



(12) 发明专利

(10) 授权公告号 CN 107002076 B

(45) 授权公告日 2021. 11. 23

(21) 申请号 201580062006.4

(22) 申请日 2015.09.15

(65) 同一申请的已公布的文献号
申请公布号 CN 107002076 A

(43) 申请公布日 2017.08.01

(30) 优先权数据
62/050,549 2014.09.15 US
62/051,832 2014.09.17 US

(85) PCT国际申请进入国家阶段日
2017.05.15

(86) PCT国际申请的申请数据
PCT/US2015/050119 2015.09.15

(87) PCT国际申请的公布数据
W02016/044227 EN 2016.03.24

(73) 专利权人 阿布维特罗有限责任公司
地址 美国马萨诸塞州

(72) 发明人 弗朗索瓦·维格纳尔特
阿德里安·阮格汉姆布里格斯
克里斯多佛·赖安·克劳泽
斯蒂芬·雅各布·戈德弗莱斯
索尼娅·汀布莱克

(74) 专利代理机构 北京安信方达知识产权代理有限公司 11262

代理人 郑霞

(51) Int.Cl.
C12N 15/10 (2006.01)
C12Q 1/6874 (2018.01)

(56) 对比文件
CN 103732738 A, 2014.04.16

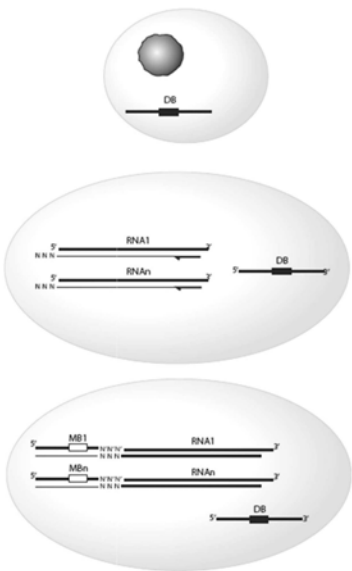
审查员 熊壮壮

权利要求书32页 说明书83页
序列表17页 附图12页

(54) 发明名称
高通量核苷酸文库测序

(57) 摘要

本文提供了用于免疫组库测序和单细胞条码化的方法和组合物。该方法和组合物可用于匹配来源于单细胞任何两个序列,诸如重链和轻链抗体序列, α 和 β 链T细胞受体序列,或 γ 和 δ 链T细胞受体序列,以供抗体和T细胞受体发现、疾病和免疫诊断以及低错误测序。



1. 一种非诊断目的的方法,其包括:

(a) 形成多个容器,每个容器包含

(i) 来自包含多个细胞的样品的单细胞,

(ii) 多个分子条码化多核苷酸,以及

(iii) 容器条码化多核苷酸;

(b) 产生:

(i) 第一互补多核苷酸,其与来自所述单细胞的第一细胞多核苷酸互补,以及

(ii) 第二互补多核苷酸,其与来自所述单细胞的第二细胞多核苷酸互补;

(c) 衔接:

(i) 所述多个分子条码化多核苷酸的第一分子条码化多核苷酸至所述第一互补多核苷酸,以及

(ii) 所述多个分子条码化多核苷酸的第二分子条码化多核苷酸至所述第二互补多核苷酸,

从而形成第一和第二单细胞单条码化多核苷酸;以及

(d) 将所述容器条码化多核苷酸或其扩增产物衔接至

(i) 所述第一单细胞单条码化多核苷酸或其扩增产物,以及

(ii) 所述第二单细胞单条码化多核苷酸或其扩增产物,

从而形成第一和第二单细胞双条码化多核苷酸。

2. 一种非诊断目的的方法,其包括:

(a) 从来自样品的多个免疫细胞产生来自重链免疫球蛋白 (IgH) 多核苷酸的第一互补多核苷酸和来自轻链免疫球蛋白 (IgL) 多核苷酸的第二互补多核苷酸,其中采用:

(i) 第一目标引物,其包含与来自所述多个免疫细胞的所述IgH多核苷酸的相同区域互补的区域;

(ii) 第二目标引物,其包含与来自所述多个免疫细胞的所述IgL多核苷酸的相同区域互补的区域;

(iii) 包含非模板末端转移酶活性的逆转录酶,其中将3个或更多个相同的非模板核苷酸添加至所述第一和第二互补多核苷酸的3' 端;

(iv) 容器条码化多核苷酸;以及

(v) 多个分子条码化多核苷酸,每个包含:

(A) 分子条码,

(B) 与所述容器条码化多核苷酸的区域互补的5' 端区域,以及

(C) 与所述3个或更多个非模板核苷酸互补的3' 端区域;

从而形成第一和第二单细胞单条码化多核苷酸;

(b) 扩增所述容器条码化多核苷酸,从而形成第一和第二单细胞双条码化多核苷酸;

(c) 扩增所述第一和第二单细胞双条码化多核苷酸,从而形成包含所述IgH或IgL多核苷酸的可变区或其组合的序列的文库;以及

(d) 对所述文库的一个或多个所述序列进行测序,

其中(a) 在多个容器的容器中进行,其中所述容器包含来自所述多个免疫细胞的单个免疫细胞。

3. 一种非诊断目的的方法,其包括:

(a) 从来自样品的多个免疫细胞产生来自T细胞受体 α (TCR α) 多核苷酸的第一互补多核苷酸和来自T细胞受体 β (TCR β) 多核苷酸的第二互补多核苷酸,其中采用:

(i) 第一目标引物,其包含与来自所述多个免疫细胞的所述TCR α 多核苷酸的相同区域互补的区域;

(ii) 第二目标引物,其包含与来自所述多个免疫细胞的所述TCR β 多核苷酸的相同区域互补的区域;

(iii) 包含非模板末端转移酶活性的逆转录酶,其中将3个或更多个相同的非模板核苷酸添加至所述第一和第二互补多核苷酸的3' 端;

(iv) 容器条码化多核苷酸;以及

(v) 多个分子条码化多核苷酸,每个包含:

(A) 分子条码,

(B) 与所述容器条码化多核苷酸的区域互补的5' 端区域,以及

(C) 与所述3个或更多个非模板核苷酸互补的3' 端区域;

从而形成第一和第二单细胞单条码化多核苷酸;

(b) 扩增所述容器条码化多核苷酸,从而形成第一和第二单细胞双条码化多核苷酸;

(c) 扩增所述第一和第二单细胞双条码化多核苷酸,从而形成包含所述TCR α 或TCR β 多核苷酸的可变区或其组合的序列的文库;以及

(d) 对所述文库的一个或多个所述序列进行测序,

其中(a) 在多个容器的容器中进行,其中所述容器包含来自所述多个免疫细胞的单个免疫细胞。

4. 一种非诊断目的的方法,其包括:

(a) 从来自样品的多个免疫细胞产生来自T细胞受体 γ (TCR γ) 多核苷酸的第一互补多核苷酸和来自T细胞受体 δ (TCR δ) 多核苷酸的第二互补多核苷酸,其中采用:

(i) 第一目标引物,其包含与来自所述多个免疫细胞的所述TCR γ 多核苷酸的相同区域互补的区域;

(ii) 第二目标引物,其包含与来自所述多个免疫细胞的所述TCR δ 多核苷酸的相同区域互补的区域;

(iii) 包含非模板末端转移酶活性的逆转录酶,其中将3个或更多个相同的非模板核苷酸添加至所述第一和第二互补多核苷酸的3' 端;

(iv) 容器条码化多核苷酸;以及

(v) 多个分子条码化多核苷酸,每个包含:

(A) 分子条码,

(B) 与所述容器条码化多核苷酸的区域互补的5' 端区域,以及

(C) 与所述3个或更多个非模板核苷酸互补的3' 端区域;

从而形成第一和第二单细胞单条码化多核苷酸;

(b) 扩增所述容器条码化多核苷酸,从而形成第一和第二单细胞双条码化多核苷酸;

(c) 扩增所述第一和第二单细胞双条码化多核苷酸,从而形成包含所述TCR γ 或TCR δ 多核苷酸的可变区或其组合的序列的文库;以及

(d) 对所述文库的一个或多个所述序列进行测序，

其中 (a) 在多个容器的容器中进行，其中所述容器包含来自所述多个免疫细胞的单个免疫细胞。

5. 根据权利要求2-4中任一项所述的方法，其中所述文库代表所述样品的免疫状态。

6. 根据权利要求1-4中任一项所述的方法，其中所述第一和第二单细胞双条码化多核苷酸为第一和第二单细胞双条码化多核苷酸的文库。

7. 根据权利要求1所述的方法，其中所述第一和第二分子条码化多核苷酸的分子条码不同。

8. 根据权利要求1-4中任一项所述的方法，其中所述第一和第二单细胞单条码化多核苷酸包含不同的分子条码。

9. 根据权利要求1-4中任一项所述的方法，其中所述第一和第二单细胞双条码化多核苷酸包含不同的分子条码。

10. 根据权利要求1-4中任一项所述的方法，其中所述第一和第二单细胞双条码化多核苷酸包含相同的容器条码。

11. 根据权利要求1-4中任一项所述的方法，其中所述多个分子条码化多核苷酸不是扩增产物。

12. 根据权利要求2-4中任一项所述的方法，其中第一容器中分子条码化多核苷酸的分子条码与第二容器中分子条码化多核苷酸的分子条码不同。

13. 根据权利要求2-4中任一项所述的方法，其中所述多个容器的第一容器中每个分子条码化多核苷酸的分子条码是独特的。

14. 根据权利要求13所述的方法，其中所述多个容器的第二容器中每个分子条码化多核苷酸的分子条码是独特的。

15. 根据权利要求2-4中任一项所述的方法，其中第一容器和第二容器中每个分子条码化多核苷酸的分子条码是独特的。

16. 根据权利要求14所述的方法，其中所述多个容器的第三容器中每个分子条码化多核苷酸的分子条码是独特的。

17. 根据权利要求16所述的方法，其中所述第一容器、所述第二容器和所述第三容器中每个分子条码化多核苷酸的分子条码是独特的。

18. 根据权利要求2-4中任一项所述的方法，其中所述多个容器的任何单个容器中每个分子条码化多核苷酸的分子条码是独特的。

19. 根据权利要求2-4中任一项所述的方法，其中所述多个容器的任何一个容器中每个分子条码化多核苷酸的分子条码与所述多个容器的任何其他一个容器中每个分子条码化多核苷酸的分子条码不同。

20. 根据权利要求2-4中任一项所述的方法，其中第一容器中分子条码化多核苷酸的分子条码与第二容器中分子条码化多核苷酸的分子条码相同。

21. 根据权利要求2-4中任一项所述的方法，其中第一容器中分子条码化多核苷酸的分子条码与所述第一容器中分子条码化多核苷酸的分子条码相同。

22. 根据权利要求21所述的方法，其中第二容器中分子条码化多核苷酸的分子条码与所述第二容器中分子条码化多核苷酸的分子条码相同。

23. 根据权利要求1-4中任一项所述的方法, 其中所述多个容器的第一容器中容器条码化多核苷酸或其扩增子的容器条码与所述多个容器的第二容器中容器条码化多核苷酸或其扩增子的容器条码不同。

24. 根据权利要求1-4中任一项所述的方法, 其中所述多个容器的第一容器中容器条码化多核苷酸或其扩增子的容器条码为第一相同容器条码。

25. 根据权利要求24所述的方法, 其中所述多个容器的第二容器中每个容器条码化多核苷酸或其扩增子的容器条码为第二相同容器条码。

26. 根据权利要求25所述的方法, 其中所述第一相同容器条码与所述第二相同容器条码不同。

27. 根据权利要求1-4中任一项所述的方法, 其中所述多个容器的单个容器中每个容器条码化多核苷酸或其扩增子的容器条码包含相同的容器条码。

28. 根据权利要求1-4中任一项所述的方法, 其中所述多个容器的任何单个容器中每个容器条码化多核苷酸及其扩增子的容器条码对于所述多个容器的任何其他单个容器中每个容器条码化多核苷酸及其扩增子的容器条码而言是独特的。

29. 根据权利要求1-4中任一项所述的方法, 其中 (a) 中的容器条码化多核苷酸以单分子存在于容器中。

30. 根据权利要求1-4中任一项所述的方法, 其中 (a) 中的容器条码化多核苷酸以单分子存在于所述多个容器的每个容器中。

31. 根据权利要求1-4中任一项所述的方法, 其中 (a) 中的容器条码化多核苷酸以至少单分子存在于所述多个容器的容器中。

32. 根据权利要求1-4中任一项所述的方法, 其中 (a) 中的容器条码化多核苷酸以至少单分子存在于所述多个容器的每个容器中。

33. 根据权利要求1-4中任一项所述的方法, 其中所述多个容器的第一容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述第一容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。

34. 根据权利要求33所述的方法, 其中所述多个容器的所述第一容器中所述第一容器条码化多核苷酸或其扩增子的第二共同容器序列与所述第一容器中第二容器条码化多核苷酸或其扩增子的第二共同容器序列相同。

35. 根据权利要求33所述的方法, 其中所述多个容器的任何单个容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述单个容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。

36. 根据权利要求33所述的方法, 其中所述多个容器的单个容器中每个容器条码化多核苷酸包含相同的第一共同容器序列。

37. 根据权利要求36所述的方法, 其中所述多个容器的单个容器中每个容器条码化多核苷酸包含相同的第二共同容器序列。

38. 根据权利要求1-4中任一项所述的方法, 其中所述多个容器的第一容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述多个容器的第二容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。

39. 根据权利要求38所述的方法, 其中所述第一容器条码化多核苷酸或其扩增子的第

二共同容器序列与所述第二容器条码化多核苷酸或其扩增子的第二共同容器序列相同。

40. 根据权利要求1-4中任一项所述的方法, 其中所述多个容器的任何一个容器中每个容器条码化多核苷酸或其扩增子包含第一共同容器序列, 该第一共同容器序列包含与所述多个容器的任何其他一个容器中容器条码化多核苷酸或其扩增子的第一共同容器序列相同的序列。

41. 根据权利要求40所述的方法, 其中所述多个容器的任何一个容器中每个容器条码化多核苷酸或其扩增子包含第二共同容器序列, 该第二共同容器序列包含与所述多个容器的任何其他一个容器中容器条码化多核苷酸或其扩增子的第二共同容器序列相同的序列。

42. 根据权利要求1-4中任一项所述的方法, 其中所述多个容器的第一容器中第一分子条码化多核苷酸的第一共同分子序列与所述第一容器中第二分子条码化多核苷酸的第一共同分子序列相同。

43. 根据权利要求42所述的方法, 其中所述多个容器的所述第一容器中所述第一分子条码化多核苷酸的第二共同分子序列与所述第一容器中第二分子条码化多核苷酸的第二共同分子序列相同。

44. 根据权利要求42所述的方法, 其中所述多个容器的任何单个容器中第一分子条码化多核苷酸的第一共同分子序列与所述单个容器中第二分子条码化多核苷酸的第一共同分子序列相同。

45. 根据权利要求42所述的方法, 其中所述多个容器的单个容器中每个分子条码化多核苷酸包含相同的第一共同分子序列。

46. 根据权利要求45所述的方法, 其中所述多个容器的单个容器中每个分子条码化多核苷酸包含相同的第二共同分子序列。

47. 根据权利要求42所述的方法, 其中所述多个容器的第一容器中第一分子条码化多核苷酸的第一共同分子序列与所述多个容器的第二容器中第二分子条码化多核苷酸的第一共同分子序列相同。

48. 根据权利要求47所述的方法, 其中所述第一分子条码化多核苷酸的第二共同分子序列与所述第二分子条码化多核苷酸的第二共同分子序列相同。

49. 根据权利要求42所述的方法, 其中所述多个容器的任何一个容器中每个分子条码化多核苷酸包含第一共同分子序列, 该第一共同分子序列包含与所述多个容器的任何其他一个容器中分子条码化多核苷酸的第一共同分子序列相同的序列。

50. 根据权利要求49所述的方法, 其中所述多个容器的任何一个容器中每个分子条码化多核苷酸包含第二共同分子序列, 该第二共同分子序列包含与所述多个容器的任何其他一个容器中分子条码化多核苷酸的第二共同分子序列相同的序列。

51. 根据权利要求33所述的方法, 其中所述第一共同容器序列包含的序列包含与所述第一共同分子序列相同的序列。

52. 根据权利要求33所述的方法, 其中所述第一共同容器序列包含与所述第一共同分子序列或其互补体互补的序列。

53. 根据权利要求50所述的方法, 其中所述第二共同分子序列包含与添加至所述第一互补多核苷酸的3'端的三个或更多个非模板核苷酸互补的区域。

54. 根据权利要求53所述的方法, 其中所述与添加至所述第一互补多核苷酸的3'端的

三个或更多个非模板核苷酸互补的区域为末端区域。

55. 根据权利要求1-4中任一项所述的方法, 其中所述第一和第二单细胞单条码化多核苷酸不融合在一起。

56. 根据权利要求1-4中任一项所述的方法, 其中所述第一和第二单细胞双条码化多核苷酸不融合在一起。

57. 根据权利要求1所述的方法, 其中所述第一细胞多核苷酸为DNA。

58. 根据权利要求57所述的方法, 其中所述第二细胞多核苷酸为DNA。

59. 根据权利要求1所述的方法, 其中所述第一细胞多核苷酸为RNA。

60. 根据权利要求59所述的方法, 其中所述第二细胞多核苷酸为RNA。

61. 根据权利要求59或60所述的方法, 其中所述RNA为mRNA。

62. 根据权利要求1-4中任一项所述的方法, 其中所述第一互补多核苷酸为cDNA。

63. 根据权利要求62所述的方法, 其中所述第二互补多核苷酸为cDNA。

64. 根据权利要求1所述的方法, 其中 (b) 包括延伸与所述第一细胞多核苷酸杂交的第一目标引物, 以及延伸与所述第二细胞多核苷酸杂交的第二目标引物。

65. 根据权利要求64所述的方法, 其中所述延伸包括用所述第一目标引物逆转录所述第一细胞多核苷酸, 以及用所述第二目标引物逆转录所述第二细胞多核苷酸。

66. 根据权利要求64所述的方法, 其中所述第一目标引物包含与所述第一细胞多核苷酸的目标序列互补的序列。

67. 根据权利要求66所述的方法, 其中所述第二目标引物包含与所述第二细胞多核苷酸的目标序列互补的序列。

68. 根据权利要求2-4和64中任一项所述的方法, 其中所述第一目标引物包含聚(T)序列。

69. 根据权利要求68所述的方法, 其中所述第二目标引物包含聚(T)序列。

70. 根据权利要求66或67所述的方法, 其中所述第一细胞多核苷酸的所述目标序列为重链免疫球蛋白(IgH)序列、TCR α 序列、TCR γ 序列或其组合。

71. 根据权利要求66或67所述的方法, 其中所述第一细胞多核苷酸的所述目标序列为重链恒定区(C_H)序列、TCR α 恒定区(C α)序列、TCR γ 恒定区(C γ)序列或其组合。

72. 根据权利要求67所述的方法, 其中所述第二细胞多核苷酸的所述目标序列为轻链免疫球蛋白(IgL)序列、TCR β 序列、TCR δ 序列或其组合。

73. 根据权利要求67所述的方法, 其中所述第二细胞多核苷酸的所述目标序列为轻链恒定区(C_L)序列、TCR β 恒定区(C β)序列、TCR δ 恒定区(C δ)序列或其组合。

74. 根据权利要求2-4和64中任一项所述的方法, 其中所述第一目标引物包含多个第一目标引物。

75. 根据权利要求2-4和64中任一项所述的方法, 其中所述第二目标引物包含多个第二目标引物。

76. 根据权利要求74所述的方法, 其中所述多个第一目标引物包含与多个重链免疫球蛋白(IgH)序列、TCR α 序列、TCR γ 序列或其组合互补的多个序列。

77. 根据权利要求76所述的方法, 其中所述多个重链免疫球蛋白(IgH)序列、TCR α 序列或TCR γ 序列包含多个重链恒定区(C_H)序列、TCR α 恒定区(C α)序列、TCR γ 恒定区(C γ)序列

或其组合。

78. 根据权利要求77所述的方法, 其中所述多个重链恒定区 (C_H) 序列包含两个或更多个选自来自IgM、IgD、IgA、IgE、IgG及其组合的重链恒定区 (C_H) 序列的序列。

79. 根据权利要求75所述的方法, 其中所述多个第二目标引物包含与多个轻链免疫球蛋白 (IgL) 序列、TCR β 序列、TCR δ 序列或其组合互补的多个序列。

80. 根据权利要求79所述的方法, 其中所述多个轻链免疫球蛋白 (IgL) 序列、TCR β 序列或TCR δ 序列包含多个轻链恒定区 (C_L) 序列、TCR β 恒定区 ($C\beta$) 序列、TCR δ 恒定区 ($C\delta$) 序列或其组合。

81. 根据权利要求80所述的方法, 其中所述多个轻链恒定区 (C_L) 序列包含两个或更多个选自来自Ig κ 、Ig λ 及其组合的轻链恒定区 (C_L) 序列的序列。

82. 根据权利要求1-4中任一项所述的方法, 其中权利要求1的 (b) 或权利要求2-4之一的 (a) 包括通过使用非模板末端转移酶进行延伸, 其中将三个或更多个非模板核苷酸添加至所述第一互补多核苷酸的3' 端。

83. 根据权利要求82所述的方法, 其中所述非模板末端转移酶为逆转录酶或聚合酶。

84. 根据权利要求82所述的方法, 其中所述非模板末端转移酶为逆转录酶, 并且其中该逆转录酶选自SuperscriptII逆转录酶、Maxima逆转录酶、ProtoscriptII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。

85. 根据权利要求82所述的方法, 其中将三个或更多个非模板核苷酸添加至所述第二互补多核苷酸的3' 端。

86. 根据权利要求82所述的方法, 其中权利要求1的 (c) 中的所述附接或权利要求2-4之一的 (b) 中的所述扩增包括使第一分子条码化多核苷酸的区域与添加至所述第一互补多核苷酸的3' 端的所述三个或更多个非模板核苷酸杂交。

87. 根据权利要求82所述的方法, 其中在权利要求1的 (c) 中的所述附接或在权利要求2-4之一的 (b) 中的所述扩增包括使第二分子条码化多核苷酸的区域与添加至所述第二互补多核苷酸的3' 端的所述三个或更多个非模板核苷酸杂交。

88. 根据权利要求82所述的方法, 其中在权利要求1的 (c) 或权利要求2-4之一的 (a) 中, 附接至所述第一互补多核苷酸的第一分子条码化多核苷酸包含与所述第一互补多核苷酸的3' 端上的所述三个或更多个非模板核苷酸互补的区域。

89. 根据权利要求88所述的方法, 其中在权利要求1的 (c) 或权利要求2-4之一的 (a) 中, 附接至所述第二互补多核苷酸的第二分子条码化多核苷酸包含与所述第二互补多核苷酸的3' 端上的三个或更多个非模板核苷酸互补的区域。

90. 根据权利要求82所述的方法, 其中所述三个或更多个非模板核苷酸是相同的。

91. 根据权利要求82所述的方法, 其中所述三个或更多个非模板核苷酸中的至少一个与所述三个或更多个非模板核苷酸中的另一个核苷酸不相同。

92. 根据权利要求86所述的方法, 其中所述第一分子条码化多核苷酸的杂交区域的至少一个核苷酸与所述第一分子条码化多核苷酸的杂交区域的另一个核酸不相同。

93. 根据权利要求87所述的方法, 其中所述第二分子条码化多核苷酸的杂交区域的至少一个核苷酸与所述第二分子条码化多核苷酸的杂交区域的另一个核酸不相同。

94. 根据权利要求92所述的方法, 其中所述至少一个不相同的核苷酸为脱氧核糖核苷酸或其类似物。

95. 根据权利要求92所述的方法, 其中所述至少一个不相同的核苷酸不是核糖核苷酸或其类似物。

96. 根据权利要求92所述的方法, 其中所述至少一个不相同的核苷酸为脱氧核糖鸟苷。

97. 根据权利要求92所述的方法, 其中所述至少一个不相同的核苷酸为脱氧核糖鸟苷类似物。

98. 根据权利要求92所述的方法, 其中所述至少一个不相同的核苷酸为所述第一或第二分子条码化多核苷酸的末端核苷酸。

99. 根据权利要求92所述的方法, 其中所述至少一个不相同的核苷酸为核糖核苷酸或其类似物。

100. 根据权利要求86所述的方法, 其中所述第一分子条码化多核苷酸的杂交区域的末端核苷酸为脱氧核糖核苷酸或其类似物。

101. 根据权利要求86所述的方法, 其中所述第一分子条码化多核苷酸的杂交区域的末端核苷酸不是核糖核苷酸或其类似物。

102. 根据权利要求86所述的方法, 其中所述第一分子条码化多核苷酸的杂交区域的末端核苷酸为脱氧核糖鸟苷。

103. 根据权利要求86所述的方法, 其中所述第一分子条码化多核苷酸的杂交区域的末端核苷酸为脱氧核糖鸟苷类似物。

104. 根据权利要求86所述的方法, 其中所述第一分子条码化多核苷酸的杂交区域的末端核苷酸为核糖核苷酸或其类似物。

105. 根据权利要求86所述的方法, 其中所述第一分子条码化多核苷酸的杂交区域的至少两个非末端核苷酸为核糖核苷酸或其类似物。

106. 根据权利要求86所述的方法, 其中所述第一或第二分子条码化多核苷酸的杂交区域的至少两个非末端核苷酸不是脱氧核糖核苷酸或其类似物。

107. 根据权利要求86所述的方法, 其中所述第一或第二分子条码化多核苷酸的杂交区域的至少两个非末端核苷酸为脱氧核糖核苷酸或其类似物。

108. 根据权利要求1所述的方法, 其中(c)进一步包括在所述附接后延伸所述第一互补多核苷酸和所述第二互补多核苷酸。

109. 根据权利要求1-4中任一项所述的方法, 其中所述第一互补多核苷酸包含与第一分子条码化多核苷酸互补的区域。

110. 根据权利要求1-4中任一项所述的方法, 其中所述第二互补多核苷酸包含与第二分子条码化多核苷酸互补的区域。

111. 根据权利要求109所述的方法, 其中所述第一互补多核苷酸包含与第二分子条码化多核苷酸互补的区域。

112. 根据权利要求111所述的方法, 其中与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的区域不与分子条码序列互补。

113. 根据权利要求111所述的方法, 其中与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的区域不与所述容器条码化多核苷酸或其扩增产物的区域互补。

114. 根据权利要求111所述的方法,其中与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的区域包含添加至所述第一互补多核苷酸的3'端的三个或更多个非模板核苷酸。

115. 根据权利要求114所述的方法,其中与所述第二分子条码化多核苷酸互补的所述第二互补多核苷酸的区域包含添加至所述第二互补多核苷酸的3'端的三个或更多个非模板核苷酸。

116. 根据权利要求1-4中任一项所述的方法,其中所述第一互补多核苷酸不与所述容器条码化多核苷酸互补。

117. 根据权利要求116所述的方法,其中所述第二互补多核苷酸不与所述容器条码化多核苷酸互补。

118. 根据权利要求1-4中任一项所述的方法,其中第一分子条码化多核苷酸的互补体的区域与所述容器条码化多核苷酸的区域互补。

119. 根据权利要求118所述的方法,其中第二分子条码化多核苷酸的互补体的区域与所述容器条码化多核苷酸的区域互补。

120. 根据权利要求1-4中任一项所述的方法,其中所述第一单细胞单条码化多核苷酸的区域与所述容器条码化多核苷酸的区域互补。

121. 根据权利要求120所述的方法,其中所述第二单细胞单条码化多核苷酸的区域与所述容器条码化多核苷酸的区域互补。

122. 根据权利要求121所述的方法,其中所述第一单细胞单条码化多核苷酸的区域与所述容器条码化多核苷酸中与所述第二单细胞单条码化多核苷酸互补的区域互补。

123. 根据权利要求1-4中任一项所述的方法,其中所述方法进一步包括用第一引物组扩增所述容器条码化多核苷酸,其中所述扩增在附接所述容器条码化多核苷酸之前或与附接所述容器条码化多核苷酸同时进行。

124. 根据权利要求1-4中任一项所述的方法,其中所述容器条码化多核苷酸包含第一和第二容器条码化多核苷酸,其选自所述容器条码化多核苷酸、所述容器条码化多核苷酸的互补体、来自所述容器条码化多核苷酸的扩增产物及其任意组合。

125. 根据权利要求123所述的方法,其中附接所述容器条码化多核苷酸包括:

(i) 使所述容器条码化多核苷酸或其扩增产物的区域与所述第一单细胞单条码化多核苷酸的区域杂交,以及

(ii) 使所述容器条码化多核苷酸或其扩增产物的区域与所述第二单细胞单条码化多核苷酸的区域杂交。

126. 根据权利要求125所述的方法,其中所述方法进一步包括在附接所述容器条码化多核苷酸之后延伸所述第一单细胞单条码化多核苷酸和所述第二单细胞单条码化多核苷酸,从而形成所述第一和第二单细胞双条码化多核苷酸。

127. 根据权利要求126所述的方法,其中所述第一单细胞双条码化多核苷酸包含与所述容器条码化多核苷酸互补的区域。

128. 根据权利要求127所述的方法,其中所述第二单细胞双条码化多核苷酸包含与所述容器条码化多核苷酸互补的区域。

129. 根据权利要求128所述的方法,其中与所述容器条码化多核苷酸互补的所述第一

和第二单细胞双条码化多核苷酸的区域为相同序列。

130. 根据权利要求129所述的方法,其中与所述第一或第二分子条码化多核苷酸互补的所述第一单细胞单条码化多核苷酸的区域不与所述容器条码化多核苷酸或其扩增产物的区域互补。

131. 根据权利要求123所述的方法,其中所述第一引物组的第一引物与第一分子条码化多核苷酸、所述第一单细胞单条码化多核苷酸的互补体、所述第一单细胞双条码化多核苷酸的互补体或其任意组合的区域互补。

132. 根据权利要求131所述的方法,其中所述第一引物组的所述第一引物与第二分子条码化多核苷酸、所述第二单细胞单条码化多核苷酸的互补体、所述第二单细胞双条码化多核苷酸的互补体或其任意组合的区域互补。

133. 根据权利要求123所述的方法,其中所述第一引物组的第一引物不与所述第一细胞多核苷酸或其互补体互补。

134. 根据权利要求123所述的方法,其中所述第一引物组的所述第一引物不与所述第二细胞多核苷酸或其互补体互补。

135. 根据权利要求123所述的方法,其中所述第一引物组的第一引物与所述第一单细胞单条码化多核苷酸的互补体在所述分子条码的下游的区域互补。

136. 根据权利要求123所述的方法,其中所述第一引物组的所述第一引物与所述第二单细胞单条码化多核苷酸的互补体在所述分子条码的下游的区域互补。

137. 根据权利要求123所述的方法,其中所述第一引物组的第一引物与所述第一单细胞双条码化多核苷酸的互补体在所述容器条码的上游的区域互补。

138. 根据权利要求123所述的方法,其中所述第一引物组的所述第一引物与所述第二单细胞双条码化多核苷酸的互补体在所述容器条码的上游的区域互补。

139. 根据权利要求123所述的方法,其中所述第一引物组的第二引物不与所述第一细胞多核苷酸或其互补体、所述第一互补多核苷酸或其互补体、第一分子条码化多核苷酸或其互补体、所述第一单细胞单条码化多核苷酸或其互补体或其任意组合的区域互补。

140. 根据权利要求139所述的方法,其中所述第一引物组的所述第二引物不与所述第二细胞多核苷酸或其互补体、所述第二互补多核苷酸或其互补体、第二分子条码化多核苷酸或其互补体、所述第二单细胞单条码化多核苷酸或其互补体或其任意组合的区域互补。

141. 根据权利要求123所述的方法,其中所述第一引物组的第二引物与所述第一单细胞双条码化多核苷酸的区域互补。

142. 根据权利要求123所述的方法,其中所述第一引物组的第二引物与所述第二单细胞双条码化多核苷酸的区域互补。

143. 根据权利要求123所述的方法,其中所述第一引物组的第二引物与所述第一单细胞双条码化多核苷酸在所述分子条码的上游的区域互补。

144. 根据权利要求143所述的方法,其中所述第一引物组的所述第二引物与所述第二分子条码化多核苷酸在所述分子条码的上游的区域互补。

145. 根据权利要求123所述的方法,其中所述第一引物组的第二引物与所述第一单细胞双条码化多核苷酸在所述容器条码的上游的区域互补。

146. 根据权利要求145所述的方法,其中所述第一引物组的所述第二引物与所述第二

分子条码化多核苷酸在所述容器条码的上游的区域互补。

147. 根据权利要求1-4中任一项所述的方法,其中所述方法进一步包括打破所述多个容器中的两个或更多个容器。

148. 根据权利要求147所述的方法,其中所述方法进一步包括合并来自所述两个或更多个打破的容器的所述第一和第二单细胞双条码化多核苷酸。

149. 根据权利要求1-4中任一项所述的方法,其中所述方法进一步包括

(e) 扩增所述第一和第二单细胞双条码化多核苷酸。

150. 根据权利要求149所述的方法,其中所述扩增所述第一和第二单细胞双条码化多核苷酸在所述多个容器的容器之外进行。

151. 根据权利要求123所述的方法,其中所述方法进一步包括

(e) 用第二引物组扩增所述第一和第二单细胞双条码化多核苷酸。

152. 根据权利要求151所述的方法,其中所述第二引物组的第一引物不与所述第一细胞多核苷酸或其互补体、所述第一互补多核苷酸或其互补体、第一分子条码化多核苷酸或其互补体、所述第一单细胞单条码化多核苷酸或其互补体或其任意组合的区域互补。

153. 根据权利要求152所述的方法,其中所述第二引物组的所述第一引物不与所述第二细胞多核苷酸或其互补体、所述第二互补多核苷酸或其互补体、第二分子条码化多核苷酸或其互补体、所述第二单细胞单条码化多核苷酸或其互补体或其任意组合的区域互补。

154. 根据权利要求151所述的方法,其中所述第二引物组的第一引物与所述第一单细胞双条码化多核苷酸的区域互补。

155. 根据权利要求154所述的方法,其中所述第二引物组的所述第一引物与所述第二单细胞双条码化多核苷酸的区域互补。

156. 根据权利要求151所述的方法,其中所述第二引物组的第一引物与所述第一单细胞双条码化多核苷酸在所述分子条码的上游的区域互补。

157. 根据权利要求156所述的方法,其中所述第二引物组的所述第一引物与所述第二单细胞双条码化多核苷酸在所述分子条码的上游的区域互补。

158. 根据权利要求151所述的方法,其中所述第二引物组的第一引物与所述第一单细胞双条码化多核苷酸在所述容器条码的上游的区域互补。

159. 根据权利要求158所述的方法,其中所述第二引物组的所述第一引物与所述第二单细胞双条码化多核苷酸在所述容器条码的上游的区域互补。

160. 根据权利要求152所述的方法,其中所述第一引物组的所述第二引物为所述第二引物组的所述第一引物。

161. 根据权利要求151所述的方法,其中所述第二引物组的第二引物与所述第一和第二细胞多核苷酸、所述第一和第二互补多核苷酸的互补体、所述第一和第二单细胞单条码化多核苷酸的互补体、所述第一和第二单细胞双条码化多核苷酸的互补体或其任意组合的区域互补。

162. 根据权利要求161所述的方法,其中所述第二引物组的所述第二引物包含聚(T)序列。

163. 根据权利要求151所述的方法,其中所述第二引物组的第二引物与所述第一或第二细胞多核苷酸、所述第一或第二互补多核苷酸的互补体、所述第一或第二单细胞单条码

化多核苷酸的互补体、所述第一或第二单细胞双条码化多核苷酸的互补体或其任意组合的区域互补。

164. 根据权利要求163所述的方法, 其中所述第二引物组的所述第二引物不与第一或第二分子条码化多核苷酸或其互补体、所述容器条码化多核苷酸或其互补体或其任意组合互补。

165. 根据权利要求151所述的方法, 其中所述第二引物组的第三引物与所述第二细胞多核苷酸、所述第二互补多核苷酸的互补体、所述第二单细胞单条码化多核苷酸的互补体、所述第二单细胞双条码化多核苷酸的互补体或其任意组合的区域互补。

166. 根据权利要求163所述的方法, 其中所述第二引物组的所述第二引物与所述第一细胞多核苷酸、所述第一互补多核苷酸的互补体、所述第一单细胞单条码化多核苷酸的互补体、所述第一单细胞双条码化多核苷酸的互补体或其任意组合的区域互补。

167. 根据权利要求165所述的方法, 其中所述第二引物组的所述第三引物不与所述第一细胞多核苷酸、所述第一互补多核苷酸的互补体、所述第一单细胞单条码化多核苷酸的互补体、所述第一单细胞双条码化多核苷酸的互补体或其任意组合的区域互补。

168. 根据权利要求165所述的方法, 其中所述第二引物组的所述第三引物不与第一或第二分子条码化多核苷酸或其互补体、所述容器条码化多核苷酸或其互补体或其任意组合互补。

169. 根据权利要求151所述的方法, 其中所述第二引物组的所述第二引物包含目标特异性序列。

170. 根据权利要求165所述的方法, 其中所述第二引物组的所述第三引物包含目标特异性序列。

171. 根据权利要求165所述的方法, 其中所述第二引物组的所述第二引物的所述目标特异性序列靶向重链免疫球蛋白 (IgH) 序列、TCR α 序列、TCR γ 序列或其组合。

172. 根据权利要求169所述的方法, 其中所述第二引物组的所述第二引物的所述目标特异性序列靶向重链恒定区序列(C_H)、TCR α 恒定区(C α)序列、TCR γ 恒定区(C γ)序列或其组合。

173. 根据权利要求169所述的方法, 其中所述第二引物的所述目标特异性序列选自GGGTTGGGCGGATGCAC、CATCCGGAGCCTTGTTGG、CCTTGGGGCTGGTCGGGG、CGGATGGGCTCTGTGTGG、CCGATGGGCCCTTGTTGG、GGATTTAGAGTCTCTCAGCTG、CACGGCAGGGTCAGGGTTC和GGGGAACATCTGCATCAAGT。

174. 根据权利要求170所述的方法, 其中所述第二引物组的所述第三引物的所述目标特异性序列靶向轻链免疫球蛋白 (IgL) 序列、TCR β 序列、TCR δ 序列或其组合。

175. 根据权利要求170所述的方法, 其中所述第二引物组的所述第三引物的所述目标特异性序列靶向轻链恒定区序列(C_L)、TCR β 恒定区(C β)序列、TCR δ 恒定区(C δ)序列或其组合。

176. 根据权利要求170所述的方法, 其中所述第三引物的所述目标特异性序列选自TTTGATCTCCACCTTGTTCCCTCCGC、TTTGATCTCCAGCTTGTTCCCTGG、TTTGATATCCACTTTGGTCCCAGGGC、TTTGATTTCCACCTTGTTCCCTTGGC、TTTAATCTCCAGTCGTGTTCCCTTGGC、GAGGACGGTCACCTTGTTGCCA、TAGGACGGTCAGCTTGTTCCCTCC、

GAGGACGGTCAGCTGGGTGCC、TAAAATGATCAGCTGGGTTCTCCAC、TAGGACGGTGACCTTGGTCCCAG、GGGAGATCTCTGCTTCTGATG、CGACCTCGGGTGGGAACAC和CGGATGGTTTGGTATGAGGC。

177. 根据权利要求163所述的方法,其中所述第二引物组的所述第二引物包括多个第二引物。

178. 根据权利要求165所述的方法,其中所述第二引物组的所述第三引物包括多个第三引物。

179. 根据权利要求177所述的方法,其中所述多个第二引物的所述目标特异性序列靶向多个重链免疫球蛋白(IgH)序列、TCR α 序列、TCR γ 序列或其组合。

180. 根据权利要求179所述的方法,其中所述多个重链免疫球蛋白(IgH)序列、TCR α 序列或TCR γ 序列包括多个重链恒定区(C_H)、TCR α 恒定区(C α)序列、TCR γ 恒定区(C γ)序列或其组合。

181. 根据权利要求180所述的方法,其中所述多个重链恒定区(C_H)序列包括两个或更多个序列,所述序列选自来自IgM、IgD、IgA、IgE、IgG及其组合的重链恒定区(C_H)序列。

182. 根据权利要求178所述的方法,其中所述多个第三引物的所述目标特异性序列靶向多个轻链免疫球蛋白(IgL)序列、TCR β 序列、TCR δ 序列或其组合。

183. 根据权利要求182所述的方法,其中所述多个轻链免疫球蛋白(IgL)序列、TCR β 序列或TCR δ 序列包括多个轻链恒定区(C_L)序列、TCR β 恒定区(C β)序列、TCR δ 恒定区(C δ)序列或其组合。

184. 根据权利要求183所述的方法,其中所述多个轻链恒定区(C_L)序列包括两个或更多个序列,所述序列选自来自Ig κ 、Ig λ 及其组合的轻链恒定区(C_L)序列。

185. 根据权利要求1-4中任一项所述的方法,其中第一目标引物、第二目标引物、所述容器条码化多核苷酸、分子条码化多核苷酸或其任意组合不与固体支持体附接。

186. 根据权利要求151所述的方法,其中第一目标引物、第二目标引物、所述第一引物组的引物、所述第二引物组的引物或其任意组合不包括分子条码、容器条码、条码或其任意组合。

187. 根据权利要求151所述的方法,其中第一目标引物、第二目标引物、所述第一引物组的引物、所述第二引物组的引物或其任意组合不包括突出端区域。

188. 根据权利要求1-4中任一项所述的方法,其中所述多个容器的每个容器不包含固体支持体。

189. 根据权利要求1-4中任一项所述的方法,其中所述容器条码化多核苷酸附接至固体支持体。

190. 根据权利要求1-4中任一项所述的方法,其中所述容器条码化多核苷酸附接至珠子。

191. 根据权利要求1-4中任一项所述的方法,其中所述容器条码化多核苷酸、分子条码化多核苷酸或其任意组合不是引物。

192. 根据权利要求1-4中任一项所述的方法,其中所述容器条码化多核苷酸、分子条码化多核苷酸或其任意组合不是延伸的。

193. 根据权利要求1-4中任一项所述的方法,其中(a)-(d)在所述单个容器中进行。

194. 根据权利要求1-4中任一项所述的方法,其中(a)-(d)在单个反应中进行。

195. 根据权利要求1-4中任一项所述的方法,其中所述方法进一步包括裂解所述单细胞或单个免疫细胞。

196. 根据权利要求195所述的方法,其中所述裂解从所述单细胞或单个免疫细胞释放所述第一和第二细胞多核苷酸。

197. 根据权利要求195所述的方法,其中所述单细胞或单个免疫细胞在(a)后裂解。

198. 根据权利要求195所述的方法,其中所述单细胞或单个免疫细胞在(b)前裂解。

199. 根据权利要求195所述的方法,其中所述单细胞或单个免疫细胞在所述容器中裂解。

200. 根据权利要求195所述的方法,其中所述裂解包括冻融。

201. 根据权利要求1-4中任一项所述的方法,其中所述容器条码化多核苷酸的容器条码在(d)前扩增。

202. 根据权利要求1-4中任一项所述的方法,其中所述容器条码化多核苷酸的容器条码与(d)同时扩增。

203. 根据权利要求1-4中任一项所述的方法,其中所述容器条码化多核苷酸的容器条码和所述第一单细胞条码化多核苷酸同时扩增或延伸。

204. 根据权利要求1-4中任一项所述的方法,其中所述容器条码化多核苷酸的容器条码、所述第一单细胞单条码化多核苷酸和所述第二单细胞单条码化多核苷酸同时扩增或延伸。

205. 根据权利要求1-4中任一项所述的方法,其中所述第一单细胞条码化多核苷酸和所述第二单细胞单条码化多核苷酸同时扩增或延伸。

206. 根据权利要求1-4中任一项所述的方法,其中所述第一单细胞双条码化多核苷酸和所述第二单细胞双条码化多核苷酸同时扩增或延伸。

207. 根据权利要求1-4中任一项所述的方法,其中所述多个容器包含多个孔。

208. 根据权利要求1-4中任一项所述的方法,其中所述多个容器包含多份乳液。

209. 根据权利要求208所述的方法,其中所述多份乳液的每份乳液的体积为0.01皮升至10微升。

210. 根据权利要求1-4中任一项所述的方法,其中所述多个容器包含多个器皿。

211. 根据权利要求151所述的方法,其中所述第一目标引物、第二目标引物、所述第一引物组的引物或所述第二引物组的引物包括样品条码。

212. 根据权利要求1-4中任一项所述的方法,其中所述方法进一步包括从所述容器回收所述第一单细胞双条码化多核苷酸、所述第二单细胞双条码化多核苷酸及其扩增产物。

213. 根据权利要求1-4中任一项所述的方法,其中所述方法进一步包括对所述第一单细胞双条码化多核苷酸、所述第二单细胞双条码化多核苷酸、其扩增产物或其任意组合测序。

214. 根据权利要求213所述的方法,其中所述第一单细胞双条码化多核苷酸、所述第二单细胞双条码化多核苷酸、其扩增产物或其任意组合同时测序。

215. 根据权利要求213所述的方法,其中所述第一单细胞双条码化多核苷酸、所述第二单细胞双条码化多核苷酸、其扩增产物或其任意组合在相同的反应中测序。

216. 根据权利要求1-4中任一项所述的方法,进一步包括基于所述容器条码,确定所述

第一细胞多核苷酸和所述第二细胞多核苷酸的所述细胞来源是相同的。

217. 根据权利要求216所述的方法, 其中所述确定包括匹配所述第一单细胞双条码化多核苷酸或其扩增产物的所述容器条码的所述序列与所述第二单细胞双条码化多核苷酸或其扩增产物的所述容器条码的所述序列。

218. 根据权利要求1所述的方法, 进一步包括基于所述分子条码, 确定具有所述第一细胞多核苷酸、所述第二细胞多核苷酸或二者的序列的起始分子的数目。

219. 根据权利要求216所述的方法, 其中所述确定包括确定具有相同第一分子条码、相同第二分子条码或二者的序列的所述数目。

220. 根据权利要求1-4中任一项所述的方法, 其中当单细胞双条码化多核苷酸或其扩增产物的第一序列和单细胞双条码化多核苷酸或其扩增产物的第二序列含有相同的容器条码或其互补体时, 它们来自相同的单个容器或单细胞。

221. 根据权利要求220所述的方法, 其中当单细胞双条码化多核苷酸或其扩增产物的所述第一序列和单细胞双条码化多核苷酸或其扩增产物的所述第二序列含有不同的分子条码或其互补体时, 它们来自不同的细胞多核苷酸分子。

222. 根据权利要求220所述的方法, 其中当单细胞双条码化多核苷酸或其扩增产物的所述第一序列和单细胞双条码化多核苷酸或其扩增产物的所述第二序列含有相同的分子条码或其互补体时, 它们来自相同的细胞多核苷酸分子。

223. 根据权利要求220所述的方法, 其中当单细胞双条码化多核苷酸或其扩增产物的所述第一序列和单细胞双条码化多核苷酸或其扩增产物的所述第二序列含有不同的容器条码或其互补体时, 它们来自不同的单个容器或单细胞。

224. 根据权利要求1所述的方法, 其中所述单细胞包括免疫细胞。

225. 根据权利要求1所述的方法, 其中所述多个细胞包括多个免疫细胞。

226. 根据权利要求2-4中任一项所述的方法, 其中所述多个免疫细胞的所述单个免疫细胞为淋巴细胞或其亚型、B细胞或其亚型、T细胞或其亚型或其组合。

227. 根据权利要求2-4和225中任一项所述的方法, 其中所述多个免疫细胞针对记忆B细胞、初始B细胞、成浆B细胞、初始T细胞、成浆T细胞、B细胞的任何亚型、T细胞的任何亚型或其任意组合进行了富集。

228. 根据权利要求1所述的方法, 其中所述单细胞包括癌细胞。

229. 根据权利要求1所述的方法, 其中所述多个细胞包括多个癌细胞。

230. 根据权利要求228所述的方法, 其中所述癌细胞为鳞状细胞癌细胞、腺癌细胞、移行细胞癌细胞、骨肉瘤细胞、软骨肉瘤细胞、肌肉肉瘤细胞、白血病细胞、淋巴瘤细胞、神经胶质瘤细胞或其任意组合。

231. 根据权利要求229所述的方法, 其中所述多个癌细胞针对循环癌细胞、内皮癌细胞、上皮癌细胞、罕见癌细胞或癌细胞的任何类型或亚型进行了富集。

232. 根据权利要求1-4中任一项所述的方法, 其中所述样品为生物样品。

233. 根据权利要求232所述的方法, 其中所述生物样品来自受试者。

234. 根据权利要求233所述的方法, 其中所述受试者为动物。

235. 根据权利要求234所述的方法, 其中所述动物为人类。

236. 根据权利要求233所述的方法, 其中所述方法进一步包括确定所述受试者对于等

位基因是纯合还是杂合的。

237. 根据权利要求1-4中任一项所述的方法,其中所述样品为血液样品。

238. 根据权利要求1所述的方法,其中从所述样品分离所述第一或第二细胞多核苷酸。

239. 根据权利要求1所述的方法,其中不从所述样品分离所述第一或第二细胞多核苷酸。

240. 根据权利要求1-4中任一项所述的方法,其中所述样品包括多个包含第一样品和第二样品的样品。

241. 根据权利要求240所述的方法,其中所述多个样品包括至少3、4、5、10、20、30、40、50、60、70、80、90或100个或更多样品。

242. 根据权利要求240所述的方法,其中所述多个样品包括至少100、200、300、400、500、600、700、800、900或1000个或更多样品。

243. 根据权利要求240所述的方法,其中所述多个样品包括至少1000、2000、3000、4000、5000、6000、7000、8000个样品、9000或10,000个样品、或100,000个样品、或1,000,000个或更多样品。

244. 根据权利要求240所述的方法,其中所述多个样品包括至少10,000个样品。

245. 根据权利要求240所述的方法,其中所述第一样品来自第一受试者,且所述第二样品来自第二受试者。

246. 根据权利要求245所述的方法,其中所述第一受试者为具有疾病或病况的受试者。

247. 根据权利要求245所述的方法,其中所述第二受试者为不具有疾病或病况的受试者。

248. 根据权利要求1所述的方法,其中所述第一或第二细胞多核苷酸包含变异序列。

249. 根据权利要求248所述的方法,其中所述变异序列包括突变、多态性、缺失或插入。

250. 根据权利要求249所述的方法,其中所述多态性为单核苷酸多态性。

251. 根据权利要求1所述的方法,其中所述第一或第二细胞多核苷酸为对于疾病或病况的生物标志物。

252. 根据权利要求1所述的方法,其中所述第一或第二细胞多核苷酸来自病原体。

253. 根据权利要求252所述的方法,其中所述病原体为病毒、细菌或真菌。

254. 根据权利要求1-4中任一项所述的方法,其中所述方法进一步包括比较来自受试者的所述第一和第二单细胞双条码化多核苷酸的文库与在不同时间点来自相同受试者的所述第一和第二单细胞双条码化多核苷酸的文库的所述序列。

255. 根据权利要求1-4中任一项所述的方法,其中所述方法进一步包括比较来自具有疾病或病况的受试者的所述第一和第二单细胞双条码化多核苷酸的文库与来自没有所述疾病或病况的受试者的所述第一和第二单细胞双条码化多核苷酸的文库的所述序列。

256. 根据权利要求1所述的方法,其中所述方法进一步包括确定所述第一细胞多核苷酸、所述第二细胞多核苷酸或二者的种系序列;其中所述第一细胞多核苷酸包括IgH或V_H序列,且其中所述第二细胞多核苷酸包括IgL或V_L序列。

257. 根据权利要求256所述的方法,进一步包括根据所述种系序列中的其序列确定所述IgL、IgH、V_H、V_L或其任意组合的序列的变异。

258. 根据权利要求256或257所述的方法,其中所述方法进一步包括确定以下至少一

个：

- (a) 独特的IgH序列的总数目；
- (b) 独特的IgL序列的总数目；
- (c) 独特的IgH和IgL序列的总数目；
- (d) 独特的成对IgL和IgH序列的总数目；
- (e) IgH序列或IgL序列的频率；或
- (f) IgH序列和IgL序列的组合相对于其他一个或多个组合的频率。

259. 根据权利要求1所述的方法，其中所述方法进一步包括确定所述第一细胞多核苷酸、所述第二细胞多核苷酸或二者的种系序列；其中所述第一细胞多核苷酸包括TCR α 或V α 序列，且其中所述第二细胞多核苷酸包括TCR β 或V β 序列。

260. 根据权利要求259所述的方法，进一步包括根据所述种系序列中的其序列确定所述TCR α 、TCR β 、V α 、V β 或其任意组合的序列的变异。

261. 根据权利要求259或260所述的方法，其中所述方法进一步包括确定以下至少一个：

- (a) 独特的TCR α 序列的总数目；
- (b) 独特的TCR β 序列的总数目；
- (c) 独特的TCR α 和TCR β 序列的总数目；
- (d) 独特的成对TCR β 和TCR α 序列的总数目；
- (e) TCR α 序列或TCR β 序列的频率；或
- (f) TCR α 序列和TCR β 序列的组合相对于其他一个或多个组合的频率。

262. 根据权利要求1所述的方法，其中所述方法进一步包括确定所述第一细胞多核苷酸、所述第二细胞多核苷酸或二者的种系序列；其中所述第一细胞多核苷酸包括TCR γ 或V γ 序列，且其中所述第二细胞多核苷酸包括TCR δ 或V δ 序列。

263. 根据权利要求262所述的方法，进一步包括根据所述种系序列中的其序列确定所述TCR γ 、TCR δ 、V γ 、V δ 或其任意组合的序列的变异。

264. 根据权利要求262或263所述的方法，其中所述方法进一步包括确定以下至少一个：

- (a) 独特的TCR γ 序列的总数目；
- (b) 独特的TCR δ 序列的总数目；
- (c) 独特的TCR γ 和TCR δ 序列的总数目；
- (d) 独特的成对TCR δ 和TCR γ 序列的总数目；
- (e) TCR γ 序列或TCR δ 序列的频率；或
- (f) TCR γ 序列和TCR δ 序列的组合相对于其他一个或多个组合的频率。

265. 根据权利要求1-4中任一项所述的方法，其中所述方法进一步包括确定以下至少一个：

- (a) 来自第一基因的序列的总数目；
- (b) 来自第二基因的序列的总数目；
- (c) 来自第一基因的独特序列的总数目；
- (d) 来自第二基因的独特序列的总数目；或

(e) 来自第一基因的序列或来自第二基因的序列的频率。

266. 根据权利要求256、259或262所述的方法, 其中所述方法进一步包括基于分别成对的IgL和IgH序列、或TCR α 和TCR β 序列、或TCR γ 和TCR δ 序列的一对或多对的总量, 以及来自种系的变异, 来选择抗体或TCR。

267. 根据权利要求256、259或262所述的方法, 其中所述方法进一步包括基于一个或多个IgL或IgH序列、或TCR α 和TCR β 序列、或TCR γ 和TCR δ 序列, 以及来自种系的变异, 来选择抗体或TCR。

268. 根据权利要求256、259或262所述的方法, 其中所述方法进一步包括基于一个或多个序列模式、变异分析、动力学或频率来选择抗体或TCR。

269. 根据权利要求268所述的方法, 其中所述方法进一步包括基于频率选择抗体或TCR。

270. 根据权利要求266所述的方法, 其中所述选定的抗体或TCR以小于或等于 1×10^{-7} 、 1×10^{-8} 、 1×10^{-9} 、 1×10^{-10} 、 1×10^{-11} 或 1×10^{-12} M的 K_D 与表位结合。

271. 根据权利要求266所述的方法, 其中所述选定的抗体或TCR为人类治疗性抗体或TCR。

272. 根据权利要求266所述的方法, 其中所述选定的抗体或TCR为中和抗体或TCR。

273. 根据权利要求266所述的方法, 其中所述选定的抗体或TCR结合的目标是未知的。

274. 根据权利要求266所述的方法, 其中所述选定的抗体或TCR结合的目标在选择所述选定的抗体或TCR时是未知的。

275. 根据权利要求266所述的方法, 其中所述方法进一步包括使所述选定的抗体或TCR与至少一个生物标志物候选物接触以发现生物标志物。

276. 根据权利要求275所述的方法, 其中所述至少一个生物标志物候选物在固体支持体上。

277. 根据权利要求275所述的方法, 其中所述至少一个生物标志物在溶液中。

278. 根据权利要求266所述的方法, 其中所述选定的抗体或TCR在固体支持体上。

279. 根据权利要求266所述的方法, 其中所述选定的抗体或TCR在溶液中。

280. 根据权利要求276所述的方法, 其中所述固体支持体为阵列。

281. 根据权利要求276所述的方法, 其中所述固体支持体为珠子。

282. 根据权利要求1所述的方法, 其中所述方法进一步包括将所述第一细胞多核苷酸插入载体中。

283. 根据权利要求282所述的方法, 进一步包括将所述第二细胞多核苷酸插入所述载体中。

284. 根据权利要求282或283所述的方法, 其中所述载体为克隆载体。

285. 根据权利要求282或283所述的方法, 其中所述载体为表达载体。

286. 根据权利要求1-4中任一项所述的方法, 其中所述方法进一步包括匹配具有相同分子条码的序列。

287. 根据权利要求1-4中任一项所述的方法, 其中所述方法进一步包括从所述文库形成共有序列。

288. 根据权利要求2-4中任一项所述的方法, 其中测序和PCR错误被最小化、消除或小

于0.01%、0.001%、0.0001%、0.00001%、0.000001%或0.0000001%。

289. 根据权利要求1-4中任一项所述的方法,其中扩增反应中循环的数目限于1-40个循环的任何一个。

290. 根据权利要求266所述的方法选定的抗体或TCR或其片段在制备用于治疗有需要的受试者的药物中的用途。

291. 根据权利要求290所述的用途,其中所述抗体或TCR或其片段鉴别为来自所述有需要的受试者。

292. 根据权利要求290所述的用途,其中所述抗体或TCR或其片段未鉴别为来自所述有需要的受试者。

293. 根据权利要求290所述的用途,其中所述有需要的受试者展示疾病的一种或多种症状。

294. 根据权利要求290所述的用途,其中所述有需要的受试者具有疾病。

295. 根据权利要求293所述的用途,其中所述疾病是未知的。

296. 根据权利要求293所述的用途,其中所述疾病是已知的。

297. 根据权利要求1-4中任一项所述的方法,其中所述样品包括在第一时间点从受试者获取的第一样品和在第二时间点从所述受试者获取的第二样品。

298. 根据权利要求297所述的方法,进一步包括确定来自在所述第一和第二时间点获取的样品的所述第一或第二细胞多核苷酸的量中的增加或减少。

299. 根据权利要求298所述的方法,其中所述量中的增加或减少为范围至少:0.1倍、0.2倍、0.3倍、0.4倍、0.5倍、0.6倍、0.7倍、0.8倍、0.9倍、1.5倍、2倍、3倍、5倍、10倍、50倍、100倍、1,000倍、10,000倍、100,000倍、1,000,000倍或更多的增加或减少。

300. 根据权利要求297所述的方法,其中所述第一和第二时间点之间的所述时间为约1小时、2小时、3小时、4小时、5小时、6小时、7小时、8小时、9小时、10小时、11小时、12小时、13小时、14小时、15小时、16小时、17小时、18小时、19小时、20小时、21小时、22小时、23小时、1天、2天、3天、4天、5天、6天、1周、2周、3周、4周、5周、6周、7周、8周、9周、10周、11周、12周、4个月、5个月、6个月、7个月、8个月、9个月、10个月、11个月、12个月,或至少:1小时、2小时、3小时、4小时、5小时、6小时、7小时、8小时、9小时、10小时、11小时、12小时、13小时、14小时、15小时、16小时、17小时、18小时、19小时、20小时、21小时、22小时、23小时、1天、2天、3天、4天、5天、6天、1周、2周、3周、4周、5周、6周、7周、8周、9周、10周、11周、12周、4个月、5个月、6个月、7个月、8个月、9个月、10个月、11个月、12个月或更久。

301. 根据权利要求2-4中任一项所述的方法,其中所述测序为高通量的。

302. 根据权利要求1-4中任一项所述的方法,其中所述方法不包括引物的复用和/或衔接至固体支持体的引物的复用。

303. 根据权利要求1-4中任一项所述的方法,其中所述方法不使用包含与单功能性V区段或V区段的小家族互补的序列的多重V-区段引物。

304. 根据权利要求1所述的方法,其中所述方法不使用分离所述第一或第二细胞多核苷酸的步骤。

305. 根据权利要求2-4中任一项所述的方法,其中所述测序通过大规模平行合成完成。

306. 根据权利要求2-4中任一项所述的方法,其中所述方法进一步包括比较序列读取

与种系序列,并确定所述序列读取的体细胞超突变积累。

307.根据权利要求1-4中任一项所述的方法,其中所述方法进一步包括确定抗体序列的同种型分布以选择特异性同种型。

308.根据权利要求266所述的方法,其中所选定的抗体包括特异性Ig同种型。

309.根据权利要求308所述的方法,其中所述Ig同种型为IgA、IgG、IgM、IgD或IgE。

310.根据权利要求1-4中任一项所述的方法,其中所述方法进一步包括生成成对的IgH和IgL抗体序列或TCR α 和TCR β 序列的文库。

311.根据权利要求310所述的方法,其中所述文库为数据库。

312.根据权利要求1-4中任一项所述的方法,其中所述第一和第二单细胞双条码化多核苷酸包括遍布抗体或TCR编码序列中的CDR1、CDR2、CDR3和/或超突变区域。

313.根据权利要求266所述的方法,其中所述方法进一步包括将所述选定的抗体或TCR直接克隆用于表面展示技术。

314.根据权利要求266所述的方法,其中所述方法进一步包括通过定向演化来演化所述选定的抗体或TCR。

315.根据权利要求266所述的方法,其中所述方法进一步包括对于功能特异性、亲和力或中和能力筛选所述选定的抗体或TCR。

316.根据权利要求1-4中任一项所述的方法,其中以99%或更高的置信度确定体细胞突变。

317.根据权利要求1-4中任一项所述的方法,其中鉴别来自每个多核苷酸分子的V、D和J区段。

318.根据权利要求1-4中任一项所述的方法,其中所述容器条码化多核苷酸的容器条码包含至少2个核苷酸。

319.根据权利要求318所述的方法,其中所述容器条码包含至少3、4、5、6、7、8或9个核苷酸。

320.根据权利要求318所述的方法,其中所述容器条码包含至少10个核苷酸。

321.根据权利要求318所述的方法,其中所述容器条码包含至少15个核苷酸。

322.根据权利要求318所述的方法,其中所述容器条码包含最多50个核苷酸。

323.根据权利要求318所述的方法,其中所述容器条码包含10-30个核苷酸。

324.根据权利要求318所述的方法,其中所述容器条码化多核苷酸的容器条码包含简并序列。

325.根据权利要求318所述的方法,其中所述容器条码化多核苷酸的容器条码包含完整或部分的简并序列。

326.根据权利要求318所述的方法,其中所述容器条码化多核苷酸的容器条码包含所述序列NNNNNNNNNNNNNN,其中N为任何核酸。

327.根据权利要求318所述的方法,其中所述容器条码化多核苷酸的容器条码包含所述序列NNNNNWNNNNNWNNNNN,其中N为任何核酸,且W为腺嘌呤或胸腺嘧啶。

328.根据权利要求318所述的方法,其中所述容器条码化多核苷酸的容器条码包含所述序列NNNNNXNNNNXNNNNN,其中N为任何核酸,且X为任何已知的核苷酸。

329.根据权利要求318所述的方法,其中所述容器条码化多核苷酸的容器条码包含所

述序列NNNNNNNNNNNNNNNNNN，其中N为任何核酸，且序列中至少一个或两个N为W，其中W为腺嘌呤或胸腺嘧啶。

330. 根据权利要求318所述的方法，其中所述容器条码化多核苷酸的容器条码包含所述序列NNNNNNNNNNNNNNNNNN，其中N为任何核酸，且序列中至少一个或两个N为X，其中X为任何已知的核苷酸。

331. 根据权利要求1-4中任一项所述的方法，其中所述分子条码化多核苷酸的分子条码包含至少2个核苷酸。

332. 根据权利要求331所述的方法，其中所述分子条码包含至少3、4、5、6、7、8或9个核苷酸。

333. 根据权利要求331所述的方法，其中所述分子条码包含至少10个核苷酸。

334. 根据权利要求331所述的方法，其中所述分子条码包含至少15个核苷酸。

335. 根据权利要求331所述的方法，其中所述分子条码包含最多50个核苷酸。

336. 根据权利要求331所述的方法，其中所述分子条码包含10-30个核苷酸。

337. 根据权利要求331所述的方法，其中所述分子条码包含简并序列。

338. 根据权利要求331所述的方法，其中所述分子条码包含完整或部分的简并序列。

339. 根据权利要求331所述的方法，其中所述分子条码包含序列NNNNNNNN，其中N为任何核酸。

340. 根据权利要求331所述的方法，其中所述分子条码包含序列NNTNNANN，其中N为任何核酸。

341. 根据权利要求331所述的方法，其中所述分子条码包含序列NNWNNWNN，其中N为任何核酸，且W为腺嘌呤或胸腺嘧啶。

342. 根据权利要求331所述的方法，其中所述分子条码包含序列NNXNNXNN，其中N为任何核酸，且X为任何已知的核苷酸。

343. 根据权利要求331所述的方法，其中所述分子条码包含序列NNNNNNNN，其中N为任何核酸，且序列中至少一个或两个N为W，其中W为腺嘌呤或胸腺嘧啶。

344. 根据权利要求331所述的方法，其中所述分子条码包含序列NNNNNNNN，其中N为任何核酸，且序列中至少一个或两个N为X，其中X为任何已知的核苷酸。

345. 根据权利要求1-4中任一项所述的方法，其中所述方法进一步包括修正扩增错误。

346. 根据权利要求2-4中任一项所述的方法，其中所述方法进一步包括修正测序错误。

347. 根据权利要求2-4中任一项所述的方法，其中所述方法进一步包括将包含具有相同核酸序列的分子条码的序列分箱或分组。

348. 根据权利要求2-4中任一项所述的方法，其中所述方法进一步包括使用计算机或算法将包含具有相同核酸序列的分子条码的序列分箱或分组。

349. 根据权利要求2-4中任一项所述的方法，其中所述方法进一步包括使用计算机或算法将包含具有相同核酸序列的容器条码的序列分箱或分组。

350. 根据权利要求2-4中任一项所述的方法，其中所述方法进一步包括簇集具有至少90%、95%或99%序列同源性的序列。

351. 根据权利要求2-4中任一项所述的方法，其中所述方法进一步包括对具有至少90%、95%或99%序列同源性的序列进行比对。

352. 根据权利要求350所述的方法, 其中所述簇集或比对在计算机或算法的帮助下进行。

353. 根据权利要求306所述的方法, 包括确定含有具有相同核酸序列的分子条码的序列读取的数目。

354. 根据权利要求306所述的方法, 包括确定含有具有相同核酸序列的分子条码和具有至少90%、95%或99%序列同源性的相同第一细胞多核苷酸序列的序列读取的数目。

355. 根据权利要求354所述的方法, 包括确定含有具有相同核酸序列的分子条码和具有至少90%、95%或99%序列同源性的相同第二细胞多核苷酸序列的序列读取的数目。

356. 根据权利要求1所述的方法, 包括确定所述样品中所述第一或第二细胞多核苷酸的所述量。

357. 根据权利要求1-4中任一项所述的方法, 包括形成来自两个或更多个序列、序列读取、扩增子序列、分箱序列、比对序列、簇集序列、或包含具有相同核酸序列的分子条码或容器条码或二者的扩增子集序列的共有序列。

358. 根据权利要求1所述的方法, 包括确定以至少80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.5%、99.6%、99.7%、99.8%、99.9%、99.99%或100%的准确度或置信度确定第一或第二细胞多核苷酸序列。

359. 根据权利要求2-4中任一项所述的方法, 其中测序和PCR错误被最小化、消除或小于0.01%、0.001%、0.0001%、0.00001%、0.000001%或0.0000001%。

360. 根据权利要求2-4中任一项所述的方法, 其中所述测序的错误率小于或等于0.00001%、0.0001%、0.001%、0.01%或0%。

361. 根据权利要求2-4中任一项所述的方法, 其中所述测序的错误率不是0。

362. 根据权利要求2-4中任一项所述的方法, 其中对至少1,000、5,000、10,000、20,000、30,000、40,000、50,000、100,000、500,000或 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 或 9×10^{12} 个多核苷酸测序。

363. 根据权利要求1-4中任一项所述的方法, 其中所述方法在小于或等于4周、3周、2周、1周、6天、5天、4天、3天、2天、1天、18小时、12小时、9小时、6小时、3小时、2小时或1小时的为正值的时间内进行。

364. 根据权利要求306所述的方法, 其中用于达到特定置信度或碱基判定精确度的序列读取的数目比使用相似的方法但不使用分子条码、容器条码或二者而用于达到所述相同、相似或更高的置信度或碱基判定精确度的序列读取的数目少至少1.1、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、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000倍。

365. 根据权利要求306所述的方法, 其中用于达到特定置信度或碱基判定精确度的序列读取的数目比使用相似的方法但不使用分子条码、容器条码或二者而用于达到所述相

同、相似或更高的置信度或碱基判定精确度的序列读取的数目至少1、2、3、4、5、5.5、6、6.5、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 或 9×10^{12} 个读取。

366. 根据权利要求1-4中任一项所述的方法,其中所述多个容器包括至少3、4、5、5.5、6、6.5、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 或 9×10^{12} 个或更多容器。

367. 根据权利要求1所述的方法,其中所述多个细胞多核苷酸包括至少3、4、5、5.5、6、6.5、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 或 9×10^{12} 个或更多细胞多核苷酸。

368. 一种组合物,包含:

(a) 多个容器,每个包含

(i) 来自包含多个细胞的样品的单细胞,

(ii) 多个分子条码化多核苷酸,

(iii) 容器条码化多核苷酸,

(iv) 第一互补多核苷酸,其与来自所述单细胞的第一细胞多核苷酸互补,以及

(v) 第二互补多核苷酸,其与来自所述单细胞的第二细胞多核苷酸互补;

其中所述第一互补多核苷酸包含所述多个分子条码化多核苷酸的第一分子条码和所述容器条码化多核苷酸的扩增产物或所述容器条码化多核苷酸的所述容器条码,以及

其中所述第二互补多核苷酸包含所述多个分子条码化多核苷酸的第二分子条码和所述容器条码化多核苷酸的扩增产物或所述容器条码化多核苷酸的所述容器条码。

369. 根据权利要求368所述的组合物,其中所述第一和第二分子条码化多核苷酸的所述分子条码不同。

370. 根据权利要求368所述的组合物,其中所述第一和第二互补多核苷酸包含不同的分子条码。

371. 根据权利要求368所述的组合物,其中所述第一和第二互补多核苷酸包含相同的容器条码。

372. 根据权利要求368所述的组合物,其中所述多个分子条码化多核苷酸不是扩增产物。

373. 根据权利要求368所述的组合物,其中第一容器中分子条码化多核苷酸的所述分子条码与第二容器中分子条码化多核苷酸的所述分子条码不同。

374. 根据权利要求368所述的组合物,其中所述多个容器的第一容器中每个分子条码化多核苷酸的所述分子条码是独特的。

375. 根据权利要求374所述的组合物,其中所述多个容器的第二容器中每个分子条码化多核苷酸的所述分子条码是独特的。

376. 根据权利要求368所述的组合物,其中第一容器和第二容器中每个分子条码化多核苷酸的所述分子条码是独特的。

377. 根据权利要求375或376所述的组合物,其中所述多个容器的第三容器中每个分子条码化多核苷酸的所述分子条码是独特的。

378. 根据权利要求377所述的组合物,其中所述第一容器、所述第二容器和所述第三容器中每个分子条码化多核苷酸的所述分子条码是独特的。

379. 根据权利要求368所述的组合物,其中所述多个容器的任何单个容器中每个分子条码化多核苷酸的所述分子条码是独特的。

380. 根据权利要求368所述的组合物,其中所述多个容器的任何一个容器中每个分子条码化多核苷酸的所述分子条码与所述多个容器的任何其他一个容器中每个分子条码化多核苷酸的所述分子条码不同。

381. 根据权利要求368所述的组合物,其中第一容器中分子条码化多核苷酸的所述分子条码与第二容器中分子条码化多核苷酸的所述分子条码相同。

382. 根据权利要求368所述的组合物,其中第一容器中分子条码化多核苷酸的所述分子条码与所述第一容器中分子条码化多核苷酸的所述分子条码相同。

383. 根据权利要求382所述的组合物,其中第二容器中分子条码化多核苷酸的所述分子条码与所述第二容器中分子条码化多核苷酸的所述分子条码相同。

384. 根据权利要求368所述的组合物,其中所述多个容器的第一容器中容器条码化多核苷酸或其扩增子的所述容器条码与所述多个容器的第二容器中容器条码化多核苷酸或其扩增子的所述容器条码不同。

385. 根据权利要求368所述的组合物,其中所述多个容器的第一容器中容器条码化多核苷酸或其扩增子的所述容器条码为第一相同容器条码。

386. 根据权利要求385所述的组合物,其中所述多个容器的第二容器中每个容器条码化多核苷酸或其扩增子的所述容器条码为第二相同容器条码。

387. 根据权利要求385或386所述的组合物,其中所述第一相同容器条码与所述第二相同容器条码不同。

388. 根据权利要求368所述的组合物,其中所述多个容器的单个容器中每个容器条码化多核苷酸或其扩增子的所述容器条码包含相同的容器条码。

389. 根据权利要求368所述的组合物,其中所述多个容器的任何单个容器中每个容器条码化多核苷酸及其扩增子的所述容器条码对于所述多个容器的任何其他单个容器中每个容器条码化多核苷酸及其扩增子的所述容器条码而言是独特的。

390. 根据权利要求368所述的组合物,其中所述容器条码化多核苷酸以单分子存在于容器中。

391. 根据权利要求368所述的组合物,其中所述容器条码化多核苷酸以单分子存在于所述多个容器的每个容器中。

392. 根据权利要求368所述的组合物,其中所述容器条码化多核苷酸以至少单分子存在于所述多个容器的容器中。

393. 根据权利要求368所述的组合物,其中所述容器条码化多核苷酸以至少单分子存在于所述多个容器的每个容器中。

394. 根据权利要求368所述的组合物,其中所述多个容器的第一容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述第一容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。

395. 根据权利要求394所述的组合物,其中所述多个容器的所述第一容器中所述第一容器条码化多核苷酸或其扩增子的第二共同容器序列与所述第一容器中第二容器条码化多核苷酸或其扩增子的第二共同容器序列相同。

396. 根据权利要求394或395所述的组合物,其中所述多个容器的任何单个容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述单个容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。

397. 根据权利要求394所述的组合物,其中所述多个容器的单个容器中每个容器条码化多核苷酸包含相同的第一共同容器序列。

398. 根据权利要求397所述的组合物,其中所述多个容器的单个容器中每个容器条码化多核苷酸包含相同的第二共同容器序列。

399. 根据权利要求368所述的组合物,其中所述多个容器的第一容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述多个容器的第二容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。

400. 根据权利要求399所述的组合物,其中所述第一容器条码化多核苷酸或其扩增子的第二共同容器序列与所述第二容器条码化多核苷酸或其扩增子的第二共同容器序列相同。

401. 根据权利要求368所述的组合物,其中所述多个容器的任何一个容器中每个容器

条码化多核苷酸或其扩增子包含第一共同容器序列,所述第一共同容器序列包含与所述多个容器的任何其他一个容器中容器条码化多核苷酸或其扩增子的第一共同容器序列相同的序列。

402. 根据权利要求401所述的组合物,其中所述多个容器的任何一个容器中每个容器条码化多核苷酸或其扩增子包含第二共同容器序列,所述第二共同容器序列包含与所述多个容器的任何其他一个容器中容器条码化多核苷酸或其扩增子的第二共同容器序列相同的序列。

403. 根据权利要求368所述的组合物,其中所述多个容器的第一容器中第一分子条码化多核苷酸的第一共同分子序列与所述第一容器中第二分子条码化多核苷酸的第一共同分子序列相同。

404. 根据权利要求403所述的组合物,其中所述多个容器的所述第一容器中所述第一分子条码化多核苷酸的第二共同分子序列与所述第一容器中第二分子条码化多核苷酸的第二共同分子序列相同。

405. 根据权利要求403或404所述的组合物,其中所述多个容器的任何单个容器中第一分子条码化多核苷酸的第一共同分子序列与所述单个容器中第二分子条码化多核苷酸的第一共同分子序列相同。

406. 根据权利要求368所述的组合物,其中所述多个容器的单个容器中每个分子条码化多核苷酸包含相同的第一共同分子序列。

407. 根据权利要求406所述的组合物,其中所述多个容器的单个容器中每个分子条码化多核苷酸包含相同的第二共同分子序列。

408. 根据权利要求368所述的组合物,其中所述多个容器的第一容器中第一分子条码化多核苷酸的第一共同分子序列与所述多个容器的第二容器中第二分子条码化多核苷酸的第一共同分子序列相同。

409. 根据权利要求408所述的组合物,其中所述第一分子条码化多核苷酸的第二共同分子序列与所述第二分子条码化多核苷酸的第二共同分子序列相同。

410. 根据权利要求368所述的组合物,其中所述多个容器的任何一个容器中每个分子条码化多核苷酸包含第一共同分子序列,所述第一共同分子序列包含与所述多个容器的任何其他一个容器中分子条码化多核苷酸的第一共同分子序列相同的序列。

411. 根据权利要求410所述的组合物,其中所述多个容器的任何一个容器中每个分子条码化多核苷酸包含第二共同分子序列,所述第二共同分子序列包含与所述多个容器的任何其他一个容器中分子条码化多核苷酸的第二共同分子序列相同的序列。

412. 根据权利要求394所述的组合物,其中所述第一共同容器序列包含的序列包含与所述第一共同分子序列相同的序列。

413. 根据权利要求412所述的组合物,其中所述第一共同容器序列包含与所述第一共同分子序列或其互补体互补的序列。

414. 根据权利要求404所述的组合物,其中所述第二共同分子序列包含与添加至所述第一互补多核苷酸的3'端的三个或更多个非模板核苷酸互补的区域。

415. 根据权利要求414所述的组合物,其中与添加至所述第一互补多核苷酸的所述3'端的三个或更多个非模板核苷酸互补的所述区域为末端区域。

416. 根据权利要求368所述的组合物, 其中第一和第二分子条码化多核苷酸不融合在一起。

417. 根据权利要求368所述的组合物, 其中所述第一和第二互补多核苷酸不融合在一起。

418. 根据权利要求368所述的组合物, 其中所述第一细胞多核苷酸为DNA。

419. 根据权利要求418所述的组合物, 其中所述第二细胞多核苷酸为DNA。

420. 根据权利要求368所述的组合物, 其中所述第一细胞多核苷酸为RNA。

421. 根据权利要求420所述的组合物, 其中所述第二细胞多核苷酸为RNA。

422. 根据权利要求420或421所述的组合物, 其中所述RNA为mRNA。

423. 根据权利要求420或421所述的组合物, 其中所述第一互补多核苷酸为cDNA。

424. 根据权利要求423所述的组合物, 其中所述第二互补多核苷酸为cDNA。

425. 根据权利要求368所述的组合物, 其中所述组合物进一步包含非模板末端转移酶、逆转录酶、聚合酶或其任意组合。

426. 根据权利要求425所述的组合物, 其中所述第一和/或第二互补多核苷酸包含添加至所述3'端的三个或更多个非模板核苷酸。

427. 根据权利要求425或426所述的组合物, 其中所述非模板末端转移酶为逆转录酶, 并且其中所述逆转录酶选自SuperscriptII逆转录酶、Maxima逆转录酶、ProtoscriptII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。

428. 根据权利要求425所述的组合物, 其中第一分子条码化多核苷酸包含与所述第一互补多核苷酸的所述3'端上的所述三个或更多个非模板核苷酸互补的区域。

429. 根据权利要求428所述的组合物, 其中第二分子条码化多核苷酸包含与所述第二互补多核苷酸的所述3'端上的三个或更多个非模板核苷酸互补的区域。

430. 根据权利要求426所述的组合物, 其中所述三个或更多个非模板核苷酸为相同的。

431. 根据权利要求426所述的组合物, 其中所述三个或更多个非模板核苷酸的至少一个与所述三个或更多个非模板核苷酸的另一个核苷酸不相同。

432. 根据权利要求428所述的组合物, 其中所述第一分子条码化多核苷酸的所述互补区域的至少一个核苷酸与所述第一分子条码化多核苷酸的所述互补区域的另一个核苷酸不相同。

433. 根据权利要求428所述的组合物, 其中所述第二分子条码化多核苷酸的所述互补区域的至少一个核苷酸与所述第二分子条码化多核苷酸的所述互补区域的另一个核苷酸不相同。

434. 根据权利要求432或433所述的组合物, 其中所述至少一个不相同的核苷酸为脱氧核糖核苷酸或其类似物。

435. 根据权利要求432或433所述的组合物, 其中所述至少一个不相同的核苷酸不是核糖核苷酸或其类似物。

436. 根据权利要求432或433所述的组合物, 其中所述至少一个不相同的核苷酸为脱氧核糖鸟苷。

437. 根据权利要求432或433所述的组合物, 其中所述至少一个不相同的核苷酸为脱氧

核糖鸟苷类似物。

438. 根据权利要求432或433所述的组合物,其中所述至少一个不相同的核苷酸为所述第一或第二分子条码化多核苷酸的末端核苷酸。

439. 根据权利要求432或433所述的组合物,其中所述至少一个不相同的核苷酸为核糖核苷酸或其类似物。

440. 根据权利要求428所述的组合物,其中所述第一或第二分子条码化多核苷酸的所述互补区域的末端核苷酸为脱氧核糖核苷酸或其类似物。

441. 根据权利要求428所述的组合物,其中所述第一或第二分子条码化多核苷酸的杂交区域的末端核苷酸不是核糖核苷酸或其类似物。

442. 根据权利要求428所述的组合物,其中所述第一或第二分子条码化多核苷酸的杂交区域的末端核苷酸为脱氧核糖鸟苷。

443. 根据权利要求428所述的组合物,其中所述第一或第二分子条码化多核苷酸的杂交区域的末端核苷酸为脱氧核糖鸟苷类似物。

444. 根据权利要求428所述的组合物,其中所述第一或第二分子条码化多核苷酸的杂交区域的末端核苷酸为核糖核苷酸或其类似物。

445. 根据权利要求428所述的组合物,其中所述第一或第二分子条码化多核苷酸的杂交区域的至少两个非末端核苷酸为核糖核苷酸或其类似物。

446. 根据权利要求428所述的组合物,其中所述第一或第二分子条码化多核苷酸的杂交区域的至少两个非末端核苷酸不是脱氧核糖核苷酸或其类似物。

447. 根据权利要求428所述的组合物,其中所述第一或第二分子条码化多核苷酸的杂交区域的至少两个非末端核苷酸为脱氧核糖核苷酸或其类似物。

448. 根据权利要求368所述的组合物,其中所述第一互补多核苷酸包含与第一分子条码化多核苷酸互补的区域。

449. 根据权利要求448所述的组合物,其中所述第二互补多核苷酸包含与第二分子条码化多核苷酸互补的区域。

450. 根据权利要求449所述的组合物,其中所述第一互补多核苷酸包含与所述第二分子条码化多核苷酸互补的区域。

451. 根据权利要求450所述的组合物,其中与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的所述区域不与分子条码序列互补。

452. 根据权利要求450所述的组合物,其中与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的所述区域不与所述容器条码化多核苷酸或其扩增产物的区域互补。

453. 根据权利要求448所述的组合物,其中与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的所述区域包含添加至所述第一互补多核苷酸的3'端的三个或更多个非模板核苷酸。

454. 根据权利要求453所述的组合物,其中与所述第二分子条码化多核苷酸互补的所述第二互补多核苷酸的所述区域包含添加至所述第二互补多核苷酸的3'端的三个或更多个非模板核苷酸。

455. 根据权利要求448所述的组合物,其中所述第一互补多核苷酸不与所述容器条码

化多核苷酸互补。

456. 根据权利要求455所述的组合物,其中所述第二互补多核苷酸不与所述容器条码化多核苷酸互补。

457. 根据权利要求368所述的组合物,其中第一分子条码化多核苷酸的互补体的区域与所述容器条码化多核苷酸的区域互补。

458. 根据权利要求457所述的组合物,其中第二分子条码化多核苷酸的互补体的区域与所述容器条码化多核苷酸的区域互补。

459. 根据权利要求368所述的组合物,其中所述组合物进一步包含来自以上所述方法的任何一个或多个引物。

460. 根据权利要求368所述的组合物,其中所述多个容器的每个容器不包含固体支持体。

461. 根据权利要求368所述的组合物,其中所述容器条码化多核苷酸附接至固体支持体。

462. 根据权利要求368所述的组合物,其中所述容器条码化多核苷酸附接至珠子。

463. 根据权利要求368所述的组合物,其中所述容器条码化多核苷酸、分子条码化多核苷酸或其任意组合不是引物。

464. 根据权利要求368所述的组合物,其中所述容器条码化多核苷酸、分子条码化多核苷酸或其任意组合不是延伸的多核苷酸。

465. 根据权利要求368所述的组合物,其中所述单细胞为裂解的。

466. 根据权利要求368所述的组合物,其中所述多个容器包含多个孔。

467. 根据权利要求368所述的组合物,其中所述多个容器包含多份乳液。

468. 根据权利要求467所述的组合物,其中所述多份乳液的每份乳液的体积为0.01皮升至10微升。

469. 根据权利要求368所述的组合物,其中所述单细胞包括免疫细胞。

470. 根据权利要求368所述的组合物,其中所述多个细胞包括多个免疫细胞。

471. 根据权利要求469所述的组合物,其中所述免疫细胞为淋巴细胞或其亚型、B细胞或其亚型,或T细胞或其亚型。

472. 根据权利要求470所述的组合物,其中所述多个细胞针对记忆B细胞、初始B细胞、成浆B细胞、初始T细胞、成浆T细胞、B细胞的任何亚型、T细胞的任何亚型或其任意组合进行富集。

473. 根据权利要求368所述的组合物,其中所述单细胞包括癌细胞。

474. 根据权利要求368所述的组合物,其中所述多个细胞包括多个癌细胞。

475. 根据权利要求473所述的组合物,其中所述癌细胞为鳞状细胞癌细胞、腺癌细胞、移行细胞癌细胞、骨肉瘤细胞、软骨肉瘤细胞、肌肉肉瘤细胞、白血病细胞、淋巴瘤细胞或神经胶质瘤细胞。

476. 根据权利要求474所述的组合物,其中所述多个癌细胞针对循环癌细胞、内皮癌细胞、上皮癌细胞、罕见癌细胞或癌细胞的任何类型或亚型进行富集。

477. 根据权利要求368所述的组合物,其中所述第一或第二细胞多核苷酸包含变异序列。

478. 根据权利要求477所述的组合物,其中所述变异序列包括突变、多态性、缺失或插入。

479. 根据权利要求478所述的组合物,其中所述多态性为单核苷酸多态性。

480. 根据权利要求368所述的组合物,其中所述第一或第二细胞多核苷酸为对于疾病或病况的生物标志物。

481. 根据权利要求368所述的组合物,其中所述第一或第二细胞多核苷酸来自病原体。

482. 根据权利要求368所述组合物法,其中所述第一和第二互补多核苷酸包括遍布抗体或TCR编码序列中的CDR1、CDR2、CDR3和/或超突变区域。

483. 根据权利要求368所述的组合物,其中所述容器条码包含至少2个核苷酸。

484. 根据权利要求368所述的组合物,其中所述容器条码包含至少3、4、5、6、7、8或9个核苷酸。

485. 根据权利要求368所述的组合物,其中所述容器条码包含至少10个核苷酸。

486. 根据权利要求368所述的组合物,其中所述容器条码包含至少15个核苷酸。

487. 根据权利要求368所述的组合物,其中所述容器条码包含最多50个核苷酸。

488. 根据权利要求368所述的组合物,其中所述容器条码包含10-30个核苷酸。

489. 根据权利要求368所述的组合物,其中所述容器条码包含简并序列。

490. 根据权利要求368所述的组合物,其中所述容器条码包含完整或部分的简并序列。

491. 根据权利要求490所述的组合物,其中所述容器条码包含序列NNNNNNNNNNNNNNN,其中N为任何核酸。

492. 根据权利要求490所述的组合物,其中所述容器条码包含序列NNNNNNNNNNNNNNNN,其中N为任何核酸,且W为腺嘌呤或胸腺嘧啶。

493. 根据权利要求490所述的组合物,其中所述容器条码包含序列NNNNNNNNNNNNNNNN,其中N为任何核酸,且X为任何已知的核苷酸。

494. 根据权利要求490所述的组合物,其中所述容器条码包含序列NNNNNNNNNNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为W,其中W为腺嘌呤或胸腺嘧啶。

495. 根据权利要求490所述的组合物,其中所述容器条码包含序列NNNNNNNNNNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为X,其中X为任何已知的核苷酸。

496. 根据权利要求368所述的组合物,其中所述分子条码包含至少2个核苷酸。

497. 根据权利要求368所述的组合物,其中所述分子条码包含至少3、4、5、6、7、8或9个核苷酸。

498. 根据权利要求368所述的组合物,其中所述分子条码包含至少10个核苷酸。

499. 根据权利要求368所述的组合物,其中所述分子条码包含至少15个核苷酸。

500. 根据权利要求368所述的组合物,其中所述分子条码包含最多50个核苷酸。

501. 根据权利要求368所述的组合物,其中所述分子条码包含10-30个核苷酸。

502. 根据权利要求368所述的组合物,其中所述分子条码包含简并序列。

503. 根据权利要求368所述的组合物,其中所述分子条码包含完整或部分的简并序列。

504. 根据权利要求503所述的组合物,其中所述分子条码包含序列NNNNNNNN,其中N为

任何核酸。

505. 根据权利要求503所述的组合物,其中所述分子条码包含序列NNTNNANN,其中N为任何核酸。

506. 根据权利要求503所述的组合物,其中所述分子条码包含序列NNWNNWNN,其中N为任何核酸,且W为腺嘌呤或胸腺嘧啶。

507. 根据权利要求503所述的组合物,其中所述分子条码包含序列NNXNNXNN,其中N为任何核酸,且X为任何已知的核苷酸。

508. 根据权利要求503所述的组合物,其中所述分子条码包含序列NNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N 为W,其中W为腺嘌呤或胸腺嘧啶。

509. 根据权利要求503所述的组合物,其中所述分子条码包含序列NNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为X,其中X为任何已知的核苷酸。

510. 根据权利要求368所述的组合物,其中所述多个容器包括至少3、4、5、5.5、6、6.5、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 或 9×10^{12} 个或更多容器。

511. 根据权利要求368所述的组合物,其中所述多个细胞多核苷酸包括至少3、4、5、5.5、6、6.5、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 或 9×10^{12} 个或更多细胞多核苷酸。

512. 一种非诊断目的的多核苷酸条码化方法,包括:

(a) 使分子条码化多核苷酸与来自单细胞的多个多核苷酸中的每个杂交,其中所述杂交分子条码化多核苷酸来自包含所述单细胞的容器内多个独特的分子条码化多核苷酸;

(b) 延伸与所述分子条码化多核苷酸杂交的来自所述单细胞的多核苷酸,以形成分子条码细胞多核苷酸;

(c) 使容器条码化多核苷酸与所述分子条码细胞多核苷酸杂交,其中所述容器条码化

多核苷酸对于多个容器中的单个容器是独特的；

(d) 延伸与所述容器条码化多核苷酸杂交的所述分子条码细胞多核苷酸，以形成双条码化细胞多核苷酸；以及

(e) 对所述双条码化细胞多核苷酸测序。

513. 根据权利要求512所述的方法，其中 (a) 中的所述杂交不是通过在来自所述单细胞的所述多核苷酸上天然存在的序列的碱基配对进行的。

514. 根据权利要求512或513所述的方法，其中与所述分子条码细胞多核苷酸杂交的所述容器条码化多核苷酸为扩增产物。

515. 根据权利要求512或513所述的方法，其中 (c) 中的所述杂交不是通过在来自所述单细胞的所述多核苷酸上天然存在的序列的互补体的碱基配对进行的。

516. 根据权利要求512或513所述的方法，其中 (c) 中的所述杂交不是通过与来自 (b) 中延伸的所述单细胞的所述多核苷酸的区域的碱基配对进行的。

517. 根据权利要求1所述的方法，其中所述第一和第二分子条码化多核苷酸不融合在一起。

518. 根据权利要求93所述的方法，其中所述至少一个不相同的核苷酸为脱氧核糖核苷酸或其类似物。

519. 根据权利要求92所述的方法，其中所述至少一个不相同的核苷酸为所述第一或第二分子条码化多核苷酸的末端核苷酸。

520. 根据权利要求87所述的方法，其中所述第二分子条码化多核苷酸的杂交区域的末端核苷酸为脱氧核糖核苷酸或其类似物。

521. 根据权利要求195所述的方法，其中所述裂解包括化学裂解。

522. 根据权利要求351所述的方法，其中所述比对在计算机或算法的帮助下进行。

523. 根据权利要求470所述的组合物，其中所述免疫细胞为淋巴细胞或其亚型、B细胞或其亚型、T细胞或其亚型或其组合。

524. 根据权利要求474所述的组合物，其中所述癌细胞为鳞状细胞癌细胞、腺癌细胞、移行细胞癌细胞、骨肉瘤细胞、软骨肉瘤细胞、肌肉肉瘤细胞、白血病细胞、淋巴瘤细胞、神经胶质瘤细胞或其任意组合。

高通量核苷酸文库测序

[0001] 交叉引用

[0002] 本申请要求于2014年9月15日提交的美国临时申请号62/050,549和于2014年9月17日提交的美国临时申请号62/051,832的优先权,其各自通过引用而全文并入本文。

背景技术

[0003] 目前的抗体展示技术(噬菌体、酵母、核糖体、哺乳动物等)是有局限性的,因为所选择的抗体候选物的质量受限于它们从中生成的起始库。诸如组合和“智能”抗体设计途径以及杂交瘤发现途径等途径通常产生表现出下游并发问题的合成抗体,该并发问题包括大规模表达的困难,患者中高风险的免疫原性,以及除了高结合亲和力以外缺乏足够的免疫功能。极少数来源于展示技术的抗体在过去十年中已经成功通过了临床试验,甚至证明有积极的临床前特性。目前,预测或了解特定抗体序列识别并激活对抗外来目标的免疫应答的机制的能力仍然是难以把握。因此,在本领域中对于发现并生成具有高结合亲和力、可大规模生成并具有足够免疫功能的抗体的方法存在需求。本文所述的方法旨在利用数百万年的免疫组库(immune repertoire)演化来满足这些需要,并促进对这些概念和它们怎样参与抗体生成的理解。本文所述的方法可用来产生用于选择高质量的抗体候选物的抗体序列和/或抗体文库。

[0004] 人类抗体组库在其复杂性和大小上几乎是无限的。因此,已在统计学上证明了组合文库极少产生正确的重链(V_H)或轻链(V_L)配对。其他研究者着眼于仅仅改组最常表达的互补决定区(CDR)的框架家族(诸如V3-23、V1-69或匹配 V_H 和 V_L 的频率),并因此将组库多样性限制在可控制的大小。根据预期,最常表达的家族在免疫应答期间将更频繁地被选择并演化。令人惊讶的是,通过使用人类抗体组库的免疫测序,已经发现在抗体框架表达频率与抗体响应于免疫激发的激活潜力之间没有关联。本文所述的方法可用于设计和/或生成非限制性抗体文库以克服这些对于抗体发现和选择的挑战。可针对个性化医疗生成自身免疫、癌症、传染性和正常/健康供体库,以应对基本的尚未得到满足的生物需求。

发明内容

[0005] 在一个方面,本文提供了一种方法,其包括:形成多个容器,每个容器包含来自包含多个细胞的样品的单细胞,多个分子条码化多核苷酸,以及容器条码化多核苷酸;产生:与来自所述单细胞的第一细胞多核苷酸互补的第一互补多核苷酸,以及与来自所述单细胞的第二细胞多核苷酸互补的第二互补多核苷酸;附接:所述多个分子条码化多核苷酸的第一分子条码化多核苷酸至所述第一互补多核苷酸,以及第二分子条码化多核苷酸至所述第二互补多核苷酸,从而形成第一和第二单细胞单条码化多核苷酸;以及将所述容器条码化多核苷酸或其扩增产物附接至所述第一单细胞单条码化多核苷酸,以及所述第二单细胞单条码化多核苷酸,从而形成第一和第二单细胞双条码化序列。

[0006] 在一个方面,本文提供了一种组合物,其包含:多个容器,每个容器包含来自包含多个细胞的样品的单细胞,多个分子条码化多核苷酸,容器条码化多核苷酸;与来自所述单

细胞的第一细胞多核苷酸互补的第一互补多核苷酸,以及与来自所述单细胞的第二细胞多核苷酸互补的第二互补多核苷酸;其中所述第一互补多核苷酸包含所述多个分子条码化多核苷酸的第一分子条码和所述容器条码化多核苷酸的扩增产物或所述容器条码化多核苷酸的容器条码,并且其中所述第二互补多核苷酸包含所述多个分子条码化多核苷酸的第二分子条码和所述容器条码化多核苷酸的扩增产物或所述容器条码化多核苷酸的容器条码。

[0007] 在一个方面,本文提供了一种方法,其包括:(a)形成多个容器,每个容器包含来自包含多个细胞的样品的单细胞,多个分子条码化多核苷酸,以及容器条码化多核苷酸;(b)产生:与来自所述单细胞的第一细胞多核苷酸互补的第一互补多核苷酸,以及与来自所述单细胞的第二细胞多核苷酸互补的第二互补多核苷酸;(c)衔接:所述多个分子条码化多核苷酸的第一分子条码化多核苷酸至所述第一互补多核苷酸,以及第二分子条码化多核苷酸至所述第二互补多核苷酸,从而形成第一和第二单细胞单条码化多核苷酸;以及(d)将所述容器条码化多核苷酸或其扩增产物衔接至所述第一单细胞单条码化多核苷酸或其扩增产物,以及所述第二单细胞单条码化多核苷酸或其扩增产物,从而形成第一和第二单细胞双条码化序列。

[0008] 在一个方面,本文提供了一种方法,其包括:(a)从来自样品的多个免疫细胞产生来自重链免疫球蛋白(IgH)多核苷酸的第一互补多核苷酸和来自轻链免疫球蛋白(IgL)多核苷酸的第二互补多核苷酸,其中采用:第一目标引物,其包含与来自所述多个免疫细胞的所述IgH多核苷酸的相同区域互补的区域;第二目标引物,其包含与来自所述多个免疫细胞的所述IgL多核苷酸的相同区域互补的区域;包含非模板末端转移酶活性的逆转录酶,其中将3个或更多个相同的非模板核苷酸添加至所述第一和第二互补多核苷酸的3'端;多个分子条码化多核苷酸,每个包含:分子条码,与容器条码化多核苷酸的区域互补的5'端区域,以及与所述3个或更多个非模板核苷酸互补的3'端区域;以及容器条码化多核苷酸,从而形成第一和第二单细胞单条码化多核苷酸;(b)扩增所述容器条码化多核苷酸,从而形成第一和第二单细胞双条码化多核苷酸;(c)扩增所述第一和第二单细胞双条码化多核苷酸,从而形成包含所述IgH或IgL多核苷酸的可变区或其组合的序列的文库;以及(d)对所述文库的一个或多个所述序列进行测序,其中(a)在多个容器的容器中进行,其中所述容器包含来自所述多个免疫细胞的单个免疫细胞。

[0009] 在一个方面,本文提供了一种方法,其包括:(a)从来自样品的多个免疫细胞产生来自T细胞受体 α (TCR α)多核苷酸的第一互补多核苷酸和来自T细胞受体 β (TCR β)多核苷酸的第二互补多核苷酸,其中采用:第一目标引物,其包含与来自所述多个免疫细胞的所述TCR α 多核苷酸的相同区域互补的区域;第二目标引物,其包含与来自所述多个免疫细胞的所述TCR β 多核苷酸的相同区域互补的区域;包含非模板末端转移酶活性的逆转录酶,其中将3个或更多个相同的非模板核苷酸添加至所述第一和第二互补多核苷酸的3'端;多个分子条码化多核苷酸,每个包含:分子条码,与容器条码化多核苷酸的区域互补的5'端区域,以及与所述3个或更多个非模板核苷酸互补的3'端区域;以及容器条码化多核苷酸,从而形成第一和第二单细胞单条码化多核苷酸;(b)扩增所述容器条码化多核苷酸,从而形成第一和第二单细胞双条码化多核苷酸;(c)扩增所述第一和第二单细胞双条码化多核苷酸,从而形成包含所述TCR α 或TCR β 多核苷酸的可变区或其组合的序列的文库;以及(d)对所述文库的一个或多个所述序列进行测序,其中(a)在多个容器的容器中进行,其中所述容器包含来

自所述多个免疫细胞的单个免疫细胞。

[0010] 在一个方面,本文提供了一种方法,其包括:(a)从来自样品的多个免疫细胞产生来自T细胞受体 γ (TCR γ)多核苷酸的第一互补多核苷酸和来自T细胞受体 δ (TCR δ)多核苷酸的第二互补多核苷酸,其中采用:第一目标引物,其包含与来自所述多个免疫细胞的所述TCR γ 多核苷酸的相同区域互补的区域;第二目标引物,其包含与来自所述多个免疫细胞的所述TCR δ 多核苷酸的相同区域互补的区域;包含非模板末端转移酶活性的逆转录酶,其中将3个或更多个相同的非模板核苷酸添加至所述第一和第二互补多核苷酸的3'端;多个分子条码化多核苷酸,每个包含:分子条码,与容器条码化多核苷酸的区域互补的5'端区域,以及与所述3个或更多个非模板核苷酸互补的3'端区域;以及容器条码化多核苷酸,从而形成第一和第二单细胞单条码化多核苷酸;(b)扩增所述容器条码化多核苷酸,从而形成第一和第二单细胞双条码化多核苷酸;(c)扩增所述第一和第二单细胞双条码化多核苷酸,从而形成包含所述TCR γ 或TCR δ 多核苷酸的可变区或其组合的序列的文库;以及(d)对所述文库的一个或多个所述序列进行测序,其中(a)在多个容器的容器中进行,其中所述容器包含来自所述多个免疫细胞的单个免疫细胞。

[0011] 在一些实施方案中,所述文库代表所述样品的免疫状态。在一些实施方案中,所述第一和第二单细胞双条码化序列为第一和第二单细胞双条码化序列的文库。在一些实施方案中,所述第一和第二分子条码化多核苷酸的分子条码不同。在一些实施方案中,所述第一和第二单细胞单条码化多核苷酸包含不同的分子条码。在一些实施方案中,所述第一和第二单细胞双条码化序列包含不同的分子条码。在一些实施方案中,所述第一和第二单细胞双条码化序列包含相同的容器条码。在一些实施方案中,所述多个分子条码化多核苷酸不是扩增产物。在一些实施方案中,第一容器中分子条码化多核苷酸的分子条码与第二容器中分子条码化多核苷酸的分子条码不同。在一些实施方案中,所述多个容器的第一容器中每个分子条码化多核苷酸的分子条码是独特的。在一些实施方案中,所述多个容器的第二容器中每个分子条码化多核苷酸的分子条码是独特的。在一些实施方案中,第一容器和第二容器中每个分子条码化多核苷酸的分子条码是独特的。在一些实施方案中,所述多个容器的第三容器中每个分子条码化多核苷酸的分子条码是独特的。在一些实施方案中,所述第一容器、所述第二容器和所述第三容器中每个分子条码化多核苷酸的分子条码是独特的。在一些实施方案中,所述多个容器的任何单个容器中每个分子条码化多核苷酸的分子条码是独特的。在一些实施方案中,所述多个容器的任何一个容器中每个分子条码化多核苷酸的分子条码与所述多个容器的任何其他一个容器中每个分子条码化多核苷酸的分子条码不同。在一些实施方案中,第一容器中分子条码化多核苷酸的分子条码与第二容器中分子条码化多核苷酸的分子条码相同。在一些实施方案中,第一容器中分子条码化多核苷酸的分子条码与所述第一容器中分子条码化多核苷酸的分子条码相同。在一些实施方案中,第二容器中分子条码化多核苷酸的分子条码与所述第二容器中分子条码化多核苷酸的分子条码相同。在一些实施方案中,所述多个容器的第一容器中容器条码化多核苷酸或其扩增子的容器条码与所述多个容器的第二容器中容器条码化多核苷酸或其扩增子的容器条码不同。在一些实施方案中,所述多个容器的第一容器中容器条码化多核苷酸或其扩增子的容器条码为第一相同容器条码。在一些实施方案中,所述多个容器的第二容器中每个容器条码化多核苷酸或其扩增子的容器条码为第二相同容器条码。在一些实施方案中,所

述第一相同容器条码与所述第二相同容器条码不同。在一些实施方案中,所述多个容器的单个容器中每个容器条码化多核苷酸或其扩增子的容器条码包含相同的容器条码。在一些实施方案中,所述多个容器的任何单个容器中每个容器条码化多核苷酸及其扩增子的容器条码对于所述多个容器的任何其他单个容器中每个容器条码化多核苷酸及其扩增子的容器条码而言是独特的。

[0012] 在一些实施方案中,(a)中的容器条码化多核苷酸以单分子存在于容器中。在一些实施方案中,(a)中的容器条码化多核苷酸以单分子存在于所述多个容器的每个容器中。在一些实施方案中,(a)中的容器条码化多核苷酸以至少单分子存在于所述多个容器的容器中。在一些实施方案中,(a)中的容器条码化多核苷酸以至少单分子存在于所述多个容器的每个容器中。

[0013] 在一些实施方案中,所述多个容器的第一容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述第一容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。在一些实施方案中,所述多个容器的所述第一容器中所述第一容器条码化多核苷酸或其扩增子的第二共同容器序列与所述第一容器中第二容器条码化多核苷酸或其扩增子的第二共同容器序列相同。在一些实施方案中,所述多个容器的任何单个容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述单个容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。在一些实施方案中,所述多个容器的单个容器中每个容器条码化多核苷酸包含相同的第一共同容器序列。在一些实施方案中,所述多个容器的单个容器中每个容器条码化多核苷酸包含相同的第二共同容器序列。在一些实施方案中,所述多个容器的第一容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述多个容器的第二容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。在一些实施方案中,所述第一容器条码化多核苷酸或其扩增子的第二共同容器序列与所述第二容器条码化多核苷酸或其扩增子的第二共同容器序列相同。在一些实施方案中,所述多个容器的任何一个容器中每个容器条码化多核苷酸或其扩增子包含第一共同容器序列,该第一共同容器序列包含与所述多个容器的任何其他一个容器中容器条码化多核苷酸或其扩增子的第一共同容器序列相同的序列。在一些实施方案中,所述多个容器的任何一个容器中每个容器条码化多核苷酸或其扩增子包含第二共同容器序列,该第二共同容器序列包含与所述多个容器的任何其他一个容器中容器条码化多核苷酸或其扩增子的第二共同容器序列相同的序列。在一些实施方案中,所述多个容器的第一容器中第一分子条码化多核苷酸的第一共同分子序列与所述第一容器中第二分子条码化多核苷酸的第一共同分子序列相同。在一些实施方案中,所述多个容器的所述第一容器中所述第一分子条码化多核苷酸的第二共同分子序列与所述第一容器中第二分子条码化多核苷酸的第二共同分子序列相同。在一些实施方案中,所述多个容器的任何单个容器中第一分子条码化多核苷酸的第一共同分子序列与所述单个容器中第二分子条码化多核苷酸的第一共同分子序列相同。在一些实施方案中,所述多个容器的单个容器中每个分子条码化多核苷酸包含相同的第一共同分子序列。在一些实施方案中,所述多个容器的单个容器中每个分子条码化多核苷酸包含相同的第二共同分子序列。在一些实施方案中,所述多个容器的第一容器中第一分子条码化多核苷酸的第一共同分子序列与所述多个容器的第二容器中第二分子条码化多核苷酸的第一共同分子序列相同。在一些实施方案中,所述第一分子

条码化多核苷酸的第二共同分子序列与所述第二分子条码化多核苷酸的第二共同分子序列相同。在一些实施方案中,所述多个容器的任何一个容器中每个分子条码化多核苷酸包含第一共同分子序列,该第一共同分子序列包含与所述多个容器的任何其他一个容器中分子条码化多核苷酸的第一共同分子序列相同的序列。在一些实施方案中,所述多个容器的任何一个容器中每个分子条码化多核苷酸包含第二共同分子序列,该第二共同分子序列包含与所述多个容器的任何其他一个容器中分子条码化多核苷酸的第二共同分子序列相同的序列。在一些实施方案中,所述第一共同容器序列包含的序列包含与所述第一共同分子序列相同的序列。在一些实施方案中,所述第一共同容器序列包含与所述第一共同分子序列或其互补体互补的序列。在一些实施方案中,所述第二共同分子序列包含与添加至所述第一互补多核苷酸的3'端的三个或更多个非模板核苷酸互补的区域。在一些实施方案中,所述与添加至所述第一互补多核苷酸的3'端的三个或更多个非模板核苷酸互补的区域为末端区域。

[0014] 在一些实施方案中,第一和第二分子条码化多核苷酸不融合在一起。在一些实施方案中,所述第一和第二单细胞单条码化多核苷酸不融合在一起。在一些实施方案中,所述第一和第二单细胞双条码化序列不融合在一起。

[0015] 在一些实施方案中,所述第一细胞多核苷酸为DNA。在一些实施方案中,所述第二细胞多核苷酸为DNA。在一些实施方案中,所述第一细胞多核苷酸为RNA。在一些实施方案中,所述第二细胞多核苷酸为RNA。在一些实施方案中,所述RNA为mRNA。在一些实施方案中,(b)的第一互补多核苷酸为cDNA。在一些实施方案中,(b)的第二互补多核苷酸为cDNA。

[0016] 在一些实施方案中,(b)包括延伸与所述第一细胞多核苷酸杂交的第一目标引物,以及延伸与所述第二细胞多核苷酸杂交的第二目标引物。在一些实施方案中,所述延伸包括用第一目标引物逆转录所述第一细胞多核苷酸,以及用第二目标引物逆转录所述第二细胞多核苷酸。在一些实施方案中,所述第一目标引物包含与所述第一细胞多核苷酸的目标序列互补的序列。在一些实施方案中,所述第二目标引物包含与所述第二细胞多核苷酸的目标序列互补的序列。在一些实施方案中,所述第一目标引物包含聚(T)序列。在一些实施方案中,所述第二目标引物包含聚(T)序列。在一些实施方案中,所述第一细胞多核苷酸的目标序列为重链免疫球蛋白(IgH)序列、TCR α 序列、TCR γ 序列或其组合。在一些实施方案中,所述第一细胞多核苷酸的目标序列为重链恒定区(C_H)序列、TCR α 恒定区(C α)序列、TCR γ 恒定区(C γ)序列或其组合。在一些实施方案中,所述第二细胞多核苷酸的目标序列为轻链免疫球蛋白(IgL)序列、TCR β 序列、TCR δ 序列或其组合。在一些实施方案中,所述第二细胞多核苷酸的目标序列为轻链恒定区(C_L)序列、TCR β 恒定区(C β)序列、TCR δ 恒定区(C δ)序列或其组合。在一些实施方案中,所述第一目标引物包含多个第一目标引物。在一些实施方案中,所述第二目标引物包含多个第二目标引物。在一些实施方案中,所述多个第一目标引物包含与多个重链免疫球蛋白(IgH)序列、TCR α 序列、TCR γ 序列或其组合互补的多个序列。在一些实施方案中,所述多个重链免疫球蛋白(IgH)序列、TCR α 序列或TCR γ 序列包含多个重链恒定区(C_H)序列、TCR α 恒定区(C α)序列、TCR γ 恒定区(C γ)序列或其组合。在一些实施方案中,所述多个重链恒定区(C_H)序列包含两个或更多个选自来自IgM、IgD、IgA、IgE、IgG及其组合的重链恒定区(C_H)序列的序列。在一些实施方案中,所述多个第二目标引物包含与多个轻链免疫球蛋白(IgL)序列、TCR β 序列、TCR δ 序列或其组合互补的多个序列。在一些实

施方案中,所述多个轻链免疫球蛋白(IgL)序列、TCRB序列或TCR δ 序列包含多个轻链恒定区(C_L)序列、TCR β 恒定区(C β)序列、TCR δ 恒定区(C δ)序列或其组合。在一些实施方案中,所述多个轻链恒定区(C_L)序列包含两个或更多个选自来自Ig κ 、Ig λ 及其组合的轻链恒定区(C_L)序列的序列。在一些实施方案中,在(b)中,所述延伸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添加至所述第一互补多核苷酸的3'端。在一些实施方案中,所述非模板末端转移酶为逆转录酶或聚合酶。在一些实施方案中,所述非模板末端转移酶为逆转录酶,并且其中该逆转录酶选自Superscript II逆转录酶、Maxima逆转录酶、Protoscript II逆转录酶、莫洛尼(moloney)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。在一些实施方案中,将三个或更多个非模板核苷酸添加至所述第二互补多核苷酸的3'端。

[0017] 在一些实施方案中,在(c)中,所述附接包括使第一分子条码化多核苷酸的区域与添加至所述第一互补多核苷酸的3'端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,在(c)中,所述附接包括使第二分子条码化多核苷酸的区域与添加至所述第二互补多核苷酸的3'端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,在(c)中,附接至所述第一互补多核苷酸的第一分子条码化多核苷酸包含与所述第一互补多核苷酸的3'端上的所述三个或更多个非模板核苷酸互补的区域。在一些实施方案中,在(c)中,附接至所述第二互补多核苷酸的第二分子条码化多核苷酸包含与所述第二互补多核苷酸的3'端上的三个或更多个非模板核苷酸互补的区域。在一些实施方案中,所述三个或更多个非模板核苷酸是相同的。在一些实施方案中,所述三个或更多个非模板核苷酸中的至少一个与所述三个或更多个非模板核苷酸中的另一个核苷酸不相同。在一些实施方案中,所述第一分子条码化多核苷酸的杂交区域的至少一个核苷酸与所述第一分子条码化多核苷酸的杂交区域的另一个核苷酸不相同。在一些实施方案中,所述第二分子条码化多核苷酸的杂交区域的至少一个核苷酸与所述第二分子条码化多核苷酸的杂交区域的另一个核苷酸不相同。在一些实施方案中,所述至少一个不相同的核苷酸为脱氧核糖核苷酸或其类似物。在一些实施方案中,所述至少一个不相同的核苷酸不是核糖核苷酸或其类似物。在一些实施方案中,所述至少一个不相同的核苷酸为脱氧核糖鸟苷。在一些实施方案中,所述至少一个不相同的核苷酸为脱氧核糖鸟苷类似物。在一些实施方案中,所述至少一个不相同的核苷酸为所述第一或第二分子条码化多核苷酸的末端核苷酸。在一些实施方案中,所述至少一个不相同的核苷酸为核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的杂交区域的末端核苷酸为脱氧核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的杂交区域的末端核苷酸不是核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的杂交区域的末端核苷酸为脱氧核糖鸟苷。在一些实施方案中,所述第一或第二分子条码化多核苷酸的杂交区域的末端核苷酸为脱氧核糖鸟苷类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的杂交区域的末端核苷酸为核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的杂交区域的至少两个非末端核苷酸为核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的杂交区域的至少两个非末端核苷酸不是脱氧核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸

的杂交区域的至少两个非末端核苷酸为脱氧核糖核苷酸或其类似物。在一些实施方案中，(c) 进一步包括在所述附接后延伸所述第一互补多核苷酸和所述第二互补多核苷酸。在一些实施方案中，所述第一互补多核苷酸包含与第一分子条码化多核苷酸互补的区域。在一些实施方案中，所述第二互补多核苷酸包含与第二分子条码化多核苷酸互补的区域。在一些实施方案中，所述第一互补多核苷酸包含与第二分子条码化多核苷酸互补的区域。在一些实施方案中，与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的区域不与分子条码序列互补。在一些实施方案中，与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的区域不与所述容器条码化多核苷酸或其扩增产物的区域互补。在一些实施方案中，与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的区域包含添加至所述第一互补多核苷酸的3' 端的三个或更多个非模板核苷酸。在一些实施方案中，与所述第二分子条码化多核苷酸互补的所述第二互补多核苷酸的区域包含添加至所述第二互补多核苷酸的3' 端的三个或更多个非模板核苷酸。在一些实施方案中，所述第一互补多核苷酸不与所述容器条码化多核苷酸互补。在一些实施方案中，所述第二互补多核苷酸不与所述容器条码化多核苷酸互补。在一些实施方案中，第一分子条码化多核苷酸的互补体的区域与所述容器条码化多核苷酸的区域互补。在一些实施方案中，第二分子条码化多核苷酸的互补体的区域与所述容器条码化多核苷酸的区域互补。在一些实施方案中，所述第一单细胞单条码化多核苷酸的区域与所述容器条码化多核苷酸的区域互补。在一些实施方案中，所述第二单细胞单条码化多核苷酸的区域与所述容器条码化多核苷酸的区域互补。在一些实施方案中，所述第一单细胞单条码化多核苷酸的区域与所述容器条码化多核苷酸中与所述第二单细胞单条码化多核苷酸互补的区域互补。在一些实施方案中，所述方法进一步包括用第一引物组扩增所述容器条码化多核苷酸，其中所述扩增在附接所述容器条码化多核苷酸之前或与附接所述容器条码化多核苷酸同时进行。在一些实施方案中，所述容器条码化多核苷酸包含第一和第二容器条码化多核苷酸，其选自所述容器条码化多核苷酸、所述容器条码化多核苷酸的互补体、来自所述容器条码化多核苷酸的扩增产物及其任意组合。在一些实施方案中，附接所述容器条码化多核苷酸包括：使所述容器条码化多核苷酸或其扩增产物的区域与所述第一单细胞单条码化多核苷酸的区域杂交，以及使所述容器条码化多核苷酸或其扩增产物的区域与所述第二单细胞单条码化多核苷酸的区域杂交。

[0018] 在一些实施方案中，所述方法进一步包括在附接所述容器条码化多核苷酸之后延伸所述第一单细胞单条码化序列和所述第二单细胞单条码化序列多核苷酸，从而形成所述第一和第二单细胞双条码化序列。在一些实施方案中，所述第一单细胞双条码化序列包含与所述容器条码化多核苷酸互补的区域。在一些实施方案中，所述第二单细胞双条码化序列包含与所述容器条码化多核苷酸互补的区域。在一些实施方案中，与所述容器条码化多核苷酸互补的所述第一和第二单细胞双条码化序列的区域为相同序列。在一些实施方案中，与所述第一或第二分子条码化多核苷酸互补的所述第一单细胞单条码化多核苷酸的区域不与所述容器条码化多核苷酸或其扩增产物的区域互补。在一些实施方案中，所述第一引物组的第一引物与第一分子条码化多核苷酸、所述第一单细胞单条码化多核苷酸的互补体、所述第一单细胞双条码化序列的互补体或其任意组合的区域互补。在一些实施方案中，所述第一引物组的第一引物与第二分子条码化多核苷酸、所述第二单细胞单条码化多核苷

酸的互补体、所述第二单细胞双条码化序列的互补体或其任意组合的区域互补。在一些实施方案中,所述第一引物组的第一引物不与所述第一细胞多核苷酸或其互补体互补。在一些实施方案中,所述第一引物组的第一引物不与所述第二细胞多核苷酸或其互补体互补。在一些实施方案中,所述第一引物组的第一引物与所述第一单细胞单条码化序列的互补体在所述分子条码的下游的区域互补。在一些实施方案中,所述第一引物组的第一引物与所述第二单细胞单条码化多核苷酸的互补体在所述分子条码的下游的区域互补。在一些实施方案中,所述第一引物组的第一引物与所述第一单细胞双条码化序列的互补体在所述容器条码的上游的区域互补。在一些实施方案中,所述第一引物组的第一引物与所述第二单细胞双条码化多核苷酸的互补体在所述容器条码的上游的区域互补。在一些实施方案中,所述第一引物组的第二引物不与所述第一细胞多核苷酸或其互补体、所述第一互补多核苷酸或其互补体、第一分子条码化多核苷酸或其互补体、所述第一单细胞单条码化多核苷酸或其互补体或其任意组合的区域互补。在一些实施方案中,所述第一引物组的第二引物不与所述第二细胞多核苷酸或其互补体、所述第二互补多核苷酸或其互补体、第二分子条码化多核苷酸或其互补体、所述第二单细胞单条码化多核苷酸或其互补体或其任意组合的区域互补。在一些实施方案中,所述第一引物组的第二引物与所述第一单细胞双条码化序列的区域互补。在一些实施方案中,所述第一引物组的第二引物与所述第二单细胞双条码化序列的区域互补。在一些实施方案中,所述第一引物组的第二引物与所述第一单细胞双条码化序列在所述分子条码的上游的区域互补。在一些实施方案中,所述第一引物组的第二引物与所述第二分子条码化多核苷酸在所述分子条码的上游的区域互补。在一些实施方案中,所述第一引物组的第二引物与所述第一单细胞双条码化序列在所述容器条码的上游的区域互补。在一些实施方案中,所述第一引物组的第二引物与所述第二分子条码化多核苷酸在所述容器条码的上游的区域互补。

[0019] 在一些实施方案中,所述方法进一步包括打破所述多个容器中的两个或更多个容器。在一些实施方案中,所述方法进一步包括合并来自所述两个或更多个打破的容器的所述第一和第二单细胞双条码化序列。

[0020] 在一些实施方案中,所述方法进一步包括扩增所述第一和第二单细胞双条码化序列。在一些实施方案中,所述扩增第一和第二单细胞双条码化序列在所述多个容器的容器之外进行。在一些实施方案中,所述方法进一步包括用第二引物组扩增所述第一和第二单细胞双条码化序列。在一些实施方案中,所述第二引物组的第一引物不与所述第一细胞多核苷酸或其互补体、所述第一互补多核苷酸或其互补体、第一分子条码化多核苷酸或其互补体、所述第一单细胞单条码化多核苷酸或其互补体或其任意组合的区域互补。在一些实施方案中,所述第二引物组的所述第一引物不与所述第二细胞多核苷酸或其互补体、所述第二互补多核苷酸或其互补体、第二分子条码化多核苷酸或其互补体、所述第二单细胞单条码化多核苷酸或其互补体或其任意组合的区域互补。在一些实施方案中,所述第二引物组的第一引物与所述第一单细胞双条码化序列的区域互补。在一些实施方案中,所述第二引物组的所述第一引物与所述第二单细胞双条码化序列的区域互补。在一些实施方案中,所述第二引物组的第一引物与所述第一单细胞双条码化序列在所述分子条码的上游的区域互补。在一些实施方案中,所述第二引物组的所述第一引物与所述第二单细胞双条码化序列在所述分子条码的上游的区域互补。在一些实施方案中,所述第二引物组的第一引物

与所述第一单细胞双条码化序列在所述容器条码的上游的区域互补。在一些实施方案中,所述第二引物组的所述第一引物与所述第二单细胞双条码化序列在所述容器条码的上游的区域互补。在一些实施方案中,所述第一引物组的所述第二引物为所述第二引物组的所述第一引物。在一些实施方案中,所述第二引物组的第二引物与所述第一和第二细胞多核苷酸、所述第一和第二互补多核苷酸的互补体、所述第一和第二单细胞单条码化多核苷酸的互补体、所述第一和第二单细胞双条码化序列的互补体或其任意组合的区域互补。在一些实施方案中,所述第二引物组的所述第二引物包含聚(T)序列。在一些实施方案中,所述第二引物组的第二引物与所述第一或第二细胞多核苷酸、所述第一或第二互补多核苷酸的互补体、所述第一或第二单细胞单条码化多核苷酸的互补体、所述第一或第二单细胞双条码化序列的互补体或其任意组合的区域互补。在一些实施方案中,所述第二引物组的所述第二引物不与第一或第二分子条码化多核苷酸或其互补体、所述容器条码化多核苷酸或其互补体或其任意组合互补。在一些实施方案中,所述第二引物组的第三引物与所述第二细胞多核苷酸、所述第二互补多核苷酸的互补体、所述第二单细胞单条码化多核苷酸的互补体、所述第二单细胞双条码化序列的互补体或其任意组合的区域互补。在一些实施方案中,所述第二引物组的所述第二引物与所述第一细胞多核苷酸、所述第一互补多核苷酸的互补体、所述第一单细胞单条码化多核苷酸的互补体、所述第一单细胞双条码化序列的互补体或其任意组合的区域互补。在一些实施方案中,所述第二引物组的所述第三引物不与所述第一细胞多核苷酸、所述第一互补多核苷酸的互补体、所述第一单细胞单条码化多核苷酸的互补体、所述第一单细胞双条码化序列的互补体或其任意组合互补。在一些实施方案中,所述第二引物组的所述第三引物包含目标特异性序列。在一些实施方案中,所述第二引物组的所述第三引物包含目标特异性序列。在一些实施方案中,所述第二引物组的所述第二引物的所述目标特异性序列靶向重链免疫球蛋白(IgH)序列、TCR α 序列、TCR γ 序列或其组合。在一些实施方案中,所述第二引物组的所述第二引物的所述目标特异性序列靶向重链恒定区序列(C_H)、TCR α 恒定区(C α)序列、TCR γ 恒定区(C γ)序列或其组合。在一些实施方案中,所述第二引物的所述目标特异性序列选自GGGTTGGGCGGATGCAC、CATCCGGAGCCTTGGTGG、CCTTG GGGCTGGTCGGGG、CGGATGGGCTCTGTGTGG、CCGATGGGCCCTTGGTGG、GGATTTAGAGTCTCTCAGCTG、CACGGCAGGGTCAGGGTTC和GGGGAACATCTGCATCAAGT。在一些实施方案中,所述第二引物组的所述第三引物的所述目标特异性序列靶向轻链免疫球蛋白(IgL)序列、TCR β 序列、TCR δ 序列或其组合。在一些实施方案中,所述第二引物组的所述第三引物的所述目标特异性序列靶向轻链恒定区序列(C_L)、TCR β 恒定区(C β)序列、TCR δ 恒定区(C δ)序列或其组合。在一些实施方案中,所述第三引物的所述目标特异性序列选自TTTGATCTCCACCTTGGTCCCTCCGC、TTTGATCTCCAGCTTGGTCCCTGG、TTTGATATCCACTTTGGTCCCAGGGC、TTTGATTTCCACCTTGGTCCCTTGGC、TTTAATCTCCAGTCGTGCCCTTGGC、GAGGACGGTCACCTTGGTGCCA、TAGGACGGTCAGCTTGGTCCCTCC、GAGGACGGTCAGCTGGGTGCC、TAAATGATCAGCTGGGTTCCTCCAC、TAGGACGGTGACCTTGGTCCCAG、GGGAGATCTCTGCTTCTGATG、CGACCTCGGGTGGGAACAC和CGGATGGTTTGGTATGAGGC。在一些实施方案中,所述第二引物组的所述第二引物包括多个第二引物。在一些实施方案中,所述第二引

物组的所述第三引物包括多个第三引物。在一些实施方案中,所述多个第二引物的所述目标特异性序列靶向多个重链免疫球蛋白(IgH)序列、TCR α 序列、TCR γ 序列或其组合。在一些实施方案中,所述多个重链免疫球蛋白(IgH)序列、TCR α 序列或TCR γ 序列包括多个重链恒定区(C_H)、TCR α 恒定区(C α)序列、TCR γ 恒定区(C γ)序列或其组合。在一些实施方案中,所述多个重链恒定区(C_H)序列包括两个或更多个序列,所述序列选自来自IgM、IgD、IgA、IgE、IgG及其组合的重链恒定区(C_H)序列。在一些实施方案中,所述多个第三引物的所述目标特异性序列靶向多个轻链免疫球蛋白(IgL)序列、TCR β 序列、TCR δ 序列或其组合。在一些实施方案中,所述多个轻链免疫球蛋白(IgL)序列、TCR β 序列或TCR δ 序列包括多个轻链恒定区(C_L)序列、TCR β 恒定区(C β)序列、TCR δ 恒定区(C δ)序列或其组合。在一些实施方案中,所述多个轻链恒定区(C_L)序列包括两个或更多个序列,所述序列选自来自Ig κ 、Ig λ 及其组合的轻链恒定区(C_L)序列。

[0021] 在一些实施方案中,第一目标引物、第二目标引物、所述容器条码化多核苷酸、分子条码化多核苷酸或其任意组合不与固体支持体附接。在一些实施方案中,第一目标引物、第二目标引物、所述第一引物组的引物、所述第二引物组的引物或其任意组合不包含分子条码、容器条码、条码或其任意组合。在一些实施方案中,第一目标引物、第二目标引物、所述第一引物组的引物、所述第二引物组的引物或其任意组合不包含突出端区域。在一些实施方案中,所述多个容器的每个容器不包含固体支持体。在一些实施方案中,所述容器条码化多核苷酸附接至固体支持体。在一些实施方案中,所述容器条码化多核苷酸附接至珠子。在一些实施方案中,所述容器条码化多核苷酸、分子条码化多核苷酸或其任意组合不是引物。在一些实施方案中,不延伸所述容器条码化多核苷酸、分子条码化多核苷酸或其任意组合。

[0022] 在一些实施方案中,(a) - (d)在所述单个容器中进行。

[0023] 在一些实施方案中,(a) - (d)在单个反应中进行。

[0024] 在一些实施方案中,所述方法进一步包括裂解所述单细胞。在一些实施方案中,所述裂解从所述单细胞释放所述第一和第二细胞多核苷酸。在一些实施方案中,所述单细胞在(a)后裂解。在一些实施方案中,所述单细胞在(b)前裂解。在一些实施方案中,所述单细胞在所述容器中裂解。在一些实施方案中,所述裂解包括化学裂解。在一些实施方案中,所述裂解包括冻融。

[0025] 在一些实施方案中,所述容器条码在(d)前扩增。在一些实施方案中,所述容器条码与(d)同时扩增。在一些实施方案中,所述容器条码和所述第一单细胞条码化多核苷酸同时扩增或延伸。在一些实施方案中,所述容器条码、所述第一单细胞单条码化多核苷酸和所述第二单细胞单条码化多核苷酸同时扩增或延伸。在一些实施方案中,所述第一单细胞条码化多核苷酸和所述第二单细胞单条码化多核苷酸同时扩增或延伸。在一些实施方案中,所述第一单细胞双条码化多核苷酸和所述第二单细胞双条码化多核苷酸同时扩增或延伸。在一些实施方案中,所述多个容器包含多个孔。在一些实施方案中,所述多个容器包含多份乳液。在一些实施方案中,所述多份乳液的每份乳液的体积为约0.01皮升至10微升。在一些实施方案中,所述多个容器包含多个器皿。在一些实施方案中,所述第一目标引物、第二目标引物、所述第一引物组的引物或所述第二引物组的引物包含样品条码。在一些实施方案中,所述方法进一步包括从所述容器回收所述第一单细胞双条码化多核苷酸、所述第二单

细胞双条码化多核苷酸及其扩增产物。在一些实施方案中,所述方法进一步包括对所述第一单细胞双条码化多核苷酸、所述第二单细胞双条码化多核苷酸、其扩增产物或其任意组合进行测序。在一些实施方案中,所述第一单细胞双条码化多核苷酸、所述第二单细胞双条码化多核苷酸、其扩增产物或其任意组合同时测序。在一些实施方案中,所述第一单细胞双条码化多核苷酸、所述第二单细胞双条码化多核苷酸、其扩增产物或其任意组合在相同的反应中测序。

[0026] 在一些实施方案中,所述方法进一步包括基于所述容器条码,确定所述第一细胞多核苷酸和所述第二细胞多核苷酸的细胞来源是相同的。在一些实施方案中,所述确定包括匹配所述第一单细胞双条码化多核苷酸或其扩增产物的所述容器条码的序列与所述第二单细胞双条码化多核苷酸或其扩增产物的所述容器条码的序列。在一些实施方案中,所述方法进一步包括基于所述分子条码,确定具有所述第一细胞多核苷酸、所述第二细胞多核苷酸或二者的序列的起始分子的数目。在一些实施方案中,所述确定包括确定具有相同第一分子条码、相同第二分子条码或二者的序列的所述数目。在一些实施方案中,当单细胞双条码化多核苷酸或其扩增产物的第一序列和单细胞双条码化多核苷酸或其扩增产物的第二序列含有相同的容器条码或其互补体时,它们来自相同的单个容器或单细胞。在一些实施方案中,当单细胞双条码化多核苷酸或其扩增产物的所述第一序列和单细胞双条码化多核苷酸或其扩增产物的所述第二序列含有不同的分子条码或其互补体时,它们来自不同的细胞多核苷酸分子。在一些实施方案中,当单细胞双条码化多核苷酸或其扩增产物的所述第一序列和单细胞双条码化多核苷酸或其扩增产物的所述第二序列含有相同的分子条码或其互补体时,它们来自相同的细胞多核苷酸分子。在一些实施方案中,当单细胞双条码化多核苷酸或其扩增产物的所述第一序列和单细胞双条码化多核苷酸或其扩增产物的所述第二序列含有不同的容器条码或其互补体时,它们来自不同的单个容器或单细胞。

[0027] 在一些实施方案中,所述单细胞包括免疫细胞。在一些实施方案中,所述多个细胞包括多个免疫细胞。在一些实施方案中,所述免疫细胞为淋巴细胞或其亚型、B细胞或其亚型、T细胞或其亚型或其组合。在一些实施方案中,所述多个细胞针对记忆B细胞、初始B细胞(naïve B cells)、成浆B细胞(plasmablast B cells)、初始T细胞(naïve T cells)、成浆T细胞(plasmablast T cells)、B细胞的任何亚型、T细胞的任何亚型或其任意组合进行了富集。在一些实施方案中,所述单细胞包括癌细胞。在一些实施方案中,所述多个细胞包括多个癌细胞。在一些实施方案中,所述癌细胞为鳞状细胞癌细胞、腺癌细胞、移行细胞癌细胞、骨肉瘤细胞、软骨肉瘤细胞、肌肉肉瘤细胞、白血病细胞、淋巴瘤细胞、神经胶质瘤细胞或其任意组合。在一些实施方案中,所述多个癌细胞针对循环癌细胞、内皮癌细胞、上皮癌细胞、罕见癌细胞或癌细胞的任何类型或亚型进行了富集。在一些实施方案中,所述样品为生物样品。在一些实施方案中,所述生物样品来自受试者。在一些实施方案中,所述方法进一步包括诊断所述受试者为具有疾病或病况。在一些实施方案中,所述受试者为动物。在一些实施方案中,所述动物为人类。在一些实施方案中,所述方法进一步包括确定受试者对于等位基因是纯合还是杂合的。在一些实施方案中,所述方法进一步包括对具有疾病或病况的受试者进行诊断、预后或治疗。在一些实施方案中,所述样品为血液样品。在一些实施方案中,从所述样品分离所述第一或第二细胞多核苷酸。在一些实施方案中,不从所述样品分离所述第一或第二细胞多核苷酸。

[0028] 在一些实施方案中,所述样品包括多个包含第一样品和第二样品的样品。在一些实施方案中,所述多个样品包括至少3、4、5、10、20、30、40、50、60、70、80、90或100个或更多样品。在一些实施方案中,所述多个样品包括至少约100、200、300、400、500、600、700、800、900或1000个或更多样品。在一些实施方案中,所述多个样品包括至少约1000、2000、3000、4000、5000、6000、7000、8000个样品、9000或10,000个样品、或100,000个样品、或1,000,000个或更多样品。在一些实施方案中,所述多个样品包括至少约10,000个样品。在一些实施方案中,所述第一样品来自第一受试者,且所述第二样品来自第二受试者。在一些实施方案中,所述第一受试者为具有疾病或病况的受试者。在一些实施方案中,所述第二受试者为不具有疾病或病况的受试者。在一些实施方案中,所述第一或第二细胞多核苷酸包含变异序列。在一些实施方案中,所述变异序列包含突变、多态性、缺失或插入。在一些实施方案中,所述多态性为单核苷酸多态性。在一些实施方案中,所述第一或第二细胞多核苷酸为对于疾病或病况的生物标志物。在一些实施方案中,所述第一或第二细胞多核苷酸来自病原体。在一些实施方案中,所述病原体为病毒、细菌或真菌。

[0029] 在一些实施方案中,所述方法进一步包括比较来自受试者的所述第一和第二单细胞双条码化多核苷酸的文库与在不同时间点来自相同受试者的所述第一和第二单细胞双条码化多核苷酸的文库的序列。在一些实施方案中,所述方法进一步包括比较来自具有疾病或病况的受试者的所述第一和第二单细胞双条码化多核苷酸的文库与来自没有所述疾病或病况的受试者的所述第一和第二单细胞双条码化多核苷酸的文库的序列。

[0030] 在一些实施方案中,所述方法进一步包括确定所述第一细胞多核苷酸、所述第二细胞多核苷酸或二者的种系序列,其中所述第一细胞多核苷酸包括IgH或V_H序列,且其中所述第二细胞多核苷酸包括IgL或V_L序列,或其任意组合。在一些实施方案中,所述方法进一步包括根据所述种系序列中的其序列确定所述IgL、IgH、V_H、V_L或其任意组合的序列的变异。在一些实施方案中,所述方法进一步包括确定以下至少一个:独特的IgH序列的总数目;独特的IgL序列的总数目;独特的IgH和IgL序列的总数目;独特的成对IgL和IgH序列的总数目;IgH序列或IgL序列的频率;或IgH序列和IgL序列的组合相对于其他一个或多个组合的频率。在一些实施方案中,所述方法进一步包括确定所述第一细胞多核苷酸、所述第二细胞多核苷酸或二者的种系序列,其中所述第一细胞多核苷酸包括TCR α 或V α 序列,且其中所述第二细胞多核苷酸包括TCR β 或V β 序列,或其任意组合。在一些实施方案中,所述方法进一步包括根据所述种系序列中的其序列确定所述TCR α 、TCR β 、V α 、V β 或其任意组合的序列的变异。在一些实施方案中,所述方法进一步包括确定以下至少一个:独特的TCR α 序列的总数目;独特的TCR β 序列的总数目;独特的TCR α 和TCR β 序列的总数目;独特的成对TCR β 和TCR α 序列的总数目;TCR α 序列或TCR β 序列的频率;或TCR α 序列和TCR β 序列的组合相对于其他一个或多个组合的频率。在一些实施方案中,所述方法进一步包括确定所述第一细胞多核苷酸、所述第二细胞多核苷酸或二者的种系序列,其中所述第一细胞多核苷酸包括TCR γ 或V γ 序列,且其中所述第二细胞多核苷酸包括TCR δ 或V δ 序列,或其任意组合。在一些实施方案中,所述方法进一步包括根据所述种系序列中的其序列确定所述TCR γ 、TCR δ 、V γ 、V δ 或其任意组合的序列的变异。在一些实施方案中,所述方法进一步包括确定以下至少一个:独特的TCR γ 序列的总数目;独特的TCR δ 序列的总数目;独特的TCR γ 和TCR δ 序列的总数目;独特的成对TCR δ 和TCR γ 序列的总数目;TCR γ 序列或TCR δ 序列的频率;或TCR γ 序列和TCR δ 序列的

组合相对于其他一个或多个组合的频率。在一些实施方案中,所述方法进一步包括确定以下至少一个:来自第一基因的序列的总数目;来自第二基因的序列的总数目;来自第一基因的独特序列的总数目;来自第二基因的独特序列的总数目;或来自第一基因的序列或来自第二基因的序列的频率。在一些实施方案中,所述方法进一步包括基于分别成对的IgL和IgH序列、或TCR α 和TCR β 序列、或TCR γ 和TCR δ 序列的一对或多对的总量,以及来自种系的变异,来选择抗体或TCR。在一些实施方案中,所述方法进一步包括基于一个或多个IgL或IgH序列、或TCR α 和TCR β 序列、或TCR γ 和TCR δ 序列,以及来自种系的变异,来选择抗体或TCR。在一些实施方案中,所述方法进一步包括基于一个或多个序列模式、变异分析、动力学或频率来选择抗体或TCR。在一些实施方案中,所述方法进一步包括基于频率选择抗体或TCR。

[0031] 在一些实施方案中,所述选定的抗体或TCR以小于约或等于 1×10^{-7} 、 1×10^{-8} 、 1×10^{-9} 、 1×10^{-10} 、 1×10^{-11} 或 1×10^{-12} M的 K_D 与表位结合。

[0032] 在一些实施方案中,所述选定的抗体或TCR为人类治疗性抗体或TCR。在一些实施方案中,所述选定的抗体或TCR为中和抗体或TCR。在一些实施方案中,所述选定的抗体或TCR结合的目标是未知的。在一些实施方案中,所述选定的抗体或TCR结合的目标在选择所述选定的抗体或TCR时是未知的。

[0033] 在一些实施方案中,所述方法进一步包括使所述选定的抗体或TCR与至少一个生物标志物候选物接触以发现生物标志物。在一些实施方案中,所述生物标志物候选物在固体支持体上。在一些实施方案中,所述生物标志物在溶液中。在一些实施方案中,所述抗体或TCR在固体支持体上。在一些实施方案中,所述抗体或TCR在溶液中。在一些实施方案中,所述固体支持体为阵列。在一些实施方案中,所述固体支持体为珠子。

[0034] 在一些实施方案中,所述方法进一步包括将所述第一细胞多核苷酸插入载体中。在一些实施方案中,所述方法进一步包括将所述第二细胞多核苷酸插入所述载体中。在一些实施方案中,所述载体为克隆载体。在一些实施方案中,所述载体为表达载体。

[0035] 在一些实施方案中,所述方法进一步包括匹配具有相同分子条码的序列。在一些实施方案中,所述方法进一步包括从所述文库形成共有序列。在一些实施方案中,测序和PCR错误被最小化、消除或小于0.01%、0.001%、0.0001%、0.00001%、0.000001%或0.0000001%。在一些实施方案中,扩增反应中循环的数目限于1-40个循环的任何一个。

[0036] 在一个方面,本文提供了通过本文所述的任何方法鉴别的分离、纯化的抗体或TCR。在一个方面,本文提供了通过本文所述的任何方法鉴别的分离、纯化的抗体IgL、TCR β 或TCR δ 。在一个方面,本文提供了通过本文所述的任何方法鉴别的分离、纯化的抗体IgH、TCR α 或TCR γ 。在一个方面,本文提供了通过本文所述的任何方法鉴别的分离、纯化的抗体或TCR的Fab片段。在一个方面,本文提供了通过本文所述的任何方法鉴别的分离、纯化的抗体或TCR的Fab₂片段。在一个方面,本文提供了通过本文所述的任何方法鉴别的分离、纯化的抗体或TCR的Fv片段。在一个方面,本文提供了通过本文所述的任何方法鉴别的分离、纯化的抗体的ScFv片段。在一个方面,本文提供了一种治疗有需要的受试者的方法,包括将所述选定的抗体或TCR或其片段施用给有需要的受试者。在一些实施方案中,所述抗体、TCR或其片段鉴别为来自所述有需要的受试者。在一些实施方案中,所述抗体、TCR或其片段未鉴别为来自所述有需要的受试者。在一些实施方案中,所述有需要的受试者展示疾病的一种或多种症状。在一些实施方案中,所述有需要的受试者具有疾病。在一些实施方案中,所述

疾病是未知的。在一些实施方案中,所述疾病是已知的。在一些实施方案中,所述样品包括在第一时间点从受试者获取的第一样品和在第二时间点从所述受试者获取的第二样品。在一些实施方案中,所述方法进一步包括确定来自在所述第一和第二时间点获取的样品的所述第一或第二细胞多核苷酸的量的增加或减少。在一些实施方案中,所述量的增加或减少为范围至少约:0.1倍、0.2倍、0.3倍、0.4倍、0.5倍、0.6倍、0.7倍、0.8倍、0.9倍、1.5倍、2倍、3倍、5倍、10倍、50倍、100倍、1,000倍、10,000倍、100,000倍、1,000,000倍或更多的增加或减少。在一些实施方案中,所述第一和第二时间点之间的时间为约或至少约:1小时、2小时、3小时、4小时、5小时、6小时、7小时、8小时、9小时、10小时、11小时、12小时、13小时、14小时、15小时、16小时、17小时、18小时、19小时、20小时、21小时、22小时、23小时、1天、2天、3天、4天、5天、6天、1周、2周、3周、4周、5周、6周、7周、8周、9周、10周、11周、12周、4个月、5个月、6个月、7个月、8个月、9个月、10个月、11个月、12个月或更久。

[0037] 在一些实施方案中,所述测序为高通量的。在一些实施方案中,所述方法不包括引物的复用(multiplex)和/或附接至固体支持体的引物的复用。在一些实施方案中,所述方法不使用包含与单功能性V区段或V区段的小家族互补的序列的多重V-区段引物。在一些实施方案中,所述方法不使用分离所述第一或第二细胞多核苷酸的步骤。在一些实施方案中,所述测序通过大规模平行合成完成。

[0038] 在一些实施方案中,所述方法进一步包括比较所述序列读取(read)与种系序列,并确定所述序列读取的体细胞超突变积累。在一些实施方案中,所述方法进一步包括确定抗体序列的同种型分布以选择特异性同种型。在一些实施方案中,选定的抗体包括特异性Ig同种型。在一些实施方案中,所述Ig同种型为IgA、IgG、IgM、IgD或IgE。

[0039] 在一些实施方案中,所述方法进一步包括生成成对的IgH和IgL抗体序列或TCR α 和TCR β 序列的文库。在一些实施方案中,所述文库为数据库。在一些实施方案中,所述第一和第二单细胞双条码化多核苷酸包括遍布抗体或TCR编码序列中的CDR1、CDR2、CDR3和/或超突变区域。

[0040] 在一些实施方案中,所述方法进一步包括将所述选定的抗体或TCR直接克隆用于表面展示技术。在一些实施方案中,所述方法进一步包括通过定向演化来演化所述选定的抗体或TCR。在一些实施方案中,所述方法进一步包括对于功能特异性、亲和力或中和能力筛选所述选定的抗体或TCR。在一些实施方案中,以99%或更高的置信度确定体细胞突变。在一些实施方案中,鉴别来自每个多核苷酸分子的各个V、D和J区段。

[0041] 在一些实施方案中,所述容器条码包含至少2个核苷酸。在一些实施方案中,所述容器条码包含至少3、4、5、6、7、8或9个核苷酸。在一些实施方案中,所述容器条码包含至少10个核苷酸。在一些实施方案中,所述容器条码包含至少15个核苷酸。在一些实施方案中,所述容器条码包含最多50个核苷酸。在一些实施方案中,所述容器条码包含10-30个核苷酸。在一些实施方案中,所述容器条码包含简并序列。在一些实施方案中,所述容器条码包含完整或部分的简并序列。在一些实施方案中,所述容器条码包含序列NNNNNNNNNNNNNNNN,其中N为任何核酸。在一些实施方案中,所述容器条码包含序列NNNNNWNNNNNWNNNNN,其中N为任何核酸,且W为腺嘌呤或胸腺嘧啶。在一些实施方案中,所述容器条码包含序列NNNNNXNNNNXNNNN,其中N为任何核酸,且X为任何已知的核苷酸。在一些实施方案中,所述容器条码包含序列NNNNNNNNNNNNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为W,

其中W为腺嘌呤或胸腺嘧啶。在一些实施方案中,所述容器条码包含序列NNNNNNNNNNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为X,其中X为任何已知的核苷酸。在一些实施方案中,所述分子条码包含至少2个核苷酸。在一些实施方案中,所述分子条码包含至少3、4、5、6、7、8或9个核苷酸。在一些实施方案中,所述分子条码包含至少10个核苷酸。在一些实施方案中,所述分子条码包含至少15个核苷酸。在一些实施方案中,所述分子条码包含最多50个核苷酸。在一些实施方案中,所述分子条码包含10-30个核苷酸。在一些实施方案中,所述分子条码包含简并序列。在一些实施方案中,所述分子条码包含完整或部分的简并序列。在一些实施方案中,所述分子条码包含序列NNNNNNNN,其中N为任何核酸。在一些实施方案中,所述分子条码包含序列NNTNNANN,其中N为任何核酸。在一些实施方案中,所述分子条码包含序列NNWNNWNN,其中N为任何核酸,且W为腺嘌呤或胸腺嘧啶。在一些实施方案中,所述分子条码包含序列NNXNNXNN,其中N为任何核酸,且X为任何已知的核苷酸。在一些实施方案中,所述分子条码包含序列NNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为W,其中W为腺嘌呤或胸腺嘧啶。在一些实施方案中,所述分子条码包含序列NNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为X,其中X为任何已知的核苷酸。

[0042] 在一些实施方案中,所述方法进一步包括修正扩增错误。在一些实施方案中,所述方法进一步包括修正测序错误。在一些实施方案中,所述方法进一步包括将包含所述相同分子条码的序列分箱(bin)或分组。在一些实施方案中,所述方法进一步包括使用计算机或算法将包含所述相同分子条码的序列分箱或分组。在一些实施方案中,所述方法进一步包括使用计算机或算法将包含所述相同容器条码的序列分箱或分组。在一些实施方案中,所述方法进一步包括簇集具有至少约90%、95%或99%序列同源性的序列。在一些实施方案中,所述方法进一步包括对具有至少约90%、95%或99%序列同源性的序列进行比对。在一些实施方案中,所述簇集或比对在计算机或算法的帮助下进行。在一些实施方案中,所述方法包括确定含有所述相同分子条码的序列读取的数目。在一些实施方案中,所述方法包括确定含有所述相同分子条码和具有至少约90%、95%或99%序列同源性的相同第一细胞多核苷酸序列的序列读取的数目。在一些实施方案中,所述方法包括确定含有所述相同分子条码和具有至少约90%、95%或99%序列同源性的相同第二细胞多核苷酸序列的序列读取的数目。在一些实施方案中,所述方法包括确定所述样品中第一或第二细胞多核苷酸的量。在一些实施方案中,所述方法包括根据两个或更多个序列、序列读取、扩增子序列、分箱序列、比对序列、簇集序列、或包含所述相同分子条码或容器条码或二者的扩增子集序列形成共有序列。在一些实施方案中,所述方法包括以至少约80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.5%、99.6%、99.7%、99.8%、99.9%、99.99%或100%的准确度或置信度确定第一或第二细胞多核苷酸序列。在一些实施方案中,测序和PCR错误被最小化、消除或小于0.01%、0.001%、0.0001%、0.00001%、0.000001%或0.0000001%。在一些实施方案中,所述测序的错误率小于或等于0.00001%、0.0001%、0.001%、0.01%或0%。在一些实施方案中,所述测序的错误率不是0。在一些实施方案中,对至少1000、100000、 1×10^6 、 1×10^7 、 1×10^8 、 1×10^9 、 1×10^{10} 、 1×10^{11} 、 1×10^{12} 或 9×10^{12} 个多核苷酸进行测序。在一些实施方案中,所述方法在小于或等于4周、3周、2周、1周、6天、5天、5天、4天、3天、2天、1天、18小时、12小时、9小时、6小

时、3小时、2小时或1小时的为正值的时间内进行。在一些实施方案中,用于达到特定置信度或碱基判定精确度的读取的数目比使用相似的方法但不使用分子条码、容器条码或二者而用于达到所述相同、相似或更高置信度或碱基判定精确度的读取的数目少至少约1.1、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、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000倍。在一些实施方案中,用于达到特定置信度或碱基判定精确度的读取的数目比使用相似的方法但不使用分子条码、容器条码或二者而用于达到所述相同、相似或更高置信度或碱基判定精确度的读取的数目少至少约1000、100000、 1×10^6 、 1×10^7 、 1×10^8 、 1×10^9 、 1×10^{10} 、 1×10^{11} 、 1×10^{12} 或 9×10^{12} 个读取。在一些实施方案中,所述多个容器包括至少1000、100000、 1×10^6 、 1×10^7 、 1×10^8 、 1×10^9 、 1×10^{10} 、 1×10^{11} 、 1×10^{12} 或 9×10^{12} 个或更多容器。在一些实施方案中,所述多个细胞多核苷酸包括至少1000、100000、 1×10^6 、 1×10^7 、 1×10^8 、 1×10^9 、 1×10^{10} 、 1×10^{11} 、 1×10^{12} 或 9×10^{12} 个或更多细胞多核苷酸。

[0043] 在一个方面,本文提供了一种组合物,其包含:多个容器,每个容器包含来自包含多个细胞的样品的单细胞,多个分子条码化多核苷酸,容器条码化多核苷酸;与来自所述单细胞的第一细胞多核苷酸互补的第一互补多核苷酸,以及与来自所述单细胞的第二细胞多核苷酸互补的第二互补多核苷酸;其中所述第一互补多核苷酸包含所述多个分子条码化多核苷酸的第一分子条码和所述容器条码化多核苷酸的扩增产物或所述容器条码化多核苷酸的所述容器条码,并且其中所述第二互补多核苷酸包含所述多个分子条码化多核苷酸的第二分子条码和所述容器条码化多核苷酸的扩增产物或所述容器条码化多核苷酸的所述容器条码。

[0044] 在一些实施方案中,所述第一和第二分子条码化多核苷酸的所述分子条码不同。在一些实施方案中,所述第一和第二互补多核苷酸包含不同的分子条码。在一些实施方案中,所述第一和第二互补多核苷酸包含相同的容器条码。在一些实施方案中,所述多个分子条码化多核苷酸不是扩增产物。在一些实施方案中,第一容器中分子条码化多核苷酸的所述分子条码与第二容器中分子条码化多核苷酸的所述分子条码不同。在一些实施方案中,所述多个容器的第一容器中每个分子条码化多核苷酸的所述分子条码是独特的。在一些实施方案中,所述多个容器的第二容器中每个分子条码化多核苷酸的所述分子条码是独特的。在一些实施方案中,第一容器和第二容器中每个分子条码化多核苷酸的所述分子条码是独特的。在一些实施方案中,所述多个容器的第三容器中每个分子条码化多核苷酸的所述分子条码是独特的。在一些实施方案中,所述第一容器、所述第二容器和所述第三容器中每个分子条码化多核苷酸的所述分子条码是独特的。在一些实施方案中,所述多个容器的任何单个容器中每个分子条码化多核苷酸的所述分子条码是独特的。在一些实施方案中,所述多个容器的任何一个容器中每个分子条码化多核苷酸的所述分子条码与所述多个容器的任何其他一个容器中每个分子条码化多核苷酸的所述分子条码不同。在一些实施方案中,第一容器中分子条码化多核苷酸的所述分子条码与第二容器中分子条码化多核苷酸的所述分子条码相同。在一些实施方案中,第一容器中分子条码化多核苷酸的所述分子条码与所述第一容器中分子条码化多核苷酸的所述分子条码相同。在一些实施方案中,第二容器中分子条码化多核苷酸的所述分子条码与所述第二容器中分子条码化多核苷酸的所述分子条码相同。在一些实施方案中,所述多个容器的第一容器中容器条码化多核苷酸或其

扩增子的所述容器条码与所述多个容器的第二容器中容器条码化多核苷酸或其扩增子的所述容器条码不同。在一些实施方案中,所述多个容器的第一容器中容器条码化多核苷酸或其扩增子的所述容器条码为第一相同容器条码。在一些实施方案中,所述多个容器的第二容器中每个容器条码化多核苷酸或其扩增子的所述容器条码为第二相同容器条码。在一些实施方案中,所述第一相同容器条码与所述第二相同容器条码不同。在一些实施方案中,所述多个容器的单个容器中每个容器条码化多核苷酸或其扩增子的所述容器条码包含相同的容器条码。在一些实施方案中,所述多个容器的任何单个容器中每个容器条码化多核苷酸及其扩增子的所述容器条码对于所述多个容器的任何其他单个容器中每个容器条码化多核苷酸及其扩增子的所述容器条码而言是独特的。在一些实施方案中,所述容器条码化多核苷酸以单分子存在于容器中。在一些实施方案中,所述容器条码化多核苷酸以单分子存在于所述多个容器的每个容器中。在一些实施方案中,所述容器条码化多核苷酸以至少单分子存在于所述多个容器的容器中。在一些实施方案中,所述容器条码化多核苷酸以至少单分子存在于所述多个容器的每个容器中。在一些实施方案中,所述多个容器的第一容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述第一容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。在一些实施方案中,所述多个容器的所述第一容器中所述第一容器条码化多核苷酸或其扩增子的第二共同容器序列与所述第一容器中第二容器条码化多核苷酸或其扩增子的第二共同容器序列相同。在一些实施方案中,所述多个容器的任何单个容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述单个容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。在一些实施方案中,所述多个容器的单个容器中每个容器条码化多核苷酸包含相同的第一共同容器序列。在一些实施方案中,所述多个容器的单个容器中每个容器条码化多核苷酸包含相同的第二共同容器序列。在一些实施方案中,所述多个容器的第一容器中第一容器条码化多核苷酸或其扩增子的第一共同容器序列与所述多个容器的第二容器中第二容器条码化多核苷酸或其扩增子的第一共同容器序列相同。在一些实施方案中,所述第一容器条码化多核苷酸或其扩增子的第二共同容器序列与所述第二容器条码化多核苷酸或其扩增子的第二共同容器序列相同。在一些实施方案中,所述多个容器的任何一个容器中每个容器条码化多核苷酸或其扩增子包含第一共同容器序列,所述第一共同容器序列包含与所述多个容器的任何其他一个容器中容器条码化多核苷酸或其扩增子的第一共同容器序列相同的序列。在一些实施方案中,所述多个容器的任何一个容器中每个容器条码化多核苷酸或其扩增子包含第二共同容器序列,所述第二共同容器序列包含与所述多个容器的任何其他一个容器中容器条码化多核苷酸或其扩增子的第二共同容器序列相同的序列。在一些实施方案中,所述多个容器的第一容器中第一分子条码化多核苷酸的第一共同分子序列与所述第一容器中第二分子条码化多核苷酸的第一共同分子序列相同。在一些实施方案中,所述多个容器的所述第一容器中所述第一分子条码化多核苷酸的第二共同分子序列与所述第一容器中第二分子条码化多核苷酸的第二共同分子序列相同。在一些实施方案中,所述多个容器的任何单个容器中第一分子条码化多核苷酸的第一共同分子序列与所述单个容器中第二分子条码化多核苷酸的第一共同分子序列相同。在一些实施方案中,所述多个容器的单个容器中每个分子条码化多核苷酸包含相同的第一共同分子序列。在一些实施方案中,所述多个容器的单个容器中每个分子条码化多核苷酸包含相同的第二共同分子

序列。在一些实施方案中,所述多个容器的第一容器中第一分子条码化多核苷酸的第一共同分子序列与所述多个容器的第二容器中第二分子条码化多核苷酸的第一共同分子序列相同。在一些实施方案中,所述第一分子条码化多核苷酸的第二共同分子序列与所述第二分子条码化多核苷酸的第二共同分子序列相同。在一些实施方案中,所述多个容器的任何一个容器中每个分子条码化多核苷酸包含第一共同分子序列,所述第一共同分子序列包含与所述多个容器的任何其他一个容器中分子条码化多核苷酸的第一共同分子序列相同的序列。在一些实施方案中,所述多个容器的任何一个容器中每个分子条码化多核苷酸包含第二共同分子序列,所述第二共同分子序列包含与所述多个容器的任何其他一个容器中分子条码化多核苷酸的第二共同分子序列相同的序列。在一些实施方案中,所述第一共同容器序列包含的序列包含与所述第一共同分子序列相同的序列。在一些实施方案中,所述第一共同容器序列包含与所述第一共同分子序列或其互补体互补的序列。在一些实施方案中,所述第二共同分子序列包含与添加至所述第一互补多核苷酸的3'端的三个或更多个非模板核苷酸互补的区域。在一些实施方案中,与添加至所述第一互补多核苷酸的3'端的三个或更多个非模板核苷酸互补的所述区域为末端区域。

[0045] 在一些实施方案中,第一和第二分子条码化多核苷酸不融合在一起。在一些实施方案中,所述第一和第二互补多核苷酸不融合在一起。

[0046] 在一些实施方案中,所述第一细胞多核苷酸为DNA。在一些实施方案中,所述第二细胞多核苷酸为DNA。在一些实施方案中,所述第一细胞多核苷酸为RNA。在一些实施方案中,所述第二细胞多核苷酸为RNA。在一些实施方案中,所述RNA为mRNA。在一些实施方案中,所述第一互补多核苷酸为cDNA。在一些实施方案中,所述第二互补多核苷酸为cDNA。

[0047] 在一些实施方案中,所述组合物进一步包含非模板末端转移酶、逆转录酶、聚合酶或其任意组合。在一些实施方案中,所述第一和/或第二互补多核苷酸包含添加至3'端的三个或更多个非模板核苷酸。在一些实施方案中,所述非模板末端转移酶为逆转录酶,并且其中该逆转录酶选自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。在一些实施方案中,第一分子条码化多核苷酸包含与所述第一互补多核苷酸的3'端上的所述三个或更多个非模板核苷酸互补的区域。在一些实施方案中,第二分子条码化多核苷酸包含与所述第二互补多核苷酸的3'端上的三个或更多个非模板核苷酸互补的区域。在一些实施方案中,所述三个或更多个非模板核苷酸为相同的。在一些实施方案中,所述三个或更多个非模板核苷酸中的至少一个与所述三个或更多个非模板核苷酸中的另一个核苷酸不相同。在一些实施方案中,所述第一分子条码化多核苷酸的所述互补区域的至少一个核苷酸与所述第一分子条码化多核苷酸的所述互补区域的另一个核苷酸不相同。在一些实施方案中,所述第二分子条码化多核苷酸的所述互补区域的至少一个核苷酸与所述第二分子条码化多核苷酸的所述互补区域的另一个核苷酸不相同。在一些实施方案中,所述至少一个不相同的核苷酸为脱氧核糖核苷酸或其类似物。在一些实施方案中,所述至少一个不相同的核苷酸不是核糖核苷酸或其类似物。在一些实施方案中,所述至少一个不相同的核苷酸为脱氧核糖鸟苷。在一些实施方案中,所述至少一个不相同的核苷酸为脱氧核糖鸟苷类似物。在一些实施方案中,所述至少一个不相同的核苷酸为所述第一或第二分子条码化多核苷酸的末端核苷酸。在一

些实施方案中,所述至少一个不相同的核苷酸为核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的所述互补区域的末端核苷酸为脱氧核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的所述杂交区域的末端核苷酸不是核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的所述杂交区域的末端核苷酸为脱氧核糖鸟苷。在一些实施方案中,所述第一或第二分子条码化多核苷酸的所述杂交区域的末端核苷酸为脱氧核糖鸟苷类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的所述杂交区域的末端核苷酸为核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的所述杂交区域的至少两个非末端核苷酸为核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的所述杂交区域的至少两个非末端核苷酸不是脱氧核糖核苷酸或其类似物。在一些实施方案中,所述第一或第二分子条码化多核苷酸的所述杂交区域的至少两个非末端核苷酸为脱氧核糖核苷酸或其类似物。在一些实施方案中,所述第一互补多核苷酸包含与第一分子条码化多核苷酸互补的区域。在一些实施方案中,所述第二互补多核苷酸包含与第二分子条码化多核苷酸互补的区域。在一些实施方案中,所述第一互补多核苷酸包含与第二分子条码化多核苷酸互补的区域。在一些实施方案中,与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的所述区域不与分子条码序列互补。在一些实施方案中,与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的所述区域不与所述容器条码化多核苷酸或其扩增产物的区域互补。在一些实施方案中,与所述第一或第二分子条码化多核苷酸互补的所述第一互补多核苷酸的所述区域包含添加至所述第一互补多核苷酸的3'端的三个或更多个非模板核苷酸。在一些实施方案中,与所述第二分子条码化多核苷酸互补的所述第二互补多核苷酸的所述区域包含添加至所述第二互补多核苷酸的3'端的三个或更多个非模板核苷酸。在一些实施方案中,所述第一互补多核苷酸不与所述容器条码化多核苷酸互补。在一些实施方案中,所述第二互补多核苷酸不与所述容器条码化多核苷酸互补。在一些实施方案中,第一分子条码化多核苷酸的互补体的区域与所述容器条码化多核苷酸的区域互补。在一些实施方案中,第二分子条码化多核苷酸的互补体的区域与所述容器条码化多核苷酸的区域互补。在一些实施方案中,所述组合物进一步包含来自以上方法的任何一个或多个引物。在一些实施方案中,所述多个容器的每个容器不包含固体支持体。在一些实施方案中,所述容器条码化多核苷酸附接至固体支持体。在一些实施方案中,所述容器条码化多核苷酸附接至珠子。在一些实施方案中,所述容器条码化多核苷酸、分子条码化多核苷酸或其任意组合不是引物。在一些实施方案中,所述容器条码化多核苷酸、分子条码化多核苷酸或其任意组合不是延伸的多核苷酸。在一些实施方案中,所述细胞为裂解的。在一些实施方案中,所述多个容器包含多个孔。在一些实施方案中,所述多个容器包含多份乳液。在一些实施方案中,所述多份乳液的每份乳液的体积为约0.01皮升至10微升。

[0048] 在一些实施方案中,所述单细胞包括免疫细胞。在一些实施方案中,所述多个细胞包括多个免疫细胞。在一些实施方案中,所述免疫细胞为淋巴细胞或其亚型、B细胞或其亚型、T细胞或其亚型或其组合。在一些实施方案中,所述多个细胞针对记忆B细胞、初始B细胞、成浆B细胞、初始T细胞、成浆T细胞、B细胞的任何亚型、T细胞的任何亚型或其任意组合进行富集。在一些实施方案中,所述单细胞包括癌细胞。在一些实施方案中,所述多个细胞

包括多个癌细胞。在一些实施方案中,所述癌细胞为鳞状细胞癌细胞、腺癌细胞、移行细胞癌细胞、骨肉瘤细胞、软骨肉瘤细胞、肌肉肉瘤细胞、白血病细胞、淋巴瘤细胞、神经胶质瘤细胞或其任意组合。在一些实施方案中,所述多个癌细胞针对循环癌细胞、内皮癌细胞、上皮癌细胞、罕见癌细胞或癌细胞的任何类型或亚型进行富集。在一些实施方案中,所述第一或第二细胞多核苷酸包含变异序列。在一些实施方案中,所述变异序列包含突变、多态性、缺失或插入。在一些实施方案中,所述多态性为单核苷酸多态性。在一些实施方案中,所述第一或第二细胞多核苷酸为对于疾病或病况的生物标志物。在一些实施方案中,所述第一或第二细胞多核苷酸来自病原体。在一些实施方案中,所述第一和第二互补多核苷酸包括遍布抗体或TCR编码序列中的CDR1、CDR2、CDR3和/或超突变区域。

[0049] 在一些实施方案中,所述容器条码包含至少2个核苷酸。在一些实施方案中,所述容器条码包含至少3、4、5、6、7、8或9个核苷酸。在一些实施方案中,所述容器条码包含至少10个核苷酸。在一些实施方案中,所述容器条码包含至少15个核苷酸。在一些实施方案中,所述容器条码包含最多50个核苷酸。在一些实施方案中,所述容器条码包含10-30个核苷酸。在一些实施方案中,所述容器条码包含简并序列。在一些实施方案中,所述容器条码包含完整或部分的简并序列。在一些实施方案中,所述容器条码包含序列NNNNNNNNNNNNNNN,其中N为任何核酸。在一些实施方案中,所述容器条码包含序列NNNNNWNNNNWNNNNN,其中N为任何核酸,且W为腺嘌呤或胸腺嘧啶。在一些实施方案中,所述容器条码包含序列NNNNNXNNNNXNNNNN,其中N为任何核酸,且X为任何已知的核苷酸。在一些实施方案中,所述容器条码包含序列NNNNNNNNNNNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为W,其中W为腺嘌呤或胸腺嘧啶。在一些实施方案中,所述容器条码包含序列NNNNNNNNNNNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为X,其中X为任何已知的核苷酸。在一些实施方案中,所述分子条码包含至少2个核苷酸。在一些实施方案中,所述分子条码包含至少3、4、5、6、7、8或9个核苷酸。在一些实施方案中,所述分子条码包含至少10个核苷酸。在一些实施方案中,所述分子条码包含至少15个核苷酸。在一些实施方案中,所述分子条码包含最多50个核苷酸。在一些实施方案中,所述分子条码包含10-30个核苷酸。在一些实施方案中,所述分子条码包含简并序列。在一些实施方案中,所述分子条码包含完整或部分的简并序列。在一些实施方案中,所述分子条码包含序列NNNNNNNN,其中N为任何核酸。在一些实施方案中,所述分子条码包含序列NNTNNANN,其中N为任何核酸。在一些实施方案中,所述分子条码包含序列NNWNNWNN,其中N为任何核酸,且W为腺嘌呤或胸腺嘧啶。在一些实施方案中,所述分子条码包含序列NNXNNXNN,其中N为任何核酸,且X为任何已知的核苷酸。在一些实施方案中,所述分子条码包含序列NNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为W,其中W为腺嘌呤或胸腺嘧啶。在一些实施方案中,所述分子条码包含序列NNNNNNNN,其中N为任何核酸,且序列中至少一个或两个N为X,其中X为任何已知的核苷酸。

[0050] 在一些实施方案中,所述多个容器包括至少1000、100000、 1×10^6 、 1×10^7 、 1×10^8 、 1×10^9 、 1×10^{10} 、 1×10^{11} 、 1×10^{12} 或 9×10^{12} 个或更多容器。在一些实施方案中,所述多个细胞多核苷酸包括至少1000、100000、 1×10^6 、 1×10^7 、 1×10^8 、 1×10^9 、 1×10^{10} 、 1×10^{11} 、 1×10^{12} 或 9×10^{12} 个或更多细胞多核苷酸。

[0051] 在一个方面,本文提供了一种多核苷酸条码化方法,其包括(a)使分子条码化多核

核苷酸与来自单细胞的多个多核苷酸中的每个杂交,其中所述杂交分子条码化多核苷酸来自包含所述单细胞的容器内的多个独特分子条码化多核苷酸;(b)延伸与分子条码化多核苷酸杂交的来自所述单细胞的多核苷酸,以形成分子条码细胞多核苷酸;(c)使容器条码化多核苷酸与分子条码细胞多核苷酸杂交,其中所述容器条码化多核苷酸对于多个容器中的单个容器是独特的;(d)延伸与容器条码化多核苷酸杂交的分子条码细胞多核苷酸,以形成双条码化细胞多核苷酸;以及(e)对所述双条码化细胞多核苷酸进行测序。在一些实施方案中,(a)中的所述杂交不是通过在来自单细胞的所述多核苷酸上天然存在的序列的碱基配对进行的。在一些实施方案中,与所述分子条码细胞多核苷酸杂交的所述容器条码化多核苷酸为扩增产物。在一些实施方案中,(c)中的杂交不是通过在来自单细胞的所述多核苷酸上天然存在的序列的互补体的碱基配对进行的。在一些实施方案中,(c)中的杂交不是通过与(b)中延伸的来自所述单细胞的所述多核苷酸的区域的碱基配对进行的。在一些实施方案中,(a)-(d)在所述单个容器中进行。在一些实施方案中,(a)-(d)在单个反应中进行。

[0052] 援引并入

[0053] 本说明书中提到的所有出版物、专利和专利申请均为了所有目的通过引用以其全文并入本文,其程度犹如特别地且单独地指出每个单独的出版物、专利或专利申请通过引用而并入。例如,本文提到的所有出版物和专利均为了描述和公开出版物中所述的试剂盒、组合物和方法学的目的通过引用以其全文并入本文,其可与本文所述的方法、试剂盒和组合物结合使用。本文所讨论的文件仅针对其在本申请的申请日之前的公开内容而提供。本文的任何内容都不应解释为承认本文所述的发明人无权因在先发明或由于任何其他原因而先于此公开。

附图说明

[0054] 本文所述的新颖特征在随附权利要求中具体阐述。通过参考以下对应用本文所述特征的原理的说明性实例加以阐述的详细描述以及附图,将会获得对本文所述特征的特点和优点的更好理解,在这些附图中:

[0055] 图1A描绘了本文所述的示例性方法的条码化(barcoding)期的示意图。该简图表示对两个或更多个多核苷酸进行扩增和条码化的方法,所述多核苷酸例如是成对的可变Ig(例如, V_H 和 V_L mRNA)和TCR序列(例如, $V\alpha/V\beta$ 和 $V\gamma/V\delta$ mRNA),例如以供文库制备和免疫测序。容器条码(DB);分子条码(MD)。(上图)含有单细胞和其他反应组分(例如,酶、缓冲液、寡核苷酸)的乳液中的(多个小滴的)单个小滴。(中图)细胞裂解和裂解的细胞RNA的逆转录。(下图)在逆转录期间,单个分子的分子条码(MB)标记。

[0056] 图1B描绘了本文所述的示例性方法的扩增期的示意图。该简图表示对两个或更多个多核苷酸进行扩增和条码化的方法,所述多核苷酸例如是成对的可变Ig(例如, V_H 和 V_L mRNA)和TCR序列(例如, $V\alpha/V\beta$ 和 $V\gamma/V\delta$ mRNA),例如以供文库制备和免疫测序。(上图)容器条码(VB)的独立扩增在每个小滴中生成相同VB的多个拷贝。在扩增的退火和延伸期期间,用VB同时标记cDNA-MB分子。(中图)在扩增循环期间,双条码化cDNA分子的同时扩增。(下图)从乳液中的小滴回收的扩增产物。

[0057] 图2示出的示意图显示,容器条码(DB)的序列同一性允许对每个RNA鉴别起源细胞。

[0058] 图3示出的示意图显示,如果发现相同的分子条码(MB)与同一个相同RNA序列附接,则此RNA-MB-DB种类可能是PCR复制的结果。当发现两个不同的MB与同一个相同RNA序列附接时,则这些RNA1-MB1-DB和RNA1-MB2-DB为两个独立的起源RNA分子的实际观察结果,而不是PCR复制的实际观察结果。

[0059] 图4A描绘了本文所述的示例性方法的示意图。该简图表示对成对的可变Ig(例如, V_H 和 V_L 序列)和TCR序列(例如, $V\alpha/V\beta$ 和 $V\gamma/V\delta$ 序列)进行扩增和条码化的方法,以供文库制备和免疫测序。容器条码(DB);分子条码(MD)。每个示出的反应可在单个乳液相中完成,且为了易于呈现而分别示出。

[0060] 图4B描绘了本文所述的示例性方法的示意图。该简图表示对 V_H 和 V_L 抗体mRNA进行扩增和条码化的方法,以供文库制备和免疫测序。

[0061] 图4C描绘了本文所述的示例性方法的示意图。该简图例示了对 V_H 和 V_L 抗体mRNA进行扩增和条码化的方法,以供文库制备和免疫测序。

[0062] 图4D描绘了本文所述的示例性方法的示意图。该简图例示了对成对的可变Ig(例如, V_H 和 V_L 序列)和TCR序列(例如, $V\alpha/V\beta$ 和 $V\gamma/V\delta$ 序列)进行扩增和条码化的方法,以供文库制备和免疫测序。在用容器条码(DB)标记cDNA之前,可选的cDNA扩增步骤。

[0063] 图5示出的示意图显示,容器条码(DB)的序列同一性允许对每个RNA鉴别起源细胞。该方法可使用含有多个各含单细胞的小滴的乳液,以在单个反应中产生双条码化cDNA。

[0064] 图6示出的示意图显示,如果发现相同的分子条码(MB)与同一个相同RNA序列附接,则此RNA-MB-DB种类可能是PCR复制的结果。当发现两个不同的MB与同一个相同RNA序列附接时,则这些RNA1-MB1-DB和RNA1-MB2-DB为两个独立的起源RNA分子的实际观察结果,而不是PCR复制的实际观察结果。

[0065] 图7A描绘了本文所述的示例性方法的示意图。该简图表示权利要求中的术语的图例。

[0066] 图7B描绘了本文所述的示例性方法的示意图。该简图表示对两个或更多个多核苷酸进行扩增和条码化的方法,所述多核苷酸例如是成对的可变Ig(例如, V_H 和 V_L mRNA)和TCR序列(例如, $V\alpha/V\beta$ 和 $V\gamma/V\delta$ mRNA),诸如以供文库制备和免疫测序。

具体实施方式

[0067] 参考用于说明的实例应用,多个方面描述如下。应当理解,许多具体细节、关系和方法的阐述是为了提供对本文所述的特征的全面了解。然而,相关领域普通技术人员将容易认识到,本文所述的特征可以在不存在一个或多个具体细节或采用其他方法的情况下实践。本文所述的特征不受限于所阐述的行为或事件的排序,因为一些行为可能以不同顺序发生和/或与其他行为或事件同时发生。此外,实施依照本文所述的特征的方法并非需要全部所阐述的行为或事件。

[0068] 本文使用的术语仅为了描述具体情况,而并非意在限制。如本文所用的,除非上下文另有明确说明,否则单数形式“一个”、“一种”和“该”意在还包括复数形式。此外,在术语“包括”、“具有”或其变化形式用于具体实施方式和/或权利要求时,这样的术语意为包含性的,类似于术语“包含”。

[0069] 术语“约”或“大约”可意指如本领域普通技术人员所确定的,位于特定值的可接受

的误差范围内,其将部分取决于该值如何测量或确定,即测量系统的局限性。例如,“约”可意指按照本领域中的实践,在1个或超过1个标准差内。或者,“约”可意指给定值的多达20%、多达10%、多达5%或多达1%的范围。或者,特别是对于生物系统或过程,该术语可意指在数值的数量级内,在5倍以内,更优选在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,应当假定术语“约”意指对于该特定值在可接受的误差范围内。

[0070] T细胞受体链对和抗体免疫球蛋白链对均为免疫受体的类型,且在演化上相关。本发明的目的在于生成多核苷酸文库以供高通量测序和诊断。本发明的目的还在于从患者或具有特定共同属性的队列开发人源文库组 (panel) 以供抗体和/或TCR发现。起始材料可为外周血或来自组织活检,如果需要的话,从中总体分离免疫细胞或亚分选为幼稚、记忆和ASC。本公开发明可应用于多种不同类型的成对可变序列,例如,T细胞受体链对和抗体免疫球蛋白链对。

[0071] 分离的细胞,诸如免疫细胞,可封装在容器如油包水乳液(小滴)中,封装方式使得创造每个小滴含有单个或更少的免疫细胞的单独皮升区室。对于每个样品,诸如来自受试者的生物样品,可处理数百万个细胞,从而允许单细胞测序技术中的高通量。使用本文所述的方法可避免使用固体支持体,诸如珠子。使用本文所述的方法也可避免对生成分隔的容器群体的需要。例如,序列的文库可在相同的或单个反应中,或在单独的多个容器或容器群体中生成。在容器形成期间引入与细胞多核苷酸(诸如 V_H 和 V_L 抗体链和/或 $V\alpha/V\beta$ 和 $V\gamma/V\delta$ T细胞受体(TCR)链)互补的多核苷酸。在容器形成期间也可引入带有容器条码的多核苷酸。这些容器条码化多核苷酸可携带简并条码,使得每个含有容器条码的细胞多核苷酸含有对应于它们所处的容器的独特识别码。在容器形成期间还可引入多个带有分子条码的多核苷酸。这些分子条码化多核苷酸可携带简并条码,使得每个含有分子条码的细胞多核苷酸分子含有对应于它们所来自的单细胞多核苷酸分子的独特识别码。数百万个单个免疫细胞可在乳液内裂解,细胞转录物,如 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 链转录物,可使用引物进行逆转录或拷贝,随后用容器条码和分子条码标记,并PCR扩增该条码化多核苷酸。起源于单个免疫细胞(例如,B细胞或T细胞)的每个 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 链可借助相同的容器条码标识虚拟地彼此关联。

[0072] 然后,可从容器中回收 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 链并PCR富集,以便添加下一代测序(NGS)标签。可使用高通量测序平台对该文库进行测序,然后分析组库多样性、抗体频率、CDR3表征、体细胞超突变系统发生分析等。可通过对容器和分子条码序列进行去卷积来生成正确匹配的 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 对的数据库。因为每个单个免疫细胞都隔离在它们各自的容器中,所以对于每个观察到两次的容器条码,测序的转录物来源于相同的乳液小滴,并因此来自独特的单细胞。对于观察到的每个不同分子条码,对于含有相同容器条码的序列,测序的转录物来源于来自单细胞的不同转录物分子。对于观察到的每个相同分子条码,对于含有相同容器条码的序列,测序的转录物来源于来自单细胞的相同转录物分子(例如,PCR复制物)。

[0073] 与测序并行地,从容器中回收的 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 链的文库可克隆到抗体表达载体中并共转染以供酵母展示筛选。与在开始时就分裂生物样品相比,克隆此相同的文库集合体是优选的方法,因为一些罕见的免疫细胞将只在一个或另一个分析中被捕获。人源的 V_H 及 V_L 和/或 $V\alpha$ 及 $V\beta$ 和/或 $V\gamma$ 及 $V\delta$ 链的文库不论正确或错误的成对匹配均可表

达,如经典展示分析一样。然后,可针对一个或多个抗原靶标进行酵母展示,以富集潜在的抗体候选物。

[0074] 从展示技术如酵母展示得到的阳性候选抗体可进行测序并针对匹配对的条码数据库进行查询。每个酵母展示的 V_H 和/或 V_α 和/或 V_γ 链可分别回溯匹配于其各自的 V_L 或 V_β 或 V_δ 链,且每个酵母展示的 V_L 和/或 V_β 和/或 V_δ 链可分别回溯匹配于其各自的 V_H 或 V_α 或 V_γ 链。这些正确配对的候选物可在哺乳动物细胞系中基因合成并表达,并针对感兴趣的靶标进行功能性验证。这些候选物可以是完全人类抗体和/或TCR。

[0075] “抗体”是指天然的或部分或完全合成产生的免疫球蛋白(Ig)。“T细胞受体”(“TCR”)是指天然的或部分或完全合成产生的分子,其发现于识别与主要组织相容性复合体(MHC)分子结合的抗原的T淋巴细胞(T细胞)的表面。包括具有结合域的多肽或蛋白质,该结合域为抗原结合域或与抗原结合域同源。该术语进一步包括“抗原结合片段”和其他对于相似的结合片段的可互换术语,如下所述。这些术语还涉及互补决定区(CDR)移植抗原和TCR以及其他人源化抗体和TCR(包括CDR修饰和构架区修饰)。应当注意,虽然可能只提及了免疫球蛋白链(例如,重链和轻链),但是本公开发明可应用于多种其他不同类型的成对序列,例如T细胞受体链对(TCR α 和TCR β 链以及TCR γ 和TCR δ 链),且不限于免疫球蛋白。

[0076] 天然抗体和天然免疫球蛋白通常是约150,000道尔顿的异四聚体糖蛋白,由两个相同的轻链(L)和两个相同的重链(H)组成。每个轻链一般都通过一个共价二硫键与重链连接,而在不同免疫球蛋白同种型的重链之间二硫键的数目不同。每个重链和轻链还具有规律间隔的链内二硫桥。每个重链在一端具有可变域(V_H),随后是许多恒定域(C_H)。每个轻链在一端具有可变域(V_L),并且在其另一端具有恒定域(C_L);轻链的恒定域与重链的第一恒定域对齐,且轻链可变域与重链的可变域对齐。特定的氨基酸残基被认为形成轻链和重链可变域之间的界面。根据其重链的恒定域的氨基酸序列,抗体可归为不同的类别,包括IgA、IgD、IgE、IgG和IgM,并且其中几种可进一步分为亚类(同种型),例如IgG₁、IgG₂、IgG₃、IgG₄、IgA和IgA₂。抗体的重链(IgH)基于其恒定域的氨基酸序列而对应于免疫球蛋白的不同类别,分别称为 α 、 δ 、 ϵ 、 γ 和 μ 。来自任何脊椎动物物种的抗体的轻链(IgL)基于其恒定域的氨基酸序列,可归为两种明显不同的类型之一,称为kappa(κ)和lambda(λ)。

[0077] T细胞识别与多种癌症或传染性生物体相关的抗原的能力由其TCR赋予,TCR由alpha(α)链和beta(β)链或由gamma(γ)和delta(δ)链构成。构成这些链的蛋白质由DNA编码,其采用独特的机理生成了TCR的巨大多样性。这种多亚单元免疫识别受体与CD3复合体相关联,并与抗原呈递细胞(APC)表面上的MHC I类和II类蛋白质所呈递的肽相结合。TCR与APC上的抗原肽的结合是T细胞活化的核心事件,其在T细胞与APC之间的接触点上的免疫突触处发生。

[0078] 每个TCR含有可变互补决定区(CDR),以及构架区(FR)和恒定区。 α 和 β 链可变域的第三互补决定区(CDR3)环的氨基酸序列很大程度上分别决定了由 β 链基因座中可变性(V_β)、多样性(D_β)和连接(J_β)基因区段之间,以及 α 链基因座中类似的 V_α 和 J_α 基因区段之间的重组引起的 $\alpha\beta$ T细胞的序列多样性。多种这类基因区段在TCR α 和 β 链基因座中的存在允许大量不同的CDR3序列得到编码。在TCR基因重排过程期间,在 V_β - D_β 、 D_β - J_β 和 V_α - J_α 连接处的核苷酸的独立添加和缺失进一步增加了CDR3序列多样性。在此方面,免疫活性反映在TCR的多样性中。

[0079] $\gamma\delta$ TCR与 $\alpha\beta$ TCR的区别在于其编码的受体与先天免疫系统紧密相互作用。TCR $\gamma\delta$ 在发育的早期表达,具有特定的解剖分布,具有独特的病原体和小分子特异性,且具有广谱的先天性和适应性细胞相互作用。在个体发生早期,因为TCR $\gamma\delta$ 细胞的有限的子集在产前占据了各种组织,所以建立了TCR γ V和J区段表达的偏向模式。因此,通过环境暴露于病原体和毒性分子而刺激后,广泛的外围扩充在成人组织中产生了许多多样性的TCR γ 组库。

[0080] 由B细胞表达的Ig为由四条多肽链组成的蛋白质,即两条重链(IgH)和两条轻链(IgL),形成了 H_2L_2 结构。每一对IgH和IgL链含有高变域和恒定域,该高变域由 V_L 和 V_H 区组成。Ig的IgH链有多个类型, μ 、 δ 、 γ 、 α 和 β 。个体内的Ig的多样性主要由高变域决定。与TCR相似,IgH链的V结构域通过 V_H 、 D_H 和 J_H 基因区段的组合连接而创建。在Ig基因重排过程期间,在 V_H - D_H 、 D_H - J_H 和 V_H - J_H 连接处的核苷酸的独立添加或缺失进一步增加了高变域序列多样性。在此,免疫活性反映在Ig的多样性中。

[0081] 关于抗体链如重链和轻链或TCR链如 α 和 β 链或 γ 和 δ 链的“可变”是指抗体或TCR链的部分,其在抗体或TCR之间的序列有所不同,并且参与每个特定抗体或TCR对其特定抗原的结合和特异性。这样的可变性集中于三个被称为高变区的区段,其在轻链和重链可变域二者或 α 和 β 可变域二者内。可变域更高度保守的部分称为构架区(FR)。天然重链和轻链的可变域各自包含四个FR(分别为FR1、FR2、FR3和FR4),通过三个高变区相连接。每条链中的高变区通过FR很靠近地保持在一起,并且与来自另一条链的高变区一起,促进了抗体的抗原结合位点的形成(参见Kabat等人,Sequences of Proteins of Immunological Interest,第五版,Public Health Service,National Institutes of Health,Bethesda,Md. (1991),647-669页)。恒定域不直接参与抗体或TCR与抗原的结合,但展现出多种效应子功能,例如,抗体参与抗体依赖性细胞毒性。

[0082] “高变区”是指抗体或TCR中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。“构架”或“FR”残基是除了如本文所定义的高变区残基之外的那些可变域残基。

[0083] “抗体片段”和“TCR片段”包含全长抗体或TCR的一部分,通常是其抗原结合域或可变域。抗体和TCR片段的实例包括但不限于Fab、Fab'、 $F(ab')_2$ 、Fv和scFv片段、线性抗体或TCR、单链抗体或TCR分子、双抗体,以及由抗体或TCR片段形成的多特异性抗体或TCR。

[0084] “单克隆抗体”是指由免疫细胞的单个克隆合成的抗体分子。修饰语“单克隆”表明了抗体的特性是从基本上均质的抗体群体获得的,且不应解释为需要通过任何特定方法产生抗体。因此,单克隆抗体可通过首先由Kohler和Milstein,Nature 256:495 (1975); Eur.J.Immunol.6:511 (1976)描述的杂交瘤方法、通过重组DNA技术制备,或者还可从噬菌体抗体文库中分离。

[0085] “多克隆抗体”是指由免疫细胞群体合成的抗体分子群体。

[0086] “单链Fv”或“scFv”是指包含抗体的可变重链(V_H)和可变轻链(V_L)域或TCR的可变 α 或 γ 链($V\alpha$ 或 $V\gamma$)和可变 β 或 δ 链($V\beta$ 或 $V\delta$)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多肽进一步在 V_H 和 V_L 结构域或 $V\alpha$ 和 $V\beta$ 结构域或 $V\gamma$ 和 $V\delta$ 结构域之间包含多肽连接体,其使scFv能够形成用于抗原结合的所需结构。

[0087] “双抗体”是指具有两个抗原结合位点的小抗体和/或TCR片段,该片段包含同一多肽链中的与 V_L 连接的 V_H (V_H - V_L),或同一多肽链中的与 $V\beta$ 连接的 $V\alpha$ ($V\alpha$ - $V\beta$),或同一多肽链中

的与V δ 连接的V γ (V γ -V δ)。通过使用因太短而无法允许在同一链上的两个结构域之间配对的连接体,该结构域被迫与另一条链的互补结构域配对,并形成了两个抗原结合位点。示例性的双抗体更充分地描述于例如EP404097和W093111161中。

[0088] “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。具体地,如本文使用的该术语包括但不限于,对靶抗原和帮助向特定组织递送的另一个靶标显示出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。

[0089] “线性抗体”或“线性TCR”是指形成一对抗原结合区的一对串联Fd区段(例如,V_H-C_{H1}-V_H-C_{H1}或V α -C α_1 -V α -C α_1)。线性抗体和TCR可为双特异性或单特异性的,例如,如Zapata等人,Protein Eng.8(10):1057-1062(1995)所述。

[0090] “抗原结合域”是指抗体或TCR的一个或多个片段,其保持了与抗原特异性结合的能力。这类术语内包含的抗体片段的非限制性实例包括但不限于:(i) Fab片段,即由V_L、V_H、C_L和C_{H1}结构域组成的单价片段;(ii) F(ab')₂片段,即含有在铰链区处由二硫桥连接的两个Fab片段的二价片段;(iii) 由V_H和C_{H1}结构域组成的Fd片段;(iv) 含有抗体的单臂的V_L和V_H结构域的Fv片段,(v) dAb片段(Ward等人,(1989)Nature 341:544-546),其含有V_H结构域;以及(vi) 分离的CDR。此定义中还包括包含单个重链和单个轻链的抗体或具有单个 α 链或单个 β 链的TCR。

[0091] “F(ab')₂”和“Fab”部分可通过用蛋白酶如胃蛋白酶和木瓜蛋白酶处理Ig而产生,并包括通过在两条重链各自的铰链区之间存在的二硫键附近消化免疫球蛋白生成的抗体片段。例如,木瓜蛋白酶在两条重链各自的铰链区之间存在的二硫键的上游切割IgG,以生成两个同源抗体片段,其中由V_L和C_L组成的轻链和由V_H和C_{H1} (重链的恒定区中的 γ 1区)组成的重链片段通过二硫键在它们的C末端区处相连接。这两个同源抗体片段各自称为'Fab'。胃蛋白酶还在两条重链各自的铰链区之间存在的二硫键的下游切割IgG,以生成比该片段略大的抗体片段,其中两个上述'Fab'在铰链区处相连接。此抗体片段被称为F(ab')₂。Fab片段还含有轻链的恒定域和重链的第一恒定域(C_{H1})。'Fab'片段与Fab片段的不同在于在重链C_{H1}结构域的羧基末端添加少量残基,包括来自抗体铰链区的一个或多个半胱氨酸。Fab'-SH在此是指其中恒定域的半胱氨酸残基带有游离巯基的Fab'。F(ab')₂抗体片段最初以其间具有铰链半胱氨酸的Fab'片段对的形式产生。

[0092] “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、非共价缔合的一个重链和一个轻链可变域或一条TCR α 链和一条TCR β 链或一条TCR γ 链和一条TCR δ 链的二聚体组成。在此构型中,每个可变域的三个CDR相互作用,在V_H-V_L二聚体或V α -V β 二聚体或V γ -V δ 二聚体的表面上限定了抗原结合位点。共同地,来自每条V_H和V_L链或V α -V β 链或V γ -V δ 链的一个或多个CDR的组合给抗体或TCR赋予了抗原结合特异性。例如,应当理解,例如,当转移到受体选择的抗体、TCR或其抗原结合片段的V_H和V_L链或V α 和V β 链或V γ -V δ 链时,CDRH3和CDRL3可能足以给抗体或TCR赋予抗原结合特异性,并且可检测此CDR组合的结合、亲和力等。即使单个可变域(或只含有对抗原具有特异性的三个CDR的一半Fv)也具有识别并结合抗原的能力,但是可能比与第二可变域组合时具有更低的亲和力。此外,虽然Fv片段的两个结构域(V_L和V_H或V α 和V β 或V γ 和V δ)由单独的基因编码,但是它们可使用重组方法通过合成的连接体连接,这使它们能够成为单个蛋白质链,其中V_L和V_H或V α 和V β 或V γ 和V δ

δ 链区配对以形成单价分子(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)。这样的scFv还意在包含在术语抗体的“抗原结合部分”内。特定scFv的任何 V_H 和 V_L 序列都可与Fc区cDNA或基因组序列相连接,以便生成编码完整的Ig(例如IgG)分子或其他同种型的表达载体。 V_H 和 V_L 还可用于使用蛋白质化学或重组DNA技术生成Fab、Fv或Ig的其他片段。

[0093] 抗原结合多肽还包括重链二聚体,例如,来自驼类和鲨鱼的抗体。驼类和鲨鱼抗体包含V-样和C-样结构域(均不具有轻链)的两条链的同源二聚体对。因为驼类中重链二聚体IgG的 V_H 区不需要与轻链进行疏水相互作用,所以在驼类中通常与轻链接触的重链区域变为亲水氨基酸残基。重链二聚体IgG的 V_H 结构域被称为 V_{HH} 结构域。鲨鱼Ig-NAR包含一个可变域(称为V-NAR结构域)和五个C-样恒定域(C-NAR结构域)的同源二聚体。在驼类中,抗体组库的多样性由 V_H 或 V_{HH} 区中的CDR1、2和3决定。驼类 V_{HH} 区中的CDR3的特征为其相对较长的长度,平均为16个氨基酸(Muyldermans等人,1994,Protein Engineering 7(9):1129)。

[0094] 非人类(例如,鼠)抗体或TCR的“人源化”形式包括含有最少的来源于非人类Ig或TCR的序列的嵌合抗体或TCR。在大多数情况下,人源化抗体或TCR为人类Ig或TCR(受体抗体或TCR),其中受体的一个或多个CDR被来自具有所需特异性、亲和力和结合功能的非人类物种(诸如小鼠、大鼠、兔或非人灵长类)抗体或TCR(供体抗体或TCR)的CDR所替代。在一些情况下,人类Ig或TCR的一个或多个FR氨基酸残基被相应的非人类氨基酸残基所替代。此外,人源化抗体或TCR可含有未发现于受体抗体或TCR中或者供体抗体或TCR中的残基。如果需要的话,可进行这些修饰以便改善抗体或TCR的性能。人源化抗体或TCR可包含基本上所有的至少一个(在一些情况下为两个)可变域,其中所有或基本上所有的高变区对应于非人类免疫球蛋白或TCR的高变区,并且所有或基本上所有的FR为人类免疫球蛋白或TCR序列的FR。任选地,人源化抗体或TCR还可包含免疫球蛋白或TCR恒定区(Fc)的至少一部分,一般为人类免疫球蛋白或TCR的一部分。参见例如,Jones等人,Nature 321:522-525(1986);Reichmann等人,Nature 332:323-329(1988);以及Presta,Curr.Op.Struct.Biol.2:593-596(1992)。

[0095] “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)的基因序列。种系DNA含有编码单个Ig重链或轻链、或单个TCR α 或TCR β 链、或单个TCR γ 或TCR δ 链的多个基因区段。这些基因区段携带于生殖细胞中,但在排列为功能性基因之前不能转录和翻译。在骨髓中的B细胞和T细胞分化期间,这些基因区段被能够生成超过 10^8 种特异性的动态遗传系统随机改组。这些基因区段中的大部分由种系数据库公布并收集。

[0096] “亲和力”是指两种物质可逆结合的平衡常数,且表示为 K_D 。结合蛋白质对配体的亲和力,诸如抗体对表位的亲和力可为,例如,约100纳摩尔(nM)至约0.1nM、约100nM至约1皮摩尔(pM)或约100nM至约1飞摩尔(fM)。术语“亲合力”是指两种或多种物质的复合体在稀释后对解离的阻力。

[0097] “表位”是指能够与抗体或TCR的可变区结合袋形成结合相互作用的抗原或其他大分子的一部分。这样的结合相互作用可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。抗原结合可涉及,例如,CDR3、CDR3对,或在一些情况下,涉及多达 V_H 和 V_L 链的所有六个CDR的相互作用。表位可为线性肽序列(即“连续的”)或可由非邻接的氨基酸序列

构成(即“构象的”或“不连续的”)。抗体或TCR可识别一个或多个氨基酸序列;因此表位可限定超过一个不同的氨基酸序列。被抗体和TCR识别的表位可通过本领域技术人员公知的肽作图和序列分析技术来确定。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。

[0098] “特异性”是指这样一种情况:除了含有被该抗体或TCR识别的表位的抗原以外,抗体或TCR对其他分子将不显示出任何明显的结合。该术语还适用于,例如,抗原结合域对大量抗原携带的特定表位具有特异性的情况,在这种情况下,携带该抗原结合域的选定的抗体、TCR或抗原结合片段将能够与携带该表位的各种抗原结合。术语“优先结合”或“特异性结合”意指该抗体、TCR或其片段与表位结合的亲和力大于它与无关氨基酸序列结合的亲和力,并且,如果与含有该表位的其他多肽具有交叉反应性,则在其配制用于人类给药应用的水平下无毒。在一个方面,这样的亲和力比该抗体、TCR或其片段对无关氨基酸序列的亲和力大至少1倍、至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、10倍、至少20倍、至少30倍、至少40倍、至少50倍、至少60倍、至少70倍、至少80倍、至少90倍、至少100倍或至少1000倍。术语“结合”是指在生理条件下两个分子之间由于例如共价、静电、疏水和离子和/或氢键相互作用的直接缔合,且包括诸如盐桥和水桥等相互作用,以及结合的任何其他常规方式。

[0099] “药学上可接受的”是指生理上可忍受,且在施用给人类时通常不产生变应性或类似的不良反应,诸如胃不适、头晕等的分子实体和组合物。

[0100] 当与治疗性组合物结合使用时,“单位剂量”是指适合作为用于人类的单一剂量的物理上离散的单元,每个单元含有经计算将产生所需治疗效果的预定量的活性物质以及需要的稀释剂,即载体或媒介物。

[0101] “包装材料”是指容纳试剂盒的组分的物理结构。包装材料可无菌地保持组分,并可由通常用于此类目的的材料(例如,纸、波纹纤维、玻璃、塑料、箔、安瓿等)制成。标签或包装插页可包含适当的书面说明。因此,试剂盒还可包含关于在本发明的任何方法中使用试剂盒组分的标签或说明。试剂盒可包含包装中的化合物,或分配器,以及关于以本文所述的方法施用化合物的说明。

[0102] “防止”是指预防、阻止症状的发作,防止与蛋白质水平过量相关或与蛋白质活性相关的疾病或病症的进展。

[0103] “抑制”、“治疗”和“处理”可互换使用,并且是指,例如,停滞症状、延长存活、部分或完全减轻症状,以及部分或全部消除与蛋白质水平过量相关或与蛋白质活性相关的病症、疾病或病症。例如,癌症的治疗包括但不限于停滞、部分或完全消除癌生长或肿瘤。治疗或部分消除包括,例如,生长或肿瘤大小和/或体积的成倍减少,诸如约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,治疗或部分消除可包括约1%、2%、3%、4%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、95%的生长或肿瘤大小和/或体积减少百分比,或之间的任何减少百分比。

[0104] “中和抗体”或“中和TCR”是指抑制病原体如病毒或细菌复制的任何抗体或TCR,而无论实现中和的机理如何。

[0105] “抗体组库”或“TCR组库”是指抗体、TCR或其片段的集合。例如,抗体组库可用于选择特定的抗体或针对诸如结合能力、结合特异性、胃肠转运能力、稳定性、亲和力等特定性

质进行筛选。该术语特别地包括抗体和TCR文库,包括所有形式的组合文库,例如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,包括原始、合成和半合成文库。

[0106] “目标核酸分子”、“目标分子”、“目标多核苷酸”、“目标多核苷酸分子”是指任何感兴趣的核酸。

[0107] 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,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。采用两条引物的结果是,对于每次循环,两条链的模板多核苷酸拷贝数的指数增加,因为每次循环两条链都得到复制。多核苷酸双链体具有对应于所用引物的末端的末端。PCR可包括使模板多核苷酸变性、使引物与引物结合位点退火以及在核苷酸的存在下通过DNA或RNA聚合酶延伸引物的一次或多次重复。特定温度、每个步骤的持续时间和步骤之间的变化率取决于本领域普通技术人员公知的许多因素。(McPherson等人, IRL Press, Oxford (1991和1995))。例如,在使用Taq DNA聚合酶的传统PCR中,双链模板多核苷酸可在 $>90^{\circ}\text{C}$ 的温度下变性,引物可在 $50-75^{\circ}\text{C}$ 范围内的温度下退火,且引物可在 $72-78^{\circ}\text{C}$ 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包括在PCR反应之前的逆转录反应,并扩增所产生的cDNA。巢式PCR包括两级PCR,其中使用第一引物组的第一PCR反应的扩增子成为使用第二引物组的第二PCR反应的样品,至少一个第二引物与第一PCR反应的扩增子的内部位置相结合。多重PCR包括PCR反应,其中多个多核苷酸序列在同一反应混合物中同时经历PCR。PCR反应体积可为 $0.2\mu\text{L}-1000\mu\text{L}$ 之间的任何体积。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。定量测量可包括将一个或多个参考序列或标准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))。

[0108] 在其他实施方案中,本文公开的方法、试剂盒和组合物可包含支持体。在其他实施方案中,本文公开的方法、试剂盒和组合物不包含支持体。通常,固体支持体包含一种或多种材料,该材料包含一个或多个刚性或半刚性表面。在一些实施方案中,该支持体为非固体支持体。该支持体或基底可包括膜、纸、塑料、涂层表面、平面、玻璃、载片、芯片或其任意组合。在一些实施方案中,支持体的一个或多个表面基本上为平面,然而在一些实施方案中,可期望其对于不同的化合物在物理上分隔合成区域,例如采用孔、凸起区、针、蚀刻沟槽等。在一些实施方案中,固体支持体包括珠子、树脂、凝胶、微球或其他几何构型。或者,固体支持体可包括二氧化硅芯片、微粒、纳米粒子、平板和阵列。固体支持体可包括使用在微孔中自组装的珠子。例如,固体支持体包括Illumina的BeadArray技术。或者,固体支持体包括Abbott Molecular的Bead Array珠阵列技术和Applied Microarray的FlexiPlexTM系统。在其他情况下,固体支持体为板。板的实例包括但不限于MSD多阵列板、MSD Multi-Spot[®]板、微板、ProteOn微板、AlphaPlate、DELFI板、IsoPlate和LumaPlate。在一些实施方案中,

支持体可包括多个珠子。在一些实施方案中,支持体可包括阵列。在一些实施方案中,支持体可包括载玻片。适用于聚合物的方法、基底和技术(美国专利号5,744,305、5,143,854、5,242,974、5,252,743、5,324,633、5,384,261、5,405,783、5,424,186、5,451,683、5,482,867、5,491,074、5,527,681、5,550,215、5,571,639、5,578,832、5,593,839、5,599,695、5,624,711、5,631,734、5,795,716、5,831,070、5,837,832、5,856,101、5,858,659、5,936,324、5,968,740、5,974,164、5,981,185、5,981,956、6,025,601、6,033,860、6,040,193、6,090,555、6,136,269、6,269,846和6,428,752;美国专利公开号20090149340、20080038559、20050074787;以及PCT公开号WO 00/58516、WO 99/36760和WO 01/58593)。多核苷酸与支持体的附接可包括胺-巯基交联、马来酰亚胺交联、N-羟基琥珀酰亚胺或N-羟基磺基琥珀酰亚胺、Zenon或SiteClick。将标记的核酸附接至支持体可包括将生物素与多个多核苷酸附接,并用链霉亲和素涂覆一个或多个珠子。在一些实施方案中,固体支持体为珠子。珠子的实例包括但不限于链霉亲和素珠子、琼脂糖珠子、磁珠、Dynabeads®、MACS®微珠、抗体偶联的珠子(例如,抗免疫球蛋白微珠)、蛋白A偶联的珠子、蛋白G偶联的珠子、蛋白A/G偶联的珠子、蛋白L偶联的珠子、多核苷酸dT偶联的珠子、二氧化硅珠子、二氧化硅样珠子、抗生物素微珠、抗荧光染料微珠和BcMag™羧基封端磁珠。珠子的直径可为约5μm、10μm、20μm、25μm、30μm、35μm、40μm、45μm或50μm。固体支持体可为阵列或微阵列。固体支持体可包含离散的区域。固体支持体可为阵列,例如可寻址阵列。

[0109] 如本文所用的,“核苷酸”、“核苷”、“核苷酸残基”和“核苷残基”可意指能够作为适于在扩增反应(例如PCR反应)中使用的引物的组分的脱氧核糖核苷酸或核糖核苷酸残基,或其他相似的核苷类似物。除非另有说明,这类核苷及其衍生物可用作本文所述的引物的构件。本申请中的任何内容都不意在排除使用经化学修饰以增强它们的稳定性或在扩增反应中的有效性的核苷衍生物或碱基,条件是化学修饰不干扰它们被聚合酶适当地识别为脱氧鸟嘌呤、脱氧胞嘧啶、脱氧胸腺嘧啶或脱氧腺嘌呤。在一些实施方案中,核苷酸类似物可以使杂合体形成稳定化。在一些实施方案中,核苷酸类似物可以使杂合体形成去稳定化。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。

[0110] “核酸”或语法上的等同物是指单个核苷酸或共价连接在一起的至少两个核苷酸。

[0111] “多核苷酸”或语法上的等同物是指共价连接在一起的至少两个核苷酸。多核苷酸包括含有两个或更多个核苷酸的分子。多核苷酸包括任何长度的核苷酸的聚合形式,该核苷酸为核糖核苷酸、脱氧核糖核苷酸或肽核酸(PNA),其包含嘌呤和嘧啶碱基,或其他天然的、化学或生物修饰的、非天然的核苷酸碱基或其衍生物。多核苷酸的骨架可包含糖和磷酸基团,或修饰或取代的糖或磷酸基团。多核苷酸可包含修饰的核苷酸,诸如甲基化的核苷酸和核苷酸类似物。核苷酸的序列可被非核苷酸组分中断。多核苷酸可包括其他分子,诸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实例包括基因、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、序列的分离的DNA、序列的分离的RNA、核酸探针和引物。多核苷酸可从天然来源中分离、重组或人工合成。

[0112] 多核苷酸包含四种核苷酸碱基的特定序列:腺嘌呤(A);胞嘧啶(C);鸟嘌呤(G);和

胸腺嘧啶 (T) (当多核苷酸为RNA时尿嘧啶 (U) 代替胸腺嘧啶 (T))。因此,多核苷酸序列为多核苷酸分子的字母表示;或者,该术语可应用于多核苷酸分子本身。此字母表示可输入具有中央处理器的计算机中的数据库中,并用于生物信息学应用,诸如功能基因组学、同源性搜索、分箱序列、比对序列和确定共有序列。

[0113] 多核苷酸可包括非标准核苷酸,诸如核苷酸类似物或修饰的核苷酸。在一些实施方案中,非标准核苷酸使杂合体形成稳定化。在一些实施方案中,非标准核苷酸可使杂合体形成去稳定化。在一些实施方案中,非标准核苷酸可增强杂交特异性。在一些实施方案中,非标准核苷酸可降低杂交特异性。非标准核苷酸修饰的实例包括2' O-Me、2' O-烯丙基、2' O-炔丙基、2' O-烷基、2' 氟代、2' 阿拉伯糖基、2' 木糖基、2' 氟阿拉伯糖基、硫代磷酸、二硫代磷酸、磷酰胺化物、2' 氨基、5-烷基-取代嘧啶、3' 脱氧鸟苷、5-卤代-取代嘧啶、烷基-取代嘌呤、卤代-取代嘌呤、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、4-乙酰胞嘧啶、5-(羧基羟甲基)尿嘧啶、5-羧甲基氨基甲基-2-硫代尿苷、5-羧甲基氨基甲基尿嘧啶、二氢尿嘧啶、 β -D-半乳糖基queosine、肌苷、N6-异戊烯基腺嘌呤、1-甲基鸟嘌呤、1-甲基肌苷、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-硫尿嘧啶、 β -D-甘露糖基queosine、5'-甲氧基羧甲基尿嘧啶、5-甲氧基尿嘧啶、2-甲硫基-D46-异戊烯基腺嘌呤、尿嘧啶-5-氧基乙酸(v)、wybutoxosine、假尿嘧啶、queosine、2-硫代胞嘧啶、5-甲基-2-硫尿嘧啶、2-硫尿嘧啶、4-硫尿嘧啶、5-甲基尿嘧啶、尿嘧啶-5-氧基乙酸甲酯、尿嘧啶-5-氧基乙酸(v)、5-甲基-2-硫尿嘧啶、3-(3-氨基-3-N-2-羧丙基)尿嘧啶、(acp3)w、2,6-二氨基嘌呤及其衍生物。

[0114] “受试者”、“个体”、“宿主”或“患者”是指活生物体,诸如哺乳动物。受试者和宿主的实例包括但不限于马、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,非哺乳类脊椎动物,诸如禽类(例如,鸡或鸭)、鱼(例如鲨鱼)或蛙(例如爪蟾),和非哺乳类无脊椎动物,以及其转基因物种。在某些方面,受试者是指单个生物体(例如人)。在某些方面,提供了一组个体,其组成具有研究的共同免疫因子和/或疾病的小群组(cohort),和/或不具有该疾病的个体的群组(例如,阴性/正常对照)。从中获得样品的受试者可患有疾病和/或病症(例如,一种或多种变态反应、感染、癌症或自身免疫疾病等),并可与未患有该疾病的阴性对照受试者相比较。

[0115] “试剂盒”是指为了实施本文所公开的方法而递送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,适当容器中的探针、酶等)和/或支持材料(例如,缓冲液,关于进行分析的书面说明等)储存、运输或从一个位置递送到另一个位置的系统。例如,试剂盒包括一个或多个含有相关反应试剂和/或支持材料的包装物(例如,盒子)。这类内容物可一起或分开递送给预期的接受者。例如,第一容器可含有供在分析中使用的酶,而第二容器含有多个引物。

[0116] “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,多肽可包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000个肽或氨基酸。多肽的实例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。

[0117] “样品”是指生物、环境、医学、受试者或患者样品或含有多核苷酸如目标多核苷酸的样品。

[0118] 样品

[0119] 含有多核苷酸的任何生物样品都可在本文所述的方法中使用。例如，样品可为来自含有RNA或DNA的受试者的生物样品。可从生物样品中提取多核苷酸，或者可直接对样品进行该方法而无需提取或纯化多核苷酸。样品可为提取或分离的DNA或RNA。样品还可为从生物试样中提取的总RNA或DNA、cDNA文库、病毒或基因组DNA。在一个实施方案中，从含有多其他组分如蛋白质、脂质和非模板核酸的生物样品中分离多核苷酸。可从任何细胞材料获得核酸模板分子，该细胞材料从动物、植物、细菌、真菌或任何其他细胞生物体获得。在某些实施方案中，从单细胞获得多核苷酸。多核苷酸可从生物体直接获得或从由生物体获得的生物样品获得。任何组织或体液试样都可用作用于本发明中的核酸的来源。多核苷酸还可从培养的细胞如原代细胞培养或细胞系中分离。可用病毒或其他细胞内病原体感染从中获得模板核酸的细胞或组织。

[0120] 在某些实施方案中，产生抗体或TCR的免疫细胞可从受试者或宿主的血液或其他生物样品中分离，以鉴别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，该受试者或宿主例如是人或其他动物，诸如已经免疫或患有感染、癌症、自身免疫病况或任何其他疾病的人或其他动物。例如，该人可被诊断为患有疾病、表现出疾病的症状、未被诊断为患有疾病或未表现出疾病的症状。例如，该人可为暴露于传染原（例如，病毒、细菌、寄生虫、朊病毒等）、抗原或疾病和/或可产生对其有用的抗体或TCR的人。例如，该动物可为暴露于传染原（例如，病毒、细菌、寄生虫、朊病毒等）、抗原或疾病和/或可产生对其有用的抗体或TCR的动物。来自经免疫的宿主的某些免疫细胞产生针对一种或多种所述靶抗原和/或一种或多种未知抗原的抗体或TCR。在本发明中，可通过任何合适的方法，诸如使用荧光激活细胞分选（FACS）、磁性激活细胞分选、淘选或其他筛选方法来筛选并分选细胞，以生成来自样品的多种免疫细胞，诸如免疫细胞文库，从而针对所需免疫细胞对淋巴细胞集合体进行富集，之后对抗体链进行测序、制备抗体，或制备表达文库。与只提供表达不同抗体的免疫细胞的很少子集，并因此只提供很少的天然存在的可变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表达不同抗体或TCR的单独的免疫细胞或其至少2个子集。例如，本发明的免疫细胞文库可含有表达不同抗体或TCR的单独的免疫细胞或其至少5、10、100、250、500、750、1000、2500、5000、10000、25000、50000、75000、100000、250000、500000、750000、1000000、2500000、5000000、7500000或10000000个子集。本发明的方法使免疫细胞回收最大化，并提供了非常高的多样性。

[0121] 在一些实施方案中，采用来自非免疫的人类或非人类供体的免疫细胞。动物的起始组库（抗原激发之前的组库）向动物提供了能以适度的亲和力（ K_A 为约 1×10^{-6} 至 1×10^{-7} M）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在种系中直接编码，而是以组合的方式从V基因区段装配。免疫引发任何免疫细胞产生 V_H-V_L 或 $V\alpha-V\beta$ 或 $V\gamma-V\delta$ 组合，该组合与免疫原结合以增殖（克隆扩充）并分泌如上所述的相应抗体。然而，使用来自未经免疫的受试者的脾细胞和/或免疫细胞或其他外周血淋巴细胞（PBL）可提供对可能的抗体或TCR组库的更好呈现，并且还允许使用任何动物物种构建随后的B细胞或T细

胞抗体或TCR文库。

[0122] 在一些情况下,为了获得足够的核酸以供测试,抽取至少0.001、0.005、0.01、0.05、0.1、0.5、1、2、3、4、5、10、20、25、30、35、40、45或50mL的血液体积。

[0123] 在一些情况下,起始材料为外周血。外周血细胞可针对特定细胞类型(例如,单个核细胞;红细胞;CD4⁺细胞;CD8⁺细胞;免疫细胞;T细胞,NK细胞等)进行富集。外周血细胞还可选择性消耗特定细胞类型(例如,单个核细胞;红细胞;CD4⁺细胞;CD8⁺细胞;免疫细胞;T细胞,NK细胞等)。

[0124] 在一些情况下,起始材料可为组织样品,包括固体组织,其非限制性实例包括脑、肝、肺、肾、前列腺、卵巢、脾、淋巴结(包括扁桃体)、甲状腺、胰腺、心脏、骨骼肌、肠、喉、食道和胃。在其他情况下,起始材料可为含有核酸的细胞、免疫细胞,并且特别是B细胞或T细胞。在一些情况下,起始材料可为来自可从中获得遗传物质的任何生物体的含有核酸的样品。在一些情况下,样品为流体,例如血液、唾液、淋巴或尿液。

[0125] 可从具有病况的受试者取得样品。在一些情况下,从中取得样品的受试者可为患者,例如,癌症患者或疑似患有癌症的患者。受试者可为哺乳动物,例如人类,并可为雄性或雌性。在一些情况下,该雌性为妊娠的。该样品可为肿瘤活检物。活检可由例如医疗服务人员进行,包括内科医师、医师助理、护士、兽医、牙医、脊椎推拿师、急救医师、皮肤科医师、肿瘤科医师、胃肠科医师或外科医师。

[0126] 在一些情况下,可使用酶处理(如蛋白酶消化)从起始材料中去除非核酸物质。

[0127] 在一些情况下,血液可收集到含有镁螯合剂(包括但不限于EDTA)的装置中,并在4℃下储存。任选地,可添加钙螯合剂,包括但不限于EGTA。在另一种情况下,向血液中添加细胞裂解抑制剂,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、巯基添加剂或二硫化物还原剂、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。

[0128] 在一些情况下,当提取的材料包含单链RNA、双链RNA或DNA-RNA杂合体时,可使用本领域已知的技术将这些分子转化为双链DNA。例如,可使用逆转录酶从RNA分子合成DNA。在一些情况下,RNA到DNA的转化可能需要预先的连接步骤,以将连接体片段与RNA相连接,从而允许使用通用引物来启动逆转录。在其他情况下,mRNA分子的聚-A尾例如可用来启动逆转录。在一些情况下,在转化为DNA后,可使用本文详细描述的方法来进行进一步捕获、选择、标记或分离所需序列。

[0129] 核酸分子包括脱氧核糖核酸(DNA)和/或核糖核酸(RNA)。核酸分子可为合成的或来源于天然存在的来源。在一个实施方案中,从含有多种其他组分如蛋白质、脂质和非模板核酸的生物样品中分离核酸分子。可从任何细胞材料获得核酸模板分子,该细胞材料从动物、植物、细菌、真菌或任何其他细胞生物体获得。在某些实施方案中,从单细胞获得核酸分子。用于在本发明中使用的生物样品包括病毒颗粒或制品。核酸分子可从生物体直接获得,或从自生物体获得的生物样品获得,例如,从血液、尿液、脑脊液、精液、唾液、痰、粪便和组织获得。任何组织或体液试样都可用作供本发明使用的核酸的来源。核酸分子还可从培养的细胞如原代细胞培养物或细胞系中分离。可用病毒或其他细胞内病原体感染从中获得模板核酸的细胞或组织。

[0130] 样品还可以是从生物试样中提取的总RNA、cDNA文库、病毒或基因组DNA。在某些实施方案中,核酸分子与诸如蛋白质、酶、底物、抗体、结合剂、珠子、小分子、肽或任何其他分子等其他目标分子结合。通常,可通过多种技术从生物样品中提取核酸,诸如Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*, 第三版, Cold Spring Harbor, N.Y. (2001) 描述的技术。核酸分子可为单链的、双链的或具有单链区的双链(例如,茎结构和环结构)。

[0131] DNA提取方法是本领域公知的。经典的DNA分离方案基于使用有机溶剂如苯酚和氯仿的混合物来提取,然后用乙醇沉淀(J.Sambrook等人,“*Molecular Cloning: A Laboratory Manual*,”1989,第二版,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)。用于从生物样品中提取DNA的多种试剂盒可商购获得(例如, BD Biosciences Clontech (Palo Alto, CA); Epicentre Technologies (Madison, WI); Gentra Systems, Inc. (Minneapolis, MN); MicroProbe Corp. (Bothell, WA); Organon Teknika (Durham, NC); 和 Qiagen Inc. (Valencia, CA))。

[0132] RNA提取方法也是本领域公知的(例如, J.Sambrook等人, “*Molecular Cloning: A Laboratory Manual*” 1989, 211d Ed., Cold Spring Harbour Laboratory Press:New York)且用于从体液中提取RNA的试剂盒可商购获得(例如, Ambion, Inc. (Austin, TX); Amersham Biosciences (Piscataway, NJ); BD Biosciences Clontech (Palo Alto, CA); BioRad Laboratories (Hercules, CA); Dynal Biotech Inc. (Lake Success, NY); Epicentre Technologies (Madison, WI); Gentra Systems, Inc. (Minneapolis, MN); GIBCO BRL (Gaithersburg, MD); Invitrogen Life Technologies (Carlsbad, CA); MicroProbe Corp. (Bothell, WA); Organon Teknika (Durham, NC); Promega, Inc. (Madison, WI); 和 Qiagen Inc. (Valencia, CA))。

[0133] 一个或多个样品可来自一个或多个来源。一个或多个样本可来自两个或更多个来源。一个或多个样本可来自一个或多个受试者。一个或多个样本可来自两个或更多个受试者。一个或多个样本可来自相同的受试者。一个或多个受试者可来自相同的物种。一个或多个受试者可来自不同的物种。该一个或多个受试者可为健康的。该一个或多个受试者可患有疾病、病症或病况。

[0134] 在一些实施方案中,样品为流体,诸如血液、唾液、淋巴、尿液、脑脊液、精液、痰、粪便或组织匀浆。

[0135] 可从具有病况的受试者取得样品。在一些实施方案中,从中取得样品的受试者可为患者,例如,癌症患者或疑似患有癌症的患者。受试者可为哺乳动物,例如人,并可为雄性或雌性。在一些实施方案中,该雌性为妊娠的。该样品可为肿瘤活检物。活检可由例如医疗服务人员进行,包括内科医师、医师助理、护士、兽医、牙医、脊椎推拿师、急救医师、皮肤科医师、肿瘤科医师、胃肠科医师或外科医师。

[0136] 在一些实施方案中,多核苷酸与诸如蛋白质、酶、底物、抗体、结合剂、珠子、小分子、肽或任何其他分子等其他目标分子结合。在一些实施方案中,多核苷酸不与固体支持体

结合。可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))。

[0137] 在一些实施方案中,样品为唾液。在一些实施方案中,样品为全血。在一些实施方案中,为了获得足量的多核苷酸以供测试,取得至少0.001、0.005、0.01、0.05、0.1、0.5、1、2、3、4、5、10、20、25、30、35、40、45或50mL的血液体积。在一些实施方案中,血液可收集到含有镁螯合剂(包括但不限于EDTA)的装置中,并在4℃下储存。任选地,可添加钙螯合剂,包括但不限于EGTA。

[0138] 在一些实施方案中,向血液中添加细胞裂解抑制剂,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、巯基添加剂或二硫化物还原剂、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。在一些实施方案中,可使用酶处理(诸如蛋白酶消化)从起始材料中去除非核苷酸物质。

[0139] 多个样品可包括至少2、3、4、5、10、20、30、40、50、60、70、80、90或100个或更多样品。多个样品可包括至少约100、200、300、400、500、600、700、800、900或1000个或更多样品。多个样品可包括至少约1000、2000、3000、4000、5000、6000、7000、8000个样品、9000或10,000个样品、或100,000个样品、或1,000,000个或更多样品。多个样品可包括至少约10,000个样品。

[0140] 第一样品中的一个或多个多核苷酸可与第二样品中的一个或多个多核苷酸不同。第一样品中的一个或多个多核苷酸可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包含至少约80%、85%、90%、95%、96%、97%、98%、99%或100%的序列同一性。在一些实施方案中,样品中的一个或多个多核苷酸可以有少于约100、90、80、70、60、50、40、30、25、20、25、10、9、8、7、6、5、4、3、2或1个核苷酸或碱基对不同。多个样品中的一个或多个样品中的多个多核苷酸可包含两个或更多个相同序列。多个样品中的一个或多个样品中的总多核苷酸的至少约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%可包含相同的序列。多个样品中的一个或多个样品中的多个多核苷酸可包含至少两个不同的序列。多个样品中的一个或多个样品中的总多核苷酸的至少约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%可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为彼此的变体。例如,一个或多个多核苷酸可含有单核苷酸多态性或其他类型的突变。在另一个实例中,一个或多个多核苷酸为剪接变体。

[0141] 第一样品可包含一个或多个细胞,且第二样品可包含一个或多个细胞。第一样品中的一个或多个细胞与第二样品中的一个或多个细胞可属于相同的细胞类型。第一样品中的一个或多个细胞与所述多个样品中的一个或多个不同细胞可属于不同的细胞类型。

[0142] 所述多个样品可同时获得。多个样品可在相同时间获得。多个样品可顺序获得。多个样品可在获得一个或多个不同样品的数年过程中获得,例如100年、10年、5年、4年、3年、2年或1年。一个或多个样品可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、4个

月、3个月、2个月或1个月内获得。一个或多个样品可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。一个或多个样品可在获得一个或多个不同样品的少于一秒内获得。

[0143] 样品的不同多核苷酸可以不同浓度或量存在于样品中(例如,不同的分子数目)。例如,样品中一个多核苷酸的浓度或量可大于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是样品中至少一个其他多核苷酸的浓度或量的至少约1.5、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000倍或更多倍。在另一个实例中,样品中一个多核苷酸的浓度或量小于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、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000倍或更多倍。

[0144] 在一些实施方案中,两个或更多个样品可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,血液样品可比尿液样品含有更高的特定多核苷酸的量。或者,单个样品可分为两个或更多个亚样品。亚样品可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是另一个样品中相同多核苷酸的浓度或量的至少约1.5、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000倍或更多倍。或者,一个样品中一个多核苷酸的浓度或量可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比另一个样品中相同多核苷酸的浓度或量小至少约1.5、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000倍或更多倍。

[0145] 目标多核苷酸

[0146] 在一些情况下,本文提供的方法涉及目标多核苷酸分子(诸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目标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两个或更多个目标多核苷酸分子的扩增和测序。在一个方面,目标多核苷酸为RNA。在一个方面,目标多核苷酸为基因组核酸。来源于特定生物体的染色体中的遗传物质的DNA可为基因组DNA。在优选的实施方案中,目标多核苷酸包括包含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,目标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,目标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,目标多核苷酸包括包含由免疫细胞产生的TCR的 α 链可变区的序列。在一些实施方案中,目标多核苷酸包括包含由免疫细胞产生的TCR的 β 链可变区的序列。在一些实施方案中,目标多核苷酸包括包含由免疫细胞产生的TCR的 γ 链可变区的序列。在一些实施方案中,目标

多核苷酸包括包含由免疫细胞产生的TCR的 δ 链可变区的序列。

[0147] 目标多核苷酸可从实质上任何来源获得,并可使用本领域已知的方法制备。例如,目标多核苷酸可直接分离,而无需使用本领域已知的方法扩增,该方法包括但不限于从生物体或细胞(例如免疫细胞)中提取基因组DNA或mRNA的片段以获得目标多核苷酸。目标多核苷酸还可包含通过逆转录-PCR从RNA(诸如mRNA)生成的cDNA。在一些情况下,目标多核苷酸为RNA分子。在一些情况下,目标多核苷酸为mRNA分子,或从mRNA分子产生的cDNA。在一些情况下,目标多核苷酸为来自单个免疫细胞的mRNA分子或从mRNA分子产生的cDNA分子。在一些情况下,目标多核苷酸为来自单独免疫细胞的mRNA分子或从mRNA分子产生的cDNA分子。在一些情况下,目标多核苷酸为编码来自单个免疫细胞的抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的重链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单个免疫细胞的重链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的轻链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单个免疫细胞的轻链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的抗体可变序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单个免疫细胞的可变抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变轻链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变轻链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变重链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变重链抗体序列的mRNA分子。在一些情况下,目标多核苷酸可为无细胞的核酸,例如DNA或RNA。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变 α 、 β 、 γ 和/或 δ 链TCR序列的mRNA分子。

[0148] 本文所述的方法可用来由一个或多个目标多核苷酸生成多核苷酸文库以供测序。目标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,目标多核苷酸可包括生物样品中的多核苷酸。例如,目标多核苷酸不包括PCR反应的产物。例如,目标多核苷酸可包括用于生成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,目标多核苷酸可包括用于生成逆转录反应或引物延伸反应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,目标多核苷酸包括可经历逆转录反应或引物延伸反应的感兴趣的多核苷酸。例如,目标多核苷酸包括RNA或DNA。例如,目标多核苷酸包括cDNA。在一些实施方案中,目标RNA多核苷酸为mRNA。在一些实施方案中,目标RNA多核苷酸为聚腺苷酸化的。在一些实施方案中,RNA多核苷酸不是聚腺苷酸化的。在一些实施方案中,目标多核苷酸为DNA多核苷酸。该DNA多核苷酸可为基因组DNA。DNA多核苷酸可包括外显子、内含子、非翻译区或其任意组合。

[0149] 在一些实施方案中,文库可从目标多核苷酸的两个或更多个区域生成。在一些实施方案中,方法文库可从两个或更多个目标多核苷酸生成。在一些实施方案中,目标多核苷酸为基因组核酸或来源于染色体的DNA。在一些实施方案中,目标多核苷酸包括包含诸如多态性或突变等变体的序列。在一些实施方案中,目标多核苷酸包括DNA且不包括RNA。在一些实施方案中,目标多核苷酸包括RNA且不包括DNA。在一些实施方案中,目标多核苷酸包括DNA和RNA。在一些实施方案中,目标多核苷酸为mRNA分子。在一些实施方案中,目标多核苷酸为DNA分子。在一些实施方案中,目标多核苷酸为单链多核苷酸。在一些实施方案中,目标

多核苷酸为双链多核苷酸。在一些实施方案中,目标多核苷酸为双链多核苷酸的单链。

[0150] 目标多核苷酸可从任何生物样品中获得,并使用本领域已知的方法制备。在一些实施方案中,目标多核苷酸直接分离而无需扩增。直接分离方法是本领域已知的。非限制性实例包括从生物样品、生物体或细胞中提取基因组DNA或mRNA。

[0151] 在一些实施方案中,一个或多个目标多核苷酸从生物样品中纯化。在一些实施方案中,目标多核苷酸不从包含它的生物样品中纯化。在一些实施方案中,目标多核苷酸从生物样品中分离。在一些实施方案中,目标多核苷酸不从包含它的生物样品中分离。在一些实施方案中,目标多核苷酸可为无细胞的核酸。在一些实施方案中,目标多核苷酸可为成片段化的核酸。在一些实施方案中,目标多核苷酸可为经转录的核酸。在一些实施方案中,目标多核苷酸为经修饰的多核苷酸。在一些实施方案中,目标多核苷酸为未经修饰的多核苷酸。

[0152] 在一些实施方案中,目标多核苷酸为来自单细胞的多核苷酸。在一些实施方案中,目标多核苷酸来自单独的细胞。在一些实施方案中,目标多核苷酸为来自含有多个细胞的样品的多核苷酸。

[0153] 在一些实施方案中,目标多核苷酸编码生物标志物序列。在一些实施方案中,目标多核苷酸编码两个或更多个生物标志物序列。在一些实施方案中,多个目标多核苷酸编码生物标志物序列。在一些实施方案中,多个目标多核苷酸编码两个或更多个生物标志物序列。在一些实施方案中,多个目标多核苷酸编码3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100个或更多个生物标志物序列。

[0154] 在一些实施方案中,多个目标多核苷酸包含一组免疫球蛋白序列。在一些实施方案中,多个目标多核苷酸包含一组TCR序列。例如,一组免疫球蛋白序列可为 V_H 和/或 V_L 序列。在一些实施方案中,一组免疫球蛋白或TCR序列含有1、2、3、4、5、6、7、8、9或10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序列含有至少约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 或 9×10^{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序列含有最多约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、

4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 或 9×10^{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、10-30、10-40、10-30、10-40、10-50、10-60、10-70、10-80、10-90、10-100、50-60、50-70、50-80、50-90、50-100、100-200、100-300、100-400、100-300、100-400、100-500、100-600、100-700、100-800、100-900、100-1000、500-600、500-700、500-800、500-900、500-1000、1000-2000、1000-3000、1000-4000、1000-3000、1000-4000、1000-5000、1000-6000、1000-7000、1000-8000、1000-9000、1000-10000、5000-6000、5000-7000、5000-8000、5000-9000、5000-10000、 $1-1 \times 10^5$ 、 $1-2 \times 10^5$ 、 $1-3 \times 10^5$ 、 $1-4 \times 10^5$ 、 $1-5 \times 10^5$ 、 $1-6 \times 10^5$ 、 $1-7 \times 10^5$ 、 $1-8 \times 10^5$ 、 9×10^5 、 $1-1 \times 10^6$ 、 $1-2 \times 10^6$ 、 $1-3 \times 10^6$ 、 $1-4 \times 10^6$ 、 $1-5 \times 10^6$ 、 $1-6 \times 10^6$ 、 $1-7 \times 10^6$ 、 $1-8 \times 10^6$ 、 9×10^6 、 1×10^7 、 $1-2 \times 10^7$ 、 $1-3 \times 10^7$ 、 $1-4 \times 10^7$ 、 $1-5 \times 10^7$ 、 $1-6 \times 10^7$ 、 $1-7 \times 10^7$ 、 $1-8 \times 10^7$ 、 $1-9 \times 10^7$ 、 $1-1 \times 10^8$ 、 $1-2 \times 10^8$ 、 $1-3 \times 10^8$ 、 $1-4 \times 10^8$ 、 $1-5 \times 10^8$ 、 $1-6 \times 10^8$ 、 $1-7 \times 10^8$ 、 $1-8 \times 10^8$ 、 $1-9 \times 10^8$ 、 $1-1 \times 10^9$ 、 $1-2 \times 10^9$ 、 $1-3 \times 10^9$ 、 $1-4 \times 10^9$ 、 $1-5 \times 10^9$ 、 $1-6 \times 10^9$ 、 $1-7 \times 10^9$ 、 $1-8 \times 10^9$ 、 $1-9 \times 10^9$ 、 $1-1 \times 10^{10}$ 、 $1-2 \times 10^{10}$ 、 $1-3 \times 10^{10}$ 、 $1-4 \times 10^{10}$ 、 $1-5 \times 10^{10}$ 、 $1-6 \times 10^{10}$ 、 $1-7 \times 10^{10}$ 、 $1-8 \times 10^{10}$ 、 $1-9 \times 10^{10}$ 、 $1-1 \times 10^{11}$ 、 $1-2 \times 10^{11}$ 、 $1-3 \times 10^{11}$ 、 $1-4 \times 10^{11}$ 、 $1-5 \times 10^{11}$ 、 $1-6 \times 10^{11}$ 、 $1-7 \times 10^{11}$ 、 $1-8 \times 10^{11}$ 、 $1-9 \times 10^{11}$ 、 $1-1 \times 10^{12}$ 、 $1-2 \times 10^{12}$ 、 $1-3 \times 10^{12}$ 、 $1-4 \times 10^{12}$ 、 $1-5 \times 10^{12}$ 、 $1-6 \times 10^{12}$ 、 $1-7 \times 10^{12}$ 、 $1-8 \times 10^{12}$ 或 $1-9 \times 10^{12}$ 个免疫球蛋白或TCR序列。

[0155] 在一些实施方案中,目标多核苷酸的长度为约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000或20,000个碱基或碱基对。在一些实施方案中,目标多核苷酸的长度为至少约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000或20,000个碱基或碱基对。在一些实施方案中,目标多核苷酸的长度为最多约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000或20,000个碱基或碱基对。在一些实施方案中,目标多核苷酸的长度为约10-20、10-30、10-40、10-30、10-40、10-50、10-60、10-70、10-80、10-90、10-100、50-60、50-70、50-80、50-90、50-100、100-200、100-300、100-400、100-300、100-400、100-500、100-600、100-700、100-800、100-900、100-1000、500-600、500-700、500-800、500-900、500-1000、1000-2000、1000-3000、1000-4000、1000-3000、1000-4000、1000-5000、1000-6000、1000-7000、1000-8000、1000-9000、1000-10000、5000-6000、5000-7000、5000-8000、5000-9000或5000-10000个碱基或碱基对。在一些实施方案中,目标多核苷酸或其片段的平均长度可为小于约100、200、300、400、500或800个碱基对,或小于约5、10、20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190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(诸如含有目标多核苷酸的样品)

的目标序列为约40、45、50、55、60、65、70、75、80、85、90、95或100个碱基。在某些实施方案中,使用含有与疾病或病况相关联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的序列进行比对。

[0156] 免疫组库测序

[0157] 本发明采用了操作核酸以便生成多核苷酸文库以供测序的步骤。在一些实施方案中,本发明采用了操作核酸以便产生重组单克隆抗体的步骤。在一般意义上,在本发明的一些实施方案中,使用免疫细胞和/或T细胞遗传物质的扩增,例如逆转录聚合酶链反应(逆转录-PCR),来生成免疫细胞遗传物质的cDNA扩增。对于抗体分子来说,可从免疫细胞或T细胞的基因组DNA或mRNA获得免疫球蛋白基因。RNA可为重链(V、D、J区段)或轻链(V、J区段)。在一些实施方案中,起始材料为来自免疫细胞的RNA,其由编码抗体的V、D、J基因区段构成,并含有恒定区。

[0158] 可使用一种多核苷酸或多核苷酸集合体将多核苷酸起始材料如RNA逆转录为cDNA。多核苷酸可包含与RNA的区域(诸如在恒定区中)或与mRNA的聚A尾互补的部分。容器条码可为具有或不具有已知嵌入碱基位置的一段约20个简并核苷酸,诸如NNNNWNNNNWNNNNWNNNNW,其中W意指A或T。

[0159] 可用一个或多个条码,例如,用容器条码和分子条码标记由逆转录产生的cDNA。具有特定设计的各种寡核苷酸可用于标记。逆转录产生的经标记的cDNA可扩增一次或多次,诸如通过PCR扩增。具有特定设计的各种引物可用于扩增。可使用第二扩增反应来扩增第一扩增反应如PCR的产物,诸如第一或第二PCR期。多种引物可用于扩增步骤。可使用本文所述的方法生成经扩增的多核苷酸的文库。产生的文库可包含具有适当分子和容器条码的完整或部分的抗体或TCR。

[0160] 在其他实施方案中,可使用模板转换来生成文库,诸如用于免疫组库测序。例如,可在逆转录期间使用模板转换来生成逆转录产物上与带有条码的多核苷酸(诸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。可在逆转录期间使用模板转换来去除PCR偏向的问题。这些方法可用于抗体测序,诸如通过使用高通量测序平台。

[0161] 起始材料可为RNA或DNA,诸如来自包含编码抗体的V、D、J基因区段且含有恒定区的免疫细胞或T细胞。在一些实施方案中,目标多核苷酸包含重链区段(V、D、J区段)或轻链区段(V、J区段)。

[0162] 可使用一种多核苷酸或多核苷酸集合体将目标多核苷酸逆转录为cDNA。用于逆转录目标多核苷酸的多核苷酸集合体中引物的实例可包含与目标多核苷酸的区域互补的部分。在一些实施方案中,与目标多核苷酸的区域互补的部分可与目标多核苷酸如mRNA的恒定区或与聚A尾互补。多种多核苷酸,诸如引物,可用来使一个或多个恒定区退火。可使用逆转录酶来进行逆转录反应。在特定实施方案中,逆转录酶可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的末端时,其可添加三个或更多个非模板残基,诸如三个或更多个非模板胞嘧啶残基。在一些实施方案中,Superscript II™逆转录酶用于此目的。在一些实施方案中,Maxima™逆转录酶用于此目的。在一些实施方案中,Protoscript II™逆转录酶用于此目的。在一些实施方案中,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,HighScriber™逆转录酶用于此目的。在一些实施方案中,末端脱氧核苷酸转移酶用于此目的。在一些实施方案中,禽成髓细胞瘤病毒

(AMV) 逆转录酶用于此目的。可使用能够转录RNA的具有非模板末端转移酶活性的任何逆转录酶。可使用能够转录RNA的具有非模板末端转移酶活性的任何逆聚合酶。可使用能够转录DNA的具有非模板末端转移酶活性的任何逆聚合酶。

[0163] 如上所述的逆转录反应可在3' 标记多核苷酸的存在下进行。3' 标记多核苷酸可为用于向目标多核苷酸如cDNA的3' 端添加核酸的多核苷酸。3' 标记多核苷酸可为用作向目标多核苷酸如cDNA的3' 端添加核酸的模板的多核苷酸。3' 标记多核苷酸可为与目标多核苷酸如cDNA的3' 端杂交的多核苷酸。3' 标记多核苷酸可为含有与目标多核苷酸如cDNA的3' 端杂交的3' 区域如3' 末端区域的多核苷酸。例如, 3' 标记多核苷酸可包含区段, 诸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' 标记多核苷酸为分子条码多核苷酸。在一些实施方案中, 3' 标记多核苷酸可包含分子条码。在一些实施方案中, 3' 标记多核苷酸可在3' 端上包含3个核糖-鸟嘌呤残基或其类似物(rGrGrG) (RNA碱基), 后者与通过逆转录酶产生的链互补且退火。在一些实施方案中, 可使用三个或更多个鸟嘌呤残基而不是核糖-鸟嘌呤(DNA核苷酸而不是RNA核苷酸)。在一些实施方案中, 3' 标记多核苷酸可包含在3' 端上的1个或2个核糖-鸟嘌呤残基和在3' 端上的脱氧核糖-鸟嘌呤残基或其类似物(rGrGG), 其与通过逆转录酶产生的链互补且退火。

[0164] 在反应中, 一旦3' 标记多核苷酸与cDNA链的CCC退火, 逆转录酶即可继续将cDNA延伸至标记多核苷酸, 从而将分子条码或其互补体附接至多核苷酸的目标群体如cDNA。例如, 3' 标记多核苷酸可为这样的多核苷酸, 其含有与目标多核苷酸的3' 端杂交的3' 区域的5' 侧区域。所述与目标多核苷酸的3' 端杂交的3' 区域的5' 侧区域可包含不与目标多核苷酸如cDNA互补的区域。所述与目标多核苷酸的3' 端杂交的3' 区域的5' 侧区域可包含分子条码。所述与目标多核苷酸的3' 端杂交的3' 区域的5' 侧区域可包含与容器条码化多核苷酸或其互补体互补的区域。在其他实施方案中, 模板转换可在单独的反应中进行。例如, 可在逆转录反应后添加3' 标记多核苷酸, 并可使用诸如逆转录酶或聚合酶等酶延伸至标记多核苷酸中。因为标记多核苷酸可在容器中的每个分子上带有独特的简并分子条码, 所以容器中的每个cDNA都可用分子条码独特地标记。在一些实施方案中, 模板转换可在逆转录反应进行的同时进行。

[0165] 在一些实施方案中, 3' 标记多核苷酸, 诸如分子条码化多核苷酸, 可进一步包含5' 区域, 诸如与含有另一个条码如容器条码的3' 标记多核苷酸或其互补体互补的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目标多核苷酸, 诸如标记的cDNA分子, 可包含3' 区域, 诸如与含有另一个条码如容器条码的3' 标记多核苷酸或其互补体互补的3' 末端区域。

[0166] 在一些实施方案中, 3' 标记多核苷酸为容器条码化多核苷酸。一旦从目标多核苷酸生成含有分子条码或其互补体的多核苷酸, 容器条码即可添加至分子条码化目标多核苷酸。3' 标记多核苷酸可为用于向目标多核苷酸如分子条码化目标多核苷酸的3' 端添加核酸的多核苷酸。3' 标记多核苷酸可为用作向目标多核苷酸如分子条码化目标多核苷酸的3' 端添加核酸的模板的多核苷酸。3' 标记多核苷酸可为与目标多核苷酸如分子条码化目标多核苷酸的3' 端杂交的多核苷酸。3' 标记多核苷酸可为含有与目标多核苷酸如分子条码化目标多核苷酸的3' 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包含与分子条码化目标多核苷酸的3' 端杂交的3' 区域, 诸如3' 末端区域。

[0167] 在反应中,一旦3'标记多核苷酸与分子条码化目标多核苷酸退火,逆转录酶即可继续将cDNA延伸至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多核苷酸的目标群体,如分子条码化目标多核苷酸。例如,3'标记多核苷酸可为这样的多核苷酸,其含有与分子条码化目标多核苷酸的3'端杂交的3'区域的5'侧区域。所述与分子条码化目标多核苷酸的3'端杂交的3'区域的5'侧区域可包含不与目标多核苷酸或分子条码化目标多核苷酸互补的区域。所述与分子条码化目标多核苷酸的3'端杂交的3'区域的5'侧区域可包含容器条码。

[0168] 在一些实施方案中,3'标记多核苷酸为扩增产物。在一些实施方案中,3'标记多核苷酸为来源于单分子的扩增产物。在一些实施方案中,3'标记多核苷酸为容器条码化多核苷酸的扩增产物。在一些实施方案中,3'标记多核苷酸为来源于单个容器条码化多核苷酸的扩增产物。与分子条码化目标多核苷酸的3'端杂交的3'区域的5'侧区域可包含与引物或其互补体互补的区域。与分子条码化目标多核苷酸的3'端杂交的3'区域的5'侧区域可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。

[0169] 然后可通过例如PCR扩增双条码化目标多核苷酸,诸如含有分子条码和容器条码的cDNA。然后可通过例如使用引物组进行PCR。然后上述PCR反应的产物可例如通过一轮或多轮PCR扩增一次或多次,或直接进行测序。

[0170] 根据本文所述的方法产生的文库可为包含经测序的大量或全部抗体或TCR序列的文库,该序列具有适当的条码,诸如容器条码和分子条码。在一些实施方案中,根据本文所述的方法产生的文库可含有适当的簇集区段以供测序。在一些实施方案中,可生成相同分子条码的许多拷贝。在一些实施方案中,对每个起始的独特目标多核苷酸分子,都可生成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,对用容器条码标记的每个起始的独特目标多核苷酸分子,都可生成含有相同分子条码的多核苷酸的许多拷贝。

[0171] 一经测序,即可匹配或配对具有相同分子条码的序列。一经测序,即可匹配或配对具有相同容器条码的序列。一经测序,即可匹配或配对具有相同目标序列的序列。在一些实施方案中,序列读取可分解为共有序列。将匹配或配对的序列读取分解为共有序列可由此减少或消除测序和PCR错误。对于第一读取可使用第一引物位点进行测序。对于第二读取可使用第二引物位点进行测序。对于第二读取可使用第二引物位点进行测序。

[0172] 含有相同容器条码的抗体重链和轻链可以配对,并且在一些实施方案中,可在哺乳动物载体系统中克隆。可在其他人类或哺乳动物宿主细胞系中表达该抗体构建体。然后可通过对表达的感兴趣的抗体或TCR进行瞬时转染分析和Western印迹分析来验证该构建体。

[0173] 扩增RNA或DNA的方法是本领域公知的,并可基于本文所提出的教导和指导根据本发明来使用,而无需过多的实验。已知的DNA或RNA扩增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4,795,699和4,921,794;授予Innis的5,142,033;授予Wilson等人的5,122,464;授予Innis的5,091,310;授予Gyllensten等人的5,066,584;授予Gelfand等人的4,889,818;授予Silver等人的4,994,370;授予Biswas的4,766,067;授予Ringold的4,656,134)以及RNA介导的扩增,其使用目标序列的反义RNA作为双链DNA合成的模板(授予Malek等人的美国专利号5,130,238,商标名为NASBA),这些参考文献的全部内

容均通过引用并入本文。(参见例如,Ausubel,同上;或Sambrook,同上。)

[0174] 适宜地,本文所述的方法步骤,诸如扩增、测序等,可以按照或不按照使用固相如阵列的复用分析形式进行,在该固相上固定有多种底物,例如抗原等。在一些实施方案中,该阵列为蛋白质生物芯片。使用蛋白质生物芯片,可筛选上百种甚至上千种抗原。如本文所用的,“阵列”、“微阵列”或“生物芯片”是指具有附着有吸附剂的、通常为平面的表面的固体基底。经常地,生物芯片的表面包含多个可寻址的位置,其中每个位置结合有吸附剂。生物芯片可适合于接合探针界面,并因此起到探针的作用。“蛋白质生物芯片”是指适合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. 270:103-111;Ge,2000,Nucleic Acids Res. 28,e3,1-V_H;MacBeath和Schreiber,2000,Science 289:1760-1763;WO 01/40803和WO 99/51773A1。阵列的使用允许许多步骤如筛选自动地和/或以高通量方式进行。用于阵列的多肽可在高速率下点样,例如使用可商购获得的机器人装置,例如,来自Genetic MicroSystems或BioRobotics。阵列基底可为,例如,硝化纤维、塑料、玻璃,例如表面改性的玻璃。阵列还可包括多孔基质,例如,丙烯酰胺、琼脂糖或其他聚合物。一旦在生物芯片上捕获,即可通过多种检测方法检测分析物,该方法选自,例如,气相离子光谱法、光学法、电化学法、原子力显微术和射频法。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、透光率、双折射率或折射率(例如,表面等离子体共振、椭圆测量术、共振镜法、光栅耦合器波导法或干涉测量法)。光学法包括显微术(共聚焦和非共聚焦)、成像方法和非成像方法。多种形式的免疫测定法(例如,ELISA)是用于检测在固相上捕获的分析物的流行方法。电化学法包括伏安测量法和电流滴定法。射频法包括多级共振光谱法。

[0175] 在本发明的一些实施方案中,例如,用于制备单克隆抗体的自然多样性途径,使用已经针对用单细胞运行而建立的技术。一种技术包含可在FACS中使用以使单细胞偏转到单独的容器中的特殊配件。这类配件可商购获得,且为本领域公知的。这类配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的区室内。或者,细胞可通过有限稀释而沉积在微量滴定板内以确保单细胞沉积。

[0176] 第二种技术为在单个免疫细胞上进行的PCR,用以扩增V_H和V_L区段。在自然多样性途径中,使用单细胞PCR来保持单细胞中V_L和V_H的天然配对。抗体的特异性取决于V_L区和V_H区内的互补决定区(CDR)。

[0177] 用于进行单细胞PCR的方法是本领域公知的(例如,Larrick,J.W.等人,Bio/Technology 7:934(1989))。例如,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细胞可用固定剂溶液或含有诸如甲醛、戊二醛等化学品的溶液固定。然后用含有例如去污剂的透化溶液使细胞透化。固定和透化过程应该提供足够的多孔性以允许酶、核苷酸和其他试剂进入细胞内,而不会过度破坏细胞区室或其中的核酸。然后添加的酶和核苷酸可进入细胞以将细胞的V_H和V_L或V_α和V_β或V_γ和V_δmRNA逆转录为例如相应的cDNA序列。逆转录可在单个步骤中进行,或任选地与PCR程序一起进行,其进行使用逆转录酶、足量的四种dNTP以及与mRNA结合的引物,从而为逆转录酶提供3' 羟基以启动聚合。可使用与mRNA互补的任何引物,但优选使用与V_H和V_L或V_α和V_β或V_γ和V_δ分子的3' 末端互补的引物,以便帮助选择可变区mRNA。大量研究已经表明,可制备简并多核苷酸作为用于V_H和V_L或V_α和

V β 或V γ 和V δ 的5'端引物。制备目标分子的组合文库方法依赖于这样的引物。此外,大量实验已经显示,PCR可从单细胞扩增感兴趣的基因区段,诸如V $_H$ 和V $_L$ 或V α 和V β 或V γ 和V δ 。因为甚至采用单细胞也能工作,所以该PCR方法甚至在感兴趣的免疫细胞以低频率存在的情况下也能生成抗体。

[0178] 在高度多样性的实施方案中,在FACS分选后,合并免疫细胞文库的细胞,并在整个细胞集合体上进行逆转录-PCR。通过用于制备并表征抗体或TCR的公知程序易于完成了克隆抗体或TCR目的而生成mRNA(参见例如,Antibodies:A Laboratory Manual,1988;通过引用而并入本文)。例如,通过本领域标准且常规的适当方法提取来自B细胞文库的总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,如以上所解释的,这些是本领域技术人员已知的过程。来源于B细胞或T细胞文库的核酸分子的文库,例如,来源于这类B或T淋巴细胞的RNA或cDNA分子的文库,可克隆至表达载体中以形成表达文库。在一些实施方案中,只扩增来源于免疫细胞文库的V $_H$ 或V α 或V γ 结构域以生成V $_H$ 或V α 或V γ 结构域的文库。来自另一个来源的V $_L$ 或V β 或V δ 文库与V $_H$ 或V α 或V γ 文库组合使用,以利用本文所述的方法生成抗体或TCR。可通过熟练技术人员已知的许多方式,通过将V $_H$ 和V $_L$ 或V α 和V β 或V γ 和V δ 文库组合在一起构建抗体或TCR片段的文库。例如,每个文库可在不同载体中创建,且这些载体在体外或体内重组。或者,文库可顺序地克隆至同一个载体中,或通过PCR装配在一起,然后克隆。PCR装配还可用来将V $_H$ 和V $_L$ 或V α 和V β 或V γ 和V δ DNA与编码柔性肽间隔区的DNA连接起来以形成如本文他处描述的单链Fv(scFv)文库。在另一种技术中,使用细胞内PCR装配来通过PCR在淋巴细胞内组合V $_H$ 和V $_L$ 或V α 和V β 或V γ 和V δ 基因,然后克隆连接的基因的组库。

[0179] 单细胞条码化

[0180] 对于采用容器条码和分子条码的单细胞条码化,可以用这样的方式创建容器,诸如油包水乳液:产生的容器含有每个容器1个或更少的细胞。可以用这样的方式创建容器:产生的容器还含有每个容器1个容器条码。可以用这样的方式创建容器:产生的容器还含有每个容器1个分子条码化多核苷酸。可以用这样的方式创建容器:产生的容器还含有每个容器两个或更多个或多种分子条码化多核苷酸。细胞/容器可经历如本文所述的RNA或DNA单条码化方案,且每个容器的容器条码和一个或多个分子条码可与感兴趣的靶标如细胞多核苷酸融合。在一些实施方案中,匹配容器条码化多核苷酸可与同所述一个或多个细胞条码化多核苷酸存在于相同容器中的细胞组分融合。在测序之后,可使用容器条码和细胞条码去卷积来鉴别哪个RNA(或DNA)来源于哪个细胞。在一些实施方案中,可以用这样的方式创建容器,诸如油包水乳液:产生的乳液含有每份乳液1个或更多个细胞。在一些实施方案中,可以用这样的方式创建油包水乳液:产生的乳液含有每个容器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。在一些实施方案中,可以用这样的方式创建容器:产生的容器含有每个容器超过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。在一些实施方案中,当在溶液中时,可将容器条码和分子条码引入容器中。在一些实施方案中,当不与固体支持体如珠子附接时,可将容器条码和分子条码引入容器中。

[0181] 在一些方面,单细胞可隔离在乳液内,该乳液可充当区室。可裂解细胞,并且可对来自该细胞的转录物进行条码化。每个转录物可与分子条码或容器条码融合,以便当检测

到两个或更多个RNA转录物具有相同容器条码时,可确定它们来源于相同的起始细胞。这可应用于许多不同类型的序列。一种具体的应用可为连接抗体和TCR序列的 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 链。

[0182] 可在容器条码和分子条码的存在下在一份或多份乳液中隔离一个或多个单细胞,以便该一份或多份乳液的一滴可含有最多1个或更少的细胞。可通过乳液中含有的缓冲液以化学方式裂解或通过冻融来裂解细胞,从而在乳液中释放细胞的内容物。

[0183] 单细胞的RNA可逆转录为cDNA。可使用如上所述添加约3个胞嘧啶残基的具有非模板末端转移活性的逆转录酶完成逆转录反应。所有逆转录缓冲液、酶和核苷酸可在形成乳液时存在。在一些实施方案中,可使引物通用化(诸如含有聚dT序列的多核苷酸)以靶向所有mRNA。在一些实施方案中,可使用DNA。在一些实施方案中,可靶向超过2个RNA。

[0184] 在一些实施方案中,在逆转录期间,容器条码可与RNA相连接。在一些实施方案中,在逆转录期间,分子条码可与RNA相连接。在一些实施方案中,在逆转录期间,容器条码和分子条码可与RNA相连接。

[0185] 逆转录反应可在3' 标记多核苷酸的存在下进行。3' 标记多核苷酸可包含可用于使测序引物退火的P7区段。3' 标记多核苷酸可包含容器条码或分子条码。3' 标记多核苷酸可在3' 端上包含3个核糖-鸟嘌呤残基(rGrGrG) (RNA碱基),后者可与通过逆转录酶产生的链互补且退火。从而,容器条码和分子条码可在此相同乳液中通过逆转录酶添加到cDNA的末端。在一些实施方案中,可使用鸟嘌呤残基而不是核糖-鸟嘌呤(DNA核苷酸而不是RNA核苷酸)。在反应中,一旦3' 标记多核苷酸与cDNA链的CCC退火,则逆转录酶继续将cDNA延伸至标记3' 多核苷酸内,从而对所有cDNA创建分子条码化标签。在反应中,一旦3' 标记多核苷酸与分子条码化cDNA的区域退火,则逆转录酶或聚合酶继续将分子条码化cDNA延伸至另一个3' 标记多核苷酸内,从而对所有cDNA创建容器条码化标签。在一些实施方案中,模板转换可在单独的反应中进行,而不是在可进行逆转录反应的同时进行。在一些实施方案中,可在逆转录反应后添加3' 标记多核苷酸,并可以用相似的方式使用诸如逆转录酶或聚合酶等酶将其延伸到标记多核苷酸内。因为3' 标记多核苷酸可在每个单分子上带有独特的简并分子条码,所以每个cDNA可用分子条码独特地标记。因为3' 标记多核苷酸可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可用对于该容器独特的容器条码标记。

[0186] B细胞文库遗传物质的克隆和表达

[0187] 如本文所用的“抗体表达文库”或“TCR表达文库”或“表达文库”可指在核酸或蛋白质水平上的分子的集合(即,两个或更多个分子)。因此,该术语可指编码多种抗体或TCR分子的表达载体的集合(即,在核酸水平上),或可指已在适当的表达系统中表达后的抗体或TCR分子的集合(即,在蛋白质水平上)。或者,表达载体/表达文库可被包含在可表达它们的合适的宿主细胞内。在本发明的表达文库中编码或表达的抗体分子可为任何适当的形式,例如,可为完整的抗体或TCR分子,或者可为抗体或TCR片段,例如,单链抗体(例如,scFv抗体)、Fv抗体、Fab' 抗体、(Fab')₂片段、双抗体等。如“编码”特定酶的核酸序列或特定酶的DNA编码序列或“编码”特定酶的核苷酸序列中的术语“编码”,以及其他同义术语,是指当置于适当调节序列的控制下时转录并翻译为酶的DNA序列。“启动子序列”是能够与细胞中的RNA聚合酶结合并启动下游(3' 方向)编码序列的转录的DNA调节区。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。启动子序列包含最小数目的碱基,以及以高于背景的可检测水平启动转录所必需的元件。然而,在RNA聚合酶与序列结合并且转录在起始密码子处(具有启动子的3'末端)启动之后,转录以3'方向向下游进行。在启动子序列内将发现转录起始位点(方便地通过用核酸酶S1作图来定义)以及负责RNA聚合酶结合的蛋白质结合域(共有序列)。

[0188] 由本发明的抗体或TCR表达文库鉴别、衍生、选择或获得的抗体或TCR分子构成本发明的又一方面。同样,这些抗体或TCR分子可为蛋白质或编码抗体或TCR分子的核酸,该核酸进而可并入适当的表达载体中和/或包含在合适的宿主细胞内。

[0189] cDNA集合体可同与抗体基因的重链的恒定区杂交的多核苷酸以及与抗体或TCR基因的 V_H 或 V_α 或 V_γ 链区域的5'端杂交的多核苷酸一起经历PCR反应。cDNA集合体可同与抗体或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包含抗体或TCR序列的条码化多核苷酸的 V_H 或 V_α 或 V_γ 链区域5'端的5'侧区域杂交的多核苷酸一起经历PCR反应。还可设置PCR反应用于扩增例如 κ 和 λ 类别的 V_L 或 V_β 或 V_δ 链集合体。cDNA集合体可同与抗体基因的轻链的恒定区杂交的多核苷酸以及与抗体或TCR基因的 V_L 或 V_β 或 V_δ 链区域的5'端杂交的多核苷酸一起经历PCR反应。cDNA集合体可同与抗体基因的轻链的恒定区杂交的多核苷酸以及与包含抗体或TCR序列的条码化多核苷酸的 V_L 或 V_β 或 V_δ 链区域5'端的5'侧区域杂交的多核苷酸一起经历PCR反应。可基于已知且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这类寡核苷酸或引物。

[0190] 在一些实施方案中,可从使用对重链或轻链基因,特别是对 V_H 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 多核苷酸的末端区域之一或两者均无特异性的一个或多个引物通过PCR扩增产生的 V_H 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 序列的文库便利地获得 V_H 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 序列。在一些实施方案中,可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_H 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 序列的文库便利地获得 V_H 和 V_L 序列。在一些实施方案中,可从使用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_H 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 序列的文库便利地获得 V_H 和 V_L 序列。在一些实施方案中,可从使用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和作为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的引物组通过PCR扩增产生的 V_H 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 序列的文库便利地获得 V_H 和 V_L 序列。在一些实施方案中,可从使用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和对通用序列具有特异性的第二引物的引物组通过PCR扩增产生的 V_H 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 序列的文库便利地获得 V_H 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 序列。

[0191] 在一些实施方案中,一旦逆转录,可使用对免疫球蛋白基因,特别是对 V_H 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 多核苷酸的末端区域之一或两者均具有特异性的一个或多个引物,通过PCR来扩增所得到的cDNA序列。在一些实施方案中,可从使用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_H 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 序列的文库获得 V_H 和 V_L 序列(Nicholls等人,J.Immunol.Meth.,1993,165:81;W093/12227),或基于可获得的序列信息根据本领域已知的标准方法设计 V_H 和 V_L 序列。 $(V_H$ 和 V_L 或 V_α 和 V_β 或 V_γ 和 V_δ 序列可进行连接,通常采用介于中间的间隔区序列(例如,编码符合读框的柔性肽间隔区),从而形成编码单链抗体的盒)。对于免疫球蛋白表达细胞,V区域序列可便利地克隆为cDNA或PCR扩增产物。任选地,在本文所述的方法中,并且特别是在指出的某些步骤之后(例如,在单细胞PCR后;

在哺乳动物或其他细胞表面展示后,在FACS筛选后,等等),对 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 区域进行测序。除其他原因外,测序可用来证实多样性水平为可接受的水平。测序可包括高通量测序、深度测序(其中从多个单独的样品对相同的基因进行测序,以鉴别序列中的差异)或二者的组合。

[0192] 在一些实施方案中,不必要使用本文所述的方法在物理上连接天然的 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 组合。在一些实施方案中,cDNA、条码化多核苷酸或PCR扩增的条码化cDNA未物理连接。在一些实施方案中,cDNA、条码化多核苷酸或PCR扩增的条码化cDNA未在相同的反应或容器中物理连接。

[0193] 在一些实施方案中,除了cDNA引物之外,还使用针对 V_H 或 $V\alpha$ 或 $V\gamma$ 基因的5'端的一个引物或多个引物和针对 V_L 或 $V\beta$ 或 $V\delta$ 基因的5'端的另一个引物或多个引物,对天然的 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 组合进行物理连接。这些引物还含有额外序列的互补尾,以允许 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 基因的自装配。在PCR扩增和连接后,得到混合产物即混合可变区的机会减至最小,因为扩增和连接反应在每个细胞内进行。可通过利用大体积试剂如地高辛标记的核苷酸来进一步确保V区域cDNA对不离开细胞区室并相互混合,而是停留在细胞内以供PCR扩增和连接,从而进一步降低混合的风险。扩增的序列通过互补末端序列的杂交而连接。在连接后,可从细胞中回收序列,以供在本文所述的进一步方法步骤中使用。例如,如果必要的话,回收的DNA可使用末端引物进行PCR扩增,并克隆入载体中,该载体可以是如下详述的质粒、噬菌体、粘粒、噬菌粒、病毒载体或其组合。可将适宜的限制酶位点并入杂交的序列中以帮助克隆。这些载体还可保存为连接的可变区的文库供以后使用。

[0194] 在期望提供附加的 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 组合的一些实施方案中,选择表达系统以帮助这一点。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是本领域技术人员已知的。

[0195] 应当注意,在 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 序列来源于非人类的情况下,在一些实施方案中,可优选将这些序列与完全人类Fc嵌合。如本文所用的,“嵌合”是指这样的免疫球蛋白或TCR,其中重链和轻链可变区或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 区不是人源的,且其中重链和轻链或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 链的恒定区是人源的。这通过将可变域扩增并克隆到人类Fc中来实现。该人类Fc可为载体的一部分,或在单独的分子中,并且还可使用Fc的文库。在优选的实施方案中,嵌合分子在哺乳动物细胞如CHO细胞中生长,用FACS筛选两次,以针对表达感兴趣的抗体的细胞来富集该细胞群体。通过测序然后功能性表征或通过直接功能性表征或动力学来表征嵌合抗体或TCR。生长、筛选和表征详细描述如下。

[0196] 重要的是要注意到,以上所述的PCR反应均针对克隆IgG形式的抗体而描述。这些是优选的,因为它们通常与更成熟的免疫应答相关,并且通常比IgM抗体展现出更高的亲和力,从而使它们更理想地用于某些治疗和诊断应用。然而,显然可设计多核苷酸,如果需要或适当的话,其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。

[0197] 一旦鉴别了抗体或TCR,并且适当的所述细胞的群体在适当的时间已经分离并任选地如上所述进行富集,则抗体或TCR表达文库不需要立刻生成,这使得包含在细胞内的遗传物质可保持完整,从而使文库能够在以后的日期制备。因此,可通过适当的方法,例如,通过冷冻,将例如细胞、细胞裂解物或核酸,例如由其衍生的RNA或DNA储存到以后的日期,且

表达文库在需要时在以后的日期生成。

[0198] 一旦生成了表达载体的文库,则编码的抗体分子可在适当的表达系统中表达,并使用本领域公知和记载的适当技术进行筛选。因此,以上定义的本发明方法可包括在适当的表达系统中表达表达载体的文库,并且针对具有所需性质的抗体筛选表达文库的进一步的步骤,进一步详细解释如下。

[0199] 如本文所表明的,通过本发明的方法制备的、包括编码抗体或TCR序列的多核苷酸的多核苷酸可包括但不限于,编码抗体或TCR片段的氨基酸序列的多核苷酸本身,整个抗体或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,以及附加序列,诸如至少一个信号前导区或融合肽的编码序列,其具有或不具有上述的附加编码序列,诸如至少一个内含子,以及附加的非编码序列,包括但不限于非编码5'和3'序列,诸如在转录、RNA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)中起作用的转录的非翻译序列;编码附加氨基酸的附加编码序列,诸如提供附加功能性的序列。因此,编码抗体的序列可融合至标记序列,诸如编码帮助纯化包含抗体或TCR片段或部分的融合抗体或TCR的肽的序列。

[0200] 然后初级PCR产物可任选地经历使用新的多核苷酸组的次级PCR反应,该多核苷酸组与抗体或TCR可变域 V_H 、 $V_L\kappa$ 和 $V_L\lambda$ 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 的5'和3'端杂交(视情况取决于使用新多核苷酸组的初级PCR反应是否经过设计以扩增重链或轻链抗体基因或 $V\alpha$ 或 $V\beta$ TCR基因或 $V\gamma$ 或 $V\delta$ TCR基因的部分)。这些多核苷酸有利地包含对限定的一组限制酶具有特异性的DNA序列(即,限制酶位点)以供后续克隆。选定的限制酶必须进行选择以便不在人类抗体或TCR V-基因区段内切割。可基于已知且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这样的多核苷酸。然而,所包括的优选限制酶位点为NcoI、Hind III、MluI和NotI。这样的次级PCR反应的产物为多种V-重、V-轻 κ 和V-轻 λ 抗体片段/结构域的组库。因此,这种类型的次级PCR反应通常在以下情况时进行:感兴趣的表达文库形式为scFv或Fv形式,其中只存在抗体或TCR的 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 结构域。

[0201] PCR产物还可经历采用与条码化多核苷酸的5'和3'端杂交的新引物组的PCR反应。这些多核苷酸可有利地包含对限定的一组限制酶具有特异性的DNA序列(即,限制酶位点)以供后续克隆。选定的限制酶必须进行选择以便不在人类抗体或TCR V-基因区段内切割。可基于已知且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这样的多核苷酸。然而,所包括的优选限制酶位点为NcoI、Hind III、MluI和NotI。这样的次级PCR反应的产物为各种 V_H 、 $V_L\kappa$ 和 $V_L\lambda$ 抗体片段/结构域或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ TCR片段/结构域的组库。

[0202] 本领域技术人员将会认识到,该系统还可使用重链或轻链或 $V\alpha$ 或 $V\beta$ 链或 $V\gamma$ 或 $V\delta$ 链Fv或Fab片段或者单链抗体或TCR。重链或轻链或 $V\alpha$ 或 $V\beta$ 链或 $V\gamma$ 或 $V\delta$ 链可经诱变,然后向溶液中添加互补链。然后使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 $V\alpha$ 或 $V\beta$ 链或 $V\gamma$ 或 $V\delta$ 链序列的添加允许产生组合系统以生成多样性成员的文库。

[0203] 这类克隆片段组库的文库构成了本发明的其他方面,所述片段包含由如本文所定义的受免疫攻击的宿主的B或T淋巴细胞衍生的抗体或TCR基因的可变重链或 $V\alpha$ 链或 $V\gamma$ 链区域或其片段,和/或可变轻链或 $V\beta$ 链或 $V\delta$ 链区域或其片段。包含克隆的可变区的这些文库可任选地插入表达载体中以形成表达文库。

[0204] 在一些实施方案中,可设置PCR反应以便保持分离的免疫细胞群体中含有的各种抗体或TCR链的全部或部分恒定区。这在以下情况时是理想的:表达文库形式为Fab形式,其中重链或 α 或 γ 链组分包含 V_H 或 V_α 或 V_γ 和 C_H 或 C_α 或 C_γ 结构域,且轻链或 V_B 链或 V_δ 链组分包含 V_L 或 V_B 或 V_δ 链和 C_L 或 C_B 或 C_δ 结构域。同样,包含抗体或TCR链的所有或部分恒定区的这类克隆片段的文库构成了本发明的进一步的方面。

[0205] 这些核酸可便利地包含除了本发明的多核苷酸之外的序列。例如,可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,以帮助多核苷酸的分离。同样,可插入可翻译的序列以帮助分离本发明的翻译的多核苷酸。例如,六聚组氨酸标记序列提供了纯化本发明的蛋白质的便利方式。本发明的核酸,排除编码序列,任选地为用于克隆和/或表达本发明的多核苷酸的载体、衔接子或连接体。

[0206] 附加序列可添加至这样的克隆和/或表达序列上,以优化它们在克隆和/或表达中的功能,帮助多核苷酸的分离,或改善多核苷酸向细胞内的引入。克隆载体、表达载体、衔接子和连接体的使用是本领域公知的。(参见例如,Ausubel,同上;或Sambrook,同上)。

[0207] 本文所公开的文库可用于多种应用。如本文所用的,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、扩增子或扩增子集的多个序列读取。文库可多次储存并多次使用以生成用于分析的样品。一些应用包括,例如,对多态性进行基因型分型,研究RNA加工,和选择克隆代表来根据本文所提供的方法进行测序。可以生成包含多个多核苷酸的文库,诸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、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、50,000,000、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,多核苷酸的文库包含多个至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、50,000,000、100,000,000个或更多个独特的多核苷酸,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。

[0208] 条码

[0209] 条码可为分子条码或容器条码。在一些实施方案中,条码,诸如分子条码或容器条码,可各自具有2至36个核苷酸、4至36个核苷酸、或6至30个核苷酸、或8至20个核苷酸、2至20个核苷酸、4至20个核苷酸、或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彼此在10℃以内、彼此在5℃以内或彼此在2℃以内。在某些方面,组内条码的解链温度彼此不在10℃以内、彼此不在5℃以内或彼此不在2℃以内。在其他方面,条码是最低交叉杂交组的成员。例如,这样的组的每个成员的核苷酸序列可与该组的每个其他成员的核苷酸序列充分不同,使得在严格杂交条件下,没有成员可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,最低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有至少两个核苷酸不同。条码技术描述于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* 101:11046; 以及Brenner (2004) *Genome Biol.* 5:240。

[0210] 如本文所用的,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独特的信息。如本文所用的,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单个容器的多核苷酸独特的信息。在一些实施方案中,独特的信息包含核苷酸的独特序列。例如,分子条码或容器条码的序列可通过确定含有该分子条码或容器条码的核苷酸的独特或随机序列的身份和顺序来确定。在一些实施方案中,独特的信息不能用于鉴别目标多核苷酸的序列。例如,分子条码可与一个目标多核苷酸相附接,但该分子条码不能用于确定其附接的目标多核苷酸。在一些实施方案中,独特的信息不是与目标多核苷酸的序列的身份相关联的已知序列。例如,容器条码可与一个或多个目标多核苷酸相附接,但该容器条码不能用于确定其附接的是该一个或多个目标多核苷酸中的哪一个。在一些实施方案中,独特的信息包含核苷酸的随机序列。在一些实施方案中,独特的信息包含多核苷酸上核苷酸的一个或多个独特序列。在一些实施方案中,独特的信息包含简并核苷酸序列或简并条码。简并条码可包含可变核苷酸碱基组成或序列。例如,简并条码可为随机序列。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。

[0211] 分子条码或容器条码可包含任何长度的核苷酸。例如,分子条码或容器条码可包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸。例如,分子条码或容器条码可包含最多约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸。在一些实施方案中,分子条码或容器条码具有特定的核苷酸长度。例如,分子条码或容器条码的长度可为约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸。

[0212] 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的长度可为至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500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最多约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的长度可为最多约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500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的核苷酸长度。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的长度可为2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500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的核苷酸长度。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具有约或至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸,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具有约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500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与所述一个或多个第二分子条码或容器条码的多核苷酸的数目不同。

[0213] 分子条码的数目可超过将在多个容器中标记的分子的总数目。容器条码的数目可超过将在多个容器中标记的分子的总数目。例如,分子条码或容器条码的数目可以是将在多个容器中标记的分子的总数目的至少约2、3、4、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0214] 不同分子条码的数目可超过将在多个容器中标记的分子的总数目。在一些实施方案中,不同分子条码的数目是将在多个容器中标记的分子的总数目的至少约1、1.5、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0215] 单个容器中不同分子条码的数目可超过将在单个容器中标记的不同分子的数目。在一些实施方案中,单个容器中不同分子条码的数目是将在单个容器中标记的不同分子的数目的至少约1、1.5、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0216] 不同容器条码的数目可小于将在多个容器中标记的分子的总数目。在一些实施方案中,不同容器条码的数目比将在多个容器中标记的分子的总数目小至少约1、1.5、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0217] 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目可超过将在单个容器中标记的不同分子的数目。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目是将在单个容器中标记的不同分子的数目的至少约1、1.5、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0218] 单个容器中容器条码化多核苷酸分子的数目可小于将在单个容器中标记的不同分子的数目。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目比将在单个容器中标记的不同分子的数目小至少约1、1.5、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0219] 单个容器中容器条码化多核苷酸分子的数目可为一个分子。单个容器中非扩增的容器条码化多核苷酸分子的数目可为一个分子。

[0220] 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。

[0221] 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。

[0222] 分子条码或容器条码群体中的分子条码或容器条码可具有至少10、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000个或更多个不同的序列。例如,群体中的分子条码或容器条码可具有至少2,000、3,000、4,000、5,000、6,000、7,000、8,000、9,000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000个或更多个不同的序列。因此,多个分子条码或容器条码可用于从一个或多个多核苷酸如目标多核苷酸生成至少10、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000个或更多个不同的序列。例如,多个分子条码或容器条码可用于从一个或多个多核苷酸如目标多核苷酸生成至少2,000、3,000、4,000、5,000、6,000、7,000、8,000、9,000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 、 9×10^{12} 个或更多个不同的序列。例如,多个分子条码或容器条码可用于从至少约10、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 、 9×10^{12} 个或更多个目标多核苷酸生成至少约10、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、

900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 、 9×10^{12} 个或更多个不同的序列。

[0223] 在一些实施方案中,一个或多个分子条码用于将序列分组或分箱。在一些实施方案中,一个或多个分子条码用于将序列分组或分箱,其中每个箱元中的序列含有相同的分子条码。在一些实施方案中,一个或多个分子条码或容器条码用于将序列分组或分箱,其中每个箱元中的序列包含扩增子集。在一些实施方案中,一个或多个分子条码用于将序列分组或分箱,其中每个箱元中的序列包含多个序列,其中从中生成该多个序列的多核苷酸来源于扩增反应中的相同多核苷酸分子。

[0224] 在一些实施方案中,一个或多个容器条码用于将序列分组或分箱。在一些实施方案中,一个或多个容器条码用于将序列分组或分箱,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,一个或多个容器条码用于将序列分组或分箱,其中每个箱元中的序列包含一个或多个扩增子集。在一些实施方案中,一个或多个容器条码用于将序列分组或分箱,其中每个箱元中的序列包含多个序列,其中从中生成该多个序列的多核苷酸来源于来自单个容器或单细胞的多核苷酸。

[0225] 在一些实施方案中,一个或多个分子条码和容器条码用于将序列分组或分箱。在一些实施方案中,一个或多个分子条码和容器条码用于将序列分组或分箱,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,一个或多个分子条码和容器条码用于将序列分组或分箱,其中每个箱元中的序列包含一个或多个扩增子集。在一些实施方案中,一个或多个分子条码和容器条码用于将序列分组或分箱,其中每个箱元中的序列包含多个序列,其中从中生成该多个序列的多核苷酸来源于扩增反应中的相同多核苷酸且来自相同的单细胞或容器。在一些实施方案中,一个或多个分子条码和容器条码不用于比对序列。

[0226] 在一些实施方案中,一个或多个分子条码不用于比对序列。在一些实施方案中,一个或多个分子条码用于比对序列。在一些实施方案中,一个或多个分子条码用于将序列分组或分箱,且目标特异性区域用于比对序列。在一些实施方案中,一个或多个容器条码不用于比对序列。在一些实施方案中,一个或多个容器条码用于比对序列。在一些实施方案中,一个或多个容器条码用于将序列分组或分箱,且目标特异性区域用于比对序列。在一些实施方案中,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,一个或多个分子条码和容器条码用于将序列分组或分箱,且目标特异性区域用于比对序列。

[0227] 在一些实施方案中,比对的序列含有相同的分子条码。在一些实施方案中,比对的序列含有相同的容器条码。在一些实施方案中,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,一个或多个分子条码或容器条码用于比对序列,其中比对的序列包

含来自扩增子集的两个或更多个序列。在一些实施方案中,一个或多个分子条码或容器条码用于比对序列,其中比对的序列包含多个序列,其中从中生成该多个序列的多核苷酸来源于扩增反应中的相同多核苷酸分子。在一些实施方案中,一个或多个分子条码或容器条码用于比对序列,其中比对的序列包含多个序列,其中从中生成该多个序列的多核苷酸来源于单细胞或单个容器。

[0228] 小滴生成

[0229] 将多个细胞的样品分成小反应体积,结合来自或来源于来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,可使得能够对序列如生物标志物序列的组库进行高通量测序。

[0230] 将多个细胞的样品分成小反应体积,结合来自或来源于来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,可使得能够对序列(诸如代表生物体的转录物组的百分比的序列)的组库进行高通量测序。例如,序列组库可包含多个序列,所述序列代表生物体的转录物组的至少约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%。

[0231] 将免疫细胞的样品分成小反应体积,结合来自或来源于来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,可使得能够对重链和轻链序列的组库进行高通量测序。这些方法还可允许在基于条码化序列的测序后将重链和轻链配对。如本文所述将样品分成小反应体积还可使得能够使用减少的试剂量,从而降低该分析的材料成本。

[0232] 在一些情况下,逆转录反应和/或扩增反应(例如,PCR)在小滴中进行,诸如在小滴数字PCR中。在某些方面,本发明提供了流体区室以包含所有或部分的目标材料。在一些实施方案中,区室为小滴。当在本说明书各处提及“小滴”时,除非另有说明,该术语与流体区室和流体分区可互换使用。除非另有说明,为了方便起见使用“小滴”,并且可使用任何流体分区或区室。本文所使用的小滴可包括乳液组合物(或两种或多种不混溶液体的混合物),诸如美国专利号7,622,280中所述。可通过W0/2010/036352中所述的装置生成小滴。如本文所述,术语乳液可指不混溶液体(诸如油和水)的混合物。油相和/或油包水乳液允许水性小滴内反应混合物的区室化。乳液可含有连续油相中的水性小滴。本文所提供的乳液可为水包油乳液,其中小滴为连续水相中的油小滴。本文所提供的小滴被设计用于防止区室间的混合,每个区室保护其内容物不蒸发且不与其他区室的内容物聚结。

[0233] 本文所述的混合物或乳液可为稳定或不稳定的。该乳液可为相对稳定的,且具有最小的聚结。当小的小滴合并以形成逐渐更大的小滴时发生聚结。在一些情况下,由小滴生成器生成的少于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%的小滴与其他小滴聚结。该乳液还可具有有限的絮凝,这是分散相从悬浮液中以薄片形式生成的过程。

[0234] 可生成具有约、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微米的平均直径的小滴。小滴可具有约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、约1至约100微米的平均直径。使用微通道交叉流聚焦或物理搅拌产生乳液小滴的微流体方法已知用于产生单分散或多分散乳液。该小滴可为单分散小滴。可生成小滴以使得小滴的大小变化不超过小滴平均大小的加或减5%。在一些情况下,生成小滴以使得小滴的大小变化不超过小滴平均大小的加或减2%。小滴生成器可从单个样品生成小滴群体,其中小滴的大小变化都不超过小滴总群体的平均大小的加或减约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%。

[0235] 更高的机械稳定性对微流体操作和更高剪切的流体处理可能是有用的(例如,在微流体毛细管中或通过流体路径中的90度转角,诸如阀门时)。预先和随后热处理的小滴或胶囊对于标准移液管操作和离心可以是机械稳定的。

[0236] 可通过使油相流动穿过水性样品来形成小滴。水相可包含用于进行扩增反应的缓冲溶液和试剂,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、模板核酸和酶,诸如DNA聚合酶、RNA聚合酶和/或逆转录酶。

[0237] 水相可包含具有或不具有固体表面如珠子的用于进行扩增反应的缓冲溶液和试剂。缓冲溶液可包含约、多于约或少于约1、5、10、15、20、30、50、100或200mM Tris。在一些情况下,氯化钾的浓度可为约、大于约或小于约10、20、30、40、50、60、80、100或200mM。缓冲溶液可包含约15mM Tris和50mM KCl。核苷酸可包括脱氧核苷三磷酸分子,包括dATP、dCTP、dGTP和dTTP,每种的浓度为约、大于约或小于约50、100、200、300、400、500、600或700 μ m。在一些情况下,dUTP在水相中添加到约、大于约或小于约50、100、200、300、400、500、600或700、800、900或1000 μ m的浓度。在一些情况下,将氯化镁($MgCl_2$)或乙酸镁以约、大于约或小于约1.0、2.0、3.0、4.0或5.0mM的浓度添加到水相。 $MgCl_2$ 的浓度可为约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。

[0238] 可使用诸如BSA或来自牛皮肤的明胶的非特异性封闭剂,其中明胶或BSA的浓度范围为约0.1-0.9%w/v。其他可能的封闭剂可包括 β 乳球蛋白、酪蛋白、奶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。

[0239] 水相中用于扩增的引物可具有约、大于约或小于约0.05、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1.0、1.2、1.5、1.7或2.0 μ m的浓度。水相中的引物浓度可为约0.05至约2、约0.1至约1.0、约0.2至约1.0、约0.3至约1.0、约0.4至约1.0或约0.5至约1.0 μ m。引物的浓度可为约0.5 μ m。在PCR中对于目标核酸浓度适用的范围包括但不限于约1pg和约500ng之间。

[0240] 在一些情况下,水相还可包含添加剂,包括但不限于非特异性背景/封闭核酸(例如,鲑精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)。其他添加剂可包括,例如,二甲亚砜(DMSO)、甘油、甜菜碱(单)水合物(N,N,N-三甲基甘氨酸=[羧甲基]三甲基铵)、海藻糖、7-脱氮-2'-脱氧鸟苷三磷酸(dC7GTP或7-脱氮-2'-dGTP)、BSA(牛血清蛋白)、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-Cl)和四丙基氯化铵(TPrA-Cl)、非离子型去污剂(例如,Triton X-100、吐温20、Nonidet P-40(NP-40))或PREXCEL-Q。在一些情况下,水相可包含0、1、2、3、4、5、6、7、8、9或10种不同添加剂。在其他情况下,水相可包含至少0、1、2、3、4、5、6、7、8、9或10种不同添加剂。

[0241] 在一些情况下,非离子型环氧乙烷/环氧丙烷嵌段共聚物可以约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。常见的生物表面活性剂包括非离子型表面活性剂,诸如Pluronic F-68、Tetronics和Zonyl FSN。Pluronic F-68可以约0.5%w/v的浓度存在。

[0242] 在一些情况下,硫酸镁可以相似浓度替代氯化镁。来自多个供应商的广泛的常见、商用PCR缓冲液可替代缓冲溶液。

[0243] 可配制乳液以产生高度单分散的具有液体样界面膜的小滴,该小滴可通过加热转化为具有固体样界面膜的微胶囊;这样的微胶囊可表现为能够保持它们的内容物经过反应过程如PCR扩增的生物反应器。转化为微胶囊形式一经加热即可发生。例如,这样的转化可在大于约50℃、60℃、70℃、80℃、90℃或95℃的温度下发生。在一些情况下,使用热循环仪发生该加热。在加热过程中,可使用流体或矿物油覆盖物防止蒸发。在加热之前可去除或可不去除过量的连续相油。生物相容的胶囊可在广泛的热和机械处理中抵抗聚结和/或絮凝。在转化后,胶囊可储存于约、大于约或小于约3℃、4℃、5℃、6℃、7℃、8℃、9℃、10℃、15℃、20℃、25℃、30℃、35℃或40℃下。这些胶囊可在生物医学应用中有用,诸如稳定、数字化封装大分子,特别是含有核酸或蛋白质或二者的混合物的水性生物流体;药物和疫苗递送;生物分子文库;临床成像应用和其他。

[0244] 微胶囊可含有一种或多种多核苷酸,并可抵抗聚结,特别是在高温下。因此,PCR扩增反应可以非常高的密度发生(例如,每单位体积的反应数目)。在一些情况下,每ml可发生大于100,000、500,000、1,000,000、1,500,000、2,000,000、2,500,000、5,000,000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,在反应体积之间没有相互混合。微胶囊还可含有使逆转录、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。这些胶囊在广泛的热和机械处理中展现出对聚结和絮凝的抗性。

[0245] 在一些情况下,通过进行数字PCR,诸如基于微流体的数字PCR或小滴数字PCR,进行扩增步骤。

[0246] 可使用微流体系统或装置生成小滴。如本文所用的,“微-”前缀(例如,“微通道”或“微流体”)通常是指具有小于约1mm,且在一些情况下小于约100微米(micron/micrometer)的宽度或直径的元件或物品。在一些情况下,该元件或物品包括流体可流动穿过的通道。另外,如本文所用的“微流体”是指包含至少一个微型通道的设备、装置或系统。

[0247] 微流体系统和装置已经在多个情境下描述,典型地是在微型化实验室(例如,临床)分析的情境下。其他用途也已经描述。例如,国际专利申请公开号WO 01/89788;WO 2006/040551;WO 2006/04055;WO 2004/002627;WO 2008/063227;WO 2004/091763;WO 2005/021151;WO 2006/096571;WO 2007/089541;WO 2007/081385和WO 2008/063227。

[0248] 小滴通常在第二载流中包含一定量的第一样品流。本发明的方法可使用本领域已知的用于形成小滴的任何技术。示例性的方法包括使含有目标材料(例如,免疫细胞)的样品流体流动,以便其与流动载流的两股相对流相交。载流与样品流不混溶。样品流与流动载流的两股相对流的相交导致将样品流体分隔为含有目标材料的单独样品小滴。

[0249] 载流可为与样品流体不混溶的任何流体。示例性的载流为油。在某些实施方案中,载流包含表面活性剂。

[0250] 相同的方法可用于创建含有其他试剂的单个小滴,该其他试剂例如是用于扩增反应如聚合酶链反应(PCR)、或非基于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扩增反应的合适的试剂为本领域普通技术人员已知的,且包括但不限于DNA聚合酶、正向和反向引物、脱氧核苷三磷酸(dNTP)和一种或多种缓冲液。

[0251] 在某些实施方案中,通过提供包含目标材料(例如,免疫细胞和/或固体支持体,如珠子)的第一流体分区(例如,小滴)和第二流体(例如,作为液流或位于小滴内)形成流体隔室。合并第一和第二流体以形成小滴。可通过将电场施加于两种流体来完成合并。在某些实施方案中,第二流体含有用于进行诸如聚合酶链反应或扩增反应的扩增反应的试剂。

[0252] 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 α 和 β 链TCR序列和/或 γ 和 δ 链TCR序列的文库的方法,包括获得多个核酸构建体,其中每个构建体包含独特的N-mer和功能性N-mer。功能性N-mer可为随机N-mer、PCR引物、通用引物、抗体、粘性末端或任何其他序列。该方法可包括制备M组N个流体隔室,每个隔室含有独特构建体的一个或多个拷贝。该方法可通过向组中的每个隔室添加另外的构建体来创建具有更高复杂性的条码文库,并对每个组重复此步骤,以产生 $N \times M$ 个隔室,每个含有独特的一对构建体。可杂交或连接该对,以产生新构建体。在条码文库中的每个构建体中,每种独特的N-mer可适用于通过测序、探针杂交、其他方法或方法的组合来鉴别。

[0253] 小滴文库

[0254] 通常,小滴文库由在单个集合中合并在一起的许多文库元件组成。文库的复杂性可从单个文库元件到 1×10^{15} 个文库元件或更多不等。每个文库元件是固定浓度的一种或多种给定的组分。该元件可为但不限于细胞、珠子、氨基酸、蛋白质、多肽、核酸、多核苷酸或小分子化合物。该元件可含有标识符,诸如分子条码或容器条码或二者。

[0255] 细胞文库元件可包括但不限于杂交瘤、B细胞、T细胞、原代细胞、培养的细胞系、癌症细胞、干细胞或任何其他细胞类型。通过将一千到上万个数目的细胞封装在单个小滴中来制备细胞文库元件。封装细胞的数目通常由来自细胞的数量密度和小滴体积的Poisson统计给出。然而,在一些情况下,该数目偏离Poisson统计,如Edd等人,“Controlled encapsulation of single-cells into monodisperse picolitre drops.”Lab Chip,8(8):1262-1264,2008中所述。细胞的离散性质允许用多个细胞变体诸如各自产生一种抗体或TCR的免疫细胞大量制备文库,所有都存在于单一的起始介质中,然后该介质分解为含有最多一个细胞的单个小滴胶囊。然后裂解单个小滴胶囊中的细胞,用分子条码和容器条码对来自裂解细胞的重链和轻链多核苷酸和/或 α 和 β 链多核苷酸和/或 γ 和 δ 链多核苷酸进行条码化,并扩增,然后组合或合并以形成由重链和轻链和/或 α 和 β 链和/或 γ 和 δ 链文库元件组成的文库。

[0256] 基于珠子的文库元件含有一个或多个珠子,并且还可含有其他试剂,诸如抗体、酶或其他蛋白质。在所有文库元件都含有不同类型的珠子但含有相同的周围介质的情况下,所有文库元件都可从单一起始流体制备,或具有多个起始流体。在从变体的集合大量制备细胞文库的情况下,将从多个起始流体制备文库元件。当以多个细胞开始时,期望每滴恰好具有一个细胞,只有少量小滴含有超过一个细胞。在一些情况下,可实现相对于Poisson统计的变化以提供增强的小滴装载,以便更多小滴恰好每个小滴具有一个细胞,并且空小滴

或含有超过一个小滴的小滴的例外极少。

[0257] 在一些实施方案中,当从多个容器条码化多核苷酸开始时,期望每滴恰好具有一个容器条码化多核苷酸,只有少量小滴含有超过一个容器条码化多核苷酸。在一些情况下,可实现相对于Poisson统计的变化以提供增强的小滴装载,以便更多小滴恰好每个小滴具有一个容器条码化多核苷酸,并且空小滴或含有超过一个容器条码化多核苷酸的小滴的例外极少。

[0258] 小滴文库的实例是具有不同内容物的小滴的集合,该内容物的范围包括珠子、细胞、小分子、DNA、引物、抗体和分子条码。小滴的直径大小范围从大致0.5微米到500微米,其相当于约1皮升到1纳升。然而,小滴可小到5微米且大到500微米。优选地,小滴的直径为小于100微米,约1微米至约100微米。最优选的直径大小为约20至40微米(10至100皮升)。对小滴文库检查的优选性质包括渗透压平衡、均匀大小和大小范围。

[0259] 由本发明提供的小滴文库中包含的小滴优选大小均匀。也就是说,当与相同文库中其他小滴的直径相比时,文库中任何小滴的直径的变化将小于5%、4%、3%、2%、1%或0.5%。文库中小滴的均匀大小对维持小滴的稳定性和完整性可能非常关键,并且可能还对文库中的小滴随后用于本文所述的各种生物和化学分析非常重要。

[0260] 本发明提供了包含不混溶流体内的多个水性小滴的小滴文库,其中每个小滴优选是基本上大小均匀的,且包含不同的文库元件。本发明提供了用于形成小滴文库的方法,其包括提供包含不同文库元件的单个水性流体,将每个文库元件封装进不混溶流体中的水性小滴内。

[0261] 在某些实施方案中,将不同类型的元件(例如,细胞或珠子)合并到相同介质中含有的单个来源中。在初始合并后,将元件封装在小滴中,以形成小滴的文库,其中具有不同类型珠子或细胞的每个小滴均为不同的文库元件。对初始溶液的稀释允许封装过程。在一些实施方案中,形成的小滴将含有单个元件或将不含任何物体,即为空的。在其他实施方案中,形成的小滴将含有文库元件的多个拷贝。封装的元件通常为某种类型的变体。在一个实例中,元件为血液样品的免疫细胞,且封装每个免疫细胞以对免疫细胞中核苷酸的抗体序列进行扩增并条码化。

[0262] 例如,在一种类型的乳液文库中,存在的文库元件具有不同的粒子,即不同介质中的细胞或条码化多核苷酸,并且在合并之前封装。在一个实例中,在不同介质中包含指定数目的文库元件,即 n 个数目的不同细胞或条码化多核苷酸。每个文库元件分开乳化并合并,在该点处, n 个合并的不同文库元件中的每一个组合并合并成单个集合体。产生的集合体含有多个油包水乳液小滴,每个含有不同类型的粒子。

[0263] 在一些实施方案中,形成的小滴将含有单个文库元件或将不含任何物体,即为空的。在其他实施方案中,形成的小滴将含有文库元件的多个拷贝。珠子的内容物遵循Poisson分布,其中存在离散概率分布,其表达了如果若干事件以已知的平均率发生,且独立于自上一次事件以来的时间,则这些事件在固定的时间段内发生的概率。用于创建文库的油和表面活性剂防止了小滴之间文库内容物的交换。

[0264] 逆转录

[0265] 在一些情况下,通过逆转录从RNA制备目标多核苷酸。在一些情况下,通过引物延伸,诸如使用聚合酶,从DNA制备目标多核苷酸。

[0266] 本文所述的方法可在耦合的逆转录-PCR(逆转录-PCR)中使用。例如,逆转录和PCR可在两个不同的步骤中进行。首先,可使用多核苷酸dT引物、序列特异性引物、通用引物或本文所述的任何引物合成样品mRNA的cDNA拷贝。

[0267] 可在单个密闭容器反应中进行逆转录和PCR。例如,可使用三种引物,一种用于逆转录,而两种用于PCR。用于逆转录的引物可与PCR扩增子的位置的3'侧的mRNA结合。虽然不是必要的,但逆转录引物可包含RNA残基或修饰的类似物,诸如2'-O-甲基RNA碱基,当其与mRNA杂交时将不形成RNA酶H的底物。

[0268] 进行逆转录反应的温度取决于使用的逆转录酶。在一些情况下,使用热稳定的逆转录酶,且逆转录反应在约37℃至约75℃下、在约37℃至约50℃下、在约37℃至约55℃下、在约37℃至约60℃下、在约55℃至约75℃下、在约55℃至约60℃下、在约37℃下或在约60℃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。

[0269] 本文所述的逆转录反应和PCR反应可在本领域已知的多种形式中进行,诸如在管、微量滴定板、微流体装置或优选地在小滴中进行。

[0270] 逆转录反应可在范围为5μL至100μL或10μL至20μL反应体积的体积内进行。在小滴中,反应体积可为1pL至100nL或10pL至1nL。在一些情况下,逆转录反应在具有约或小于1nL体积的小滴中进行。在一些情况下,PCR反应在具有1pL至100nL,优选10pL至1nL的反应体积范围的小滴中进行。在一些情况下,PCR反应在具有约或小于1nL体积的小滴中进行。在一些情况下,逆转录反应和PCR反应在相同的具有1pL至100nL或10pL至1nL的反应体积范围的相同小滴中进行。在一些情况下,逆转录反应和PCR反应在具有约或小于1nL的体积或约或小于1pL的体积的小滴中进行。在一些情况下,逆转录反应和PCR反应在不同的小滴中进行。在一些情况下,逆转录反应和PCR反应在多个小滴中进行,每个小滴具有1pL至100nL或10pL至1nL的反应体积范围。在一些情况下,逆转录反应和PCR反应在多个小滴中进行,每个小滴具有约或小于1nL的体积。

[0271] 在一些情况下,第一PCR反应在具有1pL至100nL、优选10pL至1nL的反应体积范围的第一小滴内进行,且第二PCR反应在具有1pL至100nL、优选10pL至1nL的反应体积范围的第二小滴内进行。在一些情况下,第一PCR反应在具有约或小于1nL的体积的第一小滴内进行,且第二PCR反应在具有约或小于1nL的体积的第二小滴内进行。

[0272] 在一些情况下,第一PCR反应和第二PCR反应在多个小滴中进行,每个小滴具有1pL至100nL或10pL至1nL的反应体积范围。在一些情况下,第一PCR反应和第二PCR反应在多个小滴中进行,每个小滴具有约或小于1nL的体积。

[0273] 可使用一种或多种逆转录引物将目标多核苷酸如RNA逆转录为cDNA。该一种或多种逆转录引物可包含与RNA的区域如恒定区(例如,重链或轻链恒定区,或mRNA的聚A尾)互补的区域。在一些实施方案中,逆转录引物可包括第一逆转录引物和第二逆转录引物,第一逆转录引物具有与第一RNA的恒定区互补的区域,第二逆转录引物具有与第二RNA的恒定区互补的区域。在一些实施方案中,逆转录引物可包括第一逆转录引物和一个或多个逆转录引物,第一逆转录引物具有与第一RNA的恒定区互补的区域,所述一个或多个逆转录引物分别具有与一个或多个RNA的恒定区互补的区域。

[0274] 在一些实施方案中,逆转录引物不包含条码。

[0275] 逆转录引物可进一步包含不与RNA的区域互补的区域。在一些实施方案中,不与RNA的区域互补的区域在与RNA互补的引物区域的5'侧。在一些实施方案中,不与RNA的区域互补的区域在与RNA互补的引物区域的3'侧。在一些实施方案中,不与RNA的区域互补的区域为5'突出端区域。在一些实施方案中,不与RNA的区域互补的区域包含用于扩增和/或测序反应的引发位点。通过使用本文所述的一种或多种引物,使用本领域已知的合适试剂逆转录RNA分子。

[0276] 在进行RNA分子的逆转录反应后,产生的cDNA分子可用分子条码和容器条码进行条码化,并通过一个或多个PCR反应扩增,诸如第一和/或第二PCR反应。第一和/或第二PCR反应可利用一对引物或多个引物对。第一和/或第二PCR反应可利用多个正向/反向引物和反向引物。第一和/或第二PCR反应可利用多个正向/反向引物和正向引物。多个正向/反向引物的第一和/或第二引物可为含有与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/反向引物的第一和/或第二引物可为含有与条码化cDNA分子互补的区域的正向/反向引物。

[0277] 在一些实施方案中,多个正向/反向引物包括一个或多个正向/反向引物,其中多个正向/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/反向引物包括包含与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,以及包含与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正向/反向引物。例如,多个正向/反向引物包括包含与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正向/反向引物,以及包含与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正向/反向引物。例如,多个正向/反向引物包括包含与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正向/反向引物,包含与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正向/反向引物,以及包含与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正向/反向引物,等等。多个正向/反向引物中的引物可用来与样品中的细胞(诸如免疫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。

[0278] 在一些实施方案中,多个正向/反向引物包括一个或多个正向/反向引物,其中多个正向/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/反向引物包括包含与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,以及包含与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正向/反向引物。例如,多个正向/反向引物包括包含与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正向/反向引物,以及包含与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正向/反向引物。例如,多个正向/反向引物包括包含与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正向/反向引物,包含与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正向/反向引物,以及包含与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正向/反向引物,等等。多个正向/反向引物中的引物可用来与样品中的细胞(诸如免疫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。

[0279] 在一些实施方案中,多个正向/反向引物包括一个或多个正向/反向引物,其中多个正向/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/反向引物包括包含与条码化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,以及包含与条码化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/反向引物包括包含与条码化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正向/反向引物,以及包含与条码化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/反向引物包括包含与条码化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正向/反向引物,包含与条码化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,以及包含与条码化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。多个正向/反向引物可用来与样品中的细胞(诸如免疫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。

[0280] 在一些实施方案中,多个正向/反向引物包括一个或多个正向/反向引物,其中多个正向/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/反向引物包括包含与条码化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,以及包含与条码化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/反向引物包括包含与条码化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正向/反向引物,以及包含与条码化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/反向引物包括包含与条码化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正向/反向引物,包含与条码化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,以及包含与条码化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。多个正向/反向引物中的引物可用来与样品中的细胞(诸如免疫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。

[0281] 多个正向/反向引物中的正向/反向引物进一步包含不与RNA的区域互补的区域。在一些实施方案中,不与RNA的区域互补的区域在与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'侧。在一些实施方案中,不与RNA的区域互补的区域在与RNA互补的正向/反向引物区域的3'侧。在一些实施方案中,不与RNA的区域互补的区域为5'突出端区域。在一些实施方案中,不与RNA的区域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,不与RNA的区域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,不与RNA的区域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。通过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,区域互补于RNA的区域,诸如恒定区或mRNA的聚A尾。

[0282] 扩增

[0283] 含有目标多核苷酸的样品可包含mRNA或其片段,其可进行扩增。在一些情况下,mRNA或其片段的平均长度可为小于约100、200、300、400、500或800个碱基对,或小于约5、

10、20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190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,从相对较短的模板,诸如含有约40、45、50、55、60、65、70、75、80、85、90、95或100个碱基的模板的样品扩增目标序列。

[0284] 扩增反应可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为二甲亚砜(DMSO)、甘油、甜菜碱(单)水合物(N,N,N-三甲基甘氨酸=[羧甲基]三甲基铵)、海藻糖、7-脱氮-2'-脱氧鸟苷三磷酸(dC7GTP或7-脱氮-2'-dGTP)、BSA(牛血清蛋白)、甲酰胺、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-Cl)和四丙基氯化铵(TPrA-Cl)、非离子型去污剂(例如,Triton X-100、吐温20、Nonidet P-40(NP-40))或PREXCEL-Q。在一些情况下,扩增反应包含0、1、2、3、4、5、6、7、8、9或10种不同添加剂。在其他情况下,扩增反应包含至少0、1、2、3、4、5、6、7、8、9或10种不同添加剂。

[0285] 可在包含于反应体积(例如,小滴)中的样品上进行热循环反应。小滴可为多分散的或优选单分散的,通过搅拌、超声处理或穿过T通道接合点以微流体方式或本领域熟知的其他手段生成。密度可超过20,000个小滴/40 μ l(1nL小滴)、200,000个小滴/40 μ l(100pL小滴)。在热循环期间,小滴可保持完整。在热循环期间,小滴可以大于约10,000个小滴/ μ L、100,000个小滴/ μ L、200,000个小滴/ μ L、300,000个小滴/ μ L、400,000个小滴/ μ L、500,000个小滴/ μ L、600,000个小滴/ μ L、700,000个小滴/ μ L、800,000个小滴/ μ L、900,000个小滴/ μ L或1,000,000个小滴/ μ L的密度保持完整。在其他情况下,两个或更多个小滴在热循环期间不聚结。在其他情况下,大于100个或大于1,000个小滴在热循环期间不聚结。

[0286] 可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I的Klenow片段、T7DNA聚合酶、T4 DNA聚合酶、Taq聚合酶、Pfu DNA聚合酶、Vent DNA聚合酶、噬菌体29、REDTaqTM、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。还可进行热启动PCR,其中在添加聚合酶之前将反应加热到95 $^{\circ}$ C保持两分钟,或聚合酶可保持无活性直到循环1中的第一加热步骤。热启动PCR可用于使非特异性扩增最小化。可以使用任何数目的PCR循环来扩增DNA,例如约、大于约或小于约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44或45个循环。扩增循环的数目可为约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。

[0287] 可通过本领域已知的任何手段进行目标核酸的扩增。可通过聚合酶链反应(PCR)或等温DNA扩增来扩增目标核酸。可使用的PCR技术的实例包括但不限于定量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、原位聚合酶克隆PCR、原位滚环扩增(RCA)、数字PCR(dPCR)、小滴数字PCR(ddPCR)、桥式PCR、皮升PCR和乳液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、自动维持序列复制、目标多核苷酸序列的选择性扩增、共有序列引物聚合酶链反应(CP-PCR)、任意引物聚合酶链反应(AP-PCR)、简并多核苷酸引物PCR(DOP-PCR)和基于核酸的序列扩增(NABSA)。本文可使用的其他扩增方法包括美国专利号5,242,794、5,494,810、4,988,617和6,582,938中描述的方法,以及包括Q β 复制酶介导的RNA扩增。扩增可为恒温扩增,例如,恒温线性扩增。

[0288] 在一些实施方案中,扩增不在固体支持体上发生。在一些实施方案中,扩增不在小滴中的固体支持体上发生。在一些实施方案中,当不在小滴中扩增时,扩增在固体支持体上发生。

[0289] 扩增反应可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为二甲亚砜(DMSO)、甘油、甜菜碱(单)水合物(N,N,N-三甲基甘氨酸=[羧甲基]三甲基铵)、海藻糖、7-脱氮-2'-脱氧鸟苷三磷酸(dC7GTP或7-脱氮-2'-dGTP)、BSA(牛血清蛋白)、甲酰胺、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-Cl)和四丙基氯化铵(TPrA-Cl)、非离子型去污剂(例如,Triton X-100、吐温20、Nonidet P-40(NP-40))或PREXCEL-Q。在一些实施方案中,扩增反应可包含0、1、2、3、4、5、6、7、8、9或10种不同添加剂。在其他情况下,扩增反应可包含至少0、1、2、3、4、5、6、7、8、9或10种不同添加剂。

[0290] 引物

[0291] 通常,在扩增反应中可使用一对或多对引物;引物对的一个引物可为正向引物且引物对的一个引物可为反向引物。

[0292] 在一些情况下,在扩增反应中可使用第一对引物;第一对的一个引物可为与第一目标多核苷酸分子的序列互补的正向引物,且第一对的一个引物可为可与第一目标多核苷酸分子的第二序列互补的反向引物,且第一目标基因座可位于第一序列与第二序列之间。在一些实施方案中,第一目标基因座包含 V_H 或 V_α 或 V_γ 序列。

[0293] 在一些情况下,在扩增反应中可使用第二对引物;第二对的一个引物可为与第二目标多核苷酸分子的第一序列互补的正向引物,且第二对的一个引物可为与第二目标多核苷酸分子的第二序列互补的反向引物,且第二目标基因座可位于第一序列与第二序列之间。在一些实施方案中,第二目标基因座包含 V_L 或 V_β 或 V_δ 序列。

[0294] 在一些情况下,在扩增反应中可使用第三对引物;第三对的一个引物可为与第三目标多核苷酸分子的第一序列互补的正向引物,且第三对的一个引物可为与第三目标多核苷酸分子的第二序列互补的反向引物,且第三目标基因座可位于第一序列与第二序列之间。在一些实施方案中,第三目标基因座包含条码,诸如分子条码或容器条码。

[0295] 正向引物和反向引物的长度可取决于目标多核苷酸的序列和目标基因座。例如,可优化正向引物和反向引物的长度和/或 T_m 。在一些情况下,引物的长度可为约、大于约或小于约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59或60个核苷酸。在一些情况下,引物的长度为约15至约20、约15至约25、约15至约30、约15至约40、约15至约45、约15至约50、约15至约55、约15至约60、约20至约25、约20至约30、约20至约35、约20至约40、约20至约45、约20至约50、约20至约55或约20至约60个核苷酸。

[0296] 在与模板多核苷酸结合之前,引物可为单链DNA。在一些情况下,引物最初含有双链序列。引物的适当长度可取决于引物的预期用途,但范围可为约6至约50个核苷酸,或约15至约35个核苷酸。短引物分子通常可需要更冷的温度以与模板形成足够稳定的杂合复合体。在一些实施方案中,引物不需要反映模板核酸的准确序列,但可足够互补以与模板杂交。在一些情况下,在与模板多核苷酸结合前,引物可为部分双链的。具有双链序列的引物可具有约、大于约或小于约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基的发夹环。引物的双链部分可为约、大于约或小于约3、4、5、6、7、8、9、10、11、12、13、14、15、

16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个碱基对。用于扩增给定目标序列的合适引物的设计是本领域公知的。

[0297] 引物可包含附加特征,该特征允许检测或固定引物,但不改变引物的基本性质(例如,作为DNA合成的起始点)。例如,引物可在5'端含有附加核酸序列,该序列不与目标核酸杂交,但其促进克隆或进一步的扩增,或对扩增产物的测序。例如,该附加序列可包含引物结合位点,诸如通用引物结合位点。本文可将与模板足够互补以杂交的引物区域称作杂交区域。

[0298] 在另一种情况下,本文所述的方法和组合物中所采用的引物可包含一个或多个通用核苷。通用核苷的非限制性实例为5-硝基吡啶和肌苷,如美国申请公开号2009/0325169和2010/0167353中所述。

[0299] 可根据已知参数设计引物,以便避免二级结构和自身杂交。不同的引物对可在大约相同的温度下退火和解链,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内。在一些情况下,最初使用多于1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、100、200、500、1000、5000、10,000个或更多个引物。这样的引物可与本文所述的目标多核苷酸杂交。

[0300] 可通过多种方法制备引物,包括但不限于使用本领域公知的方法克隆适当的序列和直接化学合成(Narang等人,Methods Enzymol.68:90(1979);Brown等人,Methods Enzymol.68:109(1979))。还可从商业来源获得引物。引物可具有相同的解链温度。引物可具有不相同的解链温度。引物的长度可在5'端或3'端延伸或缩短,以产生具有所需解链温度的引物。引物的一个引物可比另一个引物长。引物对中引物的3'退火长度可不同。同样,可设计引物对的退火位置,以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等式为Wallace规则($T_M = 2(A+T) + 4(G+C)$)。还可使用计算机程序设计引物。可使用软件程序计算每个引物的 T_M (解链或退火温度)。在任何扩增循环(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)后,引物的退火温度可重新计算并提高。在扩增的初始循环后,引物的5'一半可从每个感兴趣的基因座并入产物中;因此可基于每个引物的5'一半和3'一半两者的序列重新计算 T_M 。

[0301] 进行本文公开的方法的一个或多个反应可包括使用一个或多个引物。如本文所用的,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。在与模板多核苷酸结合之前,引物可为单链DNA。在一些实施方案中,引物最初含有双链序列。引物位点包括引物杂交的模板区域。在一些实施方案中,引物能够作为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试剂或酶如DNA或RNA聚合酶或逆转录酶时,引物可启动模板指导的核酸合成。引物对包含2个引物:具有与模板序列的5'端杂交的5'上游区的第一引物,以及具有与模板序列的3'端的互补体杂交的3'下游区的第二引物。引物组包含两个或更多个引物:具有与模板序列或多个模板序列的5'端杂交的5'上游区的第一引物或第一多个引物,以及具有与模板序列或多个模板序列的3'端的互补体杂交的3'下游区的第二引物或第二多个引物。在一些实施方案中,引物包含目标特异性序列。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, 第二版 (Cold Spring Harbor Press, New York (2003))。通用引物结合位点或序列允许通用引物与多核苷酸和/或扩增子附接。通用引物是本领域公知的,并且包括但不限于-47F (M13F)、alfaMF、A0X3'、A0X5'、BGHr、CMV-30、CMV-50、CVMf、LACrmt、lamgda gt10F、lambda gt 10R、lambda gt11F、lambda gt11R、M13rev、M13Forward (-20)、M13Reverse、male、p10SEQPpQE、pA-120、pet4、pGAP Forward、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。如本文所用的,附接可指共价相互作用和非共价相互作用两者或之一。通用引物与通用引物结合位点的附接可用于多核苷酸和/或扩增子的扩增、检测和/或测序。通用引物结合位点可包含至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000个核苷酸或碱基对。在另一个实例中,通用引物结合位点包含至少约1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500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、10-20、10-30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约1-90、1-80、1-70、1-60、1-50、1-40、1-30、1-20、1-10、2-90、2-80、2-70、2-60、2-50、2-40、2-30、2-20、2-10、1-900、1-800、1-700、1-600、1-500、1-400、1-300、1-200、1-100、2-900、2-800、2-700、2-600、2-500、2-400、2-300、2-200、2-100、5-90、5-80、5-70、5-60、5-50、5-40、5-30、5-20、5-10、10-90、10-80、10-70、10-60、10-50、10-40、10-30、10-20、10-10、5-900、5-800、5-700、5-600、5-500、5-400、5-300、5-200、5-100、10-900、10-800、10-700、10-600、10-500、10-400、10-300、10-200、10-100、25-900、25-800、25-700、25-600、25-500、25-400、25-300、25-200、25-100、100-1000、100-900、100-800、100-700、100-600、100-500、100-400、100-300、100-200、200-1000、200-900、200-800、200-700、200-600、200-500、200-400、200-300、300-1000、300-900、300-800、300-700、300-600、300-500、300-400、400-1000、400-900、400-800、400-700、400-600、400-500、500-1000、500-900、500-800、500-700、500-600、600-1000、600-900、600-800、600-700、700-1000、700-900、700-800、800-1000、800-900或900-1000个核苷酸或碱基对。

[0302] 引物可具有与其在合成引物延伸产物中的应用相匹配的长度。引物可为长度为8-200个核苷酸的多核苷酸。引物的长度可取决于模板多核苷酸的序列和模板基因座。例如,可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,引物的长度可为约、大于约或小于约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59或60个核苷酸。在一些实施方案中,引物的长度为约8-100个核苷酸,例如10-75、15-60、15-40、18-30、20-40、21-50、22-45、25-40、7-9、12-15、15-20、15-25、15-30、15-45、15-50、15-55、15-60、20-25、20-30、20-35、20-45、20-50、20-55或20-60个核苷酸的长度和之间的任何长度。在一些实施方案中,引物的长度为最多约10、12、15、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、55、60、65、70、75、80、85、90、95或100个核苷酸。

[0303] 通常,在指数扩增反应中可使用一对或多对引物;引物对的一个引物可为正向引

物且引物对的一个引物可为反向引物。在一些实施方案中,在指数扩增反应中可使用第一对引物;第一对的一个引物可为与第一模板多核苷酸分子的序列互补的正向引物,且第一对的一个引物可为与第一模板多核苷酸分子的第二序列互补的反向引物,且第一模板基因座可位于第一序列与第二序列之间。在一些实施方案中,在扩增反应中可使用第二对引物;第二对的一个引物可为与第二目标多核苷酸分子的第一序列互补的正向引物,且第二对的一个引物可为与第二目标多核苷酸分子的第二序列互补的反向引物,且第二目标基因座可位于第一序列与第二序列之间。在一些实施方案中,第二目标基因座包含可变轻链抗体序列。在一些实施方案中,在扩增反应中可使用第三对引物;第三对的一个引物可为与第三模板多核苷酸分子的第一序列互补的正向引物,且第三对的一个引物可为与第三模板多核苷酸分子的第二序列互补的反向引物,且第三模板基因座可位于第一序列与第二序列之间。

[0304] 所述一个或多个引物可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与所述多个模板多核苷酸的3'端和/或5'端退火。所述一个或多个引物可与所述多个模板多核苷酸的内部区域退火。该内部区域可以距所述多个模板多核苷酸的3'端或5'端至少约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、100、150、200、220、230、240、250、260、270、280、290、300、310、320、330、340、350、360、370、380、390、400、410、420、430、440、450、460、470、480、490、500、510、520、530、540、550、560、570、580、590、600、650、700、750、800、850、900或1000个核苷酸。所述一个或多个引物可包含固定的引物组。所述一个或多个引物可包含至少一个或多个定制引物。所述一个或多个引物可包含至少一个或多个控制引物。所述一个或多个引物可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包含通用引物。该通用引物可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、目标特异性区域、其互补体或其任意组合退火。所述一个或多个引物可包含通用引物。可设计所述一个或多个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或一个或多个目标或模板多核苷酸的任何其他扩增方法。

[0305] 目标特异性区域可包含至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、100、150、200、220、230、240、250、260、270、280、290、300、310、320、330、340、350、360、370、380、390、400、410、420、430、440、450、460、470、480、490、500、510、520、530、540、550、560、570、580、590、600、650、700、750、800、850、900或1000个核苷酸或碱基对。在另一个实例中,目标特异性区域包含至少约1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500或10000个核苷酸或碱基对。在一些实施方案中,目标特异性区域包含约5-10、10-15、10-20、10-30、15-30、10-75、15-60、15-40、18-30、20-40、21-50、22-45、25-40、7-9、12-15、15-20、15-25、15-30、15-45、15-50、15-55、15-60、20-25、20-30、20-35、20-45、20-50、20-55、20-60、2-900、2-800、2-700、2-600、2-500、2-400、2-300、2-200、2-100、25-900、25-800、25-700、25-600、25-500、25-400、25-300、25-200、25-100、100-1000、100-900、100-800、100-700、100-600、100-500、100-400、100-300、100-200、200-1000、200-900、200-800、200-700、200-600、200-500、200-400、200-300、300-1000、300-900、300-800、300-700、300-600、300-500、300-400、400-1000、400-900、400-800、400-700、400-600、400-500、500-1000、500-900、500-800、500-

700、500-600、600-1000、600-900、600-800、600-700、700-1000、700-900、700-800、800-1000、800-900或900-1000个核苷酸或碱基对。

[0306] 可根据已知参数设计引物,以便避免二级结构和自身杂交。在一些实施方案中,不同的引物对可在大约相同的温度下退火和解链,例如,在另一个引物对的1℃、2℃、3℃、4℃、5℃、6℃、7℃、8℃、9℃或10℃内。在一些实施方案中,多个引物中的一个或多个引物可在大约相同的温度下退火和解链,例如,在多个引物中另一个引物的1℃、2℃、3℃、4℃、5℃、6℃、7℃、8℃、9℃或10℃内。在一些实施方案中,多个引物中的一个或多个引物可在与前述多个引物中另一个引物不同的温度下退火和解链。

[0307] 用于本文所述方法的一个或多个步骤的多个引物可包括多个引物,其包含约、最多约或至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、50,000,000、100,000,000个不同引物。例如,多个引物中的每个引物可包含不同的目标或模板特异性区域或序列。

[0308] 测序

[0309] 在进行本文所述的一个或多个方法或方法步骤后,可对生成的多核苷酸文库进行测序。

[0310] 可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,可进行高通量测序。合适的下一代测序技术包括454 Life Sciences平台(Roche,Branford,CT)(Margulies等人,Nature,437,376-380(2005));Illumina的基因组分析仪、GoldenGate甲基化分析或Infinium甲基化分析,即Infinium人类甲基化27K BeadArray或VeraCode GoldenGate甲基化分析(Illumina,San Diego,CA;Bibkova等人,Genome Res.16,383-393(2006);和美国专利号6,306,597、7,598,035、7,232,656),或DNA连接测序,SOLiD系统(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单分子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中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)和McCaughan等人,J.Pathol.,220,297-306(2010))。在一些实施方案中,通过染料修饰的探针的连接测序、焦磷酸测序或单分子测序对多核苷酸进行测序。对多核苷酸序列的测定可通过如下测序方法进行,诸如Helioscope™单分子测序、纳米孔DNA测序、Lynx Therapeutics大规模平行签名测序(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™测序、聚合酶克隆测序、DNA纳米球测序和VisiGen Biotechnologies方法。或者,测定多核苷酸的序列可采用测序平台,包括但不限于基因组分析仪Illumina、HiSeq,以及Illumina提供的MiSeq,单分子实时(SMRT™)技术,诸如Pacific Biosciences(California)提供的PacBio RS系统以及Solexa测序仪、真

正单分子测序(tSMSTM)技术,诸如Helicos Inc. (Cambridge, MA)提供的HeliScopeTM测序仪。测序可包括MiSeq测序。测序可包括HiSeq测序。在一些实施方案中,测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、鸟枪法测序、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意组合。或者,可通过电子显微术或化学敏感性场效应晶体管(chemFET)阵列测定多核苷酸的序列。

[0311] 方法可进一步包括对文库中的一个或多个多核苷酸进行测序。方法可进一步包括将文库中的一个或多个多核苷酸序列、序列读取、扩增子序列或扩增子集序列彼此进行比对。

[0312] 如本文所用的,比对包括将测试序列如序列读取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,比对可用于从多个序列或比对的序列确定共有序列。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,为了比较目的而比对的序列的长度为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%。可通过公知的方法,例如,使用数学算法完成对两个或更多个序列的实际比较。这样的数学算法的非限制性实例描述于Karlin, S. 和Altschul, S., Proc. Natl. Acad. Sci. USA, 90:5873-5877 (1993) 中。这样的算法并入NBLAST和XBLAST程序(版本2.0)中,如Altschul, S. 等人, Nucleic Acids Res., 25:3389-3402 (1997) 中所述。当利用BLAST和Gapped BLAST程序时,可使用相应程序(例如, NBLAST)的任何相关参数。例如,用于序列比较的参数可设置为分数=100, 字长=12, 或可变化(例如, W=5或W=20)。其他实例包括Myers和Miller, CABIOS (1989) 的算法、ADVANCE、ADAM、BLAT和FASTA。在一些实施方案中,可使用例如GCG软件包中的GAP程序(Accelrys, Cambridge, UK) 完成两个氨基酸序列之间的百分比同一性。

[0313] 测序可包括对多核苷酸的至少约10、20、30、40、50、60、70、80、90、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、300、400、500、600、700、800、900、1000个或更多个核苷酸或碱基对进行测序。在其他情况下,测序包括对多核苷酸的至少约1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000个或更多个核苷酸或碱基对进行测序。

[0314] 测序可包括每次运行至少约200、300、400、500、600、700、800、900、1000个或更多个测序读取。如本文所用的,序列读取包括从通过测序技术生成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、2000、3000、4000、5000、6000、7000、8000、9000、10,000个或更多个测序读取进行测序。测序可包括每次运行多于、少于或等于约1,000,000,000个测序读取。测序可包括每次运行多于、少于或等于约200,000,000个读取。

[0315] 在一些实施方案中,用于确定共有序列的序列读取的数目为约2-1000个序列读取。例如,用于确定共有序列的序列读取的数目可为约2-900、2-800、2-700、2-600、2-500、2-400、2-300、2-200、2-100、25-900、25-800、25-700、25-600、25-500、25-400、25-300、25-200、25-100、100-1000、100-900、100-800、100-700、100-600、100-500、100-400、100-300、100-200、200-1000、200-900、200-800、200-700、200-600、200-500、200-400、200-300、300-1000、300-900、300-800、300-700、300-600、300-500、300-400、400-1000、400-900、400-

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

[0316] 方法可包括对错读进行测序。方法可包括确定错读的数目,诸如用于确定反应条件或设计引物序列。比较在一种或多种第一条件或条件组下生成的错读的数目可用于确定优选的条件或条件组。例如,可在PCR反应期间在高盐浓度下进行第一方法,并且可在PCR反应期间在低盐浓度下进行第二方法,其中除了盐浓度差异之外,第一和第二方法基本相同地进行。如果第一方法导致更高数目的错读,诸如对特定目标多核苷酸序列或引物的更高数目的错读,则可确定对于该特定目标多核苷酸序列或引物优选更低的盐反应条件。

[0317] 诊断

[0318] 在一些实施方案中,方法可进一步包括对受试者疾病、病症、症状和/或病况的诊断、预后、监测、治疗、减轻和/或预防。在一些实施方案中,方法可进一步包括基于目标多核苷酸的存在、不存在或水平对受试者疾病、病症、症状和/或病况的诊断、预后、监测、治疗、减轻和/或预防。在一些实施方案中,方法可进一步包括基于一个或多个目标多核苷酸的存在、不存在或水平对受试者疾病、病症、症状和/或病况的诊断、预测、监测、治疗、减轻和/或预防。

[0319] 在一些实施方案中,方法可进一步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、水平或序列对受试者疾病、病症、症状和/或病况的诊断、预后、监测、治疗、减轻和/或预防。例如,可基于使用本文所述的方法获得的变异序列的存在、不存在、水平或序列进行疾病的诊断。在一些实施方案中,方法可进一步包括基于使用本文所述的方法获得的一个或多个序列读取的存在、不存在、水平或序列对受试者疾病、病症、症状和/或病况的诊断、预后、监测、治疗、减轻和/或预防。在一些实施方案中,方法可进一步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、水平或序列对受试者疾病、病症、症状和/或病况的诊断、预后、监测、治疗、减轻和/或预防。在一些实施方案中,方法可进一步包括基于样品中目标多核苷酸的水平(例如,量或浓度)的确定对受试者疾病、病症、症状和/或病况的诊断、预后、监测、治疗、减轻和/或预防。可基于一个或多个序列读取、序列、共有序列或其任意组合来确定样品中目标多核苷酸的水平。可使用本文所述

的方法确定样品中多个目标多核苷酸中的每一个的水平。可基于多个目标多核苷酸中每个目标多核苷酸的多个序列读取、序列、共有序列或其任意组合来确定样品中多个目标多核苷酸中的每一个的水平。例如,可使用本文所述的方法确定第一目标多核苷酸的水平 and 第二目标多核苷酸的水平。

[0320] 在一些实施方案中,多个目标多核苷酸的第一和第二目标多核苷酸是相同的。例如,第一目标多核苷酸可包含mRNA分子的第一拷贝,而第二目标多核苷酸可包含mRNA分子的第二拷贝。在一些实施方案中,第一和第二目标多核苷酸是不同的。例如,第一目标多核苷酸可包含第一mRNA分子,而第二目标多核苷酸可包含从与第一mRNA分子不同的基因转录的第二mRNA分子。例如,第一目标多核苷酸可包含第一等位基因,而第二目标多核苷酸可包含第二等位基因。例如,第一目标多核苷酸可包含野生型序列,而第二目标多核苷酸可包含变异序列。

[0321] 在一些实施方案中,方法可进一步包括以至少50%的置信度将受试者诊断或预后为患有疾病、病症、症状和/或病况。例如,可以以至少50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%的置信度确定对受试者患有疾病、病症、症状和/或病况的诊断或预后。在一些实施方案中,可以以50%-100%的置信度确定对受试者患有疾病、病症、症状和/或病况的诊断或预后。例如,可以以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%的置信度确定对受试者患有疾病、病症、症状和/或病况的诊断或预后。

[0322] 在一些实施方案中,可以以至少50%的置信度确定受试者的目标多核苷酸(诸如生物标志物)的存在、不存在、水平、序列或其任意组合。例如,可以以至少50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%的置信度确定受试者的目标多核苷酸的存在、不存在、水平、序列或其任意组合。在一些实施方案中,可以以50%-100%的置信度确定受试者的目标多核苷酸的存在、不存在、水平、序列或其任意组合。例如,可以以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的目标多核苷酸的存在、不存在、水平、序列或其任意组合。

[0323] 酶

[0324] 本文所公开的方法和试剂盒可包含一种或多种酶。酶的实例包括但不限于连接酶、逆转录酶、聚合酶和限制酶。

[0325] 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。连接酶的实例包括但不限于DNA连接酶,诸如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,诸如T4RNA连接酶I和T4RNA连接酶II。

[0326] 本文所公开的方法和试剂盒可进一步包括使用一种或多种逆转录酶。在一些实施方案中,该逆转录酶为HIV-1逆转录酶、M-MLV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,该逆转录酶为M-MLV逆转录酶。

[0327] 在一些实施方案中,本文所公开的方法和试剂盒包括使用一种或多种蛋白酶。

[0328] 在一些实施方案中,本文所公开的方法和试剂盒包括使用一种或多种聚合酶。聚合酶的实例包括但不限于DNA聚合酶和RNA聚合酶。在一些实施方案中,该DNA聚合酶为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、Bst 2.0WarmStart™DNA聚合酶、Bst DNA聚合酶、硫化叶菌DNA聚合酶IV、Taq DNA聚合酶、9°N™m DNA聚合酶、Deep VentR™(exo-) DNA聚合酶、Deep VentR™DNA聚合酶、Hemo KlenTaq™、**LongAmp®** Taq DNA聚合酶、**OneTaq®** DNA聚合酶、**Phusion®** DNA聚合酶、Q5™高保真DNA聚合酶、Therminator™γ DNA聚合酶、Therminator™DNA聚合酶、Therminator™II DNA聚合酶、Therminator™III DNA聚合酶、**VentR®** DNA聚合酶、**VentR®** (exo-) 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®** Taq聚合酶、KAPA Taq DNA聚合酶和KAPA Taq热启动DNA聚合酶。

[0329] 在一些实施方案中,所述聚合酶为RNA聚合酶,诸如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T7 RNA聚合酶。

[0330] 附加试剂

[0331] 本文所公开的方法和试剂盒可包括使用一种或多种试剂。试剂的实例包括但不限于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、固体支持体如珠子,以及用于核酸纯化和/或分离的试剂。

[0332] 固体支持体可包含实际上任何不溶性或固体材料,并且通常选择不溶于水的固体支持体材料组合物。例如,固体支持体可包含以下成分或基本由以下成分组成:硅胶、玻璃(例如,可控孔玻璃(CPG))、尼龙、**Sephadex®**、**Sepharose®**、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、聚偏氟乙烯(PVDF))等。根据实施方案使用的珠子的实例可包含允许珠子与核酸分子相互作用的亲和部分。固相(例如,珠子)可包含结合对的成员(例如,抗生物素蛋白、链霉亲和素或其衍生物)。例如,该珠子可为链霉亲和素包被的珠子,并且用于固定在珠子上的核酸分子可包含生物素部分。在一些情况下,每个多核苷酸分子可包含两个亲和部分,如生物素,来进一步稳定多核苷酸。珠子可包含用于固定核酸的附加特征,或可用于下游筛选或选择过程的附加特征。例如,珠子可包含结合部分、荧光标记或荧光猝灭剂。在一些情况下,珠子可为磁性的。在一些情况下,固体支持体为珠子。珠子的实例包括但不限于链霉亲和素珠子、琼脂糖珠子、磁珠、**Dynabeads®**、**MACS®**微珠、抗体偶联的珠子(例如,抗免疫球蛋白微珠)、蛋白A偶联的珠子、蛋白G偶联珠子、蛋白A/G偶联的珠子、蛋白L偶联的珠子、多核苷酸dT偶联的珠子、二氧化硅珠子、二氧化硅样珠子、抗生物素微珠、抗荧光染料微珠和BcMag™羧基末端磁珠。珠子或粒子可为可膨胀的(例如,聚合物珠子,诸如Wang树脂)或不可膨胀的(例如,CPG)。在一些实施方案中,固相基本上为亲水的。在一些实施方案中,固相(例如,珠子)基本上为疏水的。在一些实施方案中,固相包含结合对的成员(例如,抗生物素蛋白、链霉亲和素或其衍生物),并且基本上为疏水的或基本上为亲水的。在一些实施方案中,固相包含结合对的成员(例如,抗生物素蛋白、链霉亲和素或其衍生物),并且具有每mg固体支持体大于约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固体支持体大于800、900、1000、1100、1200、

1250、1300、1350、1400、1450、1500、1600、1800、2000皮摩尔的游离捕获剂。适用于本发明的珠子的其他实例为金胶体或珠子,诸如聚苯乙烯珠或二氧化硅珠。可使用基本上任何珠半径。珠子的实例可包括半径范围为150纳米至10微米的珠子。还可使用其他大小。

[0333] 本文所公开的方法和试剂盒可包括使用一种或多种缓冲液。缓冲液的实例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,杂交缓冲液为可商购获得的缓冲液,诸如TMAC Hyb溶液、SSPE杂交溶液和ECONO™杂交缓冲液。本文所公开的缓冲液可包含一种或多种去污剂。

[0334] 本文所公开的方法和试剂盒可包括使用一种或多种载体。载体可增强或改善本文所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。载体可降低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,载体可通过吸附至表面而降低多核苷酸的非特异性损失。载体可降低多核苷酸对表面或基底(例如,器皿、Eppendorf管、移液管尖端)的亲和力。或者,载体可提高多核苷酸对表面或基底(例如,珠子、阵列、玻璃、载片、芯片)的亲和力。载体可保护多核苷酸免于降解。例如,载体可保护RNA分子不受核糖核酸酶作用。或者,载体可保护DNA分子不受DNA酶作用。载体的实例包括但不限于多核苷酸如DNA和/或RNA,或多肽。DNA载体的实例包括质粒、载体、聚腺苷酸化DNA和DNA多核苷酸。RNA载体的实例包括聚腺苷酸化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、短聚腺苷酸合成核糖核苷酸和RNA多核苷酸。RNA载体可为聚腺苷酸化RNA。或者,RNA载体可为非聚腺苷酸化RNA。在一些实施方案中,载体来自细菌、酵母或病毒。例如,载体可为来源于细菌、酵母或病毒的多核苷酸或多肽。例如,载体为来自枯草芽孢杆菌(*Bacillus subtilis*)的蛋白质。在另一个实例中,载体为来自大肠杆菌(*Escherichia coli*)的多核苷酸。或者,载体为来自哺乳动物(例如,人类、小鼠、山羊、大鼠、牛、绵羊、猪、狗或兔)、禽、两栖动物或爬行动物的多核苷酸或肽。

[0335] 本文所公开的方法和试剂盒可包括使用一种或多种对照剂。对照剂可包括对照多核苷酸、非活性酶、非特异性竞争剂。或者,对照剂包括亮杂交、亮探针对照、核酸模板、掺加对照、PCR扩增对照。PCR扩增对照可为阳性对照。在其他情况下,PCR扩增控制物为阴性对照。核酸模板对照可为已知浓度。对照剂可包含一种或多种标记。

[0336] 掺加对照可为向反应或样品中添加的模板。例如,掺加模板可添加到扩增反应中。掺加模板可在第一扩增循环后的任何时间添加到扩增反应中。在一些实施方案中,掺加模板在循环数2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45或50后添加到扩增反应中。掺加模板可在最终扩增循环前的任何时间添加到扩增反应中。掺加模板可包含一种或多种核苷酸或核酸碱基对。掺加模板可包含DNA、RNA或其任意组合。掺加模板可包含一种或多种标记。

[0337] 本文公开了可用于、可结合使用、可用于准备本文所公开的方法和组合物或作为其产物的分子、材料、组合物和组分。应当理解,当公开这些材料的组合、子集、相互作用、组等时,尽管这些分子和化合物的每个不同单独和共同组合和排列的具体提及不能明确公开时,但每种提及都在此具体涉及并描述。例如,如果公开并讨论了核苷酸或核酸,并且讨论了可对包括核苷酸或核酸在内的许多分子作出的大量修饰,则具体涵盖各种和每一种核苷酸或核酸组合和排列以及可能的修饰,除非有明确的相反指示。此概念适用于本申请的所有方面,包括但不限于形成及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

可进行的多种附加步骤,则可以理解,这些附加步骤中的每一种都可用所公开方法的任何具体实施方案或实施方案的组合来进行,并且具体涵盖每个这样的组合,并且应被认为是已公开的。

[0338] 当本文所述的一些实施方案在本文中已经示出并描述时,这些实施方案仅以实例的方式提供。本领域技术人员在不脱离本文所提供的公开内容的前提下现将想到多种变化、改变和替代。应当理解,本文中所述的实施方案的各种替代方案可用于实施本文所述的方法。

[0339] 除非另有解释,本文使用的所有技术和科学术语具有与本发明所属领域的普通技术人员通常所理解的相同的含义。下列参考文献含有可用于本文的方法和组合物的实施方案:The Merck Manual of Diagnosis and Therapy,第18版,Merck Research Laboratories出版,2006 (ISBN 0-9119102); Benjamin Lewin, Genes IX, Jones & Bartlett Publishing出版,2007 (ISBN-13: 9780763740634); Kendrew等人(编), The Encyclopedia of Mol. Biology, Blackwell Science Ltd.出版,1994 (ISBN 0-632-02182-9); 以及 Robert A. Meyers (ed.), Mol. Biology and Biotechnology: a Comprehensive Desk Reference, VCH Publishers出版, Inc., 1995 (ISBN 1-56081-569-8)。

[0340] 本发明的标准程序描述于例如, 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 (第二版), 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 Methods in Enzymology: Guide to Molecular Cloning Techniques Vol. 152, S.L. Berger和A.R. Kimmmerl (编), Academic Press Inc., San Diego, USA (1987)。Current Protocols in Molecular Biology (CPMB) (Fred M. Ausubel等人编, John Wiley and Sons, Inc.), Current Protocols in Protein Science (CPPS) (John E. Coligan等人编, John Wiley and Sons, Inc.), Current Protocols in Immunology (CPI) (John E. Coligan等人编, John Wiley and Sons, Inc.), Current Protocols in Cell Biology (CPCB) (Juan S. Bonifacino等人编, John Wiley and Sons, Inc.), R. Ian Freshney的 Culture of Animal Cells: A Manual of Basic Technique, 出版商: Wiley-Liss; 第5版 (2005), 以及 Animal Cell Culture Methods (Methods in Cell Biology, Vol. 57, Jennie P. Mather和David Barnes编, Academic Press, 第1版, 1998)。

[0341] 实施例

[0342] 实施例1a-为进行基于乳液的大规模高通量单细胞多核苷酸测序而准备细胞的方案。

[0343] 获得感兴趣的细胞群体。这些包括总PBMC、分选出的细胞、抗体富集的B或T细胞或其他细胞类型。细胞具有完整的质膜,因此它们不会将过多量的mRNA渗漏到周围介质中。细胞不需要是活的。

[0344] 通过在细胞缓冲液1x Dulbecco磷酸盐缓冲盐水(PBS)中离心(对于T细胞或B细胞, 200g, 10min)两次洗涤细胞。然后在细胞缓冲液中将细胞稀释到 3.5×10^6 /mL的细胞浓度。然后将悬浮液移液穿过20 μ m细胞过滤器。

[0345] 实施例1b-为进行基于乳液的大规模高通量单细胞多核苷酸测序而准备固体组织的方案。

[0346] 用包括胶原酶III (200U/mL)、DNA酶I (200U/mL) 和胰蛋白酶 (5mg/mL) 和NEDB (Invitrogen) 在内的多种蛋白酶处理固体组织 (例如, 肿瘤或非肿瘤活检样品) 以产生单独的细胞和含有超过一个细胞的聚集体的混合物。简单来说, 将从小鼠移取的肿瘤加入冷培养基中, 并去除周围的小鼠乳腺组织和脂肪。将肿瘤切成2-4mm的碎片, 然后将其与适当的解离溶液或酶在37℃下温育30min。然后每10min使用尖端切成适合于组织碎片大小的直径的1,000mL微量吸移管将肿瘤碎片上下混合。在每个温育期后, 使碎片穿过40mm尼龙网细胞过滤器过滤。将释放的细胞在1200r.p.m. 下离心2min, 并在4℃下存储于含有30%FCS的冷培养基中。将新鲜的解离溶液添加到剩余的组织碎片30min。当没有额外的细胞释放时, 停止解离。推动碎片穿过筛网, 并将来自所有温育期的所有细胞合并和计数。然后使细胞悬浮液穿过过滤器 (例如, 10、20、30、40μm) 过滤以去除大的聚集体。通过在细胞缓冲液1xDulbecco磷酸盐缓冲盐水 (PBS) 中离心 (对于T细胞或B细胞, 200g, 10分钟) 两次洗涤细胞。在通过乳液分析前, 细胞群体未染色、分选或以其他方式分离。

[0347] 还进行了用于准备移取的肿瘤的替代方法。将移取的肿瘤放入1mL解离缓冲液1 (RPMI+10%FBS中的100U/ml IV型胶原酶和100μg/mL DNA酶) 或解离缓冲液2 (补充有5%FBS、I型胶原酶 (200U/mL) 和DNA酶I (100μg/mL) 的RPMI培养基) 中, 并在37℃下温育30min。如果要随后分离髓细胞, 则分别用10%FBS和IV型胶原酶 (200U/ml) 代替解离缓冲液2中的5%FBS和I型胶原酶。使用1,000mL微量吸移管将肿瘤碎片上下混合。然后使悬浮液穿过70μm过滤器过滤, 并且为了分离髓细胞而用补充有10%FBS的MAC分离缓冲液洗涤3次。对于非常大的肿瘤 (>300mm²), 可使用密度梯度离心法 (Percoll或Ficoll) 预富集炎性细胞。然后将经过滤的细胞悬浮液在400g下离心10min。用10mL MAC缓冲液漂洗沉淀物, 并用相同设置再次离心。

[0348] 实施例2-为了进行基于乳液的大规模高通量单细胞多核苷酸测序而

[0349] 制备乳液反应混合物的方案。

[0350] 将含有以下表1中的试剂和寡核苷酸的乳液反应混合物在室温下在PCR-洁净柜中混合。

[0351] 表1

[0352]

试剂	原液浓度 (mM)	小滴中的最终浓度 (mM)	反应相中的最终浓度 (mM)	μL/200 μL
Tris-Cl, pH 8.0	500.00	50.00	100.00	40.00
MgSO ₄	100.00	3.00	6.00	12.00
DTT	1,000.00	10.00	20.00	4.00
每种 dNTP	10.00	0.50	1.00	20.00
5' 生物素寡聚-dT	1.40×10^{-2}	2.50×10^{-4}	5.00×10^{-4}	7.14
模板转换寡核苷酸	0.1	1.00×10^{-3}	2.00×10^{-3}	4.00
DB 模板分子/μL	1.00×10^6	1.75×10^4	3.50×10^4	7.00
DB 正向引物	0.2	5.00×10^{-4}	1.00×10^{-3}	1.00
DB 反向引物	0.2	7.50×10^{-4}	1.50×10^{-3}	1.50
HALT 蛋白酶抑制剂 (X)	200	1.00	2.00	2.00
RNA 酶抑制剂 (U/μL)	40	0.40	0.80	4.00
MMLV RNA 酶 H-逆转录酶				10.00
Phusion HF DNA 聚合酶				10.00
Triton X-100 (% v/v)	2.5	0.25	0.50	40.00
水				to 200
寡核苷酸序列:				
5' 生物素寡聚-dT 锚定的逆转录引物	/5BiosG//iSp18/TTT TTT TTT TTT TTT TTT TTT TTT T V N			
小滴条码模板:	ATCCATCCACGACTGACGGACGTATTAAANNNNWNNNNW NNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCACC			
模板转换寡核苷酸	AATACGTCCGTCAGTCGTGGATGNNTNNANNTGrGG			
正向容器条码	CATCCACGACTGACGGACGTATT			
反向容器条码	GTGACTGGAGTTCAGACGTGTGCT			

[0353]

[0354] /5Biosg/=5' 生物素修饰;/iSp18/=18-碳间隔区;V=A、C或G;N=任何碱基;rG=核糖鸟苷;W=A或T。

[0355] 实施例3-为了进行基于乳液的大规模高通量单细胞多核苷酸测序而生成乳液的方案。

[0356] 一旦制备了细胞和反应混合物,就形成乳液。使用100μL Hamilton微升注射器将100μL PEEK样品环加载到两个注射的各为约100μL的反应混合物中。使用100μL Hamilton气密注射器将约110μL的细胞悬浮液加载到约100μL、内径0.2mm的FEP管环中。该环附接至机械旋转器,该机械旋转器大约每1-2秒一次不断翻转细胞环以防止细胞沉淀和聚集。通过使聚焦流动喷射穿过具有内部亲氟涂层的Dolomite 2-试剂芯片而形成乳液。外部油性通道含有HFE7500 (Novec 7500) 氟碳油中0.5-5.0% (w/v) 的基于聚乙二醇的表面活性剂。在恒定流速下运行乳液喷射(细胞相和反应相通道中相等)。通过12cm、内径0.5mm的PEEK管,藉由滴入在冷模块中保持在约0°C下的聚丙烯PCR管收集乳液芯片输出物。收集四个级分,每个含有乳液中的50μL水性材料(每个级分的运行时间为5分钟)。用毛细微量吸移管从每个管的底部去除大多数沉淀的油。用40μL Overlay溶液轻轻覆盖每个乳液级分,该溶液为:50mM Na-EDTA,pH 8.0,0.002% (w/v) 甲酚红。在具有下列程序(分钟:秒)的热循环仪中温育乳液:

- [0357] 1. 42.0℃,30:00 (逆转录)
- [0358] 2. 95.0℃,05:00 (使逆转录酶和DNA模板变性)
- [0359] 3. 95.0℃,00:10
- [0360] 4. 65.0℃,00:30
- [0361] 5. 72.0℃,00:30
- [0362] 6. 去往3,总计55次循环 (扩增容器条码并与cDNA融合)
- [0363] 7. 4.0℃,无时间限制
- [0364] 将乳剂在4℃下保持过夜。
- [0365] 实施例4-为了进行基于乳液的大规模高通量单细胞多核苷酸测序而打破乳液的方案。
- [0366] 使用毛细微量吸移管的尖端移除尽可能多的Overlay溶液,而不移除乳液材料。向每个管添加12.5μL Qiagen蛋白酶溶液和2.5μL的0.5M Na-EDTA,pH 8.0。通过添加40μL的1:1FC-40:全氟辛醇并轻轻翻转约10次而打破乳液。
- [0367] 将管的内容物温和离心并在具有下列程序(分钟:秒)的热循环仪中温育:
- [0368] 1. 50℃,15:00 (蛋白酶消化)
- [0369] 2. 70℃,10:00 (蛋白酶灭活)
- [0370] 3. 95℃,03:00 (蛋白酶灭活和DNA变性)
- [0371] 4. 4.0℃,无时间限制
- [0372] 离心该管,并将上层水相和界面移到新的微量离心管中,并在15,000g下离心1分钟。将上层水相转移到新管中,而不干扰界面。
- [0373] 实施例5-为了进行基于乳液的大规模高通量单细胞多核苷酸测序而清洁来自乳液的多核苷酸的方案。
- [0374] 将0.25V的NEB链霉亲和素珠子加入2xBW (10mM Tris-Cl,pH 8.0,1mM EDTA,2M NaCl,0.2%吐温-20)中,并在室温下温育15min。然后将珠子用1xBW洗涤,用0.001%吐温-20洗涤三次,并通过添加0.25V的0.001%吐温-20洗脱,并加热至95℃3min。添加5倍体积的Qiagen缓冲液PB并加至Zyppy硅胶柱上。然后用0.7mL的Zyppy洗涤缓冲液洗涤珠子并在180μL的5mM Tris-Cl,pH 8.8,0.1mM EDTA,0.001%吐温-20中洗脱。
- [0375] 实施例6-为了进行基于乳液的大规模高通量单细胞多核苷酸测序,用于下一代测序的多核苷酸第一PCR反应(PCR1)的方案。
- [0376] 将163.2μL的纯化cDNA用于PCR1。用于第一PCR反应的示例性设置示于以下表2中。
- [0377] 表2

[0378]

PCR1 文库					
试剂	原液浓度	最终浓度	20-μL 反应	60 μL 反应	240 μL 反应
Q5 缓冲液 5X	5.00 mM	1.00 μM	4.00 μL	12.00 μL	48.00 μL
每种 dNTP	10.00 mM	0.20 μM	0.40 μL	1.20 μL	4.80 μL
Q5 热启动	125.00 mM	1.00 μM	0.16 μL	0.48 μL	1.92 μL
633 引物	10 μm		0.16 μL	0.48 μL	1.92 μL
[IgH/TCRα]-[IgL/TCRβ]-[C] 引物混合物	10 μm (各)		0.16 μL	0.48 μL	1.92 μL
cDNA			13.60 μL	40.80 μL	163.20 μL
H ₂ O			1.52 μL	4.56 μL	18.24 μL
[IgH/TCRα/TCRγ]-[IgL/TCRβ/TCRδ]-[C]引物混合物的 IgH/TCRα/TCRγ 引物序列					
IgM	GGGTTGGGGCGGATGCAC				
IgD	CATCCGGAGCCTTGGTGG				
IgA	CCTTGGGGCTGGTCGGGG				
IgE	CGGATGGGCTCTGTGTGG				
IgG	CCGATGGGCCCTTGGTGG				
TCRα1	GGATTTAGAGTCTCTCAGCTG				
TCRα2	CACGGCAGGGTCAGGGTTC				
TCRγ	AAAATAGTGGGCTTGGGG				
[IgH/TCRα/TCRγ]-[IgL/TCRβ/TCRδ]-[C]引物混合物的 IgL/TCRβ/TCRδ 引物序列					
IgKJ1	TTTGATCTCCACCTTGGTCCCCTCCGC				
IgKJ2	TTTGATCTCCAGCTTGGTCCCCTGG				
IgKJ3	TTTGATATCCACTTGGTCCCAGGGC				
IgKJ4	TTTGATTTCACCTTGGTCCCCTGGC				
IgKJ5	TTTAATCTCCAGTCGTGTCCCTTGGC				
IgLJ1	GAGGACGGTCACCTTGGTGCCA				
IgLJ2	TAGGACGGTCAGCTTGGTCCCCTCC				
IgLJ3	GAGGACGGTCAGCTGGGTGCC				
IgLJ4	TAAAATGATCAGCTGGGTTCCTCCAC				
IgLJ5	TAGGACGGTGACCTTGGTCCCAGT				
TCRβ1	GGGAGATCTCTGCTTCTGATG				
TCRβ2	CGACCTCGGGTGGGAACAC				
TCRδ	AGACAAGCGACATTTGTTCCA				
[IgH/TCRα/TCRγ]-[IgL/TCRβ/TCRδ]-[C]引物混合物的 C-引物序列					
633	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT				

[0379] 在PCR管中等分四个60 μ L反应,并在热循环仪中运行下列程序:

[0380] 1. 98 $^{\circ}$ C, 01:00

[0381] 2. 98 $^{\circ}$ C, 00:10

[0382] 3. 64 $^{\circ}$ C, 00:20

[0383] 4. 72 $^{\circ}$ C, 00:20

[0384] 5. 去往2, 总计6次循环

[0385] 6. 4 $^{\circ}$ C, 无时间限制

[0386] 用1.2倍体积的AMPure XP纯化PCR产物,用80%乙醇洗涤,并在60 μ L稀释缓冲液(10mM Tris-Cl, pH 8.0, 0.1mM EDTA)中洗脱。

[0387] 实施例7-为了进行基于乳液的大规模高通量单细胞多核苷酸测序,用于下一代测序的多核苷酸第二PCR反应(PCR2)的方案。

[0388] 将20 μ L纯化的PCR1产物用于每个子文库(例如,IgL或IgH链,或TCR α 或TCR β 链,或TCR γ 或TCR δ 链)。用于第二次PCR反应的示例性设置示于以下表3中。

[0389] 表3

PCR2 文库				
试剂	原液浓度	最终浓度	20 μ L 反应	50 μ L 反应
Q5 缓冲液 5X	5.00 mM	1.00 μ M	4.00 μ L	10.00 μ L
每种 dNTP	10.00 mM	0.20 μ M	0.40 μ L	1.00 μ L
Q5 热启动	125.00 mM	1.00 μ M	0.16 μ L	0.40 μ L
C7-索引-P7 引物	2 μ M		1.60 μ L	4.00 μ L
[P5-IgH/TCR α /TCR γ]-[P5-IgL/TCR β /TCR δ]引物混合物	1 μ M (各)		1.60 μ L	4.00 μ L
cDNA			8.00 μ L	20.00 μ L
H ₂ O			4.24 μ L	10.60 μ L
P5-IgH/TCR α /TCR γ (重) 混合物的引物序列				
IgM	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGTTGGGGCGGATGCA C			
IgD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATCCGGAGCCTTGGTG G			
IgA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTTGGGGCTGGTCGGG G			
IgE	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGATGGGCTCTGTGTG G			
IgG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGATGGGCCCTTGGTG G			
TCR α 1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGATTTAGAGTCTCTCA GCTG			
TCR α 2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACGGCAGGGTCAGGGT TC			
TCR γ	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGGAAACATCTGCATC			

[0390]

[0391]

	AAGT
P5-IgL/TCRβ/TCRδ (轻) 混合物的引物序列	
IgKJ1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGATCTCCACCTTG GTCCCTCCGC
IgKJ2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGATCTCCAGCTT GGTCCCCTGG
IgKJ3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGATATCCACTTTG GTCCCAGGGC
IgKJ4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGATTTCACCTTG GTCCCTTGGC
IgKJ5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTAATCTCCAGTCG TGTCCCTTGGC
IgLJ1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGGACGGTCACCTT GGTGCCA
IgLJ2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGACGGTCAGCTT GGTCCCTCC
IgLJ3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGGACGGTCAGCTG GGTGCC
IgLJ4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAAATGATCAGCTG GGTTCCTCCAC
IgLJ5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGACGGTGACCTT GGTCCCAGT
IgLJ6	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGACGGTCAGCTC GGTCCCC
TCR β 1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGAGATCTCTGCTT CTGATG
TCR β 2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGACCTCGGGTGGGA ACAC
TCR δ	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGATGGTTTGGTAT GAGGC

[0392] 使用“P7-索引-C7”引物,其包含串联的Illumina C7、6-碱基条码和P7序列:

[0393] 5'

[0394] CAAGCAGAAGACGGCATACGAGAT[NNNNN]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
3'。

[0395] 在热循环仪中运行下列程序:

[0396] 1. 98°C,01:00

[0397] 2. 98°C,00:10

[0398] 3. 64°C,00:20

[0399] 4. 72°C,00:20

[0400] 5. 去往3,总计6次循环

[0401] 6. 4°C,无时间限制

[0402] 用1.2倍体积的AMPure纯化PCR产物,并在40 μ L稀释缓冲液中洗脱。[0403] 实施例8-为了进行基于乳液的大规模高通量单细胞多核苷酸测序,用于下一代测
序的多核苷酸第三PCR反应(PCR3)的方案。[0404] 将8 μ L纯化的PCR2产物用于qPCR以确定扩增循环的最终数目。用于第三PCR反应的
设置示于以下表4中。

[0405] 表4

[0406]

qPCR3a 文库			
试剂	原液浓度	最终浓度	20 μ L 反应
Q5 缓冲液 5X	5.00 mM	1.00 μ M	4.00 μ L
每种 dNTP	10.00 mM	0.20 μ M	0.40 μ L
SYBR Green I 1:500	83.00 mM	1.00 μ M	0.24 μ L
Q5 热启动	125.00 mM	1.00 μ M	0.16 μ L
C5-P5 引物	10.00 μ M	0.40 μ M	0.80 μ L
C7 引物	10.00 μ M	0.40 μ M	0.80 μ L
cDNA			8.00 μ L
H ₂ O			5.60 μ L
引物序列			
P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT		
C7	CAAGCAGAAGACGGCATACGAGAT		

[0407] 在qPCR机器中运行下列程序：

[0408] 1. 98℃, 01:00

[0409] 2. 98℃, 00:10

[0410] 3. 64℃, 00:20

[0411] 4. 72℃, 00:20

[0412] 5. 读板

[0413] 6. 去往2, 总计25次循环

[0414] 7. 4℃, 无时间限制

[0415] 检查qPCR强度图以确定荧光强度最大但DNA的指数扩增还没有结束时的扩增循环。这就是对于PCR3终点的最终循环数。

[0416] 将24.0 μ L的纯化PCR2产物用于终点PCR3。用于PCR反应以确定第三PCR终点的循环数的示例性设置示于以下表5中。

[0417] 表5

[0418]

qPCR3b 文库			
试剂	原液浓度	最终浓度	60 μ L 反应
Q5 缓冲液 5X	5.00 mM	1.00 μ M	12.00 μ L
每种 dNTP	10.00 mM	0.20 μ M	1.20 μ L
H ₂ O	83.00 mM	1.00 μ M	0.72 μ L
Q5 热启动	125.00 mM	1.00 μ M	0.48 μ L
C5-P5 引物	10.00 μ M	0.40 μ M	2.40 μ L
C7 引物	10.00 μ M	0.40 μ M	2.40 μ L
cDNA			24.00 μ L
H ₂ O			16.80 μ L

[0419] 在热循环仪中运行下列程序：

[0420] 1. 98℃, 01:00

[0421] 2. 98℃, 00:10

[0422] 3. 64℃, 00:20

[0423] 4. 72℃, 00:20

[0424] 5. 去往2, 进行确定数目的循环

[0425] 6. 4℃, 无时间限制

[0426] 用1.2倍体积的AMPure纯化PCR产物,并在20 μ L的稀释缓冲液中洗脱。文库已准备好用于测序。它们按照需要合并,经过或不经过琼脂糖凝胶纯化以去除污染性的截短扩增子,然后使用下一代测序技术平台进行测序。

[0427] 实施例9-读取处理和同种型分配

[0428] 使用围绕pRESTO软件包1建立的定制管道处理Illumina MiSeq读取,以生成mRNA分子和小滴的全长共有序列,用IgBLAST和IMGT/HighV-QUEST加以注释,并用自定义脚本和Change-0软件包处理以生成统计数据 and 图片。使用Illumina软件对MiSeq读取进行多路解编(demultiplexing)。用N遮盖小于Phred质量5的位置。在扩增子中鉴别同种型特异性引物、小滴条码(DB)和分子条码(MB),并用pRESTO MaskPrimers-cut以0.2的最大错误修剪。对于每个mRNA,从根据独特分子标识符(UMI)分组的读取分别生成读取1共有序列和读取2共有序列,该标识符包含DB和MB两者,它们是来源于相同原始mRNA起源分子的PCR复制物。用MUSCLE比对UMI读取组,并使用pRESTO构建具有下列参数的共有序列: $\text{maxdiv}=0.1$; bfPRIMER ; $\text{prfreq}=0.6$; $\text{maxmiss}=0.5$; $q=5$; 对于读取组,判定的PCR引物序列一致性 $\geq 60\%$; 最大核苷酸多样性 $=0.1$; 在插入缺失位点上使用多数决定规则; 并遮盖具有低后验(共有)质量的比对列。然后分两轮缝合配对的末端共有序列。首先,使用Z-评分近似值优化每个读取对的共有序列末端的无空位比对,并用如在pRESTO AssemblePairs-align中执行的二项p值评分,采用下列参数: 最小长度 $=8$; $\alpha 1 \times 10^{-5}$; 以及最大错误 $=0.3$ 。对于未能以此方式缝合的读取对,尝试使用人类BCR和TCR种系V外显子缝合,以在缝合或空位读取连接之前支持每个读取,使用pRESTO的AssemblePairs-参考参数: 最小同一性 $=0.5$; e 值 1×10^{-5} 。

[0429] 实施例10-V(D)J区段注释和同种型证实

[0430] 使用IgBLAST、Change-0和自定义脚本来鉴别起源种系V(D)J基因、修剪mRNA序列为V(D)J区域、鉴别CDR3区域并计算来自种系V核苷酸序列的突变。IgBLAST将N计数为错配,但是具有超过6个V-区N的mRNA序列被过滤用于突变分析和交叉部分配对精确度分析。对于IG重链,通过匹配非引物C-区(恒定区外显子)与预期序列证实同种型身份,使用pRESTO MaskPrimers-评分参数: 开始 $=0$; 最大错误 $=0.2$ 。丢弃具有不一致的引物/非引物C-区判定的扩增子,除了两个引物/非引物组合,其中特定引物串扰事件通过目测来解析。

[0431] 实施例11-将V(D)J序列分组为克隆谱系

[0432] 采用加权克隆内距离使用单连接簇集将V(D)J序列分组为克隆。用Change-0软件包DefineClone-按组参数进行簇集: 模式 $=\text{mln}$; 基因 $=$ 第一; 距离 $=4.0$; $\text{norm}=\text{无}$ 。首先,将所有功能性Ig V_H链的小滴共有序列分箱到V-J连接箱元中,使得可能来源于相同初始重组事件的序列分箱在一起(基于最佳匹配的Ig V_H基因,最佳匹配的Ig J_H基因,以及通过IMGT/HighV-QUEST鉴别的接合长度)。通过使用Change-0的shm软件包的distToNearest功能生成每个Ig V_H箱元中最近邻距离的直方图,并目测该直方图的自然距离截断值(在双峰直方图的低谷中),来选择克隆内距离阈值。使用相同的距离模型和阈值定义轻链的克隆簇。

[0433] 实施例12-小滴过滤、配对保真度计算

[0434] 以两种独立方式评估重链配对置信度: 使用小滴内mRNA序列一致性和复制物对间的一致性。小滴内mRNA一致性被定义为基因座内V(D)J序列的平均成对核苷酸的差异(Nei的 $\text{pi} < 0.02$)。使用IgBLAST注释将mRNA序列修剪为V(D)J核苷酸编码序列。在每个小滴内,所

有生成性的mRNA序列均按照V基因座来分组。在每组内,使用如在pRESTO AlignSets中执行的MUSCLE,使用默认参数比对多个序列。使用pRESTO参数从每个基因座多个mRNA建立小滴共有链:BuildConsensus.py;最大div=0.2;最大miss=0.5。使用随机扰乱的小滴选择多样性截断值 $\pi \leq 0.02$ 。在扰乱的小滴中,少于0.01%的重链基因座(<0.2%的轻链基因座)符合此标准。分离包含多细胞或免疫受体的小滴以供进一步的精确度分析。

[0435] 基于对多个复制物(单独的乳液实验)中相同克隆对的观察计算配对精确度,集中于可能只含有单个谱系的那些VDJ簇,即来源于单个V(D)J和VJ重排,然后扩充。相似的VDJ重排可出现于单独的多个独立时间内,导致相同的重链V(D)J重排天然地与多个不同的轻链VJ重排配对。因为罕见的V(D)J重排将对通过本文所述方法达到的技术精确度提供更准确的量度,所以将长的重CDR3(CDR3_H)用于此分析的焦点(作为对于罕见V(D)J重排的替代)。具有>6N的序列也将去除,以提高克隆分配置信度。配对精确度随CDR3_H长度而增加,对于在整个级分中观察到的最长四分之一的克隆(接合长度 ≥ 54 nt的2,604个克隆)达到超过96%。因为克隆对一致性的概率是两个独立实验中真实对的接合概率,所以配对精确度估计为整个复制物中配对一致性的平方根,计算如下,其中 d_{hl}^f 为小滴条码的数目d,具有配对的重克隆h和轻克隆l,并在物理级分f中发现。对于每个实验的平均(平方)配对精确度通过对重克隆h和级分的所有配对(f,g)、配对的轻克隆的一致性(l,k)求平均而估计:

$$\langle \text{精确度}^2 \rangle = \text{平均值}(P_f P_g)$$

$$[0436] \quad = \frac{\text{整个级分中一致的重链对}}{\text{整个级分中观察到重克隆的总对}}$$

$$[0437] \quad \frac{\text{一致的重链对}}{\text{一致的对} + \text{不一致的对}} = \frac{\sum_h (\sum_{l=k}^{f \neq g} d_{hl}^f \cdot d_{hk}^g)}{\sum_h (\sum_{l=k}^{f \neq g} d_{hl}^f \cdot d_{hk}^g + \sum_{l \neq k}^{f \neq g} d_{hl}^f \cdot d_{hk}^g)}$$

$$[0438] \quad \langle \text{精确度}^2 \rangle = \frac{33157}{35922}$$

[0439] 因此,根据此示例性实验,每个实验的平均精确度(实验之间的精确度变化之内)为96.1%。

[0440] 实施例13-HIV系统发生分析

[0441] 通过针对与已知bNAbs的相似性挖掘我们的高通量配对抗体处理的序列来发现新的针对HIV的广泛中和抗体(bNAbs)。从文献中挖掘来自PGT供体和其他供体的先前已知的bNAbs。从乳液中回收的所有HIV IgH mRNA经由tblastx 10针对与已知CDR3氨基酸序列的相似性进行评分。使用来自健康供体的IgH mRNA序列以生成序列相似性的背景分布,使用很小的评分截止值27来分离候选bNAbs样CDR3以供进一步分析。使用MUSCLE 11以默认参数将候选序列的V(D)J序列与已知的bNAbs进行比对,并特别是除了gapopen=-15外使用默认参数与PGT-供体谱系进行比对。用PhyML默认参数生成树,用Newick Utils和Dendroscope操作并可视化,并手动检查以选择穿插有已知bNAbs序列的免疫球蛋白重链序列。如先前所述,用JALVIEW中使用的任何小滴内氨基酸冲突的比对的人工检查,建立对于每个小滴的共有序列。选择八个重链序列和它们自然配对的轻链抗体序列以供合成、克隆、表达和中和分

析。

[0442] 实施例14-数据分析和作图

[0443] 使用dplyr和ggplot2 R软件包生成图。为了只在散点图中进行可视化,对数据进行随机低取样和/或用R进行抖动。低取样最小值为每同种型20,000个小滴或如另外标注。通过添加从相同均匀概率分布提取的垂直和水平噪音来对点进行抖动,对于mRNA单位的最大值 ≤ 0.2 ,且对于突变 $\leq 0.6\%$ 。

[0001] 序 列 表(根据PCT条约第28条修改替换页)

[0002] <110> 阿布维特罗公司

[0003] <120> 高通量核苷酸文库测序

[0004] <130> 44243-709.601

[0005] <140> PCT/US2015/050119

[0006] <141> 2015-09-15

[0007] <150> 62/051,832

[0008] <151> 2014-09-17

[0009] <150> 62/050,549

[0010] <151> 2014-09-15

[0011] <160> 57

[0012] <170> PatentIn version 3.5

[0013] <210> 1

[0014] <211> 18

[0015] <212> DNA

[0016] <213> 人工序列

[0017] <220>

[0018] <221> 来源

[0019] <223> /note="人工序列的描述:合成

[0020] 引物"

[0021] <400> 1

[0022] gggttggggc ggatgcac 18

[0023] <210> 2

[0024] <211> 18

[0025] <212> DNA

[0026] <213> 人工序列

[0027] <220>

[0028] <221> 来源

[0029] <223> /note="人工序列的描述:合成

[0030] 引物"

[0031] <400> 2

[0032] catccggagc cttggtgg 18

[0033] <210> 3

[0034] <211> 18

[0035] <212> DNA

[0036] <213> 人工序列

[0037] <220>

[0038] <221> 来源

[0039] <223> /note="人工序列的描述:合成
[0040] 引物"
[0041] <400> 3
[0042] ccttggggct ggtcgggg 18
[0043] <210> 4
[0044] <211> 18
[0045] <212> DNA
[0046] <213> 人工序列
[0047] <220>
[0048] <221> 来源
[0049] <223> /note="人工序列的描述:合成
[0050] 引物"
[0051] <400> 4
[0052] cggatgggct ctgtgtgg 18
[0053] <210> 5
[0054] <211> 18
[0055] <212> DNA
[0056] <213> 人工序列
[0057] <220>
[0058] <221> 来源
[0059] <223> /note="人工序列的描述:合成
[0060] 引物"
[0061] <400> 5
[0062] ccgatgggcc cttggtgg 18
[0063] <210> 6
[0064] <211> 21
[0065] <212> DNA
[0066] <213> 人工序列
[0067] <220>
[0068] <221> 来源
[0069] <223> /note="人工序列的描述:合成
[0070] 引物"
[0071] <400> 6
[0072] ggatttagag tctctcagct g 21
[0073] <210> 7
[0074] <211> 19
[0075] <212> DNA
[0076] <213> 人工序列
[0077] <220>

[0078] <221> 来源
[0079] <223> /note="人工序列的描述:合成
[0080] 引物"
[0081] <400> 7
[0082] cacggcaggg tcagggttc 19
[0083] <210> 8
[0084] <211> 21
[0085] <212> DNA
[0086] <213> 人工序列
[0087] <220>
[0088] <221> 来源
[0089] <223> /note="人工序列的描述:合成
[0090] 引物"
[0091] <400> 8
[0092] ggggaaacat ctgcatcaag t 21
[0093] <210> 9
[0094] <211> 26
[0095] <212> DNA
[0096] <213> 人工序列
[0097] <220>
[0098] <221> 来源
[0099] <223> /note="人工序列的描述:合成
[0100] 引物"
[0101] <400> 9
[0102] tttgatctcc accttggtcc ctccgc 26
[0103] <210> 10
[0104] <211> 25
[0105] <212> DNA
[0106] <213> 人工序列
[0107] <220>
[0108] <221> 来源
[0109] <223> /note="人工序列的描述:合成
[0110] 引物"
[0111] <400> 10
[0112] tttgatctcc agcttggtcc cctgg 25
[0113] <210> 11
[0114] <211> 26
[0115] <212> DNA
[0116] <213> 人工序列

[0117] <220>
[0118] <221> 来源
[0119] <223> /note="人工序列的描述:合成
[0120] 引物"
[0121] <400> 11
[0122] ttgatatcc actttggtcc cagggc 26
[0123] <210> 12
[0124] <211> 26
[0125] <212> DNA
[0126] <213> 人工序列
[0127] <220>
[0128] <221> 来源
[0129] <223> /note="人工序列的描述:合成
[0130] 引物"
[0131] <400> 12
[0132] ttgatattcc accttggtcc cttggc 26
[0133] <210> 13
[0134] <211> 26
[0135] <212> DNA
[0136] <213> 人工序列
[0137] <220>
[0138] <221> 来源
[0139] <223> /note="人工序列的描述:合成
[0140] 引物"
[0141] <400> 13
[0142] tttaatctcc agtcgtgtcc cttggc 26
[0143] <210> 14
[0144] <211> 22
[0145] <212> DNA
[0146] <213> 人工序列
[0147] <220>
[0148] <221> 来源
[0149] <223> /note="人工序列的描述:合成
[0150] 引物"
[0151] <400> 14
[0152] gaggacggtc accttggtgc ca 22
[0153] <210> 15
[0154] <211> 24
[0155] <212> DNA

[0156] <213> 人工序列
[0157] <220>
[0158] <221> 来源
[0159] <223> /note="人工序列的描述:合成
[0160] 引物"
[0161] <400> 15
[0162] taggacgggc agcttggtcc ctcc 24
[0163] <210> 16
[0164] <211> 21
[0165] <212> DNA
[0166] <213> 人工序列
[0167] <220>
[0168] <221> 来源
[0169] <223> /note="人工序列的描述:合成
[0170] 引物"
[0171] <400> 16
[0172] gaggacgggc agctgggtgc c 21
[0173] <210> 17
[0174] <211> 26
[0175] <212> DNA
[0176] <213> 人工序列
[0177] <220>
[0178] <221> 来源
[0179] <223> /note="人工序列的描述:合成
[0180] 引物"
[0181] <400> 17
[0182] taaaatgata agctgggttc ctccac 26
[0183] <210> 18
[0184] <211> 23
[0185] <212> DNA
[0186] <213> 人工序列
[0187] <220>
[0188] <221> 来源
[0189] <223> /note="人工序列的描述:合成
[0190] 引物"
[0191] <400> 18
[0192] taggacgggtg accttggtcc cag 23
[0193] <210> 19
[0194] <211> 21

[0195] <212> DNA
[0196] <213> 人工序列
[0197] <220>
[0198] <221> 来源
[0199] <223> /note="人工序列的描述:合成
[0200] 引物"
[0201] <400> 19
[0202] gggagatctc tgcttctgat g 21
[0203] <210> 20
[0204] <211> 19
[0205] <212> DNA
[0206] <213> 人工序列
[0207] <220>
[0208] <221> 来源
[0209] <223> /note="人工序列的描述:合成
[0210] 引物"
[0211] <400> 20
[0212] cgacctcggg tgggaacac 19
[0213] <210> 21
[0214] <211> 20
[0215] <212> DNA
[0216] <213> 人工序列
[0217] <220>
[0218] <221> 来源
[0219] <223> /note="人工序列的描述:合成
[0220] 引物"
[0221] <400> 21
[0222] cggatggttt ggtatgaggc 20
[0223] <210> 22
[0224] <211> 20
[0225] <212> DNA
[0226] <213> 人工序列
[0227] <220>
[0228] <221> 来源
[0229] <223> /note="人工序列的描述:合成
[0230] 寡核苷酸"
[0231] <220>
[0232] <221> 修饰的碱基
[0233] <222> (1) .. (4)

[0234] <223> a、c、t、g、未知或其他
[0235] <220>
[0236] <221> 修饰的碱基
[0237] <222> (6) .. (9)
[0238] <223> a、c、t、g、未知或其他
[0239] <220>
[0240] <221> 修饰的碱基
[0241] <222> (11) .. (14)
[0242] <223> a、c、t、g、未知或其他
[0243] <220>
[0244] <221> 修饰的碱基
[0245] <222> (16) .. (19)
[0246] <223> a、c、t、g、未知或其他
[0247] <400> 22
[0248] nnnnwnnnnw nnnnwnnnnw 20
[0249] <210> 23
[0250] <211> 6
[0251] <212> PRT
[0252] <213> 人工序列
[0253] <220>
[0254] <221> 来源
[0255] <223> /note="人工序列的描述:合成
[0256] 6xHis标签"
[0257] <400> 23
[0258] His His His His His His
[0259] 1 5
[0260] <210> 24
[0261] <211> 27
[0262] <212> DNA
[0263] <213> 人工序列
[0264] <220>
[0265] <221> 来源
[0266] <223> /note="人工序列的描述:合成
[0267] 引物"
[0268] <220>
[0269] <221> 修饰的碱基
[0270] <222> (27) .. (27)
[0271] <223> a、c、t、g、未知或其他
[0272] <400> 24

[0273] ttttttttttt ttttttttttt tttttvn 27

[0274] <210> 25

[0275] <211> 78

[0276] <212> DNA

[0277] <213> 人工序列

[0278] <220>

[0279] <221> 来源

[0280] <223> /note="人工序列的描述:合成

[0281] 寡核苷酸"

[0282] <220>

[0283] <221> 修饰的碱基

[0284] <222> (30) .. (33)

[0285] <223> a、c、t、g、未知或其他

[0286] <220>

[0287] <221> 修饰的碱基

[0288] <222> (35) .. (38)

[0289] <223> a、c、t、g、未知或其他

[0290] <220>

[0291] <221> 修饰的碱基

[0292] <222> (40) .. (43)

[0293] <223> a、c、t、g、未知或其他

[0294] <400> 25

[0295] atccatccac gactgacgga cgtattaaan nnnwnnnwnn nnnagatcgg aagagcacac 60

[0296] gtctgaactc cagtcacc 78

[0297] <210> 26

[0298] <211> 35

[0299] <212> DNA

[0300] <213> 人工序列

[0301] <220>

[0302] <221> 来源

[0303] <223> /note="人工序列的描述:合成

[0304] 寡核苷酸"

[0305] <220>

[0306] <221> 来源

[0307] <223> /note="组合DNA/RNA分子的描述:合成

[0308] 寡核苷酸"

[0309] <220>

[0310] <221> 修饰的碱基

[0311] <222> (24) .. (25)

[0312]	<223> a、c、t、g、未知或其他
[0313]	<220>
[0314]	<221> 修饰的碱基
[0315]	<222> (27) .. (28)
[0316]	<223> a、c、t、g、未知或其他
[0317]	<220>
[0318]	<221> 修饰的碱基
[0319]	<222> (30) .. (31)
[0320]	<223> a、c、t、g、未知或其他
[0321]	<400> 26
[0322]	aatacgtccg tcagtcgtgg atgnntnnan ntggg 35
[0323]	<210> 27
[0324]	<211> 23
[0325]	<212> DNA
[0326]	<213> 人工序列
[0327]	<220>
[0328]	<221> 来源
[0329]	<223> /note="人工序列的描述:合成
[0330]	寡核苷酸"
[0331]	<400> 27
[0332]	catccacgac tgacggacgt att 23
[0333]	<210> 28
[0334]	<211> 24
[0335]	<212> DNA
[0336]	<213> 人工序列
[0337]	<220>
[0338]	<221> 来源
[0339]	<223> /note="人工序列的描述:合成
[0340]	寡核苷酸"
[0341]	<400> 28
[0342]	gtgactggag ttcagacgtg tgct 24
[0343]	<210> 29
[0344]	<211> 18
[0345]	<212> DNA
[0346]	<213> 人工序列
[0347]	<220>
[0348]	<221> 来源
[0349]	<223> /note="人工序列的描述:合成
[0350]	引物"

[0351] <400> 29
[0352] aaaatagtgg gcttgggg 18
[0353] <210> 30
[0354] <211> 24
[0355] <212> DNA
[0356] <213> 人工序列
[0357] <220>
[0358] <221> 来源
[0359] <223> /note="人工序列的描述:合成
[0360] 引物"
[0361] <400> 30
[0362] taggacggtg accttggtcc cagt 24
[0363] <210> 31
[0364] <211> 21
[0365] <212> DNA
[0366] <213> 人工序列
[0367] <220>
[0368] <221> 来源
[0369] <223> /note="人工序列的描述:合成
[0370] 引物"
[0371] <400> 31
[0372] agacaagcga catttggtcc a 21
[0373] <210> 32
[0374] <211> 34
[0375] <212> DNA
[0376] <213> 人工序列
[0377] <220>
[0378] <221> 来源
[0379] <223> /note="人工序列的描述:合成
[0380] 引物"
[0381] <400> 32
[0382] gtgactggag ttcagacgtg tgctcttccg atct 34
[0383] <210> 33
[0384] <211> 51
[0385] <212> DNA
[0386] <213> 人工序列
[0387] <220>
[0388] <221> 来源
[0389] <223> /note="人工序列的描述:合成

[0390] 引物”
[0391] <400> 33
[0392] acactctttc cctacacgac gctcttccga tctgggttgg ggcggatgca c 51
[0393] <210> 34
[0394] <211> 51
[0395] <212> DNA
[0396] <213> 人工序列
[0397] <220>
[0398] <221> 来源
[0399] <223> /note=“人工序列的描述:合成
[0400] 引物”
[0401] <400> 34
[0402] acactctttc cctacacgac gctcttccga tctcatccgg agccttggtg g 51
[0403] <210> 35
[0404] <211> 51
[0405] <212> DNA
[0406] <213> 人工序列
[0407] <220>
[0408] <221> 来源
[0409] <223> /note=“人工序列的描述:合成
[0410] 引物”
[0411] <400> 35
[0412] acactctttc cctacacgac gctcttccga tctccttggg gctggtcggg g 51
[0413] <210> 36
[0414] <211> 51
[0415] <212> DNA
[0416] <213> 人工序列
[0417] <220>
[0418] <221> 来源
[0419] <223> /note=“人工序列的描述:合成
[0420] 引物”
[0421] <400> 36
[0422] acactctttc cctacacgac gctcttccga tctcggtatgg gctctgtgtg g 51
[0423] <210> 37
[0424] <211> 51
[0425] <212> DNA
[0426] <213> 人工序列
[0427] <220>
[0428] <221> 来源

[0429] <223> /note="人工序列的描述:合成
[0430] 引物"
[0431] <400> 37
[0432] acactctttc cctacacgac gctcttccga tctccgatgg gcccttggtg g 51
[0433] <210> 38
[0434] <211> 54
[0435] <212> DNA
[0436] <213> 人工序列
[0437] <220>
[0438] <221> 来源
[0439] <223> /note="人工序列的描述:合成
[0440] 引物"
[0441] <400> 38
[0442] acactctttc cctacacgac gctcttccga tctggattta gagtctctca gctg 54
[0443] <210> 39
[0444] <211> 52
[0445] <212> DNA
[0446] <213> 人工序列
[0447] <220>
[0448] <221> 来源
[0449] <223> /note="人工序列的描述:合成
[0450] 引物"
[0451] <400> 39
[0452] acactctttc cctacacgac gctcttccga tctcacggca gggtcagggt tc 52
[0453] <210> 40
[0454] <211> 54
[0455] <212> DNA
[0456] <213> 人工序列
[0457] <220>
[0458] <221> 来源
[0459] <223> /note="人工序列的描述:合成
[0460] 引物"
[0461] <400> 40
[0462] acactctttc cctacacgac gctcttccga tctggggaaa catctgcac aagt 54
[0463] <210> 41
[0464] <211> 59
[0465] <212> DNA
[0466] <213> 人工序列
[0467] <220>

[0468] <221> 来源
[0469] <223> /note="人工序列的描述:合成
[0470] 引物"
[0471] <400> 41
[0472] acactctttc cctacacgac gctcttccga tcttttgatc tccaccttgg tccctccgc 59
[0473] <210> 42
[0474] <211> 58
[0475] <212> DNA
[0476] <213> 人工序列
[0477] <220>
[0478] <221> 来源
[0479] <223> /note="人工序列的描述:合成
[0480] 引物"
[0481] <400> 42
[0482] acactctttc cctacacgac gctcttccga tcttttgatc tccagcttgg tcccctgg 58
[0483] <210> 43
[0484] <211> 59
[0485] <212> DNA
[0486] <213> 人工序列
[0487] <220>
[0488] <221> 来源
[0489] <223> /note="人工序列的描述:合成
[0490] 引物"
[0491] <400> 43
[0492] acactctttc cctacacgac gctcttccga tcttttgata tccactttgg tcccagggc 59
[0493] <210> 44
[0494] <211> 59
[0495] <212> DNA
[0496] <213> 人工序列
[0497] <220>
[0498] <221> 来源
[0499] <223> /note="人工序列的描述:合成
[0500] 引物"
[0501] <400> 44
[0502] acactctttc cctacacgac gctcttccga tcttttgatt tccaccttgg tcccttggc 59
[0503] <210> 45
[0504] <211> 59
[0505] <212> DNA
[0506] <213> 人工序列

[0507] <220>
[0508] <221> 来源
[0509] <223> /note="人工序列的描述:合成
[0510] 引物"
[0511] <400> 45
[0512] acactctttc cctacacgac gctcttccga tcttttaatc tccagtcgtg tcccttggc 59
[0513] <210> 46
[0514] <211> 55
[0515] <212> DNA
[0516] <213> 人工序列
[0517] <220>
[0518] <221> 来源
[0519] <223> /note="人工序列的描述:合成
[0520] 引物"
[0521] <400> 46
[0522] acactctttc cctacacgac gctcttccga tctgaggacg gtcaccttgg tgcca 55
[0523] <210> 47
[0524] <211> 57
[0525] <212> DNA
[0526] <213> 人工序列
[0527] <220>
[0528] <221> 来源
[0529] <223> /note="人工序列的描述:合成
[0530] 引物"
[0531] <400> 47
[0532] acactctttc cctacacgac gctcttccga tcttaggacg gtcagcttgg tccctcc 57
[0533] <210> 48
[0534] <211> 54
[0535] <212> DNA
[0536] <213> 人工序列
[0537] <220>
[0538] <221> 来源
[0539] <223> /note="人工序列的描述:合成
[0540] 引物"
[0541] <400> 48
[0542] acactctttc cctacacgac gctcttccga tctgaggacg gtcagctggg tgcc 54
[0543] <210> 49
[0544] <211> 59
[0545] <212> DNA

- [0546] <213> 人工序列
- [0547] <220>
- [0548] <221> 来源
- [0549] <223> /note="人工序列的描述:合成
- [0550] 引物"
- [0551] <400> 49
- [0552] acactctttc cctacacgac gctcttccga tctttaaagt atcagctggg ttcctccac 59
- [0553] <210> 50
- [0554] <211> 57
- [0555] <212> DNA
- [0556] <213> 人工序列
- [0557] <220>
- [0558] <221> 来源
- [0559] <223> /note="人工序列的描述:合成
- [0560] 引物"
- [0561] <400> 50
- [0562] acactctttc cctacacgac gctcttccga tcttaggacg gtgaccttgg tcccagt 57
- [0563] <210> 51
- [0564] <211> 55
- [0565] <212> DNA
- [0566] <213> 人工序列
- [0567] <220>
- [0568] <221> 来源
- [0569] <223> /note="人工序列的描述:合成
- [0570] 引物"
- [0571] <400> 51
- [0572] acactctttc cctacacgac gctcttccga tcttaggacg gtcagctcgg tcccc 55
- [0573] <210> 52
- [0574] <211> 54
- [0575] <212> DNA
- [0576] <213> 人工序列
- [0577] <220>
- [0578] <221> 来源
- [0579] <223> /note="人工序列的描述:合成
- [0580] 引物"
- [0581] <400> 52
- [0582] acactctttc cctacacgac gctcttccga tctgggagat ctctgcttct gatg 54
- [0583] <210> 53
- [0584] <211> 52

- [0585] <212> DNA
[0586] <213> 人工序列
[0587] <220>
[0588] <221> 来源
[0589] <223> /note="人工序列的描述:合成
[0590] 引物"
[0591] <400> 53
[0592] acactctttc cctacacgac gctcttccga tctcgacctc gggtgggaac ac 52
[0593] <210> 54
[0594] <211> 53
[0595] <212> DNA
[0596] <213> 人工序列
[0597] <220>
[0598] <221> 来源
[0599] <223> /note="人工序列的描述:合成
[0600] 引物"
[0601] <400> 54
[0602] acactctttc cctacacgac gctcttccga tctcgatggg ttggtatga ggc 53
[0603] <210> 55
[0604] <211> 64
[0605] <212> DNA
[0606] <213> 人工序列
[0607] <220>
[0608] <221> 来源
[0609] <223> /note="人工序列的描述:合成
[0610] 寡核苷酸"
[0611] <220>
[0612] <221> 修饰的碱基
[0613] <222> (25) .. (30)
[0614] <223> a、c、t、g、未知或其他
[0615] <400> 55
[0616] caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgctcttccg 60
[0617] atct 64
[0618] <210> 56
[0619] <211> 58
[0620] <212> DNA
[0621] <213> 人工序列
[0622] <220>
[0623] <221> 来源

- [0624] <223> /note="人工序列的描述:合成
[0625] 引物"
[0626] <400> 56
[0627] aatgatacgg cgaccaccga gatctacact ctttcctac acgacgtct tccgatct 58
[0628] <210> 57
[0629] <211> 24
[0630] <212> DNA
[0631] <213> 人工序列
[0632] <220>
[0633] <221> 来源
[0634] <223> /note="人工序列的描述:合成
[0635] 引物"
[0636] <400> 57
[0637] caagcagaag acggcatacg agat 24

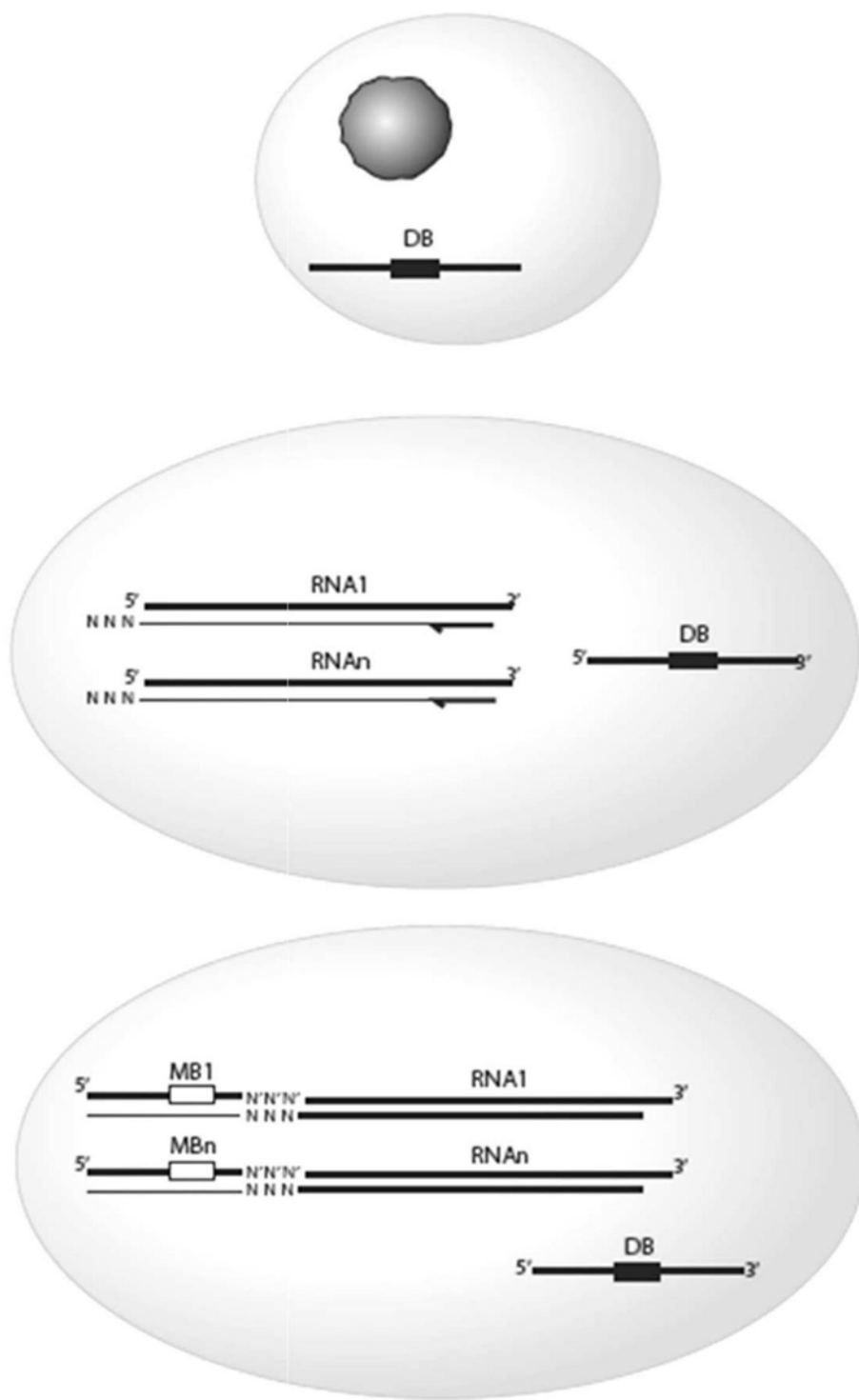


图1A

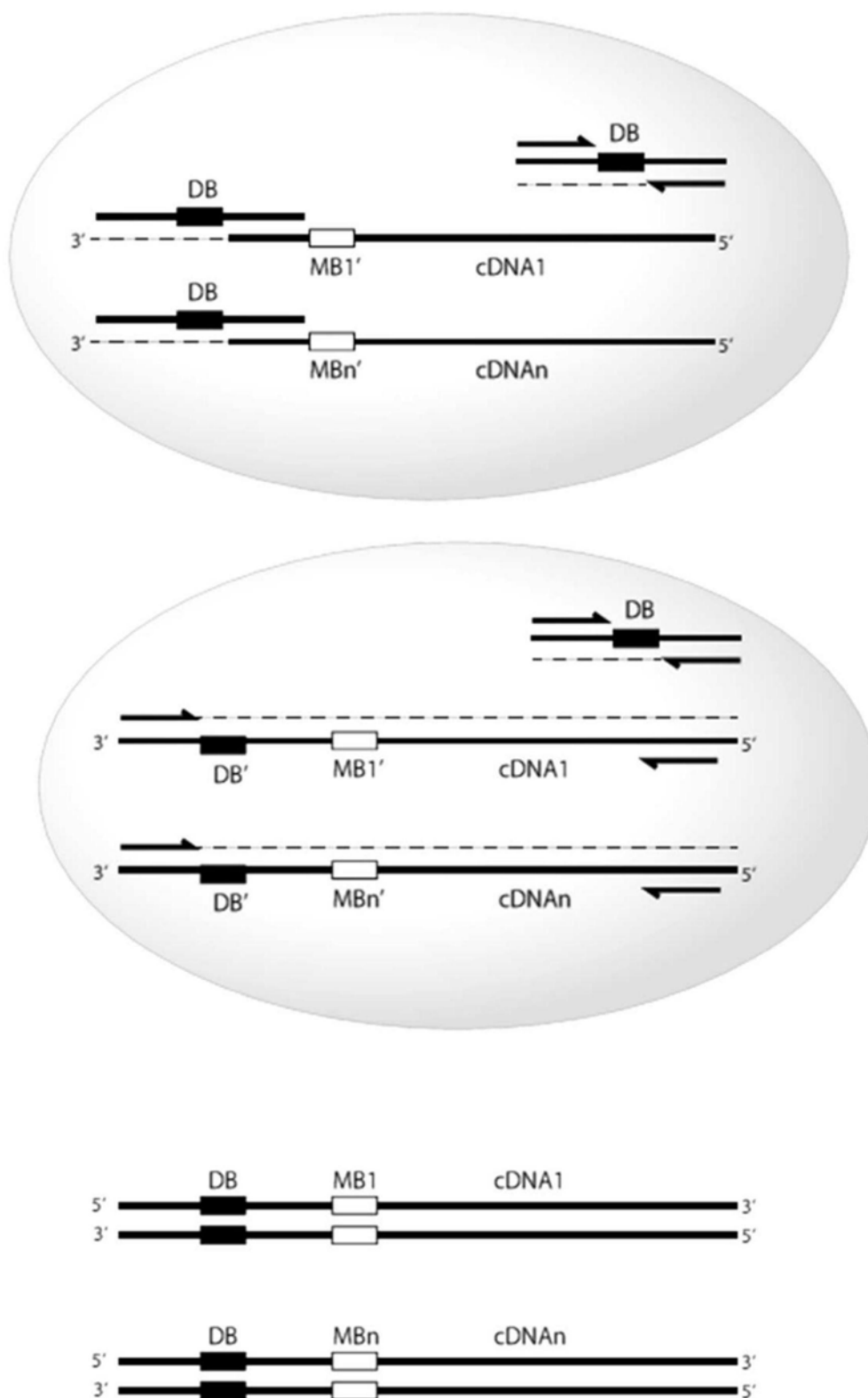


图1B

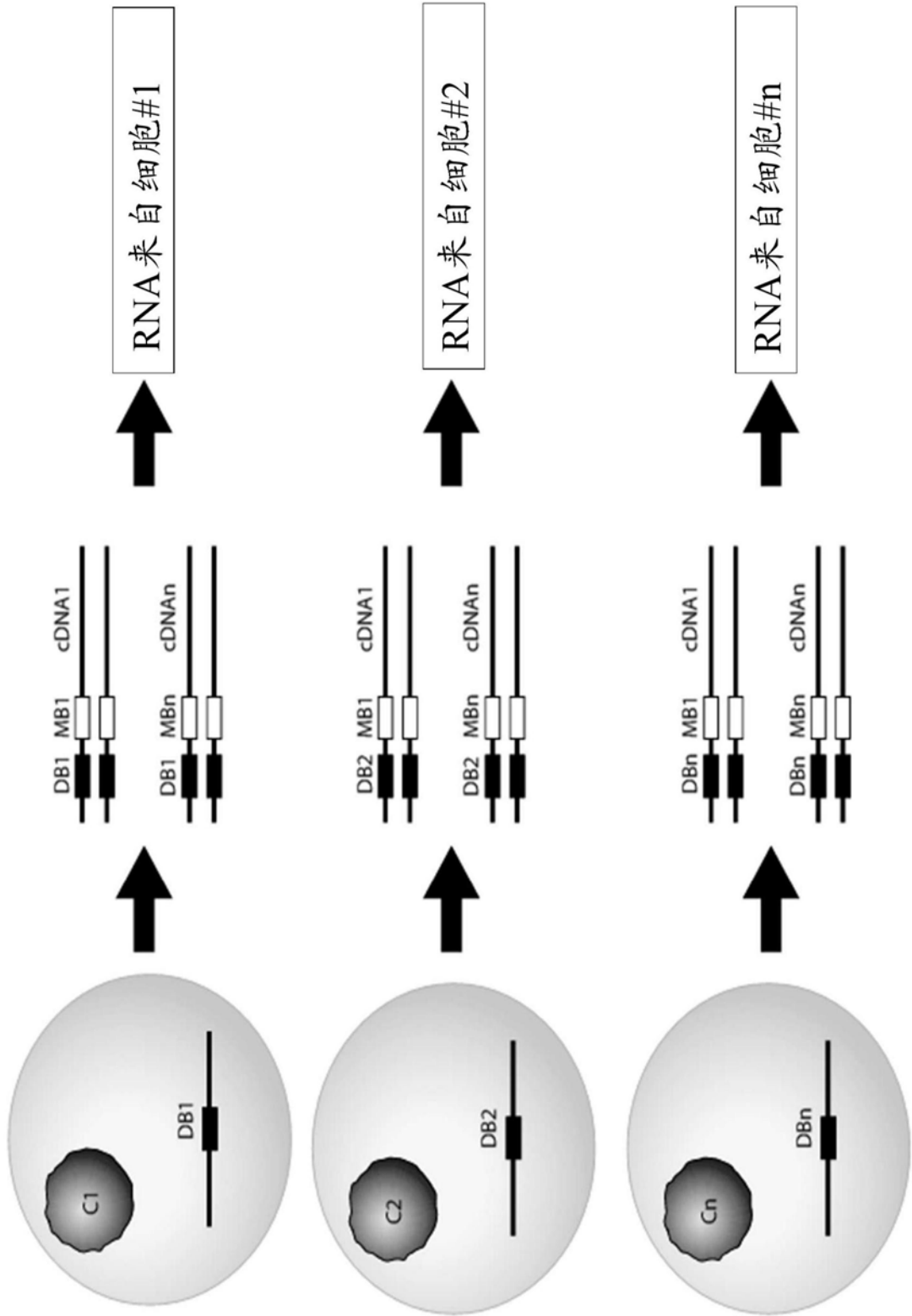


图2

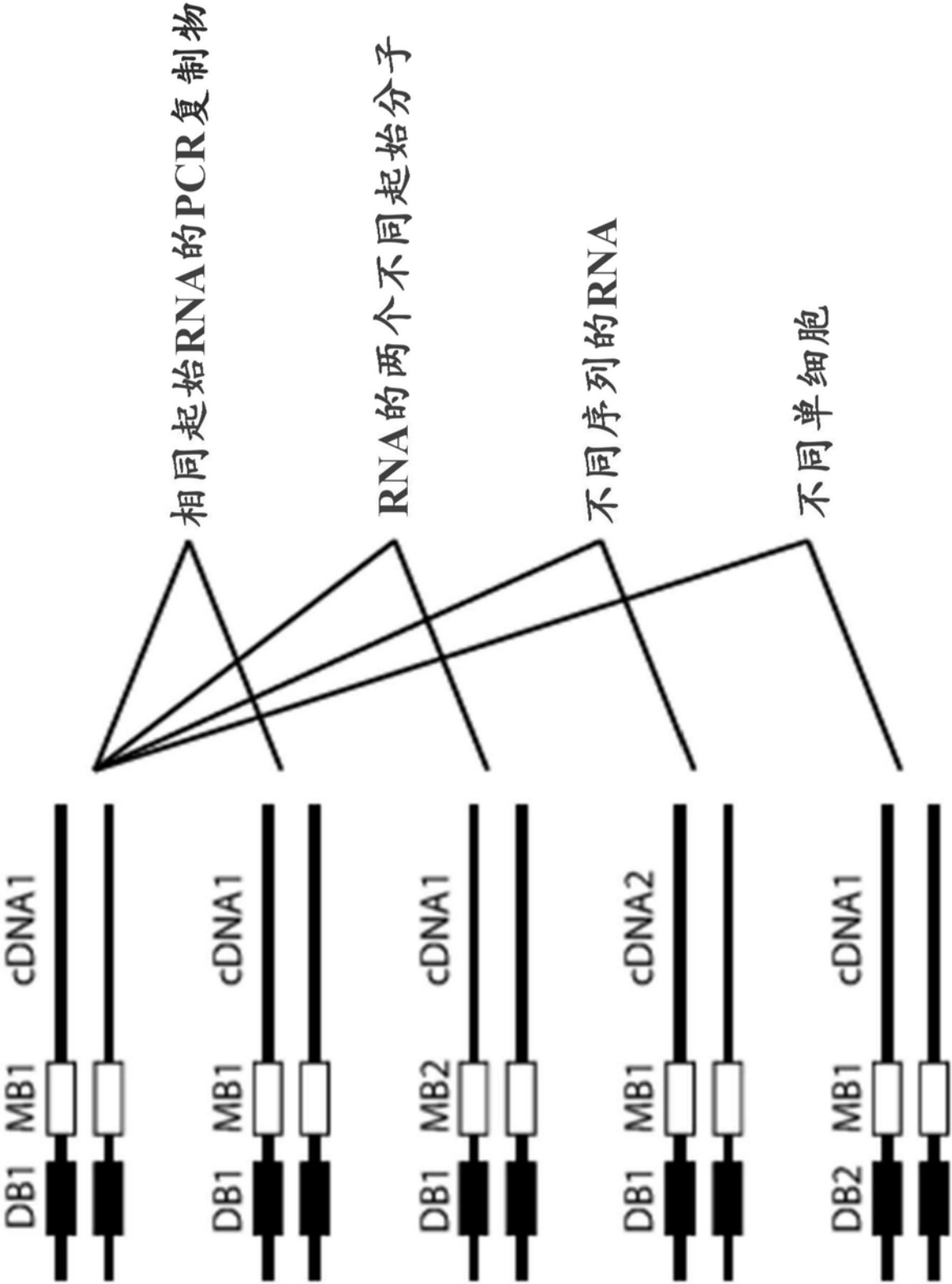


图3

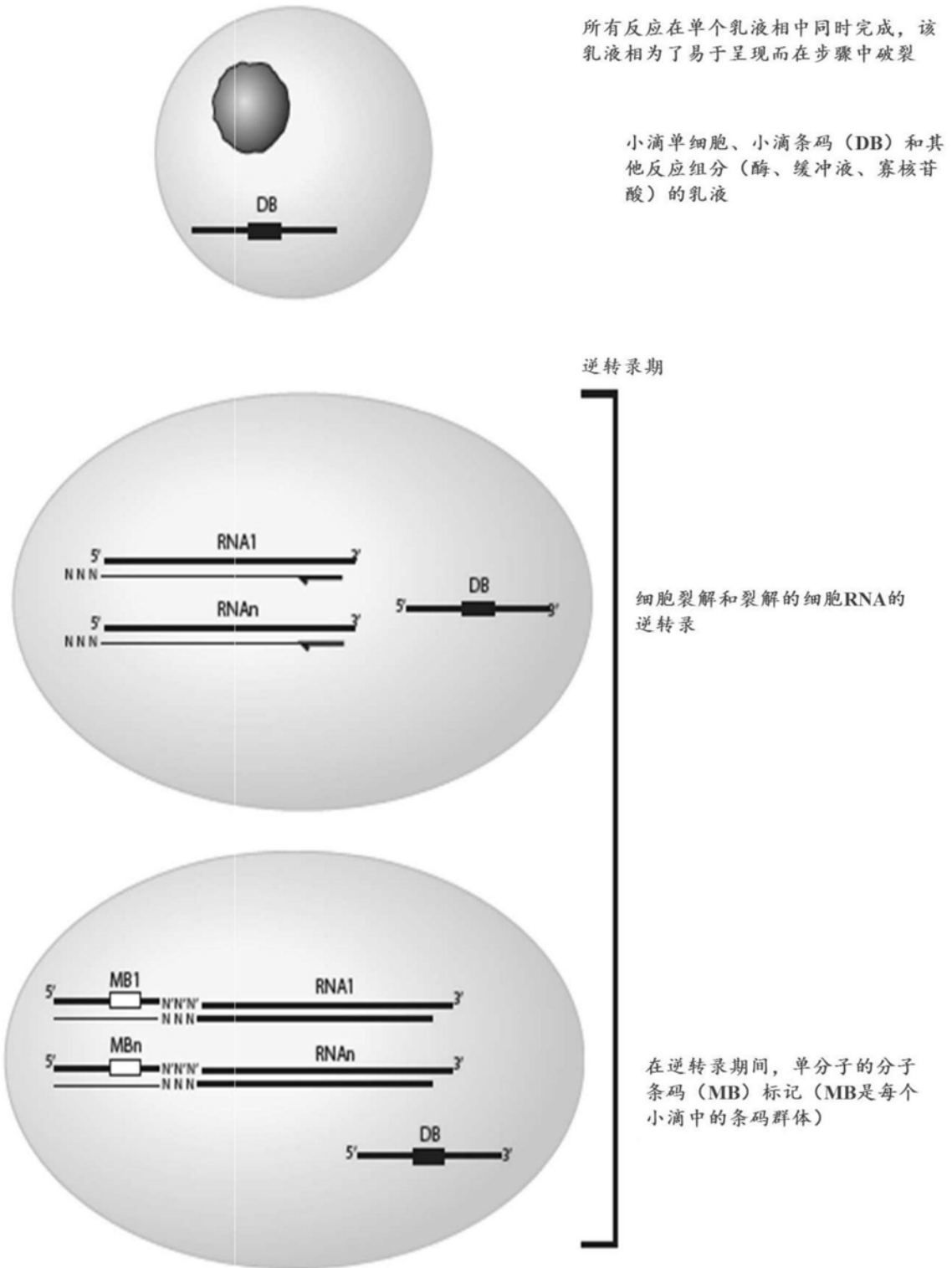


图4A

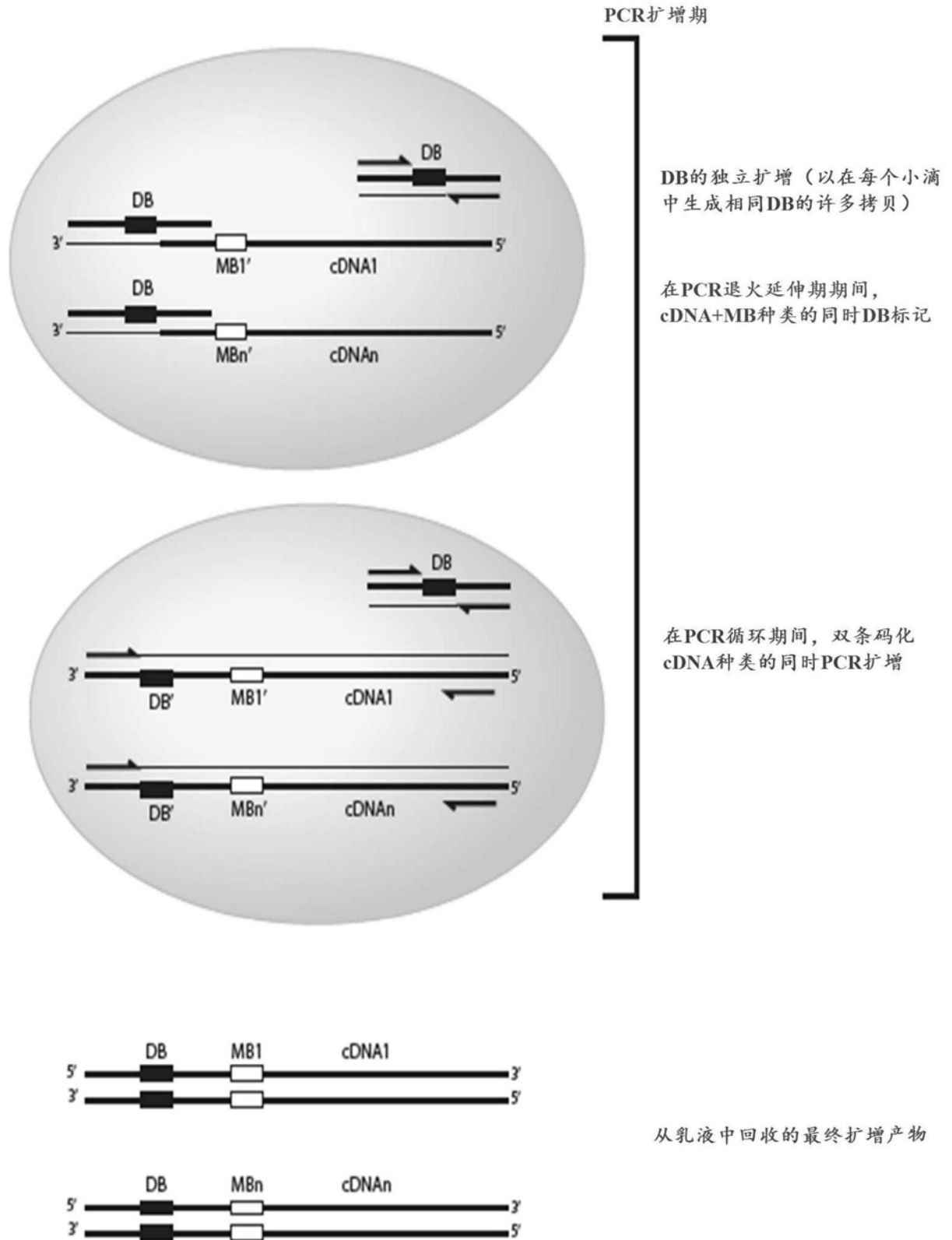


图4B

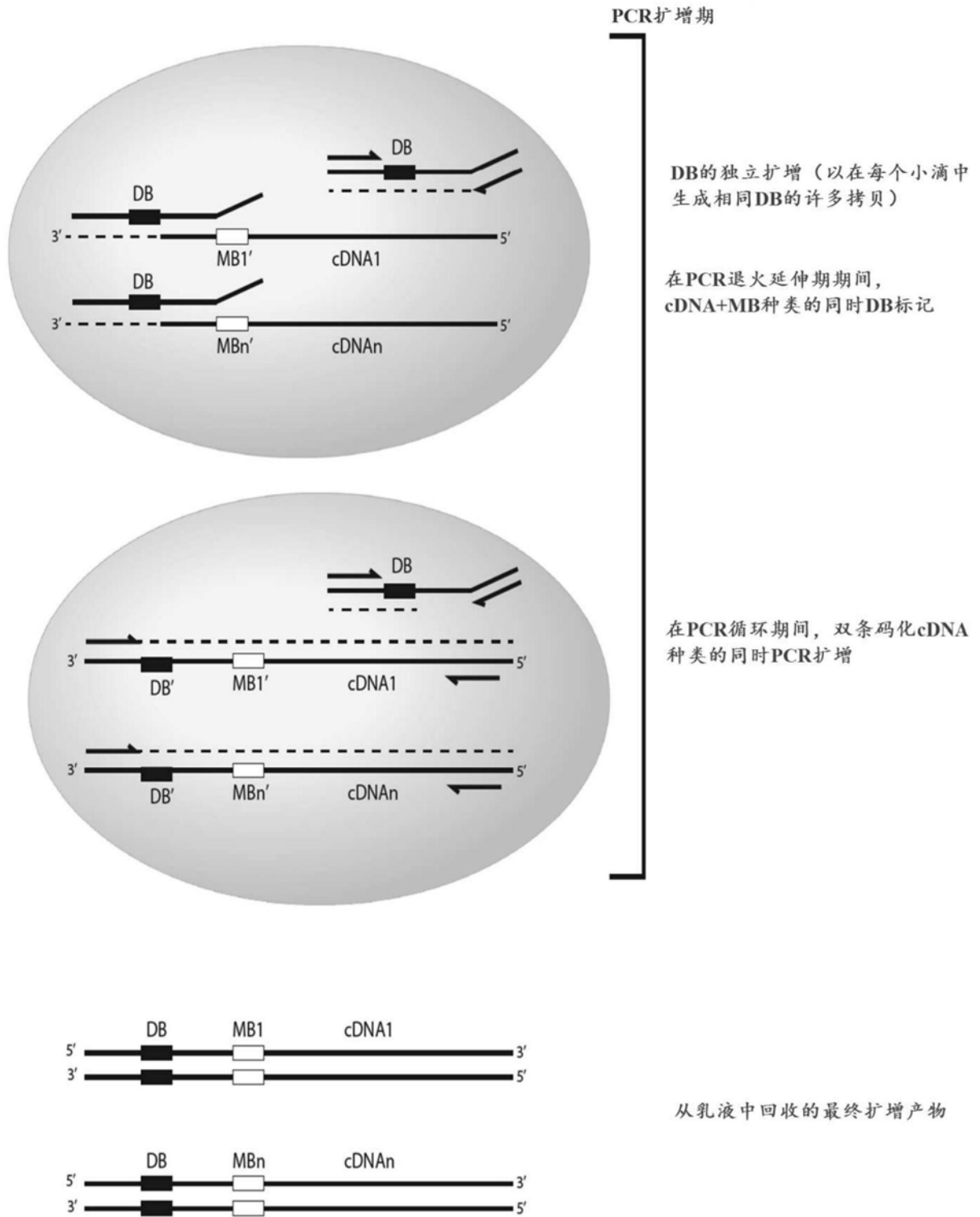


图4C

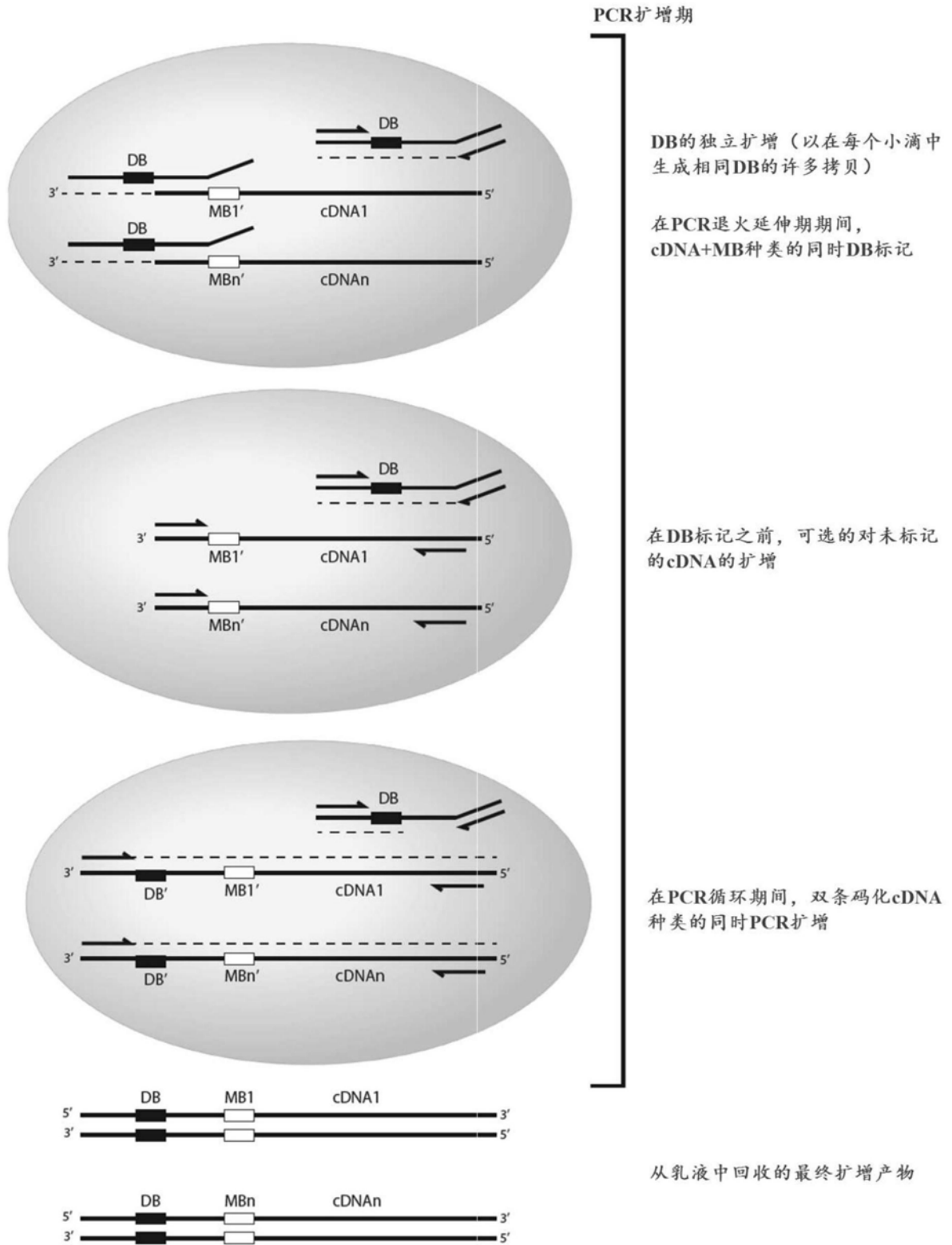


图4D

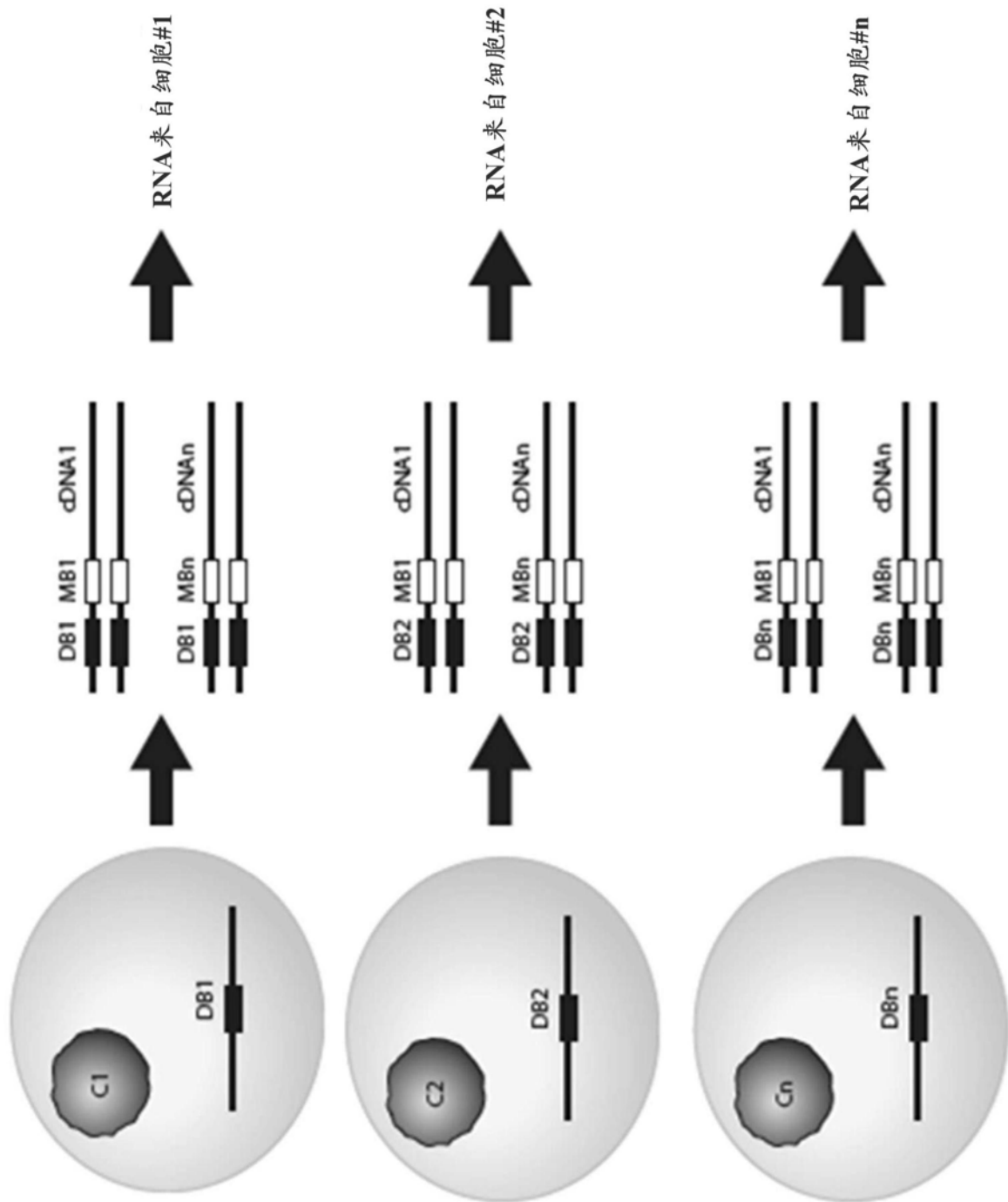


图5

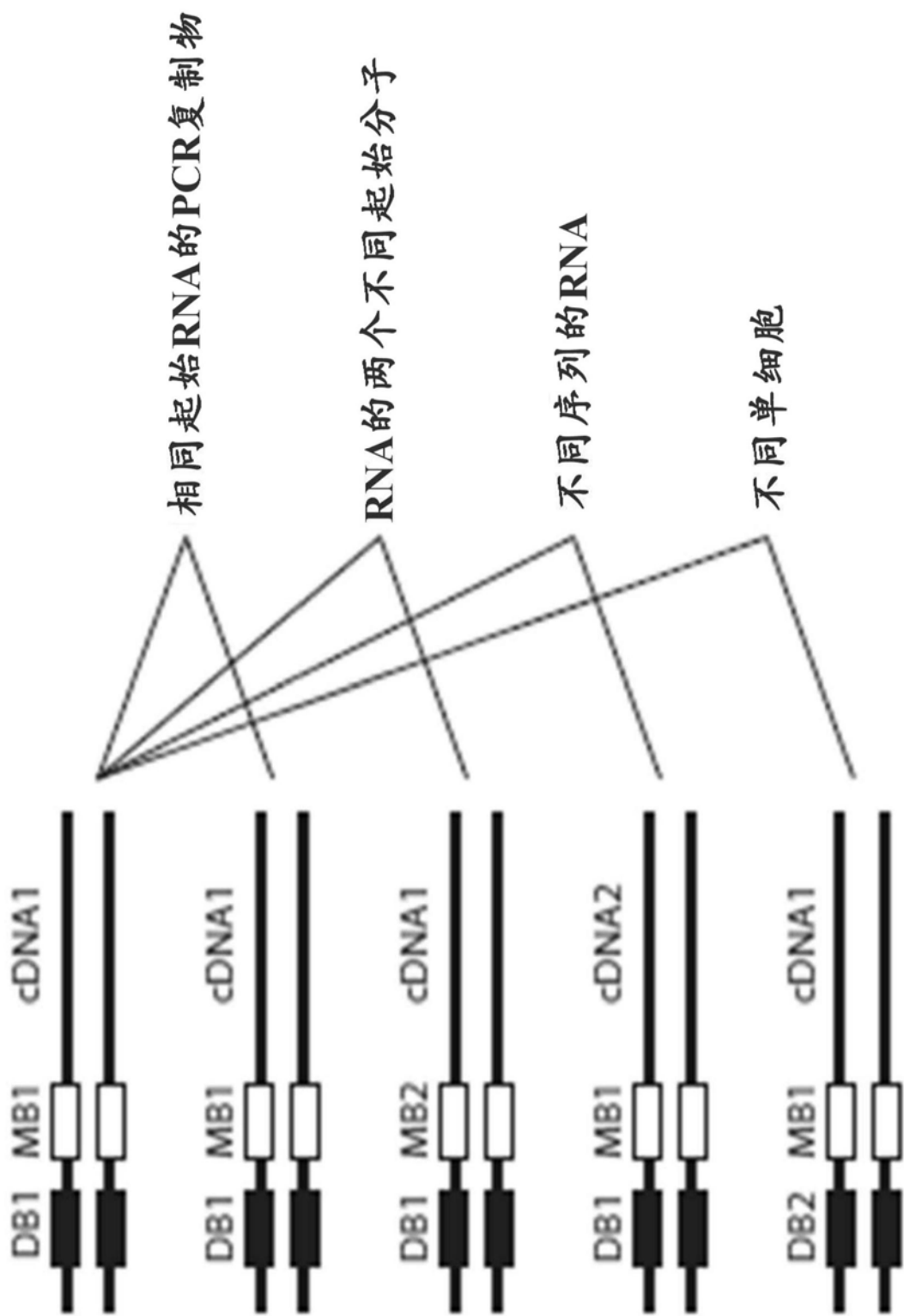


图6

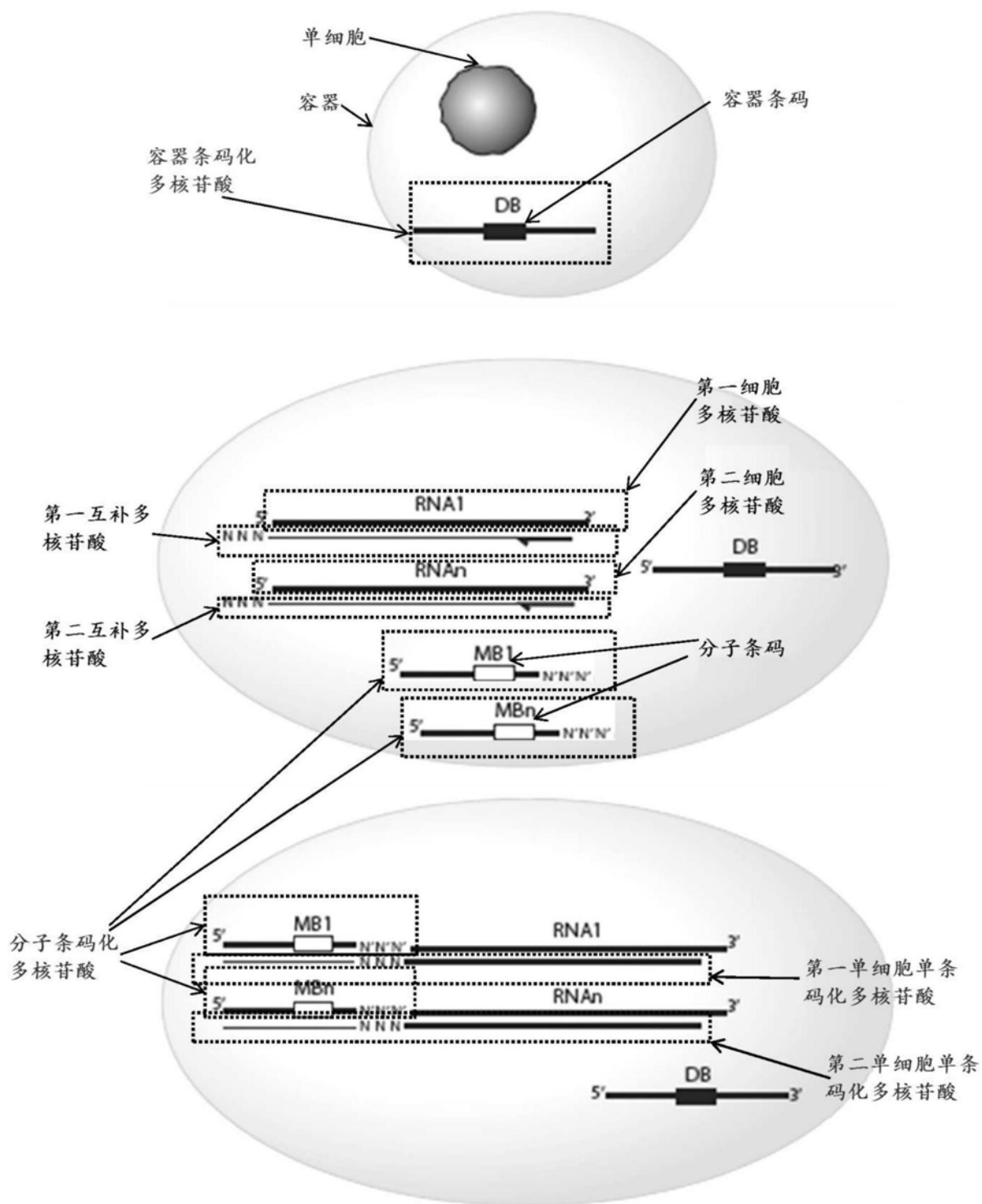


图7A

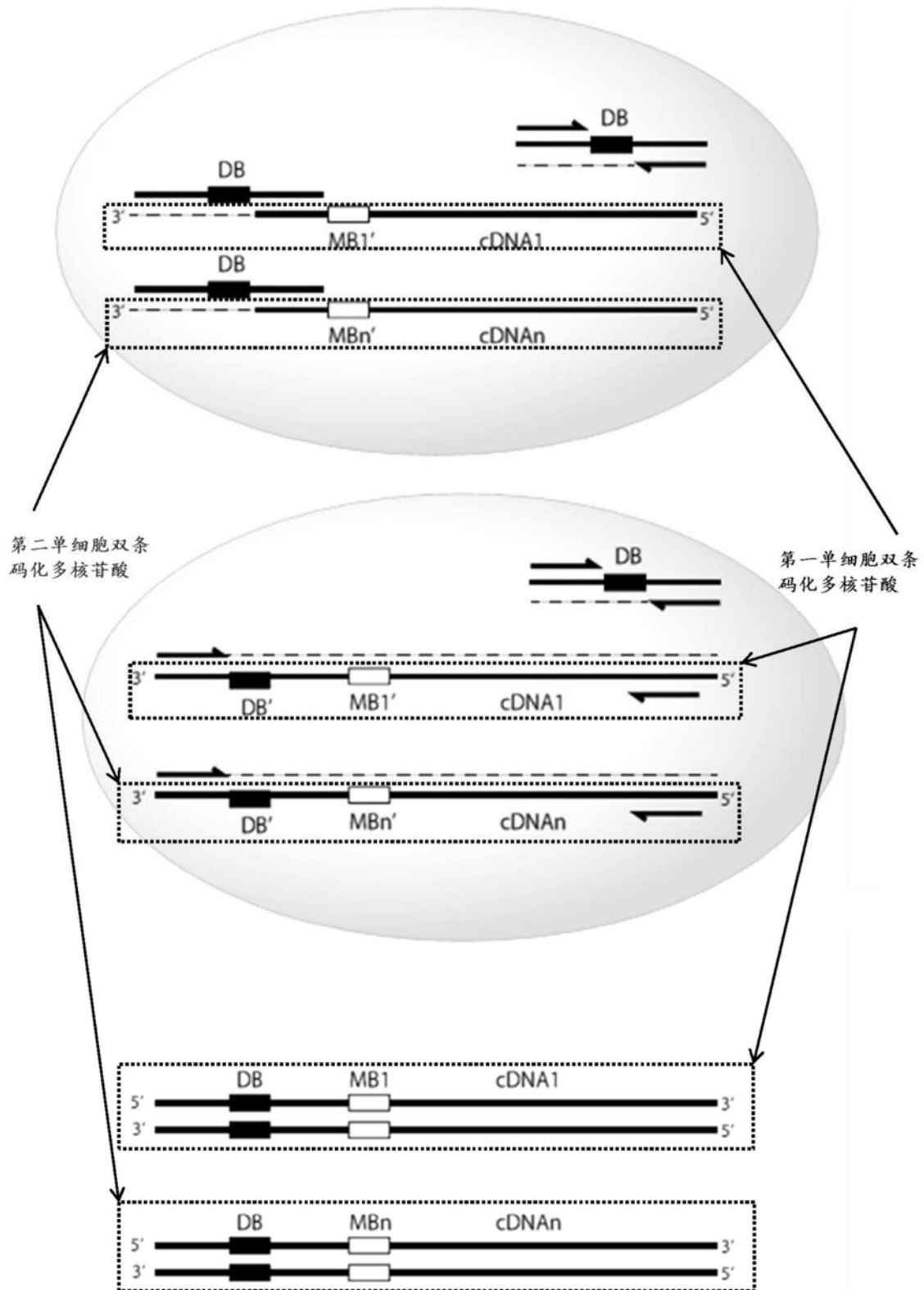


图7B