



(12) 发明专利申请

(10) 申请公布号 CN 113891935 A

(43) 申请公布日 2022.01.04

(21) 申请号 202080018775.5

W·M·戈登 C·B·迈特伽

(22) 申请日 2020.01.03

A·罗特姆 胡刚

(30) 优先权数据

62/788,678 2019.01.04 US

62/791,601 2019.01.11 US

(74) 专利代理机构 上海专利商标事务所有限公  
司 31100

代理人 陶家蓉 陈扬扬

(85) PCT国际申请进入国家阶段日

2021.09.03

(51) Int.Cl.

G12N 15/06 (2006.01)

G12N 15/81 (2006.01)

(86) PCT国际申请的申请数据

PCT/US2020/012236 2020.01.03

G12P 21/00 (2006.01)

(87) PCT国际申请的公布数据

W02020/142724 EN 2020.07.09

(71) 申请人 瑞佩尔托利免疫医药股份有限公司

地址 美国马萨诸塞州

(72) 发明人 J·F·斯温

E·L·L·阿弗扎柳斯

权利要求书5页 说明书35页 附图18页

(54) 发明名称

肽文库及其使用方法

(57) 摘要

本发明涉及肽文库及其用途。



1. 一种包含多种肽的肽文库,其中多种肽包含多于1000、多于2000、多于5000、多于10000、多于 $10^6$ 、多于 $10^7$ 、多于 $10^8$ 、多于 $10^9$ 或多于 $10^{10}$ 种独特肽。
2. 根据权利要求1所述的肽文库,其中所述多种肽包含多种抗原。
3. 根据权利要求1-2中任一权利要求所述的肽文库,其中所述多种肽包含多个pMHC多聚体。
4. 根据权利要求1-3中任一权利要求所述的肽文库,其中所述多种肽包含多个sc-pMHC。
5. 根据权利要求1-4中任一权利要求所述的肽文库,其中所述多种肽中的肽附接到核苷酸序列。
6. 根据权利要求5所述的肽文库,其中所述核苷酸序列为标识符。
7. 根据权利要求5-6中任一权利要求所述的肽文库,其中核苷酸序列编码肽的核苷酸序列。
8. 根据权利要求5-7中任一权利要求所述的肽文库,其中核苷酸序列的长度为25个核苷酸到500个核苷酸。
9. 根据权利要求5-8中任一权利要求所述的肽文库,其中核苷酸序列的长度为80个核苷酸到120个核苷酸。
10. 根据权利要求3-9中任一权利要求所述的肽文库,其中该pMHC的MHC是MHC-I。
11. 根据权利要求3-10中任一权利要求所述的肽文库,其中该pMHC的MHC是HLA-A、HLA-B、HLA-C、HLA-E、HLA-F或HLA-G。
12. 根据权利要求3-11中任一权利要求所述的肽文库,其中该pMHC的MHC是HLA-DPA1、HLA-DPB1、HLA-DQA1、HLA-DQB1、HLA-DRA、或HLA-DRB1。
13. 一种包含权利要求1-12中任一权利要求所述的多种肽的肽子集文库,其中在权利要求1-12中任一权利要求所述的多种肽中的肽结合到T细胞的TCR时诱导T细胞增殖。
14. 一种包含权利要求1-12中任一权利要求所述的多种肽的肽子集文库,其中在权利要求1-12中任一权利要求所述的多种肽中的肽结合到T细胞的TCR时诱导T细胞细胞毒性。
15. 一种包含权利要求1-12中任一权利要求所述的多种肽的肽子集文库,其中在权利要求1-12中任一权利要求所述的多种肽中的肽结合到T细胞的TCR时诱导T细胞抑制。
16. 一种包含权利要求1-12中任一权利要求所述的多种肽的肽子集文库,其中在权利要求1-12中任一权利要求所述的多种肽中的肽结合到T细胞的TCR时诱导T细胞引起的抑制。
17. 一种包含权利要求1-12中任一权利要求所述的多种肽的肽子集文库,其中在权利要求1-12中任一权利要求所述的多种肽中的肽结合到T细胞的TCR时诱导T细胞的细胞因子产生。
18. 一种分离淋巴细胞-肽对的方法,包括:
  - (a) 将多个淋巴细胞与肽文库接触,其中肽文库的多样性大于1000;和
  - (b) 产生多个隔室,其中多个隔室中的隔室包括 (i) 与肽文库的肽结合的多个淋巴细胞中的淋巴细胞,以及 (ii) 捕获载体。
19. 根据权利要求18所述的方法,其中淋巴细胞是T细胞、B细胞或NK细胞。
20. 根据权利要求18至19中任一权利要求所述的方法,其中该肽库具有大于2000、大于

5000、大于10000、大于 $10^6$ 、大于 $10^7$ 、大于 $10^8$ 、大于 $10^9$ 或大于 $10^{10}$ 个独特肽的多样性。

21. 根据权利要求18-20中任一权利要求所述的方法,其中所述多种肽包含多个抗原。

22. 根据权利要求18-21中任一权利要求所述的方法,其中所述多种肽包含多个pMHC多聚体。

23. 根据权利要求18-22中任一权利要求所述的方法,其中所述多种肽包含多个sc-pMHC。

24. 根据权利要求18-23中任一权利要求所述的肽库,其中所述多种肽中的肽附接到核苷酸序列。

25. 根据权利要求24所述的方法,其中所述核苷酸序列为标识符。

26. 根据权利要求24-25中任一权利要求所述的方法,其中核苷酸序列编码肽的核苷酸序列。

27. 根据权利要求24-26中任一权利要求所述的方法,其中核苷酸序列的长度为25个核苷酸到500个核苷酸。

28. 根据权利要求24-27中任一权利要求所述的方法,其中核苷酸序列的长度为80个核苷酸到120个核苷酸。

29. 根据权利要求22-28中任一权利要求所述的方法,其中该pMHC的MHC是MHC-I。

30. 根据权利要求22-29中任一权利要求所述的方法,其中该pMHC的MHC是HLA-A、HLA-B、HLA-C、HLA-E、HLA-F或HLA-G。

31. 根据权利要求22-30中任一权利要求所述的方法,其中该pMHC的MHC是HLA-DPA1、HLA-DPB1、HLA-DQA1、HLA-DQB1、HLA-DRA、或HLA-DRB1。

32. 一种鉴定淋巴细胞-肽对的方法,包括:

(a) 将多个淋巴细胞与肽文库接触,其中肽文库的多样性大于1000;和

(b) 将与肽文库中的肽结合的多个淋巴细胞中的淋巴细胞分隔在单个隔室中,其中肽包含独特肽标识符;和

(c) 确定与分隔的淋巴细胞结合的每条肽的独特肽标识符。

33. 根据权利要求32所述的方法,其中淋巴细胞是T细胞、B细胞或NK细胞。

34. 根据权利要求32至33中任一权利要求所述的方法,其中该肽文库具有大于2000、大于5000、大于10000、大于 $10^6$ 、大于 $10^7$ 、大于 $10^8$ 、大于 $10^9$ 或大于 $10^{10}$ 个独特肽的多样性。

35. 根据权利要求32-34中任一权利要求所述的方法,其中所述多种肽包含多个抗原。

36. 根据权利要求32-35中任一权利要求所述的方法,其中所述多种肽包含多个pMHC多聚体。

37. 根据权利要求32-36中任一权利要求所述的方法,其中所述多种肽包含多个sc-pMHC。

38. 根据权利要求36-37中任一权利要求所述的方法,其中该pMHC的MHC是MHC-I。

39. 根据权利要求36-38中任一权利要求所述的方法,其中该pMHC的MHC是HLA-A、HLA-B或HLA-C。

40. 根据权利要求36-39中任一权利要求所述的方法,其中该pMHC的MHC是HLA-DPA1、HLA-DPB1、HLA-DQA1、HLA-DQB1、HLA-DRA、或HLA-DRB1。

41. 根据权利要求32-40中任一权利要求所述的方法,其中该淋巴细胞-肽对是TCR-抗

原对。

42. 根据权利要求32-41中任一权利要求所述的方法,其进一步包括确定所述分隔的淋巴细胞上受体的身份。

43. 根据权利要求42所述的方法,其中确定身份包括对TCR、BCR或抗体的可变区、高变区或互补确定区(CDR)进行测序。

44. 根据权利要求43所述的方法,其中CDR是TCR $\alpha$ 链、TCR $\beta$ 链、TCR $\gamma$ 链、TCR $\delta$ 链、抗体重链或抗体轻链的CDR1、CDR2或CDR3。

45. 一种使用无偏肽文库的方法,该方法包括使样品接触包含多种肽的无偏肽文库,其中多种肽包含多于100、多于1000、多于2000、多于5000、多于10000、多于 $10^6$ 、多于 $10^7$ 、多于 $10^8$ 、多于 $10^9$ 或多于 $10^{10}$ 种独特肽。

46. 根据权利要求45所述的方法,其中该无偏肽文库由生物体的基因组产生。

47. 根据权利要求45所述的方法,其中该无偏肽文库由生物体的转录组产生。

48. 根据权利要求45所述的方法,其中该无偏肽文库由生物体的蛋白质组产生。

49. 根据权利要求45所述的方法,其中该无偏肽文库由生物体的肽或蛋白质产生。

50. 根据权利要求45所述的方法,其中该无偏肽文库由生物体的表位产生。

51. 根据权利要求45所述的方法,其中该无偏肽文库由基因组之间的差异序列产生。

52. 根据权利要求45所述的方法,其中该无偏肽文库由转录组之间的差异序列产生。

53. 根据权利要求45所述的方法,其中该无偏肽文库由蛋白质组之间的差异序列产生。

54. 根据权利要求51-53中任一权利要求所述的方法,其中所述差异序列包含来自患病细胞相对于健康细胞的序列。

55. 根据权利要求51-54中任一权利要求所述的方法,其中所述差异序列包含来自癌细胞相对于健康细胞的序列。

56. 根据权利要求45所述的方法,其中该无偏肽文库由基因组之间的同源序列产生。

57. 根据权利要求45所述的方法,其中该无偏肽文库由转录组之间的同源序列产生。

58. 根据权利要求45所述的方法,其中该无偏肽文库由蛋白质组之间的同源序列产生。

59. 根据权利要求56-58中任一权利要求所述的方法,其中所述同源序列包含来自患病细胞相对于健康细胞的序列。

60. 根据权利要求56-59中任一权利要求所述的方法,其中所述同源序列包含与自身免疫相关的细胞相对于健康细胞的序列。

61. 根据权利要求45所述的方法,其中该无偏肽文库包含无偏pMHC多聚体文库。

62. 根据权利要求45所述的方法,其中该无偏肽文库包含无偏sc-pMHC文库。

63. 根据权利要求61所述的方法,其中所述无偏pMHC多聚体文库的pMHC多聚体的抗原是无偏的。

64. 根据权利要求62所述的方法,其中所述无偏sc-pMHC文库的sc-pMHC的抗原是无偏的。

65. 一种组合物,包括附接到独特标识符的pMHC多聚体。

66. 根据权利要求65所述的组合物,其中,所述pMHC多聚体是sc-pMHC。

67. 根据权利要求65或66所述的组合物,其中该独特标识符是核酸。

68. 根据权利要求65-67中任一权利要求所述的组合物,其中,所述独特标识符是自身

标识符。

69. 根据权利要求65-67中任一权利要求所述的组合物,其中,所述独特标识符不是自身标识符。

70. 根据权利要求65-69中任一权利要求所述的组合物,其中独特标识符的长度为25个核苷酸到120个核苷酸。

71. 一种隔室,其包括:

- (a) 编码sc-pMHC的序列;和
- (b) T细胞。

72. 一种组合物,其包含:

- (a) 水凝胶珠;和
- (b) 附接到水凝胶珠的核酸,其中该核酸编码肽。

73. 根据权利要求72所述的组合物,其进一步包含附接到该水凝胶珠的第二核酸,其中该第二核酸包含标识符。

74. 根据权利要求73所述的组合物,其中,所述标识符是自身标识符。

75. 根据权利要求73所述的组合物,其中,所述标识符不是自身标识符。

76. 根据权利要求72-75中任一权利要求所述的组合物,其进一步包含肽,其中肽附接到水凝胶珠。

77. 根据权利要求73至75中任一权利要求所述的组合物,其进一步包含肽,其中肽附接至第三核酸且该第三核酸附接至水凝胶珠。

78. 根据权利要求77所述的组合物,其中,所述第三核酸包括标识符。

79. 根据权利要求72至78中任一权利要求所述的组合物,其中水凝胶珠包封在液滴中。

80. 根据权利要求72-79中任一权利要求所述的组合物,其中,所述肽包含sc-pMHC。

81. 根据权利要求80所述的组合物,其中sc-pMHC的MHC是MHC-I。

82. 根据权利要求80所述的组合物,其中sc-pMHC的MHC是HLA-A、HLA-B、HLA-C、HLA-E、HLA-F或HLA-G。

83. 根据权利要求80所述的方法,其中sc-pMHC的MHC是HLA-DPA1、HLA-DPB1、HLA-DQA1、HLA-DQB1、HLA-DRA、或HLA-DRB1。

84. 一种生成标识符标记的肽的方法,所述方法包括:

- (a) 提供捕获载体,其中所述捕获载体包含包括标识符的附着核酸;
- (b) 将肽附接到核酸上;
- (c) 从捕获载体分离核酸或其部分,从而释放标识符标记的肽。

85. 根据权利要求84所述的方法,其中,所述标识符是自身标识符。

86. 根据权利要求84所述的方法,其中,所述标识符不是自身标识符。

87. 根据权利要求84-86中任一权利要求所述的方法,其中所述肽包含抗原。

88. 根据权利要求84-87中任一权利要求所述的方法,其中,所述肽包含sc-pMHC。

89. 根据权利要求88所述的方法,其中sc-pMHC的MHC是MHC-I。

90. 根据权利要求88所述的方法,其中sc-pMHC的MHC是HLA-A、HLA-B、HLA-C、HLA-E、HLA-F或HLA-G。

91. 根据权利要求88所述的方法,其中sc-pMHC的MHC是HLA-DPA1、HLA-DPB1、HLA-DQA1、

HLA-DQB1、HLA-DRA、或HLA-DRB1。

92. 根据权利要求84-91中任一权利要求所述的方法,其中所述分离包括酶消化。

93. 根据权利要求84至92中任一权利要求所述的方法,其中捕获载体包封在液滴中。

94. 根据权利要求84-93中任一权利要求所述的方法,其中核酸的长度为25个核苷酸到500个核苷酸。

95. 根据权利要求84-93中任一权利要求所述的方法,其中核酸的长度为80个核苷酸到120个核苷酸。

96. 一种生成肽文库的方法,包括通过权利要求84-95中任一权利要求所述的方法生成多个标识符标记的肽。

97. 一种生成标识符标记的肽的方法,所述方法包括:

(a) 提供捕获载体,其中所述捕获载体具有附接的编码肽的第一核酸和附接的包含标识符的第二核酸;

(b) 产生肽;

(c) 将肽附接到所述第二核酸,从而产生标识符标记的肽;和

(d) 从捕获载体分离第二核酸或其部分,从而释放标识符标记的肽。

98. 根据权利要求97所述的方法,其中,所述标识符是自身标识符。

99. 根据权利要求97所述的方法,其中,所述标识符不是自身标识符。

100. 根据权利要求97-99中任一权利要求所述的方法,其中,所述产生包括体外转录和翻译。

101. 根据权利要求97-100中任一权利要求所述的方法,其中所述肽包含抗原。

102. 根据权利要求97-101中任一权利要求所述的方法,其中,所述肽包含sc-pMHC。

103. 根据权利要求102所述的方法,其中sc-pMHC的MHC是MHC-I。

104. 根据权利要求102所述的方法,其中sc-pMHC的MHC是HLA-A、HLA-B、HLA-C、HLA-E、HLA-F或HLA-G。

105. 根据权利要求102所述的方法,其中sc-pMHC的MHC是HLA-DPA1、HLA-DPB1、HLA-DQA1、HLA-DQB1、HLA-DRA、或HLA-DRB1。

106. 根据权利要求97-105中任一权利要求所述的方法,其中所述分离包括酶消化。

107. 根据权利要求97至106中任一权利要求所述的方法,其中捕获载体包封在液滴中。

108. 根据权利要求97-107中任一权利要求所述的方法,其中第二核酸的长度为25个核苷酸到500个核苷酸。

109. 根据权利要求97-107中任一权利要求所述的方法,其中第二核酸的长度为80个核苷酸到120个核苷酸。

110. 一种生成肽文库的方法,包括通过权利要求97-109中任一权利要求所述的方法生成多个标识符标记的肽。

## 肽文库及其使用方法

### 交叉引用

[0001] 本申请要求于2019年1月4日提交的美国临时申请62/788,678和2019年1月11日提交的美国临时申请62/791,601的权益,这些申请以其全部内容通过引用纳入本文。

### 背景

[0002] 在生产和使用不同的肽文库方面存在着许多挑战。

### 通过引用纳入

[0003] 本说明书中提到的所有发表物、专利和专利申请通过引用纳入本文,就好像将各篇单独的发表物、专利或专利申请专门和单独地通过引用纳入本文那样。

### 发明内容

[0004] 本发明提供肽文库的组合物和这些肽文库的使用方法。

[0005] 在一些实施方式中,本文公开了包含多种肽的肽文库,其中多种肽包含多于1000、多于2000、多于5000、多于10000、多于 $10^6$ 、多于 $10^7$ 、多于 $10^8$ 、多于 $10^9$ 或多于 $10^{10}$ 种独特肽。

[0006] 在一些实施方式中,本文公开了一种分离淋巴细胞肽对的方法,包括:(a)将多个淋巴细胞与肽文库接触,其中肽库的多样性大于1000;和(b)产生多个隔室,其中多个隔室包括(i)与肽文库的肽结合的多个淋巴细胞中的淋巴细胞,以及(ii)捕获载体。

[0007] 在一些实施方式中,本文公开了一种鉴定淋巴细胞-肽对的方法,包括:(a)将多个淋巴细胞与肽文库接触,其中肽文库的多样性大于1000;(b)将与肽文库的肽结合的多个淋巴细胞中的淋巴细胞分隔在单个隔室中,其中肽包含独特肽标识符;和(c)确定与分隔的淋巴细胞结合的每种肽的独特肽标识符。

[0008] 在一些实施方式中,本文公开了一种使用无偏肽文库的方法,该方法包括使样品接触无偏肽文库,所述文库包含多种肽,其中多种肽包含多于100、多于1000、多于2000、多于5000、多于10000、多于 $10^6$ 、多于 $10^7$ 、多于 $10^8$ 、多于 $10^9$ 或多于 $10^{10}$ 种独特肽。

[0009] 在一些实施方式中,本文公开了一种组合物,其包含附接到独特标识符上的pMHC多聚体。

[0010] 在一些实施方式中,本文公开了一种隔室,其包括:(a)编码sc-pMHC的序列;和(b)T细胞。

[0011] 为了更好地理解本发明的本质和优势,参考以下与附图结合的详细说明。本发明能够在不偏离本发明的情况下进行各个方面的修改。因此,这些实施方式的附图和描述不是限制性的。

### 附图简要说明

[0012] 本专利或申请文件包含至少一幅有色附图。本专利或专利申请公开的带彩色附图的副本将根据要求,在支付所需的费用之后由政府机关提供。所附权利要求书中具体说明了本发明的新特征。可参考以下详述更好地理解本发明的特征和优点,这些详述阐述了利用本发明原理的说明性实施方式和附图,其中:

[0013] 图1提供了不同多样性的肽文库的示范性实施方式。

[0014] 图2A显示酶切以增加肽多样性。泳道对1-2、3-4、5-6、7-8和9-10分别表示在没有切割部分的模板、具有未添加蛋白酶的切割部分的模板、具有反应完成后添加蛋白酶的切割部分的模板、具有反应期间存在蛋白酶的切割部分的模板上进行无细胞蛋白合成 (CFPS) 反应以及缺乏模板的反应。进行蛋白质印迹以确定总蛋白质产量。

[0015] 图2B显示了来自CFPS反应的样品,其中包含有或没有切割部分的多聚体和单体模板。用抗FLAG-HRP抗体对样品进行印迹分析和检测。

[0016] 图2C显示由CFPS产生并经蛋白水解切割的肽折叠成可识别的三维结构。测试CFPS蛋白的抗体构象识别。该图表明通过构象表位抗体的识别证明了蛋白酶切割的肽或未切割的肽(包含fMet)是否为正确折叠。

[0017] 图2D提供单链肽MHC (sc-pMHC) 多聚体与抗原特异性T细胞的结合。通过CFPS和蛋白酶切割产生多聚体。将T细胞与多聚体孵育,然后用荧光检测抗体染色,并通过流式细胞术进行分析。

[0018] 图3是一种印迹,显示标记有一个或多个肽标识符的肽,如肽带向上漂移所示。

[0019] 图4显示裸露CMV肽、用肽标识符标记的CMV肽或阴性对照与CMV或非CMV特异性T细胞的相对结合的流式细胞术结果。

[0020] 图5显示肽与T细胞结合后检测到的肽标识符相对量的qPCR结果。数量对管家基因标准化。

[0021] 图6说明了本发明文库中折叠肽的富集。

[0022] 图7说明了基于核酸标识符测序的文库多样性验证。

[0023] 图8显示了通过分析使用本发明的组合物和方法生成的测序数据来鉴定肽(抗原)-受体(TCR)对。

[0024] 图9说明了使用本发明的组合物和方法生成的肽(抗原)-受体(TCR)对数据的分析。图9A示出了通过TCR进行聚类以显示抗原特异性图谱。图9B示出了通过抗原的tSNE聚类以显示TCR结合收敛。图9C示出了不同TCR对象间的收敛。

[0025] 图10说明了sc-pMHC与TCR的结合以及抗原突变对结合的影响。

[0026] 图11说明了基于使用本文所述方法鉴定的感兴趣细胞的基因表达的示例性免疫表型分型。

[0027] 图12说明了sc-pMHC与使用本文所述方法鉴定的示例性抗原的抗原特异性T细胞的特异性和剂量依赖性结合。

[0028] 图13说明了响应使用本文所述方法鉴定的示例性抗原的关联sc-pMHC时T细胞的增殖和细胞因子生产。

[0029] 图14显示了示例性文库大小。图14A示出了示例性病毒组文库大小。图14B示出了示例性癌症文库大小。

[0030] 图15示出了本发明(例如,根据实施例21-25)的肽文库生产的示例性概述。

[0031] 图16提供了可用于本发明组合物和方法中的核酸构建体的示意图。该构建体可编码本发明的肽(例如,sc-pMHC)和标识符(例如,对应于肽的全部或部分编码序列的自身标识符)。指示了正向和反向引物的位置(例如,可在实施例21-23中使用的引物)。

[0032] 图17显示了如实施例22和23所述,在水凝胶上PCR扩增全长抗原编码模板(图17A)和标识符(图17B)。

[0033] 图18显示了本发明的折叠和标识符标记的sc-pMHC的生成。图18A提供了显微镜图像,其证明本发明的sc-pMHC是如实施例24所述在体外转录和翻译的。图18B提供ELISA结果,证明折叠的sc-pMHC多聚体的释放,如实施例25所述。图18C提供了蛋白质印迹,证明sc-pMHC标有一个标识符,如实施例25所述。

[0034] 图19提供了流式细胞术分析的结果,证明通过本发明方法产生的sc-pMHC与关联T细胞特异性结合,如实施例26所述。

[0035] 图20提供了单细胞测序分析的结果,证明通过本发明方法产生的sc-pMHC与关联T细胞特异性结合,如实施例26所述。

## 具体实施方式

### 定义

[0036] 如本文所使用的,术语“标识符”是指提供与标识符相对应的信息(例如特性)的数据的可读表示。

[0037] 如本文所用,术语“多聚体”指的是多个单元。一些实施方式中,多聚体包含一种或多种不同单元。在一些实施方式中,多聚体中的单元是相同的。在一些实施方式中,多聚体中的单元是不同的。在一些实施方式中,多聚体包含相同和不同单元的混合物。

[0038] 如本文所用,术语“肽文库”指的是多种肽。在一些实施方式中,文库包含具有独特序列的一种或多种肽。在一些实施方式中,文库中的每种肽具有不同的序列。在一些实施方式中,文库包含具有相同和不同序列的肽的混合物。

[0039] 如本文所使用的,术语“无偏”指缺少一个或多个选择标准。

[0040] 如本文所使用的,术语“捕获载体”指相互作用表面。在一些实施方式中,捕获载体可以是固相表面。在一些实施方式中,捕获载体可以包含基质。在一些实施方式中,捕获载体可以包含纳米粒。在一些实施方式中,捕获载体可以包含珠。在一些实施方式中,捕获载体可以包含磁珠。在一些实施方式中,捕获载体可以包含水凝胶。在一些实施方式中,捕获载体可以是油包水乳液液滴的内表面。在一些实施方式中,捕获载体可以包含核酸分子。在一些实施方式中,捕获载体可以包含蛋白质。在一些实施方式中,捕获载体可包含抗体或其衍生物。在一些实施方式中,捕获载体可以包含凝胶。在一些实施方式中,捕获载体可以包含聚合物。在一些实施方式中,捕获探针可以改变。在一些实施方式中,捕获载体可以是荧光的,例如,用一种或多种荧光染料标记。

### 引言

[0041] 本公开提供了一种肽文库。肽文库可包含具有不同氨基酸序列的多种肽。肽文库可用于一系列筛选试验,以鉴定潜在的诊断或治疗靶点或试剂。例如,肽文库可用于筛选疾病特异性、器官特异性肽、或其它隔室特异性肽,用于筛选具有治疗应用的肽,筛选具有诊断应用的肽,筛选肿瘤靶向肽,筛选抗体表位或抗原,筛选T细胞表位或抗原,筛选抗微生物肽,或其任何组合。因此,具有适当质量的多种肽文库具有许多有价值的用途。

[0042] 肽文库可以是抗原文库。抗原文库应用的非限制性示例包括用于与免疫抗原相关的分析、治疗和诊断。例如,抗原文库可用于筛选以鉴定蛋白质表位,例如抗体或T细胞表位。抗体和T细胞表位可以极大地影响适应性免疫系统的功能,因为它们是抗体、B细胞受体(BCR)和T细胞受体(TCR)识别的特定序列。免疫系统的许多效应机制可由抗体表位或T细胞

表位触发,因此抗原文库具有潜在的应用,包括但不限于在传染病、癌症免疫治疗和自身免疫中的应用。抗体和T细胞表位可适应于广泛的用途,包括,例如,用于治疗或实验室用途的抗体或抗原结合片段的产生,使用抗体或T细胞表位的疫苗生产,以及已知抗原特异性T细胞的工程化改造,例如,用于癌症免疫治疗。

[0043] 在生产和使用不同的肽文库方面存在着许多挑战。可能的肽序列绝对数量之多可能给文库的生产带来挑战。对于长度为9个残基的肽(9-聚体),蛋白质中最常见的20种氨基酸有 $20^9$ (约 $5.1 \times 10^{11}$ )种可能序列组合。感兴趣的潜在抗原肽源(例如,病原体、癌细胞、组织)可表达数百或数千种蛋白质,每种蛋白质包含多个潜在肽抗原。

[0044] 已经开发了许多技术来产生肽,但是许多技术具有局限性(例如,缺乏肽多样性的充分覆盖,倾向于高亲和力相互作用,肽生产的使用条件导致蛋白质变性或错误折叠等)。本发明提供肽文库,包括例如具有本文所述的有用应用的高多样性肽文库。还提供了制备肽文库的方法,包括例如制备无偏肽文库的方法。

#### 肽文库的组合物

[0045] 本发明提供肽文库,包括例如在一系列治疗、诊断和研究应用中有用的高多样性肽文库。例如,所提供的肽文库可用于筛选疾病特异性、器官特异性肽、或其它隔室特异性肽,用于筛选具有治疗应用的肽,筛选具有诊断应用的肽,筛选肿瘤靶向肽,筛选抗体表位或抗原,筛选T细胞表位或抗原,筛选抗微生物肽,或其任何组合。

[0046] 本发明的肽文库可以是无偏的或有偏的,并且可以包含任意数量的肽。图1中提供了各种多样性的肽文库的示例性实施方式,并在下文中描述。

#### 无偏肽文库

[0047] 本文公开的肽文库可以是无偏文库,例如,缺少一个或多个选择标准。在一些实施方式中,本发明的文库包含给定大小的肽的所有可能氨基酸组合,包含标准的20个氨基酸中的任何一个的 $20^k$ 种可能的k聚肽,例如包含标准20个氨基酸中的任何一个的 $20^9$ 种可能的9聚肽。

[0048] 在一些实施方式中,该文库包含具有一定长度范围的肽,例如,从约2个氨基酸(2聚体)到约20个氨基酸(20聚体)或任何合适范围。在一些实施方式中,该文库包含具有基本相同长度的肽,例如,2聚体、3聚体、4聚体、5聚体、6聚体、7聚体、8聚体、9聚体、10聚体、11聚体、12聚体、13聚体、14聚体、15聚体、16聚体、17聚体、18聚体、19聚体、20聚体或更长。

[0049] 在一些实施方式中,本发明的文库包括特定位置处的受约束残基,例如,位置2和9处用于对接HLA-A2的受约束残基。在一些实施例中,本发明的文库包括非约束残基处的所有可能氨基酸组合,例如,9-聚体序列的位置3、4、5、6、7和8处的 $20^6$ 个可能的6-聚序列,以及位置1、2和9处的受约束残基。在一些实施方式中,本发明的文库包含给定大小的肽中除受约束残基之外的所有可能氨基酸组合,例如,当位置2和9受到约束时, $20^7$ 个可能的9-聚体序列,其中位置1、3、4、5、6、7和8发生变化。

[0050] 受约束的残基可以约束为单个残基或任何残基子集(例如,约束为缬氨酸、亮氨酸或异亮氨酸)。在一些实施方式中,受约束残基可以是氨基酸亚类中的任何一个,例如,任何疏水性氨基酸、任何亲水性氨基酸、任何带电氨基酸、任何碱性氨基酸、任何酸性氨基酸、任何环状氨基酸、任何芳族氨基酸、任何脂肪族氨基酸、任何极性氨基酸、任何非极性氨基酸,或其任何组合。可以从所有可能的残基中,也可以从残基的任何子集中选择改变的残基。例

如,任何疏水性氨基酸、任何亲水性氨基酸、任何带电氨基酸、任何碱性氨基酸、任何酸性氨基酸、任何环状氨基酸、任何芳族氨基酸、任何脂肪族氨基酸、任何极性氨基酸、任何非极性氨基酸或其任何组合。

[0051] 在一些实施方式中,本发明的文库包含可通过任何计算机生产方法制备的所有k-聚肽。在一些实施方式中,文库可包含来自任何翻译产物的k-聚肽,例如表位、抗原、蛋白质或蛋白质组。在一些实施方式中,该文库包含从一个或多个基因组、外显子组、转录物组、蛋白质组、ORF组或其任何组合计算机衍生的k-聚肽。

[0052] 在一些实施方式中,文库包含通过转录和翻译任何感兴趣的多核苷酸序列而产生的所有k-聚肽,例如,在所有六个阅读框中计算机产生基因组或宏基因组正向和反向链的转录和翻译产物。在一些实施方式中,本发明的文库包含可从哺乳动物基因组(例如,小鼠基因组、人类基因组、患者基因组、自身免疫患者基因组或癌症基因组)计算机转录和翻译衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从微生物基因组(例如,细菌基因组、病毒基因组、原生动物基因组、原生生物基因组、酵母基因组、古细菌基因组或噬菌体基因组)计算机转录和翻译衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从病原体基因组,例如,细菌病原体基因组、病毒病原体基因组、真菌病原体基因组、机会致病病原体基因组、条件病原体基因组或真核寄生虫基因组计算机转录和翻译衍生的所有k-聚肽。在一些实施方式中,本发明的文库可来自植物基因组或真菌基因组。在一些实施方式中,本发明的文库包含衍生自基因组的计算机转录和翻译的k-聚肽,其中基因组在计算机转录和翻译期间被修饰,例如,计算机突变以产生包含突变(例如,替换、插入、删除)的k-聚肽。

[0053] 在一些实施方式中,本发明的文库包含可从感兴趣的外显子组(例如,哺乳动物外显子组、人类外显子组、小鼠外显子组、患者外显子组、自身免疫患者外显子组、癌症外显子组、病毒外显子组、原生动物外显子组、原生生物外显子组、酵母外显子组、病原体外显子组、真核寄生虫外显子组、植物外显子组或真菌外显子组)的计算机翻译中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含衍生自外显子组的计算机翻译的k-聚肽,其中外显子组在计算机翻译期间被修饰,例如,计算机突变以产生包含突变(例如,替换、插入、删除)的k-聚肽。

[0054] 在一些实施方式中,本发明的文库包含可从感兴趣的转录组(例如,哺乳动物转录组、人类转录组、小鼠转录组、患者转录组、自身免疫患者转录组、癌症转录组、微生物转录组、细菌转录组、病毒转录组、原生动物转录组、原生生物转录组、酵母转录组、古细菌转录组、噬菌体转录组、病原体转录组、真核寄生虫转录组、植物转录组、真菌转录组,来自RNA测序的转录组、微生物组转录组或来自宏基因组RNA测序的转录组)的计算机翻译中衍生的所有k-聚肽。在一些实施例中,本发明的文库包含衍生自转录组的计算机翻译的k-聚肽,其中转录组在计算机翻译期间被修饰,例如,计算机突变以产生包含突变(例如,替换、插入、删除)的k-聚肽。

[0055] 在一些实施方式中,本发明的文库包含可从感兴趣的蛋白质组(例如,哺乳动物蛋白质组、人类蛋白质组、小鼠蛋白质组、患者蛋白质组、自身免疫患者蛋白质组、癌症蛋白质组、微生物蛋白质组、细菌蛋白质组、病毒蛋白质组、原生动物蛋白质组、原生生物蛋白质组、酵母蛋白质组、古细菌蛋白质组、噬菌体蛋白质组、病原体蛋白质组、真核寄生虫蛋白质

组、植物蛋白质组或真菌蛋白质组) 衍生的所有k-聚肽。在一些实施方式中, 本发明的文库包含源自蛋白质组的k-聚肽, 其中k-聚肽从蛋白质组序列修饰而来, 例如, 包含突变(例如, 替换、插入、删除)的k-聚肽。

[0056] 在一些实施方式中, 本发明的文库包含可从感兴趣的ORF组(例如, 哺乳动物ORF组、人类ORF组、小鼠ORF组、患者ORF组、自身免疫患者ORF组、癌症ORF组、微生物ORF组、细菌ORF组、病毒ORF组、原生动物ORF组、原生生物ORF组、酵母ORF组、古细菌ORF组、噬菌体ORF组、病原体ORF组、真核寄生虫ORF组、植物ORF组、真菌ORF组, 来自RNA测序的ORF组、微生物组ORF组或来自宏基因组RNA测序的ORF组)的计算机翻译中衍生的所有k-聚肽。在一些实施方式中, 本发明的文库包含衍生自ORF组的计算机翻译的k-聚肽, 其中ORF组在计算机翻译期间被修饰, 例如, 计算机突变以产生包含突变(例如, 替换、插入、删除)的k-聚肽。

[0057] 在一些实施方式中, 本发明的文库包含可从一组基因组、蛋白质组、转录组、ORF组或其任何组合的计算机转录和翻译或翻译中衍生的所有k-聚肽。在一些实施例中, 本发明的文库包含可从一组样品(例如, 来自患者群体的临床样品或一组病原体基因组)的多核苷酸序列的计算机转录和翻译或翻译中衍生的所有k-聚肽。在一些实施方式中, 本发明的文库包含可从一组病毒基因组, 例如人病毒组的计算机转录和翻译中衍生的所有k-聚肽。在一些实施方式中, 本发明的文库包含衍生自一组基因组、蛋白质组、转录组、ORF组或其任意组合的计算机转录和翻译的所有k-聚肽, 其中来源序列在计算机翻译期间被修饰, 例如, 计算机突变以产生包含突变(例如, 替换、插入、删除)的k-聚肽。

[0058] 在一些实施方式中, 本发明的文库包含可从差异基因组、蛋白质组、转录组、ORF组或其任何组合衍生的所有k-聚肽, 其中比较两个或更多基因组、蛋白质组、转录组、ORF组或其组合以鉴定为差异序列(在其之间区分)的序列(例如, 核苷酸序列、氨基酸序列、核苷酸丰度或蛋白质丰度不同)。在一些实施方式中, 通过比较感兴趣的组织来生成基因组、蛋白质组、转录组或ORF组的差异序列。在一些实施方式中, 通过比较来自感兴趣细胞(例如, 健康细胞与癌细胞)的序列来产生基因组、蛋白质组、转录组或ORF组的差异序列。在一些实施方式中, 通过比较感兴趣生物体的序列来生成基因组、蛋白质组、转录组或ORF组的差异序列。在一些实施方式中, 通过比较感兴趣的对象(例如患病对象对健康对象)来生成基因组、蛋白质组、转录组或ORF组的差异序列。在一些实施方式中, 差异序列相差至少为100%、99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、89%、88%、87%、86%、85%、84%、83%、82%、81%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、15%、10%、5%或1%。在一些实施方式中, 差异序列相差1%至100%、5%至100%、10%至100%、15%至100%、20%至100%、25%至100%、30%至100%、40%至100%、50%至100%、60%至100%、70%至100%、80%至100%、90%至100%、95%至100%、30%至90%、40%至90%、50%至90%、60%至90%、80%至90%或60%至80%。在一些实施方式中, 差异序列在比较的序列之间具有至少为100%、99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、89%、88%、87%、86%、85%、84%、83%、82%、81%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、15%、10%、5%或1%的差异。在一些实施方式中, 在比较的序列之间, 差异序列具有1%至100%、5%至100%、10%至100%、15%至100%、20%至100%、25%至100%、30%至100%、40%至100%、50%至100%、60%至100%、70%至100%、80%至100%、90%至100%、95%至100%、30%至90%、

40%至90%、50%至90%、60%至90%、80%至90%或60%至80%的差异。

[0059] 在一些实施方式中,本发明的文库包含可从基因组、蛋白质组、转录组、ORF组或其任何组合的同源序列衍生的所有k-聚肽,其中比较两个或更多基因组、蛋白质组、转录组、ORF组,或其组合以鉴定同源序列(例如,具有一定程度的同源性)的序列,例如同源核苷酸序列、同源氨基酸序列、同源核苷酸丰度或同源蛋白质丰度。在一些实施方式中,通过比较感兴趣的组织来生成基因组、蛋白质组、转录组或ORF组的同源序列。在一些实施方式中,基因组、蛋白质组、转录组或ORF组的同源序列是通过比较来自感兴趣的细胞(例如,健康细胞与涉及自身免疫的细胞(例如,诱导自身免疫的细胞或在自身免疫期间被靶向的细胞))的序列来生成的。在一些实施方式中,通过比较感兴趣的生物体的序列来生成基因组、蛋白质组、转录组或ORF组的同源序列。在一些实施方式中,通过比较感兴趣的对象(例如患病对健康对象)来生成基因组、蛋白质组、转录组或ORF组的同源序列。在一些实施方式中,同源序列具有至少为100%、99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、89%、88%、87%、86%、85%、84%、83%、82%、81%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、15%、10%、5%或1%的同源性。在一些实施方式中,同源序列具有1%至100%、5%至100%、10%至100%、15%至100%、20%至100%、25%至100%、30%至100%、40%至100%、50%至100%、60%至100%、70%至100%、80%至100%、90%至100%、95%至100%、30%至90%、40%至90%、50%至90%、60%至90%、80%至90%或60%至80%的同源性。

[0060] 在一些实施方式中,本发明的文库包含编码与感兴趣序列(例如,如上鉴定的差异序列或同源序列)具有一定程度同源性的所有k-聚肽。在一些实施方式中,本发明的文库包含两个或多个感兴趣序列之间最接近的同源物。

[0061] 在一些实施方式中,本发明的文库包含可从感兴趣的多肽序列衍生的所有k-聚肽,例如,覆盖病毒蛋白质的完整蛋白质序列的所有可能的9-聚肽。在一些实施方式中,本发明的文库包含可从感兴趣的多肽序列产生的k-聚肽,其中感兴趣的多肽序列被修饰,例如在计算机中突变以生成包含突变(例如,取代、插入、删除)的k-聚肽。

[0062] 在一些实施方式中,本发明的文库包含可从感兴趣序列中的突变衍生的所有k-聚肽,例如,可从编码抗原或表位的多核苷酸序列中的单核苷酸突变生成的所有9-聚肽。例如,本发明的文库包含可由编码抗原或表位的多核苷酸序列中的两个、三个、四个、五个、六个、七个、八个或九个核苷酸突变产生的所有9-聚肽。在一些实施方式中,本发明的文库包含可从丙氨酸取代衍生的所有k-聚肽,例如,本文所述序列中任何位置的丙氨酸取代(例如,蛋白质、一组蛋白质、蛋白质组、计算机转录和翻译的基因组)。在一些实施方式中,本发明的文库包括位置扫描文库,其中所选氨基酸残基依次被所有其他天然氨基酸取代。在一些实施方式中,本发明的文库包括组合位置扫描文库,其中所选氨基酸残基每次在两个或多个位置依次被所有其他天然氨基酸取代。在一些实施方式中,本发明的文库包含重叠肽库,其包含来自模板序列(例如,在计算机中翻译的基因组)的重叠肽,其中设定长度的重叠肽被确定数量的残基抵消。在一些实施方式中,本发明的文库包含T细胞截短肽文库,其中文库的每个复制物包含在一末端截短的等摩尔肽混合物(例如,可从标称11-聚体的C-末端截短衍生8-聚体、9-聚体、10-聚体和11-聚体)。在一些实施方式中,本发明的文库包含定制肽组,其中定制肽组以列表形式提供。

[0063] 在一些实施方式中,本发明的基因组、外显子组、转录组、蛋白质组或ORF组是病毒基因组、外显子组、转录组、蛋白质组或ORF组。病毒的非限制性实例包括腺病毒、腺相关病毒、Aichi病毒、澳大利亚蝙蝠丽沙病毒、BK多瘤病毒、版纳(Banna)病毒、巴马森林病毒、布尼亚韦拉病毒、拉克罗斯本扬病毒(Bunyavirus La Crosse)、雪鞋兔布尼亚病毒(Bunyavirus snowshoe hare)、猕猴疱疹病毒(Cercopithecine herpesvirus)、金迪普拉病毒(Chandipura virus)、基孔肯亚病毒(Chikungunya virus)、柯萨病毒(Cosavirus)A、牛痘病毒、柯萨奇病毒(Coxsackievirus)、克里米亚-刚果出血热病毒、巨细胞病毒(CMV)、登革热病毒、多理病毒(Dhori virus)、杜比病毒(Dugbe Virus)、杜文黑基病毒(Duvenhage Virus)、东部马脑炎病毒、埃博拉病毒、埃博拉病毒、埃可病毒(Echovirus)、脑心肌炎病毒、爱泼斯坦-巴尔病毒(EBV)、欧洲蝙蝠丽沙病毒、GB病毒C/庚型肝炎病毒、汉他病毒(Hantaan virus)、亨德拉病毒(Hendra virus)、甲型肝炎病毒、乙型肝炎病毒、丙型肝炎病毒、戊型肝炎病毒、丁型肝炎病毒、马痘病毒、人腺病毒、人星状病毒、人冠状病毒、人巨细胞病毒、人内源性反转录病毒(HERV)、人肠病毒、人疱疹病毒(例如,HHV-1、HHV-2、HHV-6A、HHV-6B、HHV-7、HHV-8)、人免疫缺陷病毒(例如,HIV-1、HIV-2)、人乳头瘤病毒(如HPV-1、HPV-2、HPV-16、HPV-18)、人副流感病毒、人细小病毒B19、人呼吸道合胞病毒(RSV)、人鼻病毒、人SARS冠状病毒、人泡沫反转录病毒(spuma retrovirus)、人类嗜T淋巴细胞病毒(HTLV,如HTLV-1、HTLV-2、HTLV-3)、人 $\gamma$ 病毒、甲型流感病毒、乙型流感病毒、丙型流感病毒、伊斯法罕病毒(Isfahan virus)、JC多瘤病毒、日本脑炎病毒、胡宁沙状病毒(Juin arenavirus)、KI多瘤病毒、昆津病毒、拉各斯-蝙蝠病毒、维多利亚湖-马尔堡病毒、兰加特病毒(Langat virus)、拉沙病毒(Lassa Virus)、Lordsdale病毒、跳跃病病毒(Louping ill virus)、淋巴细胞性脉络丛脑膜炎病毒(Lymphocytic choriomeningitis virus)、马丘波病毒(Machupo virus)、马亚罗病毒(Mayaro virus)、MERS冠状病毒、麻疹病毒、门戈脑心肌炎病毒(Mengo encephalomyocarditis virus)、梅克尔细胞多瘤病毒(Merkel cell polyomavirus)、蒙古拉病毒(Mokola virus)、传染性软疣病毒(Molluscum contagiosum virus)、猴痘病毒、腮腺炎病毒、墨莱溪谷脑炎病毒(Murray valley encephalitis virus)、纽约病毒、尼帕病毒(Nipah virus)、诺如病毒(Norovirus)、诺沃克病毒、阿尼昂-尼昂病毒(O'nyong-nyong virus)、羊接触传染性化脓性口炎病毒(Orf virus)、奥罗普切病毒(Oropouche virus)、皮钦德病毒(Pichinde virus)、脊髓灰质炎病毒、蓬塔托罗白蛉病毒(Punta toro phlebovirus)、普马拉病毒(Puumala virus)、狂犬病病毒、裂谷热病毒、罗沙病毒A(Rosavirus A)、罗斯河病毒(Ross river virus)、轮状病毒(例如,轮状病毒A、轮状病毒B、轮状病毒C、轮状病毒X)、风疹病毒(Rubella virus)、鹭山病毒(Sagiyama virus)、赛利病毒(Salivirus)A、西西里白蛉热病毒(Sandfly fever sicilian virus)、札幌病毒、塞姆利基森林病毒(Semliki forest virus)、汉城病毒、猴泡沫病毒(Simian foamy virus)、猴病毒5、辛德比斯病毒(Sindbis virus)、南安普敦病毒、圣路易斯脑炎病毒、蜱传波瓦桑病毒(Tick-borne powassan virus)、细环病毒(Torque teno virus)、托斯卡纳病毒、乌库涅米病毒(Uukuniemi virus)、痘苗病毒、水痘-带状疱疹病毒(Varicella-zoster virus)、天花病毒(Variola virus)、委内瑞拉马脑炎病毒、水泡性口炎病毒(Vesicular stomatitis virus)、西部马脑炎病毒、WU多瘤病毒、西尼罗河病毒、亚巴猴肿瘤病毒(Yaba monkey tumor virus)、亚巴样病病毒、黄热病病毒、和寨卡病毒。

[0064] 在一些实施方式中,本发明的基因组、外显子组、转录组、蛋白质组或ORF组是癌基因组、外显子组、转录组、蛋白质组或ORF组。在一些实施方式中,本发明的文库包含已知的癌症新表位。在一些实施方式中,本发明的文库包含可从已知癌症抗原蛋白中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从参与上皮间质转化的基因中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从癌症相关基因中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从突变性癌症驱动基因中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从原癌基因、癌基因或肿瘤抑制基因中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从原癌基因、癌基因或肿瘤抑制基因衍生的所有k-聚肽,其中k-聚肽包含如本文所述的突变(例如,氨基酸取代、丙氨酸取代、位置扫描、组合位置扫描等)。

[0065] 癌症的非限制性示例包括急性淋巴细胞白血病(ALL)、急性髓细胞白血病(AML)、肾上腺皮质癌、艾滋病相关癌、艾滋病相关淋巴瘤、肛门癌、阑尾癌、星状细胞瘤、非典型畸胎瘤样/横纹肌样瘤、基底细胞癌、胆管癌、膀胱癌、骨癌、大脑瘤、乳腺癌、支气管肿瘤、伯基特氏淋巴瘤、类癌肿瘤、原发部位不明转移癌(Carcinoma of Unknown Primary)、心脏肿瘤、中枢神经系统癌、宫颈癌、胆管癌(Cholangiocarcinoma)、脊索瘤、慢性淋巴细胞白血病(CLL)、慢性粒细胞白血病(CML)、慢性骨髓增生性肿瘤、结直肠癌、颅咽管瘤(Craniopharyngioma)、皮肤T细胞淋巴瘤、原位乳腺导管癌(Ductal Carcinoma In Situ)、胚芽肿瘤(Embryonal Tumor)、子宫内膜癌、上皮癌、室管膜瘤、食管癌、成感觉神经细胞瘤(Esthesioneuroblastoma)、尤因氏肉瘤、颅外生殖细胞瘤、性腺外生殖细胞瘤、眼癌、输卵管癌、骨纤维组织细胞瘤、胆囊癌、胃癌、胃肠道类癌肿瘤、胃肠道间质瘤(GIST)、生殖细胞瘤、妊娠滋养细胞疾病(Gestational Trophoblastic Disease)、毛细胞白血病、头颈癌、肝细胞癌、组织细胞增多病、霍奇金氏淋巴瘤、下咽癌、眼内黑色素瘤、胰岛细胞瘤、卡波济氏肉瘤、肾(肾细胞)癌、朗格汉斯细胞组织细胞增生症、喉癌、白血病、唇和口腔癌、肝癌、肺癌(非小细胞和小细胞)、淋巴瘤、男性乳腺癌、骨的恶性纤维组织细胞瘤和骨肉瘤、黑色素瘤、默克尔细胞癌、间皮瘤、转移癌、原发性隐匿的转移性颈鳞状细胞癌、中线束癌、口癌、多发性内分泌腺瘤综合征、多发性骨髓瘤、蕈样肉芽肿、骨髓增生异常综合征、骨髓增生异常/骨髓增生性肿瘤、鼻腔癌、鼻咽癌、神经母细胞瘤、非霍奇金淋巴瘤、非小细胞肺癌、口腔癌、唇癌和口腔癌、口咽癌、骨肉瘤、卵巢癌、胰腺癌、胰腺神经内分泌肿瘤、乳头瘤、副神经节瘤、副鼻窦癌、甲状旁腺癌、阴茎癌、咽癌、嗜铬细胞瘤、垂体瘤、浆细胞瘤、胸膜肺母细胞瘤、原发性中枢神经系统(CNS)淋巴瘤、原发性腹膜癌、前列腺癌、直肠癌、复发癌、视网膜母细胞瘤、横纹肌肉瘤、唾液腺癌、肉瘤、Sézary综合征、皮肤癌、小细胞肺癌、小肠癌、软组织肉瘤、皮肤鳞状细胞癌、隐匿原发性颈部鳞状细胞癌、胃癌、T细胞淋巴瘤、睾丸癌、喉癌、胸腺瘤和胸腺癌、甲状腺癌、移行细胞癌、输尿管和肾盂癌、尿道癌、子宫癌、子宫肉瘤、阴道癌、血管瘤、外阴癌和肾母细胞瘤。

[0066] 在一些实施方式中,本发明的基因组、外显子组、转录组、蛋白质组或ORF组是炎症或自身免疫病基因组、外显子组、转录组、蛋白质组或ORF组。在一些实施方式中,本发明的文库包含已知的炎症或自身免疫性新表位或自身表位。在一些实施方式中,本发明的文库包含可从已知炎症或自身免疫性抗原蛋白中衍生的所有k-聚肽。在一些实施方式中,本发明的文库包含可从炎症或自身免疫相关基因中衍生的所有k-聚肽。在一些实施方式中,本

发明的文库包含可从炎症或自身免疫相关驱动基因的突变中衍生的所有k-聚肽。

[0067] 炎症或自身免疫病或病症的非限制性例子包括急性播散性脑脊髓炎 (ADEM); 急性坏死性出血性白质脑炎; 爱迪生病; 佐剂诱导的关节炎; 丙种球蛋白缺乏血症; 斑秃; 淀粉样变性; 强直性脊柱炎; 抗GBM/抗TBM肾炎; 抗磷脂综合征; 自身免疫性血管水肿; 自身免疫性再生障碍性贫血; 自身免疫性自主神经障碍; 自身免疫性胃萎缩; 自身免疫性溶血性贫血; 自身免疫性肝炎; 自身免疫性高脂血症; 自身免疫性免疫缺陷; 自身免疫性内耳病; 自身免疫性心肌炎; 自身免疫性卵巢炎; 自身免疫性胰腺炎; 自身免疫性视网膜病变; 自身免疫性血小板减少性紫癜; 自身免疫性甲状腺疾病; 自身免疫性荨麻疹; 轴突神经病变; 巴洛病; 白塞氏病; 大疱性类天疱疮; 心肌病; 卡斯特尔曼代病 (Castleman disease); 乳糜泻; 恰加斯病; 慢性炎性脱髓鞘性多发性神经病 (CIDP); 慢性复发性多灶性骨髓炎 (CRMO); 嗜酸性肉芽肿性多血管炎; 瘢痕性类天疱疮/良性粘膜类天疱疮; 克罗恩病; 耳蜗前庭综合征; 胶原诱导性关节炎; 冷凝集素病; 先天性心脏传导阻滞; 柯萨基病毒性心肌炎; CREST病; 原发性混合性冷球蛋白血症; 脱髓鞘性神经病; 疱疹性皮炎; 皮炎; 德维克病 (视神经脊髓炎); 盘状狼疮; 德莱斯勒综合征; 子宫内膜异位症; 嗜酸性食管炎; 嗜酸性筋膜炎; 结节性红斑; 实验性变态反应性脑脊髓炎; 实验性自身免疫性脑脊髓炎; 伊文斯综合征; 纤维肌痛; 纤维化性肺泡炎; 巨细胞动脉炎 (颞动脉炎); 巨细胞心肌炎; 肾小球肾炎; 肺出血肾炎综合征; 肉芽肿伴多血管炎 (GPA) (以前称为韦格纳肉芽肿); 格雷夫斯病; 格林-巴利综合征; 桥本脑炎; 桥本甲状腺炎; 溶血性贫血; 过敏性紫癜; 妊娠疱疹; 低丙种球蛋白血症; 特发性血小板减少性紫癜; IgA肾病; IgG4相关硬化性疾病; 免疫调节脂蛋白; 包涵体肌炎; 间质性膀胱炎; 炎症性肠病; 幼年型关节炎; 幼年少发性关节炎; 青少年糖尿病 (1型糖尿病); 青少年肌炎; 川崎综合征; 兰伯特-伊顿综合征; 白细胞分裂性脉管炎; 扁平苔癣; 硬化性苔癣; 木质性结膜炎; 线状IgA病 (LAD); 狼疮; 慢性莱姆病; 梅尼埃病; 显微镜下多血管炎; 混合性结缔组织病; 摩尔溃疡; 穆-哈二氏病; 多发性硬化; 重症肌无力; 肌炎; 嗜睡症; 视神经脊髓炎; 中性粒细胞减少; 非肥胖型糖尿病; 眼部瘢痕性类天疱疮; 视神经炎; 复发性风湿病; PANDAS (与链球菌相关的儿童自身免疫性神经精神疾病); 副肿瘤性小脑变性; 阵发性夜间血红蛋白尿 (PNH); 帕里-隆伯格综合征; 帕森纳格-特纳综合征; 扁桃体炎 (周围性葡萄膜炎); 天疱疮; 寻常性天疱疮; 周围神经病变; 静脉周围脑脊髓炎; 恶性贫血; POEMS综合征; 结节性多动脉炎; I型、II型和III型自身免疫性多腺体综合征; 风湿性多肌痛; 多发性肌炎; 心肌梗死后综合征; 心包切开术后综合征; 孕酮皮炎; 原发性胆汁性肝硬化; 原发性硬化性胆管炎; 银屑病; 斑块型银屑病; 银屑病性关节炎; 特发性肺纤维化; 坏疽性脓皮病; 纯红细胞再生障碍性贫血; 雷诺现象; 反应性关节炎; 反射性交感神经营养不良; 雷特综合征; 复发性多软骨炎; 不宁腿综合征; 腹膜后纤维化; 风湿热; 类风湿性关节炎; 类肉瘤病; 施密特综合征; 巩膜炎; 硬皮病; 硬化性胆道炎; 硬化性涎腺炎; 干燥综合征; 精子与睾丸自身免疫; 僵硬人综合征; 亚急性细菌性心内膜炎; 苏萨克综合征; 交感性眼炎; 系统性红斑狼疮; 系统性硬化; 高安动脉炎; 颞动脉炎/巨细胞动脉炎; 血小板减少性紫癜 (TTP); 疼痛性眼肌麻痹; 横贯性脊髓炎; 1型糖尿病; 溃疡性结肠炎; 未分化结缔组织病 (UCTD); 葡萄膜炎; 血管炎; 水泡性大疱性皮肤病; 白癜风; 韦格纳肉芽肿 (现称肉芽肿伴多血管炎 (GPA))。炎症或自身免疫病或病况的非限制性例子包括感染, 例如慢性感染、潜伏感染、缓慢感染、持续病毒感染、细菌感染、真菌感染、支原体感染或寄生虫感染。

[0068] 在一些实施方式中,本发明的文库可包含具有翻译后修饰的肽,包括例如乙酰化、酰胺化、生物素化、脱酰胺化、法尼基化、甲酰化、香叶基香叶基化、谷胱甘肽基化、糖化、糖基化、羟基化、甲基化、单ADP核糖基化、肉豆蔻酰化、N-乙酰化、N-糖基化、N-肉豆蔻酰化、亚硝基化、氧化、棕榈酰化、磷酸化、聚(ADP-核糖基)化、硬脂酰化、硫酸化、SUMO化、泛素化或其任何组合。在一些实施方式中,本发明的肽可包含一个或多个硒代半胱氨酸残基。

[0069] 在一些实施方式中,本发明文库中的肽是蛋白质mRNA复合物的一部分。在一些实施方式中,本发明文库中的肽是包含嘌呤霉素连接的蛋白质mRNA复合物的一部分。在一些实施方式中,本发明文库中的肽是蛋白质-mRNA-cDNA复合物的一部分。在一些实施方式中,本发明文库中的肽是蛋白质DNA复合物的一部分。在一些实施方式中,本发明文库中的肽是包含生物素-链霉亲和素连接的蛋白质DNA复合物的一部分。在一些实施方式中,本发明文库中的肽是蛋白质cDNA复合物的一部分。在一些实施方式中,本发明文库中的肽是蛋白质-核糖体-mRNA复合物的一部分。在一些实施方式中,本发明文库中的肽是蛋白质-核糖体-mRNA复合物的一部分,其中mRNA包含缺少终止密码子的间隔子序列。在一些实施方式中,本发明文库中的肽是蛋白质-核糖体-mRNA-cDNA (PRMC) 复合物的一部分。

[0070] 在一些实施方式中,本发明文库中的肽通过亲和和标签纯化(例如,使用FLAG标签)来纯化。在一些实施方式中,本发明文库中的肽包含HaloTag酶序列。在一些实施方式中,本发明文库中的肽包含亲和素或链霉亲和素。

[0071] 在一些实施方式中,本发明文库中的肽可结合或融合到另一分子。在一些实施方式中,本发明文库中的肽可结合或融合到另一多肽。在一些实施方式中,本发明文库中的肽可结合或融合到一多核酸。在一些实施方式中,本发明文库中的肽可结合或融合到一DNA。在一些实施方式中,本发明文库中的肽可结合或融合到一RNA。在一些实施方式中,本发明文库中的肽可存在于较大的支架内,例如作为较大的蛋白质序列或蛋白质复合物的一部分。

[0072] 本发明的文库可包括约100,500,1000,2000,5000, $10^4$ , $10^5$ , $10^6$ , $10^7$ , $10^8$ , $10^9$ , $10^{10}$ , $10^{11}$ , $10^{12}$ , $10^{13}$ , $10^{14}$ , $10^{15}$ , $10^{16}$ , $10^{17}$ , $10^{18}$ , $10^{19}$ , $10^{20}$ , $20^2$ , $20^3$ , $20^4$ , $20^5$ , $20^6$ , $20^7$ , $20^8$ , $20^9$ , $20^{10}$ , $20^{11}$ , $20^{12}$ , $20^{13}$ , $20^{14}$ , $20^{15}$ , $20^{16}$ , $20^{17}$ , $20^{18}$ , $20^{19}$ , $20^{20}$ , $20^{21}$ , $20^{22}$ , $20^{23}$ , $20^{24}$ , $20^{25}$ , $20^{26}$ , $20^{27}$ , $20^{28}$ , $20^{29}$ ,或 $20^{30}$ 个肽或抗原。

[0073] 本发明的文库可包括超过至少约100,500,1000,2000,5000, $10^4$ , $10^5$ , $10^6$ , $10^7$ , $10^8$ , $10^9$ , $10^{10}$ , $10^{11}$ , $10^{12}$ , $10^{13}$ , $10^{14}$ , $10^{15}$ , $10^{16}$ , $10^{17}$ , $10^{18}$ , $10^{19}$ , $10^{20}$ , $20^2$ , $20^3$ , $20^4$ , $20^5$ , $20^6$ , $20^7$ , $20^8$ , $20^9$ , $20^{10}$ , $20^{11}$ , $20^{12}$ , $20^{13}$ , $20^{14}$ , $20^{15}$ , $20^{16}$ , $20^{17}$ , $20^{18}$ , $20^{19}$ , $20^{20}$ , $20^{21}$ , $20^{22}$ , $20^{23}$ , $20^{24}$ , $20^{25}$ , $20^{26}$ , $20^{27}$ , $20^{28}$ , $20^{29}$ ,或 $20^{30}$ 个肽或抗原。

[0074] 本发明的文库可包括至多约100,500,1000,2000,5000, $10^4$ , $10^5$ , $10^6$ , $10^7$ , $10^8$ , $10^9$ , $10^{10}$ , $10^{11}$ , $10^{12}$ , $10^{13}$ , $10^{14}$ , $10^{15}$ , $10^{16}$ , $10^{17}$ , $10^{18}$ , $10^{19}$ , $10^{20}$ , $20^2$ , $20^3$ , $20^4$ , $20^5$ , $20^6$ , $20^7$ , $20^8$ , $20^9$ , $20^{10}$ , $20^{11}$ , $20^{12}$ , $20^{13}$ , $20^{14}$ , $20^{15}$ , $20^{16}$ , $20^{17}$ , $20^{18}$ , $20^{19}$ , $20^{20}$ , $20^{21}$ , $20^{22}$ , $20^{23}$ , $20^{24}$ , $20^{25}$ , $20^{26}$ , $20^{27}$ , $20^{28}$ , $20^{29}$ ,或 $20^{30}$ 个肽或抗原。

[0075] 本发明的k聚体可以是1聚体、2聚体、3聚体、4聚体、5聚体、6聚体、7聚体、8聚体、9聚体、10聚体、11聚体、12聚体、13聚体、14聚体、15聚体、16聚体、17聚体、18聚体、19聚体、20聚体、21聚体、22聚体、23聚体、25聚体、26聚体、27聚体、28聚体、29聚体、30聚体、31聚体、32聚体、33聚体、34聚体、35聚体、36聚体、37聚体、38聚体、39聚体、40聚体、41聚体、42聚体、43

聚体、44聚体、45聚体、46聚体、47聚体、48聚体、49聚体或50聚体。

[0076] 本发明的文库可包括每个肽具有约0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,110,120,130,140,150,160,170,180,190,200,225,250,275,300,350,400,450,500,600,700,800,900,或1000个受约束残基的肽。本发明的文库可包括每个肽具有约1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,110,120,130,140,150,160,170,180,190,200,225,250,275,300,350,400,450,500,600,700,800,900,或1000个可变残基的肽。

[0077] 本发明的文库可包括每个肽具有超过至少约0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,110,120,130,140,150,160,170,180,190,200,225,250,275,300,350,400,450,500,600,700,800,900,或1000个受约束残基的肽。本发明的文库可包括每个肽具有多于至少约1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,110,120,130,140,150,160,170,180,190,200,225,250,275,300,350,400,450,500,600,700,800,900,或1000个可变残基的肽。

[0078] 本发明的文库可包括每个肽具有至多约0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,110,120,130,140,150,160,170,180,190,200,225,250,275,300,350,400,450,500,600,700,800,900,或1000个受约束残基的肽。本发明的文库可包括每个肽具有至多约1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,110,120,130,140,150,160,170,180,190,200,225,250,275,300,350,400,450,500,600,700,800,900,或1000个可变残基的肽。

#### 肽文库亚组

[0079] 肽文库可以是无偏文库的亚组。在一些实施方式中,算法可用于选择本发明肽文库中的肽。例如,可以使用算法来预测最有可能折叠或停靠在MHC/HLA结合口袋中的肽,并且可以选择高于特定阈值的肽包含在文库中。在一些实施方式中,为本发明文库选择肽包括基于对特定HLA类型的预测结合亲和力对肽进行优先级排序。在一些实施方式中,为本发明文库选择肽基于群体(例如,人类人群)中的流行程度优先考虑HLA类型或等位基因。

[0080] 在一些实施方式中,本发明的文库包含基于筛选分析(例如用于折叠的功能分析)选择的肽。分析可用于测试本发明肽的成功折叠。例如,单克隆抗体可用于测定多肽的成功折叠,例如,BB7.2单克隆抗体的结合可指示HLA-A2的成功折叠。在一些实施方式中,将正确折叠的肽与错误折叠的肽分离,可以确定正确折叠和错误折叠的肽序列(例如,通过对本文公开的标识符测序),并且可以为正确折叠的肽富集后续文库(图6)。

[0081] 在一些实施方式中,本发明的文库包含来自癌症(例如,结直肠癌或非小细胞肺癌)的已知的、检测得到的或预测的新表位。在一些实施方式中,本发明的文库包含与特定疾病或状况相关或富集的序列。

### 抗原文库

[0082] 本说明书提供了肽文库,其可用于一系列筛选试验,以鉴定潜在的诊断或治疗靶点或试剂。肽文库可以是抗原文库。本发明包括抗原文库,其应用于与免疫抗原相关的分析,包括包含T细胞表位的T细胞抗原。T细胞抗原和表位可以极大地影响适应性免疫系统的功能,因为它们是T细胞受体(TCR)识别的特定序列。关联抗原的TCR识别对于例如针对病原体的免疫识别和效应物免疫应答、针对癌细胞的免疫识别和效应物免疫应答以及自身免疫应答(例如,针对导致疾病的自身组织的免疫应答)是重要的。

[0083] 因此,用于筛选T细胞表位的抗原文库在例如与传染病、癌症和自身免疫性疾病相关的研究、诊断、治疗和预防措施中具有潜在用途。鉴定T细胞识别的抗原肽序列对于例如疫苗开发(例如,鉴定特定病原体的保护性抗原或鉴定用于癌症疫苗的新抗原)、癌症治疗(例如,鉴定基于T细胞的治疗靶点包括鉴定新抗原),以及自身免疫研究和治疗(例如,鉴定自身免疫抗原)至关重要。

[0084] 然而,在T细胞抗原和表位相关的应用中,肽文库的生产和使用面临许多挑战。可能相关的肽序列绝对数量之多可能给文库生产带来挑战。对于长度为9个残基的肽(9-聚体),蛋白质中最常见的20种氨基酸有 $20^9$ (约 $5.1 \times 10^{11}$ )种可能序列组合。如下所述,一些递呈给T细胞的肽可以超过9个残基,因此TCR识别的潜在肽的多样性可以超过 $5.1 \times 10^{11}$ 。另外,感兴趣的潜在抗原肽源(例如,病原体、癌细胞、组织)可表达数千种蛋白质,每种蛋白质包含多个潜在肽抗原。需要高效、高通量的方法来生成任何具有这种多样性的文库。

[0085] 其他挑战涉及肽必须递呈到TCR的方式,以启动足够亲和力的结合,从而在实验或治疗环境中有用。在体内,TCR识别主要组织相容性复合物(MHC)分子递呈的肽。为了结合TCR,肽文库的肽抗原也必须在MHC环境下递呈。此外,为了实现在实验或治疗环境中有用的足够亲和力的结合,需要肽MHC(pMHC)复合物的多聚体。传统的制备pMHC多聚体的方法通量低,劳动强度大,所得的多肽容易发生错误折叠和肽抗原负载不良。

[0086] 本文提供了各种pMHC多聚体文库及其生产方法,包括高通量方法。在一些实施方式中,pMHC多聚体进一步包含核酸标识符,允许如本文他处所述方便地检测和量化结合。

### pMHC多聚体

[0087] 肽通过两大类MHC递呈给TCR:MHC I类(MHC-I)和MHC II类(MHC-II)。在人类中,MHC也称为人白细胞抗原(HLA)。编码HLA的基因在不同个体之间高度可变,不同的HLA基因在肽结合和抗原呈递方面可能具有不同的特征。人类有三种主要的MHC I类基因,即HLA-A、HLA-B和HLA-C。由这些基因产生的蛋白质几乎存在于所有细胞的表面。另外,非经典I类基因包括HLA-E、HLA-F和HLA-G。人类有六种主要的MHC II类基因:HLA-DPA1,HLA-DPB1,HLA-DQA1,HLA-DQB1,HLA-DRA,和HLA-DRB1。许多MHC基因有多种等位基因形式,每种等位基因都可以在一个浅抗原结合沟中递呈多种肽,该浅抗原结合沟由位于 $\beta$ 片层顶部的两个反平行 $\alpha$ -螺旋产生。MHC-I递呈的肽可以以延伸构象结合,N端和C端都结合在封闭的沟内,从而限制其大小(例如,8-10个残基)。MHC-II递呈的肽也可以以延伸构象结合,但由于沟的开放性,可以更长(例如,14-20个残基)。

[0088] 为了产生pMHC,可以产生亚基的表达构建体。例如,可以产生表达MHC-I重链和 $\beta$ -2-微球蛋白( $\beta$ 2m)的构建体。可以基因删除重链的跨膜结构域以促进纯化,并且在C末端添加生物素识别位点。重链、 $\beta$ 2m和肽都可以表达,例如,在大肠杆菌培养物中重组表达为包涵

体,或在真核细胞(如昆虫或哺乳动物细胞)中表达。例如,可以使用BirA酶对生物素识别位点进行生物素化。然后可以用变性剂处理重链、 $\beta 2m$ 和肽,重新折叠成pMHC单体,并通过尺寸排除色谱法纯化。与肽不正确结合的MHC分子可能不稳定、解离和错折叠。类似的技术可用于生成肽MHC II复合物(pMHC II)。

[0089] 然后,MHC单体可以多聚化,形成例如二聚体、四聚体、五聚体、八聚体、链霉聚体或葡聚体。二聚体可以通过MHC分子胞外结构域的遗传融合产生,例如,作为与结合第二MHC的免疫球蛋白骨架的融合物。例如,可以通过向具有生物素化C-末端的MHC单体添加链霉亲和素或亲和素“骨架”来生成四聚体。或者,链霉亲和素结构域可以表达为与MHC链的C-末端的融合物,促进自组装成四聚体。MHC四聚体和八聚体也可通过在MHC链C端的游离半胱氨酸中引入点突变来生成,该游离半胱氨酸可被含生物素的碘乙酰胺或马来酰亚胺衍生物烷基化。链霉亲和素结合物可用于寡聚。通过使用含有一个生物素和两个马来酰亚胺部分(DMG)的分枝肽,该策略允许制备八聚体MHC复合物。MHC五聚体可以通过自组装螺旋结构域与五个MHC单体复合生成。MHC葡聚体可通过将多个MHC复合物(例如,十个或更多个)附接到葡聚糖聚合物骨架来生成。可通过将MHC链的生物素化C末端连接到Strep-tactin或Strep-tag骨架来生成MHC链霉聚物,从而形成包含8-12个MHC单体的复合物。

#### 单链pMHC

[0090] MHC分子可以被设计为单链肽-MHC多肽(sc-pMHC),其包含组装的pMHC的亚基。例如,这种sc-pMHC可以简化肽抗原在MHC结合沟中的装载。例如,与可占据结合沟的其他污染肽相比,这种sc-pMHC还可促进连接的肽抗原的有效装载。Sc-pMHC多聚体可表现出与抗原特异性T细胞的特异性和剂量依赖性的结合(图12)。

[0091] sc-pMHC可包含对应于MHC-I的抗原肽、重链和/或 $\beta 2$ -微球蛋白( $\beta 2m$ )。MHC-I sc-pMHC可包含抗原肽和重链。MHC-I sc-pMHC可包含抗原肽和 $\beta 2m$ 。MHC-I sc-pMHC可包含抗原肽、 $\beta 2$ -微球蛋白( $\beta 2m$ )和重链。在一些实施方式中,MHC-I sc-pMHC可包含具有抗原肽和将抗原肽连接到重链的柔性接头的单条多肽。在一些实施方式中,MHC-I sc-pMHC可进一步包含将 $\beta 2m$ 连接到抗原肽或重链的另一柔性接头。在一些实施方式中,MHC-I sc-pMHC可包含具有抗原肽和将抗原肽连接到 $\beta 2m$ 的柔性接头的单条多肽。在一些实施方式中,MHC-I sc-pMHC可进一步包含将重链连接到抗原肽或 $\beta 2m$ 的另一柔性接头。

[0092] sc-pMHC可包含对应于MHC-II的抗原肽、 $\alpha$ 链和/或 $\beta$ 链。MHC-II sc-pMHC可包含抗原肽和 $\alpha$ 链。MHC-II sc-pMHC可包含抗原肽和 $\beta$ 链。MHC-II sc-pMHC可包含抗原肽、 $\alpha$ 链和 $\beta$ 链。MHC-II sc-pMHC可包含具有抗原肽和将抗原肽连接到 $\beta$ 链的柔性接头的单条多肽。在一些实施方式中,MHC-II sc-pMHC可进一步包含将 $\alpha$ 链连接到抗原肽或 $\beta$ 链的另一柔性接头。在一些实施方式中,MHC-II sc-pMHC可包含具有抗原肽和将抗原肽连接到 $\alpha$ 链的柔性接头的单条多肽。在一些实施方式中,MHC-II sc-pMHC可进一步包含将 $\beta$ 链连接到抗原肽或 $\alpha$ 链的另一柔性接头。

[0093] sc-pMHC可以是任何顺序,例如,单条多肽C端的肽,以允许抗原肽N端的更大多样性。如果抗原肽位于N-端,则可采用其他机制在N-端产生更大的多样性,例如,酶切N-端。sc-pMHC单体可如上述MHC单体所述进行多聚,例如,形成二聚体、四聚体、五聚体、八聚体、链霉聚体或葡聚体。

[0094] 在一些实施方式中,本发明的文库包括多个sc PMC。在一些实施方式中,本发明的

文库包含多个sc-pMHC,其中k聚体序列定位于sc-pMHC的抗原肽部分内。

[0095] 肽-MHC多聚体或特异性结合pMHC多聚体的抗体可与荧光标签结合,从而允许鉴定结合肽MHC多聚体的T细胞,例如,通过流式细胞术或显微镜检。也可以基于荧光标签通过荧光活化细胞分选的方式选择T细胞。然而,这种方法的局限性与可用的不同荧光标签和检测物的数量有关,这限制了高通量抗原文库筛选中基于荧光的方法的使用。

[0096] 在一些实施方式中,本发明的文库包含如别处所述与核苷酸标识符偶联的sc-pMHC多聚体,能够方便地检测和量化与TCR的抗原特异性结合。例如,该文库可允许检测对特定抗原特异性的T细胞,多重检测给定样品中的T细胞特异性,将TCR序列与特异性匹配(例如,通过单细胞测序),比较性TCR亲和力测定,确定给定TCR的共有特异性序列,或针对感兴趣的序列绘制T细胞的抗原反应性。

#### 接头

[0097] 在一些实施方式中,本发明提供了多肽序列,其中两个多肽序列或结构域通过接头连接。在一些实施方式中,本发明提供了多肽序列,其中三个或更多多肽序列或结构域通过接头连接。在一些实施方式中,本发明提供了多肽序列,其中四个或更多多肽序列或结构域通过接头连接。在一些实施方式中,本发明提供了多肽序列,其中五个或更多多肽序列或结构域通过接头连接。在一些实施方式中,本发明提供了多肽序列,其中六个或更多多肽序列或结构域通过接头连接。

[0098] 接头可以是化学键,例如,一个或多个共价键或非共价键。在一些实施方式中,接头是共价键。在一些实施方式中,接头是非共价键。在一些实施方式中,接头是肽接头。这种接头可以介于2-30个氨基酸之间,或更长。在一些实施方式中,接头可用于,例如,将多肽序列或结构域彼此隔开。在一些实施方式中,接头可以位于结构域之间,例如,以提供二级和三级结构的分子柔性。接头可包括柔性、刚性和/或可切割接头,如本文所述。在一些实施方式中,接头包含至少一个甘氨酸、丙氨酸和丝氨酸氨基酸以提供柔性。在一些实施方式中,接头是疏水性接头,例如包括带负电荷的磷酸根基团、聚乙二醇(PEG)基团或焦磷酸二酯基团。在一些实施方式中,接头是可切割的,以选择性地使得多肽序列或结构域相对彼此释放,但足够稳定以防止过早切割。

[0099] 柔性肽接头可具有主要由Gly和Ser残基片段组成的序列(“GS”接头)。柔性肽接头可用于连接需要一定程度运动或相互作用的域,并且可以包括小的、非极性(例如Gly)或极性(例如Ser或Thr)氨基酸。Ser或Thr的纳入还可以通过与水分子形成氢键来促进接头在水溶液中的稳定性,从而减少接头和蛋白质部分之间不利的相互作用。柔性接头可以包括,例如,表1中列出的任何序列的单拷贝或重复。

SEQ ID NO:	序列
1	GGGGS
2	GGGGG
3	GGGGGGG
4	KESGSVSSEQLAQFRSLD
5	EGKSSGSGSESKST
6	GSAGSAAGSGEF

[0100] 刚性接头可用于在结构域之间保持固定距离。当结构域的空间分隔对于保持一个

或多个组件的稳定性或生物活性至关重要时,刚性接头也很有用。刚性接头可以具有 $\alpha$ -螺旋结构或富含Pro的序列。例如,刚性接头可包含序列(EAAAK)<sub>n</sub>-A(EAAAK)<sub>n</sub>A(XP)<sub>n</sub>,其中n表示任意数量的重复(例如,2-5),X表示任意氨基酸(例如,Ala、Lys或Glu)。

[0101] 可切割的肽接头可用于移除或释放多肽序列或结构域。在一些实施方式中,接头可以在特定情况下被切割,例如存在还原剂或酶的情况下。融合体中接头的体内切割还可以通过某些条件、特定细胞或组织中、或限制在某些细胞隔室中表达的酶或蛋白酶进行。许多酶或蛋白酶的特异性提供了在限制的隔室中较慢的接头切割。例如,可切割的接头可包含基质金属蛋白酶(MMP)或另一蛋白酶的切割序列。另一个例子包括两个Cys残基之间的凝血酶敏感序列(例如,PRS)。CPRSC的体外凝血酶处理导致凝血酶敏感序列的切割,而可逆的二硫键保持完整。这种接头是已知的并描述于,例如,Chen等.2013.融合蛋白接头:特性、设计和功能。(Fusion Protein Linkers:Property,Design and Functionality.)*Adv Drug Deliv Rev.*65(10):1357-1369。

[0102] 在一些实施方式中,接头可以是肽键,例如,肽序列或结构域的C端可以通过肽键与另一肽序列或结构域的N端融合。

[0103] 接头的另一个例子包括疏水接头,例如带负电荷的磺酸根基团;脂质,例如聚(--CH<sub>2</sub>--)烃链,例如聚乙二醇(PEG)基团,其不饱和变体,其羟基化变体,其酰胺化或其他含N变体,非碳接头;碳水化合物接头;磷酸二酯接头,或能够共价连接两个或多个多肽的其他分子。也可使用非共价接头,例如生物素-链霉亲和素接头。

#### 肽标识符

[0104] 本发明提供肽文库,包括例如在一系列治疗和诊断筛选中有用的多样性肽文库。例如,提供的肽文库可用于筛选疾病特异性或器官特异性肽、筛选具有治疗应用的肽、筛选具有诊断应用的肽、筛选肿瘤靶向肽、筛选抗体表位或抗原、筛选T细胞表位或抗原、筛选抗微生物肽,或其任何组合。

[0105] 利用肽文库的分析可能受到检测和量化单个肽所需的方法的限制。核酸标识符可用于标记文库中具有独特核酸序列的每个肽,允许使用基于核酸的方法(例如PCR扩增或DNA测序)检测和定量单个肽。核酸标识符可以是独特的核酸序列。在一些实施方式中,当多种肽在共同的实验条件下合并时,核酸标识符允许检测和定量单个肽。在一些实施方式中,当多种肽在共同的实验条件下合并时,核酸标识符允许检测和定量单个肽。在一些实施方式中,核酸标识符可用于在一定实验条件下标记核酸(例如,DNA、mRNA、cDNA),用于匹配肽文库中存在于相同实验条件下且具有相同核酸标识符的肽。在一些实施方式中,核酸标识符允许对文库多样性进行验证,其中对核酸标识符进行DNA测序,并将观察到的读数映射到预测的文库序列,以鉴定文库中是否存在肽(图7)。在一些实施方式中,标识符还用于基于读数的定量的定标。在一些实施方式中,标识符是自身标识符,其对应于编码其识别的肽的核酸序列的全部或部分。在一些实施方式中,标识符不是自身标识符(例如,不包含编码其识别的肽的核酸序列)。

[0106] 标识符可以具有任何核酸序列。在一些实施方式中,标识符可以是单链或双链DNA多核苷酸。在一些实施方式中,标识符可以是RNA多核苷酸。在一些实施方式中,标识符可以是杂合的DNA和RNA多核苷酸。在一些实施方式中,标识符可包含合成或化学修饰的核苷酸或偶联物(例如,用于增强稳定性或促进与肽的附接)。

[0107] 核酸标识符可以共价或非共价附接到肽上。

[0108] 蛋白质或肽可具有通过体外翻译方法附接的标识符,以生成蛋白质-mRNA复合物、蛋白质-mRNA-cDNA复合物、蛋白质-DNA复合物、蛋白质-cDNA复合物、蛋白质-核糖体-mRNA复合物或蛋白质-核糖体-mRNA-cDNA (PRMC) 复合物,其可以包含蛋白质开放阅读框 (ORF) 5'端的合成标识符。在一些实施方式中,可以使用包含嘌呤霉素的mRNA模板和包含纯化组分的体外翻译 (IVT) 系统来执行mRNA展示。携带感兴趣蛋白质的mRNA-蛋白质复合物可通过亲和和标记纯化来富集,例如,FLAG-标记纯化。在一些实施方式中,核糖体展示可使用包含缺少终止密码子的间隔子序列的mRNA模板和包含纯化的组分的体外翻译 (IVT) 系统来执行。携带感兴趣蛋白质的蛋白质-核糖体-mRNA复合物可通过亲和和标记纯化,例如,FLAG标记纯化来富集。在一些实施方式中,核糖体展示可使用mRNA-cDNA杂合物作为模板和包含纯化的组分的体外翻译 (IVT) 系统来执行。携带感兴趣蛋白质的PRMC复合物可通过亲和和标记纯化,例如,FLAG标记纯化来富集。在一些实施方式中,DNA展示可使用生物素化DNA模板和包含纯化的组分的体外转录和翻译系统来执行。蛋白质DNA复合物可以通过生物素-链霉亲和素结合形成,通过亲和和标记纯化(例如FLAG-标记纯化)富集蛋白质DNA复合物。在一些实施方式中,mRNA作为体外转录和翻译系统的一部分从DNA合成。在一些实施方式中,在纯化包含肽和mRNA的复合物之前或之后,从mRNA反转录cDNA。

[0109] 标识符可以通过酶法附接到蛋白质或肽上。例如,可以生成包含HaloTag酶序列的融合蛋白,并且酶活性可以将融合蛋白共价偶联到HaloTag配体修饰的双链DNA上。

[0110] 标识符可通过非共价作用附接到蛋白质或肽上。例如,生物素化的多核苷酸标识符可结合到与肽结合的亲和素或链霉亲和素上。亲和素或链霉亲和素可以是肽序列的一部分,或者可以以其他方式与其结合,例如,用于组装蛋白质多聚体的链霉亲和素骨架。

[0111] 在一些实施方式中,独特标识符可用于文库中的每个独特肽。在一些实施方式中,例如当肽包含相同序列时,可在肽文库中的两个或多种肽之间共享标识符。在一些实施方式中,标识符可包括文库的多个或所有肽之间共有的序列,以及文库中的肽独特的序列。在一些实施方式中,标识符可包括文库的多个或所有肽之间共有的一个或多个序列部分,以及文库中肽独特的一个或多个其他序列部分。在一些实施方式中,标识符之间共有的序列可用于标识符扩增(例如,使用合适的引物进行PCR扩增)。在一些实施方式中,对一个标识符独特的或在标识符子集之间共有的序列可用于经由qPCR进行检测或量化(例如,水解探针的序列,例如TaqMan探针)。在一些实施方式中,一个标识符的独特序列或标识符子集之间的共有序列可用于通过测序进行检测或量化。

[0112] 在一些实施方式中,核酸标识符可包括编码氨基酸的序列。在一些实施方式中,两个或多个核酸标识符可包含编码相同氨基酸的不同序列(例如,使用不同密码子)。在一些实施方式中,核酸标识符中的差异密码子利用可允许在标识符中存储额外信息。

[0113] 在一些实施方式中,标识符可包括独特的计算机上生成的序列;每个标识符序列可分配给文库内包含独特序列的肽,并且标识符-肽的分配可存储在数据库中。在一些实施方式中,标识符可包含编码其所识别的肽的全部或部分的核苷酸序列。在一些实施方式中,标识符可包含编码开放阅读框的核苷酸序列。在一些实施方式中,标识符可包含包括启动子序列的核苷酸序列。在一些实施方式中,标识符可包括核苷酸序列,其包括DNA结合蛋白(例如转录因子或聚合酶)的结合位点。在一些实施方式中,标识符可包括核酸酶(例如限制

性内切酶)靶向的一个或多个序列。在一些实施方式中,标识符可包括序列的体外转录和翻译所需的至少一个序列元件。

[0114] 在一些实施方式中,标识符可包含编码主要组织相容性复合物(MHC)分子或其片段的核苷酸序列。在一些实施方式中,标识符可包含编码单链肽-MHC多肽(sc-pMHC)或其片段的核苷酸序列。在一些实施方式中,标识符可包含编码sc-pMHC的抗原肽的核苷酸序列。

[0115] 在一些实施方式中,标识符可以是蛋白质-mRNA复合物的一部分。在一些实施方式中,标识符可以是包含嘌呤霉素连接的蛋白质-mRNA复合物的一部分。在一些实施方式中,标识符可以是蛋白质-mRNA-cDNA复合物的一部分。在一些实施方式中,标识符可以是蛋白质-DNA复合物的一部分。在一些实施方式中,标识符可以是包含生物素-链霉亲和素连接的蛋白质-DNA复合物的一部分。在一些实施方式中,标识符可以是蛋白质-cDNA复合物的一部分。在一些实施方式中,标识符可以是蛋白质-核糖体-mRNA复合物的一部分。在一些实施方式中,标识符可以是蛋白质-核糖体-mRNA复合物的一部分,其中mRNA包含缺少终止密码子的间隔子序列。在一些实施方式中,标识符可以是mRNA-cDNA杂合物的一部分。在一些实施方式中,标识符可以是PRMC复合物的一部分。

[0116] 在一些实施方式中,标识符可包含HaloTag配体,例如,结合到官能团(例如生物素或荧光染料)的氯代烷烃接头。

[0117] 在一些实施方式中,标识符可包含生物素化核苷酸序列。在一些实施方式中,标识符可通过使用生物素化引物的PCR扩增进行生物素化。在一些实施方式中,标识符可通过使用Klenow DNA聚合酶、缺口翻译或混合引物标记RNA聚合酶(包括T7、T3和SP6 RNA聚合酶)酶促纳入生物素化标签(例如,生物素dUTP标签)来生物素化。在一些实施方式中,标识符可通过光生物素化进行生物素化,例如,可将光可活化生物素添加到样品中,且用紫外光照射样品。

[0118] 在一些实施方式中,例如通过模板DNA的PCR扩增,可从模板多核苷酸生成标识符。在一些实施方式中,标识符可以从头生成,例如通过化学合成、固相DNA合成、基于柱的寡核苷酸合成、基于微阵列的寡核苷酸合成或其他合成方法。在一些实施方式中,模板多核苷酸可包括编码开放阅读框的核苷酸序列。在一些实施方式中,模板多核苷酸可包括包括启动子序列的核苷酸序列。在一些实施方式中,模板多核苷酸可包括核苷酸序列,其包括DNA结合蛋白(例如转录因子或聚合酶)的结合位点。在一些实施方式中,模板多核苷酸可包括核酸酶(例如限制性内切酶)靶向的一个或多个序列。在一些实施方式中,模板多核苷酸可包括序列的体外转录和翻译所需的全部序列元件。在一些实施方式中,模板多核苷酸不包括序列的体外转录和翻译所需的全部序列元件。在一些实施方式中,编码两个或多个核酸标识符的模板多核苷酸可包含编码相同氨基酸的不同序列(例如,使用不同密码子)。在一些实施方式中,模板多核苷酸中的差异密码子利用可用于鉴定文库中一类肽或其子集。在一些实施方式中,核酸标识符中的差异密码子利用可允许在模板多核苷酸中存储额外信息。

[0119] 本文的标识符可以是任何长度。在一些实施方式中,标识符的长度可以是约4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105,

106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,155,160,165,170,175,180,185,190,195,200,210,220,230,240,250,260,270,280,290,300,320,340,360,380,400,420,440,460,480,500,或更多核苷酸。

[0120] 在一些实施方式中,标识符的长度可大于至少约4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,155,160,165,170,175,180,185,190,195,200,210,220,230,240,250,260,270,280,290,300,320,340,360,380,400,420,440,460,480,或500个核苷酸。

[0121] 在一些实施方式中,标识符的长度可以是至多约4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,127,128,129,130,131,132,133,134,135,136,137,138,139,140,141,142,143,144,145,146,147,148,149,150,155,160,165,170,175,180,185,190,195,200,210,220,230,240,250,260,270,280,290,300,320,340,360,380,400,420,440,460,480,500,或更多核苷酸。

[0122] 在一些实施方式中,标识符的长度可在约4-500个核苷酸的范围内。在一些实施方式中,标识符的长度可在约25-500个核苷酸的范围内。在一些实施方式中,标识符的长度可在约27-300个核苷酸的范围内。在一些实施方式中,标识符的长度可在约27-120个核苷酸的范围内。在一些实施方式中,标识符的长度可在约50-120个核苷酸的范围内。在一些实施方式中,标识符的长度可在约80-120个核苷酸的范围内。在一些实施方式中,标识符的长度可在约40-50个核苷酸的范围内。在一些实施方式中,标识符的长度可在约5-15个核苷酸的范围内。在一些实施方式中,标识符的长度可在约6-10个核苷酸的范围内。

#### 合成方法

[0123] 肽文库的肽可以通过化学方法合成,例如,茶包(tea bag)合成、数字光刻、针式(pin)合成和SPOT合成。例如,可以通过SPOT合成产生肽阵列,其中氨基酸链通过添加氨基酸和切割侧链保护基团的重复循环构建在纤维素膜上。

[0124] 生物肽文库,如噬菌体展示、细菌展示或酵母展示,涉及融合肽而非分离分子。生物肽库可能经历多样性限制和冗余性。噬菌体或细菌成分的存在可导致混杂效应,例如,分析成分与细菌成分而非感兴趣的肽结合,或在包含细胞的分析中,通过先天免疫机制免疫活化。

[0125] 可使用重组DNA技术表达肽,例如,将表达构建体引入细菌细胞、昆虫细胞或哺乳

动物细胞,并从细胞提取物中纯化重组蛋白。然而,以这种方式产生的重组蛋白通常以可溶性聚集物的形式不溶性表达,可能被蛋白水解或在细胞提取物中检测不到。

[0126] 可通过体外转录和翻译合成肽,其中合成利用无细胞环境中转录和翻译的生物学原理,例如,通过提供核酸模板、相关构建模块(例如RNA、氨基酸)、酶(例如RNA聚合酶、核糖体)和条件。体外转录和翻译可以包括无细胞蛋白质合成(CFPS)。可以使用IVTT系统合成肽,该系统既可以转录,例如,DNA构建体转录为RNA,也可以将RNA翻译为蛋白质。在一些实施方式中,DNA或RNA构建体包含嘌呤霉素。在一些实施方式中,DNA或RNA构建体包含缺少终止密码子的间隔子序列。

[0127] 在一些实施方式中,本文公开的核酸可以从头生成,例如通过化学合成、固相DNA合成、基于柱的寡核苷酸合成、基于微阵列的寡核苷酸合成或其他合成方法。在一些实施方式中,例如通过模板DNA的PCR扩增,可从模板多核苷酸生成本文公开的核酸。

[0128] 编码肽的N-末端甲硫氨酸残基和可切割部分的核苷酸序列可在DNA构建体或RNA构建体中被编码。可切割部分位于使得该肽的至少一个N-末端氨基酸残基在可切割部分之前或之内。在一些实施方式中,该方法包括编码可切割部分,其位于使得肽的一个N-末端氨基酸残基在可切割部分之前或之内。在一些实施方式中,一个N-末端氨基酸残基是甲硫氨酸残基。可以使用酶,例如对可切割部分特异的蛋白酶切割可切割部分,该蛋白酶也可以从肽的剩余部分切掉可切割部分。

[0129] 如本文所述的DNA或RNA构建体中编码的可切割部分的实例包括任何可被酶切割的可切割部分。在一些实施方式中,可由蛋白酶切割可切割部分。可使用对切割部分特异性的酶从肽中切下可切割部分。例如,该酶可以是因子Xa、人鼻病毒3C蛋白酶、AcTEV™蛋白酶, WELQut蛋白酶, Genenase™, 小泛素样修饰物(SUMO)蛋白、Ulp1蛋白酶或肠激酶。通过识别三级结构而非氨基酸序列,Ulp1蛋白酶可以特定的方式切下可切割部分。肠激酶(肠肽酶)也可用于从候选肽中切割可切割部分。肠激酶也可用于在以下切割位点的赖氨酸后切割: Asp-Asp-Asp-Asp-Lys (SEQ ID NO.:7)。肠激酶还可以在其他碱性残基处切割,取决于蛋白质底物的序列和构型。

[0130] 在编码肽的构建体翻译之后,N-末端氨基酸残基可被切割以产生用于高多样性肽文库的肽。在一些实施方式中,至少一个N-末端氨基酸残基被切割以产生肽。在一些实施方式中,一个或多个N-末端氨基酸被切割,例如2,3,4,5,6,7,8,9,10,11,12,13,14,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90,95,100,105,110,115,120,125,130,140,150,160,170,180,190,200,250或更多N-末端氨基酸残基被切割,以产生肽。该N-末端氨基酸可以是任何氨基酸残基。该N-末端氨基酸残基可以是甲硫氨酸氨基酸残基。

#### 制备和使用肽文库的方法

[0131] 本文所述肽文库可用于各种分析。

[0132] 在一些实施方式中,本文所述肽文库可用于分离细胞-肽对。在一些实施方式中,分离细胞-肽对的方法包括将多个细胞与肽文库接触,其中肽文库具有大于10、大于100、大于500、大于1000、大于2000、大于5000、大于10000、大于 $10^6$ 、大于 $10^7$ 、大于 $10^8$ 、大于 $10^9$ ,或大于 $10^{10}$ 以上的独特肽的多样性;以及产生多个隔室,其中所述多个中的隔室包括与所述肽文库的肽结合的所述多个细胞中的细胞,从而在所述隔室中分离所述细胞-肽对。在一些实施方式中,肽文库用于肽文库评估,以在从头发现靶点后分离细胞-肽对。细胞-肽对可以是

受体-配体对。细胞-肽对可以是TCR-抗原对。细胞-肽对可以是BCR-抗原对。在一些实施方式中,可转染或转导细胞以表达受体。在一些实施方式中,可转染或转导细胞以表达TCR。在一些实施方式中,可转染或转导细胞以表达BCR。在一些实施方式中,可转染或转导非淋巴细胞以表达TCR。在一些实施方式中,可转染或转导非淋巴细胞以表达BCR。在一些实施方式中,多种肽中的肽包含如本文所述的标识符。隔室可以是单独的空间,例如孔、平板、分离的边界、相漂移、容器、囊泡、细胞等。

[0133] 本文所述的方法和组合物可用于鉴定细胞-肽对。在一些实施方式中,鉴定细胞-肽对的方法包括将多个细胞与肽文库接触,其中肽文库具有大于10、大于100、大于500、1000、大于2000、大于5000、大于10000、大于 $10^6$ 、大于 $10^7$ 、大于 $10^8$ 、大于 $10^9$ 、大于 $10^{10}$ 的独特肽的多样性;分隔与所述肽库的肽在一个隔室内结合的多个细胞中的细胞,其中肽包含独特肽标识符;并确定与分隔的细胞结合的每条肽的独特肽标识符。在一些实施方式中,肽文库用于肽文库评估,以在从头发现靶点后鉴定细胞-肽对。细胞-肽对可以是受体-配体对。细胞-肽对可以是TCR-抗原对。细胞-肽对可以是BCR-抗原对。在一些实施方式中,可转染或转导细胞以表达受体。在一些实施方式中,可转染或转导细胞以表达TCR。在一些实施方式中,可转染或转导细胞以表达BCR。在一些实施方式中,可转染或转导非淋巴细胞以表达TCR。在一些实施方式中,可转染或转导非淋巴细胞以表达BCR。在一些实施方式中,肽文库是抗原文库。多种肽可以是多个抗原。多种肽可以是如本文所述的多个pMHC多聚体。多种肽可以是如本文所述的多个sc-pMHC。在一些实施方式中,多种肽中的肽包含如本文所述的标识符。

[0134] 在一些实施方式中,本文所述肽文库可用于分离淋巴细胞-肽对。在一些实施方式中,分离淋巴细胞-肽对的方法包括将多个淋巴细胞与肽文库接触,其中肽文库具有大于10、大于100、大于500、大于1000、大于2000、大于5000、大于10000、大于 $10^6$ 、大于 $10^7$ 、大于 $10^8$ 、大于 $10^9$ ,或大于 $10^{10}$ 以上的独特肽的多样性;以及产生多个隔室,其中所述多个中的隔室包括与所述肽库的肽结合的所述多个淋巴细胞中的淋巴细胞,从而在所述隔室中分离所述淋巴细胞-肽对。在一些实施方式中,肽文库用于肽文库评估,以在从头发现靶点后分离淋巴细胞-肽对。淋巴细胞可以是T细胞、B细胞或NK细胞。在一些实施方式中,肽文库是抗原文库。多种肽可以是多个抗原。多种肽可以是如本文所述的多个pMHC多聚体。多种肽可以是如本文所述的多个sc-pMHC。淋巴细胞-肽对可以是TCR-抗原对。淋巴细胞-肽对可以是BCR-抗原对。在一些实施方式中,多种肽中的肽包含如本文所述的标识符。隔室可以是单独的空间,例如孔、平板、分离的边界、相漂移、容器、囊泡、细胞等。

[0135] 本文所述的方法和组合物可用于鉴定淋巴细胞-肽对。在一些实施方式中,鉴定淋巴细胞-肽对的方法包括将多个淋巴细胞与肽文库接触,其中肽文库具有大于10、大于100、大于500、1000、大于2000、大于5000、大于10000、大于 $10^6$ 、大于 $10^7$ 、大于 $10^8$ 、大于 $10^9$ 、大于 $10^{10}$ 的独特肽的多样性;分隔与所述肽文库的肽在一个隔室内结合的多个淋巴细胞中的淋巴细胞,其中肽包含独特肽标识符;并确定与分隔的淋巴细胞结合的每条肽的独特肽标识符。在一些实施方式中,肽文库用于肽文库评估,以在从头发现靶点后鉴定淋巴细胞-肽对。淋巴细胞可以是T细胞、B细胞或NK细胞。在一些实施方式中,肽文库是抗原文库。多种肽可以是多个抗原。多种肽可以是如本文所述的多个pMHC多聚体。多种肽可以是如本文所述的多个sc-pMHC。在一些实施方式中,多种肽中的肽包含如本文所述的标识符。淋巴细胞-肽对

可以是TCR-抗原对。淋巴细胞-肽对可以是BCR-抗原对。

[0136] 在一些实施方式中,本文公开的组合物和方法可用于鉴定结合到受体、免疫受体、TCR、BCR或抗体的多种肽、配体、激动剂、拮抗剂、抗原或表位。在一些实施方式中,本文公开的组合物和方法可用于鉴定结合肽的多个受体、免疫受体、TCR、BCR或抗体。在一些实施方式中,本文公开的组合物和方法可用于鉴定结合多种肽、配体、激动剂、拮抗剂、抗原或表位的多个受体、免疫受体、TCR、BCR或抗体(例如,结合癌症文库,自身免疫文库,或病原体文库中抗原的多个TCR(图8))。

[0137] 在一些实施方式中,本文公开的组合物和方法用于鉴定受体-配体特异性。在一些实施方式中,本文公开的组合物和方法用于鉴定受体-激动剂特异性。在一些实施方式中,本文公开的组合物和方法用于鉴定受体-拮抗剂特异性。在一些实施方式中,本文公开的组合物和方法用于鉴定免疫受体-抗原特异性(例如,TCR-抗原特异性、BCR-抗原特异性)。在一些实施方式中,本文公开的组合物和方法用于鉴定抗体-抗原特异性。

[0138] 在一些实施方式中,本发明的受体、免疫受体、TCR、BCR或抗体的身份通过测序确定(例如,对TCR、BCR或抗体的可变区、高变区或互补决定区(CDR)进行测序)。在一些实施方式中,针对TCR $\alpha$ 链、TCR $\beta$ 链、TCR $\gamma$ 链、TCR $\delta$ 链、抗体重链或抗体轻链鉴定CDR1、CDR2或CDR3序列。

[0139] 在一些实施方式中,肽、配体、激动剂、拮抗剂、抗原或表位的身份通过测序(例如,使用本文所公开的标识符)确定。

[0140] 在一些实施方式中,本文公开的组合物和方法用于鉴定结合TCR的肽、抗原或表位。在一些实施方式中,本文公开的组合物和方法用于确定所鉴定的肽、抗原或表位中的突变如何影响TCR结合(图10)。在一些实施方式中,本文公开的组合物和方法用于鉴定所鉴定的肽、抗原或表位中导致TCR结合亲和力的增强或减弱的突变。在一些实施方式中,本文公开的组合物和方法用于鉴定所鉴定的肽、抗原或表位中维持TCR结合亲和力的突变。在一些实施方式中,本文公开的组合物和方法用于鉴定所鉴定的肽、抗原或表位中导致TCR结合亲和力的丧失的突变。

[0141] 在一些实施方式中,本文公开的组合物和方法用于确定使用本文所述方法鉴定的受体、免疫受体、TCR、BCR或抗体中的突变如何改变肽、配体、激动剂、拮抗剂、抗原或表位的结合。在一些实施方式中,本文公开的组合物和方法用于鉴定所鉴定的受体、免疫受体、TCR、BCR或抗体中导致肽、配体、激动剂、拮抗剂、抗原或表位的结合减少或增加的突变。在一些实施方式中,本文公开的组合物和方法可用于鉴定所鉴定的受体、免疫受体、TCR、BCR或抗体中维持肽、配体、激动剂、拮抗剂、抗原或表位的结合的突变。在一些实施方式中,本文公开的组合物和方法可用于鉴定所鉴定的受体、免疫受体、TCR、BCR或抗体中导致肽、配体、激动剂、拮抗剂、抗原或表位结合的丧失的突变。

[0142] 在一些实施方式中,本文公开的组合物和方法可用于在多种TCR群体中鉴定结合肽文库的给定肽、抗原或表位的TCR。然后,可将所鉴定的TCR与来自不同对象的样品或来自同一对象的不同样品(例如,来自不同组织的样品)的所鉴定的TCR进行比较。

[0143] 在一些实施方式中,本文公开的方法在来自多个对象的T细胞上进行。在一些实施方式中,对来自多个对象的数据分析能鉴定由多个对象识别的抗原。在一些实施方式中,对来自多个对象的数据分析能鉴定由多个TCR克隆型识别的抗原。在一些实施方式中,对来自

多个对象的数据分析能鉴定由多个患者识别的抗原,例如,多个癌症患者、多个具有自身免疫状况的患者或具有针对病原体的保护性免疫的多个患者。在一些实施方式中,对来自多个对象的数据分析能鉴定包括不同HLA型或等位基因的对象中识别的抗原。在一些实施方式中,对来自多个对象的数据分析允许鉴定显示趋集型抗原结合的不同高变或互补决定区序列。

[0144] 在一些实施方式中,本文公开的方法使用多个文库来执行。在一些实施方式中,对来自多个文库的数据分析允许鉴定文库之间共有的反应性抗原,例如,呈现TCR亲和力的抗原,其存在于病原体的多个菌株、多个癌症类型、多个癌症患者、多个自身免疫疾病或多个自身免疫状况中。在一些实施方式中,对来自多个文库的数据分析允许在文库中鉴定不同的反应性抗原,例如存在于病原体菌株、癌症、病况或患者的子集中的抗原。

[0145] 在一些实施方式中,使用本发明的肽文库调查的细胞进行基因表达分析(例如, RNA-seq、qPCR)。在一些实施方式中,对在本发明文库中被鉴定为具有对肽表现出特异性的受体的细胞进行基因表达分析(图11)。例如,确定要表达与病原体文库、癌症文库或自身免疫文库中的抗原结合的TCR的细胞进行基因表达分析。基因表达分析可以是全局性的,也可以是针对性的。分析表达的基因包括但不限于具有已知功能的基因、编码免疫效应分子(例如穿孔素、颗粒酶、细胞因子、趋化因子)、免疫检查点分子、促炎分子、抗炎分子、谱系标志物、整合素、选择素、淋巴细胞记忆标志物、死亡受体、胱冬酶、细胞周期检查点分子、酶、磷酸酶、激酶、脂肪酶和代谢基因的基因。在一些实施方式中,基因表达分析可与肽文库筛选同时进行。在一些实施方式中,基因表达分析可在肽文库筛选结果分析之后进行。在一些实施方式中,基因表达分析可在肽文库筛选结果分析之前进行。在一些实施方式中,基因表达分析允许对使用本文所述方法产生的肽受体配对中鉴定出的感兴趣的细胞进行免疫分型。

[0146] 在一些实施方式中,可针对功能特性筛选肽文库。例如,包含多种肽的肽文库,其中多种肽包含10种以上、100种以上、500种以上、1000种以上、2000种以上、5000种以上、10000种以上、 $10^6$ 种以上、 $10^7$ 种以上、 $10^8$ 种以上、 $10^9$ 种以上或 $10^{10}$ 种以上的独特肽,可以在功能分析中进行筛选。在一些实施方式中,肽文库用于肽文库评估初始功能筛选后的功能特性或额外功能特性。例如,肽文库可与样品接触,然后测试功能特性的诱导。可根据肽诱导功能特性的能力确定肽文库子集。样品可以是生物样品。样品可以是细胞样品。样品可以是T细胞样品。样品可来自对象。对象可以是哺乳动物。对象可以是人。

[0147] 本文所述的方法和组合物可用于筛选分析。例如,肽文库可包含如本文所述的与T细胞样品接触的多个pMHC多聚体。例如,肽库可包含如本文所述的与T细胞样品接触的多个sc-pMHC。接触后,可针对肽文库的pMHC多聚体或sc-pMHC确定T细胞的增殖、T细胞的细胞毒性、对T细胞的抑制、T细胞引起的抑制或T细胞的细胞因子产生(图13)。然后,可以诱导功能特性的pMHC多聚体或sc-pMHC可以制成肽文库子集。例如,肽文库子集可包含pMHC多聚体或sc-pMHC,其在与TCR结合时诱导T细胞增殖,在与TCR结合时诱导细胞毒性,在与TCR结合时抑制T细胞,在与TCR结合时T细胞引起的抑制,在与TCR结合时产生细胞因子,或其任何组合。例如,可通过染料稀释试验(例如,CFSE稀释试验)或DNA复制定量(例如,BrdU掺入试验)来确定增殖。细胞毒性可通过例如基于死亡细胞释放细胞内酶(例如乳酸脱氢酶)的分析,染料排阻分析(例如碘化丙啶),或细胞裂解标志物(例如颗粒酶、CD107a)的表达,通过流式细胞术或qPCR来确定。例如,可以通过ELISA、多重免疫分析、胞内细胞因子染色、ELISPOT、

Western印迹法或qPCR来确定细胞因子的产生。例如,可通过将T细胞克隆与效应细胞和靶抗原共同孵育,测定增殖、细胞毒性、细胞因子产生、活化标志物的表达等来确定T细胞抑制。

[0148] 在一些实施方式中,本文所述肽库可用于分离受体-肽对。在一些实施方式中,分离受体细胞-肽对的方法包括将多个受体与肽文库接触,其中肽文库具有大于10、大于100、大于500、大于1000、大于2000、大于5000、大于10000、大于 $10^6$ 、大于 $10^7$ 、大于 $10^8$ 、大于 $10^9$ ,或大于 $10^{10}$ 以上的独特肽的多样性;以及产生多个隔室,其中所述多个中的隔室包括与所述肽文库的肽结合的所述多个受体中的受体,从而在所述隔室中分离所述受体-肽对。例如,多个受体中的受体可以是TCR、BCR、受体酪氨酸激酶(RTK)、G蛋白偶联受体(GPCR)、配体门控离子通道、细胞因子受体、趋化因子受体或生长因子受体等。在一些实施方式中,受体可以是可溶的。在一些实施方式中,受体可结合到表面。在一些实施方式中,肽文库是抗原文库。多种肽可以是多个抗原。多种肽可以是如本文所述的多个pMHC多聚体。多种肽可以是如本文所述的多个sc-pMHC。受体-肽对可以是TCR-抗原对。受体-肽对可以是BCR-抗原对。在一些实施方式中,多种肽中的肽包含如本文所述的标识符。隔室可以是单独的空间,例如孔、平板、分离的边界、相漂移、容器、囊泡、细胞等。

[0149] 本文所述的方法和组合物可用于鉴定受体-肽对。在一些实施方式中,鉴定受体-肽对的方法包括将多个受体与肽文库接触,其中肽文库具有大于10、大于100、大于500、1000、大于2000、大于5000、大于10000、大于 $10^6$ 、大于 $10^7$ 、大于 $10^8$ 、大于 $10^9$ 、大于 $10^{10}$ 的独特肽的多样性;分隔与所述肽文库的肽在一个隔室内结合的多个受体中的受体,其中肽包含独特肽标识符;并确定与分隔的受体结合的每条肽的独特肽标识符。例如,多个受体中的受体可以是TCR、BCR、受体酪氨酸激酶(RTK)、G蛋白偶联受体(GPCR)、配体门控离子通道、细胞因子受体、趋化因子受体或生长因子受体等。在一些实施方式中,受体可以是可溶的。在一些实施方式中,受体可结合到表面。在一些实施方式中,肽文库是抗原文库。多种肽可以是多个抗原。多种肽可以是如本文所述的多个pMHC多聚体。多种肽可以是如本文所述的多个sc-pMHC。在一些实施方式中,多种肽中的肽包含如本文所述的标识符。受体-肽对可以是TCR-抗原对。受体-肽对可以是BCR-抗原对。

[0150] 本发明提供了用于鉴定例如肽和细胞、肽和受体、肽和免疫受体、肽和TCR、肽和BCR、肽和抗体、配体和细胞、配体和受体、配体和免疫受体、配体和TCR、配体和BCR、配体和抗体、激动剂和细胞、激动剂和受体、激动剂和免疫受体、激动剂和TCR、激动剂和BCR、激动剂和抗体、拮抗剂和细胞、拮抗剂和受体、拮抗剂和免疫受体、拮抗剂和TCR、拮抗剂和BCR、拮抗剂和抗体、抗原和细胞、抗原和受体、抗原和免疫受体、抗原和TCR、抗原和BCR、抗原和抗体、表位和细胞、表位和受体、表位和免疫受体、表位和TCR、表位和BCR、表位和抗体之间配对的组合物和方法。可以在不同的对象(横向)、同一对象在不同的时间点(纵向)或两者中鉴定配对。鉴定的肽、受体或配对可与例如健康、疾病、早期疾病、中期疾病、晚期疾病、进展性疾病、治疗反应、缓解、保护性免疫、自身免疫等相关。在一些实施方式中,缺乏肽或缺乏受体特异性与例如健康、疾病、早期疾病、中期疾病、晚期疾病、进展性疾病、治疗反应、缓解、保护性免疫、自身免疫等相关。

[0151] 在一些实施方式中,本文公开的组合物和方法用于鉴定与保护性免疫、非保护性免疫或自身免疫相关的抗原特异性T细胞效应子克隆。在一些实施方式中,本文公开的组

物和方法用于鉴定抗原特异性T细胞效应子克隆,其表现出无反应性、衰竭、耐受性、自身免疫性、炎症性或抗炎性(例如,Treg)。在一些实施方式中,本文所公开的组合物和方法用于鉴定表现出某些效应子或记忆特性(例如,未成熟、末端效应子、效应子记忆、中央记忆、驻留记忆、 $T_H1$ 、 $T_H2$ 、 $T_H17$ 、 $T_H9$ 、 $T_C1$ 、 $T_C2$ 、 $T_C17$ 、某些细胞因子的产生)的抗原特异性T细胞效应子克隆。

[0152] 在一些实施方式中,使用本文公开的组合物和方法鉴定的TCR、BCR或抗体用作治疗干预的一部分。例如,将TCR序列、TCR可变区序列或CDR序列转染或转导到T细胞中以产生具有相同特异性的T细胞。T细胞可以扩增、极化为所需的效应子表型(例如 $T_H1$ 、 $T_C1$ 、Treg),并注入对象体内。在一些实施方式中,使用本文公开的组合物和方法鉴定的多个TCR、BCR或抗体用于寡克隆治疗。

[0153] 在一些实施方式中,使用本文公开的方法鉴定的肽、配体、激动剂、拮抗剂、抗原或表位用作治疗干预的一部分。在一些实施方式中,肽、抗原或表位用于例如使用抗原呈递细胞、人工抗原呈递细胞、固定化肽或可溶性肽离体扩增细胞群。在一些实施方式中,将扩增的细胞注入患者体内。在一些实施方式中,扩增外周血淋巴细胞。在一些实施方式中,扩增肿瘤浸润性淋巴细胞(TIL)。在一些实施方式中,扩增 $T_H1$ 细胞。在一些实施方式中,扩增细胞毒性T淋巴细胞。在一些实施方式中,扩增调节性T细胞。

[0154] 在一些实施方式中,本文公开的组合物和方法用于鉴定用于疫苗开发的抗原,例如亚单位疫苗、引发覆盖一系列保护性抗原的疫苗,或通用疫苗。

[0155] 在一些实施方式中,本文公开的组合物和方法可用于诊断医学病况。在一些实施方式中,本文公开的组合物和方法用于指导临床决策制定,例如,治疗选择、预后因素的鉴定、治疗反应或疾病进展的监测,或预防措施的实施。

[0156] 本发明的组合物和方法可包括捕获载体。在一些实施方式中,本发明的肽或核酸可逆或不可逆地连接到捕获载体。在一些实施方式中,本发明的肽或核酸化学连接到捕获载体。在一些实施方式中,本发明的肽或核酸共价连接到捕获载体。在一些实施方式中,本发明的肽或核酸非共价连接到捕获载体。在一些实施方式中,本发明的肽或核酸可通过带电相互作用(例如离子键)连接到捕获载体。在一些实施方式中,本发明的肽或核酸可通过氢键连接到捕获载体。在一些实施方式中,本发明的肽或核酸可通过极性键连接到捕获载体。在一些实施方式中,本发明的肽或核酸可通过生物素-链霉亲和素或生物素-亲和素相互作用连接到捕获载体。在一些实施方式中,本发明的肽或核酸可以例如通过化学处理或酶加工从捕获载体有条件地释放。

[0157] 在一些实施方式中,捕获载体可以是固相表面。在一些实施方式中,捕获载体可以包含基质。在一些实施方式中,捕获载体可以包含纳米粒。在一些实施方式中,捕获载体可以包含珠。在一些实施方式中,捕获载体可以包含磁珠。在一些实施方式中,捕获载体可以包含水凝胶。在一些实施方式中,捕获载体可以是油包水乳液液滴的内表面。在一些实施方式中,捕获载体可以包含核酸分子。在一些实施方式中,捕获载体可以包含蛋白质。在一些实施方式中,捕获载体可包含抗体或其衍生物。在一些实施方式中,捕获载体可以包含凝胶。在一些实施方式中,捕获载体可以包含聚合物。在一些实施方式中,捕获探针可以带电。在一些实施方式中,捕获载体可以是荧光的,例如,用一种或多种荧光染料标记。

[0158] 在一些实施方式中,本发明的肽或核酸可通过酶消化从捕获载体上切割。在一些

实施方式中,本发明的肽或核酸可通过限制性内切酶消化从捕获载体上切割。在一些实施方式中,本发明的肽或核酸可通过化学处理或消化从捕获载体上切割。

#### 实施例

[0159] 以下实施例被包括以进一步描述本公开的一些方面,不应用于限制本发明的范围。

#### 实施例1:无偏9聚肽的设计

[0160] 该实施例显示了包含特定长度肽的完整化学空间的9-聚肽文库的鉴定。

[0161] 该实施例与过去的工作不同,在过去的工作中,肽文库受到感兴趣的目标的知识限制。这些文库的设计基于物理相互作用、目标/伴侣特征、生产限制或其他使文库宽度偏移的参数。因此,文库的呈现方式高度偏向。本实施例中的文库设计为包含可能来自自由遗传密码编码的20个氨基酸的所有9-聚肽。

[0162] 该文库设计成包括来自20种已知氨基酸的9聚体的所有序列组合,例如 $5 \times 10^{11}$ 独特肽序列。

#### 实施例2:偏向HLA-A2的9聚肽化学空间调查

[0163] 该实施例展示了9-聚肽文库的鉴定,该肽文库包含特定长度抗原的完整化学空间,该抗原对与主要组织相容性复合物蛋白质(例如HLA-A2)的相互作用具有特异性。

[0164] 如实施例1所示,该肽文库不受感兴趣靶点或相互作用抗原的知识限制。HLA-A2有一个众所周知的结合基序,其具有2位和9位处的关键氨基酸,其可包括I、V或L。该文库设计用于包含在两个指定位置具有任意这些序列的所有9-聚肽。得到的文库是实施例1中描述的9聚体文库的子集,因为该文库在2和9位处限制了9聚体的可变性。

[0165] 该文库设计成包括来自20种已知氨基酸的9聚体的所有序列组合,其在2位和9位有约束。共有 $1 \times 10^{10}$ 个独特的9聚肽序列被鉴定为HLA-A2限制性。其他MHC复合物也有类似的限制。

#### 实施例3:人病毒组9聚肽文库

[0166] 该实施例显示了包含特定长度的人病毒抗原的完整化学空间的9-聚肽文库的鉴定。

[0167] 如在线数据库Uniprot所示,选择纳入肽文库的病毒为综合人类定殖病毒蛋白质组,并补充了Uniprot蛋白质组程序搜索确定的毒株的分类群标识符。使用REST API从Uniprot的蛋白质组标识符下载完整和部分蛋白质组。人工验证所选毒株的蛋白质组覆盖率。该文库随后进一步扩展,以包括来自现有正汉坦病毒的所有蛋白质组,以及高度可变分类群(HIV和流感)的额外多样性。

[0168] 已鉴定的蛋白质组中每种蛋白质的每个9-聚体都包含在该文库中。该文库也是实施例1中所述9聚体无偏文库的子集,因为该文库限制了9聚体的可变性是衍生自综合人病毒蛋白质组。

[0169] 另一个文库被鉴定为限制于MHC复合蛋白之一的9-聚体,如实施例2中所述。 $3 \times 10^6$ 种肽中共有 $1.5 \times 10^5$ 个独特的9聚肽被鉴定为HLA-A2限制性。

#### 实施例4:巨细胞病毒9聚肽文库

[0170] 本实施例展示了包含巨细胞病毒(CMV)完整蛋白质组的9-聚肽文库的鉴定。

[0171] 该文库设计用于包括CMV蛋白质组中每种蛋白质的每个9-聚体。产生的文库包括

$7 \times 10^4$ 个独特的9聚肽。该文库也是实施例3中所述9聚人病毒文库的子集,因为该文库限制了9聚体的可变性衍生自CMV。

[0172] 另一个文库被鉴定为限制于MHC复合蛋白之一的9-聚体,如实施例2中所述。 $7 \times 10^4$ 种肽中共有 $4 \times 10^3$ 个独特的9聚肽被鉴定为HLA-A2限制性。

#### 实施例5:巨细胞病毒pp65蛋白的9聚肽文库

[0173] 本实施例展示了包含巨细胞病毒(CMV)蛋白pp65的完整蛋白序列的9-聚肽文库的鉴定。

[0174] 来自pp65蛋白的每个9聚体都包含在该文库中。产生的文库包括571个独特的9聚肽。这也是实施例4中所述9-聚体文库的子集,因为这些9-聚体与pp65蛋白质相关。

[0175] 另一个文库被鉴定为限制于MHC复合蛋白之一的9-聚体,如实施例2中所述。571种肽中共有26个独特的9聚肽被鉴定为HLA-A2限制性。

#### 实施例6:9聚肽文库的表位特异性位置扫描突变

[0176] 本实施例展示了9-聚肽文库的鉴定,包括表位的完整突变扫描。

[0177] 如Hoppes等人在J Immunol, 2014, 193中所述,为pp65蛋白的NLVPMVATV表位设计了一个具有单突变的9聚体库。产生的文库包括172个独特的9聚肽。这也是实施例1中所述9-聚体文库的子集,因为这些9-聚体在基于NLVPMVATV的9聚体整条序列上包括位点突变。

[0178] 为pp65蛋白的NLVPMVATV表位设计了一个含有两个突变的9-聚体的文库。产生的文库包括13,168个独特的9聚肽。

[0179] 为pp65蛋白的NLVPMVATV表位设计了一个含有三个突变的9-聚体的文库。产生的文库包括589,324个独特的9聚肽。

[0180] 为pp65蛋白的NLVPMVATV表位设计了具有所有突变的9-聚体库。产生的文库包括 $5.12 \times 10^{11}$ 个独特的9聚肽。

#### 实施例7:9-聚肽文库的产生

[0181] 根据本领域已知的方法,或由商业供应商合成生产,或根据制造商的说明使用肽合成仪,产生任何先前实施例中描述的肽文库。

#### 实施例8:肽MHC文库的体外翻译

[0182] 该实施例证明了蛋白质的无细胞合成(CFPS)。

[0183] 肽文库的无细胞合成(CFPS)能够产生范围广泛的各种肽。通过CFPS获得高产量需要使用细菌系统,其中翻译序列的第一个氨基酸是N-甲酰甲硫氨酸(fMet)。该残基与甲硫氨酸的不同之处在于包含中性甲酰基团(HCO)而不是带正电荷的氨基-末端( $\text{NH}_3^+$ )。因此,每个肽文库变体将包含fMet。然而,MHC I类分子的肽结合沟的结构被设计为特异性容纳任何给定肽的带正电的氨基末端,并且不能充分契合具有fMet序列起始的肽。肽装载不成功会影响折叠,并会导致错误折叠、无功能的MHC,因为这两个过程是相互关联的。尽管细菌能使用内源性氨肽酶来切割fMet,但其去除可能不完全或被取消,这取决于序列中第二个氨基酸的身份。例如,甲硫氨酸氨肽酶在fMet和天冬氨酸之间的切除效率低。因此,CMV衍生肽,该系统中的模型肽,最终将在单链设计中生产为fMet-NLVPMVATV;整个分子不正确折叠,且不会结合其同源T细胞受体。如果蛋白质是在由粗细胞提取物制成的细菌CFPS系统中产生的,预期是这样的结果。此外,在文库的情况下,如果处理效率低,单独模板可能会产生有或没有fMet的肽,或两者的混合物。在一个仅由纯化的组分构成且完全缺乏甲硫氨酸氨肽酶

的重构CFPS系统中,所有文库变体都将由fMet残基起始。

[0184] 为了解决这个问题,构建体被工程改造为包括编码酶促切割结构域和文库多肽的基因。去除至少初始甲硫氨酸氨基酸可以成功折叠肽并将其装载到MHC蛋白上。此外,去除至少初始甲硫氨酸氨基酸允许肽库多样性的上限更大,例如, $20^x$ ,其中x是肽的长度,而包含该残基会将文库多样性限制为 $20^{(x-1)}$ 。

[0185] 在该实施例中,在无细胞条件下合成肽。将所有CFPS组分在冰上解冻并混合,然后移至相关温度以启动反应。加入试剂:40% (v/v) PURExpress溶液A,30% (v/v) of PURExpress溶液B (E6800L,新英格兰生物实验室公司 (New England Biolabs, Inc.)), 0.8U/ $\mu$ l RNA酶抑制剂(10777019,赛默飞世尔科技公司 (ThermoFischer Scientific)), 4% (v/v) 每种二硫化物增强剂1和2 (E6820L,新英格兰生物实验室公司),0.004U/ $\mu$ l稀释于PBS的蛋白酶反应(英杰公司 (Invitrogen)),无核酸酶水和20ng/ $\mu$ l编码所需CFPS产物的相应质粒DNA的反应。试验了四种不同的CFPS温度:20、25、30和37°C。在每个指定的时间点,取样并通过将试管置于冰上并添加EDTA至终浓度2mM来停止反应。

[0186] 图2A显示酶切以增加肽多样性。泳道对1-2、3-4、5-6、7-8和9-10分别表示在没有可切割部分的模板上、具有未添加蛋白酶的可切割部分的模板、具有反应完成后添加蛋白酶的可切割部分的模板、具有反应期间存在蛋白酶的可切割部分的模板上进行CFPS反应以及缺乏模板的反应。在加入100mM DTT的还原条件下制备奇数泳道中的样品用于凝胶电泳。在室温下4小时后,通过将试管置于冰上来终止所有反应。将4U/反应的蛋白酶添加到样品3-8中。在泳道5-6上样的反应中,将管置于冰上后将蛋白酶与10mM EDTA一起加入,然后转移到室温3.5小时,然后再次置于冰上。

[0187] 进行蛋白质印迹以确定总蛋白质产量。每个CFPS样品与水、4x样品缓冲液和1M DTT混合,在95°C下煮沸5分钟,然后加载到10%SDS-PAGE凝胶上。使用HRP-抗-FLAG抗体对样品进行印迹。

[0188] 图2B显示了来自CFPS反应的样品,其中包含有或没有可切割部分的多聚体和单体模板。用抗FLAG-HRP抗体对样品进行印迹分析和检测。在室温下4小时后,通过将试管置于冰上来终止反应。

#### 实施例9:评估体外翻译蛋白质的3-D结构

[0189] 本实施例展示了CFPS蛋白折叠成可识别的三维结构。

[0190] 在本实施例中,通过抗体测试实施例8中生成的CFPS蛋白质的构象识别。错误折叠或未折叠的蛋白质不被抗体识别。下面的实施例表明,在酶切割结构域切割后,CFPS蛋白被折叠并被抗体构象识别。

[0191] 通过ELISA测量蛋白表达。用在100mM碳酸氢盐/碳酸盐包被缓冲液中稀释的抗链霉亲和素抗体(410501,白乐津公司 (Biolegend))包被板,并在4°C下孵育过夜。然后,通过用洗涤缓冲液(补充有0.05%吐温-20的PBS)填充孔将板洗涤3次,并通过用封闭缓冲液(补充有2% (V/V) BSA的洗涤缓冲液)填充孔在室温下封闭2小时。然后用封闭缓冲液中每种CFPS蛋白质的系列稀释液填充孔,然后在室温下孵育1小时。然后,用洗涤缓冲液洗涤板三次,并用封闭缓冲液中稀释的含有0.15 $\mu$ g/ml辣根过氧化物酶偶联的对蛋白质特异性的抗体室温下孵育1小时。

[0192] 再洗涤三次后,通过向每个孔中加入3,3',5,5' 四甲基联苯胺底物显色,并通过加

入市售终止溶液终止反应。使用平板阅读器测量450nm处的吸光度。值是多次重复的均值。平板用粘性塑料覆盖并在所有孵育期间在旋转器上轻轻搅拌。每个样品的浓度是从阳性对照蛋白质的标准曲线中内推的。

[0193] 图2C显示由CFPS产生并经蛋白水解切割的肽折叠成可识别的三维结构。用ELISA检测线性表位或构象表位,并计算正确折叠百分比。蛋白酶被添加到两个CFPS反应中。该图表明通过构象表位抗体的识别证明了蛋白酶切割的肽或未切割的肽(包含fMet)是否为正确折叠。

[0194] 图2D提供单链肽MHC(sc-pMHC)多聚体与抗原特异性T细胞的结合。通过CFPS和酶切割产生多聚体。将T细胞与多聚体孵育,然后用荧光检测抗体染色,并通过流式细胞术进行分析。富含CMV的T细胞(供体153,Astarte3835FE18,目录号1049)被用于FACS染色。96孔圆底微量滴定板的孔中充满T细胞,细胞用冰冷的FACS缓冲液(D-PBS、2mM EDTA和2%(V/V)胎牛血清)洗涤一次,以300g在4℃离心,去除上清液。然后,将相应孔用Fc受体封闭液在4℃温和搅拌下封闭30分钟,用FACS缓冲液洗涤,去除上清液。将FACS缓冲液添加到补偿对照孔中。

[0195] 在下一步中,将细胞与20nM阳性对照或取自实施例8的指定CFPS反应的样品稀释液在4℃下孵育30分钟,然后用FACS缓冲液洗涤一次。将在FACS缓冲液中稀释的100nM检测抗体加入每个孔中,将板在4℃下避光孵育30分钟,然后用PBS洗涤两次,并用可固定活力染料APC-efluor780(1:8000稀释,50μl/孔)染色,在室温下放置15分钟。然后将板用FACS缓冲液洗涤两次,并用固定缓冲液PBS、3.7%甲醛(v/v)、2%FBS(v/v)固定。最后,将样品转到FACS管中进行分析。

#### 实施例10:哺乳动物细胞中的肽文库产生

[0196] 如实施例8所述,肽在哺乳动物细胞中通过无细胞蛋白质合成产生,或如实施例7所述通过合成产生。

[0197] 对于哺乳动物表达,在哺乳动物表达载体中设计了一种编码CMV肽的构建体,其具有带或不带C-末端His标记的C-末端Flag标记。根据厂商的建议,通过瞬时转染Expi293F或ExpiCHO-S细胞(生命技术(Life Technologies))表达肽。

[0198] 用抗Flag亲和层析(Genscript)或Ni亲和层析从细胞培养上清液中纯化肽。在亲水性树脂(GE生命科学)上进行尺寸排阻层析(SEC),该树脂在20mM HEPES,150mM NaCl,pH 7.2中预平衡。

[0199] 或者,使用23mM磷酸钠,500mM氯化钠,500mM咪唑(pH 7.4)的柱缓冲液,通过Ni亲和层析纯化肽,而无需SEC纯化。

[0200] 哺乳动物细胞中产生的肽通过280nm处的UV进行定量,而CFPS产生的肽通过夹心ELISA相对于标准蛋白质进行定量。

#### 实施例11:将肽标识符附接到文库肽

[0201] 此实施例示范如何使用肽标识符标记文库肽。

[0202] 通过实施例7所述的合成,或通过实施例8所述的无细胞蛋白质合成,或如实施例10所述在哺乳动物细胞中产生CMV肽。

[0203] 如本文所述产生的肽用一个或多个肽标识符(例如,DNA片段)标记。商业合成(Integrated DNA Technologies)或PCR扩增每个肽标识符。通过混合50%v/v肽和50%v/v

肽标识符实现标记,并通过wester印迹法向上漂移进行确认。

[0204] 图3显示具有一个或多个肽标识符的CMV肽的western印迹。底部箭头表示裸露肽。中间的箭头表示具有一个肽标识符的肽。上部的箭头表示具有两个肽标识符的肽。

#### 实施例12:HLA-A2的9聚肽文库

[0205] 该实施例显示了包含对HLA-A2特异性的抗原一定长度的完整化学空间的9-聚肽文库的鉴定。

[0206] 这与过去的工作不同,在过去的工作中,通过物理相互作用鉴定感兴趣的特定目标,然后以高度偏向的方式呈现。HLA-A2有一个众所周知的结合基序,其具有2位和9位处的关键氨基酸,其可包括I、V或L。该实施例中的文库设计成仅包括在两个指定位置具有这些序列的所有9-聚肽,得到 $1 \times 10^{10}$ 个肽。

[0207] 构建体经过工程改造,以包括编码一个酶切结构域和 $1 \times 10^{10}$ 个独特9聚肽之一的基因。

[0208] 根据实施例8中所述的类似方法生成肽文库。然后将肽文库装载到HLA-A2分子上以生成肽/MHC (pMHC) 文库。

[0209] 得到的pMHC文库可用于T细胞筛选,以确定抗原反应性T细胞。例如,见Simon等, *Cancer Immunol Res*, 2014, 2 (12) :1230-1244。

#### 实施例13:与细胞结合的肽文库

[0210] 该实施例演示了检测与细胞结合的肽。

[0211] 为了测试用肽标识符标记的肽的功能性,获得了肽特异性 (CMV) 和非肽特异性 (HPV) T细胞 (Astarte Biologics)。冷冻T细胞根据制造商的指南解冻。在4°C,用20%v/v Fc-Block和0.1mg/ml鲑鱼精封闭细胞30分钟。然后将细胞与10%V/V肽标识符标记的肽在4°C的FACS缓冲液 (D-PBS、2mM EDTA和2% (V/V) FBS) 中孵育30分钟并洗涤。

[0212] 细胞进一步分为两部分,其中使用流式细胞术检测肽结合,并通过qPCR检测标识符。对于基于蛋白质的检测,将细胞与2%v/v的抗Flag抗体 (Biolegend) 在4°C下孵育30分钟并洗涤。最后,将细胞固定在固定缓冲液 (D-PBS、3.7%甲醛和2%FBS) 中,并在流式细胞仪上进行分析。

[0213] 图4显示裸露肽、肽标识符标记的肽或阴性对照与肽特异性或非肽特异性T细胞的相对结合。与非肽特异性T细胞相比,裸露肽和肽标识符标记的肽显示与肽特异性T细胞结合。

[0214] 对于基于肽标识符的检测,使用细胞裂解和RNA稳定试剂盒 (Life Technologies Corporation) 裂解细胞,并根据制造商的方案制备qPCR主混合物。使用管家基因特异性引物 (如Rp113) 将Ct值对内部对照定标。使用 $\delta$ - $\delta$ Ct法将相对值与来自不含肽的T细胞的相对值进行比较。

[0215] 图5显示了对管家基因定标的肽标识符标记的肽的相对量。与肽标识符标记的肽一起孵育的肽特异性T细胞比与肽标识符标记的肽孵育的非肽特异性T细胞具有更多的信号,表明T细胞与肽之间存在特异性相互作用。此外,裸露肽对肽特异性和非肽特异性T细胞几乎没有可检测的信号。

#### 实施例14:TCR抗原特异性谱的鉴定

如实施例6所述,设计包含NLVPMVATV表位的完整突变扫描的9-聚体文库。如实施

例8所述合成包含9-聚体的sc-pMHC,并如实施例11所述附接标识符。该文库与多个T细胞一起孵育,并且T细胞被分选到单细胞隔室中。裂解T细胞,并从裂解的T细胞产生包含标识符的核酸。合并这些核酸并测序。读数中的标识符允许肽标识符与来自同一隔室的T细胞序列相匹配。通过从一个隔室中鉴定TCR序列(例如,可变区、高变区或CDR)并定量来自同一隔室中的肽标识符读数来确定TCR抗原特异性谱(图9A)。鉴定出导致TCR结合亲和力增加或降低的所鉴定TCR-抗原对的抗原中的表位突变。

#### 实施例15:结合靶抗原的TCR的鉴定

[0216] 如实施例14所述产生测序数据。对于测序的每个肽标识符,鉴定相应的TCR序列(例如,可变区、高变区或CDR)。鉴定出对肽文库的肽具有结合亲和力的多个TCR,并鉴定出对特定TCR具有结合亲和力的多种肽(图9B)。

#### 实施例16:不同TCR的汇集

[0217] 使用实施例14和实施例15中描述的方法进行实验。T细胞是来源于不同对象的原代T细胞。鉴定对肽库的肽具有结合亲和力的来自不同个体的TCR(图9C)。

#### 实施例17:CMV抗原发现和疫苗设计

[0218] 该实施例展示本文所公开的组合物和方法用于发现特定抗原和T细胞受体序列,以及疫苗和细胞疗法的后续设计。

[0219] 计划接受造血干细胞移植(HSCT)的对象被纳入研究。在HSCT后第0天和第30天抽血。从血液中分离并培养T细胞。将培养的T细胞与本发明的sc-pMHC文库(例如,包含源自具有位置扫描的CMV基因组、转录组或蛋白质组的肽的文库)一起孵育,并且将细胞分选入单细胞隔室。

[0220] 裂解T细胞,并从裂解的T细胞产生包含标识符的核酸。合并这些核酸并测序。读数中的标识符允许肽标识符与来自同一隔室的T细胞序列相匹配。通过从一个隔室中鉴定TCR序列(例如,可变区、高变区或CDR)并定量来自同一隔室中的肽标识符读数来确定TCR抗原特异性谱。对于测序的每个肽标识符,鉴定相应的TCR序列。鉴定出对肽库的一种或多种肽显示结合亲和力的多个TCR,并鉴定出对一个或多个TCR显示结合亲和力的多种肽。对象分为CMV血清阳性或血清阴性,并根据CMV控制或再活化情况进行额外分类。比较来自对象的结果。已鉴定出与CMV控制相关的肽和TCR序列,并用于设计CMV疫苗和细胞疗法。

#### 实施例18:检查点抑制剂无应答者的疫苗和TCR细胞治疗

[0221] 该实施例展示本文所公开的组合物和方法用于发现与检查点抑制剂治疗的反应相关的特定抗原和TCR序列,以及疫苗和细胞疗法的后续设计。

[0222] 计划接受非小细胞肺癌(NSCLC)或结直肠癌(CRC)检查点抑制剂治疗的对象被纳入研究。在给予检查点抑制剂之前,以及给予检查点抑制剂之后对对象进行纵向活检,并留出实现治疗效果的时间。从活检中分离并培养T细胞。将培养的T细胞与本发明的sc-pMHC文库(例如,包含源自NSCLC/CRC基因组、转录组或蛋白质组的肽的文库)一起孵育,并且将细胞分选入单细胞隔室。

[0223] 裂解T细胞,并从裂解的T细胞产生包含标识符的核酸。合并这些核酸并测序。读数中的标识符允许肽标识符与来自同一隔室的T细胞序列相匹配。通过从一个隔室中鉴定TCR序列(例如,可变区、高变区或CDR)并定量来自同一隔室中的肽标识符读数来确定TCR抗原特异性谱。对于测序的每个肽标识符,鉴定相应的TCR序列。鉴定出对肽库的一种或多种肽

显示结合亲和力的多个TCR,并鉴定出对一个或多个TCR显示结合亲和力的多种肽。

[0224] 可对对象进行纵向跟踪,并在使用检查点抑制剂的两个或多个治疗周期内对活检进行分析。

[0225] 对象分为检查点抑制剂应答者和无应答者。比较来自对象的结果。鉴定了与检查点抑制剂治疗的成功响应相关的肽和TCR序列。已鉴定的肽和TCR序列可在第二个或随后的检查点抑制剂治疗周期中确认,或在随后纳入的对象中确认。已鉴定的肽和TCR序列用于设计癌症疫苗和细胞疗法。

#### 实施例19:通用流感疫苗

[0226] 该实施例展示本文所公开的组合物和方法用于发现与对流感株的免疫应答相关的特定抗原和TCR序列,以及疫苗,包括通用流感疫苗的后续设计。

[0227] 纳入对象以接种疫苗或感染各种流感株。对象感染流感,接种减毒活流感株或接种流感亚单位疫苗。

[0228] 在第7天(感染/接种前)、感染/接种后第10天和感染/接种后第45天从对象获取纵向血样。

[0229] 从血样中分离T细胞并培养。将培养的T细胞与本发明的sc-pMHC文库(例如,包含源自具有位置扫描的流感基因组、转录组或蛋白质组的肽的文库)一起孵育,并且将细胞分选入单细胞隔室。

[0230] 裂解T细胞,并从裂解的T细胞产生包含标识符的核酸。合并这些核酸并测序。读数中的标识符允许肽标识符与来自同一隔室的T细胞序列相匹配。通过从一个隔室中鉴定TCR序列(例如,可变区、高变区或CDR)并定量来自同一隔室中的肽标识符读数来确定TCR抗原特异性谱。对于测序的每个肽标识符,鉴定相应的TCR序列。鉴定出对肽文库的一种或多种肽显示结合亲和力的多个TCR,并鉴定出对一个或多个TCR显示结合亲和力的多种肽。通过分析感染/接种不同对象的肽-MHC汇集物,可以鉴定保护性抗原。保护性抗原被连接到疫苗中,以提供广泛或普遍的保护,抵御多种流感病毒株。

#### 实施例20:糖尿病的Treg治疗

[0231] 该实施例展示本文所公开的组合物和方法用于发现与自身免疫有关的特异性抗原和TCR序列,以及耐受性细胞疗法的后续设计。

[0232] 该研究的一部分利用了从0/1期1型糖尿病对象(以及匹配的健康对照组)采集的尸检组织样品。组织样品包括β岛、血液、脾脏、淋巴结和骨髓。在研究的第二部分中,纳入0/1期1型糖尿病的活对象(以及匹配的健康对照组)。定期从对象抽取血样。

[0233] 从血液和组织样品中分离T细胞并培养。将培养的T细胞与本发明的sc-pMHC文库(例如,包含源自健康或自身免疫人对象的基因组、转录组或蛋白质组的肽的文库)一起孵育,并且将细胞分选入单细胞隔室。

[0234] 裂解T细胞,并从裂解的T细胞产生包含标识符的核酸。合并这些核酸并测序。读数中的标识符允许肽标识符与来自同一隔室的T细胞序列相匹配。通过从一个隔室中鉴定TCR序列(例如,可变区、高变区或CDR)并定量来自同一隔室中的肽标识符读数来确定TCR抗原特异性谱。对于测序的每个肽标识符,鉴定相应的TCR序列。鉴定出对肽库的一种或多种肽显示结合亲和力的多个TCR,并鉴定出对一个或多个TCR显示结合亲和力的多种肽。

[0235] 比较来自对象的结果。鉴定了与1型糖尿病相关的肽和TCR序列。鉴定的肽和TCR序

列用于耐受性细胞治疗,例如,离体扩增的寡克隆Treg极化T细胞,其表达对自身免疫抗原特异性的TCR。

#### 实施例21:多孔水凝胶的制备

[0236] 该实施例显示了可用于本发明组合和方法中的多孔水凝胶的制备。水凝胶珠是通过在不同相对浓度下混合丙烯酸胺单体单元和双丙烯酸胺交联剂单元以及丙烯酸化寡核苷酸引物的混合物,使用微流体Drop-maker封装在液滴中,并孵育混合物直到交联完成而制备的。在该实施例中,预交联水性混合物包括10%TEBST(Tris-EDTA缓冲盐水加吐温-20)中的0.75%双丙烯酸胺、3%丙烯酸胺、5 $\mu$ M 5'-丙烯酸化反向引物#1、25 $\mu$ M 3'-封端(磷酸化)和5'-丙烯酸化反向引物#2(图16),0.5%过硫酸铵。引物可设计为包括用于酶切的序列,例如限制性内切酶靶向的序列,以允许部分引物从水凝胶中释放。可以使用任何合适的限制性内切酶。在该实施例中,反向引物1包括XhoI消化位点,反向引物2包括FokI消化位点。混合并搅拌水性混合物的所有试剂。向混合物中添加1.5%的TEMED和1%的008-FluoroSurfactant,封装在液滴中,在室温下孵育1小时,然后转移到60°C的烘箱中过夜孵育,从而形成水凝胶。水凝胶珠用20%的1H,1H,2H,2H-全氟-1-辛醇(PFO)洗涤一次,然后用TEBST洗涤3次,然后用低TE(1mM Tris-Cl pH 7.5,0.1mM EDTA)洗涤3次。水凝胶珠4°C储存在TEBST中直至使用。

#### 实施例22:水凝胶上全长抗原编码模板的PCR(PCR1)

[0237] 本实施例显示了水凝胶上全长抗原编码模板的PCR。编码单链多聚肽-MHC的线性DNA模板在单模板条件下以液滴形式PCR扩增到水凝胶珠上,每滴最多得到一个DNA模板。将实施例21中制备的1.4mL水凝胶珠与如下的PCR组分在2mL反应体积中混合:400 $\mu$ L Q5反应缓冲液(新英格兰生物实验室(New England Biolabs))、40 $\mu$ L 10mM dNTP、40 $\mu$ L 25 $\mu$ M正向引物#1、40 $\mu$ L 1 $\mu$ M非丙烯酸化反向引物#1(图16)、40 $\mu$ L 0.1pg/u1线性DNA模板(或模板混合物)、8 $\mu$ L 20% IGEPAL和20 $\mu$ L Q5 DNA聚合酶(新英格兰生物实验室(New England Biolabs))。以液滴形式包封混合物,并进行35轮PCR。通过添加等体积的100%全氟辛醇(PFO)使液滴裂解后,用10体积的低TE洗涤水凝胶五次。用限制性内切酶(本实施例中为XhoI)在37°C下切割反向引物#1 1小时,消化水凝胶珠的等分(10 $\mu$ L每等份),并在1.2%琼脂糖凝胶上跑胶,在水凝胶上定量扩增子的产量和质量。如图17A所示,全长抗原编码模板PCR扩增到水凝胶(“珠”)上。

#### 实施例23:标识符的PCR(PCR2)

[0238] 该实施例显示了实施例21和22中产生的水凝胶上标识符的PCR扩增。可以使用本文公开的任何合适的标识符。在该实施例中,使用自身标识符,其对应于编码其识别的肽的核酸序列的全部或部分。PCR1后经洗涤的水凝胶珠用虾碱性磷酸酶(新英格兰生物实验室(New England Biolabs))消化,以去除反向引物#2上的3'帽,然后用10体积的低TE进一步洗涤5次。将300 $\mu$ L水凝胶珠与如下PCR组分在400 $\mu$ L反应体积中混合:80 $\mu$ L Q5反应缓冲液(新英格兰生物实验室)、8 $\mu$ L 10mM dNTP、8 $\mu$ L 25 $\mu$ M 5'-生物素化正向引物#2、1.6 $\mu$ L 20% IGEPAL和4 $\mu$ L Q5 DNA聚合酶(新英格兰生物实验室)。以液滴形式包封混合物,并进行20轮PCR。通过添加等体积的100%PFO使液滴裂解后,用10体积的低TE洗涤水凝胶五次。用切割反向引物2的限制性内切酶(本实施例中为FokI)在37°C下消化水凝胶珠的小等分一小时,并在1.2%琼脂糖凝胶上跑胶,在水凝胶上确定扩增子的产量和质量。如图17B所示,将标识

符PCR扩增到水凝胶上(“自识别核酸”)。分析了三种独立的珠制备物:一个具有对应于CMV肽的模板,一个具有HPV肽,一个具有编码两种肽的模板混合物(混合物)。PCR2产生的自识别核酸片段显示为约100bp。

#### 实施例24:单链多聚肽-MHC的体外转录/翻译(IVTT)

[0239] 该实施例显示单链肽-MHC可在体外转录和翻译,例如,使用实施例21和22中产生的水凝胶上抗原编码DNA模板。120 $\mu$ L水凝胶珠与240 $\mu$ L IVTT主混合液共同包封在液滴中,包括120 $\mu$ L PURExpress溶液A(新英格兰生物实验室)、90 $\mu$ L PURExpress溶液B(NEB)、6 $\mu$ L RNA酶OUT(Invitrogen)、12 $\mu$ L的每种二硫键增强剂#1和#2(NEB)以及12 $\mu$ L Ulp1蛋白酶(Invitrogen)。液滴在22 $^{\circ}$ C下无振荡培养20小时。将D-生物素添加到IVTT反应中,使其最终浓度达到500 $\mu$ M,然后添加等体积的100%PFO,破碎液滴。用10体积的PBS加2%的BSA洗涤水凝胶珠五次。一等份的水凝胶在室温下用1:10稀释的Alexa-488标记的抗 $\beta$ -2-微球蛋白(B2M)抗体(R&D系统)在PBS+2%BSA中免疫荧光染色1小时,然后在PBS+2%BSA中进行5次10倍洗涤,并通过共聚焦显微镜(Imagexpress Micro, Molecular Devices, 图18A)显影。在21%的珠中观察到染色,证实了PCR1中的单模板条件,和单链肽-MHC的成功产生。

#### 实施例25:标识符标记的单链多聚肽-MHC从水凝胶的释放与分析

[0240] 该实施例展示了折叠的标识符标记的单链肽-MHC(sc-pMHC)多聚体从水凝胶中释放。使用实施例21、22、23和24的方法生成sc-pMHC。sc-pMHC多聚体通过DNA标识符与水凝胶结合。通过DNA与水凝胶结合的sc-pMHC可通过任何合适的核酸酶消化从水凝胶中释放。在该实施例中,在Cutsmart缓冲液(NEB)中用Benzonase核酸酶(一种非特异性核酸内切酶)或FokI(一种限制性内切酶)消化DNA,在22 $^{\circ}$ C孵育20小时。ELISA检测消化释放的蛋白质,以确定产量和折叠。采用1:1333稀释的HRP偶联的抗B2M(Biolegend)或构象敏感抗HLA抗体(Santa Cruz),以HEK生产的sc-pMHC为标准品进行检测。ELISA证实了高度折叠的sc-pMHC多聚体的释放(图18B)。还通过Western印迹进行检测,在3-8%的Tris醋酸盐凝胶上进行电泳,用硝化纤维素进行印迹,用PBS加3%BSA封闭,并用1 $\mu$ g/mL大鼠抗Flag(Biolegend)一抗和1:1000偶联Alexa647的抗大鼠IgG二抗(Invitrogen)检测,测试了消化释放的蛋白质。FokI释放的sc-pMHC多聚体相对于Benzonase核酸酶释放的sc-pMHC或相对于来自体外转录/翻译上清液的上清液的缓慢迁移证明了使用核酸标识符成功标记sc-pMHC(图18C)。

#### 实施例26:水凝胶/液滴中单链多聚肽-MHC的功能分析

[0241] 该实施例证明通过本发明方法产生的sc-pMHC特异性结合关联T细胞。通过流式细胞术和单细胞包封/测序,证实如实施例25所述从水凝胶释放的sc-pMHC与关联肽扩增的T细胞特异性结合。

[0242] 对于流动,用HPV肽或CMV肽扩增的 $10^5$ 个供体T细胞用大量溶液或如上所述的液滴中由CMV肽编码模板产生的sc-pMHC多聚体染色。与HEK细胞中产生的多聚体CMV或HPV pMHC相对应的对照蛋白也用于染色。所有pMHC均在PBS加10%FBS中稀释,并使用抗Flag-APC(Biolegend)作为二抗。如图19所示,在水凝胶/液滴中产生的CMV sc-pMHC多聚体对CMV扩增的T细胞的染色与大量溶液或HEK细胞产生的类似。尽管HPV扩增的T细胞对HEK产生的HPV-pMHC多聚体的染色接近100%阳性,但这些细胞上液滴产生的CMV pMHC多聚体没有明显的染色,证实了特异性。

[0243] 对于单细胞测序,使用T7核酸外切酶(NEB)处理液滴中产生的具有自识别核酸标

识符标记(如实施例21-25所述)的CMV-sc-pMHC多聚体。然后将CMV-sc-pMHC多聚体与HEK产生的HPV-pMHC多聚体混合,用不同的标识符标记。该抗原混合物用于对HPV和CMV扩增的T细胞的混合物进行染色,随后对其进行单细胞测序。单细胞测序显示水凝胶/液滴产生的CMV sc-pMHC多聚体具有极好的特异性。如图20所示,与液滴产生的CMV pMHC相对应的UMI与用CMV肽扩增的T细胞结合,而不与用HPV扩增的T细胞结合。

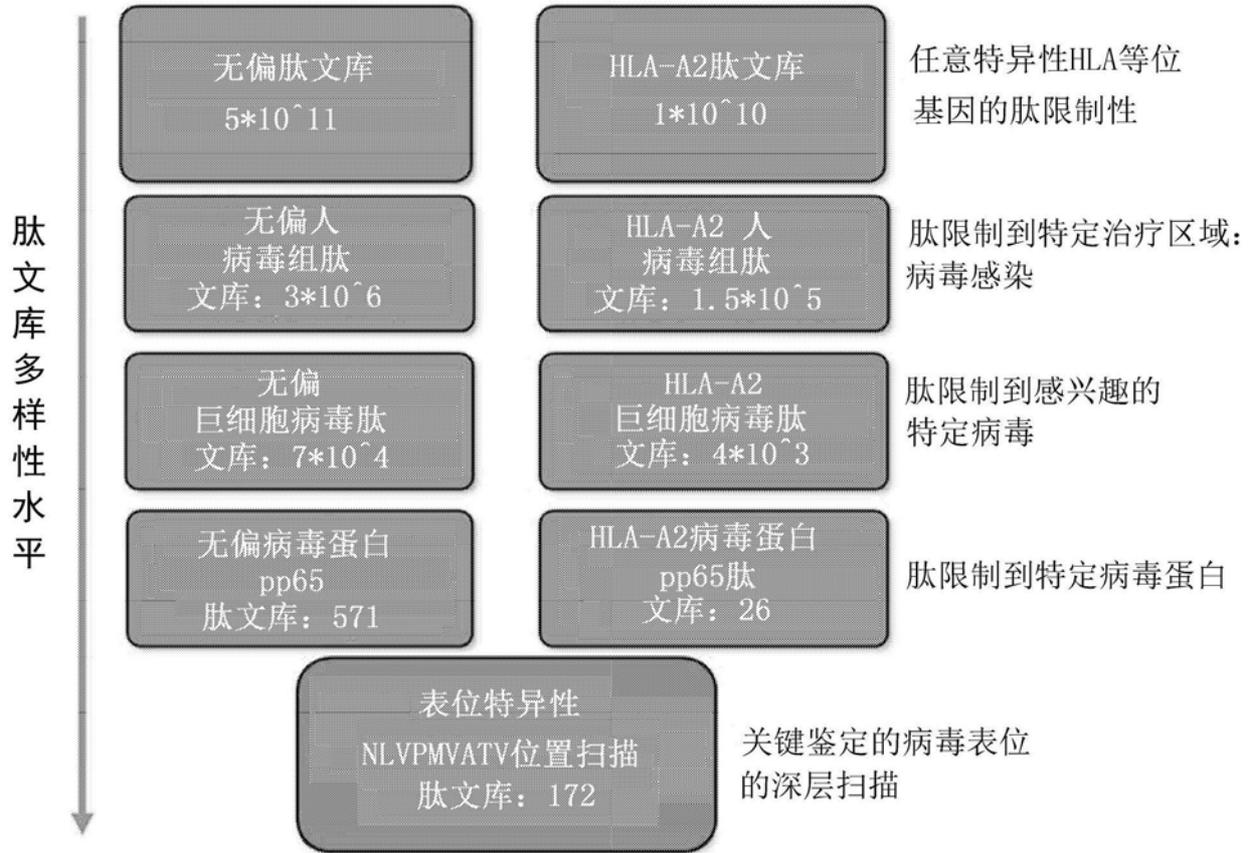


图1

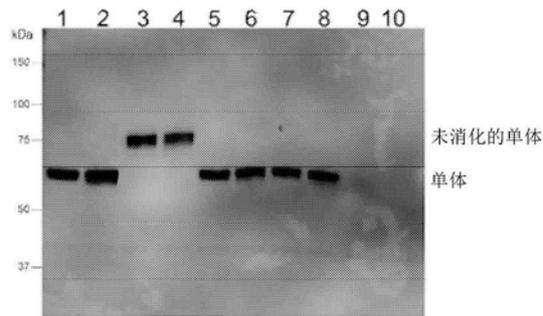


图2A

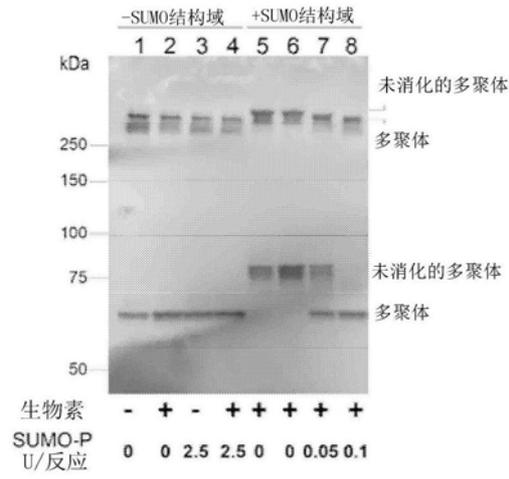


图2B

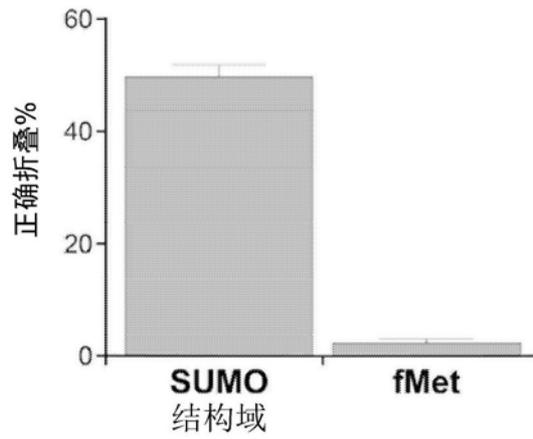


图2C

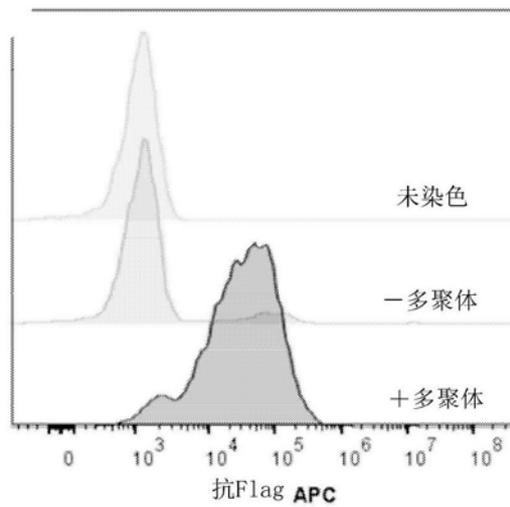


图2D

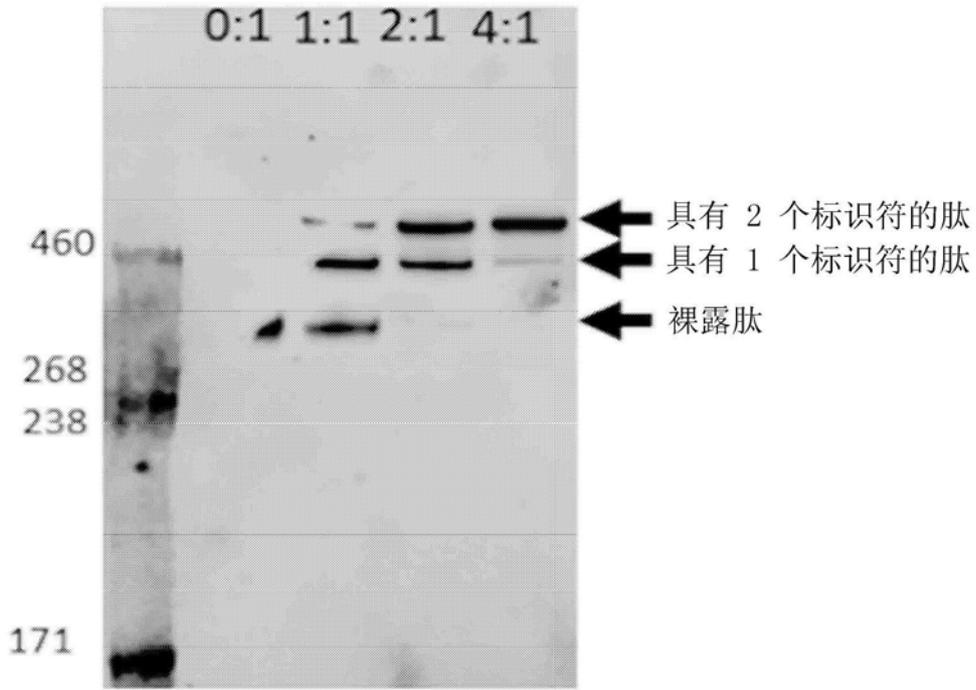


图3

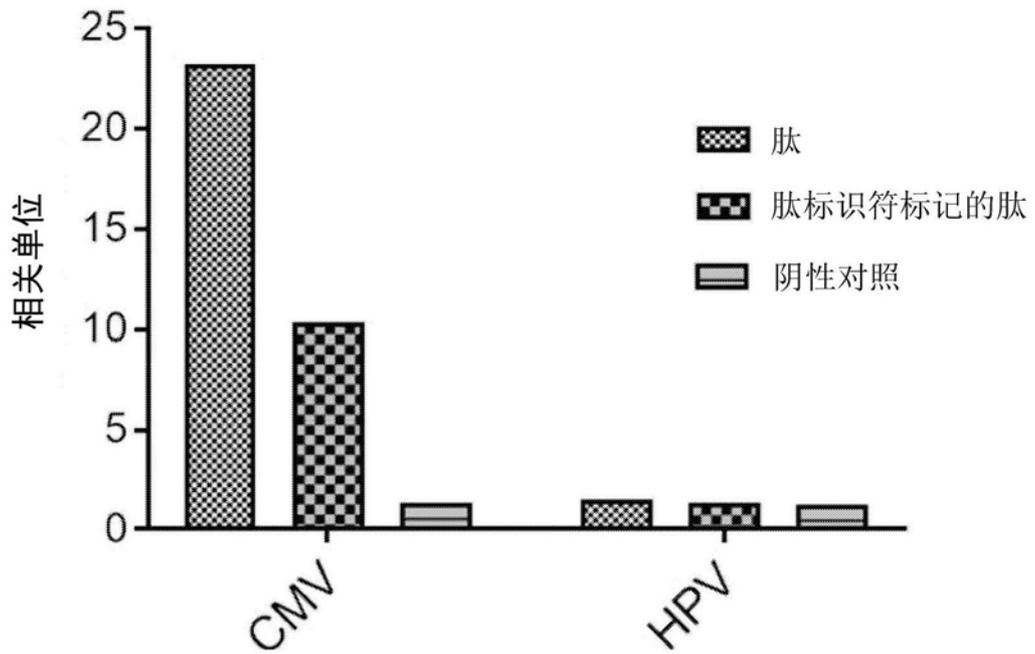


图4

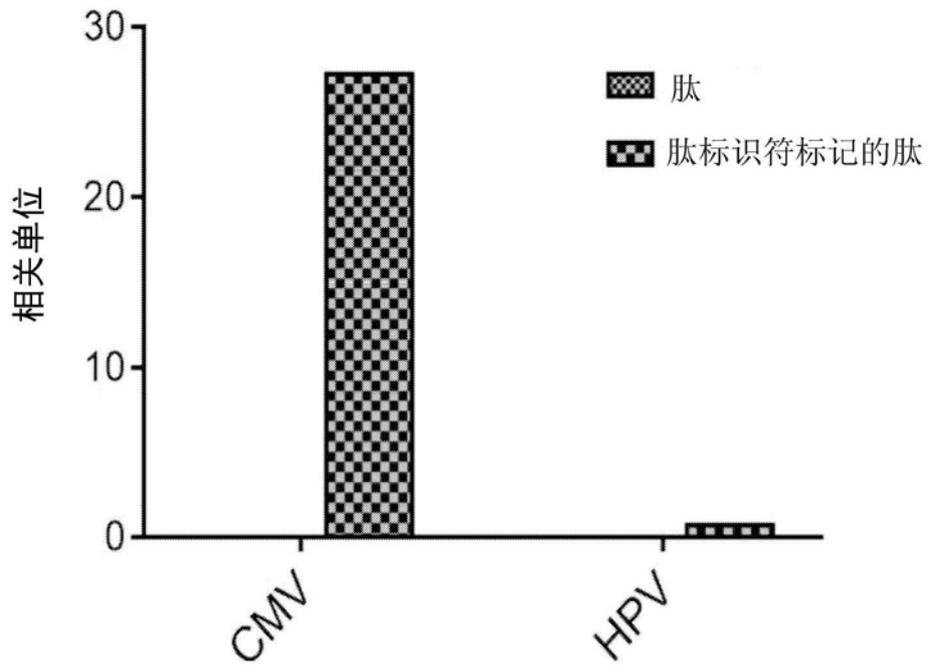


图5

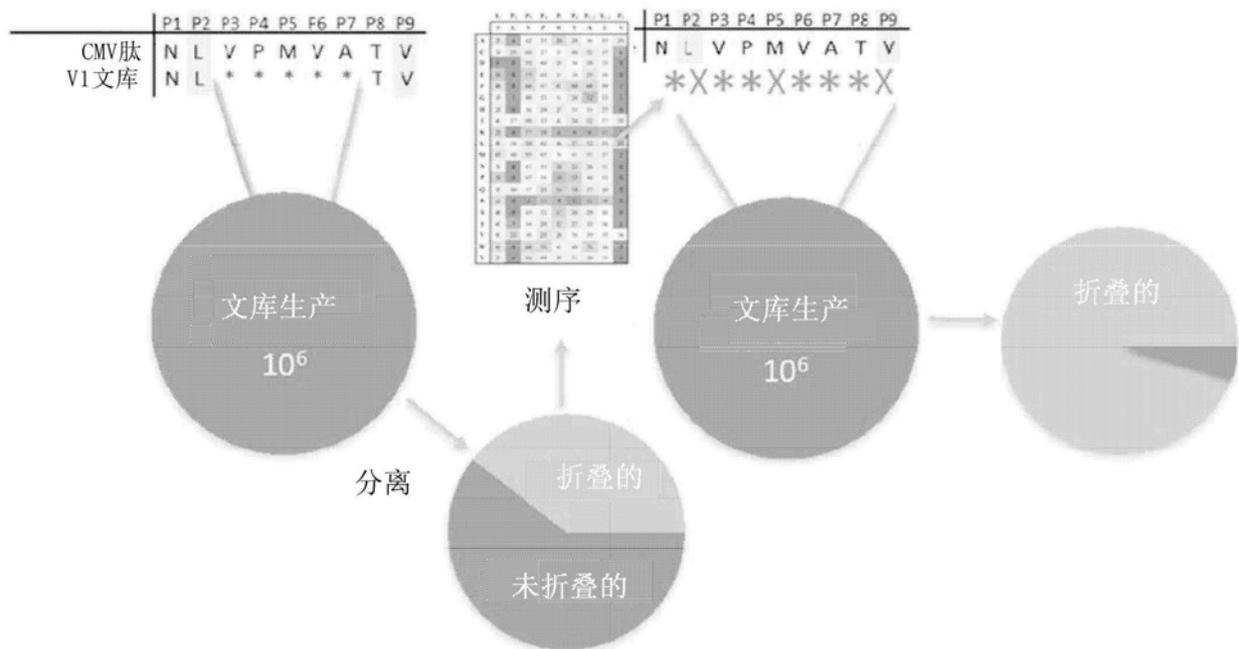


图6

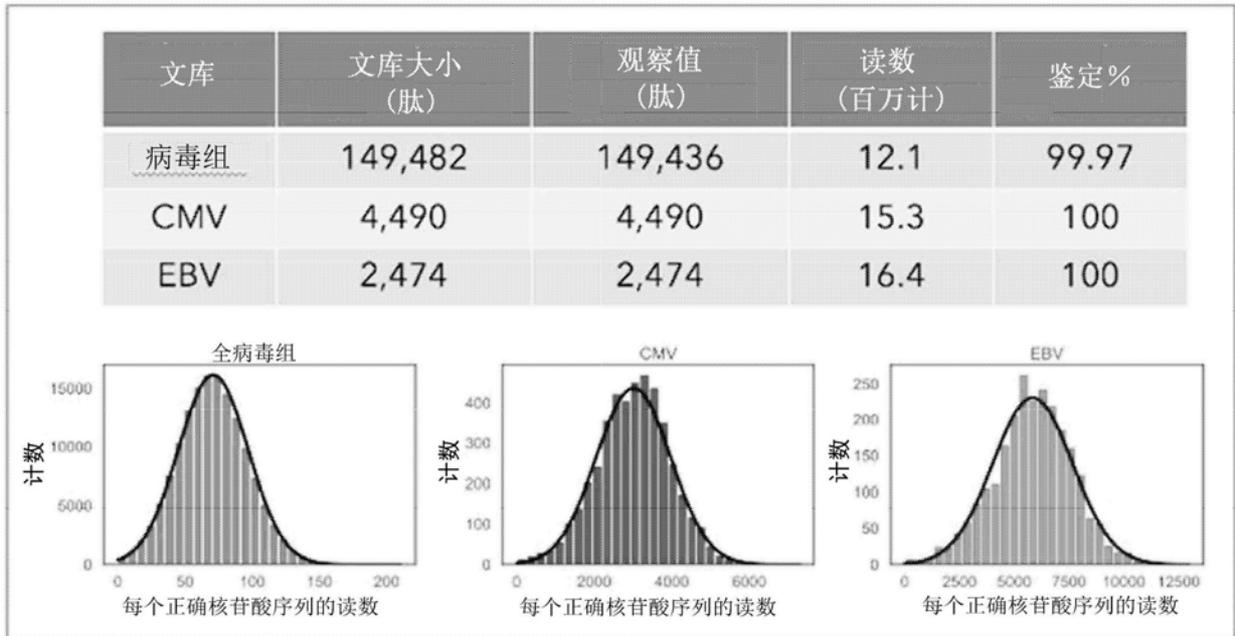


图7

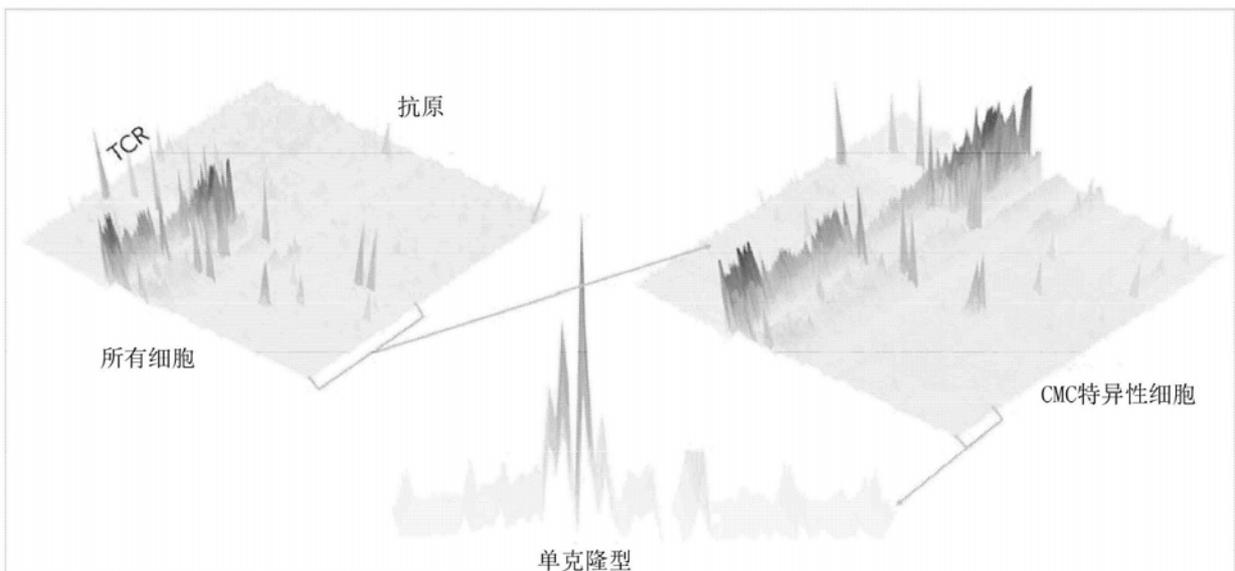
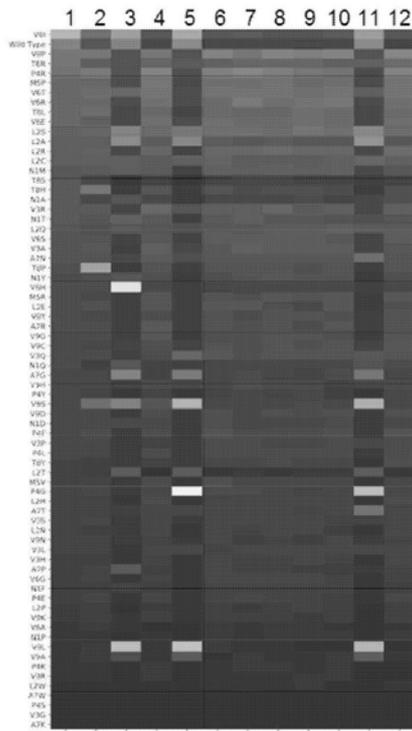


图8

患者46: 最高的12个克隆型



主要T细胞克隆型

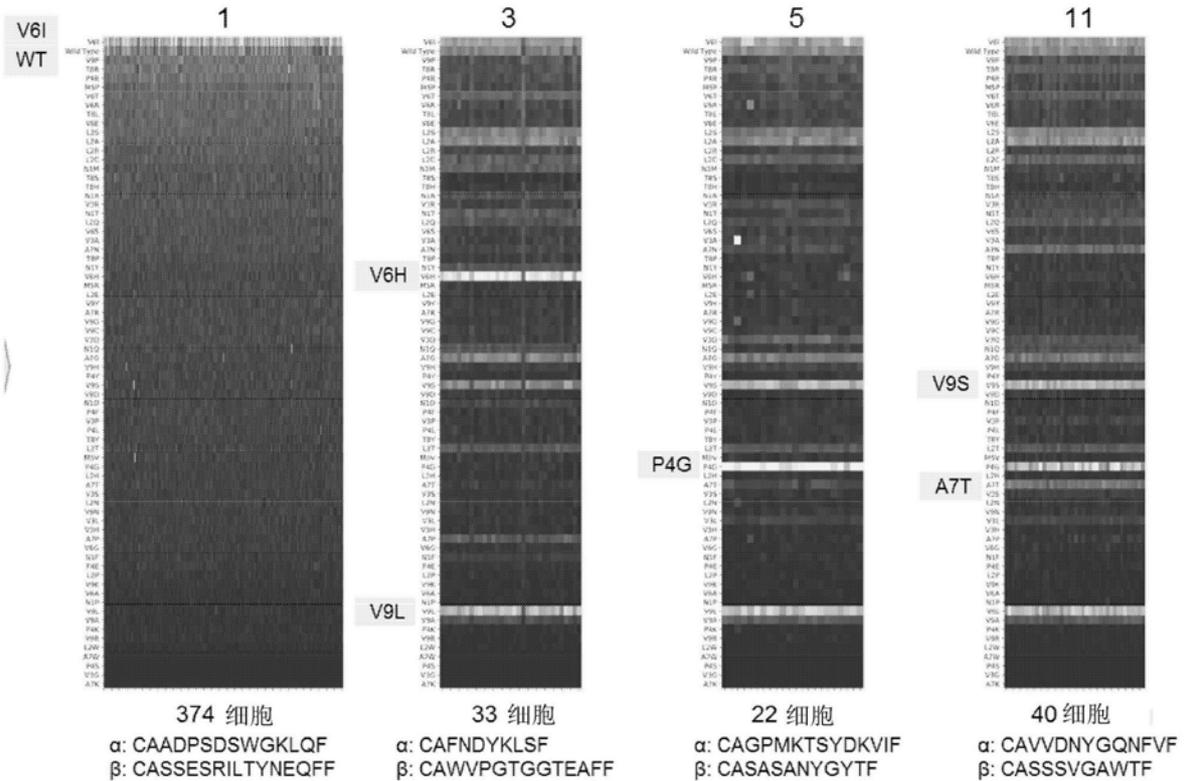


图 9A

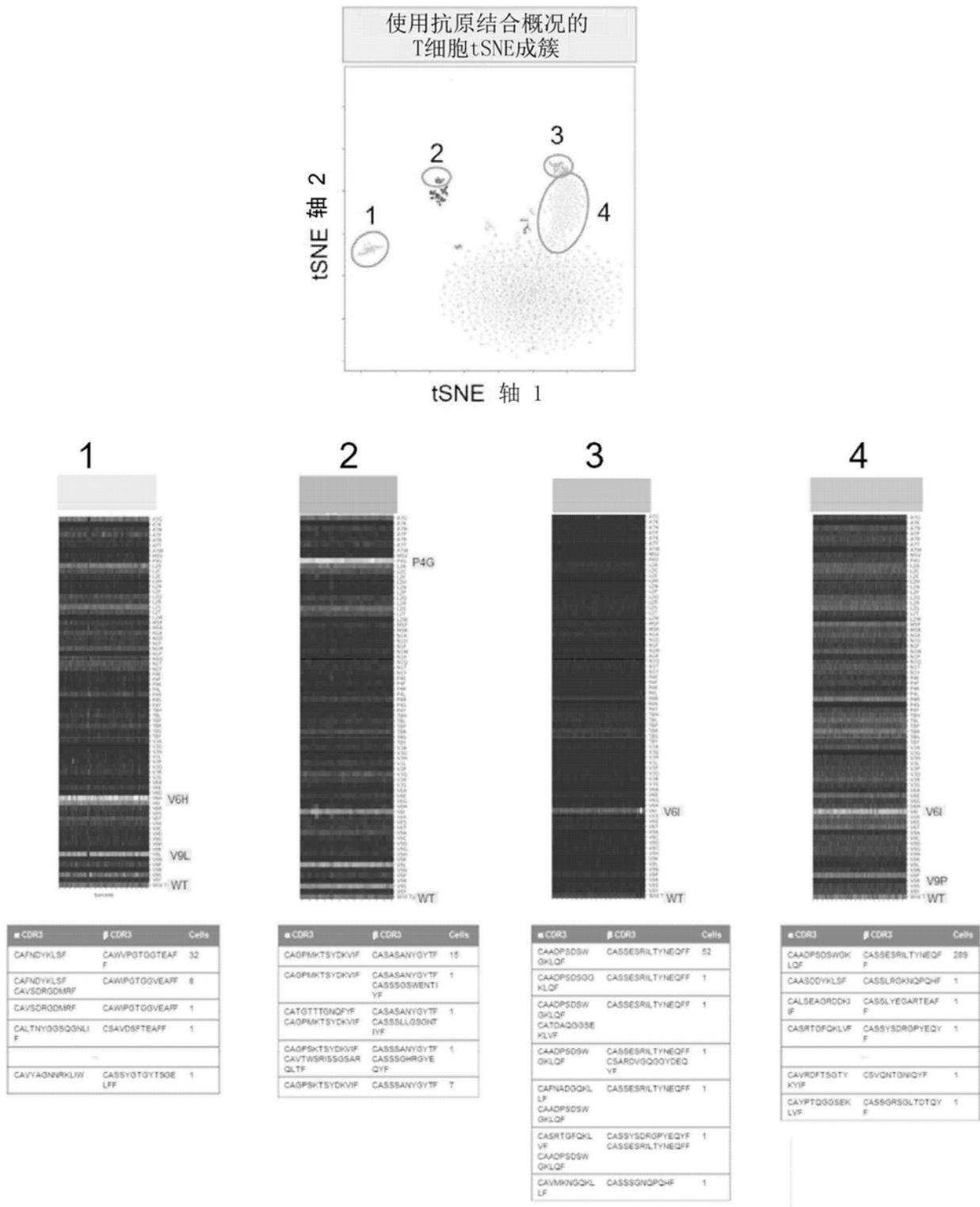


图 9B

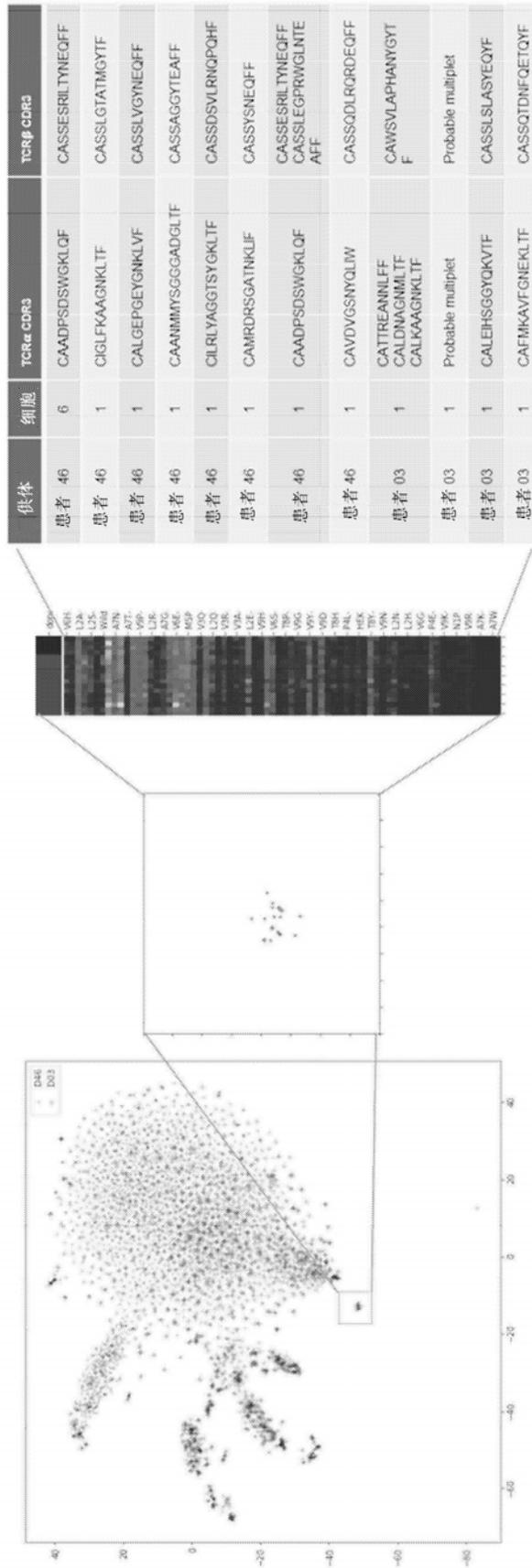


图 9C

图9

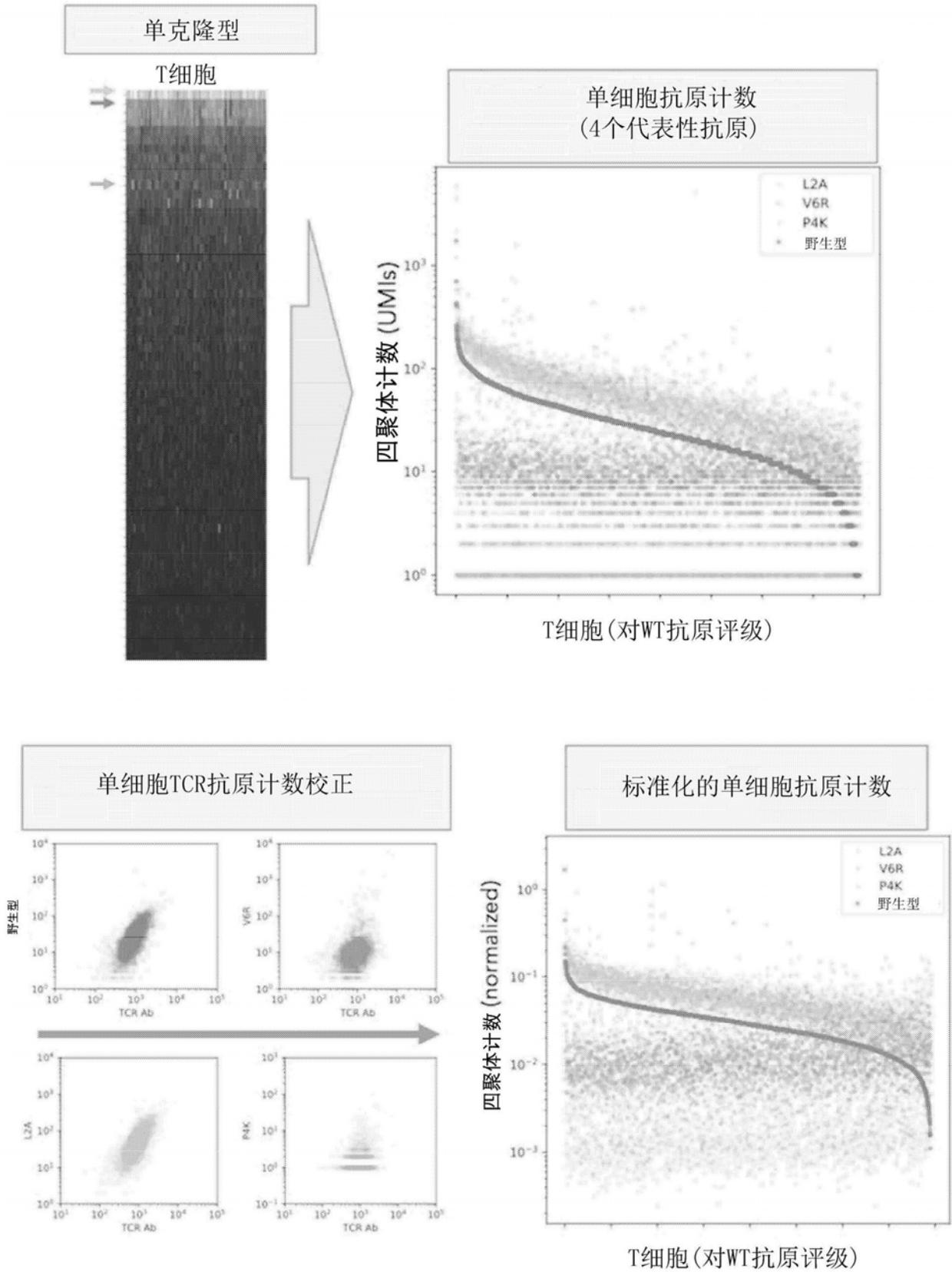


图10

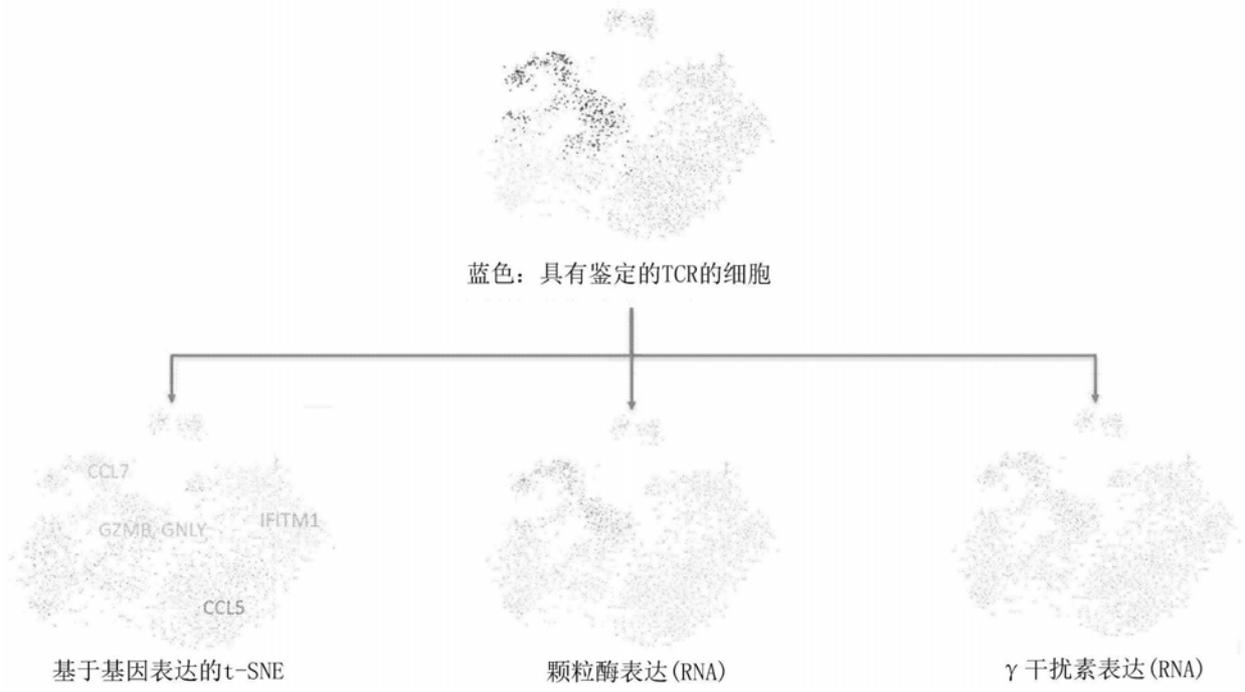


图11

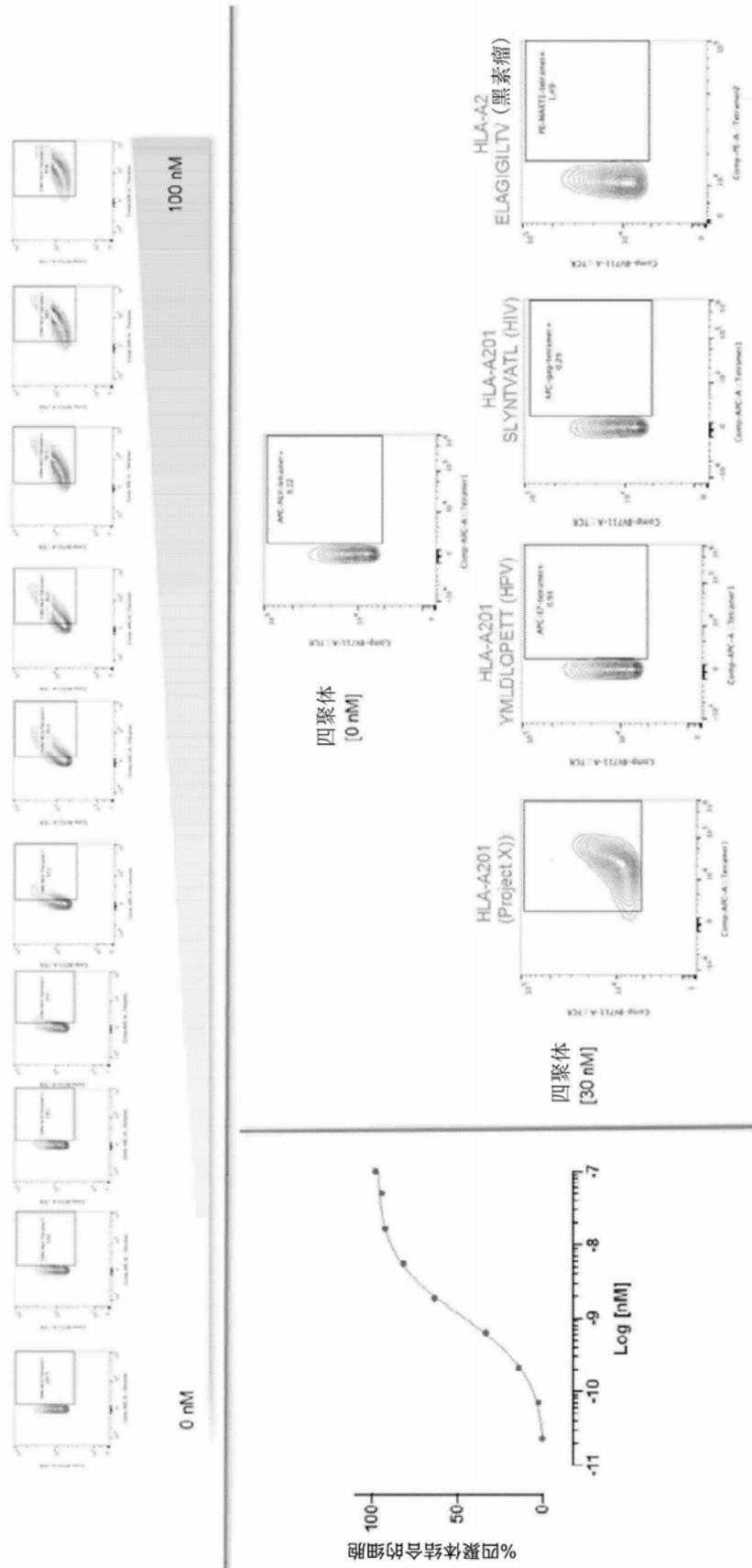


图12



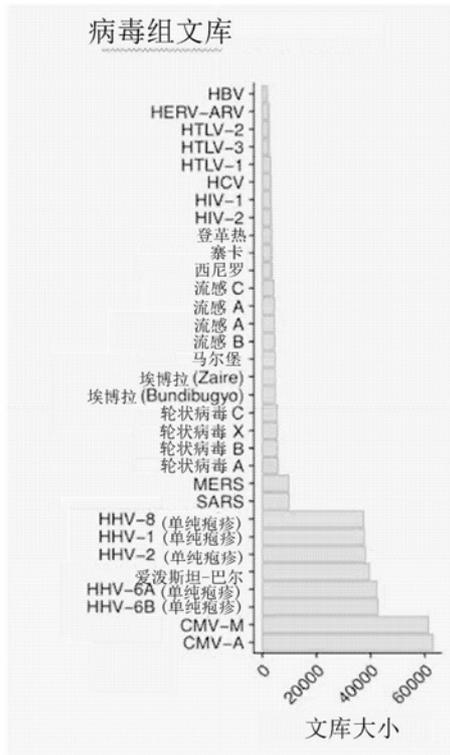


图14A

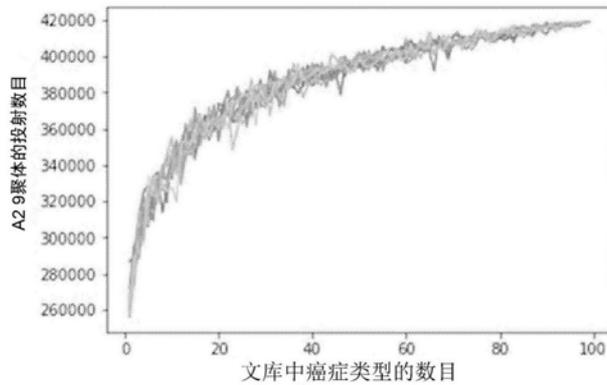
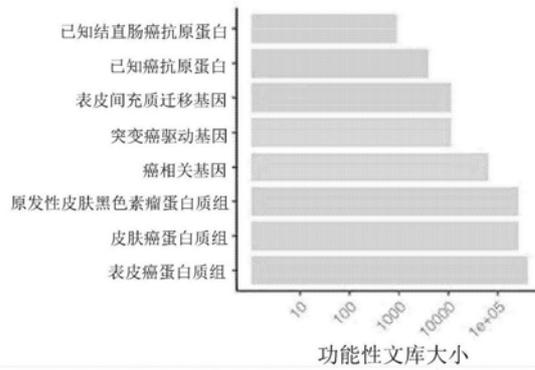


图14B

图14

单链基因构建体

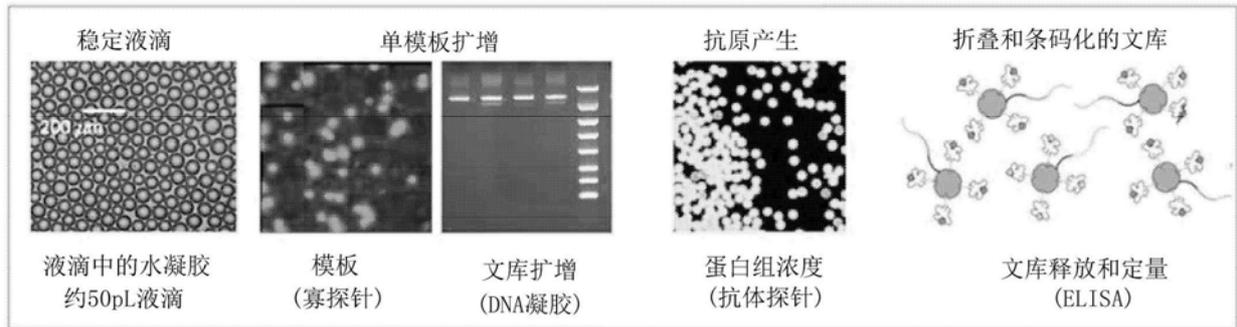
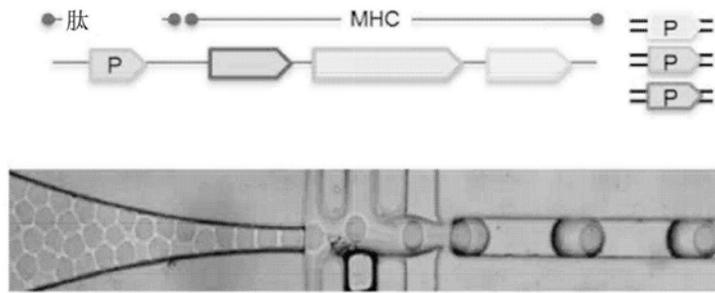


图15



图16

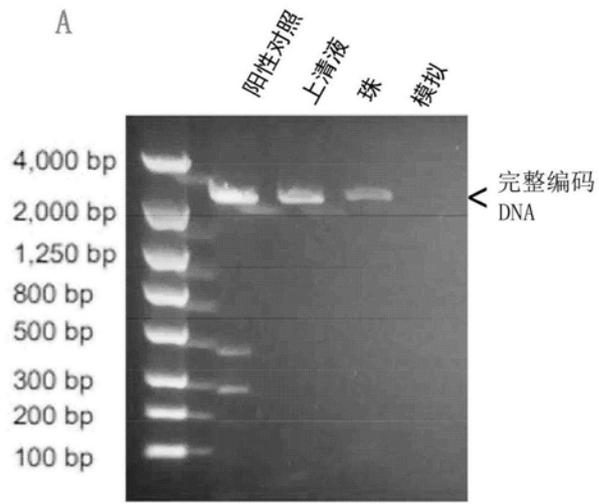


图17A

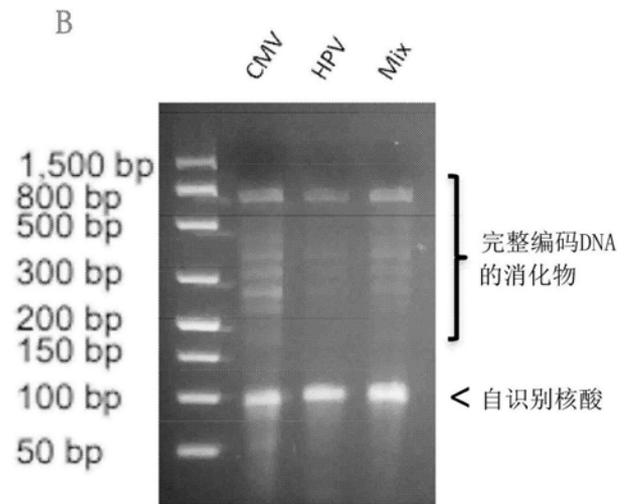


图17B

图17

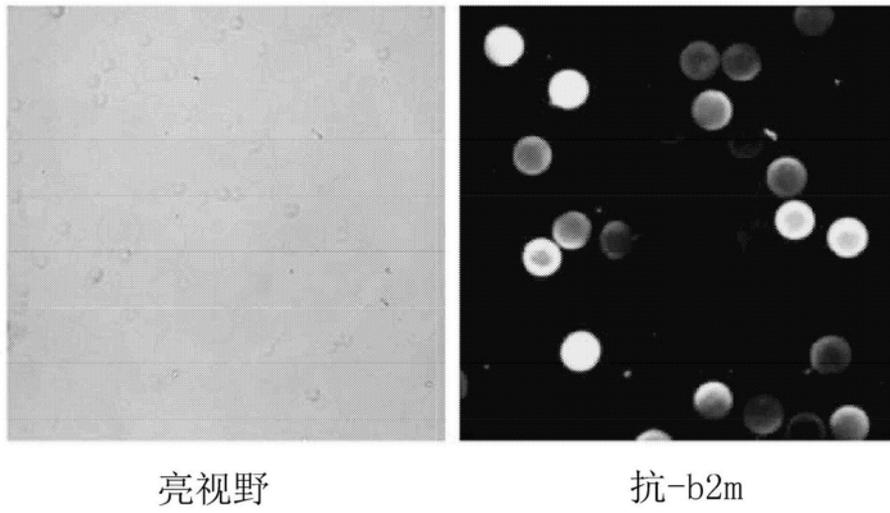


图 18A

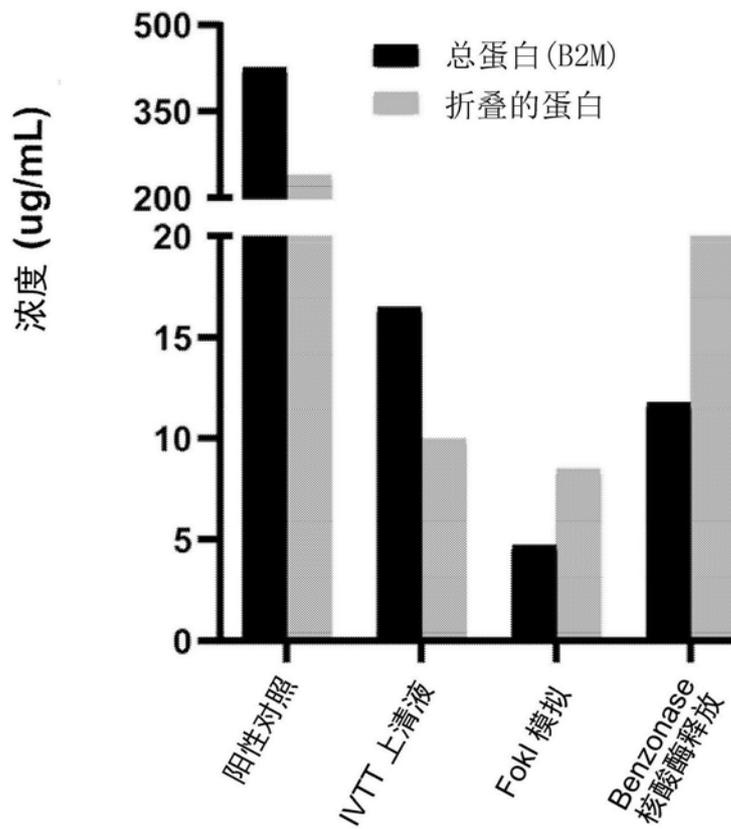


图 18B

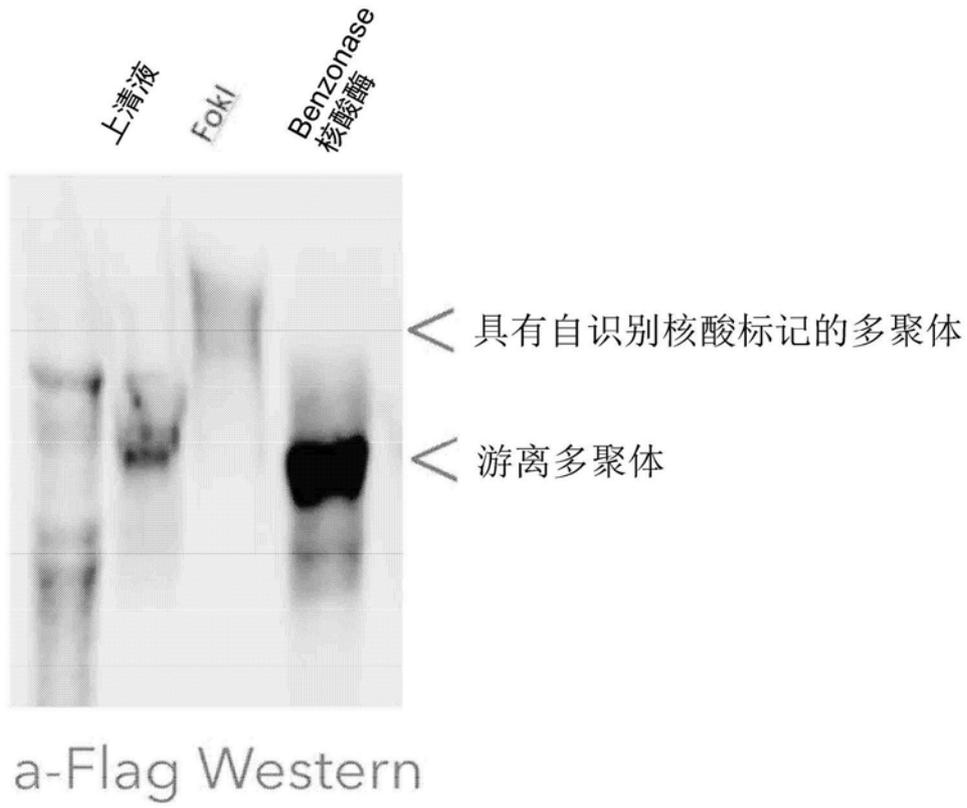


图 18C

图18

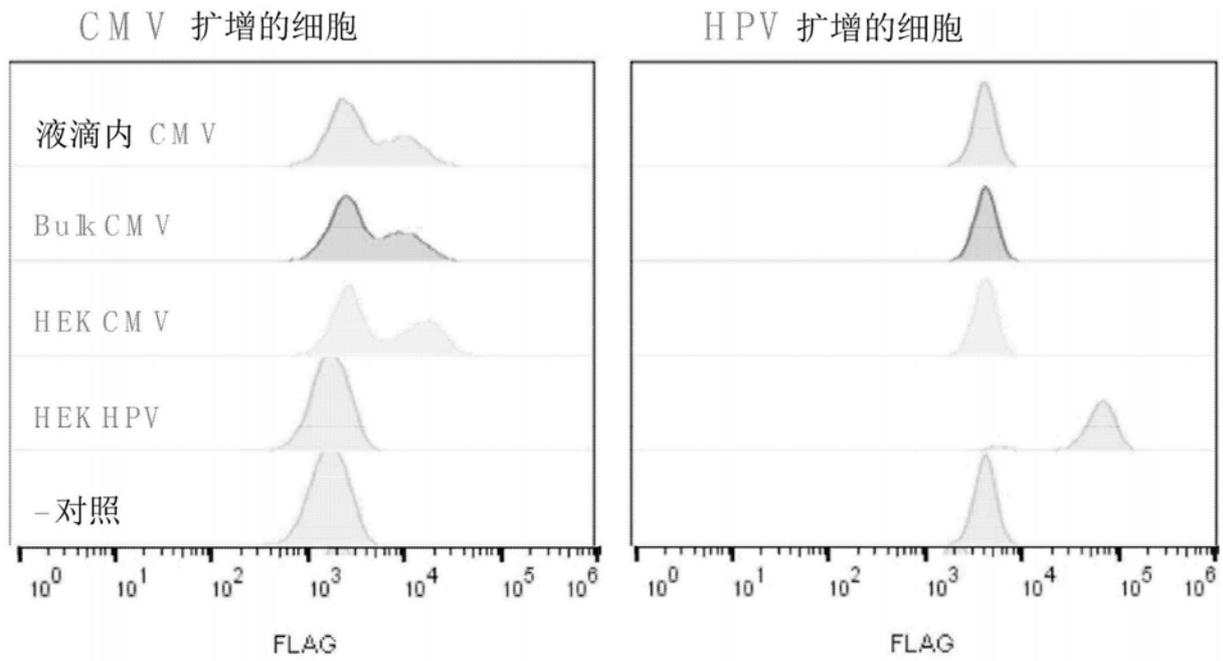


图19

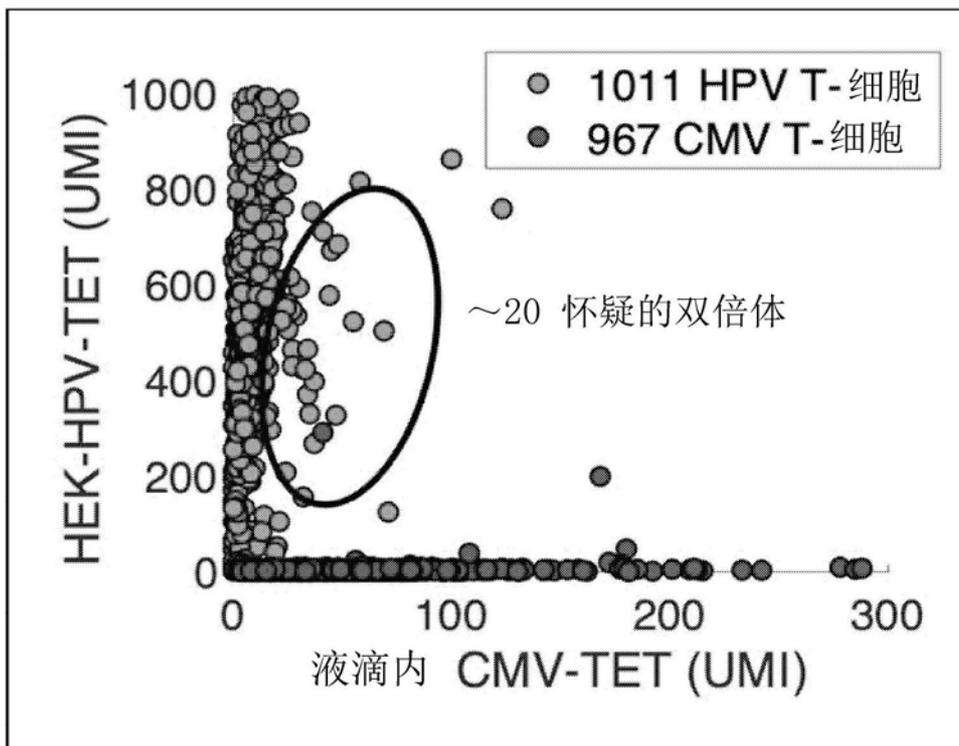


图20