



(12) 发明专利申请

(10) 申请公布号 CN 113791220 A

(43) 申请公布日 2021.12.14

(21) 申请号 202110939479.5

G01N 33/569 (2006.01)

(22) 申请日 2015.09.10

G12Q 1/6886 (2018.01)

(30) 优先权数据

A61K 39/00 (2006.01)

62/048,742 2014.09.10 US

A61P 35/00 (2006.01)

(62) 分案原申请数据

A61P 43/00 (2006.01)

201580059468.0 2015.09.10

A61P 37/04 (2006.01)

(71) 申请人 豪夫迈·罗氏有限公司

地址 瑞士巴塞尔

(72) 发明人 L·德拉马尔 P·卢帕杜斯

I·梅尔曼 M·亚达夫

S·琼琼瓦拉 J·莱尔

(74) 专利代理机构 北京坤瑞律师事务所 11494

代理人 岑晓东

(51) Int. Cl.

G01N 33/68 (2006.01)

权利要求书2页 说明书34页

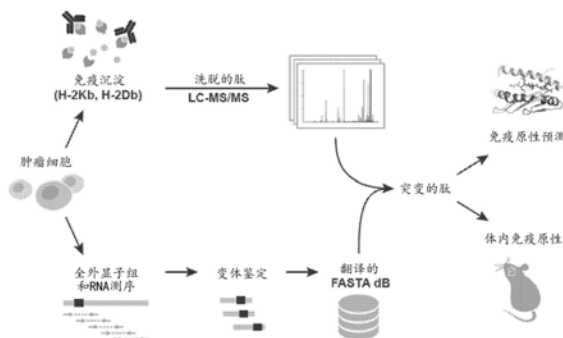
序列列表7页 附图10页

(54) 发明名称

免疫原性突变体肽筛选平台

(57) 摘要

本公开内容涉及免疫原性突变体肽筛选平台。本公开内容提供经由一系列选择步骤鉴定疾病特异性免疫原性肽的方法。通过本公开内容的方法鉴定的免疫原性表位对于在基于肽的免疫疗法,优选癌症疗法中的使用是可应用的。此外,本公开内容的方法可以以高通量方式实施,并且充当个性化疫苗开发和疗法的手段。还提供的是免疫原性肽的组合物以及包含所述组合物的治疗方法。



1. 一种自个体中的疾病组织鉴定疾病特异性免疫原性突变体肽的方法,其包括:
 - a) 提供所述个体中的所述疾病组织的一组变体编码序列,每种变体编码序列与参照样品相比具有所述序列中的变异;并且
 - b) 自该组变体编码序列选择免疫原性变体编码序列,其中所述选择步骤包括预测包含由所述变体编码序列编码的变体氨基酸的肽的免疫原性,由此鉴定所述疾病特异性免疫原性突变体肽。
2. 一种自个体中的疾病组织鉴定疾病特异性免疫原性突变体肽的方法,其包括:
 - a) 基于所述个体中的所述疾病组织的基因组序列获得第一组变体编码序列,每种变体编码序列与参照样品相比具有所述序列中的变异;
 - b) 基于所述个体中的所述疾病组织的转录物组序列自所述第一组选择第二组表达变体编码序列;
 - c) 基于由所述表达变体编码序列编码的肽结合MHC I类分子(MHCI)的预测能力自所述第二组选择第三组表位变体编码序列;
 - d) 自所述第三组选择免疫原性变体编码序列,包括预测包含由所述表位变体编码序列编码的变体氨基酸的肽的免疫原性;由此鉴定所述疾病特异性免疫原性突变体肽。
3. 一种自个体中的疾病组织鉴定疾病特异性免疫原性突变体肽的方法,其包括:
 - a) 自个体的患病组织获得多数与MHC分子结合的肽;
 - b) 将所述MHC结合的肽提交基于质谱术的测序;并
 - c) 将所述MHC结合的肽的质谱术衍生的序列信息与所述个体中的疾病组织的一组变体编码序列关联,每种变体编码序列与参照样品相比具有所述序列中的变异,由此鉴定所述疾病特异性免疫原性突变体肽。
4. 一种自个体中的疾病组织鉴定疾病特异性免疫原性突变体肽的方法,其包括:
 - a) 基于所述个体中的疾病组织的基因组序列获得第一组变体编码序列,每种变体编码序列与参照样品相比具有所述序列中的变异;
 - b) 基于所述个体中的疾病组织的转录物组序列自所述第一组选择第二组表达变体编码序列;
 - c) 基于由所述表达变体编码序列编码的肽结合MHC I类分子(MHCI)的预测能力自所述第二组选择第三组表位变体编码序列;
 - d) 自所述疾病组织获得多数与MHCI分子结合的肽;
 - e) 将所述MHCI结合的肽提交基于质谱术的测序;
 - f) 将所述MHCI结合的肽的质谱术衍生的序列信息与所述第三组表位变体编码序列关联;由此鉴定所述疾病特异性免疫原性突变体肽。
5. 权利要求1-4中任一项的方法,其进一步包括基于所鉴定的疾病特异性免疫原性突变体肽的序列合成肽。
6. 权利要求1-4中任一项的方法,其进一步包括基于所鉴定的疾病特异性免疫原性突变体肽的序列合成编码肽的核酸。
7. 一种疾病特异性免疫原性突变体肽,其通过权利要求5或6的方法获得。

8. 一种组合物,其包含权利要求7的疾病特异性免疫原性突变体肽。
9. 一种治疗个体中的疾病的方法,其包括对所述个体施用有效量的权利要求8的组合物。
10. 一种在具有疾病的个体中刺激免疫应答的方法,其包括施用权利要求8的组合物。

免疫原性突变体肽筛选平台

[0001] 本申请是申请日为2015年09月10日、中国申请号为201580059468.0、发明名称为“免疫原性突变体肽筛选平台”的发明申请的分案申请。

[0002] 对相关申请的交叉引用

[0003] 本申请要求2014年9月10日提交的美国临时申请No.62/048,742的优先权,通过提及将其全部完整收入本文用于所有目的。

[0004] ASCII文本文件上的序列表的提交

[0005] 通过提及将ASCII文本文件上的以下提交的内容整体收入本文:序列表的计算机可读形式(CRF)(文件名:146392027600SeqList.txt,记录日期:2015年9月9日,大小:6KB)。

发明领域

[0006] 本公开内容涉及鉴定可用于开发免疫治疗剂的突变体肽的方法。

[0007] 发明背景

[0008] 牵涉细胞介导的免疫的细胞毒性T淋巴细胞(细胞毒性T细胞或CD8 T细胞)通过在细胞表面上扫描肽表位或抗原监测细胞健康的变化。肽表位源自细胞蛋白,并且充当允许细胞呈现当前细胞过程的证据的显示机制。处理天然的和非天然的蛋白质(通常分别称为自身和非自体)进行肽表位呈递。自天然蛋白质周转和有缺陷的核糖体产物衍生大多数自身肽。可以自事件(诸如病毒性和细菌性感染、疾病和癌症)过程中生成的蛋白质衍生非自身肽。

[0009] 人肿瘤在特征上含有大量的体细胞突变。继而,含有突变的肽的表达可以由适应性免疫系统识别为非自身新表位。在识别非自身抗原时,细胞毒性T细胞会触发免疫应答,导致展示非自身新表位的细胞的凋亡。细胞毒性T细胞免疫应答是适应性免疫系统的高度特异性机制,并且是用于消除感染的、患病的和癌性的细胞的有效手段。在鉴定免疫原性表位中具有较大的治疗价值,因为可以使用经由疫苗接种对免疫原性表位的暴露触发期望的细胞毒性T细胞免疫应答。几十年来,科学和医学界中已经知道了免疫原性表位的作用,但是鉴定驱动有效抗肿瘤CD8 T细胞应答的抗原仍然在很大程度上是未知的。牵涉表位呈现和细胞毒性T细胞活化的复杂性已经使它们的发现和治疗用途陷入困境。

[0010] 主要组织相容性复合体(MHC) I类分子负责肽表位呈递给细胞毒性T细胞。在人中,人白细胞抗原(HLA)系统是编码MHC I类和II类分子的基因的基因座。HLA-A、-B和-C基因编码MHC I类(MHCI)蛋白质。长度通常为8-11个氨基酸的肽经由与由位于反平行β片层上方的两个α螺旋形成的沟的相互作用结合MHCI分子。肽-MHC类I(pMHCI)分子的加工和呈递牵涉一系列序贯阶段,其包括:a)蛋白质酶介导的对蛋白质的消化;b)与抗原加工有关的转运蛋白(TAP)介导的将肽转运入内质网(ER)中;c)使用新合成的MHCI分子形成pMHCI;和d)将pMHCI转运至细胞表面。

[0011] 在细胞表面上,pMHCI会经由T细胞受体(TCR)与细胞毒性T细胞相互作用。在复杂的pMHC I-TCR相互作用后,非自身抗原的鉴定可以经由由关联的酶、共受体、衔接头分子、和转录因子介导的一系列生物化学事件而导致细胞毒性T细胞活化。活化的细胞毒性T细胞

将会增殖以生成表达对鉴定的免疫原性肽表位特异性的TCR的效应T细胞群体。对鉴定的非自身表位具有TCR特异性的T细胞的扩增导致展示活化的非自身表位的细胞的免疫介导的凋亡。

[0012] 目前正在研究使用免疫原性表位激活免疫系统以用于癌症疗法。虽然癌细胞源自正常组织,但是体细胞突变会驱动癌症蛋白酶体的大量变化。继而,所得的MHCI呈递的肽表位,称为肿瘤相关抗原(TAA)或新表位,允许正常和癌组织之间的细胞毒性T细胞分化。最近的工作已经证实,突变体肽可以充当由CD4或CD8T细胞识别为非自身的表位,但是从此以后已经很少描述突变体新表位。

[0013] 基于肽的免疫疗法的使用取决于会刺激期望的细胞毒性T细胞应答的肽表位的选择。具体地,肿瘤抗原可以分成两类:肿瘤相关自身抗原(例如癌-睾丸抗原,分化抗原)和自共享的或患者特异性的突变体蛋白衍生的抗原。由于胸腺中自身抗原的呈现可以导致高亲和力T细胞的消除,突变体新抗原可能更具免疫原性。此类免疫治疗性表位的开发是一项有挑战性的追求,并且可用于鉴定有效表位的有效方法尚有待开发。

[0014] 在免疫原性肽表位的鉴定和验证中牵涉的时间和成本密集性性质妨碍了有效的基于肽的癌症疫苗的开发。为了使鉴定免疫原性表位中牵涉的问题进一步复杂化,癌细胞中的突变的排列经常是患者特异性的。突变体新表位的发现需要对患者的肿瘤浸润淋巴细胞费力筛选它们识别抗原的能力,所述抗原来自基于来自所述患者的肿瘤外显子组序列的信息构建的文库。或者,可以通过质谱术检测突变体新表位。然而,突变序列已经逃避检测,因为使用不含有患者特异性突变的公共蛋白质组数据库不允许其鉴定。使用诸如肽-MHC I结合或肽免疫原性的预测算法可以在鉴定个性化免疫原性表位中具有潜在的应用。但是,癌细胞中含有的大量体细胞突变和表达水平变化导致预测的免疫原性表位的量级对于高通量免疫原性筛选而言太大。此外,预测表位的较差的免疫原性的证据使当前方法的效用受到质疑。

[0015] 本领域中需要鉴定适合于用于基于肽的免疫疗法的免疫原性表位。具体地,本领域中需要鉴定用于基于肽的癌症疗法中的免疫原性表位。此外,本领域中需要用于基于个性化遗传和/或蛋白质组学分析预测免疫原性表位的高通量方法。

[0016] 在此通过提及收录本文中引用的所有参考文献。

[0017] 发明概述

[0018] 在一方面,本申请提供了鉴定疾病特异性免疫原性突变体肽的方法,其包括:a)提供个体中的疾病组织的变体编码序列组,每种变体编码序列与参照样品相比具有序列中的变异;并且b)自所述变体编码序列组选择免疫原性变体编码序列,其中选择步骤包括预测包含由所述变体编码序列编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,所述方法包括a)基于个体中的疾病组织的基因组序列获得变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b)基于个体中的疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c)基于由表达变体编码序列编码的肽结合MHC I类分子(MHCI)的预测能力,自第二组选择表位变体编码序列的第三组;并且d)自第三组选择免疫原性变体编码序列,包括预测包含由表位变体编码序列编码的变体氨基酸的肽的免疫原性;由此鉴定疾病特异性免疫原性突变体肽。

[0019] 在一些实施方案中,依照上文描述的任何方法,方法还包括i)自疾病组织获得多

个与MHCI分子结合的肽;ii)使MHCI结合肽受到基于质谱术的测序;iii)将MHCI结合肽的质谱术衍生的序列信息与免疫原性变体编码序列关联。

[0020] 在另一方面,本申请提供了鉴定疾病特异性免疫原性突变体肽的方法,其包括:a)自个体的患病组织获得与MHC分子结合的多个肽;b)使MHC结合的肽经受基于质谱术的测序;并且c)将MHC结合的肽的质谱术衍生的序列信息与个体中的疾病组织的变体编码序列组关联,每种变体编码序列与参照样品相比具有序列中的变异,由此鉴定疾病特异性免疫原性突变体肽。

[0021] 在一些实施方案中,所述方法包括a)基于个体中的疾病组织的基因组序列获得变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b)基于个体中的疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c)基于由表达变体编码序列编码的肽结合MHC I类分子(MHCI)的预测能力,自第二组选择表位变体编码序列的第三组;d)自疾病组织获得与MHCI分子结合的多个肽;e)使MHCI结合肽经受基于质谱术的测序;f)将MHCI结合肽的质谱术衍生的序列信息与表位变体编码序列的第三组关联;由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法还包括预测疾病特异性免疫原性突变体肽的免疫原性,其中疾病特异性免疫原性突变体肽包含变体氨基酸。

[0022] 依照包括预测免疫原性的步骤的任何上文描述的方法,预测免疫原性基于下列一项或多项参数:i)肽对MHCI分子的结合亲和力;ii)含有肽的肽前体的蛋白质水平;iii)编码肽前体的转录物的表达水平;iv)通过免疫蛋白酶体对肽前体的加工效率;v)编码肽前体的转录物的表达时机;vi)肽对TCR分子的结合亲和力;vii)变体氨基酸在肽内的位置;xiii)在与MHCI分子结合时肽的溶剂暴露;ix)在与MHCI分子结合时变体氨基酸的溶剂暴露;和x)肽中芳香族残基的含量;xi)在与野生型残基相比时变体氨基酸的特性;xii)肽前体的性质。在一些实施方案中,预测免疫原性进一步基于HLA分型分析。

[0023] 在一些实施方案中,依照包括自疾病组织获得与MHCI分子结合的多个肽的步骤的上文描述的方法之任一种,通过自疾病组织分离MHCI/肽复合物并且自MHCI洗脱肽获得与MHCI结合的肽。在一些实施方案中,通过免疫沉淀实施MHCI/肽复合物的分离。在一些实施方案中,使用对MHCI特异的抗体实施免疫沉淀。在一些实施方案中,在进行质谱术分析前通过层析进一步分开所分离的肽。

[0024] 在一些实施方案中,依照包括获得变体编码序列的第一组的步骤的上文描述的方法之任一种,获得变体编码序列的第一组包括i)基于个体中的疾病组织的基因组序列获得变体序列的第一组,每种变体序列与参照样品相比具有序列中的变异;并且ii)自变体序列的第一组鉴定变体编码序列。

[0025] 在一些实施方案中,依照上文描述的任何方法,其中方法还包括基于鉴定的疾病特异性免疫原性突变体肽的序列合成肽。在一些实施方案中,依照上文描述的任何方法,方法还包括基于鉴定的疾病特异性免疫原性突变体肽的序列合成编码肽的核酸。在一些实施方案中,方法还包括在体内对合成的肽测试免疫原性。

[0026] 在一些实施方案中,依照上文描述的任何方法,疾病是癌症。在一些实施方案中,依照上文描述的任何方法,个体是人。

[0027] 在另一方面,本申请还提供了通过本文中描述的任何方法鉴定的疾病特异性突变体肽或疾病特异性突变体肽的组合物。在一些实施方案中,组合物包含两种或更多种本文

中描述的疾病特异性免疫原性突变体肽。在一些实施方案中,组合物还包含佐剂。

[0028] 在又一方面,本申请还提供了治疗个体中的疾病的方法,其包括对个体施用有效量的组合物,所述组合物包含用本文中公开的用于鉴定疾病特异性免疫原性突变体肽的任何方法鉴定的疾病特异性突变体肽。在一些实施方案中,个体是鉴定疾病特异性免疫原性突变体肽的同一个体。

[0029] 本申请还提供了免疫原性组合物,其包含至少一种疾病特异性肽或此类疾病特异性肽的前体,其中通过本文中描述的任何方法鉴定所述疾病特异性肽。在一些实施方案中,免疫原性组合物包含多种疾病特异性肽。

[0030] 本申请还提供了包含至少一种编码疾病特异性肽的核酸的免疫原性组合物,其中通过本文中描述的任何方法鉴定所述疾病特异性肽。在一些实施方案中,免疫原性组合物包含多种各自编码至少一种疾病特异性肽的核酸。在一些实施方案中,免疫原性组合物包含编码两种或更多种(诸如3、4、5、6、7、8、9或更多种)疾病特异性肽的核酸。

[0031] 在又一方面,本申请还提供了在具有疾病的个体中刺激免疫应答的方法,其包括施用本文中描述的任何免疫原性组合物。在一些实施方案中,方法还包括施用另一种药剂。在一些实施方案中,另一种药剂是免疫调节剂。在一些实施方案中,另一种药剂是检查点蛋白。在一些实施方案中,其它药剂是PD-1的拮抗剂(诸如抗PD1抗体)。在一些实施方案中,其它药剂是PD-L1的拮抗剂(诸如抗PD-L1抗体)。

[0032] 在另一方面,本申请还提供了在具有疾病的个体中刺激免疫应答的方法,其包括:
a) 依照上文描述的鉴定方法之任一种自个体中的疾病组织鉴定疾病特异性免疫原性突变体肽,b) 基于鉴定的疾病特异性免疫原性突变体肽的序列生成包含肽或编码肽的核酸的组合物;c) 对个体施用组合物。在一些实施方案中,方法还包括对个体施用PD-1拮抗剂(诸如抗PD1抗体)。在一些实施方案中,方法还包括对个体施用PD-L1拮抗剂(诸如抗PD-L1抗体)。

[0033] 本申请涉及下述实施方案。

[0034] 1. 一种自个体中的疾病组织鉴定疾病特异性免疫原性突变体肽的方法,其包括:

[0035] a) 提供所述个体中的所述疾病组织的一组变体编码序列,每种变体编码序列与参照样品相比具有所述序列中的变异;并且

[0036] b) 自该组变体编码序列选择免疫原性变体编码序列,其中所述选择步骤包括预测包含由所述变体编码序列编码的变体氨基酸的肽的免疫原性,

[0037] 由此鉴定所述疾病特异性免疫原性突变体肽。

[0038] 2. 一种自个体中的疾病组织鉴定疾病特异性免疫原性突变体肽的方法,其包括:

[0039] a) 基于所述个体中的所述疾病组织的基因组序列获得第一组变体编码序列,每种变体编码序列与参照样品相比具有所述序列中的变异;

[0040] b) 基于所述个体中的所述疾病组织的转录物组序列自所述第一组选择第二组表达变体编码序列;

[0041] c) 基于由所述表达变体编码序列编码的肽结合MHC I类分子(MHCI)的预测能力自所述第二组选择第三组表位变体编码序列;

[0042] d) 自所述第三组选择免疫原性变体编码序列,包括预测包含由所述表位变体编码序列编码的变体氨基酸的肽的免疫原性;

[0043] 由此鉴定所述疾病特异性免疫原性突变体肽。

- [0044] 3. 实施方案1或2的方法,其中所述方法进一步包括:
- [0045] i) 自所述疾病组织获得多数与MHCI分子结合的肽;
- [0046] ii) 将所述MHCI结合的肽提交基于质谱术的测序;
- [0047] iii) 将所述MHCI结合的肽的质谱术衍生的序列信息与所述免疫原性变体编码序列关联。
- [0048] 4. 一种自个体中的疾病组织鉴定疾病特异性免疫原性突变体肽的方法,其包括:
- [0049] a) 自个体的患病组织获得多数与MHC分子结合的肽;
- [0050] b) 将所述MHC结合的肽提交基于质谱术的测序;并
- [0051] c) 将所述MHC结合的肽的质谱术衍生的序列信息与所述个体中的疾病组织的一组变体编码序列关联,每种变体编码序列与参照样品相比具有所述序列中的变异,
- [0052] 由此鉴定所述疾病特异性免疫原性突变体肽。
- [0053] 5. 一种自个体中的疾病组织鉴定疾病特异性免疫原性突变体肽的方法,其包括:
- [0054] a) 基于所述个体中的疾病组织的基因组序列获得第一组变体编码序列,每种变体编码序列与参照样品相比具有所述序列中的变异;
- [0055] b) 基于所述个体中的疾病组织的转录物组序列自所述第一组选择第二组表达变体编码序列;
- [0056] c) 基于由所述表达变体编码序列编码的肽结合MHC I类分子(MHCI)的预测能力自所述第二组选择第三组表位变体编码序列;
- [0057] d) 自所述疾病组织获得多数与MHCI分子结合的肽;
- [0058] e) 将所述MHCI结合的肽提交基于质谱术的测序;
- [0059] f) 将所述MHCI结合的肽的质谱术衍生的序列信息与所述第三组表位变体编码序列关联;
- [0060] 由此鉴定所述疾病特异性免疫原性突变体肽。
- [0061] 6. 实施方案4或5的方法,其进一步包括预测所述疾病特异性免疫原性突变体肽的免疫原性,其中所述疾病特异性免疫原性突变体肽包含变体氨基酸。
- [0062] 7. 实施方案1-3和6中任一项的方法,其中预测免疫原性基于下述参数中的一项或多项:
- [0063] i) 所述肽对所述MHCI分子的结合亲和力;
- [0064] ii) 含有所述肽的肽前体的蛋白质水平;
- [0065] iii) 编码所述肽前体的转录物的表达水平;
- [0066] iv) 免疫蛋白酶体对所述肽前体的加工效率;
- [0067] v) 所述肽前体的表达时机;
- [0068] vi) 所述肽对TCR分子的结合亲和力;
- [0069] vii) 变体氨基酸在所述肽内的位置;
- [0070] viii) 在与MHCI分子结合时所述肽的溶剂暴露;
- [0071] ix) 在与MHCI分子结合时所述变体氨基酸的溶剂暴露;和
- [0072] x) 所述肽中芳香族残基的含量;
- [0073] xi) 在与野生型残基相比时所述变体氨基酸的特性;
- [0074] xii) 所述肽前体的性质。

- [0075] 8. 实施方案7的方法,其中所述免疫原性的预测进一步包括HLA分型分析。
- [0076] 9. 实施方案3-8中任一项的方法,其中通过自所述疾病组织分离MHCI/肽复合物并自所述MHCI洗脱所述肽获得所述与MHCI结合的肽。
- [0077] 10. 实施方案9的方法,其中通过免疫沉淀进行所述分离。
- [0078] 11. 实施方案10的方法,其中使用对MHCI特异性的抗体进行所述免疫沉淀。
- [0079] 12. 实施方案9-11中任一项的方法,其中在提交质谱术分析前通过层析进一步分开所述肽。
- [0080] 13. 实施方案2、3和5-12中任一项的方法,其中获得第一组变体编码序列包括:
- [0081] i) 基于所述个体中的疾病组织的基因组序列获得第一组变体序列,每种变体序列与参照样品相比具有所述序列中的变异;并
- [0082] ii) 自所述第一组变体序列鉴定所述变体编码序列。
- [0083] 14. 实施方案1-13中任一项的方法,其进一步包括基于所鉴定的疾病特异性免疫原性突变体肽的序列合成肽。
- [0084] 15. 实施方案1-13中任一项的方法,其进一步包括基于所鉴定的疾病特异性免疫原性突变体肽的序列合成编码肽的核酸。
- [0085] 16. 实施方案1-15中任一项的方法,其进一步包括在体内对所述肽测试免疫原性。
- [0086] 17. 实施方案1-15中任一项的方法,其进一步包括在体外对所述肽测试免疫原性。
- [0087] 18. 实施方案1-17中任一项的方法,其中所述疾病是癌症。
- [0088] 19. 实施方案1-18中任一项的方法,其中所述个体是人。
- [0089] 20. 一种疾病特异性免疫原性突变体肽,其通过实施方案14或15的方法获得。
- [0090] 21. 一种组合物,其包含实施方案20的疾病特异性免疫原性突变体肽。
- [0091] 22. 实施方案21的组合物,其中所述组合物包含两种或更多种疾病特异性免疫原性突变体肽。
- [0092] 23. 实施方案21或22的组合物,其进一步包含佐剂。
- [0093] 24. 一种治疗个体中的疾病的方法,其包括对所述个体施用有效量的实施方案21-23中任一项的组合物。
- [0094] 25. 实施方案24的方法,其中所述个体与自其鉴定所述疾病特异性免疫原性突变体肽的个体是同一个体。
- [0095] 26. 一种免疫原性组合物,其包含至少一种疾病特异性肽,其中所述疾病特异性肽是通过实施方案1-19中任一项的方法鉴定的。
- [0096] 27. 一种免疫原性组合物,其包含至少一种编码疾病特异性肽的核酸,其中所述疾病特异性肽是通过实施方案1-19中任一项的方法鉴定的。
- [0097] 28. 一种免疫原性组合物,其包含多数疾病特异性肽,其中所述疾病特异性肽是通过实施方案1-19中任一项的方法鉴定的。
- [0098] 29. 一种免疫原性组合物,其包含编码多数疾病特异性肽的核酸,其中所述疾病特异性肽是通过实施方案1-19中任一项的方法鉴定的。
- [0099] 30. 一种在具有疾病的个体中刺激免疫应答的方法,其包括施用实施方案21-23中任一项的组合物。
- [0100] 31. 实施方案30的方法,其进一步包括施用另一种药剂。

- [0101] 32. 实施方案31的方法,其中所述其它药剂是免疫调节剂。
- [0102] 33. 实施方案31的方法,其中所述其它药剂是检查点蛋白的拮抗剂。
- [0103] 34. 实施方案33的方法,其中所述其它药剂是PD-1的拮抗剂。
- [0104] 35. 实施方案33的方法,其中所述其它药剂是PD-L1的拮抗剂。
- [0105] 36. 一种在具有疾病的个体中刺激免疫应答的方法,其包括:
- [0106] a) 依照实施方案1-19中任一项的方法自所述个体中的疾病组织鉴定疾病特异性免疫原性突变体肽,
- [0107] b) 基于所鉴定的疾病特异性免疫原性突变体肽的序列生成包含肽或编码所述肽的核酸的组合物;
- [0108] c) 对所述个体施用所述组合物。
- [0109] 37. 实施方案36的方法,其进一步包括对所述个体施用抗PD-1抗体。
- [0110] 38. 实施方案36的方法,其进一步包括对所述个体施用抗PD-L1抗体。
- [0111] 附图简述
- [0112] 图1显示了免疫原性肽鉴定的例示性方法。
- [0113] 图2小图A显示了相对于测量的每百万个定位读段每千碱基外显子模型的读段(RPKM),鉴定为MC-38细胞系的MHC分子上呈递的表位的鉴定基因的分布。
- [0114] 图2小图B显示了相对于测量的RPKM,鉴定为TRAMP-C1细胞系的MHC分子上呈递的表位的鉴定基因的分布。
- [0115] 图3显示了与MHC分子结合的肽的结构建模。
- [0116] 图4小图A显示了用选择肽免疫的野生型C57BL/6小鼠中肽特异性CD8 T细胞的百分比。
- [0117] 图4小图B显示了脾和肿瘤中Dextramer阳性CD8T细胞的百分比。
- [0118] 图4小图C显示了相对于肿瘤体积的CD8 T细胞和CD45 T细胞的测量。
- [0119] 图4小图D显示了在总CD8 TIL群体和Adpgk阳性CD8 TIL群体中共表达PD-1和TIM-3的肿瘤特异性CD8 TIL的百分比。
- [0120] 图5小图A显示了用MC-38肿瘤细胞的肿瘤攻击后用对照和免疫原性疫苗处理的小鼠的肿瘤体积,和疫苗接种后Adpgk阳性CD8T细胞的百分比。箭标示来自单一动物的测量。
- [0121] 图5小图B显示脾和肿瘤中肽特异性CD8 T细胞的百分比。
- [0122] 图5小图C显示了以表达CD45的T细胞和表达CD8的T细胞测量的肿瘤中活细胞的百分比。
- [0123] 图5小图D显示了疫苗接种后总CD8 T细胞群体中共表达PD-1和TIM-3的Adgpk特异性CD8 TIL的百分比。
- [0124] 图5小图E显示了疫苗接种后PD-1和TIM-3表面表达水平。
- [0125] 图5小图F显示了疫苗接种后肿瘤和脾脏中表达IFN- γ 的CD8和CD4 TIL的百分比。
- [0126] 图5小图G显示了疫苗接种后肿瘤体积的测量。
- [0127] 发明详述
- [0128] 本申请提供了用于鉴定疾病特异性免疫原性突变体肽的高效率筛选平台。通过将基于序列的变体鉴定方法与免疫原性预测和/或质谱术组合,本文中描述的方法允许自个体的疾病组织(诸如肿瘤细胞)有力且有效鉴定疾病特异性免疫原性突变体肽。这些肽或基

于核苷酸的前体(例如DNA或RNA)可用于多种不同的应用,诸如疫苗的开发、突变体肽特异性治疗剂(诸如抗体治疗剂或基于T细胞受体(“TCR”)的治疗剂)的开发、以及用于监测T细胞应答的动力学和分布的工具的开发。例如,可以利用个别的肽或肽组进行比较结合亲和力测量或将个别的肽或肽组多聚化以通过MHC多聚体流式细胞术测量抗原特异性T细胞应答。本文中描述的方法在个性化医学的背景中是特别有用的,其中可以使用自患病个体鉴定的突变体肽开发用于治疗相同个体的治疗剂(例如基于肽、DNA、或RNA的疫苗)。

[0129] 如此,在一方面,本申请提供了自个体的疾病组织鉴定疾病特异性免疫原性突变体肽的方法,其通过组合基于序列的变体鉴定方法与免疫原性预测进行。

[0130] 在另一方面,本申请提供了自个体的患病组织鉴定疾病特异性免疫原性突变体肽的方法,其通过将基于序列的变体鉴定方法与质谱术组合进行。

[0131] 还提供了可用于本文中描述方法的试剂盒和系统。进一步包括免疫原性组合物,其包含肽、呈递此类肽的细胞、和编码鉴定的此类肽的核酸。

[0132] 定义

[0133] 如本公开内容中使用的,单数形式“一个”、“一种”和“所述/该”具体还涵盖它们所指的术语的复数形式,除非上下文另有明确规定。本文中提及“约”数值或参数包括(并且描述)针对所述数值或参数本身的变化。例如,提及“约X”的描述包括“X”的描述。

[0134] 应当理解,本文中描述的本发明的方面和实施方案包括“由方面和实施方案组成”和/或“基本上由方面和实施方案组成”。

[0135] 如本文中使用的,“疾病特异性突变体肽”是指包含存在于疾病组织中但不在正常组织中的至少一种突变氨基酸的肽。“疾病特异性免疫原性突变体肽”是指能够在个体中引发免疫应答的疾病特异性突变体肽。疾病特异性突变体肽可以源自例如:导致蛋白质中不同氨基酸的非同义突变(例如点突变);连读突变,其中修饰或删除终止密码子,导致在C端具有新的肿瘤特异性序列的较长蛋白质的翻译;剪接位点突变,其导致成熟mRNA中的内含子纳入,并且如此产生独特的肿瘤特异性蛋白质序列;染色体重排,其产生在2种蛋白质的接合处具有肿瘤特异性序列的嵌合蛋白(即基因融合);和移码突变或删除,其导致具有新肿瘤特异性蛋白质序列的新可读框。参见例如Sensi和Anichini, Clin Cancer Res, 2006, v. 12, 5023-5032。

[0136] 如本文中使用的,“变体编码序列”是指与参照样品中的序列相比具有变异的序列,其中序列变异导致变体编码序列中含有的或由变体编码序列编码的氨基酸序列的变化。变体编码序列可以是具有导致编码氨基酸序列中的氨基酸变化的突变的核酸序列。或者,变体编码序列可以是含有氨基酸突变的氨基酸序列。

[0137] “表达变体编码序列”是指在个体的疾病组织中表达的变体编码序列。

[0138] “编码”肽的核酸序列是指含有肽的编码序列的核酸。“编码”肽的氨基酸序列是指含有肽序列的氨基酸序列。

[0139] “表位变体编码序列”是指编码结合或预测结合MHC分子(诸如MHC I类分子或MHCI)的肽的变体编码序列。

[0140] “免疫原性变体编码序列”是指编码预测为免疫原性的肽的变体编码序列。

[0141] 如本文中使用的,术语“疾病组织”是指与个体中的疾病相关的组织,并且包括多个细胞。“疾病组织样品”是指疾病组织的样品。

[0142] 本文中使用的“肽前体”是指存在于包含感兴趣肽的个体的疾病组织中的多肽。例如，肽前体可以是存在于疾病组织中的多肽，其可以通过免疫蛋白酶体加工以生成感兴趣的肽。

[0143] 鉴定免疫原性突变体肽的方法

[0144] 在一方面，本申请的方法将序列特异性变体鉴定方法与免疫原性预测方法组合。例如，在一些实施方案中，提供了自个体中的疾病组织鉴定疾病特异性免疫原性突变体肽的方法，其包括：a) 提供个体中的疾病组织的变体编码序列的第一组，每种变体编码序列与参照样品相比具有序列中的变异；并且b) 自变体编码序列的第一组选择免疫原性变体编码序列，其中选择步骤包括预测包含由变体编码序列编码的变体氨基酸的肽的免疫原性，由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中，提供了鉴定来自个体的疾病组织的疾病特异性免疫原性突变体肽的方法，所述疾病特异性免疫原性突变体肽充当疾病组织中的新表位。在一些实施方案中，变体编码序列组包含超过1、10、100、1000或10,000种不同变体编码序列。在一些实施方案中，提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法，其包括：a) 获得个体中的疾病组织的变体编码序列的第一组，每种变体编码序列与参照样品相比具有序列中的变异；并且b) 自变体编码序列的第一组选择免疫原性变体编码序列，其中选择步骤包括预测包含由变体编码序列编码的变体氨基酸的肽的免疫原性，由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中，选择步骤包括基于一项或多项（例如2、3、4、5、6、7、8、9、10或11之任一种）参数预测肽的免疫原性：i) 肽对MHC I分子的结合亲和力；ii) 含有肽的肽前体的蛋白质水平；iii) 编码肽前体的转录物的表达水平；iv) 通过免疫蛋白酶体对肽前体的加工效率；v) 编码肽前体的转录物的表达时机；vi) 肽对TCR分子的结合亲和力；vii) 变体氨基酸在肽内的位置；xiii) 在与MHC I分子结合时肽的溶剂暴露；ix) 在与MHC I分子结合时变体氨基酸的溶剂暴露；和x) 肽中芳香族残基的含量；xi) 在与野生型残基相比时变体氨基酸的特性（例如自带电荷变化至疏水性或者反之亦然）；xii) 肽前体的性质。

[0145] 在一些实施方案中，可以首先过滤变体编码序列的第一组以获得编码预测结合MHC分子的肽的变体编码序列（称为“表位变体编码序列”）的较小组，然后基于免疫原性的预测对变体编码序列的较小组进行选择。在此类实施方案中，方法可以包括：a) 提供个体中的疾病组织的变体编码序列的第一组，每种变体编码序列与参照样品相比具有序列中的变异；b) 基于由变体编码序列的第一组编码的肽结合MHC分子（诸如MHC I类分子或MHC I）的预测能力，自第一组选择表位变体编码序列的第二组，并且c) 自表位变体编码序列的第二组选择免疫原性变体编码序列，其中选择步骤包括预测包含由表位变体编码序列编码的变体氨基酸的肽的免疫原性，由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中，方法包括：a) 获得个体中疾病组织的变体编码序列的第一组，每种变体编码序列与参照样品相比具有序列中的变异；b) 基于由变体编码序列的第一组编码的肽结合MHC分子（诸如MHC I类分子或MHC I）的预测能力，自第一组选择表位变体编码序列的第二组，并且c) 自表位变体编码序列的第二组选择免疫原性变体编码序列，其中选择步骤包括预测包含由表位变体编码序列编码的变体氨基酸的肽的免疫原性，由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中，方法还包括通过功能性分析验证疾病特异性免疫原性突变体肽。在一些实施方案中，疾病是癌症。在一些实施方案中，个体是人个体（诸如具有癌症的人个体）。

[0146] 在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a) 基于个体中的疾病组织的基因组序列获得个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;并且b) 自变体编码序列的第一组选择免疫原性变体编码序列,其中选择步骤包括预测包含由变体编码序列编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,通过全基因组测序获得基因组序列。在一些实施方案中,通过全外显子组测序获得基因组序列。在一些实施方案中,通过靶向基因组或外显子组测序获得基因组序列。例如,可以首先通过探针组(例如对疾病关联基因特异性的探针)富集疾病组织和/或参照样品中的基因组序列,之后加工以进行变体鉴定。在一些实施方案中,可以首先过滤变体编码序列的第一组以获得表位变体编码序列的较小组,然后基于免疫原性的预测对变体编码序列的较小组进行选择。例如,在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a) 基于个体中的疾病组织的基因组序列获得个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b) 基于由变体编码序列的第一组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第一组选择表位变体编码序列的第二组,并且c) 自表位变体编码序列的第二组选择免疫原性变体编码序列,其中选择步骤包括预测包含由表位变体编码序列编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法还包括通过功能性分析验证疾病特异性免疫原性突变体肽。在一些实施方案中,该疾病是癌症。在一些实施方案中,个体是人个体(诸如具有癌症的人个体)。

[0147] 在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a) 基于个体中的疾病组织的转录物组序列获得个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;并且b) 自变体编码序列的第一组选择免疫原性变体编码序列,其中选择步骤包括预测包含由变体编码序列编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,通过全转录物组RNA-Seq测序获得转录物组序列。在一些实施方案中,通过靶向转录物组测序获得转录序列。例如,可以首先通过探针组(例如对疾病关联基因特异性的探针)富集疾病组织和/或参照样品中的RNA或cDNA序列,然后加工以进行变异鉴定。在一些实施方案中,可以首先过滤变体编码序列的第一组以获得表位变体编码序列的较小组,然后对变体编码序列的较小组进行免疫原性的预测。例如,在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a) 基于个体中的疾病组织的转录物组序列获得个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b) 基于由表位变体编码序列的第一组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第一组选择表位变体编码序列的第二组,并且c) 自变体编码序列的第二组选择免疫原性变体编码序列,其中选择步骤包括预测包含由表位变体编码序列编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。

[0148] 在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a) 基于个体中的疾病组织的基因组序列获得个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b) 基于个

体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;并且c)自表达变体编码序列的第二组选择免疫原性变体编码序列,其中选择步骤包括预测包含由表达变体编码序列编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法包括:a)基于个体中疾病组织的基因组序列获得个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b)基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;并且c)自表达变体编码序列的第二组选择免疫原性变体编码序列,其中选择步骤包括预测包含由表达变体编码序列编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。

[0149] 在一些实施方案中,可以过滤表达变体编码序列的第二组以获得表位变体编码序列的较小组,然后对变体编码序列的较小组进行免疫原性的预测。如此,例如,在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a)基于个体中疾病组织的基因组序列,提供个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b)基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c)基于由表达变体编码序列的第二组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第二组选择表位变体编码序列的第三组;并且d)自表位变体编码序列的第三组选择免疫原性变体编码序列,其中选择步骤包括预测包含由表位变体编码序列编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法包括:a)基于个体中疾病组织的基因组序列获得个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b)基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c)基于由表达变体编码序列的第二组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第二组选择表位变体编码序列的第三组,并且d)自表位变体编码序列的第三组选择免疫原性变体编码序列,其中选择步骤包括预测包含由表位变体编码序列编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法还包括通过功能性分析验证疾病特异性免疫原性突变体肽。在一些实施方案中,疾病是癌症。在一些实施方案中,个体是人个体(诸如具有癌症的人个体)。

[0150] 在一些实施方案中,通过将变体编码序列信息与同MHC分子物理结合的肽的信息关联进一步验证通过本文中描述的方法鉴定的疾病特异性免疫原性突变体肽。例如,方法可以还包括:自疾病组织获得多个与MHC分子结合的肽;对MHC结合的肽进行基于质谱术的测序;并将MHC结合的肽的质谱术衍生的序列信息与预测为免疫原性变体编码序列的肽关联。在下文部分中进一步描述质谱术和关联方法。

[0151] 另一方面,提供了将序列特异性变体鉴定方法与质谱术分析组合的方法。例如,在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a)自个体的患病组织获得多个与MHC分子结合的肽;b)对MHC结合的肽进行基于质谱术的测序;并且c)将MHC结合的肽的质谱术衍生的序列信息与个体中疾病组织的变体编码序列的组关联,每种变体编码序列与参照样品相比具有序列中的变异,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,通过自疾病组织分离MHC/肽复合物(例如通过免疫沉淀)并且自MHC洗脱肽获得与MHC结合的多个肽。在一些实施方案中,对肽进行串联

质谱术。在一些实施方案中,基于质谱术的测序包括对肽进行质谱术,并且将质谱术谱与参照谱(诸如由参照样品种中的序列编码的推定蛋白质的假设的质谱术谱)比较。在一些实施方案中,在关联步骤前通过肽长度和/或锚基序的存在过滤质谱术序列信息。在一些实施方案中,方法还包括通过功能性分析验证疾病特异性免疫原性突变体肽。在一些实施方案中,疾病是癌症。在一些实施方案中,个体是人个体(诸如具有癌症的人个体)。

[0152] 在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a)获得个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品种相比具有序列中的变异;b)自个体的患病组织获得多个与MHC分子结合的肽;c)对MHC结合的肽进行基于质谱术的测序;并且d)将MHC结合的肽的质谱术衍生的序列信息与变体编码序列的第一组关联,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,可以过滤变体编码序列的第一组以获得编码预测结合MHC分子的肽的变体编码序列(下文称为“表位变体编码序列”)的较小组,然后对变体编码序列的较小组进行关联分析。在此类实施方案中,方法可以包括:a)提供个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品种相比具有序列中的变异;b)基于由变体编码序列的第一组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第一组选择表位变体编码序列的第二组,c)自个体的患病组织获得多个与MHC分子结合的肽;d)对MHC结合的肽进行基于质谱术的测序;并且e)将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的第二组关联,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法包括:a)获得个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品种相比具有序列中的变异;b)基于由变体编码序列的第一组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第一组选择表位变体编码序列的第二组,c)自个体的患病组织获得多个与MHC分子结合的肽;d)对MHC结合的肽进行基于质谱术的测序;并且e)将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的第二组关联,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法还包括通过功能性分析验证疾病特异性免疫原性突变体肽。在一些实施方案中,疾病是癌症。在一些实施方案中,个体是人个体(诸如具有癌症的人个体)。

[0153] 在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a)基于个体中的疾病组织的基因组序列,获得个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品种相比具有序列中的变异;b)自个体的患病组织获得多个与MHC分子结合的肽;c)对MHC结合的肽进行基于质谱术的测序;并且d)将MHC结合的肽的质谱术衍生的序列信息与变体编码序列的第一组关联,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,通过全基因组测序获得基因组序列。在一些实施方案中,通过全外显子组测序获得基因组序列。在一些实施方案中,通过靶向基因组或外显子组测序获得基因组序列。例如,可以首先通过探针组(例如对疾病关联基因特异性的探针)富集疾病组织和/或参照样品种中的基因组序列,之后加工以进行变体鉴定。

[0154] 在一些实施方案中,可以过滤变体编码序列的第一组以获得编码预测结合MHC分子的肽的变体编码序列(下文称为“表位变体编码序列”)的较小组,然后对变体编码序列的较小组进行免疫原性的预测。例如,在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a)基于个体中的疾病组织的基因组序列,

获得个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b)基于由变体编码序列的第一组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第一组选择表位变体编码序列的第二组,c)自个体的患病组织获得多个与MHC分子结合的肽;d)对MHC结合的肽进行基于质谱术的测序;并且e)将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的第二组关联,由此鉴定疾病特异性免疫原性突变体肽。

[0155] 在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a)基于个体中的疾病组织的转录物组序列,获得个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b)自个体的患病组织获得多个与MHC分子结合的肽;c)对MHC结合的肽进行基于质谱术的测序;并且d)将MHC结合的肽的质谱术衍生的序列信息与变体编码序列的第一组关联,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,通过全转录物组RNA-Seq测序获得转录物组序列。在一些实施方案中,通过靶向转录物组测序获得转录物组序列。例如,可以首先通过探针组(例如对疾病关联基因特异性的探针)富集疾病组织和/或参照样品中的RNA序列或cDNA序列,之后加工以进行变体鉴定。在一些实施方案中,可以过滤变体编码序列的第一组以获得表位变体编码序列的较小组,然后对变体编码序列的较小组进行免疫原性的预测。例如,在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a)基于个体中的疾病组织的转录物组序列,获得个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b)基于由变体编码序列的第一组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第一组选择表位变体编码序列的第二组,c)自个体的患病组织获得多个与MHC分子结合的肽;d)对MHC结合的肽进行基于质谱术的测序;并且e)将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的第二组关联,由此鉴定疾病特异性免疫原性突变体肽。

[0156] 在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a)基于个体中的疾病组织的基因组序列获得个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异,每种变体编码序列与参照样品相比具有序列中的变异;b)基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c)自个体的患病组织获得多个与MHC分子结合的肽;d)对MHC结合的肽进行基于质谱术的测序;并且e)将MHC结合的肽的质谱术衍生的序列信息与表达变体编码序列的第二组关联,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法包括:a)基于个体中疾病组织的基因组序列获得个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b)基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c)自个体的患病组织获得多个与MHC分子结合的肽;d)对MHC结合的肽进行基于质谱术的测序;并且e)将MHC结合的肽的质谱术衍生的序列信息与表达变体编码序列的第二组关联,由此鉴定疾病特异性免疫原性突变体肽。

[0157] 在一些实施方案中,可以过滤表达变体编码序列的第二组以获得表位变体编码序列的较小组,然后对变体编码序列的较小组进行免疫原性的预测。如此,例如,在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包

括:a) 基于个体中疾病组织的基因组序列,提供个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b) 基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c) 基于由表达变体编码序列的第二组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第二组选择表位变体编码序列的第三组,d) 自个体的患病组织获得多个与MHC分子结合的肽;e) 对MHC结合的肽进行基于质谱术的测序;并且f) 将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的第三组关联,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法包括:a) 基于个体中疾病组织的基因组序列获得个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b) 基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c) 基于由表达变体编码序列的第二组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第二组选择表位变体编码序列的第三组,d) 自个体的患病组织获得多个与MHC分子结合的肽;e) 对MHC结合的肽进行基于质谱术的测序;并且f) 将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的第三组关联,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法还包括通过功能性分析验证疾病特异性免疫原性突变体肽。在一些实施方案中,疾病是癌症。在一些实施方案中,个体是人个体(诸如具有癌症的人个体)。

[0158] 在一些实施方案中,通过预测肽的免疫原性进一步选择通过本文中描述的基于质谱术的方法鉴定的疾病特异性免疫原性突变体肽。在一些实施方案中,选择步骤包括基于一项或多项(例如2、3、4、5、6、7、8、9、10或11之任一种)参数预测肽的免疫原性:i) 肽对MHCI分子的结合亲和力;ii) 含有肽的肽前体的蛋白质水平;iii) 编码肽前体的转录物的表达水平;iv) 通过免疫蛋白酶体对肽前体的加工效率;v) 编码肽前体的转录物的表达时机;vi) 肽对TCR分子的结合亲和力;vii) 变体氨基酸在肽内的位置;xiii) 在与MHCI分子结合时肽的溶剂暴露;ix) 在与MHCI分子结合时变体氨基酸的溶剂暴露;和x) 肽中芳香族残基的含量;和xi) 肽前体的性质。

[0159] 如此,例如,在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a) 自个体的患病组织获得多个与MHC分子结合的肽;b) 对MHC结合的肽进行基于质谱术的测序;并且c) 将MHC结合的肽的质谱术衍生的序列信息与个体中疾病组织的变体编码序列的组关联,每种变体编码序列与参照样品相比具有序列中的变异,以获得变体编码序列的第二组,并且d) 自变体编码序列的第二组选择免疫原性变体编码序列,其中选择步骤包括预测包含由变体编码序列的第二组编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a) 获得个体中疾病组织的变体编码序列的组,每种变体编码序列与参照样品相比具有序列中的变异;b) 自个体的患病组织获得多个与MHC分子结合的肽;c) 对MHC结合的肽进行基于质谱术的测序;d) 将MHC结合的肽的质谱术衍生的序列信息与变体编码序列的第一组关联,以获得变体编码序列的第二组,并且e) 自变体编码序列的第二组选择免疫原性变体编码序列,其中选择步骤包括预测包含由变体编码序列的第二组编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法包括:a) 提供个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b) 基于由变体编

码序列的第一组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第一组选择表位变体编码序列的第二组,c)自个体的患病组织获得多个与MHC分子结合的肽;d)对MHC结合的肽进行基于质谱术的测序;e)将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的第二组关联,以获得变体编码序列的第三组,并且f)自变体编码序列的第三组选择免疫原性变体编码序列,其中选择步骤包括预测包含由变体编码序列的第三组编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法包括:a)获得个体中疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品种相比具有序列中的变异;b)基于由变体编码序列的第一组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第一组选择表位变体编码序列的第二组,c)自个体的患病组织获得多个与MHC分子结合的肽;d)对MHC结合的肽进行基于质谱术的测序;并且e)将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的组关联,以获得变体编码序列的第三组,并且f)自变体编码序列的第三组选择免疫原性变体编码序列,其中选择步骤包括预测包含由变体编码序列的第三组编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法还包括通过功能性分析验证疾病特异性免疫原性突变体肽。在一些实施方案中,方法还包括通过功能性分析验证疾病特异性免疫原性突变体肽。在一些实施方案中,疾病是癌症。在一些实施方案中,个体是人个体(诸如具有癌症的人个体)。

[0160] 在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a)基于个体中的疾病组织的基因组序列提供个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品种相比具有序列中的变异;b)基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c)自个体的患病组织获得多个与MHC分子结合的肽,d)对MHC结合的肽进行基于质谱术的测序;并且e)将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的第二组关联,以获得变体编码序列的第三组,并且g)自变体编码序列的第三组选择免疫原性变体编码序列,其中选择步骤包括预测包含由变体编码序列的第三组编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法包括:a)基于个体中的疾病组织的基因组序列获得个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品种相比具有序列中的变异;b)基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c)自个体的患病组织获得多个与MHC分子结合的肽,d)对MHC结合的肽进行基于质谱术的测序;并且e)将MHC结合的肽的质谱术衍生的序列信息与表达变体编码序列的第二组关联,以获得变体编码序列的第三组,并且g)自变体编码序列的第三组选择免疫原性变体编码序列,其中选择步骤包括预测包含由变体编码序列的第三组编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法还包括通过功能性分析验证疾病特异性免疫原性突变体肽。在一些实施方案中,疾病是癌症。在一些实施方案中,个体是人个体(诸如具有癌症的人个体)。

[0161] 在一些实施方案中,提供了鉴定来自个体中的疾病组织的疾病特异性免疫原性突变体肽的方法,其包括:a)基于个体中的疾病组织的基因组序列提供个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品种相比具有序列中的变异;b)基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c)基于由表达

变体编码序列的第二组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第二组选择表位变体编码序列的第三组,d) 自个体的患病组织获得多个与MHC分子结合的肽; e) 对MHC结合的肽进行基于质谱术的测序;并且f) 将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的第三组关联,以获得变体编码序列的第四组,并且g) 自变体编码序列的第四组选择免疫原性变体编码序列,其中选择步骤包括预测包含由变体编码序列的第四组编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法包括:a) 基于个体中的疾病组织的基因组序列获得个体中的疾病组织的变体编码序列的第一组,每种变体编码序列与参照样品相比具有序列中的变异;b) 基于个体中疾病组织的转录物组序列,自第一组选择表达变体编码序列的第二组;c) 基于由表达变体编码序列的第二组编码的肽结合MHC分子(诸如MHC I类分子或MHCI)的预测能力,自第二组选择表位变体编码序列的第三组,d) 自个体的患病组织获得多个与MHC分子结合的肽; e) 对MHC结合的肽进行基于质谱术的测序;并且f) 将MHC结合的肽的质谱术衍生的序列信息与表位变体编码序列的第三组关联,以获得变体编码序列的第四组,并且g) 自变体编码序列的第四组选择免疫原性变体编码序列,其中选择步骤包括预测包含由变体编码序列的第四组编码的变体氨基酸的肽的免疫原性,由此鉴定疾病特异性免疫原性突变体肽。在一些实施方案中,方法还包括通过功能性分析验证疾病特异性免疫原性突变体肽。在一些实施方案中,疾病是癌症。在一些实施方案中,个体是人个体(诸如具有癌症的人个体)。

[0162] 本文中还提供通过本文中描述的任一种方法获得的疾病特异性免疫原性突变体肽。例如,可以使用疾病特异性免疫原性突变体肽生成用于治疗疾病的组合物(诸如疫苗组合物)。或者,可以使用疾病特异性免疫原性突变体肽生成突变体肽特异性治疗剂,诸如治疗性抗体。

[0163] 本文中描述的方法在个性化医学背景下是特别有用的,其中使用通过本文中描述的任一种方法获得的疾病特异性免疫原性突变体肽开发用于同一个体的治疗剂(诸如疫苗或治疗性抗体)。如此,例如,在一些实施方案中,提供了治疗个体中的疾病(诸如癌症)的方法,其包括:a) 鉴定个体中疾病特异性免疫原性突变体肽;并且b) 合成肽;并且c) 对个体施用肽。在一些实施方案中,提供了治疗个体中的疾病(诸如癌症)的方法,其包括:a) 自个体获得疾病组织样品;b) 鉴定个体中的疾病特异性免疫原性突变体肽;并且c) 合成肽;并且d) 对个体施用肽。在一些实施方案中,提供了治疗个体中的疾病(诸如癌症)的方法,其包括:a) 鉴定个体中的疾病特异性免疫原性突变体肽;b) 生成特异性识别突变体肽的抗体(或TCR类似物,诸如可溶性TCR);并且c) 对个体施用肽。在一些实施方案中,提供了治疗个体中的疾病(诸如癌症)的方法,其包括:a) 自个体获得疾病组织样品;b) 鉴定个体中的疾病特异性免疫原性突变体肽;c) 生成特异性识别突变体肽的抗体(或TCR类似物,诸如可溶性TCR);并且d) 对个体施用肽。在一些实施方案中,鉴定步骤将序列特异性变体鉴定方法与免疫原性预测方法组合。在一些实施方案中,鉴定步骤将序列特异性变体鉴定方法与质谱术组合。可以将本文中描述的鉴定疾病特异性免疫原性突变体肽的任何方法用于本文中描述的治疗方法。

[0164] 获得变体编码序列

[0165] 本文中在各种实施方案中描述的方法包括提供和/或获得变体编码序列。一般可以例如通过对个体的疾病组织样品中的基因组或RNA序列测序,并将序列与自参照样品获

得的序列进行比较来获得变体编码序列。

[0166] 在一些实施方案中,疾病组织是血液。在一些实施方案中,疾病组织是固体组织(诸如实体瘤)。在一些实施方案中,疾病组织是细胞的集合(例如血液中的循环癌细胞)。在一些实施方案中,疾病组织是淋巴细胞的集合。在一些实施方案中,疾病组织是白细胞的集合。在一些实施方案中,疾病组织是上皮细胞的集合。在一些实施方案中,疾病组织是结缔组织。在一些实施方案中,疾病组织是生殖细胞和/或多能细胞的集合。在一些实施方案中,疾病组织是胚细胞的集合。

[0167] 合适的疾病组织样品包括但不限于肿瘤组织、与肿瘤相邻的正常组织、肿瘤远端的正常组织、或外周血淋巴细胞。在一些实施方案中,疾病组织样品是肿瘤组织。在一些实施方案中,疾病组织样品是含有癌细胞的活检,诸如癌细胞(例如胰腺癌细胞)的细针抽吸或腹腔镜检查获得的癌细胞(例如胰腺癌细胞)。在一些实施方案中,将活检的细胞离心成团粒,固定,并且在石蜡中包埋,之后分析。在一些实施方案中,在分析前快速冷冻活检的细胞。

[0168] 在一些实施方案中,疾病组织样品包含循环转移癌细胞。在一些实施方案中,通过自血液分选循环肿瘤细胞(CTC)获得疾病组织样品。在另一个实施方案中,CTC已经自原发性肿瘤脱离并在体液中循环。在又一个实施方案中,CTC已经自原发性肿瘤脱离并在血流中循环。在另一个实施方案中,CTC是转移的指示。在一些实施方案中,CTC是胰腺癌细胞。在一些实施方案中,CTC是结肠直肠癌细胞。在一些实施方案中,CTC是非小细胞肺癌细胞。

[0169] 可以基于个体中的疾病组织中的基因组序列鉴定变异。例如,可以自个体的疾病组织获得基因组DNA并进行测序分析。然后,将如此获得的序列与自参照样品的序列比较。在一些实施方案中,对疾病样品进行全基因组测序。在一些实施方案中,将疾病样品进行全外显子组测序,即仅对基因组序列中的外显子测序。在一些实施方案中,在与参照样品比较前,对基因组序列“富集”特定序列。例如,可以设计特异性探针以在进行测序分析前富集某些期望的序列(例如疾病特异性序列)。全基因组测序、全外显子组测序和靶向测序的方法是本领域中已知的,并且报告于例如Bentley,D.R.et al.,Accurate whole human genome sequencing using reversible terminator chemistry,Nature,2008,v.456,53-59;Choi,M.et al.,Genetic diagnosis by whole exome capture and massively parallel DNA sequencing,Proceedings of the National Academy of Sciences,2009,v.106(45),19096-19101;及Ng,S.B.et al.,Targeted capture and massively parallel sequencing of 12 human exomes,Nature,2009,v.461,272-276,在此通过提及收录它们。

[0170] 在一些实施方案中,基于个体中疾病组织中的转录物组序列鉴定变异。例如,可以自个体中的疾病组织获得全部或部分转录物组序列(例如通过诸如RNA-Seq的方法)并进行测序分析。然后,将如此获得的序列与自参照样品的序列比较。在一些实施方案中,将疾病样品进行全转录物组RNA-Seq测序。在一些实施方案中,在与参照样品比较前,对转录物组序列“富集”特定序列。例如,可以设计特异性探针以在进行测序分析前富集某些期望的序列(例如疾病特异性序列)。全转录物组测序和靶向测序的方法是本领域中已知的,并且报告于例如Tang,F.et al.,mRNA-Seq whole-transcriptome analysis of a single cell,Nature Methods,2009,v.6,377-382;Ozsolak,F.,RNA sequencing:advances, challenges and opportunities,Nature Reviews,2011,v.12,87-98;German,M.A et

al., Global identification of microRNA-target RNAs by parallel analysis of RNA ends, *Nature Biotechnology*, 2008, v.26, 941-946; 及 Wang, Z. et al., RNA-Seq: a revolutionary tool for transcriptomics, *Nature Reviews*, 2009, v.10, p.57-63. 在一些实施方案中, 转录组测序技术包括但不限于RNA聚(A)文库、微阵列分析、平行测序、大规模平行测序、PCR和RNA-Seq。RNA-Seq是一种用于对部分或基本上整个转录组测序的高通量技术。简言之, 用与一个或两个末端附着的衔接头将转录组序列的分离的群体转化成cDNA片段的文库。在扩增或不扩增的情况下, 然后分析每种cDNA分子以获得序列信息的短区段, 通常为30-400个碱基对。然后, 将序列信息的这些片段与参照基因组、参照转录物比对或者从头装配以揭示转录物的结构(即转录边界)和/或表达水平。

[0171] 一旦获得, 可以将疾病组织中的序列与参照样品种中的相应序列比较。可以在核酸水平进行序列比较, 其通过将疾病组织中的核酸序列与参照样品种中的相应序列比对进行。然后, 鉴定导致编码氨基酸的一种或多种变化的序列变异。或者, 可以在氨基酸水平进行序列比较, 也就是说, 在实施比较前首先在计算机上将核酸序列转化成氨基酸序列。

[0172] 在一些实施方案中, 可以通过本领域中已知的技术, 诸如手动比对、FAST-A11 (FASTA) 和基本局部比对搜索工具 (BLAST) 完成来自疾病组织的序列与参照的序列的比较。通过BLAST完成的序列比较需要疾病序列的输入和参照序列的输入。BLAST通过首先鉴定两种序列之间的短序列匹配(称为接种的过程)将疾病序列与参照数据库比较。一旦找到序列匹配, 则使用评分矩阵实施序列比对的扩展。

[0173] 在一些实施方案中, 参照样品种是匹配的无疾病的组织样品。如本文中使用的, “匹配的”无疾病的组织样品是选自与疾病组织相同或相似组织类型的组织样品。在一些实施方案中, 匹配的无疾病的组织和疾病组织可以源自同一个体。在一些实施方案中, 本文中描述的参照样品种是来自同一个体的无疾病的样品。在一些实施方案中, 参照样品种是来自不同个体(例如没有疾病的个体)的无疾病的样品。在一些实施方案中, 参照样品种是自不同个体的群体获得的。在一些实施方案中, 参照样品种是与生物体相关的已知基因的数据库。在一些实施方案中, 参照样品种可以是与生物体相关的已知基因和来自匹配的无疾病的组织样品的基因组信息的组合。在一些实施方案中, 变体编码序列可以编码或包含氨基酸序列中的点突变。在一些实施方案中, 变体编码序列可以编码或包含氨基酸删除或插入。

[0174] 在一些实施方案中, 首先基于基因组序列鉴定变体编码序列的组。然后, 进一步过滤此初始组, 以基于转录组测序数据库中变体编码序列的存在获得较窄的表达变体编码序列组(并且如此视为“表达的”)。在一些实施方案中, 通过过滤通过转录组测序数据库将变体编码序列的组减少至少约10、20、30、40、50或更多倍。

[0175] 在一些实施方案中, 变体编码序列是源自导致蛋白质中的不同氨基酸的非同义突变(例如点突变)的序列。在一些实施方案中, 变体编码序列是源自连读突变的序列, 在所述连读突变中修饰或删除终止密码子, 导致在C端具有新的肿瘤特异性序列的较长蛋白质的翻译。在一些实施方案中, 变体编码序列是源自剪接位点突变的序列, 所述剪接位点突变导致在成熟mRNA中内含子的纳入, 并且如此导致独特的肿瘤特异性蛋白序列。在一些实施方案中, 变体编码序列是源自染色体重排的序列, 所述染色体重排产生在2种蛋白质的接合处具有肿瘤特异性序列的嵌合蛋白(即基因融合)。在一些实施方案中, 变体编码序列是源自导致具有新肿瘤特异性蛋白质序列的新可读框的移码突变或删除的序列。在一些实施方案

中,变体编码序列是源自超过一种突变的序列。在一些实施方案中,变体编码序列是源自超过一种突变机制的序列。

[0176] 获得表位变体编码序列

[0177] 过滤本文中在一些实施方案中描述的变体编码序列以获得编码预测结合MHC分子的肽的变体编码序列(“表位变体编码序列”)的较小组。在一些实施方案中,通过过滤通过MHC结合预测过程将所述组的变体编码序列减少至少约10、20、30、40、50、60、80、100、150、200、250、300或更多倍。

[0178] 可以通过预测算法(诸如NETMHC)评估由变体编码序列编码的肽结合MHC(诸如MHCI)的能力。简言之,NETMHC是一种使用来自免疫表位数据库和分析资源(IEDB)的亲合力数据和来自SYFPEITHI的洗脱数据两者对定量肽数据训练的算法。可以对长度为8和14个之间的氨基酸的肽进行MHCI结合的预测。NETMHC使用对来自55种MHC等位基因(43种人和12种非人)的9聚体氨基酸序列训练的预测器。为了允许预测较短的输入序列,NETMHC实际上会扩展8聚体氨基酸序列。对于长于9聚体的输入序列,NETMHC会产生输入序列内含有的所有可能的9聚体氨基酸序列。NETMHC然后使用训练的人工神经网络和位置特定评分矩阵来预测MHC结合。

[0179] 选择免疫原性变体编码序列

[0180] 本文中在一些实施方案中提供的方法还包括选择免疫原性变体编码序列,其包括预测包含由变体编码序列编码的变体氨基酸的肽的免疫原性。可以例如通过考虑肽和相应的肽前体的一种或多种参数预测肽为免疫原性的可能性的方法(诸如在计算机程序中)实施免疫原性的预测。这些参数包括但不限于i)肽对MHCI分子的结合亲合力;ii)含有肽的肽前体的蛋白质水平;iii)编码肽前体的转录物的表达水平;iv)通过免疫蛋白酶体对肽前体的加工效率;v)编码肽前体的转录物的表达时机;vi)肽对TCR分子的结合亲合力;vii)变体氨基酸在肽内的位置;xiii)在与MHCI分子结合时肽的溶剂暴露;ix)在与MHCI分子结合时变体氨基酸的溶剂暴露;和x)肽中芳香族残基的含量;和xi)肽前体的性质。在一些实施方案中,免疫原性基于本文中描述的参数的至少2、3、4、5、6、7、8、9或10项。

[0181] 在一些实施方案中,使用肽对MHC分子的结合亲合力预测免疫原性。肽对MHC分子的结合亲合力可以预测pMHC的稳定性,其继而可以允许pMHC的延长呈现,如此增加细胞表面暴露以与免疫细胞潜在相互作用。可以使用本领域中已知的技术预测结合亲合力,诸如RankPep、MHC Bench、nHLAPred、SVMHC、NETMHCpan和POPI,其基于诸如人工神经网络、平均相对结合矩阵、定量矩阵和稳定化矩阵法的方法。在一些实施方案中,肽对MHC的结合亲合力基于位于已知锚定位置中的特定氨基酸残基(牵涉MHC结合的氨基酸)的存在。在一些实施方案中,对肽中的每个残基评估其对结合的贡献。在一些实施方案中,用已知结合MHC的肽训练分析系统。在一些实施方案中,计算肽-MHC分子的结合能。在一些实施方案中,使用结合阈值评估肽结合MHC的预测亲合力,例如IC₅₀值<500nM。

[0182] 在一些实施方案中,使用疾病组织中肽前体的表达水平预测免疫原性。在一些实施方案中,以生物化学法(诸如Western印迹和ELISA)测量蛋白质水平。在一些实施方案中,可以通过已知的定量质谱术技术测量蛋白质水平。如La Gruta等人证明的,疾病组织中肽前体的高表达水平可以与预测的免疫原性关联。如描述的,补料入表位加工途径中的较大的肽前体的利用度与增加的表位呈递和免疫原性应答正相关(La Gruta, N.L. et al., A

virus-specific CD8+T cell immunodominance hierarchy determined by antigen dose and precursor frequencies, *Proceedings of the National Academy of Sciences*, 2006, V.103, 994-999, 在此通过提及将其收录)。

[0183] 在一些实施方案中,可以使用编码疾病组织中的肽前体的转录物的表达水平预测免疫原性。在一些实施方案中,通过RT-PCR测量RNA表达水平。在一些实施方案中,通过测序分析测量RNA表达水平。如上文讨论,增加肽前体对表位加工途径的利用度与源自所述肽前体的所述表位相关的免疫原性应答正相关。此外,已经观察到mRNA水平和蛋白质丰度之间的正关联 (Ghaemmaghami, S. et al., *Global analysis of protein expression in yeast*, *Nature*, 2003, v. 425, 737-741, 在此通过提及将其收录)。因此,可以使用编码疾病组织中的肽前体的转录物的表达水平预测免疫原性。

[0184] 在一些实施方案中,使用通过免疫蛋白酶体对肽前体的加工效率预测免疫原性。如本文中使用的,肽前体的“加工效率”是指表达、翻译、转录、消化、转运来源氨基酸序列(即较大的肽或蛋白质)以及在结合MHC I分子前的任何进一步加工的效率。“免疫蛋白酶体”是蛋白酶的集合,其将肽和/或蛋白质前体酶促消化成小的氨基酸序列,用于表位形成的最终目的。例如,如由Chen等人证明的,免疫蛋白酶体产生表位前体的能力与免疫原性直接相关 (Chen, W. et al., *Immunoproteasomes shape immunodominance hierarchies of antiviral CD8+T cell repertoire and presentation of viral antigens*, *The Journal of Experimental Medicine*, 2001, v. 193, 1319-1326, 在此通过提及将其收录)。可以使用牵涉表位加工的步骤的知识预测所得表位的免疫原性。特别地,关于免疫蛋白酶体的研究提示它有效生成MHC表位,并且对免疫蛋白酶体机制的了解会有助于预测免疫原性肽。

[0185] 在一些实施方案中,可以使用肽前体的表达时机预测免疫原性。在疾病进展中较早期表达的蛋白质更可能呈现为MHC表位 (Moutaftsi, M. et al., *A consensus epitope prediction approach identifies the breadth of murine T CD8+-cell responses to vaccinia virus*, *Nature Biotechnology*, 2006, v. 24, 817-819, 在此通过提及将其收录)。在一些实施方案中,可以使用疾病组织的比较分析测定基因产物的时间表达模式。在一些实施方案中,时间表达模式的外推可以允许鉴定与其它鉴定的表达基因产物相比在早期时间点以更大的丰度呈现的表达基因产物。

[0186] 在一些实施方案中,可以使用肽表位与T细胞受体 (TCR) 的结合亲和力预测免疫原性。预测肽表位与TCR的结合亲和力的方法是本领域中已知的,并且报告于例如Tung, C. - W. et al., *POPISK: T-cell reactivity prediction using support vector machines and string kernels*, *BMC Bioinformatics*, 2011, v. 12, 446, 在此通过提及将其收录。

[0187] 在一些实施方案中,使用变体氨基酸在肽中的位置预测免疫原性。肽表位在两个独特的锚定位置处结合MHC I分子。锚定位置之间的跨度以约6-7个氨基酸分开,如通过表位肽序列测量的,不包括占据锚定位置的氨基酸。已经报告了,两个MHC I锚定位置间的氨基酸跨度(即4-6位的氨基酸)中的突变更可能与免疫原性应答正相关 (Calis, J. J. A. et al., *Properties of MHC class I presented peptides that enhance immunogenicity*, *PLOS Computational Biology*, 2013, v. 9, 1-13, 在此通过提及将其收录)。通过开始对于末端氨基酸的序列位置计数1确定氨基酸的序列位置。

[0188] 在一些实施方案中,在MHC呈递表位上呈递的肽的结构特征可以预测免疫原性。可以通过计算机3维分析和/或蛋白质停靠程序进行MHC结合的肽的结构评估。预测pMHC分子的结构的方法是本领域中已知的,并且报告于例如在Marti-Renom,M.A.et al., Comparative protein structure modeling of genes and genomes,Annual Review of Biophysics and Biomolecular Structure,2000,v.29,291-325,Chivian,D.et al., Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection,Nucleic Acids Research,2006,v.34, e1 12,and McRobb,F.M.et al.,Homology modeling and docking evaluation of aminergic G protein-coupled receptors,Journal of Chemical Information and Modeling,2010,v.50,626-637,在此通过提及将其收录。当与MHC分子结合时预测的表位结构(诸如自Rosetta算法获得)的使用可用于评估当表位与MHC分子结合时所述表位的氨基酸残基的溶剂暴露程度。随后,可以将此信息与肽的免疫原性关联。例如,如Park等人描述,突变体肽(其中与野生型序列相比,变体氨基酸残基展现额外的溶剂暴露)与增加的免疫原性正相关(Park,M.-S.et al.,Accurate structure prediction of peptide-MHC complexes for identifying highly immunogenic antigens,Molecular Immunology, 2013,v.56,81-90,在此通过提及将其收录。)在一些实施方案中,使用在MHC复合物上呈递时整个肽的溶剂暴露预测免疫原性。在一些实施方案中,使用肽上的变体氨基酸的溶剂暴露预测免疫原性。

[0189] 在一些实施方案中,可以使用肽中的大的和/或芳香族的残基的表位含量预测免疫原性。Calis等人观察到表位氨基酸序列内的大的和/或芳香族的氨基酸残基的存在与免疫原性之间的联系。具体地,报告了苯丙氨酸和异亮氨酸含量与表位免疫原性正相关。如上文讨论,可以使用变体氨基酸的位置和结构评估两者预测表位免疫原性。除了此讨论之外,假设可以预测表位4-6位内的大的和/或芳香族的残基具有较大程度的溶剂暴露。

[0190] 在一些实施方案中,可以使用肽前体的性质预测变体编码氨基酸序列的免疫原性。例如,当疾病组织是肿瘤时,已知与癌症相关的肽前体序列在预测免疫原性时可以是有益的。然而,来自不与癌症直接相关的蛋白质的肽前体序列(例如当突变时)在本发明的方法内也是有益的。

[0191] 在一些实施方案中,可以使用本文中描述的至少两种(诸如3、4、5、6、7、8、9或10中的至少任一)参数预测肽的免疫原性。在一些实施方案中,可以使用第一预测性评估选择一组变体编码氨基酸序列,其随后通过第二预测性评估加工,导致表位免疫原性的累积预测。作为本公开内容的一部分意图可以使用多轮评估预测表位免疫原性。或者,平行评估多项参数,并且可以获得基于各种参数的评估的综合得分。例如,可以为每项参数计算得分,并且可以为每项参数分配百分比权重。然后,可以基于评估的每项参数的得分和百分比权重计算综合范围。也可以使用与平行参数评估偶联的序贯评估的组合。

[0192] 对于变体编码序列中的给定变异,可以产生各种长度的多个重叠推定突变体肽。在一个实施方案中,可以基于免疫原性对这些多个重叠推定突变体肽排序,其可以涵盖如本文中描述的一种或多种分析。此外,可以通过本领域中公知的任一种或多种手段实现推定的突变体肽的排序,所述手段可以包括在免疫原性测定中或者根据进行排序过程的那些的偏爱而认为是分开的分析。这些手段的一些非限制性例子包括前体蛋白的丰度、加工的

蛋白质组内的肽的丰度,包括此加工的效率,和肽-MHC复合物内肽的丰度。潜在排序分析的其他例子包括但不限于肽的结合亲和力、肽-MHC复合物的稳定性、和肽与自身肽的相似性或差异。或者,可以基于与测量普通技术人员公知的或本文中描述的这些或其它特征的质谱术方法产生的数据的关联对肽排序。一些代表性特征包括增加免疫原性的大的或芳香族的氨基酸的存在或缺乏和在肽内找到这些氨基酸的特定位置,对于在肽的中间位置中找到的那些氨基酸给予免疫原性影响偏爱,例如肽4到6(见例如Calis et al.PLOS,9(10): e1003266(2013))。

[0193] 在一些实施方案中,免疫原性的预测还包括HLA(人白细胞抗原)分型分析。由于MHC的多基因性(polygeny),每个人将在他或她的细胞上表达至少三种不同的抗原呈递MHC I类分子和三种(或有时四种)MHC II类分子。事实上,由于MHC的极端多态性和MHC基因产物的共显性表达,在大多数人的细胞上表达的不同MHC分子的数目是较大的。一些人MHC I类和II类基因有超过200种等位基因,每种等位基因在群体中以相对高的频率存在。因此,仅有较小的下述几率,即在个体的两条同源染色体上相应的MHC位点会具有相同的等位基因;大多数个体会在MHC基因座处是杂合的。在单个染色体上找到的MHC等位基因的特定组合称为MHC单倍型。MHC等位基因的表达是共显性的,在细胞中表达一种基因座处的两种等位基因的蛋白质产物,并且两种基因产物都能够将抗原呈递给T细胞。如此,每种基因座上的广泛多态性具有将个体中表达的不同MHC分子的数目倍增,由此增加已经经由多基因性可得多样性的潜力(关于一般概述,见例如Janeway's Immunobiology,Murphy,Kenneth编,Garland Science,New York,NY(2011))。可以使用现有技术中已知的几种方法中的任一种,诸如基于DNA的组织相容性测定法完成HLA分型。本领域中的方法的具体例子牵涉进一步分析的聚合酶链式反应(PCR)产物,诸如PCR-RFLP(限制性片段长度多态性)、PCR-SSO(序列特异性寡核苷酸)、PCR-SSP(序列特异性引物)、和PCR-SBT(基于序列的分型)技术。如此,测定牵涉感兴趣的肽呈递的特定基因多态性类型可以提供关于突变体肽的免疫原性的进一步信息。

[0194] 获得与MHC分子结合的肽

[0195] 本文中在一些实施方案中提供的方法包括自个体的疾病组织获得与MHC分子结合的肽。在一些实施方案中,通过免疫亲和技术分离MHC结合的肽。在一些实施方案中,通过亲和层析分离MHC结合的肽。在一些实施方案中,通过免疫亲和层析分离MHC结合的肽。在一些实施方案中,通过免疫沉淀技术分离MHC结合的肽。

[0196] 在一些实施方案中,使用抗MHC抗体捕捉MHC/肽分子。在一些实施方案中,可以使用任选具有不同亲和力和/或结合特征的多种抗MHC抗体捕捉MHC/肽复合物。合适的抗体包括但不限于对HLA I类特异性的单克隆抗体W6/32,和对HLA-A2特异性的单克隆抗体BB7.2。

[0197] 在一些实施方案中,可以首先分离MHC/肽复合物,随后将MHC结合的肽与MHC分子分开。在一些实施方案中,通过酸洗脱将MHC结合的肽与MHC分子分开。在一些实施方案中,可以对完整的全细胞,任选地在存在裂解的细胞和/或细胞残留物的情况中实施酸介导的与MHC分子分开MHC结合的肽。在一些实施方案中,可以在将pMHC暴露于具有酸性pH的缓冲液后与MHC分子分开MHC结合的肽。在一些实施方案中,可以通过温和酸洗脱(MAE)将MHC结合的肽与MHC分子分开。在一些实施方案中,可以通过胞外表面的温和酸洗脱(MAE)将MHC结合的肽与MHC分子分开。在一些实施方案中,可以通过pMHC分子的变性将MHC结合的肽与MCH

分子分开。

[0198] 在一些实施方案中,可以在基于质谱术的测序前进一步加工MHC结合的肽。在一些实施方案中,可以在基于质谱术的测序前将MHC结合的肽浓缩。在一些实施方案中,可以在基于质谱术的测序前纯化MHC结合的肽。在一些实施方案中,可以在基于质谱术的测序前将MHC结合的肽分级。在一些实施方案中,可以在基于质谱术的测序前将MHC结合的肽富集。在一些实施方案中,可以在进行基于质谱术的测序前将MHC结合的肽进一步酶促消化。在一些实施方案中,可以在基于质谱术的测序前更换可以含有MHC结合的肽的缓冲液。在一些实施方案中,可以在基于质谱术的测序前标记MHC结合的肽。在一些实施方案中,可以在基于质谱术的测序前共价标记MHC结合的肽。在一些实施方案中,可以在基于质谱术的测序前酶促标记MHC结合的肽。在一些实施方案中,可以在基于质谱术的测序前化学标记MHC结合的肽。在一些实施方案中,可以标记MHC结合的肽以允许在基于质谱术的测序期间增强离子化。在一些实施方案中,可以标记MHC结合的肽以允许在基于质谱术的测序期间进行定量。在一些实施方案中,分离的MHC结合的肽可以源自多种来源和/或富集方法,并且任选地可以在基于质谱术的测序前共同合并。

[0199] 基于质谱术的肽测序

[0200] 对在本文中描述的方法中与MHC分子结合的肽进行质谱测序。如本文中使用的,“基于质谱术的测序”是指通过使用质谱术鉴定肽和/或蛋白质的氨基酸序列的技术。质谱仪是能够测量个别的离子化分子的质荷(m/z)比的仪器,使研究人员能够鉴定未知化合物,量化已知化合物,并且阐明分子的结构和化学性质。可以使用本文中提供的方法获得与MHCI分子结合的肽表位的序列信息。在一些实施方案中,可以测定肽表位的整个序列。在一些实施方案中,可以测定肽表位的部分序列。在一些实施方案中,对肽进行串联质谱术,诸如串联层析质谱术(例如LC-MS或LC-MS-MS)。

[0201] 在一些实施方案中,通过将样品分离并且加载入仪器上开始质谱术分析。在一些实施方案中,可以在质谱术分析前层析加工MHC结合的肽。在一些实施方案中,层析法是液相层析。在一些实施方案中,层析是反相层析。在一些实施方案中,可以将MHC结合的肽层析分开,并且同时浓缩,之后导入质谱仪中。在一些实施方案中,层析分开可以是在线的,其中自层析源洗脱的肽直接进入质谱仪中。在一些实施方案中,层析分开可以是离线的。在一些实施方案中,可以使用离线层析分开分级分离MHC结合的肽。离线层析分开通常牵涉质谱术样品的分开和/或分级,其中在样品离开层析系统时,不将所得的分开和/或分级的样品立即导入质谱仪中。

[0202] 在一些实施方案中,可以使用已知的质谱术离子化技术(诸如基质辅助激光解吸/离子化、电喷雾离子化和/或纳米电喷雾离子化、大气压化学离子化)对MHC结合的肽测序。在一些实施方案中,MHC结合的肽可以在质谱仪外部、内部和/或在它们进入质谱仪时离子化。在一些实施方案中,可以在质谱仪中分析MHC结合的肽的正离子。随后,经由暴露于磁场,离子依照其质荷比分开。在一些实施方案中,使用扇形仪器,并且离子依照在它通过仪器的电磁场时离子轨迹的偏转幅度量,所述电磁场与离子质荷比正相关。在其它实施方案中,当离子通过四极体时,或者基于它们在三维或线性离子阱或Orbitrap中的运动,或在傅立叶变换离子回旋加速器共振质谱仪的磁场中,测量离子质荷比。仪器记录每个离子的相对丰度,其用于测定初始样品的化学、分子和/或同位素组成。在一些实施方案中,使用飞

行时间仪器,并且利用电场加速离子通过相同的电势,并且测量每个离子到达检测器花费的时间。此方法取决于均匀的每个离子的电荷,使得每个离子的动能会是相同的。在此情况中影响速度的唯一变量是质量,较轻的离子以较大的速度行进并且因此较快到达检测器。以质谱或直方图、强度对质荷比呈现所得的数据,峰表示离子化化合物或片段。

[0203] 可以通过串联质谱术获得质谱数据。在一些实施方案中,用于获得关于MHC结合的肽的肽测序的信息的质谱术获得技术可以是数据依赖性的。在一些实施方案中,用于获得关于MHC结合的肽的肽测序的信息的质谱术获得技术可以是数据非依赖性的。在一些实施方案中,用于获得关于MHC结合的肽的肽测序的信息的质谱术获得技术可以基于测量的精确质量质谱术。在一些实施方案中,用于获得关于MHC结合的肽的肽测序的信息的质谱术获得技术可以是肽质量指纹法。可以通过肽质量指纹法获得本发明中有用的质谱数据。肽质量指纹法牵涉将来自通过蛋白水解消化产生的肽混合物的谱的观察质量输入数据库中,并且将观察到的质量与计算机上的源自已知蛋白质的消化的预测片段质量关联。与样品质量对应的已知质量提供了已知蛋白质存在于测试样品中的证据。

[0204] 在一些实施方案中,串联质谱术包括使肽离子与气体碰撞并且片段化(例如由于由碰撞赋予的振动能量)的过程。片段化过程产生在沿着蛋白质的各个位点处的肽键处断裂的断裂产物。观察到的片段质量可以与许多给定肽序列之一的预测质量数据库匹配,并且可以预测蛋白质的存在。在一些实施方案中,质谱术采集技术可以利用片段化技术(诸如碰撞诱导的解离、脉冲Q解离、高能碰撞解离、电子转移解离和电子转移解离、红外多光子解离)。

[0205] 在一些实施方案中,可以使用从质谱仪获得的数据鉴定肽序列。在一些实施方案中,可以使用搜索算法(诸如SEQUEST和Mascot)将肽序列分配给获得的质谱。在一些实施方案中,分配的肽序列可以具有小于约5%的错误发现率。在一些实施方案中,分配的肽序列可以具有小于约1%的错误发现率。在一些实施方案中,分配的肽序列可以具有小于约0.5%的错误发现率。在一些实施例中,可以通过搜索算法使用数据库进行获得谱的肽序列分配。在一些实施方案中,通过搜索算法使用以进行获得谱的肽序列分配的数据库可以是生物体的已知序列的数据库。在一些实施方案中,通过搜索算法使用以进行获得谱的肽序列分配的数据库可以是生物体的已知蛋白质的数据库。在一些实施方案中,通过搜索算法使用以进行获得谱的肽序列分配的数据库可以是生物体的已知基因组序列的数据库。在一些实施方案中,通过搜索算法使用以进行获得谱的肽序列分配的数据库可以包括自疾病组织获得的序列信息。在一些实施方案中,通过搜索算法使用以进行获得谱的肽序列分配的数据库可以包括自无疾病的组织获得的序列信息。

[0206] 在一些实施方案中,可以手动验证序列分配谱以通过算法确认正确的片段离子分配。在一些实施方案中,可以使用合成肽标准品确认算法分配的序列。在一些实施方案中,可以将自MHC结合的肽产生的谱与自肽标准品产生的谱比较。例如,比较可以牵涉将片段离子的模式,和任选地片段丰度或强度(基于自疾病组织来源获得的谱的 m/z 值)与参照匹配。在一些实施方案中,手动验证可以确认完整肽序列的序列分配。在一些实施方案中,手动验证可以确认肽序列的部分片段的序列分配。

[0207] 质谱术与基因组数据的关联

[0208] 本文中在一些实施方案中提供的方法包括将MHC结合的肽的质谱术衍生的序列信

息与一组变体编码序列关联,以鉴定疾病特异性免疫原性突变体肽。例如,可以使用MHC结合的肽的质谱术序列进一步选择预测的疾病特异性免疫原性突变体肽的群体。在一些实施方案中,基于质谱术的表位鉴定可以补充基于基因组的免疫原性表位鉴定和/或预测。在一些实施方案中,基于质谱术的表位鉴定可以确认基于基因组的免疫原性表位鉴定和/或预测。

[0209] 可以将自质谱术分析获得的数据与基于疾病组织的基因组和/或转录物组序列分析的免疫原性肽预测关联。通常,将自基于质谱术的测序鉴定的氨基酸序列与预测的免疫原性肽的氨基酸序列比较,以找到包含部分序列比对的区域。在一些实施方案中,经由质谱术鉴定的肽会是与预测的免疫原性肽的序列的精确序列匹配。在一些实施方案中,氨基酸序列的长度可以在经由质谱术鉴定的肽和预测的免疫原性肽的长度之间变化。例如,与预测的免疫原性肽相比,经由质谱术鉴定的肽可以含有对肽的C和/或N端修改的别的氨基酸。或者,与预测的免疫原性肽相比,包含变体氨基酸的经由质谱术鉴定的肽可以在C和/或N端具有更少的氨基酸。在这些例示性实施方案中,变体氨基酸和变体氨基酸周围的序列在经由质谱术鉴定的肽和预测为免疫原性的肽两者中必须是相同的。在一些实施方案中,自将基于质谱术的序列与免疫原性肽预测关联获得的结果将预测的免疫原性序列与经由基于质谱术的测序确认由MHC分子物理呈现的序列匹配。

[0210] 在一些实施方案中,可以通过肽长度进一步过滤获得的质谱术序列鉴定。例如,在一些实施方案中,可以进一步过滤通过质谱术鉴定的MHC结合的肽的群体以仅包括长度为8或9个氨基酸的那些鉴定的肽序列。

[0211] 免疫原性突变体肽的功能性验证

[0212] 可以通过功能性研究进一步验证通过本文中描述的方法鉴定的疾病特异性免疫原性突变体肽。例如,可以基于激活靶向免疫应答(例如由细胞毒性T细胞介导)的能力合成和测试肽。在一些实施方案中,化学合成肽。在一些实施方案中,通过重组方法合成肽。在一些实施方案中,如下合成肽,即首先表达肽前体分子,然后加工所述肽前体分子(例如通过免疫蛋白酶体)以生成感兴趣的肽。可以在进行功能性分析前将合成的肽进行进一步纯化。

[0213] 在一些实施方案中,在体外使用预测的疾病特异性免疫原性合成肽来测试细胞毒性T细胞应答。在一些实施方案中,在体内使用预测的疾病特异性免疫原性合成肽来测试细胞毒性T细胞应答。

[0214] 在一些实施方案中,可以通过小鼠的免疫测试疾病特异性肽的免疫原性。在一些实施方案中,可以通过测量CD8 T细胞应答在免疫后测试疾病特异性肽的免疫原性。在一些实施方案中,可以使用MHCI/肽特异性的Dextramer测量CD8 T细胞应答。在一些实施方案中,可以通过分析肿瘤浸润细胞(TIL)测试疾病特异性肽的免疫原性。

[0215] 在一些实施方案中,可以测量特异性表位和/或细胞表面蛋白的存在。在一些实施方案中,可以测量源自疫苗接种的表位的存在。在一些实施方案中,可以测量干扰素 γ (IFN- γ)。在一些实施方案中,可以测量程序性细胞死亡1(PD-1)。在一些实施方案中,可以测量T细胞免疫球蛋白粘蛋白-3(TIM-3)。在一些实施方案中,可以测量表达特异性蛋白质和/或表位的细胞毒性T细胞。在一些实施方案中,可以测量展现特异性表位的细胞毒性T细胞。

[0216] 在一些实施方案中,可以如下测试疾病特异性肽的免疫原性,即首先在树突细胞

中表达肽,然后测试呈递的抗原被T细胞识别的能力。在一些实施方案中,自患者获得树突细胞,其中在所述患者中鉴定疾病特异性肽。在一些实施方案中,可以如下测试疾病特异性肽的免疫原性,即首先在B淋巴细胞中表达肽,然后测试呈递的抗原被T细胞识别的能力。在一些实施方案中,自患者获得B淋巴细胞,其中在所述患者中鉴定疾病特异性肽。参见例如美国专利号8,349,558。

[0217] 免疫原性肽的组成

[0218] 本公开内容提供了鉴定疾病特异性免疫原性肽的方法。可以基于激活靶向免疫应答(诸如由细胞毒性T细胞介导的)的能力鉴定免疫原性肽。在一些实施方案中,可以使用所鉴定的疾病特异性免疫原性肽的氨基酸序列产生药学上可接受的组合物。在一些实施方案中,组合物可以包含合成的疾病特异性免疫原性肽。在一些实施方案中,组合物可以包含合成的疾病特异性免疫原性突变体肽。在一些实施方案中,组合物可以包含两种或更多种疾病特异性免疫原性肽。在一些实施方案中,组合物可以包含两种或更多种疾病特异性免疫原性突变体肽。在一些实施方案中,两种或更多种疾病特异性免疫原性肽可以激活对两个或更多个独特表位的细胞毒性T细胞应答。

[0219] 在一些实施方案中,组合物可以包含疾病特异性免疫原性肽的前体(诸如蛋白质、肽、DNA和RNA)。在一些实施方案中,疾病特异性免疫原性肽的前体可以产生鉴定的疾病特异性免疫原性肽或针对鉴定的疾病特异性免疫原性肽产生疾病特异性免疫原性肽的前体。在一些实施方案中,疾病特异性免疫原性肽的前体可以是前药。

[0220] 在一些实施方案中,包含疾病特异性免疫原性肽的组合物可以是药学可接受的。在一些实施方案中,包含疾病特异性免疫原性肽的组合物可以进一步包含佐剂。例如,可以利用突变肽作为疫苗(参见Sahin et al., *Int. J. Cancer*, 78:387-9 (1998); Stumliolo et al., *Nature Biotechnol*, 17:555-61 (1999); Rammensee et al., *Immunol Rev* 188:164-76 (2002); 及Hannani et al. *Cancer J* 17:351-358 (2011))。此外,依照特定患者的个人需要,疫苗可以含有个体化成分。在一些实施方案中,疫苗可能对特定患者中预测的免疫原性肽是特异性的。在一些实施方案中,疫苗会含有超过一种免疫原性肽或肽前体。在一些实施方案中,疫苗中使用的肽的长度可以在长度上变化。在一些实施方案中,肽的长度是约7至50个氨基酸(诸如约8、9、10、11、12、13、14、15、17、20、22、25、30、35、40、45或50个氨基酸中的任一项)。在一些实施方案中,肽的长度是约8至12个氨基酸。在一些实施方案中,肽的长度是约8至10个氨基酸。肽可以以其分离的形式利用,或者备选可以将肽添加至MHC分离的肽的末端以生成可以证明更具免疫原性的“长肽”(参见例如Castle et al., *Cancer Res* 72:1081-1091 (2012))。在一些实施方案中,肽也可以是加标签的,或者是融合蛋白,或者是杂合分子。在一些实施方案中,肽是药学可接受盐的形式。

[0221] 在一些实施方案中,疫苗是核酸疫苗。在一些实施方案中,核酸编码免疫原性肽或肽前体。在一些实施方案中,核酸疫苗包含编码免疫原性肽或肽前体的序列侧翼的序列。在一些实施方案中,核酸疫苗包含超过一种免疫原性表位。在一些实施方案中,核酸疫苗是基于DNA的疫苗。在一些实施方案中,核酸疫苗是基于RNA的疫苗。在一些实施方案中,基于RNA的疫苗包含mRNA。在一些实施方案中,基于RNA的疫苗包含裸mRNA。在一些实施方案中,基于RNA的疫苗包含经修饰的mRNA(例如,使用鱼精蛋白提供保护而免于降解的mRNA、含有经修饰的5' CAP结构的mRNA、或含有经修饰的核苷酸的mRNA)。在一些实施方案中,基于RNA的疫

苗包含单链mRNA。

[0222] 多核苷酸可以是基本上纯的,或包含在合适的载体或投递系统中。合适的载体和投递系统包括病毒,诸如基于腺病毒、痘苗病毒、逆转录病毒、疱疹病毒、腺伴随病毒或含有超过一种病毒的元件的杂合物。非病毒投递系统包括阳离子脂质和阳离子聚合物(例如阳离子脂质体)。在一些实施方案中,可以使用物理投递,诸如用“基因枪”。

[0223] 在一些实施方案中,可以使用本文中描述的肽生成突变体肽特异性治疗剂,诸如抗体治疗剂。例如,可以使用突变体肽制备和/或鉴定特异性识别突变体肽的抗体。可以使用这些抗体作为治疗剂。已经使用合成短肽生成蛋白质反应性抗体。用合成肽免疫的优点是可以使用无限量的纯稳定抗原。此方法牵涉合成短肽序列,将它们与大载体分子偶联,并且用肽-载体分子免疫选择的动物。抗体的性质取决于一级序列信息。通常可以通过仔细选择序列和偶联方法产生对期望肽的良好响应。大多数肽可以引发良好的响应。抗肽抗体的优点是可以在测定突变体肽的氨基酸序列后立即制备它们,并且可以特异性靶向蛋白质的特定区域以进行抗体生成。由于已经对突变体肽筛选高免疫原性,存在有较高的下述几率,即所得的抗体会识别肿瘤环境中的天然蛋白质。如在疫苗情况中一样,肽的长度是考虑的另一项重要的因素。大致上,10-15个残基的肽对于抗肽抗体生产是最佳的;肽越长越好,因为可能的表位的数目随肽长度而增加。然而,长肽增加了合成、纯化和与载体蛋白偶联的困难。抗体的质量依赖于肽的质量。肽产物中含有的副产物可导致低质量的抗体。

[0224] 肽-载体蛋白偶联是牵涉高滴度抗体生产的另一项因素。大多数偶联方法依赖于氨基酸中的反应性官能团,诸如-NH₂、-COOH、-SH和酚-OH。定点偶联是选择的方法。可以与通过本发明的方法鉴定的肽一起利用在抗肽抗体生成中使用的任何合适的方法。两种此类已知的方法是多重抗原性肽系统(MAP)和脂质核心肽(LCP法)。MAP的优点是缀合方法不是必需的。不将载体蛋白或连接键引入免疫的宿主中。一项缺点是肽的纯度更难以控制。另外,MAP可以绕过一些宿主中的免疫应答系统。已知LCP方法提供比其它抗肽疫苗系统更高的滴度,如此可以是有利的。

[0225] 本文还提供了包含本文中公开的疾病特异性免疫原性突变体肽的分离的MHC/肽复合物。例如,可以使用此类MHC/肽复合物鉴定抗体、可溶性TCR、或TCR类似物。这些抗体的一种类型已经称作TCR模拟物,因为它们是在特定HLA环境的背景中结合来自肿瘤关联抗原的肽的抗体。已经显示了此类型的抗体介导在其表面上表达复合物的细胞的裂解,以及保护小鼠免于表达复合物的植入的癌细胞系(参见例如Wittman et al., J. of Immunol. 177: 4187-4195 (2006))。TCR模拟物作为IgG mAb的一项优点是可以实施亲和力成熟,并且经由目前的Fc域将分子与免疫效应器功能偶联。也可以使用这些抗体将治疗性分子靶向肿瘤,诸如毒素、细胞因子或药物产品。已经使用肽,诸如使用基于非杂交瘤的抗体生成或有结合能力的抗体片段,诸如噬菌体上的抗肽Fab分子的生成使用本发明的方法选择的那些肽开发出其它类型的分子。还可以将这些片段与用于肿瘤投递的其它治疗性分子缀合,诸如抗肽MHC Fab-免疫毒素缀合物、抗肽MHC Fab-细胞因子缀合物和抗肽MHC Fab-药物缀合物。

[0226] 包括免疫原性疫苗的治疗方法

[0227] 本公开内容提供了包含免疫原性疫苗的治疗方法。在一些实施方案中,提供了治疗疾病(诸如癌症)的方法,其可以包括对个体施用有效量的包含免疫原性肽的组合物。在一些实施方案中,提供了治疗疾病(诸如癌症)的方法,其可以包括对个体施用有效量的包

含免疫原性肽的前体的组合物。在一些实施方案中,免疫原性疫苗可以包含药学可接受的疾病特异性免疫原性肽。在一些实施方案中,免疫原性疫苗可以包含疾病特异性免疫原性肽的药学可接受的前体(诸如蛋白质、肽、DNA和RNA)。在一些实施方案中,提供了治疗疾病(诸如癌症)的方法,其可以包括对个体施用有效量的特异性识别疾病特异性免疫原性突变体肽的抗体。在一些实施方案中,提供了治疗疾病(诸如癌症)的方法,其可以包括对个体施用有效量的特异性识别疾病特异性免疫原性突变体肽的可溶性TCR或TCR类似物。

[0228] 在一些实施方案中,癌症是下列任一种:癌瘤、淋巴瘤、母细胞瘤(blastoma)、肉瘤、白血病、鳞状细胞癌、肺癌(包括小细胞肺癌、非小细胞肺癌、肺腺癌和鳞状肺癌)、腹膜癌、肝细胞癌、胃癌或胃癌(包括胃肠癌)、胰腺癌、成胶质细胞瘤、宫颈癌、卵巢癌、肝癌、膀胱癌、肝癌、乳腺癌、结肠癌、黑素瘤、子宫内膜或子宫癌、唾液腺癌、肾癌或肾癌、肝癌、前列腺癌、外阴癌、甲状腺癌、肝的癌、头和颈癌、结肠直肠癌、直肠癌、软组织肉瘤、卡波西(Kaposi)氏肉瘤、B细胞淋巴瘤(包括低级/滤泡性非何杰金淋巴瘤(NHL)、小淋巴细胞性(SL)NHL、中级/滤泡性NHL、中等级弥漫性NHL、高级成免疫细胞性NHL、高级成淋巴细胞性NHL、高级小非切割细胞NHL、大体积疾病NHL、套细胞淋巴瘤、AIDS相关淋巴瘤、和沃尔登斯特伦(waldenstrom)巨球蛋白血症)、慢性淋巴细胞性白血病(CLL)、急性成淋巴细胞性白血病(ALL)、骨髓瘤,多毛细胞白血病、慢性骨髓细胞白血病和移植后淋巴增殖性病征(PTLD)以及与斑痣性错构瘤病相关的异常血管增生、水肿(诸如与脑肿瘤相关)和麦格氏(Meigs)综合征

[0229] 本文中描述的方法在个性化医学背景下是特别有用的,其中使用通过本文中描述的任一种方法获得的疾病特异性免疫原性突变体肽开发用于同一个体的治疗剂(诸如疫苗或治疗性抗体)。因此,例如,在一些实施方案中,提供了治疗个体中的疾病(诸如癌症)的方法,其包括:a)鉴定个体中的疾病特异性免疫原性突变体肽;并且b)合成肽或肽前体;并且c)对个体施用肽。在一些实施方案中,提供了治疗个体中的疾病(诸如癌症)的方法,其包括:1)鉴定个体中的疾病特异性免疫原性突变体肽;b)生成特异性识别突变体肽的抗体;并且c)对个体施用肽。在一些实施方案中,鉴定步骤将序列特异性变体鉴定方法与免疫原性预测方法组合。在一些实施方案中,鉴定步骤将序列特异性变体鉴定方法与质谱术组合。可以使用本文中描述的鉴定疾病特异性免疫原性突变体肽的任何方法进行本文中描述的治疗方法。在一些实施方案中,方法还包括自个体获得疾病组织的样品。

[0230] 可以使用本文中提供的方法治疗已经诊断具有或怀疑具有癌症的个体(例如人)。在一些实施方案中,个体可以是人。在一些实施方案中,个体可以是至少约35、40、45、50、55、60、65、70、75、80或85岁龄的至少约任一项。在一些实施方案中,个体可以是男性。在一些实施例中,个体可以是女性。在一些实施例中,个体可以已经拒绝手术。在一些实施方案中,个体可以是医学上不能手术的。在一些实施方案中,个体可以处于Ta、Tis、T1、T2、T3a、T3b或T4的临床阶段。在一些实施方案中,癌症可以是复发性的。在一些实施方案中,个体可以是展现出与癌症相关的一种或多种症状的人。在一些实施方案中,个体可以在遗传上或在其它方面有形成癌症的素因(例如具有风险因素)。

[0231] 可以在佐剂背景中实施本文中提供的方法。在一些实施方案中,在新辅助背景中实施方法,即可以在初步/确定性疗法前实施方法。在一些实施方案中,使用方法治疗先前已经治疗的个体。可以使用本文中提供的任何治疗方法治疗先前尚未治疗的个体。在一些

实施方案中,使用该方法作为一线疗法。在一些实施方案中,使用该方法作为二线疗法。

[0232] 在一些实施方案中,提供了降低个体中预先存在的癌症肿瘤转移(诸如肺转移或转移至淋巴结)的发生或负担的方法,其包括对个体施用有效量的包含免疫原性疫苗的组合物。

[0233] 在一些实施方案中,提供了延长个体中的癌症疾病进展前时间的方法,其包括对个体施用有效量的包含免疫原性疫苗的组合物。

[0234] 在一些实施方案中,提供了延长具有癌症的个体的存活的方法,其包括对个体施用有效量的包含免疫原性疫苗的组合物。

[0235] 在一些实施方案中,在包含免疫原性疫苗的组合物外,可以施用至少一种或多种化学治疗剂。在一些实施方案中,一种或多种化学治疗剂可以(但不一定)属于不同类别的化学治疗剂。

[0236] 在一些实施方案中,提供了治疗个体中的疾病(诸如癌症)的方法,其包括施用:a)免疫原性疫苗,和b)免疫调节剂。在一些实施方案中,提供了治疗个体中的疾病(诸如癌症)的方法,其包括施用:a)免疫原性疫苗,和b)检查点蛋白的拮抗剂。在一些实施方案中,提供了治疗个体中的疾病(诸如癌症)的方法,其包括施用:a)免疫原性疫苗,和b)程序性细胞死亡1(PD-1)的拮抗剂,诸如抗PD-1。在一些实施方案中,提供了治疗个体中的疾病(诸如癌症)的方法,其包括施用:a)免疫原性疫苗,和b)程序性死亡配体1(PD-L1)的拮抗剂,诸如抗PD-L1。在一些实施方案中,提供了治疗个体中疾病(诸如癌症)的方法,其包括施用:a)免疫原性疫苗,和b)细胞毒性T淋巴细胞相关蛋白4(CTLA-4)拮抗剂,诸如抗CTLA-4。

实施例1

[0237] 此实施例演示了用于预测免疫原性肽表位的例示性方法。

[0238] 对MC-38和TRAMP-C1小鼠肿瘤细胞系实施全外显子组测序以鉴定肿瘤特异性点突变。相对于参照小鼠基因组调用编码变体以分别鉴定MC-38和TRAMP-C1中的4285和949种非同义变体。随后,通过RNA-Seq分析针对基因表达过滤数据,并且揭示了分别在MC-38和TRAMP-C1中表达1290和67种突变基因。使用NETMHC-3.4算法鉴定出MC-38中170种预测的新表位和TRAMP-C1肿瘤中6种预测的新表位。

[0239] 接着,使用转录物组生成的FASTA数据库的质谱分析揭示了由MC-38细胞系呈现的797种独特的H-2Kb表位和725种独特的H-2Db表位,以及由TRAMP-C1细胞系呈现的477种独特的H-2Kb表位和332种独特的H-2Kb表位。观察到自丰富的转录物衍生的肽更可能由MC-38(图2小图A)和TRAMP-C1(图2小图B)细胞中的MHCI呈递。

[0240] 分别在MC-38和TRAMP-C1中的1290和67种氨基酸变化中,通过质谱术仅发现在MHCI上呈递7种(MC-38中的7种和TRAMP-C1中的0种)(表1)。在MC-38细胞中通过质谱术也检测到自癌症睾丸自身抗原MAGE-D1衍生的一种表位。手动验证这些肽,并且与肽的合成生成型式比较以实现准确性。预测除了一种外的所有新表位结合MHCI($IC_{50} < 500nm$,表1)。野生型(WT)和突变体转录物两者均由肿瘤细胞表达,并且虽然预测大多数相应的WT肽也结合MHCI,但是通过质谱术仅检测到它们中的三种。

[0241] 虽然对MHCI的肽结合亲和力和免疫原性之间存在有关联,但其它因素也有所贡献。例如,突变氨基酸与TCR的相互作用可能对于将突变的肽识别为“非自身”是必需的。当

相应的WT肽也在MHC I上呈现时尤其如此。7种新表位中的5种展现出高结合预测评分(根据NETMHC3.4, $IC_{50} < 50nM$, 表1)。其它新表位展现出较低的结合预测得分,这提示了它们可以是不太有免疫原性的。利用H-2Db和H-2Kb的发表的晶体结构和基于Rosetta的算法对与MHC I的复合物中的每种突变体肽建模并且分析每种新表位中的突变体残基与T细胞受体相互作用的潜力。通常,展示的肽的TCR识别由与肽残基3到7的相互作用介导。仅在Reps1和Adpgk肽中具有高结合得分的肽在此范围内具有突变。结构建模还预测突变的残基朝向溶剂界面定向,并且如此判断为具有呈免疫原性的良好潜力(表1和图3)。另一方面,Irgq、Aatf、Dpagt1新表位中的突变存在于肽的C末端附近,其可能落在TCR结合区外部,并且提示了这些新表位不可能是免疫原性的(表1和图3)。

[0242] 接着,通过与佐剂组合用编码突变表位的长肽免疫野生型C57BL/6小鼠评估突变肿瘤抗原的免疫原性,并且使用MHC I/肽特异性Dextramer测量CD8 T细胞应答。如图4小图A中显示的,与仅佐剂的组相比,6种肽中的3种引发CD8 T细胞应答。基于结构和结合亲和力预测,预测Reps1和Adpgk是免疫原性的,并且两者都引发强的CD8 T细胞应答。在预测为非免疫原性的4种肽中,仅Dpagt1诱导弱的CD8 T细胞应答。

[0243] 通过分析肿瘤浸润细胞(TIL)在肿瘤的背景中确认了这些突变体肽的免疫原性。观察到对Reps1、Adpgk和Dpagt1特异性的T细胞在肿瘤床中是富集的(图4小图B)。虽然存在有异质性,但是Adpgk特异性CD8 T细胞在三种中是最丰富的,并且这对MC-38肿瘤是特异性的,因为在同基因TRAMP-C1肿瘤中检测不到Adpgk特异性CD8T细胞。令人感兴趣地,通过质谱术鉴定的自单一癌症睾丸自身抗原(MAGE-D1)衍生的肽显示较差的免疫原性,并且在肿瘤床中检测不到对MAGE-D1特异性的CD8T细胞(数据未显示)。

[0244] 通常分析体积TIL以监测抗肿瘤应答,这可以不提供真正的评估,因为仅一定分数的TIL是肿瘤特异性的。使用针对三种免疫原性肽的MHC I/肽特异性Dextramer检查与体积TIL相比抗肿瘤TIL的频率和表型。肿瘤特异性CD8T细胞浸润肿瘤的频率首先增加,并且随肿瘤进一步增长而下降,这提示了肿瘤生长与肿瘤中肿瘤特异性CD8 T细胞的频率呈反相关(图4小图C)。令人感兴趣地,大多数($76.9 \pm 7.1\%$)肿瘤特异性CD8 TIL共表达PD-1和TIM-3,与体积TIL相比T细胞耗尽的标志物($52.6 \pm 3.6\%$) (图4小图D)。肿瘤特异性CD8 TIL也表达较高的表面PD-1水平。

[0245] 为了测定针对新表位诱导的CD8 T细胞是否能提供保护性抗肿瘤免疫,用突变的肽疫苗免疫健康的小鼠,随后用MC-38肿瘤细胞攻击。与单独的佐剂相比,肿瘤生长在疫苗组中的大多数动物中受到完全抑制(图5小图A)。在此实验中生长出肿瘤的唯一动物实际上不响应疫苗,强烈支持对突变的肽特异性的CD8 T细胞应答赋予保护的可能性(图5小图A)。

[0246] 接着,评估新表位特异性CD8 T细胞应答以观察它们是否可以在免疫后在携带肿瘤的小鼠中进一步扩增。在单次免疫后,与未免疫的健康动物相比,Adpgk反应性CD8 T细胞的频率在携带肿瘤的小鼠的脾中显著增加(图4小图B)。也观察到在肿瘤中的总CD8 TIL中Adpgk特异性CD8 T细胞的积累的几乎三倍增加(图5小图B)。肽疫苗接种还增加了肿瘤中CD45⁺细胞和CD8⁺T细胞的总体浸润,这导致肿瘤中总活细胞中新表位特异性CD8 T细胞的频率的几乎20倍增加(图5小图C)。

[0247] 此外,分析了通过疫苗接种诱导的肽特异性细胞的表型。发现TIM-3⁺PD-1⁺Adpgk特异性CD8 TIL的频率在疫苗接种后降低,并且这些细胞上PD-1和TIM-3的表达也是降低

的(图5小图D和图5小图E)。这可以是佐剂效应,因为在仅佐剂的组中也看到它。此结果提示了,肿瘤特异性T细胞在疫苗接种后展现出消耗较少表型,并且这得到接种疫苗的肿瘤中较高的IFN- γ 表达性CD8和CD4 TIL的百分比进一步确认(图5小图F)。

[0248] 最后,评估这些疫苗诱导的肿瘤特异性CD8 T细胞的定性和定量变化是否可以转化为建立的肿瘤的消退。即使在此较为困难的治疗背景中,与未处理的对照或仅佐剂的组相比,接种疫苗的小鼠显示显著的对肿瘤生长的抑制(图5小图G)。如此,用具有预测的新表位的简单肽疫苗接种产生了足够的T细胞免疫以排斥以前建立的肿瘤。

[0249] 方法

[0250] 对H-2b-背景的两种鼠细胞系:TRAMP-C1(ATCC)和MC-38(Academisch Ziekenhuis Leiden)的H-2Kb和H-2Db配体组进行MHC I肽概况分析。如先前描述的那样制备自C57BL/6小鼠衍生的细胞。关于制备细胞系的方法的完整描述,参照美国专利申请流水号13/087,948和美国专利申请流水号11/00,474。使用两种不同的抗体免疫沉淀每份样品的MHC I分子,以分别提取H-2Kb特异性和H-2Db特异性肽。通过使用180分钟梯度的反相层析(nanoAcquity UPLC系统,Waters,Milford,MA)分离肽。通过配备有电喷雾离子化(ESI)源的LTQ-Orbitrap Velos杂合质谱仪(Thermo Fisher Scientific,Bremen,Germany)中的数据依赖性获得(DDA)分析洗脱的肽。使用下述方法获得质谱数据,所述方法包括在Orbitrap(对于前3为R=30,000,对于前5为R=60,000)中的高质量准确性的全扫描(调查扫描),接着进行对5种最丰富的前体离子(前5)在Orbitrap(R=7500)中或者对3种最丰富的前体离子(前3)在LTQ中的MS/MS(概况)扫描。对每套样品进行7份重复试样注射和分析。

[0251] 在LTQ-Orbitrap Elite质谱仪(ThermoFisher,Bremen,Germany)上分析对应于鉴定突变体MC-38和TRAMP-C1抗原肽的合成肽,并使用ADVANCE源(Michrom-Bruker,Fremont,CA)以1.2kV的喷射电压离子化。使用由以下组成的方法获得质谱数据:以60,000M/ Δ M的分辨率在m/z 400在Orbitrap中的一次完全MS扫描(375-1600m/z),接着是在肽片段离子的LTQ中的MS/MS(形心)扫描。

[0252] 使用来自MC-38和TRAMP-C1癌细胞系的1 μ g总RNA使用TruSeq RNA样品制备试剂盒(Illumina,CA)产生RNA-Seq文库。自细胞系纯化总RNA,并将其片段化成200-300个碱基对(bp),平均长度为260bp。将RNA-Seq文库多路复用(每道两种),并且按照制造商的推荐(Illumina,CA)在HiSeq 2000上测序。

[0253] 每份样品产生大于约5000万个配对末端(2x100bp)测序读段。使用SureSelect Human All Exome试剂盒(50Mb)(Agilent,CA)实施外显子组捕捉。然后,使用HiSeq测序试剂盒(200个循环)在HiSeq 2000(Illumina,CA)上对外显子组捕捉文库测序。

[0254] 自MC-38测序9290万(M)个RNA片段,并且自TRAMP-C1测序65.3M个RNA片段。对于外显子组测序,自每种细胞系测序60M个读段。使用GSNAP(Wu and Nacu,Bioinformatics,2010,v.26,873-881)将读段定位到小鼠基因组(NCBI 37或mm9版本)。仅保留独特定位的读取以进一步分析。在MC-38样品中独特定位80.6M个RNA片段,并且在TRAMP-C1样品中定位57.6M个RNA片段。独特定位MC-38中的50.9M个外显子片段和TRAMP-C1中的52M个片段。为了获得小鼠基因模型,使用GMAP将Refseq小鼠基因定位至mm9基因组,然后使用基因组序列做出基因模型。

[0255] 使用GATK1调用基于外显子组-seq的变体。保留具有10%或更大的等位频率的变

体。使用变体效应预测器工具2在对转录物的影响方面注释变体。仅保留可以解读氨基酸变化的变体。为了获得具有表达证据的变体,用RNA-Seq读取比对对基于外显子组的变体位置检查变异的证据。保留了变体,所述变体通过超过2个RNA-Seq读段证实并且基于RNA-Seq以10%或更多的等位频率表达。

[0256] 对于每种氨基酸变异,产生变体全蛋白序列以形成一组推定蛋白,以充当用于搜索LC-MS谱的参照数据库。在缺乏单体型信息的情况中,相同蛋白质中的多种变异会以数据库中的不同变体蛋白质为特征。

[0257] 使用Mascot算法版本2.3.02 (MatrixScience, London, UK) 针对多环靶物诱饵数据库Uniprot版本2011_12或转录物组生成的FASTA数据库提交串联质谱结果用于蛋白质数据库检索;包括鼠蛋白和常见的实验室污染物,诸如胰蛋白酶。在无酶特异性、甲硫氨酸氧化(+15.995Da)和20ppm前体离子质量耐受性的情况下搜索数据。

[0258] 对于在LTQ或Orbitrap中获得的MS/MS数据,分别在0.8Da或0.05Da处规定片段离子质量耐受性。使用线性判别算法(LDA)将搜索结果过滤至5%的估计的肽假发现率(FDR)。为了突变体肽鉴定中较高的置信度,通过肽长度(对于H-2Kb数据为8和对于H-2Db为9)或者采用规则表达以分离具有以下充分表征的锚定基序H-2Kb:XXXX[FY]XX[MILV](SEQ ID NO.22)和H-2Db:XXXX[N]XXX[MIL](SEQ ID NO.23)的肽进一步过滤数据。产生合成肽以验证序列。

[0259] 为了产生第一模型,基于模型结构中突变体肽和肽之间的序列相似性,自PDB选择肽-MHC复合物结构。对于每种突变体肽模型,使用以下PDB代码:Reps1、2ZOL 4;Adpgk、1HOC 5;Dpagt1、3P9L 6;Cpne1、1JUF7;Irgq、1FFN 8;Aatf、1BZ9。由于缺乏可以作为合理起始模型使用的与10聚体肽的复合物中的发表的H-2Kb晶体结构,没有对Med12肽建模。然后,使用COOT 10将肽修饰为突变体形式。然后,使用Rosetta FlexPepDock网络服务器11优化这些第一模型,并选择最高评分模型进行展示。

[0260] 还检查了每种肽的顶部评分FlexPepDock模型,并且发现主链定位对于产生的前10种模型是相似的。使用Pymol (Schrödinger, LLC) 产生肽-MHC图像。

[0261] 给龄期匹配的6-8周龄C57BL/6小鼠(The Jackson Laboratory)腹膜内注射PBS中各自与佐剂(50μg抗CD40 Ab克隆FJK45加上100mg聚(I:C)(Invivogen))组合的50mg长肽。在第0天和第14天以及最后一次注射后一周免疫小鼠,使用血液或脾细胞检测Ag特异性CD8T细胞。为了鉴定肽特异性T细胞,用PE缀合的Dextramer (MHC I/肽复合物;Immudex, Denmark)将细胞染色20分钟,接着用细胞表面标志物CD3、CD4、CD8和B220 (BD Biosciences)染色。肽序列如下

[0262] Reps1:GRVLELFRAAQLANDVVLQIMELCGATR (SEQ ID NO.1);

[0263] Adpgk:GIPVHLELASMTNMELMSSIVHQVFPPT (SEQ ID NO.2);

[0264] Dpagt1:EAGQSLVISASIIIVFNLELEGDYR (SEQ ID NO.3);

[0265] Aatf:SKLLSFMPIIDHTTMSDDARTELFRRS (SEQ ID NO.4);

[0266] Irgq:KARDETAALLNSAVLGAAPLFVPPAD (SEQ ID NO.5);

[0267] Cpne1:DFTGSNGDPSSPYSLHYLSPTGVNEY (SEQ ID NO.6);

[0268] Med12:GPQEKQQRVELSSISNFQAVSELLTFE (SEQ ID NO.7)。

[0269] 给C57BL/6小鼠在右体侧上皮植入1X10⁵个MC-38肿瘤细胞。分离整个肿瘤,并用

胶原酶和DNA酶消化以分离TIL。用Dextramer (如上文描述),接着是针对CD45、Thy1.2、CD4、CD8 (BD Biosciences)、PD-1 (eBiosciences) 和TIM-3 (R&D Systems) 的抗体染色TIL。使用活/死染色对活细胞门控。

[0270] 给所有动物皮下(右后体侧)接种汉克氏平衡盐溶液(HBSS)和无酚红基质胶(Becton Dickinson Bioscience, San Jose, CA)的悬浮液中 1×10^5 个MC-38细胞。对于预防性研究,在肿瘤接种前3周用佐剂(50mg抗CD40加100mg聚(I:C)或佐剂与各50 μ g Repls1、Adpgk和Dpagt1肽免疫小鼠。在用肿瘤细胞接种前一天,在血液中测量肽特异性CD8 T细胞的诱导。对于携带肿瘤的小鼠中的疫苗接种,在用 1×10^5 个MC-38肿瘤细胞接种后10天(在研究中包括仅在第10天时具有约100-150 mm^3 体积的肿瘤),用佐剂或佐剂与各50 μ g Repls1、Adpgk和Dpagt1肽注射小鼠。一周两次收集测量和重量。每天称重展现出动物的第一次体重的超过15%的重量减轻的动物,并且如果它们减轻其第一次体重的超过20%,那么实施安乐死。

[0271] 较频繁,根据严重性直至每天观察到展现出不利临床问题的动物,并且并且如果濒死,那么实施安乐死。如果肿瘤体积超过3,000 mm^3 ,或在3个月后(若未形成肿瘤),那么对小鼠实施安乐死。贯穿整个研究,一周两次实施所有小鼠的临床观察。

[0272] 表

[0273] 表1:在MC-38细胞系中的MHCI上呈递的突变体肽的汇总。

[0274]

基因	肽*	MHC 等位基因	IC50 (突变型) (nM)	IC50 (野生型) (nM)	突变位置	免疫原性 预测
Dpagt1	SIIVFNLV (SEQ ID NO. 8)	H-2Kb	8	34	锚定(P8)	-
	SIIVFNLL (SEQ ID NO. 9)					
Repls1	AQLPNDVVL (SEQ ID NO. 10)	H-2Db	9	100	溶剂(P4)	+
	AQLANDVVL (SEQ ID NO. 11)					

[0275]

Adpgk	ASMTNRELM (SEQ ID NO. 12)	H-2Db	2	3	溶剂(P6)	+
	ASMTNMELM (SEQ ID NO. 13)					
Cpne1	SSPDSLHYL (SEQ ID NO. 14)	H-2Db	211	685	溶剂(P4)	-
	SSPYSLHYL (SEQ ID NO. 15)					
Irgq	AALLNSAGL (SEQ ID NO. 16)	H-2Db	3	52	溶剂(P8)	-
	AALLNSAVL (SEQ ID NO. 17)					
Aatf	MAPIDHTAM (SEQ ID NO. 18)	H-2Db	30	102	溶剂(P8)	-
	MAPIDHTTM (SEQ ID NO. 19)					
Med12	DPSSSVLFED (SEQ ID NO. 20)	H-2Kb	38300	39411	无结构	-
	DPSSSVLF EY (SEQ ID NO. 21)					

[0276] *每种基因的顶部列出的序列是WT序列,并且每种基因的底部列出的序列是突变体序列。

序列表

<110> 豪夫迈·罗氏有限公司

<120> 免疫原性突变体肽筛选平台

<130> 146392027600

<140> 未指派

<141> 随本文

<150> 62/048,742

<151> 2014-09-10

<160> 23

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 28

<212> PRT

<213> 人工序列

<220>

<223> 合成构建物

<400> 1

Gly Arg Val Leu Glu Leu Phe Arg Ala Ala Gln Leu Ala Asn Asp Val

1 5 10 15

Val Leu Gln Ile Met Glu Leu Cys Gly Ala Thr Arg

20 25

<210> 2

<211> 28

<212> PRT

<213> 人工序列

<220>

<223> 合成构建物

<400> 2

Gly Ile Pro Val His Leu Glu Leu Ala Ser Met Thr Asn Met Glu Leu

1 5 10 15

Met Ser Ser Ile Val His Gln Gln Val Phe Pro Thr

20 25

<210> 3

<211> 25

<212> PRT

<213> 人工序列

<220>

<223> 合成构建物

<400> 3

Glu Ala Gly Gln Ser Leu Val Ile Ser Ala Ser Ile Ile Val Phe Asn
 1 5 10 15
 Leu Leu Glu Leu Glu Gly Asp Tyr Arg
 20 25

<210> 4

<211> 26

<212> PRT

<213> 人工序列

<220>

<223> 合成构建物

<400> 4

Ser Lys Leu Leu Ser Phe Met Ala Pro Ile Asp His Thr Thr Met Ser
 1 5 10 15
 Asp Asp Ala Arg Thr Glu Leu Phe Arg Ser
 20 25

<210> 5

<211> 26

<212> PRT

<213> 人工序列

<220>

<223> 合成构建物

<400> 5

Lys Ala Arg Asp Glu Thr Ala Ala Leu Leu Asn Ser Ala Val Leu Gly
 1 5 10 15
 Ala Ala Pro Leu Phe Val Pro Pro Ala Asp
 20 25

<210> 6

<211> 26

<212> PRT

<213> 人工序列

<220>

<223> 合成构建物

<400> 6

Asp Phe Thr Gly Ser Asn Gly Asp Pro Ser Ser Pro Tyr Ser Leu His
 1 5 10 15
 Tyr Leu Ser Pro Thr Gly Val Asn Glu Tyr
 20 25

<210> 7

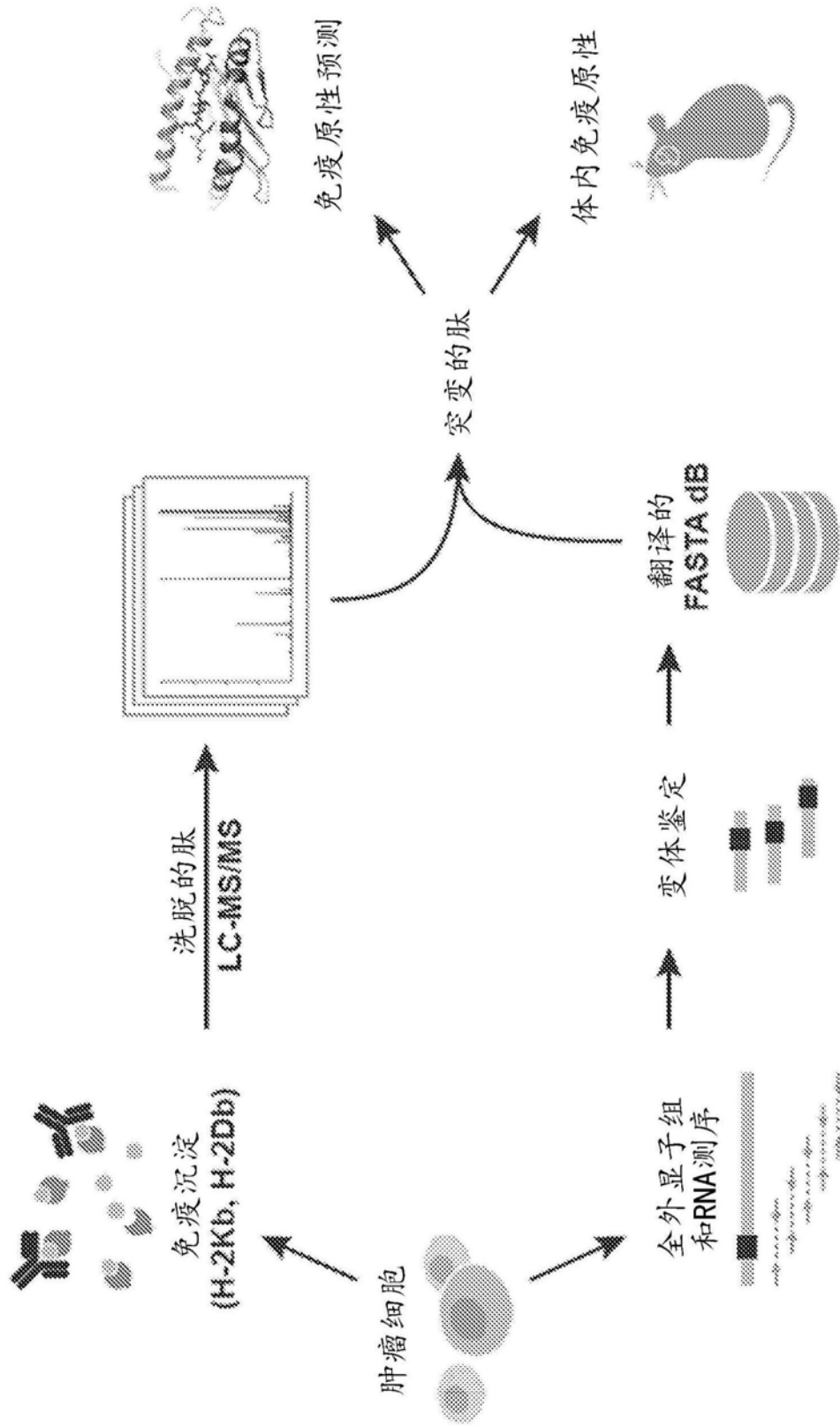


图1

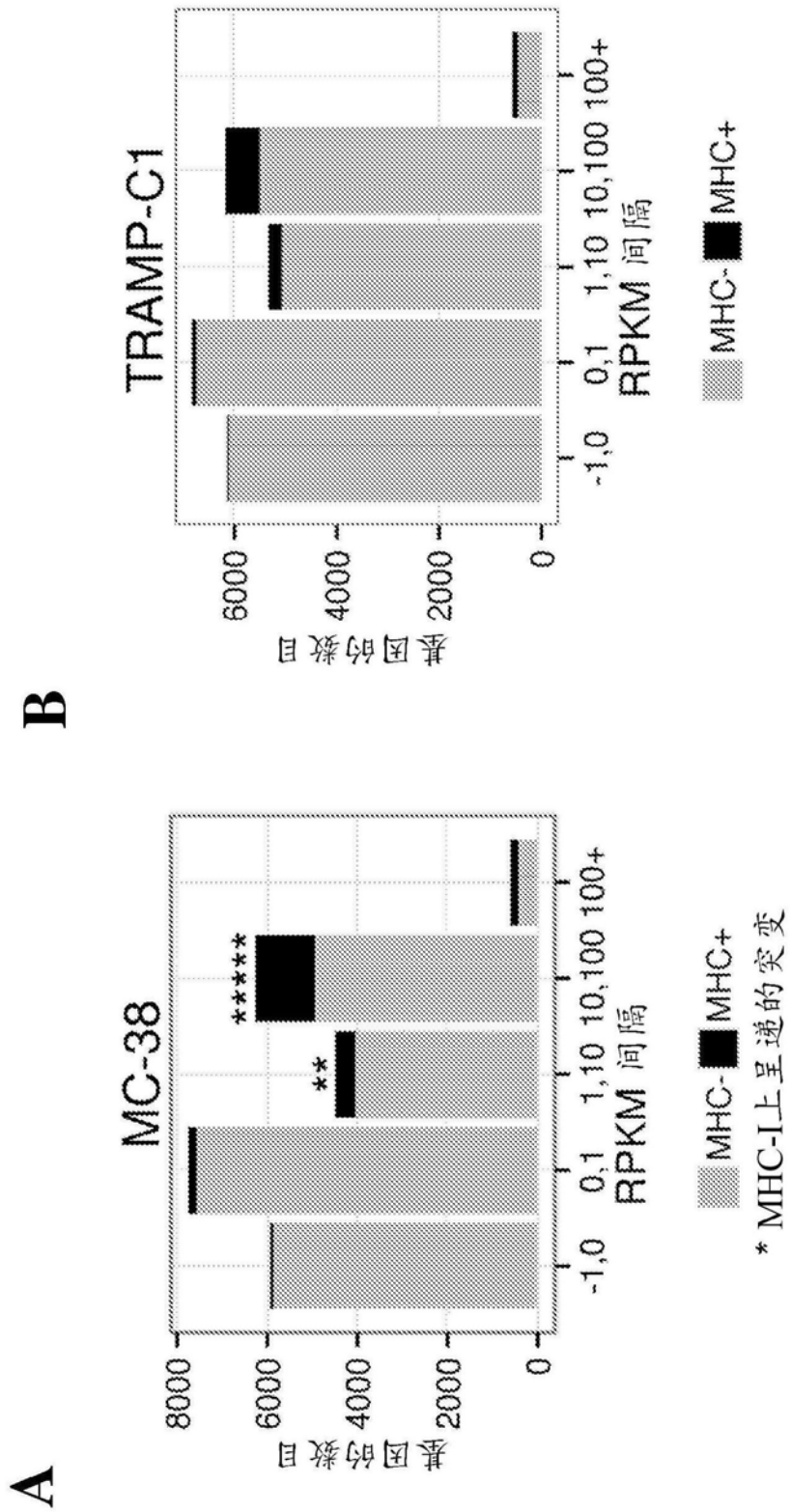


图2

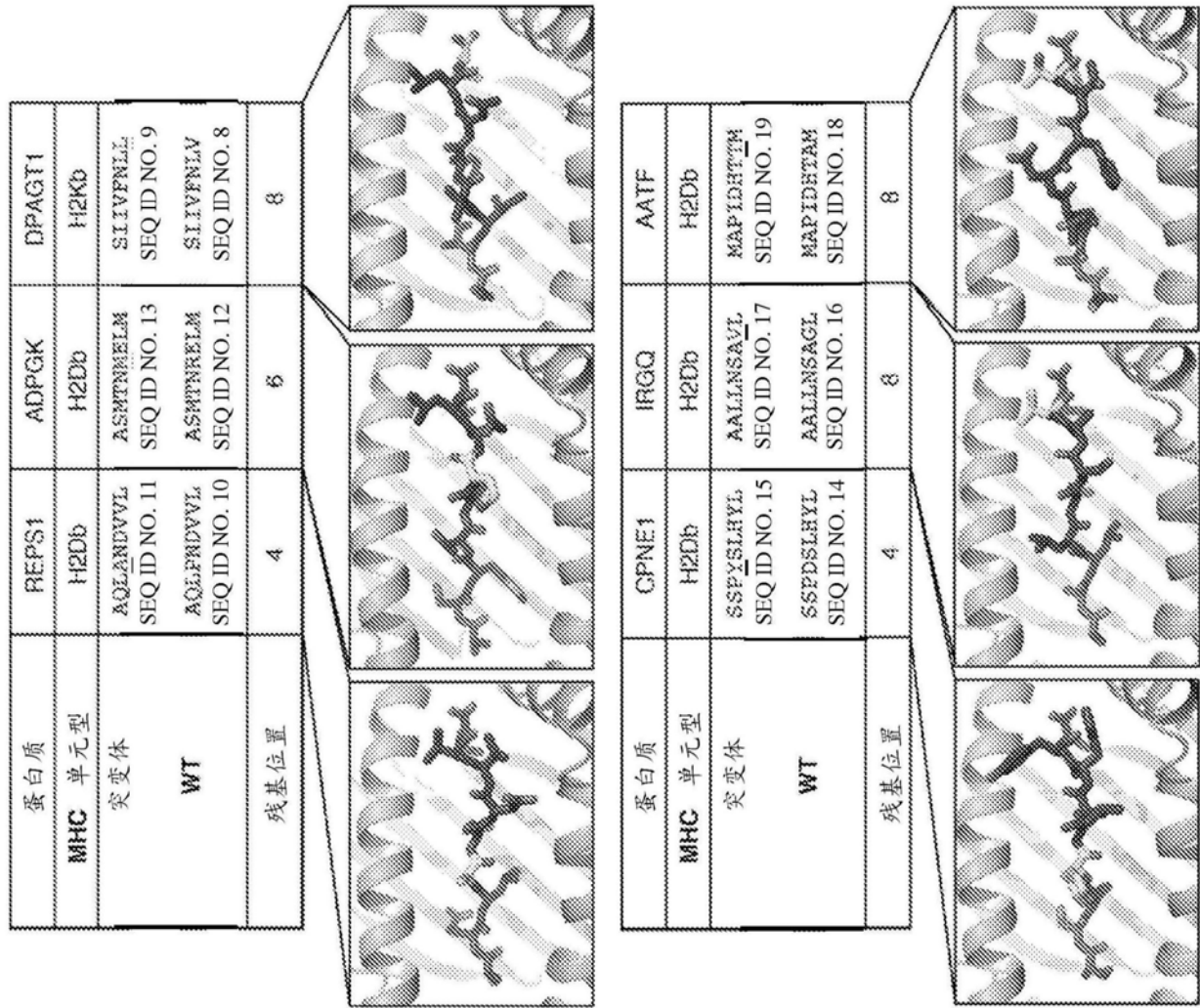


图3

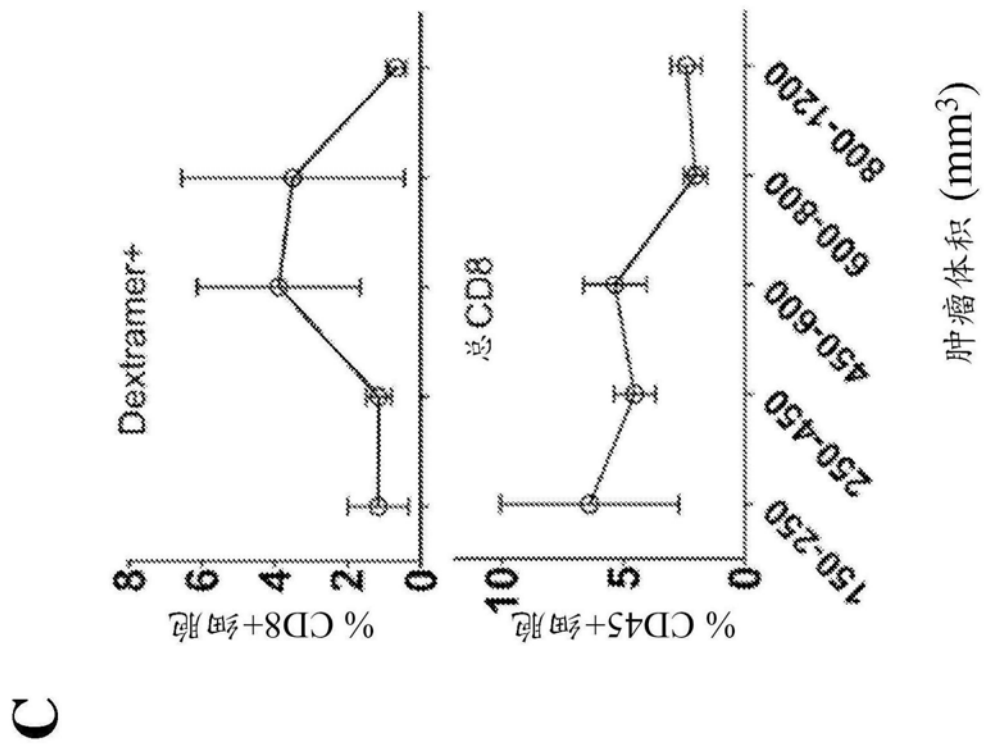
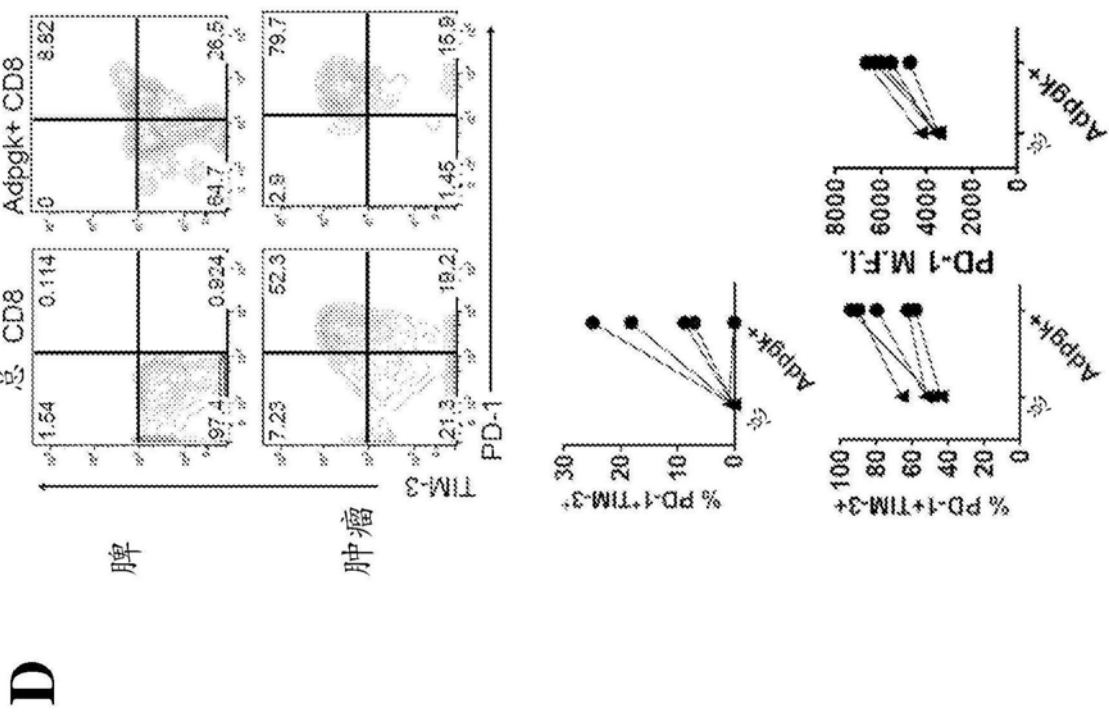


图4(续)

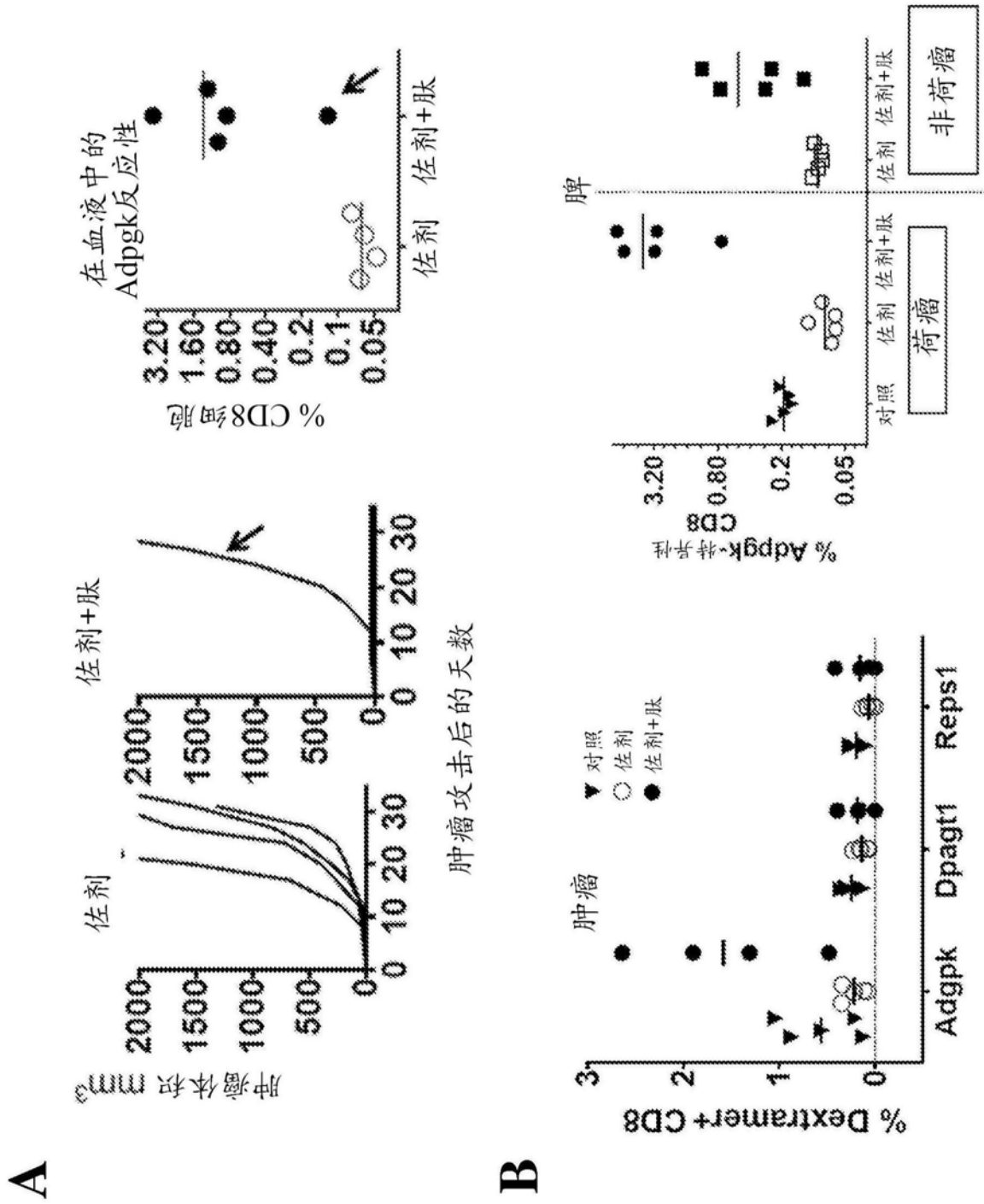


图5

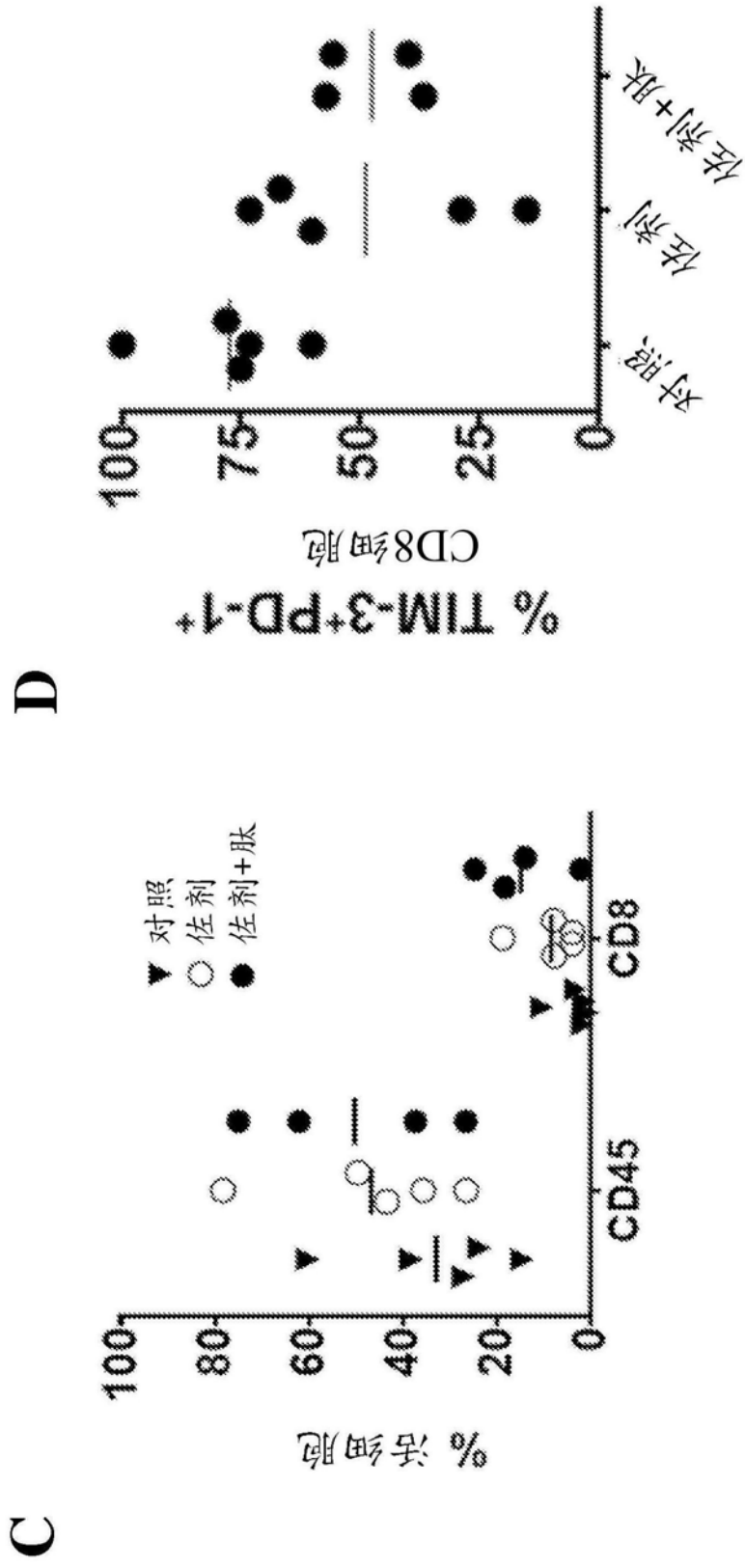


图5(续)

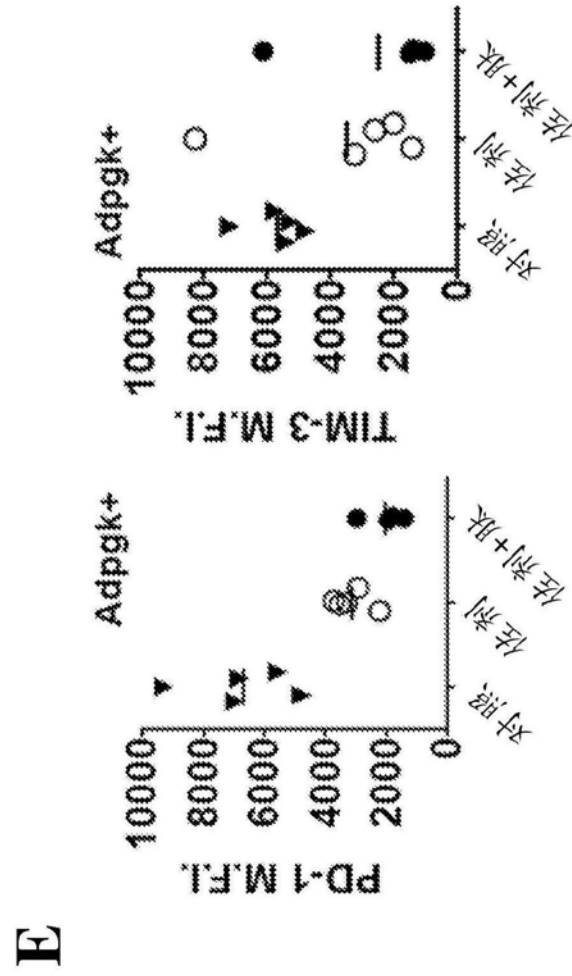


图5 (续)

F

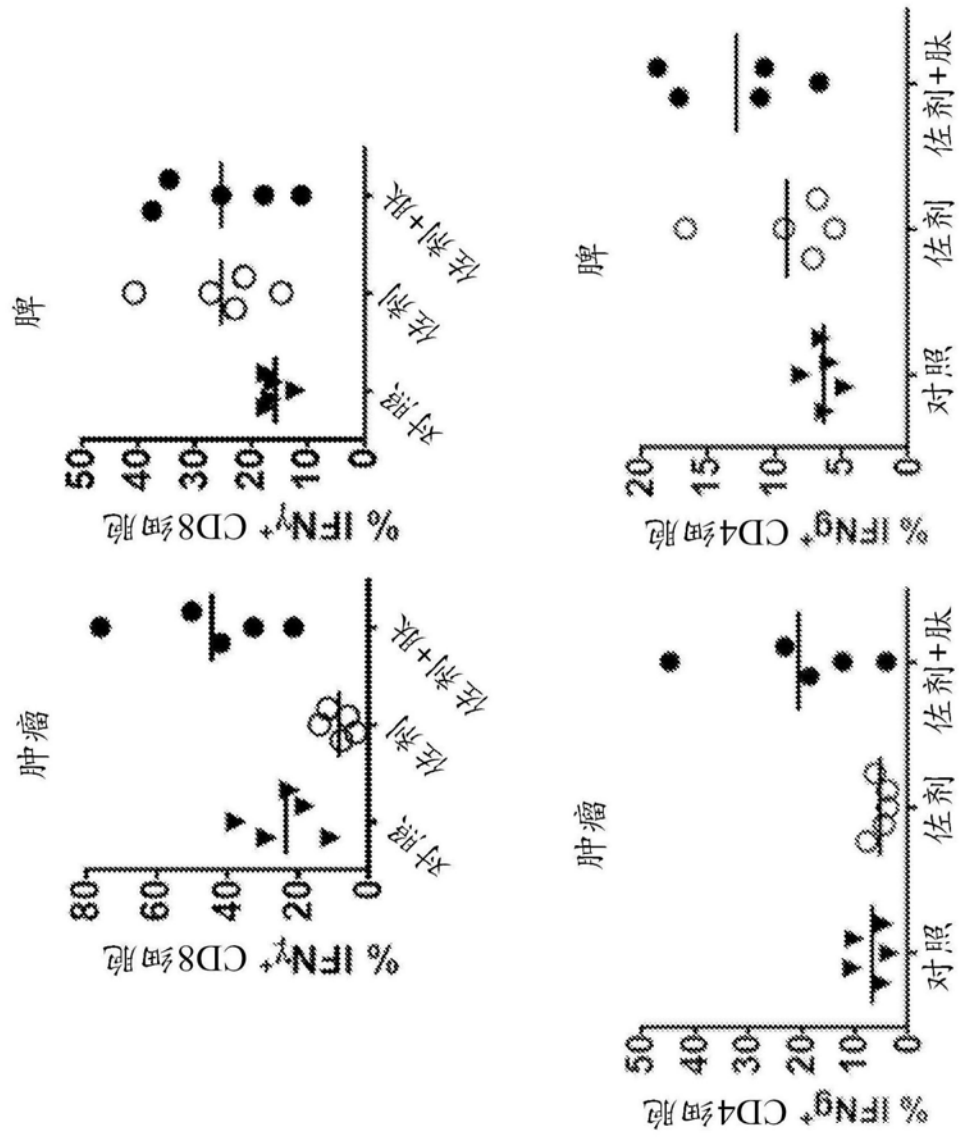


图5 (续)

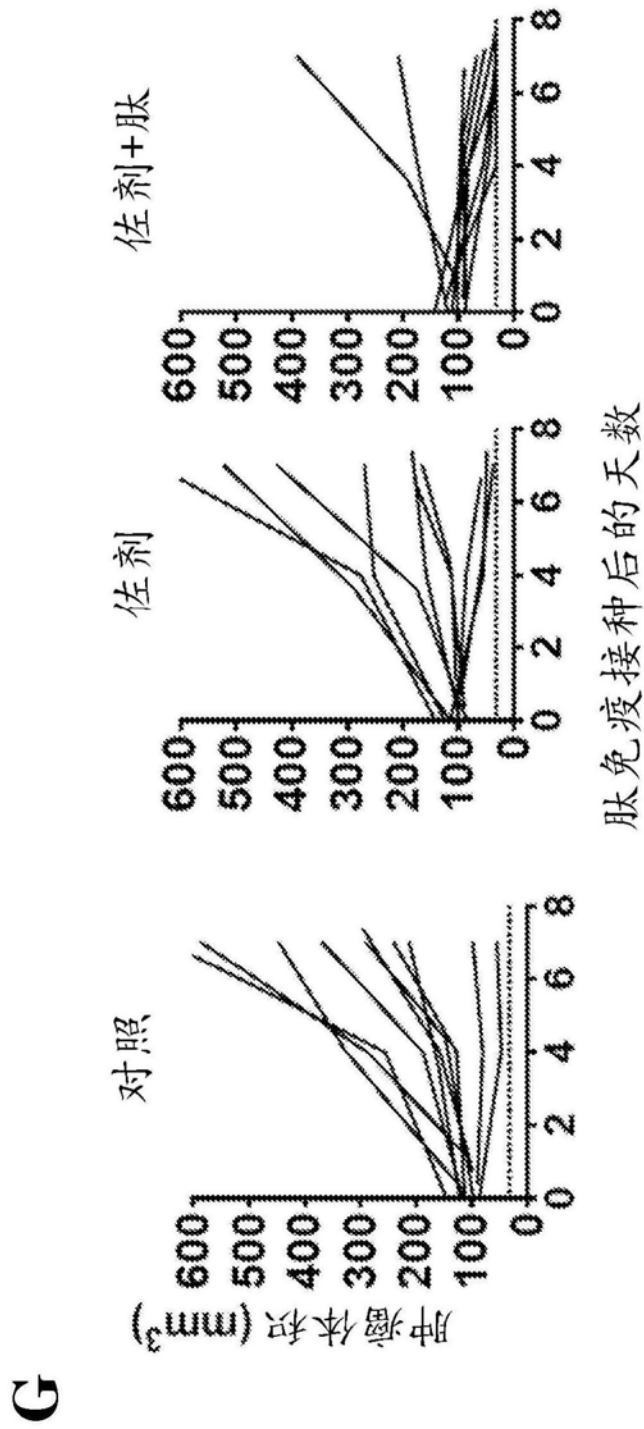


图5 (续)