



(10) 授权公告号 CN 112771214 B

(45) 授权公告日 2025.06.13

(21) 申请号 201980063254.9

(22) 申请日 2019.09.27

(65) 同一申请的已公布的文献号  
申请公布号 CN 112771214 A

(43) 申请公布日 2021.05.07

(30) 优先权数据  
18197172.2 2018.09.27 EP

(85) PCT国际申请进入国家阶段日  
2021.03.25

(86) PCT国际申请的申请数据  
PCT/EP2019/076210 2019.09.27

(87) PCT国际申请的公布数据  
W02020/065023 EN 2020.04.02

(73) 专利权人 耐考德治疗股份有限公司  
地址 挪威奥斯陆

(72) 发明人 莫妮卡·塞克利亚

卡罗琳·施杰田纳

阿涅特·布伦斯维克·弗雷德里克  
森

(74) 专利代理机构 中科专利商标代理有限责任  
公司 11021

专利代理师 高丽娜 张莹

(51) Int.Cl.

G40B 30/04 (2006.01)

A61K 39/00 (2006.01)

G16B 20/30 (2006.01)

(56) 对比文件

CN 108351916 A, 2018.07.31

CN 108513593 A, 2018.09.07

审查员 王斯奇

权利要求书3页 说明书49页  
序列表3页 附图19页

(54) 发明名称

用于选择新表位的方法

(57) 摘要

本发明涉及一种通过选择结合MHC I和/或MHC II的新表位并对它们的临床实用性进行排名来为个体选择新表位的方法。本发明还提供了通过本文所述方法获得的癌症疫苗。

1. 用于为患有或怀疑患有癌症的个体选择数目A个新表位的方法,所述方法包括以下步骤:

a. 从所述个体获得一个或多个新表位,每个新表位包含至少一个最小表位,其中每个新表位包含相对于参考序列的至少一个突变,其中所述最小表位由等于或小于新表位的氨基酸数目的氨基酸组成,并包含所述至少一个突变;其中获得新表位包括对肿瘤特异性的核酸序列中的突变进行鉴定的步骤;

b. 对于至少一个最小表位和相应的参考肽,确定MHC I和/或MHC II结合亲和力,其中结合亲和力是通过计算机预测通过计算%排名得分确定的,其中低的%排名得分表示强结合亲和力,和高的%排名得分表示弱的结合亲和力,和其中相应的参考肽是从同一个体或正常健康群体的体液或健康组织获得的;

c. 选择包含至少一个预测与MHC I和MHC II结合的最小表位的新表位,从而获得结合MHC的新表位;

d. 如下根据临床实用性的可能性对结合MHC的新表位进行排名,其中所述排名包括:

I. 确定步骤c中选择的结合MHC的新表位中包含的最小表位的数量,并且其中包含大量最小表位的新表位被排名为高于包含较少数量最小表位的新表位;

II. 确定步骤c中选择的结合MHC I的新表位的MHC I结合差异,其中MHC I结合差异由下式给出:(参考的%排名得分/(新表位的%排名得分)),其中具有高MHC I结合差异的新表位被排名为高于具有较低MHC I结合差异的新表位,并且其中在具有高MHC I结合差异的新表位中,将在锚定位置具有突变的新表位进行优先排序;

III. 根据MHC I排名,将步骤c中选择的结合MHC I的新表位进行优先排序,其中具有低%排名得分的新表位被排名为高于具有更高%排名得分的新表位,其中包含预测与MHC I结合的一个或多个最小表位的新表位的%排名得分等于所述一个或多个最小表位的最低%排名得分;和

IV. 确定BLOSUM得分,其中根据新表位的BLOSUM得分以降序将步骤c中选择的结合MHC I的新表位进行优先排序,其中新表位的BLOSUM得分等于它所包含的排名最佳的最小表位的BLOSUM得分,其中BLOSUM得分<2的最小表位被优先排序;和

e. 在步骤d的排名最高的结合MHC的新表位中选择A个新表位,从而选择可能具有临床实用性的A个新表位。

2. 根据权利要求1所述的方法,其中在步骤c中选择至少一个预测与MHC I结合的最小表位。

3. 根据权利要求1或2中任一项所述的方法,其中在步骤d.I.中,确定结合MHC I的新表位中包含的最小表位的数量。

4. 根据权利要求1或2中任一项所述的方法,其中A是1至100的整数。

5. 根据权利要求1或2中任一项所述的方法,其中A是5至50的整数。

6. 根据权利要求1或2中任一项所述的方法,其中A是5至30的整数。

7. 根据权利要求1或2中任一项所述的方法,其中步骤a包括鉴定相比于参考序列的在源自个体的一个或多个样品中的核酸序列中的突变。

8. 根据权利要求7所述的方法,其中所述一个或多个样品是源自个体的肿瘤细胞的样品。

9. 根据权利要求7所述的方法,其中所述参考序列是从需要治疗的个体的健康组织或从健康个体获得的。

10. 根据权利要求7所述的方法,其中,在源自所述个体的肿瘤细胞的至少两个样品中发现所述突变。

11. 根据权利要求1或2中任一项所述的方法,其中所述新表位包含1至50个最小表位。

12. 根据权利要求11所述的方法,其中所述新表位包含1至40个最小表位。

13. 根据权利要求11所述的方法,其中所述新表位包含1至30个最小表位。

14. 根据权利要求11所述的方法,其中所述新表位包含1至20个最小表位。

15. 根据权利要求1或2中任一项所述的方法,其中步骤c包括选择新表位,所述新表位包含至少一个预测结合MHC I的最小表位和至少一个预测结合MHC II的最小表位。

16. 根据权利要求1或2中任一项所述的方法,其中步骤 d. III包括将具有低于2.0的%排名得分的新表位进行优先排序。

17. 根据权利要求1或2中任一项所述的方法,其中在步骤d. III中将具有高于2的%排名得分的结合MHC I的新表位排除或降低其优先排序。

18. 根据权利要求1或2中任一项所述的方法,其中在步骤d. III中将具有高于10的%排名得分的结合MHC I的新表位排除或降低其优先排序。

19. 根据权利要求1或2中任一项所述的方法,其中在源自肿瘤细胞的多于一个样品中发现的新表位排名高于仅在源自肿瘤细胞的一个样品中发现的新表位。

20. 根据权利要求1或2中任一项所述的方法,其中所述突变是由至少两个不同的变体调用器鉴定的。

21. 根据权利要求1或2中任一项所述的方法,其进一步包括确定所述新表位的RNA表达水平的步骤,其中将未检测到RNA表达的新表位排除或降低其优先排序。

22. 根据权利要求1或2中任一项所述的方法,其进一步包括确定所述新表位的RNA表达水平的步骤,其中选择具有高于每百万0个转录本的可检测的归一化转录本水平表达的新表位。

23. 根据权利要求1或2中任一项所述的方法,其中步骤d进一步包括确定存在于结合MHC的新表位中的突变的等位基因频率的步骤,其中具有高等位基因频率的结合MHC的新表位排名高于具有较低等位基因频率的结合MHC的新表位,其中高等位基因频率是大于0.05的等位基因频率。

24. 根据权利要求1或2中任一项所述的方法,其中所述新表位的长度为7至40个氨基酸。

25. 根据权利要求1或2中任一项所述的方法,其中步骤d还包括将不包含任何预测结合MHC I的最小表位的新表位降低优先排序或排除的步骤。

26. 根据权利要求1或2中任一项所述的方法,其中包含至少5个结合MHC的最小表位的新表位被排名为高于包含少于5个结合MHC的最小表位的新表位。

27. 根据权利要求1所述的方法,其中步骤d, II. 和III. 的新表位通过向每个新表位分配它们包含的最小表位的最高得分来排序,其中对于新表位中包含的每个最小表位,确定突变是在最小新表位的锚定位置还是非锚定位置,和其中新表位的排序以如下1)至7)降序进行:

1) 如果突变处于锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分最高;

2) 如果突变处于非锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分低于1)的最小表位的得分;

3) 如果突变处于锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于2)的最小表位的得分;

4) 如果突变处于非锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于3)的最小表位的得分;

5) 如果突变处于锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于4)的最小表位的得分;

6) 如果突变处于非锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于5)的最小表位的得分;

7) 如果最小表位具有结合差异低于1的突变,则最小表位的得分最低,而与突变的位置无关。

28. 制备包含新表位的癌症疫苗的方法,所述方法包括使用根据权利要求1或2中任一项所述的方法选择数目A个新表位的步骤。

29. 根据权利要求28所述的方法,其中所述癌症疫苗包括:

- 靶向单元
- 二聚化单元;和
- 抗原单元,其中所述抗原单元包含A个新表位。

30. 根据权利要求29所述的方法,其中所述A是3至50的整数。

31. 根据权利要求29所述的方法,其中所述A是10至20的整数。

## 用于选择新表位的方法

### 技术领域

[0001] 本发明涉及一种通过选择结合MHC I和/或MHC II的新表位并对它们的临床实用性进行排名来为个体选择新表位的方法。本发明还提供了通过本文所述方法获得的癌症疫苗。

### 背景技术

[0002] 尽管在过去的几十年中,尤其是由于早期检测和诊断,癌症的治疗得到了改善,这大大提高了生存率,但只有大约60%的被诊断为癌症的患者在诊断后5年还活着。

[0003] 使用的大多数癌症治疗是手术治疗、放射线和细胞毒性化学疗法。但是它们都有严重的副作用。最近,也已使用针对已知癌症相关抗原或免疫调节分子的抗体进行治疗。

[0004] 在最近几年间,借助于患者自己的免疫系统靶向癌细胞的癌症免疫疗法(例如,癌症疫苗)引起了人们的兴趣,因为这种疗法可以减少甚至消除传统癌症治疗中所见的一些副作用。

[0005] 免疫学的基础是基于自身-非自身辨别。诱导感染性疾病的大多数病原体都含有可以被宿主识别并触发免疫应答的分子签名。然而,肿瘤细胞源自正常细胞,并且通常不表达任何外来分子签名,使得它们更难以与正常细胞区分开。

[0006] 然而,大多数肿瘤细胞表达不同类型的肿瘤抗原。一类肿瘤抗原是所谓的肿瘤相关抗原,例如在正常组织中以低水平表达而在肿瘤组织中以高得多的水平表达的抗原。在过去的十年中,这种肿瘤相关抗原一直是癌症疫苗的靶标。然而,针对肿瘤相关抗原的免疫治疗表现出若干挑战,因为肿瘤细胞可通过下调所讨论的抗原来逃避免疫系统,并且由于正常细胞破坏,该治疗也可能导致毒性。

[0007] 最近,已经鉴定了另一类肿瘤抗原,即所谓的肿瘤新抗原,其是肿瘤特异性抗原。肿瘤新抗原是由于肿瘤基因组中的一个或多个突变而出现的,所述突变导致所讨论蛋白质的氨基酸序列发生变化。由于这些突变在正常组织中不存在,因此针对肿瘤特异性新抗原的治疗副作用在用针对肿瘤新抗原的免疫治疗时不会出现。

[0008] 然而,为了产生有效的疫苗,重要的是选择最具免疫原性的新表位并将其用于疫苗。

### [0009] 发明概述

[0010] 本发明的发明人开发了一种新表位选择方法,以选择具有被证明对免疫原性重要的特性的新表位。通过使用这种方法,可以为疫苗选择预测结合主要组织相容性复合物(MHC)的高度免疫原性新表位,从而产生能够诱导强而有力的免疫应答的疫苗。因此可以产生适合个性化癌症治疗的疫苗。因此,本文提供了根据新表位的临床实用性对它们进行排名的方法,其特别适合于生产用于癌症治疗的个性化疫苗的情况。

[0011] 因此,本发明涉及用于为个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0012] a. 从所述个体获得一个或多个新表位,每个新表位均包含至少一个最小表位,其

中与参考序列相比,每个新表位均包含至少一个突变,诸如免疫原性突变;

[0013] b.对于每个所述新表位内的至少一个最小表位,诸如至少两个,三个或四个最小表位,确定MHC I和/或MHC II结合亲和力;

[0014] c.选择包含预测与MHC I和/或MHC II结合的至少一个最小表位的新表位,从而获得结合MHC的新表位;

[0015] d.根据临床实用性的可能性对结合MHC的新表位进行排名;

[0016] e.在排名最高的结合MHC的新表位中选择A个新表位,

[0017] 从而选择具有临床实用性的A个新表位。

[0018] 还提供了用于为个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0019] a.从所述个体获得一个或多个新表位,每个新表位均包含至少一个最小表位,其中与参考序列相比,每个新表位均包含至少一个突变,诸如免疫原性突变;

[0020] b.对于每个所述新表位,确定MHC I结合亲和力,并且对于每个所述新表位,确定结合MHC I的最小表位的数目;

[0021] c.对新表位进行如下排名:

[0022] i.将包含与MHC I和/或MHC II结合的大量最小表位的新表位进行优先排序,并选择具有高分数的第一组新表位;

[0023] ii.任选地,将来自第一组新表位的新表位如下进行优先排序:

[0024] 1)如果突变处于锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分最高;

[0025] 2)如果突变处于非锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分低于1)的最小表位的得分;

[0026] 3)如果突变处于锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于2)的最小表位的得分;

[0027] 4)如果突变处于非锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于3)的最小表位的得分;

[0028] 5)如果突变处于锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于4)的最小表位的得分;

[0029] 6)如果突变处于非锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于5)的最小表位的得分;

[0030] 7)如果最小表位具有结合差异低于1的突变,则最小表位的得分最低,而与突变的位置无关;

[0031] 其中,MHC I结合差异由下式给出:(用于参考的%排名得分(MHC I))/(用于最小表位的%排名得分(MHC I));并选择第二组得分高的新表位;

[0032] iii.任选地,基于MHC I%排名得分,将来自第二组的新表位进行优先排序,并选择具有低MHC I%排名得分的第三组新表位;

[0033] iv.任选地,从第二组或第三组中选择第四组新表位,其包括与已知被T细胞识别的表位高度相似的最小表位,

[0034] v.任选地,基于BLOSUM得分,将来自第二组、来自第三组或来自第四组的新表位进行优先排序,其中小于预定阈值的BLOSUM得分排名高于等于或大于所述阈值的BLOSUM得

分,并且选择BLOSUM得分小于所述阈值的第五组新表位,其中所述阈值优选为1;

[0035] vi. 任选地,基于在两个或更多个样品中发现的新表位,从第一组、第二组、第三组、第四组或第五组新表位中选择新表位,并选择在两个或更多个样品中发现的第六组新表位;

[0036] vii. 任选地,基于至少两个不同的变体调用器(variant caller)对突变的鉴定,从第一组、第二组、第三组、第四组、第五组或第六组新表位中选择新表位,并选择包含由至少两个不同的变体调用器所鉴定的突变的第七组新表位,

[0037] 其中第一组、第二组、第三组、第四组、第五组、第六组或第七组新表位包含所述A个表位。

[0038] 还提供了制备包含新表位的癌症疫苗的方法,所述方法包括使用本文所述的方法选择所述新表位的步骤。

[0039] 本文还描述了可通过本文描述的方法获得的癌症疫苗。

[0040] 还提供了用于为个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0041] a. 从所述个体获得一个或多个新表位,每个新表位均包含至少一个最小表位,其中与参考序列相比,每个新表位均包含至少一个突变,诸如免疫原性突变;

[0042] b. 对于每个所述新表位,确定MHC I和/或MHC II结合亲和力;

[0043] c. 选择包含预测与MHC I和/或MHC II结合的至少一个最小表位(例如至少两个,三个或四个最小表位)的新表位,从而获得结合MHC的新表位;

[0044] d. 根据临床实用性的可能性对结合MHC的新表位进行排名;

[0045] e. 在排名最高的结合MHC的新表位中选择A个新表位,其中A是整数并且A是至少3个,诸如至少4个,诸如至少5个,

[0046] 从而选择以免疫活性量施用于个体时能够诱导CD8+T细胞应答的A个新表位。

[0047] 附图简述

[0048] 图1. 免疫原性(左)与非免疫原性(右)新表位的MHC I类(A)和II类(B)分子的NetMHCpan/NetMHC IIpan预测结合亲和力%排名(在VB10.NEO疫苗接种的小鼠中进行评估)。

[0049] 图2. 对于MHC I类(A)和II类(B)分子具有低%排名的选定新表位的BLOSUM得分。免疫原性肽:左;非免疫原性肽:右。

[0050] 图3. 在CT26(A)和B16(B)模型数据集上使用NeoSELECT选择的前20个新表位的IFN- $\gamma$ 斑点的累计数目。用黑线描绘了每个数据集的1000个随机选择的新表位的IFN- $\gamma$ 斑点的平均累计数目。

[0051] 图4. 使用NeoSELECT策略从CT26(A), B16(B)和LL2(C)数据集中获得的前20个新表位(突变)和这些新表位的野生型序列(WT)的IFN- $\gamma$ 斑点的累计数目。用黑线描绘了来自CT26(A)、B16(B)和LL2(C)的1000个随机选择的新表位的IFN- $\gamma$ 斑点的平均累计数目。

[0052] 图5. 显示了MHC I结合亲和力降低(MHC I结合差异低于1), MHC I结合亲和力增加(MHC I结合差异高于3)和MHC I结合亲和力没有变化的免疫原性和非免疫原性新表位的BLOSUM得分。该图显示,当改变MHC I结合亲和力时(左图和中图),免疫原性新表位(浅灰色)的BLOSUM得分低于非免疫原性新表位(深灰色),而对于MHC I结合亲和力不变的新表位,未观察到BLOSUM得分的变化。

[0053] 图6.Kreiter等人2015年和Castle等人2012年研究了10种不同的新表位 (pep1-10),它们均预测会结合MHC I类 (CD8+T细胞应答),以研究它们是否可以在小鼠B16-F10黑色素瘤肿瘤模型中诱导CD8+T细胞应答。当以本文所述的疫苗体 (vaccibody) (“VB10.NEO”)施用,由6个新表位诱导的应答显示在上图中。当以肽加聚ICLC佐剂、RNA和疫苗体的形式施用,CD4和CD8应答总结在下图中,其中白色表示无应答,浅灰色表示弱应答,中灰色表示中应答,深灰色表示强应答。

[0054] 图7.新表位选择过程的概述。

[0055] 图8.在27个氨基酸长的新表位中,对于MHC I类和II类的预测的最小表位的总数 (在接种了VB10.NEO的小鼠中评估)。Im:免疫原性的;Non-Im:非免疫原性的。

[0056] 图9.对于在锚定位置具有突变的新表位,就MHC I类分子在WT和MT之间的结合差异而言的免疫原性和非免疫原性新表位之间的比较。

[0057] 图10.对于MHC I类 (A) 和II类 (B) 分子具有低%排名的选定新表位的BLOSUM得分。

[0058] 详细说明

[0059] 定义

[0060] 如本文所用,术语“肿瘤新抗原”或“新抗原”是指与宿主的健康组织外显子组相比,包含一个或多个突变的任何肿瘤特异性抗原。肿瘤新抗原与术语癌症新抗原同义使用。所述一个或多个突变也可以称为“新表位突变”。所述突变可以是导致至少一个氨基酸改变的任何突变。因此,该突变可能是以下之一:

[0061] -导致氨基酸改变的非同义突变

[0062] -导致移码的突变,从而在突变后的方向上形成完全不同的开放阅读框

[0063] -通读突变,其中终止密码子被修饰或缺失,导致具有肿瘤特异性新表位的更长蛋白

[0064] -导致独特的肿瘤特异性蛋白质序列的剪接突变

[0065] -染色体重排,其产生在两种蛋白的连接处带有肿瘤特异性新表位的嵌合蛋白。

[0066] 本文所理解的突变不一定指单个残基的突变,而是更通常是指给定序列 (例如,潜在的新表位) 与参考序列之间的差异。因此,突变可以指一个以上氨基酸残基的突变。在一些实施方式中,突变是免疫原性突变。

[0067] 如本文所用,术语“肿瘤新表位”或“新表位”是指肿瘤抗原中的任何免疫原性突变,并且与术语癌症新表位同义使用。通过将源自肿瘤样品的新表位序列与参考样品中存在的参考序列 (例如来自相同个体的健康组织) 进行比较来确定突变的存在。它通常是指长度为27个氨基酸的肽。如本文所定义,新表位可以包含一个或若干个最小表位。突变通常存在于新表位的中心处或中心附近,即27聚体中的位置14,但不一定存在于最小表位的中心处。

[0068] 如本文所用,术语“肿瘤新表位序列”或“新表位序列”是指在抗原亚基中包含新表位的序列,并且与术语癌症新表位序列同义使用。

[0069] 如本文所用,术语“肿瘤新表位肽”,“新表位肽”是指新表位的肽序列,其中所述肽序列包含突变。

[0070] 术语“最小表位”是指预测与MHC I或MHC II结合的新表位的亚序列,所述亚序列包含可以是免疫原性的突变。换句话说,最小表位可以是免疫原性的,即能够引发免疫应

答,例如,如果其包含向最小表位赋予免疫原性或增加最小表位免疫原性的突变。这种突变在本文中称为免疫原性突变。因此,术语最小表位可以指新表位的短亚序列,所述短亚序列预测与MHC I或MHC II结合,并包含在新表位中发现的突变。因此,在位置14处包含突变的27聚体新表位可以涵盖若干个最小结合表位,即最小结合表位,其长度短于27个氨基酸,但是每个都包含突变。例如,最小表位可以由新表位的前14个氨基酸组成,条件是它预测与MHC I或MHC II结合,或者可以由新表位的氨基酸9至18或氨基酸7至22组成。

[0071] 本文所用的术语“MHC分子”包括I类MHC(MHC I)和II类MHC(MHC II)分子。MHC I代表若干个基因座,例如HLA-A(人白细胞抗原-A),HLA-B,HLA-C,HLA-E,HLA-F,HLA-G,HLA-H,HLA-J,HLA-K,HLA-L,HLA-P和HLA-V,而MHC II代表诸如HLA-DRA,HLA-DRB 1-9,HLA-,HLA-DQA1,HLA-DQB1,HLA-DPA1,HLA-DPB1,HLA-DMA,HLA-DMB,HLA-DOA和HLA-DOB的基因座。术语“MHC分子”和“HLA分子”在本文可互换使用。

[0072] 如本文所用,术语“选择新表位”或“新表位的选择”是指用于潜在临床或治疗用途,优选用于癌症疫苗的新表位的选择。因此,当新表位被选中时,它是用于临床用途或用于癌症疫苗的潜在候选物。所选择的新表位比未被选择的新表位排名更高或优先排序。

[0073] 核苷酸在本文中定义为RNA或DNA的单体。核苷酸是附接至碱基和磷酸基团的核糖或脱氧核糖环。单磷酸、二磷酸和三磷酸核苷都称为核苷酸。

[0074] 如本文所用,术语“基因组”是指生物体或细胞的染色体中的遗传信息的总量。

[0075] 如本文所用,术语“外显子组”是指由外显子形成的基因组的一部分,被转录时该序列在通过RNA剪接除去内含子后仍保留在成熟RNA内。它由在转录组以外的任何类型的细胞中转录成成熟RNA的所有DNA组成,所述转录组是仅在特定细胞群体中转录的RNA。

[0076] 如本文所用,术语“突变”包括优选存在于编码新表位的核苷酸中的易位、倒位、缺失、重复和点突变。在一个优选的实施方式中,突变是优选存在于新表位中的氨基酸置换。

[0077] 如本文所用,术语“癌症疫苗”是指治疗现有癌症或预防癌症发生的疫苗。治疗现有癌症的疫苗被称为治疗性癌症疫苗。

[0078] 鉴定新表位

[0079] 根据本发明的方法可以包括通过从如本文所述的获自个体的样品中鉴定核酸序列中的肿瘤特异性突变来鉴定一个或多个新表位的步骤。所述一个或多个新表位优选是多个新表位;但是,在某些情况下,即使是单个新表位也可能在个性化治疗(例如,个性化的癌症治疗)的背景下有用,只要它能够诱导所需的免疫应答即可。

[0080] 优选地,所述个体是癌症患者。可以从癌症患者获得一个或多个样品,以鉴定可能是临床用途(诸如例如个性化免疫原性癌症疫苗)的潜在候选物的新表位。

[0081] 优选地,通过将获自所述个体的肿瘤样品的核苷酸序列与正常核苷酸序列进行比较来鉴定肿瘤特异性突变。正常核苷酸序列可以通过对获自所述个体的体液样品或任何非肿瘤组织的核酸进行测序而获得。优选地,正常核苷酸序列获自相同个体的健康组织。正常核苷酸序列也可以从数据库获得。术语“正常”当应用于序列时,无论是肽序列还是核苷酸序列,在本文中将与术语“参考”或“野生型”互换使用。因此,该术语可适用于在相同个体内或正常健康群体中在“正常”情况下发现的相应序列。因此,将肿瘤特异性序列与在从相同个体分离的另一个非肿瘤组织中发现的相应序列进行比较,可以减少假阳性的数量,因为由个体之间的遗传变异产生的新表位(例如包含可能是个体特异性但非肿瘤特异性的SNP)

将从结果中滤除。从个体分离的参考序列可以来自在诊断之前从相同个体获得的健康细胞样品。在治疗开始之前或治疗开始之后也可能已经获得了样品。

[0082] 如本文所用,术语“肿瘤样品”是指包含肿瘤细胞的样品。可以通过从所述个体或癌症患者进行活检获得肿瘤样品。活检可以是用针或小手术进行的肿瘤组织的小样品。样品也可以是淋巴结活检。同样,可以进行若干次活检或肿瘤活检。

[0083] 参考样品可以是体液样品。体液样品可以例如是尿液样品,粪便样品,血清样品或唾液样品。在优选的实施方式中,体液样品是血液样品。优选地,参考样品获自需要治疗的个体的健康组织。

[0084] 可以使用任何已知的测序方法对获自样品的核酸进行测序。例如,可以使用新一代测序。在一些实施方式中,将来自个体的肿瘤细胞的核酸序列与来自正常细胞(例如来自相同个体的健康细胞或参考细胞)的核酸序列进行比较,以鉴定序列差异。

[0085] MHC I和MHC II的结合亲和力

[0086] 为了选择和优先排序作为个性化癌症疫苗的最强候选者的新表位,本发明人已经开发了一种对引发强烈免疫应答的新表位进行排名和选择的方法。

[0087] 要启动免疫应答,新表位应靶向主要组织相容性复合体(MHC)限制的表位,因为只有能够结合MHC分子的肽才可以提供合格的T细胞靶标。因此,选择最有希望的新表位的第一步可能是确定它们的MHC结合亲和力。

[0088] 在体内所有有核细胞的细胞表面上都发现了MHC I。MHC I的一项功能是将细胞内的肽展示到细胞毒性T细胞。将MHC I复合物-肽复合物插入向细胞毒性T细胞呈递肽的细胞的质膜中,从而触发针对特定MHC-肽复合物的细胞毒性T细胞的激活。该肽位于MHC I分子的沟内,使肽长度通常为约8-10个氨基酸。MHC I可以与MHC I类互换使用。

[0089] MHC II分子是通常仅在抗原呈递细胞(例如树突细胞,单核吞噬细胞,一些内皮细胞,胸腺上皮细胞和B细胞)上发现的分子家族。MHC II可与MHC II类互换使用。

[0090] 与MHC I相反,由MHC II类肽呈递的抗原来源于胞外蛋白。将胞外蛋白内吞,在溶酶体中消化,然后将得到的抗原性肽加载到MHC II类分子上,然后呈递在细胞表面处。MHC II类分子的抗原结合沟在两端均打开,并且能够呈递更长的肽,通常长度为15至24个氨基酸残基。另外,通常由MHC II在树突细胞表面上呈递的外源抗原可经由交叉呈递通过MHC I途径呈递。交叉呈递对于针对大多数肿瘤和病毒的免疫力是必需的。

[0091] MHC I类分子由CD8+T细胞上的T细胞受体(TCR)和共受体识别,而MHC II类分子由CD4+T细胞上的TCR和共受体识别。

[0092] 本发明的发明人已经建立了一种选择新表位的方法,其中根据临床实用性对新表位进行排名,并且其中选择对MHC(例如,MHC I和/或MHC II)具有亲和力的新表位。

[0093] 因此,本发明的一个方面涉及用于为个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0094] a. 从所述个体获得一个或多个新表位,每个新表位均包含至少一个最小表位,其中与参考序列相比,每个新表位均包含至少一个突变,诸如免疫原性突变;

[0095] b. 对于每个所述新表位内的至少一个最小表位,诸如至少两个,三个或四个最小表位,确定MHC I和/或MHC II结合亲和力;

[0096] c. 选择包含预测与MHC I和/或MHC II结合的至少一个最小表位的新表位,从而获

得结合MHC的新表位；

[0097] d. 根据临床实用性的可能性对结合MHC的新表位进行排名；

[0098] e. 在排名最高的结合MHC的新表位中选择A个新表位，

[0099] 从而选择可能具有临床实用性的A个新表位。

[0100] 具有高临床实用性可能性的新表位是具有免疫原性并且很可能适用于癌症疫苗的新表位。因此，本发明的方法是选择高度适合用于癌症疫苗的最具免疫原性的新表位的方法。

[0101] 人群中的MHC等位基因显示出极端的多态性。每个遗传基因座包含大量的单倍型，所述单倍型包含编码不同肽的不同等位基因。因此，优选在新表位序列与从相同个体获得的MHC I和/或MHC II序列之间确定MHC I和MHC II结合亲和力。

[0102] 为了确定在个体中鉴定的新表位的MHC结合亲和力，有必要对所述个体进行HLA分型。因此，本发明的方法还可包括其中通过确定所述个体的HLA基因型来确定MHC结合亲和力的步骤。例如，所述HLA基因型是从所述个体的血液样品中确定的。

[0103] 确定个体的HLA类型的技术是众所周知的。可以通过使用任何合适的测序方法来进行HLA分型。优选地，使用新一代测序。

[0104] 例如，从所述个体的血液样品中分离的DNA可以在新一代测序平台诸如例如Illumina平台上进行测序。

[0105] 可以使用预测肽与MHC I和MHC II分子之间的结合亲和力的计算机程序来确定或预测MHC I和MHC II结合亲和力。

[0106] 因此，在一个实施方式中，MHC结合亲和力是通过计算机预测来确定的。优选地，所述计算机预测是通过使用计算机程序进行的，该计算机程序预测肽与I类MHC和/或II类MHC分子的结合。通过将预测的结合亲和力值与100,000个随机天然9聚体肽集合的预测的结合亲和力进行比较，可以将所述预测的结合亲和力值转换为百分比得分。优选地，用于预测新表位与MHC I和MHC II分子的结合亲和力的%排名得分是使用以下文献中描述的NetMHCpan进行计算的：Nielsen等人2016(Nielsen, M.等人(2016) “NetMHCpan-3.0: improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length data sets.” Genome Medicine:8:33)和Vannessa, J.等人(2017) “NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data.” Vanessa Jurtz, Sinu Paul, Massimo Andreatta, Paolo Marcatili, Bjoern Peters和Morten Nielsen. The Journal of Immunology(2017))和如Andreatta等人2015(Andreatta M等人Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. Immunogenetics .2015;67(11-12):641-650)中所描述的NetMHC IIpan计算机程序。使用NetMHCpan的30.03.2016和NetMHC IIpan的29.09.2015的以上数据库来计算本文提到的%排名得分。

[0107] 低的%排名得分表示强结合亲和力，而较高的%排名得分表示较弱的结合亲和力。因此，具有低%排名得分的新表位优选地排在具有较高%排名得分的新表位之上。

[0108] %排名得分(MHC I)是指预测肽、最小表位或新表位的MHC I结合亲和力的%排名得分。

[0109] %排名得分(MHC II)是指预测肽、最小表位或新表位的MHC II结合亲和力的%排名得分。

[0110] 如上所述,每个新表位可以包含一个或多个最小表位。还可以确定给定新表位内每个最小表位的%排名得分(MHC I)或%排名得分(MHC II)。优选地,新表位的%排名得分等于其所包含的最小表位的最低%排名得分,即,新表位的结合亲和力被认为与它所包含的最小表位中的最佳结合剂的结合亲和力相同。

[0111] 在一些实施方式中,具有小于或等于2的%排名得分(MHC I)的新表位或最小表位被认为与MHC I结合。通常,较低的%排名得分(MHC I)指示较高的结合亲和力。例如,在一些实施方式中,%排名得分(MHC I)在0.5和2之间( $0.5 < \% \text{排名得分(MHC I)} \leq 2$ )的新表位或最小表位可以被认为是弱的MHC I结合剂,而%排名得分(MHC I)等于或低于0.5表示新表位或最小表位是强MHC I结合剂。在一些实施方式中,%排名得分 $>2$ 的新表位或最小表位被认为不能与MHC I结合。在优选的实施方式中,不包含任何预测与MHC I结合的最小表位的新表位被认为不具有临床实用性并被排除或降低优先排序。换句话说,本方法的步骤e可以包括以下步骤或由以下步骤组成:排除预测不与MHC I结合的新表位或降低预测不与MHC I结合的新表位的优先排序。降低新表位的优先排序意味着该新表位被认为与临床用途的相关性远。排除新表位意味着未选择新表位用于临床用途和/或未选择新表位用于疫苗。

[0112] 通常,包括具有高MHC I结合亲和力(即低%排名得分(MHC I))的最小表位的新表位排名高于包括具有较低MHC I结合亲和力(即较高%排名得分(MHC I))的最小表位的新表位。包括具有高MHC I结合亲和力(即低%排名得分(MHC I))的最小表位的新表位排名高于包括具有较低MHC I结合亲和力(即较高%排名得分(MHC I))的最小表位的新表位。

[0113] 在一些实施方式中,%排名得分(MHC II)小于或等于10的新表位或最小表位被认为与MHC II结合。例如,在一些实施方式中,%排名得分(MHC II)在2和10之间( $2 < \% \text{排名得分(MHC II)} \leq 10$ )的新表位或最小表位可以被认为是弱的MHC II结合剂,而%排名得分(MHC II)等于或低于2表示新表位或最小表位是强MHC II结合剂。在一些实施方式中,%排名得分 $>10$ 的新表位或最小表位被认为不能结合MHC II。在优选的实施方式中,不包含任何结合性最小表位的新表位被认为不具有临床实用性并且被排除或降低优先排序。换句话说,本发明方法的步骤,例如本发明方法的步骤c或e,可以包括以下步骤或由以下步骤组成:排除仅包含预测不与MHC II结合的最小表位的新表位。降低新表位的优先排序意味着该新表位被认为与临床用途的相关性远。排除新表位意味着未选择新表位用于临床用途和/或未选择新表位用于疫苗。

[0114] 通常,具有高MHC II结合亲和力(即低%排名得分(MHC II))的新表位或最小表位的排名高于具有较低MHC I结合亲和力(即较高%排名得分高(MHC II))的新表位或最小表位。包括具有高MHC II结合亲和力(即低%排名得分(MHC II))的最小表位的新表位的排名高于包括具有较低MHC II结合亲和力(即较高%排名得分(MHC II))的最小表位的新表位。

[0115] 在一个优选的实施方式中,结合MHC II的最小表位具有低于10的%排名(MHC II)得分。因此,包含在根据本发明方法的步骤b中所选择的新表位中的结合MHC II的最小表位优选地具有低于10的%排名得分(MHC II)。这意味着选择包含最小表位的新表位用于潜在的临床应用并构成潜在的疫苗候选者,其中所述最小表位被预测或确定为以低于10的%排名得分结合MHC II。优选降低下述新表位的优先排序或排除下述新表位:包含预测以大于

10的%排名得分(MHC II)结合MHC II的最小表位的新表位。

[0116] 通常,包含具有高MHC II结合亲和力的最小表位的结合MHC II的新表位排名高于包含具有较低MHC II结合亲和力的最小表位的新表位。

[0117] 最小表位的数目

[0118] 与仅包含一个最小表位的新表位相比,可以预期包含一个以上最小表位的新表位具有更高的临床实用性。因此,在一些实施方式中,在上述方法的步骤d中,包含免疫原性突变并能够结合MHC分子的最小表位的数目是最重要的参数。然而,以下情况也是有可能的:可能具有临床实用性并且被本发明方法选择的新表位仅包含单个最小表位;如果没有其他可用的新表位,和/或如果该最小表位赋予与MHC分子的结合或强结合,则这种新表位可能是临床上相关的。

[0119] 在一个实施方式中,包含较高数目的最小表位的新表位排名高于包含较低数目的最小表位的新表位。包含较低数目的最小表位的新表位包括仅包含一个最小表位的新表位。

[0120] 因此,在一个实施方式中,本发明的方法进一步包括确定包含在结合MHC I的新表位中的最小表位的数目的步骤,其中使包含较高数目的结合性最小表位的结合MHC I的新表位排名高于包含较低数目的结合性最小表位的结合MHC I的新表位。

[0121] 在另一个实施方式中,本发明的方法包括或进一步包括确定存在于结合MHC II的新表位中的最小表位的数目的步骤,其中包含较高数目的结合性最小表位的结合MHC II的新表位排名高于包含较低数目的结合性最小表位的结合MHC II的新表位。

[0122] 在一些实施方式中,本方法包括确定能够结合MHC I分子的最小表位的数目和能够结合MHC II分子的最小表位的数目的步骤。

[0123] 为了根据新表位所包含的最小表位的总数来的函数对新表位进行排序,在一些实施方式中,与结合MHC II的最小表位的数目相比,与MHC I结合的最小表位的数目权重更高,例如两倍。这意味着结合MHC I的最小表位的数目的得分高于结合MHC II的最小表位的数目的得分,例如两倍高。例如,预测具有14个结合MHC I的最小表位和13个结合MHC II的最小表位的新表位比包含12个结合MHC I的最小表位和15个MHC II最小新表位的新表位具有更高的免疫原性。

[0124] 由此确定每个新表位的结合得分。在实践中,确定新表位中包含的最小(结合性)表位的数目,从而鉴定x个结合MHC I的最小表位和/或y个结合MHC II的最小表位。

[0125] 确定新表位的结合得分的合适评分方案可以是:

[0126]  $a*2x+b*y$

[0127] 其中较高的结合得分表示如上所述的临床实用性较高的可能性。

[0128] 例如,如果 $y>0$ ,则 $a=1$ ;如果 $y=0$ ,则 $a=0$ ;如果 $x>0$ ,则 $b=1$ ;如果 $x=0$ ,则 $b=0$ 。

[0129] 另一种合适的评分方案可以是:

[0130]  $a*x+b*y$

[0131] 其中较高的结合得分表示如上所述的临床实用性较高的可能性。

[0132] 例如, $a=2$ 和 $b=1$ ,或者如果 $y>0$ 则 $a=2$ ,如果 $y=0$ 则 $a=0$ ,如果 $x>0$ 则 $b=1$ ,如果 $x=0$ 则 $b=0$ 。

[0133] 在一些实施方式中,包含至少一个结合MHC I的最小表位和至少一个结合MHC II



包含少于5个结合MHC II的最小表位的新表位。在一些实施方式中,包含至少6个结合MHC II的最小表位的新表位排名高于包含少于6个结合MHC II的最小表位的新表位。在一些实施方式中,包含至少7个结合MHC II的最小表位的新表位排名高于包含少于7个结合MHC II的最小表位的新表位。在一些实施方式中,包含至少8个结合MHC II的最小表位的新表位排名高于包含少于8个结合MHC II的最小表位的新表位。在一些实施方式中,包含至少9个结合MHC II的最小表位的新表位排名高于包含少于9个结合MHC II的最小表位的新表位。在一些实施方式中,包含至少10个结合MHC II的最小表位的新表位排名高于包含少于10个结合MHC II的最小表位的新表位。

[0137] 结合差异

[0138] 在一些实施方式中,将包含具有与相应参考肽的MHC I结合亲和力相比更高或高得多的MHC I结合亲和力的最小表位的新表位优先于包含与相应的参考MHC I结合亲和力相比仅略微改善了MHC I结合亲和力的最小表位的新表位。

[0139] 因此,新表位与相应参考肽之间的%排名得分(MHC I)比率可用于选择新表位,以进行潜在的临床应用。在本文中,对于MHC I结合的%排名得分比率也称为MHC I结合差异。

[0140] 同样,在一些实施方式中,具有比相应参照肽的MHC II结合亲和力更高或高得多的MHC II结合亲和力的新表位优先于表达与相应的参考MHC II结合亲和力相比仅略微改善了MHC II结合亲和力的肽的新表位。

[0141] 因此,新表位与相应参考肽之间的%排名得分(MHC II)比率可用于选择新表位,以进行潜在的临床应用。在本文中,对于MHC II结合的%排名得分比率也称为MHC II结合差异。

[0142] 参考肽和新表位之间的MHC I结合差异由下式给出:(参考的%排名得分(MHC I))/(新表位的%排名得分(MHC I)),其中参考的%排名得分(MHC I)是参考肽与MHC I分子之间的预测结合亲和力,而新表位的%排名得分(MHC I)是新表位肽与MHC I分子之间的预测结合亲和力。

[0143] 在本发明的一个实施方式中,上述方法的步骤b或c进一步包括确定由下式给出的MHC I结合差异:(参考的%排名得分(MHC I))/新表位的%排名得分(MHC I),其中具有高MHC I结合差异的新表位排名高于具有较低MHC I结合差异的新表位。

[0144] 在一个实施方式中,如果MHC I结合差异(参考的%排名得分(MHC I))/(新表位的%排名得分(MHC I))低于2,则降低新表位的优先排序。

[0145] 在一个实施方式中,如果MHC I结合差异((参考的%排名得分(MHC I))/(新表位的%排名得分(MHC I))低于20(诸如低于3,诸如低于1),则降低新表位的优先排序。如果结合差异等于或大于1(诸如等于或大于3,诸如等于或大于20),则认为新表位具有潜在的临床相关性。

[0146] 以类似的方式,参考肽和最小表位之间的MHC I结合差异由下式给出:(参考的%排名得分(MHC I))/(最小表位的%排名得分(MHC I)),其中参考的%排名得分(MHC I)是参考肽与MHC I分子之间的预测结合亲和力,并且最小表位的%排名得分(MHC I)是最小表位肽与MHC I分子之间的预测结合亲和力。

[0147] 在本发明的一个实施方式中,上述方法的步骤b或c进一步包括确定由下式给出的MHC I结合差异:(参考的%排名得分(MHC I))/(最小表位的%排名得分(MHC I)),其中包

含具有高MHC I结合差异的最小表位的新表位排名高于包含具有较低MHC I结合差异的最小表位的新表位。

[0148] 在一个实施方式中,如果MHC I结合差异((最小表位的参考序列的%排名得分(MHC I))/(最小表位的%排名得分(MHC I))低于20(诸如低于3,诸如低于1),则降低新表位的优先排序。如果新表位所包含的最小表位的结合差异等于或大于1(诸如等于或大于3,诸如等于或大于20),则认为所述新表位具有潜在的临床相关性。

[0149] 如果新表位仅包含一个结合MHC I的最小表位,则所述新表位的MHC I结合差异等于所述最小表位的MHC I结合差异。如果新表位包含两个或更多个与MHC I结合的最小表位,则认为新表位的MHC I结合差异等于在所述两个或更多个与MHC I结合的最小表位中最高MHC I结合差异。同样,如果新表位仅包含一个结合MHC II的最小表位,则新表位的MHC II结合差异等于所述最小表位的MHC II结合差异。如果新表位包含两个或更多个与MHC II结合的最小表位,则认为新表位的MHC II结合差异等于在所述两个或更多个与MHC II结合的最小表位中最高MHC II结合差异。关于参考肽也是如此。

[0150] 在一些实施方式中,还考虑了突变的位置。换句话说,新表位或最小表位的临床实用性不仅根据以上所述来确定,而且还根据突变位置的函数来确定。在优选的实施方式中,突变是免疫原性的。MHC分子的肽结合沟容纳肽或肽片段,通常长度为9个氨基酸,即最小表位。最小表位与MHC分子之间的接触是通过锚定残基的侧链介导的。以长度为9个氨基酸的最小表位为例,最小表位对MHC I分子的锚定位置为位置2和9。为了与MHC II分子结合,锚定位置为位置1、4和9。

[0151] 在一些实施方式中,排除以下新表位或降低以下新表位的优先排序,而不管其结合差异如何:包含其中突变在非锚定位置处的最小表位的新表位。对于结合MHC I的最小表位,这意味着在一些实施方式中,如果突变位于MHC分子结合沟内所容纳的最小表位或部分最小表位的位置1、3、4、5、6、7或8处,则将新表位排除在外或降低其优先排序,不管其结合差异有多高。优选的新表位包含在锚定位置具有突变且结合差异等于或大于2(诸如等于或大于4,诸如等于或大于6,诸如等于或大于8,诸如等于或大于10,诸如等于或大于12,诸如等于或大于14,诸如等于或大于16,诸如等于或大于18,最优选诸如等于或大于20)的最小表位。

[0152] MHC II结合差异也可以用作选择标准。参考肽和新表位之间的MHC II结合差异由下式给出:(最小表位的参考序列的%排名得分(MHC II))/(最小表位的%排名得分(MHC II)),其中最小表位的%排名得分(MHC II)是新表位中排名最优的最小表位的预测结合亲和力,并且参考的%排名得分(MHC II)是参考序列中相应肽的预测结合亲和力。

[0153] 在本发明的一个实施方式中,上述方法的步骤b或c进一步包括确定由下式给出的MHC II结合差异:(最小表位的参考序列的%排名得分(MHC II))/最小表位的%排名得分(MHC II),其中具有高MHC II结合差异的新表位排名高于具有较低MHC II结合差异的新表位。

[0154] 以类似的方式,参考肽和最小表位之间的MHC II结合差异由下式给出:(参考的%排名得分(MHC II))/(最小表位的%排名得分(MHC II)),其中参考的%排名得分(MHC II)是参考肽与MHC II分子之间的预测结合亲和力,并且最小表位的%排名得分(MHC II)是最小表位肽与MHC II分子之间的预测结合亲和力。

[0155] 上面描述了确定MHC II结合亲和力的方法。

[0156] 在本发明的一个实施方式中,将以下新表位排除在外或降低其优先排序:该新表位包含结合MHC II的最小表位,所述最小表位预测以低于20(诸如低于15,诸如低于14,诸如例如低于13,低于12或例如低于11)的%排名得分(MHC II)结合MHC II。

[0157] 在优选的实施方式中,将以下新表位排除在外或降低其优先排序:该新表位包括具有小于10的%排名得分(MHC II)的结合MHC II的最小表位。

[0158] 在另一个实施方式中,将以下新表位排除在外或降低其优先排序:该新表位包含具有等于或小于9(诸如等于或小于8,诸如例如等于或小于7,等于或小于6,诸如等于或小于5,诸如例如等于或小于4,等于或小于3或例如等于或小于2)的%排名得分(MHC II)的结合MHC II的最小表位。

[0159] 下面给出了可以在上述方法的步骤(例如步骤d)中使用的计分方案的示例:

[0160] • 具有结合差异 $<1$ 的突变(无论突变的位置如何)的最小表位:得分等于1;

[0161] • 在非锚定位置具有突变且 $1 \leq$ 结合差异 $<3$ 的最小表位:得分等于2;

[0162] • 在锚定位置具有突变且 $1 \leq$ 结合差异 $<3$ 的最小表位:得分等于3;

[0163] • 在非锚定位置具有突变且 $3 \leq$ 结合差异 $<20$ 的最小表位:得分等于4;

[0164] • 在锚定位置具有突变且 $3 \leq$ 结合差异 $<20$ 的最小表位:得分等于5;

[0165] • 在非锚定位置具有突变且结合差异 $>20$ 的最小表位:得分等于6;

[0166] • 在锚定位置具有突变且结合差异 $>20$ 的最小表位:得分等于7;

[0167] 结合差异可以是MHCI或MHCII结合差异。

[0168] 将理解的是,以上给出的得分是任意值,并且可以由其他任意值代替,只要得分如以上示例性地增加即可。通常,可以看出,无论突变的位置如何,具有低结合差异(例如小于1)的最小表位或表位都排名最低。对于结合差异在相同范围内(例如 $1 \leq$ 结合差异 $<3$ ;或 $3 \leq$ 结合差异 $<20$ ;或结合差异 $>20$ )的最小表位或表位,在锚定位置处包含突变的最小表位或新表位评级高于在非锚定位置处包含突变的最小表位或新表位。通常,结合差异越高,得分越高。对于类似的得分,锚定位置处的突变比非锚定位置处的突变得分更高。

[0169] 在上述评分方案中,排名最低的最小表位在非锚定位置处具有突变,或在锚定位置处具有突变且结合差异小于1。

[0170] 新表位的长度

[0171] 新表位序列优选具有适合于加工和在MHC分子上呈递新表位中包含的最小表位的长度。因此,在一个实施方式中,新表位的长度为7至40个氨基酸,例如10至35个氨基酸,或更优选地15至30个氨基酸,例如25至30个氨基酸。在优选的实施方式中,新表位是27个氨基酸。新表位的长度包括结合MHC I和MHC II两者的新表位的长度。

[0172] 优选的是,突变基本上位于新表位序列的中间。

[0173] 在一些实施方式中,给定新表位的最小表位由小于或等于新表位的氨基酸数目的氨基酸数目组成。在新表位是最小表位的情况下,新表位和最小表位具有相同的长度。在这样的实施方式中,最小表位的长度为7至40个氨基酸,例如10至35个氨基酸,或更优选地15至30个氨基酸,例如25至30个氨基酸。在一个实施方式中,最小表位的长度是27个氨基酸。最小表位的长度包括结合MHC I和MHC II两者的新表位的长度。

[0174] 在其他实施方式中,新表位内包含的最小表位短于新表位。例如,新表位具有7至

40个氨基酸的长度,并且相应的最小表位具有6至39个氨基酸的长度,或者具有比新表位的长度短至少一个氨基酸(例如至少2个氨基酸,例如至少3个氨基酸,例如至少4个氨基酸,例如至少5个氨基酸,例如至少6个氨基酸,例如至少7个氨基酸,例如至少8个氨基酸,例如至少9个氨基酸,例如至少10个氨基酸,例如至少11个氨基酸,例如至少12个氨基酸,例如至少13个氨基酸,例如至少15个氨基酸,例如至少16个氨基酸,例如至少17个氨基酸,例如至少18个氨基酸,例如至少19个氨基酸,例如至少20个氨基酸)的长度。在优选的实施方式中,新表位为27个氨基酸,并且最小表位的长度为6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25或26个氨基酸。

[0175] 在一些实施方式中,最小表位结合或预测结合MHC I,并且最小表位的长度为5至20个氨基酸,例如6至19个氨基酸,例如7至18个氨基酸,例如7至17个氨基酸,例如7至16个氨基酸,例如8至15个氨基酸,例如8至14个氨基酸。通常,这种最小表位的长度为8至14个氨基酸。

[0176] 在一些实施方式中,最小表位结合或预测结合MHC II,并且最小表位的长度为5至20个氨基酸,例如6至19个氨基酸,例如7至18个氨基酸,例如7至17个氨基酸,例如8至16个氨基酸,例如9至15个氨基酸。通常,这种最小表位的长度为9至15个氨基酸。

[0177] 与已知的T细胞识别表位相似

[0178] 在一些实施方式中,该方法进一步包括选择一组新表位的步骤,所述新表位包含与已知被T细胞识别的表位高度相似的最小表位。技术人员知道在哪里可以找到这些表位的列表。例如,可以从免疫表位数据库和分析资源(IEDB数据库)获得已知被T细胞识别的表位,其中列出了人类感染性表位(<https://www.iedb.org/>)。因此可以在数据库中运行被认为可能具有临床实用性的新表位中包含的最小表位的序列,以确定最小表位是否与已知被T细胞识别的表位具有高度相似性。

[0179] 当应用于两个序列时,术语“高度相似性”是指这些序列具有显著的相似性,即相同和相似匹配的总和很高。相似性可以通过计算比对得分来确定,例如使用BLOSUM62。技术人员知道哪个比对得分阈值可用于确定两个序列是否彼此具有高度相似性。例如,对于9聚体表位,比对得分>26可以适当地用作选择具有高度相似性的表位的阈值。

[0180] 归一化的比对分数也可以用于评估两个肽序列是否彼此具有高度相似性。例如,归一化的比对得分可以是对于IEDB数据库中给定的最小表位和给定的表位获得的BLOSUM62比对得分,除以所述最小表位与自身之间的比对得分(对应于该最小表位的最大比对得分)。为了获得相似度的度量,然后从1中减去结果。换句话说,可以执行以下计算:

[0181] 与已知表位的相似度 =  $1 - (\text{最小表位比对得分}) / (\text{已知表位比对得分})$

[0182] 分数越高,与已知表位的相似度越高。

[0183] 最小表位总数

[0184] 有用的新表位包含至少一个最小表位,所述最小表位包含突变例如免疫原性突变,并且最小表位的长度为5至20个氨基酸,优选6至19个氨基酸,诸如7至18个氨基酸,例如8至17个氨基酸,诸如9至16个氨基酸,例如10至15个氨基酸,诸如11至14个氨基酸,例如12或13个氨基酸,诸如5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个氨基酸。优选地,最小表位具有8至14个氨基酸的长度。新表位可包含可以重叠的若干个最小表位。因此,新表位可以包含1至100个最小表位,例如1至90个最小表位,例如1至80个最小表位,例如1至

70个最小表位,例如1至60个最小表位,例如1至50个最小表位,例如1至40个最小表位,例如1至30个最小表位,例如1至20个最小表位,例如2至19个最小表位,例如3至18个最小表位,例如4至17个最小表位,例如5至16个最小表位,例如6至15个最小表位,例如7至14个最小表位,例如8至13个最小表位,例如9至12个最小表位,例如10或11个最小表位,所有最小表位均包含免疫原性突变。最小表位对MHC分子例如MHC I分子或MHC II分子具有亲和力。新表位中包含的突变不一定位于最小表位的中心,但是在一些实施方式中可以位于最小表位的中心。因此,突变可以在最小表位的第一个或最后一个残基处,或在它们之间的任何其他残基处。如上所述,在优选的实施方式中,突变处于锚定位置。在一些实施方式中,在非锚定位置具有突变的最小表位被排除或降低其优先排序。

[0185] 通常,新表位可以包含n个不同的最小表位,所述最小表位的长度为k个氨基酸,全部包含突变。给定的新表位可以包含不同长度的最小表位。

#### [0186] 自然发生突变的可能性

[0187] 在优选的实施方式中,本发明的方法进一步包括确定新表位中存在的突变自然发生的概率的步骤。改变蛋白质中的氨基酸可能会降低其执行其功能的能力,或甚至改变其功能。在细胞中具有重要功能的蛋白质的变化可能会导致细胞死亡。相反地,这种变化可以使细胞以不同的方式继续发挥功能,并且甚至可以导致相比于原始蛋白质的有效/有利的修饰,从而使突变传递给生物体的后代。如果这种变化没有对后代造成任何显著的生理缺点,那么这种突变将在种群中持续存在的可能性仍然存在。由于氨基酸的物理和化学性质差异很大,因此将它们分为具有相似性质的组。与用不同类别的氨基酸替代相比,用同一组中的另一种氨基酸置换对蛋白质的结构和功能的影响较小。因此,在一些实施方式中,使包含来自于与参考氨基酸不同的理化基团的突变氨基酸的新表位优先于包含来自于与参考氨基酸相同的理化基团的突变氨基酸的新表位。

[0188] 优选地,该突变引起氨基酸置换。如本文所用,术语“置换对”是指由被置换的氨基酸和置换其的氨基酸组成的对。

[0189] 因此,在本发明的优选实施方式中,本文及以上描述的方法的步骤d进一步包括确定反映突变氨基酸序列与参考氨基酸序列之间的相似性的得分。

[0190] 为了计算具体氨基酸置换的分数,可以使用置换矩阵。置换矩阵包含基于所有可用蛋白质序列中观察到的突变频率的概率或对数得分。对数几率(log odd)得分越低,在将天然存在的氨基酸序列相互比较时,观察到该氨基酸置换的可能性就越小。还已经表明,低对数几率得分与氨基酸置换对之间理化性质的高差异良好相关。例如,相比于由于新形成的突变肽(表位)与参考肽之间的理化性质差异较大而具有高对数几率得分的置换,具有低对数几率得分或从进化的角度看发生该突变的较低可能性/频率的置换,具有更高的机会被T细胞受体发现。T细胞受体不仅具有耐受自身肽的能力,而且具有耐受与自身肽高度相似的肽(即具有高对数几率的突变肽)的能力。这种机制称为中心耐受。

[0191] 矩阵计算中使用的概率是通过查看在多个蛋白质比对中发现的保守序列的“块”来计算的。假定这些保守序列在相关蛋白质中具有功能重要性,因此其置换率将低于保守程度较低的区域。为了减少密切相关的序列对置换率的偏倚,已将序列同一性高于某个阈值的区块中的区段进行了聚簇,从而降低了每个此类聚簇的权重(Henikoff, S; Henikoff, JG (1992). "Amino acid substitution matrices from protein blocks". Proceedings

of the National Academy of Sciences of the United States of America.89(22): 10915-9.)。对于BLOSUM62矩阵,此阈值已设置为62%。然后对聚簇之间的配对频率进行计数,因此仅在同一性小于62%的区段之间对配对进行计数。人们会使用编号较高的BLOSUM矩阵来比对两个紧密相关的序列,而使用较低编号的BLOSUM矩阵来用于更发散的序列。

[0192] 因此,在本发明的优选实施方式中,反映新表位中存在的氨基酸置换随机发生的概率的得分是使用基于进化的评分矩阵来确定的。在更优选的实施方式中,所述评分矩阵是对数几率矩阵。在特别优选的实施方式中,所述矩阵是BLOSUM矩阵,优选地是BLOSUM62矩阵。

[0193] 为计算BLOSUM矩阵,使用以下等式:

$$[0194] \quad S_{ij} = (1/\lambda) * \log(p_{ij}/(q_i * q_j))$$

[0195]  $p_{ij}$ 是两个氨基酸i和j在同源序列中相互替换的概率,并且 $q_i$ 和 $q_j$ 是在任何蛋白质序列中找到氨基酸i和j的背景概率。因子 $\lambda$ 是比例因子,设置为使矩阵包含易于计算的整数值。

[0196] 包含其中氨基酸置换对具有较低BLOSUM62得分的突变的新表位排名高于具有较高BLOSUM62得分的氨基酸置换对的新表位。

[0197] 在一个实施方式中,将这样的新表位降低优先排序或将其排除:该新表位不包含与BLOSUM62得分小于3(例如小于2)的氨基酸置换对连接的突变。

[0198] 在本发明的优选的实施方式中,将这样的新表位降低优先排序或将其排除:该新表位不包含BLOSUM62得分低于1的氨基酸置换对。因此,优选地将这样的新表位优先排序或选择用于临床应用:该新表位包含至少一个BLOSUM62得分低于1的氨基酸置换对。

[0199] 在本发明的另一个实施方式中,将这样的新表位降低优先排序或将其排除:该新表位不包含BLOSUM62得分低于1的氨基酸置换对。因此,在一个实施方式中,将这样的新表位优先排序或选择用于临床应用:该新表位包含至少一个BLOSUM62得分低于1的氨基酸置换对。

[0200] 对于待选择用于潜在临床应用的结合MHC I的新表位,优选的是,新表位至少满足以下两个标准i和ii中的一个:

[0201] i. 高MHC I结合差异

[0202] ii. 低BLOSUM62得分。

[0203] 在一个实施方式中,满足以下两个标准i和ii中至少一个的结合MHC I的新表位:

[0204] i. 新表位的MHC I结合差异大于1;

[0205] ii. 新表位包含BLOSUM62得分低于3(例如低于2,优选低于1)的氨基酸置换

[0206] 所述新表位的排名高于不满足标准i和ii中任一项的结合MHC I的新表位。

[0207] 在优选的实施方式中,满足以下两个标准i和ii中至少一个的结合MHC I的新表位:

[0208] i. 新表位的MHC I结合差异大于3;

[0209] ii. 新表位包含BLOSUM62得分低于3(例如低于2,优选低于1)的氨基酸置换

[0210] 所述新表位的排名高于不满足标准i和ii中任一项的结合MHC I的新表位。

[0211] 在甚至更优选的实施方式中,满足以下两个标准i和ii中至少一个的结合MHC I的新表位:

- [0212] i. 新表位的MHC I结合差异大于2;
- [0213] ii. 新表位包含BLOSUM62得分低于3(例如低于2, 优选低于1)的氨基酸置换对
- [0214] 所述新表位的排名高于不满足标准i和ii中任一项的结合MHC I的新表位。
- [0215] 在一些实施方式中, 优选地将不满足以上标准i和ii中任一项的结合MHC I的新表位排除或降低其优先排序。此外, 在一些实施方式中, 如上文所述, 将满足条件i但其中突变为非锚定位置的结合MHC I的新表位排除或降低其优先排序。
- [0216] 对于待选择用于潜在临床应用的结合MHC II的新表位, 优选的是, 满足以下两个标准i和ii中的至少一个:
- [0217] i. 高MHC II结合差异
- [0218] ii. 低BLOSUM62得分。
- [0219] 在一个实施方式中, 满足以下两个标准i和ii中至少一个的结合MHC II的新表位:
- [0220] i. 新表位的MHC II结合差异大于2;
- [0221] ii. 新表位包含BLOSUM62得分低于3(例如低于2或优选低于1)的氨基酸置换对
- [0222] 所述新表位的排名高于不满足标准i和ii中任一项的结合MHC II的新表位。
- [0223] 在优选的实施方式中, 满足以下两个标准i和ii中至少一个的结合MHC II的新表位:
- [0224] i. 新表位的MHC II结合差异大于3;
- [0225] ii. 新表位包含BLOSUM62得分低于3(例如低于2或优选低于1)的氨基酸置换对
- [0226] 所述新表位的排名高于不满足标准i和ii中任一项的结合MHC II的新表位。
- [0227] 在甚至更优选的实施方式中, 满足以下两个标准i和ii中至少一个的结合MHC II的新表位:
- [0228] i. 新表位的MHC II结合差异大于2;
- [0229] ii. 新表位包含BLOSUM62得分低于1的氨基酸置换对
- [0230] 所述新表位的排名高于不满足标准i和ii中任一项的结合MHC II的新表位。
- [0231] 优选地将不满足以上标准i和ii中任一项的结合MHC II的新表位排除或降低其优先排序。此外, 在一些实施方式中, 如上文所述, 将满足条件i但其中突变为非锚定位置的结合MHC II的新表位排除或降低其优先排序。
- [0232] RNA表达水平
- [0233] 发现突变的源基因的RNA表达水平可能是新表位免疫原性的重要因素。较高的RNA表达水平通常会导致蛋白质的较高表达以及在肿瘤细胞上MHC分子上新表位的呈递, 从而具有更高的诱发癌症特异性免疫原性的潜力。
- [0234] 因此, 在优选的实施方式中, 本发明的方法进一步包括确定新表位的RNA表达水平的步骤。
- [0235] 通常, 具有高RNA表达水平的新表位排名高于具有较低RNA表达水平的新表位。
- [0236] 在特定的实施方式中, 在本方法的步骤d中, 具有高RNA表达水平的结合MHC I的新表位排名高于具有较低RNA表达水平的结合MHC I的新表位。在另一个特定的实施方式中, 具有高RNA表达水平的结合MHC II的新表位排名高于具有较低RNA表达水平的结合MHC II的新表位。
- [0237] 优选地, 将未检测到源基因RNA的新表位排除或降低其优选排序。尤其是, 将归一

化转录本水平表达为每百万0个转录本 (TPM) 的新表位排除或降低其优选排序。不排除具有可检测的归一化转录本水平表达 (例如高于0TPM的转录本水平表达) 的转录本。

[0238] RNA表达水平可以通过对从如上定义的肿瘤样品获得的核酸进行RNA测序来确定。RNA测序可以通过本领域已知的技术进行。确定RNA表达水平的方法是技术人员众所周知的。例如,可以使用新一代测序技术例如Illumina平台来确定RNA表达水平。

[0239] 自身免疫风险

[0240] 为了使交叉反应的风险最小化,优选地将以下新表位排除或降低其优选排序:所述新表位包含与人蛋白质组中的肽序列匹配的天然存在的核心肽 (例如包含突变的长度为9个氨基酸的肽)。因此,本发明的方法可以进一步包括将新表位肽序列与人蛋白质组的肽序列进行比较的步骤。

[0241] 在此,人类蛋白质组定义为源自人类基因组已知基因的所有翻译物。已知基因发布于www.ensembl.org上。因此,为了确定新表位核心肽序列是否与人蛋白质组中的肽序列匹配,可以将该序列与从数据库获得的人基因组的肽序列进行比较。

[0242] 新表位核心肽序列包含在如上所述的新表位中。优选地,将以下新表位排除或降低其优先排序:所述新表位包含含有突变的肽序列,该肽序列与人蛋白质组中的天然存在的肽序列相匹配。

[0243] 这是因为,不受理论的束缚,预期与天然存在的肽序列匹配的突变由于中心耐受性而具有被T细胞受体耐受的更大风险,并且因此预期相比于不与天然存在的序列匹配的突变,所述突变的免疫原性更低。

[0244] 在一个实施方式中,与人基因组的肽序列比较的新表位核心肽序列包含9至27个氨基酸 (诸如例如10至25个氨基酸,诸如11至15个氨基酸或诸如12、13或14个氨基酸),或由其组成。

[0245] 在一个实施方式中,与人基因组的肽序列比较的新表位肽序列包含5至15个氨基酸 (诸如例如6至12个氨基酸,优选7至11个氨基酸,或更优选8至10个氨基酸),或由其组成。

[0246] 在优选的实施方式中,与人基因组的肽序列比较的所述新表位肽序列包含9个氨基酸或由9个氨基酸组成。

[0247] 在另一个实施方式中,与人基因组的肽序列比较的所述新表位肽序列包含15个氨基酸,14个氨基酸,13个氨基酸,12个氨基酸,11个氨基酸或例如10个氨基酸,或由其组成。

[0248] 新表位肽序列包含新表位突变。在优选的实施方式中,所述突变是氨基酸置换。对于技术人员将清楚的是,可以将所述新表位中包含的最小表位的序列与人基因组的肽序列进行比较,而不是将给定的新表位的序列与人基因组的肽序列进行比较。

[0249] 为了进一步使器官特异性自身免疫的风险最小化,在一些实施方式中,将在特定器官/组织中高表达的基因中发现的新表位排除在外或降低其优先排序。

[0250] 在一个实施方式中,将在下述基因中存在的新表位或最小表位排除在外或降低其优先排序,其中所述基因在任何器官中显示出比其他组织高至少3倍,例如高至少4倍的RNA表达水平。

[0251] 在优选的实施方式中,将在下述基因中存在的新表位或最小表位 (诸如结合MHC I 和/或MHC II的新表位) 排除在外或降低其优先排序,其中所述基因在任何器官中显示出相比于其他组织至少5倍高的RNA表达水平。

[0252] 在另一个实施方式中,将在下述基因中存在的新表位或最小表位(诸如结合MHC I和/或MHC II的新表位或最小表位)排除在外或降低其优先排序,其中所述基因在任何器官中显示出比其他组织高至少6倍,至少7倍或诸如至少8倍的RNA表达水平。

[0253] 优选的是,器官和组织来自相同物种,诸如人。RNA表达水平可以例如从数据库例如www.gtexportal.org获得。RNA表达水平也描述于Uhlén,M.等人(2015)(Uhlén,M.等人(2015)“Proteomics.Tissue-based map of the human proteome.”Science;347(6220))中。或者,从新表位所源自的个体获得RNA表达水平。

[0254] 器官可以例如选自心脏和脑。在另一个实施方式中,所述器官选自肝,肺,胃,肾,脾,结肠和肠。

#### [0255] 等位基因频率

[0256] 在一个实施方式中,本发明的方法包括确定每个新表位中存在的突变的等位基因频率的步骤。如本文所用,术语等位基因频率是指与在特定基因组区域的所有等位基因的序列总数相比,在该特定基因座处包含特定基因组突变的等位基因序列的相对频率。术语突变等位基因频率是指包含突变或新表位突变的等位基因频率,在本文中与术语变体等位基因频率(VAF)可互换使用。通常,高突变等位基因频率或VAF表示更高的临床实用性。高突变等位基因频率或VAF表示含有相同突变的癌细胞比例更高。因此,不受理论的束缚,预期具有高等位基因频率或VAF的新表位比具有低突变等位基因频率或VAF的新表位在癌细胞中以更高的比例存在。

[0257] 在特定的实施方式中,具有高突变等位基因频率或VAF的结合MHC I的新表位排名高于具有较低突变等位基因频率或VAF的结合MHC I的新表位。在另一个特定的实施方式中,具有高突变等位基因频率或VAF的结合MHC II的新表位排名高于具有较低突变等位基因频率或VAF的结合MHC II的新表位。

[0258] 突变的等位基因频率可以例如通过分析来自肿瘤样品的测序数据并将突变等位基因的频率与相同基因中存在的其他等位基因的频率进行比较来确定。

[0259] 在一些实施方式中,大于0.05的突变等位基因频率或VAF表示包含所述突变的最小表位或新表位的临床实用性。因此,在一些实施方式中,包含VAF大于0.05(例如大于0.06,例如大于0.07,例如大于0.08,例如大于0.09,例如大于0.1,例如大于0.15,例如大于0.20)的突变的新表位或最小表位,其排名分别高于包含VAF等于或小于0.05(例如等于或小于0.06,例如等于或小于0.07,例如等于或小于0.08,例如等于或小于0.09,例如等于或小于0.1,例如等于或小于0.15,例如等于或小于0.20)的突变的新表位或最小表位。

#### [0260] 变体调用

[0261] 本领域可获得许多类型的体细胞突变鉴定软件(变体调用器)。在一些实施方式中,如果新表位或最小表位包含的突变被至少两个不同的变体调用器鉴定,则认为它们可能具有临床实用性。因此,在一些实施方式中,新表位或最小表位的排名取决于多少个变体调用器鉴定它们所包含的突变。包含由至少两个不同的变体调用器鉴定的突变的新表位或最小表位的排名高于包含仅由一个变体调用器鉴定的突变的新表位或最小表位。在一些实施方式中,突变由至少3个不同的变体调用器(例如至少4个不同的变体调用器,例如至少5个不同的变体调用器,例如至少6个不同的变体调用器,例如至少7个不同的变体调用器,例如至少8个不同的变体调用器,例如至少9个不同的变体调用器,例如10个或更多个不同的

变体调用器)来鉴定。

[0262] 与已知的癌症相关基因的相似性

[0263] 在一些实施方式中,将源自已知癌症相关基因的新表位优先排序。

[0264] 癌症相关基因被定义为参与癌症发病的基因。可以在以下中找到癌症相关的基因:例如,“癌症中的体细胞突变目录”(COSMIC)数据库,或包含癌症相关突变和/或基因的首选(curated)列表的其他数据库,如本领域技术人员所知的。因此,通过将新表位的来源基因、基因组位置或变体位置(HGVS-命名)与COSMIC数据库中的相应信息进行比较,可以找到与已知癌症相关基因的匹配。

[0265] 在一些实施方式中,如果新表位所包含的基因是已知与癌症相关的基因,则认为新表位可能具有临床实用性。因此,这样的新表位优先于包含在与癌症无关的基因中的新表位。

[0266] 克隆性

[0267] 如本文所用,术语“克隆性”是指肿瘤样品中新表位的出现。如果例如在一个以上肿瘤活检中发现新表位或最小表位,则表明高克隆性,并且因此在大多数癌细胞中存在更高的可能性,并且因此更高可能性地具有临床实用性。

[0268] 新表位的克隆性可以通过从相同个体中获得至少两个肿瘤样本并确定突变是否存在于一个以上的活检中来确定。优选地,在一个以上肿瘤样品中发现的新表位优先于仅在一个活检中发现的新表位。“一个以上的肿瘤样品”被理解为源自相同个体的若干个样品,这些样品来自相同的病变或不同的病变。

[0269] 在一些实施方式中,与具有在仅一个肿瘤样品中发现的突变的最小表位或新表位相比,具有在来自同一个体的至少两个不同样品中发现的突变的最小表位或新表位可能具有更高的临床实用性。在一些实施方式中,至少两个不同的样品是来自以下的样品:

[0270] -同一肿瘤或病变,例如来自同一肿瘤或病变的多角度活检

[0271] -至少两个不同的肿瘤或病变

[0272] -至少一个肿瘤或病变和至少一个档案肿瘤材料,例如来自肿瘤或病变的档案活检或档案切除材料,

[0273] 在至少两个不同的样品中发现包含突变的表位或新表位表明具有较高的临床实用性。术语“档案肿瘤材料”是指在较早的时间点从相同个体的肿瘤组织中采样的材料,例如来自肿瘤或病变的档案活检或档案切除材料。

[0274] 在一些实施例中,至少两个不同的样品是至少三个不同的样品,例如至少四个不同的样品,例如至少五个不同的样品,例如至少六个不同的样品,例如至少七个不同的样品,例如至少八个不同的样品,例如至少九个不同的样品,例如至少十个不同的样品或更多。在一些实施方式中,至少两个不同样品中的至少一个是来自个体的档案肿瘤材料,并且至少两个不同样品中的至少一个是来自肿瘤或病变的样品。

[0275] 液体活检

[0276] 当选择包含可能具有临床实用性的最小表位的新表位时,可能有利的是确定新表位或最小表位是否也发现于个体的血浆中。因此,在一些实施方式中,该方法包括将来自无细胞DNA中(例如在获自个体的液体活检的血浆中的循环肿瘤DNA)的核酸序列进行测序。这对于鉴定、确认和/或追踪新表位可能是有用的,并且例如可以通过对无细胞DNA进行测序

来完成。

[0277] 在一些实施方式中,在来自个体的血浆样品中的循环肿瘤DNA中发现的新表位排名高于未在所述样品中发现的新表位。

[0278] 个体

[0279] 根据本发明的方法的个体优选是人。优选地,个体或人是癌症患者。因此,个体优选是患有癌症的人。癌症可以是所有类型的癌症。优选地,个体患有至少一种肿瘤。

[0280] 癌症可以是其中癌细胞包含至少一个突变的任何癌症。癌症可以是原发肿瘤,转移瘤或二者。检查突变的肿瘤可能是原发肿瘤或转移瘤。待治疗的癌症可以是已知具有高突变负荷的癌症,例如黑色素瘤或肺癌。待治疗的癌症也可以是仅以一种癌症特异性突变为特征的癌症。

[0281] 进一步的标准

[0282] 优选的是,结合MHC I和/或结合MHC II的新表位满足以下标准i-vii中的至少一个:

[0283] i. 高RNA表达水平

[0284] ii. 高等位基因频率

[0285] iii. 高克隆性

[0286] iv. 匹配已知的癌症相关基因

[0287] v. 包含由至少两个不同的变体调用器鉴定的突变

[0288] vi. 还发现于个体的血浆中,如例如通过对无细胞DNA进行测序所确定的

[0289] vii. 还发现于来自相同个体的至少两个样品中,例如来自相同病变或来自两个不同病变的两个样品,或至少一个肿瘤或病变和至少一个来自肿瘤或病变的档案肿瘤材料样品。

[0290] RNA表达水平,等位基因频率,克隆性,与已知癌症相关基因的匹配,至少两个不同变体调用器的鉴定,个体血浆中的存在以及来自相同个体的至少两个样品中的存在是如上文所定义的。

[0291] 在更优选的实施方式中,结合MHC I和/或结合MHC II的新表位满足以下标准i-vii中的至少两个:

[0292] i. 高RNA表达水平

[0293] ii. 高等位基因频率

[0294] iii. 高克隆性

[0295] iv. 匹配已知的癌症相关基因

[0296] v. 包含由至少两个不同的变体调用器鉴定的突变

[0297] vi. 还发现于个体的血浆中,如例如通过对无细胞DNA进行测序所确定的

[0298] vii. 还发现于来自相同个体的至少两个样品中,例如来自相同病变或来自两个不同病变的两个样品,或至少一个肿瘤或病变和至少一个来自肿瘤或病变的档案肿瘤材料样品。

[0299] 在甚至更优选的实施方式中,结合MHC I和/或结合MHC II的新表位满足以下标准i-vii中的至少三个:

[0300] i. 高RNA表达水平

- [0301] ii. 高等位基因频率
- [0302] iii. 高克隆性
- [0303] iv. 匹配已知的癌症相关基因
- [0304] v. 包含由至少两个不同的变体调用器鉴定的突变
- [0305] vi. 还发现于个体的血浆中, 如例如通过对无细胞DNA进行测序所确定的
- [0306] vii. 还发现于来自相同个体的至少两个样品中, 例如来自相同病变或来自两个不同病变的两个样品, 或至少一个肿瘤或病变和至少一个来自肿瘤或病变的档案肿瘤材料样品。
- [0307] 在优选的实施方式中, 结合MHC I和/或结合MHC II的新表位满足以下标准i-vii中的至少四个:
- [0308] i. 高RNA表达水平
- [0309] ii. 高等位基因频率
- [0310] iii. 高克隆性
- [0311] iv. 匹配已知的癌症相关基因
- [0312] v. 包含由至少两个不同的变体调用器鉴定的突变
- [0313] vi. 还发现于个体的血浆中, 如例如通过对无细胞DNA进行测序所确定的
- [0314] vii. 还发现于来自相同个体的至少两个样品中, 例如来自相同病变或来自两个不同病变的两个样品, 或至少一个肿瘤或病变和至少一个来自肿瘤或病变的档案肿瘤材料样品。
- [0315] 在另一个优选的实施方式中, 结合MHC I和/或结合MHC II的新表位满足以下标准i-vii中的至少五个:
- [0316] i. 高RNA表达水平
- [0317] ii. 高等位基因频率
- [0318] iii. 高克隆性
- [0319] iv. 匹配已知的癌症相关基因
- [0320] v. 包含由至少两个不同的变体调用器鉴定的突变
- [0321] vi. 还发现于个体的血浆中, 如例如通过对无细胞DNA进行测序所确定的
- [0322] vii. 还发现于来自相同个体的至少两个样品中, 例如来自相同病变或来自两个不同病变的两个样品, 或至少一个肿瘤或病变和至少一个来自肿瘤或病变的档案肿瘤材料样品。
- [0323] 在另一个优选的实施方式中, 结合MHC I和/或结合MHC II的新表位满足以下标准i-vii中的六个:
- [0324] i. 高RNA表达水平
- [0325] ii. 高等位基因频率
- [0326] iii. 高克隆性
- [0327] iv. 匹配已知的癌症相关基因
- [0328] v. 包含由至少两个不同的变体调用器鉴定的突变
- [0329] vi. 还发现于个体的血浆中, 如例如通过对无细胞DNA进行测序所确定的
- [0330] vii. 还发现于来自相同个体的至少两个样品中, 例如来自相同病变或来自两个不

同病变的两个样品,或至少一个肿瘤或病变和至少一个来自肿瘤或病变的档案肿瘤材料样品。

[0331] 在另一个优选的实施方式中,结合MHC I和/或结合MHC II的新表位满足所有以下标准i-vii:

[0332] i.高RNA表达水平

[0333] ii.高等位基因频率

[0334] iii.高克隆性

[0335] iv.匹配已知的癌症相关基因

[0336] v.包含由至少两个不同的变体调用器鉴定的突变

[0337] vi.还发现于个体的血浆中,如例如通过对无细胞DNA进行测序所确定的

[0338] vii.还发现于来自相同个体的至少两个样品中,例如来自相同病变或来自两个不同病变的两个样品,或至少一个肿瘤或病变和至少一个来自肿瘤或病变的档案肿瘤材料样品。

[0339] 上面的标准i-vii没有以任何优选的顺序列出,并且可以以任何方式优先排序。RNA表达水平,等位基因频率,克隆性,与已知癌症相关基因的匹配,至少两个不同变体调用器的鉴定,个体血浆中的存在以及在来自相同个体的至少两个样品中的存在是如上文所定义的。

[0340] 因此,优选地,具有较高RNA表达水平的结合MHC I和/或结合MHC II的新表位排名高于具有较低RNA表达水平的新表位。

[0341] 优选地,将具有较高等位基因频率的结合MHC I的和/或结合MHC II的新表位排名高于具有较低等位基因频率的新表位。

[0342] 优选地,将具有较高克隆性的结合MHC I的和/或结合MHC II的新表位排名高于具有较低克隆性的新表位。

[0343] 优选地,与已知癌症相关基因匹配的结合MHC I的和/或结合MHC II的新表位排名高于不与已知癌症相关基因匹配的新表位。

[0344] 优选地,将包含由至少两个不同的变体调用器鉴定出的突变的结合MHC I的和/或结合MHC II的新表位排名高于包含仅由一个变体调用器鉴定出的突变的新表位。

[0345] 优选地,将在个体血浆中发现的结合MHC I的和/或结合MHC II的新表位排名高于未在个体血浆中发现的新表位。

[0346] 优选地,将在来自相同个体的两个或更多个样品中发现的结合MHC I的和/或结合MHC II的新表位排名高于在来自个体的仅一个样品中发现的新表位。所述两个或更多个样品可以是来自相同病变或不同病变的样品,或来自至少一个肿瘤或病变的样品和至少一种来自肿瘤或病变的档案肿瘤材料的样品。

[0347] 选择的新表位数目

[0348] 为个体选择的新表位优选用于癌症疫苗,例如癌症疫苗构建体,优选核苷酸疫苗构建体。该疫苗也称为个性化癌症疫苗。疫苗是免疫原性的。

[0349] 在根据本发明的方法中,选择具有临床实用性的A个新表位。A个新表位优选用于癌症疫苗或癌症疫苗构建体中。因此,A个新表位优选包含在相同疫苗构建体中。疫苗构建体优选是核苷酸构建体。核苷酸构建体可以是RNA构建体和/或DNA构建体。优选地,疫苗是

DNA构建体,也称为疫苗体DNA疫苗。

[0350] A是整数。例如,A为至少1,例如至少3,例如至少5,例如至少7,或例如至少10,例如至少20,例如至少30,例如40或更大。A可以是预定的数目。

[0351] 另外,本发明的发明人已经发现,疫苗构建体中新表位的数目从1个增加至3个新表位或从3个新表位增加至10个新表位导致免疫应答的惊人增加。另外,已经发现将疫苗构建体中新表位的数目从10个新表位增加到15或20个新表位导致免疫应答的进一步增加。在一些实施方式中,疫苗构建体中的新表位的数目在20和40之间,例如30。在一些实施方式中,疫苗构建体中的新表位的数目为40或更多。

[0352] 在一个实施方式中,A为3至100的整数,例如3至75,例如3至50,例如3至30,例如3至20,例如3至15或诸如例如3至10个新表位。

[0353] 在另一个实施方式中,A是5至50,例如5至30,诸如例如5至25,例如5至20,例如5至15,例如5至10的整数。

[0354] 在进一步的实施方式中,A为10至50的整数,例如10至40,例如10至30,例如10至20,例如10至25,例如10至20或诸如例如10至15。

[0355] 通过本发明方法选择的新表位可用于设计疫苗体疫苗,这将在下面进一步详细描述。疫苗体疫苗包含至少一个抗原单元,靶向单元和二聚化单元。新表位优选包含在抗原单元中。

[0356] 本发明的发明人已经表明,与包含仅3个新表位的疫苗体DNA疫苗相比,包含10个新表位的疫苗体DNA疫苗诱导更强和更广泛的总免疫应答。类似地,与包含仅10个新表位的疫苗相比,包含20个新表位的疫苗体DNA疫苗诱导更强和更广泛的总免疫应答。然而,待治疗的癌症可能仅与一个癌症特异性新表位相关,并且可能无法构建包含一个以上表位的疫苗体DNA疫苗。

[0357] 在优选的实施方式中,A是10至20的整数。

[0358] 在另一个实施方式中,A是15至50,例如15至30或例如15至20的整数。

[0359] 在具体实施方式中,A是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或40。

[0360] 在一个实施方式中,抗原单元包含每种癌症新表位的一个拷贝,因此当疫苗中包含10种新表位时,可以引起针对10种不同新表位的细胞介导的免疫应答。

[0361] 在本发明的方法中,选择了结合MHC I的新表位中排名最高的x个新表位,而选择了结合MHC II的新表位中排名最高的y个新表位,其中 $x+y=A$ 。A是所选新表位的总数。A是如上文所定义的。

[0362] x和y是整数。

[0363] 在优选的实施方式中, $x>y$ 。例如, $x\geq 2y$ 。在另一个实施方式中, $x\geq 2.5y$ 。在一个实施例中, $x\geq 3y$ 。

[0364] 在一个实施方式中, $0.5A<x<A$ ,其中A是如上所定义的。优选地,A是10至20的整数。

[0365] 在另一个实施方式中, $0.6A<x<A$ ,例如 $0.7A<x<A$ ,诸如例如 $0.8A<x<A$ ,其中A是如上所定义的。优选地,A是10至20的整数。

[0366] 具体方法

[0367] 在特定的实施方式中,该方法是用用于为个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0368] a.从所述个体获得一个或多个新表位,每个新表位均包含至少一个最小表位,其中与参考序列相比,每个新表位均包含至少一个免疫原性突变;

[0369] b.确定每个所述新表位内的至少一个最小表位(例如至少两个,三个或四个最小表位)的MHC I结合亲和力,并确定每个所述新表位的结合MHC I的最小表位数目;从而鉴定出x个结合MHC I的新表位;

[0370] c.对新表位进行如下排名:

[0371] i.将包含大量最小表位的新表位进行优先排序,并选择具有高得分的第一组新表位;

[0372] ii.任选地,将来自第一组新表位的新表位如下进行优先排序:

[0373] 1)如果突变处于锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分最高;

[0374] 2)如果突变处于非锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分低于1)的最小表位的得分;

[0375] 3)如果突变处于锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于2)的最小表位的得分;

[0376] 4)如果突变处于非锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于3)的最小表位的得分;

[0377] 5)如果突变处于锚定位置并且最小表位的结合差异小于3,则最小表位的得分低于4)的最小表位的得分;

[0378] 6)如果突变处于非锚定位置并且最小表位的结合差异小于3,则最小表位的得分低于5)的最小表位的得分;

[0379] 7)如果最小表位具有结合差异低于1的突变,则最小表位的得分最低,而与突变的位置无关;

[0380] 其中,MHC I结合差异由下式给出:(用于参考的%排名得分(MHC I))/(用于最小表位的%排名得分(MHC I));并选择第二组得分高的新表位;

[0381] iii.任选地,基于MHC I%排名得分,将来自第二组的新表位进行优先排序,并选择具有低MHC I%排名得分的第三组新表位;

[0382] iv.任选地,选择第四组新表位,其包括与已知被T细胞识别的表位高度相似的最小表位;

[0383] v.任选地,基于BLOSUM得分,将来自第三组或来自第四组的新表位进行优先排序,其中小于预定阈值的BLOSUM得分排名高于等于或大于所述阈值的BLOSUM得分,并且选择BLOSUM得分小于所述阈值的第五组新表位,其中所述阈值优选为1;

[0384] 其中第一组,第二组,第三组,第四组或第五组新表位包含所述A个表位。

[0385] 在特定的实施方式中,该方法是用用于为个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0386] a.从所述个体获得一个或多个新表位,每个新表位均包含至少一个最小表位,其中与参考序列相比,每个新表位均包含至少一个免疫原性突变;

[0387] b. 确定每个所述新表位内的至少一个最小表位(例如至少两个,三个或四个最小表位)的MHC I结合亲和力,并确定每个所述新表位的结合MHC I的最小表位数目;从而鉴定出x个结合MHC I的新表位;

[0388] c. 对新表位进行如下排名:

[0389] i. 将包含大量最小表位的新表位进行优先排序,并选择具有高得分的第一组新表位;

[0390] ii. 任选地,将来自第一组新表位的新表位如下进行优先排序:

[0391] 1) 如果突变处于锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分最高;

[0392] 2) 如果突变处于非锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分低于1)的最小表位的得分;

[0393] 3) 如果突变处于锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于2)的最小表位的得分;

[0394] 4) 如果突变处于非锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于3)的最小表位的得分;

[0395] 5) 如果突变处于锚定位置并且最小表位的结合差异小于3,则最小表位的得分低于4)的最小表位的得分;

[0396] 6) 如果突变处于非锚定位置并且最小表位的结合差异小于3,则最小表位的得分低于5)的最小表位的得分;

[0397] 7) 如果最小表位具有结合差异低于1的突变,则最小表位的得分最低,而与突变的位置无关;

[0398] 其中,MHC I结合差异由下式给出:(用于参考的%排名得分(MHC I))/(用于最小表位的%排名得分(MHC I));并选择第二组得分高的新表位;

[0399] iii. 任选地,基于MHC I%排名得分,将来自第二组的新表位进行优先排序,并选择具有低MHC I%排名得分的第三组新表位;

[0400] iv. 任选地,选择第四组新表位,其包括与已知被T细胞识别的表位高度相似的最小表位;

[0401] v. 任选地,基于BLOSUM得分,将来自第三组或来自第四组的新表位进行优先排序,其中小于预定阈值的BLOSUM得分排名高于等于或大于所述阈值的BLOSUM得分,并且选择BLOSUM得分小于所述阈值的第五组新表位,其中所述阈值优选为1;

[0402] vi. 任选地,基于在两个或更多个样品中发现的新表位,从第一组,第二组,第三组,第四组或第五组新表位中选择新表位,并选择在两个或更多个样品中发现的第六组新表位;

[0403] vii. 任选地,基于至少两个不同的变体调用器(variant caller)对突变的鉴定,从第一组,第二组,第三组,第四组,第五组或第六组新表位中选择新表位,并选择包含由至少两个不同的变体调用器所鉴定的突变的第七组新表位,

[0404] 其中第一组,第二组,第三组,第四组,第五组,第六组或第七组新表位包含所述A个表位。

[0405] 该方法还可以进一步包括根据本文所述的任何额外参数(例如克隆性, RNA表达水

平和等位基因频率)对新表位进行排名。

[0406] 在一些实施方式中,第二组是第一组的子组。在一些实施方式中,第三组是第二组和/或第一组的子组。在一些实施方式中,第四组是第三组和/或第二组和/或第一组的子组。在一些实施方式中,第五组是第四组和/或第三组和/或第二组和/或第一组的子组。在一些实施方式中,第六组是第五组和/或第四组和/或第三组和/或第二组和/或第一组的子组。在一些实施方式中,第七组是第六组和/或第五组和/或第四组和/或第三组和/或第二组和/或第一组的子组。子组的大小可以相同。

[0407] 在一个实施方式中,该方法包括以上步骤i.。在另一个实施方式中,该方法包括以上步骤i.和ii.。在另一个实施方式中,该方法包括以上步骤i.和iii.。在另一个实施方式中,该方法包括以上步骤i.和iv.。在另一个实施方式中,该方法包括以上步骤i.,ii.和iii.。在另一个实施方式中,该方法包括以上步骤i.,ii.和iv.。在另一个实施方式中,该方法包括以上步骤i.,iii.和iv.。在另一个实施方式中,该方法包括以上步骤i.,ii.,iii.和iv.。在一些实施方式中,该方法进一步包括以上步骤v.和/或步骤vi.和/或步骤vii.。

[0408] 在一个实施方式中,用于为患有或怀疑患有癌症的个体选择数目A个新表位的方法包括以下步骤:

[0409] a.从所述个体获得一个或多个新表位,每个新表位包含至少一个最小表位,其中与参考序列相比,每个新表位包含至少一个突变例如免疫原性突变,其中所述最小表位由等于或小于新表位的氨基酸数目的氨基酸组成,并包含所述至少一个突变;其中优选地获得新表位包括鉴定对肿瘤特异性的核酸序列中的突变的步骤;

[0410] b.对于每个所述新表位内的至少一个最小表位,诸如至少两个,三个或四个最小表位,确定MHC I和/或MHC II结合亲和力,任选地,其中结合亲和力是通过计算机(in silico)预测确定的;

[0411] c.选择包含至少一个预测与MHC I和/或MHC II结合的最小表位的新表位,从而获得结合MHC的新表位;

[0412] d.如本文所述,根据临床实用性的可能性对结合MHC的新表位进行排名;

[0413] e.在排名最高的结合MHC的新表位中选择A个新表位,

[0414] 从而选择可能具有临床实用性的A个新表位,其中步骤d进一步包括如下对新表位进行排名:

[0415] i)对于新表位中包含的每个最小表位,确定突变是在最小新表位的锚定位置还是非锚定位置;

[0416] ii)将新表位进行优先排序,

[0417] 任选地,其中步骤ii)中将新表位进行优先排序是通过向每个新表位分配它们所包含的最小表位的最高得分来进行的,并且其中将新表位进行优先排序是如下进行的:

[0418] 1)如果突变处于锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分最高;

[0419] 2)如果突变处于非锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分低于1)的最小表位的得分;

[0420] 3)如果突变处于锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于2)的最小表位的得分;

[0421] 4) 如果突变处于非锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于3)的最小表位的得分;

[0422] 5) 如果突变处于锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于4)的最小表位的得分;

[0423] 6) 如果突变处于非锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于5)的最小表位的得分;

[0424] 7) 如果最小表位具有结合差异低于1的突变,则最小表位的得分最低,而与突变的位置无关。

[0425] 在一个实施方式中,用于为患有或怀疑患有癌症的个体选择数目A个新表位的方法包括以下步骤:

[0426] a. 从所述个体获得一个或多个新表位,每个新表位包含至少一个最小表位,其中与参考序列相比,每个新表位包含至少一个突变例如免疫原性突变,其中所述最小表位由等于或小于新表位的氨基酸数目的氨基酸数目组成,并包含所述至少一个突变;其中优选地获得新表位包括鉴定对肿瘤特异性的核酸序列中的突变的步骤;

[0427] b. 对于每个所述新表位内的至少一个最小表位,诸如至少两个,三个或四个最小表位,确定MHC I和/或MHC II结合亲和力,任选地,其中结合亲和力是通过计算机预测确定的;

[0428] c. 选择包含至少一个预测与MHC I和/或MHC II结合的最小表位的新表位,从而获得结合MHC的新表位;

[0429] d. 根据临床实用性的可能性对结合MHC的新表位进行排名;

[0430] e. 在排名最高的结合MHC的新表位中选择A个新表位,

[0431] 从而选择可能具有临床实用性的A个新表位,

[0432] 其中步骤d包括根据新表位所包含的最小表位的数目对所述新表位进行排名,其中较高数目的最小表位给出较高的排名。

[0433] 在另一个实施方式中,用于为患有或怀疑患有癌症的个体选择数目A个新表位的方法包括以下步骤:

[0434] a. 从所述个体获得一个或多个新表位,每个新表位包含至少一个最小表位,其中与参考序列相比,每个新表位包含至少一个突变例如免疫原性突变,其中所述最小表位由等于或小于新表位的氨基酸数目的氨基酸数目组成,并包含所述至少一个突变;其中优选地获得新表位包括鉴定对肿瘤特异性的核酸序列中的突变的步骤;

[0435] b. 对于每个所述新表位内的至少一个最小表位,诸如至少两个,三个或四个最小表位,确定MHC I和/或MHC II结合亲和力,任选地,其中结合亲和力是通过计算机预测确定的;

[0436] c. 选择包含至少一个预测与MHC I和/或MHC II结合的最小表位的新表位,从而获得结合MHC的新表位;

[0437] d. 根据临床实用性的可能性对结合MHC的新表位进行排名;

[0438] e. 在排名最高的结合MHC的新表位中选择A个新表位,

[0439] 从而选择可能具有临床实用性的A个新表位,

[0440] 其中步骤d包括根据发现新表位的样品数目对新表位进行排名,其中在较高数目

的样品中发现的新表位排名高于在较低数目的样品中发现的新表位,优选地其中样品是来自不同病变的样品。

[0441] 本节中描述的具体方法的上述每个步骤都可以如本文其他地方所述。

[0442] 疫苗

[0443] 本发明的另一方面涉及制备包含新表位的癌症疫苗的方法,所述方法包括使用本文所定义的选择新表位的方法来选择所述新表位的步骤。

[0444] 在一些实施方式中,选择10-40个新表位。即,A是10至40的整数。在其他实施方式中,选择10-30或10-20个新表位。

[0445] 在一个实施方式中,所述癌症疫苗包含核苷酸构建体,所述核苷酸构建体包含:

[0446] -靶向单元

[0447] -二聚化单元

[0448] -第一接头

[0449] 抗原单元,其中所述抗原单元包含A-1个抗原亚单元,每个亚单元包含编码至少一个所述新表位的序列和第二接头,并且所述抗原单元进一步包含编码所述新表位之一的最终序列,其中A是1至100的整数,例如3至50。

[0450] 其中所述核苷酸构建体以免疫学有效量应用于抗癌疫苗。

[0451] 整数A如本文及以上所述。上述DNA疫苗也称为疫苗体DNA疫苗。

[0452] 本发明涉及通过上述方法制备的疫苗。

[0453] 因此,本文所述方法提供的新抗原疫苗可包含编码包含三个单元的多肽的多核苷酸,所述三个单元即靶向单元、二聚化单元和抗原单元。由于二聚化单元,该多肽形成称为疫苗体的二聚体蛋白。

[0454] 编码这三个单元的基因被基因工程改造以便表达为一个基因。当在体内表达时,多肽/二聚体蛋白靶向抗原呈递细胞(APC),导致了与相同的非靶向抗原相比,疫苗效价增强。

[0455] 抗原单元

[0456] 抗原单元包含多个肿瘤新表位,其中每个新表位对应于在肿瘤新抗原中所鉴定的突变。如上所述,所述突变可以是一个或多个突变。

[0457] 在抗原单元中,除了最后一个肿瘤新表位以外的所有肿瘤新表位都排列在抗原亚单元中,其中每个亚单元由肿瘤新表位序列和第二接头组成,而最后一个亚单元仅包含新表位,即没有这种第二接头。由于肿瘤新表位序列被所述第二接头分隔,因此将每个新表位以最佳方式呈递给免疫系统,从而如下所述确保了疫苗的效率。

[0458] 癌症新表位序列优选具有适合于由上述MHC分子呈递的长度。优选的新表位长度如上所述。

[0459] 如果疫苗是针对所述基因的表达产物,则为了避免肿瘤通过关闭突变基因的表达而逃避免疫系统,优选在抗原单元中包括多个不同的新表位。通常,肿瘤必须关闭的基因越多,肿瘤能够关闭所有这些基因并且仍然能够增殖甚至存活的可能性就越小。此外,肿瘤可能是异质的,因为并非每个新抗原都是由所有肿瘤细胞表达的。因此,根据本发明,该方法是在疫苗中包括许多新表位,以便有效地攻击肿瘤。优选地,多个新表位靶向多个基因的表达。同样,为了确保所有新表位都有效地加载到相同的抗原呈递细胞上,它们被安排为一条

氨基酸链而不是离散的肽。

[0460] 疫苗构建体中选择并包括的新表位的数目如上所述。

[0461] 在一个实施方式中,抗原单元包含每种癌症新表位的一个拷贝,从而可以引起针对疫苗中包括的尽可能多的不同新表位的应答。例如,当疫苗中包括10个新表位时,可以引起针对10个不同新表位的细胞介导的免疫应答,或者当疫苗中包括20个新表位时,可以引起针对20个不同新表位的细胞介导的免疫应答。在一些实施方式中,疫苗可包含多于20个新表位,从而引起针对尽可能多的新表位的应答。

[0462] 然而,如果仅鉴定出一些相关的抗原突变,则抗原单元可包含至少一种新表位的至少两个拷贝,以便增强对这些新表位的免疫应答。同样出于制造和调节的原因,保持质粒且即抗原单元的长度恒定或长度相似可能是有利的,并且因此在抗原单元中包括相同新表位的一个以上拷贝可能是有利的。

[0463] 如上所述,保持抗原单元的长度恒定可能是有利的,并且因此在一个实施方式中优选的是所有癌症新表位序列具有相同的长度。但是,如果新表位中的一个或多个是由引起移码的突变或终止密码子突变产生的,则新表位可能具有相当长的长度,例如由蛋白质的至少突变部分、突变蛋白的最具抗原性的部分或也许整个突变蛋白组成,其中至少一个新表位的长度实质长于由非同义点突变产生的新表位。

[0464] 抗原单元的长度主要由新表位的长度和在抗原单元中排列的新表位的数目决定,并且为约21至1500,优选约30个氨基酸至约1000个氨基酸,更优选约50至约500个氨基酸,例如约100至约400个氨基酸,约100至约300个氨基酸。

[0465] 插入疫苗中的癌症新表位序列可包含在两侧由氨基酸序列侧接的突变。优选地,该突变基本上位于癌症新表位序列的中间,以确保免疫原性突变在加工后由抗原呈递细胞呈递。侧接突变的氨基酸序列优选是新抗原中侧接突变的氨基酸序列,由此癌症新表位序列是癌症新抗原氨基酸序列的真正亚序列。

[0466] 第二接头被设计为非免疫原性的,并且优选也是柔性接头,由此尽管抗原单元中存在大量抗原亚单元,肿瘤新表位仍以最佳方式呈递给T细胞。优选地,第二接头的长度为4至20个氨基酸以确保柔性。在另一个优选的实施方式中,第二接头的长度是8至20个氨基酸,例如8至15个氨基酸,例如8至12个氨基酸,或诸如例如10至15个氨基酸。在特定的实施方式中,第二接头的长度是10个氨基酸。

[0467] 在具体的实施方式中,本发明的疫苗包含10个新表位,其中第二接头的长度为8至20个氨基酸,例如8至15个氨基酸,例如8至12个氨基酸或诸如例如10至15个氨基酸。在特定的实施方式中,本发明的疫苗包含10个新表位,并且其中第二接头具有10个氨基酸的长度。

[0468] 第二接头优选是丝氨酸-甘氨酸接头,例如柔性GGGS接头,例如GGSS,GGSG,GGGS或其多个变体,例如GGGSGGGGS或(GGGGS)<sub>m</sub>,(GGSS)<sub>m</sub>,(GGSG)<sub>m</sub>,其中m是1至5、1至4或1至3的整数。在优选的实施方式中,m为2。

[0469] 靶向单元

[0470] 由于靶向单元,多肽/二聚体蛋白导致树突细胞(DC)、嗜中性粒细胞和其他免疫细胞的吸引。因此,包含靶向模块的多肽/二聚体蛋白不仅将抗原靶向特异性细胞,而且还通过将特异性免疫细胞募集到疫苗的施用部位来促进应答放大作用(佐剂作用)。由于疫苗本身具有佐剂作用,因此这种独特的机制在患者无需任何另外的佐剂就可以接受疫苗的临床

环境中非常重要。

[0471] 如本文所用,术语“靶向单元”是指这样的单元,该单元将多肽/蛋白质与其抗原一起递送至抗原呈递细胞用于向CD4<sup>+</sup>T细胞的MHC II类限制性呈递或用于提供通过MHC I类限制向CD8<sup>+</sup>T细胞的交叉呈递。

[0472] 靶向单元通过二聚化单元连接至抗原单元,其中后者在多肽/二聚体蛋白的COOH-末端或NH<sub>2</sub>-末端中。优选的是,抗原单元在多肽/二聚体蛋白的COOH-末端。

[0473] 靶向单元被设计为将本发明的多肽/二聚体蛋白靶向至在相关抗原呈递细胞上表达的表面分子,例如专门在树突细胞(DC)的子集上表达的分子。

[0474] APC上的这种靶表面分子的实例是人白细胞抗原(HLA),分化簇14(CD14),分化簇40(CD40),趋化因子受体和Toll样受体(TLR)。HLA是人主要组织相容性复合物(MHC)。Toll样受体可以例如包括TLR-2,TLR-4和/或TLR-5。

[0475] 本发明的多肽/二聚体蛋白可借助于包含以下各物的靶向单元而靶向所述表面分子:例如对CD14、CD40或Toll样受体有特异性的抗体结合区;配体,例如可溶性CD40配体;天然配体样趋化因子,例如RANTES或MIP-1 $\alpha$ ;或细菌抗原,像例如鞭毛蛋白。

[0476] 在一个实施方式中,靶向单元对MHC II类蛋白具有亲和力。因此,在一个实施方式中,编码靶向单元的核苷酸序列编码对MHC II类蛋白具有特异性的抗体可变结构域(VL和VH),所述MHC II类蛋白选自自由以下组成的组:抗-HLA-DP,抗-HLA-DR和抗-HLA-II。

[0477] 在另一个实施方式中,靶向单元对选自CD40、TLR-2、TLR-4和TLR-5组成的组的表面分子具有亲和力。因此,在一个实施方式中,编码靶向单元的核苷酸序列编码针对抗-CD40、抗-TLR-2、抗-TLR-4和抗-TLR-5有特异性的抗体可变结构域(VL和VH)。在一个实施方式中,编码靶向单元的核苷酸序列编码鞭毛蛋白。鞭毛蛋白对TLR-5具有亲和力。

[0478] 优选地,靶向单元对选自CCR1、CCR3和CCR5的趋化因子受体具有亲和力。更优选地,编码靶向单元的核苷酸序列编码趋化因子hMIP-1 $\alpha$ (LD78 $\beta$ ),该趋化因子结合至在APC的细胞表面上表达的同源受体CCR1、CCR3和CCR5。hMIP-1 $\alpha$ (人MIP-1 $\alpha$ )也称为趋化因子(C-C基序)配体3(CCL3),其在人中由CCL3基因编码。CCL3,也称为巨噬细胞炎性蛋白-1 $\alpha$ (MIP-1 $\alpha$ ),是属于CC趋化因子家族的细胞因子,在多形核白细胞的募集和激活中参与急性炎症状态。尽管小鼠CCL3是编码69个氨基酸的成熟趋化因子的单拷贝基因,但人同源物已被复制并突变以产生两个非等位基因变体LD78 $\alpha$ (CCL3)和LD78 $\beta$ (CCL3-L1),二者均显示与小鼠CCL3的74%同源性。

[0479] 本发明的多肽/二聚体蛋白质与其同源受体的结合导致APC的内化以及蛋白质降解成包括最小表位(因此也就是突变)的小肽,这些小肽被加载到MHC分子上并呈递给CD4<sup>+</sup>和CD8<sup>+</sup>T细胞以便诱导肿瘤特异性免疫应答。一旦受到刺激,并在激活的CD4<sup>+</sup>T细胞的帮助下,CD8<sup>+</sup>T细胞将靶向并杀死表达相同新抗原的肿瘤细胞。

[0480] 在本发明的一个实施方式中,靶向单元包含与SEQ ID NO:1的氨基酸序列5-70具有至少80%序列同一性的氨基酸序列。在优选的实施方式中,靶向单元包含与SEQ ID NO:1的氨基酸序列5-70具有至少85%序列同一性(例如至少86%,例如至少87%,例如至少88%,例如至少89%,例如至少90%,例如至少91%,例如至少92%,例如至少93%,例如至少94%,例如至少95%,例如至少96%,例如至少97%,例如至少98%,例如至少99%序列同一性)的氨基酸序列。

[0481] 在更优选的实施方式中,靶向单元由与SEQ ID NO:1的氨基酸序列5-70具有至少80%序列同一性(例如与SEQ ID NO:1的氨基酸序列5-70具有至少85%,例如至少86%,例如至少87%,例如至少88%,例如至少89%,例如至少90%,例如至少91%,例如至少92%,例如至少93%,例如至少94%,例如至少95%,例如至少96%,例如至少97%,例如至少98%,例如至少99%,例如至少100%序列同一性)的氨基酸序列组成。

[0482] 二聚化单元

[0483] 如本文所用,术语“二聚化单元”是指抗原单元与靶向单元之间的氨基酸序列。因此,二聚化单元用于连接抗原单元和靶向单元,并促进两个单体多肽二聚化为二聚体蛋白。此外,二聚化单元还在多肽/二聚体蛋白中提供柔性,以允许靶向单元与抗原呈递细胞(APC)上的表面分子的最佳结合,即使它们位于可变的距离。二聚化单元可以是满足这些要求的任何单元。

[0484] 因此,在一个实施方式中,二聚化单元可包括铰链区和任选地促进二聚化的另一结构域,并且铰链区和另一结构域可通过第三接头连接。

[0485] 术语“铰链区”是指二聚体蛋白的促进二聚化的肽序列。铰链区用作单元之间的柔性间隔物,从而允许两个靶向单元同时结合APC上的两个靶分子,即使它们以可变的距离表达。铰链区可以是Ig来源的,例如来源于IgG3。铰链区可通过共价键(例如二硫桥)的形成促进二聚化。因此,在一个实施方式中,铰链区具有形成一个或多个共价键的能力。共价键可以例如是二硫桥。

[0486] 在一个实施方式中,促进二聚化的另一结构域是免疫球蛋白结构域,例如羧基末端C结构域,或与C结构域或其变体基本相同的序列。优选地,促进二聚化的另一个结构域是来源于IgG的羧基末端C结构域。

[0487] 免疫球蛋白结构域通过非共价相互作用(例如疏水性相互作用)促进二聚化。例如,免疫球蛋白结构域具有经由非共价相互作用形成二聚体的能力。优选地,非共价相互作用是疏水相互作用。

[0488] 优选地,二聚化单元不包含能够结合F细胞受体的CH2结构域。在一些实施方式中,二聚化单元完全没有CH2结构域。在其他实施方式中,CH2结构域被突变,从而使其失去了与F细胞受体结合的能力。

[0489] 在优选的实施方式中,二聚化单元由通过第三接头连接至人IgG3的CH3结构域的铰链外显子h1和h4组成。

[0490] CD8+T细胞应答

[0491] 通过本发明的排名方法获得的本文所述的疫苗可以诱导针对给定新表位的免疫应答从CD4+T细胞应答转变为CD8+T细胞应答。因此,一方面,提供了用于为个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0492] a. 从所述个体获得一个或多个新表位,每个新表位均包含至少一个最小表位,其中与参考序列相比,每个新表位均包含至少一个突变,诸如免疫原性突变;

[0493] b. 对于每个所述新表位内的至少一个最小表位,诸如至少两个,三个或四个最小表位,确定MHC I和/或MHC II结合亲和力;

[0494] c. 选择预测与MHC I和/或MHC II结合的新表位,从而获得结合MHC的新表位;

[0495] d. 根据临床实用性的可能性对结合MHC的新表位进行排名;

[0496] e. 在最高排名的结合MHC的新表位中选择A个新表位,其中A是整数并且A是至少3个,诸如至少4个,诸如至少5个,

[0497] 从而选择当作为本文所述的疫苗施用能够诱导CD8+T细胞应答的A个新表位。

[0498] 在一些实施方式中,包含通过本文所述方法选择的新表位的疫苗因此能够诱导CD8+T细胞应答。在特定实施方式中,与其他疫苗形式相比,这种疫苗能够诱导显性CD8+T细胞应答。例如,当作为肽疫苗和/或作为RNA疫苗施用诱导CD4+T细胞应答的新表位在作为本文所述的疫苗体DNA疫苗施用时可诱导CD8+T细胞应答。特别地,以前可能没有发现通过本发明方法选择的新表位能够诱导CD8+T细胞应答。如实施例13和图6所示,发明人发现对于几种新表位都是这种情况。

[0499] 在一些实施方式中,A是如本文另外描述的整数。

## 实施例

[0500] 实施例1 VB10.NEO患者特异性宿主细胞系的来源、历史和生成

[0501] 新表位抗原模块的设计是基于对于每位患者鉴定的独特的肿瘤特异性序列。为了预测最具免疫原性的新表位,建立了优化的工作流程,如下所示,如图7所示。

[0502] 1) 新表位选择的第一步是鉴定所有肿瘤特异性突变。将下述新表位排除或降低其优先排序,所述新表位不在肿瘤中表达,或可能例如包含在蛋白质组中其他地方具有相同序列匹配的9个氨基酸的肽(包括突变位点)。此外,也将基因中的下述新表位排除或降低其优先排序:所述新表位在特定器官/组织中显示出与所有其他组织相比至少5倍高的RNA表达水平。

[0503] 2) 下一步是构建新表位(27个氨基酸长,突变位于中间),并基于对HLA分子的肽结合亲和力及其残基性质的最佳组合对这些新表位进行排名。

[0504] 3) 在一些实施方式中,新表位集合的完成包括基于精心选择的评估标准对新表位进行评估,所述评估标准由据报道对于定义免疫原性新表位潜在重要的新表位和/或其源基因的性质组成。该评估由靶标选择委员会(target selection board)执行,该委员会由临床医生(CMO)、免疫学家(CSO)和生物信息学家组成。

[0505] 实施例2患者组织的收集和处理

[0506] 在筛查期间从合格患者中收集实体肿瘤组织以及血液样品。

[0507] 收集至少一项核心活检用于外显子组测序和RNA测序。对于外显子组和转录组测序,冷冻保存是保持标本完整性的理想之选,因此该材料优选用于测序目的。如果在筛选时无法获得新鲜的肿瘤材料,则筛选之前保存在FFPE中的肿瘤材料对于测序是可以接受的。

[0508] 收集血液样品。对于冷冻保存的肿瘤活检样品,将样品浸入液氮中进行快速冷冻,并保存在液氮中。对于FFPE肿瘤材料,将组织样品固定在4%至10%的中性缓冲的福尔马林中。然后将样品脱水,然后包埋并在室温下保存。

[0509] 实施例3患者外显子组和RNA测序以及HLA分型

[0510] 外显子组和RNA测序数据可从血液、肿瘤以及患者的HLA类型获得。使用至少2x100 bp的读数长度对肿瘤和血液样品进行测序。原始数据输出为至少24Gb。

[0511] RNA测序是从具有至少2x100 bp读数长度的肿瘤样品中进行的。总数据输出为至少1亿读数。

[0512] 对于HLA分型,将从血液样品中分离的DNA在Illumina平台上以至少2x150 bp读数长度进行测序,覆盖度超过100x。

[0513] 实施例4-生物信息学处理和肿瘤中突变的鉴定

[0514] 外显子组和RNA表达数据作为FASTQ文件提供,可从测序提供者的内部安全服务器下载。QC传递的原始外显子组序列(FASTQ文件)的分析遵循了最佳实践工作流程,所述最佳实践工作流程是由最新技术和针对fastq处理和针对肿瘤和正常样品对的变体调用的优化算法定义的(例如Miller等人,2015;Van der Auwera等人,2013)。

[0515] RNA测序的原始FASTQ文件使用RNA测序分析中最先进的方法进行映射(例如,Dobin等人,2013;Dutton G.2016)。

[0516] 使用Illumina测序平台从EDTA血液样品中鉴定患者的HLA等位基因。

[0517] 研究了在肿瘤组织中检测到RNA表达的蛋白质编码基因中发现的所有突变(在vcf文件中鉴定的变体)(TPM>0,如Gubin等人,2015中所定义的)作为新表位的潜在实用性。

[0518] 提取在突变的每一侧跨越13个氨基酸的肽序列,形成总长度为27个氨基酸的新表位序列。对于与蛋白质C端或N端的距离比13个氨基酸更近的非同义突变,突变侧翼序列的一侧短于13个氨基酸,因此,新表位序列的总长度将短于27个氨基酸。

[0519] 实施例5-使自身免疫风险最小化的排除标准

[0520] 为了使交叉反应的风险最小化,优选地在此阶段将以下所有新表位排除在外或降低其优先排序:所述新表位具有与正常人蛋白质组中任何肽序列匹配的9个氨基酸的核心肽(包括突变)。此外,也将基因中的下述新表位排除或降低其优先排序:所述新表位在特定组织或器官中显示出与所有其他组织或器官相比至少5倍高的RNA表达水平(如Human Proteome Atlas,Uhlén等人,2015所定义的)。

[0521] 实施例6-在小鼠肿瘤模型中新表位优先排序的计算模型的开发

[0522] 为了鉴定最具免疫原性的新表位,利用来自两个体内小鼠肿瘤模型的数据和结果来开发用于预测新表位优先排序的计算模型。

[0523] 使用来自小鼠体内临床前实验的多个数据集,收集了与预测免疫原性新表位相关的特征,并评估了它们的预测潜力。使用在CT26结肠癌模型中鉴定出的新表位及其在VB10.NEO疫苗接种的BALB/c小鼠(H-2d)中观察到的免疫原性来制定基于预测的免疫原性对新表位进行优先排序的策略。使用针对在B16黑色素瘤模型和LL2肺癌模型中鉴定的新表位收集的免疫原性数据,以及在VB10.NEO疫苗接种的C57B1/6小鼠(H-2b)中观察到的免疫原性,验证了该策略的预测能力。如果通过内部进行的IFN- $\gamma$  ELISpot测定所分析的IFN- $\gamma$ 数目>阴性对照+2xSD以及斑点数目>25,则将新表位归类为免疫原性。以下各项在免疫原性和非免疫原性新表位之间表现出不同的模式:使用预测工具NetMHCpan(Nielsen等人,2016)和NetMHCIIpan(Andreatta等人,2015)的对MHC分子I和II类的结合亲和力(%排名),结合性最小表位的总数,突变和野生型表位之间的结合亲和力的差异(结合差异),和BLOcks置换矩阵(BLOSUM)得分(Henikoff等人,1992)。

[0524] 实施例7-结合亲和力(%排名),结合性表位的总数,结合差异和BLOSUM得分。

[0525] 在小鼠中,只有对MHC I类或II类具有亲和力的肽(在人类HLA I类或II类中)才可以提供合格的T细胞靶标。选择疫苗靶标的一个策略是分别使用NetMHCpan和NetMHCIIpan基于候选新表位对I类和II类MHC分子的预测亲和力来选择候选表位。对于所有27个氨基酸

的新表位,这些服务器提供了所选MHC分子的结合亲和力得分。结合亲和力预测服务器提供了多个MHC亲和力得分值,其中%排名是高度推荐的(Nielsen等人,2016和Andreatta.等人,2015)。低的%排名得分表明强的结合亲和力。

[0526] 图2说明了在VB10.NEO疫苗接种的小鼠中测试的免疫原性和非免疫原性新表位对于MHC I类和MHC II类的%排名分布。

[0527] 在27个氨基酸长的新表位中,可能有一个以上的最小表位(对于I类MHC为8-14个氨基酸和对于II类MHC为9-15个氨基酸)预测与相关MHC分子结合。含有最高数目的预测最小表位的新表位增加了在患者的MHC分子上加工和呈递一个或多个免疫原性肽的机会,从而引发了更有效的肿瘤特异性免疫应答。图8显示了B16和CT26数据中免疫原性和非免疫原性新表位之间最小表位总数的比较。

[0528] 如果新表位对MHC分子的结合亲和力高于相应的野生型序列(高结合差异),则该野生型很可能在健康组织中极少呈递或不呈递,并且因此,新表位特异性T细胞应具有识别健康细胞的低风险以及识别表达新表位的癌细胞的高潜能。为测量参考序列和突变新表位序列之间的结合差异,可以使用它们的%排名值之间的比率(参考/突变),并按残基类型分组:锚定(P2或P8/P9)和非锚定(所有其他残基位置)。图9说明如果表位在锚定位置具有突变并且具有高结合差异,则其具有更高的免疫原性机会。

[0529] BLOSUM得分描述了成对氨基酸置换发生的可能性。在我们的案例中,健康组织中表达的蛋白质中发现的氨基酸已被突变的氨基酸替代。得分越低,在相关蛋白质的比对中发生这种置换的可能性就越小。具有高结合亲和力的CT26黑色素瘤模型数据集中新表位的BLOSUM得分通常对于免疫原性新表位而言较低(图10)。

[0530] 为了预测最具免疫原性的新表位,包括了所有五个因素的组合:MHCI和MHCII的%排名值,新表位内最小表位的数目,锚定位置的结合差异和BLOSUM得分。新表位的优先排序策略称为NeoSELECT,其权衡了这五个因素,并生成了新表位的排序列表。对来自LL2、CT26和B16肿瘤模型的新表位测试NeoSELECT排名,其观察到的免疫原性如图3所示。

[0531] 对于(A)CT26,(B)B16和(C)LL2模型,图4中显示了通过NeoSELECT策略预测的前20个新表位针对参考肽序列的观察到的交叉反应性以及突变肽序列的应答。对于CT26模型,对20个新表位中的12个获得了交叉反应性测量;对于B26模型,对20个新表位中的17个获得了交叉反应性测量。对于LL2模型,对所有20个新表位获得了交叉反应性测量。因此,使用NeoSELECT策略,诱导交叉反应性T细胞的风险受到限制。

[0532] 实施例8-VB N-01临床试验中采用的新表位预测方法

[0533] 如果为患者发现了多于20个潜在新表位,则根据上述用于新表位预测的NeoSELECT优先排序策略对这些新表位进行排名。对于患者数据,在从NetMHCpan和NetMHCIIpan获取参考和WT新表位的%排名值的过程中使用了患者特异性HLA等位基因。计算结合差异并确定锚定位置。使用所有HLA等位基因的预测最小表位的完整列表来计算对于MHC I类和II类的最小表位的数目。从BLOSUM得分矩阵中提取每个突变的BLOSUM得分。这些因素随后用于NeoSELECT策略中,以获得预测的免疫原性新表位的排名列表。以%排名、结合差异、最小表位数以及BLOSUM得分测量的结合亲和力是标准值,不受其中鉴定新表位的生物体的影响。因此,基于小鼠数据开发的预测模型不需要另外的定制化即可用于人类的新表位预测。

[0534] 实施例9-最终新表位的质量控制和选择

[0535] 最后,收集通过这些排除标准的所有新表位的另外信息,以作为选择标准。这些标准包括:

[0536] • 克隆性(即,即使在患者的一个以上活检样品中发现的新表位(例如在档案肿瘤材料中,在相同病变的多角度活检中或在无细胞的DNA中发现新表位) 优先排序)

[0537] • 高RNA表达水平(以TPM为单位)

[0538] • 新表位对于MHC I类和II类的%排名低(高结合亲和力)

[0539] • 高结合差异(新表位相对于参考的%排名)

[0540] • 低BLOSUM得分:与野生型相比,新表位中置换氨基酸的残基特征

[0541] • 由VCF文件估算的突变位置处的高等位基因频率(AF)或变体等位基因频率(VAF)

[0542] • 在癌症的体细胞突变目录(Catalogue Of Somatic Mutations In Cancer)(COSMIC)数据库中,源基因是已知的癌症相关基因

[0543] • 新表位与已知被T细胞识别的表位(例如,如从免疫表位数据库和分析资源(IEDB数据库)中所获得的)具有高度相似性

[0544] • 在27个氨基酸的新表位中的大量最小表位(8-15个氨基酸)

[0545] • 突变在HLA-TCR复合物中的位置,如Fritsch等人,2014年中所述的

[0546] • 使用NetMHCstabpan (Rasmussen等人,2016)或类似方法估计的与MHC I类/MHC II类分子的高结合稳定性得分

[0547] • 通过NetChop(Nielsen等人,2005)或类似方法预测的高蛋白酶体切割得分。

[0548] 这些标准,除了根据NeoSELECT策略预测的优先排序之外,还构成了选择最终10至20个新表位集合的基础,并由靶标选择委员会进行评估。记录决策标准,并分析患者对疫苗中每种新表位的免疫应答和患者的临床反应。

[0549] 推荐的新表位集合是通过柔性的非免疫原性的富含甘氨酸/丝氨酸的接头进行组合和分离的(参见实施例11)。新表位的顺序取决于每个新表位和接头之间的连接序列,其中避免了在蛋白质组中其他地方具有相同匹配的长度为9个氨基酸的所有连接序列。

[0550] 实施例10-数据存储和分析

[0551] 原始数据的存储以及外显子组和RNA序列分析的结果均在安全、受控的计算机集群环境平台中执行,该平台满足处理和存储患者敏感数据的要求。所有患者在注册时都会获得唯一的数字,以保护患者的身份。

[0552] 实施例11-新表位抗原模块的合成

[0553] 由瓦西博迪公司基于与10个氨基酸长的富含甘氨酸/丝氨酸的接头连接的10-20个27个氨基酸长的新表位,设计了每个患者特异性的新表位抗原模块。新表位抗原模块是由DNA合成提供者从头合成的,并将其克隆到质粒中以生成VB10.NEO。

[0554] 实施例12-与其他疫苗形式相比,疫苗体可诱导强烈的显性CD8+T细胞应答

[0555] Kreiter等人2015年和Castle等人2012年研究了10种不同的新表位(pep1-10),它们均预测会结合MHC I类(CD8+T细胞应答),以研究它们是否可以诱导小鼠B16-F10黑色素瘤肿瘤模型中的CD8+T细胞应答。当以肽、RNA或本文所述的疫苗体(“VB10.NEO”)施用,由10个新表位中的6个诱导的应答显示在图6中。

[0556] 当将新表位作为合成肽施用,六分之一诱导CD8+T细胞应答;当将新表位作为 mRNA 施用,六分之二诱导CD8+T细胞应答。

[0557] 但是,当将相同的6个新表位作为靶向人MIP1 $\alpha$ 的疫苗体递送时,所有新表位均诱导CD8+T细胞应答,这清楚地表明了疫苗形式对于诱导针对新表位的强CD8+杀伤T细胞应答的重要性。

[0558] 实施例13-所选新表位的免疫原性

[0559] 使用所述方法选择新表位用于4名患有肾细胞癌或头颈部鳞状细胞癌的患者。构建个性化疫苗并将其施用于患者。在施用第一剂个性化疫苗后3至9个月收获的外周血单核细胞中测量了免疫原性。在10天预刺激ELISPOT测定中测量T细胞应答(减去背景)。疫苗诱导的应答:疫苗接种后>30SFU增加的应答。免疫原性新表位:在至少一个时间点>30SFU。

患者 ID	指征	#免疫原性新表位	%免疫原性表位	#疫苗接种后免疫应答增加的新表位	%疫苗接种后免疫应答增加的新表位
[0560] 01-001	RCC	20	100 %	17	85 %
01-002	SCCHN	20	100 %	19	95 %
01-004	SCCHN	19	95 %	12	60 %
01-005	RCC	17	85 %	14	70 %

[0561] RCC:肾细胞癌;SCCHN:头颈部鳞状细胞癌。

[0562] 平均而言,在这4名患者中,通过该方法选择的新表位中有95%具有免疫原性,即能够在其肿瘤中带有相应突变序列的患者中激活T细胞应答。当将所选的新表位掺入个性化疫苗中并施用于患者时,与疫苗接种前相比,针对平均78%的新表位诱导了增加的T细胞应答。

[0563] 该实施例表明,选择新表位的方法导致选择当施用于患者时能够增加T细胞应答的新表位。

[0564] 序列

[0565] SEQ ID NO:1

[0566] 包括信号肽和成熟肽的C-C基序趋化因子3-样1前体(LD78- $\beta$ ),aa24-93:

[0567] **MQVSTAALAVLLCTMALCNQVLSAPLAADTPTACCFSYTSRQIPQNFADYFETSSQC  
SKPSVIFLTRGRQVCADPSEEWVQKYVSDLELSA**

[0568] 参考文献

[0569] Andreatta, M. 等人 (2015) "Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification." Immunogenetics. 2015. 67(11-12): 641-50.

[0570] Castle, J.C. 等人 (2012) "Exploiting the Mutanome for Tumor Vaccination." Cancer Res; 72(5): 1081-91.

[0571] Dobin, A. 等人 (2013) "STAR: ultrafast universal RNA-seq aligner." Bioinformatics. 1; 29(1): 15-21.

[0572] Dutton, G. (2016) "From DNA to Diagnosis without Delay. Purpose-Built for Genomics, Dragen Processor Could Form Core of Clinic-Ready Data Systems." Genet

Eng Biotechnol. 36(5):8-9.

[0573] Fritsch, E.F. 等人 (2014) "HLA-binding properties of tumor neoepitopes in humans." *Cancer Immunol Res.*; 2(6):522-9.

[0574] Gubin, M.M. 等人 (2015) "Tumor neoantigens: building a framework for personalized cancer immunotherapy." *J. Clin. Invest.* 125(9):3413-21.

[0575] Henikoff, S. 等人 (1992) "Amino acid substitution matrices from protein blocks." *Proc Natl Acad Sci U S A.*; 89(22):10915-10919.

[0576] Jurtz, V. 等人 (2017) "NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data." Vanessa Jurtz, Sinu Paul, Massimo Andreatta, Paolo Marcatili, Bjoern Peters and Morten Nielsen. *The Journal of Immunology* (2017)

[0577] Kreiter S 等人. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature*. 2015 Apr 30; 520(7549):692-6

[0578] Kreiter, S. 等人 (2015) "Erratum: Mutant MHC Class II Epitopes Drive Therapeutic Immune Responses To Cancer." *Nature* 523.7560:370-370.

[0579] Miller, N.A. 等人 (2015) "26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases". *Genome Med.* 7:100.

[0580] Nielsen, M. 等人 (2005) "The role of the proteasome in generating cytotoxic T cell epitopes: Insights obtained from improved predictions of proteasomal cleavage." *Immunogenetics.* 57(1-2):33-41.

[0581] Nielsen, M. 等人 (2016) "NetMHCpan-3.0: improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length data sets." *Genome Medicine*: 8:33.

[0582] Rasmussen, M. (2016) "Pan-Specific Prediction of Peptide-MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity." *J Immunol.*; 197(4): 1517-24.

[0583] Uhlén, M. 等人 (2015) "Proteomics. Tissue-based map of the human proteome." *Science*; 347(6220)

[0584] Van der Auwera G.A. 等人 (2013) "From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline." *Curr Protoc Bioinformatics*. 43.

[0585] 项目

[0586] 1. 用于为个体选择数目A个新表位的方法, 所述方法包括以下步骤:

[0587] a. 从所述个体获得一个或多个新表位, 每个新表位均包含至少一个最小表位, 其中与参考序列相比, 每个新表位均包含至少一个突变, 诸如免疫原性突变;

[0588] b. 对于每个所述新表位内的至少一个最小表位, 诸如至少两个, 三个或四个最小表位, 确定MHC I和/或MHC II结合亲和力;

[0589] c. 选择包含至少一个预测与MHC I和/或MHC II结合的最小表位的新表位, 从而获得结合MHC的新表位;

- [0590] d. 根据临床实用性的可能性对结合MHC的新表位进行排名;
- [0591] e. 在排名最高的结合MHC的新表位中选择A个新表位,
- [0592] 从而选择可能具有临床实用性的A个新表位。
- [0593] 2. 根据项目1所述的方法,其中A个新表位中的一些或全部新表位至少结合MHC I。
- [0594] 3. 根据前述项目中任一项所述的方法,其中所述至少一个最小表位是1至50个最小表位,例如1至40个最小表位,例如1至30个最小表位,例如1至20个最小表位。
- [0595] 4. 根据前述项目中任一项所述的方法,其中每个最小表位包含突变。
- [0596] 5. 根据前述项目中任一项所述的方法,其中步骤b是通过以下进行的:确定每个所述最小表位的MHC I和/或MHC II结合亲和力,从而鉴定x个结合MHC I的最小表位和/或y个结合MHC II的最小表位。
- [0597] 6. 根据前述项目中任一项所述的方法,其中每个所述新表位的MHC I和MHC II结合亲和力计算为在每个所述新表位内包含的最小表位的最高MHC I和MHC II结合亲和力。
- [0598] 7. 根据前述项目中任一项所述的方法,其中步骤c包括以下步骤或由其组成:选择新表位,所述新表位包含至少一个预测结合MHC I的最小表位和/或至少一个预测结合MHC II的最小表位,从而获得结合MHC的新表位。
- [0599] 8. 根据前述项目中任一项所述的方法,其中步骤c包括选择新表位,所述新表位包含至少一个预测结合MHC I的最小表位和至少一个预测结合MHC II的最小表位。
- [0600] 9. 根据前述项目中任一项所述的方法,其中步骤d包括确定每个所述新表位内结合MHC I和/或MHC II的最小表位的数目的步骤,并且其中较高数目的结合MHC I和/或MHC II的最小表位指示所述新表位的临床实用性的更高可能性。
- [0601] 10. 根据前述项目中任一项所述的方法,其中步骤d包括确定所述新表位的结合得分的步骤,其中所述结合得分为 $a*2x+b*y$ ,其中优选地如果 $y=0$ 则 $a=0$ 且如果 $x=0$ 则 $b=0$ ,并且其中如果 $y>0$ 则 $a=1$ 且如果 $x>0$ 则 $b=1$ ,并且其中较高的结合得分指示所述新表位的临床实用性的较高可能性。
- [0602] 11. 根据前述项目中任一项所述的方法,其中步骤d包括确定所述新表位的结合得分的步骤,其中所述结合得分为 $a*x+b*y$ ,其中优选地 $a=2$ 且 $b=1$ ,或者如果 $y=0$ 则 $a=0$ 且如果 $x=0$ 则 $b=0$ ,并且其中如果 $y>0$ 则 $a=1$ 且如果 $x>0$ 则 $b=1$ ,并且其中较高的结合得分指示所述新表位的临床实用性的较高可能性。
- [0603] 12. 根据前述项目中任一项所述的方法,其中步骤d进一步包括确定步骤c的结合MHC I的新表位的MHC I结合差异的步骤,其中所述MHC I结合差异由下式给出:(参考的%排名得分(MHC I))/新表位的%排名得分(MHC I)),其中具有高MHC I结合差异的新表位排名高于具有较低MHC I结合差异的新表位。
- [0604] 13. 根据前述项目中任一项所述的方法,其中步骤d包括确定步骤c的结合MHC I的最小表位的MHC I结合差异的步骤,其中所述MHC I结合差异由下式给出:(参考的%排名得分(MHC I))/(最小表位的%排名得分(MHC I)),其中包含具有高MHC I结合差异的最小表位的新表位排名高于包含具有较低MHC I结合差异的最小表位的新表位。
- [0605] 14. 根据前述项目中任一项所述的方法,其中每个新表位的MHC I结合差异等于其包含的最小表位的最大MHC I结合差异。
- [0606] 15. 根据前述项目中任一项所述的方法,其中步骤d进一步包括确定步骤c的结合

MHC II的新表位的MHC II结合差异的步骤,其中所述MHC II结合差异由下式给出:(参考的%排名得分(MHC II))/新表位的%排名得分(MHC II)),其中具有高MHC II结合差异的新表位排名高于具有较低MHC II结合差异的新表位。

[0607] 16.根据前述项目中任一项所述的方法,其中步骤d包括确定步骤c的结合MHC II的最小表位的MHC II结合差异的步骤,其中所述MHC II结合差异由下式给出:(参考的%排名得分(MHC II))/(最小表位的%排名得分(MHC II)),其中包含具有高MHC II结合差异的最小表位的新表位排名高于包含具有较低MHC II结合差异的最小表位的新表位。

[0608] 17.根据前述项目中任一项所述的方法,其中每个新表位的MHC II结合差异等于其包含的最小表位的最高MHC II结合差异。

[0609] 18.根据前述项目中任一项所述的方法,其中在步骤c中选择的新表位与MHC II结合。

[0610] 19.根据前述项目中任一项所述的方法,其中仅在所述多个新表位不包含任何与MHC II结合的新表位的情况下,才在步骤e中选择与MHC I结合的新表位。

[0611] 20.根据前述项目中任一项所述的方法,其中,步骤d进一步包括如下对新表位进行排名:

[0612] i)对于新表位中包含的每个最小表位,确定突变是在最小新表位的锚定位置还是非锚定位置;

[0613] ii)将新表位进行优先排序,优选地其中具有较高结合差异的新表位优先于具有较低结合差异的新表位。

[0614] 21.根据项目20所述的方法,其中步骤ii)中将新表位进行优先排序是通过向所述新表位分配它们所包含的最小表位的最高得分来进行的。

[0615] 22.根据项目20或21中任一项所述的方法,其中具有较高结合差异的新表位的得分高于具有较低结合差异的新表位的得分,并且其中包含在锚定位置的突变的新表位的得分高于包含在非锚定位置的突变的新表位的得分。

[0616] 23.根据项目19至21中任一项所述的方法,其中,将新表位进行优先排序是如下进行的:

[0617] 1)如果突变处于锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分最高;

[0618] 2)如果突变处于非锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分低于1)的最小表位的得分;

[0619] 3)如果突变处于锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于2)的最小表位的得分;

[0620] 4)如果突变处于非锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于3)的最小表位的得分;

[0621] 5)如果突变处于锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于4)的最小表位的得分;

[0622] 6)如果突变处于非锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于5)的最小表位的得分;

[0623] 7)如果最小表位具有结合差异低于1的突变,则最小表位的得分最低,而与突变的

位置无关。

[0624] 24. 根据前述项目中任一项所述的方法,其中步骤d进一步包括根据新表位的MHC I排名对所述新表位进行优先排序的步骤。

[0625] 25. 根据项目24所述的方法,其中包含一个或多个预测结合MHC I的最小表位的新表位的%排名得分(MHC I)等于所述一个或多个最小表位的最低%排名得分(MHC I)。

[0626] 26. 根据前述项目中任一项所述的方法,其中步骤e包括选择具有低于2.0,例如低于1.5,例如低于1,优选等于或低于0.5的%排名得分(MHC I)的新表位。

[0627] 27. 根据前述项目中任一项所述的方法,其中将具有高于2的%排名得分(MHC I)的结合MHC I的新表位排除或降低其优先排序。

[0628] 28. 根据前述项目中任一项所述的方法,其中步骤d进一步包括根据新表位的MHC II排名对所述新表位进行优先排序的步骤。

[0629] 29. 根据项目24至28任一项所述的方法,其中包含一个或多个预测结合MHC II的最小表位的新表位的%排名得分(MHC II)等于所述一个或多个最小表位的最低%排名得分(MHC II)。

[0630] 30. 根据前述项目中任一项所述的方法,其中步骤e包括选择具有低于10,例如低于2的%排名得分(MHC II)的新表位。

[0631] 31. 根据前述项目中任一项所述的方法,其中将具有高于10例如高于2的%排名得分(MHC I)的结合MHC I的新表位排除或降低其优先排序。

[0632] 32. 根据前述项目中任一项所述的方法,其中,步骤d进一步包括以下步骤:根据新表位的BLOSUM得分以降序对所述新表位进行优先排序,其中,新表位的BLOSUM得分等于它所包含的排名最佳的最小表位的BLOSUM得分,其中BLOSUM得分<1的最小表位的排名高于BLOSUM得分 $\geq$ 1的最小表位的排名。

[0633] 33. 根据前述项目中任一项所述的方法,其中根据项目1的步骤d进一步包括确定新表位中存在的氨基酸置换随机发生的概率,并且其中包含以低概率随机发生的氨基酸置换的新表位排名高于包含以较高概率随机发生的氨基酸置换的新表位。

[0634] 34. 根据项目33所述的方法,其中新表位中存在的氨基酸置换随机发生的所述概率是使用基于进化的评分矩阵来确定的。

[0635] 35. 根据项目34所述的方法,其中,所述评分矩阵是对数几率矩阵。

[0636] 36. 根据项目35所述的方法,其中,所述对数几率矩阵是BLOSUM矩阵。

[0637] 37. 根据项目36所述的方法,其中,所述BLOSUM矩阵是BLOSUM62矩阵。

[0638] 38. 根据项目37所述的方法,其中,将这样的新表位排除或降低其优先排序,所述新表位不具有与BLOSUM62得分低于1的氨基酸置换对相关的突变。

[0639] 39. 根据前述项目中任一项所述的方法,进一步包括确定所述新表位的RNA表达水平,其中将未检测到RNA表达的新表位排除在外或降低其优先排序。

[0640] 40. 根据项目39所述的方法,其中,具有高RNA表达水平的新表位排名高于具有较低RNA表达水平的新表位。

[0641] 41. 根据项目39或40中任一项所述的方法,其中所述RNA表达水平是针对结合MHC I和/或结合MHC II的新表位确定的。

[0642] 42. 根据前述项目中任一项所述的方法,其进一步包括比较新表位肽序列与人蛋

白质组的肽序列的步骤,其中将这样的新表位排除或降低其优先排序:所述新表位包含与人蛋白质组中的肽序列匹配的肽序列。

[0643] 43.根据项目42所述的方法,其中所述新表位肽包含5至15个氨基酸或由5至15个氨基酸组成。

[0644] 44.根据前述项目中任一项所述的方法,其进一步包括通过鉴定所述个体中的一个或多个肿瘤特异性突变来鉴定所述多个新表位的步骤。

[0645] 45.根据项目44所述的方法,其中所述一个或多个肿瘤特异性突变是通过来自所述个体的肿瘤DNA的外显子组测序来鉴定的。

[0646] 46.根据项目44至45中任一项所述的方法,其中所述一个或多个肿瘤特异性突变导致氨基酸置换。

[0647] 47.根据前述项目中任一项所述的方法,其中所述MHC结合亲和力包括确定所述个体的HLA基因型。

[0648] 48.根据前述项目中任一项所述的方法,其中,所述HLA基因型是从所述个体的血液样品中确定的。

[0649] 49.根据前述项目中任一项所述的方法,其中所述MHC I和/或MHC II结合亲和力是通过计算机预测来确定的。

[0650] 50.根据项目49所述的方法,其中所述计算机预测是通过使用计算机程序进行的,所述计算机程序预测肽与MHC I类和/或MHC II类分子的结合。

[0651] 51.根据前述项目中任一项所述的方法,将在下述基因中存在的新表位排除在外或降低其优先排序:其中所述基因在给定器官或组织中显示出相比于其他器官或组织至少5倍高的RNA表达水平。

[0652] 52.根据项目51所述的方法,其中所述器官选自心脏和脑。

[0653] 53.根据项目52所述的方法,其中所述器官选自肝,肺,胃,肾,脾,结肠和肠。

[0654] 54.根据前述项目中任一项所述的方法,其中步骤d进一步包括确定存在于结合MHC的新表位中的突变的等位基因频率的步骤,其中具有高等位基因频率的结合MHC的新表位排名高于具有较低等位基因频率的结合MHC的新表位。

[0655] 55.根据前述项目中任一项所述的方法,其中在一个以上活检中发现的包含突变的结合MHC的新表位排名高于仅在一个活检中发现的包含突变的结合MHC的新表位。

[0656] 56.根据前述项目中任一项所述的方法,其中所述新表位的长度为7至40个氨基酸。

[0657] 57.根据前述项目中任一项所述的方法,其中所述新表位的长度为15至30个氨基酸。

[0658] 58.根据前述项目中任一项所述的方法,其中所述新表位的长度为25至30个氨基酸,例如27个氨基酸。

[0659] 59.根据前述项目中任一项所述的方法,其中所述突变实质上位于新表位序列的中间。

[0660] 60.根据前述项目中任一项所述的方法,其中,所述个体是癌症患者。

[0661] 61.根据前述项目中任一项所述的方法,其中,针对所述个体选择的新表位被用于癌症疫苗中。

- [0662] 62. 根据前述项目中任一项所述的方法,其中A是1至100,例如5至50,例如5至30的整数。
- [0663] 63. 根据前述项目中任一项所述的方法,其中A是10至20的整数。
- [0664] 64. 根据前述项目中任一项所述的方法,其中, $x > y$ 。
- [0665] 65. 根据前述项目中任一项所述的方法,其中 $x \geq 2y$ 。
- [0666] 66. 根据前述项目中任一项所述的方法,其中 $x > 2.5y$ 。
- [0667] 67. 根据前述项目中任一项所述的方法,其中 $A \geq 3$ 。
- [0668] 68. 根据前述项目中任一项所述的方法,其中A个新表位中的至少一些,例如A个新表位的A/2个,例如A个新表位的A/2至A个,能够诱导CD8+T细胞应答。
- [0669] 69. 根据前述项目中任一项所述的方法,所述方法包括以下步骤:
- [0670] a. 从所述个体获得多个新表位,每个新表位均包含至少3个最小表位,其中与参考序列相比,每个新表位均包含至少一个免疫原性突变;
- [0671] b. 确定每个所述新表位的MHC I和MHC II结合亲和力,从而鉴定x个结合MHC I的新表位和y个结合MHC II的新表位,其中 $x+y \geq 3$ ;
- [0672] c. 选择预测至少与MHC I结合并任选与MHC II结合的新表位,从而获得结合MHC的新表位;
- [0673] d. 根据临床实用性的可能性对结合MHC的新表位进行排名;其中,排名是如下进行的:
- [0674] i. 根据新表位所包含的结合MHC I和任选地结合MHC II的最小表位的数目对所述新表位进行排名,其中较高的数目给出较高的排名;
- [0675] e. 在排名最高的结合MHC的新表位中选择A个新表位,
- [0676] 从而选择可能具有临床实用性的A个新表位。
- [0677] 70. 根据项目69所述的方法,其中新表位中所包含的结合MHC II的最小表位的数目多于新表位中所包含的结合MHC II的最小表位的数目,例如是两倍多。
- [0678] 71. 根据项目69至70中任一项所述的方法,其中步骤d的步骤i.之后是步骤ii.,其中:
- [0679] ii. 确定新表位的MHC I和/或MHC II结合差异得分,其中MHC I结合差异由下式给出:(参考的%排名得分(MHC I))/(新表位的%排名得分(MHC I)),并且MHC II结合差异由下式给出:(参考的%排名得分(MHC II))/(新表位的%排名得分(MHC II));其中具有高结合差异的新表位排名高于具有较低结合差异的新表位。
- [0680] 72. 根据项目70所述的方法,其中,将所述新表位如下以降序排名:
- [0681] 1) 如果突变处于锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分最高;
- [0682] 2) 如果突变处于非锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分低于1)的最小表位的得分;
- [0683] 3) 如果突变处于锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于2)的最小表位的得分;
- [0684] 4) 如果突变处于非锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于3)的最小表位的得分;

[0685] 5) 如果突变处于锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于4)的最小表位的得分;

[0686] 6) 如果突变处于非锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于5)的最小表位的得分;

[0687] 7) 如果最小表位具有结合差异低于1的突变,则最小表位的得分最低,而与突变的位置无关。

[0688] 73. 根据项目69至72中任一项所述的方法,其中步骤d的步骤i.或步骤ii.之后是步骤iii.,其中:

[0689] iii. 根据新表位的%排名得分对其进行排名,其中具有低%排名得分的新表位排名高于具有高%排名得分的新表位。

[0690] 74. 根据项目73所述的方法,其中新表位的%排名得分等于其所包含的最小表位的最低%排名得分。

[0691] 75. 根据项目73至74中任一项所述的方法,其中新表位或最小表位被预测为:

[0692] -如果%排名得分(MHC I)  $\leq 0.5$ ,则为强MHC I结合剂;

[0693] -如果 $0.5 < \%$ 排名得分(MHC I)  $\leq 2$ ,则为弱MHC I结合剂;

[0694] -如果%排名得分(MHC I)  $> 2$ ,则为非MHC I结合剂。

[0695] 76. 根据项目73至75中任一项所述的方法,其中新表位或最小表位被预测为:

[0696] -如果%排名得分(MHC II)  $\leq 2$ ,则为强MHC I结合剂;

[0697] -如果 $2 < \%$ 排名得分(MHC II)  $\leq 10$ ,则为弱MHC I结合剂;

[0698] -如果%排名得分(MHC II)  $> 10$ ,则为非MHC I结合剂。

[0699] 77. 根据项目69至76中任一项所述的方法,其中步骤i.,ii.或iii.之后是步骤iv,其中:

[0700] iv. 根据新表位的BLOSUM分数对所述新表位进行评分,其中低BLOSUM得分给出较高的得分。

[0701] 78. 根据项目69至77中任一项所述的方法,其中优先排序具有BLOSUM得分 $< 3$ ,例如BLOSUM得分 $< 2$ ,例如BLOSUM得分 $< 1$ ,例如BLOSUM得分 $< 0$ 的新表位。

[0702] 79. 用于为个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0703] a. 从所述个体获得一个或多个新表位,每个新表位均包含至少一个最小表位,其中与参考序列相比,每个新表位均包含至少一个突变,诸如免疫原性突变;

[0704] b. 确定每个所述新表位内的至少一个最小表位(例如至少两个,三个或四个最小表位)的MHC I结合亲和力,并确定每个所述新表位的结合MHC I的最小表位数目;

[0705] c. 对新表位进行如下排名:

[0706] i. 将包含大量结合MHC I的最小表位的新表位优先排序,并选择具有高分的第一组新表位;

[0707] ii. 任选地,将来自第一组新表位的新表位如下进行优先排序:

[0708] 1) 如果突变处于锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分最高;

[0709] 2) 如果突变处于非锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分低于1)的最小表位的得分;

[0710] 3) 如果突变处于锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于2)的最小表位的得分;

[0711] 4) 如果突变处于非锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于3)的最小表位的得分;

[0712] 5) 如果突变处于锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于4)的最小表位的得分;

[0713] 6) 如果突变处于非锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于5)的最小表位的得分;

[0714] 7) 如果最小表位具有结合差异低于1的突变,则最小表位的得分最低,而与突变的位置无关;

[0715] 其中,MHC I结合差异由下式给出:(用于参考的%排名得分(MHC I))/(用于最小表位的%排名得分(MHC I));并选择第二组得分高的新表位;

[0716] iii. 任选地,基于MHC I%排名得分,将来自第二组的新表位进行优先排序,并选择具有低MHC I%排名得分的第三组新表位;

[0717] iv. 任选地,从第二组或第三组中选择第四组新表位,其包括与已知被T细胞识别的表位高度相似的最小表位;

[0718] v. 基于BLOSUM得分,将来自第二组、来自第三组或来自第四组的新表位进行优先排序,其中小于预定阈值的BLOSUM得分排名高于等于或大于所述阈值的BLOSUM得分,并且选择BLOSUM得分小于所述阈值的第五组新表位,其中所述阈值优选为1;

[0719] 其中第一组,第二组,第三组,第四组或第五组新表位包含所述A个表位。

[0720] 80. 根据项目79所述的方法,其中步骤c包括步骤i.和ii.,或i.和iii.,或i.和iv.,或i.和v.,或i.,ii.和iii.,或i.,ii.和iv.,或i.,iii.和iv.,或i.,ii.和v.,或i.,iii.和v.,或i.,iv.和v.,或i.,ii.,iii.和iv.,或i.,ii.,iii.和v.,或i.,ii.,iv.和v.,或i.,iii.,iv.和v.,或i.,ii.,iii.,iv.和v.

[0721] 81. 根据项目79至80中任一项所述的方法,其中所述方法进一步包括根据项目1至78中任一项所述的方法的任何特征。

[0722] 82. 用于为个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0723] a. 从所述个体获得一个或多个新表位,每个新表位均包含至少一个最小表位,其中与参考序列相比,每个新表位均包含至少一个突变,诸如免疫原性突变;

[0724] b. 确定每个所述新表位内的至少一个最小表位(例如至少两个,三个或四个最小表位)的MHC I结合亲和力,并确定每个所述新表位的结合MHC I的最小表位数;

[0725] c. 对新表位进行如下排名:

[0726] i. 将包含大量结合MHC I的最小表位的新表位优先排序,并选择具有高分的第一组新表位;

[0727] ii. 任选地,将来自第一组新表位的新表位如下进行优先排序:

[0728] 1) 如果突变处于锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分最高;

[0729] 2) 如果突变处于非锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分低于1)的最小表位的得分;

[0730] 3) 如果突变处于锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于2)的最小表位的得分;

[0731] 4) 如果突变处于非锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于3)的最小表位的得分;

[0732] 5) 如果突变处于锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于4)的最小表位的得分;

[0733] 6) 如果突变处于非锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于5)的最小表位的得分;

[0734] 7) 如果最小表位具有结合差异低于1的突变,则最小表位的得分最低,而与突变的位置无关;

[0735] 其中,MHC I结合差异由下式给出:(参考的%排名得分(MHC I))/(最小表位的%排名得分(MHC I));并选择第二组得分高的新表位;

[0736] iii. 任选地,基于MHC I%排名得分,将来自第二组的新表位进行优先排序,并选择具有低MHC I%排名得分的第三组新表位;

[0737] iv. 任选地,从第二组或第三组中选择第四组新表位,其包括与已知被T细胞识别的表位高度相似的最小表位;

[0738] v. 基于BLOSUM得分,将来自第二组、来自第三组或来自第四组的新表位进行优先排序,其中小于预定阈值的BLOSUM得分排名高于等于或大于所述阈值的BLOSUM得分,并且选择BLOSUM得分小于所述阈值的第五组新表位,其中所述阈值优选为1;

[0739] vi. 任选地,基于在两个或更多个样品中发现的新表位,从第一,第二,第三,第四或第五组新表位中选择新表位,并选择在两个或更多个样品中发现的第六组新表位;

[0740] vii. 任选地,基于至少两个不同的变体调用器(variant caller)对突变的鉴定,从第一,第二,第三,第四,第五或第六组新表位中选择新表位,并选择包含由至少两个不同的变体调用器所鉴定的突变的第七组新表位,

[0741] 其中第一,第二,第三,第四,第五,第六或第七组新表位包含所述A个表位。

[0742] 83. 根据项目82所述的方法,其中步骤c包括步骤i.和ii.,或i.和iii.,或i.和iv.,或i.和v.,或i.和vi.,或i.和vii.,或i.,ii.和iii.,或i.,ii.和iv.,或i.,ii.和v.,或i.,ii.和vi.,或i.,ii.和vii.,或i.,iii.和iv.,或i.,iii.和v.,或i.,iii.和vi.,或i.,iii.和vii.,或i.,iv.和v.,或i.,iv.和vi.,或i.,iv.和vii.,或i.,v.和vi.,或i.,v.和vii.,或i.,vi.和vii.,或ii.,iii.和iv.,或ii.,iii.和v.,或ii.,iii.和vi.,或ii.,iii.和vii.,或ii.,iv.和v.,或ii.,iv.和vi.,或ii.,iv.和vii.,或ii.,v.和vi.,或ii.,v.和vii.,或ii.,vi.和vii.,或iii.,iv.和v.,或iii.,iv.和vi.,或iii.,iv.和vii.,或iii.,v.和vi.,或iii.,v.和vii.,或iii.,vi.和vii.,或iv.,v.和vi.,或iv.,v.和vii.,或iv.,vi.和vii.,或v.,vi.和vii.。

[0743] 84. 根据项目82至83中任一项所述的方法,其中所述方法进一步包括根据项目1至78中任一项所述的方法的任何特征。

[0744] 85. 制备包含新表位的癌症疫苗的方法,所述方法包括使用根据项目1-78或79-84中任一项所述的方法选择所述新表位的步骤。

[0745] 86. 根据项目85所述的方法,其中选择10-20个新表位。

[0746] 87. 根据项目85至86中任一项所述的方法,其中所述癌症疫苗包含核苷酸构建体,所述核苷酸构建体包括:

[0747] -靶向单元

[0748] -二聚化单元

[0749] -第一接头

[0750] -抗原单元,其中所述抗原单元包含A-1个抗原亚单元,每个亚单元包含编码至少一个所述新表位的序列和第二接头,并且所述抗原单元进一步包含编码所述新表位之一的最终序列,其中A是1至100的整数,优选地A是3至50的整数,

[0751] 其中所述核苷酸构建体以免疫学有效量应用于抗癌疫苗。

[0752] 88. 根据项目85至87中任一项所述的方法,其中所述第二接头是丝氨酸-甘氨酸接头。

[0753] 89. 根据项目85至88中任一项所述的方法,其中所述靶向单元对选自CCR1、CCR3和CCR5的趋化因子受体具有亲和力。

[0754] 90. 根据项目89所述的方法,其中所述靶向单元包含与氨基酸序列SEQ ID NO:1 (MIP1a)的氨基酸5-70具有至少80%序列同一性的氨基酸序列。

[0755] 91. 通过项目85至90中任一项所述的方法获得的癌症疫苗。

[0756] 92. 根据项目91所述的癌症疫苗,其中所述癌症疫苗包含核苷酸构建体,所述核苷酸构建体包含:

[0757] -靶向单元

[0758] -二聚化单元

[0759] -第一接头

[0760] -抗原单元,其中所述抗原单元包含n-1个抗原亚单元,每个亚单元包含编码至少一个所述新表位的序列和第二接头,并且所述抗原单元进一步包含编码所述新表位之一的最终序列,其中n是1至100的整数,例如3至50。

[0761] 其中所述核苷酸构建体以免疫学有效量应用于抗癌疫苗。

[0762] 93. 根据项目91至92中任一项所述的癌症疫苗,其中所述第二接头是丝氨酸-甘氨酸接头。

[0763] 94. 根据项目91至93中任一项所述的癌症疫苗,其中所述靶向单元对选自CCR1、CCR3和CCR5的趋化因子受体具有亲和力。

[0764] 95. 根据项目91至94中任一项所述的癌症疫苗,其中所述靶向单元包含与氨基酸序列SEQ ID NO:1 (MIP1a)的氨基酸5-70具有至少80%序列同一性的氨基酸序列。

[0765] 96. 用于为个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0766] a. 从所述个体获得一个或多个新表位,每个新表位均包含至少一个最小表位,其中与参考序列相比,每个新表位均包含至少一个突变,诸如免疫原性突变;

[0767] B. 对于每个所述新表位,确定MHC I和/或MHC II结合亲和力;

[0768] c. 选择包含至少一个预测与MHC I和/或MHC II结合的最小表位的新表位,从而获得结合MHC的新表位;

[0769] d. 根据临床实用性的可能性对结合MHC的新表位进行排名;

[0770] e. 在最高排名的结合MHC的新表位中选择A个新表位,其中A是整数并且A是至少3

个,诸如至少4个,诸如至少5个,

[0771] 从而选择以免疫活性量施用于个体时能够诱导CD8+T细胞应答的A个新表位。

[0772] 97. 根据项目96所述的方法,其中所述方法进一步包括根据项目1至78中任一项所述的方法的任何特征。

[0773] 98. 用于为患有或怀疑患有癌症的个体选择数目A个新表位的方法,所述方法包括以下步骤:

[0774] a. 从所述个体获得一个或多个新表位,每个新表位包含至少一个最小表位,其中与参考序列相比,每个新表位包含至少一个突变例如免疫原性突变,其中所述最小表位由等于或小于新表位的氨基酸数目的氨基酸数目组成,并包含所述至少一个突变;其中优选地获得新表位包括鉴定对肿瘤特异性的核酸序列中的突变的步骤;

[0775] b. 对于每个所述新表位内的至少一个最小表位,诸如至少两个,三个或四个最小表位,确定MHC I和/或MHC II结合亲和力,任选地,其中结合亲和力是通过计算机预测确定的;

[0776] c. 选择包含至少一个预测与MHC I和/或MHC II结合的最小表位的新表位,从而获得结合MHC的新表位;

[0777] d. 根据临床实用性的可能性对结合MHC的新表位进行排名;

[0778] e. 在排名最高的结合MHC的新表位中选择A个新表位,

[0779] 从而选择可能具有临床实用性的A个新表位,其中步骤d进一步包括如下对新表位进行排名:

[0780] i) 对于新表位中包含的每个最小表位,确定突变是在最小新表位的锚定位置还是非锚定位置;

[0781] ii) 将新表位进行优先排序,

[0782] 任选地,其中步骤ii) 中将新表位进行优先排序是通过向每个新表位分配它们所包含的最小表位的最高得分来进行的,并且其中将新表位进行优先排序是如下进行的:

[0783] 1) 如果突变处于锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分最高;

[0784] 2) 如果突变处于非锚定位置并且最小表位的结合差异等于或大于20,则最小表位的得分低于1) 的最小表位的得分;

[0785] 3) 如果突变处于锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于2) 的最小表位的得分;

[0786] 4) 如果突变处于非锚定位置并且最小表位的结合差异小于20且等于或大于3,则最小表位的得分低于3) 的最小表位的得分;

[0787] 5) 如果突变处于锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于4) 的最小表位的得分;

[0788] 6) 如果突变处于非锚定位置并且最小表位的结合差异小于3且等于或大于1,则最小表位的得分低于5) 的最小表位的得分;

[0789] 7) 如果最小表位具有结合差异低于1的突变,则最小表位的得分最低,而与突变的位置无关。

[0790] 99. 用于为患有或怀疑患有癌症的个体选择数目A个新表位的方法,所述方法包括

以下步骤:

[0791] a. 从所述个体获得一个或多个新表位, 每个新表位包含至少一个最小表位, 其中与参考序列相比, 每个新表位包含至少一个突变例如免疫原性突变, 其中所述最小表位由等于或小于新表位的氨基酸数目的氨基酸数目组成, 并包含所述至少一个突变; 其中优选地获得新表位包括鉴定对肿瘤特异性的核酸序列中的突变的步骤;

[0792] b. 对于每个所述新表位内的至少一个最小表位, 诸如至少两个, 三个或四个最小表位, 确定MHC I和/或MHC II结合亲和力, 任选地, 其中结合亲和力是通过计算机预测确定的;

[0793] c. 选择包含至少一个预测与MHC I和/或MHC II结合的最小表位的新表位, 从而获得结合MHC的新表位;

[0794] d. 根据临床实用性的可能性对结合MHC的新表位进行排名;

[0795] e. 在排名最高的结合MHC的新表位中选择A个新表位,

[0796] 从而选择可能具有临床实用性的A个新表位,

[0797] 其中步骤d包括根据新表位所包含的最小表位的数目对所述新表位进行排名, 其中较高数目的最小表位给出较高的排名。

[0798] 100. 用于为患有或怀疑患有癌症的个体选择数目A个新表位的方法, 所述方法包括以下步骤:

[0799] a. 从所述个体获得一个或多个新表位, 每个新表位包含至少一个最小表位, 其中与参考序列相比, 每个新表位包含至少一个突变例如免疫原性突变, 其中所述最小表位由等于或小于新表位的氨基酸数目的氨基酸数目组成, 并包含所述至少一个突变; 其中优选地获得新表位包括鉴定对肿瘤特异性的核酸序列中的突变的步骤;

[0800] b. 对于每个所述新表位内的至少一个最小表位, 诸如至少两个, 三个或四个最小表位, 确定MHC I和/或MHC II结合亲和力, 任选地, 其中结合亲和力是通过计算机预测确定的;

[0801] c. 选择包含至少一个预测与MHC I和/或MHC II结合的最小表位的新表位, 从而获得结合MHC的新表位;

[0802] d. 根据临床实用性的可能性对结合MHC的新表位进行排名;

[0803] e. 在排名最高的结合MHC的新表位中选择A个新表位,

[0804] 从而选择可能具有临床实用性的A个新表位,

[0805] 其中步骤d包括根据发现新表位的样品数目对新表位进行排名, 其中在较高数目的样品中发现的新表位排名高于在较低数目的样品中发现的新表位, 优选地其中样品是来自不同病变的样品。

[0806] 101. 根据项目99至100中任一项所述的方法, 其中所述方法进一步包括根据项目1至78中任一项所述的方法的任何特征。

## 序列表

<110> 耐考德治疗股份有限公司

<120> 用于选择新表位的方法

<130> P4691PC00

<160> 5

<170> PatentIn version 3.5

<210> 1

<211> 93

<212> PRT

<213> 智人

<400> 1

Met Gln Val Ser Thr Ala Ala Leu Ala Val Leu Leu Cys Thr Met Ala  
1           5           10           15

[0001] Leu Cys Asn Gln Val Leu Ser Ala Pro Leu Ala Ala Asp Thr Pro Thr  
          20           25           30

Ala Cys Cys Phe Ser Tyr Thr Ser Arg Gln Ile Pro Gln Asn Phe Ile  
          35           40           45

Ala Asp Tyr Phe Glu Thr Ser Ser Gln Cys Ser Lys Pro Ser Val Ile  
          50           55           60

Phe Leu Thr Lys Arg Gly Arg Gln Val Cys Ala Asp Pro Ser Glu Glu  
65           70           75           80

Trp Val Gln Lys Tyr Val Ser Asp Leu Glu Leu Ser Ala  
          85           90

<210> 2

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 接头

<220>

<221> MISC\_FEATURE

<222> (1)..(5)

<223> linker

<400> 2

Gly Gly Gly Ser Ser

1            5

<210> 3

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 接头

[0002]

<220>

<221> MISC\_FEATURE

<222> (1)..(5)

<223> 接头

<400> 3

Gly Gly Gly Ser Gly

1            5

<210> 4

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 接头

<220>

<221> MISC\_FEATURE

<222> (1)..(5)

<223> 接头

<400> 4

Gly Gly Gly Gly Ser  
1 5

<210> 5  
<211> 10  
<212> PRT  
<213> 人工序列

[0003] <220>  
<223> 接头

<220>  
<221> MISC\_FEATURE  
<222> (1)..(10)  
<223> 接头

<400> 5

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser  
1 5 10

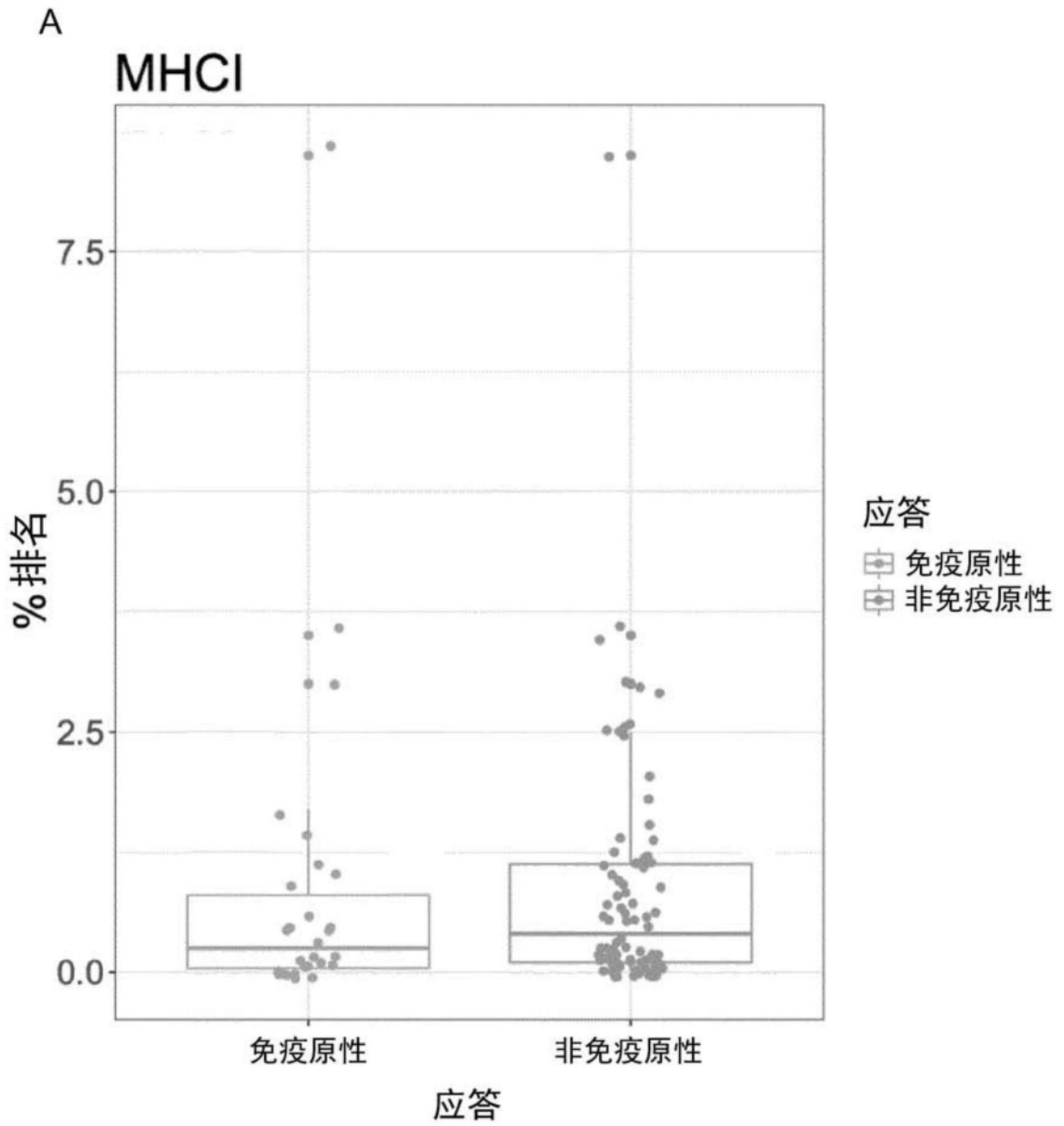


图1

B

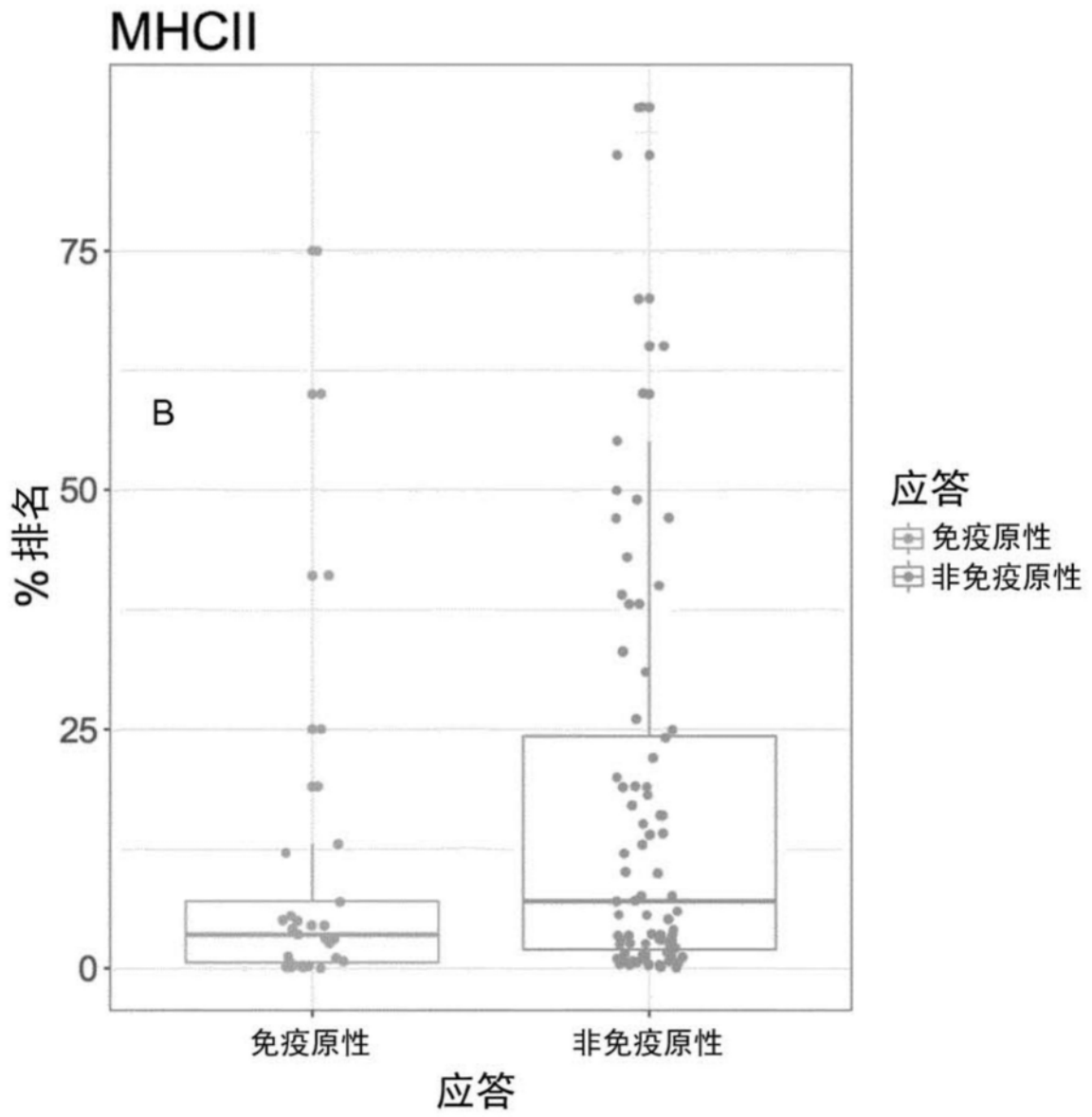


图1 (续)

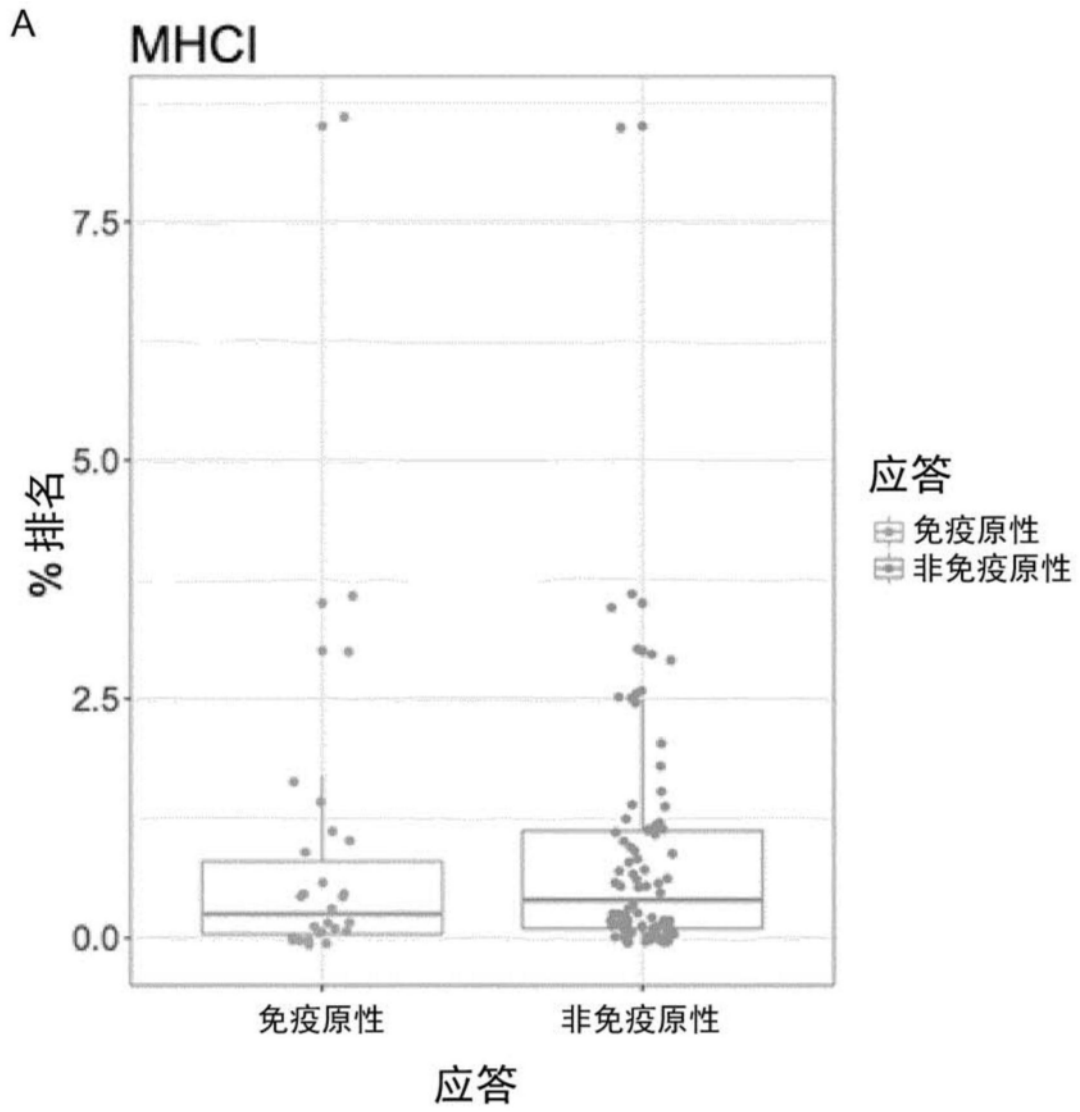


图2

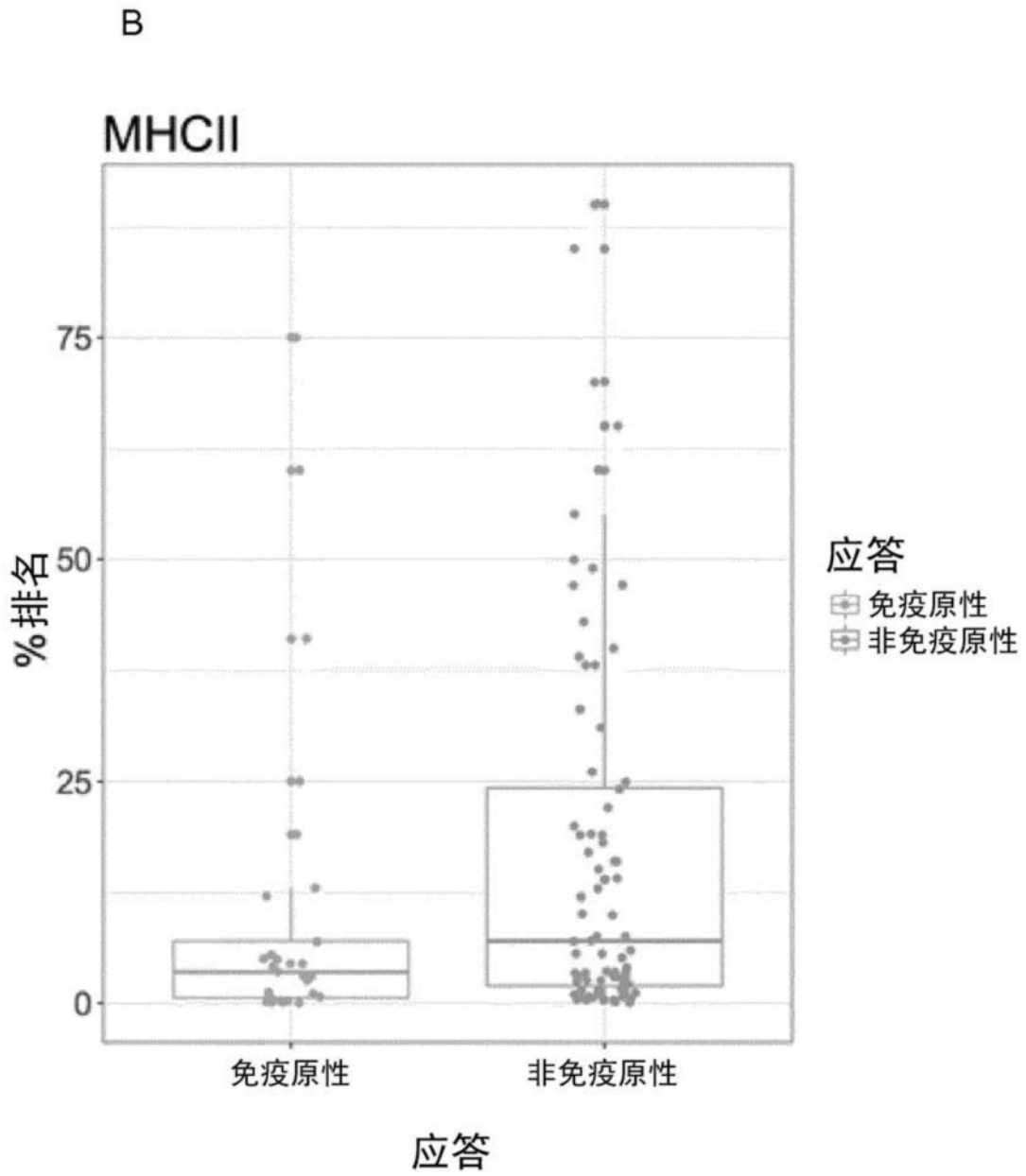


图2(续)

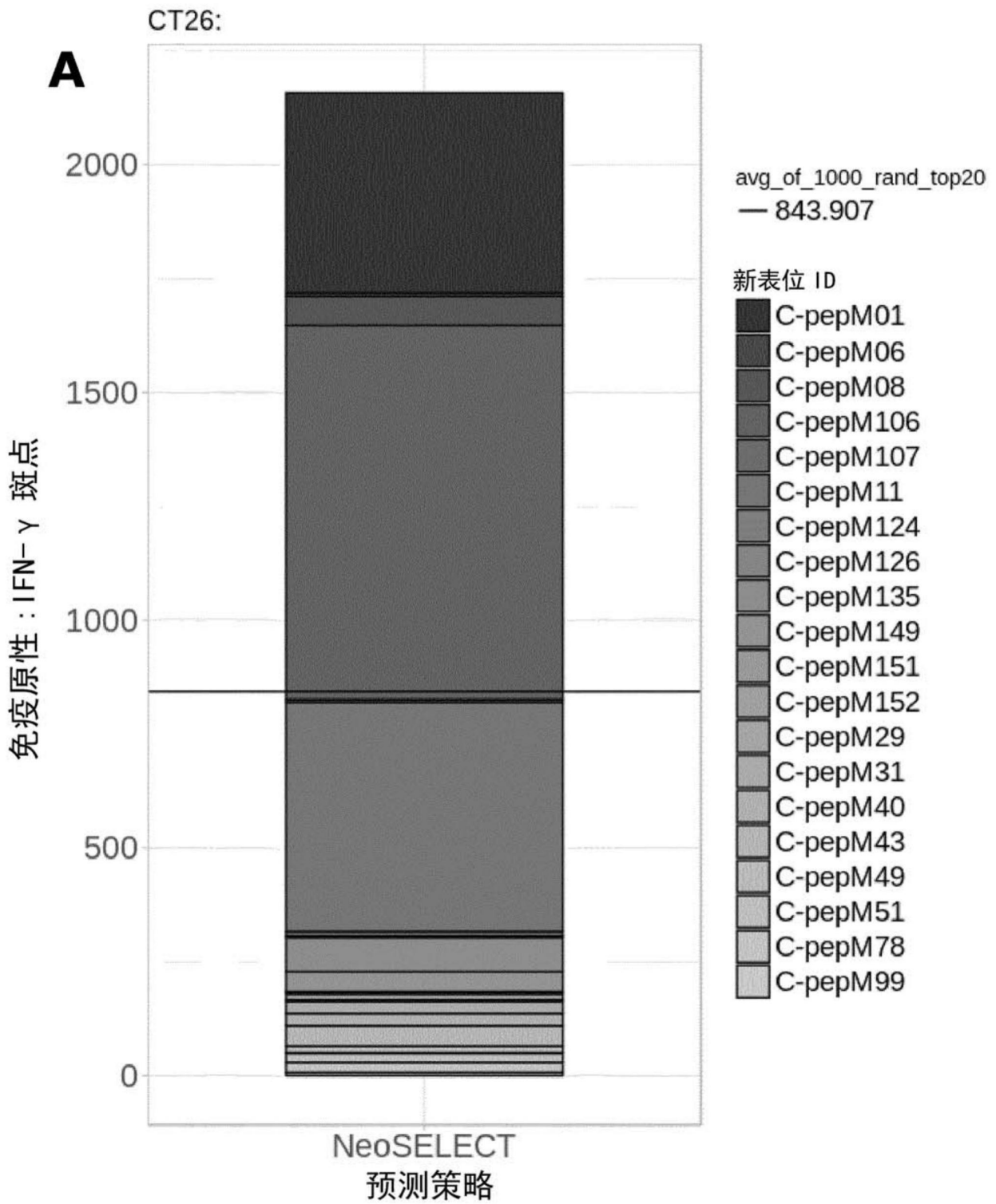


图3

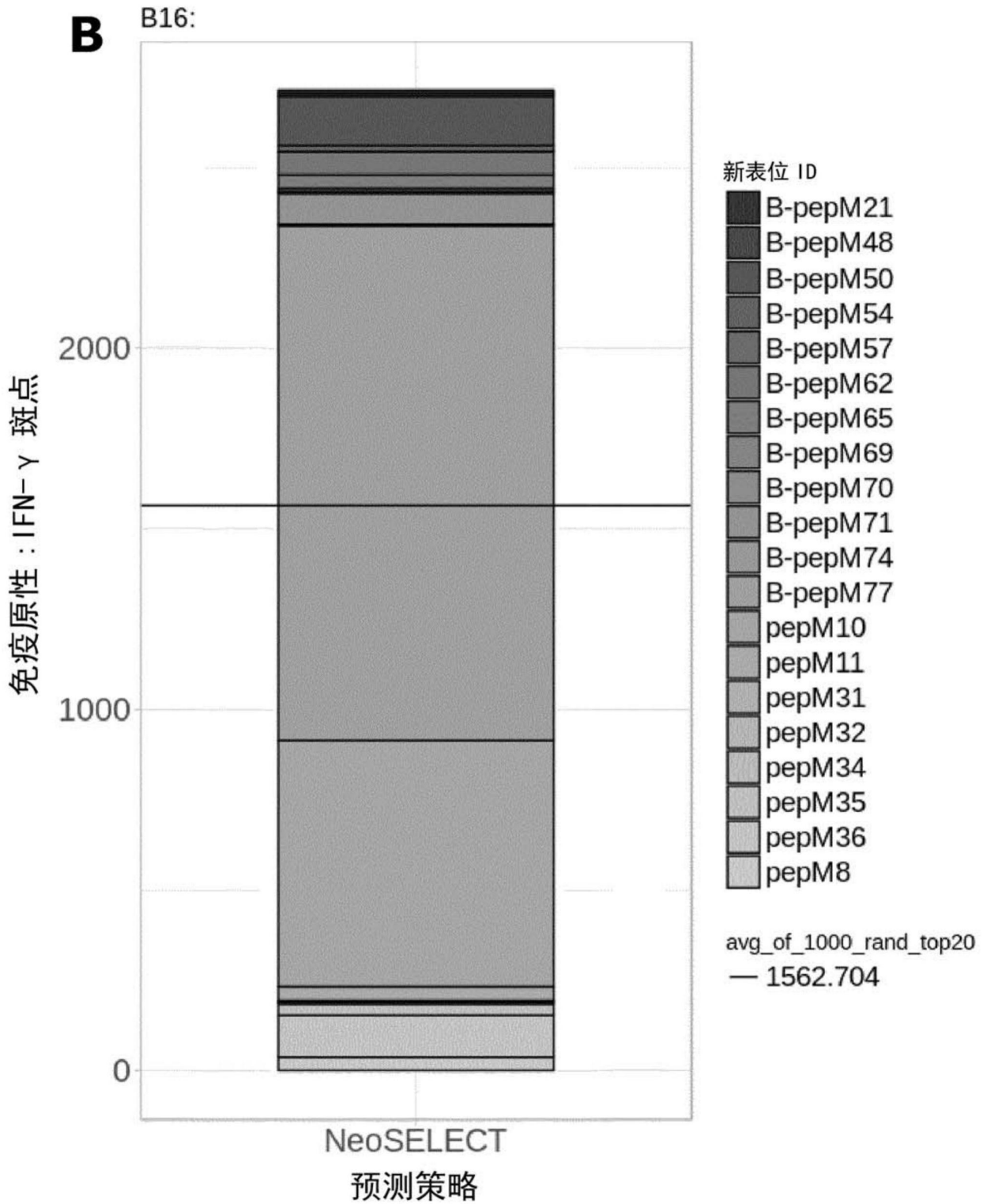


图3 (续)

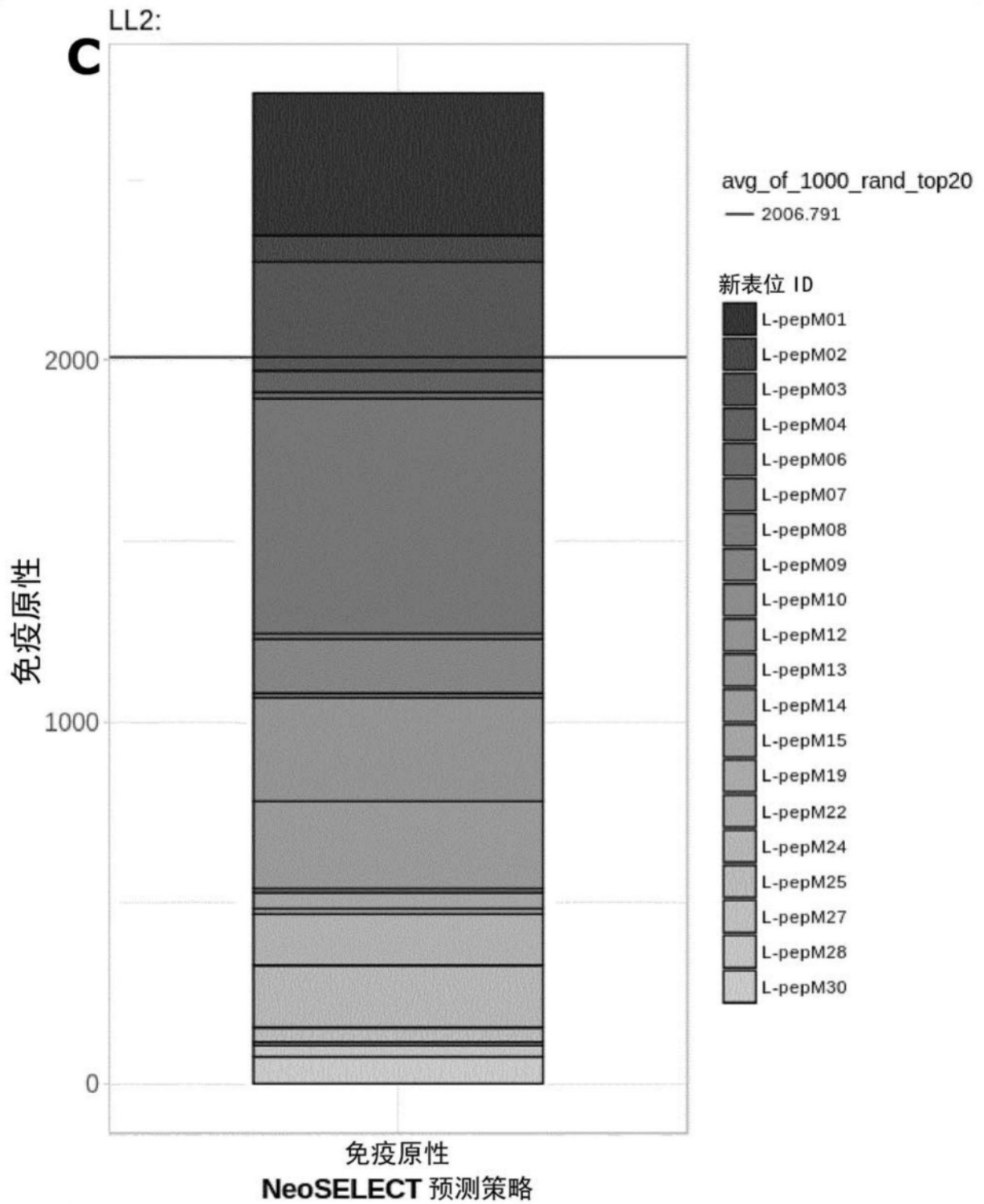


图3 (续)

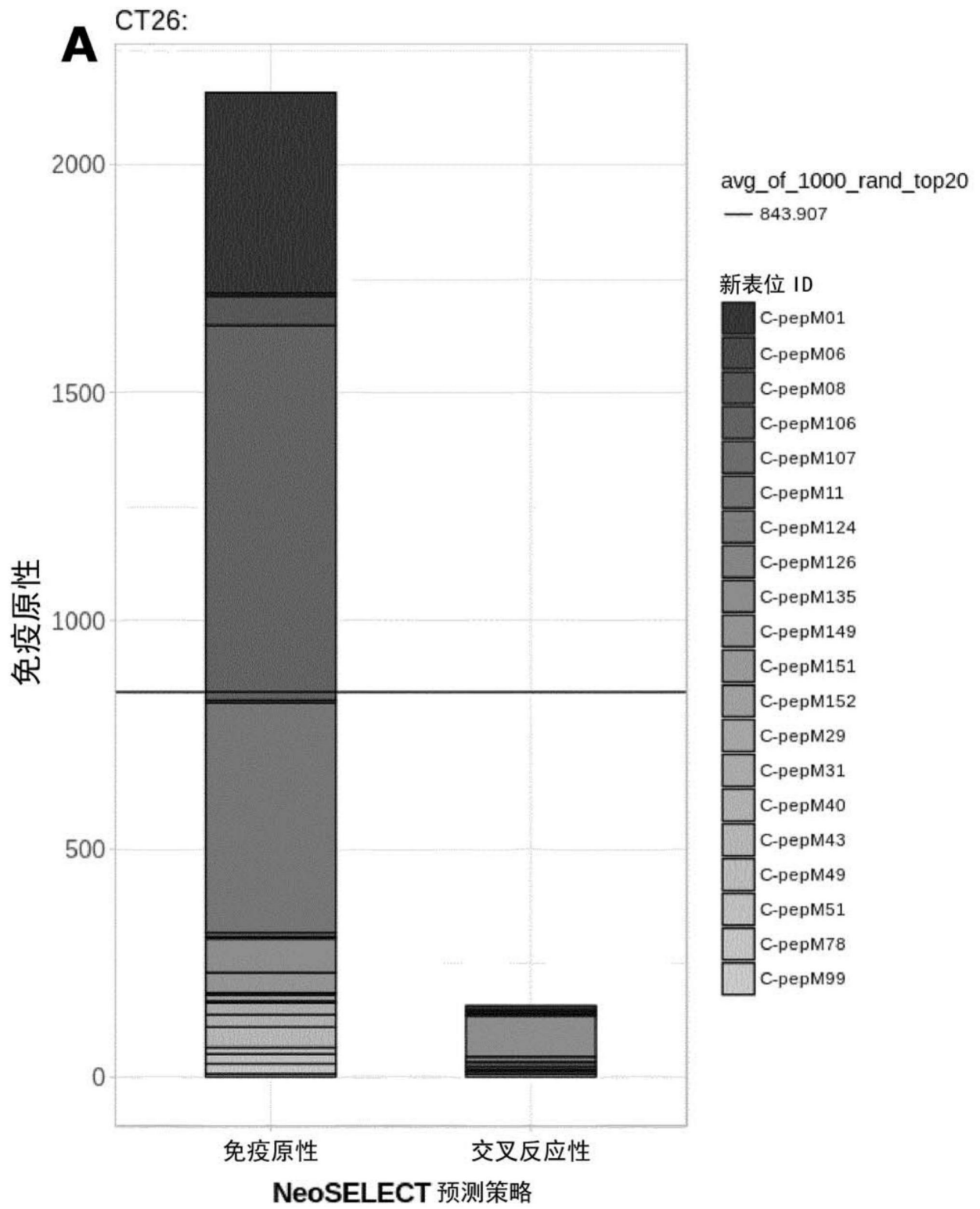


图4

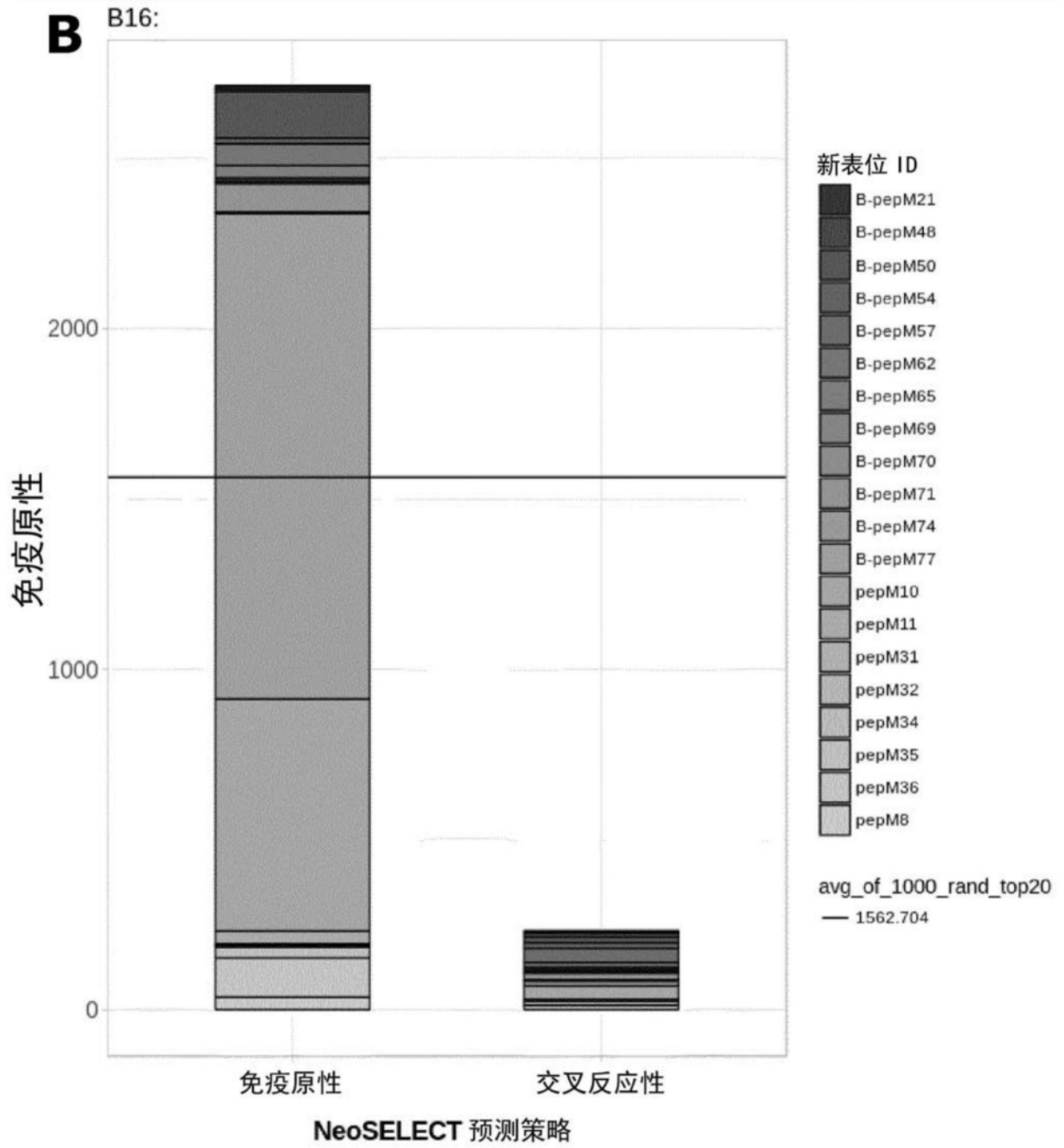


图4(续)

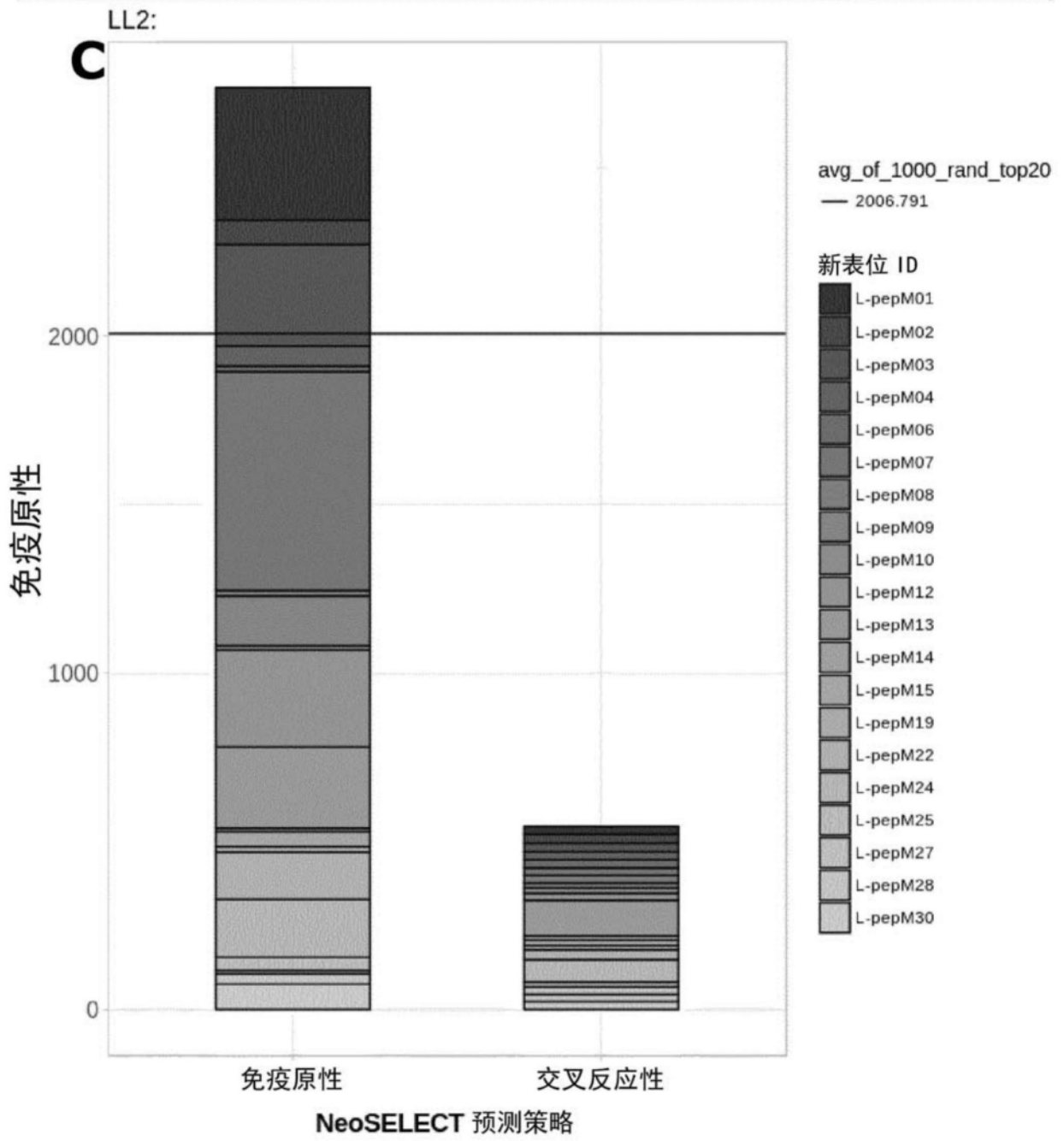


图4(续)

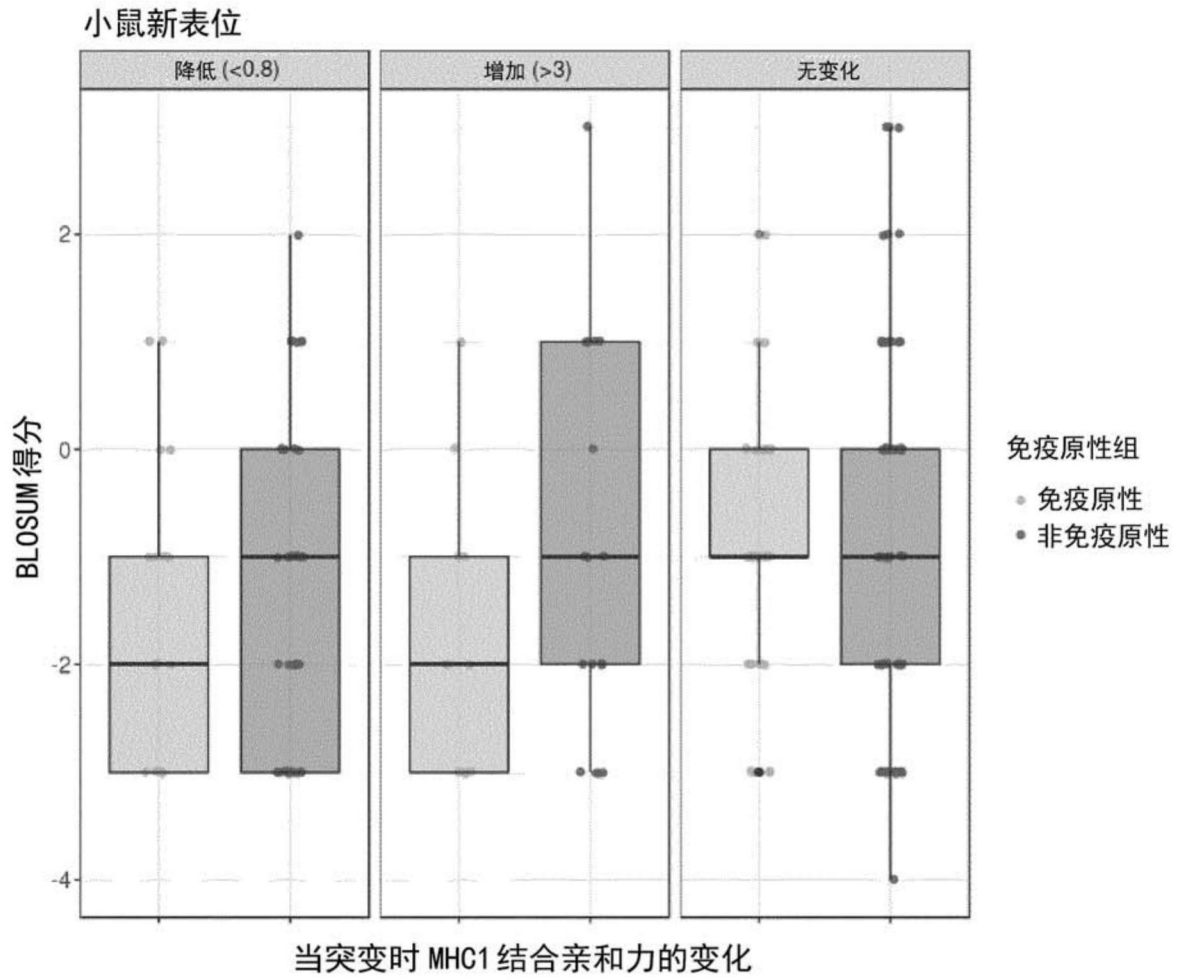


图5

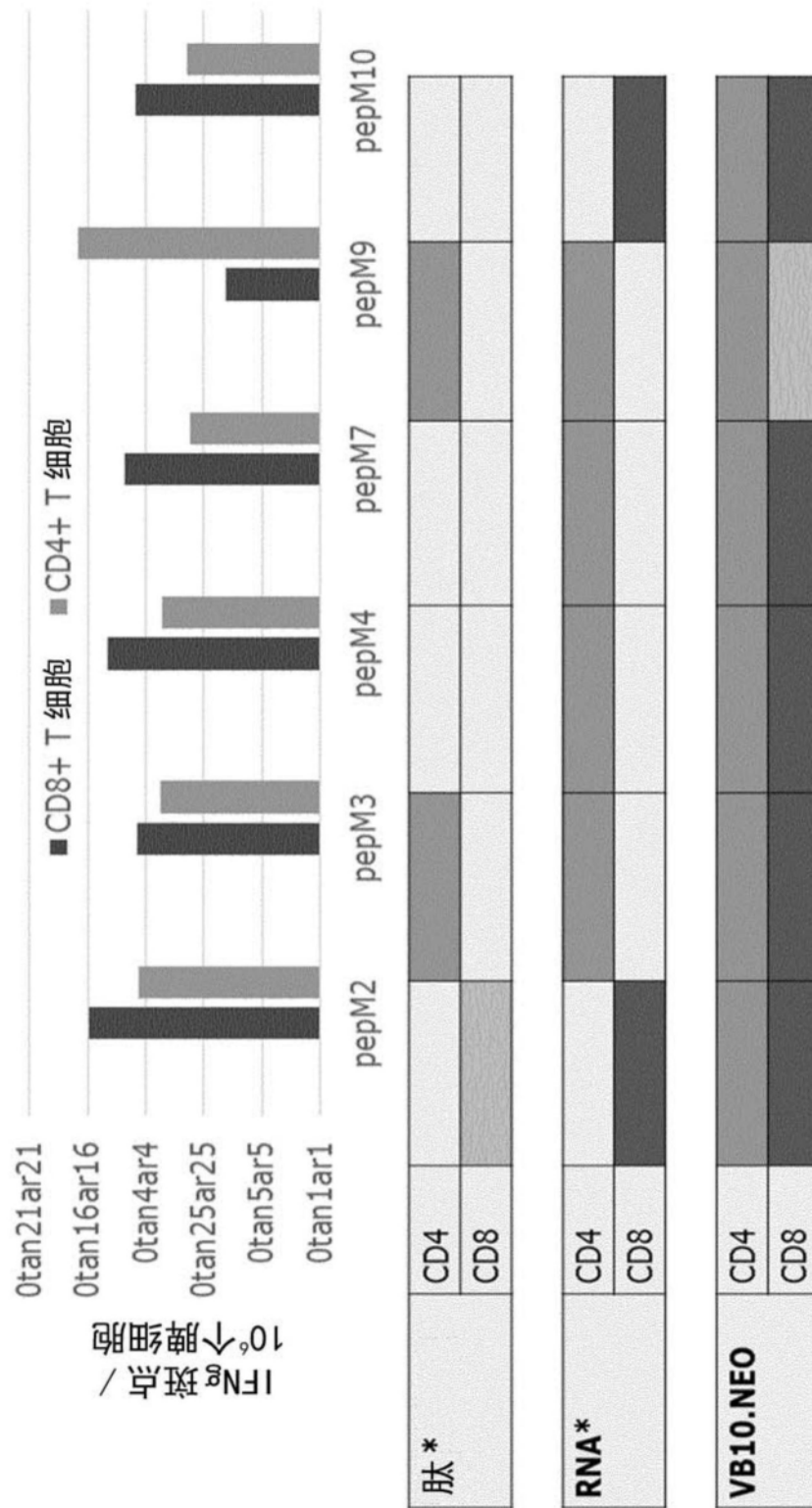


图6

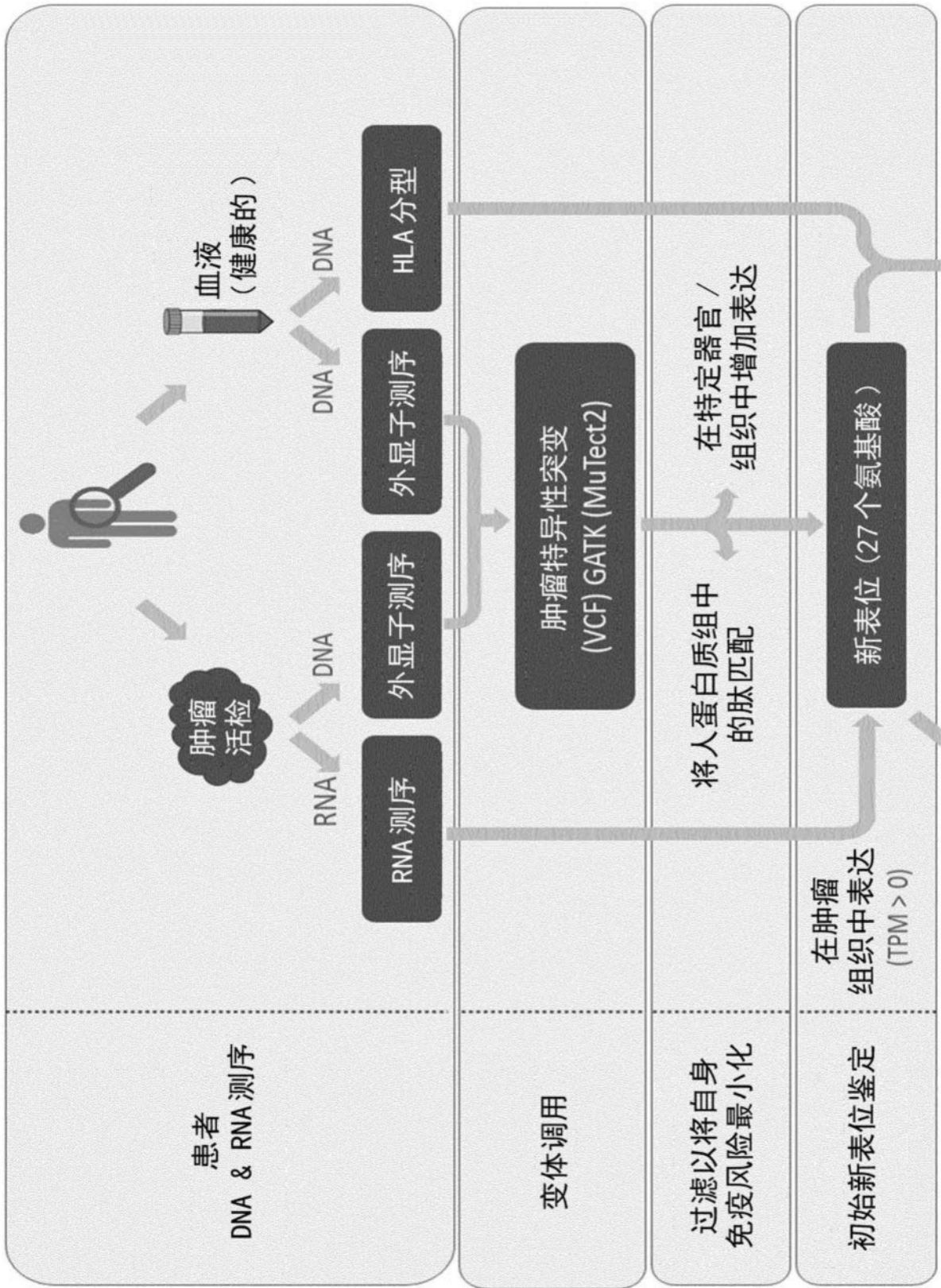


图7

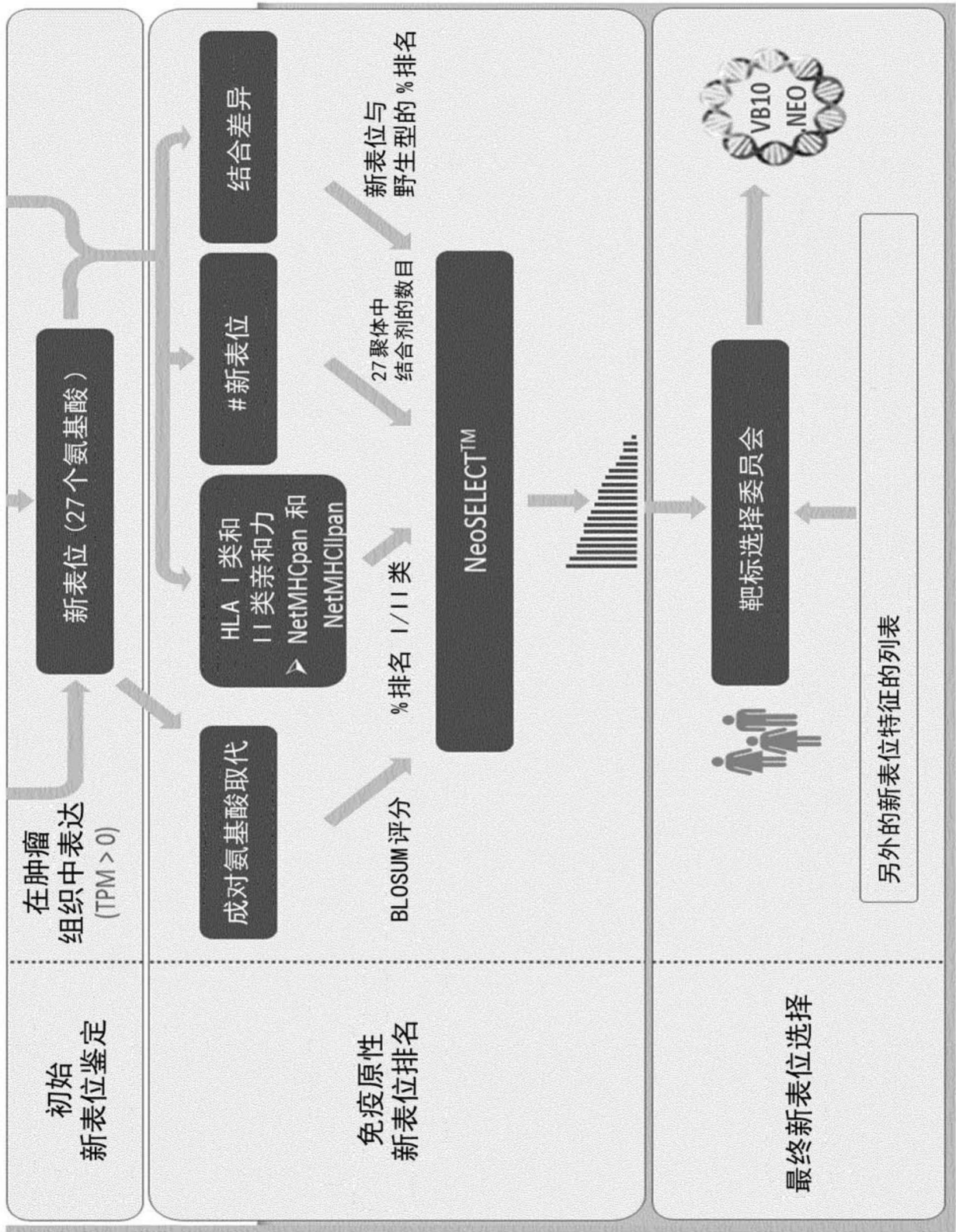


图7 (续)

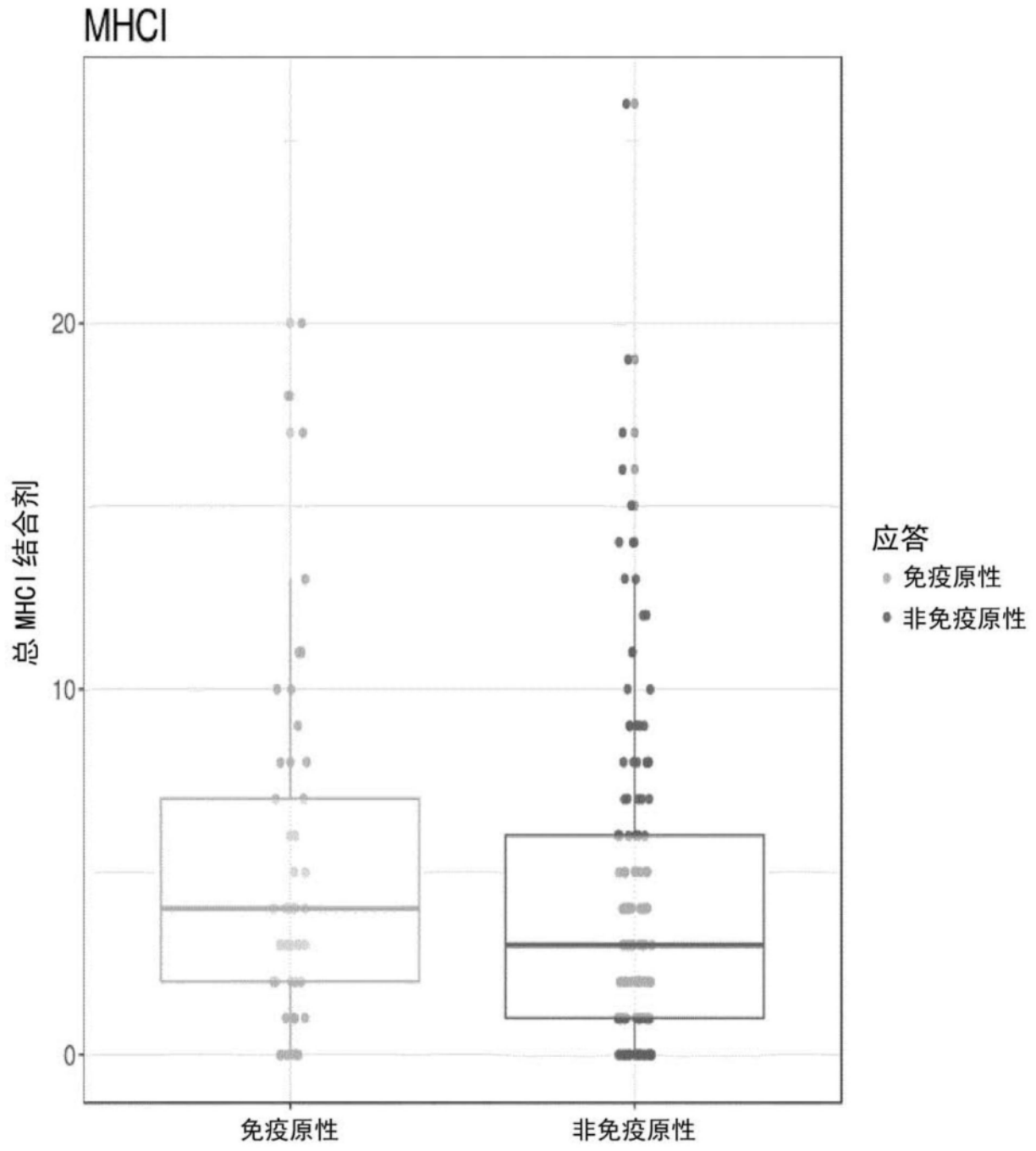


图8

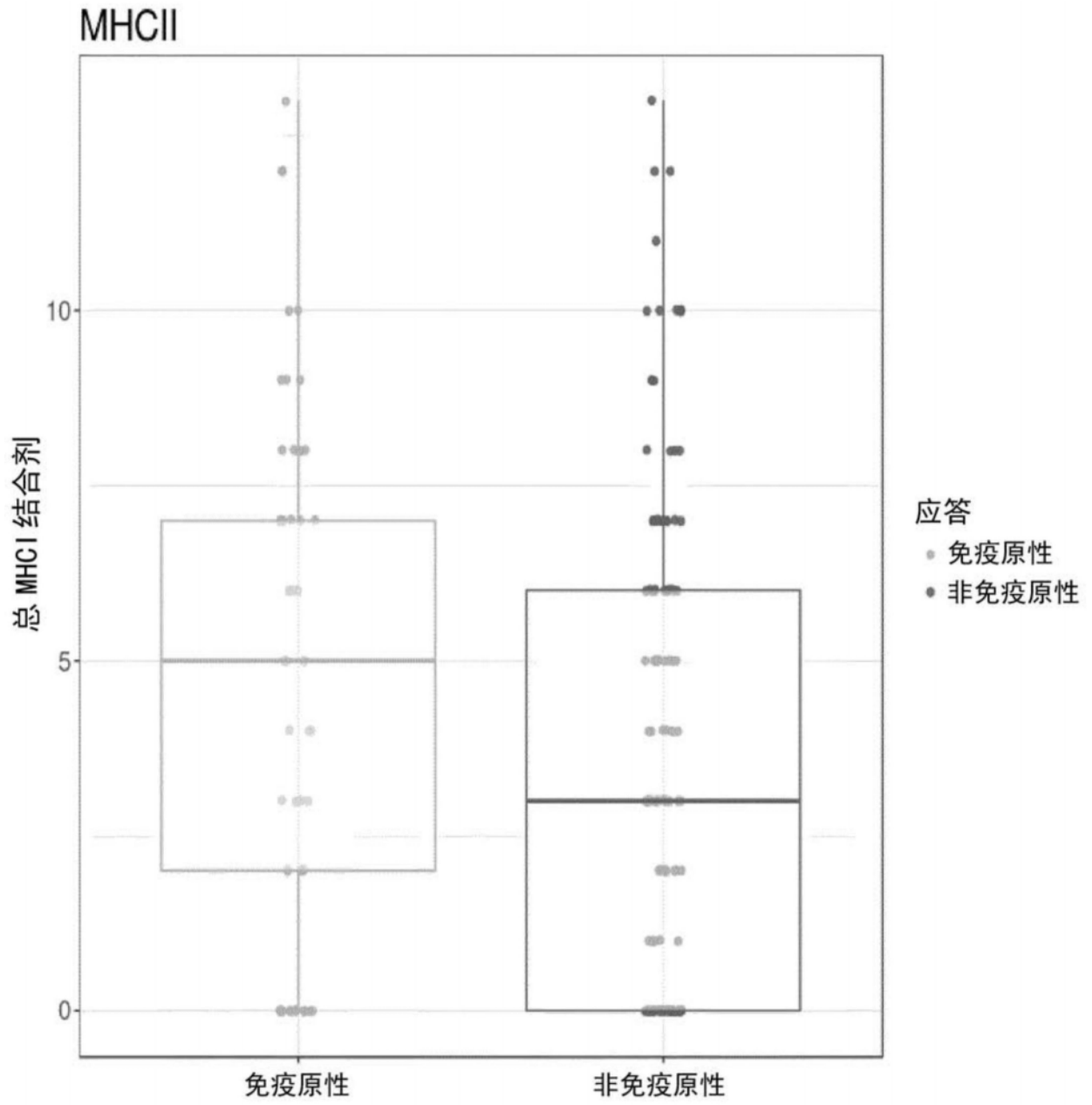


图8(续)

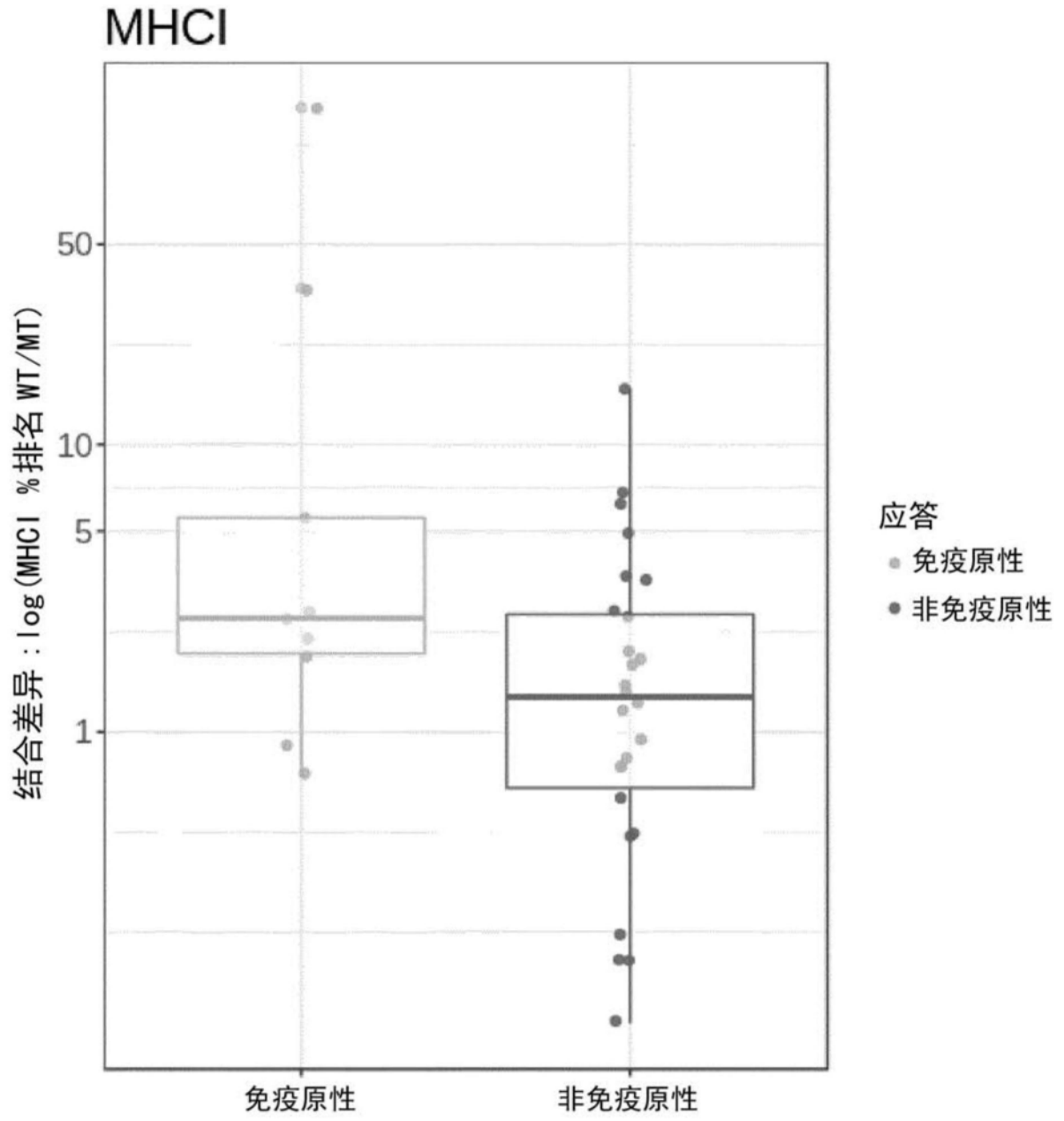


图9



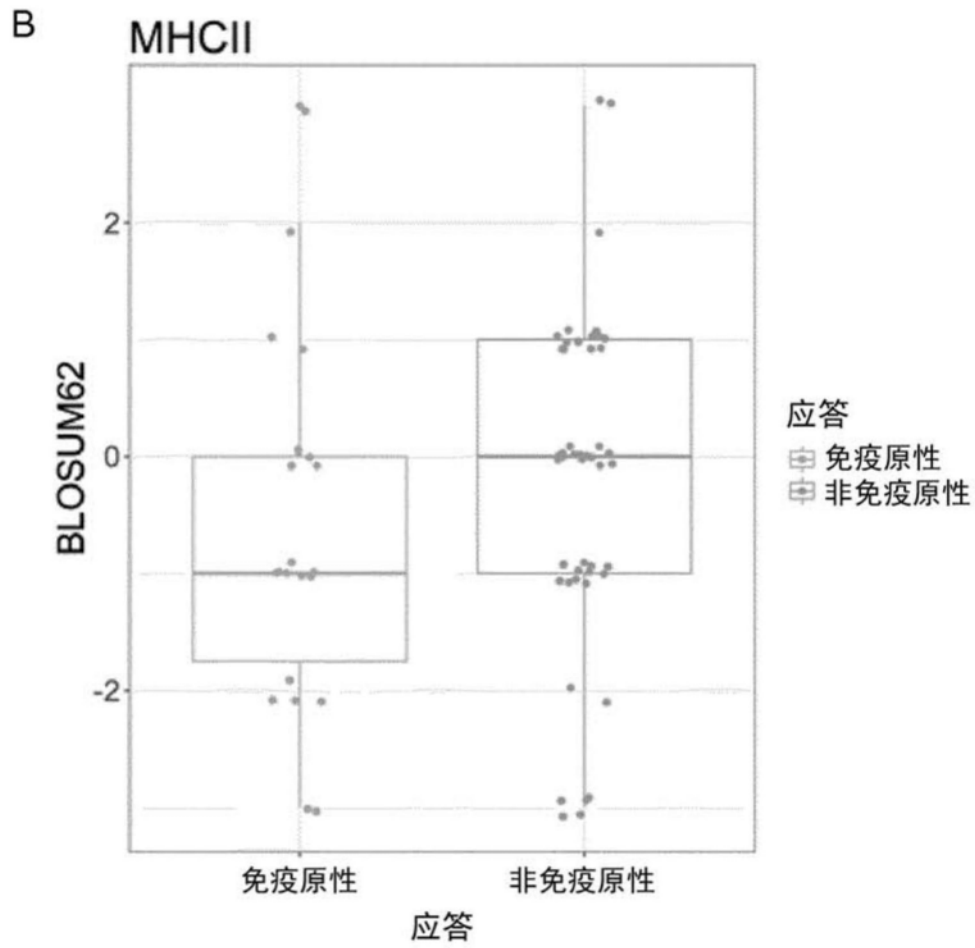


图10(续)