



(12) 发明专利申请

(10) 申请公布号 CN 118234745 A

(43) 申请公布日 2024. 06. 21

(21) 申请号 202280075075.9

(22) 申请日 2022.09.16

(30) 优先权数据

63/261,315 2021.09.17 US

(85) PCT国际申请进入国家阶段日

2024.05.10

(86) PCT国际申请的申请数据

PCT/CA2022/051377 2022.09.16

(87) PCT国际申请的公布数据

W02023/039673 EN 2023.03.23

(71) 申请人 蒙特利尔大学

地址 加拿大魁北克

(72) 发明人 克劳德·佩罗 皮埃尔·蒂伯尔

詹娜·克勒 马里-皮埃尔·哈迪

(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 刘书芝

(51) Int.Cl.

C07K 14/705 (2006.01)

A61K 38/20 (2006.01)

A61K 39/00 (2006.01)

A61P 35/00 (2006.01)

C07K 7/06 (2006.01)

C07K 14/00 (2006.01)

C12N 5/07 (2006.01)

C12N 5/078 (2006.01)

C12N 5/0783 (2006.01)

C12N 5/0784 (2006.01)

C12N 15/12 (2006.01)

C12Q 1/6809 (2006.01)

G01N 33/48 (2006.01)

G01N 33/68 (2006.01)

权利要求书5页 说明书72页

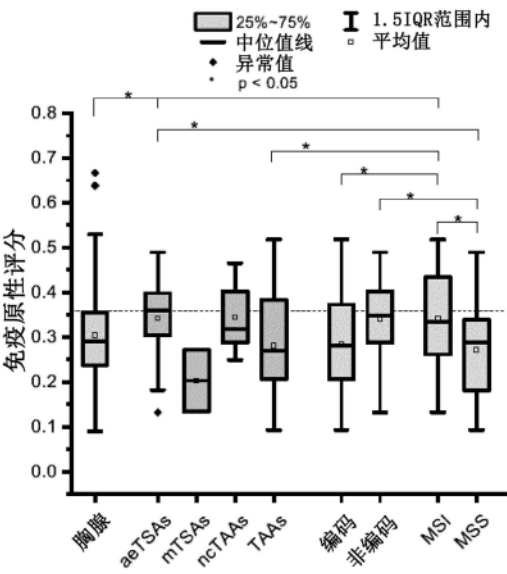
序列表(电子公布) 附图23页

(54) 发明名称

用于结直肠癌的新型肿瘤特异性抗原及其用途

(57) 摘要

结直肠癌(CRC)尚未从创新免疫疗法中获益,主要是因为缺乏可操作的免疫靶标。本文描述了由CRC细胞表达的新型肿瘤特异性抗原(TSA)和肿瘤相关抗原(TAA)。本文所述的大多数TSA源自异常表达的未突变基因组序列,诸如在正常组织中不表达的内含子和基因间序列。描述了来源于这些TSA的核酸、组合物、细胞、抗体和疫苗。还描述了TSA、核酸、组合物、抗体、细胞和疫苗用于治疗CRC的用途。



1. 一种肿瘤抗原肽 (TAP), 包括以下氨基酸序列中的一种或由以下氨基酸序列中的一种组成:

序列	SEQ ID NO:	序列	SEQ ID NO:
SIETVNSL	6	KIGEVIVTK	16
RMLLSHTGK	1	TRSTIILHL	17
LPHRALSGI	2	SILRGSLGK	18
GTNPTAAVK	3	RLVSNRVVR	19
LRHKLVLNR	4	SLKNLEPIK	20
RIGGVGVEK	5	RSAGPRAPL	21
TVNTQQYNTK	7	VLYRSVLLK	22
SVSHLHIFF	8	KAAGAGAAK	23
TTLENLPQK	9		
AQKLQVRI	10		
GQIELSIYR	11		
HGALSIRSI	12		
RLMKFLPV	13		
SLYISEERK	14		
VQTAVLNV	15		

或编码所述TAP的核酸。

2. 根据权利要求1所述的TAP或核酸, 其中, 所述TAP包括SEQ ID NO: 6、1-5和6-17中限定的序列中的一种。

3. 根据权利要求1或2所述的TAP或核酸, 其与HLA-A*02:01分子结合并包括SEQ ID NO: 6的序列。

4. 根据权利要求1或2所述的TAP或核酸, 其与HLA-A*03:01分子结合并包括SEQ ID NO:

1、11或14的序列。

5. 根据权利要求1或2所述的TAP或核酸,其与HLA-A*03:02分子结合并包括SEQ ID NO: 3、5、7、16或23的序列。

6. 根据权利要求1或2所述的TAP或核酸,其与HLA-A*11:01分子结合并包括SEQ ID NO: 9或18的序列。

7. 根据权利要求1或2所述的TAP或核酸,其与HLA-A*30:01分子结合并包括SEQ ID NO: 19、20或23的序列。

8. 根据权利要求1或2所述的TAP或核酸,其与HLA-A*32:01分子结合并包括SEQ ID NO: 8的序列。

9. 根据权利要求1或2所述的TAP或核酸,其与HLA-B*07:02分子结合并包括SEQ ID NO: 2或21的序列。

10. 根据权利要求1或2所述的TAP或核酸,其与HLA-B*13:02分子结合并包括SEQ ID NO: 13的序列。

11. 根据权利要求1或2所述的TAP或核酸,其与HLA-B*27:05分子结合并包括SEQ ID NO: 4的序列。

12. 根据权利要求1或2所述的TAP或核酸,其与HLA-B*52:01分子结合并包括SEQ ID NO: 10、12或15的序列。

13. 根据权利要求1或2所述的TAP或核酸,其与HLA-C*06:02分子结合并包括SEQ ID NO: 17的序列。

14. 根据权利要求1-13中任一项所述的TAP或核酸,其中,所述TAP由位于基因组的非蛋白编码区的序列编码。

15. 根据权利要求14所述的TAP或核酸,其中,所述基因组的非蛋白编码区是非翻译转录区(UTR)。

16. 根据权利要求14所述的TAP或核酸,其中,所述基因组的非蛋白编码区是内含子。

17. 根据权利要求14所述的TAP或核酸,其中,所述基因组的非蛋白编码区是基因间区。

18. 根据权利要求14所述的TAP或核酸,其中,所述基因组的非蛋白编码区是长非编码RNA。

19. 根据权利要求1至18中任一项所述的核酸,其中,所述核酸是mRNA。

20. 根据权利要求1至18中任一项所述的核酸,其中,所述核酸是DNA。

21. 根据权利要求1至20中任一项所述的核酸,其中,所述核酸是病毒载体的组分。

22. 一种组合,包括权利要求1-21中任一项所限定的TAP或核酸中的至少两种。

23. 一种合成肽(SLP),包括权利要求1所限定的氨基酸序列的至少一种。

24. 一种囊泡或颗粒,包括权利要求1至23中任一项所述的TAP、核酸、组合或SLP。

25. 根据权利要求24所述的囊泡或颗粒,其中,所述囊泡是脂质纳米颗粒(LNP)。

26. 根据权利要求24或25所述的囊泡或颗粒,其包括阳离子脂质。

27. 一种组合物,包括权利要求1至23中任一项所述的TAP、核酸、组合或SLP,或者权利要求24-26中任一项所述的囊泡或颗粒,以及药学上可接受的负载体。

28. 一种疫苗,包括权利要求1至23中任一项所述的TAP、核酸、组合或SLP,权利要求24-26中任一项所述的囊泡或颗粒,或者权利要求27所述的组合物,以及佐剂。

29. 一种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权利要求 1-18 中任一项所述的 TAP。

30. 根据权利要求 29 所述的分离的 MHC I 类分子, 所述分离的 MHC I 类分子为多聚体形式。

31. 根据权利要求 30 所述的分离的 MHC I 类分子, 其中, 所述多聚体是四聚体。

32. 一种分离的细胞, 包括 (i) 权利要求 1-18 中任一项所述的 TAP、(ii) 权利要求 19 所述的组合; (iii) 权利要求 23 所述的 SLP; 或 (iv) 包括编码权利要求 1-18 任一项所述的 TAP、权利要求 19 所述的组合或权利要求 23 所述的 SLP 的核苷酸序列的载体。

33. 一种分离的细胞, 所述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 I 类分子在其肽结合槽中包括权利要求 1-18 中任一项所述的 TAP 或权利要求 19 所述的组合。

34. 根据权利要求 33 所述的细胞, 所述细胞是抗原呈递细胞 (APC)。

35. 根据权利要求 34 所述的细胞, 其中, 所述 APC 是树突状细胞。

36. 一种 T 细胞受体 (TCR), 所述 T 细胞受体特异性识别权利要求 29-31 中任一项所述的分离的 MHC I 类分子和/或在权利要求 32-35 中任一项所述的细胞的表面表达的 MHC I 类分子。

37. 一种抗体或其抗原结合片段, 其特异性结合权利要求 29-31 中任一项所述的分离的 MHC I 类分子和/或在权利要求 33-35 中任一项所述的细胞的表面表达的 MHC I 类分子。

38. 根据权利要求 37 所述的抗体或其抗原结合片段, 其为双特异性抗体或其抗原结合片段。

39. 根据权利要求 38 所述的抗体或其抗原结合片段, 其中, 所述双特异性抗体或其抗原结合片段是单链双抗体 (scDb)。

40. 根据权利要求 38 或 39 所述的抗体或其抗原结合片段, 其中, 所述双特异性抗体或其抗原结合片段也特异性结合 T 细胞信号传导分子。

41. 根据权利要求 40 所述的抗体或其抗原结合片段, 其中, 所述 T 细胞信号传导分子是 CD3 链。

42. 一种分离的细胞, 所述分离的细胞在其细胞表面表达权利要求 36 所述的 TCR。

43. 根据权利要求 42 所述的分离的细胞, 所述分离的细胞是 CD8⁺T 淋巴细胞。

44. 一种细胞群, 所述细胞群包括至少 0.5% 的根据权利要求 42 或 43 所限定的分离的细胞。

45. 一种治疗受试者的结直肠癌的方法, 所述方法包括向所述受试者给药有效量的:

(a) TAP 或合成长肽 (SLP), 所述 TAP 包括 SEQ ID NO: 1-23 和 25-50 中限定的氨基酸序列中的一种或者由 SEQ ID NO: 1-23 和 25-50 中限定的氨基酸序列中的一种组成, 所述合成长肽包括 SEQ ID NO: 1-23 和 25-50 中列出的序列的至少一种;

(b) 至少一种核酸, 所述至少一种核酸编码 (a) 中限定的所述 TAP、其组合或 SLP;

(c) 囊泡或颗粒, 所述囊泡或颗粒包括 (a) 中限定的所述 TAP、其组合或 SLP 或者 (b) 中限定的所述至少一种核酸;

(d) 组合物, 所述组合物包括 (a) 中限定的所述 TAP、其组合或 SLP, (b) 中限定的所述至少一种核酸, 或 (c) 中限定的所述囊泡或颗粒, 以及药学上可接受的负载体;

(e) 疫苗,所述疫苗包括 (a) 中限定的所述TAP、其组合或SLP, (b) 中限定的所述至少一种核酸, (c) 中限定的所述囊泡或颗粒,或者 (d) 中限定的所述组合物,以及佐剂;

(f) 细胞,所述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I类分子在其肽结合槽中包括 (a) 中限定的所述TAP或其组合;

(g) 细胞,所述细胞在其细胞表面表达T细胞受体 (TCR),所述T细胞受体特异性识别在 (f) 中限定的所述细胞表面表达的MHC I类分子;或者

(h) 可溶性TCR、抗体或其抗原结合片段,其特异性结合在 (f) 中限定的细胞表面表达的所述MHC I类分子。

46. 根据权利要求45所述的方法,其中,所述TAP或核酸根据权利要求1至21中任一项所限定,所述组合物根据权利要求22所限定,所述SLP根据权利要求23所限定;所述囊泡根据权利要求24-26中任一项所限定,所述组合物根据权利要求27所限定,所述疫苗根据权利要求28所限定,所述细胞根据权利要求32-35、42和43中任一项所限定,所述细胞群根据权利要求44所述限定,和/或所述抗体或抗原结合片段根据权利要求37-41中任一项所限定。

47. 根据权利要求45或46所述的方法,其中,所述CRC是结肠癌。

48. 根据权利要求45或46所述的方法,其中,所述CRC是直肠癌。

49. 根据权利要求45-48中任一项所述的方法,进一步包括向所述受试者施用至少一种额外的抗肿瘤剂或疗法。

50. 根据权利要求49所述的方法,其中,所述至少一种额外的抗肿瘤剂或疗法是化疗剂、免疫疗法、免疫检查点抑制剂、放射疗法或外科手术。

51. 以下:

(a) TAP或合成长肽 (SLP),所述TAP包括SEQ ID NO:1-23和25-50中限定的氨基酸序列中的一种或者由SEQ ID NO:1-23和25-50中限定的氨基酸序列中的一种组成,所述合成长肽包括SEQ ID NO:1-23和25-50中列出的序列的至少一种;

(b) 至少一种核酸,所述至少一种核酸编码 (a) 中限定的所述TAP、其组合或SLP;

(c) 囊泡或颗粒,所述囊泡或颗粒包括 (a) 中限定的所述TAP、其组合或SLP或者 (b) 中限定的所述至少一种核酸;

(d) 组合物,所述组合物包括 (a) 中限定的所述TAP、其组合或SLP, (b) 中限定的所述至少一种核酸,或 (c) 中限定的所述囊泡或颗粒,以及药学上可接受的负载体;

(e) 疫苗,所述疫苗包括 (a) 中限定的所述TAP、其组合或SLP, (b) 中限定的所述至少一种核酸, (c) 中限定的所述囊泡或颗粒,或者 (d) 中限定的所述组合物,以及佐剂;

(f) 细胞,所述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I类分子在其肽结合槽中包括 (a) 中限定的所述TAP或其组合;

(g) 细胞,所述细胞在其细胞表面表达T细胞受体 (TCR),所述T细胞受体特异性识别在 (f) 中限定的所述细胞表面表达的MHC I类分子;或者

(h) 可溶性TCR、抗体或其抗原结合片段,其特异性结合在 (f) 中限定的所述细胞的表面表达的所述MHC I类分子;

用于治疗受试者的结直肠癌,或用于制备用于治疗受试者的结直肠癌的药物的用途。

52. 根据权利要求55所述的用途,其中,所述TAP或核酸根据权利要求1至21中任一项所限定,所述组合根据权利要求22所限定,所述SLP根据权利要求23所限定;所述囊泡根据权

利要求24-26中任一项所限定,所述组合物根据权利要求27所限定,所述疫苗根据权利要求28所限定,所述细胞根据权利要求32-35、42和43中任一项所限定,所述细胞群根据权利要求44所述限定,和/或所述抗体或抗原结合片段根据权利要求37-41中任一项所限定。

53. 根据权利要求51或52所述的用途,其中,所述CRC是结肠癌。

54. 根据权利要求51或52所述的用途,其中,所述CRC是直肠癌。

55. 根据权利要求51-54中任一项所述的用途,进一步包括向所述受试者使用至少一种额外的抗肿瘤剂或疗法。

56. 根据权利要求55所述的用途,其中所述至少一种额外的抗肿瘤剂或疗法是化疗剂、免疫疗法、免疫检查点抑制剂、放射疗法或外科手术。

57. 一种剂,用于治疗受试者的结直肠癌的用途,其中,所述剂是:

(a) TAP或合成长肽(SLP),所述TAP包括SEQ ID NO:1-23和25-50中限定的氨基酸序列中的一种或者由SEQ ID NO:1-23和25-50中限定的氨基酸序列中的一种组成,所述合成长肽包括SEQ ID NO:1-23和25-50中列出的序列的至少一种;

(b) 至少一种核酸,所述至少一种核酸编码(a)中限定的所述TAP、其组合或SLP;

(c) 囊泡或颗粒,所述囊泡或颗粒包括(a)中限定的所述TAP、其组合或SLP或者(b)中限定的所述至少一种核酸;

(d) 组合物,所述组合物包括(a)中限定的所述TAP、其组合或SLP, (b)中限定的所述至少一种核酸,或(c)中限定的所述囊泡或颗粒,以及药学上可接受的负载体;

(e) 疫苗,所述疫苗包括(a)中限定的所述TAP、其组合或SLP, (b)中限定的所述至少一种核酸, (c)中限定的所述囊泡或颗粒,或者(d)中限定的所述组合物,以及佐剂;

(f) 细胞,所述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I类分子在其肽结合槽中包括(a)中限定的所述TAP或其组合;

(g) 细胞,所述细胞在其细胞表面表达T细胞受体(TCR),所述T细胞受体特异性识别在(f)中限定的所述细胞表面表达的MHC I类分子;或者

(h) 可溶性TCR、抗体或其抗原结合片段,其特异性结合在(f)中限定的细胞表面表达的所述MHC I类分子。

58. 根据权利要求61所述用途的剂,其中,所述TAP或核酸根据权利要求1至21中任一项所限定,所述组合根据权利要求22所限定,所述SLP根据权利要求23所限定;所述囊泡根据权利要求24-26中任一项所限定,所述组合物根据权利要求27所限定,所述疫苗根据权利要求28所限定,所述细胞根据权利要求32-35、42和43中任一项所限定,所述细胞群根据权利要求44所述限定,和/或所述抗体或抗原结合片段根据权利要求37-41中任一项所限定。

59. 根据权利要求57或58所述用途的剂,其中,所述CRC是结肠癌。

60. 根据权利要求57或58所述用途的剂,其中,所述CRC是直肠癌。

61. 根据权利要求57-60中任一项所述用途的剂,进一步包括向所述受试者使用至少一种额外的抗肿瘤剂或疗法。

62. 根据权利要求61所述用途的剂,其中,所述至少一种额外的抗肿瘤剂或疗法是化疗剂、免疫疗法、免疫检查点抑制剂、放射疗法或外科手术。

用于结直肠癌的新型肿瘤特异性抗原及其用途

[0001] 相关申请的交叉引用

[0002] 本申请要求于2021年9月17日提交的美国临时专利申请号63/261,315的权益,其通过援引并入本文。

技术领域

[0003] 本公开内容总体上涉及肿瘤学领域,并且更具体地涉及对癌症诸如结直肠癌的治疗。

背景技术

[0004] 结直肠癌(CRC)是全球第三大最常诊断的癌症和第二大引起癌症死亡原因,仅在2018年估计就有超过180万例病例和881,000例死亡(1)。随着全球社会经济的变化,CDC的发生率预计增加,其中预计到2030年,每年将有220万例病例和110万例死亡(1,2)。这种显著的疾病负担突显了开发针对这种疾病的新的且有效治疗的必要性。

[0005] 结肠癌和直肠癌两者中肿瘤浸润淋巴细胞(TIL)的丰度与总生存率的增加之间呈正相关,这表明T细胞可以识别这些肿瘤中生物学相关的肿瘤抗原(3,4)。这些抗原的潜在免疫原性使免疫检查点抑制(ICI)成为癌症患者的有前景的治疗;然而,评价其在CDC中功效的早期临床试验产生了喜忧参半的结果。结直肠肿瘤的特征在于错配修复蛋白的缺陷导致重复DNA序列(微卫星)的积累,称为微卫星不稳定性(MSI),在抗PD1治疗的II期临床试验中表现出相对的成功(5)。相比之下,在针对不具有高突变负荷的微卫星稳定(MSS)肿瘤的临床试验中,这种治疗具有非常低的功效,其占CRC病例的大约80%(5,6)。

[0006] 鉴于免疫应答在CRC中的重要性以及ICI单独取得的有限成功,近年来一种有前景的研究途径是基于新生抗原的疫苗或基于T细胞受体的疗法,其可以与ICI一起使用或不与ICI一起使用,并将理想地弥合MSI和MSS肿瘤之间的治疗功效差距。例如,肿瘤相关抗原(TAA),即与正常细胞相比在癌细胞中过度表达的抗原,先前已在CRC中被识别(7,8)。虽然若干种TAA已经在针对CRC的疫苗和I期试验中进行了测试,但大多数都获得了“有限的成功”,这可能是由于胸腺对此类抗原的负向选择(9)。在一项研究中,用经基因工程化的抗CEA T细胞治疗转移性CRC导致一名患者肿瘤消退,但所有患者均出现“严重炎性结肠炎”,这表明不利的自身免疫应答是使用TAA的另一个可能的负面后果(10)。

[0007] 由于对TAA的应答不一,有效的基于新生抗原的疗法更有可能利用肿瘤特异性抗原(TSA),其可以通过遗传、表观遗传和翻译后变异生成,包括但不限于单核苷酸变体、异常表达的转录物或新的剪接事件,并且仅由肿瘤表达(11)。CRC中单核苷酸变体、剪接变体和INDEL突变的高患病率表明由肿瘤的主要组织相容性复合体(MHC)呈递独特抗原的可能性很高,是得有可能引起肿瘤特异性免疫应答(12)。最近在CRC中识别出了TSA,并证实在I期和II期疫苗试验中取得了一些成功。一项2015年使用源自MSI高肿瘤的移码抗原的疫苗试验表明,所有患者都产生了显著且特异性的免疫应答(13)。然而,由于这项研究使用了来源于微卫星不稳定性移码的抗原,这些发现不适用于大多数CRC患者。迄今为止,在CRC中识别

TSA的其他研究仅集中在来自基因组编码区的突变体TSA上(13,14)。一项MSS CRC类器官的调查显示,只有0.5%的非沉默突变体被识别为mTSA;这一比例显著低于由HLA结合预测软件所预测的比例(15)。另外,先前的研究没有采用质谱(MS)技术来定量肿瘤细胞上那些TSA的表达,这是可能影响靶向给定TSA的治疗潜力的信息(13,14)。

[0008] 鉴于此,迫切需要确定可再次引发针对CRC的治疗性免疫应答的肿瘤特异性抗原。此类抗原可用作疫苗(±免疫检查点抑制剂)或作为基于T细胞受体的方法(细胞疗法、双特异性生物制剂)的靶标。

[0009] 本说明书参考了许多文献,其内容通过引用以其全文并入本文。

发明内容

[0010] 在各个方面和实施方式中,本公开提供了以下项目1至62项:

[0011] 1.一种肿瘤抗原肽(TAP),包括以下氨基酸序列中的一种或由以下氨基酸序列中的一种组成:

[0012]	序列	SEQ ID NO:	序列	SEQ ID NO:
	SIHETVNSL	6	KIGEVIVTK	16
	RMLLSHTGK	1	TRSTIILHL	17
	LPHRALSGI	2	SILRGSLGK	18
	GTNPAAVK	3	RLVSNRVVR	19
	LRHKLVLNR	4	SLKNLEPIK	20
	RIGGVGVEK	5	RSAGPRAPL	21
	TVNTQQYNTK	7	VLYRSVLLLK	22
	SVSHLHIF	8	KAAGAGAAK	23
	TTLENLPQK	9		
	AQKLQVRI	10		
	GQIELSIYR	11		
	HGALSIRSI	12		
	RLMKFLPV	13		
	SLYISEERK	14		
	VQTAVLNV	15		

[0013] 或编码所述TAP的核酸。

- [0014] 2. 根据项1所述的TAP或核酸,其中,所述TAP包括SEQ ID NO:6、1-5和6-17中限定的序列中的一种。
- [0015] 3. 根据项1或2所述的TAP或核酸,其与HLA-A*02:01分子结合并包括SEQ ID NO:6的序列。
- [0016] 4. 根据项1或2所述的TAP或核酸,其与HLA-A*03:01分子结合并包括SEQ ID NO:1、11或14的序列。
- [0017] 5. 根据项1或2所述的TAP或核酸,其与HLA-A*03:02分子结合并包括SEQ ID NO:3、5、7、16或23的序列。
- [0018] 6. 根据项1或2所述的TAP或核酸,其与HLA-A*11:01分子结合并包括SEQ ID NO:9或18的序列。
- [0019] 7. 根据项1或2所述的TAP或核酸,其与HLA-A*30:01分子结合并包括SEQ ID NO:19、20或23的序列。
- [0020] 8. 根据项1或2所述的TAP或核酸,其与HLA-A*32:01分子结合并包括SEQ ID NO:8的序列。
- [0021] 9. 根据项1或2所述的TAP或核酸,其与HLA-B*07:02分子结合并包括SEQ ID NO:2或21的序列。
- [0022] 10. 根据项1或2所述的TAP或核酸,其与HLA-B*13:02分子结合并包括SEQ ID NO:13的序列。
- [0023] 11. 根据项1或2所述的TAP或核酸,其与HLA-B*27:05分子结合并包括SEQ ID NO:4的序列。
- [0024] 12. 根据项1或2所述的TAP或核酸,其与HLA-B*52:01分子结合并包括SEQ ID NO:10、12或15的序列。
- [0025] 13. 根据项1或2所述的TAP或核酸,其与HLA-C*06:02分子结合并包括SEQ ID NO:17的序列。
- [0026] 14. 根据项1-13中任一项所述的TAP或核酸,其中,所述TAP由位于基因组的非蛋白编码区的序列编码。
- [0027] 15. 根据项14所述的TAP或核酸,其中,所述基因组的非蛋白编码区是非翻译转录区(UTR)。
- [0028] 16. 根据项14所述的TAP或核酸,其中,所述基因组的非蛋白编码区是内含子。
- [0029] 17. 根据项14所述的TAP或核酸,其中,所述基因组的非蛋白编码区是基因间区。
- [0030] 18. 根据项14所述的TAP或核酸,其中,所述基因组的非蛋白编码区是长非编码RNA。
- [0031] 19. 根据项1至18中任一项所述的核酸,其中,所述核酸是mRNA。
- [0032] 20. 根据项1至18中任一项所述的核酸,其中,所述核酸是DNA。
- [0033] 21. 根据项1至20中任一项所述的核酸,其中,所述核酸是病毒载体的组分。
- [0034] 22. 一种组合,包括项1-21中任一项所限定的TAP或核酸中的至少两种。
- [0035] 23. 一种合成长肽(SLP),包括项1所限定的氨基酸序列的至少一种。
- [0036] 24. 一种囊泡或颗粒,包括项1至23中任一项所述的TAP、核酸、组合或SLP。
- [0037] 25. 根据项24所述的囊泡或颗粒,其中,所述囊泡是脂质纳米颗粒(LNP)。

- [0038] 26. 根据项24或25所述的囊泡或颗粒,其包括阳离子脂质。
- [0039] 27. 一种组合物,包括项1至23中任一项所述的TAP、核酸、组合或SLP,或者项24-26中任一项所述的囊泡或颗粒,以及药学上可接受的负载体。
- [0040] 28. 一种疫苗,包括项1至23中任一项所述的TAP、核酸、组合或SLP,或者项24-26中任一项所述的囊泡或颗粒,或者项27所述的组合物,以及佐剂。
- [0041] 29. 一种分离的主要组织相容性复合体 (MHC) I类分子,在其肽结合槽 (groove) 中包括项1-18中任一项所述的TAP。
- [0042] 30. 根据项29所述的分离的MHC I类分子,其为多聚体形式。
- [0043] 31. 根据项30所述的分离的MHC I类分子,其中,所述多聚体是四聚体。
- [0044] 32. 一种分离的细胞,包括 (i) 项1-18中任一项所述的TAP、(ii) 项19所述的组合; (iii) 项23所述的SLP; 或 (iv) 包括编码项1-18任一项所述的TAP、项19所述的组合或项23所述的SLP的核苷酸序列的载体。
- [0045] 33. 一种分离的细胞,所述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I类分子在其肽结合槽中包括项1-18中任一项所述的TAP或项19所述的组合。
- [0046] 34. 根据项33所述的细胞,其是抗原呈递细胞 (APC)。
- [0047] 35. 根据项34所述的细胞,其中,所述APC是树突状细胞。
- [0048] 36. 一种T细胞受体 (TCR),其特异性识别项29-31中任一项所述的分离的MHC I类分子和/或在项32-35中任一项所述的细胞的表面表达的MHC I类分子。
- [0049] 37. 一种抗体或其抗原结合片段,其特异性结合项29-31中任一项所述的分离的MHC I类分子和/或在项33-35中任一项所述的细胞的表面表达的MHC I类分子。
- [0050] 38. 根据项37所述的抗体或其抗原结合片段,其为双特异性抗体或其抗原结合片段。
- [0051] 39. 根据项38所述的抗体或其抗原结合片段,其中,所述双特异性抗体或其抗原结合片段是单链双抗体 (scDb)。
- [0052] 40. 根据项38或39所述的抗体或其抗原结合片段,其中,所述双特异性抗体或其抗原结合片段也特异性结合T细胞信号传导分子。
- [0053] 41. 根据项40所述的抗体或其抗原结合片段,其中,所述T细胞信号传导分子是CD3链。
- [0054] 42. 一种分离的细胞,所述分离的细胞在其细胞表面表达项36所述的TCR。
- [0055] 43. 根据项42所述的分离的细胞,所述分离的细胞是CD8⁺T淋巴细胞。
- [0056] 44. 一种细胞群,所述细胞群包括至少0.5%的根据项42或43所限定的分离的细胞。
- [0057] 45. 一种治疗受试者的癌症诸如结直肠癌的方法,所述方法包括向所述受试者给药有效量的:
- [0058] (a) TAP或合成长肽 (SLP),所述TAP包括SEQ ID NO:1-23和25-50中限定的氨基酸序列中的一种或者由SEQ ID NO:1-23和25-50中限定的氨基酸序列中的一种组成,所述合成长肽包括SEQ ID NO:1-23和25-50中列出的序列的至少一种;
- [0059] (b) 至少一种核酸,所述至少一种核酸编码 (a) 中限定的所述TAP、其组合或SLP;

[0060] (c) 囊泡或颗粒,所述囊泡或颗粒包括 (a) 中限定的所述TAP、其组合或SLP或者 (b) 中限定的所述至少一种核酸;

[0061] (d) 组合物,所述组合物包括 (a) 中限定的所述TAP、其组合或SLP, (b) 中限定的所述至少一种核酸,或 (c) 中限定的所述囊泡或颗粒,以及药学上可接受的负载体;

[0062] (e) 疫苗,所述疫苗包括 (a) 中限定的所述TAP、其组合或SLP, (b) 中限定的所述至少一种核酸, (c) 中限定的所述囊泡或颗粒,或者 (d) 中限定的所述组合物,以及佐剂;

[0063] (f) 细胞,所述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I类分子在其肽结合槽中包括 (a) 中限定的所述TAP或其组合;

[0064] (g) 细胞,所述细胞在其细胞表面表达T细胞受体 (TCR),所述T细胞受体特异性识别在 (f) 中限定的所述细胞表面表达的MHC I类分子;或者

[0065] (h) 可溶性TCR、抗体或其抗原结合片段,其特异性结合在 (f) 中限定的所述细胞表面表达的所述MHC I类分子。

[0066] 46. 根据项45所述的方法,其中,所述TAP或核酸根据项1至21中任一项所限定,所述组合根据项22所限定,所述SLP根据项23所限定;所述囊泡根据项24-26中任一项所限定,所述组合物根据项27所限定,所述疫苗根据项28所限定,所述细胞根据项32-35、42和43中任一项所限定,所述细胞群根据项44所限定,和/或所述抗体或抗原结合片段根据项37-41中任一项所限定。

[0067] 47. 根据项45或46所述的方法,其中,所述CRC是结肠癌。

[0068] 48. 根据项45或46所述的方法,其中,所述CRC是直肠癌。

[0069] 49. 根据项45-48中任一项所述的方法,进一步包括向所述受试者施用至少一种额外的抗肿瘤剂或疗法。

[0070] 50. 根据项49所述的方法,其中,所述至少一种额外的抗肿瘤剂或疗法是化疗剂、免疫疗法、免疫检查点抑制剂、放射疗法或外科手术。

[0071] 51. 以下:

[0072] (a) TAP或合成长肽 (SLP),所述TAP包括SEQ ID NO:1-23和25-50中限定的氨基酸序列中的一种或者由SEQ ID NO:1-23和25-50中限定的氨基酸序列中的一种组成,所述合成长肽包括SEQ ID NO:1-23和25-50中列出的序列的至少一种;

[0073] (b) 至少一种核酸,所述至少一种核酸编码 (a) 中限定的所述TAP、其组合或SLP;

[0074] (c) 囊泡或颗粒,所述囊泡或颗粒包括 (a) 中限定的所述TAP、其组合或SLP或者 (b) 中限定的所述至少一种核酸;

[0075] (d) 组合物,所述组合物包括 (a) 中限定的所述TAP、其组合或SLP, (b) 中限定的所述至少一种核酸,或 (c) 中限定的所述囊泡或颗粒,以及药学上可接受的负载体;

[0076] (e) 疫苗,所述疫苗包括 (a) 中限定的所述TAP、其组合或SLP, (b) 中限定的所述至少一种核酸, (c) 中限定的所述囊泡或颗粒,或者 (d) 中限定的所述组合物,以及佐剂;

[0077] (f) 细胞,所述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I类分子在其肽结合槽中包括 (a) 中限定的所述TAP或其组合;

[0078] (g) 细胞,所述细胞在其细胞表面表达T细胞受体 (TCR),所述T细胞受体特异性识别在 (f) 中限定的所述细胞表面表达的MHC I类分子;或者

[0079] (h) 可溶性TCR、抗体或其抗原结合片段,其特异性结合在 (f) 中限定的所述细胞的

表面表达的所述MHC I类分子；

[0080] 用于治疗受试者的癌症诸如结直肠癌,或用于制备用于治疗受试者的癌症诸如结直肠癌的药物的用途。

[0081] 52.根据项55所述的用途,其中,所述TAP或核酸根据项1至21中任一项所限定,所述组合物根据项22所限定,所述SLP根据项23所限定;所述囊泡根据项24-26中任一项所限定,所述组合物根据项27所限定,所述疫苗根据项28所限定,所述细胞根据项32-35、42和43中任一项所限定,所述细胞群根据项44所限定,和/或所述抗体或抗原结合片段根据项37-41中任一项所限定。

[0082] 53.根据项51或52所述的用途,其中,所述CRC是结肠癌。

[0083] 54.根据项51或52所述的用途,其中,所述CRC是直肠癌。

[0084] 55.根据项51-54中任一项所述的用途,进一步包括对所述受试者使用至少一种额外的抗肿瘤剂或疗法。

[0085] 56.根据项55所述的用途,其中,所述至少一种额外的抗肿瘤剂或疗法是化疗剂、免疫疗法、免疫检查点抑制剂、放射疗法或外科手术。

[0086] 57.一种剂,用于治疗受试者的癌症诸如结直肠癌的用途,其中,所述剂是:

[0087] (a) TAP或合成长肽 (SLP),所述TAP包括SEQ ID NO:1-23和25-50中限定的氨基酸序列中的一种或者由SEQ ID NO:1-23和25-50中限定的氨基酸序列中的一种组成,所述合成长肽包括SEQ ID NO:1-23和25-50中列出的序列的至少一种;

[0088] (b) 至少一种核酸,所述至少一种核酸编码(a)中限定的所述TAP、其组合或SLP;

[0089] (c) 囊泡或颗粒,所述囊泡或颗粒包括(a)中限定的所述TAP、其组合或SLP或者(b)中限定的所述至少一种核酸;

[0090] (d) 组合物,所述组合物包括(a)中限定的所述TAP、其组合或SLP,(b)中限定的所述至少一种核酸,或(c)中限定的所述囊泡或颗粒,以及药学上可接受的负载体;

[0091] (e) 疫苗,所述疫苗包括(a)中限定的所述TAP、其组合或SLP,(b)中限定的所述至少一种核酸,(c)中限定的所述囊泡或颗粒,或者(d)中限定的所述组合物,以及佐剂;

[0092] (f) 细胞,所述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I类分子在其肽结合槽中包括(a)中限定的所述TAP或其组合;

[0093] (g) 细胞,所述细胞在其细胞表面表达T细胞受体(TCR),所述T细胞受体特异性识别在(f)中限定的所述细胞表面表达的MHC I类分子;或者

[0094] (h) 可溶性TCR、抗体或其抗原结合片段,其特异性结合在(f)中限定的所述细胞表面表达的所述MHC I类分子。

[0095] 58.根据项61所述用途的剂,其中,所述TAP或核酸根据项1至21中任一项所限定,所述组合根据项22所限定,所述SLP根据项23所限定;所述囊泡根据项24-26中任一项所限定,所述组合物根据项27所限定,所述疫苗根据项28所限定,所述细胞根据项32-35、42和43中任一项所限定,所述细胞群根据项44所限定,和/或所述抗体或抗原结合片段根据项37-41中任一项所限定。

[0096] 59.根据项57或58所述用途的剂,其中,所述CRC是结肠癌。

[0097] 60.根据项57或58所述用途的剂,其中,所述CRC是直肠癌。

[0098] 61.根据项57-60中任一项所述用途的剂,进一步包括对所述受试者使用至少一种

额外的抗肿瘤剂或疗法。

[0099] 62. 根据项61所述用途的剂, 其中, 所述至少一种额外的抗肿瘤剂或疗法是化疗剂、免疫疗法、免疫检查点抑制剂、放射疗法或外科手术。

[0100] 本公开的其他目的、优点和特征将在阅读以下仅以实例的方式参考附图给出的其具体实施方式的非限制性描述之后变得更加显而易见。

附图说明

[0101] 在附图中:

[0102] 图1描述了在结直肠癌 (CRC) 来源的细胞系和原代肿瘤样品两者中发现肿瘤特异性抗原 (TSA) 的蛋白基因组工作流。使从CRC衍生的细胞系和正常肠来源的细胞系生成的样品以及从六名个体获得的匹配原代肿瘤/正常邻近组织 (NAT) 活检样品均经过RNA测序和主要组织相容性复合体I类 (MHC-I) 免疫沉淀 (IP) 处理。将RNA测序数据用于样品的转录组学表征和定制的全球癌症蛋白质组数据库的生成。对于每种样品, 经由IP分离的MHC-I相关肽 (MAP) 使用各自的数据库经由LC-MS/MS进行识别。在验证了对TSA候选者的识别以及TSA候选者的肿瘤特异性两者后, 通过预测其免疫原性和瘤间分布来评价其治疗潜力。附图使用BioRender.com创建。

[0103] 图2A-E示出了原代肿瘤/正常邻近组织CRC活检的转录组学谱。图2A: 在双端RNAseq和用DESeq2进行基因读段计数 (readcount) 归一化后, 对每个肿瘤/NAT样品的前500个不同基因进行主成分分析 (PCA)。将MSI组织 (由MSISensor测定) 圈起来。图2B: 与CRC组织的邻近NAT相比, 对CRC组织中上调/下调的基因的GO术语分析。使用TPM归一化值, 提交至GO术语分析的基因是具有 $|\log_2FC| > 1$ 并且在所有样品中发现差异调控的那些基因。图2C: 示出MSS NAT、MSI NAT、MSS CRC和MSI CRC的平均ESTIMATE免疫评分的条形图, 并示出了标准偏差。图2D: 示出NAT与CRC样品中可归因于5种不同转录物生物型的转录组的平均比例的堆叠条形图, 其中NAT与CRC之间非编码转录物比例的差异具有统计学显著性 ($p=0.0156$)。图2E: 显示由SNPEff基因组注释确定的MSS和MSI CRC组织的非编码RNA转录物的表达 (左)、SNV计数 (中) 和INDEL计数 (右) 的散点图, 带有平均值和标准误差棒。

[0104] 图3A-E描述了CRC衍生的细胞系和组织的免疫肽组学分析的结果。图3A: 顶小图: 显示在CRC细胞系中识别出的特有肽的数量的堆叠条形图, 并且水平线指示每种细胞系的MAP的平均数量。底小图: 示出在每种细胞系中识别出的特有MAP的数量与细胞表面MHC I的呈递之间的相关性 (皮尔逊相关系数 (Pearson's r) = 0.96) 的散点图。图3B: 显示在原代组织样品中识别出的特有肽的数量的堆叠条形图, 并且水平线指示每种组织样品的MAP的平均数量。图3A和3B中的“所有肽”表示用5% FDR识别出的肽的数量, 而“MHC I肽”表示用相应的肽评分识别的肽的数量, 长度为8-11个氨基酸, 使用netpanMHC4.1b预测等级洗脱配体阈值 $\leq 2\%$ 。图3C: 指示使用洗脱等级预测, 预测每种样品中与给定HLA等位基因结合的特有MAP的比例的条形图。图3D: 对CRC衍生细胞系和原代组织的MAP来源基因的GO术语分析。对于组织, 只有由4种或更多种组织共享的来源基因被包括在此分析中。图3E左小图: 显示衍生自蛋白编码、高变基因 (免疫球蛋白或TCR) 或非编码转录物的每种组织样品或来自未注释转录物的那些中的MAP的比例的堆叠条形图。右小图: 显示来源于经加工的转录物、保留的内含子、无休止衰变产物、无义介导的衰变产物、lncRNA或具有未注释转录物的那些的非

编码MAP的比例的堆叠条形图。

[0105] 图4A-E示出了在CRC中识别的新型TSA主要来源于非编码区,而大多数TAA来源于外显子。图4A:显示了每种样品中识别出的TSA的数量的条形图,每种组织样品中TSA的平均数量用水平线表示。图4B:一种堆叠饼图,其在内部饼图中识别TSA的基因组来源,以及TSA突变(mTSA)或异常表达(aeTSA)的比例。外部饼图说明了来自编码或非编码序列的TSA的比例。图4C:显示了每种样品中识别出的TAA的数量的条形图。图4D:一种堆叠饼图,其内部饼图中识别TAA的基因组来源,以及典型或非典型TAA的比例。外部饼图显示了来自编码或非编码序列的TAA的比例。图4E:显示了关于CRC免疫肽组学的2篇先前出版物(8,15),以及IEDB和HLA配体图谱(Atlas)(所有组织,以及仅结肠组织)中存在和不存在推定的TSA和TAA的热图。

[0106] 图5A-B示出了推定的TSA和TAA的RNA表达谱。图5A:显示了与y轴上匹配的NAT相比在CRC中以TPM表示的转录物的 $\log_2FC$ 以及给定组织样品中的平均表达(CRC和NAT的平均值)的MA图。突显的点表示推定的TAA和TSA的来源转录物。S4和S5图两者都有一个不可见的典型TAA点,因为它与另一个典型TAA来源转录物重叠。图5B:来自基因型组织表达(GTEx)门户的正常组织和合并的TEC样品中aeTSA编码序列和TAA编码序列(分为典型TAA(canTAA)和非典型TAA(non-canTAA))的以 $\log(rphm+1)$ 表示的平均RNA表达的热图。MHC1ow组织包括来自脑、神经和睾丸的组织,这些组织已被证明低表达MHC I。黑色轮廓表示平均RNA表达 $>8.55rphm$ 。

[0107] 图6A-C示出了对TSA和TAA候选者的验证。图6A:热图显示了151例TCGA COAD样品中TSA和TAA的以 $\log(rphm+1)$ 表示的平均RNA表达。左侧显示了表达TSA和TAA序列的TCGA COAD样品的比例,该比例是合并的GTEx和mTEC样品的对数转化($\log(rphm+1)$)平均表达的至少10倍高。图6B:与Adamopoulou et al.2013年报道的据推测的无免疫原性的胸腺肽相比,多组经验证的TSA和TAA的rEpitope免疫原性评分。由rEpitope建议的MHC I肽免疫原性阈值(0.36)由虚线表示。图6C:美国人群中肿瘤抗原结合MHC I类等位基因的预测患病率(IEDB)。

[0108] 图7示出了集合图(upset plot),该图显示了给定样品的交集所特有的HLA等位基因的数量,特别是给定样品所特有的MAP或2种样品独特共享的MAP。

[0109] 图8A-D示出了CRC衍生细胞系的转录组学谱,CRC组织中免疫浸润的ssGSEA分析以及所有样品的突变谱。图8A:在双端RNA seq和用DESeq2进行基因读段计数(readcount)归一化后,对CRC衍生的细胞系和一种正常肠细胞系(HIEC-6)的前500个不同基因进行主成分分析(PCA)。将已知的MSI细胞系圈出。图8B:使用在Danaher et al.2017和GSVA R程序(<https://github.com/rcastelo/GSVA>)中描述的基因对肿瘤和匹配的NAT中的免疫浸润进行ssGSEA分析。图8C:显示通过SNPEff基因组注释确定的MSS和MSI CRC衍生细胞系的SNV计数和INDEL计数的散点图,具有平均值和标准误差棒。图8D:显示通过SNPEff基因组注释确定的所有MSS和MSI样品(细胞系和组织)的SNV计数和INDEL计数的散点图,具有平均值和标准误差棒。MSS和MSI样品之间的INDEL突变数量差异具有统计学显著性($p=0.00235$)。

[0110] 图9A和B示出了对MSI和MSS原代组织样品的GO术语分析结果。图9A:与MSI肿瘤的邻近NAT相比,对MSI肿瘤中上调/下调基因的GO术语分析。用于GO术语分析的基因是那些当使用TPM归一化值与它们相应的NAT进行比较时 $|\log_2FC|>1$ 的基因,并且发现这些基因在两

种MSI肿瘤样品中特有地差异地表达(即,仅在两种MSI肿瘤中而不在任何MSS肿瘤中上调/下调的基因)。图9B:与MSS肿瘤的NAT相比,对MSS肿瘤中上调/下调的基因的GO术语分析。用于GO术语分析的基因是那些当使用TPM归一化值与它们相应的NAT进行比较时 $|\log_2FC|>1$ 的基因,并且发现这些基因在MSI肿瘤样品的3种或更多种中特有地差异地表达(即,在至少三种MSS肿瘤中而不在任何MSI肿瘤中上调/下调的基因)。

[0111] 图10A-D提供了CRC衍生细胞系和CRC/NAT组织样品中独有和共享MAP的概述。图10A:显示了四种CRC衍生细胞系的MHC I免疫肽组的MAP的重叠的维恩图。图10B:显示了六种原代组织样品的MHC I免疫肽组的MAP的重叠的维恩图。图10C:显示了给定样品的交集所特有的MAP的数量的集合R图(UpsetR plot),特别是给定样品所特有的MAP或两种样品独特共享的MAP。图10D:显示了任意两种细胞系或组织样品之间共享的MAP的数量的热图。

[0112] 图11A-E提供了CRC衍生细胞系和CRC/NAT组织样品中独有和共享MAP来源基因的概述。图11A:顶小图:显示了每种样品中识别出的特有来源基因数量的条形图。底小图:显示了每种样品中识别出的特有MAP的数量与相应的特有来源基因的数量之间的相关性(皮尔逊相关系数=0.99)的散点图。从编码序列中识别出肽的来源基因;映射到大于一种来源基因的任何肽都被排除在外。图11B:显示了给定样品的交集所特有的来源基因的数量的集合R图(UpsetR plot),特别是给定样品所特有的来源基因或由两种样品独特共享的来源基因。图11C:显示了四种CRC衍生细胞系的MHC I免疫肽组的来源基因的重叠的维恩图。图11D:显示了六种原代组织样品的MHC I免疫肽组的来源基因的重叠的维恩图。图11E:显示了任意两种细胞系或组织样品之间共享的来源基因的数量的热图。

[0113] 图12是表示每种样品中识别出的特有MAP的数量与经识别和经验证的TSA的数量之间的相关性(皮尔逊相关系数=0.76)的散点图。

[0114] 图13是描述本文所述TSA和TAA的免疫原性评分的图。与Adamopoulou et al.2013年报道的据推测的无免疫原性的胸腺肽相比,不同组的经验证的TSA和TAA的rEpitope免疫原性评分。由rEpitope建议的MHC I肽免疫原性阈值(0.36)由虚线表示。该图与图6A的不同之处在于,它包括了本工作中报告的所有TAA,而不仅仅是选择进行验证的九种TAA。

具体实施方式

[0115] 除非本文另有说明或上下文明显矛盾,否则在描述技术的上下文中(尤其是在所附权利要求的上下文中),术语“一个/一种(a/an)”和“所述/该(the)”以及类似的指代物的使用应被解释为涵盖单数和复数两者。

[0116] 除非另有说明,否则术语“包括(comprising)”、“具有”、“包括(including)”和“包含”应理解为开放式术语(即,意指“包括但不限于”)。

[0117] 除非本文另有说明或与上下文明显矛盾,否则本文所述的所有方法都可以任何合适的顺序进行。

[0118] 除非另有声明,否则本文提供的任何和所有实例或示例性语言(例如,“诸如”)的使用仅旨在更好地说明要求保护的技术的实施方式并且不对范围构成限制。

[0119] 说明书中的任何语言都不应被解释为表明任何未要求保护的元素对于实践所要求保护的技术的实施方式是必不可少的。

[0120] 在本文中,术语“约”具有其普通含义。术语“约”用于表明值包括用于确定该值的

设备或方法的固有误差变化,或涵盖接近所述值的值,例如在所述值(或值范围)的10%以内。

[0121] 除非在本文中另有说明,否则本文对数值范围的引用仅旨在用作单独提及落入该范围内的每个单独值的速记方法,并且将每个单独值并入说明书中,就好像其在本文中单独引用一样。范围内的值的所有子集也被并入到说明书中,就好像它们在本文中被单独引用一样。

[0122] 在本公开的特征或方面以马库什组或备选列表的形式描述的情况下,本领域技术人员将认识到,本公开也因此根据马库什组或备选列表的任何单独成员或成员子组来描述。

[0123] 除非另有明确定义,否则本文中使用的所有技术和科学术语应理解为与本领域(例如,在干细胞生物学、细胞培养、分子遗传学、免疫学、免疫组织化学、蛋白化学和生物化学中)中普通技术人员通常理解的含义相同。

[0124] 除非另有说明,否则本公开中利用的重组蛋白、细胞培养和免疫学技术是本领域技术人员熟知的标准程序。这些技术在诸如以下的原始文献中有完整的描述和解释: J.Perbal, A Practical Guide to Molecular Cloning, John Wiley and Sons (1984)、J.Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory Press (1989), T.A. Brown (editor), Essential Molecular Biology: A Practical Approach, Volumes 1 and 2, IRL Press (1991)、D.M. Glover and B.D. Hames (editors), DNA Cloning: A Practical Approach, Volumes 1-4, IRL Press (1995 and 1996) 以及 F.M. Ausubel et al. (editors), Current Protocols in Molecular Biology, Greene Pub. Associates and Wiley-Interscience (1988, including all updates until present)、Ed Harlow and David Lane (editors) Antibodies: A Laboratory Manual, Cold Spring Harbour Laboratory, (1988) 以及 J.E. Coligan et al. (editors) Current Protocols in Immunology, John Wiley & Sons (including all updates until present)。

[0125] 在本文所述的研究中,本发明人已经使用基于蛋白质基因组学的方法从CRC细胞系和CRC标本中识别出了TSA和TAA候选者。该TSA中有一部分源自异常表达的未突变基因组序列,这些序列在正常组织中不表达,诸如非外显子序列(例如,内含子和基因间序列)。本文所识别的新型CRC TSA和TAA候选者可以用于例如基于CRC T细胞的免疫疗法和疫苗。

[0126] 因此,在一方面,本公开涉及肿瘤抗原肽(TAP)(或肿瘤特异性肽),以及更具体地,涉及分离的TAP,诸如CRC TAP,其包括以下氨基酸序列中的一种或由以下氨基酸序列中的一种组成:

[0127]	序列	SEQ ID NO:	序列	SEQ ID NO:
	RMLLSHTGK	1	KIGEVIVTK	16

[0128]

LPHRALSGI	2	TRSTIILHL	17
GTNPAAVK	3	SILRGSLGK	18
LRHKLVLNR	4	RLVSNRVVR	19
RIGGVGVEK	5	SLKNLEPIK	20
SIETVNSL	6	RSAGPRAPL	21
TVNTQQYNTK	7	VLYRSVLLK	22
SVSHLHIF	8	KAAGAGAAK	23
TTLENLPQK	9		
AQKLQVRI	10		
GQIELSIYR	11		
HGALSIRSI	12		
RLMKFLPV	13		
SLYISEERK	14		
VQTAVLNV	15		

[0129] 在另一方面,本公开另外涉及包括以下氨基酸序列中的一种或由以下氨基酸序列中的一种组成的TAP(例如,分离的TAP)用于治疗癌症,以及更具体地用于治疗CRC的用途:

[0130]

序列	SEQ ID NO:	序列	SEQ ID NO:
RMLLSHTGK	1	EYNSDLHQF	26
LPHRALSGI	2	TYKYVDINTF	27
GTNPAAVK	3	VIDEPVRL	28
LRHKLVLNR	4	RYLEKFYGL	29
RIGGVGVEK	5	KRLPSPSLF	30
SIETVNSL	6	KSINEFWNK	31
TVNTQQYNTK	7	RVFHGNETTK	32

[0131]

SVSHLHIFF	8	RIQLPVVSK	33
TTLENLPQK	9	SIFRTPISK	34
AQKLQVRI	10	AEYHSQSRA	35
GQIELSIYR	11	IHNSDQILL	36
HGALSIRSI	12	KMHQSRISY	37
RLMKFLPV	13	QMAGLRDTY	38
SLYISEERK	14	RIKNWGSGMTF	39
VQTAVLNV	15	AQYDQASTKY	40
KIGEVIVTK	16	FVDNQYWRY	41
TRSTIHLHL	17	QLDSAVKNLY	42
SILRGSLGK	18	EQLKVTTI	43
RLVSNRVVR	19	FRSWAVQTF	44
SLKNLEPIK	20	KLYQSIIGY	45
RSAGPRAPL	21	RVAKTNSLR	46
VLYRSVLLK	22	SANVSKVSF	47
KAAGAGAAK	23	SQHDILQMI	48
VEAPHLPSF	25	VAKTPRNTY	49
		RPRLSAILF	50

。

[0132] 在一种实施方式中，TAP包括以下氨基酸序列中的一种或由以下氨基酸序列中的一种组成：SEQ ID NO:1-23。在一种实施方式中，TAP包括以下氨基酸序列中的一种或由以下氨基酸序列中的一种组成：SEQ ID NO:1-17、22、25、27、29、31、33、38、40、41和46。在另外的实施方式中，TAP包括以下氨基酸序列中的一种或由以下氨基酸序列中的一种组成：SEQ ID NO:1-17、22和25。在另外的实施方式中，TAP包括以下氨基酸序列中的一种或由以下氨基酸序列中的一种组成：SEQ ID NO:1-17和22。

[0133] 通常，在HLA I类上下文中所呈现的肽诸如肿瘤抗原肽（TAP）的长度在约7或8至约15个，或优选8至14个氨基酸残基之间变化。在本公开的方法的一些实施方式中，包括本文限定的TAP序列的更长的肽被人工负载至细胞（诸如抗原呈递细胞（APC））中，由细胞处理并且TAP由APC表面处的MHC I类分子呈递。在该方法中，可以将长度超过15个氨基酸残基的

肽/多肽负载至APC中,通过APC胞质溶胶中的蛋白酶进行处理,提供本文限定的相应TAP用于呈递。在一些实施方式中,用于生成本文限定的TAP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,使用本文所述的TAP的所有方法和过程包括使用更长的肽或多肽(包括天然蛋白),即,肿瘤抗原前体肽/多肽,以诱导在由细胞(APC)处理后呈递“最终”8-14个TAP。在一些实施方式中,本文提及的TAP为约8至14、8至13或8至12个氨基酸长(例如,8、9、10、11、12或13个氨基酸长),足够小用于直接配合HLA I类分子。在一种实施方式中,TAP包括20个氨基酸或更少,优选15个氨基酸或更少,更优选14个氨基酸或更少。在一种实施方式中TAP包括至少7个氨基酸,优选至少8个氨基酸或更少,更优选至少9个氨基酸。

[0134] 如本文所用,术语“氨基酸”包括天然存在的氨基酸以及其他氨基酸(例如,天然存在的氨基酸、非天然存在的氨基酸、不由核酸序列编码的氨基酸等)的L-和D-异构体,用于肽化学以制备TAP的合成类似物。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-氨基酸的保守置换。天然存在的非遗传编码氨基酸包括例如 β -丙氨酸、3-氨基-丙酸、2,3-二氨基丙酸、 α -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、苯基甘氨酸、环己基丙氨酸、正亮氨酸(Nle)、正缬氨酸、2-萘基丙氨酸、吡啶丙氨酸、3-苯并噻吩丙氨酸、4-氯苯丙氨酸、2-氟苯丙氨酸、3-氟苯丙氨酸、4-氟苯丙氨酸、青霉胺、1,2,3,4-四氢-异喹啉-3-羧酸、 β -2-噻吩丙氨酸、甲硫氨酸亚砷、L-高精氨酸(Hoarg)、N-乙酰赖氨酸、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、 ϵ -氨基己酸、 δ -氨基戊酸或2,3-二氨基丁酸(D-或L-)等。这些氨基酸是生物化学/肽化学领域所熟知的。在一种实施方式中,TAP仅包括天然存在的氨基酸。

[0135] 在实施方式中,本文所述的TAP包括具有改变的序列的肽,所述改变的序列含有相对于本文提及的序列的功能等效氨基酸残基的置换。例如,序列内的一个或多个氨基酸残基可以被具有相似极性(具有相似物理化学特性)的其他氨基酸置换,其充当功能等效物,导致沉默改变。序列内氨基酸的置换可以选自该氨基酸所属类别的其他成员。例如,带正电荷的(碱性)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)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。不带电荷的极性氨基酸包括丝氨酸、苏氨酸、半胱氨酸、酪氨酸、天冬酰胺和谷氨酰胺。带负电荷的(酸性)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)极性氨基酸家族中。在氨基酸家族内进行的置换通常被理解为保守置换。本文提及的TAP可以包括所有L-氨基酸、所有D-氨基酸或者L-氨基酸和D-氨基酸的混合物。在一种实施方式中,本文提及的TAP包括所有L-氨基酸。

[0136] 在一种实施方式中,在包括SEQ ID NO:1-23和25-50的序列中的一种或者由SEQ ID NO:1-23和25-50的序列中的一种组成的TAP的序列中,可以通过替换为其他氨基酸来改性基本上不有助于与T细胞受体相互作用的氨基酸残基,该其他氨基酸的掺入基本上不影响T细胞反应性并且不消除与相关MHC的结合。

[0137] TAP还可以是N-末端和/或C-末端封端的或改性的以防止降解,增加稳定性、亲和

力和/或摄取。因此,在另一方面,本公开提供了式 Z^1-X-Z^2 的改性TAP,其中X是包括以下氨基酸序列中的一种或由以下氨基酸序列中的一种组成的TAP:SEQ ID NO:1-23和25-50。

[0138] 在一种实施方式中,TAP的氨基末端残基(即,N-末端处的游离氨基基团)被改性(例如,为了防止降解),例如通过部分/化学基团(Z^1)的共价连接。 Z^1 可以是一至八个碳的直链或支链烷基基团、或者酰基基团($R-CO-$)(其中R为疏水部分(例如,乙酰基、丙酰基、丁酰基、异丙酰基或异丁酰基))、或者芳酰基基团($Ar-CO-$)(其中Ar是芳基基团)。在一种实施方式中,酰基基团是 C_1-C_{16} 或 C_3-C_{16} 酰基基团(直链或支链、饱和或不饱和的),在另一实施方式中,是饱和的 C_1-C_6 酰基基团(直链或支链)或不饱和的 C_3-C_6 酰基基团(直链或支链),例如乙酰基基团(CH_3-CO- ,Ac)。在一种实施方式中, Z^1 不存在。TAP的羧基末端残基(即TAP的C-末端处的游离羧基基团)可以被改性(例如,为了保护以免受降解),例如通过酰胺化(OH基团被 NH_2 基团取代),因此在这种情况下 Z^2 是 NH_2 基团。在一种实施方式中, Z^2 可以是异羟肟酸基团、腈基团、酰胺(伯、仲或叔)基团、具有一至十个碳的脂肪族胺(诸如甲胺、异丁胺、异戊胺或环己胺)、芳族或芳烷基胺(诸如苯胺、萘胺、苄胺、肉桂胺或苯乙胺)、醇或 CH_2OH 。在一种实施方式中, Z^2 不存在。在一种实施方式中,TAP包括以下氨基酸序列中的一种:SEQ ID NO:1-24和26-50。在一种实施方式中,TAP由以下氨基酸序列中的一种组成:SEQ ID NO:1-24和26-50,即其中 Z^1 和 Z^2 不存在。

[0139] 在另一方面,本公开提供了与HLA-A*02:01分子结合的TAP,该TAP包括SEQ ID NO:6的序列或由SEQ ID NO:6的序列组成。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-A*02:05、HLA-A*02:06和/或HLA-A*02:07分子结合。

[0140] 在另一方面,本公开提供了与HLA-A*03:01分子结合的TAP,该TAP包括以下序列或由以下序列组成:SEQ ID NO:1、11、14、38、45或46,优选SEQ ID NO:1、11或14。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-A*03:02、HLA-A*11:01或HLA-A*30:01分子结合。

[0141] 在另一方面,本公开提供了与HLA-A*03:02分子结合的TAP,该TAP包括以下序列或由以下序列组成:SEQ ID NO:3、5、7、16、23、31、32、33或34,优选SEQ ID NO:3、5、7、16或23。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-A*03:01、HLA-A*11:01或HLA-A*30:01分子结合。

[0142] 在另一方面,本公开提供了与HLA-A*11:01分子结合的TAP,该TAP包括以下序列或由以下序列组成:SEQ ID NO:9、18、33或34,优选SEQ ID NO:9或18。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-A*03:01、HLA-A*03:02、HLA-A*31:01和/或HLA-A*68:01分子结合。

[0143] 在另一方面,本公开提供了与HLA-A*23:01分子结合的TAP,该TAP包括SEQ ID NO:29的序列或由SEQ ID NO:29的序列组成。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-A*24:02分子结合。

[0144] 在另一方面,本公开提供了与HLA-A*24:02分子结合的TAP,该TAP包括SEQ ID NO:26、27或29的序列或者由SEQ ID NO:26、27或29的序列组成。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-A*23:01分子结合。

[0145] 在另一方面,本公开提供了与HLA-A*30:01分子结合的TAP,该TAP包括SEQ ID NO:

19、20或23的序列或者由SEQ ID NO:19、20或23的序列组成。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-A*30:02和/或HLA-B*15:02分子结合。

[0146] 在另一方面,本公开提供了与HLA-A*32:01分子结合的TAP,该TAP包括以下序列或由其组成:SEQ ID NO:8、37或39,优选SEQ ID NO:8。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-B*57:01和/或HLA-B*58:01分子结合。

[0147] 在另一方面,本公开提供了与HLA-B*07:02分子结合的TAP,该TAP包括以下序列或由以下序列组成:SEQ ID NO:2、21、24或50,优选SEQ ID NO:2或21。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-B*35:02、HLA-B*35:03、HLA-B*55:01和/或HLA-B*56:01分子结合。

[0148] 在另一方面,本公开提供了与HLA-B*13:02分子结合的TAP,该TAP包括以下序列或由其组成:SEQ ID NO:13或48,优选SEQ ID NO:13。

[0149] 在另一方面,本公开提供了与HLA-B*18:01分子结合的TAP,该TAP包括SEQ ID NO:25的序列或由SEQ ID NO:25的序列组成。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-B*40:01、HLA-B*44:02、HLA-B*44:03和/或HLA-B*45:01分子结合。

[0150] 在另一方面,本公开提供了与HLA-B*27:05分子结合的TAP,该TAP包括SEQ ID NO:4的序列或由SEQ ID NO:4的序列组成。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-B*27:02结合。

[0151] 在另一方面,本公开提供了与HLA-B*52:01分子结合的TAP,该TAP包括以下序列或由以下序列组成:SEQ ID NO:10、12、15或43,优选SEQ ID NO:10、12或15。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-B*51:01结合。

[0152] 在另一方面,本公开提供了与HLA-C*06:02分子结合的TAP,该TAP包括以下序列或由以下序列组成:SEQ ID NO:17或44,优选SEQ ID NO:17。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-B*27:02、HLA-C*07:01和/或HLA-C*07:02分子结合。

[0153] 在另一方面,本公开提供了与HLA-C*12:02分子结合的TAP,该TAP包括SEQ ID NO:47或49的序列或者由SEQ ID NO:47或49的序列组成。由于HLA等位基因表现出混杂性(某些HLA等位基因存在相似的表位),上面识别的TAP可以另外与HLA-B*46:01、HLA-C*03:02、HLA-C*03:03、HLA-C*03:04、HLA-C*08:01、HLA-C*12:03、HLA-C*15:02和/或HLA-C*16:01分子结合。

[0154] 在一种实施方式中,TAP由位于基因组的非蛋白编码区的序列编码。在一种实施方式中,TAP由位于非翻译转录区(UTR),即,3'-UTR或5'-UTR区中的序列编码。在另一实施方式中,TAP由位于内含子中的序列编码。在另一实施方式中,TAP由位于基因间区中的序列编码。在另一实施方式中,TAP由位于外显子中并且源自移码的序列编码。

[0155] 本公开的TAP可以通过在包括编码TAP的核酸的宿主细胞中表达(重组表达)或通过化学合成(例如,固相肽合成)来产生。肽可以通过本领域熟知的手动和/或自动固相程序

容易地合成。合适的合成可以例如通过利用“T-boc”或“Fmoc”程序进行。固相合成的技术和程序描述于例如由E. Atherton和R.C. Sheppard所著的固相肽合成：一种实用的方法，由IRL出版，牛津大学出版社，1989。替代地，MiHA肽可以通过片段缩合来制备，如以下所述，例如：Liu等人，Tetrahedron Lett. 37:933-936, 1996; Baca等人，J. Am. Chem. Soc. 117:1881-1887, 1995; Tam等人，Int. J. Peptide Protein Res. 45:209-216, 1995; Schnolzer和Kent，Science 256:221-225, 1992; Liu和Tam，J. Am. Chem. Soc. 116:4149-4153, 1994; Liu和Tam，Proc. Natl. Acad. Sci. USA 91:6584-6588, 1994; 以及Yamashiro和Li，Int. J. Peptide Protein Res. 31:322-334, 1988)。其他可用于合成TAP的方法描述于Nakagawa等人，J. Am. Chem. Soc. 107:7087-7092, 1985。在一种实施方式中，TAP是化学合成的（合成肽）。本公开的另一实施方式涉及非天然存在的肽，其中，所述肽由本文限定的氨基酸序列组成或基本上由本文限定的氨基酸序列组成并且已经合成产生（例如，合成）为药学上可接受的盐。根据本公开的TAP的盐与处于其体内一种或多种状态的肽显著不同，因为体内生成的肽不是盐。肽的非天然盐形式可以调节肽的溶解度，特别是在包括肽的药物组合物的情况下，例如，如本文公开的肽疫苗。优选地，盐是肽的药学上可接受的盐。

[0156] 在一种实施方式中，本文提及的TAP是分离的或基本上纯的。当化合物诸如肽或核酸与存在于分子的自然环境中的组分或天然来源大分子（例如，包括其他核酸、蛋白、脂质、糖，等等）分开时，该化合物是“分离的”或“基本上纯的”。通常，当化合物按重量计占样品中总材料的至少60%，更通常75%、80%或85%，优选超过90%，和更优选超过95%时，该化合物是基本上纯的。因此，例如，化学合成或通过重组技术产生的多肽通常基本上不含其天然相关组分，例如，其来源大分子的组分。当核酸分子不与编码序列紧密相邻（即共价连接）时，核酸分子是基本上纯的，在该核酸分子源自的生物体的天然存在的基因组中，该编码序列通常与该核酸分子相邻。例如，可以通过从天然来源中提取；通过表达编码肽化合物的重组核酸分子；或通过化学合成来获得基本上纯的化合物。纯度可以使用任何合适的方法测量，例如柱色谱、凝胶电泳、HPLC等。在一种实施方式中，TAP在溶液中。在另一实施方式中，TAP是固体形式，例如冻干的。

[0157] 在一种实施方式中，TAP由位于基因组的非蛋白编码区的序列编码。在一种实施方式中，TAP由位于基因间区中的序列编码。在另一实施方式中，TAP由非编码RNA（ncRNA）编码。

[0158] 在另一方面，本公开另外提供了包括本文所述TAP的至少一种的合成肽（SLP）。在一种实施方式中，SLP包括至少两种TAP，其中TAP的至少一种是本文所描述的TAP。在一种实施方式中，SLP包括本文所述的TAP的至少两种，三种，四种或五种。在一种实施方式中，SLP包括与一个或多个氨基酸或结构域连接的本文所述的TAP的至少一种，该氨基酸或结构域赋予SLP期望的特性，诸如稳定SLP和/或改善MHC分子加工和呈递的序列或结构域，例如包括可被细胞蛋白酶诸如组织蛋白酶切割的基序的序列。在另一实施方式中，SLP包括本文所述的TAP的至少一种，以及与MHC II类分子结合的TAP。TAP可以直接相互附接，或者可以由接头诸如短氨基酸接头间接附接。在实施方式中，接头包括约4至约20个氨基酸，或约4至约15个氨基酸，例如4、5、6、7、8、9、10、11、12、13、14或15个氨基酸。在一种实施方式中，接头包括甘氨酸残基、丝氨酸残基、脯氨酸残基、苏氨酸残基或其混合物。在一种实施方式中，SLP具有的长度为200、150、100、90、80、70、60或50个氨基酸或更少。在另外的实施方式中，

SLP具有的长度为20至50、45或40个氨基酸,例如20或25个氨基酸至30、35或40个氨基酸。如本文所用,“合成的”是指未从其天然来源中分离的肽或核酸分子,例如,通过重组技术或使用化学合成产生的。

[0159] 在另一方面,本公开另外提供了编码本文提及的TAP或肿瘤抗原前体-肽或SLP的核酸(例如,分离的)。在一种实施方式中,核酸包括约21个核苷酸至约45个核苷酸,约24个至约45个核苷酸,例如24、27、30、33、36、39、42或45个核苷酸。

[0160] 在一种实施方式中,编码本公开的TAP的核酸(DNA、RNA)包括以下的任何一种:SEQ ID NO:51-73和75-100,或SEQ ID NO:51-67、72、75、77、79、81、83、88、90、91和96,或SEQ ID NO:51-67、72和75(例如SEQ ID NO:51-67和72)中限定的序列,或者相应的RNA序列(即,其中胸腺嘧啶核碱基(T)被尿嘧啶核碱基(U)取代)。在一种实施方式中,编码TAP的核酸是mRNA分子。在一种实施方式中,核酸呈溶液形式。在另一实施方式中,核酸呈固体形式,例如冻干的。

[0161] 本公开的核酸可用于重组表达本公开TAP或SLP,并且可以被包括在载体或质粒中,诸如克隆载体或表达载体,其可以被转染到宿主细胞中。在一种实施方式中,本公开提供了包括编码本公开的TAP的核酸序列的克隆、表达或病毒载体或质粒。可选地,可以将编码本公开的TAP的核酸掺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TAP或蛋白。如本文所用,术语“宿主细胞”不仅指特定的受试者细胞,而且还指此类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TAP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、酵母细胞、植物细胞或哺乳动物细胞)。载体或质粒含有插入的编码序列的转录和翻译所必需的元件,以及可以含有其他组分(诸如,抗性基因、克隆位点等)。可以使用本领域技术人员熟知的方法构建表达载体,该表达载体含有编码肽或多肽的序列以及与其可操作地连接的合适的转录和翻译控制/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。此类技术描述于Sambrook.等人(1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., 和Ausubel, F.M.等人(1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y.。“可操作地连接”是指组分特别是核苷酸序列的联接(juxtaposition),使得可以进行组分的正常功能。因此,与调控序列可操作地连接的编码序列是指核苷酸序列的构型,其中编码序列可以在调控序列的调控控制下,即转录和/或翻译控制下表达。如本文所用,“调控/控制区”或“调控/控制序列”是指参与调控编码核酸的表达式非编码核苷酸序列。因此,术语调控区包括启动子序列、调控蛋白结合位点、上游激活序列等。载体(例如,表达载体)可以具有必需的5'上游和3'下游调控元件诸如启动子序列(诸如CMV、PGK和EF-1 α 启动子)、核糖体识别和结合TATA盒以及3'UTR AAUAAA转录终止序列,用于在其各自的宿主细胞中进行有效的基因转录和翻译。其他合适的启动子包括以下的组成型启动子:猿猴病毒40(SV40)早期启动子、小鼠乳腺肿瘤病毒(MMTV)、HIV LTR启动子、MoMuLV启动子、禽白血病病毒启动子、EBV立即早期启动子和劳斯肉瘤病毒启动子。还可以使用人基因启动子,包括但不限于肌动蛋白启动子、肌球蛋白启动子、血红蛋白启动子和肌酸激酶启动子。在某些实施方式中,诱导型启动子也被认为是表达TAP的载体的一部分。这提供了能够打开目的核苷酸序列的表达式或关闭表达的分子开关。诱导型启动子的实例包括,但不限于金属硫蛋白启动子、糖皮质激素启动子、孕酮启动子或四环素启动

子。载体的实例是质粒、自主复制序列和转座元件。其他示例性载体包括,但不限于质粒、噬菌粒、粘粒、人工染色体(诸如酵母人工染色体(YAC)、细菌人工染色体(BAC)或PI来源的人工染色体(PAC))、噬菌体(诸如 λ 噬菌体或M13噬菌体)和动物病毒。可用作载体的动物病毒类别的实例包括,但不限于逆转录病毒(包括慢病毒)、腺病毒、腺相关病毒、疱疹病毒(例如,单纯疱疹病毒)、痘病毒、杆状病毒、乳头瘤病毒和乳多空病毒(例如,SV40)。表达载体的实例是在哺乳动物细胞中表达的Lenti-XTM双顺反子表达系统(Neo)载体(Contech)、pCIneo载体(Promega);在哺乳动物细胞中慢病毒介导的基因转移和表达的pLenti4/V5-DESTTM、pLenti6/V5-DESTTM和pLenti6.2N5-GW/lacZ(Invitrogen)。本文公开的TAP的编码序列可以连接至此类表达载体中,用于在哺乳动物细胞中表达TAP。

[0162] 在某些实施方式中,编码本公开的TAP的核酸在病毒载体中提供。病毒载体可以是来源于腺病毒、痘苗病毒、逆转录病毒、慢病毒或泡沫病毒的那些。如本文所用,术语“病毒载体”是指核酸载体构建体,其包括至少一种病毒来源的元件并且具有被包装至病毒载体颗粒中的能力。病毒载体可以含有本文所述的各种蛋白的编码序列来代替非必需的病毒基因。在另一实施方式中,编码本公开的TAP的核酸在自扩增或自复制RNA(srRNA)载体中提供。srRNA来源于正链RNA病毒,其中的结构蛋白已被移除并被目的异源基因取代。srRNA已经成功地从黄病毒、诺达病毒(nodamura virus)、套式病毒和甲病毒中获得,其中治疗形式的技术提供了反式结构蛋白以产生单周期病毒复制子颗粒(VRP)(参见,例如,Aliahmad et al. Next generation self-replicating RNA vectors for vaccines and immunotherapies. Cancer Gene Ther (2022). <https://doi.org/10.1038/s41417-022-00435-8>)。载体和/或颗粒可用于在体外或体内将DNA、RNA或其他核酸转移至细胞中。许多形式的病毒载体是本领域已知的。

[0163] 在实施方式中,编码本公开的TAP的核酸(DNA、RNA)被包括在囊泡或纳米颗粒诸如脂质囊泡(例如,脂质体)或脂质纳米颗粒(LNP)或任何其它合适的运载体(vehicle)中。因此,在另一方面,本公开提供了囊泡或纳米颗粒,诸如脂质囊泡或纳米颗粒,其包括编码一种或多种本文所述TAP的核酸,诸如mRNA。

[0164] 如本文使用的术语脂质体与其通常含义一致,是指由双层磷脂或任何类似的两亲性脂质(例如,鞘脂)组成的包封内部水性介质的微小脂质囊泡。

[0165] 术语“脂质纳米颗粒”是指可以包括包围类似于脂质体的内部水性介质的一个或多个脂质双层环的脂质体样结构,或者是将分子(例如核酸)包封在非水性核心中的胶束样结构。脂质纳米颗粒通常含有阳离子脂质,诸如可电离的阳离子脂质。可以用于LNP的阳离子脂质的实例包括DOTMA、DOSPA、DOTAP、ePC、DLin-MC3-DMA、C12-200、ALC-0315、cKK-E12、脂质H(SM-102)、OF-Deg-Lin、A2-Iso5-2DC18、3060_{i10}、BAME-016B、TT3、9A1P9、FTT5、**COATSOME[®] SS-E、COATSOME[®] SS-EC、COATSOME[®] SS-OC以及COATSOME[®] SS-OP**(参见,例如,Hou et al., Nature Reviews Materials, volume 6, pages 1078-1094 (2021); Tenchov et al., ACS Nano, 15, 16982-17015 (2021))。

[0166] 脂质体和脂质纳米颗粒通常包括其他脂质组分,诸如脂质、脂质样物质和聚合物,它们可以改善脂质体或纳米颗粒的特性,诸如稳定性、递送效力、耐受性和生物分布。这些包括磷脂(例如,磷脂酰胆碱、磷脂酰乙醇胺、磷脂酰丝氨酸和磷脂酰甘油),诸如1,2-二硬脂酰基-sn-甘油-3-磷酸胆碱(DSPC)和DOPE、甾醇(诸如胆固醇和胆固醇衍生物)、聚乙二醇

化脂质(PEG-脂质)诸如1,2-二肉豆蔻酰基-rac-甘油-3-甲氧基聚乙二醇-2000(PEG₂₀₀₀-DMG)和1,2-二硬脂酰基-rac-甘油-3-甲氧基聚乙二醇-2000(PEG₂₀₀₀-DSG)。

[0167] 在一种实施方式中,根据本公开的脂质纳米颗粒包括一种或多种阳离子脂质,诸如可电离的阳离子脂质。可电离阳离子脂质的实例包括PCT公开号W0 2017/061150和W0 2019/188867中列出的那些,其涵盖以商品名COATSOME[®] SS-E、COATSOME[®] SS-EC、COATSOME[®] SS-OC和COATSOME[®] SS-OP商业化的可电离阳离子脂质。

[0168] 编码TAP的一种或多种的核酸(例如,mRNA)可以被修饰,例如以增加稳定性和/或降低免疫原性。例如,5'端可以被封端以稳定分子并降低免疫原性(例如,如美国专利号10519189和US10494399所述)。可以用1-甲基假尿苷修饰或取代mRNA的一个或多个核苷以增加分子的稳定性或降低先天性免疫系统对分子的识别。US9371511中描述了经修饰核苷的一种形式。可以对mRNA进行的其他类型的修饰包括掺入抗反向帽类似物(ARCA)、5'-甲基-胞苷三磷酸酯(m5CTP)、N6-甲基-腺苷-5'-三磷酸酯(m6ATP)、2-硫代-尿苷三磷酸酯(s2UTP)、假尿苷三磷酸酯、N¹甲基假尿苷三磷酸酯或5-甲氧基尿苷三磷酸酯(5moUTP)。mRNA还可以包括对5'-和/或3'-非翻译区(UTR)和聚腺苷酸化(多聚A)尾的额外修饰(参见,例如,Kim et al.,Molecular&cellular toxicology vol.18,1(2022):1-8)。对编码TAP的核酸(例如,mRNA)的所有这些修饰和其他修饰都涵盖于本公开中。

[0169] 在另一方面,本公开提供了一种MHC I类分子,其包括(即呈递或结合至)以下SEQ ID NO的TAP的一种或多种:SEQ ID NO:1-23和25-50,优选SEQ ID NO:1-23,诸如SEQ ID NO:1-17和22。

[0170] 在一种实施方式中,MHC I类分子是HLA-A2分子,在另外的实施方式中是HLA-A*02:01分子。在一种实施方式中,MHC I类分子是HLA-A3分子,在另外的实施方式中是HLA-A*03:01或HLA-A*03:02分子。在另一实施方式中,MHC I类分子是HLA-A11分子,在另外的实施方式中是HLA-A*11:01分子。在一种实施方式中,MHC I类分子是HLA-A23分子,在另外的实施方式中是HLA-A*23:01分子。在一种实施方式中,MHC I类分子是HLA-A24分子,在另外的实施方式中是HLA-A*24:02分子。在一种实施方式中,MHC I类分子是HLA-A30分子,在另外的实施方式中是HLA-A*30:01分子。在一种实施方式中,MHC I类分子是HLA-A32分子,在另外的实施方式中是HLA-A*32:01分子。在另一实施方式中,MHC I类分子是HLA-B07分子,在另外的实施方式中是HLA-B*07:02分子。在另一实施方式中,MHC I类分子是HLA-B13分子,在另外的实施方式中是HLA-B*13:02分子。在另一实施方式中,MHC I类分子是HLA-B18分子,在另外的实施方式中是HLA-B*18:01分子。在另一实施方式中,MHC I类分子是HLA-B27分子,在另外的实施方式中是HLA-B*27:05分子。在另一实施方式中,MHC I类分子是HLA-B52分子,在另外的实施方式中是HLA-B*52:01分子。在另一实施方式中,MHC I类分子是HLA-C06分子,在另外的实施方式中是HLA-C*06:02分子。在另一实施方式中,MHC I类分子是HLA-C04分子,在另外的实施方式中是HLA-C*04:01分子。在另一实施方式中,MHC I类分子是HLA-C12分子,在另外的实施方式中是HLA-C*12:02分子。

[0171] 在一种实施方式中,TAP(例如,SEQ ID NO:SEQ ID NO:1-23和25-50,优选SEQ ID NO:1-23)与MHC I类分子非共价结合(即,TAP被负载至或非共价结合至MHC I类分子的肽结合槽(groove)/口袋(pocket))。在另一实施方式中,TAP与MHC I类分子(α 链)共价附接/结

合。在此类构建体中,TAP和MHC I类分子(α 链)作为合成融合蛋白产生,通常具有短的(例如,5至20个残基,优选约8-12个,例如,10个)柔性接头或间隔子(例如,聚甘氨酸接头)。在另一方面,本公开提供了编码融合蛋白的核酸,该融合蛋白包括与MHC I类分子(α 链)融合的本文限定的TAP。在一种实施方式中,MHC I类分子(α 链)-肽复合体是多聚化的。因此,在另一方面,本公开提供了(共价或非共价地)负载有本文提及的TAP的MHC I类分子的多聚体。此类多聚体可以附接至标签,例如荧光标签,其允许检测多聚体。已经开发了大量用于产生MHC多聚体的策略,该MHC多聚体包括MHC二聚体、四聚体、五聚体、八聚体等(综述于Bakker和Schumacher,Current Opinion in Immunology 2005,17:428-433)。例如,MHC多聚体可用于检测和纯化抗原特异性T细胞。因此,在另一方面,本公开提供了用于检测或纯化(分离、富集)对本文限定的TAP具有特异性的CD8⁺T淋巴细胞的方法,该方法包括将细胞群与(共价或非共价地)负载有TAP的MHC I类分子的多聚体接触;以及检测或分离被MHC I类多聚体结合的CD8⁺T淋巴细胞。可以使用已知方法分离被MHC I类多聚体结合的CD8⁺T淋巴细胞,例如荧光激活细胞分选(FACS)或磁激活细胞分选(MACS)。

[0172] 在又一方面,本公开提供了一种细胞(例如,宿主细胞),在实施方式中是一种分离的细胞,其包括本文提及的本公开的核酸、载体或质粒,即编码一种或多种TAP的核酸或载体。在另一方面,本公开提供了在其表面表达结合或呈递根据本公开的TAP的MHC I类分子(例如,上文公开的等位基因中的MHC I类分子)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,优选人细胞,细胞系或永生化细胞。在另一实施方式中,细胞是抗原呈递细胞(APC)。在一种实施方式中,宿主细胞是原代细胞、细胞系或永生化细胞。在另一实施方式中,细胞是抗原呈递细胞(APC)。可以经由常规转化或转染技术将核酸和载体引入细胞中。术语“转化”和“转染”是指将外源核酸引入宿主细胞的技术,包括磷酸钙或氯化钙共沉淀、DEAE-葡聚糖介导的转染、脂质体转染、电穿孔、微注射和病毒介导的转染。转化或转染宿主细胞的合适方法可见于例如Sambrook等人(同上)和其他实验室手册。在体内将核酸引入哺乳动物细胞的方法也是已知的,并且可用于将本公开的载体或质粒递送至受试者以进行基因治疗。

[0173] 可以使用本领域已知的多种方法将一种或多种TAP负载至细胞(诸如APC)。如本文所用,用TAP“负载细胞”是指将编码TAP的RNA或DNA或者或TAP转染至细胞中,或替代地,用编码TAP的核酸转化APC。还可以通过将细胞与可以直接结合存在于细胞表面的MHC I类分子(例如,肽冲击细胞)的外源性TAP接触来负载细胞。还可以将TAP融合至促进其被MHC I类分子呈递的结构域或基序,例如融合至内质网(ER)回收信号、C末端Lys-Asp-Glu-Leu序列(参见Wang等人,Eur J Immunol.2004Dec;34(12):3582-94)。

[0174] 在另一方面,本公开提供了一种组合物或肽组合/库,其包括本文限定的TAP(或编码所述肽的核酸)中的任一种或其任何组合。在一种实施方式中,组合物包括本文限定的TAP的任何组合(2、3、4、5、6、7、8、9、10种或更多种TAP的任何组合),或编码所述TAP的核酸的组合)。本公开涵盖包括本文限定的TAP的任何组合/亚组合的组合物。在另一实施方式中,组合或库可以包括一种或多种已知的肿瘤抗原。

[0175] 因此,在另一方面,本公开提供了组合物,该组合物包括本文限定的TAP或者一种或多种SLP(例如,其包括以下序列或由以下序列组成:SEQ ID NO:1-23和25-50,优选SEQ ID NO:1-23,诸如SEQ ID NO:1-17和22)中的任何一种或任何组合,以及表达MHC I类分子

(例如,上文公开的等位基因之一的MHC I类分子)的细胞。用于本公开的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、巨噬细胞和B细胞,已知它们在其细胞表面呈递蛋白抗原以被CD8⁺T淋巴细胞识别。例如,可以通过从外周血单核细胞中诱导DC,然后在体外、离体或体内接触(刺激)TAP来获得APC。APC还可以被激活以在体内呈递TAP,其中向受试者给药一种或多种本公开的TAP并且在受试者体内诱导呈递TAP的APC。短语“诱导APC”或“刺激APC”包括用一种或多种TAP或编码TAP的核酸接触或负载细胞,使得由MHC I类分子在其表面呈递TAP。如本文所述,根据本公开,可以间接负载TAP,例如使用包括TAP序列的较长肽/多肽(包括天然蛋白),然后在APC内对其进行处理(例如,通过蛋白酶)以在细胞表面生成TAP/MHC I类复合体。在用TAP负载APC并允许APC呈递TAP后,APC可以作为疫苗向受试者给药。例如,离体给药可以包括以下步骤:(a)从第一受试者收集APC,(b)将步骤(a)的APC与TAP接触/负载,以在APC表面形成MHC I类/TAP复合体;以及(c)向需要治疗的第二受试者给药负载有肽的APC。

[0176] 第一受试者和第二受试者可以是相同受试者(例如,自体疫苗),或者可以是不同受试者(例如,同种异体疫苗)。替代地,根据本公开,提供了本文所述的TAP(或其组合)用于制备用于诱导抗原呈递细胞的组合物(例如,药物组合物)的用途。此外,本公开提供了制备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将TAP或其组合与药学上可接受的负载体混合或配制的步骤。负载有本文限定的TAP中的任何一种或任何组合的细胞诸如表达MHC I类分子(例如,上述HLA分子的任一种)可以用于刺激/扩增CD8⁺T淋巴细胞,例如自体CD8⁺T淋巴细胞。因此,在另一方面,本公开提供了组合物,其包括本文限定的TAP中的任一种或任何组合(或编码其的核酸或载体);表达MHC I类分子的细胞,以及T淋巴细胞,更特别地CD8⁺T淋巴细胞(例如,包括CD8⁺T淋巴细胞的细胞群)。

[0177] 在一种实施方式中,组合物还包括缓冲剂、赋形剂、负载体、稀释剂和/或介质(例如,培养基)。在另外的实施方式中,缓冲剂、赋形剂、负载体、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、一种或多种负载体、一种或多种稀释剂和/或介质((多种)介质)。如本文所用,“药学上可接受的缓冲剂、赋形剂、负载体、稀释剂和/或介质”包括任何和所有溶剂、缓冲剂、粘合剂、润滑剂、填充剂、增稠剂、崩解剂、增塑剂、包衣、阻隔层制剂、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,它们是生理学相容的,不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无毒。此类介质和试剂用于药物活性物质的用途是本领域熟知的(Rowe et al., Handbook of pharmaceutical excipients, 2003, 4th edition, Pharmaceutical Press, London UK)。除非任何常规介质或试剂与活性化合物(肽、细胞)不相容,否则预期将其用于本公开的组合物中。在一种实施方式中,缓冲剂、赋形剂、负载体和/或介质是非天然存在的缓冲剂、赋形剂、负载体和/或介质。在一种实施方式中,本文限定的一种或多种TAP,或编码所述一种或多种TAP的核酸(例如,mRNA)包括在以下中或者与以下复合:囊泡诸如脂质囊泡或脂质体(例如阳离子脂质囊泡或脂质体)(参见,例如,Vitor MT et al., Recent Pat Drug Deliv Formul. 2013 Aug; 7(2):99-110)或者其他合适的负载体。

[0178] 在另一方面,本公开提供了组合物,该组合物包括本文限定的TAP或者一种或多种SLP(例如,其包括SEQ ID NO:1-23和25-50的序列或由SEQ ID NO:1-23和25-50的序列组成,优选SEQ ID NO:1-23诸如SEQ ID NO:1-17和22)(或编码所述一种或多种肽或者一种或

多种SLP的一种或多种核酸)中的一种或多种、任何一种或任意组合,以及缓冲剂、赋形剂、负载体、稀释剂和/或介质。对于包括细胞(例如,APC、T淋巴细胞)的组合物,该组合物包括实现维持活细胞的合适介质。此类介质的典型实例包括盐水溶液、Earl's平衡盐溶液(Life Technologies®)或PlasmaLyte®(BaxterInternational®)。在一种实施方式中,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TAP或疫苗载体的组合物或制剂,并且当向受试者给药时,其能够诱导针对其中存在的一种或多种TAP的免疫应答。用于在哺乳动物中诱导免疫应答的疫苗接种方法包括使用疫苗或疫苗载体通过疫苗领域已知的任何常规途径给药,例如经由粘膜(例如,眼、鼻内、肺、口腔、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)途径,或表面给药(例如,经由经皮递送系统,诸如贴剂)。在一种实施方式中,TAP(或其组合)与负载体蛋白缀合(缀合疫苗)以增加一种或多种TAP的免疫原性。因此,本公开提供了包括TAP(或其组合)、或编码TAP或其组合的核酸以及负载体蛋白的组合物(缀合物)。例如,一种或多种TAP或者一种或多种核酸可以与以下缀合或复合:Toll样受体(TLR)配体(参见,例如,Zom等人,Adv Immunol.2012,114:177-201)或聚合物/树枝状聚合物(参见,例如,Liu等人,Biomacromolecules.2013Aug 12;14(8):2798-806)。在一种实施方式中,免疫原性组合物或疫苗还包括佐剂。“佐剂”是指当添加至免疫原性剂诸如抗原(根据本公开的TAP、核酸和/或细胞)时,在暴露于混合物后非特异性地增强或加强宿主中对该剂的免疫应答的物质。目前用于疫苗领域的佐剂的实例包括(1)矿物盐(铝盐诸如磷酸铝和氢氧化铝、磷酸钙凝胶)、角鲨烯,(2)油基佐剂,诸如基于油乳剂和表面活性剂基制剂,例如MF59(微流化洗涤剂稳定的水包油乳液)、QS21(纯化皂苷)、AS02[SBAS2](水包油乳液+MPL+QS-21),(3)颗粒佐剂,例如病毒微体(并入流感血凝素的单层脂质体运载体)、AS04([SBAS4]铝盐与MPL)、ISCOMS(皂苷和脂质的结构化复合体)、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,单磷酸脂质A(MPL)、Detox(MPL+草分枝杆菌(M.Phlei)细胞壁骨架)、AGP[RC-529](合成的酰化单糖)、DC_cho1(能够自组织成脂质体的类脂免疫刺激剂)、OM-174(脂质A衍生物)、CpG基序(含有免疫刺激性CpG基序的合成寡核苷酸)、改性LT和CT(遗传修饰的细菌毒素以提供无毒佐剂效果),(5)内源性人免疫调节剂,例如hGM-CSF或hIL-12(可以作为蛋白或编码质粒给药的细胞因子)、Immudaptin(C3d串联阵列),和/或(6)惰性运载体,诸如金颗粒等。

[0179] 在一种实施方式中,一种或多种TAP或者一种或多种SLP(例如,其包括SEQ ID NO:1-23和25-50的序列,或由SEQ ID NO:1-23和25-50的序列组成,优选SEQ ID NO:1-23)(或编码所述一种或多种肽的核酸诸如mRNA)或者包括它们的组合物是冻干形式。在另一实施方式中,一种或多种TAP、一种或多种SLP或者一种或多种核酸或包括其的组合物在液体组合物中。在另外的实施方式中,在组合物中一种或多种TAP、一种或多种SLP或者一种或多种核酸的浓度为约0.01μg/mL至约100μg/mL。在另外的实施方式中,组合物中的一种或多种TAP、一种或多种SLP或者一种或多种核酸的浓度为约0.2μg/mL至约50μg/mL,约0.5μg/mL至约10、20、30、40或50μg/mL,约1μg/mL至约10μg/mL,或约2μg/mL。

[0180] 如本文所述,表达负载有或结合有本文限定的TAP中的任一种或任何组合的MHC I类分子的细胞,诸如APC,可用于体内或离体刺激/扩增CD8⁺T淋巴细胞。因此,在另一方面,

本公开提供了能够与本文提及的MHC I类分子/TAP复合体相互作用或结合的T细胞受体(TCR)分子,以及编码此类TCR分子的核酸分子,以及包括此类核酸分子的载体。根据本公开的TCR能够特异性地与TAP相互作用或结合,该TAP负载于MHC I类分子上或由MHC I类分子呈递,优选地在体外或体内的活细胞表面上。

[0181] 如本文所用,术语TCR是指具有可变结合结构域、恒定结构域、跨膜区和短细胞质尾的免疫球蛋白超家族成员;参见,例如,Janeway et al, Immunobiology: The Immune System in Health and Disease, 3rd Ed., Current Biology Publications, p.4:33, 1997), 其能够特异性结合与MHC受体结合的抗原肽。TCR可见于细胞表面上,并且通常由具有 α 和 β 链的异二聚体(也分别称为TCR α 和TCR β)组成。如免疫球蛋白, TCR链的细胞外部分(例如, α -链、 β -链)含有两个免疫球蛋白区、可变区(例如, TCR可变 α 区或V α 和TCR可变 β 区或V β ;通常基于N-末端处的Rabat编号的氨基酸1至116), 和与细胞膜相邻的一个恒定区(例如, TCR恒定结构域 α 或C α 和通常基于Rabat的氨基酸117至259, TCR恒定结构域 β 或C β , 通常基于Rabat的氨基酸117至295)。此外,如免疫球蛋白, 可变结构域含有由框架区(FR)分隔的互补决定区(CDR, 每条链中3个)。在某些实施方式中, TCR可见于T细胞(或T淋巴细胞)的表面上并且与CD3复合体缔合。

[0182] TCR, 且特别是编码本公开的TCR的核酸可以例如应用于基因转化/改性T淋巴细胞(例如, CD8⁺T淋巴细胞)或其他类型的淋巴细胞, 产生特异性识别MHC I类/TAP复合体的新T淋巴细胞克隆。在具体实施方式中, 从患者获得的T淋巴细胞(例如, CD8⁺T淋巴细胞)被转化以表达识别TAP的一种或多种TCR, 并向患者给药转化的细胞(自体细胞输血)。在具体实施方式中, 从供体获得的T淋巴细胞(例如CD8⁺T淋巴细胞)被转化以表达识别TAP的一种或多种TCR, 并向患者给药转化的细胞(同种异体细胞输血)。在另一实施方式中, 本公开提供了T淋巴细胞, 例如, 被编码TAP特异性TCR的载体或质粒转化/转染的CD8⁺T淋巴细胞。在另一实施方式中, 本公开提供了使用用TAP特异性TCR转化的自体或同种异体细胞治疗患者的方法。在某些实施方式中, 通过使用例如CRISPR、TALEN、锌指或其他靶向破坏系统替换内源性基因座(例如内源性TRAC和/或TRBC基因座), TCR在原代T细胞(例如, 细胞毒性T细胞)中表达。

[0183] 在另一实施方式中, 本公开提供了编码上述TCR的核酸。在另外的实施方式中, 核酸存在于载体, 诸如上述载体中。在另外的实施方式中, 核酸是mRNA分子。

[0184] 在又另外的实施方式中, 提供了肿瘤抗原特异性TCR在制造用于治疗癌症(例如, 结直肠癌)的自体或同种异体细胞中的用途。

[0185] 在一些实施方式中, 用本公开的组合物(例如, 药物组合物)治疗的患者在用抗肿瘤剂和/或免疫疗法(例如, CAR疗法)治疗之前或之后接受治疗。本公开的组合物包括: 针对TAP离体激活的同种异体T淋巴细胞(例如, CD8⁺T淋巴细胞); 负载有TAP的同种异体或自体APC疫苗; TAP疫苗以及用肿瘤抗原特异性TCR转化的同种异体或自体T淋巴细胞(例如, CD8⁺T淋巴细胞)或淋巴细胞。根据本公开的提供能够识别TAP的T淋巴细胞克隆的方法可以特异性靶向受试者(例如, 移植接受者), 例如ASCT和/或供体淋巴细胞输注(DLI)接受者中表达TAP的肿瘤细胞并且可以针对其产生。因此, 本公开提供了编码和表达能够特异性识别或结合TAP/MHC I类分子复合体的T细胞受体的CD8⁺T淋巴细胞。所述T淋巴细胞(例如, CD8⁺T淋巴细胞)可以是重组的(工程化的)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⁺T淋巴细胞的至少两种方法,其包括以下步骤: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TAP/MHC I类分子复合体(通常在细胞诸如APC表面表达)接触,这可以在体外或体内(即在给药了APC疫苗的患者中,其中APC负载有TAP,或在用TAP疫苗治疗的患者中)进行。使用与MHC I类分子结合的TAP的组合或库,可能产生能够识别多种TAP的CD8⁺T淋巴细胞群。替代地,肿瘤抗原特异性或靶向肿瘤抗原的T淋巴细胞可以通过克隆编码与MHC I类分子/TAP复合体特异性结合的TCR(更特别地 α 和 β 链)的一种或多种核酸(基因)来在体外或离体生产/产生(即工程化的或重组的CD8⁺T淋巴细胞)。编码本公开的TAP特异性TCR的核酸可以使用本领域已知的方法从针对TAP离体激活的T淋巴细胞(例如,用负载有TAP的APC)获得;或从针对肽/MHC分子复合体表现出免疫应答的个体获得。本公开的TAP特异性TCR可以在从移植接受者或移植供体获得的宿主细胞和/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),来将编码TCR α 和 β 链的核酸(转基因)引入到T细胞(例如,来自待治疗的受试者或其他个体)中。可以使用熟知的培养方法在体外扩增表达对TAP具有特异性的TCR的工程化CD8⁺T淋巴细胞。

[0186] 本公开提供了用于制备表达如本文所述的TCR的免疫效应细胞的方法。在一种实施方式中,该方法包括转染或转导免疫效应细胞,例如,从受试者,诸如患有结直肠癌(例如,结肠癌、直肠癌)的受试者分离的免疫效应细胞,使得免疫效应细胞表达如本文所述的一种或多种TCR。在某些实施方式中,免疫效应细胞分离自个体并且在体外没有进一步操作的情况下进行遗传修饰。然后可以将此类细胞直接重新给药于个体中。在另外的实施方式中,免疫效应细胞在经遗传修饰以表达TCR之前首先在体外被激活和刺激以增殖。在这方面,免疫效应细胞可以在经遗传修饰(即,如本文所述被转导或转染以表达TCR)之前或之后培养。

[0187] 在本文所述的免疫效应细胞的体外操作或遗传修饰之前,细胞来源可以获自受试者。特别地,与本文所述的TCR一起使用的免疫效应细胞包括T细胞。T细胞可以获自多种来源,包括外周血单核细胞(PBMC)、骨髓、淋巴结组织、脐带血、胸腺组织、感染部位的组织、腹水、胸腔积液、脾组织和肿瘤。在某些实施方式中,使用技术人员已知的任何数量的技术,诸如FICOLLTM分离,可以从受试者收集的单位血液中获得T细胞。在一种实施方式中,来自个体循环血液的细胞通过血浆分离置换法(单采法)获得。血浆分离置换法(单采法)的产物通常含有淋巴细胞,包括T细胞、单核细胞、粒细胞、B细胞、其他有核白细胞、红细胞和血小板。在一种实施方式中,可以洗涤通过血浆分离置换法(单采法)收集的细胞以移除血浆级分并将细胞置于适当的缓冲剂或介质中用于后续处理。在本发明的一种实施方式中,细胞用PBS洗涤。在替代的实施方式中,洗涤后的溶液缺乏钙并且可能缺乏镁或可能缺乏许多(如果并非全部)二价阳离子。如本领域普通技术人员将理解的,洗涤步骤可以通过本领域技术人员已知的方法来完成,诸如通过使用半自动流通式离心机。洗涤后,可以将细胞重悬于各种生物相容性缓冲剂或其他含或不含缓冲剂的盐水溶液中。在某些实施方式中,可以在细胞直接重悬培养基中移除血浆分离置换法(单采法)样品的非期望组分。在某些实施方式中,通过裂解红细胞并耗竭单核细胞,例如通过PERCOLLTM梯度离心,从外周血单核细胞(PBMC)分离T细胞。可以通过阳性或阴性选择技术进一步分离特定的T细胞亚群,诸如CD28⁺、CD4⁺、CD8⁺、CD45RA⁺和CD45RO⁺T细胞。例如,通过负选择的T细胞群的富集可以用针对负选择细胞特有

的表面标志物的抗体组合来完成。用于本文的方法是经由负磁免疫粘附或流式细胞术的细胞分选和/或选择,其使用针对存在于负选择的细胞上的细胞表面标志物的单克隆抗体混合物。例如,为了通过负选择富集CD8⁺细胞,单克隆抗体混合物通常包括针对CD14、CD20、CD11b、CD16、HLA-DR和CD4的抗体。流式细胞术和细胞分选也可用于分离目的细胞群以用于本公开。PBMC可以直接用于使用本文所述的方法用TCR进行遗传修饰。在某些实施方式中,在分离PBMC后,进一步分离T淋巴细胞,并且在某些实施方式中,细胞毒性和辅助T淋巴细胞两者可在遗传修饰和/或扩增之前或之后分选成幼稚、记忆和效应T细胞亚群。

[0188] 本公开提供了分离的免疫细胞,诸如CD8⁺T淋巴细胞,其被TAP(即与在细胞表面表达的MHC I类分子结合的TAP)或TAP的组合特异性诱导、激活和/或扩增(扩大)。本公开还提供了组合物,其包括能够识别根据本公开的TAP或其组合(即,一种或多种与MHC I类分子结合的TAP)的CD8⁺T淋巴细胞和所述一种或多种TAP。在另一方面,本公开提供了细胞群或细胞培养物(例如,CD8⁺T淋巴细胞群),其富集特异性识别如本文所述的一种或多种MHC I类分子/TAP复合体的CD8⁺T淋巴细胞。这种富集的群体可以通过使用细胞诸如表达负载有(例如呈递)本文公开的一种或多种TAP的MHC I类分子的APC进行特异性T淋巴细胞的离体扩增来获得。如本文所用的“富集”意指群体中肿瘤抗原特异性CD8⁺T淋巴细胞的比例相对于天然细胞群(即未经特异性T淋巴细胞的离体扩增步骤)显著更高。在另外的实施方式中,细胞群中TAP特异性CD8⁺T淋巴细胞的比例为至少约0.5%,例如至少约1%、1.5%、2%或3%。在一些实施方式中,细胞群中TAP特异性CD8⁺T淋巴细胞的比例为约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%。这种富集特异性识别一种或多种目的MHC I类分子/肽(TAP)复合体的CD8⁺T淋巴细胞的细胞群或培养物(例如,CD8⁺T淋巴细胞群)可用于基于肿瘤抗原的癌症免疫疗法,如下详述。在一些实施方式中,进一步富集TAP特异性CD8⁺T淋巴细胞群,例如使用基于亲和力的系统,诸如(共价或非共价地)负载有本文限定的一种或多种TAP的MHC I类分子的多聚体。因此,本公开提供了纯化或分离的TAP特异性CD8⁺T淋巴细胞群,例如,其中TAP特异性CD8⁺T淋巴细胞的比例为至少约50%、60%、70%、80%、85%、90%、95%、96%、97%、98%、99%或100%。

[0189] 在另一方面,本公开提供了抗体或其抗原结合片段或可溶性TCR,其与包括结合到HLA分子(诸如本文定义的HLA分子)的本文所述TAP的复合体特异性结合。这种抗体通常被称为TCR样抗体。如本文所用的术语“抗体或其抗原结合片段”是指任何类型的抗体/抗体片段,其包括单克隆抗体(包括全长单克隆抗体)、多克隆抗体、多特异性抗体、人源化抗体、CDR移植抗体、嵌合抗体和抗体片段,只要它们表现出所需的抗原特异性/结合活性。抗体片段包括全长抗体的一部分,通常是其抗原结合区或可变区。抗体片段的实例包括Fab、Fab'、F(ab')₂和Fv片段、双抗体、线性抗体、单链抗体分子(例如,单链Fv、scFv)、单域抗体(例如,来自骆驼科的抗体)、鲨NAR单域抗体以及由抗体片段、单链双抗体(scDb)、双特异性T细胞衔接体(BiTE)、双亲和重靶向分子(DART)、二价scFv-Fcs和三价scFv-Fcs形成的多特异性抗体。抗体片段也可以是指包括CDR或抗原结合结构域的结合部分,包括但不限于V_H区(V_H、V_H-V_H)、抗运载蛋白(anticalin)、佩普体(PepBody)、抗体-T细胞表位融合体(Troybody)或肽体(Peptibody)。在一种实施方式中,抗体或其抗原结合片段是单链抗体,优选单链Fv(scFv)。在一种实施方式中,抗体或其抗原结合片段包括至少一个恒定结构域,例如,轻链

和/或重链的恒定结构域,或其片段。在另外的实施方式中,抗体或其抗原结合片段包括以下的片段:抗体的恒定重链的可结晶片段(Fc)。在一种实施方式中,抗体或抗原结合片段是包括Fc片段的scFv(scFv-Fc)。在一种实施方式中,scFv组分通过接头(例如,铰链)与Fc片段连接。Fc区的存在有助于诱导针对肿瘤细胞的补体依赖性细胞毒性(CDC)或抗体依赖性细胞毒性(ADCC)应答。

[0190] 在一种实施方式中,抗体或其抗原结合片段是多特异性抗体或其抗原结合片段,诸如双特异性抗体或其抗原结合片段,其中多特异性抗体或抗体片段的抗原结合结构域的至少一种识别包括与HLA分子结合的本文所述TAP的复合体。在一种实施方式中,多特异性抗体或抗体片段的抗原结合结构域的至少一种识别免疫细胞效应分子。术语“免疫细胞效应分子”是指由免疫细胞表达的分子(例如,蛋白),并且其由多特异性抗体或抗体片段引起的接合导致免疫细胞的激活。免疫细胞效应分子的实例包括T细胞诸如CD8T细胞中的CD3信号传导复合体和NK细胞上的各种激活受体(NKG2D、KIR2DS、NKp44等)。在另外的实施方式中,多特异性抗体或抗体片段的抗原结合结构域的至少一种识别并接合T细胞中的CD3信号传导复合体(例如,抗CD3)。在另外的实施方式中,多特异性抗体或抗体片段是单链双抗体(scDb)。在另一实施方式中,scDb包括第一抗体片段(例如,scFv)和第二抗体片段(例如,scFv),该第一抗体片段与包括与HLA分子结合的本文所述TAP的复合体结合,该第二抗体片段与免疫细胞效应分子诸如T细胞中的CD3信号传导复合体(例如,抗CD3 scFv)结合并接合。此类构建体可用于例如诱导对表达由多特异性抗体或抗体片段识别的肿瘤抗原/MHC复合体的肿瘤细胞的细胞毒性T细胞介导的杀伤。抗体或其抗原结合片段也可用作嵌合抗原受体(CAR)来产生CAR T细胞、CAR NK细胞等。CAR将为所需抗原(例如MHC/TAP复合体)提供特异性的配体结合结构域(例如,抗体或抗体片段)与活化的细胞内结构域(或信号转导结构域)部分(诸如T细胞或NK细胞激活结构域)相结合,从而提供初级激活信号。能够与由肿瘤细胞表达的分子结合的抗体的抗原结合片段,以及更特别地是scFv,通常用作CAR中的配体结合结构域。因此,在另一方面,本公开提供了表达本文所述抗体或抗体片段(例如,scFv)的宿主细胞,优选免疫细胞诸如T细胞或NK细胞。

[0191] 在一种实施方式中,可溶性TCR是可溶性治疗性双特异性TCR(参见,例如,Robinson et al.,FEBS J.2021Nov;288(21):6159-6173;Dilchert et al.,Antibodies (Basel).2022May 10;11(2):34)。

[0192] 本公开另外涉及包括上述免疫细胞(CD8⁺T淋巴细胞,CAR T细胞)或TAP特异性CD8⁺T淋巴细胞群的药物组合物或疫苗。此类药物组合物或疫苗可以包括一种或多种药学上可接受的赋形剂和/或佐剂,如上所述。

[0193] 本公开另外涉及以下的用途:根据本公开的任何TAP或SLP(例如,包括SEQ ID NO:1-23和25-50的任何序列,或者由SEQ ID NO:1-23和25-50组成的任何序列,优选SEQ ID NO:1-23,诸如SEQ ID NO:1-17和22)、核酸、表达载体、T细胞受体、抗体/抗体片段、细胞(例如T淋巴细胞、APC、CAR T细胞)和/或组合物或其任何组合,作为药物或在药物制造中的用途。在一种实施方式中,药物用于治疗癌症,例如,癌症疫苗。本公开涉及根据本公开的任何TAP、SLP、核酸、表达载体、T细胞受体、抗体/抗体片段、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,用于治疗癌症的用途,例如,作为癌症疫苗。本文识别的TAP序列可用于产生合成肽,以用于i)体外引发和扩增肿瘤抗原特异性T细胞以注射到

肿瘤患者和/或ii)作为疫苗以诱导或增强癌症患者(诸如CRC患者)的抗肿瘤T细胞应答。在一种实施方式中,癌症(例如,CRC)表达本文所述的TAP的一种或多种。

[0194] 在另一方面,本公开提供了本文所述TAP(例如,SEQ ID NO:1-23和25-50,优选SEQ ID NO:1-23,诸如SEQ ID NO:1-17和22)或其组合(例如,肽库),或编码一种或多种TAP的一种或多种核酸,作为用于治疗受试者的癌症(诸如CRC)的疫苗的用途。本公开还提供了本文所述的TAP或其组合(例如,肽库),或编码一种或多种TAP的一种或多种核酸作用于治疗受试者的癌症诸如CRC的疫苗。在一种实施方式中,受试者是TAP特异性T淋巴细胞(例如,CD8⁺T淋巴细胞)的接受者。因此,在另一方面,本公开提供了治疗癌症诸如CRC的方法(例如,减少肿瘤细胞的数量、杀死肿瘤细胞),所述方法包括向有此需要的受试者给药(输注)有效量的识别(即表达结合其的TCR)一种或多种MHC I类分子/TAP复合体(在细胞(诸如APC)的表面表达)的T淋巴细胞(例如,CD8⁺T淋巴细胞)。在一种实施方式中,该方法进一步包括,在给药/输注所述CD8⁺T淋巴细胞之后,向所述受试者给药有效量的以下项:TAP或其组合,或者编码一种或多种TAP的一种或多种核酸和/或表达负载有一种或多种TAP的一种或多种MHC I类分子的细胞(例如,APC,诸如树突状细胞)。在又另外的实施方式中,该方法包括向有此需要的受试者给药治疗有效量的负载有一种或多种TAP的树突状细胞。在又另外的实施方式中,该方法包括向有此需要的患者给药治疗有效量的同种异体或自体细胞,该细胞表达与由MHC I类分子呈递的TAP结合的重组TCR。

[0195] 在另一方面,本公开提供了识别负载有(呈递)TAP或其组合的一种或多种MHC I类分子的T淋巴细胞(例如,CD8⁺T淋巴细胞)用于治疗受试者的癌症(例如,减少肿瘤细胞数量、杀死肿瘤细胞)诸如CRC的用途。在另一方面,本公开提供了识别负载有(呈递)TAP或其组合的一种或多种MHC I类分子的T淋巴细胞(例如,CD8⁺T淋巴细胞)在制备/制造用于治疗受试者的癌症(例如,减少肿瘤细胞数量、杀死肿瘤细胞)诸如CRC的药物的用途。在另一方面,本公开提供了识别负载有(呈递)TAP或其组合的一种或多种MHC I类分子的T淋巴细胞(例如,CD8⁺T淋巴细胞),用于治疗受试者的癌症(例如,减少肿瘤细胞数量、杀死肿瘤细胞)诸如CRC的用途。在另外的实施方式中,该用途还包括,在使用所述TAP特异性CD8⁺T淋巴细胞之后,使用有效量的TAP(或其组合)和/或表达负载有(呈递)TAP的一种或多种MHC I类分子的细胞(例如,APC)。

[0196] 本公开还提供了在受试者中产生针对表达负载有本文公开的任何TAP(例如,SEQ ID NO:1-23和25-50,优选SEQ ID NO:1-23,诸如SEQ ID NO:1-17和22)或其组合的人I类MHC分子的肿瘤细胞(例如,结直肠癌细胞)的免疫应答的方法,该方法包括给药特异性识别负载有TAP或TAP的组合的I类MHC分子的细胞毒性T淋巴细胞。本公开还提供了特异性识别负载有本文公开的任何TAP或TAP的组合的I类MHC分子的细胞毒性T淋巴细胞用于产生针对表达负载有TAP或其组合的人I类MHC分子的肿瘤细胞的免疫应答的用途。

[0197] 在一种实施方式中,癌症是结直肠癌。在另外的实施方式中,癌症是结肠癌。在一种实施方式中,癌症是直肠癌。在一种实施方式中,结直肠癌的特征在于微卫星不稳定性(MSI)。在一种实施方式中,结直肠癌的特征在于微卫星稳定性(MSS)。在一种实施方式中,结直肠癌的特征在于RAS(例如,KRAS)和/或RAF突变。在另一实施方式中,结直肠癌对化疗具有抗性/难治性。在另一实施方式中,结直肠癌对EGFR抑制剂具有抗性/难治性。

[0198] 在一种实施方式中,本文所述的方法或用途进一步包括在治疗/使用之前确定由

患者表达的HLA I类等位基因,并给药或使用与由患者表达的一种或多种HLA I类等位基因结合的TAP。例如,如果确定患者表达HLA-A2*01和HLA-B07*02,则可以在患者中给药或使用SEQ ID NO:6(与HLA-A2*01结合)和/或SEQ ID NO:2、21、24和/或50(与HLA-B07*02结合)的TAP的任何组合。

[0199] 在一种实施方式中,根据本公开的TAP、SLP、核酸、表达载体、T细胞受体、抗体/抗体片段、细胞(例如,T淋巴细胞、CAR T或NK细胞、APC)和/或组合物或其任何组合,可以与一种或多种额外的活性剂或疗法组合使用以治疗癌症(例如,CRC),诸如化疗(例如,长春花生物碱、破坏微管形成的剂(诸如秋水仙碱及其衍生物)、抗血管生成剂、治疗性抗体、EGFR靶向剂、酪氨酸激酶靶向剂(诸如酪氨酸激酶抑制剂)、过渡金属复合体、蛋白酶体抑制剂、抗代谢药物(诸如核苷类似物)、烷化剂、铂基剂、蒽环类抗生素类、拓扑异构酶抑制剂、大环内酯类、类视黄醇类(诸如全反式视黄酸或其衍生物)、格尔德霉素(geldanamycin)或其衍生物(诸如17-AAG)、外科手术、免疫检查点抑制剂或免疫治疗剂(例如,PD-1/PD-L1抑制剂(诸如抗PD-1/PD-L1抗体)、CTLA-4抑制剂(诸如抗CTLA-4抗体)、B7-1/B7-2抑制剂(诸如抗B7-1/B7-2抗体)、TIM3抑制剂(诸如抗TIM3抗体)、BTLA抑制剂(诸如抗BTLA抗体)、CD47抑制剂(诸如抗CD47抗体)、GITR抑制剂(诸如抗GITR抗体)、针对肿瘤抗原的抗体(例如,抗CD19、抗CD22抗体)、基于细胞的疗法(例如,CAR T细胞、CAR NK细胞)以及细胞因子(诸如IL-2、IL-7、IL-21和IL-15)。在一种实施方式中,根据本公开的TAP、核酸、表达载体、T细胞受体、细胞(例如,T淋巴细胞、APC)和/或组合物与免疫检查点抑制剂组合给药/使用。在一种实施方式中,根据本公开的TAP、核酸、表达载体、T细胞受体、细胞(例如,T淋巴细胞、APC)和/或组合物与用于治疗CRC的一种或多种疗法(例如,外科手术、化疗(例如,使用5-氟尿嘧啶、卡培他滨(capecitabine)、奥沙利铂(oxaliplatin)、伊立替康(irinotecan)、雷替曲塞(raltitrexed)、曲氟尿苷(trifluridine)、替匹嘧啶(tipiracil))、放射疗法、贝伐珠单抗(bevacizumab)、西妥昔单抗(cetuximab)、帕木单抗(panitumumab)、瑞戈非尼(regorafenib))组合给药/使用。

[0200] 可以在给药根据本公开的TAP、SLP、核酸、表达载体、T细胞受体、抗体/抗体片段、细胞(例如,T淋巴细胞、CAR T或NK细胞、APC)和/或组合物之前、同时或之后给药额外的疗法。

[0201] 实施例

[0202] 通过以下非限制性实施例另外详细说明本公开。

[0203] 实施例1:实验程序

[0204] 细胞系。从美国典型培养物保藏中心(American Type Culture Collection, ATCC)获得四种结直肠癌细胞系[COLO 205(ATCC[®] CCL-222[™])、HCT 116(ATCC[®] CCL-247[™])、RK0(ATCC[®] CRL-2577[™])、SW620[SW-620](ATCC[®] CCL-227[™])]和一种正常胎儿小肠细胞系[HIEC6(ATCC[®] CRL3266[™])]。使COLO205、HCT116和SW620在补充有10%胎牛血清(FBS)的RPMI-1640(Gibco)中生长,使RK0在补充有10% FBS的伊格尔最低必需培养基(EMEM)(ATCC[®])中生长,并且使HIEC-6在补充有20mM HEPES(Gibco)、10mM GlutaMAX[®](Gibco)、10ng/mL表皮生长因子(EGF)(Gibco)和FBS至最终浓度为4%的OptiMEM[®] 1减血清培养基(Gibco)中生长。将所有细胞保持在37℃和5% CO₂下。为了收集,用温PBS漂洗细

胞,之后将其用TrypLEMTExpress酶(1X)(Gibco)在37℃和5% CO₂下胰酶消化5-15分钟。然后将收获的材料以1000rpm旋转5分钟,用温PBS漂洗一次,然后重悬于冰冷的PBS中。细胞计数后,使 2×10^8 个CRC细胞的拷贝呈团块状并在-80℃下冷冻直至进一步使用。根据制造商的说明,使用W6/32抗HLA I类抗体(BioXCell)通过Qifikit™(Agilent)测定CRC细胞系的MHC I类表面密度。

[0205] 原代组织。由匹配的结肠腺癌肿瘤和正常邻近组织(NAT)组成的六对原发性人样品购自Tissue Solutions。组织样品取自接受作为一线治疗的外科手术的患者,并在液氮中快速冷冻。关于原代组织样品的更多信息可在表2中找到。

[0206] RNA提取。对于细胞系的RNA提取,收集1-2百万个细胞并用冰冷的PBS洗涤一次。然后将细胞重悬于Trizol™(Invitrogen)中。对于细胞系和原代组织样品,使用制造商推荐的全Prep™DNA/RNA/miRNA通用试剂盒(Qiagen)或RNeasy™Mini试剂盒(Qiagen)分离总RNA。

[0207] RNA测序。将500ng的总RNA用于文库制备。使用2100Bioanalyzer™系统(Agilent Technologies)上的Bioanalyzer™RNA 6000纳米测定对RNA质量控制进行了评估,并且所有样品的RIN均高于8。使用KAPA mRNAseq Hyperprep™试剂盒(Roche)制备文库。用Illumina双端标志UMI(IDT)进行连接。在BioAnalyzer™DNA1000芯片上进行验证并通过QuBit和qPCR进行定量后,将文库合并至等摩尔浓度,并使用Nextseq™高输出150(2x75bp)循环试剂盒用Illumina Nextseq500进行测序。细胞系和组织样品分别生成了平均1.29亿和9500万个双端PF读段。在免疫学和癌症研究所(Research in Immunology and Cancer,IRIC)的基因组平台上进行文库制备和测序。

[0208] 生物信息学分析。使用Trimmomatic版本0.35对序列进行修剪(17),并使用STAR版本2.7.1a与参考人类基因组版本GRCh38(来自Gencode版本33的基因注释,基于Ensembl 99)进行比对(18)。为了获得这些链RNA文库的归一化基因和转录水平表达(以TPM值表示),基因表达既可以直接从STAR获得读段计数,也可以使用RSEM(19)计算获得。

[0209] HLA基因分型。使用OptiType(<https://github.com/FRED-2/OptiType>)对细胞系和组织进行HLA基因分型(20)。

[0210] 微卫星不稳定性预测。利用MSIsensor程序使用配对肿瘤和NAT(<https://github.com/ding-lab/msisensor>)对原代肿瘤样品的MSI状态进行预测(21)。

[0211] 差异表达分析。将DESeq2版本1.22.2(22)用于归一化基因读段计数并计算肿瘤和正常样品之间的差异表达。主成分分析(PCA)是使用前两个最有意义成分的归一化对数读段计数生成的。对于细胞系的差异表达分析,计算与正常细胞系(HIEC-6)相比四种CRC细胞系的平均表达之间的倍数变化。使用Metascape工具将显著差异表达基因(DEG)(具有padj低于0.05的基因)和具有 $|\log_2$ 倍数变化 >1 的基因视为GO术语(23)。对于组织的成对差异表达分析,将TPM归一化值用于比较肿瘤/NAT对。由于只对组织的单个复制品进行了测序,而不是通过经调整的p值进行过滤,因此只选择了在所有六名受试者中对于GO术语分析具有 $|\log_2$ 倍数变化 >1 的显著差异表达的基因。当检查MSS和MSI组织之间的差异表达基因时,应用了相同的倍数变化阈值。对于MSIDEG的GO术语分析,选择了在两种MSI组织中均排他性差异表达的基因(即在任何MSS组织中不被认为是DEG)。对于MSSDEG的GO术语分析,如果基因在3个或更多MSS组织中差异表达,则考虑这些基因。

[0212] 组织样品的转录组分析。如前所述,确定组织样品转录组中各种生物型的比例

(24)。简而言之,在由Kallisto对Ensembl注释转录物进行定量和比对后(19),使用Kallisto索引对转录物和重复元件进行注释,该Kallisto索引包括补充有来自USCS表浏览器GRCh38重复屏蔽数据库的遗传重复识别的Ensembl注释转录物(25)。

[0213] 突变谱/“遗传变体注释”。使用SNPEff (<https://pcingola.github.io/SnpEff/#snpeff>) 对细胞系和原始活检进行遗传变体识别(Genetic variant calling)(26)。

[0214] 数据库生成。如前所述构建全球癌症数据库(16)。简而言之,使用Trimmomatic版本0.35对RNA-seq读段进行修剪(17),并使用STAR版本2.7.1a与参考人基因组版本GRCh38(来自Gencode版本33的基因注释,基于Ensembl 99)进行比对(18)。将Kallisto(<https://pachterlab.github.io/kallisto>)用于以TPM定量转录物表达(19)。通过将用Freebayes(<https://github.com/ekg/freebayes>)识别出的单核苷酸变体(质量>20)整合到PyGeno中来构建样品特异性外显子组(27)。然后将具有TPM>0的注释开放阅读框在计算机中进行翻译并以fasta格式添加至典型蛋白质组中。为了生成癌症特异性蛋白质组,将RNA-seq读段切割成33个核苷酸的序列,称为kmer,并仅保留在细胞系的mTEC或组织的匹配NAT中存在的<2的kmer。将重叠的kmer被组装成contigs,然后在计算机上对其进行3-框架翻译。值得注意的是,通过kmer方法生成的短肽序列随后被串联成大约一万个氨基酸的较长序列。使用“JJ”序列作为分隔符串联这些肽,该分隔符被PeaksX+软件内部识别以便在序列出现时分割该序列。然后,将典型蛋白质组和癌症特异性蛋白质组连串联以创建全球癌症数据库。细胞系数据库平均由 3.38×10^6 个序列组成。

[0215] MAP的分离。将CRC细胞系团块样品(每次复制 2×10^8 个细胞,每种细胞系4份复制品)用PBS重悬至2mL,并且然后加入2mL冰冷的2X裂解缓冲剂(1%w/v CHAPS)进行溶解。将肿瘤和正常邻近组织样品(455mg至693mg之间)切成小块(立方体,大小~3mm),并加入5mL的冰冷的含蛋白抑制剂混合物(Sigma, cat#P8340-5mL)的PBS。首先使用设定为速度20000rpm的Ultra Turrax T25均质器(IKA-Labortechnik)将组织均质两次,持续20秒,并然后使用速度设定为25000rpm的Ultra Turrax T8均质器(IKA-Labortechnik)均质20秒。然后,将550μL冰冷的10X裂解缓冲剂(5%w/v CHAPS)添加到各个样品中。在4℃下转动孵育60分钟后,将组织样品和CRC细胞系样品在4℃下以10000g旋转30分钟。将上清液转移至含1mg的W6/32抗体共价交联的蛋白A磁珠粒的新管中,并如前所述免疫沉淀MAP(28)。然后使用Speed-Vac干燥MAP提取物,并在MS分析前保持冷冻。

[0216] TMT标记。将MAP提取物重悬于pH 8.1的200mM Hepes缓冲剂中。如下将在无水乙腈中的50μg的TMT试剂(Thermo Fisher Scientific)加入样品中:将CRC细胞系复制品用TMT6plex (lot#UG287166) 通道TMT6-126-129标记;将组织样品用TMT10plex (lot#UH285228) -126(NAT) 和-127N(肿瘤) 标记。将样品轻轻涡旋并在室温下反应1.5小时。然后在室温下用50%羟胺猝灭样品30分钟,然后用4%FA/H₂O稀释。将CRC细胞系复制品和单独NAT-肿瘤对进行组合。然后将样品在自制C18膜(Empore)柱上脱盐,并在-20℃下储存直至进样。为了定量原代组织样品中的目的MAP,将浓度为0.75至192fmol的合成肽用TMT10plex通道-128N、-128C、-129N、-129C、-130N、-130C、-131标记,而NAT和CRC组织分别用-126和127N标记。请注意,通道-127C是空的,以评估通道之间的污染。

[0217] 液相色谱-串联MS分析。将干燥的肽提取物重悬于4% FA中,并在EASY-nLC II系统上使用106-分钟梯度从0%至30% ACN(0.2% FA)和600nL/min的流速上样至自制C18分

析柱(20cm x 150 μ m i.d.,填充有C18Jupiter Phenomenex)上。使用Orbitrap Exploris 480光谱仪(Thermo Fisher Scientific)在正离子模式下使用2.8kV的Nanoflex源对样品进行分析。每个完整的MS谱,以240,000分辨率采集,随后是20个MS/MS谱,其中选择最丰富的多电荷离子进行MS/MS测序,分辨率为30,000,自动增益控制目标为100%,进样时间为700ms,以及碰撞能量为40%。

[0218] MAP识别。使用PeaksX+软件(Bioinformatics Solutions Inc.)进行数据库搜索(29)。将前体质量和碎片离子的误差容差分别设置为10.0ppm和0.01Da。使用了非特异性消化模式(digest mode)。将TMT6plex或10plex设定为固定PTM,并且可变修饰包括磷酸化(STY)、氧化(M)、脱酰胺(NQ)和TMT6plex或10plex STY。然后将峰搜索加载到MAPDP(30)中,该MAPDP用于应用以下过滤器:基于NetMHCpan-4.1b预测,使用5% FDR,选择长度为8-11个氨基酸的肽,其中等级洗脱配体阈值 $\leq 2\%$ 。

[0219] RNA-Seq数据中MAP编码序列的定量。如前所述,在RNA-seq数据中定量MAP编码序列(MCS)。简而言之,用内部python脚本(保藏于Zenodo,DOI:3739257)将MCS反向翻译成所有可能的核苷酸序列。然后用GSNAP(32)将核苷酸序列映射到基因组上以确定能够编码给定MAP的所有可能的基因组位置。MCS也被映射到转录组上以解释重叠了剪接位点的MAP,然后用GSNAP将对应于这些MAP的转录组部分也映射到参考基因组上。对于感兴趣的MAP,对包括MC的所有读段进行基因组比对。过滤GSNAP输出以仅保持序列和参考之间的完美匹配,从而生成包含能够编码给定MAP的所有可能的基因组区域的文件。对每个所需RNA-Seq样品(诸如CRC和NAT、GTEx或TCGA样品)中其各自基因组位置含有MCS的读段进行求和,上述所需RNA-Seq在参考基因组上用STAR比对。最后,对给定MAP的所有读段计数进行求和,并对每个目的样品中测序的读段总数进行归一化,以获得每亿读段(RPHM)计数。

[0220] MAP来源转录物的确定。为了研究来源于某些转录生物型的组织样品MAP的比例,确定了基于每亿kmer(KPHM)定量的最丰富的推定的来源转录物。对于来自癌症特异性(kmer)数据库的肽,将MCS反向翻译成所有可能的核苷酸序列,并识别出能够编码给定MAP的所有可能的基因组区域(参见上文“RNA-Seq数据中MAP编码序列的定量”)。最后,使用Kallisto确定在该位置表达最多的转录物,然后将其指定为给定肽的最可能的转录物。将具有大于一个推定来源转录物的肽从分析中排除。

[0221] TSA候选者的识别。TSA候选者是通过严格的TSA识别流程确定的。首先,MAP经历了其中将肽序列成员从蛋白数据库中检索出并用于提取每个肽的核苷酸序列的肽分类。使用Jellyfish 2.2.3(使用-C选项)将每种癌症和正常样品的RNA-Seq数据转化为24个核苷酸长的k-mer数据库,并用于查询每种TSA候选者编码序列的24个核苷酸长的k-mer集。使用k-mer集的最小出现次数(kmin)来估计与给定肽编码序列完全重叠的读段的数量,因为一般来说,一个k-mer总是来源于单个RNA-Seq读段。然后使用以下公式将该kmin值转化为每 10^8 个经测序的读段中检测到的若干个k-mer(kphm): $kphm = (kmin \times 10^8) / rtot$,rtot代表给定RNA-Seq实验中测序的读段总数。只有当肽的RNA编码序列在癌症中的表达是正常情况下的至少10倍高(细胞系的合并mTEC样品,组织的匹配NAT),并且在正常情况下的表达 $< 2KPHM$ 时,肽才被保留。随后的过滤去除了具有不可区分的异亮氨酸/亮氨酸变体的任何肽;只有当表达最多的变体符合上述标准时,才保留具有IL变体的肽。如上所述,在RNA-seq数据中定量剩余肽的MCS,并且仅当它们在mTEC和其他正常组织(GTEx)中的表达 $< 8.55RPHM$ 时才

被保留。每种肽的基因组定位通过使用BLAT (<https://genome.ucsc.edu/cgi-bin/hgBlat>, Kent WJ. BLAT-the BLAST-like alignment tool. *Genome Res.* 2002 Apr; 12 (4) : 656-64. PMID: 11932250) 将包含每种MCS的读段映射至参考基因组 (GRCh38.99) 来指定。如果基因组定位不明确或如果肽映射至高变区 (HLA、Ig或T细胞受体 (TCR) 基因), 则排除肽。最后, 手动验证剩余候选者的MS/MS谱。如果肽的氨基酸序列不同于参考序列, 并且如果突变不是已知的种系多态性, 则将肽分类为mTSA。如果肽在肿瘤中是相比于正常情况下的过表达 ≥ 10 倍, 并且在mTEC中的过表达 ≤ 0.2 KPHM (以及在组织中的情况下为NAT), 则被归类为aeTSA, 以及如果肽在癌症中的过表达 ≥ 10 倍, 但在mTEC和/或NAT中的表达 > 0.2 KPHM, 则被归类为TAA。最终, 通过从kallisto定量文件中选择最高表达的肽重叠转录物来确定每种TSA/TAA来源的转录物 (参见数据库生成部分)。

[0222] 瘤间共享。为了检查TSA和TAA序列在其他CRC肿瘤中的瘤间分布, 确定了来自TCGA的151例结肠腺癌 (COAD) 样品中肽编码序列的 $\log(\text{RPHM}+1)$ 表达 (参见“RNA-Seq数据中MAP编码序列的定量”)。

[0223] 免疫原性预测。用R软件包Repitope v3.0.1 (<https://github.com/masato-ogishi/Repitope>) 来测定目的MAP的预测免疫原性 (34)。

[0224] TSA肽候选者的验证。从Genscript获得TSA和选择TAA序列的合成肽。将合成肽溶解在DMSO中至浓度为 $1\text{nmol}/\mu\text{L}$, 并将所有合成肽组合在浓度为 $10\text{pmol}/\mu\text{L}$ 的储备溶液中。将储备溶液在自制C18膜 (Empore) 柱上以 150pmol 的等份脱盐, 并使用Speed-Vac干燥。将干燥的肽提取物用所述的TMT10plex通道进行标记 (参见“TMT标记”部分), 脱盐, 并在Speed-Vac中干燥。将经标记的合成肽重悬于4% FA中, 并将 1pmol 的各合成肽在EASY-nLC Ii系统上使用76-分钟梯度从0%至30% ACN (0.2% FA) 和 $600\text{nL}/\text{min}$ 的流速上样至自制C18分析柱 ($20\text{cm} \times 150\mu\text{m}$ i.d., 填充有C18Jupiter Phenomenex) 上。使用Orbitrap Exploris 480光谱仪 (Thermo Fisher Scientific) 在正离子模式下使用 2.8kV 的Nanoflex源对样品进行分析。每种完整的MS谱都以 120000 的分辨率采集, 并使用包含列表选择具有40%碰撞能量和 $1\text{m}/z$ 隔离窗口的碎片离子。以 30000 的分辨率采集MS/MS。如前所述计算MS/MS相关性 (17)。简而言之, 使用pyteomics v4.0.1 (<https://bitbucket.org/levitsky/pyteomics>) 计算预期的肽片段, 并识别出可重复检测的肽片段。这些片段的根标度强度 (Root scaled intensities) 在内源和合成肽扫描对之间相关, 并且使用SciPy v1.2.1 (<https://www.scipy.org/>) 计算皮尔逊相关系数、p值和置信区间。使用spectrum_utils v0.2.1 (https://github.com/bittremieux/spectrum_utils) 为每个肽生成了具有最低p值的扫描对的镜像图。

[0225] 相对定量。为了相对定量原代组织样品中的目的MAP, 将分别用TMT 10plex-129N、130N和131N标记的浓度为10、100或 1000fmol 的合成肽掺入分别用TMT10plex-126和127N标记的NAT和CRC组织样品的剩余纯化MAP中。注意通道TMT 10plex-127C是空的, 以评估污染。使用Orbitrap Fusion Tribrid光谱仪 (Thermo Fisher Scientific) 在正离子模式下使用 2.4kV 的Nanoflex源对样品进行分析。对于同步前体选择MS3 (SPS-MS3), 使用目的肽的包含列表, 以 $300-1000\text{m}/z$ 的范围、 120000 的轨道阱分辨率、 $5.0\text{e}5$ 的自动增益控制 (AGC) 和 50ms 的最大进样时间来采集全MS扫描。离子阱中使用了 3s 最高速度的MS2方法, 其中隔离窗口为 $0.4\text{m}/z$, 碰撞诱导的解离为35%, 离子阱扫描速率模式为“正常”, AGC目标为 $2.0\text{e}4$, 以及最

大进样时间为50ms。随后选择八个同步前体离子进行MS3采集,扫描范围为110-500m/z,轨道阱分辨率为50000,AGC为1.0e5,最大进样时间为300ms,隔离窗口为2.0m/z,以及HCD碰撞能量为65%。使用Xcalibur版本4.4(Thermo Fisher Scientific, Inc)控制LC-MS仪器。将前体质量和碎片离子的误差容差分别设置为10.0ppm和0.5Da。使用了非特异性消化模式(digest mode)。TMT10plex设定为固定PTM,并且可变修饰包括磷酸化(STY)、氧化(M)、脱酰胺(NQ)和TMT10plex STY。为了定量,对PSM进行过滤以排除TMT10plex-127C通道中具有污染物的PSM,并选择第70个强度百分位数内的PSM。手动检查所有相关通道的MS2前体谱和强度谱,以选择用于定量的肽。使用高质量PSM的平均TMT10plex-127N和TMT10plex-126强度来计算每种肽的强度比。

[0226] 数据分析和可视化。图1是用BioRender.com生成的。大多数其他图是用Python v3.7.6、R v3.6.3或Origin(Pro) 2019b创建的。R软件包包括:Repitope v3.0.1(<https://github.com/masato-ogishi/Repitope>) (34)、UpsetR v1.4.0(<https://github.com/hms-dbmi/UpSetR>) (35)、GSVA v1.38.2(<https://github.com/rcastelo/GSVA>) (36)、ESTIMATE v1.0.13(<https://bioinformatics.mdanderson.org/estimate/>) (37)。

[0227] 实验设计和统计学原理。为了有效阐明结直肠癌的MHC I免疫肽组,选择了4种CRC细胞系,并获得了由匹配的肿瘤和正常邻近组织(NAT)组成的来自人受试者的6种样品。将NAT用作健康组织的近似物,因为它是对每种肿瘤最有效的对照。由于无法获得对于细胞系的匹配的样品,因此在创建全球癌症数据库时使用了6种mTEC样品池,以获得广泛的近似正常RNA表达。使用双样品t检验来确定所有p值实例,但在确定免疫原性评分的显著性时除外,在这种情况下,使用曼-惠特尼检验,因为数据不具有正态分布,如夏皮罗检验所确定的。对于t检验,进行f检验以确定数据集是否有显著变化;如果是,则使用假设变异的t检验,否则使用假设无变异的t检验。对于CRC衍生的细胞系,制备了 2×10^8 个细胞的四份技术复制品,在注射前对其进行TMT标记和多重化(multiplexed)。由于组织材料有限,注射来自原代样品的一半纯化MAP以获得全球免疫肽组学数据,并且将剩余的样品用于合成肽的靶向分析,以确认推定的TSA的序列和丰度并选择TAA。为了选择高质量的PSM进行定量,排除了低强度或空TMT通道中有污染物的PSM。此外,只有具有有利的MS2前体和强度分布的肽才被定量。

[0228] 实施例2:使用蛋白质组学方法进行免疫肽组学分析

[0229] 为了确定结直肠癌的免疫肽组的组成,分析了由4种结直肠癌来源的细胞系和6组原发性腺癌样品组成的样品集合,该原发性腺癌样品由匹配的肿瘤和正常邻近组织组成(表1和2)。双端RNA测序(RNA-seq)允许创建全球癌症数据库,该数据库由典型癌症蛋白质组以及每个样品的癌症特异性蛋白质组组成,通过生成癌症特异性kmer来创建,这些kmer被翻译成3个阅读框以涵盖来自任何基因组来源的非典型序列(图1)。髓质胸腺上皮细胞(mTEC)在胸腺中呈递外周抗原并介导自身反应性T细胞的负选择(38)。在CRC来源的细胞系的情况下,在减去mTEC来源的序列后获得了癌症特异性的kmer,其近似于这些序列在健康组织中的表达。对于原代组织样品,从匹配的NAT中减去序列后生成癌症特异性kmer。这种方法能够确定在肿瘤中表达并在同一个体的健康结肠组织中未观察到的序列。除了数据库构建之外,RNA-seq数据还用于转录组分析,包括GO术语分析、免疫浸润研究、突变谱分析和转录物丰度测定(图1)。

[0230] 表1.CRC衍生的细胞系的描述。

细胞系	组织	形态学	疾病	生物标志物	MHC I 分子/细胞	HLA 基因分型	目的突变
Colo 205	结肠；来源于转移性部位：腹水	上皮	杜克氏 D 型，结肠腺癌	MSS, CIMP	$1.44 \times 10^5 \pm 0.00282 \times 10^5$	HLA-A*01:01 HLA-A*02:01 HLA-B*07:02 HLA-B*08:01 HLA-C*07:01 HLA-C*07:02	BRAF (V600E), SMAD4, TP53
HCT 116	结肠	上皮	大肠癌	MSI, CIMP	$5.07 \times 10^5 \pm 0.30 \times 10^5$	HLA-A*01:01 HLA-A*02:01 HLA-B*18:01 HLA-B*45:01 HLA-C*05:01 HLA-C*07:01	RAS (G13D), PI3CA, CDKN2A, CTNNB1 (B- 联蛋白)
RKO	结肠	上皮	癌	MSI, CIMP	$2.82 \times 10^5 \pm 0.11 \times 10^5$	HLA-A*03:01 HLA-B*18:01 HLA-C*07:01	BRAF (V600E), PI3CA
SW6 20	结肠；来源于转移性部位：淋巴结	上皮	杜克氏 C 型，结肠腺癌	MSS, CIN	$1.69 \times 10^5 \pm 0.0017 \times 10^5$	HLA-A*02:01 HLA-A*24:02 HLA-B*07:02 HLA-B*15:18 HLA-C*07:02 HLA-C*07:04	APC, RAS (G12V), SMAD4, TP53

[0232] 表2.原代肿瘤和匹配NAT的描述

[0233]	样品 ID	性 别	年 龄	种族背景	基质	诊断
	S1_N				结肠	NAT
	S1_T	F	73	高加索人	盲肠	癌症
	S2_N				结肠	NAT
	S2_T	M	60	高加索人	乙状结肠	癌症
	S3_N				结肠	NAT
	S3_T	F	63	高加索人	乙状结肠	癌症
	S4_N				结肠	NAT
	S4_T	F	85	高加索人	乙状结肠	癌症
	S5_N				结肠	NAT
	S5_T	F	43	高加索人	升结肠	癌症
	S6_N				结肠	NAT
	S6_T	F	48	高加索人	乙状结肠	癌症
[0234]	样品 ID	组织学 诊断	阶段	肿瘤含 量%	目的 突变	HLA
	S1_N					HLA-A*24:02 HLA-B*07:02 HLA-B*35:01 HLA-C*04:01 HLA-C*07:02

[0235]	S1_T	腺癌	IIC	100	KRAS G12D	
	S2_N					HLA-A*02:01 HLA-A*03:02 HLA-B*27:05 HLA-B*58:01 HLA-C*02:02 HLA-C*07:01
	S2_T	腺癌	IIA	95		
	S3_N					HLA-A*01:01 HLA-A*32:01 HLA-B*38:01 HLA-B*50:01 HLA-C*06:02 HLA-C*12:03
	S3_T	腺癌	IIA	100	KRAS Q61H	
	S4_N					HLA-A*01:01 HLA-A*11:01 HLA-B*15:01 HLA-B*57:01 HLA-C*03:03 HLA-C*06:02
	S4_T	腺癌	IIA	100	KRAS G12D	
	S5_N					HLA-A*03:01 HLA-A*30:01 HLA-B*13:02 HLA-B*52:01 HLA-C*06:02 HLA-C*12:02
	S5_T	腺癌	IIA	95		
	S6_N					HLA-A*03:01 HLA-A*23:01 HLA-B*07:02 HLA-B*18:01 HLA-C*07:01 HLA-C*07:02
	S6_T	腺癌	IIA	95		

[0236] 使MHC I:肽复合体通过免疫沉淀分离,并且将洗脱的MHC I相关肽(MAP)用串联质量标签(TMT)同量异序标记试剂来标记,因为TMT标记最近显示出通过增加其电荷状态和疏水性来增强对MAP的检测(37)。然后使用通过RNA-seq生成的个性化癌症数据库通过液相色谱串联质谱(LC-MS/MS)对MAP进行测序和分析。然后,已识别的MAP经过一系列严格的分类

和验证,以确定推定的TSA和TAA。然后用合成肽对CRC组织中识别出的肿瘤抗原进行验证和定量,以确定它们在肿瘤细胞表面的过表达程度,并研究它们预测的免疫原性和瘤间分布,以获得它们临床潜力的灵敏度(sence) (图1)。

[0237] 在本研究中,使用了具有不同等位基因和特征的4种CRC来源的细胞系,如表1中所总结。HCT116和RK0来源于原代肿瘤,并且其特征为微卫星不稳定性(MSI),而Colo205和SW620分别来源于腹水和淋巴结的转移,并且均为微卫星稳定性的(MSS)。这些细胞系具有范围为 1.44×10^5 至 5.07×10^5 MHC I分子/细胞的大范围MHC I表面表达和多种HLA等位基因。在这4种细胞系中有若干种关键基因的突变,诸如BRAF、RAS、SMAD4、TP53和PI3CA。如通过Qifikit所测定,这些细胞系具有范围为 1.44×10^5 至 5.07×10^5 MHC I分子/细胞的不同MHC I表面表达,以及使用OptiType并结合文献中记载的这些细胞系的HLA等位基因来识别出的多种HLA等位基因,该OptiType是一种HLA基因分型工具,其使用RNA-Seq数据来预测样品的HLA等位基因(表1) (22)。

[0238] 所有原代肿瘤样品均来源于II期腺癌,其肿瘤分级和TNM(肿瘤-淋巴结-转移)分类仅略有不同(表2)。原代样品的肿瘤含量为95-100%,以及平均质量为0.6625克。除S1(盲肠)和S5(升结肠)外,肿瘤均来源于乙状结肠。除S2外,所有患者均为女性,并且患者年龄在43-85岁,平均年龄为62岁。与细胞系相似,组织样品也具有多种HLA等位基因。细胞系和组织样品特有或共有的HLA等位基因数目的可视化可见于图7。每种细胞系和组织平均分别有1.3和3.2个特有的等位基因。

[0239] 实施例3: CRC样品的转录组特征

[0240] 由于给定疾病阶段的CRC患者的结果基于肿瘤的分子特征而有很大差异(40,41),因此使用RNA测序数据来表征样品的分子异质性。在首先检查了临床上通常用于指导治疗决策和预后的关键生物标志物(诸如KRAS、NRAS或BRAF)的突变状态后(40,41)(表1),细胞系和原代样品的微卫星状态分别从文献中确定(42,43),并使用MSIsensor软件包(46)扩展这些样品的分子特征知识。虽然在有限的CRC肿瘤的子集(即15%的散发性CRC和90%的非息肉性病结直肠癌)中发现了MSI(47),但在这项研究中,50%的致瘤细胞系和33%的原代性活检显示了这种表型(表3)。尽管文献中的几个要素表明MSI和MSS肿瘤在免疫学上是不同的(5,11,48,49),但这项研究提供了在免疫肽组学水平上对MSI和MSS结直肠肿瘤的比较。

[0241] 表3: CRC原代组织的MSIsensor结果。

	样品	总位点 的数量	覆盖率足够 高的位点的 数量	覆盖率足 够高的位 点的数量 (%)	MSI 位 点的数 量(躯体 的)	MSI 位点的 数量(躯体 的)(%)	分 类 (MSI > 3.5%)
[0242]	S1	1011195	111243	11	226	0.20	MSS
	S2	1011195	69004	7	197	0.29	MSS
	S3	1011195	56779	6	104	0.18	MSS
	S4	1011195	45848	5	177	0.39	MSS
	S5	1011195	78340	8	8267	10.55	MSI
	S6	1011195	60715	6	3085	4.08	MSI

[0243] 对正常和肿瘤活检样品(图2A)或细胞系(图9A)之间前500个变化基因的主成分分析证实了它们不同的转录组学谱。因此,对CRC来源的细胞系和活检样品两者的途径和过程富集分析揭示了就与其致癌状态相关而言富集的转录组学谱。最显著上调和下调的GO术语分别与细胞增殖(图2B上图)和肌肉表型和收缩力(图2B下小图)相关联。虽然对与增殖和细胞周期相关的GO术语的富集是癌症的一般标志(50,51),但肌肉相关途径的下调是CRC所固有的,并且是低分化肿瘤区域和高收缩性NAT之间的功能二分法的结果。在本文使用的数据集的肿瘤样品中,似乎是MSI/MSS状态解释了肿瘤间转录组差异(图2A和9A)。MSI样品往往紧密聚集在一起,而MSS肿瘤似乎更分散并且因此在转录上更具异质性。在功能上,当单独分析时,MSS和MSI CRC样品富含非常不同的基因组。当与其相应的NAT相比时,MSI肿瘤的特征为各种免疫相关GO术语的显著上调(图9A),而MSS肿瘤与Wnt和PI3K-Akt信号传导相关基因的表达增加更相关(图9B)。虽然这两种信号通路与CRC的联系是已知的(52),但没有文献支持它们在CRC中的贡献可能不同于在MSS和MSI肿瘤中的贡献。

[0244] 接下来,经由使用以下的两种独立的方法来估计每种样品的免疫浸润程度:来自ESTIMATE软件包(37)的免疫浸润评分(图2C),以及基于单样品基因组富集分析(ssGSEA)(54)的已知肿瘤浸润性白细胞(TIL)标志物(53)的富集评分(图8B)。尽管所有NAT样品都表现出相似的免疫浸润水平,但MSI和MSS肿瘤的特征分别为增加和减少的免疫浸润评分(图2C和图8B)。这种差异表明MSI肿瘤可能比其MSS同源物更具免疫原性(48,55-58)。

[0245] 由于TSA可由多种癌症特异性事件/失调引起(11,12),且每种抗原来源的免疫肽组的贡献在恶性肿瘤中差异显著(11),因此RNA测序数据也被用于确定样品中可能富含的TSA类别。通过观察转录物的基因组来源,观察到与NAT相比,肿瘤中非编码多腺苷酸化RNA的比例和绝对丰度都显著增加(图2D)。虽然平均而言绝对丰度增加限于25%,但数据表明MSI肿瘤中非编码转录物的肿瘤特异性增益可高于MSS。尽管由于MSI样品数量少($n=2$),这种比较仍然有限,但人们可以预期在MSI样品中比在MSS中识别出更多数量的来自非编码转录物的aeTSA。类似地,尽管所有CRC样品都表现出相当的单核苷酸变体(SNV)负荷,并且因此预计具有相似数量的mTSA(图2E),但与MSS相比,MSI样品中的插入/缺失(得失位)负荷显著增加,这一观察结果也在细胞系中得到注意(图8C)。事实上,同时考虑细胞系和组织样品

导致MSS和MSI样品之间INDEL突变数量的统计学显著差异($p=0.00235$) (图8D)。因为MSI和INDEL积累都是由DNA错配修复(MMR)途径的缺陷引起的(59),人们可以假设在样品中识别出的INDEL衍生的TSA(最有可能是移码衍生的抗原)的数量与其MSI水平成比例。

[0246] 实施例4:免疫肽组学分析突显了CRC抗原的多样性

[0247] 为了阐明CRC衍生细胞系的MHC I免疫肽组,对每种细胞系免疫沉淀来自4份 2×10^8 个细胞的复制品的MAP,并且对于细胞系,每份复制品用单独的TMT6Plex通道(126,127,128,129)衍生,或对于原始NAT和组织样品每份复制品分别用TMT10Plex-126和-127N衍生。将每种细胞系的四份复制品,以及来自每个受试者的相应的NAT和肿瘤MAP的一半进行多重化,并通过LC-MS/MS进行分析。细胞系和组织样品的中值标记效率分别为72.4%或87.8%。细胞系中较低的标记效率归因于MAP产量不足。在细胞系和组织数据集中分别识别出了5281和27583个特有MAP,平均每种细胞系1433个特有MAP,以及每种组织5855个特有MAP(图3A,上小图和图3B)。虽然每条线之间的识别不同,但所识别的MAP的数量与每种细胞的MHC I分子丰度密切相关(图3A,下小图;皮尔逊相关系数=0.96)。

[0248] 当将细胞系和组织样品放在一起时,总计识别出30485个特有的MAP。在每种样品的MAP库内,32-68%的肽是样品特异性的,并且即使仅比较细胞系或原代样品,也很少有共享MAP(图10A-B)。这种大比例的特有MAP可归因于样品中HLA等位基因的多样性,这是影响能够呈递在细胞表面的肽的主要因素(图3C;图7)。平均而言,任何两种细胞系或任何两种组织样品共享的MAP数量分别为59或640个MAP。存在值得注意的异常值——组织样品S1和S6共有2079个MAP(其中1673个是这些样品独有的(图10C)),代表超过它们各自的MHC I免疫肽组的三分之一(图10D)。两种组织之间MAP库的第二相似之处是两种MSI组织(S5和S6)共享的1328个MAP,对应于其库的21%。细胞系中MAP识别的减少使得这些比较不那么引人注目。例如,HCT116和RK0共享最多的MAP,尽管这些肽仅占其MAP的不到10%,并且可能是其更大肽库的一个特征(图10C)。相比之下,COL0205和SW620共享152个MAP,接近其免疫肽组的五分之一。

[0249] 为了将这些比较联系起来,可以再次考虑样品的HLA等位基因。在S1和S6共享的2079个MAP中,有1542个MAP被预测为在两个样品中结合相同的等位基因,并且在超过93%的情况下该等位基因为HLA-B*07:02。类似地,由COL0205和SW620共享的152个MA中有136个由相同的等位基因结合,并且其中113个这些情况中的等位基因为HLA-A*02:01。因此,样品的MHC I免疫肽组主要受HLA库的影响。

[0250] 在基因水平上,识别出来源于超过8000个特有来源基因的肽,每个细胞系和组织样品分别有平均1014和3168个来源基因(图11A,上小图)。在每种样品中识别出的来源基因的数量与识别出的MAP的数量高度相关(图11A,下小图)。给定免疫肽组中大约6-14%的来源基因是样品特异性的(图11B),这可能归因于样品特异性的生物学特征或它可能反映免疫肽组的不完美采样(图11E)。不期望识别出在细胞表面呈递的每一个MAP,并且由于每种样品中的大多数来源基因仅归因于一个MAP,几乎可以肯定的是,额外的来源基因对MAP库有贡献并且只是未被检测到。当比较任何两种组织样品时,它们具有平均48%的共同来源基因,而比较任何两种细胞系得到平均24%的基因是共享的(图11C-E)。因此,不同的细胞系在来源基因水平上似乎不太同质。这可能反映了样品组成的差异,因为组织样品具有来源于NAT、基质、浸润细胞等的来源基因,而细胞系仅由单一细胞类型组成。另外,细胞系的

较低的MHC I呈递和由此导致的MAP识别下降意味着采样的来源基因更少,重叠的可能性更低。无论如何,与免疫肽组水平相比,所有样品在来源基因水平上更相似,样品特异性MAP来源于共享的来源基因。

[0251] 为了获得MHC I免疫肽组的基因组来源的概况并研究来源基因的共享性质,对细胞系中识别出的所有来源基因以及4种或更多种组织中识别出的那些来源基因进行了GO术语分析。在免疫肽组水平上可以检测到细胞系和组织之间的几个共同特征,包括参与RNA代谢、核糖核蛋白复合体生物发生、翻译和细胞对应激应答的基因的显著富集(图3D)。因此,尽管细胞系和组织样品之间的异质性,包括对肽库具有巨大影响的HLA等位基因的巨大多样性,以及细胞系中的低MAP识别,但在对MHC I免疫肽组有贡献的基因方面存在显著的相似性。

[0252] 为了研究组织样品中来自非编码转录物的MAP的比例,首先确定了每种肽最丰富的推定的来源转录物(Ensembl注释99)。对于癌症特异性数据库中的肽,MCS被映射到基因组上,并确定了该位置表达最多的转录物(参见“方法”部分中的“RNA-Seq数据中MAP编码序列的定量”)。由此确定,平均而言,来自组织样品的MAP的95.3%来自蛋白编码转录物(即UTR或CDR)(图3E,左小图)。如果将来源于未注释的RNA转录物的2.8%肽包括在内,大约4.2%的肽来自非编码区,因为这些肽可能来自基因间序列。大约三分之一的非编码MAP(包括来自未注释转录物的那些MAP)来源于无义介导的衰变转录物产物,而它们中的不到1%来自lncRNA、不间断衰变产物、保留的内含子或加工转录物(不包含开放阅读框的转录物)(图3E,右小图)。

[0253] 实施例5: CRC中肿瘤特异性和肿瘤相关抗原的识别

[0254] 在识别出超过30000个特有的MAP后,对肽编码序列进行了过滤,以选择在癌症中过表达至少10倍并且在合并的mTEC样品或匹配的NAT中分别在细胞系和原始样品中表达<2rpkm的那些序列。最近对急性髓系白血病(AML)的免疫肽组研究表明,RPKM<8.55的MCS生成MAP的概率低于5%(18)。然后对RNA数据中的MAP编码序列进行定量,并且仅保留在mTEC和其他正常组织(GTEX)中表达低于8.55rpkm的那些序列。对其余肽进行人工验证后,如果肽在肿瘤中过表达至少10倍,并在mTECs中表达<0.2kphm(在组织中为NAT),则被分类为异常表达的TSA(aeTSA)。如果MAP在癌症中也过表达至少10倍,但它们在mTEC和/或NAT中的表达大于0.2kphm,则被归类为TAA。

[0255] 虽然CRC衍生细胞系中的TSA识别相对不足,可能部分由于低MAP识别,但平均每种原代组织样品中识别出3个TSA(图4A)。总的来说,在CRC衍生的细胞系中识别了1个推定的TSA,在原代组织中识别了18个推定的TSA,并且每种样品的TSA产量与识别出的MAP数量相关(皮尔逊相关系数=0.76)(图12)。其中,大约三分之一来源于编码区,而大多数识别出的推定的TSA来源于非编码区(图4B)。在来自编码区的TSA中,两个来自非典型阅读框,衍生自外显子移码序列,以及另外两个是在S2和S3的MSS组织中识别的突变TSA(图4A和4B)。在非编码TSA中,所有的TSA都异常表达,并且大比例来源于内含子或基因间区,较少数量来源于5'UTR、3'UTR或lncRNA(图4B)。九个aeTSA的序列(5个内含子、3个基因间序列、1个lncRNA)与ERE序列重叠(表4)。由于ERE普遍存在的特性,来源于异常ERE表达的TSA可能为肿瘤所共享,并显示出免疫原性(61,62)。值得注意的是,没有一个推定的TSA是在多种样品之间共享的,即使是那些共享MAP的比例很高的样品。然而,在来源于COL11A1相同转录物的不同组织

中鉴定出两个特有的TSA (一个外显子移码和一个5'UTR) ,最近示出其在CRC发展和预后中起作用 (63) 。大多数其他TSA来源基因也显示出与CRC生物学相关 (表5) 。

[0256] 表4:本研究中所识别的TSA和TAA的特征

[0257]

肽序列(SEQ ID NO:)	样品	类别	来源	基因符号	转录物 ID
RMLLSHTGK (1)	RKO	aeTSA	基因间	无	无
LPHRALSGI (2)	S1	aeTSA	内含子	NKD1 (WNT通路的抑制剂)	ENST00000268459.6
GTNPAAVK (3)	S2	aeTSA	外显子	HKDC1	ENST00000354624.6
LRHKLVLNR (4)	S2	aeTSA	3' UTR	GRIN2B	ENST00000609686.3
RIGGVGVEK (5)	S2	aeTSA	外显子移码	LY6G6F-LY6G6D 通读 (readthrough)	ENST00000503322.1
SIETVNSL (6)	S2	aeTSA	内含子	IPP	ENST00000396478.4
TVNTQQYNTK (7)	S2	aeTSA	内含子	CYP39A1	ENST00000275016.3

[0258]

SVSHLHIF (8)	S3	aeTSA	基因间	无	无
TTLENLPQK (9)	S4	aeTSA	外显子移 码	COL11A1	ENST00000370096.9
AQKLQVRI (10)	S5	aeTSA	基因间	无	无
GQIELSIYR (11)	S5	aeTSA	内含子	PATJ	ENST00000371158.6
HGALSIRSI (12)	S5	aeTSA	内含子	DHP6	ENST00000561411.1
RLMKFLPV (13)	S5	aeTSA	基因间	无	无
SLYISEERK (14)	S5	aeTSA	lncRNA	TRPC6	ENST00000526713.1
VQTAVLNV (15)	S5	aeTSA	基因间	无	无
RNRQVATAL (16)	S6	aeTSA	外显子	SUCNR1	ENST00000362032.6
VEAPHLPSF (17)	S6	aeTSA	5' UTR	COL11A1	ENST00000370096.9
KIGEIVIVTK (18)	S2	mTSA	外显子	HSPD1	ENST00000388968.8
TRSTIHL (19)	S3	mTSA	外显子	PLK1	ENST00000300093.9
SILRGSLGK (20)	S4	ncTAA	基因间	无	无
RLVSNRVVR (21)	S5	ncTAA	内含子	IDO2	ENST00000518237.6
SLKNLEPIK (22)	S5	ncTAA	基因间	无	无
RSAGPRAPL (23)	S6	ncTAA	基因间 (转录物的 反义)	无	无

[0259]

VLYRSVLLLK (24)	S6	ncTAA	内含子	FANCA	ENST00000389301.8
EYNSDLHQF (25)	S1	TAA	外显子	BUB1	ENST00000302759.1 1
TYKYVDINTF (26)	S1	TAA	外显子	MMP12	ENST00000571244.3
VIDEPVRL (27)	S1	TAA	外显子	MKI67	ENST00000368653.7
RYLEKFYGL (28)	S1, S6	TAA	外显子	MMP12	ENST00000571244.3
KRLPSPSLF (29)	S2	TAA	外显子	MGAM2	ENST00000550469.6
KSINEFWNK (30)	S2	TAA	外显子	SPC25	ENST00000282074.7
RVFHGNETTK (31)	S2	TAA	外显子	CENPE	ENST00000265148.8
RIQLPVVSK (32)	S2, S4	TAA	外显子	MKI67	ENST00000368653.7
SIFRTPISK (33)	S2, S4	TAA	外显子	ASPM	ENST00000367409.9
KAAGAGAAK (34)	S2, S5, S6	TAA	外显子	HI-5	ENST00000331442.4
AEYHSQSRA (35)	S3	TAA	外显子	ASPM	ENST00000367409.9
IHNSDQILL (36)	S3	TAA	外显子	MACC1	ENST00000400331.1 0
KMHQSRISY (37)	S3	TAA	外显子	ASPM	ENST00000367409.9
QMAGLRDTY (38)	S3	TAA	外显子	SPC25	ENST00000282074.7
RIKNWGSGMT F (39)	S3	TAA	外显子	NOS2	ENST00000313735.1 1
AQYDQASTKY (40)	S4	TAA	外显子	CENPF	ENST00000366955.8

[0260]

FVDNQYWRY (41)	S4	TAA	外显子	MMP12	ENST00000571244.3
QLDSAVKNLY (42)	S4	TAA	外显子	ZNF215	ENST00000414517.6
EQLKVTTI (43)	S5	TAA	外显子	MCM10	ENST00000378714.7
FRSWAVQTF (44)	S5	TAA	外显子	NOS2	ENST00000313735.1 1
KLYQSIIGY (45)	S5	TAA	外显子	DIAPH3	ENST00000400324.9
SANVSKVSF (46)	S5	TAA	外显子	ASPM	ENST00000367409.9
SQHDILQMI (47)	S5	TAA	外显子	MKI67	ENST00000368653.7
VAKTPRNTY (48)	S5	TAA	外显子	MKI67	ENST00000368654.8
RPRLSAILF (49)	S6	TAA	外显子	DIAPH3	ENST00000400324.9
RVAKTNSLR (50)	S5	TAA	外显子	CDCA8	ENST00000327331.2

[0261]

肽序列	基因 ID	ERE 重 叠	位置	链	预测的 HLA 等位基因
RMLLSHTGK (1)	无	无	chr8:127209405- 127209431	(-)	HLA-A*03:01
LPHRALSGI (2)	ENSG00000140 807	反义 miRc 重 叠	chr16:50564443- 50564469	(+)	HLA-B*07:02
GTNPAAVK (3)	ENSG00000156 510	无	chr10:69227303- 69227329	(+)	HLA-A*03:02
LRHKLVLNR (4)	ENSG00000273 079	无	chr12:13549419- 13549445	(-)	HLA-B*27:05

[0262]

RIGGVGVEK (5)	ENSG00000250 641	无	chr6:31717798-3 1717824	(+)	HLA-A*03:02
SIHETVNSL (6)	ENSG00000197 429	L1MEf 重叠	chr1:45713038-4 5713064	(-)	HLA-A*02:01
TVNTQQYNTK (7)	ENSG00000146 233	L1MD1 重叠	chr6:46644986-4 6645015	(-)	HLA-A*03:02
SVSHLHIFF (8)	无	反义 L1MB3 重叠	chr17:65462740- 65462766	(+)	HLA-A*32:01
TTLENLPQK (9)	ENSG00000060 718	无	chr1:103078698- 103078724	(-)	HLA-A*11:01
AQKLQVRI (10)	无	无	chr3:29988327-2 9988304	(-)	HLA-B*52:01
GQIELSIYR (11)	ENSG00000132 849	无	chr1:61808366-6 1808392	(+)	HLA-A*03:01
HGALSIRSI (12)	ENSG00000134 146	L2a 重叠	chr15:35525008- 35525034	(-)	HLA-B*52:01
RLMKFLPV (13)	无	无	chr2:170700443- 170700466	(-)	HLA-B*13:02
SLYISEERK (14)	ENSG00000137 672	LIME1 重叠	chr11:101656906 -101656932	(-)	HLA-A*03:01
VQTAVLNV (15)	无	无	chr4:93906720-9 3906697	(-)	HLA-B*52:01
RNRQVATAL (16)	ENSG00000198 829	无	chr3:151881192- 151881218	(+)	HLA-B*07:02
VEAPHLPSF (17)	ENSG00000060 718	无	chr1:103108448- 103108474	(-)	HLA-B*18:01
KIGEVIVTK (18)	ENSG00000144 381	无	chr2:197489161- 197489188	(-)	HLA-A*03:02
TRSTIHL (19)	ENSG00000166 851	无	chr16:23689617- 23689643	(+)	HLA-C*06:02
SILRGSLGK	无	CR1_Ma	chr7:140260497-	(-)	HLA-A*11:01

[0263]

(20)		m 重叠	140260523		
RLVSNRVVR (21)	ENSG00000131 203	无	chr8:39917585-3 9917611	(+)	HLA-A*30:01
SLKNLEPIK (22)	无	MSTA 重叠	chr15:97429857- 97429883	(+)	HLA-A*30:01
RSAGPRAPL (23)	无	无	chr1:145096230- 145096256	(-)	HLA-B*07:02
VLYRSVLLK (24)	ENSG00000187 741	无	chr16:89798385- 89798414	(-)	HLA-A*03:01
EYNSDLHQF (25)	ENSG00000169 679	无	chr2:110672828- 110672854	(-)	HLA-A*24:02
TYKYVDINTF (26)	ENSG00000262 406	无	chr11:102867951 -102867980	(-)	HLA-A*24:02
VIDEPVRL (27)	ENSG00000148 773	无	chr10:128122922 -128122945	(-)	HLA-C*04:01
RYLEKFYGL (28)	ENSG00000262 406	无	chr11:102873086 -102873112	(-)	HLA-A*24:02, HLA-A*23:01
KRLPSPSLF (29)	ENSG00000257 743	无	chr7:142131535- 142131561	(+)	HLA-C*07:01
KSINEFWNK (30)	ENSG00000152 253	无	chr2:168889463- 168889489	(-)	HLA-A*03:02
RVFHGNETTK (31)	ENSG00000138 778	无	chr4:103196221- 103196250	(-)	HLA-A*03:02
RIQLPVVSK (32)	ENSG00000148 773	无	chr10:128123124 -128123150	(-)	HLA-A*03:02, HLA-A*11:01
SIFRTPISK (33)	ENSG00000066 279	无	chr1:197139840- 197139866	(-)	HLA-A*03:02, HLA-A*11:01
KAAGAGAAK (34)	ENSG00000184 357	无	chr6:27867425-2 7867451	(-)	HLA-A*03:02, HLA-A*30:01, HLA-A*03:01
AEYHSQSRA (35)	ENSG00000066 279	无	chr1:197100800- 197100826	(-)	HLA-B*50:01

[0264]

IHNSDQILL (36)	ENSG00000183 742	无	chr7:20159866-2 0159892	(-)	HLA-B*38:01
KMHQSRISY (37)	ENSG00000066 279	无	chr1:197102843- 197102869	(-)	HLA-A*32:01
QMAGLRDTY (38)	ENSG00000152 253	无	chr2:168889412- 168889438	(-)	HLA-A*01:01
RIKNWGSGMT F (39)	ENSG00000007 171	无	chr17:27788839- 27788871	(-)	HLA-A*32:01
AQYDQASTKY (40)	ENSG00000117 724	无	chr1:214622255- 214629048	(+)	HLA-B*15:01
FVDNQYWRY (41)	ENSG00000262 406	无	chr11:102864249 -102865798	(-)	HLA-A*01:01
QLDSAVKNLY (42)	ENSG00000149 054	无	chr11:6943137-6 943166	(+)	HLA-A*01:01
EQLKVTTI (43)	ENSG00000065 328	无	chr10:13172417- 13172440	(+)	HLA-B*52:01
FRSWAVQTF (44)	ENSG00000007 171	无	chr17:27768980- 27769006	(-)	HLA-C*06:02
KLYQSIIGY (45)	ENSG00000139 734	无	chr13:59833182- 59833208	(-)	HLA-A*03:01
SANVSKVSF (46)	ENSG00000066 279	无	chr1:197143439- 197143465	(-)	HLA-C*12:02
SQHDILQMI (47)	ENSG00000148 773	无	chr10:128112176 -128112202	(-)	HLA-B*13:02
VAKTPRNTY (48)	ENSG00000148 773	无	chr10:128109293 -128109319	(-)	HLA-C*12:02
RPRLSAILF (49)	ENSG00000139 734	无	chr13:59879421- 59879447	(-)	HLA-B*07:02
RVAKTNSLR (50)	ENSG00000134 690	无	chr1:37692715-3 7692741	(+)	HLA-A*03:01

[0265] 表5.TSA来源基因在CRC中的生物学相关性。

来源基因	参考文献	CRC 中的生物相关性
NKD1 - 裸角质层同源物 1	PMID : 25446263; PMID : 19956716	小鼠肠道肿瘤标志物 Wnt 通路负反馈调控因子; NKD1 中的突变改变 Wnt 信号传导
ANTXR1 - 炭疽毒素受体 1	PMID : 31977138; PMID : 22340594	在胃癌中上调并与不良预后相关, 激活 mTOR 通路; ANTRXR1 阻断阻止 CRC 血管生成
HKDC1 - 含己糖激酶结构域的蛋白 1	PMID : 30005951	HKDC1 有助于增加 CRC 细胞的代谢、增殖和转移
[0266]	GRIN2B - 谷氨酸离子型受体 NMDA 型亚单位 2B	PMID : 27243824 被识别为非驱动枢纽(hub)基因, 参与进展为 II 期 CRC
LY6G6F-LY6G6D 通读 - 淋巴细胞抗原 6 家族成员 G6F 和 G6D	PMID : 26894861	在 CRC 中过表达的 LY6G6D/F, 是潜在的细胞表面标志物
IPP (KLHL27) - 脑池内 A 颗粒促进多肽	人蛋白图谱 (PMID : 28818916)	结直肠癌中的预后良好标志物; 不宜用于肾癌和肝癌
CYP39A1 - 细胞色素 P450, 家族 39, 亚家族 A, 多肽 1	PMID : 27341022	在预后不良的 CRC 中表达增加
COL11A1 - XI α 1 型胶原蛋白	PMID : 33597969	在 CRC 中上调(mRNA), 不良预后的标志物, 在 CRC 发展中的作用

[0267]	PATJ - PALS1 相关紧密连接蛋白		没有已知的关联
	DPH6 - 苯二胺生物合成 6		没有已知的关联
	TRPC6 - 瞬时受体电位阳离子通道亚家族 C 成员 6	PMID : 26422106	CRC 中 TRPC6 的 mRNA 表达低于正常组织, 可能与肿瘤发生有关
	SUCNR1 - 琥珀酸受体 1	PMID : 32365557	SUCNR1 激活诱导 Wnt 配体表达并激活 CRC 衍生细胞系中的 WNT 信号传导和 EMT
	POF1B - 卵巢早衰蛋白 1B	PMID : 29484395	可能与 CRC 细胞迁移中的 ATP5J 功能有关; 调控肠细胞系的黏附
		PMID : 25084053	
	HSPD1 - 热激蛋白家族 D (Hsp60)成员 1	PMID : 28261350	在 CRC 中差异地表达, 用于诊断的潜在生物标志物; 胞外 HSPD1
		PMID : 29246022	被识别为 CRC 的推定诊断和预后生物标志物
[0268]	PLK1 - 丝氨酸/苏氨酸蛋白激酶 PLK1 / polo 样激酶 1	PMID : 22648245	在 CRC 中过表达, 与转移和侵袭相关
	MATR3 - Matrin 3	PMID: 28580901	MATR3 示出响应 DNA 损伤, 通过与 PINCR lncRNA 和 p53 相互作用, 参与 CRC 衍生细胞的促存活活性

[0268] 虽然主要目的是在CRC中识别出推定的TSA,但在CRC组织样品中也识别出平均6.33个TAA,尽管在CRC衍生的细胞系中没有识别出TAA(图4C)。与主要的非编码推定的TSA相比之下,大多数经识别出的TAA来源于典型编码外显子序列,只有少量来源于内含子、基因间序列或lncRNA(图4D)。两个非典型TAA被ERE序列重叠(表4)。值得注意的是,在不止一种样品中识别出了4种不同的TAA,而在3种组织中识别出了一种TAA。这些重复的TAA都来源于典型的外显子,其源转录物来源于ASPM、MKI67、DIAPH3、MMP12、NOS2和SPC25,所有这些转

录物都与癌症有关(表6)。

[0269] 表6.TAA来源基因在CRC中的生物学相关性。

	基因符号	参考文献	CRC 中的生物相关性
	MMP12 - 基质金属钛酶 12	PMID: 27431388	与对照相比, 在 CRC 中过表达, CRC 中的阴性预后标志物
	MKI67 - 增殖 Ki-67 的标志物	PMID: 26281861; PMID: 27855388; PMID: 30727976	CRC 中的预后良好标志物, IHC 染色(2016); III 期和 IV 期 CRC 的预后良好标志物, IHC 染色(2016); 基于数据库 meta 分析的 CRC 预后不良标志物(2019)
[0270]	BUB1 - 有丝分裂纺锤体检查点激酶	PMID: 23747338; PMID: 11782350	BUB1 中的突变与早发性 CRC 关联; 失活可以驱动 CRC 的转移和进展
	DIAPH3 - 透明相关形成蛋白 3	人 蛋 白 图 谱 (PMID: 28818916)	DIAPH3 具有预后性, 在结直肠癌中高表达是有利的
	MGAM2 - 麦芽糖酶-葡萄糖淀粉酶 2	PMID: 30996822	在 GI 癌中表达 (TCGA 数据)
	SPC25 (动粒蛋白)	PMID: 32351050; 人 蛋 白 图 谱	在 CRC(以及其他癌症)中高度表达; 肝癌、子宫内膜癌和肺癌的预后不良

	(PMID: 28818916)	标志物
CENPE - 着丝粒相关蛋白 E		没有已知的关联
ASPM - 异常纺锤体小头畸形相关	PMID: 31966766; 人 蛋 白 图 谱 (PMID: 28818916)	在 CRC 中过表达; 建议作为预后不良标志物(涉及有丝分裂、细胞周期、肿瘤发生); 已知在肝癌、肺癌、子宫内膜癌、胰腺癌中是预后不良标志物
HI-5 - H1.5 接头组蛋白, 簇成员	PMID: 16959974	在 CRC 中频繁突变
MACC1 - 结肠癌转移相关 1	PMID: 27424982; PMID: 25003996	促进结直肠癌的生长和转移; 通过 β -联蛋白信号和 EMT 转换与致癌作用相关
[0271] NOS2 - 一氧化氮合酶 2	人 蛋 白 图 谱 (PMID: 28818916)	癌症增强(结直肠癌); RNA 数据
CENPF - 着丝粒蛋白 F	PMID: 30550624	与 CRC 恶性程度相关的磷酸化改变; 其他癌症(肝癌、肾癌等; 人蛋白图谱)的预后不良标志物。
ZNF215 - 锌指蛋白 215	人 蛋 白 图 谱 (PMID: 28818916)	免疫细胞亚群中的细胞质表达, 在胃肠道和淋巴组织中最丰富(蛋白数据)
MCM10 - 微小染色体维持 10 复制起始因子	PMID: 32597491	与正常组织相比, 结肠和直肠腺癌样品中 mRNA 表达降低
CDCA8 - 细胞分裂周期相关 8	PMID: 25260804	在 CRC 中过表达, 与癌症进展相关
IDO2 - 吲哚胺 2,3-双加氧酶 2	PMID: 18418598;	在 CRC 中表达上调

UNG - 尿嘧啶 DNA 糖基化酶 PMID: 17029639 CRC 患者中常见的种系突变

[0272] FANCA - 范可尼贫血 A 组蛋白 PMID: 27165003; 范可尼贫血易诱发某些癌症; FA 途径中的基因参与 CRC 病程(参与 HR 修复) PMID: 21286667

[0273] 最初预计在MSI组织中识别出高于平均数量的TSA和TAA。这是S5中的情况,然而对于其他MSI组织而言并非如此(图4A和4C)。这可能是由于S6的不稳定性“程度”较低,如MSIsensor结果所示(表3)。此外,具有最高数量的所识别TSA的样品是S2,一种MSS组织。因此,每种样品的TSA和TAA的产量似乎与MSI状态无关,并且可能是由于本研究范围之外的肿瘤的其他特有生物学特征。

[0274] 为了确定是否有任何推定的TSA或TAA先前已被识别,验证了肽序列是否在以下中有所报道:免疫表位数据库(IEDB)、HLA配体图谱(64)和试图从Löffler et al. 2018 (8) 和 Newey et al. 2019 (15) 中识别CRC中的肿瘤抗原的2个先前的出版物。值得注意的是,在这些资源中先前没有报道过推定的aeTSA、mTSA或非典型的TAA。在识别出的26种推定的典型TAA中,其中的24种是在IEDB、Löffler et al 2018 (PXD009602)、Newey et al 2019 (PXD014017) 或三者的某种组合中报道的(图4E)。其中的八种在HLA配体图谱中也有报道,其中的一种在健康结肠组织中有明确记录。有趣的是,先前在这些较早期出版物中发现的TAA都没有被报道为肿瘤抗原,并且与之相比,在Löffler et al. 中报道的12种目的肿瘤抗原中的6种也在本研究的免疫肽组中得到识别,尽管它们没有通过识别流程中建立的被认为是TSA或TAA的阈值,最常见的原因是NAT中的高表达(表7A)。因此,在本研究中已经识别出主要来源于非编码区的新型结直肠癌TSA,以及主要编码TAA的选择,其中一些先前已作为MAP报道,但不在其生物学相关性的背景下作为TAA。表7B和7C分别描述了本研究中所识别的TSA和TAA。

[0275] 表7A:排除Löffler et al. 2018肿瘤抗原的理由。

[0276]

肽 (SEQ ID NO)	Loffler 等人 2018 年分类	Loffler 等人 2018 年 HLA 限制	样品	排除原因
RLASRPLLL (101)	疫苗候选者	B*07	S5	癌症和 NAT 之间的 FC < 10
YRNSYEIEY (102)	疫苗候选者	C*07	S1, S2, S6	NAT 中 MCS 表达 > 2 KPHM
APTPARPVL (103)	需要进一步验证的疫苗候选者	B*07	S1	mTEC 中 MCS 表达为 8.55 KPHM
RLAEPSQMLK (104)	需要进一步验证的疫苗候选者	A*03	S6	NAT 中 MCS 表达 > 2 KPHM
SPKATGVFTTL (105)	需要进一步验证的疫苗候选者	B*07	S1, S6	NAT 中 MCS 表达 > 2 KPHM
SVLTQPPSV (106)	证据不明确的疫苗候选者	A*02	S2	NAT 中 MCS 表达 > 2 KPHM

[0277] 表7B: 本研究中所识别的TSA

[0278]

序列(氨基酸和核苷酸)	SEQ ID NO:	预测的 HLA 等位	状态 ²
-------------	------------	------------	-----------------

[0279]

		基因 ¹	
RMLLSHTGK AGGATGTTACTGAGCCACACAGGA AAA	1 51	HLA-A*03:01	新型
LPHRALSGI CTTCCTCACAGGGCTTTGTCAGGC ATT	2 52	HLA-B*07:02	新型
GTNPTAAVK GGCACCAACCCACGGCTGCAGT GAAG	3 53	HLA-A*03:02	新型
LRHKLVLNR TTAAGGCATAAGTTAGTTCTGAAC AGA	4 54	HLA-B*27:05	新型
RIGGVGVEK CGGATAGGGGGAGTAGGAGTAGA GAAG	5 55	HLA-A*03:02	新型
SIHETVNSL TCTATAATAGAAACTGTAAATTCTC TT	6 56	HLA-A*02:01	新型
TVNTQQYNTK ACTGTCAACACTCAACAATATAAT ACAAAA	7 57	HLA-A*03:02	新型
SVSHLHIF TCTGTAGCCATTTGCATATCTTCT TT	8 58	HLA-A*32:01	新型
TTLENLPQK ACCACACTGGAAAACCTGCCCCA GAAG	9 59	HLA-A*11:01	新型
AQKLQVRI GCACAAAAATTGCAAGTGAGGATT	10 60	HLA-B*52:01	新型
GQIELSIYR	11	HLA-A*03:01	新型

[0280]

GGTCAAATAGAGTTATCAATATATA GG	61		
HGALSIRSI CATGGTGCCCTGTCCATAAGAAGT ATT	12 62	HLA-B*52:01	新型
RLMKFLPV AGACTGATGAAATTCCTTCCAGTT	13 63	HLA-B*13:02	新型
SLYISEERK TCTCTGTATATCTCAGAAGAAAGG AAA	14 64	HLA-A*03:01	新型
VQTAVLNV GTTCAAACCTGCTGTTTTAAATGTG	15 65	HLA-B*52:01	新型
KIGEVIVTK AAAATTGGAGAGGTCATTGTGACC AAA	16 66	HLA-A*03:02	新型
TRSTIILHL ACCCGCAGCACCATCATCCTGCAC CTC	17 67	HLA-C*06:02	新型
RNRQVATAL AGGAATAGGCAGGTTGCTACTGCT CTG	24 74	HLA-B*07:02	先前报道于 US2022/000 2374
VEAPHLPSF GTGGAAGCCCCTCATCTGCCTTCA TTC	25 75	HLA-B*18:01	先前报道于 WO2020/13 1586

[0281] 表7C:本研究所识别的TAA

[0282]

序列	SEQ ID NO:	预测的 HLA 等位基 因	状态
SILRGS LGK TCGATTTTACGTGGATCTCTAGGGA AA	18 68	HLA-A*11:01	新型

[0283]

RLVSNRVVR AGATTGGTTTCAAATAGAGTAGTCA GG	19 69	HLA-A*30:01	新型
SLKNLEPIK AGCCTGAAAAACCTTGAGCCAATT AAA	20 70	HLA-A*30:01	新型
RSAGPRAPL CGCTCTGCGGGTCCG CGGGCCCCTCTC	21 71	HLA-B*07:02	新型
VLYRSVLLK GTGTTGTACCGTTCG GTACTTTTACTGAAA	22 72	HLA-A*03:01	新型
KAAGAGAAK AAGGCTGCCGGCGCC GGCGCTGCTAAG	23 73	HLA-A*03:02, HLA-A*30:01, HLA-A*03:01	新型
EYNSDLHQF GAGTACAACAGTGACCTCCATCAAT TT	26 76	HLA-A*24:02	先前报道于 IEDB 和 PXD014017 2
TYKYVDINTF ACCTACAAATATGTTG ACATCAACACATT	27 77	HLA-A*24:02	先前报道于 US2022/01 75898A1
VIDEPVRL GTTATTGATGAGCCTGTACGGCTA	28 78	HLA-C*04:01	先前报道于 IEDB、 PXD009602 、 PXD014017 、HLA 配体 图谱 ²
RYLEKFYGL AGATACTTAGAAAAATTTTATGGCC	29 79	HLA-A*24:02, HLA-A*23:01	先前报道于 PXD009602

[0284]

TT			²
KRLPSPSLF AAAAGGTTGCCATCACCATCTCTGT TT	30 80	HLA-C*07:01	先前报道于 PXD014017 ²
KSINEFWNK AAAAGCATAAATGAATTTTGAATA AA	31 81	HLA-A*03:02	先前报道于 IEDB、 PXD014017 和 HLA 配 体图谱 ²
RVFHGNETTK CGTGTCTTTCATGGT AATGAAACTACCAAA	32 82	HLA-A*03:02	先前报道于 IEDB 和 PXD014017 ²
RIQLPVVSK CGTATCCAGCTTCCTGTTGTGTCAA AA	33 83	HLA-A*03:02, HLA-A*11:01	先前报道于 IEDB 和 PXD014017 ²
SIFRTPISK TCAATATTCAGAACTCCAATTCTA AA	34 84	HLA-A*03:02, HLA-A*11:01	先前报道于 IEDB 和 PXD014017 ²
AEYHSQSRA GCAGAGTATCATTCTCAAAGTAGGG CT	35 85	HLA-B*50:01	先前报道于 IEDB ²
IHNSDQILL ATACATAACTCTGACCAGATCCTAC TA	36 86	HLA-B*38:01	先前报道于 IEDB ²
KMHQSRISY AAAATGCATCAGTCAAGAATAAGTT AC	37 87	HLA-A*32:01	先前报道于 IEDB 和 PXD014017 ²
QMAGLRDTY	38	HLA-A*01:01	先前报道于

[0285]

CAGATGGCGGGACTAAGAGATACCT AC	88		IEDB, PXD009602 和 PXD014017 2
RIKNWGSGMTF AGGATCAAAAAGTGGGGCAGC GGGATGACTTTC	39 89	HLA-A*32:01	先前报道于 PXD014017 2
AQYDQASTKY GCACAATACGACCAG GCGTCAACCAAGTA	40 90	HLA-B*15:01	先前报道于 IEDB 和 HLA 配体 图谱 ²
FVDNQYWRY TTTGTAGATAACCAGTATTGGAGGT AT	41 91	HLA-A*01:01	先前报道于 PXD009602 2
QLDSAVKNLY CAACTGGACTCTGCTGT AAAGAACCTGTAC	42 92	HLA-A*01:01	先前报道于 IEDB 和 HLA 配体 图谱 ²
EQLKVTTI GAGCAGCTAAAAGTAACAACAATT	43 93	HLA-B*52:01	先前报道于 IEDB ²
FRSWAVQTF TTCCGCAGCTGGGCCGTGCAAACC TTC	44 94	HLA-C*06:02	先前报道于 IEDB 和 PXD009602 2
KLYQSIIGY AAGTTATACCAGAGTATAATAGGAT AC	45 95	HLA-A*03:01	先前报道于 IEDB, PXD009602 和 PXD014017 2
SANVSKVSF	47	HLA-C*12:02	先前报道于

[0286]

AGTGCCAACGTTTCAAAGTTTCTT TT	97		IEDB、 PXD014017 和 HLA 配 体图谱 ²
SQHDILQMI AGTCAACATGATATTTTACAGATGAT A	48 98	HLA-B*13:02	先前报道于 IEDB、 PXD009602 和 HLA 配 体图谱 ²
VAKTPRNTY GTAGCAAAAACGCCCAGGAACACC TAC	49 99	HLA-C*12:02	先前报道于 IEDB ²
RPRLSAILF CGGCCACGGCTCAGTGCTATTCTCT TT	50 100	HLA-B*07:02	先前报道于 IEDB、 PXD009602 和 HLA 配 体图谱 ²
RVAKTNSLR CGGGTGGCCAAGACCAACTCCTTA CGG	46 96	HLA-A*03:01	先前报道于 IEDB、 PXD009602 和 PXD014017 ²

[0287] ¹在多于一种样品中所识别出的肽可能具有多个预测的等位基因

[0288] ²未被识别为肿瘤抗原

[0289] IEDB=免疫表位数据库;PXD=经由ProteomeXchange公布的数据集

[0290] 实施例6:推定的肿瘤特异性和肿瘤相关抗原的验证

[0291] 在确定推定的TSA和TAA后,对所有TSA和11个TAA的子集进行验证,这些TAA是基于癌症与匹配的NAT中有利的初始TMT强度比和前体离子分数选择的,然后使用合成肽进行验证。首先,研究了与匹配的NAT相比,各自肿瘤样品中的来源转录物的以TPM表示的表达,以及CRC/NAT样品中该转录物的平均值(图5A)。这种分析自然不包括来源于基因间区的肽。在被识别的样品中,推定的TSA和TAA的来源转录物的平均log2FC分别为3.6和3.2。在少数情况下(S2,S6),aeTSA的源转录物在NAT中比在肿瘤中更丰富,然而这仅反映了整个转录物的总体丰度,并且事实上肽编码序列在癌症中更丰富。对于aeTSA也是如此,其中肽编码区在NAT中完全缺失或低表达,但在癌症组织中更高表达。

[0292] 为了评价推定的肿瘤抗原的特异性,确定了由基因型-组织表达项目(GTE_x)提供的健康组织大数据集中肽编码序列的平均表达(图5B)。TSA序列在任何健康组织中的表达

都不高于8.55rphm,但RIGGVGVVEK(在S2识别的aeTSA)除外,它在睾丸中的表达高于阈值(图5B)。这表明这种TSA也可以被归类为癌症睾丸抗原(CTA),一种在男性生殖细胞中表达但也可在癌症中异常表达的aeTSA。由于睾丸中缺乏MHC I,这些抗原也是癌症免疫疗法的有希望的候选者(65)。这个推定的TSA是LY6G6F-LY6G6D外显子移码。虽然这些基因先前没有被报道为CTA,但同一基因家族的另一个成员LY6K被报道为肺癌和食管癌中的CTA(66)。TAA的表达在健康组织中低于阈值,尽管它在食管和横结肠中往往较高。这些肽中的七种在睾丸中的表达也高于阈值。

[0293] 实施例7:TSA和TAA的癌症特异性和免疫原性

[0294] 推定的TSA和TAA通过MS与其相应的合成肽进行验证。这些TAA是基于癌症与匹配的NAT中有利的初始TMT强度比和前体离子分数选择的。这些候选者具有的MS/MS与合成肽具有的MS/MS拥有良好的相关性,其皮尔逊相关性评分 ≥ 0.6 。合成肽分别以10、100和1000fmol的浓度用TMT10plex-129N、130N和131标记,并掺入用TMT 126(NAT)和127n(CRC)标记的组织样品的剩余纯化MAP中。然后使用SPS-MS3对这些样品中的目的肽进行定量。尽管SPS-MS3的灵敏度有所下降,但对七种TSA和七种TAA进行定量仍是可行的。选择质量好的PSM进行定量,与NAT相比,所有在相应的CRC中都更丰富(表8)。确定TMT127N肽相比于TMT126肽的强度的比率表明,与NAT相比,在CRC中TSA的强度中值倍数变化是NAT的16.96倍,而TAA的倍数变化为6.93。另外,具有序列RYLEKIFYGL的TSA在S1肿瘤中也过表达,尽管仅超过S6的转录组阈值。因此,有可能证明本研究中使用的TSA识别方法成功识别出了在癌细胞表面比NAT表面更丰富的TSA和TAA序列。表9A-9B列出了经用合成肽验证的TSA和TAA候选者。

[0295] 表8: CRC中已验证肿瘤抗原的相对定量比率。

序列	抗原的 性质	样品	内源性样 品比率	平均强度	SPS-MS3 比 率 (127N/126)	合成 校准 曲线 R ²
RMLLSHTGK	aeTSA	RKO	N.D.	N.D.	N.D.	N.D.
LPHRALSGI	aeTSA	S1	-0.364	N.D.	N.D.	N.D.
GTNPATAVK	aeTSA	S2	2.095	7238.425242	12.174	1.000
LRHKLVLNR	aeTSA	S2	0.307	N.D.	N.D.	N.D.
RIGGVGVEK	aeTSA	S2	1.965	29256.45	6.740	1.000
SHIETVNSL	aeTSA	S2	0.288	N.D.	N.D.	N.D.
TVNTQQYNTK	aeTSA	S2	-0.021	N.D.	N.D.	N.D.
[0296] SVSHLHIFF	aeTSA	S3	-1.100	N.D.	N.D.	N.D.
TTLENLPQK	aeTSA	S4	0.134	3140.8875	3.783	0.999
AQKLQVRI	aeTSA	S5	0.793	N.D.	N.D.	N.D.
GQIELSIYR	aeTSA	S5	0.328	N.D.	N.D.	N.D.
HGALSIRSI	aeTSA	S5	0.777	N.D.	N.D.	N.D.
RLMKFLPV	aeTSA	S5	0.171	N.D.	N.D.	N.D.
SLYISEERK	aeTSA	S5	0.046	N.D.	N.D.	N.D.
VQTAVLNV	aeTSA	S5	1.089	N.D.	N.D.	N.D.
VEAPHLPSF	aeTSA	S6	1.059	43782.84192	41.318	1.000
RNRQVATAL	aeTSA	S6	1.090	12174.6625	5.722	1.000
RNRQVATAL	未指定	S1	0.890	15514.2375	3.507	1.000
KIGEVIVTK	mTSA	S2	2.506	70659.6	13.637	1.000

	TRSTIHLHL	mTSA	S3	1.381	34365.32187	48.807	0.997
	VLYRSVLLLK	非典型 TAA	S6	0.997	N.D.	N.D.	N.D.
	TYKYVDINTF	典型 TAA	S1	1.969	29834.36875	8.226	0.998
	RYLEKFYGL	典型 TAA	S1	2.840	27614.24286	7.661	0.997
	RYLEKFYGL	典型 TAA	S6	2.970	106928.2875	16.090	0.999
	KSINEFWNK	典型 TAA	S2	2.212	56110.11667	5.238	0.999
[0297]	RIQLPVVSK	典型 TAA	S4	1.083	7612.378571	2.073	0.999
	QMAGLRDTY	典型 TAA	S3	1.140	36090.60294	2.884	0.999
	AQYDQASTKY	典型 TAA	S4	1.452	N.D.	N.D.	N.D.
	FVDNQYWRY	典型 TAA	S4	0.721	5853.986533	10.954	1.000
	SANVSKVSF	典型 TAA	S5	1.114	12780.925	2.321	0.999

[0298] 表9A和9B描述了经用合成肽验证的TSA和TAA候选者。

[0299] 表9A:经合成肽验证的TSA的列表

[0300]

序列	SEQ ID NO:
RMLLSHTGK	1
LPHRALSGI	2
GTNPTAAVK	3

[0301]

LRHKLVLNR	4
RIGGVGVEK	5
SIETVNSL	6
TVNTQQYNTK	7
SVSHLHIF	8
TTLENLPQK	9
AQKLQVRI	10
GQIELSIYR	11
HGALSIRSI	12
RLMKFLPV	13
SLYISEERK	14
VQTAVLNV	15
KIGEVIVTK	16
TRSTIILHL	17
RNRQVATAL	24
VEAPHLPSF	25

[0302] 表9B:经合成肽验证的TAA的列表

[0303]

序列	SEQ ID NO:
VLYRSVLLLK	22
TYKYVDINTF	27
RYLEKFYGL	29
KSINEFWNK	31
RIQLPVVSK	33
QMAGLRDTY	38

[0304]	AQYDQASTKY	40
	FVDNQYWRY	41
	SANVSKVSF	47

[0305] 为了检查这些TSA和TAA在其他CRC肿瘤中的瘤间分布,绘制了来自癌症基因组图谱(TCGA)(图6A)的151个结肠腺癌样品中肽编码序列的log(RPHM+1)表达。为了评价抗原的共享潜力,对于目的肽,首先计算合并的GTE_x(n=2442)和mTEC(n=8)样品的对数转换(log(rphm+1))值的平均值。总体而言,在至少5%的TCGA COAD肿瘤中,九个TSA(53%)和九个TAA(100%)具有的表达是其相应的平均GTE_x/mTEC值的 ≥ 10 倍高。这表明TAA在COAD TCGA肿瘤中比其TSA对应物更频繁地共享。然而,这也意指大多数TSA在这些样品中是高度共享的。

[0306] 肿瘤抗原识别的另一个重要考虑因素是这些肽是否能够引起有效的抗肿瘤免疫应答。免疫原性的表位预测揭露了,预计aeTSA的免疫原性显著高于推定无免疫原性的一组胸腺肽(67)(图6B)。另外,与典型TAA和编码TA总体(来源于编码区的TSA和TAA)相比,aeTSA的免疫原性评分显著更高。事实上,根据这些预测,来自典型区域的TAA的免疫原性显著低于胸腺肽($p < 0.01$)。这可能部分是由于经过验证的TAA数量较少。如果将这些预测与31种TAA的整个集合一起考虑,则不再是这种情况(图13)。考虑到所有31种TAA,MSI TA的免疫原性预计比胸腺肽更强,同时与MSS相比,MSI组织来源的TA的免疫原性预计也有统计学上的显著增加。

[0307] 最终,寻找了具有预期结合和呈递肿瘤抗原的等位基因的个体比例的近似值(图6C)。样品中的许多抗原相当普遍,并且根据IEDB人口覆盖工具的估计,预期80.64%的美国人口表达与本研究中所识别的TA相关的等位基因的至少一种。

[0308] 尽管已经通过其特定实施方式的方式在上文中描述了本发明,但是在不脱离所附权利要求中限定的本发明的精神和本质的情况下可以对其进行修改。在权利要求中,词语“包括(comprising)”用作开放式术语,基本上等同于短语“包括(including)但不限于”。除非上下文另有明确规定,否则单数形式“一个”、“一种”和“该/所述”包括相应的复数参考。

[0309] 参考文献

[0310] 1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68, 394-424

[0311] 2. Arnold, M., Sierra, M. S., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2017) Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 66, 683-691

[0312] 3. Eriksen, A. C., Sorensen, F. B., Lindebjerg, J., Hager, H., dePont Christensen, R., Kjaer-Frifeldt, S., and Hansen, T. F. (2018) The Prognostic Value of Tumor-Infiltrating lymphocytes in Stage II Colon Cancer. A Nationwide Population-Based Study. *Transl Oncol* 11, 979-987

- [0313] 4.Zhao,Y.,Ge,X.,He,J.,Cheng,Y.,Wang,Z.,Wang,J.,and Sun,L. (2019) The prognostic value of tumor-infiltrating lymphocytes in colorectal cancer differs by anatomical subsite:a systematic review and meta-analysis.*World J Surg Oncol* 17,85
- [0314] 5.Le,D.T.,Uram,J.N.,Wang,H.,Bartlett,B.R.,Kemberling,H.,Eyring,A.D.,Skora,A.D.,Luber,B.S.,Azad,N.S.,Laheru,D.,Biedrzycki,B.,Donehower,R.C.,Zaheer,A.,Fisher,G.A.,Crocenzi,T.S.,Lee,J.J.,Duffy,S.M.,Goldberg,R.M.,de la Chapelle,A.,Koshiji,M.,Bhaijee,F.,Huebner,T.,Hruban,R.H.,Wood,L.D.,Cuka,N.,Pardoll,D.M.,Papadopoulos,N.,Kinzler,K.W.,Zhou,S.,Cornish,T.C.,Taube,J.M.,Anders,R.A.,Eshleman,J.R.,Vogelstein,B.,and Diaz,L.A.,Jr. (2015) PD-1Blockade in Tumors with Mismatch-Repair Deficiency.*N Engl J Med* 372,2509-2520
- [0315] 6.Fabrizio,D.A.,George,T.J.,Jr.,Dunne,R.F.,Frampton,G.,Sun,J.,Gowen,K.,Kennedy,M.,Greenbowe,J.,Schrock,A.B.,Hezel,A.F.,Ross,J.S.,Stephens,P.J.,Ali,S.M.,Miller,V.A.,Fakih,M.,and Klempner,S.J. (2018) Beyond microsatellite testing:assessment of tumor mutational burden identifies subsets of colorectal cancer who may respond to immune checkpoint inhibition.*J Gastrointest Oncol* 9,610-617
- [0316] 7.Wagner,S.,Mullins,C.S.,and Linnebacher,M. (2018) Colorectal cancer vaccines:Tumor-associated antigens vs neoantigens.*World J Gastroenterol* 24, 5418-5432
- [0317] 8.Loffler,M.W.,Kowalewski,D.J.,Backert,L.,Bernhardt,J.,Adam,P.,Schuster,H.,Dengler,F.,Backes,D.,Kopp,H.G.,Beckert,S.,Wagner,S.,Konigsrainer,I.,Kohlbacher,O.,Kanz,L.,Konigsrainer,A.,Rammensee,H.G.,Stevanovic,S.,and Haen,S.P. (2018) Mapping the HLA Ligandome of Colorectal Cancer Reveals an Imprint of Malignant Cell Transformation.*Cancer Res* 78,4627-4641
- [0318] 9.Picard,E.,Verschoor,C.P.,Ma,G.W.,and Pawelec,G. (2020) Relationships Between Immune Landscapes,Genetic Subtypes and Responses to Immunotherapy in Colorectal Cancer.*Front Immunol* 11,369
- [0319] 10.Parkhurst,M.R.,Yang,J.C.,Langan,R.C.,Dudley,M.E.,Nathan,D.A.,Feldman,S.A.,Davis,J.L.,Morgan,R.A.,Merino,M.J.,Sherry,R.M.,Hughes,M.S.,Kammula,U.S.,Phan,G.Q.,Lim,R.M.,Wank,S.A.,Restifo,N.P.,Robbins,P.F.,Laurencot,C.M.,and Rosenberg,S.A. (2011) T cells targeting carcinoembryonic antigen can mediate regression of metastatic colorectal cancer but induce severe transient colitis.*Mol Ther* 19,620-626
- [0320] 11.Minati,R.,Perreault,C.,and Thibault,P. (2020) A Roadmap Toward the Definition of Actionable Tumor-Specific Antigens.*Front Immunol* 11,583287
- [0321] 12.Smith,C.C.,Selitsky,S.R.,Chai,S.,Armistead,P.M.,Vincent,B.G.,and Serody,J.S. (2019) Alternative tumour-specific antigens.*Nat Rev Cancer* 19,465-478

- [0322] 13.Kloor,M.,Reuschenbach,M.,Karch,J.,Rafiiyan,M.,Al-Batran,S.-E., Pauligk,C.,Jaeger,E.,and Doeberitz,M.v.K. (2015) Vaccination of MSI-H colorectal cancer patients with frameshift peptide antigens:A phase I/IIa clinical trial.Journal of Clinical Oncology 33,3020-3020
- [0323] 14.van den Bulk,J.,Verdegaal,E.M.E.,Ruano,D.,Ijsselsteijn,M.E., Visser,M.,van der Breggen,R.,Duhon,T.,van der Ploeg,M.,de Vries,N.L.,Oosting, J.,Peeters,K.,Weinberg,A.D.,Farina-Sarasqueta,A.,van der Burg,S.H.,and de Miranda,N. (2019) Neoantigen-specific immunity in low mutation burden colorectal cancers of the consensus molecular subtype 4.Genome Med11,87
- [0324] 15.Newey,A.,Griffiths,B.,Michaux,J.,Pak,H.S.,Stevenson,B.J.,Woolston, A.,Semiannikova,M.,Spain,G.,Barber,L.J.,Matthews,N.,Rao,S.,Watkins,D.,Chau, I.,Coukos,G.,Racle,J.,Gfeller,D.,Starling,N.,Cunningham,D.,Bassani-Sternberg, M.,and Gerlinger,M. (2019) Immunopeptidomics of colorectal cancer organoids reveals a sparse HLA class I neoantigen landscape and no increase in neoantigens with interferon or MEK-inhibitor treatment.J Immunother Cancer 7, 309
- [0325] 16.Laumont,C.M.,Vincent,K.,Hesnard,L.,Audemard,E.,Bonneil,E., Laverdure,J.P.,Gendron,P.,Courcelles,M.,Hardy,M.P.,Cote,C.,Durette,C.,St- Pierre,C.,Benhamadi,M.,Lanoix,J.,Vobecky,S.,Haddad,E.,Lemieux,S.,Thibault, P.,and Perreault,C. (2018) Noncoding regions are the main source of targetable tumor-specific antigens.Sci Transl Med 10
- [0326] 17.Zhao,Q.,Laverdure,J.P.,Lanoix,J.,Durette,C.,Cote,C.,Bonneil,E., Laumont,C.M.,Gendron,P.,Vincent,K.,Courcelles,M.,Lemieux,S.,Millar,D.G., Ohashi,P.S.,Thibault,P.,and Perreault,C. (2020) Proteogenomics Uncovers a Vast Repertoire of Shared Tumor-Specific Antigens in Ovarian Cancer.Cancer Immunol Res 8,544-555
- [0327] 18.Ehx,G.,Larouche,J.D.,Durette,C.,Laverdure,J.P.,Hesnard,L.,Vincent, K.,Hardy,M.P.,Theriault,C.,Rulleau,C.,Lanoix,J.,Bonneil,E.,Feghaly,A., Apavaloaei,A.,Noronha,N.,Laumont,C.M.,Delisle,J.S.,Vago,L.,Hebert,J., Sauvageau,G.,Lemieux,S.,Thibault,P.,and Perreault,C. (2021) Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes.Immunity
- [0328] 19.Bolger,A.M.,Lohse,M.,and Usadel,B. (2014) Trimmomatic:a flexible trimmer for Illumina sequence data.Bioinformatics 30,2114-2120
- [0329] 20.Dobin,A.,Davis,C.A.,Schlesinger,F.,Drenkow,J.,Zaleski,C.,Jha,S., Batut,P.,Chaisson,M.,and Gingeras,T.R. (2013) STAR:ultrafast universal RNA-seq aligner.Bioinformatics 29,15-21
- [0330] 21.Bray,N.L.,Pimentel,H.,Melsted,P.,and Pachter,L. (2016) Near-optimal probabilistic RNA-seq quantification.Nat Biotechnol 34,525-527

- [0331] 22.Szolek,A.,Schubert,B.,Mohr,C.,Sturm,M.,Feldhahn,M.,and Kohlbacher,O. (2014) OptiType:precision HLA typing from next-generation sequencing data.Bioinformatics 30,3310-3316
- [0332] 23.Jia,P.,Yang,X.,Guo,L.,Liu,B.,Lin,J.,Liang,H.,Sun,J.,Zhang,C.,and Ye,K. (2020) MSIsensor-pro:Fast,Accurate,and Matched-normal-sample-free Detection of Microsatellite Instability.Genomics Proteomics Bioinformatics 18,65-71
- [0333] 24.Love,M.I.,Huber,W.,and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.Genome Biol 15,550
- [0334] 25.Zhou,Y.,Zhou,B.,Pache,L.,Chang,M.,Khodabakhshi,A.H.,Tanaseichuk,O.,Benner,C.,and Chanda,S.K. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.Nat Commun 10,1523
- [0335] 26.Hardy,M.P.,Audemard,E.,Migneault,F.,Feghaly,A.,Brochu,S.,Gendron,P.,Boilard,E.,Major,F.,Dieude,M.,Hebert,M.J.,and Perreault,C. (2019) Apoptotic endothelial cells release small extracellular vesicles loaded with immunostimulatory viral-like RNAs.Sci Rep 9,7203
- [0336] 27.Karolchik,D.,Hinrichs,A.S.,Furey,T.S.,Roskin,K.M.,Sugnet,C.W.,Haussler,D.,and Kent,W.J. (2004) The UCSC Table Browser data retrieval tool.Nucleic Acids Res 32,D493-496
- [0337] 28.Cingolani,P.,Platts,A.,Wang le,L.,Coon,M.,Nguyen,T.,Wang,L.,Land,S.J.,Lu,X.,and Ruden,D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff:SNPs in the genome of Drosophila melanogaster strain w1118;iso-2;iso-3.Fly(Austin)6,80-92
- [0338] 29.Daouda,T.,Perreault,C.,and Lemieux,S. (2016) pyGeno:A Python package for precision medicine and proteogenomics.F1000Res 5,381
- [0339] 30.Lanoix,J.,Durette,C.,Courcelles,M.,Cossette,E.,Comtois-Marotte,S.,Hardy,M.P.,Côté,C.,Perreault,C.,and Thibault,P. (2018) Comparison of the MHC I immunopeptidome repertoire of B-cell lymphoblasts using two isolation methods.Proteomics 18,e1700251
- [0340] 31.Ma,B.,Zhang,K.,Hendrie,C.,Liang,C.,Li,M.,Doherty-Kirby,A.,and Lajoie,G. (2003) PEAKS:powerful software for peptide de novo sequencing by tandem mass spectrometry.Rapid Commun Mass Spectrom 17,2337-2342
- [0341] 32.Courcelles,M.,Durette,C.,Daouda,T.,Laverdure,J.P.,Vincent,K.,Lemieux,S.,Perreault,C.,and Thibault,P. (2020) MAPDP:A Cloud-Based Computational Platform for Immunopeptidomics Analyses.J Proteome Res 19,1873-1881
- [0342] 33.Wu,T.D.,Reeder,J.,Lawrence,M.,Becker,G.,and Brauer,M.J. (2016) GMAP and GSNAP for Genomic Sequence Alignment:Enhancements to Speed,Accuracy,and Functionality.Methods Mol Biol 1418,283-334

- [0343] 34.Ogishi,M.,and Yotsuyanagi,H. (2019) Quantitative Prediction of the Landscape of T Cell Epitope Immunogenicity in Sequence Space.Front Immunol 10,827
- [0344] 35.Conway,J.R.,Lex,A.,and Gehlenborg,N. (2017) UpSetR:an R package for the visualization of intersecting sets and their properties.Bioinformatics 33,2938-2940
- [0345] 36.Hanzelmann,S.,Castelo,R.,and Guinney,J. (2013) GSVA:gene set variation analysis for microarray and RNA-seq data.BMC Bioinformatics 14,7
- [0346] 37.Yoshihara,K.,Shahmoradgoli,M.,Martinez,E.,Vegesna,R.,Kim,H.,Torres-Garcia,W.,Trevino,V.,Shen,H.,Laird,P.W.,Levine,D.A.,Carter,S.L.,Getz,G.,Stemke-Hale,K.,Mills,G.B.,and Verhaak,R.G. (2013) Inferring tumour purity and stromal and immune cell admixture from expression data.Nat Commun 4,2612
- [0347] 38.Chan,A.Y.,and Anderson,M.S. (2015) Central tolerance to self revealed by the autoimmune regulator.Ann N Y Acad Sci 1356,80-89
- [0348] 39.Pfammatter,S.,Bonneil,E.,Lanoix,J.,Vincent,K.,Hardy,M.P.,Courcelles,M.,Perreault,C.,and Thibault,P. (2020) Extending theComprehensiveness of Immunoepitidome Analyses Using Isobaric PeptideLabeling.Anal Chem 92,9194-9204
- [0349] 40.Pira,G.,Uva,P.,Scanu,A.M.,Rocca,P.C.,Murgia,L.,Uleri,E.,Piu,C.,Porcu,A.,Carru,C.,Manca,A.,Persico,I.,Muronì,M.R.,Sanges,F.,Serra,C.,Dolei,A.,Angius,A.,and De Miglio,M.R. (2020) Landscape oftranscriptome variations uncovering known and novel driver events in colorectalcarcinoma.Sci Rep 10,432
- [0350] 41.Kawakami,H.,Zaanan,A.,and Sinicrope,F.A. (2015) Microsatelliteinstability testing and its role in the management of colorectal cancer.Curr TreatOptions Oncol 16,30
- [0351] 42.Jimeno,A.,Messersmith,W.A.,Hirsch,F.R.,Franklin,W.A.,andEckhardt,S.G. (2009) KRAS mutations and sensitivity to epidermal growthfactor receptor inhibitors in colorectal cancer:practical application of patientselection.J Clin Oncol 27,1130-1136
- [0352] 43.Van Cutsem,E.,Kohne,C.H.,Lang,I.,Folprecht,G.,Nowacki,M.P.,Cascinù,S.,Shchepotin,I.,Maurel,J.,Cunningham,D.,Tejpar,S.,Schlichting,M.,Zubel,A.,Celik,I.,Rougier,P.,and Ciardiello,F. (2011) Cetuximab plus irinotecan,fluorouracil,and leucovorin as first-line treatmentfor metastatic colorectal cancer:updated analysis of overall survival accordingto tumor KRAS and BRAF mutation status.J Clin Oncol 29,2011-2019
- [0353] 44.Ahmed,D.,Eide,P.W.,Eilertsen,I.A.,Danielsen,S.A.,Eknaes,M.,Hektoen,M.,Lind,G.E.,and Lothe,R.A. (2013) Epigenetic and geneticfeatures of 24 colon cancer cell lines.Oncogenesis 2,e71

- [0354] 45. Berg, K.C.G., Eide, P.W., Eilertsen, I.A., Johannessen, B., Bruun, J., Danielsen, S.A., Bjørnslett, M., Meza-Zepeda, L.A., Eknaes, M., Lind, G.E., Myklebost, O., Skotheim, R.I., Sveen, A., and Lothe, R.A. (2017) Multi-omic of 34 colorectal cancer cell lines—a resource for biomedical studies. *Mol Cancer* 16, 116
- [0355] 46. Niu, B., Ye, K., Zhang, Q., Lu, C., Xie, M., McLellan, M.D., Wendl, M.C., and Ding, L. (2014) MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* 30, 1015-1016
- [0356] 47. Aaltonen, L.A., Peltomäki, P., Mecklin, J.P., Jarvinen, H., Jass, J.R., Green, J.S., Lynch, H.T., Watson, P., Tallqvist, G., Juhola, M., and et al. (1994) Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. *Cancer Res* 54, 1645-1648
- [0357] 48. Llosa, N.J., Cruise, M., Tam, A., Wicks, E.C., Hechenbleikner, E.M., Taube, J.M., Blosser, R.L., Fan, H., Wang, H., Lubner, B.S., Zhang, M., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., Sears, C.L., Anders, R.A., Pardoll, D.M., and Housseau, F. (2015) The vigorous immunomicroenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. *Cancer Discov* 5, 43-51
- [0358] 49. Le, D.T., Durham, J.N., Smith, K.N., Wang, H., Bartlett, B.R., Aulakh, L.K., Lu, S., Kemberling, H., Wilt, C., Lubner, B.S., Wong, F., Azad, N.S., Rucki, A.A., Laheru, D., Donehower, R., Zaheer, A., Fisher, G.A., Crocenzi, T.S., Lee, J.J., Greten, T.F., Duffy, A.G., Ciombor, K.K., Eyring, A.D., Lam, B.H., Joe, A., Kang, S.P., Holdhoff, M., Danilova, L., Cope, L., Meyer, C., Zhou, S., Goldberg, R.M., Armstrong, D.K., Bever, K.M., Fader, A.N., Taube, J., Housseau, F., Spetzler, D., Xiao, N., Pardoll, D.M., Papadopoulos, N., Kinzler, K.W., Eshleman, J.R., Vogelstein, B., Anders, R.A., and Diaz, L.A., Jr. (2017) Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 357, 409-413
- [0359] 50. Hanahan, D., and Weinberg, R.A. (2000) The hallmarks of cancer. *Cell* 100, 57-70
- [0360] 51. Hanahan, D., and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell* 144, 646-674
- [0361] 52. Prossomariti, A., Piazzzi, G., Alquati, C., and Ricciardiello, L. (2020) Are Wnt/beta-Catenin and PI3K/AKT/mTORC1 Distinct Pathways in Colorectal Cancer? *Cell Mol Gastroenterol Hepatol* 10, 491-506
- [0362] 53. Danaher, P., Warren, S., Dennis, L., D'Amico, L., White, A., Disis, M.L., Geller, M.A., Odunsi, K., Beechem, J., and Fling, S.P. (2017) Gene expression markers of Tumor Infiltrating Leukocytes. *J Immunother Cancer* 5, 18
- [0363] 54. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., Frohling, S., Chan, E.M., Sos, M.L., Michel, K., Mermel, C., Silver, S.J., Weir, B.A., Reiling, J.H., Sheng, Q., Gupta, P.B., Wadlow, R.C., Le, H., Hoersch, S., Wittner, B.S., Ramaswamy, S., Livingston, D.M.,

- Sabatini,D.M.,Meyerson,M.,Thomas,R.K.,Lander,E.S.,Mesirov,J.P.,Root,D.E., Gilliland,D.G.,Jacks,T.,andHahn,W.C. (2009)Systematic RNA interference reveals that oncogenicKRAS-driven cancers require TBK1.Nature 462,108-112
- [0364] 55.Kim,H.,Jen,J.,Vogelstein,B.,and Hamilton,S.R. (1994)Clinicaland pathological characteristics of sporadic colorectal carcinomas with DNAreplication errors in microsatellite sequences.Am J Pathol 145,148-156
- [0365] 56.Smyrk,T.C.,Watson,P.,Kaul,K.,and Lynch,H.T. (2001)Tumor-infiltrating lymphocytes are a marker for microsatellite instability incolorectal carcinoma.Cancer 91,2417-2422
- [0366] 57.Dolcetti,R.,Viel,A.,Doglioni,C.,Russo,A.,Guidoboni,M.,Capozzi,E., Vecchiato,N.,Macri,E.,Fornasarig,M.,and Boiocchi,M. (1999)High prevalence of activated intraepithelial cytotoxic T lymphocytes andincreased neoplastic cell apoptosis in colorectal carcinomas with microsatelliteinstability.Am J Pathol 154,1805-1813
- [0367] 58.Phillips,S.M.,Banerjea,A.,Feakins,R.,Li,S.R.,Bustin,S.A., andDorudi,S. (2004)Tumour-infiltrating lymphocytes in colorectal cancer withmicrosatellite instability are activated and cytotoxic.Br J Surg 91,469-475
- [0368] 59.Boland,C.R.,Koi,M.,Chang,D.K.,and Carethers,J.M. (2008)The biochemical basis of microsatellite instability and abnormalimmunohistochemistry and clinical behavior in Lynch syndrome:from bench tobedside.Fam Cancer 7,41-52
- [0369] 60.Sarkizova,S.,Klaeger,S.,Le,P.M.,Li,L.W.,Oliveira,G.,Keshishian,H., Hartigan,C.R.,Zhang,W.,Braun,D.A.,Ligon,K.L.,Bachiredy,P.,Zervantonakis, I.K.,Rosenbluth,J.M.,Ouspenskaia,T.,Law,T.,Justesen,S.,Stevens,J.,Lane,W.J., Eisenhaure,T.,Lan Zhang,G.,Clauser,K.R.,Hacohen,N.,Carr,S.A.,Wu,C.J.,and Keskin,D.B. (2020)A large peptidome dataset improves HLA class I epitope prediction across most of the human population.Nat Biotechnol 38,199-209
- [0370] 61.Larouche,J.D.,Trofimov,A.,Hesnard,L.,Ehx,G.,Zhao,Q.,Vincent,K., Durette,C.,Gendron,P.,Laverdure,J.P.,Bonneil,E.,Cote,C.,Lemieux,S.,Thibault, P.,and Perreault,C. (2020)Widespread and tissue-specific expression of endogenous retroelements in human somatic tissues.Genome Med 12,40
- [0371] 62.Cherkasova,E.,Scrivani,C.,Doh,S.,Weisman,Q.,Takahashi,Y., Harashima,N.,Yokoyama,H.,Srinivasan,R.,Linehan,W.M.,Lerman,M.I.,and Childs, R.W. (2016)Detection of an Immunogenic HERV-E Envelope with Selective Expression in Clear Cell Kidney Cancer.Cancer Res 76,2177-2185
- [0372] 63.Patra,R.,Das,N.C.,and Mukherjee,S. (2021)Exploring the Differential Expression and Prognostic Significance of the COL11A1 Gene in Human Colorectal Carcinoma:An Integrated Bioinformatics Approach.Front Genet 12,

608313

[0373] 64. Marcu, A., Bichmann, L., Kuchenbecker, L., Kowalewski, D. J., Freudenmann, L. K., Backert, L., Mühlenbruch, L., Szolek, A., Lübke, M., Wagner, P., Engler, T., Matovina, S., Wang, J., Hauri-Hohl, M., Martin, R., Kapolou, K., Walz, J. S., Velz, J., Moch, H., Regli, L., Silginer, M., Weller, M., **Löffler, M. W.**, Erhard, F., Schlosser, A., Kohlbacher, O., **Stevanović, S.**, Rammensee, H.-G., and Neidert, M. C. (2020) The HLA Ligand Atlas-A resource of natural HLA ligands presented on benign tissues. *bioRxiv*, 778944

[0374] 65. Gjerstorff, M. F., Andersen, M. H., and Ditzel, H. J. (2015) Oncogenic cancer/testis antigens: prime candidates for immunotherapy. *Oncotarget* 6, 15772-15787

[0375] 66. Ishikawa, N., Takano, A., Yasui, W., Inai, K., Nishimura, H., Ito, H., Miyagi, Y., Nakayama, H., Fujita, M., Hosokawa, M., Tsuchiya, E., Kohno, N., Nakamura, Y., and Daigo, Y. (2007) Cancer-testis antigen lymphocyte antigen 6 complex locus K is a serologic biomarker and a therapeutic target for lung and esophageal carcinomas. *Cancer Res* 67, 11601-11611

[0376] 67. Adamopoulou, E., Tenzer, S., Hillen, N., Klug, P., Rota, I. A., Tietz, S., Gebhardt, M., Stevanovic, S., Schild, H., Tolosa, E., Melms, A., and Stoeckle, C. (2013) Exploring the MHC-peptide matrix of central tolerance in the human thymus. *Nat Commun* 4, 2039

[0377] 68. Kote, S., Pirog, A., Bedran, G., Alfaro, J., and Dapic, I. (2020) Mass Spectrometry-Based Identification of MHC-Associated Peptides. *Cancers (Basel)* 12

[0378] 69. Lin, A., Zhang, J., and Luo, P. (2020) Crosstalk Between the MSI Status and Tumor Microenvironment in Colorectal Cancer. *Front Immunol* 11, 2039

[0379] 70. Bonaventura, P., Shekarian, T., Alcazer, V., Valladeau-Guilemond, J., Valsesia-Wittmann, S., Amigorena, S., Caux, C., and Depil, S. (2019) Cold Tumors: A Therapeutic Challenge for Immunotherapy. *Front Immunol* 10, 168

[0380] 71. Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L., Edwards, K. J., Day, I. N., and Gaunt, T. R. (2013) Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 34, 57-65

[0381] 72. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., and Kircher, M. (2019) CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 47, D886-D894

[0382] 73. Sherry, S. T., Ward, M., and Sirotkin, K. (1999) dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 9, 677-679

- [0383] 74. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S.C., Kok, C.Y., Noble, K., Ponting, L., Ramshaw, C.C., Rye, C.E., Speedy, H.E., Stefancsik, R., Thompson, S.L., Wang, S., Ward, S., Campbell, P.J., and Forbes, S.A. (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 47, D941-D947
- [0384] 75. Ehx, G., and Perreault, C. (2019) Discovery and characterization of actionable tumor antigens. *Genome Med* 11, 29
- [0385] 76. Vogel, C., and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 13, 227-232
- [0386] 77. Xiang, B., Snook, A.E., Magee, M.S., and Waldman, S.A. (2013) Colorectal cancer immunotherapy. *Discov Med* 15, 301-308
- [0387] 78. Gold, P., and Freedman, S.O. (1965) Demonstration of Tumor-Specific Antigens in Human Colonic Carcinomata by Immunological Tolerance and Absorption Techniques. *J Exp Med* 121, 439-462
- [0388] 79. Zhou, F. (2009) Molecular mechanisms of IFN-gamma to up-regulate MHC class I antigen processing and presentation. *Int Rev Immunol* 28, 239-260
- [0389] 80. Deutsch, E.W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J.J., Kundu, D.J., Garcia-Seisdedos, D., Jarnuczak, A.F., Hewapathirana, S., Pullman, B.S., Wertz, J., Sun, Z., Kawano, S., Okuda, S., Watanabe, Y., Hermjakob, H., MacLean, B., MacCoss, M.J., Zhu, Y., Ishihama, Y., and Vizcaino, J.A. (2020) The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res* 48, D1145-D1152
- [0390] 81. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Perez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A.F., Ternent, T., Brazma, A., and Vizcaino, J.A. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47, D442-D450.

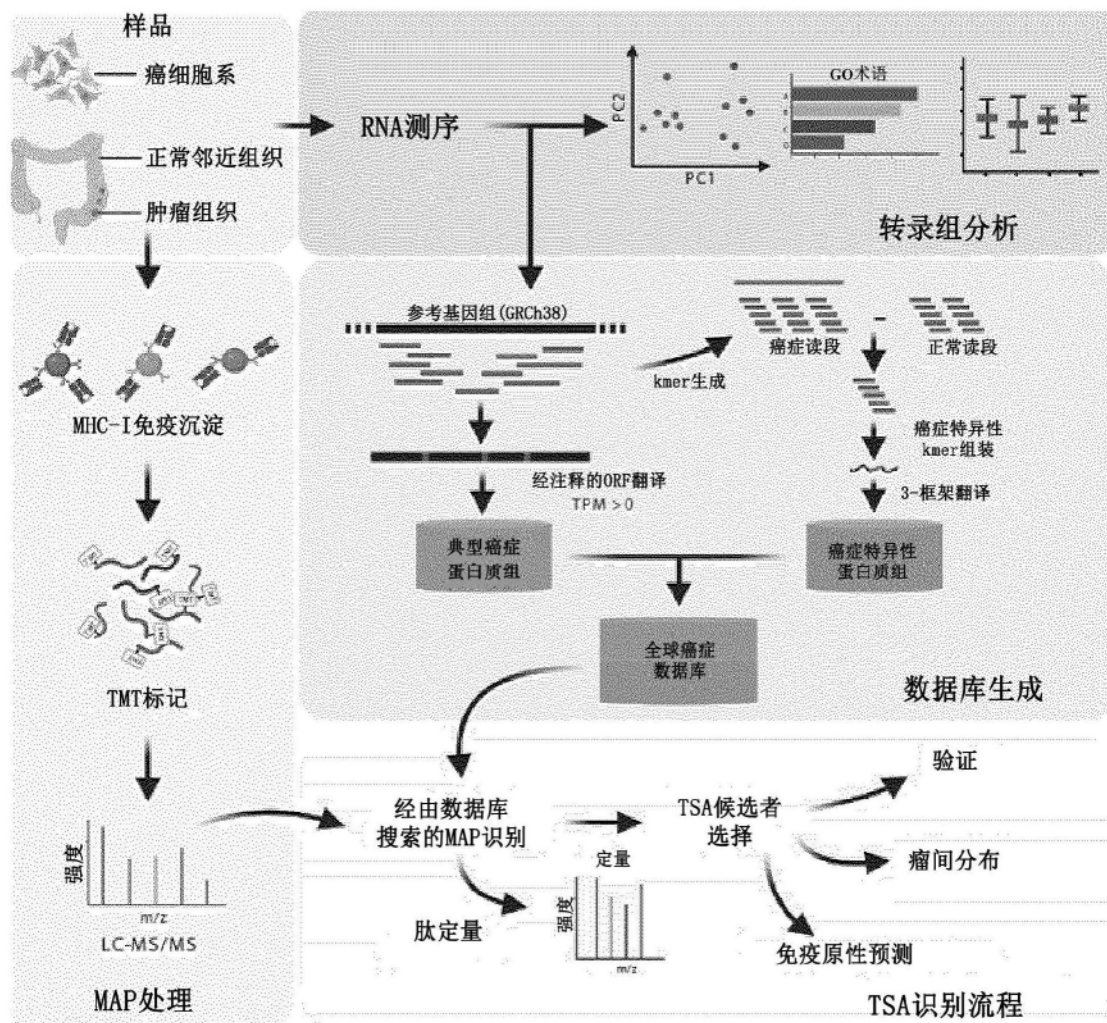


图1

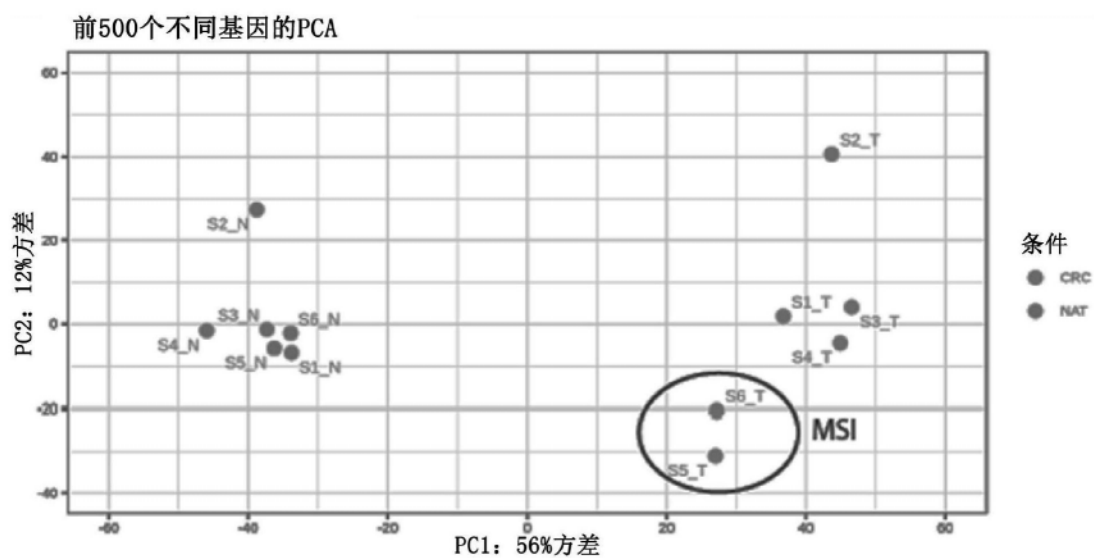


图2A

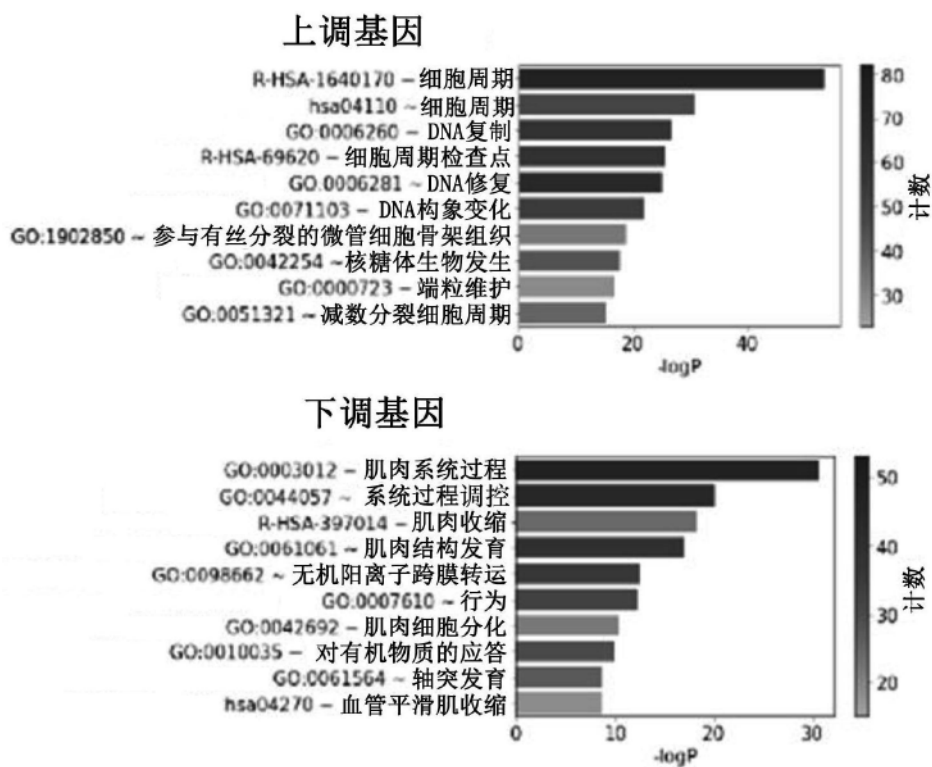


图2B

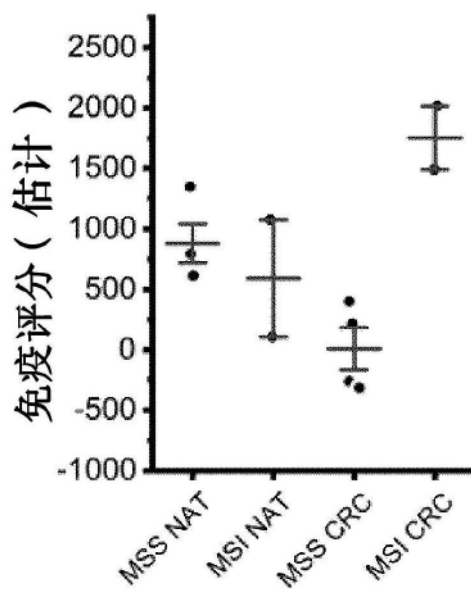


图2C

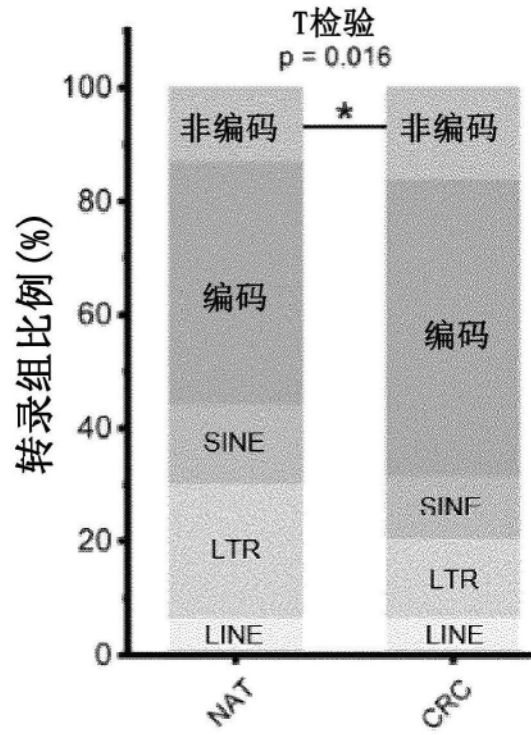


图2D

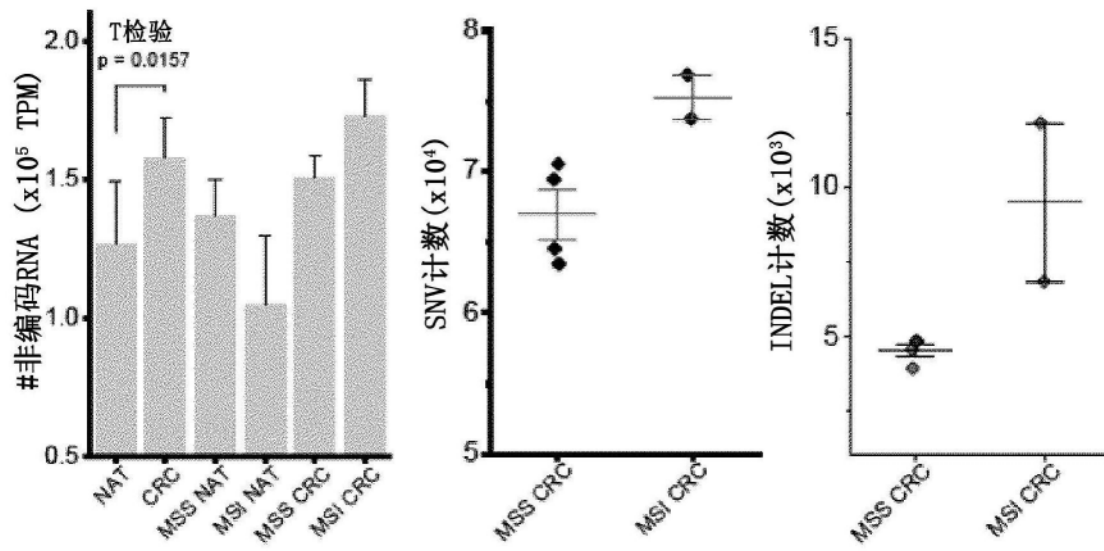


图2E

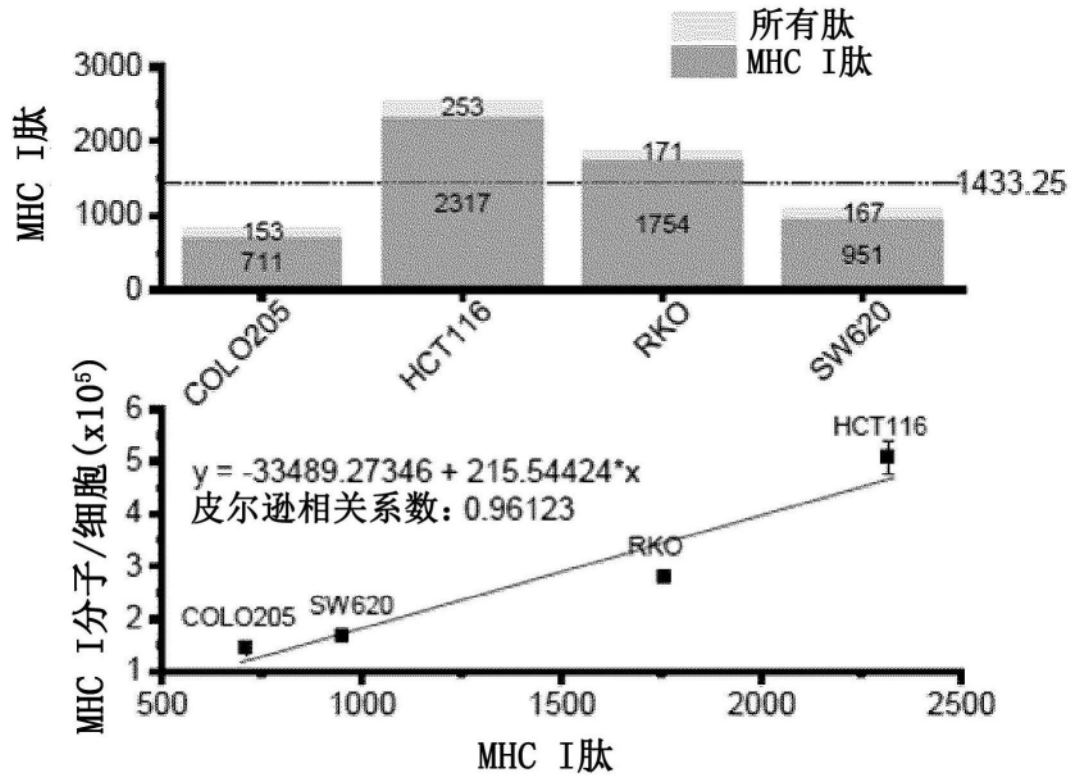


图3A

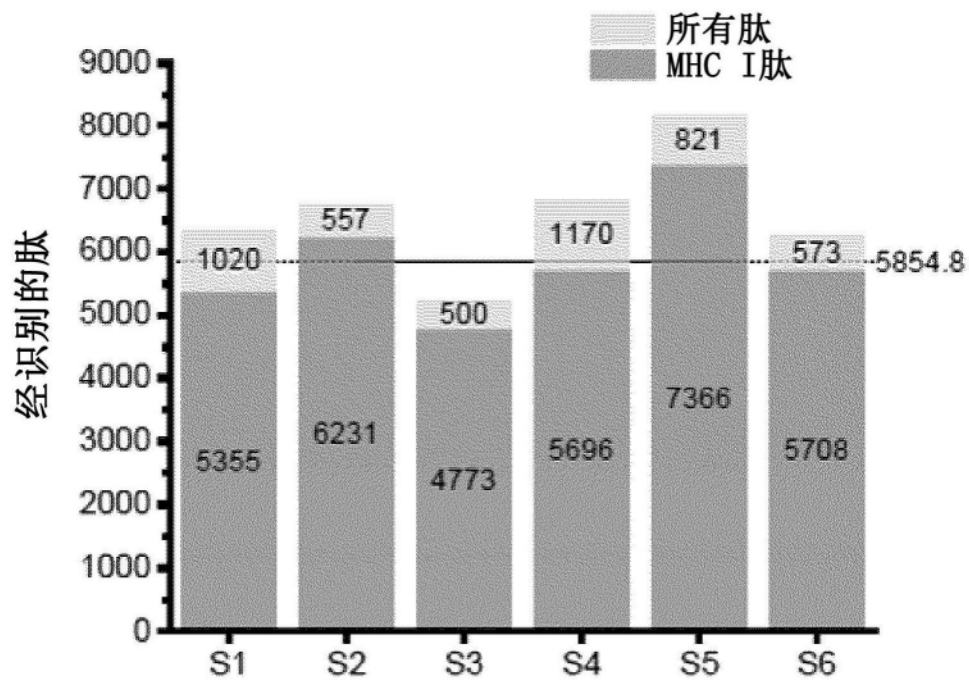


图3B

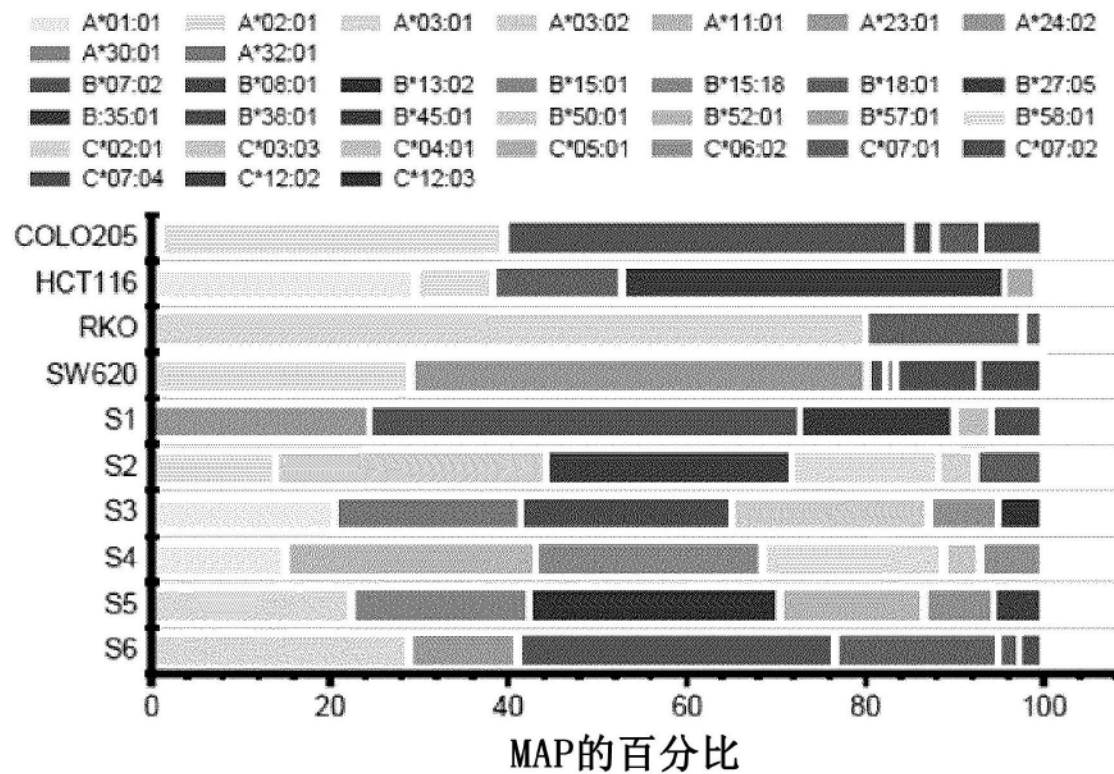
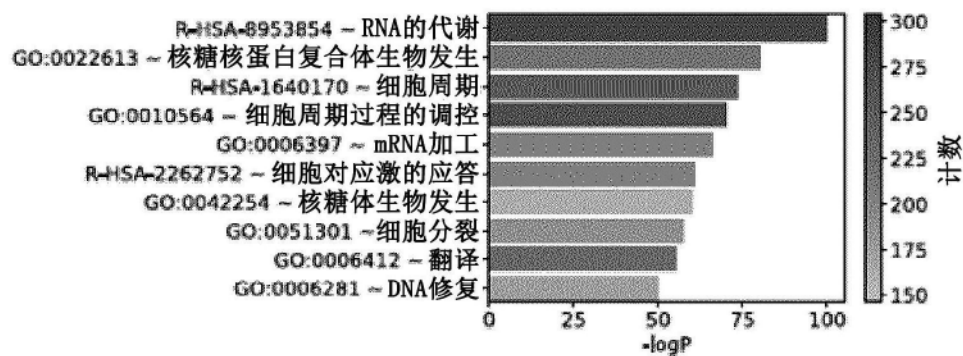


图3C



组织MAP来源基因

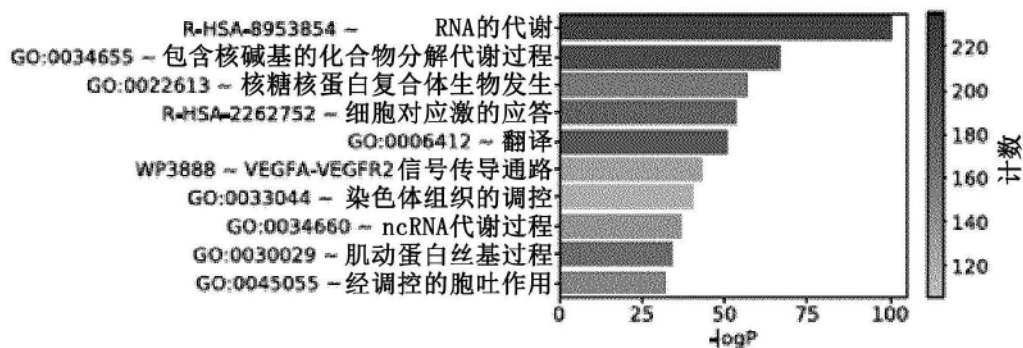


图3D

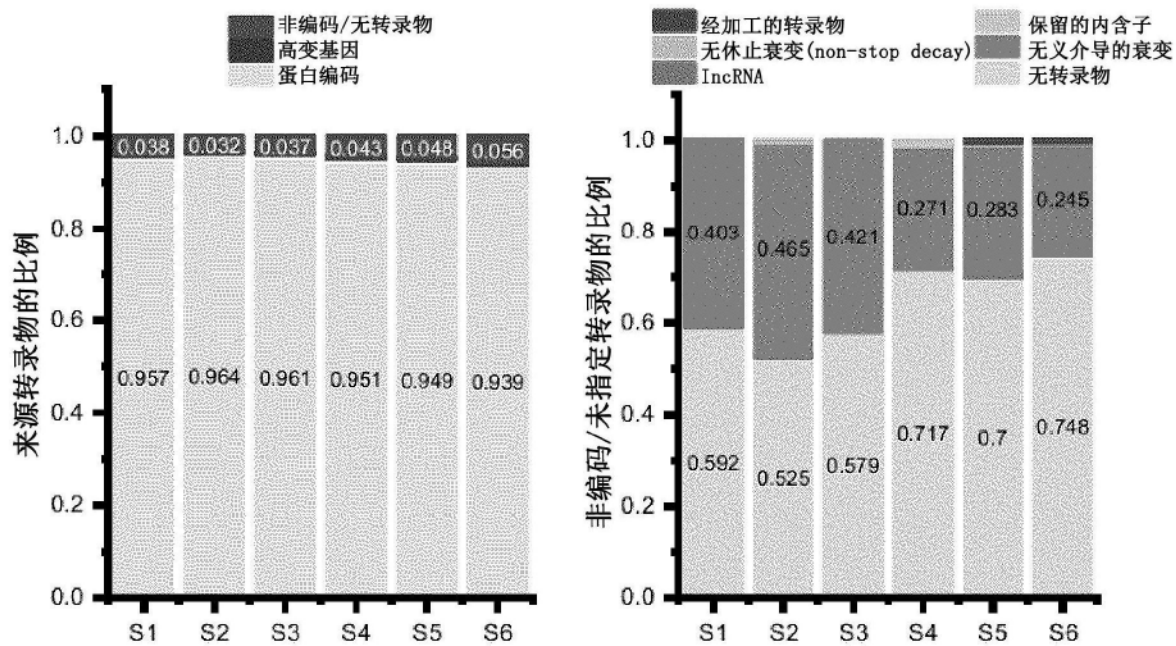


图3E

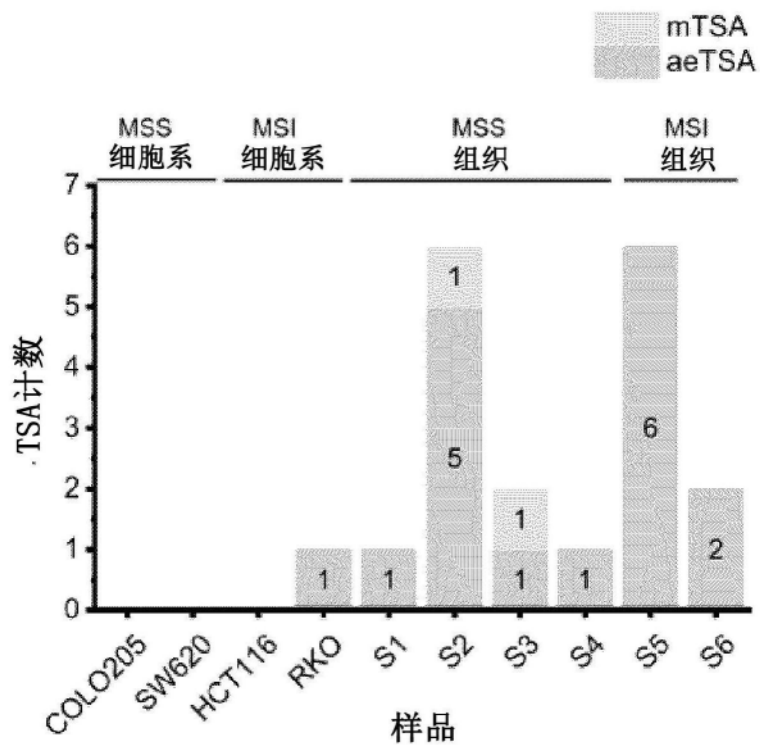


图4A

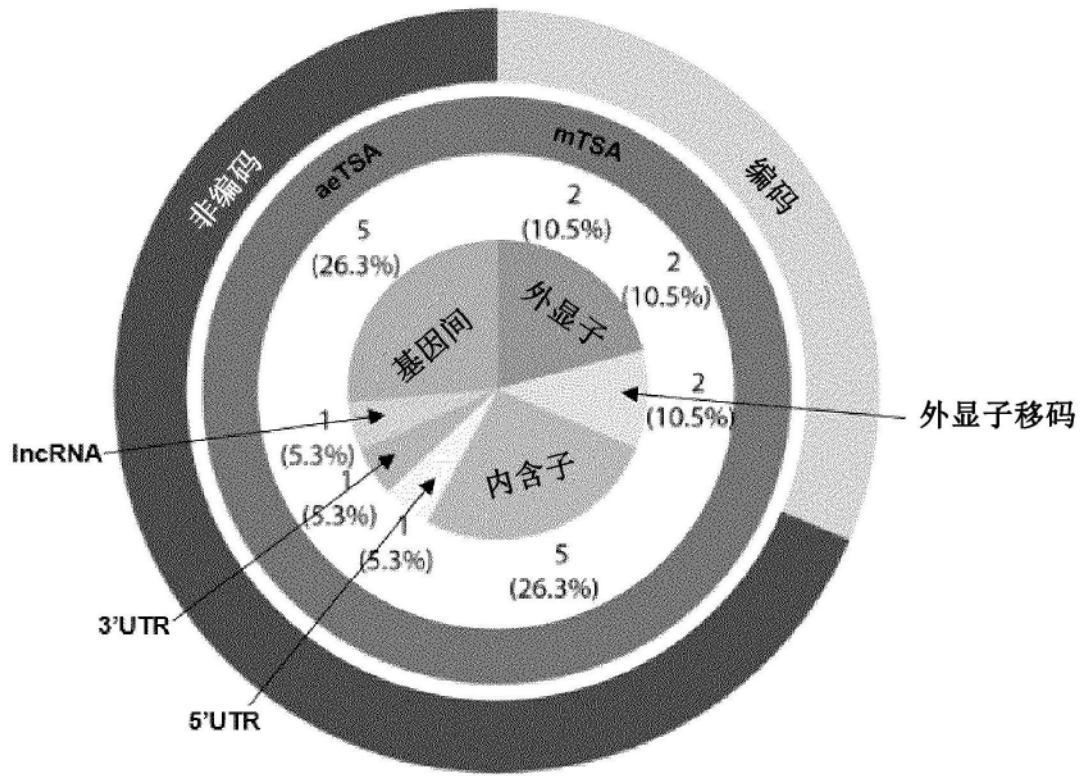


图4B

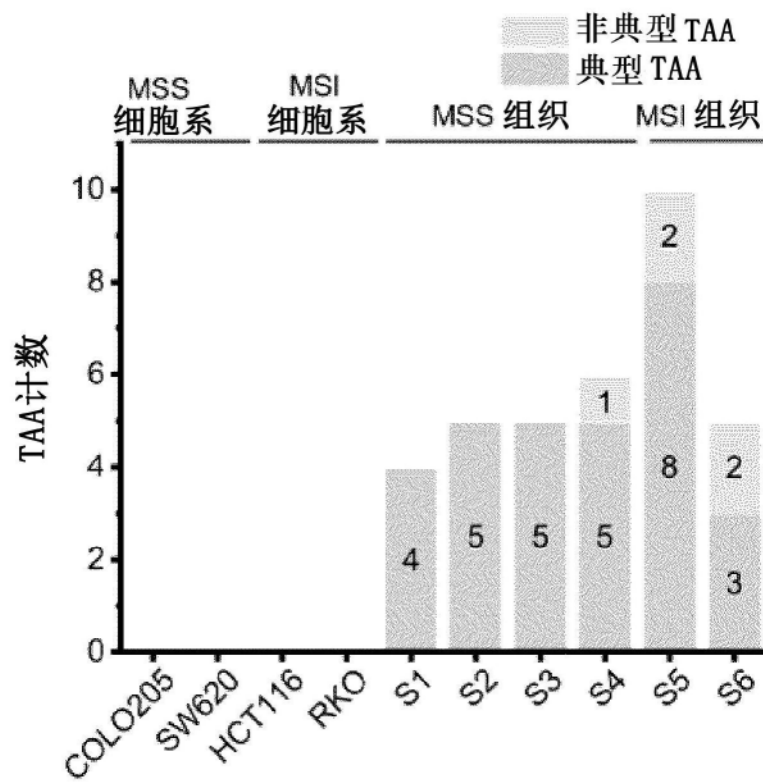


图4C

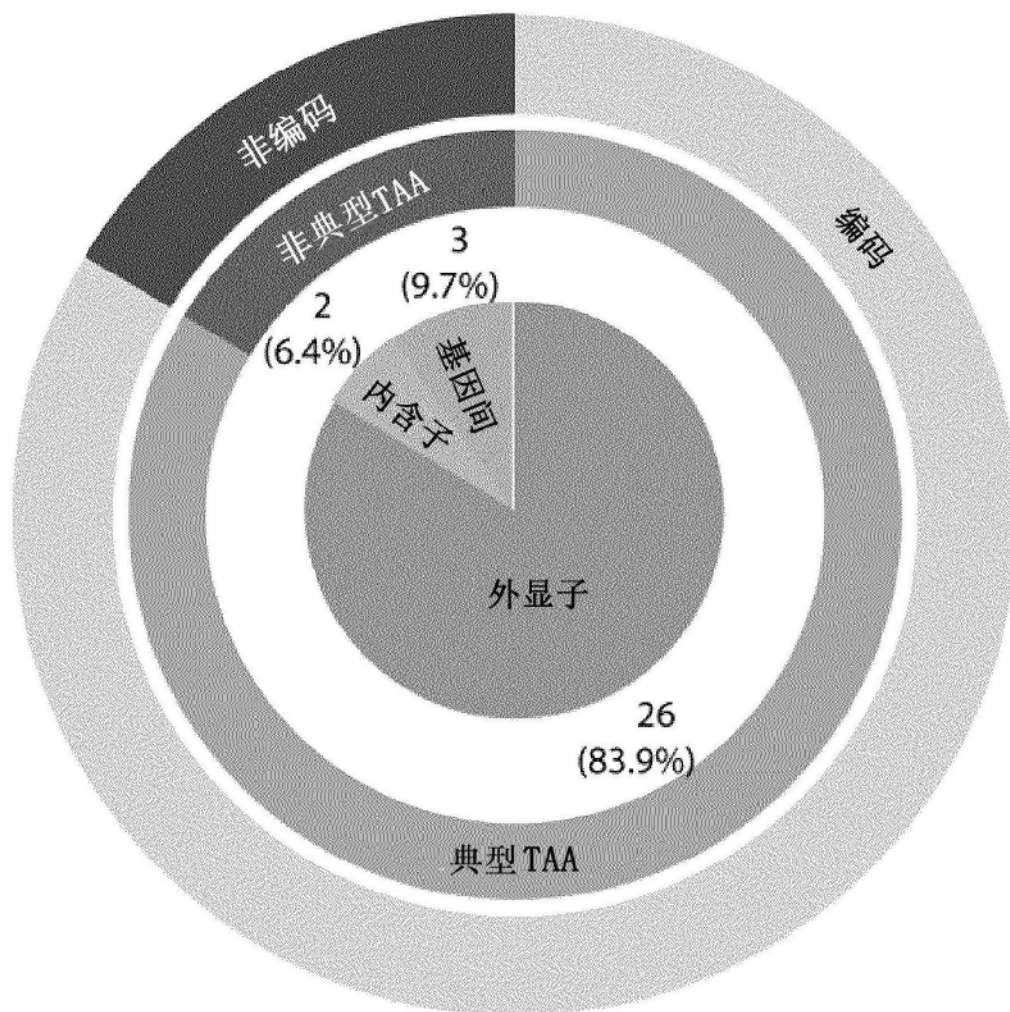


图4D

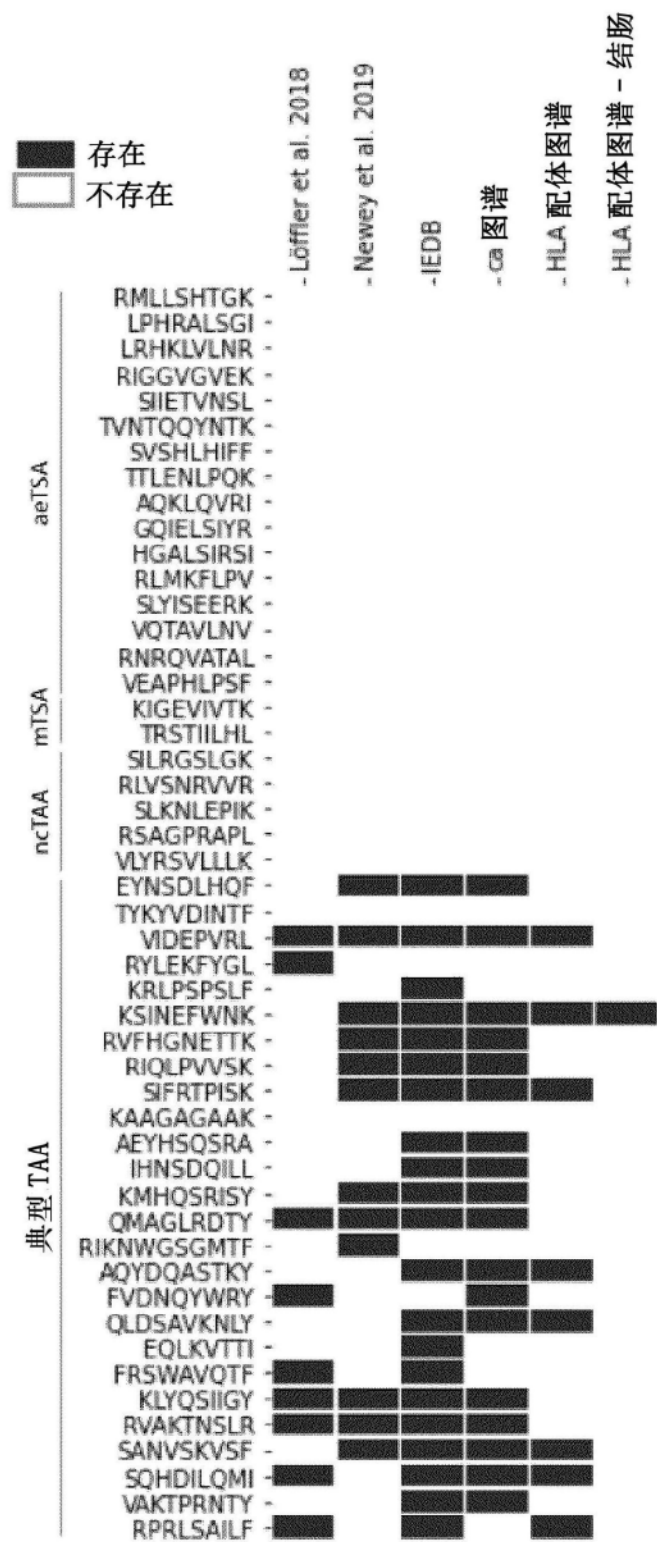


图4E



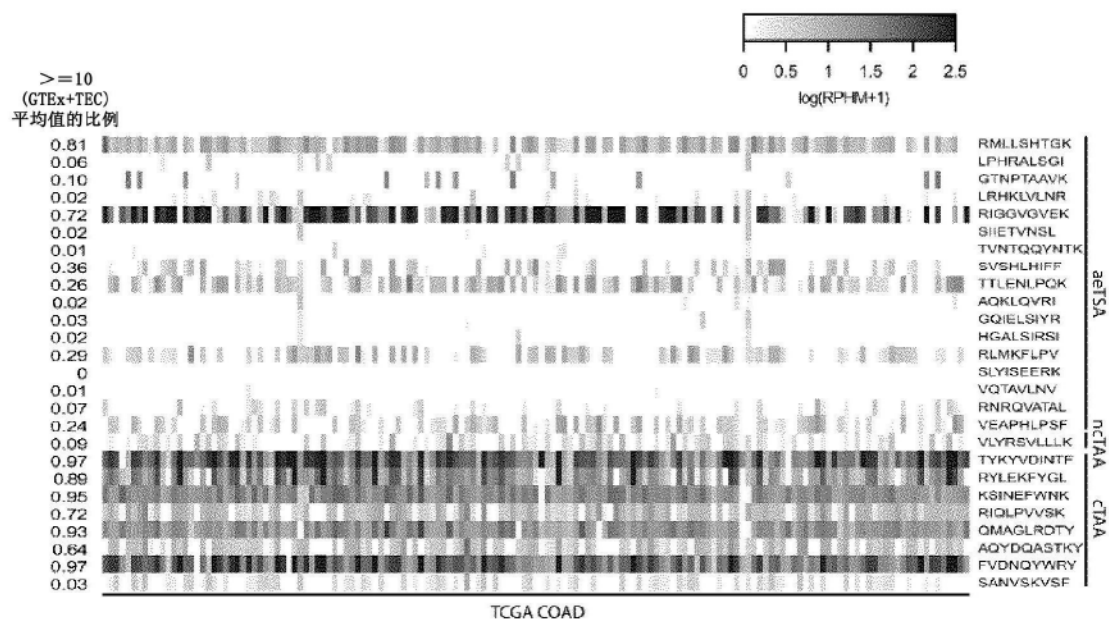


图6A

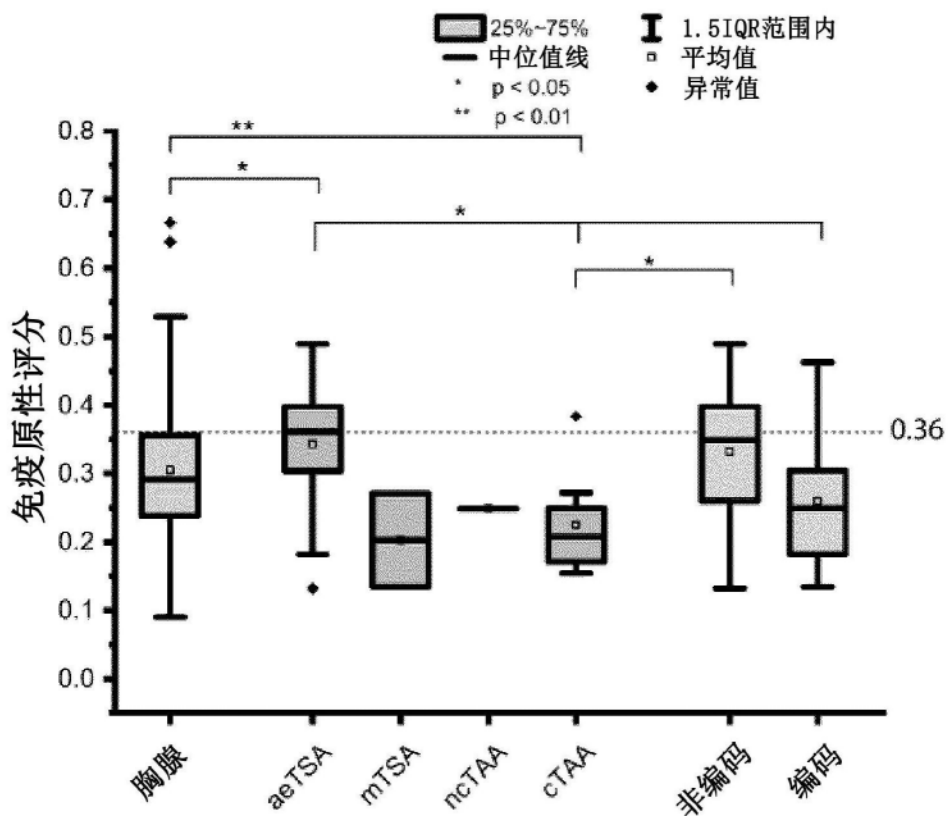


图6B

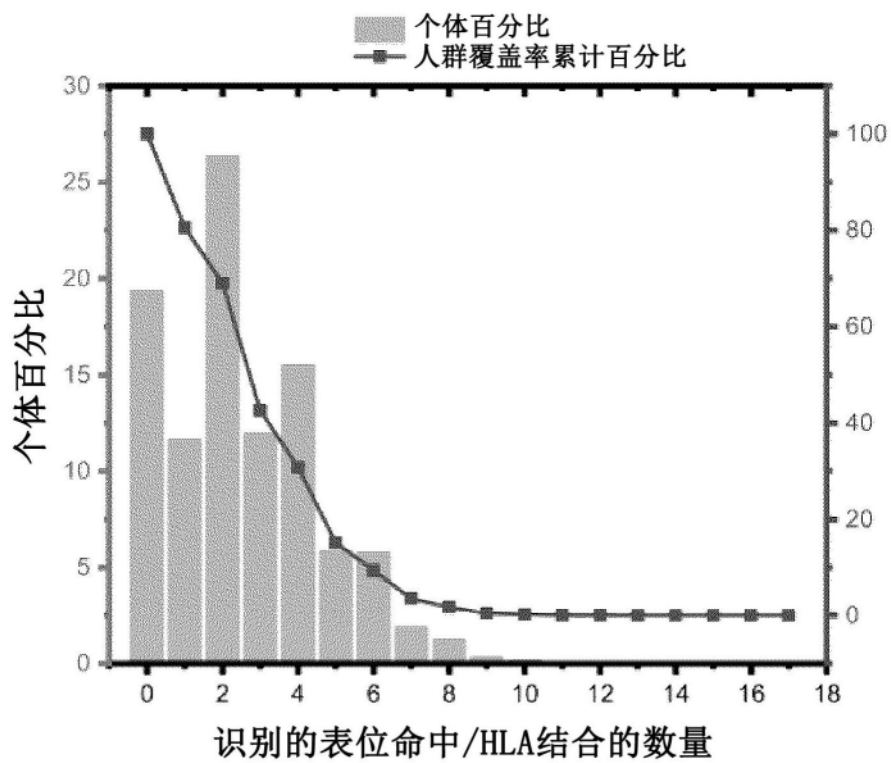


图6C

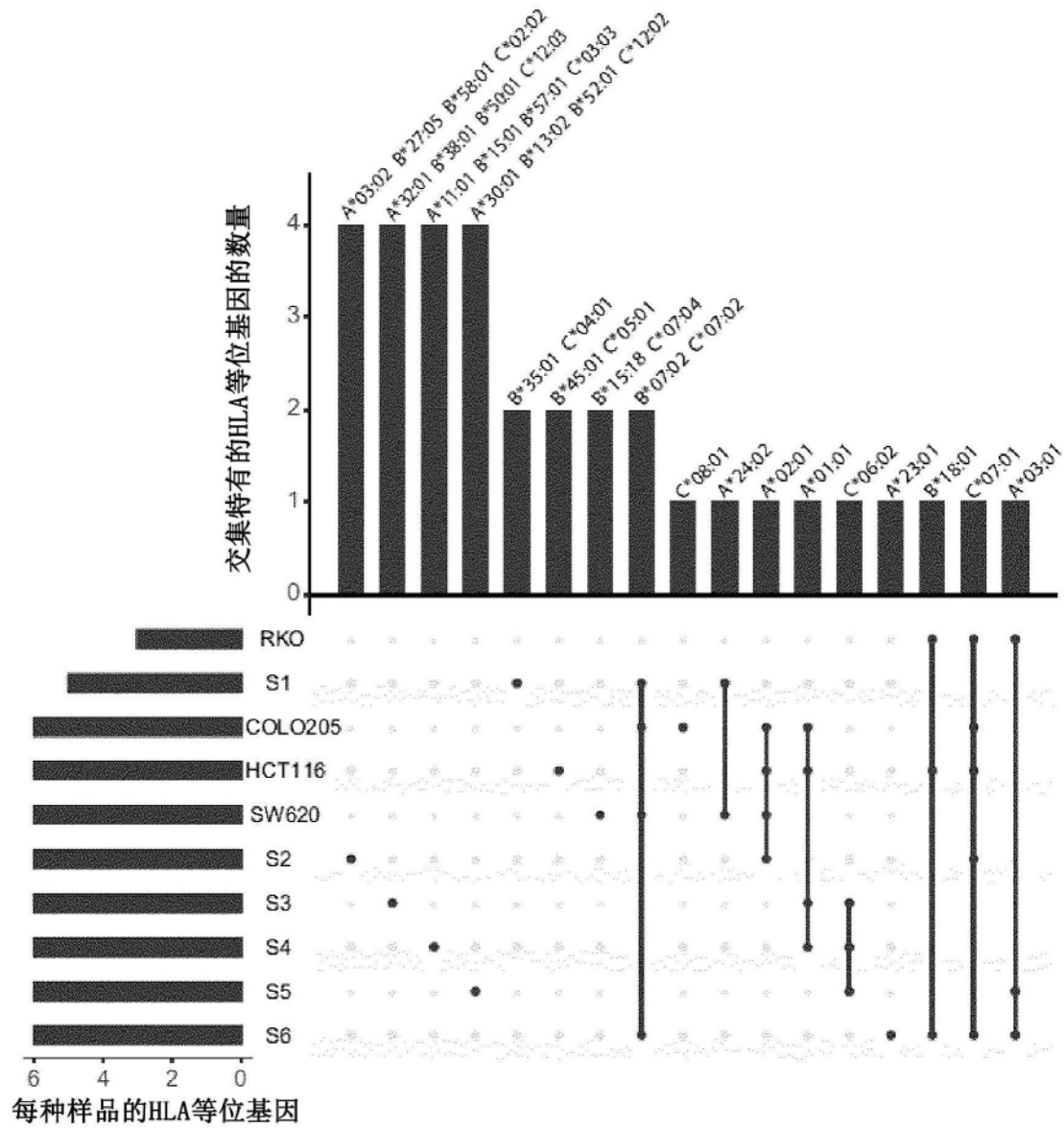


图7

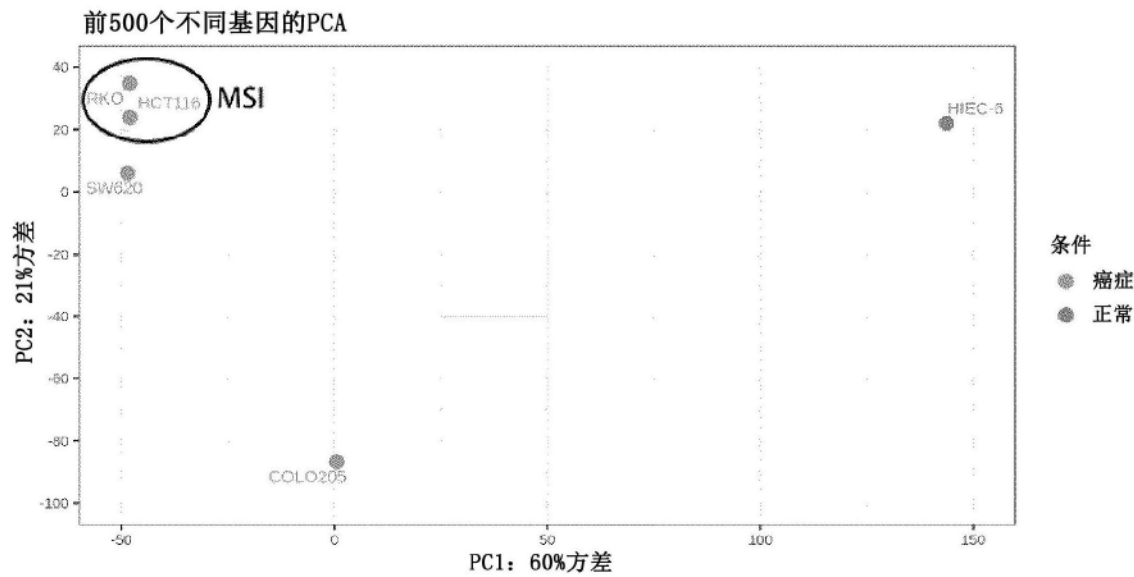


图8A

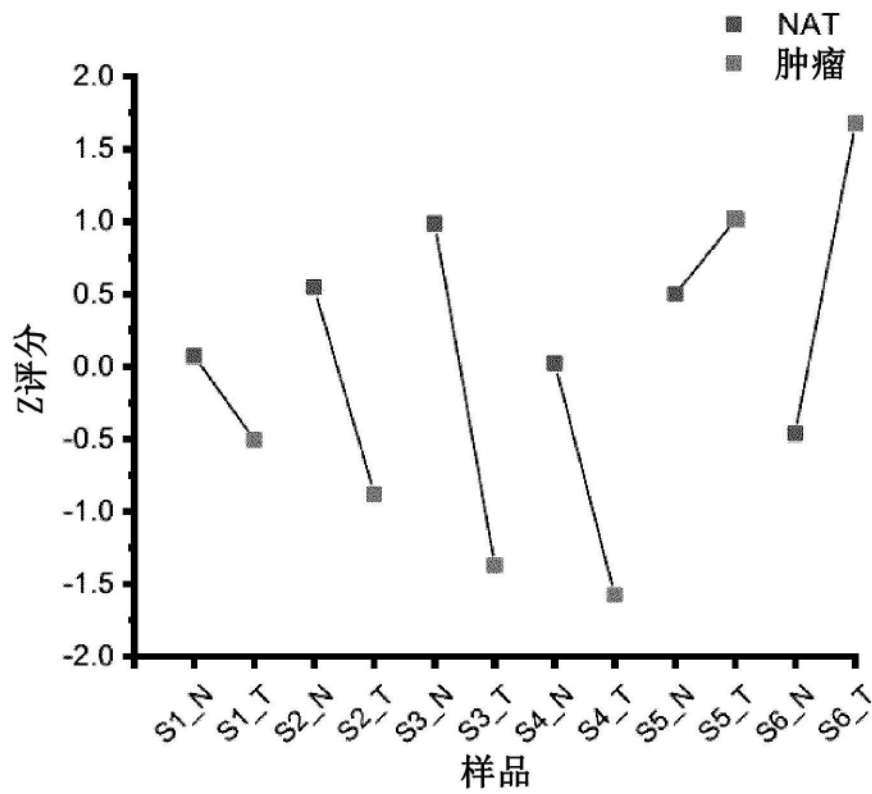


图8B

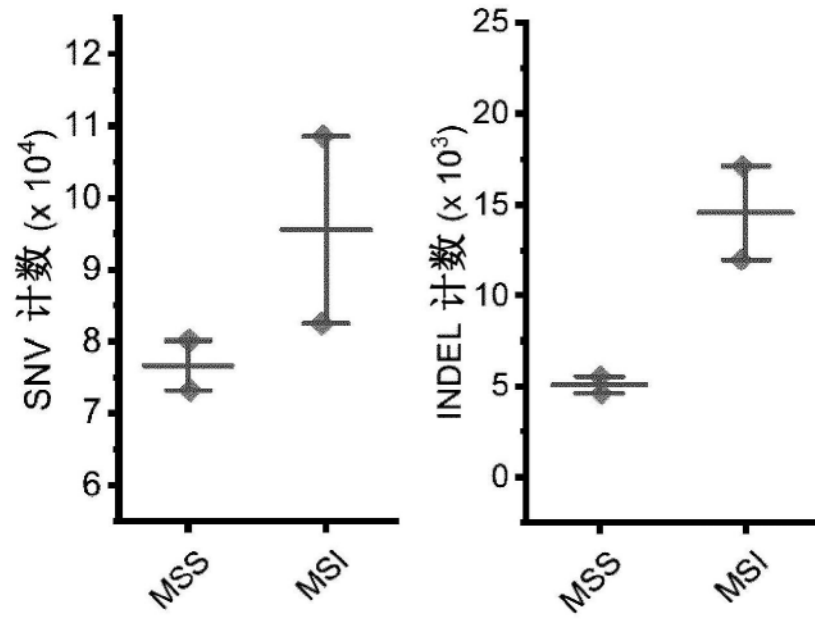


图8C

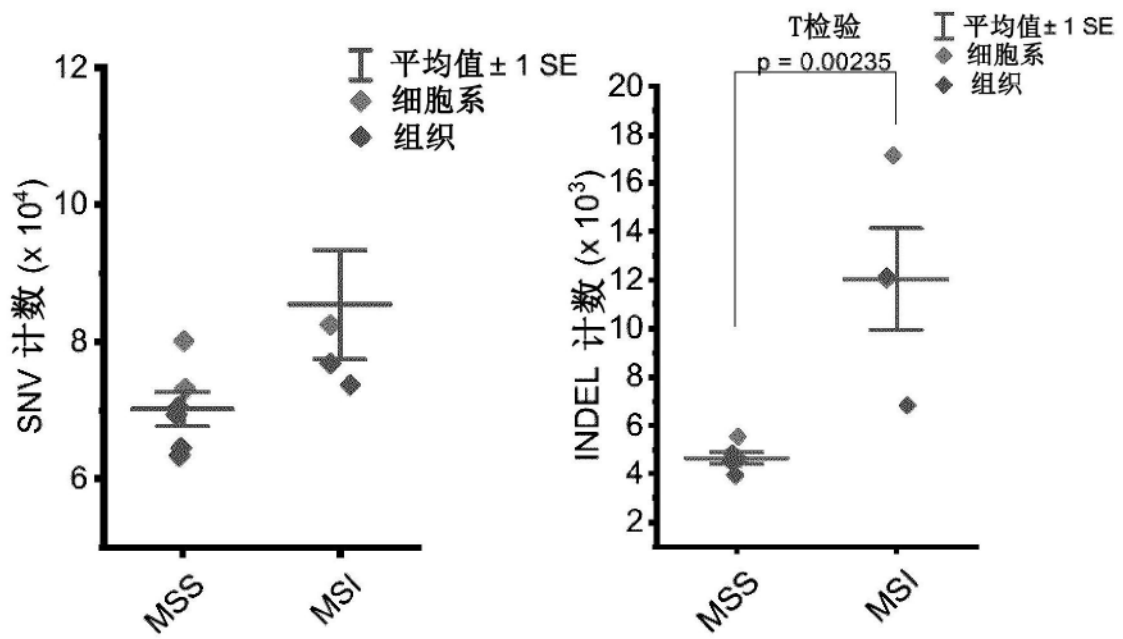
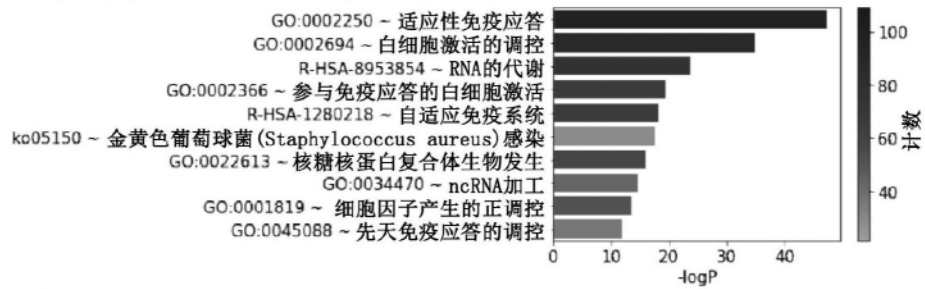


图8D

MSI上调基因(855个基因, 828个Metascape ID)



MSI下调基因(54个基因, 53个Metascape ID)

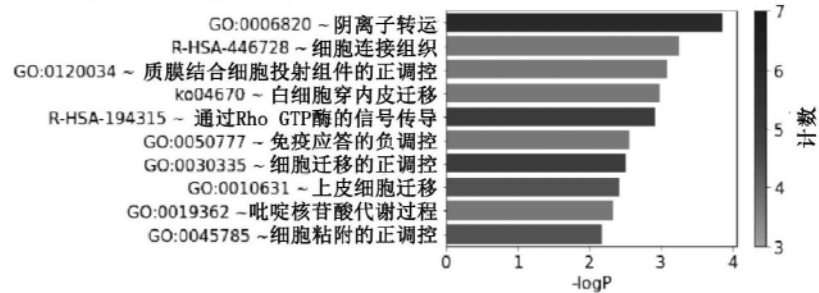
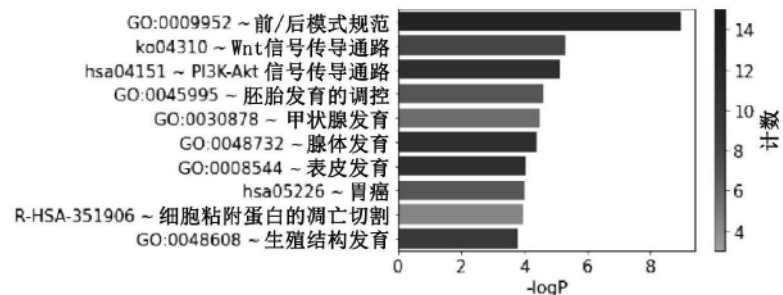


图9A

MSS上调基因(70个基因, 60个Metascape ID)



MSS下调基因(160个基因, 156个Metascape ID)

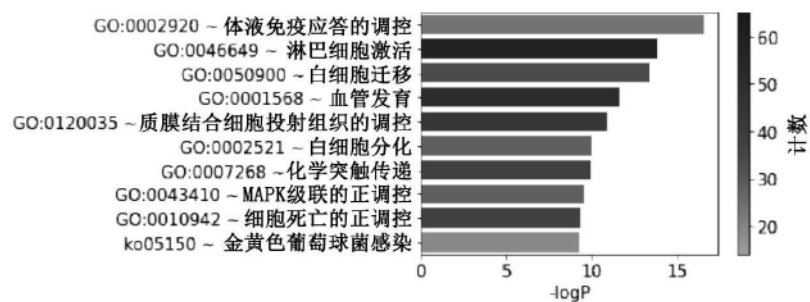


图9B

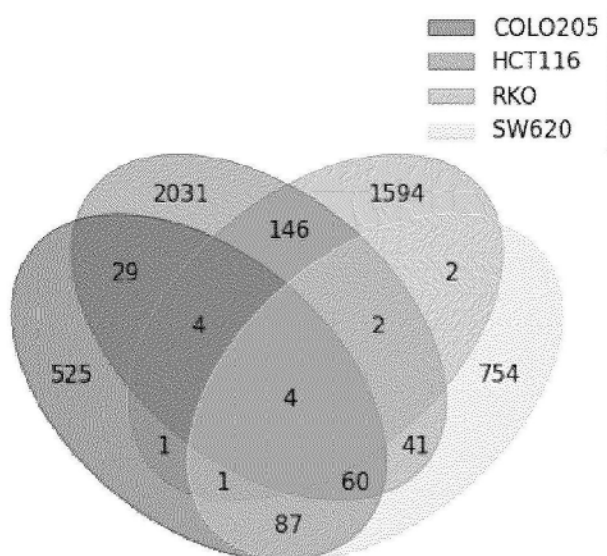


图10A

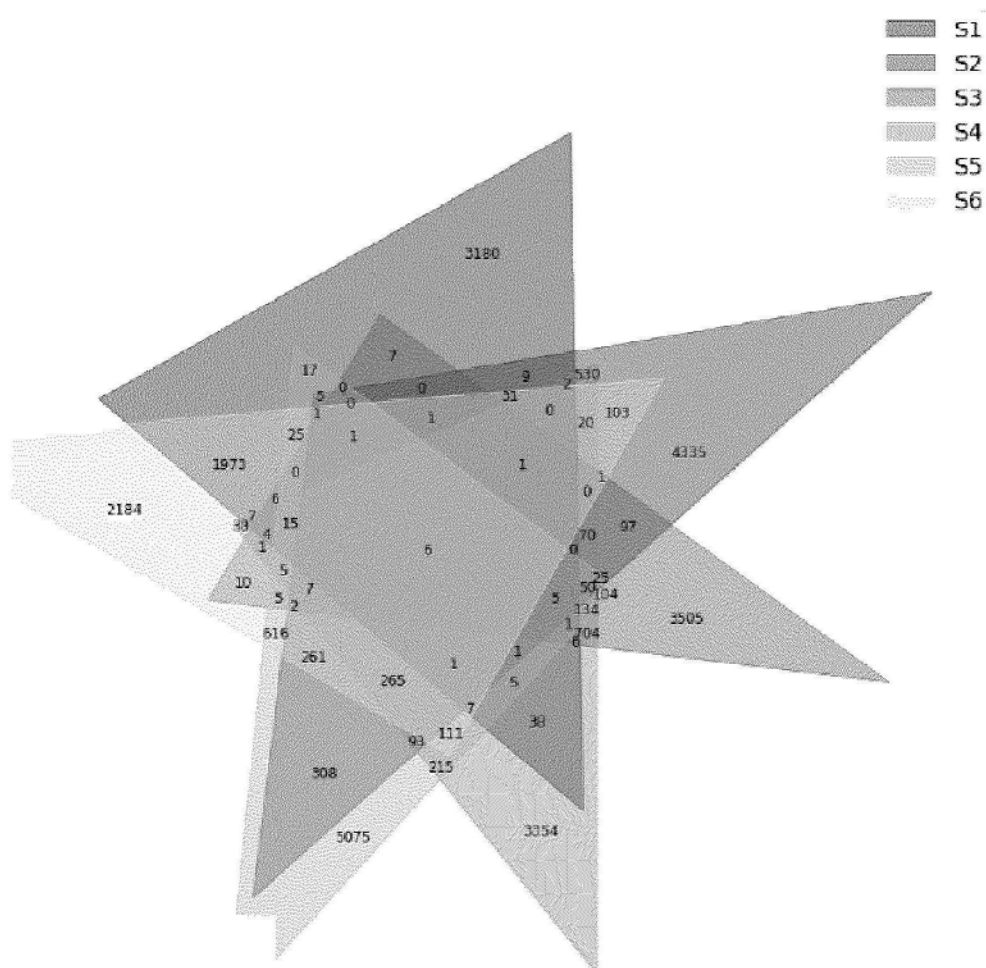


图10B

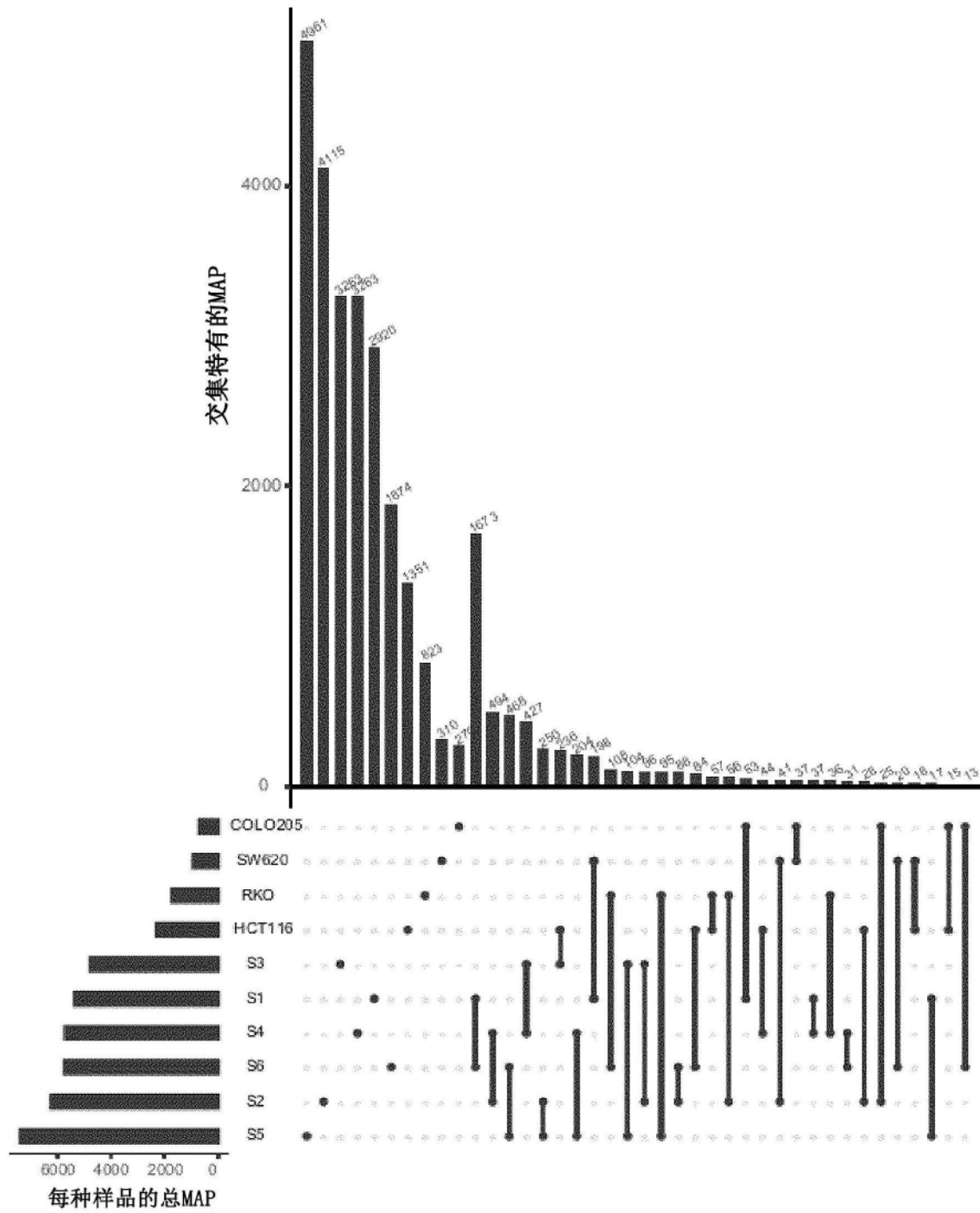


图10C

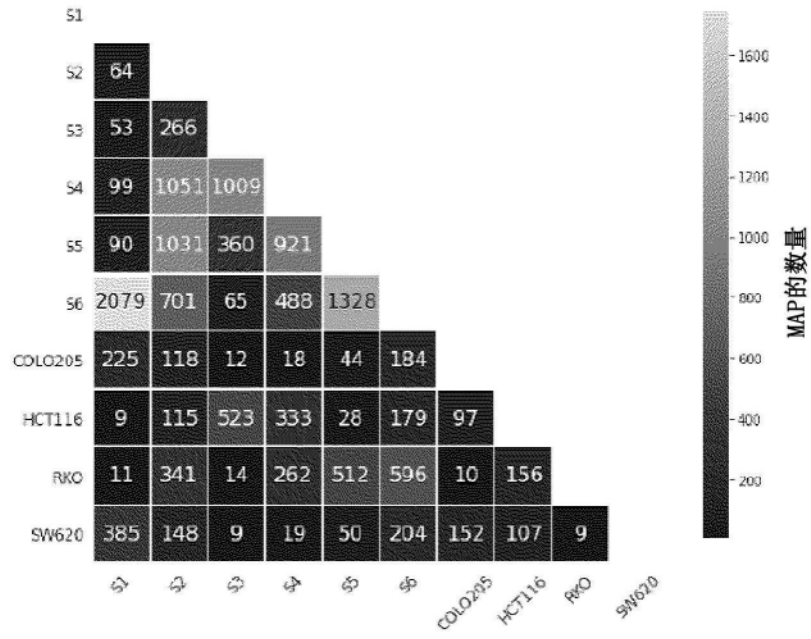


图10D

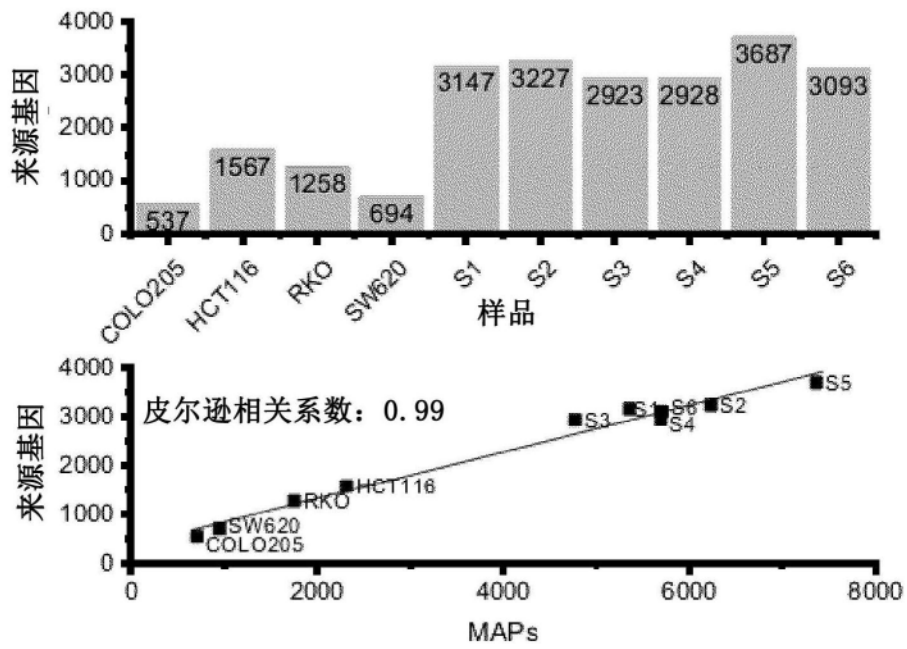


图11A

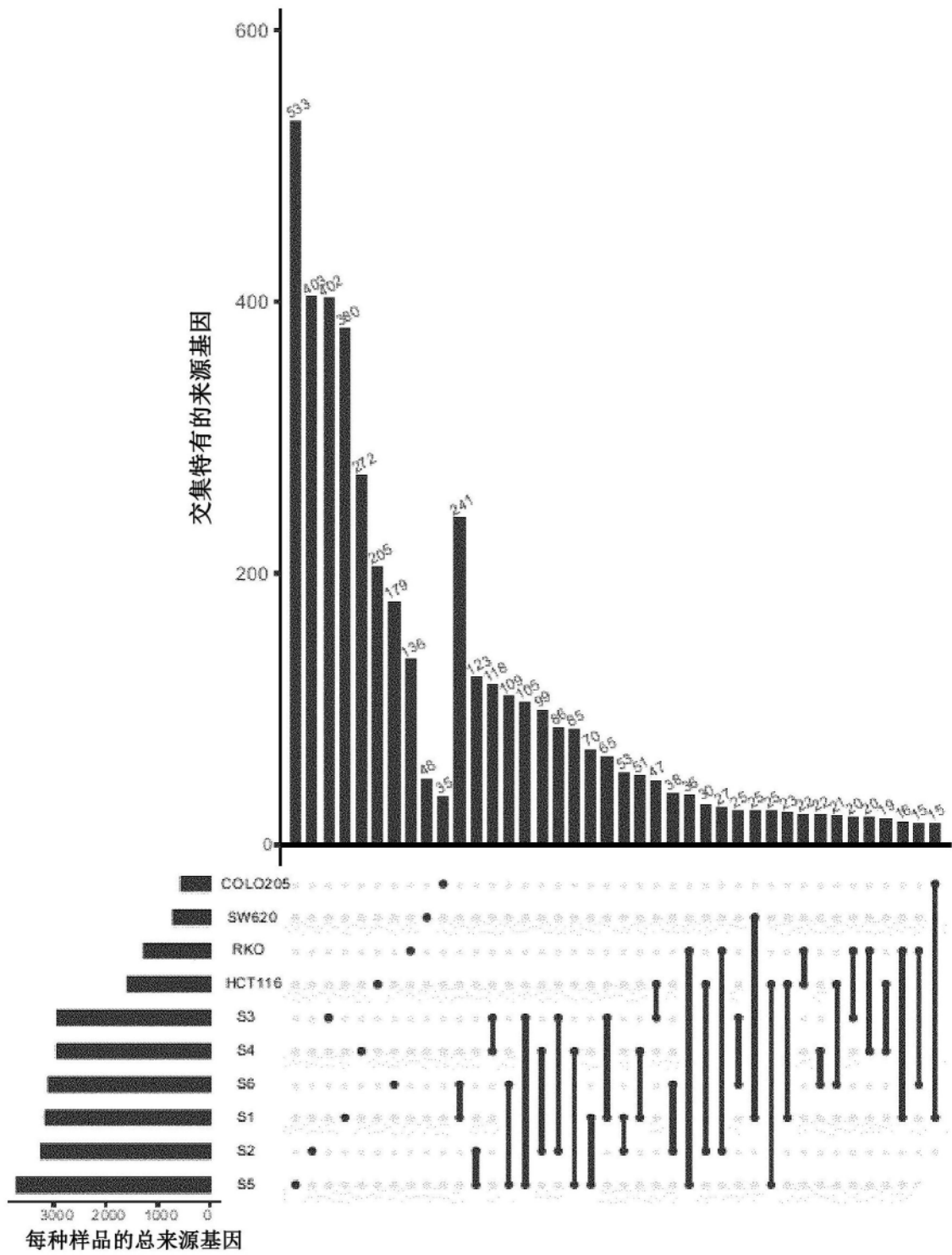


图11B

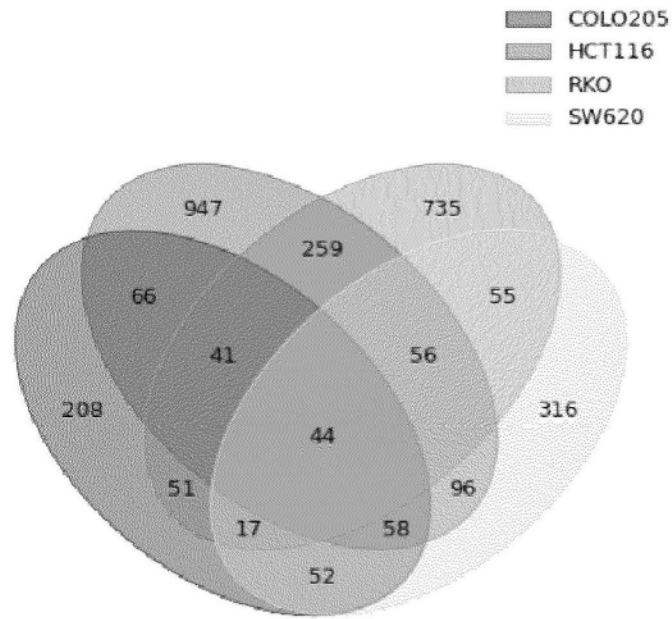


图11C

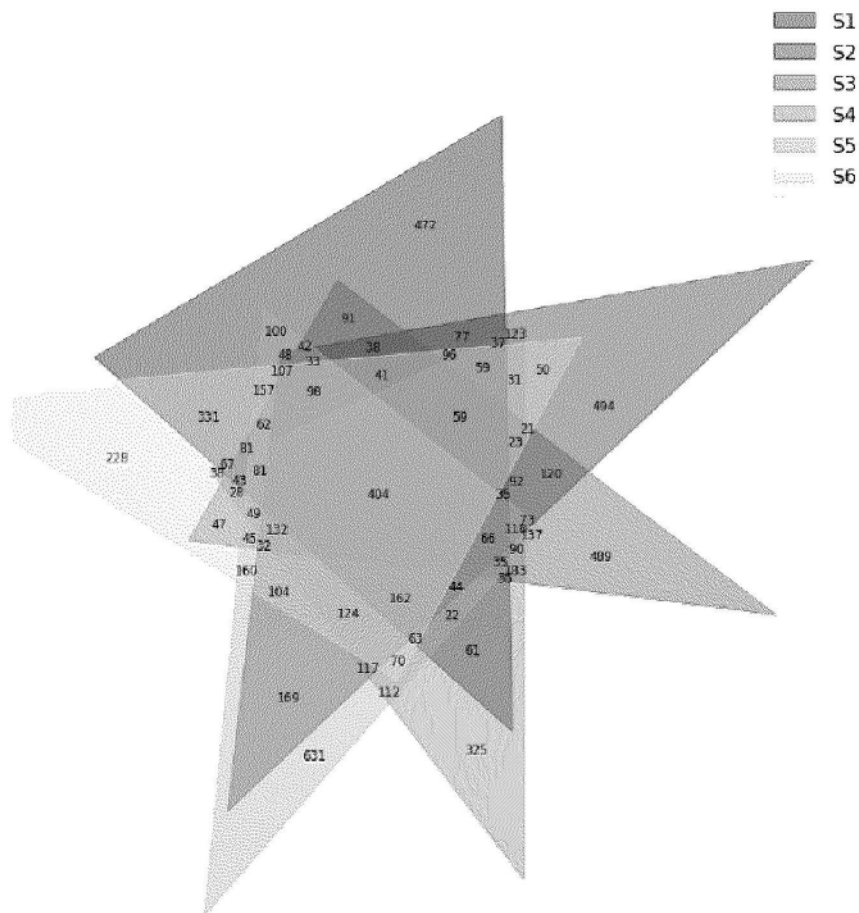


图11D

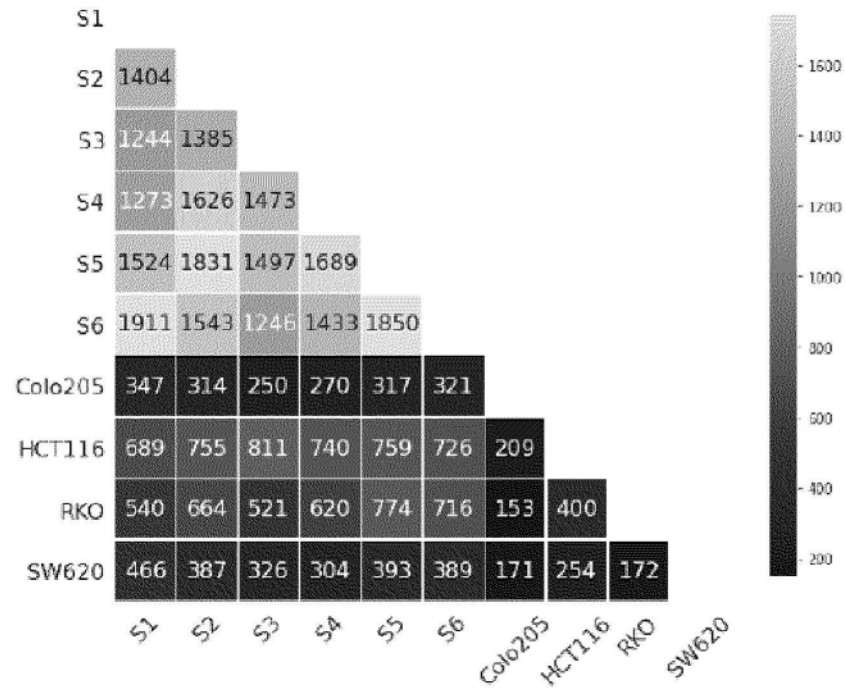


图11E

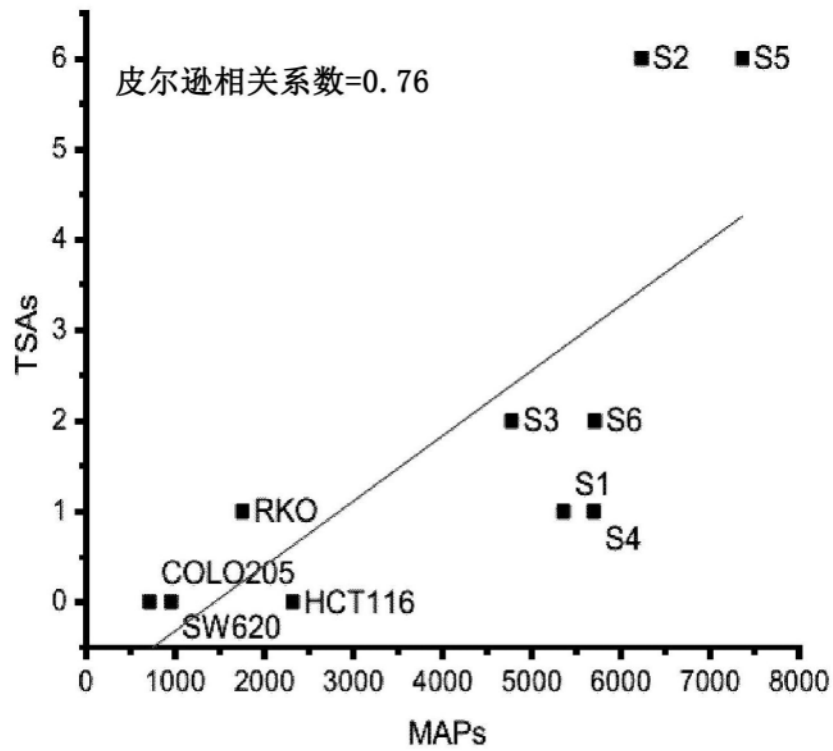


图12

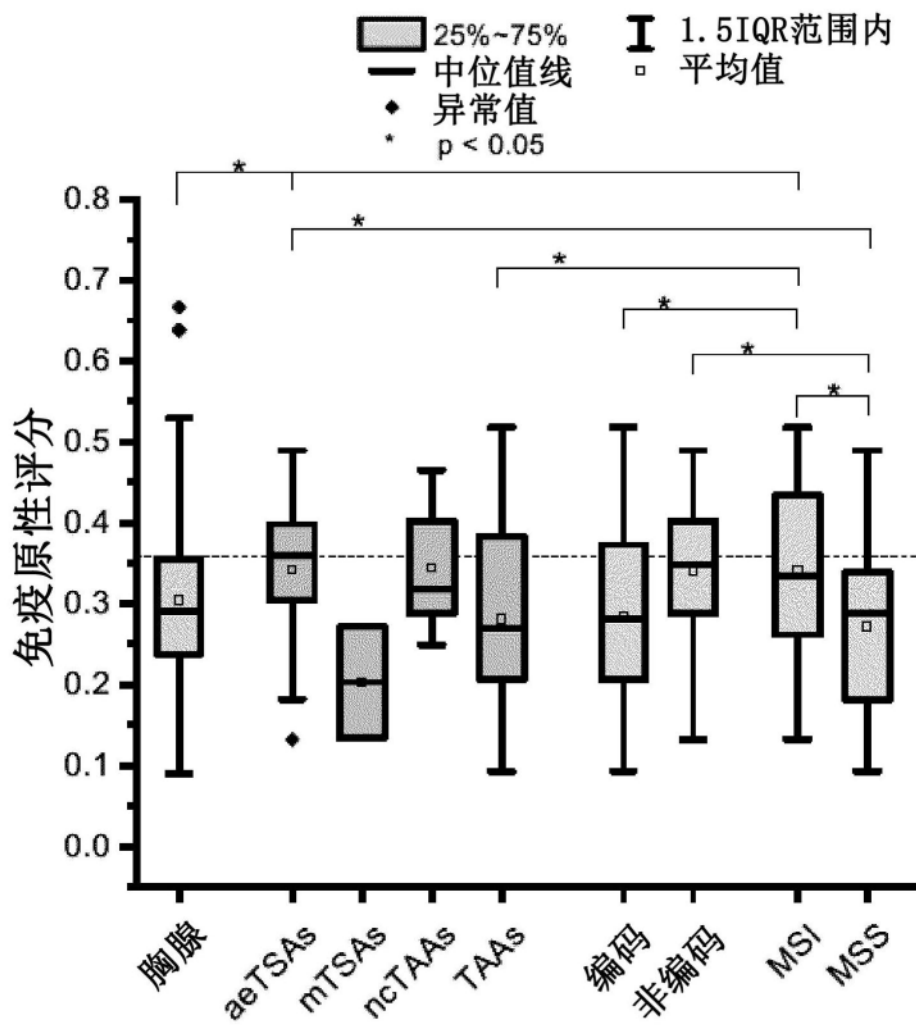


图13