



(12) 发明专利申请

(10) 申请公布号 CN 115397842 A

(43) 申请公布日 2022. 11. 25

(21) 申请号 202180028396.9

(22) 申请日 2021.03.15

(30) 优先权数据

63/009,853 2020.04.14 US

(85) PCT国际申请进入国家阶段日

2022.10.13

(86) PCT国际申请的申请数据

PCT/CA2021/050340 2021.03.15

(87) PCT国际申请的公布数据

W02021/207823 EN 2021.10.21

(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

塞巴斯蒂安·勒米厄

格雷戈里·厄克斯

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 刘明

(51) Int.Cl.

C07K 14/47 (2006.01)

A61K 39/00 (2006.01)

A61K 9/127 (2006.01)

A61P 35/00 (2006.01)

A61P 37/04 (2006.01)

C07K 14/725 (2006.01)

C07K 14/74 (2006.01)

C07K 7/06 (2006.01)

C07K 7/08 (2006.01)

C12N 15/12 (2006.01)

C12N 15/87 (2006.01)

C12N 5/071 (2006.01)

C12N 5/10 (2006.01)

权利要求书8页 说明书90页

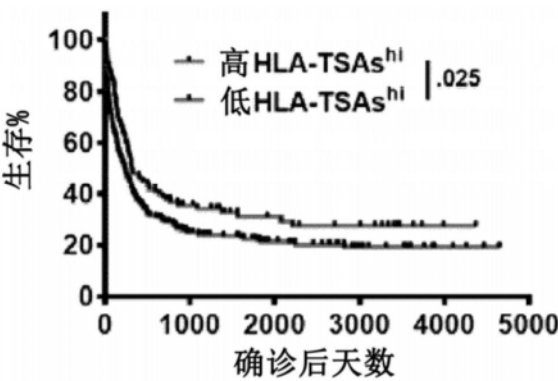
序列表40页 附图34页

(54) 发明名称

用于急性髓系白血病 (AML) 的新型肿瘤特异性抗原及其用途

(57) 摘要

急性髓系白血病 (AML) 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。本文描述了大比例AML细胞共享的新型肿瘤特异性抗原 (TSA)。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。描述了源自这些TSA的核酸、组合物、细胞和疫苗。还描述了TSA、核酸、组合物、细胞和疫苗用于治疗白血病 (诸如AML) 的用途。



1. 一种白血病肿瘤抗原肽 (TAP), 包含以下氨基酸序列中的一个:

序列 (SEQ ID NO)	序列 (SEQ ID NO)	序列 (SEQ ID NO)	序列 (SEQ ID NO.)
RQISVQASL (1)	SIQRNLSL (49)	AQDIILQAV (97)	RYLANKIHI (145)
DRELRNLEL (2)	NVSSHVHTV (50)	PPRPLGAQV (98)	SLLSGLLRA (146)
GARQQIHSW (3)	ALASHLIEA (51)	FNVALNARY (99)	SRIHLVVL (147)
SGKLRVAL (4)	ALDDITIQL (52)	GPGSRESTL (100)	SSSPVRGPSV (148)
RSASSATQVHK (5)	ALGNTVPAV (53)	IPHQRSSL (101)	STFSLYLKK (149)
SASSATQVHK (6)	ALLPAVPSL (54)	LTDRIYLT (102)	SLDLLPLSI (150)
FLLEFKPVS (7)	APAPPPVAV (55)	NLKEKKALF (103)	VTDLLALTV (151)
GPQVRGSI (8)	APDKKITL (56)	VLFGGKVSGA (104)	RTQITKVSLKK (152)
IRMKAQAL (9)	AQMNLLQKY (57)	VVFPFPVVK (105)	ILRSPLKW (153)
KIKVFSKVY (10)	DQVIRLAGL (58)	SLLIIPKKK (106)	LSTGHLSTV (154)
LLSRGLLFRI (11)	ETTSQVRKY (59)	APGAAGQRL (107)	TVEEYLVNI (155)
LPIASASLL (12)	GGSLIHPQW (60)	KLQDKEIGL (108)	QIKTKLLGSL (156)
LYFLGHGSI (13)	GLYYKLHNV (61)	SLREPQPAL (109)	LPSFSHLLL (157)
NPLQLSLSI (14)	GQKPVILTY (62)	TPGRSTQAI (110)	CLRIGPVT (158)
DLMLRESL (15)	GSLDFQRGW (63)	APRGTAAL (111)	HVSDGSTALK (159)
VTFKLSLF (16)	HHLVETLKF (64)	IASPIALL (112)	IAYSVRALR (160)
IALYKQVL (17)	HLLSETPQL (65)	ILFQNSALK (113)	PRGFLSAL (161)
IVATGSLLK (18)	HQLYRASAL (66)	ILKKNISI (114)	ISSWLISSL (162)

KIKNKTKNK (19)	HTDDIENAKY (67)	IPLAVRTI (115)	IPLNPFSSL (163)
KLLSLTIYK (20)	IAAPILHV (68)	LPRNKPLL (116)	LSDRQLSL (164)
NILKKTVL (21)	KAFPFHIIF (69)	PAPHPAAL (117)	LSHPAPSSL (165)
NPKLKDIL (22)	KATEYVHSL (70)	SPVVRVGL (118)	LRKAVDPIIL (166)
NQKKVRIL (23)	KFSNVTMLF (71)	TLNQGINVYI (119)	ILLEEQLSI (167)
PFPLVQVEPV (24)	KLLEKAFSI (72)	RPRGPRTAP (120)	LTSISIRPV (168)
SPQSGPAL (25)	KPMPTKVVF (73)	SVQLLEQAIHK (121)	TISECPLLI (169)
TSRLPKIQK (26)	NVNRPLTMK (74)	RTPKNYQHW (122)	TLKLKKIFF (170)
LLDNILQSI (27)	REPYELTVPAL (75)	ALPVALPSL (123)	ILLSNFSSL (171)
RLEVRKVIL (28)	SEAEAAKNAL (76)	SLQILVSSL (124)	LGGAWKAVF (172)
LSWGYFLFK (29)	SLWGQPAEA (77)	ISNKVPKLF (125)	LSASHLSSL (173)
TILPRILTL (30)	SPADHRGYASL (78)	TVIRIAIVNK (126)	AGDIIRLI (174)
EGKIKRNI (31)	SPQSAAAEI (79)	KEIFLELRL (127)	DRGILRNLL (175)
FLASFVEKTVL (32)	SPVVHQSL (80)	TLRSPGSSL (128)	GLRLIHVSL (176)
ILASHNLTV (33)	SPYRTPVL (81)	TVRGDVSSL (129)	GLRLLHVSL (177)
IQLTSVHLL (34)	SVFAGVVG (82)	ALDPLLLRI (130)	LHNEKGLSL (178)
LELISFLPVL (35)	SYSPAHARL (83)	ISLIVTGLK (131)	LPSFSRPSGII (179)
NFCMLHQSI (36)	THGSEQLHL (84)	KILDVNLRI (132)	LSSRLPLGK (180)
PARPAGPL (37)	TQAPPNVVL (85)	ERVYIRASL (133)	MIGIKRLL (181)
PLPIVPAL (38)	VLVPYEPPQV (86)	ILDLESRY (134)	NLKKREIL (182)
SNLIRTGSH (39)	VSFPDVRKV (87)	KTFVQQKTL (135)	RMVAYLQQL (183)

VPAPAQAI (40)	VVFDKSDLAKY (88)	LYIKSLPAL (136)	SPARALPSL (184)
KGHGGPRSW (41)	YSHHSGLEY (89)	VLKEKNASL (137)	TVPGIQRY (185)
ITSSAVTTALK (42)	YYLDWIHHY (90)	LGISLTLKY (138)	VSRNYVLLI (186)
LLLPESPSI (43)	SVYKYLKAK (91)	DLLPKKLL (139)	LTVPLSVFW (187)
VILIPLPPK (44)	IYQFIMDRF (92)	HSLISIVYL (140)	KLNQAFVL (188)
AVLLPKPPK (45)	GTLQGIRAW (93)	IAGALRSVL (141)	RLVSSTLLQK (189)
TQVSMAESI (46)	AQKVSQGAA (94)	IGNPILRVL (142)	LPSHSLLI (190)
LNHLRTSI (47)	LYPSKLTHF (95)	IYAPHIRLS (143)	
NTSHLPLIY (48)	ATQNTIIGK (96)	LRSQILSY (144)	

2. 如权利要求1所述的白血病TAP,所述白血病TAP包含SEQ ID NO:97-154中列出的所述的氨基酸序列中的一个。

3. 如权利要求1或2所述的白血病TAP,其中所述白血病TAP与HLA-A*01:01分子结合并且包含氨基酸序列NTSHLPLIY (SEQ ID NO:48)、HTDDIENAKY (SEQ ID NO:67)、YSHHSGLEY (SEQ ID NO:89)、ILDLESRY (SEQ ID NO:134)、VTDLLALTV (SEQ ID NO:151)或LSDRQLSL (SEQ ID NO:164),优选ILDLESRY (SEQ ID NO:134)或VTDLLALTV (SEQ ID NO:151)。

4. 如权利要求1或2所述的白血病TAP,其中所述白血病TAP与HLA-A*02:01分子结合并且包含氨基酸序列FLLEFKPVS (SEQ ID NO:7)、LLSRGLLFRI (SEQ ID NO:11)、LLDNILQSI (SEQ ID NO:27)、FLASFVEKTVL (SEQ ID NO:32)、ILASHNLTV (SEQ ID NO:33)、IQLTSVHLL (SEQ ID NO:34)、LELISFLPVL (SEQ ID NO:35)、LLLPESPSI (SEQ ID NO:43)、ALASHLIEA (SEQ ID NO:51)、ALDDITIQL (SEQ ID NO:52)、ALGNTVPAV (SEQ ID NO:53)、ALLPAVPSL (SEQ ID NO:54)、GLYYKLHNV (SEQ ID NO:61)、HLLSETPQL (SEQ ID NO:65)、KLEKAFSI (SEQ ID NO:72)、SLWGQPAEA (SEQ ID NO:77)、SVFAGVGV (SEQ ID NO:82)、VLVPYEPQV (SEQ ID NO:86)、VLFGGKVSGA (SEQ ID NO:104)、KLQDKEIGL (SEQ ID NO:108)、TLNQGINVYI (SEQ ID NO:119)、ALPVALPSL (SEQ ID NO:123)、ALDPLLLRI (SEQ ID NO:130)、KILDVNLRI (SEQ ID NO:132)、SLLSGLLRA (SEQ ID NO:146)、SLDLLPLSI (SEQ ID NO:150)、ILLEEQSLI (SEQ ID NO:167)、LTSISIRPV (SEQ ID NO:168)、TISECPLLI (SEQ ID NO:169)、ILLSNFSSL (SEQ ID NO:171)、RMVAYLQQL (SEQ ID NO:183)或KLNQAFVL (SEQ ID NO:188),优选VLFGGKVSGA (SEQ ID NO:104)、KLQDKEIGL (SEQ ID NO:108)、TLNQGINVYI (SEQ ID NO:119)、ALPVALPSL (SEQ ID NO:123)、ALDPLLLRI (SEQ ID NO:130)、KILDVNLRI (SEQ ID NO:132)、SLLSGLLRA (SEQ ID NO:146)或SLDLLPLSI (SEQ ID NO:150)。

5. 如权利要求1或2所述的白血病TAP,其中所述白血病TAP与HLA-A*03:01分子结合并且包含氨基酸序列RSASSATQVHK (SEQ ID NO:5)、IVATGSLLK (SEQ ID NO:18)、KIKNKTKNK (SEQ ID NO:19)、KLLSLTIYK (SEQ ID NO:20)、ITSSAVTTALK (SEQ ID NO:42)、VILIPLPPK (SEQ ID NO:44)、NVNRPLTMK (SEQ ID NO:74)、SVYKYLKAK (SEQ ID NO:91)、VVFPPVNVK (SEQ ID NO:105)、ILFQNSALK (SEQ ID NO:113)、TVIRIAIVNK (SEQ ID NO:126)、ISLIVTGLK (SEQ

ID NO:131)、HVSDGSTALK (SEQ ID NO:159)、IAYSVRALR (SEQ ID NO:160)、LSSRLPLGK (SEQ ID NO:180) 或RLVSSTLLQK (SEQ ID NO:189), 优选VVFPPFPVNK (SEQ ID NO:105)、ILFQNSALK (SEQ ID NO:113)、TVIRIAIVNK (SEQ ID NO:126) 或ISLIVTGLK (SEQ ID NO:131)。

6. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*11:01分子结合并且包含氨基酸序列SASSATQVHK (SEQ ID NO:6)、AVLLPKPPK (SEQ ID NO:45)、ATQNTIIGK (SEQ ID NO:96)、SLLIIPKKK (SEQ ID NO:106)、SVQLLEQAIHK (SEQ ID NO:121)、STFSLYLKK (SEQ ID NO:149) 或RTQITKVSLKK (SEQ ID NO:152), 优选SLLIIPKKK (SEQ ID NO:106)、SVQLLEQAIHK (SEQ ID NO:121)、STFSLYLKK (SEQ ID NO:149) 或RTQITKVSLKK (SEQ ID NO:152)。

7. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*24:02分子结合并且包含氨基酸序列LYFLGHGSI (SEQ ID NO:13)、NFCMLHQSI (SEQ ID NO:36)、KFSNVTMLF (SEQ ID NO:71)、IYQFIMDRF (SEQ ID NO:92)、LYPSKLTHF (SEQ ID NO:95) 或RYLANKIHI (SEQ ID NO:145), 优选RYLANKIHI (SEQ ID NO:145)。

8. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*26:01分子结合并且包含氨基酸序列ETTSQVRKY (SEQ ID NO:59) 或TVPGIQRY (SEQ ID NO:185)。

9. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*29:02分子结合并且包含以下氨基酸序列之一: VVFDKSDLAKY (SEQ ID NO:88)、FNVALNARY (SEQ ID NO:99) 或LGISLTLKY (SEQ ID NO:138), 优选FNVALNARY (SEQ ID NO:99) 或LGISLTLKY (SEQ ID NO:138)。

10. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*30:01分子结合并且包含氨基酸序列TSRLPKIQK (SEQ ID NO:26)、LSWGYFLFK (SEQ ID NO:29) 或LSHPAPSSL (SEQ ID NO:165)。

11. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*68:02分子结合并且包含氨基酸序列NVSSHVHTV (SEQ ID NO:50) 或SSSPVRGPSV (SEQ ID NO:148), 优选SSSPVRGPSV (SEQ ID NO:148)。

12. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*07:02分子结合并且包含氨基酸序列GPQVRGSI (SEQ ID NO:8)、SPQSGPAL (SEQ ID NO:25)、VPAPAQAI (SEQ ID NO:40)、APAPPPVAV (SEQ ID NO:55)、APDKKITL (SEQ ID NO:56)、KPMPTKVVF (SEQ ID NO:73)、SPADHRGYASL (SEQ ID NO:78)、SPQSAAAEI (SEQ ID NO:79)、SPVVHQSL (SEQ ID NO:80)、SPYRTPVL (SEQ ID NO:81)、PPRPLGAQV (SEQ ID NO:98)、GPGSRESTL (SEQ ID NO:100)、APGAAGQRL (SEQ ID NO:107)、TPGRSTQAI (SEQ ID NO:110)、APRGTAAL (SEQ ID NO:111)、SPVVRVGL (SEQ ID NO:118)、RPRGPRTAP (SEQ ID NO:120)、TLRSPGSSL (SEQ ID NO:128)、TVRGDVSSL (SEQ ID NO:129)、LPSFSHFLLL (SEQ ID NO:157)、PRGFLSAL (SEQ ID NO:161)、IPLNPFSSL (SEQ ID NO:163)、LPSFSRPSGII (SEQ ID NO:179) 或SPARALPSL (SEQ ID NO:184), 优选PPRPLGAQV (SEQ ID NO:98)、GPGSRESTL (SEQ ID NO:100)、APGAAGQRL (SEQ ID NO:107)、TPGRSTQAI (SEQ ID NO:110)、APRGTAAL (SEQ ID NO:111)、SPVVRVGL (SEQ ID NO:118)、RPRGPRTAP (SEQ ID NO:120)、TLRSPGSSL (SEQ ID NO:128) 或TVRGDVSSL (SEQ ID NO:129)。

13. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*08:01分子结合并

且包含氨基酸序列SGKLRVAL (SEQ ID NO:4)、NPLQLSLSI (SEQ ID NO:14)、DLMLRESL (SEQ ID NO:15)、IALYKQVL (SEQ ID NO:17)、NILKKTVL (SEQ ID NO:21)、NPKLKDIL (SEQ ID NO:22)、NQKKVRIL (SEQ ID NO:23)、RLEVRKVIL (SEQ ID NO:28)、EGKIKRNI (SEQ ID NO:31)、LNHLRTSI (SEQ ID NO:47)、SIQRNLSL (SEQ ID NO:49)、IPHQRSSL (SEQ ID NO:101)、NLKEKKALF (SEQ ID NO:103)、ILKKNISI (SEQ ID NO:114)、VLKEKNASL (SEQ ID NO:137)、DLLPKKLL (SEQ ID NO:139)、SRIHLVVL (SEQ ID NO:147)、QIKTKLLGSL (SEQ ID NO:156)、TLKLKKIFF (SEQ ID NO:170)、MIGIKRLL (SEQ ID NO:181) 或NLKKREIL (SEQ ID NO:182), 优选IPHQRSSL (SEQ ID NO:101)、NLKEKKALF (SEQ ID NO:103)、ILKKNISI (SEQ ID NO:114)、VLKEKNASL (SEQ ID NO:137)、DLLPKKLL (SEQ ID NO:139) 或SRIHLVVL (SEQ ID NO:147)。

14. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*14:01分子结合并且包含氨基酸序列DRELRLNEL (SEQ ID NO:2)、SNLIRTGSH (SEQ ID NO:39)、DQVIRLAGL (SEQ ID NO:58)、HQLYRASAL (SEQ ID NO:66)、SLQILVSSL (SEQ ID NO:124)、ERVYIRASL (SEQ ID NO:133)、LYIKSLPAL (SEQ ID NO:136)、IAGALRSVL (SEQ ID NO:141)、ISSWLISSL (SEQ ID NO:162)、DRGILRNLL (SEQ ID NO:175)、GLRLIHVSL (SEQ ID NO:176) 或GLRLLHVSL (SEQ ID NO:177), 优选SLQILVSSL (SEQ ID NO:124)、ERVYIRASL (SEQ ID NO:133)、LYIKSLPAL (SEQ ID NO:136) 或IAGALRSVL (SEQ ID NO:141)。

15. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*15:01分子结合并且包含氨基酸序列KIKVFSKVY (SEQ ID NO:10)、AQMNLLQKY (SEQ ID NO:57)、GQKPVILTY (SEQ ID NO:62) 或AQKVSQGAA (SEQ ID NO:94)。

16. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*27:05分子结合并且包含氨基酸序列RQISVQASL (SEQ ID NO:1) 或LRSQILSY (SEQ ID NO:144), 优选LRSQILSY (SEQ ID NO:144)。

17. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*38:01分子结合并且包含氨基酸序列TQVSMAESI (SEQ ID NO:46)、HHLVETLKF (SEQ ID NO:64) 或THGSEQLHL (SEQ ID NO:84)。

18. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*40:01分子结合并且包含氨基酸序列REPYELTVPAL (SEQ ID NO:75) 或SEAEAAKNAL (SEQ ID NO:76)。

19. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*44:03分子结合并且包含氨基酸序列KEIFLELRL (SEQ ID NO:127)。

20. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*51:01分子结合并且包含氨基酸序列LPIASASLL (SEQ ID NO:12)、PFPLVQVEPV (SEQ ID NO:24)、PLPIVPAL (SEQ ID NO:38)、IAAPILHV (SEQ ID NO:68)、IPLAVRTI (SEQ ID NO:115)、LPRNKPLL (SEQ ID NO:116) 或LPSHSLLI (SEQ ID NO:190), 优选IPLAVRTI (SEQ ID NO:115) 或LPRNKPLL (SEQ ID NO:116)。

21. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*57:01分子结合并且包含氨基酸序列GARQQIHSW (SEQ ID NO:3)、VTFKLSLF (SEQ ID NO:16)、KGHGGPRSW (SEQ ID NO:41)、GSLDFQRGW (SEQ ID NO:63)、KAFPFIHIF (SEQ ID NO:69)、GTLQGIRAW (SEQ ID NO:93)、RTPKNYQHW (SEQ ID NO:122)、ISNKVPKLF (SEQ ID NO:125)、KTFVQQKTL (SEQ ID NO:135)、ILRSPLKW (SEQ ID NO:153) 或LTVPLSVFW (SEQ ID NO:183), 优选RTPKNYQHW (SEQ ID

NO:122)、ISNKVPKLF (SEQ ID NO:125)、KTFVQQKTL (SEQ ID NO:135) 或ILRSPLKW (SEQ ID NO:153)。

22. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*57:03分子结合并且包含氨基酸序列GGSLIHPQW (SEQ ID NO:60) 或LGGAWKAVF (SEQ ID NO:172)。

23. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-C*03:03分子结合并且包含氨基酸序列PARPAGPL (SEQ ID NO:37)、IASPIALL (SEQ ID NO:112) 或HSLISIVYL (SEQ ID NO:140), 优选IASPIALL (SEQ ID NO:112) 或HSLISIVYL (SEQ ID NO:140)。

24. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-C*05:01分子结合并且包含氨基酸序列SLDLLPLSI (SEQ ID NO:150)。

25. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-C*06:02分子结合并且包含氨基酸序列IRMKAQAL (SEQ ID NO:9)、KATEYVHSL (SEQ ID NO:70)、VSFPDVRKV (SEQ ID NO:87)、IGNPILRVL (SEQ ID NO:142)、LSTGHLSTV (SEQ ID NO:154) 或LRKAVDPIIL (SEQ ID NO:166), 优选IGNPILRVL (SEQ ID NO:142) 或LSTGHLSTV (SEQ ID NO:154)。

26. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-C*07:01分子结合并且包含氨基酸序列IGNPILRVL (SEQ ID NO:142)、IYAPHIRLS (SEQ ID NO:143)、TVEEYLVNI (SEQ ID NO:155)、LHNEKGLSL (SEQ ID NO:178) 或VSRNYVLLI (SEQ ID NO:186), 优选IGNPILRVL (SEQ ID NO:142) 或IYAPHIRLS (SEQ ID NO:143)。

27. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-C*07:02分子结合并且包含氨基酸序列TILPRILTL (SEQ ID NO:30)、SYSPAHL (SEQ ID NO:83)、TQAPPNVVL (SEQ ID NO:85)、YYLDWIHHY (SEQ ID NO:90)、SLREPQPAL (SEQ ID NO:109)、PAPPHPAAL (SEQ ID NO:117) 或CLRIGPVTL (SEQ ID NO:158), 优选SLREPQPAL (SEQ ID NO:109) 或PAPPHPAAL (SEQ ID NO:117)。

28. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-C*08:02分子结合并且包含氨基酸序列AQDIILQAV (SEQ ID NO:97)、LTDRIYLT (SEQ ID NO:102) 或AGDIILQAV (SEQ ID NO:174), 优选AQDIILQAV (SEQ ID NO:97) 或LTDRIYLT (SEQ ID NO:102)。

29. 如权利要求1或2所述的白血病TAP, 其中所述白血病TAP与HLA-C*12:03分子结合并且包含氨基酸序列LSASHLSSL (SEQ ID NO:173)。

30. 如权利要求1-29中任一项所述的白血病TAP, 其由位于基因组的非蛋白质编码区的序列编码。

31. 如权利要求30所述的白血病TAP, 其中所述基因组的非蛋白质编码区是非翻译转录区 (UTR)。

32. 如权利要求30所述的白血病TAP, 其中所述基因组的非蛋白质编码区是内含子。

33. 如权利要求30所述的白血病TAP, 其中所述基因组的非蛋白质编码区是基因间区。

34. 一种包含权利要求1-33中任一项所限定的白血病TAP中的至少两种的组合。

35. 一种编码权利要求1-33中任一项所述的白血病TAP或权利要求34所述的组合的核酸。

36. 如权利要求35所述的核酸, 所述核酸是mRNA或病毒载体。

37. 一种脂质体, 所述脂质体包含权利要求1-33中任一项所述的白血病TAP、权利要求34所述的组合或者权利要求35或36所述的核酸。

38. 一种组合物,所述组合物包含权利要求1-33中任一项所述的白血病TAP、权利要求34所述的组合、权利要求35或36所述的核酸或者权利要求37所述的脂质体,以及药学上可接受的载体。

39. 一种疫苗,所述疫苗包含权利要求1-33中任一项所述的白血病TAP、权利要求34所述的组合、权利要求35或36所述的核酸、权利要求37所述的脂质体或者权利要求38所述的组合物,以及佐剂。

40. 一种分离的主要组织相容性复合体(MHC) I类分子,所述分离的MHC I类分子在其肽结合槽中包含权利要求1-33中任一项所述的白血病TAP。

41. 如权利要求40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。

42. 如权利要求41的分离的MHC I类分子,其中所述多聚体是四聚体。

43. 一种分离的细胞,所述分离的细胞包含(i) 权利要求1-33中任一项所述的白血病TAP、(ii) 权利要求34所述的组合或(iii) 包含编码权利要求1-33任一项的TAP的或权利要求34所述的组合的核苷酸序列的载体。

44. 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MHC I类分子在其肽结合槽中包含权利要求1-33中任一项所述的白血病TAP或权利要求34所述的组合。

45. 如权利要求44所述的细胞,所述细胞是抗原呈递细胞(APC)。

46. 如权利要求45所述的细胞,其中所述APC是树突状细胞。

47. 一种T细胞受体(TCR),所述TCR特异性识别权利要求40-42中任一项所述的分离的MHC I类分子和/或在权利要求44-46中任一项所述的细胞的表面表达的MHC I类分子。

48. 如权利要求47所述的TCR,其中所述TCR包含TCRB (TCRB) 链,所述TCRB链包含互补决定区3(CDR3),所述CDR3包含SEQ ID NO:191-219中列出的氨基酸序列中的一个。

49. 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47或48所述的TCR。

50. 如权利要求49所述的分离的细胞,所述分离的细胞是CD8⁺T淋巴细胞。

51. 一种细胞群,所述细胞群包含至少0.5%的如权利要求49或50所限定的分离的细胞。

52. 一种治疗受试者的白血病的方法,所述方法包括向所述受试者施用有效量的:(i) 权利要求1-33中任一项所述的白血病TAP;(ii) 权利要求34所述的组合;(iii) 权利要求35或36所述的核酸;(iv) 权利要求37所述的脂质体;(v) 权利要求38所述的组合物;(vi) 权利要求39所述的疫苗;(vii) 权利要求43-46、49和50中任一项所述的细胞;或(viii) 权利要求51所述的细胞群。

53. 如权利要求52所述的方法,其中所述白血病是髓系白血病。

54. 如权利要求53所述的方法,其中所述髓系白血病是急性髓系白血病(AML)。

55. 如权利要求52-54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另外的抗肿瘤剂或疗法。

56. 如权利要求55所述的方法,其中所述至少一种另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、放射疗法或手术。

57. (i) 权利要求1-33中任一项所述的白血病TAP;(ii) 权利要求34所述的组合;(iii) 权利要求35或36所述的核酸;(iv) 权利要求37所述的脂质体;(v) 权利要求38所述的组合

物；(vi) 权利要求39所述的疫苗；(vii) 权利要求43-46、49和50中任一项所述的细胞；或(viii) 权利要求51所述的细胞群用于治疗受试者的白血病的用途。

58. (i) 权利要求1-33中任一项所述的白血病TAP；(ii) 权利要求34所述的组合；(iii) 权利要求35或36所述的核酸；(iv) 权利要求37所述的脂质体；(v) 权利要求38所述的组合物；(vi) 权利要求39所述的疫苗；(vii) 权利要求43-46、49和50中任一项所述的细胞；或(viii) 权利要求51所述的细胞群用于制造用于治疗受试者的白血病的药物的用途。

59. 如权利要求57或58所述的用途，其中所述白血病是髓系白血病。

60. 如权利要求59所述的用途，其中所述髓系白血病是急性髓系白血病(AML)。

61. 如权利要求57-60中任一项所述的用途，进一步包括使用至少一种另外的抗肿瘤剂或疗法。

62. 如权利要求61所述的用途，其中所述至少一种另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、放射疗法或手术。

63. (i) 权利要求1-33中任一项所述的白血病TAP；(ii) 权利要求34所述的组合；(iii) 权利要求35或36所述的核酸；(iv) 权利要求37所述的脂质体；(v) 权利要求38所述的组合物；(vi) 权利要求39所述的疫苗；(vii) 权利要求43-46、49和50中任一项所述的细胞；或(viii) 权利要求51所述的细胞群，用于治疗受试者的白血病的用途。

64. 根据权利要求63所述用于用途的白血病TAP、组合、核酸、脂质体、组合物、疫苗、细胞或细胞群，其中所述白血病是髓系白血病。

65. 根据权利要求64所述用于用途的白血病TAP、组合、核酸、脂质体、组合物、疫苗、细胞或细胞群，其中所述髓系白血病是急性髓系白血病(AML)。

66. 根据权利要求63-65中任一项所述用于用途的白血病TAP、组合、核酸、脂质体、组合物、疫苗、细胞或细胞群，其用于与至少一种另外的抗肿瘤剂或疗法组合使用。

67. 根据权利要求66所述用于用途的白血病TAP、组合、核酸、脂质体、组合物、疫苗、细胞或细胞群，其中所述至少一种另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、放射疗法或手术。

用于急性髓系白血病(AML)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2020年4月14日提交的美国临时专利申请号63/009,853的权益,其通过引用以其全文并入本文。

[0003] 序列表

[0004] N/A。

技术领域

[0005] 本发明一般地涉及癌症,并且更具体地涉及可用于基于T细胞的癌症免疫疗法的急性髓系白血病的特异性肿瘤抗原。

背景技术

[0006] 急性髓系白血病(AML)是最具侵袭性的血液系统恶性肿瘤,是异质性疾病,其特征在于异常的表现遗传模式、线粒体蛋白质稳态紊乱和相对较少的突变(Li等人,2016; Ntziachristos等人,2016; Ishizawa等人,2019; Fennell等人,2019)。值得注意的是,AML的遗传和表现遗传变化可能先于诊断很多年(Abelson等人,2018; Desai等人,2018)。此外,治愈不仅需要消除大量肿瘤细胞,还需要消除白血病干细胞(Shlush等人,2017; Boyd等人,2018)。目前,大多数患者在化疗后复发,<60岁患者的5年总体生存率为40%,而年龄≥60岁的患者(占AML病例的大多数)仅为10-20%(Vasu等人,2018)。

[0007] 在过去几年中,对癌症免疫疗法的热情主要受到两项重大突破的推动:i)用于治疗黑素瘤和选定类型实体瘤的免疫检查点疗法,和ii)用于治疗淋巴恶性肿瘤的嵌合抗原受体。然而,AML尚未从这些创新中获益,主要是因为缺乏可操作的免疫靶标。根据T细胞识别的主要组织相容性复合体MHC相关肽(MAP)是抗癌反应的核心观点(Coulie等人,2014),有证据表明AML细胞应该向CD8⁺ T细胞呈递免疫原性MAP:i) AML细胞表达高密度MHC I类分子(Berlin等人,2015)和ii) AML患者的骨髓含有具有耗竭的表型和转录特征(因此具有抗原识别)的CD8⁺ T细胞(Knaus等人,2018)。然而,能够引发保护性免疫应答的AML抗原的性质仍难以捉摸。

[0008] 引起癌症免疫学家注意的第一类MAP是肿瘤相关抗原(TAA),相对于正常细胞,TAA在肿瘤细胞中过表达。由于识别自身抗原的高亲和力T细胞被胸腺选择的中心耐受过程消除,TAA基本上由低亲和力T细胞识别。因此,基于TAA的疫苗对AML进化没有令人信服的影响。研究最多的AML TAA:Wilms肿瘤1(WT1)尤其获得了令人失望的结果(Di Stasi等人,2015; Maslak等人,2018; Rashidi和Walter,2016)。重要的是,最近的报道显示,靶向WT1衍生肽的TCR基因疗法(其中T细胞经工程化表达对选定抗原具有高亲和力的TCR)可以持久地防止同种异体造血干细胞移植接受者的复发(Chapuis等人,2019)。总体上,这些研究表明WT1衍生肽的免疫原性较差,并且需要以工程化T细胞为靶点才能充分发挥其治疗潜力。

[0009] 与TAA相比,肿瘤特异性抗原(TSA)是仅由肿瘤细胞呈递的MAP。迄今为止,突变的TSA(mTSA),也称为新抗原,最近在寻求针对实体瘤的疫苗方面引起了广泛关注。事实上,

mTSA可以具有高免疫原性,因为在诱导中枢耐受的髓质胸腺细胞(mTEC)中未发现它们。然而,mTSA提示了两个警告。首先,它们通常对每名患者的肿瘤(私人新抗原)是独特的。其次,它们不像最初预测的那样普遍(Knaus等人,2018)。与AML细胞的低突变负担一致,仅有一种mTSA已通过对原代AML细胞的质谱(MS)分析得到验证(van der Lee等人,2019)。尚未对这种源自NPM1基因移码的mTSA的治疗潜力进行评价,但根据现有证据,它未引发AML患者的自发免疫应答(van der Lee等人,2019)。

[0010] 有鉴于此,迫切需要确定可再次引发针对AML的治疗性免疫应答的抗原。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。

[0011] 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

[0012] 本公开提供了以下1至67项:

[0013] 1.一种白血病肿瘤抗原肽(TAP),包含以下氨基酸序列中的一个:

[0014]

序列 (SEQ ID NO)	序列 (SEQ ID NO)	序列 (SEQ ID NO)	序列 (SEQ ID NO.)
RQISVQASL (1)	SIQRNLSL (49)	AQDIILQAV (97)	RYLANKIHI (145)
DRELRNLEL (2)	NVSSHVHTV (50)	PPRPLGAQV (98)	SLLSGLLRA (146)
GARQQIHSW (3)	ALASHLIEA (51)	FNVALNARY (99)	SRIHLVVL (147)
SGKLRVAL (4)	ALDDITIQL (52)	GPGSRESTL (100)	SSSPVRGPSV (148)
RSASSATQVHK (5)	ALGNTVPAV (53)	IPHQRSSL (101)	STFSLYLKK (149)
SASSATQVHK (6)	ALLPAVPSL (54)	LTDRIYTL (102)	SLDLLPLSI (150)
FLLEFKPVS (7)	APAPPPVAV (55)	NLKEKKALF (103)	VTDLLALTV (151)
GPQVRGSI (8)	APDKKITL (56)	VLFGGKVSGA (104)	RTQITKVSLKK (152)
IRMKAQAL (9)	AQMNLLQKY (57)	VVFPPVVK (105)	ILRSPLKW (153)
KIKVFSKVY (10)	DQVIRLAGL (58)	SLLIIPKKK (106)	LSTGHLSTV (154)
LLSRGLLFRI (11)	ETTSQVRKY (59)	APGAAGQRL (107)	TVEEYLVNI (155)
LPIASASLL (12)	GGSLIHPQW (60)	KLQDKEIGL (108)	QIKTKLLGSL (156)
LYFLGHGSI (13)	GLYYKLHNV (61)	SLREPQPAL (109)	LPSFSHLLL (157)
NPLQLSLSI (14)	GQKPVILTY (62)	TPGRSTQAI (110)	CLRIGPVTI (158)
DLMLRESL (15)	GSLDFQRGW (63)	APRGTAAL (111)	HVSDGSTALK (159)
VTFKLSLF (16)	HHLVETLKF (64)	IASPIALL (112)	IAYSVRALR (160)
IALYKQVL (17)	HLLSETPQL (65)	ILFQNSALK (113)	PRGFLSAL (161)
IVATGSLLK (18)	HQLYRASAL (66)	ILKKNISI (114)	ISSWLISL (162)

[0015]

KIKNKTKNK (19)	HTDDIENAKY (67)	IPLAVRTI (115)	IPLNPFSSL (163)
KLLSLTIYK (20)	IAAPILHV (68)	LPRNKPLL (116)	LSDRQLSL (164)
NILKKTIVL (21)	KAFPFHIIF (69)	PAPHPAAL (117)	LSHPAPSSL (165)
NPCLKDIL (22)	KATEYVHSL (70)	SPVVRVGL (118)	LRKAVIDPIL (166)
NQKKVRIL (23)	KFSNVTMLF (71)	TLNQGINVYI (119)	ILLEEQLSI (167)
PFPLVQVEPV (24)	KLLEKAISI (72)	RPRGPRTAP (120)	LTSISIRPV (168)
SPQSGPAL (25)	KPMPTKVVF (73)	SVQLLEQAIHK (121)	TISECPLLI (169)
TSRLPKIQK (26)	NVNRPLTMK (74)	RTPKNYQHW (122)	TLKLKKIFF (170)
LLDNILQSI (27)	REPYELTVPAL (75)	ALPVALPSL (123)	ILLSNFSSL (171)
RLEVRKVIL (28)	SEAEAAKNAL (76)	SLQILVSSL (124)	LGGAWKAVF (172)
LSWGYFLFK (29)	SLWGQPAEA (77)	ISNKVPKLF (125)	LSASHLSSL (173)
TILPRILTL (30)	SPADHRGYASL (78)	TVIRIAIVNK (126)	AGDIARLI (174)
EGKIKRNI (31)	SPQSAAAEI (79)	KEIFLELRL (127)	DRGILRNLL (175)
FLASFVEKTVL (32)	SPVVHQSL (80)	TLRSPGSSL (128)	GLRLIHVSL (176)
ILASHNLTV (33)	SPYRTPVL (81)	TVRGDVSSL (129)	GLRLLHVSL (177)
IQLTSVHLL (34)	SVFAGVVG (82)	ALDPLLLRI (130)	LHNEKGLSL (178)
LELISFLPVL (35)	SYSAPAHARL (83)	ISLIVTGLK (131)	LPSFSRPSGII (179)
NFCMLHQSI (36)	THGSEQLHL (84)	KILDVNLRI (132)	LSSRLPLGK (180)
PARPAGPL (37)	TQAPPNVVL (85)	ERVYIRASL (133)	MIGIKRLL (181)
PLPIVPAL (38)	VLVPYEPPQV (86)	ILDLESRY (134)	NLKKREIL (182)
SNLIRTGSH (39)	VSFPDVRKV (87)	KTFVQKTL (135)	RMVAYLQQL (183)

[0016]

VPAPAI (40)	VVFDKSDLAKY (88)	LYIKSLPAL (136)	SPARALPSL (184)
KGHGGPRSW (41)	YSHHSGLEY (89)	VLKEKNASL (137)	TVPGIQR (185)
ITSSAVTTALK (42)	YYLDWIHHY (90)	LGISLTLKY (138)	VSRNYVLLI (186)
LLLPESPSI (43)	SVYKYLKAK (91)	DLLPKKLL (139)	LTVPLSVFW (187)
VILIPPPK (44)	IYQFIMDRF (92)	HSLISIVYL (140)	KLNQAFLLV (188)
AVLLPKPPK (45)	GTLQGIRAW (93)	IAGALRSVL (141)	RLVSSTLLQK (189)
TQVSMAESI (46)	AQKVSQGAA (94)	IGNPILRVL (142)	LPSHSLLI (190)
LNHLRTSI (47)	LYPSKLTHF (95)	IYAPHIRLS (143)	
NTSHLPLIY (48)	ATQNTIIGK (96)	LRSQILSY (144)	

[0017] 2.如项1所述的白血病TAP,其包含SEQ ID NO:97-154中列出的所述氨基酸序列中的一个。

[0018] 3.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-A*01:01分子结合并且包含氨基酸序列NTSHLPLIY (SEQ ID NO:48)、HTDDIENAKY (SEQ ID NO:67)、YSHHSGLEY (SEQ ID NO:89)、ILDLESRY (SEQ ID NO:134)、VTDLLALTV (SEQ ID NO:151)或LSDRQLSL (SEQ ID NO:164),优选ILDLESRY (SEQ ID NO:134)或VTDLLALTV (SEQ ID NO:151)。

[0019] 4.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-A*02:01分子结合并且包含氨基酸序列FLLEFKPVS (SEQ ID NO:7)、LLSRGLLFRI (SEQ ID NO:11)、LLDNILQSI (SEQ ID NO:27)、FLASFVEKTVL (SEQ ID NO:32)、ILASHNLTV (SEQ ID NO:33)、IQLTSVHLL (SEQ ID NO:34)、LELISFLPVL (SEQ ID NO:35)、LLLPESPSI (SEQ ID NO:43)、ALASHLIEA (SEQ ID NO:51)、ALDDITIQL (SEQ ID NO:52)、ALGNTVPAV (SEQ ID NO:53)、ALLPAVPSL (SEQ ID NO:54)、GLYYKLHNV (SEQ ID NO:61)、HLLSETPQL (SEQ ID NO:65)、KLEKAFSI (SEQ ID NO:72)、SLWGQPAEA (SEQ ID NO:77)、SVFAGVGV (SEQ ID NO:82)、VLVPYEPQV (SEQ ID NO:86)、VLFGGKVSGA (SEQ ID NO:104)、KLQDKEIGL (SEQ ID NO:108)、TLNQGINVYI (SEQ ID NO:119)、ALPVALPSL (SEQ ID NO:123)、ALDPLLLRI (SEQ ID NO:130)、KILDVNLRI (SEQ ID NO:132)、SLLSGLLRA (SEQ ID NO:146)、SLDLLPLSI (SEQ ID NO:150)、ILLEEQSLI (SEQ ID NO:167)、LTSISIRPV (SEQ ID NO:168)、TISECPLLI (SEQ ID NO:169)、ILLSNFSSL (SEQ ID NO:171)、RMVAYLQQL (SEQ ID NO:183)或KLNQAFLLV (SEQ ID NO:188),优选VLFGGKVSGA (SEQ ID NO:104)、KLQDKEIGL (SEQ ID NO:108)、TLNQGINVYI (SEQ ID NO:119)、ALPVALPSL (SEQ ID NO:123)、ALDPLLLRI (SEQ ID NO:130)、KILDVNLRI (SEQ ID NO:132)、SLLSGLLRA (SEQ ID NO:146)或SLDLLPLSI (SEQ ID NO:150)。

[0020] 5.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-A*03:01分子结合并且包含氨基酸序列RSASSATQVHK (SEQ ID NO:5)、IVATGSLLK (SEQ ID NO:18)、KIKNKTCK (SEQ ID NO:19)、KLLSLTIYK (SEQ ID NO:20)、ITSSAVTTALK (SEQ ID NO:42)、VILIPPPK (SEQ ID NO:44)、NVNRPLTMK (SEQ ID NO:74)、SVYKYLKAK (SEQ ID NO:91)、VVFPPVVK (SEQ ID NO:105)、ILFQNSALK (SEQ ID NO:113)、TVIRIAIVNK (SEQ ID NO:126)、ISLIVTGLK (SEQ ID NO:

131)、HVSDGSTALK (SEQ ID NO:159)、IAYSVRALR (SEQ ID NO:160)、LSSRLPLGK (SEQ ID NO:180) 或RLVSSTLLQK (SEQ ID NO:189), 优选VVFPPFPVNK (SEQ ID NO:105)、ILFQNSALK (SEQ ID NO:113)、TVIRIAIVNK (SEQ ID NO:126) 或ISLIVTGLK (SEQ ID NO:131)。

[0021] 6. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*11:01分子结合并且包含氨基酸序列SASSATQVHK (SEQ ID NO:6)、AVLLPKPPK (SEQ ID NO:45)、ATQNTIIGK (SEQ ID NO:96)、SLLIIPKKK (SEQ ID NO:106)、SVQLLEQAIHK (SEQ ID NO:121)、STFSLYLKK (SEQ ID NO:149) 或RTQITKVSLKK (SEQ ID NO:152), 优选SLLIIPKKK (SEQ ID NO:106)、SVQLLEQAIHK (SEQ ID NO:121)、STFSLYLKK (SEQ ID NO:149) 或RTQITKVSLKK (SEQ ID NO:152)。

[0022] 7. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*24:02分子结合并且包含氨基酸序列LYFLGHGSI (SEQ ID NO:13)、NFCMLHQSI (SEQ ID NO:36)、KFSNVTMLF (SEQ ID NO:71)、IYQFIMDRF (SEQ ID NO:92)、LYPSKLTHF (SEQ ID NO:95) 或RYLANKIHI (SEQ ID NO:145), 优选RYLANKIHI (SEQ ID NO:145)。

[0023] 8. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*26:01分子结合并且包含氨基酸序列ETTSQVRKY (SEQ ID NO:59) 或TVPGIQRY (SEQ ID NO:185)。

[0024] 9. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*29:02分子结合并且包含氨基酸序列VVFDKSDLAKY (SEQ ID NO:88)、FNVALNARY (SEQ ID NO:99) 或LGISLTLKY (SEQ ID NO:138), 优选FNVALNARY (SEQ ID NO:99) 或LGISLTLKY (SEQ ID NO:138)。

[0025] 10. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*30:01分子结合并且包含氨基酸序列TSRLPKIQK (SEQ ID NO:26)、LSWGYFLFK (SEQ ID NO:29) 或LSHPAPSSL (SEQ ID NO:165)。

[0026] 11. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-A*68:02分子结合并且包含氨基酸序列NVSSHVHTV (SEQ ID NO:50) 或SSSPVRGPSV (SEQ ID NO:148), 优选SSSPVRGPSV (SEQ ID NO:148)。

[0027] 12. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*07:02分子结合并且包含氨基酸序列GPQVRGSI (SEQ ID NO:8)、SPQSGPAL (SEQ ID NO:25)、VPAPAQAI (SEQ ID NO:40)、APAPPPVAV (SEQ ID NO:55)、APDKKITL (SEQ ID NO:56)、KPMPTKVVF (SEQ ID NO:73)、SPADHRGYASL (SEQ ID NO:78)、SPQSAAAE (SEQ ID NO:79)、SPVVHQSL (SEQ ID NO:80)、SPYRTPVL (SEQ ID NO:81)、PPRPLGAQV (SEQ ID NO:98)、GPGSRESTL (SEQ ID NO:100)、APGAAGQRL (SEQ ID NO:107)、TPGRSTQAI (SEQ ID NO:110)、APRGTAAL (SEQ ID NO:111)、SPVVRVGL (SEQ ID NO:118)、RPRGPRTAP (SEQ ID NO:120)、TLRSPGSSL (SEQ ID NO:128)、TVRGDVSSL (SEQ ID NO:129)、LPSFSHFLLL (SEQ ID NO:157)、PRGFLSAL (SEQ ID NO:161)、IPLNPFSSL (SEQ ID NO:163)、LPSFSRPSGII (SEQ ID NO:179) 或SPARALPSL (SEQ ID NO:184), 优选PPRPLGAQV (SEQ ID NO:98)、GPGSRESTL (SEQ ID NO:100)、APGAAGQRL (SEQ ID NO:107)、TPGRSTQAI (SEQ ID NO:110)、APRGTAAL (SEQ ID NO:111)、SPVVRVGL (SEQ ID NO:118)、RPRGPRTAP (SEQ ID NO:120)、TLRSPGSSL (SEQ ID NO:128) 或TVRGDVSSL (SEQ ID NO:129)。

[0028] 13. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*08:01分子结合并且包含氨基酸序列SGKLRVAL (SEQ ID NO:4)、NPLQLSLSI (SEQ ID NO:14)、DLMLRESL (SEQ ID

NO:15)、IALYKQVL (SEQ ID NO:17)、NILKKTVL (SEQ ID NO:21)、NPKLKDIL (SEQ ID NO:22)、NQQKVRIL (SEQ ID NO:23)、RLEVRKVIL (SEQ ID NO:28)、EGKIKRNI (SEQ ID NO:31)、LNHLRTSI (SEQ ID NO:47)、SIQRNLSL (SEQ ID NO:49)、IPHQRSSL (SEQ ID NO:101)、NLKEKKALF (SEQ ID NO:103)、ILKKNISI (SEQ ID NO:114)、VLKEKNASL (SEQ ID NO:137)、DLLPKKLL (SEQ ID NO:139)、SRIHLVVL (SEQ ID NO:147)、QIKTKLLGSL (SEQ ID NO:156)、TLKLKKIFF (SEQ ID NO:170)、MIGIKRLL (SEQ ID NO:181) 或 NLKKREIL (SEQ ID NO:182), 优选 IPHQRSSL (SEQ ID NO:101)、NLKEKKALF (SEQ ID NO:103)、ILKKNISI (SEQ ID NO:114)、VLKEKNASL (SEQ ID NO:137)、DLLPKKLL (SEQ ID NO:139) 或 SRIHLVVL (SEQ ID NO:147)。

[0029] 14. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*14:01分子结合并且包含氨基酸序列DRELRLNEL (SEQ ID NO:2)、SNLIRTGSH (SEQ ID NO:39)、DQVIRLAGL (SEQ ID NO:58)、HQLYRASAL (SEQ ID NO:66)、SLQILVSSL (SEQ ID NO:124)、ERVYIRASL (SEQ ID NO:133)、LYIKSLPAL (SEQ ID NO:136)、IAGALRSVL (SEQ ID NO:141)、ISSWLISSL (SEQ ID NO:162)、DRGILRNLL (SEQ ID NO:175)、GLRLIHVSL (SEQ ID NO:176) 或 GLRLLHVSL (SEQ ID NO:177), 优选 SLQILVSSL (SEQ ID NO:124)、ERVYIRASL (SEQ ID NO:133)、LYIKSLPAL (SEQ ID NO:136) 或 IAGALRSVL (SEQ ID NO:141)。

[0030] 15. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*15:01分子结合并且包含氨基酸序列KIKVFSKVY (SEQ ID NO:10)、AQMNLLQKY (SEQ ID NO:57)、GQKPVILTY (SEQ ID NO:62) 或 AQKVSQGAA (SEQ ID NO:94)。

[0031] 16. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*27:05分子结合并且包含氨基酸序列RQISVQASL (SEQ ID NO:1) 或 LRSQILSY (SEQ ID NO:144), 优选 LRSQILSY (SEQ ID NO:144)。

[0032] 17. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*38:01分子结合并且包含氨基酸序列TQVSMAESI (SEQ ID NO:46)、HHLVETLKF (SEQ ID NO:64) 或 THGSEQLHL (SEQ ID NO:84)。

[0033] 18. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*40:01分子结合并且包含氨基酸序列REPYELTVPAL (SEQ ID NO:75) 或 SEAEAAKNAL (SEQ ID NO:76)。

[0034] 19. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*44:03分子结合并且包含氨基酸序列KEIFLELRL (SEQ ID NO:127)。

[0035] 20. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*51:01分子结合并且包含氨基酸序列LPIASASLL (SEQ ID NO:12)、PFPLVQVEPV (SEQ ID NO:24)、PLPIVPAL (SEQ ID NO:38)、IAAPILHV (SEQ ID NO:68)、IPLAVRTI (SEQ ID NO:115)、LPRNKPLL (SEQ ID NO:116) 或 LPSHSLLI (SEQ ID NO:190), 优选 IPLAVRTI (SEQ ID NO:115) 或 LPRNKPLL (SEQ ID NO:116)。

[0036] 21. 如项1或2所述的白血病TAP, 其中所述白血病TAP与HLA-B*57:01分子结合并且包含氨基酸序列GARQQIHSW (SEQ ID NO:3)、VTFKLSLF (SEQ ID NO:16)、KGHGGPRSW (SEQ ID NO:41)、GSLDFQRGW (SEQ ID NO:63)、KAPPFHIF (SEQ ID NO:69)、GTLQGIRAW (SEQ ID NO:93)、RTPKNYQHW (SEQ ID NO:122)、ISNKVPKLF (SEQ ID NO:125)、KTFVQQKTL (SEQ ID NO:135)、ILRSPLKW (SEQ ID NO:153) 或 LTVPLSVFW (SEQ ID NO:183), 优选 RTPKNYQHW (SEQ ID NO:122)、ISNKVPKLF (SEQ ID NO:125)、KTFVQQKTL (SEQ ID NO:135) 或 ILRSPLKW (SEQ ID

NO:153)。

[0037] 22.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-B*57:03分子结合并且包含氨基酸序列GGSLIHPQW (SEQ ID NO:60) 或LGGAWKAVF (SEQ ID NO:172)。

[0038] 23.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-C*03:03分子结合并且包含氨基酸序列PARPAGPL (SEQ ID NO:37)、IASPIALL (SEQ ID NO:112) 或HSLISIVYL (SEQ ID NO:140),优选IASPIALL (SEQ ID NO:112) 或HSLISIVYL (SEQ ID NO:140)。

[0039] 24.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-C*05:01分子结合并且包含氨基酸序列SLDLLPLSI (SEQ ID NO:150)。

[0040] 25.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-C*06:02分子结合并且包含氨基酸序列IRMKAQAL (SEQ ID NO:9)、KATEYVHSL (SEQ ID NO:70)、VSFPDVRKV (SEQ ID NO:87)、IGNPILRVL (SEQ ID NO:142)、LSTGHLSTV (SEQ ID NO:154) 或LRKAVDPIIL (SEQ ID NO:166),优选IGNPILRVL (SEQ ID NO:142) 或LSTGHLSTV (SEQ ID NO:154)。

[0041] 26.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-C*07:01分子结合并且包含氨基酸序列IGNPILRVL (SEQ ID NO:142)、IYAPHIRLS (SEQ ID NO:143)、TVEEYLVNI (SEQ ID NO:155)、LHNEKGLSL (SEQ ID NO:178) 或VSRNYVLLI (SEQ ID NO:186),优选IGNPILRVL (SEQ ID NO:142) 或IYAPHIRLS (SEQ ID NO:143)。

[0042] 27.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-C*07:02分子结合并且包含氨基酸序列TILPRILTL (SEQ ID NO:30)、SYSPAHLARL (SEQ ID NO:83)、TQAPPNVVL (SEQ ID NO:85)、YYLDWIHHY (SEQ ID NO:90)、SLREPQPAL (SEQ ID NO:109)、PAPPHPAAL (SEQ ID NO:117) 或CLRIGPVTL (SEQ ID NO:158),优选SLREPQPAL (SEQ ID NO:109) 或PAPPHPAAL (SEQ ID NO:117)。

[0043] 28.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-C*08:02分子结合并且包含氨基酸序列AQDIILQAV (SEQ ID NO:97)、LTDRIYLT (SEQ ID NO:102) 或AGDIIARLI (SEQ ID NO:174),优选AQDIILQAV (SEQ ID NO:97) 或LTDRIYLT (SEQ ID NO:102)。

[0044] 29.如项1或2所述的白血病TAP,其中所述白血病TAP与HLA-C*12:03分子结合并且包含氨基酸序列LSASHLSSL (SEQ ID NO:173)。

[0045] 30.如项1-29中任一项所述的白血病TAP,其由位于基因组的非蛋白质编码区的序列编码。

[0046] 31.如项30所述的白血病TAP,其中所述基因组的非蛋白质编码区是非翻译转录区(UTR)。

[0047] 32.如项30所述的白血病TAP,其中所述基因组的非蛋白质编码区是内含子。

[0048] 33.如项30所述的白血病TAP,其中所述基因组的非蛋白质编码区是基因间区。

[0049] 34.一种包含项1-33中任一项所限定的白血病TAP中的至少两种的组合。

[0050] 35.一种编码项1-33中任一项所述的白血病TAP或项34所述的组合的核酸。

[0051] 36.如项35所述的核酸,其为mRNA或病毒载体。

[0052] 37.一种脂质体,其包含项1-33中任一项所述的白血病TAP、项34所述的组合或者项35或36所述的核酸。

[0053] 38.一种组合物,其包含项1-33中任一项所述的白血病TAP、项34所述的组合、项35或36所述的核酸或者项37所述的脂质体,以及药学上可接受的载体。

- [0054] 39. 一种疫苗,其包含项1-33中任一项所述的白血病TAP、项34所述的组合、项35或36所述的核酸、项37所述的脂质体或者项38所述的组合物,以及佐剂。
- [0055] 40. 一种分离的主要组织相容性复合体 (MHC) I类分子,在其肽结合槽中包含项1-33中任一项所述的白血病TAP。
- [0056] 41. 如项40所述的分离的MHC I类分子,其为多聚体形式。
- [0057] 42. 如项41所述的分离的MHC I类分子,其中,所述多聚体是四聚体。
- [0058] 43. 一种分离的细胞,其包含 (i) 项1-33中任一项所述的白血病TAP、(ii) 项34所述的组合或 (iii) 包含编码项1-33任一项的TAP或项34所述的组合的核苷酸序列的载体。
- [0059] 44. 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MHC I类分子在其肽结合槽中包含项1-33中任一项所述的白血病TAP或项34所述的组合。
- [0060] 45. 如项44所述的细胞,其是抗原呈递细胞 (APC)。
- [0061] 46. 如项45所述的细胞,其中所述APC是树突状细胞。
- [0062] 47. 一种T细胞受体 (TCR),其特异性识别项40-42中任一项所述的分离的MHC I类分子和/或在项44-46中任一项所述的细胞的表面表达的MHC I类分子。
- [0063] 48. 如项47所述的TCR,其中所述TCR包含包含TCRB (TCR β) 链,所述TCR β 链包含互补决定区3 (CDR3),所述CDR3包含SEQ ID NO:191-219中列出的氨基酸序列中的一个。
- [0064] 49. 一种分离的细胞,所述分离的细胞在其细胞表面表达项47或48所述的TCR。
- [0065] 50. 如项49所述的分离的细胞,其是CD8⁺T淋巴细胞。
- [0066] 51. 一种细胞群,其包含至少0.5%的如项49或50所限定的分离的细胞。
- [0067] 52. 一种治疗受试者的白血病的方法,所述方法包括向所述受试者施用有效量的:
(i) 项1-33中任一项所述的白血病TAP; (ii) 项34所述的组合; (iii) 项35或36所述的核酸;
(iv) 项37所述的脂质体; (v) 项38所述的组合物; (vi) 项39所述的疫苗; (vii) 项43-46、49和50中任一项所述的细胞;或 (viii) 项51所述的细胞群。
- [0068] 53. 如项52所述的方法,其中所述白血病是髓系白血病。
- [0069] 54. 如项53所述的方法,其中所述髓系白血病是急性髓系白血病 (AML)。
- [0070] 55. 如项52-54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另外的抗肿瘤剂或疗法。
- [0071] 56. 如项55所述的方法,其中所述至少一种另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、放射疗法或手术。
- [0072] 57. (i) 项1-33中任一项所述的白血病TAP; (ii) 项34所述的组合; (iii) 项35或36所述的核酸; (iv) 项37所述的脂质体; (v) 项38所述的组合物; (vi) 项39所述的疫苗; (vii) 项43-46、49和50中任一项所述的细胞;或 (viii) 项51所述的细胞群用于治疗受试者的白血病的用途。
- [0073] 58. (i) 项1-33中任一项所述的白血病TAP; (ii) 项34所述的组合; (iii) 项35或36所述的核酸; (iv) 项37所述的脂质体; (v) 项38所述的组合物; (vi) 项39所述的疫苗; (vii) 项43-46、49和50中任一项所述的细胞;或 (viii) 项51所述的细胞群用于制造用于治疗受试者的白血病的药物的用途。
- [0074] 59. 如项57或58所述的用途,其中所述白血病是髓系白血病。

- [0075] 60. 如项59所述的用途, 其中所述髓系白血病是急性髓系白血病 (AML)。
- [0076] 61. 如项57-60中任一项所述的用途, 进一步包括使用至少一种另外的抗肿瘤剂或疗法。
- [0077] 62. 如项61所述的用途, 其中所述至少一种另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、放射疗法或手术。
- [0078] 63. (i) 项1-33中任一项所述的白血病TAP; (ii) 项34所述的组合; (iii) 项35或36所述的核酸; (iv) 项37所述的脂质体; (v) 项38所述的组合物; (vi) 项39所述的疫苗; (vii) 项43-46、49和50中任一项所述的细胞; 或 (viii) 项51所述的细胞群, 用于治疗受试者的白血病的用途。
- [0079] 64. 根据项63的用于用途的白血病TAP、组合、核酸、脂质体、组合物、疫苗、细胞或细胞群, 其中所述白血病是髓系白血病。
- [0080] 65. 根据项64的用于用途的白血病TAP、组合、核酸、脂质体、组合物、疫苗、细胞或细胞群, 其中所述髓系白血病是急性髓系白血病 (AML)。
- [0081] 66. 据项63-65中任一项的用于用途的白血病TAP、组合、核酸、脂质体、组合物、疫苗、细胞或细胞群, 其用于与至少一种另外的抗肿瘤剂或疗法组合使用。
- [0082] 67. 根据项66的用于用途的白血病TAP、组合、核酸、脂质体、组合物、疫苗、细胞或细胞群, 其中所述至少一种另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、放射疗法或手术。
- [0083] 本发明的其他目的、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

[0084] 在附图中:

[0085] 图1A-D是显示造血祖细胞在发现AML中的TSA方面是比mTEC更好的对照的图表。图1A: 通过来自6个mTEC或来自6个MPC样本的任一组合k-mer, 比较19个AML试样中每一个的k-mer组中k-mer耗竭的有效性。发生率 <2 的K-mer被忽略, 并且以典型模式生成水母数据库以进行此比较。图1B: 所有AML试样的组合k-mer与来自6个mTEC和来自图1A中使用的6个MPC样本的k-mer之间的重叠。此处重新应用了图1A中使用的数据库构建参数。图1C: 对来自指定组织的纯化细胞群中表达 (TPM ≥ 1) 的蛋白质编码基因进行T分布随机邻域嵌入 (t-SNE) 分析。图1D: 比较用于绘制小图C的指定组织和细胞群中表达的蛋白质编码基因 (TPM ≥ 1) 的总数。Pluri_stem: 多能干细胞; Ery: 红细胞; Precu: 前体; Lympho: 淋巴细胞; Granulo: 粒细胞; Mono: 单核细胞。Mann whitney U检验用于比较mTEC与其他组织 (****p <0.0001), 柱状图显示平均值和标准偏差。

[0086] 图2描绘了基于MPC的TSA发现方法的示意图。(A) 基于mTEC k-mer耗竭的TSA发现的工作流程示意图。(B) ERE衍生的MAP发现的工作流程示意图。(C) mTECs+MPCs k-mer耗竭TSA发现的工作流程示意图。(D) DKE方法的工作流程示意图。此处示出的是AML#1样本的工作流程。使用10倍变化作为最小值, 以将k-mer视为过表达 (还应用了其他过滤器, 参见方法)。至于其他三种方法, 在对从用于进行RNA测序的相同AML样本洗脱的MAP进行MS识别之前, 将获得的计算机全框翻译重叠群数据库与个性化的典型蛋白质组连接。

[0087] 图3A-H显示了基于MPC的方法识别AML中的大部分TSA^{hi}。给出 σ 的分布(以黑色绘制)。图3A:每个AML试样($n=19$)的MAP比例,得自基于TPM表达分为10个不同组(十分位)的转录物。十分位10具有表达最高的转录物,而十分位1的表达最低。方框显示分布的中位值、第25个和第75个百分位数,须线延伸至最小值和最大值。图3B:MAP(点)的累积频率的正态分布,与MS识别的AML试样中能够编码它们的RNA-seq读数(rphm)总数的对数关系。给出了分布的平均值(μ)和标准偏差(σ)(以黑色绘制)。图3C:基于图3B的正态分布参数计算RNA序列在不同指示的倍数变化(FC,原始rphm $\times$ FC)后生成MAP的概率。图3D:用于将目的MAP(MOI)分成TAA、HSA、TSA^{hi}的决策树。“正常组织”是指所有组织(GTEx、纯化的造血细胞和mTEC),以及血液/BM仅指纯化的造血细胞。图3E:通过每种指定的蛋白质组学方法获得的MOI计数的比较。图3F:比较指定方法之间的TSA^{hi}同一性的Venn图。图3G:保留时间观察值与保留时间预测值(左)或疏水性指数(右)之间的Pearson相关性。图3H:使用Comet成功重新识别指示的MAP的中位值和四分位数范围频率。

[0088] 图4A-K显示了TSA^{hi}主要源自内含子翻译并且在许多患者中共享。图4A:热图描绘了来自GTEx的总正常组织($n=12-50$,取决于可获得的样本)中、正常分选的造血细胞群($n=3-16$,取决于可获得的样本)中或mTEC($n=11$)中每个识别的TSA^{hi}的平均RNA表达(rphm+1的对数)。还报告了在临床试验中被评价为具有安全性的TAA.Prec:前体。图4B:比较19个AML试样和MPC($n=16$)中平均rphm表达之间的TSA^{hi}倍数变化。点显示每个MOI,方框显示分布的中位值、第25个和第75个百分位数,须线延伸至最小值和最大值。图4C:生成指示的MOI的生物型(基因组区域或事件)的分布。外显子-内含子:与外显子-内含子连接重叠的肽(保留内含子);ncRNA:非编码RNA;OoF翻译:框外翻译。图4D:19个AML样本和437个Leucegene患者中的TSA^{hi}RNA表达。图4E:能够呈递TSA^{hi}的HLA同种异型的群体覆盖率(呈递TSA^{hi}的19个AML试样等位基因+由MHCcluster计算的混杂结合物)。这是使用IEDB群体覆盖工具(www.iedb.org)计算的。柱状图表示世界人群中携带多达六种同种异型的个体的频率(x轴),并且群体覆盖的累积百分比显示为点。图4F:基于TSA^{hi}RNA表达(如果rphm ≥ 2 ,则认为表达)、患者的HLA等位基因(OptiType)和混杂结合物,Leucegene队列中HLA-TSA^{hi}复合物分布。显示了高(上四分位数)和低(所有其他患者)TSA^{hi}表达者之间的区别。图4G:Leucegene患者在诊断和复发时的 $\#_{pred}$ HLA-TSA^{hi}复合物。图4H:可以由AML母细胞(blasts)的HLA等位基因呈递的TSA^{hi}的RNA表达在诊断时和复发时从15名患者成对纯化(数据来自(Toffalori等人,2019))。使用Wilcoxon配对符号秩检验进行比较。图4I:另报道,TSA^{hi}样本间共有(如果rphm >0 ,则认为表达)在分选的母细胞($n=12$)或白血病干细胞(LSC, $n=8$)中的比较(Corces等人,2016)。图4J:图4I中所示的样本中HLA-ABC分子的RNA表达。显示平均值+SD。图4K:比较表达(rphm >0) $\geq$ TSA^{hi}中位数($n=207$)的Leucegene患者相比于其他($n=230$)的指示的LSC特征基因集的GSEA分析(Eppert等人,2011)。NES,归一化富集评分。

[0089] 图5A-F显示了大量TSA^{hi}的呈递与更好的生存率相关。图5A:表达高数量($n=98$,图5B中的上四分位数)相比于低数量($n=275$,所有其他患者)的HLA-TSA^{hi}复合物的Leucegene患者之间的Kaplan-Meier生存分析。统计学显著性由对数秩检验确定。图5B:用于5年总体生存率的多变量分析的森林图。HR,校正风险比;CI,置信区间;adv,不利的;fav,有利的;int,中间。NPM1/FLT3相互作用=存在NPM1^{mut}和FLT3-ITD。图5C:从(A)中进行的分析中移除指示数量的TSA^{hi}后计算的对数秩p值;对每个数值进行1000次排列,并报告平均值+SD。图

5D:在图5C中获得的显著性p值的百分比。图5E:在从图5A中的分析中可选地移除每个TSA^{hi}之后重新计算的对数秩p值的比较。图5F:在从图5A中的分析中可选地移除每个HLA等位基因之后重新计算的对数秩p值的比较。

[0090] 图6A-0显示了TSA^{hi}呈递触发细胞毒性T细胞应答。图6A:MOI、来自胸腺基质细胞的MAP和HIV MAP之间的免疫原性评分 (Repitope) 的比较。图6B:MOI的11个可获得的mTEC样本、5112个非免疫原性MAP和1411个免疫原性MAP (来自IEDB并在 (Ogishi和Yotsuyanagi, 2019) 中展出) 的平均RNA表达的中位值和四分位数范围。图6C:用指示肽脉冲的DC刺激后健康PBMC的IFN- γ ELISpot测定。将2项独立实验的结果合并。图6D:指示TSA^{hi} (单一供体) 的ELISpot测定。图6E:在指示肽存在下扩增的T细胞的细胞因子分泌的流式细胞术分析。图6F:在指示肽存在下扩增的T细胞之间指示的右旋体频率的代表性流式细胞术图。图6G:FEST测定:用3个不同TSA^{hi}池 (5种肽/池) 刺激10天后显著T细胞克隆型的扩增。图6H:在具有高计数相比于低计数的指示_{pred}HLA-MOI的Leucegene患者中每千个TCR读数的TCR CDR3 (CPK, 作为克隆型多样性的量度) (与图4F和11D相关)。图6I:ERGO预测的针对Leucegene中TSA^{hi} (n=66-164/组) 反应的克隆型频率。图6J:ERGO预测的针对Leucegene中TAA (n=74-207/组) 反应的克隆型频率。图6K:在所有抗MOI克隆型中, 识别在所考虑的样本中_{pred}呈递的MOI的克隆型的频率, 由_{pred}呈递的MOI的数量 (与I和J相关) 归一化。具有抗_{pres}MOI克隆型计数=0的患者被忽略。图6L:CD8A和CD8B基因的RNA表达与Leucegene中高于2rphm表达的TSA^{hi}数量之间的相关性。图6M:CD8A和CD8B基因的RNA表达与Leucegene中_{pred}HLA-TSA^{hi}数量之间的相关性。图6N:比较归一化TSA^{hi}_{pred}呈递高于中位值相比于低于中位值的患者的差异基因表达分析的火山图。点显示在高于中位值的患者中上调的基因。图6O:图6N中上调基因的GO term分析。

[0091] 图7A-G显示了TSA^{hi}表达与免疫编辑、AML驱动突变和表观遗传畸变相关。图7A:HE-TSA^{hi}数量与整个Leucegene队列 (n=437) 中指示基因的表达之间的Pearson相关性。第一个图的HLA-A、HLA-B和HLA-C的表达值相加。图7B:比较表达 $\geq$ HE-TSA^{hi}中位数的Leucegene患者相比于其他患者的PD-L1 (CD274) 基因表达, 分层为NPM1突变状态的函数。图7C:与HE-TSA^{hi}数量呈负相关的基因间GO term富集的网络分析。节点大小与基因集大小成正比。图7D:与HE-TSA^{hi}数量呈正相关的基因间GO term富集的网络分析。图7E:对于指示的基因, 在WT和突变患者中表达 $\geq$ HE-TSA^{hi}中位数的患者数量相比于其他患者的比较。统计学显著性使用Fisher精确检验确定 (**p<0.01, ***p<0.0001)。图7F:在NPM1、FLT3或DNMT3A中具有0至3个突变的患者之间的HE-TSA^{hi}数量比较。图7G:由IRFinder确定的Leucegene患者 (n=437, 列) 的内含子保留率的无监督一致性聚类。行代表1211个排名靠前的内含子, 对于一致性聚类这些内含子具有最高可变性和显著性, 按分层聚类。患者的FAB类型显示在热图下方, p值 (Fisher精确检验, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001) 显示与指示的一致性聚类显著相关。

[0092] 图8A是k-mer出现的概念说明。

[0093] 图8B是描绘在样本05H143中作为发生函数的k-mer频率分布的实例的图。

[0094] 图8C是描绘在仅mTEC和mTEC+MPC k-mer耗竭方法之间使用的阈值发生比较的图 (每个点是不同的AML样本)。

[0095] 图8D是描绘从在mTEC或mTEC+MPC k-mer耗竭后获得的所有19个AML试样获得的独

特k-mer组合之间k-mer同一性重叠的图。

[0096] 图9A是提供差异k-mer表达分析和MS数据库构建的详情的示意图。样本AML#1的MS数据库的构建作为实例提供。FC, 倍数变化。是显示差异k-mer表达分析和MS数据库构建的详情的图表。样本AML#1的MS数据库的构建作为实例提供。FC, 倍数变化。

[0097] 图9B是描述典型肽识别的累积数量(衍生自个性化典型蛋白质组的肽, 单独(Canon.)或在四种指示的方法中与重叠群序列连接)相对于平均数据库大小(线)的图。

[0098] 图9C描绘了比较基于每种方法识别的典型肽的同一性与仅基于典型个性化蛋白质组识别的肽的重叠的Venn图。

[0099] 图10A是描绘了通过每种TSA识别方法识别的目的MHC-I相关肽(MAP) (MOI) 的比例的比較的图。

[0100] 图10B是描绘了用mTEC+MPC k-mer耗竭或用差异k-mer表达方法识别的表达(rphm > 0) TSA^{hi}的(在本研究中用于识别TSA的19个中的) AML试样总数的比較的图。

[0101] 图11A是显示在Leucegene队列(n=437)中具有等于或高于2rphm的RNA表达的TSA^{hi}数量分布的图。

[0102] 图11B是显示呈递以等于或高于2rphm的水平表达的大量TSA^{hi}的Leucegene队列(n=372, 在诊断时测序并且可获得生存数据)的患者(左图中分布的上四分位数)相比于呈递低水平的患者(队列的其余部分)之间的生存比較的图。

[0103] 图11C-E是显示对于HSA(图11C)、TAA(图11D)和TSA^{lo}(图11E), 基于RNA表达(如果rphm ≥ 2, 则认为表达)、每名患者的HLA等位基因和HSA的混杂结合物预测(optitype和MHCcluster)获得的整个Leucegene队列的HLA-MOI复合物分布的图。

[0104] 图11F-H是显示对于HSA(图11F)、TAA(图11G)和TSA^{lo}(图11H), 呈递大量HLA-MOI复合物的Leucegene队列(n=372, 在诊断时测序并且可获得生存数据)的患者(上图中分布的上四分位数)相比于呈递低水平的患者(队列的其余部分)之间的生存比較的图。

[0105] 图12A描绘了HE-TSA^{hi}数量与整个Leucegene队列(n=437)中指示基因的表达之间的Pearson相关性。

[0106] 图12B描绘了显示具有高_{pred}呈递水平的TSA^{hi}的患者相比于其余患者的指示基因表达的比較的图(与图4F相关)。

[0107] 图12C描绘了ZNF445的表达与Leucegene队列中保留的内含子数量之间的Pearson相关性(用IRFinder分析, 并且如果在超过10%的转录物中保留, 则定义为保留)。

[0108] 图12D是显示对于指示的基因, 在WT和突变患者中表达 ≥ HE-TSA^{hi}中位数的患者相比于其他患者的比較的图。统计学显著性使用Fisher精确检验建立。

[0109] 图12E是显示在接受或未接受allo-HSCT的患者中表达 ≥ HE-TSA^{hi}中位数的患者相比于其他患者的比較的图。统计学显著性使用Fisher精确检验建立。

[0110] 图12F是显示HE-TSA^{hi}计数高于或低于整个Leucegene队列中HE-TSA^{hi}计数中位值的患者的FAB类型分布的图。

[0111] 图12G是显示HE-TSA^{hi}计数高于或低于整个Leucegene队列中HE-TSA^{hi}计数中位值的患者的WHO 2008分类分布的图。

[0112] 图12H是显示HE-TSA^{hi}计数高于或低于整个Leucegene队列中HE-TSA^{hi}计数中位值的患者的细胞遗传学谱分布的图。

具体实施方式

[0113] 本文使用的遗传学、分子生物学、生物化学和核酸的术语和符号遵循该领域的标准论文和文献的术语和符号,例如Kornberg和Baker,DNA Replication,第二版(W.H.Freeman,New York,1992);Lehninger,Biochemistry,第二版(Worth Publishers,New York,1975);Strachan和Read,Human Molecular Genetics,第二版(Wiley-Liss,New York,1999);Eckstein,编辑,Oligonucleotides and Analogs:A Practical Approach (Oxford University Press,New York,1991);Gait,编辑,Oligonucleotide Synthesis:A Practical Approach (IRL Press,Oxford,1984)等。所有术语应以其在相关领域中确立的典型含义来理解。

[0114] 冠词“一个”和“一种”在本文中用于指代冠词的一个或多于一个(即至少一个)语法对象。例如,“元素”是指一个元素或一个以上的元素。在本说明书中,除非上下文另有要求,否则词语“包含(comprise)”、“包含(comprises)”和“包含(comprising)”将被理解为暗示包括所陈述的步骤或元素或步骤或元素的组,但不排除任何其他步骤或元素或步骤或元素的组。

[0115] 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。

[0116] 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。

[0117] 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本发明并且不对本发明的范围构成限制。

[0118] 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本发明的实践是必不可少的。

[0119] 在本文中,术语“约”具有其普通含义。术语“约”用于表示值包括用于确定该值的设备或方法的固有误差变化,或包括接近所述值的值,例如在所述值(或值范围)的10%或5%以内。

[0120] 在本文所述的研究中,发明人使用基于蛋白质基因组学的方法从19个AML试样中识别了TSA候选者。这些TSA中有很大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)。这些AML TSA候选者的表达显示与表观遗传修饰因子(例如,DNMT3A)的突变和ZNF445(基因组印记的调节因子)的表达相关。研究还表明,AML TSA候选者在患者中高度共享,在母细胞和白血病干细胞中均有表达,它们的HLA表达与免疫编辑标志物和更好的总体生存率相关。因此,本文识别的新型AML TSA候选者可用于基于白血病T细胞的免疫疗法。

[0121] 因此,在一方面,本公开涉及白血病TAP(或白血病肿瘤特异性肽),其包含以下氨基酸序列中的一个或由其组成:

[0122]

序列 (SEQ ID NO)	序列 (SEQ ID NO)	序列 (SEQ ID NO)	序列 (SEQ ID NO.)
RQISVQASL (1)	SIQRNLSL (49)	AQDIILQAV (97)	RYLANKIHI (145)
DRELRNLEL (2)	NVSSHVHTV (50)	PPRPLGAQV (98)	SLLSGLLRA (146)
GARQQIHSW (3)	ALASHLIEA (51)	FNVALNARY (99)	SRIHLVVL (147)
SGKLRVAL (4)	ALDDITIQL (52)	GPGSRESTL (100)	SSSPVRGPSV (148)
RSASSATQVHK (5)	ALGNTVPAV (53)	IPHQRSSL (101)	STFSLYLKK (149)
SASSATQVHK (6)	ALLPAVPSL (54)	LTDRIYTL (102)	SLDLLPLSI (150)
FLLEFKPVS (7)	APAPPPVAV (55)	NLKEKKALF (103)	VTDLLALTV (151)
GPQVRGSI (8)	APDKKITL (56)	VLFGGKVSGA (104)	RTQITKVSLKK (152)
IRMKAQAL (9)	AQMNLLQKY (57)	VVFPPVVK (105)	ILRSPLKW (153)
KIKVFSKVY (10)	DQVIRLAGL (58)	SLLIIPKKK (106)	LSTGHLSTV (154)
LLSRGLLFRI (11)	ETTSQVRKY (59)	APGAAGQRL (107)	TVEEYLVNI (155)
LPIASASLL (12)	GGSLIHPQW (60)	KLQDKEIGL (108)	QIKTKLLGSL (156)
LYFLGHGSI (13)	GLYYKLHNV (61)	SLREPQPAL (109)	LPSFSHFLLL (157)
NPLQLSLSI (14)	GQKPVILTY (62)	TPGRSTQAI (110)	CLRIGPVTI (158)
DLMLRESL (15)	GSLDFQRGW (63)	APRGTAAL (111)	HVSDGSTALK (159)
VTFKLSLF (16)	HHLVETLKF (64)	IASPIALL (112)	IAYSVRALR (160)
IALYKQVL (17)	HLLSETPQL (65)	ILFQNSALK (113)	PRGFLSAL (161)
IVATGSLLK (18)	HQLYRASAL (66)	ILKKNISI (114)	ISSWLISSL (162)
KIKNKTKNK (19)	HTDDIENAKY (67)	IPLAVRTI (115)	IPLNPFSSL (163)
KLLSLTIYK (20)	IAAPILHV (68)	LPRNKPLL (116)	LSDRQLSL (164)

[0123]

NILKKTVL (21)	KAFPFHIIF (69)	PAPHPAAL (117)	LSHPAPSSL (165)
NPCLKDIL (22)	KATEYVHSL (70)	SPVVRVGL (118)	LRKAVDPIIL (166)
NQKKVRIL (23)	KFSNVTMLF (71)	TLNQGINVYI (119)	ILLEEQSLI (167)
PFPLVQVEPV (24)	KLLEKAFSI (72)	RPRGPRTAP (120)	LTSISIRPV (168)
SPQSGPAL (25)	KPMPTKVVF (73)	SVQLLEQAIHK (121)	TISECPLLI (169)
TSRLPKIQK (26)	NVNRPLTMK (74)	RTPKNYQHW (122)	TLKLKKIFF (170)
LLDNILQSI (27)	REPYELTVPAL (75)	ALPVALPSL (123)	ILLSNFSSL (171)
RLEVRKVIL (28)	SEAEAAKNAL (76)	SLQILVSSL (124)	LGGAWKAVF (172)
LSWGYFLFK (29)	SLWGQPAEA (77)	ISNKVPKLF (125)	LSASHLSSL (173)
TILPRILTL (30)	SPADHRGYASL (78)	TVIRIAIVNK (126)	AGDIIRLI (174)
EGKIKRNI (31)	SPQSAAAEL (79)	KEIFLELRL (127)	DRGILRNLL (175)
FLASFVEKTVL (32)	SPVVHQSL (80)	TLRSPGSSL (128)	GLRLIHVSL (176)
ILASHNLTV (33)	SPYRTPVL (81)	TVRGDVSSL (129)	GLRLLHVSL (177)
IQLTSVHLL (34)	SVFAGVVG (82)	ALDPLLLRI (130)	LHNEKGLSL (178)
LELISFLPVL (35)	SYSAPAHARL (83)	ISLIVTGLK (131)	LPSFSRPSGII (179)
NFCMLHQSI (36)	THGSEQLHL (84)	KILDVNLRI (132)	LSSRLPLGK (180)
PARPAGPL (37)	TQAPPNVVL (85)	ERVYIRASL (133)	MIGIKRLL (181)
PLPIVPAL (38)	VLVPYEPQV (86)	ILDLESRY (134)	NLKKREIL (182)
SNLIRTGSH (39)	VSFPDVRKV (87)	KTFVQQKTL (135)	RMVAYLQQL (183)
VPAPAQAI (40)	VVFDKSDLAKY (88)	LYIKSLPAL (136)	SPARALPSL (184)
KGHGGPRSW (41)	YSHHSGLEY (89)	VLKEKNASL (137)	TVPGIQRY (185)

[0124]

ITSSAVTTALK (42)	YYLDWIIHY (90)	LGISLTLKY (138)	VSRNYVLLI (186)
LLLPESPSI (43)	SVYKYLKAK (91)	DLLPKKLL (139)	LTVPLSVFW (187)
VILIPLPK (44)	IYQFIMDRF (92)	HSLISIVYL (140)	KLNQAFVLV (188)
AVLLPKPPK (45)	GTLQGIRAW (93)	IAGALRSVL (141)	RLVSSTLLQK (189)
TQVSMAESI (46)	AQKVSVGQAA (94)	IGNPILRVL (142)	LPSHSLLI (190)
LNHLRTSI (47)	LYPSKLTHF (95)	IYAPHIRLS (143)	
NTSHLPLIY (48)	ATQNTIIGK (96)	LRSQILSY (144)	

[0125] 通常,在HLA I类上下文中呈递的肽(诸如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,包含本文限定的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,用于产生本文限定的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(包括天然蛋白质),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包含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包含至少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。

[0126] 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仅包含天然存在的氨基酸。

[0127] 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(具有相似物理化学性质)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可以包含所有L-氨基酸、所有D-氨基酸或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包含所有L-氨基酸。

[0128] 在一种实施方式中,在包含SEQ ID NO:1-190,优选SEQ ID NO:97-154的序列中的一个或由其组成的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,所述其他氨基酸的并入基本上不影响T细胞反应性并且不消除与相关MHC的结合。

[0129] 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是包含以下氨基酸序列中的一个或由其组成的TAP:SEQ ID NO:1-190,优选SEQ ID NO:97-154。

[0130] 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,为了防止降解),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包含以下氨基酸序列中的一个:SEQ ID NO:1-190,优选SEQ ID NO:97-154。在一种实施方式中,TAP由以下氨基酸序列中的一个组成:SEQ ID NO:1-190,优选SEQ ID NO:97-154,即其中 Z^1 和 Z^2 不存在。

[0131] 在另一方面,本公开提供了与HLA-A*01:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:48、67、89、134、151或164,SEQ ID NO:134或151。

[0132] 在另一方面,本公开提供了与HLA-A*02:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:7、11、27、32、33、34、35、4351、52、53、54、61、65、72、77、82、86、104、108、119、123、130、132、146、150、167、168、169、171、183或188,优选SEQ ID NO:104、108、119、123、130、132、146或150。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,见表4),上面识别的TAP可以进一步与HLA-A*02:05、HLA-A*02:06和/或HLA-A*02:07分子结合。

[0133] 在另一方面,本公开提供了与HLA-A*03:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:5、18、19、20、42、44、74、91、105、113、126、131、159、160、180或189,优选SEQ ID NO:105、113、126或131。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,见表4),上面识别的TAP可以进一步与HLA-A*11:01分子结合。

[0134] 在另一方面,本公开提供了与HLA-A*11:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:6、45、96、106、121、149或152,优选SEQ ID NO:106、121、149或152。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,见表4),上面识别的TAP可以进一步与HLA-A*03:01、HLA-A*31:01和/或HLA-A*68:01分子结合。

[0135] 在另一方面,本公开提供了与HLA-A*24:02分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID No:13、36、71、92、95或145,优选SEQ ID NO:145。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,见表4),上面识别的TAP可以进一步与HLA-A*23:01分子结合。

[0136] 在另一方面,本公开提供了与HLA-A*26:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含SEQ ID NO:59或SEQ ID NO:185的序列或由其组成。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,见表4),上面识别的TAP可以进一步与HLA-A*25:01和/或HLA-A*66:01分子结合。

[0137] 在另一方面,本公开提供了与HLA-A*29:02分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID No:88、99或138,优选SEQ ID NO:99或138。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,见表4),上面识别的TAP可以进一步与HLA-A*30:02和/或HLA-B*15:02分子结合。

[0138] 在另一方面,本公开提供了与HLA-A*30:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含SEQ ID NO:26、29或165的序列或由其组成。

[0139] 在另一方面,本公开提供了与HLA-A*68:02分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:50或SEQ ID No:148,优选SEQ ID NO:148。

[0140] 在另一方面,本公开提供了与HLA-B*07:02分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:8、25、40、55、56、73、78、79、80、81、98、100、107、110、111、118、120、128、129、157、161、163、179或184,优选SEQ ID NO:98、100、107、110、111、118、120、128或129。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,见表4),上面识别的TAP可以进一步与HLA-B*35:02、HLA-B*35:03、HLA-B*55:01和/或HLA-B*56:01分子结合。

[0141] 在另一方面,本公开提供了与HLA-B*08:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:4、14、15、17、SEQ ID NO:21、22、23、28、31、47、49、101、103、114、137、139、147、156、170、181或182,优选地SEQ ID NO:101、103、114、137、139或147。

[0142] 在另一方面,本公开提供了与HLA-B*14:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:2、39、58、66、124、133、136、

141、162、175、176或177,优选SEQ ID NO:124、133、136或141。

[0143] 在另一方面,本公开提供了与HLA-B*15:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含SEQ ID NO:10、57、62或94的序列或由其组成。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,见表4),上面识别的TAP可以进一步与HLA-B*15:02、HLA-B*15:03和/或HLA-B*46:01分子结合。

[0144] 在另一方面,本公开提供了与HLA-B*27:05分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含SEQ ID No:1或144,优选SEQ ID NO:144的序列或由其组成。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,见表4),上面识别的TAP可以进一步与HLA-B*27:02分子结合。

[0145] 在另一方面,本公开提供了与HLA-B*38:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含SEQ ID NO:4、64或84的序列或由其组成。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,见表4),上面识别的TAP可以进一步与HLA-B*39:01分子结合。

[0146] 在另一方面,本公开提供了与HLA-B*40:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含SEQ ID NO:75或76的序列或由其组成。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,见表4),上面识别的TAP可以进一步与HLA-B*18:01、HLA-B*40:02、HLA-B*41:02、HLA-B*44:02、HLA-B*44:03和/或HLA-B*45:01分子结合。

[0147] 在另一方面,本公开提供了与HLA-B*44:03分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含SEQ ID NO:127的序列或由其组成。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,见表4),上面识别的TAP可以进一步与HLA-B*18:01、HLA-B*40:01、HLA-B*40:02、HLA-B*41:02、HLA-B*44:02和/或HLA-B*45:01分子结合。

[0148] 在另一方面,本公开提供了与HLA-B*51:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID No:12、24、38、68、115、116或190,优选SEQ ID NO:115或116。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,见表4),上面识别的TAP可以进一步与HLA-B*35:02、HLA-B*35:03、HLA-B*52:01、HLA-B*53:01、HLA-B*55:01和/或HLA-B*56:01分子结合。

[0149] 在另一方面,本公开提供了与HLA-B*57:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:3、16、41、63、69、93、122、125、135、153或183,优选122、125、135或153。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,见表4),上面识别的TAP可以进一步与HLA-A*32:01和/或HLA-B*58:01分子结合。

[0150] 在另一方面,本公开提供了与HLA-B*57:03分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含SEQ ID NO:60或172的序列或由其组成。

[0151] 在另一方面,本公开提供了与HLA-C*03:03分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID No:37、112或140,优选SEQ ID NO:112或140。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,见表4),上面识别的TAP可以进一步与HLA-B*46:01、HLA-C*03:02、HLA-C*03:04、HLA-C*08:01、HLA-C*08:02、HLA-C*12:02、HLA-C*12:03、HLA-C*15:02和/或HLA-C*16:01分子结合。

[0152] 在另一方面,本公开提供了与HLA-C*05:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含SEQ ID NO:150的序列或由其组成。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,见表4),上面识别的TAP可以进一步与HLA-C*08:01和/或HLA-C*08:02分子结合。

[0153] 在另一方面,本公开提供了与HLA-C*06:02分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:9、70、87、142、154或166,优选SEQ ID NO:142或154。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,见表4),上面识别的TAP可以进一步与HLA-B*27:02、HLA-C*07:01和/或HLA-C*07:02分子结合。

[0154] 在另一方面,本公开提供了与HLA-C*07:01分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:142、143、155、178或186,优选SEQ ID NO:142或143。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,见表4),上面识别的TAP可以进一步与HLA-B*27:02、HLA-C*07:01、HLA-C*07:02和/或HLA-C*14:02分子结合。

[0155] 在另一方面,本公开提供了与HLA-C*07:02分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:30、83、85、90、109、117或158,优选SEQ ID NO:109或117。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,见表4),上面识别的TAP可以进一步与HLA-B*27:02、HLA-C*07:01、HLA-C*07:02和/或HLA-C*14:02分子结合。

[0156] 在另一方面,本公开提供了与HLA-C*08:02分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含以下序列或由其组成:SEQ ID NO:97、102或174,优选SEQ ID NO:97或102。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,见表4),上面识别的TAP可以进一步与HLA-C*03:03、HLA-C*03:04、HLA-C*05:01、HLA-C*08:01和/或HLA-C*15:02分子结合。

[0157] 在另一方面,本公开提供了与HLA-C*12:03分子结合的白血病TAP(或肿瘤特异性肽),优选AML TAP,其包含SEQ ID NO:173的序列或由其组成。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,见表4),上面识别的TAP可以进一步与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。

[0158] 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。

[0159] 本公开的TAP可以通过在包含编码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。可选地,MiHA肽可以通过片段缩合来制备,例如,如以下所述: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的盐与处于其体内状态的肽显著不同,因为体内产生的肽不是盐。肽的非天然盐形式可以调节肽的溶解度,特别是在包含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。

[0160] 在一种实施方式中,本文提及的TAP是基本上纯的。当化合物与自然伴随它的组分分离时,它是“基本上纯的”。通常,当化合物按重量计占样本中总材料的至少60%,更通常75%、80%或85%,优选超过90%,且更优选超过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。

[0161] 在另一方面,本公开进一步提供了编码本文提及的TAP或肿瘤抗原前体-肽的核酸(分离的)。在一种实施方式中,核酸包含约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。如本文所用,“分离的”是指与存在于该分子的天然环境中的其他组分或天然存在的来源大分子(例如,包括其他核酸、蛋白质、脂质、糖等)分离的肽或核酸分子。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。本公开的核酸可用于重组表达本公开的TAP,并且可以被包含在载体或质粒中,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,本公开提供了包含编码本公开的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或蛋白质。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)或真核细胞(例如,昆虫细胞、酵母或哺乳动物细胞)。载体或质粒含有插入的编码序列转录和翻译所必需的元件,也可以含有其他组分(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,该表达载体包含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”是指组分的并列,特别是核苷酸序列,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

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的非编码核苷酸序列。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'下游调节元件(诸如启动子序列,诸如CMV、PGK和EF1a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多核苷酸序列的表达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)。表达载体的实例是用于在哺乳动物细胞中表达的Lenti-X™双顺反子表达系统(Neo)载体(Clontech)、pCIneo载体(Promega);pLenti4/V5-DEST™、pLenti6/V5-DEST™和pLenti6.2N5-GW/lacZ(Invitrogen),用于哺乳动物细胞中慢病毒介导的基因转移和表达。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。病毒载体可以是源自逆转录病毒、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质的编码序列来代替非必需的病毒基因。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)包含在脂质体或任何其他合适的载体中。

[0164] 在另一方面,本公开提供了一种MHC I类分子,其包含(即呈递或结合至)一个或多个SEQ ID NO:1-190,优选SEQ ID NO:97-154的TAP。在一种实施方式中,MHC I类分子是HLA-A1分子,在进一步的实施方式中是HLA-A*01:01分子。在另一实施方式中,MHC I类分子是HLA-A2分子,在进一步的实施方式中是HLA-A*02:01分子。在另一实施方式中,MHC I类分子是HLA-A3分子,在进一步的实施方式中是HLA-A*03:01分子。在另一实施方式中,MHC I类分子是HLA-A11分子,在进一步的实施方式中是HLA-A*11:01分子。在另一实施方式中,MHC I类分子是HLA-A24分子,在进一步的实施方式中是HLA-A*24:02分子。在另一实施方式中,MHC I类分子是HLA-A26分子,在进一步的实施方式中是HLA-A*26:01分子。在另一实施方式中,MHC I类分子是HLA-A29分子,在进一步的实施方式中是HLA-A*29:02分子。在另一实施方式中,MHC I类分子是HLA-A30分子,在进一步的实施方式中是HLA-A*30:01分子。在另一实施方式中,MHC I类分子是HLA-A68分子,在进一步的实施方式中是HLA-A*68:02分子。在

另一实施方式中,MHC I类分子是HLA-B07分子,在进一步的实施方式中是HLA-B*07:02分子。在另一实施方式中,MHC I类分子是HLA-B08分子,在进一步的实施方式中是HLA-B*08:01分子。在另一实施方式中,MHC I类分子是HLA-B14分子,在进一步的实施方式中是HLA-B*14:01分子。在另一实施方式中,MHC I类分子是HLA-B15分子,在进一步的实施方式中是HLA-B*15:01分子。在另一实施方式中,MHC I类分子是HLA-B27分子,在进一步的实施方式中是HLA-B*27:05分子。在另一实施方式中,MHC I类分子是HLA-B38分子,在进一步的实施方式中是HLA-B*38:01分子。在另一实施方式中,MHC I类分子是HLA-B40分子,在进一步的实施方式中是HLA-B*40:01分子。在另一实施方式中,MHC I类分子是HLA-B44分子,在进一步的实施方式中是HLA-B*44:02或HLA-B*44:03分子。在另一实施方式中,MHC I类分子是HLA-B57分子,在进一步的实施方式中是HLA-B*57:01或HLA-B*57:03分子。在另一实施方式中,MHC I类分子是HLA-C03分子,在进一步的实施方式中是HLA-C*03:03分子。在另一实施方式中,MHC I类分子是HLA-C04分子,在进一步的实施方式中是HLA-C*04:01分子。在另一实施方式中,MHC I类分子是HLA-C05分子,在进一步的实施方式中是HLA-C*05:01分子。在另一实施方式中,MHC I类分子是HLA-C06分子,在进一步的实施方式中是HLA-C*06:02分子。在另一实施方式中,MHC I类分子是HLA-C07分子,在进一步的实施方式中是HLA-C*07:01或HLA-C*07:02分子。在另一实施方式中,MHC I类分子是HLA-C08分子,在进一步的实施方式中是HLA-C*08:02分子。在另一实施方式中,MHC I类分子是HLA-C12分子,在进一步的实施方式中是HLA-C*12:03分子。

[0165] 在一种实施方式中,TAP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/袋)。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,该融合蛋白包含与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)-肽复合物是多聚化的。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 Immunology2005,17:428-433)。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)。

[0166] 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,其包含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。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一个的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)。可以通过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、脂质转染、电穿孔、显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。

[0167] 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,或可选地,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端Lys-Asp-Glu-Leu序列(参见Wang等人,Eur J Immunol.2004Dec;34(12):3582-94)。

[0168] 在另一方面,本公开提供了一种组合物或肽组合/池,其包含本文限定的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,组合物包含本文限定的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)。本公开涵盖包含本文限定的TAP的任何组合/亚组合的组合物。在另一实施方式中,组合或池可以包含一种或多种已知的肿瘤抗原。

[0169] 因此,在另一方面,本公开提供了一种组合物,其包含本文限定的TAP中的任一种或其任何组合和表达MHC I类分子(例如,上文公开的等位基因中的一个的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质抗原以被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,其中向受试者施用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,例如使用包含TAP序列的较长肽/多肽(包括天然蛋白质),然后在APC内对其进行处理(例如,通过蛋白酶)以在细胞表面产生TAP/MHC I类复合物。在用TAP负载APC并允许APC呈递TAP后,APC可以作为疫苗被施用于受试者。例如,离体施用可以包括以下步骤:(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复合物;以及(c)向需要治疗的第二受试者施用负载有肽的APC。

[0170] 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)。可选地,根据本公开,提供了本文所述的TAP(或其组合)用于制造用于诱导抗原呈递细胞的组合物(例如,药物组合物)的用途。此外,本公开提供了生产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载体混合或配制的步骤。表达负载有本文限定的TAP中的任一种或任何组合的MHC I类分子(例如,HLA-A1、HLA-A2、HLA-A3、HLA-A11、HLA-A24、HLA-A25、HLA-A29、HLA-A32、HLA-B07、HLA-B08、HLA-B14、HLA-B15、HLA-B18、HLA-B39、HLA-B40、HLA-B44、HLA-C03、HLA-C04、HLA-C05、HLA-C06、HLA-C07、HLA-C12或HLA-C14分子)的细胞,诸如APC,可用于刺激/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,其包含本文限定的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,包含CD8⁺T淋巴细胞的细胞群)。

[0171] 在一种实施方式中,组合物还包含缓冲剂、赋形剂、载体、稀释剂和/或介质(例如,培养基)。在进一步的实施方式中,缓冲剂、赋形剂、载体、稀释剂和/或介质是药学上可接受的缓冲剂、赋形剂、载体、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、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,不干扰活性成分的生物活性的有效性并且对受试者无毒。此类基质和试剂用于药物活性物质的用途是本领域熟知的(Rowe等人, Handbook of pharmaceutical excipients, 2003, 第4版, Pharmaceutical Press, London UK)。除非任何常规基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,缓冲液、赋形剂、载体和/或介质是非天然存在的缓冲液、赋形剂、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)包含在脂质体(例如阳离子脂质体)或其他合适的载体内或与其复合(参见,例如, Vitor MT等人, Recent Pat Drug Deliv Formul. 2013 Aug; 7(2): 99-110)。

[0172] 在另一方面,本公开提供了组合物,其包含本文限定的TAP中的任一种或任何组合(或编码所述肽的核酸)以及缓冲剂、赋形剂、载体、稀释剂和/或介质中的一种或多种。对于包含细胞(例如, APC、T淋巴细胞)的组合物,该组合物包含实现维持活细胞的合适介质。此类介质的典型实例包括盐溶液、Earl's平衡盐溶液(Life Technologies®)或PlasmaLyte®(Baxter International®)。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”是指包含一种或多种TAP或疫苗载体的组合物或制剂,并且当施用于受试者时,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施用,例如通过粘膜(例如,眼、鼻、肺、口腔、胃、肠、直肠、阴道或泌尿道)表面,通过肠胃外(例如,皮下、皮内、肌肉、静脉内或腹腔)途径,或外用(例如,通过经皮递送系统,诸如贴剂)。在一种实施方式中,TAP(或其组合)与载体蛋白缀合(缀合疫苗)以增加TAP的免疫原性。因此,本公开提供了包含TAP(或其组合)、或编码TAP或其组合的核酸以及载体蛋白的组合物(缀合物)。例如,TAP或核酸可以与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. 2013 Aug 12; 14(8): 2798-806)缀合或复合。在一种实施方式中,免疫原性组合物或疫苗还包含佐剂。“佐剂”是指当添加至免疫原性试剂诸如抗原(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试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的制剂,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(包含流感血凝素的单层脂质体负载体)、AS04([SBAS4]铝盐与MPL)、ISCOMS(皂苷和脂质的结构化复合物)、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 (基因改性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 (可以作为蛋白质或编码质粒施用的细胞因子)、Immudaptin (C3d串联阵列), 和/或 (6) 惰性负载体, 诸如金颗粒等。

[0173] 在一种实施方式中, TAP或包含其的组合物是冻干形式。在另一实施方式中, TAP或包含其的组合物在液体组合物中。在进一步的实施方式中, 在组合物中TAP的浓度为约0.01 $\mu\text{g/mL}$ 至约100 $\mu\text{g/mL}$ 。在进一步的实施方式中, 组合物中TAP的浓度为约0.2 $\mu\text{g/mL}$ 至约50 $\mu\text{g/mL}$, 约0.5 $\mu\text{g/mL}$ 至约10、20、30、40或50 $\mu\text{g/mL}$, 约1 $\mu\text{g/mL}$ 至约10 $\mu\text{g/mL}$, 或约2 $\mu\text{g/mL}$ 。

[0174] 如本文所述, 表达负载有或结合本文限定的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复合物相互作用或结合的T细胞受体 (TCR) 分子, 以及编码此类TCR分子的核酸分子, 以及包含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活细胞表面上。

[0175] 在一种实施方式中, 根据本公开的抗白血病 (例如, 抗AML) TCR包含包含互补决定区3 (CDR3) 的TCR β (B) 链, 该互补决定区3 (CDR3) 包含SEQ ID NO:191-219中列出的氨基酸序列中的一个。

[0176] 在一种实施方式中, TCR对以下TAP中的一个或多个具有特异性: SLLSGLLRA、ALPVALPSL、ALDPLLLRI IASPIALL和/或SLDLLPLSI, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:191-199中列出的氨基酸序列中的一个。在一种实施方式中, TCR对TAP SLLSGLLRA具有特异性, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:191-199中列出的氨基酸序列中的一个。在一种实施方式中, TCR对TAP ALPVALPSL具有特异性, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:191-199中列出的氨基酸序列中的一个。在一种实施方式中, TCR对TAP ALDPLLLRI具有特异性, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:191-199中列出的氨基酸序列中的一个。在一种实施方式中, TCR对TAP IASPIALL具有特异性, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:191-199中列出的氨基酸序列中的一个。在一种实施方式中, TCR对TAP SLDLLPLSI具有特异性, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:191-199中列出的氨基酸序列中的一个。

[0177] 在另一实施方式中, TCR对以下TAP中的一个或多个具有特异性: LTDRIYLTL、VLFGGKVSGA、LGISLTLKY、FNVALNARY和/或TLNQGINVYI, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:200-209中列出的氨基酸序列中的一个。在一种实施方式中, TCR对TAP LTDRIYLTL具有特异性, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:200-209中列出的氨基酸序列中的一个。在一种实施方式中, TCR对TAP VLFGGKVSGA具有特异性, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:200-209中列出的氨基酸序列中的一个。在一种实施方式中, TCR对TAP LGISLTLKY具有特异性, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:200-209中列出的氨基酸序列中的一个。在一种实施方式中, TCR对TAP FNVALNARY具有特异性, 并且包含包含CDR3的TCR β 链, 该CDR3包含SEQ ID NO:200-209中列出的氨基酸序列中的一个。在一种实施方式中, TCR对TAP TLNQGINVYI具有特异性, 并且包

含包含CDR3的TCR β 链,该CDR3包含SEQ ID NO:200-209中列出的氨基酸序列中的一个。

[0178] 在另一实施方式中,TCR对以下TAP中的一个或多个具有特异性:LRSQILSY、KILDVNLRI、HSLISIVYL、KLQDKEIGL和/或AQDIILQAV,并且包含包含CDR3的TCR β 链,该CDR3包含SEQ ID NO:210-219中列出的氨基酸序列中的一个。在一种实施方式中,TCR对TAP LRSQILSY具有特异性,并且包含包含CDR3的TCR β 链,该CDR3包含SEQ ID NO:210-219中列出的氨基酸序列中的一个。在一种实施方式中,TCR对TAP KILDVNLRI具有特异性,并且包含包含CDR3的TCR β 链,该CDR3包含SEQ ID NO:210-219中列出的氨基酸序列中的一个。在一种实施方式中,TCR对TAP HSLISIVYL具有特异性,并且包含包含CDR3的TCR β 链,该CDR3包含SEQ ID NO:210-219中列出的氨基酸序列中的一个。在一种实施方式中,TCR对TAP KLQDKEIGL具有特异性,并且包含包含CDR3的TCR β 链,该CDR3包含SEQ ID NO:210-219中列出的氨基酸序列中的一个。在一种实施方式中,TCR对TAP AQDIILQAV具有特异性,并且包含包含CDR3的TCR β 链,该CDR3包含SEQ ID NO:210-219中列出的氨基酸序列中的一个。

[0179] 在一种实施方式中,根据本公开的TCR识别与HLA-A*02:01、HLA-A*29:02、HLA-B*15:01、HLA-B*27:05、HLA-C*01:02和/或HLA-C*03:04分子结合的上述TAP中的一种或多种。在一种实施方式中,根据本公开的TCR识别与HLA-A*02:01分子结合的上述TAP中的一种或多种。在一种实施方式中,根据本公开的TCR识别与HLA-A*29:02分子结合的上述TAP中的一种或多种。在一种实施方式中,根据本公开的TCR识别与HLA-B*15:01分子结合的上述TAP中的一种或多种。在一种实施方式中,根据本公开的TCR识别与HLA-B*27:05分子结合的上述TAP中的一种或多种。在一种实施方式中,根据本公开的TCR识别与HLA-C*01:02分子结合的上述TAP中的一种或多种。在一种实施方式中,根据本公开的TCR识别与HLA-C*03:04分子结合的上述TAP中的一种或多种。

[0180] 如本文所用,术语TCR是指具有可变结合结构域、恒定结构域、跨膜区和短胞质尾的免疫球蛋白超家族成员;参见,例如,Janeway等人,Immunobiology:The Immune System in Health and Disease,第3版,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,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复合物缔合。

[0181] 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肽复合物的新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)被转化以表达一种或多种识别TAP的TCR,并将转化的细胞施用于患者(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)被转化以表达一种或多种识别TAP的TCR,并将转化的细胞施用于接受者(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。

[0182] 在另一实施方式中,本公开提供了编码上述TCR的核酸。在进一步的实施方式中,核酸存在于载体(诸如上述载体)中。

[0183] 在又一实施方式中,提供了肿瘤抗原特异性TCR在制造用于治疗癌症(白血病,诸如AML)的自体或同种异体细胞中的用途。

[0184] 在一些实施方式中,在用同种异体干细胞移植(ASCT)、同种异体淋巴细胞输注或自体淋巴细胞输注治疗之前或之后治疗用本公开的组合物(例如,药物组合物)治疗的患者。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类分子复合物的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,本说明书提供了至少两种用于产生本公开的CD8⁺T淋巴细胞的方法,其包括在有利于触发T细胞激活和扩增的条件下使未分化的淋巴细胞与TAP/MHC I类分子复合物(通常在细胞诸如APC表面表达)接触的步骤,这可以在体外或体内(即在施用了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的组合或池,可能产生能够识别多种TAP的CD8⁺T淋巴细胞群。可选地,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复合物特异性结合的TCR(更特别地 α 和 β 链)的一种或多种核酸(基因)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;或从对肽/MHC分子复合物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),诸如磷酸钙-DNA共沉淀、DEAE-葡聚糖介导的转染、聚凝胺介导的转染、电穿孔、显微注射、脂质体融合、脂质转染、原生质体融合、逆转录病毒感染和基因枪法,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。

[0185] 本公开提供了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,诸如患有白血病(例如,AML)的受试者分离的免疫效应细胞,使得免疫效应细胞表达一种或多种如本文所述的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基因修饰。然后将此类细胞直接重新施用于个体中。在进一步的实施方式中,免疫效应细胞在经基因修饰以表达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。

[0186] 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,包括外周血单个核细胞(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,使用本领域技术人员已知的任何数量的技术,诸如FICOLL™分离,T细胞可以获自从受试者收集的单位血液。在一种实施方式中,来自个体循环血液的细胞通过单采术获得。单采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,可以洗涤通过单采术收集的细胞以移除血浆级分并将细胞置于适当的缓冲液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。在可选的实施方式中,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液或其他含或不含缓冲液的盐水溶液中。在某些实施方式中,可以在细胞直接重悬培养基中移除单采样本的非期望组分。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PERCOLL™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。一种用于本文的方法是通过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。

[0187] 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,其包含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复合物的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,“富集”是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(即未经过特异性T淋巴细胞的离体扩增步骤)显著更高。在进一步的实施方式中,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)复合物的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%。

[0188] 本公开进一步涉及包含上述免疫细胞(CD8⁺T淋巴细胞)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含一种或多种药学上可接受的赋形剂和/或佐剂,如上所述。

[0189] 本公开进一步涉及根据本公开的任何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、核酸、表达载体、T细胞受体、细胞(例如,T淋巴细胞、APC)和/或组合物(例如,疫苗组合物)或其任何组合,用于治疗癌症的用途,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的抗肿瘤T细胞应答。

[0190] 在另一方面,本公开提供了本文所述的TAP(SEQ ID NO:1-190,优选SEQ ID NO:97-154)或其组合(例如肽池)作为用于治疗受试者的癌症的疫苗的用途。本公开还提供了本文所述的TAP或其组合(例如肽池),用作用于治疗受试者癌症的疫苗。在一种实施方式中,受试者是TAP特异性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),所述方法包括向有需要的受试者施用(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复合物(在细胞(诸如APC)的表面表达)的CD8⁺T淋巴细胞。在一种实施方式中,该方法进一步包括,在施用/输注所述CD8⁺T淋巴细胞之后,向所述受试者施用有效量的TAP或其组合和/或表达负载有TAP的MHC I类分子的细胞(例如,APC,诸如树突状细胞)。在又一实施方式中,该方法包括向有需要的受试者施用治疗有效量的负载有一种或多种TAP的树突状细胞。在又一实施方式中,该方法包括向有需要的患者施用治疗有效量的同种异体或自体细胞,该细胞表达与MHC I类分子呈递的TAP结合的重组TCR。

[0191] 在另一方面,本公开提供了识别一种或多种负载有(呈递)TAP或其组合的MHC I类分子的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)的用途。在另一方面,本公开提供了识别一种或多种负载有(呈递)TAP或其组合的MHC I类分子的CD8⁺T淋巴细胞用于制备/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)的药物的用途。在另一方面,本公开提供了识别一种或多种负载有(呈递)TAP或其组合的MHC I类分子的CD8⁺T淋巴细胞(细胞毒性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)的用途。在进一步的实施方式中,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/或表达一种或多种负载有(呈递)TAP的MHC I类分子的细胞(例如,APC)。

[0192] 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或其组合的人I类MHC分子的肿瘤细胞(白血病细胞,AML细胞)的免疫应答的方法,该方法包括施用特异性识别负载有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。

[0193] 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,并施用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-A1*01和HLA-C05*01,则可以在患者中施用或使用以下TAP的任何组合:(i) SEQ ID NO:48、67、89、134、151和/或164 (结合HLA-A1*01),和/或(iii) SEQ ID NO:150 (结合HLA-C05*01)。

[0194] 在一种实施方式中,癌症是血液癌症,优选白血病,诸如急性淋巴细胞白血病(ALL)、急性髓系白血病(AML)、慢性髓系白血病(CML)、毛细胞白血病(HCL)和骨髓增生异常综合征(MDS)。在一种实施方式中,白血病是AML。通过本文所述的方法和用途治疗的AML可以是任何类型或亚型(例如,低风险、中风险或高风险AML),例如AML伴有遗传异常,诸如AML伴有8号和21号染色体之间易位[t(8;21)]、AML伴有16号染色体易位或倒位[t(16;16)或inv(16)]、AML伴有PML-RARA融合基因、AML伴有9号和11号染色体之间易位[t(9;11)]、AML伴有6号和9号染色体之间易位[t(6;9)]、AML伴有3号染色体易位或倒位[t(3;3)或inv(3)]、AML(巨核细胞)伴有1号和22号染色体之间易位[t(1;22)]、AML伴有BCR-ABL1(BCR-ABL)融合基因、AML伴有NPM1基因突变、AML伴有CEBPA基因双等位基因突变、AML伴有RUNX1基因突变、AML伴有ASX1基因突变、AML伴有IDH1和/或IDH2基因突变、AML伴有FLT3基因突变、AML伴有骨髓增生异常相关变化以及与既往化疗或放射疗法相关的AML。

[0195] 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/或组合物或其任何组合,可以与一种或多种另外的活性剂或治疗癌症的疗法组合使用以治疗癌症,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)、过渡金属配合物、蛋白酶体抑制剂、抗代谢物(诸如核苷类似物)、烷化剂、铂类药物、蒽环类抗生素、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)、格尔德霉素或其衍生物(诸如17-AAG)、手术、放射疗法、免疫检查点抑制剂(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、基于细胞的疗法(例如,CAR T细胞、CAR NK细胞)、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肽、核酸、表达载体、T细胞受体、细胞(例如,T淋巴细胞、APC)和/或组合物与免疫检查点抑制剂组合施用/使用。在一种实施方式中,TAP、核酸、表达载体、T细胞受体、细胞(例如,T淋巴细胞、APC)和/或根据本公开的组合物与一种或多种用于治疗AML的化疗药物组合施用/使用,或与其他AML疗法(例如干细胞/骨髓移植)组合施用/使用。

[0196] 可以在施用根据本公开的TAP、核酸、表达载体、T细胞受体、细胞(例如,T淋巴细胞、APC)和/或组合物之前、同时或之后施用另外的疗法。

[0197] 实施本发明的方式

[0198] 通过以下非限制性实施例进一步详细说明本发明。

[0199] 实施例1:材料和方法

[0200] AML试样

[0201] 诊断性AML样本(DMSO冷冻白血病母细胞冷冻管)获自Banque de cellules leucé

miques du Québec计划(BCLQ,bclq.org)。样本技术和临床特征在表1中提供。将每个AML样本的一亿个细胞(14H124除外,见下文)解冻(在37℃水浴中1min)并重悬于48ml的4℃ PBS中。将200万个细胞(1ml)沉淀并重悬于1ml Trizol中用于RNA测序,而将剩余的9800万个细胞沉淀并在液氮中速冻用于质谱分析。

[0202] 表1:本研究中用于识别TSA的19个AML试样的生物学和临床特征。

[0203]

ID BCLQ	HLA 分型 (optitype)	WHO 2008	FAB	组织	母细胞%	RNA-seq 读数计数
05H14 3	A*02:01, A*34:02, B*08:01, B*57:01, C*07:01, C*06:02	急性成单核 细胞和单核 细胞白血病	M5A	外周血	85	2,14E+08
05H14 9	A*24:02, A*68:02, B*53:01, B*18:01, C*04:01, C*05:01	急性髓系白 血病, NOS	M1	外周血	80	1,19E+08
07H06 0	A*01:01, A*29:02, B*08:01, B*56:01, C*07:01, C*01:02	AML 未成熟 型	M1	外周血	90	2,04E+08
07H06 3	A*30:01, A*01:01, B*13:02, B*07:02, C*06:02, C*07:02	治疗相关的 髓系肿瘤	M5A	骨髓	96	1,92E+08
07H12 2	A*02:01, A*02:01, B*57:01, B*57:01, C*06:02, C*06:02	AML 伴有骨 髓增生异常 相关变化	NC	外周血	96	1,84E+08
07H14 1	A*02:01, A*26:01, B*38:01, B*44:02, C*05:01, C*12:03	AML 未成熟 型	M1	外周血	75	1,77E+08

[0204]

08H03 9	A*02:01, A*01:01, B*40:01, B*14:02, C*03:04, C*08:02	AML 微分化 型	M0	外周血	96	1,82E+08
08H05 3	A*01:01, A*11:01, B*14:02, B*15:01, C*03:03, C*03:03	AML 未成熟 型	M1	外周血	95	1,74E+08
10H00 5	A*11:01, A*11:01, B*13:01, B*13:01, C*07:02, C*03:04	急性粒单核 细胞白血病	M4	骨髓	71	1,84E+08
11H00 8	A*29:02, A*03:01, B*44:03, B*57:01, C*06:02, C*16:01	AML 伴有骨 髓增生异常 相关变化	NC	外周血	75	1,76E+08
11H03 5	A*03:01, A*24:02, B*07:02, B*07:02, C*07:02, C*07:02	AML 未成熟 型	M1	外周血	93	1,80E+08
12H17 2	A*24:02, A*02:01, B*15:01, B*15:01, C*03:03, C*03:03	治疗相关的 髓系肿瘤	M1	外周血	85	1,91E+08
14H12 4	A*02:01, A*26:01, B*44:05, B*27:05, C*02:02, C*02:02	AML 成熟型	M2	外周血	86	1,23E+08

[0205]

15H01 3	A*30:02, A*02:01, B*39:05, B*57:03, C*18:01, C*07:02	AML 未成熟 型	M1	外周血	90	1,61E+08
15H02 3	A*02:01, A*03:01, B*51:01, B*44:03, C*01:02, C*02:02	AML 未成熟 型	M1	外周血	90	1,79E+08
15H06 3	A*02:01, A*01:01, B*08:01, B*40:01, C*07:01, C*03:04	AML 伴有骨 髓增生异常 相关变化	NC	外周血	85	1,65E+08
15H08 0	A*02:01, A*26:01, B*44:05, B*27:05, C*02:02, C*02:02	急性成单核 细胞和单核 细胞白血病	M5A	外周血	70	1,95E+08
16H12 3	A*24:02, A*03:01, B*15:01, B*07:02, C*07:02, C*03:03	AML 未成熟 型	M1	外周血	88	1,66E+08
16H14 5	A*02:01, A*02:01, B*27:02, B*55:01, C*03:03, C*02:02	AML 微分化 型	M0	外周血	88	1,85E+08

[0206]

ID BCLQ	独特 k-mer (33)计数	mTEC 耗 竭中使用的 发生阈 值	发生过滤 后的 K-mer 计 数	mTEC 过滤后 的 K-mer 计数	mTEC+MP C 耗竭中使 用的发生阈 值	发生过 滤后的 K-mer 计数
05H14 3	1,44E+0 9	10	1,26E+08	3,01E+07	3	3,38E+0 8
05H14 9	8,89E+0 8	12	7,64E+07	2,33E+07	5	1,78E+0 8
07H06 0	1,24E+0 9	8	1,11E+08	2,02E+07	3	2,76E+0 8
07H06 3	9,55E+0 8	6	9,26E+07	1,62E+07	3	1,80E+0 8
07H12 2	9,67E+0 8	7	8,35E+07	1,85E+07	3	2,03E+0 8
07H14 1	1,08E+0 9	10	8,81E+07	1,55E+07	3	2,55E+0 8
08H03 9	1,10E+0 9	8	9,72E+07	1,85E+07	3	2,47E+0 8
08H05 3	1,14E+0 9	8	9,08E+07	1,56E+07	3	2,40E+0 8
10H00 5	1,01E+0 9	6	8,63E+07	1,56E+07	3	1,81E+0 8
11H00 8	8,30E+0 8	5	9,73E+07	1,79E+07	3	1,68E+0 8
11H03 5	1,07E+0 9	7	1,06E+08	1,86E+07	3	2,39E+0 8
12H17 2	1,12E+0 9	6	1,07E+08	2,17E+07	3	2,18E+0 8
14H12 4	7,77E+0 8	4	1,03E+08	1,76E+07	3	1,35E+0 8
15H01 3	8,65E+0 8	5	1,03E+08	1,86E+07	3	1,73E+0 8
15H02 3	1,21E+0 9	8	1,12E+08	2,28E+07	3	2,75E+0 8
15H06 3	1,02E+0 9	8	9,79E+07	2,16E+07	3	2,29E+0 8
15H08 0	1,13E+0 9	7	1,11E+08	2,23E+07	3	2,44E+0 8
16H12 3	1,06E+0 9	8	1,01E+08	1,77E+07	3	2,50E+0 8
16H14 5	1,19E+0 9	8	1,08E+08	2,12E+07	3	2,65E+0 8

ID BCLQ	mTEC+MPC 过滤后的 K-mer 计数	不同的 MAP 计数			
		mTECs depl.	ERE	mTECs+MPCs depl.	差异 k-mer 表达
05H14 3	2,53E+07	2894	2358	2897	2877
05H14 9	2,47E+07	542	519	545	539
07H06 0	1,83E+07	986	957	983	973
07H06 3	8,14E+06	1784	1777	1797	1715
07H12 2	2,53E+07	2940	2702	2911	2878
07H14 1	1,37E+07	1662	1554	1647	1544
08H03 9	2,27E+07	2594	2420	2689	2670
08H05 3	1,81E+07	317	270	328	292
[0207] 10H00 5	1,06E+07	220	136	280	159
11H00 8	9,91E+06	1163	1054	1175	1072
11H03 5	1,53E+07	1928	1899	1986	1933
12H17 2	1,59E+07	2865	2916	2861	2851
14H12 4	3,63E+06	2160	2093	2249	2116
15H01 3	8,86E+06	1761	1686	1756	1748
15H02 3	1,68E+07	1794	1733	1799	1732
15H06 3	1,25E+07	1277	1128	1285	1190
15H08 0	1,23E+07	797	722	796	760
16H12 3	1,80E+07	1424	1383	1467	1391
16H14 5	1,71E+07	755	704	765	742

[0208] NC:无法按FAB标准分类。基于每个样本的RNA-Seq数据通过optitype确定HLA。临床数据由Banque de cellules leucémiques du Québec计划(BCLQ,bclq.org)提供。

[0209] 其他数据来源

[0210] 人mTEC样本已根据我们团队先前研究的需要进行制备和测序(#GSE127825和#GSE127826) (Larouche等人,2020;Laumont等人,2018),或已由其他人公开(E-MTAB-7383) (Fergusson等人,2018)。只有我们小组先前用于TSA发现的六个mTEC样本已用于k-mer耗竭方法(Laumont等人,2018)。用作主要正常对照的11个MPC样本已由IRIC基因组平台测序,并由Leucegene小组先前公开(#GSE98310,#GSE51984)。本研究中使用的所有其他正常样本均

已从dbGap (www.ncbi.nlm.nih.gov/gap/), Arrayexpress (www.ebi.ac.uk/arrayexpress/) 或GEO (www.ncbi.nlm.nih.gov/geo/) 下载。Leucegene全队列437个RNA测序的AML样本用于研究发现的TSA^{hi}的临床意义。RNA测序数据之前已公开,并可单独获得(#GSE49642、#GSE52656、#GSE62190、#GSE66917、#GSE67039) (Lavallee等人,2015;Macrae等人,2013;Pabst等人,2016)。分类的LSC和母细胞的RNA-Seq数据已在其他地方公开,并从#GSE74246 (Corces等人,2016) 获得。复发前和复发后AML母细胞的RNA-seq数据(匹配样本)已在其他地方公开 (Toffalori等人,2019),这些样本的HLA类型由Luca Vago博士友情提供。从外部来源获得的所有数据均使用STAR v2.5.1b在GRCh38基因组上进行比对。

[0211] NSG小鼠中14H124 AML细胞的扩增

[0212] 由于该患者仅可获得2000万个细胞,因此已将患者14H124的母细胞解冻,在PBS中洗涤,并在亚致死全身辐照 (2.5Gy,137Cs- γ 源) 后24h静脉内注射至10只NOD-scid IL-2R γ null^(NSG) 小鼠 (2×10^6 /小鼠)。在第122天通过流式细胞术在外周血中评估人AML细胞移植。简言之,通过尾静脉取血收集100 μ l血液,使用RBC裂解缓冲液 (eBioscience) 移除红细胞,在染色缓冲液 (PBS+3%FBS) 中洗涤,用抗人CD45-PacificBlue (HI30,Biolegend) 和抗小鼠CD45-PECy5 (30-F11,BD) 在4℃下染色20min并在PBS中洗涤。在FACS Canto II流式细胞仪 (Becton Dickinson) 上采集数据并使用Flowjo[®] 软件7.0 (Tree Star Inc.,Ashland,OR) 进行分析。

[0213] 在8/10小鼠中发现高于1%的人细胞嵌合体。移植后第188-264天,在出现疾病体征(贫血、体重减轻>20%或明显肿瘤)时处死小鼠,并收集骨髓、脾和实体瘤(可见于肩胛间、颈部和臀部区域或肾脏、肝脏和淋巴结)。肿瘤在液氮中快速冷冻,用于未来通过质谱进行处理。通过粉碎脾并用4℃ PBS冲洗股骨和胫骨(收集骨髓)来收集AML细胞。移除细胞中的红细胞,过滤(100 μ m)以移除碎片,计数并按上述方法进行流式细胞术(5×10^5 个细胞)处理以评估其纯度,或者在Trizol[®] (Invitrogen) 中裂解和冷冻保存用于未来的RNA测序(所有剩余的细胞)。选择一个大小>1cm³且具有600万个纯度>99%的骨髓来源的母细胞(图14A-B)可用于RNA测序的肿瘤进行处理,以通过质谱识别MAP。在第122天外周血中没有人嵌合体(移植失败)的两只小鼠没有出现任何疾病体征,并在实验结束时(第264天)处死。所有处死均通过CO₂窒息和颈脱位人道地进行。每周三次评估小鼠的疾病体征,并在实验期间每天监测。

[0214] RNA提取、文库制备和测序

[0215] 使用Trizol[®]/氯仿提取和在RNeasy[®] Mini提取柱(Qiagen)上的纯化完成RNA提取。400ng的总RNA用于文库制备。使用BioAnalyzer Nano (Agilent) 评估总RNA的质量,所有样本的RIN均高于8。使用KAPA mRNAseq Hyperprep试剂盒(KAPA,货号KK8581)进行文库制备。使用终浓度为51nM的Illumina Truseq index进行连接,扩增cDNA文库需要12个PCR循环。样本14H124分别使用4M细胞、1 μ g的总RNA进行。除了扩增使用10个PCR循环而不是12个循环完成之外,文库制备如先前样本进行。通过QuBit和BioAnalyzer DNA1000对文库进行定量。所有文库均稀释至10nM,并使用KAPA文库定量试剂盒(KAPA;货号KK4973)通过qPCR归一化。将文库合并为等摩尔浓度。用Illumina Nextseq500使用Nextseq High Output Kit150个循环(2x80bp)使用2.8pM的合并文库进行测序。每个样本产生约120-200M双末端PF读数。文库制备和测序在免疫学和癌症研究所的基因组学平台(IRIC)进行。

[0216] 用于鸟枪质谱识别的数据库生成

[0217] 1) 个性化典型蛋白质组的生成。如前所述开展 (Laumont等人, 2018)。简言之, 使用 Trimmomatic v0.35 修剪 RNA-Seq 读数, 并使用 STAR v2.5.1b 与 GRCh38.88 比对 (Dobin等人, 2013), 使用除了 `--alignSJoverhangMin`、`--alignMatesGapMax`、`--alignIntronMax` 和 `--alignSJstitchMismatchNmax` 参数之外的默认参数运行, 其默认值分别替换为 10、200,000、200,000 和 “5 -1 55”, 以生成 bam 文件。使用 freeBayes 1.0.2-16-gd466dde (arXiv: 1207.3907) 识别最小交替计数设置为 5 的单碱基突变。使用默认参数, 用 kallisto v0.43.0 (Bray等人, 2016) 以每百万转录物 (tpm) 定量转录表达。最后, 使用 pyGeno 在参考外显子组中插入高质量的样本特异性单碱基突变 (freeBayes 质量 > 20), 并导出由表达的转录物 (tpm > 0) 生成的已知蛋白质的样本特异性序列, 以生成个性化的经典蛋白质组。

[0218] 2) 通过 mTEC k-mer 耗竭生成 AML 特异性蛋白质组 (图 2A)。如前所述开展 (Laumont等人, 2018)。简言之, 每个样本的 R1 和 R2 fastq 文件按照上述报告进行修剪, 并使用 FASTX-Toolkit v0.0.14 的 `fastx_reverse_complement` 函数对 R1 读数进行反向补充。使用 Jellyfish v2.2.3 (Marcais and Kingsford, 2011) 生成 K-mer 数据库 (24 或 33 长)。针对每个 AML 样本生成单个数据库, 而 6 个 mTEC 样本通过连接它们的 fastq 文件组合在唯一的数据库中。由于 k-mer 组装的持续时间 (见下文) 在超过 3000 万 k-mer 时呈指数增长, 因此每个 AML 33 个核苷酸长的 k-mer 数据库基于发生时的样本特定阈值 (给定 k-mer 在数据库中出现的次数) 过滤, 以在组装步骤中达到最多 3000 万个 k-mer (表 1)。在此过滤之后, 从每个样本数据库中移除 mTEC k-mer 数据库中至少存在一次的 k-mer, 并使用内部开发的软件 NEKTAR 将剩余的 k-mer 组装成重叠群。简言之, 随机选择提交的 33 个核苷酸长的 k-mer 中的一个作为种子, 该种子从两端延伸, 连续的 k-mer 在同一条链上重叠 32 个核苷酸 (禁用 -r 选项, 作为 k-mers 的链集使用)。当无法组装 k-mer 或有多于一个 k-mer 符合时 (-a 线性组装的 1 个选项), 组装过程将停止。如果是这样, 则选择新种子并继续组装过程, 直到提交列表中的所有 k-mers 均已使用一次。最后, 使用内部 python 脚本对重叠群进行 3 框翻译, 在内部终止密码子处拆分氨基酸序列, 并将得到的子序列与每个样本各自的个性化典型蛋白质组连接。

[0219] 3) ERE 特异性蛋白质组的生成 (图 2B)。对于每个样本, 用 STAR (Dobin等人, 2013) 使用默认参数在人参考基因组 (GRCh38.88) 上比对 RNA-Seq 读数。使用 BEDtools (PMID 20110278) 的 `intersect` 函数, 读取在两个读取数据集中分离, 完全映射到 ERE 序列或典型基因中。如果它们的序列也存在于典型读数数据集中, 则弃去 ERE 读数数据集的读数。然后使用 samtools 视图 (PMID 19505943) 从 ERE 读数数据集中弃去未映射的读取、二级比对和低质量读数。然后将剩余的 ERE 读数在计算机上翻译成所有可能阅读框中的 ERE 多肽。ERE 多肽在终止密码子位置剪接, 弃去下游序列, 仅保留 ≥ 8 个氨基酸的上游序列 (即 MAP 的最小长度)。然后将得到的 ERE 蛋白质组与相应样本的个性化典型蛋白质组连接。

[0220] 4) 通过 mTEC+MPC k-mer 耗竭生成 AML 特异性蛋白质组 (图 2C)。为了进行这种方法, 使用与 mTEC k-mer 耗竭相同的方法, 并进行以下修改: (i) 使用 Jellyfish 生成另外的正常 k-mer 数据库, 结合 11 个用作 k-mer 对照的 MPC 样本的 fastq 文件。因为这些样本没有以链模式测序, 所以 k-mer 数据库是使用 -C 选项生成的, 并且 R1 fastq 文件未反向补充。(ii) 移除 mTEC 或 MPC k-mer 数据库中存在的 AML k-mer, 因此通过此步骤过滤的 k-mer 的数量大于 mTEC k-mer 耗竭方法中的数量。(iii) 由于正常样本的 k-mer 耗竭效率较高, 因此可以使用

较低的发生阈值来预过滤AML k-mer (表1和图7C), 与仅mTEC k-mer耗竭相比, 这些数据库中存在的k-mer的同一性发生了显著变化 (图7D)。重要的是, 没有使用低于3的发生阈值来排除可能的测序错误。所有其他程序均按照“通过mTEC k-mer耗竭生成AML特异性蛋白质组”部分中报告的进行。

[0221] 5) 通过差异k-mer表达生成AML特异性蛋白质组 (图2D)。差异k-mer分析基于DE-kupl的定制使用进行, DE-kupl是一种计算流程, 用于从fastq文件生成k-mer数据库、k-mer丰度的标准化、基于k-mer的过滤它们的发生及其样本间共享、通过使用统计测试比较两种不同条件下样本之间的k-mer丰度、将差异表达的k-mer组装成重叠群、基因组上重叠群的比对和基于其基因组比对的重叠群注释 (图8) (Audoux等人, 2017)。具体地, 首先使用以下参数diff_method Ttest、kmer_length 33、gene_diff_method limma-voom、data_type WGS、lib_type unstranded、min_recurrence 6、min_recurrence_abundance 3、pvalue_threshold 0.05和log2fc_threshold 0.1进行DE-kupl运行, 以将AML样本与11个MPC对照进行比较。这返回diff-counts.tsv文件, 其中包含AML和MPC样本之间显著差异表达的33个核苷酸长k-mers (FDR<0.05) 的序列和归一化计数, 并且在至少6个样本中最少出现3个样本 (MPC或AML)。由于需要自定义k-mer过滤规则, 因此对DE-kupl (log2fc_threshold 0.1) 中的k-mer倍数变化没有应用任何限制, 并且手动过滤diff-counts.tsv文件中提供的k-mer列表以保持所有k-mer (i) 在所有MPC样本中完全不存在 (计数=0) (因此存在于至少6个AML样本中); 或 (ii) 存在于至少6个AML样本中 (>30%的样本) 并且倍数变化 ≥ 10 倍; 或 (iii) 存在于单个MPC样本中, 其丰度低于AML样本中的最低丰度; 或 (iv) 存在于至少6个AML样本中, 倍数变化 ≥ 5 倍且FDR ≤ 0.000001 。基于这些规则, 生成含有 $\sim 41 \times 10^6$ k-mer的新diff-counts.tsv文件, 并用于通过DE-kupl执行k-mer组装, 以及获得含有 $\sim 2.1 \times 10^6$ 重叠群的merge-diff-counts.tsv文件。最后, 使用DE-kupl的annot函数对GRCh38人类基因组上生成的重叠群进行映射和注释。

[0222] 为了获得每个AML样本的个性化重叠群序列, DE-kupl annot的DiffContigsInfos.tsv输出用于构建长度 ≥ 34 个核苷酸 (源自至少2k-mer的组装) 的所有重叠群的bed文件, 并且其对齐没有空位、插入或缺失 (没有N/D/I的CIGAR)。接下来, 使用此bed文件以及bedtools、samtools和bcftools套件从每个AML样本 (samtools mpileup -C50 -uf ref_genome.fasta sample.bam | bcftools call -c | vcfutils.pl vcf2fq -d 8 -D 100 | awk '/^@chr.\$|^chr..\$|^@GL.....\$|^@KI.....\$/,/^+\$/'| sed '/^+/d'| tr "@"">" consensus.fasta) 的bam文件 (用星号表示映射到GRCh38上的读数, 参见“个性化典型蛋白质组的生成。”部分) 生成的共有基因组中提取个性化重叠群序列 (bedtools getfasta -fi consensus.fasta -bed contigs.bed -name >> output.fasta)。

[0223] 使用sed (sed -E "s/NNN+/\n/g") 移除读数 (N) 未覆盖的部分重叠群, 并将所有重叠群写入fasta文件。将具有空位、插入或缺失比对 (并且无法从共有基因组中检索到) 并在DiffContigsInfos.tsv中的相关样本中报告为表达的重叠群序列添加到此fasta文件中。最后, 通过使用内部python脚本 (先前公开或包括在pyGeno中 (Daouda等人, 2016; Laumont等人, 2018)), 重叠群被6框翻译, 不明确的氨基酸序列被转化为所有可能的序列 (因为重叠群重叠的单碱基突变可以编码多个不同的氨基酸序列), 氨基酸序列在内部终止密码子处分裂, 并将得到的子序列与每个样本各自的个性化典型蛋白质组连接。

[0224] 6) 数据库大小的验证-图9B-C。鉴于与典型(个性化)蛋白质组数据库相比,本研究中使用的四种蛋白质组学方法中使用的MS数据库呈现出可变膨胀的大小,因此检查这些更大的大小如何影响MS识别。首先,在19个AML样本中比较用每种方法识别的肽的累积数量(图9B)。这表明,尽管两种方法之间的数据库大小存在显著差异,但与典型蛋白质组相比,识别的肽数量变化不大(ERE方法高达~9%)。接下来,当每个数据库与每个样本中各自的典型个性化蛋白质组连接时,推断适当大小的数据库应该允许识别相比于单独的典型蛋白质组具有相似同一性的典型蛋白质来源肽。如图9C所示,在所有AML样本的每种方法中识别的被注释为蛋白质编码的肽的绝大多数(88.2%-96.2%)与仅基于典型蛋白质组识别的那些肽相同。基于这些观察,可以得出结论,各种数据库的大小均适合可靠的MS识别。

[0225] MHC相关肽的分离

[0226] 将W6/32抗体(BioXcell)在PBS中与PureProteome蛋白A磁珠(Millipore)以1mg抗体/mL浆液的比在室温下孵育60分钟。如所述,使用二甲基庚二酸将抗体与磁珠共价交联。将珠粒在4℃下储存在PBS pH 7.2和0.02%NaN₃中。对于冷冻细胞沉淀样本(9800万个细胞/沉淀),将细胞解冻并重悬于1mL PBS pH 7.2中,并通过添加1mL含有PBS pH 7.2、1%(w/v)CHAPS(Sigma)的洗涤剂缓冲液溶解蛋白酶抑制剂混合物(Sigma)。对于肿瘤样本,将组织切成小块(立方体,~3mm大小),并添加5mL含有蛋白酶抑制剂混合物的冰冷PBS。首先使用设定为速度20000rpm的Ultra Turrax T25均质器(IKA-Labortechnik)将组织块均质两次,持续20秒,然后使用设定为25000rpm的Ultra Turrax T8均质器(IKA-Labortechnik)均质20秒。然后,将550μl冰冷的10X裂解缓冲液(5%w/v CHAPS)添加到样本中。将细胞沉淀和肿瘤样本在4℃下翻滚孵育60分钟,然后在4℃下以10000g离心20分钟。将上清液转移到每个样本含有1mg W6/32抗体共价交联蛋白A磁珠的新试管中,并在4℃下翻滚孵育180分钟。将样本置于磁铁上以将结合的MHC I复合物回收到磁珠上。首先用8×1mL PBS清洗磁珠,然后用1×1mL的0.1X PBS清洗,最后用1×1mL水清洗。MHC I复合物通过使用0.2%甲酸(FA)的酸性处理从磁珠上洗脱。为了移除任何残留的磁珠,将洗脱液转移到2.0mL Costar mL Spin-X离心管过滤器(0.45μm,Corning)中并以855g离心2分钟。使用装有20个直径为1mm的十八烷基(C-18)固相萃取盘(EMPORE)的自制载物台尖端将含有肽的滤液与MHC I亚基(HLA分子和β-2巨球蛋白)分离。先用甲醇预洗载物台吸头,然后用0.2%三氟乙酸(TFA)中的80%乙腈(ACN),最后用0.2% FA。将样本装载至载物台尖端并用0.2%FA洗涤。肽用0.1% TFA中的30% ACN洗脱,使用真空离心干燥,然后储存在-20℃直至MS分析。

[0227] 质谱分析

[0228] 将干燥的肽提取物重悬于4%甲酸中,并上样至自制C18分析柱(15cm x 150μm i.d.,填充C18 Jupiter Phenomenex),在EasynLC II系统上使用56-min梯度(10H005)或106-分钟梯度(所有其他样本),0%至30%的乙腈(0.2%甲酸)和600nL/min的流速。使用Q-Exactive HF质谱仪(Thermo Fisher Scientific)在正离子模式下使用1.6kV的Nanospray 2源对样本进行分析。每个完整的MS谱,以60,000分辨率采集,然后是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 5×10^4 (10H005)或 2×10^4 (所有其他样本),进样时间为100ms (10H005)或500ms (15H023、15H063、15H080、05H149)或800ms (所有其他样本),以及碰撞能量为25%。

[0229] 合成肽

[0230] 当可获得足够量的材料时, TSA^{hi}的氨基酸序列用合成肽进一步验证, 如前所述 (Zhao等人, Cancer Immunol Res. 2020Feb 11; doi:10.1158/2326-6066.CIR-19-0541. [Epub ahead of print])。

[0231] 生物信息学分析

[0232] 所有分析均对修剪后的数据进行, 除非另有说明, 否则所有比对均使用之前部分中所述的星号进行, 并且所有比对均在GRCh38.88上进行。

[0233] 使用PEAKS X (Bioinformatics Solution Inc.) 针对相关数据库搜索所有液相色谱 (LC) -MS/MS (LC-MS/MS) 数据。对于肽识别, 前体离子和碎片离子的容差分别设定为10ppm和0.01Da。氧化 (M) 和脱酰胺 (NQ) 的发生被设定为可变修饰。

[0234] 1) MAP的识别。肽识别后, 获得每个样本的独特肽列表, 并将5%的错误发现率 (FDR) 应用于肽评分。使用NetMHC 4.0 (Andreatta和Nielsen, 2016) 预测与样本HLA等位基因的结合亲和力, 并且仅有8至11个氨基酸长度的百分位数 $\leq 2\%$ 的肽用于进一步注释。

[0235] 2) 目的MAP (MOI) 的识别和验证。对于这两种k-mer耗竭方法, 使用与先前描述的类似方法进行 (Laumont等人, 2018)。简言之, 分别搜索每个MAP及其编码序列的相关AML和正常典型蛋白质组 (针对所有mTEC和MPC构建, 如上所述) 或癌症和正常24个核苷酸长度的k-mer数据库 (由组合mTEC或组合的MPC, 如上文所述)。无论其编码序列检测状态如何, 在正常典型蛋白质组中检测到的MAP均被排除。在正常典型蛋白质组和正常k-mer中均未检测到的MAP被标记为MOI。与正常样本相比, 两个典型蛋白质组中均不存在但存在于两个k-mer数据库中的MAP需要使其RNA编码序列在AML中过表达至少10倍, 才能被标记为MOI。最后, 只有当它们各自的编码序列始终将它们标记为MOI时, 对应于几个RNA序列 (源自不同蛋白质) 的MAP才能被标记为MOI。

[0236] 对于ERE方法, 基于其在ERE中的氨基酸序列和个性化典型蛋白质组的存在, 针对每个单独的MAP提供“是”、“可能”或“否”的ERE状态。对于“可能”候选者, 计算ERE读数和典型读数数据集中肽编码序列的表达水平 (即肽的24个核苷酸长度的k-mer集的最小出现)。只有在ERE读数数据集中表达至少高10倍的“可能”候选者才被视为ERE MAP。然后在IGV中手动验证剩余的ERE MAP候选者 (Robinson等人, Nat Biotechnol. 2011Jan; 29 (1): 24-6), 以确定肽的编码序列是否包含种系多态性, 并且与ERE序列和典型注释序列 (如适用) 相比具有适当的方向。

[0237] 对于差异k-mer方法, 搜索MAP的完整列表的AML特定蛋白质组以标记为MOI候选者。接下来, 在19个AML样本和11个用作DE-kupl中的对照的MPC中评估每个MAP的RNA表达 (按照下一部分中描述的程序), 并标记为MOI, 所有MAP在正常和癌症样本之间的最小倍数变化为5。由于MAP RNA表达评估程序基于参考基因组进行定量, 因此无法正确定量源自突变的候选MOI, 并系统地将其标记为候选MOI。为了明确地验证每个MOI在每个被识别的AML样本中存在的RNA水平, 从DE-kupl的DiffContigsInfos.tsv输出中检索MOI编码序列并搜索相关fastq文件 (正向R2 fastq中的序列和反向R1 fastq中的反向补充)。弃去未通过该检查的MOI。

[0238] 对于所有MOI候选者列表 (四种不同的方法), 由于标准MS方法无法区分亮氨酸和异亮氨酸变体, 因此检查了每个列表并弃去被标记为非MOI的现有变体, 除非它比变体呈现更高的RNA表达。手动检查所有MOI的MS/MS谱以移除任何错误识别。最后, 通过使用BLAT

(UCSC基因组浏览器的工具)在参考基因组上映射包含其编码序列的读数,将基因组位置分配给所有MOI。排除读数与一致基因组位置不匹配或与高变区(诸如MHC、Ig或TCR基因)匹配的MOI。对于具有一致基因组位置的那些,IGV用于排除具有与已知种系多态性(dbSNP149)重叠的编码序列的MOI。

[0239] 3) RNA-Seq数据中MAP编码序列的定量。为了明确评价每个MAP的RNA表达,将所有MAP氨基酸序列反向翻译成所有可能的核苷酸序列。接下来,使用GSNAP(Wu等人,2016)在基因组上映射所有这些可能的序列,并使用-n 1000000选项来定位能够编码给定MAP的所有基因组区域。为了可靠地捕获由重叠剪接位点的序列编码的MAP,还在转录组映射可能的MAP编码序列(cDNA和非编码RNA)以提取(具有--长度80选项的samtools faidx)大部分(80个核苷酸)参考转录组序列,其然后在参考基因组上映射(GSNAP,具有--use-splicing和--novelsplicing=1选项)。对于由不同TSA发现管道生成的MOI,还进行包含其编码序列的所有读数的基因组比对。过滤GSNAP的输出以仅保持序列和参考之间的完美匹配,以生成包含所有可能的基因组区域的bed文件,这些区域易受给定MAP编码的影响。通过使用samtools view(-F256选项)、grep和wc(-l选项),每个期望RNA-Seq样本(诸如AML、GTEx或正常样本)中各自基因组位置包含MAP编码序列的读取数被计数,在参考基因组上比对,用星号表示(bam文件)。最后,对给定MAP的所有读数计数(来自不同区域和编码序列)进行求和,并根据每个评估样本中测序的读数总数进行归一化,以获得每亿读数(RPM)计数。

[0240] 4) 免疫原性评估。使用Repitope进行MOI的免疫原性预测(Ogishi和Yotsuyanagi, 2019)。使用预定义的MHC_Human_MinimumFeatureSet变量进行特征计算并更新(2019年7月12日)FeatureDF_MHCI和FragmentLibrary文件在文件包的Mendeley存储库(<https://data.mendeley.com/datasets/sydw5xnxpt/1>)上提供。

[0241] 5) AML患者的MOI呈递和表达。为了识别能够呈递给给定MAP(混杂结合物)的所有可能的HLA等位基因,使用MHCcluster在线工具(<http://www.cbs.dtu.dk/services/MHCcluster/>) (Thomsen等人,2013)。聚类值 ≤ 0.4 的HLA等位基因被认为能够呈递相同的MAP。为了评价Leucegene队列中AML患者的MOI呈递,首先用Optitype确定他们的HLA类型,如果其在RNA水平的表达高于2rphm(而不是0rphm,以最大化呈递的概率)以及患者是否表达能够呈递MOI的HLA等位基因(如NetMHC4.0所预测的,用于对每个发现的MOI的呈递分子进行原始识别,而MHCcluster用于识别混杂结合物)。如果患者表达能够呈递相同MOI的两个不同HLA等位基因,则认为MOI呈递两次。

[0242] 为了评价与高TSA^高表达相关的分子特征,如果TSA^高在该患者中的表达高于其在整个队列中的中位表达(仅基于非空值计算),则认为TSA^高在给定患者中表达。对每名患者的高表达TSA^高(#HE-TSA^{hi})总数进行计数,并用于进行与基因表达的相关性分析以及与突变或其他临床特征的关联(见下一部分)。

[0243] 6) 生存分析。Leucegene队列中374名患者的生存数据由Leucegene团队(<https://leucegene.ca>)的友情馈赠。进行生存分析以评估如上所述计算的高计数HLA-TSA^{hi}复合物(HLA-TSA^{hi}的HLA限制性呈递)与临床结局(总体生存率)之间的关联。根据他们可能存在的TSA^{hi}总数将患者分为两组:高表达者(HLA-TSA^{hi}计数的上四分位数)和低表达者(所有其他患者)。使用Kaplan-Meier曲线比较两组之间的生存率,并通过GraphPad Prism v7.0中的对数秩检验评估显著性。在多变量分析中,使用R中的生存分析v0.1.1包进行,年龄被纳入

作为连续变量,突变被编码为存在/不存在(1/0),细胞遗传学风险的评估被视为单独的组,并针对中间相对于有利风险和不利相对于有利风险进行。

[0244] 7) 突变分析。NPM1、FLT3-ITD、FLT3-TKD、IDH1 (R132) 和双等位基因CEBPA的突变数据检索自先前发表的Leucegene队列数据(Audemard等人,2019;Lavallee等人,2016)。用Freebayes检测ASXL1、TP53、DNMT3A、IDH2 (仅R140和R172)、WT1、RUNX1和TET2中的突变并过滤以移除突变:(i) 变异等位基因频率(VAF) < 20%; (ii) 在COSMIC数据库(<https://cancer.sanger.ac.uk/cosmic>)中标记为SNP; (iii) 具有低推定影响(5'UTR提前起始密码子获得变体;剪接区变体和同义变体;终止保留变体;同义变体); (iv) 如FATHMM-XF (<http://fathmm.biocompute.org.uk/fathmm-xf/>) 所预测的,错义SNP对蛋白质结构和功能具有良性影响(Rogers等人,2018); (iv) 涉及AAAAA+或TTTTT+的插入和缺失; (v) 在COSMIC数据库中仅标记为种系的突变(Tate等人,2018)。

[0245] 8) 基因表达分析。所有转录物表达定量均使用默认参数的kallisto v0.43.0进行。Kallisto的转录水平计数估计值使用R包tximport转换为基因水平计数。EdgeR用于使用TMM算法对计数进行归一化,并输出每百万计数(cpm)值。只有蛋白质编码基因(如Ensembl的BioMart工具中所报告的那样,useast.ensembl.org/biomart)被保留用于进一步分析。每个基因表达和HE-TSA^{hi}计数之间的系统Pearson相关性使用R中的cor.test函数进行。对所有非NPM1/FLT3-ITD/DNMT3A突变和非FAB-M1患者进行相关性。使用Benjamini&Hochberg方法(R中p校正)对P值进行校正,并且在三个相关分析中的至少一个中,只有FDR < 0.00001的基因,在三个分析中的FDR < 0.001,在三个分析中一致的相关系数(正或负),以及在至少一个分析中的相关系数 > 0.3或 < -0.3被保留用于下游处理。

[0246] 使用Rtsne包对从Kallisto的转录水平丰度估计聚合到通过tximport的基因丰度估计获得的表达基因的同源性进行T分布随机邻域嵌入(t-SNE)分析(如果tpm ≥ 1,则表达 = 1,并且如果tpm < 1,则 = 0)。此分析仅使用蛋白质编码基因(最容易产生MAP)。

[0247] 9) GO term和富集图分析。使用Cytoscape v3.7.2中的BiNGO v3.0.3(Maere等人,2005)进行生物过程基因本体论(GO) term过度表示,使用超几何检验并应用FDR校正的P值的显著性临界值 ≤ 0.005。将BiNGO的输出导入Cytoscape中的EnrichmentMap v3.2.1(Merico等人,2010)中,以对冗余GO terms进行聚类并可视化结果。使用Jaccard相似性系数临界值0.25、P值临界值0.001和FDR校正临界值0.005生成EnrichmentMap。在Cytoscape中使用默认设定和600次迭代使用默认的“Prefuse Force-Directed Layout”可视化网络。手动圈出相似的GO terms组。

[0248] 10) 内含子保留和NMF聚类。已使用IRFinder v1.2.5(Middleton等人,2017)对整个Leucegene队列和11个主要MPC样本进行内含子保留(IR)分析。具有IRatio ≥ 10%的内含子(内含子保留在 ≥ 10%的转录物中)和3个读数的最小覆盖率被认为是保留的。过滤内含子以仅保留那些保留在至少2个AML样本中而不保留在任何MPC样本中的内含子。选择10%变化最大的内含子(通过它们在整个队列中的IRatio变异系数)进行进一步分析(6988个内含子)。无监督一致性聚类结果使用R中的NMF v0.21.0(Gaujoux和Seoighe,2010)包在选定内含子的IRatio上生成,使用默认Brunet算法,以及200次迭代用于排名调查和聚类运行。聚类结果是通过考虑共有成员矩阵的共同评分和平均轮廓宽度来选择的,用于具有3至15个聚类的聚类解决方案。

[0249] 通过识别NMF元基因(W矩阵)输出文件中排名靠前的2%内含子来生成丰度热图。移除重复名称产生1211个内含子的列表。针对每个Leucegene样本生成这些内含子IRatios的矩阵,重新排序以匹配NMF聚类输出,并使用R中的heatmap.3包进行具有中心相关距离度量 and 完整连接的内含子的分层聚类。

[0250] ELISPOT测定

[0251] 1) 单核细胞来源的树突状细胞的产生。如前所述,由冷冻PBMC产生单核细胞来源的树突状细胞(Vincent等人,Biology of Blood and Marrow Transplantation:Journal of the American Society for Blood and Marrow Transplantation,22Oct 2013,20(1):37-45;Laumont等人,Nat Commun.2016Jan 5;7:10238)。简言之,通过在补充有5%人血清(Sigma-Aldrich)、丙酮酸钠(1mM)、IL-4(100ng/mL,Peprotech)和GM-CSF(100ng/mL,Peprotech)的X-VIVOTM15培养基(Lonza Bioscience)中培养8天,从贴壁的PBMC部分制备DC。培养7天后,用IFN- γ (1000IU/mL,Gibco)和LPS(100ng/mL,Sigma Aldrich)过夜使DC成熟。在成熟过程后的2h内,DC被负载2 μ g/mL的肽,然后在它们用作T-DC培养中的APC之前进行辐照(40Gy)。对于对照组,用含有MelanA、NS3和Gag-A2肽(所有三种结合HLA-A*02:01)的混合物对DC进行脉冲处理。

[0252] 2) 体外肽特异性T细胞扩增。解冻的PBMC首先使用人CD8⁺T细胞分离试剂盒(Miltenyi Biotech)富集CD8⁺T细胞,并以1:10的APC:T细胞比与自体肽脉冲DC共同孵育。在补充有8%人血清(Sigma-Aldrich)、L-谷氨酰胺(Gibco)和细胞因子的高级RPMI培养基(Gibco)中培养扩增的T细胞4周(每7天进行一次脉冲DC再刺激)。在第一个共培养周,将IL-12(10ng/mL)和IL-21(30ng/mL)添加到培养基中。两天后,IL-2(100UI/mL)也被添加到细胞因子混合物中。第二周,向培养基中添加IL-2(100UI/mL)、IL-7(10ng/mL)、IL-15(5ng/mL)和IL-21(30ng/mL)。对于共培养的最后两周,使用IL-2(100UI/mL)、IL-7(10ng/mL)和IL-15(5ng/mL)。每两天将补充有适当细胞因子混合物的培养基添加到共培养物中。在共培养的第四周结束时,收获细胞以进行ELISPOT测定。

[0253] 3) IFN γ ELISPOT测定。根据生产商的建议使用ELISpot人IFN γ (R&D Systems,USA)试剂盒进行实验。然后将收获的CD8⁺T细胞铺板并在37℃下在用作刺激细胞的辐照肽脉冲PBMC(40Gy)存在下孵育24小时。作为阴性对照,将分选的CD8⁺T细胞与辐照的未脉冲PBMC一起孵育。如生产商方案中所述显示斑点,并使用ImmunoSpot S5 UV分析仪(Cellular Technology Ltd,Shaker Heights,OH)计数。在从阴性对照孔中减去斑点计数后,IFN- γ 的产生表示为每10⁶个CD8⁺T细胞中肽特异性斑点形成细胞(SFC)的数量。

[0254] 免疫原性预测

[0255] 使用Repitope进行MOI的免疫原性预测(Ogishi和Yotsuyanagi,2019)。使用预定义的MHCI_Human_MinimumFeatureSet变量进行特征计算并更新(2019年7月12日)FeatureDF_MHCI和FragmentLibrary文件在文件包的Mendeley存储库(<https://data.mendeley.com/datasets/sydw5xnxpt/1>)上提供。

[0256] TCR和细胞毒性T细胞特征分析

[0257] 使用TRUST4软件(Li等人,2017)和默认参数对437名Leucegene患者的RNA-seq数据进行TCR库分析。通过归一化每千克TCR读数(CPK)的TCR CDR3(完整和部分)数量来估计T细胞的克隆型多样性。通过免费门户网站(<http://tcr.cs.biu.ac.il/>)使用基于自动编码

器的模型和VDJdb作为训练数据库,对TRUST4和MOI检测到的完整TCRbeta CDR3氨基酸序列之间的相互作用进行ERGO (Springer等人,2020) 预测。

[0258] 对于细胞毒性T细胞特征分析,每名患者的预测HLA-TSA^{hi}对计数除以rphm表达 ≥ 2 的TSA^{hi}计数,以获得归一化的TSA^{hi}呈递水平。从分析中弃去没有HLA-TSA^{hi}计数或在诊断时未收集的患者样本。其余361名患者根据其归一化的TSA^{hi}呈递水平进行分组,并通过差异基因表达分析将高于分布中位值的患者与其他患者(低于中位值)进行比较。该分析在R3.6.1中进行。原始读取计数转换为每百万计数(cpm),相对于文库大小进行归一化,并通过使用edgeR 3.26.8 (Robinson等人,2010) 和limma 3.40.6 (Ritchie等人,2015) 在至少2个样本中保留cpm>1的基因,过滤掉低表达基因。随后使用limma的lmfit进行voom转换和线性建模。最后,使用eBayes计算矫正的t统计量。 $p \leq 0.01$ 和 $-0.3 \geq \log_2(FC) \geq 0.3$ 的基因被认为显著差异表达。

[0259] 细胞因子分泌测定和右旋体

[0260] 在使用载肽的单核细胞来源的树突状细胞和基于细胞因子 (Janelle等人,2015) 进行三轮刺激后,在7.5 μ g/ml的布雷非德菌素A (Sigma-Aldrich, Oakville, ON) 存在下,将 1.0×10^6 个细胞与二甲基亚砜(DMSO)、5 μ g/ml目的肽、5 μ g/ml对照肽(阴性对照)或50ng/ml佛波醇12-肉豆蔻酸酯13-乙酸酯(PMA)和500ng/ml离子霉素(阳性对照, Sigma-Aldrich) 孵育4小时。然后根据生产商的说明 (BD Biosciences, Mississauga, ON), 用细胞表面抗体对细胞进行染色,并使用Cytotfix/Cytoperm缓冲液进行固定和透化以进行细胞内染色。将透化细胞与针对IFN γ 、IL-2和TNF α (BD Biosciences) 的抗体在4 $^{\circ}$ C下孵育20分钟,然后重悬于补充有2%胎牛血清(FBS; ThermoFisher, Waltham, MA, USA) 的磷酸盐缓冲盐水(PBS)中,然后收获。使用LSRII流式细胞仪 (BD Biosciences) 进行采集,并使用FlowJoTMV10软件 (BD Biosciences) 分析数据。对于多聚体染色,将 1.0×10^6 个细胞在4 $^{\circ}$ C下用定制的荧光右旋体 (Immudex, Copenhagen, Denmark) 染色45分钟,然后在4 $^{\circ}$ C下用CD8单克隆抗体 (eBiosciences, San Diego, CA) 染色30分钟。在用LSRII细胞仪 (BD Biosciences) 采集之前,用PBS 2% FBS洗涤细胞。使用FlowJoTMV10软件 (BD Biosciences) 分析数据。

[0261] FEST测定

[0262] 对于FEST测定,T细胞如前所述培养,略作修改 (Danilova等人,2018)。简言之,在第0天,使用人泛T细胞分离试剂盒 (Miltenyi) 对来自健康供体 (BioIVT) 的解冻PBMC进行T细胞富集。T细胞以 2×10^6 /mL重悬于补充有50 μ g/mL庆大霉素 (ThermoFisher Scientific) 和1% HEPES的AIM V培养基中。T细胞阴性部分在30Gy下辐照,洗涤并以 2.0×10^6 /mL重悬于补充有50 μ g/mL庆大霉素和1% HEPES的AIM V培养基中。将每孔1mL的T细胞和经过辐照的T细胞耗竭的细胞与3个TSA^{hi}池中的一个(每个池5TSA^{hi},每种TSA的最终浓度为1 μ M)一起添加到12孔板中,或不添加肽。细胞在37 $^{\circ}$ C、5% CO₂下培养10天。在第3天和第7天,将一半培养基更换为含有100IU/mL IL-2、50ng/mL IL-7和50ng/mL IL-15(第3天)以及200IU/mL IL-2、50ng/mL IL-7和50ng/mL IL-15(第7天)的新鲜培养基。在第10天,收获细胞并使用人CD8⁺T细胞分离试剂盒 (Miltenyi) 进一步分离CD8⁺细胞。作为阴性对照,CD8⁺T细胞也从同一健康供体的新鲜解冻的未培养PBMC中分离。使用Qiagen DNA血液迷你试剂盒 (Qiagen) 从CD8⁺T细胞中提取DNA。使用ImmunoSEQ平台 (Adaptive Biotechnologies) 的测量分辨率进行TCR VBCDR3测序。从immunoSEQ门户导出的原始数据使用FEST网络工具 (www.stat-

apps.onc.jhmi.edu/FEST) 进行处理,没有最少数量的模板和“忽略基线阈值”参数。

[0263] 定量和统计分析

[0264] 除非在图例中明确提及,否则所有比较两种情况的统计检验均使用Mann-Whitney U检验进行。所有相关性均使用Pearson相关系数进行评估。除非另有说明,否则箱线图中的所有方框均显示分布的中位值、第25和第75个百分位数,并且须线延伸至第10个和第90个百分位数。除非另有说明,否则所有柱状图均显示具有标准偏差(SD)的平均值。绘图和统计检验主要使用GraphPad Prism v7.00进行。对于所有统计检验,****表示 $p < 0.0001$,***表示 $p < 0.001$,**表示 $p < 0.01$,以及*表示 $p < 0.05$ 。

[0265] 实施例2:纯化的造血祖细胞是AML中TSA发现的有价值对照。

[0266] MS是唯一可以直接识别MAP的可获得技术(Ehx和Perreault,2019;Shao等人,2018)。通常,基于MS的MAP识别通过使用将获得的串联MS谱与用户提供的蛋白质序列数据库相匹配的软件工具进行。然而,参考蛋白质数据库仅包含典型蛋白质序列,因此不能实现识别源自突变和异常表达的非典型基因组区域(其是aeTSA的主要来源)的MAP(Laumont等人,2018)。先前已经描述了构建针对全球TSA识别定制的MS数据库的蛋白质组学策略。针对每个肿瘤样本构建定制的数据库,并且必须满足两个标准:足够全面以包含所有潜在的TSA,但规模有限,因为膨胀的参考数据库增加错误发现的风险(Nesvizhskii等人,2014;Chong等人,2020)。数据库构建开始于(i)肿瘤样本的RNA测序,数据的关键,(ii)RNA-seq的计算机切片读取到33个核苷酸长度的子序列(k-mer),以及(iii)减去正常k-mer以创建仅包含癌症特异性k-mer的模块。如癌症研究的许多方面,棘手的问题是阴性对照的选择(此处,正常k-mer的来源)。在先前的研究中,来自mTEC的k-mer被用作正常对照。然而,在AML的情况下,测试另一种类型的阴性对照:分选的骨髓前体细胞(MPC,包括粒细胞/单核细胞祖细胞和各种类型的粒细胞前体)。

[0267] 为了比较mTEC和MPC作为阴性对照的价值,首先比较19个靶AML样本(特征见表1)、6个mTEC样本和6个MPC样本之间的相似性,这些样本先前已进行高覆盖率RNA-seq(Maiga等人,2016)。值得注意的是,与mTEC相比,MPC从AML中耗竭的k-mer平均多16.4%,这表明MPC和AML之间的转录组重叠比mTEC和AML之间更大(图1A)。因此,虽然mTEC和MPC获得了相似的k-mer计数(~ 8.7 vs $\sim 9.9 \times 10^8$),但与mTEC($\sim 1.9 \times 10^8$, $\sim 22\%$)相比,MPC与AML共享更多的专有k-mer($\sim 3.3 \times 10^8$, $\sim 33\%$) (图1B)。为了确定谱系差异是这种更高相似性的起源,基于表达的蛋白质编码基因的同源性,对AML样本以及从各种来源下载的一系列分选的上皮细胞和造血细胞RNA-seq进行了t-SNE聚类(参见方法)。这表明AML样本与造血细胞聚类在一起,而mTEC与上皮细胞聚类在一起(图1C)。重要的是,mTEC表达的基因多样性最高,与其生物学功能一致(图1D)。总之,这些结果表明,尽管mTEC转录组的多样性具有多样性,但对于在AML中发现TSA,MPC是比mTEC更好的正常对照。作为推论,当减去MPC k-mer而不是mTEC k-mer时,AML特定k-mer数据库的大小更小。

[0268] 实施例3:开发基于MPC的TSA发现方法

[0269] 除了捕获整个AML TSA环境外,还评估了四种策略以构建参考数据库。前两种策略先前已报道(图2A,B),其他两种是新的(图2C,D)。重要的是,AML样本的MS分析仅进行一次,因此四种不同的TSA发现方法中的每一种均在每个AML样本的相同MS谱上进行。第一个策略取决于mTEC减法(图2A)(Laumont等人,2018)。第二个特别关注由ERE编码的MAP,它们可能

是TSA的丰富来源(图2B) (Larouche等人,2020)。

[0270] 在第三种策略中,来自mTEC和MPC的k-mer被耗竭(图2C)。值得注意的是,耗竭步骤之前是基于k-mer的发生(k-mers在同一样本中发生的次数,图8A,B)的过滤,以限制k-mer的最终数量至~3000万用于重叠群组装(更k-mer的组装在计算时间方面要求太高)。因此,与仅mTEC相比,mTEC+MPC k-mer耗竭从AML样本移除更多k-mer,降低了~2-3倍发生阈值,从而能够发现mTEC k-mer耗竭方法数据库遗漏的MAP(图8C,D)。

[0271] 第四种策略旨在规避k-mer耗竭策略的主要警告:正常和癌症样本中k-mer丰度之间缺乏比较。具体地,在k-mer耗竭策略中,给定k-mer的存在,即使发生一个,在正常对照中也导致在癌症样本中过滤该k-mer,即使它在癌症中的频率是正常对照组的100倍。简言之,差异k-mer表达(DKE)分析使用DE-kupl计算协议(Audoux等人,2017)进行,并进行了一些内部校正,并且可以总结如下(图2D和图9A):(i)对至少30%的AML样本中存在(发生率 ≥ 3)的k-mer进行预过滤;(ii)k-mer丰度的归一化;(iii)通过用户定义的算法对k-mer丰度进行统计比较;(iv)将显著差异过表达的k-mer(最小倍数变化为10)组装成重叠群和(v)在基因组上比对重叠群以确定它们的起源区域。因为它们是最密切相关的正常样本,所以在本研究中选择MPC作为正常对照,并将19个AML样本与11个可获得的高覆盖MPC样本进行比较。接下来,针对每个AML样本生成个性化的重叠群序列(基于读取覆盖率和差异表达重叠群的基因组位置中的SNP调用),翻译成所有可能的阅读框并与个性化典型蛋白质组相结合以进行MAP识别(图2)。

[0272] 实施例4:基于MPC的方法识别AML中的大部分TSA^{hi}。

[0273] 四种TSA发现方法中的每一种均在19个AML样本中识别出数千个MAP(表2)。为了被认为是可操作的TSA,MAP需要由AML细胞大量呈递,并且不能由正常细胞呈递,或者以低至足以不触发T细胞识别的水平呈递,因为表位密度在CD8 T细胞消除靶标中起关键作用(Cosma和Eisenlohr,2019)。因为MAP优先来自高度丰富的转录物(图3A和Pearson等人,2016),所以确定两个重要的阈值:(I)RNA表达水平,低于该水平在正常组织中产生MAP的概率可以被认为是低的,以及(ii)RNA表达倍数变化(FC)是显著增加呈递MAP的概率所必需的。为了实现这一点,评估它们各自的AML样本中所有识别的MAP的RNA表达,发现它遵循绘制为累积频率分布的正态分布(图3B)。这证明,当表达低于8.55个读数/亿(RPHM)时,生成MAP的概率为<5%。鉴于与正常粒细胞相比,AML细胞表达的MHC分子水平相似,并且粒细胞在正常组织中表达的MHC-I水平最高(Berlin等人,2015;Boegel等人,2018),8.55RPHM被确定为所有组织的第一阈值。基于相同的分布,还可以评估不同FC对生成MAP概率的影响(图3C)。这表明从2至5的FC往往比更大的FC对概率的影响更大。因此采用五(5)作为最低FC阈值。

[0274] 基于这两个阈值,建立决策树以基于它们在AML、MPC、其他正常造血细胞和广泛正常成人组织(包括mTEC)中的RNA表达来分离MAP(图3D)。简言之,所有在正常组织中表达低于8.55RPHM并且在AML中表达水平高于MPC的MAP均被标记为TSA,因为它们的检测是它们在AML细胞表面呈递的证据,而它们被正常组织呈递的概率很低。此外,AML和MPC之间的FC至少为5的TSA被标记为TSA^{hi},因为它们最有可能被AML细胞专门呈递。相对于其他组织在造血细胞中过表达但不满足这些标准的其他MAP被分类为TAA或造血特异性抗原(HSA)(图3D)。

[0275] 在每个管道的预过滤步骤(见方法)和基于决策树的分类之后,获得了四个目的MAP(MOI)列表(表2)。mTEC耗竭方法产生了最高比例的HSA,而绝大多数TSA^{hi}通过基于MPC的方法识别(图3E、F和10A)。由于DKE方法预过滤了在最少患者中发生最少的k-mer,因此两种

基于MPC的方法之间的重叠很低。因此,与通过DKE方法确定的那些相比,通过耗竭方法识别的大多数TSA^{hi}呈现出较低的患者间共享(图9B)。总之,这些结果表明,DKE方法最适合识别AML中的TSA^{hi},并且可以补充基于MPC的k-mer耗竭方法来识别另外的较少共享的TSA^{hi}。

[0276] 表2:通过四种蛋白质组学方法识别的MOI的详细信息。

[0277]	MOI 序列(SEQ ID NO:)	分类	识别方法	生物型	基因(如果是外显子或内含子)
	RQISVQASL (1)	HSA	差异 k-mer 表达	内含子	PPP6R2
[0278]	DRELRLNEL (2)	HSA	差异 k-mer 表达	编码外显子	PYHIN1
	GARQQIHSW (3)	HSA	差异 k-mer 表达	编码外显子	FADS1
	SGKLRVAL (4)	HSA	差异 k-mer 表达	编码外显子	TBX21
	RSASSATQVHK (5)	HSA	差异 k-mer 表达	基因间	
	SASSATQVHK (6)	HSA	差异 k-mer 表达	基因间	
	FLLEFKPVS (7)	HSA	差异 k-mer 表达	内含子	CEP83
	GPQVRGSI (8)	HSA	差异 k-mer 表达	内含子	UBE3C
	IRMKAQAL (9)	HSA	差异 k-mer 表达	内含子	KMT2C
	KIKVFSKVY (10)	HSA	差异 k-mer 表达	内含子	TRPM7
	LLSRGLLFRI (11)	HSA	差异 k-mer 表达	内含子	ANKIB1
	LPIASASLL (12)	HSA	差异 k-mer 表达	内含子	MAML3
	LYFLGHGSI (13)	HSA	差异 k-mer 表达	内含子	KMT2C
	NPLQLSLSI (14)	HSA	差异 k-mer 表达	内含子	ARL15
	DLMLRESL (15)	HSA	差异 k-mer 表达	ncRNA	RP11-477N3.1
	VTFKLSELF (16)	HSA	差异 k-mer 表达& ERE	ERE	
	IALYKQVL (17)	HSA	ERE	ERE	
	IVATGSLLK (18)	HSA	ERE	ERE	
	KIKNKTKNK (19)	HSA	ERE	ERE	
	KLLSLTIYK (20)	HSA	ERE	ERE	
	NILKKTIVL (21)	HSA	ERE	ERE	
	NPKLKDIL (22)	HSA	ERE	ERE	
	NQKKVRIL (23)	HSA	ERE	ERE	
	PFPLVQVEPV (24)	HSA	ERE	ERE	
	SPQSGPAL (25)	HSA	ERE	ERE	
	TSRLPKIQK (26)	HSA	ERE	ERE	
	LLDNILQSI (27)	HSA	ERE & mTECs k-mer 耗竭	ERE	
	RLEVRKVIL (28)	HSA	mTECs k-mer 耗竭	外显子-内含子	DOCK8

[0279]

LSWGYFLFK (29)	HSA	mTECs k-mer 耗竭	基因间	
TILPRILTL (30)	HSA	mTECs k-mer 耗竭	基因间	
EGKIKRNI (31)	HSA	mTECs k-mer 耗竭	内含子	SLC25A13
FLASFVEKTVL (32)	HSA	mTECs k-mer 耗竭	内含子	OS9
ILASHNLTV (33)	HSA	mTECs k-mer 耗竭	内含子	RPS6KC1
IQLTSVHLL (34)	HSA	mTECs k-mer 耗竭	内含子	PTPRC
LELISFLPVL (35)	HSA	mTECs k-mer 耗竭	内含子	CDCA8
NFCMLHQSI (36)	HSA	mTECs k-mer 耗竭	内含子	IKZF1
PARPAGPL (37)	HSA	mTECs k-mer 耗竭	内含子	INPP5D
PLPIVPAL (38)	HSA	mTECs k-mer 耗竭	内含子	USP3
SNLIRTGSH (39)	HSA	mTECs k-mer 耗竭	内含子	INPP5D
VPAPAQAI (40)	HSA	mTECs k-mer 耗竭	内含子	PAN3
KGHGGPRSW (41)	HSA	mTECs k-mer 耗竭	ncRNA	AC104232.1
ITSSAVTTALK (42)	HSA	mTECs k-mer 耗竭	UTR	LMO2
LLLPESPSI (43)	HSA	mTECs+MPCs k-mer 耗竭 & ERE & mTECs k-mer 耗竭	ERE	
VILIPLPPK (44)	HSA	mTECs+MPCs k-mer 耗竭 & mTECs k-mer 耗竭	ERE 反义	
AVLLPKPPK (45)	HSA	mTECs+MPCs k-mer 耗竭	ERE 反义	
TQVSMAESI (46)	HSA	mTECs+MPCs k-mer 耗竭	ERE	
LNHLRTSI (47)	HSA	mTECs+MPCs k-mer 耗竭	内含子	ZNF407
NTSHLPLIY (48)	TAA	差异 k-mer 表达	内含子	COL4A5
SIQRNLSL (49)	TAA	差异 k-mer 表达	编码外显子	PIEZO2
NVSSHVHTV (50)	TAA	差异 k-mer 表达	编码外显子	TPSB2

[0280]

ALASHLIEA (51)	TAA	差异 k-mer 表达	编码外显子	EHD2
ALDDITIQL (52)	TAA	差异 k-mer 表达	编码外显子	MAMDC2
ALGNTVPAV (53)	TAA	差异 k-mer 表达	编码外显子	DNAAF3
ALLPAVPSL (54)	TAA	差异 k-mer 表达	编码外显子	WT1
APAPPPVAV (55)	TAA	差异 k-mer 表达	编码外显子	IRX3
APDKKITL (56)	TAA	差异 k-mer 表达	编码外显子	FOXC1
AQMNLLQKY (57)	TAA	差异 k-mer 表达	编码外显子	DLC1
DQVIRLAGL (58)	TAA	差异 k-mer 表达	编码外显子	PTPN14
ETTSQVRKY (59)	TAA	差异 k-mer 表达	编码外显子	MMRN1
GGSLIHPQW (60)	TAA	差异 k-mer 表达	编码外显子	TPSB2
GLYYKLHNV (61)	TAA	差异 k-mer 表达	编码外显子	GATA2
GQKPVILTY (62)	TAA	差异 k-mer 表达	编码外显子	IGSF10
GSLDFQRGW (63)	TAA	差异 k-mer 表达	编码外显子	ANGPT1
HHLVETLKF (64)	TAA	差异 k-mer 表达	编码外显子	GGT5
HLLSETPQL (65)	TAA	差异 k-mer 表达	编码外显子	IGSF10
HQLYRASAL (66)	TAA	差异 k-mer 表达	编码外显子	TTC28
HTDDIENAKY (67)	TAA	差异 k-mer 表达	编码外显子	PTPN14
IAAPILHV (68)	TAA	差异 k-mer 表达	编码外显子	GGT5
KAFPFHIIF (69)	TAA	差异 k-mer 表达	编码外显子	GUCY1B3
KATEYVHSL (70)	TAA	差异 k-mer 表达	编码外显子	MYCN
KFSNVTMLF (71)	TAA	差异 k-mer 表达	编码外显子	GUCY1A3
KLLEKAFSI (72)	TAA	差异 k-mer 表达	编码外显子	CYP7B1
KPMPTKVVF (73)	TAA	差异 k-mer 表达	编码外显子	MAMDC2

[0281]

NVNRPLTMK (74)	TAA	差异 k-mer 表达	编码外显子	GATA2
REPYELTVPAL (75)	TAA	差异 k-mer 表达	编码外显子	RBPMS
SEAEAAKNAL (76)	TAA	差异 k-mer 表达	编码外显子	RBPMS
SLWGQPAEA (77)	TAA	差异 k-mer 表达	编码外显子	COL4A5
SPADHRGYASL (78)	TAA	差异 k-mer 表达	编码外显子	SOX4
SPQSAAAEEL (79)	TAA	差异 k-mer 表达	编码外显子	FOXC1
SPVVHQSL (80)	TAA	差异 k-mer 表达	编码外显子	PTPN14
SPYRTPVL (81)	TAA	差异 k-mer 表达	编码外显子	IGSF10
SVFAGVVG (82)	TAA	差异 k-mer 表达	编码外显子	GUCY1A3
SYSPAHLARL (83)	TAA	差异 k-mer 表达	编码外显子	GATA2
THGSEQLHL (84)	TAA	差异 k-mer 表达	编码外显子	IGSF10
TQAPPNVVL (85)	TAA	差异 k-mer 表达	编码外显子	DENND6B
VLVPYEPQV (86)	TAA	差异 k-mer 表达	编码外显子	TP63
VSFDPVRKV (87)	TAA	差异 k-mer 表达	编码外显子	NEGR1
VVFDKSDLAK Y (88)	TAA	差异 k-mer 表达	编码外显子	SLC45A3
YSHHSGLEY (89)	TAA	差异 k-mer 表达	编码外显子	CAV2
YYLDWIHHY (90)	TAA	差异 k-mer 表达	编码外显子	TPSB2
SVYKYLKAK (91)	TAA	差异 k-mer 表达	UTR	IRX3
IYQFIMDRF (92)	TAA	差异 k-mer 表达	编码外显子	FOXC1
GTLQGIRAW (93)	TAA	差异 k-mer 表达& mTECs+MPCs k-mer 耗 竭	编码外显子	NKX2-3
AQKVSVGQAA (94)	TAA	mTECs k-mer 耗竭	UTR	LMO2

[0282]

LYPSKLTHF (95)	TAA	mTECs+MPCs k-mer 耗 竭& mTECs k-mer 耗竭	基因间	
ATQNTIIGK (96)	TAA	mTECs+MPCs k-mer 耗 竭& mTECs k-mer 耗竭	框外翻译	CPA3
AQDIILQAV (97)	TSA ^{hi}	差异 k-mer 表达	基因间	
PPRPLGAQV (98)	TSA ^{hi}	差异 k-mer 表达	UTR	C16orf87
FNVALNARY (99)	TSA ^{hi}	差异 k-mer 表达	编码外显子	LTBP1
GPGSRESTL (100)	TSA ^{hi}	差异 k-mer 表达	编码外显子	PLPPR3
IPHQRSSL (101)	TSA ^{hi}	差异 k-mer 表达	编码外显子	GCSAML
LTDRIYLT (102)	TSA ^{hi}	差异 k-mer 表达	编码外显子	DNAH10
NLKEKKALF (103)	TSA ^{hi}	差异 k-mer 表达	编码外显子	ST8SIA6
VLFGGKVSGA (104)	TSA ^{hi}	差异 k-mer 表达	编码外显子	MFSD2B
VVFPPVNVK (105)	TSA ^{hi}	差异 k-mer 表达	编码外显子	MYCN
SLLIIPKKK (106)	TSA ^{hi}	差异 k-mer 表达	ERE	
APGAAGQRL (107)	TSA ^{hi}	差异 k-mer 表达	基因间	
KLQDKEIGL (108)	TSA ^{hi}	差异 k-mer 表达	基因间	
SLREPQPAL (109)	TSA ^{hi}	差异 k-mer 表达	基因间	
TPGRSTQAI (110)	TSA ^{hi}	差异 k-mer 表达	基因间	
APRGTAAL (111)	TSA ^{hi}	差异 k-mer 表达	内含子	JADE1
IASPIALL (112)	TSA ^{hi}	差异 k-mer 表达	内含子	ELL
ILFQNSALK (113)	TSA ^{hi}	差异 k-mer 表达	内含子	SLC25A25
ILKKNISI (114)	TSA ^{hi}	差异 k-mer 表达	内含子	AKAP13
IPLAVRTI (115)	TSA ^{hi}	差异 k-mer 表达	内含子	FBXO28
LPRNKPLL (116)	TSA ^{hi}	差异 k-mer 表达	内含子	IGF2BP2

[0283]

PAPHPAAL (117)	TSA ^{hi}	差异 k-mer 表达	内含子	HOOK2
SPVVRVGL (118)	TSA ^{hi}	差异 k-mer 表达	内含子	DNAH7
TLNQGINVYI (119)	TSA ^{hi}	差异 k-mer 表达	内含子	SLC39A10
RPRGPRTAP (120)	TSA ^{hi}	差异 k-mer 表达	UTR	PIEZO1
SVQLLEQAIHK (121)	TSA ^{hi}	差异 k-mer 表达& ERE	ERE	
RTPKNYQHW (122)	TSA ^{hi}	差异 k-mer 表达& mTECs k-mer 耗竭	编码外显子	LINC01835
ALPVALPSL (123)	TSA ^{hi}	差异 k-mer 表达& mTECs k-mer 耗竭	内含子	RFX2
SLQILVSSL (124)	TSA ^{hi}	差异 k-mer 表达& mTECs k-mer 耗竭	内含子	LRRC8C
ISNKVPKLF (125)	TSA ^{hi}	差异 k-mer 表达& mTECs k-mer 耗竭	ncRNA	LINC02147
TVIRIAIVNK (126)	TSA ^{hi}	差异 k-mer 表达& mTECs+MPCs k-mer 耗 竭& mTECs k-mer 耗竭	内含子	FTO
KEIFLELRL (127)	TSA ^{hi}	ERE	ERE	
TLRSPGSSL (128)	TSA ^{hi}	ERE	ERE	
TVRGDVSSL (129)	TSA ^{hi}	mTECs k-mer 耗竭	基因间	
ALDPLLLRI (130)	TSA ^{hi}	mTECs k-mer 耗竭	内含子	CETP
ISLIVTGLK (131)	TSA ^{hi}	mTECs k-mer 耗竭	内含子	ASPH
KILDVNLRI (132)	TSA ^{hi}	mTECs k-mer 耗竭	内含子	SIK3
ERVYIRASL (133)	TSA ^{hi}	mTECs k-mer 耗竭	ncRNA	AL359636.1
ILDLESRY (134)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	ERE	
KTFVQQKTL (135)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	ERE	

[0284]

LYIKSLPAL (136)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	ERE	
VLKEKNASL (137)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	ERE	
LGISLTLKY (138)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	基因间	
DLLPKKLL (139)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	BACH1
HSLISIVYL (140)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	XACT
IAGALRSVL (141)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	TNS3
IGNPILRVL (142)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	EFR3A
IYAPHIRLS (143)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	ERC1
LRSQILSY (144)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	LINC00484
RYLANKIHI (145)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	ARHGAP32
SLLSGLLRA (146)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	CD34
SRIHLVVL (147)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	ZNF280C
SSSPVRGPSV (148)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	PTK2
STFSLYLKK (149)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	内含子	DTNA
SLDLLPLSI (150)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭	ncRNA	LINC00996
VTDLLALTV (151)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭& ERE & mTECs k-mer 耗竭	ERE	
RTQITKVSLKK (152)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭& mTECs k-mer 耗竭	基因间	
ILRSPLKW (153)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭& mTECs k-mer 耗竭	内含子	PIEZO2

[0285]

LSTGHLSTV (154)	TSA ^{hi}	mTECs+MPCs k-mer 耗 竭& mTECs k-mer 耗竭	内含子	MYO16
TVEEYLVNI (155)	TSA ^{lo}	差异 k-mer 表达	内含子	ECSIT
QIKTKLLGSL (156)	TSA ^{lo}	差异 k-mer 表达	内含子	KIF27
LPSFSHFLLL (157)	TSA ^{lo}	差异 k-mer 表达	基因间	
CLRIGPVTI (158)	TSA ^{lo}	差异 k-mer 表达	内含子	ZDHHC14
HVSDGSTALK (159)	TSA ^{lo}	差异 k-mer 表达	内含子	PDE4D
IAYSVRALR (160)	TSA ^{lo}	差异 k-mer 表达	内含子	ACBD3
PRGFLSAL (161)	TSA ^{lo}	差异 k-mer 表达	内含子	ZNF26
ISSWLISSL (162)	TSA ^{lo}	差异 k-mer 表达& ERE	ERE	
IPLNPFSSL (163)	TSA ^{lo}	差异 k-mer 表达& mTECs k-mer 耗竭	ERE	
LSDRQLSL (164)	TSA ^{lo}	差异 k-mer 表达& mTECs k-mer 耗竭	内含子	CDK6
LSHPAPSSL (165)	TSA ^{lo}	差异 k-mer 表达& mTECs k-mer 耗竭	内含子	RNF220
LRKAVDPII (166)	TSA ^{lo}	差异 k-mer 表达& mTECs+MPCs k-mer 耗 竭	内含子	SLC39A11
ILLEEQSLI (167)	TSA ^{lo}	ERE	ERE	
LTSISIRPV (168)	TSA ^{lo}	ERE	ERE	
TISECPLLI (169)	TSA ^{lo}	ERE	ERE	
TLKLKKIFF (170)	TSA ^{lo}	ERE	ERE	
ILLSNFSSL (171)	TSA ^{lo}	mTECs k-mer 耗竭	ERE	
LGGAWKAVF (172)	TSA ^{lo}	mTECs k-mer 耗竭	ERE	
LSASHLSSL (173)	TSA ^{lo}	mTECs k-mer 耗竭	ERE	

[0286]

AGDIARLI (174)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	ZWILCH
DRGILRNLL (175)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	ABI1
GLRLIHVSL (176)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	KLF13
GLRLLHVSL (177)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	VAV3
LHNEKGLSL (178)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	ZNF804A
LPSFSRPSGII (179)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	CCDC26
LSSRLPLGK (180)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	ZNF367
MIGIKRLL (181)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	TRIP12
NLKKREIL (182)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	POGLUT1
RMVAYLQQL (183)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	PTEN
SPARALPSL (184)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	RFX8
TVPGIQRY (185)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	ERBIN
VSRNYVLLI (186)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	SLC39A10
LTVPLSVFW (187)	TSA ^{lo}	mTECs k-mer 耗竭	假基因	SMURF2P1- LRRC37BP 1
KLNQAFLVL (188)	TSA ^{lo}	mTECs k-mer 耗竭	内含子	SNTB1
RLVSSTLLQK (189)	TSA ^{lo}	mTECs+MPCs k-mer 耗 竭& ERE & mTECs k-mer 耗竭	ncRNA	AC128707.1
LPSHSLLI (190)	TSA ^{lo}	mTECs+MPCs k-mer 耗 竭	内含子	MAPRE2

[0287]

MOI	免疫原性评 分	AML / MPC	样本中识别	HLA
RQISVQASL (1)	0,3896	0,1430968 44	15H080	HLA-B*27:0 5

[0288]

DRELRNLEL (2)	0,36072	109,65047 72	08H039	HLA-B*14:0 2
GARQQIHSW (3)	0,31568	8,8528717 55	07H122	HLA-B*57:0 1
SGKLRVAL (4)	0,38024	277,184	05H143, 07H060	HLA-B*08:0 1
RSASSATQVH K (5)	0,27736	21,150877 56	15H023	HLA-A*03:0 1
SASSATQVHK (6)	0,33728	22,045931 13	08H053	HLA-A*11:0 1
FLLEFKPVS (7)	0,32488	0,1780338 26	14H124	HLA-A*02:0 1
GPQVRGSI (8)	0,43896	0,2444368 54	07H063	HLA-B*07:0 2
IRMKAQAL (9)	0,30384	0,1412854 4	07H063	HLA-C*06:0 2
KIKVFSKVY (10)	0,15168	0,5638084 15	12H172	HLA-B*15:0 1
LLSRGLLFRI (11)	0,40744	0,3377160 83	14H124	HLA-A*02:0 1
LPIASASLL (12)	0,50848	0,9415209 82	15H023	HLA-B*51:0 1
LYFLGHGSI (13)	0,3268	0,2549318 15	16H123	HLA-A*24:0 2
NPLQLSLSI (14)	0,45888	0,4885435 37	07H060	HLA-B*08:0 1
DLMLRESL (15)	0,4704	0,5437411 44	07H060	HLA-B*08:0 1
VTFKLSLF (16)	0,38384	0,8900190 17	05H143	HLA-B*57:0 1
IALYKQVL (17)	0,37872	0,8416195 33	05H143	HLA-B*08:0 1
IVATGSLLK (18)	0,406	0,5067046 12	15H023	HLA-A*03:0 1
KIKNKTKNK (19)	0,386	0,8468225 74	16H123	HLA-A*03:0 1
KLLSLTIYK (20)	0,2392	0,3603968 52	16H123	HLA-A*03:0 1
NILKKTVL (21)	0,42712	0,9954716 68	15H063	HLA-B*08:0 1
NPKLKDIL (22)	0,37752	1,9791899 6	05H143	HLA-B*08:0 1

[0289]

NQKKVRIL (23)	0,33296	0,4567777 09	07H060	HLA-B*08:0 1
PFPLVQVEPV (24)	0,48656	0,4064113 52	15H023	HLA-B*51:0 1
SPQSGPAL (25)	0,44536	0,5406645 64	16H123	HLA-B*07:0 2
TSRLPKIQK (26)	0,29104	0,5366998 74	07H063	HLA-A*30:0 1
LLDNILQSI (27)	0,42688	2,3447396 46	15H023	HLA-A*02:0 1
RLEVRKVIL (28)	0,2716	0,6197223 53	15H063	HLA-B*08:0 1
LSWGYFLFK (29)	0,27096	2,0974786 4	07H063	HLA-A*30:0 1
TILPRILTL (30)	0,4464	0,6039300 45	16H123	HLA-C*07:0 2
EGKIKRNI (31)	0,36296	0,7945346 2	15H063	HLA-B*08:0 1
FLASFVEKTV L (32)	0,2804	0,3177042 38	14H124	HLA-A*02:0 1
ILASHNLTV (33)	0,35944	0,8365928 66	16H145	HLA-A*02:0 1
IQLTSVHLL (34)	0,37928	0,3646307 95	15H080	HLA-A*02:0 1
LELISFLPVL (35)	0,42824	0,0899276 59	14H124	HLA-A*02:0 1
NFCMLHQSI (36)	0,42648	0,5483519 18	12H172	HLA-A*24:0 2
PARPAGPL (37)	0,44312	0,8309837 29	16H145	HLA-C*03:0 3
PLPIVPAL (38)	0,51424	0,6020842 8	15H023	HLA-B*51:0 1
SNLIRTGSH (39)	0,29392	0,8344244 95	08H039	HLA-B*14:0 2
VPAPAQAI (40)	0,462	0,0856376 44	16H123	HLA-B*07:0 2
KGHGGPRSW (41)	0,25872	0,3527350 92	07H122	HLA-B*57:0 1
ITSSAVTTALK (42)	0,37744	2,0175259 59	11H008, 15H023, 16H123	HLA-A*03:0 1

[0290]

LLLPESPSI (43)	0,39184	0,1780585 78	08H039	HLA-A*02:0 1
VILIPLPPK (44)	0,40664	0,2622800 81	11H035	HLA-A*03:0 1
AVLLPKPPK (45)	0,35024	0,5454255 85	10H005	HLA-A*11:0 1
TQVSMAESI (46)	0,34456	0,7598987 81	07H141	HLA-B*38:0 1
LNHLRTSI (47)	0,4288	0,2755510 12	05H143	HLA-B*08:0 1
NTSHLPLIY (48)	0,24416	405,694	07H060	HLA-A*01:0 1
SIQRNLSL (49)	0,32568	19,187644 64	07H060	HLA-B*08:0 1
NVSSHVHTV (50)	0,27864	7,3430475 81	05H149	HLA-A*68:0 2
ALASHLIEA (51)	0,07824	15,395911 83	08H039, 15H063	HLA-A*02:0 1
ALDDITIQL (52)	0,36416	17,315176 63	05H143	HLA-A*02:0 1
ALGNTVPAV (53)	0,4856	14,047126 51	07H122	HLA-A*02:0 1
ALLPAVPSL (54)	0,47288	1261,102	05H143, 07H122, 16H145	HLA-A*02:0 1
APAPPPVAV (55)	0,53488	344,17745 37	16H123	HLA-B*07:0 2
APDKKITL (56)	0,2792	55,027516 83	11H035	HLA-B*07:0 2
AQMNLLQKY (57)	0,26056	31,386099 92	12H172	HLA-B*15:0 1
DQVIRLAGL (58)	0,44592	246,71359 86	08H039	HLA-B*14:0 2
ETTSQVRKY (59)	0,15824	9,4736372 52	07H141, 14H124	HLA-A*26:0 1
GGSLIHPQW (60)	0,45512	11,494782 71	15H013	HLA-B*57:0 3
GLYYKLHNV (61)	0,23424	13,527382 14	07H122	HLA-A*02:0 1
GQKPVILTY (62)	0,21208	5,2120851 57	16H123	HLA-B*15:0 1

[0291]

GSLDFQRGW (63)	0,31896	19,210140 06	07H122	HLA-B*57:0 1
HHLVETLKF (64)	0,23928	66,824472 84	07H141	HLA-B*38:0 1
HLLSETPQL (65)	0,51952	13,505096 09	05H143, 16H145	HLA-A*02:0 1
HQLYRASAL (66)	0,42744	9,7355695 75	08H039	HLA-B*14:0 2
HTDDIENAKY (67)	0,08944	88,634612 1	08H039	HLA-A*01:0 1
IAAPILHV (68)	0,45152	26,194800 16	15H023	HLA-B*51:0 1
KAFPFHIIF (69)	0,31792	8,7662741 99	07H122	HLA-B*57:0 1
KATEYVHSL (70)	0,14792	67,335342 62	11H008	HLA-C*06:0 2
KFSNVTMLF (71)	0,39376	8,9271050 41	05H149	HLA-A*24:0 2
KLLEKAFSI (72)	0,28016	20,377431 21	05H143, 07H122	HLA-A*02:0 1
KPMPTKVVF (73)	0,32808	20,747884 06	11H035	HLA-B*07:0 2
NVNRPLTMK (74)	0,28872	12,365496	11H008, 16H123	HLA-A*03:0 1
REPYELTVPA L (75)	0,39344	8,6641784 43	08H039	HLA-B*40:0 1
SEAEAAKNAL (76)	0,19208	15,846199 72	08H039	HLA-B*40:0 1
SLWGQPAEA (77)	0,89872	249,81397 21	15H063, 16H145	HLA-A*02:0 1
SPADHRGYAS L (78)	0,31104	28,510552 3	11H035, 16H123	HLA-B*07:0 2
SPQSAAAEL (79)	0,32288	59,618371 48	11H035, 16H123	HLA-B*07:0 2
SPVVHQSL (80)	0,4292	52,977009 97	11H035	HLA-B*07:0 2
SPYRTPVL (81)	0,4632	8,9276394 23	07H063, 11H035	HLA-B*07:0 2
SVFAGVGV (82)	0,112	8,6997700 3	05H143, 16H145	HLA-A*02:0 1
SYSPA HARL (83)	0,2644	8,8371891 3	11H035, 16H123	HLA-C*07:0 2

[0292]

THGSEQLHL (84)	0,3024	3,2933576 87	07H141	HLA-B*38:0 1
TQAPPNVVL (85)	0,44232	11,858529 97	15H013	HLA-C*07:0 2
VLVPYEPPQV (86)	0,48944	641,852	07H122	HLA-A*02:0 1
VSFPDVRKV (87)	0,2852	9,6479218 56	05H143	HLA-C*06:0 2
VVFDKSDLAK Y (88)	0,16416	9,2160675 41	07H060	HLA-A*29:0 2
YSHHSGLEY (89)	0,246	91,165677 32	08H053	HLA-A*01:0 1
YYLDWIIHY (90)	0,29696	21,437448 49	15H013, 16H123	HLA-C*07:0 2
SVYKYLKAK (91)	0,19256	8068,72	11H008, 16H123	HLA-A*03:0 1
IYQFIMDRF (92)	0,25448	86,119169 83	16H123	HLA-A*24:0 2
GTLQGIRAW (93)	0,37888	1098,148	05H143, 11H008	HLA-B*57:0 1
AQKVSQVGA A (94)	0,24528	2,6913388 11	12H172	HLA-B*15:0 1
LYPSKLTTHF (95)	0,22144	5,6276199 75	12H172	HLA-A*24:0 2
ATQNTIIGK (96)	0,22592	2,5416280 72	08H053	HLA-A*11:0 1
AQDIILQAV (97)	0,34352	230,14	08H039	HLA-C*08:0 2
PPRPLGAQV (98)	0,4844	14,354338 65	07H063	HLA-B*07:0 2
FNVALNARY (99)	0,29216	113,60010 75	07H060, 11H008	HLA-A*29:0 2
GPGSRESTL (100)	0,41576	19,381416 84	16H123	HLA-B*07:0 2
IPHQRSSL (101)	0,37912	61,729744 86	15H063	HLA-B*08:0 1
LTDRIYLTTL (102)	0,34256	203,504	08H039	HLA-C*08:0 2
NLKEKKALF (103)	0,33608	8,9681460 67	07H060	HLA-B*08:0 1
VLFGGKVSGA (104)	0,30776	18,899247 92	07H122	HLA-A*02:0 1

[0293]

VVFPPVVK (105)	0,34848	1449,166	11H035, 16H123	HLA-A*03:0 1
SLLIIPKKK (106)	0,38096	5,2158097 97	10H005	HLA-A*11:0 1
APGAAGQRL (107)	0,42808	24,617860 22	07H063, 11H035	HLA-B*07:0 2
KLQDKEIGL (108)	0,24288	343,936	16H145	HLA-A*02:0 1
SLREPQPAL (109)	0,50512	69,205538 79	15H013	HLA-C*07:0 2
TPGRSTQAI (110)	0,39992	8,3147533 82	16H123	HLA-B*07:0 2
APRGTAAL (111)	0,45936	17,290791 31	11H035	HLA-B*07:0 2
IASPIALL (112)	0,44496	20,111766 73	12H172	HLA-C*03:0 3
ILFQNSALK (113)	0,33824	16,925922 79	16H123	HLA-A*03:0 1
ILKKNISI (114)	0,37568	22,536240 85	07H060	HLA-B*08:0 1
IPLAVRTI (115)	0,4452	8,6441179 06	15H023	HLA-B*51:0 1
LPRNKPLL (116)	0,36928	6,2131736 76	15H023	HLA-B*51:0 1
PAPHPAAL (117)	0,43568	13,045462 21	07H063	HLA-C*07:0 2
SPVVRVGL (118)	0,45504	22,154004 44	16H123	HLA-B*07:0 2
TLNQGINVYI (119)	0,2868	9,5443668 47	14H124	HLA-A*02:0 1
RPRGPRTAP (120)	0,5112	29,954259 88	07H063, 11H035	HLA-B*07:0 2
SVQLLEQAIH K (121)	0,33208	56,527538 4	10H005	HLA-A*11:0 1
RTPKNYQHW (122)	0,2	53,717486 86	07H122	HLA-B*57:0 1
ALPVALPSL (123)	0,55128	16,471198 14	16H145, 14H124	HLA-A*02:0 1
SLQILVSSL (124)	0,414	6,7100192 62	08H039	HLA-B*14:0 2
ISNKVPKLF (125)	0,29072	19,420357 41	05H143	HLA-B*57:0 1

[0294]

TVIRIAIVNK (126)	0,35448	43568,4	16H123	HLA-A*03:0 1
KEIFLELRL (127)	0,29152	90,9946	11H008	HLA-B*44:0 3
TLRSPGSSL (128)	0,2984	11,271691 02	11H035	HLA-B*07:0 2
TVRGDVSSL (129)	0,23032	9,6886895 85	11H035	HLA-B*07:0 2
ALDPLLLRI (130)	0,4948	9,5635329 96	07H122	HLA-A*02:0 1
ISLIVTGLK (131)	0,39056	9,2446353 45	15H023	HLA-A*03:0 1
KILDVNLRI (132)	0,28152	8,2664123 85	15H023	HLA-A*02:0 1
ERVYIRASL (133)	0,24856	7,7053235 92	08H039	HLA-B*14:0 2
ILDLESRY (134)	0,3588	16,434	07H063	HLA-A*01:0 1
KTFVQQKTL (135)	0,1732	261,178	07H122	HLA-B*57:0 1
LYIKSLPAL (136)	0,2556	19,061178 34	08H039	HLA-B*14:0 2
VLKEKNASL (137)	0,2932	75,9224	05H143	HLA-B*08:0 1
LGISLTLKY (138)	0,29328	6,3126616 5	07H060	HLA-A*29:0 2
DLLPKKLL (139)	0,40048	141,2948	07H060	HLA-B*08:0 1
HSLISIVYL (140)	0,25104	80,1734	16H145	HLA-C*03:0 3
IAGALRSVL (141)	0,4288	221,362	08H039	HLA-B*14:0 2
IGNPILRVL (142)	0,5164	95,0262	05H143, 07H122	HLA-C*07:0 1, HLA-C*06:0 2
IYAPHIRLS (143)	0,32096	230,392	05H143	HLA-C*07:0 1
LRSQILSY (144)	0,28312	33,629989 83	15H080	HLA-B*27:0 5
RYLANKIHI (145)	0,16208	64,5122	05H149	HLA-A*24:0 2

[0295]

SLLSGLLRA (146)	0,5932	308,14	07H122	HLA-A*02:0 1
SRIHLVVL (147)	0,4008	60,2848	05H143	HLA-B*08:0 1
SSSPVRGPSV (148)	0,37792	113,1364	05H149	HLA-A*68:0 2
STFSLYLKK (149)	0,18904	98,973	10H005	HLA-A*11:0 1
SLDLLPLSI (150)	0,39032	17,236161 37	07H141, 08H039	HLA-C*05:0 1, HLA-A*02:0 1
VTDLLALTV (151)	0,42864	10,810320 29	08H039	HLA-A*01:0 1
RTQITKVSLLK (152)	0,21872	2051,04	08H053	HLA-A*11:0 1
ILRSPLKW (153)	0,41552	56,7804	07H122	HLA-B*57:0 1
LSTGHLSTV (154)	0,32176	45,7274	07H122	HLA-C*06:0 2
TVEEYLVNI (155)	0,24264	1,9117700 46	15H063	HLA-C*07:0 1
QIKTKLLGSL (156)	0,3024	1,4048449 42	07H060	HLA-B*08:0 1
LPSFSHFLLL (157)	0,38896	2,7505537	07H063	HLA-B*07:0 2
CLRIGPVTLL (158)	0,54248	2,3626752 36	11H035	HLA-C*07:0 2
HVSDGSTALK (159)	0,20528	2,4191426 18	15H023	HLA-A*03:0 1
IAYSVRALR (160)	0,38472	3,7552917 37	11H035	HLA-A*03:0 1
PRGFLSALL (161)	0,45264	4,5114225 09	07H063	HLA-B*07:0 2
ISSWLISSL (162)	0,378	4,7063025 79	08H039	HLA-B*14:0 2
IPLNPFSSL (163)	0,3876	2,6750427 3	16H123	HLA-B*07:0 2
LSDRQLSL (164)	0,42816	1,0372982 78	07H060	HLA-A*01:0 1
LSHPAPSSL (165)	0,4132	2,0605472 34	07H063	HLA-A*30:0 1

[0296]

LRKAUDPIL (166)	0,32448	3,9013281 14	07H063	HLA-C*06:0 2
ILLEEQSLI (167)	0,4304	4,3595845 24	16H145	HLA-A*02:0 1
LTSISIRPV (168)	0,39904	1,5031600 93	08H039	HLA-A*02:0 1
TISECPLLI (169)	0,52632	1,6435014 65	08H039	HLA-A*02:0 1
TLKLKKIFF (170)	0,23936	2,5272608 87	05H143	HLA-B*08:0 1
ILLSNFSSL (171)	0,33432	1,0328242 95	15H023	HLA-A*02:0 1
LGGAWKAVF (172)	0,38584	1,3975210 24	15H013	HLA-B*57:0 3
LSASHLSSL (173)	0,37704	2,2036741 56	07H141	HLA-C*12:0 3
AGDIARLI (174)	0,40696	1,8172207 19	08H039	HLA-C*08:0 2
DRGILRNLL (175)	0,46584	1,7535216 23	08H039	HLA-B*14:0 2
GLRLIHVSL (176)	0,37616	2,7386869 31	08H039	HLA-B*14:0 2
GLRLLHVSL (177)	0,388	3,3276266 89	08H039	HLA-B*14:0 2
LHNEKGLSL (178)	0,23976	1,2191219 6	05H143	HLA-C*07:0 1
LPSFSRPSGII (179)	0,38304	2,0586423 72	07H063	HLA-B*07:0 2
LSSRLPLGK (180)	0,32344	3,0161241 62	11H008	HLA-A*03:0 1
MIGIKRLL (181)	0,4116	2,4542059 25	05H143	HLA-B*08:0 1
NLKKREIL (182)	0,42808	1,7223024 38	07H060	HLA-B*08:0 1
RMVAYLQQL (183)	0,41624	1,9956015 1	12H172	HLA-A*02:0 1
SPARALPSL (184)	0,51256	1,4042109 58	16H123	HLA-B*07:0 2
TVPGIQRY (185)	0,39184	1,3040865 81	14H124	HLA-A*26:0 1
VSRNYVLLI (186)	0,24664	4,9081613 28	07H060	HLA-C*07:0 1

[0297]	LTVPLSVFW (187)	0,40472	1,1934039 76	05H143	HLA-B*57:0 1
	KLNQAFLVL (188)	0,35824	1,8204752 5	15H080	HLA-A*02:0 1
	RLVSSTLLQK (189)	0,32288	1,0911983 13	11H035	HLA-A*03:0 1
	LPSHSLLI (190)	0,39184	2,0679380 41	15H023	HLA-B*51:0 1

[0298] 在检查指定基因组位置的肽编码序列后,手动分配生物型。使用Repitope计算免疫原性评分。如netMHC4.0所预测,HLA等位基因对应于给定样本中最有可能呈递肽的基因。合成肽验证仅在TSA^{hi}上进行。

[0299] 为了评估MOI识别的稳健性,观察到的给定肽的平均保留时间(RT)与用于验证用高通量MS识别的MAP的两个同类最佳指标相关:由DeepLC算法计算的RT(Bouwmeester等人,2020)和用SSRcalc评估的疏水性指数(Krokhin,2006),均基于肽序列进行预测。这表明非经典MOI的RT分布与预测非常相关,并且与经典蛋白质组来源肽的分布没有显著差异(F检验),支持它们的正确识别(图3G)。最后,使用Comet算法重复所有MS数据库搜索(最初使用PEAKS软件进行)。重新识别的百分比显示非典型MOI和典型肽之间没有显著差异(图3H,左图)。在MOI中,重新识别58个TSA^{hi}中的52个(90%)(图3H,右图)。由两个不同的搜索引擎识别的MAP之间的这种主要重叠进一步支持了非典型MOI识别的稳健性。

[0300] 实施例5:TSA^{hi}是主要来源于内含子翻译的免疫原性MAP。

[0301] 四种TSA发现方法结果的组合产生总共47个HSA、49个TAA、36个TSA^{lo}和58个TSA^{hi}(没有冗余)。表2中列出了所有MOI的主要特征。根据定义,TSA在所有器官(来自GTEx)以及mTEC和正常造血细胞(图4A)中的表达低于阈值。重要的是,正常组织中TSA^{hi}编码RNA的表达系统性低于先前在临床试验中使用的TAA的表达,而没有脱靶毒性(Chapuis等人,2019;He等人,2020;Legat等人,2016;Qazilbash等人,2017)。与此一致,HLA Ligand Atlas中不存在任何TSA^{hi},其中包含在29个非恶性组织中识别的人MAP(<https://www.biorxiv.org/content/10.1101/778944v1>)。这支持靶向TSA^{lo}和58TSA^{hi}(无冗余)的安全性。TAA在至少一种正常组织中表达升高,而HSA表达仅限于造血区室。AML样本和MPC之间的FC比较表明,TSA^{hi}与TAA一起呈递最高的过表达(中位值为22倍),而HSA在健康细胞中以最高水平表达(中位值为0.6倍)(图4B)。总之,这些结果表明TSA^{hi}结合了两方面的优势:TSA的特异性/安全性和TAA的过表达。

[0302] 识别的TSA主要来源于据称基因组的非编码区,因为它们中仅有13%来源于典型蛋白质外显子,其中58%来源于内含子(图4C)。没有一个源自突变,与AML低突变负担一致(Lawrence等人,2013)。TAA主要来源于蛋白质编码外显子,而HSA起源也主要由非编码区主导,这与先前报道组织特异性内含子保留和ERE表达模式的研究一致(Middleton等人,2017;Larouche等人,2020)。虽然八个TSA^{hi}源自典型蛋白质编码基因,但鉴于它们在正常组织中相对于安全TAA的低表达,它们可被视为安全靶标(图4A)。支持它们作为治疗靶标的相关性,其中三个来自已知的AML生物标志物(LTBP1、MYCN和PLPPR3),其他五个具有未知功能或参与增殖、分化或耐药性(表3)。

[0303] 表3:已识别TSA^{hi}来源的典型蛋白质编码基因的特征

[0304]

基因符号	全名称	Ensembl ID	癌症相关注释
LTBP1	潜在转化生长因子 β 结合蛋白 1	ENSG00000049323	促进潜在 TGF- β 的分泌, 在 AML 中过表达(Wilson 等人, 2006)
MYCN	MYCN 原癌基因, bHLH 转录因子	ENSG00000134323	致白血病功能, 在 AML 中过表达(Liu 等人, 2017; Wilson 等人, 2006)
PLPPR3	磷脂磷酸酶相关 3	ENSG00000129951	ANPA32 致白血病基因的靶标, 在 AML 中过表达(Yang 等人, 2018)
GCSAML	生发中心相关的信号传导和运动	ENSG00000169224	与增殖相关, 该基因的甲基化在造血系统恶性肿瘤中发生改变(de Sá Machado Araújo 等人, 2018)
DNAH10	动力蛋白轴丝重链 10	ENSG00000197653	初级纤毛轴丝的形成, 与增殖和分化有关(Lagus 等人, 2019)

[0305]

ST8SIA6	ST8 α -N-乙酰神经氨酸 α -2,8-唾液酸转移酶 6	ENSG00000148488	多重耐药性(Zhang 等人, 2015)
LINC01835	C 型凝集素结构域家族 4 成员 O	ENSG00000267453	未知功能
MFSD2B	包含 2B 的主要促进子超家族结构域	ENSG00000205639	骨髓祖细胞中假定的表观遗传调节因子(Johnson 等人, 2015)

[0306] TSA的治疗价值部分取决于患者共享它的程度。为了评价原发性AML之间的TSA^{hi}共享,分析Leucegene队列,其中包括来自437名患者的纯化AML母细胞的RNA-seq数据(Lavallee等人,2015;Macrae等人,2013;Pabst等人,2016)。因为大多数MAP可以由不同的HLA同种异型呈递,所以首先通过考虑混杂结合物来评价已识别的TSA^{hi}的呈递。通过使用MHCcluster工具,它将呈递相似表位的HLA等位基因聚类在一起(Thomsen等人,2013),可以推断出能够呈递单个TSA^{hi}的全套HLA同种异型(表4)。基于这些数据,表明在世界人群中,99.92%的个体携带 ≥ 1 种HLA-I同种异型,能够呈递一种TSA^{hi}。接下来,认为只有当TSA编码转录物表达并且患者具有可以呈递该TSA的HLA同种异型时,单个TSA^{hi}才存在于给定的AML样本中。基于这些标准,可以预测在Leucegene队列中,每名患者TSA^{hi}的中位值为4,并且93.6%的患者出现至少一种TSA^{hi}(图4F)。

[0307] 表4:能够呈递MHCcluster预测的类似肽(混杂结合物)的HLA等位基因列表。

[0308]

HLA 等位基因	能够呈递类似肽的其他等位基因(混杂结合物)
HLA-A*01:01	
HLA-A*02:01	HLA-A*02:05 HLA-A*02:06 HLA-A*02:07
HLA-A*02:05	HLA-A*02:05 HLA-A*02:06 HLA-A*02:07
HLA-A*02:06	HLA-A*02:05 HLA-A*02:06 HLA-A*02:07
HLA-A*02:07	HLA-A*02:01 HLA-A*02:05 HLA-A*02:06
HLA-A*03:01	HLA-A*11:01
HLA-A*11:01	HLA-A*03:01 HLA-A*31:01 HLA-A*68:01
HLA-A*23:01	HLA-A*24:02
HLA-A*24:02	HLA-A*23:01

[0309]

HLA-A*25:01	HLA-A*26:01 HLA-A*66:01 HLA-B*15:02
HLA-A*26:01	HLA-A*25:01 HLA-A*66:01
HLA-A*29:02	HLA-A*30:02 HLA-B*15:02
HLA-A*30:01	
HLA-A*30:02	HLA-A*29:02
HLA-A*31:01	HLA-A*11:01 HLA-A*33:01 HLA-A*33:03 HLA-A*68:01
HLA-A*32:01	HLA-B*57:01 HLA-B*58:01
HLA-A*33:01	HLA-A*31:01 HLA-A*33:03 HLA-A*68:01
HLA-A*33:03	HLA-A*31:01 HLA-A*33:01 HLA-A*68:01
HLA-A*66:01	HLA-A*25:01 HLA-A*26:01
HLA-A*68:01	HLA-A*11:01 HLA-A*31:01 HLA-A*33:01 HLA-A*33:03
HLA-A*68:02	
HLA-B*07:02	HLA-B*35:02 HLA-B*35:03 HLA-B*55:01 HLA-B*56:01
HLA-B*08:01	
HLA-B*14:02	HLA-B*39:01
HLA-B*15:01	HLA-B*15:02 HLA-B*15:03 HLA-B*46:01
HLA-B*15:02	HLA-A*25:01 HLA-A*29:02 HLA-B*15:01 HLA-B*15:03 HLA-B*15:18 HLA-B*35:01 HLA-B*46:01
HLA-B*15:03	HLA-B*15:01 HLA-B*15:02 HLA-B*15:18
HLA-B*15:18	HLA-B*15:02 HLA-B*15:03
HLA-B*18:01	HLA-B*40:01 HLA-B*44:02 HLA-B*44:03 HLA-B*45:01
HLA-B*27:02	HLA-B*27:05 HLA-C*06:02 HLA-C*07:01
HLA-B*27:05	HLA-B*27:02
HLA-B*35:01	HLA-B*15:02 HLA-B*35:02 HLA-B*35:03 HLA-B*53:01
HLA-B*35:02	HLA-B*07:02 HLA-B*35:01 HLA-B*35:03 HLA-B*51:01 HLA-B*53:01 HLA-B*55:01 HLA-B*56:01
HLA-B*35:03	HLA-B*07:02 HLA-B*35:01 HLA-B*35:03 HLA-B*51:01 HLA-B*53:01 HLA-B*55:01 HLA-B*56:01
HLA-B*38:01	HLA-B*39:01
HLA-B*39:01	HLA-B*14:02 HLA-B*38:01
HLA-B*40:01	HLA-B*18:01 HLA-B*40:02 HLA-B*41:02 HLA-B*44:02 HLA-B*44:03 HLA-B*45:01
HLA-B*40:02	HLA-B*40:01 HLA-B*41:02 HLA-B*44:02 HLA-B*44:03 HLA-B*45:01
HLA-B*41:02	HLA-B*40:01 HLA-B*40:02 HLA-B*44:02 HLA-B*44:03 HLA-B*45:01
HLA-B*44:02	HLA-B*18:01 HLA-B*40:01 HLA-B*40:02 HLA-B*41:02 HLA-B*44:03 HLA-B*45:01
HLA-B*44:03	HLA-B*18:01 HLA-B*40:01 HLA-B*40:02 HLA-B*41:02 HLA-B*44:02 HLA-B*45:01
HLA-B*45:01	HLA-B*18:01 HLA-B*40:01 HLA-B*40:02 HLA-B*41:02 HLA-B*44:02 HLA-B*44:03

[0310]

HLA-B*46:01	HLA-B*15:01 HLA-B*15:02 HLA-C*03:02 HLA-C*03:03 HLA-C*03:04 HLA-C*08:01 HLA-C*12:02 HLA-C*12:03 HLA-C*16:01
HLA-B*51:01	HLA-B*35:02 HLA-B*35:03 HLA-B*52:01 HLA-B*53:01 HLA-B*55:01 HLA-B*56:01
HLA-B*52:01	HLA-B*51:01
HLA-B*53:01	HLA-B*35:01 HLA-B*35:02 HLA-B*35:03 HLA-B*51:01
HLA-B*55:01	HLA-B*07:02 HLA-B*35:02 HLA-B*35:03 HLA-B*51:01 HLA-B*56:01
HLA-B*56:01	HLA-B*07:02 HLA-B*35:02 HLA-B*35:03 HLA-B*51:01 HLA-B*55:01
HLA-B*57:01	HLA-A*32:01 HLA-B*58:01
HLA-B*58:01	HLA-A*32:01 HLA-B*57:01
HLA-C*03:02	HLA-B*46:01 HLA-C*03:03 HLA-C*03:04 HLA-C*08:01 HLA-C*12:02 HLA-C*12:03 HLA-C*16:01
HLA-C*03:03	HLA-B*46:01 HLA-C*03:02 HLA-C*03:04 HLA-C*08:01 HLA-C*08:02 HLA-C*12:02 HLA-C*12:03 HLA-C*15:02 HLA-C*16:01
HLA-C*03:04	HLA-B*46:01 HLA-C*03:02 HLA-C*03:04 HLA-C*08:01 HLA-C*08:02 HLA-C*12:02 HLA-C*12:03 HLA-C*15:02 HLA-C*16:01
HLA-C*04:01	HLA-C*07:02 HLA-C*14:02
HLA-C*05:01	HLA-C*08:01 HLA-C*08:02
HLA-C*06:02	HLA-B*27:02 HLA-C*07:01 HLA-C*07:02
HLA-C*07:01	HLA-B*27:02 HLA-C*06:02 HLA-C*07:02 HLA-C*14:02
HLA-C*07:02	HLA-C*04:01 HLA-C*06:02 HLA-C*07:01 HLA-C*14:02
HLA-C*08:01	HLA-B*46:01 HLA-C*03:02 HLA-C*03:03 HLA-C*03:04 HLA-C*05:01 HLA-C*08:02 HLA-C*12:02 HLA-C*12:03 HLA-C*15:02 HLA-C*16:01
HLA-C*08:02	HLA-C*03:03 HLA-C*03:04 HLA-C*05:01 HLA-C*08:01 HLA-C*15:02
HLA-C*12:02	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 HLA-C*16:01
HLA-C*12:03	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 HLA-C*16:01
HLA-C*14:02	HLA-C*04:01 HLA-C*07:01 HLA-C*07:02 HLA-C*16:01
HLA-C*15:02	HLA-C*03:03 HLA-C*03:04 HLA-C*08:01 HLA-C*08:02 HLA-C*12:02 HLA-C*12:03
HLA-C*16:01	HLA-B*46:01 HLA-C*03:02 HLA-C*03:03 HLA-C*03:04 HLA-C*08:01 HLA-C*12:02 HLA-C*12:03 HLA-C*14:02

[0311] 当比较在初始诊断时与复发时分析的AML样本中TSA^{hi}的数量(不匹配的样本)时,发现两组之间没有差异(图4G)。在另一项研究中比较诊断时和异基因造血细胞移植后复发时获得的AML母细胞匹配样本中患者HLA等位基因可能呈递的TSA^{hi}的RNA表达时,也没有发现差异(Toffalori等人,2019)(图4H)。由于白血病干细胞(LSC)是复发的主要介体(Shlush等人,2017),TSA^{hi}和HLA RNA表达也在从另一项研究获得的分类LSC相对于母细胞RNA-seq数据中进行评价(Corces等人,2016),并且发现两个细胞群之间没有差异(图4I-J)。尽管如此,通过使用基因集富集分析(GSEA),发现表达大量TSA^{hi}的患者也表达了更高水平的成熟LSC基因特征(Eppert等人,2011)(图4K)。总之,这些结果进一步支持了TSA^{hi}的高免疫原性,并证明它们可以针对几乎所有AML患者,无论是在诊断时还是复发时。因此可以得出结论,

可以在疾病的任何阶段设想TSA^{hi}的免疫靶向,并且将具有消除LSC的潜力。

[0312] 实施例6:大量TSA^{hi}的呈递与更好的生存率相关。

[0313] 接下来,检查了诊断时TSA^{hi}呈递对患者生存率的影响。值得注意的是,表达最高数量(上四分位数)的TSA^{hi}的患者比队列中的其余患者表现出明显更好的生存率(图5A)。在多变分析中,与多个TSA^{hi}的呈递相关的生存优势以及其他已知的预后因素(诸如年龄)、细胞遗传学风险、NPM1和FLT3-ITD突变仍然显著(图5B)。重要的是,独立于HLA_{pred}呈递进行的相同比较显示高表达和低表达之间没有差异(图11A,B),这表示TSA^{hi}的保护作用是HLA限制的。对TAA、HSA或TSA^{lo}进行的相同分析显示对生存率没有显著影响(图11C-H)。这些数据表明TSA^{hi}具有足够的免疫原性来引发自发的抗AML免疫应答。

[0314] 为了证明TSA^{hi}提供的生存优势是由于其累积的HLA_{pred}呈递,在随机移除越来越多的TSA^{hi}(58个中的1至29个)后,从分析(1000个随机排列/数值)中计算高相对于低TSA^{hi}患者的对数秩p值。相对于低表达者(所有其他患者),增加数量的TSA^{hi}的随机移除迅速导致高表达者(HLA-TSA^{hi}计数的上四分位数)缺少显著的生存优势(图5C-D)。随着个体TSA^{hi}的减少,生存优势的逐渐下降表明大多数TSA^{hi}有助于这种生存优势。更大比例的患者呈递的TSA^{hi}对p值的影响最大(图5E)。同样,从对数秩分析中移除常见HLA等位基因(超过5%的患者共享)对p值的影响大于移除低频等位基因(图5F)。总之,这些数据证明了TSA^{hi}呈递对患者生存率的HLA限制性益处。在接下来的实验中,研究了与TSA^{hi}呈递相关的生存优势的最简洁的解释:TSA^{hi}引发自发的保护性抗AML免疫应答。

[0315] 实施例7:TSA^{hi}呈递引发细胞毒性T细胞应答

[0316] 作为评价TSA^{hi}和其他MOI的免疫原性(即它们诱导免疫应答的能力)的前奏,使用Repitope,其是机器学习算法,它依赖于公共TCR数据库来预测T细胞应答的概率(Ogishi和Yotsuyanagi,2019)。使用胸腺上皮细胞(Adamopoulou等人,2013)或源自HIV的MAP分别作为阴性和阳性对照,Repitope预测表明TAA大多是非免疫原性的,而其他三组MOI与HIV肽一样具有免疫原性(图6A)。因此,相对于其他三个组和相对于在IEDB中报告为免疫原性的一组1411MAP,TAA在mTEC(~12.1rphm)中呈递高表达(图6B)。即使相对于其他免疫原性肽,非TAA MOI在mTEC中的RNA表达均非常低,这支持了它们的免疫原性。为了验证Repitope预测,以预期具有最高的免疫原性的HLA-A*02:01呈递的TSA^{hi}:ALPVALPSL开始进行体外T细胞测定。作为IFN-γELISpot的阳性对照,使用ELAGIGILTV表位是因为它是最具免疫原性的人MAP之一(Dutoit等人,2002;Hesnard等人,2016)。ALPVALPSL的免疫原性与ELAGIGILTV相似(图6C)。两个其他有前景的TSA^{hi}的ELISpot也支持它们的免疫原性(图6D)。对于这些TSA^{hi},还进行了细胞因子分泌测定和右旋体染色,这证实了ELISpot结果并支持了免疫应答的特异性(图6E-F)。为了进一步证明TSA^{hi}可以诱导自发和特异性T细胞克隆型扩增,进行特异性T细胞的功能扩增(FEST)测定,其中通过TCR测序分析用不同TSA^{hi}池刺激的外周血T细胞的短期培养(Danilova等人,2018)。5个测试的TSA^{hi}的每个池诱导9-10种不同克隆型的特异性扩增,支持它们的自发免疫原性(图6G和表5)。

[0317] 表5:响应TSA^{hi}池的特异性T细胞(FEST)测定的功能扩展。

池#	池中 TSA ^{hi}	显著扩增的克隆型(SEQ ID NO:)	FDR	比值比
1	SLLSGLLRA	CSARGDREYEQYF (191)	4.84E-52	Inf

[0319]	1	ALPVALPSL ALDPLLLRI IASPIALL SLDLLPLSI	CASTVVAGNSSPLHF (192)	1.2E-07	18.829
	1		CASSQDGIWGA YE QYF (193)	1.95E-07	Inf
	1		CASSVDAGGN YE QYF (194)	0.00066 5	12.293
	1		CASSLGGQGLSYGYTF (195)	0.01100 5	Inf
	1		CASSYRPNEQYF (196)	0.01100 5	Inf
	1		CASSGTDLNQPQHF (197)	0.03285 2	Inf
	1		CASSSNFEPLHF (198)	0.03285 2	Inf
	1		CASSWGGSN TGELFF (199)	0.03285 2	Inf
	2	LTDRIYLT VLFGGKVSG A LGISLT KY FNVALNARY TLNQGINVYI	CASSEYRALNTEAFF (200)	4.67E-30	59.789
	2		CASSES GTGGQPQHF (201)	1.14E-08	3.455
	2		CASSRTGENTE AFF (202)	0.00381 3	3.352
	2		CASSSTRQHYGYTF (203)	0.00705 2	3.764
	2		CASSDRTGGSSNEKLFF (204)	0.00803 1	8.213
	2		CATSRSGDSNQPQHF (205)	0.00803 1	8.213
	2		CASSYVLNTEAFF (206)	0.01028 4	10.837
	2		CASRESGQMNEKLFF (207)	0.01539 5	Inf
	2		CASSDGGEGTYGYTF (208)	0.02840 5	Inf
[0320]	2		CASTSWTGFGPNYGYTF (209)	0.02840 5	Inf
	3	LRSQILSY KILDVNLRI HSLISIVYL KLQDKEIGL AQDIILQAV	CASSADSSLGGYTF (210)	1.13E-10	45.434
	3		CSARDLAGGT YE QYF (211)	8.08E-10	Inf
	3		CASSPPTGEYEKLFF (212)	9.54E-09	Inf
	3		CARSFGGFF (213)	3.78E-06	Inf
	3		CASSLPSGILYE QYF (214)	3.78E-06	Inf
	3		CASAPGGMPYGYTF (215)	1.16E-05	Inf
	3		CASSLQD TDYNEQFF (216)	0.00526 5	Inf
	3		CASSDPGTSGVFTGELFF (217)	0.00907 3	Inf
	3		CSARLGTGELFF (218)	0.01680 6	Inf
	3		CASSLGRGYETQYF (219)	0.01996 1	3.55

[0321] 可以由HLA-A*02:01、-A*29:02、-B*15:01、-B27:05、-C*01:02、-C*03:04健康供体的HLA等位基因呈递的所有TSA^{hi}。TCR-seq由Adaptive Biotechnologies进行,原始数据使用FEST分析工具<http://www.stat-apps.onc.jhmi.edu/FEST>处理。此处报告了池的数量、每个池中存在的TSA^{hi}、每个池中显著扩增的每个克隆型的序列以及FEST分析工具提供的

FDR和比值比。

[0322] 接下来,由于“体内验证”,对来自437名Leucegene患者的转录组数据进行了深入分析,以评价T细胞对TSA^{hi}的体内识别潜力。首先,使用TRUST4算法(Zhang等人,2019)评价T细胞的TCR库多样性。与TAA(此处用作非免疫原性对照)相比,TSA^{hi}数量升高的_{pred}呈递与TCR库多样性降低有关,这表明抗TSA^{hi}克隆型的扩展(图6H)。为了证明这种扩展的特异性,使用ERGO算法来预测MOI-TCR相互作用(Springer等人,2020)。在识别抗MOI克隆型的ERGO概率>80%时,具有大量TSA^{hi}的患者在所有检测到的CDR3中也具有更高频率的抗TSA^{hi}克隆型(图6I)。对于TAA,未见类似的相关性(图6J)。接下来,计算能够识别由各个AML样本_{pred}呈递的MOI的抗MOI克隆型的比例(即同源TCR—MOI相互作用的频率)。该比例根据_{pred}呈递的MOI的数量进行归一化,因为否则呈递更多的MOI将自动导致检测到更高比例的抗_{pred}呈递的MOI克隆型。这表明TSA^{hi}_{pred}呈递与特异性T细胞识别频率显著高于TAA相关(图6K)。

[0323] 鉴于抗TSA^{hi}T细胞识别,推测TSA^{hi}_{pred}呈递应该与激活的CD8⁺T细胞的浸润有关。有意思的是,TSA^{hi}转录物的多样性与AML样本中的CD8A+CD8B表达呈负相关,而它们的_{pred}呈递的多样性不呈负相关(图6L-M)。这表明TSA^{hi}转录物的高度多样性反映了AML样本中略高的母细胞纯度(如TSA所预期)。为了避免这种可能的偏差,并且由于HLA-TSA^{hi}的数量在数学上与表达的TSA^{hi}的数量相关,将_{pred}呈递的TSA^{hi}的数量归一化为表达的TSA^{hi}转录物的数量,并在归一化_{pred}呈递的患者中分析差异基因表达高于中位值相比于低于中位值。值得注意的是,在与TSA^{hi}_{pred}呈正相关的123个基因中,有几个与T细胞激活和细胞溶解相关,包括CD8A、CD8B、GZMA、GZMB、IL2RB、PRF1和ZAP70(图6N)。值得注意的是,与这123个基因相关的GO terms仅与T细胞激活和分化相关(图6O)。CD4基因没有差异表达,并且没有GO term可能与下调基因显著相关。因此,得出的结论是,TSA^{hi}_{pred}呈递与较高丰度的激活CD8⁺T细胞有关。

[0324] 实施例8:TSA^{hi} RNA表达与免疫编辑、AML驱动突变和表观遗传畸变的体征有关

[0325] 鉴于TSA^{hi}的潜在治疗价值,有必要深入了解它们的生物发生。对于该分析,将高表达TSA^{hi}(HE-TSA^{hi})的计数分配给每名Leucegene患者,即在所有具有给定TSA^{hi}非空表达的患者中,他们的TSA^{hi}计数表达水平高于其中位表达。然后通过在每个蛋白质编码基因的表达和HE-TSA^{hi}计数之间进行成对Pearson相关来评估特定基因的表达是否可以与TSA^{hi}表达相关联。这表明参与MAP呈递的基因(HLA-A、HLA-B、HLA-C、B2M和NLRC5)的表达与HE-TSA^{hi}的数量之间存在一致的负相关,这表明免疫编辑响应于TSA^{hi}升高而发生表达(图7A和12A-B)。免疫编辑还得到了与CD47和CD84呈正相关的支持,CD47是免疫检查点分子,参与抑制树突状细胞吞噬作用(Majeti等人,2009),CD84促进白血病细胞表达PD-L1(Lewinsky等人,2018)。由于NPM1突变可以调节PD-L1(CD274)表达(Greiner等人,2017),因此分别分析了NPM1^{mut}和NPM1^{wt}患者。该分析显示,具有高于中位HE-TSA^{hi}计数的NPM1^{wt}患者表达的PD-L1水平显著高于具有较差HE-TSA^{hi}计数的患者(图7B)。

[0326] 接下来,分析与HE-TSA计数相关的基因途径(图7C和表6)。负相关途径包括参与细胞增殖(也包括转运和细胞组织)、线粒体OXPHOS和蛋白酶体介导的蛋白质分解代谢的生物过程。有意思的是,已显示抑制线粒体活性可降低MHC-I表达,并可用作癌细胞的免疫逃逸机制(Charni等人,2010)。类似地,抑制蛋白质降解可能导致MHC-I分子呈递的肽量减少,因此降低呈递TSA^{hi}的可能性(Tripathi等人,2016)。最后,有丝分裂相关过程的减少可能是

MHC-I下调的副作用,因为这两个过程均受NLRC5的调节(Wang等人,2019)。总之,这些数据表明TSA^{hi}表达与可能充当AML细胞免疫编辑机制的各种应答有关。

[0327] 与负相关途径相反,正相关途径仅限于调节过程(图7D)。因此,16.1%(相比于负相关的2.5%)的正相关基因是转录因子,可直接负责TSA^{hi}的转录。其中,相关性最好的基因是ZNF445(图7A),它是基因组印记的调节因子,即与DNA甲基化相关的表观遗传过程(Takahashi等人,2019)。由于ZNF445功能依赖于DNA甲基化,因此检查了TSA^{hi}表达与通常与DNA甲基化异常相关的AML突变之间的可能关联。首先测试了三种最常见的AML驱动突变(NPM1^{mut}、FLT3-ITD和DNMT3A^{mut}),发现这三种突变在表达高HE-TSA^{hi}计数(高于中位值)的患者中均显著富集(图7E)。此外,呈递两种或三种伴随突变的患者比呈递一种或没有一种突变的患者具有更高数量的HE-TSA^{hi}(图7F)。在MS分析中使用的19个AML试样中,12个呈递FLT3-ITD或NPM1突变。关于其他常见的AML突变,发现IDH2和双等位基因CEBPA突变也与HE-TSA^{hi}计数升高呈正相关,而ASXL1、SRSF2和U2AF1突变呈负相关,并且FLT3-TKD、IDH1、RUNX1、TET2、TP53和WT1不相关(图12D)。由于NPM1、DNMT3A、IDH2和CEBPA^{bi}突变与异常甲基化谱相关(Figueroa等人,2010a;Figueroa等人,2010b;Ley等人,2013),它们与HE-TSA^{hi}计数升高的相关性支持TSA^{hi}表达的表观遗传失调。

[0328] 表6:与Leucegene队列中的HE-TSA^{hi}计数呈正或负相关的GO terms列表。

	正相关/ 负相关?	GO term ID	p 值	搜索列表中的 命中计数	GO term 中的 基因 总数	名称
[0329]	负相关	278	5,63E-19	74	381	有丝分裂细胞周期
	负相关	9987	2,48E-13	599	9339	细胞过程
	负相关	51443	7,60E-13	27	71	泛素-蛋白连接酶活性的 正调控
	负相关	280	2,14E-12	48	232	核分裂
	负相关	7067	2,14E-12	48	232	有丝分裂

[0330]

负相关	7049	2,48E-12	100	795	细胞周期
负相关	51351	2,61E-12	27	74	连接酶活性的正调控
负相关	22402	2,72E-12	82	583	细胞周期过程
负相关	6091	5,20E-12	56	312	前体代谢物和能量的产生
负相关	87	7,37E-12	48	239	有丝分裂细胞周期的 M 期
负相关	51437	7,83E-12	25	65	参与有丝分裂细胞周期的泛素-蛋白连接酶活性的正调控
负相关	48285	1,04E-11	48	241	细胞器裂变
负相关	6996	2,30E-11	142	1377	细胞器组织
负相关	51301	2,86E-11	55	314	细胞分裂
负相关	51439	4,07E-11	25	69	参与有丝分裂细胞周期的泛素-蛋白连接酶活性的调控
负相关	51438	5,09E-11	27	82	泛素-蛋白连接酶活性的调控
负相关	51340	1,40E-10	27	85	连接酶活性的调控
负相关	51436	4,26E-10	23	63	参与有丝分裂细胞周期的泛素-蛋白连接酶活性的负调控
负相关	31398	5,15E-10	28	96	蛋白质泛素化的正调控
负相关	31145	6,32E-10	23	64	后期促进复合物依赖性蛋白酶体泛素依赖性蛋白质分解代谢过程
负相关	51352	1,97E-09	23	67	连接酶活性的负调控
负相关	51444	1,97E-09	23	67	泛素-蛋白连接酶活性的负调控
负相关	43161	2,40E-09	33	140	蛋白酶体泛素依赖性蛋白质分解代谢过程
负相关	10498	2,40E-09	33	140	蛋白酶体蛋白质分解代谢过程
负相关	31396	9,12E-09	30	122	蛋白质泛素化的调控
负相关	31397	1,24E-08	24	79	蛋白质泛素化的负调控
负相关	16043	1,47E-08	212	2557	细胞组成组织
负相关	279	4,28E-08	53	351	M 期

[0331]

负相关	31400	4,99E-08	31	138	蛋白质修饰过程的负调控
负相关	44237	5,13E-08	353	4974	细胞代谢过程
负相关	8152	1,63E-07	404	5938	代谢过程
负相关	22403	2,44E-07	59	435	细胞周期分期
负相关	8104	3,87E-07	97	921	蛋白质定位
负相关	45184	5,51E-07	85	767	建立蛋白质定位
负相关	15031	6,33E-07	84	756	蛋白质转运
负相关	32269	8,96E-07	36	200	细胞蛋白质代谢过程的负调控
负相关	51248	1,42E-06	37	213	蛋白质代谢过程的负调控
负相关	33036	3,32E-06	108	1111	大分子定位
负相关	6119	4,67E-06	24	102	氧化磷酸化
负相关	31401	4,68E-06	38	232	蛋白质修饰过程的正调控
负相关	51179	8,10E-06	227	2978	定位
负相关	16192	9,74E-06	68	589	囊泡介导的转运
负相关	9056	9,96E-06	99	1005	分解代谢过程
负相关	22900	1,21E-05	25	115	电子传递链
负相关	32270	3,08E-05	42	291	细胞蛋白质代谢过程的正调控
负相关	51234	4,69E-05	201	2606	建立定位
负相关	31399	5,75E-05	48	366	蛋白质修饰过程的调控
负相关	6511	7,23E-05	40	277	泛素依赖的蛋白质分解代谢过程
负相关	7059	8,48E-05	20	83	染色体分离
负相关	43632	1,21E-04	40	282	修饰依赖的大分子分解代谢过程
负相关	19941	1,21E-04	40	282	修饰依赖的蛋白质分解代谢过程
负相关	44092	1,50E-04	48	377	分子功能负调控
负相关	51247	1,69E-04	42	308	蛋白质代谢过程的正调控
负相关	9057	1,73E-04	58	502	大分子分解代谢过程
负相关	16044	2,11E-04	48	381	细胞膜组织
负相关	51246	2,15E-04	68	635	蛋白质代谢过程的调控

[0332]

负相关	51603	2,24E-04	42	311	参与细胞蛋白质分解代谢过程的蛋白水解
负相关	61024	2,29E-04	48	382	膜组织
负相关	32268	2,56E-04	62	559	细胞蛋白质代谢过程的调控
负相关	44257	2,97E-04	42	314	细胞蛋白质分解代谢过程
负相关	30163	3,43E-04	44	339	蛋白质分解代谢过程
负相关	51641	3,79E-04	92	977	细胞定位
负相关	44265	4,64E-04	52	440	细胞大分子分解代谢过程
负相关	46907	5,91E-04	69	665	细胞内转运
负相关	34613	7,00E-04	51	433	细胞蛋白质定位
负相关	51649	8,38E-04	83	865	在细胞中建立定位
负相关	70727	8,74E-04	51	436	细胞大分子定位
负相关	6810	9,25E-04	193	2572	转运
负相关	45333	1,17E-03	20	96	细胞呼吸
负相关	44248	1,28E-03	76	775	细胞分解代谢过程
负相关	42773	1,78E-03	15	57	ATP 合成耦合电子传递
负相关	42775	1,78E-03	15	57	线粒体 ATP 合成耦合电子传递
负相关	22904	1,91E-03	16	65	呼吸电子传递链
负相关	55114	2,07E-03	66	646	氧化还原
负相关	43086	6,03E-03	38	301	催化活性负调控
负相关	15980	6,23E-03	24	145	有机化合物氧化的能量衍生
负相关	7010	1,18E-02	49	448	细胞骨架组织
负相关	6886	1,75E-02	42	364	细胞内蛋白质转运
负相关	226	2,27E-02	23	145	微管细胞骨架组织
负相关	7017	2,67E-02	32	247	基于微管的过程
负相关	44093	2,95E-02	62	640	分子功能的正调控
负相关	16052	3,12E-02	19	107	碳水化合物分解代谢过程
负相关	33043	4,83E-02	31	242	细胞器组织的调控
负相关	48193	5,93E-02	21	132	高尔基囊泡转运
负相关	6508	7,37E-02	67	730	蛋白水解
负相关	65009	9,69E-02	89	1063	分子功能调控

[0333]

负相关	43085	1,04E-01	54	553	催化活性的正调控
负相关	7033	1,34E-01	11	44	液泡组织
负相关	7264	1,39E-01	34	292	小 GTPase 介导的信号转导
负相关	6007	1,64E-01	12	53	葡萄糖分解代谢过程
负相关	44238	1,93E-01	334	5270	初级代谢过程
负相关	16050	2,44E-01	12	55	囊泡组织
负相关	7040	2,93E-01	8	25	溶酶体组织
负相关	70585	2,97E-01	9	32	线粒体中的蛋白质定位
负相关	6626	2,97E-01	9	32	蛋白质靶向线粒体
负相关	910	3,26E-01	11	48	胞质分裂
负相关	46365	3,60E-01	13	66	单糖分解代谢过程
负相关	6122	3,78E-01	4	5	线粒体电子传递, 泛醇到细胞色素 c
负相关	51128	4,05E-01	51	538	细胞组成组织的调控
负相关	44267	5,10E-01	153	2146	细胞蛋白质代谢过程
负相关	6120	6,50E-01	10	43	线粒体电子传递, NADH 至泛醌
负相关	8064	7,23E-01	12	61	肌动蛋白聚合或解聚的调控
负相关	50790	7,39E-01	75	906	催化活性的调控
负相关	51656	8,00E-01	13	71	建立细胞器定位
负相关	70	8,23E-01	9	36	有丝分裂姐妹染色单体分离
负相关	44275	8,29E-01	14	81	细胞碳水化合物分解代谢过程
负相关	19320	1,01E+00	12	63	己糖分解代谢过程
负相关	30832	1,01E+00	12	63	肌动蛋白丝长度的调控
负相关	819	1,04E+00	9	37	姐妹染色单体分离
负相关	19538	1,13E+00	178	2604	蛋白质代谢过程
负相关	10324	1,18E+00	26	221	膜内陷
负相关	6897	1,18E+00	26	221	胞吞作用
负相关	32956	1,20E+00	15	94	肌动蛋白细胞骨架组织的调控
负相关	43254	1,20E+00	15	94	蛋白质复合物组装的调控
负相关	30833	1,23E+00	11	55	肌动蛋白丝聚合的调控

[0334]	负相关	10564	1,25E+00	19	138	细胞周期过程的调控
	负相关	7346	1,36E+00	22	174	有丝分裂细胞周期的调控
	负相关	60316	1,49E+00	3	3	兰尼碱敏感性钙释放通道活性的正调控
	负相关	51640	1,53E+00	15	96	细胞器定位
	负相关	32271	1,60E+00	12	66	蛋白质聚合的调控
	负相关	44283	1,92E+00	42	443	小分子生物合成过程
	负相关	32970	1,94E+00	15	98	基于肌动蛋白丝的过程的调控
	正相关	45449	1,70E-14	184	2621	转录调控
	正相关	19219	6,15E-14	201	3013	核碱基、核苷、核苷酸和核酸代谢过程的调控
	正相关	51171	7,28E-14	202	3039	氮化合物代谢过程的调控
	正相关	10468	6,82E-13	194	2926	基因表达调控
	正相关	10556	1,17E-12	191	2876	大分子生物合成过程的调控
	正相关	31326	4,39E-12	196	3020	细胞生物合成过程的调控
	正相关	9889	1,03E-11	196	3044	生物合成过程的调控
	正相关	80090	4,31E-11	217	3553	初级代谢过程的调控
	正相关	31323	1,80E-10	223	3735	细胞代谢过程的调控
	正相关	60255	1,97E-10	207	3375	大分子代谢过程的调控
	正相关	19222	6,73E-10	229	3917	代谢过程的调控
	正相关	51252	1,46E-08	130	1855	RNA 代谢过程的调控
	正相关	6355	2,28E-08	127	1806	转录调控, DNA 依赖性
	正相关	50909	5,95E-04	12	44	味觉的感官知觉

[0335] 最后,研究了TSA^{hi}表达是否与其他临床特征相关,从而可以预测它们在AML患者中的存在,诸如法国-美国-英国 (FAB) 类型 (图12E-H)。值得注意的是,在M1和M5 AML中,表达高HE-TSA^{hi}计数的患者分别过多和不足。因此,患有AML未成熟型和正常核型的患者表现出最高水平的TSA^{hi}。因此假设这是由于用于发现TSA的样本中FAB M1 AML的比例过高 (19个中的9个),并且由于大多数TSA^{hi}位于内含子区域 (37/58,包括ERE来源的TSA^{hi}位于内含子中),这可以通过FAB类型之间存在不同的内含子保留模式来解释。因此,对AML中特异性保留的内含子进行的无监督一致性聚类显示了根据他们的FAB类型的患者的明确聚类 (图7G)。总之,这些数据表明TSA^{hi}表达与特定于AML亚型的内含子保留模式有关。

[0336] 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“包含”用作开放式术语,基本上等同于短语“包括但不限于”。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。

[0337] 参考文献

[0338] Adamopoulou, E., Tenzer, S., Hillen, N., Klug, P., Rota, I. A., Tietz, S., Gebhardt, M., Stevanovic, S., Schild, H., Tolosa, E., et al. (2013). Exploring the MHC-

peptide matrix of central tolerance in the human thymus. *Nat Commun* 4,2039.

[0339] Andreatta,M.,and Nielsen,M.(2016).Gapped sequence alignment using artificial neural networks:application to the MHC class I system.*Bioinformatics*32,511-517.

[0340] Audemard,E.O.,Gendron,P.,Feghaly,A.,Lavallée,V.P.,Hébert,J.,Sauvageau,G.,and Lemieux,S.(2019).Targeted variant detection using unaligned RNA-Seq reads.*Life Sci Alliance* 2.

[0341] Audoux,J.,Philippe,N.,Chikhi,R.,Salson,M.,Gallopain,M.,Gabriel,M.,Le Coz,J.,Drouineau,E.,Commes,T.,and Gautheret,D.(2017).DE-kupl:exhaustive capture of biological variation in RNA-seq data through k-mer decomposition.*Genome Biology* 18,243.

[0342] Avigan,D.,and Rosenblatt,J.(2018).Vaccine therapy in hematologic malignancies.*Blood* 131,2640-2650.

[0343] Bamezai,S.,Rawat,V.P.,and Buske,C.(2012).Concise review:The Piwi-piRNA axis:pivotal beyond transposon silencing.*Stem Cells* 30,2603-2611.

[0344] Berlin,C.,Kowalewski,D.J.,Schuster,H.,Mirza,N.,Walz,S.,Handel,M.,Schmid-Horch,B.,Salih,H.R.,Kanz,L.,Rammensee,H.G.,et al.(2015).Mapping the HLA ligandome landscape of acute myeloid leukemia:a targeted approach toward peptide-based immunotherapy.*Leukemia* 29,647-659.

[0345] Boegel,S.,Lower,M.,Bukur,T.,Sorn,P.,Castle,J.C.,and Sahin,U.(2018).HLA and proteasome expression body map.*BMC Med Genomics* 11,36.

[0346] Bouwmeester,R.,Gabriels,R.,Hulstaert,N.,Martens,L.,and Degroeve,S.(2020).DeepLC can predict retention times for peptides that carry as-yet unseen modifications.*bioRxiv*,2020.2003.2028.013003.

[0347] Bray,N.L.,Pimentel,H.,Melsted,P.,and Pachter,L.(2016).Near-optimal probabilistic RNA-seq quantification.*Nature Biotechnology* 34,525-527.

[0348] Chapuis,A.G.,Egan,D.N.,Bar,M.,Schmitt,T.M.,McAfee,M.S.,Paulson,K.G.,Voillet,V.,Gottardo,R.,Ragnarsson,G.B.,Bleakley,M.,et al.(2019).T cell receptor gene therapy targeting WT1 prevents acute myeloid leukemia relapse post-transplant.*Nat Med* 25,1064-1072.

[0349] Charni,S.,de Bettignies,G.,Rathore,M.G.,Aguiló,J.I.,van den Elsen,P.J.,Haouzi,D.,Hipskind,R.A.,Enriquez,J.A.,Sanchez-Beato,M.,Pardo,J.,etal.(2010).Oxidative Phosphorylation Induces De Novo Expression of the MHC Class I in Tumor Cells through the ERK5 Pathway.*The Journal of Immunology* 185,3498-3503.

[0350] Chen,J.,Brunner,A.D.,Cogan,J.Z.,Nunez,J.K.,Fields,A.P.,Adamson,B.,Itzhak,D.N.,Li,J.Y.,Mann,M.,Leonetti,M.D.,and Weissman,J.S.(2020).Pervasive functional translation of noncanonical human open reading frames.*Science* 367,1140-1146.

- [0351] Chong,C.,Muller,M.,Pak,H.,Harnett,D.,Huber,F.,Grun,D.,Leleu,M.,Auger,A.,Arnaud,M.,Stevenson,B.J.,et al.(2020).Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes.Nat Commun 11,1293.
- [0352] Corces,M.R.,Buenrostro,J.D.,Wu,B.,Greenside,P.G.,Chan,S.M.,Koenig,J.L.,Snyder,M.P.,Pritchard,J.K.,Kundaje,A.,Greenleaf,W.J.,et al.(2016).Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution.Nat Genet 48,1193-1203.
- [0353] Cosma,G.L.,and Eisenlohr,L.C.(2019).Impact of epitope density on CD8(+)T cell development and function.Mol Immunol 113,120-125.
- [0354] Coulie,P.G.,Van den Eynde,B.J.,van der Bruggen,P.,and Boon,T.(2014).Tumour antigens recognized by T lymphocytes:at the core of cancer immunotherapy.Nat Rev Cancer 14,135-146.
- [0355] 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.
- [0356] Danilova,L.,Anagnostou,V.,Caushi,J.X.,Sidhom,J.W.,Guo,H.,Chan,H.Y.,Suri,P.,Tam,A.,Zhang,J.,Asmar,M.E.,et al.(2018).The Mutation-Associated Neoantigen Functional Expansion of Specific T Cells (MANAFEST) Assay:A Sensitive Platform for Monitoring Antitumor Immunity.Cancer Immunol Res 6, 888-899.
- [0357] Daouda,T.,Perreault,C.,and Lemieux,S.(2016).pyGeno:A Python package for precision medicine and proteogenomics.F1000Res 5,381.
- [0358] Di Stasi,A.,Jimenez,A.M.,Minagawa,K.,Al-Obaidi,M.,and Rezvani,K.(2015).Review of the Results of WT1 Peptide Vaccination Strategies for Myelodysplastic Syndromes and Acute Myeloid Leukemia from Nine Different Studies.Frontiers in Immunology 6.
- [0359] 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.
- [0360] Dutoit,V.,Rubio-Godoy,V.,Pittet,M.J.,Zippelius,A.,Dietrich,P.Y.,Legal,F.A.,Guillaume,P.,Romero,P.,Cerottini,J.C.,Houghten,R.A.,et al.(2002).Degeneracy of antigen recognition as the molecular basis for the high frequency of naive A2/Melan-a peptide multimer(+)CD8(+)T cells in humans.J Exp Med 196,207-216.
- [0361] Dvinge,H.,and Bradley,R.K.(2015).Widespread intron retention diversifies most cancer transcriptomes.Genome medicine 7,45-45.
- [0362] Efremova,M.,Finotello,F.,Rieder,D.,and Trajanoski,Z.(2017).Neoantigens Generated by Individual Mutations and Their Role in Cancer

Immunity and Immunotherapy.Front Immunol 8,1679.

[0363] Egen,J.G.,Ouyang,W.,and Wu,L.C.(2020).Human Anti-tumor Immunity: Insights from Immunotherapy Clinical Trials.Immunity 52,36-54.

[0364] Ehx,G.,and Perreault,C.(2019).Discovery and characterization of actionable tumor antigens.Genome Medicine 11,29.

[0365] Elias,J.E.,and Gygi,S.P.(2010).Target-decoy search strategy for mass spectrometry-based proteomics.Methods Mol Biol 604,55-71.

[0366] Eng,J.K.,Hoopmann,M.R.,Jahan,T.A.,Egertson,J.D.,Noble,W.S.,and MacCoss,M.J.(2015).A deeper look into Comet--implementation and features.J Am Soc Mass Spectrom 26,1865-1874.

[0367] Eppert,K.,Takenaka,K.,Lechman,E.R.,Waldron,L.,Nilsson,B.,van Galen,P.,Metzeler,K.H.,Poeppl,A.,Ling,V.,Beyene,J.,et al.(2011).Stem cell gene expression programs influence clinical outcome in human leukemia.Nat Med 17, 1086-1093.

[0368] Fennell,K.A.,Bell,C.C.,and Dawson,M.A.(2019).Epigenetic therapies in acute myeloid leukemia:where to from here?Blood 134,1891-1901.

[0369] Fergusson,J.R.,Morgan,M.D.,Bruchard,M.,Huitema,L.,Heesters,B.A.,van Unen,V.,van Hamburg,J.P.,van der Wel,N.N.,Picavet,D.,Koning,F.,et al.(2018).Maturing Human CD127+CCR7+PDL1+Dendritic Cells Express AIRE in the Absence of Tissue Restricted Antigens.Front Immunol 9,2902.

[0370] Figueroa,M.E.,Abdel-Wahab,O.,Lu,C.,Ward,P.S.,Patel,J.,Shih,A.,Li,Y.,Bhagwat,N.,Vasanthakumar,A.,Fernandez,H.F.,et al.(2010a).Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype,disrupt TET2 function, and impair hematopoietic differentiation.Cancer Cell 18,553-567.

[0371] Figueroa,M.E.,Lugthart,S.,Li,Y.,Erpelinck-Verschueren,C.,Deng,X.,Christos,P.J.,Schifano,E.,Booth,J.,van Putten,W.,Skrabanek,L.,et al.(2010b).DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia.Cancer Cell 17,13-27.

[0372] Garrison,E.,and Marth,G.(2012).Haplotype-based variant detection from short-read sequencing.arXiv:Genomics.

[0373] Gaujoux,R.,and Seoighe,C.(2010).A flexible R package for nonnegative matrix factorization.BMC Bioinformatics 11,367.

[0374] Greiner,J.,Hofmann,S.,Schmitt,M.,Götz,M.,Wiesneth,M.,Schrezenmeier,H.,Bunjes,D.,Döhner,H.,and Bullinger,L.(2017).Acute myeloid leukemia with mutated nucleophosmin 1:an immunogenic acute myeloid leukemia subtype and potential candidate for immune checkpoint inhibition.Haematologica 102,e499-e501.

[0375] Gu,M.,Zwiebel,M.,Ong,S.H.,Boughton,N.,Nomdedeu,J.,Basheer,F.,Nannya,Y.,Quiros,P.M.,Ogawa,S.,Cazzola,M.,et al.(2020).RNAmut:robust identification of somatic mutations in acute myeloid leukemia using RNA-

sequencing. *Haematologica* 105,e290-e293.

[0376] Gutierrez,S.E.,and Romero-Oliva,F.A.(2013).Epigenetic changes:a common theme in acute myelogenous leukemogenesis.*J Hematol Oncol* 6,57.

[0377] Hardy,M.-P.,Vincent,K.,and Perreault,C.(2019).The Genomic Landscape of Antigenic Targets for T Cell-Based Leukemia Immunotherapy.*Frontiers in Immunology* 10,2934.

[0378] He,H.,Kondo,Y.,Ishiyama,K.,Alatrash,G.,Lu,S.,Cox,K.,Qiao,N.,Clise-Dwyer,K.,St John,L.,Sukhumalchandra,P.,et al.(2020).Two unique HLA-A*0201 restricted peptides derived from cyclin E as immunotherapeutic targets in leukemia.*Leukemia*.

[0379] Hesnard,L.,Legoux,F.,Gautreau,L.,Moyon,M.,Baron,O.,Devilder,M.C.,Bonneville,M.,and Saulquin,X.(2016).Role of the MHC restriction during maturation of antigen-specific human T cells in the thymus.*Eur J Immunol* 46, 560-569.

[0380] Janelle,V.,Carli,C.,Taillefer,J.,Orio,J.,and Delisle,J.S.(2015).Defining novel parameters for the optimal priming and expansion of minor histocompatibility antigen-specific T cells in culture.*J Transl Med* 13,123.

[0381] Jung,N.,Dai,B.,Gentles,A.J.,Majeti,R.,and Feinberg,A.P.(2015).An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis.*Nature Communications* 6,8489.

[0382] Knaus,H.A.,Berglund,S.,Hackl,H.,Blackford,A.L.,Zeidner,J.F.,Montiel-Esparza,R.,Mukhopadhyay,R.,Vanura,K.,Blazar,B.R.,Karp,J.E.,et al.(2018).Signatures of CD8+T cell dysfunction in AML patients and their reversibility with response to chemotherapy.*JCI Insight* 3.

[0383] Krokhin,O.V.(2006).Sequence-specific retention calculator.Algorithm for peptide retention prediction in ion-pair RP-HPLC:application to 300-and 100-A pore size C18 sorbents.*Anal Chem* 78,7785-7795.

[0384] Lamoliatte,F.,McManus,F.P.,Maarifi,G.,Chelbi-Alix,M.K.,and Thibault,P.(2017).Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification.*Nat Commun* 8,14109.

[0385] Larouche,J.D.,Trofimov,A.,Hesnard,L.,Ehx,G.,Zhao,Q.,Vincent,K.,Durette,C.,Gendron,P.,Laverdure,J.P.,Bonneil,E.,et al.(2020).Widespread and tissue-specific expression of endogenous retroelements in human somatic tissues.*Genome Med* 12,40.

[0386] Laumont,C.M.,Daouda,T.,Laverdure,J.P.,Bonneil,E.,Caron-Lizotte,O.,Hardy,M.P.,Granados,D.P.,Durette,C.,Lemieux,S.,Thibault,P.,and Perreault,C.(2016).Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames.*Nat Commun* 7,10238.

[0387] Laumont,C.M.,Vincent,K.,Hesnard,L.,Audemard,E.,Bonneil,E.,Laverdure,

J.P., Gendron, P., Courcelles, M., Hardy, M.P., Cote, C., et al. (2018). Noncoding regions are the main source of targetable tumor-specific antigens. *Sci Transl Med* 10.

[0388] Lavalley, V.P., Baccelli, I., Kros, J., Wilhelm, B., Barabe, F., Gendron, P., Boucher, G., Lemieux, S., Marinier, A., Meloche, S., et al. (2015). The transcriptomic landscape and directed chemical interrogation of MLL-rearranged acute myeloid leukemias. *Nat Genet* 47, 1030-1037.

[0389] Lavalley, V.P., Kros, J., Lemieux, S., Boucher, G., Gendron, P., Pabst, C., Boivin, I., Marinier, A., Guidos, C.J., Meloche, S., et al. (2016). Chemo-genomic interrogation of CEBPA mutated AML reveals recurrent CSF3R mutations and subgroup sensitivity to JAK inhibitors. *Blood* 127, 3054-3061.

[0390] Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214-218.

[0391] Legat, A., Maby-El Hajjami, H., Baumgaertner, P., Cagnon, L., Abed Maillard, S., Geldhof, C., Iancu, E.M., Lebon, L., Guillaume, P., Dojcinovic, D., et al. (2016). Vaccination with LAG-3Ig (IMP321) and Peptides Induces Specific CD4 and CD8 T-Cell Responses in Metastatic Melanoma Patients--Report of a Phase I/IIa Clinical Trial. *Clin Cancer Res* 22, 1330-1340.

[0392] Lewinsky, H., Barak, A.F., Huber, V., Kramer, M.P., Radomir, L., Sever, L., Orr, I., Mirkin, V., DeZorella, N., Shapiro, M., et al. (2018). CD84 regulates PD-1/PD-L1 expression and function in chronic lymphocytic leukemia. *J Clin Invest* 128, 5465-5478.

[0393] Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., Robertson, A., Hoadley, K., Triche, T.J., Jr., Laird, P.W., Baty, J.D., et al. (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368, 2059-2074.

[0394] Li, B., Li, T., Wang, B., Dou, R., Zhang, J., Liu, J.S., and Liu, X.S. (2017). Ultrasensitive detection of TCR hypervariable-region sequences in solid-tissue RNA-seq data. *Nat Genet* 49, 482-483.

[0395] Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987-2993.

[0396] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.

[0397] Li, S., Garrett-Bakelman, F.E., Chung, S.S., Sanders, M.A., Hricik, T., Rapaport, F., Patel, J., Dillon, R., Vijay, P., Brown, A.L., et al. (2016). Distinct

evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nat Med* 22,792-799.

[0398] Loffler, M.W., Mohr, C., Bichmann, L., Freudenmann, L.K., Walzer, M., Schroeder, C.M., Trautwein, N., Hilke, F.J., Zinser, R.S., Muhlenbruch, L., et al. (2019). Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. *Genome Med* 11,28.

[0399] Logtenberg, M.E.W., Scheeren, F.A., and Schumacher, T.N. (2020). The CD47-SIRPα Immune Checkpoint. *Immunity* 52,742-752.

[0400] Luo, K., Yuan, J., Shan, Y., Li, J., Xu, M., Cui, Y., Tang, W., Wan, B., Zhang, N., Wu, Y., and Yu, L. (2006). Activation of transcriptional activities of AP1 and SRE by a novel zinc finger protein ZNF445. *Gene* 367,89-100.

[0401] Macrae, T., Sargeant, T., Lemieux, S., Hebert, J., Deneault, E., and Sauvageau, G. (2013). RNA-Seq reveals spliceosome and proteasome genes as most consistent transcripts in human cancer cells. *PLoS One* 8,e72884.

[0402] Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21,3448-3449.

[0403] Maiga, A., Lemieux, S., Pabst, C., Lavallee, V.P., Bouvier, M., Sauvageau, G., and Hebert, J. (2016). Transcriptome analysis of G protein-coupled receptors in distinct genetic subgroups of acute myeloid leukemia: identification of potential disease-specific targets. *Blood Cancer J* 6,e431.

[0404] Marcais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27,764-770.

[0405] Maslak, P.G., Dao, T., Bernal, Y., Chanel, S.M., Zhang, R., Frattini, M., Rosenblat, T., Jurcic, J.G., Brentjens, R.J., Arcila, M.E., et al. (2018). Phase 2 trial of a multivalent WT1 peptide vaccine (galinpepimut-S) in acute myeloid leukemia. *Blood Adv* 2,224-234.

[0406] Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 5,e13984.

[0407] Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J.J., Bomane, A., Cosson, B., Eyraes, E., Rasko, J.E., and Ritchie, W. (2017). IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol* 18,51.

[0408] Ntziachristos, P., Abdel-Wahab, O., and Aifantis, I. (2016). Emerging concepts of epigenetic dysregulation in hematological malignancies. *Nat Immunol* 17,1016-1024.

[0409] Ogishi, M., and Yotsuyanagi, H. (2019). Quantitative Prediction of the Landscape of T Cell Epitope Immunogenicity in Sequence Space. *Frontiers in*

Immunology 10.

[0410] Pabst,C.,Bergeron,A.,Lavallee,V.P.,Yeh,J.,Gendron,P.,Norddahl,G.L.,Krosl,J.,Boivin,I.,Deneault,E.,Simard,J.,et al.(2016).GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo.Blood 127,2018-2027.

[0411] Papaemmanuil,E.,Gerstung,M.,Bullinger,L.,Gaidzik,V.I.,Paschka,P.,Roberts,N.D.,Potter,N.E.,Heuser,M.,Thol,F.,Bolli,N.,et al.(2016).Genomic Classification and Prognosis in Acute Myeloid Leukemia.N Engl J Med 374,2209-2221.

[0412] Pearson,H.,Daouda,T.,Granados,D.P.,Durette,C.,Bonneil,E.,Courcelles,M.,Rodenbrock,A.,Laverdure,J.P.,Cote,C.,Mader,S.,et al.(2016).MHC class I-associated peptides derive from selective regions of the human genome.J Clin Invest 126,4690-4701.

[0413] Qazilbash,M.H.,Wieder,E.,Thall,P.F.,Wang,X.,Rios,R.,Lu,S.,Kanodia,S.,Ruisaard,K.E.,Giralt,S.A.,Estey,E.H.,et al.(2017).PR1 peptide vaccine induces specific immunity with clinical responses in myeloid malignancies.Leukemia 31,697-704.

[0414] Quinlan,A.R.,and Hall,I.M.(2010).BEDTools:a flexible suite of utilities for comparing genomic features.Bioinformatics 26,841-842.

[0415] Rashidi,A.,and Walter,R.B.(2016).Antigen-specific immunotherapy for acute myeloid leukemia:where are we now,and where do we go from here?Expert Rev Hematol 9,335-350.

[0416] Ritchie,M.E.,Phipson,B.,Wu,D.,Hu,Y.,Law,C.W.,Shi,W.,and Smyth,G.K.(2015).limma powers differential expression analyses for RNA-sequencing and microarray studies.Nucleic Acids Res 43,e47.

[0417] Robinson,J.T.,Thorvaldsdottir,H.,Winckler,W.,Guttman,M.,Lander,E.S.,Getz,G.,and Mesirov,J.P.(2011).Integrative genomics viewer.Nat Biotechnol 29,24-26.

[0418] Robinson,M.D.,McCarthy,D.J.,and Smyth,G.K.(2010).edgeR:a Bioconductor package for differential expression analysis of digital gene expression data.Bioinformatics 26,139-140.

[0419] Rogers,M.F.,Shihab,H.A.,Mort,M.,Cooper,D.N.,Gaunt,T.R.,and Campbell,C.(2018).FATHMM-XF:accurate prediction of pathogenic point mutations via extended features.Bioinformatics 34,511-513.

[0420] 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.,et al.(2020).A large peptidome dataset improves HLA class I epitope prediction across most of the human population.Nat Biotechnol 38,199-209.

[0421] Schulz,W.A.,Steinhoff,C.,and Florl,A.R.(2006).Methylation of

endogenous human retroelements in health and disease. *Curr Top Microbiol Immunol* 310,211-250.

[0422] Shannon,P.,Markiel,A.,Ozier,O.,Baliga,N.S.,Wang,J.T.,Ramage,D.,Amin,N.,Schwikowski,B.,and Ideker,T. (2003).Cytoscape:a software environment for integrated models of biomolecular interaction networks.*Genome Res* 13,2498-2504.

[0423] Shao,W.,Pedrioli,P.G.A.,Wolski,W.,Scurtescu,C.,Schmid,E.,Vizcaino,J.A.,Courcelles,M.,Schuster,H.,Kowalewski,D.,Marino,F.,et al. (2018).The SystemMHC Atlas project.*Nucleic Acids Res* 46,D1237-d1247.

[0424] Shlush,L.I.,Mitchell,A.,Heisler,L.,Abelson,S.,Ng,S.W.K.,Trotman-Grant,A.,Medeiros,J.J.F.,Rao-Bhatia,A.,Jaciw-Zurakowsky,I.,Marke,R.,et al. (2017).Tracing the origins of relapse in acute myeloid leukaemia to stem cells.*Nature* 547,104-108.

[0425] Smart,A.C.,Margolis,C.A.,Pimentel,H.,He,M.X.,Miao,D.,Adeegbe,D.,Fugmann,T.,Wong,K.-K.,and Van Allen,E.M. (2018).Intron retention is a source of neoepitopes in cancer.*Nature Biotechnology* 36,1056-1058.

[0426] 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.

[0427] Springer,I.,Besser,H.,Tickotsky-Moskovitz,N.,Dvorkin,S.,and Louzoun,Y. (2020).Prediction of Specific TCR-Peptide Binding From Large Dictionaries of TCR-Peptide Pairs.*Front Immunol* 11,1803.

[0428] 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.

[0429] Takahashi,N.,Coluccio,A.,Thorball,C.W.,Planet,E.,Shi,H.,Offner,S.,Turelli,P.,Imbeault,M.,Ferguson-Smith,A.C.,and Trono,D. (2019).ZNF445 is a primary regulator of genomic imprinting.*Genes Dev* 33,49-54.

[0430] Tamura,H.,Dan,K.,Tamada,K.,Nakamura,K.,Shioi,Y.,Hyodo,H.,Wang,S.D.,Dong,H.,Chen,L.,and Ogata,K. (2005).Expression of functional B7-H2 and B7.2 costimulatory molecules and their prognostic implications in de novo acute myeloid leukemia.*Clin Cancer Res* 11,5708-5717.

[0431] Tate,J.G.,Bamford,S.,Jubb,H.C.,Sondka,Z.,Beare,D.M.,Bindal,N.,Boutselakis,H.,Cole,C.G.,Creatore,C.,Dawson,E.,et al. (2018).COSMIC:the Catalogue Of Somatic Mutations In Cancer.*Nucleic Acids Research* 47,D941-D947.

[0432] Thomsen,M.,Lundegaard,C.,Buus,S.,Lund,O.,and Nielsen,M. (2013).MHCcluster,a method for functional clustering of MHC molecules.*Immunogenetics* 65,655-665.

[0433] Toffalori,C.,Zito,L.,Gambacorta,V.,Riba,M.,Oliveira,G.,Bucci,G.,

Barcella,M.,Spinelli,O.,Greco,R.,Crucitti,L.,et al.(2019).Immune signature drives leukemia escape and relapse after hematopoietic cell transplantation.Nat Med 25,603-611.

[0434] Tripathi,S.C.,Peters,H.L.,Taguchi,A.,Katayama,H.,Wang,H.,Momin,A.,Jolly,M.K.,Celiktas,M.,Rodriguez-Canales,J.,Liu,H.,et al.(2016).Immunoproteasome deficiency is a feature of non-small cell lung cancer with a mesenchymal phenotype and is associated with a poor outcome.Proc Natl Acad Sci U S A 113,E1555-1564.

[0435] van der Lee,D.I.,Reijmers,R.M.,Honders,M.W.,Hagedoorn,R.S.,de Jong,R.C.,Kester,M.G.,van der Steen,D.M.,de Ru,A.H.,Kweekel,C.,Bijen,H.M.,et al.(2019).Mutated nucleophosmin 1 as immunotherapy target in acute myeloid leukemia.J Clin Invest 129,774-785.

[0436] Vasu,S.,Kohlschmidt,J.,Mrozek,K.,Eisfeld,A.K.,Nicolet,D.,Sterling,L.J.,Becker,H.,Metzeler,K.H.,Papaioannou,D.,Powell,B.L.,et al.(2018).Ten-year outcome of patients with acute myeloid leukemia not treated with allogeneic transplantation in first complete remission.Blood Adv 2,1645-1650.

[0437] Vizcaino,J.A.,Csordas,A.,del-Toro,N.,Dianes,J.A.,Griss,J.,Lavidas,I.,Mayer,G.,Perez-Riverol,Y.,Reisinger,F.,Ternent,T.,et al.(2016).2016update of the PRIDE database and its related tools.Nucleic Acids Res 44,D447-456.

[0438] Wang,E.,Lu,S.X.,Pastore,A.,Chen,X.,Imig,J.,Chun-Wei Lee,S.,Hockemeyer,K.,Ghebrechristos,Y.E.,Yoshimi,A.,Inoue,D.,et al.(2019a).Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia.Cancer Cell 35,369-384.e367.

[0439] Wang,Q.,Ding,H.,He,Y.,Li,X.,Cheng,Y.,Xu,Q.,Yang,Y.,Liao,G.,Meng,X.,Huang,C.,and Li,J.(2019b).NLRC5 mediates cell proliferation,migration,and invasion by regulating the Wnt/beta-catenin signalling pathway in clear cell renal cell carcinoma.Cancer Lett 444,9-19.

[0440] Whiteway,A.,Corbett,T.,Anderson,R.,Macdonald,I.,and Prentice,H.G.(2003).Expression of co-stimulatory molecules on acute myeloid leukaemia blasts may effect duration of first remission.Br J Haematol 120,442-451.

[0441] Wong,J.J.L.,Gao,D.,Nguyen,T.V.,Kwok,C.-T.,van Geldermalsen,M.,Middleton,R.,Pinello,N.,Thoeng,A.,Nagarajah,R.,Holst,J.,et al.(2017).Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment.Nature Communications 8,15134.

[0442] Wu,T.D.,and Nacu,S.(2010).Fast and SNP-tolerant detection of complex variants and splicing in short reads.Bioinformatics 26,873-881.

[0443] Yang,L.,Rau,R.,and Goodell,M.A.(2015).DNMT3A in haematological malignancies.Nat Rev Cancer 15,152-165.

[0444] Zhang,J.,Hu,X.,Wang,J.,Sahu,A.D.,Cohen,D.,Song,L.,Ouyang,Z.,Fan,J.,

Wang,B.,Fu,J.,et al.(2019).Immune receptor repertoires in pediatric and adult acute myeloid leukemia.Genome Med 11,73.

[0445] Zhao,Q.,Laverdure,J.P.,Lanoix,J.,Durette,C.,Cote,C.,BonNeil,E.,Laumont,C.M.,Gendron,P.,Vincent,K.,Courcelles,M.,et al.(2020).Proteogenomics Uncovers a Vast Repertoire of Shared Tumor-Specific Antigens in Ovarian Cancer.Cancer Immunol Res.

[0446] Zhi,H.,Ning,S.,Li,X.,Li,Y.,Wu,W.,and Li,X.(2014).A novel reannotation strategy for dissecting DNA methylation patterns of human long intergenic non-coding RNAs in cancers.Nucleic Acids Res 42,8258-8270.

[0447] Zhou,J.,and Chng,W.J.(2017).Aberrant RNA splicing and mutations in spliceosome complex in acute myeloid leukemia.Stem Cell Investig 4,6.

[0448] Zhou,Y.,Lu,Y.,and Tian,W.(2012).Epigenetic features are significantly associated with alternative splicing.BMC Genomics 13,123.

[0449] Wilson CS,Davidson GS,Martin SB,Andries E,Potter J,Harvey R,et al.Gene expression profiling of adult acute myeloid leukemia identifies novel biologic clusters for risk classification and outcome prediction.Blood 2006; 108(2):685-696.

[0450] Liu L,Xu F,Chang C-K,He Q,Wu L-Y,Zhang Z,et al.MYCN contributes to the malignant characteristics of erythroleukemia through EZH2-mediated epigenetic repression of p21.Cell Death&Disease 20172017/10/01;8(10):e3126-e3126.

[0451] Yang X,Lu B,Sun X,Han C,Fu C,Xu K,et al.ANP32A regulates histone H3 acetylation and promotes leukemogenesis.Leukemia 2018 2018/07/01;32(7):1587-1597.

[0452] de SáMachado Araújo G,da Silva Francisco Junior R,dos Santos Ferreira C,Mozier Rodrigues PT,Terra Machado D,Louvain de Souza T,et al.Maternal 5mCpG Imprints at the PARD6G-AS1 and GCSAML Differentially Methylated Regions Are Decoupled From Parent-of-Origin Expression Effects in Multiple Human Tissues.Frontiers in Genetics 2018 2018-March-01;9(36).

[0453] Lagus H,Klaas M,Juteau S,Elomaa O,Kere J,Vuola J,et al.Discovery of increased epidermal DNAH10 expression after regeneration of dermis in a randomized with-in person trial—reflections on psoriatic inflammation.Scientific Reports 2019 2019/12/13;9(1):19136.

[0454] Zhang X,Dong W,Zhou H,Li H,Wang N,Miao X,et al.alpha-2,8-Sialyltransferase Is Involved in the Development of Multidrug Resistance via PI3K/Akt Pathway in Human Chronic Myeloid Leukemia.IUBMB Life 2015 Feb;67(2): 77-87.

[0455] Johnson KD,Kong G,Gao X,Chang Y-I,Hewitt KJ,Sanalkumar R,et al.Cis-regulatory mechanisms governing stem and progenitor cell transitions.Science

advances 2015;1 (8) :e1500503-e1500503。

序列表

<110> 蒙特利尔大学 (UNIVERSITÉ DE MONTRÉAL)

<120> 用于急性髓系白血病 (AML) 的新型肿瘤特异性抗原及其用途

<130> PPI22171742CA

<150> 63/009,853

<151> 2020-04-14

<160> 219

<170> PatentIn 3.5版

<210> 1

<211> 9

<212> PRT

<213> 智人

<400> 1

Arg Gln Ile Ser Val Gln Ala Ser Leu

1 5

<210> 2

<211> 9

<212> PRT

<213> 智人

<400> 2

Asp Arg Glu Leu Arg Asn Leu Glu Leu

1 5

<210> 3

<211> 9

<212> PRT

<213> 智人

<400> 3

Gly Ala Arg Gln Gln Ile His Ser Trp

1 5

<210> 4

<211> 8

<212> PRT

<213> 智人

<400> 4

Ser Gly Lys Leu Arg Val Ala Leu

1 5

<210> 5

<211> 11

<212> PRT
<213> 智人
<400> 5
Arg Ser Ala Ser Ser Ala Thr Gln Val His Lys
1 5 10
<210> 6
<211> 10
<212> PRT
<213> 智人
<400> 6
Ser Ala Ser Ser Ala Thr Gln Val His Lys
1 5 10
<210> 7
<211> 9
<212> PRT
<213> 智人
<400> 7
Phe Leu Leu Glu Phe Lys Pro Val Ser
1 5
<210> 8
<211> 8
<212> PRT
<213> 智人
<400> 8
Gly Pro Gln Val Arg Gly Ser Ile
1 5
<210> 9
<211> 8
<212> PRT
<213> 智人
<400> 9
Ile Arg Met Lys Ala Gln Ala Leu
1 5
<210> 10
<211> 9
<212> PRT
<213> 智人
<400> 10
Lys Ile Lys Val Phe Ser Lys Val Tyr

1 5
<210> 11
<211> 10
<212> PRT
<213> 智人
<400> 11
Leu Leu Ser Arg Gly Leu Leu Phe Arg Ile
1 5 10
<210> 12
<211> 9
<212> PRT
<213> 智人
<400> 12
Leu Pro Ile Ala Ser Ala Ser Leu Leu
1 5
<210> 13
<211> 9
<212> PRT
<213> 智人
<400> 13
Leu Tyr Phe Leu Gly His Gly Ser Ile
1 5
<210> 14
<211> 9
<212> PRT
<213> 智人
<400> 14
Asn Pro Leu Gln Leu Ser Leu Ser Ile
1 5
<210> 15
<211> 8
<212> PRT
<213> 智人
<400> 15
Asp Leu Met Leu Arg Glu Ser Leu
1 5
<210> 16
<211> 8
<212> PRT

<213> 智人
<400> 16
Val Thr Phe Lys Leu Ser Leu Phe
1 5
<210> 17
<211> 8
<212> PRT
<213> 智人
<400> 17
Ile Ala Leu Tyr Lys Gln Val Leu
1 5
<210> 18
<211> 9
<212> PRT
<213> 智人
<400> 18
Ile Val Ala Thr Gly Ser Leu Leu Lys
1 5
<210> 19
<211> 9
<212> PRT
<213> 智人
<400> 19
Lys Ile Lys Asn Lys Thr Lys Asn Lys
1 5
<210> 20
<211> 9
<212> PRT
<213> 智人
<400> 20
Lys Leu Leu Ser Leu Thr Ile Tyr Lys
1 5
<210> 21
<211> 8
<212> PRT
<213> 智人
<400> 21
Asn Ile Leu Lys Lys Thr Val Leu
1 5

<210> 22

<211> 8

<212> PRT

<213> 智人

<400> 22

Asn Pro Lys Leu Lys Asp Ile Leu

1 5

<210> 23

<211> 8

<212> PRT

<213> 智人

<400> 23

Asn Gln Lys Lys Val Arg Ile Leu

1 5

<210> 24

<211> 10

<212> PRT

<213> 智人

<400> 24

Pro Phe Pro Leu Val Gln Val Glu Pro Val

1 5 10

<210> 25

<211> 8

<212> PRT

<213> 智人

<400> 25

Ser Pro Gln Ser Gly Pro Ala Leu

1 5

<210> 26

<211> 9

<212> PRT

<213> 智人

<400> 26

Thr Ser Arg Leu Pro Lys Ile Gln Lys

1 5

<210> 27

<211> 9

<212> PRT

<213> 智人

<400> 27

Leu Leu Asp Asn Ile Leu Gln Ser Ile

1 5

<210> 28

<211> 9

<212> PRT

<213> 智人

<400> 28

Arg Leu Glu Val Arg Lys Val Ile Leu

1 5

<210> 29

<211> 9

<212> PRT

<213> 智人

<400> 29

Leu Ser Trp Gly Tyr Phe Leu Phe Lys

1 5

<210> 30

<211> 9

<212> PRT

<213> 智人

<400> 30

Thr Ile Leu Pro Arg Ile Leu Thr Leu

1 5

<210> 31

<211> 8

<212> PRT

<213> 智人

<400> 31

Glu Gly Lys Ile Lys Arg Asn Ile

1 5

<210> 32

<211> 11

<212> PRT

<213> 智人

<400> 32

Phe Leu Ala Ser Phe Val Glu Lys Thr Val Leu

1 5 10

<210> 33

<211> 9
<212> PRT
<213> 智人
<400> 33
Ile Leu Ala Ser His Asn Leu Thr Val
1 5
<210> 34
<211> 9
<212> PRT
<213> 智人
<400> 34
Ile Gln Leu Thr Ser Val His Leu Leu
1 5
<210> 35
<211> 10
<212> PRT
<213> 智人
<400> 35
Leu Glu Leu Ile Ser Phe Leu Pro Val Leu
1 5 10
<210> 36
<211> 9
<212> PRT
<213> 智人
<400> 36
Asn Phe Cys Met Leu His Gln Ser Ile
1 5
<210> 37
<211> 8
<212> PRT
<213> 智人
<400> 37
Pro Ala Arg Pro Ala Gly Pro Leu
1 5
<210> 38
<211> 8
<212> PRT
<213> 智人
<400> 38

Pro Leu Pro Ile Val Pro Ala Leu

1 5

<210> 39

<211> 9

<212> PRT

<213> 智人

<400> 39

Ser Asn Leu Ile Arg Thr Gly Ser His

1 5

<210> 40

<211> 8

<212> PRT

<213> 智人

<400> 40

Val Pro Ala Pro Ala Gln Ala Ile

1 5

<210> 41

<211> 9

<212> PRT

<213> 智人

<400> 41

Lys Gly His Gly Gly Pro Arg Ser Trp

1 5

<210> 42

<211> 11

<212> PRT

<213> 智人

<400> 42

Ile Thr Ser Ser Ala Val Thr Thr Ala Leu Lys

1 5 10

<210> 43

<211> 9

<212> PRT

<213> 智人

<400> 43

Leu Leu Leu Pro Glu Ser Pro Ser Ile

1 5

<210> 44

<211> 9

<212> PRT
<213> 智人
<400> 44
Val Ile Leu Ile Pro Leu Pro Pro Lys
1 5
<210> 45
<211> 9
<212> PRT
<213> 智人
<400> 45
Ala Val Leu Leu Pro Lys Pro Pro Lys
1 5
<210> 46
<211> 9
<212> PRT
<213> 智人
<400> 46
Thr Gln Val Ser Met Ala Glu Ser Ile
1 5
<210> 47
<211> 8
<212> PRT
<213> 智人
<400> 47
Leu Asn His Leu Arg Thr Ser Ile
1 5
<210> 48
<211> 9
<212> PRT
<213> 智人
<400> 48
Asn Thr Ser His Leu Pro Leu Ile Tyr
1 5
<210> 49
<211> 8
<212> PRT
<213> 智人
<400> 49
Ser Ile Gln Arg Asn Leu Ser Leu

1 5
<210> 50
<211> 9
<212> PRT
<213> 智人
<400> 50
Asn Val Ser Ser His Val His Thr Val
1 5
<210> 51
<211> 9
<212> PRT
<213> 智人
<400> 51
Ala Leu Ala Ser His Leu Ile Glu Ala
1 5
<210> 52
<211> 9
<212> PRT
<213> 智人
<400> 52
Ala Leu Asp Asp Ile Thr Ile Gln Leu
1 5
<210> 53
<211> 9
<212> PRT
<213> 智人
<400> 53
Ala Leu Gly Asn Thr Val Pro Ala Val
1 5
<210> 54
<211> 9
<212> PRT
<213> 智人
<400> 54
Ala Leu Leu Pro Ala Val Pro Ser Leu
1 5
<210> 55
<211> 9
<212> PRT

<213> 智人
<400> 55
Ala Pro Ala Pro Pro Pro Val Ala Val
1 5
<210> 56
<211> 8
<212> PRT
<213> 智人
<400> 56
Ala Pro Asp Lys Lys Ile Thr Leu
1 5
<210> 57
<211> 9
<212> PRT
<213> 智人
<400> 57
Ala Gln Met Asn Leu Leu Gln Lys Tyr
1 5
<210> 58
<211> 9
<212> PRT
<213> 智人
<400> 58
Asp Gln Val Ile Arg Leu Ala Gly Leu
1 5
<210> 59
<211> 9
<212> PRT
<213> 智人
<400> 59
Glu Thr Thr Ser Gln Val Arg Lys Tyr
1 5
<210> 60
<211> 9
<212> PRT
<213> 智人
<400> 60
Gly Gly Ser Leu Ile His Pro Gln Trp
1 5

<210> 61
<211> 9
<212> PRT
<213> 智人
<400> 61
Gly Leu Tyr Tyr Lys Leu His Asn Val
1 5
<210> 62
<211> 9
<212> PRT
<213> 智人
<400> 62
Gly Gln Lys Pro Val Ile Leu Thr Tyr
1 5
<210> 63
<211> 9
<212> PRT
<213> 智人
<400> 63
Gly Ser Leu Asp Phe Gln Arg Gly Trp
1 5
<210> 64
<211> 9
<212> PRT
<213> 智人
<400> 64
His His Leu Val Glu Thr Leu Lys Phe
1 5
<210> 65
<211> 9
<212> PRT
<213> 智人
<400> 65
His Leu Leu Ser Glu Thr Pro Gln Leu
1 5
<210> 66
<211> 9
<212> PRT
<213> 智人

<400> 66

His Gln Leu Tyr Arg Ala Ser Ala Leu

1 5

<210> 67

<211> 10

<212> PRT

<213> 智人

<400> 67

His Thr Asp Asp Ile Glu Asn Ala Lys Tyr

1 5 10

<210> 68

<211> 8

<212> PRT

<213> 智人

<400> 68

Ile Ala Ala Pro Ile Leu His Val

1 5

<210> 69

<211> 9

<212> PRT

<213> 智人

<400> 69

Lys Ala Phe Pro Phe His Ile Ile Phe

1 5

<210> 70

<211> 9

<212> PRT

<213> 智人

<400> 70

Lys Ala Thr Glu Tyr Val His Ser Leu

1 5

<210> 71

<211> 9

<212> PRT

<213> 智人

<400> 71

Lys Phe Ser Asn Val Thr Met Leu Phe

1 5

<210> 72

<211> 9
<212> PRT
<213> 智人
<400> 72
Lys Leu Leu Glu Lys Ala Phe Ser Ile
1 5
<210> 73
<211> 9
<212> PRT
<213> 智人
<400> 73
Lys Pro Met Pro Thr Lys Val Val Phe
1 5
<210> 74
<211> 9
<212> PRT
<213> 智人
<400> 74
Asn Val Asn Arg Pro Leu Thr Met Lys
1 5
<210> 75
<211> 11
<212> PRT
<213> 智人
<400> 75
Arg Glu Pro Tyr Glu Leu Thr Val Pro Ala Leu
1 5 10
<210> 76
<211> 10
<212> PRT
<213> 智人
<400> 76
Ser Glu Ala Glu Ala Ala Lys Asn Ala Leu
1 5 10
<210> 77
<211> 9
<212> PRT
<213> 智人
<400> 77

Ser Leu Trp Gly Gln Pro Ala Glu Ala

1 5

<210> 78

<211> 11

<212> PRT

<213> 智人

<400> 78

Ser Pro Ala Asp His Arg Gly Tyr Ala Ser Leu

1 5 10

<210> 79

<211> 9

<212> PRT

<213> 智人

<400> 79

Ser Pro Gln Ser Ala Ala Ala Glu Leu

1 5

<210> 80

<211> 8

<212> PRT

<213> 智人

<400> 80

Ser Pro Val Val His Gln Ser Leu

1 5

<210> 81

<211> 8

<212> PRT

<213> 智人

<400> 81

Ser Pro Tyr Arg Thr Pro Val Leu

1 5

<210> 82

<211> 9

<212> PRT

<213> 智人

<400> 82

Ser Val Phe Ala Gly Val Val Gly Val

1 5

<210> 83

<211> 9

<212> PRT

<213> 智人

<400> 83

Ser Tyr Ser Pro Ala His Ala Arg Leu

1 5

<210> 84

<211> 9

<212> PRT

<213> 智人

<400> 84

Thr His Gly Ser Glu Gln Leu His Leu

1 5

<210> 85

<211> 9

<212> PRT

<213> 智人

<400> 85

Thr Gln Ala Pro Pro Asn Val Val Leu

1 5

<210> 86

<211> 10

<212> PRT

<213> 智人

<400> 86

Val Leu Val Pro Tyr Glu Pro Pro Gln Val

1 5 10

<210> 87

<211> 9

<212> PRT

<213> 智人

<400> 87

Val Ser Phe Pro Asp Val Arg Lys Val

1 5

<210> 88

<211> 11

<212> PRT

<213> 智人

<400> 88

Val Val Phe Asp Lys Ser Asp Leu Ala Lys Tyr

1	5	10
---	---	----

<210> 89
<211> 9
<212> PRT
<213> 智人
<400> 89
Tyr Ser His His Ser Gly Leu Glu Tyr
1 5
<210> 90
<211> 9
<212> PRT
<213> 智人
<400> 90
Tyr Tyr Leu Asp Trp Ile His His Tyr
1 5
<210> 91
<211> 9
<212> PRT
<213> 智人
<400> 91
Ser Val Tyr Lys Tyr Leu Lys Ala Lys
1 5
<210> 92
<211> 9
<212> PRT
<213> 智人
<400> 92
Ile Tyr Gln Phe Ile Met Asp Arg Phe
1 5
<210> 93
<211> 9
<212> PRT
<213> 智人
<400> 93
Gly Thr Leu Gln Gly Ile Arg Ala Trp
1 5
<210> 94
<211> 10
<212> PRT

<213> 智人
<400> 94
Ala Gln Lys Val Ser Val Gly Gln Ala Ala
1 5 10
<210> 95
<211> 9
<212> PRT
<213> 智人
<400> 95
Leu Tyr Pro Ser Lys Leu Thr His Phe
1 5
<210> 96
<211> 9
<212> PRT
<213> 智人
<400> 96
Ala Thr Gln Asn Thr Ile Ile Gly Lys
1 5
<210> 97
<211> 9
<212> PRT
<213> 智人
<400> 97
Ala Gln Asp Ile Ile Leu Gln Ala Val
1 5
<210> 98
<211> 9
<212> PRT
<213> 智人
<400> 98
Pro Pro Arg Pro Leu Gly Ala Gln Val
1 5
<210> 99
<211> 9
<212> PRT
<213> 智人
<400> 99
Phe Asn Val Ala Leu Asn Ala Arg Tyr
1 5

<210> 100
<211> 9
<212> PRT
<213> 智人
<400> 100
Gly Pro Gly Ser Arg Glu Ser Thr Leu
1 5
<210> 101
<211> 8
<212> PRT
<213> 智人
<400> 101
Ile Pro His Gln Arg Ser Ser Leu
1 5
<210> 102
<211> 9
<212> PRT
<213> 智人
<400> 102
Leu Thr Asp Arg Ile Tyr Leu Thr Leu
1 5
<210> 103
<211> 9
<212> PRT
<213> 智人
<400> 103
Asn Leu Lys Glu Lys Lys Ala Leu Phe
1 5
<210> 104
<211> 10
<212> PRT
<213> 智人
<400> 104
Val Leu Phe Gly Gly Lys Val Ser Gly Ala
1 5 10
<210> 105
<211> 9
<212> PRT
<213> 智人

<400> 105
Val Val Phe Pro Phe Pro Val Asn Lys
1 5
<210> 106
<211> 9
<212> PRT
<213> 智人
<400> 106
Ser Leu Leu Ile Ile Pro Lys Lys Lys
1 5
<210> 107
<211> 9
<212> PRT
<213> 智人
<400> 107
Ala Pro Gly Ala Ala Gly Gln Arg Leu
1 5
<210> 108
<211> 9
<212> PRT
<213> 智人
<400> 108
Lys Leu Gln Asp Lys Glu Ile Gly Leu
1 5
<210> 109
<211> 9
<212> PRT
<213> 智人
<400> 109
Ser Leu Arg Glu Pro Gln Pro Ala Leu
1 5
<210> 110
<211> 9
<212> PRT
<213> 智人
<400> 110
Thr Pro Gly Arg Ser Thr Gln Ala Ile
1 5
<210> 111

<211> 8
<212> PRT
<213> 智人
<400> 111
Ala Pro Arg Gly Thr Ala Ala Leu
1 5
<210> 112
<211> 8
<212> PRT
<213> 智人
<400> 112
Ile Ala Ser Pro Ile Ala Leu Leu
1 5
<210> 113
<211> 9
<212> PRT
<213> 智人
<400> 113
Ile Leu Phe Gln Asn Ser Ala Leu Lys
1 5
<210> 114
<211> 8
<212> PRT
<213> 智人
<400> 114
Ile Leu Lys Lys Asn Ile Ser Ile
1 5
<210> 115
<211> 8
<212> PRT
<213> 智人
<400> 115
Ile Pro Leu Ala Val Arg Thr Ile
1 5
<210> 116
<211> 8
<212> PRT
<213> 智人
<400> 116

Leu Pro Arg Asn Lys Pro Leu Leu

1 5

<210> 117

<211> 9

<212> PRT

<213> 智人

<400> 117

Pro Ala Pro Pro His Pro Ala Ala Leu

1 5

<210> 118

<211> 8

<212> PRT

<213> 智人

<400> 118

Ser Pro Val Val Arg Val Gly Leu

1 5

<210> 119

<211> 10

<212> PRT

<213> 智人

<400> 119

Thr Leu Asn Gln Gly Ile Asn Val Tyr Ile

1 5 10

<210> 120

<211> 9

<212> PRT

<213> 智人

<400> 120

Arg Pro Arg Gly Pro Arg Thr Ala Pro

1 5

<210> 121

<211> 11

<212> PRT

<213> 智人

<400> 121

Ser Val Gln Leu Leu Glu Gln Ala Ile His Lys

1 5 10

<210> 122

<211> 9

<212> PRT
<213> 智人
<400> 122
Arg Thr Pro Lys Asn Tyr Gln His Trp
1 5
<210> 123
<211> 9
<212> PRT
<213> 智人
<400> 123
Ala Leu Pro Val Ala Leu Pro Ser Leu
1 5
<210> 124
<211> 9
<212> PRT
<213> 智人
<400> 124
Ser Leu Gln Ile Leu Val Ser Ser Leu
1 5
<210> 125
<211> 9
<212> PRT
<213> 智人
<400> 125
Ile Ser Asn Lys Val Pro Lys Leu Phe
1 5
<210> 126
<211> 10
<212> PRT
<213> 智人
<400> 126
Thr Val Ile Arg Ile Ala Ile Val Asn Lys
1 5 10
<210> 127
<211> 9
<212> PRT
<213> 智人
<400> 127
Lys Glu Ile Phe Leu Glu Leu Arg Leu

1 5
<210> 128
<211> 9
<212> PRT
<213> 智人
<400> 128
Thr Leu Arg Ser Pro Gly Ser Ser Leu
1 5
<210> 129
<211> 9
<212> PRT
<213> 智人
<400> 129
Thr Val Arg Gly Asp Val Ser Ser Leu
1 5
<210> 130
<211> 9
<212> PRT
<213> 智人
<400> 130
Ala Leu Asp Pro Leu Leu Leu Arg Ile
1 5
<210> 131
<211> 9
<212> PRT
<213> 智人
<400> 131
Ile Ser Leu Ile Val Thr Gly Leu Lys
1 5
<210> 132
<211> 9
<212> PRT
<213> 智人
<400> 132
Lys Ile Leu Asp Val Asn Leu Arg Ile
1 5
<210> 133
<211> 9
<212> PRT

<213> 智人
<400> 133
Glu Arg Val Tyr Ile Arg Ala Ser Leu
1 5
<210> 134
<211> 8
<212> PRT
<213> 智人
<400> 134
Ile Leu Asp Leu Glu Ser Arg Tyr
1 5
<210> 135
<211> 9
<212> PRT
<213> 智人
<400> 135
Lys Thr Phe Val Gln Gln Lys Thr Leu
1 5
<210> 136
<211> 9
<212> PRT
<213> 智人
<400> 136
Leu Tyr Ile Lys Ser Leu Pro Ala Leu
1 5
<210> 137
<211> 9
<212> PRT
<213> 智人
<400> 137
Val Leu Lys Glu Lys Asn Ala Ser Leu
1 5
<210> 138
<211> 9
<212> PRT
<213> 智人
<400> 138
Leu Gly Ile Ser Leu Thr Leu Lys Tyr
1 5

<210> 139

<211> 8

<212> PRT

<213> 智人

<400> 139

Asp Leu Leu Pro Lys Lys Leu Leu

1 5

<210> 140

<211> 9

<212> PRT

<213> 智人

<400> 140

His Ser Leu Ile Ser Ile Val Tyr Leu

1 5

<210> 141

<211> 9

<212> PRT

<213> 智人

<400> 141

Ile Ala Gly Ala Leu Arg Ser Val Leu

1 5

<210> 142

<211> 9

<212> PRT

<213> 智人

<400> 142

Ile Gly Asn Pro Ile Leu Arg Val Leu

1 5

<210> 143

<211> 9

<212> PRT

<213> 智人

<400> 143

Ile Tyr Ala Pro His Ile Arg Leu Ser

1 5

<210> 144

<211> 8

<212> PRT

<213> 智人

<400> 144
Leu Arg Ser Gln Ile Leu Ser Tyr
1 5
<210> 145
<211> 9
<212> PRT
<213> 智人
<400> 145
Arg Tyr Leu Ala Asn Lys Ile His Ile
1 5
<210> 146
<211> 9
<212> PRT
<213> 智人
<400> 146
Ser Leu Leu Ser Gly Leu Leu Arg Ala
1 5
<210> 147
<211> 8
<212> PRT
<213> 智人
<400> 147
Ser Arg Ile His Leu Val Val Leu
1 5
<210> 148
<211> 10
<212> PRT
<213> 智人
<400> 148
Ser Ser Ser Pro Val Arg Gly Pro Ser Val
1 5 10
<210> 149
<211> 9
<212> PRT
<213> 智人
<400> 149
Ser Thr Phe Ser Leu Tyr Leu Lys Lys
1 5
<210> 150

<211> 9
<212> PRT
<213> 智人
<400> 150
Ser Leu Asp Leu Leu Pro Leu Ser Ile
1 5
<210> 151
<211> 9
<212> PRT
<213> 智人
<400> 151
Val Thr Asp Leu Leu Ala Leu Thr Val
1 5
<210> 152
<211> 11
<212> PRT
<213> 智人
<400> 152
Arg Thr Gln Ile Thr Lys Val Ser Leu Lys Lys
1 5 10
<210> 153
<211> 8
<212> PRT
<213> 智人
<400> 153
Ile Leu Arg Ser Pro Leu Lys Trp
1 5
<210> 154
<211> 9
<212> PRT
<213> 智人
<400> 154
Leu Ser Thr Gly His Leu Ser Thr Val
1 5
<210> 155
<211> 9
<212> PRT
<213> 智人
<400> 155

Thr Val Glu Glu Tyr Leu Val Asn Ile

1 5

<210> 156

<211> 10

<212> PRT

<213> 智人

<400> 156

Gln Ile Lys Thr Lys Leu Leu Gly Ser Leu

1 5 10

<210> 157

<211> 10

<212> PRT

<213> 智人

<400> 157

Leu Pro Ser Phe Ser His Phe Leu Leu Leu

1 5 10

<210> 158

<211> 9

<212> PRT

<213> 智人

<400> 158

Cys Leu Arg Ile Gly Pro Val Thr Leu

1 5

<210> 159

<211> 9

<212> PRT

<213> 智人

<400> 159

His Val Ser Asp Gly Ser Thr Ala Leu

1 5

<210> 160

<211> 9

<212> PRT

<213> 智人

<400> 160

Ile Ala Tyr Ser Val Arg Ala Leu Arg

1 5

<210> 161

<211> 8

<212> PRT
<213> 智人
<400> 161
Pro Arg Gly Phe Leu Ser Ala Leu
1 5
<210> 162
<211> 9
<212> PRT
<213> 智人
<400> 162
Ile Ser Ser Trp Leu Ile Ser Ser Leu
1 5
<210> 163
<211> 9
<212> PRT
<213> 智人
<400> 163
Ile Pro Leu Asn Pro Phe Ser Ser Leu
1 5
<210> 164
<211> 8
<212> PRT
<213> 智人
<400> 164
Leu Ser Asp Arg Gln Leu Ser Leu
1 5
<210> 165
<211> 9
<212> PRT
<213> 智人
<400> 165
Leu Ser His Pro Ala Pro Ser Ser Leu
1 5
<210> 166
<211> 9
<212> PRT
<213> 智人
<400> 166
Leu Arg Lys Ala Val Asp Pro Ile Leu

1 5
<210> 167
<211> 9
<212> PRT
<213> 智人
<400> 167
Ile Leu Leu Glu Glu Gln Ser Leu Ile
1 5
<210> 168
<211> 9
<212> PRT
<213> 智人
<400> 168
Leu Thr Ser Ile Ser Ile Arg Pro Val
1 5
<210> 169
<211> 9
<212> PRT
<213> 智人
<400> 169
Thr Ile Ser Glu Cys Pro Leu Leu Ile
1 5
<210> 170
<211> 9
<212> PRT
<213> 智人
<400> 170
Thr Leu Lys Leu Lys Lys Ile Phe Phe
1 5
<210> 171
<211> 9
<212> PRT
<213> 智人
<400> 171
Ile Leu Leu Ser Asn Phe Ser Ser Leu
1 5
<210> 172
<211> 9
<212> PRT

<213> 智人
<400> 172
Leu Gly Gly Ala Trp Lys Ala Val Phe
1 5
<210> 173
<211> 9
<212> PRT
<213> 智人
<400> 173
Leu Ser Ala Ser His Leu Ser Ser Leu
1 5
<210> 174
<211> 9
<212> PRT
<213> 智人
<400> 174
Ala Gly Asp Ile Ile Ala Arg Leu Ile
1 5
<210> 175
<211> 9
<212> PRT
<213> 智人
<400> 175
Asp Arg Gly Ile Leu Arg Asn Leu Leu
1 5
<210> 176
<211> 9
<212> PRT
<213> 智人
<400> 176
Gly Leu Arg Leu Ile His Val Ser Leu
1 5
<210> 177
<211> 9
<212> PRT
<213> 智人
<400> 177
Gly Leu Arg Leu Leu His Val Ser Leu
1 5

<210> 178

<211> 9

<212> PRT

<213> 智人

<400> 178

Leu His Asn Glu Lys Gly Leu Ser Leu

1 5

<210> 179

<211> 11

<212> PRT

<213> 智人

<400> 179

Leu Pro Ser Phe Ser Arg Pro Ser Gly Ile Ile

1 5 10

<210> 180

<211> 9

<212> PRT

<213> 智人

<400> 180

Leu Ser Ser Arg Leu Pro Leu Gly Lys

1 5

<210> 181

<211> 8

<212> PRT

<213> 智人

<400> 181

Met Ile Gly Ile Lys Arg Leu Leu

1 5

<210> 182

<211> 8

<212> PRT

<213> 智人

<400> 182

Asn Leu Lys Lys Arg Glu Ile Leu

1 5

<210> 183

<211> 9

<212> PRT

<213> 智人

<400> 183
Arg Met Val Ala Tyr Leu Gln Gln Leu
1 5
<210> 184
<211> 9
<212> PRT
<213> 智人
<400> 184
Ser Pro Ala Arg Ala Leu Pro Ser Leu
1 5
<210> 185
<211> 8
<212> PRT
<213> 智人
<400> 185
Thr Val Pro Gly Ile Gln Arg Tyr
1 5
<210> 186
<211> 9
<212> PRT
<213> 智人
<400> 186
Val Ser Arg Asn Tyr Val Leu Leu Ile
1 5
<210> 187
<211> 9
<212> PRT
<213> 智人
<400> 187
Leu Thr Val Pro Leu Ser Val Phe Trp
1 5
<210> 188
<211> 9
<212> PRT
<213> 智人
<400> 188
Lys Leu Asn Gln Ala Phe Leu Val Leu
1 5
<210> 189

<211> 10

<212> PRT

<213> 智人

<400> 189

Arg Leu Val Ser Ser Thr Leu Leu Gln Lys

1 5 10

<210> 190

<211> 8

<212> PRT

<213> 智人

<400> 190

Leu Pro Ser His Ser Leu Leu Ile

1 5

<210> 191

<211> 13

<212> PRT

<213> 智人

<400> 191

Cys Ser Ala Arg Gly Asp Arg Glu Tyr Glu Gln Tyr Phe

1 5 10

<210> 192

<211> 15

<212> PRT

<213> 智人

<400> 192

Cys Ala Ser Thr Val Val Ala Gly Asn Ser Ser Pro Leu His Phe

1 5 10 15

<210> 193

<211> 16

<212> PRT

<213> 智人

<400> 193

Cys Ala Ser Ser Gln Asp Gly Ile Trp Gly Ala Tyr Glu Gln Tyr Phe

1 5 10 15

<210> 194

<211> 15

<212> PRT

<213> 智人

<400> 194

Cys Ala Ser Ser Val Asp Ala Gly Gly Asn Tyr Glu Gln Tyr Phe
1 5 10 15

<210> 195

<211> 16

<212> PRT

<213> 智人

<400> 195

Cys Ala Ser Ser Leu Gly Gly Gln Gly Leu Ser Tyr Gly Tyr Thr Phe
1 5 10 15

<210> 196

<211> 12

<212> PRT

<213> 智人

<400> 196

Cys Ala Ser Ser Tyr Arg Pro Asn Glu Gln Tyr Phe
1 5 10

<210> 197

<211> 14

<212> PRT

<213> 智人

<400> 197

Cys Ala Ser Ser Gly Thr Asp Leu Asn Gln Pro Gln His Phe
1 5 10

<210> 198

<211> 12

<212> PRT

<213> 智人

<400> 198

Cys Ala Ser Ser Ser Asn Phe Glu Pro Leu His Phe
1 5 10

<210> 199

<211> 15

<212> PRT

<213> 智人

<400> 199

Cys Ala Ser Ser Trp Gly Gly Ser Asn Thr Gly Glu Leu Phe Phe
1 5 10 15

<210> 200

<211> 15

<212> PRT

<213> 智人

<400> 200

Cys	Ala	Ser	Ser	Glu	Tyr	Arg	Ala	Leu	Asn	Thr	Glu	Ala	Phe	Phe
1				5					10				15	

<210> 201

<211> 15

<212> PRT

<213> 智人

<400> 201

Cys	Ala	Ser	Ser	Glu	Ser	Gly	Thr	Gly	Gly	Gln	Pro	Gln	His	Phe
1				5					10				15	

<210> 202

<211> 14

<212> PRT

<213> 智人

<400> 202

Cys	Ala	Ser	Ser	Arg	Thr	Gly	Glu	Asn	Thr	Glu	Ala	Phe	Phe
1				5					10				

<210> 203

<211> 15

<212> PRT

<213> 智人

<400> 203

Cys	Ala	Ser	Ser	Ser	Thr	Asp	Arg	Gln	His	Tyr	Gly	Tyr	Thr	Phe
1				5					10				15	

<210> 204

<211> 17

<212> PRT

<213> 智人

<400> 204

Cys	Ala	Ser	Ser	Asp	Arg	Thr	Gly	Gly	Ser	Ser	Asn	Glu	Lys	Leu	Phe
1				5					10					15	

Phe

<210> 205

<211> 15

<212> PRT

<213> 智人

<400> 205

Cys Ala Thr Ser Arg Ser Gly Asp Ser Asn Gln Pro Gln His Phe
1 5 10 15

<210> 206

<211> 13

<212> PRT

<213> 智人

<400> 206

Cys Ala Ser Ser Tyr Val Leu Asn Thr Glu Ala Phe Phe
1 5 10

<210> 207

<211> 15

<212> PRT

<213> 智人

<400> 207

Cys Ala Ser Arg Glu Ser Gly Gln Met Asn Glu Lys Leu Phe Phe
1 5 10 15

<210> 208

<211> 15

<212> PRT

<213> 智人

<400> 208

Cys Ala Ser Ser Asp Gly Gly Glu Gly Thr Tyr Gly Tyr Thr Phe
1 5 10 15

<210> 209

<211> 17

<212> PRT

<213> 智人

<400> 209

Cys Ala Ser Thr Ser Trp Thr Gly Phe Gly Pro Asn Tyr Gly Tyr Thr
1 5 10 15

Phe

<210> 210

<211> 14

<212> PRT

<213> 智人

<400> 210

Cys Ala Ser Ser Ala Asp Ser Ser Leu Gly Gly Tyr Thr Phe
1 5 10

<210> 211

<211> 15

<212> PRT

<213> 智人

<400> 211

Cys Ser Ala Arg Asp Leu Ala Gly Gly Thr Tyr Glu Gln Tyr Phe
1 5 10 15

<210> 212

<211> 15

<212> PRT

<213> 智人

<400> 212

Cys Ala Ser Ser Pro Pro Thr Gly Glu Tyr Glu Lys Leu Phe Phe
1 5 10 15

<210> 213

<211> 9

<212> PRT

<213> 智人

<400> 213

Cys Ala Arg Ser Phe Gly Gly Phe Phe
1 5

<210> 214

<211> 15

<212> PRT

<213> 智人

<400> 214

Cys Ala Ser Ser Leu Pro Ser Gly Ile Leu Tyr Glu Gln Tyr Phe
1 5 10 15

<210> 215

<211> 14

<212> PRT

<213> 智人

<400> 215

Cys Ala Ser Ala Pro Gly Gly Met Pro Tyr Gly Tyr Thr Phe
1 5 10

<210> 216

<211> 15

<212> PRT

<213> 智人

<400> 216

Cys Ala Ser Ser Leu Gln Asp Thr Asp Tyr Asn Glu Gln Phe Phe
1 5 10 15

<210> 217

<211> 18

<212> PRT

<213> 智人

<400> 217

Cys Ala Ser Ser Asp Pro Gly Thr Ser Gly Val Phe Thr Gly Glu Leu
1 5 10 15

Phe Phe

<210> 218

<211> 12

<212> PRT

<213> 智人

<400> 218

Cys Ser Ala Arg Leu Gly Thr Gly Glu Leu Phe Phe
1 5 10

<210> 219

<211> 12

<212> PRT

<213> 智人

<400> 219

Cys Ser Ala Arg Leu Gly Thr Gly Glu Leu Phe Phe
1 5 10

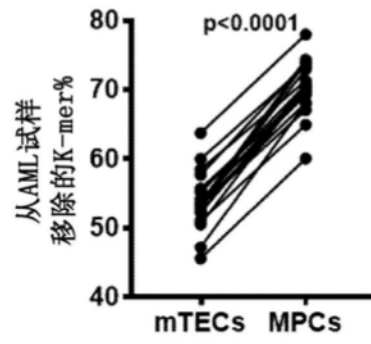


图1A

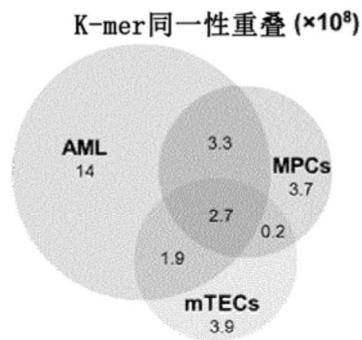


图1B

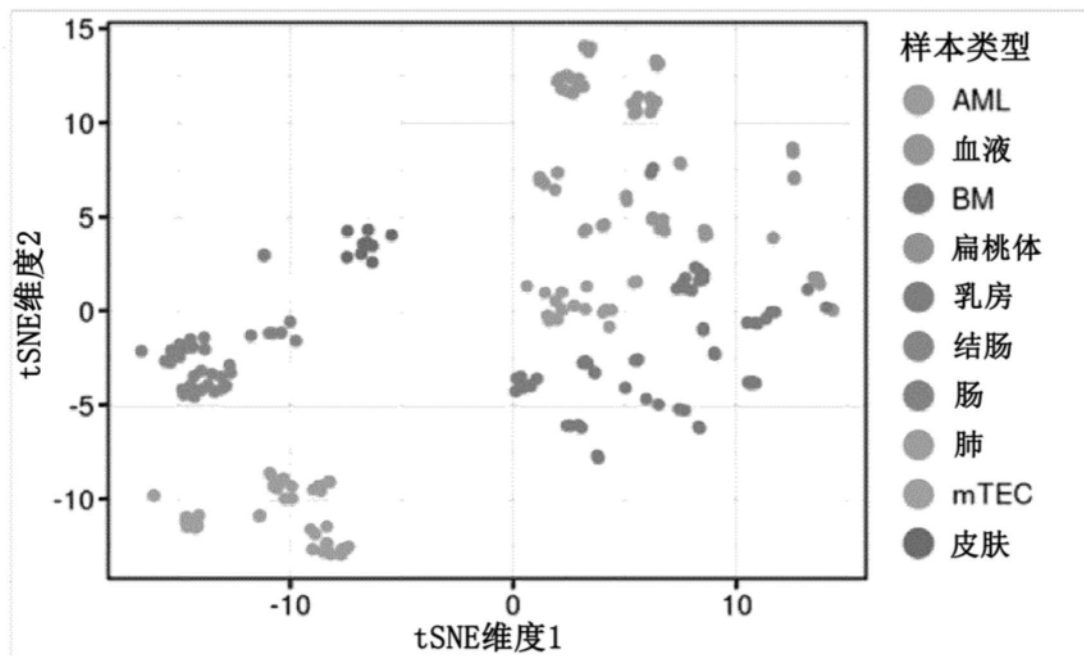


图1C

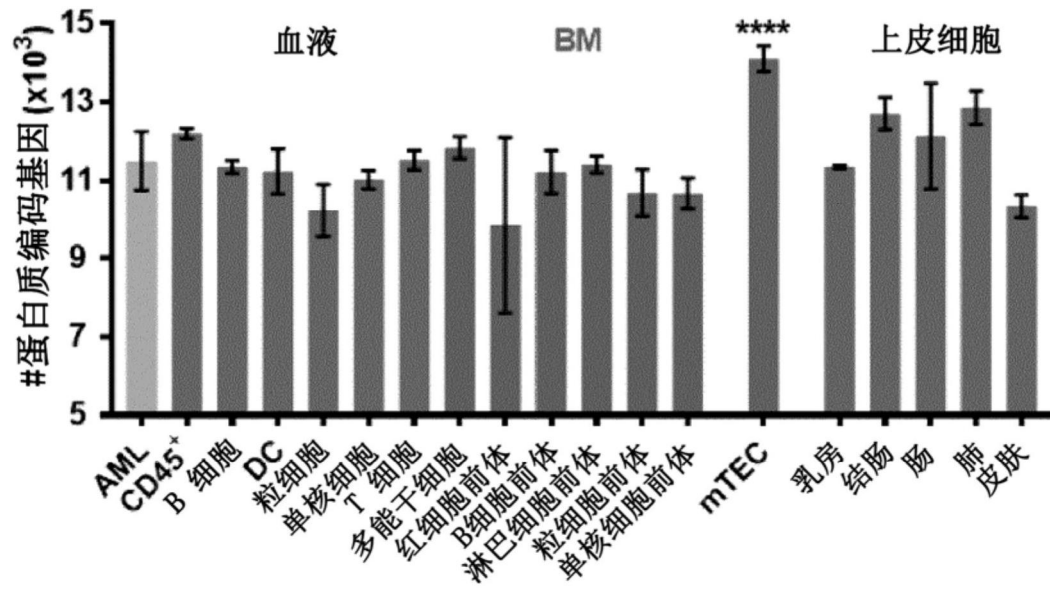


图1D

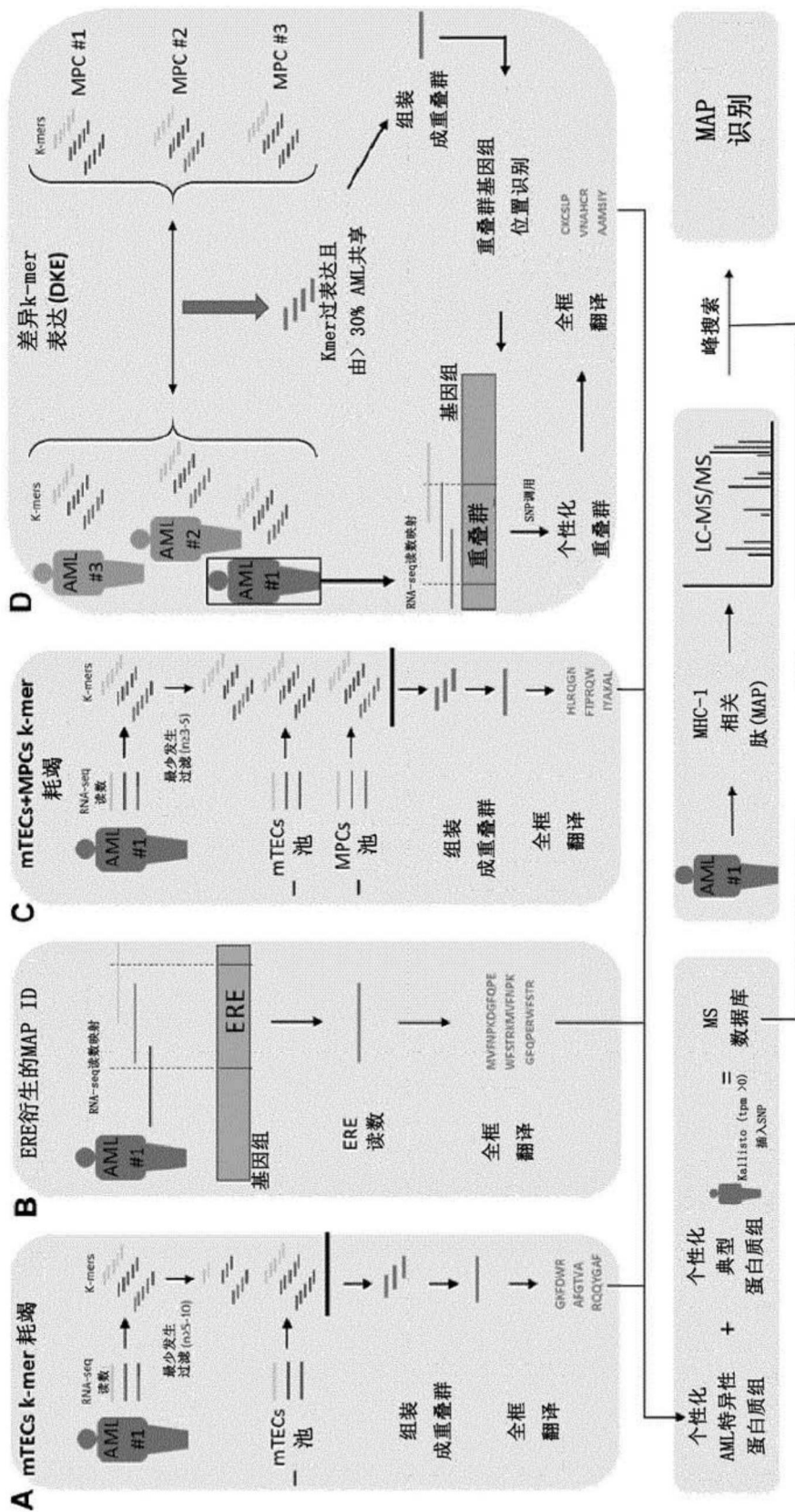


图2

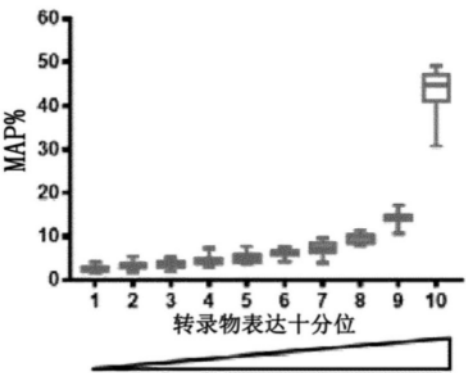


图3A

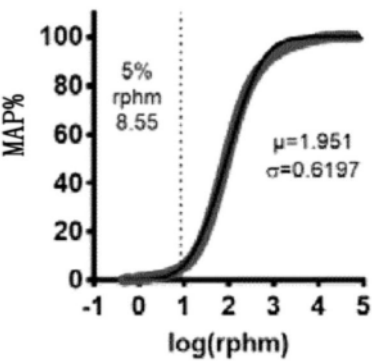


图3B

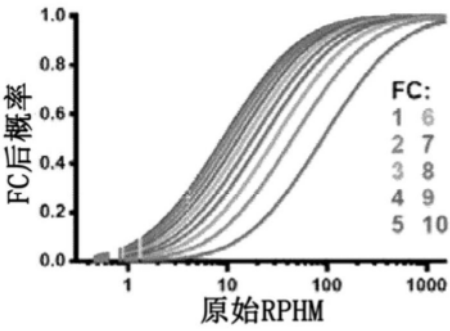


图3C

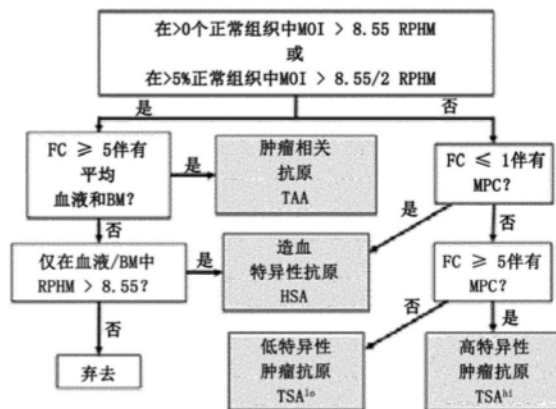


图3D

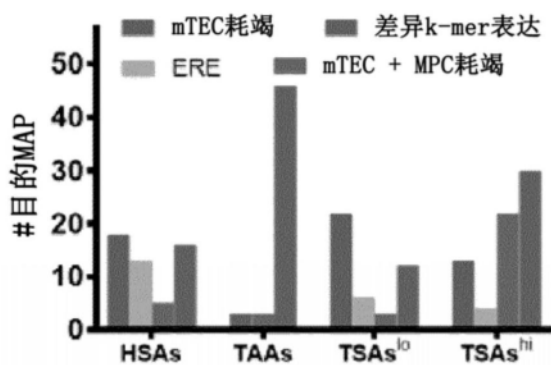


图3E

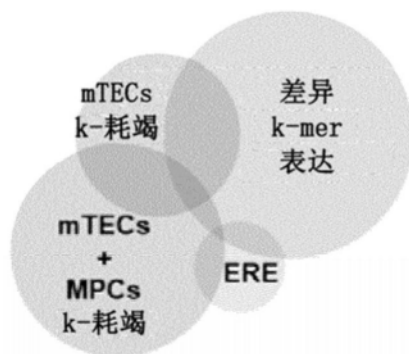
识别的TSA^{hi}

图3F

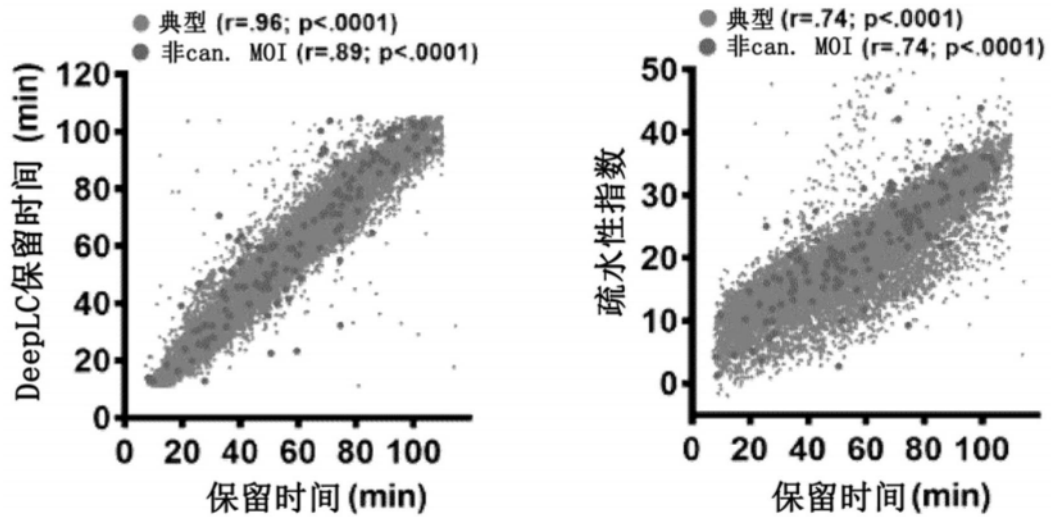


图3G

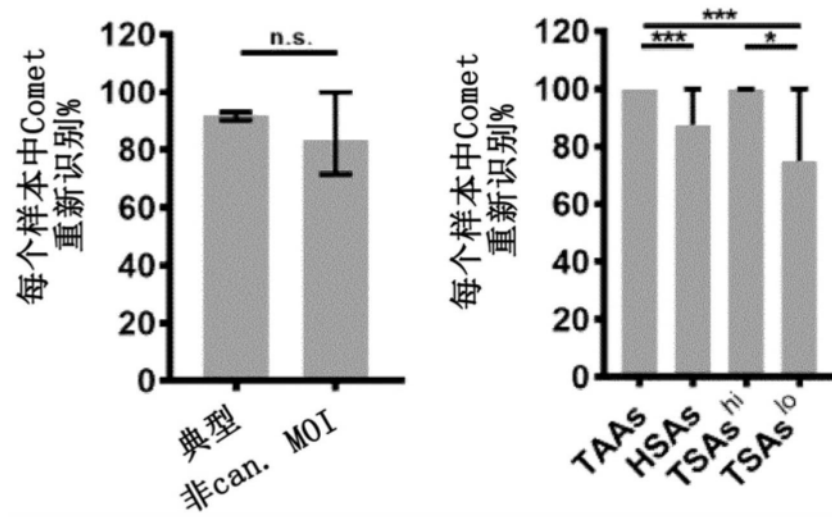


图3H

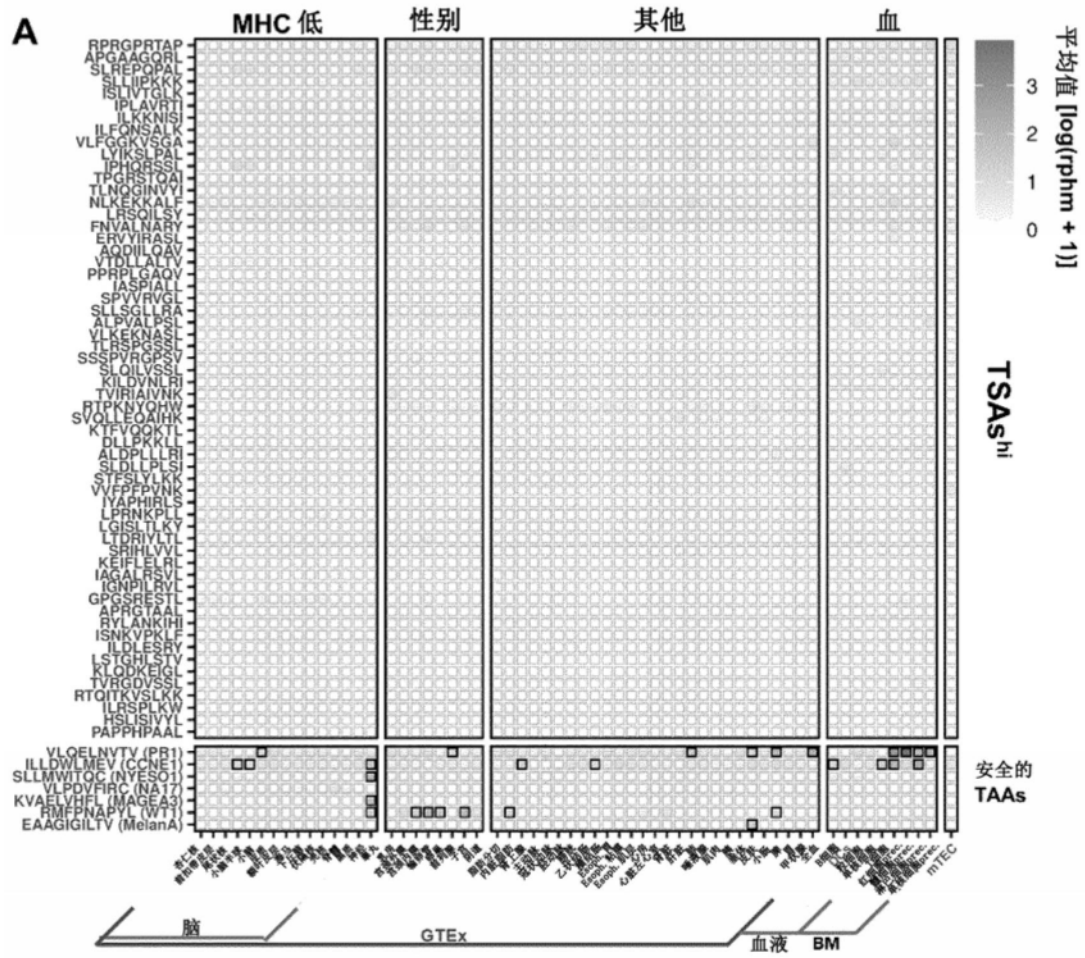


图4A

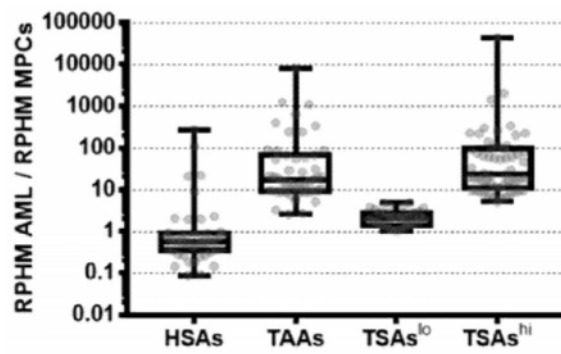


图4B

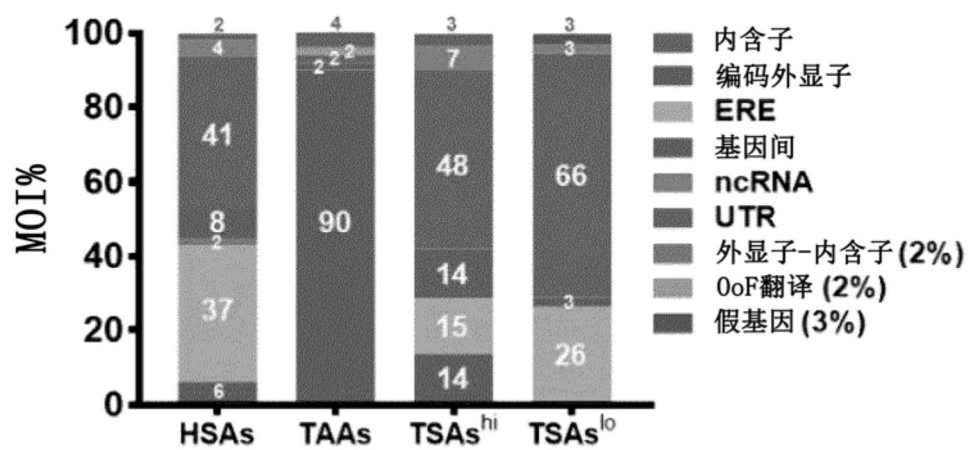


图4C

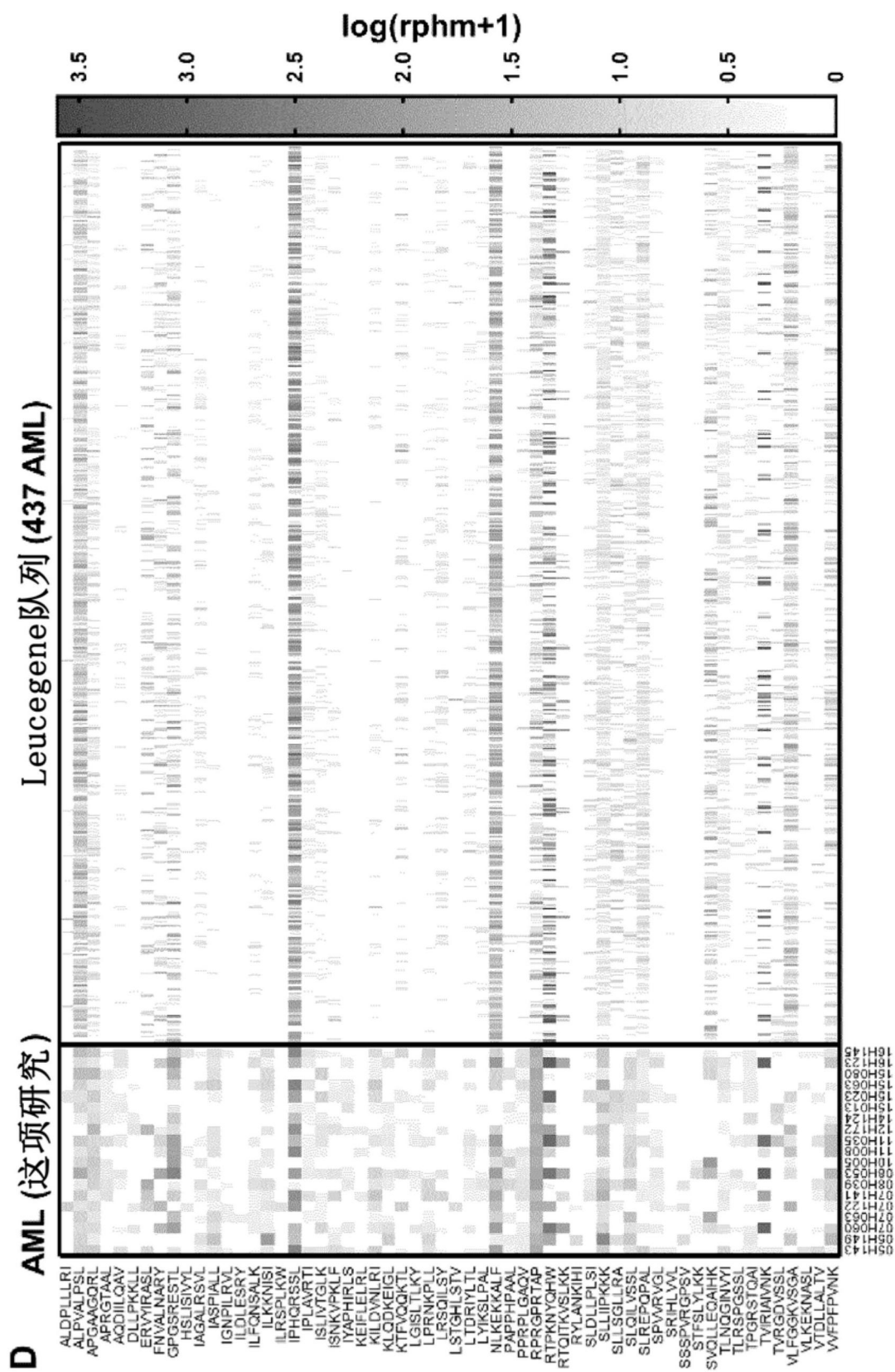


图4D

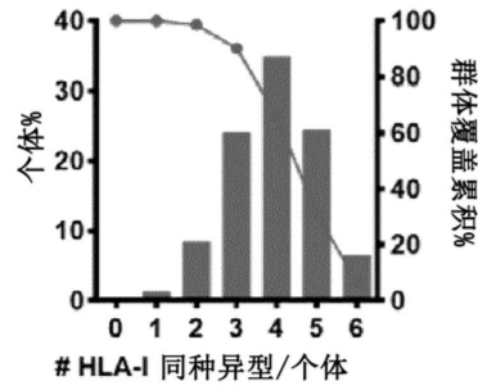


图4E

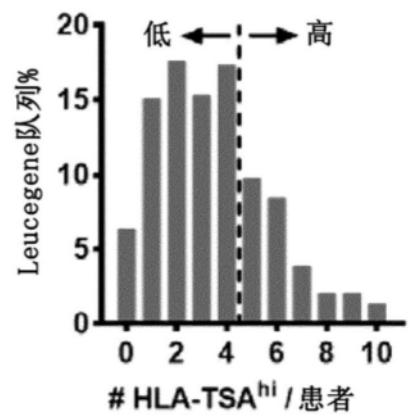


图4F

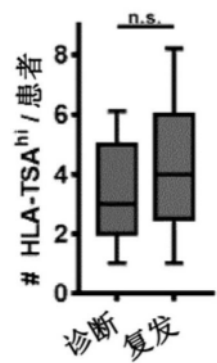


图4G

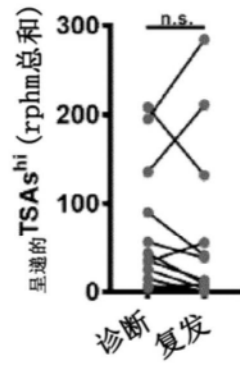


图4H

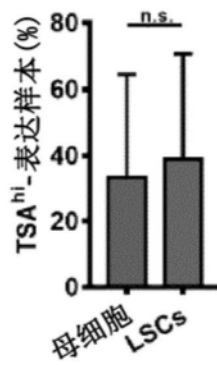


图4I

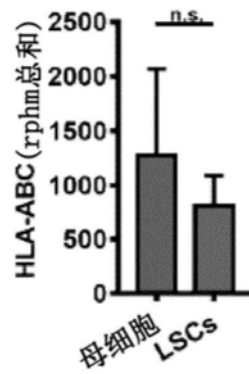


图4J

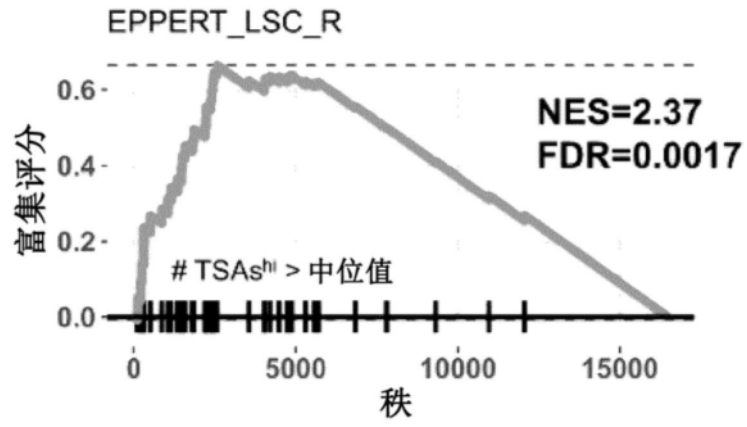


图4K

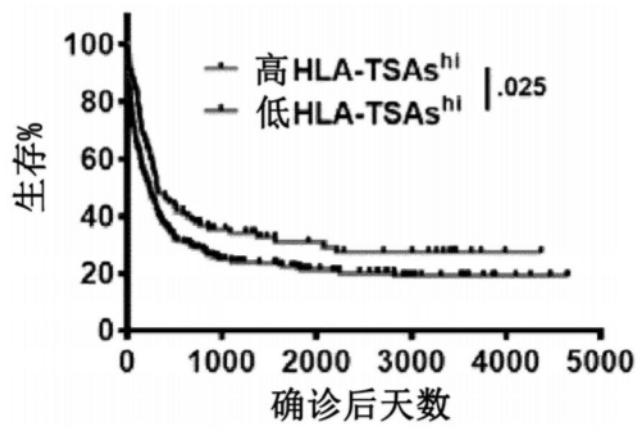


图5A

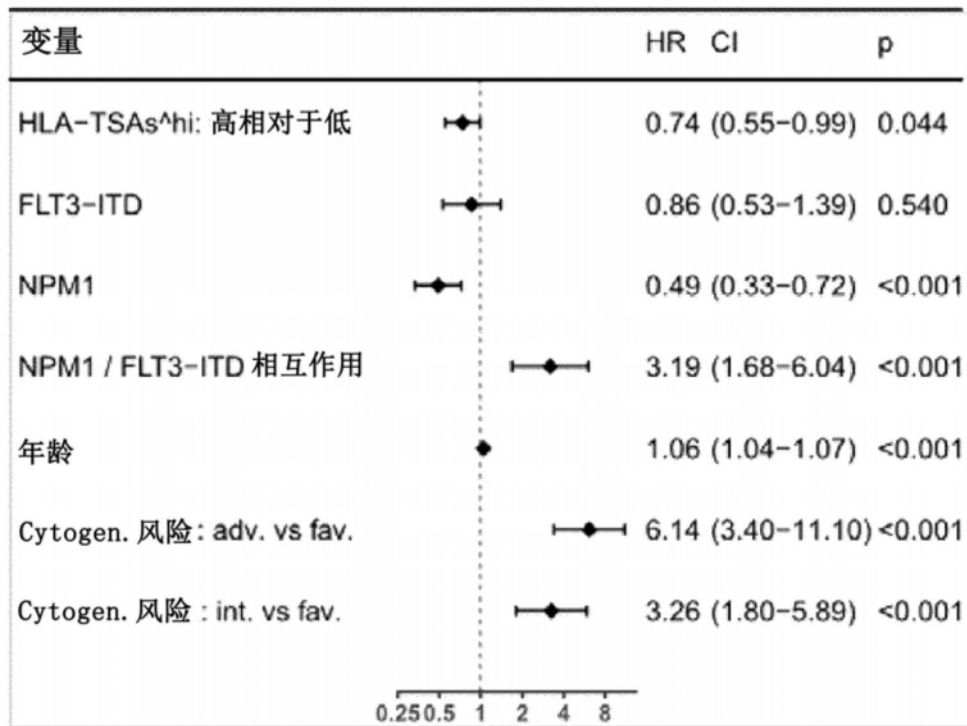


图5B

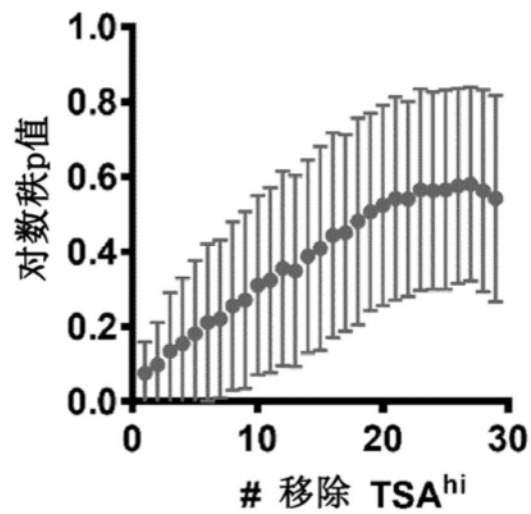


图5C

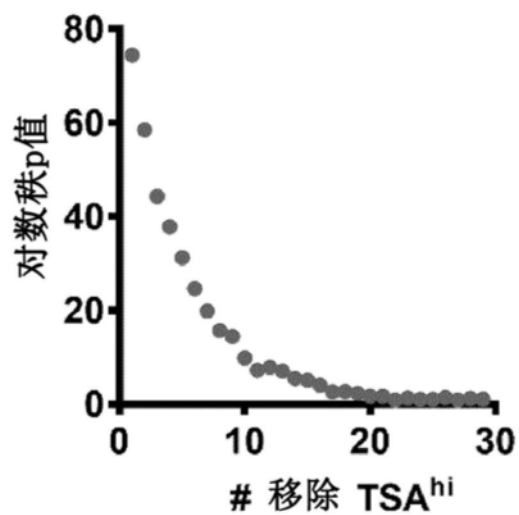


图5D

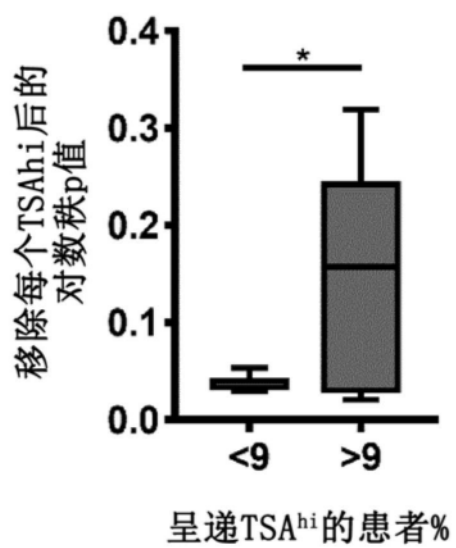


图5E

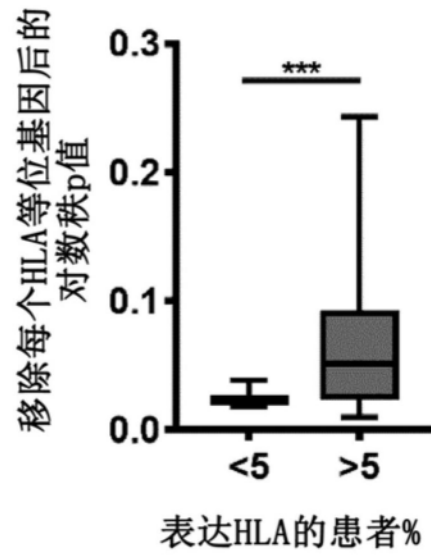


图5F

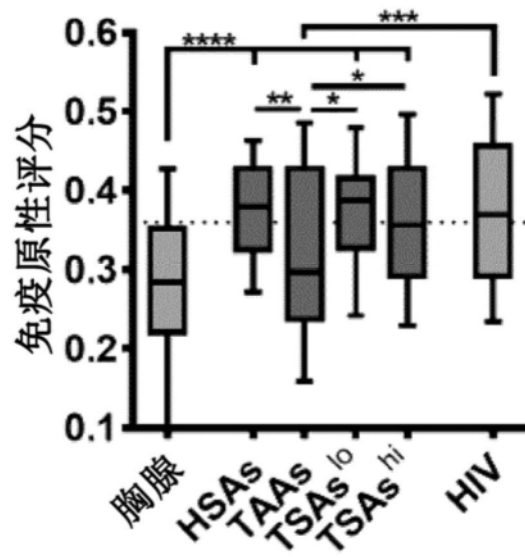


图6A

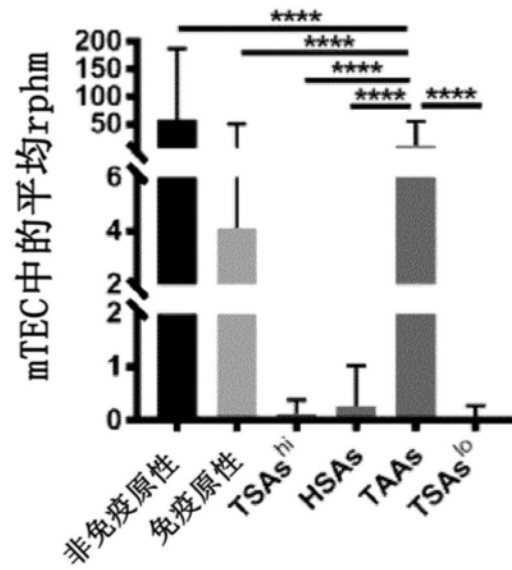


图6B

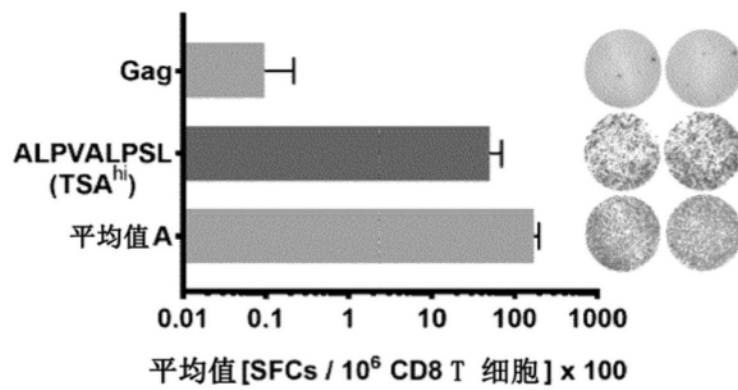


图6C

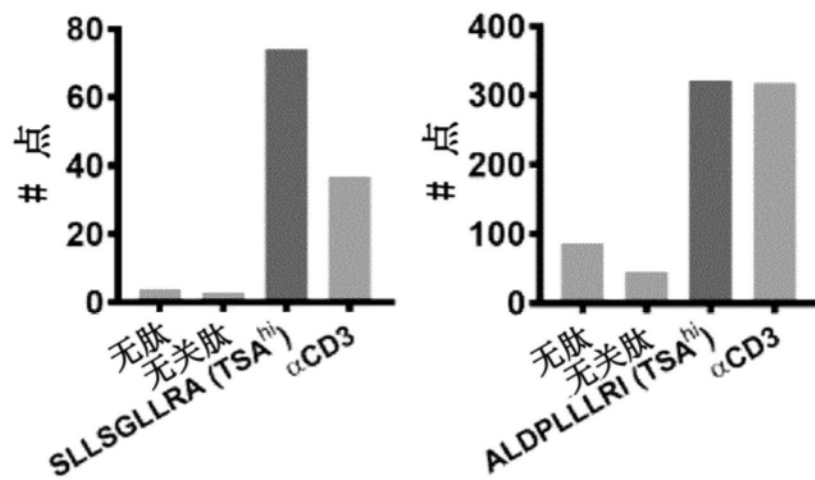


图6D

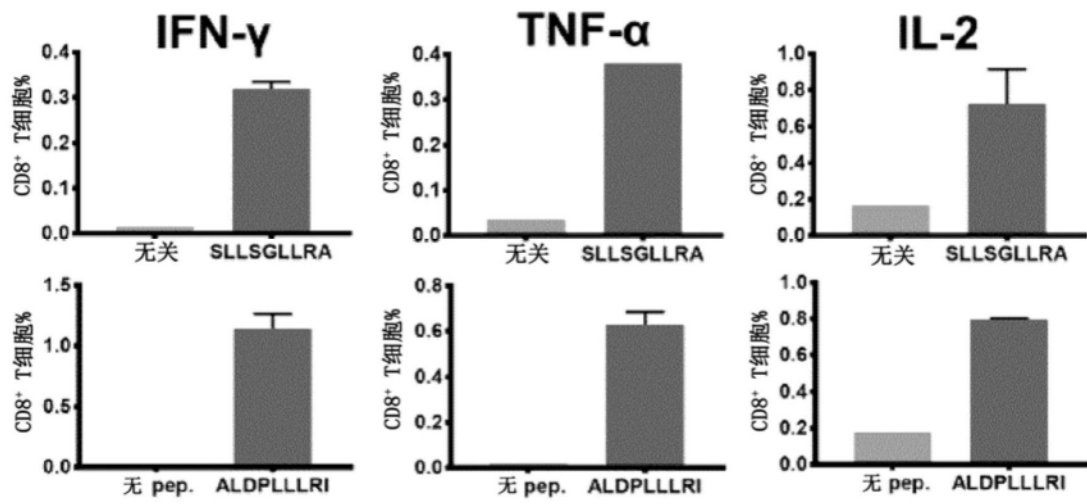


图6E

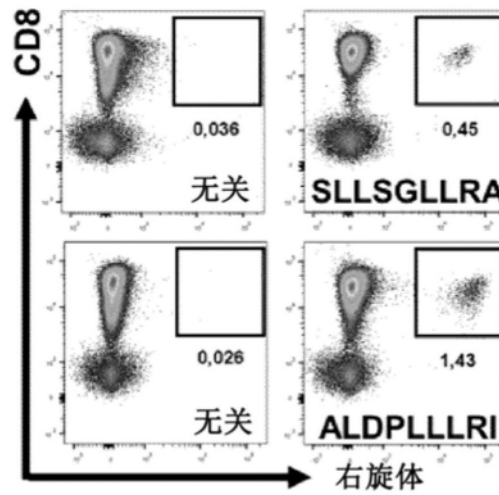


图6F

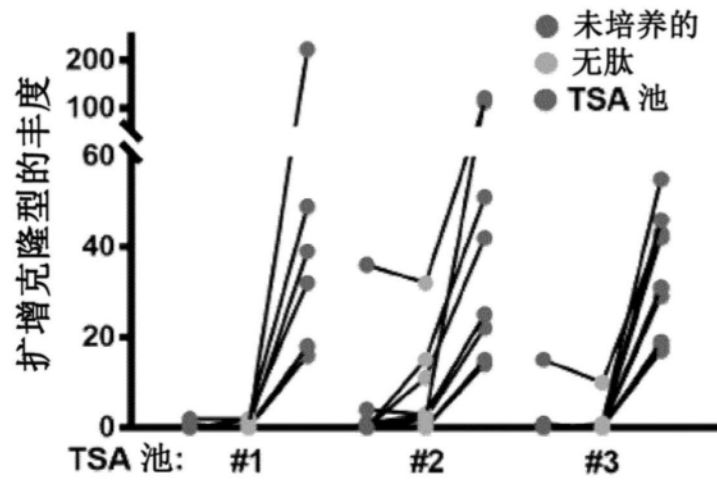


图6G

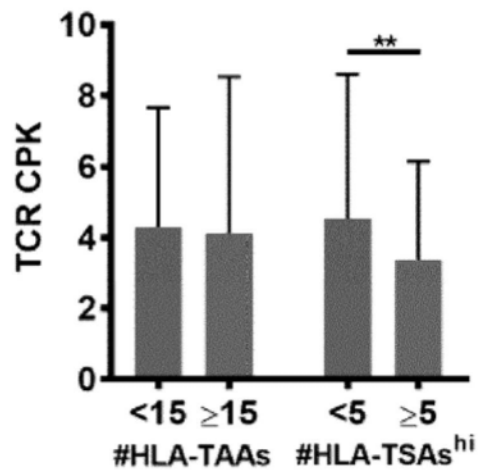


图6H

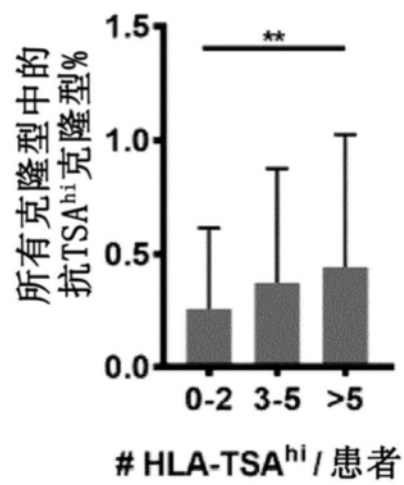


图6I

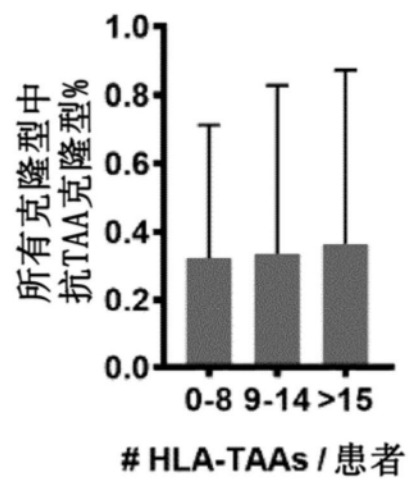


图6J

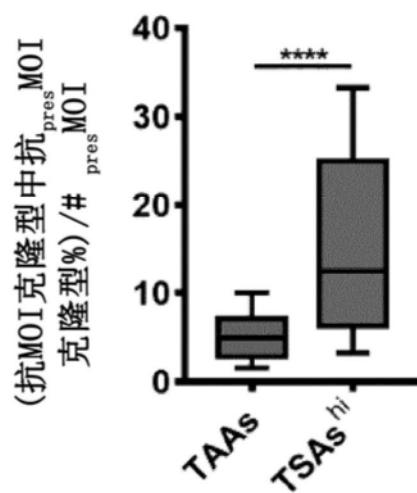


图6K

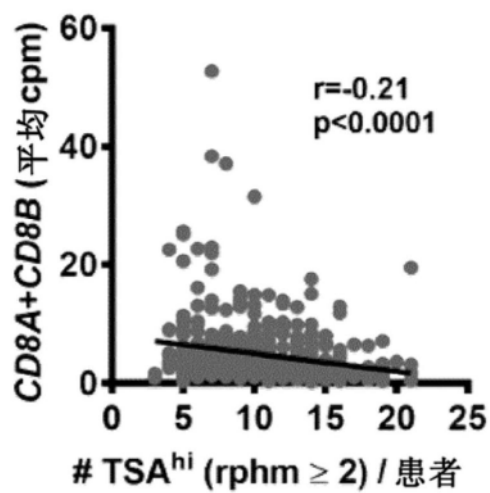


图6L

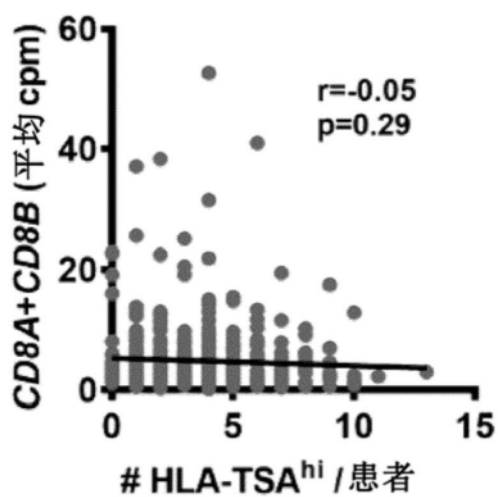


图6M

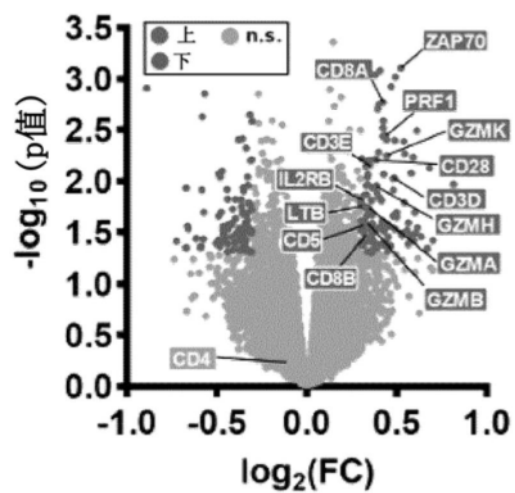


图6N

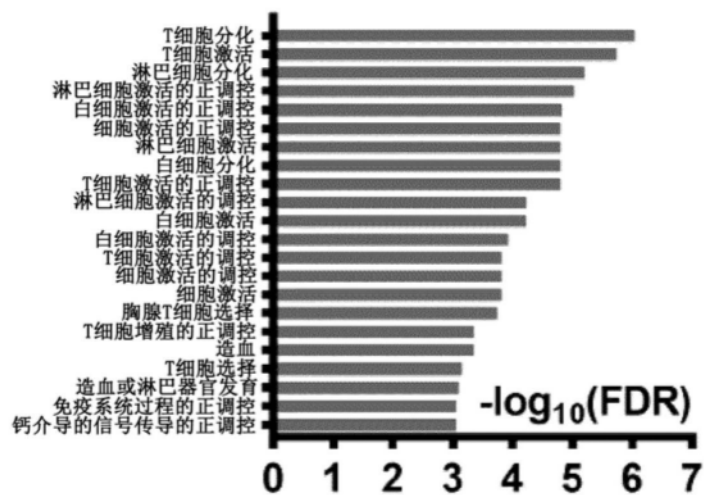


图60

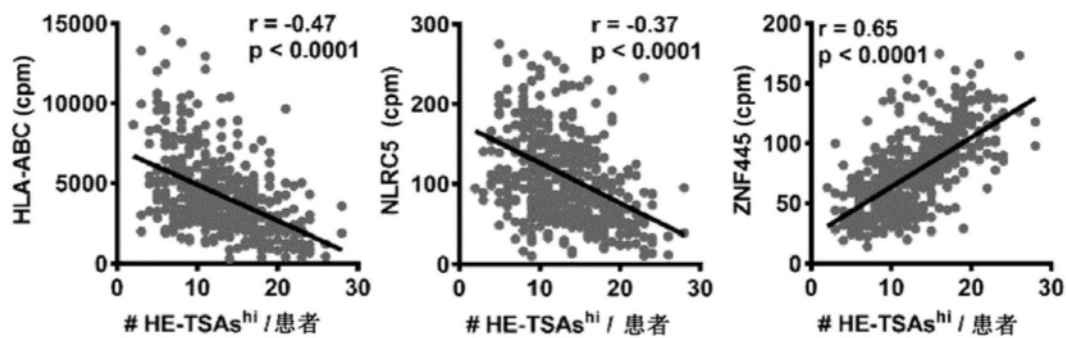


图7A

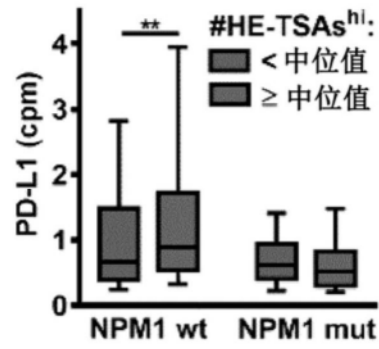


图7B

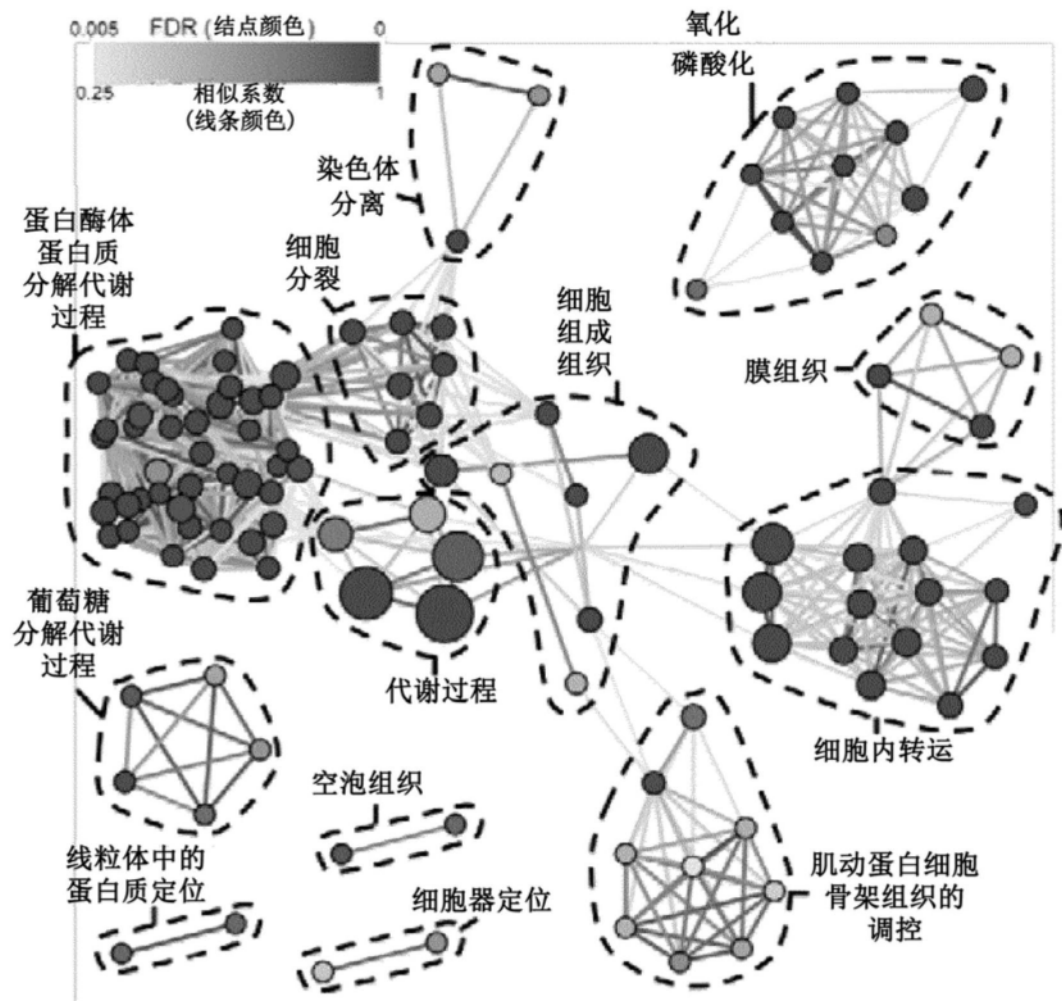


图7C

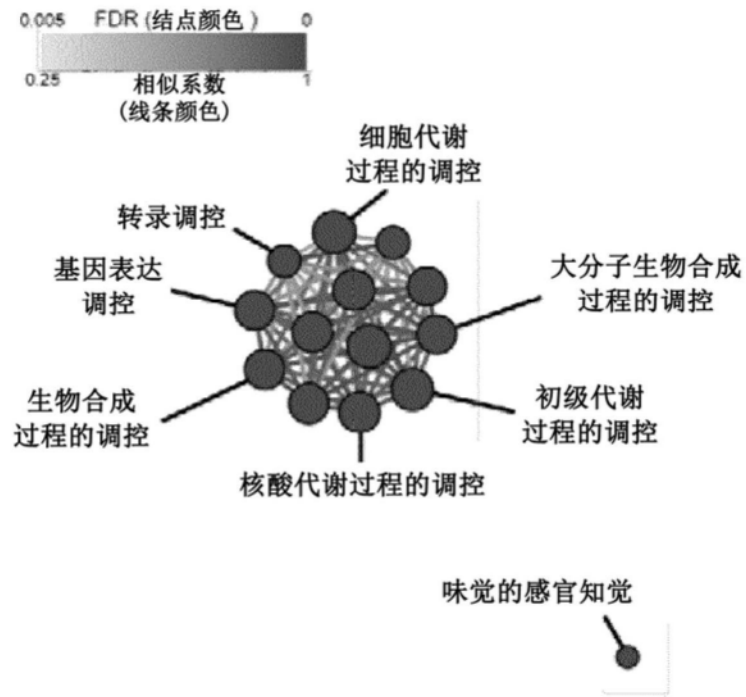


图7D

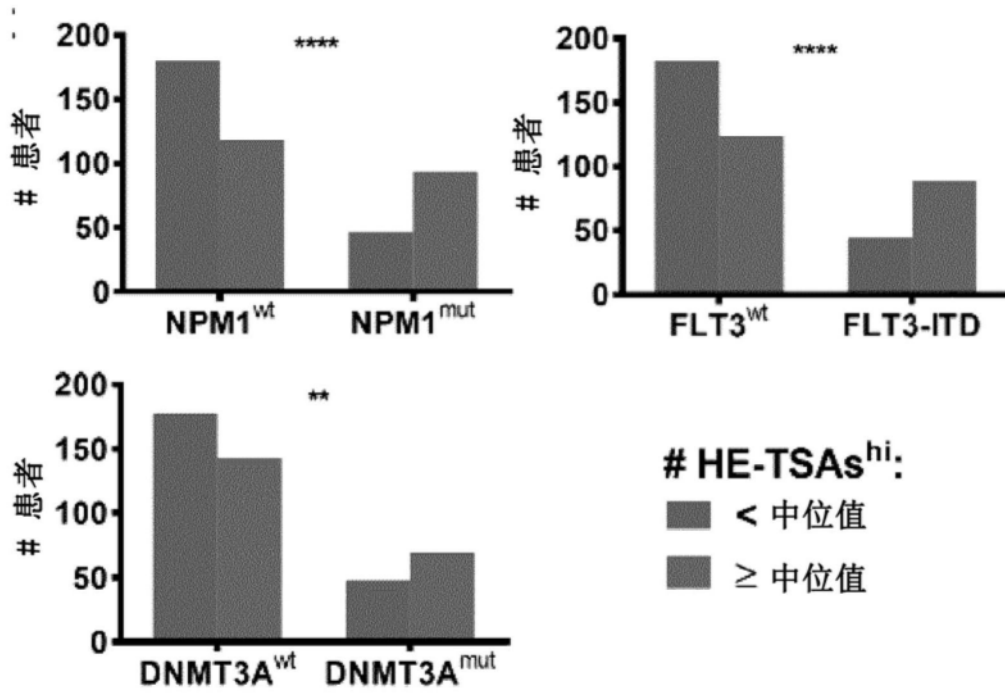


图7E

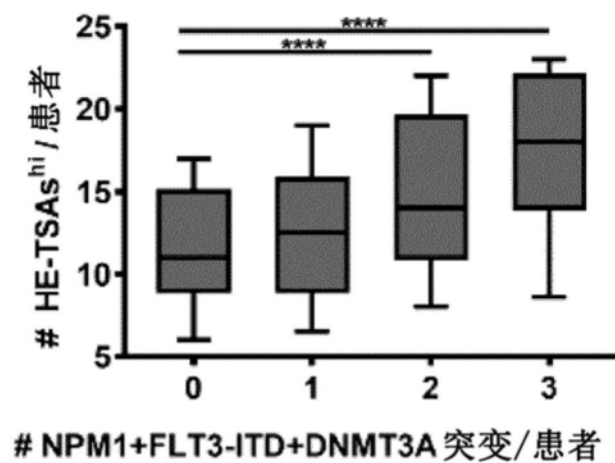


图7F

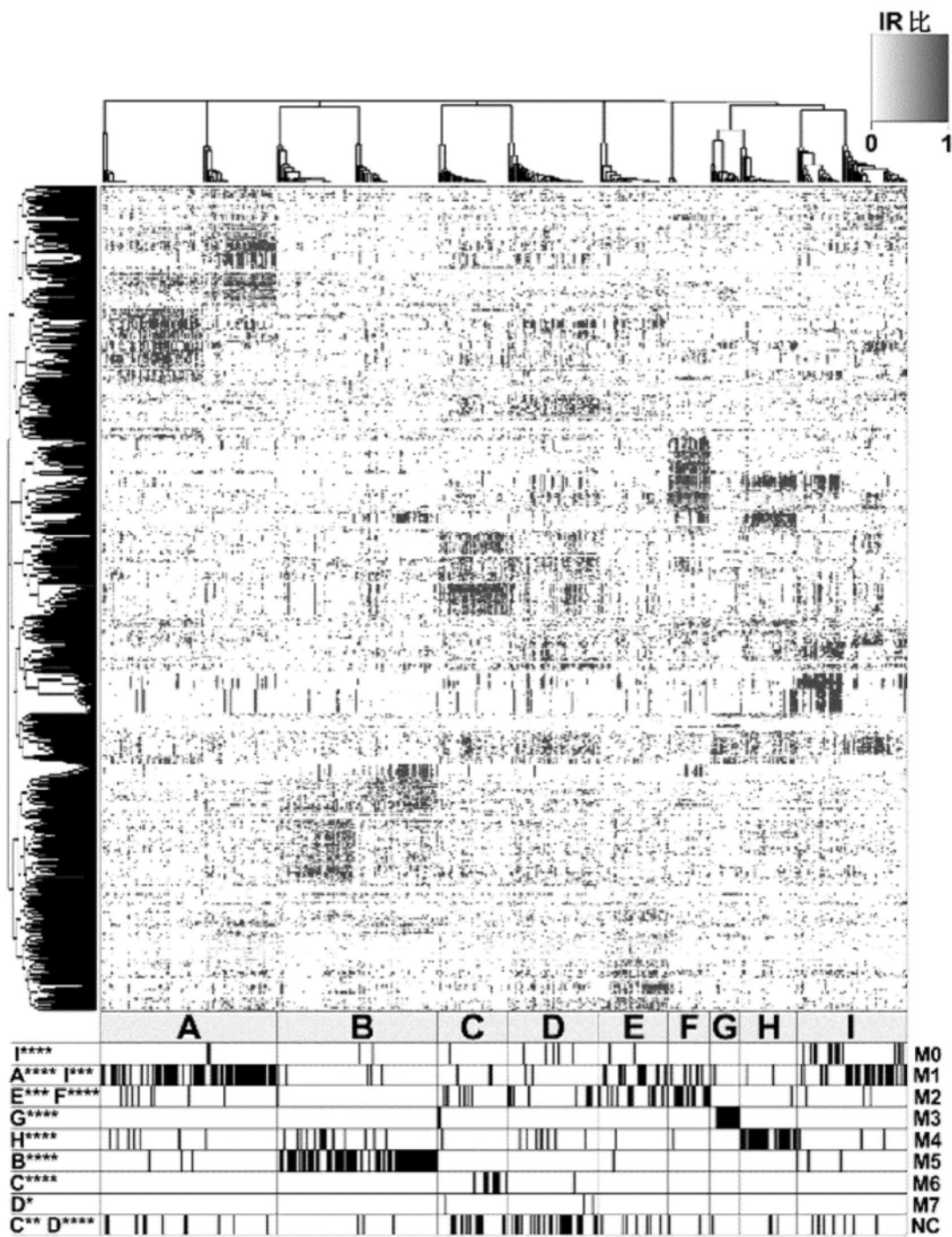


图7G

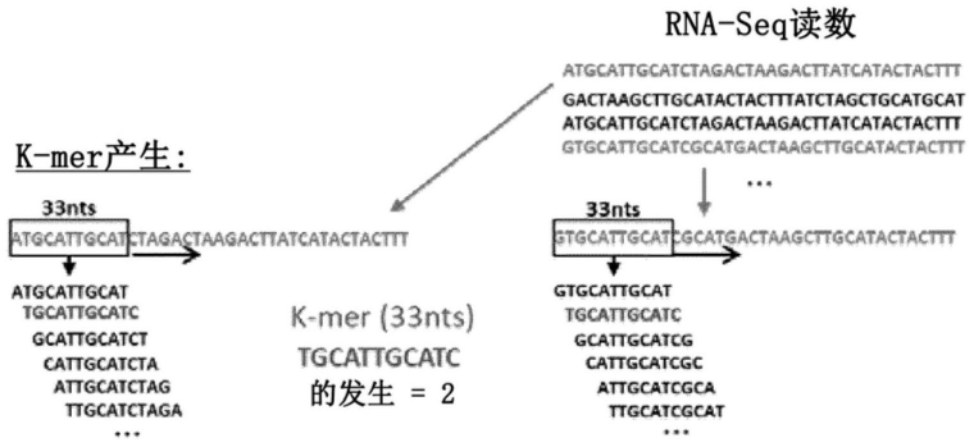


图8A

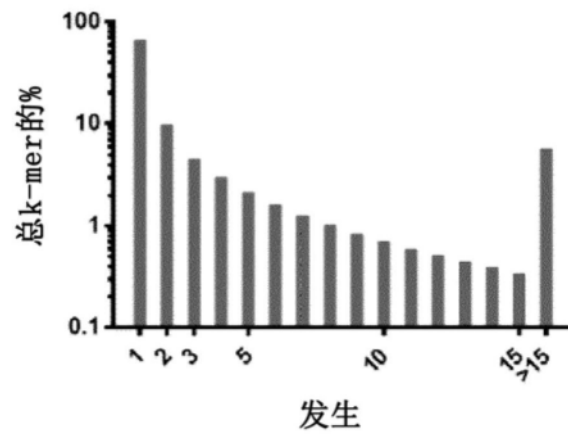


图8B

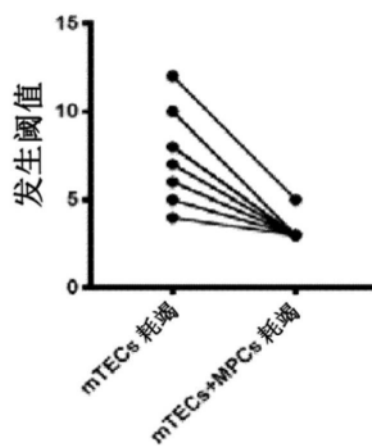


图8C

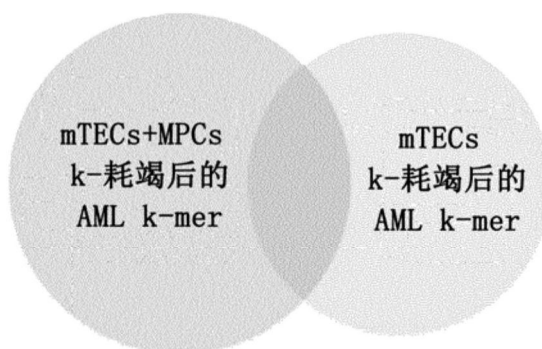


图8D

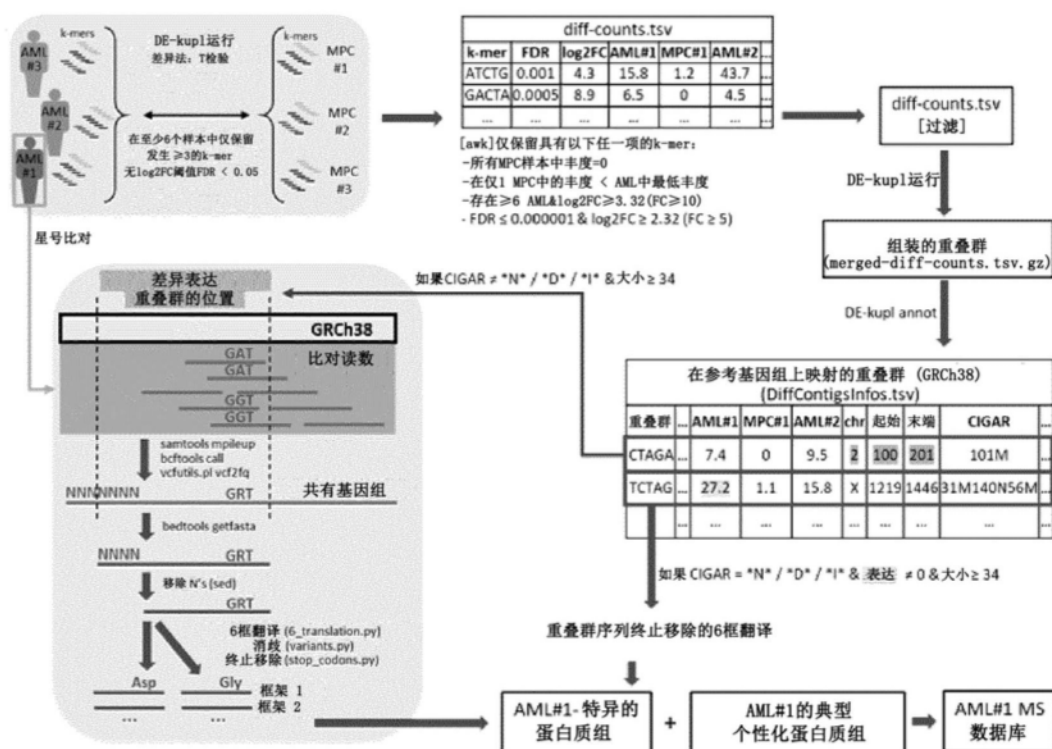


图9A

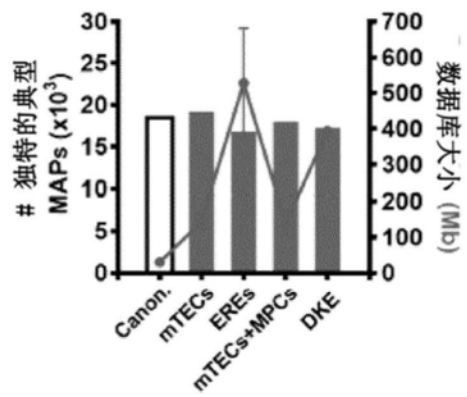


图9B

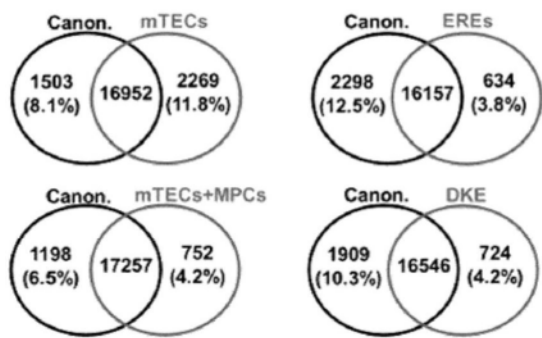


图9C

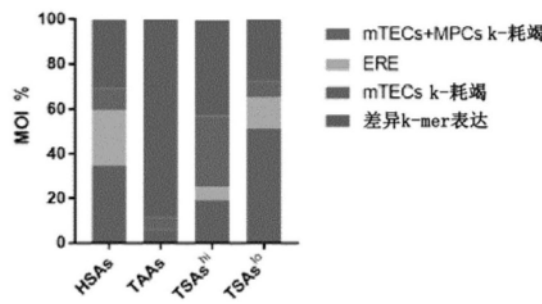


图10A

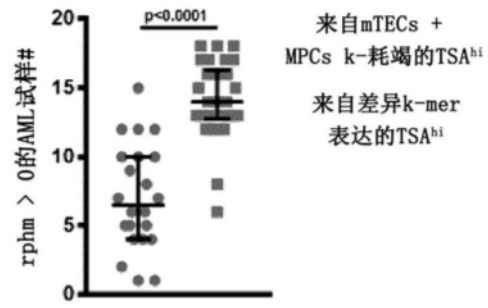


图10B

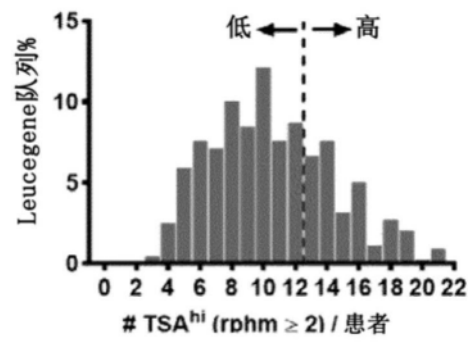


图11A

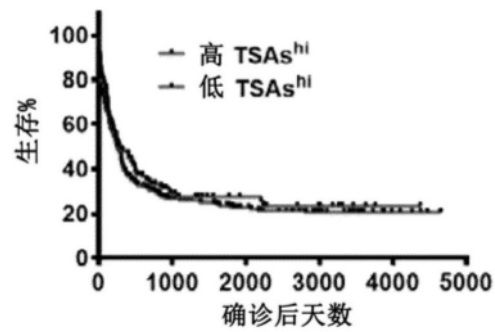


图11B

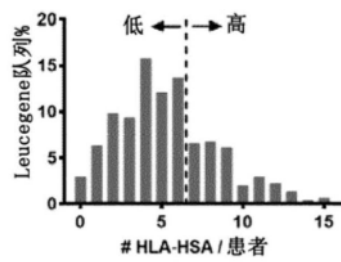


图11C

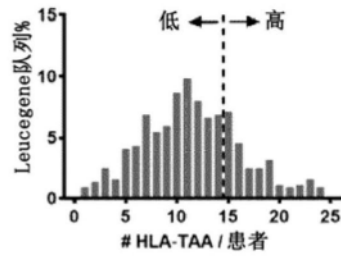


图11D

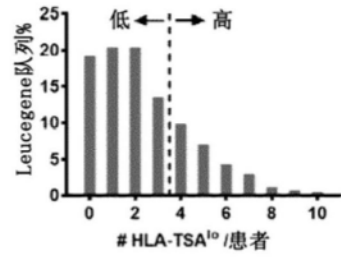


图11E

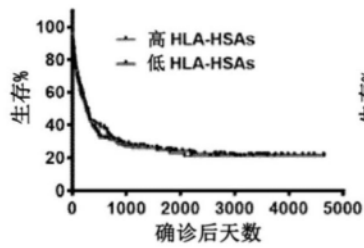


图 11F

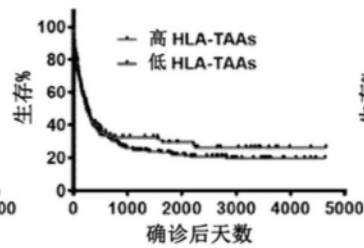


图 11G

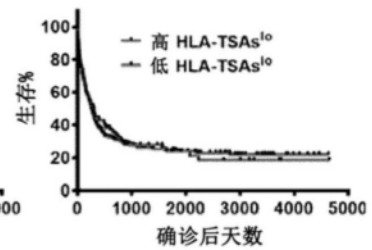


图 11H

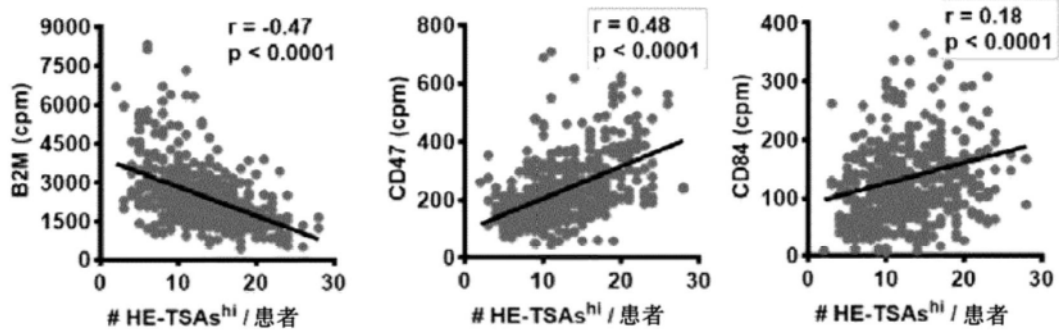


图12A

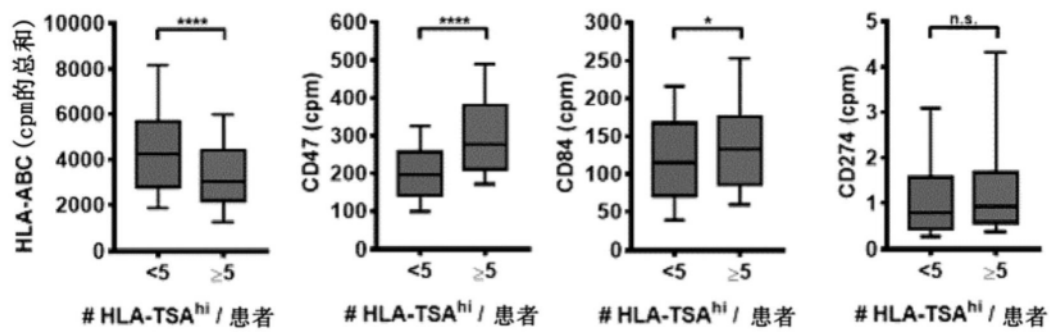


图12B

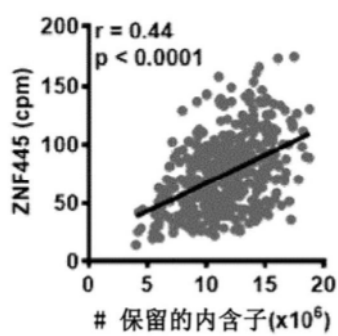


图12C

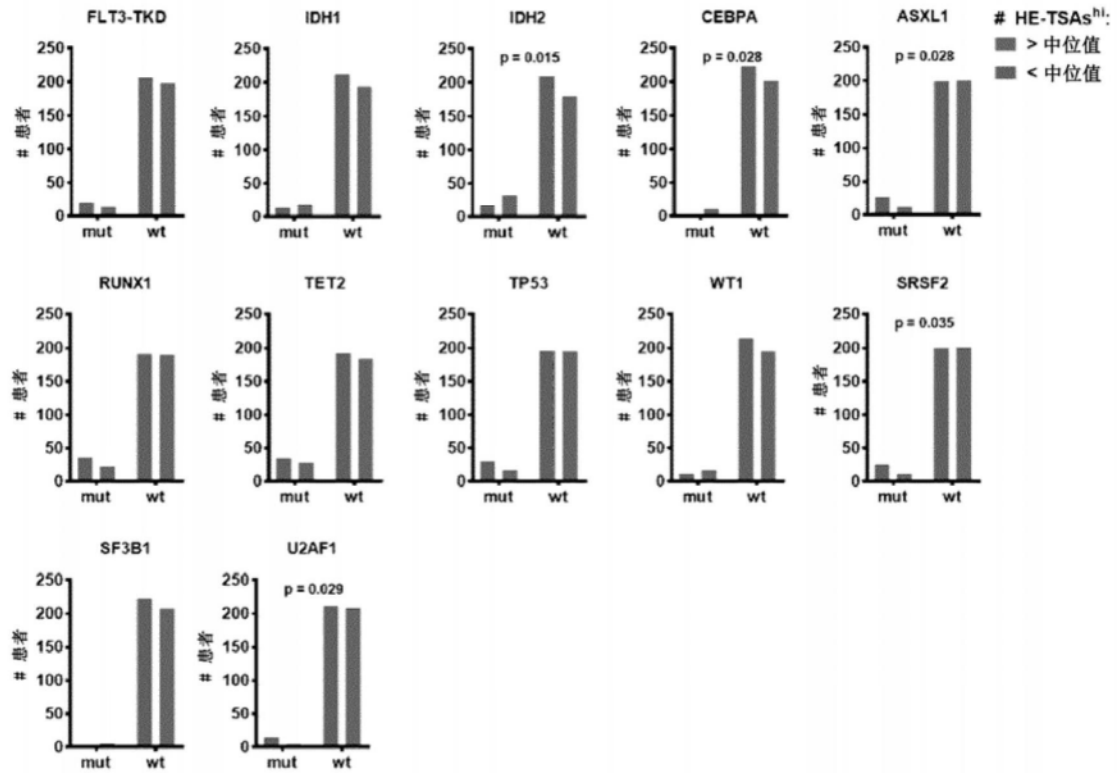


图12D

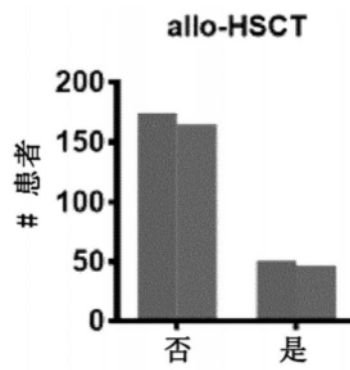


图12E

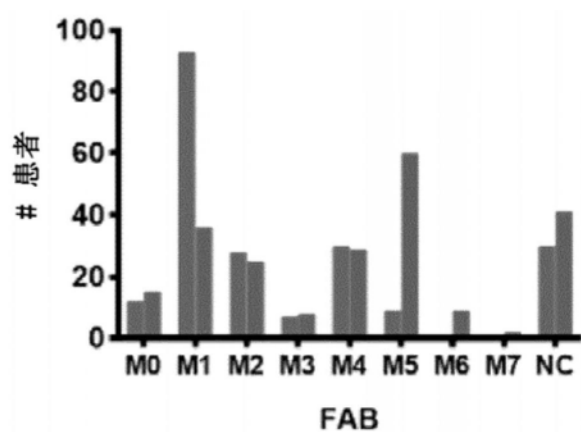


图12F

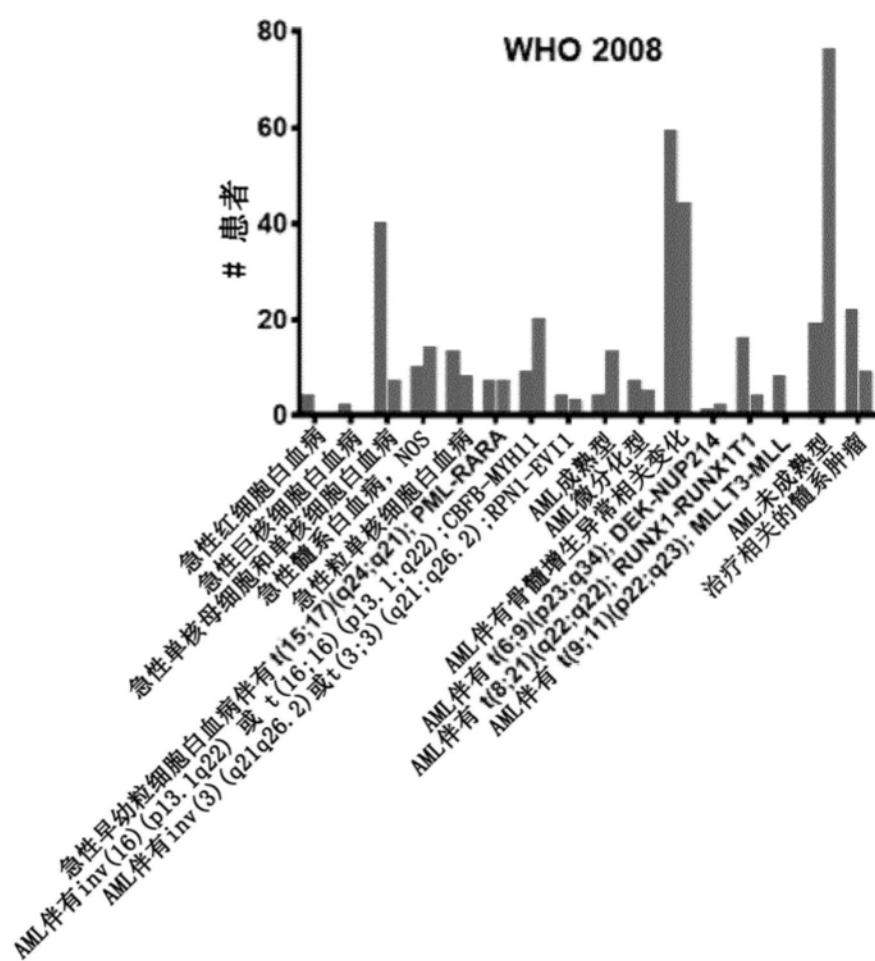


图12G

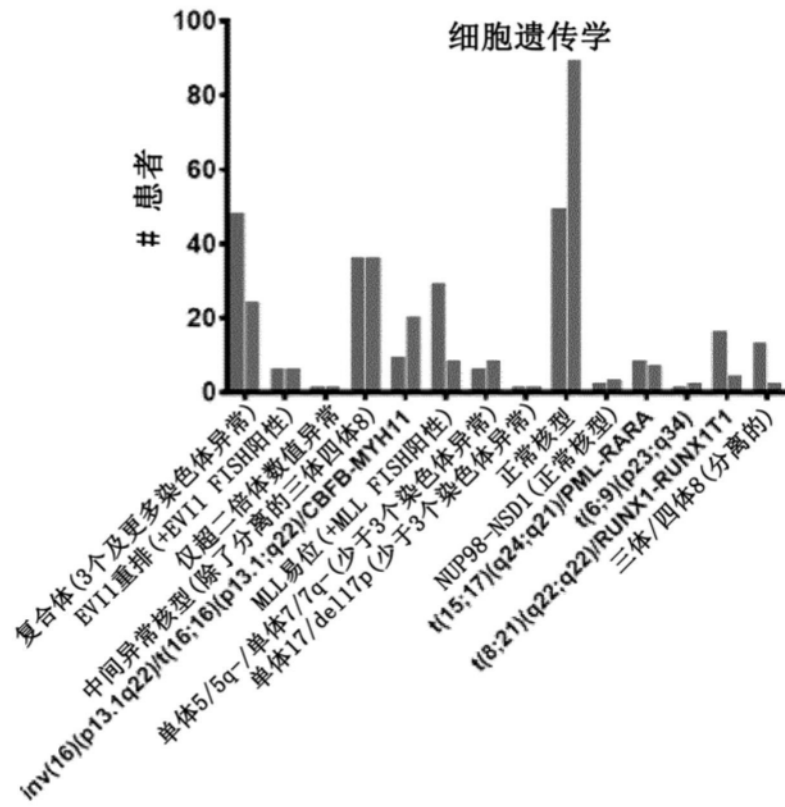


图12H