



(19)中華民國智慧財產局

(12)發明說明書公開本

(11)公開編號：TW 201923637 A

(43)公開日：中華民國 108 (2019) 年 06 月 16 日

(21)申請案號：107135870

(22)申請日：中華民國 107 (2018) 年 10 月 11 日

(51)Int. Cl. : G06F19/22 (2011.01)

G06F19/24 (2011.01)

(30)優先權：2017/10/10 美國

62/570,569

(71)申請人：美商葛利史東腫瘤科技公司 (美國) GRITSTONE ONCOLOGY, INC. (US)
美國(72)發明人：布理克 蘇利文 布萊登 BULIK-SULLIVAN, BRENDAN (US)；布雪 湯瑪士
法蘭西斯 BOUCHER, THOMAS FRANCIS (US)；葉嫩史基 羅門 YELENSKY,
ROMAN (US)

(74)代理人：陳長文

申請實體審查：無 申請專利範圍項數：31 項 圖式數：29 共 305 頁

(54)名稱

利用熱點識別新抗原

NEOANTIGEN IDENTIFICATION USING HOTSPOTS

(57)摘要

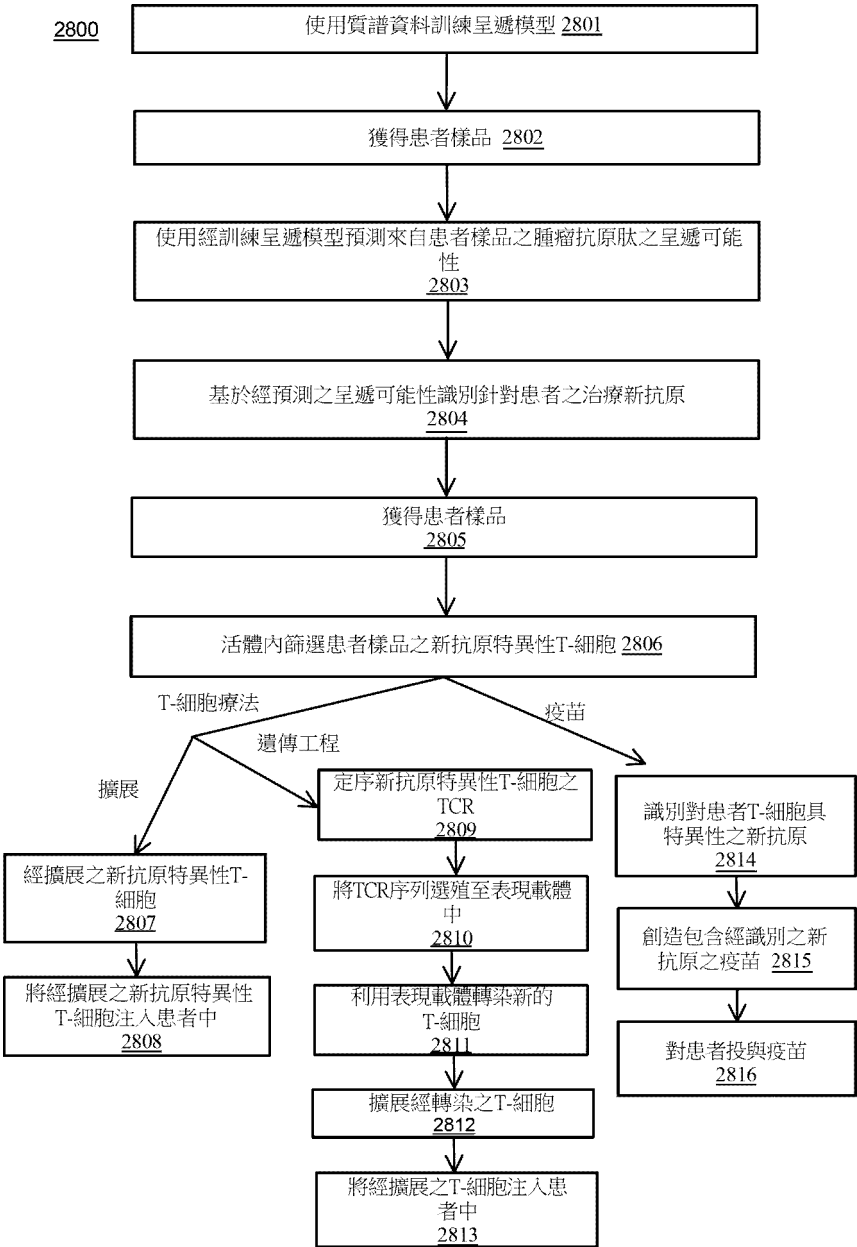
本文提供一種識別可能在受試者之腫瘤細胞表面上呈遞之新抗原之方法。藉由對該受試者之腫瘤細胞定序獲得腫瘤新抗原之肽序列。將新抗原各者之肽序列與該受試者之核苷酸定序資料之複數個 k-mer 區塊中之一或多個 k-mer 區塊相關聯；將該等肽序列及相關聯 k-mer 區塊輸入至機器學習呈遞模型中以產生針對腫瘤新抗原之呈遞可能性，各呈遞可能性表示新抗原由該受試者之腫瘤細胞表面上之 MHC 對偶基因呈遞之可能性。基於該等呈遞可能性選擇新抗原之子集。

A method for identifying neoantigens that are likely to be presented on a surface of tumor cells of a subject. Peptide sequences of tumor neoantigens are obtained by sequencing the tumor cells of the subject. The peptide sequence of each of the neoantigens is associated with one or more k-mer blocks of a plurality of k-mer blocks of the nucleotide sequencing data of the subject; The peptide sequences and the associated k-mer blocks are input into a machine-learned presentation model to generate presentation likelihoods for the tumor neoantigens, each presentation likelihood representing the likelihood that a neoantigen is presented by an MHC allele on the surfaces of the tumor cells of the subject. A subset of the neoantigens is selected based on the presentation likelihoods.

指定代表圖：

符號簡單說明：

- 2800 . . . 流程圖
- 2801 . . . 步驟
- 2802 . . . 步驟
- 2803 . . . 步驟
- 2804 . . . 步驟
- 2805 . . . 步驟
- 2806 . . . 步驟
- 2807 . . . 步驟
- 2808 . . . 步驟
- 2809 . . . 步驟
- 2810 . . . 步驟
- 2811 . . . 步驟
- 2812 . . . 步驟
- 2813 . . . 步驟
- 2814 . . . 步驟
- 2815 . . . 步驟
- 2816 . . . 步驟



【圖28】

【發明說明書】

【中文發明名稱】

利用熱點識別新抗原

【英文發明名稱】

NEOANTIGEN IDENTIFICATION USING HOTSPOTS

【技術領域】

【先前技術】

【0001】 基於腫瘤特異性新抗原之治療性疫苗及T-細胞療法作為下一代個人化癌症免疫療法有極大前景。¹⁻³具有高突變負擔之癌症(諸如非小細胞肺癌(NSCLC)及黑色素瘤)為此種療法之特別吸引人的靶，考慮到新抗原產生之相對更大可能性。^{4,5}早期證據顯示基於新抗原之疫苗接種可引起T-細胞反應⁶及新抗原靶向之T細胞療法可在選定患者之某些情況下造成腫瘤消退。⁷ MHC I類及MHC II類均對T-細胞反應有影響⁷⁰⁻⁷¹。

【0002】 然而，新抗原及識別新抗原之T細胞之識別已變成評估腫瘤反應^{77、110}、檢查腫瘤演變¹¹¹及設計下一代個人化療法¹¹²之中心挑戰。目前新抗原識別技術係費時且費力^{84、96}或不夠精確^{87、91-93}。雖然最近已證明，識別新抗原之T細胞為TIL之主要組分^{84、96、113、114}且於癌症患者之外周血中循環¹⁰⁷，但是識別新抗原反應性T細胞之目前方法具有下列三種限制之一些組合：(1)其依賴難以獲得之臨床樣品，諸如TIL^{97、98}或白血球採集物(leukaphereses)¹⁰⁷ (2)其需要不實用地篩選大的肽庫⁹⁵或(3)其依賴MHC多聚體，該等多聚體實際上可適用於僅少量MHC對偶基因。

【0003】 此外，已提出使用下一代定序、RNA基因表現及預測候選新抗原肽之MHC結合親和力併入基於突變之分析的初始方法⁸。然而，此

等提出之方法可無法建模整個抗原決定基產生過程，該過程除了基因表現及MHC結合外尚包含許多步驟(例如，TAP轉運、蛋白酶體裂解、MHC結合、肽-MHC複合體至細胞表面之轉運及/或TCR對MHC-I之識別；內吞或自噬、經由細胞外或溶酶體蛋白酶(例如，組織蛋白酶)之裂解、與CLIP肽競爭HLA-DM-催化之HLA結合、肽-MHC複合體至細胞表面之轉運及/或針對MHC-II之TCR識別)⁹。因此，現有方法可能遭受降低之低陽性預測值(PPV)。(圖1A)

【0004】 的確，由多組進行之藉由腫瘤細胞呈遞之肽之分析已顯示，使用基因表現及MHC結合親和力預測為待呈遞之肽之<5%可在腫瘤表面MHC上發現^{10、11} (圖1B)。在結合預測與MHC呈遞之間之此低相關性進一步藉由以下最近觀察結果得以加強：單獨結合限制性新抗原對於多個突變中之檢查點抑制劑反應之預測精確度提高之缺乏。¹²

【0005】 用於預測呈遞之現有方法之此低陽性預測值(PPV)提出關於基於新抗原之疫苗設計及基於新抗原之T細胞療法之問題。若使用具有低PPV之預測設計疫苗，則大多數患者不可能接受治療性新抗原及更少仍可能接受一個以上(甚至假設所有呈遞肽係免疫原性)。類似地，若基於具有低PPV之預測設計治療性T細胞，則大多數患者不可能接受對腫瘤新抗原具反應性之T細胞及使用下游實驗室技術後預測識別預測性新抗原之時間及物理資源成本可係過高。因此，利用目前方法之新抗原疫苗接種及T細胞療法於患有腫瘤之大量受試者中不可能成功。(圖1C)

【0006】 此外，先前方法僅使用順式作用突變產生候選新抗原，及很大程度上忽略考慮neo-ORF之額外來源，包括發生於多個腫瘤類型中且導致許多基因之異常剪接之剪接因子之突變¹³及產生或移除蛋白酶裂解位

點之突變。

【0007】最後，腫瘤基因組及轉錄組分析之標準方法可錯失導致候選新抗原之體細胞突變，這歸因於庫構建、外顯子組及轉錄組捕獲、定序或資料分析中之次優條件。同樣，標準腫瘤分析方法可無意地促進序列偽影或生殖系多型性作為新抗原，其各自導致低效使用疫苗容量或自體免疫風險。

【發明內容】

【0008】本文中揭示一種識別及選擇用於個人化癌症疫苗、用於T細胞療法或二者之新抗原之最佳化方法。首先，解決用於使用下一代定序(NGS)之新抗原候選識別之最佳化腫瘤外顯子組及轉錄組分析方法。此等方法建立在NGS腫瘤分析之標準方法上以確保跨所有類別之基因組改變推進最高易感性及特異性新抗原候選。其次，提出高-PPV新抗原選擇之新穎方法以克服特異性問題並確保針對疫苗納入及/或作為T細胞療法之靶推進之新抗原更可能引起抗腫瘤免疫。取決於實施例，此等方法包括訓練統計回歸或非線性深度學習模型，該模型對於跨不同長度之肽共用統計長度之多個長度之肽聯合建模肽-對偶基因圖譜分析以及按對偶基因模體。此等深度學習模型亦利用在測定肽之呈遞可能性中描述與肽序列相關之k-mer區塊之呈遞熱點之存在或不存在的參數。特定言之可設計及訓練該等非線性深度學習模型以將相同細胞中之不同MHC對偶基因處理為獨立，從而解決利用線性模型其將彼此干擾之問題。最後，解決針對基於新抗原之個人化疫苗設計及製造及針對用於T細胞療法之個人化新抗原特異性T細胞之產生的額外考慮。

【0009】本文中所揭示之模型勝過在結合親和力上訓練之申請專利

當時之技術水平之預測因子及基於MS肽資料之早期預測因子高達一個數量級。藉由更可靠地預測肽之呈遞，該模型使能使用臨床實務方法更時間有效及成本有效地識別用於個人化療法之新抗原特異性或腫瘤抗原特異性T細胞，該臨床實務方法使用有限體積之患者外周血，每例患者篩選幾個肽且不必要地依賴MHC多聚體。然而，於另一實施例中，可使用本文中所揭示之模型以使能藉由減少結合至MHC多聚體之肽之數目更時間有效及成本有效地使用MHC多聚體識別腫瘤抗原特異性T細胞，需篩選該等MHC多聚體以識別新抗原特異性T細胞或腫瘤抗原特異性T細胞。

【0010】本文中所揭示之模型對TIL新抗原決定基資料集及預期新抗原-反應性T細胞識別任務之預測性能證明，現在可藉由建模HLA處理及呈遞獲得治療上可用之新抗原決定基預測。總之，此工作提供用於靶向抗原之免疫療法之實用電腦模擬抗原識別，從而加速治癒患者之進展。

【圖式簡單說明】

【0011】本發明之此等及其他特徵、態樣及優點將關於下列描述及附圖變得更好理解，其中：

【0012】圖1A顯示新抗原識別之目前臨床方法。

【0013】圖1B顯示<5%之預測之結合肽於腫瘤細胞上呈遞。

【0014】圖1C顯示新抗原預測特異性問題之影響。

【0015】圖1D顯示結合預測不足夠用於新抗原識別。

【0016】圖1E顯示作為肽長度之函數之MHC-I呈遞概率。

【0017】圖1F顯示自普羅米加氏(Promega's)動態範圍標準產生之實例肽譜。

【0018】圖1G顯示特徵之添加如何增加模型陽性預測值。

【0019】圖2A為根據實施例之用於識別患者中之肽呈遞可能性之環境的概述。

【0020】圖2B及2C說明根據實施例之獲得呈遞資訊之方法。

【0021】圖3為說明根據一實施例之呈遞識別系統之電腦邏輯組件的高階方塊圖。

【0022】圖4說明根據一實施例之實例訓練資料集。

【0023】圖5說明與MHC對偶基因相關之實例網路模型。

【0024】圖6A說明根據一實施例之由MHC對偶基因共享之實例網路模型 $NN_H(\cdot)$ 。

【0025】圖6B說明根據另一實施例之由MHC對偶基因共享之實例網路模型 $NN_H(\cdot)$ 。

【0026】圖7說明使用實例網路模型產生與MHC對偶基因相關之肽之呈遞可能性。

【0027】圖8說明使用實例網路模型產生與MHC對偶基因相關之肽之呈遞可能性。

【0028】圖9說明使用實例網路模型產生與MHC對偶基因相關之肽之呈遞可能性。

【0029】圖10說明使用實例網路模型產生與MHC對偶基因相關之肽之呈遞可能性。

【0030】圖11說明使用實例網路模型產生與MHC對偶基因相關之肽之呈遞可能性。

【0031】圖12說明使用實例網路模型產生與MHC對偶基因相關之肽之呈遞可能性。

【0032】 圖13A說明NSCLC患者中之突變負擔之樣品頻率分佈。

【0033】 圖13B說明根據實施例之基於患者是否滿足最小突變負擔之納入標準選擇之患者之模擬疫苗中之呈遞的新抗原的數目。

【0034】 圖13C比較根據實施例之在與疫苗相關之選定患者(包括基於呈遞模型識別之治療子集)與疫苗相關之選定患者(包括通過目前申請專利當時之技術水平之模型識別之治療子集)之間模擬疫苗中之呈遞的新抗原的數目。

【0035】 圖13D比較在與疫苗相關之選定患者(包括基於針對HLA-A*02:01之單個按對偶基因呈遞模型識別之治療子集)與疫苗相關之選定患者(包括基於針對HLA-A*02:01及HLA-B*07:02之兩個按對偶基因呈遞模型識別之治療子集)之間模擬疫苗中之呈遞的新抗原的數目。根據實施例，將疫苗容量設置為 $v=20$ 個抗原決定基。

【0036】 圖13E比較根據實施例之在基於突變負擔選擇之患者與藉由期望效用分數選擇之患者之間模擬疫苗中之呈遞的新抗原的數目。

【0037】 圖14比較不同版本之MS模型及於人類腫瘤中建模HLA呈遞肽之更早期方法²⁹在40%召回下之陽性預測值(PPV)，當在包含五種不同留存測試樣品之測試集上測試各模型時，各測試樣品包括具有經呈遞肽與未經呈遞肽之1:2500比率之留存腫瘤樣品。

【0038】 圖15A比較當在五種留存測試樣品上測試模型時，跨使用呈遞熱點參數之呈遞模型及不使用呈遞熱點參數之呈遞模型之召回的平均陽性預測值(PPV)。

【0039】 圖15B比較當在留存測試樣品0上測試模型時，使用呈遞熱點參數之呈遞模型及不使用呈遞熱點參數之呈遞模型之精度及召回曲線。

【0040】圖15C比較當在留存測試樣品1上測試模型時，使用呈遞熱點參數之呈遞模型及不使用呈遞熱點參數之呈遞模型之精度及召回曲線。

【0041】圖15D比較當在留存測試樣品2上測試模型時，使用呈遞熱點參數之呈遞模型及不使用呈遞熱點參數之呈遞模型之精度及召回曲線。

【0042】圖15E比較當在留存測試樣品3上測試模型時，使用呈遞熱點參數之呈遞模型及不使用呈遞熱點參數之呈遞模型之精度及召回曲線。

【0043】圖15F比較當在留存測試樣品4上測試模型時，使用呈遞熱點參數之呈遞模型及不使用呈遞熱點參數之呈遞模型之精度及召回曲線。

【0044】圖16比較針對包含取自具有至少一個先已存在之T細胞反應之患者之測試樣品的測試集，藉由使用呈遞熱點參數之呈遞模型及藉由不使用呈遞熱點參數之呈遞模型識別之前5、10、20及30位肽之跨距由T細胞識別之體細胞突變之肽的比例。

【0045】圖17A描述對九例患者之患者-特異性新抗原肽池之T細胞反應之檢測。

【0046】圖17B描述對四例患者之個別患者-特異性新抗原肽之T細胞反應之檢測。

【0047】圖17C描述患者CU04之ELISpot孔之實例圖像。

【0048】圖18A描述於經HLA-匹配之健康供體中利用新抗原之對照實驗之結果。

【0049】圖18B描述於經HLA-匹配之健康供體中利用新抗原之對照實驗之結果。

【0050】圖19描述對圖17A中所述之各供體及各活體外擴展之PHA陽性對照之T細胞反應的檢測。

【0051】 圖20A描述針對患者CU04對池#2中之各個別患者-特異性新抗原肽之T細胞反應的檢測。

【0052】 圖20B描述針對患者CU04之三次訪問各者及針對患者1-024-002之兩次訪問各者對個別患者-特異性新抗原肽之T細胞反應的檢測，各訪問發生在不同時間點。

【0053】 圖20C描述針對患者CU04之兩次訪問各者及針對患者1-024-002之兩次訪問各者對個別患者-特異性新抗原肽及對患者特異性新抗原肽池之T細胞反應的檢測，各訪問發生在不同時間點。

【0054】 圖21描述針對圖17A之患者對兩例患者特異性新抗原肽池及對DMSO陰性對照之T細胞反應的檢測。

【0055】 圖22比較當預測新抗原決定基藉由MHC II類分子之呈遞時，使用呈遞熱點參數之呈遞模型與不使用呈遞熱點參數之呈遞模型之預測性能。

【0056】 圖23描述一種定序來自NSCLC患者之外周血之新抗原特異性記憶T細胞之TCR的方法。

【0057】 圖24描述用於將TCR引入受體細胞之TCR構築體之示例性實施例。

【0058】 圖25描述用於將TCR選殖至表現系統中用於療法開發之示例性P526構築體主鏈核苷酸序列。

【0059】 圖26描述用於將患者新抗原特異性TCR、純系型1 TCR選殖至表現系統中用於療法開發之示例性構築體序列。

【0060】 圖27描述用於將患者新抗原特異性TCR、純系型3選殖至表現系統中用於療法開發之示例性構築體序列。

【0061】圖28為根據實施例之對患者提供定製、新抗原特異性治療之方法的流程圖。

【0062】圖29說明用於實施圖1及圖3中所示之實體之實例電腦。

【實施方式】

I. 定義

【0063】總之，意欲將本申請專利範圍及本說明書中所用之術語解釋為具有由一般技術者所理解之普通含義。以下定義某些術語以提供額外清晰。在普通含義與所提供定義之間衝突之情況下，將使用所提供定義。

【0064】如本文中所用，術語「抗原」為誘導免疫反應之物質。

【0065】如本文中所用，術語「新抗原」為具有至少一種改變之抗原，該改變使其不同於對應野生型、親本抗原，例如經由腫瘤細胞中之突變或對腫瘤細胞特異性之轉譯後修飾。新抗原可包含多肽序列或核苷酸序列。突變可包括框移或非框移插入或缺失、錯義或無義取代、剪接位點改變、基因組重排或基因融合或產生neoORF之任何基因組或表現改變。突變亦可包括剪接變異體。對腫瘤細胞特異性之轉譯後修飾可包括異常磷酸化。對腫瘤細胞特異性之轉譯後修飾亦可包括產生蛋白酶體之剪接抗原。參見Liepe等人，A large fraction of HLA class I ligands are proteasome-generated spliced peptides; Science. 2016年10月21日; 354(6310):354-358。

【0066】如本文中所用，術語「腫瘤新抗原」為存在於受試者之腫瘤細胞或組織中但非存在於受試者之對應正常細胞或組織中之新抗原。

【0067】如本文中所用，術語「基於新抗原之疫苗」為基於一或多個新抗原(例如，複數個新抗原)之疫苗構築體。

【0068】如本文中所用，術語「候選新抗原」為產生可表示新抗原之新序列之突變或其他畸變。

【0069】如本文中所用，術語「編碼區」為編碼蛋白質之基因之部分。

【0070】如本文中所用，術語「編碼突變」為發生於編碼區中之突變。

【0071】如本文中所用，術語「ORF」意指開放閱讀框。

【0072】如本文中所用，術語「NEO-ORF」為自突變或其他畸變(諸如剪接)產生之腫瘤特異性ORF。

【0073】如本文中所用，術語「錯義突變」為造成自一個胺基酸至另一個胺基酸之取代之突變。

【0074】如本文中所用，術語「無義突變」為造成自胺基酸至終止密碼子之取代之突變。

【0075】如本文中所用，術語「框移突變」為造成蛋白質框架之改變之突變。

【0076】如本文中所用，術語「插入或缺失」(indel)為一或多個核酸之插入或缺失。

【0077】如本文中所用，於兩個或更多個核酸或多肽序列之上下文中，術語「同一性」百分比係指當針對最大對應度比較及比對時，具有相同核苷酸或胺基酸殘基之特定百分比之兩個或更多個序列或子序列，如使用下述序列比較演算法(例如，BLASTP及BLASTN或對熟習者可得之其他演算法)中之一者或藉由目視檢查所量測得。取決於應用，「同一性」百分比可於正比較之序列之區域上，例如，在功能域上存在，或，替代地

於待比較之兩個序列之全長上存在。

【0078】 針對序列比較，通常一個序列充當比較測試序列之參考序列。當使用序列比較演算法時，將測試及參考序列輸入至電腦中，若必要，則指定子序列座標，及指定序列演算法程式參數。然後基於所指定之程式參數，該序列比較演算法計算測試序列相對於參考序列之序列同一性百分比。或者，可藉由特定核苷酸之組合之存在或不存在或針對轉譯序列，在選定序列位置處之胺基酸(例如，序列模體)建立序列相似性或不同。

【0079】 可(例如)藉由Smith及Waterman, Adv. Appl. Math. 2:482 (1981)之局部同源演算法，藉由Needleman及Wunsch, J. Mol. Biol. 48:443 (1970)之同源比對演算法，藉由Pearson及Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988)之對相似性方法之搜索，藉由此等演算法之電腦化實施(GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.)或藉由目視檢查(一般參見Ausubel等人，在下)進行序列之最佳化比對用於比較。

【0080】 適用於測定序列同一性百分比及序列相似性之演算法之一個實例為BLAST演算法，其述於Altschul等人，J. Mol. Biol. 215:403-410 (1990)中。用於進行BLAST分析之軟體係通過國家生物技術資訊中心(National Center for Biotechnology Information)公開可得。

【0081】 如本文中所用，術語「不終止或通讀」為造成自然終止密碼子之移除之突變。

【0082】 如本文中所用，術語「抗原決定基」為通常藉由抗體或T

細胞受體結合之抗原之特定部分。

【0083】如本文中所用，術語「免疫原性」為經由T細胞、B細胞或二者引起免疫反應之能力。

【0084】如本文中所用，術語「HLA結合親和力」「MHC結合親和力」意指特異性抗原與特異性MHC對偶基因之間之結合親和力。

【0085】如本文中所用，術語「誘餌」為用於增濃來自樣品之DNA或RNA之特異性序列的核酸探針。

【0086】如本文中所用，術語「變異體」為受試者之核酸與作為對照使用之參考人類基因組之間之差異。

【0087】如本文中所用，術語「變異體調用」為通常自定序之變異體之存在的演算法測定。

【0088】如本文中所用，術語「多型性」為生殖系變異體，即，於個體之所有含DNA細胞中發現之變異體。

【0089】如本文中所用，術語「體細胞變異體」為於個體之非生殖系細胞中產生之變異體。

【0090】如本文中所用，術語「對偶基因」為基因版本或基因序列版本或蛋白質版本。

【0091】如本文中所用，術語「HLA類型」為HLA基因對偶基因之補體。

【0092】如本文中所用，術語「無義介導之衰變」或「NMD」為mRNA藉由細胞之降解，由於過早的終止密碼子。

【0093】如本文中所用，術語「軀幹突變」為源於腫瘤之早期發展之突變且存在於腫瘤細胞之大部分。

【0094】如本文中所用，術語「次選殖突變」為源於腫瘤之晚期發展之突變且僅存在於腫瘤細胞之子集中。

【0095】如本文中所用，術語「外顯子組」為編碼蛋白質之基因組之子集。外顯子組可為基因組之集體外顯子。

【0096】如本文中所用，術語「邏輯回歸」為來自統計學之二進位資料之回歸模型，其中將因變數等於一之概率之分數對數建模為因變數之線性函數。

【0097】如本文中所用，術語「神經網路」為用於分類或回歸之機器學習模型，其由線性轉換接著通常經由隨機梯度下降及反向傳播訓練之要素方式非線性之多層組成。

【0098】如本文中所用，術語「蛋白質組」為藉由細胞、細胞組或個體表現及/或轉譯之所有蛋白質之集合。

【0099】如本文中所用，術語「肽組」為藉由細胞表面上之MHC-I或MHC-II呈遞之所有肽之集合。肽組可係指細胞或細胞集合之性質(例如，腫瘤肽組，意指包含腫瘤之所有細胞之肽組之聯合)。

【0100】如本文中所用，術語「ELISPOT」意指酶聯免疫吸附點檢定，其為監測人類及動物中之免疫反應之常見方法。

【0101】如本文中所用，術語「dextramer」為用於流式細胞計中之抗原-特異性T細胞染色之基於右旋糖苷之肽-MHC多聚體。

【0102】如本文中所用，術語「MHC多聚體」為包含多個肽-MHC單體單元之肽-MHC複合體。

【0103】如本文中所用，術語「MHC四聚體」為包含四個肽-MHC單體單元之肽-MHC複合體。

【0104】如本文中所用，術語「耐受或免疫耐受」為對一或多個抗原(例如，自體抗原)之免疫非反應性之狀態。

【0105】如本文中所用，術語「中心耐受」為藉由缺失自體反應性T細胞純系或藉由促進自體反應性T細胞純系分化成免疫抑制調節T細胞(Treg)於胸腺中受影響之耐受。

【0106】如本文中所用，術語「外周耐受」為藉由下調使中心耐受存活之自體反應性T細胞或使該等自體反應性T細胞無應答或促進此等T細胞分化成Treg於外周受影響之耐受。

【0107】術語「樣品」可包括藉助包括以下取自受試者之單一細胞或多個細胞或細胞片段或體液之等分試樣：靜脈穿刺、排泄、射出、按摩、活組織檢查、針吸、灌洗樣品、刮、手術切口或介入或此項技術中已知之其他方法。

【0108】術語「受試者」涵蓋細胞、組織或生物體、人類或非人類，無論是否於活體內、離體或於活體外、雄性或雌性。術語受試者包括哺乳動物(包括人類)。

【0109】術語「哺乳動物」涵蓋人類及非人類二者且包括(但不限於)人類、非人類靈長類動物、犬科動物、貓科動物、鼠科動物、牛科動物、馬科動物及豬科動物。

【0110】術語「臨床因素」係指受試者之狀況(例如，疾病活性或嚴重度)之量度。「臨床因素」涵蓋受試者之健康狀態之所有標誌物(包括非樣品標誌物)及/或受試者之其他特徵，諸如(不限於)年齡及性別。臨床因素可為可自受試者或確定條件下之受試者之樣品(或樣品群體)之評價獲得的分數、值或值集。臨床因素亦可藉由標誌物及/或其他參數(諸如基因表

現替代物)預測。臨床因素可包括腫瘤類型、腫瘤子型及吸煙史。

【0111】縮略語：MHC：主要組織相容性複合體；HLA：人類白血球抗原或人類MHC基因座；NGS：下一代定序；PPV：陽性預測值；TSNA：腫瘤特異性新抗原；FFPE：經福馬林固定，經石蠟包埋；NMD：無義介導之衰變；NSCLC：非細胞肺癌；DC：樹突狀細胞。

【0112】應注意，如本說明書及隨附申請專利範圍中所用，除非上下文中另有明確指示，否則單數形式「一」、「一個」及「該」包括複數指代物。

【0113】應瞭解，本文中未直接定義之任何術語具有通常與如本發明技術內所理解之其相關的含義。本文中討論某些術語以於描述本發明之態樣之組合物、裝置、方法及類似者及如何製備或使用其方面對從業者提供額外指導。應瞭解，可以超過一種方式陳述相同事情。因此，替代語言及同義詞可用於本文中所討論之術語中之任一者或多者。無論本文中是否詳細闡述或討論術語，無任何意義。提供一些同義詞或可替代方法、材料及類似者。除非明確指定，否則一或幾個同義詞或等效物之詳述不排除其他同義詞或等效物之使用。使用實例(包括術語之實例)僅係用於說明目的且不限制本文中本發明之態樣之範圍及含義。

【0114】出於所有目的，本說明書主體內所引用之所有參考文獻、發行之專利及專利申請案之全文以引用的方式併入本文中。

II. 識別新抗原之方法

【0115】本文中揭示識別來自受試者之腫瘤細胞之新抗原的方法，該等新抗原可能在該等腫瘤細胞之表面上呈遞。該方法包括自受試者之腫瘤細胞以及正常細胞獲得外顯子組、轉錄組及法包括自受試者之腫瘤細資

料。使用此核苷酸定序資料以獲得新抗原集中之各新抗原之肽序列。藉由比較來自腫瘤細胞之核苷酸定序資料與來自正常細胞之核苷酸定序資料識別該新抗原集。具體而言，該新抗原集中之各新抗原之肽序列包括使其不同於自受試者之正常細胞識別之對應野生型肽序列的至少一個改變。該方法還包括將該新抗原集中之各新抗原之肽序列編碼至對應數值向量中。各數值向量包含描述組成該肽序列之胺基酸及該肽序列中之胺基酸之位置的資訊。該方法還包括使新抗原各者之肽序列與受試者之核苷酸定序資料之複數個抗原之肽序區塊中之一或多個之肽序列編區塊相關聯；該方法還包括將該等數值向量及相關聯該方法還包區塊輸入至機器學習呈遞模型中以產生該新抗原集中之各抗體之呈遞可能性。各呈遞可能性表示對應新抗原藉由受試者之腫瘤細胞表面上之酸及該肽序列中之胺基酸之位置的資訊。該方法還包括使新抗原各者之肽序列與受試者資料集識別該等複數個參數。該訓練資料集包括針對複數個樣品中之各樣品，藉由質譜法獲得的標籤，該質譜法量測結合至經識別為存在於該樣品中之MHC合至經識別中之至少一個識別為存在於該呈遞可能性表將訓練肽序列編碼為數值向量，該等數值向量包含描述組成該等肽之胺基酸及該等肽中之該等胺基酸之位置的資訊；及針對該樣品之該等訓練肽序列各者，在該訓練肽序列與該等訓練肽序列之核苷酸定序資料之複數個練肽序列各區塊中之一或多個序列各者，區塊之間相關聯。該函數表示藉由機器學習呈遞模型作為輸入接收之數值向量及相關聯數值向量遞區塊與藉由機器學習呈遞模型基於該數值向量、相關聯器學習呈遞區塊及複數個參數作為輸出產生之呈遞可能性之間之關係。該方法還包括基於呈遞可能性選擇該新抗原集之子集以產生選定之新抗原集，及返回該選定之新抗原集。

【0116】選於一些實施例中，將數值向量輸入至機器學習呈遞模型中包括將該機器學習呈遞模型應用於新抗原之肽序列以產生MHC對偶基因各者之相依性分數。MHC對偶基因之相依性分數基於肽序列之特定位置處之特定胺基酸，指示MHC對偶基因是否將呈遞新抗原。於另外實施例中，將數值向量輸入至機器學習呈遞模型中還包括轉換相依性分數以產生各MHC對偶基因之對應按對偶基因可能性，該按對偶基因可能性指示對應MHC對偶基因將呈遞對應新抗原之可能性；及將該等按對偶基因可能性組合以產生新抗原之呈遞可能性。於一些實施例中，轉換相依性分數將新抗原之呈遞建模為跨MHC對偶基因之互相排斥。於替代實施例中，將數值向量輸入至機器學習呈遞模型中還包括轉換相依性分數之組合以產生呈遞可能性。於此等實施例中，轉換相依性分數之組合將新抗原之呈遞建模為MHC對偶基因之間之干涉。

【0117】於一些實施例中，該呈遞可能性集進一步藉由一或多個對偶基因非相互作用特徵識別。於此等實施例中，該方法還包括將該機器學習呈遞模型應用於對偶基因非相互作用特徵以產生對偶基因非相互作用特徵之相依性分數。該相依性分數指示對應新抗原之肽序列是否將基於對偶基因非相互作用特徵呈遞。於一些實施例中，該一或多個對偶基因非相互作用特徵包括指示各新抗原之肽序列之各k-mer區塊之呈遞熱點之存在或不存在中之一者的值。

【0118】於一些實施例中，該方法還包括將各MHC對偶基因之相依性分數與對偶基因非相互作用特徵之相依性分數組合、轉換各MHC對偶基因之組合之相依性分數以產生各MHC對偶基因之按對偶基因可能性及組合該等按對偶基因可能性以產生呈遞可能性。MHC對偶基因之按對偶

基因可能性指示MHC對偶基因將呈遞對應新抗原之可能性。於替代實施例中，該方法還包括將MHC對偶基因之相依性分數與對偶基因非相互作用特徵之相依性分數組合，及轉換該等組合之相依性分數以產生呈遞可能性。

【0119】於一些實施例中，該等MHC對偶基因包括兩個或更多個不同MHC對偶基因。

【0120】於一些實施例中，該等肽序列包括具有除了9個胺基酸外之長度之肽序列。

【0121】於一些實施例中，編碼肽序列包括使用獨熱編碼方案編碼肽序列。

【0122】於一些實施例中，複數個樣品包括以下中之至少一者：經基因改造以表現單個MHC對偶基因之細胞株、經基因改造以表現複數個MHC對偶基因之細胞株、獲自或源自複數例患者之人類細胞株、獲自複數例患者之新鮮或冷凍腫瘤樣品及獲自複數例患者之新鮮或冷凍組織樣品。

【0123】於一些實施例中，該訓練資料集還包括與該等肽中之至少一者之肽-MHC結合親和力測量值相關聯之資料及與該等肽中之至少一者之肽-MHC結合穩定性測量值相關聯之資料中之至少一者。

【0124】於一些實施例中，該呈遞可能性集進一步藉由受試者之MHC對偶基因之表現水平(如藉由RNA-seq或質譜法所量測)識別。

【0125】於一些實施例中，該呈遞可能性集進一步藉由包括該新抗原集中之新抗原與MHC對偶基因之間之預測親和力及新抗原編碼之肽-MHC複合體之預測穩定性中之至少一者的特徵識別。

【0126】於一些實施例中，該數值可能性集進一步藉由包括其源蛋白序列內側接新抗原編碼之肽序列之C-端序列及其源蛋白序列內側接新抗原編碼之肽序列之N-端序列中之至少一者的特徵識別。

【0127】於一些實施例中，選擇該選定之新抗原集包括基於機器學習呈遞模型選擇具有相對於未經選擇之新抗原增加之在腫瘤細胞表面呈遞之可能性的新抗原。

【0128】於一些實施例中，選擇該選定之新抗原集包括基於機器學習呈遞模型選擇具有相對於未經選擇之新抗原增加之能誘導受試者之腫瘤特異性免疫反應之可能性的新抗原。

【0129】於一些實施例中，選擇該選定之新抗原集包括基於呈遞模型選擇具有相對於未經選擇之新抗原增加之能藉由專業抗原呈遞細胞(APC)呈遞至初始T細胞之可能性的新抗原。於此等實施例中，視情況該APC為樹突狀細胞(DC)。

【0130】於一些實施例中，選擇該選定之新抗原集包括基於機器學習呈遞模型選擇具有相對於未經選擇之新抗原降低之經受經由中樞或外周耐受之抑制之可能性的新抗原。

【0131】於一些實施例中，選擇該選定之新抗原集包括基於機器學習呈遞模型選擇具有相對於未經選擇之新抗原降低之能誘導受試者之正常組織之自體免疫反應之可能性的新抗原。

【0132】於一些實施例中，該一或多個腫瘤細胞係選自由以下組成之群：肺癌、黑色素瘤、乳癌、卵巢癌、前列腺癌、腎癌、胃癌、結腸癌、睪丸癌、頭頸癌、胰癌、腦癌、B細胞淋巴瘤、急性骨髓性白血病、慢性骨髓性白血病、慢性淋巴球性白血病及T細胞淋巴球性白血病、非小

細胞肺癌及小細胞肺癌。

【0133】於一些實施例中，該方法還包括產生用於自該選定之新抗原集構建個人化癌症疫苗之輸出。於此等實施例中，用於個人化癌症疫苗之輸出可包括編碼該選定之新抗原集之至少一個肽序列或至少一個核苷酸序列。

【0134】於一些實施例中，該機器學習呈遞模型為神經網路模型。於此等實施例中，該神經網路模型可包括針對MHC對偶基因之複數個網路模型，各網路模型經分配給該等MHC對偶基因之對應MHC對偶基因且包含於一或多個層中排列之一系列節點。於此等實施例中，該神經網路模型可藉由更新該神經網路模型之參數來訓練，至少兩個網路模型之參數經聯合更新用於至少一個訓練迭代。於一些實施例中，該機器學習呈遞模型可為包含一或多個節點層之深度學習模型。

【0135】於一些實施例中，該等MHC對偶基因為I類MHC對偶基因。

【0136】本文中亦揭示包括電腦處理器及記憶存儲電腦程式指令之電腦系統。當電腦程式指令藉由電腦處理器執行時，該等指令造成電腦處理器進行以上所討論之方法中之任一者。

III.識別新抗原之腫瘤特異性突變

【0137】本文中亦揭示識別某些突變(例如，於癌細胞中呈遞之變異體或對偶基因)之方法。特定言之，此等突變可於患有癌症之受試者之癌細胞之基因組、轉錄組、蛋白質組或外顯子組中呈遞，但是不於該受試者之正常組織中呈遞。

【0138】若腫瘤中之基因突變導致專門腫瘤中之蛋白質之胺基酸序

列之變化，則可認為其可用於腫瘤之免疫靶向。可用突變包括：(1)導致蛋白質中之不同胺基酸之非同義突變；(2)通讀突變，其中終止密碼子經修飾或缺失，導致具有C-端處之新穎腫瘤特異性序列之更長蛋白質之轉譯；(3)剪接位點突變，其導致成熟mRNA中之內含子之納入及因此獨特腫瘤特異性蛋白質序列；(4)染色體重排，其產生具有2個蛋白質連接處之腫瘤特異性序列之嵌合蛋白(即，基因融合)；(5)框移突變或缺失，其導致具有新穎腫瘤特異性蛋白質序列之新開放閱讀框。突變亦可包括下列中之一或多者：非框移插入或缺失、錯義或無義取代、剪接位點改變、基因組重排或基因融合或產生neoORF之任何基因組或表現改變。

【0139】產生於(例如)腫瘤細胞中之剪接位點突變、框移突變、通讀突變或基因融合突變之具有突變之肽或經突變多肽可藉由對腫瘤細胞相對於正常細胞中之DNA、RNA或蛋白質定序來識別。

【0140】突變亦可包括經先前識別之腫瘤特異性突變。已知腫瘤突變可見於癌症中之體細胞突變目錄(Catalogue of Somatic Mutations in Cancer, COSMIC)資料庫。

【0141】各種方法可用於檢測個體之DNA或RNA中之特定突變或對偶基因之存在。此領域之進展已提供精確、容易且便宜的大規模SNP基因分型。例如，已描述若干技術，包括動態對偶基因特異性雜交(DASH)、微板陣列對角線凝膠電泳(MADGE)、焦磷酸定序、寡核苷酸特異性連接、TaqMan系統以及各種DNA「晶片」技術，諸如Affymetrix SNP晶片。此等方法利用目標基因區之擴增，通常藉由PCR。仍其他方法，基於藉由侵入性裂解，接著質譜法或固定掛鎖探針及滾環擴增之小信號分子之產生。以下概述檢測特異性突變之技術中已知之方法中之若干。

【0142】基於PCR之檢測意指可包括複數個標誌物之同時多重擴增。例如，選擇PCR引子以產生大小不重疊且可經同時分析之PCR產物係此項技術中熟知。或者，可利用經差別標示及因此各者可經差別檢測之引子擴增不同標誌物。當然，基於雜交之檢測意指允許樣品中之多個PCR產物之差別檢測。其他技術係此項技術中已知以允許複數個標誌物之多重分析。

【0143】已開發若干方法以促進基因組DNA或細胞RNA中之單個核苷酸多型性之分析。例如，單鹼基多型性可藉由使用特異性抗核酸外切酶核苷酸檢測，如(例如)於Mundy, C. R. (美國專利案第4,656,127號)中所揭示。根據該方法，允許與立即至多態位點之對偶基因序列3'互補之引子與獲自特定動物或人類之靶分子雜交。若靶分子上之多態位點含有與存在之特定抗核酸外切酶核苷酸衍生物互補之核苷酸，則該將衍生物併入至雜交引子之末端。此併入使引子抗核酸外切酶，及從而允許其檢測。因為樣品之抗核酸外切酶衍生物之同一性係已知，所以引子變得抗核酸外切酶之發現結果揭示，存在於靶分子之多態位點中之核苷酸與反應中使用之核苷酸衍生物之特徵互補。此方法具有不需要測定大量外來序列資料之優點。

【0144】基於溶液之方法可用於測定多態位點之核苷酸之同一性。Cohen, D.等人(法國專利2,650,840；PCT申請案第WO91/02087號)。如於美國專利案第4,656,127號之Mundy方法中，採用與立即至多態位點之對偶基因序列3'互補之引子。該方法使用標示二去氧核苷酸衍生物測定該位點之核苷酸之同一性，若該等衍生物與多態位點之核苷酸互補，則將其併入引子之末端。藉由Goelet, P.等人(PCT申請案第92/15712號)描述被稱為Genetic Bit分析或GBA之替代方法。Goelet, P.等人之方法使用標示終

止子及引子之混合物，該引子與至多態位點之序列3'互補。因此藉由存在於正評價之靶分子之多態位點中之核苷酸及與該核苷酸互補，測定併入的經標記終止子。與Cohen等人(法國專利2,650,840；PCT申請案第WO91/02087號)之方法相反，Goelet, P.等人之方法可為非均相檢定，其中將引子或靶分子固定於固相中。

【0145】已描述用於檢定DNA中之多態位點之若干引子引導之核苷酸併入程序(Komher, J. S.等人，Nucl. Acids. Res. 17:7779-7784 (1989)；Sokolov, B. P., Nucl. Acids Res. 18:3671 (1990)；Syvanen, A.-C.等人，Genomics 8:684-692 (1990)；Kuppuswamy, M. N.等人，Proc. Natl. Acad. Sci. (U.S.A.) 88:1143-1147 (1991)；Prezant, T. R.等人，Hum. Mutat. 1:159-164 (1992)；Ugozzoli, L.等人，GATA 9:107-112 (1992)；Nyren, P.等人，Anal. Biochem. 208:171-175 (1993))。此等方法不同於GBA，因為其利用經標記去氧核苷酸之併入來區分多態位點上之鹼基。以此形式，因為信號與併入之去氧核苷酸之數目成比例，所以發生於相同核苷酸之運行中之多型性可導致與該運行之長度成比例之信號(Syvanen, A.-C.等人，Amer. J. Hum. Genet. 52:46-59 (1993))。

【0146】許多主動性直接自數百萬DNA或RNA之個別分子並行獲得序列資訊。藉由合成技術之即時單分子定序依賴螢光核苷酸之檢測，因為其經併入與正進行定序之模板互補之DNA新生股。於一方法中，將長度為30至50個鹼基之寡核苷酸在5'端處共價錨定至玻璃蓋片。此等錨定股進行兩種功能。首先，若利用與表面結合寡核苷酸互補之捕獲尾配置模板，則其充當靶向模板股之捕獲位點。其亦充當用於形成序列閱讀之基礎之模板導向引子延伸的引子。捕獲引子作為序列測定之固定位置位點起作用，

該序列測定使用多個染料-連接子之合成、檢測及化學裂解週期以移除染料。各週期由以下組成：添加聚合酶/標示核苷酸混合物、將染料沖洗、成像及裂解。於替代方法中，將聚合酶用螢光供體分子修飾及在載玻片上固定，同時將各核苷酸用連接至 γ -磷酸鹽之受體螢光部分進行顏色編碼。當將核苷酸重新併入鏈中時，系統檢測螢光標記之聚合酶與螢光修飾之核苷酸之間之相互作用。其他合成定序技術亦存在。

【0147】 可使用任何適宜合成定序平臺識別突變。如上所述，四個主要合成定序平臺係目前可得：來自Roche/454 Life Sciences之基因組定序儀、來自Illumina/Solexa之1G分析儀、來自Applied BioSystems之SOLiD系統及來自Helicos Biosciences之Heliscope系統。亦已藉由Pacific BioSciences及VisiGen Biotechnologies描述合成定序平臺。於一些實施例中，使正定序之複數個核酸分子結合至支撐物(例如，固體支撐物)。為將核酸固定在支撐物上，可在模板之3'及/或5'端添加捕獲序列/通用引發位點。核酸可藉由使捕獲序列與共價連接至支撐物之互補序列雜交結合至支撐物。捕獲序列(亦稱作通用捕獲序列)為與連接至可雙重充當通用引子之支撐物之序列互補之核酸序列。

【0148】 作為捕獲序列之替代，可將偶合對(諸如，例如，抗體/抗原、受體/配位體或抗生物素蛋白(avidin)-生物素對，如於(例如)美國專利申請案第2006/0252077號中所述)之一員連接至各片段以待在塗覆有該偶合對之各自第二成員之表面上捕獲。

【0149】 於捕獲後，可(例如)藉由(例如)如於實例中及於美國專利案第7,283,337號中所述之單分子檢測/定序(包括模板相依性合成定序)分析序列。於合成定序中，在聚合酶之存在下，將表面結合分子暴露於複數個

經標記核苷酸三磷酸酯。藉由併入生長鏈之3'端之經標記核苷酸之順序測定模板之序列。此可即時完成或可以分步重複模式完成。針對即時分析，可併入各核苷酸之不同光學標籤且可利用多鐳射來刺激併入之核苷酸。

【0150】 定序亦可包括其他大規模並行定序或下一代定序(NGS)技術及平臺。大規模並行定序技術及平臺之額外實例為Illumina HiSeq或MiSeq、Thermo PGM或Proton、Pac Bio RS II或Sequel、Qiagen's Gene Reader及Oxford Nanopore MinION。可使用額外類似目前大規模並行定序技術以及此等技術之下代。

【0151】 可利用任何細胞類型或組織獲得於本文中所述方法中使用之核酸樣品。例如，DNA或RNA樣品可獲自腫瘤或體液(例如，藉由已知技術(例如，靜脈穿刺)獲得之血液或唾液)。或者，可在乾樣品(例如，毛髮或皮膚)上進行核酸測試。此外，可自腫瘤獲得樣品用於定序及可自正常組織獲得另一樣品用於定序，其中正常組織為與腫瘤相同之組織類型。可自腫瘤獲得樣品用於定序及可自正常組織獲得另一樣品用於定序，其中正常組織為相對於腫瘤之不同組織類型。

【0152】 腫瘤可包括下列中之一或多者：肺癌、黑色素瘤、乳癌、卵巢癌、前列腺癌、腎癌、胃癌、結腸癌、睪丸癌、頭頸癌、胰癌、腦癌、B細胞淋巴瘤、急性骨髓性白血病、慢性骨髓性白血病、慢性淋巴球性白血病及T細胞淋巴球性白血病、非小細胞肺癌及小細胞肺癌。

【0153】 或者，可使用蛋白質質譜法識別或驗證結合至腫瘤細胞上之MHC蛋白之突變肽的存在。肽可自腫瘤細胞或自腫瘤免疫沉澱之HLA分子酸溶離，及然後使用質譜法識別。

IV.新抗原

【0154】新抗原可包括核苷酸或多肽。例如，新抗原可為編碼多肽序列之RNA序列。因此可用於疫苗之新抗原可包括核苷酸序列或多肽序列。

【0155】本文中揭示包括藉由本文中所揭示方法識別之腫瘤特異性突變之單離肽、包括已知腫瘤特異性突變之肽及藉由本文中所揭示方法識別之突變體多肽或其片段。新抗原肽可述於其編碼序列之上下文中，其中新抗原包括編碼相關多肽序列之核苷酸序列(例如，DNA或RNA)。

【0156】藉由新抗原核苷酸序列編碼之一或多個多肽可包括以下中之至少一者：以小於1000 nM之IC₅₀值與MHC之結合親和力，針對8至15、8、9、10、11、12、13、14或15個胺基酸長度之MHC I類肽，該肽內或附近存在促進蛋白酶體裂解的序列模體及存在促進TAP轉運之序列模體。針對6至30、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30個胺基酸長度之MHC II類肽，該肽內或附近存在促進藉由細胞外或溶酶體蛋白酶(例如，組織蛋白酶)或HLA-DM催化之HLA結合之裂解之序列模體。

【0157】一或多個新抗原可在腫瘤之表面上呈遞。

【0158】一或多個新抗原於患有腫瘤之受試者中可係免疫原性，例如，能引起受試者之T細胞反應或B細胞反應。

【0159】誘導受試者之自體免疫反應之一或多個新抗原可自針對患有腫瘤之受試者之疫苗產生背景中考慮排除。

【0160】至少一個新抗原肽分子之尺寸可包括(但不限於)約5、約6、約7、約8、約9、約10、約11、約12、約13、約14、約15、約16、約17、約18、約19、約20、約21、約22、約23、約24、約25、約26、約

27、約28、約29、約30、約31、約32、約33、約34、約35、約36、約37、約38、約39、約40、約41、約42、約43、約44、約45、約46、約47、約48、約49、約50、約60、約70、約80、約90、約100、約110、約120或更多個胺基分子殘基，及其中可衍生之任何範圍。於特定實施例中，新抗原肽分子係等於或小於50個胺基酸。

【0161】 新抗原肽及多肽可為：針對MHC I類，長度為15個殘基或更少及通常由約8與約11個殘基之間(特定言之，9或10個殘基)組成；針對MHC II類，6至30個殘基(包含端值)。

【0162】 如所需，則可以若干方式設計更長肽。於一種情況下，當HLA對偶基因上之肽之呈遞可能性經預測或已知時，更長肽可由以下中之任一者組成：(1) 2至5個胺基酸朝向各對應基因產物之N端及C端擴展之個別呈遞肽；(2)具有針對各者之擴展序列之呈遞肽中之一些或所有的濃度。於另一種情況下，當定序揭示腫瘤中存在長的(>10個殘基)新抗原決定模體列(例如，由於導致新穎肽序列之框移、通讀或內含子納入)時，更長肽將由以下組成：(3)新穎腫瘤特異性胺基酸之全拉伸—因此繞過最強呈遞HLA之較短肽之計算選擇或活體外基於測試之選擇的需要。於兩種情況下，更長肽之使用允許藉由患者-細胞之內源處理及可導致更有效抗原呈遞及T細胞反應之誘導。

【0163】 新抗原肽及多肽可在HLA蛋白上呈遞。於一些態樣中，新抗原肽及多肽在具有較野生型肽更大親和力之HLA蛋白上呈遞。於一些態樣中，新抗原肽或多肽可具有以下之IC₅₀：至少小於5000 nM、至少小於1000 nM、至少小於500 nM、至少小於250 nM、至少小於200 nM、至少小於150 nM、至少小於100 nM、至少小於50 nM或更少。

【0164】 於一些態樣中，當投與給受試者時，新抗原肽及多肽不誘導自體免疫反應及/或調用免疫耐受。

【0165】 亦提供包含至少兩種或更多種新抗原肽之組合物。於一些實施例中，該組合物含有至少兩種不同肽。至少兩種不同肽可衍生自相同多肽。不同多肽意指肽因長度、胺基酸序列或二者變化。該等肽衍生自已知或已發現含有腫瘤特異性突變之任何多肽。可衍生新抗原肽之適宜多肽可見於(例如) COSMIC資料庫中。COSMIC管理人類癌症之體細胞突變之綜合資訊。該肽含有腫瘤特異性突變。於一些態樣中，該腫瘤特異性突變為針對特定癌症類型之驅動子突變。

【0166】 可修飾具有所需活性或性質之新抗原肽及多肽以提供某些所需屬性(例如，改善之藥理學特性)，同時增加或至少實質上保留未修飾肽之所有生物活性以結合期望MHC分子並激活適宜T細胞。例如，新抗原肽及多肽可經受各種變化，諸如取代(保守或非保守)，其中此等變化可提供其使用之某些優點，諸如改善之MHC結合、穩定性或呈遞。保守取代意指用生物上及/或化學上相似之胺基酸殘基置換另一個，例如，一個疏水性殘基置換另一個，或一個極性殘基置換另一個。該等取代包括諸如下之組合：Gly、Ala；Val、Ile、Leu、Met；Asp、Glu；Asn、Gln；Ser、Thr；Lys、Arg；及Phe、Tyr。亦可使用D-胺基酸探測單胺基酸取代之效應。可使用如於(例如) Merrifield, Science 232:341-347 (1986), Barany & Merrifield, The Peptides, Gross 及 Meienhofer 編輯 (N.Y., Academic Press)，第1至284頁(1979)及Stewart及Young，Solid Phase Peptide Synthesis, (Rockford, Ill., Pierce)，第二版(1984)中所述之熟知肽合成程序作出此等修飾。

【0167】 肽及多肽利用各種胺基酸類似物或非天然胺基酸之修飾可特定言之可用於增加活體內肽及多肽之穩定性。可以許多方法檢定穩定性。例如，已使用肽酶及各種生物培養基(諸如人類血漿及血清)測試穩定性。參見，例如，Verhoef等人，Eur. J. Drug Metab Pharmacokin. 11:291-302 (1986)。可使用25%人類血清(v/v)檢定方便地測定肽之半衰期。方案一般係如下。在使用之前藉由離心破壞混合人類血清(AB型，非熱滅活)。然後將血清用RPMI組織培養基稀釋至25%並用於測試肽穩定性。在預定時間間隔移除少量反應溶液並添加至6%三氟乙酸水溶液或乙醇中。將混濁反應樣品冷卻(4℃) 15分鐘及然後旋轉以使沉澱血清蛋白成小球。然後藉由逆相HPLC使用穩定性特定層析條件測定肽之存在。

【0168】 可修飾肽及多肽以提供除了改善之血清半衰期外之所需屬性。例如，肽誘導CTL活性之能力可藉由連接至序列而增強，該序列含有能誘導T輔助細胞反應之至少一個抗原決定基。免疫原肽/T輔助偶聯物可藉由間隔子分子連接。間隔子通常包括相對小的中性分子(諸如胺基酸或胺基酸類似物)，其在生理條件下實質上不帶電。間隔子通常選自(例如)Ala、Gly或非極性胺基酸或中極性胺基酸之其他中性間隔子。應瞭解視情況存在之間隔子不需要包含相同殘基及因此可為異寡聚物或同寡聚物。當存在時，間隔子通常將為至少一或兩個殘基，更通常三至六個殘基。或者，可在無間隔子之情況下將肽連接至T輔助肽。

【0169】 可將新抗原肽直接或經由T輔助肽之胺基或羧基末端處之間隔子連接至T輔助肽。可將新抗原肽或T輔助肽之胺基末端醯化。示例性T輔助肽包括破傷風類毒素830-843、流行性感冒307-319、瘧疾環子孢子382-398及378-389。

【0170】 可藉由熟習此項技術者已知之任何技術製備蛋白質或肽，該技術包括通過標準分子生物技術之蛋白質、多肽或肽之表現、蛋白質或肽自天然來源之單離或蛋白質或肽之化學合成。對應於各種基因之核苷酸及蛋白質、多肽及肽序列先前已經揭示且可見於一般技術者已知之電腦資料庫。一個此資料庫為位於國立衛生研究院(National Institutes of Health)網站上之 National Center for Biotechnology Information's Genbank and GenPept資料庫。可使用本文中所揭示或如為一般技術者已知之技術擴增及/或表現已知基因之編碼區。或者，蛋白質、多肽及肽之各種商業製劑為一般技術者已知。

【0171】 於另一態樣中，新抗原包括編碼新抗原肽或其部分之核酸(例如，多核苷酸)。多核苷酸可為(例如) DNA、cDNA、PNA、CAN、RNA (例如，mRNA)、單股及/或雙股或初始或穩定形式之多核苷酸(諸如，例如，具有硫代磷酸酯主鏈之多核苷酸)或其組合且其可含有或不含有間隔子。仍另一態樣提供能表現多肽及其部分之表現載體。不同細胞類型之表現載體係此項技術中熟知且可在無不當實驗之情況下選擇。一般而言，以用於表現之正確方向及正確閱讀框將DNA插入表現載體(諸如質體)中。若需要，則可將DNA連接至藉由期望宿主識別之適宜轉錄及轉譯調節控制核苷酸序列，雖然此等控制一般於表現載體中可得。然後將載體通過標準技術引入該宿主中。指導可見於(例如) Sambrook等人，(1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y中。

IV.疫苗組合物

【0172】 本文中亦揭示能引起特異性免疫反應(例如，腫瘤特異性免

疫反應)之免疫原組合物(例如，疫苗組合物)。疫苗組合物通常包含複數個新抗原，例如，使用本文中所述方法選擇之新抗原。亦可將疫苗組合物稱作疫苗。

【0173】 疫苗可含有1與30個之間的肽、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或30個不同肽、6、7、8、9、10、11、12、13或14個不同肽，或12、13或14個不同肽。肽可包括轉譯後修飾。疫苗可含有1與100個之間或更多個核苷酸序列、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100個或更多個不同核苷酸序列、6、7、8、9、10、11、12、13或14個不同核苷酸序列，或12、13或14個不同核苷酸序列。疫苗可含有1與30個之間的新抗原序列、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100個或更多個不同新抗

原序列、6、7、8、9、10、11、12、13或14個不同新抗原序列，或12、13或14個不同新抗原序列。

【0174】 於一實施例中，選擇編碼其之不同肽及/或多肽或核苷酸序列使得肽及/或多肽能與不同MHC分子(諸如不同MHC I類分子及/或不同MHC II類分子)相關聯。於一些態樣中，一種疫苗組合物包含能與最頻繁發生之MHC I類分子及/或MHC II類分子相關聯之肽及/或多肽之編碼序列。因此，疫苗組合物可包含能與至少2個較佳、至少3個較佳或至少4個較佳MHC I類分子及/或MHC II類分子相關聯之不同片段。

【0175】 疫苗組合物可能引起特異性細胞毒性T細胞反應及/或特異性輔助T細胞反應。

【0176】 疫苗組合物可另外包含佐劑及/或載體。下文中給定可用佐劑及載體之實例。組合物可與諸如(例如)蛋白質或抗原呈遞細胞(諸如例如，能呈遞肽至T細胞之樹突狀細胞(DC))之載體相關聯。

【0177】 佐劑為進入疫苗組合物之其摻合物增加或以其他方式修改對新抗原之免疫反應的任何物質。載體可為新抗原能與其相關聯的支架結構(例如，多肽或多醣)。視情況，共價或非共價偶聯佐劑。

【0178】 佐劑增加對抗原之免疫反應之能力通常藉由免疫介導反應之顯著或實質增加或疾病症狀之減少而表現。例如，體液免疫之增加通常藉由提高至抗原之抗體滴度之顯著增加而表現，及通常於增加之細胞增生或細胞毒性或細胞激素分泌中顯示T細胞活性之增加。佐劑亦可(例如)藉由將主要體液或Th反應改變為主要細胞或Th反應來改變免疫反應。

【0179】 適宜佐劑包括(但不限於) 1018 ISS、明礬、鋁鹽、Amplivax、AS15、BCG、CP-870,893、CpG7909、CyaA、dSLIM、

GM-CSF、IC30、IC31、咪喹莫特(Imiquimod)、ImuFact IMP321、IS Patch、ISS、ISCOMATRIX、JuvImmune、LipoVac、MF59、單磷醯脂質 A、Montanide IMS 1312、Montanide ISA 206、Montanide ISA 50V、Montanide ISA-51、OK-432、OM-174、OM-197-MP-EC、ONTAK、PepTel載體系統、PLG微粒子、瑞喹莫德(resiquimod)、SRL172、病毒小體及其他類病毒粒子、YF-17D、VEGF拮、R848、 β -葡聚糖、Pam3Cys、衍生自皂角苷之Aquila's QS21促病毒素(Aquila Biotech, Worcester, Mass., USA)、分枝桿菌(mycobacterial)提取物及合成細菌細胞壁類似物及其他專有佐劑(諸如Ribi's Detox. Quil或Superfos)。諸如不完全弗洛因德氏(Freund's)或GM-CSF之佐劑係有用。先前已描述特定用於樹突狀細胞及其製備之若干免疫學佐劑(例如，MF59) (Dupuis M等人，Cell Immunol. 1998; 186(1):18-27；Allison A C; Dev Biol Stand. 1998; 92:3-11)。亦可使用細胞激素。已直接連接若干細胞激素以影響樹突狀細胞遷移至淋巴組織(例如，TNF- α)，加速樹突狀細胞成熟為T淋巴細胞之有效抗原呈遞細胞(例如，GM-CSF、IL-1及IL-4) (美國專利案第5,849,589號，其全文明確地以引用的方式併入本文中)及充當免疫佐劑(例如，IL-12) (Gabrilovich D I等人，J Immunother Emphasis Tumor Immunol. 1996 (6):414-418)。

【0180】 亦已報導CpG免疫刺激寡核苷酸以增強佐劑於疫苗環境中之作用。亦可使用其他TLR結合分子(諸如RNA結合TLR 7、TLR 8及/或TLR 9)。

【0181】 可用佐劑之其他實例包括(但不限於)經化學修飾之CpGs(例如，CpR、Idera)、Poly(I:C)(例如，polyi:CI2U)、非CpG細菌DNA或

RNA以及免疫活性小分子及抗體(諸如環磷醯胺、舒尼替尼(sunitinib)、貝伐單抗(bevacizumab)、西樂葆(celebrex)、NCX-4016、西地那非(sildenafil)、他達拉非(tadalafil)、伐地那非(vardenafil)、索拉非尼(sorafenib)、XL-999、CP-547632、帕唑帕尼(pazopanib)、ZD2171、AZD2171、易普利單抗(ipilimumab)、曲美目單抗(tremelimumab)及SC58175)，其可治療上起作用及/或作為佐劑。可藉由熟習技工在無不當實驗之情況下容易地確定佐劑及添加劑之量及濃度。額外佐劑包括群落刺激因子，諸如顆粒球巨噬細胞群落刺激因子(Granulocyte Macrophage Colony Stimulating Factor) (GM-CSF，沙莫司亭(sargramostim))。

【0182】 疫苗組合物可包含一種以上不同佐劑。此外，治療組合物可包含任何佐劑物質(包括以上或其組合中之任一者)。亦期望可一起或以適宜順序分開投與疫苗及佐劑。

【0183】 載體(或賦形劑)可獨立於佐劑存在。載體之功能可(例如)為增加特定言之突變體之分子量以增加活性或免疫原性，賦予穩定性，增加生物活性或增加血清半衰期。此外，載體可幫助呈遞肽至T細胞。載體可為熟習此項技術者已知之任何適宜載體，例如，蛋白質或抗原呈遞細胞。載體蛋白可為(但不限於)鑰孔血藍素、血清蛋白(諸如鐵傳遞蛋白、牛血清白蛋白、人血清白蛋白、甲狀腺球蛋白或卵白蛋白)、免疫球蛋白或激素(諸如胰島素或棕櫚酸)。針對於人類之免疫，載體一般為對人類及安全可接受之生理上可接受之載體。然而，破傷風類毒素及/或白喉類毒素為適宜載體。或者，載體可為右旋糖苷，例如，瓊脂糖。

【0184】 細胞毒性T-細胞(CTL)識別結合至MHC分子之肽之形式之抗原而非完整外來抗原本身。MHC分子本身位於抗原呈遞細胞之細胞表

面。因此，若存在肽抗原、MHC分子及APC之三聚複合體，則CTL之激活係可能。相應地，若不僅使用肽激活CTL，而且若額外添加具有各自MHC分子之APC，則其可增強免疫反應。因此，於一些實施例中，疫苗組合物額外含有至少一個抗原呈遞細胞。

【0185】新抗原亦可包含於基於病毒載體之疫苗平臺，諸如牛痘、雞痘、自我複製假病毒、馬拉病毒(marabavirus)、腺病毒(參見，例如，Tatsis等人，Adenoviruses, Molecular Therapy (2004) 10, 616-629)或慢病毒(包括但不限於第二代、第三代或雜交第二代/第三代慢病毒及經設計以靶向特異性細胞類型或受體之任何代之重組慢病毒)(參見，例如，Hu等人，Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, Immunol Rev. (2011) 239(1): 45-61，Sakuma等人，Lentiviral vectors: basic to translational, Biochem J. (2012) 443(3):603-18，Cooper等人，Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter, Nucl. Acids Res. (2015) 43 (1): 682-690，Zufferey等人，Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, J. Virol. (1998) 72 (12): 9873-9880)。取決於上述基於病毒載體之疫苗平臺之包裝能力，此方法可遞送編碼一或多個新抗原肽之一或多個核苷酸序列。該等序列可側接非突變序列，可藉由連接子分離或可在前面有靶向亞細胞隔室之一或多個序列(參見，例如，Gros等人，Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, Nat Med. (2016) 22 (4):433-8，Stronen等人，Targeting of cancer neoantigens with donor-derived T-cell

receptor repertoires, *Science*. (2016) 352 (6291):1337-41, Lu 等人, Efficient identification of mutated cancer antigens recognized by T-cells associated with durable tumor regressions, *Clin Cancer Res*. (2014) 20(13):3401-10)。引入宿主後，感染細胞表現新抗原，及從而引起對肽之宿主免疫(例如，CTL)反應。可用於免疫協定中之疫苗載體及方法述於(例如)美國專利案第4,722,848號中。另一載體為BCG (卡介苗)。BCG載體述於Stover等人(*Nature* 351:456-460 (1991))中。可用於新抗原之治療投與或免疫之各種其他疫苗載體(例如，傷寒沙門氏桿菌(*Salmonella typhi*)載體及類似者)將自本文中描述對熟習此項技術者顯而易見。

IV.A.疫苗設計及製造之額外考慮

IV.A.1.測定覆蓋所有腫瘤亞株(subclone)之肽集

【0186】 軀幹肽，意指藉由所有或大多數腫瘤亞株呈遞之彼等，將優先納入疫苗中。⁵³視情況，若不存在經預測具有高概率呈遞及免疫原性之軀幹肽，或若經預測具有高概率呈遞及免疫原性之軀幹肽之數目係足夠小使得額外非軀幹肽可包含於疫苗中，則另外肽可優先考慮評估腫瘤亞株之數目及同一性並且選擇肽以便將由疫苗覆蓋之腫瘤亞株之數目最大化。⁵⁴

IV.A.2.新抗原優先排序

【0187】 於應用所有以上新抗原篩檢程序後，更多候選新抗原仍可用於疫苗納入，而非疫苗技術可支持。此外，關於新抗原分析之各態樣之不確定可保留並且候選疫苗新抗原之不同性質之間之權衡可存在。因此，代替選擇過程之各步驟處之預定篩檢程序，可考慮整合之多維模型，該模型將候選新抗原置於具有至少下列軸之空間中及使用積分方法最佳化選

擇。

- 1.自體免疫或耐受之風險(生殖系之風險)(較低自體免疫之風險通常係較佳)
- 2.定序偽影之概率(較低偽影之概率通常係較佳)
- 3.免疫原性之概率(更高免疫原性之概率通常係較佳)
- 4.呈遞概率(更高呈遞概率通常係較佳)
- 5.基因表現(更高表現通常係較佳)
6. HLA基因之覆蓋度(涉及新抗原集之呈遞之更大數目之HLA分子可降低腫瘤經由HLA分子之下調或突變逃避免疫攻擊的概率)
7. HLA種類之覆蓋度(覆蓋HLA-I及HLA-II二者可增加治療反應之概率及降低腫瘤逃避之概率)

V · 治療及製造方法

【0188】 亦提供一種藉由對受試者投與一或多種新抗原(諸如使用本文中所示方法識別之複數個新抗原)誘導該受試者之腫瘤特異性免疫反應、抗腫瘤疫苗接種、治療或減輕該受試者之癌症症狀之方法。

【0189】 於一些態樣中，受試者已經診斷為患有癌症或由發展癌症之風險。受試者可為人類、狗、貓、馬或任何動物，其中腫瘤特異性免疫反應係所需。腫瘤可為諸如乳、卵巢、前列腺、肺、腎、胃、結腸、睪丸、頭頸、胰、腦之任何實體腫瘤、黑色素瘤及其他組織器官腫瘤及血液腫瘤(諸如淋巴瘤及白血病，包括急性骨髓性白血病、慢性骨髓性白血病、慢性淋巴球性白血病、T細胞淋巴球性白血病及B細胞淋巴瘤)。

【0190】 可以足以誘導CTL反應之量投與新抗原。

【0191】 可單獨或與其他治療劑組合投與新抗原。治療劑為(例如)

化療劑、放射或免疫療法。可投與針對特定癌症之任何適宜療法治療。

【0192】 此外，可對受試者另外投與抗免疫抑制劑/免疫刺激劑(諸如檢查點抑制劑)。例如，可對受試者另外投與抗CTLA抗體或抗PD-1或抗PD-L1。CTLA-4或PD-L1藉由抗體之阻斷可增強對患者之癌細胞之免疫反應。特定言之，當按照疫苗接種協定時，已顯示CTLA-4阻斷有效。

【0193】 可測定待包含於疫苗組合物中之各新抗原之最佳量及最佳給藥方案。例如，可製備新抗原或其變異體用於靜脈內(i.v.)注射、皮下(s.c.)注射、皮內(i.d.)注射、腹膜內(i.p.)注射、肌肉內(i.m.)注射。注射方法包括s.c.、i.d.、i.p.、i.m.及i.v.。DNA或RNA注射之方法包括i.d.、i.m.、s.c.、i.p.及i.v.。投與疫苗組合物之其他方法為熟習此項技術者已知。

【0194】 可編譯疫苗使得存在於組合物中之新抗原之選擇、數目及/或量係組織、癌症及/或患者特異性。例如，肽之精確選擇可藉由給定組織中之親本蛋白之表現模式指導。該選擇可取決於癌症之特定類型、疾病之狀態、早期治療方案及當然，患者之HLA-單倍型。此外，根據特定患者之個人需要，疫苗可含有個別組分。實例包括根據特定患者中之新抗原之表現改變新抗原之選擇或於第一輪治療或治療方案後調整二次治療。

【0195】 針對作為癌症疫苗使用之組合物，於正常組織中大量表現之具有類似正常自身肽之新抗原可避免或於本文中所述組合物中少量存在。另一方面，若已知患者之腫瘤表現大量某個新抗原，則用於治療此癌症之各自醫藥組合物可大量存在及/或可包含特異性針對此特殊新抗原或此新抗原路徑之一種以上新抗原。

【0196】 可對已患癌症之個體投與包含新抗原之組合物。於治療應

用中，以足以引起對腫瘤抗原之有效CTL反應及治癒或至少部分停止症狀及/或併發症之量對患者投與組合物。將足以實現此之量定義為「治療上有效劑量」。用於此用途之有效量將取決於(例如)組合物、投與方式、正治療之疾病之階段及嚴重度、患者之重量及健康一般狀態及處方醫師之判斷。應記住一般可於嚴重疾病狀態中(即，威脅生命或潛在威脅生命之情況，尤其當癌症已轉移時)採用組合物。於此等情況下，鑑於最小化外來物質及新抗原之相對無毒性質，藉由治療醫師投與實質上過量之此等組合物係可能且可認為可取。

【0197】 針對治療用途，投與可在檢測或手術移除腫瘤時開始。此後加強劑量直至至少症狀實質上減輕且此後持續一段時間。

【0198】 用於療法治療之醫藥組合物(例如，疫苗組合物)意欲用於非經腸、外用、鼻、口或局部投與。可非經腸(例如，經靜脈內、經皮下、經皮內或經肌內)投與醫藥組合物。可在誘導對腫瘤之局部免疫反應之手術切除的位點處投與組合物。本文中揭示包含新抗原溶液之用於非經腸投與之組合物及將疫苗組合物溶解或懸浮於可接受之載劑(例如，水性載劑)中。可使用各種水性載劑，例如，水、緩衝水、0.9%鹽水、0.3%甘胺酸、透明質酸及類似者。此等組合物可藉由習知、熟知滅菌技術滅菌或可經無菌過濾。可將所得水溶液包裝用作在投與之前與無菌溶液組合之凍乾製劑或經凍乾。該等組合物可含有如接近生理條件所需之醫藥上可接受之輔助物質，諸如pH調節劑及緩衝劑、張力調節劑、潤濕劑及類似者，例如，乙酸鈉、乳酸鈉、氯化鈉、氯化鉀、氯化鈣、脫水山梨糖醇單月桂酸酯、三乙醇胺油酸酯等。

【0199】 新抗原亦可經由脂質體投與，該等脂質體靶向其至特定細

胞組織，諸如淋巴組織。脂質體亦可用於增加半衰期。脂質體包括乳液、泡沫、膠囊、不溶性單層、液晶、磷脂分散體、片狀層及類似者。於此等製劑中，單獨或與分子結合併入待遞送之新抗原作為脂質體之部分，該分子結合至(例如)淋巴樣細胞中普遍存在之受體(諸如結合至CD45抗原之單株抗體)或與其他治療或免疫原組合物結合。因此，填充有所需新抗原之脂質體可指向淋巴樣細胞之位點，然後脂質體在該處遞送所選治療/免疫原組合物。脂質體可自標準囊泡形成脂質而形成，該等囊泡形成脂質一般包括中性及負電荷磷脂及固醇(諸如膽固醇)。脂質之選擇一般藉由以下之考慮指導：例如，脂質體大小、酸不穩定性及血流中之脂質體之穩定性。各種方法可用於製備脂質體，如於(例如) Szoka 等人，Ann. Rev. Biophys. Bioeng. 9; 467 (1980)，美國專利案第4,235,871號、第4,501,728號、第4,501,728號、第4,837,028號及第5,019,369號中所述。

【0200】 用於靶向免疫細胞，待併入脂質體之配位體可包括(例如)特定用於所需免疫系統細胞之細胞表面決定因子之抗體或其片段。脂質體懸浮液可以劑量經靜脈內、局部、外用等投與，該劑量根據尤其投與方式、正遞送之肽及所治療之疾病之階段變化。

【0201】 出於治療或免疫目的，編碼肽之核酸及視情況可選的本文中所述肽中之一或多者亦可對患者投與。許多方法方便地用於遞送核酸給患者。例如，核酸可作為「裸DNA」直接遞送。此方法述於(例如) Wolff 等人，Science 247: 1465-1468 (1990)以及美國專利案第5,580,859號及第5,589,466號。亦可使用如(例如)於美國專利案第5,204,253號中所述之彈道遞送投與核酸。可投與僅包含DNA之粒子。或者，DNA可黏附至粒子(諸如金粒子)。用於遞送核酸序列之方法可包括利用電穿孔或不利用電穿

孔之病毒載體、mRNA載體及DNA載體。

【0202】亦可遞送與陽離子化合物(諸如陽離子脂質)錯合之核酸。脂質介導之基因遞送方法述於(例如) 9618372WOAWO 96/18372、9324640WOAWO 93/24640、Mannino及Gould-Fogerite, BioTechniques 6(7): 682-691 (1988)、美國專利案第5,279,833號、Rose美國專利案第5,279,833號、9106309WOAWO 91/06309及Felgner等人, Proc. Natl. Acad. Sci. USA 84: 7413-7414 (1987)中。

【0203】新抗原亦可包含於基於病毒載體之疫苗平臺, 諸如牛痘、鳥痘、自我複製 α 病毒、馬拉病毒(marabavirus)、腺病毒(參見, 例如, Tatsis等人, Adenoviruses, Molecular Therapy (2004) 10, 616-629)或慢病毒, 包括(但不限於)第二代、第三代或雜交第二代/第三代慢病毒及經設計以靶向特異性細胞類型或受體之任何代之重組慢病毒(參見, 例如, Hu等人, Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, Immunol Rev. (2011) 239(1): 45-61, Sakuma等人, Lentiviral vectors: basic to translational, Biochem J. (2012) 443(3):603-18, Cooper等人, Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter, Nucl. Acids Res. (2015) 43 (1): 682-690, Zufferey等人, Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, J. Virol. (1998) 72 (12): 9873-9880)。取決於上述基於病毒載體之疫苗平臺之包裝能力, 此方法可遞送編碼一或多個新抗原肽之一或多個核苷酸序列。該等序列可側接非突變序列, 可藉由連接子分離或可在前面有靶向亞細胞隔室之一或多個序列(參見, 例如, Gros等人,

Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, *Nat Med.* (2016) 22 (4):433-8, Stronen等人, Targeting of cancer neoantigens with donor-derived T-cell receptor repertoires, *Science.* (2016) 352 (6291):1337-41, Lu等人, Efficient identification of mutated cancer antigens recognized by T-cells associated with durable tumor regressions, *Clin Cancer Res.* (2014) 20(13):3401-10)。引入宿主之後，經感染細胞表現新抗原及從而引起對肽之宿主免疫(例如，CTL)反應。可用於免疫協定之疫苗載體及方法述於(例如)美國專利案第4,722,848號中。另一載體為BCG (卡介苗)。BCG載體述於Stover等人(*Nature* 351:456-460 (1991))中。可用於新抗原之治療投與或免疫之各種其他疫苗載體(例如，傷寒沙門氏菌(*Salmonella typhi*)載體及類似者)將自本文中描述對熟習此項技術者顯而易見。

【0204】一種投與核酸之方式使用編碼一或多個抗原決定基之微基因構築體。為創建用於人類細胞中表現之編碼所選CTL抗原決定基(微基因)之DNA序列，該等抗原決定基之胺基酸序列經逆轉譯。使用人類密碼子使用表指導各胺基酸之密碼子選擇。此等編碼抗原決定基之DNA序列經直接連接，創建連續多肽序列。為最佳化表現及/或免疫原性，可將額外元件並入微基因設計。可經逆轉譯且包含於微基因序列中之胺基酸序列之實例包括：輔助T淋巴細胞、抗原決定基、先導(信號)序列及內質網滯留信號。此外，CTL抗原決定基之MHC呈遞可藉由包含鄰近CTL抗原決定基之合成(例如，聚丙胺酸)或天然產生之側接序列改善。藉由組裝編碼微基因之正及負股之寡核苷酸將微基因序列轉化成DNA。使用熟知技術在適宜條件下合成重疊寡核苷酸(30至100個鹼基長)，將其磷酸化、純化

及黏著。使用T4 DNA連接酶將寡核苷酸之末端連接。然後可將編碼CTL抗原決定基多肽之此合成微基因純化成所需表現載體。

【0205】 可使用各種調配物製備純化質體DNA用於注射。此等最簡單為凍乾DNA於無菌磷酸鹽緩衝鹽水(PBS)中之重新構建。已描述各種方法，且新技術可變得可得。如上所指出，利用陽離子脂質方便地調配核酸。此外，亦可將統稱為保護性、交互作用、非縮合(PINC)之糖脂、融合脂質體、肽及化合物錯合成純化質體DNA以影響變數(諸如穩定性、肌肉內分散或運輸至特定器官或細胞類型)。

【0206】 亦揭示一種製造腫瘤疫苗之方法，其包括進行本文中所揭示之方法之步驟；及產生包含複數個新抗原或複數個新抗原之子集之腫瘤疫苗。

【0207】 可使用此項技術中已知之方法製造本文中所揭示之新抗原。例如，產生本文中所揭示之新抗原或載體(例如，包含編碼一或多個新抗原之至少一個序列之載體)之方法可包括在適於表現新抗原或載體之條件下培養宿主細胞，其中該宿主細胞包含編碼新抗原或載體之至少一個多核苷酸，及將該新抗原或載體純化。標準純化方法包括層析技術、電泳、免疫學、沉澱、透析、過濾、濃縮及層析聚焦技術。

【0208】 宿主細胞可包括中國倉鼠卵巢(Chinese Hamster Ovary, CHO)細胞、NS0細胞、酵母或HEK293細胞。可將宿主細胞用包含編碼本文中所揭示之新抗原或載體之至少一個核酸序列之一或多個多核苷酸轉換，視情況其中經單離核苷酸另外包含以操作方式連接至編碼新抗原或載體之至少一個核酸序列之啟動子序列。於某些實施例中，該經單離多核苷酸可為cDNA。

VI・新抗原識別

VI.A.新抗原候選識別

【0209】已描述及於新抗原識別空間中應用腫瘤及正常外顯子及轉錄體之NGS分析之研究方法。^{6,14,15}以下實例考慮針對臨床環境中之新抗原識別之更大敏感性及特異性的某些最佳化。可將此等最佳化分成兩個區域，與實驗室製程有關之彼等及與NGS資料分析有關之彼等。

VI.A.1.實驗室製程最佳化

【0210】本文中所呈現之製程改進藉由擴展開發用於靶向癌症小組¹⁶之可靠癌症驅動基因評估至新抗原識別所需之全外顯子及轉錄體環境中之觀念來解決來自具有低腫瘤含量及小體積之臨床樣品之高精度新抗原發現的挑戰。具體而言，此等改進包括：

1.跨腫瘤外顯子組之靶向深度(>500x)獨特平均覆蓋度以檢測由於低腫瘤含量或次選殖狀態之低突變體對偶基因頻率下呈遞之突變。

2.跨腫瘤外顯子組之靶向均勻覆蓋度，在<100x下覆蓋<5%之鹼基，使得最少可能新抗原藉由(例如)以下錯失：

a.採用利用個別探針QC¹⁷之基於DNA之捕獲探針

b.包含低覆蓋區域之額外誘餌

3.跨正常外顯子組之靶向均勻覆蓋度，在<20x下覆蓋<5%之鹼基，使得最少新抗原可對體細胞/生殖系狀態仍未分類(及因此不可用作TSNA)

4.為最小化所需定序之總量，將設計序列捕獲探針僅用於基因之編碼區，因為非編碼RNA不可產生新抗原。額外最佳化包括：

a. HLA基因之增補探針，其係富含GC且藉由標準外顯子組定序差捕獲¹⁸

b.預測產生少量或無候選新抗原之基因之排除，由於諸如不充分表現、藉由蛋白酶體次最佳消化或異常序列特徵之因素。

5.腫瘤RNA同樣將在高深度(>100M讀數)下定序以使能變異體檢測、基因及剪接變異體(「同功異型物」)表現之定量及融合檢測。將使用基於探針之濃化¹⁹，利用用於捕獲DNA中之外顯子組之相同或相似探針提取來自FFPE樣品之RNA。

VI.A.2. NGS資料分析最佳化

【0211】 分析方法之改進解決常見研究突變調用方法之次最佳敏感性、及特異性，且特別考慮臨床環境中之新抗原識別相關之定製。此等包括：

1.使用HG38參考人類基因組或後期版本用於比對，因為其含有與先前基因組釋放相比更好反映群體多態性之多個MHC區域組件。

2.藉由合併不同程式⁵之結果克服單變異體調用者²⁰之限制

a.單核苷酸變異體及插入或缺失將利用一套工具自腫瘤DNA、腫瘤RNA及正常DNA檢測，該等工具包括：基於腫瘤及正常DNA之比較之程式(諸如Strelka²¹及Mutect²²)；及併入腫瘤DNA、腫瘤RNA及正常DNA之程式(諸如UNCeqR)，其於低純度樣品中係特別有利²³。

b.插入或缺失將利用進行局部再組裝之程式(諸如Strelka及ABRA²⁴)檢測。

c.結構重排將使用專用工具(諸如Pindel²⁵或Breakseq²⁶)檢測。

3.為檢測及預防樣品交換，將在選定數目之多態位點下比較來自相同患者之樣品之變異體調用。

4.將(例如)藉由以下進行人工調用之擴展過濾：

a.移除於正常DNA中發現之變異體，可在低覆蓋之情況下利用鬆弛檢測參數及在插入或缺失之情況下利用允許鄰近準則

b.移除由於低圖譜分析品質或低鹼基品質²⁷之變異體。

c.移除源自再現定序偽影之變異體，即使於對應正常²⁷中未觀察到。實例包括主要在一股上檢測到之變異體。

d.移除於不相關對照組²⁷中檢測到之變異體

5.來自正常外顯子組之精確HLA調用，其使用seq2HLA²⁸、ATHLATES²⁹或Optitype中之一者及亦將外顯子組與RNA定序資料²⁸組合。額外潛在最佳化包括採用用於HLA分型之專用檢定(諸如長閱讀DNA定序³⁰)或連接RNA片段以保留連續性之方法之調適³¹。

6.產生自腫瘤特異性剪接變異體之neo-ORF之穩健檢測將藉由組裝來自RNA-seq資料之轉錄本使用CLASS³²、Bayesembler³³、StringTie³⁴或其參考指導模式中之類似程式(即，使用除了試圖自各實驗再建其完整轉錄本外之已知轉錄本結構)進行。出於此目的，經常使用While Cufflinks³⁵，其頻繁產生難以置信之大量剪接變異體，其中之許多遠短於全長基因，且可無法回收簡單陽性對照。編碼序列及無義介導之衰變位勢將利用工具(諸如SpliceR³⁶及MAMBA³⁷)，利用再引入之突變體序列檢測。基因表現將利用工具(諸如Cufflinks³⁵或Express (Roberts及Pachter，2013))檢測。野生型及突變特異性表現計數及/或相對水平將利用出於此等目的開發之工具(諸如ASE³⁸或HTSeq³⁹)檢測。位勢過濾步驟包括：

a.移除視為不充分表現之候選neo-ORF。

b.移除預測觸發無義介導之衰變(NMD)之候選neo-ORF。

7.不能被直接證實為腫瘤特異性之僅於RNA中觀察到之候選新抗原(例如，neoORF)將根據額外參數，例如，藉由考慮以下被歸類為可能腫瘤特異性：

- a.支持僅腫瘤DNA順式作用框移或剪接位點突變之存在
- b.證實剪接因子中之僅腫瘤DNA反式作用突變之存在。例如，於利用R625-突變體SF3B1之三個獨立公開實驗中，展示最大差異剪接之基因係一致的，即使一個實驗檢查葡萄膜黑色素瘤患者⁴⁰，第二個檢查葡萄膜黑色素瘤細胞株⁴¹及第三個檢查乳癌患者⁴²。
- c.針對新穎剪接同功異型物，證實RNASeq資料中之「新穎」剪接接合閱讀之存在。
- d.針對新穎重排，證實腫瘤DNA中之近端外顯子之存在，其不存在於正常DNA中
- e.不存在基因表現簡編，諸如GTEX⁴³ (即，使生殖系起源可能性減少)

8.藉由直接比較經組裝之DNA腫瘤與正常閱讀(或來自此等閱讀之k-mer數)補充基於參考基因組比對之分析以避免基於比對及註解之誤差及人為誤差。(例如，針對生殖系變異體或重複背景插入或缺失附近產生之體細胞變異體)

【0212】 於具有聚腺苷化RNA之樣品中，將使用RNA CoMPASS⁴⁴或類似方法評估RNA-seq資料中之病毒及微生物RNA之存在，該等方法朝向可預測患者反應之額外因素之識別。

VI.B. HLA肽之單離及檢測

【0213】 於組織樣品之裂解及增溶後使用經典免疫沉澱(IP)方法進

行HLA-肽分子之單離⁵⁵⁻⁵⁸。澄清裂解物係用於HLA特異性IP。

【0214】 使用偶合至珠之抗體進行免疫沉澱，其中該抗體特異性用於HLA分子。針對泛I類HLA免疫沉澱，使用泛I類CR抗體，針對II類HLA - DR，使用HLA-DR抗體。抗體在過夜培育期間共價連接至NHS-瓊脂糖珠。於共價連接後，將該等珠洗滌及取等分試樣用於IP。^{59, 60}亦可利用不共價連接至珠之抗體進行免疫沉澱。通常使用塗覆有蛋白A及/或蛋白G之瓊脂糖或磁珠進行此以將抗體保持在管柱上。可用於選擇性濃化MHC/肽複合體之一些抗體列於以下。

抗體名稱	特異性
W6/32	I類HLA-A、B、C
L243	II類- HLA-DR
Tu36	II類- HLA-DR
LN3	II類- HLA-DR
Tu39	II類- HLA-DR、DP、DQ

【0215】 將澄清組織裂解物添加至抗體珠中用於免疫沉澱。於免疫沉澱後，將珠自裂解物移除及將裂解物儲存用於額外實驗，包括額外IP。將該等IP珠洗滌以移除非特異性結合及使用標準技術將HLA/肽複合體自珠溶離。使用分子重量自旋管柱或C18分餾將蛋白質組分自肽移除。將所得肽藉由SpeedVac蒸發至乾及於一些情況下，在MS分析之前在-20℃下儲存。

【0216】 將經乾燥肽於適用於逆相層析之HPLC緩衝液中再構成並負載於C-18微毛細管HPLC管柱上用於聚變螢光金屬氧化物半導體質譜儀(Fusion Lumos mass spectrometer，Thermo)中之梯度溶離。於Orbitrap檢測器中在高解析度下收集肽質量/電荷(m/z)之MS1光譜，接著於選定離子之HCD片段化後於離子阱檢測器中收集MS2低解析度掃描。此外，可

使用CID或ETD片段化方法或三種技術之任何組合獲得MS2光譜以達成肽之更大胺基酸覆蓋。亦可利用Orbitrap檢測器之高解析度質量精度量測MS2光譜。

【0217】 針對蛋白質資料庫使用Comet^{61, 62}搜索來自各分析之MS2光譜及使用Percolator⁶³⁻⁶⁵對肽識別評分。使用PEAKS工作室(Bioinformatics Solutions Inc.)進行額外定序及可使用其他搜尋引擎或定序方法，包括光譜匹配及重新定序⁷⁵。

VI.B.1.支持全面HLA肽定序之MS檢測極限研究

【0218】 使用肽YVYVADVAAK，使用負載於LC管柱上之不同量之肽確定檢測極限。所測試之肽之量為1 pmol、100 fmol、10 fmol、1 fmol及100 amol。(表1)結果示於圖1F中。此等結果指示最低檢測極限(LoD)係於阿托莫耳(attomol)範圍(10^{-18})內，動態範圍跨越五個數量級，及訊雜比似乎足以在毫微微莫耳(femtomol)範圍(10^{-15})下定序。

肽m/z	負載在管柱上	於1e9個細胞中之複本/細胞
566.830	1 pmol	600
562.823	100 fmol	60
559.816	10 fmol	6
556.810	1 fmol	0.6
553.802	100 amol	0.06

VII · 呈遞模型

VII.A.系統概觀

【0219】 圖2A為根據一實施例之用於識別患者中之肽呈遞可能性之環境100之概觀。環境100提供環境以引入呈遞識別系統160，其自身包含

呈遞資訊存儲165。

【0220】 呈遞識別系統160為於如下關於圖29所討論之電腦系統中所實施之一或多種電腦模型，該電腦系統接收與MHC對偶基因集相關之肽序列並確定肽序列將藉由一組相關聯MHC對偶基因中之一或多者呈遞之可能性。該呈遞識別系統160可應用於I類及II類MHC對偶基因二者。此可用於各種環境中。用於呈遞識別系統160之一種特定使用情況為能接受與來自患者110之腫瘤細胞之MHC對偶基因集相關之候選新抗原的核苷酸序列並確定該等候選新抗原將藉由腫瘤之相關聯MHC對偶基因中之一或多者呈遞及/或誘導患者110之免疫系統中之免疫原反應的可能性。可選擇如藉由系統160所確定之具有高可能性之彼等候選新抗原用於納入疫苗118，可自提供腫瘤細胞之患者110之免疫系統引出此抗腫瘤免疫反應。此外，可產生具有對具有高呈遞可能性之候選新抗原反應之TCR之T細胞用於T細胞療法中，從而亦引出來自患者110之免疫系統之抗腫瘤免疫反應。

【0221】 呈遞識別系統160通過一或多個呈遞模型確定呈遞可能性。具體而言，該等呈遞模型產生給定肽序列是否將針對相關聯MHC對偶基因集呈遞及基於存儲165中存儲之呈遞資訊產生之可能性。例如，該等呈遞模型可產生肽序列「YVYVADVAAK」是否將針對樣品之細胞表面上之對偶基因集HLA-A*02:01、HLA-A*03:01、HLA-B*07:02、HLA-B*08:03、HLA-C*01:04呈遞之可能性。呈遞資訊165含有關於肽是否結合至不同類型之MHC對偶基因使得彼等肽藉由MHC對偶基因呈遞之資訊，其於該等模型中取決於肽序列中之胺基酸之位置確定。該呈遞模型可預測未經識別之肽序列是否將與基於呈遞資訊165之相關MHC對偶基因集相關呈遞。如先前所述，該等呈遞模型可應用於I類及II類MHC對偶

基因二者。

VII.B.呈遞資訊

【0222】 圖2說明根據一實施例之獲得呈遞資訊之方法。該呈遞資訊165包含兩種一般資訊類別：對偶基因相互作用資訊及對偶基因非相互作用資訊。對偶基因相互作用資訊包括影響取決於MHC對偶基因類型之肽序列呈遞之資訊。對偶基因非相互作用資訊包括影響不取決於MHC對偶基因類型之肽序列呈遞之資訊。

VII.B.1.對偶基因相互作用資訊

【0223】 對偶基因相互作用資訊主要包括經識別之肽序列，已知該等肽序列藉由來自人類、小鼠等之一或多個經識別之MHC分子呈遞。應注意，此可包括或可不包括獲自腫瘤樣品之資料。可自表現單MHC對偶基因之細胞識別該等呈遞肽序列。於此情況下，一般自經基因改造以表現預定MHC對偶基因及隨後暴露於合成蛋白之單對偶基因細胞株收集該等呈遞肽序列。在MHC對偶基因上呈遞之肽藉由諸如酸溶離之技術單離及通過質譜法識別。圖2B顯示此之實例，其中在預定MHC對偶基因HLA-DRB1*12:01上呈遞之實例肽YEMFNDKSQRAPDDKMF經單離及通過質譜法識別。因為於此情況下，肽通過經基因改造以表現單一預定MHC蛋白之細胞識別，已確知在呈遞肽與結合其之MHC蛋白之間之直接關聯。

【0224】 該等呈遞肽序列亦可自表現多個MHC對偶基因之細胞收集。通常於人類中，細胞表現6種不同類型之MHC-I及至多12種不同類型之MHC-II分子。可自經基因改造以表現多個預定MHC對偶基因之多個對偶基因細胞株識別此等呈遞肽序列。亦可自來自正常組織樣品或腫瘤組織樣品之組織樣品識別此等呈遞肽序列。特定言之於此情況下，MHC分子

可自正常或腫瘤組織免疫沉澱。在多個MHC對偶基因上呈遞之肽可類似地藉由諸如酸溶離之技術單離及通過質譜法識別。圖2C顯示此之實例，其中六種實例肽，YEMFNDKSF、HROEIFSHDFJ、FJIEJFOESS、NEIOREIREI、JFKSIFEMMSJDSSUIFLKSJFIEIFJ及KNFLENFIESOFI在經識別之I類MHC對偶基因HLA-A*01:01、HLA-A*02:01、HLA-B*07:02、HLA-B*08:01及II類MHC對偶基因HLA-DRB1*10:01、HLA-DRB1:11:01上呈遞及經單離及通過質譜法識別。與單對偶基因細胞株相比，在呈遞肽與結合其之MHC蛋白之間之直接關聯可係未知，因為結合肽在經識別之前自MHC分子單離。

【0225】 對偶基因相互作用資訊亦可包括取決於肽-MHC分子複合體之濃度及肽之電離效率二者之質譜離子流。電離效率以序列相依方式自肽至肽變化。一般地，電離效率自肽至肽變化在約兩個數量級，同時肽-MHC複合體之濃度在較電離效率變化更大範圍變化。

【0226】 對偶基因相互作用資訊亦可包括在給定MHC對偶基因與給定肽之間之結合親和力之測量值或預測。(72, 73, 74)一或多個親和力模型可產生此等預測。例如，回到圖1D中所示之實例，呈遞資訊165可包括肽YEMFNDKSF與I類對偶基因HLA-A*01:01之間之1000 nM之結合親和力預測。具有 $IC_{50} > 1000$ nm之少數肽藉由MHC呈遞，及更低 IC_{50} 值增加呈遞之概率。呈遞資訊165可包括肽KNFLENFIESOFI與II類對偶基因HLA-DRB1:11:01之間之結合親和力預測。

【0227】 對偶基因相互作用資訊亦可包括MHC複合體之穩定性之測量值或預測。一或多個穩定性模型可產生此等預測。更穩定肽-MHC複合體(即，具有更長半衰期之複合體)更可能在腫瘤細胞上及在遇到疫苗抗原

之抗原呈遞細胞上在高複本數下呈遞。例如，回到圖2C中所示之實例，呈遞資訊165可包括針對I類分子HLA-A*01:01之1小時之半衰期之穩定性預測。呈遞資訊165亦可包括II類分子HLA-DRB1:11:01之半衰期之穩定性預測。

【0228】 對偶基因相互作用資訊亦可包括肽-MHC複合體之生成反應之量測或預測率。在更高速率下形成之複合體更可能在高濃度下在細胞表面呈遞。

【0229】 對偶基因相互作用資訊亦可包括序列及肽之長度。MHC I類分子通常較佳地呈遞具有8與15個肽之間之長度的肽。60至80%之呈遞肽具有長度9。MHC II類分子通常較佳地呈遞具有6與30個肽之間之長度的肽。

【0230】 對偶基因相互作用資訊亦可包括編碼肽之新抗原上之激酶序列模體之存在，及編碼肽之新抗原上之特定轉譯後修飾之不存在或存在。激酶模體之存在影響轉譯後修飾之概率，其可增強或干擾MHC結合。

【0231】 對偶基因相互作用資訊亦可包括參與轉譯後修飾過程之蛋白質之表現或活性水平，例如，激酶(如自RNA seq、質譜法或其他方法所量測或預測)。

【0232】 對偶基因相互作用資訊亦可包括來自表現特定MHC對偶基因之其他個體之細胞中之具有相似序列之肽的呈遞概率，如藉由質譜蛋白質組學或其他方法所評估。

【0233】 對偶基因相互作用資訊亦可包括所討論之個體中之特定MHC對偶基因的表現水平(例如，如藉由RNA-seq或質譜法所量測)。最強

結合至在高水平下表現之MHC對偶基因之肽較最強結合至在低水平下表現之MHC對偶基因之肽更可能呈遞。

【0234】對偶基因相互作用資訊亦可包括藉由表現特定MHC對偶基因之其他個體中之特定MHC對偶基因呈遞之總新抗原編碼的肽-序列-獨立概率。

【0235】對偶基因相互作用資訊亦可包括藉由其他個體中之相同家族分子(例如，HLA-A、HLA-B、HLA-C、HLA-DQ、HLA-DR、HLA-DP)中之MHC對偶基因呈遞之總肽-序列-獨立概率。例如，HLA-C分子通常較HLA-A或HLA-B分子在更低水平下表現，及因此，肽藉由HLA-C之呈遞較藉由HLA-A或HLA-B之呈遞事前更少可能。針對另一實例，HLA-DP通常較HLA-DR或HLA-DQ在更低水平下表現；因此，肽藉由HLA-DP之呈遞較藉由HLA-DR或HLA-DQ之呈遞事前更少可能。

【0236】對偶基因相互作用資訊亦可包括特定MHC對偶基因之蛋白質序列。

【0237】亦可將下節中所列之任何MHC對偶基因非相互作用資訊建模為MHC對偶基因相互作用資訊。

VII.B.2.對偶基因非相互作用資訊

【0238】對偶基因非相互作用資訊可包括其源蛋白序列內側接新抗原編碼之肽之C-端序列。針對MHC-I，C-端側接序列可影響肽之蛋白酶體處理。然而，C-端側接序列藉由蛋白酶體自肽裂解，然後將肽運送至內質網及遇到細胞表面上之MHC對偶基因。因此，MHC分子未接收關於C-端側接序列之資訊，及因此，C-端側接序列之作用不可取決於MHC對偶基因類型變化。例如，回到圖2C中所示之實例，呈遞資訊165可包括自肽

之源蛋白識別之呈遞肽 FJIEJFOESS 之 C-端側接序列 FOEIFNDKSLDKFJI。

【0239】對偶基因非相互作用資訊亦可包括mRNA定量測量值。例如，可針對提供質譜訓練資料之相同樣品獲得mRNA定量資料。如後提及圖13H所述，RNA表現經識別為肽呈遞之強預測因子。於一實施例中，自軟體工具RSEM識別mRNA定量測量值。RSEM軟體工具之詳細實施可見於Bo Li及Colin N. Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 12:323, 2011年8月。於一實施例中，以每百萬圖譜分析閱讀之轉錄本之每千鹼基對之片段數(FPKM)的單位量測mRNA定量。

【0240】對偶基因非相互作用資訊亦可包括其源蛋白序列內側接該肽之N-端序列。

【0241】對偶基因非相互作用資訊亦可包括肽序列之源基因。可將該源基因定義為肽序列之Ensembl蛋白家族。於其他實例中，可將該源基因定義為肽序列之源DNA或源RNA。例如，可將源基因表示為一串編碼蛋白質之核苷酸，或基於已知編碼特定蛋白質之已知DNA或RNA序列的命名集更分類表示。於另一實例中，對偶基因非相互作用資訊亦可包括源轉錄本或同功異型物或自資料庫(諸如Ensembl或RefSeq)提取之肽序列之潛在源轉錄本或同功異型物集。

【0242】對偶基因非相互作用資訊亦可包括肽序列來源之細胞之組織類型、細胞類型或腫瘤類型。

【0243】對偶基因非相互作用資訊亦可包括根據腫瘤細胞中之對應蛋白酶之表現視情況加權之肽中之蛋白酶裂解模體的存在(如藉由RNA-

seq或質譜法所量測)。含有蛋白酶裂解模體之肽較少可能呈遞，因為其將更容易藉由蛋白酶降解，及因此將於細胞內更不穩定。

【0244】對偶基因非相互作用資訊亦可包括源蛋白之周轉率，如於適宜細胞類型中所量測。較快周轉率(即，較低半衰期)增加呈遞概率；然而，若於不同細胞類型中量測，則此特徵之預測力係低的。

【0245】對偶基因非相互作用資訊亦可包括源蛋白之長度，視情況考慮於腫瘤細胞中最高度表現之特異性剪接變異體(「同功異型物」)，如藉由RNA-seq或蛋白質組質譜法所量測，或如自DNA或RNA序列資料中所檢測之生殖系或體細胞剪接突變之註解所預測。

【0246】對偶基因非相互作用資訊亦可包括腫瘤細胞中之蛋白酶體、免疫蛋白酶體、胸腺蛋白酶體或其他蛋白酶之表現水平(其可藉由RNA-seq、蛋白質組質譜法或免疫組織化學量測)。不同蛋白酶體具有不同裂解位點偏好。將對與其表現水平成比例之各類型之蛋白酶體之裂解偏好提供更多權重。

【0247】對偶基因非相互作用資訊亦可包括肽之源基因之表現(例如，如藉由RNA-seq或質譜法所量測)。可能最佳化包括調整經量測表現以說明腫瘤樣品內之基質細胞及腫瘤浸潤淋巴細胞之存在。更可能呈遞來自更高度表現基因之肽。來自具有無法檢測之表現水平之基因之肽可自考慮排除。

【0248】對偶基因非相互作用資訊亦可包括編碼肽之新抗原之源mRNA將遭受無義介導之衰變之概率，如藉由無義介導之衰變之模型(例如，來自Rivas等人，Science 2015之模型)所預測。

【0249】對偶基因非相互作用資訊亦可包括在細胞週期之各種階段

期間肽之源基因之典型組織特異性表現。在總體低水平(如藉由RNA-seq或質譜法蛋白質組學所量測)下表現但是已知在細胞週期之特定階段期間在高水平下表現之基因可較在極低水平下穩定表現之基因產生更多呈遞肽。

【0250】 對偶基因非相互作用資訊亦可包括如於(例如)uniProt或PDB <http://www.rcsb.org/pdb/home/home.do>中所給定之源蛋白之特徵之綜合目錄。此等特徵可尤其包括：蛋白質之二級及三級結構、亞細胞局域化11、基因本體論(GO)術語。具體而言，此資訊可含有在蛋白質水平下作用之註解(例如，5' UTR長度)及在特異性殘基之水平下作用之註解(例如，在殘基300與310之間之螺旋模體)。此等特徵亦可包括轉模體、褶板模體及無序殘基。

【0251】 對偶基因非相互作用資訊亦可包括描述含有肽之源蛋白之域之性質的特徵，例如，二級或三級結構(例如， α 螺旋相對於 β 褶板)；替代剪接。

【0252】 對偶基因非相互作用資訊亦可包括新抗原之肽序列與新抗原之源基因之複數個k-mer區塊中之一或多個k-mer區塊之間的關聯(如於受試者之核苷酸定序資料中存在)。在呈遞模型之訓練期間，將新抗原之肽序列與新抗原之核苷酸定序資料之k-mer區塊之間之此等關聯輸入至該模型中，及藉由該模型部分使用以學習模型參數，該等參數表示針對與訓練肽序列相關之k-mer區塊之呈遞熱點之存在或不存在。然後，在訓練後之模型使用期間，將測試肽序列與測試肽序列之源基因之一或多個k-mer區塊之間的關聯輸入至該模型中，及在訓練期間藉由該模型學習之參數使該呈遞模型能作出關於測試肽序列之呈遞可能性之更精確預測。

【0253】 總之，表示針對k-mer區塊之呈遞熱點之存在或不存在之模型之參數表示於控制所有其他變數(例如，肽序列、RNA表現、於HLA-結合肽中常見之胺基酸等)後，該k-mer區塊將產生呈遞肽之殘餘傾向。表示針對k-mer區塊之呈遞熱點之存在或不存在之參數可為二進位係數(例如，0或1)或沿標度之類似係數(例如，0與1之間，包含端值)。於任一情況下，較大係數(例如，更接近1或1)表示在控制其他因素下，k-mer區塊將產生呈遞肽之更大可能性，然而較低係數(例如，更接近0或0)表示k-mer區塊將產生呈遞肽之更低可能性。例如，具有低熱點係數之k-mer區塊可為來自具有高RNA表現之基因之k-mer區塊，其具有於HLA-結合肽中常見之胺基酸，其中源基因產生許多其他呈遞肽，但是呈遞肽於k-mer區塊中罕見。因為已可藉由其他參數說明肽存在之其他來源(例如，於HLA結合肽中常見之k-mer區塊或更大鹼基上之RNA表現)，此等熱點參數提供新的分離資訊，該資訊不對藉由其他參數捕獲之資訊「雙重計數」。

【0254】 對偶基因非相互作用資訊亦可包括肽自其他個體中之所討論之肽之源蛋白呈遞的概率(於調整彼等個體中之源蛋白之表現水平及彼等個體之不同HLA類型之影響後)。

【0255】 對偶基因非相互作用資訊亦可包括由於技術偏見肽將藉由質譜法未檢測到或過度表示之概率。

【0256】 如藉由基因表現檢定(諸如RNASeq、微陣列、靶向板(諸如Nanostring)或基因模組之單/多基因代表)所量測之各種基因模組/路徑之表現藉由諸如RT-PCR (其不需要含有肽之源蛋白)之檢定量測，該等檢定提供關於腫瘤細胞、基質或腫瘤浸潤淋巴細胞(TIL)之資訊。

【0257】 對偶基因非相互作用資訊亦可包括腫瘤細胞中之肽之源基因的複本數目。例如，可將來自遭受腫瘤細胞中之純合子缺失之基因的肽指定為零之呈遞概率。

【0258】 對偶基因非相互作用資訊亦可包括肽結合至TAP之概率或肽與TAP之量測或經預測之結合親和力。更可能結合至TAP之肽或結合具有更高親和力之TAP之肽更可能藉由MHC-I呈遞。

【0259】 對偶基因非相互作用資訊亦可包括腫瘤細胞中之TAP之表現水平(其可藉由RNA-seq、蛋白質組質譜法、免疫組織化學量測)。針對MHC-I，較高TAP表現水平增加所有肽之呈遞概率。

【0260】 對偶基因非相互作用資訊亦可包括腫瘤突變之存在或不存在，其包括(但不限於)：

i. 已知癌症驅動基因(諸如EGFR、KRAS、ALK、RET、ROS1、TP53、CDKN2A、CDKN2B、NTRK1、NTRK2、NTRK3)中之驅動子突變

ii. 於編碼涉及抗原呈遞機制之蛋白質之基因(例如，B2M、HLA-A、HLA-B、HLA-C、TAP-1、TAP-2、TAPBP、CALR、CNX、ERP57、HLA-DM、HLA-DMA、HLA-DMB、HLA-DO、HLA-DOA、HLA-DOBHLA-DP、HLA-DPA1、HLA-DPB1、HLA-DQ、HLA-DQA1、HLA-DQA2、HLA-DQB1、HLA-DQB2、HLA-DR、HLA-DRA、HLA-DRB1、HLA-DRB3、HLA-DRB4、HLA-DRB5或編碼蛋白酶體或免疫蛋白酶體之組分之基因中之任一者)中。呈遞依賴於遭受腫瘤中之功能喪失型突變之抗原呈遞機制之組分之肽具有降低的呈遞概率。

【0261】 功能生殖系多態性之存在或不存在包括(但不限於)：

i.於編碼涉及抗原呈遞機制之蛋白質之基因(例如，B2M、HLA-A、HLA-B、HLA-C、TAP-1、TAP-2、TAPBP、CALR、CNX、ERP57、HLA-DM、HLA-DMA、HLA-DMB、HLA-DO、HLA-DOA、HLA-DOBHLA-DP、HLA-DPA1、HLA-DPB1、HLA-DQ、HLA-DQA1、HLA-DQA2、HLA-DQB1、HLA-DQB2、HLA-DR、HLA-DRA、HLA-DRB1、HLA-DRB3、HLA-DRB4、HLA-DRB5或編碼蛋白酶體或免疫蛋白酶體之組分之基因中之任一者)中

【0262】對偶基因非相互作用資訊亦可包括腫瘤類型(例如，NSCLC，黑色素瘤)。

【0263】對偶基因非相互作用資訊亦可包括如藉由(例如) HLA對偶基因尾綴所反映之HLA對偶基因之已知功能性。例如，對偶基因名稱HLA-A*24:09N中之N尾綴指示不表現之無效對偶基因及因此不可能呈遞抗原決定基；在<https://www.ebi.ac.uk/ipd/imgt/hla/nomenclature/suffixes.html>描述全HLA對偶基因尾綴命名法。

【0264】對偶基因非相互作用資訊亦可包括臨床腫瘤亞型(例如，鱗狀肺癌相對於非鱗狀)。

【0265】對偶基因非相互作用資訊亦可包括吸煙史。

【0266】對偶基因非相互作用資訊亦可包括曬傷、陽光暴露或暴露於其他誘變劑之歷史。

【0267】對偶基因非相互作用資訊亦可包括視情況藉由驅動子突變分類之相關腫瘤類型或臨床亞型中之肽之源基因的典型表現。更可能呈遞通常於相關腫瘤類型中在高水平下表現之基因。

【0268】 對偶基因非相互作用資訊亦可包括所有腫瘤中，或相同類型之腫瘤中，或來自具有至少一個共用MHC對偶基因之個體之腫瘤中，或具有至少一個共用MHC對偶基因之個體中之相同類型之腫瘤中之突變頻率。

【0269】 於突變腫瘤特異性肽之情況下，用於預測呈遞概率之特徵之清單亦可包括突變註解(例如，錯義、通讀、框移、融合等)或預測該突變是否導致無義介導之衰變(NMD)。例如，由於純合子早期停止突變，可將來自腫瘤細胞中未轉譯之蛋白質片段之肽指定為零之呈遞概率。NMD導致降低之mRNA轉譯，其降低呈遞概率。

VII.C.呈遞識別系統

【0270】 圖3為高階方塊圖，該圖說明根據一實施例之呈遞識別系統160之電腦邏輯組件。於此示例實施例中，呈遞識別系統160包括資料管理模組312、編碼模組314、訓練模組316及預測模組320。呈遞識別系統160亦包括訓練資料存儲170及呈遞模型存儲175。模型管理系統160之一些實施例具有與本文中所述彼等不同之模組。類似地，功能可以與本文中所述不同之方式在該等模組之間分佈。

VII.C.1.資料管理模組

【0271】 該資料管理模組312自呈遞資訊165產生訓練資料集170。各組訓練資料含有複數個資料實例，其中各資料實例 i 含有包含至少一個呈遞或非呈遞肽序列 p^i 、與該肽序列 p^i 相關之一或多個相關聯MHC對偶基因 a^i 之獨立變數 z^i 集，及表示呈遞識別系統160對預測獨立變數之新值感興趣之資訊的因變數 y^i 。

【0272】 於整篇說明書之其餘部分提及之一特定實施中，因變數 y^i

為二進位標籤，其指示肽 p^i 是否藉由一或多個相關聯MHC對偶基因 a^i 呈遞。然而，應瞭解，於其他實施中，因變數 y^i 可表示任何其他種類之資訊，該資訊為該呈遞識別系統160對預測對獨立變數 z^i 之相依性感興趣。例如，於另一實施中，因變數 y^i 亦可為指示針對資料實例識別之質譜離子流之數值。

【0273】 針對資料實例 i 之肽序列 p^i 為 k_i 胺基酸之序列，其中 k_i 可在資料實例 i 之間在範圍內變化。例如，針對MHC I類範圍可為8至15或針對MHC II類範圍可為6至30。於系統160之一特定實施中，訓練資料集中之所有肽序列 p^i 可具有相同長度，例如，9。肽序列中之胺基酸之數目可取決於MHC對偶基因之類型(例如，人類中之MHC對偶基因等)變化。針對資料實例 i 之MHC對偶基因 a^i 指示存在哪些MHC對偶基因與對應肽序列 p^i 相關。

【0274】 該資料管理模組312亦可包括額外對偶基因相互作用變數，諸如結合親和力 b^i 及穩定性 s^i 預測結合訓練資料170中含有之肽序列 p^i 及相關聯MHC對偶基因 a^i 。例如，該訓練資料170可含有肽 p^i 與 a^i 中所指示之相關聯MHC分子各者之間之結合親和力預測 b^i 。作為另一實例，該訓練資料170可含有針對 a^i 中所指示之MHC對偶基因各者之穩定性預測 s^i 。

【0275】 該資料管理模組312亦可包括對偶基因非相互作用變數 w^i ，諸如C-端側接序列及mRNA定量測量值連同肽序列 p^i 。

【0276】 該資料管理模組312亦識別不藉由MHC對偶基因呈遞之肽序列以生成訓練資料170。一般地，此涉及識別源蛋白之「更長」序列，該等序列包括在呈遞之前之呈遞肽序列。當呈遞資訊含有經基因改造之細胞株時，該資料管理模組312識別合成蛋白中之一系列肽序列，該等細胞

經暴露至不在該等細胞之MHC對偶基因上呈遞之合成蛋白。當呈遞資訊含有組織樣品時，該資料管理模組312識別呈遞肽序列起源之源蛋白，且識別不在組織樣品細胞之MHC對偶基因上呈遞之源蛋白中之一系列肽序列。

【0277】 該資料管理模組312亦可利用胺基酸之隨機序列人工生成肽及識別生成之序列，因為肽不在MHC對偶基因上呈遞。此可藉由隨機生成肽序列允許該資料管理模組312容易生成不在MHC對偶基因上呈遞之肽之大量合成資料實現。因為於現實中，小百分比之肽序列藉由MHC對偶基因呈遞，以合成方式生成之肽序列高度不可能藉由MHC對偶基因呈遞，即使其包含於由細胞處理之蛋白質中。

【0278】 圖4說明根據一實施例之訓練資料170A之實例組。具體而言，訓練資料170A中之第一個3資料實例指示來自涉及對偶基因HLA-C*01:03 及 3 肽 序 列 QCEIOWAREFLKEIGJ 、 FIEUHFWI 及 FEWRHRJTRUJR之單對偶基因細胞株之肽呈遞資訊。訓練資料170A中之第四資料實例指示來自涉及對偶基因HLA-B*07:02、HLA-C*01:03、HLA-A*01:01及肽序列QIEJOEIJIE之多對偶基因細胞株之肽資訊。第一資料實例指示肽序列QCEIOWARE不藉由對偶基因HLA-DRB3:01:01呈遞。如前兩段中所討論，負標記之肽序列可藉由資料管理模組312隨機生成或自呈遞肽之源蛋白識別。該訓練資料170A亦包括針對肽序列-對偶基因對之1000 nM之結合親和力預測及1小時之半衰期之穩定性預測。該訓練資料170A亦包括對偶基因非相互作用變數，諸如肽FJELFISBOSJFIE之C-端側接序列及 10^2 TPM之mRNA定量測量值。第四資料實例指示肽序列 QIEJOEIJIE 藉由 對 偶 基 因 HLA-B*07:02 、 HLA-C*01:03 或 HLA-

A*01:01中之一者呈遞。該訓練資料170A亦包括針對對偶基因各者之結合親和力預測及穩定性預測，以及肽之C-端側接序列及肽之mRNA定量測量值。

VII.C.2.編碼模組

[illegible]

【0280】 當該訓練資料170含有不同長度之胺基酸之序列時，該編碼模組314可藉由添加PAD字元以擴展預定字母表將肽進一步編碼至等長度向量中。例如，此可藉由利用PAD字元左墊肽序列直至肽序列之長度達到具有訓練資料170中之最大長度之肽序列進行。因此，當具有最大長度之肽序列具有 $k_{\max}$ 胺基酸時，該編碼模組314數位上表示作為 $(20+1) \cdot k_{\max}$ 元之列向量之各序列。作為實例，針對擴展之字母表{PAD, A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y}及 $k_{\max}=5$ 之最大胺基酸長

[illegible]

【0281】 雖然參考具有胺基酸序列之序列描述以上編碼序列資料之方法，但是可將該方法類似地擴展至其他類型之序列資料，諸如DNA或RNA序列資料及類似者。

【0282】 編碼模組314亦將針對資料實例*i*之一或多個MHC對偶基因 $\mathbf{a}^i$ 編碼為*m*個元之列向量，其中各元 $h=1, 2, \dots, m$ 對應於經獨特識別之MHC對偶基因。對應於針對資料實例*i*識別之MHC對偶基因之元具有1之值。否則，剩餘元具有0之值。作為實例，對應於*m*=4經獨特識別之MHC對偶基因類型 {HLA-A*01:01, HLA-C*01:08, HLA-B*07:02, HLA-DRB1*10:01 }中之多對偶基因細胞株之針對資料實例*i*的對偶基因HLA-B*07:02及HLA-DRB1*10:01可由4個元 $\mathbf{a}^i=[0 \ 0 \ 1 \ 1]$ 之列向量表示，其中 $a_3^i=1$ 且 $a_4^i=1$ 。雖然本文中利用4個經識別之MHC對偶基因類型描述實例，但是於實務中MHC對偶基因類型之數目可為上百個或上千個。如先前所討論，各資料實例*i*通常含有與肽序列 $\mathbf{p}_i$ 相關之至多6種不同MHC對偶基因類型。

【0283】 編碼模組314亦將針對各資料實例i之標籤 y_i 編碼為具有{0, 1}集之值之二進位變數，其中1之值指示肽 x^i 藉由相關聯MHC對偶基因 a^i 中之一者呈遞，及0之值指示肽 x^i 不藉由相關聯MHC對偶基因 a^i 中之任一

者呈遞。當因變數 y_i 表示質譜離子流時，該編碼模組314可使用各種函數(諸如針對在 $[0, \infty)$ 之間之離子流值具有 $(-\infty, \infty)$ 之範圍之對數(log)函數)額外標度該等值。

【0284】該編碼模組314可將針對肽 p_i 之一對對偶基因相互作用變數 $\mathbf{x}_h^i$ 及相關聯MHC對偶基因 h 表示為列向量，其中將對偶基因相互作用變數之數值表示一個接一個地連在一起。例如，編碼模組314可將 $\mathbf{x}_h^i$ 表示為等於 $[p^i]$ 、 $[p^i \ b_h^i]$ 、 $[p^i \ s_h^i]$ 或 $[p^i \ b_h^i \ s_h^i]$ 之列向量，其中 b_h^i 為針對肽 p_i 及相關聯MHC對偶基因 h 之結合親和力預測，及類似地針對 s_h^i 針對穩定性。或者，可將對偶基因相互作用變數之一或多個組合個別存儲(例如，作為個別向量或矩陣)。

【0285】於一實例中，編碼模組314藉由將結合親和力之量測或預測值併入對偶基因相互作用變數 $\mathbf{x}_h^i$ 中表示結合親和力資訊。

【0286】於一實例中，編碼模組314藉由將結合穩定性之量測或預測值併入對偶基因相互作用變數 $\mathbf{x}_h^i$ 中表示結合穩定性資訊。

【0287】於一實例中，編碼模組314藉由將結合率(binding on-rate)之量測或預測值併入對偶基因相互作用變數 $\mathbf{x}_h^i$ 中表示結合率資訊。

【0288】於一實例中，針對藉由I類MHC分子呈遞之肽，編碼模組314將肽長度表示為向量 $\mathbf{T}_k=[1(L_k=8) \ 1(L_k=9) \ 1(L_k=10) \ 1(L_k=11) \ 1(L_k=12) \ 1(L_k=13) \ 1(L_k=14) \ 1(L_k=15)]$ ，其中 1 為指示函數，且 L_k 表示肽 p_k 之長度。可將向量 $\mathbf{T}_k$ 包含於對偶基因相互作用變數 $\mathbf{x}_h^i$ 中。於另一實例中，針對藉由II類MHC分子呈遞之肽，編碼模組314將肽長度表示為向量 $\mathbf{T}_k=[1(L_k=6) \ 1(L_k=7) \ 1(L_k=8) \ 1(L_k=9) \ 1(L_k=10) \ 1(L_k=11) \ 1(L_k=12) \ 1(L_k=13) \ 1(L_k=14) \ 1(L_k=15) \ 1(L_k=16) \ 1(L_k=17) \ 1(L_k=18) \ 1(L_k=19)]$

$1(L_k=20) \ 1(L_k=21) \ 1(L_k=22) \ 1(L_k=23) \ 1(L_k=24) \ 1(L_k=25) \ 1(L_k=26) \ 1(L_k=27) \ 1(L_k=28) \ 1(L_k=29) \ 1(L_k=30)]$ ，其中 1 為指示函數，且 L_k 表示肽 p_k 之長度。可將向量 T_k 包含於對偶基因相互作用變數 x_h^i 中。

【0289】 於一實例中，編碼模組314藉由將MHC對偶基因之基於RNA-seq之表現水平併入對偶基因相互作用變數 x_h^i 中表示MHC對偶基因之RNA表現資訊。

【0290】 類似地，編碼模組314可將對偶基因非相互作用變數 w^i 表示為列向量，其中將對偶基因非相互作用變數之數值表示一個接一個地連在一起。例如， w^i 可為等於 $[c^i]$ 或 $[c^i \ m^i \ w^i]$ 之列向量，其中 w^i 為表示除了肽 p^i 之C-端側接序列及與該肽相關之mRNA定量量測 m^i 外之任何其他對偶基因非相互作用變數之列向量。或者，可將對偶基因非相互作用變數之一或多個組合個別存儲(例如，作為個體向量或矩陣)。

【0291】 於一實例中，編碼模組314藉由將周轉率或半衰期併入對偶基因非相互作用變數 w^i 中表示針對肽序列之源蛋白之周轉率。

【0292】 於一實例中，編碼模組314藉由將蛋白質長度併入對偶基因非相互作用變數 w^i 中表示源蛋白或同功異型物之長度。

【0293】 於一實例中，編碼模組314藉由將包含 $\beta 1_i$ 、 $\beta 2_i$ 、 $\beta 5_i$ 亞基之免疫蛋白酶體特異性蛋白酶體亞基之平均表現併入對偶基因非相互作用變數 w^i 中表示免疫蛋白酶體之激活。

【0294】 於一實例中，編碼模組314表示肽之源蛋白或肽之基因或轉錄本之RNA-seq豐度(藉由諸如RSEM之技術以FPKM、TPM為單位定量)可將源蛋白之豐度併入對偶基因非相互作用變數 w^i 中。

【0295】 於一實例中，編碼模組314表示肽源之轉錄本將經歷無義

介導之衰變(NMD)之概率，如藉由(例如) Rivas等人，Science, 2015中之模型藉由將此概率併入對偶基因非相互作用變數 w^i 中所評估。

【0296】 於一實例中，編碼模組314表示藉由(例如)以TPM為單位定量路徑中之基因之表現，使用(例如)針對路徑中之基因各者之RSEM，然後計算概要統計(例如，跨路徑中之基因之平均值)經由RNA-seq評估之基因模組或路徑的激活狀態。可將該平均值併入對偶基因非相互作用變數 w^i 中。

【0297】 於一實例中，編碼模組314藉由將複本數目併入對偶基因非相互作用變數 w^i 中表示源基因之複本數目。

【0298】 於一實例中，編碼模組314藉由將量測或預測之TAP結合親和力(例如，以奈莫耳單位)包含於對偶基因非相互作用變數 w^i 中表示TAP結合親和力。

【0299】 於一實例中，編碼模組314藉由將藉由RNA-seq量測(及藉由(例如) RSEM以TPM為單位定量)之TAP表現水平包含於對偶基因非相互作用變數 w^i 中表示TAP表現水平。

【0300】 於一實例中，編碼模組314將腫瘤突變表示為對偶基因非相互作用變數 w^i 中之指示變數之向量(即，若肽 p^k 來自具有KRAS G12D突變之樣品，則 $d^k = 1$ 及否則， $d^k = 0$)。

【0301】 於一實例中，編碼模組314將抗原呈遞基因中之生殖系多態性表示為指示變數之向量(即，若肽 p^k 來自具有TAP中之特異性生殖系多態性之樣品，則 $d^k = 1$)。可將此等指示變數包含於對偶基因非相互作用變數 w^i 中。

【0302】 於一實例中，編碼模組314將腫瘤類型表示為在腫瘤類型

(例如，NSCLC、黑色素瘤、結腸直腸癌等)字母表內之長度1之獨熱編碼向量。可將此等獨熱編碼之變數包含於對偶基因非相互作用變數 w^i 中。

【0303】 於一實例中，編碼模組314藉由利用不同尾綴處理4-數位HLA對偶基因表示MHC對偶基因尾綴。例如，出於模型之目的，認為HLA-A*24:09N為來自HLA-A*24:09之不同等位基因。或者，針對所有肽，可將藉由N-尾綴之MHC對偶基因呈遞之概率設置為零，因為以N尾綴結束之HLA對偶基因不表現。

【0304】 於一實例中，編碼模組314將腫瘤亞型表示為內在腫瘤亞型(例如，肺腺癌、肺鱗狀細胞癌等)字母表內之長度1之獨熱編碼向量。可將此等獨熱編碼之變數包含於對偶基因非相互作用變數 w^i 中。

【0305】 於一實例中，編碼模組314將吸煙史表示二進位指示變數(若患者有吸煙史，則 $d^k = 1$ ，及否則 $d^k = 0$)，可將該變數包含於對偶基因非相互作用變數 w^i 中。或者，可將吸煙史編碼為超過吸煙嚴重度字母表之長度1之獨熱編碼之變數。例如，吸煙狀態可按1至5標度評定，其中1指示不吸煙者，及5指示目前重度吸煙者。因為吸煙史主要與肺腫瘤有關，所以當訓練多腫瘤類型之模型時，若患者有吸煙史且該腫瘤類型為肺腫瘤，則亦可將此變數定義為等於1及否則為0。

【0306】 於一實例中，編碼模組314將曬傷史表示為二進位指示變數(若患者有嚴重曬傷史，則 $d^k = 1$ 及否則 $d^k = 0$)，可將該變數包含於對偶基因非相互作用變數 w^i 中。因為嚴重曬傷主要與黑色素瘤有關，所以當訓練多腫瘤類型之模型時，若患者有嚴重曬傷史且該腫瘤類型為黑色素瘤，則亦可將此變數定義為等於1及否則為0。

【0307】 於一實例中，編碼模組314藉由使用參考資料庫(諸如

TCGA)將人類基因組中之各基因或轉錄本之特定基因或轉錄本之表現水平的分佈表示為表現水平分佈之概要統計(例如，平均值、中值)。具體而言，針對具有腫瘤類型黑色素瘤之樣品中之肽 p^k ，可不僅包含對偶基因非相互作用變數 w^i 中之肽 p^k 源之基因或轉錄本之量測之基因或轉錄本表現水平，而且包含黑色素瘤中之肽 p^k 源之基因或轉錄本之平均及/或中間基因或轉錄本表現，如藉由TCGA所量測。

【0308】 於一實例中，編碼模組314將突變類型表示為在突變類型(例如，錯義、框移、NMD-誘導等)字母表內之長度1之獨熱編碼之變數。可將此等獨熱編碼之變數包含於對偶基因非相互作用變數 w^i 中。

【0309】 於一實例中，編碼模組314將蛋白質之蛋白質級特徵表示為對偶基因非相互作用變數 w^i 中之源蛋白之註解(例如，5' UTR長度)之值。於另一實例中，編碼模組314藉由包含指示變數表示針對肽 p^i 之源蛋白之殘基級註解，若肽 p^i 與螺旋模體重疊，則該指示變數等於1及否則為0，或若肽 p^i 完全包含於對偶基因非相互作用變數 w^i 之螺旋模體內，則該指示變數等於1。於另一實例中，可將表示包含於螺旋模體註解內之肽 p^i 中之殘基之比例的特徵包含於對偶基因非相互作用變數 w^i 中。

【0310】 於一實例中，編碼模組314將人類蛋白質組中之蛋白質類型或同功異型物表示為具有等於人類蛋白質組中之蛋白質或同功異型物之數目之長度的指示向量 o^k ，及若肽 p^k 來自蛋白質 i ，則對應元 o^k_i 為1及否則為0。

【0311】 於一實例中，編碼模組314將肽 p^i 之源基因 G =基因(p^i)表示為具有 L 可能類別之分類變數，其中 L 表示索引源基因之數目之上限1、2、...、 L 。

【0312】 於一實例中，編碼模組314將肽 p^i 之組織類型、細胞類型、腫瘤類型或腫瘤組織學類型 $T = \text{組織}(p^i)$ 表示為具有 M 可能類別之分類變數，其中 M 表示索引類型之數目之上限 $1, 2, \dots, M$ 。組織之類型可包括(例如)肺組織、心臟組織、腸組織、神經組織及類似者。細胞之類型可包括樹突狀細胞、巨噬細胞、CD4 T細胞及類似者。腫瘤之類型可包括肺腺癌、肺鱗狀細胞癌、黑色素瘤、非霍奇金氏(Hodgkin)淋巴瘤及類似者。

【0313】 編碼模組314亦可將針對肽 p^i 及相關聯MHC對偶基因 h 之整組變數 z^i 表示為列向量，其中將對偶基因相互作用變數 x^i 及對偶基因非相互作用變數 w^i 之數值表示一個接一個地連在一起。例如，編碼模組314可將 z_h^i 表示為等於 $[x_h^i \ w^i]$ 或 $[w_i \ x_h^i]$ 之列向量。

VIII. 訓練模組

【0314】 訓練模組316構建一或多個呈遞模型，該等模型產生肽序列是否將藉由與肽序列相關聯之MHC對偶基因呈遞之可能性。具體而言，給定肽序列 p^k 及與該肽序列 p^k 相關之MHC對偶基因集 a^k ，各呈遞模型產生指示肽序列 p^k 將藉由相關聯MHC對偶基因 a^k 中之一或多者呈遞之可能性的估計 u_k 。

VIII.A. 概觀

【0315】 訓練模組316構建基於自165中所存儲之呈遞資訊產生之存儲170中存儲之訓練資料集的一或多個呈遞模型。一般而言，不管呈遞模型之特定類型，所有呈遞模型捕獲訓練資料170中之獨立變數與因變數之間之相依性使得損失函數最小化。具體而言，該損失函數 $\ell(y_{i \in S}, u_{i \in S}; \theta)$ 表示針對訓練資料170中之一或多個資料實例 S 之因變數 $y_{i \in S}$ 與針對藉由呈遞模型產生之資料實例 S 之預估可能性 $u_{i \in S}$ 之值之間的差異。於整篇說明

書之其餘部分提及之一特定實施中，該損失函數 $(y_{i \in S}, u_{i \in S}; \theta)$ 為藉由如下方程式(1a)給定之負對數可能性函數：

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

然而，於實務中，可使用另一損失函數。例如，當針對質譜離子流作出預測時，該損失函數為藉由如下方程式1b給定之平均平方損失：

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

【0316】 該呈遞模型可為參數模型，其中一或多個參數 θ 算數地指定獨立變數與因變數之間之相依性。通常，通過基於梯度之數值優化演算法(諸如分批梯度演算法、隨機梯度演算法及類似者)確定最小化損失函數 $(y_{i \in S}, u_{i \in S}; \theta)$ 之參數型呈遞模型之各種參數。或者，該呈遞模型可為非參數模型，其中該模型結構自訓練資料170確定且不嚴格基於固定參數集。

VIII.B.按對偶基因(Per-Allele)模型

【0317】 訓練模組316可構建呈遞模型以預測按對偶基因基礎上之肽之呈遞可能性。於此情況下，訓練模組316可訓練基於自表現單MHC對偶基因之細胞產生之訓練資料170中之資料實例S的呈遞模型。

【0318】 於一實施中，訓練模組316藉由以下建模針對肽 p^k 針對特異性對偶基因 h 之預估呈遞可能性 u_k ：

$$u_k^h = \Pr(\text{經呈遞之 } p^k : \text{MHC對偶基因 } h) = f(g_h(x_h^k; \theta_h)) \tag{2}$$

其中肽序列 x_h^k 表示針對肽 p^k 及對應MHC對偶基因 h 之編碼之對偶基因相互作用變數， $f(\cdot)$ 為任何函數，且出於描述之方便，遍及本文被稱作轉換函數。另外， $g_h(\cdot)$ 為任何函數，出於描述之方便，遍及本文被稱作相依函數，及產生基於針對MHC對偶基因 h 確定之參數集 θ_h 之對偶基因相互作用

變數 $\mathbf{x}_h^k$ 的相依性分數。各MHC對偶基因 h 之該參數集 θ_h 之值可藉由最小化關於 θ_h 之損失函數確定，其中 i 為自表現單MHC對偶基因 h 之細胞產生之訓練資料170之子集 S 中的各實例。

【0319】相依函數 $g_h(\mathbf{x}_h^k; \theta_h)$ 之輸出表示MHC對偶基因 h 之相依性分數，該分數指示MHC對偶基因 h 是否將基於至少對偶基因相互作用特徵 $\mathbf{x}_h^k$ ，及特定言之，基於肽 $\mathbf{p}^k$ 之肽序列之胺基酸的位置呈遞對應新抗原。例如，若MHC對偶基因 h 可能呈遞肽 $\mathbf{p}^k$ ，則MHC對偶基因 h 之相依性分數可具有高的值，及若呈遞係不可能，則可具有低的值。轉換函數 $f(\cdot)$ 轉換輸入，及更具體而言，將藉由於此情況下之 $g_h(\mathbf{x}_h^k; \theta_h)$ 產生之相依性分數轉換成適宜值以指示肽 $\mathbf{p}^k$ 將藉由MHC對偶基因呈遞之可能性。

【0320】於整篇說明書之其餘部分提及之一特定實施中， $f(\cdot)$ 為具有適宜域範圍之 $[0, 1]$ 內之範圍之函數。於一實例中， $f(\cdot)$ 為藉由以下給定之場外(expit)函數：

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

作為另一實例， $f(\cdot)$ 亦可為藉由以下給定之雙曲正切函數：

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

當域 z 之值等於或大於0時。或者，當對具有範圍 $[0, 1]$ 外之值之質譜離子流作出預測時， $f(\cdot)$ 可為任何函數，諸如恆等函數、指數函數、對數函數及類似者。

【0321】因此，肽序列 $\mathbf{p}^k$ 將藉由MHC對偶基因 h 呈遞之按對偶基因可能性可藉由將MHC對偶基因 h 之相依函數 $g_h(\cdot)$ 應用於編碼之肽序列 $\mathbf{p}^k$ 之版本以生成對應相依性分數產生。該相依性分數可藉由轉換函數 $f(\cdot)$ 轉換以產生肽序列 $\mathbf{p}^k$ 將藉由MHC對偶基因 h 呈遞之按對偶基因可能性。

VIII.B.1對偶基因相互作用變數之相依函數

【0322】於整篇說明書提及之一特定實施中，該相依函數 $g_h(\cdot)$ 為藉由以下給定之仿射函數：

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

該仿射函數將 x_h^k 中之各對偶基因相互作用變數與針對相關聯MHC對偶基因 h 確定之該參數集 θ_h 中之對應參數線性組合。

【0323】於整篇說明書提及之另一特定實施中，該相依函數 $g_h(\cdot)$ 為藉由以下給定之網路函數：

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

該函數藉由具有排列於一或多個層中之一系列節點之網路模型 $NN_h(\cdot)$ 表示。一個節點可通過各者具有該參數集 θ_h 中之相關參數之聯繫連接至其他節點。可將一個特定節點處之值表示為連接至藉由與特定節點相關之激活函數映射之相關參數加權之特定節點之節點之值的總和。與仿射函數相反，網路模型係有利的，因為呈遞模型可併入具有不同胺基酸序列長度之非線性及過程資料。具體而言，通過非線性建模，網路模型可捕獲肽序列中之不同位置處之胺基酸之間之相互作用及此相互作用如何影響肽呈遞。

【0324】總之，可將網路模型 $NN_h(\cdot)$ 構建為前饋網路，諸如人工神經網路(ANN)、卷積神經網路(CNN)、深層神經網路(DNN)及/或重複網路，諸如，長短期記憶網路(LSTM)、雙向重複網路、深層雙向重複網路及類似者。

【0325】於整篇說明書之其餘部分提及之一實例中， $h=1, 2, \dots, m$ 中之各MHC對偶基因係與單獨網路模型相關，及 $NN_h(\cdot)$ 表示自與MHC對偶基因 h 相關之網路模型之輸出。

【0326】圖5說明與任意MHC對偶基因 $h=3$ 相關之實例網路模型 $NN_3(\cdot)$ 。如圖5中所示，針對MHC對偶基因 $h=3$ 之網路模型 $NN_3(\cdot)$ 包括層 $l=1$ 處之三個輸入節點，層 $l=2$ 處之四個節點，層 $l=3$ 處之兩個節點及層 $l=4$ 處之一個輸出節點。該網路模型 $NN_3(\cdot)$ 係與一組十個參數 $\theta_3(1)$ 、 $\theta_3(2)$ 、...、 $\theta_3(10)$ 相關。該網路模型 $NN_3(\cdot)$ 接收針對MHC對偶基因 $h=3$ 之三個對偶基因相互作用變數 $x_3^k(1)$ 、 $x_3^k(2)$ 及 $x_3^k(3)$ 之輸入值(個別資料實例包括編碼之多肽序列資料及使用之任何其他訓練資料)並輸出值 $NN_3(x_3^k)$ 。網路函數亦可包括一或多個網路模型，該等模型各者將不同對偶基因相互作用變數作為輸入。

【0327】於另一實例中，經識別之MHC對偶基因 $h=1, 2, \dots, m$ 係與單網路模型 $NN_h(\cdot)$ 相關，且 $NN_h(\cdot)$ 表示與MHC對偶基因 h 相關之單網路模型之一或多個輸出。於此實例中，該參數集 θ_h 可對應於該單網路模型之參數集，及因此，該參數集 θ_h 可藉由所有MHC對偶基因共用。

【0328】圖6A說明藉由MHC對偶基因 $h=1, 2, \dots, m$ 共享之實例網路模型 $NN_h(\cdot)$ 。如圖6A中所示，該網路模型 $NN_h(\cdot)$ 包含各者對應於MHC對偶基因之 m 個輸出節點。該網路模型 $NN_h(\cdot)$ 接收針對MHC對偶基因 $h=3$ 之對偶基因相互作用變數 x_3^k 並輸出包含對應於MHC對偶基因 $h=3$ 之值 $NN_3(x_3^k)$ 之 m 個值。

【0329】於又一實例中，該單網路模型 $NN_h(\cdot)$ 可為輸出對偶基因相互作用變數 x_h^k 及MHC對偶基因 h 之經編碼之蛋白質序列 d_h 給定之相依性分數的網路模型。於此實例中，該參數集 θ_h 可再次對應於該單網路模型之參數集，及因此，該參數集 θ_h 可藉由所有MHC對偶基因共用。因此，於此實例中， $NN_h(\cdot)$ 可表示單網路模型之輸入 $[x_h^k \ d_h]$ 給定之單網路模型 $NN_h(\cdot)$

之輸出。此網路模型係有利的，因為訓練資料中未知之MHC對偶基因之肽呈遞概率可僅藉由其蛋白質序列之識別預測。

【0330】圖6B說明藉由MHC對偶基因共享之實例網路模型 $NN_H(\cdot)$ 。如圖6B中所示，該網路模型 $NN_H(\cdot)$ 接收對偶基因相互作用變數及MHC對偶基因 $h=3$ 之蛋白質序列作為輸入，並輸出對應於MHC對偶基因 $h=3$ 之相依性分數 $NN_3(x_3^k)$ 。

【0331】於又一實例中，可將相依函數 $g_h(\cdot)$ 表示為：

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

其中 $g'_h(x_h^k; \theta'_h)$ 為具有參數集 θ'_h 之仿射函數、具有針對MHC對偶基因之對偶基因相互作用變數之該參數集中之偏置參數 θ_h^0 的網路函數或類似者，該偏置參數表示MHC對偶基因 h 之呈遞之基線概率。

【0332】於另一實施中，偏置參數 θ_h^0 可根據MHC對偶基因 h 之基因家族共用。即，MHC對偶基因 h 之偏置參數 θ_h^0 可等於 $\theta_{\text{gene}(h)}^0$ ，其中基因 (h) 為MHC對偶基因 h 之基因家族。例如，可將I類MHC對偶基因HLA-A*02:01、HLA-A*02:02及HLA-A*02:03分配至「HLA-A」之基因家族，及可共用此等MHC對偶基因各者之偏置參數 θ_h^0 。作為另一實例，可將II類MHC對偶基因HLA-DRB1:10:01、HLA-DRB1:11:01及HLA-DRB3:01:01分配至「HLA-DRB」之基因家族，及可共用此等MHC對偶基因各者之偏置參數 θ_h^0 。

【0333】回到方程式(2)，作為實例，使用仿射相依函數 $g_h(\cdot)$ ，肽 p^k 將藉由 $m=4$ 個不同經識別之MHC對偶基因中之MHC對偶基因 $h=3$ 呈遞之可能性可藉由以下產生：

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

其中 $\mathbf{x}_3^k$ 為針對MHC對偶基因 $h=3$ 之經識別之對偶基因相互作用變數，且 θ_3 為針對MHC對偶基因 $h=3$ 通過損失函數最小化確定之參數集。

【0334】作為另一實例，使用分開網路轉換函數 $g_h(\cdot)$ ，肽 $\mathbf{p}^k$ 將藉由 $m=4$ 個不同經識別之MHC對偶基因中之MHC對偶基因 $h=3$ 呈遞之可能性可藉由以下產生：

$$u_k^3 = f(NN_3(\mathbf{x}_3^k; \theta_3)) .$$

其中 $\mathbf{x}_3^k$ 為針對MHC對偶基因 $h=3$ 之經識別之對偶基因相互作用變數，且 θ_3 為針對與MHC對偶基因 $h=3$ 相關之網路模型 $NN_3(\cdot)$ 確定之參數集。

【0335】圖7說明使用實例網路模型 $NN_3(\cdot)$ ，產生與MHC對偶基因 $h=3$ 相關之肽 $\mathbf{p}^k$ 之呈遞可能性。如圖7中所示，該網路模型 $NN_3(\cdot)$ 接收針對MHC對偶基因 $h=3$ 之對偶基因相互作用變數 $\mathbf{x}_3^k$ 並產生輸出 $NN_3(\mathbf{x}_3^k)$ 。將該輸出藉由函數 $f(\cdot)$ 映射以產生預估呈遞可能性 u_k 。

VIII.B.2.利用對偶基因非相互作用變數之按對偶基因

【0336】於一實施中，訓練模組316併入對偶基因非相互作用變數及藉由以下建模肽 $\mathbf{p}^k$ 之預估呈遞可能性 u_k ：

$$u_k^h = \Pr(\text{經呈遞之 } \mathbf{p}^k) = f\left(g_w(\mathbf{w}^k; \theta_w) + g_h(\mathbf{x}_h^k; \theta_h)\right) . \quad (8)$$

其中 $\mathbf{w}^k$ 表示肽 $\mathbf{p}^k$ 之經編碼之對偶基因非相互作用變數， $g_w(\cdot)$ 為基於針對對偶基因非相互作用變數確定之參數集 θ_w 之對偶基因非相互作用變數 $\mathbf{w}^k$ 的函數。具體而言，針對各MHC對偶基因 h 之該參數集 θ_h 及針對對偶基因非相互作用變數之該參數集 θ_w 之值可藉由最小化關於 θ_h 及 θ_w 之損失函數確定，其中 i 為自表現單MHC對偶基因之細胞產生之訓練資料170之子集 S 中的各實例。

【0337】相依函數 $g_w(\mathbf{w}^k; \theta_w)$ 之輸出表示對偶基因非相互作用變數之相依性分數，該分數指示肽 $\mathbf{p}^k$ 是否將藉由基於對偶基因非相互作用變數之影響之一或多個MHC對偶基因呈遞。例如，若肽 $\mathbf{p}^k$ 係與已知正面影響肽 $\mathbf{p}^k$ 之呈遞之C-端側接序列相關，則對偶基因非相互作用變數之相依性分數可具有高的值，及若肽 $\mathbf{p}^k$ 係與已知負面影響肽 $\mathbf{p}^k$ 之呈遞之C-端側接序列相關，則對偶基因非相互作用變數之相依性分數可具有低的值。

【0338】根據方程式(8)，肽序列 $\mathbf{p}^k$ 將藉由MHC對偶基因 h 呈遞之按對偶基因可能性可藉由將MHC對偶基因 h 之函數 $g_h(\cdot)$ 應用於肽序列 $\mathbf{p}^k$ 之經編碼版本以產生對偶基因相互作用變數之對應相依性分數而產生。亦將對偶基因非相互作用變數之函數 $g_w(\cdot)$ 應用於對偶基因非相互作用變數之經編碼版本以產生對偶基因非相互作用變數之相依性分數。將兩個分數組合，及將組合之分數藉由轉換函數 $f(\cdot)$ 轉換以產生肽序列 $\mathbf{p}^k$ 將藉由MHC對偶基因 h 呈遞之按對偶基因可能性。

【0339】或者，訓練模組316可包含對偶基因非相互作用變數 $\mathbf{w}^k$ ，藉由添加該對偶基因非相互作用變數 $\mathbf{w}^k$ 至方程式(2)中之對偶基因相互作用變數 $\mathbf{x}_h^k$ 中預測。因此，該呈遞可能性可藉由以下給定：

$$u_k^h = \Pr(\text{肽呈遞之 } \mathbf{p}^k : \text{對偶基因 } h) = f\left(g_h([\mathbf{x}_h^k \mathbf{w}^k]; \theta_h)\right). \tag{9}$$

VIII.B.3對偶基因非相互作用變數之相依函數

【0340】類似於對偶基因相互作用變數之相依函數 $g_h(\cdot)$ ，對偶基因非相互作用變數之相依函數 $g_w(\cdot)$ 可為仿射函數或網路函數，其中單獨網路模型係與對偶基因非相互作用變數 $\mathbf{w}^k$ 相關。

【0341】具體而言，該相依函數 $g_w(\cdot)$ 為藉由以下給定之仿射函數：

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

該函數將對偶基因非相互作用變數 $\mathbf{w}^k$ 與參數集 θ_w 中之對應參數線性組合。

【0342】 該相依函數 $g_w(\cdot)$ 亦可為藉由以下給定之網路函數：

$$g_w(\mathbf{w}^k; \theta_w) = NN_w(\mathbf{w}^k; \theta_w) .$$

其由具有參數集 θ_w 中之相關參數之網路模型 $NN_w(\cdot)$ 表示。該網路函數亦可包括各者將不同對偶基因非相互作用變數作為輸入之一或多個網路模型。

【0343】 於另一實例中，對偶基因非相互作用變數之相依函數 $g_w(\cdot)$ 可藉由以下給定：

$$g_w(\mathbf{w}^k; \theta_w) = g'_w(\mathbf{w}^k; \theta'_w) + h(m^k; \theta_w^m) , \tag{10}$$

其中 $g'_w(\mathbf{w}^k; \theta'_w)$ 為仿射函數、具有對偶基因非相互作用參數集 θ'_w 之網路函數或類似者， m^k 為肽 $\mathbf{p}^k$ 之mRNA定量量測， $h(\cdot)$ 為轉換定量量測之函數，且 θ_w^m 為對偶基因非相互作用變數之參數集中之參數，將該參數與mRNA定量量測組合以產生mRNA定量量測之相依性分數。於整篇說明書之其餘部分提及之一特定實施例中， $h(\cdot)$ 為對數函數，然而於實務中 $h(\cdot)$ 可為各種不同函數中之任一者。

【0344】 於又一實例中，對偶基因非相互作用變數之相依函數 $g_w(\cdot)$ 可藉由以下給定：

$$g_w(\mathbf{w}^k; \theta_w) = g'_w(\mathbf{w}^k; \theta'_w) + \theta_w^o \cdot \mathbf{o}^k , \tag{11}$$

其中 $g'_w(\mathbf{w}^k; \theta'_w)$ 為仿射函數、具有對偶基因非相互作用參數集 θ'_w 之網路函數或類似者， $\mathbf{o}^k$ 為第VII.C.2節中所述之指示向量，其表示針對肽 $\mathbf{p}^k$ 之人類蛋白質組中之蛋白質及同功異型物，且 θ_w^o 為與該指示向量組合之對偶基因非相互作用變數之參數集中之參數集。於一變化中，當 $\mathbf{o}^k$ 及參數集 θ_w^o 之維度係顯著高時，當測定該等參數之值時，可將參數正則化項(諸如 $\lambda \cdot \|\theta_w^o\|$ ，其中 $\|\cdot\|$ 表示L1規範、L2規範、組合或類似者)添加至損失函數

中。可通過適宜方法測定超參數 λ 之最佳值。

【0345】 於又一實例中，對偶基因非相互作用變數之相依函數 $g_w(\cdot)$ 可藉由以下給定：

$$g_w(\mathbf{w}^k; \boldsymbol{\theta}_w) = g'_w(\mathbf{w}^k; \boldsymbol{\theta}'_w) + \sum_{l=1}^L \mathbb{1}(\text{基因}(\mathbf{p}^k) = l) \cdot \boldsymbol{\theta}_w^l, \quad (12)$$

其中 $g'_w(\mathbf{w}^k; \boldsymbol{\theta}'_w)$ 為仿射函數、具有對偶基因非相互作用參數集 $\boldsymbol{\theta}'_w$ 之網路函數或類似者，若肽 $\mathbf{p}^k$ 係來自如上所述之關於對偶基因非相互作用變數之源基因 l ，則 $\mathbb{1}(\text{基因}(\mathbf{p}^k)=l)$ 為等於1之指示函數，且 $\boldsymbol{\theta}_w^l$ 為指示源基因 l 之「抗原性」之參數。於一變化中，當 L 係顯著高，及因此參數 $\boldsymbol{\theta}_w^{l=1, 2, \dots, L}$ 之數目係顯著高時，當測定該等參數之值時，可將參數正則化項(諸如 $\lambda \cdot \|\boldsymbol{\theta}_w^l\|$ ，其中 $\|\cdot\|$ 表示L1規範、L2規範、組合或類似者)添加至損失函數中。可通過適宜方法測定超參數 λ 之最佳值。

【0346】 於又一實例中，對偶基因非相互作用變數之相依函數 $g_w(\cdot)$ 可藉由以下給定：

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

其中 $g'_w(\mathbf{w}^k; \boldsymbol{\theta}'_w)$ 為仿射函數、具有對偶基因非相互作用參數集 $\boldsymbol{\theta}'_w$ 之網路函數或類似者，若肽 $\mathbf{p}^k$ 係來自源基因 l 及若肽 $\mathbf{p}^k$ 係來自如上所述之關於對偶基因非相互作用變數之組織類型 m ，則 $\mathbb{1}(\text{基因}(\mathbf{p}^k)=l, \text{組織}(\mathbf{p}^k)=m)$ 為等於1之指示函數，且 $\boldsymbol{\theta}_w^{lm}$ 為指示源基因 l 與組織類型 m 之組合之抗原性的參數。具體而言，針對組織類型 m 之基因 l 之抗原性可表示組織類型 m 之細胞於控制RNA表現及肽序列環境後呈遞來自基因 l 之肽之殘餘傾向。

【0347】 於一變化中，當 L 或 M 係顯著高，及因此參數 $\boldsymbol{\theta}_w^{lm=1, 2, \dots, LM}$ 之數目係顯著高時，當測定該等參數之值時，可將參數正則化項(諸如

$\lambda \cdot \|\theta_w^{lm}\|$ ，其中 $\|\cdot\|$ 表示L1規範、L2規範、組合或類似者)添加至損失函數中。可通過適宜方法測定超參數 λ 之最佳值。於另一變化中，當測定該等參數之值時，可將參數正則化項添加至損失函數中使得相同源基因之參數在組織類型之間無顯著差異。例如，諸如以下之懲罰項：

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

(其中 $\overline{\theta_w^l}$ 為跨源基因 l 之組織類型之平均抗原性)可懲罰損失函數中之跨不同組織類型之抗原性的標準差。

【0348】 於又一實例中，對偶基因非相互作用變數之相依函數 $g_w(\cdot)$ 可藉由以下給定：

$$g_w(w^k; \theta_w) = g'_w(w^k; \theta'_w) + \sum_{l=1}^L \mathbb{1}(\text{src}(p^k) = l) \cdot \theta_w^l + \sum_{m=1}^M \mathbb{1}(\text{loc}(p^k) = m) \cdot \theta_w^m \tag{12c}$$

其中 $g'_w(w^k; \theta'_w)$ 為仿射函數、具有對偶基因非相互作用參數集 θ'_w 之網路函數或類似者，若肽 p^k 係來自如上所述之關於對偶基因非相互作用變數之源基因 l ，則 $\mathbb{1}(\text{src}(p^k) = l)$ 為等於1之指示函數，且 θ_w^l 為指示源基因 l 之「抗原性」之參數，且若肽 p^k 係來自蛋白質組映射 m ，則 $\mathbb{1}(\text{loc}(p^k) = m)$ 為等於1之指示函數，且 θ_w^m 為指示蛋白質組映射 m 為呈遞「熱點」之程度之參數。於一實施例中，蛋白質組映射可包括來自相同蛋白質之 n 個鄰近肽之區塊，其中 n 為經由適宜方法(諸如網格搜索交叉驗證)確定之模型之超參數。

【0349】 於實務中，可將方程式(10)、(11)、(12a)、(12b)及(12c)中任一者之額外項組合以產生對偶基因非相互作用變數之相依函數 $g_w(\cdot)$ 。例如，可將方程式(10)中之指示mRNA定量量測之項 $h(\cdot)$ 及方程式(12)中

之指示源基因抗原性之項連同任何其他仿射函數或網路函數加在一起以產生對偶基因非相互作用變數之相依函數。

【0350】 回到方程式(8)，作為實例，使用仿射轉換函數 $g_h(\cdot)$ 、 $g_w(\cdot)$ ，肽 p^k 將藉由 $m=4$ 個不同經識別之 MHC 對偶基因中之 MHC 對偶基因 $h=3$ 呈遞的可能性可藉由以下產生：

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

其中 w^k 為肽 p^k 之經識別之對偶基因非相互作用變數，且 θ_w 為針對對偶基因非相互作用變數確定之參數集。

【0351】 作為另一實例，使用網路轉換函數 $g_h(\cdot)$ 、 $g_w(\cdot)$ ，肽 p^k 將藉由 $m=4$ 個不同經識別之 MHC 對偶基因中之 MHC 對偶基因 $h=3$ 呈遞的可能性可藉由以下產生：

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

其中 w^k 為肽 p^k 之經識別之對偶基因相互作用變數，且 θ_w 為針對對偶基因非相互作用變數確定之參數集。

【0352】 圖8說明使用實例網路模型 $NN_3(\cdot)$ 及 $NN_w(\cdot)$ 產生與 MHC 對偶基因 $h=3$ 相關之肽 p^k 之呈遞可能性。如圖8中所示，該網路模型 $NN_3(\cdot)$ 接收針對 MHC 對偶基因 $h=3$ 之對偶基因相互作用變數 x_3^k 並產生輸出 $NN_3(x_3^k)$ 。該網路模型 $NN_w(\cdot)$ 接收針對肽 p^k 之對偶基因非相互作用變數 w^k 並產生輸出 $NN_w(w^k)$ 。將該等輸出組合並藉由函數 $f(\cdot)$ 映射以產生預估呈遞可能性 u_k 。

VIII.C.多對偶基因模型

【0353】 訓練模組316亦可構建呈遞模型以預測存在兩個或更多個 MHC 對偶基因之多對偶基因環境中之肽之呈遞可能性。於此情況下，該

訓練模組316可訓練基於自表現單MHC對偶基因之細胞、表現多MHC對偶基因之細胞或其組合產生之訓練資料170中之資料實例S的呈遞模型。

VIII.C.1.實例1：按對偶基因模型之最大值

【0354】於一實施中，訓練模組316將與多MHC對偶基因集H相關之肽 p^k 之預估呈遞可能性 u_k 建模為針對基於表現單對偶基因之細胞確定之該集H中之MHC對偶基因h各者確定之呈遞可能性 $u_k^{h \in H}$ 的函數，如以上結合方程式(2)至(11)所述。具體而言，該呈遞可能性 u_k 可為 $u_k^{h \in H}$ 之任何函數。於一實施中，如方程式(12)中所示，該函數為最大值函數，及可將該呈遞可能性 u_k 確定為該組H中之各MHC對偶基因h之呈遞可能性的最大值。

$$u_k = \Pr(\text{ 經呈遞之 } p^k : \text{ 對偶基因 } H) = \max(u_k^{h \in H}).$$

VIII.C.2.實例2.1：和函數模型

【0355】於一實施中，該訓練模組316藉由以下建模肽 p^k 之預估呈遞可能性 u_k ：

$$u_k = \Pr(\text{ 經呈遞之 } p^k) = f\left(\sum_{h=1}^m a_h^k \cdot g_h(x_h^k; \theta_h)\right). \tag{13}$$

其中元 a_h^k 針對與肽序列 p^k 相關之多MHC對偶基因H為1且 x_h^k 表示針對肽 p^k 及對應MHC對偶基因之編碼之對偶基因相互作用變數。針對各MHC對偶基因h之參數集 θ_h 之值可藉由最小化關於 θ_h 之損失函數確定，其中i為自表現單MHC對偶基因之細胞及/或表現多MHC對偶基因之細胞產生之訓練資料170之子集S中的各實例。相依函數 g_h 可呈以上第VIII.B.1節中引入之相依函數 g_h 中之任一者的形式。

【0356】根據方程式(13)，肽序列 p^k 將藉由一或多個MHC對偶基因h呈遞之呈遞可能性可藉由將相依函數 $g_h(\cdot)$ 應用於針對MHC對偶基因H各

者之肽序列 $\mathbf{p}^k$ 之編碼版本以產生對偶基因相互作用變數之對應分數而產生。將各MHC對偶基因 h 之分數組合，及藉由轉換函數 $f(\cdot)$ 轉換以產生肽序列 $\mathbf{p}^k$ 將藉由該組MHC對偶基因 H 呈遞之呈遞可能性。

【0357】 方程式(13)之呈遞模型不同於方程式(2)之按對偶基因模型，因為各肽 $\mathbf{p}^k$ 之相關對偶基因之數目可大於1。換言之， $\mathbf{a}_h^k$ 中之一個以上元可針對與肽序列 $\mathbf{p}^k$ 相關之多MHC對偶基因 H 具有1之值。

【0358】 作為實例，使用仿射轉換函數 $g_h(\cdot)$ ，肽 $\mathbf{p}^k$ 將藉由 $m=4$ 個不同經識別之MHC對偶基因中之MHC對偶基因 $h=2$ 、 $h=3$ 呈遞之可能性可藉由以下產生：

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

其中 $\mathbf{x}_2^k$ 、 $\mathbf{x}_3^k$ 為針對MHC對偶基因 $h=2$ 、 $h=3$ 之經識別之對偶基因相互作用變數，且 θ_2 、 θ_3 為針對MHC對偶基因 $h=2$ 、 $h=3$ 確定之參數集。

【0359】 作為另一實例，使用網路轉換函數 $g_h(\cdot)$ 、 $g_w(\cdot)$ ，肽 $\mathbf{p}^k$ 將藉由 $m=4$ 個不同經識別之MHC對偶基因中之MHC對偶基因 $h=2$ 、 $h=3$ 呈遞之可能性可藉由以下產生：

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

其中 $NN_2(\cdot)$ 、 $NN_3(\cdot)$ 為針對MHC對偶基因 $h=2$ 、 $h=3$ 之經識別之網路模型，且 θ_2 、 θ_3 為針對MHC對偶基因 $h=2$ 、 $h=3$ 確定之參數集。

【0360】 圖9說明使用實例網路模型 $NN_2(\cdot)$ 及 $NN_3(\cdot)$ 產生與MHC對偶基因 $h=2$ 、 $h=3$ 相關之肽 $\mathbf{p}^k$ 之呈遞可能性。如圖9中所示，該網路模型 $NN_2(\cdot)$ 接收針對MHC對偶基因 $h=2$ 之對偶基因相互作用變數 $\mathbf{x}_2^k$ 並產生輸出 $NN_2(\mathbf{x}_2^k)$ 及該網路模型 $NN_3(\cdot)$ 接收針對MHC對偶基因 $h=3$ 之對偶基因相互作用變數 $\mathbf{x}_3^k$ 並產生輸出 $NN_3(\mathbf{x}_3^k)$ 。將該等輸出組合及藉由函數 $f(\cdot)$ 映射

以產生預估呈遞可能性 u_k 。

VIII.C.3.實例2.2：利用對偶基因非相互作用變數之和函數模型

【0361】 於一實施中，訓練模組316併入對偶基因非相互作用變數且藉由以下建模肽 p^k 之預估呈遞可能性 u_k ：

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

其中 w^k 表示針對肽 p^k 之經編碼之對偶基因非相互作用變數。具體而言，針對各MHC對偶基因 h 之參數集 θ_h 及針對對偶基因非相互作用變數之參數集 θ_w 之值可藉由最小化關於 θ_h 及 θ_w 之損失函數確定，其中 i 為自表現單MHC對偶基因之細胞及/或表現多MHC對偶基因之細胞產生之訓練資料170之子集S中的各實例。相依函數 g_w 可呈以上第VIII.B.3節中引入之相依函數 g_w 中之任一者的形式。

【0362】 因此，根據方程式(14)，肽序列 p^k 將藉由一或多個MHC對偶基因H呈遞之呈遞可能性可藉由將函數 $g_h(\cdot)$ 應用於針對MHC對偶基因H各者之肽序列 p^k 之編碼版本以產生針對各MHC對偶基因 h 之對偶基因相互作用變數之對應相依性分數而產生。亦將對偶基因非相互作用變數之函數 $g_w(\cdot)$ 應用於對偶基因非相互作用變數之編碼版本以產生對偶基因非相互作用變數之相依性分數。將該等分數組合，及將組合之分數藉由轉換函數 $f(\cdot)$ 轉換以產生肽序列 p^k 將藉由MHC對偶基因H呈遞之呈遞可能性。

【0363】 於方程式(14)之呈遞模型中，針對各肽 p^k 之相關對偶基因之數目可大於1。換言之， a_h^k 中之一個以上元可針對與肽序列 p^k 相關之多MHC對偶基因H具有1之值。

【0364】 作為實例，使用仿射轉換函數 $g_h(\cdot)$ 、 $g_w(\cdot)$ ，肽 p^k 將藉由 $m=4$ 個不同經識別之MHC對偶基因中之MHC對偶基因 $h=2$ 、 $h=3$ 呈遞之

可能性可藉由以下產生：

$$u_k = f(\mathbf{w}^k \cdot \boldsymbol{\theta}_w + \mathbf{x}_2^k \cdot \boldsymbol{\theta}_2 + \mathbf{x}_3^k \cdot \boldsymbol{\theta}_3).$$

其中 $\mathbf{w}^k$ 為針對肽 $\mathbf{p}^k$ 之經識別之對偶基因非相互作用變數，且 $\boldsymbol{\theta}_w$ 為針對對偶基因非相互作用變數確定之參數集。

【0365】 作為另一實例，使用網路轉換函數 $g_h(\cdot)$ 、 $g_w(\cdot)$ ，肽 $\mathbf{p}^k$ 將藉由 $m=4$ 個不同經識別之MHC對偶基因中之MHC對偶基因 $h=2$ 、 $h=3$ 呈遞之可能性可藉由以下產生：

$$u_k = f(NN_w(\mathbf{w}^k; \boldsymbol{\theta}_w) + NN_2(\mathbf{x}_2^k; \boldsymbol{\theta}_2) + NN_3(\mathbf{x}_3^k; \boldsymbol{\theta}_3))$$

其中 $\mathbf{w}^k$ 為針對肽 $\mathbf{p}^k$ 之經識別之對偶基因相互作用變數，且 $\boldsymbol{\theta}_w$ 為針對對偶基因非相互作用變數確定之參數集。

【0366】 圖10說明使用實例網路模型 $NN_2(\cdot)$ 、 $NN_3(\cdot)$ 及 $NN_w(\cdot)$ 產生與MHC對偶基因 $h=2$ 、 $h=3$ 相關之肽 $\mathbf{p}^k$ 之呈遞可能性。如圖10中所示，該網路模型 $NN_2(\cdot)$ 接收針對MHC對偶基因 $h=2$ 之對偶基因相互作用變數 $\mathbf{x}_2^k$ 且產生輸出 $NN_2(\mathbf{x}_2^k)$ 。該網路模型 $NN_3(\cdot)$ 接收針對MHC對偶基因 $h=3$ 之對偶基因相互作用變數 $\mathbf{x}_3^k$ 且產生輸出 $NN_3(\mathbf{x}_3^k)$ 。該網路模型 $NN_w(\cdot)$ 接收針對肽 $\mathbf{p}^k$ 之對偶基因非相互作用變數 $\mathbf{w}^k$ 且產生輸出 $NN_w(\mathbf{w}^k)$ 。將該等輸出組合並藉由函數 $f(\cdot)$ 映射以產生預估呈遞可能性 u_k 。

【0367】 或者，訓練模組316可包含對偶基因非相互作用變數 $\mathbf{w}^k$ ，藉由將該對偶基因非相互作用變數 $\mathbf{w}^k$ 添加至方程式(15)中之對偶基因相互作用變數 $\mathbf{x}_h^k$ 中預測。因此，呈遞可能性可藉由以下給定：

$$u_k = \Pr(\text{經呈遞之 } \mathbf{p}^k) = f\left(\sum_{h=1}^m a_h^k \cdot g_h([\mathbf{x}_h^k \mathbf{w}^k]; \boldsymbol{\theta}_h)\right). \tag{15}$$

VIII.C.4.實例3.1：使用內隱按對偶基因可能性之模型

【0368】 於另一實施中，訓練模組316藉由以下建模肽 $\mathbf{p}^k$ 之預估呈

遞可能性 u_k ：

$$u_k = \Pr(\text{呈遞肽之 } p^k) = r\left(s(v = [a_1^k \cdot u_k^{'1}(\theta) \dots a_m^k \cdot u_k^{'m}(\theta)])\right) \quad (16)$$

其中元 a_h^k 針對與肽序列 p^k 相關之多MHC對偶基因 $h \in H$ 為1， $u_k^{'h}$ 為MHC對偶基因 h 之內隱按對偶基因呈遞可能性，向量 v 為其中元 v_h 對應於 $a_h^k \cdot u_k^{'h}$ 之向量， $s(\cdot)$ 為映射 v 之元之函數，且 $r(\cdot)$ 為剪輯輸入至給定範圍之值之剪輯函數。如下更詳細所述， $s(\cdot)$ 可為和函數或二階函數，但是應瞭解，於其他實施例中， $s(\cdot)$ 可為任何函數，諸如最大值函數。針對內隱按對偶基因可能性之參數集 θ 之值可藉由最小化關於 θ 之損失函數確定，其中 i 為自表現單MHC對偶基因之細胞及/或表現多MHC對偶基因之細胞產生之訓練資料170之子集S中的各實例。

【0369】 可將方程式(17)之呈遞模型中之呈遞可能性建模為各者對應於肽 p^k 將藉由個別MHC對偶基因 h 呈遞之可能性之內隱按對偶基因呈遞可能性 $u_k^{'h}$ 的函數。該內隱按對偶基因可能性區別於第VIII.B節之按對偶基因呈遞可能性，因為內隱按對偶基因可能性之參數可自多對偶基因環境學習，其中經呈遞肽與對應MHC對偶基因之間之直接關聯係未知，除了單對偶基因環境外。因此，於多對偶基因環境中，該呈遞模型不但可評估肽 p^k 是否將藉由作為整體之MHC對偶基因集 H 呈遞，而且可提供指示哪些MHC對偶基因 h 最可能呈遞肽 p^k 之個別可能性 $u_k^{'h \in H}$ 。此之優點為該呈遞模型可在無針對表現單MHC對偶基因之細胞之訓練資料下產生內隱可能性。

【0370】 於整篇說明書之其餘部分中提及之一特定實施中， $r(\cdot)$ 為具有範圍[0, 1]之函數。例如， $r(\cdot)$ 可為剪輯函數：

$$r(z) = \min(\max(z, 0), 1) \text{。}$$

其中選擇 z 與 1 之間之最小值作為呈遞可能性 u_k 。於另一實施中，當域 z 之值等於或大於 0 時， $r(\cdot)$ 為藉由以下給定之雙曲正切函數：

$$r(z) = \tanh(z) \text{。}$$

VIII.C.5.實例3.2：和函數模型

【0371】 於一特定實施中， $s(\cdot)$ 為和函數，且呈遞可能性藉由合計內隱按對偶基因呈遞可能性給定：

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

【0372】 於一實施中，MHC對偶基因 h 之內隱按對偶基因呈遞可能性藉由以下產生：

$$u_k^{'h} = f\left(g_h(x_h^k; \theta_h)\right) \tag{18}$$

使得呈遞可能性藉由以下評估：

$$u_k = \Pr(\text{肽序列 } p^k) = r\left(\sum_{h=1}^m a_h^k \cdot f\left(g_h(x_h^k; \theta_h)\right)\right) \tag{19}$$

【0373】 根據方程式(19)，肽序列 p^k 將藉由一或多個MHC對偶基因 H 呈遞之呈遞可能性可藉由將函數 $g_h(\cdot)$ 應用於針對MHC對偶基因 H 各者之肽序列 p^k 之編碼版本以產生對偶基因相互作用變數之對應相依性分數而產生。將各相依性分數首先藉由函數 $f(\cdot)$ 轉換以產生內隱按對偶基因呈遞可能性 $u_k^{'h}$ 。將該等按對偶基因可能性 $u_k^{'h}$ 組合，及可將剪輯函數應用於該等組合可能性以剪輯值至範圍 $[0, 1]$ 以產生肽序列 p^k 將藉由該組MHC對偶基因 H 呈遞之呈遞可能性。相依函數 g_h 可呈以上第VIII.B.1節中引入之相依函數 g_h 中之任一者的形式。

【0374】 作為實例，使用仿射轉換函數 $g_h(\cdot)$ ，肽 p^k 將藉由 $m=4$ 個不同經識別之MHC對偶基因中之MHC對偶基因 $h=2$ 、 $h=3$ 呈遞之可能性可藉由以下產生：

$$u_k = r\left(f(\mathbf{x}_2^k \cdot \boldsymbol{\theta}_2) + f(\mathbf{x}_3^k \cdot \boldsymbol{\theta}_3)\right),$$

其中 $\mathbf{x}_2^k$ 、 $\mathbf{x}_3^k$ 為針對MHC對偶基因 $h=2$ 、 $h=3$ 之經識別之對偶基因相互作用變數，且 $\boldsymbol{\theta}_2$ 、 $\boldsymbol{\theta}_3$ 為針對MHC對偶基因 $h=2$ 、 $h=3$ 確定之參數集。

【0375】 作為另一實例，使用網路轉換函數 $g_h(\cdot)$ 、 $g_w(\cdot)$ ，肽 $\mathbf{p}^k$ 將藉由 $m=4$ 個不同經識別之MHC對偶基因中之MHC對偶基因 $h=2$ 、 $h=3$ 呈遞之可能性可藉由以下產生：

$$u_k = r\left(f(NN_2(\mathbf{x}_2^k; \boldsymbol{\theta}_2)) + f(NN_3(\mathbf{x}_3^k; \boldsymbol{\theta}_3))\right),$$

其中 $NN_2(\cdot)$ 、 $NN_3(\cdot)$ 為針對MHC對偶基因 $h=2$ 、 $h=3$ 之經識別之網路模型，且 $\boldsymbol{\theta}_2$ 、 $\boldsymbol{\theta}_3$ 為針對MHC對偶基因 $h=2$ 、 $h=3$ 確定之參數集。

【0376】 圖11說明使用實例網路模型 $NN_2(\cdot)$ 及 $NN_3(\cdot)$ 產生與MHC對偶基因 $h=2$ 、 $h=3$ 相關之肽 $\mathbf{p}^k$ 之呈遞可能性。如圖9中所示，該網路模型 $NN_2(\cdot)$ 接收針對MHC對偶基因 $h=2$ 之對偶基因相互作用變數 $\mathbf{x}_2^k$ 且產生輸出 $NN_2(\mathbf{x}_2^k)$ 及該網路模型 $NN_3(\cdot)$ 接收針對MHC對偶基因 $h=3$ 之對偶基因相互作用變數 $\mathbf{x}_3^k$ 且產生輸出 $NN_3(\mathbf{x}_3^k)$ 。將各輸出藉由函數 $f(\cdot)$ 映射且組合以產生預估呈遞可能性 u_k 。

【0377】 於另一實施中，當對質譜離子流之對數作出預測時， $r(\cdot)$ 為對數函數且 $f(\cdot)$ 為指數函數。

VIII.C.6.實例3.3：具有對偶基因非相互作用變數之和函數模型

【0378】 於一實施中，MHC對偶基因 h 之內隱按對偶基因呈遞可能性藉由以下產生：

$$u_k^h = f\left(g_h(\mathbf{x}_h^k; \boldsymbol{\theta}_h) + g_w(\mathbf{w}^k; \boldsymbol{\theta}_w)\right), \tag{20}$$

使得該呈遞可能性藉由以下產生：

$$u_k = \Pr(\text{經識別之 } \mathbf{p}^k) = r\left(\sum_{h=1}^m \alpha_h^k \cdot f\left(g_w(\mathbf{w}^k; \boldsymbol{\theta}_w) + g_h(\mathbf{x}_h^k; \boldsymbol{\theta}_h)\right)\right), \tag{21}$$

以併入對偶基因非相互作用變數對肽呈遞之影響。

【0379】 根據方程式(21)，肽序列 $\mathbf{p}^k$ 將藉由一或多個MHC對偶基因H呈遞之呈遞可能性可藉由將函數 $g_h(\cdot)$ 應用於針對MHC對偶基因H各者之肽序列 $\mathbf{p}^k$ 之編碼版本以產生針對MHC對偶基因h之對偶基因相互作用變數之對應相依性分數而產生。亦將對偶基因非相互作用變數之函數 $g_w(\cdot)$ 應用於對偶基因非相互作用變數之編碼版本以產生對偶基因非相互作用變數之相依性分數。將對偶基因非相互作用變數之分數組合至對偶基因相互作用變數之相依性分數各者。將該等組合分數各者藉由函數 $f(\cdot)$ 轉換以產生內隱按對偶基因呈遞可能性。將該等內隱可能性組合，及可將剪輯函數應用於組合輸出以將值剪輯至範圍[0,1]中以產生肽序列 $\mathbf{p}^k$ 將藉由MHC對偶基因H呈遞之呈遞可能性。相依函數 g_w 可呈以上第VIII.B.3節中引入之相依函數 g_w 中之任一者的形式。

【0380】 作為實例，使用仿射轉換函數 $g_h(\cdot)$ 、 $g_w(\cdot)$ ，肽 $\mathbf{p}^k$ 將藉由 $m=4$ 個不同經識別之MHC對偶基因中之MHC對偶基因 $h=2$ 、 $h=3$ 呈遞之可能性可藉由以下產生：

$$u_k = r\left(f(\mathbf{w}^k \cdot \boldsymbol{\theta}_w + x_2^k \cdot \boldsymbol{\theta}_2) + f(\mathbf{w}^k \cdot \boldsymbol{\theta}_w + x_3^k \cdot \boldsymbol{\theta}_3)\right),$$

其中 $\mathbf{w}^k$ 為針對肽 $\mathbf{p}^k$ 之經識別之對偶基因非相互作用變數，且 $\boldsymbol{\theta}_w$ 為針對對偶基因非相互作用變數確定之參數集。

【0381】 作為另一實例，使用網路轉換函數 $g_h(\cdot)$ 、 $g_w(\cdot)$ ，肽 $\mathbf{p}^k$ 將藉由 $m=4$ 個不同經識別之MHC對偶基因中之MHC對偶基因 $h=2$ 、 $h=3$ 呈遞之可能性可藉由以下產生：

$$u_k = r\left(f(NN_w(\mathbf{w}^k; \boldsymbol{\theta}_w) + NN_2(x_2^k; \boldsymbol{\theta}_2)) + f(NN_w(\mathbf{w}^k; \boldsymbol{\theta}_w) + NN_3(x_3^k; \boldsymbol{\theta}_3))\right)$$

其中 $\mathbf{w}^k$ 為針對肽 $\mathbf{p}^k$ 之經識別之對偶基因相互作用變數，且 $\boldsymbol{\theta}_w$ 為針對對偶基

因非相互作用變數確定之參數集。

【0382】圖12說明使用實例網路模型 $NN_2(\cdot)$ 、 $NN_3(\cdot)$ 及 $NN_w(\cdot)$ 產生與MHC對偶基因 $h=2$ 、 $h=3$ 相關之肽 p^k 之呈遞可能性。如圖12中所示，該網路模型 $NN_2(\cdot)$ 接收針對MHC對偶基因 $h=2$ 之對偶基因相互作用變數 x_2^k 且產生輸出 $NN_2(x_2^k)$ 。該網路模型 $NN_w(\cdot)$ 接收針對肽 p^k 之對偶基因非相互作用變數 w^k 且產生輸出 $NN_w(w^k)$ 。將該等輸出組合及藉由函數 $f(\cdot)$ 映射。該網路模型 $NN_3(\cdot)$ 接收針對MHC對偶基因 $h=3$ 之對偶基因相互作用變數 x_3^k 且產生輸出 $NN_3(x_3^k)$ ，將該輸出再次與相同網路模型 $NN_w(\cdot)$ 之輸出 $NN_w(w^k)$ 組合及藉由函數 $f(\cdot)$ 映射。將兩種輸出組合以產生預估呈遞可能性 u_k 。

【0383】於另一實施中，MHC對偶基因 h 之內隱按對偶基因呈遞可能性藉由以下產生：

$$u_k^{'h} = f\left(g_h([x_h^k w^k]; \theta_h)\right). \tag{22}$$

使得該呈遞可能性藉由以下產生：

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

VIII.C.7.實例4：二階模型

【0384】於一實施中， $s(\cdot)$ 為二階函數，且肽 p^k 之預估呈遞可能性 u_k 藉由以下給定：

$$u_k = \Pr(p^k) = \sum_{h=1}^m a_h^k \cdot u_k^{'h}(\theta) - \sum_{h=1}^m \sum_{j < h} a_h^k \cdot a_j^k \cdot u_k^{'h}(\theta) \cdot u_k^{'j}(\theta) \tag{23}$$

其中元 $u_k^{'h}$ 為MHC對偶基因 h 之內隱按對偶基因呈遞可能性。針對內隱按對偶基因可能性之參數集 θ 之值可藉由最小化關於 θ 之損失函數確定，其中 i 為自表現單MHC對偶基因之細胞及/或表現多MHC對偶基因之細胞

產生之訓練資料170之子集S中的各實例。該等內隱按對偶基因呈遞可能性可呈上述方程式(18)、(20)及(22)中所示之任何形式。

【0385】 於一態樣中，方程式(23)之模型可暗示存在肽 p^k 將同時藉由兩個MHC對偶基因呈遞之可能性，其中藉由兩個HLA對偶基因之呈遞係統計上獨立的。

【0386】 根據方程式(23)，肽序列 p^k 將藉由一或多個MHC對偶基因H呈遞之呈遞可能性可藉由組合內隱按對偶基因呈遞可能性並減去各對MHC對偶基因將同時呈遞來自總和之肽 p^k 之可能性以產生肽序列 p^k 將藉由MHC對偶基因H呈遞之呈遞可能性而產生。

【0387】 作為實例，使用仿射轉換函數 $g_h(\cdot)$ ，肽 p^k 將藉由 $m=4$ 個不同經識別之HLA對偶基因中之HLA對偶基因 $h=2$ 、 $h=3$ 呈遞之可能性可藉由以下產生：

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

其中 x_2^k 、 x_3^k 為針對HLA對偶基因 $h=2$ 、 $h=3$ 之經識別之對偶基因相互作用變數，且 θ_2 、 θ_3 為針對HLA對偶基因 $h=2$ 、 $h=3$ 確定之參數集。

【0388】 作為另一實例，使用網路轉換函數 $g_h(\cdot)$ 、 $g_w(\cdot)$ ，肽 p^k 將藉由 $m=4$ 個不同經識別之HLA對偶基因中之HLA對偶基因 $h=2$ 、 $h=3$ 呈遞之可能性可藉由以下產生：

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

其中 $NN_2(\cdot)$ 、 $NN_3(\cdot)$ 為針對HLA對偶基因 $h=2$ 、 $h=3$ 之經識別之網路模型，且 θ_2 、 θ_3 為針對HLA對偶基因 $h=2$ 、 $h=3$ 確定之參數集。

IX.實例5：預測模組

【0389】 預測模組320使用呈遞模型接收序列資料並選擇該序列資

料中之候選新抗原。具體而言，序列資料可為自患者之腫瘤組織細胞提取之DNA序列、RNA序列及/或蛋白質序列。預測模組320將序列資料處理成針對MHC-I具有8至15個胺基酸或針對MHC-II具有6至30個胺基酸之複數個肽序列 p^k 。例如，預測模組320可將給定序列「IEFROEIFJEF」處理成具有9個胺基酸之三種肽序列「IEFROEIFJ」、「EFROEIFJE」及「FROEIFJEF」。於一實施例中，預測模組320可藉由將自患者之正常組織細胞提取之序列資料與自患者之腫瘤組織細胞提取之序列資料比較以識別含有一或多個突變之部分來識別候選新抗原，該等新抗原為突變的肽序列。

【0390】預測模組320將呈遞模型中之一或多者應用於經處理之肽序列以評估肽序列之呈遞可能性。具體而言，預測模組320可藉由將呈遞模型應用於候選新抗原選擇更可能在腫瘤HLA分子上呈遞之一或多個候選新抗原肽序列。於一實施中，預測模組320選擇評估在預定臨限值以上之呈遞可能性之候選新抗原序列。於另一實施中，該呈遞模型選擇具有最高預估呈遞可能性之 v 個候選新抗原序列(其中 v 一般為可於疫苗中遞送之抗原決定基之最大數目)。可將包含針對給定患者之選定之候選新抗原之疫苗注射至該患者中以誘導免疫反應。

X.實例6：患者選擇模組

【0391】患者選擇模組324基於患者是否滿足納入標準選擇用於疫苗治療及/或T-細胞療法之患者子集。於一實施例中，基於如藉由呈遞模型產生之患者新抗原候選之呈遞可能性確定該納入標準。藉由調整該納入標準，患者選擇模組324可基於其新抗原候選之呈遞可能性調整將接受疫苗及/或T-細胞療法之患者之數目。具體而言，嚴格納入標準導致將利用

疫苗及/或T-細胞療法治療之患者之更少數目，但是可導致接受有效治療(例如，1或多個腫瘤特異性新抗原(TSNA)及/或1或多個新抗原響應之T-細胞)之經疫苗及/或T-細胞療法治療之患者之更高比例。另一方面，寬鬆納入標準導致將利用疫苗及/或利用T-細胞療法治療之患者之更高數目，但是可導致接受有效治療之經疫苗及/或T-細胞療法治療之患者之更低比例。患者選擇模組324基於將接受治療之患者之目標比例與接受有效治療之患者之比例之間的所需平衡修改納入標準。

【0392】於一些實施例中，用於接受疫苗治療之患者之選擇之納入標準與用於接受T-細胞療法之患者之選擇之納入標準相同。然而，於替代實施例中，用於接受疫苗治療之患者之選擇之納入標準可與用於接受T-細胞療法之患者之選擇之納入標準不同。下列第X.A節及第X.B節各自討論用於接受疫苗治療之患者之選擇之納入標準及用於接受T-細胞療法之患者之選擇之納入標準。

X.A.針對疫苗治療之患者選擇

【0393】於一實施例中，患者係與 v 個新抗原候選之對應治療子集相關，該等新抗原候選可潛在地包含於針對具有疫苗容量 v 之患者之定製疫苗中。於一實施例中，患者之治療子集為具有如藉由呈遞模型所測定之最高呈遞可能性之新抗原候選。例如，若疫苗可包含 $v=20$ 個抗原決定基，則該疫苗可包含具有如藉由呈遞模型所測定之最高呈遞可能性之各患者之治療子集。然而，應瞭解於其他實施例中，患者之治療子集可基於其他方法確定。例如，患者之治療子集可自患者之新抗原候選組隨機選擇或可基於建模肽序列之結合親和力或穩定性之目前申請專利當時之技術水平模型或包含來自呈遞模型之呈遞可能性及關於彼等肽序列之親和力或穩定

性資訊之因素的一些組合部分確定。

【0394】於一實施例中，患者選擇模組324確定若患者之腫瘤突變負擔等於最小突變負擔或在最小突變負擔以上，則該患者滿足納入標準。患者之腫瘤突變負擔(TMB)指示腫瘤外顯子組中之非同義突變之總數目。於一實施中，若患者之TMB之絕對數目等於預定臨限值或在預定臨限值以上，則患者選擇模組324可選擇該患者用於疫苗治療。於另一實施中，若患者之TMB係於針對患者組測定之TMB中之臨限值百分位內，則患者選擇模組324可選擇該患者用於疫苗治療。

【0395】於另一實施例中，患者選擇模組324確定若基於患者之治療子集之患者之效用得分等於最小效用得分或在最小效用得分以上，則該患者滿足納入標準。於一實施中，該效用得分為呈遞來自治療子集之新抗原之預估數目的量測。

【0396】呈遞新抗原之預估數目可藉由將新抗原呈遞建模為一或多個概率分佈之隨機變數預測。於一實施中，患者i之效用得分為來自治療子集或其一些函數之呈遞新抗原候選之期望數目。作為實例，可將各新抗原之呈遞建模為伯努利(Bernoulli)隨機變數，其中呈遞(成功)概率藉由新抗原候選之呈遞可能性給定。具體而言，針對各者具有最高呈遞可能性 u_{i1} 、 u_{i2} 、...、 u_{iv} 之v個新抗原候選 p^{i1} 、 p^{i2} 、...、 p^{iv} 之治療子集 S_i ，新抗原候選 p^{ij} 之呈遞藉由隨機變數 A_{ij} 給定，其中：

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

呈遞新抗原之期望數目藉由合計各新抗原候選之呈遞可能性給定。換言之，可將患者i之效用得分表示為：

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

患者選擇模組324選擇具有等於最小功效或在最小功效以上之效用得分之患者子集用於疫苗治療。

【0397】 於另一實施中，患者*i*之效用得分為將呈遞至少臨限值數目之新抗原*k*之概率。於一實例中，將新抗原候選之治療子集*S_i*中之呈遞新抗原之數目建模為泊松二項(Poisson Binomial)隨機變數，其中呈遞(成功)概率藉由抗原決定基各者之呈遞可能性給定。具體而言，患者*i*之呈遞新抗原之數目可藉由隨機變數*N_i*給定，其中：

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

其中PBD(·)表示泊松二項分佈。將呈遞至少臨限值數目之新抗原*k*之概率藉由合計呈遞新抗原之數目*N_i*將等於*k*或在*k*以上之概率給定。換言之，可將患者*i*之效用得分表示為：

$$\text{util}_i(S_i) = \mathbb{P}[N_i \geq k] = \sum_{m=1}^k \mathbb{P}[N_i = m] \tag{27}$$

患者選擇模組324選擇具有等於最小功效或在最小功效以上之效用得分之患者子集用於疫苗治療。

【0398】 於另一實施中，患者*i*之效用得分為具有對患者之HLA對偶基因中之一或多者之固定臨限值(例如，500nM)以下之結合親和力或預測親和力之新抗原候選之治療子集*S_i*中之新抗原的數目。於一實例中，該固定臨限值為自1000 nM至10 nM之範圍。視情況，該效用得分可僅計算如經由RNA-seq表現之檢測到之彼等新抗原。

【0399】 於另一實施中，患者*i*之效用得分為具有對患者之HLA對偶基因中之一或多者之結合親和力在隨機肽對HLA對偶基因之結合親和力之臨限值百分位或臨限值百分位以下之新抗原候選之治療子集*S_i*中之新抗

原的數目。於一實例中，該臨限值百分位數為自第10個百分位至第0.1個百分位之範圍。視情況，該效用得分可僅計算如經由RNA-seq表現之檢測到之彼等新抗原。

【0400】應瞭解，關於方程式(25)及(27)說明之產生效用得分之實例僅係說明性，且患者選擇模組324可使用其他統計或概率分佈以產生效用得分。

X.B.針對T-細胞療法之患者選擇

【0401】於另一實施例中，代替或除了接受疫苗治療外，患者可接受T-細胞療法。如同疫苗治療，於患者接受T-細胞療法之實施例中，該患者可與如上所述之 v 個新抗原候選之對應治療子集相關。 v 個新抗原候選之此治療子集可用於活體外識別患者之T細胞，該患者對 v 個新抗原候選中之一或多者反應。然後可將此等經識別之T細胞擴展及融合至患者中用於定製之T-細胞療法。

【0402】可在兩個不同時間點選擇患者接受T-細胞療法。第一點係於針對使用模型之患者預測 v 個新抗原候選之治療子集後，但是在對 v 個新抗原候選之預測治療子集特異性之T細胞之活體外篩選之前。第二點係於對 v 個新抗原候選之預測治療子集特異性之T細胞之活體外篩選後。

【0403】首先，可於針對患者預測 v 個新抗原候選之治療子集後，但是在對 v 個新抗原候選之預測子集特異性之患者之T細胞的活體外識別之前選擇患者接受T-細胞療法。具體而言，因為來自患者之新抗原特異性T細胞之活體外篩選可係昂貴的，所以若患者更可能具有新抗原特異性T細胞，則可期望僅選擇篩選新抗原特異性T細胞之患者。為在活體外T-細胞篩選步驟之前選擇患者，可使用用於選擇用於疫苗治療之患者之相同標

準。具體而言，於一些實施例中，若患者之腫瘤突變負擔等於如上所述之最小突變負擔或在最小突變負擔以上，則患者選擇模組324可選擇該患者接受T-細胞療法。於另一實施例中，若基於患者之 v 個新抗原候選之治療子集之患者之效用得分等於如上所述之最小效用得分或在最小效用得分以上，則該患者選擇模組324可選擇該患者接受T-細胞療法。

【0404】 其次，除了或代替在對 v 個新抗原候選之預測子集特異性之患者之T細胞的活體外識別之前選擇患者接受T-細胞療法，亦可於對 v 個新抗原候選之預測治療子集特異性之T細胞的活體外識別後選擇患者接受T-細胞療法。具體而言，若在用於新抗原識別之患者之T-細胞之活體外篩選期間，針對患者識別至少臨限值數量之新抗原特異性TCR，則可選擇該患者接受T-細胞療法。例如，只有在針對患者識別至少兩個新抗原特異性TCR之情況下，或只有在針對兩個不同新抗原識別新抗原特異性TCR之情況下，可選擇該患者接受T-細胞療法。

【0405】 於另一實施例中，只有在藉由患者之TCR識別患者之 v 個新抗原候選之治療子集之至少臨限值數量之新抗原的情況下，可選擇該患者接受T-細胞療法。例如，只有在藉由患者之TCR識別患者之 v 個新抗原候選之治療子集之至少一個新抗原的情況下，可選擇該患者接受T-細胞療法。於其他實施例中，只有在患者之至少臨限值數量之TCR經識別為對特定HLA限制類別之新抗原肽具有新抗原特異性的情況下，可選擇該患者接受T-細胞療法。例如，只有在患者之至少一個TCR經識別為新抗原特異性HLA I類限制之新抗原肽的情況下，可選擇該患者接受T-細胞療法。

【0406】 於甚至其他實施例中，只有在藉由患者之TCR識別至少臨限值數量之特定HLA限制類別之新抗原肽的情況下，可選擇該患者接受

T-細胞療法。例如，只有在藉由患者之TCR識別至少一個HLA I類限制之新抗原肽的情況下，可選擇該患者接受T-細胞療法。作為另一實例，只有在藉由患者之TCR識別至少兩個HLA II類限制之新抗原肽的情況下，可選擇該患者接受T-細胞療法。以上標準之任何組合亦可用於對患者之 v 個新抗原候選之預測治療子集特異性之T-細胞的活體外識別後選擇患者接受T-細胞療法。

XI.實例7：顯示實例患者選擇表現之實驗結果

【0407】 第X節中所述之患者選擇方法的有效性藉由對模擬患者集進行患者選擇來測試，該患者各者與模擬新抗原候選之測試集相關，其中已知模擬新抗原之子集待以質譜資料呈遞。具體而言，測試集中之各模擬新抗原候選係與標籤相關，該標籤指示新抗原是否於來自Bassani-Sternberg資料集（資料集「D1」）（資料可見於www.ebi.ac.uk/pride/archive/projects/PXD0000394中）之多對偶基因JY細胞株HLA-A*02:01及HLA-B*07:02質譜資料集中呈遞。如下結合圖13A更詳細所述，基於非小細胞肺癌(NSCLC)患者中之突變負擔之已知頻率分佈，模擬患者之多個新抗原候選係自人類蛋白質組取樣。

【0408】 使用訓練集訓練相同HLA對偶基因之按對偶基因呈遞模型，該訓練集為來自IEDB資料集（資料集「D2」）（資料可見於http://www.iedb.org/doc/mhc_ligand_full.zip中）之單對偶基因HLA-A*02:01及HLA-B*07:02質譜資料之子集。具體而言，各對偶基因之呈遞模型為方程式(8)中所示之按對偶基因模型，該方程式利用網路相依函數 $g_h(\cdot)$ 及 $g_w(\cdot)$ 及場外函數 $f(\cdot)$ 併入N-端及C-端側接序列作為對偶基因非相互作用變數。對偶基因HLA-A*02:01之呈遞模型產生給定肽將在對偶基因

HLA-A*02:01上呈遞之呈遞可能性、給定肽序列作為對偶基因相互作用變數及N-端及C-端側接序列作為對偶基因非相互作用變數。對偶基因HLA-B*07:02之呈遞模型產生給定肽將在對偶基因HLA-B*07:02上呈遞之呈遞可能性、給定肽序列作為對偶基因相互作用變數及N-端及C-端側接序列作為對偶基因非相互作用變數。

【0409】 如下列實例及參考圖13A至13E所述，將各種模型(諸如用於肽結合預測之訓練呈遞模型及目前申請專利當時之技術水平模型)應用於各模擬患者之新抗原候選之測試集以基於該等預測識別患者之不同治療子集。滿足納入標準之患者經選擇用於疫苗治療，且與包含患者之治療子集中之抗原決定基的定製疫苗相關。治療子集之大小根據不同疫苗容量變化。在用於訓練呈遞模型之訓練集與類比新抗原候選之測試集之間不引入重疊。

【0410】 於下列實例中，分析具有包含於疫苗中之抗原決定基中之至少一定數目之呈遞新抗原之選定患者的比例。此統計指示模擬疫苗遞送將引起患者之免疫反應之潛在新抗原的有效性。具體而言，若於質譜資料集D2中呈遞新抗原，則呈遞測試集中之模擬新抗原。具有呈遞新抗原之高比例患者指示經由新抗原疫苗藉由誘導免疫反應成功治療之潛力。

XI.A.實例7A：NSCLC癌症患者之突變負擔之頻率分佈

【0411】 圖13A說明NSCLC患者之突變負擔之樣品頻率分佈。可(例如)在癌症基因組圖譜(TCGA) (<https://cancergenome.nih.gov>)發現突變負擔及不同腫瘤類型(包括NSCLC)之突變。x-軸表示各患者中之非同義突變之數目，且y-軸表示具有給定數目之非同義突變之樣品患者之比例。圖13A中之樣品頻率分佈顯示3至1786個突變之範圍，其中30%之患者具

有少於100個突變。雖然於圖13A中尚未顯示，但是研究指示突變負擔於吸煙者中相較於非吸煙者中係更高，及突變負擔可為負載於患者中之新抗原之強指標。

【0412】 如在以上第XI節之開始所介紹，將許多模擬患者各者與新抗原候選之測試集相關聯。各患者之測試集藉由自針對各患者之圖13A中所示之頻率分佈對突變負擔 m_i 取樣產生。針對各突變，隨機選擇來自人類蛋白質組之21-mer肽序列以表示模擬突變序列。針對患者 i 之新抗原候選序列之測試集藉由識別跨21-mer之突變之各(8, 9, 10, 11)-mer肽序列產生。各新抗原候選係與標籤相關，該標籤指示新抗原候選序列是否存在於質譜D1資料集中。例如，存在於資料集D1中之新抗原候選序列可與標籤「1」相關，而不存在於資料集D1中之序列可與標籤「0」相關。如下更詳細所述，圖13B至13E說明基於測試集之患者之經呈遞之新抗原之患者選擇的實驗結果。

XI.B.實例7B：具有基於突變負擔納入標準之新抗原呈遞之選定患者的比例

【0413】 圖13B說明針對基於患者是否滿足最小突變負擔之納入標準選擇之患者之模擬疫苗中之經呈遞之新抗原的數目。識別於對應測試中具有至少一定數量之經呈遞之新抗原之選擇之患者的比例。

【0414】 於圖13B中， x -軸指示自基於最小突變負擔之疫苗治療排除之患者之比例，如藉由標籤「突變之最小數目」所指示。例如，「突變之最小數目」200處之資料點指示患者選擇模組324僅選擇具有至少200個突變之突變負擔之模擬患者之子集。作為另一實例，「突變之最小數目」300處之資料點指示患者選擇模組324選擇具有至少300個突變之較低比例

之模擬患者。y-軸指示與測試集中之至少一定數量之經呈遞之新抗原相關而無任何疫苗容量 v 之選定患者之比例。具體而言，頂部圖顯示呈遞至少1個新抗原之選定患者之比例，中間圖顯示呈遞至少2個新抗原之選定患者之比例，及底部圖顯示呈遞至少3個新抗原之選定患者之比例。

【0415】如圖13B中所指示，具有經呈遞之新抗原之選定患者之比例隨著更高突變負擔顯著增加。此指示作為納入標準之突變負擔可有效用於選定患者，新抗原疫苗對該等患者更可能成功誘導免疫反應。

XI.C.實例7C：藉由呈遞模型識別之疫苗之新抗原呈遞相對於藉由申請專利當時之技術水平模型識別之疫苗之新抗原呈遞的比較

【0416】圖13C比較與包含基於呈遞模型識別之治療子集之疫苗相關之選定患者與與包含通過目前申請專利當時之技術水平模型識別之治療子集之疫苗相關之選定患者之間之模擬疫苗中之經呈遞之新抗原的數目。左圖假設受限疫苗容量 $v=10$ ，及右圖假設受限疫苗容量 $v=20$ 。基於指示經呈遞之新抗原之預期數目之效用得分選定患者。

【0417】於圖13C中，實線指示與包括基於對偶基因HLA-A*02:01及HLA-B*07:02之呈遞模型識別之治療子集之疫苗相關之患者。針對各患者之治療子集藉由將呈遞模型各者應用於測試集中之序列及識別具有最高呈遞可能性之 v 個新抗原候選來識別。虛線指示與包含基於單對偶基因HLA-A*02:01之目前申請專利當時之技術水平模型NETMHCpan識別之治療子集之疫苗相關之患者。在<http://www.cbs.dtu.dk/services/NetMHCpan>詳細提供NETMHCpan之實施細節。針對各患者之治療子集藉由將NETMHCpan模型應用於測試集中之序列及識別具有最高預估結合親和力之 v 個新抗原候選來識別。兩個

圖之x-軸指示基於預期效用得分自疫苗治療排除之患者之比例，該等評分指示基於呈遞模型識別之治療子集中之經呈遞之新抗原之預期數目。如參考第X節中之方程式(25)所述測定預期效用得分。y-軸指示呈遞包含於疫苗中之至少一定數量之新抗原(1、2或3個新抗原)之選定患者的比例。

【0418】如圖13C中所指示，與包含基於呈遞模型之治療子集之疫苗相關之患者較與包含基於申請專利當時之技術水平模型之治療子集之疫苗相關之患者以顯著更高速率接受含有經呈遞之新抗原之疫苗。例如，如右圖中所示，與基於呈遞模型之疫苗相關之80%之選定患者相較於與基於目前申請專利當時之技術水平模型之疫苗相關之僅40%之選定患者接受疫苗中之至少一種經呈遞之新抗原。結果指示如本文中所述之呈遞模型有效用於選擇可能引起治療腫瘤之免疫反應之疫苗之新抗原候選。

XI.D.實例7D：HLA覆蓋度對通過呈遞模型識別之疫苗之新抗原呈遞的影響

【0419】圖13D比較與包含基於HLA-A*02:01之單按對偶基因呈遞模型識別之治療子集之疫苗相關之選定患者與與包含基於HLA-A*02:01及HLA-B*07:02之按對偶基因呈遞模型二者識別之治療子集之疫苗相關之選定患者之間之模擬疫苗中之經呈遞之新抗原的數目。將疫苗容量設置為 $v=20$ 個抗原決定基。針對各實驗，基於不同治療子集測定之預期效用得分選擇患者。

【0420】於圖13D中，實線指示與包含基於HLA對偶基因HLA-A*02:01及HLA-B*07:02之呈遞模型二者之治療子集之疫苗相關之患者。針對各患者之治療子集藉由將呈遞模型各者應用於測試集中之序列及識別具有最高呈遞可能性之 v 個新抗原候選來識別。虛線指示與包含基於HLA

對偶基因HLA-A*02:01之單呈遞模型之治療子集之疫苗相關之患者。針對各患者之治療子集藉由將僅單HLA對偶基因之呈遞模型應用於測試集中之序列及識別具有最高呈遞可能性之 v 個新抗原候選來識別。針對實線圖， x -軸指示基於藉由兩種呈遞模型識別之治療子集之預期效用得分自疫苗治療排除之患者的比例。針對虛線圖， x -軸指示基於藉由單呈遞模型識別之治療子集之預期效用得分自疫苗治療排除之患者的比例。 y -軸指示呈遞至少一定數量之新抗原(1、2或3個新抗原)之選定患者的比例。

【0421】如圖13D中所指示，與包含藉由針對HLA對偶基因經呈遞之新抗原二者之呈遞模型識別之治療子集之疫苗相關之患者較與包含藉由單呈遞模型識別之治療子集之疫苗相關之患者在顯著更高速率下經呈遞之新抗原。結果指示建立具有高HLA對偶基因覆蓋度之呈遞模型之重要性。

XI.E.實例7E：藉由突變負擔選擇之患者之新抗原呈遞相對於藉由經呈遞之新抗原之預期數目選擇之患者之新抗原呈遞的比較

【0422】圖13E比較基於突變負擔選擇之患者與藉由預期效用得分選擇之患者之間之模擬疫苗中之經呈遞之新抗原的數目。基於藉由具有 $v=20$ 個抗原決定基之大小之呈遞模型識別之治療子集測定預期效用得分。

【0423】於圖13E中，實線指示基於與包含藉由呈遞模型識別之治療子集之疫苗相關之預期效用得分選擇的患者。針對各患者之治療子集藉由將呈遞模型應用於測試集中之序列及識別具有最高呈遞可能性之 $v=20$ 個新抗原候選來識別。基於第X節中之方程式(25)之治療子集識別之呈遞可能性測定預期效用得分。虛線指示基於與亦包含藉由呈遞模型識別之治療子集之疫苗相關之突變負擔選擇的患者。 x -軸指示基於針對實線圖之預期

效用得分自疫苗治療排除之患者的比例，及基於針對虛線圖之突變負擔排除之患者的比例。y-軸指示接受含有至少一定數量之經呈遞之新抗原(1、2或3個新抗原)之疫苗之選定患者的比例。

如圖13E中所指示，基於預期效用得分選擇之患者較基於突變負擔選擇之患者在更高速率下接受含有經呈遞之新抗原之疫苗。然而，基於突變負擔選擇之患者較未經選擇之患者在更高速率下接受含有經呈遞之新抗原之疫苗。因此，突變負擔為用於成功新抗原疫苗治療之有效患者選擇標準，儘管預期效用得分係更有效。

XII.實例8：質譜訓練模型對留存質譜資料之評價

【0424】因為藉由腫瘤細胞之HLA肽呈遞為抗腫瘤免疫^{91,96,97}之關鍵需求，以便使用此等及公開可得資料^{92,98,99}產生具有成對I類HLA肽序列、HLA類型及轉錄組RNA-seq (方法)之人類腫瘤及正常組織樣品之大的(N=74例患者)整合資料集以訓練新穎深度學習模型¹⁰⁰以預測人類癌症中之抗原呈遞。在免疫療法發展所關注之若干腫瘤類型中及基於組織可得性選擇樣品。質譜法在肽水平FDR<0.1 (範圍344-11,301)下識別平均3,704個肽/樣品。該等肽遵循特徵I類HLA長度分佈：長度8至15個aa，具有9之模態長度(56%之肽)。與先前報導一致，預測大多數肽(中值79%)在藉由MHCflurry⁹⁰之標準500 nM親和力臨限值下結合至少一例患者HLA對偶基因，但是具有跨樣品之實質上可變性(例如，一個樣品中之33%之肽具有經預測之親和力>500nM)。50 nM之常用¹⁰¹「強結合劑」臨限值捕獲中值僅42%之經呈遞肽。轉錄組定序產生平均131M獨特閱讀/樣品及68%之基因在至少一個樣品中之至少1個轉錄本/百萬(TPM)之水平下表現，其強調大及不同樣品集之值以觀察最大數目之基因表現。藉由HLA之肽呈遞

與mRNA表現強烈相關。觀察到超出可單獨藉由RNA表現或序列之差異解釋之肽呈遞率之顯著且可複製基因間差異。觀察到之HLA類型匹配來自患者之歐洲血統組佔優勢之樣品之期望。

【0425】 使用此等及公開可得HLA肽資料^{92,98,99}訓練神經(NN)模型以預測HLA抗原呈遞。為自腫瘤質譜資料學習對偶基因-特異性模型，其中各肽可藉由六個HLA對偶基因中之任一者呈遞，開發能共同學習對偶基因-肽圖譜分析及對偶基因-特異性呈遞模體(方法)之新穎網路架構。針對各患者，經陽性標記之資料點為經由質譜法檢測之肽，及經陰性標記之資料點為來自該樣品中不經由質譜法檢測之參考蛋白質組(SwissProt)之肽。將資料分成訓練、驗證及測試集(方法)。該訓練集由來自101個樣品(69個於此研究中新近描述及32個先前公開)之142,844個經HLA呈遞之肽(FDR<~0.02)組成。該驗證集(用於早期停止)由來自相同101個樣品之18,004個經呈遞之肽組成。將兩個質譜資料集用於測試：(1)腫瘤樣品測試集，其由來自從訓練資料中剔除之5個額外腫瘤樣品(2個肺、2個結腸、1個卵巢)之571個經呈遞之肽組成，及(2)單對偶基因細胞株測試集，其由來自鄰近但是區別於包含於訓練資料中之單對偶基因肽之位置之基因組定位窗(區塊)之2,128個經呈遞的肽組成(參見關於訓練/測試劃分之另外細節的方法)。

【0426】 該訓練資料識別針對53個HLA對偶基因之預測模型。與先前工作相反^{92,104}，此等模型捕獲HLA呈遞對多個長度之肽之各序列位置的相依性。該模型亦利用mRNA豐度正確學習對基因RNA表現及基因特異性呈遞傾向之臨界相依性及學習獨立組合以產生最低表現之最少呈遞傾向與最高表現之最多呈遞傾向基因之間之呈遞率之至多~60倍差異之呈遞

的每基因傾向。另外觀察到該模型預測IEDB⁸⁸中之HLA/肽複合體之量測之穩定性(針對10個對偶基因， $p < 1e-10$)，甚至於針對經預測之結合親和力之控制後(針對所測試之8/10個對偶基因， $p < 0.05$)。共同地，此等特徵形成改善之免疫原性HLA I類肽之預測之基礎。

【0427】 評價此NN模型作為HLA呈遞之預測因子在保留質譜測試集上之表現。具體而言，圖14比較當在五種不同保留測試樣品上測試各模型時，在MS模型及建模自質譜之溶離肽之最近公開之方法(MixMHCPred)之不同版本之40%召回下的陽性預測值(PPV)。圖14亦描述針對五種測試樣品之模型之40%召回下的平均PPV。

【0428】 圖14中所測試之模型為(自左至右)：「全MS模型」：於方法中所述之全NN模型；「MS模型，無側接序列」：與全NN模型相同，除了移除側接序列特徵外；「MS模型，無側接序列或每基因參數」：與全NN模型相同，除了移除側接序列及每基因參數特徵外；「僅肽MS模型，聯合訓練之所有長度」：與全NN模型相同，除了僅使用之特徵為肽序列及HLA類型外；「僅肽MS模型，分開訓練之各長度」：針對此模型，該模型架構與僅肽MS模型相同，除了訓練針對9及10mer之分開模型外；「線性僅肽MS模型(利用系集)」：與具有分開訓練之各肽長度之僅肽MS模型相同；除了代替使用神經網路建模肽序列外，使用用於全模型之相同最佳化程序訓練及於方法中所述之線性模型之系集；「MixMHCPred 1.1」為具有默認設置之MixMHCPred；「結合親和力」為MHCflurry 1.2.0。

【0429】 「全MS模型」、「MS模型，無側接序列」、「MS模型，無側接序列或每基因參數」、「僅肽MS模型，聯合訓練之所有長

度」、「僅肽MS模型，分開訓練之所有長度」及「線性僅肽MS模型」為在如上所述之質譜資料上訓練之所有神經網路模型。然而，使用樣品之不同特徵訓練及測試各模型。「MixMHCpred 1.1」模型及「結合親和力」模型為建模經HLA呈遞之肽之早期方法¹⁰⁴。於比較中使用僅9及10mer，因為MixMHCpred目前不建模除了9及10之長度之肽。後5種模型(通過「結合親和力」之僅肽MS模型，聯合訓練之所有長度)具有相同輸入：僅肽序列及HLA類型。特定言之，後5種模型中無一者使用RNA豐度作出預測。

【0430】表現最佳之僅肽模型(「僅肽MS模型，聯合訓練之所有長度」)達成40%召回下之0.41之平均PPV，而在質譜資料上訓練之表現最差之僅肽模型(「線性僅肽MS模型」)達成僅28%之平均PPV(僅稍微高於18%下之MixMHCpred 1.1之平均PPV)，其強調改良之肽序列之NN建模的值。注意MixMHCpred 1.1在與線性僅肽MS模型不同之資料上訓練，但是具有許多相同建模特徵(例如，其為線性模型，其中分開訓練針對各肽長度之模型)。

【0431】總之，NN模型達成顯著改善之HLA肽呈遞之預測，具有高於對腫瘤測試集之標準結合親和力+基因表現至多9倍之PPV。基於MS之NN模型之大的PPV優勢堅持跨各種召回臨限值及係統計上顯著(針對所有腫瘤樣品， $p < 10^{-6}$)。針對HLA肽呈遞之標準結合親和力+基因表現之陽性預測值達到低至6%，與先前評估^{87,93}一致。應注意的是，然而，此~6% PPV仍表示超過基線流行之>100倍濃化，因為僅檢測到經呈遞之小比例之肽(例如，腫瘤MS測試資料集中之~1/2500)。

【0432】藉由比較在僅使用HLA類型及肽序列作為全MS模型之輸

入之質譜資料上訓練之減少的模型，確定超過結合親和力預測之PPV之~30%之增加來自建模肽外在特徵(RNA豐度、側接序列、每基因參數)，該等特徵可利用質譜法但非結合親和力檢定捕獲。增加之其他~70%來自改良之肽序列之建模。不僅為訓練資料集(經HLA呈遞之肽)之性質，而且有助於改善之性能之整體模型架構，因為其亦勝過建模人類腫瘤中之經HLA呈遞之肽¹⁰⁴的早期方法。新模型架構使能經由端至端訓練過程學習對偶基因特異性模型，該過程不需要使用結合親和力預測或困難群集方法之肽至聲稱之呈遞對偶基因之事前分配¹⁰⁴⁻¹⁰⁶。重要的是，其亦避免對對偶基因特異性子模型施加精度降低限制作為去卷積(諸如線性或各肽長度之分開考慮¹⁰⁴)之先決條件。全模型勝過若干簡化模型及施加此等限制之經先前公開之方法。

XIII.實例9：誘導呈遞熱點建模之實驗結果

【0433】 為具體評價於建模HLA呈遞中使用呈遞熱點參數之效益，將併入呈遞熱點參數之神經網路呈遞模型之表現與不併入呈遞熱點參數之神經網路呈遞模型之表現相比。基礎神經網路架構針對兩種模型係相同及與以上第VII節中所述之呈遞模型相同。簡言之，該等模型包含肽及側接胺基酸序列參數、RNA定序轉錄資料(TPM)、蛋白質家族資料、每樣品識別及HLA-A、B、C類型。針對各模型使用5個網路系集。包含呈遞熱點參數之模型使用以上第VIII.B.3節中所述之方程式12c，具有10之每基因蛋白質組區塊大小及肽長度8至12。

【0434】 藉由進行實驗使用以上第XII節中所述之質譜資料集比較該兩種模型。具體而言，出於公平評價競爭模型之目的，自模型訓練及驗證保留五種樣品。將其餘樣品隨機分成用於模型訓練之90%及用於驗證該

訓練之10%。

【0435】圖15A比較當在五種保留測試樣品上測試該等模型時，跨使用呈遞熱點參數之呈遞模型及不使用呈遞熱點參數之呈遞模型之召回之平均陽性預測值(PPVs)。併入呈遞熱點參數之模型勝過在個別樣品各者上不併入呈遞熱點參數之模型，含有呈遞熱點參數之平均精度為0.82及不含呈遞熱點參數之平均精度為0.77。

【0436】圖15B至F比較當在五種保留測試樣品各者上測試該等模型時，使用呈遞熱點參數之呈遞模型及不使用呈遞熱點參數之呈遞模型之精度及召回曲線。

XIV.實例10：用於識別T-細胞抗原決定基之呈遞熱點參數之評價

【0437】亦直接測試使用呈遞熱點參數建模HLA呈遞以識別人類腫瘤CD8 T-細胞抗原決定基(即，免疫療法靶)之效益。定義用於此評價之適宜測試資料集係挑戰性，因為該測試資料集應含有藉由T細胞識別及藉由腫瘤細胞表面上之HLA呈遞二者之肽。此外，正規表現評估不僅要求經陽性標記之(即，經T-細胞識別之)肽，而且要求足夠數目之經陰性標記之(即，經測試之但是未經識別之)肽。質譜資料集定址腫瘤呈遞，但是非T-細胞識別；相反，引發或接種後之T-細胞檢定定址T-細胞識別但是非腫瘤呈遞。

【0438】為獲得適宜資料集，吾人收集經公開之來自滿足所需標準之5個最近研究之CD8 T-細胞抗原決定基：研究A⁹⁶檢查患有胃腸腫瘤之9例患者中之TIL並報導於自體DC中使用串聯微基因(TMG)方法藉由IFN- γ ELISPOT測試之12/1,053體細胞SNV突變之T-細胞識別。研究B⁸⁴亦使用TMG並報導6/574 SNV藉由來自5個黑色素瘤患者之CD8+PD-1+循環淋巴

細胞之T-細胞識別。研究C⁹⁷使用脈衝肽刺激評估來自3個黑色素瘤患者之TIL及發現對5/381經測試之SNV突變之反應。研究D¹⁰⁸使用TMG檢定與具有最小抗原決定基肽之脈衝之組合評估來自乳癌患者之TIL並報導2/62 SNV之識別。研究E¹⁶⁰評估來自國家癌症研究所(National Cancer Institute)之具有52 TSNA之17例患者中之TIL。經組合之資料集包含來自33例患者之4,843個經檢定之SNV，該等患者包含具有先已存在之T-細胞反應之75 TSNA。重要的是，因為該資料集主要包括藉由腫瘤浸潤之淋巴細胞之新抗原識別，對此資料集之成功預測證實該模型具有不僅識別能引發如先前節中之T細胞之新抗原，而且識別藉由腫瘤呈遞至T細胞之新抗原的能力。

【0439】 為模擬用於個人化免疫療法之抗原之選擇，使用兩種方法將體細胞突變以呈遞概率之順序排名：(1)包含熱點特徵(如具有區塊大小 $n=10$ 之方程式12c中所述)之MS模型，及(2)無熱點特徵之傳統MS模型。因為抗原特異性免疫療法之容量受限於靶向特異性之數目(例如，目前個人化疫苗編碼~10至20個突變^{6, 81-82})，預測方法藉由計算針對各患者之前5、10、20或30位肽中之先已存在之T-細胞反應來比較。圖16中描述結果。

【0440】 具體而言，圖16比較針對藉由使用呈遞熱點參數之呈遞模型及藉由不使用呈遞熱點參數之呈遞模型識別之前5、10、20及30位肽，針對包含取自具有至少一個先已存在之T-細胞反應之患者之測試樣品的測試集，藉由T-細胞識別之跨體細胞突變之肽之比例。如圖16中所說明，具有熱點特徵之模型與無該特徵之模型表現相當，其中兩種模型各自預測於前20及10位肽中之45及31個T-細胞反應。然而，當預測前30及前5位肽

時，熱點模型顯示改善，其中該熱點模型各自包含6及4個更多T-細胞反應。

XIII.A.資料

【0441】吾人自 Gros 等人⁸⁴、Tran 等人¹⁴⁰、Stronen 等人¹⁴¹、Zacharakis 等人及 Koşaloğlu-Yalçın 等人¹⁶⁰之補充資訊獲得突變細胞、HLA 類型及 T-細胞識別資料。

【0442】針對突變-水平分析(圖 16)，Gros 等人、Tran 等人、Zacharakis 等人¹⁰⁸及 Koşaloğlu-Yalçın 等人¹⁶⁰之經陽性標記之資料點為藉由 TMG 檢定或最小抗原決定基肽-脈衝檢定二者中之患者 T-細胞識別之突變。經陰性標記之資料點為於 TMG 檢定中所測試之所有其他突變。針對 Stronen 等人，經陽性標記之突變為藉由至少一個經識別之肽跨越之突變，及陰性資料點為於四聚體檢定中測試但未識別之所有突變。針對 Gros、Tran 及 Zacharakis 資料，將突變藉由總計呈遞概率或取跨所有突變-跨越肽之最小結合親和力排名，因為突變之 25mer TMG 檢定測試跨越該突變之所有肽之 T-細胞識別。針對 Stronen 資料，將突變藉由總計呈遞概率或取跨四聚體檢定中所測試之所有突變-跨越肽之最小結合親和力排名。突變及特徵之完整清單於補充表 1 中可得。

【0443】針對抗原決定基-水平分析，經陽性標記之資料點為藉由肽脈衝或四聚體檢定中之患者 T-細胞識別之所有最小抗原決定基，及陰性資料點為未藉由肽-脈衝或四聚體檢定中之 T-細胞識別之所有最小抗原決定基及來自未藉由患者 T-細胞識別之經測試之 TMG 之所有突變-跨越肽。於 Gros 等人、Tran 等人及 Zacharakis 等人之情況下，自分析移除未經由肽-脈衝檢定測試之 TMG 分析中識別之最小抗原決定基肽跨越突變，因為未

實驗上測定此等肽之T-細胞識別狀態。

XV.實例11：癌症患者中之新抗原反應性T-細胞之識別

【0444】 此實例證實改善之預測可使自常規患者樣品之新抗原識別。為做到此，自患有轉移性NSCLC之經歷抗-PD(L)1療法之9例患者分析存檔FFPE腫瘤活組織檢查及5至30ml外周血(補充表2：圖17A至C中研究之N=9例患者之患者人口統計學及治療資訊。關鍵區域包括腫瘤階段及亞型、接受之抗-PD1療法及NGS結果之概述)。腫瘤全外顯子組定序、腫瘤轉錄組定序及經匹配之正常外顯子組定序導致平均198個體細胞突變/患者(SNV及短插入或缺失)，其中平均118個表現(方法，補充表2)。將全MS模型應用於優先考慮之20個新抗原決定基/患者用於對先已存在之抗腫瘤T-細胞反應測試。為將分析集中在可能CD8反應上，合成優先考慮之肽作為8-11mer最小抗原決定基(方法)，及然後利用合成肽培養外周血單核細胞(PBMC)簡而言之活體外刺激(IVS)培養以擴展新抗原-反應性T-細胞(補充表3)。於兩周後，使用針對優先考慮之新抗原決定基之IFN- γ ELISpot評估抗原特異性T-細胞之存在。於充足PBMC係可得之7例患者中，亦進行分開實驗以完全或部分去卷積經識別之特異性抗原。圖17A至C及18A至21中描述結果。

【0445】 圖17A描述對9例患者之患者特異性新抗原肽池之T-細胞反應的檢測。針對各患者，將經預測之新抗原組合至10個肽之2個池中，各者根據模型排名及任何序列相同性(將相同肽分開至不同池中)。然後，針對各患者，將患者之活體外擴展之PBMC用IFN- γ ELISpot中之2例患者-特異性新抗原肽池刺激。將圖17A中之數據表示為斑點形成單元(SFU)/ 10^5 個經平板接種之細胞，其中減去背景(對應DMSO陰性對照)。

圖21中顯示背景量測(DMSO陰性對照)。針對患者1-038-001、1-050-001、1-001-002、CU04、1-024-001、1-024-002及CU05顯示單孔(患者1-038-001、CU02、CU03及1-050-001)或具有平均及標準差之複製(所有其他患者)對同源肽池#1及#2之反應。針對患者CU02及CU03，細胞數目允許僅對特異性肽池#1測試。認為具有值>背景以上之2倍增加之樣品係陽性及用星星指定(反應性供體包括患者1-038-001、CU04、1-024-001、1-024-002及CU02)。無反應供體包括患者1-050-001、1-001-002、CU05及CU03。圖17C描述具有自於IFN- γ ELISpot中利用DMSO陰性對照、PHA陽性對照、CU04特異性新抗原肽池#1、CU04特異性肽1、CU04特異性肽6及CU04特異性肽8刺激之患者CU04之活體外擴展之PBMC之ELISpot孔的照片。

【0446】 圖18A至B描述來自利用HLA匹配之健康供體中之患者新抗原之對照實驗之結果。此等實驗之結果證實活體外培養條件僅擴展先已存在之活體內引發之記憶T-細胞，而非使能重新活體外引發。

【0447】 圖19描述針對各供體及圖17A中所述之各活體外擴展之PHA陽性對照之T-細胞反應的檢測。針對各供體及圖17A中之各活體外擴展，將經活體外擴展之患者PBMC用PHA刺激用於最大T-細胞激活。將圖19中之數據表示為斑點形成單元(SFU)/ 10^5 個經平板接種之細胞，其中減去背景(對應DMSO陰性對照)。針對患者1-038-001、1-050-001、1-001-002、CU04、1-024-001、1-024-002、CU05及CU03顯示單孔或生物複製之反應。針對患者CU02不進行利用PHA之測試。將來自患者CU02之細胞包含至分析中，作為指示活性及功能T-細胞之對肽池#1之陽性反應(圖17A)。如圖17A中所示，對肽池具反應性之供體包括患者1-038-001、

CU04、1-024-001及1-024-002。亦如圖17A中所示，對肽池不具反應性之供體包括患者1-050-001、1-001-002、CU05及CU03。

【0448】 圖20A描述對患者CU04之池#2中之各個別患者特異性新抗原肽之T-細胞反應的檢測。圖20A亦描述對患者CU04之PHA陽性對照之T-細胞反應的檢測。(此為陽性對照資料，亦顯示於圖19中。)針對患者CU04，將該患者之經活體外擴展之PBMC於IFN- γ ELISpot中用來自患者CU04之池#2之患者特異性個別新抗原肽刺激。亦將該患者之經活體外擴展之PBMC於IFN- γ ELISpot中用PHA刺激作為陽性對照。將數據表示為斑點形成單元(SFU)/ 10^5 個經平板接種之細胞，其中減去背景(對應DMSO陰性對照)。

【0449】 圖20B描述對患者CU04之三次訪問各者及對患者1-024-002之兩次訪問各者(各訪問發生在不同時間點)之個別患者特異性新抗原肽之T-細胞反應的檢測。針對兩例患者，將該患者之經活體外擴展之PBMC於IFN- γ ELISpot中用患者特異性個別新抗原肽刺激。針對各患者，將各訪問之數據表示為累積(附加)斑點形成單元(SFU)/ 10^5 個經平板接種之細胞，其中減去背景(對應DMSO對照)。將患者CU04之資料顯示為減去背景之自3次訪問之累積SFU。針對患者CU04，針對首次訪問(T0)及於首次訪問(T0)後2個月(T0 + 2個月)及14個月(T0 + 14個月)之隨後訪問顯示減去背景之SFU。將患者1-024-002之資料顯示為減去背景之自2次訪問之累積SFU。針對患者1-024-002，針對首次訪問(T0)及於首次訪問(T0)後1個月(T0 + 1個月)之隨後訪問顯示減去背景之SFU。認為具有值>背景以上之2倍增加之樣品係陽性及用星星指定。

【0450】 圖20C描述對患者CU04之兩次訪問各者及對患者1-024-

002之兩次訪問各者(各訪問發生在不同時間點)之個別患者特異性新抗原肽及患者特異性新抗原肽池之T-細胞反應的檢測。針對兩例患者，將該患者之經活體外擴展之PBMC於IFN- γ ELISpot中用患者特異性個別新抗原肽以及患者特異性新抗原肽池刺激。具體而言，針對患者CU04，將患者CU04之經活體外擴展之PBMC於IFN- γ ELISpot中用CU04-特異性個別新抗原肽6及8以及用CU04-特異性新抗原肽池刺激，及針對患者1-024-002，將患者1-024-002之經活體外擴展之PBMC於IFN- γ ELISpot中用1-024-002-特異性個別新抗原肽16以及用1-024-002-特異性新抗原肽池刺激。將圖20C之數據表示為斑點形成單元(SFU)/ 10^5 個經平板接種之細胞，其中減去背景(對應DMSO對照)用於具有平均及範圍之各技術複製。將患者CU04之資料顯示為減去背景之自2次訪問之SFU。針對患者CU04，針對首次訪問(T0；技術一式三份)及於首次訪問(T0)後2個月(T0 + 2個月；技術一式三份)之隨後訪問顯示減去背景之SFU。將患者1-024-002之資料顯示為減去背景之自2次訪問之SFU。針對患者1-024-002，針對首次訪問(T0；技術一式三份)及於首次訪問(T0)後1個月(T0 + 1個月；技術一式兩份，除了用患者1-024-002-特異性新抗原肽池刺激之樣品外)之隨後訪問顯示減去背景之SFU。

【0451】 圖21描述對圖17A之患者之兩種患者特異性新抗原肽池及DMSO陰性對照之T-細胞反應的檢測。針對各患者，將該患者之經活體外擴展之PBMC用IFN- γ ELISpot中之兩個患者特異性新抗原肽池刺激。針對各供體及各活體外擴展，亦將經活體外擴展之患者PBMC於IFN- γ ELISpot中用DMSO刺激作為陰性對照。將圖21中之數據表示為斑點形成單元(SFU)/ 10^5 個經平板接種之細胞，其中包含背景(對應DMSO陰性對

照)用於患者特異性新抗原肽池及對應DMSO對照。針對患者1-038-001、1-050-001、1-001-002、CU04、1-024-001、1-024-002及CU05顯示對同源肽池#1及#2之單孔(1-038-001、CU02、CU03及1-050-001)或具有生物複製之標準差之平均(所有其他樣品)的反應。針對患者CU02及CU03，細胞數目允許僅對特異性肽池#1測試。認為具有值>背景以上之2倍增加之樣品係陽性及用星星指定(反應性供體包括患者1-038-001、CU04、1-024-001、1-024-002及CU02)。無反應性供體包括患者1-050-001、1-001-002、CU05及CU03。

【0452】 如上關於圖18A至B簡要討論，為證實活體外培養條件僅擴展先已存在之活體內引發之記憶T-細胞，而非使能重新活體外引發，利用HLA匹配之健康供體中之新抗原進行一系列對照實驗。於圖18A至B中及於補充表5中描述此等實驗之結果。使用IVS培養技術，此等實驗之結果證實健康供體中不存在重新引發及不存在可檢測之新抗原特異性T-細胞反應。

【0453】 相比之下，使用IFN- γ ELISpot於利用患者特異性肽池(圖17A及19至21)測試之大多數(5/9, 56%)患者中識別先已存在之新抗原反應性T-細胞。在細胞數目允許個別新抗原同源肽之完全或部分測試之7例患者中，4例患者對經測試之新抗原肽中之至少一者反應，及所有此等患者具有對應池反應(圖17B)。利用個別新抗原測試之其餘3例患者(患者1-001-002、1-050-001及CU05)不具有對單一肽之可檢測之反應(資料未顯示)，其證實此等患者對新抗原池缺少反應(圖17A)。在4例反應性患者中，來自單次訪問之樣品對具有反應之2例患者(患者1-024-001及1-038-001)可得，而來自多次訪問之樣品對具有反應之其他2例患者(CU04及1-

024-002)可得。針對具有來自多次訪問之樣品之2例患者，來自3次訪問(患者CU04)或2次訪問(患者1-024-002)之累積(附加)斑點形成單元(SFU)示於圖17B中及於圖20B中藉由訪問分解。來自相同訪問之額外PBMC樣品亦對患者1-024-002及CU04可得，及重複IVS培養及ELISpot證實對患者特異性新抗原反應(圖20C)。

【0454】總之，在如藉由對圖17A中之10個肽之池之反應所示識別至少一個經T-細胞識別之新抗原決定基的患者中，經識別之新抗原決定基之數目平均為至少2個/患者(5例患者中識別之最少10個抗原決定基，計算不可作為1個經識別之肽去卷積之經識別之池)。除了藉由ELISpot測試IFN- γ 反應外，亦測試培養上清液用於藉由ELISA之顆粒酶B及用於藉由MSD細胞激素多重檢定之TNF- α 、IL-2及IL-5。來自具有陽性ELISpot之4例或5例患者之細胞分泌3種或更多種分析物，包括顆粒酶B(補充表4)，指示新抗原特異性T-細胞之多功能性。重要的是，因為組合預測及IVS方法不依賴於有限組之可得MHC多聚體，跨限制HLA對偶基因廣泛測試反應。此外，與識別經識別之突變之串聯微基因篩選相比，此方法直接識別最小抗原決定基，及需要單獨去卷積步驟以識別最小抗原決定基。總之，新抗原識別產量可與先前最佳方法⁹⁶相比，該等最佳方法利用機採樣品測試對所有突變之TIL，同時僅利用常規5至30 mL全血篩選20個合成肽。

XV.A.肽

【0455】定製、重組凍乾肽係購自JPT Peptide Technologies (Berlin, Germany)或Genscript (Piscataway, NJ, USA)及於無菌DMSO中在10至50 mM下再構成(VWR International, Pittsburgh, PA, USA)，等分及在-80℃下儲存。

XV.B.人類外周血單核細胞(PBMC)

【0456】 來自健康供體之經冷藏保存之HLA-型PBMC (證實HIV、HCV及HBV血清反應陰性)係購自Precision for Medicine (Gladstone, NJ, USA)或Cellular Technology, Ltd. (Cleveland, OH, USA)及儲存於液氮中直至使用。新鮮血液樣品係購自Research Blood Components (Boston, MA, USA)，將來自AllCell (Boston, MA, USA)及PBMC之leukopak在冷藏保存之前藉由Ficoll-Paque密度梯度(GE Healthcare Bio, Marlborough, MA, USA)單離。根據本地臨床標準操作程序(SOP)及協定批准之IRB，將患者PBMC在本地臨床處理中心處理。批准IRB為Quorum Review IRB、Comitato Etico Interaziendale A.O.U. San Luigi Gonzaga di Orbassano及Comité Ético de la Investigación del Grupo Hospitalario Quirón en Barcelona。

【0457】 簡言之，將PBMC通過密度梯度離心單離、洗滌、計數及於CryoStor CS10 (STEMCELL Technologies, Vancouver, BC, V6A 1B6, Canada)中以 5×10^6 個細胞/ml冷藏保存。將經冷藏保存之細胞於cryoport中裝運並轉移以在到達時於LN₂中儲存。於補充表2中列出患者人口統計學。將經冷藏保存之細胞解凍及於含有全能核酸酶(Benzonase) (EMD Millipore, Billerica, MA, USA)之OpTmizer T-細胞擴展基礎培養基(Gibco, Gaithersburg, MD, USA)中洗滌兩次及於不含Benzonase之OpTmizer T-細胞擴展基礎培養基中洗滌一次。使用Guava ViaCount試劑及Guava easyCyte HT細胞計(EMD Millipore)上之模組評估細胞計數及活力。隨後將細胞以濃度及於適用於進程檢定之培養基中再懸浮(參見下節)。

XV.C.活體外刺激(IVS)培養

【0458】 於同源肽及IL-2之存在下，以藉由Ott等人⁸¹應用之類似方法將來自健康供體或患者樣品之先已存在之T-細胞擴展。簡言之，使解凍之PBMC休息過夜及於24孔組織培養板中於肽池(10 μ M/肽，10個肽/池)之存在下於含有10 IU/ml rhIL-2 (R&D Systems Inc., Minneapolis, MN)之ImmunoCult™-XF T-細胞擴展培養基(STEMCELL Technologies)中刺激14天。將細胞以2 x 10⁶個細胞/孔接種及藉由置換2/3之培養基每2至3天供給。一個患者樣品顯示偏離協定及應認為潛在假陰性：解凍後，患者CU03不產生足夠數目之細胞及將細胞以2 x 10⁵個細胞/肽池(比每個協定少10倍)接種。

XV.D. IFN γ 酶聯免疫斑點(ELISpot)檢定

【0459】 藉由ELISpot檢定¹⁴²進行產生IFN γ 之T-細胞之檢測。簡言之，收穫PBMC（離體或活體外擴展後），於無血清RPMI (VWR International)中洗滌及於對照或同源肽之存在下於OpTmizer T-細胞擴展培養基中(離體)或於ImmunoCult™-XF T-細胞擴展培養基中(擴展培養)於塗覆有抗人類IFN γ 捕獲抗體(Mabtech, Cincinnati, OH, USA)之ELISpot Multiscreen板(EMD Millipore)中培養。於5% CO₂、37℃加濕培育器中於18小時培育後，將細胞自板移除及使用抗人類IFN γ 檢測抗體(Mabtech)、維沙坦(Vectastain)抗生物素蛋白過氧化物酶複合體(Vector Labs, Burlingame, CA, USA)及AEC受質(BD Biosciences, San Jose, CA, USA)檢測膜結合IFN γ 。允許將ELISpot板乾燥，避光儲存及送至Zellnet Consulting, Inc., Fort Lee, NJ, USA)用於標準化評價¹⁴³。將數據表示為斑點形成單元(SFU)/經平板接種之細胞數目。

XV.E.顆粒酶B ELISA及MSD多重檢定

【0460】 使用3重檢定MSD U-PLEX Biomarker檢定(目錄號K15067L-2)進行ELISpot上清液中之分泌之IL-2、IL-5及TNF- α 的檢定。根據製造商之說明進行檢定。使用針對各細胞激素之已知標準之連續稀釋計算分析物濃度(pg/ml)。針對圖形資料表示，將標準曲線之最小範圍以下之值表示為等於零。根據製造商之說明使用顆粒酶B DuoSet® ELISA (R & D Systems, Minneapolis, MN)進行ELISpot上清液中之顆粒酶B之檢測。簡言之，將ELISpot上清液於樣品稀釋劑中1:4稀釋及並行顆粒酶B標準之連續稀釋以計算濃度(pg/ml)。針對圖形資料表示，將標準曲線之最小範圍以下之值表示為等於零。

XV.F. IVS檢定之陰性對照實驗-來自健康供體中所測試之腫瘤細胞株之新抗原

【0461】 圖18A說明來自健康供體中所測試之腫瘤細胞株之新抗原之IVS檢定的陰性對照實驗。於IVS培養中利用肽池刺激健康供體PBMC，該等肽池含有陽性對照肽(先前暴露於傳染性疾病)、源自腫瘤細胞株之HLA匹配之新抗原(未暴露)及源自病原體之肽，供體對該等病原體係血清反應陰性。於利用DMSO (陰性對照，黑色圓圈)、PHA及常見傳染性疾病肽(陽性對照，紅色圓圈)、新抗原(未暴露，淺藍色圓圈)或HIV及HCV肽(證實供體係血清反應陰性，海軍藍，A及B)刺激後，隨後將擴展之細胞藉由IFN γ ELISpot (10^5 個細胞/孔)分析。將資料顯示為斑點形成單元(SFU)/ 10^5 個經接種之細胞。顯示具有平均及SEM之生物複製。未觀察到對新抗原或對源自病原體之肽之反應，供體未暴露於該等病原體(血清反應陰性)。

XV.G. IVS檢定之陰性對照實驗—來自健康供體中所測試之患者之新抗原

【0462】 圖18A說明來自針對健康供體中之反應性所測試之患者之新抗原之IVS檢定的陰性對照實驗。評估健康供體中對HLA匹配之新抗原肽池之T-細胞反應。左圖：將健康供體PBMC於離體IFN- γ ELISpot中用對照(DMSO、CEF及PHA)或HLA匹配之源自患者之新抗原肽刺激。將數據表示為斑點形成單元(SFU)/針對一式三份孔之 2×10^5 個經平板接種之細胞。右圖：將於新抗原池或CEF池之存在下擴展之IVS培養後之健康供體PBMC於IFN- γ ELISpot中利用對照(DMSO、CEF及PHA)或HLA匹配之源自患者之新抗原肽池刺激。將數據表示為SFU/針對一式三份孔之 1×10^5 個經平板接種之細胞。未看到對健康供體中之新抗原反應。

XV.H.補充表3：用於NSCLC患者中之T-細胞識別所測試之肽

【0463】 圖17A至C中研究針對N=9例患者所測試之新抗原肽之細節(來自NSCLC患者之新抗原反應性T-細胞之識別)。關鍵區域包括源突變、肽序列及池及觀察到之個別肽響應。「最可能限制」行指示哪些對偶基因預測模型最可能呈遞各肽。亦包含各患者之所有突變肽中之此等肽的排名，如利用結合親和力預測(方法)所計算。

【0464】 存在藉由全MS模型高排名及藉由CD8 T-細胞識別之四種肽，該等肽具有低的經預測之結合親和力或藉由結合親和力預測低排名。

【0465】 針對此等肽中之三者，此由模型與MHCflurry 1.2.0之間之HLA覆蓋之差異造成。預測肽YEHEDVKEA藉由HLA-B*49:01呈遞，該HLA-B*49:01未藉由MHCflurry 1.2.0覆蓋。類似地，預測肽SSAAAPFPL及FVSTSDIKSM藉由HLA-C*03:04呈遞，該HLA-C*03:04

亦未由MHCflurry 1.2.0覆蓋。線上NetMHCpan 4.0 (BA)預測因子(原則上覆蓋所有對偶基因之泛特異性結合親和力預測因子)將SSAAAPFPL列為HLA-C*03:04之強結合劑(23.2 nM，針對患者1-024-002排名第2)，預測FVSTSDIKSM對HLA-C*03:04之弱結合(943.4 nM，針對患者1-024-002排名第39)及YEHEDVKEA對HLA-B*49:01之弱結合(3387.8 nM)，但是對HLA-B*41:01之更強結合(208.9 nM，針對患者1-038-001排名第11)，該HLA-B*41:01亦存在於此患者中但是不由該模型覆蓋。因此，在此等三種肽中，FVSTSDIKSM將藉由結合親和力預測錯失，SSAAAPFPL將經捕獲，及YEHEDVKEA之HLA限制係不確定。

【0466】將肽特異性T-細胞反應去卷積之其餘五種肽來自患者，其中如藉由模型所測定之最可能呈遞對偶基因亦藉由MHCflurry 1.2.0覆蓋。在此等五種肽中，4/5具有強於標準500 nM臨限值之經預測之結合親和力且排名前20，雖然具有較自模型之排名稍微更低排名(肽DENITTIQF、QDVSVQVER、EVADAATLTM、DTVEYPYTSF藉由模型各自排名第0、4、5、7相對於藉由MHCflurry之2、14、7及9)。肽GTKKDVDVLK藉由CD8 T-細胞識別及藉由模型排名第1，但是具有藉由MHCflurry之排名70及經預測之結合親和力2169 nM。

【0467】總之，藉由全MS模型高排名之個別經識別之肽之6/8亦使用結合親和力預測高排名且具有經預測之結合親和力<500nM，而個別經識別之肽之2/8將錯失，若使用結合親和力預測代替全MS模型。

XV.I.補充表4：對來自NSCLC新抗原肽之ELISpot上清液之MSD細胞激素多重及ELISA檢定

【0468】針對顆粒酶B (ELISA)、TNF α 、IL-2及IL-5 (MSD)顯示來

自陽性ELISpot (IFN γ)之上清液中所檢測之分析物。將值顯示為來自技術重複之平均pg/ml。將陽性值用斜體顯示。顆粒酶B ELISA：認為值 $\geq$ DMSO背景之1.5倍係陽性。U-Plex MSD檢定：認為相對於DMSO背景值 $\geq$ 1.5倍係陽性。

XV.J.補充表5：IVS對照實驗中之新抗原及傳染性疾病抗原決定基

【0469】 於IVS對照實驗中所測試之腫瘤細胞株新抗原及病毒肽之細節示於圖18A至B中。關鍵區域包括源細胞株或病毒、肽序列及經預測之呈遞HLA對偶基因。

XV.K.資料

【0470】 用於訓練及測試預測模型之MS肽資料集(圖16)係在MassIVE檔案(massive.ucsd.edu)，登錄號MSV000082648下可得。藉由ELISpot測試之新抗原肽(圖17A至C及18A至B)包含於手稿(補充表3及5)中。

XVI.實例8至11之方法

XVI.A.質譜法

XVI.A.1.樣品

【0471】 用於質譜分析之歸檔之冷凍組織樣品獲自商業來源，包括BioServe (Beltsville, MD)、ProteoGenex (Culver City, CA)、iSpecimen (Lexington, MA)及Indivumed (Hamburg, Germany)。亦自患者在Hopital Marie Lannelongue (Le Plessis-Robinson, France)下在藉由Comité de Protection des Personnes, Ile-de-France VII批准之研究協定下前瞻性收集樣品子集。

XVI.A.2. HLA免疫沉澱

【0472】 使用於裂解後建立之免疫沉澱(IP)方法及組織樣品之溶解^{87,124-126}進行HLA-肽分子之單離。將新鮮冷凍組織粉碎(CryoPrep ; Covaris, Woburn, MA)，添加裂解緩衝液(1% CHAPS、20 mM Tris-HCl、150 mM NaCl、蛋白酶及磷酸酶抑制劑，pH=8)以溶解組織及將所得溶液在4℃下離心2小時成為球碎片。將澄清裂解物用於HLA特異性IP。如先前所述使用抗體W6/32¹²⁷進行免疫沉澱。將裂解物添加至抗體珠中及在4℃下旋轉過夜用於免疫沉澱。於免疫沉澱後，將該等珠自裂解物移除。將IP珠洗滌以移除非特異性結合及將HLA/肽複合體用2N乙酸自該等珠溶離。使用分子量自旋管柱將蛋白質組分自肽移除。將所得肽藉由SpeedVac蒸發至乾及在MS分析之前在-20℃下儲存。

XVI.A.3.肽定序

【0473】 將經乾燥之肽於HPLC緩衝液A中再構成及負載於C-18微毛細管HPLC管柱上用於梯度溶離進入質譜儀中。使用0至40%B之梯度(溶劑A – 0.1%甲酸，溶劑B- 0.1%甲酸/ 80%乙腈)於180分鐘中溶離肽至Fusion Lumos質譜儀(Thermo)中。於具有120,000解析度之Orbitrap檢測器中收集肽質量/電荷(m/z)之MS1光譜，接著於選定離子之HCD破碎後於Orbitrap或離子阱檢測器中收集20個MS2低解析度掃描。使用資料相依性獲取模式及於離子之MS2選擇後30秒之動態排除進行MS2離子之選擇。將針對MS1掃描之自動增益控制(AGC)設置為 4×10^5 及針對MS2掃描設置為 1×10^4 。針對定序HLA肽，可選擇+1、+2及+3電荷狀態用於MS2片段化。

【0474】 針對蛋白質資料庫使用Comet^{128,129}搜索來自各分析之MS2光譜及使用Percolator¹³⁰⁻¹³²對肽識別評分。

XVI.B.機器學習

XVI.B.1.資料編碼

【0475】 針對各樣品，訓練資料點為來自參考蛋白質組之所有8至11mer (包含端值)肽，該蛋白質組對樣品中所表現之一個基因精確圖譜分析。完整訓練資料集藉由連接來自各訓練樣品之訓練資料集形成。選擇長度8至11，因為此長度範圍捕獲所有HLA I類呈遞肽之~95%；然而，可使用相同方法學，以計算需求之適度增加為代價實現將長度12至15添加至模型中。使用獨熱編碼方案將肽及側接序列向量化。藉由利用填補字元擴大胺基酸字母表及將所有肽填充至最大長度11將多個長度(8至11)之肽表示為固定長度向量。將訓練肽之源蛋白之RNA豐度表示為獲自RSEM¹³³之同功異型物-水平轉錄本之對數/百萬(TPM)評估。針對各肽，以對含有該肽之同功異型物各者之每同功異型物TPM評估之總和計算按肽TPM。將來自0 TPM下表現之基因之肽自訓練資料排除，及在測試時間，將來自非表現基因之肽指定為0之呈遞概率。最後，將各肽指定為Ensembl蛋白質家族ID，及各獨特Ensembl蛋白質家族ID對應於每基因呈遞傾向截距(參見下節)。

XVI.B.2.模型架構之說明書

【0476】 全呈遞模型具有下列功能形式：

$$Pr(\text{肽} i \text{ 呈遞}) = \sum_{k=1}^m a_k^i \cdot Pr(\text{對偶基因} k \text{ 呈遞})$$

其中k指示資料集中之HLA對偶基因，其自1至m運行，且 a_k^i 為指示變數，若對偶基因k存在於衍生肽i之樣品中，則其值為1及否則為0。請注意針對給定肽i，所有但是 a_k^i 中之至多6個(6對應於肽i之起源之樣品之HLA類型)將為零。在 $1 - \epsilon$ 下，例如，利用 $\epsilon = 10^{-6}$ 剪輯概率之和。

【0477】 將按對偶基因呈遞概率如下建模：

$$\Pr(\text{藉由對偶基因}a\text{呈遞之肽}i) = \text{S形}\{NN_a(\text{肽}_i) + NN_{\text{側接}}(\text{側接}_i) + NN_{\text{RNA}}(\log(\text{TPM}_i)) + \alpha_{\text{樣品}(i)} + \beta_{\text{蛋白質}(i)}\},$$

其中該等變數具有下列含義： S形 為s形(亦稱場外)函數， 肽_i 為肽i之獨熱編碼之中間填充之胺基酸序列， NN_a 為建模肽序列對呈遞概率之貢獻之具有線性最後層激活的神經網路， 側接_i 為其源蛋白中之肽i之獨熱編碼之側接序列， $NN_{\text{側接}}$ 為建模側接序列對呈遞概率之貢獻之具有線性最後層激活的神經網路， TPM_i 為TPM單元中之肽i之源mRNA之表現， $\text{樣品}(i)$ 為肽i之起源之樣品(即，患者)， $\alpha_{\text{樣品}(i)}$ 為每樣品截距， $\text{蛋白質}(i)$ 為肽i之源蛋白，及 $\beta_{\text{蛋白質}(i)}$ 為每蛋白質截距(亦稱每基因呈遞傾向)。

【0478】 針對結果部分中所述之模型，組件神經網路具有下列架構：

- NN_a 各者為具有輸入尺寸231 (11個殘基x 21個可能字元/殘基，包含填補字元)，寬度256，隱藏層中之整流線性單元(ReLU)激活，輸出層中之線性激活之單隱層多層感知器(MLP)之一個輸出節點，及訓練資料集中之每HLA對偶基因a之一個輸出節點。
- $NN_{\text{側接}}$ 為具有輸入尺寸210 (N-端側接序列之5個殘基+ C-端側接序列之5個殘基x 21個可能字元/殘基，包含填補字元)，寬度32，隱藏層中之整流線性單元(ReLU)激活及輸出層中之線性激活之單隱層MLP。
- NN_{RNA} 為具有輸入尺寸1，寬度16，隱藏層中之整流線性單元(ReLU)激活及輸出層中之線性激活之單隱層MLP。

【0479】 請注意，模型之一些組分(例如， NN_a)取決於特定HLA對偶基因，但是許多組分($NN_{\text{側接}}$ 、 NN_{RNA} 、 $\alpha_{\text{樣品}(i)}$ 、 $\beta_{\text{蛋白質}(i)}$)不取決於特定HLA對偶基因。將前者稱作「對偶基因相互作用」及將後者稱作「對偶基因非

相互作用」。基於生物事前知識選擇建模將特徵建模為對偶基因相互作用或非相互作用：HLA對偶基因參見肽，因此應將肽序列建模為對偶基因相互作用，但是無關於源蛋白之資訊，將RNA表現或側接序列傳遞給HLA分子(因為肽在其遇到內質網中之HLA時自其源蛋白分離)，因此應將此等特徵建模為對偶基因非相互作用。於Keras v2.0.4¹³⁴及Theano v0.9.0¹³⁵中實施模型。

【0480】 肽MS模型使用相同去卷積程序作為全MS模型(方程式1)，但是使用僅考慮肽序列及HLA對偶基因之降階按對偶基因模型產生按對偶基因呈遞概率：

$$\Pr(\text{蛋白對偶基因呈遞之肽} i) = \frac{\exp\{NN_{\alpha}(\text{肽}_i)\}}{\sum_j \exp\{NN_{\alpha}(\text{肽}_j)\}}.$$

【0481】 肽MS模型使用與結合親和力預測相同之特徵，但是模型之權重在不同資料類型(即，質譜資料相對於HLA-肽結合親和力資料)上訓練。因此，比較肽MS模型與全MS模型之預測表現揭示非肽特徵(即，RNA豐度、側接序列、基因ID)對總體預測表現之貢獻，及比較肽MS模型與結合親和力模型之預測表現揭示改良之肽序列之建模對總體預測表現之重要性。

XVI.B.3.訓練/驗證/測試劃分

【0482】 吾人使用下列程序確保無肽出現於一個以上訓練/驗證/測試集中：首先藉由自出現於一個以上蛋白質中之參考蛋白質組移除所有肽，然後藉由將該蛋白質組劃分成10個鄰近肽之區塊。將各區塊唯一分配給訓練、驗證或測試集。以此方式，無肽出現於一個以上訓練、驗證或測試集中。該驗證集僅用於早期停止。圖14至16中之腫瘤樣品測試資料表示來自被完全排除在訓練及驗證集外之五個腫瘤樣品之測試集肽(即，來

自唯一分配給測試集之鄰近肽之區塊之肽)。

XVI.B.4.模型訓練

【0483】 針對模型訓練，獨立建模所有肽，其中按肽損失為陰性伯努利對數可能性損失函數(亦稱作對數損失)。正式地，肽*i*對總體損失之貢獻為

$$\text{Loss}(i) = -\log \left(\Pr(y_i | \Pr(\text{呈遞之肽})) \right),$$

其中 y_i 為肽*i*之標籤；即，若呈遞肽*i*，則 $y_i = 1$ 及否則為0，及 $\Pr(y|p)$ 表示獨立同分佈二元觀察向量 y 給定之參數 $p \in [0, 1]$ 之伯努利可能性。模型藉由最小化損失函數來訓練。

【0484】 為減少訓練時間，藉由隨機移除經陰性標記之訓練資料之90%，產生一個呈遞肽/ $\sim$ 2000非呈遞肽之總體訓練集類別平衡來調整類別平衡。使用Glorot統一程式61初始化及使用ADAM62隨機優化器利用Nvidia Maxwell TITAN X GPU上之標準參數訓練模型權重。包含總資料之10%之驗證集係用於早期停止。每季度在驗證集上評價模型及於第一季度後停止模型訓練，其中驗證損失(即，驗證集上之陰性伯努利對數可能性)未減少。

【0485】 全呈遞模型為10個模型複本之系集，其中各複本利用系集內之每個模型之模型權重之不同隨機初始化在相同訓練資料之雜亂無章複本上獨立訓練。在測試時間，藉由取模型複本之概率輸出之平均值產生預測。

XVI.B.5.模體標識

【0486】 使用weblogolib Python API v3.5.0¹³⁸產生模體標識。為產生結合親和力標識，自2017年7月之免疫抗原決定基資料庫(IEDB⁸⁸)下載

mhc_ligand_full.csv檔案及保留滿足下列標準之肽：量度以奈莫耳(nM)為單位，參考日期在2000年後，目標類型等於「線性肽」及來自權威20-字母胺基酸字母表之肽中之所有殘基。使用具有習知結合臨限值500 nM以下之經量測之結合親和力之經過濾肽的子集產生標識。針對於IEDB中具有太少結合劑之對偶基因對，不產生標識。為產生表示學習呈遞模型之標識，針對各對偶基因及各肽長度預測2,000,000個隨機肽之模型預測。針對各對偶基因及各長度，藉由學習呈遞模型使用排名前1% (即，前20,000名)之肽產生標識。重要的是，來自IEDB之此結合親和力資料不用於模型訓練或測試中，而是僅用於學習之模體之比較。

XVI.B.6.結合親和力預測

【0487】 吾人使用來自MHCflurry v1.2.0¹³⁹(開放源)之僅結合親和力預測因子、具有比得上模型之NetMHC家族之表現之GPU相容性HLA I類結合親和力預測因子預測肽-MHC結合親和力。為組合針對跨多個HLA對偶基因之單肽之結合親和力預測，選擇最小結合親和力。為組合跨多個肽之結合親和力(即，為將藉由如圖16中之多個突變肽跨越之突變排名)，選擇跨該等肽之最小結合親和力。針對T-細胞資料集上之RNA表現臨限值，使用自TCGA至TPM>1下之臨限值之腫瘤類型匹配之RNA-seq資料。將所有原始T-細胞資料集於原始出版物中在TPM>0上過濾，因此不使用在TPM>0上過濾之TCGA RNA-seq資料。

XVI.B.7.呈遞預測

【0488】 為組合跨多個HLA對偶基因之單肽之呈遞概率，如於方程式1中識別概率之和。為組合跨多個肽之呈遞概率(即，為將藉由如圖16中之多個肽跨越之突變排名)，識別呈遞概率之和。概率上，若將肽之呈遞

視作獨立同分佈伯努利隨機變數，則概率之和對應於經呈遞之突變肽之預期數目：

$$E[\text{經呈遞之跨越突變之新抗原數目}] = \sum_{j=1}^{n_i} \text{Pr}[\text{經呈遞之抗原決定基} j],$$

其中 $\text{Pr}[\text{經呈遞之抗原決定基} j]$ 藉由將訓練呈遞模型應用於抗原決定基 j 獲得，且 n_i 表示跨越突變 i 之突變抗原決定基之數目。例如，針對遠離其源基因末端之 SNV i ，針對跨越突變抗原決定基之總共 $n_i = 38$ ，存在 8 個跨越 8-mer、9 個跨越 9-mer、10 個跨越 10-mer 及 11 個跨越 11-mer。

XVI.C. 下一代定序

XVI.C.1. 樣品

【0489】 針對冷凍切除腫瘤之轉錄組分析，RNA 係獲自與用於 MS 分析相同之組織樣品(腫瘤或鄰近正常)。針對進行抗-PD1 療法之患者之新抗原外顯子組及轉錄組分析，DNA 及 RNA 係獲自存檔 FFPE 腫瘤活組織檢查。針對正常外顯子組及 HLA 分型，使用鄰近正常匹配之血液或 PBMC 獲得正常 DNA。

XVI.C.2. 核酸提取及庫構建

【0490】 於製造商建議之程序後，使用 Qiagen DNeasy 管柱(Hilden, Germany)將源自血液之正常/生殖系 DNA 單離。於製造商建議之程序後，使用 Qiagen Allprep DNA/RNA 單離套組將來自組織樣品之 DNA 及 RNA 單離。將 DNA 及 RNA 藉由 Picogreen 及 Ribogreen Fluorescence (Molecular Probes) 定量，將各自具有 >50 ng 產量之樣品提前至庫構建。按照製造商建議之方案，藉由聲學剪切(Covaris, Woburn, MA)，接著 DNA Ultra II (NEB, Beverly, MA) 庫製備套組產生 DNA 定序庫。藉由熱片段化及具有 RNA Ultra II (NEB) 之庫構建產生腫瘤 RNA 定序庫。將所得庫藉由

Picogreen (Molecular Probes)定量。

XVI.C.3.全外顯子組捕獲

【0491】使用 xGEN 全外顯子組小組 (Integrated DNA Technologies)進行DNA及RNA定序庫二者之外顯子濃化。將1至1.5 µg源自正常DNA或腫瘤DNA或RNA之庫用作輸入並允許雜交大於12小時，接著鏈黴抗生物素純化。將捕獲之庫藉由PCR最低程度擴增及藉由NEBNext Library Quant Kit (NEB)定量。將捕獲之庫在等莫耳濃度下彙集及使用c-bot (Illumina)聚集及在HiSeq4000 (Illumina)上在75個鹼基對端定序至>500x腫瘤外顯子組、>100x正常外顯子組及>100M閱讀腫瘤轉錄組之目標獨特平均覆蓋度。

XVI.C.4.分析

【0492】使用 BWA-MEM¹⁴⁴ (v. 0.7.13-r1126)將外顯子組閱讀 (FFPE腫瘤及匹配正常)與參考人類基因組(hg38)比對。使用STAR (v. 2.5.1b)將 RNA-seq 閱讀 (FFPE 及 冷凍腫瘤組織樣品)與基因組及 GENCODE轉錄本(v. 25)比對。使用具有相同參考轉錄本之RSEM¹³³ (v. 1.2.31)將RNA表現定量。Picard (v. 2.7.1)係用於標記重複比對及計算比對度量。針對利用GATK¹⁴⁵ (v. 3.5-0)之基本質量分數再校準後之FFPE腫瘤樣品，使用成對腫瘤-正常外顯子組利用FreeBayes¹⁴⁶ (1.0.2)測定取代及短插入或缺失變數。篩檢程序包括對偶基因頻率>4%；中間鹼基品質>25，支持閱讀之最小圖譜分析品質30及具有獲得之充分覆蓋之正常中之交替閱讀計數≤2。亦必須在兩股上檢測變數。排除發生於重複區域中之體細胞變化。使用RefSeq轉錄本利用snpEff¹⁴⁷ (v. 4.2)進行轉譯及註釋。將腫瘤RNA比對中驗證之非同義、非停止變異體提前至新抗原預測。使

用Optitype¹⁴⁸ 1.3.1產生HLA類型。

XVI.C.5.圖18A至B：針對IVS對照實驗之腫瘤細胞株及匹配之正常

【0493】 按照銷售者之說明，使均購自ATCC (Manassas, VA)之腫瘤細胞株H128、H122、H2009、H2126、Colo829及其正常供體匹配之對照細胞株BL128、BL2122、BL2009、BL2126及Colo829BL生長至 10^{83} 至 10^{84} 個細胞，然後速凍用於核酸提取及定序。一般如上所述進行NGS過程，不同之處在於使用MuTect¹⁴⁹ (3.1-0)僅用於取代突變檢測。補充表5中列出IVS對照檢定中所使用之肽。

XVI.D.針對MHC II類分子之呈遞熱點建模

【0494】 吾人亦評價當使用呈遞熱點參數時及當不使用呈遞熱點參數時，本文中所揭示之II類HLA肽呈遞模型之性能。當I類複合體呈遞胞漿蛋白及存在於人類所有有核細胞之表面時，II類複合體主要存在於抗原呈遞細胞上及主要負責呈遞細胞外(外源)蛋白。I類及II類在其結合機制及肽長度之間亦存在差異。

【0495】 為評價當使用呈遞熱點特徵時及當不使用呈遞熱點特徵時，本文中所揭示之II類HLA肽呈遞模型之性能，針對兩種細胞株獲得公開之II類質譜資料，該等細胞株各者表現單HLA I類對偶基因。一種細胞株表現HLA-DRB1*15:01及另一種表現HLA-DRB5*01:01¹⁵⁰。此等兩種細胞株係用於訓練資料。針對測試資料，II類質譜資料係獲自表現HLA-DRB1*15:01及HLA-DRB5*01:01二者之分開細胞株。¹⁵¹ RNA定序資料對訓練或測試細胞株不可得，因此取代來自不同B-細胞株(B721.221⁹²)之RNA定序資料。

【0496】 使用與針對HLA I類資料相同之程序將肽集分成訓練集、

驗證集及測試集，不同之處在於包含長度在9與20之間之II類資料肽。訓練資料包括藉由HLA-DRB1*15:01呈遞之330個肽及藉由HLA-DRB5*01:01呈遞之103個肽。測試資料集包括藉由HLA-DRB1*15:01或HLA-DRB5*01:01連同4708個非呈遞肽呈遞之223個肽。

【0497】用於產生圖22中所述之結果之呈遞模型為本文中所揭示之MHC II類呈遞預測模型。該呈遞模型為預測HLA II類肽呈遞之在訓練資料集上訓練之10個模型之系集。此等模型之架構及訓練程序與用於預測I類呈遞之彼等相同，不同之處在於II類模型將作為獨熱編碼及零填充之輸入肽序列帶至長度20而非11。圖22比較當預測新抗原決定基藉由MHC II類分子之呈遞時，使用呈遞熱點參數之呈遞模型與不使用呈遞熱點參數之呈遞模型之預測表現。具體而言，圖22描述此等兩個版本之呈遞模型之接收器操作特性(ROC)曲線。熱點模型產生改善之性能，達成0.96之ROC曲線下面積(ROC AUC)，而無熱點之模型產生僅0.93之ROC AUC。

XVII.實例12：對來自NSCLC患者之外周血之新抗原特異性記憶T-細胞之TCR定序

【0498】圖23描述對來自NSCLC患者之外周血之新抗原特異性記憶T-細胞之TCR定序的方法。於ELISpot培育後收集來自NSCLC患者CU04之外周血單核細胞(PBMC) (如以上關於圖17A至21所述)。具體而言，如上所討論，將來自2次訪問之患者CU04之活體外擴展之PBMC於IFN- γ ELISpot中利用CU04-特異性個別新抗原肽(圖20C)、利用CU04-特異性新抗原肽池(圖20C)及利用DMSO陰性對照(圖21)刺激。於培育後及在添加檢測抗體之前，將PBMC轉移至新培養板中及在ELISpot檢定完成期間保持於培育器中。基於ELISpot結果識別陽性(反應性)孔。如圖20中所示，

經識別之陽性孔包括利用CU04-特異性個別新抗原肽8刺激之孔及利用CU04-特異性新抗原肽池刺激之孔。將來自此等陽性孔及陰性對照(DMSO)孔之細胞組合及用磁力標記之抗體針對CD137染色用於使用Miltenyi磁性單離管柱濃化。

【0499】 使用10x基因組單細胞解析成對免疫TCR剖析方法將如上所述經單離及擴展之經CD137-濃化及經耗盡之T-細胞片段定序。具體而言，將活的T細胞分隔成單細胞乳液用於隨後單細胞cDNA生成及全長TCR剖析(通過恆定區之5' UTR，確保 α 及 β 成對)。一種方法利用在轉錄本之5'端處轉換寡核苷酸之分子條碼模板，第二種方法利用3'端處之分子條碼恆定區寡核苷酸，及第三種方法將RNA聚合酶啟動子與TCR之5'端或3'端偶合。所有此等方法使能在單細胞水平下識別及去卷積 α 及 β TCR對。所得條碼cDNA轉錄本經歷最佳化酶促及庫構建工作流程以減少偏差並確保細胞之池內之純系型的精確表示。將庫在Illumina之MiSeq或HiSeq4000儀器(成對-端150個循環)上定序用於約5至5萬個閱讀/細胞之目標定序深度。補充表6中描述所得TCR核酸序列。藉由基於正交錨-PCR之TCR定序方法(Archer)證實補充表6中所述之TCRa及TCRb鏈之存在。此特定方法具有使用有限細胞數目作為輸入及當與基於10x基因組之TCR定序相比時更少酶促操作之優點。

【0500】 使用10x軟體及定製生物資訊學管線分析定序輸出以識別亦如補充表6中所示之T-細胞受體(TCR) α 及 β 鏈對。補充表6另外列出最普遍TCR純系型之 α 及 β 可變(V)區、連接(J)區、恆定(C)區及 β 多樣性(D)區及CDR3胺基酸序列。將純系型定義為獨特CDR3胺基酸序列之 α 、 β 鏈對。針對以2個細胞以上頻率存在之單 α 及單 β 鏈對，將純系型過濾以產生

患者CU04中之純系型/目標肽之最終清單(補充表6)。

【0501】總言之，使用以上關於圖23所述之方法，識別來自患者CU04之外周血之記憶CD8⁺ T-細胞，該等細胞對如上關於第XIV節中之實例10所討論之經識別之患者CU04之腫瘤新抗原係新抗原特異性。將此等經識別之新抗原-特異性T-細胞之TCR定序。及此外，識別經定序之TCR，該等TCR對如藉由以上呈遞模型所識別之患者CU04之腫瘤新抗原係新抗原特異性。

XVIII.實例13：使用用於T-細胞療法之新抗原特異性記憶T-細胞

【0502】於識別對藉由患者之腫瘤呈遞之新抗原係新抗原特異性的T-細胞及/或TCR後，可將此等經識別之新抗原特異性T-細胞及/或TCR用於患者之T-細胞療法。具體而言，可使用此等經識別之新抗原特異性T-細胞及/或TCR產生治療數量之新抗原特異性T-細胞用於在T-細胞療法期間融入患者中。本文中第XVII.A.及XVII.B.節中討論產生用於患者之T-細胞療法之治療數量之新抗原特異性T-細胞的兩種方法。第一種方法包括擴展來自患者樣品之經識別之新抗原特異性T-細胞(第XVII.A.節)。第二種方法包括對經識別之新抗原特異性T-細胞之TCR定序及將經定序之TCR選殖至新的T-細胞中(第XVII.B.節)。亦可使用本文中未明確提及之產生用於T-細胞療法之新抗原特異性T-細胞的替代方法以產生用於T-細胞療法之治療數量之新抗原特異性T-細胞。一旦經由此等方法中之一或多者獲得新抗原特異性T-細胞，就可將此等新抗原特異性T-細胞融合至患者中。

XVIII.A.用於T-細胞療法之來自患者樣品之新抗原特異性記憶T-細胞之識別及擴展

【0503】用於產生用於患者之T-細胞療法之治療數量之新抗原特異

性T-細胞的第一種方法包括擴展來自患者樣品之經識別之新抗原特異性T-細胞。

【0504】 具體而言，為將新抗原特異性T-細胞擴展至用於患者之T-細胞療法之治療數量，使用如上所述之呈遞模型識別最可能藉由患者之癌症細胞呈遞之新抗原肽集。此外，自患者獲得含有T-細胞之患者樣品。該患者樣品可包括患者之外周血、腫瘤浸潤淋巴細胞(TIL)或淋巴結細胞。

【0505】 於患者樣品包括患者之外周血之實施例中，可使用下列方法將新抗原特異性T-細胞擴展至治療數量。於一實施例中，可進行引發。於另一實施例中，可使用上述方法中之一或多者識別已經激活之T-細胞。於另一實施例中，可進行已經激活之T-細胞之引發及識別二者。引發及識別已經激活之T-細胞二者之優點為將表示之特異性之數目最大化。引發及識別已經激活之T-細胞二者之缺點為此方法係困難且費時。於另一實施例中，可將不一定被激活之新抗原特異性細胞單離。於此等實施例中，亦可進行此等新抗原特異性細胞之抗原特異性或非特異性擴展。於收集此等經引發之T-細胞後，可使該等經引發之T-細胞接受快速擴展協定。例如，於一些實施例中，可使該等經引發之T-細胞接受Rosenberg快速擴展協定(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2978753/>、<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2305721/>)^{153、154}。

【0506】 於患者樣品包括患者之TIL之實施例中，可使用下列方法將新抗原特異性T-細胞擴展至治療數量。於一實施例中，可將新抗原特異性TIL離體四聚體/多聚體分類，及然後可使經分類之TIL接受如上所述之快速擴展協定。於另一實施例中，可進行TIL之新抗原非特異性擴展，然後可將新抗原特異性TIL四聚體分類，及然後使經分類之TIL接受如上所

述之快速擴展協定。於另一實施例中，可在使TIL接受快速擴展協定之前進行抗原特異性培養。
(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4607110/>、
<https://onlinelibrary.wiley.com/doi/pdf/10.1002/eji.201545849>)^{155、156}。

【0507】於一些實施例中，可修改Rosenberg快速擴展協定。例如，可將抗PD1及/或抗41BB添加至TIL培養中以模擬更快速擴展。
(<https://jitc.biomedcentral.com/articles/10.1186/s40425-016-0164-7>)¹⁵⁷。

XVIII.B.識別新抗原特異性T-細胞、對經識別之新抗原特異性T-細胞之TCR定序及將經定序之TCR選殖至新的T-細胞中

【0508】用於產生用於患者之T-細胞療法之治療數量之新抗原特異性T-細胞的第二種方法包括識別來自患者樣品之新抗原特異性T-細胞，對經識別之新抗原特異性T-細胞之TCR定序，及將經定序之TCR選殖至新的T-細胞中。

【0509】首先，自患者樣品識別新抗原特異性T-細胞，及將經識別之新抗原特異性T-細胞之TCR定序。可自其單離T細胞之患者樣品可包括血液、淋巴結或腫瘤中之一或多者。更具體而言，可自其單離T細胞之患者樣品可包括外周血單核細胞(PBMC)、腫瘤浸潤細胞(TIL)、解離腫瘤細胞(DTC)、經活體外引發之T細胞及/或自淋巴結單離之細胞中之一或多者。此等細胞可係新鮮及/或經冷凍。PBMC及經活體外引發之T細胞可獲自癌症患者及/或健康受試者。

【0510】於獲得患者樣品後，可將該樣品擴展及/或引發。可實施各種方法擴展及引發患者樣品。於一實施例中，可在肽或串聯微基因之存在

下模擬新鮮及/或冷凍PBMC。於另一實施例中，可在肽或串聯微基因之存在下利用抗原呈遞細胞(APC)模擬及引發新鮮及/或冷凍經單離之T-細胞。APC之實例包括B-細胞、單核細胞、樹突狀細胞、巨噬細胞或人工抗原呈遞細胞(諸如於<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2929753>中評論之呈遞相關HLA及共同刺激分子之細胞或珠)。於另一實施例中，可在細胞激素(例如，IL-2、IL-7及/或IL-15)之存在下刺激PBMC、TIL及/或經單離之T-細胞。於另一實施例中，可在最大刺激物、細胞激素及/或飼養細胞之存在下刺激TIL及/或經單離之T-細胞。於此等實施例中，T細胞可藉由激活標誌物及/或多聚體(例如，四聚體)單離。於另一實施例中，可利用刺激因子及/或共同刺激標誌物(例如，CD3抗體、CD28抗體及/或珠(例如，DynaBead)刺激TIL及/或經單離之T細胞。於另一實施例中，可使用快速擴展協定在飼養細胞上在IL-2之高劑量下於富有培養基中擴展DTC。

【0511】 然後，識別及單離新抗原特異性T細胞。於一些實施例中，將T細胞自患者樣品離體單離無需先前擴展。於一實施例中，可使用以上關於第XVI.節中所述之方法識別來自患者樣品之新抗原特異性T細胞。於替代實施例中，藉由陽性選擇之特定細胞群體之濃化或藉由陰性選擇之特定細胞群體之耗盡進行單離。於一些實施例中，藉由利用一或多個抗體或其他結合劑培育細胞實現陽性或陰性選擇，該等結合劑特異性結合至各自在經陽性或陰性選擇之細胞上表現或在相對較高水平(標誌物^高)下表現(標誌物+)之一或多個表面標誌物。

【0512】 於一些實施例中，藉由在非T細胞(諸如B細胞、單核細胞或其他白細胞(諸如CD14))上表現之標誌物之陰性選擇將T細胞自PBMC樣

品分離。於一些態樣中，使用CD4+或CD8+選擇步驟分離CD4+輔助及CD8+細胞毒性T-細胞。可藉由對在一或多個初始T-細胞、記憶T-細胞及/或效應T-細胞子群體上表現或相對較高程度表現之標誌物之陽性或陰性選擇將此等CD4+及CD8+群體進一步分成子群體。

【0513】 於一些實施例中，CD8+細胞(諸如)藉由基於與各自子群體相關之表面抗原之陽性或陰性選擇進一步濃化或耗盡初始幹細胞、中樞記憶幹細胞、效應記憶幹細胞及/或中樞記憶幹細胞。於一些實施例中，進行中樞記憶T (TCM)細胞之濃化以增加功效(諸如改善長期生存、擴展及/或於投與後之移植物移入)，該功效於一些態樣中於此等子群體中係特別穩健的。參見Terakura等人(2012) *Blood*. 1:72-82；Wang等人(2012) *J Immunother*. 35(9):689-701。於一些實施例中，組合經TCM濃化之CD8+ T-細胞及CD4+ T-細胞進一步增強功效。

【0514】 於實施例中，記憶T細胞存在於CD8+外周血淋巴細胞之CD62L+及CD62L-子集二者中。PBMC可(諸如)使用抗-CD8抗體及抗-CD62L抗體濃化或耗盡CD62L-CD8+及/或CD62L+CD8+片段。

【0515】 於一些實施例中，中樞記憶T (TCM)細胞之濃化係基於CD45RO、CD62L、CCR7、CD28、CD3及/或CD127之陽性或高表面表現；於一些態樣中，其係基於對表現或高度表現CD45RA及/或顆粒酶B之細胞之陰性選擇。於一些態樣中，藉由表現CD4、CD14、CD45RA之細胞之耗盡及對表現CD62L之細胞之陽性選擇或濃化進行對TCM細胞濃化之CD8+群體之單離。於一態樣中，以基於CD4表現選擇之細胞之陰性片段開始進行中樞記憶T (TCM)細胞之濃化，使該濃化接受基於CD14及CD45RA之表現之陰性選擇及基於CD62L之陽性選擇。於一些態樣中，同

時進行此等選擇及於其他態樣中，依序(以任一順序)進行此等選擇。於一些態樣中，於製備CD8+細胞群體或子群體中使用之相同基於CD4表現之選擇步驟亦用於產生CD4+細胞群體或子群體，使得來自基於CD4之分離之陽性及陰性片段二者經保留及於方法之隨後步驟中(視情況，於一或多個其他陽性或陰性選擇步驟後)使用。

【0516】 於一特定實例中，使PBMC樣品或其他白細胞樣品接受CD4+細胞之選擇，其中保留陰性及陽性片段二者。然後使陰性片段接受基於CD14及CD45RA或ROR1之表現之陰性選擇，及基於中樞記憶T-細胞(諸如CD62L或CCR7)之標誌物特徵之陽性選擇，其中以任一順序進行該等陽性及陰性選擇。

【0517】 藉由識別具有細胞表面抗原之細胞群體將CD4+ T輔助細胞分成初始細胞、中樞記憶細胞及效應細胞。CD4+淋巴細胞可藉由標準方法獲得。於一些實施例中，初始CD4+ T淋巴細胞為CD45RO⁻、CD45RA⁺、CD62L⁺、CD4+ T-細胞。於一些實施例中，中樞記憶CD4+細胞為CD62L⁺及CD45RO⁺。於一些實施例中，效應CD4+細胞為CD62L⁻及CD45RO⁻。

【0518】 於一實例中，為藉由陰性選擇對CD4+細胞濃化，單株抗體混合物通常包括CD14、CD20、CD11b、CD16、HLA-DR及CD8之抗體。於一些實施例中，使抗體或結合搭檔結合至固體支撐物或基質(諸如磁珠或順磁珠)以允許分離細胞進行陽性及/或陰性選擇。例如，於一些實施例中，使用免疫磁性(或親和力-磁性)分離技術(於Methods in Molecular Medicine，第58卷：Metastasis Research Protocols，第2卷：Cell Behavior In Vitro and In Vivo，第17至25頁藉由S. A. Brooks and

U. Schumacher Humana Press Inc., Totowa, N.J.編輯中所評論)將細胞及細胞群體分離或單離。

【0519】於一些態樣中，將待分離之細胞樣品或組合物用小的、可磁化或磁回應物質(諸如磁回應粒子或微粒，諸如順磁珠(例如，諸如Dynabead或MACS珠))培育。磁回應物質(例如，粒子)一般直接或間接連接至結合搭檔，例如，特異性結合至分子之抗體，例如，在細胞上呈遞之表面標誌物、期望分離(例如，期望陰性或陽性選擇)之細胞或細胞群體。

【0520】於一些實施例中，磁性粒子或磁珠包括結合至特異性結合成員(諸如抗體或其他結合搭檔)之磁回應物質。存在許多熟知於磁性分離方法中使用之磁回應物質。適宜磁性粒子包括於Molday美國專利案第4,452,773號中及於歐洲專利說明書EP 452342 B中(其以引用的方式併入)所述之彼等。膠體大小粒子(諸如於Owen美國專利案第4,795,698號及Liberti等人，美國專利案第5,200,084號中所述之彼等)為其他實例。

【0521】一般在抗體或結合搭檔或分子，諸如二級抗體或特異性結合至此等抗體或結合搭檔之其他試劑，連接至特異性結合至細胞表面分子(若在樣品內之細胞上呈遞)之磁性粒子或磁珠之其他試劑之條件下進行培育。

【0522】於一些態樣中，將樣品放置於磁場中，及具有連接其上之磁回應或可磁化粒子之彼等細胞將被磁鐵吸引及自未經標記之細胞分離。針對陽性選擇，保留被磁鐵吸引之細胞；針對陰性選擇，保留未被吸引之細胞(未經標記之細胞)。於一些態樣中，在相同選擇步驟期間進行陽性及陰性選擇之組合，其中將該等陽性及陰性片段保留及進一步處理或接受進一步分離步驟。

【0523】於某些實施例中，將該等磁回應粒子於初級抗體或其他結合搭檔、二級抗體、凝集素、酵素或鏈黴抗生物素中塗覆。於某些實施例中，將該等磁性粒子經由對一或多個標誌物特異性之初級抗體之塗覆連接至細胞。於某些實施例中，利用初級抗體或結合搭檔標記細胞而非珠，及然後添加經細胞類型特異性二級抗體或其他結合搭檔(例如，鏈黴抗生物素)塗覆之磁性粒子。於某些實施例中，將經鏈黴抗生物素塗覆之磁性粒子聯合經生物素化初級或二級抗體使用。

【0524】於一些實施例中，將該等磁回應粒子左連接至待隨後培育、培養及/或基因改造之細胞中；於一些態樣中，將粒子左連接至細胞用於對患者投與。於一些實施例中，自細胞移除可磁化或磁回應粒子。自細胞移除可磁化粒子之方法係已知且包括(例如)使用競爭未經標記之抗體、可磁化粒子或偶聯至可裂解連接子之抗體等。於一些實施例中，該等可磁化粒子係可生物降解。

【0525】於一些實施例中，基於親和力之選擇係經由經磁性激活之細胞分選(MACS) (Miltenyi Biotec, Auburn, Calif.)。經磁性激活之細胞分選(MACS)系統能使具有連接其上之磁化粒子之細胞高純度選擇。於某些實施例中，MACS以模式操作，其中於應用外部磁場後，將非目標及目標物質依序溶離。即，將連接至磁化粒子之細胞保持於位置中，同時將未經連接之物質溶離。然後，於此第一個溶離步驟完成後，將困在磁場中及防止被溶離之物質以某種方式釋放使得其可經溶離及恢復。於某些實施例中，將非大T細胞標記及自不同細胞群體耗盡。

【0526】於某些實施例中，使用進行方法之單離、細胞製備、分離、處理、培育、培養及/或調配步驟中之一或多者之系統、設備或裝置

進行單離或分離。於一些態樣中，使用該系統於封閉或無菌環境中進行此等步驟各者(例如)以最小化誤差、使用者操作及/或污染。於一實例中，該系統係如國際專利申請案公開號WO2009/072003或US 20110003380 A1中所述之系統。

【0527】 於一些實施例中，該系統或裝置於系集或自含系統中及/或以自動化或可程序設計方式進行單離、處理、基因改造及調配步驟中之一或多者(例如，所有)。於一些態樣中，該系統或裝置包含與系統或裝置交流之電腦及/或電腦程序，其允許使用者程序設計、控制、評估處理、單離、基因改造及調配步驟之結果及/或調整處理、單離、基因改造及調配步驟之各種態樣。

【0528】 於一些態樣中，例如，針對細胞於封閉且無菌系統中在臨床級別水平下之自動化分離，使用CliniMACS系統(Miltenyi Biotec)進行分離及/或其他步驟。元件可包括積體微電腦、磁分離裝置、蠕動泵及各種夾緊閥。於一些態樣中，積體電腦控制儀器之所有元件及指導系統以標準化序列進行重複程序。於一些態樣中，磁分離裝置包括包括可移動之永久磁鐵及選擇管柱之支架。蠕動泵控制整個管組之流率及與夾緊閥一起確保緩衝液通過系統之控制流動及細胞之連續懸浮。

【0529】 於一些態樣中，該CliniMACS系統使用於無菌非熱原溶液中供給之抗體偶合之可磁化粒子。於一些實施例中，於將細胞利用磁性粒子標記後，將該等細胞洗滌以移除過量粒子。然後將細胞製備袋連接至管組，該管組轉而連接至含有緩衝液之袋及細胞收集袋。該管組由預組裝之無菌管(包含預管柱及分離管柱)組成且僅用於單一用途。於啟動分離程序後，系統自動將細胞樣品應用於分離管柱。將經標記之細胞保留於管柱

內，同時藉由一系列洗滌步驟移除未經標記之細胞。於一些實施例中，利用本文中所述方法使用之細胞群體未經標記且不保留於管柱中。於一些實施例中，利用本文中所述方法使用之細胞群體經標記且保留於管柱中。於一些實施例中，於移出磁場後，將利用本文中所述方法使用之細胞群體自管柱溶離，及收集於細胞收集袋內。

【0530】於某些實施例中，使用CliniMACS Prodigy系統(Miltenyi Biotec)進行分離及/或其他步驟。於一些態樣中，該CliniMACS Prodigy系統裝備有允許自動洗滌及細胞藉由離心分離之細胞處理裝置。該CliniMACS Prodigy系統亦可包含車載相機及圖像識別軟體，該軟體藉由識別源細胞產品之宏觀層確定最佳細胞分離終點。例如，可將外周血自動分離成紅細胞、白細胞及血漿層。該CliniMACS Prodigy系統亦可包含積體細胞培植室，其實現細胞培養協定，諸如，例如，細胞分化及擴展、抗原負載及長期細胞培養。輸入埠可允許培養基之無菌移除及補充及可使用積體顯微鏡監測細胞。參見，例如，Klebanoff等人(2012) J Immunother. 35(9): 651-660，Terakura等人(2012) Blood. 1:72-82及Wang等人(2012) J Immunother. 35(9):689-701。

【0531】於一些實施例中，經由流式細胞計收集及濃化(或耗盡)本文中所述細胞群體，其中將經多細胞表面標誌物染色之細胞帶入流體流中。於一些實施例中，經由製備級(FACS)分選收集及濃化(或耗盡)本文中所述細胞群體。於某些實施例中，藉由使用微機電系統(MEMS)晶片與基於FACS之檢測系統組合收集及濃化(或增濃)本文中所述細胞群體(參見，例如，WO 2010/033140，Cho等人(2010) Lab Chip 10, 1567-1573；及Godin等人(2008) J Biophoton. 1(5):355-376)。於兩種情況下，可將細胞

用多個標誌物標記，允許經良好定義之T-細胞子集在高純度下之單離。

【0532】 於一些實施例中，可將抗體或結合搭檔用一或多個可檢測標誌物標記以促進陽性及/或陰性選擇之分離。例如，分離可基於結合至經螢光標記之抗體。於一些實例中，可於流體流中，諸如藉由經螢光激活之細胞分選(FACS) (包括製備級(FACS)及/或微機電系統(MEMS)晶片) (例如)與流式細胞計檢測系統組合進行基於對一或多個細胞表面標誌物特异性之抗體或其他結合搭檔之結合之細胞的分離。此等方法允許同時基於多個標誌物之陽性及陰性選擇。

【0533】 於一些實施例中，該等製備方法包括在單離、培育及/或基因改造之前或之後冷凍(例如，冷藏保存)細胞之步驟。於一些實施例中，冷凍及隨後解凍步驟移除細胞群體中之粒細胞及在某種程度上，單核細胞。於一些實施例中，例如，於移除血漿及血小板之洗滌步驟後，將細胞懸浮於冷凍溶液中。於一些態樣中，可使用任何各種已知冷凍溶液及參數。一個實例涉及使用含有20% DMSO及8%人血清白蛋白(HSA)之PBS或其他適宜細胞冷凍培養基。然後可將此用培養基1:1稀釋使得DMSO及HAS之最終濃度各自為10%及4%。其他實例包括Cryostor®、CTL-Cryo™ ABC冷凍培養基及類似者。然後以1度/分鐘之速率將細胞冷凍至-80℃並儲存於液氮儲罐之蒸汽相中。

【0534】 於一些實施例中，所提供之方法包括培植、培育、培養及/或基因改造步驟。例如，於一些實施例中，提供培育及/或基因改造經耗盡之細胞群體及培養引發組合物之方法。

【0535】 因此，於一些實施例中，將細胞群體於培養引發組合物中培育。可於培養容器(諸如裝置、室、孔、管柱、管、管組、閥、小瓶、

培養皿、袋)或用於培養或培植細胞之其他容器中進行培育及/或基因改造。

【0536】 於一些實施例中，在基因改造之前或連同基因改造培育及/或培養細胞。培育步驟可包括培養、培植、刺激、激活及/或繁殖。於一些實施例中，在刺激條件或刺激劑之存在下培育組合物或細胞。此等條件包括經設計以誘導群體中細胞之增生、擴展、激活及/或生存，模擬抗原暴露及/或引發細胞進行基因改造(諸如引入重組抗原受體)之彼等。

【0537】 該等條件可包括下列中之一或多者：特定培養基、溫度、氧含量、二氧化碳含量、時間、物劑(例如，營養素、胺基酸、抗生素)、離子及/或刺激因子(諸如細胞激素、趨化因子、抗原、結合搭檔、融合蛋白、重組可溶性受體)及經設計以激活細胞之任何其他物劑。

【0538】 於一些實施例中，刺激條件或物劑包括能激活TCR複合體之細胞內信號傳導域之一或多種物劑(例如，配位體)。於一些態樣中，該物劑打開或引發T-細胞中之TCR/CD3細胞內信號傳導級聯。此等物劑可包括抗體(諸如對TCR組分及/或(例如)結合至固體支撐物(諸如珠)之共同刺激受體(例如，抗-CD3、抗-CD28)特異性之彼等)及/或一或多種細胞激素。視情況，擴展方法可另外包括將抗-CD3及/或抗CD28抗體添加至培養基中之步驟(例如，以至少約0.5 ng/ml之濃度)。於一些實施例中，刺激劑包括IL-2及/或IL-15，例如，至少約10個單位/mL之IL-2濃度。

【0539】 於一些態樣中，根據諸如頒與Riddell等人之美國專利案第6,040,177號，Klebanoff等人(2012) J Immunother. 35(9): 651-660，Terakura等人(2012) Blood. 1:72-82及/或Wang等人(2012) J Immunother. 35(9):689-701中所述之彼等之技術進行培育。

【0540】 於一些實施例中，藉由添加至培養引發組合物飼養細胞，諸如非分裂外周血單核細胞(PBMC)中(例如，使得所得細胞群體含有至少約5、10、20或40或更多個待擴展之初始群體中之各T淋巴細胞之PBMC飼養細胞)及培育培養物(例如，持續足以擴展T-細胞數目之時間)擴展T-細胞。於一些態樣中，該等非分裂飼養細胞可包括經 γ -照射之PBMC飼養細胞。於一些實施例中，將PBMC用約3000至3600雷德(rad)範圍之 γ 射線照射以防止細胞分裂。於一些實施例中，將PBMC飼養細胞用絲裂黴素(Mytomicin) C滅活。於一些態樣中，在添加T-細胞群體之前，將飼養細胞添加至培養基中。

【0541】 於一些實施例中，刺激條件包括適用於人類T淋巴細胞生長之溫度，例如，至少約25攝氏度，一般至少約30度，及一般在或約37攝氏度。視情況，培育可另外包括添加非分裂EBV-轉化之淋巴母細胞(LCL)作為飼養細胞。可將LCL用約6000至10,000雷德範圍之 γ 射線照射。於一些態樣中，以任何適宜量(諸如LCL飼養細胞與初始T淋巴細胞之比率為至少約10:1)提供LCL飼養細胞。

【0542】 於實施例中，藉由利用抗原刺激初始或抗原特異性T淋巴細胞獲得抗原特異性T-細胞(諸如抗原特異性CD4+及/或CD8+ T-細胞)。例如，可藉由自感染受試者單離T-細胞及利用相同抗原刺激活體外細胞產生對巨細胞病毒抗原抗原特異性T-細胞株或純系。

【0543】 於一些實施例中，於利用功能檢定(例如，ELISpot)刺激後識別及/或單離新抗原特異性T-細胞。於一些實施例中，藉由細胞內細胞激素染色分選多功能細胞將新抗原特異性T-細胞單離。於一些實施例中，使用激活標誌物(例如，CD137、CD38、CD38/HLA-DR雙陽性及/或

CD69)識別及/或單離新抗原特異性T-細胞。於一些實施例中，使用I類或II類多聚體及/或激活標誌物識別及/或單離新抗原特異性CD8+、自然殺手T-細胞、記憶T-細胞及/或CD4+ T-細胞。於一些實施例中，使用記憶標誌物(例如，CD45RA、CD45RO、CCR7、CD27及/或CD62L)識別及/或單離新抗原特異性CD8+及/或CD4+ T-細胞。於一些實施例中，識別及/或單離增生細胞。於一些實施例中，識別及/或單離經激活之T-細胞。

【0544】 於識別來自患者樣品之新抗原特異性T-細胞後，將經識別之新抗原特異性T-細胞之新抗原特異性TCR定序。為對新抗原特異性TCR定序，首先必須識別該TCR。識別T-細胞之新抗原特異性TCR之一種方法可包括使該T-細胞與包含至少一個新抗原之HLA-多聚體(例如，四聚體)接觸；及經由該HLA-多聚體與該TCR之間之結合識別該TCR。識別新抗原特異性TCR之另一種方法可包括獲得包含該TCR之一或多個T-細胞；利用在至少一個抗原呈遞細胞(APC)上呈遞之至少一個新抗原激活該一或多個T-細胞；及經由與至少一個新抗原相互作用激活之一或多個細胞之選擇識別該TCR。

【0545】 於識別新抗原特異性TCR後，可將該TCR定序。於一實施例中，以上關於第XVI節所述之方法可用於對TCR定序。於另一實施例中，可將TCR之TCRa及TCRb批量定序及然後基於頻率配對。於另一實施例中，可使用Howie等人，Science Translational Medicine 2015 (doi: 10.1126/scitranslmed.aac5624)之方法將TCR定序及配對。於另一實施例中，可使用Han等人，Nat Biotech 2014 (PMID 24952902, doi 10.1038/nbt.2938)之方法將TCR定序及配對。於另一實施例中，可使用藉由 <https://www.biorxiv.org/content/early/2017/05/05/134841> 及

<https://patents.google.com/patent/US20160244825A1/>^{158、159}所述之方法獲得成對TCR序列。

【0546】於另一實施例中，T細胞純系群體可藉由限制稀釋產生，及然後可將T細胞純系群體之TCRa及TCRb定序。於又一實施例中，可將T-細胞分類至具有孔之板上使得每孔存在一個T細胞，及然後可將各孔中之各T細胞之TCRa及TCRb定序及配對。

【0547】接下來，於自患者樣品識別新抗原特異性T-細胞及將經識別之新抗原特異性T-細胞之TCR定序後，將經定序之TCR選殖至新的T-細胞中。此等經選殖之T-細胞含有抗原特異性受體，例如，含有包含TCR之細胞外域。亦提供此等細胞之群體及含有此等細胞之組合物。於一些實施例中，將此等細胞之組合物或群體濃化，諸如其中表現TCR之細胞組成組合物或某個類型之細胞(諸如T-細胞或CD8+或CD4+細胞)中之總細胞之至少1、5、10、20、30、40、50、60、70、80、90、91、92、93、94、95、96、97、98、99或超過99%。於一些實施例中，組合物包含至少一個含有本文中所揭示之TCR之細胞。其中組合物為用於投與(諸如用於授受性細胞療法)之醫藥組合物及調配物。亦提供對受試者(例如，患者)投與細胞及組合物之治療方法。

【0548】因此亦提供表現TCR之經基因改造之細胞。該等細胞一般為真核細胞，諸如哺乳動物細胞及通常為人類細胞。於一些實施例中，該等細胞係源自血液、骨髓、淋巴或淋巴器官，為免疫系統之細胞，諸如先天或適應性免疫之細胞，例如，骨髓或淋巴細胞，通常T-細胞及/或NK細胞。其他示例性細胞包括幹細胞，諸如多能及多潛能幹細胞，包括誘導性多潛能幹細胞(iPSC)。該等細胞通常為初始細胞，諸如直接自受試者單離

及/或自受試者單離並冷凍之彼等。於一些實施例中，該等細胞包括T-細胞或其他細胞類型(諸如全T-細胞群體、CD4+細胞、CD8+細胞及其子群體)之一或多個子集，諸如藉由以下所定義之彼等：功能、啟動狀態、成熟、分化潛能、擴展、再循環、局域化及/或持久性能力、抗原特異性、抗原受體之類型、存在於特定器官或隔室中、標誌物或細胞激素分泌特性及/或分化程度。關於所治療之受試者，該等細胞可係異基因及/或自體同源。其中該等方法包括現成方法。於一些態樣中，諸如針對現成技術，該等細胞係多潛能及/或多能，諸如幹細胞，諸如誘導性多潛能幹細胞(iPSC)。於一些實施例中，該等方法包括如上所述自患者單離細胞、製備、處理、培養及/或基因改造細胞，及在冷藏保存之前或之後將細胞再移入相同患者中。

【0549】 其中T-細胞及/或CD4+及/或CD8+ T-細胞之子類型及子群體為初始T (TN)細胞、效應T-細胞(TEFF)、記憶T-細胞及其子類型(諸如幹細胞記憶T (TSCM)、中樞記憶T (TCM)、效應記憶T (TEM)或終期分化之效應記憶T-細胞)、腫瘤浸潤淋巴細胞(TIL)、不成熟T-細胞、成熟T-細胞、輔助T-細胞、細胞毒性T-細胞、黏膜相關不變異T (MALT)細胞、天然存在及適應性調節T (Treg)細胞、輔助T-細胞(諸如TH1細胞、TH2細胞、TH3細胞、TH17細胞、TH9細胞、TH22細胞、濾泡輔助T-細胞、 α/β T-細胞及 δ/γ T-細胞)。

【0550】 於一些實施例中，該等細胞為自然殺手(NK)細胞。於一些實施例中，該等細胞為單核細胞或粒細胞，例如，骨髓細胞、巨噬細胞、嗜中性白血球、樹突狀細胞、肥大細胞、嗜酸性白血球及/或嗜堿細胞。

【0551】 可將該等細胞基因修飾以減少表現或剔除內源性TCR。此

等修飾述於Mol Ther Nucleic Acids. 2012 Dec; 1(12): e63; Blood. 2011年8月11日 ; 118(6):1495-503 ; Blood. 2012年6月14日 ; 119(24): 5697-5705 ; Torikai, Hiroki等人, 「HLA and TCR Knockout by Zinc Finger Nucleases: Toward 「off-the-Shelf」 Allogeneic T-Cell Therapy for CD19+ Malignancies..」 Blood 116.21 (2010): 3766 ; Blood. 2018年1月18 日 ; 131(3):311-322. doi: 10.1182/blood-2017-05-787598 及 WO2016069283中, 其全文以引用的方式併入。

【0552】 可將該等細胞基因修飾以促進細胞激素分泌。此等修飾述於Hsu C、Hughes MS、Zheng Z、Bray RB、Rosenberg SA、Morgan RA. Primary human T lymphocytes engineered with a codon-optimized IL-15 gene resist cytokine withdrawal-induced apoptosis and persist long-term in the absence of exogenous cytokine. J Immunol. 2005;175:7226-34 ; Quintarelli C、Vera JF、Savoldo B、Giordano Attianese GM、Pule M、Foster AE, Co-expression of cytokine and suicide genes to enhance the activity and safety of tumor-specific cytotoxic T lymphocytes. Blood. 2007;110:2793-802 ; 及Hsu C、Jones SA、Cohen CJ、Zheng Z、Kerstann K、Zhou J, Cytokine-independent growth and clonal expansion of a primary human CD8+ T-cell clone following retroviral transduction with the IL-15 gene. Blood. 2007;109:5168-77中。

【0553】 已顯示T-細胞上之趨化因子受體與腫瘤分泌之趨化因子之錯配導致T-細胞至腫瘤微環境之次優遷移。為提高治療功效, 可將該等細胞基因修飾以增加腫瘤微環境中之趨化因子之識別。此等修飾之實例述於

Moon 、 EKCarpenito 、 Csun 、 Jwang 、 LCKapoor 、 VPredina, J Expression of a functional CCR2 receptor enhances tumor localization and tumor eradication by retargeted human T-cells expressing a mesothelin-specific chimeric antibody receptor.Clin Cancer Res. 2011; 17: 4719-4730 ； 及 Craddock 、 JALu 、 Abear 、 Apule 、 Mbrenner 、 MKRooney 、 CM等人， Enhanced tumor trafficking of GD2 chimeric antigen receptor T-cells by expression of the chemokine receptor CCR2b.J Immunother. 2010; 33: 780-788中。

【0554】 可將該等細胞基因修飾以增強共同刺激/增強受體(諸如CD28及41BB)之表現。

【0555】 T-細胞療法之不利作用可包括細胞激素釋放症候群及延長之B-細胞耗盡。於受體細胞中引入自殺/安全開關可提高基於細胞之療法之安全特性。因此，可將該等細胞基因修飾以包含自殺/安全開關。自殺/安全開關可為基於表現基因之細胞賦予對物劑(例如，藥物)敏感性之基因，且當細胞與該物劑接觸或暴露於該物劑中時，該基因造成細胞死亡。示例性自殺/安全開關述於Protein Cell. 2017年8月；8(8): 573-589中。該自殺/安全開關可為HSV-TK。該自殺/安全開關可為胞嘧啶脫胺酶、嘌呤核苷酸磷酸化酶或硝基還原酶。該自殺/安全開關可為美國專利申請公開案第US20170166877A1號中所述之RapaCIDETM。該自殺/安全開關系統可為Haematologica. 2009年9月；94(9): 1316-1320中所述之CD20/利妥昔(Rituximab)。此等參考文獻之全文以引用的方式併入。

【0556】 可將TCR引入受體細胞中作為分裂受體，該分裂受體僅在雜二聚小分子之存在下組裝。此等系統述於Science. 2015年10月16日；

350(6258): aab4077及美國專利案第9,587,020號中，其以引用的方式併入。

【0557】於一些實施例中，該等細胞包含一或多個核酸(例如，編碼本文中所揭示之TCR之多核苷酸)，其中經由基因改造引入該多核苷酸，及從而表現如本文中所揭示之重組或經基因改造之TCR。於一些實施例中，核酸係異源，即，正常不存在於細胞或獲自該細胞之樣品中，諸如獲自另一生物體或細胞者，其(例如)不通常見於經基因改造之細胞及/或衍生此細胞之生物體中。於一些實施例中，核酸係非天然存在(諸如自然界中未發現之核酸)，包括包含編碼來自多個不同細胞類型之各種域之核酸之嵌合組合者。

【0558】核酸可包含經密碼子最佳化之核苷酸序列。不受特定理論或機制之約束，據信核苷酸序列之密碼子最佳化增加mRNA轉錄本之轉譯效率。核苷酸序列之密碼子最佳化可涉及用另一密碼子取代原生密碼子，該另一密碼子編碼相同胺基酸，但是可藉由於細胞內更容易獲得之tRNA轉譯，因此增加轉譯效率。核苷酸序列之最佳化亦可減少將干擾轉譯之二級mRNA結構，因此增加轉譯效率。

【0559】可使用構築體或載體將TCR引入受體細胞中。本文中描述示例性構築體。編碼TCR之 α 及 β 鏈之多核苷酸可於單一構築體中或於分開構築體中。可將編碼 α 及 β 鏈之多核苷酸以操作方式連接至啟動子，例如，異源啟動子。該異源啟動子可為強啟動子，例如，EF1 α 、CMV、PGK1、Ubc、 β 肌動蛋白、CAG啟動子及類似者。該異源啟動子可為弱啟動子。該異源啟動子可為可誘導啟動子。示例性可誘導啟動子包括(但不限於) TRE、NFAT、GAL4、LAC及類似者。其他示例性可誘導表現系統

述於美國專利案第5,514,578號、第6,245,531號、第7,091,038號及歐洲專利案第0517805號，其全文以引用的方式併入。

【0560】 用於將TCR引入受體細胞中之構築體亦可包含編碼信號肽(信號肽元件)之多核苷酸。該信號肽可促進經引入之TCR之表面遷移。示例性信號肽包括(但不限於) CD8信號肽、免疫球蛋白信號肽，其中特定實例包括GM-CSF及IgG κ 。此等信號肽述於Trends Biochem Sci. 2006年10月；31(10):563-71. Epub 2006年8月21日；及An等人「Construction of a New Anti-CD19 Chimeric Antigen Receptor and the Anti-Leukemia Function Study of the Transduced T-cells.」 Oncotarget 7.9 (2016): 10638-10649. PMC. Web. 2018年8月16日中；其以引用的方式併入。

【0561】 於一些情況(例如，自單一構築體或開放閱讀框架表現 α 及 β 鏈之情況或標誌物基因包含於構築體中之情況)下，該構築體可包含核糖體跳躍序列。該核糖體跳躍序列可為2A肽，例如，P2A或T2A肽。示例性P2A及T2A肽述於Scientific Reports第7卷，文章編號：2193 (2017)中，其全文以引用的方式併入。於一些情況下，在2A元件之上游引入FURIN/PACE裂解位點。FURIN/PACE裂解位點述於(例如)<http://www.nuolan.net/substrates.html>中。裂解肽亦可為因子Xa裂解位點。於自單一構築體或開放閱讀框架表現 α 及 β 鏈之情況下，該構築體可包含內部核糖體進入位點(IRES)。

【0562】 該構築體可另外包含一或多個標誌物基因。示例性標誌物基因包括(但不限於) GFP、螢光素酶、HA、lacZ。該標誌物可為可選擇標誌物，諸如抗生素耐藥性標誌物、重金屬耐藥性標誌物或除生物劑耐藥性標誌物，如為熟習此項技術者所知。該標誌物可為於營養缺陷型宿主中

使用之互補標誌物。示例性互補標誌物及營養缺陷型宿主述於Gene. 2001年1月24日；263(1-2):159-69中。此等標誌物可經由IRES、框移序列、2A肽連接子、與TCR之融合表現或自單獨啟動子分開表現。

【0563】 用於將TCR引入受體細胞中之示例性載體或系統包括(但不限於) Adeno相關病毒、腺病毒、腺病毒+經改性之牛痘、安卡拉(Ankara)病毒(MVA)、腺病毒+逆轉錄病毒、腺病毒+仙台(Sendai)病毒、腺病毒+牛痘病毒、 α 病毒(VEE)複製子疫苗、反義寡核苷酸、長雙歧桿菌(*Bifidobacterium longum*)、CRISPR-Cas9、大腸桿菌(*E. coli*)、黃病毒屬(*Flavivirus*)、基因槍(Gene gun)、疱疹病毒屬(*Herpesviruses*)、單純疱疹病毒(*Herpes simplex virus*)、乳酸乳球菌(*Lactococcus lactis*)、電穿孔、慢病毒、脂質轉染法、單核細胞增生李斯特氏菌(*Listeria monocytogenes*)、麻疹病毒(*Measles virus*)、經改性之牛痘安卡拉病毒(MVA)、mRNA電穿孔、裸/質體DNA、裸/質體DNA +腺病毒、裸/質體DNA +經改性之牛痘安卡拉病毒(MVA)、裸/質體DNA + RNA轉移、裸/質體DNA +牛痘病毒、裸/質體DNA +水疱性口炎病毒(*Vesicular stomatitis virus*)、新堡病病毒(*Newcastle disease virus*)、非病毒、PiggyBacTM (PB)轉位子、基於奈米粒子之系統、脊髓灰質炎病毒(*Poliovirus*)、痘病毒(*Poxvirus*)、痘病毒+牛痘病毒、逆轉錄病毒、RNA轉移、RNA轉移+裸/質體DNA、RNA病毒、釀酒酵母(*Saccharomyces cerevisiae*)、鼠傷寒沙門氏桿菌(*Salmonella typhimurium*)、塞姆利基森林病毒(*Semliki forest virus*)、仙台病毒、痢疾志賀氏菌(*Shigella dysenteriae*)、猿腎病毒(*Simian virus*)、siRNA、睡美人(*Sleeping Beauty*)轉位子、變形鏈球菌(*Streptococcus mutans*)、牛痘病毒、委內瑞

拉馬腦炎病毒複製子(Venezuelan equine encephalitis virus replicon)、水
胞性口炎病毒及霍亂弧菌(*Vibrio cholera*)。

【0564】 於較佳實施例中，將TCR經由腺相關病毒(AAV)、腺病
毒、CRISPR-CAS9、疱疹病毒、慢病毒、脂質轉染法、mRNA電穿孔、
PiggyBac™ (PB)轉位子、逆轉錄病毒、RNA轉移或睡美人轉位子引入受
體細胞中。

【0565】 於一些實施例中，用於將TCR引入受體細胞中之載體為病
毒載體。示例性病毒載體包括腺病毒載體、腺相關病毒(AAV)載體、慢病
毒載體、疱疹病毒載體、逆轉錄病毒載體及類似者。本文中描述此等載
體。

【0566】 用於將TCR引入受體細胞中之TCR構築體之示例性實施例
示於圖24中。於一些實施例中，TCR構築體自5'-3'方向包含下列多核苷
酸序列：啟動子序列、信號肽序列、TCR β 可變(TCR β v)序列、TCR β 恆
定(TCR β c)序列、裂解肽(例如，P2A)、信號肽序列、TCR α 可變
(TCR α v)序列及TCR α 恆定(TCR α c)序列。於一些實施例中，構築體之
TCR β c及TCR α c序列包含一或多個鼠科區域，例如，如本文中所述之全鼠
科恆定序列或人類→鼠科胺基酸交換。於一些實施例中，該構築體還包含
3'-TCR α c序列、裂解肽序列(例如，T2A)，接著報告基因。於一實施例
中，該構築體自5'-3'方向包含下列多核苷酸序列：啟動子序列、信號肽
序列、TCR β 可變(TCR β v)序列、含有一或多個鼠科區域之TCR β 恆定
(TCR β c)序列、裂解肽(例如，P2A)、信號肽序列、TCR α 可變(TCR α v)
序列及含有一或多個鼠科區域之TCR α 恆定(TCR α c)序列、裂解肽(例如，
P2A)及報告基因。

【0567】 圖25描述用於將TCR選殖至表現系統中用於療法開發之示例性P526構築體主鏈核苷酸序列。

【0568】 圖26描述用於將患者新抗原特異性TCR，純系型1選殖至表現系統中用於療法開發之示例性構築體序列。

【0569】 圖27描述用於將患者新抗原特異性TCR，純系型3選殖至表現系統中用於療法開發之示例性構築體序列。

【0570】 亦提供編碼TCR之單離核酸、包含該等核酸之載體及包含該等載體及核酸之宿主細胞以及用於產生該等TCR之重組技術。

【0571】 核酸可係重組。可在活細胞外面藉由將天然或合成核酸片段連接至可於活細胞中複製之核酸分子或其複製產物構建重組核酸。出於本文之目的，該複製可為活體外複製或活體內複製。

【0572】 用於TCR之重組產生，可將編碼該TCR之核酸單離並插入可複製載體中用於進一步選殖(即，DNA之擴增)或表現。於一些態樣中，核酸可藉由同源重組(例如，如美國專利案第5,204,244號中所述，其全文以引用的方式併入)產生。

【0573】 許多不同向量於此項技術中已知。向量分量一般包含下列中之一或多者：信號序列、複製起源、一或多個標誌物基因、增強子元件、啟動子及轉錄終止序列，例如，如美國專利案第5,534,615號中所述，其全文以引用的方式併入。

【0574】 適用於表現TCR、抗體或其抗原結合片段之示例性載體或構築體包括(例如) pUC系列(Fermentas Life Sciences)、pBluescript系列(Stratagene, LaJolla, CA)、pET系列(Novagen, Madison, WI)、pGEX系列(Pharmacia Biotech, Uppsala, Sweden)及pEX系列(Clontech, Palo

Alto, CA)。噬菌體載體(諸如AGT10、AGT1 1、AZapII (Stratagene)、AEMBL4及ANM1 149)亦適用於表現本文中所揭示之TCR。

XIX.治療概況流程圖

【0575】 圖28為根據一實施例之方法之流程圖，該方法提供對患者之定製新抗原特異性治療。於其他實施例中，該方法可包括與圖28中所示彼等不同及/或額外步驟。此外，可於各種實施例中以與結合圖28所述之順序不同之順序進行該方法之步驟。

【0576】 如上所述，使用質譜資料訓練呈遞模型2801。獲得患者樣品2802。於一些實施例中，該患者樣品包括腫瘤活組織檢查及/或患者之外周血。為預測將呈遞來自患者樣品之腫瘤抗原肽之可能性，將步驟2802中獲得之患者樣品定序以識別資料以輸入呈遞模型。使用訓練呈遞模型預測來自步驟2802中獲得之患者樣品之腫瘤抗原肽之呈遞可能性2803。基於經預測之呈遞可能性識別患者之治療新抗原2804。接下來，獲得另一患者樣品2805。該患者樣品可包括患者之外周血、腫瘤浸潤淋巴細胞(TIL)、淋巴、淋巴結細胞及/或T-細胞之任何其他來源。針對新抗原特異性T-細胞活體內篩選於步驟2805中獲得之患者樣品。

【0577】 在治療過程之此點，患者可接受T-細胞療法及/或疫苗治療。為接受疫苗治療，識別對患者之T-細胞係特異性之新抗原2814。然後，創建包含經識別之新抗原之疫苗2815。最後，對患者投與該疫苗2816。

【0578】 為接受T-細胞療法，將經歷擴展之新抗原特異性T-細胞及/或新的新抗原特異性T-細胞基因改造。為擴展新抗原特異性T-細胞用於T-細胞療法，將該等細胞簡單擴展2807及融合2808至患者中。

【0579】為基因改造新的新抗原特異性T-細胞用於T-細胞療法，將活體內識別之新抗原特異性T-細胞之TCR定序2809。接下來，將此等TCR序列選殖2810至表現載體中。然後將該表現載體2810轉染2811至新的T-細胞中。擴展該等經轉染之T-細胞2812。及最後，將該等經擴展之T-細胞融合2813至患者中。

【0580】患者可接受T-細胞療法及疫苗療法二者。於一實施例中，患者首先接受疫苗療法，然後接受T-細胞療法。此方法之一個優點為疫苗療法可增加腫瘤特異性T-細胞之數目及藉由可檢測水平之T-細胞識別之新抗原的數目。

【0581】於另一實施例中，患者可接受T-細胞療法，接著接受疫苗療法，其中包含於該疫苗中之抗原決定基之組包括藉由T-細胞療法靶向之抗原決定基中之一或多者。此方法之一個優點為疫苗之投與可促進治療性T-細胞之擴展及持久。

XX.示例電腦

【0582】圖29說明用於實施圖1及3中所示之實體之示例電腦2900。電腦2900包含耦合至晶片組2904之至少一個處理器2902。晶片組2904包含記憶體控制器集線器2920及輸入/輸出(I/O)控制器集線器2922。記憶體2906及圖形配接器2912耦合至記憶體控制器集線器2920，及顯示器2918耦合至圖形配接器2912。存儲裝置2908、指向裝置2914及網路配接器2916耦合至I/O控制器集線器2922。電腦2900之其他實施例具有不同架構。

【0583】存儲裝置2908為非暫時性電腦可讀存儲媒體，諸如硬碟驅動器、光碟唯讀記憶體(CD-ROM)、DVD或固態記憶體設備。記憶體

2906持有藉由處理器2902使用之說明及資料。輸入介面2914為觸屏介面、滑鼠、跟蹤球或其他類型之指向裝置、鍵盤或其一些組合，及用於輸入資料至電腦2900中。於一些實施例中，可配置電腦2900以經由使用者之手勢接受來自輸入介面2914之輸入(例如，命令)。圖形配接器2912顯示顯示器2918上之圖像及其他資訊。網路配接器2916將電腦2900連接至一或多個電腦網路。

【0584】 電腦2900適應執行電腦程式模組以提供本文中所述功能。如本文中所用，術語「模組」係指係指用於提供指定功能之電腦程式邏輯。因此，可在硬體、固件及/或軟體中實施模組。於一實施例中，將程式模組存儲在存儲裝置2908上，載入至記憶體2906中及藉由處理器2902執行。

【0585】 藉由圖1之實體使用之電腦2900之類型可取決於實施例及實體所需之處理能力變化。例如，呈遞識別系統160可於單個電腦2900或彼此通過網路通訊之多個電腦2900中(諸如於伺服器場中)運行。電腦2900可缺少上述組件中之一些，諸如圖形配接器2912及顯示器2918。

【0586】

参考文献

1. Desrichard, A.、Snyder, A.及Chan, T. A. Cancer Neoantigens and Applications for Immunotherapy. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. (2015). doi:10.1158/1078-0432.CCR-14-3175
2. Schumacher, T. N.及Schreiber, R. D. Neoantigens in cancer immunotherapy. Science 348, 69-74 (2015)。
3. Gubin, M. M.、Artyomov, M. N.、Mardis, E. R.及Schreiber, R.

D. Tumor neoantigens: building a framework for personalized cancer immunotherapy. J. Clin. Invest. 125, 3413-3421 (2015)。

4. Rizvi, N. A.等人，Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 348, 124-128 (2015)。

5. Snyder, A.等人，Genetic basis for clinical response to CTLA-4 blockade in melanoma. N. Engl. J. Med. 371, 2189-2199 (2014)。

6. Carreno, B. M.等人，Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T-cells. Science 348, 803-808 (2015)。

7. Tran, E. 等人，Cancer immunotherapy based on mutation-specific CD4+ T-cells in a patient with epithelial cancer. Science 344, 641-645 (2014)。

8. Hacohen, N. 及 Wu, C. J.-Y. 美國專利申請案：0110293637 - COMPOSITIONS AND METHODS OF IDENTIFYING TUMOR SPECIFIC NEOANTIGENS. (A1). 在

<<http://appft1.uspto.gov/netacgi/nph->

Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20110293637.PGNR.>

9. Lundegaard, C.、Hoof, I.、Lund, O. 及 Nielsen, M. State of the art and challenges in sequence based T-cell epitope prediction. Immunome Res. 6 Suppl 2, S3 (2010)。

10. Yadav, M.等人，Predicting immunogenic tumour mutations by

combining mass spectrometry and exome sequencing. Nature 515, 572-576 (2014)。

11. Bassani-Sternberg, M.、Pletscher-Frankild, S.、Jensen, L. J.及Mann, M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol. Cell. Proteomics MCP 14, 658-673 (2015)。

12. Van Allen, E. M.等人，Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science 350, 207-211 (2015)。

13. Yoshida, K.及Ogawa, S. Splicing factor mutations and cancer. Wiley Interdiscip. Rev. RNA 5, 445-459 (2014)。

14. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543-550 (2014)。

15. Rajasagi, M.等人，Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. Blood 124, 453-462 (2014)。

16. Downing, S. R. 等人，美國專利申請案：0120208706 - OPTIMIZATION OF MULTIGENE ANALYSIS OF TUMOR SAMPLES. (A1). 在 <<http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20120208706.PG NR.>>

17. Target Capture for NextGen Sequencing - IDT. 在

<<http://www.idtdna.com/pages/products/nextgen/target-capture>>

18. Shukla, S. A. 等人，Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* 33, 1152-1158 (2015)。

19. Cieslik, M.等人，The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* 25, 1372-1381 (2015)。

20. Bodini, M. 等人，The hidden genomic landscape of acute myeloid leukemia: subclonal structure revealed by undetected mutations. *Blood* 125, 600-605 (2015)。

21. Saunders, C. T.等人，Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinforma. Oxf. Engl.* 28, 1811-1817 (2012)。

22. Cibulskis, K. 等人，Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213-219 (2013)。

23. Wilkerson, M. D.等人，Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. *Nucleic Acids Res.* 42, e107 (2014)。

24. Mose, L. E.、Wilkerson, M. D.、Hayes, D. N.、Perou, C. M.及 Parker, J. S. ABRA: improved coding indel detection via assembly-based realignment. *Bioinforma. Oxf. Engl.* 30, 2813-2815 (2014)。

25. Ye, K.、Schulz, M. H.、Long, Q.、Apweiler, R.及 Ning, Z.

Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinforma. Oxf. Engl.* 25, 2865-2871 (2009)。

26. Lam, H. Y. K. 等人，Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nat. Biotechnol.* 28, 47-55 (2010)。

27. Frampton, G. M.等人，Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* 31, 1023-1031 (2013)。

28. Boegel, S.等人，HLA typing from RNA-Seq sequence reads. *Genome Med.* 4, 102 (2012)。

29. Liu, C.等人，ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res.* 41, e142 (2013)。

30. Mayor, N. P.等人，HLA Typing for the Next Generation. *PloS One* 10, e0127153 (2015)。

31. Roy, C. K.、Olson, S.、Graveley, B. R.、Zamore, P. D.及 Moore, M. J. Assessing long-distance RNA sequence connectivity via RNA-templated DNA-DNA ligation. *eLife* 4, (2015)。

32. Song, L.及Florea, L. CLASS: constrained transcript assembly of RNA-seq reads. *BMC Bioinformatics* 14 Suppl 5, S14 (2013)。

33. Maretty, L.、Sibbesen, J. A.及 Krogh, A. Bayesian transcriptome assembly. *Genome Biol.* 15, 501 (2014)。

34. Pertea, M.等人，StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290-295 (2015)。
35. Roberts, A.、Pimentel, H.、Trapnell, C. 及 Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinforma. Oxf. Engl.* (2011). doi:10.1093/bioinformatics/btr355
36. Vitting-Seerup, K.、Porse, B. T.、Sandelin, A. 及 Waage, J. spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* 15, 81 (2014)。
37. Rivas, M. A. 等人，Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* 348, 666-669 (2015)。
38. Skelly, D. A.、Johansson, M.、Madeoy, J.、Wakefield, J. 及 Akey, J. M. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res.* 21, 1728-1737 (2011)。
39. Anders, S.、Pyl, P. T. 及 Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* 31, 166-169 (2015)。
40. Furney, S. J. 等人，SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov.* (2013). doi:10.1158/2159-8290.CD-13-0330

41. Zhou, Q.等人，A chemical genetics approach for the functional assessment of novel cancer genes. *Cancer Res.* (2015). doi:10.1158/0008-5472.CAN-14-2930
42. Maguire, S. L. 等人，SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* 235, 571-580 (2015)。
43. Carithers, L. J. 等人，A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation Biobanking* 13, 311-319 (2015)。
44. Xu, G.等人，RNA CoMPASS: a dual approach for pathogen and host transcriptome analysis of RNA-seq datasets. *PloS One* 9, e89445 (2014)。
45. Andreatta, M.及Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinforma. Oxf. Engl.* (2015). doi:10.1093/bioinformatics/btv639
46. Jørgensen, K. W.、Rasmussen, M.、Buus, S.及Nielsen, M. NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* 141, 18-26 (2014)。
47. Larsen, M. V.等人，An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur. J. Immunol.* 35, 2295-2303 (2005)。
48. cytotoxic T-cell epitopes: insights obtained from improved

predictions of proteasomal cleavage. Immunogenetics 57, 33-41 (2005)。

49. Boisvert, F.-M. 等人，A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. Mol. Cell. Proteomics 11, M111.011429-M111.011429 (2012)。

50. Duan, F. 等人，Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. J. Exp. Med. 211, 2231-2248 (2014)。

51. Janeway's Immunobiology: 9780815345312: Medicine & Health Science Books，在 Amazon.com，在 <<http://www.amazon.com/Janeways-Immunobiology-Kenneth-Murphy/dp/0815345313>>

52. Calis, J. J. A. 等人，Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. PLoS Comput. Biol. 9, e1003266 (2013)。

53. Zhang, J. 等人，Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science 346, 256-259 (2014)

54. Walter, M. J. 等人，Clonal architecture of secondary acute myeloid leukemia. N. Engl. J. Med. 366, 1090-1098 (2012)。

55. Hunt DF、Henderson RA、Shabanowitz J、Sakaguchi K、Michel H、Sevilir N、Cox AL、Appella E、Engelhard VH. Characterization of peptides bound to the class I MHC molecule HLA-

A2.1 by mass spectrometry. *Science* 1992. 255: 1261-1263。

56. Zarling AL、Polefrone JM、Evans AM、Mikesh LM、Shabanowitz J、Lewis ST、Engelhard VH、Hunt DF. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc Natl Acad Sci U S A*. 2006 年 10 月 3 日；103(40):14889-94。

57. Bassani-Sternberg M、Pletscher-Frankild S、Jensen LJ、Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics*. 2015 年 3 月；14(3):658-73. doi: 10.1074/mcp.M114.042812。

58. Abelin JG、Trantham PD、Penny SA、Patterson AM、Ward ST、Hildebrand WH、Cobbold M、Bai DL、Shabanowitz J、Hunt DF. Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nat Protoc*. 2015 年 9 月；10(9):1308-18. doi: 10.1038/nprot.2015.086. 2015 年 8 月 6 日 电子出版

59. Barnstable CJ、Bodmer WF、Brown G、Galfre G、Milstein C、Williams AF、Ziegler A. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell*. 1978 年 5 月；14(1):9-20。

60. Goldman JM、Hibbin J、Kearney L、Orchard K、Th'ng KH. HLA-DR monoclonal antibodies inhibit the proliferation of normal and

chronic granulocytic leukaemia myeloid progenitor cells. Br J Haematol. 1982年11月 ; 52(3):411-20。

61. Eng JK、Jahan TA、Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. Proteomics. 2013 年 1 月 ; 13(1):22-4. doi: 10.1002/pmic.201200439.2012 年12月4日电子出版。

62. Eng JK、Hoopmann MR、Jahan TA、Egertson JD、Noble WS、MacCoss MJ. A deeper look into Comet--implementation and features. J Am Soc Mass Spectrom. 2015年11月;26(11):1865-74. doi: 10.1007/s13361-015-1179-x. 2015年6月27号电子出版。

63. Lukas Käll、Jesse Canterbury、Jason Weston、William Stafford Noble 及 Michael J. MacCoss. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nature Methods 4:923 - 925，2007年11月

64. Lukas Käll、John D. Storey、Michael J. MacCoss 及 William Stafford Noble. Assigning confidence measures to peptides identified by tandem mass spectrometry. Journal of Proteome Research, 7(1):29-34，2008年1月。

65. Lukas Käll、John D. Storey 及 William Stafford Noble. Nonparametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. Bioinformatics, 24(16):i42-i48，2008年8月。

66. Bo Li 及 C. olin N. Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a referenfe genome.

BMC Bioinformatics, 12:323 , 2011年8月。

67. Hillary Pearson 、 Tariq Daouda 、 Diana Paola Granados 、 Chantal Durette 、 Eric Bonneil 、 Mathieu Courcelles 、 Anja Rodenbrock 、 Jean-Philippe Laverdure 、 Caroline Côté 、 Sylvie Mader 、 Sébastien Lemieux 、 Pierre Thibault 及 Claude Perreault. MHC class I-associated peptides derive from selective regions of the human genome. The Journal of Clinical Investigation, 2016 ,

68. Juliane Liepe 、 Fabio Marino 、 John Sidney 、 Anita Jeko 、 Daniel E. Bunting 、 Alessandro Sette 、 Peter M. Klotzel 、 Michael P. H. Stumpf 、 Albert J. R. Heck 、 Michele Mishto. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. Science, 21 , 2016年10月。

69. Mommen GP. 、 Marino, F. 、 Meiring HD. 、 Poelen, MC. 、 van Gaans-van den Brink, JA. 、 Mohammed S. 、 Heck AJ. 及 van Els CA. Sampling From the Proteome to the Human Leukocyte Antigen-DR (HLA-DR) Ligandome Proceeds Via High Specificity. Mol Cell Proteomics 15(4): 1412-1423 , 2016年4月。

70. Sebastian Kreiter 、 Mathias Vormehr 、 Niels van de Roemer 、 Mustafa Diken 、 Martin Löwer 、 Jan Diekmann 、 Sebastian Boegel 、 Barbara Schrörs 、 Fulvia Vascotto 、 John C. Castle 、 Arbel D. Tadmor 、 Stephen P. Schoenberger 、 Christoph Huber 、 Özlem Türeci 及 Ugur Sahin. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. Nature 520, 692-696 , 2015年4月。

71. Tran E. 、Turcotte S. 、Gros A. 、Robbins P.F. 、Lu Y.C. 、Dudley M.E. 、Wunderlich J.R. 、Somerville R.P. 、Hogan K. 、Hinrichs C.S. 、Parkhurst M.R. 、Yang J.C. 、Rosenberg S.A. Cancer immunotherapy based on mutation-specific CD4+ T-cells in a patient with epithelial cancer. *Science* 344(6184) 641-645 , 2014年5月。

72. Andreatta M. 、Karosiene E. 、Rasmussen M. 、Stryhn A. 、Buus S. 、Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* 67(11-12) 641-650 , 2015年11月。

73. Nielsen, M. 、Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 10:296 , 2009年9月。

74. Nielsen, M. 、Lundegaard, C. 、Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 8:238 , 2007年7月。

75. Zhang, J. 等人 , PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics*. 11(4):1-8. 1/2/2012。

76. Snyder, A.等人 , Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189-2199 (2014)。

77. Rizvi, N. A.等人 , Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348, 124-128 (2015)。

78. Gubin, M. M. 、 Artyomov, M. N. 、 Mardis, E. R.及Schreiber, R. D. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J. Clin. Invest.* 125, 3413-3421 (2015) 。
79. Schumacher, T. N.及Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* 348, 69-74 (2015) 。
80. Carreno, B. M.等人 , Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T-cells. *Science* 348, 803-808 (2015) 。
81. Ott, P. A.等人 , An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 547, 217-221 (2017) 。
82. Sahin, U.等人 , Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* 547, 222-226 (2017) 。
83. Tran, E. 等 人 , T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N. Engl. J. Med.* 375, 2255-2262 (2016) 。
84. Gros, A.等人 , Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat. Med.* 22, 433-438 (2016) 。
85. The problem with neoantigen prediction. *Nat. Biotechnol.* 35, 97-97 (2017) 。
86. Vitiello, A.及Zanetti, M. Neoantigen prediction and the need for validation. *Nat. Biotechnol.* 35, 815-817 (2017) 。
87. Bassani-Sternberg, M. 、 Pletscher-Frankild, S. 、 Jensen, L. J.及

Mann, M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics MCP* 14, 658-673 (2015)。

88. Vita, R. 等人，The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* 43, D405-412 (2015)。

89. Andreatta, M.及Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinforma. Oxf. Engl.* 32, 511-517 (2016)。

90. O'Donnell, T. J.等人，MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. *Cell Syst.* (2018).
doi:10.1016/j.cels.2018.05.014

91. Bassani-Sternberg, M.等人，Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* 7, 13404 (2016)。

92. Abelin, J. G. 等人，Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity* 46, 315-326 (2017)。

93. Yadav, M.等人，Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* 515, 572-576 (2014)。

94. Stranzl, T.、Larsen, M. V.、Lundegaard, C.及Nielsen, M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics* 62, 357-368 (2010)。

95. Bentzen, A. K.等人，Large-scale detection of antigen-specific T-cells using peptide-MHC-I multimers labeled with DNA barcodes. Nat. Biotechnol. 34, 1037-1045 (2016)。
96. Tran, E.等人，Immunogenicity of somatic mutations in human gastrointestinal cancers. Science 350, 1387-1390 (2015)。
97. Stronen, E.等人，Targeting of cancer neoantigens with donor-derived T-cell receptor repertoires. Science 352, 1337-1341 (2016)。
98. Trolle, T.等人，The Length Distribution of Class I-Restricted T-cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference. J. Immunol. Baltim. Md 1950 196, 1480-1487 (2016)。
99. Di Marco, M.等人，Unveiling the Peptide Motifs of HLA-C and HLA-G from Naturally Presented Peptides and Generation of Binding Prediction Matrices. J. Immunol. Baltim. Md 1950 199, 2639-2651 (2017)。
100. Goodfellow, I.、Bengio, Y.及Courville, A. Deep Learning. (MIT Press, 2016)。
101. Sette, A.等人，The relationship between class I binding affinity and immunogenicity of potential cytotoxic T-cell epitopes. J. Immunol. Baltim. Md 1950 153, 5586-5592 (1994)。
102. Fortier, M.-H.等人，The MHC class I peptide repertoire is molded by the transcriptome. J. Exp. Med. 205, 595-610 (2008)。
103. Pearson, H.等人，MHC class I-associated peptides derive

from selective regions of the human genome. *J. Clin. Invest.* 126, 4690-4701 (2016)。

104. Bassani-Sternberg, M.等人，Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allosteric regulating HLA specificity. *PLoS Comput. Biol.* 13, e1005725 (2017)。

105. Andreatta, M.、Lund, O. 及 Nielsen, M. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinforma. Oxf. Engl.* 29, 8-14 (2013)。

106. Andreatta, M.、Alvarez, B. 及 Nielsen, M. GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx248

107. Gros, A. 等人，Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat. Med.* 22, 433-438 (2016)。

108. Zacharakis, N.等人，Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nat. Med.* 24, 724-730 (2018)。

109. Chudley, L.等人，Harmonisation of short-term in vitro culture for the expansion of antigen-specific CD8+ T-cells with detection by ELISPOT and HLA-multimer staining. *Cancer Immunol. Immunother.* 63, 1199-1211 (2014)。

110. Van Allen, E. M.等人，Genomic correlates of response to

CTLA-4 blockade in metastatic melanoma. *Science* 350, 207-211 (2015)。

111. Anagnostou, V. 等人，Evolution of Neoantigen Landscape during Immune Checkpoint Blockade in Non-Small Cell Lung Cancer. *Cancer Discov.* 7, 264–276 (2017)。

112. Carreno, B. M.等人，Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T-cells. *Science* 348, 803-808 (2015)。

113. Stevanović, S. 等人，Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. *Science* 356, 200-205 (2017)。

114. Pasetto, A. 等人，Tumor- and Neoantigen-Reactive T-cell Receptors Can Be Identified Based on Their Frequency in Fresh Tumor. *Cancer Immunol. Res.* 4, 734-743 (2016)。

115. Gillette, M. A.及Carr, S. A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat. Methods* 10, 28-34 (2013)。

116. Boegel, S.、Löwer, M.、Bukur, T.、Sahin, U.及Castle, J. C. A catalog of HLA type, HLA expression, and neo-epitope candidates in human cancer cell lines. *Oncoimmunology* 3, e954893 (2014)。

117. Johnson, D. B.等人，Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nat. Commun.* 7, 10582 (2016)。

118. Robbins, P. F. 等人，A Pilot Trial Using Lymphocytes Genetically Engineered with an NY-ESO-1-Reactive T-cell Receptor: Long-term Follow-up and Correlates with Response. *Clin. Cancer Res.* 21, 1019-1027 (2015)。

119. Snyder, A.等人，Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189-2199 (2014)。

120. Calis, J. J. A. 等人，Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput. Biol.* 9, e1003266 (2013)。

121. Duan, F. 等人，Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* 211, 2231–2248 (2014)。

122. Glanville, J.等人，Identifying specificity groups in the T-cell receptor repertoire. *Nature* 547, 94-98 (2017)。

123. Dash, P.等人，Quantifiable predictive features define epitope-specific T-cell receptor repertoires. *Nature* 547, 89-93 (2017)。

124. Hunt, D. F.等人，Pillars article: Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* 1992. 255: 1261-1263. *J. Immunol. Baltim. Md* 1950 179, 2669-2671 (2007)。

125. Zarling, A. L.等人，Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14889-14894 (2006)。

126. Abelin, J. G.等人，Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. Nat. Protoc. 10, 1308-1318 (2015)。

127. Barnstable, C. J.等人，Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. Cell 14, 9-20 (1978)。

128. Eng, J. K.、Jahan, T. A.及Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. Proteomics 13, 22-24 (2013)。

129. Eng, J. K.等人，A deeper look into Comet--implementation and features. J. Am. Soc. Mass Spectrom. 26, 1865-1874 (2015)。

130. Käll, L.、Storey, J. D.、MacCoss, M. J.及Noble, W. S. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. J. Proteome Res. 7, 29-34 (2008)。

131. Käll, L.、Storey, J. D.及Noble, W. S. Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. Bioinforma. Oxf. Engl. 24, i42-48 (2008)。

132. Käll, L.、Canterbury, J. D.、Weston, J.、Noble, W. S.及MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat. Methods 4, 923-925 (2007)。

133. Li, B.及Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.

BMC Bioinformatics 12, 323 (2011)。

134. Chollet, F.及他人，Keras. (2015)。

135. Bastien, F.等人，Understanding the difficulty of training deep feedforward neural networks. Proc. Thirteenth. Int. Conf. Artif. Intell. Stat. 249-256 (2010)。

136. Glorot, X. 及 Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics 249-256 (2010)。

137. Kingma, D. 及 Ba, J. Adam: A method for stochastic optimization. ArXiv Prepr. ArXiv14126980 (2014)。

138. Schneider, T. D.及Stephens, R. M. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18, 6097-6100 (1990)。

139. Rubinsteyn, A. 、 O'Donnell, T. 、 Damaraju, N. 及 Hammerbacher, J. Predicting Peptide-MHC Binding Affinities With Imputed Training Data. biorxiv (2016).
doi:<https://doi.org/10.1101/054775>

140. Tran, E.等人，Immunogenicity of somatic mutations in human gastrointestinal cancers. Science 350, 1387-1390 (2015)。

141. Stronen, E. 等人，Targeting of cancer neoantigens with donor-derived T-cell receptor repertoires. Science 352, 1337-1341 (2016).

142. Janetzki, S. 、 Cox, J. H. 、 Oden, N. 及 Ferrari, G. Standardization and validation issues of the ELISPOT assay. *Methods Mol. Biol.* Clifton NJ 302, 51-86 (2005) 。
143. Janetzki, S.等人， Guidelines for the automated evaluation of Elispot assays. *Nat. Protoc.* 10, 1098-1115 (2015) 。
144. Li, H.及 Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* 25, 1754-1760 (2009) 。
145. DePristo, M. A.等人， A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491-498 (2011) 。
146. Garrison, E.及 Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* (2012) 。
147. Cingolani, P.等人， A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80-92 (2012) 。
148. Szolek, A.等人， OptiType: precision HLA typing from next-generation sequencing data. *Bioinforma. Oxf. Engl.* 30, 3310-3316 (2014) 。
149. Cibulskis, K. 等人， Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213-219 (2013) 。

150. Scholz, E. M. 等人， Human Leukocyte Antigen (HLA)-DRB1*15:01 and HLA-DRB5*01:01 Present Complementary Peptide Repertoires. *Front. Immunol.* 8, 984 (2017)。

151. Ooi, J. D. 等人， Dominant protection from HLA-linked autoimmunity by antigen-specific regulatory T-cells. *Nature* 545, 243-247 (2017)。

152. Karosiene, E.等人， NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics* 65, 711-724 (2013)。

153. Dudley ME、Gross CA、Lanahan MM等人， CD8+ enriched 「 young 」 tumor infiltrating lymphocytes can mediate regression of metastatic melanoma. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2010;16(24):6122-6131. doi:10.1158/1078-0432.CCR-10-1297。

154. Dudley ME、Wunderlich JR、Shelton TE、Even J、Rosenberg SA. Generation of Tumor-Infiltrating Lymphocyte Cultures for Use in Adoptive Transfer Therapy for Melanoma Patients. *Journal of immunotherapy (Hagerstown, Md : 1997).* 2003;26(4):332-342。

155. Cohen CJ、Gartner JJ、Horovitz-Fried M等人， Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. *The Journal of Clinical Investigation.* 2015;125(10):3981-3991. doi:10.1172/JCI82416。

156. Kelderman, S. 、Heemskerk, B. 、Fanchi, L. 、Philips, D. 、Toebe, M. 、Kvistborg, P. 、Buuren, M. M. 、Rooij, N. 、Michels, S. 、Germeroth, L. 、Haanen, J. B. 及 Schumacher, N. M. (2016), Antigen - specific TIL therapy for melanoma: A flexible platform for personalized cancer immunotherapy. Eur. J. Immunol., 46: 1351-1360. doi:10.1002/eji.201545849 。

157. Hall M 、Liu H 、Malafa M 等人 , Expansion of tumor-infiltrating lymphocytes (TIL) from human pancreatic tumors. Journal for Immunotherapy of Cancer. 2016;4:61. doi:10.1186/s40425-016-0164-7 。

158. Briggs A 、Goldfless S 、Timberlake S 等人 , Tumor-infiltrating immune repertoires captured by single-cell barcoding in emulsion. bioRxiv. 2017. doi.org/10.1101/134841 。

159.美國專利申請案第20160244825A1號 。

160. Koşaloğlu-Yalçın, Z.等人 , Predicting T cell recognition of MHC class I restricted neoepitopes. J. OncoImmunology, 1-15 (2018) 。

補充表1

具有先已存在之反應之突變之預測排名

突變ID	患者ID	MHCFlurry ， TPM >0	MHCFlurry ， TPM >1	MHCFlurry ， TPM >2	肽MS模型， TPM > 1	全MS模型
KARS_D356H	3942	81	44	36	26	5
NUP98_A359D	3942	13	8	7	0	0
CASP8_F67V	3971	13	3	2	3	1
KRAS_G12D	3995	36	21	18	2	2
RNF213_N1702S	3995	0	0	0	7	7
TUBGCP2_P293	3995	2	2	2	8	6
H3F3B_A48T	4007	33	23	21	13	0
SKIV2L_R653H	4007	2	1	1	15	17
API5_R243Q	4032	52	31	27	10	1
PHLPP1_G566E	4032	54	33	29	72	67
RNF10_E572K	4032	43	23	22	46	46
ZFYVE27_R6H	4069	35	23	22	0	0
CADPS2_R1266H	4136	23	22	22	4	5
KIAA0368_S186F	4136	2	2	2	1	0
FLNA_R2049C	NCI-3784	91	85	81	31	5

補充表1

具有先已存在之反應之突變之預測排名

突變ID	患者ID	MHCFlurry ， TPM >0	MHCFlurry ， TPM >1	MHCFlurry ， TPM >2	肽MS模型， TPM > 1	全MS模型
KIF16B_L1009P	NCI-3784	22	21	19	74	69
SON_R1927C	NCI-3784	37	35	32	105	83
KIF1BP_P246S	NCI-3903	66	35	32	22	7
MAGEA6_E168K	NCI-3998	15	10	9	1	0
MED13_P1691S	NCI-3998	5	3	2	0	1
PDS5A_Y1000F	NCI-3998	13	8	7	6	4
CDK4_R71L	患者1	56	23	20	5	0
DNAH17_H8302Y	患者1	42	80	59	112	77
GCN1_L2330P	患者1	59	25	22	3	1
BRWD1_R925W	患者2	80	62	58	74	75
PARG_Y427N	患者2	88	69	65	51	49
	中值	35.5	23	21.5	9	5

補充表2

NSCLS患者之人口統計學

患者ID	年齡範圍(年)	性別	種族	初始(肺癌)診斷年份	腫瘤階段(募集時)	原發腫瘤位置	組織學類型
1-001-002	81-90	男性	白種人	2010	IIIB	肺	非鱗狀 類肉瘤肺癌
1-024-001	81-90	男性	白種人	2016	IV	肺	
1-024-002	51-60	女性	白種人	2016	IV	肺	腺癌
1-038-001	61-70	男性	白種人	2016	IV	肺	腺癌

補充表2

NSCLS患者之人口統計學

全身NSCLC導引療法	目前抗PD(L)-1療法	HLA-A	HLA-A	HLA-B	HLA-B	HLA-C	HLA-C	表現之突變
卡鉑(Carboplatin)	納武單抗(Nivolumab)	A*01:01	A*01:01	B*08:01	B*51:01	C*01:02	C*07:01	122
	派姆單抗(Pembrolizumab)	A*32:01	A*03:01	B*27:05	B*27:05	C*02:02	C*02:02	83
多西他奇(DOCEtaxel)、 貝伐單抗(Bevacizumab)、 雷莫蘆單抗(Ramucirumab)、 培美曲塞(Pemetrexed)二鈉	納武單抗	A*68:01	A*68:01	B*40:02	B*40:27	C*03:04	C*03:04	38
培美曲塞(premetexed)、 順鉑(Cisplatin)	納武單抗	A*69:01	A*01:02	B*41:01	B*49:01	C*17:01	C*07:01	158

補充表2

NSCLS患者之人口統計學

非同義 突變	正常DNA中值 外顯子覆蓋度	腫瘤DNA中值外 顯子覆蓋度	RNA PF獨 特閱讀(M)	已知驅動子	可能驅動子	中值VAF
232	145	552	173	KRAS_G12D、 TP53_R213*	STK11_G52fs	0.22
143	165	508	131.9	KRAS_G12C、 TP53_R280T	PML_E43*、 NF2_R341*	0.093
69	190	454	114.4	KRAS_G12S、 TP53_Q331*	STK11_E199*	0.182
265	158	983	311.8	KRAS_G12V	KDM5C_E303*	0.19

補充表2

NSCLS患者之人口統計學

患者ID	年齡範圍 (歲)	性別	種族	初始(肺癌)診斷之年份	腫瘤階段(募集時)	原發腫瘤之 位置	組織學類型
1-050-001	71-80	女性	白種人	2015	IIIB	肺	腺癌
CU05	71-80	女性	白種人	2013	IV	肺	肺鱗狀
CU04	61-70	女性	西班牙裔或拉丁美洲人	2013	I	肺	腺癌
CU03	61-70	男性	非洲人 美國人	2016	I	肺	肺鱗狀
CU02	61-70	男性	白種人	2016	I	肺	肺鱗狀

補充表2

NSCLS患者之人口統計學

全身NSCLC導 引療法	目前抗PD(L)- 1療法	HLA-A	HLA-A	HLA-B	HLA-B	HLA-C	HLA-C	表現之突變
依託泊苷 (ETOPOSIDE)、 順鉑	納武單抗	A*29:02	A*26:01	B*44:03	B*07:05	C*16:01	C*15:05	53
卡鉑加培美曲塞	納武單抗	A*24:02	A*68:02	B*14:02	B*15:17	C*07:01	C*08:02	65
	德瓦魯單抗 (durvalumab)加 曲美目單抗 (tremelimumab)	A*24:26	A*26:01	B*18:01	B*38:01	C*12:03	C*12:03	336
	n/a	A*23:01	A*01:01	B*08:01	B*15:03	C*01:02	C*12:03	105
卡鉑+吉西他濱 (gemcitabine)	n/a	A*02:01	A*03:01	B*07:02	B*57:01	C*07:02	C*06:02	102

補充表2

NSCLS患者之人口統計學

	正常DNA中 值外顯子覆 蓋度	腫瘤DNA中 值外顯子覆 蓋度	RNA PF 獨特閱讀 (M)			
非同義突變				已知驅動子	可能驅動子	中值VAF
92	117	556	119			0.059
109	191	448	83.6			0.095
511	213	552	240.4	TP53_R158G	NFKBIE_G41fs、 CDH1_Q346*、 NF1_D2163fs、 MED12_R730*	0.224
187	114	830	182.1			0.242
174	105	738	185.3	TP53_R175H	ATR_Q195*	0.32

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
1-001-002	HSPFTATSL	N		1-001-002_池_1	N	chr15_28215653_C_A
1-001-002	DPEEVLVTV	N		1-001-002_池_1	N	chr17_59680958_C_T
1-001-002	ELDPDIQLEY	N		1-001-002_池_1	N	chr13_30210371_C_A
1-001-002	TPLTKDVTL	N		1-001-002_池_1	N	chr5_78100974_A_T
1-001-002	DGVGKSAL	N		1-001-002_池_1	N	chr12_25245350_C_T
1-001-002	YTTVRALTL	N		1-001-002_池_1	N	chr17_28339664_G_T
1-001-002	TPSAAVKLI	N		1-001-002_池_1	N	chr15_81319417_T_C
1-001-002	WPVLLLN	N		1-001-002_池_1	N	chr3_179025167_AAC_A
1-001-002	ELNARRCSF	N		1-001-002_池_1	N	chr18_79943341_G_A

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變類型	基因	蛋白質 效應	TPM	藉由全MS模型覆蓋 之最可能限制	全MS 模型排 名	MHCFlurry 排名	MHCFlurry (nM)	藉由MHCFlurry 覆蓋之最可能限 制
snp	HERC2	A2060S	41.9	HLA-C*01:02	0	95	5169.68205	假
snp	CLTC	S989L	272.1	HLA-B*51:01	1	61	3455.25069	真
snp	KATNAL 1	D407Y	12.81	HLA-A*01:01	2	1	24.2177849	真
snp	AP3B1	S817T	44.4	HLA-B*08:01	3	2	48.9740194	真
snp	KRAS	G12D	40.75	HLA-B*08:01	4	89	4714.29522	真
snp	TNFAIP1	R48L	45.62	HLA-B*08:01	5	26	973.417701	真
snp	STARD5	M108V	1.95	HLA-B*51:01	6	39	2030.48603	真
del_fs	ZMAT3	V240fs	14.99	HLA-B*51:01	7	16	600.564752	真
snp	PQLC1	R109C	33.89	HLA-B*08:01	8	5	62.0439997	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應 (任何時間點)	個別肽反應註釋	池ID	池反應 (任何時間點)	突變
1-001-002	QMKNPILEL	N		1-001-002_池_1	N	chr9_127663287_G_T
1-001-002	LTEKVSLLK	N		1-001-002_池_2	N	chr9_92719180_C_T
1-001-002	SPFTATSL	N		1-001-002_池_2	N	chr15_28215653_C_A
1-001-002	NVDMRTISF	N		1-001-002_池_2	N	chr9_121353262_T_A
1-001-002	TSIVVSQTL	N		1-001-002_池_2	N	chr4_39205691_C_T
1-001-002	HIKIEPVAI	N		1-001-002_池_2	N	Chr13_73062087_C_T
1-001-002	DSPDGSNGL	N		1-001-002_池_2	N	chr20_44197575_C_T
1-001-002	YTAVHYAASY	N		1-001-002_池_2	N	chr12_56248788_C_A
1-001-002	VGADGVGKSAL	N		1-001-002_池_2	N	chr12_25245350_C_T

補充表3
用於NSCLC患者中之T-細胞識別所測試之肽

突變類型	基因	蛋白質 效應	TPM	藉由全MS模 型覆蓋之最可 能限制	全MS 模型排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可能 限制
snp	STXBP1	R171L	38.76	HLA-B*08:01	9	20	674.64733	真
snp	BICD2	E489K	42.66	HLA-A*01:01	10	10	428.744925	真
snp	HERC2	A2060S	41.9	HLA-B*08:01	11	4	59.1155419	真
snp	STOM	K93N	360.6	HLA-B*08:01	12	30	1490.72261	真
snp	WDR19	A282V	18.12	HLA-B*08:01	13	176	9862.33009	真
snp	KLF5	T163I	25.77	HLA-B*08:01	14	27	1122.27455	真
snp	OSER1	S119N	20.7	HLA-C*01:02	15	471	21598.414	假
snp	ANKRD5 2	A559S	18.32	HLA-A*01:01	16	0	11.5906737	真
snp	KRAS	G12D	40.75	HLA-C*01:02	17	370	17985.3612	假

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
1-001-002	MMPPLPGI	N		1-001-002_池_2	N	chr17_32369404_A_T
1-001-002	FPYPGMTNQ	N		1-001-002_池_2	N	chr5_109186272_G_T
1-024-001	VTNHAPLSW	N		1-024-001_池_1	Y	chr3_125552370_C_A
1-024-001	GTKKDVDVLK	Y		1-024-001_池_1	Y	chr20_56513366_G_A
1-024-001	GLNVPVQSNK	N		1-024-001_池_1	Y	chr4_88390868_G_T
1-024-001	VVVGACGVGK	N		1-024-001_池_1	Y	chr12_25245351_C_A
1-024-001	AQFAGKDQTY	N		1-024-001_池_1	Y	chr9_89045819_C_A
1-024-001	KVVLPDVTSY	N		1-024-001_池_1	Y	chr3_48591778_G_T
1-024-001	MLMKNISTK	N		1-024-001_池_1	Y	chr12_6959976_G_A

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變類型	基因	蛋白質 效應	TPM	藉由全MS模 型覆蓋之最可 能限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可 能限制
snp	ZNF207	Q409L	186	HLA-B*51:01	18	136	7609.76602	真
snp	FER	C759F	67.36	HLA-B*51:01	19	38	1999.07208	真
snp	OSBPL1 1	G489W	24.12	HLA-A*32:01	0	7	77.009026	真
snp	RTFDC1	E177K	61.32	HLA-A*03:01	1	70	2168.51668	真
snp	HERC6	R218L	8.7	HLA-A*03:01	2	4	59.675168	真
snp	KRAS	G12C	40.05	HLA-A*03:01	3	11	133.648023	真
snp	SHC3	E376D	8.88	HLA-A*32:01	4	91	3715.42819	真
snp	COL7A1	R468S	25.42	HLA-A*32:01	6	85	3234.15772	真
snp	PTPN6	E471K	105.4	HLA-A*03:01	7	0	12.2301919	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
1-024-001	DLAGGTFDV	N		1-024-001_池_1	Y	chr11_123059991_C_G
1-024-001	LIFDLAGGTF	N		1-024-001_池_1	Y	chr11_123059991_C_G
1-024-001	NVLIFDLA	N		1-024-001_池_1	Y	chr11_123059991_C_G
1-024-001	VVGACGVGKN	N		1-024-001_池_2	N	chr12_25245351_C_A
1-024-001	VIMLNGTKK	N		1-024-001_池_2	N	chr20_56513366_G_A
1-024-001	LAGGTFDV	N		1-024-001_池_2	N	chr11_123059991_C_G
1-024-001	LRNSGGEVF	N		1-024-001_池_2	N	chr14_80906012_TC_T
1-024-001	VVLPSDVTSYN	N		1-024-001_池_2	N	chr3_48591778_G_T
1-024-001	IFDLAGGTF	N		1-024-001_池_2	N	chr11_123059991_C_G

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變類型	基因	蛋白質 效應	TPM	藉由全MS模型 覆蓋之最可能 限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可 能限制
snp	HSPA8	G201A	736.6	HLA-B*27:05	9	353	18290.7955	真
snp	HSPA8	G201A	736.6	HLA-C*02:02	11	57	1716.74204	假
snp	HSPA8	G201A	736.6	HLA-A*32:01	17	621	27984.1357	真
snp	KRAS	G12C	40.05	HLA-A*03:01	5	19	197.846108	真
snp	RTFDC1	E177K	61.32	HLA-A*03:01	8	10	122.750322	真
snp	HSPA8	G201A	736.6	HLA-C*02:02	10	632	28384.8834	假
del_fs	CEP128	R102fs	11.31	HLA-B*27:05	12	46	1020.95087	真
snp	COL7A1	R468S	25.42	HLA-A*32:01	13	62	1925.29397	真
snp	HSPA8	G201A	736.6	HLA-C*02:02	14	427	21255.2074	假

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別 肽反 應(任 何時	個別 肽反 應註 釋	池ID	池反 應(任 何時 間點)	突變
1-024-001	GLLDEAKRLLY	N		1-024-001_池_2	N	chr19_57575861_G_T
1-024-001	SVLLPENYITK	N		1-024-001_池_2	N	chr11_122789248_G_T
1-024-001	DLAGGTFDVS	N		1-024-001_池_2	N	chr11_123059991_C_G
1-024-001	IFDLAGGTFDV	N		1-024-001_池_2	N	chr11_123059991_C_G
1-024-002	AEWRNGSTSS	N		1-024-002_池_1	Y	chr3_122703943_C_G
	L					
1-024-002	YVSEKDVISAK	N		1-024-002_池_1	Y	chr2_43889858_G_A
1-024-002	EGSLGISHTR	N		1-024-002_池_1	Y	chr18_62157782_C_A
1-024-002	IPASVSAPK	N		1-024-002_池_1	Y	chr13_109784018_C_A
1-024-002	QDVSVQVER	Y		1-024-002_池_1	Y	chr9_64411223_T_G

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白質 效應	TPM	藉由全MS模 型覆蓋之最可 能限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry覆 蓋之最可能限 制
snp	ZNF416	Q49K	11.89	HLA-A*03:01	15	24	354.82068	真
snp	UBASH3 B	G307V	12.11	HLA-A*03:01	16	23	228.127132	真
snp	HSPA8	G201A	736.6	HLA-A*32:01	18	487	23357.3292	真
snp	HSPA8	G201A	736.6	HLA-C*02:02	19	563	25887.4267	假
snp	PARP14	P1095A	129.5	HLA-A*68:01	0	8	126.397714	真
snp	LRPPRC	T1335I	79.08	HLA-A*68:01	1	9	136.482978	真
snp	PIGN	W83L	20.74	HLA-A*68:01	2	6	88.2623459	真
snp	IRS2	S679I	63.55	HLA-A*68:01	3	16	224.278982	真
snp	ANKRD2 0A4	M646R	8.92	HLA-A*68:01	4	14	193.974327	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
1-024-002	LVVVGASGVGK	N		1-024-002_池_1	Y	Chr12_25245351_C_T
1-024-002	RATIVPEL	N		1-024-002_池_1	Y	chr7_131463253_A_T
1-024-002	SSAAAPFPL	Y		1-024-002_池_1	Y	chr6_13711102_T_A
1-024-002	GVSKIIGGNPK	N		1-024-002_池_1	Y	chr4_10116175_C_T
1-024-002	EQNFVSTSDIK	未個別測試		1-024-002_池_1	Y	chr3_25791346_A_C
1-024-002	RTQDVSVQVE					
1-024-002	R	N		1-024-002_池_2	Y	chr9_64411223_T_G
1-024-002	EAGNNSRVPR	N		1-024-002_池_2	Y	chr2_74046630_G_T
1-024-002	RYVLHVVA	N		1-024-002_池_2	Y	chr3_122703943_C_G
1-024-002	VSKIIGGNPK	N		1-024-002_池_2	Y	chr4_10116175_C_T

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變類型	基因	蛋白質效應	TPM	藉由全MS模型覆蓋之最可能限制	全MS模型排名	MHCFlurry排名	MHCFlurry (nM)	藉由MHCFlurry覆蓋之最可能限制
snp	KRAS	G12S	72.77	HLA-A*68:01	6	41	1238.56407	真
snp	MKLN1	D521V	84.08	HLA-C*03:04	7	266	16010.7063	假
snp	RANBP9	H135L	43.5	HLA-C*03:04	8	103	4565.97417	假
snp	WDR1	D26N	134.5	HLA-A*68:01	9	125	6797.60699	真
snp	OXSM	K109T	12.82	HLA-A*68:01	17	156	9099.70986	真
snp	ANKRD20A4	M646R	8.92	HLA-A*68:01	5	53	1847.42359	真
snp	TET3	G238V	56.35	HLA-A*68:01	10	13	161.242762	真
snp	PARP14	P1095A	129.5	HLA-A*68:01	11	176	10453.627	真
snp	WDR1	D26N	134.5	HLA-A*68:01	12	38	954.724495	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應 (任何時間點)	個別肽反應 註釋	池ID	池反應(任 何時間點)	突變
1-024-002	QPSGVPTSL	N		1-024-002_池_2	Y	chr12_14478436_GG_T T
1-024-002	DVSVQVER	N		1-024-002_池_2	Y	chr9_64411223_T_G
1-024-002	FVSTSDIKSM	Y		1-024-002_池_2	Y	chr3_25791346_A_C
1-024-002	FPVVNSHSL	N		1-024-002_池_2	Y	chr1_116062776_G_C
1-024-002	APFPLGDSAL	N		1-024-002_池_2	Y	chr6_13711102_T_A
1-024-002	ATIVPELNEI	N		1-024-002_池_2	Y	chr7_131463253_A_T
1-038-001	QEFAPLGTV	N	見池結果	1-038-001_池_1	Y	chr2_219501883_G_T
1-038-001	MNQVLHAY	未個別測試	見池結果	1-038-001_池_1	Y	chr14_100354547_C_G
1-038-001	HEDVKEAI	未個別測試	見池結果	1-038-001_池_1	Y	chr8_96231911_C_G

補充表3
用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白質 效應	TPM	藉由全MS模 型覆蓋之最可 能限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可 能限制
mnp	ATF7IP	G1021L	123.2	HLA-A*68:01	13	139	7795.97025	真
snp	ANKRD2 0A4	M646R	8.92	HLA-A*68:01	14	7	123.489687	真
snp	OXSM	K109T	12.82	HLA-C*03:04	15	128	7025.56581	假
snp	SLC22A1 5	A396P	8.57	HLA-C*03:04	16	155	9082.40652	假
snp	RANBP9	H135L	43.5	HLA-A*68:01	18	196	11590.601	真
snp	MKLN1	D521V	84.08	HLA-A*68:01	19	365	19785.1419	真
snp	GMPPA	G92V	21.6	HLA-B*49:01	0	31	3481.07375	假
snp	WARS	D148H	757.2	HLA-C*07:01	12	422	27180.1513	假
snp	UQCRB	D41H	174.8	HLA-B*49:01	16	300	24830.2411	假

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
1-038-001	GPYPFVQAV	未個別測試	見池結果	1-038-001_池_1	Y	chr1_111242326_C_T
1-038-001	YEHEDVKEAI	未個別測試	見池結果	1-038-001_池_1	Y	chr8_96231911_C_G
1-038-001	EESVMLLTV	未個別測試	見池結果	1-038-001_池_1	Y	chr1_15583354_CC_AG
1-038-001	IEEDSAEKI	未個別測試	見池結果	1-038-001_池_1	Y	chr6_84215849_C_A
1-038-001	TEEDVKIKF	未個別測試	見池結果	1-038-001_池_1	Y	chr7_93105459_C_A
1-038-001	NEQSKLLKV	未個別測試	見池結果	1-038-001_池_1	Y	chrX_70375298_C_G
1-038-001	VDNIIQSI	未個別測試	見池結果	1-038-001_池_1	Y	chr20_2654879_G_T
1-038-001	YEHEDVKEA	Y		1-038-001_池_2	Y	chr8_96231911_C_G
1-038-001	YVSEVPVSV	未個別測試		1-038-001_池_2	Y	chr17_2330604_G_A

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白質效應	TPM	藉由全MS模型覆蓋之 最可能限制	全MS 模型排名	MHCFlurry 排名	MHCFlurry (nM)	藉由MHCFlurry覆蓋 之最可能限制
snp	CHI3L2	L379F	122.3	HLA-B*49:01	1	19	1176.97782	假
snp	UQCRB	D41H	174.8	HLA-B*49:01	2	212	22559.0306	假
mnp	AGMAT	G105L	1.03	HLA-B*49:01	3	109	17185.8013	假
snp	CEP162	E82D	15.62	HLA-B*49:01	4	171	20568.515	假
snp	SAMD9	M213I	68.23	HLA-B*49:01	5	226	22894.2742	假
snp	KIF4A	L625V	19.51	HLA-B*49:01	6	141	19054.8385	假
snp	NOP56	M167I	89.39	HLA-B*49:01	7	119	17928.6022	假
snp	UQCRB	D41H	174.8	HLA-B*49:01	9	250	23419.567	假
snp	TSR1	H561Y	48.21	HLA-C*17:01	10	0	6.07874308	假

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應 (任何時間)	個別肽反應註 釋	池ID	池反應 (任何時間 點)	突變
1-038-001	SELTVHQRI	未個別 測試		1-038-001_池_2	Y	chr19_37564705_G_C
1-038-001	VGVGKSAL	未個別 測試		1-038-001_池_2	Y	chr12_25245350_C_A
1-038-001	DMNQVLHAY	未個別 測試		1-038-001_池_2	Y	chr14_100354547_C_G
1-038-001	NEKGKAUY	未個別 測試		1-038-001_池_2	Y	chr17_51294040_G_T
1-038-001	TEYKLVVVGAV	未個別 測試		1-038-001_池_2	Y	chr12_25245350_C_A
1-038-001	QEFAPLGTVG	未個別 測試		1-038-001_池_2	Y	chr2_219501883_G_T
1-038-001	QEVRLTLLNV	未個別 測試		1-038-001_池_2	Y	chr17_4085728_C_A
1-038-001	VEMLGLISC	未個別 測試		1-038-001_池_2	Y	chr4_168427109_C_A
1-050-001	LFHDMNVSY	N		1-050-001_池_1	N	chr1_193097666_T_C

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白質效應	TPM	藉由全MS模型覆蓋之最 可能限制	全MS 模型排名	MHCFlurry 排名	MHCFlurry (nM)	藉由MHCFlurry覆蓋 之最可能限制
snp	ZNF571	L575V	19.07	HLA-B*49:01	11	159	19886.0407	假
snp	KRAS	G12V	91.89	HLA-C*17:01	13	388	26432.7668	假
snp	WARS	D148H	757.2	HLA-C*07:01	14	64	10286.4383	假
snp	UTP18	M547I	63.21	HLA-C*07:01	15	339	25564.2874	假
snp	KRAS	G12V	91.89	HLA-B*49:01	17	233	23113.572	假
snp	GMPPA	G92V	21.6	HLA-B*49:01	18	338	25558.5468	假
snp	ZZEF1	G863V	63	HLA-B*49:01	19	124	18359.7482	假
snp	DDX60L	A631S	44.71	HLA-B*49:01	8	267	23949.2398	假
snp	GLRX2	N94S	17.92	HLA-A*29:02	0	1	44.54051	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
1-050-001	ISTFRQCAL	未個別測試		1-050-001_池_1	N	chr17_80346815_G_T
1-050-001	YNTDDIEFY	未個別測試		1-050-001_池_1	N	chr15_26580447_G_T
1-050-001	EETPPFSNY	N		1-050-001_池_1	N	chr21_31266125_T_A
1-050-001	QASGNHHVW	未個別測試		1-050-001_池_1	N	chr22_30893501_T_C
1-050-001	EEVTPILAI	未個別測試		1-050-001_池_1	N	chr18_5419733_G_A
1-050-001	IEHNIRNAKY	未個別測試		1-050-001_池_1	N	chr3_52617347_T_G
1-050-001	AERLDVKAI	未個別測試		1-050-001_池_1	N	chr14_103339252_G_T
1-050-001	LFQQGKDLQQY	未個別測試		1-050-001_池_1	N	chr17_80346815_G_T
1-050-001	DTSPVAVAL	未個別測試		1-050-001_池_1	N	chr5_73074790_T_C

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白質 效應	TPM	藉由全MS模 型覆蓋之最可 能限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可 能限制
snp	RNF213	R2827L	330.6	HLA-C*16:01	10	322	22721.4424	假
snp	GABRB3	T185N	2.2	HLA-A*29:02	16	20	447.152559	真
snp	TIAM1	Y283F	13.99	HLA-B*44:03	1	26	537.02592	真
snp	OSBP2	Y677H	7.86	HLA-B*44:03	19	109	7506.81856	真
snp	EPB41L3	S495L	51.69	HLA-B*44:03	2	17	390.306194	真
snp	PBRM1	D578A	65.68	HLA-B*44:03	3	10	186.953378	真
snp	EIF5	M275I	89.97	HLA-B*44:03	5	34	1075.19965	真
snp	RNF213	R2827L	330.6	HLA-A*29:02	6	54	2855.46701	真
snp	FCHO2	L543S	43.6	HLA-A*26:01	8	91	5750.39585	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應 (任何時間)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
1-050-001	AEETPPFSNY	N		1-050-001_池_2	N	chr21_31266125_T_A
1-050-001	AAKAALEDF	未個別測試		1-050-001_池_2	N	chr3_47661451_C_G
1-050-001	EVTPILAIR	未個別測試		1-050-001_池_2	N	chr18_5419733_G_A
1-050-001	DVKAIGPLV	未個別測試		1-050-001_池_2	N	chr14_103339252_G_T
1-050-001	NETPVAVLTI	未個別測試		1-050-001_池_2	N	chr7_79453094_C_A
1-050-001	LFVVFQTVY	未個別測試		1-050-001_池_2	N	chr1_159535913_A_T
1-050-001	AEAERLDVKAI	未個別測試		1-050-001_池_2	N	chr14_103339252_G_T
1-050-001	ASGNHHVW	未個別測試		1-050-001_池_2	N	chr22_30893501_T_C
1-050-001	KLFHDMNVSY	未個別測試		1-050-001_池_2	N	chr1_193097666_T_C

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白質 效應	TPM	藉由全MS模 型覆蓋之最可 能限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可 能限制
snp	TIAM1	Y283F	13.99	HLA-B*44:03	9	16	364.187996	真
snp	SMARCC1	E721D	39.53	HLA-C*16:01	11	307	22125.437	假
snp	EPB41L3	S495L	51.69	HLA-A*26:01	12	125	9269.11767	真
snp	EIF5	M275I	89.97	HLA-A*26:01	13	90	5692.75283	真
snp	MAG12	G76V	2.29	HLA-B*44:03	14	13	253.431553	真
snp	OR10J5	L32Q	0.9	HLA-A*29:02	15	9	139.510048	真
snp	EIF5	M275I	89.97	HLA-B*44:03	17	38	1465.22509	真
snp	OSBP2	Y677H	7.86	HLA-C*16:01	18	173	13216.9384	假
snp	GLRX2	N94S	17.92	HLA-A*29:02	4	21	453.621334	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時	突變
1-050-001	ETPPFSNYNTL	未個別測試		1-050-001_池_2	N	chr21_31266125_T_A
CU04	DENITTIQF	Y		CU04_池_1	Y	chr4_22413213_C_A
CU04	MELKVESF	N		CU04_池_1	Y	chr1_37874128_G_C
CU04	EHIPESAGF	N		CU04_池_1	Y	chr3_9943508_G_C
CU04	YHGDPMPCL	N		CU04_池_1	Y	chr12_7066530_C_T
CU04	DEERIPVL	N		CU04_池_1	Y	chr7_5752914_T_C
CU04	EVADAATLTMY	Y		CU04_池_1	Y	chr1_52268541_A_C
CU04	IEVEVNEI	N		CU04_池_1	Y	chr7_135598004_C_G
CU04	DTVEYPYTSF	Y		CU04_池_1	Y	chr14_34713369_C_A

補充表3
用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白 質效 應	TPM	藉由全MS模型 覆蓋之最可能 限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可 能限制
snp	TIAM1	Y283F	13.99	HLA-A*26:01	7	172	13162.6216	真
snp	ADGRA3	C734F	20.67	HLA-B*18:01	0	2	8.27203164	真
snp	INPP5B	Q606E	36.85	HLA-B*18:01	1	5	13.0510076	真
snp	CRELD1	Q347H	29.9	HLA-B*38:01	2	103	4218.0095	真
snp	C1S	P295L	157.5	HLA-B*38:01	3	12	76.7416543	真
snp	RNF216	M45V	49.2	HLA-B*18:01	4	29	387.328968	真
snp	ZFYVE9	K845T	70.08	HLA-A*26:01	5	7	38.7340629	真
snp	NUP205	L691V	42.37	HLA-B*18:01	6	21	209.301169	真
snp	CFL2	D66Y	16.65	HLA-A*26:01	7	9	42.7267485	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
CU04	VEIEQLTY	N		CU04_池_1	Y	chr11_62827178_C_G
CU04	LELKAVHAY	N		CU04_池_1	Y	chr7_138762364_G_T
CU04	EEADFLLAY	N		CU04_池_2	N	chr6_10556704_C_T
CU04	ENITTIQFY	N		CU04_池_2	N	chr4_22413213_C_A
CU04	FHATNPLNL	N		CU04_池_2	N	chr14_75117203_C_G
CU04	VFKDLSVTL	N		CU04_池_2	N	chrX_40597563_G_A
CU04	QAVAAVQKL	N		CU04_池_2	N	chr17_42104792_T_A
CU04	IQDQIQNCI	N		CU04_池_2	N	chr2_67404159_G_C
CU04	VAKGFISRM	N		CU04_池_2	N	chr2_85395579_C_T

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白 質效 應	TPM	藉由全MS模 型覆蓋之最 可能限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可 能限制
snp	STX5	E134Q	83.43	HLA-B*18:01	8	3	11.6727539	真
snp	ATP6V0 A4	P163H	47.21	HLA-B*18:01	9	0	3.63590379	真
snp	GCNT2	P94L	25.19	HLA-B*18:01	10	1	6.48490966	真
snp	ADGRA3	C734F	20.67	HLA-A*26:01	11	16	135.44155	真
snp	NEK9	D252H	20.29	HLA-B*38:01	12	8	39.1165673	真
snp	ATP6AP 2	E145K	88.26	HLA-B*38:01	13	45	1080.8332	真
snp	DHX58	M513L	35.87	HLA-C*12:03	14	136	6872.44	真
snp	ETAA1	E493Q	38.47	HLA-B*38:01	15	59	1665.0162	真
snp	CAPG	E314K	151.7	HLA-C*12:03	16	107	5236.61406	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
CU04	QTKPASLLY	N		CU04_池_2	N	chr2_32487684_AG_A
CU04	DHFETIIKY	N		CU04_池_2	N	chr1_220024376_C_G
CU04	VEYPYTSF	N		CU04_池_2	N	chr14_34713369_C_A
CU05	SVSDISEYRV	N		CU05_池_1	N	chr12_15670870_G_C
CU05	YTFEIQGVNG					
CU05	V	N		CU05_池_1	N	chr1_22865138_C_G
CU05	IYTSSGQLQLF	N		CU05_池_1	N	chr10_73293336_T_C
CU05	FATPSLHTSV	N		CU05_池_1	N	chr17_80345147_A_T
CU05	AVSKPGLDYEL	N		CU05_池_1	N	chr14_77026556_T_A
CU05	KYINKTIRV	N		CU05_池_1	N	chr19_2328426_C_T

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白質 效應	TPM	藉由全MS模 型覆蓋之最 可能限制	全MS 模型排 名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可 能限制
del_fs	BIRC6	G2619fs	111.7	HLA-A*26:01	17	47	1143.73481	真
snp	EPRS	M277I	76.64	HLA-B*18:01	18	6	29.8996386	真
snp	CFL2	D66Y	16.65	HLA-B*18:01	19	4	12.3783994	真
snp	EPS8	Q64E	52.56	HLA-A*68:02	0	1	6.0399624	真
snp	EPHB2	A410G	74.99	HLA-A*68:02	1	22	132.877429	真
snp	CFAP70	E636G	30.45	HLA-A*24:02	2	17	46.3526841	真
snp	RNF213	D2271V	735.3	HLA-A*68:02	4	16	43.8761927	真
snp	IRF2BPL	M413L	58.51	HLA-A*68:02	5	274	13566.6012	真
snp	LSM7	D20N	76.01	HLA-A*24:02	8	32	318.671051	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
CU05	ETTEEMKYVL	N		CU05_池_1	N	chr6_80040624_G_A
CU05	VVSHPHLVYW	N		CU05_池_1	N	chr4_106232956_C_G
CU05	DIFQVVKAI	N		CU05_池_1	N	chr1_198754369_C_A
CU05	FAFDAVSKPGLN			CU05_池_1	N	chr14_77026556_T_A
CU05	SVSDISEYR	N		CU05_池_2	N	chr12_15670870_G_C
CU05	YTFEIQGV	N		CU05_池_2	N	chr1_22865138_C_G
CU05	ATPSLHTSV	N		CU05_池_2	N	chr17_80345147_A_T
CU05	DFATPSLHTSV	N		CU05_池_2	N	chr17_80345147_A_T
CU05	KYINKTIRVKF	N		CU05_池_2	N	chr19_2328426_C_T

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白質 效應	TPM	藉由全MS模 型覆蓋之最 可能限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可 能限制
snp	TTK	G804E	17.14	HLA-A*68:02	9	37	398.324158	真
snp	TBCK	D478H	71.17	HLA-A*68:02	11	235	10875.8686	真
snp	PTPRC	L1204I	104.6	HLA-A*68:02	13	36	394.198029	真
snp	IRF2BPL	M413L	58.51	HLA-A*68:02	18	65	1067.11951	真
snp	EPS8	Q64E	52.56	HLA-A*68:02	3	94	2050.45825	真
snp	EPHB2	A410G	74.99	HLA-A*68:02	6	11	26.6362167	真
snp	RNF213	D2271V	735.3	HLA-A*68:02	7	25	177.027506	真
snp	RNF213	D2271V	735.3	HLA-A*68:02	10	185	7619.02631	真
snp	LSM7	D20N	76.01	HLA-A*24:02	12	42	538.209517	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
CU05	SVKPHLCSL	N		CU05_池_2	N	chr17_35363437_C_T
CU05	DISEYRVEHL	N		CU05_池_2	N	chr12_15670870_G_C
CU05	WVVSHPHLV	N		CU05_池_2	N	chr4_106232956_C_G
CU05	KVFKLGNGKV	N		CU05_池_2	N	chrX_24810777_G_A
CU05	VSKPGLDYEL	N		CU05_池_2	N	chr14_77026556_T_A
CU02	SPSKTSLTL	未個別測試	見池結果	CU02_池_1	Y	chr12_132750694_G_T
CU02	ASADGTVKLW	未個別測試	見池結果	CU02_池_1	Y	chr16_1977246_A_G
CU02	LVGPAQLSHW	未個別測試	見池結果	CU02_池_1	Y	chr8_143930249_G_A
CU02	QTAAAVGVLK	未個別測試	見池結果	CU02_池_1	Y	chr7_77773271_A_G

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白質 效應	TPM	藉由全MS模 型覆蓋之最 可能限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry覆 蓋之最可能限 制
snp	SLFN11	R124H	91.5	HLA-A*68:02	14	88	1897.58723	真
snp	EPS8	Q64E	52.56	HLA-A*68:02	15	59	885.161001	真
snp	TBCK	D478H	71.17	HLA-A*68:02	16	15	40.725305	真
snp	POLA1	E1017K	19.31	HLA-A*68:02	17	61	954.869111	真
snp	IRF2BPL	M413L	58.51	HLA-A*68:02	19	258	12457.5646	真
snp	ANKLE2	P266T	43.78	HLA-B*07:02	0	7	20.5140939	真
snp	TBL3	I545V	26.23	HLA-B*57:01	1	20	77.5504026	真
snp	PLEC	P863L	528.5	HLA-B*57:01	4	42	287.473059	真
snp	RSBN1L	T584A	25.89	HLA-A*03:01	5	19	76.1012011	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應(任何時間點)	個別肽反應註釋	池ID	池反應(任何時間點)	突變
CU02	FPSPSKTSLTL	未個別測試	見池結果	CU02_池_1	Y	chr12_132750694_G_T
CU02	SSTSNRSSTW	未個別測試	見池結果	CU02_池_1	Y	chr10_96604023_G_A
CU02	LVYGPLGAGK	未個別測試	見池結果	CU02_池_1	Y	chr13_33821175_C_T
CU02	HSYSELCTW	未個別測試	見池結果	CU02_池_1	Y	chr8_119802006_C_G
CU02	VTLDVILER	未個別測試	見池結果	CU02_池_1	Y	chr9_108979413_T_G
CU02	HSKPEDTDAW	未個別測試	見池結果	CU02_池_1	Y	chr12_133057238_A_G
CU03	IAASRSVVM	未個別測試		CU03_池_1	N	chr1_230868472_G_A
CU03	AAIAASRSV	未個別測試		CU03_池_1	N	chr1_230868472_G_A
CU03	AASRSVVM	未個別測試		CU03_池_1	N	chr1_230868472_G_A

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白質 效應	TPM	藉由全MS模 型覆蓋之最 可能限制	全MS 模型 排名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry覆 蓋之最可能限 制
snp	ANKLE2	P266T	43.78	HLA-B*07:02	6	26	131.765585	真
snp	PIK3AP1	R733W	9.84	HLA-B*57:01	7	30	162.029882	真
snp	RFC3	S44L	9.76	HLA-A*03:01	8	2	8.21211585	真
snp	TAF2	D194H	29.74	HLA-B*57:01	9	3	10.120376	真
snp	CTNNAL 1	E323D	32.44	HLA-B*57:01	10	136	2107.24068	真
snp	ZNF84	T175A	29.84	HLA-B*57:01	11	23	90.7546185	真
snp	Clorf19 8	A14 V	36.47	HLA-C*12:03	0	19	146.699014	真
snp	Clorf19 8	A14 V	36.47	HLA-C*12:03	2	42	492.404622	真
snp	Clorf19 8	A14 V	36.47	HLA-C*12:03	6	116	3437.73836	真

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

患者	肽	個別肽反應 (任何時間 點)	個別 肽反 應註 釋	池ID	池反應 (任何時 間點)	突變
CU03	EMDMHLSDY	未個別測試		CU03_池_1	N	chr5_37180032_T_A
CU03	VENQKHSL	未個別測試		CU03_池_1	N	chr12_30728769_C_T
CU03	QYMDSSLVKI	未個別測試		CU03_池_1	N	chr10_60788061_G_T
CU03	SASLHPATV	未個別測試		CU03_池_1	N	chr2_25929006_C_T
CU03	VPDQKSKQL	未個別測試		CU03_池_1	N	chr6_63685063_T_G
CU03	IVFIATSEF	未個別測試		CU03_池_1	N	chr11_65976483_A_T
CU03	YPAPQPPVL	未個別測試		CU03_池_1	N	chr20_44066022_C_A

補充表3

用於NSCLC患者中之T-細胞識別所測試之肽

突變 類型	基因	蛋白 質效 應	TPM	藉由全MS模 型覆蓋之最可 能限制	全MS 模型排 名	MHCFlurry 排名	MHCFlurry (nM)	藉由 MHCFlurry 覆蓋之最可 能限制
snp	C5orf42	I1908L	14.78	HLA-A*01:01	8	7	35.7275148	真
snp	CAPRIN2	S554N	6.69	HLA-B*08:01	10	124	3970.47602	真
snp	CDK1	S107I	26.84	HLA-A*23:01	7	8	50.3301427	真
snp	KIF3C	R785H	17.29	HLA-C*12:03	9	30	260.370195	真
snp	PHF3	N447K	47.53	HLA-B*08:01	13	130	4071.14261	真
snp	SART1	N554I	70.53	HLA-B*15:03	5	3	17.4168253	真
snp	TOX2	S382Y	11.56	HLA-B*08:01	11	101	2455.95947	真

補充表4

		供體ID				
分析物(平均)	刺激因數	1-038-001	CU04	1-024-001	1-024-002	CU02
顆粒酶B (pg/ml)*	DMSO	1786.73	1383.53	2639.03	854.78	1449.74
	肽池1	1672.60	4269.64	2449.23	1281.54	1132.49
	DMSO	1874.02	3747.71	2382.01	626.20	n/a
	肽池2	3118.30	3191.90	2006.73	872.89	n/a
TNFα (pg/ml)#	DMSO	37.58	34.64	21.76	38.07	1.22
	肽池1	53.02	217.57	42.05	57.13	7.44
	DMSO	16.58	80.81	24.98	24.77	n/a
	肽池2	61.54	75.70	33.70	48.84	n/a
IL-2 (pg/ml)#	DMSO	1.78	3.86	4.24	0.23	6.67
	肽池1	15.53	9.88	7.75	0.00	0.00
	DMSO	26.66	27.25	5.72	10.20	n/a
	肽池2	0.00	19.15	11.48	0.00	n/a
IL-5 (pg/ml)#	DMSO	26.47	5.20	20.92	11.96	18.91
	肽池1	10.48	14.65	26.72	9.42	17.64
	DMSO	27.31	19.65	11.01	29.93	n/a
	肽池2	26.47	25.43	20.11	40.11	n/a
以斜體顯示陽性值。*顆粒酶B ELISA：相對於DMSO背景，值≥1.5倍被認為係陽性。# U-Plex MSD檢定：相對於DMSO背景，值≥1.5倍被認為係陽性						

補充表5

IVS對照實驗中之TSNA及傳染性疾病抗原決定基

肽名稱	序列	起源(細胞株，基因)	經預測之HLA限制	經預測之結合親和力	突變位置	突變核苷酸
新抗原_A1	APKKKSIKL	H2009 PPFIA3	B*07:02	125	chr19-49140014	C-至-T
新抗原_A2	LLLEVWHL	H128 FANCA	A*02:01	6	chr16-89808348	C-至-T
新抗原_A3	FTDEKVKAY	H2122 PDE10A	A*01:01	41	chr6-165543564	G-至-T
新抗原_A6	RTAKQNPLTK	H2122 GPR183	A*03:01	138	chr13-99295446	G-至-A
新抗原_A7	FLAPTGVVPV	H128 NTM	A*02:01	8	chr11-131911555	T-至-C
新抗原_A10	RLADAEKLFQL	H128 PLEKHG4	A*02:01	201	chr16-67284435	G-至-A
新抗原_A11	RTAKQNPLTKK	H2122 GPR183	A*03:01	131	chr13-99295446	G-至-A
新抗原_B2	IMYLTGMVNK	H2009 GSPT1	A*03:01	33	chr16-11891120	G-至-A
新抗原_B3	TLQELSHAL	H128 PRPF19	A*02:01	106	chr11-60902829	G-至-T
新抗原_B6	VSQPVAPSY	Colo829 KIAA0319L	A*01:01	948	chr1-35479047	C-至-T

補充表5

IVS對照實驗中之TSNA及傳染性疾病抗原決定基

肽名稱	序列	起源(細胞株，基因)	經預測之HLA限制	經預測之結合親和力	突變位置	突變核苷酸
新抗原_B7	RLFTPISAGY	H2126 CYP26B1	A*03:01	157	chr2-72133060	G-至-C
新抗原_B8	ITEEPILMTY	H2122 RP1L1	A*01:01	308	chr8-10611205	C-至-A
新抗原_B10	KVTGHRWLK	H2009 BSG	A*03:01	51	chr19-579577	G-至-A
新抗原_B12	KLSEQILKK	H2009 TLR5	A*03:01	39	chr1-223110532	C-至-G
新抗原_C3	GTKPNPHVY	H2126 OAS3	A*03:01	7336	chr12-112961105	G-至-T
新抗原_C4	QQQQVVTNK	H2126 LRP1	A*03:01	2361	chr12-57162861	G-至-T
新抗原_C5	KVLGKGSAFAK	H2126 PLK2	A*03:01	40	chr5-58459089	G-至-A
新抗原_C6	SVQAPVPPK	H2009 ENGASE	A*03:01	279	chr17-79084548	C-至-G
EBV RAKF	RAKFKQLL	EBV BZLF-1	B*08:01	457	Nan	Nan
流感(Flu) CTCL	CTELKLSDY	流感(Influenza) NP	A*01:01	39	Nan	Nan
流感ELRS	ELRSRYWAI	流感A	B*08:01	12	Nan	Nan
CMV NLVP	NLVPMVATV	CMV pp65	A*02:01	45	Nan	Nan
流感GILG	GILGFVFTL	流感MP	A*02:01	20	Nan	Nan

補充表5

IVS對照實驗中之TSNA及傳染性疾病抗原決定基

肽名稱	序列	起源(細胞株，基因)	經預測之HLA限制	經預測之結合親和力	突變位置	突變核苷酸
HCV KLVA	KLVALGINAV	HCV NS3	A*02:01	49	Nan	Nan
HIV ILKE	ILKEPVHGV	HIV pol	A*02:01	144	Nan	Nan
RSV NPKA	NPKASLLSL	RSV NP	B*07:02	60	Nan	Nan
*新抗原序列中之突變肽帶下劃線。				NaN	Nan	Nan
**腫瘤細胞株： Colo829、H128、 H2009、H2122、 H2126				NaN	Nan	Nan

補充表6

純系型	頻率	比例	TRAV	TRAJ	TRAC	TRBV	TRBD	TRBJ	TRBC
純系型1	386	0.49171975	TRAV8-4	TRAJ5	TRAC	TRBV2	TRBD2	TRBJ2-5	TRBC2
純系型3	53	0.06751592	TRAV6	TRAJ31	TRAC	TRBV6-1	TRBD2	TRBJ1-1	TRBC1
純系型9	7	0.0089172	TRAV22	TRAJ33	TRAC	TRBV20-1	TRBD1	TRBJ1-5	TRBC1
純系型10	5	0.00636943	TRAV17	TRAJ57	TRAC	TRBV7-6	TRBD1	TRBJ2-3	TRBC2
純系型14	4	0.00509554	TRAV13-1	TRAJ33	TRAC	TRBV28	TRBD2	TRBJ2-7	TRBC2

補充表6

α CDR3	β CDR3	全長 α VJ
CAVTVTGRRALTF	CASNPPDAARGQETQYF	MLLLLVPVLEVIFTLGGTRAQSVTQLGSHVSVSEGALVLLRCNYSSSVPPYLFWYV QYPNQGLQLLLKYTTGATLVKGINGFEAEFKKSETSFHLTKPSAHMSDAAEYFCAV TVTGRRALTFGSGTRLQVQ
CALNARLMF	CASSYREYNTEAFF	MAFWLRRRLGLHFRPHLGRRMESFLGGVLLILWLQVDWVKSQKIEQNSEALNIQE GKTATLTCNYTNYSPAYLQWYRQDPGRGPVFLLLIRENEKEKRKERLKVTFDITLK QSLFHITASQPADSATYLCALNARLMFGDGTQLVVK
CAVVLDSDNYQLIW	CSATRGHLSNQPQHF	MKRILGALLGLLSAQVCCVRGIQVEQSPPDULQEGANSTLRCNFSDSVNNLQWF HQNPWGQLINLFYIPSGTKQNGRLSATTVATERYSLLYISSSQTTDSGVYFCAVVLD SDNYQLIWGAGTKLIK
CATASRQGGSEKLVF	CASSRGGGTDTQYF	METLLGVSLVILWLQLARVNSQQGEEDPQALSIQEGENATMNCSYKTSINNQLW YRQNSGRGLVHLILIRSNEREKHSGRRLRVTLDTSKKSSSLITASRAADTASYFCATA SRQGGSEKLVFGKGTKLTVN
CAASSNYQLIW	CASSLGLAYEQYF	MTSIRAVFIFLWLQLDLVNGENVEQHPSTLSVQEGDSAVIKCTYSDSASNYFPWYK QELGKGPQLIIDIRSNVGEKKDQRIAVTLNKTAKHFSLHITETQPEDSAVYFCAASS NYQLIWGAGTKLIK

補充表6

全長β V(D)J
MDTWLVCWAI FSLLKAGLTEPEVTQT PSHQVTQM GQEVILRCVPISNHL YFYWYR QILGQKVEFLVS FYNNEISEKSEIFDDQFS VERPDGSNFTLKIRSTKLED SAMYFCAS NPPDAARGQETQYFGPGTRLLVL
MSIGLLCCVAFSLLWASPVNAGVTQT PKFQVLKTGQSMTLQCAQDMNHNSMY WYRQDPGMGLRLIYYSASEGTTDKG EVPNGYNVSRLNKREFSLRLESAAPSQTSVY FCASSYREYNTEAFFGQGTRLTVV
MLLLLLLLGPGSGLGAVVSQHPSRVICKSGTSVKIECRSLDFQATTMFWYRQFPKQ SLMLMATSNEGSKATYEQGVEKDKFLINHASLTLSTLTVTSAHPEDSSFYICSATRG HLSNQPQHFGDGTRLSIL
MGTSLLCWVVLGFLGTDHTGAGVSQSPRYKVTKRGQDVALRCDPISGHVS LYWY RQALGQGPEFLTYFN YEAAQQDKSGLPNDRFSAERPEGSISTLTIQRTEQRDSAMYR CASSRGGGTD TQYFGPGTRLTVL
MGIRLLCRVAF CFLAVGLVDVKVTQSSRYLVKRTGEKVFLECVQDMDHENMFWY RQDPGLGLRLIYFSYDVKMKEKGD IPEGYSVSREKKERFSLILESASTNQTSMYLCAS SLGLAYEQYFGPGTRLTVT

【符號說明】

【0587】

110	患者
114	候選新抗原序列
160	呈遞識別系統
165	呈遞資訊
118	疫苗
170	訓練資料存儲
170A	訓練資料
175	呈遞模型
312	資料管理模組
314	編碼模組
316	訓練模組
320	預測模組
324	患者選擇模組
2800	流程圖
2801	步驟
2802	步驟
2803	步驟
2804	步驟
2805	步驟
2806	步驟
2807	步驟

2808	步驟
2809	步驟
2810	步驟
2811	步驟
2812	步驟
2813	步驟
2814	步驟
2815	步驟
2816	步驟
2900	電腦
2902	處理器
2904	晶片組
2906	記憶體
2908	存儲裝置
2910	鍵盤
2912	圖形適配器
2914	指向裝置/輸入介面
2916	網路適配器
2918	顯示器
2920	記憶體控制器集線器
2922	輸入/輸出(I/O)控制器集線器



201923637

【發明摘要】**【中文發明名稱】**

利用熱點識別新抗原

【英文發明名稱】

NEOANTIGEN IDENTIFICATION USING HOTSPOTS

【中文】

本文提供一種識別可能在受試者之腫瘤細胞表面上呈遞之新抗原之方法。藉由對該受試者之腫瘤細胞定序獲得腫瘤新抗原之肽序列。將新抗原各者之肽序列與該受試者之核苷酸定序資料之複數個k-mer區塊中之一或多個k-mer區塊相關聯；將該等肽序列及相關聯k-mer區塊輸入至機器學習呈遞模型中以產生針對腫瘤新抗原之呈遞可能性，各呈遞可能性表示新抗原由該受試者之腫瘤細胞表面上之MHC對偶基因呈遞之可能性。基於該等呈遞可能性選擇新抗原之子集。

【英文】

A method for identifying neoantigens that are likely to be presented on a surface of tumor cells of a subject. Peptide sequences of tumor neoantigens are obtained by sequencing the tumor cells of the subject. The peptide sequence of each of the neoantigens is associated with one or more k-mer blocks of a plurality of k-mer blocks of the nucleotide sequencing data of the subject; The peptide sequences and the associated k-mer blocks are input into a machine-learned presentation model to generate presentation likelihoods for the tumor neoantigens, each presentation likelihood representing the likelihood that a neoantigen is

presented by an MHC allele on the surfaces of the tumor cells of the subject. A subset of the neoantigens is selected based on the presentation likelihoods.

【指定代表圖】

圖28

【代表圖之符號簡單說明】

- 2800 流程圖
- 2801 步驟
- 2802 步驟
- 2803 步驟
- 2804 步驟
- 2805 步驟
- 2806 步驟
- 2807 步驟
- 2808 步驟
- 2809 步驟
- 2810 步驟
- 2811 步驟
- 2812 步驟
- 2813 步驟
- 2814 步驟
- 2815 步驟
- 2816 步驟

【發明申請專利範圍】

【第1項】

一種識別來自受試者之一或多個腫瘤細胞之一或多個新抗原的方法，該等新抗原可能在該等腫瘤細胞之表面上呈遞，該方法包括以下步驟：

自該受試者之該等腫瘤細胞及正常細胞獲得外顯子組、轉錄組或全基因組核苷酸定序資料中之至少一者，其中使用該核苷酸定序資料以獲得代表藉由比較來自該等腫瘤細胞之該核苷酸定序資料與來自該等正常細胞之該核苷酸定序資料識別之新抗原集各者之肽序列的資料，其中各新抗原之該肽序列包含至少一個改變，該改變使其不同於自該受試者之該等正常細胞識別之對應野生型肽序列；

將該等新抗原各者之該肽序列編碼至對應數值向量中，各數值向量包含關於組成該肽序列之複數個胺基酸及該肽序列中之該等胺基酸之位置集的資訊；

使該等新抗原各者之該肽序列與該受試者之該核苷酸定序資料之複數個k-mer區塊中之一或多個k-mer區塊關聯；

使用電腦處理器，將該等數值向量及該一或多個相關聯k-mer區塊輸入至機器學習呈遞模型中以產生該新抗原集之呈遞可能性集，該集中之各呈遞可能性表示對應新抗原藉由該受試者之該等腫瘤細胞表面上之一或多個MHC對偶基因呈遞的可能性，該機器學習呈遞模型包括：

至少基於訓練資料集識別之複數個參數，該訓練資料集包括：

針對複數個樣品中之各樣品，藉由質譜法獲得的標記，該質

譜法量測結合至經識別為存在於該樣品中之MHC對偶基因集中之至少一個MHC對偶基因之肽之存在；

針對該等樣品各者，將訓練肽序列編碼為數值向量，該等數值向量包含關於組成該等肽之複數個胺基酸及該等肽中之該等胺基酸之位置集的資訊；及

針對該等樣品各者，針對該樣品之該等訓練肽序列各者，在該訓練肽序列與該等訓練肽序列之該核苷酸定序資料之複數個k-mer區塊中之一或多個k-mer區塊之間相關聯，

其中該等複數個參數之子集表示該一或多個k-mer區塊之呈遞熱點之存在或不存在；及

函數，該函數表示作為輸入接收之該數值向量及該一或多個k-mer區塊與基於該數值向量、該一或多個k-mer區塊及該等參數作為輸出產生之該呈遞可能性之間之關係；

基於該呈遞可能性集選擇該新抗原集之子集以產生選定之新抗原集；及

返回該選定之新抗原集。

【第2項】

如請求項1之方法，其中將該數值向量輸入至該機器學習呈遞模型中包括：

將該機器學習呈遞模型應用於該新抗原之該肽序列以產生該一或多個MHC對偶基因各者之相依性分數，該分數指示該MHC對偶基因是否將基於該肽序列之特定位置處之特定胺基酸呈遞該新抗原。

【第3項】

如請求項2之方法，其中將該數值向量輸入至該機器學習呈遞模型中還包括：

轉換該等相依性分數以產生各MHC對偶基因之對應按對偶基因可能性，該按對偶基因可能性指示該對應MHC對偶基因將呈遞該對應新抗原之可能性；及

將該等按對偶基因可能性組合以產生該新抗原之該呈遞可能性。

【第4項】

如請求項3之方法，其中該轉換該等相依性分數將該新抗原之該呈遞建模為跨一或多個MHC對偶基因互相排斥。

【第5項】

如請求項2之方法，其中將該數值向量輸入至該機器學習呈遞模型中還包括：

轉換該等相依性分數之組合以產生該呈遞可能性，其中轉換該等相依性分數之該組合將該新抗原之該呈遞建模為該一或多個MHC對偶基因之間之干涉。

【第6項】

如請求項2至5中任一項之方法，其中該呈遞可能性集進一步藉由至少一或多個對偶基因非相互作用特徵識別，且還包括：

將該機器學習呈遞模型應用至該等對偶基因非相互作用特徵以產生該等對偶基因非相互作用特徵之相依性分數，該分數指示該對應新抗原之該肽序列是否將基於該等對偶基因非相互作用特徵進行呈遞。

【第7項】

如請求項6之方法，其還包括：

將該一或多個MHC對偶基因中之各MHC對偶基因之該相依性分數與該等對偶基因非相互作用特徵之該相依性分數組合；

轉換各MHC對偶基因之該等組合之相依性分數以產生各MHC對偶基因之按對偶基因可能性，該按對偶基因可能性指示該對應MHC對偶基因將呈遞該對應新抗原之可能性；及

將該等按對偶基因可能性組合以產生該呈遞可能性。

【第8項】

如請求項6之方法，其還包括：

將該等MHC對偶基因各者之該等相依性分數與該等對偶基因非相互作用特徵之該相依性分數組合；及

轉換該等組合之相依性分數以產生該呈遞可能性。

【第9項】

如請求項6至8中任一項之方法，其中該至少一或多個對偶基因非相互作用特徵包括該新抗原之該肽序列與該新抗原之該核苷酸定序資料之複數個k-mer區塊中之一或多個k-mer區塊之間的關聯。

【第10項】

如請求項1至9中任一項之方法，其中該一或多個MHC對偶基因包括兩個或更多個不同MHC對偶基因。

【第11項】

如請求項1至10中任一項之方法，其中該等肽序列包括具有除了9個胺基酸外之長度之肽序列。

【第12項】

如請求項1至11中任一項之方法，其中編碼該肽序列包括使用獨熱編

碼方案編碼該肽序列。

【第13項】

如請求項1至12中任一項之方法，其中該等複數個樣品包括下列中之至少一者：

- (a)經基因改造以表現單個MHC對偶基因之一或多個細胞株；
- (b)經基因改造以表現複數個MHC對偶基因之一或多個細胞株；
- (c)獲自或源自複數例患者之一或多個人類細胞株；
- (d)獲自複數例患者之新鮮或冷凍腫瘤樣品；及
- (e)獲自複數例患者之新鮮或冷凍組織樣品。

【第14項】

如請求項1至13中任一項之方法，其中該訓練資料集還包括下列中之至少一者：

- (a)與該等肽中之至少一者之肽-MHC結合親和力測量值相關聯之資料；及
- (b)與該等肽中之至少一者之肽-MHC結合穩定性測量值相關聯之資料。

【第15項】

如請求項1至14中任一項之方法，其中該呈遞可能性集進一步藉由該受試者中之該一或多個MHC對偶基因之至少表現水平識別，該表現水平如藉由RNA-seq或質譜法所量測。

【第16項】

如請求項1至15中任一項之方法，其中該呈遞可能性集進一步藉由包括下列中之至少一者之特徵識別：

(a)該新抗原集中之新抗原與該一或多個MHC對偶基因之間之預測親和力；及

(b)該新抗原編碼之肽-MHC複合體之預測穩定性。

【第17項】

如請求項1至16中任一項之方法，其中該數值可能性集進一步藉由包括下列中之至少一者之特徵識別：

(a)其源蛋白序列內側接該新抗原編碼之肽序列之C-端序列；及

(b)其源蛋白序列內側接該新抗原編碼之肽序列之N-端序列。

【第18項】

如請求項1至17中任一項之方法，其中選擇該選定之新抗原集包括基於該機器學習呈遞模型選擇具有相對於未經選擇之新抗原增加之在該腫瘤細胞表面呈遞之可能性的新抗原。

【第19項】

如請求項1至18中任一項之方法，其中選擇該選定之新抗原集包括基於該機器學習呈遞模型選擇具有相對於未經選擇之新抗原增加之能誘導該受試者之腫瘤特異性免疫反應之可能性的新抗原。

【第20項】

如請求項1至19中任一項之方法，其中選擇該選定之新抗原集包括基於該呈遞模型選擇具有相對於未經選擇之新抗原增加之能藉由專業抗原呈遞細胞(APC)呈遞至初始T細胞之可能性的新抗原，視情況，其中該APC為樹突狀細胞(DC)。

【第21項】

如請求項1至20中任一項之方法，其中選擇該選定之新抗原集包括基於該機器學習呈遞模型選擇具有相對於未經選擇之新抗原降低之經受經由

中樞或外周耐受之抑制之可能性的新抗原。

【第22項】

如請求項1至21中任一項之方法，其中選擇該選定之新抗原集包括基於該機器學習呈遞模型選擇具有相對於未經選擇之新抗原降低之能誘導該受試者之正常組織之自體免疫反應之可能性的新抗原。

【第23項】

如請求項1至22中任一項之方法，其中該一或多種腫瘤細胞係選自由以下組成之群：肺癌、黑色素瘤、乳癌、卵巢癌、前列腺癌、腎癌、胃癌、結腸癌、睪丸癌、頭頸癌、胰癌、腦癌、B細胞淋巴瘤、急性骨髓性白血病、慢性骨髓性白血病、慢性淋巴球性白血病及T細胞淋巴球性白血病、非小細胞肺癌及小細胞肺癌。

【第24項】

如請求項1至23中任一項之方法，其還包括產生用於自該選定之新抗原集構建個人化癌症疫苗之輸出。

【第25項】

如請求項24之方法，其中用於該個人化癌症疫苗之該輸出包括編碼該選定之新抗原集之至少一個肽序列或至少一個核苷酸序列。

【第26項】

如請求項1至25中任一項之方法，其中該機器學習呈遞模型為神經網路模型。

【第27項】

如請求項26之方法，其中該神經網路模型包括針對該等MHC對偶基因之複數個網路模型，各網路模型經分配給該等MHC對偶基因之對應

MHC對偶基因且包含於一或多個層中排列之一系列節點。

【第28項】

如請求項27之方法，其中該神經網路模型藉由更新該神經網路模型之該等參數來訓練，且其中至少兩個網路模型之該等參數經聯合更新用於至少一個訓練迭代。

【第29項】

如請求項26至28中任一項之方法，其中該機器學習呈遞模型為包含一或多個節點層之深度學習模型。

【第30項】

如請求項1至29中任一項之方法，其中該一或多個MHC對偶基因為I類MHC對偶基因。

【第31項】

一種電腦系統，其包含：

電腦處理器；

記憶存儲電腦程式指令，當藉由該電腦處理器執行時使該電腦處理器進行以下：

自該受試者之該等腫瘤細胞及正常細胞獲得外顯子組、轉錄組或全基因組核苷酸定序資料中之至少一者，其中使用該核苷酸定序資料以獲得代表新抗原集中各者之肽序列的資料，該新抗原集藉由比較來自該等腫瘤細胞之該核苷酸定序資料與來自該等正常細胞之該核苷酸定序資料識別，其中各新抗原之該肽序列包含至少一個改變，該改變使其不同於自該受試者之正常細胞識別之對應野生型肽序列；

將該等新抗原各者之該肽序列編碼至對應數值向量中，各數值向量包含關於組成該肽序列之複數個胺基酸及該肽序列中之該等胺基酸之位置集的資訊；

使該等新抗原各者之該肽序列與該受試者之該核苷酸定序資料之複數個k-mer區塊中之一或多個k-mer區塊關聯；

將該等數值向量及該一或多個相關聯k-mer區塊輸入至機器學習呈遞模型中以產生該新抗原集之呈遞可能性集，該集中之各呈遞可能性表示對應新抗原藉由該受試者之該等腫瘤細胞表面上之一或多個MHC對偶基因呈遞的可能性，該機器學習呈遞模型包括：

至少基於訓練資料集識別之複數個參數，該訓練資料集包括：

針對複數個樣品中之各樣品，藉由質譜法獲得的標記，該質譜法量測結合至經識別為於該樣品中呈遞之MHC對偶基因集中之至少一個MHC對偶基因之肽之存在；

針對該等樣品各者，將訓練肽序列編碼為數值向量，該等數值向量包含關於組成該等肽之複數個胺基酸及該等肽中之該等胺基酸之位置集的資訊；及

針對該等樣品各者，針對該樣品之該等訓練肽序列各者，在該訓練肽序列與該等訓練肽序列之該核苷酸定序資料之複數個k-mer區塊中之一或多個k-mer區塊之間相關聯，

其中該等複數個參數之子集表示該一或多個k-mer區塊之呈遞熱點之存在或不存在；及

函數，該函數表示作為輸入接收之該數值向量及該一或多個k-mer區塊與基於該數值向量、該一或多個k-mer區塊及該等參數作

為輸出產生之該呈遞可能性之間之關係；

基於該呈遞可能性集選擇該新抗原集之子集以產生選定之新抗原集；及

返回該選定之新抗原集。

