



(12) 发明专利



(10) 授权公告号 CN 107849560 B

(45) 授权公告日 2021.10.22

(21) 申请号 201680039185.4

(22) 申请日 2016.04.27

(65) 同一申请的已公布的文献号  
申请公布号 CN 107849560 A

(43) 申请公布日 2018.03.27

(30) 优先权数据  
62/153,041 2015.04.27 US

(85) PCT国际申请进入国家阶段日  
2017.12.27

(86) PCT国际申请的申请数据  
PCT/US2016/029556 2016.04.27

(87) PCT国际申请的公布数据  
W02016/176322 EN 2016.11.03

(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)  
C12N 15/85 (2006.01)  
C12Q 1/6869 (2018.01)

(56) 对比文件  
CN 103732738 A, 2014.04.16  
WO 2014144495 A1, 2014.09.18  
WO 2014145992 A1, 2014.09.18  
CN 103703143 A, 2014.04.02  
余馨等.高通量DNA测序技术在抗体新药研发中的应用.《药学学报》.2012,  
George Georgiou等.The promise and challenge of high-throughput sequencing of the antibody repertoire.《Nature Biotechnology》.2014,

审查员 赵鹏

权利要求书13页 说明书87页  
序列表14页 附图26页

(54) 发明名称

对治疗剂和疾病特异性抗原进行测序、确定、配对和验证的方法

(57) 摘要

本文提供了用于免疫组库测序和单个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个细胞的任何两个序列,诸如重链和轻链抗体序列,以供抗体发现、疾病和免疫诊断以及低错误测序。

1. 一种用于对肿瘤浸润性淋巴细胞 (TIL) 测序, 并从所述 TIL 细胞中选择免疫球蛋白 (Ig) 多核苷酸序列或 T 细胞受体 (TCR) 多核苷酸序列的非诊断目的的方法, 其包括:

(a) 形成多个包含以下的容器:

(i) 来自第一受试者的生物样品的单个 TIL 细胞或非 TIL 细胞, 其中所述 TIL 细胞和非 TIL 细胞不在测序之前进行分选;

(ii) 多个分子条码化多核苷酸; 和

(iii) 容器条码化多核苷酸;

(b) 产生第一互补多核苷酸, 所述第一互补多核苷酸与来自所述单个细胞的编码 Ig 多肽或 TCR 多肽的第一多核苷酸互补, 并产生第二互补多核苷酸, 所述第二互补多核苷酸与来自所述单个细胞的编码 Ig 多肽或 TCR 多肽的第二多核苷酸互补;

(c) 将所述多个分子条码化多核苷酸中的第一分子条码化多核苷酸附接至第一互补多核苷酸, 并将所述多个分子条码化多核苷酸中的第二分子条码化多核苷酸附接至第二互补多核苷酸, 从而形成第一 TIL 或非 TIL 单条码化多核苷酸和第二 TIL 或非 TIL 单条码化多核苷酸;

(d) 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 TIL 或非 TIL 单条码化多核苷酸和所述第二 TIL 或非 TIL 单条码化多核苷酸, 从而形成第一和第二双条码化序列的文库;

(e) 对以下测序:

来自所述 TIL 细胞或非 TIL 细胞的所述编码 Ig 多肽或 TCR 多肽的多核苷酸、被附接的分子条码化多核苷酸和所述容器条码化多核苷酸或所述容器条码化多核苷酸的扩增产物, 从而获得序列信息; 和

(f) 基于所述序列信息从所述 TIL 细胞中选择 Ig 多核苷酸序列或 TCR 多核苷酸序列, 其中选择所述 Ig 多核苷酸序列或 TCR 多核苷酸序列基于以下:

(i) 所述 Ig 多核苷酸或 TCR 多核苷酸的表达水平,

(ii) 所述 Ig 多核苷酸或 TCR 多核苷酸的种序列的突变模式或突变水平,

(iii) 所述 Ig 多核苷酸或 TCR 多核苷酸在所述序列信息中的存在, 以及所述 Ig 多核苷酸或 TCR 多核苷酸在来自正常细胞的一组序列信息中的不存在,

(iv) 所述 Ig 多核苷酸或 TCR 多核苷酸在所述序列信息中的富集, 以及所述 Ig 多核苷酸或 TCR 多核苷酸在来自正常细胞的一组序列信息中的不存在,

(v) 所述 Ig 多核苷酸或 TCR 多核苷酸在所述序列信息中的同种型谱;

(vi) 所述序列信息中所述 Ig 多核苷酸或 TCR 多核苷酸的系统发生簇或其大小;

(vii) 所述序列信息中所述 Ig 多核苷酸或 TCR 多核苷酸与来自患病生物样品的一组序列信息的序列之间的相似性; 或

(viii) 所述序列信息中所述 Ig 多核苷酸或 TCR 多核苷酸与来自正常生物样品的一组序列信息的序列之间缺乏相似性。

2. 一种用于对肿瘤浸润性淋巴细胞 (TIL) 测序, 并从所述 TIL 细胞中选择免疫球蛋白 (Ig) 多核苷酸序列或 T 细胞受体 (TCR) 多核苷酸序列的非诊断目的的方法, 其包括:

(a) 形成多个包含以下的容器:

(i) 来自第一受试者的生物样品的单个 TIL 细胞或非 TIL 细胞, 其中所述 TIL 细胞和非

TIL细胞不在测序之前进行分选；

(ii) 多个分子条码化多核苷酸；和

(iii) 容器条码化多核苷酸；

(b) 产生第一互补多核苷酸，所述第一互补多核苷酸与来自所述单个细胞的编码Ig多肽或TCR多肽的第一多核苷酸互补，并产生第二互补多核苷酸，所述第二互补多核苷酸与来自所述单个细胞的编码Ig多肽或TCR多肽的第二多核苷酸互补；

(c) 将所述多个分子条码化多核苷酸中的第一分子条码化多核苷酸附接至第一互补多核苷酸，并将所述多个分子条码化多核苷酸中的第二分子条码化多核苷酸附接至第二互补多核苷酸，从而形成第一TIL或非TIL单条码化的多核苷酸和第二TIL或非TIL单条码化的多核苷酸；

(d) 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TIL或非TIL单条码化的多核苷酸和第二TIL或非TIL单条码化的多核苷酸，从而形成第一和第二双条码化的序列的文库；

(e) 对以下测序：

来自所述TIL细胞或非TIL细胞的所述编码Ig多肽或TCR多肽的多核苷酸、被附接的分子条码化多核苷酸和所述容器条码化多核苷酸或所述容器条码化多核苷酸的扩增产物，

从而获得序列信息；

(f) 将所述序列信息与从来自所述第一受试者的相应的正常邻近组织样品获得的序列信息进行比较；以及

(g) 基于所述比较从所述TIL细胞中选择Ig多核苷酸序列或TCR多核苷酸序列，

其中选择所述Ig多核苷酸序列或所述TCR多核苷酸序列基于以下：

(i) 所述Ig多核苷酸或所述TCR多核苷酸的表达水平，

(ii) 所述Ig多核苷酸或所述TCR多核苷酸的种系序列的突变模式或突变水平，

(iii) 所述Ig多核苷酸或所述TCR多核苷酸在所述序列信息中的存在，以及所述Ig多核苷酸或所述TCR多核苷酸在来自正常细胞的一组序列信息中的不存在，

(iv) 所述Ig多核苷酸或所述TCR多核苷酸在所述序列信息中的富集，以及所述Ig多核苷酸或所述TCR多核苷酸在来自正常细胞的一组序列信息中的不存在，

(v) 所述Ig多核苷酸或所述TCR多核苷酸在所述序列信息中的同种型谱；

(vi) 所述序列信息中所述Ig多核苷酸或所述TCR多核苷酸的系统发生簇或其大小；

(vii) 所述序列信息中所述Ig多核苷酸或所述TCR多核苷酸与来自患病生物样品的一组序列信息的序列之间的相似性；或

(viii) 所述序列信息中所述Ig多核苷酸或所述TCR多核苷酸与来自正常生物样品的一组序列信息的序列之间缺乏相似性。

3. 一种用于对肿瘤浸润性淋巴细胞(TIL)测序，并从所述TIL细胞中选择免疫球蛋白(Ig)多核苷酸序列或T细胞受体(TCR)多核苷酸序列的非诊断目的的方法，其包括：

(a) 形成多个包含以下的容器：

(i) 来自第一受试者的生物样品的单个TIL细胞或非TIL细胞，其中所述TIL细胞和非TIL细胞不在测序之前进行分选；

(ii) 多个分子条码化多核苷酸；和

(iii) 容器条码化多核苷酸;

(b) 产生第一互补多核苷酸,所述第一互补多核苷酸与来自所述单个细胞的编码Ig多肽或TCR多肽的第一多核苷酸互补,并产生第二互补多核苷酸,所述第二互补多核苷酸与来自所述单个细胞的编码Ig多肽或TCR多肽的第二多核苷酸互补;

(c) 将所述多个分子条码化多核苷酸中的第一分子条码化多核苷酸附接至第一互补多核苷酸,并将所述多个分子条码化多核苷酸中的第二分子条码化多核苷酸附接至第二互补多核苷酸,从而形成第一TIL或非TIL单条码化的多核苷酸和第二TIL或非TIL单条码化的多核苷酸;

(d) 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TIL或非TIL单条码化的多核苷酸和所述第二TIL或非TIL单条码化的多核苷酸,从而形成第一和第二双条码化的序列的文库;

(e) 对以下测序:

来自所述TIL细胞或非TIL细胞的所述编码Ig多肽或TCR多肽的多核苷酸、被附接的分子条码化多核苷酸和所述容器条码化多核苷酸或所述容器条码化多核苷酸的扩增产物,

从而获得序列信息;

(f) 将所述序列信息与从来自第二受试者的生物样品获得的序列信息进行比较,其中所述第一和第二受试者患有相同的疾病;

(g) 基于所述比较从所述TIL细胞中选择Ig多核苷酸序列或TCR多核苷酸序列,

其中选择所述Ig多核苷酸序列或所述TCR多核苷酸序列基于以下:

(i) 所述Ig多核苷酸或所述TCR多核苷酸的表达水平,

(ii) 所述Ig多核苷酸或所述TCR多核苷酸的种系序列的突变模式或突变水平,

(iii) 所述Ig多核苷酸或所述TCR多核苷酸在所述序列信息中的存在,以及所述Ig多核苷酸或所述TCR多核苷酸在来自正常细胞的一组序列信息中的不存在,

(iv) 所述Ig多核苷酸或所述TCR多核苷酸在所述序列信息中的富集,以及所述Ig多核苷酸或所述TCR多核苷酸在来自正常细胞的一组序列信息中的不存在,

(v) 所述Ig多核苷酸或所述TCR多核苷酸在所述序列信息中的同种型谱;

(vi) 所述序列信息中所述Ig多核苷酸或所述TCR多核苷酸的系统发生簇或其大小;

(vii) 所述序列信息中所述Ig多核苷酸或所述TCR多核苷酸与来自患病生物样品的一组序列信息的序列之间的相似性;或

(viii) 所述序列信息中所述Ig多核苷酸或所述TCR多核苷酸与来自正常生物样品的一组序列信息的序列之间缺乏相似性。

4. 根据权利要求1-3中任一项所述的方法,其中所述第一受试者的生物样品为组织样品。

5. 根据权利要求1-4中任一项所述的方法,其进一步包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽。

6. 根据权利要求5所述的方法,其进一步包括鉴别所产生的Ig多肽或TCR多肽的靶抗原。

7. 根据权利要求4所述的方法,其中所述方法包括提供组织样品。

8. 根据权利要求1-7中任一项所述的方法,其中所述TIL细胞的形态是未知的。

9. 根据权利要求1-8中任一项所述的方法,其中所述非TIL细胞的形态是未知的。
10. 根据权利要求1-9中任一项所述的方法,其中所述TIL细胞的表型是未知的。
11. 根据权利要求1-10中任一项所述的方法,其中所述非TIL细胞的表型是未知的。
12. 根据权利要求1-11中任一项所述的方法,其中所述TIL细胞和所述非TIL细胞以1:10,000或更小的比例存在于所述第一受试者的生物样品中。
13. 根据权利要求1-11中任一项所述的方法,其中所述TIL细胞和所述非TIL细胞以1:100,000或更小的比例存在于所述第一受试者的生物样品中。
14. 根据权利要求1-11中任一项所述的方法,其中所述TIL细胞和所述非TIL细胞以1:1,000,000或更小的比例存在于所述第一受试者的生物样品中。
15. 根据权利要求1-14中任一项所述的方法,其中来自所述TIL细胞的所述编码Ig多肽或TCR多肽的多核苷酸和来自所述非TIL细胞的所述编码Ig多肽或TCR多肽的多核苷酸在所述测序期间以1:10,000或更小的比例存在。
16. 根据权利要求1-14中任一项所述的方法,其中来自所述TIL细胞的所述编码Ig多肽或TCR多肽的多核苷酸和来自所述非TIL细胞的编码所述Ig多肽或TCR多肽的多核苷酸在所述测序期间以1:100,000或更小的比例存在。
17. 根据权利要求1-14中任一项所述的方法,其中来自所述TIL细胞的所述编码Ig多肽或TCR多肽的多核苷酸和来自所述非TIL细胞的所述编码Ig多肽或TCR多肽的多核苷酸在所述测序期间以1:1,000,000或更小的比例存在。
18. 根据权利要求1-17中任一项所述的方法,其中所述选择包括进行所述序列信息的生物信息学分析。
19. 根据权利要求1-18中任一项所述的方法,其中所述选择包括确定所述序列信息的多核苷酸的表达水平。
20. 根据权利要求1-19中任一项所述的方法,其中所述选择包括将所述序列信息的多核苷酸序列进行比对。
21. 根据权利要求1或权利要求3-20中任一项所述的方法,其中来自所述第一受试者的所述正常生物样品为正常邻近组织样品。
22. 根据权利要求1-21中任一项所述的方法,其中所述方法包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并确定所产生的Ig多肽或TCR多肽对患病生物样品或患病细胞的特异性。
23. 根据权利要求22所述的方法,其中所述确定特异性包括确定所产生的Ig多肽或TCR多肽对患病生物样品或患病细胞的亲和力以及所产生的Ig多肽或TCR多肽对相应的正常邻近组织或相应的正常生物样品或相同组织类型的相应正常细胞的亲和力。
24. 根据权利要求22或权利要求23所述的方法,其中所述方法包括鉴别杀死患病细胞的所产生的Ig多肽或TCR多肽。
25. 根据权利要求24所述的方法,其中所述鉴别的产生的Ig多肽或TCR多肽通过直接与所述患病细胞结合杀死所述患病细胞。
26. 根据权利要求22-25中任一项所述的方法,其中所述产生包括合成或重组表达所述Ig多肽或所述TCR多肽。
27. 根据权利要求1-26中任一项所述的方法,其中所述非TIL细胞包括上皮细胞、淋巴

细胞、癌细胞或其组合。

28. 根据权利要求1-26中任一项所述的方法, 其中所述TIL细胞包括T细胞、B细胞或其组合。

29. 根据权利要求1-28中任一项所述的方法, 其中所述第一受试者的生物样品为癌症活检物。

30. 根据权利要求1-29中任一项所述的方法, 其中所述第一受试者的生物样品包括血管外组织。

31. 根据权利要求1-30中任一项所述的方法, 其中所选择的Ig多核苷酸序列或TCR多核苷酸序列所编码的多肽结合的目标分析物对患病生物样品或患病组织具有特异性。

32. 根据权利要求1-30中任一项所述的方法, 其中所选择的Ig多核苷酸序列或TCR多核苷酸序列所编码的多肽结合的目标分析物对所述第一受试者的生物样品的患病细胞具有特异性。

33. 根据权利要求1-30中任一项所述的方法, 其中所选择的Ig多核苷酸序列或TCR多核苷酸序列所编码的多肽结合的目标分析物对癌细胞具有特异性。

34. 根据权利要求1-33中任一项所述的方法, 其中所述测序为高通量测序。

35. 根据权利要求1-34中任一项所述的方法, 其中所述测序为合成测序、杂交测序或连接测序。

36. 根据权利要求1-35中任一项所述的方法, 其中所述测序不包括对整个免疫组库进行测序。

37. 根据权利要求1-36中任一项所述的方法, 其中所述测序为大规模平行测序。

38. 根据权利要求1-37中任一项所述的方法, 其中所述方法不包括使用多重引物或附接至固体支持体的多重引物。

39. 根据权利要求1-38中任一项所述的方法, 其中所述方法不使用包含与Ig可变域区或TCR可变域区互补的序列的多重引物。

40. 根据权利要求1-39中任一项所述的方法, 其中所述方法在所述测序之前不使用从所述TIL细胞或所述非TIL细胞分离多核苷酸的步骤。

41. 根据权利要求1-40中任一项所述的方法, 其中所述第一受试者的生物样品不是血液。

42. 根据权利要求1-41中任一项所述的方法, 其中所述第一受试者的生物样品为固体组织样品。

43. 根据权利要求1-42中任一项所述的方法, 其中所述第一受试者的生物样品来自器官。

44. 根据权利要求1-43中任一项所述的方法, 其中所述第一受试者的生物样品包括三维结构。

45. 根据权利要求1-44中任一项所述的方法, 其中所述第一受试者的生物样品包括癌细胞或癌前细胞。

46. 根据权利要求1-45中任一项所述的方法, 其中所述第一受试者的生物样品包括被所述第一受试者的免疫系统异常靶向的健康组织。

47. 根据权利要求1-46中任一项所述的方法, 其中所述多个容器包括1,000个或更多个

非TIL细胞。

48. 根据权利要求1-47中任一项所述的方法,其中所述多个容器包括1,000个或更多个TIL细胞。

49. 根据权利要求1-48中任一项所述的方法,其中所选择的Ig多核苷酸序列或TCR多核苷酸序列包括来自T细胞的TCR多核苷酸序列。

50. 根据权利要求1-48中任一项所述的方法,其中所选择的Ig多核苷酸序列或TCR多核苷酸序列包括来自B细胞的Ig多核苷酸序列。

51. 根据权利要求1-50中任一项所述的方法,其中所述测序的多核苷酸包括编码Ig重链(IgH)的Ig多核苷酸。

52. 根据权利要求51所述的方法,其中所述方法进一步包括将所述IgH与来自相同B细胞的Ig轻链(IgL)进行配对。

53. 根据权利要求1-50中任一项所述的方法,其中所述测序的多核苷酸包括编码IgL的Ig多核苷酸。

54. 根据权利要求53所述的方法,其中所述方法进一步包括将所述IgL与来自相同B细胞的IgH进行配对。

55. 根据权利要求1-50中任一项所述的方法,其中所述测序的多核苷酸包括编码IgH的Ig多核苷酸和编码IgL的Ig多核苷酸。

56. 根据权利要求55所述的方法,其中所述IgL与来自相同B细胞的IgH配对。

57. 根据权利要求1-50中任一项所述的方法,其中所述测序的多核苷酸包括编码TCR $\alpha$ 链的TCR多核苷酸。

58. 根据权利要求57所述的方法,其中所述方法进一步包括将所述TCR $\alpha$ 链与来自相同T细胞的TCR $\beta$ 链进行配对。

59. 根据权利要求1-50中任一项所述的方法,其中所述测序的多核苷酸包括编码TCR $\beta$ 链的TCR多核苷酸。

60. 根据权利要求59所述的方法,其中所述方法进一步包括将所述TCR $\beta$ 链与来自相同T细胞的TCR $\alpha$ 链进行配对。

61. 根据权利要求1-50中任一项所述的方法,其中所述测序的多核苷酸包括编码TCR $\alpha$ 链的TCR多核苷酸和编码TCR $\beta$ 链的TCR多核苷酸。

62. 根据权利要求61所述的方法,其中所述TCR $\alpha$ 链与来自相同T细胞的TCR $\beta$ 链配对。

63. 根据权利要求52、权利要求54或权利要求56所述的方法,其中所述方法进一步包括生成配对的IgL和IgH的数据库。

64. 根据权利要求58、权利要求60或权利要求62所述的方法,其中所述方法进一步包括生成配对的TCR $\alpha$ 和TCR $\beta$ 链的数据库。

65. 根据权利要求1-50中任一项所述的方法,其中所述测序的多核苷酸包括编码TCR $\gamma$ 链的TCR多核苷酸。

66. 根据权利要求65所述的方法,其中所述方法进一步包括将所述TCR $\gamma$ 链与来自相同T细胞的TCR $\delta$ 链进行配对。

67. 根据权利要求1-50中任一项所述的方法,其中所述测序的多核苷酸包括编码TCR $\delta$ 链的TCR多核苷酸。

68. 根据权利要求67所述的方法, 其中所述方法进一步包括将所述TCR $\delta$ 链与来自相同T细胞的TCR  $\gamma$  链进行配对。

69. 根据权利要求1-50中任一项所述的方法, 其中所述测序的多核苷酸包括编码TCR  $\gamma$  链的TCR多核苷酸和编码TCR $\delta$ 链的TCR多核苷酸。

70. 根据权利要求69所述的方法, 其中所述TCR  $\gamma$  链与来自相同T细胞的TCR $\delta$ 链配对。

71. 根据权利要求1-50中任一项所述的方法, 其中所述编码Ig多肽或TCR多肽的多核苷酸包含可变区。

72. 根据权利要求71所述的方法, 其中所述编码Ig多肽的多核苷酸包含重链可变区( $V_H$ )。

73. 根据权利要求71或权利要求72所述的方法, 其中所述编码Ig多肽的多核苷酸包含轻链可变区( $V_L$ )。

74. 根据权利要求71所述的方法, 其中所述编码TCR多肽的多核苷酸包含TCR $\alpha$ 链可变区( $V\alpha$ )。

75. 根据权利要求71或权利要求74所述的方法, 其中所述编码TCR多肽的多核苷酸包含TCR $\beta$ 链可变区( $V\beta$ )。

76. 根据权利要求71所述的方法, 其中所述编码TCR多肽的多核苷酸包含TCR  $\gamma$  链可变区( $V\gamma$ )。

77. 根据权利要求71或76所述的方法, 其中所述编码TCR多肽的多核苷酸包含TCR $\delta$ 链可变区( $V\delta$ )。

78. 根据权利要求1-77中任一项所述的方法, 其中所述编码Ig多肽或TCR多肽的多核苷酸包含TCR恒定域区。

79. 根据权利要求78所述的方法, 其中所述TCR恒定域区包含TCR $\alpha$ 恒定域、TCR $\beta$ 恒定域、TCR  $\gamma$  恒定域、TCR $\delta$ 恒定域或其组合。

80. 根据权利要求1-77中任一项所述的方法, 其中所述编码Ig多肽或TCR多肽的多核苷酸包括Ig恒定域区。

81. 根据权利要求80所述的方法, 其中所述Ig恒定域区包含选自 $C_H1$ 、 $C_H2$ 、 $C_H3$ 和 $C_H4$ 的IgH恒定域。

82. 根据权利要求81所述的方法, 其中所述Ig恒定域区包含选自 $C_H1$ 、 $C_H2$ 、 $C_H3$ 和 $C_H4$ 的两个、三个或四个IgH恒定域。

83. 根据权利要求80-82中任一项所述的方法, 其中所述Ig恒定域区包含来自选自IgM、IgD、IgA、IgE、IgG及其组合的Ig同种型的IgH恒定域。

84. 根据权利要求80所述的方法, 其中所述Ig恒定域区包含IgL恒定域( $C_L$ )。

85. 根据权利要求84所述的方法, 其中所述Ig恒定域区包含来自选自Ig $\kappa$ 、Ig $\lambda$ 及其组合的IgL同种型的 $C_L$ 。

86. 根据权利要求85所述的方法, 其中所述IgL同种型为Ig $\kappa$ 。

87. 根据权利要求1-86中任一项所述的方法, 其中所述编码Ig多肽或TCR多肽的多核苷酸包含含有种系构架序列的构架区序列。

88. 根据权利要求1-87中任一项所述的方法, 其中所述编码Ig多肽或TCR多肽的多核苷酸包含含有种系 $V_H$ 序列的 $V_H$ 序列、含有种系 $V_L$ 序列的 $V_L$ 序列、含有种系TCR $\alpha$ 可变区序列的

TCR $\alpha$ 可变区序列、含有种系TCR $\beta$ 可变区序列的TCR $\beta$ 可变区序列、含有种系TCR $\gamma$ 可变区序列的TCR $\gamma$ 可变区序列、含有种系TCR $\delta$ 可变区序列的TCR $\delta$ 可变区序列或其组合。

89. 根据权利要求1-88中任一项所述的方法,其中所述编码Ig多肽或TCR多肽的多核苷酸包含含有一个或多个突变的构架残基构架残基的构架区序列。

90. 根据权利要求89所述的方法,其中所述一个或多个突变的构架残基是在来自两名或更多名患有疾病的受试者的前5%最大表达的IgH、IgL、TCR $\alpha$ 、TCR $\beta$ 、TCR $\gamma$ 或TCR $\delta$ 多核苷酸中的一种或多种中发现的。

91. 根据权利要求1-90中任一项所述的方法,其中由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽包括特异性Ig同种型。

92. 根据权利要求91所述的方法,其中所述特异性Ig同种型为IgA、IgG、IgM、IgD或IgE。

93. 根据权利要求92所述的方法,其中所述特异性同种型为IgG。

94. 根据权利要求1-93中任一项所述的方法,其中所述序列信息包括至少1,000个独特的Ig多核苷酸序列或TCR多核苷酸序列。

95. 根据权利要求1-93中任一项所述的方法,其中所述序列信息包括至少1,000个序列读取。

96. 根据权利要求1-95中任一项所述的方法,其中所述序列信息包括来自所述TIL细胞的至少一个Ig多核苷酸序列或TCR多核苷酸序列,该至少一个Ig多核苷酸序列或TCR多核苷酸序列编码对疾病相关蛋白具有 $1 \times 10^{-7}$ M或更小的 $K_D$ 的Ig多肽或TCR多肽。

97. 根据权利要求1-96中任一项所述的方法,其中所述序列信息不包括来自所述非TIL细胞的Ig多核苷酸序列或TCR多核苷酸序列,该Ig多核苷酸序列或TCR多核苷酸序列编码对疾病相关蛋白具有 $1 \times 10^{-7}$ M或更小的 $K_D$ 的Ig多肽或TCR多肽。

98. 根据权利要求1-97中任一项所述的方法,还包括由所选择的Ig多核苷酸序列或TCR多核苷酸序列产生Ig多肽或TCR多肽,其中所产生的Ig多肽或TCR多肽对疾病相关蛋白具有 $1 \times 10^{-7}$ M或更小的 $K_D$ 。

99. 根据权利要求1-98中任一项所述的方法,其中所选择的多核苷酸序列编码的Ig多肽或TCR多肽对疾病相关蛋白具有 $1 \times 10^{-7}$ M或更小的 $K_D$ 。

100. 根据权利要求1-99中任一项所述的方法,其中所述测序的错误率小于或等于0.01%。

101. 根据权利要求1-100中任一项所述的方法,其中所述测序包括以至少80%的准确度或置信度确定序列。

102. 根据权利要求1-101中任一项所述的方法,其中扩增错误被最小化、消除或小于0.01%。

103. 根据权利要求1-102中任一项所述的方法,其中所述方法在少于或等于4周内进行。

104. 根据权利要求1所述的方法,其中所述第一多核苷酸编码 $V_H$ 或TCR $\alpha$ ,并且所述第二多核苷酸编码 $V_L$ 或TCR $\beta$ ;其中所述产生通过以下进行:

- (A) 第一引物,其包含与编码Ig多肽或TCR多肽的多核苷酸的相同区域互补的区域;
- (B) 第二引物,其包含与编码Ig多肽或TCR多肽的多核苷酸的相同区域互补的区域;和
- (C) 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；

其中所述多个分子条码化多核苷酸的每个分子条码化多核苷酸包含：

(1) 分子条码，

(2) 与容器条码化多核苷酸的区域互补的5'端区域，以及

(3) 与所述3个或更多个非模板核苷酸互补的3'端区域；其中所述附接包括：

在将所述分子条码化多核苷酸附接至所述第一和第二互补多核苷酸之后形成第一和第二单条码化多核苷酸；和扩增所述容器条码化多核苷酸，从而形成第一和第二双条码化多核苷酸；其中所述方法还包括：

扩增所述第一和第二双条码化多核苷酸，从而形成包含所述 $V_H$ 、 $V_L$ 、TCR $\alpha$ 或TCR $\beta$ 多核苷酸的可变区的序列的文库；并且其中所述测序包括对所述文库的一个或多个所述序列进行测序，其中所述文库代表所述样品的免疫状态。

105. 根据权利要求104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。

106. 根据权利要求1-105中任一项所述的方法，其中所述第一受试者为动物。

107. 根据权利要求106所述的方法，其中所述动物为哺乳动物。

108. 根据权利要求107所述的方法，其中所述哺乳动物为人类。

109. 根据权利要求106-108中任一项所述的方法，其中所述第一受试者患有自身免疫病。

110. 根据权利要求106-108中任一项所述的方法，其中所述第一受试者患有癌症。

111. 根据权利要求106-108中任一项所述的方法，其中所述第一受试者患有癌前疾病。

112. 根据权利要求1-111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。

113. 根据权利要求1-112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。

114. 根据权利要求1-113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相同条码序列的序列分箱或分组。

115. 根据权利要求1-114中任一项所述的方法，其中所述方法进一步包括将具有至少90%的序列同源性的序列簇集。

116. 根据权利要求1-115中任一项所述的方法，其中所述方法进一步包括将具有至少90%的序列同源性的序列进行比对。

117. 根据权利要求115所述的方法，其中借助于计算机或算法进行所述簇集。

118. 根据权利要求116所述的方法，其中借助于计算机或算法进行所述比对。

119. 根据权利要求1-118中任一项所述的方法，其中所述方法进一步包括将序列读取与种系序列进行比较，并确定所述序列读取的体细胞超突变积累。

120. 根据权利要求1-119中任一项所述的方法，其中所述方法进一步包括确定所述序列的同种型分布。

121. 根据权利要求1-120中任一项所述的方法，其中由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽不与正常邻近组织的细胞相互作用。

122. 根据权利要求1-121中任一项所述的方法，其中由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽不与来自未患疾病的受试者的相同组织的细胞结合。

123. 根据权利要求1-122中任一项所述的方法, 还包括产生Ig多肽或TCR多肽, 其中所述产生包括表达由所选择的多核苷酸序列编码的重组Ig多肽或TCR多肽。

124. 根据权利要求123所述的方法, 其中所述产生包括表达各自由选择的多核苷酸序列编码的1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、300、400或500个或更多个重组Ig多肽或TCR多肽。

125. 根据权利要求123或权利要求124所述的方法, 其中所述产生包括将所选择的多核苷酸序列中的序列克隆到载体中。

126. 根据权利要求123-125中任一项所述的方法, 其中所述产生包括使细胞与包含编码Ig多肽或TCR多肽的所选择的多核苷酸序列中的序列的多核苷酸接触。

127. 根据权利要求126所述的方法, 其中所述接触包括转染。

128. 根据权利要求123-127中任一项所述的方法, 其中所述产生包括在细胞中表达由所选择的多核苷酸序列编码的重组Ig多肽或TCR多肽。

129. 根据权利要求128所述的方法, 其中所述细胞为哺乳动物细胞。

130. 根据权利要求129所述的方法, 其中所述哺乳动物细胞为中国仓鼠卵巢(CHO)细胞或HEK293细胞。

131. 根据权利要求123-130中任一项所述的方法, 其中所述方法进一步包括纯化所产生的由所选择的多核苷酸序列编码的重组Ig多肽或TCR多肽。

132. 根据权利要求123-131中任一项所述的方法, 其中所述方法进一步包括分离所产生的由所选择的多核苷酸序列编码的重组Ig多肽或TCR多肽。

133. 根据权利要求131或权利要求132所述的方法, 其中由所选择的多核苷酸序列编码的重组Ig多肽或TCR多肽还包含异源标签。

134. 根据权利要求133所述的方法, 其中所述异源标签为纯化标签。

135. 根据权利要求128所述的方法, 其中所述细胞为细菌细胞或昆虫细胞。

136. 根据权利要求1-135中任一项所述的方法, 还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原, 其中所述鉴别包括将所产生的Ig多肽或TCR多肽的序列与包含Ig多肽或TCR多肽序列数据的数据库进行比较。

137. 根据权利要求1-136中任一项所述的方法, 还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原, 其中所述鉴别包括进行全基因组siRNA筛选。

138. 根据权利要求1-137中任一项所述的方法, 还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原, 其中所述鉴别包括采用由所选择的多核苷酸序列编码的Ig多肽或TCR多肽进行蛋白质展示筛选。

139. 根据权利要求138所述的方法, 其中所述蛋白质展示筛选为噬菌体展示筛选。

140. 根据权利要求138所述的方法, 其中所述蛋白质展示筛选为核糖体展示筛选。

141. 根据权利要求6所述的方法, 其中所述鉴别包括进行酵母双杂交筛选。

142. 根据权利要求1-128中任一项所述的方法, 还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原, 其中所产生的Ig多肽或TCR多肽的靶抗原的所述鉴别包括进行2D凝胶电泳。

143. 根据权利要求1-128中任一项所述的方法, 还包括产生由所选择的多核苷酸序列

编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原,其中所述鉴别包括采用蛋白质阵列对由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽进行筛选。

144. 根据权利要求143所述的方法,其中所述蛋白质阵列包含至少10%的人类蛋白质组蛋白质。

145. 根据权利要求1-128中任一项所述的方法,还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原,其中所述鉴别包括针对由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽进行蛋白质组筛选。

146. 根据权利要求1-128中任一项所述的方法,还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原,其中所述鉴别包括采用由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽进行免疫沉淀。

147. 根据权利要求1-128中任一项所述的方法,还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原,其中所述鉴别包括进行质谱分析。

148. 根据权利要求1-128中任一项所述的方法,还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原,其中所述鉴别包括采用由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽进行抗体依赖性细胞介导的细胞毒性(ADCC)测定。

149. 根据权利要求1-128中任一项所述的方法,还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原,其中所述鉴别包括确定由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽的特异性。

150. 根据权利要求1-128中任一项所述的方法,还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原,其中所述鉴别包括进行结合测定。

151. 根据权利要求1-128中任一项所述的方法,还包括产生由所选择的多核苷酸序列编码的Ig多肽或TCR多肽并鉴别所产生的Ig多肽或TCR多肽的靶抗原,其中所述鉴别包括使由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽与目标分析物候选物接触。

152. 根据权利要求151所述的方法,其中所述目标分析物候选物在固体支持体上。

153. 根据权利要求151所述的方法,其中所述目标分析物候选物在溶液中。

154. 根据权利要求151-153中任一项所述的方法,其中由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽在固体支持体上。

155. 根据权利要求151-153中任一项所述的方法,其中由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽在溶液中。

156. 根据权利要求152或权利要求154所述的方法,其中所述固体支持体为阵列。

157. 根据权利要求152或权利要求154所述的方法,其中所述固体支持体为珠子。

158. 根据权利要求1-157中任一项所述的方法,其中在选择所选择的多核苷酸序列时,与由所选择的多核苷酸序列编码的Ig多肽或TCR多肽结合的目标分析物是未知的。

159. 根据权利要求1-158中任一项所述的方法,还包括鉴别由所选择的多核苷酸编码的Ig多肽或TCR多肽的目标分析物。

160. 根据权利要求159所述的方法,其中所述鉴别的目标分析物为疾病相关性目标分

析物。

161. 根据权利要求159或权利要求160所述的方法, 其中所述鉴别的目标分析物为具有细胞外区域的多肽。

162. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列编码的Ig多肽或TCR多肽。

163. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列中的Ig多核苷酸编码的IgL多肽。

164. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列中的Ig多核苷酸编码的IgH多肽。

165. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列中的IgH和IgL多核苷酸编码的抗体。

166. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸编码的Ig多肽的Fab片段。

167. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列编码的Ig多肽的F(ab)<sub>2</sub>片段。

168. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列编码的Ig多肽的Fv片段。

169. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列编码的Ig多肽的scFv片段。

170. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列编码的TCR  $\gamma$  多肽。

171. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列编码的TCR $\delta$ 多肽。

172. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列编码的TCR  $\gamma$  多肽和TCR $\delta$ 多肽。

173. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列编码的TCR $\alpha$ 多肽。

174. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列编码的TCR $\beta$ 多肽。

175. 根据权利要求1-158中任一项所述的方法, 还包括分离和纯化由所选择的多核苷酸序列编码的TCR $\alpha$ 多肽和TCR $\beta$ 多肽。

176. 根据权利要求1-158中任一项所述的方法, 还包括产生细胞, 所述细胞经遗传修饰以表达嵌合抗原受体(CAR), 该嵌合抗原受体包含抗原结合域、跨膜结构域和细胞内信号传导结构域, 其中所述抗原结合域包括由所选择的Ig多核苷酸序列或TCR多核苷酸序列编码的抗原结合域。

177. 根据权利要求1-158中任一项所述的方法, 还包括产生细胞, 所述细胞经遗传修饰以表达嵌合抗原受体(CAR), 该嵌合抗原受体包含抗原结合域、跨膜结构域和细胞内信号传导结构域, 其中所述抗原结合域识别所产生的Ig多肽或TCR多肽的鉴别的靶抗原的表位, 所产生的Ig多肽或TCR多肽由所选择的多核苷酸序列编码。

178. 根据权利要求1-158中任一项所述的方法,还包括产生细胞,所述细胞经遗传修饰以表达由所选择的多核苷酸编码的Ig多肽或TCR多肽的抗原结合域。

179. 根据权利要求1-158中任一项所述的方法,还包括产生细胞,所述细胞经遗传修饰以表达所选择的Ig多肽或TCR多肽。

180. 根据权利要求1-158中任一项所述的方法,还包括产生细胞,所述细胞经遗传修饰以表达抗原结合域,该抗原结合域识别所产生的Ig多肽或TCR多肽的鉴别的靶抗原的表位,所产生的Ig多肽或TCR多肽由所选择的多核苷酸序列编码。

181. 根据权利要求1-158中任一项所述的方法,还包括产生细胞,所述细胞经遗传修饰以表达抗原结合域,该抗原结合域识别由所选择的多核苷酸序列编码的Ig多肽或TCR多肽的靶抗原。

182. 根据权利要求176-181中任一项所述的方法,其中经遗传修饰的所述细胞不是来自受试者。

183. 根据权利要求176-182中任一项所述的方法,其中经遗传修饰的所述细胞是分离的。

184. 根据权利要求3所述的方法,其中由所选择的多核苷酸序列编码的Ig多肽或TCR多肽的鉴别的目标分析物为所述疾病的生物标志物。

185. 由根据权利要求1-184中任一项所述的方法选择的多核苷酸序列编码的Ig多肽或TCR多肽或其抗原结合片段在制备用于治疗有需要的受试者的药物中的用途。

186. 根据权利要求185所述的用途,其中由所选择的多核苷酸序列编码的所述Ig多肽或所述TCR多肽为人类治疗性多肽。

187. 根据权利要求185或权利要求186所述的用途,其中由所选择的多核苷酸序列编码的Ig多肽或TCR多肽为中和多肽。

188. 由根据权利要求1-184中任一项所述的方法选择的多核苷酸序列编码的Ig多肽或TCR多肽的鉴别的目标分析物的抑制剂或其抗原结合片段在制备用于治疗有需要的受试者的药物中的用途。

189. 根据权利要求188所述的用途,其中所述抑制剂选自小分子、核酸、多肽及其组合。

190. 根据权利要求189所述的用途,其中所述抑制剂为多肽抑制剂,其中所述多肽抑制剂为由所选择的多核苷酸序列编码的Ig多肽或TCR多肽。

191. 根据权利要求189所述的用途,其中所述抑制剂为核酸抑制剂,其中所述核酸抑制剂为siRNA核酸。

192. 根据权利要求189所述的用途,其中所述抑制剂为核酸抑制剂,其中所述核酸抑制剂用于基因疗法。

193. 根据权利要求176-183的任一项所述的方法所产生的经遗传修饰的所述细胞在制备用于治疗有需要的受试者的药物中的用途。

## 对治疗剂和疾病特异性抗原进行测序、确定、配对和验证的方法

[0001] 交叉引用

[0002] 本申请要求于2015年4月27日提交的美国临时申请号62/153,041的优先权,该美国临时申请通过引用而整体并入本文。

### 背景技术

[0003] 免疫系统采用多种策略来生成T细胞和B细胞抗原受体的组库。这些受体的多样性足以识别全部的潜在病原体。使B淋巴细胞成熟以便表达重(H)轻(L)链多肽的异二聚体抗体(免疫球蛋白,Ig),而T淋巴细胞表达异二聚体T细胞受体(TCR)。免疫系统还充当肿瘤细胞必须躲避才得以存活的外在肿瘤抑制系统。然而,肿瘤细胞可利用机制来逃避免疫识别,并且可导致肿瘤生长。这些机制包括免疫编辑(通过免疫编辑能消除表达高度免疫原性肿瘤抗原的肿瘤细胞),以及免疫原性肿瘤抗原的下调。免疫系统缺乏肿瘤特异性、肿瘤细胞的抗原调节和MHC分子的异常表达以及其他因素阻碍了对肿瘤的检测。

### 发明内容

[0004] (本文)所公开的方法包括确定由例如免疫细胞如肿瘤浸润性淋巴细胞(TIL)外源或内源表达的抗原结合分子如免疫球蛋白(Ig)或TCR或其结合部分的序列。在一些方面,所述抗原结合分子对组织的抗原,例如疾病特异性抗原具有高亲和力。在一些方面,所公开的方法可用于从患病生物样品(例如,患病组织样品)确定、检测和/或选择对组织上或组织中表达的抗原和/或组织的抗原如疾病特异性抗原具有高亲和力的TIL。在一些方面,所公开的方法可用于发现和/或鉴别抗体,包括全长抗体、TCR、治疗靶标和生物标志物的抗原结合部分。在所公开的方法中,采用高通量、准确且最小偏差的测序方法对多核苷酸,如淋巴细胞多核苷酸,例如,Ig和TCR多核苷酸进行测序。在一些方面,方法采用准确的测序方法,如在W02014144495、W02012048340和W02012048341;以及美国临时申请号:62/050549、62/051832、61/938227和62/031405中描述的方法,其中每个申请的内容均通过引用而整体并入本文。所公开的方法可用于例如天然重链和轻链和/或 $\alpha$ TCR和 $\beta$ TCR链或 $\gamma$ TCR和 $\delta$ TCR链序列的配对,例如以便鉴别存在于例如在单个细胞中天然存在和/或在细胞表面上一起在复合体中表达的此类链的配对内的此类序列。在一些实施方案中公开的方法包括对获自患病生物体的生物样品的多核苷酸进行测序。所述生物样品可以是患病样品,例如,实体瘤样品。在一些情况下,所述生物样品包含多个含有待测序的多核苷酸的TIL。

[0005] 在一些实施方案中,所述方法进一步包括选择含有所述测序的多核苷酸如Ig或TCR多核苷酸的淋巴细胞的一个或多个多核苷酸,例如,配对的重链和轻链抗体多核苷酸或配对的 $\alpha$ 链和 $\beta$ 链TCR多核苷酸。所述选择基于从上述测序步骤获得的测序数据。在一些方面,所述方法进一步包括产生由所述选择的多核苷酸编码的多肽,例如,由所述选择的多核苷酸编码的Ig或TCR多肽。在一些实施方案中,所述方法进一步包括,例如通过使用重组表达或合成的Ig或TCR多肽来鉴别由所述选择的淋巴细胞的多核苷酸编码的多肽的抗原。

[0006] 在一些方面,所公开的方法包括提供包含至少一个肿瘤浸润性淋巴细胞(TIL)和至少一个非TIL细胞的生物样品。在一些方面,所公开的方法包括对编码来自所述至少一个TIL和来自所述至少一个非TIL细胞的Ig或TCR多肽的多核苷酸进行测序,从而获得序列信息。在一些方面,所公开的方法进一步包括以下步骤:基于所述序列信息从所述至少一个TIL中的TIL和至少一个非TIL细胞中选择Ig或TCR多核苷酸序列,和/或产生由所述选择的多核苷酸序列编码的Ig或TCR多肽。在一些方面,所述方法进一步包括鉴别所述产生的Ig或TCR多肽的靶抗原。

[0007] 在一些方面,所公开的方法包括对编码来自受试者的生物样品的至少一个TIL的Ig或TCR多肽的多核苷酸和编码来自所述受试者的生物样品的至少一个非TIL细胞的Ig或TCR多肽的多核苷酸进行测序,从而获得序列信息;将所述获得的序列信息与从相应的正常邻近组织样品获得的序列信息进行比较;以及基于所述比较从所述至少一个TIL中的TIL和至少一个非TIL细胞中选择Ig或TCR多核苷酸序列。在一些实施方案中,所公开的方法进一步包括以下步骤:产生由所述选择的多核苷酸序列编码的Ig或TCR多肽;和/或鉴别所述产生的Ig或TCR多肽的靶抗原。

[0008] 在一些实施方案中,所公开的方法包括鉴别从TIL产生的Ig或TCR多肽的靶抗原。在一些实施方案中,所公开的方法包括对编码来自受试者的生物样品的至少一个TIL的Ig或TCR多肽的多核苷酸和编码来自所述受试者的生物样品的至少一个非TIL细胞的Ig或TCR多肽的多核苷酸进行测序,从而获得序列信息;将所述获得的序列信息与从相应的正常邻近组织样品获得的序列信息进行比较;以及基于所述比较从所述至少一个TIL中的TIL和至少一个非TIL细胞中选择Ig或TCR多核苷酸序列;产生由所述选择的多核苷酸序列编码的Ig或TCR多肽;以及鉴别所述产生的Ig或TCR多肽的靶抗原。

[0009] 在一些方面,所公开的方法包括提供来自第一受试者的生物样品,该生物样品包含至少一个肿瘤浸润性淋巴细胞(TIL)和至少一个非TIL细胞;对编码来自所述至少一个TIL和来自所述至少一个非TIL细胞的Ig或TCR多肽的多核苷酸进行测序,从而获得序列信息;将所述序列信息与从来自第二受试者的生物样品获得的序列信息进行比较,其中所述第一和第二受试者患有相同的疾病;基于所述比较从所述至少一个TIL中的TIL和至少一个非TIL细胞中选择Ig或TCR多核苷酸序列;产生由所述选择的多核苷酸序列编码的Ig或TCR多肽;以及鉴别所述产生的Ig或TCR多肽的靶抗原。

[0010] 在一些方面,所述至少一个TIL的形态是未知的。在一些方面,所述至少一个TIL的形态是未知的。在一些方面,所述至少一个TIL的表型是未知的。在一些方面,所述至少一个非TIL细胞的表型是未知的。

[0011] 在一些方面,所述至少一个TIL和所述至少一个非TIL细胞以1:10,000或更小的比例存在于所述生物样品中。在一些方面,所述TIL和所述非TIL细胞以1:100,000或更小的比例存在于所述生物样品中。在一些方面,所述TIL和所述非TIL细胞以1:1,000,000或更小的比例存在于所述生物样品中。

[0012] 在一些实施方案中,编码来自受试者的生物样品的至少一个TIL的Ig或TCR多肽的多核苷酸和编码来自受试者的生物样品的至少一个非TIL细胞的Ig或TCR多肽的多核苷酸以1:10,000或更小、1:100,000或更小、或者1:1,000,000或更小的比例存在于所述生物样品中。在本公开方法的一些实施方案中,编码来自受试者的生物样品的至少一个TIL的Ig或

TCR多肽的多核苷酸和编码来自受试者的生物样品的至少一个非TIL细胞的Ig或TCR多肽的多核苷酸在所述测序步骤期间以1:10,000或更小、1:100,000或更小、或者1:1,000,000或更小的比例存在。

[0013] 在一些方面,所述选择包括进行所述序列信息的生物信息学分析。在一些方面,所述选择包括确定所述序列信息的多核苷酸的表达水平。在一些方面,所述选择包括将所述序列信息的多核苷酸序列进行比对。在一些方面,所述选择基于编码Ig或TCR多肽的多核苷酸的表达水平。在一些方面,所述选择基于编码Ig或TCR多肽的多核苷酸的种系序列的突变模式。在一些方面,所述选择基于所述序列信息中编码Ig或TCR多肽的多核苷酸的种系序列的突变水平。在一些方面,所述选择基于所述序列信息中编码Ig或TCR多肽的多核苷酸的存在以及来自正常细胞的一组序列信息中所述选择的多核苷酸序列的不存在。在一些方面,所述选择基于所述序列信息中编码Ig或TCR多肽的多核苷酸的富集以及来自正常细胞的第二组序列信息中所述选择的多核苷酸序列的不存在。在一些方面,所述选择基于所述序列信息中编码Ig或TCR多肽的多核苷酸的同种型谱。在一些方面,所述选择基于所述序列信息中编码Ig或TCR多肽的多核苷酸的系统发生簇(phylogenetic cluster)。在一些方面,所述选择基于所述序列信息中编码Ig或TCR多肽的多核苷酸的系统发生簇的大小。在一些方面,所述选择基于所述序列信息中编码Ig或TCR多肽的多核苷酸的序列与来自患病生物样品的另一组序列信息的序列之间的相似性。

[0014] 在一些方面,所述患病生物样品包括来自患有所述疾病的第一受试者的患病生物样品的多个淋巴细胞,以及来自患有所述疾病的第二受试者的患病生物样品的多个淋巴细胞。

[0015] 在一些方面,所述选择基于所述序列信息中编码Ig或TCR多肽的多核苷酸的序列与来自正常生物样品的另一组序列信息的序列之间缺乏相似性。

[0016] 在一些方面,所述正常生物样品为正常邻近组织样品。在一些方面,所述正常生物样品包括来自未患所述疾病的第一受试者的正常生物样品的多个淋巴细胞,以及来自未患所述疾病的第二受试者的正常生物样品的多个淋巴细胞。

[0017] 在一些方面,所述方法包括确定所述产生的Ig或TCR多肽对患病组织或患病生物样品或患病细胞的特异性。在一些方面,所述确定特异性包括确定所述产生的Ig或TCR多肽对所述患病组织或所述患病生物样品或所述患病细胞的亲和力以及所述产生的Ig或TCR多肽对相应的正常邻近组织或相同组织类型的相应正常细胞的亲和力。

[0018] 在一些方面,所述方法包括鉴别能杀死患病细胞的所述产生的Ig或TCR多肽。在一些方面,所述鉴别的产生的Ig或TCR多肽通过直接与所述患病细胞结合杀死所述患病细胞。

[0019] 在一些方面,所述产生包括合成或重组表达所述Ig或TCR多肽。在一些方面,所述至少一个非TIL细胞包括上皮细胞、淋巴细胞、癌细胞或其组合。在一些方面,所述至少一个TIL包括至少一个T细胞、至少一个B细胞或其组合。

[0020] 在一些方面,所述生物样品为癌症活检物。在一些方面,所述生物样品为正常组织活检物。在一些方面,所述生物样品包括血管外组织。在一些方面,由所述选择的多核苷酸序列编码的Ig或TCR多肽为重组多肽。在一些方面,所述目标分析物对患病生物样品具有特异性。在一些方面,所述目标分析物对所述生物样品的患病细胞具有特异性。在一些方面,所述目标分析物对癌细胞具有特异性。

[0021] 在一些方面,所述测序为高通量测序。在一些方面,所述测序为合成测序、杂交测序或连接测序。在一些方面,所述测序不包括对整个免疫组库进行测序。在一些方面,所述测序为大规模平行测序。

[0022] 在一些方面,所述方法不包括使用多重引物或附接至固体支持体的多重引物。在一些方面,所述方法不使用包含与Ig或TCR可变域区互补的序列的多重引物。在一些方面,所述方法在所述测序之前不使用从所述至少一个TIL或所述至少一个非TIL细胞分离多核苷酸的步骤。

[0023] 在一些方面,所述生物样品不是血液。在一些方面,所述生物样品为固体组织样品。在一些方面,所述生物样品来自器官。在一些方面,所述生物样品包括三维结构。在一些方面,所述生物样品包括癌细胞或癌前细胞。在一些方面,所述生物样品包括被所述受试者的免疫系统异常靶向的健康组织。

[0024] 在一些方面,所述至少一个非TIL细胞包括1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个或更多个非TIL细胞。

[0025] 在一些方面,所述至少一个TIL包括1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个或更多个TIL。

[0026] 在一些方面,所述选择的多核苷酸序列包括1-500个独特的Ig或TCR多核苷酸序列。在一些方面,所述选择的多核苷酸序列包括至多1、2、3、4、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90、100、125、150、175、200、225、250、275、300、325、350、375、400、425、450、475或500个独特的Ig或TCR多核苷酸序列。

[0027] 在一些方面,所述选择的多核苷酸序列包括来自T细胞的TCR多核苷酸序列。在一些方面,所述选择的多核苷酸序列包括来自B细胞的Ig多核苷酸序列。在一些方面,所述生物样品的所述至少一个TIL和至少一个非TIL细胞不在所述测序之前基于细胞外细胞标志物进行分选。在一些方面,所述生物样品的所述至少一个TIL和至少一个非TIL细胞不在所述测序之前基于细胞标志物进行分选。在一些方面,所述生物样品的所述至少一个TIL和至少一个非TIL细胞不在所述测序之前进行分选。

[0028] 在一些方面,所述测序的多核苷酸包括编码Ig重链(IgH)的Ig多核苷酸。在一些方面,所述方法进一步包括将所述IgH与来自相同B细胞的Ig轻链(IgL)进行配对。在一些方面,所述测序的多核苷酸包括编码IgL的Ig多核苷酸。在一些方面,所述方法进一步包括将所述IgL与来自相同B细胞的IgH进行配对。在一些方面,所述测序的多核苷酸包括编码IgH

的Ig多核苷酸和编码IgL的Ig多核苷酸。在一些方面,所述IgL与来自相同B细胞的IgH配对。在一些方面,所述方法进一步包括将IgL与来自相同B细胞的IgH进行配对。在一些方面,所述测序的多核苷酸包括编码TCR $\alpha$ 链的TCR多核苷酸。在一些方面,所述方法进一步包括将所述TCR $\alpha$ 链与来自相同T细胞的TCR $\beta$ 链进行配对。在一些方面,所述测序的多核苷酸包括编码TCR $\beta$ 链的TCR多核苷酸。在一些方面,所述方法进一步包括将所述TCR $\beta$ 链与来自相同T细胞的TCR $\alpha$ 链进行配对。在一些方面,所述测序的多核苷酸包括编码TCR $\alpha$ 链的TCR多核苷酸和编码TCR $\beta$ 链的TCR多核苷酸。

[0029] 在一些方面,所述TCR $\alpha$ 链与来自相同T细胞的TCR $\beta$ 链配对。在一些方面,所述方法进一步包括将TCR $\alpha$ 链与来自相同T细胞的TCR $\beta$ 链进行配对。在一些方面,所述方法进一步包括生成配对的IgL和IgH的数据库。在一些方面,所述方法进一步包括生成配对的TCR $\alpha$ 和TCR $\beta$ 链的数据库。

[0030] 在一些方面,所述编码Ig或TCR多肽的多核苷酸包含可变区。在一些方面,所述编码Ig的多核苷酸包含重链可变区( $V_H$ )。在一些方面,所述编码Ig的多核苷酸包含轻链可变区( $V_L$ )。在一些方面,所述编码TCR的多核苷酸包含TCR $\alpha$ 链可变区。在一些方面,所述编码TCR的多核苷酸包含TCR $\beta$ 链可变区。在一些方面,所述编码TCR的多核苷酸包含TCR $\gamma$ 链可变区。在一些方面,所述编码TCR的多核苷酸包含TCR $\delta$ 链可变区。在一些方面,所述可变区包含CDR1、CDR2、CDR3、超突变区或其任意组合。在一些方面,所述可变区包含V区段、D区段、J区段或其任意组合。在一些方面,所述编码Ig或TCR多肽的多核苷酸包含TCR恒定域区。在一些方面,所述TCR恒定域区包含TCR $\alpha$ 恒定域、TCR $\beta$ 恒定域或其组合。在一些方面,编码Ig或TCR多肽的多核苷酸的区域包括Ig恒定域区。在一些方面,所述Ig恒定域区包含选自CH<sub>1</sub>、CH<sub>2</sub>、CH<sub>3</sub>和CH<sub>4</sub>的IgH恒定域。

[0031] 在一些方面,所述Ig恒定域区包含选自CH<sub>1</sub>、CH<sub>2</sub>、CH<sub>3</sub>和CH<sub>4</sub>的两个、三个或四个IgH恒定域。在一些方面,所述Ig恒定域区包含来自选自IgM、IgD、IgA、IgE、IgG及其组合的Ig同种型的IgH恒定域。在一些方面,所述选择的多核苷酸序列的Ig同种型为IgG同种型序列。在一些方面,所述Ig恒定域区包含IgL恒定域( $C_L$ )。在一些方面,所述Ig恒定域区包含来自选自Ig $\kappa$ 、Ig $\lambda$ 及其组合的IgL同种型的 $C_L$ 。在一些方面,所述选择的多核苷酸序列的Ig同种型为Ig $\kappa$ 。

[0032] 在一些方面,编码所述Ig或TCR多肽的多核苷酸包含含有种系构架序列的构架区序列。

[0033] 在一些方面,编码所述Ig或TCR多肽的多核苷酸包含含有种系 $V_H$ 序列的 $V_H$ 序列、含有种系 $V_L$ 序列的 $V_L$ 序列、含有种系TCR $\alpha$ 可变区序列的TCR $\alpha$ 可变区序列、含有种系TCR $\beta$ 可变区序列的TCR $\beta$ 可变区序列、含有种系TCR $\gamma$ 可变区序列的TCR $\gamma$ 可变区序列、含有种系TCR $\delta$ 可变区序列的TCR $\delta$ 可变区序列或其组合。

[0034] 在一些方面,编码所述Ig或TCR多肽的多核苷酸包含含有一个或多个突变的构架残基的构架区序列。在一些方面,由所述选择的多核苷酸序列编码的Ig或TCR多肽的一个或多个突变的构架残基中的突变的构架残基是在来自两名或更多名患有疾病的受试者的前5%最大表达的IgH、IgL、TCR $\alpha$ 、TCR $\beta$ 、TCR $\gamma$ 或TCR $\delta$ 多核苷酸中的一种或多种中发现的残基。在一些方面,由所述选择的多核苷酸序列编码的Ig或TCR多肽包括特异性Ig同种型。在一些方面,所述特异性Ig同种型为IgA、IgG、IgM、IgD或IgE。在一些方面,所述特异性同种型为

IgG。

[0035] 在一些方面,所述序列信息包括至少约1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个独特的Ig或TCR序列。

[0036] 在一些方面,所述序列信息包括至少约1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个序列读取。

[0037] 在一些方面,所述序列信息包括来自所述至少一个TIL的至少一个Ig或TCR序列,该至少一个Ig或TCR序列编码对疾病相关蛋白或疾病特异性蛋白具有约 $1 \times 10^{-7}$ M、 $1 \times 10^{-8}$ M、 $1 \times 10^{-9}$ M、 $1 \times 10^{-10}$ M、 $1 \times 10^{-11}$ M、 $1 \times 10^{-12}$ M或更小的 $K_d$ 的Ig或TCR多肽。在一些方面,所述序列信息不包括来自所述至少一个非TIL细胞的Ig或TCR序列,该Ig或TCR序列编码对疾病相关蛋白或疾病特异性蛋白具有 $1 \times 10^{-7}$ M、 $1 \times 10^{-8}$ M、 $1 \times 10^{-9}$ M、 $1 \times 10^{-10}$ M、 $1 \times 10^{-11}$ M、 $1 \times 10^{-12}$ M或更小的 $K_d$ 的Ig或TCR多肽。在一些方面,由所述选择的多核苷酸序列编码的产生的Ig或TCR多肽对疾病相关蛋白或疾病特异性蛋白具有约 $1 \times 10^{-7}$ M、 $1 \times 10^{-8}$ M、 $1 \times 10^{-9}$ M、 $1 \times 10^{-10}$ M、 $1 \times 10^{-11}$ M、 $1 \times 10^{-12}$ M或更小的 $K_d$ 。在一些方面,编码所述产生的Ig或TCR多肽的所述选择的多核苷酸序列编码对疾病相关蛋白或疾病特异性蛋白具有约 $1 \times 10^{-7}$ M、 $1 \times 10^{-8}$ M、 $1 \times 10^{-9}$ M、 $1 \times 10^{-10}$ M、 $1 \times 10^{-11}$ M、 $1 \times 10^{-12}$ M或更小的 $K_d$ 的Ig或TCR多肽。

[0038] 在一些方面,包含所述选择的多核苷酸序列的TIL以所述生物样品的所述至少一个TIL和所述至少一个非TIL细胞的每个总数目约1-500的量存在。在一些方面,包含所述选择的多核苷酸序列的TIL以所述生物样品的所述至少一个TIL和所述至少一个非TIL细胞的每个总数目约一个、两个、三个、四个或五个的量存在。在一些方面,包含所述选择的多核苷酸序列的所述至少一个TIL中的TIL在所述生物样品中以每至少1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个的总免疫细胞约一个的量存在。在一些方面,包含所述选择的多核苷酸序列的所述至少一个TIL中的TIL在所述生物样品中以每至少1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、

$6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个的总淋巴细胞约一个的量存在。在一些方面,包含所述选择的多核苷酸序列的所述至少一个TIL中的TIL在所述生物样品中以每至少1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个的所述至少一个非TIL细胞约一个的量存在。在一些方面,所述至少一个TIL在所述生物样品中以每至少1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个的所述至少一个非TIL细胞约一个的量存在。在一些方面,所述生物样品中疾病相关或疾病特异性淋巴细胞与总淋巴细胞的比例为每至少1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个并非疾病相关或疾病特异性淋巴细胞的生物样品中的淋巴细胞约一个。

[0039] 在一些方面,所述测序的错误率小于或等于0.00001%、0.0001%、0.001%或0.01%。

[0040] 在一些方面,所述测序包括以至少约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序列。

[0041] 在一些方面,扩增错误被最小化、消除或小于0.01%、0.001%、0.0001%、0.00001%、0.000001%或0.0000001%。

[0042] 在一些方面,所述测序包括对所述至少一个TIL和所述至少一个非TIL细胞的至少1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、

$6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个中的编码Ig或TCR多肽的多核苷酸进行测序。

[0043] 在一些方面,对至少1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个编码Ig或TCR多肽的多核苷酸进行测序。

[0044] 在一些方面,所述方法在少于或等于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小时的正时间量内进行。

[0045] 在一些方面,来自所述生物样品的编码Ig或TCR的多核苷酸包括至少1,000、5,000、10,000、20,000、30,000、40,000、50,000、1000,000、500,000或 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个编码Ig或TCR的多核苷酸。

[0046] 在一些方面,在所述测序之前,所述方法包括:形成多个第一容器,每个第一容器包含:来自所述生物样品的所述至少一个TIL或所述至少一个非TIL细胞的单个细胞,以及单个固体支持体;将以下的多核苷酸拷贝到所述单个固体支持体上:来自所述单个细胞的编码Ig或TCR的第一多核苷酸的第一拷贝,以及来自所述单个细胞的编码Ig或TCR的第二多核苷酸的第二拷贝;形成多个第二容器,每个第二容器包含来自所述多个第一容器的单个固体支持体,以及条码化多核苷酸;以及采用以下引物对所述第一拷贝、所述第二拷贝和所述条码进行扩增:第一组引物,以及第二组引物,其中所述第一组引物与所述第二组引物互补;从而形成第一和第二单个TIL或非TIL条码化序列的文库。

[0047] 在一些方面,所述多个第一容器包含在单一反应环境中。在一些实施方案中,单一反应环境指示所述容器不通过物理屏障,例如进入板的单个孔中而彼此隔开。

[0048] 在一些方面,所述第一和第二单个TIL或非TIL条码化序列包含相同的条码。在一些方面,所述方法进一步包括将所述第一和第二单个TIL或非TIL条码化序列进行融合。在一些方面,将所述第一和第二单个TIL或非TIL条码化序列进行融合。

[0049] 在一些方面,在所述测序之前,所述方法进一步包括:形成多个第一容器,每个第一容器包含:来自所述生物样品的所述至少一个TIL或所述至少一个非TIL细胞的单个细胞,以及固体支持体;将以下多核苷酸拷贝到所述固体支持体上:来自所述单个细胞的编码Ig或TCR的第一多核苷酸的第一拷贝,其中所述第一拷贝附接至第一条码化多核苷酸;以及来自所述单个细胞的编码Ig或TCR的第二多核苷酸的第二拷贝,其中所述第二拷贝附接至第二条码化多核苷酸;采用以下引物扩增所述第一拷贝和所述第一条码以及所述第二拷贝

和所述第二条码：正向引物，以及反向引物，从而形成来自所述单个细胞的独特配对的条码化序列的文库；形成多个第二容器，每个第二容器包含来自所述多个第一容器的单个固体支持体；在所述第二容器中扩增：具有第一正向条码引物和第一反向条码引物的第一条码，以及具有第二正向条码引物和第二反向条码引物的第二条码，从而形成扩增的第一和第二条码的文库；其中第一条码引物与第二条码引物互补，或者第一条码引物序列是第二条码引物序列的回文序列。

[0050] 在一些方面，所述方法进一步包括将来自(e)的扩增的第一和第二条码进行融合。

[0051] 在一些方面，所述融合的扩增的第一和第二条码在所述第二容器中融合。

[0052] 在一些方面，所述第一和第二条码包括不同的条码。

[0053] 在一些方面，所述不同的条码是独特的。

[0054] 在一些方面，所述不同的条码是独特的条码对。

[0055] 在一些方面，所述第一和第二条码包括相同的条码。

[0056] 在一些方面，所述第一和第二条码的相同的条码是独特的。

[0057] 在一些方面，在所述测序之前，所述方法进一步包括：形成多个容器，每个容器包含来自所述生物样品的所述至少一个TIL或所述至少一个非TIL细胞的单个细胞；多个分子条码化多核苷酸；以及容器条码化多核苷酸；产生：第一互补多核苷酸，其与来自所述单个细胞的编码Ig或TCR的第一多核苷酸互补，以及第二互补多核苷酸，其与来自所述单个细胞的编码Ig或TCR的第二多核苷酸互补；将所述多个分子条码化多核苷酸中的第一分子条码化多核苷酸附接至所述第一互补多核苷酸，以及将第二分子条码化多核苷酸附接至所述第二互补多核苷酸，从而形成第一和第二单个TIL或非TIL单条码化(single-barcoded)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个TIL或非TIL单条码化多核苷酸，以及所述第二单个TIL或非TIL单条码化多核苷酸，从而形成第一和第二单个细胞双条码化序列的文库。

[0058] 在一些方面，在所述测序之前，所述方法进一步包括：采用以下物质产生来自所述生物样品的所述至少一个TIL或所述至少一个非TIL细胞的编码 $V_H$ 或TCR $\alpha$ 或TCR $\gamma$ 的多核苷酸的第一互补多核苷酸，以及来自所述生物样品的所述至少一个TIL或所述至少一个非TIL细胞的编码 $V_L$ 或TCR $\beta$ 或TCR $\delta$ 的多核苷酸的第二互补多核苷酸，从而形成第一和第二单个TIL或非TIL单条码化多核苷酸：第一引物，其包含与来自所述生物样品的所述至少一个TIL或所述至少一个非TIL细胞的编码Ig或TCR的多核苷酸的相同区域互补的区域；第二引物，其包含与编码Ig或TCR的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，每个分子条码化多核苷酸包含：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；扩增所述容器条码化多核苷酸，从而形成第一和第二单个TIL或非TIL双条码化多核苷酸；以及扩增所述第一和第二单个TIL或非TIL双条码化多核苷酸，从而形成包含 $V_H$ 、 $V_L$ 、TCR $\alpha$ 、TCR $\beta$ 、TCR $\gamma$ 或TCR $\delta$ 多核苷酸的可变区的序列的文库；以及对所述文库的一个或多个所述序列进行测序，其中所述文库代表所述样品的免疫状态，其中所述产生在多个容器的容器中进行，其中所述容器包含来自所述生物样品的所述至少一个TIL或所述至少一个非TIL细胞的单个细胞。在一些方面，所述第一和第二分子条码化多核

核苷酸的分子条码不同。在一些方面,所述第一和第二单个TIL或非TIL单条码化多核苷酸包含不同的分子条码。在一些方面,所述第一和第二单个TIL或非TIL双条码化序列包含不同的分子条码。在一些方面,所述第一和第二单个TIL或非TIL双条码化序列包含相同的容器条码。在一些方面,所述多个分子条码化多核苷酸不是扩增产物。

[0059] 在一些方面,所述至少一个TIL和所述至少一个非TIL细胞来自患有疾病的受试者的生物样品。在一些方面,所述受试者为动物。在一些方面,所述动物为哺乳动物。在一些方面,所述哺乳动物为人类。在一些方面,从所述生物样品分离所述编码Ig或TCR多肽的多核苷酸。在一些方面,不从所述生物样品分离所述编码Ig或TCR多肽的多核苷酸。

[0060] 在一些方面,来自患有所述疾病的受试者的生物样品包括来自2个或更多个患有所述疾病的受试者的多个生物样品。在一些方面,所述多个生物样品包括至少3、4、5、10、20、30、40、50、60、70、80、90或100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、100,000或1,000,000个或更多个样品。

[0061] 在一些方面,所述疾病为自身免疫病。在一些方面,所述疾病为癌症。在一些方面,所述疾病为癌前疾病。

[0062] 在一些方面,所述方法进一步包括修正扩增错误。在一些方面,所述方法进一步包括修正测序错误。在一些方面,所述方法进一步包括将包含相同条码序列的序列分箱(bin)或分组。在一些方面,所述方法进一步包括使用计算机或算法将包含相同条码序列的序列分箱或分组。在一些方面,所述方法进一步包括簇集具有至少约90%、95%或99%的序列同源性的序列。在一些方面,所述方法进一步包括将具有至少约90%、95%或99%的序列同源性的序列进行比对。在一些方面,借助于计算机或算法进行所述簇集或比对。在一些方面,所述方法进一步包括将序列读取与种系序列进行比较,并确定所述序列读取的体细胞超突变积累。在一些方面,所述方法进一步包括确定所述序列的同种型分布。

[0063] 在一些方面,由所述选择的多核苷酸序列编码的Ig或TCR多肽基本不与正常邻近组织的细胞相互作用。在一些方面,由所述选择的多核苷酸序列编码的Ig或TCR多肽基本不与来自未患所述疾病的受试者的相同组织的细胞结合。

[0064] 在一些方面,所述产生包括表达由所述选择的多核苷酸序列编码的重组Ig或TCR多肽。在一些方面,所述产生包括表达各自由选择的多核苷酸序列编码的1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、300、400或500个或更多个重组Ig或TCR多肽。在一些方面,所述产生包括将所述选择的多核苷酸序列中的序列克隆到载体中。在一些方面,所述载体为克隆载体。在一些方面,所述载体为表达载体。在一些方面,所述产生包括使细胞与包含编码Ig或TCR多肽的所述选择的多核苷酸序列中的序列的多核苷酸接触。在一些方面,所述接触包括转染。在一些方面,所述产生包括在细胞中表达由所述选择的多核苷酸序列编码的重组Ig或TCR多肽。在一些方面,所述细胞为哺乳动物细胞。在一些方面,所述哺乳动物细胞为中国仓鼠卵巢(CHO)细胞或HEK293细胞。在一些方面,所述方法进一步包括纯化所述产生的由所述选择的多核苷酸序列编码的重组Ig或TCR多肽。在一些方面,所述方法进一步包括分离所述产生的由所述选择的多核苷酸序列编码的重组Ig或TCR多肽。在一些方面,由所述选择的多核苷酸序列编码的重组Ig或TCR多肽包含异源标签。在一些方面,所述异源标签为纯化标签。在一些方面,所述细胞为细菌细胞或昆虫细胞。

[0065] 在一些方面,所述鉴别包括将Ig或TCR序列与包含Ig或TCR序列数据的数据库进行

比较。在一些方面,所述鉴别包括进行全基因组siRNA筛选。在一些方面,所述鉴别包括采用由所述选择的多核苷酸序列编码的Ig或TCR多肽进行蛋白质展示筛选。在一些方面,所述蛋白质展示筛选为噬菌体展示筛选。在一些方面,所述蛋白质展示筛选为核糖体展示筛选。在一些方面,所述鉴别包括进行酵母双杂交筛选。在一些方面,所述鉴别包括进行2D凝胶电泳。在一些方面,所述鉴别包括采用蛋白质阵列对由所述选择的多核苷酸序列编码的Ig或TCR多肽进行筛选。在一些方面,所述蛋白质阵列包含至少约10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、96%、97%、98%或99%或更多的人类蛋白质组蛋白质。在一些方面,所述鉴别包括针对由所述选择的多核苷酸序列编码的Ig或TCR多肽进行蛋白质组筛选。在一些方面,所述鉴别包括采用由所述选择的多核苷酸序列编码的Ig或TCR多肽进行免疫沉淀。在一些方面,所述鉴别包括进行质谱分析。在一些方面,所述鉴别包括采用由所述选择的多核苷酸序列编码的Ig或TCR多肽进行抗体依赖性细胞介导的细胞毒性(ADCC)测定。在一些方面,所述鉴别包括确定由所述选择的多核苷酸序列编码的Ig或TCR多肽的特异性。在一些方面,所述鉴别包括进行结合测定。在一些方面,所述鉴别包括使由所述选择的多核苷酸序列编码的Ig或TCR多肽与至少一个目标分析物候选物接触。

[0066] 在一些方面,所述目标分析物候选物在固体支持体上。在一些方面,所述目标分析物候选物在溶液中(例如,核糖体展示)。在一些方面,由所述选择的多核苷酸序列编码的Ig或TCR多肽在固体支持体上。在一些方面,由所述选择的多核苷酸序列编码的Ig或TCR多肽在溶液中。在一些方面,所述固体支持体为阵列。在一些方面,所述固体支持体为珠子。

[0067] 在一些方面,与由所述选择的多核苷酸序列编码的Ig或TCR多肽结合的所述目标分析物是未知的。

[0068] 在一些方面,在选择所述选择的多核苷酸序列时,与由所述选择的多核苷酸序列编码的Ig或TCR多肽结合的所述目标分析物是未知的。在一些方面,描述了目标分析物,其包括通过本文公开的方法鉴别的目标分析物。在一些方面,所述鉴别的目标分析物为疾病相关或疾病特异性目标分析物。在一些方面,所述鉴别的目标分析物为具有细胞外区域的多肽。在一些方面,描述了分离的纯化的Ig或TCR多肽,其中该分离的纯化的Ig或TCR多肽由本文所述方法中所述选择的多核苷酸序列编码。在一些方面,描述了分离的纯化的IgL多肽,其中该分离的纯化的IgL多肽由本文所述方法中所述选择的多核苷酸序列中的Ig多核苷酸编码。在一些方面,描述了分离的纯化的IgH多肽,其中该分离的纯化的IgH多肽由本文所述方法中所述选择的多核苷酸序列中的Ig多核苷酸编码。在一些方面,描述了由IgH和IgL多核苷酸编码的分离的纯化的抗体,该抗体包括由本文所述方法中所述选择的多核苷酸序列编码的分离的纯化的抗体。在一些方面,描述了Ig多肽的分离的纯化的Fab片段,该Fab片段包括由本文所述方法中所述选择的多核苷酸序列编码的Ig多肽的分离的纯化的Fab片段。在一些方面,描述了Ig多肽的分离的纯化的F(ab)<sub>2</sub>片段,该F(ab)<sub>2</sub>片段包括由本文所述方法中所述选择的多核苷酸序列编码的Ig多肽的分离的纯化的F(ab)<sub>2</sub>片段。在一些方面,描述了Ig多肽的分离的纯化的Fv片段,该Fv片段包括由本文所述方法中所述选择的多核苷酸序列编码的Ig多肽的分离的纯化的Fv片段。在一些方面,描述了Ig多肽的分离的纯化的ScFv片段,该ScFv片段包括由本文所述方法中所述选择的多核苷酸序列编码的Ig多肽的分离的纯化的ScFv片段。在一些方面,描述了TCR $\alpha$ 多肽的分离的纯化的片段,该片段包括由本文所述方法中所述选择的多核苷酸序列编码的TCR $\alpha$ 多肽的分离的纯化的片段。在一些

方面,描述了TCR $\beta$ 多肽的分离的纯化的片段,该片段包括由本文所述方法中所述选择的多核苷酸序列编码的TCR $\beta$ 多肽的分离的纯化的片段。在一些方面,描述了TCR $\alpha$ 和TCR $\beta$ 多肽的分离的纯化的片段,该片段包括由本文所述方法中所述选择的多核苷酸序列编码的TCR $\alpha$ 和TCR $\beta$ 多肽的分离的纯化的片段。在一些方面,由所述选择的多核苷酸序列编码的Ig或TCR多肽的所述鉴别的目标分析物为所述疾病的生物标志物。

[0069] 在一些方面,描述了一种治疗有需要的受试者的方法,该方法包括向患有所述疾病的受试者施用由本文所述方法中选择的多核苷酸序列编码的Ig或TCR多肽或其片段。在一些方面,由所述选择的多核苷酸序列编码的Ig或TCR多肽为人类治疗性多肽。在一些方面,由所述选择的多核苷酸序列编码的Ig或TCR多肽为中和多肽。在一些方面,描述了一种治疗有需要的受试者的方法,该方法包括向患有所述疾病的受试者施用由本文所述方法中选择的多核苷酸序列编码的Ig或TCR多肽的所述鉴别的目标分析物的抑制剂或其片段。在一些方面,所述抑制剂选自小分子、核酸、多肽及其组合。在一些方面,所述抑制剂为多肽抑制剂,其中所述多肽抑制剂为由所述选择的多核苷酸序列编码的Ig或TCR多肽。在一些方面,所述抑制剂为核酸抑制剂,其中所述核酸抑制剂为siRNA核酸。在一些方面,所述抑制剂为核酸抑制剂,其中所述核酸抑制剂用于基因疗法。

[0070] 援引并入

[0071] 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。

[0072] 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

[0073] 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:

[0074] 图1A例示了显示抗体组库的图,其揭示了胰腺肿瘤样品在其可与正常样品区分开的抗体谱中显示出相似性。

[0075] 图1B例示了多个PDAC肿瘤样品的抗体谱的相关性的热图。

[0076] 图2A例示了证明在大多数分析的前列腺导管腺癌(PDAC)样品中肿瘤切除组织受表达IgG抗体的B细胞的存在支配(从而表明靶特异性免疫应答)的图。匹配的正常组织样品与具有主要IgA应答的正常胰腺组织样品类似。这些IgG抗体的表达规模表示显著的寡克隆应答,其通常指示强烈且特异性的免疫应答。

[0077] 图2B例示了显示在多发性硬化样品的脑活检中的异常应答和表达高频寡克隆抗体的B细胞的存在图。

[0078] 图3A例示了两个PDAC肿瘤样品和相应正常邻近组织样品的免疫测序的mRNA/克隆

总数目与克隆等级的图。结果证明,对PDAC的总体免疫应答受IgG同种型支配,而在正常邻近组织(NAT)中,免疫细胞受IgA同种型支配。

[0079] 图3B例示了mRNA/克隆总数目与克隆等级以及mRNA克隆的PDAC级分与mRNA克隆的NAT级分的图。结果证明,对于最大量表达的抗体,肿瘤样品也几乎完全受表达IgG的B细胞的存在支配。

[0080] 图4例示了本文公开的示例性方法的流程图,该方法包括免疫测序、抗体选择、抗体产生、抗体验证、鉴别抗体的抗原和确定抗体的治疗功效。

[0081] 图5例示了本文公开的示例性方法的步骤流程图。使用来自正常和肿瘤组织样品的序列数据选择会对肿瘤特异性抗原显示高亲和力和特异性的一种或多种抗体。重组产生抗体并将其用于免疫荧光测定,以通过对患病组织进行染色并与正常邻近组织的染色进行比较来验证选定的抗体。然后在FFPE人类组织的免疫组织化学荧光测定中测定由免疫荧光测定验证的选定抗体。通过该验证阶段的抗体可用于本文公开的方法中,以鉴别生物标志物或疾病特异性抗原。

[0082] 图6A例示了用显示出与导管腺癌、腺鳞癌和神经内分泌癌强结合但对正常胰腺组织的染色最小的抗体进行的免疫染色。

[0083] 图6B例示了被形成用于检测抗体与人类组织的FFPE载玻片的结合的示例性免疫偶联物的示意图。

[0084] 图7A例示了用来自上皮细胞PDAC样品和基质细胞PDAC样品的抗体A1-22进行的免疫染色。坎妥珠单抗(Cantuzumab)显示无基质细胞染色,仅上皮细胞染色。这指示,靶向基质抗原可对PDAC具有临床益处。

[0085] 图7B例示了用来自鳞状细胞肺癌样品的抗体A1-22进行的免疫染色,该样品具有与PDAC相似的表型细胞进化,从而显示出相对于正常肺组织的强烈且特异性的染色。坎妥珠单抗显示任何组织均未染色。

[0086] 图8A例示了用来自鳞状细胞肺癌样品的抗体A1-22进行的免疫染色,该样品具有与PDAC相似的表型细胞进化,从而显示出相对于正常肺组织的强烈的特异性的染色。坎妥珠单抗显示任何组织均未染色。

[0087] 图8B例示了通过对图8A中所示样品进行免疫荧光染色获得的平均荧光的图。

[0088] 图9A例示了用来自慢性胰腺炎样品的抗体A1-99进行的免疫染色。在所示的所有正常组织中均看到强染色,从而表明该疾病可能是自身免疫的。

[0089] 图9B例示了用来自PDAC样品的抗体A1-108进行的免疫染色。在所示的所有正常组织中均看到最小染色。

[0090] 图10例示了用于生成含有多个小滴的乳液的系统的示意图,每个小滴均含有单个细胞。可在这些单独的区室中裂解细胞。

[0091] 图11例示了以高通量形式对多核苷酸进行测序的示例性方法的示意图,其中将B细胞从生物样品分离成单独的乳液,在该乳液中对B细胞进行裂解和测序。

[0092] 图12例示了以高通量形式对多核苷酸进行测序的示例性方法的示意图,其中将B细胞从生物样品分离成单独的乳液,在该乳液中对B细胞进行裂解和测序,使得在通过使用小滴条码和分子条码对单独的链进行测序之后使单个细胞的重链和轻链配对。

[0093] 图13例示了由执行用于幼稚、记忆和血浆B细胞配对的本文所述方法所得到的准

确度和严格性(stringency)数据。

[0094] 图14例示了显示出在所示时间段高严格性配对与总配对的比例的预期增加以及预期配对的高严格性配对和抗体对的总数目的示意图。

[0095] 图15例示了比较本文公开的方法提供的相对于免疫组库测序和抗体配对的已知方法的一些示例性优点和技术方案的示意图。

[0096] 图16例示了从患病样品的免疫测序获得的TCR的等级丰度表达图。绘制每个克隆的TCR $\alpha$ 和TCR $\beta$ 的mRNA数目与所示T细胞克隆的等级的图。

[0097] 图17例示了本文公开的示例性方法的步骤流程图。收集来自正常和肿瘤组织样品的序列数据,然后对序列数据进行生物信息学处理。基于多种标准,包括例如mRNA丰度、克隆扩增和体细胞超突变选择TIL的一种或多种抗体或TCR。然后重组产生选定的抗体。然后使用免疫荧光测定通过对患病组织进行染色并与正常邻近组织的染色进行比较来测试重组抗体或TCR对肿瘤特异性抗原的高亲和力和特异性。然后在FFPE人类组织的免疫组织化学荧光测定中测定由免疫荧光测定验证的选定抗体。然后将通过该验证阶段的抗体用于免疫沉淀测定,并使用质谱来确定抗体靶向的抗原。

[0098] 图18A例示了来自测序数据的小滴条码数目与指定受体链组合的图,该测序数据获自含有约400,000个肿瘤解离细胞的未分离TIL的卵巢癌样品(处理的样品含有正常上皮细胞、癌细胞和TIL),其中同时对B细胞和T细胞进行测序。该图显示了准确的B细胞和T细胞受体配对,而没有串扰(crosstalk)或污染。

[0099] 图18B例示了来自图18A中所述样品的总B细胞的百分比与所示Ig同种型的图。肿瘤显示了衍生出TIL的活化Ig浸润物的显著富集。

[0100] 图18C例示了来自图18A中所述样品的轻链可变基因突变百分比与重链可变基因突变百分比以及指定同种型的重链密度与指定配对同种型的轻链密度的图。肿瘤显示了衍生出TIL的重度突变Ig浸润物的显著富集。

[0101] 图19A例示了来自用作选择并分级TIL的标准的TIL免疫组库测序分析的mRNA丰度与克隆等级的图。

[0102] 图19B例示了来自用作选择并分级TIL的标准的TIL免疫组库测序分析的扩增量(细胞数目)与克隆等级的图。

[0103] 图19C例示了来自用作选择并分级TIL的标准的TIL免疫组库测序分析的体细胞超突变率与克隆等级的图。

[0104] 图20例示了来自用作选择并分级TIL的标准的TIL免疫组库测序分析的指定Ig同种型的体细胞超突变率与扩增量(细胞数目)的图。

[0105] 图21A例示了来自TIL免疫组库测序分析的指定Ig同种型的CD21<sup>lo</sup>克隆的体细胞超突变率与mRNA丰度的图。数据表明,CD21<sup>lo</sup>克隆在分析的肺肿瘤样品中扩增。数据表明,可从测序数据鉴别第二细胞标志物并使用该第二细胞标志物选择TIL。

[0106] 图21B例示了图21A中所示克隆的密度与平均CD21表达的图。

[0107] 图22例示了用于鉴别选定TIL的靶抗原的本文公开的示例性方法的步骤流程图。产生来自选定TIL的重组抗体,并将其用于与质谱分析相结合的免疫沉淀测定。

[0108] 图23例示了来自用对照抗体(CH59)、帕尼单抗或蛋白-G单独进行免疫沉淀的Mia PaCa-2细胞裂解物样品的蛋白质染色凝胶。当通过质谱进行分析时,在用帕尼单抗进行免

疫沉淀的样品中,160kDa条带(EGFR)的强度百分比是免疫沉淀的全部组合物的98.5%。该数据表明,可以以高准确度鉴别选定TIL的靶抗原。

### 具体实施方式

[0109] 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施。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。

[0110] 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”。

[0111] 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。

### [0112] 定义

[0113] 当提及多核苷酸或多肽分子时,术语“候选物”是指基于如本文公开的方法所述的测序信息选择的来自淋巴细胞的多核苷酸或多肽。

[0114] 术语“浸润”或“肿瘤浸润”免疫细胞是指来自生物样品,如患病或肿瘤组织样品的免疫细胞的异质群体。浸润免疫细胞包括髓系细胞(粒细胞、巨噬细胞和髓源抑制性细胞)和淋巴细胞谱系的细胞(T细胞、B细胞和自然杀伤(NK)细胞)。

[0115] 术语“抗体”是指天然的或部分或完全合成产生的免疫球蛋白(Ig)。术语“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)。

[0116] 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。

[0117] 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 $\alpha$ 链和 $\beta$ 链或由 $\gamma$ 和 $\delta$ 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。

[0118] 每个TCR含有可变互补决定区(CDR),以及构架区(FR)和恒定区。 $\alpha$ 和 $\beta$ 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 $\beta$ 链基因座中可变性( $V\beta$ )、多样性( $D\beta$ )和连接( $J\beta$ )基因区段之间,以及 $\alpha$ 链基因座中类似的 $V\alpha$ 和 $J\alpha$ 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 $\alpha$ 和 $\beta$ 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 $V\beta$ - $D\beta$ 、 $D\beta$ - $J\beta$ 和 $V\alpha$ - $J\alpha$ 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。

[0119]  $\gamma\delta$ 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 $\gamma$ 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 $\gamma$ 组库。

[0120] 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, $\mu$ 、 $\delta$ 、 $\gamma$ 、 $\alpha$ 和 $\beta$ 。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。

[0121] 关于抗体链如重链和轻链或TCR链如 $\alpha$ 和 $\beta$ 链或 $\gamma$ 和 $\delta$ 链的术语“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,均在轻链和重链可变域或 $\alpha$ 和 $\beta$ 可变域内。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。

[0122] 术语“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。

[0123] 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。

[0124] 对应于免疫球蛋白的不同类别的重链恒定域分别被称为 $\alpha$ 、 $\delta$ 、 $\epsilon$ 、 $\gamma$  和 $\mu$ 。

[0125] 来自任何脊椎动物物种的抗体的轻链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 ( $\kappa$ ) 和lambda ( $\lambda$ )。

[0126] 术语“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。例如,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。

[0127] 术语“多克隆抗体”是指由免疫细胞群体合成的抗体分子群体。

[0128] “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')<sub>2</sub>、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。

[0129] 术语“单链Fv”或“scFv”是指包含抗体的可变重链(V<sub>H</sub>)和可变轻链(V<sub>L</sub>)域或TCR的可变 $\alpha$ 链(V $\alpha$ )和可变 $\beta$ 链(V $\beta$ )域或TCR的可变 $\gamma$ 链(V $\gamma$ )和可变 $\delta$ 链(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<sub>H</sub>和V<sub>L</sub>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,使其scFv能够形成用于抗原结合的所需结构。

[0130] 术语“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<sub>L</sub>连接的V<sub>H</sub>(V<sub>H</sub>-V<sub>L</sub>),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 $\delta$ 连接的V $\gamma$ (V $\gamma$ -V $\delta$ )。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。

[0131] 术语“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。

[0132] 术语“线性抗体”和线性“TCR”是指形成一对抗原结合区的一对串联Fd区段(V<sub>H</sub>-C<sub>H1</sub>-V<sub>H</sub>-C<sub>H1</sub>或V $\alpha$ -Ca<sub>1</sub>-V $\alpha$ -Ca<sub>1</sub>)。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。

[0133] 术语“抗体文库”或“TCR文库”是指抗体或TCR或抗体片段或TCR片段的集合。例如,抗体或TCR组库可用来选择特定的抗体或TCR,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抗体和TCR噬菌体展示文库,包括原始、合成和半合成文库。

[0134] 术语“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。

[0135] 术语“肿瘤浸润性淋巴细胞”(TIL)是指浸润到癌结节的基质中的淋巴细胞。

[0136] 术语“合成的多核苷酸”或“合成的多肽”是指与相当的天然序列相比,相应的多核苷酸或多肽序列或其一部分或氨基酸序列或其一部分衍生自己被设计或从头合成或修饰

的序列。可通过本领域已知的方法,包括但不限于核酸或氨基酸序列的化学合成来制备合成的多核苷酸或多肽。

[0137] 术语“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包含单个 $\alpha$ 链或单个 $\beta$ 链的TCR。

[0138] “ $F(ab')_2$ ”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 $C_{H\gamma 1}$ (重链的恒定区中的 $\gamma 1$ 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 $F(ab')_2$ 。

[0139] 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。

[0140] “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 $\alpha$ 链和一条TCR $\beta$ 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 $V\alpha$ - $V\beta$ 二聚体或 $V\gamma$ - $V\delta$ 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 $V\alpha$ - $V\beta$ 链或 $V\gamma$ - $V\delta$ 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 $V\alpha$ 和 $V\beta$ 链或 $V\gamma$ 和 $V\delta$ 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 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ )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 $V\alpha$ 和 $V\beta$ 或 $V\delta$ 和 $V\gamma$ 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。

[0141] 抗原结合多肽还包括重链二聚体,例如,来自骆驼和鲨鱼的抗体。骆驼和鲨鱼抗体包含V-样和C-样结构域(均不具有轻链)的两条链的同源二聚体对。因为骆驼中重链二聚体

IgG的 $V_H$ 区不需要与轻链进行疏水相互作用,所以在骆驼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。在骆驼中,抗体组库的多样性由 $V_H$ 或 $V_{HH}$ 区中的CDR1、2和3决定。骆驼 $V_{HH}$ 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)。

[0142] 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)。

[0143] 术语“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 $\alpha$ 或TCR $\beta$ 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。

[0144] 术语“亲和力”是指两种物质可逆结合的平衡常数,且表示为Kd。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解离的阻力。

[0145] 术语“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组成(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。

[0146] 术语“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。

[0147] 术语“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。

[0148] 当与治疗性组合物结合使用时,术语“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。

[0149] 术语“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。

[0150] 术语“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。

[0151] 术语“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况、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。

[0152] 鉴别淋巴细胞靶标的方法

[0153] 在健康的免疫前受试者中,识别疾病相关抗原的T细胞大部分(如果不是唯一的话)在幼稚T细胞区室中。幼稚型个体中抗原特异性T细胞的频率为约 $1 \times 10^{-5}$ ,而利用白血病的TIL(ATCI)的过继性免疫疗法需要约 $1 \times 10^9$ 个T细胞。因此,抗原特异性ATCI需要抗原特异性T细胞的大规模扩增,其必须在人类中离体(ex vivo)进行。遗憾的是,用于离体扩增的大多数方法导致抗原致敏T细胞的耗尽,这些T细胞具有缩短的端粒并丧失功能属性。

[0154] 多于50%的晚期黑素瘤患者对ATCI产生反应(Dudley等人,2005)。然而,由于上述原因并且由于可分离的TIL的数目很少,因此将这种方法应用到其他癌症是困难的。可在离体扩增之前从组织分离的TIL的总数目可少于约 $50 \times 10^8$ 、 $25 \times 10^8$ 、 $10 \times 10^8$ 、 $5 \times 10^8$ 、 $1 \times 10^8$ 、 $50 \times 10^7$ 、 $25 \times 10^7$ 、 $10 \times 10^7$ 、 $5 \times 10^7$ 、 $1 \times 10^7$ 、 $50 \times 10^6$ 、 $25 \times 10^6$ 、 $10 \times 10^6$ 、 $5 \times 10^6$ 、 $1 \times 10^6$ 、 $50 \times 10^5$ 、 $25 \times 10^5$ 、 $10 \times 10^5$ 、 $5 \times 10^5$ 、 $1 \times 10^5$ 或更少。在从组织分离的细胞的总数目中,TIL的百分比可小于约70%、65%、60%、55%、50%、45%、40%、35%、30%、25%、20%、15%、10%、5%或更小。甚至确

定已经浸润患病的组织如实体瘤的不同淋巴细胞亚型的数目和比例的当前方法仍然低效且困难。在实体瘤中,不准确的比例导致疾病错误预后。

[0155] 由于T细胞发育期间的胸腺选择,并且由于许多肿瘤抗原是自身抗原,因此在携带癌症的患者中,循环T细胞已经极大地暴露于肿瘤抗原。在循环T细胞上表达的天然TCR通常对自身抗原具有低亲和力( $K_D$ 范围为1-100 $\mu$ M)。因为癌细胞通常在其表面上表达少量的表位/HLA复合体,所以这样的循环T细胞对自体癌细胞的反应性较差。此外,对癌症特异性抗原具有高亲和力的TIL的数目与缺乏这种特异性的免疫细胞的巨大数目相比是不可思议的少。对癌症特异性抗原具有高亲和力的TIL的百分比与缺乏这种特异性的免疫细胞相比可小于约50%、40%、30%、20%、15%、10%、9%、8%、7%、6%、5%、4%、3%、2%、1%或更小。因此,对于在含有指数级更大量的T细胞和其他非疾病特异性的免疫细胞的样品中鉴别对疾病特异性抗原具有高亲和力的浸润T细胞的方法存在巨大需求。

[0156] 可基于使用生物样品检测功能性重排的免疫球蛋白和TCR编码基因来实现B细胞和T细胞数目的定量表征,从所述生物样品可容易地分离出大量的此类淋巴细胞,所述生物样品诸如血液、淋巴液或其他生物流体。在这些样品中,B细胞和T细胞在流体悬浮液中以颗粒存在。然而,目前对其中B细胞和T细胞无法被容易地分离的组织或器官中的淋巴细胞进行定量的方法受到诸多限制。例如,为了检测实体组织和实体瘤样品中的淋巴细胞,需要在小的非代表性样品中进行组织学检测。这些劳动密集型的半定量技术通常使用在固定或冷冻的活检标本切片上的免疫组织化学或原位杂交。由于在处理过程中部分样品的损失或破坏,这种费时费力的步骤可阻止从样品中回收淋巴细胞。当前方法的这些局限性和相关局限性都会损害可获得的定量数据的质量。

[0157] 从这些方法获得有意义的定量数据的努力在可浸润组织的淋巴细胞的数目方面具有严重的局限性。例如,当数据收集取决于可通过观察显微镜载玻片上有限数目的小场检测到的事件数目时,无法实现高的统计显著性。此外,组织样品必须被机械解离和/或酶促解离,以产生适于流动免疫细胞荧光分析的单个细胞悬浮液。

[0158] 尽管已经在一些癌症类型(例如,前列腺导管腺癌(PDAC))中对肿瘤浸润T细胞淋巴细胞进行了深入研究,但对肿瘤环境中的癌症特异性B细胞的检测和表征迄今为止仍然在很大程度上未知。此外,免疫组织化学技术应用于T细胞的限制大部分来自回顾性临床研究,并且无法对不同TIL群体的详细分析以及对它们在肿瘤微环境中的功能特性进行评估。尽管浸润T细胞数目的定量分析可用于预后,但仍然存在对鉴别疾病特异性T细胞的TCR序列的方法的需要。

[0159] 尽管浸润B细胞数目的定量分析也可用于预后,但仍然存在对鉴别疾病特异性B细胞的Ig序列的方法的需要。为发现肿瘤浸润B淋巴细胞的Ig序列所做的尝试少之又少。为发现这些淋巴细胞的靶标(相对于T细胞)所做出的尝试更少,而且都没有成功。可通过寻找相关抗体的低统计学机会来解释其他人的不成功尝试,这是由于天然存在的大量B细胞对患病和正常样品中存在的疾病抗原没有特异性。目前用来研究浸润淋巴细胞的方法是低通量的,并且由于生物体中免疫组库的庞大规模而无法有效地产生治疗性抗体。为了鉴别新型疾病相关抗原,通常进行漫长的基因和蛋白质功能研究。因此,对于在含有指数级更大量的非疾病特异性B细胞和其他免疫细胞的样品中鉴别对疾病特异性抗原具有高亲和力的浸润B细胞的方法存在巨大需求。

[0160] 显然,需要用于在复杂的生物样品中鉴别并选择疾病特异性淋巴细胞的方法,该复杂的生物样品含有许多非疾病特异性的淋巴细胞和其他不是淋巴细胞的细胞。此外,需要鉴别疾病特异性淋巴细胞靶标的方法。目前描述的实例利用与生物信息学和蛋白质组学方法相结合的高通量且高度准确的无偏差免疫组库测序,从而满足这些需求并提供其他相关优势。

#### [0161] 免疫测序

[0162] 提供了操作来自样品的核酸以便生成多核苷酸文库用于测序的方法。在一般意义上,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cDNA并扩增免疫细胞(包括淋巴细胞)的遗传物质。在一些情况下,从B细胞的核酸获得免疫球蛋白序列。在一些情况下,从T细胞的核酸获得T细胞受体序列。在一些情况下,核酸为RNA。在一些情况下,核酸包含IgH或TCR $\beta$ 链或(V、D、J区段)核酸、IgL或TCR $\alpha$ 链(V、J区段)核酸或这两者。在一些情况下,核酸包含TCR  $\gamma$  链核酸、TCR $\delta$ 链核酸或这两者。

#### [0163] 样品

[0164] 样品包括但不限于生物、环境、医学、受试者或患者样品或含有多核苷酸如多核苷酸的样品。含有淋巴细胞的任何生物样品均可在所公开的方法中使用。含有多核苷酸的任何生物样品均可在所公开的方法中使用。例如,样品可以为来自受试者的含有淋巴细胞的生物样品,该淋巴细胞包含编码Ig或TCR多肽的RNA或DNA。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。在一种情况下,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。在某些情况下,从单个细胞获得多核苷酸。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。组织或体液试样可用作使用所公开的方法进行测序的核酸的来源。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。从中获得核酸的细胞或组织可以是患病的或者感染有病毒或其他细胞内病原体。

[0165] 在某些情况下,免疫细胞如TIL可从受试者或宿主中分离,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,该受试者或宿主诸如是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或表现出疾病的症状。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Ig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抗原,例如一种或多种未知抗原的Ig或TCR。在一些情况下,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(pool)进行富集,之后对Ig或TCR链进行测序、产生Ig或TCR,或产生表达文库。在一些情况下,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,免疫细胞文库可含有表达不同Ig或TCR的单独的免疫细胞或其至少5;10;100;250;500;750;1,000;2,500;5,000;10,000;25,000;50,000;75,000;10,000;250,000;500,000;750,000;1,000,000;2,500,000;

5,000,000;7,500,000或10,000,000个或更多个子集。本发明的方法使免疫细胞测序最大化,并提供了非常高的多样性。

[0166] 在一些情况下,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 $1 \times 10^{-6}$ 至 $1 \times 10^{-7}$ M)与基本上任何非自身分子结合的抗体。抗体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 $V_H-V_L$ 组合或 $V\alpha-V\beta$ 组合或 $V\gamma-V\delta$ 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或TCR。然而,使用来自未经免疫或未患病的受试者或来自患病受试者正常邻近组织的脾细胞和/或免疫细胞或其他淋巴细胞可提供对照抗体或TCR文库的呈现。这还允许将患病文库与未患病文库进行比较,以选择如本文一些实例中所述的淋巴细胞。这还可允许使用任何动物物种构建随后的B细胞抗体文库或T细胞TCR文库。

[0167] 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)。

[0168] 在一些情况下,起始材料可为组织样品,包括血管外组织或固体组织,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,并且特别是免疫细胞。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。

[0169] 在一些情况下,样品为流体,例如血液、唾液、淋巴或尿液。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。在一些情况下,样品不是血液样品。在一些情况下,样品不是流体样品。在一些情况下,样品为固体样品。

[0170] 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。

[0171] 在一些情况下,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。

[0172] 核酸分子包括但不限于脱氧核糖核酸(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。在一种情况下,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。在某些情况下,从单个细胞获得核酸分子。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,例如,血管外组织或实体瘤活检物获得。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。

[0173] 样品还包括从生物试样中提取的总RNA、cDNA文库、病毒或基因组DNA。在某些情况下,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)。

[0174] 一个或多个样品可来自一个或多个来源。一个或多个样品可来自两个或更多个来源。一个或多个样品可来自一个或多个受试者。一个或多个样品可来自两个或更多个受试者。一个或多个样品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。

[0175] 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。

[0176] 在一些情况下, 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分子结合。在一些情况下, 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))。

[0177] 多个样品可包括至少2、3、4、5、10、20、30、40、50、60、70、80、90或100个或更多个样品。多个样品可包括至少约100、200、300、400、500、600、700、800、900或1,000个或更多个样品。多个样品可包括至少约1,000; 2,000; 3,000; 4,000; 5000; 6,000; 7,000; 8,000; 9,000; 10,000; 100,000; 1,000,000个或更多个样品。例如, 多个样品可包括至少约10,000个样品。

[0178] 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。第一样品的一个或多个细胞与第二样品的一个或多个细胞可属于相同的细胞类型。第一样品的一个或多个细胞与所述多个样品的一个或多个不同细胞可属于不同的细胞类型。

[0179] 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, 即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。

[0180] 用于测序的多核苷酸

[0181] 所公开的方法包括多核苷酸分子(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, 所公开的方法包括两个或更多个多核苷酸分子的扩增和测序。在一个方面, 多核苷酸为RNA。在一个方面, 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。在优选的情况下,多核苷酸包括包含由免疫细胞产生的抗体的可变区的序列。在一些情况下,多核苷酸包括包含由免疫细胞产生的抗体或TCR $\alpha$ 链的重链可变区的序列。在一些情况下,多核苷酸包括包含由免疫细胞产生的抗体或TCR $\beta$ 链的轻链可变区的序列。

[0182] 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,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的片段以获得多核苷酸。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,多核苷酸为RNA分子。在一些情况下,多核苷酸为mRNA分子,或从mRNA分子产生的cDNA。在一些情况下,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,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,多核苷酸为编码来自单个免疫细胞的抗体或TCR序列的mRNA分子。在一些情况下,多核苷酸为编码来自单独免疫细胞的重链抗体或TCR $\alpha$ 链序列的mRNA分子。在一些情况下,多核苷酸为编码来自单独免疫细胞的重链抗体或TCR $\alpha$ 链序列的mRNA分子。在一些情况下,多核苷酸为编码来自单独免疫细胞的轻链抗体或TCR $\beta$ 链序列的mRNA分子。在一些情况下,多核苷酸为编码来自单独免疫细胞的轻链抗体或TCR $\beta$ 链序列的mRNA分子。在一些情况下,多核苷酸为编码来自单独免疫细胞的抗体或TCR可变序列的mRNA分子。在一些情况下,多核苷酸为编码来自单独免疫细胞的可变抗体或TCR序列的mRNA分子。在一些情况下,多核苷酸为编码来自单独免疫细胞的可变轻链抗体或TCR $\beta$ 链序列的mRNA分子。在一些情况下,多核苷酸为编码来自单独免疫细胞的可变轻链抗体或TCR $\beta$ 链序列的mRNA分子。在一些情况下,多核苷酸为编码来自单独免疫细胞的可变重链抗体或TCR $\alpha$ 链序列的mRNA分子。在一些情况下,多核苷酸为编码来自单个免疫细胞的可变重链抗体或TCR $\alpha$ 链序列的mRNA分子。在一些情况下,多核苷酸可为无细胞的核酸,例如DNA或RNA。

[0183] 在一些情况下,对多个Ig和/或TCR多核苷酸进行测序。例如,对多个V<sub>H</sub>和/或V<sub>L</sub>和/或V $\alpha$ 和/或V $\beta$ 和/或V $\gamma$ -V $\delta$ 多核苷酸进行测序。在一些情况下,对1、2、3、4、5、6、7、8、9或10个免疫球蛋白或TCR多核苷酸进行测序。在一些情况下,对至少约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、1x10<sup>6</sup>、2x10<sup>6</sup>、3x10<sup>6</sup>、4x10<sup>6</sup>、5x10<sup>6</sup>、6x10<sup>6</sup>、7x10<sup>6</sup>、8x10<sup>6</sup>、9x10<sup>6</sup>、1x10<sup>7</sup>、2x10<sup>7</sup>、3x10<sup>7</sup>、4x10<sup>7</sup>、5x10<sup>7</sup>、6x10<sup>7</sup>、7x10<sup>7</sup>、8x10<sup>7</sup>、9x10<sup>7</sup>、1x10<sup>8</sup>、2x10<sup>8</sup>、3x10<sup>8</sup>、4x10<sup>8</sup>、5x10<sup>8</sup>、6x10<sup>8</sup>、7x10<sup>8</sup>、8x10<sup>8</sup>、9x10<sup>8</sup>、1x10<sup>9</sup>、2x10<sup>9</sup>、3x10<sup>9</sup>、4x10<sup>9</sup>、5x10<sup>9</sup>、6x10<sup>9</sup>、7x10<sup>9</sup>、8x10<sup>9</sup>、9x10<sup>9</sup>、1x10<sup>10</sup>、2x10<sup>10</sup>、3x10<sup>10</sup>、4x10<sup>10</sup>、5x10<sup>10</sup>、6x10<sup>10</sup>、7x10<sup>10</sup>、8x10<sup>10</sup>、9x10<sup>10</sup>、1x10<sup>11</sup>、2x10<sup>11</sup>、3x10<sup>11</sup>、4x10<sup>11</sup>、5x10<sup>11</sup>、6x10<sup>11</sup>、7x10<sup>11</sup>、8x10<sup>11</sup>、9x10<sup>11</sup>、1x10<sup>12</sup>、2x10<sup>12</sup>、3x10<sup>12</sup>、4x10<sup>12</sup>、5x10<sup>12</sup>、6x10<sup>12</sup>、7x10<sup>12</sup>、8x10<sup>12</sup>或9x10<sup>12</sup>个免疫球蛋白或TCR多核苷酸进行测序。在一些情况下,对至多约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、 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 或 $9 \times 10^{12}$ 个免疫球蛋白或TCR多核苷酸进行测序。在一些情况下,对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、 $1-1 \times 10^5$ 、 $1-2 \times 10^5$ 、 $1-3 \times 10^5$ 、 $1-4 \times 10^5$ 、 $1-5 \times 10^5$ 、 $1-6 \times 10^5$ 、 $1-7 \times 10^5$ 、 $1-8 \times 10^5$ 、 $9 \times 10^5$ 、 $1-1 \times 10^6$ 、 $1-2 \times 10^6$ 、 $1-3 \times 10^6$ 、 $1-4 \times 10^6$ 、 $1-5 \times 10^6$ 、 $1-6 \times 10^6$ 、 $1-7 \times 10^6$ 、 $1-8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $1-2 \times 10^7$ 、 $1-3 \times 10^7$ 、 $1-4 \times 10^7$ 、 $1-5 \times 10^7$ 、 $1-6 \times 10^7$ 、 $1-7 \times 10^7$ 、 $1-8 \times 10^7$ 、 $1-9 \times 10^7$ 、 $1-1 \times 10^8$ 、 $1-2 \times 10^8$ 、 $1-3 \times 10^8$ 、 $1-4 \times 10^8$ 、 $1-5 \times 10^8$ 、 $1-6 \times 10^8$ 、 $1-7 \times 10^8$ 、 $1-8 \times 10^8$ 、 $1-9 \times 10^8$ 、 $1-1 \times 10^9$ 、 $1-2 \times 10^9$ 、 $1-3 \times 10^9$ 、 $1-4 \times 10^9$ 、 $1-5 \times 10^9$ 、 $1-6 \times 10^9$ 、 $1-7 \times 10^9$ 、 $1-8 \times 10^9$ 、 $1-9 \times 10^9$ 、 $1-1 \times 10^{10}$ 、 $1-2 \times 10^{10}$ 、 $1-3 \times 10^{10}$ 、 $1-4 \times 10^{10}$ 、 $1-5 \times 10^{10}$ 、 $1-6 \times 10^{10}$ 、 $1-7 \times 10^{10}$ 、 $1-8 \times 10^{10}$ 、 $1-9 \times 10^{10}$ 、 $1-1 \times 10^{11}$ 、 $1-2 \times 10^{11}$ 、 $1-3 \times 10^{11}$ 、 $1-4 \times 10^{11}$ 、 $1-5 \times 10^{11}$ 、 $1-6 \times 10^{11}$ 、 $1-7 \times 10^{11}$ 、 $1-8 \times 10^{11}$ 、 $1-9 \times 10^{11}$ 、 $1-1 \times 10^{12}$ 、 $1-2 \times 10^{12}$ 、 $1-3 \times 10^{12}$ 、 $1-4 \times 10^{12}$ 、 $1-5 \times 10^{12}$ 、 $1-6 \times 10^{12}$ 、 $1-7 \times 10^{12}$ 、 $1-8 \times 10^{12}$ 或 $1-9 \times 10^{12}$ 个免疫球蛋白或TCR多核苷酸进行测序。

[0184] 在一些情况下,测序的免疫球蛋白或TCR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。在一些情况下,测序的免疫球蛋白或TCR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。在一些情况下,测序的免疫球蛋白或TCR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。在一些情况下,测序的免疫球蛋白或TCR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、500-1,000、1,000-2,000、1,000-3,000、1,000-4,000、1,000-5,000、1,000-6,000、1,000-7,000、1,000-8,000、1,000-9,000、1,000-10,000、5,000-6,000、5,000-7,000、5,000-8,000、5,000-9,000或5,000-10,000个碱基或碱基对。在一些情况下,测序的

免疫球蛋白或TCR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、3、4、5、10、20、30、40、50、60、70、80、90、100千碱基。在一些情况下,来自相对较短模板的测序的免疫球蛋白或TCR多核苷酸为约40、45、50、55、60、65、70、75、80、85、90、95或100个碱基。在某些情况下,使用含有与疾病或病况相关联的序列或免疫球蛋白序列的数据库将测序数据与已知或预期的序列进行比对。

[0185] 在一个方面,公开了一种方法,其包括确定来自免疫细胞的多个多核苷酸中每一个的序列,例如序列文库。在一些情况下,该多核苷酸来自正常(即未患病)样品的免疫细胞(正常文库)。在一个方面,公开了一种方法,其包括确定来自浸润免疫细胞的多个多核苷酸中每一个的序列。该多核苷酸可来自患病样品的免疫细胞(患病文库)。

[0186] 单个细胞条码化

[0187] 在一些情况下,方法可包括确定患病样品的多个多核苷酸中每一个的序列以及确定正常样品的多个多核苷酸中每一个的序列。该方法可包括将来自患病样品的序列信息与正常样品的序列信息进行比较。例如,使用与将文库中的多核苷酸条码化的方法相结合的高通量测序技术,生物信息学可用于将数百万至数万亿的患病样品序列读取与正常样品进行比较。

[0188] 采用容器条码和/或分子条码进行单个细胞条码化。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:产生的容器含有每个容器2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。在一些情况下,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。在一些情况下,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。在一些情况下,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2个或更多个分子条码化多核苷酸。在一些情况下,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2个或更多个分子条码化多核苷酸。在一些情况下,当在溶液中时,可将容器条码和分子条码引入容器中。在一些情况下,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。

[0189] 在一些方面,单个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2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序列的 $V_H$ 链和 $V_L$ 链。一种具体的应用可为连接TCR序列的 $V\alpha$ 链和 $V\beta$ 链。一种具体的应用可为连接TCR序列的 $V\gamma$ 链和 $V\delta$ 链。

[0190] 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个细胞,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。

[0191] 单个细胞的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。在一些情况下,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。在一些情况下,可使用DNA。在一些情况下,可靶向超过2个RNA。

[0192] 在一些情况下,在逆转录期间,容器条码可与RNA相连接。在一些情况下,在逆转录期间,分子条码可与RNA相连接。在一些情况下,在逆转录期间,容器条码和分子条码可与RNA相连接。

[0193] 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。在一些情况下,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。在一些情况下,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。在一些情况下,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个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个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。

[0194] 条码

[0195] 条码可为分子条码或容器条码。在一些情况下,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。在一些情况下,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。

[0196] 如本文所用的,分子条码包含对于来自单个细胞或来自单个容器的单个分子、或来自两个或更多个单个细胞或来自2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,与来自不同单个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个细胞或单个容器的多核苷酸独特的信息。在一些情况下,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。在一些情况下,独特的信息不能用于鉴别多核苷酸的序列。例如,分子条码可与一个多核苷酸相附接,但该分子条码不能用于确定其附接的多核苷酸。在一些情况下,独特的信息不是与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多核苷酸中的哪一个。在一些情况下,独特的信息包含核苷酸的随机序列。在一些情况下,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。在一些情况下,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。在一些情况下,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。

[0197] 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。在一些情况下,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。

[0198] 在一些情况下,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。在一些情况下,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。在一些情况下,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。在一些情况下,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。

[0199] 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。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。在一些情况下,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。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。在一些情况下,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。

[0200] 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。在一些情况下,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。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。在一些情况下,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。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。在一些情况下,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。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。

[0201] 在一些情况下,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。在一些情况下,至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%的不同容器条码具有相同的浓度。在一些情况下,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。在一些情况下,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。

[0202] 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、 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 、 $9 \times 10^{12}$ 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、 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 、 $9 \times 10^{12}$ 个或更多个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、 $1 \times 10^6$ 、 $2 \times 10^6$ 、 $3 \times 10^6$ 、 $4 \times 10^6$ 、 $5 \times 10^6$ 、 $6 \times 10^6$ 、 $7 \times 10^6$ 、 $8 \times 10^6$ 、 $9 \times 10^6$ 、 $1 \times 10^7$ 、 $2 \times 10^7$ 、 $3 \times 10^7$ 、 $4 \times 10^7$ 、 $5 \times 10^7$ 、 $6 \times 10^7$ 、 $7 \times 10^7$ 、 $8 \times 10^7$ 、 $9 \times 10^7$ 、 $1 \times 10^8$ 、 $2 \times 10^8$ 、 $3 \times 10^8$ 、 $4 \times 10^8$ 、 $5 \times 10^8$ 、 $6 \times 10^8$ 、 $7 \times 10^8$ 、 $8 \times 10^8$ 、 $9 \times 10^8$ 、 $1 \times 10^9$ 、 $2 \times 10^9$ 、 $3 \times 10^9$ 、 $4 \times 10^9$ 、 $5 \times 10^9$ 、 $6 \times 10^9$ 、 $7 \times 10^9$ 、 $8 \times 10^9$ 、 $9 \times 10^9$ 、 $1 \times 10^{10}$ 、 $2 \times 10^{10}$ 、 $3 \times 10^{10}$ 、 $4 \times 10^{10}$ 、 $5 \times 10^{10}$ 、 $6 \times 10^{10}$ 、 $7 \times 10^{10}$ 、 $8 \times 10^{10}$ 、 $9 \times 10^{10}$ 、 $1 \times 10^{11}$ 、 $2 \times 10^{11}$ 、 $3 \times 10^{11}$ 、 $4 \times 10^{11}$ 、 $5 \times 10^{11}$ 、 $6 \times 10^{11}$ 、 $7 \times 10^{11}$ 、 $8 \times 10^{11}$ 、 $9 \times 10^{11}$ 、 $1 \times 10^{12}$ 、 $2 \times 10^{12}$ 、 $3 \times 10^{12}$ 、 $4 \times 10^{12}$ 、 $5 \times 10^{12}$ 、 $6 \times 10^{12}$ 、 $7 \times 10^{12}$ 、 $8 \times 10^{12}$ 、 $9 \times 10^{12}$ 个或更多个不同的序列。

[0203] 在一些情况下，一个或多个分子条码用于将序列分组或分箱。在一些情况下，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。在一些情况下，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。在一些情况下，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。

[0204] 在一些情况下，一个或多个容器条码用于将序列分组或分箱。在一些情况下，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。在一些情况下，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。在一些情况下，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个细胞的多核苷酸。

[0205] 在一些情况下,一个或多个分子条码和容器条码用于将序列分组或分箱。在一些情况下,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。在一些情况下,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。在一些情况下,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个细胞或容器。在一些情况下,一个或多个分子条码和容器条码不用于比对序列。

[0206] 在一些情况下,一个或多个分子条码不用于比对序列。在一些情况下,一个或多个分子条码用于比对序列。在一些情况下,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。在一些情况下,一个或多个容器条码不用于比对序列。在一些情况下,一个或多个容器条码用于比对序列。在一些情况下,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。在一些情况下,一个或多个分子条码和容器条码用于比对序列。在一些情况下,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。

[0207] 在一些情况下,比对的序列含有相同的分子条码。在一些情况下,比对的序列含有相同的容器条码。在一些情况下,比对的序列含有相同的分子条码和容器条码。在一些情况下,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。在一些情况下,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。在一些情况下,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个细胞或单个容器。

[0208] 小滴生成

[0209] 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。

[0210] 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%。

[0211] 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或TCR $\alpha$ 和TCR $\beta$ 链序列的文库或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或TCR $\alpha$ 和TCR $\beta$ 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。

[0212] 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。在一些情况下,区室为小滴。当在本说明书各处提及“小滴”时,除非另有说明,该术语与流体区室或容器以及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流体的混合物),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。

[0213] 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。

[0214] 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%。

[0215] 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。

[0216] 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。

[0217] 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 $\mu$ 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 $\mu$ m的浓度。在一些情况下,将氯化镁(MgCl<sub>2</sub>)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。在一些情况下,使用乙酸镁或镁。在一些情况下,使用硫酸镁。

[0218] 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0.1-1%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。

[0219] 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2.0μm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1.0μm。引物的浓度可为约0.5μm。在PCR中对于核酸浓度适用的范围包括但不限于约1pg至约500ng。

[0220] 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。

[0221] 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、Zonyl FSN。Pluronic F-68可以约0.5%w/v的浓度存在。

[0222] 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。

[0223] 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。

[0224] 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。

[0225] 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。

[0226] 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。

[0227] 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W02001/89788; W02006/040551; W02006/040554; W02004/002627; W0 2008/063227; W02004/091763; W02005/021151; W02006/096571; W02007/089541; W02007/081385和W02008/063227。

[0228] 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。

[0229] 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。在某些情况下,载流包含表面活性剂。

[0230] 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、脱氧核糖核苷三磷酸(dNTP)和一种或多种缓冲液。

[0231] 在某些情况下,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)的一种或多种形成流体隔室。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。在某些情况下,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。

[0232] 逆转录

[0233] 在一些情况下,通过逆转录从RNA制备用于测序的多核苷酸。在一些情况下,通过引物延伸,诸如使用聚合酶,从DNA制备用于测序的多核苷酸。

[0234] 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。例如,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。

[0235] 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。

[0236] 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。

[0237] 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。

[0238] 逆转录反应可在范围为5 $\mu$ L至100 $\mu$ L或10 $\mu$ L至20 $\mu$ L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。

[0239] 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。

[0240] 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。

[0241] 可使用一种或多种逆转录引物将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。在一些情况下,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。在一些情况下,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。

[0242] 在一些情况下,逆转录引物不包含条码。在一些情况下,逆转录引物包含条码。

[0243] 逆转录引物可进一步包含不与RNA的区域互补的区域。在一些情况下,不与RNA的区域互补的区域在与RNA互补的引物区域的5'侧。在一些情况下,不与RNA的区域互补的区域在与RNA互补的引物区域的3'侧。在一些情况下,不与RNA的区域互补的区域为5'突出端区域。在一些情况下,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。在一些情况下,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反应进行扩增,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。

[0244] 在一些情况下,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cDNA或条码化cDNA的V区段的第一和/或第二上游或下游区域互补的区域的第二正向/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第二正向/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第二正向/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(诸如免疫细胞,例如B细胞和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。

[0245] 在一些情况下,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cDNA或条码化cDNA的C区段的第一和/或第二上游或下游区域互补的区域的第二正向/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第二正向/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第二正向/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(诸如免疫细胞,例如B细胞和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。

[0246] 在一些情况下,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第二正向/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第二正向/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,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/反向引物中的引物可用来与样品中的细胞(诸如免疫细胞,例如B细胞和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。

[0247] 在一些情况下,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(诸如免疫细胞,例如B细胞和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。

[0248] 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。在一些情况下,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。在一些情况下,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。在一些情况下,不与RNA的区域互补的区域为5'突出端区域。在一些情况下,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。在一些情况下,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。在一些情况下,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。在一些情况下,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。在一些情况下,区域互补于RNA的区域,诸如恒定区或mRNA的聚A尾。

#### [0249] 扩增

[0250] 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。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1的Klenow片段、T7 DNA聚合酶、T4DNA聚合酶、Taq聚合酶、Pfu DNA聚合酶、Vent DNA聚合酶、噬菌体29、REDTaq<sup>TM</sup>、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℃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。

[0251] 可通过本领域已知的任何手段进行核酸的扩增。可通过聚合酶链反应(PCR)或等温DNA扩增来扩增核酸。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)、单个细胞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、PicoTiter 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、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)。RNA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 $\beta$ 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。

[0252] 在一些情况下,扩增不在固体支持体上发生。在一些情况下,扩增不在小滴中的固体支持体上发生。在一些情况下,当不在小滴中扩增时,扩增在固体支持体上发生。

[0253] 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。

[0254] 引物

[0255] 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。

[0256] 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多核苷酸分子的第二序列互补的反向引物,且第一基因座可位于第一序列与第二序列之间。在一些情况下,第一基因座包含V<sub>H</sub>序列。在一些情况下,第二基因座包含V <sub>$\alpha$</sub> 序列。在一些情况下,第二基因座包含V <sub>$\gamma$</sub> 序列。

[0257] 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多核苷酸分子的第二序列互补的反向引物,且第二基因座可位于第一序列与第二序列之间。在一些情况下,第二基因座包含V<sub>L</sub>序列。在一些情况下,第二基因座包含V<sub>B</sub>序列。在一些情况下,第二基因座包含V <sub>$\delta$</sub> 序列。

[0258] 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多核苷酸分子的第二序列互补的反向引物,且第三基因座可位于第一序列与第二序列之间。在一些情况下,第三基因座包含条码,诸如分子条码或容器条码。

[0259] 正向引物和反向引物的长度可取决于多核苷酸的序列和基因座。例如,可优化正向引物和反向引物的长度和/或T<sub>m</sub>。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。

[0260] 在与多核苷酸结合之前,引物可为单链DNA。在一些情况下,引物最初含有双链序列。短引物分子通常可需要更冷的温度以与多核苷酸形成足够稳定的杂合复合体。在一些情况下,引物不需要反映多核苷酸的准确序列,但可足够互补以与该多核苷酸杂交。在一些情况下,在与多核苷酸结合前,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。用于扩增给定多核苷酸的合适引物的设计是本领域公知的。

[0261] 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。本文可将与互补体足够互补以杂交的引物区域称作杂交区域。

[0262] 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,如美国公开号20090325169和20100167353中所述。

[0263] 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℃、2℃、3℃、4℃、5℃、6℃、7℃、8℃、9℃或10℃内。在一些情况下,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。

[0264] 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( $T_d = 2(A+T) + 4(G+C)$ )。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 $T_m$ 。

[0265] 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多核苷酸杂交的双链、单链或部分单链的多核苷酸。在与多核苷酸结合之前,引物可为单链DNA。在一些情况下,引物最初含有双链序列。引物位点包括引物杂交的多核苷酸区域。在一些情况下,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:具有与序列的5'端杂交的5'上游区的第一引物,以及具有与多核苷酸序列的3'端的互补体杂交的3'下游区的第二引物。引物组包含2个或

更多个引物:具有与一个多核苷酸序列或多个多核苷酸序列的5'端杂交的5'上游区的第一引物或第一多个引物,以及具有与所述一个多核苷酸序列或多个多核苷酸序列的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。在一些情况下,引物包含目标特异性序列。在一些情况下,引物包含样品条码序列。在一些情况下,引物包含通用引发序列。在一些情况下,引物包含PCR引发序列。在一些情况下,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)、 $\alpha$ MF、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。在一些情况下,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。在一些情况下,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。

[0266] 所述一个或多个引物可与多个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多核苷酸的内部区域退火。该内部区域可以距所述多个多核苷酸的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。在一些情况下,所述一个或多个定制引物与特异性区域、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或一个或多个多核苷酸的任何其他扩增方法。

[0267] 与引物结合的多核苷酸的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。在一些情况下,与引物结合的多核苷酸的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。

[0268] 可根据已知参数设计引物,以便避免二级结构和自身杂交。在一些情况下,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。在一些情况下,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。在一些情况下,多个引物中的一个或多个引物可在与所述多个引物中另一个引物不同的温度下退火和解链。

[0269] 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、至多约或至少约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个相同或不同的引物。例如,多个引物中的每个引物可包含与多核苷酸的特异性区域结合的相同或不同的序列。

[0270] 酶

[0271] 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。

[0272] 在一些情况下,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。

[0273] 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。在一些情况下,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。在一些情况下,该逆转录酶为M-MLV逆转录酶。

[0274] 在一些情况下,本文所公开的方法和试剂盒包括使用一种或多种蛋白酶。

[0275] 在一些情况下,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。在一些情况下,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™ 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、T4DNA聚合酶、T7DNA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。

[0276] 在一些情况下,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、phi6RNA聚合酶(RdRP)、聚(U)聚合酶、SP6RNA聚合酶和T7RNA聚合酶。

[0277] 附加试剂

[0278] 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。

[0279] 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®、Sephacrose®、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)。在一些情况下,固相基本上为亲水的。在一些情况下,固相(例如,珠子)基本上为疏水的。在一些情况下,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。在一些情况下,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。在一些情况下,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150nm至10 $\mu$ m的珠子。还可使用其他大小。

[0280] 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。在一些情况下,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。

[0281] 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)的亲合力。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)的亲合力。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、短聚腺苷酸化合成核糖核苷酸和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。在一些情况下,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(*E.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。

[0282] 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。核酸对照可为已知浓度。对照剂可包含一种或多种标记。

[0283] 掺加对照可为向反应或样品中添加的模板。例如,掺加多核苷酸可添加到扩增反应中。掺加多核苷酸可在第一扩增循环后的任何时间添加到扩增反应中。在一些情况下,掺加多核苷酸在循环数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。掺加多核苷酸可在最终扩增循环前的任何时间添加到扩增反应中。掺加多核苷酸可包含一种或多种核苷酸或酸碱基对。掺加多核苷酸可包含DNA、RNA或其任意组合。掺加多核苷酸可包含一种或多种标记。

[0284] 从测序信息中选择淋巴细胞

[0285] 所公开的方法进一步包括基于测序信息的分析选择淋巴细胞或其多核苷酸。所述选择可包括分析从免疫测序步骤获得的测序数据,例如对测序数据进行生物信息学分析。可基于测序信息中包含的一个或多个参数或一条或多条信息选择淋巴细胞或其多核苷酸。可用于选择淋巴细胞的测序信息(即淋巴细胞的抗体或TCR多核苷酸序列)中包含的示例性参数或信息包括但不限于抗体或TCR多核苷酸序列的表达量,抗体或TCR多核苷酸序列的突变水平或模式,包含抗体或TCR多核苷酸序列的TIL在患病组织中相比于在正常(未患病)组织例如正常邻近组织中的富集,抗体或TCR多核苷酸序列的同种型或同种型谱,抗体或TCR多核苷酸序列的系统发生簇,抗体或TCR多核苷酸序列的系统发生簇的大小,来自多个患有相同疾病的患者的样品之间的抗体或TCR多核苷酸序列相关性,来自多个患有相同疾病的患者的样品之间的抗体或TCR多核苷酸序列相似性(或缺乏相似性),及其组合。

[0286] 选择抗体或TCR多核苷酸序列可包括准确有效地对多个多核苷酸进行测序,以鉴别多核苷酸序列或包含用于选择的多核苷酸序列的淋巴细胞。在一些情况下,该方法包括从肿瘤浸润性淋巴细胞选择多核苷酸序列。在一些情况下,该方法包括选择编码靶向疾病相关或疾病特异性多肽的候选多肽的多核苷酸序列。例如,该方法可包括从靶向癌基因表达产物的B细胞选择编码免疫球蛋白的 $V_H$ 或 $V_L$ 的多核苷酸。例如,该方法可包括鉴别编码靶向疾病相关或疾病特异性抗原的T细胞的TCR的 $V\alpha$ 或 $V\beta$ 的多核苷酸。

[0287] 在一些情况下,所述方法包括从B细胞选择多核苷酸序列。例如,所选的多核苷酸可编码来自肿瘤浸润B细胞的候选多肽。在一些情况下,候选多肽包含抗体或其片段。例如,候选多肽可包含抗体的可变域。在一些情况下,候选多肽包含免疫球蛋白重链。在一些情况下,候选多肽包含免疫球蛋白轻链。在一些情况下,候选多肽包含免疫球蛋白重链和免疫球蛋白轻链。在一些情况下,候选多肽包含 $V_H$ 结构域。在一些情况下,候选多肽包含 $V_L$ 结构域。在一些情况下,该方法包括选择包含 $V_H$ 序列和 $V_L$ 序列的多核苷酸。例如,该方法可包括选择包含来自单个B细胞的 $V_H$ 序列和 $V_L$ 序列的多核苷酸。

[0288] 在一些情况下,所述方法包括选择包含来自第一样品的 $V_H$ 序列和来自第二样品的 $V_L$ 序列的多核苷酸。在一些情况下,该方法包括选择包含来自第一样品的 $V_H$ 序列和来自第二样品的多个 $V_L$ 序列的多核苷酸。在一些情况下,该方法包括选择包含来自第一样品的 $V_H$ 序列和来自多个第二样品的 $V_L$ 序列的多核苷酸。

[0289] 在一些情况下,所述方法包括从T细胞选择多核苷酸序列。例如,所选的多核苷酸可编码来自肿瘤浸润T细胞的候选多肽。在一些情况下,候选多肽为TCR或其片段。例如,候选多肽可包含TCR的可变域。在一些情况下,候选多肽包含TCR $\alpha$ 链。在一些情况下,候选多肽包含TCR $\beta$ 链。在一些情况下,候选多肽包含TCR $\gamma$ 链。在一些情况下,候选多肽包含TCR $\delta$ 链。在一些情况下,候选多肽包含TCR的 $V\alpha$ 结构域。在一些情况下,候选多肽包含TCR的 $V\beta$ 结构域。在一些情况下,候选多肽包含TCR的 $V\gamma$ 结构域。在一些情况下,候选多肽包含TCR的 $V\delta$ 结构域。在一些情况下,该方法包括选择包含 $V\alpha$ 序列和 $V\beta$ 序列的多核苷酸。例如,该方法可包括选择包含来自单个T细胞的 $V\alpha$ 序列和 $V\beta$ 序列的多核苷酸。在一些情况下,该方法包括选择

包含V $\gamma$ 序列和V $\delta$ 序列的多核苷酸。例如,该方法可包括选择包含来自单个T细胞的V $\gamma$ 序列和V $\delta$ 序列的多核苷酸。

[0290] 所公开的方法包括基于测序信息选择免疫细胞或其多核苷酸,如肿瘤浸润性淋巴细胞或其多核苷酸。可通过基于测序信息选择浸润免疫细胞的多核苷酸序列来选择免疫细胞或其多核苷酸,如浸润免疫细胞或其多核苷酸。可通过确定浸润免疫细胞多核苷酸的序列,例如通过对来自包含浸润免疫细胞的组织样品的多个免疫细胞进行高通量测序来选择浸润免疫细胞的多核苷酸。由于与非浸润免疫细胞的数目相比,样品中浸润免疫细胞的绝对数目较少和/或样品中浸润免疫细胞的数目较少,因此本文提供的对用于选择浸润免疫细胞的多核苷酸进行测序的方法通常使用高通量测序。可对包含一种或多种浸润免疫细胞的组织样品进行测序,而无需在测序步骤之前提取所述一种或多种浸润免疫细胞。

[0291] 在一些情况下,来自样品的选定的淋巴细胞的测序的多核苷酸可以以不同的浓度或量(例如,不同的分子数目)存在于样品中。例如,来自选定的淋巴细胞的测序的一种多核苷酸的浓度或量可小于或大于来自样品中的淋巴细胞的测序的另一种多核苷酸的浓度或量。例如,来自选定的淋巴细胞的测序的一种多核苷酸的浓度或量可比来自样品中的淋巴细胞的测序的至少一种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。例如,来自选定的淋巴细胞的测序的一种多核苷酸的浓度或量可比来自样品中的淋巴细胞的测序的至少一种多核苷酸的浓度或量大至少约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。在一些情况下,样品中至少一种多核苷酸的浓度或量比至少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%的测序多核苷酸的浓度或量少至少约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,样品中一种多核苷酸的浓度或量小于另一种多核苷酸的浓度或量。

[0292] 在一些疾病中,由于正在进行的体细胞超突变或由于疾病相关的体细胞突变,如碱基置换、异常重排等,单个淋巴细胞祖细胞可产生许多相关的淋巴细胞子代,每个淋巴细胞子代具有和/或表达稍有差异的TCR或抗体,因此是不同的系统发生克隆(phylogenetic clone)。一组系统发生克隆,如相关的系统发生克隆,可称为系统发生簇。在一个方面,选择抗体或TCR多核苷酸序列包括基于系统发生克隆的频率进行选择。在一个方面,选择抗体或TCR多核苷酸序列包括基于系统发生簇的频率(即,该簇的组成系统发生克隆型的频率的总和)而非单个系统发生克隆的频率进行选择。

[0293] 可通过与亲本克隆相关的一个或多个测量来鉴别系统发生克隆。在一种情况下,例如,可通过百分比同源性将系统发生克隆分组至相同的簇中。在另一种情况下,通过V区、J区和/或D区的常见用法来鉴别系统发生克隆或系统发生簇。例如,可通过具有共同的J区和D区但具有不同的V区的克隆来定义簇;或者可通过具有相同的V区和J区但具有不同的D区的克隆来定义簇;或者可通过经历了1-10个碱基、或1-5个碱基或1-3个碱基的一个或多个插入和/或缺失的克隆来定义簇,从而产生簇成员。可将单个样品的系统发生克隆分组为簇,并且可将来自在不同时间获取的连续样品的簇相互进行比较。在本发明的一个方面,在

从各时间点的样品确定的克隆中鉴别含有与疾病如癌症相关的克隆的簇。可将来自该时间点的相关克隆的簇与先前样品的克隆簇进行比较,以选择Ig或TCR多核苷酸,例如,在连续簇中确定特定克隆的频率是增加还是降低,是否出现已知来自群体研究或待关联数据库等的新相关克隆。

[0294] 在一些情况下,基于测序信息的分析选择淋巴细胞或其多核苷酸可包括基于测序信息的分析根据抗体或TCR多核苷酸的同种型进行选择。例如,基于测序信息的分析选择淋巴细胞或其多核苷酸可包括选择IgG、IgM、IgA、IgE或IgD抗体。例如,基于测序信息的分析选择淋巴细胞或其多核苷酸可包括选择IgG抗体。

[0295] 由选定淋巴细胞的多核苷酸编码的多肽的克隆与表达

[0296] 重组与合成方法以及组合物

[0297] 可使用合成和/或重组方法以及组合物产生由来自选定的淋巴细胞的多核苷酸编码的抗体和TCR(参见例如,美国专利号4,816,567)。在一些情况下,提供了编码多肽的分离的选定多核苷酸。这样的核酸可编码包含例如抗体的 $V_L$ 的氨基酸序列和/或包含抗体的 $V_H$ 的氨基酸序列。在另一种情况下,提供了包含这样的核酸的一种或多种载体。“载体”是能够使与其连接的另一核酸增殖的核酸分子。该术语包括作为自我复制的核酸结构的载体以及整合到其被引入的宿主细胞的基因组中的载体。某些载体能够指导与其可操作地连接的核酸的表达。

[0298] 在另一种情况下,提供了包含这样的核酸的宿主细胞。宿主细胞是已被引入外源核酸的细胞,包括此类细胞的子代。宿主细胞包括“转化体”和“转化的细胞”,其包括初级转化的细胞和由其衍生的子代,而不考虑传代次数。子代在核酸内含物方面不能与亲本细胞完全相同,但可含有突变。本文包括如在最初转化的细胞中筛选或选择的具有相同功能或生物学活性的突变子代。在一个这样的实例中,宿主细胞包含(例如已被转化具有)含有编码包含抗体的 $V_L$ 的氨基酸序列和包含抗体的 $V_H$ 的氨基酸序列的核酸的载体,或含有编码包含抗体的 $V_L$ 的氨基酸序列的核酸的第一载体和含有编码包含抗体的 $V_H$ 的氨基酸序列的核酸的第二载体。在一些情况下,宿主细胞是真核细胞,例如,中国仓鼠卵巢(CHO)细胞或淋巴样细胞(例如Y0、NS0、Sp20细胞)。在一些情况下,提供了制备由选定的多核苷酸编码的多肽的方法,其中该方法包括在适用于表达所述多肽的条件下培养包含编码所述多肽的选定核酸的宿主细胞,并且任选地从所述宿主细胞或宿主细胞培养基回收所述多肽。

[0299] 为了重组产生由选定的多核苷酸编码的多肽,将编码由选定的多核苷酸编码的多肽例如抗体的分离的核酸插入到一个或多个载体中,用于在宿主细胞中进一步克隆和/或表达。可使用常规程序容易地对这样的核酸进行分离和测序。

[0300] 用于克隆或表达编码多肽的载体的合适宿主细胞包括本文所述的原核或真核细胞。例如,可以例如在不需要糖基化和Fc效应子功能时,在细菌中产生由选定的多核苷酸编码的多肽(参见例如,美国专利号5,648,237、5,789,199和5,840,523;Charlton,Methods in Molecular Biology,Vol.248,pp.245-254(2003))。表达后,可从细菌细胞糊中以可溶级分形式分离由选定的多核苷酸编码的多肽,并且可将其进一步纯化。

[0301] 除原核生物之外,真核微生物如丝状真菌或酵母是编码多肽的载体的合适克隆或表达宿主(参见例如,Gerngross,Nat.Biotech.22:1409-1414(2004);以及Li等人,Nat.Biotech.24:210-215(2006))。用于表达糖基化多肽例如抗体的合适宿主细胞还来源

于多细胞生物体,包括无脊椎动物和脊椎动物。无脊椎动物的实例包括植物和昆虫细胞(参见例如,美国专利号5,959,177、6,040,498、6,420,548、7,125,978和6,417,429)。脊椎动物细胞的实例包括哺乳动物细胞系,经SV40转化的猴肾CV1系(COS-7);人胚胎肾系(如例如在Graham等人,J.Gen Virol.36:59(1977)中描述的293或293细胞);幼仓鼠肾细胞(BHK);小鼠支持细胞(TM4细胞);猴肾细胞(CV1);非洲绿猴肾细胞(VERO-76);人宫颈癌细胞(HELA);犬肾细胞(MDCK);水牛鼠肝细胞(BRL 3A);人肺细胞(W138);人肝细胞(Hep G2);小鼠乳房肿瘤(MMT 060562);TR1细胞;MRC 5细胞;FS4细胞;中国仓鼠卵巢(CHO)细胞,包括DHFR CHO细胞;以及骨髓瘤细胞系如Y0、NS0和Sp2/0。(参见例如,Yazaki和Wu,Methods in Molecular Biology,Vol.248,pp.255-268(2003))。

[0302] 术语“抗体表达文库”、“TCR表达文库”、“重组抗体文库”、“重组TCR文库”、“合成抗体文库”和“合成TCR文库”是指来自两种或更多种选定的淋巴细胞的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,这些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),或可指在已在适当的表达系统中重组产生(例如,表达)后或者例如使用肽合成器合成后的抗体或TCR分子的集合(即,在蛋白质水平上)。表达载体文库可被包含在可表达它们的合适的宿主细胞内。在表达文库中编码或表达的抗体或TCR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抗体或TCR、Fab'抗体或TCR、(Fab')<sub>2</sub>片段、双抗体等。如“编码”特定多肽的核酸序列或特定多肽的DNA编码序列或“编码”特定多肽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,例如启动子序列的控制下时转录并翻译为多肽的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。在启动子序列内存在转录起始位点(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)。

[0303] 由抗体或TCR表达或合成文库鉴别、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,该核酸进而可被合成或并入适当的表达载体中和/或包含在合适的宿主细胞内。

[0304] cDNA集合体可同与抗体或TCR基因的重链或TCR $\alpha$ 链的恒定区杂交的多核苷酸以及同与抗体或TCR基因的V<sub>H</sub>链或V $\alpha$ 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TCR $\alpha$ 链的恒定区杂交的多核苷酸以及同与包含抗体或TCR序列的条码化多核苷酸的V<sub>H</sub>或V $\alpha$ 链区域5'端的5'侧区域杂交的多核苷酸一起经历PCR反应。还设置PCR反应用于扩增例如 $\kappa$ 和 $\lambda$ 类别的V<sub>L</sub>或V $\beta$ 链集合体。cDNA集合体可同与抗体或TCR基因的轻链或TCR $\beta$ 链的恒定区杂交的多核苷酸以及同与抗体或TCR基因的V<sub>L</sub>或TCR $\beta$ 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基因的轻链或TCR $\beta$ 链的恒定区杂交的多核苷酸以及同与包含抗体或TCR序列的条码化多核苷酸的V<sub>L</sub>或V $\beta$ 链区域5'端的5'侧区域杂交的多核苷酸一起经历PCR反应。可基于免疫球蛋白基因序列数据库信息来设计这类寡核苷酸或引物。

[0305] 在一些情况下,可从使用对重链或轻链或TCR $\alpha$ 和TCR $\beta$ 基因,特别是对V<sub>H</sub>和V<sub>L</sub>或V $\alpha$ 和

VB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 $V_\alpha$ 和 $V_\beta$ 序列的文库便利地获得 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 序列。在一些情况下,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 $V_\alpha$ 和 $V_\beta$ 序列的文库便利地获得 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 序列。在一些情况下,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 $V_\alpha$ 和 $V_\beta$ 序列的文库便利地获得 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 序列。在一些情况下,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 $V_\alpha$ 和 $V_\beta$ 序列的文库便利地获得 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 序列。在一些情况下,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 $V_\alpha$ 和 $V_\beta$ 序列的文库便利地获得 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 序列。

[0306] 在一些情况下,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 $V_\alpha$ 和 $V_\beta$ 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。在一些情况下,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 $V_L$ 或非天然的 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 序列的文库获得 $V_H$ 和 $V_L$ 或非天然的 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 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 $V_L$ 或非天然的 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 序列。(天然或非天然的 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 序列可进行连接,诸如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)。对于免疫球蛋白或TCR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,在单个细胞PCR后;在哺乳动物或其他细胞表面展示后,在FACS筛选后,等等),对天然或非天然的 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 区域进行测序。除其他原因外,测序可用于证实多样性水平为可接受的水平。测序包括但不限于高通量测序、深度测序或二者的组合。

[0307] 在一些情况下,不必要使用本文所述的表达或合成方法来物理连接天然或非天然的 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 组合。在一些情况下,编码来自选定淋巴细胞的多肽的cDNA未物理连接。在一些情况下,cDNA、条码化多核苷酸或PCR扩增的条码化cDNA未在相同的表达载体中物理连接。

[0308] 在一些情况下,除了cDNA引物之外,还使用针对 $V_H$ 或 $V_\alpha$ 区基因的5'端的一个引物或多个引物和针对 $V_L$ 或 $V_\beta$ 基因的5'端的另一个引物或多个引物,对天然或非天然的 $V_H$ 和 $V_L$ 或 $V_\alpha$ 和 $V_\beta$ 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 $V_\beta$ 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。

[0309] 在期望提供附加的 $V_H$ 和 $V_L$ 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 组合的一些情况下,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。

[0310] 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,在一些情况下,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Ig链和轻Ig链或 $\alpha$ 和 $\beta$ TCR链可变区不是人源的,且其中恒定区是人源的。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。在优选的情况下,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多肽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和TCR。生长、筛选和表征详细描述如下。

[0311] 一旦生成了表达载体的文库,则编码的抗体或TCR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。

[0312] 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,单独编码抗体或TCR片段的氨基酸序列的多核苷酸,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,编码抗体或TCR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。

[0313] 任选地,初级PCR产物然后可经历采用与抗体或TCR可变域的5'和3'端杂交的新多核苷酸组的次级PCR反应。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。这样的次级PCR反应的产物为多种可变抗体或TCR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的可变域。

[0314] PCR产物还可经历采用与多核苷酸的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。

[0315] 包含衍生自淋巴细胞的可变区或其片段的这类克隆片段组库的文库构成了本发明的进一步的方面。这些包含克隆的可变区的文库可以任选地插入表达载体中以形成表达文库。

[0316] 在一些情况下,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链的

全部或部分恒定区。这在表达文库形式为Fab形式时是理想的。同样,包含这些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。

[0317] 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,除了编码序列之外,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。

[0318] 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)。

[0319] 例如,可将 $V_H$ 和 $V_L$ 链或 $V\alpha$ 和 $V\beta$ 链或 $V\gamma$ 和 $V\delta$ 链克隆到表达载体中,用于以例如完整的人IgG形式在例如293K细胞中表达。例如,可将100-500个Ig或TCR链克隆到表达载体中,用于在细胞中表达。

[0320] 在一些实施方案中,与测序并行地,例如,可从容器中回收 $V_H$ 和 $V_L$ 链或 $V\alpha$ 和 $V\beta$ 链或 $V\gamma$ 和 $V\delta$ 链的文库,并且可将该文库克隆到表达载体中并进行共转染,例如以供酵母展示筛选。与在开始时就分裂生物样品相比,克隆此一致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。例如,人源的 $V_H$ 和 $V_L$ 链或 $V\alpha$ 和 $V\beta$ 链的文库不论正确或错误的 $V_H$ 和 $V_L$ 成对匹配或 $V\alpha$ 和 $V\beta$ 成对匹配均可表达。例如,随后可针对一个或多个抗原靶标进行酵母展示筛选,以富集潜在的抗体或TCR候选物。可对从展示技术如酵母展示得到的阳性候选抗体和TCR进行测序,并可查询候选抗体和TCR的配体。

[0321] 在一些实施方案中,可使用由Kohler等人,Nature,256:495 (1975) 首次描述的杂交瘤方法制备单克隆抗体,或者可通过重组DNA方法(U.S.4,816,567)制备单克隆抗体。在杂交瘤方法中,如上所述对小鼠或其他合适的宿主动物如仓鼠进行免疫,以得到产生或能够产生将会与用于免疫的蛋白质特异性结合的抗体的淋巴细胞。或者,可在体外对淋巴细胞进行免疫。然后使用合适的融合剂如聚乙二醇使淋巴细胞与骨髓瘤细胞融合,以形成杂交瘤细胞(Goding,Monoclonal Antibodies:Principles and Practice,pp.59-103 (Academic Press,1986))。接种如此制备的杂交瘤细胞并使其在合适的培养基中生长,该培养基优选含有一种或多种抑制未融合的亲本骨髓瘤细胞的生长或存活物质。例如,如果亲本骨髓瘤细胞缺乏酶——次黄嘌呤鸟嘌呤磷酸核糖转移酶(HGPRT或HPRT),则用于杂交瘤的培养基通常将包括次黄嘌呤、氨基蝶呤和胸苷(HAT培养基),这些物质阻止HGPRT缺陷型细胞的生长。优选的骨髓瘤细胞是那些有效融合,支持由选定的产生抗体的细胞稳定高水平地产生抗体,并对培养基如HAT培养基敏感的细胞。在这些细胞中,优选的骨髓瘤细胞系是鼠骨髓瘤系,如衍生自可从美国加州圣地亚哥索尔克研究所细胞分配中心(Salk Institute Cell Distribution Center, San Diego, California USA)获得的MOPC-21和MPC-11小鼠肿瘤以及可从美国马里兰州罗克维尔美国典型培养物保藏中心(American Type Culture Collection, Rockville, Maryland USA)获得的SP-2或X63-Ag8-653细胞的骨髓瘤细胞系。还描述了人骨髓瘤和小鼠-人异源骨髓瘤细胞系,用于产生人单克隆抗体(Kozbor, J. Immunol., 133:3001 (1984); Brodeur等人, Monoclonal Antibody Production Techniques and Applications, pp.51-63 (Marcel Dekker, Inc., New York, 1987))。对其

中生长杂交瘤细胞的培养基中针对抗原的单克隆抗体的产生进行测定。优选地,通过免疫沉淀或通过体外结合测定如放射免疫测定(RIA)或酶联免疫吸附测定(ELISA)来确定由杂交瘤细胞产生的单克隆抗体的结合特异性。可以例如通过Munson等人,Anal.Biochem., 107:220(1980)的斯卡查德分析(Scatchard analysis)来确定单克隆抗体的结合亲和力。在鉴定出产生具有期望特异性、亲和力和/或活性的抗体的杂交瘤细胞后,可通过有限稀释程序对这些克隆进行亚克隆并通过标准方法使其生长(Goding, Monoclonal Antibodies: Principles and Practice, pp.59-103(Academic Press, 1986))。用于此目的的合适培养基包括,例如,D-MEM或RPMI-1640培养基。另外,杂交瘤细胞可作为动物中的腹水瘤在体内生长。通过常规的免疫球蛋白纯化程序例如蛋白A-Sepharose、羟基磷灰石层析、凝胶电泳、透析或亲和层析从培养基、腹水或血清恰当地分离由亚克隆分泌的单克隆抗体。容易使用常规程序(例如,通过使用能够与编码鼠抗体重链和轻链的基因特异性结合的寡核苷酸探针)对编码单克隆抗体的DNA进行分离和测序。一旦分离,可将DNA置于表达载体中,然后将表达载体转染到宿主细胞如大肠杆菌细胞、猿COS细胞、中国仓鼠卵巢(CHO)细胞或不另外产生免疫球蛋白的骨髓瘤细胞中,以获得重组宿主细胞中单克隆抗体的合成。关于在编码抗体的DNA的细菌中重组表达的综述文章包括Skerra等人,Curr.Opinion in Immunol., 5: 256-262(1993)和Pliickthun, Immunol.Revs., 130:151-188(1992)。在另一种情况下,可使用在McCafferty等人,Nature, 348:552-554(1990)中描述的技术从生成的抗体噬菌体文库分离抗体或抗体片段。Clackson等人,Nature, 352:624-628(1991)和Marks等人,J.Mol.Biol., 222:581-597(1991)分别描述了使用噬菌体文库分离鼠抗体和人抗体。随后的出版物描述了通过链改组(Marks等人,Bio/Technology, 10:779-783(1992)),以及作为用于构建超大型噬菌体文库的策略的组合感染和体内重组(Waterhouse等人,Nuc.Acids.Res., 21:2265-2266(1993))来产生高亲和力(nM范围)人抗体。因此,这些技术是用于分离单克隆抗体的传统单克隆抗体杂交瘤技术的可行替代方案。例如,还可通过用人重链和轻链恒定域的编码序列取代同源鼠序列(U.S.4,816,567;Morrison等人,Proc.Natl Acad.Sci.USA, 81:6851(1984))或通过将免疫球蛋白编码序列与非免疫球蛋白多肽的编码序列的全部或部分共价连接,来对DNA进行修饰。典型地,这样的非免疫球蛋白多肽代替了抗体恒定域,或者其代替了抗体的一个抗原结合位点的可变域,以产生包含对抗原具有特异性的一个抗原结合位点和对不同抗原具有特异性的另一个抗原结合位点的嵌合二价抗体。

[0322] 或者,可使用噬菌体展示技术(McCafferty等人,Nature, 348:552-553(1990))从免疫球蛋白或TCR可变(V)结构域基因组库在体外产生人抗体或TCR以及抗体片段或TCR。根据该技术,抗体或TCR V结构域基因被框内克隆到丝状噬菌体的主要或次要外壳蛋白基因如M13或fd中,并展示为在噬菌体颗粒的表面上的功能性抗体或TCR片段。因为丝状颗粒含有噬菌体基因组的单链DNA拷贝,所以基于抗体或TCR的功能特性的选择还导致编码显示出这些特性的抗体或TCR的基因的选择。因此,噬菌体模拟了B细胞或T细胞的一些特性。噬菌体展示可以以多种形式进行;有关它们的综述参见例如,Johnson和Chiswell,Current Opinion in Structural Biology, 3:564-571(1993)。可构建V基因组库,并且可基本按照由Marks等人,J.Mol.Biol., 222:581-597(1991),或Griffith等人,EMBO J., 12:725-734(1993)描述的技术分离针对各种抗原阵列(包括自身抗原)的抗体。还参见U.S.5,565,332

和5,573,905。还可通过体外活化的B细胞生成抗体(参见U.S.5,567,610和5,229,275)。

[0323] 已经开发了多种技术,用于产生抗体或TCR片段。传统上,这些片段经由完整抗体或TCR的蛋白水解消化而获得(参见例如,Morimoto等人,J.Biochem.Biophys.Methods,24:107-117(1992)和Brennan等人,Science,229:81(1985))。然而,这些片段现在可直接通过重组宿主细胞产生。例如,可从上文讨论的抗体噬菌体文库分离抗体或TCR片段。

[0324] 在一些实施方案中,抗体或TCR可变域与免疫球蛋白恒定域序列融合。在一些实施方案中,融合物包含含有铰链区、CH2区和CH3区的至少一部分的免疫球蛋白重链恒定域。在一些实施方案中,融合物包含存在于至少一个融合物中的含有轻链结合所必需的位点的第一重链恒定区(CH1)。编码免疫球蛋白重链融合物以及如果需要还编码免疫球蛋白轻链的DNA可插入到分开的表达载体中,并共转染到合适的宿主生物体中。这在当构建中使用的不同比例的多肽链提供最佳产量的情况下,为调节多肽链的相互比例提供了巨大的灵活性。然而,当相同比例的至少两条多肽链的表达导致高产量时或者当比例没有特别意义时,将多肽链的编码序列插入到一个表达载体中是可能的。

[0325] 根据U.S.5,731,168中描述的另一种方法,可将一对抗体分子之间的界面工程化以使从重组细胞培养物回收的异二聚体的百分比最大化。优选的界面包含抗体或TCR恒定域的CH3结构域的至少一部分。在该方法中,来自第一抗体或TCR分子的界面的一个或多个小氨基酸侧链被更大的侧链(例如,酪氨酸或色氨酸)取代。通过用较小的氨基酸侧链(例如,丙氨酸或苏氨酸)取代大的氨基酸侧链,在第二抗体或TCR分子的界面上产生与大侧链相同或相似大小的补偿“腔”。这提供了相对于其他不需要的终产物如同二聚体增加异二聚体的产量的机制。

[0326] 可以例如通过饱和硫酸铵沉淀、免疫球蛋白沉淀法、己酸法、辛酸法、离子交换层析(DEAE或DE52)或使用抗-Ig柱或蛋白A、G或L柱的亲和层析从培养物上清液或其他培养物中分离并纯化抗体和TCR。

[0327] 在另一方面,编码一个或多个CDR的氨基酸序列的核苷酸可以例如通过重组技术插入到编码抗体、抗原结合片段或结合蛋白的现有多核苷酸的限制性内切核酸酶位点中。

[0328] 对于高水平生产,最广泛使用的哺乳动物表达系统是利用二氢叶酸还原酶缺陷型(“dhfr-”)中国仓鼠卵巢细胞提供的基因扩增程序的表达系统。该系统是技术人员公知的。该系统基于二氢叶酸还原酶“dhfr”基因,该基因编码催化二氢叶酸转变为四氢叶酸的DHFR酶。为了实现高产量,用含有功能性DHFR基因的表达式载体与编码期望蛋白质的基因一起转染dhfr-CHO细胞。

[0329] 通过增加竞争性DHFR抑制剂甲氨蝶呤(MTX)的量,通过扩增dhfr基因使重组细胞发展出抗性。在标准情况下,所采用的扩增单位远大于dhfr基因的大小。

[0330] 当需要蛋白质如抗体或TCR链的大规模生产时,考虑所采用的细胞的表达水平和稳定性。在长期培养中,重组CHO细胞群体在扩增期间失去了关于其特异性抗体或TCR生产力的均一性,即使它们衍生自单个亲本克隆。

[0331] 提供了一种组合物,其包含编码来自选定淋巴细胞的抗体、TCR或其抗原结合片段的分离的多核苷酸,含有此类多核苷酸的载体,以及用于将此类多核苷酸转录并翻译成多肽的宿主细胞和表达系统。

[0332] 本申请还提供了包含如上所述的至少一种多核苷酸的质粒、载体、转录盒或表达

盒形式的构建体。

[0333] 本申请还提供了包含如上所述的一种或多种构建体的重组宿主细胞。编码来自选定淋巴细胞的任何抗体、TCR或其抗原结合片段的核酸本身构成本申请的一个方面,同样抗体、TCR或其抗原结合片段的的生产方法也构成本申请的方面,所述方法包括从编码来自选定淋巴细胞的抗体、TCR或其抗原结合片段的核酸进行表达。可通过在适当条件下培养含有该核酸的重组宿主细胞来实现表达。在通过表达产生后,抗体、TCR或其抗原结合片段可使用任何合适的技术进行分离和/或纯化,然后视情况用于例如验证。

[0334] 来自选定淋巴细胞的特异性抗体、TCR、抗原结合片段和编码核酸分子以及载体可以以基本纯的或均一的形式,或者在核酸,不含或基本不含除编码具有所需功能的多肽的序列以外的核酸或基因来源的情况下提供为例如从其天然环境分离和/或纯化的。核酸可包括DNA或RNA,并且可以是全部或部分合成的。

[0335] 用于在多种不同的宿主细胞中克隆并表达多肽的系统可用于所公开的方法中。合适的宿主细胞包括但不限于细菌、哺乳动物细胞、酵母和杆状病毒系统。本领域中用于表达异源多肽的可用的哺乳动物细胞系包括中国仓鼠卵巢细胞、HeLa细胞、幼仓鼠肾细胞、NS0小鼠黑素瘤细胞和许多其他细胞。常见的细菌宿主为大肠杆菌。

[0336] 在一些情况下,可使用抗体、TCR及其片段在原核细胞如大肠杆菌中的表达。关于综述,参见例如Plückthun, A. *Bio/Technology* 9:545-551 (1991)。在培养物中的真核细胞中的表达对于本领域技术人员而言也可用作产生本文所述的抗体和抗原结合片段的选项,关于近期综述,参见例如Raff, M.E. (1993) *Curr. Opin. Biotech.* 4:573-576; Trill J.J. 等人 (1995) *Curr. Opin. Biotech.* 6:553-560, 其中每一篇文献均通过引用而整体并入本文。

[0337] 可选择或构建合适的载体,该载体含有适当的调节序列,包括启动子序列、终止子序列、聚腺苷酸化序列、增强子序列、标志物基因和其他适当的序列。载体可以是质粒、病毒,例如,视情况为噬菌体或噬菌粒。进一步的细节参见例如, *Molecular Cloning: a Laboratory Manual*: 第2版, Sambrook等人, 1989, Cold Spring Harbor Laboratory Press。用于操控核酸的许多已知技术和方案,例如关于核酸构建体的制备、诱变、测序、DNA向细胞中的引入和基因表达以及蛋白质分析,详细描述于 *Short Protocols in Molecular Biology*, 第二版, Ausubel等人编著, John Wiley & Sons, 1992中。Sambrook等人和Ausubel等人的公开内容通过引用而整体并入本文。

[0338] 因此,另一个方面提供了含有来自选定淋巴细胞的多核苷酸的宿主细胞。又一个方面提供了一种方法,其包括将这样的核酸引入宿主细胞中。该引入可采用任何可用的技术。对于真核细胞,合适的技术可包括例如,磷酸钙转染、DEAE葡聚糖、电穿孔、脂质体介导的转染,以及使用逆转录病毒或其他病毒例如牛痘病毒或对于昆虫细胞使用杆状病毒的转导。对于细菌细胞,合适的技术可包括例如,氯化钙转化、电穿孔和使用噬菌体的转染。

[0339] 引入之后可以例如通过在用于基因表达的条件下培养宿主细胞引起或允许从核酸表达。

[0340] 在一种情况下,核酸整合到宿主细胞的基因组(例如染色体)中。可按照标准技术通过包含促进与基因组重组的序列来促进整合。

[0341] 本申请还提供了一种方法,该方法包括在表达系统中使用如上所述的构建体,以

便表达抗体、TCR或其抗原结合片段。

[0342] 本申请还涉及编码与使用本文所述方法鉴别的(抗原)结合的来自选定淋巴细胞的抗体、TCR或其抗原结合序列的分离的核酸,如重组DNA分子或克隆基因或其简并变体,其突变体、类似物或片段。

[0343] 另一个特征是本文公开的DNA序列的表达。如本领域公知的,DNA序列可通过将该DNA序列可操作地连接至合适的表达载体中的表达控制序列并采用该表达载体转化合适的单细胞宿主来表达。

[0344] 当然,DNA序列与表达控制序列的这种可操作连接包括(如果并非已作为DNA序列的一部分)在DNA序列上游的正确阅读框中提供起始密码子ATG。

[0345] 多核苷酸和载体可以以分离的和/或纯化的形式(例如,不含或基本不含除编码具有所需功能的多肽的多核苷酸以外来源的多核苷酸)提供。如本文所用的,“基本纯的”和“基本不含”是指含有少于,例如,20%或更少的外来物质、10%或更少的外来物质、5%或更少的外来物质、4%或更少的外来物质、3%或更少的外来物质、2%或更少的外来物质或者1%或更少的外来物质的溶液或悬浮液。

[0346] 可在表达本发明的DNA序列时采用很多种宿主/表达载体组合。例如,有用的表达载体可由染色体序列、非染色体序列和合成DNA序列的区段组成。合适的载体包括SV40的衍生物和已知的细菌质粒,例如,大肠杆菌质粒col E1、Pcr1、Pbr322、Pmb9及其衍生物,质粒如RP4;噬菌体DNA,例如,噬菌体λ的许多衍生物,例如,NM989和其他噬菌体DNA,例如,M13和丝状单链噬菌体DNA;酵母质粒如2μ质粒或其衍生物;用于真核细胞的载体,如用于昆虫或哺乳动物细胞的载体;衍生自质粒和噬菌体DNA的组的载体,如已被修饰以采用噬菌体DNA或其他表达控制序列的质粒等。

[0347] 多种表达控制序列(控制与其可操作地连接的DNA序列的表达的序列)中的任何序列均可用于这些载体中以表达DNA序列。这样的有用的表达控制序列包括例如,SV40、CMV、牛痘病毒、多瘤病毒或腺病毒、lac系统、trp系统、TAC系统、TRC系统、LTR系统的早期或晚期启动子,噬菌体λ的主要操纵子和启动子区域,fd外壳蛋白的控制区域,3-磷酸甘油酸激酶或其他糖酵解酶的启动子,酸性磷酸酶(例如Pho5)的启动子,酵母交配因子的启动子,以及已知控制原核或真核细胞或其病毒的基因表达的其他序列,及其各种组合。

[0348] 很多种单细胞宿主细胞也用于表达DNA序列。这些宿主包括公知的真核和原核宿主,如大肠杆菌、假单胞菌(Pseudomonas)、芽孢杆菌(Bacillus)、链霉菌(Streptomyces),真菌如酵母,和动物细胞如CHO、YB/20、NS0、SP2/0、R1.1、B-W和L-M细胞、非洲绿猴肾细胞(例如,COS 1、COS 7、BSC1、BSC40和BMT10)、昆虫细胞(例如,Sf9)以及组织培养中的人类细胞和植物细胞。

[0349] 应当理解,并非所有的载体、表达控制序列和宿主都将会同样充分地发挥作用以表达DNA序列。所有的宿主也将不会在相同的表达系统中同样充分地发挥作用。然而,本领域技术人员将能够在不偏离本申请范围的情况下选择合适的载体、表达控制序列和宿主,而无需过度的实验来实现期望的表达。例如,在选择载体时,必须考虑宿主,因为载体必须在该宿主中发挥作用。还将考虑载体的拷贝数、控制该拷贝数的能力以及由该载体编码的任何其他蛋白质如抗生素标志物的表达。本领域普通技术人员可在不偏离本申请范围的情况下选择合适的载体、表达控制序列和宿主,以实现期望的表达。例如,在选择载体时,对宿

主加以考虑,因为载体在该宿主中发挥作用。还可考虑载体的拷贝数、控制该拷贝数的能力以及由该载体编码的任何其他蛋白质如抗生素标志物的表达。

[0350] 本申请还提供了包含来自选定淋巴细胞的至少一种多核苷酸的如本文其他各处所述的质粒、载体、转录盒或表达盒形式的构建体。可选择或构建合适的载体,该载体含有适当的调节序列,包括启动子序列、终止子序列、聚腺苷酸化序列、增强子序列、可选择标志物和其他适当的序列。载体可以是质粒、病毒,例如,视情况为噬菌体、噬菌粒等。进一步的细节参见例如,Molecular Cloning:a Laboratory Manual:第2版,Sambrook等人,1989,Cold Spring Harbor Laboratory Press。用于操控核酸的许多已知技术和方案,例如关于核酸构建体的制备、诱变、测序、DNA向细胞中的引入和基因表达以及蛋白质分析,详细描述于Short Protocols in Molecular Biology,第二版,Ausubel等人编著,John Wiley& Sons,1992中。Sambrook等人和Ausubel等人的公开内容通过引用并入本文。

[0351] 在选择表达控制序列时,通常将会考虑多种因素。这些因素包括例如,系统的相对强度、其可控性及其与待表达的特定DNA序列或基因的相容性(特别是关于潜在的二级结构)。合适的单细胞宿主将通过考虑例如其与所选载体的相容性、其分泌特性、其正确折叠蛋白质的能力及其发酵需求,以及由待表达的DNA序列编码的产物对宿主的毒性,以及表达产物纯化的便利性来进行选择。

[0352] 除克隆之外或者除了克隆,还可重组/合成制备编码来自选定淋巴细胞的抗体、TCR或其抗原结合片段的多核苷酸。多核苷酸可设计为具有用于抗体、TCR或其抗原结合片段的适当密码子。一般而言,如果序列将用于表达,则将选择用于预期宿主的优选密码子。完整的多核苷酸可从通过标准方法制备的重叠寡核苷酸进行装配并装配成完整的编码序列。参见例如,Edge,Nature,292:756(1981);Nambair等人,Science,223:1299(1984);Jay等人,J.Biol.Chem.,259:6311(1984)。

[0353] 将非天然氨基酸位点特异性并入蛋白质的一般方法描述于Christopher J.Noren,Spencer J.Anthony-Cahill,Michael C.Griffith,Peter G.Schultz,Science,244:182-188(1989年4月)。该方法可用于采用非天然氨基酸产生类似物。

[0354] 如上所述,除克隆以外还可合成制备编码抗体、TCR或其抗原结合片段的DNA序列。

[0355] 变体

[0356] 在一些情况下,涉及由本文提供的选定多核苷酸编码的多肽的氨基酸序列变体。变体通常在一个或多个置换、缺失、添加和/或插入方面不同于本文具体公开的多肽。这样的变体可以是天然存在的,或者可以是例如通过修饰本发明上述多肽序列中的一种或多种并评估本文所述多肽的一种或多种生物学活性和/或使用本领域公知的多种技术中的任何技术而合成产生的。例如,可能需要改善由选定多核苷酸编码的多肽的结合亲和力和/或其他生物学特性。可通过将适当的修饰引入编码多肽的选定核苷酸序列中或通过肽合成来制备由选定多核苷酸编码的多肽的氨基酸序列变体。这样的修饰包括,例如,抗体氨基酸序列内的残基的缺失和/或插入和/或置换。可进行缺失、插入和置换的任意组合以得到最终构建体,条件是最终构建体具有期望的特征,例如抗原结合。

[0357] 在一些情况下,提供了由具有一个或多个氨基酸置换的选定多核苷酸的变体编码的多肽。通过抗体多肽置换进行的诱变的感兴趣位点包括CDR和FR。可将氨基酸置换引入由感兴趣的选定多核苷酸编码的多肽中,并筛选产物以获得期望的活性,例如,保留/改善的

抗原结合、降低的免疫原性或改善的ADCC或CDC。

[0358]

| 原始残基    | 示例性的保守置换                      |
|---------|-------------------------------|
| Ala (A) | Val; Leu; Ile                 |
| Arg (R) | Lys; Gln; Asn                 |
| Asn (N) | Gln; His; Asp; Lys; Arg       |
| Asp (D) | Glu; Asn                      |
| Cys (C) | Ser; Ala                      |
| Gln (Q) | Asn; Glu                      |
| Glu (E) | Asp; Gln                      |
| Gly (G) | Ala                           |
| His (H) | Asn; Gln; Lys; Arg            |
| Ile (I) | Leu; Val; Met; Ala; Phe; 正亮氨酸 |
| Leu (L) | 正亮氨酸; Ile; Val; Met; Ala; Phe |
| Lys (K) | Arg; Gln; Asn                 |
| Met (M) | Leu; Phe; Ile                 |
| Phe (F) | Trp; Leu; Val; Ile; Ala; Tyr  |
| Pro (P) | Ala                           |

[0359]

| 原始残基    | 示例性的保守置换                      |
|---------|-------------------------------|
| Ser (S) | Thr                           |
| Thr (T) | Val; Ser                      |
| Trp (W) | Tyr; Phe                      |
| Tyr (Y) | Trp; Phe; Thr; Ser            |
| Val (V) | Ile; Leu; Met; Phe; Ala; 正亮氨酸 |

[0360] 疏水氨基酸包括：正亮氨酸、Met、Ala、Val、Leu和Ile。中性亲水氨基酸包括：Cys、Ser、Thr、Asn和Gln。酸性氨基酸包括：Asp和Glu。碱性氨基酸包括：His、Lys和Arg。具有影响链取向的残基的氨基酸包括：Gly和Pro。芳香族氨基酸包括：Trp、Tyr和Phe。

[0361] 在一些情况下，置换、插入或缺失可发生在一个或多个CDR内，其中该置换、插入或缺失基本不降低与抗原的结合。例如，可在CDR中进行基本不降低结合亲和力的保守置换。这样的改变可在CDR“热点”或SDR之外。在变体V<sub>H</sub>和V<sub>L</sub>序列的一些情况下，每个CDR或者是未改变的，或者含有不超过一个、两个或三个氨基酸置换。

[0362] 可在CDR中进行改变（例如，置换），例如以改善亲和力。可在体细胞成熟期间在具有高突变率的编码CDR的密码子中进行这样的改变（参见例如，Chowdhury, Methods

Mol. Biol. 207:179-196 (2008)), 并且可对所得变体的结合亲和力进行测试。亲和力成熟(例如, 使用易错PCR、链改组、CDR的随机化或寡核苷酸定向诱变)可用于改善亲和力(参见例如, Hoogenboom等人in *Methods in Molecular Biology* 178:1-37 (2001))。可以例如使用丙氨酸扫描诱变或建模, 来特异性鉴别参与靶标或抗原结合的CDR残基(参见例如, Cunningham and Wells *Science*, 244:1081-1085 (1989))。特别是CDR-H3和CDR-L3经常被靶向。备选地或附加地, 使用抗原-抗体或抗原-TCR复合体的晶体结构来鉴别抗体或TCR与抗原之间的接触点。这样的接触残基和相邻残基可作为用于置换的候选物而被靶向或消除。可筛选变体以确定其是否含有期望的特性。

[0363] 氨基酸序列插入和缺失包括长度从一个残基到含有一百个或更多个残基的多肽的氨基末端和/或羧基末端融合, 以及单个或多个氨基酸残基的序列内插入和缺失。末端插入的实例包括N末端甲硫氨酰残基。分子的其他插入变体包括抗体或TCR的N末端或C末端与增加抗体或TCR的血清半衰期的酶(例如, 用于ADEPT)或多肽的融合。

[0364] 在一些情况下, 改变由来自选定淋巴细胞的多核苷酸编码的多肽, 以增加或降低其糖基化(例如, 通过改变氨基酸序列使得产生或去除一个或多个糖基化位点)。例如, 可改变与Fc区附接的碳水化合物。来自哺乳动物细胞的天然抗体通常包含通过N连接与Fc区的CH2结构域的Asn297附接的分支双触角寡糖(参见例如, Wright等人TIBTECH 15:26-32 (1997))。寡糖可以是与双触角寡糖结构的茎中的GlcNAc附接的各种碳水化合物, 例如, 甘露糖、N-乙酰葡萄糖胺(GlcNAc)、半乳糖、唾液酸、岩藻糖。可对抗体中的寡糖进行修饰, 例如以产生具有某些改善特性的抗体变体。抗体和TCR糖基化变体可具有改善的ADCC和/或CDC功能。

[0365] 因此, 由选定淋巴细胞的多核苷酸编码的多肽可由具有一种或多种外源和/或高内源糖基转移酶活性的宿主细胞产生。具有糖基转移酶活性的基因包括 $\beta$ (1,4)-N-乙酰基葡萄糖胺基转移酶III(GnTIII)、 $\alpha$ -甘露糖苷酶II(ManII)、 $\beta$ (1,4)-半乳糖基转移酶(GalT)、 $\beta$ (1,2)-N-乙酰基葡萄糖胺基转移酶I(GnTI)和 $\beta$ (1,2)-N-乙酰基葡萄糖胺基转移酶II(GnTII)。糖基转移酶可包含含有高尔基体定位结构域的融合物(参见例如, Lifely等人, *Glycobiology* 318:813-22 (1995); Schachter, *Biochem. Cell Biol.* 64:163-81 (1986); 美国临时专利申请号60/495,142和60/441,307; 专利公开号US 2003/0175884和US 2004/0241817; 以及W004/065540)。在一些情况下, 可在包含破坏的或失活的糖基转移酶基因的宿主细胞中表达由选定淋巴细胞的多核苷酸编码的多肽。因此, 在一些情况下, 本发明涉及宿主细胞, 其包含(a)分离的核酸, 该核酸含有编码具有糖基转移酶活性的多肽的序列; 和(b)来自选定淋巴细胞的分离的多核苷酸, 该多核苷酸编码结合人类靶标如人类疾病特异性靶标的抗体或TCR多肽。在特定的情况下, 由宿主细胞产生的选定淋巴细胞的多核苷酸编码的多肽的修饰多肽具有包含Fc区的IgG恒定区或其片段。在另一种特定的情况下, 由选定淋巴细胞的多核苷酸编码的多肽是包含Fc区的人源化抗体或其片段。分离的核酸包括通常含有核酸分子的细胞中包含的核酸分子, 但该核酸分子存在于染色体外或位于与其天然染色体位置不同的染色体位置。

[0366] 由宿主细胞产生、具有改变的糖基化、由来自选定淋巴细胞的多核苷酸编码的多肽可显示出增加的Fc受体结合亲和力(例如, 增加的与Fc $\gamma$ 激活受体如Fc $\gamma$ RIIIa受体的结合)和/或增加的效应子功能。增加的效应子功能可以是以下的一项或多项的增加: 抗体依

赖性细胞毒性增加、抗体依赖性细胞吞噬作用 (ADCP) 增加、细胞因子分泌增加、通过抗原呈递细胞的免疫复合体介导的抗原摄取增加、Fc介导的细胞毒性增加、与NK细胞的结合增加、与巨噬细胞的结合增加、与多形核细胞 (PMN) 的结合增加、与单核细胞的结合增加、与靶标结合的抗体或TCR的交联增加、诱导凋亡的直接信号传导增加、树突细胞成熟增加、以及T细胞引发增加。因此,在一个方面,本发明提供了与未经糖工程化的多肽相比,具有增加的效应子功能、由来自选定淋巴细胞的多核苷酸编码的多肽的糖型。(参见例如,Tang等人, J.Immunol.179:2815-2823 (2007))。

[0367] 本发明还涉及用于产生由来自选定淋巴细胞的多核苷酸编码的具有经修饰的寡糖的多肽的方法,该方法包括 (a) 在允许产生由来自选定淋巴细胞的多核苷酸编码的多肽的条件下培养宿主细胞,该宿主细胞被工程化以表达至少一种编码具有糖基转移酶活性的多肽的核酸,其中具有糖基转移酶活性的多肽以足以修饰由来自所述宿主细胞产生的选定淋巴细胞的多核苷酸编码的多肽的Fc区中寡糖的量进行表达;以及 (b) 分离由来自选定淋巴细胞的多核苷酸编码的多肽。在另一种情况下,存在两种具有糖基转移酶活性的多肽。本发明方法产生的由来自选定淋巴细胞的多核苷酸编码的多肽可具有增加的Fc受体结合亲和力和/或增加的效应子功能。

[0368] 在一些情况下,由来自选定淋巴细胞的多核苷酸编码的多肽的Fc区中平分的N-连接寡糖的百分比为总寡糖的至少约10%至约100%,具体地至少约50%,更具体地至少约60%、至少约70%、至少约80%或至少约90-95%。

[0369] 在另一种情况下,提供了包含由来自选定淋巴细胞的多核苷酸编码的多肽的组合物,该多肽被工程化以具有通过本文所述方法产生的增加的效应子功能和/或增加的Fc受体结合亲和力。在一些情况下,抗体或TCR是完整的抗体或TCR。在一些情况下,抗体或TCR是含有Fc区的抗体或TCR片段,或包含与免疫球蛋白或TCR链的Fc区等同的区域的融合蛋白。

[0370] 在一个方面,本发明提供了用于生成具有修饰的糖基化模式的本发明抗体和TCR的宿主细胞表达系统。特别地,本发明提供了用于生成具有改善的治疗价值的本发明抗体和TCR的糖型的宿主细胞系统。因此,本发明提供了被选择或工程化以表达具有糖基转移酶活性的多肽的宿主细胞表达系统。

[0371] 通常,包括上文所述细胞系在内的任何类型的培养细胞系均可用作将本发明的宿主细胞系工程化的背景。在一些情况下,CHO细胞、BHK细胞、NS0细胞、SP2/0细胞、Y0骨髓瘤细胞、P3X63小鼠骨髓瘤细胞、PER细胞、PER.C6细胞或杂交瘤细胞、其他哺乳动物细胞、酵母细胞、昆虫细胞或植物细胞用作生成本发明的工程化宿主细胞的背景细胞系。

[0372] 在一些情况下,可进一步修饰本文提供的抗体或TCR,以包含本领域已知且容易获得的附加非蛋白质部分。适用于抗体衍生的部分包括但不限于水溶性聚合物。水溶性聚合物的非限制性实例包括但不限于聚乙二醇 (PEG)、乙二醇/丙二醇的共聚物、羧甲基纤维素、葡聚糖、聚乙烯醇、聚乙烯吡咯烷酮、聚-1,3-二氧戊环、聚-1,3,6-三氧杂环己烷、乙烯/马来酸酐共聚物、聚氨基酸 (均聚物或无规共聚物) 和葡聚糖或聚 (n-乙烯基吡咯烷酮) 聚乙二醇、聚丙二醇均聚物、聚环氧丙烷/环氧乙烷共聚物、聚氧乙烯多元醇 (例如,甘油)、聚乙烯醇及其混合物。聚乙二醇丙醛由于其在水中的稳定性而可在制造中具有优势。

[0373] 聚合物可以具有任何分子量,并且可以是分支的或无分支的。与抗体或TCR附接的聚合物的数目可以变化,并且如果附接了两个或更多个聚合物,则这些聚合物可以是相同

或不同的分子。

[0374] 在另一种情况下,提供了可通过暴露于辐射而被选择性加热的抗体或TCR与非蛋白质部分的偶联物。在一些情况下,非蛋白质部分是碳纳米管(参见例如,Kam等人,Proc.Natl.Acad.Sci.USA 102:11600-11605(2005))。辐射可以具有任何波长,并且包括但不限于不伤害普通细胞但将非蛋白质部分加热至接近抗体-非蛋白质部分或TCR-非蛋白质部分的细胞被杀死的温度的波长。

[0375] 突变频率

[0376] 抗体或TCR可包含来自相应种系序列、具有至少约5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%或20%或更高的突变频率的重链、轻链、TCR $\alpha$ 或TCR $\beta$ 序列。例如,由来自选定淋巴细胞的多核苷酸编码的抗体可包含CDR3区,该CDR3区是来自种系序列、具有至少约5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%或20%或更高的突变频率的轻链序列。例如,本发明的抗体可包含来自种系序列、具有至少约5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%或20%或更高的突变频率的重链和轻链序列。

[0377] 在一些情况下,由来自选定淋巴细胞的多核苷酸编码的抗体或TCR是人类抗体或TCR。可使用本领域已知的各种技术产生人类抗体(参见例如, van Dijk和van de Winkel, Curr.Opin.Pharmacol.5:368-74(2001);以及Lonberg, Curr.Opin.Immunol.20:450-459(2008))。人类抗体或TCR是具有与由人类或人类细胞产生的或衍生自利用人类抗体组库或其他人类抗体编码序列的非人类来源的抗体或TCR的氨基酸序列相对应的氨基酸序列的人类抗体或TCR。可来自选定淋巴细胞的多核苷酸,例如包含来自选定淋巴细胞的多核苷酸的序列的载体制备人类抗体。

[0378] 还可通过基于杂交瘤的方法制备人类抗体。例如,可使用人类B细胞杂交瘤技术和其他方法,从人骨髓瘤和小鼠-人异源骨髓瘤细胞系产生人类抗体(参见例如,Kozbor J.Immunol.,133:3001(1984);Brodeur等人, Monoclonal Antibody Production Techniques and Applications, pp.51-63(1987);Boerner等人, J.Immunol.,147:86(1991);Li等人, Proc.Natl.Acad.,103:3557-3562(2006);美国专利号7,189,826;Ni, Xiandai Mianyixue,26(4):265-268(2006);Vollmers和Brandlein, Histology and Histopathology,20(3):927-937(2005);以及Vollmers和Brandlein, Methods and Findings in Experimental and Clinical Pharmacology,27(3):185-91(2005))。还可通过分离选自人源噬菌体展示文库的Fv克隆可变域序列来生成人类抗体和TCR。然后可将这样的可变域序列与期望的人类恒定域组合。

[0379] 由选定淋巴细胞的多核苷酸编码的多肽的靶标的鉴别及其表征

[0380] 本文所公开的方法包括验证或表征由来自选定淋巴细胞的多核苷酸编码的重组或合成产生的多肽。可通过本领域已知的各种测定法来测定、筛选或表征由来自选定淋巴细胞的多核苷酸编码的多肽的物理/化学特性和/或生物学活性。公开了用于确定由选定淋巴细胞的多核苷酸编码的多肽的靶标的身份的方法。所公开的方法可包括通过各种测定来筛选或表征由选定淋巴细胞的多核苷酸编码的多肽的物理/化学特性和/或生物学活性。靶标可以是蛋白质或抗原,如组织特异性蛋白或抗原。在一些情况下,该蛋白质或抗原可以是疾病特异性蛋白或抗原,如癌症特异性蛋白或抗原。

[0381] 在一个方面,例如通过ELISA、蛋白质印迹等来测试由来自选定淋巴细胞的多核苷酸编码的多肽,例如抗体或TCR的抗原结合活性。例如,可以例如使用免疫沉淀和质谱或包含待测试的多种蛋白质或抗原的蛋白质阵列来测试由来自选定淋巴细胞的多核苷酸编码的多肽与多种抗原或蛋白质的结合活性。蛋白质阵列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-VHH;MacBeath和Schreiber,2000,Science 289:1760-1763;WO 01/40803和WO 99/51773A1。阵列的使用允许靶标的鉴别由机器人和/或以高通量方式进行。

[0382] 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。

[0383] 在一个方面,竞争性测定可用于鉴别与由选定多核苷酸编码的多肽竞争与靶标结合的分子,如多肽、抗体或小分子。在一些情况下,这样的竞争性分子与由选定多核苷酸编码的多肽所结合的相同表位(例如,线性或构象表位)结合。示例性表位作图方法是已知的(参见例如,Morris“Epitope Mapping Protocols,”in Methods in Molecular Biology vol.66 (1996))。在示例性竞争性测定中,在包含第一标记多肽和第二未标记多肽的溶液中温育固定的靶标,该第一标记多肽由选定多核苷酸编码并与靶标结合,该第二未标记多肽正在被测试其与由选定多核苷酸编码的多肽竞争与靶标结合的能力。第二抗体可存在于杂交瘤上清液中。作为对照,在包含由选定多核苷酸编码的第一标记多肽但不包含第二未标记分子的溶液中温育固定的靶标。在允许由选定多核苷酸编码的多肽与靶标结合的条件下温育后,去除过量的未结合多肽,并测量与固定的靶标相关的标记的量。如果测试样品中与固定的靶标相关的标记的量相对于对照样品大幅降低,则表明第二分子与由选定淋巴细胞的选定多核苷酸编码的第一多肽竞争与靶标的结合(参见例如,Harlow and Lane Antibodies:A Laboratory Manual Ch.14(1996))。

[0384] 在一些情况下,由选定淋巴细胞的多核苷酸编码的抗体或TCR多肽具有约1 $\mu$ M、100nM、10nM、5nM、2nM、1nM、0.5nM、0.1nM、0.05nM、0.01nM或0.001nM或更小(例如,10<sup>-8</sup>M或更低,例如,10<sup>-8</sup>M至10<sup>-13</sup>M,例如,10<sup>-9</sup>M至10<sup>-13</sup>M)的解离常数(K<sub>D</sub>)。本发明的另一个方面提供了由来自选定淋巴细胞的多核苷酸编码的多肽,该多肽对其靶标具有增加的亲和力,例如,亲和力成熟的抗体或TCR。亲和力成熟的抗体或TCR是与不具有改变的亲本抗体或TCR相比在一个或多个高变区(HVR)中具有一个或多个改变的抗体或TCR,这样的改变导致抗体或TCR对抗原或靶标的亲和力改善。这些抗体和TCR可与靶标结合,并具有约5 $\times$ 10<sup>-9</sup>M、2 $\times$ 10<sup>-9</sup>M、1

$\times 10^{-9}$  M、 $5 \times 10^{-10}$  M、 $2 \times 10^{-9}$  M、 $1 \times 10^{-10}$  M、 $5 \times 10^{-11}$  M、 $1 \times 10^{-11}$  M、 $5 \times 10^{-12}$  M、 $1 \times 10^{-12}$  M或更小的 $K_D$ 。在一些情况下,本发明提供了由来自选定淋巴细胞的多核苷酸编码的抗体或TCR,该抗体或TCR具有相比于种系抗体或TCR至少1.5倍、2倍、2.5倍、3倍、4倍、5倍、10倍、20倍或更高倍的增加的亲和力。在一些情况下,由来自选定淋巴细胞的多核苷酸编码的多肽显示出效应子功能活性,例如,Fc介导的细胞毒性,包括ADCC活性。

[0385] 可通过任何合适的测定来测量 $K_D$ 。例如,可通过放射性标记的抗原结合测定(RIA)来测量 $K_D$ (参见例如,Chen等人,J.Mol.Biol.293:865-881(1999);Presta等人,Cancer Res.57:4593-4599(1997))。例如,可使用表面等离子体共振测定(例如,使用BIACORE®-2000或BIACORE®-3000)来测量 $K_D$ 。

[0386] 在一些情况下,本文提供的抗体或TCR是多特异性抗体或TCR,例如,双特异性抗体或TCR。多特异性抗体或TCR可以是对至少两个不同位点具有结合特异性的抗体或TCR(参见例如,美国专利公开号US 2008/0069820)。在一些情况下,其中一种结合特异性是针对第一靶标,而另一种是针对任何其他靶标。在一些情况下,双特异性抗体或TCR可与靶标的两个不同表位结合。双特异性抗体或TCR还可用于将细胞毒性剂定位至患病细胞或感染细胞。双特异性抗体或TCR可制备成全长抗体或TCR,或者抗体或TCR片段。

[0387] 用于制备多特异性抗体或TCR的示例性技术包括具有不同特异性的两个免疫球蛋白重链-轻链对或TCR $\alpha$ 链-TCR $\beta$ 链对的重组共表达,用于制备Fc-异二聚体分子的工程静电转向效应,两种或更多种抗体、TCR或其片段的交联,使用亮氨酸拉链产生双特异性抗体或TCR,使用“双抗体”技术用于制备双特异性抗体或TCR片段,使用单链Fv(scFv)二聚体制备三特异性抗体或TCR,以及“杵臼(knob-in-hole)”工程化(参见例如,Milstein和Cuellar,Nature 305:537(1983);W009/089004A1;W093/08829;Traunecker等人,EMBO J.10:3655(1991);美国专利号4,676,980和5,731,168;Brennan等人,Science,229:81(1985);Kostelny等人,J.Immunol.,148(5):1547-1553(1992);Hollinger等人,Proc.Natl.Acad.Sci.USA,90:6444-6448(1993);Gruber等人,J.Immunol.,152:5368(1994);以及Tutt等人J.Immunol.147:60(1991))。具有三个或更多个功能性抗原结合位点的工程化抗体或TCR也包括在内(参见例如,US 2006/0025576)。

[0388] 在一个方面,提供了用于鉴别具有生物学活性、由来自选定淋巴细胞的多核苷酸编码的一种或多种多肽的测定。在一些情况下,提供了用于鉴别对靶标具有中和活性、由来自选定淋巴细胞的多核苷酸编码的多肽的测定。还提供了在体内和/或体外具有这样的生物学活性、由选定多核苷酸编码的多肽。在一些情况下,测试由本发明的选定多核苷酸编码的多肽的这样的生物学活性。

[0389] 在一个方面,提供了用于鉴别由来自选定淋巴细胞的多核苷酸编码的一种或多种多肽的测定,该多肽对与取样的患病组织的病理学相关或对该患病组织的病理学特异的抗原或靶标具有反应性、高亲和力和/或高特异性。

[0390] 在一个方面,提供了使用噬菌体、核糖体或RNA展示技术鉴别由来自选定淋巴细胞的多核苷酸编码的一种或多种多肽的测定。例如,这些技术可用于选择具有相关反应性、由来自选定淋巴细胞的多核苷酸编码的多肽。选择之前和之后的反应性比较可鉴别那些具有反应性并因此可能是病理性的多肽。在另一种情况下,可以以阵列形式使用特定展示技术(例如噬菌体、核糖体或RNA展示)。例如,携带来自由选定淋巴细胞的多核苷酸编码的一种

或多种多肽的单个序列的单个分子(或这些单个分子的扩增)可排列为噬菌体、核糖体或RNA。然后可研究特异性靶标或抗原,以鉴别编码与其结合的肽(例如,Ig或TCR多肽)的序列。然后可基于由来自选定淋巴细胞的多核苷酸编码的一种或多种多肽的靶标的身份,来选择抑制与疾病相关的抗原的药物。

[0391] 在一些方面,提供了使用患病和或未患病样品如组织样品或FFPE样品或NAT样品,通过免疫测定技术来鉴别由来自选定淋巴细胞的多核苷酸编码的一种或多种多肽的测定。可沿着基本平行的平面将组织样品切割成多个连续的组织学切片,以用于通过许多已知的组织学、组织化学、免疫组织学、组织病理学、显微镜(包括形态测定分析和/或三维重建)、细胞学、生物化学、药理学、分子生物学、免疫化学、成像或其他分析技术中的任何技术进行分析。参见例如,Bancroft和Gamble,Theory and Practice of Histological Techniques(第6版),2007 Churchill Livingstone,Oxford,UK;Kiemann,Histological and Histochemical Methods:Theory and Practice,2001 Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.;M.A.Hayat(编著),Cancer Imaging一卷1和卷2,2007 Academic Press,NY。

[0392] 使用方法

[0393] 参与免疫系统介导的疾病特异性细胞例如癌细胞破坏的疾病特异性抗原例如癌抗原的分子鉴别用于鉴别已知药物和/或开发特异性活性药物,例如肽、核酸、抗体和小分子,这些特异性活性药物抑制使用本文所述方法鉴别的靶标。参与免疫系统介导的疾病特异性细胞例如癌细胞破坏的疾病特异性抗原例如癌抗原的分子鉴别用于开发针对疾病的特异性主动免疫策略(例如,癌症疫苗)以及用于体外产生用于过继性免疫疗法的淋巴细胞。使用在体外对人类癌症抗原具有反应性的淋巴细胞,筛选cDNA或基因组文库以鉴别编码这些抗原的基因是可能的。

[0394] 由选定淋巴细胞的多核苷酸编码的Ig或TCR多肽,如本文所述的抗体和抗原结合片段的组合物可用作非治疗剂(例如,作为亲和纯化剂)。通常,在一个这样的实例中,使用本领域已知的常规方法将感兴趣的蛋白质固定在固相如Sephadex树脂或滤纸上。使固定的蛋白质与含有待纯化的感兴趣靶标(或其片段)的样品接触,然后用合适的溶剂洗涤支持体,该溶剂将会基本去除样品中除与固定的Ig或TCR多肽(例如抗体)结合的靶蛋白之外的所有材料。最后,用另一种合适的溶剂,如pH 5.0的甘氨酸缓冲液洗涤支持体,该支持体将释放靶蛋白。除了纯化之外,组合物还可用于检测、诊断并治疗与靶蛋白相关的疾病和病症。

[0395] 根据本申请的一个实例的患者是表现出疾病或病症,例如,与患病样品相同的疾病或病症的一种或多种临床表现和/或症状的哺乳动物(例如人)。在某些情况下,患者可以是无症状的,但仍具有该疾病或病症的临床表现。

[0396] 选定的抗体、TCR或其抗原结合片段可与治疗部分偶联或者可以是含有治疗部分的融合蛋白。选定的抗体、TCR或其抗原结合片段可与可检测部分偶联或者可以是含有可检测部分的融合蛋白。在一种情况下,选定的抗体、TCR或其抗原结合片段可与治疗部分和可检测部分两者偶联。选定的抗体、TCR或其抗原结合片段可与亲和标签(例如,纯化标签)偶联或被重组工程化为具有该亲和标签。

[0397] 本文提供的抗体、TCR或其抗原结合片段设置为可与治疗部分和/或成像或可检测

部分和/或亲和标签偶联或连接。用于偶联或连接多肽的方法是本领域公知的。化合物与标记之间的关联(结合)包括本领域已知的任何方式,包括但不限于共价和非共价相互作用、化学偶联以及重组技术。

#### [0398] 诊断

[0399] 抗蛋白质抗体、TCR及其片段可用于体内和体外检测、诊断和/或监测目的。靶蛋白(以及在一些情况下,过量或突变的蛋白质)可参与多种疾病和病症。靶蛋白相关疾病和病况的治疗部分取决于其诊断,并且本文所述的抗体、TCR及其抗原结合片段用于诊断过量或突变的靶蛋白或用于诊断与靶蛋白活性相关的疾病和病况。

[0400] 本文提供了检测样品或受试者的靶蛋白水平的方法,该方法包括(i)使抗体、TCR或其抗原结合片段与来自受试者的样品接触,以及(ii)检测选定抗体、TCR或其抗原结合片段与蛋白质的复合体。

[0401] 在一种情况下,选定的抗体、TCR或抗原结合片段进一步包含可检测部分。检测可在体外、体内或离体进行。用于采用选定的抗体、TCR或其抗原结合片段检测和/或确定(定量、定性等)靶蛋白的体外测定包括但不限于例如ELISA、RIA和蛋白质印迹。可通过从患者获得样品(例如,活检样品)并在例如标准ELISA测定中测试样品来进行靶蛋白的体外检测、诊断或监测。例如,可将96孔微量滴定板用本文所述的选定抗体、TCR或其抗原结合片段包被,洗涤并用PBS-吐温/BSA包被以抑制非特异性结合。可将样品连续稀释并置于与连续稀释的靶蛋白标准曲线相比重复的孔中。温育并洗涤孔之后,可添加用生物素标记的抗靶蛋白抗体或TCR,然后添加链霉亲和素-碱性磷酸酶。可洗涤孔并添加底物(辣根过氧化物酶)以使板显影。可使用常规的读板器和软件读板。

[0402] 当检测发生在体内时,使用任何常规方法如本文其他各处所述的方法通过施用抗体、TCR或其抗原结合片段进行接触。在这样的方法中,样品或受试者中的靶蛋白(以及在一些情况下过量水平的靶蛋白)的检测可用于诊断与靶蛋白活性相关或相关联的疾病或病症,如本文所述的那些疾病和病症。

[0403] 在靶蛋白的体内检测、诊断或监测中,向患者施用与靶蛋白结合的选定抗体、TCR或其抗原结合片段,该选定抗体、TCR或其抗原结合片段与可检测部分结合。可使用本领域公认的方法使可检测部分可视化,这些方法诸如但不限于磁共振成像(MRI),荧光,放射成像,由内窥镜、腹腔镜或血管内导管提供的光源(即,经由光敏剂的检测),光扫描,正电子发射断层成像(PET)扫描,全身核磁共振(NMR),放射扫描成像,单光子发射计算机断层成像(SPECT),靶向近红外区(NIR)扫描,X射线,超声波等,例如,描述于美国专利号6,096,289、美国专利号7,115,716、美国专利号7,112,412、美国专利申请号20030003048和美国专利申请号20060147379中的方法,其中每一篇专利均通过引用而整体并入本文。用于使用此类方法检测化合物的标记也是本领域已知的,并描述于这些专利和申请中,并通过引用并入本文。可检测部分的可视化可允许检测、诊断和/或监测与靶蛋白相关的病况或疾病。

[0404] 利用对期望靶蛋白,即靶蛋白具有特异性的抗体或TCR的另外的诊断测定是本领域已知的并且也在本文中涉及。

[0405] 在病况和疾病的检测、诊断或监测中,可向受试患者施用选定抗体、TCR或其抗原结合片段的组合物,该选定Ig或TCR多肽与可检测部分偶联。可使用本领域公认的方法如上述方法使该部分可视化。可检测部分的可视化可允许检测、诊断和/或监测病况和疾病。

[0406] 因此,提供了包含针对靶蛋白的Ig或TCR多肽(例如抗体、TCR及其抗原结合片段)的组合物,该组合物用于检测或诊断与疾病或病症相关的靶蛋白的过量水平或靶蛋白,从而潜在地指示对治疗处理的需求。在某些情况下,抗体或TCR包括本文所述的选定的和任选人源化的抗靶蛋白抗体或TCR。在其他情况下,由选定淋巴细胞的多核苷酸编码的抗体或TCR多肽进一步包括第二试剂。这样的试剂可以是分子或部分,例如,报道分子或可检测标记。用于此类检测方法的可检测标记/部分是本领域已知的并在下文更详细地描述。报道分子是可使用测定来检测的任何部分。已与多肽偶联的报道分子的非限制性实例包括酶、放射性标记、半抗原、荧光标记、磷光分子、化学发光分子、发色团、发光分子、光亲和分子、有色粒子或配体如生物素。可检测标记包括可由于其特定功能特性和/或化学特征而被检测到的化合物和/或元件,使用该化合物和/或元件允许检测与其附接的多肽,和/或如果需要进一步定量。本领域已知许多合适的可检测(成像)剂,以及它们与多肽附接的方法(参见例如,美国专利号5,021,236;4,938,948;和4,472,509,其中每一篇专利均通过引用并入本文)。

[0407] 连接多肽如抗体或TCR与可检测部分的方法是本领域已知的,并且包括例如,用来形成融合蛋白的重组DNA技术和偶联(例如,化学偶联)。通过化学偶联或重组工程制备融合蛋白的方法是本领域公知的。共价和非共价连接组分的方法也是本领域已知的。参见例如,Williams(1995) *Biochemistry* 34:1787-1797;Dobeli(1998) *Protein Expr.Purif.* 12:404-414;和Kroll(1993) *DNA Cell.Biol.* 12:441-453。

[0408] 在一些情况下,可能需要在标记或部分与抗体、TCR或其抗原结合片段的一个或多个部分之间引入非结构化多肽连接体区域。连接体可促进增强灵活性,和/或减少任何两个片段之间的空间位阻。连接体还可促进每个片段发生适当的折叠。连接体可以是天然来源的,如被确定存在于蛋白质的两个结构域之间的无规卷曲中的序列。一个连接体序列是在RNA聚合酶 $\alpha$ 亚基的C末端与N末端结构域之间发现的连接体。天然存在的连接体的其他实例包括在1CI和LexA蛋白质中发现的连接体。

[0409] 在连接体内,氨基酸序列可根据如凭经验确定的或通过建模揭示的连接体特征而变化。选择连接体的考虑因素包括连接体的灵活性、连接体的电荷和连接体的一些氨基酸在天然存在的亚基中的存在。还可设计连接体使得连接体中的残基接触脱氧核糖核酸(DNA),从而影响结合亲和力或特异性,与其他蛋白质相互作用。在一些情况下,诸如当有必要跨越亚基之间的较长距离时或者当结构域必须保持特定的构型时,连接体可任选地含有额外的折叠域。在一些情况下,连接体的设计可涉及要求连接体跨越相对较短距离,例如小于约10埃( $\text{\AA}$ )的结构域的排列。然而,在某些情况下,连接体跨越长达约50埃的距离。

[0410] 在连接体内,氨基酸序列可根据如凭经验确定的或通过建模揭示的连接体特征而变化。选择连接体的考虑因素包括连接体的灵活性、连接体的电荷和连接体的一些氨基酸在天然存在的亚基中的存在。还可设计连接体使得连接体中的残基接触DNA,从而影响结合亲和力或特异性,与其他蛋白质相互作用。在一些情况下,当有必要跨越亚基之间的较长距离时或者当结构域必须保持特定的构型时,连接体可任选地含有额外的折叠域。

[0411] 用于将多肽(游离的或与细胞结合的)与珠子偶联的方法是本领域已知的。用于选择偶联的多肽或展示多肽的细胞的方法也是本领域已知的。简言之,顺磁性聚苯乙烯微粒可商购获得(Spherotech, Inc., Libertyville, IL; Invitrogen, Carlsbad, CA), 其使肽与用

官能团修饰或用各种抗体或配体,例如,抗生物素蛋白、链霉亲和素或生物素包被的微粒表面偶联。

[0412] 微粒的顺磁特性允许使用磁铁将其从溶液中分离出来。当从磁铁移除时,微粒可容易地重悬。多肽可与管中包被有聚氨酯层的顺磁性聚苯乙烯微粒偶联。微粒表面上的羟基基团通过与对甲苯磺酰氯反应而活化(Nilsson K和Mosbach K.“p-Toluenesulfonyl chloride as an activating agent of agarose for the preparation of immobilized affinity ligands and proteins.”*Eur.J.Biochem.*1980;112:397-402)。或者,可用碳化二亚胺激活含有表面羧酸的顺磁性聚苯乙烯微粒,然后将该微粒与多肽偶联,从而在多肽的伯氨基与微粒表面上的羧酸基团之间产生稳定的酰胺键(Nakajima N和Ikade Y, Mechanism of amide formation by carbodiimide for bioconjugation in aqueous media,*Bioconjugate Chem.*1995,6(1),123-130;Gilles MA,Hudson AQ和Borders CL Jr, Stability of water-soluble carbodiimides in aqueous solution,*Anal Biochem.*1990年2月1日;184(2):244-248;Sehgal D和Vijay IK,a method for the high efficiency of water-soluble carbodiimide-mediated amidation,*Anal Biochem.*1994年4月;218(1):87-91;Szajani B等人,Effects of carbodiimide structure on the immobilization of enzymes,*Appl Biochem Biotechnol.*1991年8月;30(2):225-231)。另一种选择是将生物素化多肽与表面已与单层链霉亲和素共价连接的顺磁性聚苯乙烯微粒偶联。(Argarana CE,Kuntz ID,Birken S,Axel R,Cantor CR.Molecular cloning and nucleotide sequence of the streptavidin gene.*Nucleic Acids Res.*1986;14(4):1871-82;Pahler A,Hendrickson WA,Gawinowicz Kolks MA,Aragana CE,Cantor CR.Characterization and crystallization of core streptavidin.*J Biol Chem* 1987;262(29):13933-7)。

[0413] 多肽可与很多种荧光染料、猝灭剂和半抗原,如荧光素、R-藻红蛋白和生物素偶联。偶联可在多肽合成期间或多肽合成并纯化之后进行。生物素是小分子(244kDa)维生素,其以高亲和力与抗生物素蛋白以及链霉亲和素蛋白结合,并且可与大多数肽偶联而不改变这些肽的生物学活性。使用固定的链霉亲和素和抗生物素蛋白亲和凝胶容易从未标记的多肽纯化生物素标记的多肽,并且链霉亲和素或抗生物素蛋白偶联的探针可用于在例如ELISA、斑点印迹或蛋白质印迹应用中检测生物素化多肽。生物素的N-羟基琥珀酰亚胺酯是最常用的生物素化试剂类型。N-羟基琥珀酰亚胺活化的生物素在生理缓冲液中与伯氨基有效反应,形成稳定的酰胺键。多肽在N末端具有伯胺,并且还可在赖氨酸残基的侧链中具有数个伯胺,这些伯胺可用作采用N-羟基琥珀酰亚胺活化的生物素试剂进行标记的靶标。生物素的数个不同的N-羟基琥珀酰亚胺酯是可获得的,它们具有不同的性质和间隔臂长度(Pierce,Rockford,IL)。磺基-N-羟基琥珀酰亚胺酯试剂是水溶性的,从而使得反应能够在缺乏有机溶剂的情况下进行。

[0414] 可采用本领域公认的技术使用2-(4'-羟基偶氮苯-2-羧酸)测定来估计生物素与多肽的摩尔比(Green,NM,(1975)“Avidin.In *Advances in Protein Chemistry.*”Academic Press,New York.29,85-133;Green,NM,(1971)“The use of bifunctional biotinyl compounds to determine the arrangement of subunits in avidin.”*Biochem J.*125,781-791;Green,NM.,(1965)“A spectrophotometric assay for avidin

and biotin based on binding of dyes by avidin.”Biochem.J.94:23c-24c)。数个生物素分子可与多肽偶联,并且每个生物素分子可结合一个抗生物素蛋白分子。生物素-抗生物素蛋白键形成在有机溶剂、极端pH和变性试剂中是非常快速且稳定的。为了对生物素化进行定量,将含有生物素化多肽的溶液添加到2-(4'-羟基偶氮苯-2-羧酸)和抗生物素蛋白的混合物中。由于生物素对抗生物素蛋白具有较高的亲和力,因此其取代了2-(4'-羟基偶氮苯-2-羧酸),并且在500nm处的吸光度成比例降低。可通过在添加含生物素的肽之前和之后测量2-(4'-羟基偶氮苯-2-羧酸)-抗生物素蛋白溶液的吸光度,在单个吸收池中对溶液中生物素的量进行定量。吸光度的变化通过2-(4'-羟基偶氮苯-2-羧酸)-抗生物素蛋白复合体的消光系数与样品中生物素的量关联。

[0415] 或者,由选定淋巴细胞的多核苷酸编码的Ig或TCR多肽可与荧光部分偶联。可使用本领域公认的技术实现Ig或TCR多肽与荧光部分(例如,R-藻红蛋白、异硫氰酸荧光素(FITC)等)的偶联,该技术描述于例如,Glazer,AN和Stryer L.(1984).Trends Biochem.Sci.9:423-7;Kronick,MN和Grossman,PD(1983)Clin.Chem.29:1582-6;Lanier,LL和Loken,MR(1984)J.Immunol.,132:151-156;Parks,DR等人(1984)Cytometry 5:159-68;Hardy,RR等人(1983)Nature 306:270-2;Hardy RR等人(1984)J.Exp.Med.159:1169-88;Kronick,MN(1986)J.Immuno.Meth.92:1-13;Der-Balian G,Kameda,N和Rowley,G.(1988)Anal.Biochem.173:59-63中。

[0416] 在一个非限制性实例中,由选定淋巴细胞的多核苷酸编码的Ig或TCR多肽可与可检测标记,如放射性核素、铁相关化合物、染料、成像剂或用于靶蛋白免疫检测的荧光剂相关联(偶联),该可检测标记可用于在体外和/或体内使Ig或TCR多肽与靶蛋白的结合可视化。

[0417] 放射性标记的非限制性实例包括例如, $^{32}\text{P}$ 、 $^{33}\text{P}$ 、 $^{43}\text{K}$ 、 $^{52}\text{Fe}$ 、 $^{57}\text{Co}$ 、 $^{64}\text{Cu}$ 、 $^{67}\text{Ga}$ 、 $^{67}\text{Cu}$ 、 $^{68}\text{Ga}$ 、 $^{71}\text{Ge}$ 、 $^{75}\text{Br}$ 、 $^{76}\text{Br}$ 、 $^{77}\text{Br}$ 、 $^{77}\text{As}$ 、 $^{77}\text{Br}$ 、 $^{81}\text{Rb}$ 、 $^{81}\text{MKr}$ 、 $^{87}\text{MSr}$ 、 $^{90}\text{Y}$ 、 $^{97}\text{Ru}$ 、 $^{99}\text{Tc}$ 、 $^{100}\text{Pd}$ 、 $^{101}\text{Rh}$ 、 $^{103}\text{Pb}$ 、 $^{105}\text{Rh}$ 、 $^{109}\text{Pd}$ 、 $^{111}\text{Ag}$ 、 $^{111}\text{In}$ 、 $^{113}\text{In}$ 、 $^{119}\text{Sb}$ 、 $^{121}\text{Sn}$ 、 $^{123}\text{I}$ 、 $^{125}\text{I}$ 、 $^{127}\text{Cs}$ 、 $^{128}\text{Ba}$ 、 $^{129}\text{Cs}$ 、 $^{131}\text{I}$ 、 $^{131}\text{Cs}$ 、 $^{143}\text{Pr}$ 、 $^{153}\text{Sm}$ 、 $^{161}\text{Tb}$ 、 $^{166}\text{Ho}$ 、 $^{169}\text{Eu}$ 、 $^{177}\text{Lu}$ 、 $^{186}\text{Re}$ 、 $^{188}\text{Re}$ 、 $^{189}\text{Re}$ 、 $^{191}\text{Os}$ 、 $^{193}\text{Pt}$ 、 $^{194}\text{Ir}$ 、 $^{197}\text{Hg}$ 、 $^{199}\text{Au}$ 、 $^{203}\text{Pb}$ 、 $^{211}\text{At}$ 、 $^{212}\text{Pb}$ 、 $^{212}\text{Bi}$ 和 $^{213}\text{Bi}$ 。可使用常规化学方法使放射性标记与化合物衔接。放射性标记的化合物用于体外诊断技术和体内放射成像技术以及放射免疫疗法。例如,在体内成像的情况下,Ig或TCR多肽可与成像剂而非放射性同位素(包括但不限于磁共振成像增强剂)偶联,其中例如由选定淋巴细胞的多核苷酸编码的Ig或TCR多肽通过螯合基团加载有大量顺磁性离子。螯合基团的实例包括EDTA、卟啉、多胺冠醚和聚脲。顺磁性离子的实例包括钆、铁、锰、铈、镧、铈和铕。这样的可检测部分还包括:金属;金属螯合剂;镧系元素;镧系螯合剂;放射性金属;放射性金属螯合剂;正电子发射核;微泡(用于超声波);脂质体;在脂质体或纳米球中微囊化的分子;单晶氧化铁纳米化合物;磁共振成像造影剂;光吸收剂、光反射剂和/或光散射剂;胶体粒子;荧光团,如近红外荧光团。在许多情况下,这样的第二官能团/部分将是相对较大的,例如,至少25个原子质量单位(amu)的大小,并且在许多情况下可以是至少50、100或250amu的大小。在某些情况下,第二官能团是用于螯合金属的螯合部分,例如,放射性金属或顺磁性离子的螯合剂。在实例中,放射性核素的螯合剂用于放射疗法或成像程序。

[0418] 疗法

[0419] 本文提供了预防或治疗与鉴别的靶蛋白相关的一种或多种疾病或病症的方法,该

方法包括施用包含药物的组合物,该药物抑制或结合由选定淋巴细胞的多核苷酸编码的多肽的鉴别的靶标,例如选定的抗体、TCR或其抗原结合片段;肽;核酸;或小分子;该药物结合与疾病或病症相关的鉴别的靶蛋白。

[0420] 本文提供了预防或治疗与靶蛋白相关的一种或多种疾病或病症的方法,该方法包括施用包含药物的组合物,该药物抑制或结合由选定淋巴细胞的多核苷酸编码的多肽的鉴别的靶标,例如选定的抗体、TCR或其抗原结合片段;肽;核酸;或小分子;该药物结合与疾病或病症相关的蛋白质,减少所鉴别的靶蛋白与所鉴别的蛋白质的结合配偶体例如配体之间的复合体形成。

[0421] 可以以治疗有效量将包含药物的组合物施用于患者(例如,哺乳动物如人类或非人类动物,如灵长类动物、啮齿类动物、牛、马、猪、羊等),该药物抑制或结合由选定淋巴细胞的多核苷酸编码的多肽的鉴别的靶标,例如选定的抗体、TCR或其抗原结合片段;肽;核酸;或小分子;所述组合物通过抑制与所鉴别的靶蛋白相关的疾病或病症以适用于任何医学治疗的合理的获益/风险比有效地产生一些期望的治疗效果。为了向人类患者施用本发明组合物,可通过本领域普通技术人员已知的方法来配制组合物。治疗有效量是在器官或组织中实现至少部分期望的治疗或预防效果的量。在一个实例中,抑制或结合由选定淋巴细胞的多核苷酸编码的多肽的鉴别的靶标、实现疾病或病症的预防和/或治疗所必需的药物,例如选定的抗体、TCR或其抗原结合片段;肽;核酸;或小分子的量本身不是固定的。抑制或结合由选定淋巴细胞的多核苷酸编码的多肽的鉴别的靶标的药物,例如选定的抗体、TCR或其抗原结合片段;肽;核酸;或小分子的施用量将随着药物的类型、疾病的类型、疾病的广泛程度以及患有疾病或病症的哺乳动物的大小而变化。在一些情况下,向患者组合施用两种或更多种抑制或结合由选定淋巴细胞的多核苷酸编码的多肽的鉴别的靶标的药物,例如选定的抗体、TCR或其抗原结合片段;肽;核酸;或小分子。组合包括同时或随后施用抑制或结合由选定淋巴细胞的多核苷酸编码的多肽的鉴别的靶标的药物,例如选定的抗体、TCR或其抗原结合片段;肽;核酸;或小分子。

[0422] 当患者经历部分或全部缓解或者疾病的体征或症状减轻时,获得应答,并且具体包括但不限于生存期的延长。预期的无进展生存时间可测量为数月至数年,取决于预后因素,包括复发次数、疾病阶段和其他因素。延长的生存期包括但不限于至少1个月、约至少2个月、约至少3个月、约至少4个月、约至少5个月、约至少1年、约至少2年、约至少3年等时间。总生存期也可测量为数月至数年。患者的症状可保持静止或者可减少。

[0423] 具有本领域普通技术的医师或兽医可容易地确定并指定所需组合物的有效量(ED50)。例如,医师或兽医可以以低于所需水平的水平开始组合物中所用化合物的剂量,以便达到期望的治疗效果并逐渐增加剂量,直到达到期望的效果。

[0424] 可通过如上所述的任何方便的途径向患者施用组合物。不论所选择的施用途径如何,组合物均被配制成如下所述的可接受的剂型或通过本领域技术人员已知的其他常规方法进行配制。

[0425] 可改变组合物中活性成分的实际剂量水平,以便获得有效实现对特定患者、组合物和给药模式的期望治疗反应而对患者没有毒性的活性成分的量。所选择的剂量水平将取决于多种因素,包括所用特定化合物的活性,给药途径,给药时间,所用特定化合物的排泄率,治疗持续时间,与所用特定组合物组合使用的其他药物、化合物和/或材料,所治疗患者

的年龄、性别、体重、状况、总体健康状况和既往病史,以及医学领域公知的类似因素。

[0426] 可使用本领域已知的方法例如化学偶联、共价或非共价键或者重组技术将Ig或TCR多肽或其他药物与治疗部分或可检测(成像)部分相组合,以产生偶联物或如下文更详细描述融合蛋白。或者,可在单独的组合物中将Ig或TCR多肽和/或其他药剂组合起来,以供同时或顺序施用。

[0427] 药物组合物

[0428] 本文所述的每种化合物在与可接受的载体或赋形剂组合时均可作为组合物使用。这样的组合物用于体外或体内分析,或者用于体内或离体施用于受试者以用所公开化合物治疗受试者。

[0429] 因此,除了活性成分以外,药物组合物还可包含药学上可接受的赋形剂、载体、缓冲剂、稳定剂或本领域技术人员公知的其他物质。这样的物质应该是无毒的,并且不应干扰活性成分的功效。载体或其他物质的确切性质将取决于给药途径。

[0430] 可以以冻干制剂或水溶液的形式,通过将具有期望纯度的Ig或TCR多肽与任选的生理学上可接受的载体、赋形剂或稳定剂混合来制备包含由本文所述方法鉴别的由淋巴细胞的选定多核苷酸编码的感兴趣蛋白质(例如Ig或TCR多肽)的药物制剂用于储存(Remington's Pharmaceutical Sciences第16版,Oslo,A编著(1980))。可接受的载体、赋形剂或稳定剂是那些在所用剂量和浓度下对受体无毒的载体、赋形剂或稳定剂,并且包括缓冲剂如磷酸盐、柠檬酸盐和其他有机酸;抗氧化剂包括抗坏血酸和甲硫氨酸;防腐剂(如十八烷基二甲基苄基氯化铵;六甲氯铵;氯化苄甲烃铵、苄索氯铵;苯酚、丁醇或苯甲醇;对羟基苯甲酸烷基酯如对羟基苯甲酸甲酯或对羟基苯甲酸丙酯;邻苯二酚;间苯二酚;环己醇;3-戊醇;和间甲酚);低分子量(少于约10个残基)多肽;蛋白质,如血清白蛋白、明胶或免疫球蛋白;亲水聚合物如聚乙烯吡咯烷酮;氨基酸如甘氨酸、谷氨酰胺、天冬酰胺、组氨酸、精氨酸或赖氨酸;单糖、二糖和其他碳水化合物,包括葡萄糖、甘露糖或糊精;螯合剂如EDTA;糖如蔗糖、甘露糖醇、海藻糖或山梨糖醇;成盐抗衡离子如钠;金属络合物(例如,Zn-蛋白质络合物);和/或非离子表面活性剂如TWEEN<sup>®</sup>、PLURONICS<sup>®</sup>或聚乙二醇(PEG)。

[0431] 可接受的载体对给药的患者在生理学上是可接受的,并且保持所施用化合物的治疗特性。可接受的载体及其制剂通常描述于例如,Remington's pharmaceutical Sciences(第18版,A.Gennaro编著,Mack Publishing Co.,Easton,PA 1990)中。一种示例性载体为生理盐水。如本文所用的短语“药学上可接受的载体”意指参与从一个器官或身体的一部分的给药部位到另一个器官或身体的一部分,或在体外测定系统中携带或运输主题化合物的药学上可接受的物质、组合物或媒介物,如液体或固体填充剂、稀释剂、赋形剂、溶剂或包封材料。每种载体在与制剂的其他成分相容的意义上是可接受的,并且对被施用的受试者无害。可接受的载体也不应改变主题化合物的比活性。

[0432] 在一个方面,本文提供了药学上可接受的或生理学上可接受的组合物,该组合物包含与药物施用相容的溶剂(水性或非水性)、溶液、乳液、分散介质、包衣材料、等渗剂和吸收促进剂或延迟剂。因此,药物组合物或药物制剂是指适合于在受试者中药物使用的组合物。药物组合物和制剂包含一定量的本文所述化合物和药学上或生理学上可接受的载体。

[0433] 组合物可配制与特定给药途径(即全身或局部)相容。因此,组合物包含适合于

通过各种途径给药的载体、稀释剂或赋形剂。

[0434] 在另一种情况下,如果需要,组合物可进一步包含可接受的添加剂,以便改善化合物在组合物中的稳定性和/或控制组合物的释放速率。可接受的添加剂不改变主题化合物的比活性。示例性的可接受的添加剂包括但不限于糖如甘露糖醇、山梨糖醇、葡萄糖、木糖醇、海藻糖、山梨糖、蔗糖、半乳糖、葡聚糖、右旋糖、果糖、乳糖及其混合物。可接受的添加剂可与可接受的载体和/或赋形剂如右旋糖组合。或者,示例性的可接受的添加剂包括但不限于表面活性剂如用来增加肽的稳定性并减少溶液的胶凝的聚山梨醇酯20或聚山梨醇酯80。表面活性剂可以以溶液的0.01%至5%的量添加至组合物中。这样的可接受的添加剂的添加增加了组合物在储存中的稳定性和半衰期。

[0435] 可以例如通过注射施用药物组合物。用于注射的组合物包括水溶液(其为水溶性的)或分散液和用于临时制备无菌可注射溶液或分散液的无菌粉末。对于静脉内施用,合适的载体包括生理盐水、抑菌水、Cremophor EL™ (BASF, Parsippany, N.J.) 或磷酸盐缓冲盐水(PBS)。载体可以是含有例如水、乙醇、多元醇(例如,甘油、丙二醇和液体聚乙二醇等)的溶剂或分散介质及其合适的混合物。例如,通过使用包衣材料如卵磷脂,通过在分散的情况下维持所需的颗粒大小以及通过使用表面活性剂来维持流动性。抗菌剂和抗真菌剂包括例如,对羟基苯甲酸酯、氯丁醇、苯酚、抗坏血酸和硫柳汞。可在组合物中包含等渗剂,例如糖、多元醇如甘露糖醇、山梨糖醇和氯化钠。所得到的溶液可进行包装以按原样使用或冻干;随后可在施用之前将冻干制剂与无菌溶液组合。对于静脉内注射,或在痛苦部位注射,活性成分将为肠胃外可接受的水溶液形式,其无热原并具有合适的pH、等渗性和稳定性。本领域相关技术人员能够熟练使用例如等渗媒介物如氯化钠注射液、林格氏注射液、乳酸林格氏注射液来制备合适的溶液。根据需要,可包含防腐剂、稳定剂、缓冲剂、抗氧化剂和/或其他添加剂。可通过将所需量的活性成分与上述列举的一种成分或成分的组合(如果需要)一起并入合适的溶剂中,然后过滤灭菌来制备无菌可注射溶液。通常,通过将活性成分并入含有碱性分散介质和上述列举的所需其他成分的无菌媒介物中来制备分散液。在用于制备无菌可注射溶液的无菌粉末的情况下,优选的制备方法是真空干燥和冷冻干燥,该方法从其先前无菌过滤的溶液中得到活性成分加任何其他期望成分的粉末。

[0436] 例如,组合物可按照惯例静脉内施用,如通过注射单位剂量。对于注射,活性成分可以是胃肠外可接受的水溶液形式,其基本不含热原并具有合适的pH、等渗性和稳定性。可使用例如等渗媒介物如氯化钠注射液、林格氏注射液、乳酸林格氏注射液来制备合适的溶液。根据需要,可包含防腐剂、稳定剂、缓冲剂、抗氧化剂和/或其他添加剂。此外,可经由雾化施用组合物。(Lahn等人, Aerosolized Anti-T-cell-Receptor Antibodies Are Effective against Airway Inflammation and Hyperreactivity, Int. Arch. Allergy Immunol., 134:49-55 (2004))。

[0437] 在一种情况下,将组合物冻干,例如,以便增加储存保质期。当考虑将组合物用于药品或本文提供的任何方法时,预期组合物可基本不含热原,从而使得组合物在施用于人类患者时不会引起炎症反应或不安全的过敏反应。测试组合物的热原并制备基本不含热原的组合物是本领域普通技术人员公知的,并且可使用可商购获得的试剂盒来完成。

[0438] 可接受的载体可含有使吸收或清除稳定、增加或延迟的化合物。这样的化合物包括例如,碳水化合物,如葡萄糖、蔗糖或葡聚糖;低分子量蛋白质;减少肽的清除或水解的组

合物;或赋形剂或其他稳定剂和/或缓冲剂。延迟吸收的试剂包括例如,单硬脂酸铝和明胶。洗涤剂也可用于稳定或增加或减少药物组合物(包括脂质体载体)的吸收。为了保护免于消化,可将化合物与组合物复合以使其抗酸和酶促水解,或者可将化合物在适当的抗性载体如脂质体中复合。保护化合物免于消化的方法是本领域已知的(参见例如,Fix (1996) Pharm Res.13:1760 1764;Samanen (1996) J.Pharm.Pharmacol.48:119 135;和美国专利号5,391,377,其描述了用于口服递送治疗剂的脂质组合物)。

[0439] 可以以与剂量制剂相容的方式并以治疗有效量施用组合物。待施用的量取决于待治疗的受试者、受试者免疫系统利用活性成分的能力以及期望的结合能力的程度。需要施用的活性成分的精确量取决于从业者的判断,并且对于每个个体是特定的。用于初始施用和加强注射的合适方案也是可变的,但典型的是先初始施用,然后通过随后的注射或其他施用以一个或多个小时间隔进行重复给药。或者,考虑足以维持血液浓度的连续静脉内输注。

[0440] 一种情况考虑使用本文所述的组合物来制备用于治疗病况、疾病或病症的药品。例如,在鉴别由来自选定淋巴细胞的多核苷酸编码的多肽(的靶标)之后,将抑制所鉴别的(靶标)的药物,例如抗体、肽、核酸或小分子制成用于治疗病况、疾病或病症的药品。药品可基于需要治疗的患者/受试者的身体特征进行配制,并且可基于病况、疾病或病症的阶段以单一制剂或多个制剂进行配制。药品可包装在合适的包装中,并具有合适的标记以供分发到医院和诊所,其中标记用于指示患有本文所述疾病的受试者的治疗。药品可包装成单个或多个单位。针对组合物的剂量和施用的说明书可包括在如下所述的包装内。本发明进一步涉及上文所述的人源化抗靶蛋白抗体、TCR或其抗原结合片段和药学上可接受的载体的药品。

#### [0441] 制品

[0442] 在本发明的一个方面,提供了含有用于治疗、预防和/或诊断上述病症的材料的制品。该制品包含容器以及在容器上或与容器相关联的标记或包装插页。合适的容器包括例如,瓶子、小瓶、注射器、IV输液袋等。容器可由多种材料如玻璃或塑料形成。该容器容纳组合物,该组合物本身或与另一组合物组合对于治疗、预防和/或诊断病况是有效的,并且可具有无菌入口(例如,该容器可以是静脉输液袋或具有可被皮下注射针刺穿的塞子的小瓶)。组合物中的至少一种活性剂是由选定多核苷酸编码的Ig或TCR多肽。标记或包装插页指示组合物用于治疗所选择的病况。此外,该制品可包含(a)其中含有组合物的第一容器,其中该组合物包含由本发明的选定多核苷酸编码的Ig或TCR多肽;和(b)其中含有组合物的第二容器,其中该组合物包含另外的细胞毒性剂或其他治疗剂。本发明的在这种情况下下的制品可进一步包含包装插页,其指示组合物可用于治疗特定病况。备选地或额外地,该制品可进一步包含第二(或第三)容器,该容器包含药学上可接受的缓冲剂,如注射用抑菌水(BWFI)、磷酸盐缓冲盐水、林格氏溶液和右旋糖溶液。该制品可进一步包含从商业和用户角度来看所需的材料,包括其他缓冲剂、稀释剂、过滤器、针和注射器。

#### [0443] 包装和试剂盒

[0444] 在另一些情况下,本申请涉及与上述化合物一起使用的试剂盒。可在试剂盒中提供结合靶蛋白的选定的抗体、TCR或其抗原结合片段。因此,试剂盒将在合适的容器装置中包含含有结合靶蛋白的Ig或TCR多肽的组合物。试剂盒可在合适的容器装置中包含结合靶

蛋白的Ig或TCR多肽。

[0445] 试剂盒的容器装置通常将包括至少一个小瓶、试管、烧瓶、瓶子、注射器和/或其他容器装置,至少一种多肽可置于这些容器装置中,和/或优选恰当地在这些容器装置中等分。试剂盒可包括用于容纳至少一种融合蛋白、可检测部分、报道分子和/或严格限制用于商业销售的任何其他试剂容器的装置。这样的容器可包括其中保留所需小瓶的注射和/或吹塑成型的塑料容器。试剂盒还可包括针对试剂盒中的材料的使用的印刷材料。

[0446] 包装和试剂盒可另外包含药物制剂中的缓冲剂、防腐剂 and/或稳定剂。试剂盒的每个组分均可封装在单独的容器内,并且所有不同的容器均可位于单个包装内。本发明试剂盒可被设计用于冷藏或室温储存。

[0447] 此外,制剂可含有稳定剂,以增加试剂盒的保存期,并且包括例如牛血清白蛋白(BSA)。当组合物被冻干时,试剂盒可含有另外的溶液制剂以重构冻干制剂。可接受的重构溶液是本领域公知的,并且包括例如,药学上可接受的磷酸盐缓冲盐水(PBS)。

[0448] 此外,本文提供的包装或试剂盒可进一步包括本文提供的任何其他部分,例如,一种或多种报道分子和/或一种或多种可检测部分/药剂。

[0449] 包装和试剂盒可进一步包括用于测定,例如ELISA测定的一个或多个组件。本申请中的待测试样品包括例如,活检物和组织切片。包装和试剂盒可进一步包括用于样品收集的一个或多个组件(例如,注射器、杯、拭子等)。

[0450] 包装和试剂盒可进一步包括指定例如产品描述、给药模式和/或治疗指示的标记。本文提供的包装可包括如本文所述的任何组合物。该包装可进一步包括用于治疗疾病的标记。

[0451] 说明书可包括用于实施本文所述的任何方法(包括治疗方法)的说明书。说明书还可包括满意的临床终点的适应症或可能发生的任何不良症状,或监管机构如食品与药品监督管理局要求的针对人类受试者使用的附加信息。

[0452] 说明书可在“印刷品”上,例如,位于在试剂盒内或贴在试剂盒上的纸或纸板上,或位于贴在试剂盒或包装材料上的标记上,或附在含有试剂盒组分的小瓶或管上。说明书还可包含在计算机可读介质,如磁盘(软盘或硬盘)、光学CD如CD-ROM/RAM或DVD-ROM/RAM、磁带、电子存储介质如RAM和ROM、IC chip以及这些介质的杂合物如磁/光存储介质中。

[0453] 另外的实施方案

[0454] 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个其他步骤,则可以理解,这些其他步骤中的每一个均可用所公开方法的任何具体实例或实例的组合来进行,并且每个这样的组合均被具体涉及,并且应被认为是已公开的。

[0455] 虽然本文中已经示出并描述了本文所述的一些实例,但这些实例仅以示例的方式

提供。本领域技术人员在不脱离本文所提供的公开内容的情况下现将想到多种变化、改变和替代。应当理解,本文中所述的实例的各种替代方案可用于实施本文所述的方法。

[0456] 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例: The Merck Manual of Diagnosis and Therapy,第18版,Merck Research Laboratories 出版,2006; Benjamin Lewin, Genes IX, Jones & Bartlett Publishing 出版,2007; Kendrew 等人(编), The Encyclopedia of Mol. Biology, Blackwell Science Ltd. 出版, 1994; 以及 Robert A. Meyers (编), Mol. Biology and Biotechnology: a Comprehensive Desk Reference, VCH Publishers 出版, Inc., 1995。

[0457] 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. Kimmerl (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)。

[0458] 实施例

[0459] 实施例1-为进行基于乳液的大规模高通量单个细胞多核苷酸测序而准备细胞的方案。

[0460] 获得来自包含肿瘤浸润性淋巴细胞 (TIL) 的患病组织样品的细胞群体。还获得相应的正常组织样品或 NAT 样品。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。样品包含淋巴细胞以及正常细胞和/或患病细胞 (例如, 癌细胞)。

[0461] 将 T 细胞或 B 细胞通过在细胞缓冲液: 1x Dulbecco 磷酸盐缓冲盐水 (PBS) 中以 200g 离心 10min (两次) 来进行洗涤。然后在细胞缓冲液中将细胞稀释到  $3.5 \times 10^6$  个细胞/mL 的细胞浓度。然后将悬浮液移液穿过 20  $\mu$ m 细胞过滤器。

[0462] 实施例2-为进行基于乳液的大规模高通量单个细胞多核苷酸测序而制备乳液反应混合物的方案。

[0463] 将含有以下表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。

|        | 试剂                           | 原液浓度<br>(mM)          | 小滴中的最终浓<br>度 (mM)     | 反应相中的最<br>终浓度 (mM)    | $\mu\text{L}/200$<br>$\mu\text{L}$ |
|--------|------------------------------|-----------------------|-----------------------|-----------------------|------------------------------------|
|        |                              |                       |                       |                       |                                    |
| [0464] | Tris-Cl, pH 8.0              | 500.00                | 50.00                 | 100.00                | 40.00                              |
|        | MgSO <sub>4</sub>            | 100.00                | 3.00                  | 6.00                  | 12.00                              |
|        | DTT                          | 1,000.00              | 10.00                 | 20.00                 | 4.00                               |
|        | 各种 dNTP                      | 10.00                 | 0.50                  | 1.00                  | 20.00                              |
|        | 5'生物素寡聚-dT                   | $1.40 \times 10^{-2}$ | $2.50 \times 10^{-4}$ | $5.00 \times 10^{-4}$ | 7.14                               |
|        | 模板转换寡核苷酸                     | 0.1                   | $1.00 \times 10^{-3}$ | $2.00 \times 10^{-3}$ | 4.00                               |
|        | DB 模板分子/ $\mu\text{L}$       | $1.00 \times 10^6$    | $1.75 \times 10^4$    | $3.50 \times 10^4$    | 7.00                               |
|        | DB 正向引物                      | 0.2                   | $5.00 \times 10^{-4}$ | $1.00 \times 10^{-3}$ | 1.00                               |
|        | DB 反向引物                      | 0.2                   | $7.50 \times 10^{-4}$ | $1.50 \times 10^{-3}$ | 1.50                               |
|        | HALT 蛋白酶抑制剂 (X)              | 200                   | 1.00                  | 2.00                  | 2.00                               |
|        | RNA 酶抑制剂 (U/ $\mu\text{L}$ ) | 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                            |                       |                       |                       | 至 200                              |

[0465] 寡核苷酸序列:

|        |                        |                                                                                    |
|--------|------------------------|------------------------------------------------------------------------------------|
| [0466] | 5'生物素寡聚-dT 锚定<br>逆转录引物 | /5BiosG//iSp18/TTT TTT TTT TTT TTT TTT TTT T V N                                   |
|        | 小滴条码模板:                | ATCCATCCACGACTGACGGACGTATTAAANNNNWNNNNWN<br>NNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCACC |
|        | 模板转换寡核苷酸               | AATACGTCCGTCAGTCGTGGATGNNTNNANNTGrGG                                               |
|        | 正向容器条码                 | CATCCACGACTGACGGACGTATT                                                            |
|        | 反向容器条码                 | GTGACTGGAGTTCAGACGTGTGCT                                                           |
|        |                        |                                                                                    |

[0467] /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。

[0468] 实施例3-为进行基于乳液的大规模高通量单个细胞多核苷酸测序而生成乳液的方案。

[0469] 一旦制备了细胞和反应混合物,就形成乳液。使用100 $\mu\text{L}$  Hamilton微升注射器将100 $\mu\text{L}$  PEEK样品环加载到两个注射的各为约100 $\mu\text{L}$ 的反应混合物中。使用100 $\mu\text{L}$  Hamilton气密注射器将约110 $\mu\text{L}$ 的细胞悬浮液加载到约100 $\mu\text{L}$ 、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,通过滴入在冷模块中保持在约0 $^{\circ}\text{C}$ 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50 $\mu\text{L}$ 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40 $\mu\text{L}$  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：

[0470] 1. 42.0℃, 30:00 (逆转录)

[0471] 2. 95.0℃, 05:00 (使逆转录酶和DNA模板变性)

[0472] 3. 95.0℃, 00:10

[0473] 4. 65.0℃, 00:30

[0474] 5. 72.0℃, 00:30

[0475] 6. 去往3, 总计55次循环 (扩增容器条码并与cDNA融合)

[0476] 7. 4.0℃, 无时间限制

[0477] 将乳液在4℃下保持过夜。

[0478] 实施例4- 为进行基于乳液的大规模高通量单个细胞多核苷酸测序而打破乳液的方案。

[0479] 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。

[0480] 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：

[0481] 1. 50℃, 15:00 (蛋白酶消化)

[0482] 2. 70℃, 10:00 (蛋白酶灭活)

[0483] 3. 95℃, 03:00 (蛋白酶灭活和DNA变性)

[0484] 4. 4.0℃, 永远

[0485] 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。

[0486] 实施例5- 为进行基于乳液的大规模高通量单个细胞多核苷酸测序而清洁来自乳液的多核苷酸的方案。

[0487] 将0.25V的NEB链霉亲和素珠子加入2x BW (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1x BW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并加至硅胶柱上。然后用0.7mL的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。

[0488] 实施例6- 为进行基于乳液的大规模高通量单个细胞多核苷酸测序, 用于下一代测序的多核苷酸第一PCR反应 (PCR1) 的方案。

[0489] 将163.2μL的纯化cDNA用于PCR1。

[0490]

| PCR1 文库             |         |      |             |          |            |
|---------------------|---------|------|-------------|----------|------------|
| 试剂                  | 原液 (mM) | 最终浓度 | μL/20-μL 反应 | 60-μL 反应 | 4 60-μL 反应 |
| Q5 缓冲液 5X           | 5.00    | 1.00 | 4.00        | 12.00    | 48.00      |
| 各种 dNTP             | 10.00   | 0.20 | 0.40        | 1.20     | 4.80       |
| Q5 热启动              | 125.00  | 1.00 | 0.16        | 0.48     | 1.92       |
| 633 (10μm)          |         |      | 0.16        | 0.48     | 1.92       |
| Ig-C 引物混合物 (各 10μm) |         |      | 0.16        | 0.48     | 1.92       |
| cDNA                |         |      | 13.60       | 40.80    | 163.20     |
| H2O                 |         |      | 1.52        | 4.56     | 18.24      |

[0491] 引物序列

[0492] “Ig-C”混合物:

|        |     |                    |
|--------|-----|--------------------|
| [0493] | IgM | GGGTTGGGGCGGATGCAC |
|        | IgD | CATCCGGAGCCTTGGTGG |
|        | IgA | CCTTGGGGCTGGTCGGGG |

|        |     |                    |
|--------|-----|--------------------|
| [0494] | IgE | CGGATGGGCTCTGTGTGG |
|        | IgG | CCGATGGGCCCTTGGTGG |

|        |       |                            |
|--------|-------|----------------------------|
| [0495] | IGKJ1 | TTTGATCTCCACCTTGGTCCCTCCGC |
|        | IGKJ2 | TTTGATCTCCAGCTTGGTCCCCTGG  |
|        | IGKJ3 | TTTGATATCCACTTTGGTCCCAGGGC |
|        | IGKJ4 | TTTGATTTCCACCTTGGTCCCTTGGC |
|        | IGKJ5 | TTTAATCTCCAGTCGTGTCCCTTGGC |
|        | IGLJ1 | GAGGACGGTCACCTTGGTGCCA     |
|        | IGLJ2 | TAGGACGGTCAGCTTGGTCCCTCC   |
|        | IGLJ3 | GAGGACGGTCAGCTGGGTGCC      |
|        | IGLJ4 | TAAAATGATCAGCTGGGTTCCTCCAC |
|        | IGLJ5 | TAGGACGGTGACCTTGGTCCCAGT   |

|        |     |                                    |
|--------|-----|------------------------------------|
| [0496] | 633 | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT |
|--------|-----|------------------------------------|

[0497] 在PCR管中等分4x60μL反应,并在热循环仪中运行下列程序:

[0498] 1. 98°C, 01:00

[0499] 2. 98°C, 00:10

[0500] 3. 64°C, 00:20

[0501] 4. 72°C, 00:20

[0502] 5. 去往2, 总计6次循环

[0503] 6. 4°C, 无时间限制

[0504] 用1.2倍体积的AMPure XP纯化PCR产物,用80%乙醇洗涤,并在60μL稀释缓冲液(10mM Tris-Cl, pH 8.0, 0.1mM EDTA)中洗脱。

[0505] 实施例7-为进行基于乳液的大规模高通量单个细胞多核苷酸测序,用于下一代测序的多核苷酸第二PCR反应(PCR2)的方案。

[0506] 将20μL纯化的PCR1产物用于每个子文库(例如, IgL链或IgH链, 或TCRα链或TCRβ链)。

| [0507] | PCR2 文库   |         |      |             |             |
|--------|-----------|---------|------|-------------|-------------|
|        | 试剂        | 原液 (mM) | 最终浓度 | μL/20-μL 反应 | 用于 50-μL 反应 |
|        | Q5 缓冲液 5X | 5.00    | 1.00 | 4.00        | 10.00       |
|        | 各种 dNTP   | 10.00   | 0.20 | 0.40        | 1.00        |

|        |                                   |        |      |      |       |
|--------|-----------------------------------|--------|------|------|-------|
| [0508] | Q5 热启动                            | 125.00 | 1.00 | 0.16 | 0.40  |
|        | C7-索引-P7 (2 $\mu$ m)              |        |      | 1.60 | 4.00  |
|        | P5-IgH 或 P5-IgL 混合物 (各 1 $\mu$ m) |        |      | 1.60 | 4.00  |
|        | cDNA                              |        |      | 8.00 | 20.00 |
|        | H2O                               |        |      | 4.24 | 10.60 |

[0509] 引物序列

[0510] P5-IgH (重链) 混合物

|        |     |                                                     |
|--------|-----|-----------------------------------------------------|
| [0511] | IgM | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGTTGGGGCGGATGCAC |
|        | IgD | ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATCCGGAGCCTTGGTGG |
|        | IgA | ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTTGGGGCTGGTCGGGG |
|        | IgE | ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGATGGGCTCTGTGTGG |
|        | IgG | ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGATGGGCCCTTGGTGG |

[0512] P5-IgL (轻链) 混合物

|        |       |                                                             |
|--------|-------|-------------------------------------------------------------|
| [0513] | IGKJ1 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGATCTCCACCTTGGTCCCTCCGC |
|        | IGKJ2 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGATCTCCAGCTTGGTCCCCTGG  |
|        | IGKJ3 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGATATCCACTTTGGTCCCAGGGC |
|        | IGKJ4 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTGATTTCCACCTTGGTCCCTTGGC |
|        | IGKJ5 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTAATCTCCAGTCGTGTCCCTTGGC |
|        | IGLJ1 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGGACGGTCACCTTGGTGCCA     |
|        | IGLJ2 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGACGGTCAGCTTGGTCCCTCC   |
|        | IGLJ3 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGGACGGTCAGCTGGGTGCC      |
|        | IGLJ4 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAATGATCAGCTGGGTTCCTCCAC  |
|        | IGLJ5 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGACGGTGACCTTGGTCCCAGT   |
|        | IGLJ6 | ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGGACGGTCAGCTCGGTCCCC     |

[0514] 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:

[0515] n5'

[0516] CAAGCAGAAGACGGCATACGAGAT[NNNNN]GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT。

[0517] 在热循环仪中运行下列程序:

[0518] 1. 98°C, 01:00

[0519] 2. 98°C, 00:10

[0520] 3. 64°C, 00:20

[0521] 4. 72°C, 00:20

[0522] 5. 去往2, 总计6次循环

[0523] 6. 4°C, 无时间限制

[0524] 用1.2倍体积的AMPure纯化PCR产物,并在40 $\mu$ L稀释缓冲液中洗脱。

[0525] 实施例8-为进行基于乳液的大规模高通量单个细胞多核苷酸测序,用于下一代测序的多核苷酸第三PCR反应(PCR2)的方案。

[0526] 将8 $\mu$ L纯化的PCR2产物用于中试qPCR以确定扩增循环的最终数目。

[0527]

| qPCR3a 文库               |              |      |                                        |
|-------------------------|--------------|------|----------------------------------------|
| 试剂                      | 原液 (mM)      | 最终浓度 | $\mu\text{L}/20\text{-}\mu\text{L}$ 反应 |
| Q5 缓冲液 5X               | 5.00         | 1.00 | 4.00                                   |
| 各种 dNTP                 | 10.00        | 0.20 | 0.40                                   |
| SYBR Green I 1:500      | 83.00        | 1.00 | 0.24                                   |
| Q5 热启动                  | 125.00       | 1.00 | 0.16                                   |
| C5-P5 ( $\mu\text{M}$ ) | <b>10.00</b> | 0.40 | 0.80                                   |
| C7 ( $\mu\text{M}$ )    | <b>10.00</b> | 0.40 | 0.80                                   |
| cDNA                    |              |      | 8.00                                   |
| H2O                     |              |      | 5.60                                   |

[0528] 引物序列

[0529] P5:

[0530] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

[0531] C7:CAAGCAGAAGACGGCATACGAGAT

[0532] 在qPCR机器中运行下列程序:

[0533] 1. 98°C, 01:00

[0534] 2. 98°C, 00:10

[0535] 3. 64°C, 00:20

[0536] 4. 72°C, 00:20

[0537] 5. 读板

[0538] 6. 去往2, 总计25次循环

[0539] 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。

[0540] 将24.0 $\mu\text{L}$ 的纯化PCR2产物用于终点PCR3。

[0541]

| qPCR3b 文库 |         |      |                                        |
|-----------|---------|------|----------------------------------------|
| 试剂        | 原液 (mM) | 最终浓度 | $\mu\text{L}/60\text{-}\mu\text{L}$ 反应 |
| Q5 缓冲液 5X | 5.00    | 1.00 | 12.00                                  |
| 各种 dNTP   | 10.00   | 0.20 | 1.20                                   |
| H2O       | 83.00   | 1.00 | 0.72                                   |
| Q5 热启动    | 125.00  | 1.00 | 0.48                                   |

[0542]

|                         |              |      |       |
|-------------------------|--------------|------|-------|
| C5-P5 ( $\mu\text{M}$ ) | <b>10.00</b> | 0.40 | 2.40  |
| C7 ( $\mu\text{M}$ )    | <b>10.00</b> | 0.40 | 2.40  |
| cDNA                    |              |      | 24.00 |
| H2O                     |              |      | 16.80 |

[0543] 在热循环仪中运行下列程序:

[0544] 1. 98°C, 01:00

[0545] 2. 98°C, 00:10

[0546] 3. 64°C, 00:20

[0547] 4. 72°C, 00:20

[0548] 5. 去往2, 进行确定数目的循环

[0549] 6. 4℃, 永远

[0550] 用1.2倍体积的AMPure纯化PCR产物,并在20μL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。

[0551] 实施例9-人源化选定抗体的表达

[0552] 载体构建体

[0553] 合成含有人源化选定抗体V<sub>H</sub>(H1)和V<sub>L</sub>(K1)区的密码子的两个dsDNA序列。这些合成的序列还含有在5'和3'端添加或保留限制性内切核酸酶位点所必需的核苷酸。所有密码子均针对中国仓鼠卵巢(CHO)细胞中的表达进行优化。用于完成重链和轻链的信号肽和恒定区序列来源于cDNA。所有构建体的编码区序列均通过DNA测序进行确认。蛋白质产物为选定抗体#1(对于IgG<sub>4</sub>)和选定抗体#2(对于IgG<sub>1</sub>)。

[0554] 表达载体构建体

[0555] 将来自上述载体构建体的重链和轻链编码区亚克隆至双顺反子表达载体中。设计引物以生成具有末端限制性位点的编码区,以促进插入到双顺反子表达载体的多克隆位点(MCS)中。此外,添加8个碱基对的限制性位点以促进生成未来构建体。将κ链连接到MCS1的限制性位点中。将IgG1重链连接到MCS的合适的限制性位点中。

[0556] 已经充分确定,IgG<sub>4</sub>可表达为一条重链和一条轻链。为了稳定IgG<sub>4</sub>,将其铰链区用IgG<sub>1</sub>的铰链区取代。因此,在三向连接中,将含有V<sub>H</sub>-C<sub>H</sub>1和铰链区的IgG<sub>1</sub>片段与含有IgG<sub>4</sub>Fc区的IgG<sub>4</sub>片段连接。将合适的引物用于PCR,并将免疫球蛋白序列从载体构建体转移至双顺反子表达载体。

[0557] 实施例10-测量选定抗体对靶蛋白的亲和力

[0558] 可使用常规技术例如表面等离子体共振(SPR;Biacore)评估本文所述抗体及其抗原结合片段对靶蛋白的亲和力。

[0559] 通过使用例如配备有CM5传感器芯片(BIAcore AB)的BIAcore™ 3000分析系统的SPR来确定各种选定抗体和抗原结合片段与靶蛋白的结合的亲和常数。选定抗体或抗原结合片段与CM5传感器芯片共价偶联达到1500个共振单位(使用在10mM醋酸盐缓冲液中的10μg/mL的浓度和适合用于所测试的特异性选定抗体或抗原结合片段的pH)。将靶蛋白以30μL/min的流速、5至250nM的浓度注入(40μL)。在每个循环之后,使用10微升的10mM HCl溶液使芯片再生。采用BIAcore™ 3000(Langmuir结合模型)的软件计算结合和解离速率常数。

[0560] 实施例11-选定抗体对不同种类靶蛋白的亲和力

[0561] 通过P-ELISA确定选定抗体#3与大鼠、小鼠、兔和人靶蛋白的结合。选定抗体的相对亲和力为人=兔>小鼠>大鼠。通过ELISA的初步评估得到对人靶蛋白的亲和力相对于亲本小鼠抗体高大约2至4倍。对人靶蛋白和兔靶蛋白的亲和力可能似乎相对于亲本小鼠抗体高4至5倍。

[0562] 人源化抗体与小鼠靶蛋白结合。选定抗体对小鼠靶蛋白的相对亲和力和小鼠亲本抗体与兔靶蛋白结合的亲和力大约相同。由于亲本抗体在兔疾病模型中显示出功效,因此可预期选定抗体在小鼠疾病模型中显示功效。人源化过程中接近CDR的变化导致对人靶蛋白更高的亲和力,以及对小鼠和大鼠靶蛋白的显著反应性。选定抗体#2对小鼠靶蛋白的亲和力似乎相对于亲本小鼠抗靶蛋白高超过10倍。

[0563] 实施例12-选定的抗体与靶蛋白的结合常数的测定

[0564] 进行该实验以测定人源化选定抗体和相应的亲本小鼠抗体的结合常数。

[0565] 以5种不同的表面密度将人源化选定抗体捕获到抗人IgG表面上。将人源化选定抗体#1和亲本小鼠抗体#1稀释至100nM的起始浓度,并使用具有0.005%吐温-20和0.1mg/ml BSA的PBS以3倍稀释系列进行测试。在25℃下收集结合数据。监测结合阶段5min,并监测解离阶段2.5hr。将5个不同密度抗体表面上的每种抗原的响应数据全局拟合为简单的1:1相互作用模型。确定数据的拟合,并在25℃下确定结合常数。下表中提供了示例性结合常数的汇总。

| [0566] |           | $K_a (M^{-1}s^{-1})$ | $K_d (s^{-1})$     | $K_D$ |
|--------|-----------|----------------------|--------------------|-------|
|        | 人源化选定抗体#1 | $8 \times 10^5$      | $8 \times 10^{-6}$ | 15pM  |
|        | 亲本小鼠抗体#1  | $4 \times 10^5$      | $4 \times 10^{-6}$ | 10nM  |

[0567] 实施例13-选定抗体与靶蛋白的结合常数的测定

[0568] 利用若干生物分析测定来支持最终药物候选物的选择和初始药代动力学评估。这些测定包括由固定至链霉亲和素包被的微量滴定孔的n末端生物素标记的靶蛋白组成的靶蛋白ELISA (P-ELISA)。用HRP偶联的抗人抗体检测选定抗体的结合。该测定的灵敏度被确定为10-20ng/ml。

[0569] 使用Neutravidin™包被板的ELISA方案

[0570] 将所有试剂置于室温下,并在洗涤缓冲液(1X TBS,0.1%BSA,0.05%吐温)中进行稀释。简言之,方案步骤如下:将100μL的Neutravidin™ Pierce#31000 (0.5μg/ml,在TBS中)添加至96孔Immulon-4板。在室温下温育1小时。用200μl洗涤缓冲液将孔洗涤3次。添加50μL的生物素化靶蛋白(0.06nM)。在室温下温育1hr。在洗涤缓冲液中将板洗涤3次。添加100μl的选定抗体。在室温下温育30min。在洗涤缓冲液中将板洗涤3次。添加50μl的第二Ab-HRP (1:10,000)。在室温下温育30min。在洗涤缓冲液中将板洗涤4次。添加100μl TMB试剂(底物)。在室温下温育。添加100μL的2M硫酸,以终止底物的显影。使用450nm滤光片,并以615-620nm滤光片作为参照进行读板。

[0571] 实施例14-免疫组织化学染色

[0572] 将切割至Vectabond包被的载玻片上的冷冻切片(10μm)固定在甲醇(-20℃,5min)中并使用如本领域先前描述的三步过氧化物酶法进行染色。简言之,将这些切片用一级选定抗体在4℃下标记过夜或用针对靶蛋白#1、磷酸化靶蛋白#2、非磷酸化靶蛋白#2或CD45的抗体在室温下标记1hr。随后用合适的辣根过氧化物酶(HRP)偶联物进行温育。将针对CD45染色的切片用Mayer苏木精进行复染。通常将缺乏一级抗体、二级抗体或抗生物素蛋白生物素复合体用作对照。

[0573] 实施例15-蛋白质提取和蛋白质印迹法

[0574] 将动物的速冻患病样品称重、细切,并以1:10g湿重/ml重悬于pH 7.4的Tris-HCl缓冲液(100mM Tris,5mM EDTA,150mM NaCl,含有1%Triton X-100)中。使用高强度超声处理器将样品均质化并在冰上温育30min。在Eppendorf离心机中将组织悬浮液在15,000g、4℃下旋转60min,收集上清液并在-70℃下等分储存。通过Folin酚法(Lowry等人,J.Biol.Chem,193:265-75(1951))确定脊髓匀浆的总蛋白浓度。

[0575] 对于蛋白质印迹分析,在Tris-HCl十二烷基硫酸钠聚丙烯酰胺凝胶上解析40μg的

上清液蛋白,并将其转移至Immobilon-P聚偏二氟乙烯膜。将膜上的非特异性结合位点用溶于Tris缓冲盐水(TBST) (10mM Tris,pH 7.4,150mM NaCl和0.1%吐温20)的5% Marvel®干脱脂牛奶在室温下封闭1hr,然后用以1:1000稀释于TBST中的5% Marvel®的一级抗体在室温下温育2hr。表1总结了一级抗体来源和运行条件。在TBST中洗涤后,将膜用二级抗体在室温下温育1hr,该二级抗体与HRP:抗小鼠IgG HRP、抗兔IgG HRP或抗山羊IgG HRP偶联。经过三次最终洗涤后,通过增强的化学发光使印迹显影。为了获得特定蛋白质的半定量测量值,使用分析软件包分析产生的印迹,并以任意单位测量条带密度。为了确保相等的蛋白质加载,将膜分成条带并用对照抗体进行探测,以相对于针对此类合适的测定的表达水平进行归一化。

[0576] 实施例16-用于评估靶蛋白突变体的结合亲和力的酶联免疫吸附测定 (ELISA)

[0577] 在4℃下,将Costar 96孔板用4μg/ml的针对靶蛋白#1、靶蛋白#1突变体-1或靶蛋白#1突变体2的小鼠抗体包被48hr。用在1x PBS中的1%BSA将孔在4℃下封闭过夜,然后用1x PBS吐温80 (0.004%) 洗涤板。将蛋白提取物样品和标准品稀释于含有0.004%吐温80、0.1%BSA和5mM EDTA的1x PBS中,并且每孔添加180μl,并在4℃下温育过夜。生成标准曲线。洗涤后,添加生物素化二级抗体在37℃下1hr。在添加ABC复合体(载体)在室温下1hr后,使用邻苯二胺使板显影,并使用4M硫酸终止反应。在490nm处读取吸光度,并读取650nm处的参照读数。通过ELISA并根据制造商的说明书进行靶抗原底物活性的评估。

[0578] 实施例17-免疫组织化学染色

[0579] 在室温下将患病样品的冷冻切片(5μm厚)在0.07M磷酸盐缓冲盐水(PBS;pH 7.0)中的1%多聚甲醛中固定5min或在丙酮中固定10min,然后用一级抗体温育。在室温下将选定抗体温育1小时。在PBS中洗涤切片(三次,每次10min),并添加与辣根过氧化物酶(HRP)或异硫氰酸四甲基罗丹明(TRITC)偶联的合适的二级抗体并应用30min。对于免疫染色,添加一滴3-氨基-9-乙基咔唑(AEC+;Dako),并在苏木精中将切片复染1min。对于免疫荧光染色,在PBS中洗涤3次(每次10min)并在10mM Tris-HCl缓冲液(pH 8.8)中进行最终漂洗后,在配备有落射荧光光学器件的倒置显微镜下分析标记。通过同时分析正常样品或正常邻近组织与患病样品来评估染色的特异性。在计算机(Prism 3.0;GraphPad, San Diego CA)上分析数据。使用曼-惠特尼检验(Mann-Whitney test)来确定不同实验条件之间是否存在显著( $P < 0.05$ )差异。

[0580] 实施例18-免疫荧光筛选

[0581] 使用多个选定抗体进行免疫荧光筛选测定。来自七名不同胰腺癌患者的患病样品显示,106个抗体中88个(83%)在全部7名患者中表现出强染色。与正常邻近组织相比,多个抗体中的这些选定抗体中的许多也表现出对胰腺肿瘤组织的高特异性。进行与质谱分析相结合的免疫沉淀实验,以阐明未知的靶蛋白身份。多个选定抗体的抗体A1-108在用于甲醛固定石蜡包埋(FFPE)癌症组织样品的免疫组织荧光染色时,也表现出胰腺导管腺癌组织相对于正常邻近组织的强特异性染色。此外,A1-108使胰腺的腺鳞癌和神经内分泌癌均染色。另外14个正常组织的A1-108染色仍然很弱或完全缺乏。

[0582] 实施例19-靶蛋白的选定抗体中和

[0583] 确定选定抗体中和靶蛋白、抑制底物蛋白酶活性的能力。数据表明,人源化选定抗体#1的中和活性等同于亲本小鼠抗体#1。人类抗体对照不中和靶蛋白活性。在最少三次测

定中比较选定抗体和变体的中和活性。

[0584] 实施例20-靶蛋白中和测定

[0585] 抗体及其抗原结合片段的功能特性可通过利用靶蛋白中和测定评估其抑制活性靶蛋白的能力进行确定。

[0586] 使用酶偶联显色法确定靶蛋白活性。简言之,在96孔微量滴定板的孔中将25 $\mu$ L靶蛋白(50ng/mL活性靶蛋白)与等体积的TBS缓冲液(0.05M Tris-HCl,0.01M NaCl,pH 7.4,含有0.01%吐温80)或者选定抗体或其抗原结合片段的连续2倍稀释液一起温育,从而导致摩尔过量(抗体:靶蛋白)在1至128之间。使混合物在室温下反应2hr。随后,添加50 $\mu$ L靶抗原底物(20IU/mL或40ng/mL),并将该板在37 $^{\circ}$ C下温育15min。然后,添加100 $\mu$ L含有酶(1 $\mu$ M)、CNBr消化的酶底物(1 $\mu$ M)和发色团(0.6mM)的溶液。记录405nm处的吸光度变化,以测量残余靶蛋白活性。100%靶蛋白活性是在抗体缺乏的情况下观察到的靶蛋白活性。由抗体存在的情况下测量的残余靶蛋白活性计算抗体的抑制百分比(即靶蛋白活性的中和)。

[0587] 实施例21-抗体中和测定

[0588] 测试针对选定抗体的靶标的活性测定,该测定测量选定抗体对其靶标活性的抑制。该测定可用于确定抗体中和靶标的效率。

[0589] 将所有试剂置于室温下,并将读板器预热至37 $^{\circ}$ C。在测定缓冲液(0.15M NaCl,0.05M Tris(pH 7.5),0.01%吐温,100 $\mu$ g/ml BSA)中进行所有稀释。最终条件如下:100 $\mu$ L--重复孔,1.5U靶蛋白的酶底物/孔,8nM活性人类野生型靶蛋白,25 $\mu$ L显色底物,0-80 $\mu$ g/ml选定抗体。测定步骤如下:将50 $\mu$ L选定抗体稀释液置于96孔中;添加25 $\mu$ L靶蛋白底物酶(1.5U),在读板器上摇动3sec;在37 $^{\circ}$ C下温育5min;添加25 $\mu$ L显色底物以使该板显影。将板摇动3sec,并在37 $^{\circ}$ C下在具有405nm滤光片的读板器上每5分钟读数一次达到30min。由平均值V计算活性百分比(%)。

[0590] 实施例22-靶蛋白失活的测定

[0591] 可使用常规技术确定本文所述的选定抗体或其抗原结合片段对靶蛋白失活速率的影响。例如,可计算在选定抗体或其抗原结合片段存在的情况下靶蛋白的半衰期。

[0592] 在37 $^{\circ}$ C下将靶蛋白(在PBS中40 $\mu$ g/mL)与3倍摩尔过量的选定抗体或其抗原结合片段一起温育。在不同的时间间隔,取出等分试样并将其与2倍摩尔过量的靶抗原底物一起温育(37 $^{\circ}$ C下25min)。通过SDS-PAGE分析反应产物,然后进行银染色。通过随后的光密度扫描进行反应产物的定量。基于在每个时间点的活性靶蛋白的量,可计算在选定抗体或其抗原结合片段存在的情况下靶蛋白的半衰期。

[0593] 实施例23-测量由靶抗原活性生成的反应产物对底物的抑制

[0594] 可使用常规技术评估本文所述的选定抗体或其抗原结合片段对在靶蛋白与底物相互作用期间生成的反应产物的影响。

[0595] 简言之,在缺乏(对照)或存在8倍摩尔过量的选定抗体或抗原结合片段的情况下将靶蛋白(40 $\mu$ g/mL,在PBS中)在37 $^{\circ}$ C下温育10min。然后将样品与2倍摩尔过量的底物一起温育(37 $^{\circ}$ C下25min)。通过添加SDS(终浓度1%)并在100 $^{\circ}$ C下加热30sec终止反应。通过SDS-PAGE分析反应产物,然后用考马斯亮蓝染色。通过随后的光密度扫描进行反应产物的定量。

[0596] 实施例24-选定抗体作为疾病治疗剂的体内评估。

[0597] 将动物分成不同的治疗组,每个治疗组放置有多个动物。诱导疾病。然后在如试验

期间确立的多个给药方案中预先确定的时间点向动物测试组施用抗靶蛋白抗体或抗原结合片段的剂量。通过在整个治疗期间经由ELISA或HPLC确定与疾病进展或消退相关的分析物的水平或水平变化来评估治疗功效。在整个治疗期间处死动物,以检查来自动物的各种生物样品如器官,以证明与疾病相关的形态学变化。此外,在整个治疗期间进行免疫组织化学染色,以证明与疾病相关的分子变化。可经由疾病的动物模型测试本文所述的抗蛋白抗体和抗原结合片段用于治疗疾病的功效。

#### [0598] 实施例25-血浆中蛋白质抗体的检测

[0599] P-ELISA可用于监测PK和功效研究中选定抗体的血浆水平。与对照IgG、缺乏血浆的抗体或抗体+EDTA相比,P-ELISA能够检测掺入的血浆样品中的蛋白抗体。可确定实现通过P-ELISA对血浆样品中选定抗体的检测的变量的影响。这包括样品处理和储存条件。

#### [0600] 实施例26-免疫测序V2

[0601] 使用独特标识符 (UID) 条码标记每一个单个RNA分子。然后将UID扩增许多拷贝,从而在测序后多个测序读取折叠成具有更高碱基准确性的单一序列,并相对于PCR或测序错误而揭示真正抗体序列和突变。UID也用于追踪多个样品中的污染。

#### [0602] 起始材料

[0603] 将由编码抗体的V、D、J基因区段组成且含有恒定区的免疫细胞的RNA或DNA用作起始材料。在一些实验中, RNA来自T细胞。在一些实验中, RNA为重链 (V、D、J区段) 或轻链 (仅V、J区段)。

#### [0604] 逆转录

[0605] 使用一个或大量由以下部分组成的多核苷酸将RNA逆转录成cDNA: 与RNA区域互补的部分 (通常在mRNA的恒定区或聚A尾)。UID为具有或不具有已知嵌入碱基位置的约20个简并核苷酸的片段 (如NNNNWNNNNWNNNNWNNNNW, 其中W意指A或T)。当将每个RNA分子条码化时, 随着UID长度增加, 其不太可能会被检测到两次。突出端尾 (P5) 充当下游的读取-1测序启动位点。使用多个多核苷酸与各种恒定区退火。每个多核苷酸均带有完整的独特UID, 从而使得每个RNA分子实际上均被UID独特地条码化。

[0606] 在20 $\mu$ l反应中用500ng总RNA进行逆转录, 该反应含有: 5pmol IGHC-UID-P5引物混合物, dNTP各500 $\mu$ M, 5mM DTT, 1 $\mu$ l RNA酶抑制剂 (Enzymatics, Beverly, MA), 1 $\mu$ l在1x First Strand缓冲液 (Life Technologies, Carlsbad, CA) 中的SuperScript II逆转录酶。将反应在55 $^{\circ}$ C下温育45min, 然后在85 $^{\circ}$ C下再温育5min, 以使酶失活。然后添加1 $\mu$ l外切核酸酶I (Enzymatics), 并将反应在37 $^{\circ}$ C下温育15min。在85 $^{\circ}$ C下温育15min后, 添加1 $\mu$ l RNA酶H (Enzymatics), 并将反应在37 $^{\circ}$ C下再温育15min。

#### [0607] PCR1

[0608] 使用以下引物对cDNA进行PCR扩增: (1) 与具有充当读取-2测序和读取-3测序启动位点的突出端尾 (P7) 的V区段上游的RNA互补的正向引物库, 以及 (2) 由具有突出端 (C5) 的P5序列组成的反向引物, 以在Illumina测序平台上簇集。在一些实验中, 正向引物是许多多核苷酸的库, 以供与由免疫细胞表达的所有可能的V区退火。在其他实验中, 正向引物具有P7、SBC和C7突出端。反向引物位于UID之后, 从而使得每个独特的UID得以扩增。

[0609] 在50 $\mu$ l PCR反应中扩增上述制备的20 $\mu$ l逆转录反应物, 该50 $\mu$ l PCR反应含有: 1 $\mu$ M P5/C5引物, 1 $\mu$ M IGHV-P7引物混合物, dNTP各200 $\mu$ M, 1单位在1x Phusion HF缓冲液 (Thermo

Fischer Scientific, Waltham, MA) 中的 Phusion Hotstart II 聚合酶。将反应在 98℃ 下温育 1 个循环, 随后进行 12 个如下的循环: 98℃ 下 10sec; 62℃ 下 20sec; 72℃ 下 20sec; 随后进行一个在 72℃ 下的 3min 循环。

[0610] qPCR

[0611] 然后添加 1μl 外切核酸酶 I (Enzymatics), 并将反应在 37℃ 下温育 20min, 随后在 80℃ 下温育 15min。

[0612] PCR2

[0613] 使用第二 PCR 阶段用以下引物扩增 PCR1 产物: 在 PCR1 中使用的相同 P5C5 反向引物, 以及由 P7 序列和样品条码 (SBC) 组成且具有第二突出端 (C7) 的正向引物, 以在 Illumina 测序平台上簇集。实验中处理的每个样品的样品条码是不同的, 从而使得可在一次测序运行中将多个样品合并在一起。由于在 PCR1 反应中使用了多重引物库, 因此 PCR1 可能引入偏差。通过限制 PCR1 循环的次数并在 PCR2 时普遍扩增, 限制了所引入的偏差。PCR2 还加载了样品条码和簇集标签以供测序。

[0614] 装配 25μl Sybr green qPCR, 其含有 1μM P5-C5 引物, 1μM P7-C7 引物, dNTP 各 200μM, 1x Sybr Green, 和 0.5 单位在 1x Phusion HF 缓冲液 (Thermo Fischer Scientific, Waltham, MA) 中的 Phusion Hotstart II 聚合酶。将反应在 98℃ 下温育 1 个循环, 随后进行 35 个如下的循环: 98℃ 下 10sec; 62℃ 下 20sec; 72℃ 下 20sec; 随后进行一个在 72℃ 下的 3min 循环。

[0615] 在 50μl PCR 反应中扩增 25μl 的 PCR-1 反应物, 该 50μl PCR 反应含有 1μM P5-C5 引物, 1μM P7-SBC-C7, dNTP 各 200μM, 1 单位在 1x Phusion HF 缓冲液 (Thermo Fischer Scientific, Waltham, MA) 中的 Phusion Hotstart II 聚合酶。将反应在 98℃ 下温育 1 个循环, 随后通过 qPCR 分析来确定 PCR 循环次数。循环; N 个如下的循环: 98℃ 下 10sec; 62℃ 下 20sec; 72℃ 下 20sec; 随后为一个在 72℃ 下的 3min 循环。根据制造商方案使样品在 Illumina Miseq 或 HiSeq 系统上进行高通量测序。

[0616] 最终文库

[0617] 得到的文库由具有合适的标签的完整抗体序列和经测序的簇集区段组成。对于每个起始独特 RNA 分子均存在生成的相同 UID 的许多拷贝。一经测序, 即匹配相同的 UID, 并将测序读取折叠成共有序列, 由此消除测序和 PCR 错误。对于读取 -1 (C、J、D、V) 从 P5 位点进行测序, 随后对于读取 -2 (UID 和 VDJ) 从 P7 位点进行测序, 最后对于 SBC 的索引读取 -3 从反向 P7 位点进行测序。

[0618] 实施例 26-免疫测序 V3

[0619] 该实施例描述了在逆转录期间使用模板转换来消除多重 V 引物库的使用, 从而解决了 PCR 偏差的问题。该方法用于抗体下一代测序, 以及独特标识符多核苷酸 (UID) 的并入。

[0620] 起始材料

[0621] 起始材料为来自免疫细胞或 T 细胞的由编码抗体的 V、D、J 基因区段组成且含有恒定区的 RNA 或 DNA。在一些实验中, RNA 包含重链区段 (V、D、J 区段) 或轻链区段 (V、J 区段)。

[0622] 逆转录

[0623] 为了在不需要基因特异性可变区段引物的情况下生成免疫球蛋白重排重链和轻链 cDNA 文库, 首先在模板-开关 (TS) 多核苷酸存在的情况下进行 RNA 样品的逆转录。TS 多核

昔酸含有三个末端核糖鸟苷残基,该残基允许该多核苷酸充当模板用于通过逆转录酶将末端胞嘧啶残基添加至逆转录延伸产物的末端。这在所有cDNA片段的3'端产生通用序列端。关键地,由于TS多核苷酸携带约15个碱基的简并条码序列(通用标识符或UID),因此所有cDNA分子都将携带允许在测序结果中鉴别PCR复制物的不同条码,从而提供如前所述的许多优点。使用一个或大量由以下部分组成的多核苷酸将RNA逆转录成cDNA:与RNA区域互补的部分。在这种情况下,与RNA区域互补的部分与恒定区或mRNA的聚A尾互补。使用多个多核苷酸与各种恒定区退火。本文使用的逆转录酶包含非模板末端转移酶活性。当逆转录酶到达模板末端时,其自然地添加3个非模板胞嘧啶残基。Superscript II (Invitrogen) 用于此目的。

#### [0624] 模板转换

[0625] 先前的逆转录反应在5'标记多核苷酸存在的情况下进行,该5'标记多核苷酸由以下部分组成:用于与测序引物退火的P7区段、UID、与由逆转录酶产生的链互补并退火的3个3'端核糖-鸟嘌呤残基(rGrGrG)(RNA碱基)。在一些实验中,使用3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)。在反应中,一旦标记多核苷酸与cDNA链的CCC退火,则逆转录酶继续将cDNA延伸至标记多核苷酸内,从而对所有cDNA创建通用标签。在其他实验中,模板转换在单独的反应中进行,而不是在进行逆转录反应的同时进行。在这些实验中,在逆转录反应后添加5'标记多核苷酸,并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。因为标记多核苷酸在每个单个分子上带有独特的简并UID,所以每个cDNA均被UID独特地标记。

[0626] 在20 $\mu$ l反应中用模板转换使来自外周血单核细胞(PBMC)的200ng总RNA进行逆转录,该20 $\mu$ l反应含有50mM Tris-HCl,75mM KCl,3mM MgCl<sub>2</sub>,3mM MnCl<sub>2</sub>,10mM二硫苏糖醇,dATP、dGTP、dCTP、dTTP各250 $\mu$ M,2单位/ $\mu$ l RNA酶抑制剂(Enzymatics),10单位/ $\mu$ l MuMLV逆转录酶RNA酶H(NEB),500nM多核苷酸dT(18)引物和500nM TS多核苷酸。建立反应并在42℃下温育45分钟。在AMPure XP珠(Beckman Coulter)上纯化产物,并将产物在20 $\mu$ l H<sub>2</sub>O中洗脱。

#### [0627] PCR1

[0628] 使用由以下部分组成的引物进行PCR:与UID上游的标记多核苷酸末端互补的正向引物(P7),由与RNA(C)互补的区段和用于测序的突出端(P5)组成的反向引物。C区段被嵌套至逆转录多核苷酸并导致对正确RNA靶标的反应特异性增加。在其他实验中,C7突出端和样品条码已存在于正向P7引物上。

[0629] 使用与免疫球蛋白重链或轻链的恒定区段互补的引物以及与在cDNA片段3'端的模板转换区互补的引物使纯化的逆转录产物进行第一轮PCR。

[0630] 将全部20 $\mu$ l纯化的逆转录产物置于50 $\mu$ l PCR反应中,该50 $\mu$ l PCR反应含有1X Q5缓冲液(EB),dATP、dGTP、dCTP、dTTP各200 $\mu$ M,重链/轻链恒定引物(IGHC、IGKC、IGLC)各65nM,40nM长模板转换引物,800nM短模板转换引物和0.02单位/ $\mu$ l Q5热启动聚合酶(NEB)。反应在98℃下进行1分钟,随后进行12个如下的循环:98℃下10sec;64℃下30sec;72℃下15sec。通过AMPure XP纯化产物,并将产物在25 $\mu$ l H<sub>2</sub>O中洗脱。

#### [0631] PCR1产物的定量

[0632] 接下来通过SYBR green定量PCR(qPCR)对纯化的PCR1产物的等分试样进行定量。

将5 $\mu$ l纯化的PCR1产物置于25 $\mu$ l PCR反应中,该25 $\mu$ l PCR反应含有1X Q5缓冲液(EB),dATP、dGTP、dCTP、dTTP各200 $\mu$ M,400nM Illumina相容性正向引物(P5-C5),400nM Illumina相容性配对末端引物(P7-SBC-C7)和0.02单位/ $\mu$ l Q5热启动聚合酶(NEB)。反应在98 $^{\circ}$ C下进行1分钟,随后进行20个如下的循环:98 $^{\circ}$ C下10sec;72 $^{\circ}$ C下45sec。

#### [0633] 索引PCR2

[0634] 使用第二PCR阶段用以下引物扩增PCR1产物:在PCR1中使用的相同P5C5反向引物,以及由P7序列和样品条码(SBC)组成且具有第二突出端(C7)的正向引物,以在Illumina测序平台上簇集。实验中处理的每个样品的样品条码是不同的,从而使得可在一次测序运行中将多个样品合并在一起。由于在PCR1反应中使用了多重引物库,因此PCR1可能引入偏差。通过限制PCR1循环的次数并在PCR2时普遍扩增,限制了所引入的偏差。PCR2还加载了样品条码和簇集标签以供测序。

[0635] 然后在PCR中扩增剩余的PCR1产物,以将全部Illumina衔接子序列添加至文库,包括用于合并测序的样品特异性索引。基于qPCR结果,选择理想的PCR循环次数,以防止PCR进入平台期,在该时间点可能产生不期望的PCR产品。

[0636] 对于索引PCR,将10 $\mu$ l纯化的PCR1产物置于50 $\mu$ l PCR反应中,该50 $\mu$ l PCR反应含有1X Q5缓冲液(EB),dATP、dGTP、dCTP、dTTP各200 $\mu$ M,0.25X SYBR green I(Invitrogen),400nM Illumina相容性正向引物(P5-C5),400nM Illumina相容性配对末端引物(P7-SBC-C7)和0.02单位/ $\mu$ L Q5热启动聚合酶(NEB)。反应在98 $^{\circ}$ C下进行1分钟,随后进行如下的循环:98 $^{\circ}$ C下10sec;72 $^{\circ}$ C下45sec,并且基于先前qPCR的结果决定循环次数。使用AMPure XP珠纯化产物,将产物在25 $\mu$ l TE缓冲液中洗脱,并在高通量Illumina测序和分析之前通过凝胶电泳进行可视化。

#### [0637] 最终文库

[0638] 得到的文库由具有合适的标签的完整抗体序列和经测序的簇集区段组成。对于每个起始独特RNA分子均存在生成的相同UID的许多拷贝。该UID位于与实施例1中描述的位置相比不同的位置。一经测序,即匹配相同的UID,并将测序读取分解成共有序列,由此消除测序和PCR错误。对于读取-1(C、J、D、V)从P5位点进行测序,随后对于读取-2(UID和VDJ)从P7位点进行测序,最后对于样品条码(SBC)的索引读取-3从反向P7位点进行测序。

#### [0639] 实施例27-直接从肿瘤样品对TIL进行测序

[0640] 制备包含400,000个卵巢肿瘤解离细胞的未分离TIL的卵巢肿瘤样品(即样品包含正常上皮细胞、癌细胞和TIL),用于进行如上所述的基于乳液的大规模高通量单个细胞多核苷酸测序。在不根据细胞类型进行事先分离的情况下,对来自样品中的B细胞和T细胞的编码Ig和TCR的多核苷酸进行测序。通过在细胞缓冲液:1x Dulbecco磷酸盐缓冲盐水(PBS)中以200g离心10min(两次)来对细胞进行洗涤。然后在细胞缓冲液中将细胞稀释到3.5 $\times 10^6$ 个细胞/mL的细胞浓度。然后将悬浮液移液通过20 $\mu$ m细胞过滤器。

[0641] 然后如上所述制备用于进行基于乳液的大规模高通量单个细胞多核苷酸测序的乳液反应混合物。一旦制备了细胞和反应混合物,就形成乳液。使用100 $\mu$ L Hamilton微升注射器将100 $\mu$ L PEEK样品环超负荷加载到两个各为约100 $\mu$ L的反应混合物的注射液中。使用100 $\mu$ L Hamilton气密注射器将约110 $\mu$ L的细胞悬浮液加载到约100 $\mu$ L、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秒一次不断地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)。采用12cm、内径0.5mm的PEEK管,通过滴入在冷模块中保持在大约0℃下的聚丙烯PCR管收集乳液芯片输出物。收集四个级分,每个含有在乳液中的50μL水性物质 (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:

[0642] 1. 42.0℃,30:00 (逆转录)

[0643] 2. 95.0℃,05:00 (使逆转录酶和DNA模板变性)

[0644] 3. 95.0℃,00:10

[0645] 4. 65.0℃,00:30

[0646] 5. 72.0℃,00:30

[0647] 6. 去往3,总计55次循环 (扩增容器条码并与cDNA融合)

[0648] 7. 4.0℃,无时间限制

[0649] 将乳液在4℃下保持过夜。然后将乳液破开。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。将管的内容物温和地离心并在热循环仪中采用下列程序 (分钟:秒) 进行温育:

[0650] 1. 50℃,持续15:00 (蛋白酶消化)

[0651] 2. 70℃,持续10:00 (蛋白酶灭活)

[0652] 3. 95℃,持续03:00 (蛋白酶灭活和DNA变性)

[0653] 4. 4.0℃,永远

[0654] 对管进行离心,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。在不干扰界面的情况下将上层水相转移到新管中。然后将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并加热至95℃3min进行洗脱。添加5体积的Qiagen缓冲液PB并加至硅胶柱上。然后用0.7mL的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中进行洗脱。然后通过三轮PCR扩增多核苷酸。用1.2体积的AMPure纯化最终的PCR产物,并在20μL的稀释缓冲液中进行洗脱。然后使用下一代测序技术平台对文库进行测序。

[0655] 如图18A-C中所示,当对B细胞和T细胞同时测序时,对6,056个IgH-IgL受体链组合和5,217个TCRα-TCRβ受体链组合进行测序,并具有准确的B细胞和T细胞受体配对,而没有串扰或污染 (图18A)。在6,056个IgH-IgL受体链组合中,5,152个是IgG同种型,从而表明肿瘤显示活化IgG浸润物的显著富集 (图18B)。肿瘤显示了衍生出TIL的重度突变浸润物的显著富集 (图18C)。

<110> 阿布维特罗有限责任公司

<120> 对治疗剂和疾病特异性抗原进行测序、确定、配对和验证的方法

<130> 44243-708.601

<140> PCT/US2016/029556

<141> 2016-04-27

<150> 62/153,041

<151> 2015-04-27

<160> 44

<170> PatentIn 版本 3.5

<210> 1

<211> 6

<212> PRT

<213> 人工序列

[0001]

<220>

<223> 人工序列的描述：合成 6xHis 标签

<400> 1

His His His His His His

1 5

<210> 2

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<220>

<221> 修饰的碱基

<222> (27)..(27)

<223> a、c、t、g、未知或其他

|        |                                                                                                    |    |
|--------|----------------------------------------------------------------------------------------------------|----|
|        | <p>&lt;400&gt; 2<br/>tttttttttt tttttttttt tttttvn</p>                                             | 27 |
|        | <p>&lt;210&gt; 3<br/>&lt;211&gt; 78<br/>&lt;212&gt; DNA<br/>&lt;213&gt; 人工序列</p>                   |    |
|        | <p>&lt;220&gt;<br/>&lt;223&gt; 人工序列的描述：合成寡核苷酸</p>                                                  |    |
|        | <p>&lt;220&gt;<br/>&lt;221&gt; 修饰的碱基<br/>&lt;222&gt; (30).. (33)<br/>&lt;223&gt; a、c、t、g、未知或其他</p> |    |
|        | <p>&lt;220&gt;<br/>&lt;221&gt; 修饰的碱基<br/>&lt;222&gt; (35).. (38)<br/>&lt;223&gt; a、c、t、g、未知或其他</p> |    |
| [0002] | <p>&lt;220&gt;<br/>&lt;221&gt; 修饰的碱基<br/>&lt;222&gt; (40).. (43)<br/>&lt;223&gt; a、c、t、g、未知或其他</p> |    |
|        | <p>&lt;400&gt; 3<br/>atccatccac gactgacgga cgtattaaan nnnwnnnnwn nnnagatcgg aagagcacac</p>         | 60 |
|        | <p>gtctgaactc cagtcacc</p>                                                                         | 78 |
|        | <p>&lt;210&gt; 4<br/>&lt;211&gt; 35<br/>&lt;212&gt; DNA<br/>&lt;213&gt; 人工序列</p>                   |    |
|        | <p>&lt;220&gt;<br/>&lt;223&gt; 人工序列的描述：合成寡核苷酸</p>                                                  |    |
|        | <p>&lt;220&gt;<br/>&lt;223&gt; 组合 DNA/RNA 分子的描述：合成寡核苷酸</p>                                         |    |

|        |                                                                                                                                                                                                                                                                                                                                                                          |    |
|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
|        | <p>&lt;220&gt;<br/>&lt;221&gt; 修饰的碱基<br/>&lt;222&gt; (24).. (25)<br/>&lt;223&gt; a、c、t、g、未知或其他</p> <p>&lt;220&gt;<br/>&lt;221&gt; 修饰的碱基<br/>&lt;222&gt; (27).. (28)<br/>&lt;223&gt; a、c、t、g、未知或其他</p> <p>&lt;220&gt;<br/>&lt;221&gt; 修饰的碱基<br/>&lt;222&gt; (30).. (31)<br/>&lt;223&gt; a、c、t、g、未知或其他</p> <p>&lt;400&gt; 4<br/>aatacgtccg tcagtcgtgg atgnntnnan ntggg</p> | 35 |
| [0003] | <p>&lt;210&gt; 5<br/>&lt;211&gt; 23<br/>&lt;212&gt; DNA<br/>&lt;213&gt; 人工序列</p> <p>&lt;220&gt;<br/>&lt;223&gt; 人工序列的描述：合成寡核苷酸</p> <p>&lt;400&gt; 5<br/>catccacgac tgacggacgt att</p>                                                                                                                                                                                    | 23 |
|        | <p>&lt;210&gt; 6<br/>&lt;211&gt; 24<br/>&lt;212&gt; DNA<br/>&lt;213&gt; 人工序列</p> <p>&lt;220&gt;<br/>&lt;223&gt; 人工序列的描述：合成寡核苷酸</p> <p>&lt;400&gt; 6<br/>gtgactggag ttcagacgtg tgct</p>                                                                                                                                                                                   | 24 |
|        | <p>&lt;210&gt; 7</p>                                                                                                                                                                                                                                                                                                                                                     |    |

|                     |    |
|---------------------|----|
| <211> 18            |    |
| <212> DNA           |    |
| <213> 人工序列          |    |
| <220>               |    |
| <223> 人工序列的描述: 合成引物 |    |
| <400> 7             |    |
| gggttggggc ggatgcac | 18 |
| <210> 8             |    |
| <211> 18            |    |
| <212> DNA           |    |
| <213> 人工序列          |    |
| <220>               |    |
| <223> 人工序列的描述: 合成引物 |    |
| <400> 8             |    |
| catccggagc cttgttg  | 18 |
| [0004]              |    |
| <210> 9             |    |
| <211> 18            |    |
| <212> DNA           |    |
| <213> 人工序列          |    |
| <220>               |    |
| <223> 人工序列的描述: 合成引物 |    |
| <400> 9             |    |
| ccttggggct ggtcgggg | 18 |
| <210> 10            |    |
| <211> 18            |    |
| <212> DNA           |    |
| <213> 人工序列          |    |
| <220>               |    |
| <223> 人工序列的描述: 合成引物 |    |
| <400> 10            |    |
| cggatgggct ctgtgtgg | 18 |

|        |                              |    |
|--------|------------------------------|----|
|        | <210> 11                     |    |
|        | <211> 18                     |    |
|        | <212> DNA                    |    |
|        | <213> 人工序列                   |    |
|        | <220>                        |    |
|        | <223> 人工序列的描述: 合成引物          |    |
|        | <400> 11                     |    |
|        | ccgatgggcc cttggtgg          | 18 |
|        | <210> 12                     |    |
|        | <211> 26                     |    |
|        | <212> DNA                    |    |
|        | <213> 人工序列                   |    |
|        | <220>                        |    |
|        | <223> 人工序列的描述: 合成引物          |    |
|        | <400> 12                     |    |
| [0005] | tttgatctcc accttggtcc ctccgc | 26 |
|        | <210> 13                     |    |
|        | <211> 25                     |    |
|        | <212> DNA                    |    |
|        | <213> 人工序列                   |    |
|        | <220>                        |    |
|        | <223> 人工序列的描述: 合成引物          |    |
|        | <400> 13                     |    |
|        | tttgatctcc agcttggtcc cctgg  | 25 |
|        | <210> 14                     |    |
|        | <211> 26                     |    |
|        | <212> DNA                    |    |
|        | <213> 人工序列                   |    |
|        | <220>                        |    |
|        | <223> 人工序列的描述: 合成引物          |    |

|        |                                                                                   |    |
|--------|-----------------------------------------------------------------------------------|----|
|        | <p>&lt;400&gt; 14<br/>tttgatatcc actttggtcc cagggc</p>                            | 26 |
|        | <p>&lt;210&gt; 15<br/>&lt;211&gt; 26<br/>&lt;212&gt; DNA<br/>&lt;213&gt; 人工序列</p> |    |
|        | <p>&lt;220&gt;<br/>&lt;223&gt; 人工序列的描述: 合成引物</p>                                  |    |
|        | <p>&lt;400&gt; 15<br/>tttgatttcc accttggtcc cttggc</p>                            | 26 |
|        | <p>&lt;210&gt; 16<br/>&lt;211&gt; 26<br/>&lt;212&gt; DNA<br/>&lt;213&gt; 人工序列</p> |    |
| [0006] | <p>&lt;220&gt;<br/>&lt;223&gt; 人工序列的描述: 合成引物</p>                                  |    |
|        | <p>&lt;400&gt; 16<br/>tttaatctcc agtcgtgtcc cttggc</p>                            | 26 |
|        | <p>&lt;210&gt; 17<br/>&lt;211&gt; 22<br/>&lt;212&gt; DNA<br/>&lt;213&gt; 人工序列</p> |    |
|        | <p>&lt;220&gt;<br/>&lt;223&gt; 人工序列的描述: 合成引物</p>                                  |    |
|        | <p>&lt;400&gt; 17<br/>gaggacggtc accttggtgc ca</p>                                | 22 |
|        | <p>&lt;210&gt; 18<br/>&lt;211&gt; 24<br/>&lt;212&gt; DNA<br/>&lt;213&gt; 人工序列</p> |    |

|        |                             |  |    |
|--------|-----------------------------|--|----|
| [0007] | <220>                       |  |    |
|        | <223> 人工序列的描述: 合成引物         |  |    |
|        | <400> 18                    |  |    |
|        | taggacggtc agcttggtcc ctcc  |  | 24 |
|        | <210> 19                    |  |    |
|        | <211> 21                    |  |    |
|        | <212> DNA                   |  |    |
|        | <213> 人工序列                  |  |    |
|        | <220>                       |  |    |
|        | <223> 人工序列的描述: 合成引物         |  |    |
|        | <400> 19                    |  |    |
|        | gaggacggtc agctgggtgc c     |  | 21 |
|        | <210> 20                    |  |    |
|        | <211> 26                    |  |    |
|        | <212> DNA                   |  |    |
|        | <213> 人工序列                  |  |    |
|        | <220>                       |  |    |
|        | <223> 人工序列的描述: 合成引物         |  |    |
|        | <400> 20                    |  |    |
|        | taaaatgac agctgggttc ctccac |  | 26 |
|        | <210> 21                    |  |    |
|        | <211> 24                    |  |    |
|        | <212> DNA                   |  |    |
|        | <213> 人工序列                  |  |    |
|        | <220>                       |  |    |
|        | <223> 人工序列的描述: 合成引物         |  |    |
|        | <400> 21                    |  |    |
|        | taggacggtg accttggtcc cagt  |  | 24 |
|        | <210> 22                    |  |    |
|        |                             |  |    |

|                                                          |    |
|----------------------------------------------------------|----|
| <211> 34                                                 |    |
| <212> DNA                                                |    |
| <213> 人工序列                                               |    |
| <220>                                                    |    |
| <223> 人工序列的描述: 合成引物                                      |    |
| <400> 22                                                 |    |
| gtgactggag ttcagacgtg tgcctctccg atct                    | 34 |
| <210> 23                                                 |    |
| <211> 51                                                 |    |
| <212> DNA                                                |    |
| <213> 人工序列                                               |    |
| <220>                                                    |    |
| <223> 人工序列的描述: 合成引物                                      |    |
| <400> 23                                                 |    |
| acactctttc cctacacgac gctcttccga tctgggttg ggcggatgca c  | 51 |
| [0008]                                                   |    |
| <210> 24                                                 |    |
| <211> 51                                                 |    |
| <212> DNA                                                |    |
| <213> 人工序列                                               |    |
| <220>                                                    |    |
| <223> 人工序列的描述: 合成引物                                      |    |
| <400> 24                                                 |    |
| acactctttc cctacacgac gctcttccga tctcatccgg agccttggtg g | 51 |
| <210> 25                                                 |    |
| <211> 51                                                 |    |
| <212> DNA                                                |    |
| <213> 人工序列                                               |    |
| <220>                                                    |    |
| <223> 人工序列的描述: 合成引物                                      |    |
| <400> 25                                                 |    |
| acactctttc cctacacgac gctcttccga tctccttggg gctggtcggg g | 51 |

|        |                                                                                                                                                                                                                                              |    |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
|        | <p>&lt;210&gt; 26</p> <p>&lt;211&gt; 51</p> <p>&lt;212&gt; DNA</p> <p>&lt;213&gt; 人工序列</p> <p>&lt;220&gt;</p> <p>&lt;223&gt; 人工序列的描述: 合成引物</p> <p>&lt;400&gt; 26</p> <p>acactctttc cctacacgac gctcttccga tctcggatgg gctctgtgtg g</p>         | 51 |
|        | <p>&lt;210&gt; 27</p> <p>&lt;211&gt; 51</p> <p>&lt;212&gt; DNA</p> <p>&lt;213&gt; 人工序列</p> <p>&lt;220&gt;</p> <p>&lt;223&gt; 人工序列的描述: 合成引物</p> <p>&lt;400&gt; 27</p> <p>acactctttc cctacacgac gctcttccga tctcggatgg gcccttggg g</p>          | 51 |
| [0009] | <p>&lt;210&gt; 28</p> <p>&lt;211&gt; 59</p> <p>&lt;212&gt; DNA</p> <p>&lt;213&gt; 人工序列</p> <p>&lt;220&gt;</p> <p>&lt;223&gt; 人工序列的描述: 合成引物</p> <p>&lt;400&gt; 28</p> <p>acactctttc cctacacgac gctcttccga tcttttgatc tccaccttgg tccctccgc</p> | 59 |
|        | <p>&lt;210&gt; 29</p> <p>&lt;211&gt; 58</p> <p>&lt;212&gt; DNA</p> <p>&lt;213&gt; 人工序列</p> <p>&lt;220&gt;</p> <p>&lt;223&gt; 人工序列的描述: 合成引物</p>                                                                                               |    |

|        |                                                                  |    |
|--------|------------------------------------------------------------------|----|
|        | <400> 29                                                         |    |
|        | acactctttc cctacacgac gctcttccga tcttttgatc tccagcttgg tcccctgg  | 58 |
|        | <210> 30                                                         |    |
|        | <211> 59                                                         |    |
|        | <212> DNA                                                        |    |
|        | <213> 人工序列                                                       |    |
|        | <220>                                                            |    |
|        | <223> 人工序列的描述: 合成引物                                              |    |
|        | <400> 30                                                         |    |
|        | acactctttc cctacacgac gctcttccga tcttttgata tccactttgg tcccagggc | 59 |
|        | <210> 31                                                         |    |
|        | <211> 59                                                         |    |
|        | <212> DNA                                                        |    |
|        | <213> 人工序列                                                       |    |
|        | <220>                                                            |    |
| [0010] | <223> 人工序列的描述: 合成引物                                              |    |
|        | <400> 31                                                         |    |
|        | acactctttc cctacacgac gctcttccga tcttttgatt tccaccttgg tcccttggc | 59 |
|        | <210> 32                                                         |    |
|        | <211> 59                                                         |    |
|        | <212> DNA                                                        |    |
|        | <213> 人工序列                                                       |    |
|        | <220>                                                            |    |
|        | <223> 人工序列的描述: 合成引物                                              |    |
|        | <400> 32                                                         |    |
|        | acactctttc cctacacgac gctcttccga tcttttaatc tccagtcgtg tcccttggc | 59 |
|        | <210> 33                                                         |    |
|        | <211> 55                                                         |    |
|        | <212> DNA                                                        |    |
|        | <213> 人工序列                                                       |    |

|        |                                                                  |    |
|--------|------------------------------------------------------------------|----|
|        | <220>                                                            |    |
|        | <223> 人工序列的描述: 合成引物                                              |    |
|        | <400> 33                                                         |    |
|        | acactctttc cctacacgac gctcttccga tctgaggacg gtcaccttgg tgcca     | 55 |
|        | <210> 34                                                         |    |
|        | <211> 57                                                         |    |
|        | <212> DNA                                                        |    |
|        | <213> 人工序列                                                       |    |
|        | <220>                                                            |    |
|        | <223> 人工序列的描述: 合成引物                                              |    |
|        | <400> 34                                                         |    |
|        | acactctttc cctacacgac gctcttccga tcttaggacg gtcagcttgg tccctcc   | 57 |
|        | <210> 35                                                         |    |
|        | <211> 54                                                         |    |
|        | <212> DNA                                                        |    |
| [0011] | <213> 人工序列                                                       |    |
|        | <220>                                                            |    |
|        | <223> 人工序列的描述: 合成引物                                              |    |
|        | <400> 35                                                         |    |
|        | acactctttc cctacacgac gctcttccga tctgaggacg gtcagctggg tgcc      | 54 |
|        | <210> 36                                                         |    |
|        | <211> 59                                                         |    |
|        | <212> DNA                                                        |    |
|        | <213> 人工序列                                                       |    |
|        | <220>                                                            |    |
|        | <223> 人工序列的描述: 合成引物                                              |    |
|        | <400> 36                                                         |    |
|        | acactctttc cctacacgac gctcttccga tcttaaaatg atcagctggg ttctctcac | 59 |
|        | <210> 37                                                         |    |

|                                                                   |    |
|-------------------------------------------------------------------|----|
| <211> 57                                                          |    |
| <212> DNA                                                         |    |
| <213> 人工序列                                                        |    |
| <220>                                                             |    |
| <223> 人工序列的描述: 合成引物                                               |    |
| <400> 37                                                          |    |
| acactctttc cctacacgac gctcttccga tcttaggacg gtgaccttgg tcccagt    | 57 |
| <210> 38                                                          |    |
| <211> 55                                                          |    |
| <212> DNA                                                         |    |
| <213> 人工序列                                                        |    |
| <220>                                                             |    |
| <223> 人工序列的描述: 合成引物                                               |    |
| <400> 38                                                          |    |
| acactctttc cctacacgac gctcttccga tcttaggacg gtcagctcgg tcccc      | 55 |
| [0012]                                                            |    |
| <210> 39                                                          |    |
| <211> 64                                                          |    |
| <212> DNA                                                         |    |
| <213> 人工序列                                                        |    |
| <220>                                                             |    |
| <223> 人工序列的描述: 合成引物                                               |    |
| <220>                                                             |    |
| <221> 修饰的碱基                                                       |    |
| <222> (25).. (30)                                                 |    |
| <223> a、c、t、g、未知或其他                                               |    |
| <400> 39                                                          |    |
| caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgctcttccg | 60 |
| atct                                                              | 64 |
| <210> 40                                                          |    |
| <211> 58                                                          |    |

|        |                                                                |    |
|--------|----------------------------------------------------------------|----|
|        | <212> DNA                                                      |    |
|        | <213> 人工序列                                                     |    |
|        | <220>                                                          |    |
|        | <223> 人工序列的描述: 合成引物                                            |    |
|        | <400> 40                                                       |    |
|        | aatgatacgg cgaccaccga gatctacact ctttccctac acgacgtct tccgatct | 58 |
|        | <210> 41                                                       |    |
|        | <211> 24                                                       |    |
|        | <212> DNA                                                      |    |
|        | <213> 人工序列                                                     |    |
|        | <220>                                                          |    |
|        | <223> 人工序列的描述: 合成引物                                            |    |
|        | <400> 41                                                       |    |
|        | caagcagaag acggcatacg agat                                     | 24 |
| [0013] | <210> 42                                                       |    |
|        | <211> 20                                                       |    |
|        | <212> DNA                                                      |    |
|        | <213> 人工序列                                                     |    |
|        | <220>                                                          |    |
|        | <223> 人工序列的描述: 合成寡核苷酸                                          |    |
|        | <220>                                                          |    |
|        | <221> 修饰的碱基                                                    |    |
|        | <222> (1).. (4)                                                |    |
|        | <223> a、c、t、g、未知或其他                                            |    |
|        | <220>                                                          |    |
|        | <221> 修饰的碱基                                                    |    |
|        | <222> (6).. (9)                                                |    |
|        | <223> a、c、t、g、未知或其他                                            |    |
|        | <220>                                                          |    |
|        | <221> 修饰的碱基                                                    |    |
|        | <222> (11).. (14)                                              |    |
|        | <223> a、c、t、g、未知或其他                                            |    |

|        |                      |    |
|--------|----------------------|----|
|        | <220>                |    |
|        | <221> 修饰的碱基          |    |
|        | <222> (16).. (19)    |    |
|        | <223> a、c、t、g、未知或其他  |    |
|        | <400> 42             |    |
|        | nnnnwnnnnw nnnwnnnnw | 20 |
|        | <210> 43             |    |
|        | <211> 18             |    |
|        | <212> DNA            |    |
|        | <213> 人工序列           |    |
|        | <220>                |    |
| [0014] | <223> 人工序列的描述：合成引物   |    |
|        | <400> 43             |    |
|        | tttttttttt tttttttt  | 18 |
|        | <210> 44             |    |
|        | <211> 10             |    |
|        | <212> DNA            |    |
|        | <213> 人工序列           |    |
|        | <220>                |    |
|        | <223> 人工序列的描述：合成寡核苷酸 |    |
|        | <400> 44             |    |
|        | aaaaaaaaaa           | 10 |

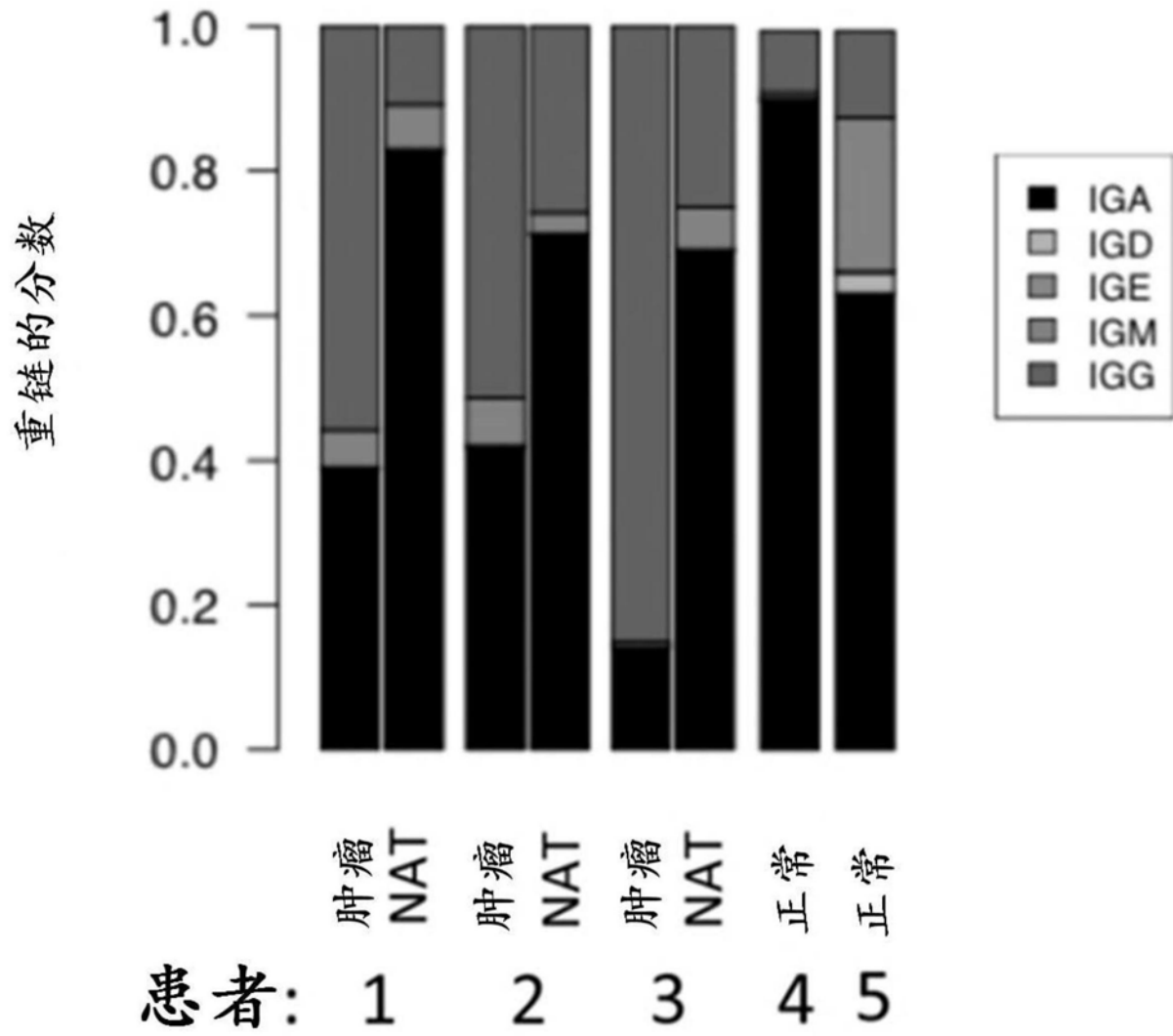


图1A

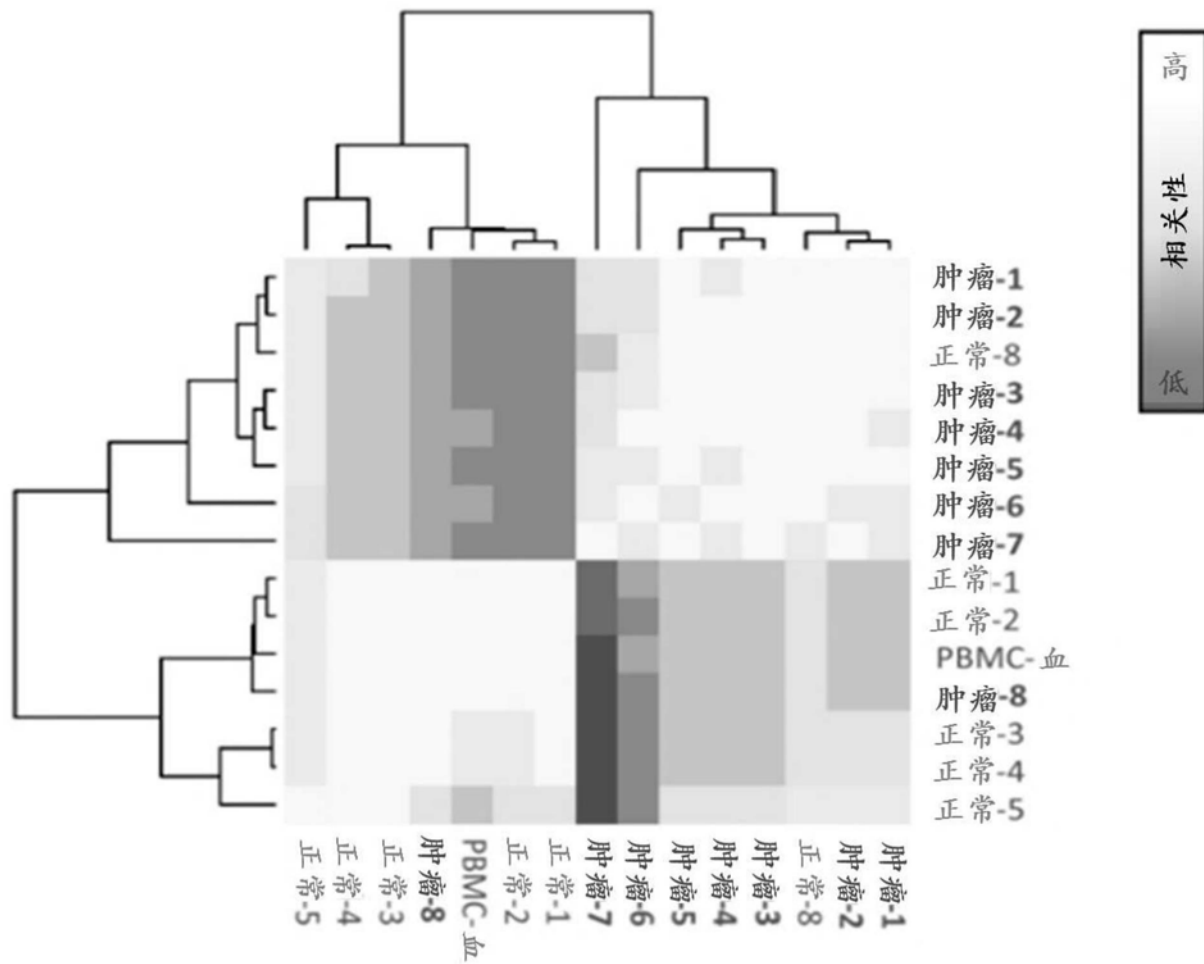


图1B

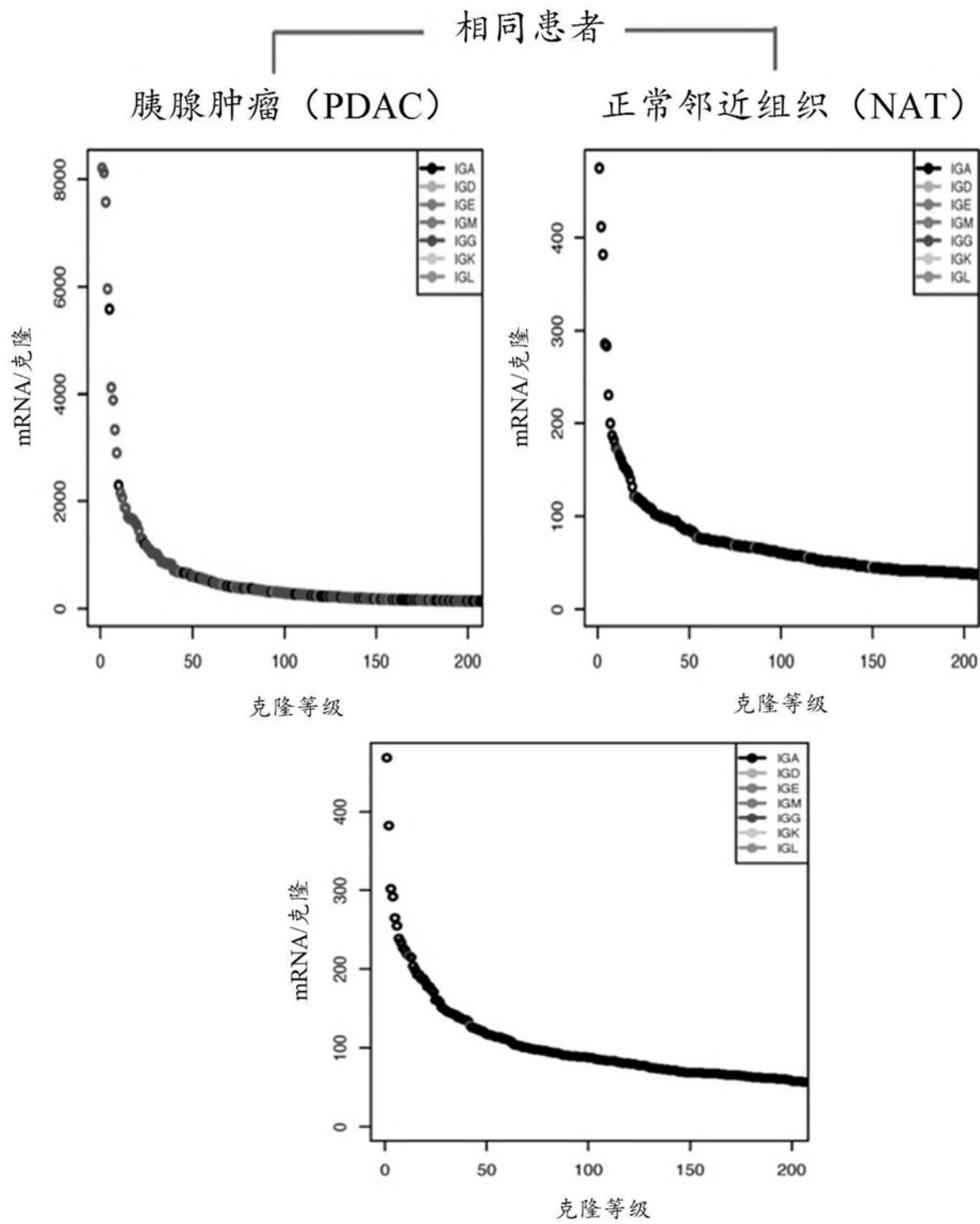


图2A

## MS脑B细胞组库-等级

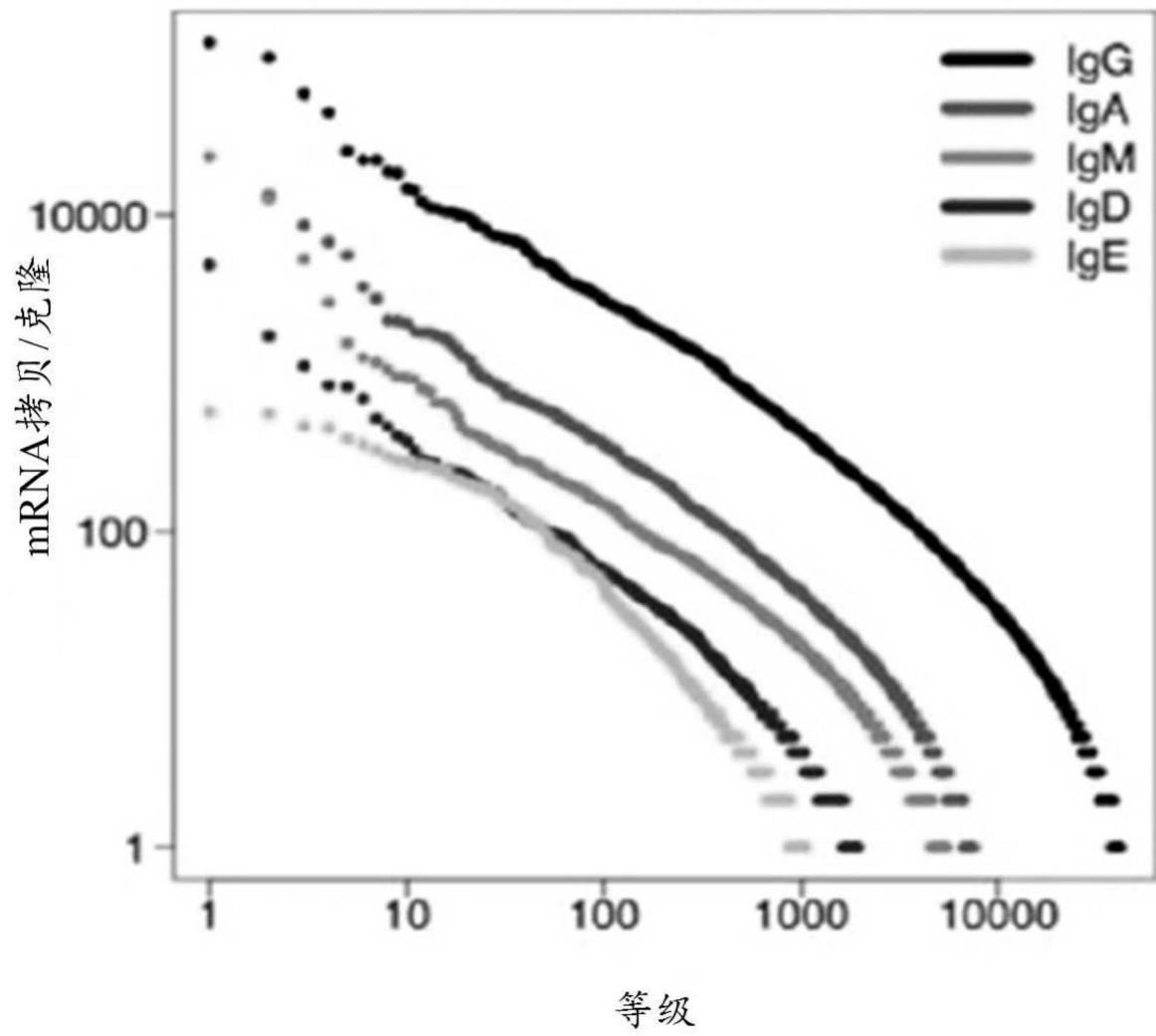


图2B

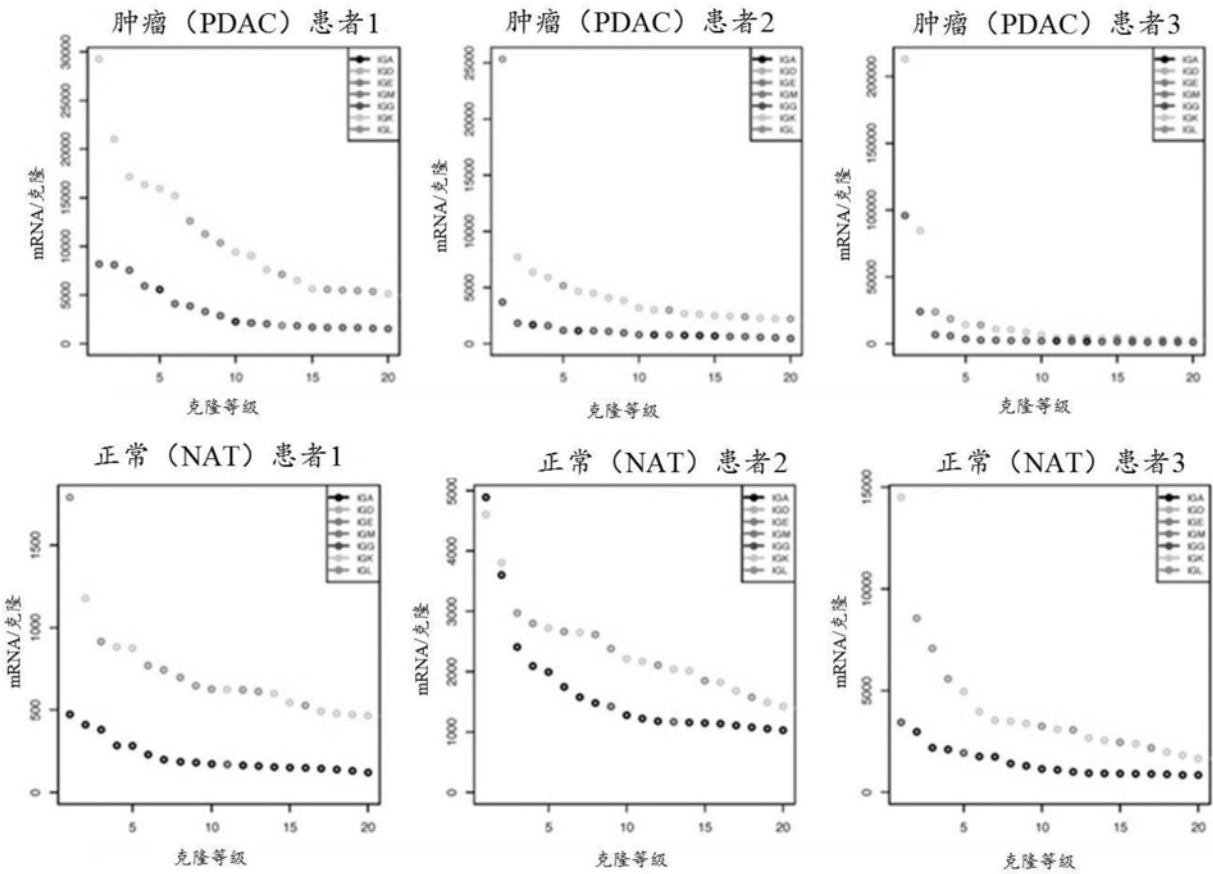


图3A

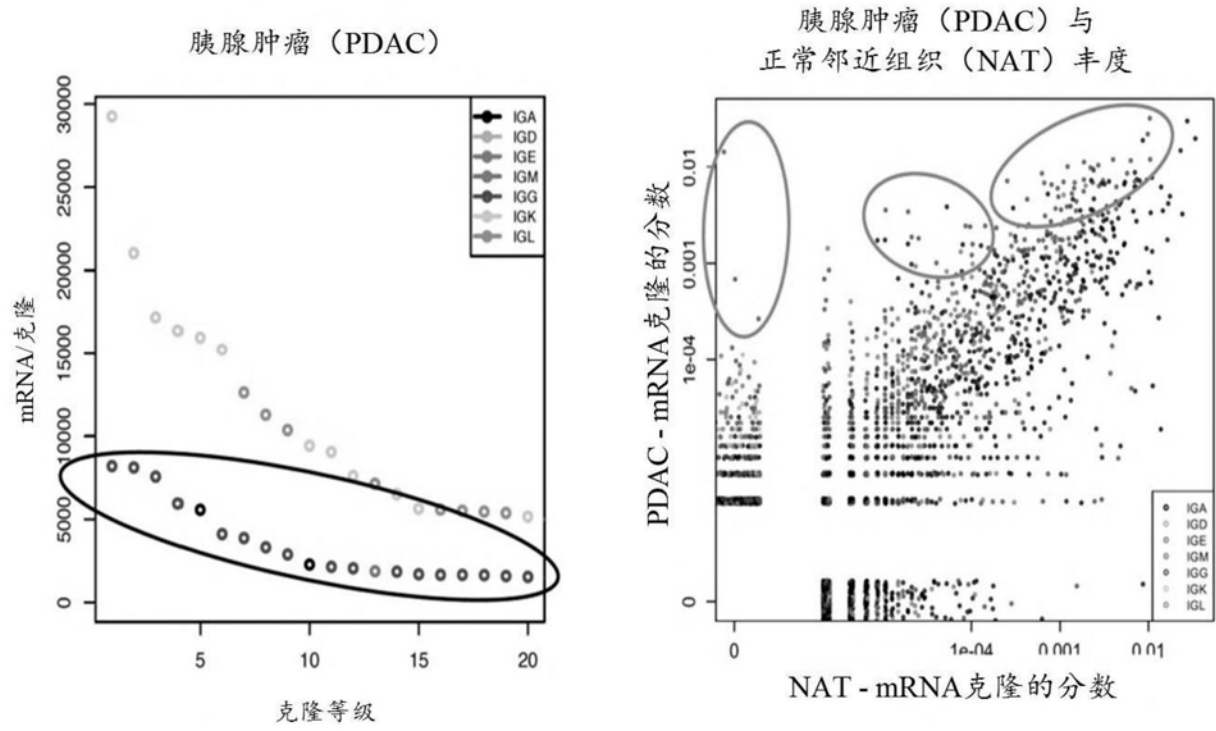


图3B

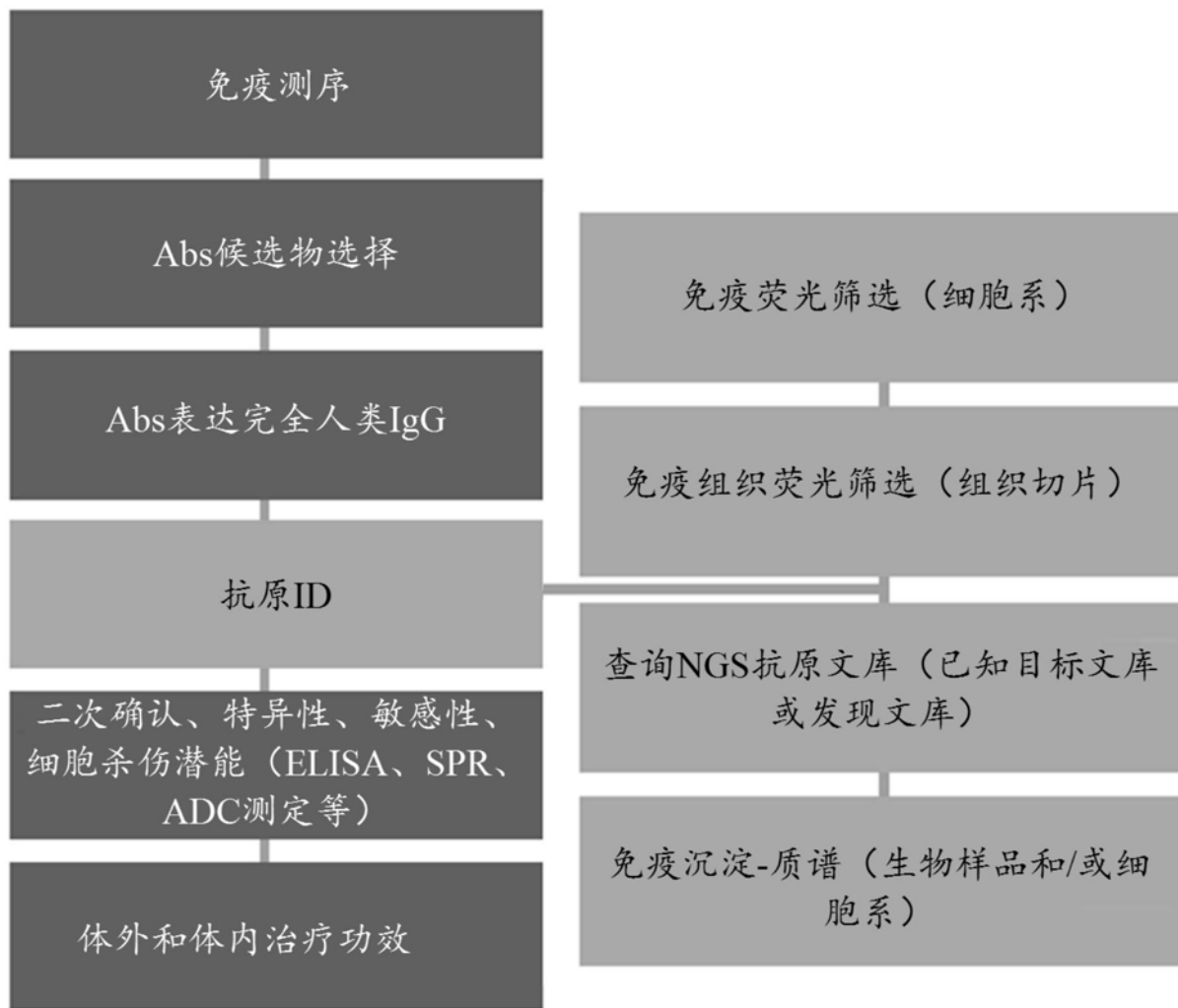


图4

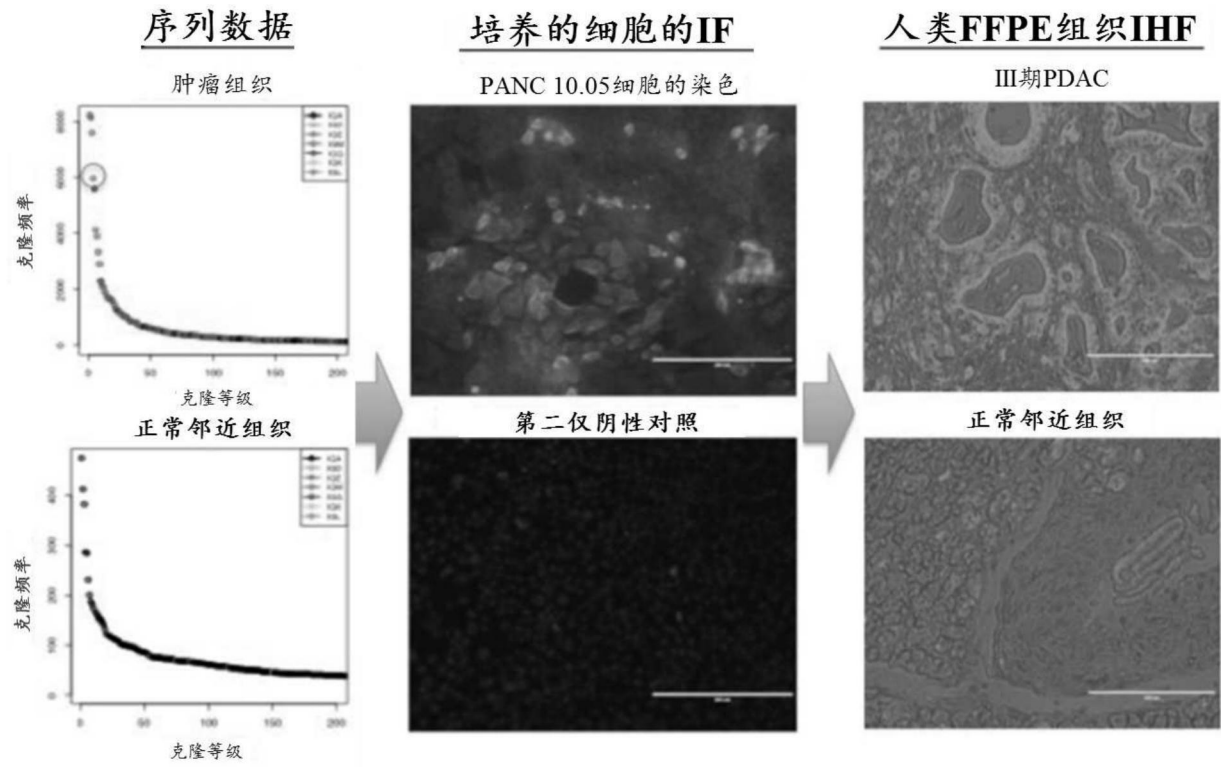


图5

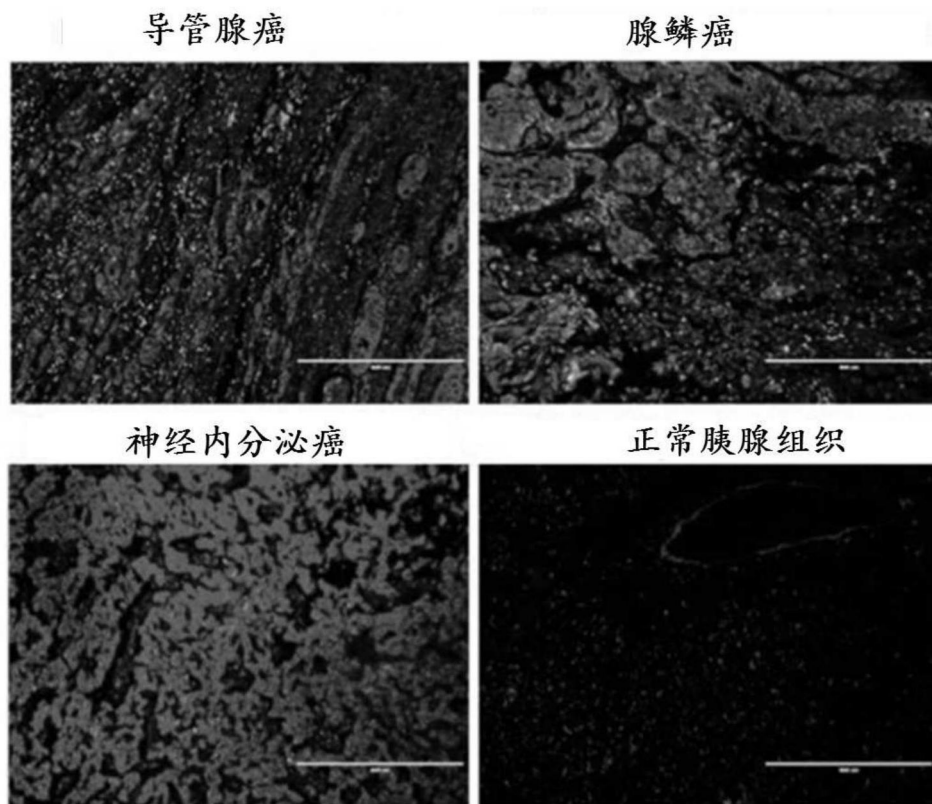


图6A

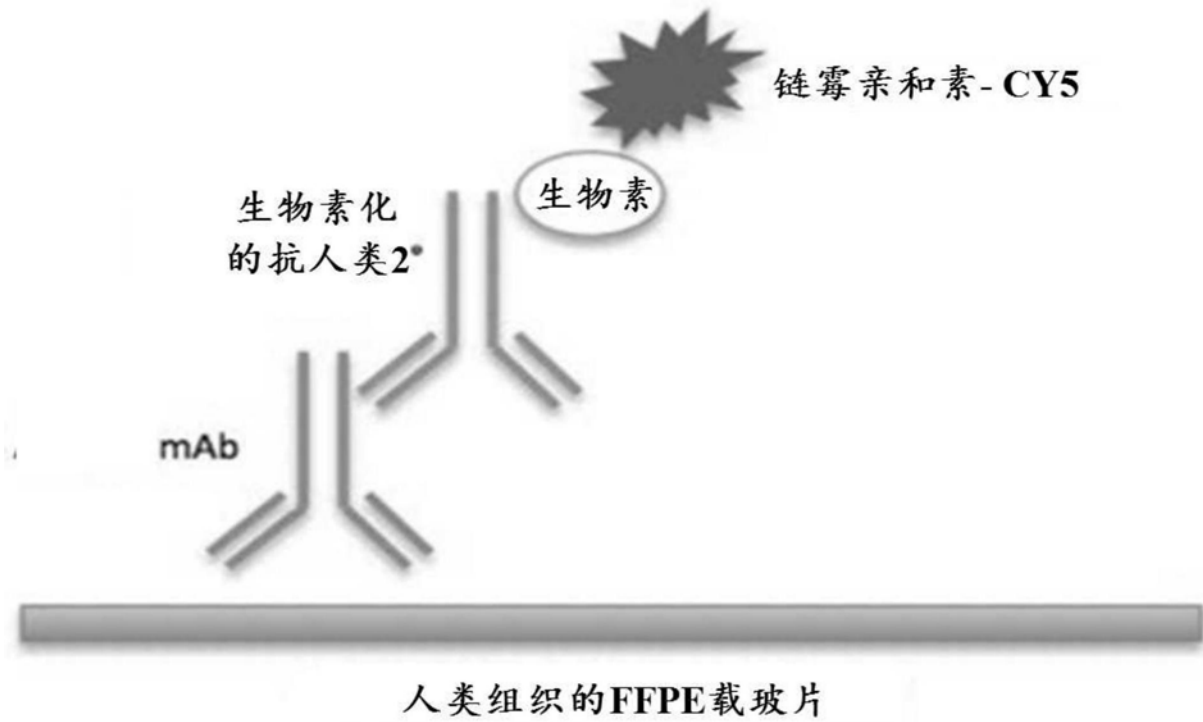
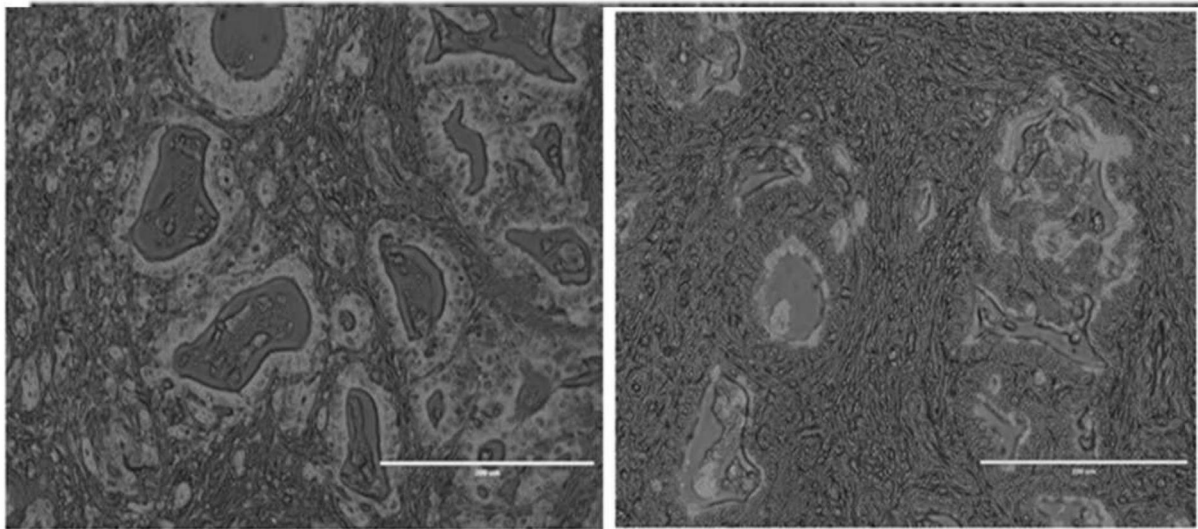


图6B

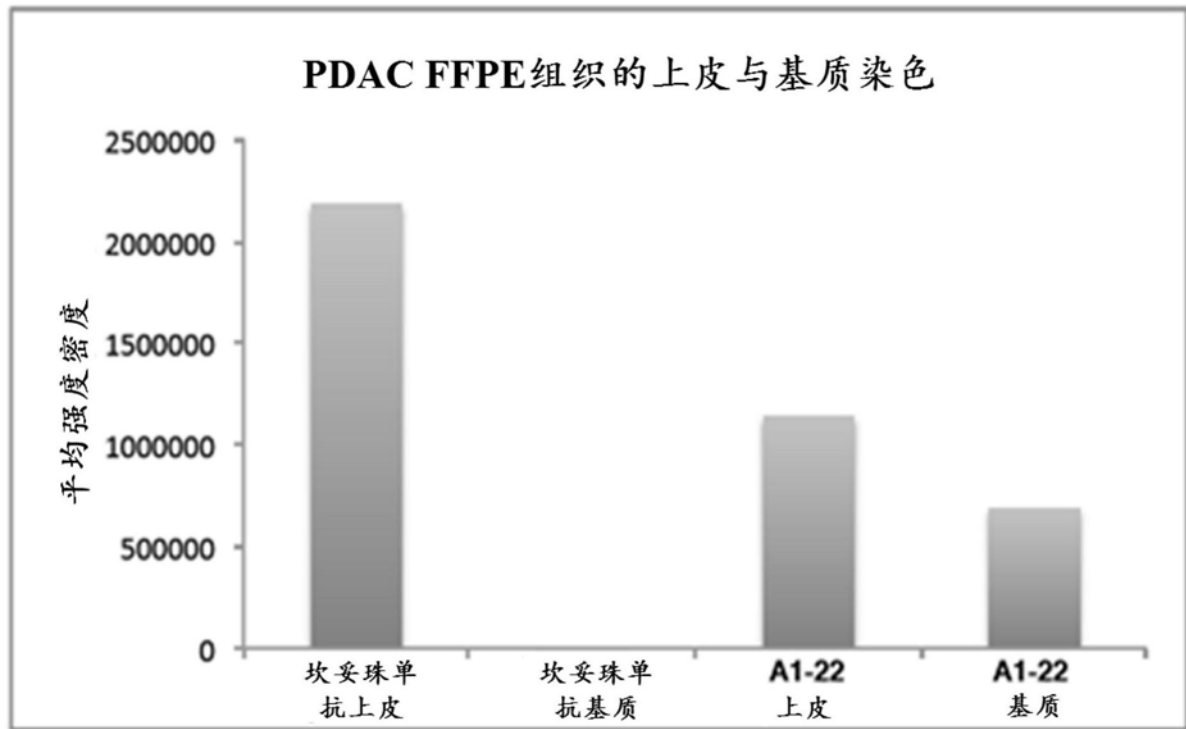
### III期PDAC组织



A1-22

坎妥珠单抗

图7A



\* 如通过ImageJ软件以减去背景的平均强度密度所测量的荧光

图7B

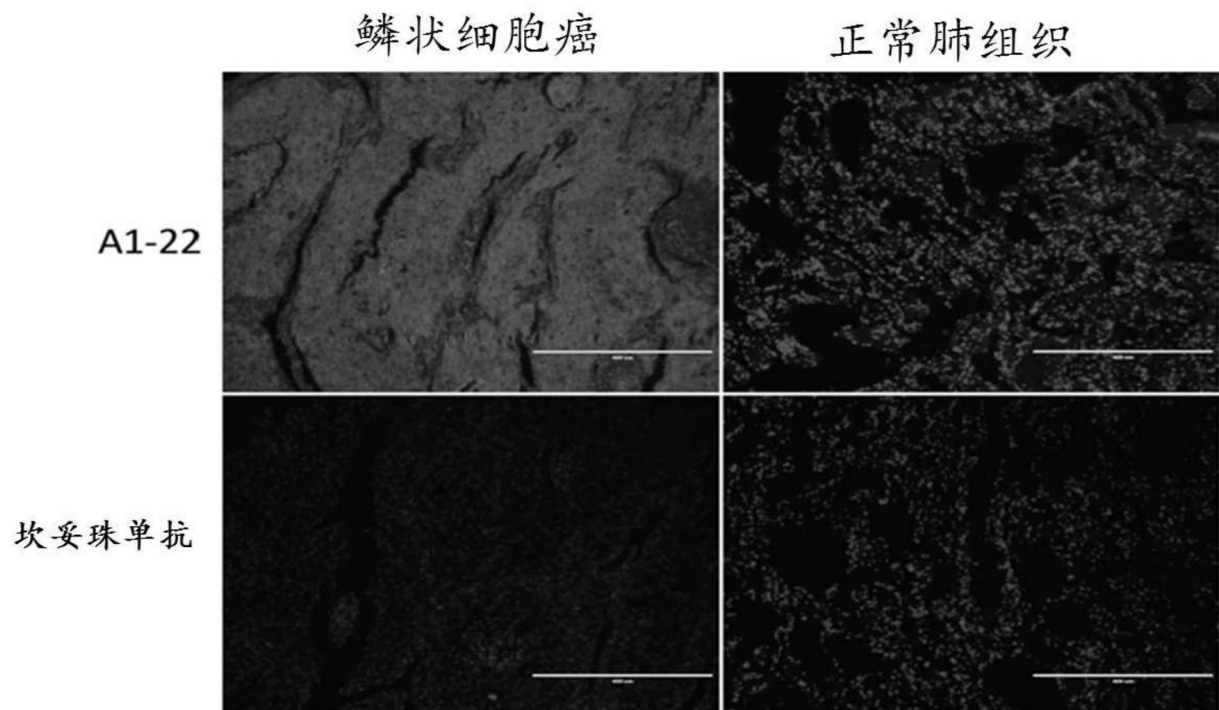
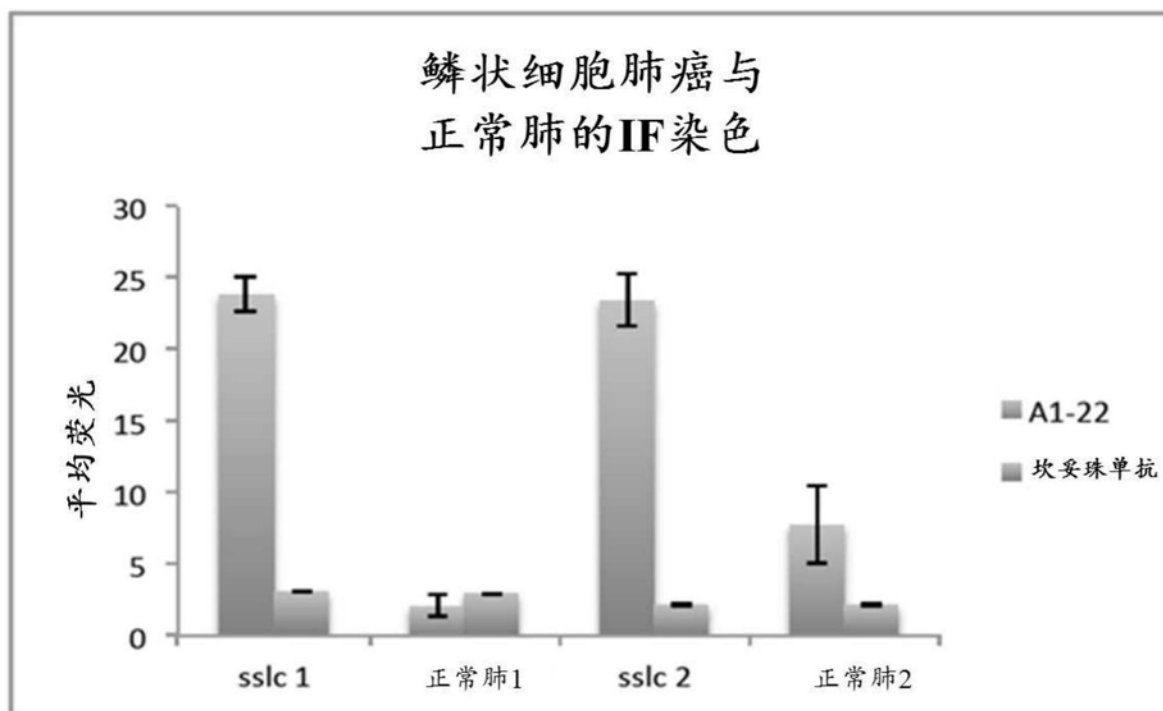


图8A



\* 通过ImageJ软件测量的荧光

图8B

### 正常组织的慢性胰腺炎mAb染色

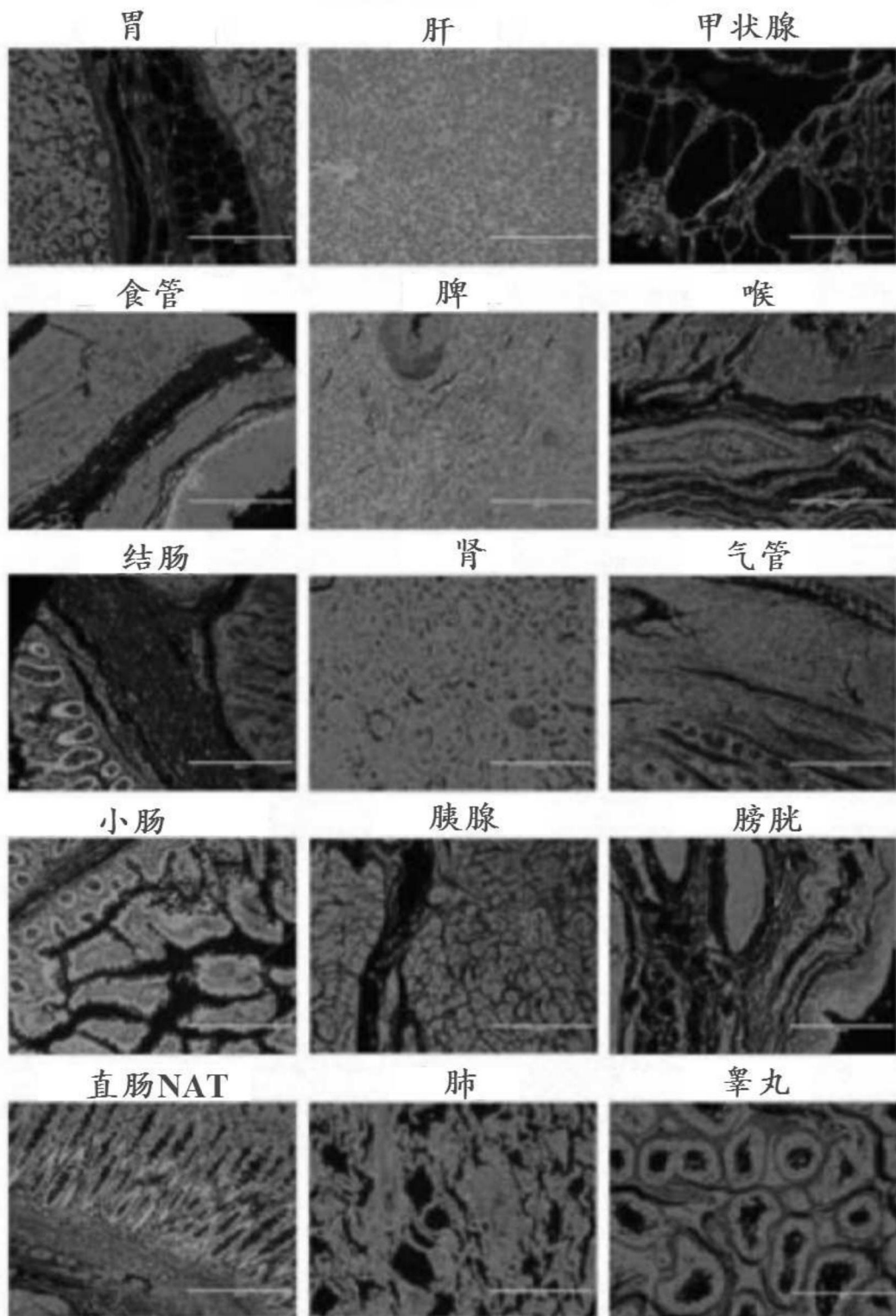


图9A

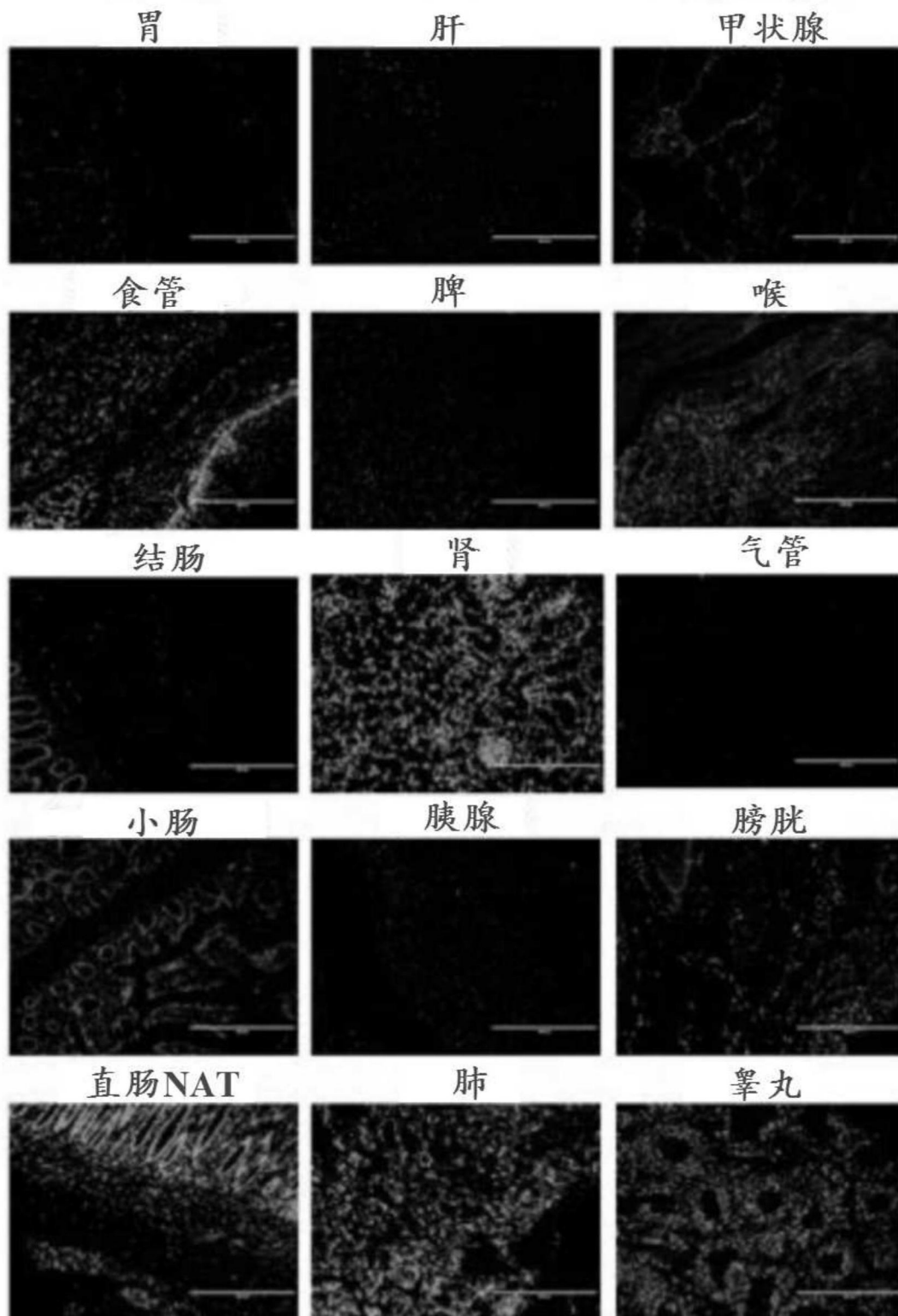
正常组织的PDAC mAb A1-108染色

图9B

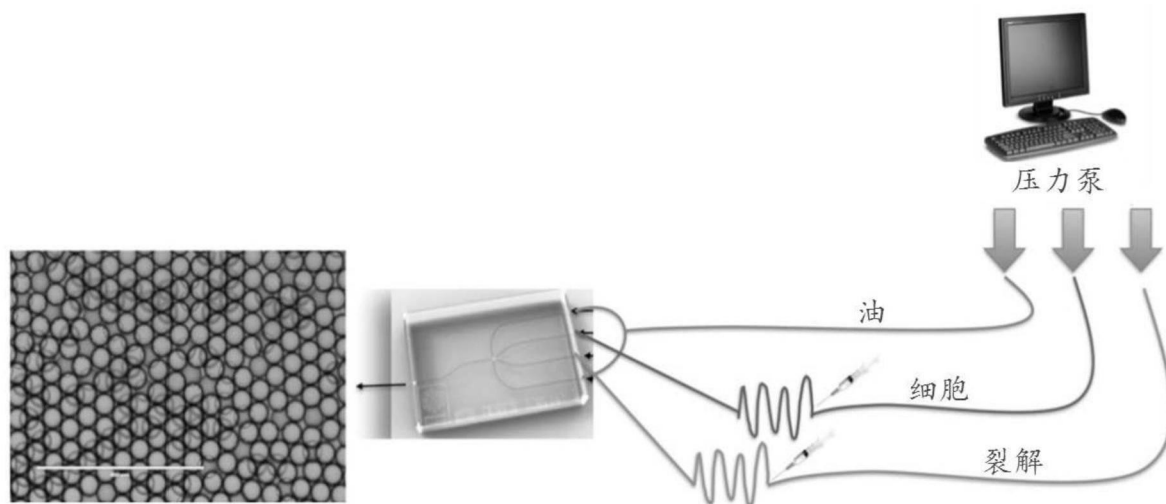


图10

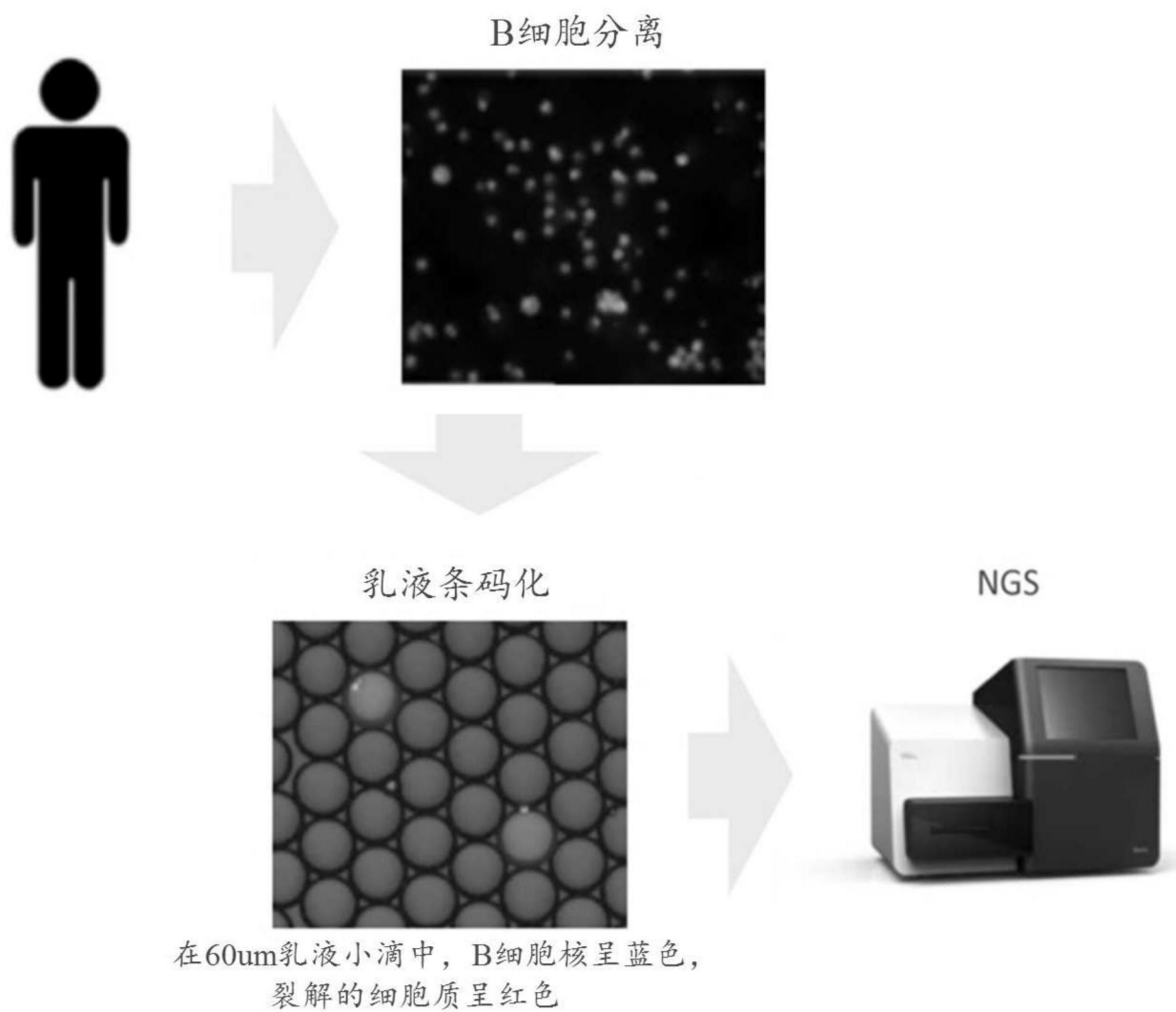


图11

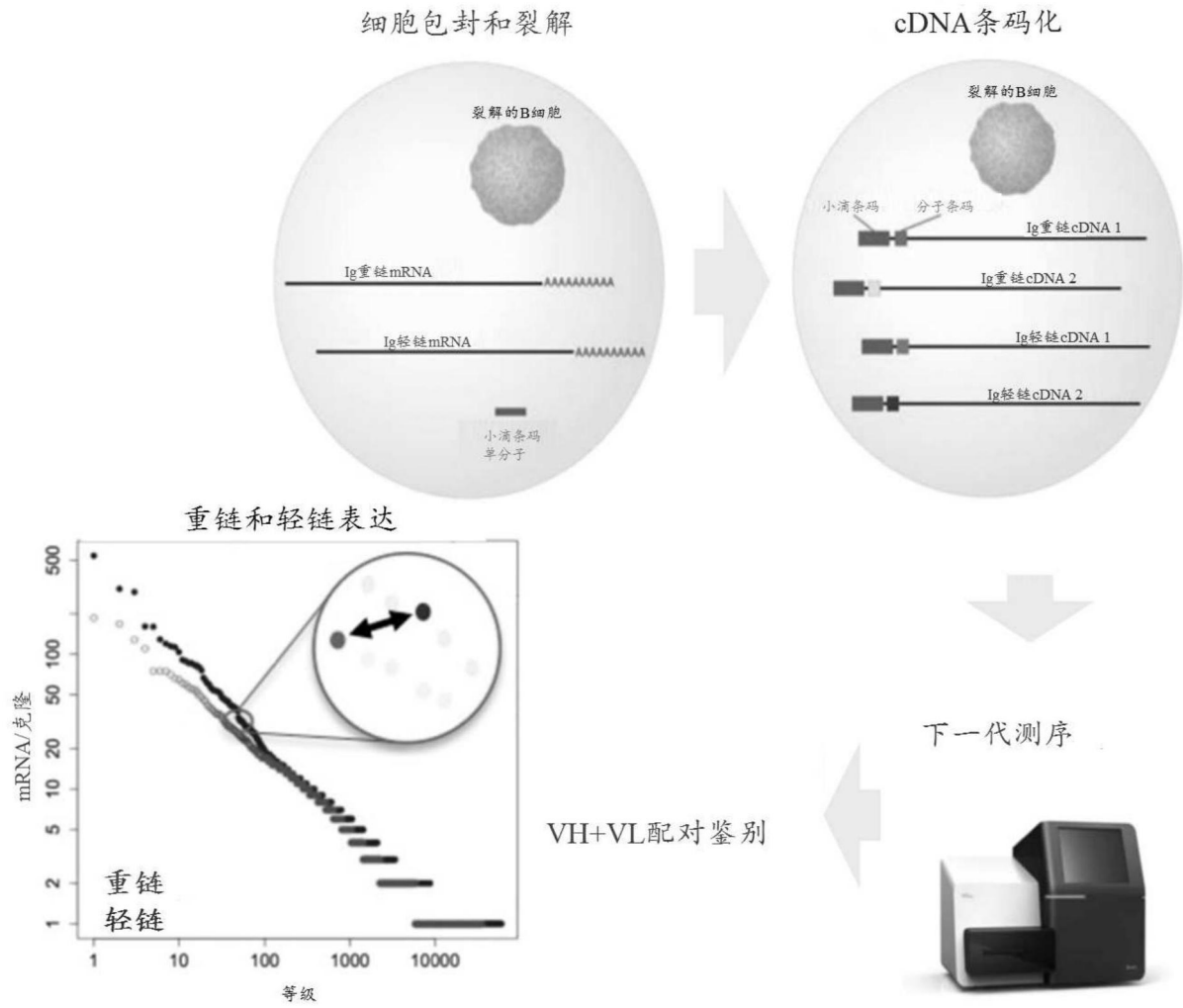


图12

## B细胞（幼稚、记忆和浆细胞）的配对

|                           |            |
|---------------------------|------------|
| 转化成乳液的B细胞                 | 500,000    |
| Illumina 2x300bp原始序列读取    | 13,670,330 |
| 正确条码化的VH/VL读取             | 11,594,330 |
| 观察到的不同的小滴条码               | 358,780    |
| 总配对（90%准确度 <sup>1</sup> ） | 197,865    |
| 高严格性配对（ $p=0.0006$ ）      | 72,030     |
| 总配对效率                     | 40%        |

<sup>1</sup>-通过细胞系掺加进行测量

图13

## 每500,000个B细胞的抗体链配对

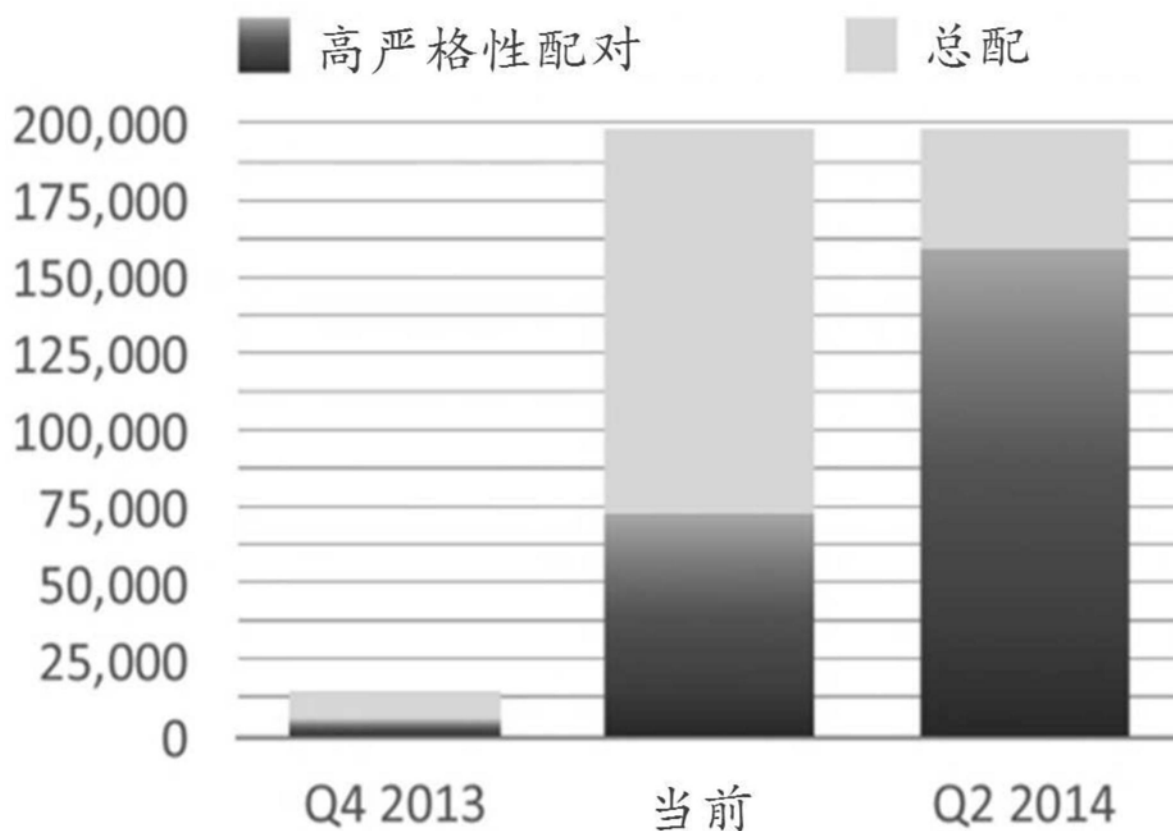


图14





|          |                   |                   |                    |
|----------|-------------------|-------------------|--------------------|
| 细胞分离区室   | 微流体乳液             | 125K孔微量滴定板        | 96孔板               |
| 典型的细胞输入  | 500,000<br>(1次运行) | 61,000<br>(4张载玻片) | 960个细胞<br>(96x10板) |
| 细胞类型     | 所有                | 所有                | 仅浆母细胞              |
| 总配对      | 200,000           | 2,716             | 500                |
| 高严格性配对   | 75,000            | 0                 | 500                |
| 检索到的抗体序列 | 全长，具有同种型          | 部分长度（不能表达）；无同种型   | 全长，仅IgG            |
| 抗原筛选能力   | 是                 | 否                 | 否                  |

图15

## T细胞受体的等级丰度表达

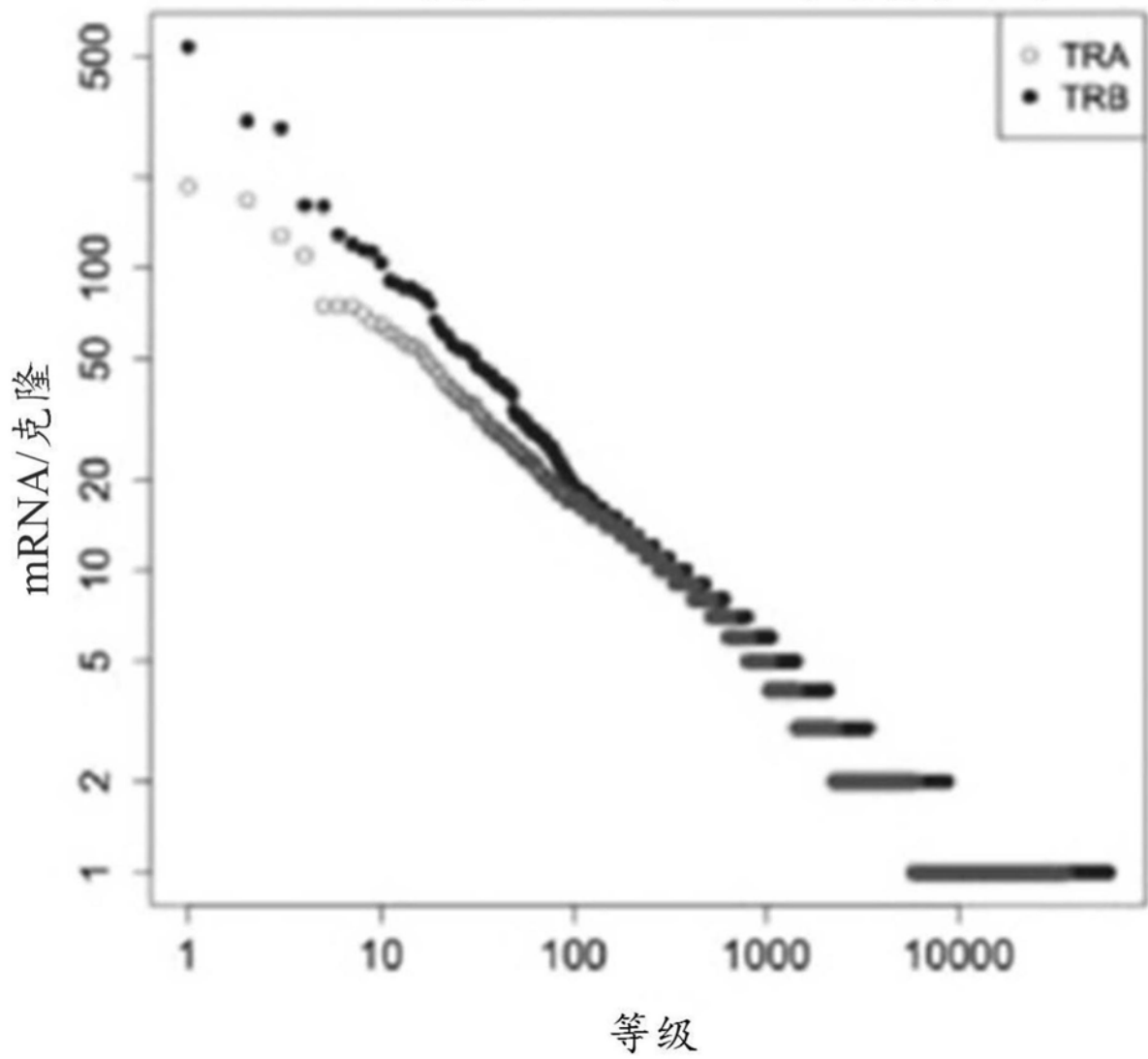


图16

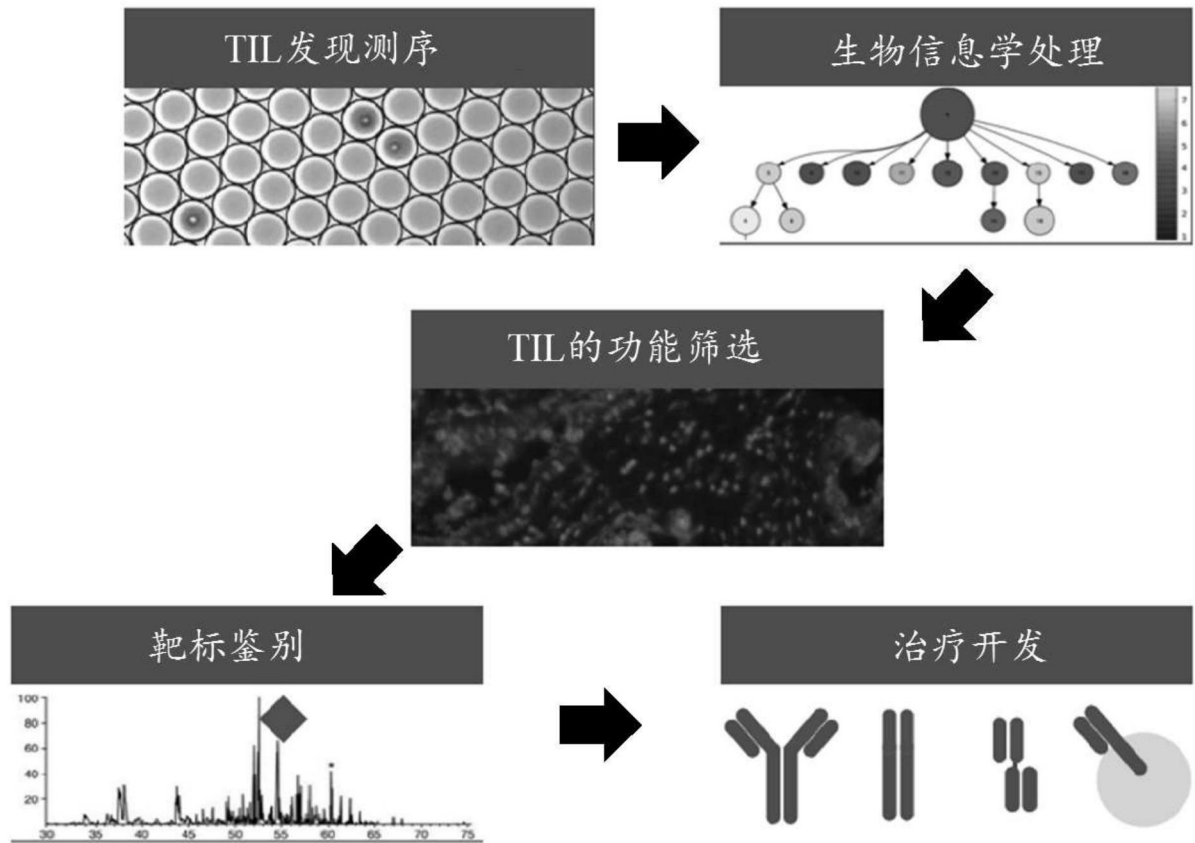


图17

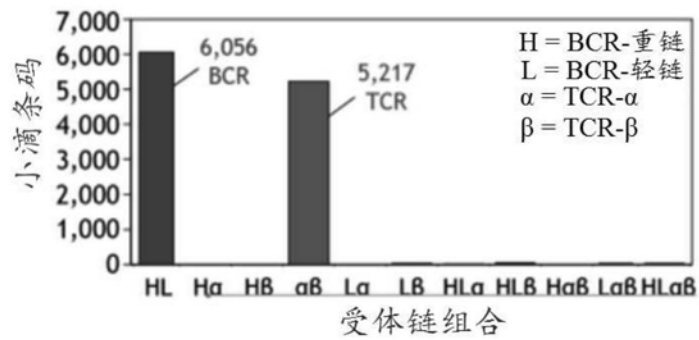


图18A

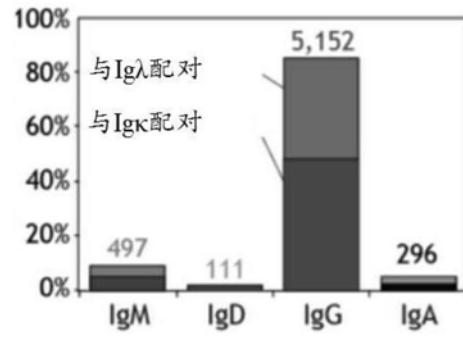


图18B

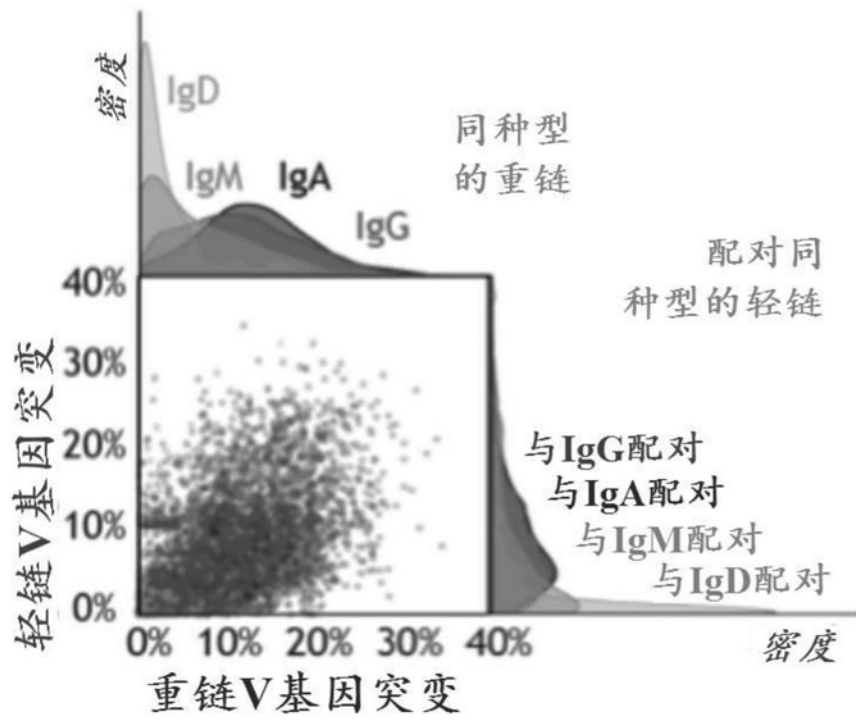


图18C

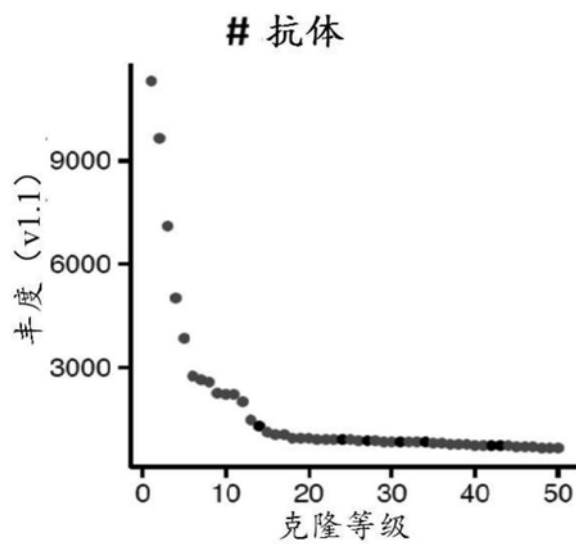


图19A

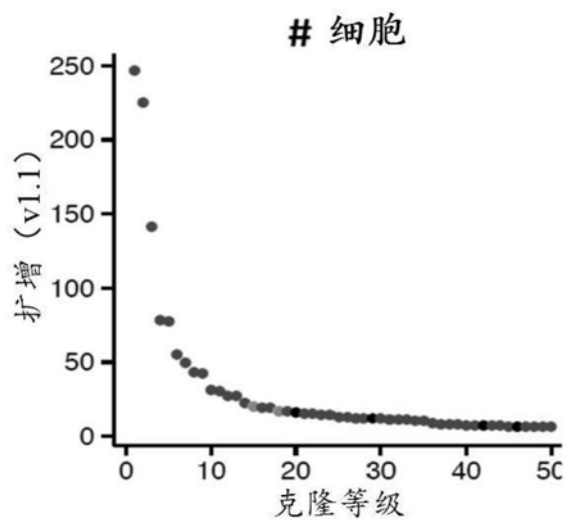


图19B

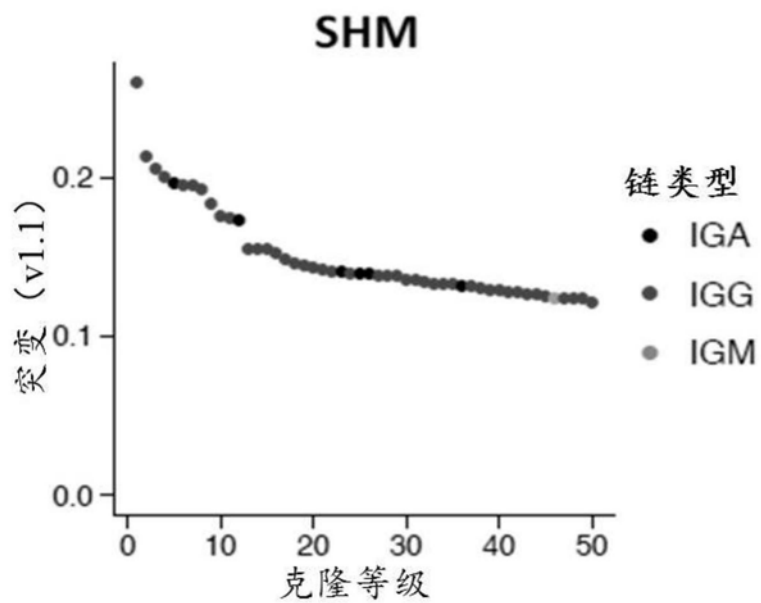


图19C

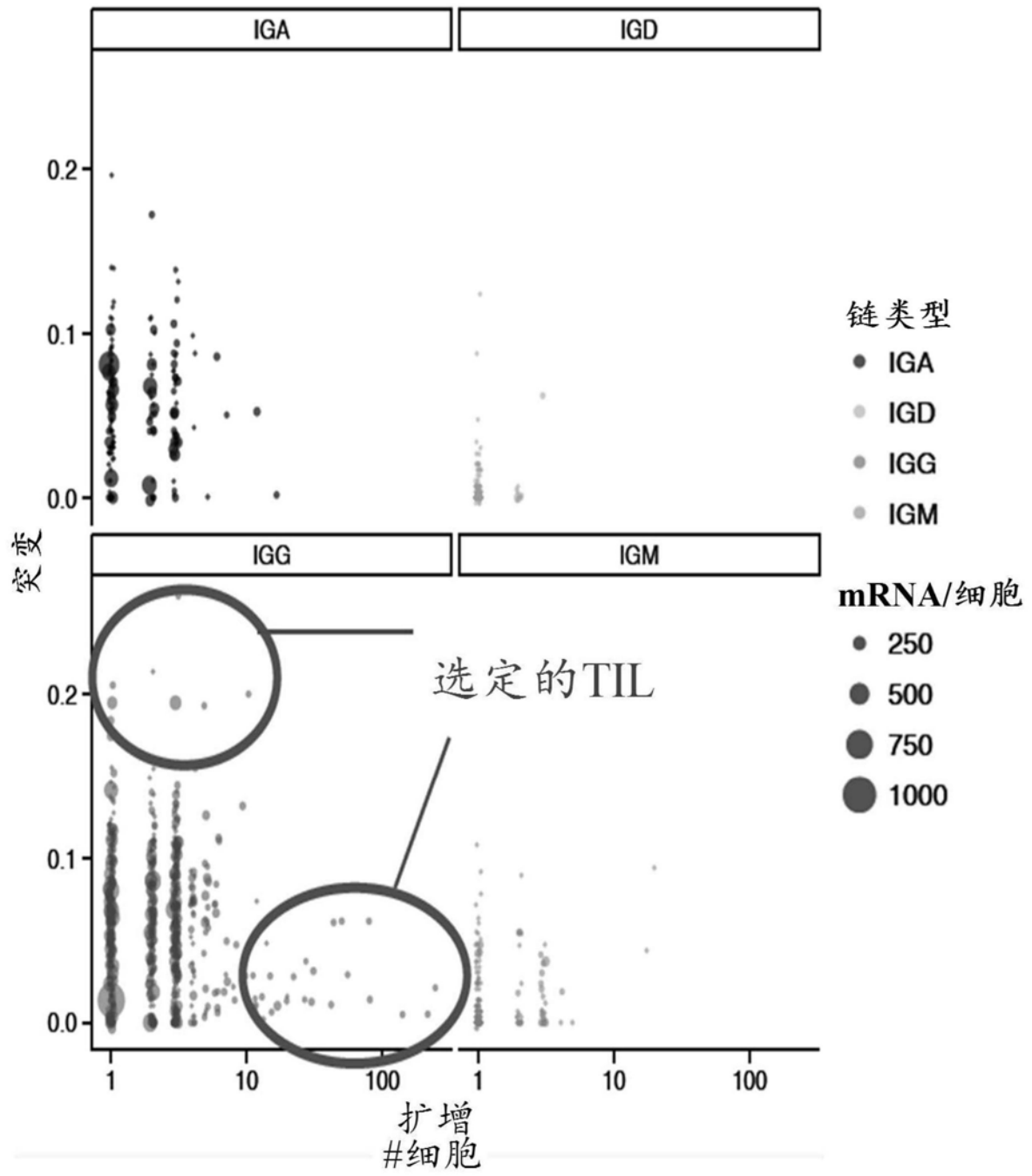


图20

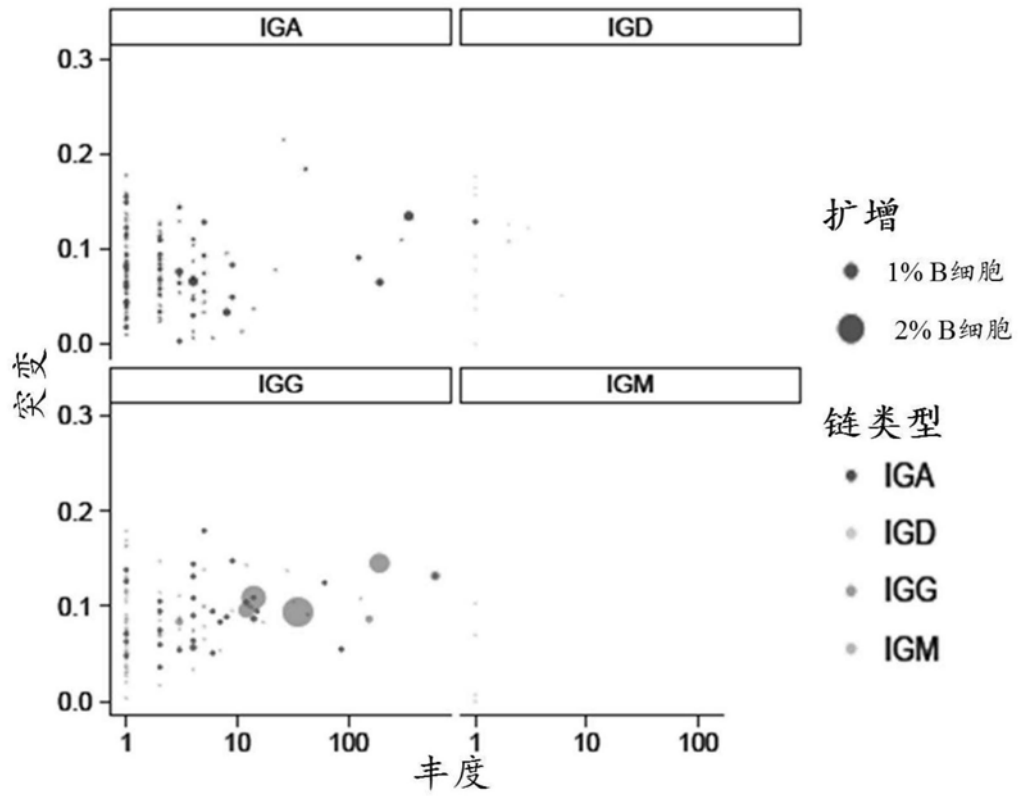


图21A

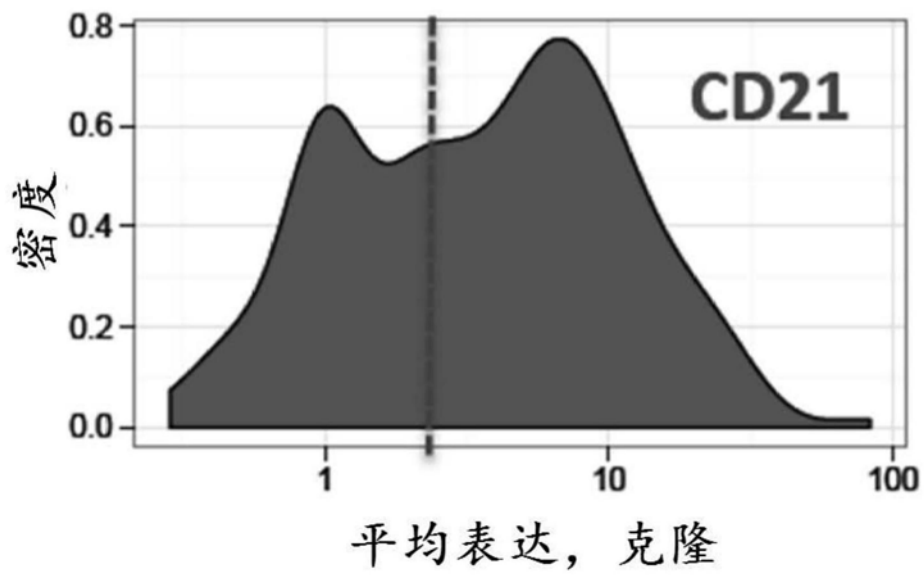


图21B

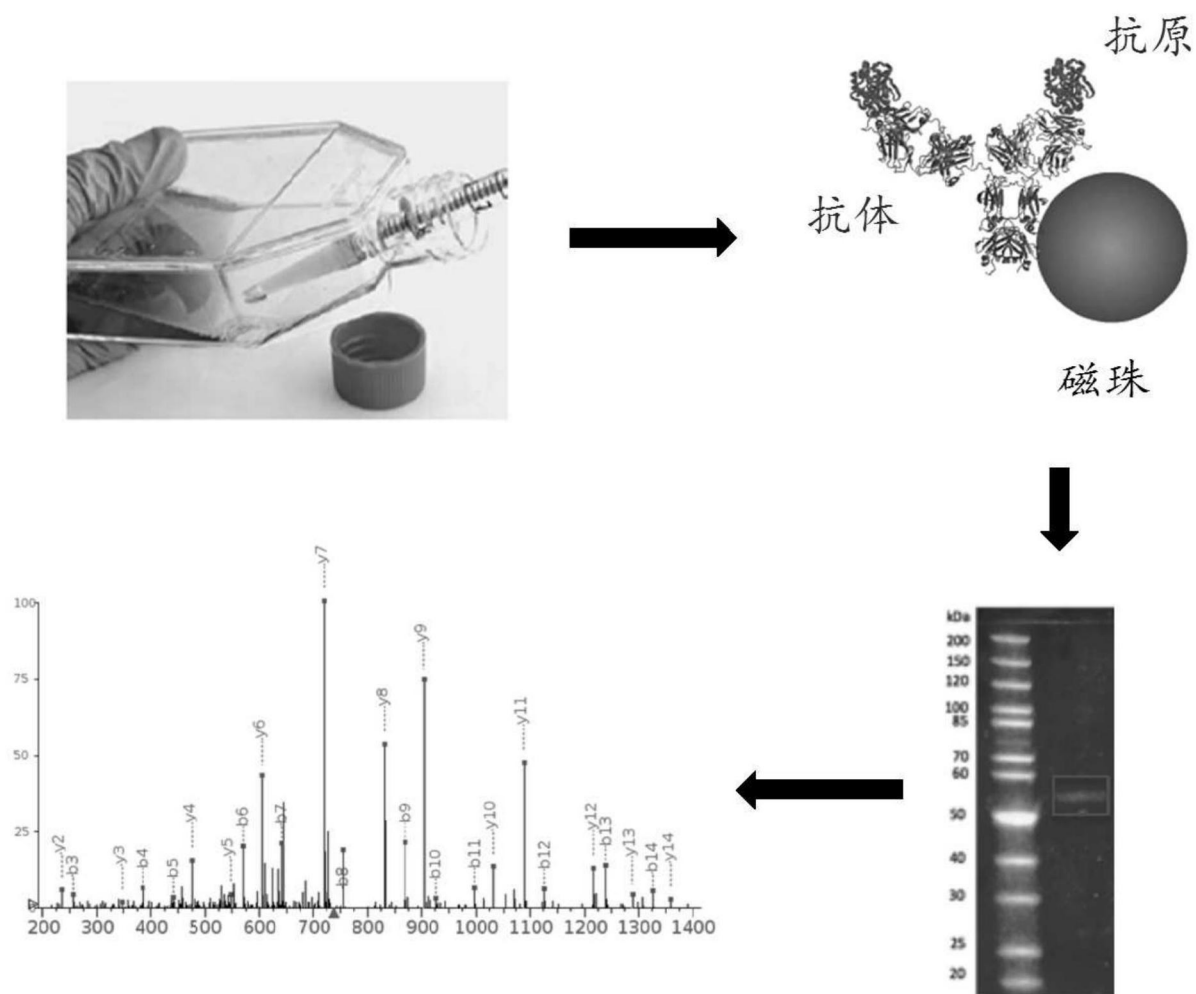


图22

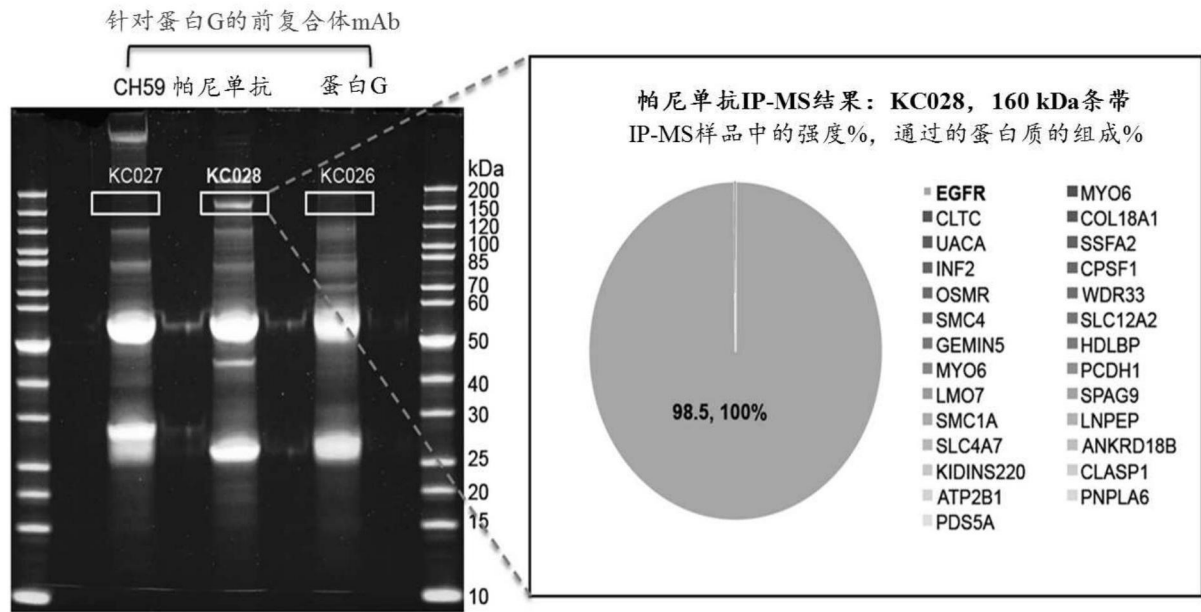


图23