の比で使用することを意味し、ここで、デコイは、ヒットと長さを釣り合わせたものであり、少なくとも１つのヒットが生成された遺伝子のセットからランダムにサンプリングされたものである。ｎｅｏｎｍｈｃ２を使用して結合潜在性を算出し、これらの試料は複対立遺伝子であったので、各ペプチド候補ペプチドについての結合スコアを、各ドナーの遺伝子型によって示される１～４種のＤＲ対立遺伝子の平均になるように取得した。この当てはめプロセスにより、前方予測において結合および切断変数に乗せられる相対的な重みが決定された。

【０６９３】

前方予測の性能を決定するために、ヒットおよびデコイ（１：１９の比）の評価を、ちょうど記載したものと同じプロトコールを使用して「試験」区分から得た。ＰＰＶを図４についてと同じ様式で算出した。いくつかの異なる切断予測器を表３に示されている通り評価した。

【表３－１】

表3

既知の切断部位を有するPSM
未知の切断部位を有するPSM 候補ペプチドをスコア化するために、9merのコアをまず帰属させた。次に、最大PSMスコアを 3つの領域:コアの10アミノ酸上流 (真のN末端切断部位の位置にかかわらず)、コア配列、およびコアの後ろの10アミノ酸にわたって算出した。ロジスティック回帰をチューニング区分 (上記を参照されたい)に対して*neonmhc2*結合スコアならびに3つの領域特異的PSMスコアを使用して訓練した。
新規のニューラルネットワーク、未知の切断部位
カテプシンモチーフ、未知の切断部位
構造的特色 SCRATCH suite (Cheng J, Randall AZ, Sweredoski MJ & Baldi P (2005) SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res. 33: W72-6 Available at: dx.doi.org/10.1093/nar/gki396)内で、ツールACCpro20を使用して相対的な溶媒露出度 (RSA)を予測し、ツールSSpro8を使用して二次構造の特色を予測した(H:アルファヘリックス G: 310-ヘリックス I: pi-ヘリックス E: 伸長鎖 B: ベータブリッジ T: ターン S: ベンド C: その他)。配列障害の残基毎のスコア (d2p2.pro/download)をプロテオーム全体にわたって同様に決定し、当該位置を不規則と標識した予測エンジンの数に従って0-5の尺度でスコア化した (使用したサーバー: anchor、espritz-d、espritz-n、espritz-x、iupred-l、およびiupred-s)。最後に、Uniprot (uniprot_sprot.dat)を使用してトポロジカルドメイン(膜貫通、シグナル、細胞外、内腔、およびジスルフィド)を残基レベルで決定した。各候補ペプチドについて、平均RSAおよび平均障害スコアを、予測される9merのコアを中心にした21merのウインドウにわたって算出した。同じ21残基のウインドウにわたって、二次構造特色およびトポロジカル特色についてパーセント組成を算出した。4つのロジスティック回帰モデルを、以下のインプット特色を使用して「チューニング」区分に対して訓練した: 1. *Neonmhc2*結合スコアおよび溶媒露出度(RSA) 2. *Neonmhc2*結合スコアおよび障害スコア 3. *Neonmhc2*結合スコアおよび二次構造を表す8つのSSpro8由来の特色 4. *Neonmhc2*結合スコアおよびトポロジーを表す5つのUniprot特色
前に観察されたペプチドとの重複 パブリックドメインから多くの複対立遺伝子クラスII実験にわたってMSに基づくペプチド同定をプールした。新しい候補ペプチドが前に観察されたペプチドのうちの1つと重複するかどうかを表す新しい特色を創出した。具体的には、この特色を、前に観察されたMHC IIリガンドの大きなセット内の任意のペプチドと少なくとも1つの9merを共有する場合には1に設定し、他の場合には、特色を0に設定した。ロジスティック回帰を「チューニング」区分に対して*neonmhc2*結合スコアおよび重複特色を使用して訓練した。

【表３－２】

【０６９４】
ＭＨＣクラスＩＩ提示と発現の関係
ヒト卵巣組織のＨＬＡ－ＤＲリガンドームがプロファイリングされた以前に公開されたＭＳ実験にわたってペプチドをプールした。ＲＮＡ－Ｓｅｑデータが入手可能な各試料について、生のｆａｓｔｑをＳＲＡからダウンロードし、ｂｏｗｔｉｅ２を使用してＵＣＳＣ　ｈｇ１９トランスクリプトームにアラインメントした。ＲＳＥＭによって算出される百万当たりの転写物数（ＴＰＭ）を使用して転写レベルでの遺伝子数量化を実施した。発現推定値を、遺伝子レベルを合計すること、非コード遺伝子を外すこと、および総ＴＰＭが合計１０００００００になるように再正規化すること（ｎｃＲＮＡ存在量のライブラリー間の変動の原因となるタンパク質をコードする遺伝子にわたって再正規化すること）によってさらに処理した。
【０６９５】
各組織試料中の各遺伝子に関して、試料中の発現レベルおよび試料中で少なくとも１つのペプチドを産生するかどうかを検討した。全てのＭＳ実験にわたって、これらの知見を発現レベルおよびペプチド生成の状態に従ってビンに入れた（図１８Ａを参照されたい）。
【０６９６】
過剰に表された遺伝子および過少に表された遺伝子の同定
ＭＨＣ　ＩＩリガンドームにおいて過剰に表された遺伝子および過少に表された遺伝子を同定するために、卵巣組織、結腸直腸組織、ならびに皮膚黒色腫、肺がんおよび頭頸部がんがプロファイリングされた５つの以前の試験からのデータをコンパイルした。各遺伝子について、我々のベースライン仮定は、ペプチドがその長さにその発現レベルを掛けた割合でもたらされるはずであるというものであった。各遺伝子の長さを決定するために、全ての転写物アイソフォームにわたる独特の９ｍｅｒを数え上げた。転写物アイソフォームにわたって合計することによって遺伝子レベル発現を得た。観察された、各遺伝子にマッピングされるペプチドの数を入れ子セットレベルで決定した（例えば、ペプチド　ＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤＶ（配列番号１６）、ＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１７）、およびＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤＶ（配列番号１８）を単一の観察として計数した）。
【０６９７】
卵巣試験からの多くの試料には対応するＲＮＡ－Ｓｅｑデータがあったが、一部にはなかった。これらの場合には、ＲＮＡ－Ｓｅｑデータが入手可能な試料にわたって平均することによって発現を推定した。結腸直腸試験および黒色腫試験に関しては、いずれの試料についても対応するＲＮＡ－Ｓｅｑが存在せず、したがって、ＧＴＥｘおよびＴＣＧＡからのデータを使用して代理試料にわたって平均を算出した。全ての場合において、生のｆａｓｔｑを得、それらを、卵巣試験のＲＮＡ－Ｓｅｑに関して上に記載されているものと同じプロトコールに従ってアラインメントし、数量化した。
【０６９８】
それぞれＥおよびＯと称される予測される計数および観察された計数を表す２つの行列を創出し、行は遺伝子に対応し、列は試料に対応する。まず、各試料中の各遺伝子の長さにその発現を掛けることによって行列Ｅをポピュレートし、次いで、Ｅの列を再尺度化して、Ｅの列の合計とＯの列の合計がマッチするようにした。最後に、Ｅの行の合計とＯの行の合計を比較することによって遺伝子レベル解析を行った。遺伝子をヒト血漿中のそれらの存在および濃度に従って強調表示した。
【０６９９】
オートファジーに関連する遺伝子の解析
２つのオートファジー関連遺伝子セットを定義した。第１のセットは、公知のオートファジー関連遺伝子の物理的相互作用パートナーとして実験的に同定されたタンパク質を含むものであった。オートファジー相互作用ネットワークデータベース（ｂｅｓｒａ．ｈｍｓ．ｈａｒｖａｒｄ．ｅｄｕ／ｉｐｍｓｍｓｄｂｓ／ｃｇｉ－ｂｉｎ／ｄｏｗｎｌｏａｄｓ．ｃｇｉからアクセスされる）に寄託されている、ＩＰ－ＭＳ実験においてベイトとし

て使用された規準のオートファジー関連遺伝子各々（ｇｅｎｅｎａｍｅｓ．ｏｒｇ／ｃｇｉ－ｂｉｎ／ｇｅｎｅｆａｍｉｌｉｅｓ／ｓｅｔ／１０２２）について、「ＷＤ」信頼度スコア（ｂｅｓｒａ．ｈｍｓ．ｈａｒｖａｒｄ．ｅｄｕ／ｉｐｍｓｍｓｄｂｓ／ｃｇｉ－ｂｉｎ／ｔｕｔｏｒｉａｌ．ｃｇｉ）に従った上位１００種のタンパク質同定を同定した。２２の実験にわたってプーリングを行い、少なくとも１つの規準のオートファジー関連遺伝子に確信的に関連付けられる１００４種の独特の遺伝子のセットを得た（図１８Ｃ）。

【０７００】

オートファジー関連遺伝子の第２のセットを、仔マウス腎臓上皮（ｉＢＭＫ）細胞におけるＡＴＧ５ノックアウト前後の汎プロテオームタンパク質存在量をＳＩＬＡＣ（ｓｃｉｅｎｃｅｄｉｒｅｃｔ．ｃｏｍ／ｓｃｉｅｎｃｅ／ａｒｔｉｃｌｅ／ｐｉｉ／Ｓ１０９７２７６５１４００６１２１）を使用して測定した試験を使用して同定した。ｔ－統計値が＞５である遺伝子を、ＡＴＧ５ノックアウトによって安定化されるものと分類した（事前飢餓条件；補足的データファイルｍｍｃ２．ｘｌｓの変数「Ｉｎｔｅｒｃｅｐｔ＿ｔ」）。各マウスＵｎｉｐｒｏｔ　ＩＤをｈｇ１９　ＵＣＳＣ　ＩＤにマッピングするために、マウスＵｎｉｐｒｏｔ配列と大多数の９ｍｅｒを共有するヒトＵＣＳＣ　タンパク質配列を決定した（図１８Ｃ）。

【０７０１】

図１８Ｂに基づいて、遺伝子当たりの観察されたペプチド対予測されたペプチドの比Ｒを算出し、分子および分母の一方または両方に疑似カウントを足した。例えば、Ｒ＝（Ｏ＋１）／（Ｅ＋１）。これらの遺伝子についての相対的富化／枯渇を数量化するためには情報が不十分であると感じられた各遺伝子の相対的富化（Ｌｏｇ（Ｒ）＞０）または枯渇（Ｌｏｇ（Ｒ）＜０）を表すためにＬｏｇ（Ｒ）を取った。有効なＬｏｇ（Ｒ）算出を有する遺伝子の間で、Ｌｏｇ（Ｒ）分布を、ＩＰ－ＭＳデータセット内のもの、ＳＩＬＡＣデータセット内のもの、およびいずれのデータセットにもないものについてプロットした（図１８Ｃ）。

【０７０２】

供給源遺伝子の局在の解析、図１８Ｄに関連する

上記と同じｌｏｇ（Ｒ）スコアを使用して、各供給源遺伝子の局在に従った分布をプロットした（図１８Ｄ）。供給源遺伝子の局在を、Ｕｎｉｐｒｏｔ（ｕｎｉｐｒｏｔ＿ｓｐｒｏｔ．ｄａｔ）を使用して決定した。

【０７０３】

単一細胞ＲＮＡ－ＳｅｑデータにおけるクラスＩＩ発現データの解析、図１９Ａに関連する

単一細胞ＲＮＡ－Ｓｅｑデータを、３つの、ヒト腫瘍試料がプロファイリングされた以前に公開されたデータセットから得た。

【０７０４】

第１の試験には皮膚黒色腫からのデータを含めた。ファイル「ＧＳＥ７２０５６＿ｍｅｌａｎｏｍａ＿ｓｉｎｇｌｅ＿ｃｅｌｌ＿ｒｅｖｉｓｅｄ＿ｖ２．ｔｘｔ」をＧｅｎｅ　Ｅｘｐｒｅｓｓｉｏｎ　Ｏｍｎｉｂｕｓ（ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｅｏ／；ａｃｃｅｓｓｉｏｎ：ＧＳＥ７２０５６）からダウンロードした。腫瘍の状態フラグ「２」を有する細胞を腫瘍細胞として扱い、腫瘍の状態フラグ「１」、および「１」から「６」までと等しい免疫細胞型フラグで標識された細胞をそれぞれＴ細胞、Ｂ細胞、マクロファージ、内皮、線維芽細胞、およびＮＫとして扱った。他の細胞は全て外した。データをｌｏｇ２（ＴＰＭ／１０＋１）の単位でネイティブなまま示し、したがって、ＴＰＭ尺度に数学的に変換した。ＴＰＭ尺度になったら、各細胞についてのデータを、タンパク質をコードするＵＣＳＣ遺伝子の符号のセットにわたって再正規化して、１，０００，０００まで合計した（発現行列に現れないタンパク質をコードする遺伝子は暗黙のうちにゼロ発現を有するものとして扱った）。最後に、同じ細胞型および同じ供給源の生検材料に対応する単一細胞観察を平均して患者－細胞型レベルの発現推定値を生じさせた。

【０７０５】

第２の試験には、頭頸部腫瘍からのデータを含めた。ファイル「ＧＳＥ１０３３２２＿ＨＮＳＣＣ＿ａｌｌ＿ｄａｔａ．ｔｘｔ」をＧｅｎｅ Ｅｘｐｒｅｓｓｉｏｎ Ｏｍｎｉｂｕｓ（ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｅｏ／；ａｃｃｅｓｓｉｏｎ：ＧＳＥ１０３３２２）からダウンロードした。表中のデータは、ｌｏｇ２（ＴＰＭ／１０＋１）単位である；したがって、値をＴＰＭ単位に数学的に変換した。黒色腫試験と同様に、各細胞についてのデータを、タンパク質をコードするＵＣＳＣ遺伝子の符号のセットにわたって再正規化して、１，０００，０００まで合計し、同じ細胞型および同じ供給源の生検材料に対応する単一細胞観察を平均した。リンパ節生検材料に対応するデータを除外した。

第３の試験には、無処置の非小細胞肺からのデータを含めた。ファイル「ＲａｗＤａｔａＬｕｎｇ．ｔａｂｌｅ．ｒｄｓ」および「ｍｅｔａｄａｔａ．ｘｌｓｘ」をＡｒｒａｙ Ｅｘｐｒｅｓｓ（ｅｂｉ．ａｃ．ｕｋ／ａｒｒａｙｅｘｐｒｅｓｓ／；ａｃｃｅｓｓｉｏｎｓ：Ｅ－ＭＴＡＢ－６１４９およびＥ－ＭＴＡＢ－６６５３）からダウンロードした。データ（すでにＴＰＭ）単位を、以前に記載されているタンパク質をコードする遺伝子のセットにわたって再尺度化して、１，０００，０００まで合計した。最後に、同じ細胞型および同じ供給源の生検材料に対応する単一細胞観察を平均して患者－細胞型レベルの発現推定値を生じさせた。結腸直腸がんおよび卵巣がんにおける同様の試験を実施した。結果を図１９Ａに示す。

【０７０６】

簡単にするために、表４に、細胞型をマージしてネイティブなまま報告されたものよりも粒度を粗くした。

【表４】

表４

大雑把な名称	構成物の細胞型
肺胞	「肺胞」、「立方肺胞2型(AT2)細胞」は除く
FO B細胞	「濾胞性B細胞」
形質細胞	「形質B細胞」
CLEC9A+ DCs	「交差提示性樹状細胞」
monoDCs	「単球由来樹状細胞」
pDCs	「形質細胞様樹状細胞」
ランゲルハンス	「ランゲルハンス細胞」
マクロファージ	「マクロファージ」
顆粒球	「顆粒球」
内皮	「正常な内皮細胞」、「腫瘍内皮細胞」、および「低品質内皮細胞」、「リンパ管EC」は除く
上皮	「上皮細胞」および「低品質上皮細胞」
線維芽細胞	「COL12A1発現線維芽細胞」、「COL4a2発現線維芽細胞」、「GABARAP発現線維芽細胞」、「低品質線維芽細胞」、「正常な肺線維芽細胞」、「PLA2G2A発現線維芽細胞」、および「TFPI2発現線維芽細胞」
T細胞	「調節性T細胞」、「CD4+ T細胞」および「CD8+ T細胞」
NKs	「ナチュラルキラー細胞」
腫瘍	「がん細胞」
解析から除外	「赤芽球」および「MALT B細胞」

【０７０７】

５つの試験におけるＨＬＡ－ＤＲＢ１の発現レベルを図１９Ａにプロットした。

【０７０８】

腫瘍由来クラスＩＩ発現と間質由来クラスＩＩ発現の特徴付け

腫瘍に起因するＭＨＣクラスＩＩ結合性ペプチド発現と間質に起因するＭＨＣクラスＩＩ結合性ペプチド発現の相対量を決定するために、ＴＣＧＡ患者におけるクラスＩＩ経路

遺伝子の変異を同定し（ＤＮＡに基づいてコールされる）、クラスＩＩ変異を有する各患者について、遺伝子の変異したコピーおよび変異していないコピーの相対的な発現量を対応するＲＮＡ－Ｓｅｑで数量化した。さらに、以下を仮定した：

１．変異リードは腫瘍から生じる

２．非変異リードは間質または腫瘍における野生型対立遺伝子について生じる

３．腫瘍は変異コピーとほぼ等しい発現を有する野生型コピーを保持する

【０７０９】

これに基づいて、観察された変異体対立遺伝子分率ｆに関して、腫瘍に起因するクラスＩＩ発現の分率はおよそ２ｆであり、１００％以下であることが決定された。３種の遺伝子、つまりＣＩＩＴＡ、ＣＤ７４、およびＣＴＳＳを中核クラスＩＩ経路遺伝子として選択し、ＴＣＧＡで変異について評価した（同義変異およびＵＴＲ変異を除外しない）（データをＴｕｍｏｒＰｏｒｔａｌ（ｔｕｍｏｒｐｏｒｔａｌ．ｏｒｇ／）：ＢＲＣＡ、ＣＲＣ、ＨＮＳＣ、ＤＬＢＣＬ、ＭＭ、ＬＵＡＤ；ＴＣＧＡバルクダウンロード（ｔｃｇａ－ｄａｔａ．ｎｃｉ．ｎｉｈ．ｇｏｖ）：ＣＥＳＣ、ＬＩＨＣ、ＰＡＡＤ、ＰＲＡＤ、ＫＩＲＰ、ＴＧＣＴ、ＵＣＳ；Ｓｙｎａｐｓｅ（ｓｙｎａｐｓｅ．ｏｒｇ／＃！Ｓｙｎａｐｓｅ：ｓｙｎ１７２９３８３）：ＧＢＭ、ＫＩＲＣ、ＬＡＭＬ、ＵＣＥＣ、ＬＵＳＣ、ＯＶ、ＳＫＣＭ；または元のＴＣＧＡ公開（ｃａｎｃｅｒｇｅｎｏｍｅ．ｎｉｈ．ｇｏｖ／ｐｕｂｌｉｃａｔｉｏｎｓ）：ＢＬＣＡ、ＫＩＣＨ、ＳＴＡＤ、およびＴＨＣＡからダウンロードした）。これらの遺伝子をクラスＩＩ発現におけるそれらの公知の役割、および８５００のＧＴＥｘサンプルのコホートにわたるそれらのＨＬＡ－ＤＲＢ１との密接な相関に基づいて選択した。ＨＬＡ－ＤＲＢ１（ＨＬＡ－ＤＲＡ１、ＨＬＡ－ＤＰＡ１、ＨＬＡ－ＤＱＡ１、ＨＬＡ－ＤＱＢ１、およびＨＬＡ－ＤＰＢ１）と等価相関を有する他の遺伝子は、それらの多型的性質により偽陽性変異コールが起こりやすくなるので除外した。天然に、患者のごく一部のみがＣＩＩＴＡ、ＣＤ７４、またはＣＴＳＳに変異を有し、また、一部の腫瘍型に関しては、解析に利用可能な患者が存在しなかった。

【０７１０】

Ｂｉｎａｒｙ Ｓｅｑｕｅｎｃｅ Ａｌｉｇｎｍｅｎｔ／Ｍａｐ（ＢＡＭ）フォーマットの元の全エクソーム配列決定（ＷＥＳ）の配列を視覚的に評価して（ＩＧＶツール）、変異が腫瘍試料には存在するが正常な試料には存在しないことを確認した。変異体リード計数と野生型リード計数とを対応するＲＮＡ－Ｓｅｑからｐｙｓａｍを使用して得た。全体的なＨＬＡ－ＤＲＢ１発現をＧｅｎｏｍｉｃ Ｄａｔａ Ｃｏｍｍｏｎｓ（ｇｄｃ．ｃａｎｃｅｒ．ｇｏｖ）からダウンロードした発現データに基づいて決定し、それを、タンパク質をコードする遺伝子のセットにわたって再正規化して、１，０００，０００まで合計した。腫瘍に起因するＨＬＡ－ＤＲＢ１発現の分率（図１９Ｂ）をｍｉｎ（１，２ｆ）として推定し、ここで、ｆは、変異を示すＣＩＩＴＡ、ＣＤ７４、またはＣＴＳＳにおけるＲＮＡ－Ｓｅｑリードの分率である。

【０７１１】

天然のドナー組織に対する予測の全体的な性能の評価

７体の健康なドナー由来の末梢血を、上の節「抗体に基づくＨＬＡ－ペプチド複合体の単離」に記載されている通り、ＤＲ特異的抗体を用いてプロファイリングした。これらの結果に基づいて、２つのデータセットを定義した：一方は多変量ロジスティック回帰を当てはめるためのものであり、他方は回帰の予測性能を評価するものであった。

【０７１２】

第１のデータセットを、図４との関連で前に記載されているヒットおよびデコイ選択アルゴリズムを使用することによって築いた。手短に言えば、これは、各入れ子セットを１つのヒットペプチド（入れ子セット内で最も短いペプチド）で表すことおよび遺伝子にわたって長さを釣り合わせたデコイをタイリングすることを意味し、したがって、ヒットと最小に重複し、互いに最小に重複する。しかし、２つの重要な詳細が図４に関して概説したアルゴリズムとは異なる。第１に、ヒットおよびデコイは、「チューニング」区分内の遺伝子（「試験」区分ではなく）から選択され、第２に、デコイを、ゼロヒットを示した

遺伝子にマッピングさせた。ＭＨＣ結合スコア（ＮｅｔＭＨＣＩＩｐａｎまたはｎｅｏｎｍｈｃ２に由来するもの）ならびに他の入力特色（発現、遺伝子の偏りなど）を用いたロジスティック回帰モデルをこのデータセットに対して訓練した。

【０７１３】

第２のデータセット（評価のために使用する）を、「試験」区分から引き出されたヒットおよびデコイを使用した以外は同一の様式で築いた。表５に示す通り、結合スコアに加えて以下の変数を回帰のサブセットに使用した。

【表５】

表５

NetMHCIIpan親和性	NetMHCIIpanから導き出されたもの。各候補ペプチドについて、ドナーの遺伝子型における全てのDR対立遺伝子にわたって最も強いスコアを取得した。
MSに基づく結合スコア	*neonmhc2*から導き出されたもの。各候補ペプチドについて、ドナーの遺伝子型における全てのDR対立遺伝子にわたって最も強いスコアを取得した。
発現	遺伝子発現の推定値をbowtie2、RSEM、およびタンパク質をコードする遺伝子のみにわたる再正規化からのデータを解析することによって得、値をN個の試料にわたって平均した。
ホットスポット	候補ペプチドが以前に公開された複対立遺伝子データセットのうちのいずれかと少なくとも1つの9merを共有するかどうかについての指示変数(0/1)
遺伝子の偏り	解析当たりの(1+観察値)/(1+予測値)
交差提示	候補ペプチドがDCによりMHCクラスII上に交差提示されることが観察された遺伝子に由来するものかどうかについての指示変数(0/1)。

【０７１４】

性能評価の目的で、ｎ個のヒットペプチド全てを、４９９ｎ個のデコイ（デコイの完全なセットから、元に戻さずにランダムにサンプリングしたもの）のセットと一緒に所与のロジスティック回帰によって評価し、スコア化した。組み合わせたセットのペプチドの上位０．２％を陽性コールと標識し、陽性適中率（ＰＰＶ）を、ヒットであった陽性コールの分率として算出した。この評価シナリオでは、陽性の数はヒットの数と等しくなるように制約されるので、再現率はＰＰＶと正確に等しいことに留意されたい。対立遺伝子にわたって一貫した１：４９９の比を適用することは、そうでなければ各ドナーについて観察されるヒットの数の影響を高度に受ける性能値の安定化に役立つ。ヒットの数が、ドナー細胞の内在性特性よりも実験条件に大きく関連すると仮定されるので、これは妥当であると思われた。１：４９９の比は、ダウンサンプリングがインプリメントされなかった場合に使用されるものとそう遠くなかった。

【０７１５】

本発明の好ましい実施形態が本明細書において示され、記載されているが、そのような実施形態は単に例として提供されていることは当業者には明白であろう。本発明は、本明細書の中で提供される特定の実施例によって限定されるものではない。本発明が上記の明細書に関連して記載されているが、本明細書の実施形態の説明および図表は、限定の意味で解釈されることを意図していない。当業者は、本発明から逸脱することなく多数の変形、変化および置換をすぐに思いつくであろう。さらに、本発明の全ての態様は、本明細書に記載の特定の描写、構成または種々の条件および変化するものに依存する相対的な割合に限定されないことが理解されるべきである。本明細書に記載の発明の実施形態に対する種々の代替を本発明の実施に使用することができることが理解されるべきである。したがって、本発明は、任意のそのような代替物、改変、変形形態または等価物も包含することが意図されている。以下の特許請求の範囲により本発明の範囲が定義されること、および、それにより、これらの特許請求の範囲内に入る方法および構造ならびにそれらの等価物

も包含されることが意図されている。

【０７１６】

（実施例１１）

ＨＬＡクラスＩＩ対立遺伝子結合性エピトープのハイスループットな同定および検証

本実施例では、新規のＭＨＣ－ＩＩ対立遺伝子結合性ペプチドを同定および検証するための、時間分解蛍光エネルギー移動（ＴＲ－ＦＲＥＴ）を使用した代表的な信頼できるハイスループット方法を記載する。アッセイには以下のいくつかの部分がある。（１）細胞に、ＦＲＥＴアッセイ用の蛍光タグを有するＭＨＣ－ＩＩα鎖およびβ鎖の発現および分泌に適したベクター構築物をトランスフェクトすること、（２）分泌されたＭＨＣ－ＩＩ構築物のタンパク質産物を精製すること、（３）ペプチド交換アッセイを実施すること（図２２Ａ）。図２２Ｂおよび図２３に設計および手順をさらに例証する。本明細書に記載のアッセイでは、安定な細胞株を必要としなくてよく、また、単純な単離戦略を包含するので、高速かつ効率的な検出および検証プロトコールが促進される。さらに、四量体または多量体を使用して、例えばｎｅｏｎｍｈｃ２により予測されたエピトープをｉｎ　ｖｉｖｏで投与した後に抗原特異的ＣＤ４細胞を検出することができ、その後に生じた免疫応答を使用してＣＤ４＋　Ｔ細胞応答を検証する。

【０７１７】

高親和性ＭＨＣクラスＩＩ結合性ペプチドを同定するためのＣＬＩＰ－ＴＲ－ＦＲＥＴアッセイ

単一の構築物におけるＣＭＶプロモーターによって駆動されるＨＬＡクラスＩＩα鎖およびβ鎖の発現のための例示的なベクター、適当にフォールディングされたα鎖およびβ鎖対がもたらされるタンパク質産物が本明細書で提示される。適当にフォールディングされたα鎖およびβ鎖の形態では、α１サブユニットおよびβ１サブユニットは二量体の形態であり、α１サブユニットおよびβ１サブユニットは、生理的立体配置が似ているペプチドを受容することができる開いた受容末端を形成する。このアッセイの目的に関しては、これらのベクターにより発現された、適当にフォールディングされたα鎖およびβ鎖の形態を有するＨＬＡタンパク質産物は、ＨＬＡ単量体と称される。発現構築物は、リンカー、１つまたは複数のペプチド切断部位、分泌シグナル、二量体形成因子、例えば、ｃ－ＦｏｓおよびＪｕｎ、これらと連結したビオチン化モチーフ（ＢＡＰ）および１０×－Ｈｉｓタグ（配列番号２０）を含む。単量体を安定化し、分泌を補助するためにプレースホルダーペプチドを使用する。プレースホルダーペプチドは、ＣＭＶペプチドであり得る。プレースホルダーペプチドは、ＣＬＩＰペプチドであり得る。プレースホルダーペプチドは、対立遺伝子についてのＭＳに基づくリガンドームによって同定されたペプチドであり得る。プレースホルダーペプチドは、開いたα１－β１ペプチド受容末端においてＨＬＡペプチドと共有結合により結合したものであり得る。

【０７１８】

本明細書で使用される例示的な構築物は、図２３（上のパネル）に示されている通り、ＣＬＩＰとβ鎖の間にトロンビン切断部分が配置されたＣＬＩＰプレースホルダーペプチドをコードする。トランスフェクションを行い、トランスフェクトされた細胞を最適な成長に到達するまで培養したら、分泌型タンパク質（単量体）を含む細胞培養上清（培地）を収集し、精製のためにニッケル（Ni^{2+}）カラムを通過させた（図２４）。発現レベルおよび精製をクーマシー染色によって調査した。２８ｍｅｒのＣＬＩＰペプチドはβ鎖と会合したままであり、これはトロンビンによる処理によって切断され（図２４）、その後、試験（例えば、候補）ＨＬＡ－クラスＩＩ結合性ペプチドと競合させることによって取り除くことができる。ＣＬＩＰペプチドを首尾よく取り除く試験ペプチドは、ＩＣ５０によって測定される、ＣＬＩＰペプチドに置き換わる能力に基づいて、説明可能に、構築物のＭＨＣ－ＩＩヘテロ二量体への結合に関して同類ペプチドである。新規の試験ペプチドを上記の競合置換え反応に使用することができ、次いで、質量分析（ＭＳ）によって同定することができる。

【０７１９】

ＣＬＩＰプレースホルダーペプチドを用いてＨＬＡ－ＤＲヘテロ二量体構築物の大きな集合を作出し、それらは首尾よく分泌され、ペプチド交換アッセイを実施した。

【０７２０】

ＣＤ７４に由来するペプチドプレースホルダーＣＬＩＰがＨＬＡクラスＩＩ単量体の分泌に対して有意な効果を有することが観察された。ＣＤ７４配列ＰＶＳＫＭＲＭＡＴＰＬＬＭＱＡ（配列番号１）を有する編集された規準のＣＬＩＰペプチド（図２５Ａ～２５ＣではＣＬＩＰ０と称される）が一般にプレースホルダー配列として使用された。しかし、一部のＨＬＡ－ＤＲペプチド、例えば、ＤＲＢ１＊１２：０１およびＤＲＢ１＊１３：０２は規準のペプチドを用いると収量が低いことが認められた（表６）。ある特定のＨＬＡＤＲＢ対立遺伝子二量体が、配列：ＬＰＫＰＰＫＰＶＳＫＭＲＭＡＴＰＬＬＭＱＡＬＰＭ（配列番号２１）（ＣＬＩＰ１）のアミノ酸の全部または一部を包含する配列よりも長い結合性配列を有することが観察された（図２５Ａ）。実際に、ＤＲＢ１＊１２：０１およびＤＲＢ１＊１３：０２の場合、ＣＬＩＰ０配列の代わりにＣＬＩＰ１配列を使用することにより、ＨＬＡ二量体の分泌収量が改善された（図２５Ｂ～２５Ｃ）。

【表６】

表 6

DRB1*	ペプチドプレースホルダー	配列番号	収量(mg/L)
01:01	PVSKMRMATPLLMQA (**CLIP**)	1	20
	LPLKMLNIPSINVH (**CMV**)	22	100
	PKYVKQNTLKLAT (**HA**)	23	85
12:01	PVSKMRMATPLLMQA (**CLIP**)	1	<3
13:02	PVSKMRMATPLLMQA (**CLIP**)	1	<3

【０７２１】

ＳＴＩＩ－ＴＲ－ＦＲＥＴを使用した上首尾のペプチド交換アッセイによるペプチドの新規のスクリーニング

ペプチドを、上記のＨＬＡ－単量体タンパク質をＥｘｐｉ２９３細胞などの細胞株において発現させ、上清から収集および精製し、ペプチド交換アッセイに供することを伴うアッセイを使用して新規にスクリーニングすることができる。予測アルゴリズムによって予測されたＨＬＡクラスＩＩ結合性ペプチドを、ペプチド交換アッセイを使用して試験することができた。ペプチド交換アッセイは、蛍光偏光を伴う方法を使用して実施することができる。例えば、プレースホルダーペプチドを標識するため、または試験ペプチドを標識するため、または２つの異なるフルオロフォアを使用してその両方を標識するために任意のフルオロフォアを使用することができる。置換え反応を定量的に評価するために、フルオロフォアで予め標識されていたプレースホルダーペプチドの結合の喪失によるか、または、そうでなければＨＬＡと結合した形態では生化学的反応によってクエンチされる、遊離したフルオロフォアの蛍光発光による蛍光の変化を記録することができる。あるいは、非蛍光プレースホルダーペプチドが標識された蛍光ペプチドに置き換わったことを記録して、置換え反応を定量的に決定することができる。例示的なアッセイでは、ＦＩＴＣで標識されたプレースホルダーＣＬＩＰペプチドを使用して、ＣＭＶペプチドなどの既存の共有結合したペプチドを置き換える。ＦＩＴＣで標識されたペプチドは、ＨＬＡと結合している場合、高い極性化を誘導する。試験ペプチドを用いてＦＩＴＣ－プレースホルダーペプチドの力価測定を行った場合、試験ペプチドによりＦＩＴＣ－ＣＬＩＰが置き換えられ、それにより、蛍光の低減が導かれる。

【０７２２】

本明細書に記載の蛍光偏光の代わりに時間分解ＦＲＥＴ（ＴＲ－ＦＲＥＴ）技術を使用してペプチド交換アッセイを実施することもできる。本明細書に記載の例示的なＴＲ－Ｆ

ＲＥＴアッセイでは、細胞に、Ｓｔｒｅｐｔａｇ　ＩＩ（ＳＴＩＩ）部分を含むプレースホルダーペプチドを有するＨＬＡ単量体構築物をトランスフェクトした。ＳＴＩＩに対するＡｌｅｘａ－６４７タグが付された抗体によってＳＴＩＩ部分を検出した。同時に、本実施例において前に記載した単量体構築物のＪｕｎ末端に付着した、ＨＬＡペプチドのα２－β２末端の付近に存在するＨｉｓタグを抗Ｈｉｓ抗体とカップリングしたユウロピウムＩＩＩ（Ｅｕ）化合物によって検出した（図２６Ａ）。プレースホルダーペプチドがＨＬＡ単量体と結合したままの場合、ＦＲＥＴ反応においてＥｕ複合体はエネルギー供与体として作用し、一方、Ａｌｅｘａ６４７は受容体として作用する。試験ペプチドによりＳＴＩＩ－ＣＬＩＰプレースホルダーが置き換えられると、Ａｌｅｘａ－６４７－αＳＴＩＩペプチドは遊離するが、もはや蛍光によって検出されない。ＴＲ－ＦＲＥＴアッセイは蛍光偏光よりも信頼できることが見いだされた。さらに、当該アッセイでは、バックグラウンドシグナルがはるかに低減した（蛍光読み取りデータが図２６Ｂ～２６Ｅに示されている）。当該アッセイにより、ＨＬＡ－ペプチド対に対するハイスループットな同定プラットフォームがもたらされる。以下の表７に示されている通り、試験ペプチド（または候補ペプチド）Ｐ－１５６～Ｐ－１９１は広範囲の置換え能力、および、各実行で算出されたＩＣ５０によって決定された結合親和性を示した。ＩＣ５０が低いほど、高い置換え能力、および高い結合親和性が実証される。

【表７】

表７

ペプチド	IC50 (nm) 実行 1	IC50 (nm) 実行 2	平均 IC50 (nM)	標準偏差 (nM)	所見
P- 156	36294	38246	37270	976	
P- 157	2550	2243	2397	154	
P- 158	5786	5815	5800	15	
P-159	58	94	76	18	
P-160	1668	2020	1844	176	
P-161	13541	14401	13971	430	
P-162	3298	3636	3467	169	
P-163	>40,000	>40,000	>40,000	N/A	40μM で<50%の置き換わり
P-164	4553	4353	4453	100	
P-165	357	422	389	32	
P-166	3448	3104	3276	172	
P-167	6612	5906	6259	353	
P-189	4597	----	----	---	
P-190	1137	-----	----	---	
P-191	5167	-----	----	---	
P-192	>40,000				40μM まで置き換わりなし
HA	23	23	23	0	
ASP51	>40,000	>40,000	N/A	N/A	40μM まで置き換わりなし。

【０７２３】

示差走査蛍光光度法（ＤＳＦ）を使用したペプチド交換検証

この方法では、特定のＨＬＡ対立遺伝子に結合し得るペプチドをスクリーニングするためのハイスループットアッセイ、およびまた、ＨＬＡ二量体とのペプチド結合の強度を決定する（図２６Ｆ）。このアッセイでは、熱の適用によって対立遺伝子が互いに解離した場合にのみタンパク質の疎水性残基に結合し、したがって、ＭＨＣ対立遺伝子に結合し得る蛍光プローブを使用する。ＭＨＣクラスＩＩ二量体が同類ペプチドに結合すると、二量体はその二量体形態で一緒に保持される。結合した形態のＭＨＣ二量体－ペプチドに熱を適用すると、弱く結合しているペプチドはより速くＭＨＣクラスＩＩタンパク質二量体から解離し、それにより、フルオロフォアが解離したＭＨＣアルファ鎖およびベータ鎖に結

合することが可能になり、高い蛍光が生じる。蛍光を温度に応じて記録する。代表的な融解曲線を図２６Ｆに示す。融解曲線を比較して、強力な結合体（より高い温度で蛍光が検出される）と弱い結合体（より低い温度で蛍光が検出される）とを決定することができる。

【０７２４】

可溶性ＨＬＡ－ＤＭ（ＨＬＡ－ｓＤＭ）の、ＭＨＣクラスＩＩペプチド交換の触媒としての使用：ＨＬＡ－ＤＭは、天然のシャペロンであり、また、ＨＬＡ－ＤＲ、ＨＬＡ－ＤＰ、およびＨＬＡ－ＤＱ分子のペプチド交換触媒である。ＨＬＡ－ＤＭは内在性膜タンパク質であり、アルファポリペプチド鎖とベータポリペプチド鎖のヘテロ二量体として存在する（ＤＭＡおよびＤＭＢ）。この節に記載されているペプチド交換は、可溶性形態のＨＬＡ－ＤＭ（例えば、ＨＬＡ－ｓＤＭタンパク質）をＨＬＡ－ＤＲ、ＨＬＡ－ＤＰ、およびＨＬＡ－ＤＱ交換のシャペロンとして使用して実施する。図２６Ｇに示されている通り、ＨＬＡ－ｓＤＭタンパク質はＥｘｐｉ－ＣＨＯ細胞における一過性トランスフェクションによって産生される。簡単に述べると、図２６Ｇの上半分に図表で示されている通り、組換えＨＬＡ－ｓＤＭ構築物を設計する。組換えＨＬＡ－ｓＤＭ構築物は、リーダー配列の上流にあり、プロモーターに作動可能に（ｏｐｅｒａｂｌｙ）連結したＣＭＶ構成的プロモーターを含む。リーダー配列は、産物の分泌を補助する（分泌シグナル）。リーダー配列の３’末端に、ＨＬＡ－ＤＭベータ鎖外部ドメインをコードする（および膜貫通ドメインを欠く）配列を導入する。ビオチン化モチーフ（ＢＡＰ）をコードする配列を、ベータ鎖をコードする配列の３’にライゲーションする。ＨＬＡ－ＤＭアルファ鎖外部ドメインをコードする（および膜貫通ドメインを欠く）配列を分泌配列（リーダー）と共に５’末端に配置し、ＢＡＰ配列とは、リボソームスキッピング配列を介在させることによって分離する。ＨＬＡ－ＤＭアルファ鎖配列を３’末端において１０×Ｈｉｓタグ(配列番号２０）とライゲーションする。ヘテロ二量体ＨＬＡ－ｓＤＭが形成されたら、細胞の外部に分泌される。この構築物をＥｘｐｉ－ＣＨＯ細胞において発現させると、ＨＬＡ－ｓＤＭタンパク質が培養培地中に分泌される。

【０７２５】

Ｅｘｐｉ－ＣＨＯ細胞に、ＨＬＡ－ｓＤＭ構築物を発現するプラスミドベクターをトランスフェクトし、約１４日間の期間にわたって培養した。培養期間にわたってタンパク質を培養培地中に分泌させた。ＨＬＡ－ｓＤＭタンパク質を培養物からＭＨＣ－ＩＩタンパク質の精製と非常に類似したプロセスで精製した。ＭＨＣ－ＩＩペプチド交換は、酸およびＨＬＡ－ｓＤＭを用いて、または酸を用いずに、およびオクチルグルコシドを用いて効率的に実施することができる。分子ふるいクロマトグラフィーを実施して、ペプチド交換を評価した。結果が図２６Ｈに示されている。ペプチド交換アッセイは全て、ＨＬＡ－ｓＤＭまたはオクチル－グルコシドを触媒として使用して実施した。

【０７２６】

ＨＬＡ－クラスＩＩ四量体（または多量体）レパートリー

エピトープ：ＨＬＡ結合および解離カイネティクスを生化学的アッセイにおいて試験するために、ＨＬＡクラスＩＩ四量体の大きなレパートリーを生成した。したがって、これらの生成されたクラスＩＩ四量体を、ペプチド結合および提示をアッセイするために使用する。例えば、四量体をペプチド交換アッセイに使用した。図２７Ａに示されている通り、１２種の四量体を生成し、６種の四量体を１５ｍｇ／ｍｌよりも高い濃度で保存し、４種の四量体を＜５ｍｇ／ｍｌで保存した。ＨＬＡ四量体をフローサイトメトリーに使用して、ネオ抗原反応性ＣＤ４＋　Ｔ細胞を同定する。インフルエンザウイルスエピトープ（ＨＡ）およびＨＩＶエピトープを、存在する場合にはＴ細胞認識についてＨＬＡ四量体によって試験した（図２７Ｅ）。

【０７２７】

図２７Ｂ～２７Ｄは、生成および精製されたＨＬＡクラスＩＩ四量体の種々のサブセットを示す。図２７Ｂに示されている通り、ＤＲＢ１ヘテロ二量体四量体の大きなレパートリーを構築し、１５ｍｇ／Ｌよりも高い濃度で精製した。図２７Ｃおよび２７Ｄは、蛍光に基づくペプチド結合アッセイのために作製され、検証されたヒトＭＨＣクラスＩＩ対立

遺伝子構築物のカバレッジを要約するものである。表８Ａ、表８Ｂおよび表８Ｃには、製造された対立遺伝子の四量体の一覧が、精製された産物の対応する分泌収量濃度と共に提示されている。

【表８Ａ】

表 8A

HLA ヘテロ二量体	分泌収量
DRB1*01:01	>15 mg/L
DRB1*04:01	>15 mg/L
DRB1*04:02	>15 mg/L
DRB1*04:04	>15 mg/L
DRB1*04:05	>15 mg/L
DRB1*08:01	>15 mg/L
DRB1*09:01	>15 mg/L
DRB1*11:01	>15 mg/L
DRB1*13:03	>15 mg/L
DRB1*14:01	>15 mg/L
DRB1*15:03	>15 mg/L
DRB1*01:02	>15 mg/L
DRB1*11:04	>15 mg/L
DRB1*15:02	>15 mg/L
DRB4*01:01	>15 mg/L
DRB1*07:01	5-15 mg/L
DRB1*13:01	5-15 mg/L
DRB1*13:02	5-15 mg/L
DRB1*15:01	5-15 mg/L
DRB1*15:02	5-15 mg/L
DRB3*01:01	5-15 mg/L
DRB1*08:03	5-15 mg/L
DRB1*11:02	5-15 mg/L
DRB1*16:02	5-15 mg/L
DRB3*02:01	5-15 mg/L
DRB3*02:02	5-15 mg/L
DRB3*03:01	5-15 mg/L
マウス I-Ab	5-15 mg/L
DRB1*03:01	<5 mg/L
DRB1*12:01	<5 mg/L
DRB5*01:01	<5 mg/L
DPA*01:03/DPB*04:01	<5 mg/L

【表８Ｂ】

表 8B

HLA ヘテロ二量体	分泌収量
DPB1*05:01	>15 mg/L
DPB1*13:01	>15 mg/L
DPB1*03:01	5-15 mg/L
DPB1*04:02	5-15 mg/L
DPB1*06:01	5-15 mg/L
DPB1*11:01	5-15 mg/L
DPB1*01:01	<5 mg/L
DPB1*02:01	<5 mg/L
DPB1*02:02	<5 mg/L
DPB1*04:01	<5 mg/L
DPB1*17:01	<5 mg/L

【表８Ｃ】

表 8C

HLA ヘテロ二量体	分泌収量
A1*02:02+B1*06:02	>15 mg/L
A1*02:01+B1*02:02	>15 mg/L
A1*01:03+B1*06:03	>15 mg/L
A1*02:01+B1*03:03	>15 mg/L
A1*01:02+B1*06:04	5-15 mg/L
A1*05:01+B1*02:01	<5 mg/L
A1*05:05+B1*03:01	<5 mg/L
A1*01:01+B1*05:01	<5 mg/L
A1*03:01+B1*03:02	<5 mg/L
A1*03:03+B1*03:01	<5 mg/L

【０７２８】

ＭＨＣ－ＩＩ四量体産物パイプラインは、ＤＲＢ３、ＤＲＢ４、およびＤＲＢ５対立遺伝子、ならびにＤＰおよびＤＱ対立遺伝子をさらに含む。

【０７２９】

蛍光偏光（ＦＰ）を使用したペプチド交換検証

ＭＨＣクラスＩＩタンパク質に結合したペプチドと遊離のペプチドを区別するためのアッセイに蛍光偏光顕微鏡を使用した。蛍光タグが付されたプレースホルダーペプチドは、ＭＨＣクラスＩＩ二量体と結合すると、その遊離形態、フルオロフォアタグ付けされていない競合エピトープペプチドがプレースホルダーペプチドに置き換わることによってＭＨＣクラスＩＩ二量体に結合したままの場合と比較して、蛍光偏光（ＦＰ）顕微鏡による高度に極性化した光をもたらす。図２８Ａは原理をグラフ表示によって示す。簡単に述べると、アッセイを以下の一般化された方法で実施し、変形形態がそれぞれの説明に示されて

いるか、または当業者に容易に理解される。

【０７３０】

表９（以下）に記載されている試薬を反応チューブ（例えば、１．５ｍｌのＥｐｐｅｎｄｏｒｆチューブ）中でアセンブルし、十分に混合し、３７℃で２時間インキュベートする。インキュベーション時間の最後に１０×ＰＢＳ、２５ｍｌを混合物に添加してペプチド交換反応を中和する。

【表９】

表9

成分	ストック濃度	最終濃度
トロンビンにより消化される MHC class II 対立遺伝子	不定	5 μM
交換ペプチド/目的のペプチド	10mM	50 μM
酢酸ナトリウム、pH5.2	1M	100 mM
塩化ナトリウム	5M	50 mM
可溶性 DM	不定	5 μM
MiliQ 水		最大 100 μl

【０７３１】

交換されたペプチドを、例えば染色によって検出する；または後で評価するために液体窒素でスナップ凍結させることによって－８０℃で保管する。

【０７３２】

図２８Ｂおよび２８Ｃは、ＨＬＡ　ＤＲＢ１＊０１：０１についてのＦＰを使用したアッセイ展開、および使用した種々の条件の概要を提示する。一部の実施例では、アッセイに対するｐＨの影響を決定した。手短に言えば、全長対立遺伝子および可溶性対立遺伝子の両方を細胞において発現させる。膜に結合した全長対立遺伝子の形態は膜を透過処理することによって回収し、一方、分泌形態は細胞の上清から回収する。回収されたＨＬＡクラスＩＩタンパク質を、ニッケル（Ni^{2+}）カラムを通過させることによって精製する。一部の実施例では、全長ＭＨＣＩＩ対立遺伝子を回収するための膜透過処理における界面活性剤（１％オクチルグルコシド対１．６％ＮＰ４０）の影響を評価した。一部の実施例では、温度、または使用するプローブ、または精製方法もしくは標的フォーマットの影響を個別に評価した（図２８Ｂ）。

【０７３３】

コンフォメーション特異的抗体Ｌ２４３またはＨｉｓタグ精製のいずれかを使用した精製方法の影響を評価した。結果を図２８Ｄに示す。各データポイントが表の左側に示されており、右側のグラフに単一のドットとして表されている。グラフのドットは大まかにどちら軸にも４５度の角度に沿って整列しており、ｒ値は０．９６２１であり、これは、どちらの精製方法によるＩＣ５０値も互いにと一致することを示す。これはまた、ペプチド効力の順位が精製方法間で変わらないことも示す。

【０７３４】

可溶性形態のＨＬＡクラスＩＩタンパク質（ｓＤＲ１）か全長形態のＨＬＡクラスＩＩタンパク質（ｆＤＲ１）の選択の影響を評価し、図２８Ｅは、標的フォーマットの選択がペプチド効力に影響を及ぼさないことを示す。左側にはｓＤＲ１形態またはｆＤＲ１を使用した実験からの平均ＩＣ５０値が示されている。これらのデータをプロットして右側のグラフを得る。各々が単一のドットによって示されているデータ点は、どちらの軸にも大

まかに４５度の角度に沿って整列しており、ｒ値は０．９３６５であり、これは、使用した形態のどちらによるＩＣ５０値も互いに十分に相関することを示す。これはまた、ペプチド効力の順位が精製方法間で変わらないことも示す。

【０７３５】

図２８Ｆは、ｎｅｏｎｍｈｃ２およびＮｅｔＭＨＣＩＩｐａｎにより予測されたペプチドの結合アッセイにおける例示的な評価方法および不調和のペプチドの同定の図解を示す。Ｎｅｏｎｍｈｃ２およびＮｅｔＭＨＣＩＩｐａｎにより予測されたペプチドを実際のペプチド結合アッセイにおいて評価するために蛍光偏光アッセイを使用した。アッセイのために、６０ｎＭのトロンビンにより消化された可溶性ＨＬＡ－ＤＲＢ１＊１５：０１をＦＩＴＣタグが付された超結合体プローブペプチド（ＰＶＶＨＦＦＫ（ＦＩＴＣ）ＮＩＶＴＰＲＴＰＰＹ（配列番号２４））（アッセイ当たり１０ｎＭ）およびアッセイペプチドと一緒にアッセイ緩衝剤（ｐＨ５．２）中、３７℃で５時間インキュベートした。蛍光偏光を調査し、それから％プローブ置換えを算出した。図２８Ｇに示されている通り、超結合体蛍光ペプチドの阻害は予測されたペプチドの濃度に比例し、これにより、アッセイの特異度が良好であることが示される。Ｎｅｏｎｍｈｃ２によるペプチドの予測とＮｅｔＭＨＣＩＩｐａｎによるペプチドの予測の性能に著しい差異が見られる。Ｎｅｏｎｍｈｃ２によるペプチドの予測では、より多くのペプチドが陽性に結合し、阻害の程度がより高かったが、一方、ＮｅｔＭＨＣＩＩｐａｎによるペプチドの予測はＮｅｏｎｍｈｃ２ペプチドと比較して全体的に不十分であった。図２８Ｈは、Ｎｅｏｎｍｈｃ２により予測されたペプチドの結合アッセイにおける評価を要約する。円グラフによって示されている通り、ダブルネガティブペプチド（ＮｅｔＭＨＣＩＩｐａｎとＮｅｏｎｍｈｃ２のいずれによっても予測されなかったペプチド）のうち、結合体であることが判明したのは５％だけであり、９５％が非結合体であった。ＮｅｔＭＨＣＩＩにより予測されたペプチドのうち、４０％がプローブ置換えアッセイの蛍光偏光検出により結合体であり、Ｎｅｏｎｍｈｃ２により結合体であると予測されたペプチドのうち、１００％が真の結合体であることがプローブ置換えアッセイによって見いだされた。

【０７３６】

Ｓｅｔｔｅらによって特定の対立遺伝子に結合することが示された以前に公開されたペプチドを精査することにより、ＦＩＴＣ標識されたプローブを調製した。次いで、これらのペプチド配列を、予測されたクラスＩＩ結合性コアを使用して解析して、ペプチドおよびアンカー残基の最小の９ｍｅｒのコアを同定した。次いで、この情報を、リシン置換およびＦＩＴＣ標識のための残基位置を選択する際に考慮した。例えば、下記の表（表１０）には、Sette et al.（Sidney J, Southwood S, Moore C, et al. Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture. Curr Protoc Immunol. 2013;Chapter 18:Unit-18.3. doi:10.1002/0471142735.im1803s100）（以下「Ｓｅｔｔｅ配列」））に記載されている配列が列挙されている。特定の対立遺伝子に関して、各ペプチドについての予測されたクラスＩＩ結合性コアに下線が引かれている。太字は、エピトープ改善の結果として同定されたアンカー位置を示す。一部の場合では、異なる対立遺伝子に対して同じペプチド配列を使用することができる。

【表１０】

表 10

略記 ID	対立遺伝子	Sette 配列	配列番号	選択されたプローブ配列	配列番号
SB-DR7/11	DRB1*07:01	YATF**F**IK**A**NSKF**I**GITE	25	YATFFI***K***ANSKFIGITE	25
SB-DR7/11	DRB1*11:01	YATFF**I**KANS**K**FIGITE	25	YATFFI***K***ANSKFIGITE	25
SB-DR9	DRB1*09:01	TLSVTFIGAAPLILSY	26	TLSVTFIGAAP***K***ILSY	29
SB-DR4/15	DRB1*15:01	PV**V**HF**F**K**N**IV**T**PRTPPY	27	PVVHFF***K***NIVTPRTPPY	27
SB-DR4/15	DRB1*04:01	PVVHF**F**KNIV**T**PR**T**PPY	27	PVVHFF***K***NIVTPRTPPY	27
SB-DR3	DRB1*03:01	YARI**R**RD**D**GCLL**R**LVD	28	YARI***K***RDGCLLRLVD	30

【０７３７】

上記の配置に基づいて、青色でない結合性コア内の位置（下線が引かれている）に焦点を当てることによってＦＩＴＣコンジュゲーションのための内部リシンを選択した－赤色で強調表示されているのがこれらの位置であり、必要に応じてＦＩＴＣコンジュゲーションのための位置である。ＦＩＴＣコンジュゲーションのための内部リシンを有さない配列に関しては、対立遺伝子の結合性モチーフとペプチド配列の比較を実施し、ＤＲＢ１＊０９：０１およびＤＲＢ１＊０３：０１ペプチドについての内部リシン置換のための位置を選択する手動による手法を行った（上記の表を参照されたい）。より詳細には、ＦＩＴＣコンジュゲーションを可能にするために、ＤＲＢ１＊０９：０１についてはロイシン残基、およびＤＲＢ１＊０３：０１についてはアルギニン残基をリシンで置換した。コンジュゲートしたフルオロフォアは結合した時に極性化した光を放出する可能性がより高い（すなわち、フルオロフォアの運動がより大きく制限される）ので、この置換戦略は、位置の手動での同定が強力なアミノ酸優先性を伴わない（ｎｅｏｎｍｈｃ２により予測された９ｍｅｒのコアの中間にも存在する）ＭＡＰＴＡＣ由来のモチーフに基づくものであった。

【０７３８】

（実施例１２）

治療的に標的化可能ながん抗原を同定するためのＨＬＡクラスＩＩ結合およびプロセシング規則

増加しているエビデンスにより、ＣＤ４＋　Ｔ細胞はがん特異的抗原を認識し、腫瘍成長を制御することができることが示されている。しかし、依然としてヒト白血球抗原クラスＩＩ分子（ＨＬＡクラスＩＩ）によって提示される抗原を予測することは難しく、それにより、それらの抗原を治療的に最適に標的化する試みが妨害されている。障害には、不正確なペプチド結合予測およびＨＬＡクラスＩＩ経路の未解決の複雑さが含まれる。本実施例では、ＨＬＡクラスＩＩ結合性モチーフを発見するための改善された技術を記載する。さらに、腫瘍微小環境（ＴＭＥ）における関連性のあるプロセシング規則を学習するために行う腫瘍－リガンドームの包括的な解析が本明細書に記載されている。

【０７３９】

４０種のＨＬＡクラスＩＩ対立遺伝子をプロファイリングし、結合性モチーフがペプチドローディングシャペロンであるＨＬＡ－ＤＭに対して高い感受性を持つことが示された。がん細胞ではなくプロフェッショナル抗原提示細胞（ＡＰＣ）による腫瘍内ＨＬＡクラスＩＩ提示が優位を占めることが明らかになった。これらの知見を統合して、ファゴサイトーシスされたがん細胞由来のペプチドを含めたＡＰＣリガンドームを正確に予測する本

明細書に記載のアルゴリズムを開発した。これらのツールおよび生物学的洞察により、ＨＬＡクラスＩＩを対象とするがん治療を増強することができる。

【０７４０】

有望な新しいクラスの治療は、がん抗原および新抗原と称される体細胞変異した配列に対するＴ細胞応答を誘導することによってがんを処置しようとするものである。現在のところ、これらの試みでは、ＨＬＡクラスＩ（ＨＬＡクラスＩ）により提示されるリガンドに対するＣＤ８＋　Ｔ細胞応答を引き出すことに主に焦点が当てられている。しかし、いくつかの最近の試験により、ＣＤ４＋　Ｔ細胞がＨＬＡクラスＩＩにより提示されたリガンドを認識し、腫瘍の制御に寄与することができることも示されている。がんワクチンおよび他の免疫治療はＣＤ４＋　Ｔ細胞応答を方向付けることを利用することが理想的であるが、現行の試みでは、現行の予測ツールの正確度が不十分なことから、ＨＬＡクラスＩＩ抗原予測が完全に無視されている。

【０７４１】

ＨＬＡクラスＩＩがん抗原の正確な同定を妨げる重要な因子は、ペプチド結合の規則を学習するために必要な包括的な高品質データの利用可能性である。各対立遺伝子バリアントが別個のペプチド結合優先性を示す３種の高度に多型的な規準のＨＬＡクラスＩＩ遺伝子座、ＨＬＡ－ＤＲ、ＨＬＡ－ＤＰ、およびＨＬＡ－ＤＱに関してデータが必要である。ペプチド結合性モチーフを定義するために広く使用されている方法は、ＨＬＡ－ＤＭなどの生理的シャペロンの非存在下で単一のペプチドの親和性を測定する生化学的アッセイである。測定される親和性データカバレッジは一般的なコーカサス人ＨＬＡ－ＤＲ対立遺伝子に限定され、これらの対立遺伝子についてさえ、予測の正確度はＨＬＡクラスＩの予測に著しく後れをとっている。原理上は、質量分析（ＭＳ）に基づくリガンドミクスにより、スケーラビリティおよび内因性ペプチドローディング条件がもたらされることによって予測の改善が可能になるはずである。それにもかかわらず、天然のリガンドームは複対立遺伝子であり、それにより、正確な訓練用データを得るために必要なペプチドと対立遺伝子間のマッピング情報が隠れる。ＨＬＡクラスＩに関してはこの問題を解決する進歩があり、そこでは、逆畳み込みと、ロースループットトランスジェニックマウスモデルＨＬＡクラスＩＩ欠損細胞株、またはホモ接合性ＨＬＡ－ＤＲ対立遺伝子を有する細胞株を使用して生成された単一対立遺伝子ＨＬＡクラスＩＩ細胞株単一対立遺伝子ＨＬＡクラスＩＩリガンドームデータセットの両方が使用される。

【０７４２】

別の難題は、腫瘍抗原が入る可能性が最も高いＨＬＡクラスＩＩ提示経路の不明確性である。最近のＭＳに基づく試験により、腫瘍試料のＨＬＡクラスＩＩリガンドームが調査されたが、プロフェッショナルＡＰＣまたはがん細胞が治療的に関連性のあるＨＬＡクラスＩＩ抗原を提示するかどうかについては対処されていない。さらに、ＨＬＡクラスＩＩによる腫瘍抗原のプロセシングが主にファゴサイトーシスに依存するのかオートファジーに依存するのかは現在のところ分かっていない。関連性のある細胞型において優位を占める経路に応じて、ＨＬＡクラスＩＩペプチドリガンドの供給源としていずれのタンパク質が好ましいかに関して強烈な差異が存在し得る。問題を複合すると、一般的な理論によりタンパク質配列特色がＨＬＡクラスＩＩプロセシングの潜在性に影響を及ぼすはずであることが保持されるにもかかわらず、タンパク質内のいずれの領域がＨＬＡクラスＩＩリガンドを産生する可能性が最も高いかを決定するための系統的な手法は存在しない。

【０７４３】

治療的に標的化可能なＨＬＡクラスＩＩ抗原のプロセシングおよび提示規則を調査するために、ｉ）ペプチド結合予測を改善すること、およびｉｉ）ＴＭＥにおいてＨＬＡクラスＩＩリガンドがどのようにプロセシングされ、提示されるかを決定することの２方向からの手法に従った。対立遺伝子特異的ペプチド結合規則を学習するために、ＭＡＰＴＡＣ（商標）（Ｍｏｎｏ－Ａｌｌｅｌｉｃ Ｐｕｒｉｆｉｃａｔｉｏｎ ｗｉｔｈ Ｔａｇｇｅｄ Ａｌｌｅｌｅ Ｃｏｎｓｔｒｕｃｔｓ）と称される、内因性に提示されるＨＬＡクラスＩＩリガンドの配列決定を行うためにＭＳを利用するスケーラブルな単一対立遺伝子ＨＬ

Ａリガンドームプロファイリングワークフローを開発した。ＭＡＰＴＡＣ（商標）により、ペプチド結合性モチーフを４０種のＨＬＡクラスＩＩ対立遺伝子に明白に分解し、免疫原性ウイルスエピトープおよび新抗原を正確に同定することができる結合予測アルゴリズムを訓練することが可能になる。ＨＬＡクラスＩＩプロセシング予測を改善するために、腫瘍試料を解析し、それにより、プロフェッショナルＡＰＣを腫瘍内ＨＬＡクラスＩＩ発現の主要な供給源として確立し、これらの細胞によって優先的にプロセシングされる遺伝子のセットおよび遺伝子領域を定義した。次いで、結合特色およびプロセシング特色を組み込むアルゴリズムにより、天然のＡＰＣリガンドーム、およびより重要なことに、エンドサイトーシスによって取り込まれたがん細胞に由来するＨＬＡクラスＩＩリガンドのサブセットを予測することができることを実証した。治療的に関連性のあるＨＬＡクラスＩＩ抗原のプロセシングおよび提示規則の理解におけるこれらの進歩により、ＣＤ４＋　Ｔ細胞応答を利用することを目的とする治療が可能になる。

【０７４４】

実験手順

ＭＡＰＴＡＣ（商標）構築物の設計および細胞培養

ＨＬＡクラスＩに関しては、α鎖を、Ｃ末端ＧＳＧリンカー、その後、ビオチン－アクセプター－ペプチド（ＢＡＰ）配列、終止コドン、および変動するＤＮＡバーコードと融合し、ｐＳＦ　Ｌｅｎｔｉベクター（Ｏｘｆｏｒｄ　Ｇｅｎｅｔｉｃｓ）にクローニングした。ＨＬＡクラスＩＩ構築物をｐＳＦ　Ｌｅｎｔｉに同様にクローニングし、これは、Ｃ末端において融合した同じリンカー－ＢＡＰ配列を有し、その後に別の短いＧＳＧリンカー、Ｆ２Ａリボソームスキッピング配列、Ｃ末端ＨＡタグを有するα鎖の配列、終止コドン、および変動するＤＮＡバーコードが続くβ鎖配列からなった。ＭＡＰＴＡＣ（商標）構築物をＥｘｐｉ２９３細胞、ＨＥＫ２９３Ｔ細胞、Ａ３７５細胞、ＨｅＬａ細胞、ＫＧ－１細胞、Ｋ５６２細胞およびＢ７２１．２２１細胞にトランスフェクトまたは形質導入した。

【０７４５】

ＨＬＡ－ペプチド単離プロトコール

ＢＡＰタグが付されたＨＬＡを発現する細胞５０×１０^{6}個を含有する急速冷凍した細胞ペレットを氷上で２０分間にわたって解凍し、低温溶解緩衝剤１．２ｍＬ中に手動でピペッティングすることによって穏やかに溶解させた。ＤＮＡ、ＲＮＡ、および細胞デブリを一掃した後、上清を新しい１．５ｍＬチューブに移し、０．５６μＭのビオチン、１ｍＭのＡＴＰ、および３μＭのＢｉｒＡと一緒に室温で１０分間インキュベートすることにより、ＢＡＰタグが付されたＨＬＡをビオチン化した。ビオチン化溶解物を、ＮｅｕｔｒＡｖｉｄｉｎ樹脂２００μＬと一緒に４℃で３０分間インキュベートして、ビオチン化ＨＬＡ－ペプチド複合体を親和性富化させた。洗浄後、ＨＬＡが結合した樹脂を４℃、１，５００×ｇで１分間遠心分離することによってペレット化し、－８０℃で保存した、またはＳｅｐ－Ｐａｋ固相抽出を使用してすぐにＨＬＡ－ペプチドの溶出および脱塩に供した。健康なドナー材料の内因性ＨＬＡクラスＩＩリガンドームをプロファイリングするために、社内で生成した抗ＨＬＡ－ＤＲ抗体Ｌ２４３を使用して、または市販のＴＡＬ　１Ｂ５抗体を用いてＨＬＡ－ペプチド複合体を単離した。

【０７４６】

タンデム質量分析によるＨＬＡ－ペプチド配列決定

全てのｎａｎｏＬＣ－ＥＳＩ－ＭＳ／ＭＳ分析に同じＬＣ分離条件、機器パラメータ、およびデータ分析法を使用した。簡単に述べると、試料を、社内でＣ１８　Ｒｅｐｒｏｓｉｌビーズを充填し、６０℃で加熱したＰｉｃｏＦｒｉｔカラムに合わせたＰｒｏｘｅｏｎ　Ｅａｓｙ　ＮａｎｏＬＣ　１２００を使用してクロマトグラフィーにより分離した。データ依存性取得の間、溶出したペプチドを、Ｎａｎｏｓｐｒａｙ　Ｆｌｅｘ　Ｉｏｎ　ｓｏｕｒｃｅを備えたＯｒｂｉｔｒａｐ　Ｆｕｓｉｏｎ　Ｌｕｍｏｓ質量分析計に導入した。Ｓｐｅｃｔｒｕｍ　Ｍｉｌｌ　ｓｏｆｔｗａｒｅ　ｐａｃｋａｇｅ　ｖ６．０　ｐｒｅ－Ｒｅｌｅａｓｅを使用して質量スペクトルを解釈した。ＰＳＭ　ＦＤＲ推定値＜１％を通過し

た同定されたペプチドを、参照データベース内の２６４種の一般的な夾雑物タンパク質に割り当てられた全てのペプチドを除去することによって、および陰性対照ＭＡＰＴＡＣ（商標）親和性プルダウンにおいて同定されたペプチドを除去することによって夾雑物についてフィルタリングした。さらに、参照データベースのｉｎ　ｓｉｌｉｃｏトリプシン消化物にマッピングされた全てのペプチドを除去してトリプシン試料キャリーオーバーを説明した。生の質量分析データセットをＭａｓｓＩＶＥに受け入れられ次第寄託する（ｍａｓｓｉｖｅ．ｕｃｓｄ．ｅｄｕ）。

【０７４７】

結合性モチーフおよび結合予測のための機械学習手法

各対立遺伝子について、畳み込みニューラルネットワークのアンサンブルを、ＭＡＰＴＡＣ（商標）ペプチドとスクランブルデコイを区別するように訓練した。各ネットワークは、ＲｅＬＵ活性化畳み込み層を２つ含み、各々が５０の６ワイドフィルターを有する。層当たりフィルター当たりの最大活性化および平均活性化を経由してシグモイド活性化を用いる最終的な高密度層に送る。Ｌ２－ｎｏｒｍ、各畳み込み層後の２０％空間ドロップアウト、および早期停止によって正則化を実現し、対立遺伝子毎に重複していないペプチド（約１２．５％）のホールドアウト区分に従ってチューニングした。性能ベンチマーキングでは、ＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１による予測を各クエリペプチド内の最大スコア化１５ｍｅｒとして算出し、これは、ネイティブなＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１による予測よりも良好に一様に実施される手法である。

【０７４８】

ＣＤ４＋　Ｔ細胞誘導アッセイ

ＰＢＭＣを、１：１０の比で合計３回の刺激にわたってペプチドをパルスしたｍＤＣと共培養した。次いで、誘導されたＴ細胞を、以前に記載されている一意的な２色バーコードで標識し、ペプチドをパルスし、成熟させた自己ｍＤＣと１：１０の比で終夜培養した。その後、細胞を、ペプチドに応答したＩＦＮ－γの産生についてフローサイトメトリーによって評価した。ペプチドに対して正に応答した誘導試料は、ＩＦＮ－γ産生をペプチドなし対照よりも３％高く誘導した試料であった。

【０７４９】

ＳＩＬＡＣで標識された腫瘍細胞のＡＰＣによるエンドサイトーシス

Ｋ５６２細胞（ＡＴＣＣ、Ｍａｎａｓｓａｓ、ＶＡ）を、重同位元素アミノ酸、Ｌ－リシン２ＨＣｌ$^{13}C_6{}^{15}N_2$（Ｌｉｆｅ　Ｔｅｃｈｎｏｌｏｇｉｅｓ）およびＬ－ロイシン$^{13}C_6$（Ｌｉｆｅ　Ｔｅｃｈｎｏｌｏｇｉｅｓ）を含有するＳＩＬＡＣ用ＲＰＭＩ培地（ＴｈｅｒｍｏＦｉｓｈｅｒ）中で５倍加に成長させた。単球由来樹状細胞（ｍＤＣ）を、ＵＶ処理したＫ５６２細胞と一緒に終夜、またはＨＯＣｌ処理後に生成された溶解物と一緒に５時間にわたって、１：３の比で共培養した。プロテオミクス解析のために、細胞を回収し、ペレット化し、液体窒素で急速冷凍した。

【０７５０】

結果

ＭＡＰＴＡＣ（商標）：単一対立遺伝子ＨＬＡクラスＩＩリガンドプロファイリングのためのスケーラブルなプラットフォーム

ＨＬＡクラスＩＩ結合性モチーフに関する現行の知見は、主に２種の生化学的結合アッセイを使用して生成されたデータに基づく。そのような以前の手法の１つでは、アッセイペプチドおよび放射性標識した競合ペプチドを細胞質由来のＨＬＡ抽出物と共インキュベートしてＩＣ５０を決定する。別の手法では、ＥＣ５０を決定するために、コンフォメーション的に特異的な抗体によりアッセイペプチドと結合したＨＬＡの割合を測定する。これらのアッセイからのデータを免疫エピトープデータベース（ＩＥＤＢ）にコンパイルし、ＮｅｔＭＨＣＩＩｐａｎなどのＨＬＡクラスＩＩ予測アルゴリズムを訓練するために使用する。５種の最も一般的なコーカサス人ＨＬＡ－ＤＲＢ１対立遺伝子はＩＥＤＢでは十分に支持されるが（各々３３２６～８９６７ペプチド）、強力な結合体（親和性＜１００ｎＭ）であるのはたった約２９％であり、ＩＥＤＢペプチド全体の８５％はきっかり１５

ｍｅｒである（図１２Ｂ、図１２Ｅ）。ＨＬＡ－ＤＰおよびＨＬＡ－ＤＱ対立遺伝子および非コーカサス人ＨＬＡ－ＤＲ対立遺伝子（例えば、ＨＬＡ－ＤＲＢ１＊１５：０２）は相当少ないデータによって支持される。

【０７５１】

多様な患者集団を支持する対立遺伝子幅を有する高品質データセットを創出するために、ＭＳに基づく同定のための単一の対立遺伝子に結合するＨＬＡクラスＩＩペプチドの効率的な単離を可能にする技術であるＭＡＰＴＡＣ（商標）を開発した（図１１Ａ）。選択されたＨＬＡクラスＩＩヘテロ二量体のアルファ鎖およびベータ鎖は、ビオチン－アクセプターペプチド（ＢＡＰ）配列がベータ鎖のＣ末端に配置された遺伝子構築物によってコードされる。ＨＬＡ－ＤＲＡは機能的にインバリアントであるので、ＭＡＰＴＡＣ（商標）では、外因性ベータ鎖と内因性アルファ鎖の潜在的な対形成にかかわらず単一対立遺伝子ＨＬＡ－ＤＲ結果がもたらされる。機能的なアルファ鎖バリアントの限定されたセットを示すＨＬＡ－ＤＰおよびＨＬＡ－ＤＱに関しては、細胞株を、マッチするまたは発現されないアルファ鎖対立遺伝子を有するように選択する。重要なことに、この手法は、ＢＡＰタグをＨＬＡクラスＩ重鎖に付加すればＨＬＡクラスＩに対しても機能する。

【０７５２】

４８時間のトランスフェクションにより、ＭＡＰＴＡＣ（商標）構築物の頑強な発現が実現され（図１２Ｃ）、細胞表面提示は正常レベルであった（図３４Ａおよび３４Ｂ）。これは、７種の別個の細胞株（ｅｘｐｉ２９３、Ａ３７５、ＫＧ－１、Ｋ５６２、ＨｅＬａ、ＨＥＫ２９３、およびＢ７２．２２１）において、および４０種のＨＬＡクラスＩＩ対立遺伝子について実証され、３種の規準のＨＬＡクラスＩＩ遺伝子座：ＨＬＡ－ＤＲ、ＨＬＡ－ＤＰ、およびＨＬＡ－ＤＱの全てについてのデータがもたらされた。品質管理フィルターを通過した反復実験（細胞約５０００万個）当たりの独特のペプチド同定の平均数は対立遺伝子にわたって２３６種から２５８０種までにわたり（図２９）、中央値は１３１９ペプチドであった。ＨＬＡ－ＤＭ過剰発現、ならびにペプチドの還元およびアルキル化を含めたいくつかのプロセス変形形態を使用してデータの深さを増大させた。ほんの小さなパーセンテージのＭＳヒットが既知の夾雑物であるトリプシンペプチド、およびニセトランスフェクション（空のプラスミド）に対応した（図１１Ｂおよび図２９）。ＭＡＰＴＡＣ（商標）ＨＬＡクラスＩおよびＨＬＡクラスＩＩペプチドの長さ分布は抗体に基づくプルダウンを利用した以前のＭＳ試験で観察されたものとマッチした（図１１Ｃ）。

【０７５３】

ＭＡＰＴＡＣ（商標）ＨＬＡクラスＩＩペプチドの中で、大多数のアミノ酸が供給源プロテオーム頻度と一致するレベルで表された（図１２Ｄおよび図１２Ｆ）。例外にはＣ、Ｍ、およびＷが含まれ、これらはそれぞれ８５％、３４％、および４２％枯渇し、これはＨＬＡクラスＩＩペプチドの以前のＭＳに基づく試験と一致した。ＨＬＡクラスＩＩペプチドの還元およびアルキル化は、Ｃの頻度のほぼ３倍であったが、それでもプロテオームに関しては過少に表された（図１２Ｆ）。ＩＥＤＢからの対立遺伝子を釣り合わせた高親和性ペプチド（＜１００ｎＭ）ではＣ、Ｍ、およびＷの枯渇は観察されなかった。逆に、ＩＥＤＢ結合体は、ＩＥＤＢ非結合体（＞５０００ｎＭ）と比較してＤの枯渇（－３９％）およびＥの枯渇（－３７％）ならびにＭの富化（＋６５％）を示した。したがって、ＭＡＰＴＡＣ（商標）は、他の技術を用いて観察されたものと一致する定義された偏りを示す。

【０７５４】

ＭＡＰＴＡＣ（商標）によりＨＬＡクラスＩＩペプチド結合性モチーフが分解される

ＭＡＰＴＡＣ（商標）を使用して、対立遺伝子特異的ＨＬＡクラスＩＩ結合性モチーフを分解した。４０種のＨＬＡクラスＩＩ対立遺伝子をプロファイリングし、そのうち１５種は、非コーカサス人集団において一般的な対立遺伝子（ＤＲＢ１＊１２：０２、ＤＲＢ１＊１５：０３、およびＤＲＢ１＊０４：０７）を含め、以前に特徴付けられていないものであった（ＩＥＤＢにおいて１００ｎＭ未満の親和性を有する３０種未満のペプチド）。ＨＬＡクラスＩＩペプチドは結合溝の残基数より長い可能性があるので、各ペプチドの

いずれの部分がＨＬＡと相互作用するかはすぐには明らかにならない（「コア」）とそれに対して突出；しかし、結合性コアの分解は結合性モチーフの特徴付けに極めて重要である。結合性コアを同定するために、各ペプチドに対して結合レジスターを繰り返し指名し、結合性モチーフをペプチドにわたって再学習するために期待値最大化アルゴリズムを使用するツールＧｉｂｂｓＣｌｕｓｔｅｒ－２．０を使用してコンセンサス結合性コアに対するペプチドをアラインメントした。若干の例外を除いて、一般的なＨＬＡ－ＤＲ対立遺伝子に対する結合性コアモチーフはＩＥＤＢに基づくモチーフとの強力な一致を示した（図３５）。ＭＡＰＴＡＣ（商標）により観察されたペプチドは一般的な対立遺伝子に対して必ずしも強力なＮｅｔＭＨＣＩＩｐａｎスコアを示さなかったが（図３６Ａ）、それでも、ＮｅｔＭＨＣＩＩｐａｎによって不十分に予測された、観察された結合体は、非常に強力な測定された親和性を有することが示され（図３６Ｂ）、これにより、これらの知見が偽陽性である可能性が低いことが示される。とりわけ、ＭＡＰＴＡＣ（商標）モチーフは多数の細胞株にわたって常に安定していた（図３６Ｃ）。

【０７５５】

一般には、ＭＡＰＴＡＣ（商標）とＩＥＤＢは、アンカー位置（約４つの最も高度に保存された位置）における最も高頻度のアミノ酸については一致したが、ＭＡＰＴＡＣ（商標）モチーフは、一般に、より低いエントロピーを示した（配列ロゴの文字の高さが高いことによって顕在化する）。興味深いことに、細胞にＭＡＰＴＡＣ（商標）構築物およびＨＬＡ－ＤＭを同時トランスフェクトした場合、アンカー位置のエントロピーは大多数の対立遺伝子についてさらにはるかに低下した（図３０Ａおよび図３７Ａ）。これは１２種のＨＬＡ－ＤＲ対立遺伝子にわたって一貫して観察され、それにより、レパートリー「編集者」としてのＨＬＡ－ＤＭの行き渡っている効果が示され、ＨＬＡ－ＤＭおよび他のローディングシャペロンを欠く親和性アッセイに基づくモデルではｉｎ　ｖｉｖｏには当てはまらない結合規則を学習することができることが示唆される。ＨＬＡ－ＤＭの効果はＣＬＩＰペプチドの存在に関しても明らかであった。ＨＬＡ－ＤＭ同時トランスフェクションを伴わない場合、ＣＤ７４由来ペプチドが１０種のＨＬＡ－ＤＲ対立遺伝子において観察され、公知のＣＬＩＰバリアントとマッチした（図３７Ｃ）。一方、我々のＨＬＡ－ＤＭ同時トランスフェクション実験のいずれにおいてもＣＬＩＰペプチドは観察されなかった。

【０７５６】

ＨＬＡ－ＤＭの効果は解析したＨＬＡ－ＤＰ対立遺伝子については明らかでなかった（図３０Ａおよび図３７Ａ）。これは、以前には報告されていない普通でない正に荷電したＰ１アンカーの存在に関係する可能性がある。ＨＬＡ－ＤＭは結合したペプチドのＮ末端側に主に作用すると考えられており、したがって、疎水性Ｐ１アンカーを有するＨＬＡ－ＤＰモチーフもＨＬＡ－ＤＭの存在によって変更されなかったので、普通でないＰ１アンカーはＨＬＡ－ＤＭ非感受性のマーカーではなかった（図３７Ａ）。他方では、ＨＬＡ－ＤＱＢ１＊０６：０４／Ａ１＊０１：０２はＨＬＡ－ＤＭによって大いに影響を受けた（図３０Ａ）。ＨＬＡ－ＤＭ同時トランスフェクションを伴わない場合、この対立遺伝子の結合性モチーフは識別可能ではなく、それにより、一部のＨＬＡ－ＤＱ対立遺伝子へのシャペロンを伴わないローディングにより非生理的結合体の大部分がもたらされることが示される。

【０７５７】

公開された複対立遺伝子ＨＬＡクラスＩＩデータセットの利用可能性を考慮して、我々の対立遺伝子特異的ペプチドが有効に同定されたかどうかを、ｉｎ　ｓｉｌｉｃｏ逆畳み込み法を使用して調査した。いくつかのグループが複対立遺伝子ＨＬＡクラスＩデータからのＨＬＡクラスＩ対立遺伝子モチーフの逆畳み込みにおいて成功を示しているが、ＨＬＡクラスＩＩモチーフの逆畳み込みは、各ペプチドの結合性コアおよび対立遺伝子割り当ての両方を同時に分解する必要性によって複雑になる。ＨＬＡクラスＩＩ逆畳み込みの正確度を評価するために、汎ＤＲ抗体によってプロファイリングされた８つの試料（ＰＢＭＣおよび公開された細胞株）からＨＬＡ－ＤＲリガンドームを解析した。各データセット

について、試料の遺伝子型の各対立遺伝子とマッチする単一対立遺伝子データのうちの２０ペプチドをスパイクインした（１～２種のＤＲ１対立遺伝子プラス０～２種のＤＲ３／４／５対立遺伝子、ハプロタイプおよび接合性に応じて）。ＧｉｂｂｓＣｌｕｓｔｅｒツール（逆畳み込みに使用することもできる）を使用して、ペプチドを群に区分化し、スパイクインペプチドがそれらの既知の対立遺伝子起源に従って適正に共クラスター化されたかどうかを観察した。全ての場合において、ペプチドは多様なクラスターにわたって分布し、正しい供給源対立遺伝子との関連はわずかしか示されず（図３０Ｂ）、逆畳み込みに基づくＨＬＡクラスＩＩ訓練用データは大きな誤差を有する可能性が高いことが示される。

【０７５８】

逆畳み込みの不十分な性能を理解するために、単一対立遺伝子ＭＡＰＴＡＣ（商標）データを精査して、ＧｉｂｂｓＣｌｕｓｔｅｒのガイドポストとしての役割を果たし得る「明白な」アンカーの頻度を決定した。したがって、各ＨＬＡクラスＩＩ対立遺伝子について各アンカー位置（エントロピーが最も低い４つの位置）における明白なアミノ酸（頻度が＞１０％）を定義した。ペプチドのたった１０～２０％が４つのアンカー位置全てにおいて理想的な残基を示し、５０％もの多くが２つまたはそれよりも少ない明白なアンカーを示した（図３０Ｃ）。予測されたアンカーの大部分を示すペプチドの頻度が低いことを考慮すると、ペプチドの大部分が、純粋にコンピュータ計算に基づいて分類することが困難であることは驚くべきことではない。したがって、ＭＡＰＴＡＣ（商標）では、ｉｎｓｉｌｉｃｏ方法で解決することが重要な不明確性の主要な供給源に対処する。

【０７５９】

ＭＡＰＴＡＣ（商標）を使用してＨＬＡクラスＩ対立遺伝子のモチーフを定義することもできる。これには、結合プロファイルが以前に定義されていない対立遺伝子（例えば、Ｂ＊５２：０１、日本で一般的）が含まれた。以前に特徴付けられた対立遺伝子に関しては、親和性に基づく方法で導き出されたモチーフと以前の単一対立遺伝子ＭＳ試験に良好な対応があることが認められた。それにもかかわらず、モチーフを定義するために逆畳み込み法を使用する複対立遺伝子のＭＳに基づく試験に関してはいくらかの不一致が存在することが認められた（図３７Ｂ）。

【０７６０】

ＭＡＰＴＡＣ（商標）データに対して訓練されたアルゴリズムにより免疫原性が予測される

ＭＡＰＴＡＣ（商標）データにより正確度が改善されたＨＬＡクラスＩＩ結合予測器を生成することができるかどうかを検討した。ＨＬＡクラスＩＩペプチドのＨＬＡ結合性部分配列はＮ末端またはＣ末端に対して固定された位置にないので、学習アルゴリズムでは、各ペプチドについて異なる結合性コアの可能性を動的に検討しなければならない。この制約に対処するために、翻訳不変パターン認識に関して熟練していることからコンピュータビジョンの分野で成功している畳み込みニューラルネットワーク（ＣＮＮ）を使用した。各対立遺伝子について、ＣＮＮのアンサンブルを訓練し（図３１Ａ）、全体的な予測器を「ｎｅｏｎｍｈｃ２」と名付けた。

【０７６１】

ＭＳでは特にＣに対してある程度のアミノ酸残基の偏りが示されるという事実を説明するために、観察された結合体（ヒットと称される）の配列をランダムに並び替えることによって陰性訓練例（デコイと称される）を生成した。この手法には天然のタンパク質の配列特性を学習するというリスクがあるので、デコイをＨＬＡクラスＩＩリガンドのペプチド供給源遺伝子の観察されていない部分配列からランダムにサンプリングした。各対立遺伝子についての陽性適中率（ＰＰＶ）を算出するために、ｎ個のＭＳにより観察されたペプチドを、供給源遺伝子の同じセットからサンプリングした１９ｎ個の長さを釣り合わせたデコイと併せてスコア化し、各予測器のｎ個の上位に順位付けされたペプチド（すなわち、上位５％）を陽性とコールした。偽陽性の数と偽陰性の数が等しいので、この場合のＰＰＶは再現率と同一である。ヒット対デコイ比１：１９での陽性適中率（ＰＰＶ）の算出により、ｎｅｏｎｍｈｃ２ではＮｅｔＭＨＣＩＩｐａｎと比較して、ＭＡＰＴＡＣ（商

標）により観察されたペプチドの予測に関するＰＰＶが改善されたことが示された（図３１Ｂ；表１１）。

【表１１－１】

表 11

MHC クラス II 対立遺伝子	NetMHCIIpan の PPV	*neonmhc2* の PPV
DRB1*16:01	0.13	0.66
DRB1*15:01	0.17	0.61
DRB4*01:01	0.23	0.62
DPB1*02:01/DPA1*01:03	0.12	0.52
DRB1*11:04	0.22	0.60
DRB1*14:01	0.14	0.59
DRB1*13:03	0.08	0.58
DPB1*06:01/ DPA1*01:03	0.01	0.48
DRB3*03:01	0.21	0.57
DRB1*03:01	0.35	0.55
DRB1*01:01	0.34	0.56
DRB5*01:01	0.35	0.54
DRB1*01:02	0.27	0.54
DRB3*01:01	0.43	0.54
DPB1*01:01/DPA1*01:03	0.08	0.53
DRB1*07:01	0.29	0.54
DRB1*04:04	0.22	0.55
DRB1*11:01	0.17	0.52
DRB1*15:03	0.02	0.50
DPB1*04:01/DPA1*01:03	0.27	0.52
DPB1*04:02/DPA1*01:03	0.25	0.53
DRB1*15:02	0.17	0.50
DRB1*10:01	0.15	0.48
DRB1*08:02	0.21	0.50
DRB1*13:01	0.16	0.49
DRB1*04:05	0.13	0.50
DRB1*09:01	0.19	0.47
DQB1*06:02/DQA1*01:02	0.14	0.49
DRB1*11:02	0.07	0.46
DRB1*12:01	0.1	0.46
DRB1*04:01	0.27	0.45
DRB1*04:02	0.1	0.46
DRB1*04:03	0.18	0.45
DRB3*02:02	0.22	0.45
DRB1*08:04	0.1	0.44

【表１１－２】

DRB1*12:02	0.12	0.44
DPB1*03:01/DPA1*01:03	0.03	0.48
DRB1*04:07	0.21	0.40
DRB1*08:01	0.11	0.37
DQB1*06:04/DQA1*01:02	0.06	0.37
DRB1*03:02	0.31	0.29
DRB1*08:03	0.1	0.24
DRB1*13:02	0.06	0.22

【０７６２】
訓練用データセットのサイズを種々の程度までダウンサンプリングする飽和実験により、ｎｅｏｎｍｈｃ２の性能がデータにより限定されており、より多くのデータで改善されることが示唆される（図３８Ａｉ）。
【０７６３】
図３０Ｂにおける低忠実度のＨＬＡクラスＩＩ逆畳み込みの観察の解析により、同等の予測性能が単一対立遺伝子データなしでは実現することができていないことが示唆される。これを試験するために、逆畳み込みを使用して対立遺伝子特異的結合予測器を複対立遺伝子ＭＳデータに対して訓練する、最近公開されたコンピュータによるワークフロー（Barra et al., 2018）に従った。ＧｉｂｂｓＣｌｕｓｔｅｒロゴを１１種の複対立遺伝子試料（図３０Ｂと同じ試料）について検査することにより、多くのクラスター（１３／３２）が試料中に存在することが分かっている対立遺伝子といかなる類似点も有さない（図３８Ａｉｉ）ことが観察された。モチーフがどのように見えるはずであるかに関する既存の知見を使用して、合法的なクラスターのみ（図３８Ａｉｉにおいて印づけられている）を選択し、同じＣＮＮアーキテクチャを有する予測器を築いた。次いで、これらのモデルをｎｅｏｎｍｈｃ２と並行して真正な単一対立遺伝子データ（訓練に使用しなかったＭＡＰＴＡＣ（商標）データのホールドアウト区分）に対して評価した。逆畳み込みされた複対立遺伝子データに対して訓練されたモデルは通常ＮｅｔＭＨＣＩＩｐａｎにまさるが、ＭＡＰＴＡＣ（商標）で訓練されたｎｅｏｎｍｈｃ２には劣る（図３１Ｅ）。単一対立遺伝子データの優位性は、それぞれの訓練用データセットのサイズが同一になるようにＭＡＰＴＡＣ（商標）データセットをダウンサンプリングした場合であっても維持された。
【０７６４】
明らかな予測改善が非ＭＳデータに対して評価した場合にも保持されることを確実にするために、四量体によりガイドされるエピトープマッピング（ＴＧＥＭ）によって検出された対立遺伝子特異的ＣＤ４＋メモリーＴ細胞応答の大きなデータセットをキュレートした。とりわけ、これらの四量体データは、シャペロンを伴わないペプチド交換に依拠し、したがって、従来の親和性アッセイと同じ偏りを受けやすい（Archila and Kwok, 2017）。それにもかかわらず、ｎｅｏｎｍｈｃ２は、評価のために十分なデータ（少なくとも２０の陽性例）を有する全ての対立遺伝子についてＮｅｔＭＨＣＩＩｐａｎよりも優れた（図３１Ｃ）。ＮｅｔＭＨＣＩＩｐａｎの性能（ＰＰＶによって測定される）は変動し、ＤＲＢ１＊１５：０１については５％の低さまで降下し（対照的に、ｎｅｏｎｍｈｃ２の性能は３０％ＰＰＶを下回ることはなかった）、ｎｅｏｎｍｈｃ２の性能に近づいたのは、十分に試験されているＨＬＡ－ＤＲＢ１＊０１：０１を含めたたった２種の対立遺伝子だけであった。他方では、ｎｅｏｎｍｈｃ２では、２種の最も一般的なコーカサス人ＨＬＡ－ＤＲ対立遺伝子（ＤＲＢ１＊０７：０１およびＤＲＢ１＊１５：０１）を含めた他の全ての評価した対立遺伝子に関して確かな改善が示された。これらの結果から、ｎｅｏｎｍｈｃ２のＮｅｔＭＨＣＩＩｐａｎに対する予測改善をＭＳに基づくものではないベンチマークにおいて確証することができ、大多数の対立遺伝子にわたって拡大することがで

きることが示される。

【０７６５】

ｎｅｏｎｍｈｃ２の治療的関連性を評価するために、ｎｅｏｎｍｈｃ２により、ｅｘｖｉｖｏ誘導アッセイ（方法を参照されたい）においてＣＤ４＋　Ｔ細胞応答を引き出すことができる新抗原を同定することができるかどうかを決定した。ＩＥＤＢにおける多くの親和性アッセイにより確認された結合体を有する一般的な対立遺伝子（ＤＲＢ１＊０１：０１およびＤＲＢ１＊０７：０１によってのみ超えられる；図１２Ｅ）であるＤＲＢ１＊１１：０１に焦点を当て、Ｃａｎｃｅｒ　Ｇｅｎｏｍｅ　Ａｔｌａｓ（ＴＣＧＡ）により観察された新抗原配列のセットをスコア化し、ｎｅｏｎｍｈｃ２では優先された（予測の上位１％）がＮｅｔＭＨＣＩＩｐａｎによっては選択されなかった（予測の下位９０％）サブセットを選択した。このセットを、誘導材料中に存在する他のＨＬＡ－ＤＲ対立遺伝子に結合し得るペプチドを除去することによってさらに精緻化した。大多数のｎｅｏｎｍｈｃ２により選択されたペプチド（８／１２）により、ペプチドの再現率に応答して、ＩＦＮγ発現によって測定されるＣＤ４＋　Ｔ細胞応答がもたらされた（図３１Ｄ、図３８Ｂおよび図３８Ｃ）。これらの結果から、ＭＡＰＴＡＣ（商標）により訓練された予測器により、ＮｅｔＭＨＣＩＩｐａｎによって同定されない免疫原性ＨＬＡクラスＩＩ新抗原配列を同定することができることが実証される。

【０７６６】

プロフェッショナルＡＰＣは腫瘍微小環境に存在する優性ＨＬＡクラスＩＩである

ＨＬＡクラスＩＩ対立遺伝子特異的ペプチド結合優先性の特徴付けおよび予測の両方を可能にする技術を開発したので、ＨＬＡクラスＩＩがん抗原を生じる可能性が最も高いタンパク質配列の優先順位付けに極めて重要である抗原プロセシングに関するさらなる洞察を用いて結合予測改善を補完しようと試みた。ＴＭＥの状況でこれらの疑問に対処するために、腫瘍におけるＨＬＡクラスＩＩリガンドームを調査した単一細胞ＲＮＡ－Ｓｅｑおよび公開されたＭＳに基づく試験を含めた非ＭＡＰＴＡＣ（商標）データセットを解析した。微小環境においていずれの細胞型が治療的に標的化可能ながん抗原を提示する最も可能性が高いかを検討した。現在、がん抗原が、エンドサイトーシスによって取り込まれた腫瘍タンパク質を有するプロフェッショナルＡＰＣによって提示されるのか、または腫瘍細胞自体によって提示されるのかに関しては意見が一致していない。そのために、肺がん、頭頸部がん、結腸直腸がん、卵巣がん、および黒色腫をプロファイリングし、規準のＡＰＣ（マクロファージ、樹状細胞、およびＢ細胞）がＴＭＥ内の腫瘍細胞および他の間質細胞型よりもはるかに高いレベルのＨＬＡクラスＩＩを発現することが見いだされた５つの公開された単一細胞ＲＮＡ－Ｓｅｑデータセットにおいて ＨＬＡ－ＤＲＢ１発現を解析した。この観察は、多数の患者および腫瘍型にわたって一貫している（図１９Ａ）。腫瘍細胞はＴＭＥ内ではＡＰＣに数で勝り得るので、腫瘍細胞によるＨＬＡクラスＩＩ発現のレベルがより低いことには、それにもかかわらず、免疫学的関連性がある可能性がある。いかほどの全体的なＨＬＡクラスＩＩ発現が腫瘍細胞によってもたらされるか、対していかほどの全体的なＨＬＡクラスＩＩ発現が間質によってもたらされるかを評価するために、ＨＬＡクラスＩＩ特異的遺伝子に変異を有するＴＣＧＡ患者（ＣＩＩＴＡ、ＣＤ７４、およびＣＴＳＳ）を同定し、ＨＬＡ－ＤＲＢ１発現のどの画分が腫瘍に由来するか、対してどの画分が間質に由来するかを帰属させるために、体細胞バリアントを示したＲＮＡ－Ｓｅｑリードの画分を決定した（図１９Ｂ、方法を参照されたい）。１７種の別個の腫瘍型を表す１５３名の患者において同定された変異に基づいて、大多数のＨＬＡクラスＩＩ発現が非腫瘍細胞から生じると思われた。実際、患者の４５％パーセントでゼロの腫瘍由来ＨＬＡクラスＩＩ発現が示された。最高レベルのＴ細胞浸潤を有する患者に焦点を当て（上位１０％、以前に公開された１８遺伝子シグネチャー（Ayers et al., 2017）を使用して同定されたもの）、腫瘍のＨＬＡ－ＤＲ低発現はそれでも正常であると思われ、１６名の患者のうち３名のみが＞１０００ＴＰＭを発現した。免疫治療によりこの傾向が乱されるかどうかを探索するために、チェックポイント遮断応答性腫瘍型からの追加的な単一細胞ＲＮＡ－Ｓｅｑを解析し、処置の前後のＨＬＡ－ＤＲＢ１発現を評価した。１つの

確認された応答者を含んだ黒色腫コホートでは、治療前生検材料および治療後生検材料のどちらにおいても腫瘍細胞による一様に低いＨＬＡ－ＤＲＢ１発現が示された（図１９Ｃ）。抗ＰＤ１治療に対して５５％臨床応答率を示した基底細胞癌コホートは、同様に、時点にかかわらず、腫瘍細胞由来ＨＬＡ－ＤＲＢ１の低発現を示した（図１９Ｃ）。これらの結果により、大多数の腫瘍内ＨＬＡクラスＩＩ提示が主にプロフェッショナルＡＰＣによって駆動され、「ホット」ＴＭＥ条件ではこの一般的なパターンとの相違が保証されないことが示唆される。

【０７６７】

特定の遺伝子はＨＬＡクラスＩＩ提示経路へのアクセス特権を有する

腫瘍に存在するＡＰＣによって優先的に提示されるエピトープの供給源遺伝子、およびそれらが、オートファジーから生じるのかまたはエンドサイトーシスから生じるのかを決定するために、腫瘍組織を使用して実施された３つの公開されたＨＬＡクラスＩＩリガンドーム試験を解析した。

【０７６８】

まず、各遺伝子の観察の数がその長さと発現レベルの積に比例するはずであると仮定して、各遺伝子が腫瘍ＨＬＡクラスＩＩリガンドームにおいて表される程度を数量化した（図１８Ｂ）。ネイティブな組織では発現されないにもかかわらず、ヒト血漿において発現されるタンパク質、特にアルブミン、フィブリノーゲン、補体因子、アポリポタンパク質、およびトランスフェリンの明らかな富化が観察された。これらの同定がＨＬＡリガンドームにおける非特異的結合を表すことを懸念して、ｎｅｏｎｍｈｃ２結合スコアを、４つのＰＢＭＣ　ＨＬＡ－ＤＲリガンドーム（図３９Ａ）における血漿由来ペプチドについて評価した；ペプチドは強力な結合スコアを示し、これにより、それらがＨＬＡと結合したことが示唆される。腫瘍ＨＬＡクラスＩリガンドームデータでは血漿由来タンパク質は有意には富化されなかった（図３９Ｂ）。ＨＬＡクラスＩＩリガンドームにおける血漿遺伝子の富化は、ＡＰＣが組織血清由来の細胞外タンパク質を微飲作用によって「すること」と一致する。ＩＴＧＡＭ（１１×富化）、ＬＣＰ１（８×）、ＩＴＧＡＶ（６×）、およびＩＣＡＭ１（６×）などの白血球細胞接着に関与する遺伝子の追加的な富化も観察され、これにより、ＡＰＣがそれら自体の膜を積極的に再利用することが示唆される。卵巣がんＨＬＡクラスＩリガンドームにおいて富化されることが最近報告されたＭＵＣ１６は過剰には表されなかった。

【０７６９】

ＨＬＡクラスＩＩ抗原提示経路における遺伝子の偏りをさらに調べるために細胞内局在も考慮した。遺伝子を局在によって群分けした場合、分泌型遺伝子および膜遺伝子は遺伝子発現に基づいて予想されたよりも２倍多く表され、これにより、ＨＬＡクラスＩＩリガンドームを形づくることにおけるマクロピノサイトーシスの重要な役割が強調される。それにもかかわらず、ＨＬＡクラスＩＩペプチドの半分よりも多くは核および細胞質などのマクロピノサイトーシスと一致しない区画から生じる。これらの遺伝子の多くがオートファジーを介して提示される場合、プロテアソームによって取り除かれることが分かっている遺伝子の対応する欠乏が存在するはずであることは理にかなっていた。実際に、ユビキチン部位を含有することが分かっているタンパク質からそれらの長さおよび発現に基づいて予測されたペプチドが生成される頻度は低かった（図３２Ｃ）。プロテアソームが阻害されるとレベルが上昇することが分かっているタンパク質の枯渇も観察された。これらは、オートファジーに関しては予測されるが、ファゴサイトーシスに関しては必ずしも予測されないというパターンであり、これにより、ＡＰＣペプチドリガンドームがそれら自体の細胞内プロテオームを部分的に表すことが示唆される。

【０７７０】

ＴＭＥにおいてＡＰＣによって提示されるＨＬＡクラスＩＩ抗原の起源に対処するために、核および細胞質基質ペプチド同定がＡＰＣ特異的またはバルク腫瘍遺伝子発現プロファイルとより大きく一致するかどうかを決定することによって供給源遺伝子の起源の直接逆畳み込みを行うことが可能であり得るかどうかを検討した（図３９Ｃ）。推定値に著し

い不確実性が存在したが（回帰に基づくモデルおよびブートストラップリサンプリングによって評価した；補足的方法）、ＨＬＡクラスＩＩリガンドームは腫瘍およびＡＰＣ遺伝子発現プロファイルの両方の混合物によって最良に説明された。総合すると、プロテアソームにより取り除かれるタンパク質の枯渇が観察され、この結果により、腫瘍内ＡＰＣが外因性タンパク質および内因性タンパク質の両方の混合物を提示することが示唆される。

【０７７１】

一部の遺伝子領域は優先的にプロセシングされるが、明らかな切断モチーフを欠く

いずれの配列が抗原プロセシングのために好ましいかに関する理論が多数存在する（図３２Ｄ）。１つのモデルによると、ＨＬＡクラスＩの場合と同様に、供給源タンパク質はＨＬＡクラスＩＩに結合する前に酵素により切断される（Sercarz and Maverakis, 2003）。第２のモデルでは、まずペプチド結合が起こり、その後、結合したペプチドがエキソペプチダーゼ酵素により、ＨＬＡクラスＩＩによって立体的に妨げられるまでトリミングされることが提起される。第３のモデルでは、ペプチド切断事象はＨＬＡ結合の前後どちらでも起こる。ＨＬＡクラスＩＩペプチドがどのように生成されるかに関する競合モデルが存在するので、予測のための３つの異なるフレームワークを生成した（図３２Ｄ）。第１のモデルでは、エンドペプチダーゼが優位を占めると仮定する（「まず切断」）；第２のモデルでは、ＨＬＡクラスＩＩが全長タンパク質とかみ合い、それがその後、エキソペプチダーゼによって内向きにトリミングされると仮定する（「まず結合」）；および第３のモデルでは、酵素的消化がＨＬＡ結合の前後の両方で起こると提起する（「ハイブリッド」）。各モデルは異なるアルゴリズム手法を必要とする。具体的には、まず切断展望によって動機づけられるアルゴリズムは、ＭＳにより観察されたリガンドの端のアミノ酸モチーフに焦点を当てるべきであるが、まず結合展望によって動機づけられるアルゴリズムでは、これらのモチーフを無視し、ＨＬＡ結合接近可能性を規定する局所的なタンパク質構造特性に焦点を当てる方がよい。ハイブリッドモデルにより呼び起こされるアルゴリズムは、観察されたＨＬＡクラスＩＩペプチドの上流および下流を候補前駆体カット部分として調べるべきである。

【０７７２】

検討した３つの手法のうち、ベースラインモデルに対する測定可能な改善がもたらされたのはまず切断アルゴリズムだけであった（図３２Ｅおよび図４０Ｂ）。しかし、この手法では、クエリペプチドの正確なカット部分が遮蔽されている場合には値を加えることができないので、陽性例ペプチドに存在するエキソペプチダーゼトリミングの特質（例えば、最後から２番目のプロリンシグネチャー（Barra et al., 2018））が学習される（ＳＴＡＲ法）と思われる。

【０７７３】

純粋に経験的な手法を軸にして、公開されたＨＬＡ－ＤＱリガンドームにおいて観察されたタンパク質領域（Bergseng et al., 2015）を分類整理し、重複を使用してＨＬＡ－ＤＲリガンドを予測した。重複変数により予測性能のわずかな改善がもたらされた（ｎｅｏｎｍｈｃ２単独に対してＰＰＶの平均で３．１％の増大）（図３２Ｅ）。ＨＬＡ－ＤＱ対立遺伝子とＨＬＡ－ＤＲ対立遺伝子が、同じＨＬＡ－ＩＩプロセシング環境を共有するが結合性モチーフは共有しないと仮定して、この結果から、ある特定の遺伝子領域が実際にプロセシングに関して優先されるが、切断モチーフまたはコンフォメーション特性に明白には関係付けられないことが示される。

【０７７４】

ＭＳにより観察されたペプチドの観察された末端を使用して、「まず切断」モデルを仮定する手法であるプロセシングアルゴリズムを訓練することの陽性の結果が複数のグループにより報告されている。しかし、多数の別個の細胞株および組織型におけるペプチド末端に隣接するアミノ酸富化の精査において（図４０Ａ）、結合後トリミングとより大きく一致すると思われるパターンが観察された。これらには、ＨＬＡクラスＩＩプロセシング酵素、カテプシンＳの既知のモチーフとの対応の欠如、およびＰによりトリミング酵素の進行が遮断された場合に生じうるモチーフである、最後から２番目のペプチド位における

不十分に切断可能なＰの富化が含まれた。「まず切断」仮定が正しいかどうかを試験するために、ニューラルネットワークモデルをペプチド末端に対して訓練し、それらを２つの異なるやり方で評価した：ｉ）各ペプチドの正確なＮ末端およびＣ末端における切断性をスコア化する、またはｉｉ）各ペプチドの予測された結合性コアから±１５ＡＡの範囲内の最良の部位をスコア化する（補足的方法）。どちらの手法でも、まず切断モデルが正しい場合には予測値が足されるはずであることが仮定されたが、それがなされたのは第１の手法のみであった（図３２Ｅおよび図４０Ｂ）。したがって、ニューラルネットワークでは、リガンドの曝露特色（例えば、最後から２番目のＰ）に基づいてＨＬＡクラスＩＩをデコイと見分けることができるが、カット部分が先験的に分かっていない場合には、一次タンパク質配列から免疫原性ペプチドを予測する場合にはいつもそうである通り、見当違いになる。このわずかな違いには当該分野における錯乱を引き起こす潜在性がある。

【０７７５】

「まず結合」理論を用いると、ＭＳにより観察されたデコイペプチドを溶媒露出度について、ならびに内因的に不規則なドメインについてスコア化した。タンパク質構造によりＨＬＡ結合への利用可能性が規定される場合には溶媒露出性または不規則なドメインがＨＬＡクラスＩＩリガンドにおいて富化される可能性がある。しかし、これらの特色はまた、予測的ではないことも証明された（図３２Ｅ）。次いで、ペプチド結合の前に酵素によりタンパク質が部分的に消化され、その後、追加的なトリミングが起こるハイブリッドモデルを検討した。このモデルでは、ＭＳリガンドの観察された末端のさらに上流および下流に前駆体切断部位が存在する。したがって、ＣＮＮを、拡張されたタンパク質コンテキスト（±３０ＡＡ）に基づいて訓練して、前駆体カットに対応する遠位シグナルを検出した。このモデルでは、いずれの予測値も示されなかった（図３２Ｅ）。最後に、プロセシング優先領域は一次配列に基づいて予測することが難しいことが証明されているので、公開されたＨＬＡ－ＤＱリガンドームにおける観察されたタンパク質領域を分類整理し、重複を使用してＨＬＡ－ＤＲリガンドを予測した。重複変数により予測性能のわずかな改善がもたらされた（ｎｅｏｎｍｈｃ２単独に対してＰＰＶの平均で３．１％の増大）（図３２Ｅ）。ＨＬＡ－ＤＱとＨＬＡ－ＤＲ対立遺伝子が、同じＨＬＡクラスＩＩプロセシング環境を共有するが、ペプチド結合性モチーフは共有しないと仮定して、この結果から、ある特定の遺伝子領域が実際にプロセシングに関して優先されるが、明白な切断モチーフまたは特別なコンフォメーション特性は示さないことが示される。

【０７７６】

提示規則を組み込むことによりＨＬＡ－ＤＲリガンドーム予測が著しく増強される

結合規則がプロセシング関連特色とどのように協同するかを数量化するために、ＨＬＡクラスＩＩを提示する細胞株、樹状細胞、および健康なドナー末梢血単核細胞（ＰＢＭＣ）のＨＬＡ－ＤＲリガンドームを予測するための多変量モデルを創出した。提示されるペプチドは変異していないが、予測シナリオでは、ランダムにサンプリングされたゲノムの遺伝子座をそれらのＨＬＡクラスＩＩペプチドを産生する能力に関して評価しなければならない新抗原予測を模倣する。ヒットとデコイを１：４９９の比で使用し、タンパク質をコードするエクソームからデコイをランダムにサンプリングして、ｎｅｏｎｍｈｃ２に基づくモデルの性能およびＮｅｔＭＨＣＩＩｐａｎに基づくモデル、ならびに、ＲＮＡ－Ｓｅｑ由来発現、遺伝子レベルの偏り（図３２Ａに従って、関連する図３９Ｂを参照されたい）、および以前に観察されたＨＬＡ－ＤＱペプチドとの重複を含めた追加的なプロセシング特色が組み入れられたモデルの性能を評価した。変異腫瘍エピトープ標的の、がんの処置に関する優先順位付けのされ方と一致するモデルを作出するために、遺伝子レベルの偏り特色を、新抗原の供給源に関連しない血漿遺伝子の優先性が中和されるように改変した。

【０７７７】

これらの統合的アルゴリズムにより、結合予測およびプロセシング予測のどちらに関しても実質的な改善が確認された（図２１Ａ）。具体的には、完全モデルでは、評価されるデータセットに応じて、ＮｅｔＭＨＣＩＩｐａｎ結合予測を単独で使用したモデルに対し

て７．４×～６１×の変化倍率の改善が示された。発現および遺伝子の偏りのどちらによっても予測の正確度に対する実質的な独立した寄与がもたらされた。ＤＱ重複特色による寄与はより小さかったが、一貫して正の改善がもたらされた。重要なことに、親和性に基づくモデルの正確度は、プロセシング関連予測変数の十分な利益がもたらされた場合であっても、ＭＡＰＴＡＣ（商標）に基づくモデルの半分であった。

【０７７８】

プロフェッショナルＡＰＣによって提示される腫瘍由来ＨＬＡクラスＩＩペプチドを使用した予測の正確度のベンチマーキング

ＨＬＡクラスＩＩリガンドームの予測における我々の正確度を評価したので、プロフェッショナルＡＰＣによりエンドサイトーシスによって取り込まれた腫瘍由来のリガンドを予測することができるかどうかを試験することに注意をシフトした。ＴＭＥにおける大多数のＨＬＡクラスＩＩ発現がプロフェッショナルＡＰＣに由来するという我々の知見により、このプロセシング経路が腫瘍抗原提示に関して最も関連性がある経路の可能性が高いことが示される。残念ながら、腫瘍組織の従来のＭＳに基づくリガンドームでは、いずれのペプチドがエンドサイトーシスによって取り込まれた腫瘍タンパク質を起源とするものであるかは同定されない。したがって、ＳＩＬＡＣで標識された腫瘍細胞を「フィーディング」した樹状細胞（ＤＣ）のＨＬＡ－ＤＲリガンドームをプロファイリングする実験を考案した（図３３Ａ）。

【０７７９】

腫瘍由来タンパク質を標識するために、ＨＬＡクラスＩＩ欠損がん細胞株（Ｋ５６２）を同位体で標識されたＬおよびＫを含有する培地中で成長させ、９５％を超える標識効率を実現した。ＤＣに、溶解させた腫瘍細胞（腫瘍デブリのマクロピノサイトーシスを模倣するため）またはＵＶ処理した全腫瘍細胞（全細胞のファゴサイトーシスを模倣するため）のいずれかをフィーディングした。ＭＳを使用してＨＬＡ－ＤＲ結合性ペプチドをプロファイリングして、重標識または軽標識されたアミノ酸を有するペプチドを同定した。この実験では２９種の重標識されたペプチドがもたらされ、細胞全体実験では、溶解物実験およびＵＶ実験についてそれぞれ５６種の重標識されたペプチドがもたらされた（表１０（データＳ１Ｂ））。１つよりも多くのＬまたはＫを有するペプチドは２つの場合を除く全てにおいて完全な標識を示し、これにより、重標識されたペプチドが腫瘍細胞を起源とし、不調和標識を示す新しく翻訳されたＤＣタンパク質を起源とするものではないことが示される。無処理のＤＣ、および溶解物と一緒に１０分間インキュベートした後に回収したＤＣのどちらによっても重標識されたペプチドはもたらされなかった。

【０７８０】

本明細書に開示される統合された予測アルゴリズムを使用して、腫瘍由来ペプチドを予測する能力を評価した。天然のＨＬＡクラスＩＩリガンドームの予測に関する我々の以前の結果と一致して、ｎｅｏｎｍｈｃ２に基づくモデルではＮｅｔＭＨＣＩＩｐａｎに基づくモデルよりもはるかに高い予測の正確度が実現された（図３３Ｄ）。

【０７８１】

遺伝子発現とは異なり、遺伝子の偏りおよびＤＱ重複特色によりエンドサイトーシスによって取り込まれた抗原の予測は改善されず、これにより、バルク組織リガンドームから学習されたパターンがこのクラスのエピトープとは関連しないことが示唆される。重標識されたペプチドの供給源遺伝子を解析することにより、リガンドーム実験において見られた分泌型タンパク質および膜タンパク質の優性とは対照的に、ＲＮＡ結合性タンパク質（ＲＢＰ）、ＤＮＡ結合性タンパク質（ＤＢＰ）、熱ショックタンパク質（ＨＳＰ）およびミトコンドリアタンパク質（図２１Ｄ）が認められた（図３２Ａ）。これが別個のプロセシング優先性を表すものであるかどうかは不明であった。実際に、供給源タンパク質はＫ５６２において一般には高度に発現され（発現中央値４３０ＴＰＭ、これと比較して、標識されていないペプチドでは１３０ＴＰＭ）、これにより、観察される遺伝子優先性が検出限界により駆動される可能性が示唆される。

【０７８２】

明確にするために、ロジスティック回帰モデルを築いて、遺伝子の局在および機能的カテゴリーにより、遺伝子発現がすでに考慮されているモデルを超えてペプチド予測を改善することができるかどうかを試験した。結合および発現を考慮に入れた場合には、ＲＢＰ、ＤＢＰおよびＨＳＰはもはや有意ではなかったが、ミトコンドリアタンパク質は有意なままであった（ｐ＝２．６ｅ－４：図３３Ｅ）。とりわけ、富化のパターンは軽標識されたペプチドにおいて観察されたものとは完全に別個のものであった。

【０７８３】

ミトコンドリア富化により予測を改善することができるかどうかを決定するために、データを、より深いカバレッジを目的として、細胞入力を増加させること、ＵＶ処理プロトコールのみに焦点を当てること、および終夜時点に加えて２４時間のインキュベーション時点を追加することによって、新しいドナーから収集した。この実験では、終夜時点および２４時間時点についてそれぞれ７７種および５９種の重標識されたペプチドがもたらされ、共同で、７８種の独特の供給源遺伝子が同定された。ミトコンドリア優先性を考慮に入れたロジスティック回帰モデルを使用して（元のＳＩＬＡＣデータに対して訓練した）、結合および発現のみを含むモデルに対して８～１２％の正味の上昇でＰＰＶを改善することができた（図３３Ｇ）。これらの改善は有意であった（１６ｈおよび２４ｈについて、それぞれｐ＝１．１ｅ－９およびｐ＝１．５ｅ＝－８）。これらの優先性はバルクリガンドームから学習することはできておらず、より正確なエピトープ予測を可能にするために使用することができる。

【０７８４】

ＴＭＥにおけるＨＬＡクラスＩＩ提示の存在は、がん免疫治療を用いた処置を受けた患者における正の転帰に関連付けられている。残念ながら、ＨＬＡクラスＩＩリガンド予測の不正確性およびＴＭＥにおいて腫瘍抗原がどのように提示されるかの不明確性さにより、ＨＬＡクラスＩＩ抗原を標的とする治療の開発が遅れている。したがって、ＭＡＰＴＡＣ（商標）と称される単一対立遺伝子プロファイリング技術を本明細書に記載の通り開発し、腫瘍リガンドームを包括的に解析してＨＬＡクラスＩＩリガンドプロセシング規則を定義した。ＭＡＰＴＡＣ（商標）により、米国の患者の９５％を網羅する３５種のＨＬＡ－ＤＲＢ１対立遺伝子を含めた４０種のＨＬＡクラスＩＩ対立遺伝子の迅速なプロファイリングが可能になった。さらに、ＭＡＰＴＡＣ（商標）データに対して訓練された我々の結合予測アルゴリズムであるｎｅｏｎｍｈｃ２は、メモリーＣＤ４＋　Ｔ細胞応答の予測に関して、ＮｅｔＭＨＣＩＩｐａｎ訓練に利用可能な大多数の既存の親和性測定値を有する対立遺伝子に関してであっても、ＮｅｔＭＨＣＩＩｐａｎよりも優れた。ｎｅｏｎｍｈｃ２はＴＧＥＭ検証用データセットにおけるメモリーＣＤ４$^{+}$　Ｔ細胞応答の同定に関してＮｅｔＭＨＣＩＩｐａｎよりも性能が優れることが観察された。さらに、本明細書に開示されるアルゴリズムはｅｘ　ｖｉｖｏで誘導される新抗原に対するＣＤ４$^{+}$　Ｔ細胞応答の予測に関してもまさり、ＮｅｔＭＨＣＩＩｐａｎによって優先順位が付けられていなかった免疫原性ネオエピトープが首尾よく同定された。一方、単一細胞ＲＮＡ－Ｓｅｑ腫瘍データの解析により、最も関連性がある腫瘍抗原が、腫瘍細胞をファゴサイトーシスする浸潤性ＡＰＣによって優勢に発現される可能性が高いことが明らかになった。したがって、いずれの遺伝子および遺伝子領域がＴＭＥにおいて優先的に提示されるかを調査し、ＨＬＡ－ＤＲリガンドームおよびファゴサイトーシス性ＡＰＣによって提示される腫瘍由来のリガンドを正確に予測する多変量のモデルを創出した。これらのモデルはＮｅｔＭＨＣＩＩｐａｎの陽性適中率を著しく超えた。

【０７８５】

内因性にプロセシングされ、提示されたＨＬＡクラスＩＩリガンドを従来のペプチド結合アッセイとは対照的にＭＡＰＴＡＣ（商標）を使用して直接プロファイリングすることの利点は、ＨＬＡ－ＤＭなどのペプチドローディングシャペロンが存在することである。ＨＬＡ－ＤＭは、ＡＰＣのＨＬＡクラスＩＩペプチドレパートリーの編集において役割を果たすことが公知であり、そのことが、ＨＬＡ－ＤＭの示差的な発現のＨＬＡクラスＩＩリガンドームに対する影響を試験する動機づけになった。ＨＬＡ－ＤＲ　ＭＡＰＴＡＣ（

商標）実験においてＨＬＡ－ＤＭを過剰発現させたところ、結合性モチーフが、ＨＬＡ－ＤＭ過剰発現を伴わない実験よりも明白に分解された。驚いたことに、ＨＬＡ－ＤＭはＨＬＡ－ＤＱＢ１＊０６：０４／Ａ１＊０１：０２に対して極めて大きな影響を及ぼし、これにより、一部のＨＬＡ－ＤＱ対立遺伝子に関して正確なペプチド結合規則を学習することには、このペプチドローディングシャペロンの存在が必要になり得ることが実証される。逆に、２種のＨＬＡ－ＤＰ対立遺伝子では影響は示されず（Yin et al., 2015）、これにより、ＨＬＡ－ＤＭ感受性とＰ１アンカー優先性の関係がこれらの２種のＨＬＡ－ＤＰ対立遺伝子に対しては普通でないことが示唆される。ＨＬＡ－ＤＭを超えて、ＭＡＰＴＡＣ（商標）プラットフォームにより、他の重要なシャペロン、およびＣＤ７４またはＨＬＡ－ＤＯなどのＨＬＡクラスＩＩ経路に関与するタンパク質がＨＬＡクラスＩＩ対立遺伝子のペプチド結合レパートリーにどのように影響を及ぼし得るかを迅速に学習するやり方がもたらされる。

【０７８６】

腫瘍生物学に関して、我々の大多数の結果として起こる観察は、ＡＰＣが、評価された腫瘍型に関してＴＭＥにおける優性ＨＬＡクラスＩＩ発現を担うというものであった。これにより、治療的に関連性のある腫瘍抗原の提示がアポトーシス性腫瘍細胞のファゴサイトーシスまたは分泌された腫瘍タンパク質のマクロピノサイトーシスに依存する可能性が高いことが示唆される。直接ＣＤ４　Ｔ細胞死滅の報告も存在するが、もたらされたデータから、ＣＤ４　Ｔ細胞がより一般にはＴＭＥにおいて支持的役割を果たし、主に、浸潤性白血球上に提示される腫瘍抗原を認識することが示唆される。したがって、ＣＤ４　Ｔ細胞の抗腫瘍効果は、おそらく大部分が、直接細胞溶解機能を有するものを含めた他の免疫細胞の輸送および活性化を調節するケモカインおよびサイトカインの分泌によって媒介される。これは機構的により複雑になるので、利点の１つは、腫瘍が、ＨＬＡクラスＩＩ抗原が提示されるかどうかに関しての制御性が低いことであり、これにより、腫瘍がＨＬＡクラスＩ提示を回避する一般的な機構である機能喪失型変異を介した免疫エスケープがＨＬＡクラスＩＩではそれほど頻繁ではないことが示唆される。いずれのＡＰＣ集団がエンドサイトーシスによって取り込まれた腫瘍抗原の提示を担うか、およびこれらの食細胞の動員をＴＭＥに増強するやり方が存在するかどうかを慎重に定義する今後の試験が当該分野において有益になる。さらに、低酸素、化学療法、および放射線などの腫瘍細胞死滅の種々の方式により、これらのＡＰＣによる腫瘍抗原捕捉の種々のレベルがどのようにもたらされるかを理解することが有用になり、それにより、ＨＬＡクラスＩＩ標的化治療との最適な治療的組合せが導かれ得る。

【０７８７】

最後に、ＨＬＡクラスＩＩリガンドームの包括的な解析により、特定の遺伝子がそれらの転写物発現レベルにより予測されるよりも高頻度で提示されると思われるという観察が導かれた。腫瘍細胞由来の遺伝子レベルの偏りを学習することにより、ＡＰＣ　ＨＬＡクラスＩＩリガンドームの予測の改善が容易になったが、これらのシグナルのうちの一部は新抗原予測について関連性が低い可能性がある。例えば、外因性抗原の取り込みではなくＡＰＣにおけるオートファジーおよび膜再利用に関すると思われる富化が検出された。興味深いことに、ｉｎ　ｖｉｔｒｏにおいて腫瘍細胞が樹状細胞に「フィーディング」された場合、その代わりに、供給源遺伝子の同定によりＲＮＡ結合性タンパク質の富化が示された。そのような機構により、病原体エピトープの提示が促進され、全身性エリテマトーデスおよび他の自己免疫性状態において観察されるＲＮＡ結合性タンパク質に対する反応性が潜在的に説明されるので、ＲＮＡ結合性タンパク質が優先的に提示されると推測したくなる。いずれの場合でも、我々のＳＩＬＡＣに基づくＨＬＡリガンドミクスワークフローは、感染性疾患および自己免疫に関与する抗原の試験にも適用することができるので、その有用性は腫瘍抗原に限定されないことに留意することが重要である。

【０７８８】

要約すると、ＨＬＡクラスＩＩプロセシングおよび提示の規則は、ＨＬＡクラスＩよりも著しく複雑である。このような理由で、ＣＤ４＋　Ｔ細胞応答を駆動する抗原は多くの

場合に未定義のままになっている。ＨＬＡクラスＩＩ結合およびプロセシング規則を定義することに関する我々の進歩により、より有効な治療に変換することができる標的化可能ながん抗原および他の疾患関連エピトープを同定することが可能になる。

【０７８９】

（実施例１３）

補足情報

関連するメタデータを用いた実験およびデータ供給源の要約

ＭＡＰＴＡＣ（商標）データ、非ＭＡＰＴＡＣ（商標）原稿データ、および以前に公開されたデータを含めたデータセットの網羅的な一覧。適切な場合には、試料遺伝子型などの関連性のある付随する特色が提供される。Ｂ）実験的ＭＡＰＴＡＣ（商標）反復実験、ＰＢＭＣドナー、細胞株、およびＳＩＬＡＣフィーディング実験にわたってマージした独特のペプチド同定。夾雑物および完全なトリプシンペプチドを除去する。例えば、少なくとも、とりわけ図１２Ｅ～１２Ｆ、３４Ａ～３４Ｂを参照されたい。

【０７９０】

逆畳み込み解析のためのスパイクインペプチド

スパイクイン解析において対立遺伝子毎に使用した２０種のペプチド例の例示的な一覧。ペプチドを、最小ＳＰＩ、７０、長さ１２アミノ酸から２０アミノ酸の間を要求することによって、および他のＭＡＰＴＡＣ（商標）によりプロファイリングされたＤＲ対立遺伝子について観察されたあらゆる結合体との９ｍｅｒの重複を許容しないことによって選択した。さらに、所与の対立遺伝子について、９ｍｅｒを共有する２つのスパイクインペプチドは存在しない。例えば、少なくとも、とりわけ図３５、３６Ａ～３６Ｃを参照されたい。

【０７９１】

選択された対立遺伝子の補足実験手順のための照合されたＴＧＥＭデータセット

多様な病原体およびアレルゲンペプチドに関するＤＲＢ１＊０１：０１、ＤＲＢ１＊０３：０１、ＤＲＢ１＊０４：０１、ＤＲＢ１＊０７：０１、ＤＲＢ１＊１１：０１、およびＤＲＢ１＊１５：０１についてのＨＬＡクラスＩＩ四量体の結果ならびにそれらの対応するＮｅｔＭＨＣＩＩｐａｎ予測およびｎｅｏｎｍｈｃ２予測。データをＫｗｏｋおよび共同研究者によって公開された論文からキュレートした。とりわけ、図３８Ａ～３８Ｃを参照されたい。

【０７９２】

補足的方法

ＨＬＡクラスＩＩ対立遺伝子頻度および親和性データ統計値、図１２Ａおよび１２Ｅに関連する

対立遺伝子頻度を、リソース、ｂｉｏｉｎｆｏｒｍａｔｉｃｓ．ｂｅｔｈｅｍａｔｃｈｃｌｉｎｉｃａｌ．ｏｒｇ／ｈｌａ－ｒｅｓｏｕｒｃｅｓ／ｈａｐｌｏｔｙｐｅ－ｆｒｅｑｕｅｎｃｉｅｓ／ｈｉｇｈ－ｒｅｓｏｌｕｔｉｏｎ－ｈｌａ－ａｌｌｅｌｅｓ－ａｎｄ－ｈａｐｌｏｔｙｐｅｓ－ｉｎ－ｔｈｅ－ｕｓ－ｐｏｐｕｌａｔｉｏｎから得た。ｍｈｃ＿ｌｉｇａｎｄ＿ｆｕｌｌ．ｃｓｖデータセットを２０１８年９月２１日にＩＥＤＢデータ（ｉｅｄｂ．ｏｒｇ／ｄａｔａｂｓｅ＿ｅｘｐｏｒｔ＿ｖ３．ｐｈｐ）からダウンロードした。「細胞ＭＨＣ／競合／蛍光」、「細胞ＭＨＣ／競合／放射活性」、「細胞ＭＨＣ／直接／蛍光」、「精製されたＭＨＣ／競合／蛍光」、「精製されたＭＨＣ／競合／放射活性」、または「精製されたＭＨＣ／直接／蛍光」と同等の「方法／技法」および「解離定数ＫＤ」、「解離定数ＫＤ（およそのＥＣ５０）」、「解離定数ＫＤ（およそのＩＣ５０）」、「最大半量の有効濃度（ＥＣ５０）」、または「最大半数阻害濃度（ＩＣ５０）」と同等の「アッセイ群」を有するためには有効な親和性測定が必要であった。「著者」欄に列「Ｂｕｕｓ」が表示されている場合、測定はＳｏｒｅｎ Ｂｕｕｓグループ（Ｕｎｉｖｅｒｓｉｔｙ ｏｆ Ｃｏｐｅｎｈａｇｅｎ、Ｄｅｎｍａｒｋ）によるものである。他の場合では、著者欄に列「Ｓｅｔｔｅ」または「Ｓｉｄｎｅｙ」が含まれる場合、測定はＡｌｅｓｓａｎｄｒｏ Ｓｅｔｔｅ ｇｒｏｕｐ（Ｌａ Ｊｏｌｌａ Ｉｎｓｔｉｔｕｔｅ

ｆｏｒ Ｉｍｍｕｎｏｌｏｇｙ、Ｕ．Ｓ．Ａ）によるものである。他の全ての測定は「他」と標識した。強力な結合体を数え上げるために、測定された親和性が１００ｎＭよりも強力であるペプチドのみを計数した。

【０７９３】

ＭＡＰＴＡＣ（商標）プロトコールの概要、図２に関連する：ＤＮＡ構築物の設計

ＨＬＡクラスＩおよびＨＬＡクラスＩＩ対立遺伝子の遺伝子配列をＩＰＤ－ＩＭＧＴ／ＨＬＡウェブページ（ｅｂｉ．ａｃ．ｕｋ／ｉｐｄ／ｉｍｇｔ／ｈｌａ）によって同定し、組換え発現構築物の設計に使用した。ＨＬＡクラスＩに関しては、α鎖をＣ末端ＧＳＧＧＳＧＧＳＡＧＧリンカー（配列番号１０）と融合し、その後にビオチン－アクセプター－ペプチド（ＢＡＰ）タグ配列ＧＬＮＤＩＦＥＡＱＫＩＥＷＨＥ（配列番号１１）、終止コドン、および変動するＤＮＡバーコードを続け、ｐＳＦ Ｌｅｎｔｉベクター（Ｏｘｆｏｒｄ Ｇｅｎｅｔｉｃｓ、Ｏｘｆｏｒｄ、ＵＫ）にＮｃｏＩおよびＸｂａＩ制限部位を介してクローニングした。ＨＬＡクラスＩＩ構築物（ＤＲ、ＤＰおよびＤＱ）を同様にｐＳＦ ＬｅｎｔｉにＮｃｏＩおよびＸｂａＩ制限部位を介してクローニングし、これは、Ｃ末端においてＨＬＡクラスＩ構築物由来のリンカー－ＢＡＰ配列（ＳＧＧＳＧＧＳＡＧＧＬＮＤＩＦＥＡＱＫＩＥＷＨＥ（配列番号１２））、その後、別の短いＧＳＧリンカー、Ｆ２Ａリボソームスキッピング配列（ＶＫＱＴＬＮＦＤＬＬＫＬＡＧＤＶＥＳＮＰＧＰ（配列番号１３））、α鎖の配列、ＨＡタグ（ＧＳＹＰＹＤＶＰＤＹＡ（配列番号１４））、終止コドン、および変動するＤＮＡバーコードと融合したβ鎖配列からなるものであった。全てのＤＲ対立遺伝子について、ベータ鎖をＤＲＡ＊０１：０１と対形成させた。ＨＬＡ－ＤＭ構築物を、ＢＡＰ配列およびＨＡタグを欠いた以外は同様にＨＬＡクラスＩＩ構築物にクローニングした。ＨＬＡ－ＤＭをＨＬＡクラスＩＩ実験のサブセットに加えた。全てのＤＮＡ配列の同一性をサンガーシーケンシングによって検証した。

【０７９４】

細胞培養および一過性トランスフェクション

Ｅｘｐｉ２９３細胞（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）を、Ｅｘｐｉ２９３培地（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）中、８％CO_2、３７℃、オービタルシェーカーにおいて１２５ｒｐｍで成長させた。Ｅｘｐｉ２９３細胞を、定期的な隔週での継代を伴って、０．５×10^6個／ｍＬから６×10^6個／ｍＬの間の細胞密度で維持した。Ｅｘｐｉ２９３細胞懸濁液３０ｍＬを、およそ３×10^6個／ｍＬの細胞密度および＞９０％生存能力での一過性トランスフェクションのために使用した。簡単に述べると、ＤＮＡ３０μｇ（細胞懸濁液１ｍＬ当たりＤＮＡ １μｇ）を１つのチューブ中、Ｏｐｔｉ－ＭＥＭ培地（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）１．５ｍＬに希釈し、一方、ＥｘｐｉＦｅｃｔａｍｉｎｅ（商標）２９３トランスフェクション試薬（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）８０μＬをＯｐｔｉ－ＭＥＭ、１．５ｍＬを含有する第２のチューブ中に希釈した。これらの２つのチューブを室温で５分間インキュベートし、組み合わせ、穏やかに混合し、室温で３０分間インキュベートした。ＤＮＡとＥｘｐｉＦｅｃｔａｍｉｎｅの混合物をＥｘｐｉ２９３細胞に添加し、３７℃、８％CO_2、８０％相対湿度でインキュベートした。４８時間後、トランスフェクトされた細胞を４回の技術的反復実験においてチューブ当たり細胞５０×10^6個で回収し、遠心分離し、１×Ｇｉｂｃｏ ＤＰＢＳ（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）で１回洗浄し、質量分析のために液体窒素で急速冷凍した。細胞１×10^6個の一定分量を各トランスフェクションバッチから収集し、ウエスタンブロット分析を使用して抗ＢＡＰ（Ｒｏｃｋｌａｎｄ Ｉｍｍｕｎｏｃｈｅｍｉｃａｌｓ Ｉｎｃ．、Ｌｉｍｅｒｉｃｋ、ＰＡ）または抗ＨＡ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）によって解析して、親和性タグ付けされたＨＬＡタンパク質発現を検証した。Ｅｘｐｉ２９３の内因性ＨＬＡクラスＩＩ遺伝子型はＤＲＢ１＊１５：０１、ＤＲＢ１＊０１：０１、ＤＰＢ１＊０４：０２、ＤＰＡ１＊０１：０３、ＤＱＢ１＊０６：０２、ＤＱＡ１＊０１：０２であることが決定された（Ｌａｂｏｒａｔｏｒｙ Ｃｏｒｐｏｒａｔｉｏｎ ｏｆ Ａｍｅｒｉｃａ、Ｂｕｒｌｉｎｇｔｏｎ、ＮＣ）。一部の実験では、ＨＬＡクラスＩＩ対立遺伝子にＨＬＡ－ＤＭを同時トランスフェクトし、その場合、両

方のプラスミドに使用するＤＮＡ濃度を細胞懸濁液１ｍＬ当たりＤＮＡ　０．５μｇまで低下させた。

【０７９５】

Ａ３７５細胞（ＡＴＣＣ）を、１０％ＦＢＳを伴うＤＭＥＭ中で成長させ、定期的な継代を伴って、８０％以下の集密での培養で維持した。質量分析実験のために、Ａ３７５細胞を５００ｃｍ^{2}のプレートにおいて、７０％集密の細胞数から算出された通り、１００ｍＬ中、細胞１８．５×１０^{6}個／ｍＬの播種密度で培養した。２４時間後、細胞に、ＴｒａｎｓＩＴ－Ｘ２（Ｍｉｒｕｓ　Ｂｉｏ）を総培養体積に関して調整したＴｒａｎｓＩＴシステムプロトコールに従ってトランスフェクトした。４８時間後、細胞培地を吸引し、細胞を１×Ｇｉｂｃｏ　ＤＰＢＳ（Ｔｈｅｒｍｏ　Ｓｃｉｅｎｔｉｆｉｃ）で洗浄した。回収のために、Ａ３７５細胞を非酵素的細胞解離溶液（Ｓｉｇｍａ－Ａｌｄｒｉｃｈ）３０ｍＬと一緒に３７℃で１０分間インキュベートし、遠心分離し、１×ＤＰＢＳで洗浄し、試料当たり５０×１０^{6}個の細胞で分注した。２９３Ｔ細胞およびＨｅＬａ細胞をＡＴＣＣから購入し、それぞれＤＭＥＭ、１０％ＦＢＳ、２ｍＭのＬ－グルタミン中、またはＤＭＥＭ＋１０％ＦＢＳ中、３７℃、５％ＣＯ２で培養した。両方の細胞株にＨＬＡ構築物を、ＴｒａｎｓＩＴ　ＬＴ１試薬（Ｍｉｒｕｓ　Ｂｉｏ）を使用して製造者の指示に従ってトランスフェクトし、トランスフェクションの４８時間後に、Ａ３７５細胞について記載されているのと同様に処理した。全ての試料から、細胞１×１０^{6}個の一定分量を各トランスフェクションで収集し、抗ＢＡＰ（Ｒｏｃｋｌａｎｄ　Ｉｍｍｕｎｏｃｈｅｍｉｃａｌｓ　Ｉｎｃ．、Ｌｉｍｅｒｉｃｋ、ＰＡ）または抗ＨＡ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）ウエスタンブロットによって解析して、親和性タグ付けされたＨＬＡタンパク質発現を検証した。Ｂ７２１．２２１細胞をＦｒｅｄ　Ｈｕｔｃｈｉｓｏｎ　Ｃａｎｃｅｒ　Ｃｅｎｔｅｒ（Ｓｅａｔｔｌｅ、ＷＡ）から得、１０％熱失活ウシ胎仔血清プラス１％ペニシリン／ストレプトマイシン（どちらもＴｈｅｒｍｏ　Ｆｉｓｈｅｒ　Ｓｃｉｅｎｔｉｆｉｃ）を伴うＲＰＭＩ－１６４０プラスｇｌｕｔａｍａｘ（Ｔｈｅｒｍｏ　Ｆｉｓｈｅｒ　Ｓｃｉｅｎｔｉｆｉｃ）中で培養した。細胞を週２回培養し、２５回の継代後に廃棄した。Ｋ５６２細胞およびＫＧ－１細胞（ＡＴＣＣ、Ｍａｎａｓｓａｓ、ＶＡ）を、ＩＭＤＭ（Ｔｈｅｒｍｏ　Ｆｉｓｈｅｒ　Ｓｃｉｅｎｔｉｆｉｃ）培地プラス１０％熱失活ＦＢＳ、１％ペニシリン／ストレプトマイシン、１％ピルビン酸ナトリウム、および１％ＭＥＭ－ＮＥＡＡ中で成長させた。細胞を週２回培養し、２５回の継代後に廃棄した。

【０７９６】

Ｂ７２１．２２１細胞、ＫＧ－１細胞、およびＫ５６２細胞への形質導入のためのレンチウイルスを、８０％集密度まで成長させたＨＥＫ２９３Ｔ細胞において産生させた。ＨＬＡクラスＩまたはＨＬＡクラスＩＩをコードするゲノムベクターｐｓＦＬｅｎｔｉ（前の節に記載されている）６マイクログラムをレンチウイルスパッケージングベクターｐｓＰＡＸ２　５．３μｇおよびエンベロープベクターｐＭＤ．２　１．８１μｇと混合した。ＤＮＡをＯｐｔｉ－ＭＥＭ（Ｔｈｅｒｍｏ　Ｆｉｓｈｅｒ　Ｓｃｉｅｎｔｉｆｉｃ）およびトランスフェクション試薬Ｆｕｇｅｎｅ　ＨＤ（Ｐｒｏｍｅｇａ、Ｍａｄｉｓｏｎ、ＷＩ）と混合し、混合物を室温で１５分間インキュベートした。次いで、混合物をＨＥＫ２９３Ｔ細胞のディッシュに液滴で添加し、７２時間インキュベートした。次いで、上清を回収し、Ｌｅｎｔｉ－Ｘ　ＧｏＳｔｉｘ（Ｔａｋａｒａ　Ｂｉｏ　Ｉｎｃ．、Ｊａｐａｎ）を使用してレンチウイルス力価を試験した。形質導入のために、細胞を１２ウェル平底プレート（Ｃｏｒｎｉｎｇ　Ｉｎｃ．、Ｃｏｒｎｉｎｇ、ＮＹ）に播種し、６μｇ／ｍｌのポリブレン（Ｓｉｇｍａ－Ａｌｄｒｉｃｈ）を伴うレンチウイルス上清と混合した。レンチウイルスと混合した細胞を３２℃、８００×ｇで９０分間高速回転させた。細胞を温暖な培地中に再懸濁させ、３７℃のインキュベーター中、５％CO_2で７２時間インキュベートした。次いで、細胞を、１μｇ／ｍｌのピューロマイシンを２週間使用して選択した。選択後、少なくとも５０００万個の細胞を回収し、遠心分離し、１×Ｇｉｂｃｏ　ＤＰＢＳ（Ｔｈｅｒｍｏ　Ｓｃｉｅｎｔｉｆｉｃ）で１回洗浄し、質量分析のために液体窒素で急速冷凍した。

【０７９７】
ＢｉｒＡタンパク質の発現および精製
Ｃ末端ヘキサ－ヒスチジンタグ（配列番号１５）と融合したＥ．ｃｏｌｉ ＢｉｒＡをコードするｐＥＴ１９ベクターを使用した。化学的コンピテントＥ．ｃｏｌｉ ＢＬ２１（ＤＥ３）細胞（Ｎｅｗ Ｅｎｇｌａｎｄ Ｂｉｏｌａｂｓ）を、ＢｉｒＡ発現プラスミド（Ｃ末端ヘキサ－ヒスチジン（配列番号１５）と融合したＥ．ｃｏｌｉ ＢｉｒＡをコードするｐＥＴ１９ベクター）で形質転換し、ＬＢブロスプラス１００μｇ／ｍｌのアンピシリン中、３７℃で、$ＯＤ_{６００}$が０．６～０．８になるまで成長させ、３０℃まで冷却した後、０．４ｍＭのイソプロピル－β－Ｄ－チオガラクトピラノシドを添加することによって発現を誘導した。Ｅ．ｃｏｌｉ細胞の成長を３０℃で４時間継続した。Ｅ．ｃｏｌｉ細胞を４℃、８０００×ｇで３０分間遠心分離することによってＥ．ｃｏｌｉ細胞を回収し、使用するまで－８０℃で保存した。組換えＢｉｒＡを発現する凍結細胞ペレットを、５ｍＭのイミダゾールを伴うＩＭＡＣ緩衝剤（５０ｍＭの$ＮａＨ_{２}ＰＯ_{４}$、ｐＨ８．０、３００ｍＭのＮａＣｌ）に再懸濁し、１ｍｇ／ｍｌのリゾチームと一緒に氷上で２０分間インキュベートし、超音波処理によって溶解させた。４℃、１６，０００×ｇで３０分間遠心分離することによって細胞デブリおよび不溶性材料を除去した。清澄化した上清をその後、ＡＫＴＡ ｐｕｒｅ ｃｈｒｏｍａｔｏｇｒａｐｈｙ ｓｙｓｔｅｍ（ＧＥ Ｈｅａｌｔｈｃａｒｅ）を使用してＨｉｓＴｒａｐ ＨＰ ５ｍＬカラムにローディングし、ＩＭＡＣ緩衝剤プラス２５ｍＭおよび５０ｍＭのイミダゾールで洗浄した後、５００ｍＭのイミダゾールを用いて溶出させた。ＢｉｒＡを含有する画分をプールし、２０ｍＭのＴｒｉｓ－ＨＣｌ、ｐＨ８．０に対して２５ｍＭのＮａＣｌを用いて透析しＨｉＴｒａｐ Ｑ ＨＰ ５ｍＬカラム（ＧＥ Ｈｅａｌｔｈｃａｒｅ、Ｃｈｉｃａｇｏ、ＩＬ）にローディングし、２５ｍＭのＮａＣｌから６００ｍＭのＮａＣｌまでの直線勾配を適用することによって溶出した。高度に純粋なＢｉｒＡを含有する画分をプールし、緩衝剤を保存緩衝剤（２０ｍＭのＴｒｉｓ－ＨＣｌ、ｐＨ８．０、１００ｍＭのＮａＣｌ、５％グリセロール）に交換し、およそ５～１０ｍｇ／ｍＬに濃縮し、分注し、－８０℃での保存のために液体窒素で急速冷凍した。ＢｉｒＡタンパク質濃度を、ＵＶ分光法により、算出された吸光係数ε＝４７，４４０$Ｍ^{－１}ｃｍ^{－１}$を使用して$ＯＤ_{２８０}$ｎｍにおいて決定した。
【０７９８】
ウエスタンブロッティングプロトコール
試料をＸＴ Ｓａｍｐｌｅ ＢｕｆｆｅｒおよびＸＴ Ｒｅｄｕｃｉｎｇ Ａｇｅｎｔ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）に添加し、９５℃で５分間加熱し、次いで、約１００，０００個の細胞に対応する体積を１０％Ｃｒｉｔｅｒｉｏｎ ＸＴ Ｂｉｓ－Ｔｒｉｓゲル（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）にローディングし、Ｐｏｗｅｒ Ｐａｃ Ｂａｓｉｃ Ｐｏｗｅｒ Ｓｕｐｐｌｙ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）を使用し、ＸＴ ＭＥＳ Ｒｕｎｎｉｎｇ Ｂｕｆｆｅｒ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）を用いて２００Ｖで３５分間電気泳動した。ゲルを水で簡単にすすぎ、次いで、タンパク質を、Ｉｎｖｉｔｒｏｇｅｎ ｉＢｌｏｔ２ Ｇｅｌ Ｔｒａｎｓｆｅｒ Ｄｅｖｉｃｅ（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）で設定Ｐ３を使用してＩｎｖｉｔｒｏｇｅｎ ｉＢｌｏｔ Ｔｒａｎｓｆｅｒ Ｓｔａｃｋｓ（Ｔｈｅｒｍｏ Ｆｉｓｈｅｒ Ｓｃｉｅｎｔｉｆｉｃ）内のＰＶＤＦ膜に移した。Ｐｒｅｃｉｓｉｏｎ Ｐｌｕｓ Ｐｒｏｔｅｉｎ Ａｌｌ Ｂｌｕｅ Ｓｔａｎｄａｒｄ（Ｂｉｏ－Ｒａｄ、Ｈｅｒｃｕｌｅｓ、ＣＡ）を使用して分子量をモニタリングした。次に、膜を、Ｐｉｅｒｃｅ ＴＢＳ Ｔｗｅｅｎ２０緩衝剤［（ＴＢＳＴ）２５ｍＭのＴｒｉｓ、０．１５ｍＭのＮａＣｌ、０．０５％（ｖ／ｖ）Ｔｗｅｅｎ２０、ｐＨ７．５、Ｔｈｅｒｍｏ Ｆｉｓｈｅｒ Ｓｃｉｅｎｔｉｆｉｃ）］を用いて３×５分間洗浄し、ＴＢＳＴ－Ｍ［５％（ｗ／ｖ）無脂肪インスタントドライミルクを含有するＴＢＳＴ］中、室温で１時間ブロッキングし、次いで、ＴＢＳＴ－Ｂ［５％（ｗ／ｖ）Ｂｏｖｉｎｅ Ｓｅｒｕｍ Ａｌｂｕｍｉｎ（Ｓｉｇｍａ Ａｌｄｒｉｃｈ）を含有するＴＢＳＴ］ならびにウサギ抗ベータチューブリン抗体（カタログ番号ａｂ６０４６、Ａｂｃａｍ、Ｃａｍｂｒｉｄｇｅ、ＭＡ）およびウサギ抗ビオチンリガー

ゼエピトープタグ抗体（カタログ番号１００－４０１－Ｂ２１、Ｒｏｃｋｌａｎｄ Ｉｍｍｕｎｏｃｈｅｍｉｃａｌｓ、Ｌｉｍｅｒｉｃｋ、ＰＡ）の１：５，０００希釈物中、４℃で終夜インキュベートした。次に、膜をＴＢＳＴで３×５分間洗浄し、ヤギ抗ウサギＩｇＧ（Ｈ＋Ｌ－西洋ワサビペルオキシダーゼとコンジュゲートした抗体（カタログ番号１７０－６５１５、Ｂｉｏ－Ｒａｄ）の１：１０，０００希釈物を含有するＴＢＳＴ－Ｍ中、室温で１時間インキュベートし、次いで、ＴＢＳＴを用いて室温で３×５分間洗浄した。最後に、膜をＰｉｅｒｃｅ ＥＣＬ Ｗｅｓｔｅｒｎ Ｂｌｏｔｔｉｎｇ Ｓｕｂｓｔｒａｔｅ（Ｔｈｅｒｍｏ Ｆｉｓｈｅｒ Ｓｃｉｅｎｔｉｆｉｃ）に浸し、ＣｈｅｍｉＤｏｃ ＸＲＳ＋Ｉｍａｇｅｒ（Ｂｉｏ－Ｒａｄ）を使用して展開し、Ｉｍａｇｅ Ｌａｂ ｓｏｆｔｗａｒｅ（Ｂｉｏ－Ｒａｄ）を使用して可視化した。

【０７９９】

親和性タグ付けされたＨＬＡ－ペプチド複合体の単離

ＢＡＰタグが付されたＨＬＡ対立遺伝子を発現する細胞およびＢＡＰタグを伴わない内因性ＨＬＡ－ペプチド複合体のみを発現する陰性対照細胞株からの親和性タグ付けされたＨＬＡ－ペプチド複合体の単離を実施した。ＮｅｕｔｒＡｖｉｄｉｎビーズを付したアガロース樹脂を低温ＰＢＳ １ｍＬで３回洗浄した後、ＨＬＡ－ペプチドアフィニティー精製に使用した。ＢＡＰタグが付されたＨＬＡ分子を発現する細胞５０×１０^{6}個を含有する凍結ペレットを氷上で２０分間にわたって解凍し、低温溶解緩衝剤［２０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１００ｍＭのＮａＣｌ、６ｍＭのＭｇＣｌ$_{2}$、１．５％（ｖ／ｖ）Ｔｒｉｔｏｎ Ｘ－１００、６０ｍＭのオクチルグルコシド、０．２ｍＭの２－ヨードアセトアミド、１ｍＭのＥＤＴＡ、ｐＨ８、１ｍＭのＰＭＳＦ、１×完全ＥＤＴＡフリープロテアーゼ阻害剤カクテル（Ｒｏｃｈｅ）１．２ｍＬ中に手動でピペッティングすることによって穏やかに溶解させた。溶解物を、≧２５０単位のベンゾナーゼヌクレアーゼ（Ｓｉｇｍａ－Ａｌｄｒｉｃｈ）と一緒に４℃で１５分間、回転させながらインキュベートして、ＤＮＡ／ＲＮＡを分解し、４℃、１５，０００×ｇで２０分間遠心分離して細胞デブリおよび不溶性材料を除去した。清澄化した上清を新しいチューブに移し、０．５６μＭのビオチン、１ｍＭのＡＴＰ、および３μＭのＢｉｒＡを伴う１．５ｍＬチューブ中、室温で１０分間、回転させながらインキュベートすることにより、ＢＡＰタグが付されたＨＬＡ分子をビオチン化した。上清を、２００μＬに対応する体積のＰｉｅｒｃｅ ｈｉｇｈ－ｃａｐａｃｉｔｙ ＮｅｕｔｒＡｖｉｄｉｎ ｂｅａｄｅｄ ａｇａｒｏｓｅ ｒｅｓｉｎ（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）スラリーと一緒に４℃で３０分間、回転させながらインキュベートして、ビオチン化ＨＬＡ－ペプチド複合体を親和性富化した。最後に、ＨＬＡが結合した樹脂を低温洗浄緩衝剤（２０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１００ｍＭのＮａＣｌ、６０ｍＭのオクチルグルコシド、０．２ｍＭの２－ヨードアセトアミド、１ｍＭのＥＤＴＡ、ｐＨ８）１ｍＬで４回洗浄し、次いで、低温の１０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１ｍＬで４回洗浄した。洗浄間に、ＨＬＡが結合した樹脂を手動で穏やかに混合し、次いで、４℃、１，５００×ｇで１分間遠心分離することによってペレット化した。洗浄済みのＨＬＡが結合した樹脂を－８０℃で保存したかまたはすぐにＨＬＡ－ペプチドの溶出および脱塩に供した。

【０８００】

抗体に基づくＨＬＡ－ペプチド複合体の単離

ＨＬＡ ＤＲ－ペプチド複合体を健康なドナー末梢血単核細胞（ＰＢＭＣ）から単離した。７５μＬに対応する体積のＧａｍｍａＢｉｎｄ Ｐｌｕｓ Ｓｅｐｈａｒｏｓｅ樹脂を低温ＰＢＳ １ｍＬで３回洗浄し、抗体１０μｇと一緒に４℃で終夜、回転させながらインキュベートし、次いで、低温ＰＢＳ１ｍＬで３回洗浄した後、ＨＬＡ－ペプチド免疫沈降に使用した。細胞５０×１０^{6}個を含有する凍結ＰＢＭＣペレットを氷上で２０分間にわたって解凍し、低温溶解緩衝剤［２０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１００ｍＭのＮａＣｌ、６ｍＭのＭｇＣｌ２、１．５％（ｖ／ｖ）Ｔｒｉｔｏｎ Ｘ－１００、６０ｍＭのオクチルグルコシド、０．２ｍＭの２－ヨードアセトアミド、１ｍＭのＥＤＴＡ、ｐＨ８、１ｍＭのＰＭＳＦ、１×完全ＥＤＴＡフリープロテアーゼ阻害剤カクテル（Ｒｏｃｈ

ｅ）１．２ｍＬ中にピペッティングすることによって穏やかに溶解させた。溶解物を、≧２５０単位のベンゾナーゼヌクレアーゼ（Ｓｉｇｍａ－Ａｌｄｒｉｃｈ）と一緒に４℃で１５分間、回転させながらインキュベートして、ＤＮＡ／ＲＮＡを分解し、４℃、１５，０００×ｇで２０分間遠心分離して細胞デブリおよび不溶性材料を除去した。次いで、上清を、ＧａｍｍａＢｉｎｄ　Ｐｌｕｓ　Ｓｅｐｈａｒｏｓｅ樹脂（ＧＥ　Ｌｉｆｅ　Ｓｃｉｅｎｃｅｓ）に結合させた抗ＨＬＡ　ＤＲ抗体（ＴＡＬ　１Ｂ５、製品番号ｓｃ－５３３１９；Ｓａｎｔａ　Ｃｒｕｚ　Ｂｉｏｔｅｃｈｎｏｌｏｇｙ、Ｄａｌｌａｓ、ＴＸ）と一緒に４℃で３時間、回転させながらインキュベートしてＨＬＡ　ＤＲ－ペプチド複合体を免疫沈降させた。最後に、ＨＬＡが結合した樹脂を低温洗浄緩衝剤（２０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１００ｍＭのＮａＣｌ、６０ｍＭのオクチルグルコシド、０．２ｍＭの２－ヨードアセトアミド、１ｍＭのＥＤＴＡ、ｐＨ８）１ｍＬで４回洗浄し、次いで、低温の１０ｍＭのＴｒｉｓ－Ｃｌ、ｐＨ８、１ｍＬで４回洗浄した。洗浄間に、ＨＬＡが結合した樹脂を穏やかに混合し、次いで、４℃、１，５００×ｇで１分間遠心分離することによってペレット化した。洗浄済みのＨＬＡが結合した樹脂を－８０℃で保存したかまたはすぐにＨＬＡ－ペプチドの溶出および脱塩に供した。

【０８０１】

ＨＬＡ－ペプチドの溶出および脱塩

ＨＬＡ－ペプチドを親和性タグ付けされた内因性ＨＬＡ複合体から溶出させ、同時にＳｅｐ－Ｐａｋ（Ｗａｔｅｒｓ）固相抽出系を使用して脱塩した。簡単に述べると、Ｓｅｐ－Ｐａｋ　Ｖａｃ　１ｃｃ（５０ｍｇ）３７～５５μｍ粒子サイズｔＣ１８カートリッジを２４位置抽出マニホールド（Ｒｅｓｔｅｋ）に付着させ、ＭｅＯＨ、２００μＬで２回活性化し、その後、５０％（ｖ／ｖ）ＡＣＮ／１％（ｖ／ｖ）ＦＡ、１００μＬ、次いで、１％（ｖ／ｖ）ＦＡ、５００μＬで４回洗浄した。ＨＬＡ－ペプチドを親和性タグ付けされたＨＬＡ分子から解離させ、ｔＣ１８固相へのペプチド結合を容易にするために、３％（ｖ／ｖ）ＡＣＮ／５％（ｖ／ｖ）ＦＡ、４００μＬをＨＬＡが結合したビーズが付されたアガロース樹脂を含有するチューブに添加した。ピペッティングすることによってスラリーを混合し、次いで、Ｓｅｐ－Ｐａｋカートリッジに移した。チューブおよびピペットチップを１％（ｖ／ｖ）ＦＡ（２×２００μＬ）ですすぎ、すすぎ液をカートリッジに移した。Ｐｉｅｒｃｅ　Ｐｅｐｔｉｄｅ　Ｒｅｔｅｎｔｉｏｎ　Ｔｉｍｅ　Ｃａｌｉｂｒａｔｉｏｎ（ＰＲＴＣ）ｍｉｘｔｕｒｅ（Ｔｈｅｒｍｏ　Ｓｃｉｅｎｔｉｆｉｃ）１００ｆｍｏｌをローディング対照としてカートリッジに添加した。ビーズが付されたアガロース樹脂を１０％（ｖ／ｖ）ＡｃＯＨ、２００μＬと一緒に２回、５分間にわたってインキュベートして、ＨＬＡ－ペプチドを親和性タグ付けされたＨＬＡ分子からさらに解離させ、次いで、１％（ｖ／ｖ）ＦＡ、５００μＬで４回洗浄した。ＨＬＡ－ペプチドをｔＣ１８から溶出させて新しい１．５ｍＬマイクロチューブ（Ｓａｒｓｔｅｄｔ）に１５％（ｖ／ｖ）ＡＣＮ／１％（ｖ／ｖ）ＦＡ、２５０μＬ、その後、３０％（ｖ／ｖ）ＡＣＮ／１％（ｖ／ｖ）ＦＡ、２×２５０μＬを用いたステップ分画によって入れた。活性化、試料ローディング、洗浄、および溶出に使用した溶液は重力によって流出させたが、残りの溶出液をカートリッジから取り出すために真空（≦－２．５　ＰＳＩ）を使用した。ＨＬＡ－ペプチドを含有する溶出液を凍結させ、真空遠心分離によって乾燥させ、－８０℃で保存した後、第２の脱塩ワークフローに供した。以前に記載されている通りＥｍｐｏｒｅ　Ｃ１８固相抽出ディスク（３Ｍ、Ｓｔ．Ｐａｕｌ、ＭＮ）の１６ゲージの穿孔を２つ使用して充填した社内で築いたＳｔａｇｅＴｉｐｓを用いてＨＬＡ－ペプチド試料の二次脱塩を実施した。ＳｔａｇｅＴｉｐｓを、ＭｅＯＨ、１００μＬを用い、その後、５０％（ｖ／ｖ）ＡＣＮ／０．１％（ｖ／ｖ）ＦＡ、５０μＬを用いて、２回活性化し、次いで、１％（ｖ／ｖ）ＦＡ、１００μＬで３回洗浄した。乾燥ＨＬＡ－ペプチドを、３％（ｖ／ｖ）ＡＣＮ／５％（ｖ／ｖ）２００μＬを添加し、次いで、ＳｔａｇｅＴｉｐｓにローディングすることによって可溶化した。チューブおよびピペットチップを１％（ｖ／ｖ）ＦＡ（２×１００μＬ）ですすぎ、すすぎ体積をＳｔａｇｅＴｉｐｓに移し、次いで、ＳｔａｇｅＴｉｐｓを１％（ｖ／ｖ）ＦＡ、１００μＬで５回洗浄した。ペプチドを、１５％（ｖ／

ｖ）ＡＣＮ／１％（ｖ／ｖ）ＦＡ、２０μＬ、その後、３０％（ｖ／ｖ）ＡＣＮ／１％（ｖ／ｖ）ＦＡの２つの２０μＬのカットの段階的勾配を使用して溶出させた。試料ローディング、洗浄、および溶出は、卓上遠心分離機において１，５００～３，０００×ｇの最大スピードで実施した。溶出液を凍結させ、真空遠心分離によって乾燥させ、－８０℃で保存した。

【０８０２】

タンデム質量分析によるＨＬＡ－ペプチド配列決定

全てのｎａｎｏＬＣ－ＥＳＩ－ＭＳ／ＭＳ分析に同じ下記のＬＣ分離条件を使用した。試料を、ＰｉｃｏＦｒｉｔ（Ｎｅｗ Ｏｂｊｅｃｔｉｖｅ，Ｉｎｃ．、Ｗｏｂｕｒｎ、ＭＡ）７５μｍ内径キャピラリーにフィットさせたＰｒｏｘｅｏｎ Ｅａｓｙ ＮａｎｏＬＣ １２００（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ、Ｓａｎ Ｊｏｓｅ、ＣＡ）を使用してクロマトグラフィーにより分離した。１０μｍのエミッターに、１．９μｍ粒子サイズ／２００Å孔径のＣ１８ Ｒｅｐｒｏｓｉｌビーズを用い、１０００ｐｓｉの圧力でＨｅを約３０～４０ｃｍまで充填し（Ｄｒ．Ｍａｉｓｃｈ ＧｍｂＨ、Ａｍｍｅｒｂｕｃｈ、Ｇｅｒｍａｎｙ）、分離の間、６０℃で加熱した。カラムを１０×ベッド体積の緩衝剤Ａ［０．１％（ｖ／ｖ）ＦＡおよび３％（ｖ／ｖ）ＡＣＮ］で平衡化し、試料を３％（ｖ／ｖ）ＡＣＮ／５％（ｖ／ｖ）ＦＡ、４μＬにローディングし、７～３０％の緩衝剤Ｂ［０．１％（ｖ／ｖ）ＦＡおよび８０％（ｖ／ｖ）ＡＣＮ］を８２分間、３０～９０％の緩衝剤Ｂを６分間の直線勾配を用いてペプチドを溶出させ、次いで、９０％緩衝剤Ｂで１５分間保持してカラムを洗浄した。試料のサブセットを、６～４０％の緩衝剤Ｂを８４分間、４０～６０％の緩衝剤Ｂを９分間の直線勾配を用いて溶出させ、次いで、９０％緩衝剤Ｂで５分間および５０％緩衝剤Ｂで９分間保持してカラムを洗浄した。試料溶出のための直線勾配は毎分２５０ｎＬの速度で流し、約１３秒のピーク幅中央値がもたらされた。

【０８０３】

データ依存性取得の間、溶出したペプチドを、Ｎａｎｏｓｐｒａｙ Ｆｌｅｘ Ｉｏｎ ｓｏｕｒｃｅ（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ、Ｓａｎ Ｊｏｓｅ、ＣＡ）を備えたＯｒｂｉｔｒａｐ Ｆｕｓｉｏｎ Ｌｕｍｏｓ質量分析計（Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ、Ｓａｎ Ｊｏｓｅ、ＣＡ）に２．２－２．５ｋＶで導入した。フルスキャンＭＳを分解能６０，０００、３００ｍ／ｚから１，７００ｍ／ｚまで（ＡＧＣ標的４ｅ５、５０ｍｓ ｍａｘ ＩＴ）で取得した。各フルスキャンの後に、２秒間のサイクル時間、または、単離幅１．０ｍ／ｚ、衝突エネルギー３４（ＨＬＡクラスＩデータ）および３８（ＨＬＡクラスＩＩデータ）、ＡＣＧ標的５ｅ４、および最大充填時間２５０ｍｓ最大イオン時間を使用した分解能１５，０００でのデータ依存性ＭＳ２スキャンの上位１０を続けた。単離幅１．０ｍ／ｚを使用した理由は、ＨＬＡクラスＩＩペプチドはより長い傾向があり（＞４０アミノ酸のペプチドのサブセットを用いた中央値１６アミノ酸）、したがって、モノアイソトピックピークが常に同位体クラスターにおいて最も高いピークであり、質量分析計取得ソフトウェアにより、指定のオフセットの非存在下で最も高い同位体ピークが単離ウインドウの中央に置かれるとは限らないからである。したがって、１．０ｍ／ｚ単離ウインドウにより、同位体クラスターにおける最も高いピークではない場合であっても、ＨＬＡクラスＩＩペプチドの電荷の状態は多くの場合＋２またはそれよりも高いので、モノアイソトピックピークの共単離が可能になる。選択された前駆体当たり約３ＰＳＭを可能にするための繰り返し回数１および排除持続時間５秒間で動的排除が可能になった。ＨＬＡ－ペプチド同定はＰＳＭ品質に依拠し、したがって、電荷の状態が異なる多数のＰＳＭによりペプチド同定の信頼度がさらに増大するので、前駆体当たり単一の電荷の状態に対する依存的スキャンが無効の間は同位体を排除した。ＨＬＡクラスＩＩデータ収集のための電荷の状態のスクリーニングを、電荷の状態が１である（塩基性アンカー残基を有する対立遺伝子についてのみ）、＞７である、または割り当てられていない前駆イオンに対するＭＳ／ＭＳの誘発を防止するためのＰｅｐｔｉｄｅ Ｍｏｄｅを使用したモノアイソトピック前駆体選択（ＭＩＰＳ）と一緒に可能にした。ＨＬＡクラスＩに関してはデータ収集、電荷の状態が１である前駆イオン（質量範囲８００～１７００ｍ／ｚ）および

２～４である前駆イオンを選択し、一方、電荷の状態が＞４である前駆イオンおよび割り当てられていない前駆イオンは除外した。

【０８０４】

ＬＣ－ＭＳ／ＭＳデータの解釈、図２９に関連する

質量スペクトルを、Ｓｐｅｃｔｒｕｍ Ｍｉｌｌ ｓｏｆｔｗａｒｅ ｐａｃｋａｇｅ ｖ６．０ ｐｒｅ－Ｒｅｌｅａｓｅ（Ａｇｉｌｅｎｔ Ｔｅｃｈｎｏｌｏｇｉｅｓ、Ｓａｎｔａ Ｃｌａｒａ、ＣＡ）を使用して解釈した。ＭＳ／ＭＳスペクトルの前駆体ＭＨ＋が６００～２０００（ＨＬＡクラスＩ）／６００～４０００（ＨＬＡクラスＩＩ）の範囲内に入らない、前駆体電荷が＞５（ＨＬＡクラスＩ）／＞７（ＨＬＡクラスＩＩ）である、または検出ピークが最低＜５である場合にはそのＭＳ／ＭＳスペクトルを検索から除外した。同じクロマトグラフィーピークで取得された同じ前駆体ｍ／ｚを有する同様のスペクトルのマージは無効にした。ＭＳ／ＭＳスペクトルを、２６４種の一般的な夾雑物と組み合わせた、ゲノムのｈｇ１９アノテーションおよびそのタンパク質コード転写物を伴う全てのＵＣＳＣ Ｇｅｎｏｍｅ Ｂｒｏｗｓｅｒ遺伝子を含有するデータベース（６３，６９１種のエントリー；１０，９１７，８６７種の独特の９ｍｅｒのペプチド）に対して検索した。データベース検索の前に、全てのＭＳ／ＭＳを、配列タグの長さが＞２である、例えば、最小で３つの質量がアミノ酸の鎖内質量から分離されるスペクトル品質フィルターを通過させなければならなかった。最小の骨格切断スコア（ＢＣＳ）を５に設定し、ＥＳＩ ＱＥｘａｃｔｉｖｅ ＨＬＡｖ２スコア化スキームを使用した。還元もアルキル化もされていないネイティブなＨＬＡ－ペプチド試料からのスペクトルの全てを、酵素特異性なし、システイニル化として固定されたシステインの修飾を使用し、以下の可変の修飾：酸化メチオニン（ｍ）、ピログルタミン酸（Ｎ末端ｑ）、カルバミドメチル化（ｃ）を用いて検索した。還元およびアルキル化されたＨＬＡ－ペプチド試料を、酵素特異性なし、カルバミドメチル化として固定されたシステインの修飾を使用し、以下の可変の修飾：酸化メチオニン（ｍ）、ピログルタミン酸（Ｎ末端ｑ）、システイニル化（ｃ）を用いて検索した。前駆体質量許容誤差±１０ｐｐｍ、生成物質量許容誤差±１０ｐｐｍ、および最小のスコア化ピーク強度３０％をネイティブなＨＬＡ－ペプチドデータセットならびに還元およびアルキル化されたＨＬＡ－ペプチドデータセットの両方に対して使用した。個々のスペクトルについてのペプチドスペクトルマッチ（ＰＳＭ）を、Ｓｐｅｃｔｒｕｍ Ｍｉｌｌ自動検証モジュールを使用して標的－デコイのＰＳＭ順位でのＦＤＲ推定に基づいた利用に確信的に割り当てられたものと自動的に指定して、スコア化閾値基準を設定した。ＨＬＡ対立遺伝子に対して最小の配列長７、自動可変範囲前駆体質量フィルタリング、ならびに全てのＬＣ－ＭＳ／ＭＳ実行にわたって最適化されたスコアおよびデルタ順位１－順位２スコア閾値を使用した自動閾値戦略により、各前駆体電荷の状態についてＰＳＭ ＦＤＲ推定値＜１％がもたらされた。

【０８０５】

ＰＳＭ ＦＤＲ推定値＜１．０％を通過した同定されたペプチドを参照データベース内の２６４種の一般的な夾雑物タンパク質に割り当てられた全てのペプチドを除去することによって、および陰性対照ＭＡＰＴＡＣ（商標）親和性プルダウンにおいて同定されたペプチドを除去することによって、夾雑物に対してさらにフィルタリングした。さらに、参照データベースのｉｎ ｓｉｌｉｃｏトリプシン消化物にマッピングされたペプチドはｕＰＬＣカラムの試料キャリーオーバーからのトリプシン夾雑物として除外することができないので、これらのペプチド同定の全てを除去した。

【０８０６】

潜在的な偽陽性ＰＳＭ同定をＳＩＬＡＣ ＤＣフィーディング実験から除去するために、上記の方法を使用して同定されたＰＳＭに追加的な品質フィルターを適用した。高品質ＰＳＭのために、以下の閾値を使用してＦＤＲ＜１％を有する全てのペプチドをフィルタリングした：ｉ）スコア化ピーク強度＞６０％ｉｉ）骨格切断スコア≧８およびｉｉｉ）同じＬＣ－ＭＳ／ＭＳ反復実験における全てのＰＳＭ同定にわたって観察された中央値ｐｐｍからのｐｐｍ質量許容誤差±１ｐｐｍ。

【０８０７】
ＭＡＰＴＡＣ（商標）プロトコールを使用したＨＬＡ－ＤＲ、ＨＬＡ－ＤＱ、ＨＬＡ－ＤＰヘテロ二量体の単一対立遺伝子割り当て
ＭＡＰＴＡＣ（商標）プロトコールではＨＬＡクラスＩＩのベータ鎖のみにグ付けするので、ペプチド－ＭＨＣ複合体がノックインまたは内因性アルファ鎖を含有するかどうかにかかわらず、ペプチド－ＭＨＣ複合体をプルダウンステップにより単離する。ＨＬＡ－ＤＲの場合では、アルファ鎖の対立遺伝子バリエーションはペプチド結合に影響すると考えられない；したがって、内因性アルファ対形成を伴う対形成の相対的な程度はデータ解釈には関連しない、つまり、データは有効に単一対立遺伝子性である。しかし、ＨＬＡ－ＤＰおよびＨＬＡ－ＤＱ遺伝子座に関しては、アルファ鎖は、重要な対立遺伝子バリアントを示し、したがって、ノックインおよび内因性アルファ鎖対立遺伝子の両方が存在することにより、１～３種の別個の特異性の潜在性が生じる（細胞株が１つまたは２つのアルファ鎖対立遺伝子を有するかどうか、およびいずれかがノックイン対立遺伝子とマッチするかどうかに応じて）。この問題は、原理上は、ノックインアルファ鎖を用いて、および用いずにプロトコールを実行し、アルファを用いた実験に特異的なペプチドのセットを同定することによって軽減することができる。本明細書ではノックインアルファ対立遺伝子とマッチする単一のアルファ対立遺伝子を発現する細胞株を使用する手法をとった。
【０８０８】
以前に公開されたＭＳデータの解析、図１２Ａ～１２Ｆ、図３０Ａ～３０Ｃ、図３１Ａ～３１Ｄ、図２１Ａ、図３９Ａ～３９Ｂ、および図４０Ａ～Ｂに関連する
．ｒａｗファイルをもたらした、公開されたＬＣ－ＭＳ／ＭＳデータセットを、Ｓｐｅｃｔｒｕｍ Ｍｉｌｌ ｓｏｆｔｗａｒｅ ｐａｃｋａｇｅ ｖ６．０ ｐｒｅ－Ｒｅｌｅａｓｅ（Ａｇｉｌｅｎｔ Ｔｅｃｈｎｏｌｏｇｉｅｓ、Ｓａｎｔａ Ｃｌａｒａ、ＣＡ）を使用して再処理した。ＨＣＤ断片化ならびにオービトラップ（高分解能）へのＭＳおよびＭＳ／ＭＳデータ収集を利用したＴｈｅｒｍｏ Ｏｒｂｉｔｒａｐ機器（例えば、Ｖｅｌｏｓ、ＱＥｘａｃｔｉｖｅ、Ｆｕｓｉｏｎ、Ｌｕｍｏｓ）に収集されたデータセットを、上記の節「ＬＣ－ＭＳ／ＭＳデータの解釈」に記載されているパラメータを使用して解析した。ＣＩＤ断片化を利用したＭＳおよびＭＳ／ＭＳ高分解能データセットに対して、ＥＳＩオービトラップスコア化スキームを用い、上記と同じパラメータを使用した。オービトラップへのＭＳデータ収集およびイオントラップへのＭＳ／ＭＳデータ収集を用いたデータセットに対しても同様に上記と同じ以下のパラメータを使用し、以下の偏差を用いた。ＨＣＤデータに対してはＥＳＩ ＱＥｘａｃｔｉｖｅ ＨＬＡｖ２スコア化スキームを使用し、一方ＣＩＤデータに対してはＥＳＩオービトラップスコア化スキームを使用した。前駆体質量許容誤差±１０ｐｐｍ、生成物質量許容誤差±０．５Ｄａを使用した。高分解能ＭＳ／ＭＳデータセットおよび低分解能ＭＳ／ＭＳデータセットの両方に対して、個々のスペクトルについてのペプチドスペクトルマッチ（ＰＳＭ）を、Ｓｐｅｃｔｒｕｍ Ｍｉｌｌ自動検証モジュールを使用して標的－デコイのＰＳＭ順位でのＦＤＲ推定に基づいた利用に確信的に割り当てられたものと自動的に指定して、スコア化閾値基準を設定した。ＨＬＡ対立遺伝子に対して最小の配列長７、自動可変範囲前駆体質量フィルタリング、ならびに全てのＬＣ－ＭＳ／ＭＳ実行にわたって最適化されたスコアおよびデルタ順位１－順位２スコア閾値を使用した自動閾値戦略により、各前駆体電荷の状態についてＰＳＭＦＤＲ推定値＜１．０％がもたらされた。いくつかの以前に公開されたデータからのペプチド同定の解析により、高い率の９ｍｅｒ（＞１０％）が明らかになった。これらは潜在的に夾雑したＨＬＡクラスＩリガンドを表し得るので、短いペプチド（長さ＜１２）を全ての外部データセットから外した。
【０８０９】
ペプチドの遺伝子および「入れ子セット」へのマッピング、図３０Ａ～３０Ｃ、図３１Ａ～３１Ｄ、および図３２Ａ～３２Ｅに関連する
各ペプチドを、ＵＣＳＣ ｈｇ１９遺伝子アノテーション内の１つまたは複数のタンパク質をコードする転写物に割り当てた（ｇｅｎｏｍｅ．ｕｃｓｃ．ｅｄｕ／ｃｇｉ－ｂｉ

ｎ／ｈｇＴａｂｌｅｓ）。多くのペプチド同定は他と重複し、したがって、大部分が重複する情報を構成するので、ペプチドを、各々が約１つの独特の結合事象に対応することが意図されている「入れ子セット」に群分けした。例えば、ペプチドＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤＶ（配列番号１６）、ＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１７）、およびＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤＶ（配列番号１８）は全て、保存された配列ＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１９）を含有し、おそらく全てが同じレジスター内のＭＨＣに結合する。所与のデータセットのペプチドを入れ子状にするために、各ノードが独特のペプチドに対応するグラフを築き、少なくとも１つの９ｍｅｒを共有し、少なくとも１つの共通の転写物にマッピング可能なペプチドの任意の対間にエッジを創出した。Ｒパッケージｉｇｒａｐｈ内のクラスターコマンド（Team, 2014）（ｃｒａｎ．ｒ－ｐｒｏｊｅｃｔ．ｏｒｇ／ｗｅｂ／ｐａｃｋａｇｅｓ／ｉｇｒａｐｈ／ｃｉｔａｔｉｏｎ．ｈｔｍｌ）を使用して、接続したノードのクラスターを同定し、各クラスターを入れ子セットと定義した。この手順により、エッジ基準（共通の９ｍｅｒが１つ以上、共通の転写物が１つ以上）を満たす任意の２つのペプチドが同じ入れ子セット内に入ることが保証される。入れ子を配列ロゴ生成のために使用した（各入れ子セット内の最も短いペプチドを使用してロゴを生成した；図３０Ａ～３０Ｃ、機械学習（合計１になる、入れ子セット内のペプチドにわたる重要性の重み付け；図３１Ａ～３１Ｄ）、および遺伝子の偏り解析（各入れ子セットを個々のペプチドではなく１つの観察として計数した；図３２Ａ～３２Ｅ）。

【０８１０】

アミノ酸頻度の解析、図１２Ｆに関連する

ヒトプロテオームにおけるアミノ酸頻度を、ＵＣＳＣ　ｈｇ１９アノテーション内のタンパク質をコードする遺伝子全ての配列に基づいて算出した（多数の転写物アイソフォームによって表される遺伝子に対して１つの転写物をランダムに選択する）。ＩＥＤＢ頻度を、少なくとも１つの親和性観察≦１００ｎＭを有するペプチドの独特のセットを同定することによって決定した（Ｃ末端に六価ポリヒスチジンを有するペプチドは除外する）。ＭＡＰＴＡＣ（商標）頻度をまず、入れ子セット当たりただ１つのペプチド（最長のもの）を使用し、５種のＤＲＢ１対立遺伝子（ＤＲＢ１＊０１：０１、ＤＲＢ１＊０３：０１、ＤＲＢ１＊０９：０１、およびＤＲＢ１＊１１：０１）にわたって標準の順相プロトコールに関して検討した。さらに、還元およびアルキル化プロトコールによってプロセシングされた試料のサブセットについてＭＡＰＴＡＣ（商標）頻度を別々に算出した。外部データセットからのＭＳデータを、潜在的対立遺伝子起源を考慮せずに、同様に入れ子セット当たり最長ペプチドを使用して解析した。

【０８１１】

ＨＬＡクラスＩ配列ロゴを築くこと、図３７Ｂに関連する

各ＨＬＡクラスＩ対立遺伝子について、対応するペプチドの最初の５つの位置（ロゴ位置１～５にマッピングされる）および最後の４つの位置（ロゴ位置６～９にマッピングされる）におけるアミノ酸頻度をプロファイリングすることによって長さ９の配列ロゴを創出した。このように、ペプチドは、長さにかかわらず配列ロゴに寄与した。ＨＬＡクラスＩＩロゴの場合と同様に、文字の高さは各位置における各アミノ酸の頻度に比例し、頻度が≧１０％のアミノ酸について色分けを使用した。

【０８１２】

ＨＬＡクラスＩＩペプチドの逆畳み込みの性能の評価、図３０Ｂに関連する

ＧｉｂｂｓＣｌｕｓｔｅｒ（ｖ２．０）ツールの、複対立遺伝子ＨＬＡクラスＩＩペプチドデータを対立遺伝子起源毎にクラスター化する能力を評価するために、ＰＢＭＣ試料４つ、黒色腫細胞株（Ａ３７５）１つ、および以前に公開されたリンパ芽球様細胞株３つを含む８つの試料に対するその性能を解析した。各試料遺伝子型に存在する各ＤＲＢ１／３／４／５対立遺伝子について、我々の単一対立遺伝子ＭＡＰＴＡＣ（商標）データからの２０種のペプチドをスパイクインした。スパイクされたペプチドはＳＰＩ≧７０の１２～２０ｍｅｒに制限し、他のＨＬＡ－ＤＲ対立遺伝子についてのＭＡＰＴＡＣ（商標）デ

ータ内のいずれのペプチドとも９ｍｅｒを共有しなかった、または目的の対立遺伝子についてのスパイクされたペプチドのいずれとも９ｍｅｒを共有しなかった。次いで、これらの強化されたデータセットをＧｉｂｂｓＣｌｕｓｔｅｒ－ｖ２．０に、逆畳み込みのために、その他は以前から有しているのと同様に１位における疎水性優先性を強制した以外はデフォルトＨＬＡクラスＩＩ設定を使用してサブミットした。各試料について、溶液中のクラスターの数を手動で指定し、遺伝子型に存在するＨＬＡ－ＤＲ対立遺伝子の数と等しくなるように設定した。

【０８１３】

好ましいアンカー残基を有するペプチドの分率の算出、図３０Ｃに関連する

アンカー位置を、最も低いエントロピーを有する４つの位置と定義し、それらの位置内の「優先される」アミノ酸には、≧１０％の頻度を有するアミノ酸全てが含まれた。ｎ位に優先されるアミノ酸を有するペプチドの分率を算出する場合、入れ子セット当たりただ１つのペプチドを使用した（最も短いもの）。

【０８１４】

ＭＳにより観察されたペプチドの予測される親和性、図３６Ａに関連する

各ＨＬＡクラスＩＩ対立遺伝子について、１４から１７までの独特のペプチド長を全て同定し、結合潜在性についてＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１を使用してスコア化した。比較のために、５０，０００種のランダムな長さを釣り合わせたペプチドをヒトプロテオームからサンプリングした。密度分間布を対数変換された値に基づいて決定した。

【０８１５】

ＭＳにより観察されたペプチドについての測定された親和性、図３６Ｂに関連する

ペプチドを、予測されるＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１結合親和性が不十分なものである場合（ＤＲＢ１＊０１：０１については＞１００ｎＭまたはＤＲＢ１＊１１：０１については＞５００ｎＭ）、または≦２のヒューリスティックに定義されたアンカーを示す場合、親和性測定値について選択した。

【０８１６】

交差検証区分の確立、図３１Ａに関連する

各ノードがタンパク質をコードする転写物を表し、少なくとも５つの独特の９ｍｅｒのアミノ配列内容を共有する転写物の全ての対の間にエッジが存在する（ＵＣＳＣ　ｈｇ１９遺伝子アノテーション）グラフを創出した。Ｒパッケージｉｇｒａｐｈ内のクラスターコマンド（Team, 2014）（ｃｒａｎ．ｒ－ｐｒｏｊｅｃｔ．ｏｒｇ／ｗｅｂ／ｐａｃｋａｇｅｓ／ｉｇｒａｐｈ／ｃｉｔａｔｉｏｎ．ｈｔｍｌ）を使用して、接続したノードのクラスターを同定し、各クラスターを「転写物群」と定義した。このように、２つの転写物がエッジを共有する場合（共有される９ｍｅｒが５つ以上）、同じ転写物群に入ることが保証された。転写物群をランダムにサンプリングし、プロテオームを８つの大まかに同等のサイズにした区分に分割した。ＭＳにより観察されたペプチド（および観察されていないデコイペプチド）をそれらの供給源転写物の区分に従って区分に入れ、これらの区分を交差検証およびハイパーパラメータチューニングのために使用した。プロテオームの区分化のグラフに基づく手法を使用して、訓練および評価中に同様のペプチド配列が現れ、それにより予測性能が人為的に膨張し得る可能性を最小限にした。

【０８１７】

ＣＮＮに基づくＨＬＡクラスＩＩ結合予測器、ｎｅｏｎｍｈｃ２のアーキテクチャおよび訓練、図３１Ａに関連する

訓練のために陰性例（デコイ）をヒットペプチドの配列をランダムにシャッフリングすることによって生成した。一般的なアミノ酸優先性をもたらす可能性があるＭＳの偏りを排除するために、プロテオームから観察されていない領域を選択するのではなく、このデコイ生成方法を選択した。このように、我々の結合予測器では、例えば、システインの相対的枯渇を知らない（図１２Ｆ）。同様に、これにより、我々のモデルが、全体的な疎水性などのペプチドの全体的特性に関連するＭＳの偏りを学習することが防止される。この方法は、図３１Ａに示されている結果に関連する。

【０８１８】
モデルを２つの適用シナリオについて訓練した：内部ＭＡＰＴＡＣ（商標）データに対する検証（図３１Ｂ）および外部データに対する検証（図３１Ｃ、図２１Ａ、および図２１Ｂ）。モデルを前者について訓練した場合、６つのデータの区分（訓練区分）を使用してネットワーク重み最適化を学習し、第７の区分（チューニング区分）を使用してハイパーパラメータ最適化および早期停止を実施し、モデル設計を完成させた後に第８の区分（評価区分）に対して最終的な検証を実施する、単純な訓練手順を採用した。外部検証の場合では、交差検証を使用し、データの各区分について、ハイパーパラメータチューニングおよび早期停止のためにその区分をホールドアウトし、残りの７区分をネットワーク重み最適化に使用する、モデルのアンサンブルを築く。さらに、非ＭＳデータをスコア化する場合（図３１Ｃおよび図３１Ｄ）、標的ペプチドの１２～２０ｍｅｒの部分列の各々をスコア化し、最も高いスコアを保持した。
【０８１９】
我々のモデルを訓練する場合、損失関数における各ヒットおよびデコイについて、その供給源入れ子セットのサイズによって重みを減らし、したがって、各入れ子セットが全体として同等の重みを有した。モデルをハイパーパラメータチューニングについて評価する場合、各入れ子セットからの最も短いペプチドを関連性のある区分に陽性例として使用し、これらのヒットのスクランブルバージョンをデコイとして使用した。さらに、全体的な重み付け因子を、訓練時のヒットの合計された重みがデコイの合計された重みと等しくなるように適用した。モデルの最終的な評価のために、図３１Ｂに示されている通り、各入れ子セットから最も短いペプチドを再度選択して評価区分（区分８）に入れたが、デコイはペプチド供給源遺伝子の観察されていない部分配列からランダムにサンプリングした（「天然のデコイ」、後の節に記載されている）。このように、単に天然の配列とスクランブル配列を区別するためにモデルによって学習された偏りはいずれも評価区分に対する我々の性能を膨張させるものではなかった。
【０８２０】
モデルを、Ａｄａｍ最適化器を使用し、最初の学習率０．００３、ベータ＿１値０．９、ベータ＿２値０．９９９および減衰なし（学習率以外はデフォルトＫｅｒａｓパラメータ）で訓練し、バイナリ交差エントロピー損失関数を使用した。最初のモデルの重みを、Ｈｅ初期化を使用して設定した。５エポックの訓練毎に、チューニング区分に対する陽性適中率（ＰＰＶ、後の節に記載されている）を測定し、最大値を追跡した。各エポック後に、訓練損失が減少していない場合には、学習率に１／３を掛けた。同様に、毎回ＰＰＶをチューニング区分に対して測定し、ＰＰＶが最大実行と比較して増加していない場合、学習率に１／３を掛けた。訓練損失を３回の連続したエポックにわたって減少できなかった場合、またはチューニングＰＰＶを３回の連続したチェックにわたって最大実行よりも上に増加できなかった場合には訓練を停止する、早期停止スキームをインプリメントした。モデルを訓練した際、固定されたヒット対デコイ比１：３９を訓練セットに使用し、および１：１９をチューニング区分に使用した。
【０８２１】
特色付け：アミノ酸は「ワンホット」符号によって表すことができるが、他方ではＰＭＢＥＣ行列およびＢＬＯＳＵＭ行列を使用してアミノ酸を符号化する選択をし（Henikoff and Henikoff, 1992）、その場合、同様のアミノ酸は同様の特色プロファイルを有する。我々のペプチド特色付けの目的のために、解析されたタンパク質構造においてアミノ酸近接性に基づく新規の行列を生成した。この手法の概念は、アミノ酸の典型的な近隣アミノ酸に化学的特性が反映されるはずであるというものである。約１００，０００種のＤＳＳＰタンパク質構造の各々の各アミノ酸について（ｃｄｎ．ｒｃｓｂ．ｏｒｇ／ｅｔｌ／ｋａｂｓｃｈＳａｎｄｅｒ／ｓｓ．ｔｘｔ．ｇｚ）、３Ｄ空間では最も近いが、一次配列では少なくとも１０アミノ酸離れている残基を決定した。このデータを使用して、アラニンの最も近くの近隣アミノ酸がアラニンである回数、アラニンの最も近くの近隣アミノ酸がシステインなどである回数を決定して、近接計数の２０×２０行列を創出した。行列

の各要素をその対応する列および行の合計の積で割り、行列全体を対数変換した。最後に、行列全体の平均値を各要素から引いた。

【０８２２】

各アミノ酸をまた、酸性であるか（Ｎ、Ｑ）、脂肪族であるか（Ｉ、Ｌ、Ｖ）、芳香族であるか（Ｈ、Ｆ、Ｗ、Ｙ）、塩基性であるか（Ｈ、Ｋ、Ｒ）、荷電しているか（Ｄ、Ｅ、Ｈ、Ｋ、Ｒ）、疎水性であるか（Ａ、Ｃ、Ｆ、Ｈ、Ｉ、Ｋ、Ｌ、Ｍ、Ｔ、Ｖ、Ｗ、Ｙ）、ヒドロキシルであるか（Ｓ、Ｔ）、極性であるか（Ｃ、Ｓ、Ｎ、Ｑ、Ｔ、Ｄ、Ｅ、Ｈ、Ｋ、Ｒ、Ｙ、Ｗ）、小さいか（Ｖ、Ｐ、Ａ、Ｇ、Ｃ、Ｓ、Ｔ、Ｎ、Ｄ）、非常に小さいか（Ａ、Ｇ、Ｃ、Ｓ）、または硫黄を含有するか（Ｍ、Ｃ）などの、アミノ酸の特性を記述する１１種のバイナリ特色を用いて符号化した。各アミノ酸の位置を記述するために２つの特色を使用した。一方はペプチドにわたって単調に増加し、一方はペプチドの中心から絶対的距離を示し、どちらも単位は位置である（物理的距離ではない）。最後に、短いペプチドの端を越えて起こり得る、アミノ酸のその位置からの「欠損」が生じているかどうかを示すために単一のバイナリ特色を含めた。結果は、各アミノ酸が２０のアミノ酸近接特色、１１のアミノ酸文字特色、２の位置特色、および１の欠損文字特色の合計３４の特色によって符号化されるというものである。全てのペプチドを２０ｍｅｒとして符号化し、より長いペプチドについて中心的な２０アミノ酸を使用し、２０アミノ酸よりも短いペプチドの端に欠損文字値を対称的に加えた。

【０８２３】

例をニューラルネットワークに入力する際、訓練および評価の両方について、３４の特色の各々を、それらの平均を引き算し、それらの標準偏差で割ることによって正規化する。平均および標準偏差は訓練セットのみに基づいて算出し、ペプチド内部の位置は考慮に入れない。

【０８２４】

結合を予測するために、各対立遺伝子について畳み込みニューラルネットワークのアンサンブルを訓練した。モデルアーキテクチャの略図が図３１Ａに示されており、各々カーネルサイズが６でありフィルターが５０である２つの畳み込み層が示されている。各層の後、グローバル最大および平均プーリングを適用し、得られた値を、シグモイド活性化を用いる最終的な出力ニューロンに入力した。ＲｅＬＵ活性化、バッチ正規化（Ioffe and Szegedy, 2015）、および２０％空間ドロップアウトを各畳み込み層の直後に適用したことが含まれるが示していない。

モデルのアンサンブルを各対立遺伝子について訓練する際、アーキテクチャは固定したが、Ｌ２正則化の量は変動させた。基本のＬ２正則化の重み０．０５を第１の畳み込み層に使用し、および第２の畳み込み層には０．１を使用した。Ｌ２正則化の量を変動させるために、これらの値に０．１、０．５、および１を掛けた。アンサンブルにおける各反復について、正則化レベル当たり１つのモデルを訓練し、チューニング区分に対する性能に基づいて最良のモデルを保持した。

【０８２５】

ＭＡＰＴＡＣ（商標）により観察されたペプチドに関する予測性能のベンチマーキング、図７Ａに関連する

ＨＬＡ対立遺伝子によってコードされる所与のペプチドまたはタンパク質についての予測性能値の一部の例示的な評価では、「スクランブルデコイ」を含む方法を使用することができる。スクランブルデコイは、例えば質量分析データに基づいて所与のＨＬＡペプチドまたはタンパク質と結合することが分かっているペプチドと同じペプチド長およびアミノ酸を有するが、アミノ酸の配列がスクランブルされているペプチドである。図７Ａに示されている通り、質量分析によって同定されたすべての単一のペプチドについて、そのようなスクランブルペプチドデコイを１９使用した（ヒット：デコイ１：１９）。提示予測モデルを試験し、ＰＰＶを、試験区分におけるペプチドの最良スコア化５％を解析し、どの画分が陽性であるかを調べることによって決定した。このように生成されたＰＰＶが図７Ａおよび以下の表１２に示されている。

【表１２－１】

表 12

MHC クラス II 対立遺伝子	*neonmhc2* の PPV
DPB1-0101_DPA1-0103	0.48
DPB1-0101_DPA1-0201	0.41
DPB1-0101_DPA1-0202	0.56
DPB1-0201_DPA1-0103	0.48
DPB1-0202_DPA1-0103	0.27
DPB1-0301_DPA1-0103	0.36
DPB1-0401_DPA1-0103	0.52
DPB1-0402_DPA1-0103	0.54
DPB1-0501_DPA1-0201	0.59
DPB1-0501_DPA1-0202	0.46
DPB1-0601_DPA1-0103	0.43
DPB1-0901_DPA1-0201	0.53
DPB1-1001_DPA1-0201	0.49
DPB1-1101_DPA1-0201	0.31
DPB1-1301_DPA1-0201	0.58
DPB1-1401_DPA1-0201	0.49
DPB1-1701_DPA1-0201	0.39
DQB1-0201_DQA1-0201	0.49
DQB1-0201_DQA1-0501	0.08
DQB1-0202_DQA1-0201	0.3
DQB1-0301_DQA1-0501	0.27
DQB1-0301_DQA1-0505	0.13
DQB1-0301_DQA1-0601	0.28
DQB1-0302_DQA1-0301	0.1
DQB1-0303_DQA1-0201	0.46
DQB1-0303_DQA1-0301	0.39
DQB1-0401_DQA1-0301	0.47
DQB1-0402_DQA1-0401	0.53
DQB1-0501_DQA1-0101	0.62
DQB1-0502_DQA1-0101	0.26
DQB1-0502_DQA1-0102	0.19
DQB1-0601_DQA1-0102	0.17
DQB1-0601_DQA1-0103	0.35
DQB1-0602_DQA1-0102	0.42
DQB1-0603_DQA1-0103	0.56

【表１２－２】

DQB1-0604_DQA1-0102	0.28
DRB1_0101	0.48
DRB1_0102	0.49
DRB1_0301	0.47
DRB1_0302	0.29
DRB1_0401	0.38
DRB1_0402	0.49
DRB1_0403	0.43
DRB1_0404	0.42
DRB1_0405	0.41
DRB1_0407	0.36
DRB1_0410	0.61
DRB1_0701	0.49
DRB1_0801	0.36
DRB1_0802	0.46
DRB1_0803	0.30
DRB1_0804	0.36
DRB1_0901	0.35
DRB1_1001	0.54
DRB1_1101	0.43
DRB1_1102	0.42
DRB1_1104	0.51
DRB1_1201	0.47
DRB1_1202	0.55
DRB1_1301	0.42
DRB1_1302	0.23
DRB1_1303	0.56
DRB1_1401	0.56
DRB1_1501	0.53
DRB1_1502	0.43
DRB1_1503	0.41
DRB1_1601	0.61
DRB3_0101	0.57
DRB3_0201	0.6
DRB3_0202	0.48
DRB3_0301	0.53

【表１２－３】

DRB4_0103	0.67
DRB5_0101	0.54
DRB5_0102	0.65
DRB5_0202	0.63

【０８２６】

ＭＡＰＴＡＣ（商標）により観察されたペプチドに関する予測性能のベンチマーキング、図３１Ｂに関連する
所与の対立遺伝子についての予測性能を評価するために、観察することができたが（プロテオームに存在することが理由で）、ＭＳデータでは観察されなかったペプチドのセットを定義することが必要であった。これらの陰性例は訓練された「天然のデコイ」であった（上記の「スクランブルデコイ」とは対照的に）。指導原理の通り、以下を決定した：天然のデコイの長さ分布はＭＳにより観察されたヒットの長さ分布とマッチすべきである、天然のデコイは、他の天然のデコイと重複する配列を含有すべきではない、天然のデコイは重複ヒットであるべきではない、および／または天然のデコイは、少なくとも１つのヒットを生じた遺伝子に由来すべきである。

【０８２７】

以下の疑似コードは、これらの原理を満たす評価を創出するためにインプリメントしたプロセスを表す：

ヒットの空リスト２つ、$H_{minimal}$および$H_{exhaustive}$を初期化する
ＭＳにより観察されたペプチドの各入れ子セットＳについて：
Ｓ内に訓練またはチューニング区分内の転写物にマッピングすることができるペプチドがない場合：
Ｓ内の最も短いペプチドを$H_{minimal}$に加える
Ｓ内の全てのペプチドを$H_{exhaustive}$に加える
デコイペプチドの空リスト、Ｄを初期化する
試験区分内の各タンパク質をコードする転写物（最長が最初、最短が最後）について：
$H_{exhaustive}$内に転写物にマップされるペプチドがない場合：
飛ばして次の転写物に進む
転写物のタンパク質配列に重複ペプチドＰのセットをかぶせる。ペプチド長は$H_{minimal}$の長さ分布からランダムにサンプリングする。重複は８アミノ酸である（Ｐ内の最後のペプチドが一般にはタンパク質の終わりからぶら下がる）。
Ｐ内の最後のペプチドがまだぶら下がっている間に：
Ｐ内の最長ペプチドの長さから１アミノ酸を差し引く
Ｐ内の各ペプチドについて：
$H_{exhaustive}$内のペプチドと９ｍｅｒを共有せず、Ｄ内のペプチドのいずれかに観察される９ｍｅｒのいずれも有さない場合：
当該ペプチドをＤに加える
他の場合には：
当該ペプチドをはねる
$H_{minimal}$およびＤは評価データセットを構成する

【０８２８】

このセットに対する性能を評価するために、ｎ個のヒットペプチド全てを、１９ｎ個のデコイのセット（デコイの完全なセットから、元に戻さずにランダムにサンプリングしたもの）と一緒に予測器（ｎｅｏｎｍｈｃ２またはＮｅｔＭＨＣＩＩｐａｎ）によって評価し、スコア化した。組み合わせたセット内のペプチドの上位５％を陽性コールと標識し、陽性適中率（ＰＰＶ）を、ヒットであった陽性コールの分率として算出した。陽性の数はヒットの数と等しくなるように制約されるので、この評価シナリオでは、再現率はＰＰＶと正確に等しいことに留意されたい。対立遺伝子にわたって一貫した１：１９の比を適用することは、そうでなければ各対立遺伝子について観察されるヒットの数の影響を受ける性能値の安定化に役立つ。ヒットの数が対立遺伝子の内在性特性よりも実験条件および反復実験数に大きく関連することが仮定されたので、これは妥当であると思われた。

【０８２９】

非１５ｍｅｒに対するＮｅｔＭＨＣＩＩｐａｎ親和性の算出、図３１Ａ～Ｄ、図４０Ａ～Ｂ、および３３Ａ～Ｄに関連する
初期の解析では、非１５ｍｅｒに対するＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１親和性およ

びパーセント順位予測はベンチマークに対して不十分に実施された。しかし、以下の手法により性能が著しく改善された：ペプチドが１５アミノ酸よりも長い場合、全ての構成物を１５ｍｅｒとスコア化し、最も強い予測を全体的なペプチドスコアとして選択した；ペプチドが１５アミノ酸より短い場合、Ｎ末端にＧを当ててペプチドの長さを１５にし、得られた伸長したペプチドをスコア化した。

【０８３０】

訓練セットのサイズに応じた性能、図３８Ａに関連する

我々のモデルの性能が我々のデータセットのサイズによってどのように限定されるかを理解するために、飽和解析を実施した。これは、使用する訓練用データの画分を変動させることがホールドアウト区分に対する性能にどのように影響を及ぼすかを理解するために、使用する訓練用データの画分を変動させながらモデルのアンサンブルを再訓練することを伴うものであった。図３８Ａは、訓練セットに使用したヒットペプチドの数に応じた評価区分（区分８）ＰＰＶを示す。各データポイントは、１０モデルの集合にわたる平均ＰＰＶを示し、エラーバーは標準偏差を示す。

【０８３１】

天然のＣＤ４＋　Ｔ細胞応答の予測性能のベンチマーキング、図３１Ｃに関連する

ＩＥＤＢにおいて実証されたＣＤ４＋　Ｔ細胞応答の大部分（ｔｃｅｌｌ＿ｆｕｌｌ＿ｖ３．ｚｉｐ　ａｔ　ｉｅｄｂ．ｏｒｇ／ｄａｔａｂａｓｅ＿ｅｘｐｏｒｔ＿ｖ３．ｐｈｐ）は未知のまたはコンピュータにより帰属されたＨＬＡクラスＩＩ対立遺伝子制限を有するので、ＨＬＡクラスＩＩ四量体によって実験的に確認された記録のサブセットに焦点を当てた。そのような記録はほぼ全てがＷｉｌｌｉａｍ　Ｋｗｏｋ　Ｌａｂｏｒａｔｏｒｙ（Ｂｅｎａｒｏｙａ　Ｒｅｓｅａｒｃｈ　Ｉｎｓｔｉｔｕｔｅ、Ｓｅａｔｔｌｅ、ＷＡ）により寄託されたものであり、免疫反応性個体の血液を使用して、多様な病原体およびアレルゲンの四量体によりガイドされるエピトープマッピング（ＴＧＥＭ）を実施している。陰性ペプチドは一部の試験に関してはポストされているが、他の試験に関してはポストされていないので、ソース刊行物を精査して、陽性および陰性ペプチド反応性の完全なセットを再構築した。一部の場合では、陰性ペプチドはソース刊行物に明確に列挙されている。他の場合では、陰性を刊行物の方法で指定されているタイリング手順に従って帰属させ、それにより、ペプチド境界が既知の陽性例と一致することを確認する。図３１Ｃに示されているこのアッセイでは、インフルエンザウイルスおよびライノウイルスのウイルス遺伝子からウイルスエピトープをマッピングし、エピトープを含むペプチド配列を使用して、各エピトープに対するＨＬＡクラスＩＩタンパク質結合体を予測した。この場合、ペプチドに対するＣＤ４＋メモリーＴ細胞応答についてのＰＰＶをそれぞれのＨＬＡ－ＤＲＢ１タンパク質について予測した。ＰＰＶを、陽性結合体に関して、上位に順位付けられるエピトープのいずれの画分が真のヒットであったかを問うことによって決定した。ＨＬＡクラスＩＩタンパク質分子とペプチドの陽性対形成、およびそれぞれのウイルスを感染させた対象においてＨＬＡ分子が存在する場所を考慮して、対象においてＣＤ４応答が生じる。Ｎｅｏｎｍｈｃ２と公的に入手可能な予測器（ＮｅｔＭＨＣＩＩｐａｎ）の予測効率（ＰＰＶ、言い換えれば、真のヒット数の予測）の比較が図３１Ｃにこの例示的試験において試験した各ＤＲＢ１タンパク質毎に示されている。試験した６種の対立遺伝子の各々についてｎｅｏｎｍｈｃ２の方がＮｅｔＭＨＣＩＩｐａｎよりも優れていた。

【０８３２】

２０ｍｅｒのペプチド全てをｎｅｏｎｍｈｃ２によって、およびＮｅｔＭＨＣＩＩｐａｎ－ｖ３．１によってスコア化した。ＰＰＶを、上位にスコア化されたペプチドｎ個の間で実験的に確認された陽性の分率として算出し、ここで、実験的に確認されたペプチドが合計ｎ個存在した（図３１Ｃ）。

【０８３３】

Ｔ細胞誘導プロトコールおよび免疫原性読み取り、図３１Ｄに関連する

単球由来樹状細胞（ｍＤＣ）を生成するために、ＣＤ１４＋単球を、ＨＬＡ－ＤＲＢ１＊１１：０１＋健康ドナー末梢血単球（ＰＢＭＣ）から、ヒトＣＤ１４マイクロビーズを

製造者のプロトコール（Ｍｉｌｔｅｎｙｉ　Ｂｉｏｔｅｃ）の通り使用し、磁気分離によって単離した。単離されたＣＤ１４＋細胞を、８００Ｕ／ｍｌのｒｈＧＭ－ＣＳＦおよび４００Ｕ／ｍｌのｒｈＩＬ－４を補充したＣｅｌｌｇｅｎｉｘ　ＧＭＰ　ＤＣ培地（Ｃｅｌｌｇｅｎｉｘ）中で５日間にわたって分化させた。５日目に、ｍＤＣを回収し、０．４μＭのペプチドを３７摂氏温度で１時間にわたってパルスし、その後、１０ｎｇ／ｍｌのＴＮＦ－α、１０ｎｇ／ｍｌのＩＬ－１β、１０ｎｇ／ｍｌのＩＬ－６（Ｃｅｌｌｇｅｎｉｘ）、および０．５μｇ／ｍｌのＰＧＥ１（Ｃａｙｍａｎ　Ｐｈａｒｍａ）を使用して成熟化させた。４８時間後、ｍＤＣを、ＡＩＭＶ／ＲＰＭＩ（ＴｈｅｒｍｏＦｉｓｈｅｒ）、１０％ヒト血清（Ｓｉｇｍａ－Ａｌｄｒｉｃｈ）、１％Ｐｅｎ／Ｓｔｒｅｐ（ＴｈｅｒｍｏＦｉｓｈｅｒ）を含有し、５ｎｇ／ｍｌのＩＬ７およびＩＬ１５（Ｃｅｌｌｇｅｎｉｘ）を補充した培地中、自己ＰＢＭＣと１：１０の比で共培養した。１２日目に、Ｔ細胞を回収し、０．４μＭのペプチドをパルスした成熟ＤＣで７日間、２回の追加的な刺激で合計３回の刺激にわたって再刺激した。

【０８３４】

誘導されたＴ細胞を以前に記載されている一意的な２色バーコード標識系で標識し、ペプチドをパルスし、上記の通りＣＤ１４＋単球に由来する成熟させた自己ｍＤＣと１：１０の比で終夜培養した。翌朝、細胞をペプチドに応答したＩＦＮ－γの産生についてフローサイトメトリーによって評価した。細胞をＧｏｌｇｉ　Ｐｌｕｇ／Ｇｏｌｇｉ　Ｓｔｏｐ（ＢＤ　Ｂｉｏｓｃｉｅｎｃｅｓ）を用いて３７℃で４時間にわたって処理した。次いで、細胞を、ＣＤ１９、ＣＤ１６、ＣＤ１４、ＣＤ３、ＣＤ４、ＣＤ８に対する表面マーカー抗体（ＢＤ　Ｂｉｏｓｃｉｅｎｃｅｓ、Ｓａｎ　Ｊｏｓｅ、ＣＡ）、ならびにＬｉｖｅ／Ｄｅａｄ　Ｆｉｘａｂｌｅ　Ｄｅａｄ　Ｃｅｌｌ　ｓｔａｉｎ（ＴｈｅｒｍｏＦｉｓｈｅｒ）で染色した；以下の表１３を参照されたい。次いで、試料を透過処理し、ＢＤ　Ｃｙｔｏｆｉｘ／Ｃｙｔｏｐｅｒｍ　ｋｉｔ（ＢＤ　Ｂｉｏｓｃｉｅｎｃｅｓ）を用い、製造者のプロトコールに従って固定し、ＩＦＮ－γ（ＢＤ　Ｂｉｏｓｃｉｅｎｃｅｓ）に対する細胞内抗体を用いて染色した。試料を、ＢＤ　Ｆｏｒｔｅｓｓａ　Ｘ－２０フローサイトメーターに流し、ＦｌｏｗＪｏソフトウェア（Ｔｒｅｅｓｔａｒ）を使用して解析した。ペプチドに対して正に応答した誘導試料は、ＩＦＮ－ガンマ産生をペプチドなし対照よりも３％高く誘導する試料であった。

【表１３】

表 13

マーカー	フルオロフォア	ベンダー	カタログ番号	クローン
Live dead 染色	IR 色素	ThermoFisher	L34976	
CD19	BUV395	BD	563551	SJ25C1
CD16	BUV395	BD	563784	3G8
CD14	BUV395	BD	563562	MφP9
CD3	BUV805	BD	565511	SK7
CD4	AF700	BD	557922	RPA-T4
CD8	PerCP-Cy5.5	BD	565310	SK1
IFN-γ	APC	BD	554702	B27

【０８３５】

単一細胞ＲＮＡ－ＳｅｑでのＨＬＡクラスＩＩ発現データの解析、図１９Ａに関連する

単一細胞ＲＮＡ－Ｓｅｑデータを、３つの、ヒト腫瘍試料がプロファイリングされた以前に公開されたデータセットから得た。第１の試験には皮膚黒色腫からのデータを含めた。ファイル「ＧＳＥ７２０５６＿ｍｅｌａｎｏｍａ＿ｓｉｎｇｌｅ＿ｃｅｌｌ＿ｒｅｖｉ

ｓｅｄ＿ｖ２．ｔｘｔ」をＧｅｎｅ Ｅｘｐｒｅｓｓｉｏｎ Ｏｍｎｉｂｕｓ（ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｅｏ／；ａｃｃｅｓｓｉｏｎ：ＧＳＥ７２０５６）からダウンロードした。腫瘍の状態フラグ「２」を有する細胞を腫瘍細胞として扱い、腫瘍の状態フラグ「１」、および「１」から「６」までと等しい免疫細胞型フラグで標識された細胞をそれぞれＴ細胞、Ｂ細胞、マクロファージ、内皮、線維芽細胞、およびＮＫとして扱った。他の細胞は全て外した。データをｌｏｇ２（ＴＰＭ／１０＋１）の単位でネイティブなまま示し、したがって、ＴＰＭ尺度に数学的に変換した。ＴＰＭ尺度になったら、各細胞についてのデータを、タンパク質をコードするＵＣＳＣ遺伝子の符号のセットにわたって再正規化して、１，０００，０００まで合計した（発現行列に現れないタンパク質をコードする遺伝子は暗黙のうちにゼロ発現を有するものとして扱った）。最後に、同じ細胞型および同じ供給源の生検材料に対応する単一細胞観察を平均して患者－細胞型レベルの発現推定値を生じさせた。

【０８３６】

第２の試験には、頭頸部腫瘍からのデータを含めた。ファイル「ＧＳＥ１０３３２２＿ＨＮＳＣＣ＿ａｌｌ＿ｄａｔａ．ｔｘｔ」をＧｅｎｅ Ｅｘｐｒｅｓｓｉｏｎ Ｏｍｎｉｂｕｓ（ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｅｏ／；ａｃｃｅｓｓｉｏｎ：ＧＳＥ１０３３２２）からダウンロードした。Ｉｔａｙ Ｔｉｒｏｓｈ（August 22, 2018）との個人の対応に従って、この表中のデータはまた、ｌｏｇ２の単位（ＴＰＭ／１０＋１）であった。したがって、値をＴＰＭ単位に数学的に変換した。黒色腫試験と同様に、各細胞についてのデータを、タンパク質をコードするＵＣＳＣ遺伝子の符号のセットにわたって再正規化して、１，０００，０００まで合計し、同じ細胞型および同じ供給源の生検材料に対応する単一細胞観察を平均した。リンパ節生検材料に対応するデータを除外した。

【０８３７】

第３の試験には、無処置の非小細胞肺からのデータを含めた。ファイル「ＲａｗＤａｔａＬｕｎｇ．ｔａｂｌｅ．ｒｄｓ」および「ｍｅｔａｄａｔａ．ｘｌｓｘ」をＡｒｒａｙ Ｅｘｐｒｅｓｓ（ｅｂｉ．ａｃ．ｕｋ／ａｒｒａｙｅｘｐｒｅｓｓ／；ａｃｃｅｓｓｉｏｎｓ：Ｅ－ＭＴＡＢ－６１４９およびＥ－ＭＴＡＢ－６６５３）からダウンロードした。データ（すでにＴＰＭ）単位を、以前に記載されているタンパク質をコードする遺伝子のセットにわたって再尺度化して、１，０００，０００まで合計した。最後に、同じ細胞型および同じ供給源の生検材料に対応する単一細胞観察を平均して患者－細胞型レベルの発現推定値を生じさせた。簡単にするために、以下の表１４に、細胞型をマージしてネイティブなまま報告されたものよりも粒度を粗くした。

【表１４】

表 14

大雑把な名称	構成物の細胞型
肺胞	「肺胞」、「立方肺胞2型(AT2)細胞」は除く
FO B細胞	「濾胞性B細胞」
形質細胞	「形質B細胞」
CLEC9A+ DCs	「交差提示性樹状細胞」
monoDCs	「単球由来樹状細胞」
pDCs	「形質細胞様樹状細胞」
ランゲルハンス	「ランゲルハンス細胞」
マクロファージ	「マクロファージ」
顆粒球	「顆粒球」
内皮	「正常な内皮細胞」、「腫瘍内皮細胞」、および「低品質内皮細胞」、「リンパ管EC」は除く
上皮	「上皮細胞」および「低品質上皮細胞」
線維芽細胞	「COL12A1発現線維芽細胞」、「COL4a2発現線維芽細胞」、「GABARAP発現線維芽細胞」、「低品質線維芽細胞」、「正常な肺線維芽細胞」、「PLA2G2A発現線維芽細胞」、および「TFPI2発現線維芽細胞」
T細胞	「調節性T細胞」、「CD4+ T細胞」および「CD8+ T細胞」
NKs	「ナチュラルキラー細胞」
腫瘍	「がん細胞」
解析から除外	「赤芽球」および「MALT B細胞」

【０８３８】

第４の試験には、結腸直腸腫瘍からのデータを含めた。ファイル「ＧＳＥ８１８６１＿ＣＲＣ＿ｔｕｍｏｒ＿ａｌｌ＿ｃｅｌｌｓ＿ＦＰＫＭ．ｃｓｖ」をＧｅｎｅ Ｅｘｐｒｅｓｓｉｏｎ Ｏｍｎｉｂｕｓ（ｎｃｂｉ．ｎｌｍ．ｎｉｈ．ｇｏｖ／ｇｅｏ／；ａｃｃｅｓｓｉｏｎ：ＧＳＥ８１８６１）からダウンロードした。データ（すでにＴＰＭ）単位を、以前に記載されているタンパク質をコードする遺伝子のセットにわたって再尺度化して、１，０００，０００まで合計した。最後に、同じ細胞型および同じ供給源の生検材料に対応する単一細胞観察を平均して患者－細胞型レベルの発現推定値を生じさせた。この試験に関しては、「上皮」と標識された細胞は腫瘍細胞と正常な上皮の混合物を表すと推定される。

【０８３９】

第５の試験には、漿液性卵巣がん腫瘍からのデータを含めた。悪性度が低い漿液性卵巣がん患者２例（ＬＧ１、ＬＧ２）および悪性度が高い漿液性卵巣がん患者４例（ＨＧ１、ＨＧ２Ｆ、ＨＧ３、ＨＧ４）の６例の卵巣上皮がんに関する単一細胞ＲＮＡ配列決定データを他の箇所から得た。品質フィルタリング、クラスタリングおよび解析はShih et al., 2018に概説されているステップに従った。簡単に述べると、Ｓｅｕｒａｔ解析ツールを

使用して、品質フィルタリングを通過する細胞をクラスター化した（最小２００種の発現された遺伝子、各遺伝子は少なくとも３つの異なる細胞において検出されなければならない；全部で２２５８の細胞）。細胞周期および独特の転写物計数の影響を取り除いた。細胞を主成分分析に従ってクラスター化し、クラスターを元の刊行物からの遺伝子シグネチャーの発現に基づいて細胞型に割り当てた。各患者の各細胞型について、ＨＬＡ－ＤＲＢ１遺伝子についてのＴＰＭを、タンパク質コード遺伝子の正規化された独特の転写物計数から算出した。

【０８４０】

４つの試験におけるＨＬＡ－ＤＲＢ１の発現レベルを図１９Ａにプロットした。

腫瘍由来ＨＬＡクラスＩＩ発現と間質由来ＨＬＡクラスＩＩ発現との特徴付け、図１９Ｂに関連する

腫瘍に起因するＨＬＡクラスＩＩ発現と間質に起因するＨＬＡクラスＩＩ発現の相対量を決定するために、ＴＣＧＡ患者におけるＨＬＡクラスＩＩ経路遺伝子のＤＮＡ配列決定からコールされた変異を同定し、ＨＬＡクラスＩＩ変異を有する各患者について、遺伝子の変異したコピーおよび変異していないコピーの相対的な発現量を対応するＲＮＡ－Ｓｅｑで数量化した。さらに、変異リードが腫瘍から生じ、変異していないリードが間質または腫瘍における野生型対立遺伝子について生じ、および腫瘍が変異コピーとほぼ等しい発現を有する野生型コピーを保持することが仮定された。

【０８４１】

これに基づいて、観察された変異体対立遺伝子分率ｆに関して、腫瘍に起因するＨＬＡクラスＩＩ発現の分率がおよそ２ｆであり、１００％以下であったことが決定された。３つの遺伝子、つまりＣＩＩＴＡ、ＣＤ７４、およびＣＴＳＳを、コアＨＬＡクラスＩＩ経路遺伝子として選択し、ＴＣＧＡの変異について評価した（同義のおよびＵＴＲ変異を除外せず）（データをＴｕｍｏｒＰｏｒｔａｌ（ｔｕｍｏｒｐｏｒｔａｌ．ｏｒｇ／）：ＢＲＣＡ、ＣＲＣ、ＨＮＳＣ、ＤＬＢＣＬ、ＭＭ、ＬＵＡＤ；ＴＣＧＡバルクダウンロード（ｔｃｇａ－ｄａｔａ．ｎｃｉ．ｎｉｈ．ｇｏｖ）：ＣＥＳＣ、ＬＩＨＣ、ＰＡＡＤ、ＰＲＡＤ、ＫＩＲＰ、ＴＧＣＴ、ＵＣＳ；シナプス（ｓｙｎａｐｓｅ．ｏｒｇ／＃！Ｓｙｎａｐｓｅ：ｓｙｎ１７２９３８３）：ＧＢＭ、ＫＩＲＣ、ＬＡＭＬ、ＵＣＥＣ、ＬＵＳＣ、ＯＶ、ＳＫＣＭ；または元のＴＣＧＡ公開（ｃａｎｃｅｒｇｅｎｏｍｅ．ｎｉｈ．ｇｏｖ／ｐｕｂｌｉｃａｔｉｏｎｓ）：ＢＬＣＡ、ＫＩＣＨ、ＳＴＡＤ、およびＴＨＣＡからダウンロードした）。これらの遺伝子を、ＨＬＡクラスＩＩ発現におけるそれらの公知の役割、および８５００ＧＴＥｘ試料のコホートにわたるそれらのＨＬＡ－ＤＲＢ１との密接な相関に基づいて選択した。ＨＬＡ－ＤＲＢ１（ＨＬＡ－ＤＲＡ１、ＨＬＡ－ＤＰＡ１、ＨＬＡ－ＤＱＡ１、ＨＬＡ－ＤＱＢ１、およびＨＬＡ－ＤＰＢ１）と等価相関を有する他の遺伝子は、それらの多型的性質により偽陽性変異コールが起こりやすくなるので除外した。天然に、患者のごく一部のみがＣＩＩＴＡ、ＣＤ７４、またはＣＴＳＳに変異を有し、また、一部の腫瘍型に関しては、解析に利用可能な患者が存在しなかった。

【０８４２】

元の全エクソーム配列決定（ＷＥＳ）ＢＡＭを視覚的に評価して（ＩＧＶ）、変異が腫瘍試料には存在するが正常な試料には存在しないことを確認した。変異体リード計数と野生型リード計数とを対応するＲＮＡ－Ｓｅｑからｐｙｓａｍを使用して得た。全体的なＨＬＡ－ＤＲＢ１発現をＧｅｎｏｍｉｃ　Ｄａｔａ　Ｃｏｍｍｏｎｓ（ｇｄｃ．ｃａｎｃｅｒ．ｇｏｖ／）からダウンロードされた発現データに基づいて決定し、それを、タンパク質をコードする遺伝子のセットにわたって再正規化して、１，０００，０００まで合計した。腫瘍に起因するＨＬＡ－ＤＲＢ１発現の分率（図１９Ｂ）をｍｉｎ（１，２ｆ）として推定し、ここで、ｆは、変異を示すＣＩＩＴＡ、ＣＤ７４、またはＣＴＳＳにおけるＲＮＡ－Ｓｅｑリードの分率である。

【０８４３】

過剰に表された遺伝子および過少に表された遺伝子の同定、図３２Ａおよび図３９Ｂに関連する

卵巣がん、結腸直腸がん、および黒色腫のＭＨＣ－ＩＩリガンドームがプロファイリングされた以前に公開されたＭＳ実験からの試料を解析した。卵巣がんデータセットからの多くの試料はＲＮＡ－Ｓｅｑが利用可能なものであり、これらの試料のデータをＳＲＡ（ＮＣＢＩ　ＢｉｏＰｒｏｊｅｃｔ　ＰＲＪＮＡ３９８１４１）からダウンロードし、ＳＴＡＲアライナーを使用してＵＣＳＣ　ｈｇ１９トランスクリプトームとアラインメントした。ＲＮＡ－Ｓｅｑが利用可能でなかった卵巣試料については、ＲＮＡ－Ｓｅｑが利用可能であった試料にわたって平均することによって発現を推定した。結腸直腸試験および黒色腫試験に関しては、いずれの試料についても対応するＲＮＡ－Ｓｅｑが存在せず、したがって、ＴＣＧＡ（Ｔｈｅ　Ｃａｎｃｅｒ　Ｇｅｎｏｍｅ　Ａｔｌａｓ　Ｎｅｔｗｏｒｋ）からのデータを使用して代理試料にわたって平均を算出した。ＲＳＥＭバージョン－１．２．３１によって算出される百万当たりの転写物数（ＴＰＭ）を使用して転写レベルでの遺伝子数量化を実施した。発現推定値を、遺伝子レベルを合計すること、非コード遺伝子を外すこと、および総ＴＰＭが合計１０００００００になるように再正規化すること（ｎｃＲＮＡ存在量のライブラリー間の変動の原因となるタンパク質をコードする遺伝子にわたって再正規化すること）によってさらに処理した。

【０８４４】

ＨＬＡクラスＩＩリガンドームにおいて過剰に表された遺伝子および過少に表された遺伝子を同定するために、発現解析に使用したものと同じ３つのデータセットを解析した。各遺伝子について、我々のベースライン仮定は、ペプチドがその長さにその発現レベルを掛けた割合でもたらされるはずであるというものであった。各遺伝子の長さを決定するために、全ての転写物アイソフォームにわたる独特の９ｍｅｒを数え上げた。転写物アイソフォームにわたって合計することによって遺伝子レベル発現を得た。観察された、各遺伝子にマッピングされるペプチドの数を入れ子セットレベルで決定した（例えば、ペプチドＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤＶ（配列番号１６）、ＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１７）、およびＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤＶ（配列番号１８）を単一の観察として計数した）。

【０８４５】

それぞれＥおよびＯと称される、予測される計数および観察された計数を表す２つの行列を創出し、ここで、行は遺伝子に対応し、列は試料に対応する。Ｏ内の値は、入れ子セットレベルで試料当たりのペプチド数を計数することによって決定した。まず、各試料中の各遺伝子の長さにその発現を掛けることによって行列Ｅをポピュレートし、次いで、Ｅの列を再尺度化して、Ｅの列の合計とＯの列の合計がマッチするようにした。最後に、Ｅの行の合計とＯの行の合計を比較することによって遺伝子レベル解析を行った（図３２Ａ）。遺伝子をヒト血漿中のそれらの存在および濃度に従って強調表示した。同じ試験のセットからの黒色腫、結腸直腸がん、および卵巣がんデータを使用してＨＬＡクラスＩデータにおける過剰に表された遺伝子および過少に表された遺伝子を同定するために同一の手法を使用した。ＨＬＡクラスＩ解析に関しては、入れ子を適用せず、独特のペプチドのみを計数した。

【０８４６】

過剰に表された遺伝子の結合スコアの評価、図３９Ａに関連する

過剰に表された遺伝子の多くが血漿遺伝子であることが観察された。血清遺伝子の包括的な一覧を得、血漿遺伝子に由来するＨＬＡ　ＤＲ結合ペプチドについて、非血清遺伝子に由来するＨＬＡ－ＤＲ結合ペプチドと、ならびに、免疫ペプチドームにおいて表された遺伝子からサンプリングした長さを釣り合わせた非結合性（例えば、ＭＳでは観察されていない）ペプチドとｎｅｏｎｍｈｃ２結合スコアを比較した。汎ＤＲ抗体を用いてプロファイリングされたＨＬＡクラスＩＩペプチドを有する、遺伝子型決定された複対立遺伝子データセットについて（図３０Ｂにおける解析と同じ試料）、ｎｅｏｎｍｈｃ２で各ＤＲ対立遺伝子について試料が発現されたペプチドをスコア化した。発現された対立遺伝子全てにわたるｎｅｏｎｍｈｃ２による最良のスコア出力を各ペプチドの代表的なスコアとした。データを全ての使用可能なデータセットにわたってプールし、ペプチドの各カテゴリ

ーについてスコアの分布を箱ひげ図で可視化した。

【０８４７】

タンパク質ターンオーバーに関連する遺伝子の解析、図３２Ｃに関連する

ターンオーバーがプロテアソームによって調節されるタンパク質を表すことが意図された２つの遺伝子セットを同定した。第１の遺伝子セットは、細胞株ＫＧ１、Ｊｕｒｋａｔ、またはＭＭ１Ｓにおける少なくとも１つの観察されたユビキチン化部位を有する遺伝子を含むものであった。第２のセットは、ｐ値フィルター０．０１を適用し、上向きの変化倍率が最も大きな３００種の遺伝子を選択して、公開された論文のプロテアソーム阻害剤ボルテゾミブ（ＢＴＺ）を適用するとレベルが上昇する遺伝子を含むものであった。

【０８４８】

バルク腫瘍と抗原提示細胞遺伝子発現の説明力の比較、図３９Ｃに関連する

４つの遺伝子発現プロファイルを創出した。第１の遺伝子発現プロファイルは、ＡＰＣを表すことが意図され、上記の単一細胞ＲＮＡ－Ｓｅｑ実験からの細胞型特異的プロファイルを平均することによって推定されたものである。平均には「マクロファージ」（頭頸部試験、肺試験、および黒色腫試験からのもの）、「ＣＬＥＣ９Ａ　ＤＣ」（肺試験からのもの）、および「ｍｏｎｏＤＣ」（肺試験からのもの）を含めた。他の３つの発現プロファイルは卵巣がん、結腸直腸がん、および黒色腫からのバルク腫瘍プロファイルに対応するものである（データ、図１９Ａ）。卵巣プロファイルはＳｃｈｕｓｔｅｒらによって公開された試料の平均であり、他のプロファイルは、腫瘍型毎の、「絶対的」アルゴリズムを使用して以前に推定された最も高い腫瘍細胞充実性を有する５つのＴＣＧＡ試料に由来するものである。各腫瘍型について、遺伝子当たりのペプチドの数（入れ子セットレベルで）を計数し、遺伝子の長さ、ＡＰＣ特異的遺伝子発現、および腫瘍特異的遺伝子発現に応じた各遺伝子のペプチド計数を線形回帰を使用してモデル化した。出力変数および全ての入力変数をｌｏｇ（ｘ＋１）によって転換した。モデルのパラメータ推定値を使用し、腫瘍の寄与を$\beta_{腫瘍}$／（$\beta_{腫瘍}$＋β_{APC}）として算出し、ＡＰＣの寄与をβ_{APC}／（$\beta_{腫瘍}$＋β_{APC}）として算出した。各試料について、遺伝子レベルでのブートストラップリサンプリング（Ｍ＝１００）を使用して、説明的割合についての信頼区間を算出した。

【０８４９】

ＨＬＡクラスＩＩペプチドの観察された切断部位の特徴付け、図４０Ａに関連する

天然にプロセシングされ、提示されるＨＬＡクラスＩＩペプチドを６つのデータセット：ＰＢＭＣドロー、ＤＣ様ＭＵＴＺ３細胞株、結腸直腸がん組織、黒色腫、卵巣がん、およびｅｘｐｉ２９３細胞株から解析した。多くのペプチドが同じＮ末端（例えば、ＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤＶ（配列番号１６）およびＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１７））または同じＣ末端（例えば、ＧＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１７）およびＫＡＰＩＬＩＡＴＤＶＡＳＲＧＬＤ（配列番号１９））を共有するので、Ｎ末端にについて１つ、およびＣ末端について１つの、２セットの重複していないカット部分をキュレートした。図４１に示されている命名システムを使用してペプチドの上流、ペプチド内部、およびペプチドの下流の位置を指す。上流および下流の頻度（．．．Ｕ１およびＤ１．．．）、をプロテオームアミノ酸頻度と比較し、有意な偏差をカイ二乗検定によってスコア化した。ペプチド位（Ｎ１．．．Ｃ１）をＭＳペプチドにおいて観察された頻度と比較した。

【０８５０】

種々のＨＬＡクラスＩＩ切断予測器の性能のベンチマーキング、図４０Ｂに関連する

４つのＰＢＭＣ試料および公開されたデータセットを使用して、切断関連変数／予測器の、提示されたＨＬＡクラスＩＩエピトープの同定を増強する能力をベンチマークした。

【０８５１】

結合潜在性および切断潜在性の両方を使用してペプチドの提示を予測する統合予測器を築くために、まず、図３１Ｂに関して記載されているものと同じ手法を使用してデータセットを構築した。これは、ヒット対デコイの比１：１９を使用することを意味し、デコイは、ヒットと長さを釣り合わせたものであり、少なくとも１つのヒットが生成された遺伝

子のセットからランダムにサンプリングされたものである。３つの異なる目的のために異なるデータセットをこのように築いた：

１．溶媒露出度に基づく切断予測器および障害に基づく切断予測器に関しては、ヒト腫瘍組織からのＨＬＡクラスＩＩリガンドームデータを使用してロジスティックモデルを当てはめた。リガンドーム実験において観察されたペプチドについては、首尾よくプロセシングされたに違いないことが推定された（ニューラルネットワークおよびＣＮＮに基づく切断予測器に関しては、訓練用データを、下記の表において説明されている通り、同じデータセットを使用し、別個の様式で生成した）。

２．所与の切断予測器では結合単独に対して性能がブーストされるかどうかを評価するために、Ｂ７２１細胞およびＫＧ１細胞を用いて生成した単一対立遺伝子ＭＡＰＴＡＣ（商標）データを使用してモデルを当てはめ、大多数の機能的にＡＰＣ様の細胞株を調べた。ｎｅｏｎｍｈｃ２を使用して結合潜在性を算出し、ロジスティック回帰により前方予測において結合および切断変数に乗せられる相対的な重みを決定した。

３．前方予測の性能を評価するために、ＰＢＭＣ試料についてのデータセットおよび公開されたデータセットを前と同じ様式で構築した。しかし、これらの試料は複対立遺伝子であったので、各ペプチド候補ペプチドについての結合スコアを、各ドナーの遺伝子型によって示される１～４種のＤＲ対立遺伝子の最大のスコア化のために取った。ＰＰＶを図３１Ｂについて記載されている通り算出した。

【０８５２】

いくつかの異なる切断予測器を評価した。

【０８５３】

まず切断モデル、カット部分既知（ニューラルネットワーク）

ＭＳにより観察されたカット部分からの切断シグナルを学習するために、腫瘍組織由来ＨＬＡクラスＩＩリガンドームにおけるＵ３からＮ３まで、およびＣ３からＤ３まで（「ＨＬＡクラスＩＩペプチドの観察された切断部位の特徴付け、図４０Ａに関連する」の節で導入された命名法を使用）の独特の６ｍｅｒのアミノ酸配列の全てを、それぞれＮ末端カットおよびＣ末端カットをモデル化する２つの別個のニューラルネットワークを訓練するための陽性例として使用した。前と同様に、同等数の独特の観察されていないＮ末端およびＣ末端カット部分（陰性例）を、そのコンテキストについてのプロテオームのアミノ酸頻度から、およびペプチドについてＭＳにより観察されたリガンドームから合成により生成した。アミノ酸配列を、ｎｅｏｎｍｈｃ２に使用したものと同じ特色のサブセット、具体的には、タンパク質構造に基づくアミノ酸近接性、およびアミノ酸特性（例えば、酸性、脂肪族など）で符号化した。Ｎ末端およびＣ末端で観察されたカット部分モデルの各々について（ｌｒ＝０．０００５、Ａｄａｍ最適化器、損失関数としてバイナリ交差エントロピー）、次いで、全結合ニューラルネットワークを、ＲｅＬｕ活性化を用いた２つの隠れ層（１つの層に２０個のニューロン、その後、次の層に１０個のニューロン）を用いて訓練し、その後、最終的なシグモイド層を用いて訓練した。正則化のために、ドロップアウト率２０％、Ｌ２ノルム０．００１（Ｃ末端モデルについてのみ）、および最大ノルム制約４を使用した。

【０８５４】

候補ペプチドをスコア化するために、Ｎ末端モデルをペプチドに対して６ｍｅｒ配列Ｕ３～Ｎ３に適用し、Ｃ末端モデルをＣ３～Ｄ３に適用した。Ｎ末端モデルおよびＣ末端モデルの両方を候補ペプチドにわたって６ｍｅｒ配列タイリングにも適用して、ペプチド自体内の配列の切断傾向を評価した。ｎｅｏｎｍｈｃ２結合スコア、ならびに、Ｎ末端、Ｃ末端ならびにペプチド内のＮ末端およびＣ末端モデルの最大スコア化カット部分に対応する４つのニューラルネットワーク出力を使用してロジスティック回帰をＭＡＰＴＡＣ（商標）データに対して訓練した。

【０８５５】

まず切断モデル、カット部分未知（＋／－１５ＡＡ）（ニューラルネットワーク）

ペプチドの厳密な末端が分かっていない場合に、観察されたカット部分から学習した切

断モデルが予測的であるかどうかを決定するために、上で学習した同じニューラルネットワークを、ペプチド末端を１５アミノ酸越えて、拡張されたコンテキストに適用した。この場合の候補ペプチドをスコア化するために、３つの領域：Ｎ末端モデルを用いてスコア化した、ペプチドの１５アミノ酸上流（真のＮ末端切断部位の位置にかかわらず）、Ｎ末端モデルおよびＣ末端モデルの両方を用いてスコア化したペプチド配列、およびＣ末端モデルを用いてスコア化した、ペプチドの１５アミノ酸下流にわたって最大のスコアを算出した。ｎｅｏｎｍｈｃ２結合スコアならびに４つの領域特異的スコア（ペプチド自体がＮ末端およびＣ末端モデルからの値の２つのセットに寄与するので）を使用してロジスティック回帰をＭＡＰＴＡＣ（商標）データに対して訓練した。

【０８５６】

まず結合モデル、溶媒露出度

ＳＣＲＡＴＣＨ　ｓｕｉｔｅ内で、ＡＣＣｐｒｏ２０ツールを使用して、総体的な溶媒露出度を予測した。次いで、ペプチドの平均溶媒露出度スコアを考慮してペプチドがプロセシングされる可能性を、腫瘍組織データを使用してロジスティック回帰に当てはめた。最後に、ｎｅｏｎｍｈｃ２結合スコアおよび腫瘍組織により訓練された予測器からの出力を使用してロジスティック回帰を単一対立遺伝子データに対して訓練した。

【０８５７】

まず結合モデル、障害

配列障害の残基毎のスコアをプロテオーム全体にわたって決定し、当該位置を不規則と標識した予測エンジンの数に従って０～５の尺度でスコア化した（使用したサーバー：ａｎｃｈｏｒ、ｅｓｐｒｉｔｚ－ｄ、ｅｓｐｒｉｔｚ－ｎ、ｅｓｐｒｉｔｚ－ｘ、ｉｕｐｒｅｄ－ｌ、およびｉｕｐｒｅｄ－ｓ）。各候補ペプチドにわたって平均障害スコアを算出し、６つの障害予測器の出力を合計した。溶媒露出度と同様に、まず、ロジスティックモデルを、この全体的な障害スコアを使用して腫瘍組織データに当てはめた。この後に、ｎｅｏｎｍｈｃ２結合スコアおよび腫瘍組織により訓練された予測器からの出力を使用してロジスティック回帰の単一対立遺伝子データに対する訓練を行った。

【０８５８】

ハイブリッドモデル、前駆体カットスキャン（＋／－３０ＡＡ）（ＣＮＮ）

ヒットに対する訓練用データを、「まず切断、カット部分既知」切断予測器について記載されている通り、独特の６ｍｅｒ配列Ｕ３からＮ３まで、およびＣ３からＤ３までを使用する代わりに、ペプチドにフランキングする３０アミノ酸（Ｕ３０～Ｕ１、およびＤ１～Ｄ３０）をモデル入力としたことを例外として生成した。さらに、３０ｍｅｒの配列がＮ末端フランクに由来するのかＣ末端側フランクから由来するのかは区別せず、その代わりに、観察されたペプチドの両側に存在し得ると仮定された前駆体カットシグナルを学習させるために、データをプールして単一のモデルを訓練した。この状況では、合成デコイを使用する代わりに、同じ供給源遺伝子から引き出された観察されていないペプチド由来のフランキング配列を陰性例として使用した。配列を前と同様に、タンパク質構造に基づくアミノ酸近接性、およびアミノ酸特性（例えば、酸性の、脂肪族のなど）を使用して符号化した。ＣＮＮのアーキテクチャは、２つの畳み込み層、カーネルサイズが２であり、フィルターを４８個有する第１の層と、その後の、カーネルサイズが３であり、フィルターを４０個有する層とからなった。これらの層はＲｅＬｕ活性化を有した。畳み込み層の後にグローバル最大プーリング層が続き、その後、シグモイド活性化を用いる最終的な高密度層が続いた。ＣＮＮを学習率０．００１で、Ａｄａｍ最適化およびバイナリ交差エントロピーを損失関数として用いて訓練した。

【０８５９】

候補ペプチドをスコア化するために、ＣＮＮをペプチドの３０アミノ酸上流および３０アミノ酸下流に適用し、Ｎ末端フランキングスコアおよびＣ末端フランキングスコアを生じさせた。ｎｅｏｎｍｈｃ２結合スコアおよび２つのＣＮＮスコアを使用してロジスティック回帰をＭＡＰＴＡＣ（商標）データに対して訓練した。

【０８６０】

ＤＱ重複

ＭＳに基づくペプチド同定をBergseng et al., 2015からのＨＬＡ－ＤＱリガンドームにわたってプールした。新しい候補ペプチドが前に観察されたペプチドのうちの１つと重複するかどうかを表す新しい特色を創出した。具体的には、以前に観察されたＨＬＡ－ＤＱリガンドのセット内の任意のペプチドと少なくとも１つの９ｍｅｒを共有する場合には特色を１に設定し、他の場合には特色を０に設定した。ｎｅｏｎｍｈｃ２結合スコアおよび重複特色を使用してロジスティック回帰を単一対立遺伝子データに対して訓練した。

【０８６１】

統合された結合および切断モデルついても全てに図４０Ｂにおけるものの代わりにＮｅｔＭＨＣＩＩｐａｎを結合予測器として使用して当てはめおよび評価を行った。

【０８６２】

天然のドナー組織に対する予測の全体的な性能の評価、図２１Ａ～２１Ｂに関連する

７体の健康なドナー由来の末梢血を、上の節「抗体に基づくＨＬＡ－ペプチド複合体の単離」に記載されている通り、ＤＲ特異的抗体を用いてプロファイリングした。訓練データセットおよび評価データセットを、図３１Ｂとの関連で前に記載されているヒットおよびデコイ選択アルゴリズムを使用して構築した。手短に言えば、これは、各入れ子セットを１つのヒットペプチド（入れ子セット内で最も短いペプチド）で表すことおよび遺伝子にわたって長さを釣り合わせたデコイをタイリングすることを意味し、したがって、ヒットと最小に重複し、互いに最小に重複する。この状況では、デコイ選択はＭＳにより観察された遺伝子に制約されず、その代わりに、デコイを、プロテオーム全体から、元に戻すことなくランダムにサンプリングした。プロテオームにおけるＨＬＡ－ＤＲにより提示されたペプチドの大まかな推定頻度を反映する、ヒットとデコイの率１：４９９を利用した。ＭＨＣ結合スコア（ＮｅｔＭＨＣＩＩｐａｎまたはｎｅｏｎｍｈｃ２からのもの）ならびに他の入力特色（発現、遺伝子の偏り、およびＤＱ重複）を用いたロジスティック回帰モデルをＫＧ１細胞株およびＢ７２１細胞株からのＭＡＰＴＡＣ（商標）データに対して訓練した。

【０８６３】

以下の表１５中の変数を回帰のサブセットに使用した。

【表１５】

表 15

NetMHCIIpanパーセント順位	NetMHCIIpan-v3.1から導き出されたもの。各候補ペプチドについて、ドナーの遺伝子型における全てのDR対立遺伝子にわたって最も強いスコアを取得した。
*neonmhc2*パーセント順位	*Neonmhc2*から導き出されたもの。各候補ペプチドについて、ドナーの遺伝子型における全てのDR対立遺伝子にわたって最も強いスコアを取得した。
発現	遺伝子発現の推定値を(bowtie2、RSEM、およびタンパク質をコードする遺伝子のみにわたる再正規化)からのデータを解析することによって得、値をN個の試料にわたって平均した。発現を閾値で表したか(0/1、発現がゼロでない場合を示す)または連続変数として処理した。
遺伝子の偏り	**図32A**の解析当たりの(1+観察値)/(1+予測値)
DQ重複	候補ペプチドがHLA-DQデータセットのうちのいずれかと少なくとも1つの9merを共有するかどうかについての指示変数(0/1)。HLA-DQ結合性ペプチドはHLA-DR結合性ペプチド由来の別個の結合性モチーフを有し、したがって、結合の傾向はこの特色では学習されないはずである。

【０８６４】

次いで、これらのモデルの天然のドナー組織（ＰＢＭＣ試料など）由来のＨＬＡ－ＤＲリガンドームに対する性能を評価した。デコイを、入手可能なデコイ配列をほぼ飽和させるヒット対デコイの比１：４９９が実現されるように、プロテオームからランダムにサンプリングした（ＭＳにより観察されたペプチドを決して産生しない遺伝子を含む）。ヒットとデコイの率１：４９９を評価（ならびに訓練）のために使用した。評価されたデータセット内のスコア化されたペプチドの上位０．２％を陽性コールとして標識し、ＰＰＶをヒットであった陽性コールの分率として算出した（例えば、図２１Ａおよび表１５を参照されたい）。この評価シナリオでは、陽性の数はヒットの数と等しくなるように制約されるので、再現率はＰＰＶと正確に等しいことに留意されたい。対立遺伝子にわたって一貫した１：４９９の比を適用することは、そうでなければ各ドナーについて観察されるヒットの数の影響を高度に受ける性能値の安定化に役立つ。ヒットの数はドナー細胞の内在性特性よりも実験条件に大きく関連すると仮定されたので、これは妥当であると思われた。

【表１６】

表 16

MS 試料	NetMHCIIpan の PPV	*neonmhc2* の PPV
肺	0.05	>0.3
脾臓	0.025	約 0.3
Heyder LCLs	0.01	約 0.2
A375	0.01	約 0.3
mDCs	0.05	>0.44
PBMC 試料#1	0.01	約 0.3
PBMC 試料#2	0.01	>0.3
PBMC 試料#3	0.025	>0.3
PBMC 試料#4	0.0025	約 0.275
Ritz DOHH2	0.01	0.3
Ritz Maver1	0.015	約 0.175

【０８６５】

ＤＣにより提示された腫瘍ペプチドのＳＩＬＡＣに基づく同定、図３３Ａに関連する

単球由来樹状細胞（ｍＤＣ）を生成するために、ＣＤ１４＋単球を、健康ドナー末梢血単球（ＰＢＭＣ）から、ヒトＣＤ１４マイクロビーズを製造者のプロトコール（Ｍｉｌｔｅｎｙｉ Ｂｉｏｔｅｃ）の通り使用し、磁気分離によって単離した。単離された細胞を、８００Ｕ／ｍｌのｒｈＧＭ－ＣＳＦおよび４００Ｕ／ｍｌのｒｈＩＬ－４（ＣｅｌｌＧｅｎｉｘ、Ｇｅｒｍａｎｙ）を補充したＣｅｌｌＧｅｎｉｘ ＧＭＰ ＤＣ培地中で６日間にわたって分化させた。Ｋ５６２細胞（ＡＴＣＣ、Ｍａｎａｓｓａｓ、ＶＡ）を、細胞培養物中、アミノ酸を伴う安定な同位元素標識を使用して同位体標識した（ＳＩＬＡＣ）。細胞を、重同位元素アミノ酸、Ｌ－リシン２ＨＣｌ１３Ｃ６１５Ｎ２（Ｌｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ、Ｃａｒｌｓｂａｄ、ＣＡ）およびＬ－ロイシン１３Ｃ６（Ｌｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ、Ｃａｒｌｓｂａｄ、ＣＡ）を１５％の熱失活透析ウシ胎仔血清（ＴｈｅｒｍｏＦｉｓｈｅｒ）と共に含有するＳＩＬＡＣ（ＴｈｅｒｍｏＦｉｓｈｅｒ）についてＲＰＭＩ １６４０培地の存在下で５倍に成長させた。ＳＩＬＡＣ標識されたＫ５６２細胞を、６０μＭの次亜塩素酸（ＨＯＣｌ）を使用して以前に記載されている通り溶解させた、または室温で３時間にわたってＵＶ処理してアポトーシスを誘導し、終夜静置した。７５００万個のｍＤＣをＵＶ処理したＳＩＬＡＣ標識Ｋ５６２細胞と１：３の比で、３７℃で１４時間共培養したか、またはＨＯＣｌを用いて溶解させたＫ５６２と１：３の比で、３７℃で１０分間または５時間培養した。共培養後、プロテオミクス解析のために、細胞を回収し、ペレット化し、液体窒素で急速冷凍した。

【０８６６】

ＤＣにより提示された腫瘍ペプチドの予測および発現解析、図３３Ｂおよび図２１Ｃに関連する

重標識された（腫瘍由来）ペプチドの予測についてのＰＰＶを算出するために、図３３Ｂにおいて使用されているものと同じモデルおよび評価手法を使用した。Ｋ５６２細胞株についての発現を、ＥＮＣＯＤＥ（ｅｎｃｏｄｅｐｒｏｊｅｃｔ．ｏｒｇ／ｅｘｐｅｒｉｍｅｎｔｓ／ＥＮＣＳＲ５４５ＤＫＹ／；ライブラリーＥＮＣＬＢ０７５ＧＥＫおよびＥＮＣＬＢ３６５ＡＵＹ；（ＥＮＣＯＤＥ Ｐｒｏｊｅｃｔ Ｃｏｎｓｏｒｔｉｕｍ、２０１２））からのデータに基づいて決定した。樹状細胞についての発現をＧＳＥ１１６４１２

（受託ＧＳＭ３２３１１０２、ＧＳＭ３２３１１１１、ＧＳＭ３２３１１２１、ＧＳＭ３２３１１３３、ＧＳＭ３２３１１４５の平均）に基づいて決定した。

【０８６７】

（実施例１４）

トリプシンペプチドを用いたＦＡＩＭＳのベンチマーキング

本実施例では、高磁場非対称波形イオン移動度分光分析（ＦＡＩＭＳ）の使用のための標準のＨＬＡ－ペプチドームワークフローを記載する。Ａ３７５細胞により内因性にプロセシングされ、提示されたＨＬＡクラスＩおよびＨＬＡクラスＩＩペプチドを特徴付けた。ペプチドを酸性逆相（ａＲＰ）および塩基性逆相（ｂＲＰ）オフライン分画の両方に供した後、ＦＡＩＭＳ　Ｐｒｏインターフェースを備えていない（－）および備えた（＋）Ｔｈｅｒｍｏ　Ｓｃｉｅｎｔｉｆｉｃ　Ｏｒｂｉｔｒａｐ　Ｆｕｓｉｏｎ　Ｌｕｍｏｓ　Ｔｒｉｂｒｉｄ質量分析計を使用してｎＬＣ－ＭＳ／ＭＳによる分析を行った。図４２Ａに示されている図にワークフローが示されている。図４２Ｂは、ＦＡＩＭＳによりトリプシン試料におけるペプチド検出が１０ｎｇの低さまで改善されることを示す結果を示す。ＦＡＩＭＳにより、ＭＳ１強度が低いにもかかわらずＨＬＡ－１およびＨＬＡクラスＩＩペプチド検出がＬＣ勾配全体を通して増大する（図４３Ａおよび図４４Ａ、それぞれＨＬＡ－１およびＨＬＡクラスＩＩペプチドについてのデータ）。著しいことに、ＦＡＩＭＳ評価を用いて、酸性および塩基性逆相試料において独特のペプチド検出の増大が観察される（図４３Ｂ、および図４４Ｂ、それぞれＨＬＡ－１およびＨＬＡクラスＩＩペプチドについてのデータ）。この試験により、ＦＡＩＭＳを用いるとＭＳ１強度が低いにもかかわらずＨＬＡクラスＩおよびＨＬＡクラスＩＩペプチドの検出がＬＣ勾配全体を通して増大することが示された。図４５および図４６（それぞれＨＬＡ－１およびＨＬＡクラスＩＩペプチド）に示されている通り、オフライン分画とＦＡＩＭＳを組み合わせることにより、ＨＬＡペプチドレパートリーの解析の深さが増大する。

【０８６８】

（実施例１５）

示差走査蛍光光度法（ＤＳＦ）ペプチド交換アッセイ

ペプチド交換アッセイを以下の通り実施した：以下の試薬（表１７）を組み合わせ、３７℃で１８時間にわたって混合した。

【表１７】

表 17

成分	ストック濃度	最終濃度
トロンビンにより消化される DRB1*15:01 (CLIP0)、「DR15」	85.6mM	5 mM
交換ペプチド/目的のペプチド[A]	10mM	100 mM
酢酸ナトリウム	1M	100mM
塩化ナトリウム	5M	50mM
オクチルグルコシド	10%	1%
MiliQ 水	-	最大 100 ml

[A]DMSO または以下の配列を有するペプチドを使用した: PPIDGYPNHPCFEPE (配列番号 31)(M230), PQILPYPAPEEAQEN (配列番号 32) (M231), PQLRQWWAQGADPLA (配列番号 33) (M247), LLRPGQIVAFDSTAQ (配列番号 34) (M248) または ASLRSWPSTWAPWAS (配列番号 35) (M371。

【０８６９】

次いで、ＰＤ ｍｉｎｉｔｒａｐ Ｇ－２５脱塩カラムを使用して緩衝剤を交換した。Ｓｙｐｒｏオレンジ色色素（Ｆｉｓｈｅｒ Ｓ６６５１）を１００％ＤＭＳＯ中に１０００×まで希釈した。１００×のＳｙｐｒｏオレンジ色色素のワーキングストック５０μＬを脱塩緩衝剤中に調製した。１００×ｓｙｐｒｏオレンジ色色素２μＬおよび脱塩ペプチド交換試料１８μＬを３８４白色ＰＣＲマイクロプレートのウェルに移し、混合した。次いで、プレートを透明なプレートシーラーで覆い、プレートをＲｏｃｈｅ ｌｉｇｈｔｃｙｃｌｅｒ ４８０において以下のプログラムに供した：（１）２５℃まで加熱、１０秒間保持する；（２）温度を９９℃まで上昇させ、プレートを１℃当たり２０回読み取る（３）温度を２５℃まで下げ、１０秒間保持する。次いで、融解温度を算出した。例示的な結果を以下の表１８に示す。

【表１８】

表 18

試料	DT_m (試料 T_m-対照 T_m)	平均 T_m	傾き	初期	低	ピーク
DR15 + M230	21.4	79.7	0.5	2.8	2.0	4.8
DR15 + M231	15.6	73.9	0.3	2.7	1.9	5.8
DR15 + M247	20.1	78.4	0.4	3.0	2.3	5.0
DR15 + M248	5.6	63.9	0.4	1.8	1.4	5.1
DR15 + M371	19.9	78.2	0.4	2.5	1.9	4.9
DR15 + DMSO (対照)	NA	58.3	0.8	2.1	1.7	9.2

特定の実施形態では、例えば、以下が提供される：

（項目１）

（ａ）複数の提示予測を生成するために、機械学習ＨＬＡペプチド提示予測モデルを使用して複数の候補ペプチド配列のアミノ酸情報を処理するステップであって、前記複数の候補ペプチド配列の各候補ペプチド配列が、対象のゲノムもしくはエクソーム、または前記対象における病原体もしくはウイルスによってコードされ、前記複数の提示予測が、前記複数の候補ペプチド配列の各々についてのＨＬＡ提示予測を含み、各ＨＬＡ提示予測により、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質により前記複数の候補ペプチド配列のうちの所与の候補ペプチド配列が提示され得る可能性が示され、

前記機械学習ＨＬＡペプチド提示予測モデルが、訓練用細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された訓練用ペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップと、

（ｂ）少なくとも前記複数の提示予測に基づいて、前記複数のペプチド配列のうち、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる前記１つまたは複数のタンパク質のうちの少なくとも１つによって提示されるペプチド配列を同定するステップと、

を含み、前記機械学習ＨＬＡペプチド提示予測モデルの陽性適中率（ＰＰＶ）が、提示ＰＰＶ決定法に従うと少なくとも０．０７である、方法。

（項目２）

（ａ）複数の結合予測を生成するために、機械学習ＨＬＡペプチド結合予測モデルを使用して対象のゲノムまたはエクソームによってコードされる複数のペプチド配列のアミノ酸情報を処理するステップであって、前記複数の結合予測が、複数の候補ペプチド配列の各々についてのＨＬＡ結合予測を含み、各結合予測により、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる１つまたは複数のタンパク質が前記複数の候補ペプチド配列のうちの所与の候補ペプチド配列に結合する可能性が示され、

前記機械学習ＨＬＡペプチド結合予測モデルが、ＨＬＡクラスＩＩタンパク質またはＨＬＡクラスＩＩタンパク質類似体に結合することが同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップと、

（ｂ）少なくとも前記複数の結合予測に基づいて、前記複数のペプチド配列のうち、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる前記１つまたは複数のタンパク質のうちの少なくとも１つとの結合に関する閾値結合予測確率値よりも大きな確率を有するペプチド配列を同定するステップと

を含み、

前記機械学習ＨＬＡペプチド結合予測モデルの陽性適中率（ＰＰＶ）が、結合ＰＰＶ決

定法に従うと少なくとも０．１である、方法。
（項目３）
前記機械学習ＨＬＡペプチド提示予測モデルが、訓練用細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された訓練用ペプチドの配列の配列情報を含む訓練用データを使用して訓練される、項目２に記載の方法。
（項目４）
前記提示予測に基づいて、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる前記１つまたは複数のタンパク質のうちの少なくとも１つによって提示されることが同定された少なくとも２つのペプチドに順位を付けることを含む、前記項目のいずれか一項に記載の方法。
（項目５）
前記２つまたはそれよりも多くの順位付けされたペプチドのうちの１つまたは複数のペプチドを選択することを含む、前記項目のいずれか一項に記載の方法。
（項目６）
前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる前記１つまたは複数のタンパク質のうちの少なくとも１つによって提示されることが同定された前記複数のうちの１つまたは複数のペプチドを選択することを含む、前記項目のいずれか一項に記載の方法。
（項目７）
前記提示予測に基づいて順位付けされた２つまたはそれよりも多くのペプチドのうちの１つまたは複数のペプチドを選択することを含む、前記項目のいずれか一項に記載の方法。
（項目８）
複数の試験ペプチド配列のアミノ酸情報を処理して、複数の試験提示予測を生成し、各試験提示予測により、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる前記１つまたは複数のタンパク質により前記複数の試験ペプチド配列のうちの所与の試験ペプチド配列が提示され得る可能性が示される場合に、前記機械学習ＨＬＡペプチド提示予測モデルの陽性適中率（ＰＰＶ）が少なくとも０．０７であり、前記複数の試験ペプチド配列が、（ｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された少なくとも１つのヒットペプチド配列と、（ｉｉ）生物体のゲノムによってコードされるタンパク質内に含有される少なくとも４９９のデコイペプチド配列とを含む少なくとも５００の試験ペプチド配列を含み、前記生物体と前記対象が同じ種であり、前記複数の試験ペプチド配列が、前記少なくとも１つのヒットペプチド配列と前記少なくとも４９９のデコイペプチド配列を１：４９９の比で含み、前記機械学習ＨＬＡペプチド提示予測モデルにより、前記複数の試験ペプチド配列の上位パーセンテージが、細胞において発現される前記ＨＬＡタンパク質によって提示されることが予測される、前記項目のいずれか一項に記載の方法。
（項目９）
複数の試験ペプチド配列のアミノ酸情報を処理して、複数の試験結合予測を生成し、各試験結合予測により、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる前記１つまたは複数のタンパク質が前記複数の試験ペプチド配列のうちの所与の試験ペプチド配列に結合する可能性が示される場合に、前記機械学習ＨＬＡペプチド提示予測モデルの陽性適中率（ＰＰＶ）が少なくとも０．１であり、前記複数の試験ペプチド配列が、（ｉ）細胞において発現されるＨＬＡタンパク質によって提示されることが質量分析によって同定された少なくとも１つのヒットペプチド配列と、（ｉｉ）細胞において発現されるＨＬＡタンパク質、例えば、細胞において発現される単一ＨＬＡタンパク質（例えば、単一対立遺伝子細胞）によって提示されることが質量分析によって同定された少なくとも１つのペプチド配列を含むタンパク質内に含有される少なくとも１９のデコイペプチド配列とを含む少なくとも２０の試験ペプチド配列を含み、前記複数の試験ペプチド配列が、前記少なくとも１つのヒットペプチド配列と前記少なくとも１９のデコイペプチド配列を１：１９の比で含み、前記機械学習ＨＬＡペプチド提示予測モデルにより、前記複数

の試験ペプチド配列の上位パーセンテージが細胞において発現されるHLAタンパク質に結合することが予測される、前記項目のいずれか一項に記載の方法。
（項目10）
前記少なくとも1つのヒットペプチド配列と前記デコイペプチド配列の間にアミノ酸配列の重複が存在しない、前記項目のいずれか一項に記載の方法。
（項目11）
前記機械学習HLAペプチド提示予測モデルの陽性適中率（PPV）が、少なくとも0.08、0.09、0.1、0.11、0.12、0.13、0.14、0.15、0.16、0.17、0.18、0.19、0.2、0.21、0.22、0.23、0.24、0.25、0.26、0.27、0.28、0.29、0.3、0.31、0.32、0.33、0.34、0.35、0.36、0.37、0.38、0.39、0.4、0.41、0.42、0.43、0.44、0.45、0.46、0.47、0.48、0.49、0.5、0.51、0.52、0.53、0.54、0.55、0.56、0.57、0.58、0.59、0.6、0.61、0.62、0.63、0.64、0.65、0.66、0.67、0.68、0.69、0.7、0.71、0.72、0.73、0.74、0.75、0.76、0.77、0.78、0.79、0.8、0.81、0.82、0.83、0.84、0.85、0.86、0.87、0.88、0.89、0.9、0.91、0.92、0.93、0.94、0.95、0.96、0.97、0.98または0.99である、項目1から10のいずれか一項に記載の方法。
（項目12）
前記少なくとも1つのヒットペプチド配列が、少なくとも5、10、20、50または100のヒットペプチド配列を含む、項目1および項目3から11のいずれか一項に記載の方法。
（項目13）
前記少なくとも499のデコイペプチド配列が、少なくとも2500、5000、10000、25000、50000または100000のデコイペプチド配列を含む、項目1および項目3から12のいずれか一項に記載の方法。
（項目14）
前記少なくとも500の試験ペプチド配列が、少なくとも2500、5000、10000、25000、50000または100000の試験ペプチド配列を含む、項目1および項目3から13のいずれか一項に記載の方法。
（項目15）
前記上位パーセンテージが、上位0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1.00%または2.00%である、項目1および項目3から14のいずれか一項に記載の方法。
（項目16）
前記少なくとも1つのヒットペプチド配列が、少なくとも5、10、20、50または100のヒットペプチド配列を含む、項目2から11のいずれか一項に記載の方法。
（項目17）
前記少なくとも19のデコイペプチド配列が、少なくとも500、1000、2000、5000または10000のデコイペプチド配列を含む、項目2から11および16のいずれか一項に記載の方法。
（項目18）
前記少なくとも20の試験ペプチド配列が、少なくともを含み、前記少なくとも500の試験ペプチド配列が、少なくとも500、1000、2000、5000または10000の試験ペプチド配列試験ペプチド配列を含む、項目2から11、16および17のいずれか一項に記載の方法。
（項目19）
前記上位パーセンテージが、上位5%、6%、7%、8%、9%、10%、15%、または20%である、項目2から11および16から18のいずれか一項に記載の方法。

（項目２０）
前記ＰＰＶが、表１１の対応するＨＬＡ対立遺伝子によってコードされるタンパク質についての表１１の２列目のＰＰＶよりも大きい、前記項目のいずれか一項に記載の方法。
（項目２１）
前記ＰＰＶが、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質についての表１６の２列目のＰＰＶよりも大きい、前記項目のいずれか一項に記載の方法。
（項目２２）
前記対象が、単一の対象である、前記項目のいずれか一項に記載の方法。
（項目２３）
前記細胞が、前記対象の細胞のクラスＩＩ　ＨＬＡ対立遺伝子によってコードされる単一のタンパク質を発現する細胞である、前記項目のいずれか一項に記載の方法。
（項目２４）
前記細胞が、単一対立遺伝子ＨＬＡ細胞またはアフィニティータグを有するＨＬＡ対立遺伝子を発現する細胞である、項目１から２３のいずれか一項に記載の方法。
（項目２５）
前記複数のペプチド配列の各ペプチド配列が、がんに関連付けられる、前記項目のいずれか一項に記載の方法。
（項目２６）
前記複数のペプチド配列のうちの少なくとも１つのペプチド配列が、前記対象のがん細胞によって過剰発現される、前記項目のいずれか一項に記載の方法。
（項目２７）
前記複数のペプチド配列の各ペプチド配列が、前記対象のがん細胞によって過剰発現される、前記項目のいずれか一項に記載の方法。
（項目２８）
前記複数のペプチド配列のうちの少なくとも１つのペプチド配列が、がん細胞特異的ペプチドである、前記項目のいずれか一項に記載の方法。
（項目２９）
前記複数のペプチド配列の各ペプチド配列が、がん細胞特異的ペプチドである、前記項目のいずれか一項に記載の方法。
（項目３０）
前記複数のペプチド配列の各ペプチド配列が、前記対象のがん細胞によって発現される、前記項目のいずれか一項に記載の方法。
（項目３１）
前記複数のペプチド配列のうちの少なくとも１つのペプチド配列が、前記対象の非がん細胞によってコードされない、前記項目のいずれか一項に記載の方法。
（項目３２）
前記複数のペプチド配列の各ペプチド配列が、前記対象の非がん細胞によってコードされない、前記項目のいずれか一項に記載の方法。
（項目３３）
前記複数のペプチド配列のうちの少なくとも１つのペプチド配列が、前記対象の非がん細胞では発現されない、前記項目のいずれか一項に記載の方法。
（項目３４）
前記複数のペプチド配列の各ペプチド配列が、前記対象の非がん細胞では発現されない、前記項目のいずれか一項に記載の方法。
（項目３５）
前記対象の前記複数のペプチド配列を得ることを含む、前記項目のいずれか一項に記載の方法。
（項目３６）
前記対象の複数のポリヌクレオチド配列を得ることを含む、前記項目のいずれか一項に記載の方法。

（項目３７）
対象のゲノムまたはエクソームによってコードされる複数のペプチド配列をコードする前記対象の複数のポリヌクレオチド配列を得ることを含む、前記項目のいずれか一項に記載の方法。
（項目３８）
対象のゲノムまたはエクソームによってコードされる複数のペプチド配列をコードする前記対象の複数のポリヌクレオチド配列をコンピュータプロセッサによって得ることを含む、前記項目のいずれか一項に記載の方法。
（項目３９）
前記対象の複数のポリヌクレオチド配列をゲノム配列決定またはエクソーム配列決定によって得ることを含む、前記項目のいずれか一項に記載の方法。
（項目４０）
前記対象の複数のポリヌクレオチド配列を全ゲノム配列決定または全エクソーム配列決定によって得ることを含む、前記項目のいずれか一項に記載の方法。
（項目４１）
前記対象に、ペプチド配列の選択されたサブセットの１つまたは複数を含む組成物を投与するステップをさらに含む、前記項目のいずれか一項に記載の方法。
（項目４２）
前記ＨＬＡクラスＩＩタンパク質が、ＨＬＡ－ＤＲタンパク質、ＨＬＡ－ＤＱタンパク質、またはＨＬＡ－ＤＰタンパク質を含む、前記項目のいずれか一項に記載の方法。
（項目４３）
前記ＨＬＡクラスＩＩタンパク質が、ＨＬＡ－ＤＰＢ１＊０１：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０２：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０３：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０２／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０６：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＱＢ１＊０２：０１／ＨＬＡ－ＤＱＡ１＊０５：０１、ＨＬＡ－ＤＱＢ１＊０２：０２／ＨＬＡ－ＤＱＡ１＊０２：０１、ＨＬＡ－ＤＱＢ１＊０６：０２／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＱＢ１＊０６：０４／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０１：０１、ＨＬＡ－ＤＲＢ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０３：０１、ＨＬＡ－ＤＲＢ１＊０３：０２、ＨＬＡ－ＤＲＢ１＊０４：０１、ＨＬＡ－ＤＲＢ１＊０４：０２、ＨＬＡ－ＤＲＢ１＊０４：０３、ＨＬＡ－ＤＲＢ１＊０４：０４、ＨＬＡ－ＤＲＢ１＊０４：０５、ＨＬＡ－ＤＲＢ１＊０４：０７、ＨＬＡ－ＤＲＢ１＊０７：０１、ＨＬＡ－ＤＲＢ１＊０８：０１、ＨＬＡ－ＤＲＢ１＊０８：０２、ＨＬＡ－ＤＲＢ１＊０８：０３、ＨＬＡ－ＤＲＢ１＊０８：０４、ＨＬＡ－ＤＲＢ１＊０９：０１、ＨＬＡ－ＤＲＢ１＊１０：０１、ＨＬＡ－ＤＲＢ１＊１１：０１、ＨＬＡ－ＤＲＢ１＊１１：０２、ＨＬＡ－ＤＲＢ１＊１１：０４、ＨＬＡ－ＤＲＢ１＊１２：０１、ＨＬＡ－ＤＲＢ１＊１２：０２、ＨＬＡ－ＤＲＢ１＊１３：０１、ＨＬＡ－ＤＲＢ１＊１３：０２、ＨＬＡ－ＤＲＢ１＊１３：０３、ＨＬＡ－ＤＲＢ１＊１４：０１、ＨＬＡ－ＤＲＢ１＊１５：０１、ＨＬＡ－ＤＲＢ１＊１５：０２、ＨＬＡ－ＤＲＢ１＊１５：０３、ＨＬＡ－ＤＲＢ１＊１６：０１、ＨＬＡ－ＤＲＢ３＊０１：０１、ＨＬＡ－ＤＲＢ３＊０２：０２、ＨＬＡ－ＤＲＢ３＊０３：０１、ＨＬＡ－ＤＲＢ４＊０１：０１およびＨＬＡ－ＤＲＢ５＊０１：０１；ＤＰＡ＊０１：０３／ＤＰＢ＊０４：０１、ＤＲＢ１＊０１：０１、ＤＲＢ１＊０１：０２、ＤＲＢ１＊０３：０１、ＤＲＢ１＊０４：０１、ＤＲＢ１＊０４：０２、ＤＲＢ１＊０４：０４、ＤＲＢ１＊０４：０５、ＤＲＢ１＊０７：０１、ＤＲＢ１＊０８：０１、ＤＲＢ１＊０８：０２、ＤＲＢ１＊０８：０３、ＤＲＢ１＊０９：０１、ＤＲＢ１＊１１：０１、ＤＲＢ１＊１１：０２、ＤＲＢ１＊１１：０４、ＤＲＢ１＊１２：０１、ＤＲＢ１＊１３：０１、ＤＲＢ１＊１３：０２、ＤＲＢ１＊１３：０３、ＤＲＢ１＊１４：０１、ＤＲＢ１＊１５：０１、ＤＲＢ１＊１５：０２、ＤＲＢ１＊１５：０３、ＤＲＢ１＊１６：０２、ＤＲＢ３＊０１：０１、ＤＲＢ３＊０２：０１、ＤＲＢ３＊０２：０２、ＤＲＢ３＊０３：０１、ＤＲＢ４＊０１：０１、

ＤＲＢ４＊０１：０３およびＤＲＢ５＊０１：０１からなる群から選択されるＨＬＡ－ＤＲタンパク質；ＤＰＢ１＊０１：０１、ＤＰＢ１＊０２：０１、ＤＰＢ１＊０２：０２、ＤＰＢ１＊０３：０１、ＤＰＢ１＊０４：０１、ＤＰＢ１＊０４：０２、ＤＰＢ１＊０５：０１、ＤＰＢ１＊０６：０１、ＤＰＢ１＊１１：０１、ＤＰＢ１＊１３：０１、ＤＰＢ１＊１７：０１からなる群から選択されるＨＬＡ－ＤＰタンパク質；またはＡ１＊０１：０１＋Ｂ１＊０５：０１、Ａ１＊０１：０２＋Ｂ１＊０６：０２、Ａ１＊０１：０２＋Ｂ１＊０６：０４、Ａ１＊０１：０３＋Ｂ１＊０６：０３、Ａ１＊０２：０１＋Ｂ１＊０２：０２、Ａ１＊０２：０１＋Ｂ１＊０３：０３、Ａ１＊０３：０１＋Ｂ１＊０３：０２、Ａ１＊０３：０３＋Ｂ１＊０３：０１、Ａ１＊０５：０１＋Ｂ１＊０２：０１およびＡ１＊０５：０５＋Ｂ１＊０３：０１からなる群から選択されるＨＬＡ－ＤＱタンパク質複合体、
からなる群から選択される、前記項目のいずれか一項に記載の方法。
（項目４４）
前記ＨＬＡタンパク質によって提示されるペプチドの長さが、１５～４０アミノ酸である、前記項目のいずれか一項に記載の方法。
（項目４５）
個別化がん治療が、アジュバントをさらに含む、前記項目のいずれか一項に記載の方法。
（項目４６）
前記個別化がん治療が、免疫チェックポイント阻害剤をさらに含む、前記項目のいずれか一項に記載の方法。
（項目４７）
対象に対する免疫治療のためのＨＬＡクラスＩＩ特異的ペプチドを同定する方法であって、
（ａ）コンピュータプロセッサにより、エピトープを含む候補ペプチド、およびそれぞれが前記エピトープを含む複数のペプチド配列を得るステップと、
（ｂ）免疫細胞に対する前記複数のペプチド配列の各々についての提示予測を生成するために、コンピュータプロセッサにより、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して前記複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、ＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質により前記複数のペプチド配列のうちの所与のペプチド配列が提示され得る可能性が示され、前記機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練される、ステップと、
（ｃ）前記対象の細胞の前記ＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質から、前記候補ペプチドに結合することが前記機械学習ＨＬＡ－ペプチド提示予測モデルによって予測されたタンパク質を選択するステップであって、前記タンパク質が、免疫細胞に対する前記候補ペプチドの提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップと、
（ｄ）前記候補ペプチドを前記選択されたタンパク質と接触させ、したがって、前記候補ペプチドを、前記選択されたタンパク質と会合したプレースホルダーペプチドと競合させるステップと、
（ｅ）前記候補ペプチドが前記プレースホルダーペプチドに置き換わるかどうかに基づいて、前記候補ペプチドを前記選択されたタンパク質に特異的な免疫治療のためのペプチドとして同定するステップと
を含む方法。
（項目４８）
ＭＨＣクラスＩＩ結合性ペプチドの免疫原性をアッセイするための方法であって、
（ａ）前記ＭＨＣクラスＩＩ結合性ペプチドに結合することが機械学習ＨＬＡ－ペプチド提示予測モデルによって予測された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質を選択するステップであって、前記機械学習ＨＬＡ－ペプチド提示予測モデルが、所与のペプチド配列についての提示予測を生成するように構成されており、前記提

示予測により、前記ＨＬＡクラスＩＩ対立遺伝子によってコードされる１つまたは複数のタンパク質により前記所与のペプチド配列が提示され得る可能性が示され、前記タンパク質が、前記ＭＨＣクラスＩＩ結合性ペプチドの提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップと、

（ｂ）前記ペプチドを前記選択されたタンパク質と接触させ、したがって、前記ペプチドを、前記選択されたタンパク質と会合したプレースホルダーペプチドと競合させ、前記プレースホルダーペプチドを置き換え、それにより、前記ＨＬＡクラスＩＩタンパク質と前記ＭＨＣクラスＩＩ結合性ペプチドとを含む複合体を形成するステップと、

（ｃ）前記複合体をＣＤ４＋　Ｔ細胞と接触させるステップと、

（ｄ）サイトカインの誘導、ケモカインの誘導、および細胞表面マーカーの発現からなる群から選択されるＣＤ４＋　Ｔ細胞の活性化パラメータの１つまたは複数についてアッセイするステップと

を含む方法。

（項目４９）

がん免疫治療のために対象におけるＣＤ４＋　Ｔ細胞活性化を誘導するための方法であって、

（ａ）がんに関連付けられ、がん変異を含むペプチド配列を同定するステップであって、前記ペプチド配列の同定が、前記対象のがん細胞由来のＤＮＡ、ＲＮＡ、またはタンパク質配列を前記対象の正常細胞由来のＤＮＡ、ＲＮＡ、またはタンパク質配列と比較することを含む、ステップと、

（ｂ）前記対象の細胞によって通常発現され、前記ペプチドに結合することが機械学習ＨＬＡ－ペプチド提示予測モデルによって予測された、ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質を選択するステップであって、前記予測モデルの陽性適中率が、少なくとも０．１％、０．１％～５０％または最大で５０％の再現率で少なくとも０．１であり、前記タンパク質が、前記同定されたペプチド配列の提示についての提示予測確率値の閾値よりも大きな確率を有する、ステップと、

（ｃ）前記同定されたペプチドを、前記選択された、前記ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質と接触させて、前記同定されたペプチドが、５００ｎＭ未満のＩＣ５０値で、前記選択された、前記ＨＬＡクラスＩＩ対立遺伝子によってコードされるタンパク質と会合したプレースホルダーペプチドと競合して前記プレースホルダーペプチドに置き換わるかどうかを検証するステップと、

（ｄ）必要に応じて、前記同定されたペプチドを精製するステップと、

（ｅ）前記同定されたペプチドの配列を含むポリペプチドまたは前記ポリペプチドをコードするポリヌクレオチドを前記対象に有効量で投与するステップと

を含む方法。

（項目５０）

ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、

（ａ）コンピュータプロセッサにより、前記ポリペプチド配列の複数のペプチド配列を得るステップと、

（ｂ）前記複数のペプチド配列の各々についての提示予測を生成するために、コンピュータプロセッサにより、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して前記複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、前記対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により前記複数のペプチド配列のうちの所与のペプチド配列のエピトープ配列が提示され得る可能性が示され、前記機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質に関連する配列情報を含む訓練用データを使用して訓練される、ステップと、

（ｃ）前記複数の提示予測に基づいて、前記ポリペプチド配列の前記複数のペプチド配列の各々が前記対象に対して免疫原性ではないことを決定または予測するステップと、

（ｄ）前記対象に、前記薬物を含む組成物を投与するステップと
を含む方法。
（項目５１）
４つの個々のＨＬＡタンパク質アルファ１およびベータ１ヘテロ二量体をコンジュゲートすることによってＨＬＡクラスＩＩ四量体を製造するための方法であって、
（ａ）真核細胞において、ＨＬＡタンパク質のアルファ鎖およびベータ鎖をコードする核酸配列、分泌シグナル、ビオチン化モチーフおよび同定用または精製用の少なくとも１つのタグを含むベクターを発現させ、したがって、各ＨＬＡタンパク質アルファ１およびベータ１ヘテロ二量体を、二量体を形成した状態で分泌させるステップであって、前記ヘテロ二量体が、プレースホルダーペプチドと会合している、ステップと、
（ｂ）前記分泌されたヘテロ二量体を細胞培地から精製するステップと、
（ｃ）ペプチド交換アッセイを使用して前記ペプチド結合活性を検証するステップと、
（ｄ）ストレプトアビジンを添加し、それにより、ヘテロ二量体をコンジュゲートして四量体にするステップと、
（ｅ）前記四量体を精製し、１ｍｇ／Ｌを超える収量を得るステップと
を含む方法。
（項目５２）
ＨＬＡ－ＤＲヘテロ二量体、またはＨＬＡ－ＤＰヘテロ二量体、またはＨＬＡ－ＤＱヘテロ二量体のいずれかを含むＨＬＡクラスＩＩ四量体または多量体であって、各ヘテロ二量体がアルファ鎖とベータ鎖とを含み、前記ヘテロ二量体が精製されており、１ｍｇ／Ｌを超える濃度で存在する、ＨＬＡクラスＩＩ四量体または多量体。
（項目５３）
ＨＬＡタンパク質が、ＨＬＡ－ＤＰＢ１＊０１：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０２：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０３：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０４：０２／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＰＢ１＊０６：０１／ＨＬＡ－ＤＰＡ１＊０１：０３、ＨＬＡ－ＤＱＢ１＊０２：０１／ＨＬＡ－ＤＱＡ１＊０５：０１、ＨＬＡ－ＤＱＢ１＊０２：０２／ＨＬＡ－ＤＱＡ１＊０２：０１、ＨＬＡ－ＤＱＢ１＊０６：０２／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＱＢ１＊０６：０４／ＨＬＡ－ＤＱＡ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０１：０１、ＨＬＡ－ＤＲＢ１＊０１：０２、ＨＬＡ－ＤＲＢ１＊０３：０１、ＨＬＡ－ＤＲＢ１＊０３：０２、ＨＬＡ－ＤＲＢ１＊０４：０１、ＨＬＡ－ＤＲＢ１＊０４：０２、ＨＬＡ－ＤＲＢ１＊０４：０３、ＨＬＡ－ＤＲＢ１＊０４：０４、ＨＬＡ－ＤＲＢ１＊０４：０５、ＨＬＡ－ＤＲＢ１＊０４：０７、ＨＬＡ－ＤＲＢ１＊０７：０１、ＨＬＡ－ＤＲＢ１＊０８：０１、ＨＬＡ－ＤＲＢ１＊０８：０２、ＨＬＡ－ＤＲＢ１＊０８：０３、ＨＬＡ－ＤＲＢ１＊０８：０４、ＨＬＡ－ＤＲＢ１＊０９：０１、ＨＬＡ－ＤＲＢ１＊１０：０１、ＨＬＡ－ＤＲＢ１＊１１：０１、ＨＬＡ－ＤＲＢ１＊１１：０２、ＨＬＡ－ＤＲＢ１＊１１：０４、ＨＬＡ－ＤＲＢ１＊１２：０１、ＨＬＡ－ＤＲＢ１＊１２：０２、ＨＬＡ－ＤＲＢ１＊１３：０１、ＨＬＡ－ＤＲＢ１＊１３：０２、ＨＬＡ－ＤＲＢ１＊１３：０３、ＨＬＡ－ＤＲＢ１＊１４：０１、ＨＬＡ－ＤＲＢ１＊１５：０１、ＨＬＡ－ＤＲＢ１＊１５：０２、ＨＬＡ－ＤＲＢ１＊１５：０３、ＨＬＡ－ＤＲＢ１＊１６：０１、ＨＬＡ－ＤＲＢ３＊０１：０１、ＨＬＡ－ＤＲＢ３＊０２：０２、ＨＬＡ－ＤＲＢ３＊０３：０１、ＨＬＡ－ＤＲＢ４＊０１：０１およびＨＬＡ－ＤＲＢ５＊０１：０１からなる群から選択されるＨＬＡクラスＩＩタンパク質である、項目５２に記載のＨＬＡクラスＩＩ四量体または多量体。
（項目５４）
前記項目のいずれか一項のＨＬＡタンパク質のアルファ鎖およびベータ鎖をコードする核酸配列、分泌シグナル、ビオチン化モチーフおよび同定用または精製用の少なくとも１つのタグを含み、したがって、各ＨＬＡタンパク質アルファ１およびベータ１ヘテロ二量体を、二量体を形成した状態で分泌させるベクターであって、必要に応じて、前記分泌されるヘテロ二量体が、プレースホルダーペプチドと会合している、ベクター。

（項目５５）
項目５４に記載のベクターを含む細胞。
（項目５６）
ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、
（ａ）コンピュータプロセッサにより、前記ポリペプチド配列の複数のペプチド配列を得るステップと、
（ｂ）前記複数のペプチド配列の各々についての提示予測を生成するために、コンピュータプロセッサにより、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して前記複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、前記対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により前記複数のペプチド配列のうちの所与のペプチド配列のエピトープ配列が提示され得る可能性が示され、前記機械学習ＨＬＡ－ペプチド提示予測モデルが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報を含む訓練用データを使用して訓練されるステップと、
（ｃ）前記複数の提示予測に基づいて、前記ポリペプチド配列の複数のペプチド配列のうちの少なくとも１つが前記対象に対して免疫原性であることを決定または予測するステップと
を含む方法。
（項目５７）
ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、
（ａ）コンピュータプロセッサを使用して前記ポリペプチド配列のペプチド配列のアミノ酸情報を機械学習ＨＬＡ－ペプチド提示予測モデルに入力して、前記ペプチド配列についての提示予測のセットを生成するステップであって、各提示予測により、前記対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により所与のペプチド配列のエピトープ配列が提示される確率が示され、前記機械学習ＨＬＡ－ペプチド提示予測モデルが、少なくとも訓練用データに基づいて同定された複数の予測変数を含み、前記訓練用データが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；アミノ酸位置情報を含む訓練用ペプチド配列情報であって、前記細胞において発現されるＨＬＡタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られた前記アミノ酸位置情報と前記アミノ酸位置情報および前記予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む、ステップと、
（ｂ）前記提示予測のセットに基づいて、前記ポリペプチド配列のペプチド配列の各々が前記対象に対して免疫原性ではないことを決定または予測するステップと、
（ｃ）前記対象に、前記薬物を含む組成物を投与するステップと
を含む方法。
（項目５８）
ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、
（ａ）コンピュータプロセッサを使用して前記ポリペプチド配列のペプチド配列のアミノ酸情報を機械学習ＨＬＡ－ペプチド提示予測モデルに入力して、前記ペプチド配列についての提示予測のセットを生成するステップであって、各提示予測により、前記対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により所与のペプチド配列のエピトープ配列が提示される確率が示され、前記機械学習ＨＬＡ－ペプチド提示予測モデルが、少なくとも訓練用データに基づいて同定された複数の予測変数を含み、前記訓練用データが、細胞において発現されるＨＬＡタンパク質によって提示され、質量分析によって同定されたペプチドの配列の配列情報；アミノ酸位置情報を含む訓練用ペプチド配列情報であって、前記細胞において発現されるＨＬ

Ａタンパク質に関連付けられる、訓練用ペプチド配列情報；ならびに、入力として受け取られた前記アミノ酸位置情報と前記アミノ酸位置情報および前記予測変数に基づいて出力として生成された提示可能性との関係を表す関数を含む、ステップと、
（ｂ）前記提示予測のセットに基づいて、前記ポリペプチド配列のペプチド配列のうちの少なくとも１つが前記対象に対して免疫原性であることを決定または予測するステップとを含む方法。
（項目５９）
ポリペプチド配列を含む薬物を対象における免疫原性についてスクリーニングする方法であって、
（ａ）コンピュータプロセッサにより、前記ポリペプチド配列の複数のペプチド配列を得るステップと、
（ｂ）前記複数のペプチド配列の各々についての提示予測を生成するために、コンピュータプロセッサにより、機械学習ＨＬＡ－ペプチド提示予測モデルを使用して前記複数のペプチド配列のアミノ酸情報を処理するステップであって、各提示予測により、前記対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣ対立遺伝子によってコードされる１つまたは複数のタンパク質により前記複数のペプチド配列のうちの所与のペプチド配列のエピトープ配列が提示され得る可能性が示され、前記機械学習ＨＬＡ－ペプチド提示予測モデルが、前記細胞において発現されるＨＬＡタンパク質に関連する配列情報を含む訓練用データを使用して訓練されるステップと、
（ｃ）前記複数の提示予測に基づいて、前記ポリペプチド配列の前記複数のペプチド配列の各々が前記対象に対して免疫原性ではないことを決定または予測するステップと、
（ｄ）前記対象に、前記薬物を含む組成物を投与するステップと
を含む方法。
（項目６０）
自己免疫疾患または状態を有する対象を処置する方法であって、
（ａ）前記対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣによって提示される発現されたタンパク質のエピトープを同定または予測するステップであって、前記同定または予測されたエピトープと前記クラスＩまたはクラスＩＩ　ＭＨＣとを含む複合体が、前記対象のＣＤ８　Ｔ細胞またはＣＤ４　Ｔ細胞の標的とされる、ステップと、
（ｂ）前記複合体に結合するＴ細胞受容体（ＴＣＲ）を同定するステップと、
（ｃ）前記ＴＣＲを前記対象由来の調節性Ｔ細胞または同種異系の調節性Ｔ細胞において発現させるステップと、
（ｄ）前記ＴＣＲを発現する前記調節性Ｔ細胞を前記対象に投与するステップと
を含む方法。
（項目６１）
自己免疫疾患または状態を有する対象を処置する方法であって、前記対象に、（ｉ）前記対象の細胞のクラスＩまたはクラスＩＩ　ＭＨＣによって提示されることが同定または予測された発現されたタンパク質のエピトープと（ｉｉ）前記クラスＩまたはクラスＩＩ　ＭＨＣとを含む複合体であって、前記対象のＣＤ８Ｔ細胞またはＣＤ４　Ｔ細胞の標的とされる、複合体に結合するＴ細胞受容体（ＴＣＲ）を発現する調節性Ｔ細胞を投与するステップを含む方法。

【図面】

【図１Ａ】

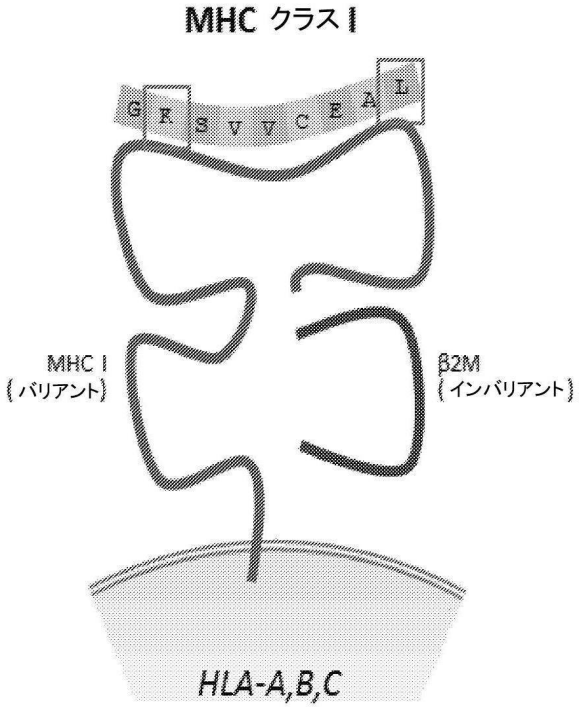


FIG. 1A

【図１Ｂ】

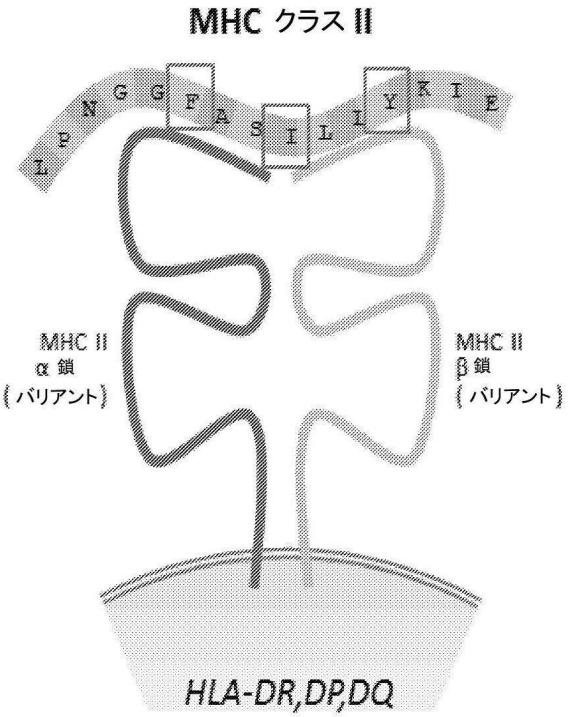


FIG. 1B

【図２】

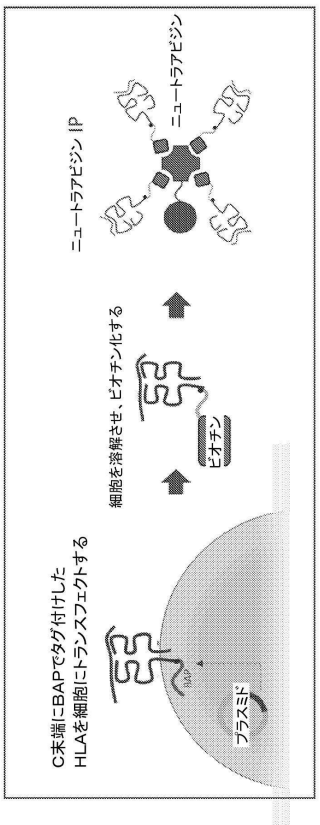


FIG. 2

【図３】

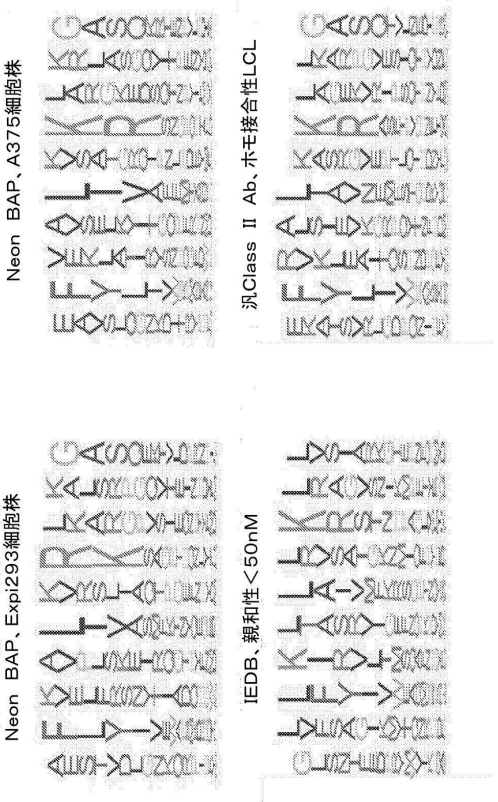


FIG. 3

【図４】

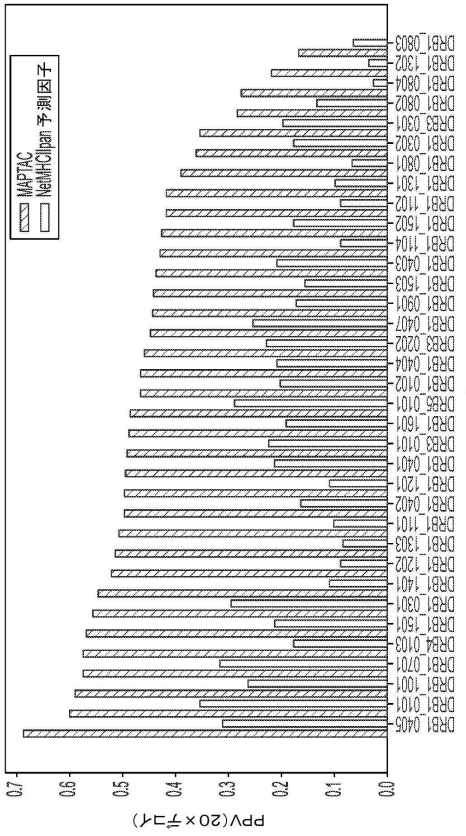


FIG. 4

【図５】

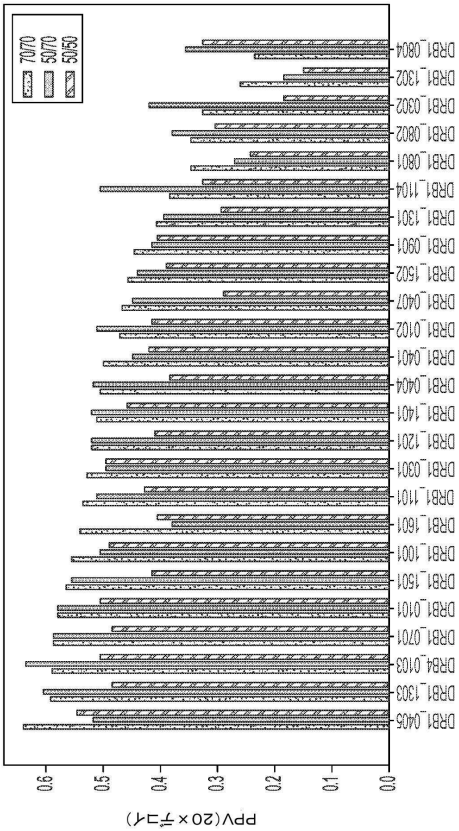


FIG. 5

【図６】

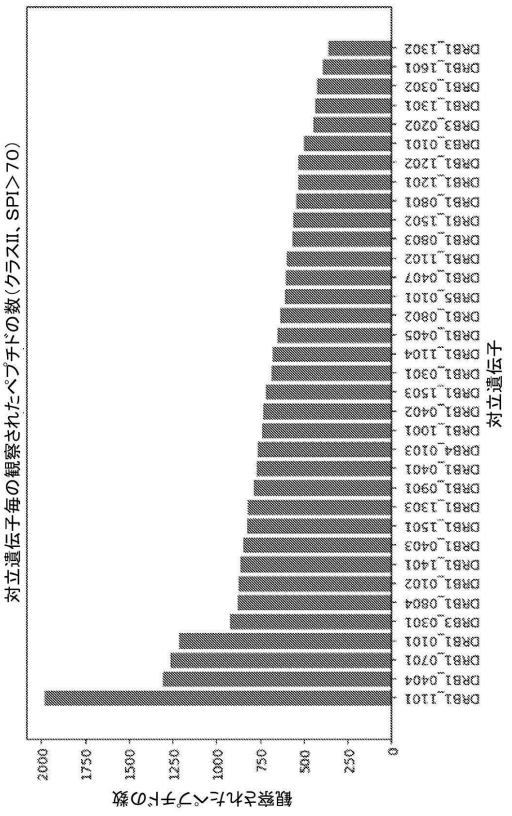


FIG. 6

【図７Ａ】

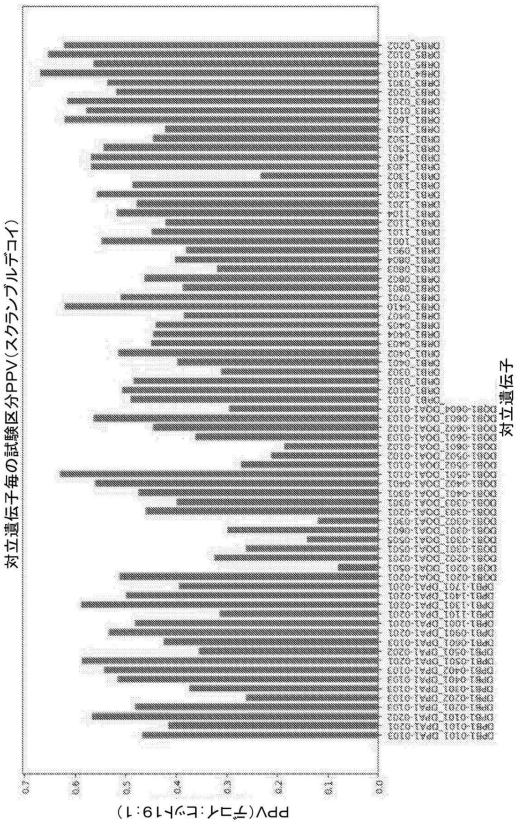


FIG. 7A

【図７Ｂ－１】

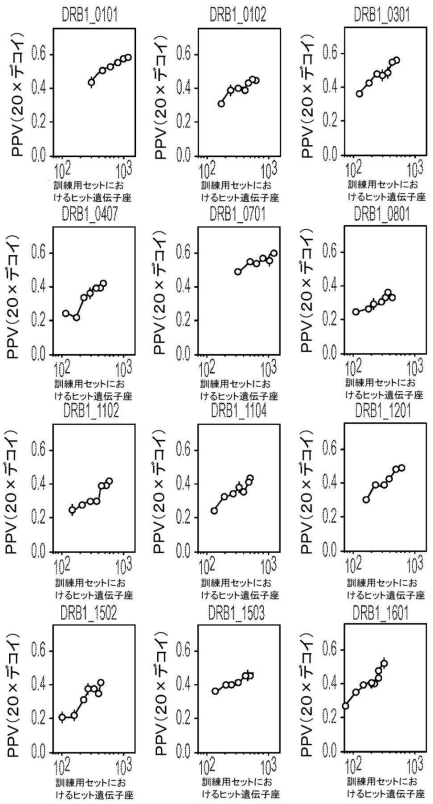


FIG. 7B

【図７Ｂ－２】

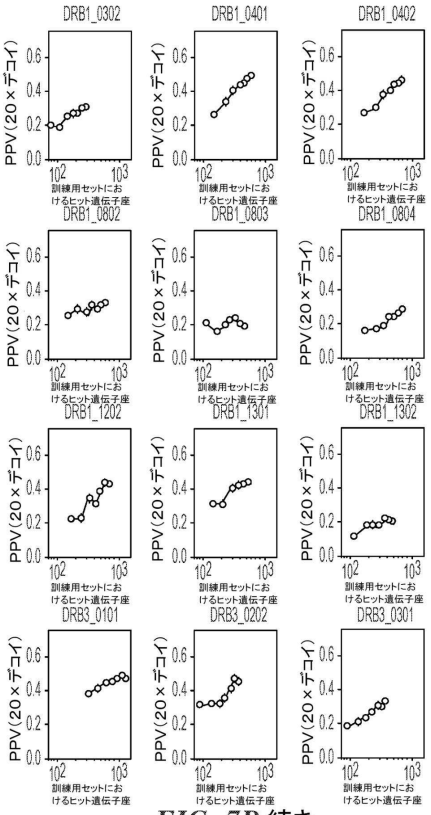


FIG. 7B 続き

【図７Ｂ－３】

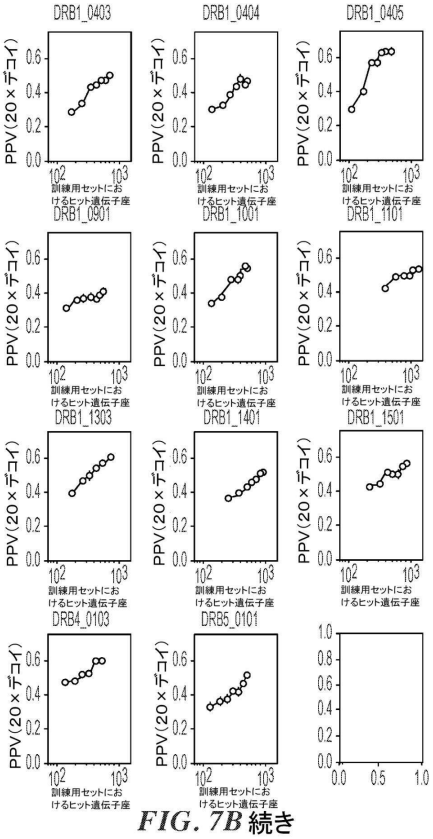


FIG. 7B 続き

【図８】

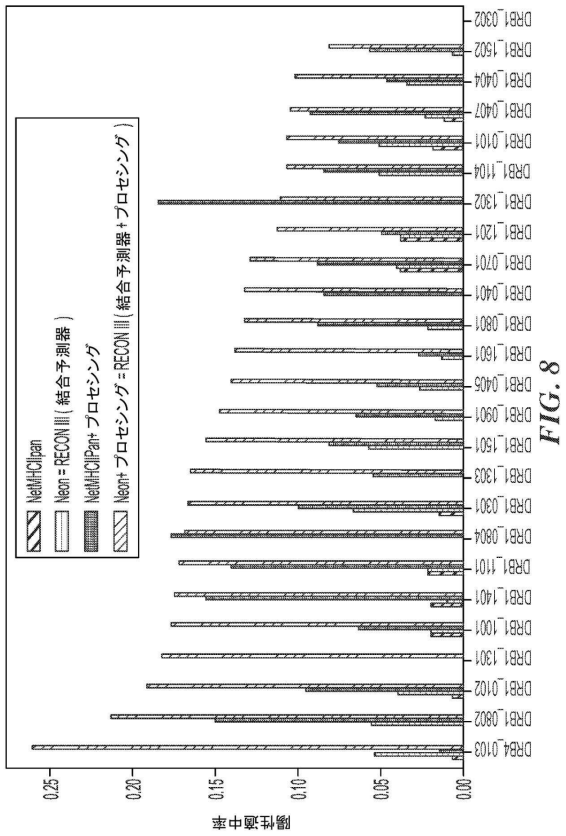


FIG. 8

【図９】

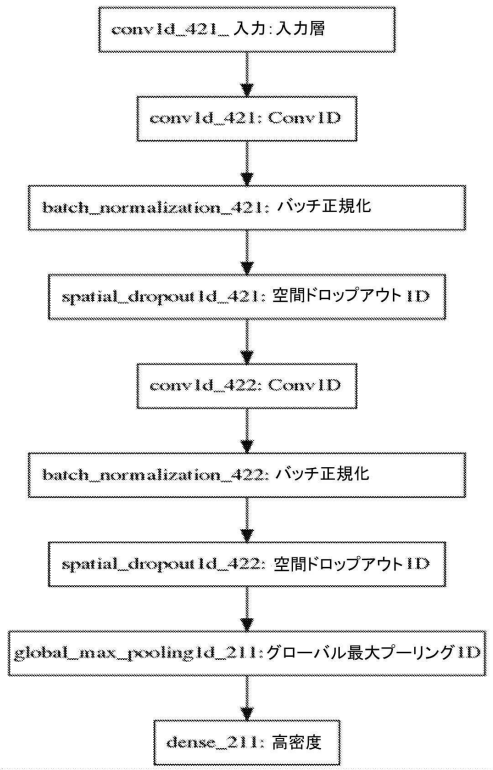


FIG. 9

【図１０】

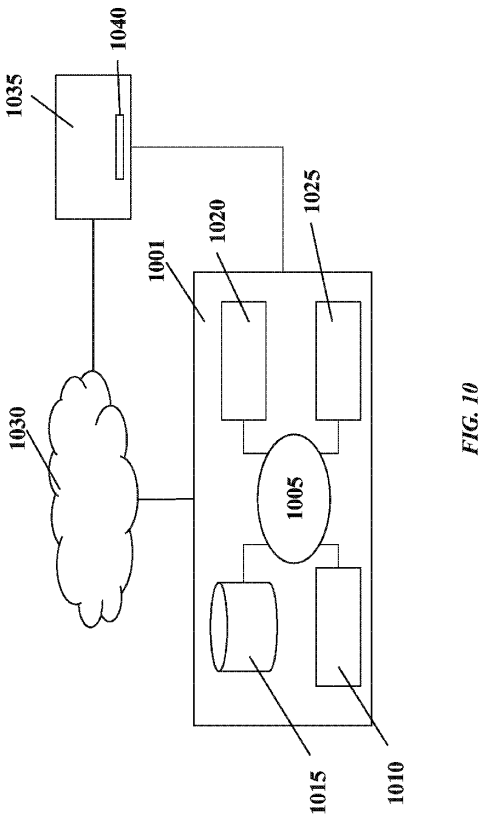


FIG. 10

【図１１Ａ】

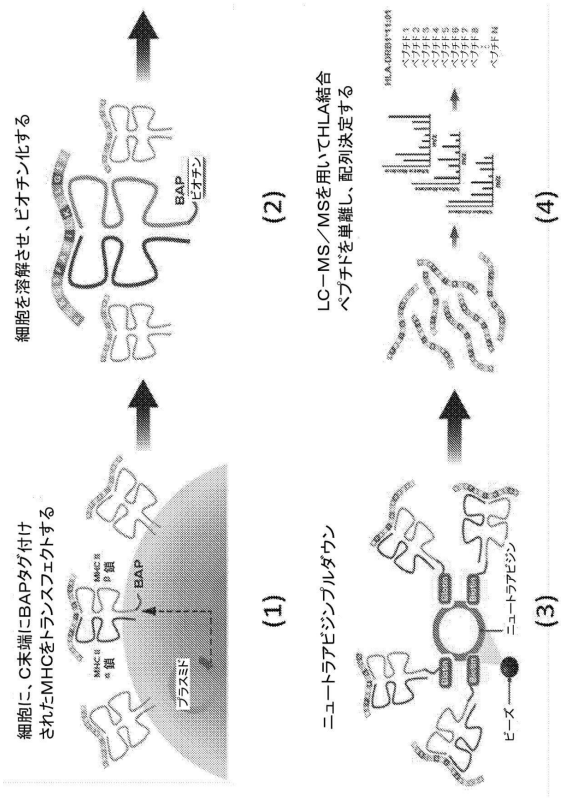


FIG. 11A

【図１１Ｂ】

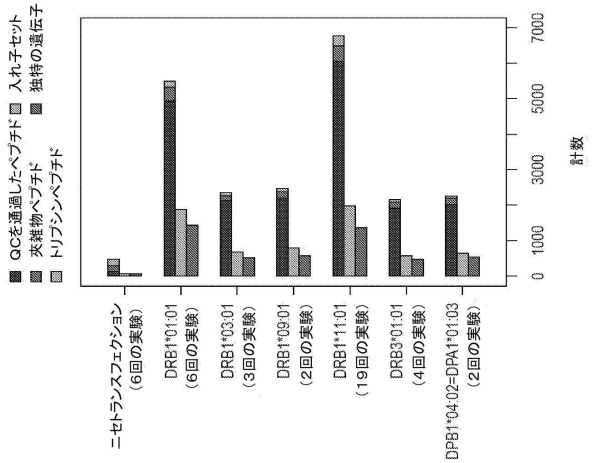


FIG. 11B

【図１１Ｃ】

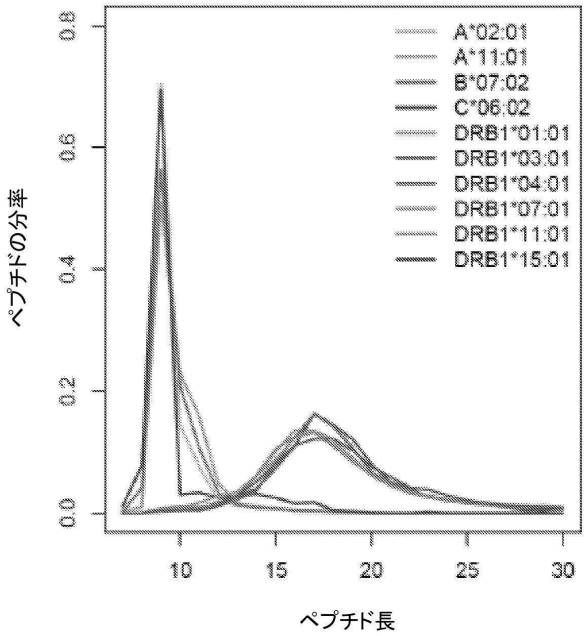


FIG. 11C

【図１１Ｄ】

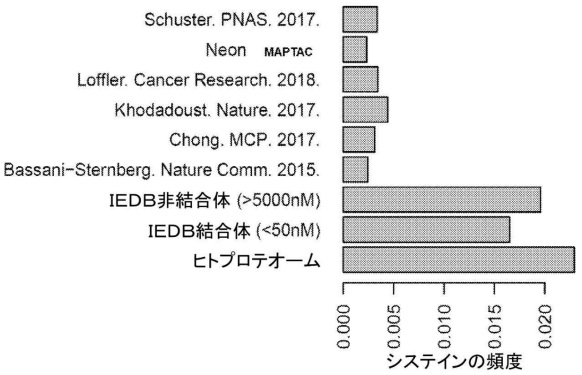


FIG. 11D

【図１２Ａ】

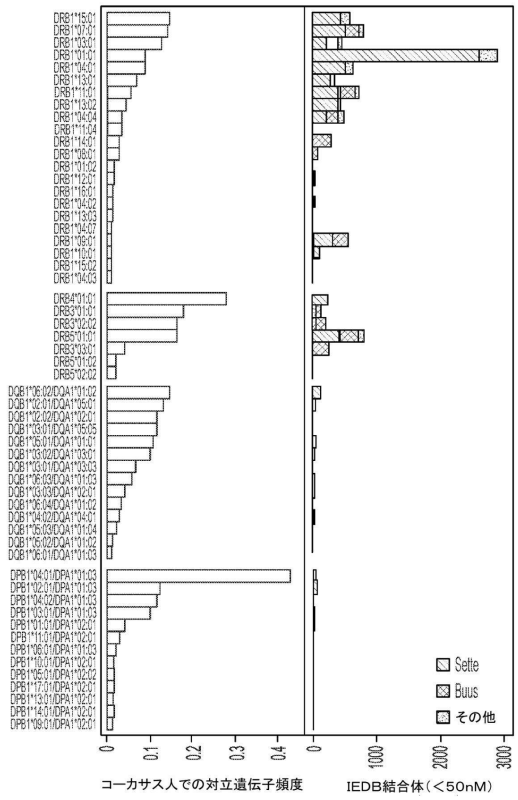


FIG. 12A

【図１２Ｂ】

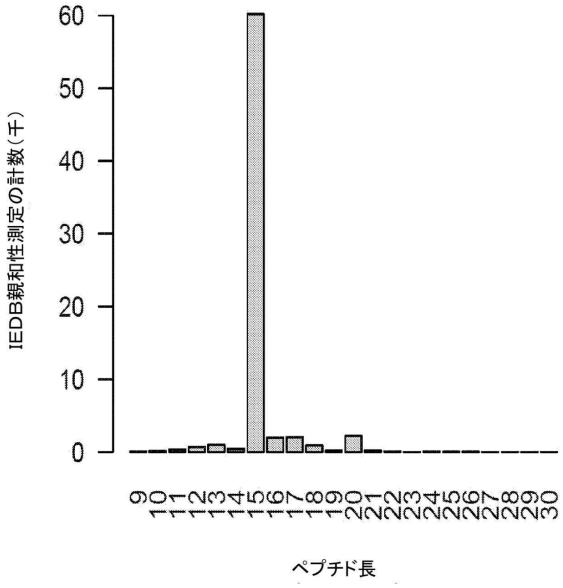


FIG. 12B

【図１２Ｃ】

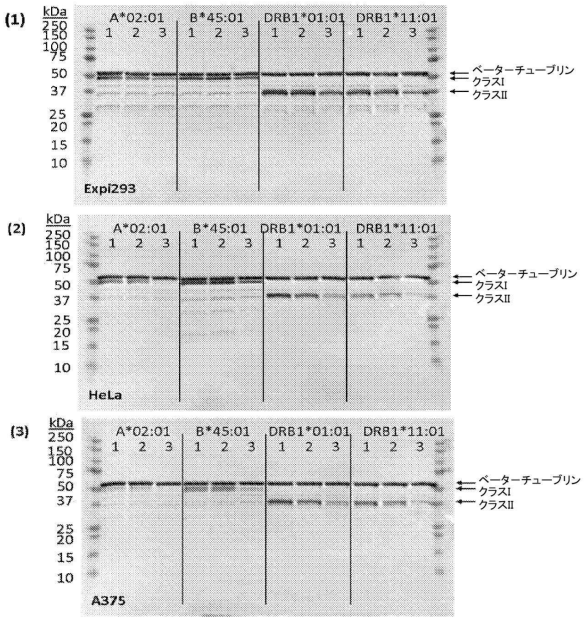


FIG. 12C

【図１２Ｄ】

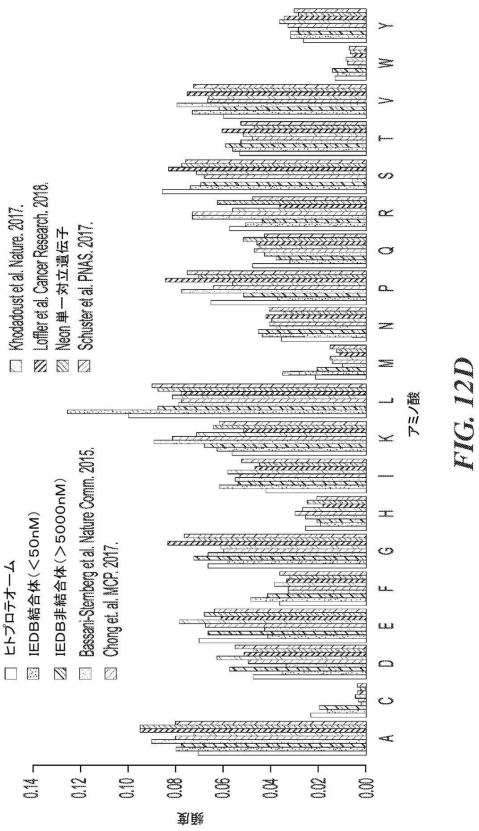


FIG. 12D

【図１２Ｅ】

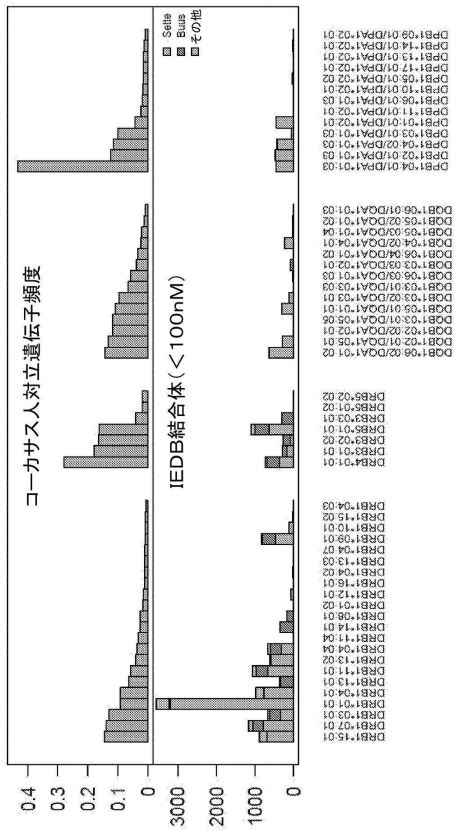


FIG. 12E

【図１２Ｆ】

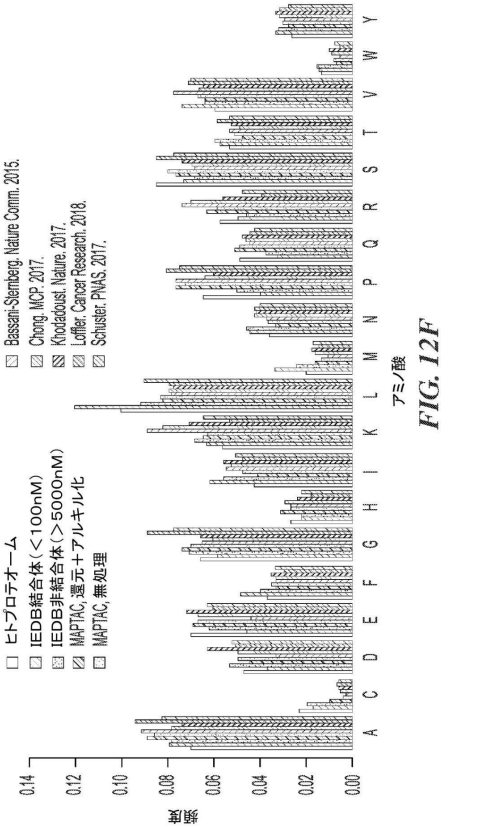


FIG. 12F

【図１３】

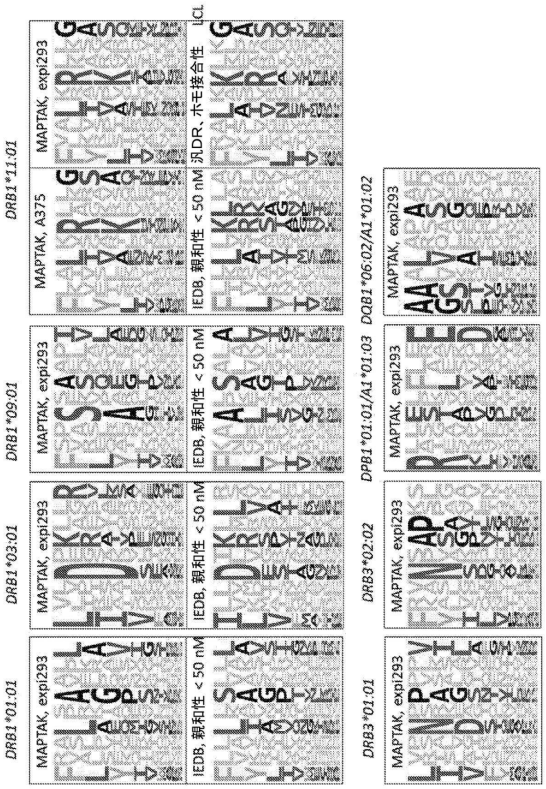


FIG. 13

【図１４Ａ】

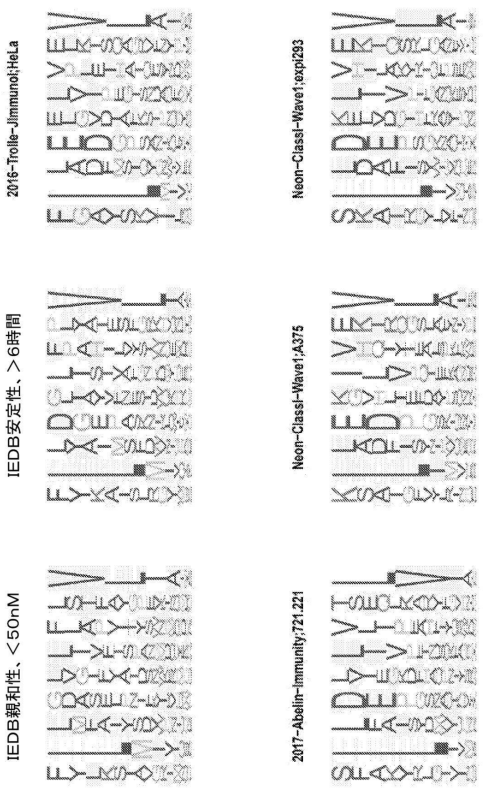


FIG. 14A

【図１４Ｂ】

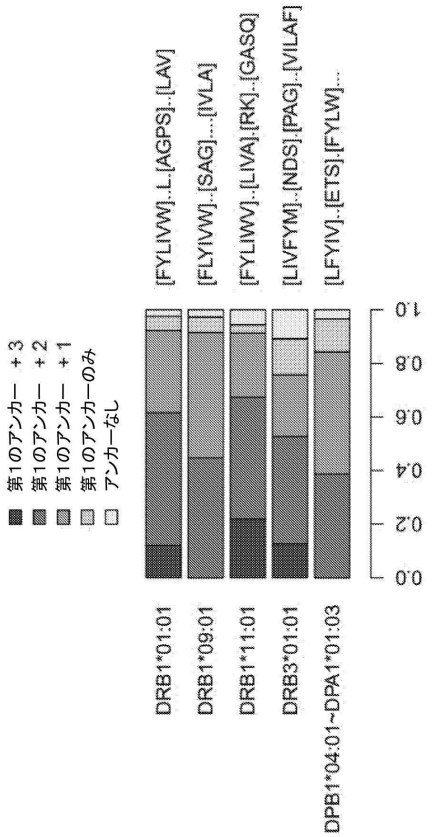


FIG. 14B

【図１４Ｃ】

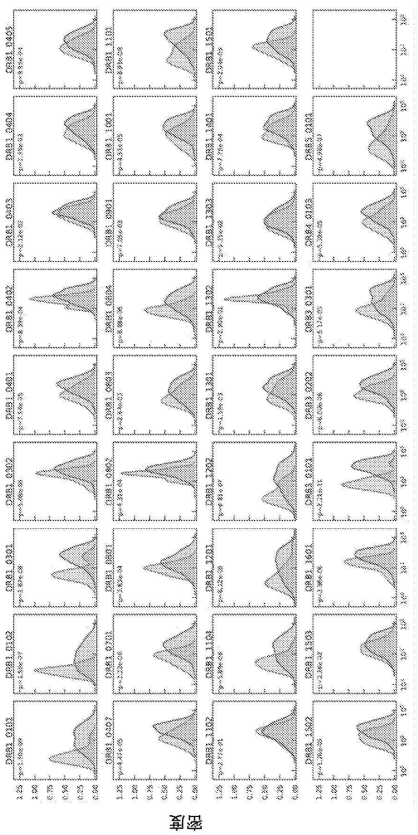


FIG. 14C

【図１５Ａ】

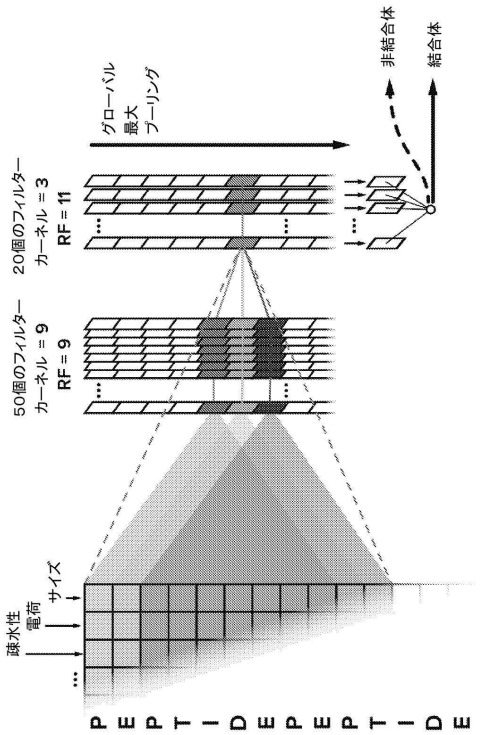


FIG. 15A

【図１５Ｂ】

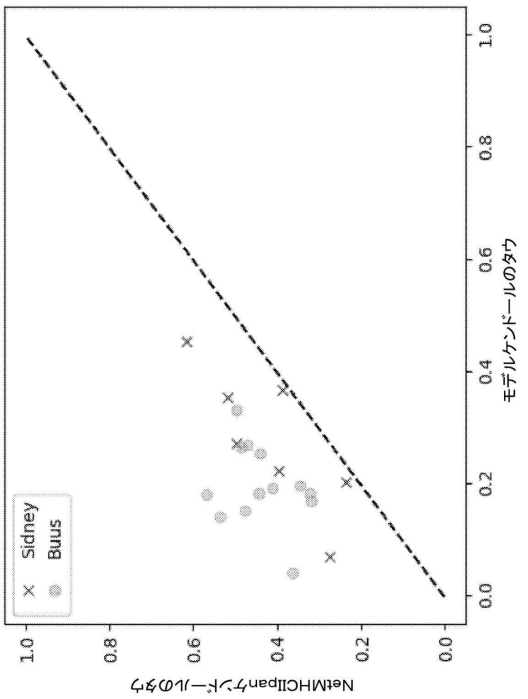


FIG. 15B

【図１６】

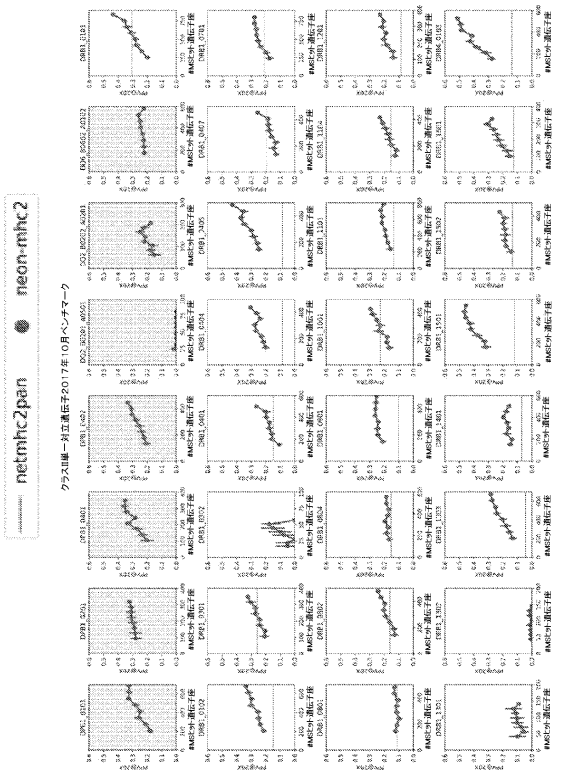


FIG. 16

【図１７Ａ】

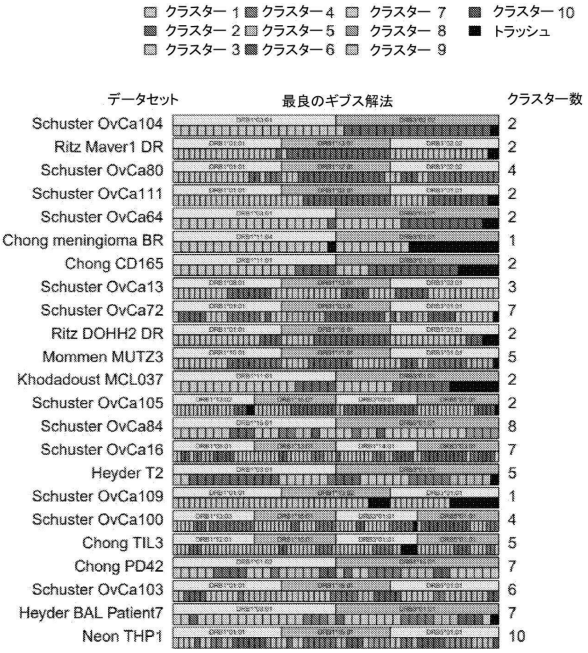


FIG. 17A

【図１７Ｂ】

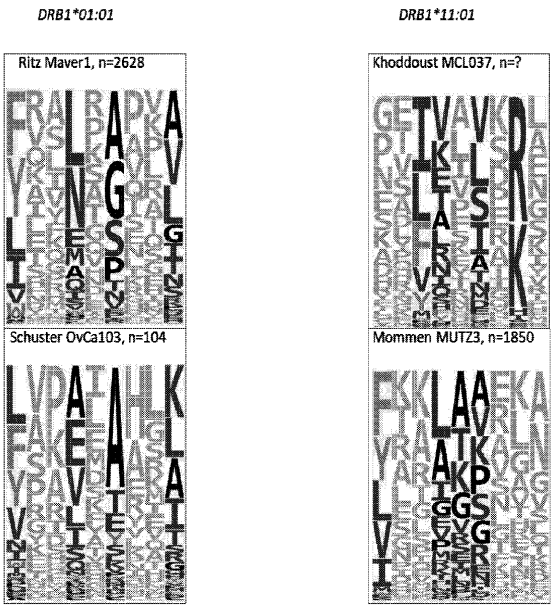


FIG. 17B

【図１７Ｃ】

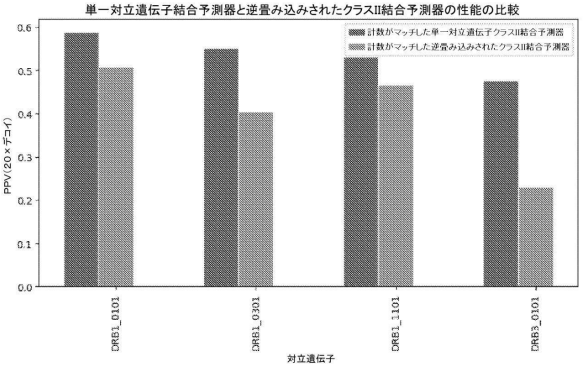


FIG. 17C

【図１７Ｄ】

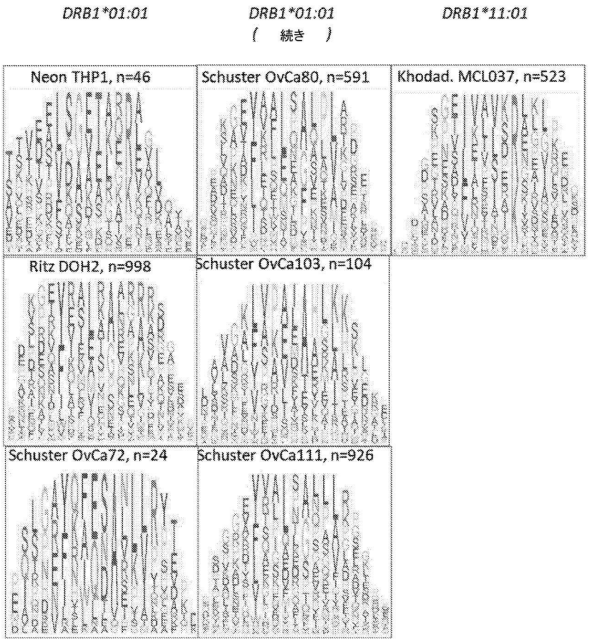


FIG. 17D

【図１８Ａ】

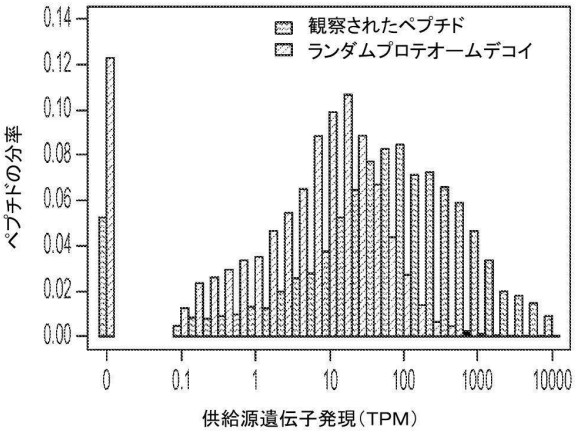


FIG. 18A

【図１８Ｂ】

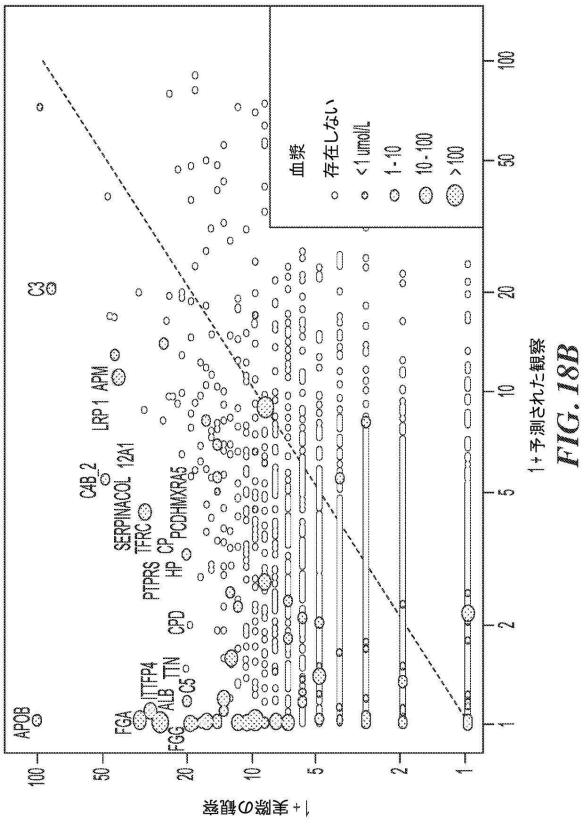


FIG. 18B

【図１８Ｃ】

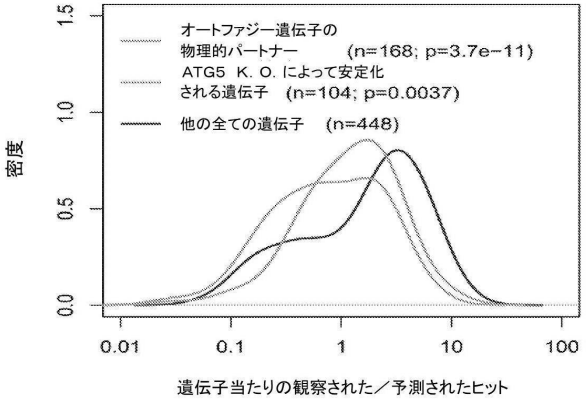


FIG. 18C

【図１８Ｄ】

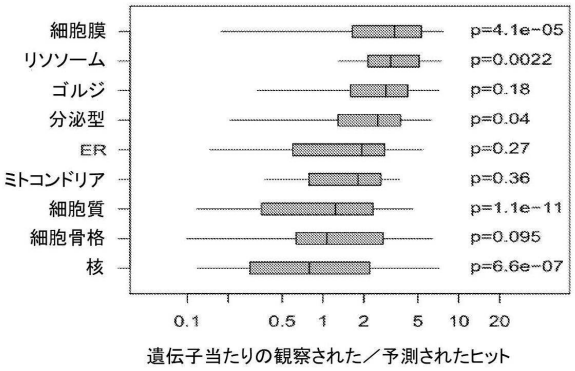


FIG. 18D

【図１８Ｅ】

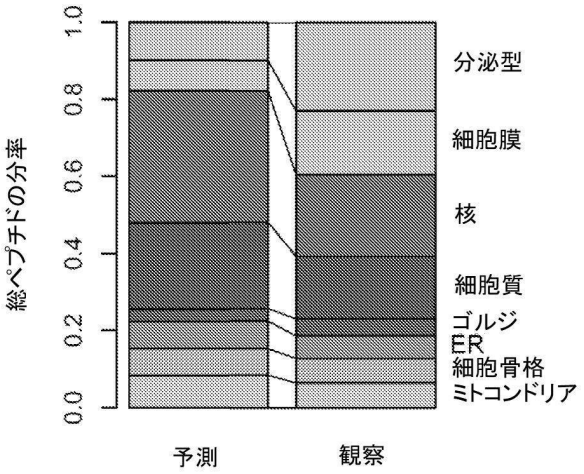


FIG. 18E

【図１８Ｆ】

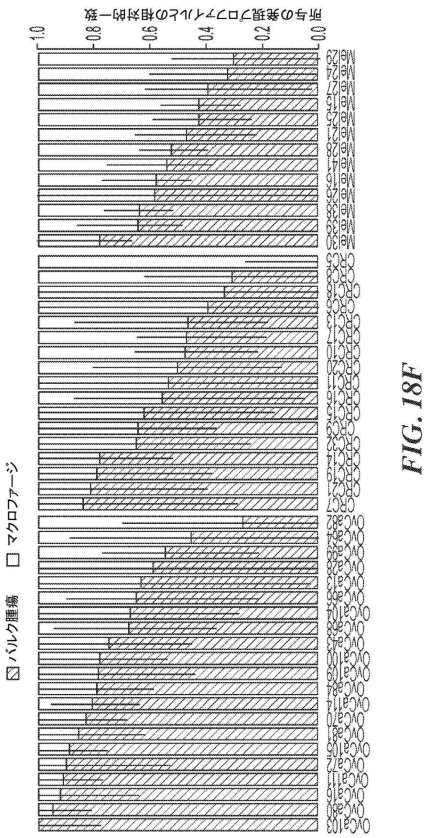


FIG. 18F

【図１９Ａ】

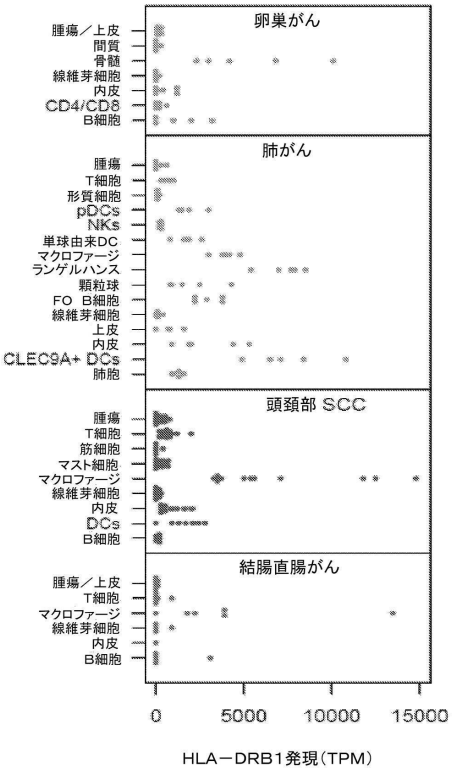


FIG. 19A

【図１９Ｂ】

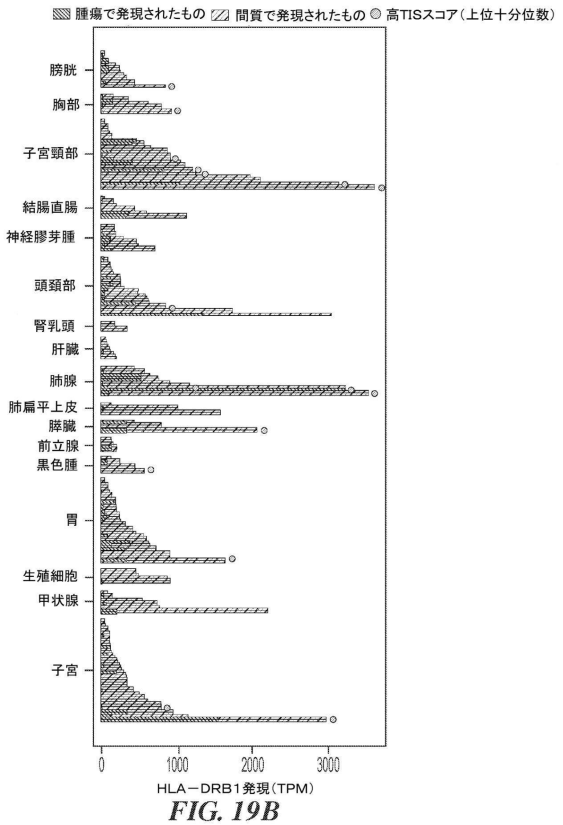


FIG. 19B

【図１９Ｃ】

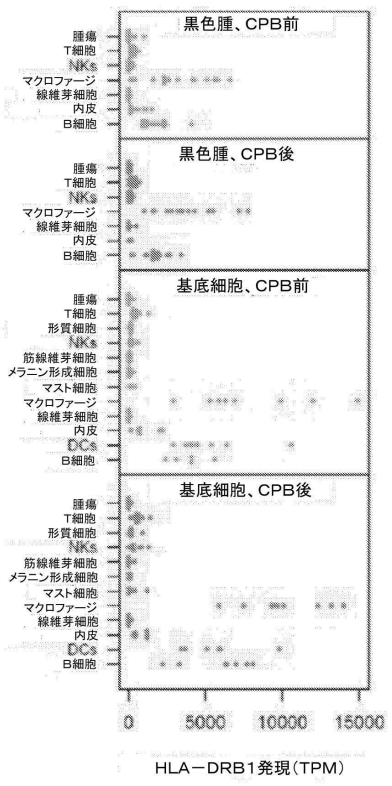


FIG. 19C

【図２０】

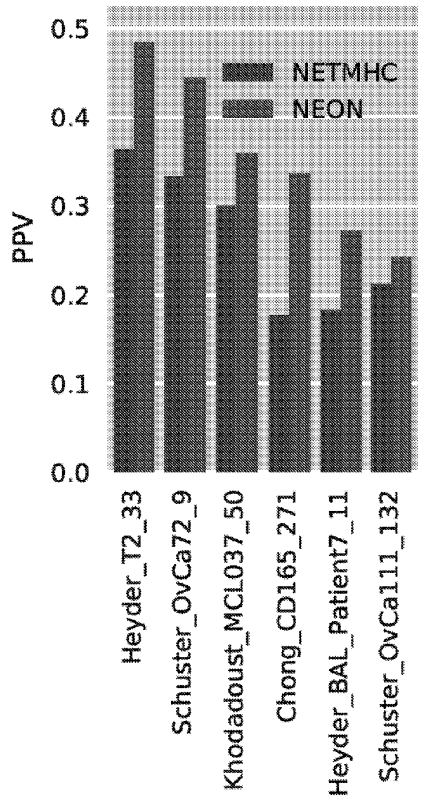


FIG. 20

【図２１Ａ】

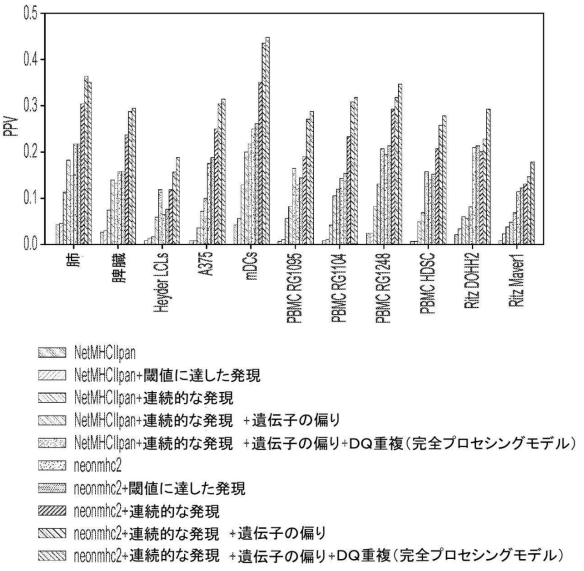


FIG. 21A

【図２１Ｂ】

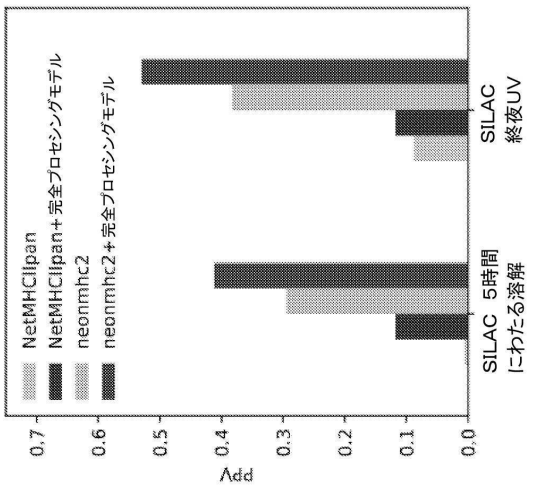


FIG. 21B

【図２１Ｃ】

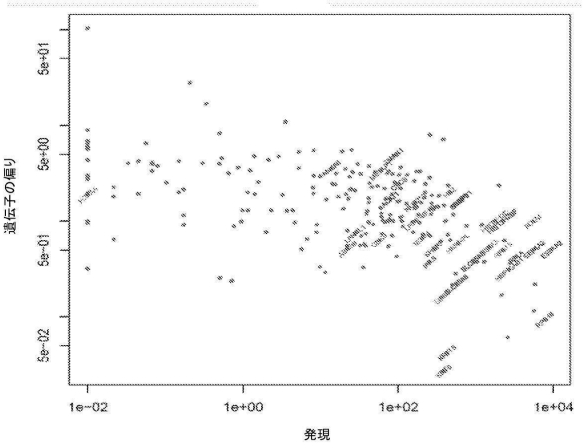


FIG. 21C

【図２１Ｄ】

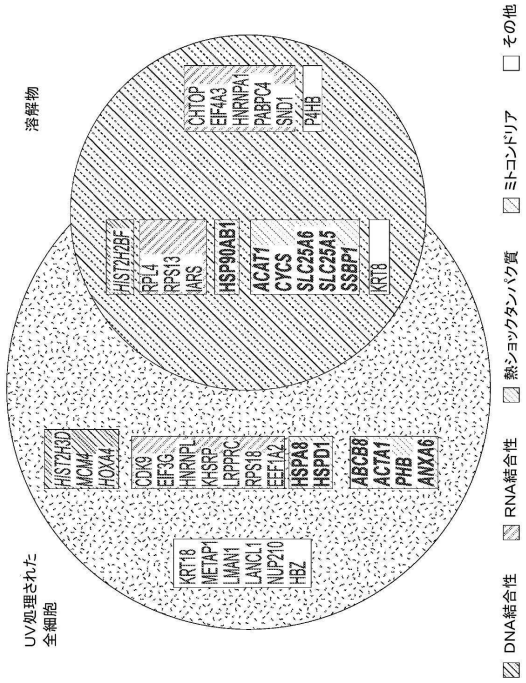


FIG. 21D

【図２２Ａ】

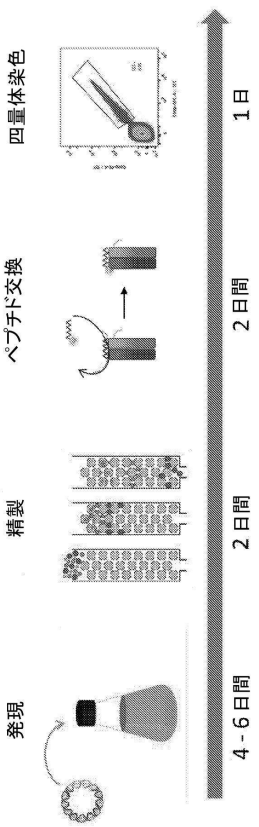


FIG. 22A

【図２２Ｂ】

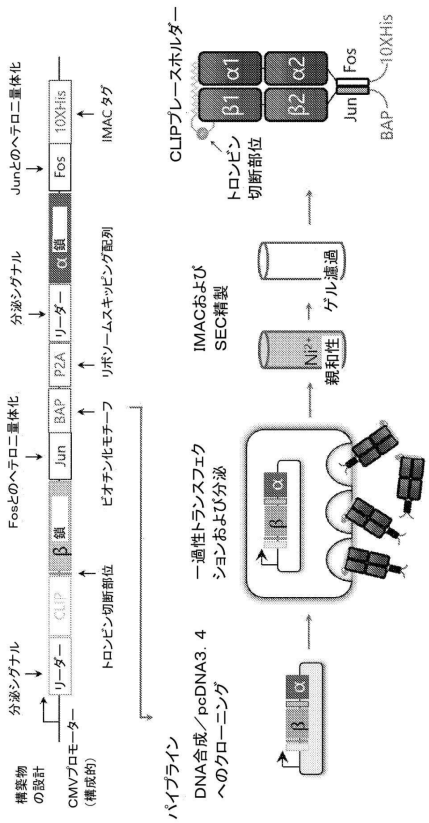


FIG. 22B

【図２３】

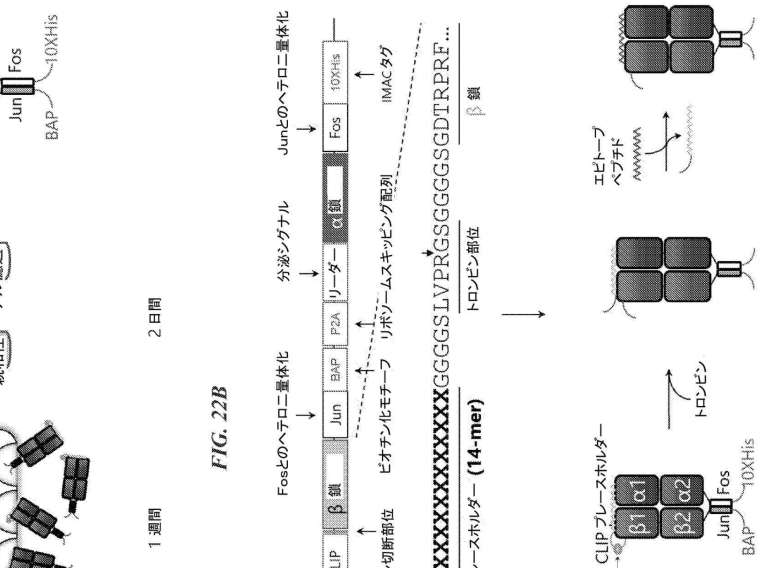


FIG. 23

【図２４】

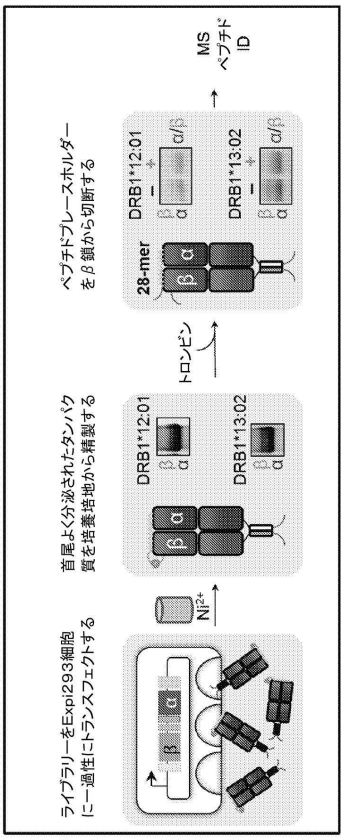


FIG. 24

【図２５Ａ】

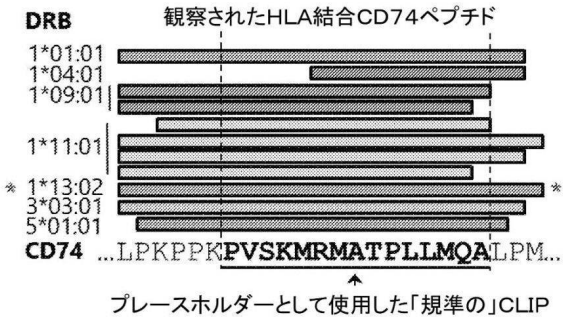


FIG. 25A

【図２５Ｂ】

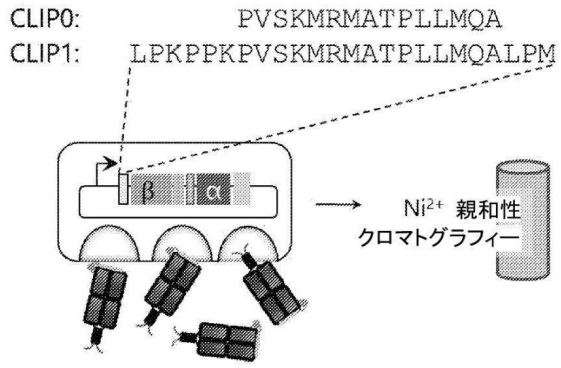


FIG. 25B

【図２５Ｃ】

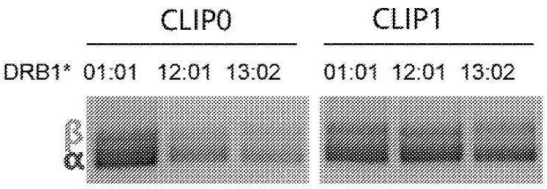


FIG. 25C

【図２６Ａ】

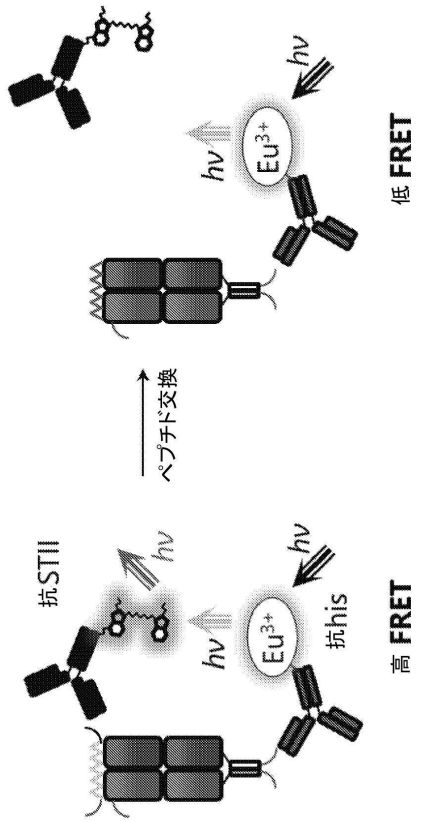


FIG. 26A

【図２６Ｂ】

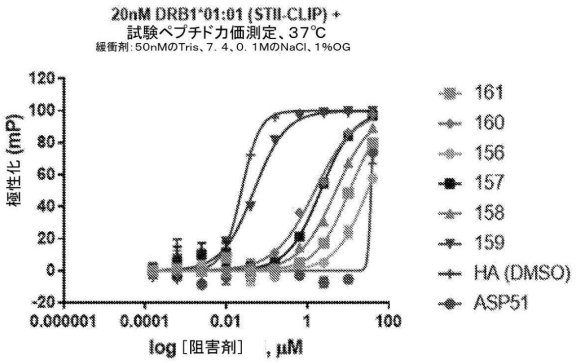


FIG. 26B

【図２６Ｃ】

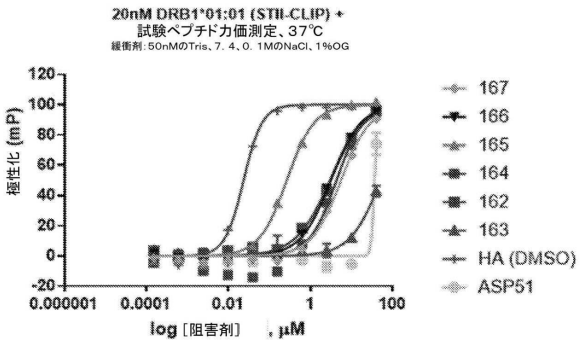


FIG. 26C

【図２６Ｄ】

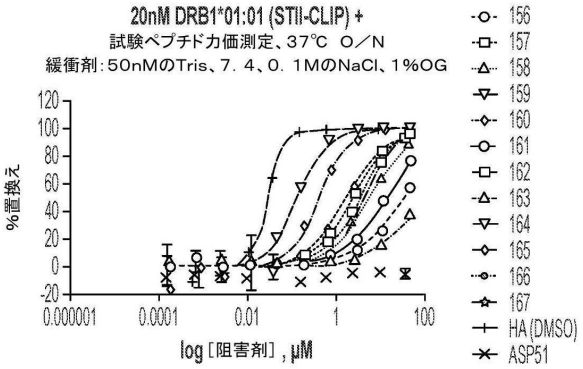


FIG. 26D

【図２６Ｅ】

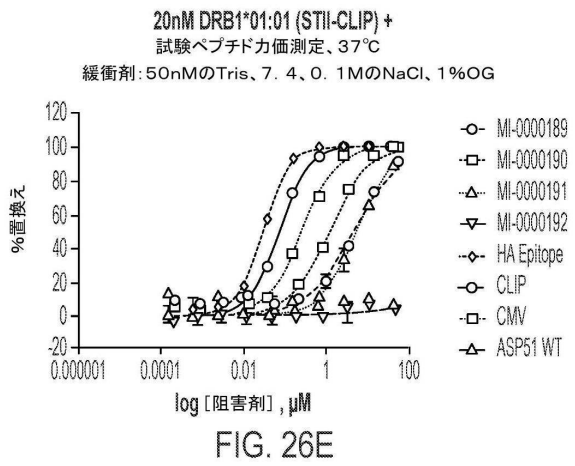


FIG. 26E

【図２６Ｆ】

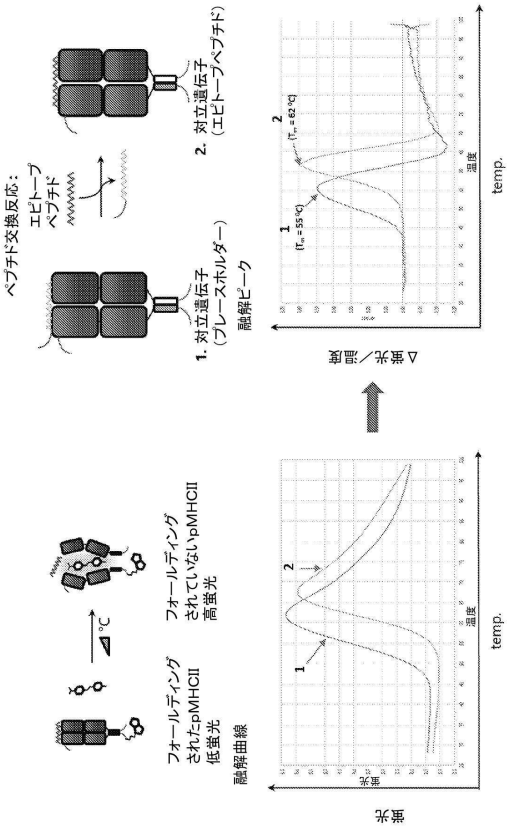


FIG. 26F

【図２６Ｇ】

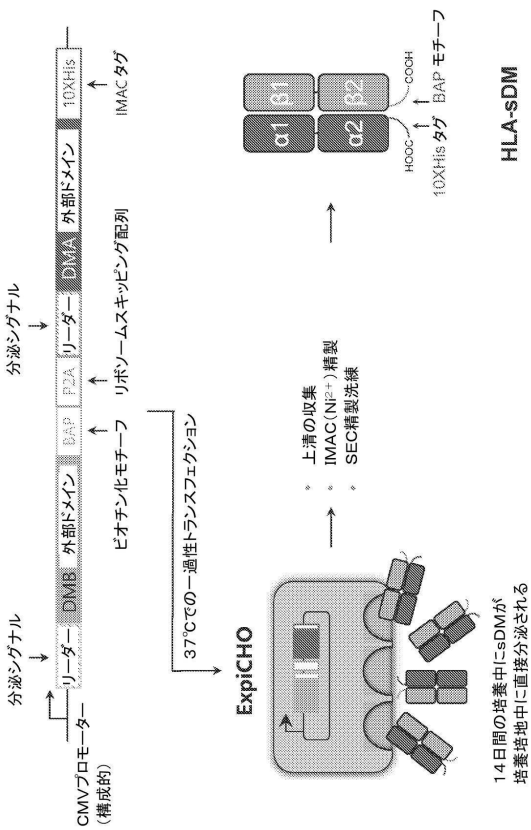


FIG. 26G

【図２６Ｈ】

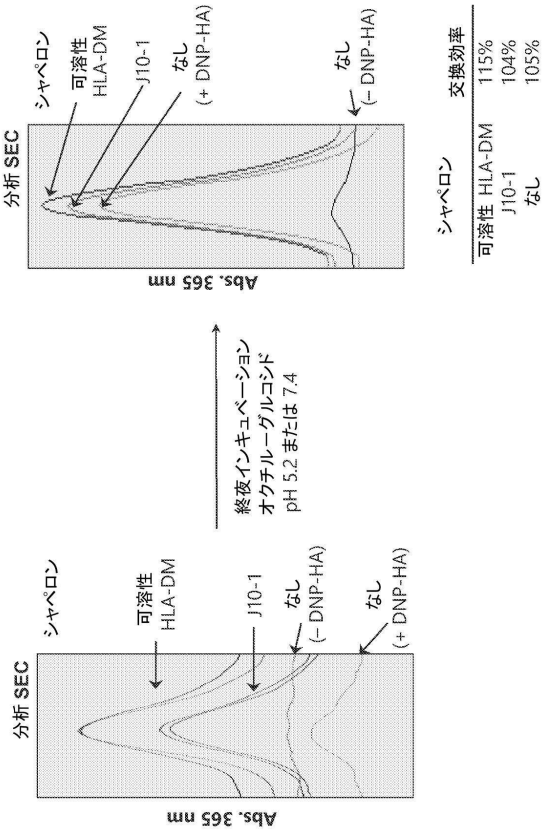


FIG. 26H

【図２７Ａ】

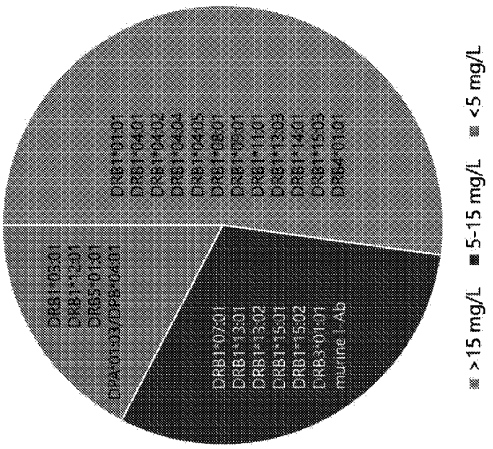


FIG. 27A

【図２７Ｂ】

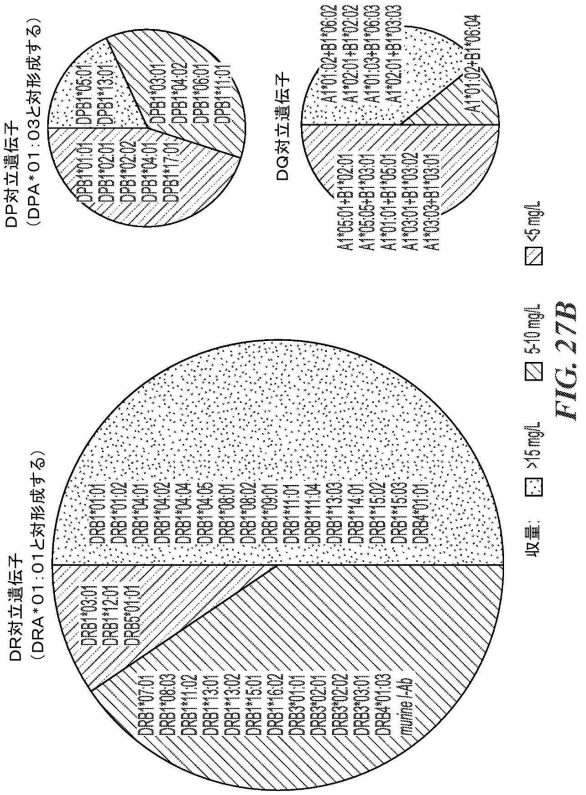


FIG. 27B

【図２７Ｃ】

対立遺伝子	交換について検証されたペプチド
DRB1*07:01	3
DRB1*15:01	7
DRB1*01:01	8
DRB1*11:01	12
DRB1*13:02	1

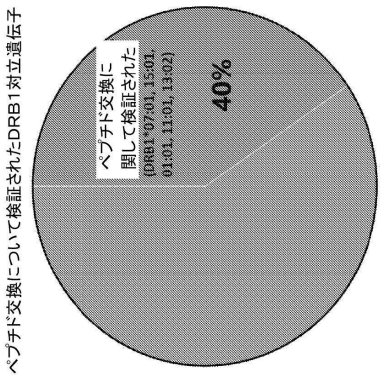


FIG. 27C

【図２７Ｄ】

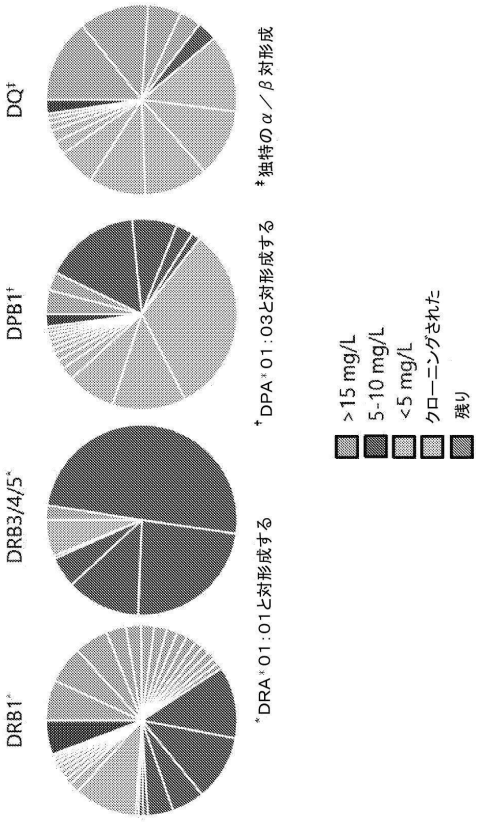


FIG. 27D

【図２７Ｅ】

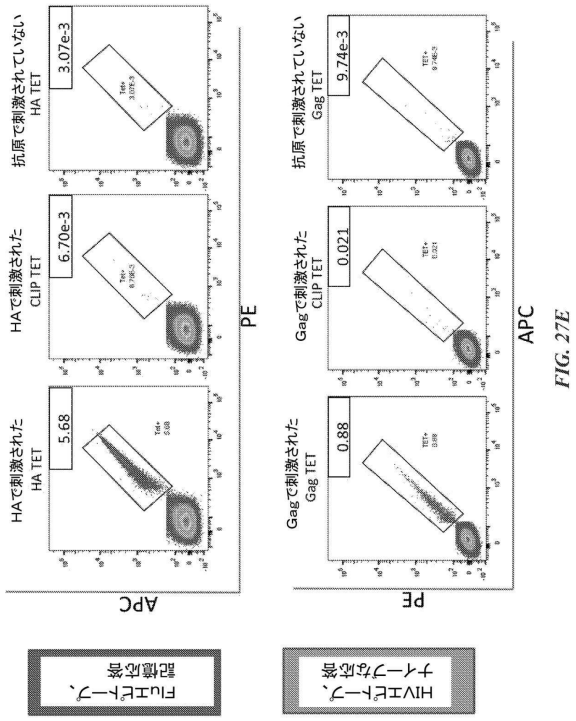


FIG. 27E

【図２８Ａ】

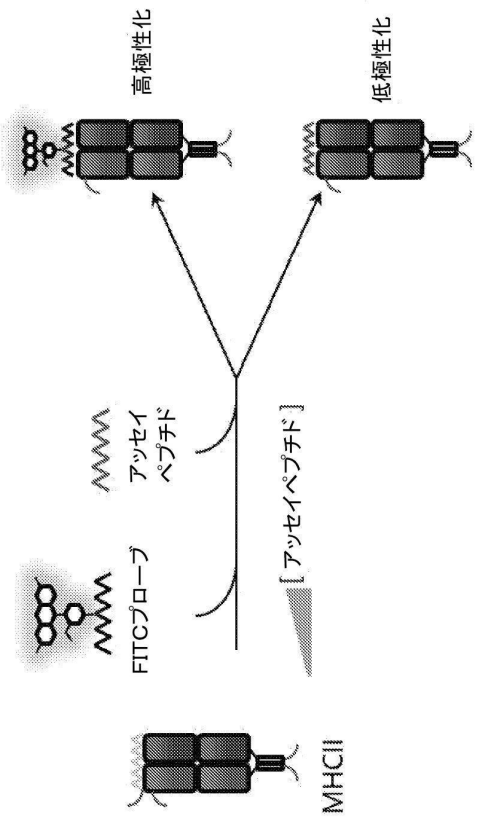


FIG. 28A

【図２８Ｂ】

変数	評価した条件
pH	5. 2対7. 4
界面活性剤	1%オクチルグルコシド対1. 6%NP40
温度	37℃対室温
プローブ	FITC－CLIP0対FITC－HA（超結合体）
精製方法	Hisタグ対コンフォメーション特異的抗体（L243）
標的フォーマット	可溶性対立遺伝子対全長対立遺伝子

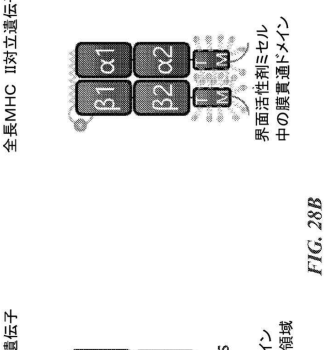


FIG. 28B

【図２８Ｃ】

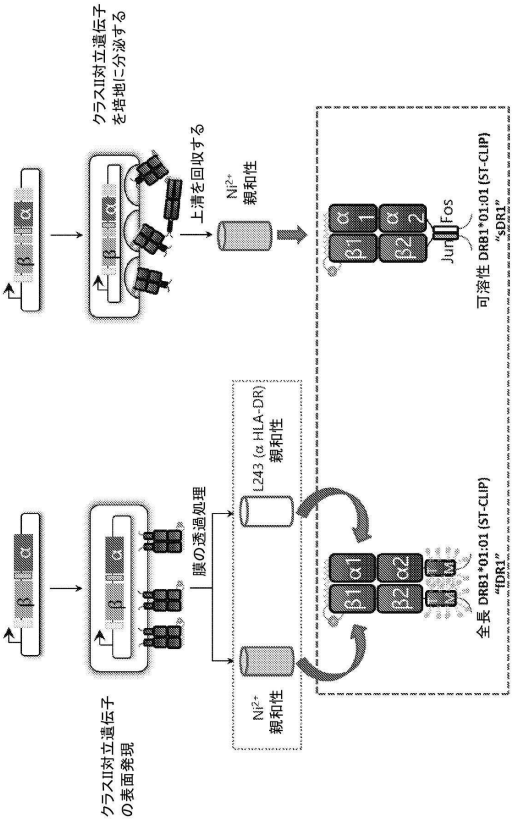
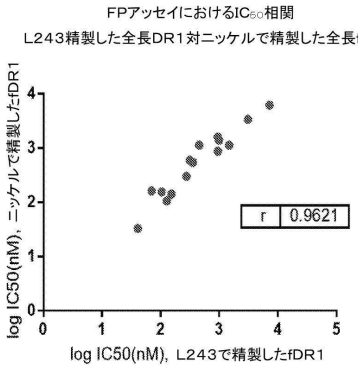


FIG. 28C

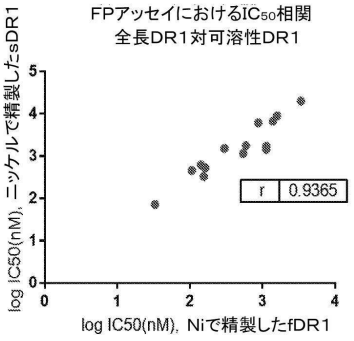
【図２８Ｄ】



	平均IC50、nM（n＝2）		IC50変化倍率
ペプチド	L243で精製したfDR1	Niで精製したfDR1	Ni fDR1に対するL243 fDR1
ペプチド 166	453	1128	2.5
ペプチド 165	70	163	2.3
ペプチド 157	315	591	1.9
ペプチド 163	941	1585	1.7
ペプチド 158	357	542	1.5
ペプチド 159	104	156	1.5
ペプチド 190	992	1398	1.4
ペプチド 162	3073	3384	1.1
ペプチド 164	278	300	1.1
ペプチド 161	946	875	0.9
ペプチド 167	153	142	0.9
ペプチド 189	7173	6132	0.9
ペプチド 160	128	106	0.8
ペプチド 191	1470	1124	0.8
HA（ 対照 ）	41	33	0.8

FIG. 28D

【図２８Ｅ】



	平均IC50、nM（n＝2）		IC50変化倍率
ペプチド	全長DR1 (fDR1)	可溶性DR1 (sDR1)	fDR1に対するsDR1
ペプチド 161	875	6076	**6.9**
ペプチド 162	3384	20076	**5.9**
ペプチド 163	1585	8915	**5.6**
ペプチド 164	300	1517	**5.1**
ペプチド 190	1398	6695	**4.8**
ペプチド 160	106	457	**4.3**
ペプチド 167	142	614	**4.3**
ペプチド 165	163	529	**3.2**
ペプチド 157	591	1773	**3**
HA（ 対照 ）	33	72	**2.2**
ペプチド 158	542	1155	**2.1**
ペプチド 159	156	331	**2.1**
ペプチド 166	1128	1706	**1.5**
ペプチド 191	1124	1444	**1.3**
ペプチド 189	6132	>50,000	ND

FIG. 28E

【図２８Ｆ】

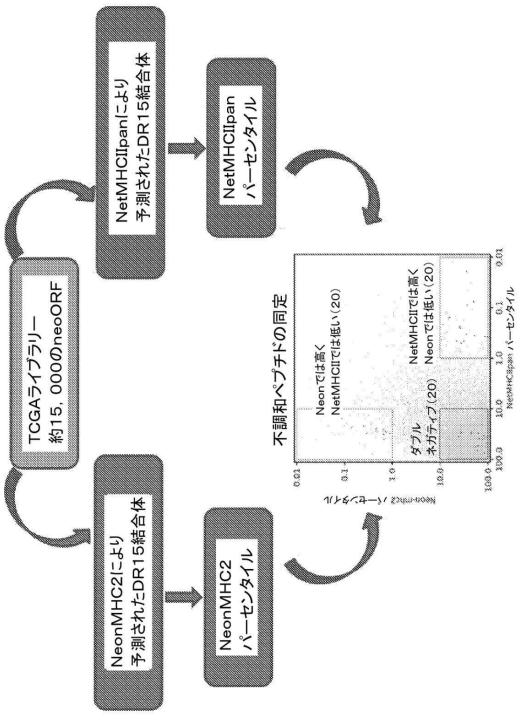


FIG. 28F

【図２８Ｇ】

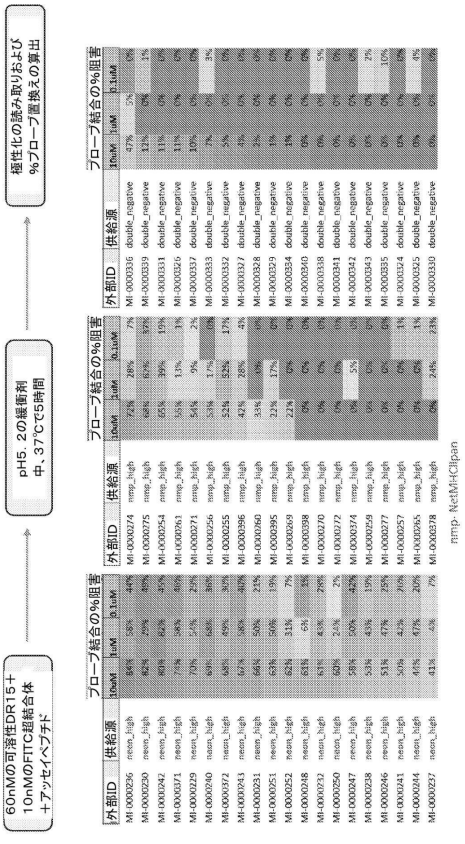


FIG. 28G

【図２８Ｈ】

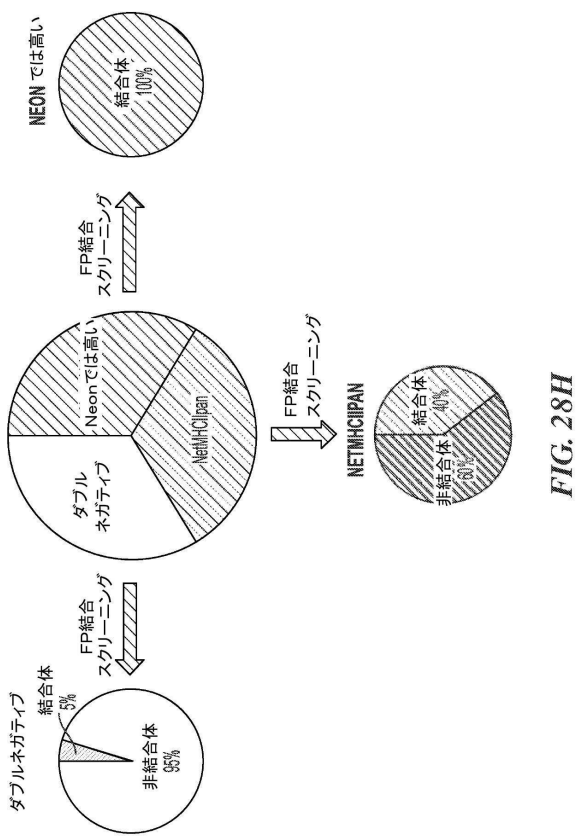


FIG. 28H

【図２９】

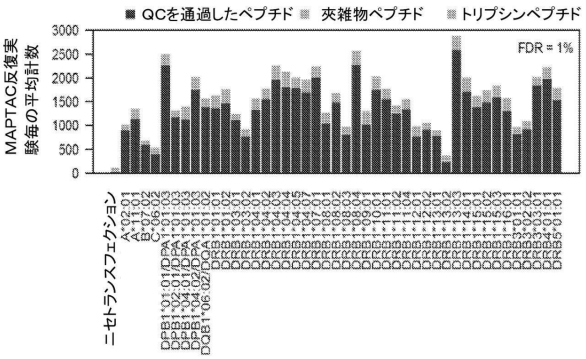


FIG. 29

【図３０Ａ－Ｂ】

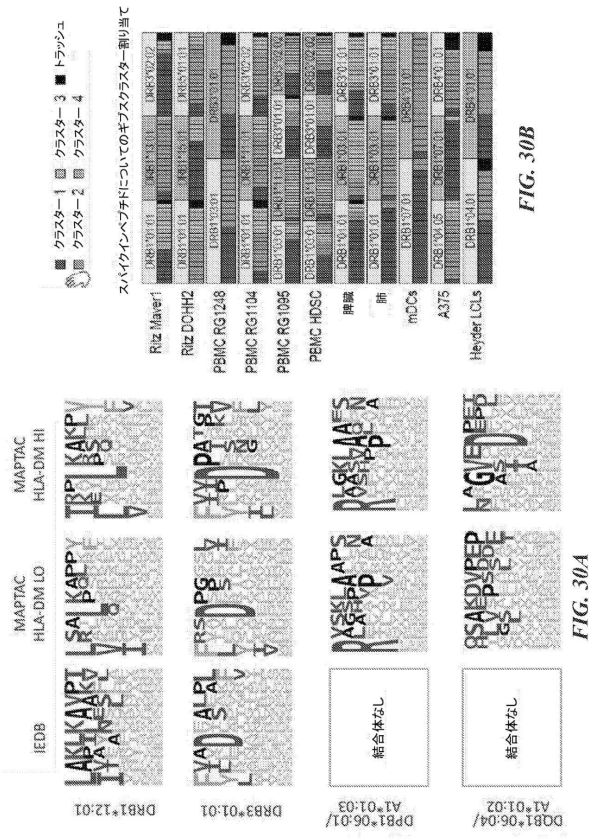


FIG. 30B

FIG. 30A

【図３０Ｃ】

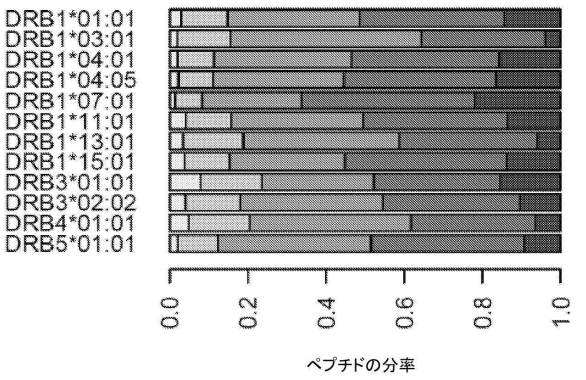


FIG. 30C

【図３１Ａ】

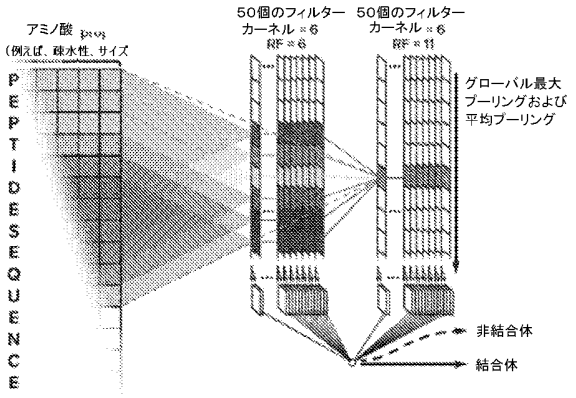


FIG. 31A

【図３１Ｂ】

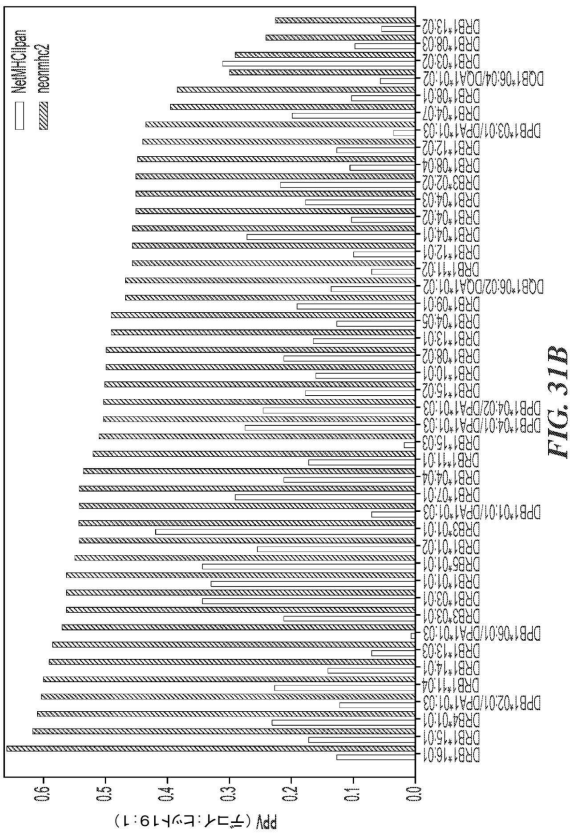


FIG. 31B

【図３１Ｃ】

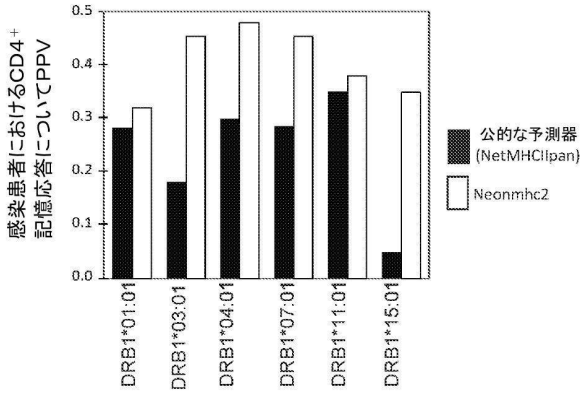


FIG. 31C

【図３１Ｄ】

新抗原ペプチド	NetMHCIIpan による親和性 (nM)	NetMHCIIpan による%順位	neonmhc2 による%順位	状態
GEGLEQPAHRYPDAG	744	39.00%	0.43%	ヒット
RNQWKCLGKPVGHEM	426	29.00%	0.10%	アッセイされた
VPAWTRAWRNSSPKG	395	27.00%	0.51%	ヒット
DGGSYFSLWKWTQV	364	26.00%	0.11%	ヒット
LHPELLPLWRLLPDG	367	26.00%	0.23%	ヒット
PEASLYGALSKGSGG	353	26.00%	0.24%	ヒット
HEGGFPPLLRRAAED	325	24.00%	0.63%	ヒット
LSPVWCLQWKLSGTD	293	23.00%	0.31%	ヒット
KNTVYTKQVQGSCQ	259	21.00%	0.57%	アッセイされた
PATYILLKEFCLVG	246	20.00%	0.32%	アッセイされた
NNEQFQWKIRHVGPE	222	19.00%	0.50%	ヒット
YKGGYELVKKSQTEL	200	17.00%	0.37%	アッセイされた

FIG. 31D

【図３１Ｅ－Ｆ】

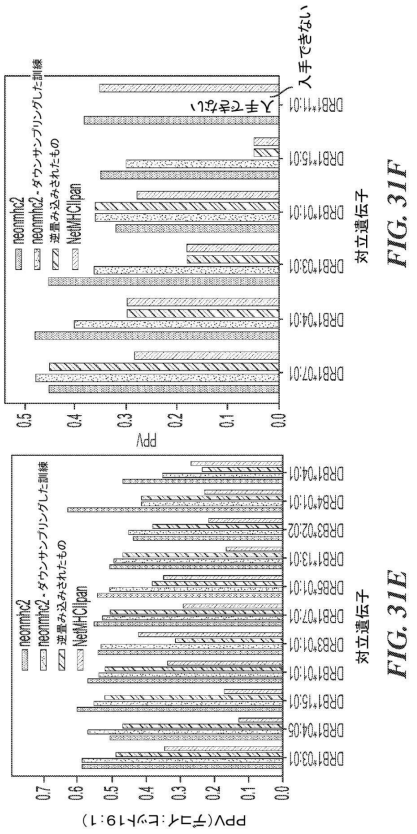


FIG. 31F

FIG. 31E

【図３２Ａ－Ｅ】

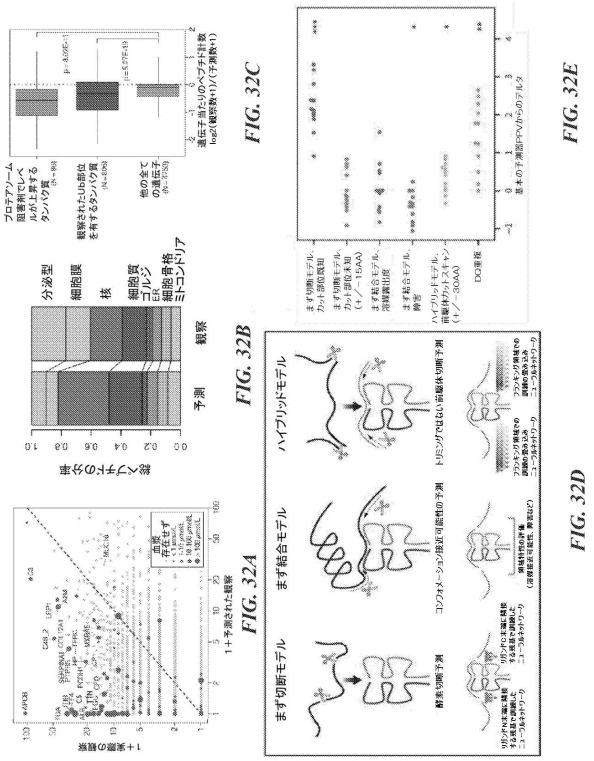


FIG. 32A

FIG. 32B

FIG. 32C

FIG. 32D

FIG. 32E

【図３３Ａ－Ｃ】

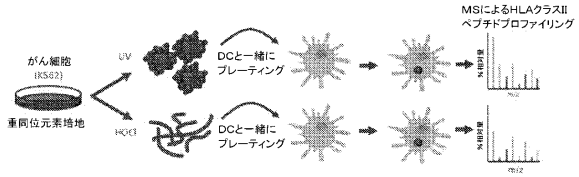


FIG. 33A

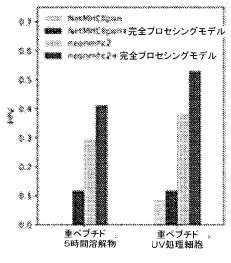


FIG. 33B

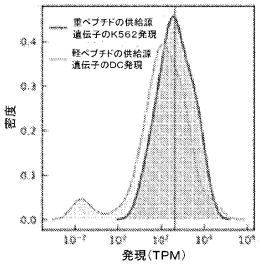


FIG. 33C

【図３３Ｄ－Ｅ】

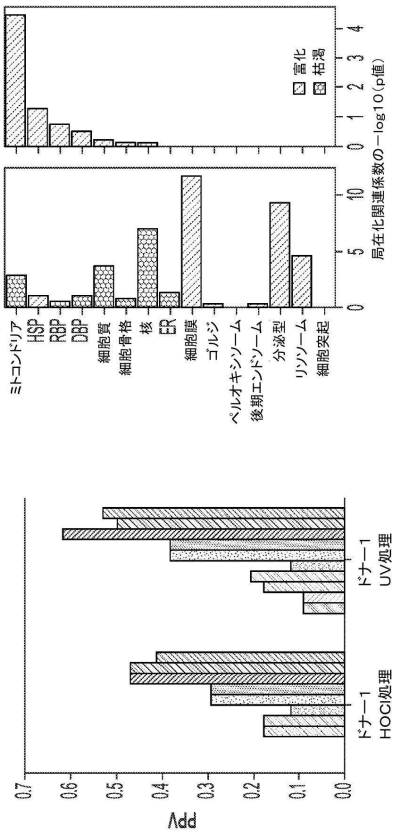


FIG. 33E

FIG. 33D

【図３３Ｆ－Ｇ】

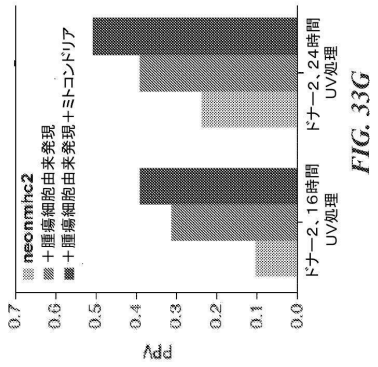


FIG. 33G

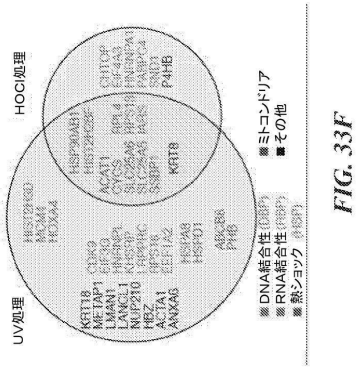


FIG. 33F

【図３４Ａ－Ｂ】

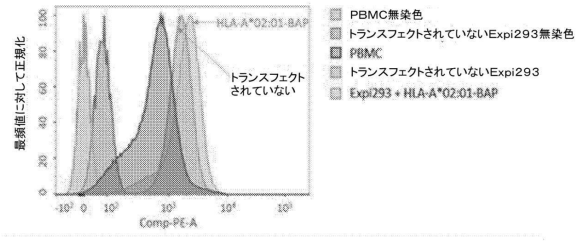


FIG. 34A

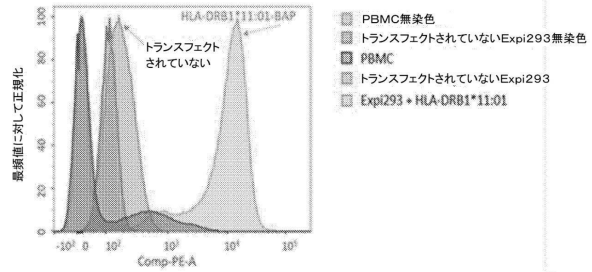


FIG. 34B

【図３５】



FIG. 35

【図３６Ａ】

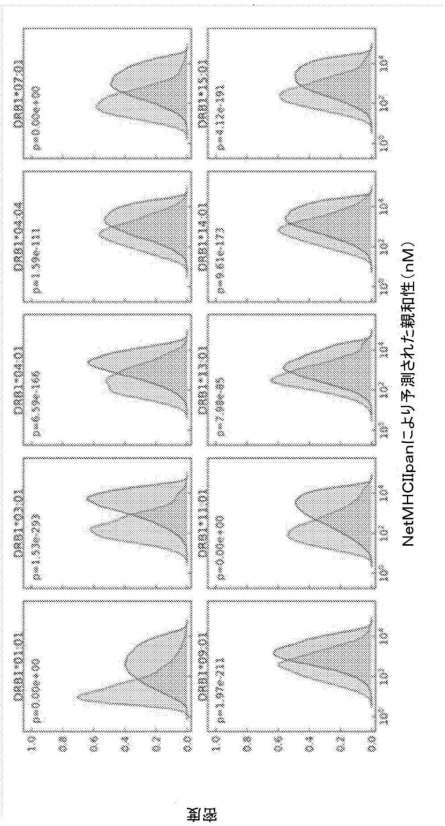


FIG. 36A

【図３６Ｂ】

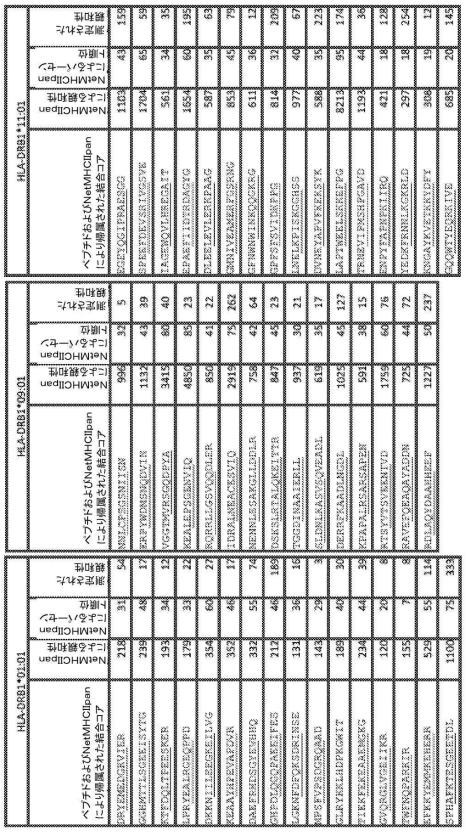


FIG. 36B

【図３６Ｃ】

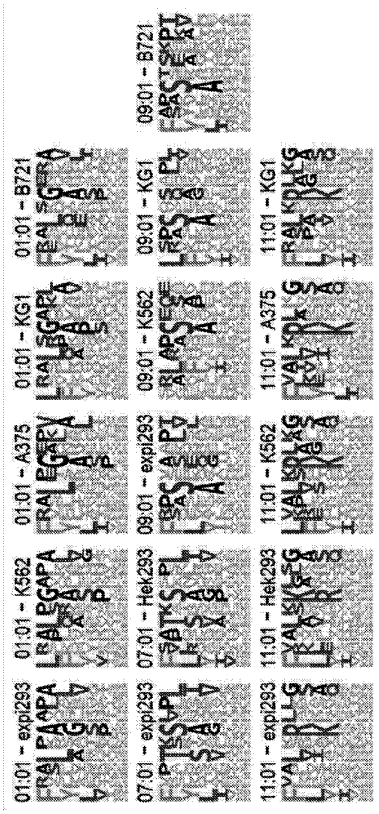


FIG. 36C

【図３７Ａ－Ｂ】



FIG. 37A

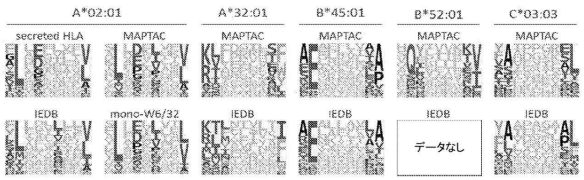


FIG. 37B

【図３７Ｃ】

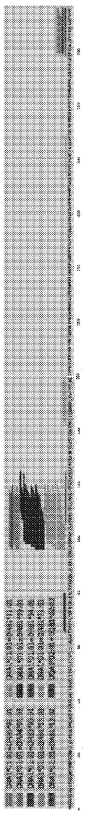


FIG. 37C

【図３８Ａｉ】

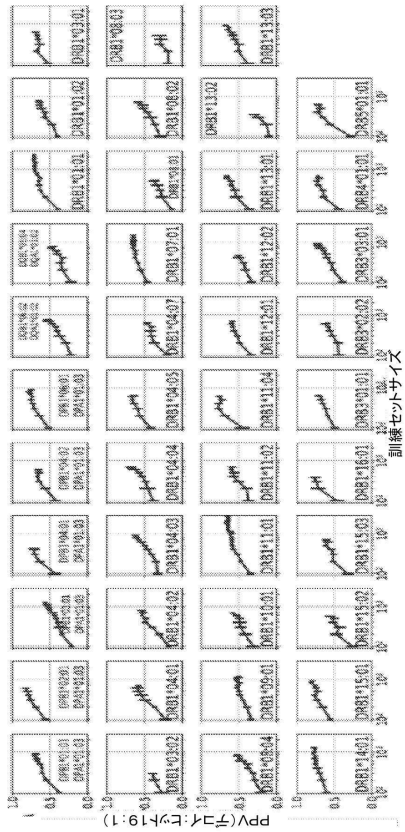


FIG. 38Ai

【図３８Ａｉｉ】

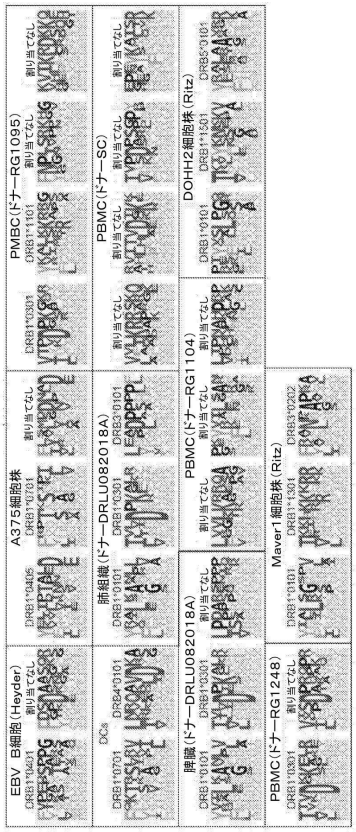


FIG. 38Aii

【図３８Ｂ－Ｃ】

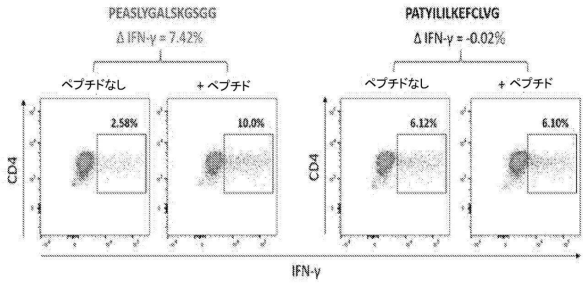


FIG. 38B

ペプチド	状態	陽性ウェル数	ウェル1	ウェル2	ウェル3	ウェル4	ウェル5	ウェル6	ウェル7	ウェル8
GEGLFQPAHRYPDAG	誘導_ヒット	4	3.28	3.5	3.87	3.33	0.81			
RNQWKCLGKPVGAEM	アッセイされた	0	1.88	-1.44	-1.27	-15.43	-10.18	-0.09		
VPAWTRAWRNSSPKG	誘導_ヒット	2	3.53	0.4	-0.36	2.93	2.2	6.3		
LHPELLPLWRLLPDG	誘導_ヒット	1	-1.26	0.46	-0.36	-1.74	0.75	3.92		
DGGSYFSLWKIWTQV	誘導_ヒット	5	4.83	3.84	4.34	13.8	8.1			
PEASLYGALSKGSGG	誘導_ヒット	5	1.96	7.42	1.09	5.24	6.03	1.72	12.93	9.74
HEGGFPPLLRRAAED	誘導_ヒット	5	-0.36	3.58	2.58	-22.75	6.59	3.5	8.44	10.39
LSPVWCLQWKLSGTD	誘導_ヒット	2	-0.08	2.52	0.71	13.27	11.32			
KNTIVYTTKQVQSCQ	アッセイされた	0	-1.19							
PATYILILKEFCLVG	アッセイされた	0	-0.02	0.69	1.47	-0.17				
NNEQFQWKIRHVGPE	誘導_ヒット	4	5.44	5.7	0.93	2.97	5.1	4.5		
YKGGYELVKKSQTEL	アッセイされた	0	2.28	-2.01	-2.31					

FIG. 38C

【図３８Ｄ】

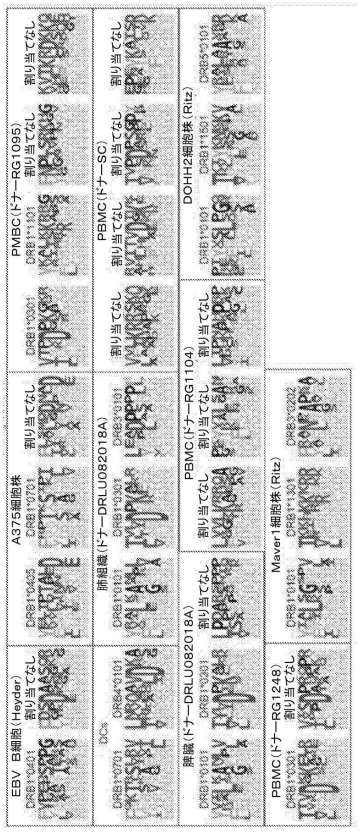


FIG. 38D

【図３９Ａ－Ｃ】

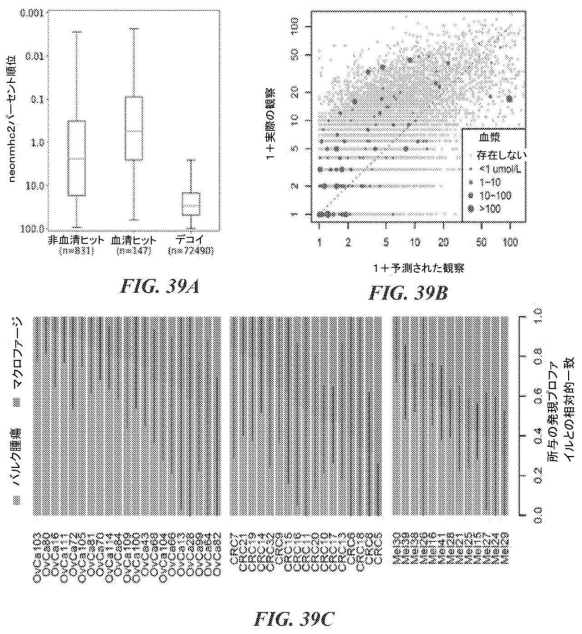


FIG. 39A

FIG. 39B

FIG. 39C

【図４０Ａ－Ｂ】

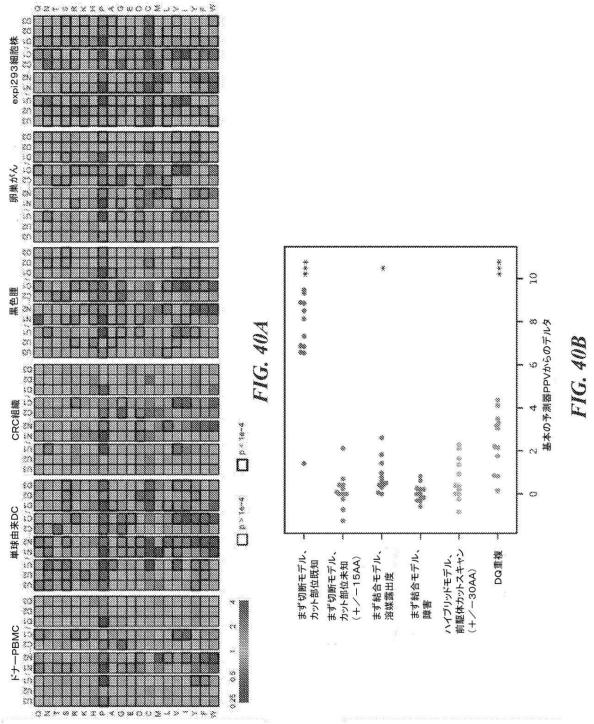


FIG. 40A

FIG. 40B

【図４１】

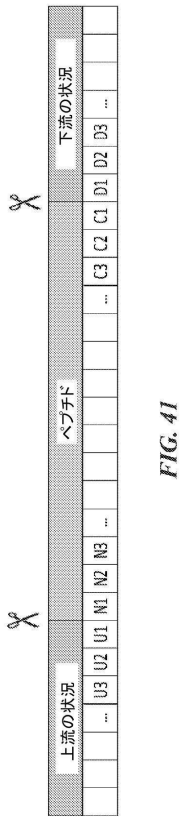


FIG. 41

【図４２Ａ】

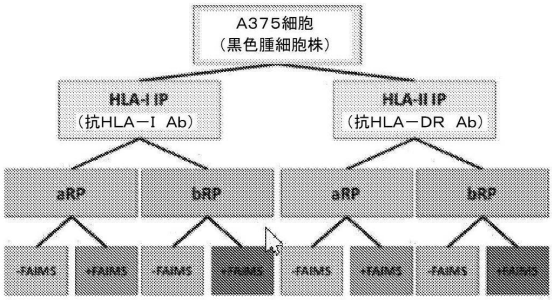


FIG. 42A

【図４２Ｂ】

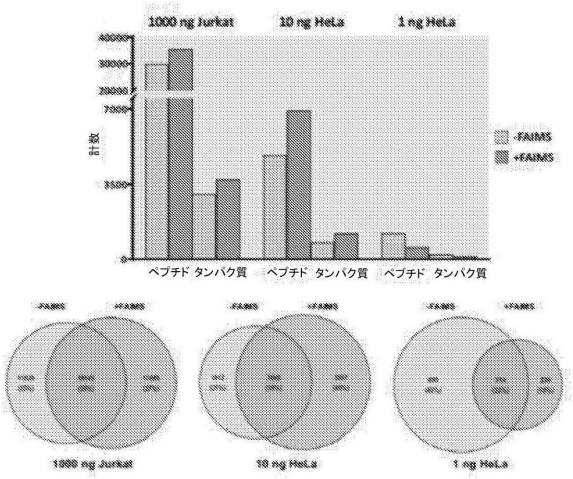


FIG. 42B

【図４３Ａ】

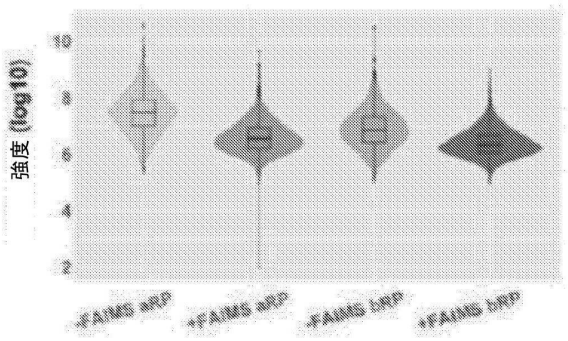


FIG. 43A

【図４３Ｂ】

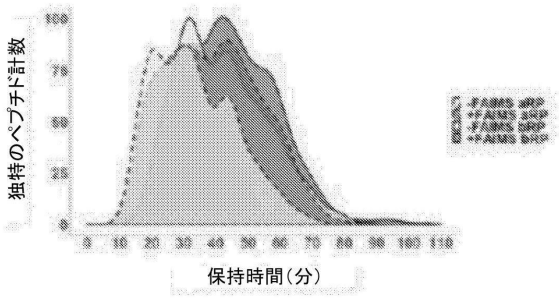


FIG. 43B

【図４４Ａ】

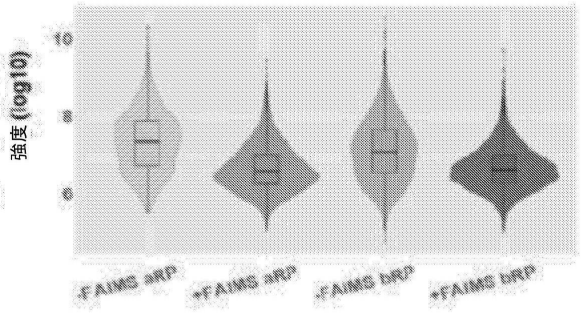


FIG. 44A

【図４４Ｂ】

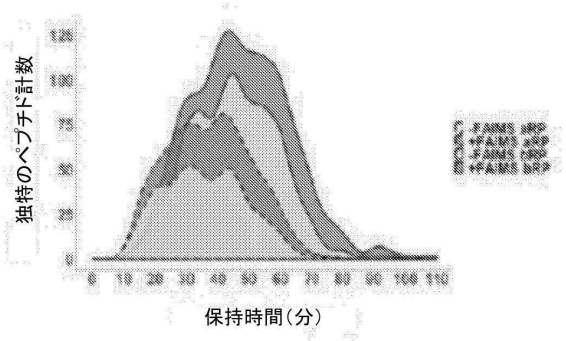


FIG. 44B

【図４５－１】

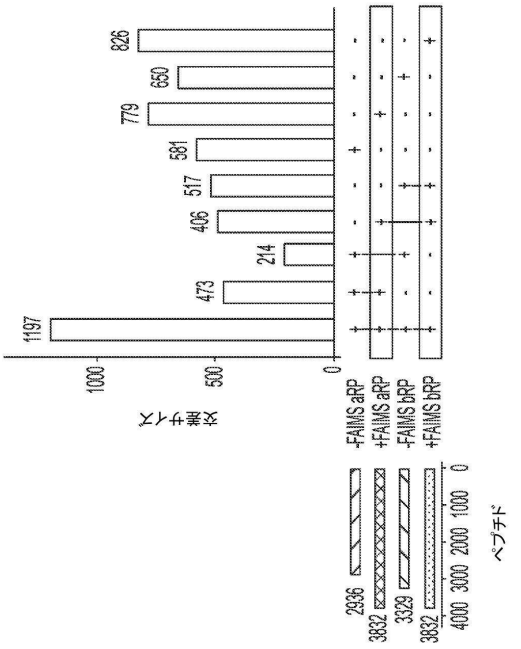


FIG. 45

【図４５－２】

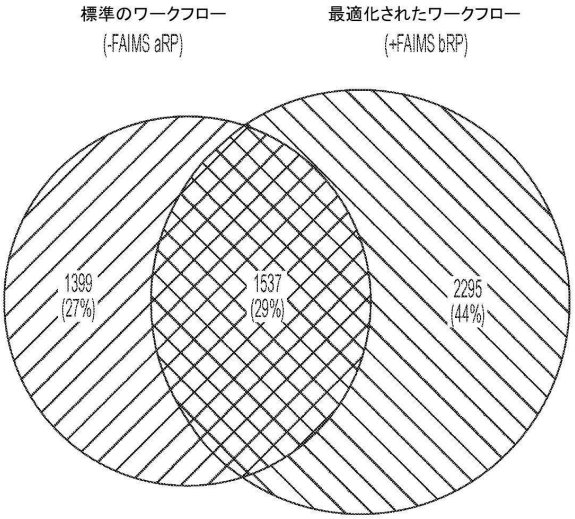


FIG. 45(　続き　)

【図４６－１】

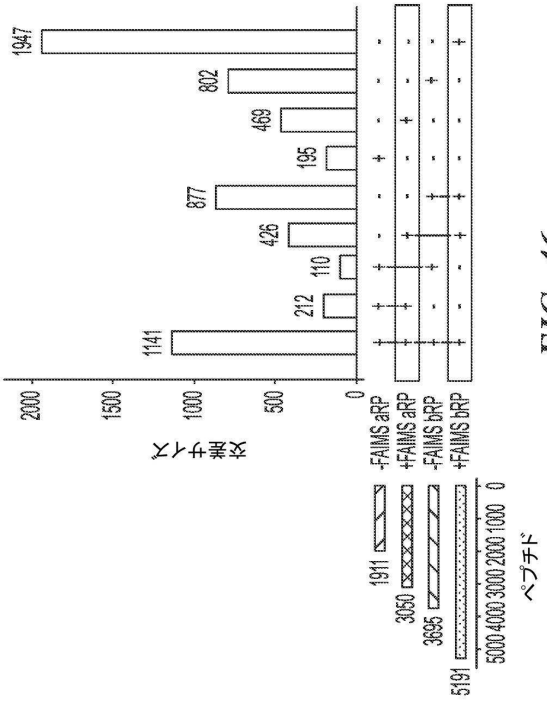


FIG. 46

【図４６－２】

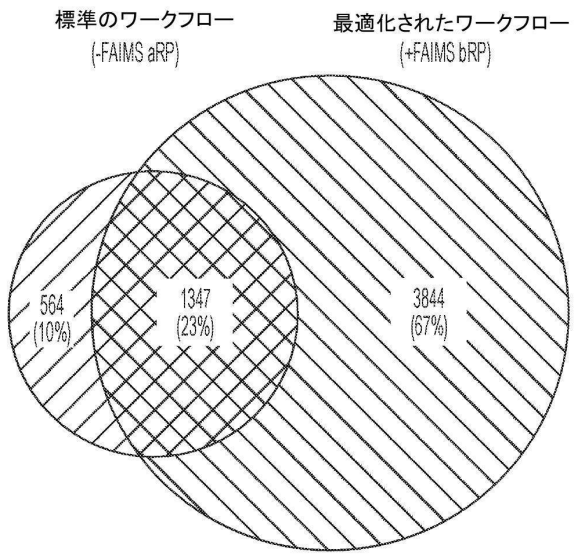


FIG. 46(　続き　)

【配列表】

0007236543000001.app

フロントページの続き

(51)国際特許分類		FI		
C12N 15/33 (2006.01)		C12N	15/33	
C07K 14/005 (2006.01)		C07K	14/005	
C12N 5/10 (2006.01)		C12N	5/10	
C07K 7/08 (2006.01)		C07K	7/08	
C07K 14/74 (2006.01)		C07K	14/74	
C12N 15/63 (2006.01)		C12N	15/63	Z
A61K 39/00 (2006.01)		A61K	39/00	H
A61P 37/04 (2006.01)		A61P	37/04	
A61P 35/00 (2006.01)		A61P	35/00	

(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号 62/855,379
(32)優先日 令和1年5月31日(2019.5.31)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号 62/891,101
(32)優先日 令和1年8月23日(2019.8.23)
(33)優先権主張国・地域又は機関
米国(US)
早期審査対象出願

弁護士 山本 健策
(72)発明者 ルーニー， マイケル スティーブン
アメリカ合衆国 マサチューセッツ ０２１３９， ケンブリッジ， エリー ストリート ４０， スイート １１０
(72)発明者 アベリン， ジェニファー グレイス
アメリカ合衆国 マサチューセッツ ０２１３９， ケンブリッジ， エリー ストリート ４０， スイート １１０
(72)発明者 バーセルミ， ドミニク
アメリカ合衆国 マサチューセッツ ０２１３９， ケンブリッジ， エリー ストリート ４０， スイート １１０
(72)発明者 カメン， ロバート
アメリカ合衆国 マサチューセッツ ０２１３９， ケンブリッジ， エリー ストリート ４０， スイート １１０
審査官 塩田 徳彦

(56)参考文献 特開２０１７－１８１２９（ＪＰ，Ａ）
特表２０２０－５１８０８３（ＪＰ，Ａ）
米国特許出願公開第２０１７／０２１２９８４（ＵＳ，Ａ１）
国際公開第２０１７／１８４５９０（WO，Ａ１）
国際公開第２０１６／１２８０６０（WO，Ａ１）
米国特許出願公開第２００９／０１３０１３４（ＵＳ，Ａ１）

(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ１６Ｂ ５／００ － ９９／００
Ｃ１２Ｎ １５／１２
Ｃ０７Ｋ １４／４７
Ｃ１２Ｑ １／０２
Ｃ１２Ｑ １／６８
Ｃ１２Ｎ １５／３３
Ｃ０７Ｋ １４／００５
Ｃ１２Ｎ ５／１０

C07K　7／08
C07K　14／74
C12N　15／63
A61K　39／00
A61P　37／04
A61P　35／00