



## (12)发明专利申请

(10)申请公布号 CN 111148849 A

(43)申请公布日 2020.05.12

(21)申请号 201880049800.9

(74)专利代理机构 北京坤瑞律师事务所 11494

(22)申请日 2018.05.25

代理人 封新琴

(30)优先权数据

(51)Int.Cl.

62/511,949 2017.05.26 US

C12Q 1/6855(2006.01)

(85)PCT国际申请进入国家阶段日

2020.01.23

(86)PCT国际申请的申请数据

PCT/US2018/034768 2018.05.25

(87)PCT国际申请的公布数据

W02018/218222 EN 2018.11.29

(71)申请人 阿布维托有限责任公司

地址 美国特拉华州

(72)发明人 S·J·戈德弗莱斯

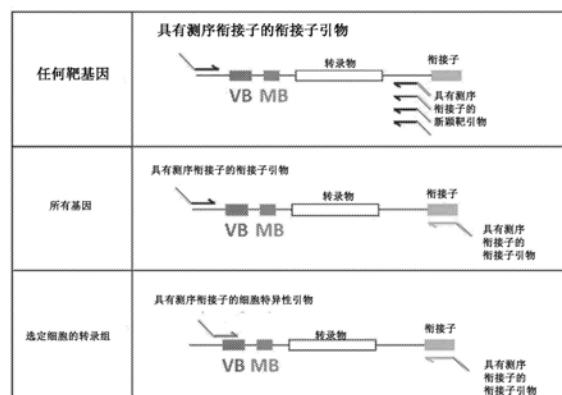
A·W·布里格斯 R·查里 Y·蒋  
R·豪斯 F·维尼奥特权利要求书13页 说明书119页  
序列表33页 附图5页

(54)发明名称

高通量多核苷酸文库测序和转录组分析

(57)摘要

本文提供了结合单细胞中基因表达的分析进行靶基因测序和单细胞条码化的方法。在一些实施方案中，所述靶基因是免疫分子，如抗体或TCR。在一些实施方案中，所述方法可以用于进行转录组测序，例如RNA测序，以捕获与完整受体免疫受体序列配对的单细胞的转录组，使得可以确定关于细胞的免疫库和转录组的信息。还提供了用于在进行转录组分析和免疫分子，例如抗体或TCR，测序中使用的多核苷酸文库。



1. 一种产生多核苷酸文库的方法,所述方法包括在与第一衔接子相对的末端处或附近向多个单链条码化多核苷酸中的每一个添加第二衔接子,所述第一衔接子与所述单链条码化多核苷酸中的每一个附接,其中所述第一衔接子包含容器条码并且所述多个单链条码化多核苷酸包含:

(i) 一个或多个靶单链多核苷酸,所述一个或多个靶单链多核苷酸包含在细胞群的细胞中存在的一个或多个靶多核苷酸或其一个或多个互补体的扩增子;以及

(ii) 单链多核苷酸的集合,所述单链多核苷酸各自包含所述细胞群的细胞中的多核苷酸或其互补体的扩增子;并且

其中来自所述细胞群的同一细胞的来自 (i) 和 (ii) 的多核苷酸中的每一个包含相同的容器条码序列。

2. 根据权利要求1所述的方法,其中所述多个单链条码化多核苷酸包含所述细胞群中的多个细胞的(i)和(ii)的多核苷酸。

3. 根据权利要求1或权利要求2所述的方法,其中所述多个单链条码化多核苷酸中的每一个还包含对于每个单链多核苷酸或其扩增产物而言独特的分子条码。

4. 根据权利要求1-3中任一项所述的方法,其中来自所述细胞群的每个细胞的单链条码化多核苷酸的所述集合共同包含转录组或部分转录组的互补DNA (cDNA) 链。

5. 根据权利要求4所述的方法,其中所述转录组或所述部分转录组共同包含在所述细胞的基因组中存在的至少60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的转录物。

6. 根据权利要求1-5中任一项所述的方法,其中所述单链条码化多核苷酸中的每一个具有大于或等于约50个碱基对、100个碱基对或200个碱基对的大小。

7. 根据权利要求1-6中任一项所述的方法,其中所述单链条码化多核苷酸中的每一个具有从或从约50个碱基对 (bp) 至1500bp、50bp至1250bp、50bp至1000bp、50bp至750bp、50bp至500bp、100bp至1500bp、100bp至1250bp、100bp至1000bp、100bp至750bp、100bp至500bp、200bp至1500bp、200bp至1250bp、200bp至1000bp、200bp至750bp或250bp至500bp的大小。

8. 根据权利要求1-7中任一项所述的方法,其中添加所述第二衔接子在包含所述多个单链条码化多核苷酸的均质混合物中进行。

9. 一种产生多核苷酸文库的方法,所述方法包括:

(a) 裂解多个容器中的每一个内的细胞,其中所述容器中的每一个包含来自细胞群的细胞;

(b) 在每个容器中产生多个互补多核苷酸,所述产生包括 (i) 使用一个或多个靶特异性引物产生与所述细胞中存在的一个或多个靶多核苷酸互补的一个或多个靶多核苷酸;并且 (ii) 产生多核苷酸的集合,所述多核苷酸中的每一个与所述细胞中的多核苷酸转录物互补。

10. 根据权利要求9所述的方法,其中 (ii) 中的多核苷酸的所述集合是使用随机寡聚引物产生的。

11. 根据权利要求9或权利要求10所述的方法,其中所述容器中的每一个还包含多个分子条码化寡核苷酸、一个容器条码化寡核苷酸或容器条码化寡核苷酸池以及第一衔接子或第一衔接子池,并且所述方法还包括:

(c) 将所述多个分子条码化寡核苷酸之一附接至多个互补多核苷酸、任选地所述多个互补多核苷酸中的每一个,从而产生各自包含分子条码的多个分子条码化多核苷酸,任选地其中所述分子条码化多核苷酸中的每一个的所述分子条码与在所述多个分子条码化多核苷酸中的其他分子条码化多核苷酸所包含的分子条码不同和/或是独特的分子条码;

(d) 将所述一个容器条码化寡核苷酸或容器条码化寡核苷酸池之一或其扩增产物以及所述第一衔接子或所述第一衔接子池中的一个第一衔接子或其扩增产物附接至所述分子条码化多核苷酸中的多个、任选地所述分子条码化多核苷酸中的每一个,从而产生多个双条码化多核苷酸、任选地单链双条码化多核苷酸,每个双条码化多核苷酸包含分子条码和容器条码,其中在同一容器中的所述双条码化多核苷酸中的每一个包含相同的容器条码。

12. 根据权利要求11所述的方法,其还包括 (e) 产生所述多个双条码化多核苷酸中的多个、任选地所述多个双条码化多核苷酸中的每一个的单链扩增子。

13. 根据权利要求12所述的方法,其还包括 (f) 向所述单链扩增子中的每一个添加第二衔接子,其中所述第一衔接子和所述第二衔接子存在于所述双条码化单链多核苷酸中的每一个的相对末端处或附近。

14. 一种产生多核苷酸文库的方法,所述方法包括:

(a) 裂解多个容器中的每一个内的细胞,其中所述容器中的每一个包含来自包含细胞群的样品的细胞、多个分子条码化寡核苷酸、一个容器条码化寡核苷酸或容器条码化寡核苷酸池以及第一衔接子或第一衔接子池;

(b) 在每个容器中产生多个互补多核苷酸,所述多个互补多核苷酸的所述产生包括 (i) 产生与所述细胞中存在的一个或多个靶多核苷酸转录物互补的一个或多个靶多核苷酸;并且 (ii) 产生多核苷酸的集合,所述多核苷酸各自单独与所述细胞中的多核苷酸转录物互补;

(c) 将所述多个分子条码化寡核苷酸之一附接至每个互补多核苷酸,从而产生各自包含独特分子条码的多个条码化多核苷酸;

(d) 将所述一个容器条码化寡核苷酸或容器条码化寡核苷酸池之一或其扩增产物以及所述第一衔接子或所述第一衔接子池中的一个第一衔接子或其扩增产物附接至所述条码化多核苷酸中的每一个,从而产生多个双条码化多核苷酸、任选地单链双条码化多核苷酸,其中所述双条码化多核苷酸中的每一个包含分子条码和容器条码,并且在同一容器中的所述双条码化多核苷酸中的每一个包含相同的容器条码;

(e) 产生所述多个双条码化多核苷酸中的每一个的单链扩增子;以及

(f) 向所述单链扩增子中的每一个添加第二衔接子,从而将所述第二衔接子添加到双条码化单链多核苷酸中,其中所述第一衔接子和所述第二衔接子存在于所述双条码化单链多核苷酸中的每一个的相对末端处或附近。

15. 根据权利要求11-14中任一项所述的方法,其中所述第一衔接子包含所述容器条码化寡核苷酸。

16. 根据权利要求9-15中任一项所述的方法,其中来自所述细胞群的每个细胞的多核苷酸的所述集合共同包含与细胞的转录组或部分转录组的转录物互补的序列。

17. 根据权利要求16所述的方法,其中所述转录组或所述部分转录组包含在所述细胞的基因组中存在的至少60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或

100%的转录物。

18. 根据权利要求1-17中任一项所述的方法,其中衍生出(i)中的扩增子的来自(i)的所述一个或多个靶多核苷酸和/或衍生出(ii)中的扩增子的来自(ii)的所述多核苷酸是DNA。

19. 根据权利要求1-18中任一项所述的方法,其中衍生出(i)中的扩增子的来自(i)的所述一个或多个靶多核苷酸和/或衍生出(ii)中的扩增子的来自(ii)的所述多核苷酸是RNA。

20. 根据权利要求19所述的方法,其中所述RNA是mRNA。

21. 根据权利要求9-20中任一项所述的方法,其中(b)的所述互补多核苷酸中的每一个或者一个或多个是cDNA。

22. 根据权利要求1-21中任一项所述的方法,其中所述一个或多个条码化单链多核苷酸中的每一个或者一个或多个是cDNA的链。

23. 根据权利要求1-22中任一项所述的方法,其中所述第一衔接子和/或所述第二衔接子包含至少一个通用引发位点。

24. 根据权利要求1-23中任一项所述的方法,其中:

所述第一衔接子和所述第二衔接子是不同的;和/或

所述第一衔接子包含第一通用引发位点,并且所述第二衔接子包含第二通用引发位点,任选地其中所述第一通用引发位点和所述第二通用引发位点是不同的。

25. 根据权利要求24所述的方法,其中所述第一通用引发位点和/或所述第二通用引发位点是或包含P7引发位点(C7)或其连续部分或者P5引发位点(C5)或其连续部分,任选地其中所述其连续部分足够与互补序列退火。

26. 根据权利要求24或权利要求25所述的方法,其中所述第一通用引发位点是或包含所述P7引发位点(C7)或其连续部分,并且所述第二通用引发位点是或包含所述P5引发位点(C5)或其连续部分。

27. 根据权利要求25或权利要求26所述的方法,其中所述P7引发位点(C7)包含序列AGATCGGAAGAGCACACGTCTGAACCTCCA (SEQ ID NO:77),或者是其连续部分。

28. 根据权利要求25或权利要求26所述的方法,其中所述P5引发位点包含序列

AGATCGGAAGAGCGCTGTAGGGAAAGAGTGTAGATCTCGGTGGTCCGTATCATT (SEQ ID NO:78),或者是其连续部分。

29. 根据权利要求25-28中任一项所述的方法,其中所述连续部分在长度上包含至少或至少约15个、20个、25个或30个核苷酸。

30. 根据权利要求25、26、28或29中任一项所述的方法,其中所述P5引发位点是如SEQ ID NO:25 (AGATCGGAAGAGCGCTGT) 所示的连续部分。

31. 根据权利要求1-8和13-30中任一项所述的方法,其中添加所述第二衔接子包括在包含第二通用引发位点的寡核苷酸的存在下使夹板寡核苷酸与所述条码化单链多核苷酸中的每一个杂交,其中所述夹板寡核苷酸包含(i)与所述第二通用引发位点互补的序列和(ii)能够与所述条码化单链多核苷酸的3'末端随机退火的简并悬突序列。

32. 根据权利要求31所述的方法,其中,在所述杂交之前,使所述夹板寡核苷酸和包含所述第二通用引发位点的寡核苷酸退火以形成夹板-衔接子双链体。

33. 根据权利要求31或权利要求32所述的方法,其中所述简并悬突序列包含序列(N)<sub>3-12</sub>,其中N是任何核苷酸。

34. 根据权利要求31-33中任一项所述的方法,其中所述简并悬突序列包含序列NNNNNN,其中N是任何核苷酸(SEQ ID NO:24)。

35. 根据权利要求31-34中任一项所述的方法,其中所述夹板寡核苷酸包含序列ACACGACGCTTCCGATCTNNNNNN,其中N是任何氨基酸(SEQ ID NO:26)。

36. 根据权利要求31-35中任一项所述的方法,其中包含所述第二通用引发位点的寡核苷酸包含序列AGATCGGAAGAGCGTCGTGT(SEQ ID NO:25)。

37. 根据权利要求1-8和11-36中任一项所述的方法,其中所述容器条码化寡核苷酸包含至少或约至少3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、30个、40个或50个核苷酸。

38. 根据权利要求1-8和11-37中任一项所述的方法,其中所述容器条码化寡核苷酸包含从或从约10个至30个核苷酸。

39. 根据权利要求1-8和11-38中任一项所述的方法,其中所述容器条码化寡核苷酸包含简并序列。

40. 根据权利要求1-8和11-39中任一项所述的方法,其中所述容器条码化寡核苷酸包含序列(N)<sub>14-17</sub>,其中N是任何核苷酸,任选地其中所述序列中的至少一个或两个N是W,其中W是腺嘌呤或胸腺嘧啶。

41. 根据权利要求1-8和11-40中任一项所述的方法,其中所述容器条码化寡核苷酸包含序列NNNNWNNNNWNNNN(SEQ ID NO:80)、WNNNNWNNNNWNNNN(SEQ ID NO:81)、NWNNNWNNNNWNNNN(SEQ ID NO:82)或NNWNNNNWNNNNWNNNN(SEQ ID NO:83),其中N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶。

42. 根据权利要求11-41中任一项所述的方法,其中每个容器包含第一衔接子池,所述第一衔接子池包含容器条码化寡核苷酸池,其中与所述池中的其他容器条码化寡核苷酸中的至少一个相比,所述第一衔接子池的每个容器条码化寡核苷酸包含至少一个碱基位移或碱基添加。

43. 根据权利要求42所述的方法,其中所述第一衔接子池的所述容器条码化寡核苷酸包含序列NNNNWNNNNWNNNN(SEQ ID NO:80)、WNNNNWNNNNWNNNN(SEQ ID NO:81)、NWNNNWNNNNWNNNN(SEQ ID NO:82)和NNWNNNNWNNNNWNNNN(SEQ ID NO:83),其中N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶。

44. 根据权利要求11-43中任一项所述的方法,其中在步骤(d)中,所述方法还包括扩增:

所述一个容器条码化寡核苷酸或所述容器条码化寡核苷酸池;或者

一个所述第一衔接子或所述第一衔接子的池,其中所述第一衔接子包含所述一个容器条码化寡核苷酸或所述容器条码化寡核苷酸池,其中:

所述扩增是在附接所述容器条码化寡核苷酸之前或与之同时进行的。

45. 根据权利要求11-44中任一项所述的方法,其中附接所述容器条码化寡核苷酸包括使所述容器条码化寡核苷酸的区域与所述互补多核苷酸中的每一个的区域或与所述分子条码化多核苷酸中的每一个的区域杂交。

46. 根据权利要求45所述的方法,其中所述区域包含与所述分子条码化多核苷酸的分子条码的3'末端区域互补的3'标记多核苷酸。

47. 根据权利要求9-46中任一项所述的方法,其中在步骤(b)中:

所述一个或多个靶多核苷酸通过在逆转录酶和一个或多个靶特异性引物的存在下将所述一个或多个靶多核苷酸逆转录来产生,所述一个或多个靶特异性引物与所述一个或多个靶多核苷酸的靶序列互补;和/或

多核苷酸的所述集合通过在逆转录酶和一个或多个转录组引物的存在下将所述细胞中的多核苷酸转录物逆转录来产生,所述一个或多个转录组引物与所述细胞中的多核苷酸转录物互补。

48. 根据权利要求1-47中任一项所述的方法,其中所述一个或多个靶多核苷酸包含免疫分子或其链的多核苷酸。

49. 根据权利要求1-48中任一项所述的方法,其中所述一个或多个靶多核苷酸包含至少两个靶多核苷酸,所述至少两个靶多核苷酸各自包含免疫分子链的多核苷酸。

50. 根据权利要求1-49中任一项所述的方法,其中所述一个或多个靶多核苷酸包含TCR或其链的多核苷酸。

51. 根据权利要求1-50中任一项所述的方法,其中所述一个或多个靶多核苷酸包含T细胞受体 $\alpha$  (TCR $\alpha$ ) 的第一多核苷酸和T细胞受体 (TCR $\beta$ ) 的第二多核苷酸。

52. 根据权利要求1-50中任一项所述的方法,其中所述一个或多个靶多核苷酸包含T细胞受体  $\gamma$  (TCR  $\gamma$ ) 的第一多核苷酸和T细胞受体 $\delta$  (TCR $\delta$ ; TCR $\delta$ ) 的第二多核苷酸。

53. 根据权利要求1-49中任一项所述的方法,其中所述一个或多个靶多核苷酸包含抗体或其链的多核苷酸。

54. 根据权利要求1-49和53中任一项所述的方法,其中所述一个或多个靶多核苷酸包含重链免疫球蛋白 (IgH) 多核苷酸的第一多核苷酸和轻链免疫球蛋白 (IgL) 多核苷酸的第二多核苷酸。

55. 根据权利要求47-54中任一项所述的方法,其中所述一个或多个靶特异性引物和/或所述一个或多个转录组引物包含聚 (T) 序列。

56. 根据权利要求47-54中任一项所述的方法,其中所述一个或多个转录组引物包含随机六聚体寡核苷酸引物的混合物。

57. 根据权利要求47-56中任一项所述的方法,其中所述一个或多个靶特异性引物包含与所述靶多核苷酸的一个或多个靶序列的一个或多个序列互补的一个或多个引物。

58. 根据权利要求57所述的方法,其中所述一个或多个靶特异性引物包含至少第一引物和第二引物。

59. 根据权利要求57或权利要求58所述的方法,其中所述一个或多个靶特异性引物包含针对多个靶多核苷酸的靶序列的引物,所述多个靶多核苷酸各自编码免疫分子或其链。

60. 根据权利要求59所述的方法,其中所述免疫分子是T细胞受体或抗体。

61. 根据权利要求58-60中任一项所述的方法,其中至少所述第一引物与免疫分子的第一链的多核苷酸的靶序列互补,并且所述第二引物与所述免疫分子的第二链的多核苷酸的靶序列互补。

62. 根据权利要求58-61中任一项所述的方法,其中所述第一引物和所述第二引物与

TCR的不同TCR链多核苷酸的靶序列互补。

63.根据权利要求58-62中任一项所述的方法,其中:

所述第一引物与TCR $\alpha$ 多核苷酸序列的靶序列互补,并且所述第二引物与TCR $\beta$ 多核苷酸序列的靶序列互补;或者

所述第一引物与TCR  $\gamma$  多核苷酸序列的靶序列互补,并且所述第二引物与TCR $\delta$ 多核苷酸序列的靶序列互补。

64.根据权利要求62或权利要求63所述的方法,其中所述TCR链多核苷酸的靶序列是恒定区序列。

65.根据权利要求58-64中任一项所述的方法,其中:

所述第一引物与TCR $\alpha$ 恒定区多核苷酸序列的靶序列互补,并且所述第二引物与TCR $\beta$ 恒定区多核苷酸序列的靶序列互补;或者

所述第一引物与TCR  $\gamma$  恒定区多核苷酸序列的靶序列互补,并且所述第二引物与TCR $\delta$ 恒定区多核苷酸序列的靶序列互补。

66.根据权利要求58-61中任一项所述的方法,其中至少所述第一引物和所述第二引物与抗体的不同抗体链多核苷酸的靶序列互补。

67.根据权利要求58-61和66中任一项所述的方法,其中所述第一引物与重链免疫球蛋白(IgH)多核苷酸序列的靶序列互补,并且所述第二引物与轻链免疫球蛋白(IgL)多核苷酸序列的靶序列互补。

68.根据权利要求66或权利要求67所述的方法,其中所述抗体链多核苷酸的靶序列是恒定区序列。

69.根据权利要求58-61和66-68中任一项所述的方法,其中所述第一引物与重链恒定区(CH)多核苷酸序列的靶序列互补,并且所述第二引物与轻链恒定区(CL)多核苷酸序列的靶序列互补。

70.根据权利要求68或权利要求69所述的方法,其中:

所述C<sub>H</sub>多核苷酸的靶序列来自IgM、IgD、IgA、IgE或IgG或其组合;和/或

所述C<sub>L</sub>多核苷酸序列的靶序列来自Ig $\kappa$ 、Ig $\lambda$ 或其组合。

71.根据权利要求1-70中任一项所述的方法,其中所述一个或多个靶多核苷酸包含全长编码序列。

72.根据权利要求9-71中任一项所述的方法,其中所述一个或多个靶多核苷酸和多核苷酸的所述集合在所述容器中以相同的反应体积产生。

73.根据权利要求9-72中任一项所述的方法,其中,在步骤(b)中,产生所述多个互补多核苷酸包括使用非模板末端转移酶,其中将三个或更多个非模板核苷酸、核糖核苷酸或其类似物添加到每个产生的互补多核苷酸的3'末端。

74.根据权利要求73所述的方法,其中所述非模板末端转移酶是逆转录酶或聚合酶。

75.根据权利要求73或权利要求74所述的方法,其中所述非模板末端转移酶是逆转录酶,并且其中所述逆转录酶选自Superscript II逆转录酶、Maxima逆转录酶、Protoscript II逆转录酶、莫洛尼鼠白血病病毒逆转录酶(MMLV-RT)、HighScriber逆转录酶、禽成髓细胞瘤病毒(AMV)逆转录酶、包含末端脱氧核苷酸转移酶活性的任何逆转录酶及其组合。

76.根据权利要求11-75中任一项所述的方法,其中,在步骤(c)中,所述附接包括使所

述多个分子条码化寡核苷酸之一的区域与所述互补多核苷酸中的每一个的所述三个或更多个非模板核苷酸杂交。

77. 根据权利要求76中任一项所述的方法,其中提供所述多个分子条码化寡核苷酸作为多个模板转换寡核苷酸,所述多个模板转换寡核苷酸各自包含与所述三个或更多个非模板核苷酸互补的3'部分。

78. 根据权利要求77所述的方法,其中所述模板转换寡核苷酸还包含与所述第一衔接子的一部分互补的5'末端区域,其中所述第一衔接子包含所述容器条码化寡核苷酸。

79. 根据权利要求47-78中任一项所述的方法,其中:

所述逆转录酶具有模板转换活性;

所述多个产生的互补多核苷酸的至少一些链包含含有三个或更多个非模板核苷酸的3'悬突;

提供所述多个分子条码化寡核苷酸作为多个模板转换寡核苷酸,所述多个模板转换寡核苷酸各自包含(1)与包含所述第一衔接子和所述容器条码化寡核苷酸的3'标记寡核苷酸互补的5'末端区域、(2)所述分子条码和(3)与所述3'悬突的三个或更多个非模板核苷酸互补的3'部分;并且所述模板转换寡核苷酸用作所述逆转录酶的模板,使得将所述分子条码掺入每个互补多核苷酸中以产生所述分子条码化多核苷酸。

80. 根据权利要求77-79中任一项所述的方法,其中与所述三个或更多个非模板核苷酸互补的3'部分包含核苷酸、核糖核苷酸或其类似物。

81. 根据权利要求73-80中任一项所述的方法,其中所述三个或更多个非模板核苷酸包含三个或更多个C核苷酸,并且与三个或更多个非模板核苷酸互补的所述3'部分包含一个或多个G核苷酸或核糖核苷酸或其类似物。

82. 根据权利要求73-81中任一项所述的方法,其中所述模板转换寡核苷酸还包含3'修饰的核苷酸,其阻断通过逆转录酶或DNA聚合酶延伸所述模板转换寡核苷酸。

83. 根据权利要求82所述的方法,其中所述修饰是3'末端核苷酸的脱氧、磷酸、氨基或烷基修饰。

84. 根据权利要求11-83中任一项所述的方法,其中步骤(d)还包括在所述附接以产生所述多个双条码化多核苷酸之后延伸所述多个分子条码化多核苷酸中的每一个。

85. 根据权利要求1-84中任一项所述的方法,其中所述容器是孔、乳剂、微滴或微胶囊。

86. 根据权利要求12-85中任一项所述的方法,其包括在步骤(e)之前组合所述多个容器中的两个或更多个的内容物,从而产生均质混合物,所述均质混合物包含所述多个单链双条码化多核苷酸中的两个或更多个。

87. 根据权利要求86所述的方法,其中组合所述多个容器的内容物包括破坏所述多个容器中的两个或更多个,并且将来自所述两个或更多个破坏的容器的单链双条码化多核苷酸合并。

88. 根据权利要求86或权利要求87所述的方法,其包括在步骤(e)之前选择或纯化具有大于或大于约50个碱基对、大于100个碱基对或大于200个碱基对的大小的单链双条码化多核苷酸。

89. 根据权利要求86-88中任一项所述的方法,其包括在步骤(e)之前选择或纯化具有以下大小的单链双条码化多核苷酸:从或从约50个碱基对(bp)至1500bp、50bp至1250bp、

50bp至1000bp、50bp至750bp、50bp至500bp、100bp至1500bp、100bp至1250bp、100bp至1000bp、100bp至750bp、100bp至500bp、200bp至1500bp、200bp至1250bp、200bp至1000bp、200bp至750bp或250bp至500bp。

90. 根据权利要求1-89中任一项所述的方法,其中所述单链双条码化多核苷酸按顺序(5'至3')包含:所述第一衔接子、所述容器条码、所述分子条码和所述第二衔接子。

91. 根据权利要求1-90中任一项所述的方法,其中所述第一衔接子位于所述单链条码化多核苷酸、任选地单链双条码化多核苷酸的5'区域处或附近。

92. 根据权利要求1-91中任一项所述的方法,其中所述第二衔接子位于所述单链条码化多核苷酸、任选地单链双条码化多核苷酸的3'区域处或附近。

93. 根据权利要求13-92中任一项所述的方法,其中步骤(a)-(f)中的一个或多个是在溶液中进行和/或不在固体支持物的存在下进行,任选地其中所述固体支持物是或包含珠。

94. 根据权利要求11-93中任一项所述的方法,其中至少步骤(c)和(d)是在溶液中进行和/或不在固体支持物的存在下进行,任选地其中所述固体支持物是珠。

95. 根据权利要求12-94中任一项所述的方法,其中步骤(a)-(e)中的每一个是在溶液中进行和/或不在固体支持物的存在下进行,任选地其中所述固体支持物是珠。

96. 根据权利要求1-95中任一项所述的方法,其中所述细胞群包含至少或约至少 $1 \times 10^3$ 个、 $5 \times 10^3$ 个、 $1 \times 10^4$ 个、 $5 \times 10^4$ 个、 $1 \times 10^5$ 个、 $5 \times 10^5$ 个、 $1 \times 10^6$ 个或 $5 \times 10^6$ 个细胞。

97. 根据权利要求1-96中任一项所述的方法,其中所述细胞群来自受试者的生物样品。

98. 根据权利要求97所述的方法,其中所述生物样品是或包含全血样品、血沉棕黄层样品、外周血单核细胞(PBMC)样品、未分级T细胞样品、淋巴细胞样品、白细胞样品、单采术产物或白细胞单采术产物。

99. 根据权利要求1-98中任一项所述的方法,其中所述细胞群包含免疫细胞。

100. 根据权利要求99所述的方法,其中所述免疫细胞包含淋巴细胞或抗原呈递细胞。

101. 根据权利要求99或权利要求100所述的方法,其中所述免疫细胞是淋巴细胞或其亚型、B细胞或其亚型、T细胞或其亚型、或其组合。

102. 根据权利要求101所述的方法,其中所述免疫细胞是T细胞,所述T细胞是CD4+和/或CD8+T细胞。

103. 根据权利要求1-102中任一项所述的方法,其中所述细胞群富集或包含中枢记忆T细胞、效应记忆T细胞、幼稚T细胞、干细胞样中枢记忆T细胞、效应T细胞和调节性T细胞。

104. 根据权利要求1-101中任一项所述的方法,其中所述细胞群富集记忆B细胞、幼稚B细胞或浆母细胞B细胞。

105. 根据权利要求97-104中任一项所述的方法,其中所述受试者是人类受试者。

106. 根据权利要求97-105中任一项所述的方法,其中所述受试者患有癌症、感染或自身免疫病症。

107. 根据权利要求106所述的方法,其中所述感染是病毒、细菌或真菌感染。

108. 根据权利要求1-107中任一项所述的方法,其还包括扩增所述多个条码化单链多核苷酸,从而产生多个多核苷酸模板。

109. 根据权利要求108所述的方法,其中所述多个条码化单链多核苷酸的扩增是在第

一引物组的存在下进行,所述第一引物组包含与所述第一衔接子序列互补的第一引物和与所述第二衔接子序列互补的第二引物。

110. 根据权利要求109所述的方法,其中所述第一引物和/或所述第二引物是通用引物。

111. 根据权利要求110所述的方法,其中所述第一引物和/或所述第二引物与所述P7引发位点 (C7) 或其连续部分或者所述P5引发位点 (C5) 或其连续部分互补。

112. 根据权利要求110或权利要求111所述的方法,其中所述第一引物与所述P7引发位点 (C7) 或其连续部分互补,并且所述第二引物与所述P5引发位点 (C5) 或其连续部分互补。

113. 根据权利要求111或权利要求112所述的方法,其中:

与所述P7引发位点 (C7) 或其连续部分互补的引物具有或包含序列 CAAGCAGAACGGCATACGAGAT (SEQ ID NO:39); 和/或

与所述P5引发位点 (C5) 或其连续部分互补的引物包含序列 ACACGACGCTTCCGATCT (SEQ ID NO:27)。

114. 根据权利要求109-113中任一项所述的方法,其中所述第一引物和/或所述第二引物还包含测序衔接子。

115. 根据权利要求114所述的方法,其中:

与所述P7引发位点 (C7) 或其连续部分互补的引物还包含序列 CAAGCAGAACGGCATACGAGAT [NNNNNN] GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT (SEQ ID NO: 28); 和/或

与所述P5引发位点 (C5) 或其连续部分互补的引物包含序列 AATGATAACGGCGACCACCGAGA TCTACACTCTTCCCTACACGACGCTCTCC (SEQ ID NO:76)。

116. 根据权利要求1-115中任一项所述的方法,其还包括纯化所述单链条码化多核苷酸、任选地单链双条码化多核苷酸中的每一个。

117. 一种多核苷酸文库,其包含通过根据权利要求1-116中任一项所述的方法产生的多个条码化多核苷酸。

118. 一种多核苷酸文库,其包含多个条码化多核苷酸,其中所述多个条码化多核苷酸包含 (i) 一个或多个靶多核苷酸,所述一个或多个靶多核苷酸包含在细胞群的细胞中存在的一个或多个靶多核苷酸的扩增子; 以及

(ii) 多核苷酸的集合,所述多核苷酸各自包含所述细胞中的多核苷酸的扩增子,其中每个条码化多核苷酸包含:

包含与第一通用引物互补的第一通用引发位点的第一衔接子;

包含容器条码的容器条码化寡核苷酸,其中所述容器条码对于来自所述细胞群的同一细胞的来自 (i) 和 (ii) 的所有条码化多核苷酸而言是相同的; 以及

包含与第二通用引物互补的第二通用引发位点的第二衔接子序列。

119. 根据权利要求118所述的多核苷酸文库,其中所述多个条码化多核苷酸中的每一个包含分子条码,所述分子条码对于每个多核苷酸而言是独特的。

120. 根据权利要求118或权利要求119所述的多核苷酸文库,其中来自所述细胞群的每个细胞的条码化多核苷酸模板的所述集合共同包含与转录组或部分转录组对应的互补DNA (cDNA) 链。

121. 根据权利要求118-120中任一项所述的多核苷酸文库,其中所述转录组或所述部分转录组共同包含在所述细胞的基因组中存在的至少60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的转录物。

122. 根据权利要求118-121中任一项所述的多核苷酸文库,其中所述条码化多核苷酸模板中的每一个具有大于或大于约50个碱基对、大于100个碱基对或大于200个碱基对的大小。

123. 根据权利要求118-122中任一项所述的多核苷酸文库,其中所述条码化多核苷酸中的每一个具有从或从约50个碱基对 (bp) 至1500bp、50bp至1250bp、50bp至1000bp、50bp至750bp、50bp至500bp、100bp至1500bp、100bp至1250bp、100bp至1000bp、100bp至750bp、100bp至500bp、200bp至1500bp、200bp至1250bp、200bp至1000bp、200bp至750bp或250bp至500bp的大小。

124. 根据权利要求118-123中任一项所述的多核苷酸文库,其中所述第一衔接子包含所述容器条码。

125. 根据权利要求118-124中任一项所述的多核苷酸文库,其中所述条码化多核苷酸是单链的。

126. 根据权利要求118-125中任一项所述的多核苷酸文库,其中所述第一衔接子和所述第二衔接子是不同的。

127. 根据权利要求118-126中任一项所述的多核苷酸文库,其中所述第一通用引发位点和/或所述第二通用引发位点是或包含P7引发位点 (C7) 或其连续部分或者P5引发位点 (C5) 或其连续部分,任选地其中所述其连续部分足够与互补序列退火。

128. 根据权利要求118-127中任一项所述的多核苷酸文库,其中所述第一通用引发位点是或包含所述P7引发位点 (C7) 或其连续部分,并且所述第二通用引发位点是或包含所述P5引发位点 (C5) 或其连续部分。

129. 根据权利要求127或权利要求128所述的多核苷酸文库,其中所述P7引发位点 (C7) 包含序列AGATCGGAAGAGCACACGTCTGAACCTCCA (SEQ ID NO:77),或者是其连续部分。

130. 根据权利要求127-129中任一项所述的多核苷酸文库,其中所述P5引发位点包含序列

AGATCGGAAGAGCGTCGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT (SEQ ID NO:78),或者是其连续部分。

131. 根据权利要求127-130中任一项所述的多核苷酸文库,其中所述连续部分在长度上包含至少或至少约15个、20个、25个或30个核苷酸。

132. 根据权利要求127-131中任一项所述的多核苷酸文库,其中所述P5引发位点是如SEQ ID NO:25 (AGATCGGAAGAGCGTCGT) 所示的连续部分。

133. 根据权利要求118-132中任一项所述的多核苷酸文库,其中所述容器条码化寡核苷酸包含至少或约至少3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、30个、40个或50个核苷酸。

134. 根据权利要求118-133中任一项所述的多核苷酸文库,其中所述容器条码化寡核苷酸包含从或从约10个至30个核苷酸。

135. 根据权利要求118-134中任一项所述的多核苷酸文库,其中所述一个或多个靶多

核苷酸包含免疫分子或其链的多核苷酸。

136. 根据权利要求118-135中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含至少两个靶多核苷酸,所述至少两个靶多核苷酸各自包含免疫分子链的多核苷酸。

137. 根据权利要求118-136中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含TCR或其链的一个或多个多核苷酸。

138. 根据权利要求118-137中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含T细胞受体 $\alpha$  (TCR $\alpha$ )的第一多核苷酸和T细胞受体 (TCR $\beta$ )的第二多核苷酸。

139. 根据权利要求118-137中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含T细胞受体 $\gamma$  (TCR $\gamma$ )的第一多核苷酸和T细胞受体 $\delta$  (TCR $\delta$ )的第二多核苷酸。

140. 根据权利要求118-136中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含抗体或其链的一个或多个多核苷酸。

141. 根据权利要求118-136和140中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含重链免疫球蛋白 (IgH) 多核苷酸的第一多核苷酸和轻链免疫球蛋白 (IgL) 多核苷酸的第二多核苷酸。

142. 根据权利要求118-141中任一项所述的多核苷酸文库,其中所述条码化多核苷酸按顺序(5'至3')包含:所述第一衔接子、所述容器条码寡核苷酸、所述分子条码和所述第二衔接子。

143. 根据权利要求118-142中任一项所述的多核苷酸文库,其中所述第一衔接子位于所述双条码化单链多核苷酸的5'区域处或附近。

144. 根据权利要求118-143中任一项所述的多核苷酸文库,其中所述第二衔接子位于所述双条码化单链多核苷酸的3'区域处或附近。

145. 一种用于对一个或多个靶多核苷酸和/或一个或多个细胞的完整或部分转录组测序的方法,其包括对通过权利要求1-116中任一项产生或来自根据权利要求117-142中任一项所述的多核苷酸文库的所述多个条码化多核苷酸、任选地双条码化多核苷酸中的一个或多个进行测序。

146. 根据权利要求145所述的方法,其中对一个或多个细胞的所述完整或部分转录组进行测序。

147. 根据权利要求146所述的方法,其还包括在所述测序之前扩增所述完整转录组或其部分。

148. 根据权利要求147所述的方法,其中使用第一引物组进行扩增,所述第一引物组包含分别对所述第一衔接子序列和所述第二衔接子序列具有特异性的第一引物和第二引物。

149. 根据权利要求144-148中任一项所述的方法,其中对来自所述多个条码化多核苷酸的一个或多个靶多核苷酸进行测序。

150. 根据权利要求149所述的方法,其还包括在所述测序之前扩增来自所述多个条码化多核苷酸的一个或多个靶多核苷酸。

151. 根据权利要求150所述的方法,其中扩增所述一个或多个靶多核苷酸的一个或多个全长序列。

152. 根据权利要求150或权利要求151所述的方法,其中在第二引物组的存在下进行扩

增,所述第二引物组包含与一个或多个靶多核苷酸互补的一个或多个第一引物和与所述第一衔接子序列互补的第二引物。

153. 根据权利要求152所述的方法,其中所述第二引物组的第二引物与所述P7引发位点 (C7) 或其连续部分或者所述P5引发位点 (C5) 或其连续部分互补。

154. 根据权利要求152或权利要求153所述的方法,其中所述第二引物组的第二引物与所述P7引发位点 (C7) 或其连续部分互补。

155. 根据权利要求152-154中任一项所述的方法,其中所述第二引物组的第二引物具有或包含序列CAAGCAGAAGACGGCATACTGAGAT (SEQ ID NO:39) 或

CAAGCAGAAGACGGCATACTGAGAT [NNNNNN] GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT (SEQ ID NO:28)。

156. 根据权利要求152-155中任一项所述的方法,其中与所述一个或多个靶多核苷酸互补的一个或多个第一引物对免疫分子或其链的靶序列具有特异性。

157. 根据权利要求156所述的方法,其中所述免疫分子是T细胞受体或抗体。

158. 根据权利要求156或权利要求157所述的方法,其中所述一个或多个第一引物对所述免疫分子的恒定区的靶序列具有特异性。

159. 根据权利要求156-158中任一项所述的方法,其中所述免疫分子是TCR,并且所述一个或多个第一引物包含 AGTCTCTCAGCTGGTACACGG (SEQ ID NO:37)、ATGGCTCAAACACAGCGACCTC (SEQ ID NO:38) 或其组合。

160. 根据权利要求156-159中任一项所述的方法,其中所述免疫分子是抗体,并且所述一个或多个第一引物包含SEQ ID NO:29-36中的任何一个或其组合。

161. 根据权利要求145-160中任一项所述的方法,其还包括确定所述一个或多个条码化多核苷酸、任选地双条码化多核苷酸的细胞起源。

162. 根据权利要求161所述的方法,其中确定细胞起源包括鉴定双条码化多核苷酸的序列,所述双条码化多核苷酸因为来自同一细胞而具有相同的容器条码。

163. 根据权利要求161或权利要求162所述的方法,其中所述靶多核苷酸是包含第一多核苷酸链和第二多核苷酸链的免疫分子,并且所述方法包括通过在经测序的双条码化多核苷酸中存在相同的容器条码将所述第一多核苷酸链和所述第二多核苷酸链与同一细胞进行匹配。

164. 根据权利要求155-163中任一项所述的方法,其还包括定量或确定具有相同条码、任选地相同分子条码的多核苷酸的数量。

165. 根据权利要求155-164中任一项所述的方法,其中所述多个条码化多核苷酸是包含分子条码和容器条码的双条码化多核苷酸,并且所述方法还包括鉴定具有相同容器条码的转录组序列和靶多核苷酸序列,从而鉴定带有所述一个或多个靶多核苷酸的细胞的转录组信息。

166. 一种用于转录组分析的方法,所述方法包括:

(a) 对来自通过根据权利要求1-116中任一项所述的方法产生的所述多个条码化多核苷酸中的多个条码化多核苷酸的一个或多个靶多核苷酸或来自根据权利要求117-142中任一项所述的多核苷酸文库的所述多个条码化多核苷酸的一个或多个靶多核苷酸进行测序,其中所述条码化多核苷酸是包含分子条码和容器条码的双条码化多核苷酸,从而产生来自

所述多个细胞的靶多核苷酸的序列信息；

(b) 对来自通过根据权利要求1-116中任一项所述的方法产生的所述多个条码化多核苷酸或来自根据权利要求117-142中任一项所述的多核苷酸文库的所述多个条码化多核苷酸的全转录组或其部分进行测序，其中所述条码化多核苷酸是包含分子条码和容器条码的双条码化多核苷酸，从而产生来自所述多个细胞的转录组数据；以及

(c) 鉴定因为来自同一细胞而具有相同容器条码的来自 (a) 和来自 (b) 的序列信息。

167. 一种分析选定单细胞的转录组的方法，其包括：

(a) 对来自通过根据权利要求1-116中任一项所述的方法产生的所述多个条码化多核苷酸的一个或多个靶多核苷酸或来自根据权利要求117-142中任一项所述的多核苷酸文库的所述多个条码化多核苷酸中的多个条码化多核苷酸的一个或多个靶多核苷酸进行扩增和测序，其中所述条码化多核苷酸是包含分子条码和容器条码的双条码化多核苷酸，从而产生在所述多个细胞中的至少一个中的所述靶多核苷酸中的每一个的序列信息；

(b) 鉴定与在 (a) 中测序的所述靶多核苷酸之一相关的所述一个或多个容器条码，从而鉴定带有所述靶多核苷酸的选定单细胞；

(c) 对来自带有 (b) 中鉴定的容器条码的细胞的所述多个条码化多核苷酸的转录组或其部分进行扩增和测序，从而产生来自表达靶多肽的所述选定细胞的转录组数据。

168. 根据权利要求167所述的方法，其中使用对 (b) 中鉴定的容器条码具有特异性的引物和对所述条码化多核苷酸的第二衔接子序列具有特异性的引物对来自所述选定细胞的转录组或其部分扩增或测序。

169. 一种用于转录组分析的方法，其包括将所述转录组或其部分的序列信息与来自同一细胞的所述一个或多个靶多核苷酸中的至少一个进行匹配，其中所述序列信息是从通过根据权利要求1-114中任一项所述的方法产生的所述多个条码化多核苷酸或从根据权利要求117-142中任一项所述的多核苷酸文库的所述多个条码化多核苷酸确定，其中所述条码化多核苷酸是包含分子条码和容器条码的双条码化多核苷酸，或者是从根据权利要求155-165中任一项所述的方法确定。

170. 根据权利要求169所述的方法，其中具有相同容器条码的序列因为来自同一细胞而相匹配。

171. 根据权利要求166-170中任一项所述的方法，其中所述转录组数据包含与所述细胞的功能或活性相关的参数、特征、特性或表型。

172. 根据权利要求171所述的方法，其中所述转录组数据与所述细胞的激活、耗竭或增殖活性相关。

## 高通量多核苷酸文库测序和转录组分析

### 相关申请的交叉引用

本申请要求2017年5月26日提交的标题为“高通量多核苷酸文库测序和转录组分析 (HIGH-THROUGHPUT POLYNUCLEOTIDE LIBRARY SEQUENCING AND TRANSCRIPTOME ANALYSIS)”的美国临时申请号62/511,949的优先权,将其内容通过引用以其整体并入。

### 通过引用并入序列表

本申请是与电子格式的序列表一起提交的。所述序列表以2018年5月25日创建的标题为735042011740SeqList.txt的文件提供,其大小为444,416字节。将在电子格式的序列表中的信息通过引用以其整体并入。

### 技术领域

本公开文本涉及结合单细胞中基因表达的分析进行靶基因测序和单细胞条码化的方法。在一些实施方案中,所述靶基因是免疫分子,如抗体或TCR。在一些实施方案中,所述方法可以用于进行转录组测序(例如, RNA测序),以捕获与完整受体免疫受体序列配对的单细胞的转录组,使得可以确定关于细胞的免疫库和转录组的信息。本公开文本还涉及用于在进行转录组分析和免疫分子(例如,抗体或TCR)测序中使用的多核苷酸文库。

### 背景技术

细胞或组织的转录组含量的确定(即,“基因表达谱分析”)提供了对正常和患病细胞或组织进行功能分析的方法,包括提供关于细胞或组织的“状态”的表征信息或鉴定细胞亚群的特征。单细胞转录组测序的现有工具受限于其通量和/或不能捕获全长免疫受体序列。因此,需要改进的方法。提供了满足此类需求的方法和组合物。

### 发明内容

提供了从单个细胞或多个细胞产生多核苷酸文库的方法,其中产生了一个或多个全长条码化靶多核苷酸序列或其选定片段,同时共产生大量条码化多核苷酸序列,其集合基本上代表一个细胞或多个细胞的转录组或基因组。条码化允许分析和定量来自同一细胞的多核苷酸的表达或存在。还提供了通过本文的方法产生的多核苷酸文库。还提供了从单细胞或大量细胞进行转录组分析并将转录组分析与一个或多个靶序列的分析相结合的方法。

在一些实施方案中,所述方法包括产生多核苷酸文库,其包括在与第一衔接子相对的末端处或附近向多个条码化单链多核苷酸中的每一个添加第二衔接子,所述第一衔接子附接至所述条码化单链多核苷酸中的每一个,所述多个条码化单链多核苷酸含有: (i) 一个或多个靶单链多核苷酸,所述一个或多个靶单链多核苷酸包括在细胞群的细胞中存在的一个或多个靶多核苷酸或其一个或多个互补体的扩增子;以及 (ii) 单链多核苷酸的集合,所述单链多核苷酸各自含有所述细胞中的多核苷酸或其互补体的扩增子;并且其中所述多个条码化单链多核苷酸中的每一个包括容器条码,所述容器条码对于来自所述细胞群的同一细胞的来自 (i) 和 (ii) 的所有多核苷酸而言是相同的。

在一些实施方案中,所述多个条码化单链多核苷酸包括所述细胞群中多个细胞的(i)和(ii)的多核苷酸。在一些实施方案中,所述多个条码化单链多核苷酸还含有对于每个单链多核苷酸或其扩增产物而言独特的分子条码。

在一些实施方案中,来自所述细胞群的每个细胞的单链多核苷酸的所述集合共同含有转录组或部分转录组的互补DNA(cDNA)链。在一些实施方案中,所述转录组或所述部分转录组共同含有所述细胞的基因组中存在的至少60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的转录物。

在所述方法的一些实施方案中,所述条码化单链多核苷酸中的每一个具有大于或大于约50个碱基对、大于100个碱基对或大于200个碱基对的大小。在一些实施方案中,所述条码化单链多核苷酸中的每一个具有从或从约50个碱基对(bp)至1500bp、50bp至1250bp、50bp至1000bp、50bp至750bp、50bp至500bp、100bp至1500bp、100bp至1250bp、100bp至1000bp、100bp至750bp、100bp至500bp、200bp至1500bp、200bp至1250bp、200bp至1000bp、200bp至750bp或250bp至500bp的大小。

在所述方法的一些实施方案中,将所述第二衔接子添加到含有所述多个条码化单链多核苷酸的均质混合物中。

在一些实施方案中,所述第一衔接子含有所述容器条码。

在产生多核苷酸文库的方法的一些实施方案中,所述方法包括:(a)裂解多个容器中的每一个内的细胞,其中所述容器中的每一个含有来自含有细胞群的样品的细胞;(b)在每个容器中产生多个互补多核苷酸,所述多个互补多核苷酸的所述产生包括(i)使用一个或多个靶特异性引物产生与所述细胞中存在的一个或多个靶多核苷酸互补的一个或多个靶多核苷酸;并且(ii)产生多核苷酸的集合,所述多核苷酸中的每一个与所述细胞中的多核苷酸互补。在一些实施方案中,(ii)中的多核苷酸的所述集合是使用随机和/或简并和/或非特异性寡聚引物产生的。在一些实施方案中,(i)中的多核苷酸的所述集合是使用寡聚dT引物产生的。

在一些实施方案中,所述容器中的每一个还含有多个分子条码化寡核苷酸、容器条码化寡核苷酸池以及任选地第一衔接子,并且所述方法还包括:(c)将所述多个分子条码化寡核苷酸之一附接至多个互补多核苷酸、任选地所述多个互补多核苷酸中的每一个,从而产生各自含有分子条码的多个条码化多核苷酸,如分子条码化多核苷酸,任选地其中所述分子条码与在所述多个条码化多核苷酸内的其他条码化多核苷酸所包含的分子条码不同和/或是独特的分子条码;(d)将所述一个容器条码化寡核苷酸或容器条码化寡核苷酸池之一或其扩增产物附接至所述条码化多核苷酸中的多个、任选地所述条码化多核苷酸中的每一个,从而产生多个双条码化多核苷酸,其中在同一容器中的所述双条码化多核苷酸中的每一个含有相同的容器条码。在一些实施方案中,步骤(d)包括将所述一个容器条码化寡核苷酸或容器条码化寡核苷酸池之一或其扩增产物以及所述第一衔接子或所述第一衔接子池中的一个第一衔接子或其扩增产物附接至所述分子条码化多核苷酸中的多个、任选地所述分子条码化多核苷酸中的每一个,从而产生多个双条码化多核苷酸、任选地单链双条码化多核苷酸,每个双条码化多核苷酸包含分子条码和容器条码,其中在同一容器中的所述双条码化多核苷酸中的每一个包含相同的容器条码。

在一些实施方案中,所述方法还包括(e)产生所述多个双条码化多核苷酸中的多个、任

选地所述多个双条码化多核苷酸中的每一个的单链扩增子,和/或(f)向所述单链扩增子中的每一个添加第二衔接子,从而将所述衔接子添加到双条码化单链多核苷酸中,其中所述第一衔接子和所述第二衔接子存在于所述双条码化单链多核苷酸中的每一个的相对末端处或附近。

在一些实施方案中,产生多核苷酸文库的方法包括以下步骤: (a) 裂解多个容器中的每一个内的细胞,其中所述容器中的每一个含有来自含有细胞群的样品的细胞、多个分子条码化寡核苷酸和一个容器条码化寡核苷酸或容器条码化寡核苷酸池以及任选地第一衔接子或第一衔接子池; (b) 在每个容器中产生多个互补多核苷酸,所述多个互补多核苷酸的所述产生包括(i)产生与所述细胞中存在的一个或多个靶多核苷酸互补的一个或多个靶多核苷酸;并且(ii)产生多核苷酸的集合,所述多核苷酸各自单独与所述细胞中的多核苷酸互补; (c) 将所述多个分子条码化寡核苷酸之一附接至每个互补多核苷酸,从而产生各自含有独特分子条码的多个条码化多核苷酸; (d) 将所述一个容器条码化寡核苷酸或容器条码化寡核苷酸池之一或其扩增产物以及所述第一衔接子或所述第一衔接子池中的一个第一衔接子或其扩增产物附接至所述条码化多核苷酸中的每一个,从而产生多个双条码化多核苷酸、任选地单链双条码化多核苷酸,其中所述双条码化多核苷酸中的每一个包含分子条码和容器条码,并且在同一容器中的所述双条码化多核苷酸中的每一个包含相同的容器条码; (e) 产生所述多个双条码化多核苷酸中的每一个的单链扩增子;以及(f)向所述单链扩增子中的每一个添加第二衔接子,从而将所述第二衔接子添加到双条码化单链多核苷酸中,其中所述第一衔接子和所述第二衔接子存在于所述双条码化单链多核苷酸中的每一个的相对末端处或附近。

在一些实施方案中,所述衔接子(如所述第一衔接子)含有所述容器条码化寡核苷酸。在一些实施方案中,所述容器条码化寡核苷酸含有所述第一衔接子。在一些实施方案中,来自所述细胞群的每个细胞的多核苷酸的所述集合共同含有与细胞的转录组或部分转录组的转录物互补的序列。在一些实施方案中,所述转录组或所述部分转录组含有所述细胞的基因组中存在的至少60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的转录物。

在一些实施方案中,在所述细胞中的所述一个或多个靶多核苷酸和/或所述多核苷酸是DNA。在一些实施方案中,衍生出(i)中的扩增子的来自(i)的所述一个或多个靶多核苷酸和/或衍生出(ii)中的扩增子的来自(ii)的所述多核苷酸是DNA。在一些实施方案中,在所述细胞中的所述一个或多个靶多核苷酸和/或所述多核苷酸是RNA,如mRNA。在一些实施方案中,衍生出(i)中的扩增子的来自(i)的所述一个或多个靶多核苷酸和/或衍生出(ii)中的扩增子的来自(ii)的所述多核苷酸是RNA,如mRNA。

在一些实施方案中,(b)的所述互补多核苷酸中的每一个或者一个或多个是cDNA。在一些实施方案中,所述一个或多个条码化单链多核苷酸中的每一个或者一个或多个是cDNA的链。

在所述方法的一些实施方案中,所述第一衔接子和/或所述第二衔接子含有至少一个通用引发位点。在一些实施方案中,所述第一衔接子和所述第二衔接子是不同的;和/或所述第一衔接子含有第一通用引发位点,并且所述第二衔接子含有第二通用引发位点,任选地其中所述第一通用引发位点和所述第二通用引发位点是不同的。在一些实施方案中,所

述第一通用引发位点和/或所述第二通用引发位点是或含有P7引发位点 (C7) 或其连续部分或者P5引发位点或其连续部分,任选地其中所述其连续部分足够与互补序列退火。在一些实施方案中,所述第一通用引发位点是或含有所述P7引发位点 (C7) 或其连续部分,并且所述第二通用引发位点是或含有所述P5引发位点 (C5) 或其连续部分。在一些实施方案中,所述P7引发位点 (C7) 含有序列AGATCGGAAGAGCACACGTCTGAACCTCCA (SEQ ID NO:77),或者是其连续部分。在一些实施方案中,所述P5引发位点含有序列AGATCGGAAGAGCGTCGTAGGGAAAG AGTGTAGATCTCGGTGGTCGCC GTATCATT (SEQ ID NO:78),或者是其连续部分。在一些实施方案中,所述连续部分在长度上含有至少或至少约15个、20个、25个或30个核苷酸。在一些实施方案中,所述P5引发位点是如SEQ ID NO:25 (AGATCGGAAGAGCGTCGTGT) 所示的连续部分。

在一些实施方案中,添加所述第二衔接子包括在包括第二通用引发位点的寡核苷酸的存在下使夹板寡核苷酸与所述条码化单链多核苷酸中的每一个杂交,其中所述夹板寡核苷酸含有(i)与所述第二通用引发位点互补的序列和(ii)能够与所述条码化单链多核苷酸的3'末端随机退火的简并悬突序列。在一些情况下,在所述杂交之前,所述夹板寡核苷酸和含有所述第二通用引发位点的寡核苷酸退火以形成夹板-衔接子双链体。在一些方面,所述简并悬突序列含有序列(N)<sub>3-12</sub>,其中N是任何核苷酸。

在一些任何此类实施方案中,所述简并悬突序列含有序列NNNNNN,其中N是任何核苷酸 (SEQ ID NO:24)。在一些实施方案中,所述夹板寡核苷酸含有序列ACACGACGCTCTCCGATCTNNNNNN,其中N是任何氨基酸 (SEQ ID NO:26)。在一些任何此类实施方案中,含有所述第二通用引发位点的寡核苷酸含有序列AGATCGGAAGAGCGTCGTGT (SEQ ID NO:25)。

在一些任何此类实施方案中,所述容器条码化寡核苷酸含有至少或约至少3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、30个、40个或50个核苷酸。在一些任何此类实施方案中,所述容器条码化寡核苷酸含有从或从约10个至30个核苷酸。在一些任何此类实施方案中,所述容器条码化寡核苷酸含有简并序列。在一些任何此类实施方案中,所述容器条码化寡核苷酸含有序列(N)<sub>14-17</sub>,其中N是任何核苷酸,任选地其中所述序列中的至少一个或两个N是W,其中W是腺嘌呤或胸腺嘧啶。在一些任何此类实施方案中,所述容器条码化寡核苷酸含有序列NNNNWNNNNWNNNN (SEQ ID NO:80)、WNNNNWNNNNWNNNN (SEQ ID NO:81)、NWNNNWNNNNWNNNN (SEQ ID NO:82)或NNWNNNWNNNNWNNNN (SEQ ID NO:83),其中N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶。

在一些任何此类实施方案中,每个容器含有第一衔接子池,其中与所述池中的所述其他容器条码化寡核苷酸中的至少一个相比,所述第一衔接子池的每个容器条码化寡核苷酸含有至少一个碱基位移或碱基添加。在一些任何此类实施方案中,所述第一衔接子池的容器条码化寡核苷酸含有序列NNNNWNNNNWNNNN (SEQ ID NO:80)、WNNNNWNNNNWNNNN (SEQ ID NO:81)、NWNNNWNNNNWNNNN (SEQ ID NO:82)和NNWNNNWNNNNWNNNN (SEQ ID NO:83),其中N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶。

在一些任何此类实施方案中,在步骤(d)中,所述方法还包括扩增所述一个容器条码化寡核苷酸或所述容器条码化寡核苷酸池或者一个所述第一衔接子或所述第一衔接子池,其中所述第一衔接子包含所述一个容器条码化寡核苷酸或所述容器条码化寡核苷酸池,其中在将所述容器条码化寡核苷酸附接至所述分子条码化多核苷酸之前或与之同时进行所述

扩增。在一些实施方案中,附接所述容器条码化寡核苷酸包括使所述容器条码化寡核苷酸的区域与所述互补多核苷酸中的每一个的区域或与含有分子条码的所述分子条码化多核苷酸中的每一个的区域杂交。在一些情况下,所述区域含有与所述条码化多核苷酸的分子条码的3'末端区域互补的3'标记多核苷酸。在一些实施方案中,所述区域含有与所述分子条码化寡核苷酸的5'末端区域互补的3'标记多核苷酸。

在一些任何此类实施方案中,在步骤(b)中,所述一个或多个靶多核苷酸通过在逆转录酶和一个或多个靶特异性引物的存在下将所述一个或多个靶多核苷酸逆转录来产生,所述一个或多个靶特异性引物与所述一个或多个靶多核苷酸的靶序列互补;和/或多核苷酸的所述集合通过在逆转录酶和一个或多个转录组引物的存在下将所述细胞中的多核苷酸(如多核苷酸转录物)逆转录来产生,所述一个或多个转录组引物与所述细胞中的多核苷酸(如多核苷酸转录物)互补。

在一些任何此类实施方案中,所述一个或多个靶多核苷酸含有免疫分子或其链的多核苷酸。在一些任何此类实施方案中,所述一个或多个靶多核苷酸含有至少两个靶多核苷酸,所述至少两个靶多核苷酸各自含有免疫分子链的多核苷酸。

在一些任何此类实施方案中,所述一个或多个靶多核苷酸含有TCR或其链的多核苷酸。在一些任何此类实施方案中,所述一个或多个靶多核苷酸含有T细胞受体 $\alpha$  (TCR $\alpha$ )的第一多核苷酸和T细胞受体 (TCR $\beta$ ) 的第二多核苷酸。在一些实施方案中,所述一个或多个靶多核苷酸含有T细胞受体 $\gamma$  (TCR $\gamma$ ) 的第一多核苷酸和T细胞受体 $\delta$  (TCR $\delta$ ) 的第二多核苷酸。

在一些任何此类实施方案中,所述一个或多个靶多核苷酸含有抗体或其链的多核苷酸。在一些实施方案中,所述一个或多个靶多核苷酸含有重链免疫球蛋白 (IgH) 多核苷酸的第一多核苷酸和轻链免疫球蛋白 (IgL) 多核苷酸的第二多核苷酸。在一些实施方案中,所述一个或多个靶特异性引物和/或所述一个或多个转录组引物包括聚 (T) 序列。

在一些任何此类实施方案中,所述一个或多个转录组引物含有随机六聚体寡核苷酸引物的混合物。在一些实施方案中,所述一个或多个靶特异性引物含有与所述靶多核苷酸的一个或多个靶序列的一个或多个序列互补的一个或多个引物。在一些情况下,所述一个或多个靶特异性引物含有至少两个引物,例如第一引物和第二引物。在一些实施方案中,所述一个或多个靶特异性引物含有针对多个靶多核苷酸的靶序列的引物,所述多个靶多核苷酸各自编码免疫分子或其链。在一些方面,所述免疫分子是T细胞受体或抗体。

在一些任何此类实施方案中,至少所述第一引物与免疫分子的第一链的多核苷酸的靶序列互补,并且第二引物与所述免疫分子的第二链的多核苷酸的靶序列互补。在一些实施方案中,所述第一引物和所述第二引物与TCR的不同TCR链多核苷酸的靶序列互补。

在一些任何此类实施方案中,所述第一引物与TCR $\alpha$ 多核苷酸序列的靶序列互补,并且所述第二引物与TCR $\beta$ 多核苷酸序列的靶序列互补;或者所述第一引物与TCR $\gamma$ 多核苷酸序列的靶序列互补,并且所述第二引物与TCR $\delta$ 多核苷酸序列的靶序列互补。在一些方面,所述TCR链多核苷酸的靶序列是恒定区序列。

在一些任何此类实施方案中,所述第一引物与TCR $\alpha$ 恒定区多核苷酸序列的靶序列互补,并且所述第二引物与TCR $\beta$ 恒定区多核苷酸序列的靶序列互补;或者所述第一引物与TCR $\gamma$ 恒定区多核苷酸序列的靶序列互补,并且所述第二引物与TCR $\delta$ 恒定区多核苷酸序列的靶序列互补。在一些实施方案中,至少所述第一引物和所述第二引物与抗体的不同抗体链多

核苷酸的靶序列互补。在一些实施方案中,所述第一引物与重链免疫球蛋白(IgH)多核苷酸序列的靶序列互补,并且所述第二引物与轻链免疫球蛋白(IgL)多核苷酸序列的靶序列互补。在一些实施方案中,所述抗体链多核苷酸的靶序列是恒定区序列。

在一些任何此类实施方案中,所述第一引物与重链恒定区(CH)多核苷酸序列的靶序列互补,并且所述第二引物与轻链恒定区(CL)多核苷酸序列的靶序列互补。在一些情况下,所述CH多核苷酸的靶序列来自IgM、IgD、IgA、IgE或IgG或其组合;和/或所述CL多核苷酸序列的靶序列来自Igκ、Igλ或其组合。

在一些任何此类实施方案中,所述一个或多个靶多核苷酸含有全长编码序列。在一些任何此类实施方案中,所述一个或多个靶多核苷酸和多核苷酸的所述集合在所述容器中以相同的反应体积产生。

在所述方法的一些任何此类实施方案中,在步骤(b)中,产生所述多个互补多核苷酸包括使用非模板末端转移酶,其中将三个或更多个非模板核苷酸、核糖核苷酸或其类似物添加到每个产生的互补多核苷酸的3'末端。在一些情况下,所述非模板末端转移酶是逆转录酶或聚合酶。在一些方面,所述非模板末端转移酶是逆转录酶,并且其中所述逆转录酶选自Superscript II逆转录酶、Maxima逆转录酶、Protoscript II逆转录酶、莫洛尼鼠白血病病毒逆转录酶(MMLV-RT)、HighScriber逆转录酶、禽成髓细胞瘤病毒(AMV)逆转录酶、包括末端脱氧核苷酸转移酶活性的任何逆转录酶及其组合。

在所述方法的一些任何此类实施方案中,在步骤(c)中,所述附接包括使所述多个分子条码化寡核苷酸之一的区域与所述互补多核苷酸中的每一个的所述三个或更多个非模板核苷酸杂交。在一些实施方案中,提供所述多个分子条码化寡核苷酸作为多个模板转换寡核苷酸,所述多个模板转换寡核苷酸各自包括与所述三个或更多个非模板核苷酸互补的3'部分。在一些情况下,所述模板转换寡核苷酸还包括与包括所述容器条码的第一衔接子的3'标记多核苷酸互补的5'末端区域。在一些情况下,所述模板转换寡核苷酸还包含与所述第一衔接子的一部分互补的5'末端区域,其中所述第一衔接子包含所述容器条码化寡核苷酸。

在一些任何此类实施方案中,所述逆转录酶具有模板转换活性;所述多个产生的互补多核苷酸的至少一些链包含含有三个或更多个非模板核苷酸的3'悬突;提供所述多个分子条码化寡核苷酸作为多个模板转换寡核苷酸,所述多个模板转换寡核苷酸各自含有(1)与包含所述第一衔接子和所述容器条码化寡核苷酸的3'标记寡核苷酸互补的5'末端区域、(2)所述分子条码和(3)与所述3'悬突的三个或更多个非模板核苷酸互补的3'部分;并且所述模板转换寡核苷酸用作所述逆转录酶的模板,使得将所述分子条码掺入每个互补多核苷酸中以产生所述分子条码化多核苷酸。

在一些任何此类实施方案中,与所述三个或更多个非模板核苷酸互补的3'部分包括核苷酸、核糖核苷酸或其类似物。在一些任何此类实施方案中,所述三个或更多个非模板核苷酸包括三个或更多个C核苷酸,并且与三个或更多个非模板核苷酸互补的所述3'部分含有一个或多个G核苷酸或核糖核苷酸或其类似物。

在一些任何此类实施方案中,所述模板转换寡核苷酸还含有3'修饰的核苷酸,其阻断通过逆转录酶或DNA聚合酶延伸所述模板转换寡核苷酸。在一些情况下,所述修饰是3'末端核苷酸的脱氧、磷酸、氨基或烷基修饰。

在所述方法的一些任何此类实施方案中,步骤(d)还包括在所述附接之后延伸所述多个互补多核苷酸中的每一个。在一些实施方案中,步骤(d)还包括在所述附接以产生所述多个双条码化多核苷酸之后延伸所述多个分子条码化多核苷酸中的每一个。

在一些任何此类实施方案中,所述容器是孔、乳剂、微滴或微胶囊。

在一些任何此类实施方案中,所述方法包括在步骤(e)之前组合所述多个容器中的两个或更多个的内容物,从而产生均质混合物,所述均质混合物包括所述多个双条码化单链多核苷酸中的两个或更多个。在一些方面,组合所述多个容器的内容物包括破坏所述多个容器中的两个或更多个,并且将来自所述两个或更多个破坏的容器的双条码化单链多核苷酸合并。

在一些实施方案中,所述方法包括在步骤(e)之前选择或纯化具有大于或大于约50个碱基对、大于100个碱基对或大于200个碱基对的大小的双条码化单链多核苷酸。在一些实施方案中,所述方法包括在步骤(e)之前选择或纯化具有以下大小的双条码化单链多核苷酸:从或从约50个碱基对(bp)至1500bp、50bp至1250bp、50bp至1000bp、50bp至750bp、50bp至500bp、100bp至1500bp、100bp至1250bp、100bp至1000bp、100bp至750bp、100bp至500bp、200bp至1500bp、200bp至1250bp、200bp至1000bp、200bp至750bp或250bp至500bp。

在一些任何此类实施方案中,所述双条码化单链多核苷酸按顺序(5'至3')含有:所述第一衔接子、所述容器条码、所述分子条码和所述第二衔接子。在一些任何此类实施方案中,所述第一衔接子位于所述单链条码化多核苷酸、任选地所述双条码化单链多核苷酸的5'区域处或附近。在一些任何此类实施方案中,所述第二衔接子位于所述单链条码化多核苷酸、任选地所述双条码化单链多核苷酸的3'区域处或附近。

在所述方法的一些任何此类实施方案中,步骤(a)-(f)中的一个或多个是在溶液中进行和/或不在固体支持物的存在下进行,任选地其中所述支持物是珠。在一些任何此类实施方案中,至少步骤(c)和(d)是在溶液中进行和/或不在固体支持物的存在下进行,任选地其中所述支持物是珠。在一些实施方案中,步骤(a)-(e)中的每一个是在溶液中进行和/或不在固体支持物的存在下进行,任选地其中所述支持物是珠。

在一些任何此类实施方案中,所述细胞群含有至少或约至少 $1 \times 10^3$ 个、 $5 \times 10^3$ 个、 $1 \times 10^4$ 个、 $5 \times 10^4$ 个、 $1 \times 10^5$ 个、 $5 \times 10^5$ 个、 $1 \times 10^6$ 个或 $5 \times 10^6$ 个细胞。在一些实施方案中,所述细胞群来自受试者的生物样品。在一些例子中,所述生物样品是或含有全血样品、血沉棕黄层样品、外周血单核细胞(PBMC)样品、未分级T细胞样品、淋巴细胞样品、白细胞样品、单采术产物或白细胞单采术产物。

在一些任何此类实施方案中,所述细胞群含有免疫细胞。在一些实施方案中,所述免疫细胞含有淋巴细胞或抗原呈递细胞。在一些任何此类实施方案中,所述免疫细胞是淋巴细胞或其亚型、B细胞或其亚型、T细胞或其亚型、或其组合。在一些例子中,所述免疫细胞是T细胞,所述T细胞是CD4+和/或CD8+T细胞。

在一些任何此类实施方案中,所述细胞群富集或含有中枢记忆T细胞、效应记忆T细胞、幼稚T细胞、干细胞样中枢记忆T细胞(stem central memory T cell)、效应T细胞和调节性T细胞。在一些任何此类实施方案中,所述细胞群富集记忆B细胞、幼稚B细胞或浆母细胞B细胞。

在一些任何此类实施方案中,所述受试者是人类受试者。在一些实施方案中,所述受试

者患有癌症、感染或自身免疫病症。在一些情况下，所述感染是病毒、细菌或真菌感染。

在一些任何此类实施方案中，所述方法还包括扩增所述多个条码化单链多核苷酸，从而产生多个多核苷酸模板。在一些任何此类实施方案中，所述多个条码化单链多核苷酸的扩增是在第一引物组的存在下进行，所述第一引物组包括与所述第一衔接子序列互补的第一引物和与所述第二衔接子序列互补的第二引物。在一些方面，所述第一引物和/或所述第二引物是通用引物。在一些情况下，所述第一引物和/或所述第二引物与所述P7引发位点(C7)或其连续部分或者所述P5引发位点(C5)或其连续部分互补。在一些情况下，所述第一引物与所述P7引发位点(C7)或其连续部分互补，并且所述第二引物与所述P5引发位点(C5)或其连续部分互补。在一些实施方案中，与所述P7引发位点(C7)或其连续部分互补的引物具有或含有序列CAAGCAGAAGACGGCATACGAGAT (SEQ ID NO:39)；和/或与所述P5引发位点(C5)或其连续部分互补的引物含有序列ACACGACGCTCTTCCGATCT (SEQ ID NO:27)。在一些任何此类实施方案中，所述第一引物和/或所述第二引物还含有测序衔接子。在一些实施方案中，与所述P7引发位点(C7)或其连续部分互补的引物还含有序列CAAGCAGAAGACGGCATACGAGAT [NNNNNN] GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT (SEQ ID NO:28)；和/或与所述P5引发位点(C5)或其连续部分互补的引物含有序列AATGATACGGCGACCAC CGAGATCTACACTCTTCCCTACACGACGCTCTTCC (SEQ ID NO:76)。

在一些实施方案中，所述方法还包括纯化所述多个单链条码化多核苷酸、任选地单链双条码化多核苷酸中的每一个。

提供了一种多核苷酸文库，其含有通过所描述的任何实施方案的方法产生的多个条码化多核苷酸。还提供了一种多核苷酸文库，其含有多个条码化多核苷酸，其中所述多个条码化多核苷酸含有(i)一个或多个靶多核苷酸，所述一个或多个靶多核苷酸含有在细胞群的细胞中存在的一个或多个靶多核苷酸的扩增子；以及(ii)多核苷酸的集合，所述多核苷酸各自含有所述细胞中的多核苷酸的扩增子，其中每个条码化多核苷酸含有包括与第一通用引物互补的第一通用引发位点的第一衔接子；包含容器条码的容器条码化寡核苷酸，其中所述容器条码对于来自所述细胞群的同一细胞的来自(i)和(ii)的所有条码化多核苷酸而言是相同的；以及含有与第二通用引物互补的第二通用引发位点的第二衔接子序列。

在一些实施方案中，所述多个条码化多核苷酸中的每一个含有分子条码，所述分子条码对于每个多核苷酸或多核苷酸模板而言是独特的。在一些方面，来自所述细胞群的每个细胞的条码化多核苷酸模板的所述集合共同含有转录组或部分转录组的互补DNA(cDNA)链或其互补体。在一些实施方案中，所述转录组或所述部分转录组共同含有所述细胞的基因组中存在的至少60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的转录物。

在一些任何此类实施方案中，所述条码化多核苷酸中的每一个具有大于或大于约50个碱基对、大于100个碱基对或大于200个碱基对的大小。在一些实施方案中，所述条码化单链多核苷酸中的每一个具有从或从约50个碱基对(bp)至1500bp、50bp至1250bp、50bp至1000bp、50bp至750bp、50bp至500bp、100bp至1500bp、100bp至1250bp、100bp至1000bp、100bp至750bp、100bp至500bp、200bp至1500bp、200bp至1250bp、200bp至1000bp、200bp至750bp或250bp至500bp的大小。

在一些任何此类实施方案中，所述第一衔接子含有所述容器条码。在一些实施方案中，

所述条码化多核苷酸是单链的。在一些实施方案中，所述条码化多核苷酸是双链的。在一些实施方案中，所述第一衔接子和所述第二衔接子是不同的。

在一些任何此类实施方案中，所述第一通用引发位点和/或所述第二通用引发位点是或含有P7引发位点(C7)或其连续部分或者P5引发位点(C5)或其连续部分，任选地其中所述其连续部分足够与互补序列退火。在一些实施方案中，所述第一通用引发位点是或含有所述P7引发位点(C7)或其连续部分，并且所述第二通用引发位点是或含有所述P5引发位点(C5)或其连续部分。在一些方面，所述P7引发位点(C7)含有序列AGATCGGAAGAGCACACGTCTGAACCTCCA(SEQ ID NO:77)，或者是其连续部分。在一些例子中，所述P5引发位点含有序列AGATCGGAAGAGCGTCGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCA TT(SEQ ID NO:78)，或者是其连续部分。

在一些任何此类实施方案中，所述连续部分在长度上含有至少或至少约15个、20个、25个或30个核苷酸。在一些实施方案中，所述P5引发位点是如SEQ ID NO:25(AGATCGGAAGAGCGTCGTGT)所示的连续部分。

在一些任何此类实施方案中，所述容器条码化寡核苷酸含有至少或约至少3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、30个、40个或50个核苷酸。在一些实施方案中，所述容器条码化寡核苷酸含有从或从约10个至30个核苷酸。

在一些任何此类实施方案中，所述一个或多个靶多核苷酸含有免疫分子或其链的多核苷酸。在一些实施方案中，所述一个或多个靶多核苷酸含有至少两个靶多核苷酸，所述至少两个靶多核苷酸各自含有免疫分子链的多核苷酸。

在一些任何此类实施方案中，所述一个或多个靶多核苷酸含有TCR或其链的一个或多个多核苷酸。在一些实施方案中，所述一个或多个靶多核苷酸含有T细胞受体 $\alpha$ (TCR $\alpha$ )的第一多核苷酸和T细胞受体(TCR $\beta$ )的第二多核苷酸。在一些实施方案中，所述一个或多个靶多核苷酸含有T细胞受体 $\gamma$ (TCR $\gamma$ )的第一多核苷酸和T细胞受体 $\delta$ (TCR $\delta$ )的第二多核苷酸。

在一些任何此类实施方案中，所述一个或多个靶多核苷酸含有抗体或其链的一个或多个多核苷酸。在一些实施方案中，所述一个或多个靶多核苷酸含有重链免疫球蛋白(IgH)多核苷酸的第一多核苷酸和轻链免疫球蛋白(IgL)多核苷酸的第二多核苷酸。

在一些任何此类实施方案中，所述条码化多核苷酸按顺序(5'至3')含有：所述第一衔接子、所述容器条码、所述分子条码和所述第二衔接子。在一些实施方案中，所述第一衔接子位于所述双条码化单链多核苷酸的5'区域处或附近。在一些实施方案中，所述第二衔接子位于所述双条码化单链多核苷酸的3'区域处或附近。

提供了用于测序的方法，其包括对以下进行测序：通过所描述的任何实施方案产生或从所描述的多核苷酸文库的任何实施方案产生的所述多个多核苷酸(如条码化多核苷酸(例如，双条码化多核苷酸))中的一个或多个。在一些例子中，对来自所述多个多核苷酸(如多核苷酸模板)的转录组进行测序。在一些情况下，所述方法还包括在所述测序之前扩增全转录组或其部分。在一些方面，使用第一引物组进行扩增，所述第一引物组含有分别对所述第一衔接子序列和所述第二衔接子序列具有特异性的第一引物和第二引物。

在一些实施方案中，对来自所述多个多核苷酸模板的一个或多个靶多核苷酸进行测序。在一些情况下，所述方法还包括在所述测序之前扩增来自所述多个多核苷酸模板的一个或多个靶多核苷酸。在一些例子中，扩增所述一个或多个靶多核苷酸的一个或多个全长

序列。

在一些实施方案中,在第二引物组的存在下进行扩增,所述第二引物组合有与一个或多个靶多核苷酸互补的一个或多个第一引物和与所述第一衔接子序列互补的第二引物。在一些情况下,所述第二引物组的第二引物与所述P7引发位点(C7)或其连续部分或者所述P5引发位点(C5)或其连续部分互补。在一些方面,所述第二引物组的第二引物与所述P7引发位点(C7)或其连续部分互补。在一些实施方案中,所述第二引物组的第二引物具有或含有序列CAAGCAGAAGACGGCATACGAGAT (SEQ ID NO:39) 或CAAGCAGAAGACGGCATACGAGAT [NNNNNN] GTGACTGGAGTTCAGACGTGTGCTTCCGATCT (SEQ ID NO:28)。

在一些实施方案中,与所述一个或多个靶多核苷酸互补的一个或多个第一引物对免疫分子或其链的靶序列具有特异性。在一些情况下,所述免疫分子是T细胞受体或抗体。在一些方面,所述一个或多个第一引物对所述免疫分子的恒定区的靶序列具有特异性。

在一些实施方案中,所述免疫分子是TCR,并且所述一个或多个第一引物包括AGTCTCTCAGCTGGTACACGG (SEQ ID NO:37)、ATGGCTAAACACAGCGACCTC (SEQ ID NO:38) 或其组合。在一些实施方案中,所述免疫分子是抗体,并且所述一个或多个第一引物包括SEQ ID NO:29-36中的任何一个或其组合。

在一些任何此类实施方案中,所述方法包括确定所述一个或多个条码化多核苷酸、任选地双条码化多核苷酸的细胞起源。在一些情况下,确定细胞起源包括鉴定双条码化多核苷酸的序列信息或序列,所述双条码化多核苷酸因为来自同一细胞而具有相同的容器条码。

在一些任何此类实施方案中,所述靶多核苷酸是含有第一多核苷酸链和第二多核苷酸链的免疫分子,并且所述方法包括通过在经测序的双条码化多核苷酸中存在相同的容器条码将所述第一多核苷酸链和所述第二多核苷酸链与同一细胞进行匹配。在一些实施方案中,所述方法还包括定量或确定具有相同条码、任选地相同分子条码和/或容器条码的多核苷酸的数量。

在一些任何此类实施方案中,所述多个条码化多核苷酸是包含分子条码和容器条码的双条码化多核苷酸,并且所述方法还包括鉴定具有相同容器条码的转录组序列和靶多核苷酸序列,从而鉴定带有所述一个或多个靶多核苷酸的细胞的转录组信息。

提供了用于转录组分析的方法,其包括(a)对来自通过所描述的任何方法产生的所述多个条码化多核苷酸或来自所描述的任何多核苷酸文库的所述多个条码化多核苷酸的一个或多个靶多核苷酸进行测序,其中所述条码化多核苷酸是包含分子条码和容器条码的双条码化多核苷酸,从而产生来自所述多个细胞的靶多核苷酸的序列信息;(b)对来自通过所描述的任何方法产生的所述多个条码化多核苷酸或来自所描述的任何多核苷酸文库的所述多个条码化多核苷酸的全转录组或其部分进行测序,其中所述条码化多核苷酸是包含分子条码和容器条码的双条码化多核苷酸,从而产生来自所述多个细胞的转录组数据;以及(c)鉴定因为来自同一细胞而具有相同容器条码的来自(a)和来自(b)的序列信息。

提供了用于分析选定单细胞的转录组的方法,其包括(a)对来自通过所描述的任何方法产生的所述多个条码化多核苷酸中的多个条码化多核苷酸的一个或多个靶多核苷酸或来自所描述的任何多核苷酸文库的所述多个条码化多核苷酸的一个或多个靶多核苷酸进行扩增和测序,其中所述条码化多核苷酸是包含分子条码和容器条码的双条码化多核苷

酸,从而产生在所述多个细胞中的至少一个中的所述靶多核苷酸中的每一个的序列信息; (b) 鉴定与在 (a) 中测序的所述靶多核苷酸之一相关的一个或多个容器条码,从而鉴定带有所述靶多核苷酸的选定单细胞; (c) 对来自带有 (b) 中鉴定的容器条码的细胞的所述多个条码化多核苷酸的转录组或其部分进行扩增和测序,从而产生来自表达靶多肽的所述选定细胞的转录组数据。在一些实施方案中,使用对 (b) 中鉴定的容器条码具有特异性的引物和对所述条码化多核苷酸的第二衔接子序列具有特异性的引物对来自所述选定细胞的转录组或其部分扩增或测序。

在一些实施方案中,所述方法包括将所述转录组或其部分的序列信息与来自同一细胞的所述一个或多个靶多核苷酸中的至少一个进行匹配,其中所述序列信息是从通过所描述的任何方法产生的所述多个条码化多核苷酸或从所描述的任何多核苷酸文库的所述多个条码化多核苷酸确定或者是从所描述的任何方法确定,其中所述条码化多核苷酸是包含分子条码和容器条码的双条码化多核苷酸。在一些情况下,具有相同容器条码的序列因为来自同一细胞而相匹配。在一些实施方案中,所述转录组数据包括与所述细胞的功能或活性相关的参数、特征、特性或表型。在一些情况下,所述转录组数据与所述细胞的激活、耗竭或增殖活性相关。

## 附图说明

图1A描绘了条码化阶段的示意图:本文所述的示例性方法。草图代表将用于文库制备和免疫测序的两个或更多个多核苷酸(如靶多核苷酸和一个或多个基因组或转录组多核苷酸)或成对序列(如成对可变Ig(例如,V<sub>H</sub>和V<sub>LmRNA</sub>)或TCR序列(例如,V<sub>α</sub>/V<sub>β</sub>和V<sub>γ</sub>/V<sub>δmRNA</sub>))进行扩增和条码化的方法。容器条码(VB);分子条码(MB)。(顶图)含有单细胞和其他反应组分(例如,酶、缓冲液、寡核苷酸)的乳剂(作为示例性容器)的单微滴(多个微滴中的单微滴)。(中图)细胞裂解和利用靶特异性、随机和/或寡聚dT逆转录引物进行裂解细胞RNA的逆转录的示例性方法。(底图)在逆转录阶段期间单分子的模板转换阶段和分子条码(MB)标记。

图1B描绘了本文所述的示例性方法的扩增阶段的示意图。草图代表将用于文库制备和免疫测序的两个或更多个多核苷酸(如靶多核苷酸和一个或多个基因组或转录组多核苷酸)或成对序列(如成对可变Ig(例如,V<sub>H</sub>和V<sub>LmRNA</sub>)和TCR序列(例如,V<sub>α</sub>/V<sub>β</sub>和V<sub>γ</sub>/V<sub>δmRNA</sub>))进行扩增和条码化的方法。(顶图)容器条码(VB)的独立扩增在每个微滴中产生相同VB的多个拷贝。在扩增的退火和延伸阶段期间,分子条码(MB)cDNA分子同时被标记上VB。(中图)在扩增循环期间双条码化cDNA分子的同时扩增。(底图)示例性双条码化cDNA分子准备好进行进一步处理(例如,纯化、大小选择、衔接子连接、扩增和测序)。

图2描绘了本文所述的示例性方法的双条码化转录物的扩增和测序。(顶图)使用对第一衔接子的通用引发序列具有特异性的引物和靶特异性引物(每个引物与测序衔接子连接用于测序)对编码感兴趣的靶基因的转录物的扩增和测序。(中图)使用对第一衔接子的通用引发序列具有特异性的引物和对第二衔接子的通用引发序列具有特异性的引物对文库中的所有转录物(例如,一个或多个细胞的转录组或其部分)的扩增和测序。(底图)使用对来自同一细胞的多核苷酸共有的所确定的容器条码(VB)具有特异性的引物和对第二衔接子的通用引发序列具有特异性的引物对选定的感兴趣的细胞(如确定含有感兴趣的靶基因的mRNA转录物的细胞)的转录组的扩增和测序。

图3描绘了根据Seurat软件工具通过推断的簇身份进行着色(A;虚线表示主要展现相同颜色的事件的簇)或通过所测序的全长免疫受体的类型(B细胞受体(BCR)或T细胞受体(TCR))进行着色(B;虚线表示主要展现相同颜色的事件的簇)的单细胞转录组数据的t-SNE图。

图4描绘了图3的单细胞转录组数据的t-SNE图,被着色以鉴定表达以下所鉴定序列的细胞:TLR7样受体7(TLR7;A)、T细胞表面糖蛋白CD3 $\epsilon$ 链(CD3E;B)、自然杀伤细胞颗粒蛋白7(NKG7;C)、甘露糖受体C型1(MRC1;D)。

## 具体实施方式

本文提供了用于分析单细胞中或多个单细胞中的基因表达的方法和组合物。所提供的方法允许有效地产生多个单细胞高质量多核苷酸(例如,DNA)测序文库,所述测序文库含有来自完整转录组或其部分的多核苷酸以及感兴趣的靶基因(如全长配对的免疫受体产物,例如TCR或抗体)的一个或多个全长多核苷酸序列,由此可以鉴定一个或多个靶多核苷酸和来自源自于同一细胞的完整或部分转录组的多核苷酸。在一些实施方案中,与必须在进行实验后决定的特定基因相反,文库中的所述多个多核苷酸中的每一个含有允许对总回收产物进行下一代测序的衔接子序列(例如,第一衔接子和第二衔接子)。因此,在所提供的方法的一些方面,可以使用对这些衔接子序列具有特异性的引物和/或对感兴趣的靶多核苷酸具有特异性的引物和/或细胞特异性引物进行随后的PCR扩增。在一些实施方案中,所提供的方法允许与全长免疫分子序列(例如,抗体或TCR)结合以有效且高通量的方式在单个实验中处理成千上万个细胞,从而产生单细胞测序数据,例如RNA-seq数据(如mRNA计数)。

在一些情况下,越来越多地使用组织的全部或部分基因组和mRNA含量的直接测序,使得能够分析可变剪接、调节/启动子区域和聚腺苷酸化信号,而无需预先选择先前已知的基因(Clooney等人,Nat Methods 5(7):613-9(2008))。然而,当前通过直接测序分析细胞mRNA含量的方法依赖于分析从通常含有数百万个细胞的组织样品中获得的大量mRNA。这意味着在大量mRNA中分析基因表达时,单细胞中存在的许多功能信息会丢失或难以区分。此外,在动态过程(如细胞周期)期间的基因表达难以在群体平均值中观察到。因此,在一些应用中,直接测序的基于单细胞的方法比大量样品更可取。

在一些情形中,期望分析选定细胞(如表达特定基因或感兴趣的基因(如靶基因)的细胞)的基因组或mRNA含量。在一些情形中,期望获得选定细胞的基因组或转录组含量,同时还获得靶基因(如免疫受体)的全长序列。用于单细胞转录组测序的现有工具包括微阵列、基于96孔的方法(如传统的FACS分选到孔中)和微流体仪器(如Fluidigm C1)。这些工具可以用于制备全转录组和靶文库,但是它们的通量受到限制,因为它们受限于可以分析的细胞数量(例如,数百到数千个细胞)。

在一些情形中,期望分析选定细胞(如表达特定基因或感兴趣的基因、如靶基因(例如,免疫受体)的细胞)的基因组或mRNA含量。已经描述了使用微孔阵列或乳剂的超高通量方法以允许全单细胞转录组测序(参见例如,Klein等人,Cell (2015) 161(5):1187-1201;Macosko等人,Cell (2015) 161(5):1202-1214;WO/2015/164212;WO/2016/040476),但使用现有技术不可能有效捕获全长靶序列,如全长免疫受体序列。由于基于珠的方法所施加的限制,这些方法还受限于较少的细胞数量(通常为五千个以下),所述基于珠的方法需要对

每个细胞而言更大的微滴和更大的反应体积。

在所提供的方法的实施方案中,靶序列(如靶免疫分子(例如,抗体或TCR))和多个细胞的基因组或转录组序列在一个同时反应中产生,并且提供了用于联系衍生自同一细胞的序列的序列信息的机制。在一些方面,当与高通量测序技术结合时,当前公开的方法允许分析大量的单细胞并在一个单反应测定中实现所述分析。原则上,人们可以测序任何数量的细胞和每个细胞而言任何数量的靶向区域。在一些方面,可以处理的单细胞的数量仅受限于实际约束,如高通量测序的速度。在一些实施方案中,本文公开的方法适用于与珠一起使用。在其他实施方案中,本文公开的方法不包括基于珠的测序或扩增步骤。

在一些方面,所提供的方法通过提供制备可以用于分析多个单细胞中的基因表达的cDNA文库的方法克服或减少了现有方法的问题。在一些实施方案中,所提供的方法导致用于超高通量测序的多核苷酸文库的产生,这允许现成合成的全长靶序列(包括成对的异二聚体或多聚体靶的序列)的回收,同时捕获被鉴定为表达一个或多个靶序列的细胞的互补定量基因组或转录组信息。

具体地,所提供的方法用于从多个单细胞制备多核苷酸文库,例如cDNA文库。所述方法是基于从单个细胞的群体中确定基因表达水平,其可以用于按细胞水平鉴定细胞上基因表达的自然变化。所述方法还可以用于鉴定和表征细胞群的细胞组成,包括在没有合适的细胞表面标记的情况下。本文所述的方法还提供了使用单细胞产生代表细胞群中RNA含量的cDNA文库的优点,然而通过经典方法制备的cDNA文库通常需要从大群体中分离总RNA。因此,在一些方面,使用所提供的方法产生的cDNA文库通过利用单个细胞的较小亚群以及如本文所述的另外的优点允许至少等价地表示细胞群中的RNA含量。

所提供的实施方案是基于用于单细胞靶测序(如免疫靶测序)和转录组分析的基于乳剂的方法。在一些方面,例如使用基于微流体乳剂的方法,将来自含有细胞群的样品的细胞封装在单细胞容器(例如微滴)中。然后,进行方法以将容器(例如微滴)特异性条码附接至乳剂微滴内的靶多核苷酸(例如,靶mRNA转录物的扩增子)和/或多个多核苷酸(例如,完整或部分转录组的mRNA转录物的扩增子),从而允许对微滴内包含的单细胞进行高通量遗传和/或表达分析。在一些方面,容器条码最初作为具有侧翼为已知引物位点的随机化中心序列部分的单分子DNA模板存在。可以在容器(如微滴)内将模板逆转录以产生扩增子和/或PCR扩增以产生一个或多个扩增子,并且可以通过序列重叠与细胞衍生的核酸附接。在所述方法的一些方面,使用基因特异性PCR引物扩增细胞衍生的核酸,以扩增一个或多个感兴趣的靶基因,例如免疫分子,如抗体或T细胞受体(TCR)。在一些情况下,扩增几个靶基因的能力可以提供关于细胞各种特性的信息,如细胞的表型、活性或其他特性。

在一些情况下,添加另外的靶基因可以将关于特定免疫分子(例如TCR)的信息与细胞的细胞表型信息联系起来。在所述方法的另外方面,可以使用允许对细胞的完整转录组或其部分进行测序以及对源自于同一细胞的转录物进行匹配的引物来扩增条码化全转录组细胞衍生的核酸。在一些方面,将随机逆转录引物用于产生与转录组的转录物对应的扩增子。在某些实施方案中,可以例如通过添加的第一衔接子和第二衔接子将通用引发位点添加到条码化多核苷酸的末端,使得可以批量扩增整个文库并以高通量散弹枪的方式测序。

在一些实施方案中,扩增细胞衍生的核酸的反应可以在一锅反应中进行,所述反应可以进行a)细胞裂解;b)靶mRNA逆转录;c)每个cDNA的分子条码化;d)容器特异性DNA条码的

PCR扩增；以及e)将一个拷贝的容器条码附接至每个cDNA。在一些实施方案中，可以将产物进行回收和测序，例如使用多种测序平台中的任何一种。例如，可以使用325 x 300bp的 Illumina MiSeq平台对每种产物的全长进行测序。在一些方面，微滴条码允许鉴定来自每个单细胞的所有产物。在某些方面，分子条码允许对每个细胞的表达定量，并且在一些情况下，测序和RT-PCR错误的消除。

在所提供的实施方案中，转录组测序的过程可以与一些选定的转录物（如TCR）的靶向测序分开进行，因为可以单独处理扩增的文库。这允许对同一扩增的条码化文库在单独等分试样上进行许多次所述方法。因为对同一扩增的条码化文库在不同时间进行不同实验的能力，通过所提供的方法可以实现多种有用的方法。在一个例子中，所述方法可以用于首先进行对样品中的感兴趣的靶分子（例如TCR）的靶向测序，这可能导致鉴定一些特定的感兴趣的细胞。在一些方面，然后可以将PCR或捕获寡核苷酸设计成靶向代表那些感兴趣的细胞的容器条码，从而允许仅捕获并测序那些细胞的所有条码化转录物。在一些方面，这大大降低了测序整个文库的测序成本。

在某些方面，本文所述的方法和组合物可用于单细胞分析，例如像用于研究复杂细胞样品的基因组、转录组、蛋白质组、代谢途径等。在研究复杂样品或混合物时，对异质细胞群中的多个细胞的分析特别有用。复杂样品或细胞混合物包括例如外周血单核细胞（PBMC）样品、宏基因组样品、正常和癌性组织切片、胚胎和干细胞集落。需要基因组和转录组测序来鉴定发散细胞类型或在某些阶段（如激活、耗竭或增殖的不同阶段）的细胞。特定应用包括分子T细胞受体和B细胞受体谱分析、单体分型和HLA分型。

“宏基因组样品”是指含有来自多个起源（如物种）的基因组的样品。例如，本方法可以应用于细菌物种的混合物，以允许在一个测定中对来自多种细菌的核酸进行测序，随后将序列与同一细菌细胞关联。类似地，一种细胞类型的核酸序列在另一种细胞类型的细胞（如浸润肿瘤的免疫细胞）中。在此类情形中，使用本文提供的方法可以将浸润肿瘤的免疫细胞的核酸与免疫受体或其结合片段的一个或多个全长序列关联。

所提供的实施方案还允许对大量的单细胞进行采样。使用表达模式的相似度，可以构建细胞图，显示细胞之间的关系。通过检测紧密相关细胞的簇，此图可以用于经由电脑模拟区分细胞类型。通过不是仅采样少数单细胞，而是采样大量单细胞，表达模式的相似度可以用于构建细胞图以及它们之间的关系。此方法允许获得来自群体中存在的每种不同类型的细胞的未稀释的表达数据，而无需事先纯化那些细胞类型。此外，在可获得已知标记的情况下，这些可以用于经由电脑模拟描绘感兴趣的细胞。

所提供的实施方案包括通过从每个单细胞中释放mRNA以提供多个单独样品而从多个单细胞制备多核苷酸文库（例如，cDNA文库）的方法，其中每个单独mRNA样品中的mRNA来自单细胞，从而从每个单独mRNA样品中的mRNA合成cDNA的第一链（即，扩增子）并将核苷酸条码（例如，分子条码和/或容器条码）掺入cDNA扩增子中以提供多个条码化（例如，双条码化）cDNA样品（即，其中每个条码化cDNA样品中的cDNA与来自合并了条码化cDNA样品的单细胞的mRNA互补，并且扩增合并的cDNA样品以产生包含条码化双链cDNA的cDNA文库）。在一些实施方案中，所得的扩增子是双条码化单链cDNA，如与一个或多个mRNA模板对应的双条码化单链cDNA。在一些实施方案中，将条码化双链cDNA变性以产生条码化单链cDNA，以促进添加衔接子（如促进测序的衔接子）。通过利用上述方法，可行的是在短时间内从几百个单细胞

制备用于测序的样品。从RNA制备用于测序的片段文库的传统方法包括费力的凝胶切除步骤。在本文所述的方法的一些方面,将多个细胞制备为单个样品(在cDNA合成之后),这使得制备多个(如数百个)细胞用于测序是可行的。此外,因为将每个组的所述多个细胞一起制备(在单个管中),因此可以最小化技术变异。

在本发明的一些方面,将从单细胞获得的每个cDNA样品用条码标记,这允许在单细胞的水平上分析基因表达。这允许研究动态过程(如细胞周期)期间的表达并分析复杂组织(例如,脑)中的不同细胞类型。在一些方面,可以在分析之前合并cDNA样品。合并样品简化了来自每个单细胞的样品的处理,并减少了分析单细胞中的基因表达所需的时间,这允许对基因表达进行高通量分析。在扩增之前合并cDNA样品还提供了以下优点,即实际上消除了样品之间的技术变异。此外,因为在扩增之前合并cDNA样品,因此与单独扩增和处理来自每个单细胞的cDNA样品相比,产生足够量的cDNA用于后续分析所需的扩增较少。这减少了扩增偏差,并且也意味着在用于提供合并cDNA样品的所有细胞中任何偏差都将相似。也不需要RNA纯化、储存和处理,这有助于消除由RNA的不稳定性质引起的问题。

T细胞受体链对和抗体免疫球蛋白链对是预期使用当前公开的方法进行测序的两种免疫受体类型。在一些实施方案中,所提供的方法允许产生用于高通量测序和基因表达分析的多核苷酸文库,所述多核苷酸文库包括一个或多个靶序列的序列(如免疫受体的一个或多个序列)以及可以组合以提供基因组和/或转录组测序信息的序列。在一些实施方案中,可以开发多核苷酸文库,所述多核苷酸文库是人衍生的文库组,用于从具有特定共同属性的患者或群组中发现抗体和/或TCR。在所提供的方法的一些实施方案中,起始材料可以是含有感兴趣的细胞群的任何来源,所述感兴趣的细胞群确实或可能含有感兴趣的靶多核苷酸,如免疫分子或受体,例如抗体或TCR。在一些实施方案中,起始材料可以是外周血或来自组织活检,如果需要,从其中可以将免疫细胞整体分离或针对幼稚细胞、记忆细胞和/或抗体分泌细胞(ASC)进行亚分选。在一些实施方案中,所提供的方法可以应用于多种不同类型的单数或成对可变序列,例如T细胞受体链对和抗体免疫球蛋白链对。

在一些方面,通过所提供的方法产生的cDNA文库适合于通过直接测序分析单细胞的基因表达谱,并且有可能使用这些文库研究基因的表达,包括与以下相关的或者在带有以下的细胞中的基因的表达:感兴趣的特定靶多核苷酸,如免疫分子或受体,例如抗原、抗体或TCR。在一些实施方案中,可以分析先前未知的基因表达谱。在一些实施方案中,所提供的方法可以用于表征或比较来自样品的多个细胞中的每一个的转录细胞状态,例如细胞的激活、耗竭、增殖或其他所需的参数或属性。在一些实施方案中,通过查看带有对感兴趣的抗原具有特异性的特定TCR的特定细胞的反应,所提供的方法可以用于促进治疗候选物(如TCR)的发现。在所提供的方法的一些实施方案中,有可能鉴定表达与所需反应(例如, T细胞激活的性质或程度)相关的TCR的细胞。在一些实施方案中,与需要先验知识和/或选择较小候选基因组的现有方法相反,所提供的方法使得可以通过分析全转录组来捕获更丰富的数据集。

### I. 用于靶和转录组分析的多核苷酸文库

在一些实施方案中,本文提供的方法涉及对一种或多种靶多核苷酸分子的扩增和测序,以及对多核苷酸的集合的扩增和测序,如来自单细胞或细胞群的一种或多种靶分子以及多核苷酸的集合。在一些实施方案中,本文所述的方法和组合物可用于单细胞分析,例如像用于研究复杂细胞样品的基因组、转录组、蛋白质组、代谢途径等。在其他方面,本文所述

的方法和组合物可以用于免疫受体发现(例如,通过在单个B和T细胞中将重和轻免疫球蛋白或T细胞受体链配对)、以及用于HLA分型。在其他实施方案中,可以将抗体配对信息和单细胞分析组合,以使细胞功能或细胞状态信息与所鉴定的免疫受体序列的表达相关联。在仍其他方面,本文所述的方法和组合物可以用于监测小分子和药物或免疫疗法的影响、以及它们在用于诊断的复杂正常或癌性样品中的一种或多种作用或在新药物或治疗方案的发现中的一种或多种作用。在又其他实施方案中,所述方法和组合物可以用于检测和分析靶病原体,如生物样品中的细菌或病毒。

在一些实施方案中,所提供的方法可以包括如以下国际公开号中所述的方法的各种特性:WO 2012/048341、WO 2014/144495、WO 2016/044227、WO 2016/176322或WO 2017/053902中,将每一个通过引用以其整体并入。

本发明利用操纵核酸以便产生用于测序的多核苷酸的文库的步骤。在一些实施方案中,本发明利用操纵核酸以便产生靶多核苷酸分子的步骤,所述靶多核苷酸分子包括包含免疫受体(如由免疫细胞产生的抗体或TCR)的可变区的序列。在一些情况下,所述靶多核苷酸分子包括重组单克隆抗体。在一些实施方案中,本发明利用操纵核酸以便产生代表一个或多个细胞的转录组或基因组的多核苷酸的步骤。从一般意义上来说,在本发明的一些实施方案中,将细胞的(如免疫细胞和/或T细胞的)遗传材料的扩增(例如,逆转录聚合酶链反应(逆转录-PCR))用于产生免疫细胞遗传材料的cDNA扩增。

在一些实施方案中,所述方法可以用于获得关于细胞内感兴趣的靶多核苷酸(如TCR或抗体)的序列信息。靶基因可以从来自样品或细胞群的细胞的基因组DNA或mRNA获得。样品或细胞群可以包括免疫细胞。例如,对于靶抗体分子,免疫球蛋白基因可以从免疫细胞或T细胞的基因组DNA或mRNA获得。RNA可以是重链(V、D、J区段)或轻链(V、J区段)。在一些实施方案中,起始材料是来自免疫细胞的由V、D、J基因区段构成的RNA,所述RNA编码抗体并含有恒定区。

在一些实施方案中,除了获得感兴趣的靶多核苷酸(例如,免疫分子,如抗体或TCR)的全长序列数据外,所提供的方法还允许从全转录组产物和一个或多个全长靶(包括多亚基靶)多核苷酸(例如,抗体或TCR,包括全长配对的免疫受体产物)有效产生高质量的DNA测序文库。

在一些实施方案中,此类方法包括将衔接子DNA序列添加(例如,连接)到单链多核苷酸产物中,这可以允许对单细胞或多个单细胞的转录组进行扩增和下一代测序。

#### A. 多核苷酸文库

根据本文所述的方法产生的文库可以是包含具有适当条码(如容器条码和分子条码)的大或全长靶序列(如抗体或TCR序列)的文库。在一些实施方案中,所述文库含有大或全长靶序列(例如,抗体或TCR序列,包括抗体或TCR的两条链)以及与靶序列(例如,抗体或TCR)所源于的细胞的部分或完整转录组的一个或多个转录物对应的序列,每个序列都具有容器条码和分子条码。在此类实施方案中,大或全长靶序列(例如,抗体或TCR序列)以及与靶序列(例如,抗体或TCR)所源于的细胞的部分或完整转录组的一个或多个转录物对应的一个或多个序列含有相同的容器条码并且含有对于所述转录组的每个原始转录物而言独特的分子条码。在一些方面,容器条码被包括在第一衔接子中,所述第一衔接子被附接至每个靶多核苷酸和代表转录组的转录物的多核苷酸的集合的每个多核苷酸。

在一些实施方案中,提供了用于产生多核苷酸文库的方法,由此将衔接子(在下文中也称为第二衔接子)添加到多个先前或第一衔接子标记的条码化单链多核苷酸中的每一个中,使得所述衔接子位于多核苷酸的相对末端处,其中所述多个条码化单链多核苷酸包括(i)与在细胞群的细胞中存在的一个或多个靶多核苷酸互补的一个或多个靶单链多核苷酸;以及(ii)单链多核苷酸的集合,所述单链多核苷酸各自与细胞中的多核苷酸互补,其中所述多个条码化单链多核苷酸中的每一个含有容器条码,所述容器条码对于来自细胞群的同一细胞的所有互补多核苷酸而言是相同的。衔接子(如第一衔接子和第二衔接子中的每一个)含有通用引发序列,所述通用引发序列可以用于对衔接子标记的双条码化多核苷酸进行扩增或测序。

在一些实施方案中,可以对多核苷酸文库进行测序。在一些实施方案中,根据本文所述的方法产生的文库可以含有用于测序的适当成簇的区段。在一些实施方案中,可以产生相同分子条码的许多拷贝。在一些实施方案中,对于每个起始独特靶多核苷酸分子,可以产生含有相同分子条码的多核苷酸的许多拷贝。在一些实施方案中,对于标记有容器条码的每个起始独特靶多核苷酸分子,可以产生含有相同分子条码的多核苷酸的许多拷贝。可以对任何或所有序列进行测序和配对,例如以确定表达一个或多个靶序列的细胞的完整或部分转录组。

起始材料可以是来自细胞(如来自免疫细胞或T细胞)的RNA或DNA。在一些情况下,细胞可以是已知或怀疑含有所需的靶多核苷酸(如免疫受体,例如TCR或抗体)的细胞。例如,在抗体的情况下,靶细胞是包含编码抗体并含有恒定区的V、D、J基因区段的细胞。在一些实施方案中,靶多核苷酸包含重链区段(V、D、J区段)或轻链区段(V、J区段)。

可以使用一个多核苷酸或多核苷酸池将多核苷酸起始材料(如RNA)逆转录成cDNA。用于逆转录靶多核苷酸的多核苷酸池中的引物的例子可以包含与靶多核苷酸的区域互补的部分和/或可以包含用于逆转录全转录组或其部分的序列。在一些情况下,多核苷酸可以包含与靶RNA的区域互补的部分,例如在靶的恒定区中或与mRNA的聚A尾互补。在一些情况下,多个寡核苷酸(如引物)可以用于退火一个或多个靶序列(如恒定区)。在一些方面,所述一个或多个多核苷酸包括序列特异的、聚dT和/或随机六聚体引物。

可以采用逆转录酶进行逆转录反应。在特定实施方案中,逆转录酶可以包含非模板末端转移酶活性。当包含非模板末端转移酶活性的逆转录酶到达模板末端时,它可以添加三个或更多个非模板残基,如三个或更多个非模板胞嘧啶残基。在一些实施方案中,将Superscript II<sup>TM</sup>逆转录酶用于此目的。在一些实施方案中,将Maxima<sup>TM</sup>逆转录酶用于此目的。在一些实施方案中,将Protoscript II<sup>TM</sup>逆转录酶用于此目的。在一些实施方案中,将莫洛尼鼠白血病病毒逆转录酶(MMLV-RT)用于此目的。在一些实施方案中,将HighScriber<sup>TM</sup>逆转录酶用于此目的。在一些实施方案中,将末端脱氧核苷酸转移酶用于此目的。在一些实施方案中,将禽成髓细胞瘤病毒(AMV)逆转录酶用于此目的。可以使用具有非模板末端转移酶活性的能够转录RNA的任何逆转录酶。可以使用具有非模板末端转移酶活性的能够转录RNA的任何逆聚合酶。可以使用具有非模板末端转移酶活性的能够转录DNA的任何逆聚合酶。可以用一个或多个条码标记由逆转录产生的cDNA。在一些例子中,可以用容器条码和分子条码标记由逆转录产生的cDNA。特殊设计的各种寡核苷酸可以用于条码标记。

在一些实施方案中,模板转换可以用于产生文库,所述文库例如用于免疫库测序和/或

转录组分析。例如,可以在逆转录期间采用模板转换以在逆转录产物上产生与具有条码的多核苷酸(如容器条码化多核苷酸或分子条码化多核苷酸)互补的区域。可以在逆转录期间采用模板转换,以去除PCR偏差的问题。这些方法可以用于抗体测序,例如通过使用高通量测序平台。

在一些实施方案中,容器条码包括侧翼为已知引物位点的随机化序列部分。例如,可以用容器条码标记cDNA,所述容器条码可以包括一段具有或不具有一个或多个已知插入碱基位置的约20个简并核苷酸,如NNNNWN>NNWN>NNWN>NN (SEQ ID NO:99;其中N是任何核苷酸并且W是作为A或T的已知的插入碱基)或NNNNWISCN>NNWISCN>NN (SEQ ID NO:100;其中N是任何核苷酸;W是作为A或T的已知的插入碱基;I是作为A、T、G或C(即,N)的已知的插入碱基;S是作为G或C的插入碱基;并且C是已知的插入胞嘧啶)。在容器条码化寡核苷酸中包括的其他示例性序列包括NNNNWN>NNWN>NN (SEQ ID NO:80)、WNNNNWN>NNWN>NN (SEQ ID NO:81)、NW>NNWN>NNWN>NN (SEQ ID NO:82)或NNW>NNWN>NNWN>NN (SEQ ID NO:83),其中N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶。容器条码还可以包括已知的引物位点,所述引物位点在其例如通过退火与转录物附接或标记之前能够由反应混合物中用于扩增容器条码的正向和反向引物识别。示例性容器条码引物如SEQ ID NO:4和SEQ ID NO:5或者SEQ ID NO:10和SEQ ID NO:11所示。在一些情况下,将含有相同引物位点但在简并部分中发生碱基位移的容器条码池(如SEQ ID NO:80、81、83、99或100中任何一个所示的两个或更多个容器条码)提供给容器以产生标记的多核苷酸产物,所述多核苷酸产物发生了碱基位移以增加测序期间的多样性。在一些实施方案中,含有容器条码的寡核苷酸是含有通用引物位点(例如P7)的衔接子(如第一衔接子)的一部分。如本文所述的第一衔接子可以包括通用引发位点和容器条码。含有容器条码的此类寡核苷酸(包括通用引物位点、简并部分和引物)的例子如SEQ ID NO:2、6、7、8或9所示,或者是SEQ ID NO:6、7、8或9中任何两个或更多个的池。在一些情况下,容器条码或容器条码池可以用于标记在同一容器中处理的cDNA分子。在特定例子中,在同一容器中处理的cDNA分子与来自同一细胞的RNA分子互补。

在一些实施方案中,分子条码包括简并序列以独特地标记来自逆转录反应的多核苷酸转录物。在一些方面,分子条码是模板转换寡核苷酸的一部分,由此模板转换寡核苷酸包括逆转录酶的模板序列,使得分子条码被掺入每个互补多核苷酸中。在一些实施方案中,模板转换寡核苷酸可以含有(1)与含有容器条码的第一衔接子的3'标记多核苷酸互补的5'末端区域、(2)分子条码和(3)与3'悬突互补的3'部分。在一些实施方案中,模板转换分子(如含有条码(例如,分子条码)的模板转换寡核苷酸)可以掺入经修饰的碱基以最小化伪影形成(artifact formation)。含有示例性分子条码的示例性模板转换寡核苷酸如SEQ ID NO:3所示。

可以在3'标记多核苷酸的存在下进行逆转录反应,如上文所述的那些。3'标记多核苷酸可以是如下多核苷酸,所述多核苷酸用于将核酸添加到选定多核苷酸(如靶cDNA)的3'末端,或添加到与细胞中的转录组的转录物互补的多核苷酸(例如,cDNA)中。3'标记多核苷酸可以是用作模板以将核酸添加到靶多核苷酸(如cDNA)的3'末端的多核苷酸。3'标记多核苷酸可以是与靶多核苷酸(如cDNA)的3'末端杂交的多核苷酸。3'标记多核苷酸可以是含有与靶多核苷酸(如cDNA)的3'末端杂交的3'区域(如3'末端区域)的多核苷酸。例如,3'标记多核苷酸可以包含区段,如与三个或更多个非模板残基退火的区段。在一些实施方案中,3'标

记多核苷酸是分子条码多核苷酸。在一些实施方案中,3'标记多核苷酸可以包含分子条码。在一些实施方案中,3'标记多核苷酸可以包含与由逆转录酶产生的链(例如,序列CCC)互补并退火的3'末端上的3'核糖鸟苷残基或其类似物(rGrGrG)(RNA碱基)。在一些实施方案中,可以使用三个或更多个鸟嘌呤残基代替核糖鸟苷残基(DNA核苷酸代替RNA核苷酸)。在一些实施方案中,3'标记多核苷酸可以包含与由逆转录酶产生的链(例如,CCC)互补并退火的3'末端上的1个或2个核糖鸟苷残基和3'末端上的核糖鸟苷残基或其类似物(rGrGG)。

在3'标记多核苷酸与cDNA链的CCC退火后,逆转录酶可以继续将cDNA延伸到标记多核苷酸中,从而将分子条码或其互补体附接至反应中的靶多核苷酸群体(如cDNA)。例如,3'标记多核苷酸可以是含有与靶多核苷酸的3'末端杂交的在3'区域5'的区域的多核苷酸。与靶多核苷酸的3'末端杂交的在3'区域5'的区域可以包含不与靶多核苷酸(如cDNA)互补的区域。与靶多核苷酸的3'末端杂交的在3'区域5'的区域可以包含分子条码。与靶多核苷酸的3'末端杂交的在3'区域5'的区域可以包含与容器条码化多核苷酸或其互补体互补的区域。在其他实验中,模板转换可以在单独的反应中进行。例如,可以在逆转录反应之后添加3'标记多核苷酸,并且可以使用诸如逆转录酶或聚合酶等酶来延伸到标记多核苷酸中。因为标记多核苷酸可以在容器中的每个分子上具有独特的简并分子条码,所以可以用分子条码独特地标记容器中的每个cDNA。在一些实施方案中,可以在进行逆转录反应的同时进行模板转换。

在一些实施方案中,3'标记多核苷酸(如分子条码化多核苷酸)还可以包含与含有另一种条码(如容器条码)的3'标记多核苷酸或其互补体互补的5'区域,如5'末端区域。在一些实施方案中,含有分子条码或其互补体的靶多核苷酸(如标记的cDNA分子)可以包含与含有另一种条码(如容器条码)的3'标记多核苷酸或其互补体互补的3'区域,如3'末端区域。

在一些实施方案中,3'标记多核苷酸是容器条码化多核苷酸。在从靶多核苷酸产生含有分子条码或其互补体的多核苷酸后,可以将容器条码添加到分子条码化靶多核苷酸中。3'标记多核苷酸可以是用于将核酸添加到靶多核苷酸(如分子条码化靶多核苷酸)的3'末端的多核苷酸。3'标记多核苷酸可以是用作模板以将核酸添加到靶多核苷酸(如分子条码化靶多核苷酸)的3'末端的多核苷酸。3'标记多核苷酸可以是与靶多核苷酸(如分子条码化靶多核苷酸)的3'末端杂交的多核苷酸。3'标记多核苷酸可以是含有与靶多核苷酸(如分子条码化靶多核苷酸)的3'末端杂交的3'区域(如3'末端区域)的多核苷酸。容器条码化多核苷酸可以包含与分子条码化靶多核苷酸的3'末端杂交的3'区域,如3'末端区域。

在3'标记多核苷酸与分子条码化靶多核苷酸退火后,逆转录酶可以继续将cDNA延伸到3'标记多核苷酸(如容器条码化多核苷酸)中,从而将容器条码或其互补体附接至反应中的靶多核苷酸群体(如分子条码化靶多核苷酸)。例如,3'标记多核苷酸可以是含有与分子条码化靶多核苷酸的3'末端杂交的在3'区域5'的区域的多核苷酸。与分子条码化靶多核苷酸的3'末端杂交的在3'区域5'的区域可以包含不与靶多核苷酸或分子条码化靶多核苷酸互补的区域。与分子条码化靶多核苷酸的3'末端杂交的在3'区域5'的区域可以包含容器条码。

在一些实施方案中,3'标记多核苷酸是扩增产物。在一些实施方案中,3'标记多核苷酸是源自于单分子的扩增产物。在一些实施方案中,3'标记多核苷酸是容器条码化多核苷酸的扩增产物。在一些实施方案中,3'标记多核苷酸是源自于单容器条码化多核苷酸的扩增

产物。与分子条码化靶多核苷酸的3'末端杂交的在3'区域5'的区域可以包含与引物或其互补体互补的区域。与分子条码化靶多核苷酸的3'末端杂交的在3'区域5'的区域可以包含与用于扩增容器条码化多核苷酸的引物或其互补体互补的区域。

在一些实施方案中,3'标记多核苷酸可以充当用于扩增分子条码化cDNA的引物(如正向引物),并且DNA聚合酶可以延伸序列以产生与cDNA互补并且含有容器条码、分子条码和靶基因或转录物的编码序列的双条码化单链多核苷酸分子。在一些实施方案中,双条码化单链多核苷酸分子从5'至3'含有:容器条码、分子条码、靶基因或转录物的编码序列和第一衔接子。在一些实施方案中,含有容器条码的寡核苷酸含有第二衔接子,并且双条码化单链多核苷酸分子从5'至3'含有:第二衔接子、容器条码、分子条码、靶基因或转录物的编码序列和第一衔接子。

可以例如通过PCR扩增将由逆转录产生的标记的cDNA扩增一次或多次。特殊设计的各种引物可以用于扩增。可以使用第二扩增反应(如第一或第二PCR阶段)扩增第一扩增反应(如PCR)的产物。各种引物可以用于扩增步骤。可以使用本文所述的方法产生扩增的多核苷酸文库。在一些例子中,所得的文库可以包含具有适当分子条码和容器条码的完整或部分抗体或TCR序列。所述文库还可以含有与具有适当分子条码和容器条码的部分或完整转录组的一个或多个转录物对应的序列。

然后可以例如通过PCR扩增双条码化靶多核苷酸,如含有分子条码和容器条码的cDNA。然后可以例如通过使用引物组进行PCR。然后可以将上述PCR反应的产物扩增一次或多次,例如通过一轮或多轮PCR,或者直接测序。在一些实施方案中,引物组可以包括对第一衔接子和/或第二衔接子具有特异性的正向引物和/或反向引物。在一些实施方案中,引物组可以包括含有容器条码的寡核苷酸作为正向引物和对第一衔接子具有特异性的反向引物。在一些实施方案中,引物组可以包括如下引物组,其包括结合第二衔接子和第一衔接子的正向引物和反向引物以及含有容器条码的寡核苷酸作为另外的引物(如另外的正向引物)。在本文的实施例中描述了示例性引物。

在测序时,具有相同分子条码的序列可以匹配或配对。在测序时,具有相同容器条码的序列可以匹配或配对。在测序时,具有相同靶序列的序列可以匹配或配对。在一些实施方案中,可以将测序读段分解成共有序列。将匹配或配对的测序读段分解成共有序列从而可以减少或消除测序和PCR错误。对于第一读段,可以使用第一引物位点进行测序。对于第二读段,可以使用第二引物位点进行测序。对于第三读段,可以使用第三引物位点进行测序。

在一些实施方案中,可以将含有相同容器条码的免疫受体的链(如TCR或抗体的链)配对。在一些情况下,可以将含有相同容器条码的抗体重链和轻链配对。在一些实施方案中,可以将配对的链克隆到哺乳动物载体系统中。免疫受体(如抗体)构建体可以在其他人或哺乳动物宿主细胞系中表达。然后,可以通过瞬时转染测定和表达的兴趣抗体或TCR的Western印迹分析来验证构建体。

在某些方面,本发明提供了制备独特条码化重链和轻链抗体序列和/或 $\alpha$ 链和 $\beta$ 链TCR序列和/或 $\gamma$ 链和 $\delta$ 链TCR序列的文库的方法,其包括获得多个核酸构建体,其中每个构建体包括独特N聚体和功能性N聚体。功能性N聚体可以是随机N聚体、PCR引物、通用引物、抗体、粘性末端或任何其他序列。所述方法可以包括制备M个组的N个数量的流体区室,所述流体区室各自含有一个或多个拷贝的独特构建体。所述方法可以通过将另外的构建体添加到组中

的每个区室并为每个组重复所述过程以产生M个区室(每个区室含有独特的构建体对)来产生更高复杂度的条码文库。可以将所述对杂交或连接以产生新的构建体。在条码文库中的每个构建体中,每个独特的N聚体可以适于通过测序、探针杂交、其他方法或方法组合进行鉴定。

扩增RNA或DNA的方法是熟知的,并且可以基于本文呈现的教导和指导根据本发明使用,而无需进行过多的实验。DNA或RNA扩增的已知方法包括但不限于聚合酶链反应(PCR)和相关的扩增过程(参见例如,授予Mullis等人的美国专利号4,683,195、4,683,202、4,800,159、4,965,188;授予Tabor等人的美国专利号4,795,699和4,921,794;授予Innis的美国专利号5,142,033;授予Wilson等人的美国专利号5,122,464;授予Innis的美国专利号5,091,310;授予Gyllensten等人的美国专利号5,066,584;授予Gelfand等人的美国专利号4,889,818;授予Silver等人的美国专利号4,994,370;授予Biswas的美国专利号4,766,067;授予Ringold的美国专利号4,656,134)以及使用靶序列的反义RNA作为双链DNA合成的模板的RNA介导的扩增(授予Malek等人的美国专利号5,130,238,商品名为NASBA),将所述参考文献的全部内容通过引用并入本文(参见例如,Current Protocols in Molecular Biology (CPMB) (Fred M. Ausubel等人编辑,John Wiley and Sons, Inc.) ;或J. Sambrook等人,“Molecular Cloning:A Laboratory Manual,”1989,第2版,Cold Spring Harbour Laboratory Press:New York,N.Y.)。

方便地,可以按照或不按照采用其上固定了多种底物(例如,抗原等)的固相(如阵列)的多重测定形式进行本文所述的方法步骤(如扩增、测序等)。在一些实施方案中,阵列是蛋白质生物芯片。使用蛋白质生物芯片,可以筛选数百甚至数千的抗原。如本文所用,“阵列”、“微阵列”或“生物芯片”是指具有附接了吸附剂的大致平坦表面的固体基底。通常,生物芯片的表面包含多个可寻址位置,每个所述位置都有吸附剂结合在此。生物芯片可以适于接合探针接口,并且因此充当探针。“蛋白质生物芯片”是指适于捕获多肽的生物芯片。本领域中描述了许多蛋白质生物芯片。产生多肽阵列的方法描述于例如De Wildt等人,2000,Nat.Biotechnol.18:989-994;Lueking等人,1999,Anal.Biochem.270:103-1 11;Ge,2000,Nucleic Acids Res.28,e3,1-VH;MacBeath和Schreiber,2000,Science 289:1760-1763;WO 01/40803和WO 99/51773A1中。阵列的使用允许许多步骤(如筛选)自动和/或以高通量的方式进行。可以例如使用可商购获得的自动装置(例如,来自Genetic MicroSystems或BioRobotics)高速点样用于阵列的多肽。阵列基底可以是例如硝化纤维素、塑料、玻璃(例如,表面改性的玻璃)。阵列还可以包括多孔基质,例如丙烯酰胺、琼脂糖或另一种聚合物。

在捕获到生物芯片上后,可以通过多种检测方法来检测分析物,所述检测方法选自例如气相离子光谱法、光学法、电化学法、原子力显微镜检查和射频法。特别感兴趣的是使用质谱法,特别是SELDI。光学法包括例如检测荧光、发光、化学发光、吸光度、反射率、透射率、双折射或折射率(例如,表面等离子体共振、椭圆偏振术、共振镜法、光栅耦合波导法或干涉法)。光学法包括显微镜检查(共聚焦和非共聚焦二者)、成像法和非成像法。各种形式的免疫测定(例如,ELISA)是检测固相上捕获的分析物的流行方法。电化学法包括伏安法和安培法。射频法包括多极共振波谱法。

在本发明的一些实施方案中,采用已经建立用于处理单细胞或选择特定细胞群的技术。一种示例性技术结合了特殊的附件,所述附件可以在FACS中使用以将单细胞偏转到单

独的容器中。此类附件是可商购获得的并且在本领域中是熟知的。此类附件可用于将单细胞分配到例如标准96孔微量滴定培养板的选定区室中。可替代地,可以按有限稀释度将细胞沉积到微量滴定板中,以确保单细胞沉积。

第二种技术是对单个免疫细胞进行PCR以扩增V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>区段,加上对单个免疫细胞的转录组进行任选扩增。在一些实施方案中,将单细胞PCR用于保留单细胞中V<sub>L</sub>和V<sub>H</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>的天然配对。抗体或TCR的特异性分别由V<sub>L</sub>区和V<sub>H</sub>区内的互补决定区(CDR)、或者V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>区确定。

进行单细胞PCR的方法是熟知的(例如,Larrick,J.W.等人,Bio/Technology 7:934(1989))。例如,可以用固定液或含有诸如甲醛、戊二醛等的化学品的溶液固定来自B细胞文库的产生抗体的B细胞或来自T细胞文库的产生TCR的T细胞。然后,用包含例如洗涤剂的透化溶液来透化细胞。固定和透化过程应当提供足够的孔隙度,以允许酶、核苷酸和其他试剂进入细胞而不会过度破坏其中的细胞区室或核酸。然后,酶和核苷酸的添加可以进入细胞以将细胞mRNA(包括V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>mRNA)逆转录成例如相应的cDNA序列。在其他例子中,如本文所述,可以在溶液中进行来自裂解的非固定细胞的单细胞PCR。

可以使用逆转录酶、足够能量的四种dNTP以及与为逆转录酶提供3'羟基基团以启动聚合反应的mRNA结合的引物在单个步骤中或任选地与PCR程序一起进行逆转录。靶特异性引物和/或随机六聚体寡核苷酸引物可以用于启动逆转录反应并产生高质量的测序文库。

对于靶序列,可以使用与靶mRNA互补的任何引物,但优选使用与V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>分子的3'末端互补的引物以便促进可变区mRNA的选择。大量研究已经表明,可以制备简并多核苷酸来用作V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>的5'末端引物。制备靶向分子的组合文库方法依赖于此类引物。此外,大量实验已经表明,PCR可以从单细胞中扩增感兴趣的基因区段,如V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>。因为能够处理甚至单细胞,因此即使在感兴趣的免疫细胞以低频率出现的情况下,这种PCR方法也可以产生抗体。

在一些实施方案中,在FACS分选之后,将免疫细胞文库的细胞合并,并对整个细胞池进行逆转录-PCR。出于克隆抗体或TCR目的而产生mRNA通过用于制备和表征抗体或TCR的熟知的程序容易地实现(参见例如,Antibodies:A Laboratory Manual,1988;通过引用并入本文)。例如,通过本领域中标准和常规的适当方法从B细胞文库中提取总RNA。然后,通过适当的方法,例如使用随机六聚体多核苷酸、或C基因或C基因家族特异性引物、或V基因或V基因家族特异性引物,从RNA合成cDNA。同样,如上所解释,这些是本领域技术人员已知的过程。可以将从B细胞或T细胞文库(例如,从此类B或T淋巴细胞衍生的RNA或cDNA分子的文库)衍生的核酸分子的文库克隆到表达载体中以形成表达文库。在一些实施方案中,仅扩增从免疫细胞文库衍生的V<sub>H</sub>或V<sub>α</sub>或V<sub>γ</sub>结构域,以产生V<sub>H</sub>或V<sub>α</sub>或V<sub>γ</sub>结构域的文库。使用本文所述的方法,将来自另一个来源的V<sub>L</sub>或V<sub>β</sub>或V<sub>δ</sub>文库与V<sub>H</sub>或V<sub>α</sub>或V<sub>γ</sub>文库组合使用以产生抗体或TCR。可以通过以任何数量的如熟练技术人员已知的方式将V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>文库组合在一起构建抗体或TCR片段的文库。例如,可以在不同的载体中产生每个文库,并且将所述载体在体外或在体内重组。可替代地,可以将文库依次克隆到同一载体中,或者通过PCR组装在一起,然后克隆。PCR组装也可以用于将V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>DNA与编码柔性肽间隔子的DNA连接以形成单链Fv(scFv)文库,如本文其他地方所述。在又另一种技术中,将细胞内PCR组装用于通过PCR将淋巴细胞内的V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>基因组合,然后克隆

连接基因的库。

### 1. 靶多核苷酸

在实施方案中,本文提供的方法涉及对靶多核苷酸分子(如来自细胞的多核苷酸分子)的扩增和测序。在一些情况下,本文提供的方法涉及对靶多核苷酸分子的两个或更多个区域的扩增和测序。在一些情况下,本文提供的方法涉及对两种或更多种靶多核苷酸分子(如两种或更多种天然配对分子)的扩增和测序。在一方面,靶多核苷酸是RNA。在一方面,靶多核苷酸是基因组核酸。从特定生物染色体中的遗传材料衍生的DNA可以是基因组DNA。

在一些实施方案中,提及“靶核酸分子”、“靶多核苷酸”、“靶多核苷酸分子”是指任何感兴趣的核酸。

在一些实施方案中,靶多核苷酸包括包含免疫受体(如由免疫细胞产生的抗体或TCR)的可变区的序列。

在一些实施方案中,靶多核苷酸包括免疫受体的两条或更多条链,所述两条或更多条链天然配对以产生免疫受体或其结合片段。在一些实施方案中,靶多核苷酸包括包含由免疫细胞产生的抗体的重链可变区的序列。在一些实施方案中,靶多核苷酸包括包含由免疫细胞产生的抗体的轻链可变区的序列。在一些实施方案中,靶多核苷酸包括包含由同一免疫细胞产生的抗体的重链可变区的序列和包含可变轻链的序列。

因此,本文的术语“抗体”在最广泛的意义上使用,并且包括多克隆和单克隆抗体,包括完整抗体及其功能性(抗原结合)抗体片段,包括片段抗原结合(Fab)片段、F(ab')2片段、Fab'片段、Fv片段、重组IgG(rIgG)片段、单链抗体片段(包括单链可变片段(scFv))和单结构域抗体(例如, sdAb、sdFv、纳米抗体)片段。所述术语涵盖免疫球蛋白的基因工程化的和/或以其他方式修饰的形式,如胞内抗体、肽体(peptibody)、嵌合抗体、全人抗体、人源化抗体和异缀合抗体(heteroconjugate antibody)、多特异性(例如,双特异性)抗体、双抗体、三抗体和四抗体、串联二-scFv、串联三-scFv。除非另有说明,否则术语“抗体”应当理解为涵盖其功能性抗体片段。所述术语还涵盖完整或全长抗体,包括任何类别或亚类(包括IgG及其亚类、IgM、IgE、IgA和IgD)的抗体。

术语“互补决定区”和“CDR”与“高变区”或“HVR”同义,在本领域中是已知的,指代抗体可变区内的氨基酸的非连续序列,其赋予抗原特异性和/或结合亲和力。通常,在每个重链可变区(CDR-H1、CDR-H2和CDR-H3)中有三个CDR,并且在每个轻链可变区(CDR-L1、CDR-L2和CDR-L3)中有三个CDR。“框架区”和“FR”在本领域中是已知的,指代重链和轻链的可变区的非CDR部分。通常,在每个全长重链可变区(FR-H1、FR-H2、FR-H3和FR-H4)中有四个FR,并且在每个全长轻链可变区(FR-L1、FR-L2、FR-L3和FR-L4)中有四个FR。

给定CDR或FR的精确氨基酸序列边界可以使用许多熟知的方案中的任何一种容易地确定,所述方案包括由以下参考文献描述的那些:Kabat等人(1991),“Sequences of Proteins of Immunological Interest,”第5版Public Health Service, National Institutes of Health, Bethesda, MD(“Kabat”编号方案);Al-Lazikani等人,(1997) JMB 273, 927-948(“Chothia”编号方案);MacCallum等人,J.Mol.Biol.262:732-745(1996),“Antibody-antigen interactions:Contact analysis and binding site topography,”J.Mol.Biol.262, 732-745.”(“Contact”编号方案);Lefranc MP等人,“IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig

superfamily V-like domains,”Dev Comp Immunol, 2003年1月;27 (1):55-77 (“IMGT”编号方案);以及Honegger A和Plückthun A, “Yet another numbering scheme for immunoglobulin variable domains:an automatic modeling and analysis tool,”J Mol Biol, 2001年6月8日;309 (3):657-70 (“Aho”编号方案)。

因此,除非另有规定,否则应当理解给定抗体或其区域(如其可变区)的“CDR”或“互补决定区”或单个指定的CDR(例如,“CDR-H1、CDR-H2)涵盖如由任何上述方案所定义的一个(或特定的)互补决定区。例如,在声明特定的CDR(例如,CDR-H3)含有给定V<sub>H</sub>或V<sub>L</sub>氨基酸序列中的相应CDR的氨基酸序列的情况下,应理解,这样的CDR具有在可变区内的相应CDR(例如,CDR-H3)的序列,如由任何上述方案所定义的。在一些实施方案中,规定了指定的CDR序列。

同样,除非另有规定,否则应当理解给定抗体或其区域(如其可变区)的FR或单独指定的一个或多个FR(例如,FR-H1、FR-H2)涵盖如由任何已知方案所定义的一个(或特定的)框架区。在一些情形中,规定了用于鉴定特定的CDR、FR或者FR或CDR的方案,如通过Kabat、Chothia或Contact方法定义的CDR。在其他情况下,给出了CDR或FR的特定氨基酸序列。

术语“可变区”或“可变结构域”是指抗体重链或轻链的参与抗体与抗原的结合的结构域。天然抗体的重链和轻链的可变结构域(分别为V<sub>H</sub>和V<sub>L</sub>)通常具有相似的结构,每个结构域包含四个保守的框架区(FR)和三个CDR。(参见例如,Kindt等人Kuby Immunology,第6版,W.H.Freeman and Co.,第91页(2007))。单个V<sub>H</sub>或V<sub>L</sub>结构域可以足以赋予抗原结合特异性。此外,可以使用来自结合抗原的抗体的V<sub>H</sub>或V<sub>L</sub>结构域来分离结合特定抗原的抗体,以分别筛选互补的V<sub>L</sub>或V<sub>H</sub>结构域的文库。参见例如,Portolano等人,J. Immunol. 150:880-887 (1993); Clarkson等人,Nature 352:624-628 (1991)。

所提供的抗体包括抗体片段。“抗体片段”是指不同于完整抗体的分子,其包含完整抗体的结合完整抗体所结合的抗原的一部分。抗体片段的例子包括但不限于Fv、Fab、Fab'、Fab'-SH、F(ab')<sub>2</sub>;双抗体;线性抗体;单链抗体分子(例如,scFv);和从抗体片段形成的多特异性抗体。在特定实施方案中,抗体是包含可变重链区和/或可变轻链区的单链抗体片段,如scFv。

由B细胞表达的免疫球蛋白(Ig)在一些方面是由四条多肽链(两条重链(IgH)和两条轻链(IgL))组成的蛋白质,从而形成H2L2结构。每对IgH和IgL链含有由VL和VH区组成的高变结构域以及恒定结构域。Ig的IgH链具有几种类型:μ、δ、γ、α和β。Ig在个体内的多样性主要由高变结构域决定。与TCR相似,IgH链的V结构域是通过VH、DH和JH基因区段的组合连接产生的。在Ig基因重排过程期间在VH-DH、DH-JH和VH-JH连接处独立添加和缺失核苷酸进一步增加了高变结构域序列多样性。这里,免疫能力反映在Ig的多样性上。

术语“可变区”或“可变结构域”是指抗体重链或轻链的参与抗体与抗原的结合的结构域。天然抗体的重链和轻链的可变结构域(分别为V<sub>H</sub>和V<sub>L</sub>)通常具有相似的结构,每个结构域包含四个保守的框架区(FR)和三个CDR。(参见例如,Kindt等人Kuby Immunology,第6版,W.H.Freeman and Co.,第91页(2007))。单个V<sub>H</sub>或V<sub>L</sub>结构域可以足以赋予抗原结合特异性。此外,可以使用来自结合抗原的抗体的V<sub>H</sub>或V<sub>L</sub>结构域来分离结合特定抗原的抗体,以分别筛选互补的V<sub>L</sub>或V<sub>H</sub>结构域的文库。参见例如,Portolano等人,J. Immunol. 150:880-887 (1993); Clarkson等人,Nature 352:624-628 (1991)。

“高变区”是指抗体或TCR的负责抗原结合的氨基酸残基。高变区包含来自互补决定区

或CDR的氨基酸残基。框架或FR残基是除如本文定义的高变区残基之外的那些可变结构域残基。

所提供的抗体包括抗体片段。“抗体片段”是指不同于完整抗体的分子,其包含完整抗体的结合完整抗体所结合的抗原的一部分。抗体片段的例子包括但不限于Fv、Fab、Fab'、Fab'-SH、F(ab')2;双抗体;线性抗体;单链抗体分子(例如,scFv);和从抗体片段形成的多特异性抗体。在特定实施方案中,抗体是包含可变重链区和/或可变轻链区的单链抗体片段,如scFv。

单结构域抗体是包含抗体的全部或部分重链可变结构域或者全部或部分轻链可变结构域的抗体片段。在某些实施方案中,单结构域抗体是人单结构域抗体。

抗体片段可以通过各种技术制备,包括但不限于完整抗体的蛋白水解消化以及通过重组宿主细胞产生。在一些实施方案中,抗体是重组产生的片段,如包含天然不存在的排列的片段(如具有通过合成接头(例如,肽接头)连接的两个或更多个抗体区或链的那些),和/或可以不通过酶消化天然存在的完整抗体产生的片段。在一些方面,抗体片段是scFv。

抗原结合多肽还包括重链二聚体,例如像来自骆驼和鲨鱼的抗体。骆驼和鲨鱼抗体包含V样和C样结构域(都不具有轻链)的两条链的同二聚体对。由于骆驼中重链二聚体IgG的VH区不必与轻链发生疏水性相互作用,因此重链中通常与轻连接触的区域被变成骆驼中的亲水性氨基酸残基。重链二聚体IgG的VH结构域被称为VHH结构域。鲨鱼Ig-NAR包含一个可变结构域(称为V-NAR结构域)和五个C样恒定结构域(C-NAR结构域)的同二聚体。在骆驼中,抗体库的多样性由VH或VHH区中的CDR 1、2和3决定。骆驼VHH区中的CDR3的特征在于其相对较长的长度,平均为16个氨基酸(Muyldermans等人,1994,Protein Engineering 7 (9) : 1129)。

“人源化”抗体是如下抗体,其中所有或基本上所有CDR氨基酸残基都衍生自非人CDR并且所有或基本上所有FR氨基酸残基都衍生自人FR。人源化抗体任选地可以包括衍生自人抗体的抗体恒定区的至少一部分。非人抗体的“人源化形式”是指非人抗体的变体,其经历人源化以通常降低对人的免疫原性,同时保留亲本非人抗体的特异性和亲和力。在一些实施方案中,人源化抗体中的一些FR残基被来自非人抗体(例如,衍生出CDR残基的抗体)的相应残基取代,例如以恢复或改善抗体特异性和亲和力。

所提供的抗体包括人抗体。“人抗体”是具有与人或人细胞或者利用人抗体库或其他人抗体编码序列(包括人抗体文库)的非人来源产生的抗体的氨基酸序列对应的氨基酸序列的抗体。所述术语不包括包含非人抗原结合区的非人抗体的人源化形式,如所有或基本上所有CDR都是非人的那些。

人抗体可以通过将免疫原给予转基因动物来制备,所述转基因动物已被修饰以响应于抗原激发产生完整人抗体或具有人可变区的完整抗体。此类动物通常含有人免疫球蛋白基因座的全部或部分,其替代内源免疫球蛋白基因座,或者其存在于在染色体外或随机整合到动物的染色体中。在此类转基因动物中,通常已经使内源免疫球蛋白基因座失活。人抗体也可以衍生自含有衍生自人库的抗体编码序列的人抗体文库,包括噬菌体展示和无细胞文库。

所提供的抗体包括单克隆抗体,包括单克隆抗体片段。如本文所用的术语“单克隆抗体”是指从基本上同质抗体的群体(即,构成所述群体的单独抗体是相同的,但含有天然存

在的突变或在单克隆抗体制剂的产生期间产生的可能变体除外,此类变体通常以少量存在)获得或在所述群体内的抗体。与通常包括针对不同表位的不同抗体的多克隆抗体制剂相反,单克隆抗体制剂的每种单克隆抗体针对抗原上的单个表位。所述术语不应解释为需要通过任何特定方法产生抗体。单克隆抗体可以通过多种技术制备,包括但不限于从杂交瘤产生、重组DNA方法、噬菌体展示和其他抗体展示方法。

在一些实施方案中,靶多核苷酸包括包含由免疫细胞产生的TCR的 $\alpha$ 链的可变区的序列。在一些实施方案中,靶多核苷酸包括包含由免疫细胞产生的TCR的 $\beta$ 链的可变区的序列。在一些实施方案中,靶多核苷酸包括包含由同一免疫细胞产生的TCR的 $\alpha$ 链的可变区的序列和包含TCR的 $\beta$ 链的可变区的序列。在一些实施方案中,靶多核苷酸包括包含由免疫细胞产生的TCR的 $\gamma$ 链的可变区的序列。在一些实施方案中,靶多核苷酸包括包含由免疫细胞产生的TCR的 $\delta$ 链的可变区的序列。在一些实施方案中,靶多核苷酸包括包含由同一免疫细胞产生的TCR的 $\gamma$ 链的可变区的序列和包含TCR的 $\delta$ 链的可变区的序列。

在一些实施方案中,TCR涵盖完整的TCR及其抗原结合部分或抗原结合片段(也称为MHC-肽结合片段)。在一些实施方案中,TCR是完整或全长TCR。在一些实施方案中,TCR是抗原结合部分,所述抗原结合部分小于全长TCR,但是与结合至MHC分子(即,在MHC分子的背景下)的特定抗原肽(即,MHC-肽复合物)结合。在一些情况下,TCR的抗原结合部分或片段可以仅含有全长或完整TCR的结构性结构域的一部分,但是仍能够结合与完整TCR结合的表位(例如,MHC-肽复合物)。在一些情况下,TCR的抗原结合部分或片段含有TCR的可变结构域(如TCR的可变 $\alpha$ 链和可变 $\beta$ 链),其足以形成与特定MHC-肽复合物结合的结合位点,例如通常其中每条链含有三个互补决定区。包括具有结合结构域的多肽或蛋白质,所述结合结构域是抗原结合结构域或与抗原结合结构域同源。这些术语还涵盖互补决定区(CDR)枝接的抗体和TCR以及其他人源化抗体和TCR(包括CDR修饰和框架区修饰)。应当注意,尽管可以仅参考免疫球蛋白链(例如,重链和轻链),但是所公开的发明可以应用于多种其他不同类型的配对序列(例如,T细胞受体链对(TCR $\alpha$ 和TCR $\beta$ 链以及TCR $\gamma$ 和TCR $\delta$ 链)),并且不限于免疫球蛋白。

T细胞识别与各种癌症或感染性生物相关的抗原的能力由其TCR赋予,所述TCR由阿尔法( $\alpha$ )链和贝塔( $\beta$ )链或者伽马( $\gamma$ )和德尔塔( $\delta$ )链构成。构成这些链的蛋白质由DNA编码,所述DNA采用独特的机制来产生TCR的巨大多样性。这种多亚基免疫识别受体与CD3复合物缔合并结合由MHC I类和II类蛋白在抗原呈递细胞(APC)表面上呈递的肽。TCR与APC上的抗原肽的结合是T细胞激活中的主要事件,其发生在T细胞与APC之间的接触点的免疫突触处。

每个TCR包含可变互补决定区(CDR)以及框架区(FR)。 $\alpha$ 和 $\beta$ 链可变结构域的第三个互补决定区(CDR3)环的氨基酸序列分别在很大程度上决定了由 $\beta$ 链基因座中的可变(V $\beta$ )、多样性(D $\beta$ )和接合(J $\beta$ )基因区段之间、以及 $\alpha$ 链基因座中的类似V $\alpha$ 和J $\alpha$ 基因区段之间的重组引起的 $\alpha$ BT细胞的序列多样性。TCR $\alpha$ 和 $\beta$ 链基因座中多个此类基因区段的存在允许对大量不同的CDR3序列进行编码。在TCR基因重排过程期间在V $\beta$ -D $\beta$ 、D $\beta$ -J $\beta$ 和V $\alpha$ -J $\alpha$ 连接处独立添加和缺失核苷酸进一步增加了CDR3序列多样性。在这方面,免疫能力反映在TCR的多样性上。

还提供了TCR片段,包括抗原结合片段。在一些实施方案中,TCR是其抗原结合部分,如不含有其一个或多个跨膜和/或胞质区的全长TCR变体,其可以被称为完全可溶性TCR。在一些实施方案中,TCR是二聚体TCR(dTCR)。在一些实施方案中,TCR是单链TCR(scTCR),如具有

如PCT专利公开号WO 03/020763、WO 04/033685或WO 2011/044186中所述的结构的scTCR。在某些实施方案中,TCR是包含与 $\beta$ 链可变区连接的 $\alpha$ 链可变区的单链TCR片段,如scTv。在一些实施方案中,scTv也被称为scFv。

在一些方面,单链Fv或scFv是指包含抗体的可变重链(VH)和可变轻链(VL)结构域或者TCR的可变 $\alpha$ 或 $\gamma$ 链(V $\alpha$ 或V $\gamma$ )和可变 $\beta$ 或 $\delta$ 链(V $\beta$ 或V $\delta$ )结构域的抗体或TCR片段,其中这些结构域存在于单条多肽链中。通常,Fv多肽还包含在VH与VL结构域之间或V $\alpha$ 与V $\beta$ 结构域之间或V $\gamma$ 与V $\delta$ 结构域之间的多肽接头,所述多肽接头使得scFv能够形成所需的抗原结合结构。

在一些方面,双抗体是指具有两个抗原结合位点的小抗体和/或TCR片段,所述片段包含与同一多肽链中的VL连接的VH(VH-VL)或与同一多肽中的V $\beta$ 连接的V $\alpha$ (V $\alpha$ -V $\beta$ )或与同一多肽链中的V $\delta$ 连接的V $\gamma$ (V $\gamma$ -V $\delta$ )。通过使用过短而不允许在同一链上的两个结构域之间配对的接头,迫使结构域与另一条链的互补结构域配对,并产生两个抗原结合位点。在例如EP404097和WO 93111161中更充分地描述了示例性双抗体。

在一些方面,双特异性抗体或双特异性TCR是指对两种不同类型的抗原显示出特异性的抗体或TCR。如本文所用的术语具体地包括但不限于显示对靶抗原的结合特异性和与促进递送至特定组织的另一种靶的结合特异性的抗体和TCR。类似地,多特异性抗体和TCR具有两种或更多种结合特异性。

在一些方面,线性抗体或“线性TC是指形成一对抗原结合区的一对串联Fd区段(例如,VH-CH1-VH-CH1或V $\alpha$ -Ca1-V $\alpha$ -Ca1)。线性抗体和TCR可以是双特异性的或单特异性的,例如如Zapata等人,Protein Eng. 8 (10) :1057-1062 (1995)所述。

在一些方面,抗原结合结构域是指抗体或TCR的保留特异性结合抗原的能力的一个或多个片段。在此类术语内包括的抗体片段的非限制性例子包括但不限于:(i) Fab片段,即由VL、VH、CL和CH1结构域组成的单价片段;(ii) F(ab')2片段,即含有通过铰链区的二硫键连接的两个Fab片段的二价片段;(iii) 由VH和CH1结构域组成的Fd片段;(iv) 含有抗体单臂的VL和VH结构域的Fv片段,包括scFv;(v) dAb片段(Ward等人,(1989) Nature 341:544546),其含有VH结构域;和(vi) 分离的CDR。此外,在此定义中包括包含单条重链和单条轻链的抗体或具有单条 $\alpha$ 链或单条 $\beta$ 链的TCR。

“F(ab')2”和“Fab”部分可以通过用诸如胃蛋白酶和木瓜蛋白酶等蛋白酶处理Ig来产生,并且包括通过在两条重链的每一条中的铰链区之间存在的二硫键附近消化免疫球蛋白而产生的抗体片段。例如,木瓜蛋白酶在两条重链的每一条中的铰链区之间存在的二硫键上游切割IgG,以产生两个同源抗体片段,其中由VL和CL构成的轻链和由VH和CH $\gamma$ 1(重链的恒定区中的 $\gamma$ 1区域)构成的重链片段通过二硫键在其C末端区域连接。这两个同源抗体片段中的每一个都被称为'Fab'。胃蛋白酶还在两条重链的每一条中的铰链区之间存在的二硫键下游切割IgG,以产生比两个上述'Fab'在铰链区连接的片段稍大的抗体片段。此抗体片段被称为F('ab')2。Fab片段还含有轻链的恒定结构域和重链的第一恒定结构域(CH1)。'Fab'片段与Fab片段的不同之处在于重链CH1结构域的羧基末端处的几个残基的添加,包括来自抗体铰链区的一个或多个半胱氨酸。Fab'-SH是本文中对恒定结构域的一个或多个半胱氨酸残基带有游离巯基基团的Fab'的名称。F(ab')2抗体片段最初是作为Fab'片段对产生的,所述Fab'片段对具有在它们之间的铰链半胱氨酸。

在一些方面,Fv是指含有完整的抗原识别和抗原结合位点的抗体或TCR片段。此区域由

紧密的非共价缔合的一个重链和一个轻链可变结构域或者一条TCR $\alpha$ 链和一条TCR $\beta$ 链或者一条TCR $\gamma$ 链和一条TCR $\delta$ 链的二聚体组成。在此构型中,每个可变结构域的三个CDR相互作用以将抗原结合位点限定在VH-VL二聚体或V $\alpha$ -V $\beta$ 二聚体或V $\gamma$ -V $\delta$ 二聚体的表面上。共同地,来自V $H$ 和V $L$ 链或V $\alpha$ 和V $\beta$ 链或V $\gamma$ 和V $\delta$ 链中的每一个的一个或多个CDR的组合为抗体或TCR赋予了抗原结合特异性。例如,应理解,例如当转移至受体选定抗体、TCR或其抗原结合片段的V $H$ 和V $L$ 链或V $\alpha$ 和V $\beta$ 链或V $\gamma$ 和V $\delta$ 链时,CDRH3和CDRL3可能足以使抗体或TCR赋予抗原结合特异性,并且可以针对结合、亲和力等对这种CDR组合进行测试。甚至单个可变结构域(或Fv的仅包含对抗原具有特异性的三个CDR的一半)具有识别和结合抗原的能力,但是可能低于当与第二可变结构域结合时的亲和力。此外,尽管Fv片段的两个结构域(VL和VH或V $\alpha$ 和V $\beta$ 或V $\gamma$ 和V $\delta$ )由单独的基因编码,但是它们可以使用重组方法通过合成接头连接,所述合成接头使得它们能够成为单条蛋白质链,其中VL和VH或V $\alpha$ 和V $\beta$ 或V $\gamma$ 和V $\delta$ 链区配对以形成单价分子(被称为单链Fv(scFv);Bird等人(1988)Science 242:423-426;Huston等人(1988)Proc.Natl.Acad.Sci.USA 85:5879-5883;以及Osbourne等人(1998)Nat.Biotechnol.16:778)。此类scFv也旨在涵盖在术语抗体的“抗原结合部分”内。可以将特定scFv的任何VH和VL序列与Fc区cDNA或基因组序列连接,以便产生编码完整Ig(例如,IgG)分子或其他同种型的表达载体。在使用蛋白质化学或重组DNA技术产生Fab、Fv或其他Ig片段中也可以使用VH和VL。

“种系序列”是指来自种系(单倍体配子和形成它们的那些二倍体细胞)的遗传序列。种系DNA含有编码单条Ig重链或轻链、或单条TCR $\alpha$ 或TCR $\beta$ 链、或单条TCR $\gamma$ 或TCR $\delta$ 链的多个基因区段。这些基因区段携带在生殖细胞中,但是不能进行转录和翻译,直至它们被排列成功能性基因。在骨髓中的B细胞和T细胞分化期间,通过能够产生超过108种特异性的动态遗传系统将这些基因区段随机改组。这些基因区段中的大多数都由种系数据库发布和收集。

在一些实施方案中,免疫分子可以是或可能是中和抗体或中和TCR。在一些方面,中和抗体或TCR是抑制病原体(如病毒或细菌)复制的抗体或TCR,而不管实现中和的机制。

在一些实施方案中,样品(如细胞群或单细胞)可以含有免疫库,例如抗体库或TCR库,并且这样的免疫库可以通过所提供的方法进行阐明。在一些实施方案中,抗体库或TCR库是指抗体、TCR或其片段的集合。在一些实施方案中,抗体库可以例如用于选择特定抗体或筛选特定特性,如结合能力、结合特异性、胃肠道运输能力、稳定性、亲和力等。所述术语具体地包括抗体和TCR文库,包括所有形式的组合文库,例如像抗体噬菌体展示文库,包括但不限于来自任何来源的单链Fv(scFv)和Fab抗体噬菌体展示文库,包括初始、合成和半合成文库。

靶多核苷酸可以从几乎任何来源获得,并且可以使用本领域已知的方法制备。例如,可以使用本领域已知的方法直接分离靶多核苷酸而无需扩增,所述方法包括但不限于从生物或细胞(例如,免疫细胞)中提取基因组DNA或mRNA的片段以获得靶多核苷酸。靶多核苷酸还可以涵盖通过逆转录-PCR从RNA(如mRNA)产生的cDNA。在一些情况下,靶多核苷酸是RNA分子。在一些情况下,靶多核苷酸是mRNA分子或从所述mRNA分子产生的cDNA。在一些情况下,靶多核苷酸是来自单免疫细胞的mRNA分子或从所述mRNA分子产生的cDNA分子。在一些情况下,靶多核苷酸是来自单个免疫细胞的mRNA分子或从所述mRNA分子产生的cDNA分子。在一些情况下,靶多核苷酸是来自单免疫细胞的编码抗体序列的mRNA分子。在一些情况下,靶多

核苷酸是来自单个免疫细胞的编码重链抗体序列的mRNA分子。在一些情况下,靶多核苷酸是来自单个免疫细胞的编码重链抗体序列的mRNA分子。在一些情况下,靶多核苷酸是来自单个免疫细胞的编码轻链抗体序列的mRNA分子。在一些情况下,靶多核苷酸是来自单个免疫细胞的编码轻链抗体序列的mRNA分子。在一些情况下,靶多核苷酸是来自单个免疫细胞的编码抗体可变序列的mRNA分子。在一些情况下,靶多核苷酸是来自单个免疫细胞的编码可变抗体序列的mRNA分子。在一些情况下,靶多核苷酸是来自单个免疫细胞的编码可变轻链抗体序列的mRNA分子。在一些情况下,靶多核苷酸是来自单个免疫细胞的编码可变轻链抗体序列的mRNA分子。在一些情况下,靶多核苷酸是来自单个免疫细胞的编码可变重链抗体序列的mRNA分子。在一些情况下,靶多核苷酸是来自单个免疫细胞的编码可变重链抗体序列的mRNA分子。在一些情况下,靶多核苷酸可以是无细胞的核酸,例如DNA或RNA。在一些情况下,靶多核苷酸是来自单个免疫细胞的编码可变 $\alpha$ 、 $\beta$ 、 $\gamma$  和/或 $\delta$ 链TCR序列的mRNA分子。

本文所述的方法可以用于从一个或多个靶多核苷酸产生多核苷酸文库用于测序。靶多核苷酸包括不是扩增反应产物的任何感兴趣的多核苷酸。例如,靶多核苷酸可以包括生物样品中的多核苷酸。例如,靶多核苷酸不包括PCR反应的产物。例如,靶多核苷酸可以包括用于产生扩增反应产物的多核苷酸模板,但不包括扩增产物本身。例如,靶多核苷酸可以包括用于产生逆转录反应或引物延伸反应的产物的多核苷酸模板,并且还包括逆转录反应或引物延伸反应产物本身。例如,靶多核苷酸包括可以进行逆转录反应或引物延伸反应的感兴趣的多核苷酸。例如,靶多核苷酸包括RNA或DNA。例如,靶多核苷酸包括cDNA。在一些实施方案中,靶RNA多核苷酸是mRNA。在一些实施方案中,靶RNA多核苷酸被聚腺苷酸化。在一些实施方案中,RNA多核苷酸未被聚腺苷酸化。在一些实施方案中,靶多核苷酸是DNA多核苷酸。DNA多核苷酸可以是基因组DNA。DNA多核苷酸可以包含外显子、内含子、非翻译区或其任何组合。

在一些实施方案中,文库可以从靶多核苷酸的两个或更多个区域产生。在一些实施方案中,方法文库可以从两个或更多个靶多核苷酸产生。在一些实施方案中,靶多核苷酸是基因组核酸或衍生自染色体的DNA。在一些实施方案中,靶多核苷酸包括包含变体(如多态性或突变)的序列。在一些实施方案中,靶多核苷酸包括DNA而非RNA。在一些实施方案中,靶多核苷酸包括RNA而非DNA。在一些实施方案中,靶多核苷酸包括DNA和RNA。在一些实施方案中,靶多核苷酸是mRNA分子。在一些实施方案中,靶多核苷酸是DNA分子。在一些实施方案中,靶多核苷酸是单链多核苷酸。在一些实施方案中,靶多核苷酸是双链多核苷酸。在一些实施方案中,靶多核苷酸是双链多核苷酸的单链。

靶多核苷酸可以从任何生物样品获得,并且可以使用本领域已知的方法制备。在一些实施方案中,将靶多核苷酸直接分离而无需扩增。直接分离的方法在本领域中是已知的。非限制性例子包括从生物样品、生物或细胞中提取基因组DNA或mRNA。

在一些实施方案中,一个或多个靶多核苷酸是从生物样品中纯化的。在一些实施方案中,靶多核苷酸不是从含有其的生物样品中纯化的。在一些实施方案中,靶多核苷酸是从生物样品中分离的。在一些实施方案中,靶多核苷酸不是从含有其的生物样品中分离的。在一些实施方案中,靶多核苷酸可以是无细胞核酸。在一些实施方案中,靶多核苷酸可以是片段化核酸。在一些实施方案中,靶多核苷酸可以是经转录的核酸。在一些实施方案中,靶多核苷酸是经修饰的多核苷酸。在一些实施方案中,靶多核苷酸是未经修饰的多核苷酸。

在一些实施方案中，靶多核苷酸是来自单细胞的多核苷酸。在一些实施方案中，靶多核苷酸来自单个细胞。在一些实施方案中，靶多核苷酸是来自含有多个细胞的样品的多核苷酸。

在一些实施方案中，靶多核苷酸编码生物标记序列。在一些实施方案中，靶多核苷酸编码两个或更多个生物标记序列。在一些实施方案中，多个靶多核苷酸编码生物标记序列。在一些实施方案中，多个靶多核苷酸编码两个或更多个生物标记序列。在一些实施方案中，多个靶多核苷酸编码3个、4个、5个、6个、7个、8个、9个、10个、20个、30个、40个、50个、60个、70个、80个、90个或100个或更多个生物标记序列。

$10^{10}$ 个、 $5 \times 10^{10}$ 个、 $6 \times 10^{10}$ 个、 $7 \times 10^{10}$ 个、 $8 \times 10^{10}$ 个、 $9 \times 10^{10}$ 个、 $1 \times 10^{11}$ 个、 $2 \times 10^{11}$ 个、 $3 \times 10^{11}$ 个、 $4 \times 10^{11}$ 个、 $5 \times 10^{11}$ 个、 $6 \times 10^{11}$ 个、 $7 \times 10^{11}$ 个、 $8 \times 10^{11}$ 个、 $9 \times 10^{11}$ 个、 $1 \times 10^{12}$ 个、 $2 \times 10^{12}$ 个、 $3 \times 10^{12}$ 个、 $4 \times 10^{12}$ 个、 $5 \times 10^{12}$ 个、 $6 \times 10^{12}$ 个、 $7 \times 10^{12}$ 个、 $8 \times 10^{12}$ 个或 $9 \times 10^{12}$ 个免疫球蛋白或TCR序列。在一些实施方案中,一组免疫球蛋白或TCR序列含有约10-20个、10-30个、10-40个、10-30个、10-40个、10-50个、10-60个、10-70个、10-80个、10-90个、10-100个、50-60个、50-70个、50-80个、50-90个、50-100个、100-200个、100-300个、100-400个、100-300个、100-400个、100-500个、100-600个、100-700个、100-800个、100-900个、100-1000个、500-600个、500-700个、500-800个、500-900个、500-1000个、1000-2000个、1000-3000个、1000-4000个、1000-3000个、1000-4000个、1000-5000个、1000-6000个、1000-7000个、1000-8000个、1000-9000个、1000-10000个、5000-6000个、5000-7000个、5000-8000个、5000-9000个、5000-10000个、 $1-1 \times 10^5$ 个、 $1-2 \times 10^5$ 个、 $1-3 \times 10^5$ 个、 $1-4 \times 10^5$ 个、 $1-5 \times 10^5$ 个、 $1-6 \times 10^5$ 个、 $1-7 \times 10^5$ 个、 $1-8 \times 10^5$ 个、 $9 \times 10^5$ 个、 $1-1 \times 10^6$ 个、 $1-2 \times 10^6$ 个、 $1-3 \times 10^6$ 个、 $1-4 \times 10^6$ 个、 $1-5 \times 10^6$ 个、 $1-6 \times 10^6$ 个、 $1-7 \times 10^6$ 个、 $1-8 \times 10^6$ 个、 $9 \times 10^6$ 个、 $1-1 \times 10^7$ 个、 $1-2 \times 10^7$ 个、 $1-3 \times 10^7$ 个、 $1-4 \times 10^7$ 个、 $1-5 \times 10^7$ 个、 $1-6 \times 10^7$ 个、 $1-7 \times 10^7$ 个、 $1-8 \times 10^7$ 个、 $1-9 \times 10^7$ 个、 $1-1 \times 10^8$ 个、 $1-2 \times 10^8$ 个、 $1-3 \times 10^8$ 个、 $1-4 \times 10^8$ 个、 $1-5 \times 10^8$ 个、 $1-6 \times 10^8$ 个、 $1-7 \times 10^8$ 个、 $1-8 \times 10^8$ 个、 $1-9 \times 10^8$ 个、 $1-1 \times 10^9$ 个、 $1-2 \times 10^9$ 个、 $1-3 \times 10^9$ 个、 $1-4 \times 10^9$ 个、 $1-5 \times 10^9$ 个、 $1-6 \times 10^9$ 个、 $1-7 \times 10^9$ 个、 $1-8 \times 10^9$ 个、 $1-9 \times 10^9$ 个、 $1-1 \times 10^{10}$ 个、 $1-2 \times 10^{10}$ 个、 $1-3 \times 10^{10}$ 个、 $1-4 \times 10^{10}$ 个、 $1-5 \times 10^{10}$ 个、 $1-6 \times 10^{10}$ 个、 $1-7 \times 10^{10}$ 个、 $1-8 \times 10^{10}$ 个、 $1-9 \times 10^{10}$ 个、 $1-1 \times 10^{11}$ 个、 $1-2 \times 10^{11}$ 个、 $1-3 \times 10^{11}$ 个、 $1-4 \times 10^{11}$ 个、 $1-5 \times 10^{11}$ 个、 $1-6 \times 10^{11}$ 个、 $1-7 \times 10^{11}$ 个、 $1-8 \times 10^{11}$ 个、 $1-9 \times 10^{11}$ 个、 $1-1 \times 10^{12}$ 个、 $1-2 \times 10^{12}$ 个、 $1-3 \times 10^{12}$ 个、 $1-4 \times 10^{12}$ 个、 $1-5 \times 10^{12}$ 个、 $1-6 \times 10^{12}$ 个、 $1-7 \times 10^{12}$ 个、 $1-8 \times 10^{12}$ 个或 $1-9 \times 10^{12}$ 个免疫球蛋白或TCR序列。

在一些实施方案中，靶多核苷酸在长度上为约10个、20个、30个、40个、50个、60个、70个、80个、90个、100个、150个、200个、250个、300个、350个、400个、450个、500个、550个、600个、650个、700个、750个、800个、850个、900个、1000个、1500个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个、10,000个、11,000个、12,000个、13,000个、14,000个、15,000个、16,000个、17,000个、18,000个、19,000个或20,000个碱基或碱基对。在一些实施方案中，靶多核苷酸在长度上为至少约10个、20个、30个、40个、50个、60个、70个、80个、90个、100个、150个、200个、250个、300个、350个、400个、450个、500个、550个、600个、650个、700个、750个、800个、850个、900个、1000个、1500个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个、10,000个、11,000个、12,000个、13,000个、14,000个、15,000个、16,000个、17,000个、18,000个、19,000个或20,000个碱基或碱基对。在一些实施方案中，靶多核苷酸在长度上为至多约10个、20个、30个、40个、50个、60个、70个、80个、90个、100个、150个、200个、250个、300个、350个、400个、450个、500个、550个、600个、650个、700个、750个、800个、850个、900个、1000个、1500个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个、10,000个、11,000个、12,000个、13,000个、14,000个、15,000个、16,000个、17,000个、18,000个、19,000个或20,000个碱基或碱基对。在一些实施方案中，靶多核苷酸在长度上为约10-20个、10-30个、10-40个、10-30个、10-40个、10-50个、10-60个、10-70个、10-80个、10-90个、10-100个、50-60个、50-70个、50-80个、50-90个、50-100个、100-200个。

个、100-300个、100-400个、100-300个、100-400个、100-500个、100-600个、100-700个、100-800个、100-900个、100-1000个、500-600个、500-700个、500-800个、500-900个、500-1000个、1000-2000个、1000-3000个、1000-4000个、1000-3000个、1000-4000个、1000-5000个、1000-6000个、1000-7000个、1000-8000个、1000-9000个、1000-10000个、5000-6000个、5000-7000个、5000-8000个、5000-9000个或5000-10000个碱基或碱基对。在一些实施方案中，靶多核苷酸或其片段的平均长度可以小于约100个、200个、300个、400个、500个或800个碱基对，或小于约5个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个、110个、120个、130个、140个、150个、160个、170个、180个、190个或200个核苷酸，或小于约1、2、5、10、20、30、40、50、60、70、80、90、100千碱基。在一些实施方案中，来自相对较短的模板(如含有靶多核苷酸的样品)的靶序列为约40个、45个、50个、55个、60个、65个、70个、75个、80个、85个、90个、95个或100个碱基。在某些实施方案中，使用含有与疾病或病症相关的序列或免疫球蛋白或TCR序列的数据库，将测序数据与已知或预期序列进行比对。

## 2. 多核苷酸的集合，例如转录组

与基因组或转录组多核苷酸对应的多核苷酸的集合可以从几乎任何来源(如一个细胞或多个细胞)获得，并且可以使用本领域已知的方法来制备。例如，可以使用本领域已知的方法从单细胞或多个细胞直接分离多核苷酸的集合而无需扩增，所述方法包括但不限于从生物或细胞(例如，免疫细胞)中提取基因组DNA或mRNA的片段以获得多核苷酸的集合。基因组或转录组多核苷酸的集合还可以涵盖通过逆转录-PCR从RNA(如mRNA)产生的cDNA。在一些情况下，多核苷酸的集合是RNA分子的集合。在一些情况下，多核苷酸的集合是mRNA分子的集合或由所述mRNA分子产生的cDNA分子的集合。在一些情况下，多核苷酸的集合是来自单免疫细胞的mRNA分子的集合或由所述mRNA分子产生的cDNA分子的集合。在一些情况下，多核苷酸的集合是来自单个免疫细胞的mRNA分子的集合或从所述mRNA分子产生的cDNA分子的集合。

本文所述的方法可以用于产生含有来自一个或多个细胞的多核苷酸的集合的文库用于测序。多核苷酸的集合可以衍生自基因组DNA或RNA，如来自生物样品的一个或多个细胞的mRNA转录物。例如，基因组DNA或细胞RNA(如mRNA)可以用作模板，以产生扩增反应(如逆转录反应或引物延伸反应)的产物。在一些例子中，可以从cDNA产生多肽的集合。在一些实施方案中，多核苷酸的集合是从作为mRNA的RNA多核苷酸产生的，并且所述集合基本上代表来自生物样品的一个或多个细胞的转录组。在一些实施方案中，多核苷酸的集合是从被聚腺苷酸化的RNA多核苷酸产生的。在一些实施方案中，RNA多核苷酸未被聚腺苷酸化。在一些实施方案中，多核苷酸的集合是从DNA多核苷酸产生的。DNA多核苷酸可以是基因组DNA。DNA多核苷酸可以包含外显子、内含子、非翻译区或其任何组合。例如，本发明的多核苷酸的集合可以含有至少5个、10个、100个、250个、500个、750个、1,000个、2,500个、5,000个、10,000个、25,000个、50,000个、75,000个、10,000个、250,000个、500,000个、750,000个、1,000,000个、2,500,000个、5,000,000个、7,500,000个或10,000,000个细胞亚组或单个细胞(如表达不同抗体或TCR的免疫细胞亚组或单个免疫细胞)的基因组或转录组信息。

在一些实施方案中，可以使用随机六聚体引物通过逆转录酶或引物延伸反应产生多核苷酸的集合。在一些实施方案中，可以使用针对聚A核苷酸序列的引物通过逆转录酶或引物延伸反应产生多核苷酸的集合。在一些例子中，可以使用寡聚dT通过逆转录酶或引物延伸

反应产生多核苷酸的集合。在一些例子中,可以将引物生物素化。任选地,可以在逆转录或引物延伸反应后纯化多核苷酸集合。例如,可以任选地使用链霉亲和素纯化技术纯化使用生物素化引物产生的多核苷酸的集合。在其他实施方案中,可以通过亲和纯化、琼脂糖凝胶电泳中的一种或多种纯化多核苷酸。

### 3. 微滴文库

通常,微滴文库由在单个集合中合并在一起的许多文库元件构成。从单个文库元件到 $1 \times 10^{15}$ 个文库元件或更多个文库元件,文库的复杂度可能不同。每个文库元件都是在固定浓度下的一种或多种给定组分。元件可以是但不限于细胞、珠、氨基酸、蛋白质、多肽、核酸、多核苷酸或小分子化学化合物。元件可以含有标识符,如分子条码、容器条码或两者。

细胞文库元件可以包括但不限于杂交瘤、B细胞、T细胞、原代细胞、培养细胞系、癌细胞、干细胞或任何其他细胞类型。通过将从一千到数万个的许多细胞封装在单个微滴中来制备细胞文库元件。通常通过泊松统计从细胞的数量密度和微滴的体积给出被封装的细胞的数量。然而,在一些情况下,所述数量会偏离泊松统计,如在Edd等人,“Controlled encapsulation of single-cells into monodisperse picoliter drops.”Lab Chip, 8 (8):1262-1264, 2008中所述。细胞的离散性质允许使用多个细胞变体(如各自产生一种抗体或TCR的免疫细胞)大规模制备文库,所有细胞变体都存在于单一起始介质中,然后将所述介质分解成含有至多一个细胞的单个容器(如微滴或胶囊)。然后裂解单个容器(例如,微滴或胶囊)内的细胞,并且将来自裂解细胞的释放到容器内的多核苷酸(如包括靶mRNA或DNA在内的细胞mRNA和基因组DNA(例如,重链和轻链多核苷酸和/或 $\alpha$ 和 $\beta$ 链多核苷酸和/或 $\gamma$ 和 $\delta$ 链多核苷酸))用分子条码和容器条码进行条码化并进行扩增。然后,将双条码化多核苷酸产物组合或合并以形成由转录组或基因组和靶(例如,重链和轻链和/或 $\alpha$ 和 $\beta$ 链和/或 $\gamma$ 和 $\delta$ 链)文库元件组成的文库。具体地,将转录组和靶文库合并。

基于珠的文库元件含有一个或多个珠,并且还可以含有其他试剂,如抗体、酶或其他蛋白质。在所有文库元件含有不同类型的珠,但周围介质相同的情况下,文库元件可以全部由单一起始流体制备或具有多种起始流体。在从变体的集合大规模制备的细胞文库的情况下,文库元件将从多种起始流体制备。当以多个细胞开始时,期望每个微滴恰好具有一个细胞,而只有少数微滴含有超过一个细胞。在一些情况下,可以实现相对于泊松统计的变化以提供增强的微滴负载,使得有更多微滴恰好具有一个细胞/微滴,而空微滴或含有超过一个细胞的微滴的例外很少。

在一些实施方案中,当以多个容器条码化多核苷酸开始时,期望每个微滴恰好具有一个容器条码化多核苷酸,只有少数微滴含有超过一个容器条码化多核苷酸。在一些情况下,可以实现相对于泊松统计的变化以提供增强的微滴负载,使得有更多微滴恰好具有一个容器条码化多核苷酸/微滴,而空微滴或含有超过一个容器条码化多核苷酸的微滴的例外很少。

微滴文库的例子是具有不同内容物的微滴的集合,所述内容物的范围为珠、细胞、小分子、DNA、引物、抗体和条码化多核苷酸。微滴的大小范围为直径大约0.5微米至500微米,这对应于约1皮升至1纳升。然而,微滴可以小至5微米和大至500微米。优选地,微滴的直径小于100微米,为约1微米至约100微米。最优选的大小为直径约20至40微米(10至100皮升)。检查的微滴文库的优选特性包括渗透压平衡、均匀大小和大小范围。

本发明提供的微滴文库内包含的微滴的大小优选地是均匀的。也就是说，当与同一文库内其他微滴的直径相比时，文库内任何微滴的直径将变化小于5%、4%、3%、2%、1%或0.5%。文库中微滴的均匀大小对于维持微滴的稳定性和完整性可能是关键的，并且对于随后将文库内的微滴用于本文所述的各种生物学和化学测定可能也是至关重要的。

本发明提供了微滴文库，所述微滴文库包含在不混溶流体内的多个水性微滴，其中每个微滴的大小优选地是基本上均匀的并且包含不同的文库元件。本发明提供了用于形成微滴文库的方法，其包括提供包含不同文库元件的单一水性流体，将每个文库元件封装到不混溶流体内的水性微滴中。

在某些实施方案中，将不同类型的元件（例如，细胞或珠）合并到在同一介质中包含的单一来源中。在初始合并之后，然后将元件封装在微滴中以产生微滴文库，其中具有不同类型的珠或细胞的每个微滴都是不同的文库元件。初始溶液的稀释使得能够进行封装过程。在一些实施方案中，所形成的微滴将含有单个元件，或将不含有任何元件，即是空的。在其他实施方案中，所形成的微滴将含有文库元件的多个拷贝。被封装的元件通常是一种类型的变体。在一个例子中，元件是血液样品的免疫细胞，并且封装每个免疫细胞以扩增免疫细胞中的核苷酸的抗体序列并将其条码化。

例如，在一种类型的乳剂文库中，存在具有不同颗粒（即，在不同介质中的细胞或条码化多核苷酸）并且在合并之前被封装的文库元件。在一个例子中，在不同的介质内含有指定数量的文库元件，即n个数量的不同细胞或条码化多核苷酸。每个文库元件分别被乳化和合并，这时，n个数量的合并的不同文库元件中的每一个都被组合和合并到单个池中。所得的池含有多个油包水乳剂微滴，每个微滴含有不同类型的颗粒。

在一些实施方案中，所形成的微滴将含有单个文库元件，或将不含有任何元件，即是空的。在其他实施方案中，所形成的微滴将含有文库元件的多个拷贝。珠的含量遵循泊松分布，其中存在离散的概率分布，所述概率分布表示如果许多事件以已知的平均比率并且独立于自上次事件以来的时间而发生，在固定时间段内这些事件发生的概率。用于产生文库的油和表面活性剂阻止了微滴之间文库内容物的交换。

#### B. 产生单细胞双条码化多核苷酸文库的方法

在一些实施方案中，提供了用于产生多核苷酸文库的方法，所述方法包括以下步骤：  
(a) 裂解多个容器中的每一个内的细胞，其中所述容器中的每一个包含来自包含细胞群的样品的细胞、多个分子条码化寡核苷酸和包含容器条码化寡核苷酸的第一衔接子；(b) 在每个容器中产生多个单链多核苷酸，所述多个单链多核苷酸包含 (i) 与所述细胞中存在的一一个或多个靶多核苷酸互补的一个或多个靶单链多核苷酸；以及 (ii) 单链多核苷酸的集合，所述单链多核苷酸各自与所述细胞中的多核苷酸互补；(c) 将所述多个分子条码化寡核苷酸之一附接至每个单链多核苷酸，从而产生各自包含独特分子条码的多个条码化单链多核苷酸；(d) 将包含所述容器条码化寡核苷酸的第一衔接子或其扩增产物附接至所述条码化单链多核苷酸中的每一个，从而产生多个双条码化单链多核苷酸，其中在同一容器中的所述双条码化单链多核苷酸中的每一个包含相同的容器条码；以及 (e) 向所述双条码化单链多核苷酸中的每一个添加第二衔接子，其中所述第一衔接子和所述第二衔接子存在于所述双条码化单链多核苷酸中的每一个的相对末端处或附近。产生了多核苷酸文库的示例性容器包括孔、乳剂、微滴或微胶囊。

## 1. 样品制备

含有多核苷酸的任何生物样品(包括含有细胞群的样品)都可以用于本文所述的方法中。通常,含有细胞的任何样品均可以用于本文所述的方法中。例如,样品可以是如下生物样品,所述生物样品来自受试者或来自由其衍生的含有RNA或DNA的样品。可以从生物样品中提取多核苷酸,或者可以直接将样品进行所述方法而无需提取或纯化多核苷酸。可以对样品提取或分离DNA或RNA。样品也可以是从生物标本、cDNA文库、病毒或基因组DNA中提取的总RNA或DNA。在一个实施方案中,从生物样品中分离多核苷酸,所述生物样品含有多种其他组分,如蛋白质、脂质和非模板核酸。核酸模板分子可以从获得自动物、植物、细菌、真菌或任何其他细胞生物的任何细胞材料获得。

在某些实施方案中,多核苷酸从单细胞(如存在于细胞群中的细胞)获得。多核苷酸可以直接从生物获得或从获得自生物的生物样品获得。任何组织或体液标本都可以用作在本发明中使用的核酸的来源。还可以从培养的细胞(如原代细胞培养物或细胞系)中分离多核苷酸。在一些实施方案中,细胞可以是血细胞、免疫细胞、组织细胞或肿瘤细胞。在一些实施方案中,细胞是免疫细胞,如B细胞或T细胞。B细胞可以是浆母细胞、记忆B细胞或浆细胞。从其获得模板核酸的细胞或组织可能感染有病毒或其他细胞内病原体。

在一些实施方案中,可以从受试者或宿主的血液或其他生物样品中分离细胞群(如含有免疫细胞的群体),所述受试者或宿主例如为人或其他动物,如已进行免疫或正患有感染、癌症、自身免疫病症或任何其他疾病的人或其他动物。在一些实施方案中,所述人可能被诊断出患有疾病,展现出疾病的症状,未被诊断出患有疾病,或未展现出疾病的症状。在一些实施方案中,所述受试者或宿主(例如,人类受试者)可以是暴露于和/或可以产生TCR的受试者或宿主,所述TCR针对感染原(例如,病毒、细菌、寄生虫、朊病毒等)、抗原、疾病或与疾病或病症相关的抗原(例如,肿瘤相关抗原)。在一些实施方案中,免疫细胞可以来自含有T细胞的任何生物样品,所述T细胞例如为PBMC、脾脏或其他淋巴器官中存在的细胞。在一些实施方案中,免疫细胞来自正常健康受试者的T细胞来源。在一些实施方案中,免疫细胞来自患病受试者的T细胞来源。在一些实施方案中,可以分离或获得CD4<sup>+</sup>或CD8<sup>+</sup>细胞。在一些情况下,可以分离或获得外周血单核细胞(PBMC)。在一些情况下,可以分离或获得肿瘤浸润性淋巴细胞(TIL)。

在某些实施方案中,可以从受试者或宿主的血液或其他生物样品中分离产生抗体或TCR的免疫细胞,以鉴定出具有潜在临床意义的病原体、肿瘤和/或疾病特异性抗体或TCR,所述受试者或宿主例如为人或其他动物,如已进行免疫或正患有感染、癌症、自身免疫病症或任何其他疾病的人或其他动物。例如,所述人可能被诊断出患有疾病,展现出疾病的症状,未被诊断出患有疾病,或未展现出疾病的症状。例如,所述人可以是暴露于和/或可以产生有用的抗体或TCR的人,所述抗体或TCR针对感染原(例如,病毒、细菌、寄生虫、朊病毒等)、抗原或疾病。例如,所述动物可以是暴露于和/或可以产生有用的抗体或TCR的动物,所述抗体或TCR针对感染原(例如,病毒、细菌、寄生虫、朊病毒等)、抗原或疾病。在一些例子中,所述动物(如人)不再展现出疾病或病症的症状。来自经免疫的宿主的某些免疫细胞产生针对一种或多种靶抗原和/或一种或多种未知抗原的抗体或TCR。在本发明中,在对抗体链进行测序、制备抗体或制备一个或多个表达文库之前,可以通过任何合适的方法(如使用荧光激活细胞分选(FACS)、磁激活细胞分选(MACS)筛选和分选细胞、淘选或其他筛选方法)

使淋巴细胞池富集所需的免疫细胞,以从样品(如免疫细胞文库)产生多个免疫细胞。与仅提供表达不同抗体的少数免疫细胞亚组,并且因此仅提供少数天然存在的可变结构域组合的现有技术富集方法相比,本发明的免疫细胞文库含有至少2个亚组的或单个的表达不同抗体或TCR的免疫细胞。例如,本发明的免疫细胞文库可以含有至少5个、10个、100个、250个、500个、750个、1,000个、2,500个、5,000个、10,000个、25,000个、50,000个、75,000个、10,000个、250,000个、500,000个、750,000个、1,000,000个、2,500,000个、5,000,000个、7,500,000个或10,000,000个亚组的或单个的表达不同抗体或TCR的免疫细胞。本发明的方法使免疫细胞回收最大化,并且提供非常高的多样性。

T细胞可以从许多来源获得,所述来源包括外周血单核细胞、骨髓、胸腺、组织活检、肿瘤、淋巴结组织、肠相关淋巴组织、粘膜相关淋巴组织、脾脏组织或任何其他淋巴组织以及肿瘤。T细胞可以从T细胞系以及从自体或同种异体来源获得。T细胞可以获得自单一个体或个体群体,例如全部患有相同疾病(如癌症或感染性疾病)的个体群体。在一些实施方案中,来自个体的循环血液的细胞通过单采术或白细胞单采术获得。单采术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。在一个实施方案中,可以将通过单采术或白细胞单采术收集的细胞洗涤,以除去血浆部分并将细胞置于适当的缓冲液或介质中进行后续处理步骤。在本发明的一个实施方案中,用磷酸盐缓冲盐水(PBS)洗涤细胞。在一个替代性实施方案中,洗涤溶液缺少钙,并且可能缺少镁,或者可能缺少许多(如果不是全部)二价阳离子。如本领域普通技术人员将容易理解的,洗涤步骤可以通过本领域技术人员已知的方法(例如通过使用半自动“流通式”离心机)来实现。

在洗涤之后,可以将细胞重悬于多种生物相容性缓冲液(例如像不含Ca<sup>++</sup>/Mg<sup>++</sup>的PBS)中。可替代地,可以去除单采术样品的不希望组分,并将细胞直接重悬于培养基中。在其他实施方案中,通过裂解红细胞并通过PERCOLL<sup>TM</sup>梯度离心从外周血淋巴细胞中分离T细胞。可以通过阳性或阴性选择技术进一步分离T细胞的特定亚群,如CD28<sup>+</sup>、CD4<sup>+</sup>、CD8<sup>+</sup>、CD45RA<sup>+</sup>和CD45RO<sup>+</sup> T细胞。例如,CD3<sup>+</sup>、CD28<sup>+</sup> T细胞可以使用CD3/CD28缀合的磁珠(例如,DYNABEADS<sup>®</sup> M-450 CD3/CD28 T细胞扩增器)进行阳性选择。

在一些实施方案中,可以使用针对对于阴性选择细胞而言独特的表面标记的抗体组合实现通过阴性选择对T细胞群的富集。一种这样的方法是使用针对在阴性选择的细胞上存在的细胞表面标记的单克隆抗体混合物经由负磁免疫粘附或流式细胞术进行细胞分选和/或选择。例如,为了通过阴性选择富集CD4<sup>+</sup>细胞,单克隆抗体混合物通常包括针对CD14、CD20、CD11b、CD16、HLA-DR和CD8的抗体。制备用于刺激的T细胞的另一种方法是在洗涤步骤之后冷冻细胞,这不需要单核细胞去除步骤。冷冻和随后的解冻步骤可以通过去除细胞群中的粒细胞和在一定程度上的单核细胞来提供更均匀的产物。在去除血浆和血小板的洗涤步骤之后,可以将细胞悬浮在冷冻溶液中。尽管许多冷冻溶液和参数在本领域中是已知的并且将在此背景下有用,但是一种方法涉及使用含有20%DMSO和8%人血清白蛋白(HSA)的PBS或其他合适的细胞冷冻介质。然后将其用介质1:1稀释,使得DMSO和HSA的最终浓度分别为10%和4%。然后将细胞以1°C/分钟的速率冷冻至-80°C并储存在液氮储罐的气相中。

在一些实施方案中,从样品中富集细胞群。在一些实施方案中,针对细胞的特定亚组或亚型富集细胞。在一些实施方案中,细胞群富集或含有T细胞或B细胞。在一些实施方案中,细胞群富集或含有CD4<sup>+</sup>或CD8<sup>+</sup>细胞。在一些实施方案中,细胞群富集或含有中枢记忆T细

胞、效应记忆T细胞、幼稚T细胞、干细胞样中枢记忆T细胞、效应T细胞和调节性T细胞。在一些实施方案中,细胞群富集或含有记忆B细胞、幼稚B细胞或浆母细胞B细胞。

在一些实施方案中,可以基于来自细胞的免疫受体对选定靶抗原或复合物的亲和力来选择免疫细胞。在一些方面,亲和力是指两种试剂的可逆结合的平衡常数,并表示为 $K_D$ 。结合蛋白对配体的亲和力(如抗体对表位的亲和力或者如TCR对MHC-肽复合物的亲和力)可以是例如从约100纳摩尔(nM)至约0.1nM、从约100nM至约1皮摩尔(pM)、或从约100nM至约1飞摩尔(fM)。术语“亲合力”是指在稀释之后,两种或更多种试剂的复合物对于解离的抗性。

在一些实施方案中,表位在一些方面是指抗原或其他大分子的能够与抗体或TCR的可变区结合口袋形成结合相互作用的一部分。此类结合相互作用可以表现为与一个或多个CDR的一个或多个氨基酸残基的分子间接触。抗原结合可以涉及例如CDR3、CDR3对、或在一些情形中 $V_H$ 和 $V_L$ 链的多达全部六个CDR的相互作用。表位可以是线性肽序列(即,“连续的”)或者可以由非连续的氨基酸序列构成(即,“构象的”或“不连续的”)。抗体或TCR可以识别一个或多个氨基酸序列;因此表位可以定义超过一个不同的氨基酸序列。在一些方面,TCR可以识别在MHC背景下的一个或多个氨基酸序列或表位。由抗体和TCR识别的表位可以通过本领域技术人员熟知的肽作图和序列分析技术确定。结合相互作用表现为与CDR的一个或多个氨基酸残基的分子间接触。

在一些实施方案中,提及免疫受体(如在免疫细胞上表达的,例如抗体或TCR),特异性结合是指抗体或TCR将与除含有由抗体或TCR识别的表位的抗原以外的分子不表现出任何显著结合的情境。所述术语也适用于例如抗原结合结构域对许多抗原携带的特定表位具有特异性的情况,在这种情况下,携带抗原结合结构域的选定抗体、TCR或其抗原结合片段将能够结合携带表位的各种抗原。

术语“优先结合”或“特异性结合”意指抗体、TCR或其片段以比其结合不相关的氨基酸序列更大的亲和力结合表位,并且,如果与含有所述表位的其他多肽发生交叉反应,在配制它们用于给予人类用途的水平下是无毒的。在一方面,这种亲和力比抗体、TCR或其片段对不相关的氨基酸序列的亲和力大至少1倍、大至少2倍、大至少3倍、大至少4倍、大至少5倍、大至少6倍、大至少7倍、大至少8倍、大至少9倍、大10倍、大至少20倍、大至少30倍、大至少40倍、大至少50倍、大至少60倍、大至少70倍、大至少80倍、大至少90倍、大至少100倍或大至少1000倍。术语“结合”是指在生理条件下由于例如共价、静电、疏水和离子和/或氢键相互作用在两个分子之间的直接缔合,并且包括诸如盐桥和水桥等相互作用以及任何其他常规结合方式。

在一些实施方案中,术语“结合”是指在生理条件下由于例如共价、静电、疏水和离子和/或氢键相互作用在两个分子之间的直接缔合,并且包括诸如盐桥和水桥等相互作用以及任何其他常规结合方式。

在一些实施方案中,可以基于来自细胞的免疫受体(例如,TCR)对四聚体或其他MHC-肽多聚体的亲和力来选择免疫细胞。在一些实施方案中,术语“四聚体”可以指代包含与链霉亲和素的单分子结合的四个亚基的复合物,其可以结合并因此鉴定细胞群。亚基可以是MHC-肽复合物。亚基可以是不具有相关肽的MHC。亚基可以是B细胞受体抗原。由四聚体鉴定的细胞群可以是表达与四聚体的亚基结合的受体(如TCR或BCR)的群体。细胞群可以是抗原特异性T细胞。细胞群可以是抗原特异性B细胞。四聚体可以是荧光标记的。如本文所用,

MHC-肽四聚体可以与pMHC互换使用。

在一些例子中,可以基于对亲和寡核苷酸缀合物的亲和力来选择免疫细胞(参见例如, WO 2017/053905)。可以通过本文所述的阳性或阴性选择技术进一步分离基于选定靶抗原或复合物(如亲和寡核苷酸缀合物)的结合或识别而选择的细胞。

在一些实施方案中,利用来自非免疫的人或非人供体的免疫细胞。动物的初始库(抗原激发之前的库)为动物提供了能以中等亲和力(约 $1 \times 10^{-6}$ 至 $1 \times 10^{-7}$  M的 $K_d$ )与基本上任何非自身分子结合的抗体或TCR。抗体或TCR结合位点的序列多样性不是直接在种系中编码,而是以组合方式从V基因区段组装的。免疫触发产生结合免疫原的 $V_H$ - $V_L$ 或 $V_{\alpha}$ - $V_{\beta}$ 或 $V_{\gamma}$ - $V_{\delta}$ 组合的任何免疫细胞,以增殖(克隆扩增)并分泌如上所述的相应的抗体。然而,使用来自未经免疫的受试者的脾脏细胞和/或免疫细胞或其他外周血淋巴细胞(PBL)可以提供可能的抗体或TCR库的更好代表,并且还允许使用任何动物物种来构建随后的B细胞或T细胞抗体或TCR文库。

在一些实施方案中,样品是唾液。在一些实施方案中,样品是全血。在一些实施方案中,为了获得足够量的多核苷酸进行测试,抽取至少约0.001mL、0.005mL、0.01mL、0.05mL、0.1mL、0.5mL、1mL、2mL、3mL、4mL、5mL、10mL、20mL、25mL、30mL、35mL、40mL、45mL或50mL的血液体积。在一些情况下,为了获得足够的核酸进行测试,抽取至少0.001mL、0.005mL、0.01mL、0.05mL、0.1mL、0.5mL、1mL、2mL、3mL、4mL、5mL、10mL、20mL、25mL、30mL、35mL、40mL、45mL或50mL的血液体积。

在一些情况下,起始材料是外周血。外周血细胞可以富集特定的细胞类型(例如,单核细胞;红细胞;CD4+细胞;CD8+细胞;免疫细胞;T细胞;NK细胞等)。外周血细胞也可以选择性地耗尽特定的细胞类型(例如,单核细胞;红细胞;CD4+细胞;CD8+细胞;免疫细胞;T细胞;NK细胞等)。

在一些情况下,起始材料可以是包含实体组织的组织样品,所述实体组织的非限制性例子包括皮肤、脑、肝、肺、肾、前列腺、卵巢、脾脏、淋巴结(包括扁桃体)、甲状腺、胰腺、心脏、骨骼肌、肠、喉、食管和胃。在其他情况下,起始材料可以是含有核酸的细胞、免疫细胞,特别是B细胞或T细胞。在一些情况下,起始材料可以是来自任何生物的含有核酸的样品,可以从其中获得遗传材料。在一些情况下,样品是流体,例如血液、唾液、淋巴液或尿液。

样品可以取自患有病症的受试者。在一些情况下,从其取得样品的受试者可以是患者,例如癌症患者或怀疑患有癌症的患者。受试者可以是哺乳动物,例如人,并且可以是雄性或雌性。在一些情况下,雌性是怀孕的。样品可以是肿瘤活检。活检可以由例如医疗保健提供者进行,所述医疗保健提供者包括医师、医师助理、护士、兽医、牙医、脊医、护理人员、皮肤科医生、肿瘤科医生、胃肠病医生或外科医生。

在一些情况下,可以使用酶促处理(如蛋白酶消化)从起始材料中去除非核酸材料。

在一些情况下,可以将血液收集到含有镁螯合剂(包括但不限于EDTA)的装置中,并且保存在4°C下。任选地,可以添加钙螯合剂,包括但不限于EGTA。在另一种情况下,将细胞裂解抑制剂添加到血液中,包括但不限于甲醛、甲醛衍生物、福尔马林、戊二醛、戊二醛衍生物、蛋白质交联剂、核酸交联剂、蛋白质和核酸交联剂、伯胺反应性交联剂、巯基反应性交联剂、巯基加成或二硫化物还原、碳水化合物反应性交联剂、羧基反应性交联剂、光反应性交联剂或可裂解交联剂。

在一些情况下,当提取的材料包含单链RNA、双链RNA或DNA-RNA杂交体时,可以使用本领域已知的技术将这些分子转化成双链DNA。例如,逆转录酶可以用于从RNA分子合成DNA。在一些情况下,将RNA转化成DNA可能需要先前连接步骤,以将接头片段连接至RNA,从而允许使用通用引物来启动逆转录。在其他情况下,例如,mRNA分子的聚A尾可以用于启动逆转录。在转化成DNA后,在一些情况下,本文详述的方法可以用于进一步捕获、选择、标记或分离所需的序列。

核酸分子包括脱氧核糖核酸(DNA)和/或核糖核酸(RNA)。核酸分子可以是合成的或衍生自天然存在的来源。在一个实施方案中,从生物样品中分离核酸分子,所述生物样品含有多种其他组分,如蛋白质、脂质和非模板核酸。核酸模板分子可以从获得自动物、植物、细菌、真菌或任何其他细胞生物的任何细胞材料获得。在某些实施方案中,核酸分子从单细胞获得。用于在本发明中使用的生物样品包括病毒颗粒或制剂。核酸分子可以直接从生物获得或从获得自生物的生物样品(例如,从血液、尿液、脑脊液、精液、唾液、痰、粪便和组织)获得。任何组织或体液标本都可以用作在本发明中使用的核酸的来源。还可以从培养的细胞(如原代细胞培养物或细胞系)中分离核酸分子。从其获得模板核酸的细胞或组织可能感染有病毒或其他细胞内病原体。

样品也可以是从生物标本、cDNA文库、病毒或基因组DNA中提取的总RNA。在某些实施方案中,核酸分子与其他靶分子(如蛋白质、酶、底物、抗体、结合剂、珠、小分子、肽或任何其他分子)结合。通常,可以通过多种技术从生物样品中提取核酸,所述技术例如为由以下参考文献所述的那些:Sambrook和Russell,Molecular Cloning:A Laboratory Manual,第三版,Cold Spring Harbor,N.Y.(2001)。核酸分子可以是单链的、双链的或具有单链区的双链的(例如,茎结构和环结构)。

DNA提取的方法在本领域中是熟知的。经典的DNA分离方案是基于使用有机溶剂(如苯酚和氯仿的混合物)提取,随后用乙醇沉淀(J.Sambrook等人,“Molecular Cloning:A Laboratory Manual,”1989,第2版,Cold Spring Harbour Laboratory Press:New York,N.Y.)。其他方法包括:盐析DNA提取(P.Sunnucks等人,Genetics,1996,144:747-756;S.M.Aljanabi等人,Nucl.Acids Res.1997,25:4692-4693)、三甲基溴化铵盐DNA提取(S.Gustincich等人,BioTechniques,1991,11:298-302)和硫氰酸胍DNA提取(J.B.W.Hammond等人,Biochemistry,1996,240:298-300)。多种试剂盒可商购获得用于从生物样品中提取DNA(例如,BD Biosciences Clontech(加利福尼亚州帕洛阿尔托(Palo Alto,CA));Epicentre Technologies(威斯康星州麦迪逊(Madison,WI));Gentra Systems,Inc.(明尼苏达州明尼阿波利斯(Minneapolis,MN));MicroProbe Corp.(华盛顿州波塞尔(Bothell,WA));Organon Teknika(北卡罗来纳州达勒姆(Durham,NC));和Qiagen Inc.(加利福尼亚州巴伦西亚(Valencia,CA)))。

RNA提取的方法在本领域中也是熟知的(例如,J.Sambrook等人,“Molecular Cloning:A Laboratory Manual”1989,第2版,Cold SpringHarbour Laboratory Press:New York),并且从体液中进行RNA提取的试剂盒是可商购获得的(例如,Ambion,Inc.(德克萨斯州奥斯汀(Austin,TX));Amersham Biosciences(新泽西州皮斯卡塔韦(Piscataway,NJ));BD Biosciences Clontech(加利福尼亚州帕洛阿尔托(Palo Alto,CA));BioRad Laboratories(加利福尼亚州赫拉克勒斯(Hercules,CA));Dynal Biotech Inc.(纽约州成

功湖(Lake Success, NY)) ; Epicentre Technologies(威斯康星州麦迪逊(Madison, WI)) ; Gentra Systems, Inc. (明尼苏达州明尼阿波利斯(Minneapolis, MN)) ; GIBCO BRL(马里兰州盖瑟斯堡(Gaithersburg, MD)) ; Invitrogen Life Technologies(加利福尼亚州卡尔斯巴德(Carlsbad, CA)) ; MicroProbe Corp. (华盛顿州波塞尔(Bothell, WA)) ; Organon Teknica(北卡罗来纳州达勒姆(Durham, NC)) ; Promega, Inc. (威斯康星州麦迪逊(Madison, WI)) ; 和Qiagen Inc. (加利福尼亚州巴伦西亚(Valencia, CA))。

一个或多个样品可以来自一个或多个来源。一个或多个样品可以来自两个或更多个来源。一个或多个样品可以来自一名或多名受试者。一个或多个样品可以来自两名或更多名受试者。一个或多个样品可以来自同一受试者。一名或多名受试者可以来自同一物种。一名或多名受试者可以来自不同物种。所述一名或多名受试者可以是健康的。所述一名或多名受试者可能受疾病、障碍或病症的影响。

在一些实施方案中,样品是流体,如血液、唾液、淋巴液、尿液、脑脊液、精液、痰、粪便或组织匀浆。

样品可以取自患有病症的受试者。在一些实施方案中,从其取得样品的受试者可以是患者,例如癌症患者或怀疑患有癌症的患者。受试者可以是哺乳动物,例如人,并且可以是雄性或雌性。在一些实施方案中,雌性是怀孕的。样品可以是肿瘤活检。活检可以由例如医疗保健提供者进行,所述医疗保健提供者包括医师、医师助理、护士、兽医、牙医、脊医、护理人员、皮肤科医生、肿瘤科医生、胃肠病医生或外科医生。

在一些实施方案中,多核苷酸与其他靶分子(如蛋白质、酶、底物、抗体、结合剂、珠、小分子、肽或任何其他分子)结合。在一些实施方案中,多核苷酸不与固体支持物结合。可以通过多种技术从生物样品中提取核酸(Sambrook等人, Molecular Cloning: A Laboratory Manual, 第三版, Cold Spring Harbor, N.Y. (2001))。

在一些实施方案中,可以在分析之前将细胞悬浮液预热。在一些实施方案中,紧接着在乳剂产生之前将细胞悬浮液加热(在下面第B.2节中描述)到一定温度并持续足够的持续时间,以增强细胞内部DNA聚合酶的活性,但最小化不希望的影响(如RNA降解)。因此,将细胞加热以优化本文提供的方法的产率。在一些例子中,将细胞加热到大约30°C至70°C(如30°C至60°C、25°C至60°C、30°C至60°C、40°C至60°C、45°C至55°C),持续1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20分钟的持续时间。在加热细胞之后,可以在形成乳剂之前将细胞悬浮液保持在室温下或置于冰上持续30秒至长达4小时,如30秒、45秒、1分钟、2分钟、3分钟、4分钟、5分钟、6分钟、7分钟、8分钟、9分钟、10分钟、15分钟、20分钟、25分钟、30分钟、35分钟、40分钟、45分钟、1小时、1.5小时、2小时、2.5小时、3小时、3.5小时或4小时。

多个样品可以包含至少2个、3个、4个、5个、10个、20个、30个、40个、50个、60个、70个、80个、90个或100个或更多个样品。所述多个样品可以包含至少约100个、200个、300个、400个、500个、600个、700个、800个、900个或1000个或更多个样品。所述多个样品可以包含中至少约1000个、2000个、3000个、4000个、5000个、6000个、7000个、8000个样品、9000个或10,000个样品、或100,000个样品、或1,000,000个或更多个样品。所述多个样品可以包含至少约10,000个样品。

第一样品中的所述一个或多个多核苷酸可以与第二样品中的一个或多个多核苷酸不同。第一样品中的所述一个或多个多核苷酸可以与多个样品中的一个或多个多核苷酸不

同。样品中的一个或多个多核苷酸可以包含至少约80%、85%、90%、95%、96%、97%、98%、99%或100%序列同一性。在一些实施方案中,样品中的一个或多个多核苷酸可以相差少于约100个、90个、80个、70个、60个、50个、40个、30个、25个、20个、25个、10个、9个、8个、7个、6个、5个、4个、3个、2个或1个核苷酸或碱基对。所述多个样品中的一个或多个样品中的多个多核苷酸可以包含两个或更多个相同序列。所述多个样品中的一个或多个中至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%的总多核苷酸可以包含相同的序列。所述多个样品中的一个或多个样品中的多个多核苷酸可以包含至少两个不同的序列。所述多个样品中的一个或多个中至少约5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的总多核苷酸可以包含至少两个不同的序列。在一些实施方案中,一个或多个多核苷酸是彼此的变体。例如,一个或多个多核苷酸可以含有单核苷酸多态性或其他类型的突变。在另一个例子中,一个或多个多核苷酸是剪接变体。

第一样品可以包含一个或多个细胞,并且第二样品可以包含一个或多个细胞。第一样品的所述一个或多个细胞可以与第二样品的所述一个或多个细胞属于相同的细胞类型。第一样品的所述一个或多个细胞可以与所述多个样品的一个或多个不同细胞属于不同的细胞类型。

可以并行获得所述多个样品。可以同时获得多个样品。可以依次获得所述多个样品。可以在数年的过程中(例如,获得一个或多个不同样品的100年、10年、5年、4年、3年、2年或1年)获得多个样品。可以在获得一个或多个不同样品的约一年内获得一个或多个样品。可以在获得一个或多个不同样品的12个月、11个月、10个月、9个月、8个月、7个月、6个月、5个月、4个月、3个月、2个月或1个月内获得一个或多个样品。可以在获得一个或多个不同样品的30天、28天、26天、24天、21天、20天、18天、17天、16天、15天、14天、13天、12天、11天、10天、9天、8天、7天、6天、5天、4天、3天、2天或1天内获得一个或多个样品。可以在获得一个或多个不同样品的约24小时、22小时、20小时、18小时、16小时、14小时、12小时、10小时、8小时、6小时、4小时、2小时或1小时内获得一个或多个样品。可以在获得一个或多个不同样品的约60秒、45秒、30秒、20秒、10秒、5秒、2秒或1秒内获得一个或多个样品。可以在获得一个或多个不同样品的少于一秒内获得一个或多个样品。

样品的不同多核苷酸能以不同的浓度或量(例如,不同数量的分子)存在于样品中。例如,一个多核苷酸的浓度或量可以大于样品中另一个多核苷酸的浓度或量。在一些实施方案中,样品中至少一个多核苷酸的浓度或量比样品中至少一个其他多核苷酸的浓度或量大至少约1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、20倍、25倍、30倍、35倍、40倍、45倍、50倍、60倍、70倍、80倍、90倍、100倍、200倍、300倍、400倍、500倍、600倍、700倍、800倍、900倍、1000倍或更多倍。在另一个例子中,一个多核苷酸的浓度或量小于样品中另一个多核苷酸的浓度或量。样品中至少一个多核苷酸的浓度或量可以比样品中至少一个其他多核苷酸的浓度或量小至少约1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、20倍、25倍、30倍、35倍、40倍、45倍、50倍、60倍、70倍、80倍、90倍、100倍、200倍、300倍、400倍、500倍、600倍、700倍、800倍、900倍、1000倍或

更多倍。

在一些实施方案中,两个或更多个样品可以含有不同量或浓度的多核苷酸。在一些实施方案中,一个样品中一个多核苷酸的浓度或量可以大于不同样品中相同多核苷酸的浓度或量。例如,血液样品可能含有比尿液样品更高量的特定多核苷酸。可替代地,可以将单一样品分成两个或更多个子样品。子样品可以含有不同量或浓度的相同多核苷酸。一个样品中至少一个多核苷酸的浓度或量可以比另一个样品中相同多核苷酸的浓度或量大至少约1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、20倍、25倍、30倍、35倍、40倍、45倍、50倍、60倍、70倍、80倍、90倍、100倍、200倍、300倍、400倍、500倍、600倍、700倍、800倍、900倍、1000倍或更多倍。可替代地,一个样品中一个多核苷酸的浓度或量可以小于不同样品中相同多核苷酸的浓度或量。例如,一个样品中至少一个多核苷酸的浓度或量可以比另一个样品中相同多核苷酸的浓度或量小至少约1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、20倍、25倍、30倍、35倍、40倍、45倍、50倍、60倍、70倍、80倍、90倍、100倍、200倍、300倍、400倍、500倍、600倍、700倍、800倍、900倍、1000倍或更多倍。

## 2. 微滴产生和单细胞条码化

对于使用容器条码和分子条码的单细胞条码化,可以按这样的方式产生容器(如油包水乳剂):使所得容器含有1个或更少个细胞/容器。可以按这样的方式产生容器:使所得容器还含有1个容器条码/容器。可以按这样的方式产生容器:使所得容器还含有1个分子条码化多核苷酸/容器。可以按这样的方式产生容器:使所得容器还含有两个或更多个或者多个分子条码化多核苷酸/容器。可以使细胞/容器经受如本文所述的RNA或DNA单一条码化方案,并且可以将每个容器的容器条码和一个或多个分子条码与感兴趣的靶(如细胞多核苷酸)融合。在一些实施方案中,可以将匹配的容器条码化多核苷酸与所述一个或多个分子条码化多核苷酸相同的容器中存在的细胞组分融合。在测序后,可以使用容器条码和分子条码解卷积来鉴定哪个RNA(或DNA)源自于哪个细胞。在一些实施方案中,可以按这样的方式产生容器(如油包水乳剂):使所得乳剂含有1个或更多个细胞/乳剂。在一些实施方案中,可以按这样的方式产生油包水乳剂:使所得乳剂含有1个容器条码化多核苷酸和两个或更多个分子条码化多核苷酸/容器。在一些实施方案中,可以按这样的方式产生容器:使所得容器含有超过1个容器条码化多核苷酸和两个或更多个分子条码化多核苷酸/容器。在一些实施方案中,当在溶液中时,可以将容器条码和分子条码引入容器中。在一些实施方案中,当未附接至固体支持物(如珠)时,可以将容器条码和分子条码引入容器中。示例性容器包括孔、乳剂、微滴和微胶囊。

在一些方面,可以在乳剂内部分离单细胞,所述乳剂可以充当区室(例如,容器)。可以裂解细胞,并且可以将来自细胞的转录物条码化。每个转录物都可以按这样的方式与分子条码或容器条码融合:使当使用相同的容器条码检测到两个或更多个RNA转录物时,就可以确定它们源自于同一起始细胞。这可以应用于许多不同类型的序列。一种特定的应用可以是连接抗体和TCR序列的VH和VL或V $\alpha$ 和V $\beta$ 或V $\gamma$ 和V $\delta$ 链。

在容器条码和分子条码的存在下,可以在一个或多个乳剂中分离一个或多个单细胞,使得所述一个或多个乳剂的一个容器(如微滴)可以含有最多1个或更少个细胞。可以通过乳剂中含有的缓冲液或通过冻融来化学裂解细胞,从而释放乳剂中的细胞的内含物。

可以将单细胞的RNA逆转录成cDNA。逆转录反应可以用拥有非模板末端转移酶活性的逆转录酶完成,所述逆转录酶如上所述添加约3个胞嘧啶残基。当形成乳剂时,所有逆转录缓冲液、酶和核苷酸都可以存在。在一些实施方案中,可以将引物泛化(如包含聚dT序列的多核苷酸)以靶向所有mRNA。在一些实施方案中,可以使用DNA。在一些实施方案中,可以靶向超过2个RNA。

在一些实施方案中,在逆转录期间可以将容器条码与RNA连接。在一些实施方案中,在逆转录期间可以将分子条码与RNA连接。在一些实施方案中,在逆转录期间可以将容器条码和分子条码与RNA连接。将多个细胞的样品分成小反应体积结合来自或衍生自所述多个细胞的单个细胞的多核苷酸的分子和容器条码化可以使得能够对序列(如生物标记序列)库进行高通量测序。

将多个细胞的样品分成小反应体积或含有一个或多个细胞的容器结合来自或衍生自所述多个细胞的单个细胞的多核苷酸的分子和容器条码化可以使得能够对序列(如代表生物的一定百分比的转录组的序列)库进行高通量测序。例如,序列库可以包含多个序列,所述多个序列代表生物的至少约0.00001%、0.00005%、0.00010%、0.00050%、0.001%、0.005%、0.01%、0.05%、0.1%、0.5%、1%、2%、2.5%、3%、3.5%、4%、4.5%、5%、6%、7%、8%、9%、10%、15%、20%、30%、35%、40%、45、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%的转录组。

将免疫细胞的样品分成小反应体积或含有一个或多个免疫细胞的容器结合来自或衍生自所述多个免疫细胞的单个免疫细胞的多核苷酸的分子和容器条码化可以使得能够对重链和轻链序列库进行高通量测序。这些方法还可以允许在基于条码化序列的测序之后重链和轻链进行配对。如本文所述将样品分成小反应体积还可以使得能够使用减少量的试剂,从而降低分析的材料成本。

在一些情况下,逆转录反应和/或扩增反应(例如,PCR)在微滴中进行,例如在液滴数字PCR中进行。在某些方面,本发明提供了流体区室或容器,以容纳全部或部分靶材料。在一些实施方案中,区室或容器是微滴。尽管在整个说明书中都提及“微滴”,但是除非另有说明,否则这些术语可与流体区室和流体分区互换使用。容器可以包含这样的流体区室或流体分区或者由其组成。除了在另有说明的情况下,为方便起见使用“微滴”,并且可以使用任何流体分区或区室。本文所用的微滴可以包括乳剂组合物(或两种或更多种不混溶流体的混合物),例如美国专利号7,622,280中所述。微滴可以通过WO/2010/036352中描述的设备产生。如本文所用,术语乳剂可以指代不混溶液体(如油和水)的混合物。油相和/或油包水乳剂允许在水性微滴内的反应混合物的区室化。乳剂可以包含在连续油相内的水性微滴。本文提供的乳剂可以是水包油乳剂,其中微滴是在连续水相内的油性微滴。本文所述的微滴被设计为防止区室之间的混合,每个区室保护其内容物免于蒸发和/或其他区室的内容物聚结。

本文所述的混合物或乳剂可以是稳定的或不稳定的。乳剂可以是相对稳定的并且具有最小的聚结。当小微滴组合形成逐渐变大的微滴时发生聚结。在一些情况下,少于0.00001%、0.00005%、0.00010%、0.00050%、0.001%、0.005%、0.01%、0.05%、0.1%、0.5%、1%、2%、2.5%、3%、3.5%、4%、4.5%、5%、6%、7%、8%、9%或10%的由微滴产生器产生的微滴与其他微滴聚结。乳剂也可能具有有限的絮凝,絮凝是分散相从悬浮液中以

薄片出来的过程。

可以产生具有约、小于约或大于约、或至少约0.001、0.01、0.05、0.1、1、5、10、20、30、40、50、60、70、80、100、120、130、140、150、160、180、200、300、400或500微米的平均直径的微滴。微滴可以具有约0.001至约500、约0.01至约500、约0.1至约500、约0.1至约100、约0.01至约100或约1至约100微米的平均直径。已知使用微通道错流聚焦或物理搅拌产生乳剂微滴的微流体方法产生单分散或多分散乳剂。微滴可以是单分散的微滴或容器。可以产生微滴，使得微滴的大小变化不超过微滴平均大小的正负5%。在一些情况下，产生微滴，使得微滴的大小变化不超过微滴平均大小的正负2%。微滴产生器可以从单一样品产生微滴群，其中没有一个微滴的大小变化超过总微滴群的平均大小的正负约0.1%、0.5%、1%、1.5%、2%、2.5%、3%、3.5%、4%、4.5%、5%、5.5%、6%、6.5%、7%、7.5%、8%、8.5%、9%、9.5%或10%。

可以通过使油相流过水性样品来形成微滴或容器。水相可以包含缓冲溶液和用于进行扩增反应的试剂，包括细胞、核苷酸、核苷酸类似物、分子条码化多核苷酸、容器条码化多核苷酸引物、模板核酸和酶（如DNA聚合酶、RNA聚合酶和/或逆转录酶）。在一些实施方案中，水相可以含有细胞裂解试剂，如化学细胞裂解试剂。

水相可以包含缓冲溶液和用于进行扩增反应的试剂，所述扩增反应使用或不使用固体表面（如珠）。缓冲溶液可以包含约、大于约或小于约1mM、5mM、10mM、15mM、20mM、30mM、50mM、100mM或200mM Tris。在一些情况下，氯化钾的浓度可以是约、大于约或小于约10mM、20mM、30mM、40mM、50mM、60mM、80mM、100mM、200mM。缓冲溶液可以包含约15mM Tris和50mM KC1。核苷酸可以包含脱氧核糖核苷酸三磷酸分子，包括dATP、dCTP、dGTP和dTTP，各自的浓度为约、大于约或小于约50μM、100μM、200μM、300μM、400μM、500μM、600μM或700μM。在一些情况下，将dUTP添加到水相内至约、大于约或小于约50μM、100μM、200μM、300μM、400μM、500μM、600μM、或700μM、800μM、900μM或1000μM的浓度。在一些情况下，将氯化镁或乙酸镁（MgCl<sub>2</sub>）按以下浓度添加到水相中：约、大于约或小于约1.0mM、2.0mM、3.0mM、4.0mM或5.0mM）。MgCl<sub>2</sub>的浓度可以为约3.2mM。在一些情况下，使用乙酸镁或镁。在一些情况下，使用硫酸镁。

可以使用非特异性封闭剂（如BSA或牛皮肤的明胶），其中明胶或BSA以大约0.1%–0.9%w/v的浓度范围存在。其他可能的封闭剂可以包括β乳球蛋白、酪蛋白、乳粉或其他常见的封闭剂。在一些情况下，BSA和明胶的优选浓度为约0.1%w/v。

用于在水相内扩增的引物可以具有以下的浓度：约、大于约或小于约0.05μM、0.1μM、0.2μM、0.3μM、0.4μM、0.5μM、0.6μM、0.7μM、0.8μM、0.9μM、1.0μM、1.2μM、1.5μM、1.7μM或2.0μM。水相内的引物浓度可以为约0.05至约2μM、约0.1至约1.0μM、约0.2至约1.0μM、约0.3至约1.0μM、约0.4至约1.0μM或约0.5至约1.0μM。引物的浓度可以为约0.5μM。PCR中靶核酸浓度的合适范围包括但不限于在约1 pg与约500 ng之间。

在一些情况下，水相还可以包含添加剂，包括但不限于非特异性背景/封闭核酸（例如，鲑鱼精子DNA）、生物防腐剂（例如，叠氮化钠）、PCR增强剂（例如，甜菜碱、海藻糖等）和抑制剂（例如，RNA酶抑制剂）。其他添加剂可以包括例如二甲亚砜（DMSO）、甘油、甜菜碱（一）水合物（N,N,N-三甲基甘氨酸=[羧甲基]三甲基铵）、海藻糖、7-去氮杂-2'-脱氧鸟苷三磷酸（dC7GTP或7-去氮杂-2'-dGTP）、BSA（牛血清白蛋白）、甲酰胺（formamide/methanamide）、四甲基氯化铵（TMAC）、其他四烷基铵衍生物（例如，四乙基氯化铵（TEA-C1）和四丙基氯化铵

(TPrAC1))、非离子洗涤剂(例如, Triton X-100、TWEEN®20、Nonidet P-40 (NP-40)) 或 PREXCEL-Q。在一些情况下, 水相可以包含0种、1种、2种、3种、4种、5种、6种、7种、8种、9种或10种不同的添加剂。在其他情况下, 水相可以包含至少0种、1种、2种、3种、4种、5种、6种、7种、8种、9种或10种不同的添加剂。

在一些情况下, 可以将非离子环氧乙烷/环氧丙烷嵌段共聚物以约0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%或1.0%的浓度添加到水相中。常见的生物表面活性剂包括非离子表面活性剂, 如Pluronic F-68、Tetronics和Zonyl FSN。Pluronic F-68可以按约0.5%w/v的浓度存在。

在一些情况下, 硫酸镁可以按类似的浓度代替氯化镁。来自不同供应商的大范围的常见商业PCR缓冲液可以代替缓冲溶液。

考虑了与周围相展现出液体样或固体样界面的容器。例如, 在一些实施方案中, 可以将乳剂配制成产生具有液体样界面膜的高度单分散微滴(用作容器), 所述高度单分散微滴可以通过加热转化成具有固体样界面膜的微胶囊; 此类微胶囊可以充当生物反应器, 所述生物反应器能够通过诸如PCR扩增等反应过程保留其内容物。在加热时可以发生容器向微胶囊形式的转化。例如, 这种转化可以在大于约50°C、60°C、70°C、80°C、90°C或95°C的温度下发生。在一些情况下, 使用热循环仪发生这种加热。在加热过程期间, 可以使用流体或矿物油覆盖物来防止蒸发。在加热之前可以去除或不去除过多的连续相油。生物相容性胶囊可以在大范围的热和机械过程中抵抗聚结和/或絮凝。在转化后, 可以将胶囊储存在约、大于约或小于约3°C、4°C、5°C、6°C、7°C、8°C、9°C、10°C、15°C、20°C、25°C、30°C、35°C或40°C下。胶囊形式中的这些容器可以用于生物医学应用, 如大分子(特别是含有核酸或蛋白质或两者一起的混合物的水性生物流体)的稳定的数字化封装; 药物和疫苗递送; 生物分子文库; 临床成像应用等。

微胶囊可以含有一个或多个多核苷酸并且可以抵抗聚结, 特别是在高温下。因此, PCR扩增反应能以非常高的密度(例如, 每单位体积的反应数)发生。在一些情况下, 每mL可以发生大于100,000个、500,000个、1,000,000个、1,500,000个、2,000,000个、2,500,000个、5,000,000个或10,000,000个单独的反应。在一些情况下, 反应在单个孔(例如, 微量滴定板的孔)中发生, 而没有反应体积之间的相互混合。微胶囊还可以含有使得逆转录、引物延伸和/或PCR反应能够发生所必需的其他组分, 例如引物、探针、dNTP、DNA或RNA聚合酶等。胶囊形式中的这些容器在大范围的热和机械过程中展现出对聚结和絮凝的抗性。

在一些情况下, 扩增步骤通过进行数字PCR来进行, 所述数字PCR例如为基于微流体的数字PCR或液滴数字PCR。

在一些实施方案中, 容器可以是微滴。可以使用微流体系统或设备来产生微滴。如本文所用, “微”前缀(例如, 作为“微通道”或“微流体”)通常是指元件或物品具有小于约1mm且在一些情况下小于约100微米(micron/micrometer)的宽度或直径。在一些情况下, 元件或物品包括流体可以流过的通道。另外, 如本文所用, “微流体”是指设备、装置或系统包括至少一个微尺度通道。

已经在多种背景下、通常在小型化实验室(例如, 临床)分析的背景下描述了微流体系统和设备。还已经描述了其他用途。例如, 国际专利申请公开号W0 01/89788; WO 2006/040551; WO 2006/040554; WO 2004/002627; WO 2008/063227; WO 2004/091763; WO 2005/

021 151; WO 2006/096571; WO 2007/089541; WO 2007/081385 和 WO 2008/063227。

微滴通常包括在第二载流中的一定量的第一样品流体。本领域已知的用于形成微滴的任何技术都可以与本发明的方法一起使用。示例性方法涉及使含有靶材料(例如,免疫细胞)的样品流体流动,使得其与流动的载流的两个相对流相交。载流不与样品流体混溶。样品流体与流动载流的两个相对流的相交导致将样品流体分成含有靶材料的单个样品微滴,其可以用作容器。

载流可以是与样品流体不混溶的任何流体。示例性载流是油。在某些实施方案中,载流包括表面活性剂。

相同的方法可以应用于产生含有其他试剂的单个微滴或容器,所述试剂例如为用于扩增反应(如聚合酶链反应(PCR))或基于非PCR的扩增反应(如多链置换扩增)或本领域普通技术人员已知的其他方法的试剂。用于进行基于PCR的扩增反应的合适试剂是本领域普通技术人员已知的,并且包括但不限于DNA聚合酶、正向引物和反向引物、脱氧核苷酸三磷酸(dNTP)和一种或多种缓冲液。

在某些实施方案中,通过提供包含靶材料(例如,免疫细胞和/或固体支持物如珠)的第一流体分区(例如,微滴)和第二流体(例如,作为流体流或在微滴内)来形成流体区室。将第一流体和第二流体合并以形成微滴,所述微滴可以用作所提供的方法的容器。合并可以通过对两种流体施加电场来实现。在某些实施方案中,第二流体含有用于进行扩增反应(如聚合酶链反应或扩增反应)的试剂。

较高的机械稳定性可以用于微流体操纵和较高剪切的流体处理(例如,在微流体毛细管中或在流体路径中通过90度转弯,如阀门)。热处理前后的微滴容器或胶囊容器对于标准移液器操纵和离心而言可以是机械稳定的。

### 3. 逆转录

在一些情况下,靶多核苷酸通过逆转录从RNA(如mRNA)制备。在一些情况下,靶多核苷酸通过引物延伸(例如使用聚合酶)从DNA制备。在逆转录反应期间,将细胞RNA(如mRNA)逆转录产生互补DNA(cDNA),并将独特分子条码添加到每个cDNA中,以产生与细胞转录物互补的条码化单链多核苷酸。可以为转录组中的每个转录物产生此类条码化单链多核苷酸。

本文所述的方法可以用于偶联的逆转录-PCR(逆转录-PCR)中。例如,逆转录和PCR可以在两个不同的步骤中进行。首先,可以使用多核苷酸dT引物、序列特异性引物、通用引物、随机六聚体寡核苷酸引物的混合物或本文所述的任何引物来合成样品mRNA的cDNA拷贝。在一些例子中,可以使用引物的混合物(如序列特异性引物以及随机六聚体寡核苷酸引物的混合物)来产生RNA的cDNA拷贝,例如以捕获细胞的特异性靶RNA分子,还捕获基本上与同一细胞的转录组对应的多核苷酸的集合。

逆转录和PCR可以在单个密闭容器反应中进行。例如,在同一密闭容器中可以采用大量引物,用于逆转录的一个或多个引物以及用于PCR的两个或更多个引物。用于逆转录的一个或多个引物可以与第一PCR扩增子的位置3'的mRNA结合。在一些实施方案中,可以修改PCR条件以使用对其具有特异性的引物将扩增基本限制在第一衔接子或第一衔接子池中,并限制较大分子条码化cDNA的扩增。尽管不是必需的,所述一个或多个逆转录引物可以包括RNA残基或经修饰的类似物,如2'-O-甲基RNA碱基,当与mRNA杂交时,它们将不形成RNA酶H的底物。

进行逆转录反应的温度取决于所使用的逆转录酶。在一些情况下,使用热稳定的逆转录酶,并且逆转录反应在约37°C至约75°C、在约37°C至约50°C、在约37°C至约55°C、在约37°C至约60°C、在约55°C至约75°C、在约55°C至约60°C、在约37°C或在约60°C下进行。在一些情况下,使用将3个或更多个非模板末端核苷酸转移到转录产物末端的逆转录酶。

本文所述的逆转录反应和PCR反应能以本领域已知的各种形式进行,例如在试管、微量微滴定板、微流体设备或优选地微滴中进行。

逆转录反应可以在5μL至100μL范围内的体积中或在10μL至20μL反应体积中进行。在微滴中,反应体积可以在从1pL至100nL或10pL至1nL的范围内。在一些情况下,逆转录反应在具有为约或小于1nL体积的微滴中进行。

可以使用一个或多个逆转录引物将靶多核苷酸(如RNA)逆转录成cDNA。所述一个或多个逆转录引物可以包含与RNA区域(如恒定区(例如,mRNA的重链或轻链恒定区或聚A尾))互补的区域。在一些实施方案中,逆转录引物可以包含具有与第一RNA的恒定区互补的区域的第一逆转录引物以及具有与第二RNA的恒定区互补的区域的第二逆转录引物。在一些实施方案中,逆转录引物可以包含具有与第一RNA的恒定区互补的区域的第一逆转录引物以及具有分别与一个或多个RNA的恒定区互补的区域的一个或多个逆转录引物。

在一些实施方案中,可以对逆转录引物进行修饰,以通过在反应中指数扩增引物-二聚体或引物-模板转换产物来最小化伪影形成。在一些实施方案中,通过添加引物的一个或多个碱基的2'-0-甲基化来修饰逆转录引物。在一些实施方案中,所述一个或多个2'-0-甲基化碱基位于引物序列的中心附近。此类经修饰的引物通常用于含有不能掺入与2'0-甲基修饰的残基相对的碱基的DNA聚合酶的反应中。示例性2'0-甲基修饰的引物如SEQ ID NO:12-22所示。

在一些实施方案中,逆转录引物是随机六聚体寡核苷酸的混合物。此类引物可以在随机位置结合RNA,从而引发未知序列的逆转录反应。在此类例子中,供给充足的随机六聚体引物用于实现细胞的基本上转录组的逆转录。因此,在此类实施方案中,产生了多核苷酸的集合,如与细胞的转录组对应的cDNA多核苷酸的集合。

在一些实施方案中,逆转录引物不包含条码。

逆转录引物还可以包含与RNA的区域不互补的区域。在一些实施方案中,与RNA的区域不互补的区域在与RNA互补的引物区域的5'。在一些实施方案中,与RNA的区域不互补的区域在与RNA互补的引物区域的3'。在一些实施方案中,与RNA的区域不互补的区域是5'悬突区域。在一些实施方案中,与RNA的区域不互补的区域包含用于扩增和/或测序反应的引发位点,如衔接子。使用本文所述的一个或多个引物,使用本领域已知的合适试剂将RNA分子逆转录。

在特定实施方案中,逆转录酶可以包含非模板末端转移酶活性。当包含非模板末端转移酶活性的逆转录酶到达模板末端时,它可以添加三个或更多个非模板残基,如三个或更多个非模板胞嘧啶残基。在一些实施方案中,将Superscript II<sup>TM</sup>逆转录酶用于此目的。在一些实施方案中,将Maxima<sup>TM</sup>逆转录酶用于此目的。在一些实施方案中,将Protoscript II<sup>TM</sup>逆转录酶用于此目的。在一些实施方案中,将莫洛尼鼠白血病病毒逆转录酶(MMLV-RT)用于此目的。在一些实施方案中,将HighScriber<sup>TM</sup>逆转录酶用于此目的。在一些实施方案中,将末端脱氧核苷酸转移酶用于此目的。在一些实施方案中,将禽成髓细胞瘤病毒(AMV)

逆转录酶用于此目的。可以使用具有非模板末端转移酶活性的能够转录RNA的任何逆转录酶。可以使用具有非模板末端转移酶活性的能够转录RNA的任何逆聚合酶。可以使用具有非模板末端转移酶活性的能够转录DNA的任何逆聚合酶。

可以在3'标记多核苷酸的存在下进行逆转录反应,如上文所述的那些。3'标记多核苷酸可以是用于将核酸添加到靶多核苷酸(如cDNA)的3'末端的多核苷酸。3'标记多核苷酸可以是用作模板以将核酸添加到靶多核苷酸(如cDNA)的3'末端的多核苷酸。3'标记多核苷酸可以是与靶多核苷酸(如cDNA)的3'末端杂交的多核苷酸。3'标记多核苷酸可以是含有与靶多核苷酸(如cDNA)的3'末端杂交的3'区域(如3'末端区域)的多核苷酸。例如,3'标记多核苷酸可以包含区段,如与三个或更多个非模板残基退火的区段。在一些实施方案中,3'标记多核苷酸是分子条码多核苷酸。在一些实施方案中,3'标记多核苷酸可以包含分子条码。在一些实施方案中,3'标记多核苷酸可以包含与由逆转录酶产生的链互补并退火的3'末端上的3'核糖鸟苷残基或其类似物(rGrGrG)(RNA碱基)。在一些实施方案中,可以使用三个或更多个鸟嘌呤残基代替核糖鸟苷(DNA核苷酸代替RNA核苷酸)。在一些实施方案中,3'标记多核苷酸可以包含与由逆转录酶产生的链互补并退火的3'末端上的1个或2个核糖鸟苷残基和3'末端上的核糖鸟苷残基或其类似物(rGrGG)。

在3'标记多核苷酸与cDNA链的CCC退火后,逆转录酶可以继续将cDNA延伸到标记多核苷酸中,从而将分子条码或其互补体附接至反应中的靶多核苷酸群体(如cDNA)。例如,3'标记多核苷酸可以是含有与靶多核苷酸的3'末端杂交的在3'区域5'的区域的多核苷酸。与靶多核苷酸的3'末端杂交的在3'区域5'的区域可以包含不与靶多核苷酸(如cDNA)互补的区域。与靶多核苷酸的3'末端杂交的在3'区域5'的区域可以包含分子条码。与靶多核苷酸的3'末端杂交的在3'区域5'的区域可以包含与容器条码化多核苷酸或其互补体互补的区域。在其他实验中,模板转换可以在单独的反应中进行。例如,可以在逆转录反应之后添加3'标记多核苷酸,并且可以使用诸如逆转录酶或聚合酶等酶来延伸到标记多核苷酸中。因为标记多核苷酸可以在容器中的每个分子上具有独特的简并分子条码,所以可以用分子条码独特地标记容器中的每个cDNA。在一些实施方案中,可以在进行逆转录反应的同时进行模板转换。

逆转录反应可以在3'标记多核苷酸的存在下进行。3'标记多核苷酸可以包含P7区段,其可以用于退火测序引物。3'标记多核苷酸可以包含容器条码或分子条码。3'标记多核苷酸可以包含与由逆转录酶产生的链可以互补并退火的3'末端上的3'核糖鸟苷残基(rGrGrG)(RNA碱基)。因此,可以通过逆转录酶将容器条码和分子条码添加到所述相同乳剂中的cDNA的末端。在一些实施方案中,可以使用鸟嘌呤残基代替核糖鸟苷(DNA核苷酸代替RNA核苷酸)。在3'标记多核苷酸与cDNA链的CCC退火后,逆转录酶继续将cDNA延伸到3'标记多核苷酸中,从而为反应中的所有cDNA产生分子条码化标签。在3'标记多核苷酸与分子条码化cDNA的区域退火后,逆转录酶或聚合酶继续将分子条码化cDNA延伸到另一个3'标记多核苷酸中,从而为反应中的所有cDNA产生容器条码化标签。

在一些实施方案中,模板转换可以在单独的反应中完成,而不是与可以进行逆转录反应的同时完成。在一些实施方案中,可以在逆转录反应之后添加3'标记多核苷酸,并且可以使用诸如逆转录酶或聚合酶等酶以类似的方式延伸到标记多核苷酸中。因为3'标记多核苷酸可以在每个单分子上带有独特的简并分子条码,所以每个cDNA都可以用分子条码独特地

标记。因为3'标记多核苷酸可以在来自单个容器的每个单分子上带有相同的简并容器条码,所以每个cDNA都可以用对于所述容器而言独特的容器条码来标记。

在一些实施方案中,模板转换分子(如含有条码(例如,分子条码)的模板转换寡核苷酸)可以掺入经修饰的碱基以最小化伪影形成。在一些例子中,模板转换寡核苷酸可以含有2'脱氧尿苷,其可以被逆转录,但是不能被DNA聚合酶拷贝。在一些实施方案中,可以将核糖鸟苷掺入模板转换寡核苷酸中。在一些实施方案中,可以在3'末端修饰模板转换寡核苷酸,以防止被逆转录酶或DNA聚合酶延伸。此类修饰包括3'脱氧、3'磷酸、3'氨基和3'烷基修饰以实现引物延伸的阻断。

#### 4. 聚合酶链反应(PCR)

在进行RNA分子的逆转录反应之后,可以将所得的单条码化cDNA分子同时用容器条码进行条码化,并通过一个或多个PCR反应扩增以产生一个或多个扩增子。在一些例子中,将PCR用于产生双条码化cDNA链,所述双条码化cDNA链含有与mRNA转录物序列对应的编码序列并且与在逆转录酶反应期间产生的cDNA链互补。用于PCR的酶和引物设计是已知的,并且本文描述了此类试剂的非限制性例子。可以选择任何此类试剂,并将其用于所提供方法中的PCR。

在一些情况下,PCR反应是在具有从1pL至100nL、优选地10pL至1nL范围的反应体积的微滴中进行。在一些情况下,PCR反应在具有为约或小于1nL体积的微滴中进行。在一些情况下,逆转录反应和PCR反应在具有从1pL至100nL或10pL至1nL范围的反应体积的同一微滴中进行。在一些情况下,逆转录反应和PCR反应在具有为约或小于1nL的体积或者为约或小于1pL的体积的微滴中进行。在一些情况下,逆转录反应和PCR反应在不同的微滴中进行。在一些情况下,逆转录反应和PCR反应在各自具有从1pL至100nL或10pL至1nL范围的反应体积的多个微滴中进行。在一些情况下,逆转录反应和PCR反应在各自具有为约或小于1nL体积的多个微滴中进行。

在一些情况下,第一PCR反应在具有从1pL至100nL、优选地10pL至1nL范围的反应体积的第一微滴中进行,并且第二PCR反应在具有1pL至100nL、优选地10pL至1nL范围的反应体积的第二微滴中进行。在一些情况下,第一PCR反应在具有为约或小于1nL体积的第一微滴中进行,并且第二PCR反应在具有为约或小于1nL体积的第二微滴中进行。

在一些情况下,第一PCR反应和第二PCR反应在各自具有从1pL至100nL或10pL至1nL范围的反应体积的多个微滴中进行。在一些情况下,第一PCR反应和第二PCR反应在各自具有为约或小于1nL体积的多个微滴中进行。

在本文提供的方法的一些实施方案中,可以修改反应条件以实现选定序列的扩增并最小化其他序列的扩增。此类修改可以包括在PCR热循环期间改变温度,如解链温度。例如,PCR的“冷”循环可以用于选择性扩增较短的寡核苷酸。“冷”PCR循环在其变性温度上不同于“热”PCR循环。PCR的“冷”循环实现在较低温度下的变性,以优选地扩增比较长的双链序列更容易变性的较短序列。因此,将PCR的“冷”循环用于扩增较短序列,同时限制较长序列的扩增。在一些例子中,“冷”循环和“热循环”的组合用于产生所需的扩增子。

在一些实施方案中,可以修改PCR的变性、引发和/或延伸步骤的持续时间以选择性地或优选地扩增反应体积中的特定序列。在一些例子中,可以选择或修饰用于PCR扩增的引物,以降低或增强在特定条件下的PCR扩增。在一些例子中,可以经由磷酸三酯键将热不稳

定辅助基团添加到大多数碱基的3'末端,以使寡核苷酸引物在较低温度下失活,但一旦暴露于较热温度下即具有活性。在一些例子中,在所提供的方法中使用在3'末端连接至热不稳定辅助基团的寡核苷酸,使得寡核苷酸不能在较低温度(如发生逆转录酶反应的温度)下进行引物延伸,但是在暴露于较高温度(如在PCR之前或期间)时会变得有活性。

在进行RNA分子的逆转录反应或基因组分子的引物延伸之后,通过聚合酶链反应扩增含有容器条码的寡核苷酸,以产生多个拷贝来附加到分子条码化多核苷酸上。在一些例子中,使用已通过在3'末端添加热不稳定辅助基团进行修饰的引物扩增含有容器条码的寡核苷酸,以防止在逆转录酶反应期间进行引物延伸,但使得能够在PCR期间进行引物延伸和随后的扩增。在一些实施方案中,使用如本文所述的“冷”热循环进行含有容器条码的寡核苷酸的扩增。

在进行RNA分子的逆转录反应之后,可以将所得的cDNA分子用分子条码和容器条码进行条码化,并通过一个或多个PCR反应(如第一和/或第二PCR反应)进行扩增。第一和/或第二PCR反应可以利用一对引物或多个引物对。第一和/或第二PCR反应可以利用多个正向/反向引物和反向引物。第一和/或第二PCR反应可以利用多个正向/反向引物和正向引物。多个正向/反向引物中的第一和/或第二引物可以是含有与cDNA分子或条码化cDNA分子互补的区域的正向/反向引物。多个正向/反向引物中的第一和/或第二引物可以是含有与条码化cDNA分子互补的区域的正向/反向引物。

在一些实施方案中,多个正向/反向引物包含一个或多个正向/反向引物,其中所述多个正向/反向引物中的每个正向/反向引物包含与cDNA或条码化cDNA的V区段的一个或多个上游或下游区域互补的区域。例如,多个正向/反向引物包含含有与cDNA或条码化cDNA的V区段的上游或下游区域互补的区域的正向/反向引物以及含有与cDNA或条码化cDNA的V区段的一个或多个其他上游或下游区域互补的区域的一个或多个其他正向/反向引物。例如,多个正向/反向引物包含含有与cDNA或条码化cDNA的V区段的第一和/或第二上游或下游区域互补的区域的第一和/或第二正向/反向引物以及含有与cDNA或条码化cDNA的V区段的第二上游或下游区域互补的区域的第二正向/反向引物。例如,多个正向/反向引物包含含有与cDNA或条码化cDNA的V区段的第一和/或第二正向/反向引物、含有与cDNA或条码化cDNA的V区段的第二上游或下游区域互补的区域的第二正向/反向引物以及含有与cDNA或条码化cDNA的V区段的第三上游或下游区域互补的区域的第三正向/反向引物等。所述多个正向/反向引物中的引物可以用于与由样品中的细胞(如免疫B细胞或T细胞)表达的所有V区段的所有可能的上游或下游区域退火。

在一些实施方案中,多个正向/反向引物包含一个或多个正向/反向引物,其中所述多个正向/反向引物中的每个正向/反向引物包含与cDNA或条码化cDNA的C区段的一个或多个上游或下游区域互补的区域。例如,多个正向/反向引物包含含有与cDNA或条码化cDNA的C区段的上游或下游区域互补的区域的正向/反向引物以及含有与cDNA或条码化cDNA的C区段的一个或多个其他上游或下游区域互补的区域的一个或多个其他正向/反向引物。例如,多个正向/反向引物包含含有与cDNA或条码化cDNA的C区段的第一和/或第二上游或下游区域互补的区域的第一和/或第二正向/反向引物以及含有与cDNA或条码化cDNA的C区段的第二上游或下游区域互补的区域的第二正向/反向引物。例如,多个正向/反向引物包含含有与cDNA或条码化cDNA的C区段的第一和/或第二正向/反向引物、含有与cDNA或条码化cDNA的C区段的第二上游或下游区域互补的区域的第二正向/反向引物等。所述多个正向/反向引物中的引物可以用于与由样品中的细胞(如免疫B细胞或T细胞)表达的所有C区段的所有可能的上游或下游区域退火。

二正向/反向引物、含有与cDNA或条码化cDNA的C区段的第二上游或下游区域互补的区域的第二正向/反向引物以及含有与cDNA或条码化cDNA的C区段的第三上游或下游区域互补的区域的第三正向/反向引物等。所述多个正向/反向引物中的引物可以用于与由样品中的细胞(如免疫B细胞或T细胞)表达的所有C区段的所有可能的上游或下游区域退火。

在一些实施方案中,多个正向/反向引物包含一个或多个正向/反向引物,其中所述多个正向/反向引物中的每个正向/反向引物包含与条码化cDNA的分子条码的一个或多个上游或下游区域互补的区域。例如,多个正向/反向引物包含含有与条码化cDNA的分子条码的上游或下游区域互补的区域的正向/反向引物以及含有与条码化cDNA的分子条码的一个或多个其他上游或下游区域互补的区域的一个或多个其他正向/反向引物。例如,多个正向/反向引物包含含有与条码化cDNA的分子条码的第一和/或第二上游或下游区域互补的区域的第一和/或第二正向/反向引物以及含有与条码化cDNA的分子条码的第二上游或下游区域互补的区域的第二正向/反向引物。例如,多个正向/反向引物包含含有与条码化cDNA的分子条码的第一和/或第二上游或下游区域互补的区域的第一和/或第二正向/反向引物、含有与条码化cDNA的分子条码的第二上游或下游区域互补的区域的第二正向/反向引物以及含有与条码化cDNA的分子条码的第三上游或下游区域互补的区域的第三正向/反向引物等。所述多个正向/反向引物可以用于与由样品中的细胞(如免疫B细胞或T细胞)表达的所有分子条码的所有可能的上游或下游区域退火。

在一些实施方案中,多个正向/反向引物包含一个或多个正向/反向引物,其中所述多个正向/反向引物中的每个正向/反向引物包含与条码化cDNA的容器条码的一个或多个上游或下游区域互补的区域。例如,多个正向/反向引物包含含有与条码化cDNA的容器条码的上游或下游区域互补的区域的正向/反向引物以及含有与条码化cDNA的容器条码的一个或多个其他上游或下游区域互补的区域的一个或多个其他正向/反向引物。例如,多个正向/反向引物包含含有与条码化cDNA的容器条码的第一和/或第二上游或下游区域互补的区域的第一和/或第二正向/反向引物以及含有与条码化cDNA的容器条码的第二上游或下游区域互补的区域的第二正向/反向引物。例如,多个正向/反向引物包含含有与条码化cDNA的容器条码的第一和/或第二上游或下游区域互补的区域的第一和/或第二正向/反向引物、含有与条码化cDNA的容器条码的第二上游或下游区域互补的区域的第二正向/反向引物以及含有与条码化cDNA的容器条码的第三上游或下游区域互补的区域的第三正向/反向引物等。所述多个正向/反向引物中的引物可以用于与由样品中的细胞(如免疫B细胞或T细胞)表达的所有容器条码的所有可能的上游或下游区域退火。

所述多个正向/反向引物中的正向/反向引物还包含与RNA的区域不互补的区域。在一些实施方案中,与RNA的区域不互补的区域在与RNA互补的正向/反向引物区域(即,V区段的上游或下游区域)的5'。在一些实施方案中,与RNA的区域不互补的区域在与RNA互补的正向/反向引物区域的3'。在一些实施方案中,与RNA的区域不互补的区域是5'悬突区域。在一些实施方案中,与RNA的区域不互补的区域包含用于扩增和/或第二测序反应的引发位点。在一些实施方案中,与RNA的区域不互补的区域包含用于扩增和/或第三测序反应的引发位点。在一些实施方案中,与RNA的区域不互补的区域包含用于第二和第三测序反应的引发位点。在一些实施方案中,用于第二和第三测序反应的引发位点的序列是相同的。使用如本文所述的一个或多个正向/反向引物和反向引物,使用本领域已知的合适试剂扩增cDNA分子。

在一些实施方案中,区域与RNA的区域(如mRNA的恒定区或聚A尾)互补。

### 5. 条码

条码可以是分子条码或容器条码。在一些实施方案中,条码(如分子条码或容器条码)可以各自具有从2至36个核苷酸、4至36个核苷酸、或从6至30个核苷酸、或从8至20个核苷酸、2至20个核苷酸、4至20个核苷酸或从6至20个核苷酸范围内的长度。在某些方面,组内条码的解链温度在彼此的10°C内、在彼此的5°C内或在彼此的2°C内。在某些方面,组内条码的解链温度不在彼此的10°C内、不在彼此的5°C内或不在彼此的2°C内。在其他方面,条码是最小交叉杂交组的成员。例如,这样的组的每个成员的核苷酸序列可以与所述组的每个其他成员的核苷酸序列充分不同,以至于在严格的杂交条件下没有成员可以与任何其他成员的互补体形成稳定的双链体。在一些实施方案中,最小交叉杂交组的每个成员的核苷酸序列与每个其他成员的核苷酸序列相差至少两个核苷酸。在Winzeler等人(1999) *Science* 285: 901; Brenner(2000) *Genome Biol.* 1:1; Kumar等人(2001) *Nature Rev.* 2:302; Giaever等人(2004) *Proc. Natl. Acad. Sci. USA* 101:793; Eason等人(2004) *Proc. Natl. Acad. Sci. USA* 101:1 1046; 以及Brenner(2004) *Genome Biol.* 5:240中描述了条码技术。

如本文所用,分子条码包含对于来自单细胞或来自单个容器的单分子、或者来自两个或更多个单细胞或来自两个或更多个单个容器的多个分子或分子文库的两个或更多个分子而言独特的信息。如本文所用,与来自不同单细胞或来自不同单个容器的多核苷酸相比,容器条码包含对于来自单细胞或来自单个容器的多核苷酸而言独特的信息。在一些实施方案中,独特信息包含独特的核苷酸序列。例如,可以通过确定包含分子条码或容器条码的独特或随机核苷酸序列的身份和顺序来确定分子条码或容器条码的序列。在一些实施方案中,第一衔接子包括容器条码序列。

在一些实施方案中,独特信息不能用于鉴定靶多核苷酸的序列。例如,可以将分子条码附接至一个靶多核苷酸,但是分子条码不能用于确定其附接的靶多核苷酸。在一些实施方案中,独特信息不是与靶多核苷酸序列的身份相联系的已知序列。例如,可以将容器条码附接至一个或多个靶多核苷酸,但是容器条码不能用于确定其附接至一个或多个靶多核苷酸中的哪一个。在一些实施方案中,独特信息包含随机核苷酸序列。在一些实施方案中,独特信息包含多核苷酸上的一个或多个独特核苷酸序列。在一些实施方案中,独特信息包含简并核苷酸序列或简并条码。简并条码可以包含可变核苷酸组合物或序列。例如,简并条码可以是随机序列。在一些实施方案中,分子条码或容器条码的互补体序列也是分子条码或容器条码序列。

分子条码或容器条码可以包含任何长度的核苷酸。例如,分子条码或容器条码可以包含至少约2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个、100个、200个、500个或1000个核苷酸。例如,分子条码或容器条码可以包含至多约5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个、100个、200个、500个或1000个核苷酸。

酸。在一些实施方案中,分子条码或容器条码具有特定长度的核苷酸。例如,分子条码或容器条码在长度上可以是约2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个、100个、200个、500个或1000个核苷酸。

在一些实施方案中,多个分子条码或容器条码中的每个分子条码或容器条码具有至少约2个核苷酸。例如,多个分子条码或容器条码中的每个分子条码或容器条码在长度上可以是至少约2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个、100个、200个、500个或1000个核苷酸。在一些实施方案中,多个分子条码或容器条码中的每个分子条码或容器条码具有至多约1000个核苷酸。例如,多个分子条码或容器条码中的每个分子条码或容器条码在长度上可以是至多约5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个、100个、200个、500个或1000个核苷酸。在一些实施方案中,多个分子条码或容器条码中的每个分子条码或容器条码具有相同长度的核苷酸。例如,多个分子条码或容器条码中的每个分子条码或容器条码在长度上可以是2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个、100个、200个、500个或1000个核苷酸。在一些实施方案中,多个分子条码或容器条码中的一个或多个分子条码或容器条码具有不同长度的核苷酸。例如,多个分子条码或容器条码中的一个或多个第一分子条码或容器条码可以具有约或至少约2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个、100个、200个、500个或1000个核苷酸,并且多个分子条码或容器条码中的一个或多个第二分子条码或容器条码可以具有约2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、60个、70个、80个、90个、100个、200个、500个或1000个核苷酸,其中所述一个或多个第一分子条码或容器条码的核苷酸数量不同于所述一个或多个第二分子条码或容器条码。

分子条码的数量可以超过多个容器中将被标记的分子的总数。容器条码的数量可以超过多个容器中将被标记的分子的总数。例如,分子条码或容器条码的数量可以比多个容器中将被标记的分子的总数大至少约2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、

30倍、40倍、50倍、60倍、70倍、80倍、90倍或100倍。

不同分子条码的数量可以超过多个容器中将被标记的分子的总数。在一些实施方案中,不同分子条码的数量比多个容器中将被标记的分子的总数大至少约1倍、1.5倍、2倍、2.5倍、3倍、3.5倍、4倍、4.5倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍、40倍、50倍、60倍、70倍、80倍、90倍或100倍。

单个容器中不同分子条码的数量可以超过所述单个容器中将被标记的不同分子的数量。在一些实施方案中,单个容器中不同分子条码的数量比所述单个容器中将被标记的不同分子的数量大至少约1倍、1.5倍、2倍、2.5倍、3倍、3.5倍、4倍、4.5倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍、40倍、50倍、60倍、70倍、80倍、90倍或100倍。

不同容器条码的数量可以小于多个容器中将被标记的分子的总数。在一些实施方案中,不同容器条码的数量比多个容器中将被标记的分子的总数小至少约1倍、1.5倍、2倍、2.5倍、3倍、3.5倍、4倍、4.5倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍、40倍、50倍、60倍、70倍、80倍、90倍或100倍。

来自单个容器中容器条码化多核苷酸分子的扩增产物分子的数量可以超过所述单个容器中将被标记的不同分子的数量。在一些实施方案中,来自单个容器中容器条码化多核苷酸分子的扩增产物分子的数量比所述单个容器中将被标记的不同分子的数量大至少约1倍、1.5倍、2倍、2.5倍、3倍、3.5倍、4倍、4.5倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍、40倍、50倍、60倍、70倍、80倍、90倍或100倍。

单个容器中容器条码化多核苷酸分子的数量可以小于所述单个容器中将被标记的不同分子的数量。在一些实施方案中,单个容器中容器条码化多核苷酸分子的数量比所述单个容器中将被标记的不同分子的数量小至少约1倍、1.5倍、2倍、2.5倍、3倍、3.5倍、4倍、4.5倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍、40倍、50倍、60倍、70倍、80倍、90倍或100倍。

单个容器中容器条码化多核苷酸分子的数量可以是一个分子。单个容器中未扩增的容器条码化多核苷酸分子的数量可以是一个分子。

在一些实施方案中,至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%的不同分子条码具有相同的浓度。在一些实施方案中,至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%的不同容器条码具有相同的浓度。

在一些实施方案中,至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%的不同分子条码具有不同的浓度。在一些实施方案中,至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%的不同容器条码具有不同的浓度。

分子条码或容器条码群体中的分子条码或容器条码可以具有至少10个、15个、20个、25个、30个、35个、40个、45个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500



000个、90,000个、100,000个、200,000个、300,000个、400,000个、500,000个、600,000个、700,000个、800,000个、900,000个、 $1 \times 10^6$ 个、 $2 \times 10^6$ 个、 $3 \times 10^6$ 个、 $4 \times 10^6$ 个、 $5 \times 10^6$ 个、 $6 \times 10^6$ 个、 $7 \times 10^6$ 个、 $8 \times 10^6$ 个、 $9 \times 10^6$ 个、 $1 \times 10^7$ 个、 $2 \times 10^7$ 个、 $3 \times 10^7$ 个、 $4 \times 10^7$ 个、 $5 \times 10^7$ 个、 $6 \times 10^7$ 个、 $7 \times 10^7$ 个、 $8 \times 10^7$ 个、 $9 \times 10^7$ 个、 $1 \times 10^8$ 个、 $2 \times 10^8$ 个、 $3 \times 10^8$ 个、 $4 \times 10^8$ 个、 $5 \times 10^8$ 个、 $6 \times 10^8$ 个、 $7 \times 10^8$ 个、 $8 \times 10^8$ 个、 $9 \times 10^8$ 个、 $1 \times 10^9$ 个、 $2 \times 10^9$ 个、 $3 \times 10^9$ 个、 $4 \times 10^9$ 个、 $5 \times 10^9$ 个、 $6 \times 10^9$ 个、 $7 \times 10^9$ 个、 $8 \times 10^9$ 个、 $9 \times 10^9$ 个、 $1 \times 10^{10}$ 个、 $2 \times 10^{10}$ 个、 $3 \times 10^{10}$ 个、 $4 \times 10^{10}$ 个、 $5 \times 10^{10}$ 个、 $6 \times 10^{10}$ 个、 $7 \times 10^{10}$ 个、 $8 \times 10^{10}$ 个、 $9 \times 10^{10}$ 个、 $1 \times 10^{11}$ 个、 $2 \times 10^{11}$ 个、 $3 \times 10^{11}$ 个、 $4 \times 10^{11}$ 个、 $5 \times 10^{11}$ 个、 $6 \times 10^{11}$ 个、 $7 \times 10^{11}$ 个、 $8 \times 10^{11}$ 个、 $9 \times 10^{11}$ 个、 $1 \times 10^{12}$ 个、 $2 \times 10^{12}$ 个、 $3 \times 10^{12}$ 个、 $4 \times 10^{12}$ 个、 $5 \times 10^{12}$ 个、 $6 \times 10^{12}$ 个、 $7 \times 10^{12}$ 个、 $8 \times 10^{12}$ 个、 $9 \times 10^{12}$ 个或更多个不同的序列。

在一些实施方案中,将一个或多个分子条码用于对序列分组(group)或分箱(bin)。在一些实施方案中,将一个或多个分子条码用于对序列分组或分箱,其中每个箱元(bin)中的序列含有相同的分子条码。在一些实施方案中,将一个或多个分子条码或容器条码用于对序列分组或分箱,其中每个箱元中的序列包含扩增子组。在一些实施方案中,将一个或多个分子条码用于对序列分组或分箱,其中每个箱元中的序列包含多个序列,其中从其产生所述多个序列的多核苷酸衍生自扩增反应中的相同多核苷酸分子。

在一些实施方案中,将一个或多个容器条码用于对序列分组或分箱。在一些实施方案中,将一个或多个容器条码用于对序列分组或分箱,其中每个箱元中的序列含有相同的容器条码。在一些实施方案中,将一个或多个容器条码用于对序列分组或分箱,其中每个箱元中的序列包含一个或多个扩增子组。在一些实施方案中,将一个或多个容器条码用于对序列分组或分箱,其中每个箱元中的序列包含多个序列,其中从其产生所述多个序列的多核苷酸衍生自单个容器或单细胞的多核苷酸分子。

在一些实施方案中,将一个或多个分子条码和容器条码用于对序列分组或分箱。在一些实施方案中,将一个或多个分子条码和容器条码用于对序列分组或分箱,其中每个箱元中的序列含有相同的分子条码和相同的容器条码。在一些实施方案中,将一个或多个分子条码和容器条码用于对序列分组或分箱,其中每个箱元中的序列包含一个或多个扩增子组。在一些实施方案中,将一个或多个分子条码和容器条码用于对序列分组或分箱,其中每个箱元中的序列包含多个序列,其中从其产生所述多个序列的多核苷酸衍生自扩增反应中的相同多核苷酸和衍生自同一单细胞或容器。在一些实施方案中,不使用一个或多个分子条码和容器条码来比对序列。

在一些实施方案中,不使用一个或多个分子条码来比对序列。在一些实施方案中,将一个或多个分子条码用于比对序列。在一些实施方案中,将一个或多个分子条码用于对序列分组或分箱,并且将靶特异性区域用于比对序列。在一些实施方案中,不使用一个或多个容器条码来比对序列。在一些实施方案中,将一个或多个容器条码用于比对序列。在一些实施方案中,将一个或多个容器条码用于对序列分组或分箱,并且将靶特异性区域用于比对序列。在一些实施方案中,将一个或多个分子条码和容器条码用于比对序列。在一些实施方案中,将一个或多个分子条码和容器条码用于对序列分组或分箱,并且将靶特异性区域用于比对序列。

在一些实施方案中,所比对的序列含有相同的分子条码。在一些实施方案中,所比对的

序列含有相同的容器条码。在一些实施方案中,所比对的序列含有相同的分子条码和容器条码。在一些实施方案中,将一个或多个分子条码或容器条码用于比对序列,其中所比对的序列包含来自扩增子组的两个或更多个序列。在一些实施方案中,将一个或多个分子条码或容器条码用于比对序列,其中所比对的序列包含多个序列,其中从其产生所述多个序列的多核苷酸衍生自扩增反应中的相同多核苷酸分子。在一些实施方案中,将一个或多个分子条码或容器条码用于比对序列,其中所比对的序列包含多个序列,其中从其产生所述多个序列的多核苷酸衍生自单细胞或单个容器。

### C. 衔接子连接

在衔接子连接之前,可以对双条码化多核苷酸或扩增子进行纯化和/或选择大小。可以选择双条码化多核苷酸的大小,以优化选定的测序方法。所需的多核苷酸大小由测序仪器的限制和由特定的测序应用决定。在一些例子中,所需的多核苷酸大小为0个碱基对 (bp) 至 100,000bp (100千碱基 (kb))、50bp至50kb、100bp至25kb。在一些实施方案中,使用短读段测序仪对本文产生的多核苷酸进行测序。通常,短读段测序仪的最佳多核苷酸大小的范围在长度上为:从约20个碱基对 (bp) 至2000bp、50bp至1500bp、50bp至1250bp、50bp至1000bp、50bp至750bp、50bp至500bp、100bp至1500bp、100bp至1250bp、100bp至1000bp、100bp至750bp、100bp至500bp、200bp至1500bp、200bp至1250bp、200bp至1000bp、200bp至750bp或250bp至500bp。

在一些实施方案中,使用长读段测序仪对本文产生的多核苷酸进行测序。通常,短读段测序仪的最佳多核苷酸大小的范围在长度上为:从约1千碱基 (kb) 至100kb,如1kb至50kb、5kb至25kb、5kb至20kb、或大约1kb、5kb、10kb、15kb或20kb。

为了产生所需大小的多核苷酸的集合,可以通过修改逆转录或引物延伸反应的条件(例如修改反应的延伸步骤的时间)来对多核苷酸的集合调整大小。在一些实施方案中,可以通过物理方法(即,声学剪切和超声处理)或酶促方法(即,非特异性核酸内切酶混合物和转座酶标签化反应)将多核苷酸的集合片段化或调整大小至所需长度。可以通过琼脂糖凝胶电泳(如变性凝胶电泳)、尺寸排阻方法或自动化方法或商业试剂盒分离所需大小的多核苷酸(Quail等人, *Electrophoresis* (2012) 33 (23) :3521-3528; Duhaime等人, *Environ Microbiol* (2012) 14 (9) :2526-2537)。

在一些实施方案中,在大小选择之前,例如通过亲和纯化来纯化双链双条码化多核苷酸。在一些实施方案中,在大小选择之前将双链双条码化多核苷酸变性。在一些实施方案中,通过破坏DNA互补链之间的氢键将双链双条码化多核苷酸变性。在一些实施方案中,通过应用酸或碱、浓缩的无机盐、有机溶剂(例如,醇或氯仿)、辐射或热来实现双链DNA的变性。在一些实施方案中,通过暴露于化学剂(如甲酰胺、胍、水杨酸钠、二甲亚砜(DMS)、丙二醇、脲或NaOH)来实现双链DNA的变性。在一些实施方案中,将双链DNA分子用NaOH(如0.1M NaOH)处理以产生单链分子。

在大小选择和/或纯化后,可以将第二衔接子添加到衔接子标记的双条码化多核苷酸中,所述双条码化多核苷酸是通过所述方法产生的含有具有通用引发序列的第一衔接子、容器条码和分子条码的多核苷酸。衔接子含有通用引发序列,所述通用引发序列可以用于对衔接子标记的双条码化多核苷酸进行扩增或测序。衔接子可以含有任何已知的通用引发序列或其片段。示例性通用引发序列包括P7、C7、P5或C5引发序列。

可以使用任何已知的方法实现第二衔接子的添加。可以将衔接子添加到单链多核苷酸或双链多核苷酸中。在一些例子中,将衔接子添加到单链多核苷酸中。在其他例子中,将衔接子(如双链衔接子)添加到双链多核苷酸中。在一些实施方案中,将连接酶用于连接单链衔接子。例如,可以使用Thermostable App连接酶(NEB)或CircLigase II(Epicentre)将第二衔接子或单链衔接子连接到单链的衔接子标记的双条码化多核苷酸上。

在一些实施方案中,可以通过将简并的夹板衔接子退火来将第二衔接子添加到单链的衔接子标记的双条码化多核苷酸中。例如,可以通过在单链的衔接子标记的双条码化多核苷酸的末端添加夹板衔接子双链体来添加第二衔接子。此类夹板衔接子双链体含有在分子的一个末端具有简并悬突的成对双链寡核苷酸。简并悬突可以为1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个碱基。将分子悬突部分的简并核苷酸与第一衔接子序列末端相对的单链的衔接子标记的双条码化多核苷酸的末端退火。在所述方法的一些实施方案中,将具有3'悬突的夹板衔接子双链体与第一衔接子相对的单链的衔接子标记的双条码化多核苷酸的3'末端退火。在所述方法的一些实施方案中,将具有5'悬突的夹板衔接子双链体与第一衔接子相对的单链的衔接子标记的双条码化多核苷酸的5'末端退火。连接酶(如blunt/TA连接酶)可以促进夹板衔接子双链体与单链的衔接子标记的双条码化多核苷酸退火。

在所述方法的一些实施方案中,可以将非模板核苷酸的酶促添加物添加到与衔接子退火的单链的衔接子标记的双条码化多核苷酸的末端。在一些实施方案中,使用互补碱基配对使第二衔接子直接与非模板核苷酸退火。在一些实施方案中,可以使夹板衔接子双链体与非模板核苷酸退火,以实现将衔接子添加到分子末端。在一些实施方案中,将衔接子添加到单链的衔接子标记的双条码化多核苷酸的3'末端。在一些实施方案中,将衔接子添加到单链的衔接子标记的双条码化多核苷酸的5'末端。连接酶(如blunt/TA连接酶)可以促进第二衔接子或夹板衔接子双链体与单链的衔接子标记的双条码化多核苷酸退火。

第二衔接子含有通用引发序列、或通用引发位点、或通用引发序列或通用引发位点足以与互补序列退火的连续部分。通用引发序列或通用引发位点含有与通用引物或其连续部分互补的寡核苷酸序列。在下面第D节中列出了示例性通用引物。

#### D. 扩增和测序

可以将如在上述程序中产生的与一个或多个靶多核苷酸序列和/或细胞的全部或部分转录组对应的含有双条码化多核苷酸的样品(在双条码化多核苷酸的相对末端处或附近具有第一和第二衔接子)扩增,以产生双条码化多核苷酸文库的多个拷贝。可以在测序之前进行扩增。在一些实施方案中,具有测序衔接子的引物可以用于扩增文库中的一个或多个序列。在一些实施方案中,将选定的转录物扩增并制备用于测序。

在一些实施方案中,将具有测序衔接子的衔接子引物用于扩增文库中的所有转录物,并制备所述转录物用于测序。在一些实施方案中,将具有测序衔接子的衔接子引物和具有测序衔接子的靶特异性引物用于扩增靶基因,并制备所述靶基因用于测序。在一些实施方案中,将具有测序衔接子的细胞特异性引物(如针对容器条码的引物)和具有测序衔接子的在来自容器条码的转录物的相对末端的针对衔接子的引物用于扩增选定细胞的转录组,并制备所述转录组用于测序。图2描绘了通过使用如本文所述的选定引物对产生的文库中的多核苷酸的示例性选择扩增。

## 1. 扩增

含有靶多核苷酸的样品可以包含与完整mRNA转录物或其一个或多个片段对应的DNA，其可以被扩增。在一些情况下，相应的mRNA转录物或其一个或多个片段的平均长度可以小于约100个、200个、300个、400个、500个或800个碱基对，或小于约5个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个、110个、120个、130个、140个、150个、160个、170个、180个、190个或200个核苷酸，或小于约1、2、5、10、20、30、40、50、60、70、80、90、100千碱基。在一些情况下，扩增来自相对较短的模板（如含有模板的样品）的靶序列，所述靶序列为约40个、45个、50个、55个、60个、65个、70个、75个、80个、85个、90个、95个或100个碱基。

扩增反应可以包含一种或多种添加剂。在一些情况下，所述一种或多种添加剂是二甲亚砜(DMSO)、甘油、甜菜碱(一)水合物(N,N,N-三甲基甘氨酸=[羧甲基]三甲基铵)、海藻糖、7-去氮杂-2'-脱氧鸟苷三磷酸(dC7GTP或7-去氮杂-2'-dGTP)、BSA(牛血清白蛋白)、甲酰胺(formamide/methanamide)、四甲基氯化铵(TMAC)、其他四烷基铵衍生物(例如，四乙基氯化铵(TEA-C1)和四丙基氯化铵(TPrA-C1)、非离子洗涤剂(例如，Triton X-100、TWEEN® 20、Nonidet P-40(NP-40))或PREXCEL-Q。在一些情况下，扩增反应包含0种、1种、2种、3种、4种、5种、6种、7种、8种、9种或10种不同的添加剂。在其他情况下，扩增反应包含至少0种、1种、2种、3种、4种、5种、6种、7种、8种、9种或10种不同的添加剂。

可以对反应体积(例如，微滴)中含有的样品进行热循环反应。微滴可以是多分散的或优选地是单分散的，由本领域技术人员通过T通道连接或其他方式通过搅拌、超声处理或微流体地产生。密度可以超过20,000个微滴/40μL(1nL微滴)、200,000个微滴/40μL(100pL微滴)。在热循环期间，微滴可以保持完整。在大于约10,000个微滴/μL、100,000个微滴/μL、200,000个微滴/μL、300,000个微滴/μL、400,000个微滴/μL、500,000个微滴/μL、600,000个微滴/μL、700,000个微滴/μL、800,000个微滴/μL、900,000个微滴/μL或1,000,000个微滴/μL的密度下，在热循环期间，微滴可以保持完整。在其他情况下，在热循环期间，两个或更多个微滴不会聚结。在其他情况下，在热循环期间，大于100个或大于1,000个微滴不会聚结。

可以使用催化引物延伸的任何DNA聚合酶，包括但不限于大肠杆菌DNA聚合酶、大肠杆菌DNA聚合酶1的Klenow片段、T7DNA聚合酶、T4DNA聚合酶、Taq聚合酶、Pfu DNA聚合酶、Vent DNA聚合酶、噬菌体29、REDTaq™、基因组DNA聚合酶或测序酶。在一些情况下，使用热稳定的DNA聚合酶。还可以进行热启动PCR，其中在添加聚合酶之前将反应加热至95°C持续两分钟，或者可以使聚合酶保持无活性直到循环1中的第一个加热步骤。热启动PCR可以用于最小化非特异性扩增。

任何数量的PCR循环可以用于扩增DNA，例如约、大于约或小于约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个或45个循环。扩增循环的数量可以为约1-45、10-45、20-45、30-45、35-45、10-40、10-30、10-25、10-20、10-15、20-35、25-35、30-35或35-40。

可以通过任何已知的方式进行靶核酸的扩增。可以通过聚合酶链反应(PCR)或等温DNA扩增来扩增靶核酸。可以使用的PCR技术的例子包括但不限于定量PCR、定量荧光PCR(QF-PCR)、多重荧光PCR(MF-PCR)、实时PCR(逆转录-PCR)、单细胞PCR、限制性片段长度多态性

PCR (PCR-RFLP)、PCR-RFLP/逆转录-PCR-RFLP、热启动PCR、巢式PCR、原位polony PCR、原位滚环扩增 (RCA)、数字PCR (dPCR)、液滴数字PCR (ddPCR)、桥式PCR、皮升PCR和乳剂PCR。其他合适的扩增方法包括连接酶链反应 (LCR)、转录扩增、分子倒置探针 (MIP) PCR、自我维持序列复制、靶多核苷酸序列选择性扩增、共有序列引发的聚合酶链反应 (CP-PCR)、任意引发的聚合酶链反应 (AP-PCR)、简并多核苷酸引发的PCR (DOP-PCR) 和基于核酸的序列扩增 (NABSA)。在本文中可以使用的其他扩增方法包括在美国专利号5,242,794; 5,494,810; 4,988,617; 和6,582,938中所述的那些，并且包括QB复制酶介导的RNA扩增。扩增可以是等温扩增，例如等温线性扩增。

在一些实施方案中，扩增不在固体支持物上发生。在一些实施方案中，在微滴中扩增不在固体支持物上发生。在一些实施方案中，当扩增不在微滴中时，扩增确实在固体支持物上发生。

扩增反应可以包含一种或多种添加剂。在一些实施方案中，所述一种或多种添加剂是二甲亚砜 (DMSO)、甘油、甜菜碱 (一) 水合物 (N,N,N-三甲基甘氨酸=[羧甲基]三甲基铵)、海藻糖、7-去氮杂-2'-脱氧鸟昔三磷酸 (dC7GTP或7-去氮杂-2'-dGTP)、BSA (牛血清白蛋白)、甲酰胺 (formamide/methanamide)、四甲基氯化铵 (TMAC)、其他四烷基铵衍生物 (例如，四乙基氯化铵 (TEA-C1) 和四丙基氯化铵 (TPrA-C1)、非离子洗涤剂 (例如，Triton X-100、TWEEN® 20、Nonidet P-40 (NP-40) 或PREXCEL-Q)。在一些实施方案中，扩增反应可以包含0种、1种、2种、3种、4种、5种、6种、7种、8种、9种或10种不同的添加剂。在其他情况下，扩增反应可以包含至少0种、1种、2种、3种、4种、5种、6种、7种、8种、9种或10种不同的添加剂。

通常，可以在扩增反应中使用一对或多对引物；引物对的一个引物可以是正向引物，并且引物对的一个引物可以是反向引物。

在一些情况下，可以在扩增反应中使用第一对引物；第一对的一个引物可以是与第一靶多核苷酸分子的序列互补的正向引物，并且第一对的一个引物可以是与第一靶多核苷酸分子的第二序列互补的反向引物，并且第一靶基因座可以位于第一序列与第二序列之间。在一些实施方案中，第一靶基因座包含V<sub>H</sub>或V<sub>α</sub>或V<sub>γ</sub>序列。

在一些情况下，可以在扩增反应中使用第二对引物；第二对的一个引物可以是与第二靶多核苷酸分子的第一序列互补的正向引物，并且第二对的一个引物可以是与第二靶多核苷酸分子的第二序列互补的反向引物，并且第二靶基因座可以位于第一序列与第二序列之间。在一些实施方案中，第二靶基因座包含V<sub>L</sub>或V<sub>β</sub>或V<sub>δ</sub>序列。

在一些情况下，可以在扩增反应中使用第三对引物；第三对的一个引物可以是与第三靶多核苷酸分子的第一序列互补的正向引物，并且第三对的一个引物可以是与第三靶多核苷酸分子的第二序列互补的反向引物，并且第三靶基因座可以位于第一序列与第二序列之间。在一些实施方案中，第三靶基因座包含条码，如分子条码或容器条码。

在一些情况下，可以在扩增反应中使用多对引物。在一些例子中，引物对的一个引物可以是与第一衔接子序列互补的正向引物，并且引物对的一个引物可以是针对第二衔接子序列的反向引物。在一些例子中，引物对的一个引物可以是与第二衔接子序列互补的正向引物，并且引物对的一个引物可以是针对第一衔接子序列的反向引物。

正向引物和反向引物的长度可以取决于靶多核苷酸和靶基因座的序列。例如，可以优化正向引物和反向引物的长度和/或T。在一些情况下，引物在长度上可以是约、大于约或小

于约10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、51个、52个、53个、54个、55个、56个、57个、58个、59个或60个核苷酸。在一些情况下,引物在长度上是约15至约20个、约15至约25个、约15至约30个、约15至约40个、约15至约45个、约15至约50个、约15至约55个、约15至约60个、约20至约25个、约20至约30个、约20至约35个、约20至约40个、约20至约45个、约20至约50个、约20至约55个或约20至约60个核苷酸。

在结合模板多核苷酸之前,引物可以是单链DNA。在一些情况下,引物最初包含双链序列。引物的适当长度可以取决于引物的预期用途,但是可以在从约6至约50个核苷酸或从约15至约35个核苷酸的范围内。短引物分子通常可能需要较低的温度以与模板形成足够稳定的杂交体复合物。在一些实施方案中,引物不必反映模板核酸的确切序列,但是可以充分互补以与模板杂交。在一些情况下,在与模板多核苷酸结合之前,引物可以是部分双链的。具有双链序列的引物可以具有约、大于约或小于约3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个碱基的发夹环。引物的双链部分可以是约、大于约、小于约或至少约3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个或50个碱基对。用于扩增给定靶序列的合适引物的设计在本领域中是熟知的。

引物可以掺入允许检测或固定引物但不改变引物的基本特性的另外功能(例如,充当DNA合成的起始点)。例如,引物可以在5'末端含有不与靶核酸杂交但促进克隆或进一步扩增或对扩增产物进行测序的另外的核酸序列。例如,另外的序列可以包含引物结合位点,如可以是衔接子的通用引物结合位点。与模板充分互补以杂交的引物区域在本文中可以被称为杂交区域。

在另一种情况下,在本文所述的方法和组合物中使用的引物可以包含一个或多个通用核苷。通用核苷的非限制性例子是5-硝基吲哚和肌苷,如美国申请公开号2009/0325169和2010/0167353中所述。

可以根据已知的参数设计引物,以避免二级结构和自杂交。不同的引物对可以在大约相同的温度下进行退火和解链,例如在另一个引物对的1°C、2°C、3°C、4°C、5°C、6°C、7°C、8°C、9°C或10°C之内。在一些情况下,最初使用大于1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、25个、30个、35个、40个、45个、50个、100个、200个、500个、1000个、5000个、10,000个或更多个引物。此类引物可以与本文所述的靶多核苷酸杂交。

引物可以通过多种方法制备,所述方法包括但不限于克隆适当的序列和使用本领域熟知的方法直接化学合成(Narang等人,Methods Enzymol.68:90(1979);Brown等人,Methods Enzymol.68:109(1979))。引物也可以从商业来源获得。引物可以具有相同的解链温度。引物可以具有不相同的解链温度。引物的长度可以在5'末端或3'末端延长或缩短,以产生具有所需解链温度的引物。引物对的一个引物可以比另一个引物长。引物对内引物的3'退火长度可以不同。同样,可以设计每个引物对的退火位置,使得引物对的序列和长度产生所需的解链温度。用于确定小于25个碱基对的引物的解链温度的方程式是Wallace法则( $T = 2(A + T) + 4(G + C)$ )。

$+T) + 4(G+C)$ )。计算机程序也可以用于设计引物。可以使用软件程序计算每个引物的 $T_m$ (解链或退火温度)。在扩增的任何循环(包括但不限于循环1、2、3、4、5、循环6-10、循环10-15、循环15-20、循环20-25、循环25-30、循环30-35或循环35-40)之后,可以重新计算和增加引物的退火温度。在扩增的初始循环之后,可以将5'一半的引物掺入来自每个感兴趣基因座的产物中;因此,可以基于每个引物的5'一半和3'一半的序列重新计算 $T_m$ 。

进行本文公开的方法的一个或多个反应可以包括使用一个或多个引物。如本文所用,引物包含足够互补以与模板多核苷酸杂交的双链、单链或部分单链的多核苷酸。在结合模板多核苷酸之前,引物可以是单链DNA。在一些实施方案中,引物最初包含双链序列。引物位点包括引物与其杂交的模板区域。在一些实施方案中,引物能够充当模板指导的核酸合成的起始点。例如,当存在四种不同的核苷酸和聚合剂或酶(如DNA或RNA聚合酶或逆转录酶)时,引物可以启动模板指导的核酸合成。

引物对包括2个引物:第一引物具有与模板序列的5'末端杂交的5'上游区域,并且第二引物具有与模板序列的3'末端的互补体杂交的3'下游区域。引物组包括两个或更多个引物:第一引物或第一多个引物具有与一个模板序列或多个模板序列的5'末端杂交的5'上游区域,并且第二引物或第二多个引物具有与所述模板序列或所述多个模板序列的3'末端的互补体杂交的3'下游区域。

在一些实施方案中,引物包含靶特异性序列。在一些实施方案中,引物包含样品条码序列。在一些实施方案中,引物包含通用引发序列。在一些实施方案中,引物包含PCR引发序列。在一些实施方案中,引物包含用于启动多核苷酸扩增的PCR引发序列。(Dieffenbach, PCR Primer: A Laboratory Manual, 第2版 (Cold Spring Harbor Press, New York (2003))。通用引物结合位点或序列允许将通用引物与多核苷酸和/或扩增子附接。通用引物在本领域中是熟知的,并且包括但不限于-47F (M13F)、 $\alpha$ MF、AOX3'、AOX5'、BGHr、CMV-30、CMV-50、CVMf、LACrmt、 $\lambda$ gt 10F、 $\lambda$ gt 10R、 $\lambda$ gt 11F、 $\lambda$ gt 11R、M13rev、M13正向(-20)、M13反向、male、pQEproseq、pQE、pA-120、pet4、pGAP正向、pGLRVpr3、pGLpr2R、pKLAC14、pQEFS、pQERS、pucU1、pucU2、reversA、seqIREStam、seqIRESzpet、seqori、seqPCR、seqpIRES-、seqpIRES+、seqpSecTag、seqpSecTag+、seqretro+PSI、SP6、T3-prom、T7-prom和T7-termInv。

如本文所用,附接可以指代共价相互作用和非共价相互作用中的两者或任一者。通用引物与通用引物结合位点的附接可以用于对多核苷酸和/或扩增子进行扩增、检测和/或测序。通用引物结合位点可以包含至少约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个或1000个核苷酸或碱基对。在另一个例子中,通用引物结合位点包含至少约1500个、2000个、2500个、3000个、3500个、4000个、4500个、5000个、5500个、6000个、6500个、7000个、7500个、8000个、8500个、9000个、9500个或10000个核苷酸或碱基对。在一些实施方案中,通用引物结合位点包含1-10个、10-20个、10-30个或10-100个核苷酸或碱基对。在一些实施方案中,通用引物结合位点包含从约1-90个、1-80个、1-70个、1-60个、1-50个、1-40个、1-30个、1-20个、1-10个、2-90个、2-80个、2-70个、2-60个、2-50个、2-40个、2-30个、2-20个、2-10个、1-900个、1-800个、1-700个、1-600个、1-500个、1-400个、1-300个、1-200个、1-100个、2-900个、2-800个、2-700

个、2-600个、2-500个、2-400个、2-300个、2-200个、2-100个、5-90个、5-80个、5-70个、5-60个、5-50个、5-40个、5-30个、5-20个、5-10个、10-90个、10-80个、10-70个、10-60个、10-50个、10-40个、10-30个、10-20个、10-10个、5-900个、5-800个、5-700个、5-600个、5-500个、5-400个、5-300个、5-200个、5-100个、10-900个、10-800个、10-700个、10-600个、10-500个、10-400个、10-300个、10-200个、10-100个、25-900个、25-800个、25-700个、25-600个、25-500个、25-400个、25-300个、25-200个、25-100个、100-1000个、100-900个、100-800个、100-700个、100-600个、100-500个、100-400个、100-300个、100-200个、200-1000个、200-900个、200-800个、200-700个、200-600个、200-500个、200-400个、200-300个、300-1000个、300-900个、300-800个、300-700个、300-600个、300-500个、300-400个、400-1000个、400-900个、400-800个、400-700个、400-600个、400-500个、500-1000个、500-900个、500-800个、500-700个、500-600个、600-1000个、600-900个、600-800个、600-700个、700-1000个、700-900个、700-800个、800-1000个、800-900个或900-1000个核苷酸或碱基对。

引物可以具有与其在引物延伸产物的合成中的使用相容的长度。引物可以是长度为8个至200个核苷酸的多核苷酸。引物的长度可以取决于模板多核苷酸和模板基因座的序列。例如,可以优化引物或引物组的长度和/或解链温度( $T_m$ )。在一些情况下,引物在长度上可以是约、大于约或小于约10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、51个、52个、53个、54个、55个、56个、57个、58个、59个或60个核苷酸。在一些实施方案中,引物在长度上为约8-100个核苷酸,例如在长度上为10-75个、15-60个、15-40个、18-30个、20-40个、21-50个、22-45个、25-40个、7-9个、12-15个、15-20个、15-25个、15-30个、15-45个、15-50个、15-55个、15-60个、20-25个、20-30个、20-35个、20-45个、20-50个、20-55个或20-60个核苷酸以及它们之间的任何长度。在一些实施方案中,引物在长度上为至多约10个、12个、15个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、35个、40个、45个、50个、55个、60个、65个、70个、75个、80个、85个、90个、95个或100个核苷酸。

通常,可以在指数扩增反应中使用一对或多对引物;引物对的一个引物可以是正向引物,并且引物对的一个引物可以是反向引物。在一些实施方案中,可以在指数扩增反应中使用第一对引物;第一对的一个引物可以是与第一模板多核苷酸分子的序列互补的正向引物,并且第一对的一个引物可以是与第一模板多核苷酸分子的第二序列互补的反向引物,并且第一模板基因座可以位于第一序列与第二序列之间。在一些实施方案中,可以在扩增反应中使用第二对引物;第二对的一个引物可以是与第二靶多核苷酸分子的第一序列互补的正向引物,并且第二对的一个引物可以是与第二靶多核苷酸分子的第二序列互补的反向引物,并且第二靶基因座可以位于第一序列与第二序列之间。在一些实施方案中,第二靶基因座包含可变轻链抗体序列。在一些实施方案中,可以在扩增反应中使用第三对引物;第三对的一个引物可以是与第三模板多核苷酸分子的第一序列互补的正向引物,并且第三对的一个引物可以是与第三模板多核苷酸分子的第二序列互补的反向引物,并且第三模板基因座可以位于第一序列与第二序列之间。

所述一个或多个引物可以与多个模板多核苷酸的至少一部分退火。所述一个或多个引物可以与所述多个模板多核苷酸的3'末端和/或5'末端退火。所述一个或多个引物可以与

所述多个模板多核苷酸的内部区域退火。内部区域可以距所述多个模板多核苷酸的3'末端或5'末端至少约10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、100个、150个、200个、210个、220个、230个、240个、250个、260个、270个、280个、290个、300个、310个、320个、330个、340个、350个、360个、370个、380个、390个、400个、410个、420个、430个、440个、450个、460个、470个、480个、490个、500个、510个、520个、530个、540个、550个、560个、570个、580个、590个、600个、650个、700个、750个、800个、850个、900个或1000个核苷酸。所述一个或多个引物可以包含固定引物组。所述一个或多个引物可以包含至少一个或多个定制引物。所述一个或多个引物可以包含至少一个或多个对照引物。所述一个或多个引物可以包含至少一个或多个管家基因引物。所述一个或多个引物可以包含通用引物。通用引物可以与通用引物结合位点退火。在一些实施方案中，所述一个或多个定制引物与SBC、靶特异性区域、其互补体或其任何组合退火。所述一个或多个引物可以包含通用引物。可以设计所述一个或多个引物引物以扩增或进行引物延伸、逆转录、线性延伸、非指数扩增、指数扩增、PCR或一个或多个靶或模板多核苷酸的任何其他扩增方法。

靶特异性区域可以包含至少约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个、100个、150个、200个、220个、230个、240个、250个、260个、270、280个、290个、300个、310个、320个、330个、340个、350个、360个、370个、380个、390个、400个、410个、420个、430个、440个、450个、460个、470个、480个、490个、500个、510个、520个、530个、540个、550个、560个、570个、580个、590个、600个、650个、700个、750个、800个、850个、900个或1000个核苷酸或碱基对。在另一个例子中，靶特异性区域包含至少约1500个、2000个、2500个、3000个、3500个、4000个、4500个、5000个、5500个、6000个、6500个、7000个、7500个、8000个、8500个、9000个、9500个或10000个核苷酸或碱基对。在一些实施方案的中，靶特异性区域包含从约5-10个、10-15个、10-20个、10-30个、15-30个、10-75个、15-60个、15-40个、18-30个、20-40个、21-50个、22-45个、25-40个、7-9个、12-15个、15-20个、15-25个、15-30个、15-45个、15-50个、15-55个、15-60个、20-25个、20-30个、20-35个、20-45个、20-50个、20-55个、20-60个、2-900个、2-800个、2-700个、2-600个、2-500个、2-400个、2-300个、2-200个、2-100个、25-900个、25-800个、25-700个、25-600个、25-500个、25-400个、25-300个、25-200个、25-100个、100-1000个、100-900个、100-800个、100-700个、100-600个、100-500个、100-400个、100-300个、100-200个、200-1000个、200-900个、200-800个、200-700个、200-600个、200-500个、200-400个、200-300个、300-1000个、300-900个、300-800个、300-700个、300-600个、300-500个、300-400个、400-1000个、400-900个、400-800个、400-700个、400-600个、400-500个、500-1000个、500-900个、500-800个、500-700个、500-600个、600-1000个、600-900个、600-800个、600-700个、700-1000个、700-900个、700-800个、800-1000个、800-900个或900-1000个核苷酸或碱基对。

可以根据已知的参数设计引物，以避免二级结构和自杂交。在一些实施方案中，不同的引物对可以在大约相同的温度下进行退火和解链，例如在另一个引物对的1°C、2°C、3°C、4

°C、5°C、6°C、7°C、8°C、9°C或10°C之内。在一些实施方案中,多个引物中的一个或多个引物可以在大约相同的温度下退火和解链,例如在所述多个引物中的另一个引物的1°C、2°C、3°C、4°C、5°C、6°C、7°C、8°C、9°C或10°C之内。在一些实施方案中,多个引物中的一个或多个引物可以在与所述多个引物中的另一个引物不同的温度下退火和解链。

用于本文所述的方法的一个或多个步骤的多个引物可以包含含有约、至多约或至少约1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个、1500个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个、10,000个、11,000个、12,000个、13,000个、14,000个、15,000个、16,000个、17,000个、18,000个、19,000个、20,000个、30,000个、40,000个、50,000个、60,000个、70,000个、80,000个、90,000个、100,000个、200,000个、300,000个、400,000个、500,000个、600,000个、700,000个、800,000个、900,000个、1,000,000个、50,000,000个、100,000,000个不同引物的多个引物。例如,多个引物中的每个引物可以包含不同的靶或模板特异性区域或序列。

图2描绘了通过本文提供的方法产生的本文提供的多核苷酸文库的扩增的示例性扩增反应。在本文提供的示例性方法中,出于测序目的的文库扩增使用与将被用于测序(如下一代测序)的测序衔接子连接的引物。此类引物是已知的并且在本文中有所描述。将测序衔接子标记的引物用于下面提供的示例性应用中。

在一些实施方案中,将靶基因扩增用于测序。在一些实施方案中,使用针对在多核苷酸一个末端的衔接子序列的引物和定位以对全长靶多肽或其选定部分进行测序的靶特异性引物来扩增靶基因。在一些例子中,靶序列可以存在于来自单细胞或多个细胞的文库中。在一些实施方案中,使用对多核苷酸的通用引发序列具有特异性的引物和一个或多个靶特异性引物来扩增一个或多个靶序列。在一些实施方案中,扩增了两个或更多个靶序列,每个靶序列使用如所述的通用序列和靶特异性引物。在一些实施方案中,所述两个或更多个靶序列是连接在一起的,如在细胞中共表达的两个靶序列,例如表达为二聚体(例如,异源二聚体)的靶序列。因此,使用所提供的实施方案,可以使用所提供的方法获得配对序列信息,如全长配对序列信息。

在一些实施方案中,可以使用对第一衔接子的通用引发序列和第二衔接子的通用引发序列具有特异性的引物扩增完整制备的多核苷酸文库用于测序。使用对文库的多核苷酸的两个末端的通用引发序列具有特异性的引物扩增本文提供的多核苷酸文库可以提供用于制备文库的所有细胞的转录组或基因组或其部分。此类转录组信息可以用于在以后的时间点进行挖掘和/或用于评价在制备样品的细胞群内几个基因的表达(在转录物水平上)。在一些实施方案中,可以分析所有细胞的转录组信息,并将其用于从产生文库的细胞总群产生具有相似转录物表达谱的细胞簇。

在一些实施方案中,可以使用对容器条码序列具有特异性的引物和对第二衔接子中存在的通用引发位点具有特异性的引物对来自单细胞的多核苷酸进行特异性扩增和测序。在此类实施方案中,如上所述扩增一个或多个靶序列,并且在被鉴定为感兴趣的一个或多个靶序列中鉴定一个或多个容器条码。由于来自同一细胞的所有多核苷酸都用相同的容器条码进行条码化,因此所述方法的此应用产生来自一个或多个选定细胞的所有多核苷酸的序

列信息。然后,来自一个或多个选定细胞的文库中所有多核苷酸的扩增可以提供表达所述一个或多个特定靶序列的一个或多个细胞的表达谱或遗传谱。

## 2. 测序

在进行本文所述的一种或多种方法或一个或多个方法步骤之后,可以对产生的多核苷酸文库进行测序。

可以通过本领域已知的任何测序方法进行测序。在一些实施方案中,能以高通量进行测序。合适的下一代测序技术包括454Life Sciences平台(Roche,康涅狄格州布兰福德(Branford, CT)) (Margulies等人, *Nature*, 437, 376-380 (2005)) ; Illumina's Genome Analyzer, GoldenGate Methylation Assay, 或 Infinium Methylation Assays, 即, Infinium HumanMethylation 27K BeadArray 或 VeraCode GoldenGate 甲基化测定(Illumina, 加利福尼亚州圣地亚哥(San Diego, CA); Bibkova等人, *Genome Res.* 16, 383-393 (2006); 以及美国专利号6,306,597、7,598,035、7,232,656), 或通过连接进行DNA测序, SOLiD系统(DNA Sequencing by Ligation, SOLiD System) (Applied Biosystems/Life Technologies; 美国专利号6,797,470、7,083,917、7,166,434、7,320,865、7,332,285、7,364,858和7,429,453); 或 Helicos True Single Molecule DNA测序技术(Harris等人, *Science*, 320, 106-109 (2008); 以及美国专利号7,037,687、7,645,596、7,169,560和7,769,400), Pacific Biosciences的单分子实时(SMRT™)技术, 和测序(Soni等人, *Clin. Chem.* 53, 1996-2001 (2007))。这些系统允许对从样品分离的许多多核苷酸进行多重平行测序(Dear, *Brief Funct. Genomic Proteomic*, 1 (4), 397-416 (2003) 以及 McCaughan等人, *J. Pathol.* 220, 297-306 (2010))。

在一些实施方案中,通过经由染料修饰探针连接的测序、焦磷酸测序或单分子测序来对多核苷酸进行测序。确定多核苷酸的序列可以通过测序方法进行,所述测序方法例如为 Helioscope™单分子测序、纳米孔DNA测序、Lynx Therapeutics的大规模平行标记测序(MPSS)、454焦磷酸测序、单分子实时(RNAP)测序、Illumina(Solexa)测序、SOLiD测序、Ion Torrent™、离子半导体测序、单分子SMRT(TM)测序、Polony测序、DNA纳米球测序和VisiGen Biotechnologies方法。可替代地,确定多核苷酸的序列可以使用测序平台,包括但不限于由 Illumina 提供的 Genome Analyzer IIx, HiSeq 和 MiSeq; 单分子实时(SMRT™)技术, 如由 Pacific Biosciences(加利福尼亚州)提供的 PacBio RS 系统, 以及 Solexa 测序仪; 真正单分子测序(tSMSTM)技术, 如由 Helicos Inc. (马萨诸塞州剑桥(Cambridge, MA)) 提供的 HeliScope™测序仪。测序可以包括 MiSeq 测序。测序可以包括 HiSeq 测序。在一些实施方案中,确定多核苷酸的序列包括配对末端测序、纳米孔测序、高通量测序、散弹枪测序、染料-终止子测序、多引物DNA测序、引物步移、Sanger双脱氧测序、Maxim-Gilbert测序、焦磷酸测序、真正单分子测序或其任何组合。可替代地,可以通过电子显微镜检查或化学敏感场效应晶体管(chemFET)阵列来确定多核苷酸的序列。

方法还可以包括对文库中的一个或多个多核苷酸进行测序。方法还可以包括将文库中的一个或多个多核苷酸序列、序列读段、扩增子序列或扩增子组序列彼此进行比对。

如本文所用,比对包括将测试序列(如序列读段)与一个或多个其他测试序列、参考序列或其组合进行比较。在一些实施方案中,比对可以用于从多个序列或比对序列确定共有序列。在一些实施方案中,比对包括从各自具有相同的分子条码或容器条码的多个序列确

定共有序列。在一些实施方案中,出于比较目的比对的序列的长度是参考序列长度的至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少95%。两个或更多个序列的实际比较可以通过熟知的方法(例如使用数学算法)来实现。这样的数学算法的非限制性例子描述于Karlin,S.和Altschul,S.,Proc.Natl.Acad.Sci.USA,90:5873-5877(1993)中。将这样的算法并入NBLAST和XBLAST程序(2.0版)中,如Altschul,S.等人,Nucleic Acids Res.,25:3389-3402(1997)中所述。当使用BLAST和Gapped BLAST程序时,可以使用相应程序(例如,NBLAST)的任何相关参数。例如,用于序列比较的参数可以设置为得分=100、字长=12,或者可以改变(例如,W=5或W=20)。其他例子包括Myers和Miller的算法,CABIOS(1989)、ADVANCE、ADAM、BLAT和FASTA。在一些实施方案中,两个氨基酸序列之间的百分比同一性可以使用例如GCG软件包(Accelrys,英国剑桥(Cambridge,UK))中的GAP程序来实现。

测序可以包括对多核苷酸的至少约10个、20个、30个、40个、50个、60个、70个、80个、90个、100个或更多个核苷酸或碱基对进行测序。在一些实施方案中,测序包括对多核苷酸的至少约200个、300个、400个、500个、600个、700个、800个、900个、1000个或更多个核苷酸或碱基对进行测序。在其他情形中,测序包括对多核苷酸的至少约1500个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个、10,000个或更多个核苷酸或碱基对进行测序。

测序可以包括每次运行至少约200个、300个、400个、500个、600个、700个、800个、900个、1000个或更多个测序读段。如本文所用,序列读段包含由通过测序技术产生的序列或数据流确定的核苷酸序列。在一些实施方案中,测序包括每次运行对至少约1500个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个、10,000个或更多个测序读段进行测序。测序可以包括每次运行大于、小于或等于约1,000,000,000个测序读段。测序可以包括每次运行大于、小于或等于约200,000,000个读段。

在一些实施方案中,用于确定共有序列的序列读段的数量为从约2-1000个序列读段。例如,用于确定共有序列的序列读段的数量可以为从约2-900个、2-800个、2-700个、2-600个、2-500个、2-400个、2-300个、2-200个、2-100个、25-900个、25-800个、25-700个、25-600个、25-500个、25-400个、25-300个、25-200个、25-100个、100-1000个、100-900个、100-800个、100-700个、100-600个、100-500个、100-400个、100-300个、100-200个、200-1000个、200-900个、200-800个、200-700个、200-600个、200-500个、200-400个、200-300个、300-1000个、300-900个、300-800个、300-700个、300-600个、300-500个、300-400个、400-1000个、400-900个、400-800个、400-700个、400-600个、400-500个、500-1000个、500-900个、500-800个、500-700个、500-600个、600-1000个、600-900个、600-800个、600-700个、700-1000个、700-900个、700-800个、800-1000个、800-900个或900-1000个序列读段。在一些实施方案中,用于确定共有序列的序列读段的数量为至少约1000个、1500个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个、10,000个、11,000个、12,000个、13,000个、14,000个、15,000个、16,000个、17,000个、18,000个、19,000个、20,000个、25,000个、30,000个、35,000个、40,000个、45,000个、50,000个、55,000个、60,000个、65,000个、70,000个、75,000个、80,000个、85,000个、90,000个、95,000个、100,000个、150,000个、200,000个、250,000个、300,000个、350,000个、400,000个、450,000个、500,000个、550,000个、600,

000个、650,000个、700,000个、750,000个、800,000个、850,000个、900,000个、950,000个、1,000,000个、50,000,000个或100,000,000个读段。在一些实施方案中,用于确定共有序列的序列读段的数量为至多约1000个、1500个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个、10,000个、11,000个、12,000个、13,000个、14,000个、15,000个、16,000个、17,000个、18,000个、19,000个、20,000个、25,000个、30,000个、35,000个、40,000个、45,000个、50,000个、55,000个、60,000个、65,000个、70,000个、75,000个、80,000个、85,000个、90,000个、95000个、100,000个、150,000个、200,000个、250,000个、300,000个、350,000个、400,000个、450,000个、500,000个、550,000个、600,000个、650,000个、700,000个、750,000个、800,000个、850,000个、900,000个、950,000个、1,000,000个、50,000,000个或100,000,000个读段。

方法可以包括对错误读段进行测序。方法可以包括确定错误读段的数量,例如用于确定反应条件或设计引物序列。比较在一个或多个第一条件或条件组下产生的错误读段的数量可以用于确定优选条件或条件组。例如,第一方法可以在PCR反应期间以高盐浓度进行,并且第二方法可以在PCR反应期间以低盐浓度进行,其中所述第一方法和所述第二方法除了盐浓度差异之外基本上相同地进行。如果第一方法产生较高质量的错误读段,如针对特定靶多核苷酸序列或引物而言较高质量的错误读段,则可以确定较低的盐反应条件对于所述特定靶多核苷酸序列或引物是优选的。

## II. 靶基因的克隆和表达

在一些实施方案中,可以将根据所提供的方法鉴定和测序的靶基因克隆到在细胞内表达或从细胞表达的载体中。可以产生靶基因(如免疫受体,例如抗体或TCR)的表达文库。

如本文所用的“抗体表达文库”或“TCR表达文库”或“表达文库”可以指代在核酸或蛋白质水平上的分子(即,两个或更多个分子)的集合。因此,此术语可以指代编码多个抗体或TCR分子的表达载体的集合(即,在核酸水平上),或者可以指代它们已在适当表达系统中表达之后的抗体或TCR分子的集合(即,在蛋白质水平上)。可替代地,表达载体/表达文库可以包含在它们可以在其中表达的合适的宿主细胞中。在本发明的表达文库中编码或表达的抗体分子可以呈任何适当的形式,例如可以是完整抗体或TCR分子或者可以是抗体或TCR片段,例如单链抗体(例如,scFv抗体)、Fv抗体、Fab'抗体、(Fab')<sub>2</sub>片段、双抗体等。如“编码”(“encoding' V' coding for”)特定酶的核酸序列或特定酶的DNA编码序列或“编码”(“encoding' V' coding for”)特定酶的核苷酸序列中的术语“编码”(“encoding”和“coding for”)以及其他同义术语是指当置于适当调节序列的控制下时被转录并翻译成酶的DNA序列。“启动子序列”是能够结合细胞中的RNA聚合酶并启动下游(3'方向)编码序列的转录的DNA调节区域。启动子是DNA序列的一部分。此序列区域在其3'末端具有起始密码子。启动子序列包括最小数量的碱基与在高于背景的可检测水平上启动转录所必需的元件。然而,在RNA聚合酶结合序列并在起始密码子(带有启动子的3'末端)启动转录之后,转录在3'方向上向下游进行。在启动子序列内将发现转录起始位点(通过用核酸酶SI作图方便地界定)以及负责结合RNA聚合酶的蛋白质结合结构域(共有序列)。

通过本发明的抗体或TCR表达文库鉴定、衍生自、选自或可获得自本发明的抗体或TCR表达文库的抗体或TCR分子形成本发明的又另一方面。同样,这些抗体或TCR分子可以是蛋白质或编码抗体或TCR分子的核酸,所述核酸转而又可以掺入适当的表达载体中和/或被包

含在合适的宿主细胞中。

使用与抗体基因重链的恒定区杂交的多核苷酸以及与抗体或TCR基因的V<sub>H</sub>或V<sub>α</sub>或V<sub>γ</sub>链区的5'末端杂交的多核苷酸可以将cDNA池进行PCR反应。使用与抗体或TCR基因的重链或α或γ链的恒定区杂交的多核苷酸以及与在包含抗体或TCR序列的条码化多核苷酸的V<sub>H</sub>或V<sub>α</sub>或V<sub>γ</sub>链区的5'末端的5'的区域杂交的多核苷酸可以将cDNA池进行PCR反应。还可以建立PCR反应来扩增例如κ和λ类别的V<sub>L</sub>或V<sub>β</sub>或V<sub>γ</sub>链池。使用与抗体基因轻链的恒定区杂交的多核苷酸以及与抗体或TCR基因的V<sub>L</sub>或V<sub>β</sub>或V<sub>γ</sub>链区的5'末端杂交的多核苷酸可以将cDNA池进行PCR反应。使用与抗体基因轻链的恒定区杂交的多核苷酸以及与在包含抗体或TCR序列的条码化多核苷酸的V<sub>L</sub>或V<sub>β</sub>或V<sub>γ</sub>链区的5'末端的5'的区域杂交的多核苷酸可以将cDNA池进行PCR反应。可以基于已知的和可公开获得的免疫球蛋白或TCR基因序列数据库信息来设计此类寡核苷酸或引物。

在一些实施方案中,可以从使用一个或多个引物通过PCR扩增产生的V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>序列的文库方便地获得V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>序列,所述一个或多个引物对重链或轻链基因不具有特异性,并且具体地,对V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>多核苷酸的一个或两个末端区域不具有特异性。在一些实施方案中,可以从使用对容器条码化多核苷酸的区域具有特异性的引物通过PCR扩增产生的V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>序列的文库方便地获得V<sub>H</sub>和V<sub>L</sub>序列。在一些实施方案中,可以从使用C基因家族特异性引物或C基因特异性引物通过PCR扩增产生的V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>序列的文库方便地获得V<sub>H</sub>和V<sub>L</sub>序列。在一些实施方案中,可以从使用引物组通过PCR扩增产生的V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>序列的文库方便地获得V<sub>H</sub>和V<sub>L</sub>序列,所述引物组具有对容器条码化多核苷酸的区域具有特异性的第一引物以及为C基因家族特异性引物或C基因特异性引物的第二引物或多个第二引物。在一些实施方案中,可以从使用引物组通过PCR扩增产生的V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>序列的文库方便地获得V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>序列,所述引物组具有对容器条码化多核苷酸的区域具有特异性的第一引物以及对通用序列具有特异性的第二引物。

在一些实施方案中,在逆转录后,可以使用一个或多个引物通过PCR来扩增所得的cDNA序列,所述一个或多个引物对免疫球蛋白基因具有特异性,并且具体地,对V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>多核苷酸的一个或两个末端区域具有特异性。在一些实施方案中,V<sub>H</sub>和V<sub>L</sub>序列可以从使用V基因家族特异性引物或V基因特异性引物通过PCR扩增产生的V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>序列的文库获得(Nicholls等人,J. Immunol. Meth., 1993, 165:81; WO 93/12227)或者基于可用序列信息根据标准技术已知的方法进行设计。(通常可以将V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>序列与插入间隔子序列(例如,编码框内柔性肽间隔子)连接在一起,形成编码单链抗体的盒)。对于表达免疫球蛋白的细胞,可以方便地将V区域序列作为cDNA或PCR扩增产物克隆。V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>区域任选地在本文所述的方法中进行测序,并且特别是在所述的某些步骤之后(例如,在单细胞PCR之后;在哺乳动物或其他细胞表面展示之后;在FACS筛选之后,等等)。除其他原因之外,还可以使用测序来验证多样性水平是否在可接受的水平上。测序可以包括高通量测序、深度测序(其中对来自多个个体样品的相同基因进行测序以鉴定序列差异)或两者的组合。

在一些实施方案中,没有必要使用本文所述的方法物理地连接天然V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>组合。在一些实施方案中,没有将cDNA、条码化多核苷酸或PCR扩增的条码化cDNA物

理地连接。在一些实施方案中,没有将cDNA、条码化多核苷酸或PCR扩增的条码化cDNA在同一反应或容器中物理地连接。

在一些实施方案中,除cDNA引物之外,还使用针对VH或V $\alpha$ 或V $\gamma$ 基因5'末端的一个引物或多个引物以及针对VL或V $\beta$ 或V $\delta$ 基因5'末端的另一个引物或多个引物来物理地连接天然VH和VL或V $\alpha$ 和V $\beta$ 或V $\gamma$ 和V $\delta$ 组合。这些引物还含有额外序列的互补尾,以允许VH和VL或V $\alpha$ 和V $\beta$ 或V $\gamma$ 和V $\delta$ 基因的自组装。在PCR扩增和连接之后,因为扩增和连接反应是在每个细胞内进行的,获得混合产物(换言之,混合可变区)的机会最小。通过使用体积大的试剂(如地高辛标记的核苷酸)可以进一步降低混合的风险,以进一步确保V区域cDNA对不离开细胞区室并相互混合,而是保留在细胞内进行PCR扩增和连接。通过互补末端序列杂交连接扩增的序列。在连接之后,可以从细胞中回收序列,以用于在本文所述的另外的方法步骤中使用。例如,如果需要,可以将回收的DNA使用末端引物进行PCR扩增,并克隆到载体中,所述载体可以是质粒、噬菌体、粘粒、噬菌粒、病毒载体或其组合,如下文详述。可以将便利的限制性酶切位点掺入杂交序列中来促进克隆。也可以将这些载体保存为连接可变区的文库,供以后使用。

在希望提供另外的VH和VL或V $\alpha$ 和V $\beta$ 或V $\gamma$ 和V $\delta$ 组合的一些实施方案中,选择表达式系统来促进此操作。例如,噬菌体表达系统允许重链和轻链序列的随机重组。其他合适的表达系统对本领域技术人员而言是已知的。

应当注意,在衍生自非人的VH和VL或V $\alpha$ 和V $\beta$ 或V $\gamma$ 和V $\delta$ 序列的情况下,在一些实施方案中,可能优选的是将这些序列与完全人Fc嵌合。如本文所用,“嵌合的”是指如下免疫球蛋白或TCR,其中重链和轻链可变区或V $\alpha$ 和V $\beta$ 或V $\gamma$ 和V $\delta$ 区域不是人起源的并且其中重链和轻链或V $\alpha$ 和V $\beta$ 或V $\gamma$ 和V $\delta$ 链的恒定区是人起源的。这受将可变结构域扩增并克隆到人Fc中的影响。人Fc可以是载体的一部分,或者在单独的分子中,或者也可以使用Fc文库。在一个优选的实施方案中,在哺乳动物细胞(如CHO细胞)中生长的嵌合分子用FACS筛选两次以使细胞群富集表达感兴趣抗体的细胞。嵌合抗体或TCR通过测序随后进行功能表征、或直接进行功能表征或动力学来表征。在下面详细描述生长、筛选和表征。

重要的是注意,出于克隆呈IgG形式的抗体而描述上述PCR反应。这些是优选的,因为它们通常与更成熟的免疫反应相关并且通常展现出比IgM抗体更高的亲和力,从而使它们对于某些治疗和诊断应用而言更加理想。然而,显然,可以设计如下多核苷酸,如果需要或适当的话,所述多核苷酸将允许克隆一种或多种其他形式的免疫球蛋白分子(例如,IgM、IgA、IgE和IgD)。

一旦鉴定出抗体或TCR,并且所述细胞的适当群体已经在适当时间被分离并任选地如上所述进行富集,就无需立即产生抗体或TCR表达文库,条件是细胞中含有的遗传材料可以保持完整,从而使得在以后的日期能够制备文库。因此,例如,可以通过适当的方法(例如,通过冷冻)将细胞、细胞裂解物或核酸(例如,源自其中的RNA或DNA)保存直到以后的日期,并且在需要时在以后的日期产生表达文库。

一旦已经产生表达载体文库,然后编码的抗体分子就可以在适当的表达系统中表达并使用本领域中熟知和记录的适当技术进行筛选。因此,本发明的上文定义的方法可以包括另外的步骤,即在适当的表达系统中使表达载体文库表达,并在表达的文库中筛选具有所需特性的抗体,如下文进一步详细解释。

如本文所指示,通过本公开文本的方法制备的多核苷酸(其包含编码抗体或TCR序列的多核苷酸)可以包括但不限于:编码抗体或TCR片段的氨基酸序列的那些自身;完整抗体或TCR或其部分的非编码序列;抗体或TCR、片段或部分的编码序列;以及另外的序列,如至少一个信号前导序列或融合肽的编码序列,具有或不具有上述另外的编码序列,如至少一个内含子,连同另外的非编码序列,包括但不限于非编码5'和3'序列,如在转录、mRNA加工(包括剪接和聚腺苷酸化信号(例如核糖体结合和mRNA的稳定性))中起作用的经转录的非翻译序列;编码另外的氨基酸(如提供另外的功能的那些)的另外的编码序列。因此,可以将编码抗体的序列与标记序列融合,所述标记序列例如为编码促进融合抗体或包含抗体的TCR或TCR片段或部分纯化的肽的序列。

然后可以任选地将初级PCR产物使用新的多核苷酸组进行二级PCR反应,所述新的多核苷酸组与抗体或TCR可变结构域V<sub>H</sub>、V<sub>LK</sub>和V<sub>L</sub>λ或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>的5'和3'末端杂交(视情况取决于与新的多核苷酸组一起使用的初级PCR反应是否被设计成扩增重链或轻链抗体基因或V<sub>α</sub>或V<sub>β</sub>TCR基因或V<sub>γ</sub>或V<sub>δ</sub>TCR基因的部分)。这些多核苷酸有利地包括用于随后克隆的对一组限定的限制酶具有特异性的DNA序列(即,限制性酶切位点)。必须选择所选定的限制酶,以便不会在人抗体或TCR V基因区段内切割。可以基于已知的和可公开获得的免疫球蛋白或TCR基因序列和限制酶数据库信息来设计此类多核苷酸。然而,将要包括的优选限制性酶切位点是NcoI、Hind III、M和NotI。此类二级PCR反应的产物是各种V-重、V-轻κ和V-轻λ抗体片段/结构域的库。因此,通常当感兴趣的表达文库形式为scFv或Fv形式(其中仅存在抗体或TCR的V<sub>H</sub>和V<sub>L</sub>或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>结构域)时进行这种类型的二级PCR反应。

也可以将PCR产物使用新的引物组进行PCR反应,所述新的引物组与条码化多核苷酸的5'和3'末端杂交。这些多核苷酸可以有利地包括用于随后克隆的对一组限定的限制酶具有特异性的DNA序列(即,限制性酶切位点)。必须选择所选定的限制酶,以便不会在人抗体或TCR V基因区段内切割。可以基于已知的和可公开获得的免疫球蛋白或TCR基因序列和限制酶数据库信息来设计此类多核苷酸。然而,将要包括的优选限制性酶切位点是NcoI、Hind III、MluI和NotI。此类二级PCR反应的产物是各种V<sub>H</sub>、V<sub>LK</sub>和V<sub>L</sub>λ抗体片段/结构域或V<sub>α</sub>和V<sub>β</sub>或V<sub>γ</sub>和V<sub>δ</sub>TCR片段/结构域的库。

本领域技术人员将认识到,重链或轻链或V<sub>α</sub>或V<sub>β</sub>链或V<sub>γ</sub>或V<sub>δ</sub>链Fv或Fab片段或者单链抗体或TCR也可以与此系统一起使用。可以诱变重链或轻链或V<sub>α</sub>或V<sub>β</sub>链或V<sub>γ</sub>或V<sub>δ</sub>链,随后将互补链添加到溶液中。然后允许两条链组合并形成功能性抗体片段。随机非特异性轻链或重链或V<sub>α</sub>或V<sub>β</sub>链或V<sub>γ</sub>或V<sub>δ</sub>链序列的添加允许组合系统的产生,以产生不同成员的文库。

如本文所定义的包含从免疫激发的宿主的B或T淋巴细胞衍生的抗体或TCR基因的可变重链或V<sub>α</sub>链或V<sub>γ</sub>链区或其片段、和/或可变轻链或V<sub>β</sub>链或V<sub>δ</sub>链区或其片段的克隆片段的此类库的文库形成本发明的另外方面。可以任选地将包含克隆的可变区的这些文库插入表达载体中以形成表达文库。

在一些实施方案中,可以建立PCR反应以便保留分离的免疫细胞群中包含的各种抗体或TCR链的全部或部分恒定区。当表达文库形式是Fab形式,其中重链或α链或γ链组分包含V<sub>H</sub>或V<sub>α</sub>或V<sub>γ</sub>以及C<sub>H</sub>或C<sub>α</sub>或C<sub>γ</sub>结构域并且轻链或V<sub>β</sub>链或V<sub>δ</sub>链组分包含V<sub>L</sub>或V<sub>β</sub>或V<sub>δ</sub>链以及CL或β或δ结构域时,这是令人希望的。同样,包含抗体或TCR链的全部或部分恒定区的此类克隆片段文库形成本发明的另外方面。

这些核酸还可以方便地包含除本发明的多核苷酸之外的序列。例如,可以将包含一个或多个核酸内切酶限制位点的多克隆位点插入核酸中以帮助分离多核苷酸。同样,可以插入可翻译序列以帮助分离本发明的经翻译的多核苷酸。例如,六组氨酸标记序列提供了纯化本发明的蛋白质的便利手段。排除编码序列的本发明的核酸任选地是用于克隆和/或表达本发明的多核苷酸的载体、衔接子或接头。

可以将另外的序列添加到此类克隆和/或表达序列中,以优化其在克隆和/或表达中的功能,以帮助分离多核苷酸或以改善多核苷酸向细胞中的引入。克隆载体、表达载体、衔接子和接头的使用在本领域中是熟知的。(参见例如,Current Protocols in Molecular Biology (CPMB) (Fred M.Ausubel等人编辑,John Wiley and Sons,Inc.);或者J.Sambrook等人,“Molecular Cloning:A Laboratory Manual,”1989,第2版,Cold Spring Harbour Laboratory Press:New York,N.Y.)。

本文公开的文库可以用于多种应用中。如本文所用,文库包含多个分子。在一些实施方案中,文库包含多个多核苷酸。在一些实施方案中,文库包含多个引物。在一些实施方案中,文库包含来自一个或多个多核苷酸、扩增子或扩增子组的多个序列读段。可以将文库储存并使用多次来产生样品进行分析。一些应用包括例如基因分型多态性、研究RNA加工以及选择克隆代表来根据本文所提供的方法进行测序。可以产生包含多个多核苷酸的文库,如用于测序或扩增的引物或文库,其中多个多核苷酸包含至少约2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、20个、25个、30个、35个、40个、45个、50个、60个、70个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个、15000个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个、10,000个、11,000个、12,000个、13,000个、14,000个、15,000个、16,000个、17,000个、18,000个、19,000个、20,000个、30,000个、40,000个、50,000个、60,000个、70,000个、80,000个、90,000个、100,000个、200,000个、300,000个、400,000个、500,000个、600,000个、700,000个、800,000个、900,000个、1,000,000个、50,000,000个、100,000,000个或更多个分子条码或容器条码。在一些实施方案中,多核苷酸文库包含至少约2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、20个、25个、30个、35个、40个、45个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个、1500个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个、10,000个、11,000个、12,000个、13,000个、14,000个、15,000个、16,000个、17,000个、18,000个、19,000个、20,000个、30,000个、40,000个、50,000个、60,000个、70,000个、80,000个、90,000个、100,000个、200,000个、300,000个、400,000个、500,000个、600,000个、700,000个、800,000个、900,000个、1,000,000个、50,000,000个、100,000,000个或更多个独特的多核苷酸中的多个,其中每个独特的多核苷酸包含一个或多个分子条码和容器条码。

### III. 转录组分析

在一些实施方案中,所提供的方法可以用于阐明细胞的转录组信息,所述转录组信息可以与特定细胞的靶多核苷酸序列(例如,抗体或TCR)的捕获组合。在一些实施方案中,通过靶多核苷酸序列鉴定的一个细胞或多个细胞可以通过它们的转录细胞状态(包括关于细胞的特定状态、特征或属性)来表征。在一些实施方案中,可以确定细胞的个性化转录组谱,并提供与细胞的一个或多个特征有关的信息。此类特征的例子包括但不限于激活状态、增

殖状态、耗竭状态、过渡状态、细胞周期阶段或与细胞的功能或表型状态相关的其他参数。

在一些实施方案中,转录状态信息可以用于鉴定表达展现出所需或感兴趣的反应的特定靶基因(例如,抗体或TCR)的细胞。在一些方面,所提供的方法允许细胞的转录组谱与从细胞扩增或测序的靶多核苷酸(例如,抗体或TCR)相匹配。在特定实施方案中,靶多核苷酸的转录组信息和序列通过具有相同容器条码的扩增的和测序的转录组和靶多核苷酸进行匹配,所述容器条码鉴定衍生自同一细胞的转录组谱和靶多核苷酸。

用于处理转录组数据的各种方法在本领域中是已知的。在一些方面,从所述方法获得的数据可以在图上可视化。可以使用通过使用本文所述的方法产生的信息来构建来自样品的靶的数量和位置的图。所述图可以用于定位靶的物理位置。所述图可以用于鉴定多个靶的位置。

在一些实施方案中,所述系统包含计算机可读介质,所述计算机可读介质包括用于对从所提供的方法产生的序列数据集提供数据分析的代码。可以由数据分析软件提供的数据分析功能的例子包括但不限于(i)用于解码/解复用通过对多核苷酸文库测序提供的样品容器条码、分子条码和靶序列或转录组数据的算法;(ii)用于基于数据确定每个细胞每个基因的读段数量以及每个细胞每个基因的独特转录物分子数量并且产生汇总表的算法;(iii)统计分析序列数据,例如用于通过基因表达数据对细胞进行聚类,或用于预测置信区间以确定每个细胞每个基因的转录物分子数量等;(iv)用于例如使用主成分分析、分层聚类、k均值聚类、自组织图、神经网络等来鉴定稀有细胞亚群的算法;(v)用于将基因序列数据与已知参考序列进行比对并检测突变、多态性标记和剪接变体的序列比对能力;以及(vi)将分子标记自动聚类以补偿扩增或测序错误。

在一些实施方案中,可以使用计算程序来产生转录组程序集。用于短读段程序集的示例性计算程序包括在Robertson等人,Nat Methods.2010;7:909-12;Grabherr等人,Nat Biotechnol.2011;29:644-52;Schulz等人,Bioinformatics.2012;28:1086-92以及Xie等人,Bioinformatics.2014;30:1660-6中所述的那些。由于转录物之间的表达水平变异较大、测序偏差和可变剪接,因此转录组组装可能具有挑战性。因此,基于k-mer长度或使用固定的k-mer值合并转录组程序集可以用于抵消不同程度的转录物丰度并改善转录组组装(参见例如,Robertson等人,Nat Methods.2010;7:909-12;Grabherr等人,Nat Biotechnol.2011;29:644-52以及Surget-Groba Genome Res.2010;20:1432-40)。

在一些实施方案中,可以使用可商购获得的软件来进行全部或部分数据分析。在一些实施方案中,数据分析软件可以包括用于以有用的图形格式(例如,指示在细胞集合的每个细胞中出现的一个或多个基因的拷贝数的热图)输出测序结果的选项。在一些实施方案中,数据分析软件还可以包含从测序结果中提取生物学意义的算法,例如通过将细胞集合的每个细胞中出现的一个或多个基因的拷贝数与细胞的类型、稀有细胞的类型或衍生自患有特定疾病或病症的受试者的细胞相关联。在一些实施方案中,数据分析软件还可以包含用于比较不同生物样品中的细胞群的算法。

在一些实施方案中,所有的数据分析功能都可以打包在单个软件包中。在一些实施方案中,数据分析能力的完整集可以包含一套软件包。在一些实施方案中,数据分析软件可以是独立运行的软件包,其可独立于测定仪器系统提供给用户。在一些实施方案中,所述软件可以是基于网络的,并且可以允许用户共享数据。

在一些例子中,可以进行聚类分析来鉴定一个或多个不同的细胞群。在一些例子中,基于一个或多个靶基因的表达将细胞群聚类。代表转录细胞状态的已知基因或转录物的存在、上调或下调可以用于将多个细胞内的细胞聚类。

在一些实施方案中,感兴趣的转录组基因通过其容器条码进行鉴定,并与感兴趣的全长靶分子(例如,抗体或TCR)的扩增和测序的输出相匹配或连接。因此,通过实践所提供的方法,每个容器条码都被匹配或者可以用基因计数(例如,转录组)和靶分子(例如,全长抗体或TCR)信息进行注释。

在一些实施方案中,可以产生由转录组中存在的核酸序列代表的表达谱。可以产生细胞亚组的表达谱,所述细胞例如为表达一个或多个靶基因并共享容器条码的细胞。

#### IV. 诊断

在一些实施方案中,方法还可以包括在受试者中诊断、预后、监测、治疗、改善和/或预防疾病、障碍、症状和/或病症。在一些实施方案中,方法还可以包括基于靶多核苷酸的存在、不存在或水平和/或特定的转录细胞状态(如本文上面所述的细胞状态)在受试者中诊断、预后、监测、治疗、改善和/或预防疾病、障碍、症状和/或病症。在一些实施方案中,方法还可以包括基于一个或多个靶多核苷酸的存在、不存在或水平和/或一个或多个细胞的转录状态在受试者中诊断、预后、监测、治疗、改善和/或预防疾病、障碍、症状和/或病症。

在一些实施方案中,方法还可以包括基于使用本文所述的方法获得的一个或多个序列的存在、不存在、水平或序列在受试者中诊断、预后、监测、治疗、改善和/或预防疾病、障碍、症状和/或病症。例如,可以基于使用本文所述的方法获得的变体序列的存在、不存在、水平或序列来进行疾病的诊断。在一些实施方案中,方法还可以包括基于使用本文所述的方法获得的一个或多个序列读段的存在、不存在、水平或序列在受试者中诊断、预后、监测、治疗、改善和/或预防疾病、障碍、症状和/或病症。在一些实施方案中,方法还可以包括基于使用本文所述的方法获得的一个或多个共有序列的存在、不存在、水平或序列在受试者中诊断、预后、监测、治疗、改善和/或预防疾病、障碍、症状和/或病症。在一些实施方案中,方法还可以包括基于样品中靶多核苷酸的水平(例如,量或浓度)的确定在受试者中诊断、预后、监测、治疗、改善和/或预防疾病、障碍、症状和/或病症。可以基于一个或多个序列读段、序列、共有序列或其任何组合来确定样品中靶多核苷酸的水平。可以使用本文所述的方法确定样品中多个靶多核苷酸中的每一个的水平。可以基于所述多个靶多核苷酸中的每一个靶多核苷酸的序列读段、序列、共有序列或其任何组合的数量来确定样品中多个靶多核苷酸中的每一个的水平。例如,可以使用本文所述的方法确定第一靶多核苷酸的水平和第二靶多核苷酸的水平。

在一些实施方案中,多个靶多核苷酸中的第一靶多核苷酸和第二靶多核苷酸是相同的。例如,第一靶多核苷酸可以包含mRNA分子的第一拷贝,并且第二靶多核苷酸可以包含mRNA分子的第二拷贝。在一些实施方案中,第一靶多核苷酸和第二靶多核苷酸是不同的。例如,第一靶多核苷酸可以包含第一mRNA分子,并且第二靶多核苷酸可以包含从与第一mRNA分子不同的基因转录的第二mRNA分子。例如,第一靶多核苷酸可以包含第一等位基因,并且第二靶多核苷酸可以包含第二等位基因。例如,第一靶多核苷酸可以包含野生型序列,并且第二靶多核苷酸可以包含变体序列。

在一些实施方案中,方法还可以包括以至少50%的置信度对患有疾病、障碍、症状和/

或病症的受试者进行诊断或预后。例如,能以至少50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%的置信度确定患有疾病、障碍、症状和/或病症的受试者的诊断或预后。在一些实施方案中,能以50%-100%的置信度确定患有疾病、障碍、症状和/或病症的受试者的诊断或预后。例如,能以60%-100%、70%-100%、80%-100%、90%-100%、50%-90%、50%-80%、50%-70%、50%-60%、60%-90%、60%-80%、60%-70%、70%-90%、70%-80%或80%-90%的置信度确定患有疾病、障碍、症状和/或病症的受试者的诊断或预后。

在一些实施方案中,能以至少50%的置信度确定受试者中靶多核苷酸(如生物标记)的存在、不存在、水平、序列或其任何组合。例如,能以至少50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%的置信度确定受试者中靶多核苷酸的存在、不存在、水平、序列或其任何组合。在一些实施方案中,能以50%-100%的置信度确定受试者中靶多核苷酸的存在、不存在、水平、序列或其任何组合。例如,能以60%-100%、70%-100%、80%-100%、90%-100%、50%-90%、50%-80%、50%-70%、50%-60%、60%-90%、60%-80%、60%-70%、70%-90%、70%-80%或80%-90%的置信度确定受试者中靶多核苷酸的存在、不存在、水平、序列或其任何组合。

#### V. 酶

本文公开的方法和试剂盒可以包括一种或多种酶。酶的例子包括但不限于连接酶、逆转录酶、聚合酶和限制性核酸酶。

在一些实施方案中,衔接子与多核苷酸的附接包括使用一种或多种连接酶。连接酶的例子包括但不限于DNA连接酶,如DNA连接酶I、DNA连接酶III、DNA连接酶IV和T4DNA连接酶;以及RNA连接酶,如T4RNA连接酶I和T4RNA连接酶II。

本文公开的方法和试剂盒还可以包括使用一种或多种逆转录酶。在一些实施方案中,逆转录酶是HIV-1逆转录酶、MMLV逆转录酶、AMV逆转录酶和端粒酶逆转录酶。在一些实施方案中,逆转录酶是M-MLV逆转录酶。

在一些实施方案中,本文公开的方法和试剂盒包括使用一种或多种蛋白酶。

在一些实施方案中,本文公开的方法和试剂盒包括使用一种或多种聚合酶。聚合酶的例子包括但不限于DNA聚合酶和RNA聚合酶。在一些实施方案中,DNA聚合酶是DNA聚合酶I、DNA聚合酶II、DNA聚合酶III全酶和DNA聚合酶IV。可商购获得的DNA聚合酶包括但不限于Bst 2.0DNA聚合酶、Bst2.0WarmStart<sup>TM</sup> DNA聚合酶、Bst DNA聚合酶、Sulfolobus DNA聚合酶IV、Taq DNA聚合酶、9°NTMm DNA聚合酶、Deep VentR<sup>TM</sup>(外切) DNA聚合酶、Deep VentR<sup>TM</sup> DNA聚合酶、Hemo KlenTaq<sup>TM</sup>、LongAmp<sup>®</sup> Taq DNA聚合酶、OneTaq<sup>®</sup> DNA聚合酶、Phusion<sup>®</sup> DNA聚合酶、Q5TM高保真DNA聚合酶、Therminator<sup>TM</sup> γ DNA聚合酶、Therminator<sup>TM</sup> DNA聚合酶、Therminator<sup>TM</sup> II DNA聚合酶、Therminator<sup>TM</sup> III DNA聚合酶、VentR<sup>®</sup> DNA聚合酶、VentR<sup>®</sup> (外切) DNA聚合酶、Bsu DNA聚合酶、phi29 DNA聚合酶、T4 DNA聚合酶、T7 DNA聚合酶、末端转移酶、Titanium<sup>®</sup> Taq聚合酶、KAPA Taq DNA聚合酶和KAPA Taq热启动DNA聚合酶。

在一些实施方案中,聚合酶是RNA聚合酶,如RNA聚合酶I、RNA聚合酶II、RNA聚合酶III、大肠杆菌聚(A)聚合酶、phi6 RNA聚合酶(RdRP)、聚(U)聚合酶、SP6 RNA聚合酶和T7RNA聚合

酶。

## VI. 试剂盒和另外的试剂

本文提供了包含用于进行所提供方法的一种或多种试剂的制品或试剂盒。试剂盒可以任选地包括一个或多个组件,如使用说明书、设备和另外的试剂(例如,用于稀释和/或重构试剂或组分的无菌水或盐水溶液);以及诸如用于实践所述方法的试管、容器和注射器等组件。在一些实施方案中,所述试剂盒还可以含有用于收集样品或制剂和处理样品的试剂。在一些实施方案中,所述试剂盒可以作为制品提供,所述制品包括用于包装试剂盒的试剂和组件的包装材料。例如,所述试剂盒可以含有容器、瓶子、管、小瓶以及适合用于分离或组织试剂盒的组件的任何包装材料。

本文公开的方法和试剂盒可以包括使用一种或多种试剂。试剂的例子包括但不限于PCR试剂、连接试剂、逆转录试剂、酶试剂、杂交试剂、样品制备试剂、亲和捕获试剂、固体支持物(如珠)以及用于核酸纯化和/或分离的试剂。

固体支持物可以包含几乎任何不溶性或固体材料,并且通常选择不溶于水中的固体支持物组合物。例如,固体支持物可以包含硅胶、玻璃(例如,可控孔度玻璃(CPG))、尼龙、Sephadex®、Sephadose®、纤维素、金属表面(例如,钢、金、银、铝、硅和铜)、磁性材料、塑料材料(例如,聚乙烯、聚丙烯、聚酰胺、聚酯、聚偏二氟乙烯(PVDF))等,或者基本上由其组成。根据实施方案使用的珠的例子可以包括允许珠与核酸分子相互作用的亲和部分。固相(例如,珠)可以包含结合对成员(例如,抗生物素蛋白、链霉亲和素或其衍生物)。例如,珠可以是包被链霉亲和素的珠,并且用于固定在珠上的核酸分子可以包括生物素部分。在一些情况下,每个多核苷酸分子可以包括两个亲和部分(如生物素),以进一步稳定多核苷酸。珠可以包括用于在固定核酸中使用或可以用于在下游筛选或选择过程中使用的另外特征。例如,珠可以包括结合部分、荧光标记或荧光淬灭剂。在一些情况下,珠可以是磁性的。在一些情形中,固体支持物是珠。珠的例子包括但不限于链霉亲和素珠、琼脂糖珠、磁珠、Dynabeads®、MACS®微珠、抗体缀合珠(例如,抗免疫球蛋白微珠)、蛋白A缀合珠、蛋白G缀合珠、蛋白A/G缀合珠、蛋白L缀合珠、多核苷酸-dT缀合珠、二氧化硅珠、二氧化硅样珠、抗生物素微珠、抗荧光色素微珠和BcMag™羧基封端磁珠。珠或颗粒可以是可溶胀的(例如,聚合珠,如Wang树脂)或不可溶胀的(例如,CPG)。在一些实施方案中,固相是基本上亲水的。在一些实施方案中,固相(例如,珠)是基本上疏水的。在一些实施方案中,固相包含结合对成员(例如,抗生物素蛋白、链霉亲和素或其衍生物),并且是基本上疏水的或基本上亲水的。在一些实施方案中,固相包含结合对成员(例如,抗生物素蛋白、链霉亲和素或其衍生物),并且具有每mg固体支持物大于约1350皮摩尔游离捕获剂(例如,游离生物素)的结合能力。在一些实施方案中,包含结合对成员的固相的结合能力为每mg固体支持物大于800、900、1000、1100、1200、1250、1300、1350、1400、1450、1500、1600、1800、2000皮摩尔自由捕获剂。适合用于本发明的珠的其他例子是金胶体或珠,如聚苯乙烯珠或二氧化硅珠。可以使用基本上任何珠半径。珠的例子可以包括半径在从150纳米至10微米范围内的珠。也可以使用其他大小。

本文公开的方法和试剂盒可以包括使用一种或多种缓冲液。缓冲液的例子包括但不限于洗涤缓冲液、连接缓冲液、杂交缓冲液、扩增缓冲液和逆转录缓冲液。在一些实施方案中,杂交缓冲液是可商购获得的缓冲液,如TMACHyb溶液、SSPE杂交溶液和ECONOTM杂交缓冲液。

本文公开的缓冲液可以包含一种或多种洗涤剂。

本文公开的方法和试剂盒可以包括使用一种或多种载体。载体可以增强或改善本文公开的一种或多种反应(例如,连接反应、逆转录、扩增、杂交)的效率。载体可以减少或防止分子或其任何产物(例如,多核苷酸和/或扩增子)的非特异性损失。例如,载体可以通过吸收到表面来减少多核苷酸的非特异性损失。载体可以降低多核苷酸对表面或基底(例如,容器、Eppendorf管、移液器吸头)的亲和力。可替代地,载体可以增加多核苷酸对表面或基底(例如,珠、阵列、玻璃、载玻片或芯片)的亲和力。载体可以保护多核苷酸免于降解。例如,载体可以保护RNA分子免受核糖核酸酶降解。可替代地,载体可以保护DNA分子免受DNA酶降解。载体的例子包括但不限于多核苷酸(如DNA和/或RNA)或多肽。DNA载体的例子包括质粒、载体、聚腺苷酸化DNA和DNA多核苷酸。RNA载体的例子包括聚腺苷酸化RNA、噬菌体RNA、噬菌体MS2RNA、大肠杆菌RNA、酵母RNA、酵母tRNA、哺乳动物RNA、哺乳动物tRNA、短的聚腺苷酸化合成核糖核苷酸和RNA多核苷酸。RNA载体可以是聚腺苷酸化RNA。可替代地,RNA载体可以是非聚腺苷酸化RNA。在一些实施方案中,载体来自细菌、酵母或病毒。例如,载体可以是衍生自细菌、酵母或病毒的多核苷酸或多肽。例如,载体是来自枯草芽孢杆菌(*Bacillus subtilis*)的蛋白质。在另一个例子中,载体是来自大肠杆菌(*Escherichia coli*)的多核苷酸。可替代地,载体是来自哺乳动物(例如,人、小鼠、山羊、大鼠、牛、绵羊、猪、狗或兔)、禽类、两栖动物或爬行动物的多核苷酸或肽。

本文公开的方法和试剂盒可以包括使用一种或多种对照剂。对照剂可以包括对照多核苷酸、非活性酶和/或非特异性竞争剂。可替代地,对照剂包含亮杂交(bright hybridization)、亮探针(bright probe)对照、核酸模板、掺加(spike-in)对照、PCR扩增对照。PCR扩增对照可以是阳性对照。在其他情形中,PCR扩增对照是阴性对照。核酸模板对照可以具有已知浓度。对照剂可以包含一个或多个标记。

掺加对照可以是添加到反应或样品中的模板。例如,可以将掺加模板添加到扩增反应中。可以在第一扩增循环之后的任何时间将掺加模板添加到扩增反应中。在一些实施方案中,在循环数2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45或50之后,将掺加模板添加到扩增反应中。可以在最后一个扩增循环之前的任何时间将掺加模板添加到扩增反应中。掺加模板可以包含一个或多个核苷酸或核酸碱基对。掺加模板可以包含DNA、RNA或其任何组合。掺加模板可以包含一个或多个标记。

本文公开了分子、材料、组合物和组分,它们可以用于本文公开的方法和组合物,可以与本文公开的方法和组合物结合使用,可以用于本文公开的方法和组合物的产物的制备中,或者是本文公开的方法和组合物的产物。应理解,当公开了这些材料的组合、亚组、相互作用、组等时并且在不能明确公开这些分子和化合物的每个不同单独和共同的组合以及排列的具体提及的时候,在本文中具体地考虑和描述了每一种。例如,如果公开和讨论了核苷酸或核酸并且讨论了可以对包括核苷酸或核酸的许多分子进行的许多修饰,则除非明确相反地说明,否则明确考虑了核苷酸或核酸的每一种组合和排列以及可能的修饰。此概念适用于本申请的所有方面,包括但不限于制备和使用所公开的方法和组合物的方法中的步骤。因此,如果存在可以进行的多种另外步骤,则应理解,这些另外步骤中的每一个都可以用所公开方法的任何特定实施方案或实施方案组合来进行,并且每个这样的组合都被明确考虑并且应当视为已公开。

尽管本文已经示出和描述了本文所述的一些实施方案,但是此类实施方案仅以举例的方式提供。在不背离本文提供的公开文本的情况下,本领域技术人员现在将想到许多变化、改变和替换。应当理解,本文所述的实施方案的各种替代方案可以用于实践本文所述的方法中。

除非另有解释,否则本文使用的所有技术和科学术语具有与本公开文本所属领域的普通技术人员通常理解的相同含义。以下参考文献含有本文可以使用的方法和组合物的实施方案: The Merck Manual of Diagnosis and Therapy, 第18版, 由Merck Research Laboratories出版, 2006 (ISBN 0-91 19102) ; Benjamin Lewin, Genes IX, 由Jones & Bartlett Publishing出版, 2007 (ISBN-13: 9780763740634) ; Kendrew等人(编辑), The Encyclopedia of Mol. Biology, 由Blackwell Science Ltd. 出版, 1994 (ISBN 0-632-02182-9) ; 以及 Robert A. Meyers(编辑), Mol. Biology and Biotechnology: a Comprehensive Desk Reference, 由VCH Publishers, Inc. 出版, 1995 (ISBN 1-56081-569-8)。

本公开文本的标准程序描述于例如以下参考文献中: Maniatis等人, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., USA (1982) ; Sambrook等人, Molecular Cloning: A Laboratory Manual (第2版), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., USA (1989) ; Davis等人, Basic Methods in Molecular Biology, Elsevier Science Publishing, Inc., New York, USA (1986) ; 或者 Methods in Enzymology: Guide to Molecular Cloning Techniques 第152卷, S.L. Berger 和 A.R. Kimmerl (编辑), Academic Press Inc., San Diego, USA (1987) ; Current Protocols in Molecular Biology (CPMB) (Fred M. Ausubel等人编辑, John Wiley and Sons, Inc.) ; Current Protocols in Protein Science (CPPS) (John E. Coligan等人编辑, John Wiley and Sons, Inc.) ; Current Protocols in Immunology (CPI) (John E. Coligan等人编辑, John Wiley and Sons, Inc.) ; Current Protocols in Cell Biology (CPCB) (Juan S. Bonifacino等人编辑, John Wiley and Sons, Inc.) ; Culture of Animal Cells: A Manual of Basic Technique 由 R. Ian Freshney 出版商出版: Wiley-Liss; 第5版 (2005) ; 以及 Animal Cell Culture Methods (Methods in Cell Biology, 第57卷, Jennie P. Mather 和 David Barnes 编辑, Academic Press, 第1版, 1998)。

## VII. 定义

本文使用的术语仅出于描述特定情况的目的,并不旨在进行限制。除非另有定义,否则本文使用的所有领域术语、符号和其他技术和科学术语或用辞旨在具有与所要求保护的主题所属领域的普通技术人员通常理解的相同含义。在一些情况下,为了清楚和/或为了便于引用而在本文中定义具有通常理解的含义的术语,并且本文中包含此类定义不一定应当被解释为代表与本领域通常所理解的含义有实质性差异。

如本文所用,单数形式“一种/一个”(“a”)、“一种/一个”(“an”)和“所述”(“the”)也旨在包括复数形式,除非上下文另有明确说明。此外,在术语“包括”(“including”)、“包括”(“includes”)、“具有”(“having”)、“具有”(“has”)、“具有”(“with”)或其变体在具体实施方式和/或权利要求书中使用的程度上,此类术语旨在以与术语“包含”(“comprising”)相

似的方式是包含性的。

术语“约”或“大约”可以意指在如通过本领域普通技术人员确定的特定值的可接受误差范围内,其部分取决于如何测量或确定所述值,即测量系统的局限性。例如,根据本领域的实践,“约”可以意指在1个或超过1个标准偏差内。可替代地,“约”可以意指给定值的至多20%、至多10%、至多5%或至多1%的范围。可替代地,特别是关于生物系统或过程,所述术语可以意指在值的数量级以内、在值的5倍以内、更优选地在值的2倍以内。在本申请和权利要求书中描述特定值的情况下,除非另有说明,否则应当假设术语“约”意指所述特定值在可接受的误差范围内。

术语“多肽”和“蛋白质”可互换使用以指代氨基酸残基的聚合物,并且不限于最小长度。多肽(包括所提供的抗体和抗体链以及其他肽(例如,接头和结合肽))可以包括含有天然和/或非天然氨基酸残基的氨基酸残基。所述术语还包括多肽的表达后修饰,例如糖基化、唾液酸化、乙酰化、磷酸化等。在一些方面,多肽可以含有关于原生或天然序列的修饰,只要蛋白质保持所需活性即可。这些修饰可能是故意的(如通过定点诱变),或者可能是偶然的(如通过产生蛋白质的宿主的突变或由于PCR扩增引起的错误)。

聚合酶链反应(PCR)是指通过双链多核苷酸的互补链的同时引物延伸进行多核苷酸序列的体外扩增反应。PCR反应产生侧翼为引物结合位点的模板多核苷酸的拷贝。使用两个引物的结果是每个循环中两条链的模板多核苷酸拷贝数呈指数增加,因为在每个循环中两条链均被复制。多核苷酸双链体具有与所使用引物的末端对应的末端。PCR可以包括以下的一次或多次重复:使模板多核苷酸变性,使引物与引物结合位点退火,并在核苷酸的存在下通过DNA或RNA聚合酶延伸引物。特定的温度、每个步骤的持续时间以及步骤之间的变化率取决于本领域普通技术人员熟知的许多因素。(McPherson等人,IRL Press, Oxford (1991和1995))。例如,在使用Taq DNA聚合酶的常规PCR中,双链模板多核苷酸可以在>90°C的温度下变性,引物可以在50°C-75°C范围内的温度下退火,并且引物可以在72°C-78°C范围内的温度下延伸。在一些实施方案中,PCR包括逆转录PCR(RT-PCR)、实时PCR、巢式PCR、定量PCR、多重PCR等。在一些实施方案中,PCR不包括RT-PCR。(美国专利号5,168,038、5,210,015、6,174,670、6,569,627和5,925,517; Mackay等人, Nucleic Acids Research, 30:1292-1305 (2002))。RT-PCR包括先前为逆转录反应的PCR反应,并扩增所得cDNA。巢式PCR包括两阶段PCR,其中使用第一组引物的第一PCR反应的扩增子变为使用第二引物组的第二PCR反应的样品,所述第二引物组的至少一个引物与第一PCR反应的扩增子的内部位置结合。多重PCR包括其中多个多核苷酸序列在同一反应混合物中同时进行PCR的PCR反应。PCR反应体积可以为从0.2pL-1000μL的任何值。定量PCR包括设计用于测量样品中一个或多个序列的绝对或相对量、丰度或浓度的PCR反应。定量测量可以包括将一个或多个参考序列或标准物与感兴趣的多核苷酸序列进行比较。(Freeman等人, Biotechniques, 26:112-126 (1999); Becker-Andre等人, Nucleic Acids Research, 17:9437-9447 (1989); Zimmerman等人, Biotechniques, 21:268-279 (1996); Diviacco等人, Gene, 122:3013-3020 (1992); Becker-Andre等人, Nucleic Acids Research, 17:9437-9446 (1989))。

如本文所用,“核苷酸”、“核苷”、“核苷酸残基”和“核苷残基”可以意指脱氧核糖核苷酸或核糖核苷酸残基、或能够用作适合用于在扩增反应(例如,PCR反应)中使用的引物的组分的其他类似核苷类似物。除了在另有说明的情况下,此类核苷及其衍生物可以用作本文所

述的引物的结构单元。本申请中的任何内容都不意在排除已被化学修饰以增强其在扩增反应中的稳定性或有用性的核苷衍生物或碱基的使用,条件是所述化学修饰不干扰它们被聚合酶识别为脱氧鸟嘌呤、脱氧胞嘧啶、脱氧胸苷或脱氧腺嘌呤(视情况而定)。在一些实施方案中,核苷酸类似物可以稳定杂交体形成。在一些实施方案中,核苷酸类似物可以使杂交体形成不稳定。在一些实施方案中,核苷酸类似物可以增强杂交特异性。在一些实施方案中,核苷酸类似物可以降低杂交特异性。

“核酸”或语法等效物是指单个核苷酸或共价连接在一起的至少两个核苷酸。

“多核苷酸”或语法等效物是指共价连接在一起的至少两个核苷酸。多核苷酸包含含有两个或更多个核苷酸的分子。多核苷酸包含任何长度的核苷酸(核糖核苷酸、脱氧核糖核苷酸或肽核酸(PNA))的聚合形式,所述核苷酸包含嘌呤和嘧啶碱基、或其他天然的、经化学或生物化学修饰的、非天然的核苷酸碱基或核苷酸碱基的衍生物。多核苷酸的主链可以包含糖和磷酸基团、或经修饰的或经取代的糖或磷酸基团。多核苷酸可以包含经修饰的核苷酸,如甲基化的核苷酸和核苷酸类似物。核苷酸的序列可以被非核苷酸组分中断。多核苷酸可以包括其他分子,如另一种杂交的多核苷酸。多核苷酸包括脱氧核糖核酸(DNA)、核糖核酸(RNA)或两者的序列。多核苷酸的非限制性例子包括基因、基因片段、外显子、内含子、基因间DNA(包括但不限于异染色质DNA)、信使RNA(mRNA)、转移RNA、核糖体RNA、核酶、小干扰RNA(siRNA)、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、序列的分离DNA、序列的分离RNA、核酸探针和引物。多核苷酸可以从天然来源分离,可以是重组的,或者可以是人工合成的。

多核苷酸可以包括非标准核苷酸,如核苷酸类似物或经修饰的核苷酸。在一些实施方案中,非标准核苷酸可以稳定杂交体形成。在一些实施方案中,非标准核苷酸可以使杂交体形成不稳定。在一些实施方案中,非标准核苷酸可以增强杂交特异性。在一些实施方案中,非标准核苷酸可以降低杂交特异性。非标准核苷酸修饰的例子包括2'0-Me、2'0-烯丙基、2'0-炔丙基、2'0-烷基、2'氟、2'阿拉伯糖基、2'木糖基、2'氟阿拉伯糖基、硫代磷酸酯、二硫代磷酸酯、磷酸酰胺、2'氨基、5-烷基-取代的嘧啶、3'脱氧鸟苷、5-卤代-取代的嘧啶、烷基-取代的嘌呤、卤代-取代的嘌呤、二环核苷酸、2'MOE、PNA分子、LNA分子、LNA样分子、二氨基嘌呤、S2T、5-氟尿嘧啶、5-溴尿嘧啶、5-氯尿嘧啶、5-碘尿嘧啶、次黄嘌呤、黄嘌呤(xantine)、4-乙酰胞嘧啶、5-(羧基羟基甲基)尿嘧啶、5-羧基甲基氨基甲基-2-硫代尿苷、5-羧基甲基氨基甲基尿嘧啶、二氢尿嘧啶、β-D-半乳糖基辫昔(queosine)、肌昔、N6-异戊烯基腺嘌呤、1-甲基鸟嘌呤、1-甲基肌昔、2,2-二甲基鸟嘌呤、2-甲基腺嘌呤、2-甲基鸟嘌呤、3-甲基胞嘧啶、5-甲基胞嘧啶、N6-腺嘌呤、7-甲基鸟嘌呤、5-甲基氨基甲基尿嘧啶、5-甲氧基氨基甲基-2-硫代尿嘧啶、β-D-甘露糖基辫昔、5'5'-甲氧基羧基甲基尿嘧啶、5-甲氧基尿嘧啶、2-甲硫基-D46-异戊烯基腺嘌呤、尿嘧啶-5-氨基乙酸(v)、wybutoxosine、假尿嘧啶、辫昔、2-硫代胞嘧啶、5-甲基-2-硫代尿嘧啶、2-硫代尿嘧啶、4-硫代尿嘧啶、5-甲基尿嘧啶、尿嘧啶-5-氨基乙酸甲酯、尿嘧啶-5-氨基乙酸(v)、5-甲基-2-硫代尿嘧啶、3-(3-氨基-3-N-2-羧基丙基)尿嘧啶、(acp3)w、2,6-二氨基嘌呤及其衍生物。

“受试者”、“个体”、“宿主”或“患者”是指活生物,如哺乳动物。受试者和宿主的例子包括但不限于马、牛、骆驼、绵羊、猪、山羊、狗、猫、兔、豚鼠、大鼠、小鼠(例如,人源化小鼠)、沙鼠、非人灵长类动物(例如,猕猴)、人等;非哺乳动物,包括例如非哺乳动物脊椎动物,如禽类(例如,鸡或鸭)、鱼(例如,鲨鱼)或青蛙(例如,非洲爪蟾蜍),和非哺乳动物无脊椎动物;

及其转基因物种。在某些方面,受试者是指单个生物(例如,人)。在某些方面,或提供了构成具有研究的共同免疫因子和/或疾病的小群组的一组个体和/或没有疾病的个体群组(例如,阴性/正常对照)。从其获得样品的受试者可能患有疾病和/或障碍(例如,一种或多种过敏、感染、癌症或自身免疫障碍等),并且可以与不受疾病影响的阴性对照受试者进行比较。

“试剂盒”是指用于递送用于进行本文公开的方法的材料或试剂的递送系统。在一些实施方案中,试剂盒包括允许将反应试剂(例如,在适当容器中的探针、酶等)和/或支持材料(例如,缓冲液、用于进行测定的书面说明等)从一个地点储存、运输或递送到另一个地点。例如,试剂盒包括含有相关反应试剂和/或支持材料的一个或多个外壳(例如,盒子)。可以将此类内容物一起或单独递送至预期的受体。例如,第一容器可以含有用于在测定中使用的酶,而第二容器含有多个引物。

在一些方面,“多肽”是指包含至少两个氨基酸的分子。在一些实施方案中,多肽由单个肽组成。在一些实施方案中,多肽包含两个或更多个肽。例如,多肽可以包含至少约2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个或1000个肽或氨基酸。多肽的例子包括但不限于氨基酸链、蛋白质、肽、激素、多肽糖、脂质、糖脂、磷脂、抗体、酶、激酶、受体、转录因子和配体。

在一些方面,“样品”是指生物学、环境、医学、受试者或患者样品或含有多核苷酸(如靶多核苷酸)的样品。

“药学上可接受的”是指分子实体和组合物在给予人时在生理上可耐受并且通常不会产生过敏反应或类似的不良反应(如胃部不适、头晕等)。

“预防 (prevention)”是指预防 (prophylaxis, prevention) 症状发作、预防 (prevention) 与蛋白质过量水平相关或与蛋白质活性相关的疾病或障碍的进展。

“抑制”、“治疗”(“treatment”和“treating”)可互换使用,并且是指例如症状停滞、生存期延长、症状部分或完全改善、以及与蛋白质过量水平相关或与蛋白质活性相关的病症、疾病或障碍的部分或完全根除。例如,癌症的治疗包括但不限于癌性生长或肿瘤的停滞、部分或完全消除。治疗或部分消除包括例如生长或肿瘤大小和/或体积的减少倍数,如约2倍、约3倍、约4倍、约5倍、约10倍、约20倍、约50倍或之间的任何倍数减少。类似地,治疗或部分消除可以包括以下的生长或肿瘤大小和/或体积的减少百分比:约1%、2%、3%、4%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%或之间的任何百分比减少。

本申请中提及的所有出版物(包括专利文献、科学文章和数据库)出于所有目的通过引用以其整体并入,在程度上如同每个单独的出版物通过引用单独并入。如果本文所述的定义与通过引用并入本文的专利、申请、公开申请和其他出版物中阐述的定义相反或在其他方面不一致,则本文所述的定义优先于通过引用并入本文的定义。

出于说明,下面参考实施例应用来描述几个方面。应当理解,阐述了许多具体细节、关系和方法以提供对本文所述特征的全面理解。然而,相关领域的普通技术人员将容易地认识到,可以在没有一个或多个具体细节或使用其他方法的情况下实践本文所述的特征。本文所述的特征不受动作或事件的所说明顺序的限制,因为一些动作能以不同的顺序和/或与其他动作或事件同时发生。此外,根据本文所述的特征,并不需要所有所说明动作或事件来实现方法。

本文使用的章节标题仅用于组织目的,而不应解释为限制所描述的主题。

## VIII. 示例性实施方案

所提供的实施方案包括:

1. 一种产生多核苷酸文库的方法,所述方法包括在与第一衔接子相对的末端处或附近向多个条码化单链多核苷酸中的每一个添加第二衔接子,所述第一衔接子与所述条码化单链多核苷酸中的每一个附接,所述多个条码化单链多核苷酸包含:

(i) 一个或多个靶单链多核苷酸,所述一个或多个靶单链多核苷酸包含在细胞群的细胞中存在的一一个或多个靶多核苷酸或其一个或多个互补体的扩增子;以及

(ii) 单链多核苷酸的集合,所述单链多核苷酸各自包含所述细胞中的多核苷酸或其互补体的扩增子;并且

其中所述多个条码化单链多核苷酸中的每一个包含容器条码,所述容器条码对于来自所述细胞群的同一细胞的来自(i)和(ii)的所有多核苷酸而言是相同的。

2. 根据实施方案1所述的方法,其中所述多个条码化单链多核苷酸包含所述细胞群中的多个细胞的(i)和(ii)的多核苷酸。

3. 根据实施方案1或实施方案2所述的方法,其中所述多个条码化单链多核苷酸中的每一个还包含对于每个单链多核苷酸而言独特的分子条码。

4. 根据实施方案1-3中任一项所述的方法,其中来自所述细胞群的每个细胞的单链多核苷酸的所述集合共同包含转录组或部分转录组的互补DNA(cDNA)链。

5. 根据实施方案1-4中任一项所述的方法,其中所述转录组或所述部分转录组共同包含在所述细胞的基因组中存在的至少60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的转录物。

6. 根据实施方案1-5中任一项所述的方法,其中所述条码化单链多核苷酸中的每一个具有大于或大于约50个碱基对、大于100个碱基对或大于200个碱基对的大小。

7. 根据实施方案1-6中任一项所述的方法,其中所述条码化单链多核苷酸中的每一个具有从或从约50个碱基对(bp)至1500bp、50bp至1250bp、50bp至1000bp、50bp至750bp、50bp至500bp、100bp至1500bp、100bp至1250bp、100bp至1000bp、100bp至750bp、100bp至500bp、200bp至1500bp、200bp至1250bp、200bp至1000bp、200bp至750bp或250bp至500bp的大小。

8. 根据实施方案1-7中任一项所述的方法,其中添加所述第二衔接子在包含所述多个条码化单链多核苷酸的均质混合物中进行。

9. 根据实施方案1-8中任一项所述的方法,其中所述第一衔接子包含所述容器条码。

10. 一种产生多核苷酸文库的方法,所述方法包括:

(a) 裂解多个容器中的每一个内的细胞,其中所述容器中的每一个包含来自包含细胞群的样品的细胞;

(b) 在每个容器中产生多个互补多核苷酸,所述多个互补多核苷酸的所述产生包括(i)使用一个或多个靶特异性引物产生与所述细胞中存在的一个或多个靶多核苷酸互补的一个或多个靶多核苷酸;并且(ii)使用随机寡聚引物产生多核苷酸的集合,所述多核苷酸中的每一个与所述细胞中的多核苷酸互补。

11. 根据实施方案10所述的方法,其中所述容器中的每一个还包含多个分子条码化寡核苷酸、一个容器条码化寡核苷酸或容器条码化寡核苷酸池以及任选地第一衔接子,并且

所述方法还包括：

(c) 将所述多个分子条码化寡核苷酸之一附接至多个互补多核苷酸、任选地所述多个互补多核苷酸中的每一个,从而产生各自包含分子条码的多个分子条码化多核苷酸,任选地其中所述分子条码化多核苷酸中的每一个的所述分子条码与在所述多个分子条码化多核苷酸中的其他分子条码化多核苷酸所包含的分子条码不同和/或是独特的分子条码;

(d) 将所述一个容器条码化寡核苷酸或容器条码化寡核苷酸池之一和所述第一衔接子或其扩增产物附接至所述条码化多核苷酸中的多个、任选地所述条码化多核苷酸中的每一个,从而产生多个双条码化多核苷酸,其中在同一容器中的所述双条码化多核苷酸中的每一个包含相同的容器条码。

12. 根据实施方案11所述的方法,其还包括(e)产生所述多个双条码化多核苷酸中的多个、任选地所述多个双条码化多核苷酸中的每一个的单链扩增子。

13. 根据实施方案11所述的方法,其还包括(f)向所述单链扩增子中的每一个添加第二衔接子,其中所述第一衔接子和所述第二衔接子存在于所述双条码化单链多核苷酸中的每一个的相对末端处或附近。

14. 一种产生多核苷酸文库的方法,所述方法包括:

(a) 裂解多个容器中的每一个内的细胞,其中所述容器中的每一个包含来自包含细胞群的样品的细胞、多个分子条码化寡核苷酸和一个容器条码化寡核苷酸或容器条码化寡核苷酸池以及任选地第一衔接子;

(b) 在每个容器中产生多个互补多核苷酸,所述多个互补多核苷酸的所述产生包括(i)产生与所述细胞中存在的一个或多个靶多核苷酸互补的一个或多个靶多核苷酸;并且(ii)产生多核苷酸的集合,所述多核苷酸各自单独与所述细胞中的多核苷酸互补;

(c) 将所述多个分子条码化寡核苷酸之一附接至每个互补多核苷酸,从而产生各自包含独特分子条码的多个条码化多核苷酸;

(d) 将所述一个容器条码化寡核苷酸或容器条码化寡核苷酸池之一或其扩增产物附接至所述条码化多核苷酸中的每一个,从而产生多个双条码化多核苷酸,其中在同一容器中的所述双条码化多核苷酸中的每一个包含相同的容器条码;

(e) 产生所述多个双条码化多核苷酸中的每一个的单链扩增子;以及

(f) 向所述单链扩增子中的每一个添加第二衔接子,从而将所述第二衔接子添加到双条码化单链多核苷酸中,其中所述第一衔接子和所述第二衔接子存在于所述双条码化单链多核苷酸中的每一个的相对末端处或附近。

15. 根据实施方案11-14中任一项所述的方法,其中所述第一衔接子包含所述容器条码化寡核苷酸。

16. 根据实施方案10-15中任一项所述的方法,其中来自所述细胞群的每个细胞的多核苷酸的所述集合共同包含与细胞的转录组或部分转录组的转录物互补的序列。

17. 根据实施方案16中任一项所述的方法,其中所述转录组或所述部分转录组包含在所述细胞的基因组中存在的至少60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的转录物。

18. 根据实施方案1-17中任一项所述的方法,其中在所述细胞中的所述一个或多个靶多核苷酸和/或所述多核苷酸是DNA。

19. 根据实施方案1-18中任一项所述的方法,其中在所述细胞中的所述一个或多个靶多核苷酸和/或所述多核苷酸是RNA。
20. 根据实施方案19所述的方法,其中所述RNA是mRNA。
21. 根据实施方案10-20中任一项所述的方法,其中(b)的所述互补多核苷酸中的每一个或者一个或多个是cDNA。
22. 根据实施方案1-21中任一项所述的方法,其中所述一个或多个条码化单链多核苷酸中的每一个或者一个或多个是cDNA的链。
23. 根据实施方案1-22中任一项所述的方法,其中所述第一衔接子和/或所述第二衔接子包含至少一个通用引发位点。
24. 根据实施方案1-23中任一项所述的方法,其中:  
所述第一衔接子和所述第二衔接子是不同的;和/或  
所述第一衔接子包含第一通用引发位点,并且所述第二衔接子包含第二通用引发位点,任选地其中所述第一通用引发位点和所述第二通用引发位点是不同的。
25. 根据实施方案24所述的方法,其中所述第一通用引发位点和/或所述第二通用引发位点是或包含P7引发位点(C7)或其连续部分或者P5引发位点(C5)或其连续部分,任选地其中所述其连续部分足够与互补序列退火。
26. 根据实施方案24或实施方案25所述的方法,其中所述第一通用引发位点是或包含所述P7引发位点(C7)或其连续部分,并且所述第二通用引发位点是或包含所述P5引发位点(C5)或其连续部分。
27. 根据实施方案25或实施方案26所述的方法,其中所述P7引发位点(C7)包含序列AGATCGGAAGAGCACACGTCTGAACCTCCA (SEQ ID NO:77),或者是其连续部分。
28. 根据实施方案25或实施方案26所述的方法,其中所述P5引发位点包含序列AGATCGGAAGAGCGTCGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT (SEQ ID NO:78),或者是其连续部分。
29. 根据实施方案25-28中任一项所述的方法,其中所述连续部分在长度上包含至少或至少约15个、20个、25个或30个核苷酸。
30. 根据实施方案25、26、28或29中任一项所述的方法,其中所述P5引发位点是如SEQ ID NO:25 (AGATCGGAAGAGCGTCGTGT) 所示的连续部分。
31. 根据实施方案1-9和13-30中任一项所述的方法,其中添加所述第二衔接子包括在包含所述第二通用引发位点的寡核苷酸的存在下使夹板寡核苷酸与所述条码化单链多核苷酸中的每一个杂交,其中所述夹板寡核苷酸包含(i)与所述第二通用引发位点互补的序列和(ii)能够与所述条码化单链多核苷酸的3'末端随机退火的简并悬突序列。
32. 根据实施方案31所述的方法,其中在所述杂交之前,使所述夹板寡核苷酸和包含所述第二通用引发位点的寡核苷酸退火以形成夹板-衔接子双链体。
33. 根据实施方案31或实施方案32所述的方法,其中所述简并悬突序列包含序列(N)<sub>3-12</sub>,其中N是任何核苷酸。
34. 根据实施方案31-33中任一项所述的方法,其中所述简并悬突序列包含序列NNNNNN,其中N是任何核苷酸(SEQ ID NO:24)。
35. 根据实施方案31-34中任一项所述的方法,其中所述夹板寡核苷酸包含序列

ACACGACGCTTCCGATCTNNNNNN, 其中N是任何氨基酸 (SEQ ID NO:26)。

36. 根据实施方案31-35中任一项所述的方法, 其中包含所述第二通用引发位点的寡核苷酸包含序列AGATCGGAAGAGCGTCGTGT (SEQ ID NO:25)。

37. 根据实施方案1-9和11-36中任一项所述的方法, 其中所述容器条码化寡核苷酸包含至少或约至少3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、30个、40个或50个核苷酸。

38. 根据实施方案1-9和11-37中任一项所述的方法, 其中所述容器条码化寡核苷酸包含从或从约10个至30个核苷酸。

39. 根据实施方案1-38中任一项所述的方法, 其中所述容器条码化寡核苷酸包含简并序列。

40. 根据实施方案1-9和11-39中任一项所述的方法, 其中所述容器条码化寡核苷酸包含序列 (N)<sub>14-17</sub>, 其中N是任何核苷酸, 任选地其中所述序列中的至少一个或两个N是W, 其中W是腺嘌呤或胸腺嘧啶。

41. 根据实施方案1-9和11-40中任一项所述的方法, 其中所述容器条码化寡核苷酸包含序列NNNNWNNNNWNNNN (SEQ ID NO:80)、WNNNNWNNNNWNNNN (SEQ ID NO:81)、NWNNNWNNNNWNNNN (SEQ ID NO:82) 或NNWNNNWNNNNWNNNN (SEQ ID NO:83), 其中N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶。

42. 根据实施方案11-41中任一项所述的方法, 其中每个容器包含第一衔接子池, 其中与所述池中的其他容器条码化寡核苷酸中的至少一个相比, 所述第一衔接子池的每个容器条码化寡核苷酸包含至少一个碱基位移或碱基添加。

43. 根据实施方案42所述的方法, 其中所述第一衔接子池的所述容器条码化寡核苷酸包含序列NNNNWNNNNWNNNN (SEQ ID NO:80)、WNNNNWNNNNWNNNN (SEQ ID NO:81)、NWNNNWNNNNWNNNN (SEQ ID NO:82) 和NNWNNNWNNNNWNNNN (SEQ ID NO:83), 其中N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶。

44. 根据实施方案11-43中任一项所述的方法, 其中在步骤(d)中, 所述方法还包括扩增所述一个容器条码化寡核苷酸或所述容器条码化寡核苷酸池或者包含所述一个容器条码化寡核苷酸或所述容器条码化寡核苷酸池的第一衔接子, 其中所述扩增是在附接所述容器条码化寡核苷酸之前或与之同时进行的。

45. 根据实施方案11-44中任一项所述的方法, 其中附接所述容器条码化寡核苷酸包括使所述容器条码化寡核苷酸的区域与所述互补多核苷酸中的每一个的区域或与包含分子条码的所述分子条码化多核苷酸中的每一个的区域杂交。

46. 根据实施方案45所述的方法, 其中所述区域包含与所述分子条码化多核苷酸的分子条码的5'末端区域互补的3'标记多核苷酸。

47. 根据实施方案10-46中任一项所述的方法, 其中在步骤(b)中:

所述一个或多个靶多核苷酸通过在逆转录酶和一个或多个靶特异性引物的存在下将所述一个或多个靶多核苷酸逆转录来产生, 所述一个或多个靶特异性引物与所述一个或多个靶多核苷酸的靶序列互补; 和/或

多核苷酸的所述集合通过在逆转录酶和一个或多个转录组引物的存在下将所述细胞中的多核苷酸逆转录来产生, 所述一个或多个转录组引物与所述细胞中的多核苷酸互补。

48. 根据实施方案1-47中任一项所述的方法,其中所述一个或多个靶多核苷酸包含免疫分子或其链的多核苷酸。

49. 根据实施方案1-48中任一项所述的方法,其中所述一个或多个靶多核苷酸包含至少两个靶多核苷酸,所述至少两个靶多核苷酸各自包含免疫分子链的多核苷酸。

50. 根据实施方案1-49中任一项所述的方法,其中所述一个或多个靶多核苷酸包含TCR或其链的多核苷酸。

51. 根据实施方案1-50中任一项所述的方法,其中所述一个或多个靶多核苷酸包含T细胞受体 $\alpha$  (TCR $\alpha$ ) 的第一多核苷酸和T细胞受体 (TCR $\beta$ ) 的第二多核苷酸。

52. 根据实施方案1-50中任一项所述的方法,其中所述一个或多个靶多核苷酸包含T细胞受体 $\gamma$  (TCR $\gamma$ ) 的第一多核苷酸和T细胞受体 $\delta$  (TCR $\delta$ ) 的第二多核苷酸。

53. 根据实施方案1-49中任一项所述的方法,其中所述一个或多个靶多核苷酸包含抗体或其链的多核苷酸。

54. 根据实施方案1-49和53中任一项所述的方法,其中所述一个或多个靶多核苷酸包含重链免疫球蛋白 (IgH) 多核苷酸的第一多核苷酸和轻链免疫球蛋白 (IgL) 多核苷酸的第二多核苷酸。

55. 根据实施方案47-54中任一项所述的方法,其中所述一个或多个靶特异性引物和/或所述一个或多个转录组引物包含聚 (T) 序列。

56. 根据实施方案47-54中任一项所述的方法,其中所述一个或多个转录组引物包含随机六聚体寡核苷酸引物的混合物。

57. 根据实施方案47-56中任一项所述的方法,其中所述一个或多个靶特异性引物包含与所述靶多核苷酸的一个或多个靶序列的一个或多个序列互补的一个或多个引物。

58. 根据实施方案57所述的方法,其中所述一个或多个靶特异性引物包含至少第一引物和第二引物。

59. 根据实施方案57或实施方案58所述的方法,其中所述一个或多个靶特异性引物包含针对多个免疫分子或其链的靶序列的引物。

60. 根据实施方案59所述的方法,其中所述免疫分子是T细胞受体或抗体。

61. 根据实施方案58-60中任一项所述的方法,其中至少所述第一引物与免疫分子的第一链的多核苷酸的靶序列互补,并且第二引物与所述免疫分子的第二链的多核苷酸的靶序列互补。

62. 根据实施方案58-61中任一项所述的方法,其中所述第一引物和所述第二引物与TCR的不同TCR链多核苷酸的靶序列互补。

63. 根据实施方案58-62中任一项所述的方法,其中:

所述第一引物与TCR $\alpha$ 多核苷酸序列的靶序列互补,并且所述第二引物与TCR $\beta$ 多核苷酸序列的靶序列互补;或者

所述第一引物与TCR $\gamma$ 多核苷酸序列的靶序列互补,并且所述第二引物与TCR $\delta$ 多核苷酸序列的靶序列互补。

64. 根据实施方案62或实施方案63所述的方法,其中所述TCR链多核苷酸的靶序列是恒定区序列。

65. 根据实施方案58-64中任一项所述的方法,其中:

所述第一引物与TCR $\alpha$ 恒定区多核苷酸序列的靶序列互补，并且所述第二引物与TCR $\beta$ 恒定区多核苷酸序列的靶序列互补；或者

所述第一引物与TCR $\gamma$ 恒定区多核苷酸序列的靶序列互补，并且所述第二引物与TCR $\delta$ 恒定区多核苷酸序列的靶序列互补。

66. 根据实施方案58-61中任一项所述的方法，其中至少所述第一引物和所述第二引物与抗体的不同抗体链多核苷酸的靶序列互补。

67. 根据实施方案58-61和66中任一项所述的方法，其中所述第一引物与重链免疫球蛋白(IgH)多核苷酸序列的靶序列互补，并且所述第二引物与轻链免疫球蛋白(IgL)多核苷酸序列的靶序列互补。

68. 根据实施方案66或实施方案67所述的方法，其中所述抗体链多核苷酸的靶序列是恒定区序列。

69. 根据实施方案58-61和66-68中任一项所述的方法，其中所述第一引物与重链恒定区(CH)多核苷酸序列的靶序列互补，并且所述第二引物与轻链恒定区(CL)多核苷酸序列的靶序列互补。

70. 根据实施方案68或实施方案69所述的方法，其中：

所述CH多核苷酸的靶序列来自IgM、IgD、IgA、IgE或IgG或其组合；和/或

所述CL多核苷酸序列的靶序列来自Ig $\kappa$ 、Ig $\lambda$ 或其组合。

71. 根据实施方案1-70中任一项所述的方法，其中所述一个或多个靶多核苷酸包含全长编码序列。

72. 根据实施方案10-71中任一项所述的方法，其中所述一个或多个靶多核苷酸和多核苷酸的所述集合在所述容器中以相同的反应体积产生。

73. 根据实施方案10-72中任一项所述的方法，其中，在步骤(b)中，产生所述多个互补多核苷酸包括使用非模板末端转移酶，其中将三个或更多个非模板核苷酸、核糖核苷酸或其类似物添加到每个产生的互补多核苷酸的3'末端。

74. 根据实施方案73所述的方法，其中所述非模板末端转移酶是逆转录酶或聚合酶。

75. 根据实施方案73或实施方案74所述的方法，其中所述非模板末端转移酶是逆转录酶，并且其中所述逆转录酶选自Superscript II逆转录酶、Maxima逆转录酶、Protoscript II逆转录酶、莫洛尼鼠白血病病毒逆转录酶(MMLV-RT)、HighScriber逆转录酶、禽成髓细胞瘤病毒(AMV)逆转录酶、包含末端脱氧核苷酸转移酶活性的任何逆转录酶及其组合。

76. 根据实施方案11-75中任一项所述的方法，其中，在步骤(c)中，所述附接包括使所述分子条码化寡核苷酸之一的区域与所述互补多核苷酸中的每一个的所述三个或更多个非模板核苷酸杂交。

77. 根据实施方案73-75中任一项所述的方法，其中提供所述多个分子条码化寡核苷酸作为多个模板转换寡核苷酸，所述多个模板转换寡核苷酸各自包含与所述三个或更多个非模板核苷酸互补的3'部分。

78. 根据实施方案77所述的方法，其中所述模板转换寡核苷酸还包含与包含所述容器条码的第一衔接子的3'标记寡核苷酸互补的5'末端区域。

79. 根据实施方案11-78中任一项所述的方法，其中：

所述逆转录酶具有模板转换活性；

所述多个产生的互补多核苷酸的至少一些链包含含有三个或更多个非模板核苷酸的3'悬突；

提供所述多个分子条码化寡核苷酸作为多个模板转换寡核苷酸，所述多个模板转换寡核苷酸各自包含(1)与包含所述容器条码的第一衔接子的3'标记寡核苷酸互补的5'末端区域、(2)所述分子条码和(3)与所述3'悬突的三个或更多个非模板核苷酸互补的3'部分；并且

所述模板转换寡核苷酸用作所述逆转录酶的模板，使得将所述分子条码掺入每个互补多核苷酸中。

80. 根据实施方案77-79中任一项所述的方法，其中与所述三个或更多个非模板核苷酸互补的3'部分包含核苷酸、核糖核苷酸或其类似物。

81. 根据实施方案73-80中任一项所述的方法，其中所述三个或更多个非模板核苷酸包含三个或更多个C核苷酸，并且与三个或更多个非模板核苷酸互补的所述3'部分包含一个或多个G核苷酸或核糖核苷酸或其类似物。

82. 根据实施方案73-77中任一项所述的方法，其中所述模板转换寡核苷酸还包含阻断通过逆转录酶或DNA聚合酶进行延伸的3'修饰的核苷酸。

83. 根据实施方案82所述的方法，其中所述修饰是3'核苷酸的脱氧、磷酸、氨基或烷基修饰。

84. 根据实施方案11-83中任一项所述的方法，其中步骤(d)还包括在所述附接之后延伸所述多个互补分子条码化多核苷酸中的每一个。

85. 根据实施方案1-84中任一项所述的方法，其中所述容器是孔、乳剂或微滴。

86. 根据实施方案12-85中任一项所述的方法，其包括在步骤(e)之前组合所述多个容器中的两个或更多个的内容物，从而产生均质混合物，所述均质混合物包含所述多个双条码化单链多核苷酸中的两个或更多个。

87. 根据实施方案86所述的方法，其中组合所述多个容器的内容物包括破坏所述多个容器中的两个或更多个，并且将来自所述两个或更多个破坏的容器的双条码化单链多核苷酸合并。

88. 根据实施方案86或实施方案87所述的方法，其包括在步骤(e)之前选择或纯化具有大于或大于约50个碱基对、大于100个碱基对或大于200个碱基对的大小的双条码化单链多核苷酸。

89. 根据实施方案86-88中任一项所述的方法，其包括在步骤(e)之前选择或纯化具有以下大小的双条码化单链多核苷酸：从或从约50个碱基对(bp)至1500bp、50bp至1250bp、50bp至1000bp、50bp至750bp、50bp至500bp、100bp至1500bp、100bp至1250bp、100bp至1000bp、100bp至750bp、100bp至500bp、200bp至1500bp、200bp至1250bp、200bp至1000bp、200bp至750bp或250bp至500bp。

90. 根据实施方案1-89中任一项所述的方法，其中所述双条码化单链多核苷酸按顺序(5'至3')包含：所述第一衔接子、所述容器条码、所述分子条码和所述第二衔接子。

91. 根据实施方案1-90中任一项所述的方法，其中所述第一衔接子位于所述双条码化单链多核苷酸的5'区域处或附近。

92. 根据实施方案1-91中任一项所述的方法，其中所述第二衔接子位于所述双条码化

单链多核苷酸的3'区域处或附近。

93. 根据实施方案10-92中任一项所述的方法,其中步骤(a)-(f)中的一个或多个是在溶液中进行和/或不在固体支持物、任选地珠的存在下进行。

94. 根据实施方案11-93中任一项所述的方法,其中至少步骤(c)和(d)是在溶液中进行和/或不在固体支持物、任选地珠的存在下进行。

95. 根据实施方案10-94中任一项所述的方法,其中步骤(a)-(e)中的每一个是在溶液中进行和/或不在固体支持物、任选地珠的存在下进行。

96. 根据实施方案1-95中任一项所述的方法,其中所述细胞群包含至少或约至少 $1 \times 10^3$ 个、 $5 \times 10^3$ 个、 $1 \times 10^4$ 个、 $5 \times 10^4$ 个、 $1 \times 10^5$ 个、 $5 \times 10^5$ 个、 $1 \times 10^6$ 个或 $5 \times 10^6$ 个细胞。

97. 根据实施方案1-96中任一项所述的方法,其中所述细胞群来自受试者的生物样品。

98. 根据实施方案97所述的方法,其中所述生物样品是或包含全血样品、血沉棕黄层样品、外周血单核细胞(PBMC)样品、未分级T细胞样品、淋巴细胞样品、白细胞样品、单采术产物或白细胞单采术产物。

99. 根据实施方案1-98中任一项所述的方法,其中所述细胞群包含免疫细胞。

100. 根据实施方案1-99中任一项所述的方法,其中所述免疫细胞包含淋巴细胞或抗原呈递细胞。

101. 根据实施方案1-100中任一项所述的方法,其中所述免疫细胞是淋巴细胞或其亚型、B细胞或其亚型、T细胞或其亚型、或其组合。

102. 根据实施方案101所述的方法,其中所述免疫细胞是T细胞,所述T细胞是CD4+和/或CD8+T细胞。

103. 根据实施方案1-102中任一项所述的方法,其中所述细胞群富集或包含中枢记忆T细胞、效应记忆T细胞、幼稚T细胞、干细胞样中枢记忆T细胞、效应T细胞和调节性T细胞。

104. 根据实施方案1-101中任一项所述的方法,其中所述细胞群富集记忆B细胞、幼稚B细胞或浆母细胞B细胞。

105. 根据实施方案97-104中任一项所述的方法,其中所述受试者是人类受试者。

106. 根据实施方案97-105中任一项所述的方法,其中所述受试者患有癌症、感染或自身免疫病症。

107. 根据实施方案106所述的方法,其中所述感染是病毒、细菌或真菌感染。

108. 根据实施方案1-107中任一项所述的方法,其还包括扩增所述多个条码化单链多核苷酸,从而产生多个多核苷酸模板。

109. 根据实施方案1-108中任一项所述的方法,其中扩增是在第一引物组的存在下进行,所述第一引物组包含与所述第一衔接子序列互补的第一引物和与所述第二衔接子序列互补的第二引物。

110. 根据实施方案109所述的方法,其中所述第一引物和/或所述第二引物是通用引物。

111. 根据实施方案110所述的方法,其中所述第一引物和/或所述第二引物与所述P7引发位点(C7)或其连续部分或者所述P5引发位点(C5)或其连续部分互补。

112. 根据实施方案110或实施方案111所述的方法,其中所述第一引物与所述P7引发位

点 (C7) 或其连续部分互补, 并且所述第二引物与所述P5引发位点 (C5) 或其连续部分互补。

113. 根据实施方案111或实施方案112所述的方法, 其中:

与所述P7引发位点 (C7) 或其连续部分互补的引物具有或包含序列 CAAGCAGAAGACGGCATACGAGAT (SEQ ID NO:39); 和/或

与所述P5引发位点 (C5) 或其连续部分互补的引物包含序列 ACACGACGCTCTCCGATCT (SEQ ID NO:27)。

114. 根据实施方案109-113中任一项所述的方法, 其中所述第一引物和/或所述第二引物还包含测序衔接子。

115. 根据实施方案114所述的方法, 其中:

与所述P7引发位点 (C7) 或其连续部分互补的引物还包含序列 CAAGCAGAAGACGGCATACGAGAT [NNNNNN] GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT (SEQ ID NO:28); 和/或

与所述P5引发位点 (C5) 或其连续部分互补的引物包含序列 AATGATAACGGCGACCACCGAGA TCTACACTTTCCCTACACGACGCTCTTCC (SEQ ID NO:76)。

116. 根据实施方案115所述的方法, 其还包括纯化所述多个条码化多核苷酸中的每一个。

117. 一种多核苷酸文库, 其包含通过根据实施方案1-116中任一项所述的方法产生的多个条码化多核苷酸。

118. 一种多核苷酸文库, 其包含多个条码化多核苷酸, 其中所述多个条码化多核苷酸包含 (i) 一个或多个靶多核苷酸, 所述一个或多个靶多核苷酸包含在细胞群的细胞中存在的一个或多个靶多核苷酸的扩增子; 以及 (ii) 多核苷酸的集合, 所述多核苷酸各自包含所述细胞中的多核苷酸的扩增子, 其中每个条码化多核苷酸包含:

包含与第一通用引物互补的第一通用引发位点的第一衔接子;

对于来自所述细胞群的同一细胞的来自 (i) 和 (ii) 的所有条码化多核苷酸而言相同的容器条码; 以及

包含与第二通用引物互补的第二通用引发位点的第二衔接子序列。

119. 根据实施方案118所述的多核苷酸文库, 其中所述多个条码化多核苷酸模板中的每一个包含分子条码, 所述分子条码对于每个多核苷酸模板而言是独特的。

120. 根据实施方案118或实施方案119所述的多核苷酸文库, 其中来自所述细胞群的每个细胞的条码化多核苷酸模板的所述集合共同包含转录组或部分转录组的互补DNA (cDNA) 链。

121. 根据实施方案118-120中任一项所述的多核苷酸文库, 其中所述转录组或所述部分转录组共同包含在所述细胞的基因组中存在的至少 60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99% 或 100% 的转录物。

122. 根据实施方案118-121中任一项所述的多核苷酸文库, 其中所述条码化多核苷酸模板中的每一个具有大于或大于约 50 个碱基对、大于 100 个碱基对或大于 200 个碱基对的大小。

123. 根据实施方案118-122中任一项所述的多核苷酸文库, 其中所述条码化单链多核苷酸中的每一个具有从或从约 50 个碱基对 (bp) 至 1500 bp、50 bp 至 1250 bp、50 bp 至 1000 bp、

50bp至750bp、50bp至500bp、100bp至1500bp、100bp至1250bp、100bp至1000bp、100bp至750bp、100bp至500bp、200bp至1500bp、200bp至1250bp、200bp至1000bp、200bp至750bp或250bp至500bp的大小。

124. 根据实施方案118-123中任一项所述的多核苷酸文库,其中所述第一衔接子包含所述容器条码。

125. 根据实施方案118-124中任一项所述的多核苷酸文库,其中所述第一衔接子和所述第二衔接子是不同的。

126. 根据实施方案118-125中任一项所述的多核苷酸文库,其中所述第一通用引发位点和/或所述第二通用引发位点是或包含P7引发位点(C7)或其连续部分或者P5引发位点(C5)或其连续部分,任选地其中所述其连续部分足够与互补序列退火。

127. 根据实施方案118-126中任一项所述的多核苷酸文库,其中所述第一通用引发位点是或包含所述P7引发位点(C7)或其连续部分,并且所述第二通用引发位点是或包含所述P5引发位点(C5)或其连续部分。

128. 根据实施方案126或实施方案127所述的多核苷酸文库,其中所述P7引发位点(C7)包含序列AGATCGGAAGAGCACACGTCTGAACCTCCA (SEQID NO:77),或者是其连续部分。

129. 根据实施方案126或实施方案127所述的多核苷酸文库,其中所述P5引发位点包含序列AGATCGGAAGAGCGTCGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT (SEQ ID NO:78),或者是其连续部分。

130. 根据实施方案126-129中任一项所述的多核苷酸文库,其中所述连续部分在长度上包含至少或至少约15个、20个、25个或30个核苷酸。

131. 根据实施方案129或实施方案130所述的多核苷酸文库,其中所述P5引发位点是如SEQ ID NO:25 (AGATCGGAAGAGCGTCGTGT) 所示的连续部分。

132. 根据实施方案118-131中任一项所述的多核苷酸文库,其中所述容器条码化寡核苷酸包含至少或约至少3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、30个、40个或50个核苷酸。

133. 根据实施方案118-132中任一项所述的多核苷酸文库,其中所述容器条码化寡核苷酸包含从或从约10个至30个核苷酸。

134. 根据实施方案118-133中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含免疫分子或其链的多核苷酸。

135. 根据实施方案118-134中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含至少两个靶多核苷酸,所述至少两个靶多核苷酸各自包含免疫分子链的多核苷酸。

136. 根据实施方案118-135中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含TCR或其链的多核苷酸。

137. 根据实施方案118-136中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含T细胞受体 $\alpha$  (TCR $\alpha$ ) 的第一多核苷酸和T细胞受体 (TCR $\beta$ ) 的第二多核苷酸。

138. 根据实施方案118-136中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含T细胞受体 $\gamma$  (TCR $\gamma$ ) 的第一多核苷酸和T细胞受体 $\delta$  (TCR $\delta$ ) 的第二多核苷酸。

139. 根据实施方案118-135中任一项所述的多核苷酸文库,其中所述一个或多个靶多

核苷酸包含抗体或其链的多核苷酸。

140. 根据实施方案118-135和139中任一项所述的多核苷酸文库,其中所述一个或多个靶多核苷酸包含重链免疫球蛋白(IgH)多核苷酸的第一多核苷酸和轻链免疫球蛋白(IgL)多核苷酸的第二多核苷酸。

141. 根据实施方案118-140中任一项所述的多核苷酸文库,其中所述条码化多核苷酸按顺序(5'至3')包含:所述第一衔接子、所述容器条码、所述分子条码和所述第二衔接子。

142. 根据实施方案118-1141中任一项所述的多核苷酸文库,其中所述第一衔接子位于所述双条码化单链多核苷酸的5'区域处或附近。

143. 根据实施方案118-137中任一项所述的多核苷酸文库,其中所述第二衔接子位于所述双条码化单链多核苷酸的3'区域处或附近。

144. 一种用于测序的方法,其包括对通过实施方案1-116中任一项产生或来自根据实施方案118-141中任一项所述的多核苷酸文库的所述多个条码化多核苷酸中的一个或多个进行测序。

145. 根据实施方案144所述的方法,其中对来自所述多个条码化多核苷酸的转录组进行测序。

146. 根据实施方案145所述的方法,其还包括在所述测序之前扩增全转录组或其部分。

147. 根据实施方案146所述的方法,其中使用第一引物组进行扩增,所述第一引物组包含分别对所述第一衔接子序列和所述第二衔接子序列具有特异性的第一引物和第二引物。

148. 根据实施方案146或实施方案147所述的方法,其中对来自所述多个条码化多核苷酸的一个或多个靶多核苷酸进行测序。

149. 根据实施方案148所述的方法,其还包括在所述测序之前扩增来自所述多个核苷酸模板的一个或多个靶多核苷酸。

150. 根据实施方案149所述的方法,其中扩增所述一个或多个靶多核苷酸的一个或多个全长序列。

151. 根据实施方案149或实施方案150所述的方法,其中在第二引物组的存在下进行扩增,所述第二引物组包含与一个或多个靶多核苷酸互补的一个或多个第一引物和与所述第一衔接子序列互补的第二引物。

152. 根据实施方案151所述的方法,其中所述第二引物组的第二引物与所述P7引发位点(C7)或其连续部分或者所述P5引发位点(C5)或其连续部分互补。

153. 根据实施方案151或实施方案152所述的方法,其中所述第二引物组的第二引物与所述P7引发位点(C7)或其连续部分互补。

154. 根据实施方案151-153中任一项所述的方法,其中所述第二引物组的第二引物具有或包含序列CAAGCAGAAGACGGCATACGAGAT (SEQ ID NO:39) 或CAAGCAGAAGACGGCATACGAGAT [NNNNNN] GTGACTGGAGTTCAGACGTGTGCTTCCGATCT (SEQ ID NO:28)。

155. 根据实施方案151-154中任一项所述的方法,其中与所述一个或多个靶多核苷酸互补的一个或多个第一引物对免疫分子或其链的靶序列具有特异性。

156. 根据实施方案155所述的方法,其中所述免疫分子是T细胞受体或抗体。

157. 根据实施方案155或实施方案156所述的方法,其中所述一个或多个第一引物对所述免疫分子的恒定区的靶序列具有特异性。

158. 根据实施方案155-157中任一项所述的方法,其中所述免疫分子是TCR,并且所述一个或多个第一引物包含AGTCTCTCAGCTGGTACACGG (SEQ ID NO:37)、ATGGCTCAAACACAGCGACCTC (SEQ ID NO:38) 或其组合。

159. 根据实施方案155-158中任一项所述的方法,其中所述免疫分子是抗体,并且所述一个或多个第一引物包含SEQ ID NO:29-36中的任何一个或其组合。

160. 根据实施方案144-159中任一项所述的方法,其包括确定所述一个或多个条码化多核苷酸的细胞起源。

161. 根据实施方案160所述的方法,其中确定细胞起源包括鉴定因为来自同一细胞而具有相同容器条码的序列信息。

162. 根据实施方案144-161中任一项所述的方法,其中所述靶多核苷酸是包含第一多核苷酸链和第二多核苷酸链的免疫分子,并且所述方法包括通过相同的容器条码的存在将所述第一多核苷酸链和所述第二多核苷酸链与同一细胞进行匹配。

163. 根据实施方案154-162中任一项所述的方法,其还包括定量或确定具有相同分子条码的多核苷酸的数量。

164. 根据实施方案154-163中任一项所述的方法,其中所述方法包括鉴定具有相同容器条码的转录组序列和靶多核苷酸序列,从而鉴定带有所述一个或多个靶多核苷酸的细胞的转录组信息。

165. 一种用于转录组分析的方法,所述方法包括:

(a) 对来自通过根据实施方案1-116中任一项所述的方法产生的所述多个条码化多核苷酸或来自根据实施方案118-141中任一项所述的多核苷酸文库的所述多个条码化多核苷酸的条码化多核苷酸的靶多核苷酸进行测序,从而产生来自所述多个细胞的靶多核苷酸的序列信息;

(b) 对来自通过根据实施方案1-116中任一项所述的方法产生的所述多个条码化多核苷酸或来自根据实施方案118-141中任一项所述的多核苷酸文库的所述多个条码化多核苷酸的全转录组或其部分进行测序,从而产生来自所述多个细胞的转录组数据;以及

(c) 鉴定因为来自同一细胞而具有相同容器条码的来自 (a) 和来自 (b) 的序列信息。

166. 一种分析选定单细胞的转录组的方法,其包括:

(a) 对来自通过根据实施方案1-116中任一项所述的方法产生的所述多个条码化多核苷酸的靶多核苷酸或来自根据实施方案118-141中任一项所述的多核苷酸文库的所述多个条码化多核苷酸中的多个条码化多核苷酸的靶多核苷酸进行扩增和测序,从而产生在所述多个细胞中的至少一个中的所述靶多核苷酸中的每一个的序列信息;

(b) 鉴定与在 (a) 中测序的所述靶多核苷酸之一相关的一个或多个容器条码,从而鉴定带有所述靶多核苷酸的选定单细胞;

(c) 对来自带有所述容器条码的细胞的所述多个条码化多核苷酸的转录组或其部分进行扩增和测序,从而产生来自表达靶多肽的所述选定细胞的转录组数据。

167. 根据实施方案166所述的方法,其中使用对 (b) 中鉴定的容器条码具有特异性的引物和对所述条码化多核苷酸的第二衔接子序列具有特异性的引物对来自所述选定细胞的转录组或其部分扩增或测序。

168. 一种用于转录组分析的方法,其包括将所述转录组或其部分的序列信息与来自同

一细胞的所述一个或多个靶多核苷酸中的至少一个进行匹配,其中所述序列信息是从通过根据实施方案1-114中任一项所述的方法产生的所述多个条码化多核苷酸或从根据实施方案118-141中任一项所述的多核苷酸文库的所述多个多核苷酸模板确定,或者是从根据实施方案154-164中任一项所述的方法确定。

169. 根据实施方案168所述的方法,其中具有相同容器条码的序列因为来自同一细胞而相匹配。

170. 根据实施方案165-169中任一项所述的方法,其中所述转录组数据包含与所述细胞的功能或活性相关的参数、特征、特性或表型。

171. 根据实施方案170所述的方法,其中所述转录组数据与所述细胞的激活、耗竭或增殖活性相关。

## IX. 实施例

以下实施例被包括在内仅用于说明目的,并不旨在限制本发明的范围。

### 实施例1-对乳剂中的转录物条码化以进行单细胞多核苷酸测序

#### A. 细胞制备

从总外周血单核细胞(PBMC)获得用于进行单细胞多核苷酸测序的单细胞悬浮液。抽取大约50mL血液到具有肝素钠(BD)的Vacutainer CPT细胞制备管中,在1800×g下离心20分钟,在细胞制备缓冲液(补充有2%胎牛血清和2mM EDTA的1x PBS)中洗涤两次,使用200×g下的旋转去除血小板,并且将所得的PBMC在RPMI-1640培养基(Life Technologies)+20%胎牛血清+10%DMSO中冷冻保存在-80℃下直至需要。在乳剂产生之前,将PBMC解冻,通过在细胞缓冲液1 x杜尔贝科氏(Dulbecco's)磷酸盐缓冲盐水(PBS)中离心(200×g,持续10分钟)洗涤两次。然后将细胞稀释在细胞缓冲液中至 $3.5 \times 10^6$ 个细胞/mL的细胞浓度。然后通过20μm细胞过滤器来移液吸取悬浮液。

#### B. 乳剂中的条码化

形成含有制备的细胞和反应混合物的乳剂。将反应混合物制备为2×浓缩物,在微滴形成过程期间将所述浓缩物与细胞悬浮液以1:1的体积比混合。

#### 1. 乳剂反应混合物的制备

将含有下表E1中的试剂和寡核苷酸的乳剂反应混合物在PCR清洁罩中在室温下混合。

表E1. 乳剂反应混合物	
	试剂
Tris-Cl, pH 8.0	
MgSO <sub>4</sub>	
DTT	
各dNTP	
5'生物素寡聚dT	

模板转换寡聚物	
VB模板分子/ $\mu$ L	
VB正向引物	
VB反向引物	
蛋白酶抑制剂 (X)	
酶促RNA酶抑制剂 (U/ $\mu$ L)	
MMLV RNA酶H-逆转录酶	
DNA聚合酶	
Triton X-100 (% v/v)	
水	
寡核苷酸序列	
5'生物素寡聚dT锚定的逆 转录引物	/5BiosG//iSpl8/TTT TTT TTT TTT TTT TTT TTT TTT T V N (SEQ ID NO: 1)
容器条码模板	ATCCATCCACGACTGACGGACGTATTAAANNNWNNNNWNN NNAGATCGGAAGAGCACACGTCTGAACTCCAGTCACC (SEQ ID NO:2)
模板转换寡聚物	AATACGTCCGTCAGTCGTGGATGNNTNNANNTrGrGG (SEQ ID NO:3)
容器条码正向引物	CATCCACGACTGACGGACGTATT (SEQ ID NO:4)
容器条码反向引物	GTGACTGGAGTTCAGACGTGTGCT (SEQ ID NO:5)

/5BiosG/=5'生物素修饰

/iSpl8/=18-碳间隔子

V=A、C或G

N=任何碱基

rG=核糖鸟苷

W=A或T

## 2. 从单细胞产生双条码化转录物文库

在图1A和图1B中描绘了产生双条码化多核苷酸文库的示例性方法的概述。使用制备的细胞和反应混合物形成乳剂。乳剂产生平台包括由单个空气压缩机驱动的三个Mitos P泵 (Dolomite Microfluidics) (每个泵都带有Mitos流速传感器)，以允许将两个水相和一个亲氟连续油相计算机控制地流入亲氟包被的石英Dolomite Small 2-试剂芯片中。一个水性输入通道含有所需密度的细胞，以产生所需的每微滴细胞占有水平(在一些实施方案中，

此所需的每微滴细胞占有水平为一),而第二水性通道含有裂解和反应混合物。

使用100 $\mu$ L Hamilton Microliter注射器在两次注射(每次大约100 $\mu$ L反应混合物)中使100 $\mu$ L内径PEEK管道样品回路过载。使用100 $\mu$ L Hamilton Gastight注射器将大约110 $\mu$ L细胞悬浮液加载到约110 $\mu$ L、内径0.2mm的FEP管道回路中。将回路连接至机械旋转器,所述旋转器大约每1-2秒一次地不断反转细胞回路,以防止细胞沉降和/或聚集。伴随从芯片中的两个油性通道中同时流油,通过以相同的流速通过Dolomite 2-试剂芯片来集中射流水相而形成乳剂。外部油性通道含有在HFE7500 (Novec 7500) 氟碳油中的0.5%-5.0% (w/v) 聚乙二醇基表面活性剂。乳剂射流以恒定流速(在细胞相和反应相通道中相等)运行。通过12cm、内径0.5mm的PEEK管,通过滴入在冷却块中保持在大约0℃下的0.2mL PCR带状管(Eppendorf)中,收集乳剂芯片输出。

用毛细管微量移液器从每个管的底部除去过量的油。每个乳剂级分均用40 $\mu$ L覆盖溶液25mM Na-EDTA (pH 8.0) 轻轻覆盖。

将乳剂在热循环仪中孵育以进行转录物标记反应。简言之,在45分钟逆转录(RT)步骤期间,用聚A特异性RT引物(寡聚dT引物) (SEQ ID NO:1),基于模板转换添加含有随机化分子条码的通用衔接子序列(SEQ ID NO:3),在42℃下将RNA逆转录(参见例如图1A)。在RT后,将乳剂进行40个循环的热循环(每个循环:82℃持续10秒、65℃持续25秒)以进行容器条码模板(SEQ ID NO:2)的PCR扩增,将所述容器条码模板稀释在初始裂解和反应混合物中至30,000个拷贝(cp)/ $\mu$ L,产生在最终混合物中15,000cp/ $\mu$ L或约1/约65pL微滴的浓度(参见例如图1B)。容器条码(在本文中也称为“微滴条码”)的一个末端含有Illumina读段2(“P7”)引物位点(SEQ ID NO:77),而另一个末端与通用衔接子寡核苷酸(SEQ ID NO:4)的通用序列匹配。因此,在PCR期间,模板转换的cDNA可以退火至扩增的容器条码链,并通过重叠延伸进行剪接,以产生含有分子条码和容器条码序列的全长产物。

上述方法可以适于如实施例5中所述的添加衔接子,并且如实施例6-8中所述的用于转录组和靶特异性分析。

#### 实施例2-使用靶特异性引物对乳剂中的转录物条码化以进行单细胞多核苷酸测序

##### A. 细胞制备

抽取50mL血液到具有肝素钠(BD)的Vacutainer CPT细胞制备管中,在1800x g下离心20分钟,在细胞制备缓冲液(补充有2%胎牛血清和2mMEDTA的1x PBS)中洗涤两次,使用200x g下的旋转去除血小板,并且将所得的PBMC在RPMI-1640培养基(Life Technologies)+20%胎牛血清+10%DMSO中冷冻保存在-80℃下直至需要。在乳剂产生之前,将PBMC解冻,在细胞制备缓冲液中洗涤两次并计数。使用基于阴性选择的人B细胞富集试剂盒(Stem Cell Technologies)分离B细胞。使细胞通过20微米细胞过滤器,并在细胞制备缓冲液中稀释至6.2E+06个细胞/ml(300万B细胞实验)或3.1E+06个细胞/ml(PGT供体和卵巢肿瘤实验)。

##### B. 乳剂中的免疫受体条码化

乳剂产生平台包括由单个空气压缩机驱动的三个Mitos P泵(Dolomite Microfluidics)(每个泵都带有Mitos流速传感器),以允许将两个水相和一个亲氟连续油相计算机控制地流入亲氟包被的石英Dolomite Small 2-试剂芯片中。一个水性输入通道含有所需密度的细胞,以产生所需的每微滴细胞占有水平,而第二个水性通道含有裂解和

反应混合物,所述裂解和反应混合物由下表E2所示的反应缓冲液和寡核苷酸、5单位/ $\mu$ L基于MuMLV的逆转录酶(Thermo Scientific)和0.1单位/ $\mu$ L Herculase II PCR聚合酶组成。使用100- $\mu$ L Hamilton Microliter注射器在两次注射(每次约100 $\mu$ L LR混合物)中使100- $\mu$ L内径PEEK管道样品回路过载。使用100- $\mu$ L Hamilton Gastight注射器将约110 $\mu$ L细胞悬浮液加载到约100- $\mu$ L、内径0.2-mm的FEP管道回路中。伴随从芯片中的两个油性通道中同时流油,通过以相同的流速通过所述2-试剂芯片来集中射流水相而形成乳剂。将离开芯片出口通道的乳剂滴入在冷却块上的0.2-ml PCR带状管(Eppendorf)中,其后通过从管底部移液除去过量的油,添加40 $\mu$ L覆盖溶液(25mM Na-EDTA, pH 8.0)并将管转移到标准热循环仪中进行转录物标记反应。

表E2. 靶特异性RT引物	
IgM-RT	/生物素/TGTGAGGTGGCTGCGTACTTG (SEQ ID NO: 84)
IgG-RT	/生物素/AGGACAGCCGGGAAGGTGT (SEQ ID NO: 85)
IgD-RT	/生物素/CACGCATTTGTACTCGCCTTG (SEQ ID NO: 86)
IgA-RT	/生物素/CTGGCTRGGTGGGAAGTTCT (SEQ ID NO: 87)
IgE-RT	/生物素/GGTGGCATAGTGACCAGAGA (SEQ ID NO: 88)
IgK-RT	/生物素/TATTCAGCAGGCACACAACAGA (SEQ ID NO: 89)
IgL-RT	/生物素/AGTGTGGCCTTGTGGCTTG (SEQ ID NO: 90)
TCR-A-RT	/生物素/GGGAGATCTCTGCTTCTGATG (SEQ ID NO: 91)
TCR-B-RT	/生物素/GGTGAATAGGCAGACAGACTTG (SEQ ID NO: 92)
CD4-RT	/生物素/GGCAGTCAATCCGAACACT (SEQ ID NO: 93)
CD-8-RT	/生物素/CTACAAAGTGGGCCCTCTG (SEQ ID NO: 94)
IgA-巢式	ACACGACGCTTCCGATCTGGCTCAGCGGGAAAGACCTTG (SEQ ID NO:44)
IgE-巢式	ACACGACGCTTCCGATCTGGGAAGACGGATGGCTCTG (SEQ ID NO:48)
IgM-巢式	ACACGACGCTTCCGATCTGAGACGAGGTGGAAAAGGGTTG (SEQ ID NO:52)
IgD-巢式	ACACGACGCTTCCGATCTGGAACACATCCGGAGCCTTG (SEQ ID NO:56)

IgG-巢式	ACACGACGCTCTCCGATCTCCAGGGGGAAAGACSGATG (SEQ ID NO:40)
IgL-巢式	ACACGACGCTCTCCGATCTAGGGYGGAACAGAGTGAC (SEQ ID NO:60)
IgK-巢式	ACACGACGCTCTCCGATCTGACAGATGGTGCAGCCACAG (SEQ ID NO:64)
TRA-巢式	ACACGACGCTCTCCGATCTCACGGCAGGGTCAGGGTTC (SEQ ID NO:68)
TRB-巢式	ACACGACGCTCTCCGATCTCGACCTCGGGTGGAACAC (SEQ ID NO:72)
CD4-巢式	ACACGACGCTCTCCGATCTTGTGGCCTGCCGAGGGAGG (SEQ ID NO:95)
CD8-巢式	ACACGACGCTCTCCGATCTTGCAGATCCCAGAGGGCCA (SEQ ID NO:96)
C7-bc-P7	CAAGCAGAAGACGGCATACGAGATNNNNNNNTGACTGGAGTT CAGACGTGTGCTCTCCGATCT (SEQ ID NO:97)
C5-P5	AATGATAACGGCGACCACCGAGATCTACACTCTTCCCTACACG ACGCTCTCCGATCT (SEQ ID NO:98)

在45分钟逆转录(RT)步骤期间,使用靶特异性RT引物(表E2),如先前所述的基于模板转换添加含有随机化分子条码的通用衔接子序列,在42°C下将RNA进行逆转录(Shugay, M. 等人 *Towards error-free profiling of immune repertoires*. *Nat. Methods* 11, 653-655 (2014); Islam, S. 等人 *Highly multiplexed and strand-specific single-cell RNA 5' end sequencing*. *Nat. Protoc.* 7, 813-828 (2012)) (参见例如图1A)。在RT后,将乳剂进行40个循环的热循环(每个循环:82°C持续10秒、65°C持续25秒)以进行微滴条码模板的PCR扩增,将所述微滴条码模板稀释在初始裂解和反应混合物中至30,000cp/μL,产生在最终混合物中15,000cp/μL或约1/约65pL微滴的浓度。容器条码(微滴条码)的一个末端包含 Illumina 读段2(“P7”)引物位点(SEQ ID NO:77),而另一个末端与通用衔接子寡核苷酸(SEQ ID NO:4)的通用序列匹配(参见例如图1B)。因此,在PCR期间,模板转换的cDNA可以退火至扩增的容器条码链,并通过重叠延伸进行剪接,以产生含有靶、分子条码和容器条码序列的全长产物。

上述方法可以适于包括引物以将转录组或其部分逆转录,例如通过在逆转录阶段期间包含随机六聚体寡核苷酸。这些方法也可以适于如实施例5中所述的添加衔接子,并且如实施例6-8中所述的用于转录组和靶特异性分析。

### 实施例3-对乳剂中的靶序列和转录组的转录物条码化以进行单细胞多核苷酸测序的方法

#### A. 细胞制备

将冷冻保存的PBMC悬浮液快速解冻，并在室温下添加到10体积的RPMI+10%FBS中。通过在350×g下离心8分钟使细胞沉淀，并且以 $2 \times 10^6$ 个细胞/mL重悬于RPMI+10%FBS中。使PBMC在组织培养箱中静置大约16小时。

将静置的PBMC与自体抗原呈递细胞以大约10:1PBMC:APC的比率共培养。在这种情况下，APC是自体单核细胞衍生的树突细胞，其已经暴露于经辐射的HSV感染的HeLa细胞。将共培养物孵育5小时。不旨在限制于本文所述的方法，预期约5小时的孵育时间足以允许抗原特异性细胞响应于抗原或响应于由其他细胞释放到培养基中的细胞因子而表达新的mRNA。

通过轻轻地上下移液并移至新的管中，从共培养物中除去PBMC。将细胞置于冰上，并在细胞缓冲液(20g/L鱼皮明胶(Biotium)、155mM KC1、0.05%叠氮化钠、5mM HEPES-Na pH 7.5)+2mM EDTA中洗涤，然后在细胞缓冲液中洗涤，并且最后通过20微米筛网过滤器进行过滤并重悬于细胞缓冲液中。对细胞计数并通过用吖啶橙-碘化丙啶染色来评估存活活力。将最终细胞密度调整为 $3.5 \times 10^6$ 个活(碘化丙啶阴性)细胞/mL，并保存在冰上。

就在乳剂产生之前，将细胞悬浮液加热并放回到冰上1分钟。

#### B. 乳剂中的条码化

与先前的实施例中一样，形成含有制备的细胞和反应混合物的乳剂，所述乳剂用于随后的转录物标记反应以将分子条码和容器条码添加到单细胞多核苷酸分子中。将反应混合物制备为2×浓缩物，在微滴形成过程期间将所述浓缩物与细胞悬浮液以1:1的体积比混合。

#### 1. 乳剂反应混合物的制备

制备含有下表E3中的试剂和寡核苷酸的反应混合物。添加VB寡聚物1(SEQ ID NO:6)、2(SEQ ID NO:7)、3(SEQ ID NO:8)和4(SEQ ID NO:9)作为等摩尔混合物，以产生扩增子的碱基移位的(交错的)集合来增加测序期间的多样性。将反应混合物加载到先前实施例中所述的乳剂产生装置的反应样品回路中。

表E3.乳剂反应混合物
组分
水
HEPES-Na, pH 8.0
Triton X-100 (Surfact-Amps, Thermo Sci)
dNTP (dATP/dCTP/dTTP/dGTP中的每一种)
VB寡聚物 (含有C7衔接子序列)
反向VB引物
正向VB引物
基因特异性RT引物 (每个)
模板转换寡聚物 (MB条码)
RNA酶抑制剂
随机六聚体寡核苷酸
蛋白酶抑制剂
MgSO <sub>4</sub>
二硫苏糖醇
碳酸氢钠
GTP
盐酸胍
硫酸铵
EvaGreen染料 (Biotium)
Herculase II融合聚合酶 (Agilent)
Maxima H-逆转录酶 (Thermo)

寡聚物	序列 (5'-3')
VB (容器条码) 寡聚物1	T*A*C*G*TCTACCGCGCTGCTCTG CCACGACTGACGGACGTATT NNNNWNNNNWNNNNAGATCGGAAG AGCACACGTCTGAACCTCCA*G*T*C*A (SEQ ID NO:6)
VB (容器条码) 寡聚物2	T*A*C*G*TCTACCGCGCTGCTCTG CCACGACTGACGGACGTATT WNNNNWNNNNWNNNNAGATCGGAAG AGCACACGTCTGAACCTCCA*G*T*C*A (SEQ ID NO:7)
VB (容器条码) 寡聚物3	T*A*C*G*TCTACCGCGCTGCTCTG CCACGACTGACGGACGTATT NWNNNNWNNNNWNNNNAGATCGGAAG AGCACACGTCTGAACCTCCA*G*T*C*A (SEQ ID NO:8)
VB (容器条码) 寡聚物4	T*A*C*G*TCTACCGCGCTGCTCTG CCACGACTGACGGACGTATT NNWNNNNWNNNNWNNNNAGATCGGAAG AGCACACGTCTGAACCTCCA*G*T*C*A (SEQ ID NO:9)
正向VB引物	GTGACTGGAGTTCAGACGTGTGCT (SEQ ID NO:10)
反向VB引物	TACGTCTACCGCGCTGCTCTG (SEQ ID NO:11)
IgG恒定RT引物	AGGACAGCC mGmGmG AAGGTGT (SEQ ID NO:12)
IgL恒定RT引物	GCTCCCGG mG T mAmG AAGTCA (SEQ ID NO:13)
IgK恒定RT引物	GGCCTCTCTG mGmGmA TAGAAGT (SEQ ID NO:14)
IgM恒定RT引物	TGTGAGGTGGCT mGmCmG TACTTG (SEQ ID NO:15)
IgA恒定RT引物	CTGGCTRGGTG mGmGmA AGTTTCT (SEQ ID NO:16)

IgD恒定RT引物	CACGCATTTGT mAmC T mC GCCTTG (SEQ ID NO:17)
IgE恒定RT引物	GATGGTGGC mA T mAmG TGACCAG (SEQ ID NO:18)
TRA恒定RT引物	TGTTTGAGAATCAA mAmA T mC GGTGAA (SEQ ID NO:19)
TRB恒定RT引物	ACGTGGTC mGmGmG GAAGAAG (SEQ ID NO:20)
TRG恒定RT引物	CAAGAAGACAAA mGmG T mA TGTTCC (SEQ ID NO:21)
TRD恒定RT引物	TCTTCTTGGAT mGmAmC ACGAGA (SEQ ID NO:22)
模板转换寡聚物 (Trilink)	AATACGTCCGTCAAGTCGTGGATGU/(N)/(N)/T/(N)/(N)/A/(N) /(N)/T/[(po)rG]/[(po)rG]/[3-脱氧鸟苷]/ (SEQ ID NO: 23)

\*表示硫代磷酸酯键

N=A/T/C/G

W=A/T

mA/mG/mC=2'0-甲基A/G/C

U=2'脱氧尿苷

(po) rG=核糖鸟苷 (RNA碱基)

## 2. 从单细胞产生双条码化转录物文库

使用制备的细胞和反应混合物形成乳剂。乳剂产生平台包括由单个空气压缩机驱动的三个Mitos P泵 (Dolomite Microfluidics) (每个泵都带有Mitos流速传感器), 以允许将两个水相和一个亲氟连续油相计算机控制地流入亲氟包被的石英Dolomite Small 2-试剂芯片中。一个水性输入通道含有所需密度的细胞, 以产生所需的每微滴细胞占有水平, 而第二个水性通道含有裂解和反应混合物。

使用100 $\mu$ L Hamilton Microliter注射器在两次注射(每次大约100 $\mu$ L反应混合物)中使100 $\mu$ L内径PEEK管道样品回路过载。使用100 $\mu$ L Hamilton Gastight注射器将大约110 $\mu$ L细胞悬浮液加载到约110 $\mu$ L、内径0.2mm的FEP管道回路中。将回路连接至机械旋转器, 所述旋转器大约每1-2秒一次地不断反转细胞回路, 以防止细胞沉降和/或聚集。伴随从芯片中的两个油性通道中同时流油, 通过以相同的流速通过Dolomite 2-试剂芯片来集中射流水相而形成乳剂。外部油性通道含有在HFE7500 (Novec 7500) 氟碳油中的0.5%-5.0% (w/v) 聚乙二醇基表面活性剂。乳剂射流以恒定流速(在细胞相和反应相通道中相等)运行。通过12cm、内径0.5mm的PEEK管, 通过滴入在冷却块中保持在大约0°C下的4个重复的0.2mL PCR

带状管 (Eppendorf) 中, 收集乳剂芯片输出。用毛细管微量移液器从每个管的底部除去过量的油。

将乳剂在热循环仪中孵育以进行转录物标记反应。简言之, 将反应在4°C下预冷10分钟。然后, 在45分钟逆转录 (RT) 步骤期间, 使用靶特异性引物或使用基于六聚体寡核苷酸结合的随机引发, 基于模板转换添加含有随机化分子条码的通用衔接子序列, 在37°C下对RNA进行逆转录, 产生各自具有独特的分子标识符 (条码) 的cDNA分子。在RT后, 将温度保持在94°C持续10分钟。然后将乳剂进行50个循环的热循环 (每个循环: 83°C持续10秒 (变性)、65°C持续25秒 (延伸)) 以扩增容器条码寡聚物。在扩增VB寡聚物之后, 将乳剂进行10个循环的较高温热循环 (每个循环: 95°C持续10秒 (变性)、63°C持续25秒 (退火)、72°C持续2分20秒 (延伸))。在完成热循环的循环之后, 将乳剂保持在4°C下。

#### 实施例4-双条码化cDNA的纯化

对于每个乳剂级分管, 在上述实施例中产生双条码化cDNA转录物之后, 通过与等体积的1:1 (v:v) 全氟辛醇:FC-40混合将乳剂破坏 (在PCR之后), 并且添加EDTA至终浓度5mM以终止DNA聚合。添加大约0.1体积的Qiagen蛋白酶, 并将破坏的乳剂在50°C下孵育10至15分钟, 随后通过将管在95°C下孵育3分钟来热灭活蛋白酶。将管短暂离心, 并将上层水相转移至新的管中。

将双条码化cDNA浓缩, 并根据制造商的指示通过用1.8体积的AMPure XP (Beckman Coulter) 纯化来脱盐。将cDNA从珠上洗脱并通过添加8μL 0.1M氢氧化钠+1mM EDTA并加热至50°C持续3分钟来变性。如果全长产物由于RT引物的5'生物素化而含有生物素, 则可以通过在链霉亲和素珠上清除将此类全长产物与过量的微滴条码PCR产物分离。

在添加2μL 6X DNA加载染料 (New England Biolabs) 之后, 将变性的单链cDNA在30 mM NaOH+1 mM EDTA中的1.5% (w/v) 琼脂糖凝胶上在5V/cm下分离35分钟。在中和凝胶的pH之后, 将含有对应于100–1000 nt大小范围的cDNA的凝胶切下, 并使用DNA回收试剂盒 (例如, Zymoclean™凝胶DNA回收试剂盒, Zymo Research) 从切下的琼脂糖中纯化cDNA, 洗脱到20μL 10 mM Tris–Cl (pH 8.5) +0.05% TWEEN® 20中。通过在95°C下加热10秒并放在冰上沉淀, 将cDNA用1.8体积的AMPure XP珠进一步脱盐并洗脱到10.5μL 10 mM Tris–Cl (pH 8.0) +0.05% TWEEN® 20中。

#### 实施例5-3'衔接子序列与双条码化cDNA转录物的连接

如实施例3中所述产生并在如实施例4中所述纯化之后, 将含有已知引发位点的衔接子序列添加到双条码化cDNA转录物中。衔接子序列的添加允许用已知引物对所有转录物进行扩增、克隆或测序 (如下一代测序)。几个衔接子序列是已知的并通常用于测序, 如本文中使用的示例性P5衔接子序列。

#### 1.3'衔接子序列添加的方法

使用几种方法将衔接子序列添加到单链的双条码化cDNA序列的未知3'末端。将连接ssDNA衔接子的连接酶 (如Thermostable App连接酶 (NEB) 和CircLigase II (Epicentre)) 用于将衔接子序列添加到双条码化cDNA转录物的3'末端。基于将非模板核苷酸酶促添加到cDNA的3'末端, 还将商业化试剂盒 (Swift Biosciences Accel-NGS 1S DNA试剂盒) 用于添加衔接子序列。此外, 采用了使用与悬突于cDNA的3'末端的衔接子退火的简并夹板的方法, 简并悬突为至多6个核苷酸 (即, NNNNNN; SEQ ID NO:24)。对于简并悬突, NNNNNN (SEQ ID

NO:24)似乎在测试的方案中作用最好。

## 2. 使用具有6-核苷酸悬突的简并夹板将3'衔接子序列添加到双条码化cDNA

通过以1.2:1的比率混合含有短P5引发序列/5Phos/AGATCGGAAGAGCGTCGTGT/3AmMO/(SEQ ID NO:25)的寡核苷酸和如ACACGACGCTTCCGATCT NNNNNN/3AmMO/(SEQ ID NO:26)所示的夹板寡核苷酸,形成夹板-衔接子双链体分子。将退火缓冲液添加到30mM HEPES-Na pH 7.5、0.1 M KC1中。将溶液在热循环仪中于85°C下加热2分钟,并允许以0.1°C/秒的速率冷却至37°C。

然后,将从上面部分A中回收的双条码化cDNA转录物与夹板-衔接子溶液混合。然后,通过向混合物中添加等体积的Blunt/TA连接酶主混合物(New England Biolabs)并在室温下孵育,将衔接子连接至cDNA转录物。通过用AMPure XP纯化混合物并在10 mM Tris-C1 (pH 8.0)+0.05% TWEEN® 20中洗脱DNA来除去过量的衔接子DNA。

### 实施例6-转录组文库的PCR扩增和测序

#### A. 扩增多核苷酸文库

使用与C7通用衔接子序列(位于转录物的双条码和编码序列的5'(cDNA转录物的5'末端))互补的正向引物(SEQ ID NO:28)以及与P5通用衔接子序列(位于转录物的双条码和编码序列的3'(cDNA转录物的3'末端))互补的反向引物(SEQ ID NO:27),将经纯化的双条码化和通用衔接子标记的序列进行PCR扩增持续8个循环(PCR0)。

首先将PCR0反应混合物(含有衔接子连接的cDNA(在实施例5中产生)、DNA聚合酶、C7正向引物、P5反向引物、dNTP和反应缓冲液)在98°C下变性1分钟,随后是8个循环的热循环(每个循环:98°C持续10秒、69°C持续20秒、72°C持续10秒)以及在72°C下2分钟的最终延伸时间。在PCR完成后,将混合物保持在4°C下。

每个PCR0产生的cDNA序列含有C7衔接子序列、容器条码序列(用于宿主细胞鉴定)、分子条码序列(用于转录物鉴定)、转录物和P5衔接子序列。然后,将扩增的文库(PCR0产物)使用AMPure XP珠纯化并洗脱在10 mM Tris-C1 (pH 8.0)+0.05% TWEEN® 20中。

然后,将经纯化的转录组文库用于测序以下中的一种或多种:一个或多个全长靶向基因,如免疫受体;乳剂中所有细胞的转录组;和/或乳剂内一个或多个选定细胞的转录组,如下所述。

#### B. 对靶向基因测序

##### 1. PCR1:一个或多个靶基因的扩增

为了扩增选定的靶基因,使用用于PCR0的通用正向引物(C7-索引-P7引物;SEQ ID NO:28)和对一个或多个所需靶具有特异性的反向引物(例如,免疫球蛋白或T细胞受体特异性引物)扩增PCR0产物。用于PCR1反应的示例性靶引物如下表E4中所示。

表E4.示例性靶特异性反向引物序列

靶	引物序列5'至3'
IgG恒定	AAGTAGTCCTTGACCAGGCAGC (SEQ ID NO:29)
IgL恒定	GGCTTGAAGCTCCTCAGAGGA (SEQ ID NO:30)
IgK恒定	AGGCACACAAACAGAGGCAGTTC (SEQ ID NO:31)
IgM恒定	CGACGGGAATTCTCACAGGAG (SEQ ID NO:32)
IgD恒定	TGTCTGCACCCTGATATGATGG (SEQ ID NO:33)
IgA恒定1	GGGTGCTGCAGAGGCTCAG (SEQ ID NO:34)
IgA恒定2	GGGTGCTGTCGAGGCTCAG (SEQ ID NO:35)
IgE恒定	GGAATGTTTGCAGCAGCGGG (SEQ ID NO:36)
TRA恒定	AGTCTCTCAGCTGGTACACGG (SEQ ID NO:37)
TRB恒定	ATGGCTCAAACACAGCGACCTC (SEQ ID NO:38)

首先将PCR1反应混合物(含有PCR0产物(在上面部分A中产生)、DNA聚合酶、C7-索引-P7正向引物、靶特异性反向引物、dNTP和反应缓冲液)在98℃下变性1分钟,随后是10个循环的热循环(每个循环:98℃持续10秒、61℃持续20秒、72℃持续20秒)以及在72℃下2分钟的最终延伸时间。在PCR1完成后,将混合物保持在4℃下。将PCR1产物(一个或多个靶基因序列)使用AMPure XP珠纯化并洗脱在10 mM Tris-C1 (pH 8.0) +0.05% TWEEN®20中。

## 2. PCR2: 靶基因的扩增,添加3'测序衔接子序列

使用C7正向引物(SEQ ID NO:39)和含有通用引发P5短序列的靶特异性反向引物,扩增经纯化的PCR1产物。用于PCR2反应的示例性引物如下表E5中所示。

表E5.示例性PCR2引物

靶	序列5'至3'
C7	CAAGCAGAAGACGGCATACGAGAT (SEQ ID NO:39)
IgG	ACACGACGCTTCCGATCTCCAGGGGAAGACSGATG (SEQ ID NO:40)
IgG	ACACGACGCTTCCGATCTNCCAGGGGAAGACSGATG (SEQ ID NO:41)
IgG	ACACGACGCTTCCGATCTNNCCAGGGGAAGACSGATG (SEQ ID NO:42)
IgG	ACACGACGCTTCCGATCTNNNCCAGGGGAAGACSGATG (SEQ ID NO:43)
IgA	ACACGACGCTTCCGATCTGGCTCAGCGGAAGACCTTG (SEQ ID NO:44)
IgA	ACACGACGCTTCCGATCTNGGCTCAGCGGAAGACCTTG (SEQ ID NO:45)
IgA	ACACGACGCTTCCGATCTNNGGCTCAGCGGAAGACCTTG (SEQ ID NO:46)
IgA	ACACGACGCTTCCGATCTNNNGGCTCAGCGGAAGACCTTG (SEQ ID NO:47)
IgE	ACACGACGCTTCCGATCTGGGAAGACGGATGGCTCTG (SEQ ID NO:48)
IgE	ACACGACGCTTCCGATCTNGGAAGACGGATGGCTCTG (SEQ ID NO:49)
IgE	ACACGACGCTTCCGATCTNNGGAAGACGGATGGCTCTG (SEQ ID NO:50)
IgE	ACACGACGCTTCCGATCTNNNGGAAGACGGATGGCTCTG (SEQ ID NO:51)

IgM	ACACGACGCTTTCCGATCTGAGACGAGGTGGAAAAGGGTTG (SEQ ID NO:52)
IgM	ACACGACGCTTTCCGATCTNGAGACGAGGTGGAAAAGGGTTG (SEQ ID NO:53)
IgM	ACACGACGCTTTCCGATCTNNAGACGAGGTGGAAAAGGGTTG (SEQ ID NO:54)
IgM	ACACGACGCTTTCCGATCTNNNGAGACGAGGTGGAAAAGGGTTG (SEQ ID NO:55)
IgD	ACACGACGCTTTCCGATCTGGAACACATCCGGAGCCTTG (SEQ ID NO:56)
IgD	ACACGACGCTTTCCGATCTNGGAACACATCCGGAGCCTTG (SEQ ID NO:57)
IgD	ACACGACGCTTTCCGATCTNNGAACACATCCGGAGCCTTG (SEQ ID NO:58)
IgD	ACACGACGCTTTCCGATCTNNNGGAACACATCCGGAGCCTTG (SEQ ID NO:59)
IgL	ACACGACGCTTTCCGATCTAGGGYGGGAACAGAGTGAC (SEQ ID NO:60)
IgL	ACACGACGCTTTCCGATCTNAGGGYGGGAACAGAGTGAC (SEQ ID NO:61)
IgL	ACACGACGCTTTCCGATCTNNAGGGYGGGAACAGAGTGAC (SEQ ID NO: 62)
IgL	ACACGACGCTTTCCGATCTNNNAGGGYGGGAACAGAGTGAC (SEQ ID NO:63)
IgK	ACACGACGCTTTCCGATCTGACAGATGGTGCAGCCACAG (SEQ ID NO:64)
IgK	ACACGACGCTTTCCGATCTNGACAGATGGTGCAGCCACAG (SEQ ID NO:65)
IgK	ACACGACGCTTTCCGATCTNNGACAGATGGTGCAGCCACAG (SEQ ID NO:66)
IgK	ACACGACGCTTTCCGATCTNNNGACAGATGGTGCAGCCACAG

	(SEQ ID NO:67)
TRA	ACACGACGCTCTCCGATCTCACGGCAGGGTCAGGGTTC (SEQ ID NO:68)
TRA	ACACGACGCTCTCCGATCTNCACGGCAGGGTCAGGGTTC (SEQ ID NO:69)
TRA	ACACGACGCTCTCCGATCTNNCACGGCAGGGTCAGGGTTC (SEQ ID NO:70)
TRA	ACACGACGCTCTCCGATCTNNNCACGGCAGGGTCAGGGTTC (SEQ ID NO:71)
TRB	ACACGACGCTCTCCGATCTCGACCTCGGTGGAACAC (SEQ ID NO:72)
TRB	ACACGACGCTCTCCGATCTNCGACCTCGGTGGAACAC (SEQ ID NO:73)
TRB	ACACGACGCTCTCCGATCTNNCGACCTCGGTGGAACAC (SEQ ID NO:74)
TRB	ACACGACGCTCTCCGATCTNNNCACCTCGGTGGAACAC (SEQ ID NO:75)

首先将PCR2反应混合物(含有PCR1产物(在上面部分B1中产生)、DNA聚合酶、C7正向引物(SEQ ID NO:39)、靶特异性短P5反向引物、dNTP和反应缓冲液)在98℃下变性1分钟,随后是6个循环的热循环(每个循环:98℃持续10秒、65℃持续20秒、72℃持续20秒)以及在72℃下2分钟的最终延伸时间。在PCR2完成后,将混合物保持在4℃下。将PCR2产物(一个或多个靶基因序列)使用AMPure XP珠纯化并洗脱在10mM Tris-Cl (pH 8.0) +0.05% TWEEN® 20中。

### 3. 靶基因的定量PCR(qPCR3)和测序

将经纯化的PCR2或PCR0产物用于定量PCR(qPCR),以确定实现qPCR3终点的扩增循环数。使用C7(正向;SEQ ID NO:39)和C5-P5(反向;SEQ ID NO:76)引物扩增来自PCR0的预扩增衔接子连接的材料或者来自PCR2的全长IG或TR材料。

简言之,首先将qPCR3反应混合物(含有PCR0或PCR2产物(分别在上面部分A或B1中产生)、DNA聚合酶、C7正向引物(SEQ ID NO:39)、C5-P5反向引物(SEQ ID NO:76)、dNTP、EvaGreen和反应缓冲液)在98℃下变性1分钟,随后是3个循环的热循环(每个循环:98℃持续10秒、60℃持续20秒、72℃持续20秒),随后是30个循环的第二轮热循环(每个循环:98℃持续10秒、70℃持续20秒、72℃持续20秒)。检查了qPCR强度图以确定荧光强度为最大但DNA扩增尚未结束的扩增循环。这被确定为qPCR3终点的最终循环数。

在确定将每个文库扩增至指数期所需的PCR循环数之后,以非定量方式将相同的PCR重

复至所需的循环数，并用1 AMPure XP纯化并洗脱在10 mM Tris-C1 (pH 8.0) +0.05% TWEEN® 20中。任选地，通过DASH将qPCR3结果归一化(Gu等人, *Genome Biology* 2016, 17:41)。在Agilent Tapestation D1000带(tape)上分析结果，使用用于Illumina的KAPA NGS Quant试剂盒进行定量，并且对于衔接子连接文库使用Illumina NextSeq高输出75循环试剂盒(例如，32个循环读段1、6个循环索引读段I7、54个循环读段2)或者对于全长IG和TR文库使用Illumina MiSeq V3 600循环试剂盒(例如，325个循环读段1、6个循环索引读段I7、300个循环读段2)进行测序。在一些情况下，NextSeq测序仪使用56个循环的读段1、6个循环的索引读段I7和33个循环的读段2。

#### C. 对来自所有细胞的转录组测序

为了产生来自所有细胞的转录组文库，除了用针对3'衔接子序列的通用反向引物(例如，SEQ ID NO:27和76)代替靶特异性反向引物用于与通用正向引物(例如，SEQ ID NO:28和39)组合进行扩增以外，如上所述进行PCR1、PCR2和qPCR3反应。因此，通用正向和反向引物可以用于对乳剂中所有细胞的所有转录物进行测序。

#### D. 对来自选定细胞的转录组测序

为了产生来自选定细胞的转录组文库，除了使用与一个或多个所需细胞(如含有在上面部分A中测序的感兴趣的Ig分子或TCR的一个或多个细胞)的容器条码(VB)互补的正向引物以及针对3'衔接子序列的通用反向引物(例如SEQ ID NO:27和76)之外，如上所述进行PCR1、PCR2和qPCR3反应。

#### 实施例7-序列数据分析

处理Illumina MiSeq读段以产生mRNA分子和微滴的全长共有序列，用IgBLAST和IMGT/HighV-QUEST进行注释，并使用自定义脚本和Change-0软件包进行处理以产生统计数据和图。将MiSeq读段使用Illumina软件解复用。将低于Phred质量5的位置用N掩盖。在扩增子中鉴定同种型特异性引物、容器条码(VB)、分子条码(MB)和衔接子序列，并使用pRESTO MaskPrimers(以最大误差0.2修剪)进行修整。

#### A. 选定免疫受体序列数据序列数据的分析

[0477] 在一个例子中，制备靶向免疫受体的全长序列并对其进行测序。对于来自同时包含VB和MB的独特分子标识符(UMI)分组的读段的每个mRNA，单独产生读段1共有序列和读段2共有序列，所述读段是由相同起源的原始mRNA分子产生的PCR复制。将UMI读段组与MUSCLE进行比对，并使用pRESTO采用以下参数建立共有序列：maxdiv=0.1; bf PRIMER; prfreq=0.6; maxmiss=0.5; q=5; 读段组的所谓PCR引物序列一致性>60%；最大核苷酸多样性=0.1；对插入加缺失(indel)位置使用多数决规则；和掩盖后验(一致)质量低的比对列。然后，以两轮缝合配对的末端共有序列。首先，如在pRESTO AssemblePairs(使用以下参数比对：最小长度=8；阿尔法1 x 10<sup>5</sup>；以及最大误差=0.3)中实施的，将每个读段对的共有序列末端的无缺口比对使用Z得分近似法优化并使用二项式p值进行评分。对于无法通过这种方式缝合的读段对，使用pRESTO's AssemblePairs(参考参数：最小同一性=0.5；e值1 x 10<sup>-5</sup>)，尝试使用人BCR和TCR种系V外显子缝合，以在缝合或有缺口的读段连接之前搭接每个读段。

#### 1. V D J区段注释和同种型确认

将IgBLAST、Change-0和自定义脚本用于鉴定起源的种系V(D)J基因，将mRNA序列修整

成V(D)J区,鉴定CDR3区并计算来自种系V核苷酸序列的突变。IgBLAST将N计数为错配,但过滤出具有超过6个V区N的mRNA序列进行突变分析和交叉部分配对精度分析。对于IG重链,使用pRESTO MaskPrimers(得分参数:起始=0;最大误差=0.2)通过将非引物C区(恒定区外显子)与预期序列进行匹配来确认同种型身份。除两个引物/非引物组合(其中特定的引物串扰事件通过可视检查解决)之外,丢弃了具有不一致的引物/非引物C区调用的扩增子。

## 2. 将V(D)J序列分组到克隆谱系中

使用单链接聚类用加权克隆内距离将V(D)J序列分组到克隆中。使用Change-0软件包DefineClones(分组参数:模型=m1n;基因=第一;距离=4.0;norm=无)进行聚类。首先,将所有功能性Ig V<sub>H</sub>链的微滴共有序列分箱到V-J连接箱元中,使得将可能由同一初始重组事件产生的序列分箱到一起(基于最佳匹配的Ig V<sub>H</sub>基因、最佳匹配的Ig J<sub>H</sub>基因和如通过IMGT/HighV-QUEST鉴定的连接长度)。通过使用Change-0的shm软件包的distToNearest功能在每个Ig V<sub>H</sub>箱元内产生最近邻距离的直方图并目视检查直方图的自然距离截止值(在双峰直方图的谷底),选择克隆内距离阈值。使用相同距离模型和阈值来限定轻链的克隆簇。

## 3. 微滴过滤,配对保真度计算

以两种独立的方式评估了重链轻链配对的置信度:使用微滴内mRNA序列一致性和复制对间的一致性。将微滴内mRNA一致性定义为基因座内V(D)J序列的平均成对核苷酸差异(Sei's pi<0.02)。使用IgBLAST注释将mRNA序列修整为V(D)J核苷酸编码序列。在每个微滴内,将所有生产性mRNA序列均按V基因座分组。在每个组内,如使用默认参数在pRESTOAlignSets中实施的,使用MUSCLE对多个序列进行比对。使用pRESTO参数(BuildConsensus.py;最大div=0.2;最大miss=0.5)按基因座从多个mRNA建立微滴共链。将随机改组的微滴用于选择多样性截止值pi<0.02。在经改组的微滴中,少于0.01%的重链基因座(<0.2%的轻链基因座)满足此标准。分离包括多细胞或免疫受体的微滴,供进一步精度分析。

基于多个复制(单独的乳剂实验)中相同克隆对的观察结果计算配对精度,集中于可能仅含有单一谱系(即,由单个V(D)J和VJ重排、随后扩增产生)的那些VDJ簇。类似的VDJ重排可以在单独的多个独立时间内出现,从而导致相同的重链V(D)J重排与多个不同的轻链VJ重排天然地配对。因为罕见的V(D)J重排将提供通过本文所述方法实现的技术精度的更准确度量,所以长的重链CDR3(CDR3H)成为此分析的焦点(作为较罕见的V(D)J重排的替代)。还去除了具有>6N的序列,以增加克隆分配的置信度。对于跨部分观察到的最长四分位数的克隆(连接长度≥54nt的2,604个克隆),配对精度随着CDR 3H长度增加到96%以上。因为克隆对一致性的概率是两个独立实验中的真实对的联合概率,因此将配对精度估算为重复中配对一致性的平方根,计算如下:其中 $d_{hl}^f$ 是具有配对的重链克隆h和轻链克隆l的容器条码d的数量,并发现于物理部分f中。每个实验的平均(平方)配对精度通过在重链克隆h和所有部分对(f,g)上对配对的轻链克隆(1,k)的一致性求平均来估算:

$$(精度^2) = \text{平均值 } (P_f P_g)$$

跨部分一致的重链轻链对

跨部分观察到重链克隆的总对

$$\frac{\text{一致的重链轻链对}}{\text{一致对 + 不一致对}} = \frac{\sum_h (\sum_{l=k}^{f+g} d_{hl}^f \cdot d_{hk}^g)}{\sum_h (\sum_{l=k}^{f+g} d_{hl}^f \cdot d_{hk}^g + \sum_{l=k}^{f+g} d_{hl}^f \cdot d_{hk}^g)}$$

$$(精度^2) = \frac{33157}{35922}$$

因此,根据此示例性实验,每个实验的平均精度(在实验之间的精度差异之内)为96.1%。

#### B. 转录组序列数据分析

对于转录组序列数据,将包含相同VB的读段分解,并将序列与人参考基因组进行比对以鉴定转录物(HiSAT2)。使用samtools操纵输出文件的比对,并按基因组位置将读段分配给转录物。

对于每个VB基因组作图,将读段按MB分解。建立MB计数的矩阵,作图到每个微滴的每个单独的参考基因上。然后,将这些数据与靶数据合并,所述靶数据例如为如上面部分A中所述处理的免疫受体序列数据。将来自每个VB(微滴)的数据用基因计数和受体信息进行注释。然后分析合并的数据集以检查单细胞RNA序列谱(scRNAseq)。使用t-SNE、Seurat、ZIFA、PCA、LDA和其他示例性程序进行数据降维、聚类和可视化。

#### 实施例8-大量单细胞的示例性高通量转录组序列数据分析和绘图

总体上如上面实施例4A和4B中所述,制备约7,000个PBMC并对转录组和全长BCR和TCR受体进行测序。

在通过Illumina NextSeq分析之前,通过D1000DNA tapestation分析如实施例4B中所述制备的乳剂的转录组测序文库。转录组由大小从大约170-900bp范围的序列代表。在NextSeq分析之前,还通过D1000DNA tapestation对由实施例4A中所述的靶向测序进行测序的全长TCR序列进行了分析,其表明TCR $\alpha$ 和 $\beta$ 峰分别位于628bp和664bp。

在NextSeq分析之后,通过t分布随机邻域嵌入(t-SNE)和Seurat聚类分析具有>1,000个读段的所有微滴细胞谱( $n=6,707$ )。使用t-SNE图将多维单细胞转录组数据可视化,并且基于具有相似转录谱的细胞的Seurat聚类(图3A)或按测序的免疫受体(即,BCR(红色/中灰色)或TCR(绿色/深灰色))的性质(图3B)对细胞进行颜色编码,展示了具有相似表型的细胞的聚类。

分析了测序的细胞中示例性基因的转录组数据,并将具有与测序的免疫受体相同的容器条码的序列信息鉴定为来自同一细胞。基于表达水平的热图对转录组中示例性基因的单细胞转录组数据进行颜色编码,并显示了Toll样受体7(TLR7;图4A)、T细胞表面糖蛋白CD3 $\epsilon$ 链(CD3E;图4B)、自然杀伤细胞颗粒蛋白7(NKG7;图4C)、MRC1甘露糖受体C型1(MRC1;图4D)的单细胞转录组数据。

这些结果表明,由于与T细胞(CD3E)或B细胞(TLR7)相关的基因标记通常分别在表达全长TCR或BCR的细胞中聚类在一起,因此能以高通量的方式在免疫受体旁边捕获全基因组

RNA表达谱。同样,与T细胞或B细胞不相关的基因标记(如示例性NK细胞标记NKG7或示例性单核细胞标记MRC1)似乎在具有全长TCR或BCR免疫受体的细胞中不聚类。

本发明在范围上并不旨在限于具体公开的实施方案,提供具体公开的实施方案例如是为了说明本发明的各个方面。根据本文的描述和传授,对组合物和方法的各种修改将变得清楚。可以在不背离本公开文本的真实范围和精神的情况下实践此类变化,并且此类变化旨在落入本公开文本的范围内。

### 序列

#	序列	注释
1	/5BiosG//iSpl8/TTT TTT TTT TTT TTT TTT TTT TTT T T V N	5'生物素寡聚dT锚定的逆转录引物
2	ATCCATCCACGACTGACGGACGTATTAAANNNNNNNNN WNNNNAGATCGGAAGAGCACACGTCTGAACCTCCAGTCA CC	容器条码模板寡聚物
3	AATACGTCCCGTCAGTCGTGGATGNNTNNANNNTrGrGG	模板转换寡聚物
4	CATCCACGACTGACGGACGTATT	容器条码正向引物/通用衔接子寡核苷酸

5	GTGACTGGAGTTCAGACGTGTGCT	容器条码反向引物
6	T*A*C*G*TCTACGCGCTGCTCTG CCACGACTGACGGACGTATT NNNNWNNNNWNNNNAGATCGGAAG AGCACACGTCTGAACCCA*G*T*C*A	容器条码寡聚物1
7	T*A*C*G*TCTACGCGCTGCTCTG CCACGACTGACGGACGTATT WNNNNWNNNNWNNNNAGATCGGAAG AGCACACGTCTGAACCCA*G*T*C*A	容器条码寡聚物2
8	T*A*C*G*TCTACGCGCTGCTCTG CCACGACTGACGGACGTATT NWNNNNWNNNNWNNNNAGATCGGAAG AGCACACGTCTGAACCCA*G*T*C*A	容器条码寡聚物3
9	T*A*C*G*TCTACGCGCTGCTCTG CCACGACTGACGGACGTATT NNWNNNNWNNNNWNNNNAGATCGGAAG AGCACACGTCTGAACCCA*G*T*C*A	容器条码寡聚物4
10	GTGACTGGAGTTCAGACGTGTGCT	容器条码正向引物
11	TACGTCTACGCGCTGCTCTG	容器条码反向引物
12	AGGACAGCC mGmGmG AAGGTGT	IgG恒定RT引物
13	GCTCCCGG mG T mAmG AAGTCA	IgL恒定RT引物
14	GGCCTCTCTG mGmGmA TAGAAGT	IgK恒定RT引物
15	TGTGAGGTGGCT mGmCmG TACTTG	IgM恒定RT引物
16	CTGGCTRGGTG mGmGmA AGTTTCT	IgA恒定RT引物
17	CACGCATTTGT mAmC T mC GCCTTG	IgD恒定RT引物
18	GATGGTGGC mA T mAmG TGACCAAG	IgE恒定RT引物
19	TGTTTGAGAATCAA mAmA T mC GGTGAA	TRA恒定RT引物
20	ACGTGGTC mGmGmG GAAGAAG	TRB恒定RT引物
21	CAAGAAGACAAA mGmG T mA TGTTCC	TRG恒定RT引物
22	TCTTCTTGGAT mGmAmC ACGAGA	TRD恒定RT引物

23	AATACGTCCGTCAGTCGTGGATGU/(N)/(N)/(N)/(N)/(A/(N)/(N)/(N)/(T/[(po)rG]/[(po)rG]/[3-脱氧鸟苷]/	模 板 转 换 寡 聚 物 (Trilink)
24	NNNNNN	简 并 悬 突
25	/5Phos/ AGATCGGAAGAGCGTCGTGT /3AmMO	短P5引 发 序 列
26	ACACGACGCTCTTCCGATCT NNNNNN /3AmMO/	夹 板 寡 核 苷 酸
27	ACACGACGCTCTTCCGATCT	短P5反 向 引 物
28	CAAGCAGAACAGGCCATACGAGAT[NNNNNN]GTGACTG GAGTTCAGACGTGTGCTCTTCCGATCT	C7-索 引 -P7正 向 引 物
29	AAGTAGTCCTTGACCAGGCAGC	IgG恒定反向引物序 列
30	GGCTTGAAGCTCCTCAGAGGA	IgL恒定反向引物序 列
31	AGGCACACAAACAGAGGCAGTTC	IgK恒定反向引物序 列
32	CGACGGGAATTCTCACAGGAG	IgM恒定反向引物序 列
33	TGTCTGCACCCGTATGATGG	IgD恒定反向引物序 列
34	GGGTGCTGCAGAGGCTCAG	IgA恒定1反向引物序 列
35	GGGTGCTGTCGAGGCTCAG	IgA恒定2反向引物序 列
36	GGAATGTTTGCGAGCAGCGGG	IgE恒定反向引物序 列
37	AGTCTCTCAGCTGGTACACGG	TRA恒定反向引物序 列
38	ATGGCTAAACACAGCGACCTC	TRB恒定反向引物序 列
39	CAAGCAGAACAGGCCATACGAGAT	C7
40	ACACGACGCTCTTCCGATCTCCAGGGGAAGACSGATG	IgG衔 接 子 标 记 的 靶 特 异 性 反 向 引 物 序 列

41	ACACGACGCTCTTCCGATCTNCCAGGGGGAAAGACSGATG	IgG 衔接子标记的靶特异性反向引物序列
42	ACACGACGCTCTTCCGATCTNNCCAGGGGGAAAGACSGATG	IgG 衔接子标记的靶特异性反向引物序列
43	ACACGACGCTCTTCCGATCTNNNCAGGGGGAAAGACSGATG	IgG 衔接子标记的靶特异性反向引物序列
44	ACACGACGCTCTTCCGATCTGGCTCAGCGGGAAAGACCTTG	IgA 衔接子标记的靶特异性反向引物序列
45	ACACGACGCTCTTCCGATCTNGGCTCAGCGGGAAAGACCTTG	IgA 衔接子标记的靶特异性反向引物序列
46	ACACGACGCTCTTCCGATCTNNGGCTCAGCGGGAAAGACCTTG	IgA 衔接子标记的靶特异性反向引物序列
47	ACACGACGCTCTTCCGATCTNNNGGCTCAGCGGGAAAGACCTTG	IgA 衔接子标记的靶特异性反向引物序列
48	ACACGACGCTCTTCCGATCTGGGAAGACGGATGGCTCTG	IgE 衔接子标记的靶特异性反向引物序列
49	ACACGACGCTCTTCCGATCTNGGAAGACGGATGGCTCTG	IgE 衔接子标记的靶特异性反向引物序列
50	ACACGACGCTCTTCCGATCTNNGGAAGACGGATGGCTCTG	IgE 衔接子标记的靶特异性反向引物序列
51	ACACGACGCTCTTCCGATCTNNNGGAAGACGGATGGCTCTG	IgE 衔接子标记的靶特异性反向引物序列
52	ACACGACGCTCTTCCGATCTGAGACGAGGTGGAAAAGGGTTG	IgM 衔接子标记的靶特异性反向引物序列
53	ACACGACGCTCTTCCGATCTNGAGACGAGGTGGAAAAGGTTG	IgM 衔接子标记的靶特异性反向引物序列
54	ACACGACGCTCTTCCGATCTNNGAGACGGAGGTGGAAAAGGGTTG	IgM 衔接子标记的靶特异性反向引物序列
55	ACACGACGCTCTTCCGATCTNNNGAGACGAGGTGGAAAAAGGGTTG	IgM 衔接子标记的靶特异性反向引物序列

56	ACACGACGCTTCCGATCTGGAACACATCCGGAGCCTTG	IgD 衔接子标记的靶特异性反向引物序列
57	ACACGACGCTTCCGATCTNGGAACACATCCGGAGCCTTG	IgD 衔接子标记的靶特异性反向引物序列
58	ACACGACGCTTCCGATCTNNGAACACATCCGGAGCCTTG	IgD 衔接子标记的靶特异性反向引物序列
59	ACACGACGCTTCCGATCTNNNGAACACATCCGGAGCCTTG	IgD 衔接子标记的靶特异性反向引物序列
60	ACACGACGCTTCCGATCTAGGGYGGAACAGAGTGAC	IgL 衔接子标记的靶特异性反向引物序列
61	ACACGACGCTTCCGATCTNAGGGYGGAACAGAGTGA	IgL 衔接子标记的靶特异性反向引物序列
62	ACACGACGCTTCCGATCTNNAGGGYGGAACAGAGTGAC	IgL 衔接子标记的靶特异性反向引物序列
63	ACACGACGCTTCCGATCTNNNAGGGYGGAACAGAGTGAC	IgL 衔接子标记的靶特异性反向引物序列
64	ACACGACGCTTCCGATCTGACAGATGGTGCAGCCACAG	IgK 衔接子标记的靶特异性反向引物序列
65	ACACGACGCTTCCGATCTNGACAGATGGTGCAGCCACAG	IgK 衔接子标记的靶特异性反向引物序列
66	ACACGACGCTTCCGATCTNNGACAGATGGTGCAGCCA	IgK 衔接子标记的靶特异性反向引物序列
67	ACACGACGCTTCCGATCTNNNGACAGATGGTGCAGCCACAG	IgK 衔接子标记的靶特异性反向引物序列
68	ACACGACGCTTCCGATCTCACGGCAGGGTCAGGGTC	TRA 衔接子标记的靶特异性反向引物序列
69	ACACGACGCTTCCGATCTNCACGGCAGGGTCAGGGTC	TRA 衔接子标记的靶特异性反向引物序列
70	ACACGACGCTTCCGATCTNNCACGGCAGGGTCAGGGTCAGGGTC	TRA 衔接子标记的靶特异性反向引物序列

71	ACACGACGCTTCCGATCTNNNCACGGCAGGGTCAGGG TTC	TRA衔接子标记的靶特异性反向引物序列
72	ACACGACGCTTCCGATCTCGACCTCGGTGGAACAC	TRB衔接子标记的靶特异性反向引物序列
73	ACACGACGCTTCCGATCTNCGACCTCGGTGGAACAC C	TRB衔接子标记的靶特异性反向引物序列
74	ACACGACGCTTCCGATCTNNCGACCTCGGTGGAAC AC	TRB衔接子标记的靶特异性反向引物序列
75	ACACGACGCTTCCGATCTNNNCACCTCGGTGGAAAC CAC	TRB衔接子标记的靶特异性反向引物序列
76	AATGATA CGCG ACCACCGAGATCTACACTCTTCC CACGACGCTCTTCC	C5-P5反向引物
77	AGATCGGAAGAGCACACGTCTGAACTCCA	P7引发位点 (C7)
78	AGATCGGAAGAGCGTCGTAGGGAAAGAGTGTAGATC TCGGTGGTCGCCGTATCATT	P5引发位点
79	ACACGACGCTTCCGATCTNNNNNN	通用衔接子; N是任何核苷酸
80	NNNNWNNNNWNNNN	通用衔接子; N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶
81	WNNNNWNNNNWNNNN	通用衔接子; N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶
82	NWNNNWNNNNWNNNN	通用衔接子; N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶
83	NNWNNNWNNNNWNNNN	通用衔接子; N是任何核苷酸并且W是腺嘌呤或胸腺嘧啶
84	/生物素/TGTGAGGTGGCTGCGTACTTG	IgM-RT引物
85	/生物素/AGGACAGCCGGGAAGGTGT	IgG-RT引物

86	/生物素/CACGCATTTGACTCGCCTTG	IgD-RT引物
87	/生物素/CTGGCTRGGTGGGAAGTTCT	IgA-RT引物
88	/生物素/GGTGGCATAGTGACCAGAGA	IgE-RT引物
89	/生物素/TATTCAGCAGGCACACAAACAGA	IgK-RT引物
90	/生物素/AGTGTGGCCTGTTGGCTTG	IgL-RT引物
91	/生物素/GGGAGATCTCTGCTTCTGATG	TCR-A-RT引物
92	/生物素/GGTGAATAGGCAGACAGACTTG	TCR-B-RT引物
93	/生物素/GGCAGTCATCCGAACACT	CD4-RT引物
94	/生物素/CTACAAAGTGGGCCCTCTG	CD-8-RT引物
95	ACACGACGCTCTCCGATCTGTGGCCTGCCGAGGGAG G	CD4-巢式引物
96	ACACGACGCTCTCCGATCTGCGGAATCCCAGAGGGCC A	CD8-巢式引物
97	CAAGCAGAAGACGGCATACGAGATNNNNNNNTGACTGG AGTTCAAGACGTGTGCTCTCCGATCT	C7-bc-P7引物
98	AATGATAACGGCGACCACCGAGATCTACACTCTTCCCTA CACGACGCTCTCCGATCT	C5-P5引物
99	NNNNWNNNNWNNNNWNNNNW	容器条码序列
100	NNNNWISCNNNWISCNNN	容器条码序列

## 序列表

<110> Goldfless, Stephen Jacob  
Briggs, Adrian Wrangham  
Chari, Rajagopal  
JIang, Yue  
Hause, Ronald  
Vigneault, Francois  
<120> 高通量多核苷酸文库测序和转录组分析  
<130> 735042011740  
<140> 尚未分配  
<141> 与此同时提交  
<150> US 62/511,949  
<151> 2017-05-26  
<160> 100  
<170> 用于Windows版本4.0的FastSEQ  
<210> 1  
<211> 27  
<212> DNA  
<213> 人工序列  
<220>  
<223> 5'生物素寡聚-dT锚定的逆转录引物  
<220>  
<221> 尚未归类的特征  
<222> 1  
<223> 具有18-碳间隔子的5'生物素修饰  
<220>  
<221> 尚未归类的特征  
<222> 27  
<223> n = A、T、C或G  
<400> 1  
ttttttttt ttttttttt ttttvn 27  
<210> 2  
<211> 78  
<212> DNA  
<213> 人工序列  
<220>  
<223> 容器条码模板寡聚物  
<220>

<221> 尚未归类的特征  
<222> 30, 31, 32, 33, 35, 36, 37, 38, 40, 41, 42, 43  
<223> n = A、T、C或G  
<400> 2  
atccatccac gactgacgga cgtattaaan nnnwnnnwn nnnagatcg aagagcacac 60  
gtctgaactc cagtcacc 78  
<210> 3  
<211> 35  
<212> DNA  
<213> 人工序列  
<220>  
<223> 模板转换寡聚物  
<220>  
<221> 尚未归类的特征  
<222> 33  
<223> 核糖鸟苷  
<220>  
<221> 尚未归类的特征  
<222> 34  
<223> 核糖鸟苷  
<220>  
<221> 尚未归类的特征  
<222> 24, 25, 27, 28, 30, 31  
<223> n = A、T、C或G  
<400> 3  
aatacgtccg tcagtcgtgg atgnntnnan ntggg 35  
<210> 4  
<211> 23  
<212> DNA  
<213> 人工序列  
<220>  
<223> 容器条码正向引物/通用衔接子寡核苷酸  
<400> 4  
catccacgac tgacggacgt att 23  
<210> 5  
<211> 24  
<212> DNA  
<213> 人工序列  
<220>

<223> 容器条码反向引物  
<400> 5  
gtgactggag ttcagacgtg tgct 24  
<210> 6  
<211> 87  
<212> DNA  
<213> 人工序列  
<220>  
<223> 容器条码寡聚物1  
<220>  
<221> 尚未归类的特征  
<222> (1) ... (5)  
<223> 硫代磷酸酯键  
<220>  
<221> 尚未归类的特征  
<222> (83) ... (87)  
<223> 硫代磷酸酯键  
<220>  
<221> 尚未归类的特征  
<222> 41, 42, 43, 44, 46, 47, 48, 49, 51, 52, 53, 54  
<223> n = A、T、C或G  
<400> 6  
tacgtctacg cgctgctctg ccacgactga cggacgtatt nnnnwnnnnw nnnnagatcg 60  
gaagagcaca cgtctgaact ccagtca 87  
<210> 7  
<211> 88  
<212> DNA  
<213> 人工序列  
<220>  
<223> 容器条码寡聚物2  
<220>  
<221> 尚未归类的特征  
<222> (1) ... (5)  
<223> 硫代磷酸酯键  
<220>  
<221> 尚未归类的特征  
<222> (84) ... (88)  
<223> 硫代磷酸酯键  
<220>

<221> 尚未归类的特征  
<222> 42, 43, 44, 45, 47, 48, 49, 50, 52, 53, 54, 55  
<223> n = A、T、C或G  
<400> 7  
tacgtctacg cgctgctctg ccacgactga cggacgtatt wnnnnwnnnn wnnnnnagatc 60  
ggaagagcac acgtctgaac tccagtca 88  
<210> 8  
<211> 89  
<212> DNA  
<213> 人工序列  
<220>  
<223> 容器条码寡聚物3  
<220>  
<221> 尚未归类的特征  
<222> (1) ... (5)  
<223> 硫代磷酸酯键  
<220>  
<221> 尚未归类的特征  
<222> (85) ... (89)  
<223> 硫代磷酸酯键  
<220>  
<221> 尚未归类的特征  
<222> 41, 43, 44, 45, 46, 48, 49, 50, 51, 53, 54, 55, 56  
<223> n = A、T、C或G  
<400> 8  
tacgtctacg cgctgctctg ccacgactga cggacgtatt nwnnnnwnnnn nwnnnnnagat 60  
cggaagagca cacgtctgaa ctccagtca 89  
<210> 9  
<211> 90  
<212> DNA  
<213> 人工序列  
<220>  
<223> 容器条码寡聚物4  
<220>  
<221> 尚未归类的特征  
<222> (1) ... (5)  
<223> 硫代磷酸酯  
<220>  
<221> 尚未归类的特征

- <222> (86) ... (90)
- <223> 硫代磷酸酯
- <220>
- <221> 尚未归类的特征
- <222> 41, 42, 44, 45, 46, 47, 49, 50, 51, 52, 54, 55, 56, 57
- <223> n = A,T,C或G
- <400> 9
- tacgtctacg cgctgctctg ccacgactga cggacgtatt nnwnnnnwnn nnwnnnnaga 60  
tcggaagagc acacgtctga actccagtca 90
- <210> 10
- <211> 24
- <212> DNA
- <213> 人工序列
- <220>
- <223> 容器条码正向引物
- <400> 10
- gtgactggag ttcagacgtg tgct 24
- <210> 11
- <211> 20
- <212> DNA
- <213> 人工序列
- <220>
- <223> 容器条码反向引物
- <400> 11
- tacgtctacg cgctgctctg 20
- <210> 12
- <211> 19
- <212> DNA
- <213> 人工序列
- <220>
- <223> IgG恒定RT引物
- <220>
- <221> 经修饰的碱基
- <222> 10
- <223> gm
- <220>
- <221> 经修饰的碱基
- <222> 11
- <223> gm

<220>  
<221> 经修饰的碱基  
<222> 12  
<223> gm  
<400> 12  
aggacagccg ggaagggtgt 19  
<210> 13  
<211> 18  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgL恒定RT引物  
<220>  
<221> 经修饰的碱基  
<222> 9  
<223> gm  
<220>  
<221> 经修饰的碱基  
<222> 11  
<223> 2'-0-甲基-腺苷  
<220>  
<221> 经修饰的碱基  
<222> 12  
<223> gm  
<400> 13  
gctcccggtt agaagtca 18  
<210> 14  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgK恒定RT引物  
<220>  
<221> 经修饰的碱基  
<222> 11  
<223> gm  
<220>  
<221> 经修饰的碱基  
<222> 12

<223> gm  
<220>  
<221> 经修饰的碱基  
<222> 13  
<223> 2'-0-甲基-腺苷  
<400> 14  
ggcctctctg ggatagaagt 20  
<210> 15  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgM恒定RT引物  
<220>  
<221> 经修饰的碱基  
<222> 13  
<223> gm  
<220>  
<221> 经修饰的碱基  
<222> 14  
<223> cm  
<220>  
<221> 经修饰的碱基  
<222> 15  
<223> gm  
<400> 15  
tgtgaggtgg ctgcgtactt g 21  
<210> 16  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgA恒定RT引物  
<220>  
<221> 经修饰的碱基  
<222> 12  
<223> gm  
<220>  
<221> 经修饰的碱基

<222> 13  
<223> gm  
<220>  
<221> 经修饰的碱基  
<222> 14  
<223> 2'-0-甲基腺苷  
<400> 16  
ctggctrggt gggaagttc t 21  
<210> 17  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgD恒定RT引物  
<220>  
<221> 经修饰的碱基  
<222> 12  
<223> 2'-0-甲基腺苷  
<220>  
<221> 经修饰的碱基  
<222> 13  
<223> gmcm  
<220>  
<221> 经修饰的碱基  
<222> 15  
<223> cm  
<400> 17  
cacgcattg tactgcctt g 21  
<210> 18  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgE恒定RT引物  
<220>  
<221> 经修饰的碱基  
<222> 10  
<223> 2'-0-甲基腺苷  
<220>

<221> 经修饰的碱基  
<222> 12  
<223> 2'-0-甲基腺苷  
<220>  
<221> 经修饰的碱基  
<222> 13  
<223> gm  
<400> 18  
gatggtaggca tagtgaccag 20  
<210> 19  
<211> 24  
<212> DNA  
<213> 人工序列  
<220>  
<223> TRA恒定RT引物  
<220>  
<221> 经修饰的碱基  
<222> 15  
<223> 2'-0-甲基腺苷  
<220>  
<221> 经修饰的碱基  
<222> 16  
<223> 2'-0-甲基腺苷  
<220>  
<221> 经修饰的碱基  
<222> 18  
<223> cm  
<400> 19  
tgtttgagaa tcaaaatcggtgaa 24  
<210> 20  
<211> 18  
<212> DNA  
<213> 人工序列  
<220>  
<223> TRB恒定RT引物  
<220>  
<221> 经修饰的碱基  
<222> 9  
<223> gm

<220>  
<221> 经修饰的碱基  
<222> 10  
<223> gm  
<220>  
<221> 经修饰的碱基  
<222> 11  
<223> gm  
<400> 20  
acgtggtcgg ggaagaag 18  
<210> 21  
<211> 22  
<212> DNA  
<213> 人工序列  
<220>  
<223> TRG恒定RT引物  
<220>  
<221> 经修饰的碱基  
<222> 13  
<223> ac4cgm  
<220>  
<221> 经修饰的碱基  
<222> 14  
<223> gm  
<220>  
<221> 经修饰的碱基  
<222> 16  
<223> 2'-0-甲基腺苷  
<400> 21  
caagaagaca aaggatgtt cc 22  
<210> 22  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> TRD恒定RT引物  
<220>  
<221> 经修饰的碱基  
<222> 12

<223> gm  
<220>  
<221> 经修饰的碱基  
<222> 13  
<223> 2'-0-甲基腺苷  
<220>  
<221> 经修饰的碱基  
<222> 14  
<223> cm  
<400> 22  
tcttcttgaa tgacacgaga 20  
<210> 23  
<211> 36  
<212> DNA  
<213> 人工序列  
<220>  
<223> 模板转换寡聚物  
<220>  
<221> 经修饰的碱基  
<222> 24  
<223> 2'-脱氧尿苷  
<220>  
<221> 尚未归类的特征  
<222> 34  
<223> 核糖鸟苷  
<220>  
<221> 尚未归类的特征  
<222> 35  
<223> 核糖鸟苷  
<220>  
<221> 经修饰的碱基  
<222> 36  
<223> 3-脱氧鸟苷  
<220>  
<221> 尚未归类的特征  
<222> 25, 26, 28, 29, 31, 32  
<223> n = A、T、C或G  
<400> 23  
aataacgtccg tcagtcgtgg atgattnna nntggg 36

- <210> 24
- <211> 6
- <212> DNA
- <213> 人工序列
- <220>
- <223> 简并悬突
- <220>
- <221> 尚未归类的特征
- <222> 1, 2, 3, 4, 5, 6
- <223> n = A、T、C或G
- <400> 24
- nnnnnn 6
- <210> 25
- <211> 20
- <212> DNA
- <213> 人工序列
- <220>
- <223> 短的P5引发序列
- <220>
- <221> 经修饰的碱基
- <222> 1
- <223> 磷酸化腺苷
- <220>
- <221> 经修饰的碱基
- <222> 20
- <223> 3'氨基修饰的 (3AmMO) 胸腺嘧啶
- <400> 25
- agatcggaaag agcgtcggt 20
- <210> 26
- <211> 26
- <212> DNA
- <213> 人工序列
- <220>
- <223> 夹板寡核苷酸
- <220>
- <221> 经修饰的碱基
- <222> 26
- <223> 3'氨基修饰的 (3AmMO) 核苷酸 (a、g、c或t)
- <220>

- <221> 尚未归类的特征  
<222> 21, 22, 23, 24, 25, 26  
<223> n = A、T、C或G  
<400> 26  
acacgacgct cttccgatct nnnnnn 26  
<210> 27  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> 短的P5反向引物  
<400> 27  
acacgacgct cttccgatct 20  
<210> 28  
<211> 64  
<212> DNA  
<213> 人工序列  
<220>  
<223> C7-索引-P7正向引物  
<220>  
<221> 尚未归类的特征  
<222> 25, 26, 27, 28, 29, 30  
<223> n = A、T、C或G  
<400> 28  
caagcagaag acggcatacg agatnnnnn gtgactggag ttcagacgtg tgctcttccg 60  
atct 64  
<210> 29  
<211> 22  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgG恒定反向引物序列  
<400> 29  
aagtagtcct tgaccaggca gc 22  
<210> 30  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>

<223> IgL恒定反向引物序列  
<400> 30  
ggcttgaagc tcctcagagg a 21  
<210> 31  
<211> 22  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgK恒定反向引物序列  
<400> 31  
aggcacacaa cagaggcagt tc 22  
<210> 32  
<211> 22  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgM恒定反向引物序列  
<400> 32  
cgacggggaa ttctcacagg ag 22  
<210> 33  
<211> 22  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgD恒定反向引物序列  
<400> 33  
tgtctgcacc ctgatatgtat gg 22  
<210> 34  
<211> 19  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgA恒定1反向引物序列  
<400> 34  
gggtgtcgca gaggctcag 19  
<210> 35  
<211> 19  
<212> DNA  
<213> 人工序列

<220>  
<223> IgA恒定2反向引物序列  
<400> 35  
gggtgctgtc gaggctcag 19  
<210> 36  
<211> 22  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgE恒定反向引物序列  
<400> 36  
ggaatgttt tgcagcagcg gg 22  
<210> 37  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> TRA恒定反向引物序列  
<400> 37  
agtctctcag ctggtagacacg g 21  
<210> 38  
<211> 22  
<212> DNA  
<213> 人工序列  
<220>  
<223> TRB恒定反向引物序列  
<400> 38  
atggctcaaa cacagcgacc tc 22  
<210> 39  
<211> 24  
<212> DNA  
<213> 人工序列  
<220>  
<223> C7  
<400> 39  
caagcagaag acggcatacg agat 24  
<210> 40  
<211> 38  
<212> DNA

<213> 人工序列  
<220>  
<223> IgG衔接子标记的靶特异性反向引物序列  
<400> 40  
acacgacgct cttccgatct ccaggggaa gacsgatg 38  
<210> 41  
<211> 39  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgG衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21  
<223> n = A、T、C或G  
<400> 41  
acacgacgct cttccgatct nccagggga agacsgatg 39  
<210> 42  
<211> 40  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgG衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22  
<223> n = A、T、C或G  
<400> 42  
acacgacgct cttccgatct nnccagggg aagacsgatg 40  
<210> 43  
<211> 41  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgG衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22, 23  
<223> n = A、T、C或G

<400> 43  
acacgacgct cttccgatct nnncagggg gaagacsgat g 41  
<210> 44  
<211> 40  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgA衔接子标记的靶特异性反向引物序列  
<400> 44  
acacgacgct cttccgatct ggctcagcgg gaagacctg 40  
<210> 45  
<211> 41  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgA衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21  
<223> n = A、T、C或G  
<400> 45  
acacgacgct cttccgatct nggctcagcg ggaagacctt g 41  
<210> 46  
<211> 42  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgA衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22  
<223> n = A、T、C或G  
<400> 46  
acacgacgct cttccgatct nnggctcagc ggaaagacct tg 42  
<210> 47  
<211> 43  
<212> DNA  
<213> 人工序列  
<220>

<223> IgA衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22, 23  
<223> n = A、T、C或G  
<400> 47  
acacgacgct cttccgatct nnnggctcag cggaaagacc ttg 43  
<210> 48  
<211> 40  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgE衔接子标记的靶特异性反向引物序列  
<400> 48  
acacgacgct cttccgatct ggaaagacgg atgggctctg 40  
<210> 49  
<211> 41  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgE衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21  
<223> n = A、T、C或G  
<400> 49  
acacgacgct cttccgatct nggaaagacg gatgggctct g 41  
<210> 50  
<211> 42  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgE衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22  
<223> n = A、T、C或G  
<400> 50  
acacgacgct cttccgatct nnggaaagac ggatgggctc tg 42

<210> 51  
<211> 43  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgE衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22, 23  
<223> n = A、T、C或G  
<400> 51  
acacgacgct cttccgatct nnnggaaga cggatggct ctg 43  
<210> 52  
<211> 42  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgM衔接子标记的靶特异性反向引物序列  
<400> 52  
acacgacgct cttccgatct gagacgaggt ggaaaagggt tg 42  
<210> 53  
<211> 43  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgM衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21  
<223> n = A、T、C或G  
<400> 53  
acacgacgct cttccgatct ngagacgagg tggaaaaggg ttg 43  
<210> 54  
<211> 44  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgM衔接子标记的靶特异性反向引物序列  
<220>

<221> 尚未归类的特征  
<222> 21, 22  
<223> n = A、T、C或G  
<400> 54  
acacgacgct cttccgatct nngagacgag gtggaaaagg gttg 44  
<210> 55  
<211> 45  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgM衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22, 23  
<223> n = A、T、C或G  
<400> 55  
acacgacgct cttccgatct nnngagacga ggtggaaaag gttg 45  
<210> 56  
<211> 40  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgD衔接子标记的靶特异性反向引物序列  
<400> 56  
acacgacgct cttccgatct ggaacacatc cggagcctg 40  
<210> 57  
<211> 41  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgD衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21  
<223> n = A、T、C或G  
<400> 57  
acacgacgct cttccgatct nggaacacat ccggagcctt g 41  
<210> 58  
<211> 42

<212> DNA  
<213> 人工序列  
<220>  
<223> IgD衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22  
<223> n = A、T、C或G  
<400> 58  
acacgacgct cttccgatct nnggaacaca tccggagcct tg 42  
<210> 59  
<211> 43  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgD衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22, 23  
<223> n = A、T、C或G  
<400> 59  
acacgacgct cttccgatct nnnggaacac atccggagcc ttg 43  
<210> 60  
<211> 39  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgL衔接子标记的靶特异性反向引物序列  
<400> 60  
acacgacgct cttccgatct agggyggaa cagagtgac 39  
<210> 61  
<211> 40  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgL衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21

<223> n = A、T、C或G  
<400> 61  
acacgacgct cttccgatct nagggyggga acagagtgac 40  
<210> 62  
<211> 41  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgL衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22  
<223> n = A、T、C或G  
<400> 62  
acacgacgct cttccgatct nnagggygg aacagagtga c 41  
<210> 63  
<211> 42  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgL衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22, 23  
<223> n = A、T、C或G  
<400> 63  
acacgacgct cttccgatct nnnagggygg gaacagagtg ac 42  
<210> 64  
<211> 40  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgK衔接子标记的靶特异性反向引物序列  
<400> 64  
acacgacgct cttccgatct gacagatggc gcagccacag 40  
<210> 65  
<211> 41  
<212> DNA  
<213> 人工序列

<220>  
<223> IgK衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21  
<223> n = A、T、C或G  
<400> 65  
acacgacgct cttccgatct ngacagatgg tgcagccaca g 41  
<210> 66  
<211> 42  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgK衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22  
<223> n = A、T、C或G  
<400> 66  
acacgacgct cttccgatct nngacagatg gtgcagccac ag 42  
<210> 67  
<211> 43  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgK衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22, 23  
<223> n = A、T、C或G  
<400> 67  
acacgacgct cttccgatct nnngacagat ggtgcagcca cag 43  
<210> 68  
<211> 39  
<212> DNA  
<213> 人工序列  
<220>  
<223> TRA衔接子标记的靶特异性反向引物序列  
<400> 68

acacgacgct cttccgatct cacggcaggg tcagggttc 39  
<210> 69

<211> 40

<212> DNA

<213> 人工序列

<220>

<223> TRA衔接子标记的靶特异性反向引物序列

<220>

<221> 尚未归类的特征

<222> 21

<223> n = A、T、C或G

<400> 69

acacgacgct cttccgatct ncacggcagg gtcagggttc 40

<210> 70

<211> 41

<212> DNA

<213> 人工序列

<220>

<223> TRA衔接子标记的靶特异性反向引物序列

<220>

<221> 尚未归类的特征

<222> 21, 22

<223> n = A、T、C或G

<400> 70

acacgacgct cttccgatct nncacggcag ggtcagggtt c 41

<210> 71

<211> 42

<212> DNA

<213> 人工序列

<220>

<223> TRA衔接子标记的靶特异性反向引物序列

<220>

<221> 尚未归类的特征

<222> 21, 22, 23

<223> n = A、T、C或G

<400> 71

acacgacgct cttccgatct nnncacggca gggtcagggc tc 42

<210> 72

<211> 39

<212> DNA  
<213> 人工序列  
<220>  
<223> TRB衔接子标记的靶特异性反向引物序列  
<400> 72  
acacgacgct cttccgatct cgacacctggg tggaaacac 39  
<210> 73  
<211> 40  
<212> DNA  
<213> 人工序列  
<220>  
<223> TRB衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21  
<223> n = A、T、C或G  
<400> 73  
acacgacgct cttccgatct ncgacacctgg gtggaaacac 40  
<210> 74  
<211> 41  
<212> DNA  
<213> 人工序列  
<220>  
<223> TRB衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22  
<223> n = A、T、C或G  
<400> 74  
acacgacgct cttccgatct nncgacacctcg ggtggaaaca c 41  
<210> 75  
<211> 42  
<212> DNA  
<213> 人工序列  
<220>  
<223> TRB衔接子标记的靶特异性反向引物序列  
<220>  
<221> 尚未归类的特征  
<222> 21, 22, 23

<223> n = A、T、C或G

<400> 75

acacgacgct cttccgatct nnncgacctc gggtgggAAC ac 42

<210> 76

<211> 53

<212> DNA

<213> 人工序列

<220>

<223> C5-P5反向引物

<400> 76

aatgatacgg cgaccacgga gatctacact cttccctac acgacgctct tcc 53

<210> 77

<211> 29

<212> DNA

<213> 人工序列

<220>

<223> P7引发位点(C7)

<400> 77

agatcggaaag agcacacgta tgaactcca 29

<210> 78

<211> 58

<212> DNA

<213> 人工序列

<220>

<223> P5引发位点

<400> 78

agatcggaaag agcgtcggtgt agggaaagag tgttagatctc ggtggtcgcc gtatcatt 58

<210> 79

<211> 26

<212> DNA

<213> 人工序列

<220>

<223> 通用衔接子

<220>

<221> 尚未归类的特征

<222> 21, 22, 23, 24, 25, 26

<223> n = A、T、C或G

<400> 79

acacgacgct cttccgatct nnnnnn 26

- <210> 80  
<211> 14  
<212> DNA  
<213> 人工序列  
<220>  
<223> 通用衔接子  
<220>  
<221> 尚未归类的特征  
<222> 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14  
<223> n = A、T、C或G  
<400> 80  
nnnnwnnnnw nnnn 14  
<210> 81  
<211> 15  
<212> DNA  
<213> 人工序列  
<220>  
<223> 通用衔接子  
<220>  
<221> 尚未归类的特征  
<222> 2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, 15  
<223> n = A、T、C或G  
<400> 81  
wnnnnwnnnn wnnnn 15  
<210> 82  
<211> 15  
<212> DNA  
<213> 人工序列  
<220>  
<223> 通用衔接子  
<220>  
<221> 尚未归类的特征  
<222> 1, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, 15  
<223> n = A、T、C或G  
<400> 82  
nwnnnwnnnn wnnnn 15  
<210> 83  
<211> 17  
<212> DNA

<213> 人工序列  
<220>  
<223> 通用衔接子  
<220>  
<221> 尚未归类的特征  
<222> 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17  
<223> n = A、T、C或G  
<400> 83  
nnwnnnnnwnnn nnwnnnnn 17  
<210> 84  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgM-RT引物  
<220>  
<221> 经修饰的碱基  
<222> 1  
<223> 生物素化胸腺嘧啶  
<400> 84  
tgtgaggtgg ctgcgtactt g 21  
<210> 85  
<211> 19  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgG-RT引物  
<220>  
<221> 经修饰的碱基  
<222> 1  
<223> 生物素化腺嘌呤  
<400> 85  
aggacagccg ggaagggtgt 19  
<210> 86  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgD-RT引物

<220>  
<221> 经修饰的碱基  
<222> 1  
<223> 生物素化胞嘧啶  
<400> 86  
cacgcattt tactgcctt g 21  
<210> 87  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgA-RT引物  
<220>  
<221> 经修饰的碱基  
<222> 1  
<223> 生物素化胞嘧啶  
<400> 87  
ctggcttgtt ggaaagttc t 21  
<210> 88  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgE-RT引物  
<220>  
<221> 经修饰的碱基  
<222> 1  
<223> 生物素化鸟嘌呤  
<400> 88  
ggggcatag tgaccagaga 20  
<210> 89  
<211> 22  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgK-RT引物  
<220>  
<221> 经修饰的碱基  
<222> 1

<223> 生物素化胸腺嘧啶  
<400> 89  
tattcagcag gcacacaaca ga 22  
<210> 90  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> IgL-RT引物  
<220>  
<221> 经修饰的碱基  
<222> 1  
<223> 生物素化腺嘌呤  
<400> 90  
agtgtggcct tttggcttg 20  
<210> 91  
<211> 21  
<212> DNA  
<213> 人工序列  
<220>  
<223> TCR-A-RT引物  
<220>  
<221> 经修饰的碱基  
<222> 1  
<223> 生物素化鸟嘌呤  
<400> 91  
gggagatctc tgcttctgat g 21  
<210> 92  
<211> 22  
<212> DNA  
<213> 人工序列  
<220>  
<223> TCR-B-RT引物  
<220>  
<221> 经修饰的碱基  
<222> 1  
<223> 生物素化鸟嘌呤  
<400> 92  
ggtaatagg cagacagact tg 22

<210> 93  
<211> 19  
<212> DNA  
<213> 人工序列  
<220>  
<223> CD4-RT引物  
<220>  
<221> 经修饰的碱基  
<222> 1  
<223> 生物素化鸟嘌呤  
<400> 93  
ggcagtcatttccggaaacact 19  
<210> 94  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> CD-8-RT引物  
<220>  
<221> 经修饰的碱基  
<222> 1  
<223> 生物素化胞嘧啶  
<400> 94  
ctacaaagtg ggcccttctg 20  
<210> 95  
<211> 40  
<212> DNA  
<213> 人工序列  
<220>  
<223> CD4-巢式引物  
<400> 95  
acacgacgct cttccgatct tgtggccttg ccgagggagg 40  
<210> 96  
<211> 40  
<212> DNA  
<213> 人工序列  
<220>  
<223> CD8-巢式引物  
<400> 96

acacgacgct cttccgatct tgcggaatcc cagagggcca 40  
<210> 97  
<211> 64  
<212> DNA  
<213> 人工序列  
<220>  
<223> C7-bc-P7引物  
<220>  
<221> 尚未归类的特征  
<222> 25, 26, 27, 28, 29, 30  
<223> n = A、T、C或G  
<400> 97  
caagcagaag acggcatacg agatnnnnn gtgactggag ttcagacgtg tgctcttccg 60  
atct 64  
<210> 98  
<211> 58  
<212> DNA  
<213> 人工序列  
<220>  
<223> C5-P5引物  
<400> 98  
aatgatacgg cgaccaccga gatctacact cttccctac acgacgctt tccgatct 58  
<210> 99  
<211> 20  
<212> DNA  
<213> 人工序列  
<220>  
<223> 容器条码序列  
<220>  
<221> 尚未归类的特征  
<222> 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19  
<223> n = A、T、C或G  
<400> 99  
nnnnwnnnnw nnnnwnnnnw 20  
<210> 100  
<211> 18  
<212> DNA  
<213> 人工序列  
<220>

〈223〉 容器条码序列

〈220〉

〈221〉 尚未归类的特征

〈222〉 1, 2, 3, 4, 6, 9, 10, 11, 13, 16, 17, 18

〈223〉 n = A、T、C或G

〈400〉 100

nnnnwnscnn nwncnnn 18

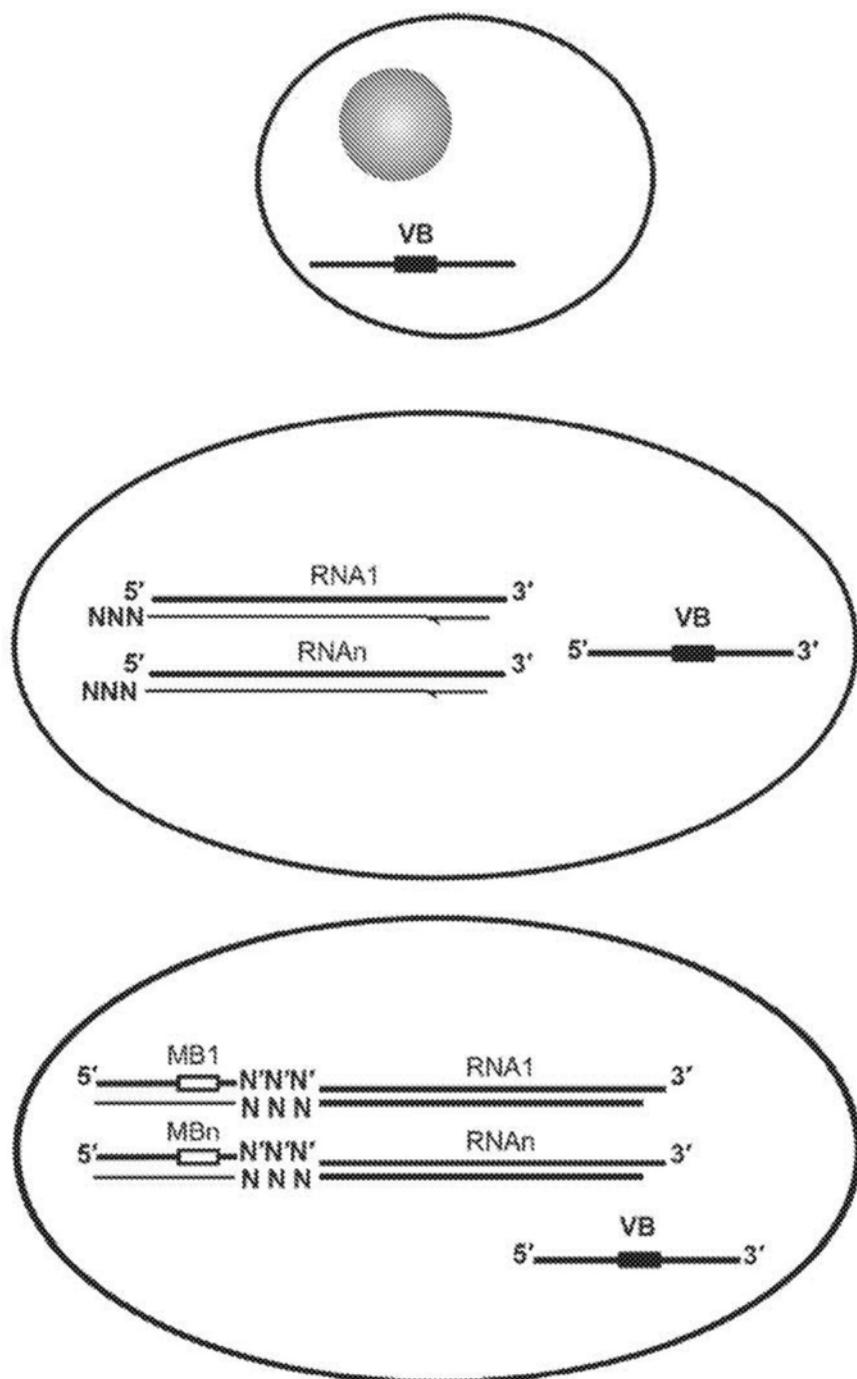


图1A

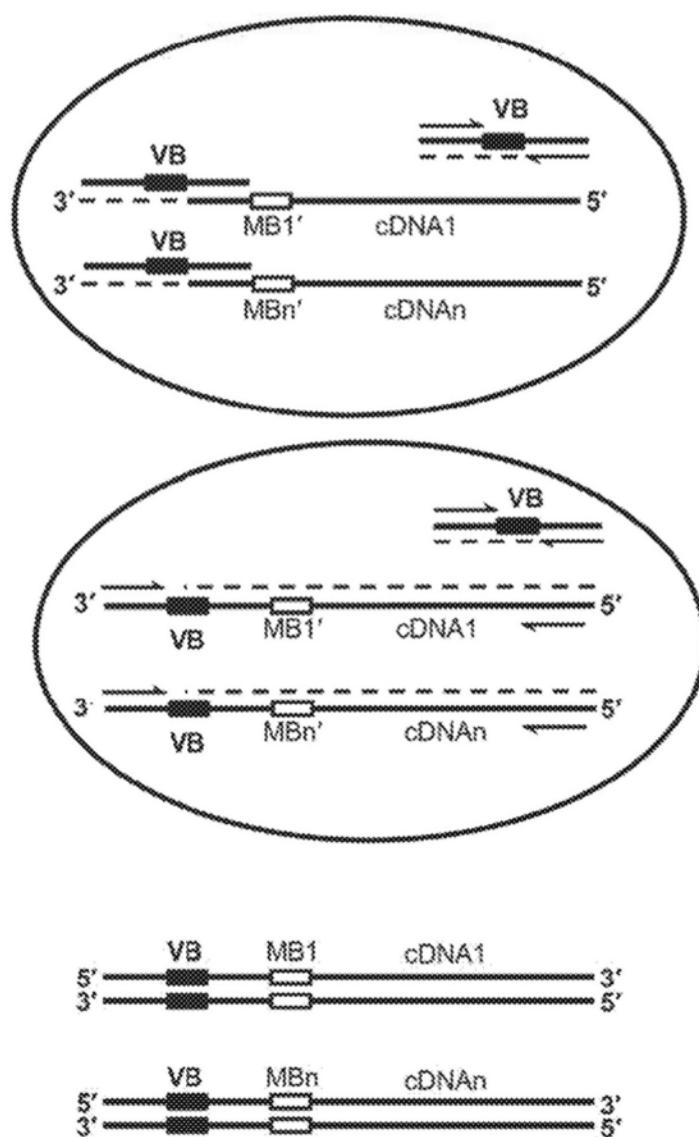


图1B

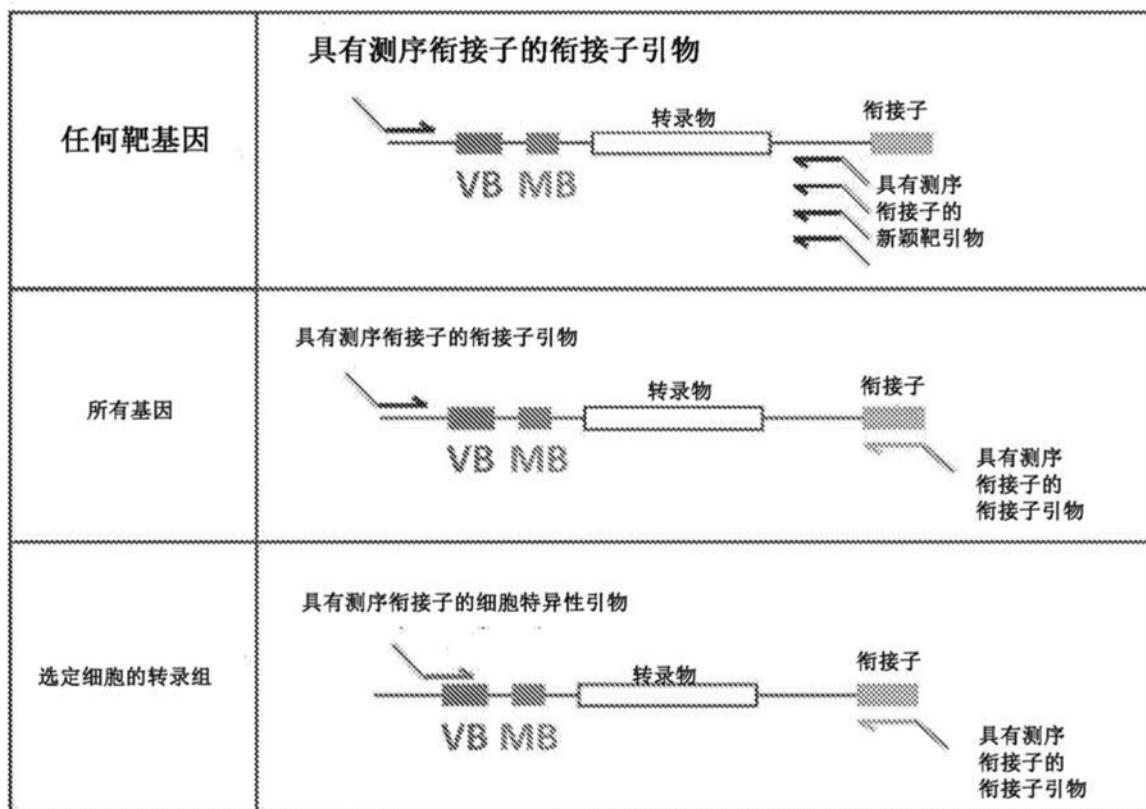


图2

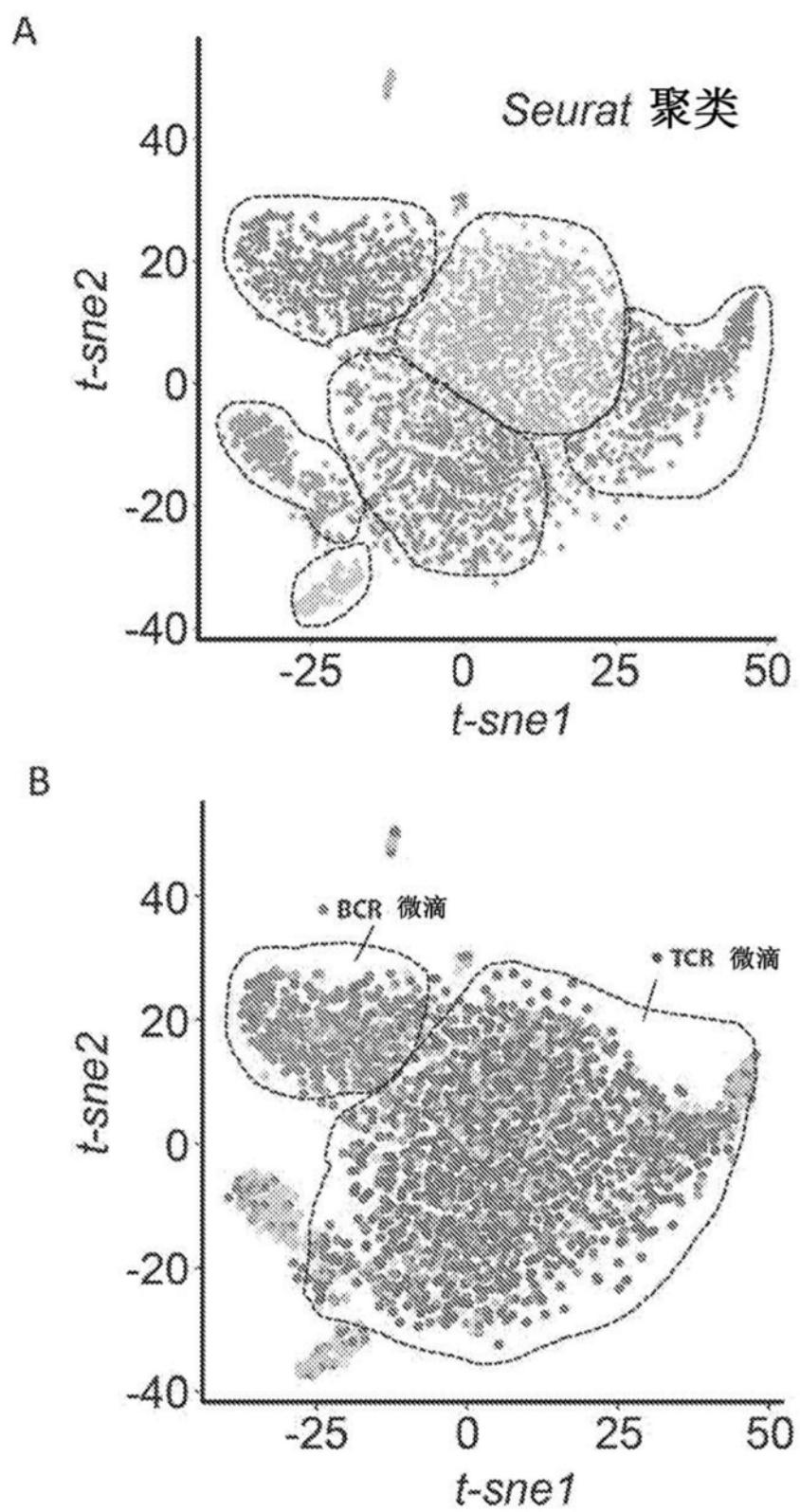


图3

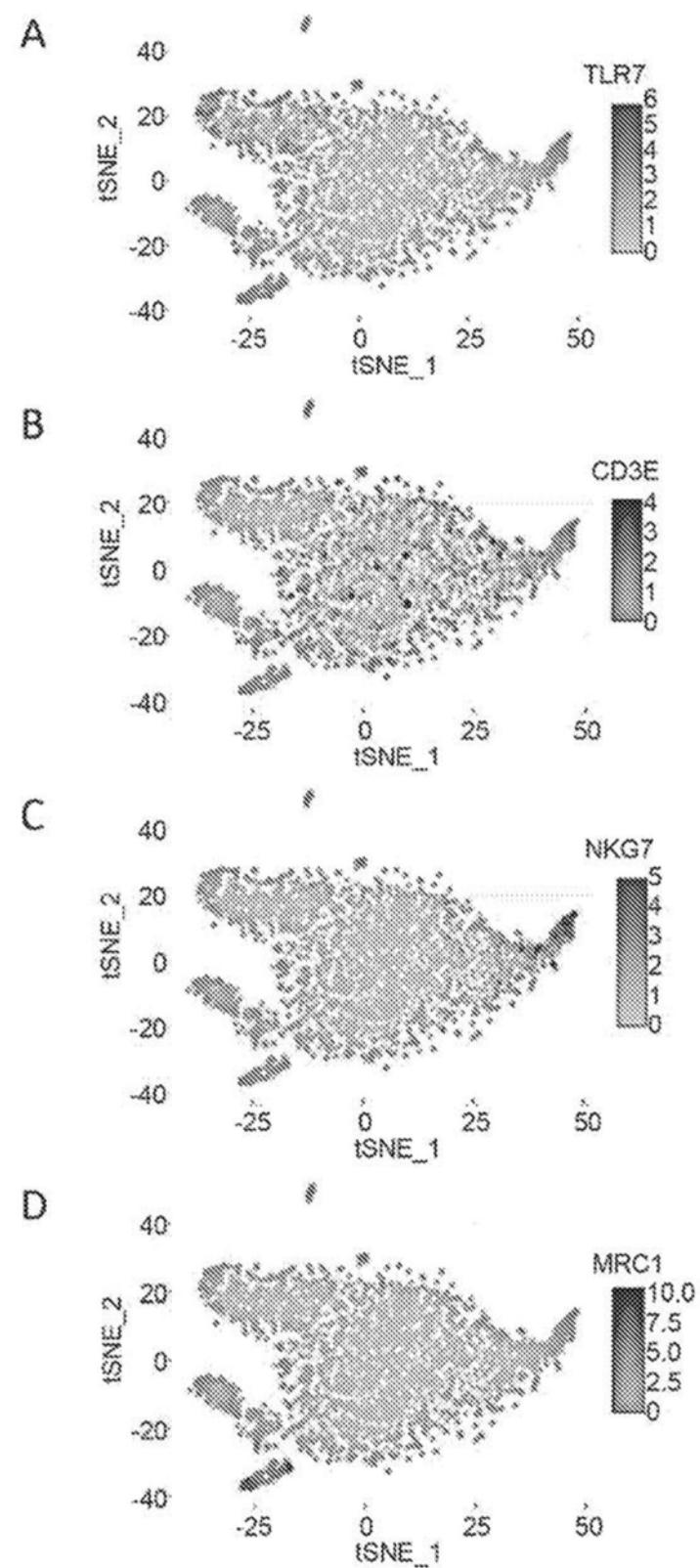


图4