



(12) 发明专利

(10) 授权公告号 CN 112639136 B

(45) 授权公告日 2024.08.02

(21) 申请号 201980056943.7	<i>C12Q 1/6809</i> (2018.01)
(22) 申请日 2019.08.28	<i>G01N 33/68</i> (2006.01)
(65) 同一申请的已公布的文献号 申请公布号 CN 112639136 A	<i>G01N 33/574</i> (2006.01)
(43) 申请公布日 2021.04.09	<i>C40B 30/00</i> (2006.01)
(30) 优先权数据 62/724,760 2018.08.30 US	<i>C40B 40/02</i> (2006.01)
(85) PCT国际申请进入国家阶段日 2021.02.26	<i>C40B 40/10</i> (2006.01)
(86) PCT国际申请的申请数据 PCT/CA2019/051186 2019.08.28	<i>C40B 50/06</i> (2006.01)
(87) PCT国际申请的公布数据 W02020/041876 EN 2020.03.05	<i>G16B 30/10</i> (2019.01)
(73) 专利权人 蒙特利尔大学 地址 加拿大魁北克	<i>C07K 7/06</i> (2006.01)
(72) 发明人 塞琳·劳蒙 皮埃尔·蒂伯尔 塞巴斯蒂安·勒米厄 克劳德·佩罗	<i>C07K 14/47</i> (2006.01)
(74) 专利代理机构 成都超凡明远知识产权代理 有限公司 51258 专利代理师 王晖 刘书芝	<i>C07K 14/725</i> (2006.01)
(51) Int. Cl.	<i>C07K 17/04</i> (2006.01)
<i>C12Q 1/6886</i> (2018.01)	<i>C12N 15/12</i> (2006.01)
	<i>C12N 5/0783</i> (2010.01)
	<i>C12N 5/10</i> (2006.01)
	<i>A61K 35/17</i> (2015.01)
	<i>A61K 39/00</i> (2006.01)
	<i>A61P 35/00</i> (2006.01)
	<i>A61P 35/02</i> (2006.01)
	<i>A61P 37/04</i> (2006.01)
	(56) 对比文件
	W0 2018102584 A1, 2018.06.07
	审查员 滕瑞芳
	权利要求书2页 说明书56页 序列表9页 附图29页

(54) 发明名称  
用于鉴定肿瘤特异性抗原的基于蛋白质基因组学的方法

(57) 摘要

由于在若干种癌症中肿瘤浸润淋巴细胞 (TIL) 的存在与良好的预后呈正相关, 因此T细胞, 尤其是CD8T细胞被认为是肿瘤根除中必不可少的参与者。为了清除肿瘤细胞, CD8T细胞识别肿瘤抗原, 该肿瘤抗原是存在于肿瘤细胞表面的MHC I相关肽, 在正常细胞上没有表达或表达极低。本文描述了使用来自癌症和正常匹配的mTEC<sup>hi</sup>样品的RNA测序数据的蛋白质基因组学方

法以便鉴定源自 (i) 基因组的编码和非编码区, (ii) 非同义单碱基突变或短插入/删除和更复杂的重排以及 (iii) 内源性逆转录元件的非耐受性肿瘤特异性抗原, 不管样品的突变负荷或复杂性如何, 所述方法均起作用。

CN 112639136 B

1. 一种用于在肿瘤细胞样品中鉴定肿瘤抗原候选物的方法,所述方法包括:

(a1) 在肿瘤细胞样品中分离主要组织相容性复合体相关肽并对其测序,其中所述分离主要组织相容性复合体相关肽包括 (i) 通过弱酸处理从所述细胞样品中释放所述主要组织相容性复合体相关肽;

以及使所述释放的主要组织相容性复合体相关肽经受具有 $3000\text{Da} \pm 10\%$ 的截留值的尺寸排阻柱处理;或 (ii) 肽-主要组织相容性复合体I类复合体的免疫沉淀,然后进行肽洗脱;

(a2) 选择具有的长度为8至11个氨基酸的主要组织相容性复合体相关肽;

(b) 对所述肿瘤细胞样品进行全转录物组测序,从而获得肿瘤RNA序列;

(c) 通过以下产生肿瘤特异性蛋白质组数据库:

(i) 从所述肿瘤RNA序列提取包括至少33个核苷酸的子序列集合;

(ii) 将 (i) 中的所述肿瘤子序列集合与从来自髓胸腺上皮细胞的RNA序列提取的包括至少33个核苷酸的相应的对照子序列集合比较;

(iii) 提取在所述相应的对照子序列中不存在或至少低表达至四分之一的所述肿瘤子序列,从而获得肿瘤特异性子序列;以及

(iv) 经计算机翻译所述肿瘤特异性子序列,从而获得所述肿瘤特异性蛋白质组数据库;

(d) 通过以下产生个性化肿瘤蛋白质组数据库:

(i) 将所述肿瘤RNA序列与参考基因组序列比较来鉴定所述肿瘤RNA序列中的单碱基突变,其中所述参考基因组是人基因组组装;

(ii) 在所述参考基因组序列中插入在 (i) 中鉴定的所述单碱基突变,从而创建个性化肿瘤基因组序列;

(iii) 经计算机翻译来自所述个性化肿瘤基因组序列的编码表达蛋白质的转录物,从而获得所述个性化肿瘤蛋白质组数据库;(e) 通过以下产生个性化正常肿瘤蛋白质组数据库:

(i) 将来自髓胸腺上皮细胞的RNA序列与所述参考基因组序列比较以鉴定所述来自髓胸腺上皮细胞的RNA序列中的单碱基突变,其中所述参考基因组是人基因组组装;

(ii) 在所述参考基因组序列中插入在 (i) 中鉴定的所述单碱基突变,从而创建个性化正常基因组序列;

(iii) 经计算机翻译来自所述个性化正常基因组序列的编码表达蛋白质的转录物,从而获得所述个性化正常蛋白质组数据库;

(f) 通过 (i) 从所述来自正常细胞的RNA序列和所述肿瘤RNA序列提取包括至少24个核苷酸的子序列集合来产生正常和肿瘤k-mer数据库;

(g) 将在 (a) 中获得的主要组织相容性复合体相关肽的序列与 (c) 的所述肿瘤特异性蛋白质组数据库以及 (d) 的所述个性化肿瘤蛋白质组数据库的序列比较以鉴定所述主要组织相容性复合体相关肽;以及

(h) 从在 (f) 中鉴定的所述主要组织相容性复合体相关肽之中鉴定肿瘤抗原候选物,其中肿瘤抗原候选物对应于以下的主要组织相容性复合体相关肽:(1) 其序列在所述个性化正常蛋白质组数据库中不存在;以及 (2) (i) 其序列在所述个性化肿瘤蛋白质组数据库中存

在;和/或(ii)其编码序列在所述肿瘤k-mer数据库中相对于所述正常k-mer数据库过表达或占比过大至少十倍,

其中所述肿瘤是白血病或肺癌。

2.根据权利要求1所述的方法,其中,所述分离主要组织相容性复合体相关肽包括通过弱酸处理从所述细胞样品中释放所述主要组织相容性复合体相关肽;以及使所述释放的主要组织相容性复合体相关肽经受具有 $3000\text{Da} \pm 10\%$ 的截留值的尺寸排阻柱处理。

3.根据权利要求1所述的方法,其中,所述分离主要组织相容性复合体相关肽包括肽-主要组织相容性复合体I类复合体的免疫沉淀,然后进行肽洗脱。

4.根据权利要求1至3中任一项所述的方法,其中,所述子序列包括33至54个碱基对。

5.根据权利要求1至3中任一项所述的方法,在步骤(c)(iii)之后进一步包括将重叠的肿瘤特异性子序列组装成较长的肿瘤子序列。

6.根据权利要求1至3中任一项所述的方法,其中,所述对主要组织相容性复合体相关肽测序包括对所述分离的主要组织相容性复合体相关肽进行质谱测序分析。

7.根据权利要求1至3中任一项所述的方法,其中,所述源自所述主要组织相容性复合体相关肽编码序列的k-mer不存在于所述正常k-mer数据库中。

8.根据权利要求1至3中任一项所述的方法,其中,所述参考基因组是基因组参考联盟人类版本38。

9.根据权利要求1至3中任一项所述的方法,在步骤(h)之后进一步包括将所述肿瘤抗原候选物的编码序列与来自正常组织的序列比较。

10.根据权利要求1至3中任一项所述的方法,在步骤(h)之后进一步包括评估所述肿瘤抗原候选物与主要组织相容性复合体分子的结合。

11.根据权利要求10所述的方法,其中,所述结合是使用主要组织相容性复合体结合预测算法评估的。

12.根据权利要求1至3中任一项所述的方法,在步骤(h)之后进一步包括评估细胞群中识别所述肿瘤抗原候选物的T细胞的频率。

13.根据权利要求12所述的方法,其中,识别所述肿瘤抗原候选物的T细胞的频率是在其肽结合槽中包括所述肿瘤抗原候选物的多聚体主要组织相容性复合体I类分子来评估的。

14.根据权利要求1至3中任一项所述的方法,在步骤(h)之后进一步包括评估所述肿瘤抗原候选物诱导T细胞激活的能力。

15.根据权利要求14所述的方法,其中,所述肿瘤抗原候选物诱导T细胞激活的能力是通过测量由T细胞与在其细胞表面上具有与主要组织相容性复合体I类分子结合的所述肿瘤抗原候选物的细胞接触而引起的细胞因子产生来评估的。

16.根据权利要求15所述的方法,其中,所述细胞因子产生包括干扰素- $\gamma$ 产生。

17.根据权利要求1至3中任一项所述的方法,在步骤(h)之后进一步包括评估所述肿瘤抗原候选物诱导T细胞介导的肿瘤细胞杀伤和/或抑制肿瘤生长的能力。

## 用于鉴定肿瘤特异性抗原的基于蛋白质基因组学的方法

[0001] 相关申请交叉引用

[0002] 本申请要求提交于2018年8月30日的美国临时专利申请No.62/724,760的优先权,其全文通过援引并入本文。

### 技术领域

[0003] 本发明总体上涉及癌症,以及更具体地涉及可用于基于T细胞的癌症免疫治疗的肿瘤抗原的鉴定。

### 背景技术

[0004] 由于在若干种癌症中肿瘤浸润淋巴细胞(TIL)的存在与良好的预后呈正相关,因此CD8 T细胞被认为是肿瘤根除中必不可少的参与者,并且响应于免疫检查点抑制剂<sup>1,2</sup>。为了消除肿瘤细胞,CD8 T细胞识别肿瘤抗原,其是由肿瘤细胞呈递的异常的MHC I相关肽(MAP)。由于CD8 T细胞识别MHC I相关肽(MAP),最重要的未解决的问题在于由CD8 TIL识别的MAP的性质<sup>3</sup>。已知CD8 TIL的丰度与肿瘤的突变负荷相关,主导范式认为CD8 TIL识别突变的肿瘤特异性抗原(mTSA),其通常被称为新抗原(neoantigen)<sup>2,4,5</sup>。mTSA的优异的免疫原性被归因于其在肿瘤上的选择性表达,这使免疫耐受的风险最小化<sup>6</sup>。尽管如此,已经显示一些TIL识别癌症限制的非突变MAP<sup>7</sup>,其被称为异常表达的TSA(aeTSA)。aeTSA可以源自导致在正常细胞中不表达的基因组序列,诸如内源性逆转录元件(ERE)的转录和翻译的各种顺式或反式作用的遗传和表观遗传变化<sup>8-10</sup>。

[0005] 正在做出相当大的努力来发现可以用于治疗性癌症疫苗的可执行的TSA。最常见的策略取决于反向免疫:i)在肿瘤细胞上进行外显子组测序以鉴定突变,以及ii)使用MHC结合预测软件工具来鉴定哪些突变的MAP可能是良好的MHC结合剂<sup>11,12</sup>。虽然反向免疫可以丰富TSA候选物,但是至少90%的这些候选物是假阳性<sup>5,13</sup>,因为可用的计算方法可以预测MHC结合,但是它们无法预测MAP加工过程中的其他步骤<sup>14,15</sup>。为了克服这种限制,一些研究将质谱(MS)分析纳入了他们的TSA发现途径<sup>16</sup>,从而提供了若干个TSA的严格分子定义<sup>17,18</sup>。然而,这些方法的收益极其贫乏:在黑色素瘤中(最易突变的肿瘤类型之一),用MS验证了每个单个肿瘤有平均2个TSA<sup>19</sup>,而对其他癌症类型只发现了极少量的TSA<sup>15</sup>。TSA的缺乏是令人迷惑的,因为如果肿瘤不表达免疫原性抗原,TIL或免疫检查点抑制剂的注射将不会引起肿瘤消退<sup>20</sup>。据推测,基于外显子突变的方法不能鉴定TSA,因为它们没有考虑到两个关键因素。第一,这些方法只关注mTSA而忽略aeTSA,本质上是因为目前没有用于aeTSA的高通量鉴定的方法。这代表了主要的缺点因为,虽然mTSA是稀有抗原,但aeTSA将成为疫苗开发的优选的靶标,因为它们可以被多种肿瘤共享<sup>7,9</sup>。第二,集中于将外显子组作为TSA的唯一来源是非常受限的。外显子组(即,所有蛋白质编码基因)只占人基因组的2%,而最多达75%的基因组可以被转录和潜在翻译<sup>22</sup>。

[0006] 因此,存在用于鉴定可用于基于T细胞的癌症免疫治疗的肿瘤抗原的新颖的方法的需要。

[0007] 急性淋巴细胞白血病 (ALL) 是发生在骨髓、血液和髓外部位的淋巴样祖细胞的恶性转化和增殖。虽然80%的ALL发生在儿童身上,但如果它发生在成人身上,则表示毁灭性的疾病。在美国,ALL的发病率估计为1.6/100 000人。虽然剂量强化策略对儿科患者的结果具有显著改善,但对于老年人的预后仍然很差。尽管对诱导化疗的反应率很高,但仅30-40%的成年ALL患者实现长期缓解。

[0008] 因此,存在用于治疗ALL的新颖的方法的需要。

[0009] 肺癌——一种高度侵袭性、迅速转移性且流行性癌症——是美利坚合众国 (USA) 男性和女性两者的头号杀手癌症。约90%的肺癌病例是由吸烟和使用烟草产品引起的。然而,其他因素诸如氡气、石棉、空气污染暴露以及慢性感染可能促进肺癌变的发生。此外,肺癌易感性的多重继承和获得机制已被提出。肺癌分为生长和扩散方式不同的两个广泛的组织学类别:小细胞肺癌 (SCLC) 和非小细胞肺癌 (NSCLC)。肺癌的治疗选择包括外科手术、放射治疗、化疗以及靶向治疗。尽管在过去的25年期间,对于肺癌的诊断和治疗有了改进,但是对于患者的预后仍然不令人满意。除了对最局部的癌症,对于当前标准治疗的反应很差。

[0010] 因此,存在用于治疗肺癌的新颖的方法的需要。

[0011] 本说明书涉及一些文档,将其全部内容通过援引以其整体并入本文。

## 发明内容

[0012] 本公开提供了以下条目1至75:

[0013] 1. 一种用于在肿瘤细胞样品中鉴定肿瘤抗原候选物的方法,所述方法包括:

[0014] (a) 通过以下产生肿瘤特异性蛋白质组数据库: (i) 从肿瘤RNA序列提取包括至少33个碱基对的子序列 (k-mers) 集合; (ii) 将 (i) 中的所述肿瘤子序列集合与来自正常细胞的RNA序列提取的包括至少33个碱基对的相应的对照子序列集合比较; (iii) 提取在所述相应的对照子序列中不存在的所述肿瘤子序列,从而获得肿瘤特异性子序列;以及 (iv) 经计算机翻译所述肿瘤特异性子序列,从而获得所述肿瘤特异性蛋白质组数据库;

[0015] (b) 通过以下产生个性化肿瘤蛋白质组数据库: (i) 将所述肿瘤RNA序列与参考基因组序列比较来鉴定所述肿瘤RNA序列中的单碱基突变; (ii) 在所述参考基因组序列中插入在 (i) 中鉴定的所述单碱基突变,从而创建个性化肿瘤基因组序列; (iii) 经计算机翻译来自所述个性化肿瘤基因组序列的编码表达蛋白质的转录物,从而获得所述个性化肿瘤蛋白质组数据库;

[0016] (c) 将来自所述肿瘤的主要组织相容性复合体 (MHC) 相关肽 (MAP) 的序列与 (a) 的所述肿瘤特异性蛋白质组数据库以及 (b) 的所述个性化肿瘤蛋白质组数据库的序列比较以鉴定所述MAP;以及

[0017] (d) 从在 (c) 中鉴定的MAP之中鉴定肿瘤抗原候选物,其中肿瘤抗原候选物是其序列和/或编码序列在肿瘤细胞中相对于正常细胞过表达或占比过大的肽。

[0018] 2. 根据条目1所述的方法,其中,上述方法进一步包括 (1) 从所述肿瘤细胞样品中分离主要组织相容性复合体 (MHC) 相关肽 (MAP) 并对其测序,和/或 (2) 对所述肿瘤细胞样品进行全转录物组测序以获得所述肿瘤RNA序列。

[0019] 3. 根据条目2所述的方法,其中,所述分离MAP包括 (i) 通过弱酸处理从所述细胞样品中释放所述MAP;以及 (ii) 对所述释放的MAP进行色谱。

[0020] 4. 根据条目3所述的方法,其中,所述方法进一步包括在所述色谱之前用尺寸排阻柱过滤所述释放的肽。

[0021] 5. 根据条目1至4中任一项所述的方法,其中,所述子序列包括33至54个碱基对。

[0022] 6. 根据条目1至5中任一项所述的方法,进一步包括将重叠的肿瘤特异性子序列组装成较长的肿瘤子序列(重叠群)。

[0023] 7. 根据条目6所述的方法,其中,所述尺寸排阻柱具有约3000Da的截留值。

[0024] 8. 根据条目1至7中任一项所述的方法,其中,所述对MAP测序包括对所述分离的MAP进行质谱(MS)测序分析。

[0025] 9. 根据条目1至8中任一项所述的方法,其中,所述方法进一步包括使用相应的正常细胞产生个性化正常蛋白质组数据库。

[0026] 10. 根据条目9所述的方法,其中,(d)中所述的鉴定包括如果在所述正常个性化蛋白质组数据库中检测到所述MAP的序列,则排除该MAP。

[0027] 11. 根据条目1至10中任一项所述的方法,其中,所述方法进一步包括从所述肿瘤RNA序列和从来自正常细胞的RNA序列产生24核苷酸或39核苷酸k-mer数据库以获得肿瘤k-mer数据库和正常k-mer数据库;以及将所述肿瘤k-mer数据库和正常k-mer数据库与源自所述MAP编码序列的24核苷酸或39核苷酸k-mer比较,其中,源自所述MAP编码序列的所述k-mer在所述肿瘤k-mer数据库中相对于所述正常k-mer数据库的过表达或占比过大表明了所述相应的MAP是肿瘤抗原候选物。

[0028] 12. 根据条目11所述的方法,其中,源自所述MAP编码序列的所述k-mer在所述肿瘤k-mer数据库中相对于所述正常k-mer数据库过表达或占比过大至少10倍。

[0029] 13. 根据条目11或12所述的方法,其中,所述源自所述MAP编码序列的k-mer不存在于所述正常k-mer数据库中。

[0030] 14. 根据条目1至13中任一项所述的方法,其中所述方法包括:

[0031] (a) 在肿瘤细胞样品中分离MAP并对其测序;

[0032] (b) 对所述肿瘤细胞样品进行全转录物组测序,从而获得肿瘤RNA序列;

[0033] (c) 通过以下产生肿瘤特异性蛋白质组数据库:(i) 从所述肿瘤RNA序列提取包括至少33个核苷酸的子序列集合;(ii) 将(i)中的所述肿瘤子序列集合与从来自正常细胞的RNA序列提取的包括至少33个核苷酸的相应的对照子序列集合比较;(iii) 提取在所述相应的对照子序列中不存在或至少低表达至四分之一的所述肿瘤子序列,从而获得肿瘤特异性子序列;以及(iv) 经计算机翻译所述肿瘤特异性子序列,从而获得所述肿瘤特异性蛋白质组数据库;

[0034] (d) 通过以下产生个性化肿瘤蛋白质组数据库:(i) 将所述肿瘤RNA序列与参考基因组序列比较来鉴定所述肿瘤RNA序列中的单碱基突变;(ii) 在所述参考基因组序列中插入在(i)中鉴定的所述单碱基突变,从而创建个性化肿瘤基因组序列;(iii) 经计算机翻译来自所述个性化肿瘤基因组序列的编码表达蛋白质的转录物,从而获得所述个性化肿瘤蛋白质组数据库;

[0035] (e) 通过以下产生个性化正常肿瘤蛋白质组数据库:(i) 将来自正常细胞的RNA序列与参考基因组序列比较以鉴定所述正常RNA序列中的单碱基突变;(ii) 在所述参考基因组序列中插入在(i)中鉴定的所述单碱基突变,从而创建个性化正常基因组序列;(iii) 经

计算机翻译来自所述个性化正常基因组序列的编码表达蛋白质的转录物,从而获得所述个性化正常蛋白质组数据库;

[0036] (f) 通过 (i) 从所述来自正常细胞的RNA序列和所述肿瘤RNA序列提取包括至少24个核苷酸的子序列集合来产生正常和肿瘤k-mer数据库;

[0037] (g) 将在 (a) 中获得的MAP的序列与 (c) 的所述肿瘤特异性蛋白质组数据库以及 (d) 的所述个性化肿瘤蛋白质组数据库的序列比较以鉴定所述MAP;以及

[0038] (h) 从所述在 (f) 中鉴定的MAP之中鉴定肿瘤抗原候选物,其中肿瘤抗原候选物对应于以下的MAP: (1) 其序列在所述个性化正常蛋白质组数据库中不存在;以及 (2) (i) 其序列在所述个性化肿瘤蛋白质组数据库中存在;和/或 (ii) 其编码序列在所述肿瘤k-mer数据库中相对于所述正常k-mer数据库过表达或占比过大。

[0039] 15. 根据条目1至14中任一项所述的方法,其中,所述方法进一步包括选择具有的长度为8至11个氨基酸的MAP。

[0040] 16. 根据条目1至15中任一项所述的方法,其中,所述正常细胞是胸腺细胞。

[0041] 17. 根据条目16所述的方法,其中,所述胸腺细胞是髓胸腺上皮细胞 (mTEC)。

[0042] 18. 根据条目1至17中任一项所述的方法,进一步包括将所述肿瘤抗原候选物的编码序列与来自正常组织的序列比较。

[0043] 19. 根据条目1至18中任一项所述的方法,其中,所述MAP具有的长度为8至11个氨基酸。

[0044] 20. 根据条目1至19中任一项所述的方法,进一步包括评估所述肿瘤抗原候选物与MHC分子的结合。

[0045] 21. 根据条目20所述的方法,其中,所述结合是使用MHC结合预测算法评估的。

[0046] 22. 根据条目1至21中任一项所述的方法,进一步包括评估细胞群中识别所述肿瘤抗原候选物的T细胞的频率。

[0047] 23. 根据条目22所述的方法,其中,识别所述肿瘤抗原候选物的T细胞的频率是使用在其肽结合槽中包括所述肿瘤抗原候选物的多聚体MHC I类分子来评估的。

[0048] 24. 根据条目1至23中任一项所述的方法,进一步包括评估所述肿瘤抗原候选物诱导T细胞激活的能力。

[0049] 25. 根据条目24所述的方法,其中,所述肿瘤抗原候选物诱导T细胞激活的能力是通过测量由T细胞与在其细胞表面上具有与MHC I类分子结合的所述肿瘤抗原候选物的细胞接触而引起的细胞因子产生来评估的。

[0050] 26. 根据条目25所述的方法,其中,所述细胞因子产生包括干扰素- $\gamma$  (IFN- $\gamma$ ) 产生。

[0051] 27. 根据条目1至26中任一项所述的方法,进一步包括评估所述肿瘤抗原候选物诱导T细胞介导的肿瘤细胞杀伤和/或抑制肿瘤生长的能力。

[0052] 28. 肿瘤抗原肽,其通过根据条目1至27中任一项所述的方法来鉴定。

[0053] 29. 肿瘤抗原肽,其包括以下或由以下组成:SEQ ID NO:1-39中任一个所示的氨基酸序列之一。

[0054] 30. 根据条目29所述的肿瘤抗原肽,其包括以下或由以下组成:SEQ ID NO:17-39中任一个所示的氨基酸序列之一。

- [0055] 31. 根据条目30所述的肿瘤抗原肽,其中,所述肿瘤抗原肽是白血病肿瘤抗原肽并且包括以下或由以下组成:SEQ ID NO:17-28中任一个所示的氨基酸序列之一。
- [0056] 32. 根据条目31所述的肿瘤抗原肽,其中,所述白血病是B细胞急性淋巴细胞白血病(B-ALL)。
- [0057] 33. 根据条目31或32所述的肿瘤抗原肽,其中,所述肿瘤抗原肽结合至HLA-A\*02:01等位基因的人白细胞抗原(HLA)并且包括以下或由以下组成:SEQ ID NO:17-19、27和28中任一个所示的氨基酸序列之一。
- [0058] 34. 根据条目31或32所述的肿瘤抗原肽,其中,所述肿瘤抗原肽结合至HLA-B\*40:01等位基因的人白细胞抗原(HLA)并且包括以下或由以下组成:SEQ ID NO:20所示的氨基酸序列。
- [0059] 35. 根据条目31或32所述的肿瘤抗原肽,其中,所述肿瘤抗原肽结合至HLA-A\*11:01等位基因的人白细胞抗原(HLA)并且包括以下或由以下组成:SEQ ID NO:21-23中任一个所示的氨基酸序列之一。
- [0060] 36. 根据条目31或32所述的肿瘤抗原肽,其中,所述肿瘤抗原肽结合至HLA-B\*08:01等位基因的人白细胞抗原(HLA)并且包括以下或由以下组成:SEQ ID NO:24或25所示的氨基酸序列。
- [0061] 37. 根据条目31或32所述的肿瘤抗原肽,其中,所述肿瘤抗原肽结合至HLA-B\*07:02等位基因的人白细胞抗原(HLA)并且包括以下或由以下组成:SEQ ID NO:26所示的氨基酸序列。
- [0062] 38. 根据条目30所述的肿瘤抗原肽,其中,所述肿瘤抗原肽是肺肿瘤抗原肽并且包括以下或由以下组成:SEQ ID NO:29-39中任一个所示的氨基酸序列之一。
- [0063] 39. 根据条目38所述的肿瘤抗原肽,其中,所述肺肿瘤是非小细胞肺癌(NSCLC)。
- [0064] 40. 根据条目38或39所述的肿瘤抗原肽,其中,所述肿瘤抗原肽结合至HLA-A\*11:01等位基因的人白细胞抗原(HLA)并且包括以下或由以下组成:SEQ ID NO:29-35中任一个所示的氨基酸序列之一。
- [0065] 41. 根据条目38或39所述的肿瘤抗原肽,其中,所述肿瘤抗原肽结合至HLA-B\*07:02等位基因的人白细胞抗原(HLA)并且包括以下或由以下组成:SEQ ID NO:36所示的氨基酸序列。
- [0066] 42. 根据条目38或39所述的肿瘤抗原肽,其中,所述肿瘤抗原肽结合至HLA-A\*24:02等位基因的人白细胞抗原(HLA)并且包括以下或由以下组成:SEQ ID NO:38或39所示的氨基酸序列。
- [0067] 43. 根据条目38或39所述的肿瘤抗原肽,其中,所述肿瘤抗原肽结合至HLA-C\*07:01等位基因的人白细胞抗原(HLA)并且包括以下或由以下组成:SEQ ID NO:37所示的氨基酸序列。
- [0068] 44. 根据条目29-43中任一项所述的肿瘤抗原,其源自基因组的非蛋白编码区。
- [0069] 45. 根据条目44所述的肿瘤抗原,其中,所述基因组的所述非蛋白编码区是基因间区、内含子区、5'非翻译区(5'UTR)、3'非翻译区(3'UTR)或内源性逆转录元件(ERE)。
- [0070] 46. 编码根据条目28-45中任一项所述的肿瘤抗原肽的核酸。
- [0071] 47. 根据条目46所述的核酸,所述核酸是mRNA或病毒载体。

- [0072] 48. 脂质体,包括根据条目28-45中任一项所述的肿瘤抗原肽或根据条目46或47所述的核酸。
- [0073] 49. 组合物,包括根据条目28-45中任一项所述的肿瘤抗原肽、根据条目46或47所述的核酸或根据条目48所述的脂质体,以及药学上可接受的载体。
- [0074] 50. 疫苗,包括根据条目28-45中任一项所述的肿瘤抗原肽、根据条目46或47所述的核酸、根据条目48所述的脂质体或根据条目49所述的组合物,以及佐剂。
- [0075] 51. 分离的主要组织相容性复合体(MHC) I类分子,在其肽结合槽中包括根据条目28-45中任一项所述的肿瘤抗原肽。
- [0076] 52. 根据条目51所述的分离的MHC I类分子,其为多聚体的形式。
- [0077] 53. 根据条目52所述的分离的MHC I类分子,其中所述多聚体是四聚体。
- [0078] 54. 分离的细胞,包括根据条目28-45中任一项所述的肿瘤抗原肽。
- [0079] 55. 分离的细胞,在其表面表达主要组织相容性复合体(MHC) I类分子,所述主要组织相容性复合体(MHC) I类分子在其肽结合槽中包括根据条目28-45中任一项所述的肿瘤抗原肽。
- [0080] 56. 根据条目55所述的细胞,其是抗原呈递细胞(APC)。
- [0081] 57. 根据条目56所述的细胞,其中所述APC是树突状细胞。
- [0082] 58. T细胞受体(TCR),其特异性识别根据条目51-53中任一项所述的分离的MHC I类分子和/或在根据条目54-57中任一项所述的细胞的表面表达的MHC I类分子。
- [0083] 59. 分离的CD8<sup>+</sup> T淋巴细胞,在其细胞表面表达根据条目58所述的TCR。
- [0084] 60. 细胞群,其包括至少0.5%的根据条目59所述的CD8<sup>+</sup> T淋巴细胞。
- [0085] 61. 治疗受试者的癌症的方法,包括向所述受试者给药有效量的:(i) 根据条目28-45中任一项所述的肿瘤抗原肽;(ii) 根据条目46或47所述的核酸;(iii) 根据条目48所述的脂质体;(iv) 根据条目49所述的组合物;(v) 根据条目50所述的疫苗;(vi) 根据条目54-57中任一项所述的细胞;(vii) 根据条目59所述的CD8<sup>+</sup> T淋巴细胞;或(viii) 根据条目60所述的细胞群。
- [0086] 62. 根据条目61所述的方法,其中,所述癌症是白血病。
- [0087] 63. 根据条目62所述的方法,其中,所述白血病是B细胞急性淋巴细胞白血病(B-ALL)。
- [0088] 64. 根据条目61所述的方法,其中,所述癌症是肺癌。
- [0089] 65. 根据条目64所述的方法,其中,所述肺肿瘤是非小细胞肺癌(NSCLC)。
- [0090] 66. 根据条目61-65中任一项所述的方法,进一步包括向所述受试者给药至少一种额外的抗肿瘤剂或治疗。
- [0091] 67. 根据条目66所述的方法,其中,所述至少一种额外的抗肿瘤剂或治疗是化学治疗剂、免疫治疗、免疫检查点抑制剂、放射治疗或外科手术。
- [0092] 68. (i) 根据条目28-45中任一项所述的肿瘤抗原肽;(ii) 根据条目46或47所述的核酸;(iii) 根据条目48所述的脂质体;(iv) 根据条目49所述的组合物;(v) 根据条目50所述的疫苗;(vi) 根据条目54-57中任一项所述的细胞;(vii) 根据条目59所述的CD8<sup>+</sup> T淋巴细胞;或(viii) 根据条目60所述的细胞群用于治疗受试者的癌症的用途。
- [0093] 69. (i) 根据条目28-45中任一项所述的肿瘤抗原肽;(ii) 根据条目46或47所述的

核酸; (iii) 根据条目48所述的脂质体; (iv) 根据条目49所述的组合物; (v) 根据条目50所述的疫苗; (vi) 根据条目54-57中任一项所述的细胞; (vii) 根据条目59所述的CD8<sup>+</sup> T淋巴细胞; 或 (viii) 根据条目60所述的细胞群制造用于治疗受试者的癌症的药物的用途。

[0094] 70. 根据条目68或69所述的用途, 其中, 所述癌症是白血病。

[0095] 71. 根据条目70所述的用途, 其中, 所述白血病是B细胞急性淋巴细胞白血病 (B-ALL)。

[0096] 72. 根据条目68或69所述的用途, 其中, 所述癌症是肺癌。

[0097] 73. 根据条目72所述的用途, 其中, 所述肺肿瘤是非小细胞肺癌 (NSCLC)。

[0098] 74. 根据条目68-73中任一项所述的用途, 进一步包括至少一种额外的抗肿瘤剂或治疗的使用。

[0099] 75. 根据条目74所述的用途, 其中, 所述至少一种额外的抗肿瘤剂或治疗是化学治疗剂、免疫治疗、免疫检查点抑制剂、放射治疗或外科手术。

[0100] 通过参考附图阅读以下仅以举例方式给出的本发明的特定实施方式的非限制性描述, 本发明的其他目的、优点和特征将变得更加明显。

## 附图说明

[0101] 在附图中:

[0102] 图1A-C示出了用于肿瘤特异性抗原 (TSA) 鉴定的靶向蛋白质基因组工作流程。图1A、B: 详述了如何创建用于每个分析样品的经典癌症蛋白质组 (图1A) 以及癌症特异性蛋白质组 (图1B) 的原理图。图1C: 然后使用这两种蛋白质组的组合 (被称为全局癌症数据库) 以鉴定来自两种充分表征的小鼠细胞系 (即CT26和EL4) 以及七种人原发样品 (即四种B-ALL和三种肺肿瘤活检 (n=2-4/样品)) 的由液相色谱-MS/MS (LC-MS/MS) 测序的MAP, 以及更特别地TSA。关于全局癌症数据库中的每一部分的统计数据可以参见表4a-b, 同时通过k-mer剖析构建癌症特异性蛋白质组的实施细节如图7所呈现。aa: 氨基酸, nts: 核苷酸, th: k-mer出现时的样品特异性阈值 (参见下文实施例1的k-mer过滤和癌症特异性蛋白质组的产生的部分)。

[0103] 图2A-D描绘了示出大部分TSA源自非编码区的翻译的实验结果。图2A: 表明TSA发现中涉及的关键验证步骤的流程图。关于每个步骤的细节在图8中找到。图2B: 大部分TSA候选物源自异常表达的序列。示出了在CT26和EL4肿瘤模型中的mTSA (m) 和aeTSA候选物 (ae) 的数目的条形图。图2C: 示出了其中RNA-Seq数据可公开获得的22个组织/器官中的aeTSA候选物的MCS表达的热图 (参见表5)。示出了先前报道的过表达的EL4 TAA<sup>40,41</sup>的MCS表达, 作为对照。表达值被归一化为rpkm (每亿个测序的读段中的读段数, 细节参见实施例1的MCS的外周表达部分), 并对每个组织可获得的所有RNA-Seq实验进行平均。加粗正方形指示检测到相关MCS>0rpkm的组织。Adip. 组织: 脂肪组织、mam. 腺: 乳腺以及s.c. adip. 组织: 皮下脂肪组织。图2D: 大多数ae/mTSA源自非编码区。描绘了源自编码外显子框内翻译 (编码-内)、编码外显子框外翻译 (编码-外) 以及据称非编码区的翻译 (非编码) 的TSA数目的条形图。条内部的数字代表了aeTSA/mTSA的数目。条上方的百分数指明了源自非典型翻译事件的TSA, 即属于编码-外和非编码类别的TSA的比例。CT26和EL4 TSA的特征可以分别在图1a和b中找到。

[0104] 图3A-C为示出了针对各个TSA的免疫赋予不同程度的对抗EL4细胞的保护的图。将C57BL/6雌性  $\bar{X}$  小鼠用以下用各个TSA冲击的DC免疫接种2次：(图3A)两种aeTSA、(图3B)两种ERE TSA(一种aeTSA或一种mTSA)以及(图3C)一种mTSA。在第0天向小鼠静脉注射 $5 \times 10^5$ 个活EL4细胞(黑三角形)以及所有存活的小鼠在第150天被再次攻击。将对照组用未冲击的DC免疫接种(黑线)。代表中位生存期。使用时序检验(log-rank test)计算免疫组相对于对照组的统计学显著性,其中ns表示不显著( $p > 0.05$ )。肽特异性免疫每组10只小鼠,对照组19只小鼠。

[0105] 图4A-D是示出幼稚(naive)和免疫小鼠中的TSA特异性T细胞的IFN- $\gamma$ 分泌的频率的图。图4A:幼稚小鼠中每 $10^6$ 个 $CD8^+$ T细胞中四聚体 $^+CD8^+$ T细胞的数目。每个圆圈代表一只小鼠( $n=5$ 至9只小鼠)。虚线代表每 $10^6$ 个 $CD8^+$ T细胞中1个四聚体 $^+$ T细胞的频率。 $p$ 值是使用双尾曼-惠特尼检验(Mann-Whitney tests)来计算的(\*\*; $p < 0.01$ 以及\*\*\*; $p < 0.001$ )。图4B:免疫之后抗原特异性 $CD8^+$ T细胞的扩增。四聚体 $^+CD8^+$ T细胞的倍数富集是通过用相关(白色条)或不相关(灰色条)肽免疫接种的小鼠的平均频率除以幼稚小鼠中的平均频率来计算的。图4C、4D:来自免疫的小鼠的分选的 $CD8^+$ T细胞在辐射的肽冲击的脾细胞中孵育48小时。图4C:将IFN- $\gamma$ 分泌抗原特异性细胞的频率表示为免疫的小鼠的斑点形成细胞(spot-forming cell)的平均频率(每 $10^6$ 个铺板的 $CD8^+$ T细胞所报告的SFC)减去幼稚小鼠的斑点形成细胞的平均频率。三个独立的实验,用圆圈表示技术复制。 $p$ 值通过双尾学生t检验计算,所有的 $p < 0.05$ (范围:0.0025-0.0143)。图4D:抗原特异性T细胞的功能亲合力是通过将SFC的频率相对于最大值归一化并通过使用剂量-反应曲线计算每个肽的 $EC_{50}$ 来计算的。 $H7^a$ 和 $H13^a$ 的功能亲合力值先前已公布并用于比较目的。三个独立的实验。在所有相关组中,完整的水平线和每个条件上方的数字代表平均值。用作对照的病毒肽用灰色突出显示。

[0106] 图5A-D是示出了源自EL4的TSA的高表达很重要但不足以诱导抗白血病反应的图。图5A、B:在第0天和第150天注射的EL4细胞上分别进行RNA和肽水平上的TSA表达的分析。图5A:表示EL4 RNA-Seq读段与编码五个EL4 TSA中的每一个的MCS完全重叠的数目的条形图。图5B:每个细胞的TSA拷贝数,如通过使用五个EL4 TSA的 $^{13}C$ 合成肽类似物的PRM MS来估计的。每个TSA的EL4细胞三次重复。TSA拷贝数/细胞的平均数在图的左侧指明。N.D.:未检出图5C:在未预先用肽冲击的DC免疫情况下,注射活EL4细胞之后的TSA特异性 $CD8^+$ T细胞的扩增。四聚体 $^+CD8^+$ T细胞的倍富集是通过用EL4注射的小鼠的平均频率除以幼稚小鼠的平均频率来计算的。识别不由EL4细胞呈递的病毒肽的T细胞的倍富集示出为阴性对照并以灰色突出显示。图5D:将C57BL/6雌性小鼠用已辐照的592d(10,000cGy)EL4细胞(蓝色线)或作为对照的未冲击的DC(黑色线)免疫接种两次,以及然后静脉注射 $5 \times 10^5$ 个活EL4细胞。 $\bar{X}$ 代表中位生存期。已辐照的EL4细胞免疫接种10只小鼠,对照组19只。

[0107] 图6A-C示出了在人原发性肿瘤中检测到的大多数TSA源自非编码区的翻译。图6A:大部分人TSA是aeTSA。示出了每个主要的分析的样品中aeTSA候选物(ae)和mTSA(m)的数目的条形图。图6B:人aeTSA候选物和TAA的外周表达。示出了27个aeTSA和24个过表达的TAA的MCS的表达式的热图,其从癌症免疫肽数据库<sup>48</sup>中获得,横跨其中RNA-Seq数据可公开获得的28种人组织的范围(参见表6)。表达值被归一化为rphm(细节参见下文实施例1的MCS的外周表达部分),并对每个组织可获得的所有RNA-Seq实验进行平均。对于每种抗原,其MCS表达>

15rphm的组织数目示出在热图的左侧。Adip.s.c.:皮下脂肪。图6C:大多数人TSA源自非编码区。描绘了源自编码外显子框内翻译(编码-内)、编码外显子框外翻译(编码-外)以及据称非编码区的翻译(非编码)的人TSA数目的条形图。在每个样品中鉴定出的人TSA的特征可以参见表2a-d和3a-c。

[0108] 图7A-D是用于k-mer剖析工作流程的代码的架构的示意图。关于用于从RNA-seq读段生成k-mers(图7A)、过滤k-mers(图7B)、将k-mers组装成重叠群(图7C)以及翻译重叠群(图7D)的代码的细节。

[0109] 图8A-C示出了TSA验证过程。图8A:详述了成对的MAP/蛋白质的免疫原性状态的计算的示意图。FC:肿瘤/同基因的mTEC<sup>hi</sup>(鼠样品)或TEC/mTEC(人样品)。图8B:用于对标记为TSA候选物的MAP进行MS相关验证的策略。图8C:概括用于为MS验证的鼠TSA候选物(CT26和EL4)以及MS验证的针对B-ALL标本和肺癌的人TSA候选物指定基因组位置的策略的示意图。

[0110] 图9A-D是示出了幼稚和预先免疫接种的小鼠中的抗原特异性CD8<sup>+</sup>T细胞的检测的图。图9A:检测离体的pMHC四聚体<sup>+</sup>CD8<sup>+</sup>T细胞的圈选策略(gating strategy)。对从每个小鼠的脾脏和淋巴结分离的单细胞悬浮液进行四聚体富集。经过双重排除(doublets exclusion)之后,分析Dump<sup>-</sup>CD3<sup>+</sup>细胞的CD8和CD4表达,以及在CD8<sup>+</sup>区室分析pMHC I四聚体<sup>+</sup>细胞。示出了在幼稚小鼠中VTPV/H-2K<sup>b</sup>-PE和M45/H-2D<sup>b</sup>-APC四聚体富集以后获得的代表性染色。指明了每种特异性检测到的四聚体<sup>+</sup>CD8<sup>+</sup>T细胞的绝对数目。Dump通道对应于死细胞、CD45R和CD19、F4/80、CD11b、CD11c阳性的混合事件。图9B、C:在幼稚(上排)以及预先免疫接种(下排)的小鼠中的抗原特异性CD8<sup>+</sup>T细胞的CD44表达的代表性分析。呈现了磁性富集前CD8<sup>+</sup>细胞的CD44状态(图9B,左侧小图)以及在离体富集之后的四聚体<sup>+</sup>病毒特异性(图9B)以及TSA特异性(图9C)。指示了CD44-阳性或阴性细胞的百分比和数目。图9D:免疫小鼠和幼稚小鼠中IFN- $\gamma$ 分泌CD8<sup>+</sup>T细胞的频率的一个代表性实验。在每个孔的下方指明了在每种条件下的斑点形成单位(SFU)的数目相对于铺板的CD8<sup>+</sup>T细胞的数目的数目。

[0111] 图10A-D是示出了抗原特异性T细胞的频率的图。图10A:在幼稚或用相关或不相关肽免疫接种的小鼠中的抗原特异性T细胞的频率。图10B、C:在第150天再次用EL4细胞攻击的针对VTPVYQHL或TVPLNHNTL(图10B)或针对VNYLHRNV或VNYIHRNV(图10C)免疫接种的小鼠中抗原特异性CD8<sup>+</sup>T细胞的频率。出于比较目的,在图10A中报告的在幼稚小鼠和免疫接种的小鼠中的抗原特异性T细胞频率再次出现。图10D:用EL4细胞注射的非免疫接种的小鼠的抗原特异性T细胞的频率。四聚体<sup>+</sup>CD8<sup>+</sup>T细胞的所有计算的频率被表达为抗原特异性CD8<sup>+</sup>T细胞的数目/10<sup>6</sup>个CD8<sup>+</sup>T细胞。每个符号代表一只小鼠(n=1至9只小鼠)。虚线代表一个四聚体<sup>+</sup>T细胞/10<sup>6</sup>个CD8<sup>+</sup>T细胞的最小检测水平。用作对照的病毒肽用灰色突出显示。p值是使用双尾曼-惠特尼检验来计算的(\*p $\leq$ 0.05)。

[0112] 图11A-C是示出了幼稚小鼠与预先免疫接种的小鼠之间的抗原特异性T细胞频率的相关性的图。在幼稚库和在免疫接种的小鼠中的抗原特异性CD8<sup>+</sup>T细胞的频率之间的相关性,如通过四聚体染色(图11A)和IFN- $\gamma$  ELISpot测定法(图11B)计算的。图11C:在免疫接种的小鼠中的抗原特异性CD8<sup>+</sup>T细胞的频率之间的相关性,如通过四聚体染色和IFN- $\gamma$  ELISpot测定法计算的。使用平均频率用于绘制数据。曲线的拟合度由决定系数(r<sup>2</sup>)确定。

[0113] 图12A-B描绘了人TEC和mTEC转录组景观(landscape)的综述。图12A:从不相关供体中分离的人TEC(062015和102015)和mTEC(S5至S11)显示类似的转录组谱。在RNA-Seq之

后,如kallisto估计的,选择在具有tpm>1的至少一个供体中表达的转录物来绘制所有一对一的散点图。斯皮尔曼等级相关系数( $\rho$ )在每个图的左上角指明,并且黑线代表转录物的相同表达。图12B:额外的人类TEC/mTEC样品的RNA-Seq应当导致信息的微小增益。使用表达的转录物的集合(至少一个样品中tpm>1),通过使用以下函数推断通过向队列(nS,参见下文实施例1的TEC和mTEC样品中检测到的转录物的累积数部分)中添加额外的样品应当能被检测到的转录物的累积数(cT):  $cT = \frac{a \times (nS-1)}{[b + (nS-1)]} + c$  其中a=23,892.73、b=0.8243389以及c=75,976.11(灰色线)。在图中指明了通过分析nS=6(目前的队列,黑色点)或nS=20个样品检测到的转录物的累积数以及应当检测到的对应于  $\lim_{nS \rightarrow \infty} \left( \frac{a \times (nS-1)}{[b + (nS-1)]} + c \right) = a + c = 99,868$  (渐近值)的转录物的总数。

[0114] 图13A-C是示出了用于通过FACS分选分离细胞的圈选策略的图。图13A:用于鼠mTEC<sup>hi</sup>的分离的圈选策略。在从C57BL/6或Balb/c小鼠的胸腺分离的单细胞悬浮液上进行mTEC<sup>hi</sup>分离。经过双重排除之后,mTEC<sup>hi</sup>细胞定义为7-AAD<sup>-</sup>、EpCAM<sup>+</sup>、CD45<sup>-</sup>(Alexa Fluor 700用于C57BL/6或FITC用于Balb/c小鼠)、UEA-1<sup>+</sup>和I-Ab<sup>+</sup>(C57BL/6小鼠)或I-A/I-E<sup>+</sup>(Balb/c小鼠)。图13B:用于人TEC和mTEC的分离的圈选策略。对从胸腺分离的单细胞悬浮液进行细胞分选,这些悬浮液是从3个月至7岁的个体进行了矫正心血管外科手术后获得的。经过双重排除之后,TEC定义为CD45<sup>-</sup>、7-AAD<sup>-</sup>、EpCAM<sup>+</sup>以及HLA-DR<sup>+</sup>。对于mTEC的分选,细胞被进一步定义为CDR2<sup>-</sup>。图13C:用于IFN- $\gamma$ ELISpot测定法的CD8<sup>+</sup> T细胞分离的圈选策略。在从幼稚或免疫接种的C57BL/6小鼠的脾脏分离的单细胞悬浮液上进行CD8<sup>+</sup> T细胞分离。经过双重排除之后,CD8a标记物用于富集CD8<sup>+</sup> T细胞。

### 具体实施方式

[0115] 本文使用的遗传学、分子生物学、生物化学和核酸的术语和符号遵循本领域的那些标准论文和文本,例如Kornberg and Baker,DNA Replication,Second Edition (W.H.Freeman&Co,New York,1992);Lehninger,Biochemistry,Sixth Edition (W.H.Freeman&Co,New York,2012);Strachan and Read,Human Molecular Genetics,fifth Edition(CRC Press,2018);Eckstein,editor,Oligonucleotides and Analogs:A Practical Approach(Oxford University Press,New York,1991);等。所有术语都应当按照相关领域中规定的典型含义来理解。

[0116] 冠词“一个或一种(a)”和“一个或一种(an)”在这里用来指冠词的一个或多个于一个语法宾语(即至少一个)。例如,“一种元素”意指一种元素或多种元素。在本说明书和权利要求书中,除非上下文另有要求,“包括(comprise)”、“包括(comprises)”和“包括(comprising)”将被理解为意指包括所陈述的步骤或要素或步骤或要素的组,但不排除任何其他步骤或要素或步骤或要素的组。

[0117] 除非在此另外指示,否则在此对数值范围的叙述仅意在用作分别提及落入范围内的每个单独数值的简写方法,并且每个单独数值被并入到说明书中,犹如其在此被单独叙述一样。范围内的所有值子集也被并入到说明书中,就像它们在本文被单独列举一样。

[0118] 本文所述的所有方法可以任何合适的顺序进行,除非另有说明或上下文明显矛

盾。

[0119] 本文提供的任何和所有实例或示例性语言(例如“诸如”)的使用仅用来更好地说明本发明,而不是对本发明的范围施加限制,除非另有说明。

[0120] 说明书中的任何语言都不应当被解释为表明了任何非权利要求的要素对于发明的实施是必不可少的。

[0121] 在本文中,术语“约”具有一般意义。术语“约”用来表明值包括用于测定该值的设备或方法的固有误差变化或涵盖接近所述值的值,例如在所列举的值(或值的范围)的10%或5%内。

[0122] 正在做出相当大的努力来发现可以用于治疗性癌症疫苗的可执行的TSA。最常见的策略取决于反向免疫:i)在肿瘤细胞上进行外显子组测序以鉴定突变,以及ii)使用MHC结合预测软件工具来鉴定哪些突变的MAP可能是良好的MHC结合剂<sup>11,12</sup>。虽然反向免疫可以丰富TSA候选物,但是至少90%的这些候选物是假阳性<sup>5,13</sup>,因为可用的计算方法可以预测MHC结合,但是它们无法预测MAP加工过程中的其他步骤<sup>14,15</sup>。为了克服这种限制,一些研究将质谱(MS)分析纳入了他们的TSA发现途径<sup>16</sup>,从而提供了若干个TSA的严格分子定义<sup>17,18</sup>。然而,这些方法的收益极其贫乏:在黑色素瘤中(最易突变的肿瘤类型之一),用MS验证了每个单个肿瘤有平均2个TSA<sup>19</sup>,而对其他癌症类型只发现了极少量的TSA<sup>15</sup>。TSA的缺乏是令人迷惑的,因为如果肿瘤不表达免疫原性抗原,TIL或免疫检查点抑制剂的注射将不会引起肿瘤消退<sup>20</sup>。据推测,基于外显子突变的方法不能鉴定TSA,因为它们没有考虑到两个关键因素。第一,这些方法只关注mTSA而忽略aeTSA,本质上是因为目前没有用于aeTSA的高通量鉴定的方法。这代表了主要的缺点因为,虽然mTSA是稀有抗原,但aeTSA将成为疫苗开发的优选的靶标,因为它们可以被多种肿瘤共享<sup>7,9</sup>。第二,集中于将外显子组作为TSA的唯一来源是非常受限的。外显子组(即,所有蛋白质编码基因)只占人基因组的2%,而最多达75%的基因组可以被转录和潜在翻译<sup>22</sup>。

[0123] 在本文所述的研究中,本发明者已经开发了能够鉴定非耐受性TSA的蛋白质基因组学工作流程,无论它们源自编码区还是非编码区,是简单重排还是复杂重排或仅仅是癌症限制的ERE。为了鉴定非耐受性序列(而不是尝试绘制所有的RNA序列读段)并重建其中的潜在突变,正确的正常匹配的信号(即mTEC<sup>hi</sup>的其中之一)减去癌症信号,并使用所得的序列的经计算机翻译作为用于MS的数据库。与其他技术相比,本文所述的k-mer分析工作流程具有以下若干个优点:(i)它是快速的。产生增大的癌症数据库的源自k-mer的部分典型地需要少于半天。(ii)它是无偏倚的。如鉴定来自非编码区的TSA以及源自7,500个碱基对的删除的一个TSA所证实的,它捕获了所有的癌症特异性序列,不管它们的性质如何。(iii)它是模块化的。为了富集非耐受性基因序列,癌症数据在mTEC<sup>hi</sup>上过滤,mTEC<sup>hi</sup>被示出为抗原外周表达的良好代用品,但是数据可能会以其他方式过滤,例如通过移除所有ENCODE数据或将dbSNP添加到混合中。产生相关的k-mer数据库并被添加到待过滤的正常样品集中。

[0124] 在一方面,本发明提供了一种用于在肿瘤细胞样品中鉴定肿瘤抗原候选物的方法,所述方法包括:

[0125] (a) 通过以下产生肿瘤特异性蛋白质组数据库:

[0126] (i) 从肿瘤RNA序列(例如,通过肿瘤细胞样品的全转录物组测序获得的RNA序列)提取包括至少33个碱基对的子序列集合(k-mers);

[0127] (ii) 将(i)中的所述肿瘤子序列集合与从来自正常细胞的RNA序列提取的包括至少33个碱基对的相应的对照子序列集合比较;

[0128] (iii) 提取在所述相应的对照子序列中不存在或至少低表达至四分之一的所述肿瘤子序列,从而获得肿瘤特异性子序列;以及

[0129] (iv) 经计算机翻译所述肿瘤特异性子序列,从而获得所述肿瘤特异性蛋白质组数据库;

[0130] (b) 通过以下产生个性化肿瘤蛋白质组数据库:

[0131] (i) 将所述肿瘤RNA序列与参考基因组序列比较来鉴定所述肿瘤RNA序列中的单碱基突变;

[0132] (ii) 在所述参考基因组序列中插入在(i)中鉴定的所述单碱基突变,从而创建个性化基因组序列;

[0133] (iii) 经计算机翻译来自所述个性化基因组序列的编码表达蛋白质的转录物,从而获得所述个性化蛋白质组数据库;

[0134] (c) 将来自所述肿瘤的主要组织相容性复合体(MHC)相关肽(MAP)的序列与(a)的所述肿瘤特异性蛋白质组数据库以及(b)的所述个性化肿瘤蛋白质组数据库的序列比较以鉴定所述MAP;以及

[0135] (d) 从在(c)中鉴定的MAP之中鉴定肿瘤抗原候选物,其中肿瘤抗原候选物是其序列和/或编码序列在肿瘤细胞中相对于正常细胞过表达的肽。

[0136] 在实施方式中,上述方法进一步包括从所述肿瘤细胞样品中分离主要组织相容性复合体(MHC)相关肽(MAP)并对其测序。

[0137] 在实施方式中,上述方法进一步包括对所述肿瘤细胞样品进行全转录物组测序,从而获得所述肿瘤RNA序列。

[0138] 如本文所用,术语“肿瘤抗原候选物”是指与主要组织相容性分子(MHC)结合且仅存在于肿瘤细胞表面或以相对于非肿瘤细胞的显著更高的水平/频率(至少2倍,优选地至少4、5或10倍)存在于肿瘤细胞表面的肽。可以靶向这样的肿瘤抗原候选物以诱导针对在其表面表达抗原的肿瘤细胞的T细胞应答。

[0139] 用于从细胞样品中分离MHC相关肽(MAP)的方法是本领域所熟知的。最常用的技术是从活细胞中洗脱MHC相关肽的弱酸洗脱(MAE),如Fortier et al. (J. Exp. Med. 205 (3) : 595-610, 2008)所描述的。另一种技术是肽-MHC I类复合体的免疫沉淀或亲和纯化,紧接着是肽洗脱(参见例如, Gebreselassie et al., Hum Immunol. 2006 November; 67 (11) : 894-906)。基于后者的两种高通量策略已被实现。第一种是基于具有编码可溶性分泌MHC(缺乏功能性跨膜结构域)的表达载体的细胞系的转染以及与分泌的MHC相关的肽的洗脱(Barnea et al., Eur J Immunol. 2002 Jan; 32 (1) : 213-22; and Hickman HD et al., J Immunol. 2004 Mar 1; 172 (5) : 2944-52)。第二种方法依赖于化学或代谢标记以提供MHC相关肽的定量谱(Weinzierl AO et al., Mol Cell Proteomics. 2007 Jan; 6 (1) : 102-13. Epub 2006 Oct 29; Lemmel C et al., Nat Biotechnol. 2004 Apr; 22 (4) : 450-4. Epub 2004 Mar 7; Milner E, Mol Cell Proteomics. 2006 Feb; 5 (2) : 357-65. Epub 2005 Nov 4)。

[0140] 在进一步分析之前,洗脱后的MAP可以经受任何纯化/富集步骤,包括尺寸排阻色谱或超滤(使用具有的截留值为约5000Da,例如约3000Da的过滤器)、反相色谱(疏水色谱)

和/或离子交换色谱(例如,阳离子交换色谱)。洗脱后的MAP的序列可以使用本领域已知的任何肽/蛋白测序方法,诸如质谱(串联质谱或MS/MS,如下文所述)和埃德曼降解反应(Edman degradation reaction)来测定。

[0141] 全转录物组测序(也称为“总RNA测序”、“RNA测序”或RNA-seq)是指对样品中(肿瘤样品,正常细胞样品)所有RNA,包括编码RNA以及多种形式的非编码RNA,诸如miRNA、snRNA和tRNA进行测序。全转录物组测序的方法,例如下一代测序(NGS)技术是本领域所熟知的。可商业获得的(例如来自Illumina(NextSeq™,HiSeq™)、ThermoFisher(Ion Total™ RNA-Seq Kit)、Clontech(SMARTer™)或文献中提到的(例如详细描述于Zhang et al.2011:The impact of next-generation sequencing on genomics.J.Genet Genomics 38(3),95-109;or in Voelkerding et al.2009:Next generation sequencing:From basic research to diagnostics.Clinical chemistry 55,641-658中的那些)的多种NGS平台可以用于本文所述的方法。

[0142] 优选地,RNA制剂作为用于NGS的起始材料。这样的核酸很容易地从样品,诸如生物材料,例如,从新鲜、快速冷冻或福尔马林固定的石蜡包埋肿瘤组织或从新鲜分离的细胞或从存在于患者外周血中的循环肿瘤细胞(CTC)中获得。正常或对照RNA可以从正常、体细胞或生殖细胞中提取。来自正常细胞的RNA序列可以对应于不同类型的正常细胞,例如,来自不同组织的正常细胞的RNA序列集。来自正常细胞的RNA序列也可以从胸腺细胞,优选地髓胸腺上皮细胞(mTEC)诸如MHC II<sup>高</sup>髓胸腺上皮细胞(mTEC<sup>hi</sup>)获得。mTEC<sup>hi</sup>细胞有利地具有独特的混杂基因表达谱,因为它们表达70至90%的体细胞的蛋白编码序列,并且它们的MAP可以诱导中枢免疫耐受。

[0143] 本文所述方法包括使用无比对RNA-seq分析工作流程,称作k-mer剖析,产生肿瘤特异性蛋白质组数据库,其包括源自结构变体(包括大插入或缺失(InDels)或融合的任何类型的突变)以及非编码区的翻译的序列。肿瘤和正常RNA序列(RNA-seq读段)被“切碎”或“分割”成k-mer,即长度为k且 $k \geq 33$ 个核苷酸的子序列。因为结合至MHC I类分子(MAP)的肽通常不多于11个氨基酸长度(以及因此由33个核苷酸长度的序列编码),将RNA序列分割成至少33个核苷酸的子序列最小化丢失潜在的MAP的风险。本领域技术人员将会理解为了最小化肿瘤特异性蛋白质组数据库的大小,将RNA序列分割成33个核苷酸的子序列(即 $k=33$ 个核苷酸)对于鉴定MHC I类限制性肿瘤抗原是优选的。本领域技术人员也将会理解为了鉴定MHC II类限制性肿瘤抗原,最小的k-mer长度应当从33增加至54个核苷酸( $k \geq 54$ 个核苷酸),MHC II类相关肽长度范围通常为从13至18个氨基酸。然后将肿瘤子序列与相应的对照子序列集合(来自正常细胞的RNA序列)比较,以提取在相应的对照子序列中不存在或至少低表达至四分之一(优选地至少低表达至五分之一、六分之一、七分之一、八分之一、九分之一或十分之一)的肿瘤子序列。在实施方式中,为了最小化k-mer空间固有的冗余,所述方法进一步包括将重叠的肿瘤特异性子序列组装成较长的肿瘤子序列(典型地称为重叠群)。然后将肿瘤特异性子序列或重叠群经计算机翻译(例如,翻译成3框或6框,取决于子序列或重叠群是源自编码链还是非编码链)以获得肿瘤特异性蛋白质组数据库。在实施方式中,从所述肿瘤特异性蛋白质组数据库中移出少于8个氨基酸(MHC I类肽的最小长度)或13个氨基酸(MHC II类肽的最小长度)的蛋白质片段。

[0144] 在实施方式中,所述方法进一步包括从RNA序列(来自正常和肿瘤细胞)产生 $k=24$

个核苷酸(对于MHC I类肽)或 $k=39$ 个核苷酸(对于MHC II类肽)的 $k$ -mer数据库以获得癌症/肿瘤和正常24(或39)个核苷酸长度的 $k$ -mer数据库。这些数据库可以用于与MAP编码序列(MCS)比较,以确定肿瘤细胞中MCS是否过表达或占比过大,如下文所述的。

[0145] 方法也包括产生个性化肿瘤蛋白质组数据库。为此,将肿瘤RNA序列(肿瘤RNA-seq读段)与参考基因组序列比较以鉴定肿瘤RNA序列中的单碱基突变。然后将这些突变插入至参考基因组中,以获得个性化肿瘤基因组,从中可以获得包含所有表达的蛋白质编码转录物序列的经典翻译产物序列的相应的个性化肿瘤蛋白质组数据库。允许鉴定WT MAP和由外显子组的经典框编码的突变TSA(新抗原)的个性化肿瘤蛋白质组数据库的产生也通过不将数据库过度偏向于肿瘤特异性序列来提高用于MS分析的数据库的可靠性,这种偏向将导致几种假阳性的鉴定。

[0146] 在实施方式中,方法也包括产生个性化正常蛋白质组数据库。为此,将来自正常细胞的RNA序列(正常RNA-seq读段)与参考基因组序列比较以鉴定正常RNA序列中的单碱基突变。然后将这些突变插入至参考基因组中,以获得个性化正常基因组,从中可以获得包含所有表达的蛋白质编码转录物序列的经典翻译产物序列的相应的个性化正常蛋白质组数据库。这种个性化正常蛋白质组数据库可以被用来过滤在正常(非肿瘤)细胞中表达的MAP,这些MAP不是合适的TSA候选物。

[0147] 如本文所用,术语“参考基因组”是指文献中报道的人基因组组装,并且包括例如Genome Reference Consortium Human Build 38(GRCh38,RefSeq:accession No.GCF\_000001405.37),Hs\_Celera\_WGSA(Celera Genomics;Istrail S.et al.,Proc Natl Acad Sci USA.2004;101(7):1916-21).Epub 2004Feb 9)、HuRef and HuRefPrime(J.Craig Venter Institute;Levy S,et al.PLoS Biology.2007;5:2113-2144)、YH1 and BGIAF(Beijing Genomics Institute;Li R,et al.Genome Research.2010;20:265-272),HsapALLPATHS1(Broad Institute)等。参考人基因组组装的列表可以在国家生物技术信息中心(NCBI)的“组装”数据库中找到。在实施方式中,参考基因组是GRCh38。

[0148] 然后将将在方法的步骤(a)中得到的MAP的序列与肿瘤特异性蛋白质组数据库以及允许MAP鉴定的个性化肿瘤蛋白质组数据库比较(例如,blast比对)。

[0149] 可以在以上鉴定的MAP中鉴定出肿瘤抗原候选物。这样的肿瘤抗原候选物对应于其序列和/或编码序列在肿瘤细胞中相对于正常细胞过表达的多肽。

[0150] 在实施方式中,方法进一步包括消除或丢弃其序列在正常的个性化蛋白质组数据库中检测到的MAP。

[0151] 在实施方式中,方法包括检索已鉴定的MAP的编码序列,即MAP编码序列(MCS)。在其他实施方式中,方法包括将MCS转化为24个(对于MHC I类肽)或39个(对于MHC II类肽)核苷酸的 $k$ -mer集合。在其他实施方式中,将源自MCS的这些 $k$ -mer集合与癌症/肿瘤和正常24(或39)个核苷酸 $k$ -mer数据库比较。

[0152] 在实施方式中,方法包括:

[0153] (a) 从肿瘤细胞样品中分离主要组织相容性复合体(MHC)相关肽(MAP)并对其测序;

[0154] (b) 对所述肿瘤细胞样品进行全转录物组测序,从而获得肿瘤RNA序列;

[0155] (c) 通过以下产生肿瘤特异性蛋白质组数据库:

- [0156] (i) 从所述肿瘤RNA序列提取包括至少33个核苷酸的子序列(k-mers)集合;
- [0157] (ii) 将(i)中的所述肿瘤子序列集合与从来自正常细胞的RNA序列提取的包括至少33个核苷酸的相应的对照子序列集合比较;
- [0158] (iii) 提取在所述相应的对照子序列中不存在或至少低表达至四分之一的所述肿瘤子序列,从而获得肿瘤特异性子序列;以及
- [0159] (iv) 经计算机翻译所述肿瘤特异性子序列,从而获得所述肿瘤特异性蛋白质组数据库;
- [0160] (d) 通过以下产生个性化肿瘤蛋白质组数据库:
- [0161] (i) 将所述肿瘤RNA序列与参考基因组序列比较来鉴定所述肿瘤RNA序列中的单碱基突变;
- [0162] (ii) 在所述参考基因组序列中插入在(i)中鉴定的所述单碱基突变,从而创建个性化肿瘤基因组序列;
- [0163] (iii) 经计算机翻译来自所述个性化肿瘤基因组序列的编码表达蛋白质的转录物,从而获得所述个性化肿瘤蛋白质组数据库;
- [0164] (e) 通过以下产生个性化正常肿瘤蛋白质组数据库:
- [0165] (i) 将来自正常细胞的RNA序列与参考基因组序列比较以鉴定所述正常RNA序列中的单碱基突变;
- [0166] (ii) 在所述参考基因组序列中插入在(i)中鉴定的所述单碱基突变,从而创建个性化正常基因组序列;
- [0167] (iii) 经计算机翻译来自所述个性化正常基因组序列的编码表达蛋白质的转录物,从而获得所述个性化正常蛋白质组数据库;
- [0168] (f) 通过(i)从所述来自正常细胞的RNA序列和所述肿瘤RNA序列提取包括至少24个核苷酸的子序列集合来产生正常和肿瘤k-mer数据库;
- [0169] (g) 将在(a)中获得的MAP的序列与(c)的所述肿瘤特异性蛋白质组数据库以及(d)的所述个性化肿瘤蛋白质组数据库的序列比较以鉴定所述MAP;以及
- [0170] (h) 从所述在(f)中鉴定的MAP之中鉴定肿瘤抗原候选物,其中肿瘤抗原候选物对应于以下的MAP:(1)其序列在所述个性化正常蛋白质组数据库中不存在;以及(2)(i)其序列在所述个性化肿瘤蛋白质组数据库中不存在;和/或(i)其编码序列在所述肿瘤k-mer数据库中相对于所述正常k-mer数据库过表达或占比过大。
- [0171] 在实施方式中,将编码序列转化成源自MAP的k-mer集合(例如,24nt的k-mer),以及确定源自MAP的k-mer在肿瘤和正常k-mer数据库中的表达或占比。如本文所用,过表达或占比过大意指该序列以相对于正常k-mer数据库的至少2倍、优选地至少3、4、或5倍以及更有选地至少10倍的水平存在于肿瘤k-mer数据库中。在实施方式中,编码序列或源自MAP的k-mer在正常的k-mer数据库中是不存在的。
- [0172] 在实施方式中,参考图7A,TSA候选物的鉴定和验证如下实现:在相关癌症和正常个性化蛋白质组或癌症和正常24个核苷酸长度k-mer数据库中查询每个MAP及其相关的一个或多个MAP编码序列(MCS)。在正常个性化蛋白质组中检测到的MAP被排除在外。只在癌症个性化蛋白质组和/或癌症k-mer数据库中存在的MAP被鉴定/选择作为TSA候选物。对于在这两种个性化蛋白质组中不存在的但在两种k-mer数据库中都存在的MAP,如果它们在癌症

细胞中相对于正常细胞过表达(例如,至少2倍、优选地至少5倍以及更优选地至少10倍),则它们被选择。如果MAP由若干种MCS编码,如果它们各自的MCS是一致的,即如果其被一致地标记为TSA候选物,则将其鉴定/选择为TSA候选物。在实施方式中,具有I/L变体的TSA候选物由于它们难以被MS区分,因此被排除在TSA候选物之外。

[0173] 在实施方式中,在比较之前,将洗脱后的MAP过滤以选择8至11个氨基酸长度的肽。在其他实施方式中,在比较之前,将洗脱后的MAP过滤以选择至少一个相关的MHC I类分子的百分等级 $\leq 2\%$ 的那些,如NetMHC software version 4.0预测的(<http://www.cbs.dtu.dk/services/NetMHC-4.0>) (Andreatta M, Nielsen M, Bioinformatics (2016) Feb 15; 32(4):511-7; Nielsen M, et al., Protein Sci., (2003) 12:1007-17)。

[0174] 在实施方式中,方法进一步包括将肿瘤抗原候选物的编码序列与正常组织的序列比较。在实施方式中,使用至少5、10、15、20或25种不同组织的序列。来自正常组织的序列可以从公共数据库诸如Expression Atlas (Petryszak et al., Nucleic Acids Research, Volume 44, Issue D1, 4 January 2016, Pages D746-D752)、scRNASeqDB (Cao Y, et al. (2017). Genes 8(12), 368)、RNA-Seq Atlas (Krupp et al., Bioinformatics, Volume 28, Issue 8, 15 April 2012, Pages 1184-1185)和Encode获得或者可以对正常组织进行RNA-seq来产生。在实施方式中,如果(1)其编码序列在任何正常组织中都不表达,或者如果它只在MHC I类阴性组织中表达或(2)其编码序列在评估的MHC I类阳性组织中表达小于50%,优选地小于45%、40%、35%或30%,则所述方法进一步包括选择肿瘤抗原候选物。在实施方式中,如果肿瘤抗原候选物的编码序列在少于7个,优选地少于6个、5个、4个或3个的评估的正常组织中表达,则所述肿瘤抗原候选物被选择。

[0175] 在实施方式中,方法进一步包括确定TSA候选物的基因编码序列的基因组位置,以及如果(1)编码序列匹配至一致的基因组位置;(2)编码序列不匹配至高可变区(诸如TCR基因的Ig, H2)或者至多个基因以及(3)没有重叠同义突变,则包括选择TSA候选物。可以使用来自UCSC Genome Browser的BLAT工具来进行这样的测定(Kent WJ. Genome Res. 2002 Apr; 12(4):656-64)和/或the Integrative genomics viewer (IGV) tool (Robinson et al., Nat Biotechnol. 2011 Jan; 29(1):24-6)。

[0176] 在实施方式中,方法进一步包括测定和预测肿瘤抗原候选物(TSA候选物)与MHC I类分子的结合。这种结合可能是肽与等位产物的预测的结合亲和性( $IC_{50}$ ),其可以通过工具诸如NetMHC获得。各种可用的MHC I类肽结合工具的综述在Peters B et al., PLoS Comput Biol 2006, 2(6):e65; Trost et al., Immunome Res 2007, 3(1):5; Lin et al., BMC Immunology 2008, 9:8)中提供。鉴定的TSA候选物与MHC I类分子的结合可以用其他已知的方法,例如T2肽结合测定法来测定。T2细胞系缺乏TAP,但仍在细胞表面表达低量的MHC I类。T2结合测定法是基于肽在T2细胞系表面稳定MHC I类复合体的能力。将T2细胞与特定的肽(例如,TSA候选物)孵育,使用pan-HLA I类抗体检测稳定的MHC I类复合体,进行分析(例如,通过流式细胞术),以及相对于非结合阴性对照评估结合。表面上的稳定的肽/MHC I类复合体的存在表明该肽(例如,候选物TSA)与MHC I类分子结合。

[0177] 也可以根据其抑制放射性标记探针肽与MHC分子结合的能力评估感兴趣的肽(例如,TSA候选物)与MHC的结合。MHC分子用表面活性剂增溶并用亲和色谱法纯化。然后在蛋白酶抑制剂混合物(cocktail)的存在下,将它们与感兴趣的肽(例如,TSA候选物)以及过量的

放射性标记探针肽在室温下孵育2天。在孵育期结束时,通过尺寸排阻凝胶过滤色谱将MHC肽复合体与未结合的放射性标记肽分离,并确定结合放射性的百分数。特定肽对MHC分子的结合亲和性可以通过各种剂量的未标记的竞争肽与MHC分子以及标记的探针肽共孵育来确定。抑制50%的标记肽的结合所需的未标记肽的浓度( $IC_{50}$ )可以通过绘制剂量对抑制%来确定(参见例如,Current Protocols in Immunology(1998)18.3.1-18.3.19,John Wiley& Sons,Inc.)。

[0178] 鉴定的TSA候选物与MHC I类分子的结合也可以使用T细胞表位发现系统/工具,诸如ProImmune REVEAL®& ProVE®T细胞表位发现系统或NetMHC工具来确定(参见例如,Desai and Kulkarni-Kale,Methods Mol Biol.2014;1184:333-64)。

[0179] 在实施方式中,方法进一步包括评估细胞群,例如,来自受试者的细胞样品(例如,PBMC)中,识别肿瘤抗原候选物的T细胞的数目或频率。识别给定抗原的T细胞的数量或频率可以使用本领域熟知的各种方法评估,例如通过将细胞群与在肽结合槽中包括所述肿瘤抗原候选物的多聚体MHC I类分子(例如,MHC四聚体)接触,并确定用多聚体MHC I类分子标记的细胞的数目。多聚体MHC I类分子可以用荧光团可检测地标记(直接标记),或用已标记的配体识别的部分来标记(间接或次级标记)。替代地,识别TSA候选物的T细胞的数目或频率可以通过在用于激活T细胞的合适的条件下,在TSA候选物的存在下确定激活的T细胞的数目/频率来评估。激活的T细胞的数目/频率可以通过检测分泌由T细胞激活所诱导的细胞因子,例如,IFN- $\gamma$ 或IL-2的细胞来评估(例如,通过ELISpot或流式细胞术)。

[0180] 在实施方式中,方法进一步包括评估肿瘤抗原候选物诱导T细胞激活的能力,例如,通过将T细胞群与具有在它们的细胞表面处与MHC I类分子结合的肿瘤抗原候选物的细胞(例如,APC,诸如树突状细胞)接触,并测量T细胞激活的至少一个参数,诸如增殖、细胞因子/趋化因子产生(例如,IFN- $\gamma$ 或IL-2产生)、细胞毒性杀伤等。

[0181] 在实施方式中,所述方法进一步包括评估肿瘤抗原候选物诱导T细胞介导的肿瘤细胞杀伤和/或抑制肿瘤生长的能力。这可以使用肿瘤细胞在体外实现,也可以使用合适的动物模型在体内实现。

[0182] 在实施方式中,肿瘤抗原候选物具有的长度为约7至20个氨基酸,以及更特别地约8至18个氨基酸,优选地长度为8至11(对于MHC I类肿瘤抗原)或13至18(对于MHC II类肿瘤抗原)个氨基酸。

[0183] 本文所述方法可用于通过对感兴趣的肿瘤/癌细胞样品进行全转录组测序来鉴定任何类型癌症的肿瘤抗原候选物。这样的癌症的实例包括但不限于癌、淋巴瘤、母细胞瘤、肉瘤、以及白血病,以及更特别地骨癌、血液/淋巴癌诸如白血病(AML、CML、ALL)、骨髓瘤、淋巴瘤、肺癌、肝癌、胰腺癌、皮肤癌、头或颈癌、皮肤或眼内黑色素瘤、子宫癌、卵巢癌、直肠癌、肛区癌、胃癌、结肠癌、乳腺癌、前列腺癌、子宫癌、性器官和生殖器官癌、食管癌、小肠癌、内分泌系统癌、甲状腺癌、甲状旁腺癌、肾上腺癌、软组织肉瘤、膀胱癌、肾癌、肾细胞癌、肾盂癌、中枢神经系统(CNS)肿瘤、神经外胚层癌、脊髓轴肿瘤、胶质瘤、脑膜瘤、以及垂体腺瘤。因此,在实施方式中,本文所述的方法的步骤(a)中使用的肿瘤细胞样品是包括上述癌症中的任何一种的细胞的样品。

[0184] 在其他方面,本公开涉及本文鉴定的肿瘤抗原肽(或肿瘤特异性肽),即,包括表1a、1b、2a-2d或3a-3c(SEQ ID NO:1-39),优选地表2a-2d或3a-3c(SEQ ID NO:17-39)公开

的氨基酸序列或其相对于SEQ ID NO:1-39的序列具有一个或多个突变的变体中的一种。

[0185] 通常,在HLA I类的情况下存在的肽,诸如肿瘤抗原肽的长度在约7或8至约15个,或优选地在8至14个氨基酸残基之间变化。在本公开所述的方法的一些实施方式中,将包括本文定义的肿瘤抗原肽序列的较长的肽人工负载至细胞诸如抗原呈递细胞(APC),由细胞加工,并且肿瘤抗原肽通过MHC I类分子呈递在APC表面。在这种方法中,长于15个氨基酸残基的肽/多肽(即,肿瘤抗原前体肽)可以被负载至APC,由在APC细胞质中的蛋白酶加工,从而提供了本文定义的相应的肿瘤抗原肽,用于呈递。在一些实施方式中,用于产生本文定义的肿瘤抗原肽的前体肽/多肽,是例如1000、500、400、300、200、150、100、75、50、45、40、35、30、25、20或15个氨基酸或更少。因此,本文所述的所有使用肿瘤抗原肽的方法和工艺包括使用较长的肽或多肽(包括天然蛋白),即肿瘤抗原前体肽/多肽,以诱导由细胞(APC)加工以后的“最终”8-14肿瘤抗原肽的呈递。在一些实施方式中,本文所提及的肿瘤抗原肽的长度是约8至14、8至13、或8至12个氨基酸长度(例如长度为8、9、10、11、12或13个氨基酸),足够小以用于直接配合HLA I类分子。在实施方式中,肿瘤抗原肽包括20个氨基酸或更少,优选地15个氨基酸或更少,更优选地14个氨基酸或更少。在实施方式中,抗原肿瘤肽包括至少7个氨基酸,优选地至少8个氨基酸,更优选地至少9个氨基酸。

[0186] 如本文所用,术语“氨基酸”包括天然存在的氨基酸以及用于多肽化学以制备肿瘤抗原肽的合成类似物的其他氨基酸(例如,天然存在的氨基酸、非天然存在的氨基酸、非核酸序列编码的氨基酸等)的L-和D-异构体。天然存在的氨基酸的实例是甘氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸等。其他氨基酸包括例如氨基酸的非遗传编码形式,以及L-氨基酸的保守性替换。天然存在的非遗传编码的氨基酸包括,例如, $\beta$ -丙氨酸、3-氨基-丙酸、2,3-二氨基丙酸、 $\alpha$ -氨基异丁酸(Aib)、4-氨基-丁酸、N-甲基甘氨酸(肌氨酸)、羟脯氨酸、鸟氨酸(例如,L-鸟氨酸)、瓜氨酸、叔丁基丙氨酸、叔丁基甘氨酸、N-甲基异亮氨酸、苯甘氨酸、环己基丙氨酸、正亮氨酸(Nle)、正缬氨酸、2-萘基丙氨酸、吡啶基丙氨酸、3-苯并噻吩基丙氨酸、4-氯苯基丙氨酸、2-氟苯基丙氨酸、3-氟苯基丙氨酸、4-氟苯基丙氨酸、青霉胺、1,2,3,4-四氢-异喹啉-3-羧酸、 $\beta$ -2-噻吩基丙氨酸、甲硫氨酸亚砷、L-高精氨酸(Hoarg)、N-乙酰基赖氨酸、2-氨基丁酸、2-氨基丁酸、2,4,-二氨基丁酸(D-或L-)、对氨基苯丙氨酸、N-甲基缬氨酸、高半胱氨酸、高丝氨酸(HoSer)、磺基丙氨酸、 $\epsilon$ -氨基己酸、 $\delta$ -氨基缬草酸或2,3-二氨基丁酸(D-或L-)等。这些氨基酸在生物化学/肽化学领域中是熟知的。在实施方式中,肿瘤抗原肽只包括天然存在的氨基酸。

[0187] 在实施方式中,本文所述的肿瘤抗原肽包括具有相对于本文提及的序列包含功能等效的氨基酸残基替换的改变的序列的变体肽。例如,序列中的一种或多种氨基酸残基可以被充当功能等效物的极性相似(具有相似的物理化学性质)的另一种氨基酸置换,导致沉默改变(silent alteration)。序列中的氨基酸的置换可以选自氨基酸所属类别的其他成员。例如,带正电荷(碱性)的氨基酸包括精氨酸、赖氨酸和组氨酸(以及高精氨酸和鸟氨酸)。非极性(疏水)的氨基酸包括亮氨酸、异亮氨酸、丙氨酸、苯丙氨酸、缬氨酸、脯氨酸、色氨酸和甲硫氨酸。不带电极性的氨基酸包括丝氨酸、苏氨酸、半胱氨酸、酪氨酸、天冬酰胺和谷氨酰胺。带负电荷(酸性)的氨基酸包括谷氨酸和天冬氨酸。氨基酸甘氨酸可以包括在非极性氨基酸家族或不带电(中性)极性的氨基酸家族。在氨基酸家族内所做的置换通常被理解为保守置换。本文所提及的肿瘤抗原肽可以包括所有的L-氨基酸、所有的D-氨基酸或者

L-和D-氨基酸的混合物。在实施方式中,本文所提及的肿瘤抗原肽包括所有的L-氨基酸。

[0188] 在实施方式中,在包括SEQ ID NO:1-39中所示的序列之一的肿瘤抗原肽的序列中,对与T细胞受体的相互作用没有实质地贡献的氨基酸残基可以被其掺入不会实质地影响T细胞的反应性,也不会消除与相关MHC分子的结合的其他氨基酸替换来修饰。在实施方式中,将肿瘤抗原肽进行变体序列优化以提高MHC结合,即包括一种或多种突变(例如,1、2或3种突变),例如,增强与MHC分子的结合的氨基酸置换。肿瘤抗原肽变体的结合亲和力可以,例如,使用MHC结合预测工具诸如NetMHC4.0;NetMHCpan4.0;以及MHCflurry 1.2.0来评估。可以考虑序列优化的肿瘤抗原肽变体,例如,如果预测序列优化的肿瘤抗原肽变体与特定的HLA的亲合结合力相比天然的肿瘤抗原肽是等效的、或优选地更强。然后可以使用本领域熟知的方法,例如使用ProImmune's REVEAL测定法筛选选择的序列优化的目标肽与特定的HLA体外结合。

[0189] 也可以将肿瘤抗原肽N-和/或C末端加盖或修饰来防止降解、提高稳定性、亲和性和/或吸收、以及因此本公开提供的肿瘤抗原肽的变体具有式 $Z^1-X-Z^2$ ,其中X为SEQ ID NO:1-39,优选地17-39中所述的肿瘤抗原肽的序列。在实施方式中,例如通过部分/化学基团( $Z^1$ )的共价连接来修饰肿瘤抗原肽的氨基末端残基(即N末端游离的氨基基团)(例如,防止降解)。 $Z^1$ 可以是一至八个碳的直链或支链烷基,或酰基基团(R-CO-),其中R是疏水部分(例如,乙酰基、丙酰基、丁烷基、异丙酰基或异丁烷基)或芳酰基基团(Ar-CO-),其中Ar是芳基基团。在实施方式中,芳基基团是 $C_1-C_{16}$ 或 $C_3-C_{16}$ 芳基基团(线性或支链,饱和或不饱和),在进一步实施方式中,饱和的 $C_1-C_6$ 芳基基团(线性或支链)或不饱和 $C_3-C_6$ 芳基基团(线性或支链),例如乙酰基基团( $CH_3-CO-$ ,Ac)。在实施方式中, $Z^1$ 是不存在的。可以例如通过部分/化学基团( $Z^2$ )的共价连接来修饰肿瘤抗原肽的羧基末端残基(即肿瘤抗原肽的C末端游离羧基)(例如,防止降解),例如通过酰胺化(OH基团被 $NH_2$ 基团替换),因此在这样的情况下, $Z^2$ 是 $NH_2$ 基团。在实施方式中, $Z^2$ 可以是羟肟酸基团、腈基基团、酰胺(伯、仲或叔)基团、一至十个碳的脂肪胺诸如甲胺、异丁胺、异戊胺或环己胺、芳族或芳基烷基胺诸如苯胺、萘胺、苄胺、桂皮胺或苯乙胺、醇或 $CH_2OH$ 。在实施方式中, $Z^2$ 是不存在的。在实施方式中,肿瘤抗原肽包括SEQ ID NO:1-39,优选地17-39公开的序列之一。在实施方式中,肿瘤抗原肽由SEQ ID NO:1-39,优选地17-39公开的序列之一组成,即,其中 $Z^1$ 和 $Z^2$ 是不存在的。

[0190] 在实施方式中,本公开提供了结合至HLA-A2等位基因,优选地HLA-A\*02:01等位基因的HLA分子的肿瘤抗原肽,并且包括以下或由以下组成:SEQ ID NO:17-19、27和28中任一序列所示的氨基酸序列之一。

[0191] 在实施方式中,本公开提供了结合至HLA-B40等位基因,优选地HLA-B\*40:01等位基因的HLA分子的肿瘤抗原肽,并且包括以下或由以下组成:SEQ ID NO:20:所示的氨基酸序列。

[0192] 在实施方式中,本公开提供了结合至HLA-A11等位基因,优选地HLA-A\*11:01等位基因的HLA分子的肿瘤抗原肽,并且包括以下或由以下组成:SEQ ID NO:21-23、29-35中任一序列所示的氨基酸序列之一。

[0193] 在实施方式中,本公开提供了结合至HLA-B08等位基因,优选地HLA-B\*08:01等位基因的HLA分子的肿瘤抗原肽,并且包括以下或由以下组成:SEQ ID NO:24或25中所示的氨基酸序列。

[0194] 在实施方式中,本公开提供了结合至HLA-B07等位基因,优选地HLA-B\*07:02等位基因的HLA分子的肿瘤抗原肽,并且包括以下或由以下组成:SEQ ID NO:26或36所示的氨基酸序列。

[0195] 在实施方式中,本公开提供了结合至HLA-A24等位基因,优选地HLA-A\*24:02等位基因的HLA分子的肿瘤抗原肽,并且包括以下或由以下组成:SEQ ID NO:38或39所示的氨基酸序列。

[0196] 在实施方式中,本公开提供了结合至HLA-C07等位基因,优选地HLA-C\*07:01等位基因的HLA分子的肿瘤抗原肽,并且包括以下或由以下组成:SEQ ID NO:37所示的氨基酸序列。

[0197] 在实施方式中,肿瘤抗原肽是白血病肿瘤抗原肽,并且包括以下或由以下组成:SEQ ID No:17-28任一个所示的氨基酸序列之一。

[0198] 在实施方式中,肿瘤抗原肽是肺肿瘤抗原肽,并且包括以下或由以下组成:SEQ ID No:29-39中的任一个所示的氨基酸序列之一。

[0199] 在实施方式中,肿瘤抗原肽是由位于基因组非编码区的序列编码的。在实施方式中,肿瘤抗原肽是由位于非翻译的转录区即3'-UTR或5'-UTR区的序列编码的。在其他实施方式中,肿瘤抗原肽是由位于内含子的序列编码的。在其他实施方式中,肿瘤抗原肽是由位于基因间区的序列编码的。在实施方式中,肿瘤抗原肽是由位于内源性逆转录元件(ERE)的序列编码的。在其他实施方式中,肿瘤抗原肽是由位于外显子且源于移码的序列编码的。

[0200] 本公开的肿瘤抗原肽可以通过在包括核酸编码肿瘤抗原肽的宿主细胞中的表达(重组表达)或通过化学合成(例如,固相肽合成)产生的。肽可以很容易地通过本领域熟知的人工和/或自动化的固相肽程序合成。例如,可以通过利用“T-boc”或“Fmoc”程序进行合适的合成。用于固相合成的技术和程序已描述于例如Solid Phase Peptide Synthesis:A Practical Approach,by E.Atherton and R.C.Sheppard,published by IRL,Oxford University Press,1989。替代地肿瘤抗原肽可以通过段缩合的方法来制备,如例如在Liu et al.,Tetrahedron Lett.37:933-936,1996;Baca et al.,J.Am.Chem.Soc.117:1881-1887,1995;Tam et al.,Int.J.Peptide Protein Res.45:209-216,1995;Schnolzer and Kent,Science 256:221-225,1992;Liu and Tam,J.Am.Chem.Soc.116:4149-4153,1994;Liu and Tam,Proc.Natl.Acad.Sci.USA 91:6584-6588,1994;以及Yamashiro and Li,Int.J.Peptide Protein Res.31:322-334,1988)中描述用于合成肿瘤抗原肽的其他方法在Nakagawa et al.,J.Am.Chem.Soc.107:7087-7092,1985中描述。在实施方式中,肿瘤抗原肽是化学合成的(合成肽)。本公开的其他实施方式涉及非天然存在的肽,其中所述肽包含以下或由以下组成:本质上由本文定义的氨基酸序列;并且已被合成生产(例如,合成的)为药物上可接受的盐。根据本公开的肿瘤抗原肽的盐与它们在体内的一种或多种状态下的肽实质上不同,因为其在体内生成的肽是无盐的。肽的非天然的盐形式可以调整肽的溶解度,特别是在包括肽的药物组合物的背景下,例如,如本文所述的肽疫苗。优选地,该盐是肽的药学上可接受的盐。

[0201] 在实施方式中,本文提及的肿瘤抗原肽是实质上纯的。当化合物与天然伴随它的成分分离时,它是“实质上纯”的。典型地,当按重量计,样品中全部材料的至少60%,更通常地75%、80%或85%,优选地超过90%以及更优选地超过95%是该化合物时,化合物是实

质上纯的。因此,例如,用化学方法合成或通过重组技术生产的多肽将通常地实质上不含其天然相关的组分,例如,其来源大分子的组分。当核酸分子不直接与通常从核酸衍生的有机体的天然存在的基因组中相邻接的编码序列相邻接(即,共价连接)时,核酸分子是实质上纯的。实质上纯的化合物可以例如通过从天然资源提取、通过编码肽化合物的重组核酸分子的表达或者通过化学合成获得。纯度可以使用任何适当的方法,诸如柱色谱法、凝胶电泳、HPLC等来测量。在实施方式中,肿瘤抗原肽为溶液形式。在其他实施方式中,肿瘤抗原肽以固态形式例如,冻干的。

[0202] 在其他方面,本公开进一步提供了编码本文所提及的肿瘤抗原肽或肿瘤抗原前体肽的核酸(分离的)。在实施方式中,核酸包括约21个核苷酸至约45个核苷酸、约24个至约45个氨基酸,例如24、27、30、33、36、39、42或45个核苷酸。如本文所用,“分离的”是指与存在于肽或核酸分子或天然存在的来源大分子(例如,包括其他核酸、蛋白质、脂类、糖等)的天然环境中的其他成分分离的肽或核酸分子。如本文所用,“合成的”是指不是从它的天然来源中分离出的肽或核酸分子,例如,它是通过重组技术或使用化学合成产生的。本公开的核酸可以用于本公开的肿瘤抗原肽的重组表达,并且可以被包括在载体或质粒,诸如可以被转染进入宿主细胞的克隆载体或者表达载体中。在实施方式中,本公开提供了包括编码本公开的肿瘤抗原肽的核酸序列的克隆或表达载体或质粒。替代地,编码本公开的肿瘤抗原肽的核酸可以被并入宿主细胞的基因组中。在任一情况下,宿主细胞表达由核酸编码的肿瘤抗原肽或蛋白质。如本文所用,“宿主细胞”不仅是指特定的受体细胞,而且也指这样的细胞的后代或潜在后代。宿主细胞可以是能够表达本文所述的肿瘤抗原肽的任何原核细胞(例如,大肠杆菌(*E. coli*))或真核细胞(例如,昆虫细胞、酵母或哺乳动物细胞)。载体或质粒包含用于插入的编码序列的转录和翻译的所需的元件,并且可以包含其他元件诸如抗性基因、克隆位点等。可以使用本领域的技术人员所熟知的方法来构建包含编码肽或多肽的序列的表达载体,以及可操作地连接至其的适当的转录和翻译控制/调控元件。这些方法包括体外重组DNA技术、合成技术和体内基因重组。这样的技术在Sambrook.et al.(1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., 以及Ausubel, F.M.et al.(1989) *Current Protocols in Molecular Biology*, John Wiley&Sons, New York, N.Y. 中描述。“可操作地连接”是指元件特别是核酸序列的并置,使得元件的正常功能得以进行。因此,可操作地连接至调控序列的编码序列是指其中编码序列可以在调控控制,即调控序列的转录和/或翻译控制下表达的核苷酸序列的配置。如本文所用,“调控/控制区”或“调控/控制序列”是指涉及编码核酸表达的调控的非编码核苷酸序列。因此,术语调控区域包括启动子序列,调控蛋白质结合位点、上游激活因子序列等。在实施方式中,本公开的编码肿瘤抗原肽的核酸(DNA、RNA)被包括或包裹在囊泡,诸如脂质体内。

[0203] 在其他方面,本公开提供了包括(即,呈递或结合至)肿瘤抗原肽的MHC I类分子。在实施方式中,MHC I类分子是HLA-A2分子,在进一步实施方式中,是HLA-A\*02:01分子。在实施方式中,MHC I类分子是HLA-A11分子,在进一步的实施方式中,是HLA-A\*11:01分子。在实施方式中,MHC I类分子是HLA-A24分子,在进一步的实施方式中,是HLA-A\*24:02分子。在其他实施方式中,MHC I类分子是HLA-B07分子,在进一步的实施方式中,是HLA-B\*07:02分子。在其他实施方式中,MHC I类分子是HLA-B08分子,在进一步的实施方式中,是HLA-B\*08:

01分子。在其他实施方式中,MHC I类分子是HLA-B40分子,在进一步的实施方式中,是HLA-B\*40:01。在其他实施方式中,MHC I类分子是HLA-C07分子,在进一步的实施方式中,是HLA-C\*07:01分子。在实施方式中,肿瘤抗原肽非共价结合至MHC I类分子(即,肿瘤抗原肽被装载至或非共价结合至MHC I类分子的肽结合槽/袋中)。在其他实施方式中,肿瘤抗原肽共价连接/结合至MHC I类分子( $\alpha$ 链)。在这样的构建体中,肿瘤抗原肽和MHC I类分子( $\alpha$ 链)被生产为合成融合蛋白,典型地具有短的(例如,5至20个残基,优选地约8-12个,例如,10个)灵活的接头或间隔子(例如,聚甘氨酸接头)。在其他方面,本公开提供了编码包括本文定义的融合至MHC I类分子( $\alpha$ 链)的肿瘤抗原肽的融合蛋白的核酸。在实施方式中,MHC I类分子( $\alpha$ 链)-肽复合物是多聚体化的。因此,在其他方面,本公开提供了负载(共价或非共价)有本文所提及的肿瘤抗原肽的MHC I类分子的多聚体。这样的多聚体可以连接至标签上,例如允许多聚体的检测的荧光标签。目前已开发了大量策略用于生产MHC多聚体,包括MHC二聚体、四聚体、五聚体、八聚体等(在Bakker and Schumacher, *Current Opinion in Immunology* 2005, 17: 428-433综述)。MHC多聚体可用于例如,检测和纯化抗原特异性T细胞。因此,在其他方面,本公开提供了用于检测或纯化(分离、富集)特异于本文定义的肿瘤抗原肽的CD8<sup>+</sup> T淋巴细胞的方法,该方法包括将细胞群与负载(共价或非共价)有肿瘤抗原肽的MHC I类分子的多聚体接触,以及检测或分离通过MHC I类多聚体结合的CD8<sup>+</sup> T淋巴细胞。通过MHC I类多聚体结合的CD8<sup>+</sup> T淋巴细胞可以通过已知方法,例如荧光激活细胞分选法(FACS)或磁性激活细胞分选法(MACS)分离。

[0204] 在又一方面,本公开提供了包括本文所提及的肿瘤抗原肽、核酸、本公开的载体或质粒,即编码一种或多种肿瘤抗原肽的核酸或载体的细胞(例如,宿主细胞),在实施方式中,是分离的细胞。在另一方面,本公开提供了在其表面表达结合至根据本公开的肿瘤抗原肽的MHC I类分子(例如,上文公开的等位基因之一的MHC I类分子)或呈递根据本公开的肿瘤抗原肽的细胞。在一种实施方式中,宿主细胞是真核细胞,诸如哺乳动物细胞,优选地人细胞、细胞系或永生化细胞。在其他实施方式中,细胞是抗原呈递细胞(APC),诸如树突状细胞(DC)或单核细胞/巨噬细胞(macropage)。在一种实施方式中,宿主细胞是原代细胞(primary cell)、细胞系或永生化细胞。可以经由常规转化或转染技术将核酸和载体引入细胞。术语“转化”和“转染”是指将外来核酸引入宿主细胞的技术,包括磷酸钙或氯化钙共沉淀、DEAE-右旋糖酐-介导的转染、脂质体转染、电穿孔、微注射以及病毒介导的转染。合适的转化或转染宿主细胞的方法可以例如参见Sambrook et al. (见上)以及其他实验室手册。用于将核酸引入哺乳动物体内细胞的方法也是已知的,并可以用于将本公开的载体或质粒递送至基因治疗的受试者。

[0205] 可以使用本领域已知的各种方法用一种或多种肿瘤抗原肽负载细胞,诸如APC。如本文所用,用肿瘤抗原肽“负载细胞”意指将编码肿瘤抗原肽的RNA(mRNA)或DNA或肿瘤抗原肽转染至细胞或替代地用编码肿瘤抗原肽的核酸转化APC。细胞也可以通过与将细胞与外来的肿瘤抗原肽接触来负载,所述肿瘤抗原肽可以直接结合至存在于细胞表面的MHC I类分子(例如,肽冲击的细胞)。也可以将肿瘤抗原肽融合到促进通过MHC I类分子的呈递的结构域或基序中,例如融合至内质网(ER)回收信号(retrieval signal),C末端Lys-Asp-Glu-Leu序列(参见Wang et al., *Eur J Immunol.* 2004Dec; 34(12):3582-94)。

[0206] 在其他方面,本公开提供了包括本文所定义的肿瘤抗原肽(或编码所述一种或多

种肽的核酸)中的任一种或任何组合的组合物或肽组合/库。在实施方式中,组合物包括本文定义的肿瘤抗原肽的任何组合(2、3、4、5、6、7、8、9、10或更多种肿瘤抗原肽的任何组合)或编码所述肿瘤抗原肽的核酸的组合。包括本文所定义的肿瘤抗原肽的组合/子组合的组合物涵盖在本公开中。在实施方式中,组合物或肽组合/库包括至少2、3、4、5、6、7、8、9、10种肿瘤抗原肽,所述肿瘤抗原肽包括以下或由以下组成:SEQ ID NO:17-28所示的序列。在实施方式中,组合物或肽组合/库包括至少2、3、4、5、6、7、8、9、10种肿瘤抗原肽,所述肿瘤抗原肽包括以下或由以下组成:SEQ ID NO:29-39所示的序列。在其他实施方式中,组合物或肽组合/库可以包括一种或多种已知的肿瘤抗原。

[0207] 因此,在其他方面,本公开提供了包括本文所定义的肿瘤抗原肽的任一种或任何组合,以及表达MHC I类分子(例如,如上文公开的等位基因之一的MHC I类分子)的细胞的组合物。本公开使用的APC不限于特定类型的细胞并且包括专职性APC,诸如树突状细胞(DC)、朗格汉斯细胞(Langerhans cell)、巨噬细胞/单核细胞以及B细胞,已知它们在细胞表面呈递蛋白质抗原以便被CD8<sup>+</sup> T淋巴细胞识别。例如,APC可以通过以下获得:从外周血单核细胞诱导DC以及然后与肿瘤抗原肽接触(刺激),无论是在体外、离体还是在体内。APC也可以被激活以在体内呈递肿瘤抗原肽,其中将本公开的肿瘤抗原肽中的一种或多种给药至受试者以及在受试者的体内诱导呈递肿瘤抗原肽的APC。短语“诱导APC”或“刺激APC”包括用一种或多种肿瘤抗原肽或编码肿瘤抗原肽的核酸接触或负载,使得肿瘤抗原肽通过MHC I类分子在其表面呈递。如上所述,根据本公开,肿瘤抗原肽可以例如使用包括肿瘤抗原肽(包括天然蛋白质)的序列的较长的肽/多肽来间接负载,然后所述肿瘤抗原肽在APC内加工(例如,通过蛋白酶)以在细胞的表面产生肿瘤抗原肽/MHC I类复合体。在用肿瘤抗原肽负载APC并允许APC呈递肿瘤抗原肽之后,APC可以作为疫苗给药至受试者。例如,离体给药可以包括以下步骤:(a)从第一受试者中收集APC,(b)将步骤(a)的APC与肿瘤抗原肽接触/负载以在APC的表面形成MHC I类/肿瘤抗原肽复合体以及(c)将肽负载的APC给药至需要治疗的第二受试者。

[0208] 第一受试者和第二受试者可以是同一受试者(例如,自体疫苗),或者可以是不同的受试者(例如,同种异体疫苗)。替代地,根据本公开,使用本文所述的肿瘤抗原肽(或其组合物)用于制造用于诱导抗原呈递细胞的组合物(例如,药物组合物)。此外,本公开提供了用于制造用于诱导抗原呈递细胞的药物组合物的方法或工艺,其中该方法或工艺包括混合或配制肿瘤抗原肽或其与药学上可接受的载体的组合。负载有本文所定义的肿瘤抗原肽的任一种或任何组合的表达MHC I类分子(例如,HLA-A2、HLA-A11、HLA-A24、HLA-B07、HLA-B08、HLA-B40或HLA-C07分子)的细胞诸如APC可以用于刺激/扩增CD8<sup>+</sup> T淋巴细胞,例如自体CD8<sup>+</sup> T淋巴细胞。因此,在其他方面,本公开提供了一种组合物,包括本文所定义的肿瘤抗原肽(或编码所述肿瘤抗原肽的核酸或载体)中的任一种或任何组合;表达MHC I类分子的细胞和T淋巴细胞,更特别地是CD8<sup>+</sup> T淋巴细胞(例如包括CD8<sup>+</sup> T淋巴细胞的细胞群)。

[0209] 在实施方式中,组合物进一步包括缓冲剂、赋形剂、载体、稀释剂和/或介质(例如,培养基)。在进一步实施方式中,缓冲剂、赋形剂、载体、稀释剂和/或介质是药学上可接受的一种或多种缓冲剂、一种或多种赋形剂、一种或多种载体、一种或多种稀释剂和/或一种或多种介质。如本文所用,“药学上可接受的缓冲剂、赋形剂、载体、稀释剂和/或介质”包括生理上相容的、不会干扰一种或多种活性成分的生物活性的有效性并且对受试者是无毒的任

何和所有溶剂、粘合剂、润滑剂、填充剂、增稠剂、崩解剂、增塑剂、包衣、阻挡层制剂、润滑剂、稳定剂、缓释剂、分散介质、包衣、抗细菌和抗真菌剂、等渗剂等。这样的用于药学上的活性物质的介质和制剂的使用是本领域所熟知的 (Rowe et al., Handbook of pharmaceutical excipients, 2003, 4<sup>th</sup> edition, Pharmaceutical Press, London UK)。除非任何常规介质或药剂与活性化合物 (肽、细胞) 不相容, 否则设想在本公开的组合物中使用它们。在实施方式中, 缓冲剂、赋形剂、载体和/或介质是非天然存在的缓冲剂、赋形剂、载体和/或介质。在实施方式中, 本文所定义的肿瘤抗原肽的一种或多种, 或编码所述一种或多种肿瘤抗原肽的核酸 (例如, mRNA) 包括在脂质体内或与脂质体复合, 例如, 阳离子脂质体 (参见例如, Vitor MT et al., Recent Pat Drug Deliv Formul. 2013 Aug; 7(2):99-110)。

[0210] 在其他方面, 本公开提供了一种组合物, 包括本文所定义的肿瘤抗原肽 (或编码所述一种或多种肽的核酸) 中的任一种或任何组合以及缓冲剂、赋形剂、载体、稀释剂和/或介质中的一种或多种。对于包括细胞 (例如, APC、T淋巴细胞) 的组合物, 该组合物包括允许维持活细胞的合适的介质。这样的介质的代表性实例包括生理盐水溶液、Earl平衡盐溶液 (Life Technologies®) 或 PlasmaLyte® (Baxter International®)。在实施方式中, 组合物 (例如, 药物组合物) 是“免疫原性组合物”、“疫苗组合物”或“疫苗”。如本文所用, 术语“免疫原性组合物”、“疫苗组合物”或“疫苗”是指包括一种或多种肿瘤抗原肽或疫苗载体的组合物或制剂, 并且当将其给药至受试者时, 其能够诱导针对其中存在的一种或多种肿瘤抗原肽的免疫应答。用于在哺乳动物中诱导免疫应答的免疫接种方法包括使用通过疫苗领域已知的任何常规路径给药的疫苗或疫苗载体, 例如, 经由黏膜 (例如, 眼、鼻、肺、口、胃、肠、直肠、阴道或泌尿道) 表面、经由肠胃外 (例如, 皮下、皮内、肌内、静脉内或腹腔内) 路径或局部给药 (例如, 经由经皮递送系统诸如贴剂)。在实施方式中, 肿瘤抗原肽 (或其组合) 缀合至载体蛋白质 (缀合物疫苗) 以增强一种或多种肿瘤抗原肽的免疫原性。因此, 本公开提供了包括肿瘤抗原肽 (或其组合) 以及载体蛋白质的组合物 (缀合物)。例如, 一种或多种肿瘤抗原肽可以与Toll样受体 (TLR) 配体 (参见例如, Zom et al., Adv Immunol. 2012, 114:177-201) 或聚合物/树枝状聚合物 (参见例如, Liu et al., Biomacromolecules. 2013 Aug 12; 14(8):2798-806) 缀合。在实施方式中, 免疫原性组合物或疫苗进一步包括佐剂。“佐剂”是指当被添加到免疫原性剂诸如抗原 (根据本公开的肿瘤抗原肽和/或细胞) 中时, 非特异性地增强或加强宿主在暴露于混合物时对药剂的免疫应答的物质。目前用于疫苗领域的实例包括 (1) 矿物盐类 (铝盐诸如磷酸铝和氢氧化铝、磷酸钙凝胶)、角鲨烯, (2) 油类佐剂, 诸如油乳液和基于表面活性剂的制剂, 诸如MF59 (微流化洗涤剂稳定的水包油型乳液)、QS21 (纯化的皂苷), AS02 [SBAS2] (水包油型乳液+MPL+QS-21), (3) 颗粒性佐剂, 例如, 病毒体 (结合流感血凝素的单层脂质体媒介物)、AS04 ([SBAS4] 铝盐与MPL)、ISCOMS (皂苷和脂类的结构复合体)、聚丙交酯-co-乙交酯 (PLG), (4) 微生物衍生物 (天然的和合成的), 例如, 单磷酸基脂质A (MPL)、Detox (MPL+M. Phlei细胞壁骨架)、AGP [RC-529] (合成的酰化单糖)、DC\_Chol (能够自组织成脂质体的脂类免疫刺激物)、OM-174 (脂质A衍生物)、CpG基序 (包含免疫刺激CpG基序的合成的寡核苷酸)、修饰的LT和CT (基因修饰的细菌毒素, 以提供无毒的佐剂作用), (5) 内源性人免疫调节剂, 例如, hGM-CSF或hIL-12 (可以以编码的蛋白质或质粒给药的细胞因子)、Immudaptin (C3d串联阵列) 和/或 (6) 惰性媒介物, 诸如金颗粒等。

[0211] 在实施方式中, 一种或多种肿瘤抗原肽或包括所述一种或多种肿瘤抗原肽的组

物是冻干的形式。在其他实施方式中,一种或多种肿瘤抗原肽或包括所述一种或多种肿瘤抗原肽的组合物是液体组合物。在进一步实施方式中,一种或多种肿瘤抗原肽在组合物中的浓度为约0.01 $\mu\text{g}/\text{mL}$ 至约100 $\mu\text{g}/\text{mL}$ 。在进一步实施方式中,一种或多种肿瘤抗原肽在组合物中的浓度为约0.2 $\mu\text{g}/\text{mL}$ 至约50 $\mu\text{g}/\text{mL}$ 、约0.5 $\mu\text{g}/\text{mL}$ 至约10、20、30、40或50 $\mu\text{g}/\text{mL}$ 、约1 $\mu\text{g}/\text{mL}$ 至约10 $\mu\text{g}/\text{mL}$ 或约2 $\mu\text{g}/\text{mL}$ 。

[0212] 如上所述,表达负载有或结合至本文所定义的肿瘤抗原肽中的任一种或任何组合的MHC I类分子的细胞诸如APC可以用于在体内或离体刺激/扩增CD8<sup>+</sup> T淋巴细胞。因此,在其他方面,本公开提供了能够与本文所提及的MHC I类分子/肿瘤抗原肽复合体相互作用或结合的T细胞受体(TCR)分子,以及编码这样的TCR分子的核酸分子,以及包括这样的核酸分子的载体。根据本公开的TCR能够与肿瘤抗原肽特异性地相互作用或结合,所述肿瘤抗原肽负载于以下或由以下呈递:MHC I类分子,优选地在体内或在体外的活细胞的表面。本公开的TCR,特别是编码TCR的核酸,可以例如被应用于遗传转化/修饰T淋巴细胞(例如,CD8<sup>+</sup> T淋巴细胞)或产生特异性识别MHC I类/肿瘤抗原肽复合体的新的T淋巴细胞克隆体的其他类型的淋巴细胞。在特定的实施方式中,转化获自患者的T淋巴细胞(例如,CD8<sup>+</sup> T淋巴细胞)以表达识别肿瘤抗原肽的一种或多种TCR,以及将转化的细胞给药至患者(自体细胞输液)。在特定的实施方式中,转化获自供体的T淋巴细胞(例如,CD8<sup>+</sup> T淋巴细胞)以表达识别肿瘤抗原肽的一种或多种TCR,以及将转化的细胞给药至受体(同种异体细胞输液)。在其他实施方式中,本公开提供了由编码肿瘤抗原肽特异性TCR的载体或质粒转化/转染的T淋巴细胞例如,CD8<sup>+</sup> T淋巴细胞。在进一步实施方式中,本公开提供了用肿瘤抗原肽特异性TCR转化的自体或同种异体细胞来治疗患者的方法。在又进一步实施方式中,提供了肿瘤抗原特异性TCR在制造用于治疗癌症的自体或同种异体细胞上的用途。

[0213] 在一些实施方式中,在(同种异体干细胞移植(ASCT)、同种异体淋巴细胞输注或自体淋巴细胞输注之前或之后,对使用本公开的组合物(例如,药物组合物)治疗的患者进行治疗。本公开的组合物包括:针对肿瘤抗原肽离体激活的同种异体T淋巴细胞(例如,CD8<sup>+</sup> T淋巴细胞)、负载有肿瘤抗原肽的同种异体或自体APC疫苗、肿瘤抗原肽疫苗和用肿瘤抗原特异性TCR转化的同种异体或自体T淋巴细胞(例如,CD8<sup>+</sup> T淋巴细胞)或淋巴细胞。提供能够识别根据本公开的肿瘤抗原肽的T淋巴细胞克隆的方法可以产生,且能够特异性靶向至在受试者(例如,移植受体),例如ASCT和/或供体淋巴细胞输注(DLI)受体中表达肿瘤抗原肽的肿瘤细胞。因此,本公开提供了编码并表达能够特异性识别或结合肿瘤抗原肽/MHC类I分子复合体的T细胞受体的CD8<sup>+</sup> T淋巴细胞。所述T淋巴细胞(例如,CD8<sup>+</sup> T淋巴细胞)可以是重组(工程化)或天然选择的T淋巴细胞。因此,本说明书提供了用于产生本公开的CD8<sup>+</sup> T淋巴细胞的至少两种方法,包括在有利于触发T细胞激活和扩增的条件下使未分化的淋巴细胞与肿瘤抗原肽/MHC I类分子复合体(典型地表达于细胞,诸如APC的表面)接触的步骤,其可以在体外或体内(即,在用APC疫苗给药的患者中,其中APC负载有肿瘤抗原肽或在用肿瘤抗原肽疫苗治疗的患者中)完成。使用结合至MHC I类分子的肿瘤抗原肽的组合或库可以产生能够识别多种肿瘤抗原肽的群体CD8<sup>+</sup> T淋巴细胞。替代地,肿瘤抗原特异性或靶向T淋巴细胞可以通过克隆编码特异性结合至MHC I类分子/肿瘤抗原肽复合体的TCR(更特别地是 $\alpha$ 和 $\beta$ 链)的一种或多种核酸(基因)在体内或离体生产/产生(即工程化的或重组CD8<sup>+</sup> T淋巴细胞)。可以使用本领域中已知的方法从针对肿瘤抗原肽离体激活(例如,用负

载有肿瘤抗原肽APC)的T淋巴细胞;或者从对肽/MHC分子复合体展现出免疫应答的个体获得编码本公开的肿瘤抗原肽特异性TCR的核酸。本公开的肿瘤抗原肽特异性TCR可以在获自移植受体或移植供体的宿主细胞和/或宿主淋巴细胞中重组表达,并且任选地在体外分化以提供细胞毒性T淋巴细胞(CTL)可以使用任何合适的方法诸如转染(例如,电穿孔)或转导(例如,使用病毒载体)将编码TCR $\alpha$ 链和 $\beta$ 链的一种或多种核酸(一个或多个转基因)引入T细胞(例如,来自待治疗的受试者或另外的个体)。可以在体外使用熟知的培养方法扩增表达对肿瘤抗原肽具有特异性的TCR的工程化CD8<sup>+</sup> T淋巴细胞。

[0214] 本公开提供了由肿瘤抗原肽(即,与在细胞表面表达的MHC I类分子结合的肿瘤抗原肽),或肿瘤抗原肽的组合特异性诱导、激活和/或扩增(扩展)的分离的CD8<sup>+</sup> T淋巴细胞。本公开也提供了包括能够识别根据本公开的肿瘤抗原肽或其组合(即,与MHC I类分子结合的一种或多种肿瘤抗原肽)的CD8<sup>+</sup> T淋巴细胞以及所述一种或多种肿瘤抗原肽的组合物。在其他方面,本公开提供了CD8<sup>+</sup> T淋巴细胞富集的细胞群或细胞培养物(例如CD8<sup>+</sup> T淋巴细胞群),所述CD8<sup>+</sup> T淋巴细胞特异性识别一种或多种MHC I类分子/肿瘤抗原肽复合体,如本文所述的。这样的富集的群通过对特异性T淋巴细胞的进行离体扩增得到,所述扩增使用表达负载有(例如,呈递)本文公开的肿瘤抗原肽的一种或多种的细胞诸如APC。如本文所用,“富集的”意指肿瘤抗原特异性CD8<sup>+</sup> T淋巴细胞在群中的比例显著高于细胞的天然群(即还没有经过特异T淋巴细胞离体扩增的步骤)。在进一步实施方式中,肿瘤抗原肽特异性CD8<sup>+</sup> T淋巴细胞在细胞群中的比例是至少约0.5%,例如至少约1%、1.5%、2%或3%。在一些实施方式中,肿瘤抗原肽特异性CD8<sup>+</sup> T淋巴细胞在细胞群中的比例是约0.5%至约10%、约0.5至约8%、约0.5至约5%、约0.5至约4%、约0.5至约3%、约1%至约5%、约1%至约4%、约1%至约3%、约2%至约5%、约2%至约4%、约2%至约3%、约3%至约5%或约3%至约4%。特异性识别感兴趣的一种或多种MHC I类分子/肽(肿瘤抗原肽)复合体的CD8<sup>+</sup> T淋巴细胞富集的这样的细胞群或培养物(例如,一个CD8<sup>+</sup> T淋巴细胞群)可能用于基于肿瘤抗原的癌症免疫治疗,如以下所述。在一些实施方式中,肿瘤抗原肽特异性CD8<sup>+</sup> T淋巴细胞的细胞群被进一步富集,例如使用基于亲和性的系统诸如负载(共价或非共价)一种或多种的本文所定义的肿瘤抗原肽的MHC I类分子的多聚体。因此,本公开提供了肿瘤抗原肽特异性CD8<sup>+</sup> T淋巴细胞的纯化或分离的群,例如,其中肿瘤抗原肽特异性CD8<sup>+</sup> T淋巴细胞的比例是至少约50%、60%、70%、80%、85%、90%、95%、96%、97%、98%、99%或100%。

[0215] 本公开进一步涉及根据本公开的任何肿瘤抗原肽、核酸、表达载体、T细胞受体、细胞(例如,T淋巴细胞、APC)和/或组合物或其任何组合作为药物或在制造药物中的用途。在实施方式中,药物用于癌症的治疗,例如,癌症疫苗。本公开涉及根据本公开的任何肿瘤抗原肽、核酸、表达载体、T细胞受体、细胞(例如,T淋巴细胞,APC)和/或组合物(例如,疫苗组合物)或其任何组合,用于治疗癌症例如,作为癌症疫苗(例如,治疗性癌症疫苗)的用途。本文鉴定的肿瘤抗原肽序列可以被用于生产适合于以下的合成肽:i)体外致敏(priming)和扩增待注射至肿瘤患者内的肿瘤抗原特异性T细胞,和/或ii)作为疫苗来诱导或加强癌症患者的抗肿瘤T细胞响应。

[0216] 在其他方面,本公开提供了本文所述的肿瘤抗原肽或其组合(例如,肽库)作为疫苗用于治疗受试者的癌症的用途。在其他方面,本公开也提供了本文所述的肿瘤抗原肽或其组合(例如,肽库),作为疫苗用于治疗受试者的癌症的用途。在实施方式中,受试者是肿

瘤抗原肽特异性CD8<sup>+</sup>T淋巴细胞的受体。因此,在其他方面,本公开提供了治疗癌症(例如,减少肿瘤细胞的数目,杀伤肿瘤细胞)的方法,所述方法包括对有此需要的受试者给药(输注)有效量的CD8<sup>+</sup>T淋巴细胞,所述CD8<sup>+</sup>T淋巴细胞识别(即表达TCR与之结合)一种或多种MHC I类分子/肿瘤抗原肽复合体(在细胞诸如APC的表面表达)。在实施方式中,所述方法进一步包括在所述CD8<sup>+</sup>T淋巴细胞的给药/输注之后将有效量的肿瘤抗原肽或其组合、和/或表达负载有一种或多种肿瘤抗原肽的一种或多种MHC I类分子的细胞(例如,APC诸如树突状细胞),给药至所述受试者。在又进一步实施方式中,方法包括将治疗有效量的负载一种或多种肿瘤抗原肽的树突状细胞给药至有此需要的受试者。在又进一步实施方式中,方法包括将治疗有效量的表达结合至由MHC I类分子呈递的肿瘤抗原肽的重组TCR的同种异体或自体细胞给药至有此需要的受试者。

[0217] 在其他方面,本公开提供了识别负载有(呈递)肿瘤抗原肽的一种或多种MHC I类分子的CD8<sup>+</sup>T淋巴细胞或其组合用于治疗受试者的癌症(例如,用于减少肿瘤细胞的数目,杀伤肿瘤细胞)的用途。在其他方面,本公开提供了识别负载有(呈递)肿瘤抗原肽的一种或多种MHC I类分子的CD8<sup>+</sup>T淋巴细胞或其组合用于制备/制造用于治疗受试者的癌症(例如,用于减少肿瘤细胞的数目,杀伤肿瘤细胞)的药物的用途。在其他方面,本公开提供了识别负载有(呈递)肿瘤抗原肽的一种或多种MHC I类分子的CD8<sup>+</sup>T淋巴细胞(细胞毒性T淋巴细胞)或其组合,用于治疗受试者的癌症(例如,用于减少肿瘤细胞的数目,杀伤肿瘤细胞)的用途。在进一步实施方式中,所述用途进一步包括在使用所述肿瘤抗原肽特异性CD8<sup>+</sup>T淋巴细胞之后,使用有效量的肿瘤抗原肽(或其组合);和/或表达负载(呈递)有肿瘤抗原肽的一种或多种MHC I类分子的细胞(例如,APC)。

[0218] 本公开也提供了在受试者中产生针对表达负载有本公开的任何肿瘤抗原肽或其组合的人MHC I类分子的肿瘤细胞的免疫响应的方法,该方法包括给药特异性识别负载有肿瘤抗原肽或肿瘤抗原肽组合的MHC I类分子的细胞毒性T淋巴细胞。本公开也提供了特异性识别负载有本文公开的肿瘤抗原肽中的任一种或肿瘤抗原肽的组合的I类MHC分子的细胞毒素T淋巴细胞用于产生针对表达负载有肿瘤抗原肽或其组合的人I类MHC分子的肿瘤细胞的免疫应答的用途。

[0219] 在实施方式中,本文所述方法或用途进一步包括在治疗/使用前确定由患者表达的HLAI类等位基因,以及给药或使用结合至由患者表达的HLA I类等位基因中的一种或多种的肿瘤抗原肽。例如,如果确定患有B-ALL的患者表达HLA-A2\*01和HLA-B\*08:01,则可以在患者中给药或使用(i)SEQ ID NO:17-19,27和/或28(与HLA-A2\*01结合)以及(ii)SEQ ID NO:24或25(与HLA-B08\*01结合)的肿瘤抗原肽的任何组合。

[0220] 在实施方式中,要治疗的癌症,例如,白血病或肺癌的肿瘤细胞表达本文(SEQ ID NO:17-39)公开的肿瘤抗原肽中的一种或多种。在其他实施方式中,本文所述方法或用途进一步包括确定来自患者的肿瘤细胞是否表达本公开的肿瘤抗原肽(SEQ ID No:17-39)中的一种或多种,以及给药或使用由来自患者的肿瘤细胞表达的肿瘤抗原肽中的一种或多种以治疗癌症。

[0221] 在实施方式中,癌症是血癌或血液系统癌症,例如,白血病、淋巴瘤和骨髓瘤。在实施方式中,癌症是白血病,包括但不限于急性淋巴细胞白血病(ALL)、急性髓性白血病(AML)、慢性淋巴细胞白血病(CLL)、慢性髓性白血病(CML)、毛细胞白血病(HCL)、T细胞性幼

淋巴细胞白血病(T-PLL)、大颗粒淋巴细胞白血病或成人T细胞性白血病。在其他实施方式中,癌症是淋巴瘤,包括但不限于霍奇金淋巴瘤(HL)、非霍奇金淋巴瘤(NHL)、伯基特淋巴瘤、前体T细胞白血病/淋巴瘤、滤泡性淋巴瘤、弥漫大B细胞淋巴瘤、套细胞淋巴瘤、B细胞慢性淋巴细胞白血病/淋巴瘤或MALT淋巴瘤。在进一步的实施方式中,癌症是B细胞白血病,诸如B-ALL。

[0222] 在其他实施方式中,癌症是实体癌,诸如肺癌。在进一步的实施方式中,肺癌是非小细胞肺癌(NSCLC)。在实施方式中,肺癌是肺鳞状细胞肺癌(SQCLC)、腺癌、或大细胞未分化性癌(LCAC)。

[0223] 在实施方式中,本文公开的肿瘤抗原肽、核酸、载体、组合物可以与其他治疗(例如,抗肿瘤疗法),诸如,化疗、免疫治疗(例如,基于CAR T/NK细胞的治疗、基于检查点抑制剂的免疫治疗、基于抗体治疗)、放射治疗或外科手术联合使用。免疫检查点抑制剂的实例包括抑制PD-1、PD-L1、CTLA-4、KIR、CD40、TIM-3或LAG-3的药剂,诸如阻断抗体。用于化疗的药剂的实例包括,例如,安吡啶、博来霉素、白消安、卡培他滨、卡铂、卡莫司汀、苯丁酸氮芥、顺铂、克拉立滨、氯法拉滨、克立他酶(crisantaspase)、环磷酰胺、阿糖胞苷、达卡巴嗪、放线菌素D、柔红霉素、多西他赛、多柔比星、表柔比星、依托泊苷、氟达拉滨、氟尿嘧啶(5-FU)、吉西他滨、gliadelimplants、羟基脲、伊达比星、异环磷酰胺、伊立替康、亚叶酸、脂质体多柔比星(liposomal doxorubicin)、脂质体柔红霉素(liposomal daunorubicin)、洛莫司汀、美法仑、巯嘌呤、美司钠、甲氨蝶呤、丝裂霉素、米托蒽醌、奥沙利铂、紫杉醇(paclitaxel)(紫杉醇(Taxol))、培美曲塞、喷司他丁、丙卡巴肼、雷替曲塞、沙铂、链佐星、替加氟-尿嘧啶、替莫唑胺、替尼泊苷、塞替派、硫鸟嘌呤、拓扑替康、曲奥舒凡、长春碱、长春新碱、长春地辛、长春瑞滨或其组合物。或者,化疗的药剂可以是生物药剂,包括针对HER2抗原的赫赛汀(Herceptin<sup>®</sup>)(曲妥珠单抗)、针对VEGF的安维汀(Avastin<sup>®</sup>)(贝伐珠单抗)或针对EGF受体的抗体,诸如爱必妥(Erbix<sup>®</sup>)(西妥昔单抗)和维克替比(Vectibix<sup>®</sup>)(帕木单抗)。可以在给药/使用本文所述的肿瘤抗原肽、核酸、载体、组合物的之前、期间和之后给药/使用这样的额外的药剂或治疗。

[0224] 用于ALL的现有的治疗典型地包括长春新碱、地塞米松或泼尼松、以及蒽环素药物诸如多柔比星(Adriamycin)或柔红霉素。同种异体干细胞移植(allo-SCT)也可以在高风险患者和具有复发/难治性疾病的患者中进行。正在临床开发的治疗B-ALL的其他药剂包括抗-CD22、抗-CD20和抗-CD19抗体,以及蛋白酶体抑制剂(硼替佐米)、JAK/STAT信号通路抑制剂(鲁索替尼(ruxolitinib)),低甲基化药物(地西他滨)以及PI3K/mTOR抑制剂(参见例如, Terwilliger and Abdul-Hay, Blood Cancer J. 2017 Jun; 7(6):e577)。

[0225] 肺癌的目前的治疗典型地包括外科手术、放射治疗、用小分子酪氨酸激酶抑制剂的化疗(厄洛替尼、克唑替尼)以及用检查点抑制剂诸如抗-PD1抗体(帕博利珠单抗)的免疫治疗(参见例如, Dholaria et al., J Hematol Oncol. 2016; 9:138)。

[0226] 进行本发明的一种或多种模式

[0227] 在以下的非限制性实施例中对本发明的进一步细节进行详细阐述。

[0228] 实施例1:材料与方法

[0229] 小鼠.C57BL/6小鼠获自Jackson实验室(巴尔港,缅因州(Bar Harbor, ME))。小鼠

被安置在特定的无病原体的条件下。

[0230] 细胞系. EL4 T-淋巴母细胞性淋巴瘤细胞系、CT26结直肠癌细胞系和B细胞杂交瘤HB-124获自美国模式培养物集存库(ATCC)。将EL4和CT26细胞在补充有10%热灭活的胎牛血清、1%L-谷氨酰胺和1%青霉素-链霉素的RPMI 1640/HEPES中培养。细胞培养基进一步补充有1%非必需氨基酸以及1%丙酮酸钠或分别仅针对EL4和CT26细胞补充有1%丙酮酸钠。为了产生抗-CDR2抗体,将HB-124细胞在补充有10%热灭活的胎牛血清的IMDM中培养。除非另有说明,所有试剂均购自Gibco®。

[0231] 人原发样品. 本研究中使用的原发白血病样品(四种B-ALL标本:07H103、10H080、10H118和12H018)收集并冷冻保存于Maisonneuve-Rosemont医院(Hôpital Maisonneuve-Rosemont)的魁北克白血病细胞库(Banque de Cellules Leucémiques du Québec)(BCLQ)中。如前所述<sup>1a</sup>,将原发白血病样品移植进入NSG小鼠后在体内扩增。简而言之,1-2x 10<sup>6</sup>个B-ALL细胞被解冻并经由静脉注射移植到8-12周龄的亚致死辐照的(250cGy, 137Cs- $\gamma$ 源)的NSG小鼠内。在出现疾病迹象时处死小鼠,并从机械破坏的脾脏制备细胞悬浮液,或,对于07H103,从脾细胞、骨髓和腹水的混合物制备。之后,使用Ficoll<sup>TM</sup>梯度用于富集在MAP分离之前的B-ALL细胞(参见MPA分离部分)。肺肿瘤活检物(1c2、1c4和1c6)购自Tissue Solutions,并且在MAP分离之前进行均质化(参见MAP分离部分)。对于所有的样品,使用Optitype 1.0版本,用默认参数运行RNA(RNA-Seq)数据获得HLA分型(参见RNA提取、文库的制备和测序部分)。

[0232] 肽. 天然的和<sup>13</sup>C标记的TSA由GenScript合成。如生产商所确定的,天然的和<sup>13</sup>C标记的肽纯度分别大于95%和75%。

[0233] 鼠mTEC<sup>hi</sup>提取. 从5-8周龄的C57BL/6或Balb/c小鼠中分离胸腺,并将其机械破坏以提取胸腺细胞。如前所述进行基质细胞富集<sup>2a</sup>。胸腺基质细胞用生物素化的荆豆(Ulex europaeus)凝集素1(UEA1; Vector Laboratories)、PE-Cy7-缀合的链霉亲和素(BD Biosciences),以及以下抗体染色:Alexa Fluor<sup>TM</sup>700抗-CD45、PE抗-I-A<sup>b</sup>(BD Biosciences)、别藻蓝蛋白-Cy7抗-EpCAM(BioLegend)。使用7-氨基放线菌素D(7-AAD; BD Biosciences)评估细胞活力。活的成熟mTEC(mTEC<sup>hi</sup>)被圈选为7-AAD<sup>-</sup>CD45<sup>+</sup>EpCAM<sup>+</sup>UEA1<sup>+</sup>MHC II<sup>hi</sup>。在三激光FACS AriaIIIu(BD Biosciences,图13A)上分选mTEC<sup>hi</sup>。

[0234] 人TEC和mTEC的提取. 胸腺从3月龄至7年龄的接受矫正心血管手术的个体获得(CHU Saint Justine Research Ethic Board, protocol and biobank#2126)。简而言之,胸腺在4°C保持于包含介质的50ml锥形管中,在手术切除后数小时内切成2-5mm的立方体。对于长期保存,胸腺立方体冷冻在包含有热灭活的人血清/10%DMSO的小瓶中冷冻,且在液氮中保存最多3年。冷冻保存的胸腺样品在干冰上转移,并按照改编自C.Stoeckle et al.<sup>3a</sup>的方案用于分离人TEC和mTEC。胸腺组织被切成小块,然后在37°C下使用2mg/mL的胶原酶A(Roche)和0.1mg DNase I/ml(Sigma-Aldrich)的溶液于RPMI-1640(Gibco)中消化三至五个周期,40min。在第二次消化之后,添加胰蛋白酶/EDTA(Gibco)的溶液,在孵育结束前15min通过添加FBS(Invitrogen)来中和其活性。对于TEC和mTEC分选(图13B),将细胞悬浮液用Pacific蓝-缀合的抗-CD45(BioLegend)、PE-缀合的抗-HLA-DR(BioLegend)、APC-缀合的抗-EpCAM(BioLegend)、Alexa 488-缀合的抗-CDR2(用HB-124杂交瘤生产-参见细胞系部

分-与来自Abcam的Dylight 488快速缀合的试剂盒缀合的,仅用于mTEC样品)染色以及使用7-AAD(BD Biosciences)评估细胞活力。

[0235] RNA提取、文库的制备和测序.对于EL4和CT26细胞,  $5 \times 10^6$  个细胞的副本被用来进行RNA测序.对于C57BL/6和Balb/c mTEC<sup>hi</sup>,从2名雌性和2名雄性中提取最少31,686或16,338个FACS分选的细胞,并一式三份进行RNA测序.对于原发白血病细胞, RNA-seq在  $2.0$  至  $4.0 \times 10^6$  个细胞的单个副本上进行.对于人TEC和mTEC,一个RNA-seq副本/供体使用33,076至84,198个FACS分选的TEC或50,058至100,719个mTEC进行.在所有的情况下,总RNA使用TRIzol(Invitrogen)分离,进一步使用如每个生产商所推荐的RNeasy试剂盒或RNeasy微试剂盒(Qiagen)纯化.对于每一个肺肿瘤活检物(总共三个),总RNA使用生产商所推荐的全Prep DNA/RNA/miRNA通用试剂盒(Qiagen)从30mg组织中分离,并且总RNA用于对每个样品进行一个副本的RNA-Seq.每个鼠类样品(EL4、CT26和鼠类mTEC<sup>hi</sup>) 在Nanodrop 2000(Thermo Fisher Scientific)上被量化,并且RNA质量在2100Bioanalyzer上(Agilent Genomics)被评估,以选择具有的RNA完整性数 $\geq 9$ 的样品.对于人样品(B-ALL、肺肿瘤活检物和人TEC/mTEC),总RNA的量化通过QuBit(ABI)完成,并且总RNA的质量在2100Bioanalyzer上(Agilent Genomics)被评估,以选择RNA完整性数 $\geq 7$ 的样品.使用TruSeq Stranded Total RNA Library Prep Kit(EL4细胞)、KAPA Stranded mRNA-Seq Kit(CT26细胞、C57BL/6mTEC<sup>hi</sup>、人mTEC、肺肿瘤和B-ALL标本)或KAPA RNA HyperPrep Kit(Balb/c mTEC<sup>hi</sup>, human TEC)从对于EL4和CT26细胞的2-4 $\mu$ g、对于鼠类mTEC<sup>hi</sup>的50-100ng、对于B-ALL标本的500ng、对于肺肿瘤活检的4 $\mu$ g、对于人TEC的8-13ng或对于人mTEC的41-68ng来制备总RNA的cDNA文库.这些文库在测序之前进一步通过9-16个循环的PCR扩增.双端RNA测序在Illumina NextSeq<sup>TM</sup>500(Balb/c mTEC<sup>hi</sup>、人TEC和mTEC)或者HiSeq<sup>TM</sup>2000(任何其他样品)上进行,并且产生了每个鼠类和人样品分别为平均175和199x  $10^6$ 的读段。

[0236] 产生经典的癌症和正常蛋白质组.对于所有样品,使用Trimmomatic版本0.35来从RNA-Seq读段上修剪掉测序接头和低质量3'碱基,以及然后使用STAR版本2.5.1b<sup>4a</sup>,在默认参数下(除了--alignSJoverhangMin、--alignMatesGapMax、--alignIntronMax、以及--alignSJstitchMismatchNmax的参数,其中默认值分别被10、200,000、200,000和5-1 5 5替换)运行来与参考基因组进行比对,对于鼠类样品参考基因组为GRCm38.87,对于人样品参考基因组为GRCh38.88.具有的最小交替计数设置为5的单碱基突变使用freeBayes版本1.0.2-16-gd466dde[arXiv:1207.3907]来鉴定,以及导出到被转换为与pyGeno<sup>5a</sup>(可在GitHub获得, <https://github.com/tariqdaouda/pyGeno>.)兼容的不可知SNP文件格式的VCF中.最后用kallisto版本0.43.1[Nicolas L Bray, Harold Pimentel, Páll Melsted and Lior Pachter, Near-optimal probabilistic RNA-seq quantification, Nature Biotechnology 34, 525-527 (2016)]以默认参数运行,以转录物/百万(tpm)来量化转录子表达.值得注意的是, kallisto索引是使用索引功能以及使用从Ensembl下载的适当的\*.cdna.all.fa.gz文件构建的.为了构建每个样品的经典蛋白质组,使用pyGeno来(i)在参考基因组中插入高质量样品特异性单碱基突变(freeBayes质量 $> 20$ ),从而创建个性化外显子组,以及(ii)输出由表达的转录物(tpm $> 0$ )产生的已知蛋白质的一种或多种样品特异性序列.这些蛋白质序列被写入fasta文件,其随后用于质谱(MS)数据库检索(癌症经典蛋白质组)和/或MHC I相关肽(MAP)类别(癌症和正常经典蛋白质组).参见原理图图1B以及统计

数据表4a-b。

[0237] 产生癌症和正常k-mer数据库。对于所有的癌症和正常样品,R1和R2 fastq文件都是独立下载,并且使用Trimmomatic版本0.35以修剪掉测序接头和低质量3'碱基。为了确保所有的读段在转录物编码链上,R1读段是使用FASTX-Toolkit版本0.0.14的fastx\_reverse\_complement功能进行了反向互补。使用jellyfish 2.2.3<sup>6a</sup>、生成了分别用于k-mer剖析和MAP分类的k-mer数据库的33和24个核苷酸长的k-mer数据库(细节参见图7A)。值得注意的是,当多个生物副本(鼠类mTEC<sup>hi</sup>)或当不相关的供体的多个样品(人TEC和mTEC)是可用的,fastq文件被级联起来以为每个条件(C57BL/6,Balb/c或人)生成单个正常k-mer数据库。

[0238] k-mer过滤和癌症特异性蛋白质组的产生.为了提取可能会引起TSA的33个核苷酸长的k-mer,分析只限于在EL4或CT26 k-mer细胞中观察到至少4次、肺肿瘤活检物中至少7次以及原发白血病样品中至少10次的k-mer。然后通过对在相关的mTEC<sup>hi</sup>或人TEC/mTEC k-mer数据库中未表达的那些进行选择来获得癌症特异性k-mer(参见图7B)。这种癌症特异性的k-mer集合被进一步组装成较长的线性序列,称为重叠群。简而言之,提交的33个核苷酸长的k-mer之一被随机选择作为种子,然后从两端延伸,其中连续的k-mer与同一条链上的32个核苷酸重叠(禁用-r选项,因为使用了k-mers的链集(stranded sets))。当没有k-mer可以组装时,即,没有32个核苷酸重叠的k-mer被找到,或当多于一个k-mer符合时,组装过程停止(-a 1线性组装的选项)。如果是这样,则选择新的种子,并且组装过程重新恢复,直到提交的列表中的所有k-mer都使用过一次(参见图7C)。这一步骤是由内部开发的软件NEKTAR(<https://bitbucket.org/eaudemard/nektar>)的kmer\_assembly功能完成的。为了获得蛋白质,使用内部python脚本对至少34个核苷酸长的重叠群进行3框翻译。癌症特异性蛋白随后在内部停止密码子处分开,并且任何得到的至少8个氨基酸长的序列在被纳入相关数据库之前都给予唯一的ID(参见图7D)。参见原理图图1B以及统计数据表4a-b。

[0239] MAP分离.对于EL4和CT26细胞,250x 10<sup>6</sup>个指数型生长细胞的三个生物副本是从指数型生长细胞中制备的。对于所有的原发白血病样品,450至700x 10<sup>6</sup>个细胞的三个生物副本是从新鲜收获的白血病细胞中制备的(参见人类原发样品部分)。MAP如前所述<sup>7a</sup>来获得,但有少许修改:弱酸洗脱(MAE)、然后将肽用Oasis HLB柱脱盐(30mg,Waters)并以3kDa分子量的截留值(Amicon Ultra-4,Millipore)过滤以移除β2-微球蛋白(β<sub>2</sub>M)蛋白质。对于原发白血病样品的其中之一(标本10H080),制备100x 10<sup>6</sup>个细胞的四个额外的副本,通过免疫沉淀(IP)分离MAP,如前所述<sup>1a</sup>。最后,将肺肿瘤活检物(湿重范围为771至1,825mg,参见人类初始样品部分)切成小块(3mm大小的立方体),并向每个组织样品中添加5ml的包含蛋白质抑制剂混合物(Sigma)的冰冷的PBS。首先使用Ultra Turrax T25匀化器(在20,000rpm下20秒,IKA-Labortechnik)将组织均质化两次,然后使用Ultra Turrax T8匀化器(在25,000rpm下20秒,IKA-Labortechnik)均质化一次。然后,将550μl的冰冷的10X裂解缓冲液(10%w/v CHAPS)添加到每个样品中,并且MAP如前所述<sup>1</sup>使用每个样品1mg(1ml)中的针对蛋白A磁珠的共价交联的W6/32抗体的免疫沉淀。无论是何种MAP分离技术,MAP提取物全部使用Speed-Vac干燥并在MS分析之前保持冷冻。

[0240] 质谱分析.将干燥的MAP提取物全部重新悬浮于0.2%甲酸中。对于EL4和CT26,MAP提取物负载在自制的C<sub>18</sub>预处理柱(5mm x 360μm i.d.用C<sub>18</sub> Jupiter Phenomenex填充的)

上、在自制的C<sub>18</sub>分析柱(15cm x 150 $\mu$ m i.d.由C<sub>18</sub> Jupiter Phenomenex填充的)上,其中用0-40%乙腈(0.2%甲酸)56-min梯度,600nl.min<sup>-1</sup>流速,使用nEasy-LC II系统。对于所有的人样品,MAP提取物被负载在自制的C<sub>18</sub>分析柱(15cm x 150 $\mu$ m i.d.由C<sub>18</sub> Jupiter Phenomenex填充的),其中0-40%乙腈(0.2%甲酸、07H103、10H080-MAE、10H118和12H018)56min梯度,或5-28%乙腈(0.2%甲酸、肺肿瘤活检和10H080-IP)100-min梯度,600nl.min<sup>-1</sup>流速,使用nEasy-LC II系统。样品用Q-Exactive Plus (EL4,Thermo Fisher Scientific)或HF(所有其他样品,Thermo Fisher Scientific)分析。对于Q-Exactive Plus,以70,000的分辨率采集的每个完整的MS谱,然后是12MS/MS谱,其中最丰富的多电荷离子被选择进行MS/MS测序,分辨率为17,500、自动增益控制目标为1e6、注入时间为50ms以及碰撞能量为25%。对于Q-Exactive HF,以60,000的分辨率采集的每个完整的MS谱,伴然后是20MS/MS谱,其中最丰富的多电荷离子被选择进行MS/MS测序,分辨率为15,000(CT26、07H103、10H080-MAE、10H118、12H018)或为30,000(肺肿瘤活检物、10H080-IP)、自动增益控制目标为5x 10<sup>4</sup>、注入时间为100ms以及碰撞能量为25%。使用Peaks 8.5 (Bioinformatics Solution Inc.)鉴定肽,并且肽序列从通过连接癌症蛋白质组和癌症特异性蛋白质组获得的相关的全局癌症数据库中检索(参见生成典范的癌症和正常蛋白质组,以及k-mer过滤和产生癌症特异性蛋白质组部分经典)。对于肽鉴定,前体离子和片段离子的耐受分别设置为10ppm和0.01Da。氧化(M)和脱酰胺作用(NQ)的发生被认为是翻译后修饰。

[0241] MAP的鉴定。为了选择MAP,对从Peaks获得的唯一标识的列表过滤,以包括8至11个氨基酸长的肽,所述肽对于相关的MHC I类分子上的至少一个具有的百分等级 $\leq 2\%$ ,如NetMHC 4.0<sup>8a</sup>所预测的。而且,应用本地5%的错误发现率(FDR),其定义为诱饵鉴定的数目除以超过给定Peaks分数阈值的目标鉴定的数目,以限制最终MAP列表中假阳性鉴别的数目。

[0242] TSA候选物的鉴定和验证。为了在所有的鉴定的MAP中鉴定TSA候选物,对每一对MAP/蛋白质进行免疫原性状态分配。为此,将每个MAP及其相关的一个或多个MAP编码序列(MCS)分别在相关癌症和正常个性化蛋白质组或癌症和正常24个核苷酸长度k-mer数据库中查询。在正常经典的蛋白质组中检测到的MAP,不论它们的MCS检测状态如何,都被排除在外,因为它们可能是致耐受性的。真正的癌症特异性的MAP,即,在正常经典的蛋白质组和正常k-mer中均未检测到,被标记为TSA候选物。在两种经典的蛋白质组都不存在但在两种k-mer数据库都存在的MAP,需要使它们的MCS在癌细胞中相对正常细胞过表达至少10倍,这样才可以被标记为TSA候选物(参见图8A)。最后,通过若干种MCS编码的MAP(来自不同的蛋白质)仅在它们各自的MCS一致的情况下才能被标记为TSA候选物,即,如果一致地标记此MAP作为TSA候选物。人工检查所有TSA候选物的MS/MS谱以清除任何假鉴定。此外,具有多个基因组上可能的I/L变体的呈递的序列,当它们可以通过质谱来区分时,进行进一步检查来报告这两种变体,当它们不能通过质谱来区分时,只报告表达最多的变体(参见图8B)。最终,通过使用BLAT(来自UCSC基因组浏览器的工具)在参考基因组(GRCm38.87或GRCh38.88)上映射包含MCS的读段,将基因组定位分配给所有经MS验证的TSA候选物。那些其读段与一致的基因组定位不匹配或匹配至高可变区(诸如MHC、Ig或TCR基因)或多个基因的TSA候选物被排除在外。对于具有一致的基因组定位的那些,使用Integrative Genome Viewer(IGV)<sup>9a</sup>来排除以下TSA候选物:具有关于它们的相关的正常对应体的MCS重叠同义突变的TSA候选

物或对于人TSA候选物,重叠已知种系多态性的那些TSA候选物(即,在dbSNP v.149中列出,图8C)。其余的肽被分类为mTSA或aeTSA候选物,取决于它们MCS是否与癌症特异性突变重叠或不重叠。

[0243] MCS的外周表达.为了评估TAA的和aeTSA候选物的MCS的外周表达,使用了RNA-Seq数据,其来自(1)用ENCODE联合体<sup>10a,11a</sup>对RNA进行测序的22个鼠类组织(表5)或(2)28个外周人组织(50个供体/组织),其已经被GTEx联合体测序,以及已经从GTEx Portal on 04/16/2018(phs000424.v7.p2,图6)下载。简而言之,来自每个组织的RNA测序数据转化为用Jellyfish 2.2.3(使用-C选项)的24个核苷酸长的k-mer数据库,以及用于查询每个MCS的24个核苷酸长k-mer集合。对于每个RNA-Seq实验,与给定MCS完全重叠的读段的数目( $r_{\text{overlap}}$ )是使用k-mer集合的最小发生( $k_{\text{min}}$ )来评估的。确实,据推测 $k_{\text{min}} \sim r_{\text{overlap}}$ ,因为,除了可能多次生成相同的k-mer的低复杂度RNA-Seq读段之外,一个k-mer总是来源于单一的RNA-Seq读段。因此,为了比较在所有组织中的MCS表达水平,该 $r_{\text{overlap}}$ 值转化为每 $10^8$ 个读段检测到的读段的数目(rphm),使用以下公式:  $rphm = \frac{(r_{\text{overlap}} \times 10^8)}{r_{\text{tot}}}$ , 其中 $r_{\text{tot}}$ 表示在给定RNA-

Seq实验中测序的读段的总数目。然后对这样的值进行对数转换( $\log_{10}(rphm+1)$ ),并在给定组织的所有RNA-Seq实验中取其平均值。对于鼠类和人候选物,分别在10个或更少的组织中(在 $rphm > 0$ 下)或在除了肝脏的少于5个组织(在 $rphm > 15$ 下)中展现出外周表达的aeTSA候选物,被认为是真正的aeTSA。这些aeTSA,以及mTSA的特征在表1a-b,2a-d和3a-c中报告。

[0244] TSA候选物的MS验证。对于CT26 TSA候选物和两个EL4 TSA候选物(ATQQFQQL-SEQ ID NO:11 and SSPRGSSTL-SEQ ID NO:13),将先前获得的MS/MS谱与相关的<sup>12</sup>C类似物比较。对于在体内测试的其他五种EL4 TSA候选物(IILEFHSL-SEQ ID NO:12, TVPLNHNTL-SEQ ID NO:14, VNYIHRNV-SEQ ID NO:15, VNYLHRNV-SEQ ID NO:15, VTPVYQHL-SEQ ID NO:16),来自六个额外的EL4副本(450至1,400x  $10^6$ 个细胞/副本)的MAP按照如前所述的方法洗脱和完全加工(参见MAP分离和质谱分析部分)。对于绝对量化,6个EL4副本中的3个各自掺入500fmol的<sup>13</sup>C标记的TSA。对于序列验证,在用PRM MS分析样品之前,采集<sup>12</sup>C TSA候选物的MS/MS谱。简而言之,按预定时间监测了5种肽(每种肽仅在以洗脱时间为中心的10分钟窗口内监测)的PRM采集,由以下组成:一个MS1扫描,紧接着是以HCD模式的靶向MS/MS扫描。对于全谱扫描和串联质谱,自动增益控制和注入时间分别为 $3e6-50ms$ 和 $2e5-100ms$ 。在所有的情况下,Skyline<sup>12a</sup>用于提取每种TSA候选物的内源性MS/MS谱以及将其与相关的<sup>12</sup>C MS/MS谱相比较(序列验证)或用于提取内源的和相关合成的<sup>13</sup>C标记肽的强度(绝对量化)。使用以下式,这些强度被进一步用作对每个副本计算每个细胞的TSA拷贝的数目:  $(n_{\text{synthetic}} \times I_{\text{endogenous}} \times N_A / I_{\text{synthetic}}) \times (1/N_{\text{cells}})$ , 其中 $n_{\text{synthetic}}$ 是合成的<sup>13</sup>C标记的TSA的加标初始摩尔数; $I_{\text{endogenous}}$ 和 $I_{\text{synthetic}}$ 分别是相关内源的和<sup>13</sup>C标记的TSA的强度; $N_A$ 是阿伏伽德罗数; $N_{\text{cells}}$ 是用于弱酸洗脱的初始细胞数。

[0245] 在人TEC和mTEC样品中检测到的转录物的累积数目。将分析限制在六个样品中(2个TEC和6个mTEC)中的至少一个在 $tpm > 1$ 下表达的转录物上,计算每个1-对-1TEC/mTEC比较的斯皮尔曼等级相关系数。然后,使用这些表达的转录物的相同的集合,在分析每个额外的样品时计算检测到的转录物的累积数目(cT)。因为在分析中样品被引入的顺序可能影响cT值,所有样品排列方式的cT值被平均化,并且这些平均数据点用于拟合以下预测曲线(R的

‘nls’ 函数) :  $cT = \frac{a \times (nS-1)}{b + (nS-1)} + c$ , 其中cT是转录物的累积数目以及nS是分析样品的数目。随后将这个方程用于推断通过研究最多达20个样品已经检测的, 以及可以通过简单计算  $\lim_{nS \rightarrow \infty} (cT)$  估计的转录物的数目。

[0246] 产生源自骨髓的树突状细胞 (DC)、小鼠免疫接种以及EL4细胞注射。源自骨髓的DC如前所述<sup>13a, 14a</sup>产生。对于小鼠免疫接种, 来自雄性C57BL/6小鼠的DC用2 $\mu$ M的选择的肽冲击3小时, 然后洗涤。在第14天和第7天, 向8至12周龄的雌性C57BL/6小鼠静脉注射10<sup>6</sup>个单独地肽冲击的DC或辐照的EL4细胞 (10,000cGy)。作为阴性对照, 向C57BL/6雌性小鼠免疫接种未冲击的DC。在第0天和第150天, 向小鼠静脉注射5x 10<sup>5</sup>个EL4细胞, 并且监测其体重减轻、瘫痪或肿瘤生长情况。

[0247] IFN- $\gamma$  ELISpot和亲合力测定法. 如前所述<sup>14a</sup>进行ELISpot亲合力测定法。简而言之, 将Millipore MultiScreen PVDF板用35%乙醇渗透、洗涤、以及使用小鼠IFN- $\gamma$  ELISpot Ready-SET-Go! 试剂组 (eBioscience) 涂覆过夜。小鼠免疫接种后第0天, 从免疫接种的小鼠或幼稚小鼠收获脾细胞。将30x 10<sup>6</sup>个脾细胞/mL用FITC-缀合的抗-CD8 $\alpha$  (BD Biosciences) 在4 $^{\circ}$ C下染色30分钟, 洗涤, 并且使用FACSAria<sup>TM</sup>IIu或FACSAria<sup>TM</sup>IIIu装置分选 (BD Biosciences, 图13C)。铺板分选的CD8<sup>+</sup> T细胞, 并在37 $^{\circ}$ C下在辐照的脾细胞 (4000cGy) 存在的下孵育48小时, 这些脾细胞来自用相关肽 (对ELISpot测定法是4 $\mu$ M, 而对亲合力测定法是10<sup>-4</sup>至10<sup>-14</sup>M) 冲击的同系小鼠。作为阴性对照, 来自幼稚小鼠的CD8<sup>+</sup> T细胞用肽冲击的脾细胞孵育。使用试剂组制造商方案来显示斑点, 以及使用ImmunoSpot S5 UV Analyzer (Cellular Technology Ltd) 来枚举。将IFN- $\gamma$  产生表达为形成斑点单位的数目/10<sup>6</sup>个CD8<sup>+</sup> T细胞以及使用剂量-反应曲线来计算EC<sub>50</sub>。

[0248] 淋巴组织中的细胞分离和基于四聚体的富集方案脾脏和腹股沟、腋羽、腕板、颈和肠系膜淋巴结均从C57BL/6小鼠收获。将单细胞悬浮液用Fc block和10nM的PE-或APC标记的pMHC I四聚体 (NIH Tetramer Core Facility) 在4 $^{\circ}$ C下染色30分钟。在使用冰冷的分选缓冲液 (PBS与2%FBS) 洗涤之后, 将细胞重新悬浮在200 $\mu$ L的分选缓冲液中以及50 $\mu$ L的抗-PE和/或抗-APC抗体缀合的磁性微球 (Miltenyi Biotech), 随后将细胞在4 $^{\circ}$ C下孵育20分钟。如前所述<sup>15a, 16a</sup>, 细胞随后被洗涤以及四聚体<sup>+</sup>细胞被磁性富集。得到的四聚体富集的部分用以下染色: APC Fire 750-缀合的抗-B220、F4/80、CD19、CD11b、CD11c (BioLegend); PerCP-缀合的抗-CD4 (BioLegend); BV421-缀合的抗-CD3 (BD Biosciences); BB515-缀合的抗-CD8 (BD Biosciences); BV510-缀合的抗-CD44 (BD Biosciences) 抗体和Zombie NIR Fixable Viability Kit (BioLegend)。抗CD11b和CD11c在免疫后接种库的分析中不被考虑, 因为这些标记可能通过一些激活的T细胞来表达<sup>17a, 18a</sup>。然后在FACSCanto<sup>TM</sup>II细胞仪 (BD Biosciences) 上分析整个染色样品, 并且使用荧光计数珠子 (Thermo Fisher Scientific) 来归一化结果。作为阴性对照, 靶向以下3种源自病毒的抗原的抗原特异性CD8<sup>+</sup> T细胞库被富集: 来自淋巴细胞脉络丛脑膜炎病毒 (LCMV) 蛋白gp-33 (KAVYNFATC-SEQ ID NO:40; H-2D<sup>b</sup>) 的gp-33、来自鼠类巨细胞病毒蛋白M45 (HGIRNASFI-SEQ ID NO:41; H-2D<sup>b</sup>) 的M45以及来自牛痘病毒蛋白B8R (TSYKFESV-SEQ ID NO:42; H-2K<sup>b</sup>) 的B8R。

[0249] 数据. 关于本研究中使用的的所有样品的信息列举在表7中。在图1中使用的测序和

表达数据,已经存入NCBI's Sequen0e Read Archive和GEO中,其均可以在GEO的SuperSeries登录码GSE113992下访问,分别包含鼠类或人测序和表达数据的GSE111092和GSE113972登录码。SuperSeries记录可以经由<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113992>访问,通过在方框中输入通行证cnutscacjkbkzteb。在图1中使用的MS原始数据和相关的数据库,已经经由PRIDE<sup>19a</sup>伙伴储存库存入ProteomeXchange Consortium,其具有以下数据集合标识符:PXD009065和10.6019/PXD009065(CT26细胞系)、PXD009064和10.6019/PXD009064(EL4细胞系)、PXD009749和10.6019/PXD009749(07H103)、PXD009753和10.6019/PXD009753(10H080、弱酸洗脱)、PXD007935-测定#81756和10.6019/PXD007935(10H080、免疫沉淀反应)<sup>1a</sup>、PXD009750和10.6019/PXD009750(10H118)、PXD009751和10.6019/PXD009751(12H018)、PXD009752和10.6019/PXD009752(1c2)、PXD009754和10.6019/PXD009754(1c4)和PXD009755和10.6019/PXD009755(1c6)。

[0250] 实施例2:用于TSA发现的蛋白质基因组学方法的基本原理和设计。

[0251] 使用各种算法在计算上预测TSA的尝试充满了极高的假发现率<sup>27</sup>。因此,MAP库的系统水平的分子定义只有通过高通量MS研究才能实现<sup>3</sup>。目前的方法使用MS/MS软件工具,诸如Peaks<sup>28</sup>,其依赖于用户定义的蛋白质数据库来匹配每个采集的MS/MS谱和肽序列。由于参考蛋白质组不包含TSA,基于MS的TSA发现工作流程必须使用蛋白质基因组学策略来构建定制化的数据库,其源自肿瘤RNA测序(RNA-Seq)数据<sup>29</sup>,该数据库应当理想地包含所有在考虑的肿瘤样品中表达的蛋白质,甚至是未注释的蛋白质。由于目前的MS/MS软件工具无法处理全框翻译所有RNA-Seq读段<sup>30,31</sup>所创建的巨大搜索空间,富集癌症特异性序列的蛋白质基因组学策略被设计出来,以便全面地表征由所有基因组区编码的TSA的景观。得到的被称为全局癌症数据库的数据库由两个可定制化的部分组成。第一部分,被称为经典的癌症蛋白质组(图1A),其通过在它们的经典框中的编码表达蛋白质的转录物的计算机翻译来获得;其因此包含由正常或包含单碱基突变的外显子序列编码的蛋白质。第二部分,被称为癌症特异性蛋白质组(图1B),其是使用无比对RNA-Seq工作流程产生的,称为k-mer剖析,因为当前的映射器和变体读取器很难鉴定结构变体。这第二个数据集合使得能够检测任何基因组来源的任何阅读框编码的肽(包括结构变体),只要他们是癌症特异性的(即,正常细胞不存在的)。在此,选择使用mTEChi作为“正常对照”,因为它们表达大多数已知的基因,并诱导对其庞大的转录组编码的MAP的中枢耐受<sup>32</sup>。因此,为了鉴定癌症特异性的RNA序列,癌症RNA-Seq读段被切碎成33个核苷酸长的序列,称为k-mers<sup>33</sup>,其来自从同系mTEChi中移出的k-mer(图7A-B)。通过将重叠的癌症特异性k-mer组装成较长的序列(称为重叠群)来除去k-mer空间固有的冗余,其进一步经计算机3框翻译(图1B和图7C-D)。经典的癌症蛋白质组和癌症特异性蛋白质组随后被级联以为每一个分析样品创建一个全局癌症数据库。使用这样的优化的数据库,鉴定出从经MS测序的两种充分表征的小鼠肿瘤细胞系,即CT26(来自Ba1b/c小鼠的结肠癌)和EL4(来自C57BL/6小鼠的T-淋巴母细胞性淋巴瘤)洗脱出的MAP(图1C)。

[0252] 实施例3:非编码区是TSA的主要来源。

[0253] 在5%的假发现率下,鉴定了在CT26细胞上的1,875个MAP和EL4细胞的上783个MAP。在那些中,如果(i)它们的源自全癌症限制的33个核苷酸长的k-mer的33个核苷酸长的MAP编码序列(MCS)在mTEChi转录组中不存在或如果(ii)它们的源自癌症限制的33个核苷酸长的k-mer的截短版本的24-至-30核苷酸长的MCS在癌症的转录组中相比于mTEChi<sup>hi</sup>细胞过

表达至少10倍(图8A),则mTEC<sup>hi</sup>蛋白质组中不存在的MAP被视为TSA候选物。在MS相关的验证步骤和基因组定位的分配(图8B-C)之后,共获得6种mTSA候选物和15种aeTSA候选物:14种由CT26细胞呈递以及7种由EL4细胞呈递(图2A-B)。既突变也异常表达的MAP被包括在mTSA类别中。据信所有这些MAP是新颖的以及在Immune Epitope Database<sup>36</sup>是不存在的,除了这一个:AH1肽(SPSYVYHQF),在CT26细胞上使用反向免疫先前鉴定的唯一的150aeTSA<sup>9,37</sup>。

[0254] 为了评估基于从癌症k-mer中移出mTEC<sup>hi</sup> k-mer的数据库构建策略的严密性,评估了在在一组22个组织中编码aeTSA的MCS的外周表达<sup>38,39</sup>(表5)。15个aeTSA候选物中的四个具有与先前报道的“过表达”肿瘤相关抗原(TAA)<sup>40,41</sup>相似的表达谱,因为他们的MCS在大多数或所有组织中表达(图2C)。这四种肽因此被排除在TSA列表之外。与此形成对比,11种MAP被认为是真正的aeTSA,因为它们的MCS要么完全不存在,要么在一些组织中以痕量存在(图2C)。确实,低转录物水平的检测是无关紧要的,因为MAP优先源自高丰度的转录物<sup>42,43</sup>。该构思由AH1 TSA举例说明,AH1 TSA引发强烈的抗肿瘤反应但无不良反应<sup>9,37</sup>,即使它的MCS在肝脏、胸腺和膀胱中弱表达(图2C)。这些结果证明了mTEC<sup>hi</sup>中发现消减mRNA序列强烈地富集癌症限制性MCS。当整个鼠类TSA数据集合(6种mTSA和11种aeTSA)被考虑时,最显著的发现是它们大多数源自非典型翻译事件:编码外显子的框外翻译或非编码区的翻译(图2D)。此外,所有被鉴定的TSA,除了其中的两种都被典型的基于外显子组的方法忽略了,因为它们的源序列没有被注释为蛋白质编码。有趣的是,也注意到任何类型的非编码区可以产生TSA(表1):基因间和内含子间序列、非编码外显子、UTR/外显子连接、以及ERE,其似乎是TSA的特别丰富的来源(8种aeTSA和1种mTSA)。最终,鉴定了本文所述的有效捕获源自EL4细胞中非常大的基因间删除(7,500bp)的作为抗原的结构变体,VTPVYQHL(表1b)的方法。总之,这些观察证实非编码区是TSA的主要源,并且他们有潜力来大幅扩展肿瘤的TSA景观。

[0255] 对似乎最有吸引力的TSA中的一些进行进一步的研究,即,由EL4细胞呈递的那些,其MCS不被任何正常组织表达(表2C和表1b)。为了评估他们的免疫原性,在使用活EL4细胞攻击前,将C57BL/6小鼠分别用未冲击的(对照组)或TSA冲击的DC免疫接种2次。针对IILEFHSL或TVPLNHNTL的致敏(priming)可以延长10%的小鼠生存期,只有TVPLNHNTL免疫接种的小鼠存活最长达150天(图3A)。其他三种TSA示出了第150天的存活率分别为20%(VNYIHRNV)、30%(VTPVYQHL)和100%(VNYLHRNV)(图3B,C)。为了评估TSA疫苗接种的长期功效,存活的小鼠在第150天再次接受活EL4的攻击,监测疾病迹象。两个VNYIHRNV免疫接种的幸存者在50天内死于白血病,然而所有其他(针对TVPLNHNTL、VTPVYQHL或VNYLHRNV免疫接种的)在再次攻击个中幸存(图3)。因此,可以推断针对单独的TSA的免疫接种提供了不同程度的对抗EL4细胞的保护,而且在大多数情况下,这种保护是长效的。

[0256] 实施例4:在幼稚和免疫接种的小鼠中的TSA特异性T细胞的频率。

[0257] 在各种模型中,在体内的免疫应答的强度是通过抗原反应性T细胞的数目的调节的<sup>44,45</sup>。因此,使用基于四聚体的富集方案评估在幼稚和免疫的小鼠的TSA特异性T细胞的频率<sup>46,47</sup>,圈选策略和一个有代表性的实验可以在图9A-C中找到。作为阳性对照,使用特异于三种病毒表位(gp-33、M45B8R)的高丰度的CD8 T细胞,并且经证实它们的频率在先前的研究中观察到的范围内<sup>45</sup>(图4A)。在幼稚小鼠中,特异于TVPLNHNTL、VTPVYQHL和IILEFHSL的CD8T细胞是罕见的(少于1个四聚体+细胞/10<sup>6</sup>个CD8 T细胞),而特异于ERE TSA(VNYIHRNV和VNYLHRNV)的CD8 T细胞显示的频率与病毒对照组的那些相似(图4A和图10A)。因此,在用

TSA冲击的DC免疫接种的小鼠中,发现如四聚体染色或IFN- $\gamma$  ELISpot测定所评估的,针对两个ERE TSA的T细胞频率(图9C-D和10A),显著高于针对TVPLNHNTL、VTPVYQHL和IILEFHSL的频率(图4B-C)。而且,在幼稚和免疫接种的小鼠中,发现抗原特异性T细胞的频率高度相关(图11A-C)。最终,据估计特异于VNYIHRNV和VNYLHRNV的T细胞的功能亲合力与特异于两个高免疫原性的非自身抗原(次要组织相容性抗原H7a和H13a)的T细胞的功能亲合力相似(图4D)。因此,源自所谓非编码区的这些TSA是通过高丰度的具有高功能亲合力的T细胞来鉴定的。这对于VNYIHRNV aeTSA是特别需要注意的,因为其具有未突变的种系序列。

[0258] 合在一起,这些结果示出了TSA特异性T细胞的频率总体上是TSA免疫原性的显著参数。然而,VTPVYQHL对EL4攻击提供了第二好的保护,即使其同源T细胞以非常低的频率出现(图3和4A-C)。为了更好地评估T细胞扩增在白血病保护中的重要性,评估了在第150天用EL4细胞再次攻击后,长期的幸存者中的四聚体<sup>+</sup>CD8<sup>+</sup> T细胞的频率(图3)。这些分析在第210天或在处死时(在VNYIHRNV致敏的小鼠的情况下)进行。所有长期的幸存者,包括VTPVYQHL免疫的小鼠示出了TSA特异性(四聚体<sup>+</sup>)CD8<sup>+</sup>T细胞的明显的群(图10B-C)。尽管VNYIHRNV由四聚体<sup>+</sup>细胞大的群所识别,但在本文使用的实验条件下,这不足以在再次攻击时保护小鼠。

[0259] 实施例5:抗原表达对对抗EL4细胞的保护的重要性。

[0260] 接下来,通过评估第0天注射的EL4细胞群中的RNA水平上的TSA的丰度来评估抗原表达对免疫原性的影响(图3)。发现编码提供最好的对抗EL4细胞的保护的TSA的序列(VNYLHRNV)比其他TSA以高得多的水平表达(图5A)。这表明VNYLHRNV可能是“克隆的”(由所有EL4细胞表达)以及高表达的,而其他TSA是亚克隆的和/或以低水平表达。接下来,使用平行反应监测(PRM)MS来分析用于再次攻击(第150天,图5B)的EL4细胞群上的每个细胞的TSA拷贝数。在RNA和肽水平上的TSA丰度之间没有线性关系<sup>40</sup>(图5A-B)。显著地,最好的TSA,VNYLHRNV是两个丰度最大的TSA(>500拷贝/细胞)的其中之一,而VNYIHRNV,在本文使用的实验条件下再次攻击时未提供显著保护(图3B),在EL4细胞上其不再被检测到。这个观察表明VNYIHRNV是亚克隆的TSA,并且抗原丢失最可能解释再次攻击时显著保护的缺乏。最后,指出当通过DC呈递时TSA是免疫原性,而通过EL4细胞呈递时则不是:i)没有预先免疫的情况下,活EL4细胞的注射不会诱导TSA特异性T细胞的显著扩增,以及ii)辐照的EL4细胞的免疫接种不会提供对抗活EL4细胞的显著保护(图5C-D和图10D)。这表明在不存在免疫的情况下,高免疫原性TSA(诸如VNYLHRNV)被忽视了,因为它们没有被DC有效地交叉呈递,这突出了有效的T细胞致敏在癌症免疫治疗中的重要性。

[0261] 实施例6:非编码区扩展人原发性肿瘤TSA的景观。

[0262] 已经论证了非编码区是TSA在两种鼠类细胞系中的主要来源,本文所述的蛋白质基因组学方法被应用至七种人原发性肿瘤样品:四种B系ALL和三种肺癌。为此,不使用来自鼠类同系mTEC<sup>hi</sup>的RNA-Seq数据,来自六个不相关的接受矫正心血管手术的供体的总TEC(n=2)和纯化的mTEC(n=4)的转录组被测序。显著地,发现了最小的个体间的差异,并且这个队列的大小被示出足以几乎涵盖mTEC转录组景观的全部宽度(图12A-B)。使用这些RNA-Seq数据作为图1所述的工作流程的正常k-mer的库,3种mTSA和27种aeTSA候选物被鉴定出来(图6A)。除了被广泛验证,还确保了mTSA不与已知的种系多态性相交。为了进一步验证aeTSA候选物的状态,分析了在来自28种组织(6-50个个体/组织,图6B和表6)的RNA-Seq数

据的aeTSA MCS的表达,与对鼠类aeTSA所做的类似(图2C)。基于这些数据,六种aeTSA候选物被排除在外:i)三种是广泛表达的,类似于大多数先前报道过的过度表达的TAA<sup>48</sup>,以及ii)三种在单个器官(肝脏)中以显著水平表达(图6B)。因此,总共三种mTSA和20种非冗余的aeTSA候选物被鉴定出(图6C和表2a-d和3a-c)。值得注意的是,SLTALVFHV aeTSA被两种HLA-A\*02:01-阳性ALL共享(表2a和2d)。这种aeTSA源自TCL1A的3' UTR,一种涉及淋巴恶性肿瘤的基因。总之,结果示出了本文所述的蛋白质基因组学方法可以在大约两周内表征在各个肿瘤上mTSA和aeTSA的库。

[0263] 表

[0264] 表1a:CT26 TSA

TSA 序列 (SEQ ID NO)	基因组位置	Ensembl 转录物 id	TSA 起始 框	MHC I 类分 子	百分等 级
GYQKMKALL (SEQ ID NO:1)	chr8:123429315-123429 341	MuLV	ERE	H-2-K <sup>d</sup>	0,06
KPLK/EAPLD L (SEQ ID NO:2)	chr1:173783238-173783 264	ENSMUST00000155 076	内含子	H-2-L <sup>d</sup>	0,01
KYLSVQS/GQ L (SEQ ID NO:3)	chr17:29332770-293327 78   chr17:29333514-293335 31	ENSMUST00000095 427	编码外显 子 框内	H-2-K <sup>d</sup>	0,01
KYLSVQS/GQ LF (SEQ ID NO:4)	chr17:29332767-293327 78   chr17:29333514-293335 31	ENSMUST00000095 427	编码外显 子 框内	H-2-K <sup>d</sup>	0,25
[0265] LPQELPGLVV L (SEQ ID NO:5)	chr8:123427101-123427 133	MuLV	ERE	H-2-L <sup>d</sup>	0,5
MPHSLLPLVT F (SEQ ID NO:6)	chr7:89664573-8966460 5	ENSMUST00000159 167	内含子	H-2-L <sup>d</sup>	0,2
QGPMALRI/LF (SEQ ID NO:7)	chr9:66126885-6612691 1	ENSMUST00000034 945	编码外显 子 框外	H-2-D <sup>d</sup>	0,05
SGPPYYKGI (SEQ ID NO:8)	chr8:121839803-121839 829	MMERGLN_I 或 ENSMUST00000127 664	ERE 或内 含子	H-2-D <sup>d</sup>	0,25
SPHQVFNL (SEQ ID NO:9)	chr8:123428239-123428 262	MuLV	ERE	H-2-L <sup>d</sup>	0,01
[0266] SPSYVYHQF (SEQ ID NO:10)	chr8:123426985-123427 011	MuLV	ERE	H-2-L <sup>d</sup>	0,5

[0267] 表1b:EL4 TSA

TSA 序列 (SEQ ID NO)	基因组位置	Ensembl 转录物 id	TSA 起始框	MHC I 类分子	百分等级
ATQQFQQL (SEQ ID NO:11)	chr8:123426867-123426844	MuLV	ERE	H-2-K <sup>b</sup>	0,2
IILEFHSL (SEQ ID NO:12)	chr10:116678525-116678548	ENSMUST00000181656	非编码外显子	H-2-K <sup>b</sup>	0,02
SSPRGSSTL (SEQ ID NO:13)	chr6:114732754-114732780	B3A 或 ENSMUST00000032457	ERE 或内含子	H-2-D <sup>b</sup>	0,3
TVPLNHNT L (SEQ ID NO:14)	chr4:83615597-83615624	ENSMUST00000053414	新的反义	H-2-D <sup>b</sup>	0,12
VNYI/LHRN V (SEQ ID NO:15)	chr4:46583174-46583197	MMTV	ERE	H-2-K <sup>b</sup>	0,01
VNYI/LHRN V (SEQ ID NO:15)	chr4:46583174-46583197	MMTV	ERE	H-2-K <sup>b</sup>	0,01
VTPVYQ HL (SEQ ID NO:16)	chr2:75078751-75078756   chr2:75086270-75086287	N/A	基因间	H-2-K <sup>b</sup>	0,01

[0268] 表2a:B-ALL标本检测的人TSA-07H103 TSA的特征

TSA 序列 (SEQ ID NO)	基因组位置	Ensembl 转录物 id	TSA 起始框	MHC I 类分子	百分等级
KILLLQSL (SEQ ID NO:17)	chr5:132450600-132450626	L1ME3G 或 ENST00000407797	ERE 或内含子	A*02:01	1,8
KISLYLPAL (SEQ ID NO:18)	chr8:144861684-144861710	LTR46-int	ERE	A*02:01	0,5
SLTALVFH V (SEQ ID NO:19)	chr14:95710533-95710559	ENST00000554012	3' UTR	A*02:01	0,06

[0270] 表2b:B-ALL标本检测的人TSA-10H080 TSA的特征

TSA 序列 (SEQ ID NO)	基因组位置	Ensembl 转录物 id	TSA 起始框	MHC I 类分子	百分等级
HETLRLLL (SEQ ID NO:20)	chr6:106197722-106197745	ENST00000369076	内含子	B*40:01	1,2
RIFGFRLWK (SEQ ID NO:21)	chr1:80641339-80641365	ENST00000418041	内含子	A*11:01	0,01
TSAETWM K (SEQ ID NO:22)	chr7:43947484-43947510	L1PA6 或 ENST00000427076	内含子或 ERE	A*11:01	0,01
TSIPKPNLK (SEQ ID NO:23)	chr2:237428272-237428298	N/A	基因间	A*11:01	0,15

[0272] 表2c:B-ALL标本检测的人TSA-10H118 TSA的特征

TSA 序列 (SEQ ID NO)	基因组位置	Ensembl 转录物 id	TSA 起始框	MHC I 类分子	百分等级
[0274] LPFEQKSL (SEQ ID NO:24)	chr2:47522843-47522866	ENST0000032787 6	内含子	B*08:01	0,7
SLREKGFSL (SEQ ID NO:25)	chr1:175955400-175955426	ENST0000036766 7	内含子	B*08:01	0,09
VPAALRSL (SEQ ID NO:26)	chr7:106886341-106886364	ENST0000035919 5	内含子	B*07:02	0,3

[0275] 表2d: B-ALL标本检测的人TSA-12H018 TSA的特征

TSA 序列 (SEQ ID NO)	基因组位置	Ensembl 转录物 id	TSA 起始框	MHC I 类分子	百分等级
[0276] LLAATILLSV (SEQ ID NO:27)	chr2:174631426-174631455	ENST0000039254 7	内含子	A*02:01	0,2
SLFVA/VSLSL (SEQ ID NO:28)	chr6:106971679-106971705	ENST0000060601 7	编码外显子 框内	A*02:01	0,6
SLTALVFHV (SEQ ID NO:19)	chr14:95710533-95710559	ENST0000040239 9	3' UTR	A*02:01	0,06

[0277] 表3a: 在肺肿瘤活检中检测的人TSA-1c2 TSA的特征

TSA 序列 (SEQ ID NO)	基因组位置	Ensembl 转录物 id	TSA 起始框	MHC I 类分子	百分等级
[0278] IIAPPPPK (SEQ ID NO:29)	chr14:21098919-21098945	ENST0000042109 3	5' UTR	A*11:01	0,15
LVFNILHR (SEQ ID NO:30)	chr6:6800963-6800989	N/A	基因间	A*11:01	0,25
MISPVLALK (SEQ ID NO:31)	chr19:41751004-41751030	ENST0000059574 0	5'UTR	A*11:01	0,03
SLSYLILKK (SEQ ID NO:32)	chrX:107212979-10721305	ENST0000037245 3	编码外显子 框外	A*11:01	0,05
SSASQLPSK (SEQ ID NO:33)	chr16:19430493-19430519	L4_B_Mam 或 ENST0000054258 3	ERE 或 5'UTR	A*11:01	0,07
SVIQTGHLAK (SEQ ID NO:34)	chr3:169840282-169840311	ENST0000031642 8	编码外显子 框内	A*11:01	0,1
TTLKYLWKK (SEQ ID NO:35)	chr3:169381477-169381503	ENST0000048595 7	5'UTR	A*11:01	0,03

[0279] 表3b: 在肺肿瘤活检中检测的人TSA-1c4 TSA的特征

TSA 序列 (SEQ ID NO)	基因组位置	Ensembl 转录物 id	TSA 起始框	MHC I 类分子	百分等级
[0280] KPSVFPLSL (SEQ ID NO:36)	chr14:37589683-37589708	N/A	基因间	B*07:02	0,15

[0281] 表3c:在肺肿瘤活检中检测的人TSA-1c6 TSA的特征

TSA 序列 (SEQ ID NO)	基因组位置	Ensembl id	转录物	TSA 起始 框	MHC I 类分 子	百分等 级
QR/KF/LQGRVT M (SEQ ID NO:37)	chr15:19972868-199728 94	N/A		基因间	C*07:01	0,02
SRFSGVPDRF (SEQ ID NO:38)	chr2:89234284-8923431 3	N/A		基因间	A*24:02	0,9
TYTQN/DFNKF (SEQ ID NO:39)	chr11:14968916-149689 42	ENST0000033158 7		编码外显 子 框内	A*24:02	0,03

[0283] 表4a:与全局癌症数据库的生成相关的统计数据-鼠样品

		EL4	mTEChi_C57BL/6	CT26	mTEChi_Balb/c	
经典的蛋 白质	转录物	表达的(tpm > 0)	64318	86947	65242	82420
		蛋白质编码	34171	47086	35104	44943
蛋白质 组	蛋白质	不同的	35280	50304	37810	54456
	读段	总	240372644	456991966	247522370	455625158
癌症特 异性蛋 白质 组	k-mers (k = 33nts)	总	14862978110	28980506746	15026027458	21482018335
		不同的	429163639	1084732266	507092097	1115569754
	重叠群	计数 ≥ 4	116852296		104699335	
		癌症特异性	19091379		22892864	
蛋白质	不同的	895313		1845144		
	≥ 34nts	715161		1377631		
蛋白质	不同的, ≥ 8aa	2153996		3701717		

[0285] 表4b:与全局癌症数据库的生成相关的统计数据-人样品

			<b>07H103</b>	<b>10H080</b>	<b>10H118</b>
经典的蛋白质组	转录物	表达的(tpm > 0)	107590	115494	116981
		蛋白质编码	57931	62280	63133
	蛋白质	不同的	59082	64150	63921
癌症特异性蛋白质组	读段	总	105 863 640	129 444 492	226 508 070
		总	6 739 820 561	8 297 285 105	13 804 699 469
	k-mers (k = 33nts)	不同的	633 011 468	761 444 095	1 119 514 550
		计数 ≥ 7 或 10	77 745 744	98 652 247	135 682 880
		癌症特异性	11 694 475	20 210 820	32 070 840
	重叠群	不同的	492 273	778 594	1 412 680
		≥ 34nts	440 367	697 184	1 246 048
蛋白质	不同的, ≥8aa	1 326 854	2 156 187	3 708 759	

[0286]

			<b>12H018</b>	<b>lc2</b>	<b>lc4</b>
经典的蛋白质组	转录物	表达的(tpm > 0)	113438	116600	117476
		蛋白质编码	61481	66874	67549
	蛋白质	不同的	63767	70493	71734
癌症特异性蛋白质组	读段	总	161 724 658	268 396 930	262 531 548
		总	9 981 973 250	17 197 030 205	17 341 587 177
	k-mers (k = 33nts)	不同的	868 719 740	669 751 679	727 571 721
		计数 ≥ 7 或 10	96 193 003	78 611 668	81 410 185
		癌症特异性	17 879 385	9 003 814	9 918 787
	重叠群	不同的	758 491	669 145	749 712
		≥ 34nts	666 164	513 928	581 510
蛋白质	不同的, ≥ 8aa	2 014 334	1 401 735	1 554 082	

			lc6	102015	062015
经典的蛋白质组	转录物	表达的(tpm > 0)	119870	62976	85686
		蛋白质编码	67135	37073	49155
	蛋白质	不同的	71526	46181	67497
癌症特异性蛋白质组	读段	总	246 868 078	134 624 214	136 558 238
		总	16 284 413 566 a		a
	k-mers (k = 33nts)	不同的	864 050 270		
		计数 ≥ 7 或 10	97 121 823		
		癌症特异性	17 663 050		
	重叠群	不同的	1 113 278		
		≥ 34nts	886 470		
蛋白质	不同的, ≥ 8aa	2 431 066			

[0287]

			S5	S9	S10	S11
经典的蛋白质组	转录物	表达的(tpm > 0)	95090	118246	112739	119225
		蛋白质编码	55103	66223	63113	66695
	蛋白质	不同的	70276	79469	75384	80996
癌症特异性蛋白质组	读段	总	200363532	229281098	231185678	251770122
		总	a	a	a	a
	k-mers (k = 33nts)	不同的	b	b	b	b
		计数 ≥ 7 或 10				
		癌症特异性				
	重叠群	不同的				
≥ 34nts						
蛋白质	不同的, ≥8aa					

[0288] 表5:本研究使用的ENCODE数据集合的登录号

组织	登录号(SRA)
脂肪组织	SRR5171088, SRR5171089
肾上腺	SRR5171111, SRR5171112, SRR5047957, SRR5047958, SRR5047959, SRR5047960, SRR5047961, SRR5047962
脑	SRR5171101, SRR5171102
结肠	SRR5047913, SRR5047914, SRR5047915, SRR5047916, SRR5047917, SRR5047918
十二指肠	SRR5047963, SRR5047964, SRR5047965, SRR5047966, SRR5047967, SRR5047968, SRR5047969
性腺脂垫 (Gonadal fat pad)	SRR5047970, SRR5047971, SRR5047972, SRR5047973
心脏	SRR5171076, SRR5171077, SRR5047921, SRR5047922, SRR5047923, SRR5047924
肾脏	SRR5047925, SRR5047926, SRR5047927, SRR5047928, SRR5047929, SRR5047930, SRR5171094, SRR5171095
大肠	SRR5047975, SRR5047976, SRR5047977, SRR5047978
肝脏	SRR3192469, SRR3192470, SRR5171078, SRR5171079, SRR5047931,

[0289]

	SRR5047932, SRR5047933, SRR5047934, SRR5047935, SRR5047936
肺	SRR5171113, SRR5171114, SRR5047937, SRR5047938, SRR5047939, SRR5047940
乳腺	SRR5047979, SRR5047980, SRR5047981, SRR5047982, SRR5047983, SRR5047984
卵巢	SRR5047985, SRR5047986, SRR5047987, SRR5047988, SRR5047989, SRR5047990, SRR5047991, SRR5047992, SRR5047993, SRR5047994, SRR5171100
胰腺	SRR5171086, SRR5171087
乙状结肠	SRR5171098, SRR5171099
[0290] 小肠	SRR5048001, SRR5048002, SRR5048003, SRR5048004, SRR5048005, SRR5048006, SRR5048007, SRR5048008, SRR5048009, SRR5048010, SRR171080, SRR5171081
脾	SRR5047941, SRR5047942, SRR5047943, SRR5047944, SRR5047945, SRR5047946, SRR5171241, SRR5171242
胃	SRR5047997, SRR5047996, SRR5047995, SRR5047998, SRR5048000, SRR5047999
皮下脂肪组织	SRR5048011, SRR5048012, SRR5048013, SRR5048014
睾丸	SRR5047953, SRR5047954, SRR5047955, SRR5047956, SRR5171085, SRR5171084
胸腺	SRR5047947, SRR5047948, SRR5047949, SRR5047950, SRR5047951, SRR5047952
膀胱	SRR5048035, SRR5048036

[0291] 表6:本研究使用的GTEX数据集合的登录号

组织	随机选择的供体的登录号(SRA)
脂肪 - 皮下	SRR599313 SRR608150 SRR608198 SRR612263 SRR612707 SRR612815 SRR612863 SRR612935 SRR613150 SRR613234 SRR613342 SRR613390 SRR613533 SRR613550 SRR613639 SRR613675 SRR613855 SRR613896 SRR613915 SRR613927 SRR614119 SRR614191 SRR614395 SRR614419 SRR614864 SRR615069 SRR615237 SRR615431 SRR615659 SRR615778 SRR615874 SRR615946 SRR617841 SRR654730 SRR654862 SRR654898 SRR655182 SRR655531 SRR655637 SRR655768 SRR655816 SRR656059 SRR656946 SRR657599 SRR657949 SRR658081 SRR658754 SRR658941 SRR658953 SRR659109
肾上腺	SRR1069421 SRR1070913 SRR1072626 SRR1073365 SRR1073775 SRR1074474 SRR1075314 SRR1076632 SRR1076823 SRR1082035 SRR1082616 SRR1082733 SRR1083824 SRR1083892 SRR1085590 SRR1085951 SRR1086046 SRR1087297 SRR1087511 SRR1087606 SRR1088365 SRR1088461 SRR1089479 SRR1089950 SRR1091476 SRR1092160 SRR1092329 SRR1092686 SRR1093625 SRR1093721 SRR1093954 SRR1094144 SRR1099378 SRR1099427 SRR1099598 SRR1099694 SRR1100496 SRR1100728 SRR808862 SRR809873 SRR810129 SRR810713 SRR811237 SRR811631 SRR812246 SRR814407 SRR816495 SRR816865 SRR817649 SRR818694
[0292] 动脉 - 主动脉	SRR1069376 SRR1070111 SRR1070641 SRR1071644 SRR1072078 SRR1072749 SRR1073705 SRR1074478 SRR1074622 SRR1075028 SRR1075579 SRR1076343 SRR1077090 SRR1078586 SRR1079023 SRR1079998 SRR1080148 SRR1081137 SRR1081519 SRR1081910 SRR1082283 SRR1083076 SRR1083286 SRR1083604 SRR1084276 SRR1084460 SRR1085159 SRR654850 SRR808044 SRR808152 SRR808351 SRR808836 SRR808914 SRR809320 SRR809470 SRR809785 SRR809831 SRR810201 SRR810367 SRR811333 SRR811471 SRR811819 SRR812673 SRR813632 SRR815092 SRR816565 SRR817744 SRR818232 SRR818999 SRR819293
膀胱	SRR1071717 SRR1079830 SRR1081765 SRR1085402 SRR1086236 SRR1092208 SRR1093930 SRR1097296 SRR1099957 SRR1120296 SRR2135324 SRR2135407
大脑 - 皮层	SRR1081741 SRR1082262 SRR1083632 SRR1085975 SRR1310008 SRR1310136 SRR1311400 SRR1311575 SRR1311794 SRR1312428 SRR1314958 SRR1315269 SRR1315866 SRR1316815 SRR1320280 SRR1323043 SRR1323746 SRR1324371 SRR1327593 SRR1328487 SRR598332 SRR601006 SRR601669 SRR602927 SRR603333 SRR604026 SRR608662 SRR612575 SRR614310 SRR615213 SRR615838 SRR627421 SRR627425 SRR627449 SRR627455 SRR654874 SRR656745 SRR659555 SRR660626 SRR660933 SRR663320 SRR663753 SRR664854 SRR808614 SRR810319 SRR810877 SRR812012 SRR812436 SRR816770 SRR820078
乳房 - 乳	SRR1068977 SRR1068999 SRR1070208 SRR1070260 SRR1070738 SRR1071084

腺组织	SRR1071905 SRR1074860 SRR1075484 SRR1076219 SRR1076441 SRR1077139 SRR1077920 SRR1078258 SRR1079948 SRR1081023 SRR1082859 SRR1083052 SRR1083959 SRR1084079 SRR1084674 SRR1086538 SRR1086772 SRR615910 SRR655447 SRR655852 SRR656911 SRR656970 SRR657018 SRR657528 SRR658105 SRR658319 SRR658409 SRR659223 SRR660248 SRR660283 SRR662306 SRR662378 SRR662811 SRR808428 SRR808942 SRR811073 SRR811285 SRR812198 SRR813868 SRR815208 SRR816336 SRR818873 SRR820571 SRR821498
宫颈 - 外 子宫颈	SRR1075223 SRR1088832 SRR1089562 SRR1096876 SRR1097035 SRR1097574
结肠 - 乙 状结肠	SRR1069943 SRR1074337 SRR1077380 SRR1081068 SRR1083504 SRR1083678 SRR1084505 SRR1086020 SRR1087271 SRR1090431 SRR1091524 SRR1092493 SRR1093366 SRR1102198 SRR1102224 SRR1102998 SRR1308269 SRR1312577 SRR1312666 SRR1312784 SRR1317110 SRR1317653 SRR1318624 SRR1319038 SRR1320445 SRR1320490 SRR1321377 SRR1322070 SRR1323002 SRR1323215 SRR1324473 SRR1327454 SRR1327505 SRR1327527 SRR1327570 SRR1328528 SRR1328980 SRR1329642 SRR1329663 SRR1330176 SRR1330770 SRR1330831 SRR1332467 SRR1333167 SRR1333287 SRR1334011 SRR1334055 SRR1334181 SRR1336617 SRR1336863
食管 - 黏 膜	SRR1069231 SRR1069255 SRR1069328 SRR1069666 SRR1069871 SRR1070036 SRR1070060 SRR1070620 SRR1070665 SRR1071207 SRR1071499 SRR1072055 SRR1072297 SRR1072388 SRR1072480 SRR1073631 SRR1074450 SRR1074502 SRR1074578 SRR1075458 SRR1075603 SRR1076195 SRR1076705 SRR1076801 SRR1077310 SRR1077356 SRR1077619 SRR1077850 SRR1078140 SRR1078538 SRR807679 SRR807703 SRR809406 SRR809919 SRR812294 SRR812318 SRR813283 SRR813505 SRR813536 SRR814467 SRR815116 SRR815568 SRR816403 SRR817306 SRR819124 SRR819559 SRR819637 SRR820280 SRR820689 SRR821282
输卵管	SRR1071359 SRR1074140 SRR1076584 SRR1082520 SRR1083776 SRR1101693 SRR811938
心脏 - 左 心室	SRR598148 SRR598509 SRR598589 SRR599025 SRR599086 SRR599249 SRR599380 SRR600474 SRR600829 SRR600852 SRR600924 SRR601239 SRR601613 SRR601645 SRR601868 SRR601986 SRR602106 SRR602437 SRR602461 SRR603449 SRR603918 SRR603968 SRR604122 SRR604174 SRR604206 SRR604230 SRR606939 SRR607252 SRR607313 SRR607970 SRR608096 SRR608480 SRR612335 SRR612719 SRR612875 SRR613186 SRR613462 SRR613510 SRR613759 SRR614215 SRR614683 SRR614996 SRR615335 SRR615359 SRR615898 SRR615970 SRR655792 SRR657903 SRR658283 SRR658331
肾脏 - 皮 层	SRR1071807 SRR1080366 SRR1085759 SRR1089504 SRR1105272 SRR1314940 SRR1317086 SRR1325483 SRR1328447 SRR1329154 SRR1340662 SRR1362263 SRR1377578 SRR1380931 SRR1396700 SRR1416516 SRR1420649 SRR1432650 SRR1433066 SRR1435730 SRR1437274 SRR1442708 SRR1443092 SRR1445835 SRR1447631 SRR1452888 SRR1456711 SRR1465871 SRR1468426 SRR1469746 SRR1486080 SRR1490658 SRR1500261 SRR2135353 SRR2135396 SRR809943 SRR810007 SRR821356
肝脏	SRR1069141 SRR1070689 SRR1071668 SRR1073435 SRR1075102 SRR1075804 SRR1076022 SRR1080117 SRR1080294 SRR1081184 SRR1082151 SRR1083983 SRR1086256 SRR1087007 SRR1087321 SRR1089446 SRR1090095 SRR1090556 SRR1091865 SRR1093861 SRR1095383 SRR1095913 SRR1098737 SRR1100991 SRR1101883 SRR1102152 SRR1102899 SRR1105248 SRR1120939 SRR1310433 SRR1312266 SRR1313807 SRR1316096 SRR1317532 SRR1317554 SRR1321877 SRR1322312 SRR1322477 SRR1323491 SRR1324295 SRR1324412 SRR1325290 SRR1328760 SRR1331488 SRR1334866 SRR1335236 SRR1336314 SRR815140 SRR815711 SRR821043
肺	SRR1070015 SRR1070358 SRR1071568 SRR1072150 SRR1073119 SRR1074769 SRR1081283 SRR1084602 SRR1084766 SRR1086728 SRR1087559 SRR1091670 SRR1095695 SRR1098074 SRR1098785 SRR1098998 SRR1099286 SRR1099546 SRR1102079 SRR1102804 SRR1307123 SRR1307615 SRR1308239 SRR1308504 SRR1308939 SRR1309452 SRR1309468 SRR1309490 SRR1310313 SRR1310520

[0293]

	SRR1310797 SRR1310959 SRR1310975 SRR1312209 SRR1312522 SRR1312558 SRR813043 SRR814244 SRR814703 SRR817004 SRR817070 SRR817166 SRR817488 SRR818499 SRR819186 SRR819318 SRR819658 SRR820596 SRR821302 SRR821525
小 唾 液 腺	SRR1071105 SRR1078392 SRR1080790 SRR1081589 SRR1097245 SRR1100608 SRR1315412 SRR1318089 SRR1321897 SRR1325201 SRR1328715 SRR1330723 SRR1331771 SRR1338384 SRR1339987 SRR1340260 SRR1348929 SRR1353600 SRR1356057 SRR1358391 SRR1376380 SRR1376450 SRR1376741 SRR1381185 SRR1382978 SRR1385690 SRR1386927 SRR1388459 SRR1389955 SRR1397720 SRR1400931 SRR1404339 SRR1405147 SRR1406135 SRR1406348 SRR1407044 SRR1413307 SRR1416141 SRR1416188 SRR1416841 SRR1418225 SRR1418473 SRR1418747 SRR1419561 SRR1429429 SRR1429540 SRR1431823 SRR1432868 SRR1432958 SRR1433493
肌 肉 - 骨 骼 的	SRR1068855 SRR1071231 SRR1071594 SRR1071955 SRR1074359 SRR1074670 SRR1074719 SRR1077288 SRR1077805 SRR1080766 SRR1084369 SRR1084417 SRR1085519 SRR1087245 SRR1087825 SRR1088581 SRR1089424 SRR1089901 SRR1090265 SRR1092349 SRR1092985 SRR1094051 SRR1095720 SRR1096174 SRR1096662 SRR1098474 SRR1098879 SRR1100588 SRR1102830 SRR1105057 SRR812773 SRR813656 SRR813802 SRR813983 SRR815020 SRR815044 SRR815470 SRR815783 SRR815825 SRR816015 SRR816226 SRR816382 SRR817282 SRR817421 SRR818600 SRR818773 SRR818901 SRR819054 SRR819261 SRR820907
神 经 - 胫 的	SRR1070086 SRR1070159 SRR1070597 SRR1072724 SRR1073553 SRR1074550 SRR1075384 SRR1075825 SRR1076559 SRR1079636 SRR1079850 SRR1080093 SRR1082059 SRR1082809 SRR1086417 SRR1087079 SRR1088706 SRR1090070 SRR1091184 SRR1092062 SRR1095334 SRR1096007 SRR1096222 SRR1096478 SRR1096500 SRR1096806 SRR1097055 SRR1098385 SRR1310455 SRR1310645 SRR1311131 SRR1311308 SRR1312370 SRR1312464 SRR813704 SRR814052 SRR814996 SRR815422 SRR815685 SRR817026 SRR817397 SRR817539 SRR817609 SRR818939 SRR818961 SRR820350 SRR820402 SRR821096 SRR821124 SRR821255
卵 巢	SRR1071475 SRR1073389 SRR1073878 SRR1075360 SRR1078042 SRR1078636 SRR1078735 SRR1081987 SRR1082352 SRR1082471 SRR1085565 SRR1085736 SRR1086212 SRR1086656 SRR1088856 SRR1089134 SRR1090698 SRR1090928 SRR1091164 SRR1092038 SRR1093601 SRR1093747 SRR1096458 SRR1097124 SRR1097148 SRR1098807 SRR1099310 SRR1099669 SRR1101453 SRR1101859 SRR1102005 SRR1102780 SRR1120276 SRR1312446 SRR1315495 SRR1316513 SRR1319793 SRR1336244 SRR1339699 SRR1340598 SRR1341583 SRR1342849 SRR1347518 SRR1350891 SRR1351641 SRR1353537 SRR814293 SRR814892 SRR816629 SRR821072
胰 腺	SRR1069352 SRR1070403 SRR1070764 SRR1071519 SRR1072007 SRR1072104 SRR1072972 SRR1073021 SRR1073167 SRR1073991 SRR1074090 SRR1074385 SRR1075174 SRR1075336 SRR1076244 SRR1076868 SRR1078066 SRR1079754 SRR1080624 SRR1082080 SRR1082544 SRR1084128 SRR1084323 SRR1085187 SRR1085310 SRR1086070 SRR1087728 SRR1088291 SRR1088413 SRR1088537 SRR1089537 SRR1089688 SRR1091032 SRR1091144 SRR1092937 SRR1093340 SRR1093434 SRR1093577 SRR1095407 SRR1095479 SRR1095651 SRR1097777 SRR1097883 SRR812745 SRR813208 SRR816541 SRR819771 SRR821050 SRR821231 SRR821666
垂 体	SRR1076393 SRR1077455 SRR1077708 SRR1077968 SRR1082664 SRR1082685 SRR1089785 SRR1096101 SRR1096339 SRR1101612 SRR1309119 SRR1309638 SRR1310817 SRR1311599 SRR1311709 SRR1311958 SRR1317963 SRR1318026 SRR1319946 SRR1321650 SRR1323977 SRR1324141 SRR1324184 SRR1325161 SRR1325944 SRR1326408 SRR1326797 SRR1328143 SRR1331962 SRR1332024 SRR1332904 SRR1336029 SRR1336529 SRR1337321 SRR1339007 SRR1340241 SRR1343012 SRR1343221 SRR1343720 SRR1343778 SRR1345329 SRR1347236 SRR1347278 SRR1347389 SRR813959 SRR815920 SRR816517 SRR816609 SRR816677 SRR821573
前 列 腺	SRR1069209 SRR1069514 SRR1073069 SRR1074410 SRR1075126 SRR1075530 SRR1076120 SRR1077429 SRR1078164 SRR1078684 SRR1078855 SRR1080318

[0294]

	SRR1080696	SRR1081789	SRR1082496	SRR1083732	SRR1086441	SRR1086514
	SRR1086869	SRR1091645	SRR1091990	SRR1092444	SRR1092468	SRR1092636
	SRR1092913	SRR1093075	SRR1093697	SRR1096081	SRR1097344	SRR1098686
	SRR1099402	SRR1105441	SRR1308860	SRR1310939	SRR1312002	SRR1315353
	SRR1317751	SRR1323699	SRR1324314	SRR1326100	SRR1332360	SRR1335605
	SRR1335964	SRR813108	SRR815280	SRR815542	SRR815845	SRR816818
	SRR820234					SRR816969
皮肤 - 非阳光暴露的(耻骨上的)	SRR1069048	SRR1070232	SRR1070888	SRR1073605	SRR1074289	SRR1075247
	SRR1076292	SRR1077263	SRR1077898	SRR1079434	SRR1083215	SRR1083579
	SRR1084299	SRR1087801	SRR1091597	SRR1094216	SRR1095503	SRR1096408
	SRR1098216	SRR1100703	SRR1309920	SRR1309985	SRR1310053	SRR1311153
	SRR1311224	SRR1311916	SRR1312124	SRR1312244	SRR1312645	SRR1312934
	SRR1313494	SRR1314036	SRR1314137	SRR1314728	SRR1314810	SRR1315912
	SRR1316438	SRR1316747	SRR1316833	SRR1317022	SRR814491	SRR815164
	SRR815759	SRR815805	SRR818372	SRR818440	SRR819844	SRR820427
	SRR820810					
小肠 - 末端回肠	SRR1070133	SRR1071181	SRR1072602	SRR1074934	SRR1076046	SRR1076465
	SRR1077728	SRR1079973	SRR1084154	SRR1085378	SRR1087680	SRR1310497
	SRR1311731	SRR1313664	SRR1319059	SRR1319301	SRR1321483	SRR1326449
	SRR1326845	SRR1329508	SRR1330371	SRR1337749	SRR1337930	SRR1338402
	SRR1339086	SRR1340762	SRR1340782	SRR1343136	SRR1344079	SRR1344364
	SRR1351907	SRR1354400	SRR1356327	SRR1358803	SRR1359027	SRR1359587
	SRR1360321	SRR1361391	SRR1365655	SRR1365767	SRR1366102	SRR1366412
	SRR1367520	SRR1375371	SRR1378199	SRR1379036	SRR1380358	SRR1380436
	SRR1384312	SRR1387745				
胃	SRR1068953	SRR1069166	SRR1069714	SRR1069778	SRR1070382	SRR1070549
	SRR1070884	SRR1071761	SRR1072199	SRR1072700	SRR1072821	SRR1072920
	SRR1073459	SRR1074066	SRR1075874	SRR1076268	SRR1076417	SRR1076990
	SRR1078090	SRR1078759	SRR1079900	SRR1080672	SRR1081092	SRR1081235
	SRR1081717	SRR1081935	SRR1082933	SRR1082957	SRR1083149	SRR1083191
	SRR1083262	SRR1083360	SRR1083408	SRR1084252	SRR1085450	SRR1087101
	SRR1088068	SRR1088117	SRR808542	SRR810689	SRR810829	SRR811193
	SRR812152	SRR813234	SRR814195	SRR814268	SRR814820	SRR815326
	SRR815970	SRR819719				
睾丸	SRR1068788	SRR1068905	SRR1069734	SRR1070479	SRR1071379	SRR1071429
	SRR1072845	SRR1073531	SRR1075607	SRR1076490	SRR1077753	SRR1078299
	SRR1078612	SRR1079455	SRR1079612	SRR1080022	SRR1080811	SRR1080859
	SRR1081357	SRR1081401	SRR1081449	SRR1081614	SRR1081663	SRR1081688
	SRR1082307	SRR1083554	SRR1084347	SRR1087055	SRR1087535	SRR1088241
	SRR1308288	SRR1309425	SRR1311329	SRR1312288	SRR1314014	SRR807517
	SRR808065	SRR809667	SRR810531	SRR810899	SRR811447	SRR812912
	SRR813431	SRR814082	SRR814943	SRR815588	SRR817512	SRR818850
	SRR820839	SRR821518				
甲状腺	SRR597952	SRR598068	SRR598100	SRR598364	SRR598565	SRR598645
	SRR599122	SRR599346	SRR599412	SRR601157	SRR601359	SRR601525
	SRR601549	SRR601843	SRR601962	SRR602338	SRR602389	SRR602951
	SRR602978	SRR603036	SRR603268	SRR603726	SRR603834	SRR603942
	SRR604148	SRR604294	SRR604342	SRR607502	SRR607679	SRR607705
	SRR608064	SRR608120	SRR608512	SRR613018	SRR613258	SRR613402
	SRR613711	SRR613795	SRR613975	SRR614023	SRR614107	SRR614275
	SRR614743	SRR614912	SRR615285	SRR615347	SRR615491	SRR615886
	SRR654969	SRR655696				
子宫	SRR1069466	SRR1071737	SRR1073483	SRR1074430	SRR1075850	SRR1077159
	SRR1077211	SRR1077996	SRR1078114	SRR1078188	SRR1078212	SRR1079213
	SRR1079408	SRR1079874	SRR1080342	SRR1082128	SRR1084553	SRR1085358
	SRR1086369	SRR1309745	SRR1313991	SRR1319242	SRR1319991	SRR1321720
	SRR1323234	SRR1329423	SRR1330082	SRR1336682	SRR1338468	SRR1339258
	SRR1343943	SRR1353686	SRR1358126	SRR1360280	SRR1361138	SRR1361838
	SRR1363718	SRR1374543	SRR1381372	SRR1382780	SRR1383237	SRR1387132
	SRR1388257	SRR808704	SRR810105	SRR815256	SRR817817	SRR818139
	SRR818646	SRR820026				

[0295]

[0296] 表7a:关于本研究使用的样品的信息-鼠类样品

样品名称	生物样品类型(BiosampleType)	株	H-2-D	H-2-K	H-2-L	副本类型	NbCells
EL4	细胞系	C57BL/6	b	b	-	未重复的	5 000 000
mTEChi_1	原代细胞	C57BL/6	b	b	-	生物学的	51 237
mTEChi_2	原代细胞	C57BL/6	b	b	-	生物学的	31 686
mTEChi_3	原代细胞	C57BL/6	b	b	-	生物学的	31 702
CT26	细胞系	Balb/c	d	d	d	未重复的	5 000 000
mTEChi_1	原代细胞	Balb/c	d	d	d	生物学的	16 338
mTEChi_2	原代细胞	Balb/c	d	d	d	生物学的	19 782
mTEChi_3	原代细胞	Balb/c	d	d	d	生物学的	23 130

样品名称	BioAnalyser-RIN	输入 RNA_ng	核酸类型	链特异性的	平台	读段类型
EL4	9,95	4 000	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
mTEChi_1	10	100	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
mTEChi_2	9,2	100	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
mTEChi_3	9,9	100	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
CT26	10	2 000	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
mTEChi_1	9,5	50	多腺苷酸化的 mRNA	链特异性的	NextSeq 500	双端
mTEChi_2	9,4	50	多腺苷酸化的 mRNA	链特异性的	NextSeq 500	双端
mTEChi_3	9,2	50	多腺苷酸化的 mRNA	链特异性的	NextSeq 500	双端

样品名称	读段长度_bp	总 Nb 读段	登录号_GEO	NbRepMS	NbCellsMS (x10 <sup>6</sup> )	登录码_MS 数据
EL4	100	240372644	GSE111092	3	250	PXD009064
mTEChi_1	100	159208840	GSE111092	N/A	N/A	N/A
mTEChi_2	100	145643202	GSE111092	N/A	N/A	N/A
mTEChi_3	100	152139924	GSE111092	N/A	N/A	N/A
CT26	100	247522370	GSE111092	3	250	PXD009065
mTEChi_1	80	156128844	GSE111092	N/A	N/A	N/A
mTEChi_2	80	161566962	GSE111092	N/A	N/A	N/A
mTEChi_3	80	137929352	GSE111092	N/A	N/A	N/A

[0298] Table 7b:关于本研究使用的样品的信息-人样品

样品名称	生物样品类型	HLA-A	HLA-B	HLA-C	副本类型	NbCells
07H103	原发性白血病细胞	01:01   02:01	40:01   44:02	03:04   05:01	未重复的	2 600 000
10H080	原发性白血病细胞	02:01   11:01	40:01   44:03	03:04   16:01	未重复的	2 000 000
10H118	原发性白血病细胞	01:01   02:01	07:02   08:01	07:01   07:17	未重复的	3 400 000
12H018	原发性白血病细胞	02:01   11:01	07:02   35:03	07:02   12:03	未重复的	4 000 000
1c2	肿瘤活检物	11:01   23:01	35:01   44:03	04:01	未重复的	N/A

lc4	肿瘤活检物	02:01   03:01	07:02	07:02	未重复的	N/A
lc6	肿瘤活检物	01:01   24:02	08:01   27:13	02:02   07:01	未重复的	N/A
102015	原代 TEC	N/A	N/A	N/A	未重复的	33 076
062015	原代 TEC	N/A	N/A	N/A	未重复的	84 198
S5	原代 mTEC	N/A	N/A	N/A	未重复的	59 197
S9	原代 mTEC	N/A	N/A	N/A	未重复的	100 719
S0	原代 mTEC	N/A	N/A	N/A	未重复的	50 058
S11	原代 mTEC	N/A	N/A	N/A	未重复的	100 506

样品名称	BioAnalyser_RIN	输入 RNA_ng	核酸类型	链特异性的	平台	读段类型
07H103	10	500	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
10H080	10	500	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
10H118	9	500	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
12H018	9	500	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
lc2	9,2	4 000	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
lc4	9,4	4 000	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
[0300] lc6	8,9	4 000	多腺苷酸化的 mRNA	链特异性的	HiSeq 2000	双端
102015	7	8	多腺苷酸化的 mRNA	链特异性的	NextSeq 500	双端
062015	7	13	多腺苷酸化的 mRNA	链特异性的	NextSeq 500	双端
S5	7	41	多腺苷酸化的 mRNA	链特异性的	NextSeq 500	双端
S9	8	56	多腺苷酸化的 mRNA	链特异性的	NextSeq 500	双端
S0	8	68	多腺苷酸化的 mRNA	链特异性的	NextSeq 500	双端
S11	7	59	多腺苷酸化的 mRNA	链特异性的	NextSeq 500	双端

样品名称	读段长度_bp	总 Nb 读段	登录号_GEO	NbRepMS	NbCellsMS (x10 <sup>6</sup> )	登录码_MS 数据
07H103	100	105 863 640	GSE113972	3	650	PXD009749
10H080	100	129 444 492	GSE113972	3 / 4	500 / 100	PXD009753 / PXD007935
10H118	100	226 508 070	GSE113972	3	700	PXD009750
12H018	100	161 724 658	GSE113972	3	465	PXD009751
lc2	100	268 396 930	GSE113972	2	N/A	PXD009752
lc4	100	262 531 548	GSE113972	2	N/A	PXD009754
lc6	100	246 868 078	GSE113972	2	N/A	PXD009755
102015	80	134 624 214	N/A	链特异性的	NextSeq 500	双端

	062015	80	136 558 238	N/A	链特异性的	NextSeq 500	双端
	S5	80	200 363 532	N/A	链特异性的	NextSeq 500	双端
[0301]	S9	80	229 281 098	N/A	链特异性的	NextSeq 500	双端
	S0	80	231 185 678	N/A	链特异性的	NextSeq 500	双端
	S11	80	251 770 122	N/A	链特异性的	NextSeq 500	双端

[0302] 虽然本发明在上文已通过其具体实施例进行了描述,但可以在不背离所附权利要求书中所定义的主体发明的精神和性质的情况下,对其进行修改。在权利要求书中,词语“包括(comprising)”作为开放式术语使用,实质上等同于短语“包括但不限于”。除非上下文另有明确指示,单数形式“一个或一种(a)”、“一个或一种(an)”和“所述(The)”包括相应的复数指代。

[0303] 参考文献

[0304] 1.Mlecnik,B.,et al.The tumor microenvironment and immunoscore are critical determinants of dissemination to distant metastasis.Sci Transl Med 8,327ra326(2016) .

[0305] 2.Charoentong,P.,et al.Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade.Cell Rep 18,248-262(2017) .

[0306] 3.Shao,W.,et al.The systemMHC Atlas project.Nucleic Acids Res 46, D1237-D1247(2018) .

[0307] 4.Martin,S.D.,Coukos,G.,Holt,R.A.&Nelson,B.H.Targeting the undruggable:Immunotherapy meets personalized oncology in the genomic era.Ann Oncol 26,2367-2374(2015) .

[0308] 5.Marty,R.,et al.MHC-I genotype restricts the oncogenic mutational landscape.Cell 171,1272-1283e1215(2017) .

[0309] 6.Zhong,S.,et al.T-cell receptor affinity and avidity defines antitumor response and autoimmunity in T-cell immunotherapy.Proc Natl Acad Sci U S A 110,6973-6978(2013) .

[0310] 7.Sahin,U.,et al.Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer.Nature 547,222-226(2017) .

[0311] 8.Turajlic,S.,et al.Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype:a pan-cancer analysis.Lancet Oncol 18,1009-1021(2017) .

[0312] 9.Yadav,M.,et al.Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing.Nature 515,572-576(2014) .

[0313] 10.Pearson,H.,et al.MHC class I-associated peptides derive from selective regions of the human genome.J Clin Invest 126,4690-4701(2016) .

[0314] 11.Tran,E.,et al.Immunogenicity of somatic mutations in human gastrointestinal cancers.Science 350,1387-1390(2015) .

[0315] 12.Gros,A.,et al.Prospective identification of neoantigen-specific

lymphocytes in the peripheral blood of melanoma patients. *Nat Med* 22,433-438 (2016) .

[0316] 13. Bassani-Sternberg, M., et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* 7,13404(2016) .

[0317] 14. Mertens, F., Johansson, B., Fioretos, T. & Mitelman, F. The emerging complexity of gene fusions in cancer. *Nat Rev Cancer* 15,371-381(2015) .

[0318] 15. Baca, S. C., et al. Punctuated evolution of prostate cancer genomes. *Cell* 153,666-677(2013) .

[0319] 16. Hayward, N. K., et al. Whole-genome landscapes of major melanoma subtypes. *Nature* 545,175-180(2017) .

[0320] 17. Khurana, E., et al. Role of non-coding sequence variants in cancer. *Nat Rev Genet* 17,93-108(2016) .

[0321] 18. Laumont, C. M., et al. Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nat Commun* 7,10238(2016) .

[0322] 19. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160,48-61(2015) .

[0323] 20. Charoentong, P., et al. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* 18,248-262(2017) .

[0324] 21. Anwar, S. L., Wulaningsih, W. & Lehmann, U. Transposable Elements in Human Cancer: Causes and Consequences of Deregulation. *Int J Mol Sci* 18(2017) .

[0325] 22. Kassiotis, G. & Stoye, J. P. Immune responses to endogenous retroelements: taking the bad with the good. *Nat Rev Immunol* 16,207-219(2016) .

[0326] 23. Kershaw, M. H., et al. Immunization against endogenous retroviral tumor-associated antigens. *Cancer Res* 61,7920-7924(2001) .

[0327] 24. Sacha, J. B., et al. Vaccination with cancer- and HIV infection-associated endogenous retrotransposable elements is safe and immunogenic. *J Immunol* 189,1467-1479(2012) .

[0328] 25. Malarkannan, S., Serwold, T., Nguyen, V., Sherman, L. A. & Shastri, N. The mouse mammary tumor virus env gene is the source of a CD8+ T-cell-stimulating peptide presented by a major histocompatibility complex class I molecule in a murine thymoma. *Proc Natl Acad Sci U S A* 93,13991-13996(1996) .

[0329] 26. Huang, A. Y., et al. The immunodominant major histocompatibility complex class I-restricted antigen of a murine colon tumor derives from an endogenous retroviral gene product. *Proc Natl Acad Sci U S A* 93,9730-9735 (1996) .

- [0330] 27. Schiavetti, F., Thonnard, J., Colau, D., Boon, T. & Coulie, P. G. A human endogenous retroviral sequence encoding an antigen recognized on melanoma by cytolytic T lymphocytes. *Cancer Res* 62, 5510-5516 (2002).
- [0331] 28. Takahashi, Y., et al. Regression of human kidney cancer following allogeneic stem cell transplantation is associated with recognition of an HERV-E antigen by T cells. *J Clin Invest* 118, 1099-1109 (2008).
- [0332] 29. Kim, M. J., Miller, C. M., Shadrach, J. L., Wagers, A. J. & Serwold, T. Young, proliferative thymic epithelial cells engraft and function in aging thymuses. *J Immunol* 194, 4784-4795 (2015).
- [0333] 30. Dobin, A., et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21 (2013).
- [0334] 31. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842 (2010).
- [0335] 32. Caron, E., et al. The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. *Mol Syst Biol* 7, 533 (2011).
- [0336] 33. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 32, 511-517 (2016).
- [0337] 34. Robinson, J. T., et al. Integrative genomics viewer. *Nat Biotechnol* 29, 24-26 (2011).
- [0338] 35. Bereman, M. S., et al. An Automated Pipeline to Monitor System Performance in Liquid Chromatography-Tandem Mass Spectrometry Proteomic Experiments. *J Proteome Res* 15, 4763-4769 (2016).
- [0339] 36. Yue, F., et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515, 355-364 (2014).
- [0340] 37. Barbosa-Morais, N. L., et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587-1593 (2012).
- [0341] 38. Patenaude, J. & Perreault, C. Thymic Mesenchymal Cells Have a Distinct Transcriptomic Profile. *J Immunol* 196, 4760-4770 (2016).
- [0342] 39. St-Pierre, C., Trofimov, A., Brochu, S., Lemieux, S. & Perreault, C. Differential Features of AIRE-Induced and AIRE-Independent Promiscuous Gene Expression in Thymic Epithelial Cells. *J Immunol* 195, 498-506 (2015).
- [0343] 40. Dumont-Lagace, M., St-Pierre, C. & Perreault, C. Sex hormones have pervasive effects on thymic epithelial cells. *Sci Rep* 5, 12895 (2015).
- [0344] 41. Dumont-Lagace, M., Brochu, S., St-Pierre, C. & Perreault, C. Adult thymic epithelium contains nonsenescent label-retaining cells. *J Immunol* 192, 2219-2226 (2014).
- [0345] 42. de Verteuil, D. A., et al. Immunoproteasomes shape the transcriptome and regulate the function of dendritic cells. *J Immunol* 193, 1121-1132 (2014).

- [0346] 43.de Verteuil,D.,et al.Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules.Mol Cell Proteomics 9,2034-2047 (2010) .
- [0347] 44.Moon,J.J.,et al.Naive CD4(+)T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude.Immunity 27,203-213(2007) .
- [0348] 45.Legoux,F.P.&Moon,J.J.Peptide:MHC tetramer-based enrichment of epitope-specific T cells.J Vis Exp(2012) .
- [0349] 46.McFarland,H.I.,Nahill,S.R.,Maciaszek,J.W.&Welsh,R.M.CD11b(Mac-1):a marker for CD8+cytotoxic T cell activation and memory in virus infection.J Immunol 149,1326-1333(1992) .
- [0350] 47.Chadburn,A.,Inghirami,G.&Knowles,D.M.Hairy cell leukemia-associated antigen LeuM5 (CD11c) is preferentially expressed by benign activated and neoplastic CD8 T cells.Am J Pathol 136,29-37(1990) .
- [0351] 48.Nesvizhskii,A.I.Proteogenomics:concepts,applications and computational strategies.Nat Methods 11,1114-1125(2014) .
- [0352] 49.Noble,W.S.Mass spectrometrists should search only for peptides they care about.Nat Methods 12,605-608(2015) .
- [0353] 50.Murphy,J.P.,et al.MHC-I Ligand Discovery Using Targeted Database Searches of Mass Spectrometry Data:Implications for T-Cell Immunotherapies.J Proteome Res 16,1806-1816(2017) .
- [0354] 51.Granados,D.P.,et al.Impact of genomic polymorphisms on the repertoire of human MHC class I-associated peptides.Nat Commun 5,3600(2014) .
- [0355] 52.Bassani-Sternberg,M.,Pletscher-Frankild,S.,Jensen,L.J.&Mann,M.Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation.Mol Cell Proteomics 14,658-673(2015) .
- [0356] 53.Fortier,M.H.,et al.The MHC class I peptide repertoire is molded by the transcriptome.J Exp Med 205,595-610(2008) .
- [0357] 54.Jenkins,M.K.&Moon,J.J.The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude.J Immunol188,4135-4140 (2012) .
- [0358] 55.Obar,J.J.,Khanna,K.M.&Lefrancois,L.Endogenous naive CD8+ T cell precursor frequency regulates primary and memory responses to infection.Immunity 28,859-869(2008) .
- [0359] 56.La Gruta,N.L.,et al.Primary CTL response magnitude in mice is determined by the extent of naive T cell recruitment and subsequent clonal expansion.J Clin Invest 120,1885-1894(2010) .

- [0360] 57. Mueller, S.N., Gebhardt, T., Carbone, F.R. & Heath, W.R. Memory T cell subsets, migration patterns, and tissue residence. *Annu Rev Immunol* 31, 137-161 (2013).
- [0361] 58. Baaten, B.J., Tinoco, R., Chen, A.T. & Bradley, L.M. Regulation of Antigen-Experienced T Cells: Lessons from the Quintessential Memory Marker CD44. *Front Immunol* 3, 23 (2012).
- [0362] 59. Laugel, B., et al. Different T cell receptor affinity thresholds and CD8 coreceptor dependence govern cytotoxic T lymphocyte activation and tetramer binding properties. *J Biol Chem* 282, 23799-23810 (2007).
- [0363] 60. Richards, D.M., Kyewski, B. & Feuerer, M. Re-examining the Nature and Function of Self-Reactive T cells. *Trends Immunol* 37, 114-125 (2016).
- [0364] 61. McGranahan, N., et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* 351, 1463-1469 (2016).
- [0365] 62. Assarsson, E., et al. A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. *J Immunol* 178, 7890-7901 (2007).
- [0366] 63. Martin, S.D., et al. Low Mutation Burden in Ovarian Cancer May Limit the Utility of Neoantigen-Targeted Vaccines. *PLoS One* 11, e0155189 (2016).
- [0367] 64. Rudensky, A., Preston-Hurlburt, P., Hong, S.C., Barlow, A. & Janeway, C.A., Jr. Sequence analysis of peptides bound to MHC class II molecules. *Nature* 353, 622-627 (1991).
- [0368] 65. Meydan, C., Otu, H.H. & Sezerman, O.U. Prediction of peptides binding to MHC class I and II alleles by temporal motif mining. *BMC Bioinformatics* 14 Suppl 2, S13 (2013).
- [0369] 66. Szpakowski, S., et al. Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements. *Gene* 448, 151-167 (2009).
- [0370] 67. Capietto, A.H., Jhunjhunwala, S. & Delamarre, L. Characterizing neoantigens for personalized cancer immunotherapy. *Curr Opin Immunol* 46, 58-65 (2017).
- [0371] 68. Helft, J., et al. GM-CSF Mouse Bone Marrow Cultures Comprise a Heterogeneous Population of CD11c(+) MHCII(+) Macrophages and Dendritic Cells. *Immunity* 42, 1197-1211 (2015).
- [0372] 69. Wimmers, F., Schreiber, G., Skold, A.E., Figdor, C.G. & De Vries, I.J. Paradigm Shift in Dendritic Cell-Based Immunotherapy: From in vitro Generated Monocyte-Derived DCs to Naturally Circulating DC Subsets. *Front Immunol* 5, 165 (2014).
- [0373] 70. Williams, M. & Malissen, B. A Death Notice for In-Vitro-Generated GM-CSF Dendritic Cells? *Immunity* 42, 988-990 (2015).

- [0374] 71.Melief,C.J.,van Hall,T.,Arens,R.,Ossendorp,F.&van der Burg,S.H.Therapeutic cancer vaccines.J Clin Invest 125,3401-3412(2015).
- [0375] 72.Guo,C.,et al.Therapeutic cancer vaccines:past,present,and future.Adv Cancer Res 119,421-475(2013).
- [0376] 73.Melero,I.,et al.Therapeutic vaccines for cancer:an overview of clinical trials.Nat Rev Clin Oncol 11,509-524(2014).
- [0377] 74.Baruch,E.N.,Berg,A.L.,Besser,M.J.,Schachter,J.&Markel,G.Adoptive T cell therapy:An overview of obstacles and opportunities.Cancer 123,2154-2162 (2017).
- [0378] 75.Rosenberg,S.A.&Restifo,N.P.Adoptive cell transfer as personalized immunotherapy for human cancer.Science 348,62-68(2015).
- [0379] 76.Stoeckle,C.,et al.Isolation of myeloid dendritic cells and epithelial cells from human thymus.J Vis Exp,e50951(2013).
- [0380] 77.Marcais,G.&Kingsford,C.A fast,lock-free approach for efficient parallel counting of occurrences of k-mers.Bioinformatics 27,764-770(2011).
- [0381] 1a.Lanoix,J.,et al.Comparison of the MHC I immunopeptidome repertoire of B-cell lymphoblasts using two isolation methods.Proteomics,e1700251(2018).
- [0382] 2a.Kim,M.J.,Miller,C.M.,Shadrach,J.L.,Wagers,A.J.&Serwold,T.Young, proliferative thymic epithelial cells engraft and function in aging thymuses.J Immunol 194,4784-4795(2015).
- [0383] 3a.Stoeckle,C.,et al.Isolation of myeloid dendritic cells and epithelial cells from human thymus.J Vis Exp,e50951(2013).
- [0384] 4a.Dobin,A.,et al.STAR:Ultrafast universal RNA-seq aligner.Bioinformatics 29,15-21(2013).
- [0385] 5a.Daouda,T.,Perreault,C.&Lemieux,S.pyGeno:A python package for precision medicine and proteogenomics.F1000Res 5,381(2016).
- [0386] 6a.Marcais,G.&Kingsford,C.A fast,lock-free approach for efficient parallel counting of occurrences of k-mers.Bioinformatics 27,764-770(2011).
- [0387] 7a.Caron,E.,et al.The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation.Mol Syst Biol 7,533(2011).
- [0388] 8a.Andreatta,M.&Nielsen,M.Gapped sequence alignment using artificial neural networks:Application to the MHC class I system.Bioinformatics32,511-517(2016).
- [0389] 9a.Robinson,J.T.,et al.Integrative genomics viewer.Nat Biotechnol 29, 24-26(2011).
- [0390] 10a.Yue,F.,et al.A comparative encyclopedia of DNA elements in the mouse genome.Nature 515,355-364(2014).
- [0391] 11a.Sloan,C.A.,et al.ENCODE data at the ENCODE portal.Nucleic Acids Res 44,D726-732(2016).

- [0392] 12a. Bereman, M.S., et al. An automated pipeline to monitor system performance in liquid chromatography-tandem mass spectrometry proteomic experiments. *J Proteome Res* 15, 4763-4769 (2016).
- [0393] 13a. de Verteuil, D., et al. Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules. *Mol Cell Proteomics* 9, 2034-2047 (2010).
- [0394] 14a. Vincent, K., et al. Rejection of leukemic cells requires antigen-specific T cells with high functional avidity. *Biol Blood Marrow Transplant* 20, 37-45 (2014).
- [0395] 15a. Moon, J.J., et al. Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* 27, 203-213 (2007).
- [0396] 16a. Legoux, F.P. & Moon, J.J. Peptide:MHC tetramer-based enrichment of epitope-specific T cells. *J Vis Exp* 68, 4420 (2012).
- [0397] 17a. McFarland, H.I., Nahill, S.R., Maciaszek, J.W. & Welsh, R.M. CD11b (Mac-1): A marker for CD8+ cytotoxic T cell activation and memory in virus infection. *J Immunol* 149, 1326-1333 (1992).
- [0398] 18a. Chadburn, A., Inghirami, G. & Knowles, D.M. Hairy cell leukemia-associated antigen LeuM5 (CD11c) is preferentially expressed by benign activated and neoplastic CD8 T cells. *Am J Pathol* 136, 29-37 (1990)
- [0399] 19a. Vizcaino, J.A., et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 44, 11033 (2016).

- [0001] 序列列表
- [0002] <110> 蒙特利尔大学 (UNIVERSITÉ DE MONTRÉAL)
- [0003] <120> 用于鉴定肿瘤特异性抗原的基于蛋白质基因组学的方法
- [0004] <130> PPI21170041CA
- [0005] <150> US 62/724,760
- [0006] <151> 2018-08-30
- [0007] <160> 42
- [0008] <170> PatentIn version 3.5
- [0009] <210> 1
- [0010] <211> 9
- [0011] <212> PRT
- [0012] <213> 小鼠 (Mus musculus)
- [0013] <400> 1
- [0014] Gly Tyr Gln Lys Met Lys Ala Leu Leu
- [0015] 1 5
- [0016] <210> 2
- [0017] <211> 9
- [0018] <212> PRT
- [0019] <213> 小鼠 (Mus musculus)
- [0020] <220>
- [0021] <221> 变体
- [0022] <222> (4) .. (4)
- [0023] <223> Xaa 是 Lys 或 Glu
- [0024] <400> 2
- [0025] Lys Pro Leu Xaa Ala Pro Leu Asp Leu
- [0026] 1 5
- [0027] <210> 3
- [0028] <211> 9
- [0029] <212> PRT
- [0030] <213> 小鼠 (Mus musculus)
- [0031] <220>
- [0032] <221> 变体
- [0033] <222> (7) .. (7)
- [0034] <223> Xaa 是 Ser 或 Gly
- [0035] <400> 3
- [0036] Lys Tyr Leu Ser Val Gln Xaa Gln Leu
- [0037] 1 5
- [0038] <210> 4



[0078] <400> 8  
[0079] Ser Gly Pro Pro Tyr Tyr Lys Gly Ile  
[0080] 1 5  
[0081] <210> 9  
[0082] <211> 8  
[0083] <212> PRT  
[0084] <213> 小鼠 (Mus musculus)  
[0085] <400> 9  
[0086] Ser Pro His Gln Val Phe Asn Leu  
[0087] 1 5  
[0088] <210> 10  
[0089] <211> 9  
[0090] <212> PRT  
[0091] <213> 小鼠 (Mus musculus)  
[0092] <400> 10  
[0093] Ser Pro Ser Tyr Val Tyr His Gln Phe  
[0094] 1 5  
[0095] <210> 11  
[0096] <211> 8  
[0097] <212> PRT  
[0098] <213> 小鼠 (Mus musculus)  
[0099] <400> 11  
[0100] Ala Thr Gln Gln Phe Gln Gln Leu  
[0101] 1 5  
[0102] <210> 12  
[0103] <211> 8  
[0104] <212> PRT  
[0105] <213> 小鼠 (Mus musculus)  
[0106] <400> 12  
[0107] Ile Ile Leu Glu Phe His Ser Leu  
[0108] 1 5  
[0109] <210> 13  
[0110] <211> 9  
[0111] <212> PRT  
[0112] <213> 小鼠 (Mus musculus)  
[0113] <400> 13  
[0114] Ser Ser Pro Arg Gly Ser Ser Thr Leu  
[0115] 1 5  
[0116] <210> 14

- [0117] <211> 9  
[0118] <212> PRT  
[0119] <213> 小鼠 (Mus musculus)  
[0120] <400> 14  
[0121] Thr Val Pro Leu Asn His Asn Thr Leu  
[0122] 1 5  
[0123] <210> 15  
[0124] <211> 8  
[0125] <212> PRT  
[0126] <213> 小鼠 (Mus musculus)  
[0127] <220>  
[0128] <221> 变体  
[0129] <222> (4) .. (4)  
[0130] <223> Xaa 是 Ile 或 Leu  
[0131] <400> 15  
[0132] Val Asn Tyr Xaa His Arg Asn Val  
[0133] 1 5  
[0134] <210> 16  
[0135] <211> 8  
[0136] <212> PRT  
[0137] <213> 小鼠 (Mus musculus)  
[0138] <400> 16  
[0139] Val Thr Pro Val Tyr Gln His Leu  
[0140] 1 5  
[0141] <210> 17  
[0142] <211> 9  
[0143] <212> PRT  
[0144] <213> 智人 (Homo sapiens)  
[0145] <400> 17  
[0146] Lys Ile Leu Ile Leu Leu Gln Ser Leu  
[0147] 1 5  
[0148] <210> 18  
[0149] <211> 9  
[0150] <212> PRT  
[0151] <213> 智人 (Homo sapiens)  
[0152] <400> 18  
[0153] Lys Ile Ser Leu Tyr Leu Pro Ala Leu  
[0154] 1 5  
[0155] <210> 19

[0156] <211> 9  
[0157] <212> PRT  
[0158] <213> 智人 (Homo sapiens)  
[0159] <400> 19  
[0160] Ser Leu Thr Ala Leu Val Phe His Val  
[0161] 1 5  
[0162] <210> 20  
[0163] <211> 8  
[0164] <212> PRT  
[0165] <213> 智人 (Homo sapiens)  
[0166] <400> 20  
[0167] His Glu Thr Leu Arg Leu Leu Leu  
[0168] 1 5  
[0169] <210> 21  
[0170] <211> 9  
[0171] <212> PRT  
[0172] <213> 智人 (Homo sapiens)  
[0173] <400> 21  
[0174] Arg Ile Phe Gly Phe Arg Leu Trp Lys  
[0175] 1 5  
[0176] <210> 22  
[0177] <211> 9  
[0178] <212> PRT  
[0179] <213> 智人 (Homo sapiens)  
[0180] <400> 22  
[0181] Thr Ser Phe Ala Glu Thr Trp Met Lys  
[0182] 1 5  
[0183] <210> 23  
[0184] <211> 9  
[0185] <212> PRT  
[0186] <213> 智人 (Homo sapiens)  
[0187] <400> 23  
[0188] Thr Ser Ile Pro Lys Pro Asn Leu Lys  
[0189] 1 5  
[0190] <210> 24  
[0191] <211> 8  
[0192] <212> PRT  
[0193] <213> 智人 (Homo sapiens)  
[0194] <400> 24

- [0195] Leu Pro Phe Glu Gln Lys Ser Leu  
 [0196] 1 5  
 [0197] <210> 25  
 [0198] <211> 9  
 [0199] <212> PRT  
 [0200] <213> 智人 (Homo sapiens)  
 [0201] <400> 25  
 [0202] Ser Leu Arg Glu Lys Gly Phe Ser Ile  
 [0203] 1 5  
 [0204] <210> 26  
 [0205] <211> 8  
 [0206] <212> PRT  
 [0207] <213> 智人 (Homo sapiens)  
 [0208] <400> 26  
 [0209] Val Pro Ala Ala Leu Arg Ser Leu  
 [0210] 1 5  
 [0211] <210> 27  
 [0212] <211> 10  
 [0213] <212> PRT  
 [0214] <213> 智人 (Homo sapiens)  
 [0215] <400> 27  
 [0216] Leu Leu Ala Ala Thr Ile Leu Leu Ser Val  
 [0217] 1 5 10  
 [0218] <210> 28  
 [0219] <211> 9  
 [0220] <212> PRT  
 [0221] <213> 智人 (Homo sapiens)  
 [0222] <220>  
 [0223] <221> 变体  
 [0224] <222> (5) .. (5)  
 [0225] <223> Xaa 是 Ala 或 Val  
 [0226] <400> 28  
 [0227] Ser Leu Phe Val Xaa Ser Leu Ser Leu  
 [0228] 1 5  
 [0229] <210> 29  
 [0230] <211> 9  
 [0231] <212> PRT  
 [0232] <213> 智人 (Homo sapiens)  
 [0233] <400> 29

[0234] Ile Ile Ala Pro Pro Pro Pro Pro Lys  
 [0235] 1 5  
 [0236] <210> 30  
 [0237] <211> 9  
 [0238] <212> PRT  
 [0239] <213> 智人 (Homo sapiens)  
 [0240] <400> 30  
 [0241] Leu Val Phe Asn Ile Ile Leu His Arg  
 [0242] 1 5  
 [0243] <210> 31  
 [0244] <211> 9  
 [0245] <212> PRT  
 [0246] <213> 智人 (Homo sapiens)  
 [0247] <400> 31  
 [0248] Met Ile Ser Pro Val Leu Ala Leu Lys  
 [0249] 1 5  
 [0250] <210> 32  
 [0251] <211> 9  
 [0252] <212> PRT  
 [0253] <213> 智人 (Homo sapiens)  
 [0254] <400> 32  
 [0255] Ser Leu Ser Tyr Leu Ile Leu Lys Lys  
 [0256] 1 5  
 [0257] <210> 33  
 [0258] <211> 9  
 [0259] <212> PRT  
 [0260] <213> 智人 (Homo sapiens)  
 [0261] <400> 33  
 [0262] Ser Ser Ala Ser Gln Leu Pro Ser Lys  
 [0263] 1 5  
 [0264] <210> 34  
 [0265] <211> 10  
 [0266] <212> PRT  
 [0267] <213> 智人 (Homo sapiens)  
 [0268] <400> 34  
 [0269] Ser Val Ile Gln Thr Gly His Leu Ala Lys  
 [0270] 1 5 10  
 [0271] <210> 35  
 [0272] <211> 9

- [0273] <212> PRT  
[0274] <213> 智人 (Homo sapiens)  
[0275] <400> 35  
[0276] Thr Thr Leu Lys Tyr Leu Trp Lys Lys  
[0277] 1 5  
[0278] <210> 36  
[0279] <211> 9  
[0280] <212> PRT  
[0281] <213> 智人 (Homo sapiens)  
[0282] <400> 36  
[0283] Lys Pro Ser Val Phe Pro Leu Ser Leu  
[0284] 1 5  
[0285] <210> 37  
[0286] <211> 9  
[0287] <212> PRT  
[0288] <213> 智人 (Homo sapiens)  
[0289] <220>  
[0290] <221> 变体  
[0291] <222> (2) .. (2)  
[0292] <223> Xaa 是 Arg 或 Lys  
[0293] <220>  
[0294] <221> 变体  
[0295] <222> (3) .. (3)  
[0296] <223> Xaa 是 Phe 或 Leu  
[0297] <400> 37  
[0298] Gln Xaa Xaa Gln Gly Arg Val Thr Met  
[0299] 1 5  
[0300] <210> 38  
[0301] <211> 10  
[0302] <212> PRT  
[0303] <213> 智人 (Homo sapiens)  
[0304] <400> 38  
[0305] Ser Arg Phe Ser Gly Val Pro Asp Arg Phe  
[0306] 1 5 10  
[0307] <210> 39  
[0308] <211> 9  
[0309] <212> PRT  
[0310] <213> 智人 (Homo sapiens)  
[0311] <220>

- [0312] <221> 变体  
[0313] <222> (5) .. (5)  
[0314] <223> Xaa 是 Asn 或 Asp  
[0315] <400> 39  
[0316] Thr Tyr Thr Gln Xaa Phe Asn Lys Phe  
[0317] 1 5  
[0318] <210> 40  
[0319] <211> 9  
[0320] <212> PRT  
[0321] <213> 淋巴细胞脉络丛脑膜炎病毒  
[0322] <400> 40  
[0323] Lys Ala Val Tyr Asn Phe Ala Thr Cys  
[0324] 1 5  
[0325] <210> 41  
[0326] <211> 9  
[0327] <212> PRT  
[0328] <213> 鼠类巨细胞病毒  
[0329] <400> 41  
[0330] His Gly Ile Arg Asn Ala Ser Phe Ile  
[0331] 1 5  
[0332] <210> 42  
[0333] <211> 8  
[0334] <212> PRT  
[0335] <213> 牛痘病毒  
[0336] <400> 42  
[0337] Thr Ser Tyr Lys Phe Glu Ser Val  
[0338] 1 5

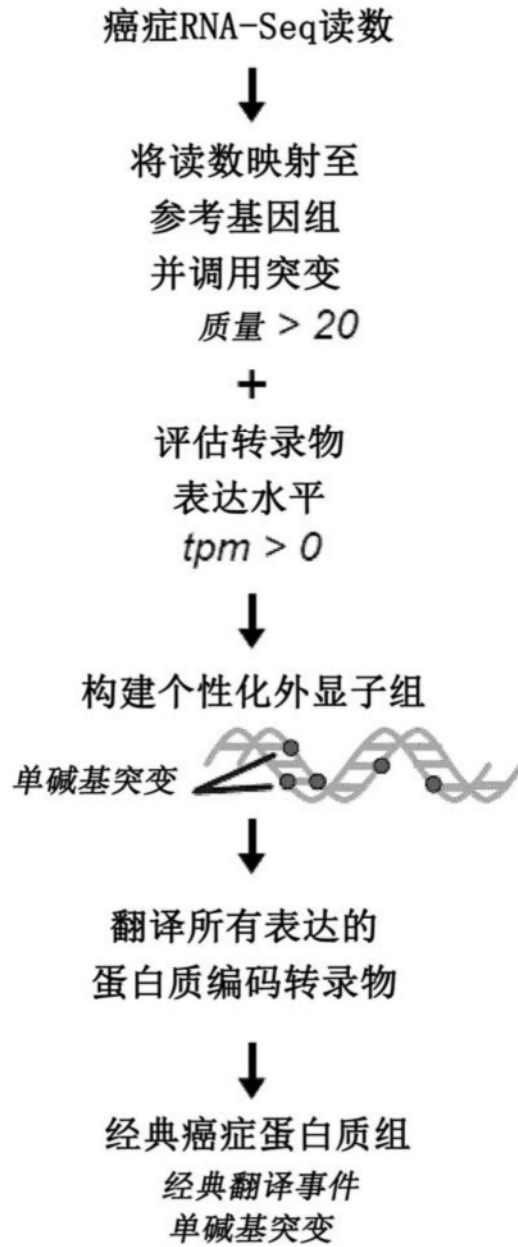


图1A

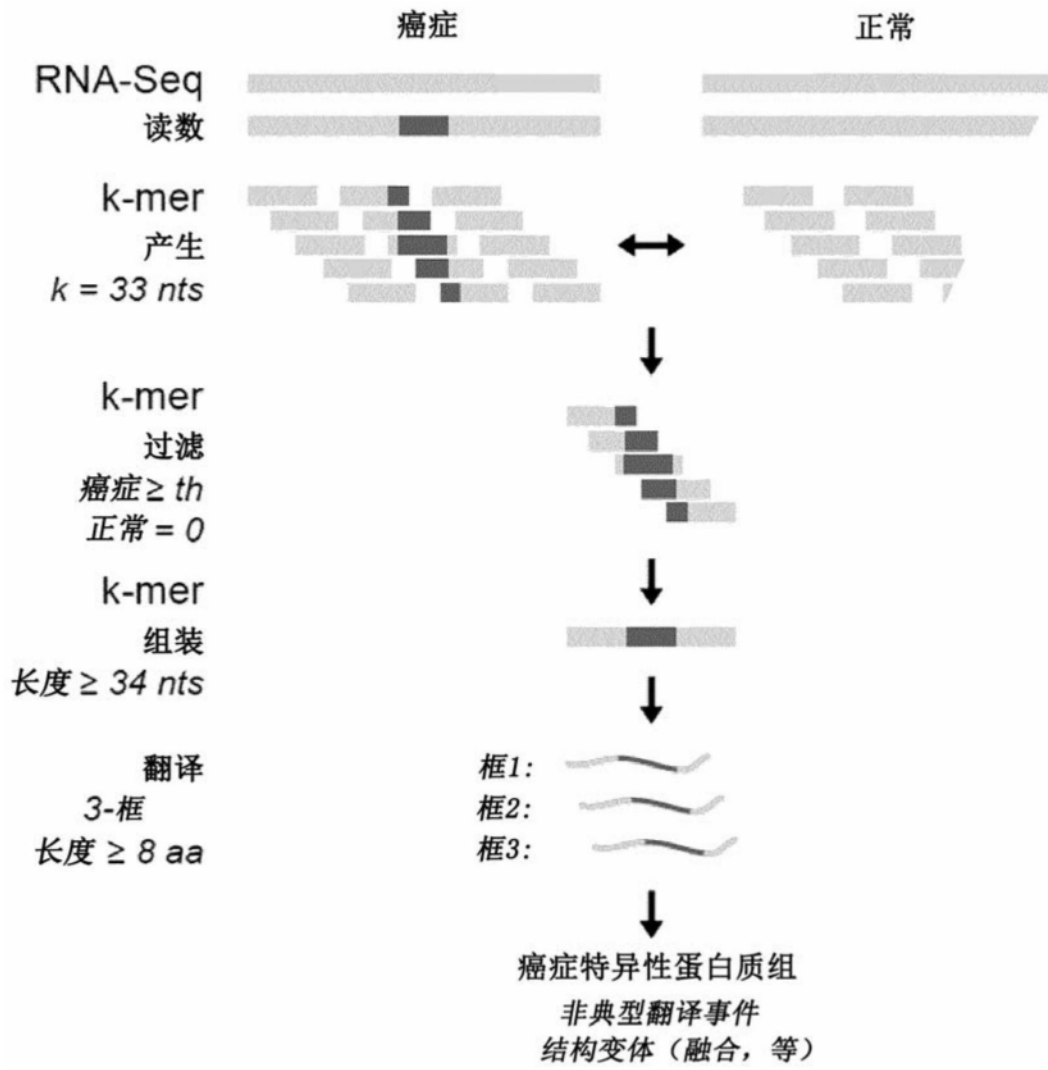


图1B

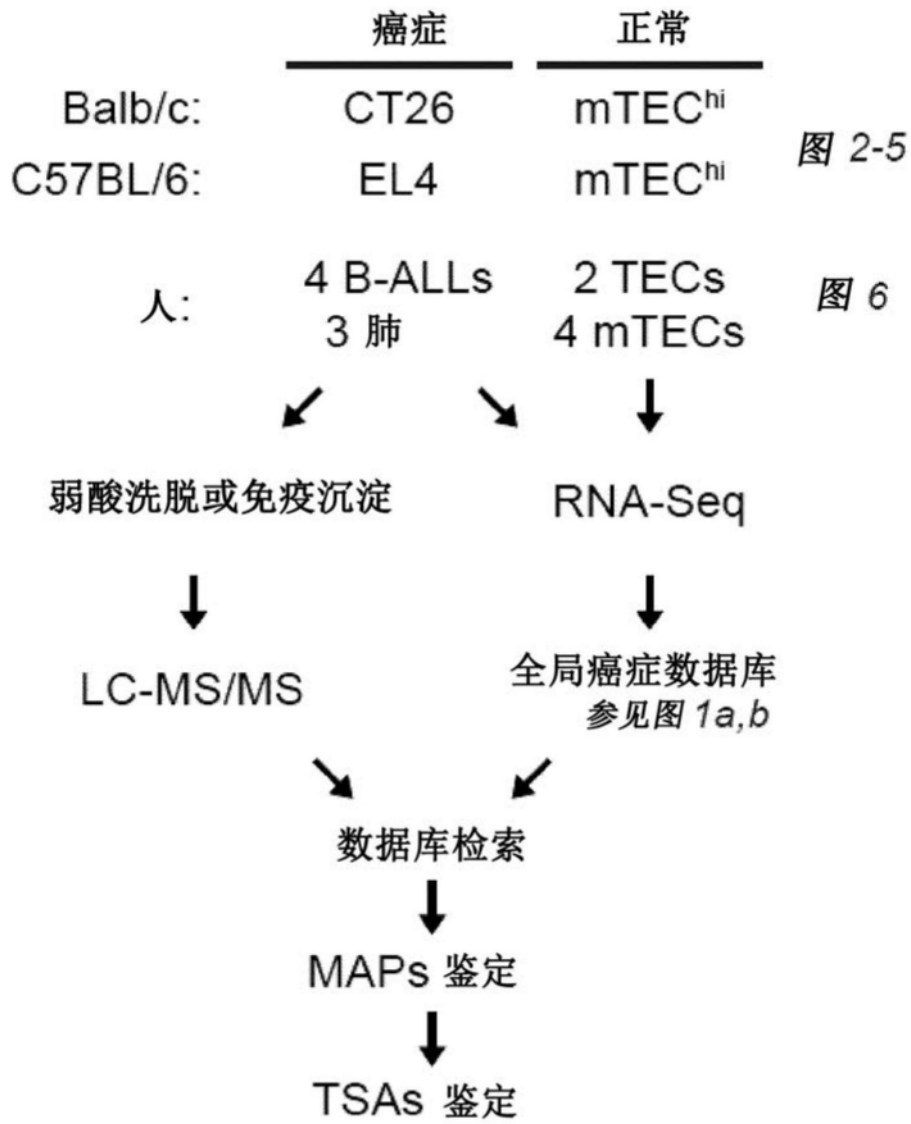


图1C

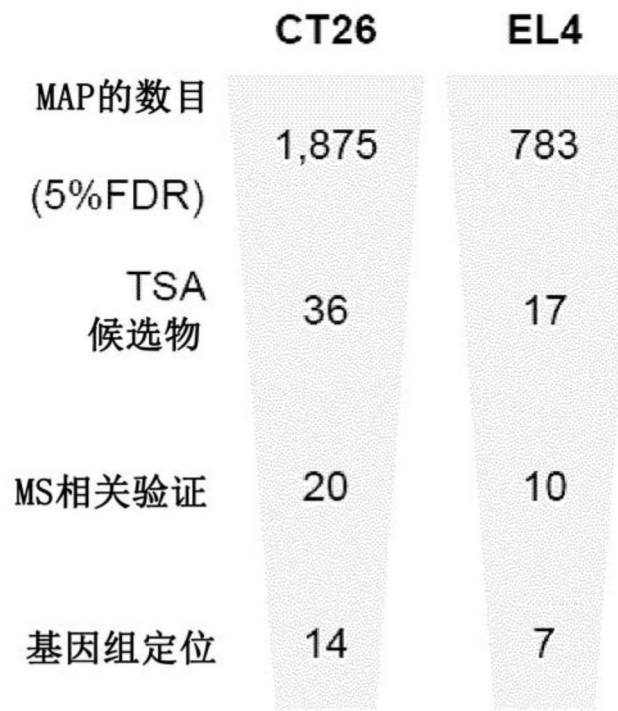


图2A

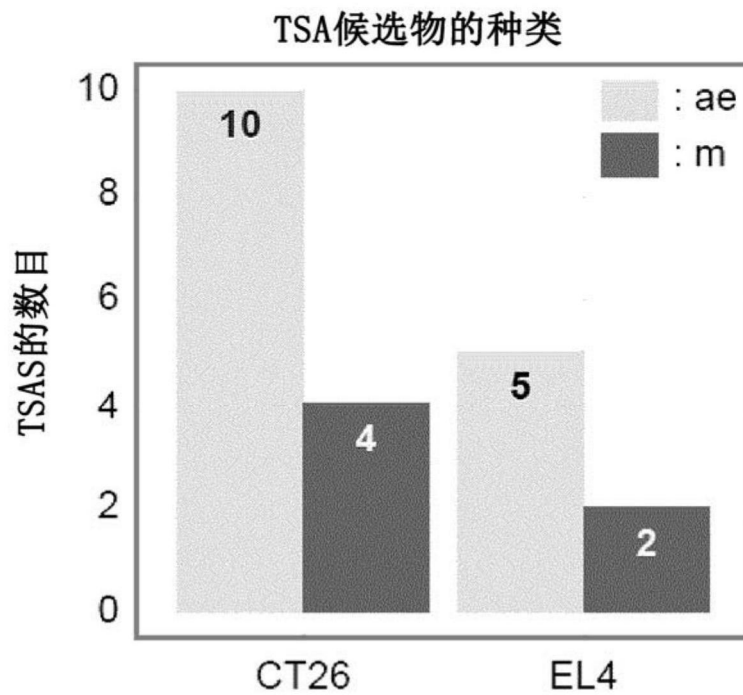


图2B

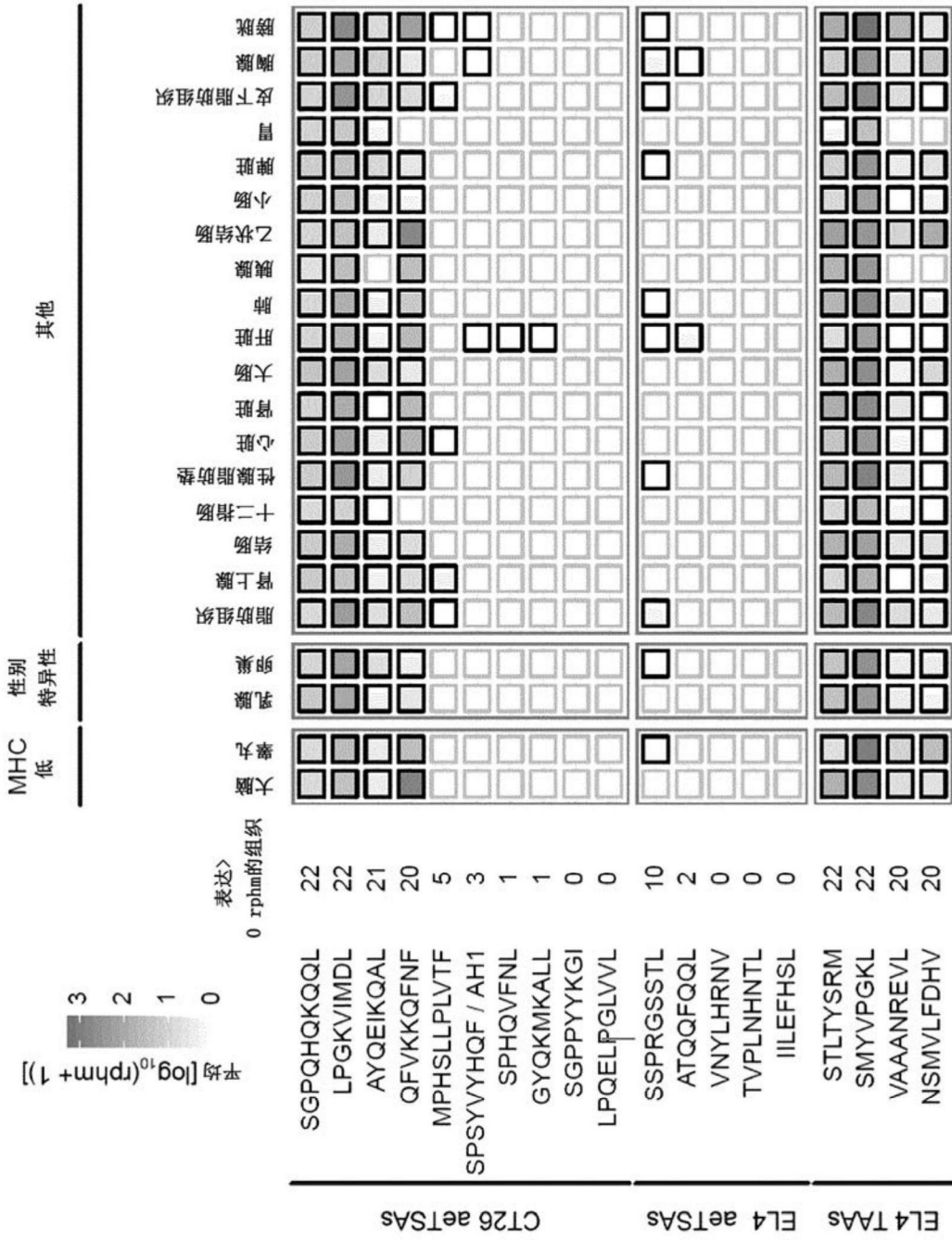


图2C

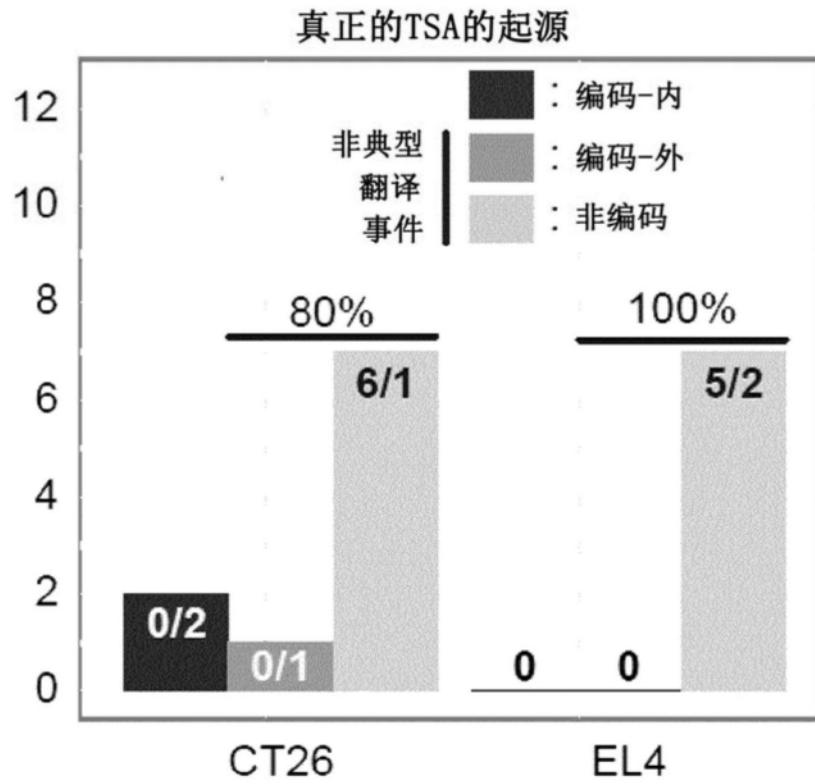


图2D

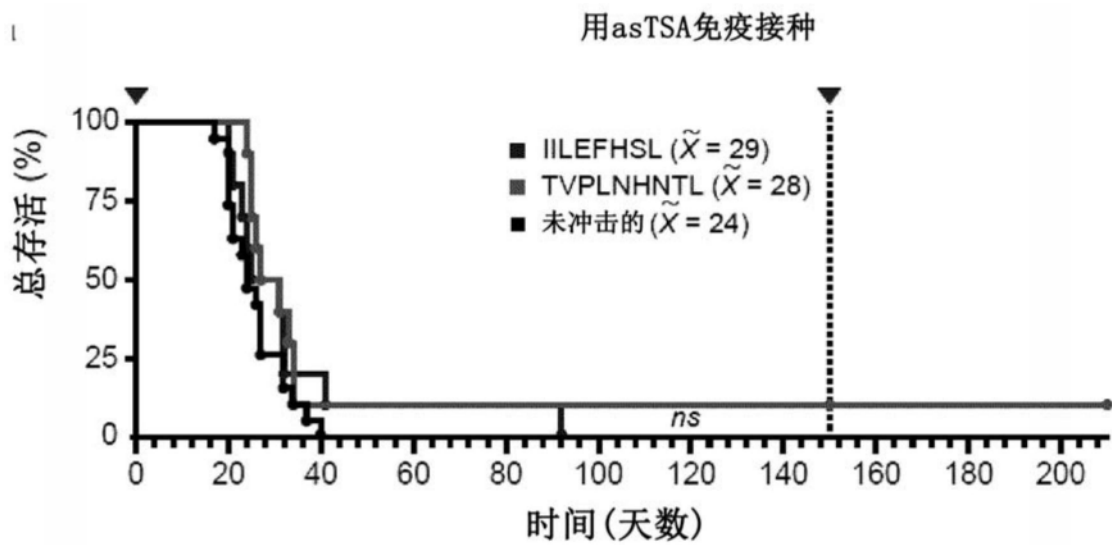


图3A

用ERE TSA免疫接种

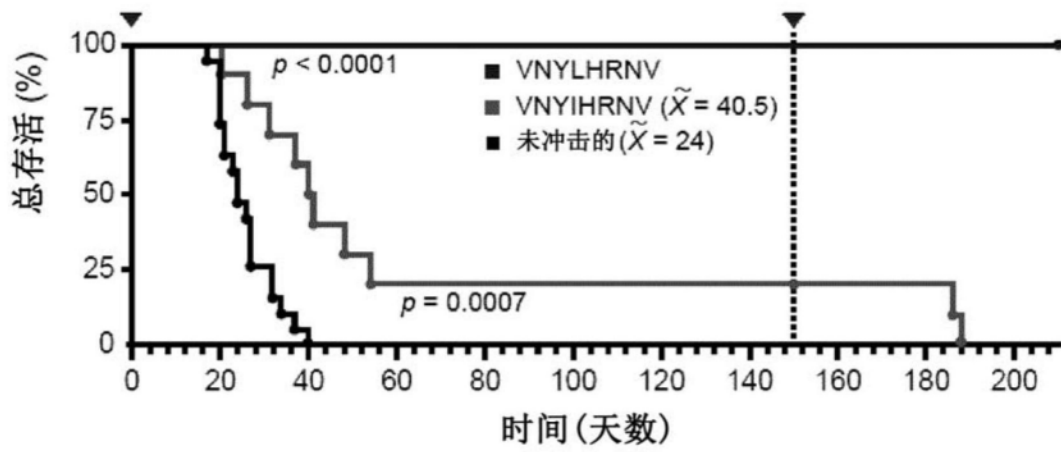


图3B

用mTSA免疫接种

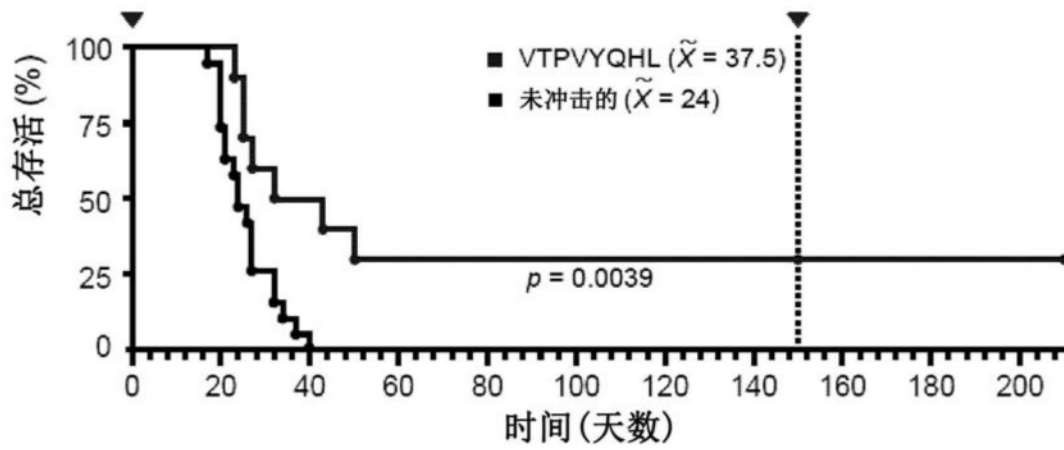


图3C

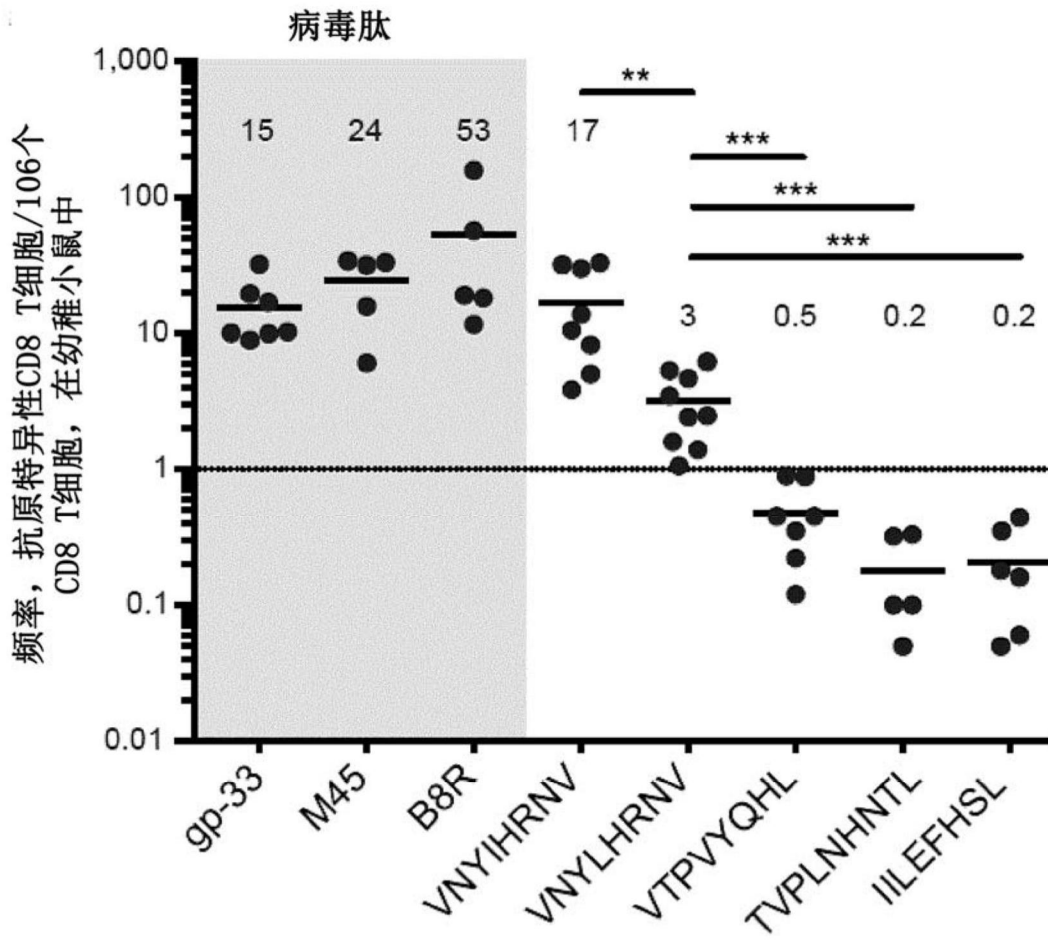


图4A

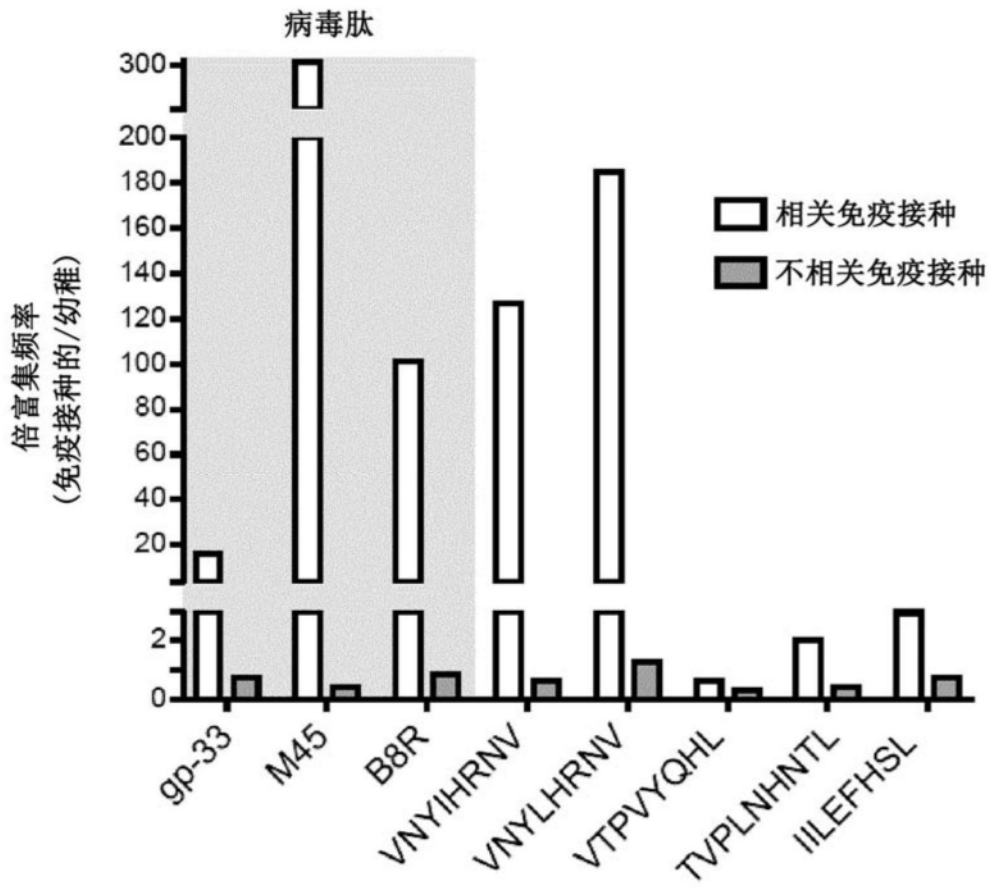


图4B

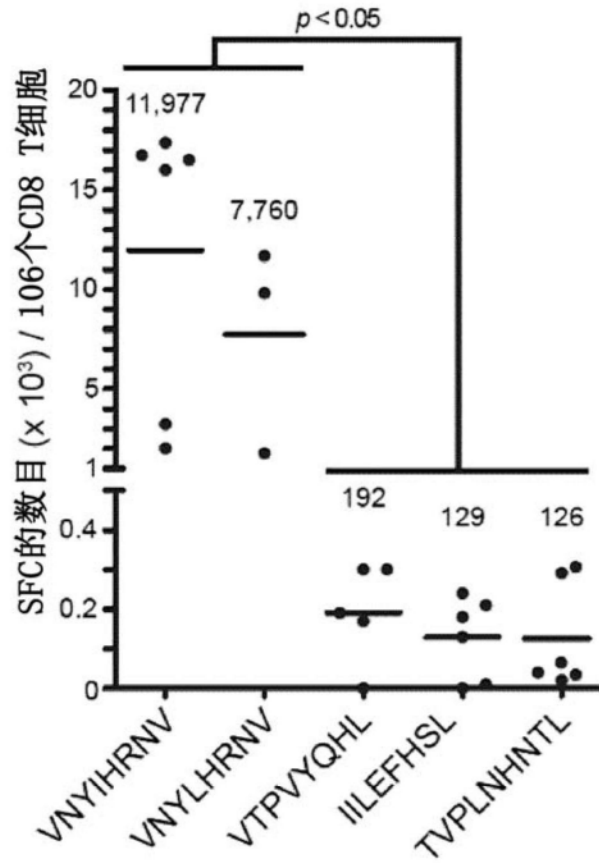


图4C

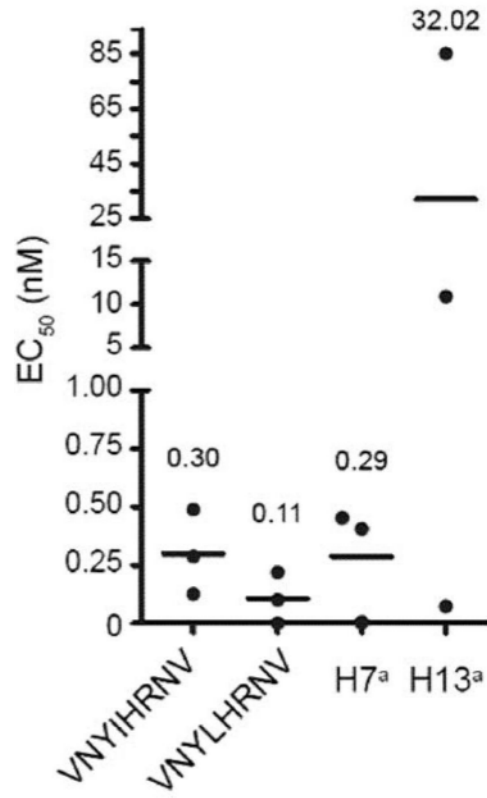


图4D

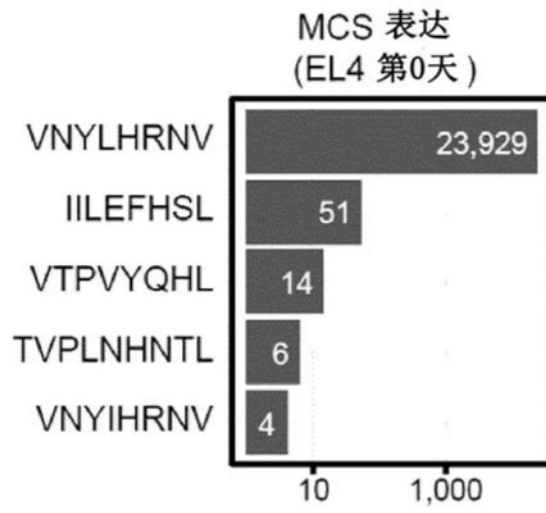


图5A

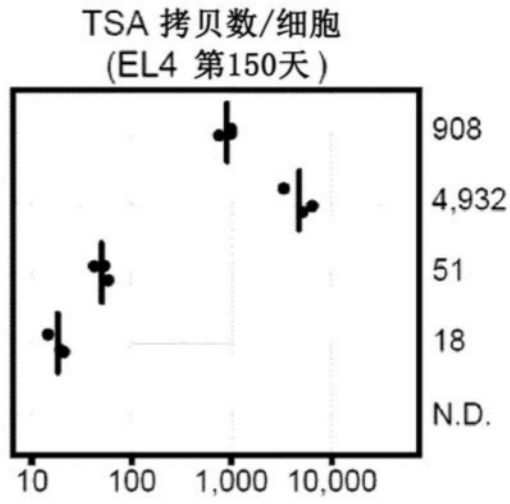


图5B

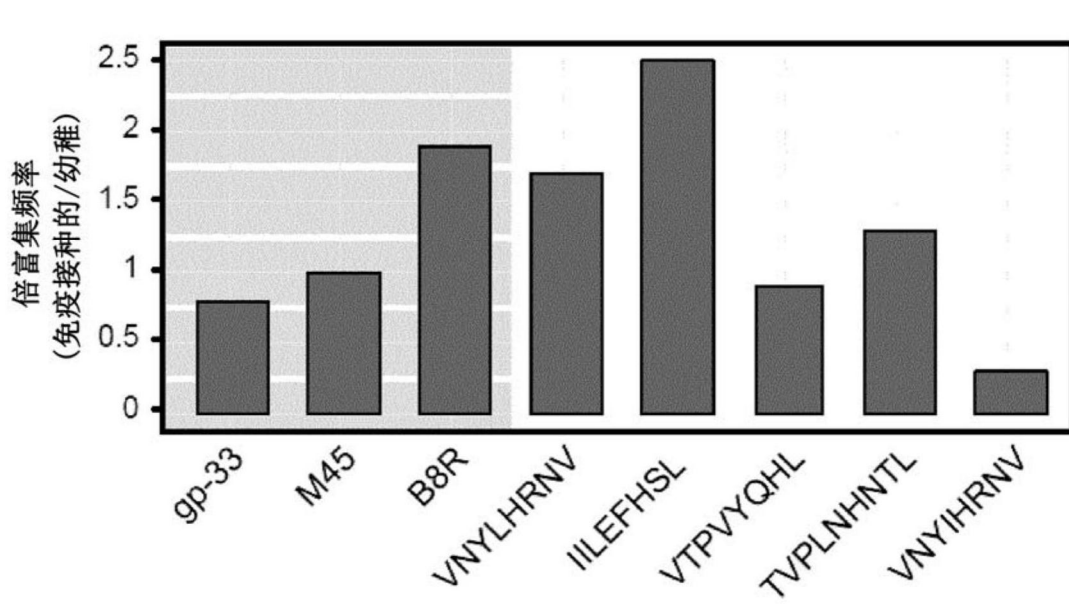


图5C

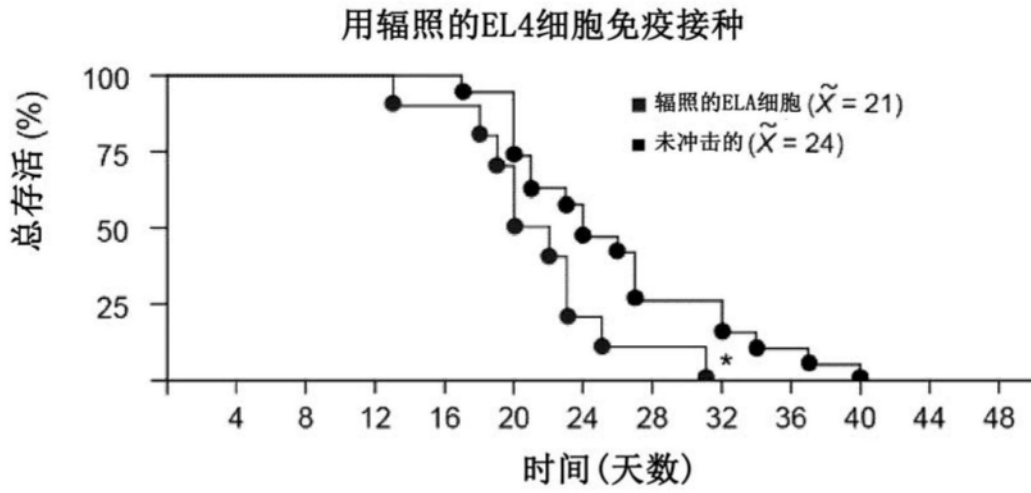


图5D

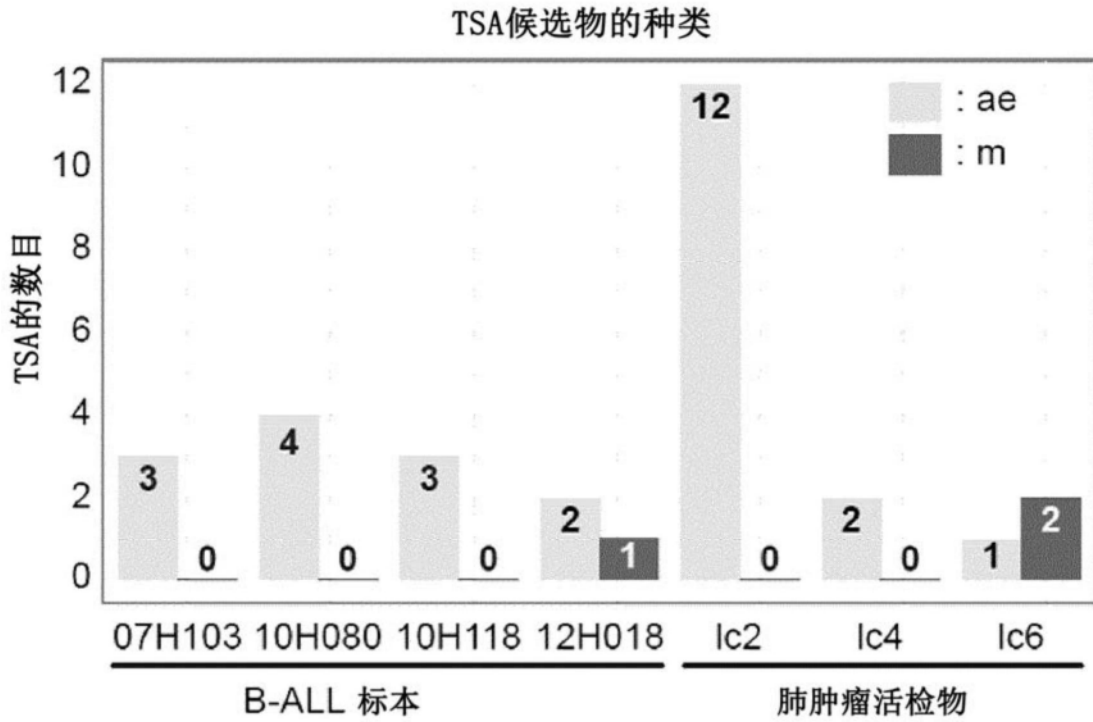


图6A

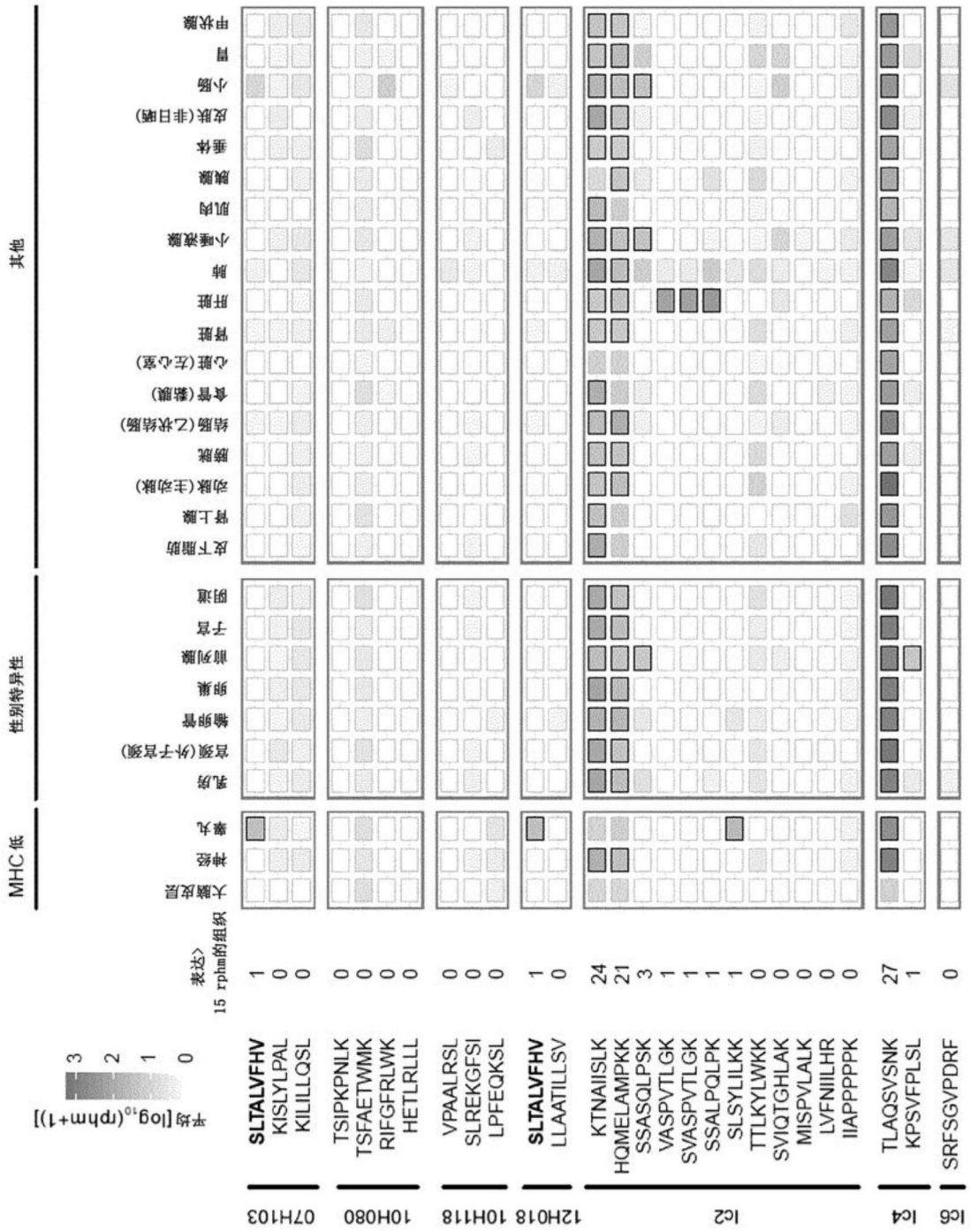


图6B

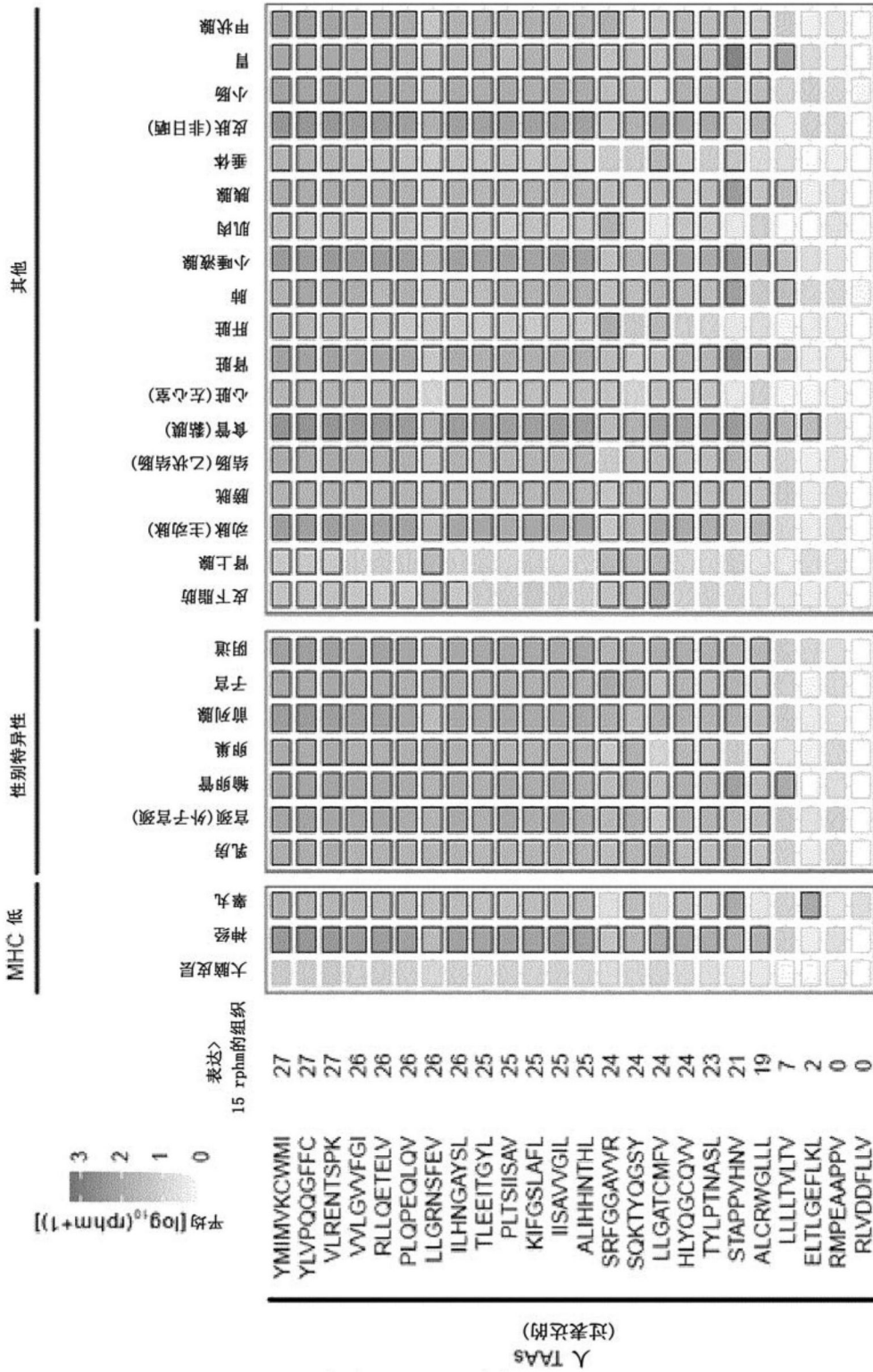


图6B(续)

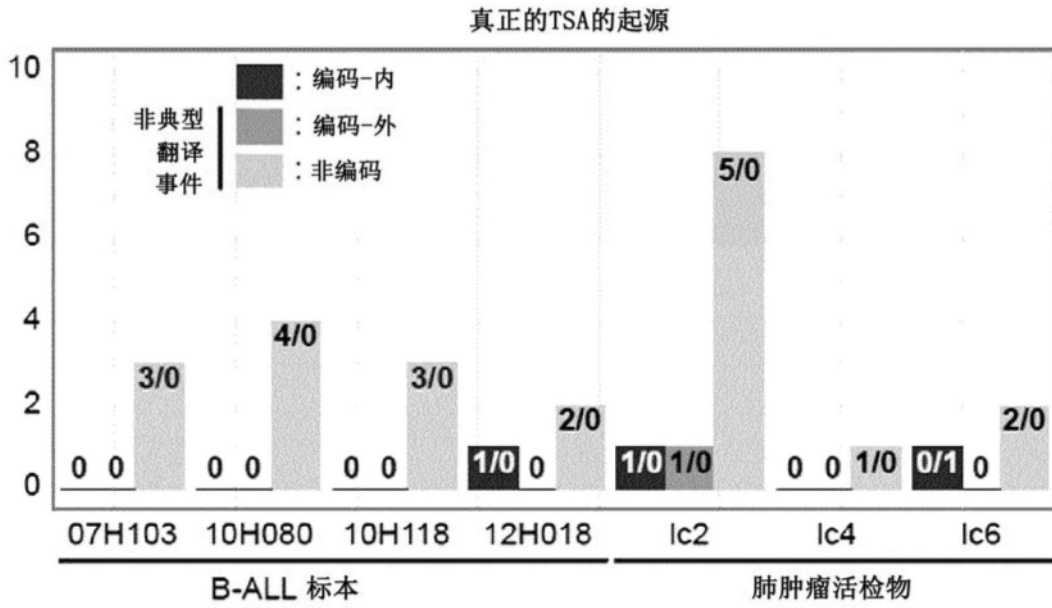


图6C

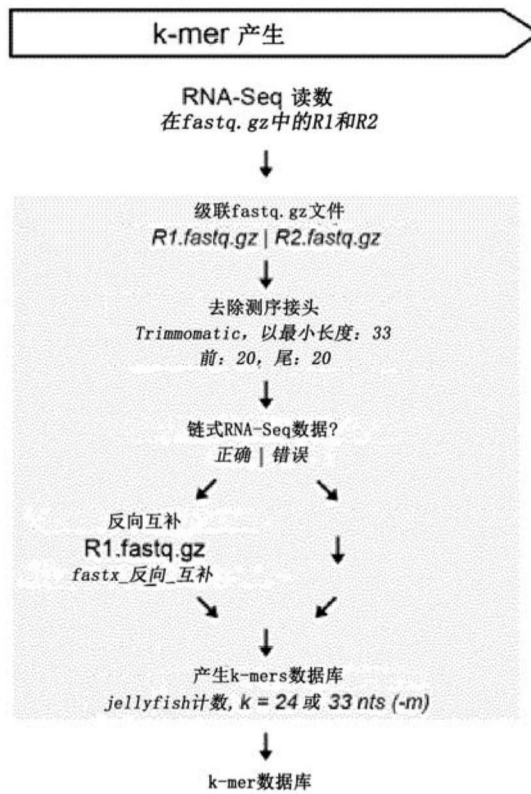


图7A



图7B

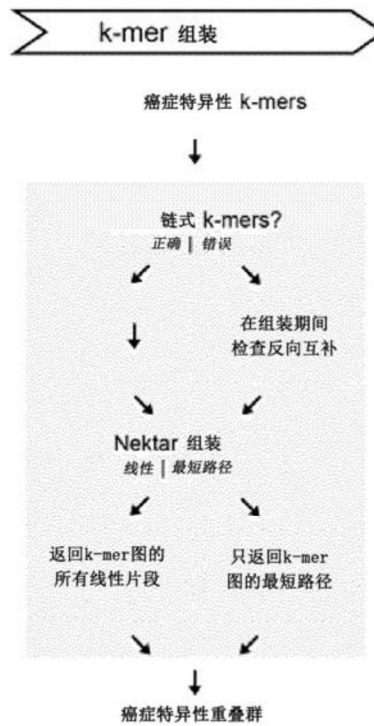


图7C

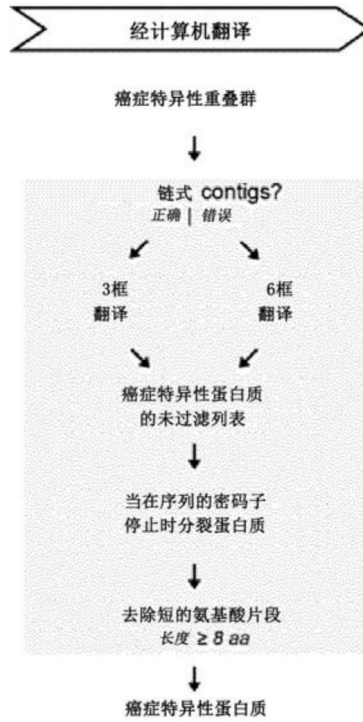


图7D

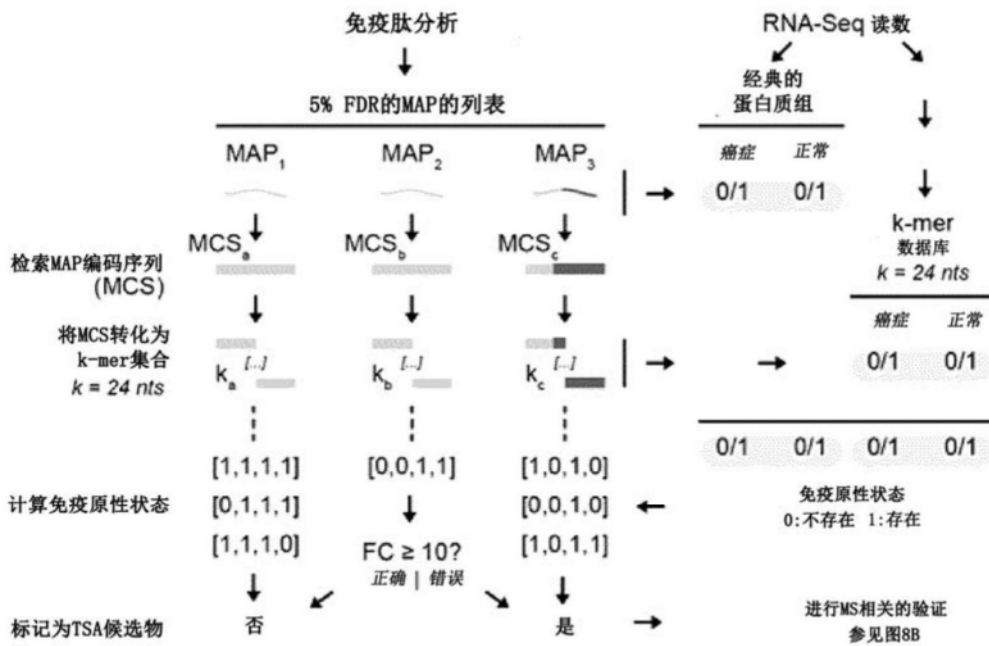


图8A



图8B

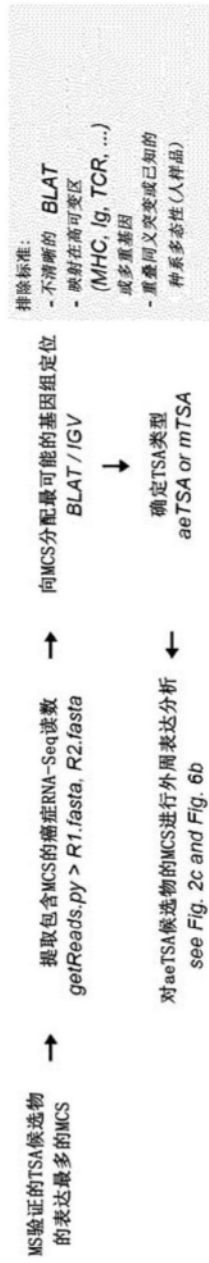


图8C

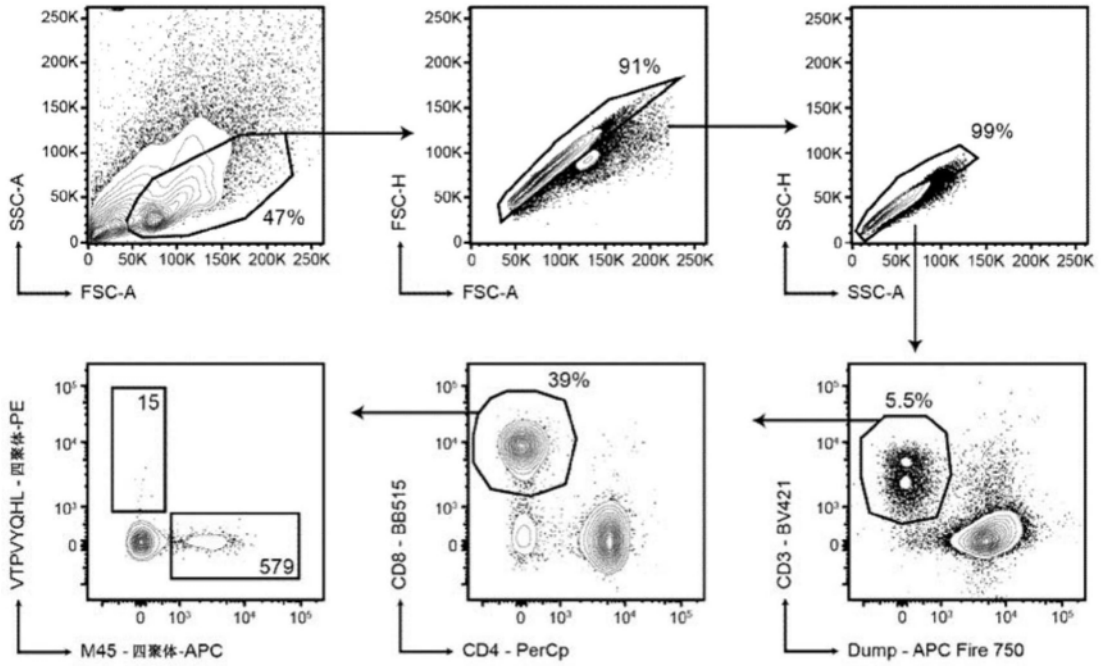


图9A

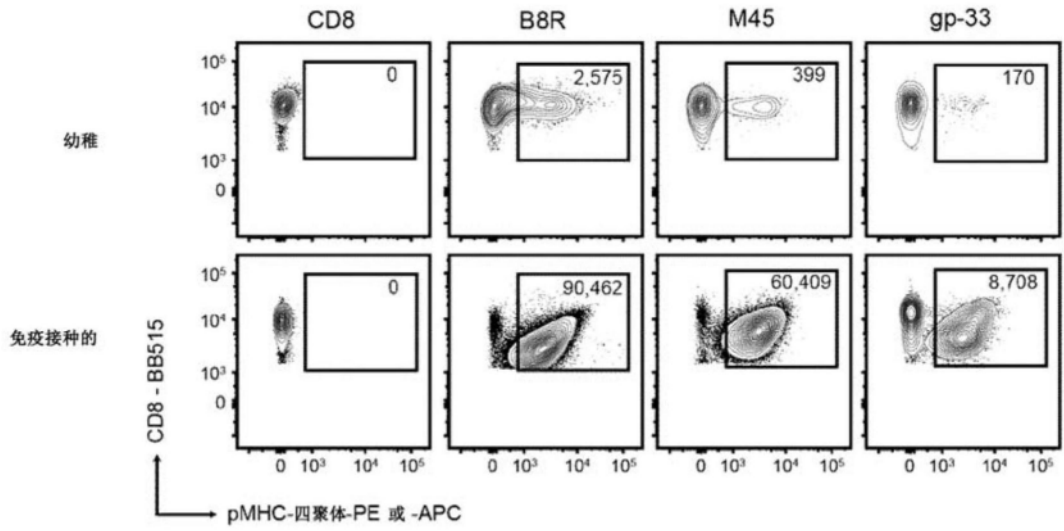


图9B

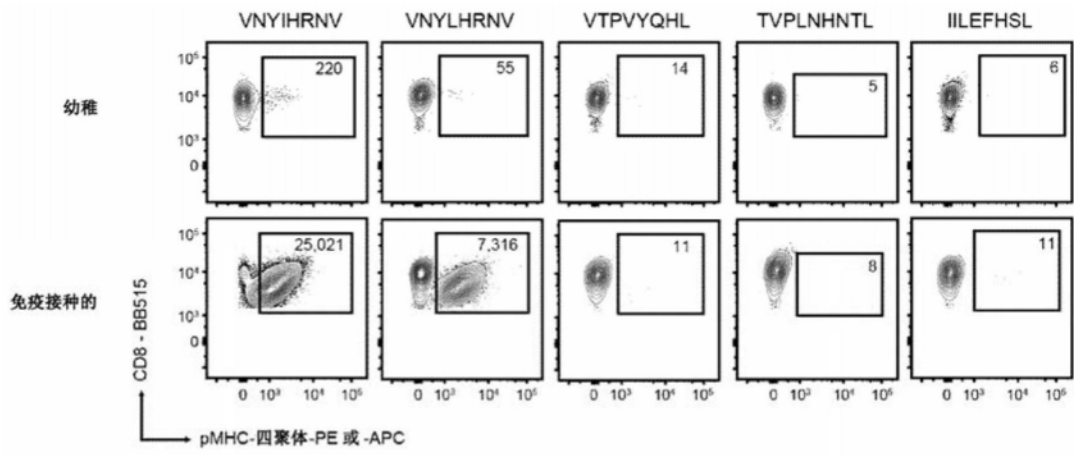


图9C

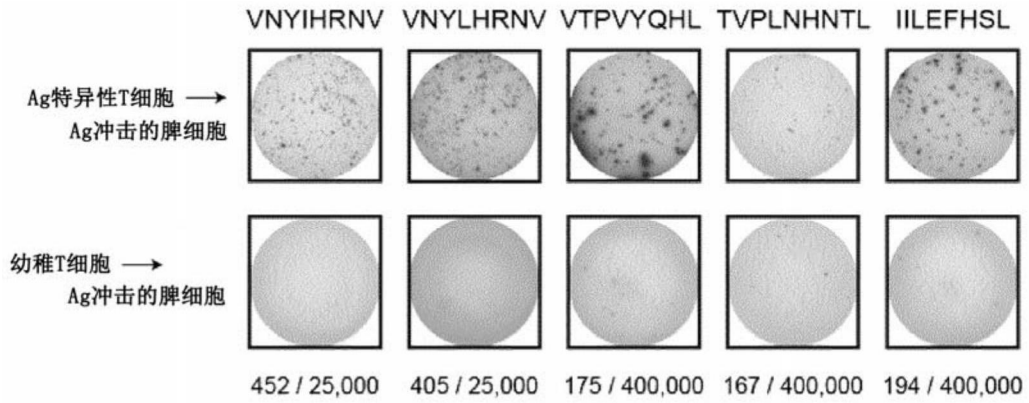


图9D

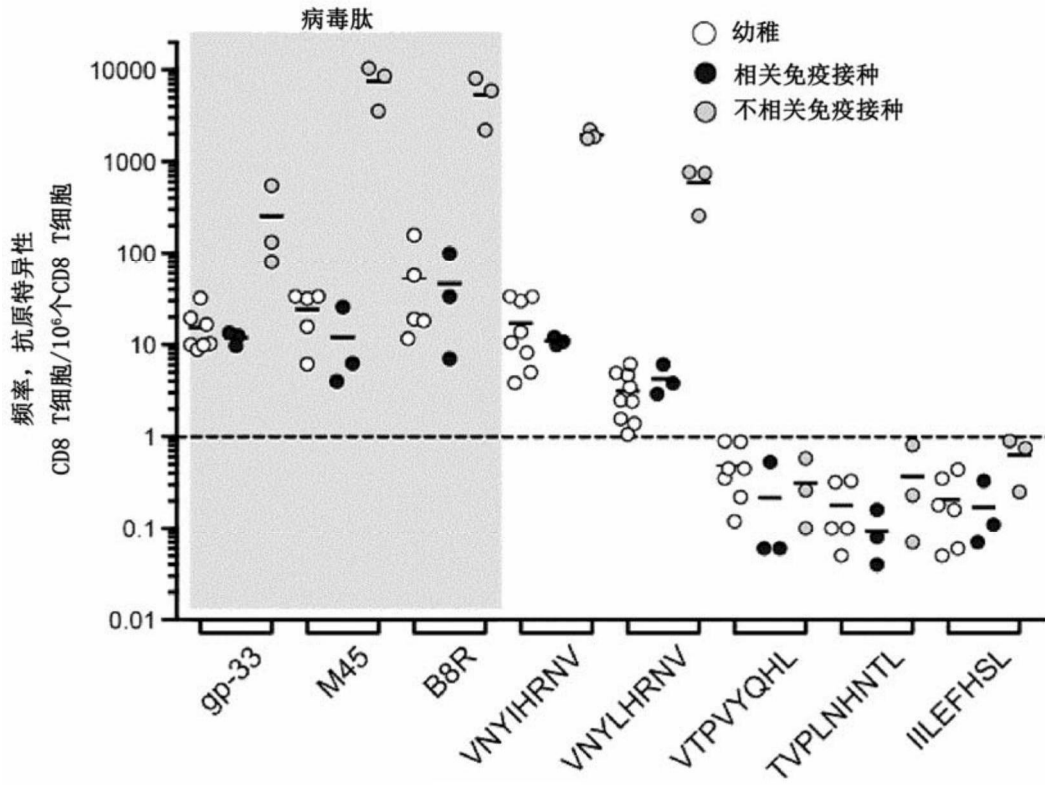


图10A

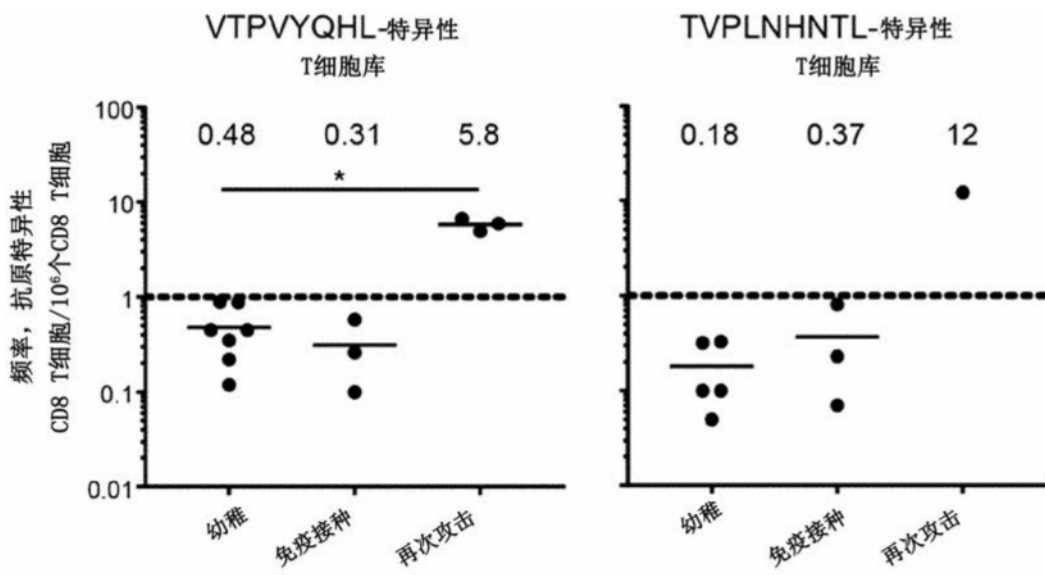


图10B

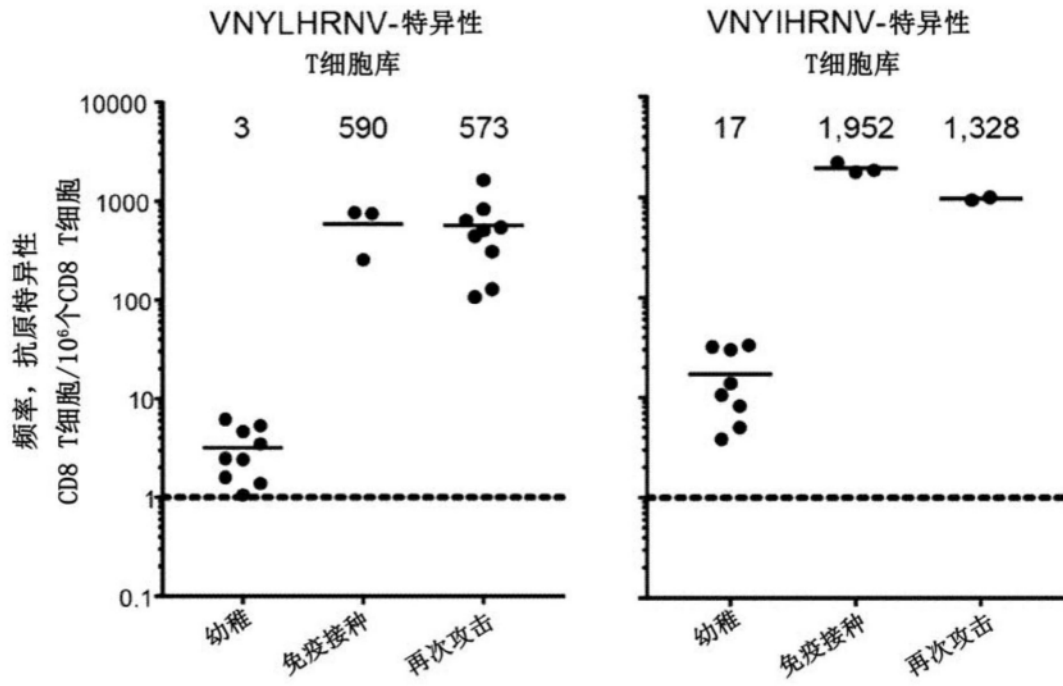


图10C

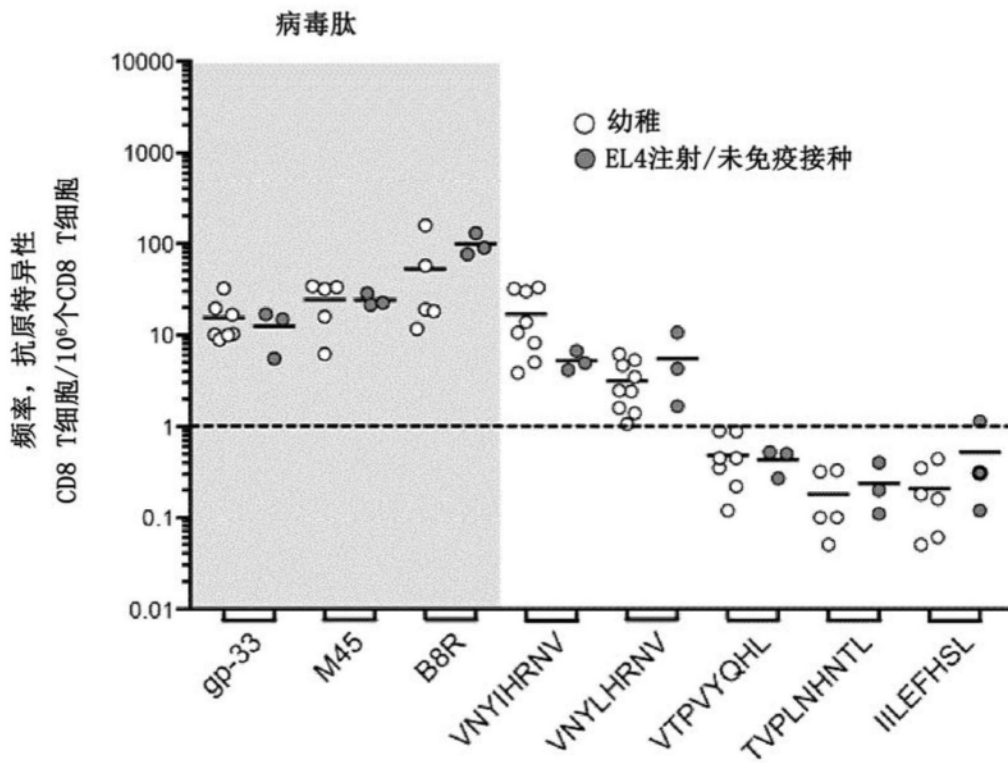


图10D

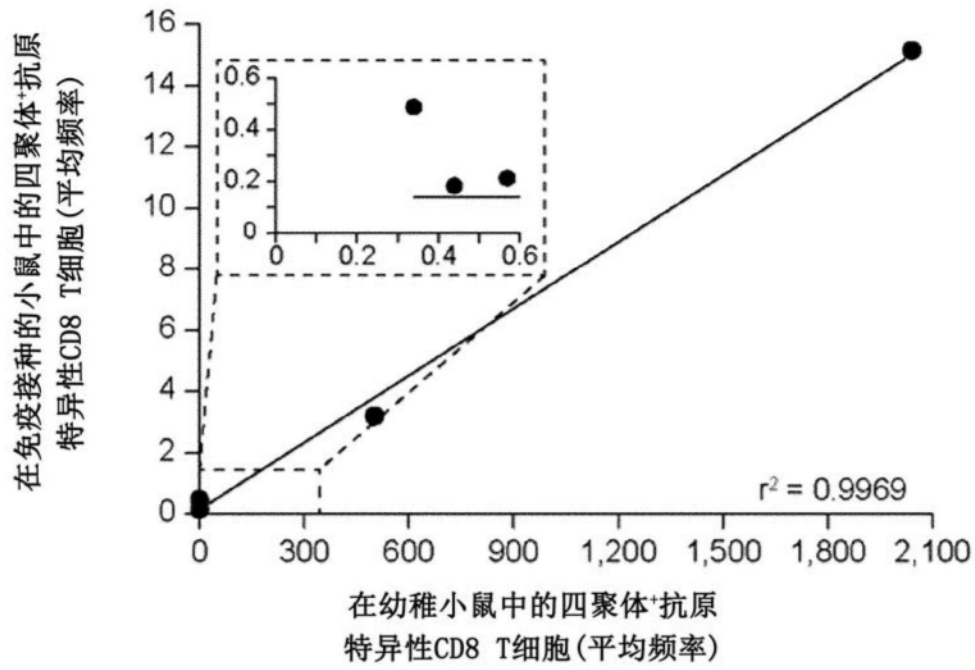


图11A

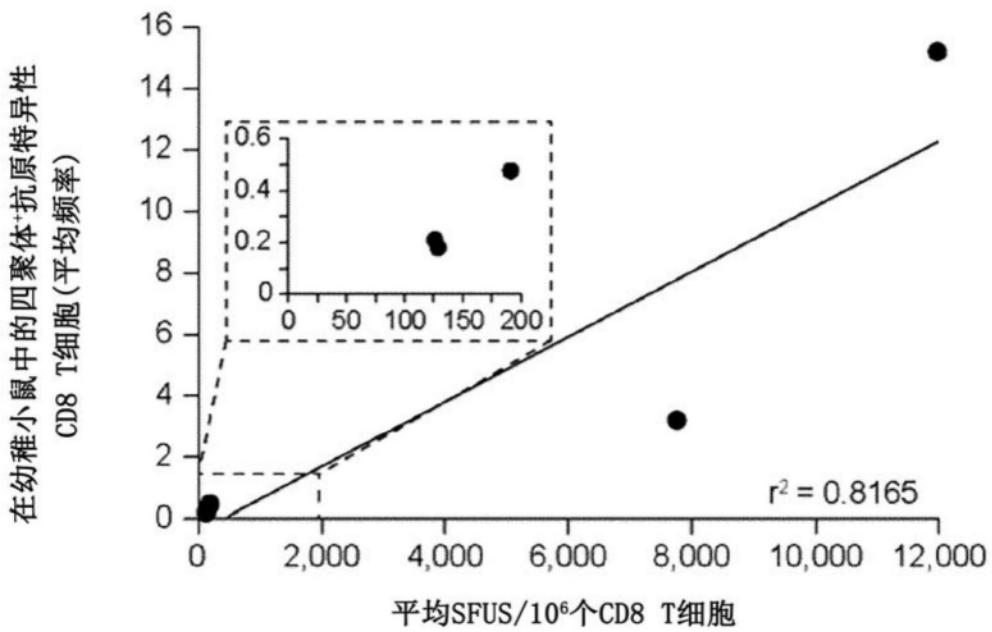


图11B

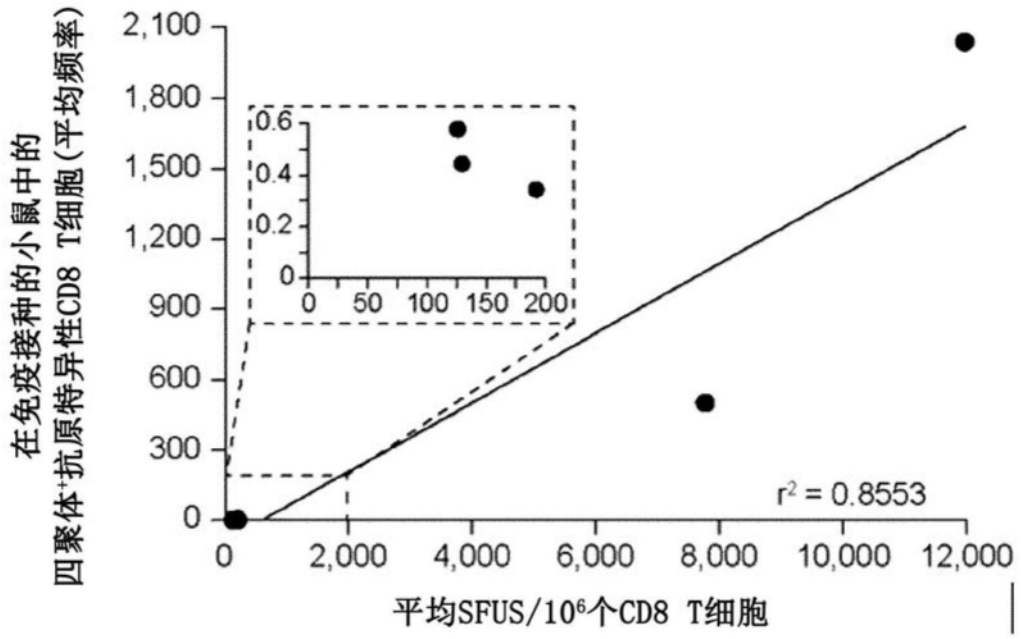


图11C

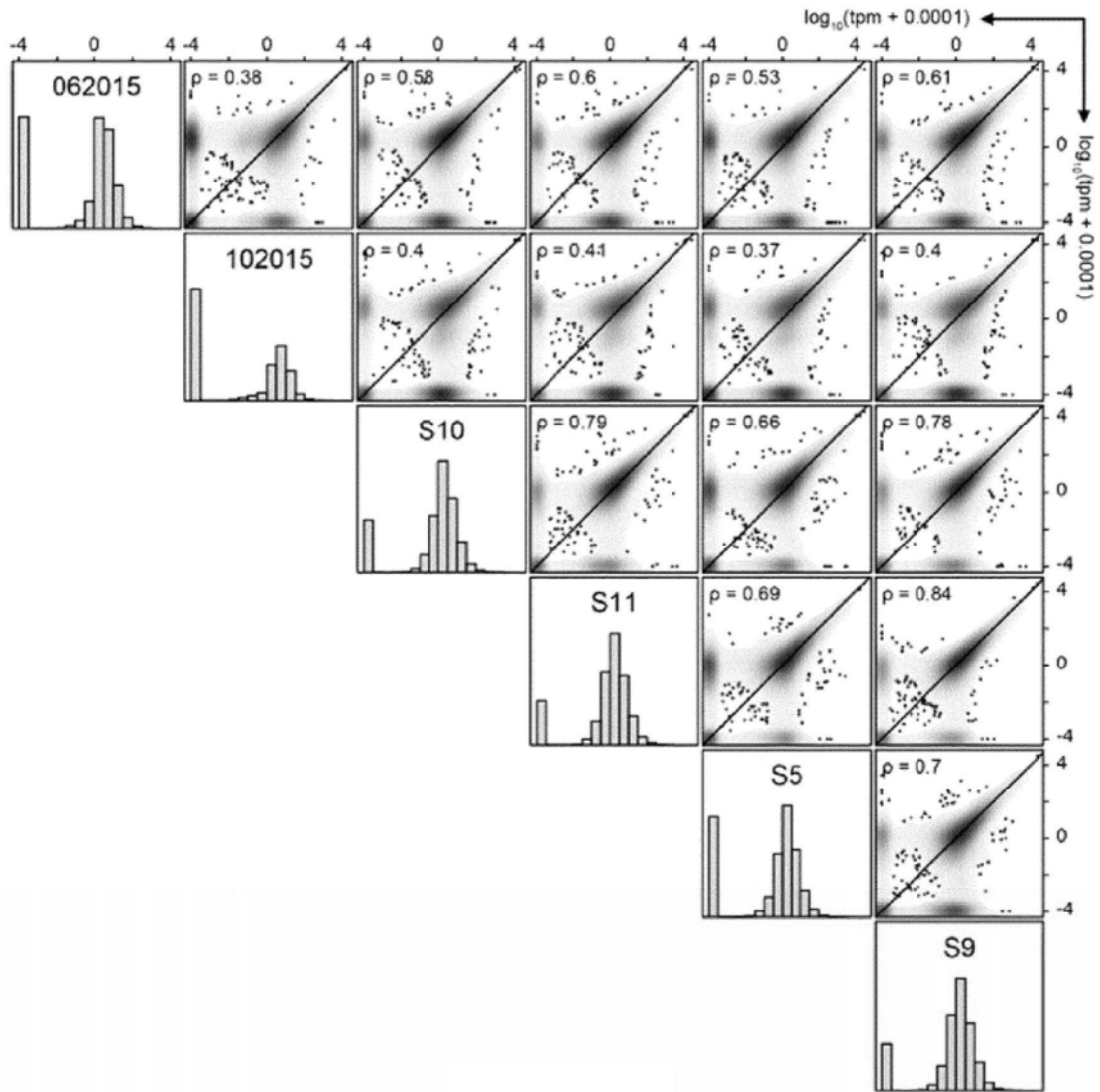


图12A

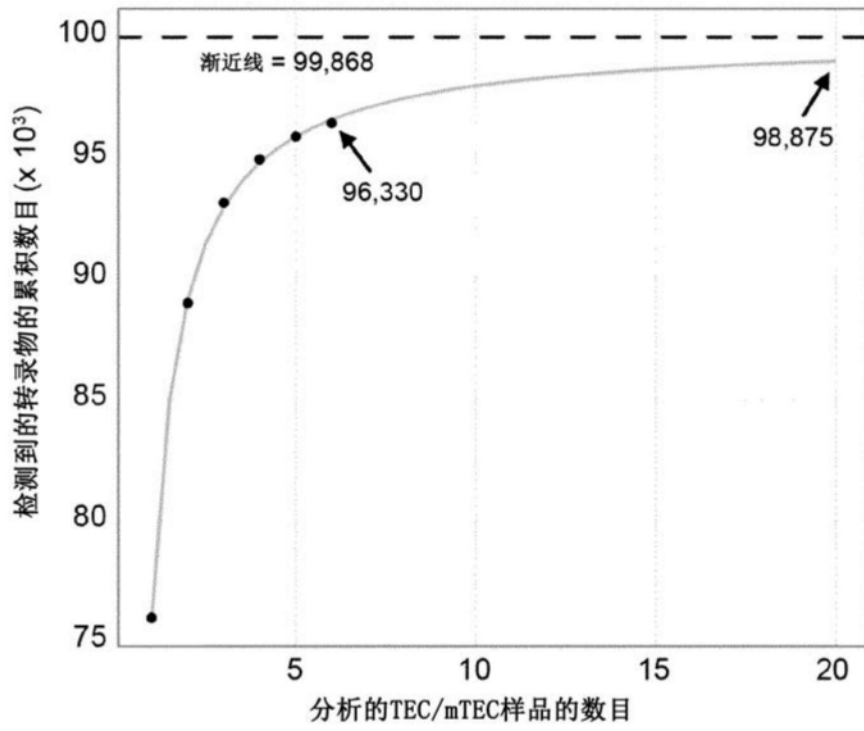


图12B

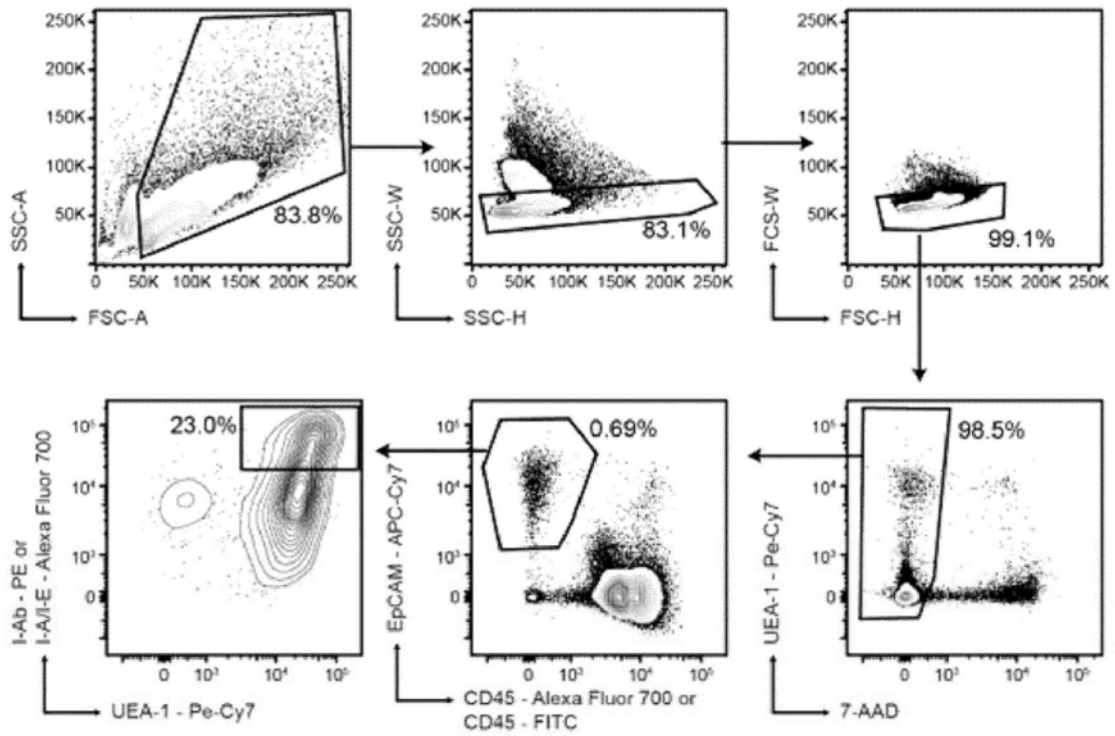


图13A

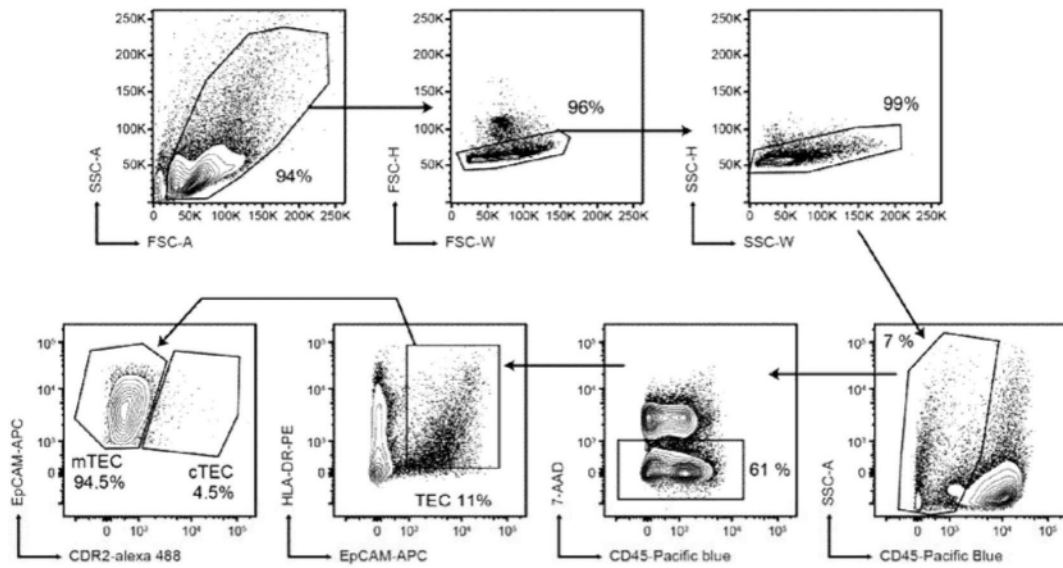


图13B

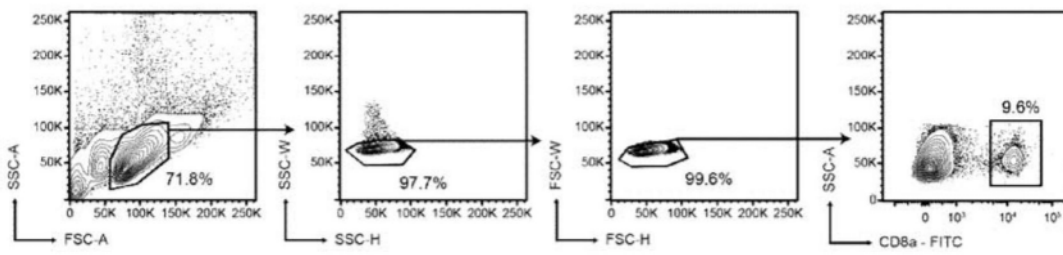


图13c