



(12) 发明专利申请

(10) 申请公布号 CN 113774495 A

(43) 申请公布日 2021. 12. 10

(21) 申请号 202111003652.7

(51) Int. Cl.

(22) 申请日 2016.09.23

G40B 40/10 (2006.01)

(30) 优先权数据

G07K 14/725 (2006.01)

62/232,511 2015.09.25 US

A61P 35/00 (2006.01)

A61P 37/02 (2006.01)

(62) 分案原申请数据

G01N 33/566 (2006.01)

201680069115.3 2016.09.23

(71) 申请人 阿布维特罗有限责任公司

地址 美国华盛顿州

(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

斯蒂芬·J·戈德弗莱斯

布莱恩·J·贝尔蒙特

(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 王玮玮 郑霞

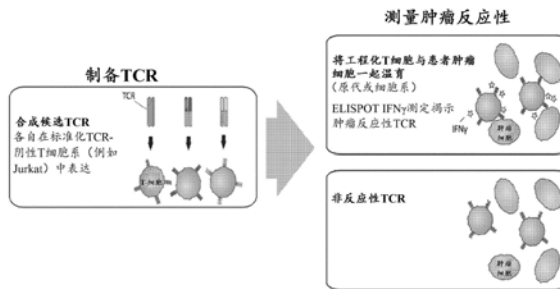
权利要求书1页 说明书74页 附图23页

(54) 发明名称

用于对天然配对T细胞受体序列进行T细胞受体靶向鉴别的高通量方法

(57) 摘要

本申请涉及一种用于对天然配对T细胞受体序列进行T细胞受体靶向鉴别的高通量方法。本文提供了用于对天然配对的T细胞受体序列进行高通量T细胞受体靶向鉴别的方法和组合物。



1. T细胞受体 (TCR) 多肽的文库, 所述文库包含第一TCR多肽和第二TCR多肽, 其中所述第一TCR多肽识别由第一多个抗原呈递细胞 (APC) 中的一个或多个APC表达的第一MHC类型, 并且所述第二TCR多肽识别由第二多个APC中的一个或多个APC表达的第二MHC类型。

2. 权利要求1所述的文库, 其中

(a) 所述第一TCR多肽对包含 (i) 由所述第一多个APC中的一个或多个APC呈递的第一抗原多肽和 (ii) 所述第一MHC类型的MHC的复合体具有特异性, 并且

(b) 所述第二TCR多肽对包含 (i) 由所述第二多个APC中的一个或多个APC呈递的第二抗原多肽和 (ii) 所述第二MHC类型的MHC的复合体具有特异性。

3. 权利要求1所述的文库, 其中所述第一TCR多肽包含一对包括TCR- α 链和TCR- β 链的天然配对的TCR链。

4. 权利要求1或权利要求3所述的文库, 其中所述第二TCR多肽包含一对包括TCR- α 链和TCR- β 链的天然配对的TCR链。

5. 权利要求1所述的文库, 其中所述第一TCR多肽包含一对包括TCR- γ 链和TCR- δ 链的天然配对的TCR链。

6. 权利要求1或权利要求5所述的文库, 其中所述第二TCR多肽包含一对包括TCR- γ 链和TCR- δ 链的天然配对的TCR链。

7. 权利要求2所述的文库, 其中所述第一抗原多肽与疾病、癌症或病毒感染相关。

8. 权利要求2或权利要求7所述的文库, 其中所述第二抗原多肽与疾病、癌症或病毒感染相关。

9. 一种制品, 所述制品包含来自第一受试者的第一多个工程化细胞和来自第二受试者的第二多个工程化细胞, 其中所述第一多个工程化细胞中的一个或多个细胞表达权利要求1所述的文库的第一TCR多肽, 并且所述第二多个工程化细胞中的一个或多个细胞表达权利要求1所述的文库的第二TCR多肽。

10. 权利要求9所述的制品, 其中由所述第一TCR多肽识别的第一MHC类型与所述第一受试者的MHC类型相似。

11. 权利要求9或权利要求10所述的制品, 其中由所述第二TCR多肽识别的第二MHC类型与所述第二受试者的MHC类型相似。

12. 权利要求9所述的制品, 其中所述第一多个APC中的一个或多个APC被工程化为表达与所述第一受试者匹配的MHC基因。

13. 权利要求9或权利要求12所述的制品, 其中所述第二多个APC中的一个或多个APC被工程化为表达与所述第二受试者匹配的MHC基因。

14. 权利要求9所述的制品, 其中所述第一多个APC中的一个或多个APC被工程化为不表达所述第一受试者的内源性MHC多肽。

15. 权利要求9或权利要求14所述的制品, 其中所述第二多个APC中的一个或多个APC被工程化为不表达所述第二受试者的内源性MHC多肽。

用于对天然配对T细胞受体序列进行T细胞受体靶向鉴别的高通量方法

[0001] 本申请是申请日为2016年09月23日,申请号为201680069115.3,发明名称为“用于对天然配对T细胞受体序列进行T细胞受体靶向鉴别的高通量方法”的申请的分案申请。

交叉引用

[0002] 本申请要求于2015年9月25日提交的美国临时申请号62/232,511的优先权;该美国临时申请通过引用以其全文并入本文。

背景技术

[0003] T淋巴细胞或T细胞是人和其他动物的适应性免疫系统的不可缺少的组分。它们构成生物体内细胞介导免疫的重要部分。存在许多T细胞亚群,包括细胞毒性T细胞和辅助性T细胞。细胞毒性T细胞(也被称为 $CD8^+$ T细胞)杀伤异常细胞,例如病毒感染细胞或肿瘤细胞。辅助性T细胞(也被称为 $CD4^+$ T细胞)有助于其他免疫细胞的激活和成熟。

[0004] 细胞毒性T细胞和辅助性T细胞二者在识别触发其各自的应答的特异性靶抗原之后执行其功能。T细胞的抗原特异性由在T细胞表面上表达的T细胞受体(TCR)限定。T细胞受体是由两个多肽链(最常见为 α 链和 β 链)构成的异二聚体蛋白质,但少数T细胞表达 γ 链和 δ 链。TCR的特定氨基酸序列和所产生的三维结构限定了TCR抗原特异性和亲和力。由于存在大量可能的TCR序列,因此任何单个T细胞的TCR链的氨基酸和编码DNA序列在生物体的整个TCR组库中几乎总是唯一的或处于非常低的丰度。这种大的序列多样性通过多种细胞机制在T细胞发育期间实现,并且是免疫系统响应各种各样潜在抗原的能力的关键方面。

[0005] TCR靶抗原是由抗原呈递细胞(APC)展示的肽。在APC的细胞表面上,肽与在人类中被称为人类白细胞抗原(HLA)的主要组织相容性复合体(MHC)蛋白质复合而呈递至T细胞。MHC I类蛋白质向细胞毒性T细胞展示肽抗原,而MHC II类蛋白质向辅助性T细胞展示肽抗原。MHC基因具有高度多态性,因此任何生物体的特定MHC蛋白质组通常基本不同于该物种内大多数或所有其他生物体。

[0006] 为了发生TCR-抗原识别,T细胞与呈递TCR同源肽的APC相互作用。该肽必须与由TCR识别的MHC分子复合,并且与T细胞类型相容(MHC-I蛋白质与 $CD8^+$ T细胞相容或者MHC-II蛋白质与 $CD4^+$ T细胞相容)。在正确的TCR-抗原/MHC识别后,T细胞得以激活,从而导致各种细胞效应。

发明内容

[0007] 本文公开了筛选和确定来源于高通量单细胞分析的大量TCR的抗原特异性的方法和组合物。在一个方面,本文提供了一种方法,该方法包括:使用高通量平行单细胞测序来自受试者的生物样品鉴别多个天然配对的TCR链;在一个或多个工程化细胞中表达TCR多肽,其中所述TCR多肽包含至少一对所述天然配对TCR链;使所述一个或多个工程化细胞与多个工程化抗原呈递细胞(APC)中的一个或多个相接触,其中所述多个工程化APC中的每一个工程化APC包含不同的抗原多肽;鉴别所述多个工程化APC中的反应性工程化APC,其中所

述反应性工程化APC对表达所述TCR多肽的所述工程化细胞具有反应性和/或表现出应答；以及确定所选择的APC的抗原多肽，从而确定所述天然配对TCR链的靶抗原。

[0008] 在一个方面，本文提供了一种方法，该方法包括：使用高通量平行单细胞测序来自受试者的生物样品鉴别多个天然配对的TCR链；在一个或多个工程化细胞中表达TCR多肽，其中所述TCR多肽包含至少一对所述天然配对TCR链；使所述一个或多个工程化细胞与多个工程化抗原呈递细胞（APC）中的一个或多个相接触，其中所述多个工程化APC中的每一个工程化APC包含不同的抗原多肽；鉴别所述多个工程化APC中的活化工程化APC，其中所述活化工程化APC激活表达所述TCR多肽的所述工程化细胞；以及确定所选择的工程化APC的抗原多肽，从而确定所述天然配对TCR链的靶抗原。

[0009] 在一个方面，本文提供了一种方法，该方法包括：使用高通量平行单细胞测序来自受试者的生物样品鉴别多个天然配对的TCR链；在一个或多个工程化细胞中表达TCR多肽，其中所述TCR多肽包含至少一对所述天然配对TCR链；使所述一个或多个工程化细胞与多个工程化抗原呈递细胞（APC）中的一个或多个相接触，其中所述多个工程化APC中的每一个工程化APC呈递不同的抗原多肽；鉴别所述多个工程化APC中的反应性工程化APC，其中所述反应性工程化APC对表达所述TCR多肽的所述工程化细胞具有反应性和/或表现出应答；以及确定在所选择的工程化APC的细胞表面上呈递的抗原多肽，从而确定所述天然配对TCR链的靶抗原。

[0010] 在一个方面，本文提供了一种方法，该方法包括：使用高通量平行单细胞测序来自受试者的生物样品鉴别多个天然配对的TCR链；在一个或多个工程化细胞中表达TCR多肽，其中所述TCR多肽包含至少一对所述天然配对TCR链；使所述一个或多个工程化细胞与多个工程化抗原呈递细胞（APC）中的一个或多个相接触，其中所述多个工程化APC中的每一个工程化APC呈递不同的抗原多肽；鉴别所述多个工程化APC中的活化工程化APC，其中所述活化工程化APC激活表达所述TCR多肽的所述工程化细胞；以及确定在所选择的工程化APC的细胞表面上呈递的抗原多肽，从而确定所述天然配对TCR链的靶抗原。

[0011] 在一个方面，本文提供了一种方法，该方法包括：使用高通量平行单细胞测序来自受试者的生物样品鉴别多个天然配对的TCR链；在一个或多个工程化细胞中表达TCR多肽，其中所述TCR多肽包含至少一对所述天然配对TCR链；使所述一个或多个工程化细胞与多个工程化抗原呈递细胞（APC）相接触，其中所述多个工程化APC中的每一个工程化APC包含不同的抗原多肽；鉴别所述多个工程化APC中的反应性工程化APC，其中所述反应性工程化APC对表达所述TCR多肽的所述工程化细胞具有反应性和/或表现出应答；以及确定所选择的APC的抗原多肽，从而确定所述天然配对TCR链的靶抗原。

[0012] 在一个方面，本文提供了一种方法，该方法包括：使用高通量平行单细胞测序来自受试者的生物样品鉴别多个天然配对的TCR链；在一个或多个工程化细胞中表达TCR多肽，其中所述TCR多肽包含至少一对所述天然配对TCR链；使所述一个或多个工程化细胞与多个工程化抗原呈递细胞（APC）中的一个或多个相接触，其中所述多个工程化APC中的每一个工程化APC包含不同的抗原多肽；确定对表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞具有反应性和/或表现出应答的工程化APC的抗原多肽；确定所述天然配对TCR链的靶抗原，其中所确定的靶抗原是对表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞具有反应性和/或表现出应答的工程化APC的抗原多肽。

[0013] 在一个方面,本文提供了一种方法,该方法包括:使用高通量平行单细胞测序来自受试者的生物样品鉴别多个天然配对的TCR链;在一个或多个工程化细胞中表达TCR多肽,其中所述TCR多肽包含至少一对所鉴别的天然配对TCR链;使所述一个或多个工程化细胞与多个工程化抗原呈递细胞(APC)中的一个或多个相接触,其中所述多个工程化APC中的每一个工程化APC呈递不同的抗原多肽;确定在工程化APC的细胞表面上呈递的抗原多肽,该工程化APC对表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞具有反应性和/或表现出应答;确定所述天然配对TCR链的靶抗原,其中所确定的靶抗原是在工程化APC的细胞表面上呈递的抗原多肽,所述工程化APC对表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞具有反应性和/或表现出应答。

[0014] 在一个方面,本文提供了一种方法,该方法包括:使用高通量平行单细胞测序来自受试者的生物样品鉴别多个天然配对的TCR链;在一个或多个工程化细胞中表达TCR多肽,其中所述TCR多肽包含至少一对所鉴别的天然配对TCR链;使所述一个或多个工程化细胞与多个工程化抗原呈递细胞(APC)中的一个或多个相接触,其中所述多个工程化APC中的每一个工程化APC包含不同的抗原多肽;确定激活表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞的工程化APC的抗原多肽;确定所述天然配对TCR链的靶抗原,其中所确定的靶抗原是激活表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞的工程化APC的抗原多肽。

[0015] 在一个方面中,本文提供了一种方法,该方法包括:使用高通量平行单细胞测序来自受试者的生物样品鉴别多个天然配对的TCR链;在一个或多个工程化细胞中表达TCR多肽,其中所述TCR多肽包含至少一对所鉴别的天然配对TCR链;使所述一个或多个工程化细胞与多个工程化抗原呈递细胞(APC)中的一个或多个相接触,其中所述多个工程化APC中的每一个工程化APC呈递不同的抗原多肽;确定在工程化APC的细胞表面上呈递的抗原多肽,该工程化APC激活表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞;确定所述天然配对TCR链的靶抗原,其中所确定的靶抗原是在工程化APC的细胞表面上呈递的抗原多肽,所述工程化APC激活表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞。

[0016] 在一个方面,本文提供了一种方法,该方法包括:使一个或多个工程化细胞与多个抗原呈递细胞(APC)中的一个或多个相接触,其中所述一个或多个工程化细胞表达T细胞受体(TCR)多肽,所述T细胞受体(TCR)多肽包含来自受试者的单个免疫细胞的至少一对天然配对的TCR链;测定所述多个APC中的APC对表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞的反应性;选择对所述表达TCR多肽的一个或多个工程化细胞中的工程化细胞具有反应性的APC;以及确定所选择的APC的抗原多肽,从而确定所述天然配对TCR链的靶抗原。

[0017] 在一个方面,本文提供了一种方法,该方法包括:使一个或多个工程化细胞与多个抗原呈递细胞(APC)中的一个或多个相接触,其中所述一个或多个工程化细胞表达T细胞受体(TCR)多肽,所述T细胞受体(TCR)多肽包含来自受试者的单个免疫细胞的至少一对天然配对的TCR链;测定所述多个APC中的APC被表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞的激活;选择激活表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞的APC;以及确定所选择的APC的抗原多肽,从而确定所述天然配对TCR链的靶抗原。

[0018] 在一个方面,本文提供了一种方法,该方法包括:使一个或多个工程化细胞与多个

抗原呈递细胞 (APC) 中的一个或多个相接触,其中所述一个或多个工程化细胞表达T细胞受体 (TCR) 多肽,所述T细胞受体 (TCR) 多肽包含来自受试者的单个免疫细胞的至少一对天然配对的TCR链;确定由所述多个APC中对表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞具有反应性的APC呈递的抗原多肽;以及确定所述天然配对TCR链的靶抗原,其中所确定的靶抗原是由所述多个APC中对表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞具有反应性的APC呈递的所确定抗原多肽。

[0019] 在一个方面,本文提供了一种方法,该方法包括:使一个或多个工程化细胞与多个抗原呈递细胞 (APC) 相接触,其中所述一个或多个工程化细胞表达T细胞受体 (TCR) 多肽,所述T细胞受体 (TCR) 多肽包含来自受试者的单个免疫细胞的至少一对天然配对的TCR链;确定由所述多个APC中激活的APC呈递的抗原多肽,所述激活的APC由所述表达TCR多肽的一个或多个工程化细胞中的工程化细胞激活;以及确定所述天然TCR链对的靶抗原,其中所确定的靶抗原是由所述多个APC中激活的APC呈递的所确定抗原多肽,所述激活的APC由表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞激活。

[0020] 在一个方面,本文提供了一种方法,该方法包括:使一个或多个工程化细胞与呈递与疾病相关的抗原多肽的多个工程化抗原呈递细胞 (APC) 相接触,其中所述一个或多个工程化细胞表达T细胞受体 (TCR) 多肽,所述T细胞受体 (TCR) 多肽包含来自受试者的单个免疫细胞的至少一对天然配对的TCR链;确定所述多个APC中由所述一个或多个工程化细胞中的工程化细胞激活或对所述一个或多个工程化细胞中的工程化细胞具有反应性的工程化APC;选择所述多个APC中由所述一个或多个工程化细胞中的工程化细胞激活或对所述一个或多个工程化细胞中的工程化细胞具有反应性的工程化APC。

[0021] 在一个方面,本文提供了一种方法,该方法包括:使一个或多个工程化细胞与包含多个细胞的样品相接触,其中所述一个或多个工程化细胞表达T细胞受体 (TCR) 多肽,所述T细胞受体 (TCR) 多肽包含来自受试者的单个免疫细胞中的至少一对天然配对的TCR链;对于包含所述多个细胞的样品中的细胞,测定所述多个细胞中的细胞对表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞的反应性;从包含所述多个细胞的所述样品中选择对表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞具有反应性的细胞;以及确定所选择的细胞的抗原多肽,从而确定所述天然配对TCR链的靶抗原。

[0022] 在一个方面,本文提供了一种方法,该方法包括:使一个或多个工程化细胞与包含多个细胞的样品相接触,其中所述一个或多个工程化细胞表达T细胞受体 (TCR),所述T细胞受体 (TCR) 包含来自受试者的单个免疫细胞中的至少一对天然配对的TCR链;对于包含所述多个细胞的样品中的细胞,测定所述多个细胞中的细胞被表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞的激活;从包含所述多个细胞的所述样品中选择激活表达TCR多肽的所述一个或多个工程化细胞中的工程化细胞的细胞;以及确定所选择的细胞的抗原多肽,从而确定所述天然配对TCR链的靶抗原。

[0023] 在一个方面,本文提供了一种方法,该方法包括:使一个或多个工程化细胞与包含多个细胞的样品相接触,其中所述一个或多个工程化细胞表达T细胞受体 (TCR) 多肽,所述T细胞受体 (TCR) 多肽包含来自受试者的单个免疫细胞的至少一对天然配对的TCR链;确定由包含所述多个细胞的所述样品中的细胞呈递的抗原多肽,所述细胞对所述表达TCR多肽的一个或多个工程化细胞中的工程化细胞具有反应性;确定所述天然配对TCR链的靶抗原,其

中所确定的靶抗原是由包含所述多个细胞的所述样品中的细胞呈递的所确定抗原多肽,所述细胞对所述表达TCR多肽的一个或多个工程化细胞中的工程化细胞具有反应性。

[0024] 在一个方面,本文提供了一种方法,该方法包括:使一个或多个工程化细胞与包含多个细胞的样品相接触,其中所述一个或多个工程化细胞表达T细胞受体(TCR)多肽,所述T细胞受体(TCR)多肽包含来自受试者的单个免疫细胞的至少一对天然配对的TCR链;测定由包含所述多个细胞的样品的细胞呈递的抗原多肽,所述抗原多肽激活所述表达TCR多肽的一个或多个工程化细胞中的工程化细胞;确定所述天然配对TCR链的靶抗原,其中所确定的靶抗原是由包含所述多个细胞的样品中的细胞呈递的所确定抗原多肽,所述细胞激活所述表达TCR多肽的一个或多个工程化细胞中的工程化细胞。

[0025] 在一些实施方案中,所述多个天然配对的TCR链被同时鉴别或从单个反应鉴别。

[0026] 在一些实施方案中,所述APC为工程化APC。

[0027] 在一些实施方案中,所述TCR多肽由外源性多核苷酸表达。在一些实施方案中,所述抗原多肽由外源性多核苷酸序列表达。在一些实施方案中,所述多个工程化APC中的每一个工程化APC均包含由编码不同抗原的外源性多核苷酸序列表达的抗原多肽。在一些实施方案中,所述一个或多个工程化细胞从外源性多核苷酸序列表达TCR多肽。

[0028] 在一些实施方案中,所述外源性多核苷酸序列编码第一T细胞受体(TCR)链和第二TCR链。在一些实施方案中,所述第一TCR链和第二TCR链包含所述至少一对天然配对的TCR链。

[0029] 在一些实施方案中,由所述工程化细胞激活或对所述工程化APC具有反应性的所述工程化APC用作免疫治疗剂。在一些实施方案中,所述工程化APC包含识别多个工程化细胞的多个工程化APC。

[0030] 在一些实施方案中,所述多个工程化APC表达至少一种MHC类型。在一些实施方案中,所述多个工程化APC表达至少两种MHC类型。在一些实施方案中,由所述工程化细胞激活或对所述工程化细胞具有反应性的工程化APC表达对所述工程化细胞的TCR多肽具有特异性的MHC类型。在一些实施方案中,所述多个工程化细胞表达对由在所述多个工程化APC中的工程化APC表达的MHC类型具有特异性的TCR多肽。

[0031] 在一些实施方案中,与疾病相关的所述抗原多肽是已知的。在一些实施方案中,所述一个或多个工程化细胞包括来自第一受试者的第一多个工程化细胞和来自第二受试者的第二多个工程化细胞,并且其中所述多个工程化APC包含第一多个工程化APC和第二多个工程化APC。在一些实施方案中,所述方法包括使所述第一多个工程化细胞与所述第一多个APC相接触,以及使所述第二多个工程化细胞与所述第二多个APC相接触。在一些实施方案中,所述第一多个工程化APC和第二多个工程化APC呈递相同的抗原多肽。在一些实施方案中,所述第一多个工程化APC表达第一MHC类型,并且所述第二多个工程化APC表达第二MHC类型。

在一些实施方案中,所述第一多个工程化细胞表达对由所述第一多个工程化APC表达的所述第一MHC类型具有特异性的第一TCR多肽,并且所述第二多个工程化细胞表达对由所述第二多个工程化APC表达的所述第二MHC类型具有特异性的第二TCR多肽。在一些方面,提供了包含第一TCR多肽和第二TCR多肽的TCR多肽的文库。在一些实施方案中,所述第一TCR多肽对包含所述抗原多肽和所述第一MHC类型的MHC的复合物具有特异性,并且其中

所述第二TCR多肽对包含所述抗原多肽和所述第二MHC类型的MHC的复合体具有特异性。

[0032] 在一些实施方案中,所述TCR多肽用作免疫治疗剂。

[0033] 在一些实施方案中,所述方法进一步包括使用高通量平行单细胞测序来自受试者的生物样品鉴别多个天然配对的TCR链,其中所述多个天然配对的TCR链被同时鉴别或从单个反应鉴别。在一些实施方案中,所述方法进一步包括生成一个或多个工程化细胞,所述一个或多个工程化细胞表达被同时鉴别的或从单个反应鉴别的所鉴别的天然配对TCR链中的至少一个。

[0034] 在一些实施方案中,所述APC来自患病受试者。

[0035] 在一些实施方案中,所述样品细胞来自受试者。在一些实施方案中,所述样品细胞为细胞系。在一些实施方案中,所述样品细胞来自患病受试者。在一些实施方案中,所述抗原呈递细胞来自患有癌症的受试者。在一些实施方案中,所述抗原呈递细胞来自受病毒感染的受试者。在一些实施方案中,所述抗原呈递细胞与所述单个免疫细胞来自相同的受试者。在一些实施方案中,所述抗原呈递细胞来自具有主要组织相容性复合体(MHC)组库的受试者,所述主要组织相容性复合体(MHC)组库与包含所述单个免疫细胞的所述受试者的MHC组库相似。在一些实施方案中,所述抗原呈递细胞来自具有主要组织相容性复合体(MHC)单体型的受试者,所述主要组织相容性复合体(MHC)单体型与包含所述单个免疫细胞的所述受试者的MHC单体型相似。在一些实施方案中,所述抗原呈递细胞来自具有主要组织相容性复合体(MHC)组库的受试者,所述主要组织相容性复合体(MHC)组库与包含所述单个免疫细胞的所述受试者相容。在一些实施方案中,将所述工程化APC工程化为表达与所述受试者匹配的MHC基因。在一些实施方案中,所述工程化APC来源于K562细胞系。在一些实施方案中,表达所述TCR多肽的所述工程化细胞来源于Jurkat细胞系。在一些实施方案中,所述抗原呈递细胞为原代细胞。在一些实施方案中,所述抗原呈递细胞为细胞系。在一些实施方案中,来自受试者的抗原呈递细胞(APC)对表达所述TCR多肽的所述工程化细胞具有反应性。在一些实施方案中,来自受试者的抗原呈递细胞(APC)由表达所述TCR多肽的所述工程化细胞激活。

[0036] 在一些实施方案中,所述方法进一步包括分离工程化APC。在一些实施方案中,所述确定所述工程化APC的抗原包括对编码由所述工程化APC呈递的抗原的所述工程化APC的序列进行测序。在一些实施方案中,所述工程化APC通过荧光激活细胞分选来分离。

[0037] 在一些实施方案中,所述方法进一步包括对编码第一T细胞受体(TCR)链和第二TCR链的外源性多核苷酸序列进行转染。在一些实施方案中,所述方法进一步包括对编码不同抗原的外源性多核苷酸序列进行转染。在一些实施方案中,所述方法进一步包括将不同抗原添加至所述工程化APC的细胞表面。

[0038] 在一些实施方案中,所述方法进一步包括在所述接触之前确定来自所述受试者的所述单个免疫细胞的天然配对的TCR链序列。在一些实施方案中,所述方法进一步包括确定多个天然配对的TCR链中的每一个的靶抗原。在一些实施方案中,所述工程化细胞包含多个工程化细胞,每一个均表达不同对的天然配对的TCR链序列。

[0039] 在一些实施方案中,所述工程化细胞是不表达内源性TCR多肽的工程化T细胞。在一些实施方案中,所述工程化细胞是不转录内源性TCR基因的工程化T细胞。在一些实施方案中,所述工程化细胞是不包含内源性TCR基因的工程化T细胞。在一些实施方案中,所述工

程化细胞是TCR敲除的工程化T细胞。在一些实施方案中,所述工程化细胞为T细胞。在一些实施方案中,所述工程化细胞表达CD4。在一些实施方案中,所述工程化细胞表达CD8。

[0040] 在一些实施方案中,所述抗原呈递细胞包含多个抗原呈递细胞,每一个均呈递不同的抗原。在一些实施方案中,所述测定的抗原呈递细胞在分隔的容器中。

[0041] 在一些实施方案中,所述APC包含外源性报道序列。在一些实施方案中,所述APC包含编码报道基因的外源性报道序列。在一些实施方案中,所述报道基因为萤光素酶基因。在一些实施方案中,所述报道基因编码荧光蛋白。在一些实施方案中,所述报道基因编码细胞表面蛋白质。在一些实施方案中,细胞表面表达的蛋白质配体与工程化APC上的合成Notch (synNotch) 受体的结合导致所述报道基因的表达。在一些实施方案中,所述细胞表面表达的蛋白质配体由所述工程化细胞表达。在一些实施方案中,所述细胞表面表达的蛋白质配体在诱导型启动子的控制下。在一些实施方案中,所述诱导型启动子为NFAT启动子。在一些实施方案中,将所述synNotch受体切割并且生成转录激活结构域,如GAL4-VP64结构域。在一些实施方案中,所述转录激活结构域诱导所述报道基因的表达。在一些实施方案中,所述转录激活结构域通过结合激活序列如GAL4-UAS激活序列来诱导所述报道基因的表达。在一些实施方案中,所述报道基因的表达指示所述工程化细胞被所述APC的激活。在一些实施方案中,所述报道基因的表达指示所述工程化APC对所述工程化细胞的反应性。在一些实施方案中,所述测定包括测定所述报道基因的表达。

[0042] 在一些实施方案中,所述测定包括cDNA阵列、mRNA检测或蛋白质检测。

[0043] 在一些实施方案中,所述工程化APC包含编码报道基因的外源性报道序列。在一些实施方案中,所述报道基因为萤光素酶基因。在一些实施方案中,所述报道基因编码荧光蛋白。在一些实施方案中,所述报道基因编码细胞表面蛋白质。在一些实施方案中,所述报道基因编码IL6、IL-12、IFN α 、IL-23、IL-1、TNF或IL-10。在一些实施方案中,所述细胞表面蛋白质为CD80、CD86、MHC I、MHC II、CD11c、CD11b、CD8 α 、OX40-L、ICOS-1或CD40。在一些实施方案中,所述报道基因响应于所述工程化APC上的CD80/CD86结合所述工程化细胞上的CD28而得以激活。在一些实施方案中,所述报道基因在所述APC中的Notch信号传导激活的下游得以激活。在一些实施方案中,所述报道基因在所述APC中的PI3K/AKT信号传导激活的下游得以激活。在一些实施方案中,所述报道基因在NF- κ B途径或NOTCH1信号传导途径的下游得以激活。

[0044] 在一些实施方案中,所述报道基因在第二受体的下游得以激活。在一些实施方案中,所述第二受体在结合由所述工程化细胞分泌的配体后得以激活。在一些实施方案中,所述配体的表达在结合所述工程化APC上的合成Notch (synNotch) 受体之后得以诱导。

[0045] 在一些实施方案中,所述工程化APC不表达内源性MHC多肽。在一些实施方案中,所述工程化APC不转录内源性MHC基因。在一些实施方案中,所述工程化APC不包含内源性MHC基因。

[0046] 在一些实施方案中,所述第一TCR链包含TCR- α 链,并且所述第二TCR链包含TCR- β 链。在一些实施方案中,所述第一TCR链包含TCR- γ 链,并且所述第二TCR链包含TCR- δ 链。

[0047] 在一个方面,本文提供了一种治疗方法,该方法包括向有需要的受试者施用靶向包含根据本文所述的方法确定的抗原的蛋白质的治疗剂。在一些实施方案中,所述受试者为人类。在一些实施方案中,所述治疗剂为小分子、肽、蛋白质、抗体或其片段、细胞疗法或

免疫疗法。在一些实施方案中,所述细胞疗法包括过继性细胞疗法。在一些实施方案中,所述方法包括施用T细胞,所述T细胞被工程化为表达根据本文所述所述的方法鉴别的所述多个TCR链中的一个或多个天然配对TCR链。在一些实施方案中,所述受试者患有疾病或病况。在一些实施方案中,所述疾病或病况为癌症。

附图说明

[0048] 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:

[0049] 图1描绘了针对患者来源TCR的初始肿瘤反应性的筛选方法的示例性流程图。

[0050] 图2描绘了靶向鉴别患者来源TCR的方法的示例性流程图。

[0051] 图3描绘了用于发现抗原特异性人源TCR的方法的示例性流程图。

[0052] 图4A-4D例示了来自乳液中单个免疫细胞条码化的方法的结果。

[0053] 图4A是将两种含有细胞和裂解/反应(LR)混合物的水性流传递到油中,从而以每小时超过800万个小液滴产生单分散乳液的示例性描述。

[0054] 图4B是示例性描述,显示容器内的细胞裂解并在单个反应中经历分子特异性条码化和小液滴特异性条码化。

[0055] 图4C是示例性描述,显示靶mRNA被逆转录并模板转换标记有通用衔接子序列。随后,对在开始时稀释至每个小液滴约1个分子的小液滴条码模板进行PCR扩增。通过互补重叠延伸将扩增的条码附加至模板转换的cDNA。在另外的文库处理步骤和高通量测序之前,从乳液中回收产物并使用RT引物上的生物素进行纯化。

[0056] 图4D是示例性描述,显示双条码化允许测序读取簇集到其分子和起源小液滴中,从而在尽可能减少测序错误和扩增偏差的同时重建天然受体链配对。

[0057] 图5A-5F例示了来自分离的健康B细胞回收BCR的方法的结果。

[0058] 图5A是小液滴的示例性描述,其中3百万个B细胞以0.2个细胞/小液滴传递到乳液中,导致约90%的占用细胞含有单细胞。

[0059] 图5B是 $V_H V_L$ 配对精确度的示例性描述。在乳液条码化和测序后,针对来自单细胞小液滴的数据对数据进行富集,并且使用扩增克隆间的配对一致性估计 $V_H V_L$ 配对精确度。

[0060] 图5C是小液滴条码百分比与Ig同种型的示例性图示。显示了259,368个过滤的 $V_H V_L$ 对的重链同种型(每个小液滴内最丰富的同种型)和轻链基因座使用。

[0061] 图5D是六个独立乳液级分中的每一个中100个最高频重链克隆的等级丰度的示例性图示。标出了0.05%的总频率。

[0062] 图5E是如通过每个小液滴条码内捕获的mRNA的数目所估计的细胞内 V_H 与 V_L 表达的示例性图示。每种同种型显示5,000个点。

[0063] 图5F是针对BCR对的 V_H 与 V_L 突变相关性和每种同种型内的密度分布的示例性图示。

[0064] 图6A-6C例示了来自HIV广谱中和抗体(bNAb)发现方法的结果。

[0065] 图6A是将来自HIV elite控制器的B细胞的38,620个回收的 $V_H V_L$ 对输入到乳液中的重链同种型分布示例性图示。少量比例的IgG链与先前已知的bNAb(—PGT样)良好对齐。

[0066] 图6B是已知bNAb加新回收的bNAb(通过线条连接,用小液滴条码标记)的完整VDJ

氨基酸序列的系统树的示例性描述,其中重链(左)和轻链(右)单独绘制。潜在错配的抗体是PGT122重链和PGT123轻链,以及PGT123重链和PGT122轻链。

[0067] 图6C是与PGT121的对照原种相比,8个新发现的PGT样变体针对十种HIV株的中和活性(IC_{50} , $\mu\text{g/mL}$)的示例性描述。

[0068] 图7A-7F例示了对来自卵巢肿瘤的TIL进行表征的方法的结果。

[0069] 图7A是小液滴的示例性描述,其中将来自卵巢肿瘤的400,000个未分选的解离细胞输入到乳液中,并且通过乳液条码化同时回收BCR对和TCR对。

[0070] 图7B是小液滴条码与受体链组合的示例性图示,显示在过滤后于小液滴条码内观察到的所有 V_H/V_L 和 V_α/V_β 组合的数目。

[0071] 图7C是所回收的BCR对的小液滴条码百分比与重链同种型分布的示例性图示。

[0072] 图7D是BCR对的 V_H 与 V_L 突变相关性和每种同种型内密度分布的示例性图示。

[0073] 图7E描绘了针对TCR对和BCR对整体(上图)和针对不同同种型(下图)的捕获mRNA的数目的示例性图示。

[0074] 图7F描绘了克隆分析的示例性图示,显示了六个独立的乳液级分中的每一个中30个最高频BCR重链克隆(上图)和30个最高频TCR β 链克隆的等级丰度。显示了1%和10%的总频率水平。

[0075] 图8A描绘了免疫受体序列与附加的mRNA和蛋白质靶标的示例性共同捕获。通过在乳液测序之前将细胞与DNA标记的染色抗体一起预温育而对表面蛋白质靶标进行定量。

[0076] 图8B描绘了由健康人类T细胞生成的3,682个小液滴条码TCR $V_\alpha V_\beta$ 对上的示例性CD4和CD8 mRNA和蛋白质测量。

[0077] 图8C描绘了mRNA测量与蛋白质测量之间的示例性一致性(每个点是与TCR $V_\alpha V_\beta$ 对相连接的小液滴条码)。

[0078] 图8D描绘了对来自乳液中的未分选T细胞的CD4和CD8同时进行mRNA和蛋白质检测的示例性表格。从30,000个输入T细胞中回收了3,682个TCR对。通过mRNA相比蛋白质(基于分子计数,少数服从多数原则)被判定为CD4⁺或CD8⁺的TCR对的频率显示在矩阵中。

[0079] 图9A描绘了本文所述方法中的容器的示例性示意图。

[0080] 图9B描绘了与抗体缀合的寡核苷酸标签的示例性设计。每个有色模块均代表完整寡核苷酸序列的一部分。融合序列用于乳液反应内小液滴特异性DNA条码的酶促衔接。仅示出了一种可能的序列排列,但其他排列也与所描述的方法兼容。

[0081] 图10A-10E例示了使用抗体-寡核苷酸缀合物进行免疫分型的方法。

[0082] 图10A描绘了显示2个容器的示例性示意图,每个容器均含有与抗体-寡核苷酸缀合物结合的单细胞。(DB1-小液滴条码1;DB2-小液滴条码2;MB1-分子条码1;MB2-分子条码2;AID-抗原ID条码;AMB1-抗体分子条码1;AMB2-抗体分子条码2)。

[0083] 图10B描绘了显示2个容器的示例性示意图,每个容器均含有来自图9A容器的裂解细胞的RNA分子。将RNA分子逆转录,并将非模板核苷酸添加到由逆转录产生的cDNA分子的末端。使分子条码与添加到由逆转录产生的cDNA分子末端的非模板核苷酸杂交。

[0084] 图10C描绘了显示2个容器的示例性示意图,每个容器均含有模板条码化多核苷酸,将该模板条码化多核苷酸被扩增并通过杂交衔接至来自图9B的容器的cDNA,并且cDNA被延伸(上图)。然后延伸的cDNA被扩增(下图)。

[0085] 图10D描绘了示例性示意图,显示相同分子条码(MB) 附接至相同的同一RNA序列的RNA-MB-DB种类可能是PCR复制的结果。两个不同MB附接至相同的同一RNA序列的RNA-MB-DB种类(RNA1-MB1-DB和RNA1-MB2-DB) 是来自起源而非PCR复制的两个独立RNA分子。

[0086] 图10E描绘了示例性示意图,显示相同抗体分子条码(AMB) 附接至具有相同小液滴条码(DB) 和抗原ID条码(AID) 的序列的DB-AMB-AID种类可能是PCR复制的结果。两个不同AMB附接至具有相同小液滴条码(DB) 和抗原ID条码(AID) 的序列的DB1-AMB1-AID1和DB1-AMB2-AID1种类是来自两个独立抗体寡核苷酸缀合物分子的两个独立寡核苷酸分子,所述两个独立抗体寡核苷酸缀合物分子各自具有特异性结合附接至容器中的相同单细胞的相同靶抗原的抗体,而并非来自PCR复制。两个不同AID附接至具有相同小液滴条码(DB) 和相同或不同抗体分子条码(AMB) 的序列的DB1-AMB_n-AID1和DB1-AMB_n-AID2种类是来自附接至容器中的相同单细胞的两个独立抗体寡核苷酸缀合物分子的两个独立寡核苷酸分子,其中该抗体寡核苷酸缀合物分子中的一个具有特异性结合第一靶抗原的抗体,而另一个抗体寡核苷酸缀合物分子具有特异性结合第二靶抗原的抗体。

具体实施方式

[0087] 参考用于说明的实例应用,多个方面描述如下。应当理解,许多具体细节、关系和方法的阐述是为了提供对本文所述的特征的全面了解。然而,相关领域普通技术人员将容易认识到,本文所述的特征可以在不存在一个或多个具体细节或采用其他方法的情况下实践。本文所述的特征不受限于所阐述的行为或事件的排序,因为一些行为可能以不同顺序发生和/或与其他行为或事件同时发生。此外,实施依照本文所述的特征的方法并非需要全部所阐述的行为或事件。

[0088] 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“一该”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”。

[0089] 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。

[0090] 本发明的目的在于提供使用抗体-寡核苷酸缀合物(例如,在乳液中)对单细胞(例如,免疫细胞)进行表型分型的方法和组合物。

[0091] TCR特异性由其两条TCR链(α 和 β 或 γ 和 δ)的氨基酸序列确定。为了概括TCR和进一步进行表征,期望确定两个链序列。这些多肽链独立于T细胞内的独特基因座位而表达。使用大量T细胞群体的TCR测序方法产生许多链序列,但其确定来源于相同原始细胞的TCR链的天然配对仍然较困难。因此,为了确定来自单独T细胞的完全、适当配对的TCR链序列,单细胞分析方法是重要的。

[0092] 尽管单细胞TCR测序方法已经存在多年,但它们历来具有极低的通量(每次实验1-

100个细胞)。最近,研究人员发明了新方法对大量单个T细胞的配对TCR组库进行概况分析,产生数千至数百万个天然配对的TCR序列。遗憾的是,不存在直接有效的用于确定大量TCR的抗原特异性的方法。

[0093] 在一个方面,方法包括确定哪些TCR对感兴趣的特定组织类型(例如,肿瘤细胞或病毒感染的细胞)具有反应性(图1)。在初始序列确定后,可将配对的TCR链克隆到合适的表达载体或慢病毒载体中,并在合适的T细胞系内共表达。可根据实验目的,将T细胞系(例如, Jurkat)工程化为不表达内源性TCR而表达CD4或CD8共受体。因此,这些T细胞的抗原特异性可通过所引入的TCR来确定。另外,这些T细胞可含有在TCR-抗原识别及随后T细胞激活时表达报道基因(例如,荧光素酶或荧光蛋白)的报道DNA构建体。这可涉及例如在T细胞激活时上调的NFAT蛋白响应性启动子。

[0094] 这些经工程化的、表达TCR的T细胞可与感兴趣的APC(原代细胞或细胞系)一起温育。这些APC细胞可来源于产生该TCR的原始宿主或具有相似MHC组库的另一宿主。这可以确保MHC等位基因可以被探讨中的TCR识别。

[0095] T细胞激活/TCR反应性可以通过免疫相关细胞因子(例如,IFN- γ)的工程化报道基因表达或分泌来确定。

[0096] 在一个方面,方法进一步包括鉴别候选TCR命中的抗原(图2)。在一些实施方案中,将TCR表达载体转染到如上所述的类似T细胞系中。在一些实施方案中,将TCR表达载体转染到如上所述类似的、但没有激活报道基因的T细胞系中。

[0097] 在一些实施方案中,将T细胞与基于细胞的人造或工程化APC(分别为aAPC或eAPC)混合。在一些实施方案中,aAPC细胞系(例如,基于K562细胞系)不表达内源性MHC蛋白。在一些实施方案中,用与具有探讨中的TCR的原始生物体的MHC基因匹配的MHC基因转染aAPC细胞。人类中的MHC在本文中也称为人类白细胞抗原(HLA)。在一些实施方案中,aAPC还被工程化为含有在被T细胞识别抗原时会表达报道基因(例如,荧光素酶、荧光蛋白、表面蛋白质)的报道基因构建体。

[0098] 在一些实施方案中,将抗原表达载体的文库转染到aAPC中。抗原编码序列可用于感兴趣的肽、小基因构建体或可在MHC-1结合和表面展示之前适当地加工成肽的整个cDNA编码序列。也可将肽直接添加到aAPC以供MHC装载。抗原文库可以由来自感兴趣的靶细胞的一组无偏差的蛋白质编码区构成,或者可以被更狭义地定义(例如,通过外显子组测序确定的新抗原、病毒衍生的基因)。

[0099] 在一些实施方案中,将表达TCR的T细胞与抗原呈递aAPC一起温育。成功的TCR抗原/MHC识别导致T细胞激活并引起aAPC报道基因表达。在一些实施方案中,将表达TCR的T细胞与表达内源性MHC的非工程化和/或非人造APC一起温育。在一些实施方案中,用靶抗原文库对内源性APC进行脉冲处理。

[0100] 该方法的优点在于可以针对较大的潜在抗原文库对表达单个TCR的T细胞群体进行探询。展示T细胞TCR的靶抗原的aAPC可以表达报道基因。报道基因的表达可以允许分离aAPC。例如,报道基因的表达可以允许通过荧光激活细胞分选或基于表面标志物的方法分离aAPC。可以确定在每个分离的aAPC中表达的抗原的编码序列,这将揭示所呈递抗原的身份,并因此揭示TCR的靶标。这最终允许有效地确定许多TCR针对较大的潜在抗原文库的特异性。

[0101] 本文所述的方法和组合物能够发现针对疾病靶标的治疗相关的TCR。在一些实施方案中,获得来自代表性人类群体的TCR序列。在一些实施方案中,将这些序列的文库转染或转导到T细胞中。在一些实施方案中,将T细胞与不同组的表达特定MHC和靶抗原的aAPC一起温育。在一些实施方案中,T细胞含有在抗原识别和T细胞激活时表达的报告基因构建体。在一些实施方案中,报告分子允许通过基于荧光或表面标志物的分选从群体中分离激活的细胞。在一些实施方案中,此类针对疾病相关靶标具有反应性的TCR可以是有用的并且可适用于疾病的治疗。由于人类群体中MHC序列的多样性,发现针对相同抗原具有反应性但与特定的相对常见的MHC相容的多种TCR是有利的。

[0102] 评估了来自外周血的数百万个未刺激的原始和静息记忆B细胞和成百上千个解离的肿瘤细胞,从而产生来自血液样品的全长配对BCR和来自在肿瘤中的TIL的BCR和TCR的组库水平数据集。T细胞在传统上是TIL研究的主要焦点,但B细胞在肿瘤中的作用正受到越来越多的关注。来自广泛科学界的对TIL重燃的兴趣可归因于细胞转移免疫疗法的突破性发展、CAR-T抗实体瘤的发展以及检查点抑制剂及其对肿瘤生长影响的发现。这些研究领域全都需要了解所涉及的TIL,以便监测应答患者与非应答患者中的TIL波动和演化,或使用TIL受体作为工具来发现新的肿瘤抗原。推进这一研究领域的主要障碍在于肿瘤中只有一小部分的TIL可能是抗原特异性的,以及由于一般炎症造成的较大的非特异性免疫细胞背景。

[0103] 因此,为鉴别在肿瘤环境中特别富集的TCR和BCR,采样足够数目的TCR和BCR至关重要,这一任务可能需要数千个细胞,而不是基于板的TIL分析提供的小数目。另外,由于已经报道了TIL显示出非规范性的表面表型,因此基于标志物的分选可能错过重要的TIL群体。

[0104] 最近的几项研究已经报道了在单细胞转录组分析的放大中的引人注目的进展,但迄今为止还没有展现出对来自异质样品的免疫组库进行有意义V(D)J捕获所需的读取长度或每个样品数十万个细胞的通量。特别地,Klein等人(InDrop)和Macosko等人(Dropseq)的方法允许从数千个细胞中进行基于乳液的全转录组测序。然而,全长免疫受体基因的捕获,以及每个实验需要相比于那些报道多数十至数百倍的细胞以供充分的免疫组库表征,都提出了显著不同的挑战。本文所述的方法非常适合于这种应用,原因在于其通过使用比Dropseq和InDrop的乳液小液滴小15-75倍的乳液小液滴而提供了更高的通量,以及通过PCR在乳液内生成克隆小液滴条码(其免去了通过裂分-合并(split-pool)合成来预先合成条码化小珠的需要)。另外,由于本文提出的方法涉及在原始乳液小液滴中的逆转录和PCR,因此更大范围的靶标类型可以旨在允许从单细胞中同时捕获mRNA、标记的蛋白质和潜在的基因组DNA,同时还尽可能减小条码之间乳液后交叉污染的风险。

[0105] 迄今为止,用于靶向BCR或TCR测序的所有方法均受以下限制中的至少一个所困:未配对的受体链、配对链的低通量(每个实验 $<10^4$ 个)或部分序列回收。相比之下,本文提供的乳液方法克服了这些限制,并展示出许多其他优点。首先,将链与短序列条码连接,而不是如Dekosky等人所执行的将其物理剪接成长扩增子,这允许对完整V(D)J可变区进行测序,从而确保对体细胞突变进行综合分析并允许完全精确的功能重建。其次,通过对mRNA本身和来自单细胞的所有产物进行双重条码化,可使PCR和测序错误最小化,并且特别地,每个克隆谱系中的细胞数目可由该克隆的细胞内的表达水平去卷积。这种特征在本文中首次应用于免疫组库分析,其克服了从大量mRNA进行免疫测序的限制(其测量整体表达而不测量克隆丰度)以及从基因组DNA进行免疫测序的限制(其揭示克隆丰度而不揭示关于表达的

信息)。第三,这种条码化方法可延伸至具有在mRNA和蛋白质水平下的表型值的其他基因,从而增加关于表达每个受体序列的T细胞和B细胞亚型的重要信息。本文提供的方法代表了对乳液中的单细胞同时进行RNA测序与蛋白质测量的首次证明。这种方法将通过流式细胞术收集的信息类型与免疫测序的数字解析相结合,从而允许将每个细胞的表型与其全长受体序列相联系。因此,通过揭示哪些T细胞和B细胞谱系代表特定的细胞亚型和诸如免疫抑制或“衰竭”(通常是抗肿瘤反应性的关键指标)等状态,该方法有望成为肿瘤受体概况分析的有力补充。通过在乳液分离和分析之前将B细胞与DNA条码化靶抗原一起温育,该方法还可应用于发现对所需抗原具有特异性的抗体。

[0106] 总之,从解离肿瘤样品的数千个TIL中回收了天然配对的BCR和TCR,这导致鉴别了许多经突变、类别转换和被细胞高表达的克隆扩增受体谱系,强烈提示功能活性B细胞和T细胞向肿瘤部位的特异性募集。连同用于鉴别在该肿瘤及其他肿瘤中发现的TIL受体的靶标的互补方法,本文所述的技术将能够发现新的肿瘤抗原,促进我们对肿瘤免疫生物学的基本理解,并且提供有力工具来探讨新兴免疫疗法的免疫效果。

天然配对的T细胞受体链的确定

[0107] T细胞受体链对和抗体免疫球蛋白链对均为免疫受体的类型,且在演化上相关。本发明的目的在于生成多核苷酸文库以供高通量测序和诊断。本发明的目的还在于从患者或具有特定共同属性的队列开发人源文库组(panel)以供抗体和/或TCR发现。起始材料可为外周血或来自组织活检,如果需要的话,从中总体分离免疫细胞或亚分选为幼稚、记忆和ASC。本公开发明可应用于多种不同类型的成对可变序列,例如,T细胞受体链对和抗体免疫球蛋白链对。

[0108] 分离的细胞,诸如免疫细胞,可封装在容器如油包水乳液(小液滴)中,以这样的方式创造每个小液滴含有单个或更少的免疫细胞的单独皮升区室。对于每个样品,诸如来自受试者的生物样品,可处理数百万个细胞,从而允许单细胞测序技术中的高通量。使用本文所述的方法可避免使用固体支持体,诸如珠子。使用本文所述的方法也可避免对生成两个分隔的容器群体的需要。例如,序列的文库可在相同的或单个反应中,或在单独的多个容器或容器群体中生成。在容器形成期间引入与细胞多核苷酸(诸如 V_H 和 V_L 抗体链和/或 $V\alpha/V\beta$ 和 $V\gamma/V\delta$ T细胞受体(TCR)链)互补的多核苷酸。在容器形成期间也可引入带有容器条码的多核苷酸。这些容器条码化多核苷酸可携带简并条码,使得每个含有容器条码的细胞多核苷酸含有对应于它们所处的容器的独特识别码。在容器形成期间还可引入多个带有分子条码的多核苷酸。这些分子条码化多核苷酸可携带简并条码,使得每个含有分子条码的细胞多核苷酸分子含有对应于它们所来自的单细胞多核苷酸分子的独特识别码。数百万个单个免疫细胞可在乳液内裂解,细胞转录物,如 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 链转录物,可使用引物进行逆转录或拷贝,随后用容器条码和分子条码标记,并PCR扩增该条码化多核苷酸。起源于单个免疫细胞(例如,B细胞或T细胞)的每个 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 链可借助相同的容器条码标识实际上彼此关联。

[0109] 然后,可从容器中回收 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 链并PCR富集,以便添加下一代测序(NGS)标签。可使用高通量测序平台对该文库进行测序,然后分析组库多样性、抗体频率、CDR3表征、体细胞超突变系统发生分析等。可通过对容器和分子条码序列进行去卷积来生成正确匹配的 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 对的数据库。因为每个单个免疫细胞都隔

离在它们各自的容器中,所以对于每个观察到两次的容器条码,测序的转录物来源于相同的乳液小液滴,并因此来自独特的单细胞。对于观察到的每个不同分子条码,对于含有相同容器条码的序列,测序的转录物来源于来自单细胞的不同转录物分子。对于观察到的每个相同分子条码,对于含有相同容器条码的序列,测序的转录物来源于来自单细胞的相同转录物分子(例如,PCR复制物)。

[0110] 与测序并行地,从容器中回收的 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 链的文库可克隆到抗体表达载体中并共转染以供酵母展示筛选。与在开始时就分裂生物样品相比,克隆此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。人源的 V_H 和 V_L 和/或 $V\alpha/V\beta$ 和/或 $V\gamma/V\delta$ 链的文库不论正确或错误的成对匹配均可表达,如经典展示分析一样。然后,可针对一个或多个抗原靶标进行酵母展示,以富集潜在的抗体候选物。

[0111] 从展示技术如酵母展示得到的阳性候选抗体可进行测序并查询匹配对的条码数据库。每个酵母展示的 V_H 和/或 $V\alpha$ 和/或 $V\gamma$ 链可分别回溯匹配于其各自的 V_L 或 $V\beta$ 或 $V\delta$ 链,且每个酵母展示的 V_L 和/或 $V\beta$ 和/或 $V\delta$ 链可分别回溯匹配于其各自的 V_H 或 $V\alpha$ 或 $V\gamma$ 链。这些正确匹配的候选物可在哺乳动物细胞系中基因合成并表达,并针对感兴趣的靶标进行功能性验证。这些候选物可以是完全人类抗体和/或TCR。

[0112] 因此,本文中的术语“抗体”以最广泛的含义使用,并且包括多克隆抗体和单克隆抗体,包括完整抗体及其功能性(抗原结合)抗体片段,该功能性(抗原结合)抗体片段包括片段抗原结合(Fab)片段、F(ab')₂片段、Fab'片段、Fv片段、重组IgG(rIgG)片段、单链抗体片段(包括单链可变片段(scFv))和单结构域抗体(例如,sdAb、sdFv、纳米抗体)片段。该术语包含免疫球蛋白的基因工程化形式和/或以其他方式修饰的形式,诸如胞内抗体、肽抗体、嵌合抗体、完全人类抗体、人源化抗体和异源偶联抗体、多特异性抗体(例如,双特异性抗体)、双抗体、三抗体和四抗体、串联双-scFv、串联三-scFv。除非另有说明,术语“抗体”应理解为包括其功能性抗体片段。该术语还包含完整抗体或全长抗体,包括任何类别或亚类的抗体,包括IgG及其亚类(IgM、IgE、IgA和IgD)。

[0113] 与“高变区”或“HVR”同义的术语“互补决定区”和“CDR”在本领域中是已知的,用于指代抗体可变区内氨基酸赋予抗原特异性和/或结合亲和力的非连续序列。通常,每个重链可变区中有三个CDR(CDR-H1、CDR-H2、CDR-H3),并且每个轻链可变区中有三个CDR(CDR-L1、CDR-L2、CDR-L3)。“构架区”和“FR”在本领域中是已知的,用于指代重链和轻链可变区的非CDR部分。通常,每个全长重链可变区中有四个FR(FR-H1、FR-H2、FR-H3和FR-H4),并且每个全长轻链可变区有四个FR(FR-L1、FR-L2、FR-L3和FR-L4)。

[0114] 给定CDR或FR的精确氨基酸序列边界可使用许多众所周知的方案中的任一种容易地确定,所述方案包括Kabat等人(1991),“Sequences of Proteins of Immunological Interest,”第5版.Public Health Service,National Institutes of Health,Bethesda,MD(“Kabat”编号方案)、Al-Lazikani等人,(1997)JMB 273,927-948(“Chothia”编号方案)、MacCallum等人,J.Mol.Biol.262:732-745(1996),“Antibody-antigen interactions:Contact analysis and binding site topography,”J.Mol.Biol.262,732-745.”(“Contact”编号方案)、Lefranc MP等人,“IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like

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

[0115] 给定CDR或FR的边界可根据用于鉴别的方案而变化。例如,Kabat方案是基于结构比对,而Chothia方案是基于结构信息。Kabat方案和Chothia方案二者的编号均基于最常见的抗体区序列长度,其中插入由插入字母(例如,—30a”)调节并且在一些抗体中出现缺失。这两种方案在不同位置处放置某些插入和缺失(—indel”),从而导致差异的编号。Contact方案基于对复合晶体结构的分析,并且在许多方面与Chothia编号方案相似。

[0116] 以下表A列出了分别通过Kabat、Chothia和Contact方案鉴别的CDR-L1、CDR-L2、CDR-L3和CDR-H1、CDR-H2、CDR-H3的示例性位置边界(参见,Kabat等人(1991),Sequences of Proteins of Immunological Interest,第5版.Public Health Service,National Institutes of Health,Bethesda,MD;和Al-Lazikani等人,(1997)JMB 273,927-948)。对于CDR-H1,使用Kaba编号方案和Chothia编号方案二者来列出残基编号。FR位于CDR之间,例如FR-L1位于CDR-L1与CDR-L2之间,等等。值得注意的是,由于所示的Kabat编号方案在H35A和H35B处放置了插入,因此Chothia CDR-H1环的末端在使用所示的Kabat编号惯例进行编号时根据环的长度而在H32与H34之间变化。

表A

| CDR | Kabat | Chothia | Contact |
|--------------------------------------|--------------|----------------|----------------|
| CDR-L1 | L24--L34 | L24--L34 | L30--L36 |
| CDR-L2 | L50--L56 | L50--L56 | L46--L55 |
| CDR-L3 | L89--L97 | L89--L97 | L89--L96 |
| CDR-H1 (Kabat 编号 ¹) | H31--H35B | H26--H32..34 | H30--H35B |
| CDR-H1 (Chothia 编号 ²) | H31--H35 | H26--H32 | H30--H35 |
| CDR-H2 | H50--H65 | H52--H56 | H47--H58 |
| CDR-H3 | H95--H102 | H95--H102 | H93--H101 |

[0117] 因此,除非另有规定,否则给定抗体或其区域(如其可变区)的“—CDR”或“—互补决定区”或单个指定的CDR(例如,—CDR-H1、CDR-H2)应理解为包括如由任何前述方案所定义的互补决定区(或特定互补决定区)。例如,在阐明特定CDR(例如,CDR-H3)含有给定VH或VL氨基酸序列中相应CDR的氨基酸序列的情况下,应当理解,此类CDR具有如由任何前述方案所定义的可变区内的相应CDR(例如,CDR-H3)的序列。在一些实施方案中,指定了特定的CDR序列。

[0118] 同样地,除非另有规定,否则给定抗体或其区域(如其可变区)的FR或单个指定FR

(例如,FR-H1、FR-H2)应理解为包括如由任何已知方案所定义的构架区(或特定构架区)。在一些情况下,指定了用于鉴别特定CDR、FR或FR或CDR(如由Kabat、Chothia或Contact方法所定义的CDR)的方案。在其他情况下,给出了CDR或FR的具体氨基酸序列。

[0119] 术语“可变区”或“可变域”是指涉及抗体与抗原结合的抗体重链或轻链的结构域。天然抗体的重链和轻链(分别为VH和VL)的可变域通常具有相似的结构,其中每个结构域包含四个保守的构架区(FR)和三个CDR。(参见例如,Kindt等人,Kuby Immunology,第6版,W.H.Freeman and Co.,第91页(2007))。单个VH或VL域可足以赋予抗原结合特异性。此外,结合特定抗原的抗体可使用VH或VL域从结合该抗原的抗体中分离,从而分别筛选互补的VL或VH域的文库,参见例如,Portolano等人,J.Immunol.150:880-887(1993);Clarkson等人,Nature 352:624-628(1991)。

[0120] 所提供的抗体中存在抗体片段。“抗体片段”是指除了完整抗体以外的包含完整抗体的一部分的分子,该部分结合该完整抗体所结合的抗原。抗体片段的实例包括但不限于Fv、Fab、Fab'、Fab'-SH、F(ab')₂;双抗体;线性抗体;单链抗体分子(例如,scFv);以及由抗体片段形成的多特异性抗体。在特定的实施方案中,抗体是包含可变重链区和/或可变轻链区的单链抗体片段,如scFv。

[0121] 除非另有说明,术语“TCR”应理解为包括完整TCR以及其抗原结合部分或抗原结合片段(也被称为MHC-肽结合片段)。在一些实施方案中,TCR为完整或全长TCR。在一些实施方案中,TCR为小于全长TCR但与结合至MHC分子(即,在MHC分子的语境下)的特定抗原肽(即,MHC-肽复合物)结合的抗原结合部分。在一些情况下,TCR的抗原结合部分或片段可以仅含有全长或完整TCR的结构域中的一部分,但仍能够结合该完整TCR结合的表位(例如,MHC-肽复合物)。在一些情况下,TCR的抗原结合部分或片段含有TCR的可变域,诸如TCR的可变 α 链和可变 β 链,该可变区足以形成用于与特定MHC-肽复合物结合的结合位点(诸如通常在每条链含有三个互补决定区的情况下)。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。这些术语还涉及互补决定区(CDR)移植抗体和TCR以及其他人源化抗体和TCR(包括CDR修饰和构架区修饰)。应当注意,虽然可能只提及了免疫球蛋白链(例如,重链和轻链),但是本公开发明可应用于多种其他不同类型的成对序列,例如T细胞受体链对(TCR α 和TCR β 链以及TCR γ 和TCR δ 链),且不限于免疫球蛋白。

[0122] T细胞识别与多种癌症或传染性生物体相关的抗原的能力由其TCR赋予,TCR由alpha(α)链和beta(β)链或由gamma(γ)和delta(δ)链构成。构成这些链的蛋白质由DNA编码,其采用独特的机理生成了TCR的巨大多样性。这种多亚单元免疫识别受体与CD3复合物相关联,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接触处发生。

[0123] 在一些实施方案中,TCR或其他MHC-肽结合分子识别或可能识别在MHC I类分子的环境中的T细胞表位。MHC I类蛋白质在高等脊椎动物的所有有核细胞中表达。MHC I类分子是由与12-kDa轻链 β -2微球蛋白非共价缔合的46-kDa重链组成的异二聚体。在人类中,存在若干个MHC等位基因,诸如例如HLA-A2、HLA-A1、HLA-A3、HLA-A24、HLA-A28、HLA-A31、HLA-A33、HLA-A34、HLA-B7、HLA-B45和HLA-Cw8。MHC等位基因的序列是已知的,并且可以例如在www.ebi.ac.uk/ipd/imgt/hla处可获得的IMGT/HLA数据库中找到。在一些实施方案中,MHC

I类等位基因为HLA-A2等位基因,其在一些群体中由该群体中的大约50%成员表达。在一些实施方案中,HLA-A2等位基因可以是HLA-A*0201、HLA-A*0202、HLA-A*0203、HLA-A*0206或HLA-A*0207基因产物。在一些情况下,不同群体之间亚型的频率可能存在差异。例如,在一些实施方案中,超过95%的HLA-A2阳性高加索人群为HLA-A*0201,而在中国人群中该频率被报道为大约23%HLA-A*0201、45%HLA-A*0207、8%HLA-A*0206和23%HLA-A*0203。

[0124] 在一些实施方案中,TCR或其他MHC-肽结合分子识别或可能识别在MHC II类分子的环境中的T细胞表位。MHC II类蛋白质在有核脊椎动物细胞的子集(通常被称为抗原呈递细胞(APC))中表达。在人类中,存在若干个MHC II类等位基因,诸如例如DR1、DR3、DR4、DR7、DR52、DQ1、DQ2、DQ4、DQ8和DP1。在一些实施方案中,MHC II类等位基因为HLA-DRB1*0101、HLA-DRB*0301、HLA-DRB*0701、HLA-DRB*0401和HLA-DQB1*0201。MHC等位基因的序列是已知的,并且可以例如在www.ebi.ac.uk/ipd/imgt/hla处可获得的IMGT/HLA数据库中找到。

[0125] 每个TCR均包含可变互补决定区(CDR)以及构架区(FR)。 α 和 β 链可变域的第三互补决定区(CDR3)环的氨基酸序列很大程度上分别决定了由 β 链基因座中可变性(V β)、多样性(D β)和连接(J β)基因区段之间,以及 α 链基因座中类似的V α 和J α 基因区段之间的重组引起的 $\alpha\beta$ T细胞的序列多样性。多种这类基因区段在TCR α 和 β 链基因座中的存在允许大量不同的CDR3序列得到编码。在TCR基因重排过程期间,在V β -D β 、D β -J β 和V α -J α 连接处的核苷酸的独立添加和缺失进一步增加了CDR3序列多样性。在此方面,免疫活性反映在TCR的多样性中。

[0126] 在一些方面,由B细胞表达的免疫球蛋白(Ig)是由四条多肽链——即两条重链(IgH)和两条轻链(IgL)组成的蛋白质,从而形成H₂L₂结构。每一对IgH和IgL链含有高变域和恒定域,该高变域由V_L和V_H区组成。Ig的IgH链有多个类型, μ 、 δ 、 γ 、 α 和 β 。个体内的Ig的多样性主要由高变域决定。与TCR相似,IgH链的V结构域通过V_H、D_H和J_H基因区段的组合连接而创建。在Ig基因重排过程期间,在V_H-D_H、D_H-J_H和V_H-J_H连接处的核苷酸的独立添加或缺失进一步增加了高变域序列多样性。在此,免疫活性反映在Ig的多样性中。

[0127] 术语“可变区”或“可变域”是指参与抗体与抗原结合的抗体重链或轻链的结构域。天然抗体的重链和轻链的可变域(分别为VH和VL)通常具有相似的结构,其中每个结构域包含四个保守的构架区(FR)和三个CDR。(参见例如,Kindt等人,Kuby Immunology,第6版,W.H.Freeman and Co.,第91页(2007))。单个VH或VL域可足以赋予抗原结合特异性。此外,结合特定抗原的抗体可使用VH或VL域从结合该抗原的抗体中分离,从而分别筛选互补VL或VH域的文库,参见例如,Portolano等人,J.Immunol.150:880-887(1993);Clarkson等人,Nature 352:624-628(1991)。

[0128] “亲和部分”是指与靶抗原相互作用的亲和-寡核苷酸缀合物的部分。示例性亲和部分包括抗体、肽、蛋白质、适体、小分子、药物、细胞及其他。

[0129] “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。

[0130] 所提供的抗体中存在抗体片段。“抗体片段”是指除了完整抗体以外的包含完整抗体的一部分的分子,该部分结合该完整抗体所结合的抗原。抗体片段的实例包括但不限于Fv、Fab、Fab'、Fab'-SH、F(ab')₂;双抗体;线性抗体;单链抗体分子(例如,scFv);以及由抗体片段形成的多特异性抗体。在特定的实施方案中,抗体是包含可变重链区和/或可变轻

链区的单链抗体片段,如scFv。

[0131] 单结构域抗体是包含抗体的全部或部分重链可变域或全部或部分轻链可变域的抗体片段。在某些实施方案中,单结构域抗体为人类单结构域抗体。

[0132] 抗体片段可以通过各种技术制备,包括但不限于完整抗体的蛋白质水解消化以及通过重组宿主细胞的生产。在一些实施方案中,抗体是重组产生的片段,诸如包含非天然存在的排列的片段,诸如具有两个或更多个通过合成连接体(例如,肽连接体)连接的抗体区或链和/或可能无法通过天然存在的完整抗体的酶消化而产生的片段。在一些方面,抗体片段为scFv。

[0133] 还提供了TCR片段,包括抗原结合片段。在一些实施方案中,TCR为其抗原结合部分,诸如可被称为完全可溶性TCR的不含其跨膜区和/或胞质区的全长TCR的变体。在一些实施方案中,TCR为二聚TCR(dTCR)。在一些实施方案中,TCR为单链TCR(scTCR),诸如具有如PCT专利公开号WO 03/020763、WO 04/033685或WO 2011/044186中所描述的结构scTCR。在某些实施方案中,TCR为包含与β链可变区连接的α链可变区的单链TCR片段,如scTv。在一些实施方案中,scTv也被称为scFv。

[0134] 在一些方面,单链Fv或scFv是指包含抗体的可变重链(V_H)域和可变轻链(V_L)域或TCR的可变α或γ链($V\alpha$ 或 $V\gamma$)域和可变β或δ链($V\beta$ 或 $V\delta$)域的抗体片段或TCR片段,其中这些结构域存在于单条多肽链中。通常,Fv多肽进一步在 V_H 和 V_L 结构域或 $V\alpha$ 和 $V\beta$ 结构域或 $V\gamma$ 和 $V\delta$ 结构域之间包含多肽连接体,其使sFv能够形成用于抗原结合的所需结构。

[0135] 在一些方面,双抗体是指具有两个抗原结合位点的小抗体和/或TCR片段,该片段包含同一多肽链中的与 V_L 连接的 V_H (V_H-V_L),或同一多肽链中的与 $V\beta$ 连接的 $V\alpha$ ($V\alpha-V\beta$),或同一多肽链中的与 $V\delta$ 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。

[0136] 在一些方面,双特异性抗体或双特异性TCR是指对两种不同类型的抗原显示出特异性的抗体或TCR。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。

[0137] 在一些方面,线性抗体或线性TCR是指形成一对抗原结合区的一对串联Fd区段(例如, $V_H-C_{H1}-V_H-C_{H1}$ 或 $V\alpha-C\alpha_1-V\alpha-C\alpha_1$)。线性抗体和TCR可为双特异性或单特异性的,例如,如Zapata等人,Protein Eng.8(10):1057-1062(1995)所述。

[0138] 在一些方面,抗原结合域是指抗体或TCR的一个或多个片段,其保留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,包括scFv,(v) dAb片段(Ward等人,(1989) Nature 341:544-546),其含有 V_H 结构域;以及(vi) 分离的CDR。此定义中还包括包含单个重链和单个轻链的抗体或具有单个α链或单个β链的TCR。

[0139] $-F(ab')_2$ 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 $F(ab')_2$ 。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] 在一些方面,Fv是指含有完整的抗原识别和抗原结合位点的抗体或TCR片段。此区域由紧密、非共价缔合的一个重链和一个轻链可变域或一条TCR α 链和一条TCR β 链或一条TCR γ 链和一条TCR δ 链的二聚体组成。在此构型中,每个可变域的三个CDR相互作用,在 V_H - V_L 二聚体或 $V\alpha$ - $V\beta$ 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 $V\gamma$ - $V\delta$ 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 F_v)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 $V\gamma$ 和 $V\delta$ 链区配对以形成单价分子(称为单链Fv(scFv);Bird等人(1988)Science 242:423-426;Huston等人(1988)Proc.Natl.Acad.Sci.USA 85:5879-5883;以及Osborn等人(1998)Nat.Biotechnol.16:778)。这样的scFv还意在包含在术语抗体的一“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,来自驼类和鲨鱼的抗体。驼类和鲨鱼抗体包含V-样和C-样结构域(均不具有轻链)的两条链的同源二聚体对。因为驼类中重链二聚体IgG的 V_H 区不需要与轻链进行疏水相互作用,所以在驼类中通常与轻链接触的重链区域变为亲水氨基酸残基。重链二聚体IgG的 V_H 结构域被称为 V_{HH} 结构域。鲨鱼Ig-NAR包含一个可变域(称为V-NAR结构域)和五个C-样恒定域(C-NAR结构域)的同源二聚体。在驼类中,抗体组库的多样性由 V_H 或 V_{HH} 区中的CDR1、2和3决定。驼类 V_{HH} 区中的CDR3的特征为其相对较长的长度,平均为16个氨基酸(Muyldermans等人,1994,Protein Engineering 7(9):1129)。

[0142] “一人源化”抗体是其中全部或基本上全部CDR氨基酸残基都来源于非人类CDR且全部或基本上全部FR氨基酸残基都来源于人类FR的抗体。人源化抗体可任选地包括来源于人类抗体的抗体恒定区的至少一部分。非人类抗体的一“人源化形式”是指已经历人源化的非人类抗体的变体,该人源化通常用于降低对人的免疫原性,同时保留亲代非人类抗体的特异性和亲和力。在一些实施方案中,将人源化抗体中的一些FR残基用来自非人类抗体(例如,衍生出CDR残基的抗体)的相应残基置换,例如用以恢复或提高抗体特异性或亲和力。

[0143] 所提供的抗体中存在人类抗体。“人类抗体”是具有这样的氨基酸序列的抗体,该氨基酸序列与由人或人细胞或采用人类抗体组库或其他人抗体编码序列(包括人类抗体文库)的非人来源产生的抗体的氨基酸序列相对应。该术语排除包含非人类抗原结合区的非人类抗体的人源化形式,诸如其中全部或基本上全部CDR均是非人类的人源化形式的非人类抗体。

[0144] 人类抗体可通过向转基因动物施用免疫原来制备,该转基因动物已经被修饰以响应于抗原激发而产生完整人类抗体或具有人类可变区的完整抗体。这样的动物通常含有全部或部分人类免疫球蛋白基因座,该人类免疫球蛋白基因座取代内源性免疫球蛋白基因座,或者存在于染色体外或被随机整合到动物染色体中。在这样的转基因动物中,内源性免疫球蛋白基因座通常已经失活。人类抗体还可以来源于含有源自人组库的抗体编码序列的人类抗体文库,包括噬菌体展示文库和无细胞文库。

[0145] 所提供的抗体中存在单克隆抗体,包括单克隆抗体片段。如本文所用的,术语“单克隆抗体”是指从基本均质的抗体群体(即,组成该群体的各个抗体是相同的)获得的抗体或在基本均质的抗体群体内的抗体,而含有天然存在的突变或在生产单克隆抗体制备物期间产生的可能变体(这样的变体通常以少量存在)除外。与通常包括针对不同表位的不同抗体的多克隆抗体制备物相反,单克隆抗体制备物的每个单克隆抗体是针对抗原上的单个表位。该术语不应被解释为抗体生产需要通过任何特定的方法。单克隆抗体可通过多种技术制备,包括但不限于从杂交瘤生成、重组DNA方法、噬菌体展示以及其他抗体展示方法。

[0146] 术语“多肽”和“蛋白质”可互换使用,用于指代氨基酸残基的聚合物,且不限制最小长度。包括所提供的抗体和抗体链以及其他肽(例如,连接体和结合肽)在内的多肽可包括氨基酸残基,包括天然和/或非天然的氨基酸残基。该术语还包括多肽的表达后修饰,例如糖基化、唾液酸化、乙酰化、磷酸化等。在一些方面,多肽可含有相对于原始或天然序列的修饰,只要该蛋白质仍保持所需的活性即可。这些修饰可以是故意的,诸如通过定点诱变,或可以是偶然的,诸如通过产生该蛋白质的宿主的突变或由于PCR扩增引起的错误。

[0147] “一种系序列”是指来自种系(单倍体配子和形成它们的那些二倍体细胞)的基因序列。种系DNA含有编码单个Ig重链或轻链、或单个TCR α 或TCR β 链或单个TCR γ 或TCR δ 链的多个基因区段。这些基因区段携带于生殖细胞中,但在排列为功能性基因之前不能转录和翻译。在骨髓中的B细胞和T细胞分化期间,这些基因区段被能够生成超过 10^8 种特异性的动态遗传系统随机改组。这些基因区段中的大部分由种系数据库公布并收集。

[0148] 亲和力是指两种物质可逆结合的平衡常数,且表示为 K_D 。结合蛋白质对配体的亲和力,诸如抗体对表位的亲和力可为,例如,约100纳摩尔(nM)至约0.1nM、约100nM至约1皮摩尔(pM)或约100nM至约1飞摩尔(fM)。术语“亲合力”是指两种或多种物质的复合体在稀释后解离的阻力。

[0149] 在一些方面,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。在一些方面,TCR可识别在MHC的环境中的

一个或多个氨基酸序列或表位。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。

[0150] 在一些实施方案中,提及具有特异性结合的抗体或TCR是指这样的情况:除了含有被该抗体或TCR识别的表位的抗原以外,抗体或TCR将不会对其他分子显示出任何显著的结合。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。

[0151] 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。

[0152] 当与治疗性组合物结合使用时,单位剂量是指适合作为用于人类的单一剂量的物理上离散的单元,每个单元含有经计算将产生所需治疗效果的预定量的活性物质以及需要的稀释剂,即载体或媒介物。

[0153] 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。

[0154] 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。

[0155] “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。

[0156] 在一些方面,中和抗体或中和TCR是指抑制病原体如病毒或细菌复制的任何抗体或TCR,而无论实现中和的机理如何。

[0157] 抗体组库或TCR组库是指抗体、TCR或其片段的集合。例如,抗体组库可用来选择特定的抗体或针对诸如结合能力、结合特异性、胃肠转运能力、稳定性、亲和力等特定性质进行筛选。该术语特别地包括抗体和TCR文库,包括所有形式的组合文库,例如抗体噬菌体展

示文库,包括但不限于来自任何来源的单链Fv(scFv)和Fab抗体噬菌体展示文库,包括原始、合成和半合成文库。

[0158] “一目标核酸分子”、“一目标分子”、“一目标多核苷酸”、“一目标多核苷酸分子”是指任何感兴趣的核酸。

[0159] 聚合酶链反应(PCR)是指通过双链多核苷酸的互补链的同时引物延伸,多核苷酸序列的体外扩增反应。PCR反应产生侧翼为引物结合位点的模板多核苷酸的拷贝。采用两条引物的结果是,对于每次循环,两条链的模板多核苷酸拷贝数的指数增加,因为每次循环两条链都得到复制。多核苷酸双链体具有对应于所用引物的末端的末端。PCR可包括使模板多核苷酸变性、使引物与引物结合位点退火以及在核苷酸的存在下通过DNA或RNA聚合酶延伸引物的一次或多次重复。特定温度、每个步骤的持续时间和步骤之间的变化率取决于本领域普通技术人员公知的许多因素。(McPherson等人,IRL Press,Oxford(1991和1995))。例如,在使用Taq DNA聚合酶的传统PCR中,双链模板多核苷酸可在 $>90^{\circ}\text{C}$ 的温度下变性,引物可在 $50-75^{\circ}\text{C}$ 范围内的温度下退火,且引物可在 $72-78^{\circ}\text{C}$ 范围内的温度下延伸。在一些实施方案中,PCR包括逆转录PCR(RT-PCR)、实时PCR、巢式PCR、定量PCR、多重PCR等。在一些实施方案中,PCR不包括RT-PCR。(美国专利号5,168,038、5,210,015、6,174,670、6,569,627和5,925,517;Mackay等人,Nucleic Acids Research,30:1292-1305(2002))。RT-PCR包括在PCR反应之前的逆转录反应,并扩增所产生的cDNA。巢式PCR包括两级PCR,其中使用第一引物组的第一PCR反应的扩增子成为使用第二引物组的第二PCR反应的样品,至少一个第二引物与第一PCR反应的扩增子的内部位置相结合。多重PCR包括PCR反应,其中多个多核苷酸序列在同一反应混合物中同时经历PCR。PCR反应体积可为 $0.2\mu\text{L}-1000\mu\text{L}$ 之间的任何体积。定量PCR包括设计用于测量样品中一个或多个序列的绝对或相对量、丰度或浓度的PCR反应。定量测量可包括将一个或多个参考序列或标准与感兴趣的多核苷酸序列进行比较。(Freeman等人,Biotechniques,26:112-126(1999);Becker-Andre等人,Nucleic Acids Research,17:9437-9447(1989);Zimmerman等人,Biotechniques,21:268-279(1996);Diviacco等人,Gene,122:3013-3020(1992);Becker-Andre等人,Nucleic Acids Research,17:9437-9446(1989))。

[0160] 在其他实施方案中,本文公开的方法、试剂盒和组合物可包含支持体。在其他实施方案中,本文公开的方法、试剂盒和组合物不包含支持体。通常,固体支持体包含一种或多种材料,该材料包含一个或多个刚性或半刚性表面。在一些实施方案中,该支持体为非固体支持体。该支持体或基底可包括膜、纸、塑料、涂层表面、平面、玻璃、载片、芯片或其任意组合。在一些实施方案中,支持体的一个或多个表面基本上为平面,然而在一些实施方案中,可期望其对于不同的化合物在物理上分隔合成区域,例如采用孔、凸起区、针、蚀刻沟槽等。在一些实施方案中,固体支持体包括珠子、树脂、凝胶、微球或其他几何构型。或者,固体支持体可包括二氧化硅芯片、微粒、纳米粒子、平板和阵列。固体支持体可包括使用在微孔中自组装的珠子。例如,固体支持体包括Illumina的BeadArray技术。或者,固体支持体包括Abbott Molecular的Bead Array珠阵列技术和Applied Microarray的FlexiPlex™系统。在其他情况下,固体支持体为板。板的实例包括但不限于MSD多阵列板、MSD Multi-Spot®板、微板、ProteOn微板、AlphaPlate、DELFI板、IsoPlate和LumaPlate。在一些实施方案中,支持体可包括多个珠子。在一些实施方案中,支持体可包括阵列。在一些实施方案中,支持体

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

[0161] 如本文所用的,“一核苷酸”、“一核苷”、“一核苷酸残基”和“一核苷残基”可意指能够作为适于在扩增反应(例如PCR反应)中使用的引物的组分的脱氧核糖核苷酸或核糖核苷酸残基,或其他相似的核苷类似物。除非另有说明,这类核苷及其衍生物可用作本文所述的引物的构件。本申请中的任何内容都不意在排除使用经化学修饰以增强它们的稳定性或在扩增反应中的有效性的核苷衍生物或碱基,条件是该化学修饰不干扰它们被聚合酶适当地识别为脱氧鸟嘌呤、脱氧胞嘧啶、脱氧胸腺嘧啶或脱氧腺嘌呤。在一些实施方案中,核苷酸类似物可以稳定杂合体形成。在一些实施方案中,核苷酸类似物可以去稳定杂合体形成。在一些实施方案中,核苷酸类似物可以增强杂交特异性。在一些实施方案中,核苷酸类似物可以降低杂交特异性。

[0162] “一核酸”或语法上的等同物是指单个核苷酸或共价连接在一起的至少两个核苷酸。

[0163] “一多核苷酸”或语法上的等同物是指共价连接在一起的至少两个核苷酸。多核苷酸包括含有两个或更多个核苷酸的分子。多核苷酸包括任何长度的核苷酸的聚合形式,该核苷酸为核糖核苷酸、脱氧核糖核苷酸或肽核酸(PNA),其包含嘌呤和嘧啶碱基,或其他天然的、化学或生物修饰的、非天然的核苷酸碱基或其衍生物。多核苷酸的骨架可包含糖和磷酸基团,或修饰或取代的糖或磷酸基团。多核苷酸可包含修饰的核苷酸,诸如甲基化的核苷酸和核苷酸类似物。核苷酸的序列可被非核苷酸组分中断。多核苷酸可包括其他分子,诸如另一种杂交的多核苷酸。多核苷酸包括脱氧核糖核酸(DNA)、核糖核酸(RNA)或两者的序列。多核苷酸的非限制性实例包括基因、基因片段、外显子、内含子、基因间DNA(包括但不限于异染色质DNA)、信使RNA(mRNA)、转移RNA、核糖体RNA、核酶、小干扰RNA(siRNA)、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、序列的分离的DNA、序列的分离的RNA、核酸探针和引物。多核苷酸可从天然来源中分离、重组或人工合成。

[0164] 多核苷酸可包括非标准核苷酸,诸如核苷酸类似物或修饰的核苷酸。在一些实施

方案中,非标准核苷酸可稳定杂合体形成。在一些实施方案中,非标准核苷酸可去稳定杂合体形成。在一些实施方案中,非标准核苷酸可增强杂交特异性。在一些实施方案中,非标准核苷酸可降低杂交特异性。非标准核苷酸修饰的实例包括2' O-Me、2' O-烯丙基、2' O-炔丙基、2' O-烷基、2' 氟代、2' 阿拉伯糖基、2' 木糖基、2' 氟阿拉伯糖基、硫代磷酸、二硫代磷酸、磷酸胺化物、2' 氨基、5-烷基-取代嘧啶、3' 脱氧鸟苷、5-卤代-取代嘧啶、烷基-取代嘌呤、卤代-取代嘌呤、二环核苷酸、2' MOE、PNA分子、LNA-分子、LNA-样分子、二氨基嘌呤、S2T、5-氟尿嘧啶、5-溴尿嘧啶、5-氯尿嘧啶、5-碘尿嘧啶、次黄嘌呤、黄嘌呤、4-乙酰胞嘧啶、5-(羧基羟甲基)尿嘧啶、5-羧甲基氨基甲基-2-硫代尿苷、5-羧甲基氨基甲基尿嘧啶、二氢尿嘧啶、β-D-半乳糖基queosine、肌苷、N6-异戊烯基腺嘌呤、1-甲基鸟嘌呤、1-甲基肌苷、2,2-二甲基鸟嘌呤、2-甲基腺嘌呤、2-甲基鸟嘌呤、3-甲基胞嘧啶、5-甲基胞嘧啶、N⁶-腺嘌呤、7-甲基鸟嘌呤、5-甲基氨基甲基尿嘧啶、5-甲氧基氨基甲基-2-硫尿嘧啶、β-D-甘露糖基queosine、5'-甲氧基羧甲基尿嘧啶、5-甲氧基尿嘧啶、2-甲硫基-D46-异戊烯基腺嘌呤、尿嘧啶-5-氧基乙酸(v)、wybutoxosine、假尿嘧啶、queosine、2-硫代胞嘧啶、5-甲基-2-硫尿嘧啶、2-硫尿嘧啶、4-硫尿嘧啶、5-甲基尿嘧啶、尿嘧啶-5-氧基乙酸甲酯、尿嘧啶-5-氧基乙酸(v)、5-甲基-2-硫尿嘧啶、3-(3-氨基-3-N-2-羧丙基)尿嘧啶、(acp3)_w、2,6-二氨基嘌呤及其衍生物。

[0165] “一受试者”、“一个体”、“一宿主”或“一患者”是指活生物体,诸如哺乳动物。受试者和宿主的实例包括但不限于马、牛、骆驼、绵羊、猪、山羊、狗、猫、兔、豚鼠、大鼠、小鼠(例如人源化小鼠)、沙鼠、非人灵长类动物(例如猕猴)、人等,非哺乳动物,包括例如,非哺乳类脊椎动物,诸如禽类(例如,鸡或鸭)、鱼(例如鲨鱼)或蛙(例如爪蟾),和非哺乳类无脊椎动物,以及其转基因物种。在某些方面,受试者是指单个生物体(例如人)。在某些方面,提供了一组个体,其组成具有研究的共同免疫因子和/或疾病的小队列,和/或不具有该疾病的个体的队列(例如,阴性/正常对照)。从中获得样品的受试者可患有疾病和/或病症(例如,一种或多种变态反应、感染、癌症或自身免疫疾病等),并可与未患有该疾病的阴性对照受试者相比较。

[0166] “一试剂盒”是指为了实施本文所公开的方法而递送材料或试剂的递送系统。在一些实施方案中,试剂盒包括允许将反应试剂(例如,适当容器中的探针、酶等)和/或支持材料(例如,缓冲液,关于进行分析的书面说明等)储存、运输或从一个位置递送到另一个位置的系统。例如,试剂盒包括一个或多个含有相关反应试剂和/或支持材料的包围体(例如,盒子)。这类内容物可一起或分开递送给预期的接受者。例如,第一容器可含有供在分析中使用的酶,而第二容器含有多个引物。

[0167] 在一些方面,多肽是指包含至少两个氨基酸的分子。在一些实施方案中,多肽由单个肽组成。在一些实施方案中,多肽包含两个或更多个肽。例如,多肽可包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000个肽或氨基酸。多肽的实例包括但不限于氨基酸链、蛋白质、肽、激素、多肽糖、脂质、糖脂、磷脂、抗体、酶、激酶、受体、转录因子和配体。

[0168] 在一些方面,样品是指生物、环境、医学、受试者或患者样品或含有多核苷酸如目标多核苷酸的样品。

样品

[0169] 含有多核苷酸的任何生物样品都可在本文所述的方法中使用。例如,样品可为来

自受试者的含有RNA或DNA的生物样品。可从生物样品中提取多核苷酸,或者可直接对样品进行该方法而无需提取或纯化多核苷酸。样品可为提取或分离的DNA或RNA。样品还可为从生物试样中提取的总RNA或DNA、cDNA文库、病毒或基因组DNA。在一个实施方案中,从含有多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。在某些实施方案中,从单细胞获得多核苷酸。多核苷酸可从生物体直接获得或从由生物体获得的生物样品获得。任何组织或体液试样都可用作供本发明使用的核酸的来源。多核苷酸还可从培养的细胞如原代细胞培养或细胞系中分离。可用病毒或其他细胞内病原体感染从中获得模板核酸的细胞或组织。

[0170] 在某些实施方案中,产生抗体或TCR的免疫细胞可从受试者或宿主的血液或其他生物样品中分离,以鉴别具有潜在临床意义的病原体、肿瘤和/或疾病特异性抗体或TCR,该受试者或宿主例如是人或其他动物,诸如已经免疫或患有感染、癌症、自身免疫病况或任何其他疾病的人或其他动物。例如,该人可被诊断为患有疾病、表现出疾病的症状、未被诊断为患有疾病或未表现出疾病的症状。例如,该人可为暴露于传染原(例如,病毒、细菌、寄生虫、朊病毒等)、抗原或疾病和/或可产生对其有用的抗体或TCR的人。例如,该动物可为暴露于传染原(例如,病毒、细菌、寄生虫、朊病毒等)、抗原或疾病和/或可产生对其有用的抗体或TCR的动物。来自经免疫的宿主的某些免疫细胞产生针对一种或多种所述靶抗原和/或一种或多种未知抗原的抗体或TCR。在本发明中,可通过任何合适的方法,诸如使用荧光激活细胞分选(FACS)、磁性激活细胞分选、淘选或其他筛选方法来筛选并分选细胞,以生成来自样品的多种免疫细胞,诸如免疫细胞文库,从而针对所需免疫细胞对淋巴细胞集合体进行富集,之后对抗体链进行测序、产生抗体,或产生表达文库。与只提供表达不同抗体的免疫细胞的很少子集,并因此只提供很少的天然存在的可变域组合的现有技术富集方法相比,本发明的免疫细胞文库含有表达不同抗体或TCR的单独的免疫细胞或其至少2个子集。例如,本发明的免疫细胞文库可含有表达不同抗体或TCR的单独的免疫细胞或其至少5、10、100、250、500、750、1000、2500、5000、10000、25000、50000、75000、100000、250000、500000、750000、1000000、2500000、5000000、7500000或10000000个子集。本发明的方法使免疫细胞回收最大化,并提供了非常高的多样性。

[0171] 在一些实施方案中,采用来自非免疫的人类或非人类供体的免疫细胞。动物的起始组库(抗原激发之前的组库)向动物提供了能以适度的亲和力(K_A 为约 1×10^{-6} 至 1×10^{-7} M)与基本上任何非自身分子结合的抗体或TCR。抗体或TCR结合位点的序列多样性不在种系中直接编码,而是以组合的方式从V基因区段装配。免疫引发任何免疫细胞产生 V_H-V_L 或 $V\alpha-V\beta$ 或 $V\gamma-V\delta$ 组合,该组合与免疫原结合以增殖(克隆扩充)并分泌如上所述的相应抗体。然而,使用来自未经免疫的受试者的脾细胞和/或免疫细胞或其他外周血淋巴细胞(PBL)可提供对可能的抗体或TCR组库的更好呈现,并且还允许使用任何动物物种构建随后的B细胞或T细胞抗体或TCR文库。

[0172] 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。

[0173] 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)。

[0174] 在一些情况下,起始材料可为组织样品,包括固体组织,其非限制性实例包括脑、肝、肺、肾、前列腺、卵巢、脾、淋巴结(包括扁桃腺)、甲状腺、胰腺、心脏、骨骼肌、肠、喉、食道和胃。在其他情况下,起始材料可为含有核酸的细胞、免疫细胞,并且特别是B细胞或T细胞。在一些情况下,起始材料可为来自可从中获得遗传物质的任何生物体的含有核酸的样品。在一些情况下,样品为流体,例如血液、唾液、淋巴或尿液。

[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,可使用酶处理(如蛋白酶消化)从起始材料中去除非核苷酸物质。

[0177] 在一些情况下,血液可收集到含有镁螯合剂(包括但不限于EDTA)的装置中,并在4°C下储存。任选地,可添加钙螯合剂,包括但不限于EGTA。在另一种情况下,向血液中添加细胞裂解抑制剂,包括但不限于甲醛、甲醛衍生物、福尔马林、戊二醛、戊二醛衍生物、蛋白质交联剂、核酸交联剂、蛋白质和核酸交联剂、伯胺反应性交联剂、巯基反应性交联剂、巯基添加剂或二硫化物还原剂、碳水化合物反应性交联剂、羧基反应性交联剂、光反应性交联剂或可裂解交联剂。

[0178] 在一些情况下,当提取的材料包含单链RNA、双链RNA或DNA-RNA杂合体时,可使用本领域已知的技术将这些分子转化为双链DNA。例如,可使用逆转录酶从RNA分子合成DNA。在一些情况下,RNA到DNA的转化可能需要预先的连接步骤,以将连接体片段与RNA相连接,从而允许使用通用引物来启动逆转录。在其他情况下,mRNA分子的聚-A尾例如可用来启动逆转录。在一些情况下,在转化为DNA后,可使用本文详细描述的方法来进一步捕获、选择、标记或分离所需序列。

[0179] 核酸分子包括脱氧核糖核酸(DNA)和/或核糖核酸(RNA)。核酸分子可为合成的或来源于天然存在的来源。在一个实施方案中,从含有多种其他组分如蛋白质、脂质和非模板核酸的生物样品中分离核酸分子。可从任何细胞材料获得核酸模板分子,该细胞材料从动物、植物、细菌、真菌或任何其他细胞生物体获得。在某些实施方案中,从单细胞获得核酸分子。供本发明使用的生物样品包括病毒颗粒或制品。核酸分子可从生物体直接获得,或从自生物体获得的生物样品获得,例如,从血液、尿液、脑脊液、精液、唾液、痰、粪便和组织获得。任何组织或体液试样都可用作供本发明使用的核酸的来源。核酸分子还可从培养的细胞如原代细胞培养物或细胞系中分离。可用病毒或其他细胞内病原体感染从中获得模板核酸的细胞或组织。

[0180] 样品还可以是从生物试样中提取的总RNA、cDNA文库、病毒或基因组DNA。在某些实施方案中,核酸分子与诸如蛋白质、酶、底物、抗体、结合剂、珠子、小分子、肽或任何其他分子等其他目标分子结合。通常,可通过多种技术从生物样品中提取核酸,诸如Sambrook and Russell, *Molecular Cloning: A Laboratory Manual*, 第三版, Cold Spring Harbor, N.Y. (2001) 描述的技术。核酸分子可为单链的、双链的或具有单链区的双链(例如,茎结构和环

结构)。

[0181] DNA提取方法是本领域公知的。经典的DNA分离方案基于使用有机溶剂如苯酚和氯仿的混合物来提取,然后用乙醇沉淀(J.Sambrook等人,《Molecular Cloning:A Laboratory Manual》,1989,第二版,Cold Spring Harbour Laboratory Press:New York, N.Y.)。其他方法包括:盐析DNA提取(P.Sunnucks等人,Genetics,1996,144:747-756; S.M.Aljanabi等人,Nucl.Acids Res.1997,25:4692-4693)、三甲基溴化铵DNA提取(S.Gustincich等人,BioTechniques,1991,11:298-302)和硫氰酸胍DNA提取(J.B.W.Hammond等人,Biochemistry,1996,240:298-300)。用于从生物样品中提取DNA的多种试剂盒可商购获得(例如,BD Biosciences Clontech(Palo Alto,CA);Epicentre Technologies(Madison,WI);Gentra Systems,Inc.(Minneapolis,MN);MicroProbe Corp.(Bothell,WA);Organon Teknika(Durham,NC);和Qiagen Inc.(Valencia,CA))。

[0182] RNA提取方法也是本领域公知的(例如,J.Sambrook等人,《Molecular Cloning:A Laboratory Manual》1989,211d Ed.,Cold Spring Harbour Laboratory Press:New York)且用于从体液中提取RNA的试剂盒可商购获得(例如,Ambion,Inc.(Austin,TX); Amersham Biosciences(Piscataway,NJ);BD Biosciences Clontech(Palo Alto,CA); BioRad Laboratories(Hercules,CA);DynaL Biotech Inc.(Lake Success,NY); Epicentre Technologies(Madison,WI);Gentra Systems,Inc.(Minneapolis,MN);GIBCO BRL(Gaithersburg,MD);Invitrogen Life Technologies(Carlsbad,CA);MicroProbe Corp.(Bothell,WA);Organon Teknika(Durham,NC);Promega,Inc.(Madison,WI);和 Qiagen Inc.(Valencia,CA))。

[0183] 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。

[0184] 在一些实施方案中,样品为流体,诸如血液、唾液、淋巴、尿液、脑脊液、精液、痰、粪便或组织匀浆。

[0185] 可从具有病况的受试者取得样品。在一些实施方案中,从中取得样品的受试者可为患者,例如,癌症患者或疑似患有癌症的患者。受试者可为哺乳动物,例如人,并可为雄性或雌性。在一些实施方案中,该雌性为妊娠的。该样品可为肿瘤活检物。活检可由例如医疗服务人员进行,包括内科医师、医师助理、护士、兽医、牙医、脊椎推拿师、急救医师、皮肤科医师、肿瘤科医师、胃肠科医师或外科医师。

[0186] 在一些实施方案中,多核苷酸与诸如蛋白质、酶、底物、抗体、结合剂、珠子、小分子、肽或任何其他分子等其他目标分子结合。在一些实施方案中,多核苷酸不与固体支持体结合。可通过多种技术从生物样品中提取核酸(Sambrook等人,《Molecular Cloning:A Laboratory Manual》,第三版,Cold Spring Harbor,N.Y.(2001))。

[0187] 在一些实施方案中,样品为唾液。在一些实施方案中,样品为全血。在一些实施方案中,为了获得足量的多核苷酸以供测试,取得至少0.001、0.005、0.01、0.05、0.1、0.5、1、2、3、4、5、10、20、25、30、35、40、45或50mL的血液体积。在一些实施方案中,血液可收集到含

有镁螯合剂(包括但不限于EDTA)的装置中,并在4℃下储存。任选地,可添加钙螯合剂,包括但不限于EGTA。

[0188] 在一些实施方案中,向血液中添加细胞裂解抑制剂,包括但不限于甲醛、甲醛衍生物、福尔马林、戊二醛、戊二醛衍生物、蛋白质交联剂、核酸交联剂、蛋白质和核酸交联剂、伯胺反应性交联剂、巯基反应性交联剂、巯基添加剂或二硫化物还原剂、碳水化合物反应性交联剂、羧基反应性交联剂、光反应性交联剂或可裂解交联剂。在一些实施方案中,可使用酶处理(诸如蛋白酶消化)从起始材料中去除非核苷酸物质。

[0189] 多个样品可包括至少2、3、4、5、10、20、30、40、50、60、70、80、90或100个或更多样品。多个样品可包括至少约100、200、300、400、500、600、700、800、900或1000个或更多样品。多个样品可包括至少约1000、2000、3000、4000、5000、6000、7000、8000个样品、9000或10,000个样品、或100,000个样品、或1,000,000个或更多样品。多个样品可包括至少约10,000个样品。

[0190] 第一样品中的一个或多个多核苷酸可与第二样品中的一个或多个多核苷酸不同。第一样品中的一个或多个多核苷酸可与多个样品中的一个或多个多核苷酸不同。样品中的一个或多个多核苷酸可包含至少约80%、85%、90%、95%、96%、97%、98%、99%或100%的序列同一性。在一些实施方案中,样品中的一个或多个多核苷酸可以有少于约100、90、80、70、60、50、40、30、25、20、25、10、9、8、7、6、5、4、3、2或1个核苷酸或碱基对不同。多个样品的一个或多个样品中的多个多核苷酸可包含两个或更多个相同序列。多个样品的一个或多个样品中的总多核苷酸的至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%可包含相同的序列。多个样品的一个或多个样品中的多个多核苷酸可包含至少两个不同的序列。多个样品的一个或多个样品中的总多核苷酸的至少约5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%可包含至少两个不同的序列。在一些实施方案中,一个或多个多核苷酸为彼此的变体。例如,一个或多个多核苷酸可含有单核苷酸多态性或其他类型的突变。在另一个实例中,一个或多个多核苷酸为剪接变体。

[0191] 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。

[0192] 所述多个样品可同时获得。多个样品可在相同时间获得。多个样品可顺序获得。多个样品可在获得一个或多个不同样品的数年过程中获得,例如100年、10年、5年、4年、3年、2年或1年。一个或多个样品可在获得一个或多个不同样品的约一年内获得。一个或多个样品可在获得一个或多个不同样品的12个月、11个月、10个月、9个月、8个月、7个月、6个月、4个月、3个月、2个月或1个月内获得。一个或多个样品可在获得一个或多个不同样品的30天、28天、26天、24天、21天、20天、18天、17天、16天、15天、14天、13天、12天、11天、10天、9天、8天、7天、6天、5天、4天、3天、2天或1天内获得。一个或多个样品可在获得一个或多个不同样品的约24小时、22小时、20小时、18小时、16小时、14小时、12小时、10小时、8小时、6小时、4小时、2小时或1小时内获得。一个或多个样品可在获得一个或多个不同样品的约60秒、45秒、30秒、

20秒、10秒、5秒、2秒或1秒内获得。一个或多个样品可在获得一个或多个不同样品的少于一秒内获得。

[0193] 样品的不同多核苷酸可以不同浓度或量存在于样品中(例如,不同的分子数目)。例如,样品中一个多核苷酸的浓度或量可大于另一个多核苷酸的浓度或量。在一些实施方案中,样品中至少一个多核苷酸的浓度或量是样品中至少一个其他多核苷酸的浓度或量的至少约1.5、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000倍或更多倍。在另一个实例中,样品中一个多核苷酸的浓度或量小于另一个多核苷酸的浓度或量。样品中至少一个多核苷酸的浓度或量可比样品中至少一个其他多核苷酸的浓度或量小至少约1.5、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000倍或更多倍。

[0194] 在一些实施方案中,两个或更多个样品可含有不同量或浓度的多核苷酸。在一些实施方案中,一个样品中一个多核苷酸的浓度或量可大于不同样品中相同多核苷酸的浓度或量。例如,血液样品可比尿液样品含有更高的特定多核苷酸的量。或者,单个样品可分为两个或更多个亚样品。亚样品可含有不同量或浓度的相同多核苷酸。一个样品中至少一个多核苷酸的浓度或量可以是另一个样品中相同多核苷酸的浓度或量的至少约1.5、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000倍或更多倍。或者,一个样品中一个多核苷酸的浓度或量可小于不同样品中相同多核苷酸的浓度或量。例如,一个样品中至少一个多核苷酸的浓度或量可比另一个样品中相同多核苷酸的浓度或量小至少约1.5、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000倍或更多倍。

目标多核苷酸

[0195] 在一些情况下,本文提供的方法涉及目标多核苷酸分子(诸如来自细胞的多核苷酸分子)的扩增和测序。在一些情况下,本文提供的方法涉及目标多核苷酸分子的两个或更多个区域的扩增和测序。在一些情况下,本文提供的方法涉及两个或更多个目标多核苷酸分子的扩增和测序。在一个方面,目标多核苷酸为RNA。在一个方面,目标多核苷酸为基因组核酸。来源于特定生物体的染色体中的遗传物质的DNA可为基因组DNA。在优选的实施方案中,目标多核苷酸包括包含由免疫细胞产生的抗体或TCR的可变区的序列。在一些实施方案中,目标多核苷酸包括包含由免疫细胞产生的抗体的重链可变区的序列。在一些实施方案中,目标多核苷酸包括包含由免疫细胞产生的抗体的轻链可变区的序列。在一些实施方案中,目标多核苷酸包括包含由免疫细胞产生的TCR的 α 链可变区的序列。在一些实施方案中,目标多核苷酸包括包含由免疫细胞产生的TCR的 β 链可变区的序列。在一些实施方案中,目标多核苷酸包括包含由免疫细胞产生的TCR的 γ 链可变区的序列。在一些实施方案中,目标多核苷酸包括包含由免疫细胞产生的TCR的 δ 链可变区的序列。

[0196] 目标多核苷酸可从实质上任何来源获得,并可使用本领域已知的方法制备。例如,目标多核苷酸可直接分离,而无需使用本领域已知的方法扩增,该方法包括但不限于从生物体或细胞(例如免疫细胞)中提取基因组DNA或mRNA的片段以获得目标多核苷酸。目标多核苷酸还可包含通过逆转录-PCR从RNA(诸如mRNA)生成的cDNA。在一些情况下,目标多核苷

酸为RNA分子。在一些情况下,目标多核苷酸为mRNA分子,或从mRNA分子产生的cDNA。在一些情况下,目标多核苷酸为来自单个免疫细胞的mRNA分子或从mRNA分子产生的cDNA分子。在一些情况下,目标多核苷酸为来自单独免疫细胞的mRNA分子或从mRNA分子产生的cDNA分子。在一些情况下,目标多核苷酸为编码来自单个免疫细胞的抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的重链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单个免疫细胞的重链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的轻链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的轻链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的抗体可变序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变轻链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变轻链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变重链抗体序列的mRNA分子。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变重链抗体序列的mRNA分子。在一些情况下,目标多核苷酸可为无细胞的核酸,例如DNA或RNA。在一些情况下,目标多核苷酸为编码来自单独免疫细胞的可变 α 、 β 、 γ 和/或 δ 链TCR序列的mRNA分子。

[0197] 本文所述的方法可用来由一个或多个目标多核苷酸生成多核苷酸文库以供测序。目标多核苷酸包括不是扩增反应产物的任何感兴趣的多核苷酸。例如,目标多核苷酸可包括生物样品中的多核苷酸。例如,目标多核苷酸不包括PCR反应的产物。例如,目标多核苷酸可包括用于生成扩增反应产物的多核苷酸模板,但不包括扩增产物本身。例如,目标多核苷酸可包括用于生成逆转录反应或引物延伸反应产物的多核苷酸模板,并且还包括逆转录反应或引物延伸反应产物本身。例如,目标多核苷酸包括可经历逆转录反应或引物延伸反应的感兴趣的多核苷酸。例如,目标多核苷酸包括RNA或DNA。例如,目标多核苷酸包括cDNA。在一些实施方案中,目标RNA多核苷酸为mRNA。在一些实施方案中,目标RNA多核苷酸为聚腺苷酸化的。在一些实施方案中,RNA多核苷酸不是聚腺苷酸化的。在一些实施方案中,目标多核苷酸为DNA多核苷酸。该DNA多核苷酸可为基因组DNA。DNA多核苷酸可包括外显子、内含子、非翻译区或其任意组合。

[0198] 在一些实施方案中,文库可从目标多核苷酸的两个或更多个区域生成。在一些实施方案中,方法文库可从两个或更多个目标多核苷酸生成。在一些实施方案中,目标多核苷酸为基因组核酸或来源于染色体的DNA。在一些实施方案中,目标多核苷酸包括包含诸如多态性或突变等变体的序列。在一些实施方案中,目标多核苷酸包括DNA且不包括RNA。在一些实施方案中,目标多核苷酸包括RNA且不包括DNA。在一些实施方案中,目标多核苷酸包括DNA和RNA。在一些实施方案中,目标多核苷酸为mRNA分子。在一些实施方案中,目标多核苷酸为DNA分子。在一些实施方案中,目标多核苷酸为单链多核苷酸。在一些实施方案中,目标多核苷酸为双链多核苷酸。在一些实施方案中,目标多核苷酸为双链多核苷酸的单链。

[0199] 目标多核苷酸可从任何生物样品中获得,并使用本领域已知的方法制备。在一些实施方案中,目标多核苷酸直接分离而无需扩增。直接分离方法是本领域已知的。非限制性实例包括从生物样品、生物体或细胞中提取基因组DNA或mRNA。

[0200] 在一些实施方案中,一个或多个目标多核苷酸从生物样品中纯化。在一些实施方

案中,目标多核苷酸不从包含它的生物样品中纯化。在一些实施方案中,目标多核苷酸从生物样品中分离。在一些实施方案中,目标多核苷酸不从包含它的生物样品中分离。在一些实施方案中,目标多核苷酸可为无细胞的核酸。在一些实施方案中,目标多核苷酸可为成片段化的核酸。在一些实施方案中,目标多核苷酸可为经转录的核酸。

在一些实施方案中,目标多核苷酸为经修饰的多核苷酸。在一些实施方案中,目标多核苷酸为未经修饰的多核苷酸。

[0201] 在一些实施方案中,目标多核苷酸为来自单细胞的多核苷酸。在一些实施方案中,目标多核苷酸来自单独的细胞。在一些实施方案中,目标多核苷酸为来自含有多个细胞的样品的多核苷酸。

[0202] 在一些实施方案中,目标多核苷酸编码生物标志物序列。在一些实施方案中,目标多核苷酸编码两个或更多个生物标志物序列。在一些实施方案中,多个目标多核苷酸编码生物标志物序列。在一些实施方案中,多个目标多核苷酸编码两个或更多个生物标志物序列。在一些实施方案中,多个目标多核苷酸编码3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100个或更多个生物标志物序列。

[0203] 在一些实施方案中,多个目标多核苷酸包含一组免疫球蛋白序列。在一些实施方案中,多个目标多核苷酸包含一组TCR序列。例如,一组免疫球蛋白序列可为 V_H 和/或 V_L 序列。在一些实施方案中,一组免疫球蛋白或TCR序列含有1、2、3、4、5、6、7、8、9或10个免疫球蛋白或TCR序列。在一些实施方案中,一组免疫球蛋白或TCR序列含有至少约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 或 9×10^{12} 个免疫球蛋白或TCR序列。在一些实施方案中,一组免疫球蛋白或TCR序列含有最多约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 或 9×10^{12} 个免疫球蛋白或TCR序列。在一些实施方案中,一组免

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

[0204] 在一些实施方案中,目标多核苷酸的长度为约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000或20,000个碱基或碱基对。在一些实施方案中,目标多核苷酸的长度为至少约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000或20,000个碱基或碱基对。在一些实施方案中,目标多核苷酸的长度为最多约10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000或20,000个碱基或碱基对。在一些实施方案中,目标多核苷酸的长度为约10-20、10-30、10-40、10-30、10-40、10-50、10-60、10-70、10-80、10-90、10-100、50-60、50-70、50-80、50-90、50-100、100-200、100-300、100-400、100-300、100-400、100-500、100-600、100-700、100-800、100-900、100-1000、500-600、500-700、500-800、500-900、500-1000、1000-2000、1000-3000、1000-4000、1000-3000、1000-4000、1000-5000、1000-6000、1000-7000、1000-8000、1000-9000、1000-10000、5000-6000、5000-7000、5000-8000、5000-9000或5000-10000个碱基或碱基对。在一些实施方案中,目标多核苷酸或其片段的平均长度可为小于约100、200、300、400、500或800个碱基对,或小于约5、10、20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190或200个核苷酸,或小于约1、2、5、10、20、30、40、50、60、70、80、90、100千碱基。在一些实施方案中,来自相对较短的模板(诸如含有目标多核苷酸的样品)的目标序列为约40、45、50、55、60、65、70、75、80、85、90、95或100个碱基。在某些实施方案中,使用含有与疾病或病况相关联的序列或免疫球蛋白或TCR序列的数据库将测序数据与已知或预期的序列进行比对。

B细胞文库遗传物质的克隆和表达

[0205] 在一些方面,如本文所用的一抗体表达文库”或一TCR表达文库”或一表达文库”可指在核酸或蛋白质水平上的分子的集合(即,两个或更多个分子)。因此,在一些实施方案中,该术语可指编码多种抗体或TCR分子的表达载体的集合(即,在核酸水平上),并且/或者可指已在适当的表达系统中表达后的抗体或TCR分子的集合(即,在蛋白质水平上)。或者,表达载体/表达文库可被包含在可表达它们的合适的宿主细胞内。在本发明的表达文库中编码或表达的抗体分子可为任何适当的形式,例如,可为完整的抗体或TCR分子,或者可为抗体或TCR片段,例如,单链抗体(例如,scFv抗体)、Fv抗体、Fab'抗体、(Fab')₂片段、双抗体等。如一编码”特定酶的核酸序列或特定酶的DNA编码序列或一编码”特定酶的核苷酸序列中的术语一编码”,以及其他同义术语,是指当置于适当调节序列的控制下时转录并翻译为酶的DNA序列。一启动子序列”是能够与细胞中的RNA聚合酶结合并启动下游(3'方向)编码序列的转录的DNA调节区。启动子是DNA序列的一部分。此序列区域在其3'末端具有起始密码子。启动子序列包含最小数目的碱基,以及以高于背景的可检测水平启动转录所必需的元件。然而,在RNA聚合酶与序列结合并且转录在起始密码子处(具有启动子的3'末端)启动之后,转录以3'方向向下游进行。在启动子序列内将发现转录起始位点(方便地通过用核酸酶S1作图来定义)以及负责RNA聚合酶结合的蛋白质结合域(共有序列)。

[0206] 由本发明的抗体或TCR表达文库鉴别、衍生、选择或获得的抗体或TCR分子构成本发明的又一方面。同样,这些抗体或TCR分子可为蛋白质或编码抗体或TCR分子的核酸,该核酸进而可并入适当的表达载体中和/或包含在合适的宿主细胞内。

[0207] cDNA集合体可同与抗体基因的重链的恒定区杂交的多核苷酸以及与抗体或TCR基因的V_H或V_α或V_γ链区域的5'端杂交的多核苷酸一起经历PCR反应。cDNA集合体可同与抗体或TCR基因的重链或α或γ链的恒定区杂交的多核苷酸以及与包含抗体或TCR序列的条码化多核苷酸的V_H或V_α或V_γ链区域5'端的5'侧区域杂交的多核苷酸一起经历PCR反应。还可设置PCR反应用于扩增例如κ和λ类别的V_L或V_β或V_δ链集合体。cDNA集合体可同与抗体基因的轻链的恒定区杂交的多核苷酸以及与抗体或TCR基因的V_L或V_β或V_δ链区域的5'端杂交的多核苷酸一起经历PCR反应。cDNA集合体可同与抗体基因的轻链的恒定区杂交的多核苷酸以及与包含抗体或TCR序列的条码化多核苷酸的V_L或V_β或V_δ链区域5'端的5'侧区域杂交的多核苷酸一起经历PCR反应。可基于已知且可公开获得的免疫球蛋白或TCR基因序列数据库信息来设计这类寡核苷酸或引物。

[0208] 在一些实施方案中,可从使用对重链或轻链基因,特别是对V_H和V_L或V_α和V_β或V_γ和V_δ多核苷酸的末端区域之一或两者均无特异性的一个或多个引物通过PCR扩增产生的V_H和V_L或V_α和V_β或V_γ和V_δ序列的文库便利地获得V_H和V_L或V_α和V_β或V_γ和V_δ序列。在一些实施方案中,可从使用对容器条码化多核苷酸的区域具有特异性的引物通过PCR扩增产生的V_H和V_L或V_α和V_β或V_γ和V_δ序列的文库便利地获得V_H和V_L序列。在一些实施方案中,可从使用C-基因家族特异性引物或C-基因特异性引物通过PCR扩增产生的V_H和V_L或V_α和V_β或V_γ和V_δ序列的文库便利地获得V_H和V_L序列。在一些实施方案中,可从使用具有对容器条码化多核苷酸的区域具有特异性的第一引物和作为C-基因家族特异性引物或C-基因特异性引物的第二引物或多个第二引物的引物组通过PCR扩增产生的V_H和V_L或V_α和V_β或V_γ和V_δ序列的文库便利地获得V_H和V_L序列。在一些实施方案中,可从使用具有对容器条码化多核苷酸的区域具有特异性的第一引物和对通用序列具有特异性的第二引物的引物组通过PCR扩增产生的

V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 序列的文库便利地获得 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 序列。

[0209] 在一些实施方案中,一旦逆转录,可使用对免疫球蛋白基因,特别是对 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 多核苷酸的末端区域之一或两者均具有特异性的一个或多个引物,通过PCR来扩增所得到的cDNA序列。在一些实施方案中,可从使用V-基因家族特异性引物或V-基因特异性引物通过PCR扩增产生的 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 序列的文库获得 V_H 和 V_L 序列(Nicholls等人,J.Immunol.Meth.,1993,165:81;W093/12227),或基于可获得的序列信息根据本领域已知的标准方法设计 V_H 和 V_L 序列。(V_H和V_L或V_α和V_β或V_γ和V_δ序列可进行连接,通常采用介于中间的间隔区序列(例如,编码符合读框的柔性肽间隔区),从而形成编码单链抗体的盒)。对于免疫球蛋白表达细胞,V区域序列可便利地克隆为cDNA或PCR扩增产物。任选地,在本文所述的方法中,并且特别是在指出的某些步骤之后(例如,在单细胞PCR后;在哺乳动物或其他细胞表面展示后,在FACS筛选后,等等),对 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 区域进行测序。除其他原因外,测序可用来证实多样性水平为可接受的水平。测序可包括高通量测序、深度测序(其中从多个单独的样品对相同的基因进行测序,以鉴别序列中的差异)或二者的组合。

[0210] 在一些实施方案中,不必要使用本文所述的方法在物理上连接天然的 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 组合。在一些实施方案中,cDNA、条码化多核苷酸或PCR扩增的条码化cDNA未物理连接。在一些实施方案中,cDNA、条码化多核苷酸或PCR扩增的条码化cDNA未在相同的反应或容器中物理连接。

[0211] 在一些实施方案中,除了cDNA引物之外,还使用针对 V_H 或 $V\alpha$ 或 $V\gamma$ 基因的5'端的一个引物或多个引物和针对 V_L 或 $V\beta$ 或 $V\delta$ 基因的5'端的另一个引物或多个引物,对天然的 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 组合进行物理连接。这些引物还含有额外序列的互补尾,以允许 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 基因的自装配。在PCR扩增和连接后,得到混合产物即混合可变区的机会减至最小,因为扩增和连接反应在每个细胞内进行。可通过利用大体积试剂如地高辛标记的核苷酸来进一步确保V区域cDNA对不离开细胞区室并混合,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。

[0212] 在期望提供附加的 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 组合的一些实施方案中,选择表达系统以帮助这一点。例如,噬菌体表达系统允许重链和轻链序列的随机重组。其他合适的表达系统是本领域技术人员已知的。

[0213] 应当注意,在 V_H 和 V_L 或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 序列来源于非人类的情况下,在一些实施方案中,可优选将这些序列与完全人类Fc嵌合。如本文所用的,“嵌合”是指这样的免疫球蛋白或TCR,其中重链和轻链可变区或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 区不是人源的,且其中重链和轻链或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ 链的恒定区是人源的。这通过将可变域扩增并克隆到人类Fc中来实现。该人类Fc可为载体的一部分,或在单独的分子中,并且还可使用Fc的文库。在优选的实施方案中,嵌合分子在哺乳动物细胞如CHO细胞中生长,用FACS筛选两次,以针对表达感兴趣的抗体的细胞来富集该细胞群体。通过测序然后功能性表征或通过直接功能性表征或动力学

来表征嵌合抗体或TCR。生长、筛选和表征详细描述如下。

[0214] 重要的是要注意到,以上所述的PCR反应均针对克隆IgG形式的抗体而描述。这些是优选的,因为它们通常与更成熟的免疫应答相关,并且通常比IgM抗体展现出更高的亲和力,从而使它们更理想地用于某些治疗和诊断应用。然而,显然可设计多核苷酸,如果需要或适当的话,其将允许克隆一种或多种其他形式的免疫球蛋白分子,例如IgM、IgA、IgE和IgD。

[0215] 一旦鉴别了抗体或TCR,并且适当的所述细胞的群体在适当的时间已经分离并任选地如上所述进行富集,则抗体或TCR表达文库不需要立刻生成,这使得包含在细胞内的遗传物质可保持完整,从而使文库能够在以后的日期制备。因此,可通过适当的方法,例如,通过冷冻,将例如细胞、细胞裂解物或核酸,例如由其衍生的RNA或DNA储存到以后的日期,且表达文库在需要时在以后的日期生成。

[0216] 一旦生成了表达载体的文库,则编码的抗体分子可在适当的表达系统中表达,并使用本领域公知和记载的适当技术进行筛选。因此,以上定义的本发明方法可包括在适当的表达系统中表达表达载体的文库,并且针对具有所需性质的抗体筛选表达文库的进一步的步骤,进一步详细解释如下。

[0217] 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。

[0218] 然后初级PCR产物可任选地经历使用新的多核苷酸组的次级PCR反应,该多核苷酸组与抗体或TCR可变域 V_H 、 V_L κ和 V_L λ或 $V_α$ 和 $V_β$ 或 $V_γ$ 和 $V_δ$ 的5'和3'端杂交(视情况取决于使用新多核苷酸组的初级PCR反应是否经过设计以扩增重链或轻链抗体基因或 $V_α$ 或 $V_β$ TCR基因或 $V_γ$ 或 $V_δ$ TCR基因的部分)。这些多核苷酸有利地包含对限定的一组限制酶具有特异性的DNA序列(即,限制酶位点)以供后续克隆。选定的限制酶必须进行选择以便不在人类抗体或TCR V-基因区段内切割。可基于已知且可公开获得的免疫球蛋白或TCR基因序列和限制酶数据库信息来设计这样的多核苷酸。然而,所包括的优选限制酶位点为NcoI、Hind III、MluI和NotI。这样的次级PCR反应的产物为多种V-重、V-轻κ和V-轻λ抗体片段/结构域的组库。因此,这种类型的次级PCR反应通常在以下情况时进行:感兴趣的表达文库形式为scFv或Fv形式,其中只存在抗体或TCR的 V_H 和 V_L 或 $V_α$ 和 $V_β$ 或 $V_γ$ 和 $V_δ$ 结构域。

[0219] PCR产物还可经历采用与条码化多核苷酸的5'和3'端杂交的新引物组的PCR反应。这些多核苷酸可有利地包含对限定的一组限制酶具有特异性的DNA序列(即,限制酶位点)以供后续克隆。选定的限制酶必须进行选择以便不在人类抗体或TCR V-基因区段内切割。可基于已知且可公开获得的免疫球蛋白或TCR基因序列和限制酶数据库信息来设计这样的多核苷酸。然而,所包括的优选限制酶位点为NcoI、Hind III、MluI和NotI。这样的次级PCR

反应的产物为各种 V_H 、 $V_L\kappa$ 和 $V_L\lambda$ 抗体片段/结构域或 $V\alpha$ 和 $V\beta$ 或 $V\gamma$ 和 $V\delta$ TCR片段/结构域的组库。

[0220] 本领域技术人员将会认识到,该系统还可使用重链或轻链或 $V\alpha$ 或 $V\beta$ 链或 $V\gamma$ 或 $V\delta$ 链Fv或Fab片段或者单链抗体或TCR。重链或轻链或 $V\alpha$ 或 $V\beta$ 链或 $V\gamma$ 或 $V\delta$ 链可经诱变,然后向溶液中添加互补链。然后使两条链组合并形成功能性抗体片段。随机非特异性轻链或重链或 $V\alpha$ 或 $V\beta$ 链或 $V\gamma$ 或 $V\delta$ 链序列的添加允许产生组合系统以生成多样性成员的文库。

[0221] 如本文所定义的这类克隆片段组库的文库构成了本发明的其他方面,所述片段包含由免疫激发的宿主的B或T淋巴细胞衍生的抗体或TCR基因的可变重链或 $V\alpha$ 链或 $V\gamma$ 链区域或其片段,和/或可变轻链或 $V\beta$ 链或 $V\delta$ 链区域或其片段。包含克隆的可变区的这些文库可任选地插入表达载体中以形成表达文库。

[0222] 在一些实施方案中,可设置PCR反应以便保持分离的免疫细胞群体中含有的各种抗体或TCR链的全部或部分恒定区。这在以下情况时是理想的:表达文库形式为Fab形式,其中重链或 α 或 γ 链组分包含 V_H 或 $V\alpha$ 或 $V\gamma$ 和 C_H 或 $C\alpha$ 或 $C\gamma$ 结构域,且轻链或 $V\beta$ 链或 $V\delta$ 链组分包含 V_L 或 $V\beta$ 或 $V\delta$ 链和 C_L 或 $C\beta$ 或 $C\delta$ 结构域。同样,包含抗体或TCR链的所有或部分恒定区的这类克隆片段的文库构成了本发明的进一步的方面。

[0223] 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。

[0224] 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)。

[0225] 本文所公开的文库可用于多种应用。如本文所用的,文库包含多个分子。在一些实施方案中,文库包含多个多核苷酸。在一些实施方案中,文库包含多个引物。在一些实施方案中,文库包含来自一个或多个多核苷酸、扩增子或扩增子集的多个序列读取。文库可多次储存并多次使用以生成用于分析的样品。一些应用包括,例如,对多态性进行基因型分型,研究RNA加工,和选择克隆代表来根据本文所提供的方法进行测序。可以生成包含多个多核苷酸的文库,诸如用于测序或扩增的引物或文库,其中多个多核苷酸包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、50,000,000、100,000,000个或更多个分子条码或容器条码。在一些实施方案中,多核苷酸的文库包含多个至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、

500,000、600,000、700,000、800,000、900,000、1,000,000、50,000,000、100,000,000个或更多个独特的多核苷酸,其中每个独特的多核苷酸包含一个或多个分子条码和容器条码。

条码

[0226] 条码可为分子条码或容器条码。在一些实施方案中,条码,诸如分子条码或容器条码,可各自具有2至36个核苷酸、4至36个核苷酸、或6至30个核苷酸、或8至20个核苷酸、2至20个核苷酸、4至20个核苷酸、或6至20个核苷酸范围内的长度。在某些方面,组内条码的解链温度彼此在10°C以内、彼此在5°C以内或彼此在2°C以内。在某些方面,组内条码的解链温度彼此不在10°C以内、彼此不在5°C以内或彼此不在2°C以内。在其他方面,条码是最低交叉杂交组的成员。例如,这样的组的每个成员的核苷酸序列可与该组的每个其他成员的核苷酸序列充分不同,使得在严格杂交条件下,没有成员可与任何其他成员的互补体形成稳定的双链体。在一些实施方案中,最低交叉杂交组的每个成员的核苷酸序列与每个其他成员的核苷酸序列有至少两个核苷酸不同。条码技术描述于Winzeler等人(1999) *Science* 285:901; Brenner(2000) *Genome Biol.* 1:1; Kumar等人(2001) *Nature Rev.* 2:302; Giaever等人(2004) *Proc. Natl. Acad. Sci. USA* 101:793; Eason等人(2004) *Proc. Natl. Acad. Sci. USA* 101:11046; 以及Brenner(2004) *Genome Biol.* 5:240。

[0227] 如本文所用的,分子条码包含对于来自单细胞或来自单个容器的单分子、或来自两个或更多个单细胞或来自两个或更多个单容器的多个分子或分子文库的两个或更多个分子独特的信息。如本文所用的,与来自不同单细胞或来自不同单个容器的多核苷酸相比,容器条码包含对于来自单细胞或单个容器的多核苷酸独特的信息。在一些实施方案中,独特的信息包含核苷酸的独特序列。例如,分子条码或容器条码的序列可通过确定含有该分子条码或容器条码的核苷酸的独特或随机序列的身份和顺序来确定。在一些实施方案中,独特的信息不能用于鉴别目标多核苷酸的序列。例如,分子条码可与一个目标多核苷酸相附接,但该分子条码不能用于确定其附接的目标多核苷酸。在一些实施方案中,独特的信息不是与目标多核苷酸的序列的身份相关联的已知序列。例如,容器条码可与一个或多个目标多核苷酸相附接,但该容器条码不能用于确定其附接的是该一个或多个目标多核苷酸中的哪一个。在一些实施方案中,独特的信息包含核苷酸的随机序列。在一些实施方案中,独特的信息包含多核苷酸上核苷酸的一个或多个独特序列。在一些实施方案中,独特的信息包含简并核苷酸序列或简并条码。简并条码可包含可变核苷酸碱基组成或序列。例如,简并条码可为随机序列。在一些实施方案中,分子条码或容器条码的互补体序列也是分子条码或容器条码序列。

[0228] 分子条码或容器条码可包含任何长度的核苷酸。例如,分子条码或容器条码可包含至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸。例如,分子条码或容器条码可包含最多约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸。在一些实施方案中,分子条码或容器条码具有特定的核苷酸长度。例如,分子条码或容器条码的长度可为约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、

46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸。

[0229] 在一些实施方案中,多个分子条码或容器条码中的每个分子条码或容器条码具有至少约2个核苷酸。例如,多个分子条码或容器条码中的每个分子条码或容器条码的长度可为至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸。在一些实施方案中,多个分子条码或容器条码中的每个分子条码或容器条码具有最多约1000个核苷酸。例如,多个分子条码或容器条码中的每个分子条码或容器条码的长度可为最多约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸。在一些实施方案中,多个分子条码或容器条码中的每个分子条码或容器条码具有相同的核苷酸长度。例如,多个分子条码或容器条码中的每个分子条码或容器条码的长度可为2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸。在一些实施方案中,多个分子条码或容器条码中的一个或多个分子条码或容器条码具有不同的核苷酸长度。例如,多个分子条码或容器条码中的一个或多个第一分子条码或容器条码可具有约或至少约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸,且多个分子条码或容器条码中的一个或多个第二分子条码或容器条码可具有约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、60、70、80、90、100、200、500或1000个核苷酸,其中所述一个或多个第一分子条码或容器条码与所述一个或多个第二分子条码或容器条码的多核苷酸的数目不同。

[0230] 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。

[0231] 不同分子条码的数目可超过将在多个容器中标记的分子的总数目。在一些实施方案中,不同分子条码的数目是将在多个容器中标记的分子的总数目的至少约1、1.5、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0232] 单个容器中不同分子条码的数目可超过将在单个容器中标记的不同分子的数目。在一些实施方案中,单个容器中不同分子条码的数目是将在单个容器中标记的不同分子的数目的至少约1、1.5、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0233] 不同容器条码的数目可小于将在多个容器中标记的分子的总数目。在一些实施方案中,不同容器条码的数目比将在多个容器中标记的分子的总数目小至少约1、1.5、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0234] 来自单个容器中容器条码化多核苷酸分子的扩增产物分子的数目可超过将在单

个容器中标记的不同分子的数目。在一些实施方案中,来自单个容器中容器条码化多核苷酸分子的扩增产物分子的数目是将在单个容器中标记的不同分子的数目的至少约1、1.5、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0235] 单个容器中容器条码化多核苷酸分子的数目可小于将在单个容器中标记的不同分子的数目。在一些实施方案中,单个容器中容器条码化多核苷酸分子的数目比将在单个容器中标记的不同分子的数目小至少约1、1.5、2、2.5、3、3.5、4、4.5、5、6、7、8、9、10、15、20、30、40、50、60、70、80、90或100倍。

[0236] 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。

[0237] 在一些实施方案中,至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%的不同分子条码具有相同的浓度。在一些实施方案中,至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%的不同容器条码具有相同的浓度。

[0238] 在一些实施方案中,至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%的不同分子条码具有不同的浓度。在一些实施方案中,至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%或100%的不同容器条码具有不同的浓度。

[0239] 分子条码或容器条码群体中的分子条码或容器条码可具有至少10、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000个或更多个不同的序列。例如,群体中的分子条码或容器条码可具有至少2,000、3,000、4,000、5,000、6,000、7,000、8,000、9,000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000个或更多个不同的序列。因此,多个分子条码或容器条码可用于从一个或多个多核苷酸如目标多核苷酸生成至少10、15、20、25、30、35、40、45、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000个或更多个不同的序列。例如,多个分子条码或容器条码可用于从一个或多个多核苷酸如目标多核苷酸生成至少2,000、3,000、4,000、5,000、6,000、7,000、8,000、9,000、10,000、15,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、 1×10^6 、 2×10^6 、 3×10^6 、 4×10^6 、 5×10^6 、 6×10^6 、 7×10^6 、 8×10^6 、 9×10^6 、 1×10^7 、 2×10^7 、 3×10^7 、 4×10^7 、 5×10^7 、 6×10^7 、 7×10^7 、 8×10^7 、 9×10^7 、 1×10^8 、 2×10^8 、 3×10^8 、 4×10^8 、 5×10^8 、 6×10^8 、 7×10^8 、 8×10^8 、 9×10^8 、 1×10^9 、 2×10^9 、 3×10^9 、 4×10^9 、 5×10^9 、 6×10^9 、 7×10^9 、 8×10^9 、 9×10^9 、 1×10^{10} 、 2×10^{10} 、 3×10^{10} 、 4×10^{10} 、 5×10^{10} 、 6×10^{10} 、 7×10^{10} 、 8×10^{10} 、 9×10^{10} 、 1×10^{11} 、 2×10^{11} 、 3×10^{11} 、 4×10^{11} 、 5×10^{11} 、 6×10^{11} 、 7×10^{11} 、 8×10^{11} 、 9×10^{11} 、 1×10^{12} 、 2×10^{12} 、 3×10^{12} 、 4×10^{12} 、 5×10^{12} 、 6×10^{12} 、 7×10^{12} 、 8×10^{12} 、 9×10^{12} 个或更多个不同的序列。例如,多个分子条码或容器条码可用于从至少约10、15、20、25、30、

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

[0240] 在一些实施方案中，一个或多个分子条码用于将序列分组或分箱。在一些实施方案中，一个或多个分子条码用于将序列分组或分箱，其中每个箱元中的序列含有相同的分子条码。在一些实施方案中，一个或多个分子条码或容器条码用于将序列分组或分箱，其中每个箱元中的序列包含扩增子集。在一些实施方案中，一个或多个分子条码用于将序列分组或分箱，其中每个箱元中的序列包含多个序列，其中从中生成该多个序列的多核苷酸来源于扩增反应中的相同多核苷酸分子。

[0241] 在一些实施方案中，一个或多个容器条码用于将序列分组或分箱。在一些实施方案中，一个或多个容器条码用于将序列分组或分箱，其中每个箱元中的序列含有相同的容器条码。在一些实施方案中，一个或多个容器条码用于将序列分组或分箱，其中每个箱元中的序列包含一个或多个扩增子集。在一些实施方案中，一个或多个容器条码用于将序列分组或分箱，其中每个箱元中的序列包含多个序列，其中从中生成该多个序列的多核苷酸来源于来自单个容器或单细胞的多核苷酸。

[0242] 在一些实施方案中，一个或多个分子条码和容器条码用于将序列分组或分箱。在一些实施方案中，一个或多个分子条码和容器条码用于将序列分组或分箱，其中每个箱元中的序列含有相同的分子条码和相同的容器条码。在一些实施方案中，一个或多个分子条码和容器条码用于将序列分组或分箱，其中每个箱元中的序列包含一个或多个扩增子集。在一些实施方案中，一个或多个分子条码和容器条码用于将序列分组或分箱，其中每个箱元中的序列包含多个序列，其中从中生成该多个序列的多核苷酸来源于扩增反应中的相同多核苷酸且来自相同的单细胞或容器。在一些实施方案中，一个或多个分子条码和容器条码不用于比对序列。

[0243] 在一些实施方案中,一个或多个分子条码不用于比对序列。在一些实施方案中,一个或多个分子条码用于比对序列。在一些实施方案中,一个或多个分子条码用于将序列分组或分箱,且目标特异性区域用于比对序列。在一些实施方案中,一个或多个容器条码不用于比对序列。在一些实施方案中,一个或多个容器条码用于比对序列。在一些实施方案中,一个或多个容器条码用于将序列分组或分箱,且目标特异性区域用于比对序列。在一些实施方案中,一个或多个分子条码和容器条码用于比对序列。在一些实施方案中,一个或多个分子条码和容器条码用于将序列分组或分箱,且目标特异性区域用于比对序列。

[0244] 在一些实施方案中,比对的序列含有相同的分子条码。在一些实施方案中,比对的序列含有相同的容器条码。在一些实施方案中,比对的序列含有相同的分子条码和容器条码。在一些实施方案中,一个或多个分子条码或容器条码用于比对序列,其中比对的序列包含来自扩增子集的两个或更多个序列。在一些实施方案中,一个或多个分子条码或容器条码用于比对序列,其中比对的序列包含多个序列,其中从中生成该多个序列的多核苷酸来源于扩增反应中的相同多核苷酸分子。在一些实施方案中,一个或多个分子条码或容器条码用于比对序列,其中比对的序列包含多个序列,其中从中生成该多个序列的多核苷酸来源于单细胞或单个容器。

小液滴生成

[0245] 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。

[0246] 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%。

[0247] 将免疫细胞的样品分成小反应体积,结合来自或来源于来自所述多个免疫细胞的单个免疫细胞的多核苷酸的分子和容器条码化,可使得能够对重链和轻链序列的组库进行高通量测序。这些方法还可允许在基于条码化序列的测序后将重链和轻链配对。如本文所述将样品分成小反应体积还可使得能够使用减少的试剂量,从而降低该分析的材料成本。

[0248] 在一些情况下,逆转录反应和/或扩增反应(例如,PCR)在小液滴中进行,诸如在小液滴数字PCR中。在某些方面,本发明提供了流体区室以包含所有或部分的目标材料。在一些实施方案中,区室为小液滴。当在本说明书各处提及一小液滴”时,除非另有说明,该术语与流体区室和流体分区可互换使用。除非另有说明,为了方便起见使用一小液滴”,并且可使用任何流体分区或区室。本文所使用的小液滴可包括乳液组合物(或两种或多种不混溶液体的混合物),诸如美国专利号7,622,280中所述。可通过W0/2010/036352中所述的装置生成小液滴。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液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液滴。本文所提供的乳液可为水包油乳液,其中小液滴为连续水相中的油小液滴。本文所提供的

小液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。

[0249] 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。当小的小液滴合并以形成逐渐更大的小液滴时发生聚结。在一些情况下,由小液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%的小液滴与其他小液滴聚结。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。

[0250] 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液滴。小液滴可具有约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液滴的微流体方法已知用于产生单分散或多分散乳液。该小液滴可为单分散小液滴。可生成小液滴以使得小液滴的大小变化不超过小液滴平均大小的加或减5%。在一些情况下,生成小液滴以使得小液滴的大小变化不超过小液滴平均大小的加或减2%。小液滴生成器可从单个样品生成小液滴群体,其中小液滴的大小变化都不超过小液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%。

[0251] 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液滴或胶囊对于标准移液管操作和离心可以是机械稳定的。

[0252] 可通过使油相流动穿过水性样品来形成小液滴。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。

[0253] 水相可包含具有或不具有固体表面如珠子的用于进行扩增反应的缓冲溶液和试剂。缓冲溶液可包含约、多于约或少于约1、5、10、15、20、30、50、100或200mM Tris。在一些情况下,氯化钾的浓度可为约、大于约或小于约10、20、30、40、50、60、80、100或200mM。缓冲溶液可包含约15mM Tris和50mM KCl。核苷酸可包括脱氧核苷三磷酸分子,包括dATP、dCTP、dGTP和dTTP,每种的浓度为约、大于约或小于约50、100、200、300、400、500、600或700 μ m。在一些情况下,dUTP在水相中添加到约、大于约或小于约50、100、200、300、400、500、600或700、800、900或1000 μ m的浓度。在一些情况下,将氯化镁(MgCl₂)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。MgCl₂的浓度可为约3.2mM。在一些情况下,使用乙酸镁或镁。在一些情况下,使用硫酸镁。

[0254] 可使用诸如BSA或来自牛皮肤的明胶的非特异性封闭剂,其中明胶或BSA的浓度范围为约0.1-0.9%w/v。其他可能的封闭剂可包括 β 乳球蛋白、酪蛋白、奶粉或其他常见的封闭剂。在一些情况下,BSA和明胶的优选浓度为约0.1%w/v。

[0255] 水相中用于扩增的引物可具有约、大于约或小于约0.05、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1.0、1.2、1.5、1.7或2.0 μ m的浓度。水相中的引物浓度可为约0.05至约2、约0.1至约1.0、约0.2至约1.0、约0.3至约1.0、约0.4至约1.0或约0.5至约1.0 μ m。引物的浓

度可为约0.5 μm 。在PCR中对于目标核酸浓度适用的范围包括但不限于约1pg至约500ng。

[0256] 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。

[0257] 在一些情况下,非离子型环氧乙烷/环氧丙烷嵌段共聚物可以约0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%或1.0%的浓度添加到水相。常见的生物表面活性剂包括非离子型表面活性剂,诸如Pluronic F-68、Tetronics和Zonyl FSN。Pluronic F-68可以约0.5%w/v的浓度存在。

[0258] 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。

[0259] 可配制乳液以产生高度单分散的具有液体样界面膜的小液滴,该小液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 $^{\circ}\text{C}$ 、60 $^{\circ}\text{C}$ 、70 $^{\circ}\text{C}$ 、80 $^{\circ}\text{C}$ 、90 $^{\circ}\text{C}$ 或95 $^{\circ}\text{C}$ 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 $^{\circ}\text{C}$ 、4 $^{\circ}\text{C}$ 、5 $^{\circ}\text{C}$ 、6 $^{\circ}\text{C}$ 、7 $^{\circ}\text{C}$ 、8 $^{\circ}\text{C}$ 、9 $^{\circ}\text{C}$ 、10 $^{\circ}\text{C}$ 、15 $^{\circ}\text{C}$ 、20 $^{\circ}\text{C}$ 、25 $^{\circ}\text{C}$ 、30 $^{\circ}\text{C}$ 、35 $^{\circ}\text{C}$ 或40 $^{\circ}\text{C}$ 下。这些胶囊可在生物医学应用中有用,诸如稳定、数字化封装大分子,特别是含有核酸或蛋白质或二者的混合物的水性生物流体;药物和疫苗递送;生物分子文库;临床成像应用和其他。

[0260] 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。

[0261] 在一些情况下,通过进行数字PCR,诸如基于微流体的数字PCR或小液滴数字PCR,进行扩增步骤。

[0262] 可使用微流体系统或装置生成小液滴。如本文所用的,“一微”前缀(例如,“一微通道”或“一微流体”)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,如本文所用的一“微流体”是指包含至少一个微型通道的设备、装置或系统。

[0263] 微流体系统和装置已经在多个情境下描述,典型地是在微型化实验室(例如,临床)分析的情境下。其他用途也已经描述。例如,国际专利申请公开号WO 01/89788;WO 2006/040551;WO 2006/040554;WO 2004/002627;WO 2008/063227;WO 2004/091763;WO 2005/021151;WO 2006/096571;WO 2007/089541;WO 2007/081385和WO 2008/063227。

[0264] 小液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液滴的任何技术。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液滴。

[0265] 载流可为与样品流体不混溶的任何流体。示例性的载流为油。在某些实施方案中,载流包含表面活性剂。

[0266] 相同的方法可用于创建含有其他试剂的单个小液滴,该其他试剂例如是用于扩增反应如聚合酶链反应(PCR)、或非基于PCR的扩增反应如多链置换扩增、或本领域普通技术人员已知的其他方法的试剂。用于进行基于PCR扩增反应的合适的试剂为本领域普通技术人员已知的,且包括但不限于DNA聚合酶、正向和反向引物、脱氧核苷三磷酸(dNTP)和一种或多种缓冲液。

[0267] 在某些实施方案中,通过提供包含目标材料(例如,免疫细胞和/或固体支持体,如珠子)的第一流体分区(例如,小液滴)和第二流体(例如,作为液流或位于小液滴内)形成流体隔室。合并第一和第二流体以形成小液滴。可通过将电场施加于两种流体来完成合并。在某些实施方案中,第二流体含有用于进行诸如聚合酶链反应或扩增反应的扩增反应的试剂。

[0268] 在某些方面,本发明提供了制备独特条码化重链和轻链抗体序列和/或 α 和 β 链TCR序列和/或 γ 和 δ 链TCR序列的文库的方法,包括获得多个核酸构建体,其中每个构建体包含独特的N-聚体和功能性N-聚体。功能性N-聚体可为随机N-聚体、PCR引物、通用引物、抗体、粘性末端或任何其他序列。该方法可包括制备M组N个流体隔室,每个隔室含有独特构建体的一个或多个拷贝。该方法可通过向组中的每个隔室添加另外的构建体来创建具有更高复杂性的条码文库,并对每个组重复此步骤,以产生 $N \times M$ 个隔室,每个含有独特的一对构建体。可杂交或连接该对,以产生新构建体。在条码文库中的每个构建体中,每种独特的N-聚体可适用于通过测序、探针杂交、其他方法或方法的组合来鉴别。

小液滴文库

[0269] 通常,小液滴文库由在单个集合中合并在一起的许多文库元件组成。文库的复杂性可从单个文库元件到 1×10^{15} 个文库元件或更多不等。每个文库元件是固定浓度的一种或多种给定的组分。该元件可为但不限于细胞、珠子、氨基酸、蛋白质、多肽、核酸、多核苷酸或小分子化合物。该元件可含有标识符,诸如分子条码或容器条码或二者。

[0270] 细胞文库元件可包括但不限于杂交瘤、B细胞、T细胞、原代细胞、培养的细胞系、癌症细胞、干细胞或任何其他细胞类型。通过将一千到上万个数目的细胞封装在单个小液滴中来制备细胞文库元件。封装细胞的数目通常由来自细胞的数量密度和小液滴体积的Poisson统计给出。然而,在一些情况下,该数目偏离Poisson统计,如Edd等人,“Controlled encapsulation of single-cells into monodisperse picolitre drops.” Lab Chip, 8(8):1262-1264, 2008中所述。细胞的离散性质允许用多个细胞变体诸如各自产

生一种抗体或TCR的免疫细胞大量制备文库,所有都存在于单一的起始介质中,然后该介质分解为含有最多一个细胞的单个小液滴胶囊。然后裂解单个小液滴胶囊中的细胞,用分子条码和容器条码对来自裂解细胞的重链和轻链多核苷酸和/或 α 和 β 链多核苷酸和/或 γ 和 δ 链多核苷酸进行条码化,并扩增,然后组合或合并以形成由重链和轻链和/或 α 和 β 链和/或 γ 和 δ 链文库元件组成的文库。

[0271] 基于珠子的文库元件含有一个或多个珠子,并且还可含有其他试剂,诸如抗体、酶或其他蛋白质。在所有文库元件都含有不同类型的珠子但含有相同的周围介质的情况下,所有文库元件都可从单一起始流体制备,或具有多个起始流体。在从变体的集合大量制备细胞文库的情况下,将从多个起始流体制备文库元件。当以多个细胞开始时,期望每滴恰好具有一个细胞,只有少量小液滴含有超过一个细胞。在一些情况下,可达到相对于Poisson统计的变化以提供增强的小液滴装载,以便更多小液滴恰好每个小液滴具有一个细胞,并且空小液滴或含有超过一个细胞的小液滴的例外极少。

[0272] 在一些实施方案中,当从多个容器条码化多核苷酸开始时,期望每滴恰好具有一个容器条码化多核苷酸,只有少量小液滴含有超过一个容器条码化多核苷酸。在一些情况下,可达到相对于Poisson统计的变化以提供增强的小液滴装载,以便更多小液滴恰好每个小液滴具有一个容器条码化多核苷酸,并且空小液滴或含有超过一个容器条码化多核苷酸的小液滴的例外极少。

[0273] 小液滴文库的实例是具有不同内容物的小液滴的集合,该内容物的范围包括珠子、细胞、小分子、DNA、引物、抗体和分子条码。小液滴的直径大小范围从大致0.5微米到500微米,其相当于约1皮升到1纳升。然而,小液滴可小到5微米且大到500微米。优选地,小液滴的直径为小于100微米,约1微米至约100微米。最优的直径大小为约20至40微米(10至100皮升)。对小液滴文库检查的优选性质包括渗透压平衡、均匀大小和大小范围。

[0274] 由本发明提供的小液滴文库中包含的小液滴优选大小均匀。也就是说,当与相同文库中其他小液滴的直径相比时,文库中任何小液滴的直径的变化将小于5%、4%、3%、2%、1%或0.5%。文库中小液滴的均匀大小对维持小液滴的稳定性和完整性可能非常关键,并且可能还对文库中的小液滴随后用于本文所述的各种生物和化学分析非常重要。

[0275] 本发明提供了包含不混溶流体内的多个水性小液滴的小液滴文库,其中每个小液滴优选是基本上大小均匀的,且包含不同的文库元件。本发明提供了用于形成小液滴文库的方法,其包括提供包含不同文库元件的单个水性流体,将每个文库元件封装进不混溶流体中的水性小液滴内。

[0276] 在某些实施方案中,将不同类型的元件(例如,细胞或珠子)合并到相同介质中含有的单个来源中。在初始合并后,将元件封装在小液滴中,以形成小液滴的文库,其中具有不同类型珠子或细胞的每个小液滴均为不同的文库元件。对初始溶液的稀释允许封装过程。在一些实施方案中,形成的小液滴将含有单个元件或将不含任何物体,即为空的。在其他实施方案中,形成的小液滴将含有文库元件的多个拷贝。封装的元件通常为某种类型的变体。在一个实例中,元件为血液样品的免疫细胞,且封装每个免疫细胞以对免疫细胞中核苷酸的抗体序列进行扩增并条码化。

[0277] 例如,在一种类型的乳液文库中,存在的文库元件具有不同的粒子,即不同介质中的细胞或条码化多核苷酸,并且在合并之前封装。在一个实例中,在不同介质中包含指定数

目的文库元件,即n个数目的不同细胞或条码化多核苷酸。每个文库元件分开乳化并合并,在该点处,n个合并的不同文库元件中的每一个组合并合并成单个集合体。产生的集合体含有多个油包水乳液小液滴,每个含有不同类型的粒子。

[0278] 在一些实施方案中,形成的小液滴将含有单个文库元件或将不含任何物体,即为空的。在其他实施方案中,形成的小液滴将含有文库元件的多个拷贝。珠子的内容物遵循Poisson分布,其中存在离散概率分布,其表达了如果若干事件以已知的平均率发生,且独立于自上一次事件以来的时间,则这些事件在固定的时间段内发生的概率。用于创建文库的油和表面活性剂防止了小液滴之间文库内容物的交换。

引物

[0279] 通常,在扩增反应中可使用一对或多对引物;引物对的一个引物可为正向引物且引物对的一个引物可为反向引物。

[0280] 在一些情况下,在扩增反应中可使用第一对引物;第一对的一个引物可为与第一目标多核苷酸分子的序列互补的正向引物,且第一对的一个引物可为可与第一目标多核苷酸分子的第二序列互补的反向引物,且第一目标基因座可位于第一序列与第二序列之间。在一些实施方案中,第一目标基因座包含 V_H 或 V_α 或 V_γ 序列。

[0281] 在一些情况下,在扩增反应中可使用第二对引物;第二对的一个引物可为与第二目标多核苷酸分子的第一序列互补的正向引物,且第二对的一个引物可为与第二目标多核苷酸分子的第二序列互补的反向引物,且第二目标基因座可位于第一序列与第二序列之间。在一些实施方案中,第二目标基因座包含 V_L 或 V_B 或 V_δ 序列。

[0282] 在一些情况下,在扩增反应中可使用第三对引物;第三对的一个引物可为与第三目标多核苷酸分子的第一序列互补的正向引物,且第三对的一个引物可为与第三目标多核苷酸分子的第二序列互补的反向引物,且第三目标基因座可位于第一序列与第二序列之间。在一些实施方案中,第三目标基因座包含条码,诸如分子条码或容器条码。

[0283] 正向引物和反向引物的长度可取决于目标多核苷酸的序列和目标基因座。例如,可优化正向引物和反向引物的长度和/或 T_m 。在一些情况下,引物的长度可为约、大于约或小于约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59或60个核苷酸。在一些情况下,引物的长度为约15至约20、约15至约25、约15至约30、约15至约40、约15至约45、约15至约50、约15至约55、约15至约60、约20至约25、约20至约30、约20至约35、约20至约40、约20至约45、约20至约50、约20至约55或约20至约60个核苷酸。

[0284] 在与模板多核苷酸结合之前,引物可为单链DNA。在一些情况下,引物最初含有双链序列。引物的适当长度可取决于引物的预期用途,但范围可为约6至约50个核苷酸,或约15至约35个核苷酸。短引物分子通常可需要更冷的温度以与模板形成足够稳定的杂合复合体。在一些实施方案中,引物不需要反映模板核酸的准确序列,但可足够互补以与模板杂交。在一些情况下,在与模板多核苷酸结合前,引物可为部分双链的。具有双链序列的引物可具有约、大于约或小于约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基的发夹环。引物的双链部分可为约、大于约或小于约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50个碱基对。用于扩增给定目标序列的合适引物的设计是

本领域公知的。

[0285] 引物可包含附加特征,该特征允许检测或固定引物,但不改变引物的基本性质(例如,作为DNA合成的起始点)。例如,引物可在5'端含有附加核酸序列,该序列不与目标核酸杂交,但其促进克隆或进一步的扩增,或对扩增产物的测序。例如,该附加序列可包含引物结合位点,诸如通用引物结合位点。本文可将与模板足够互补以杂交的引物区域称作杂交区域。

[0286] 在另一种情况下,本文所述的方法和组合物中所采用的引物可包含一个或多个通用核苷。通用核苷的非限制性实例为5-硝基吡啶和肌苷,如美国申请公开号2009/0325169和2010/0167353中所述。

[0287] 可根据已知参数设计引物,以便避免二级结构和自身杂交。不同的引物对可在大约相同的温度下退火和解链,例如,在另一个引物对的1°C、2°C、3°C、4°C、5°C、6°C、7°C、8°C、9°C或10°C内。在一些情况下,最初使用多于1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、100、200、500、1000、5000、10,000个或更多个引物。这样的引物可与本文所述的目标多核苷酸杂交。

[0288] 可通过多种方法制备引物,包括但不限于使用本领域公知的方法克隆适当的序列和直接化学合成(Narang等人,Methods Enzymol.68:90(1979);Brown等人,Methods Enzymol.68:109(1979))。还可从商业来源获得引物。引物可具有相同的解链温度。引物可具有不相同的解链温度。引物的长度可在5'端或3'端延伸或缩短,以产生具有所需解链温度的引物。引物的一个引物可比另一个引物长。引物对中引物的3'退火长度可不同。同样,可设计引物对的退火位置,以使得引物对的序列和长度产生所需的解链温度。用于确定小于25个碱基对的引物的解链温度的等式为Wallace规则($T_M = 2(A+T) + 4(G+C)$)。还可使用计算机程序设计引物。可使用软件程序计算每个引物的 T_M (解链或退火温度)。在任何扩增循环(包括但不限于循环1、2、3、4、5、循环6-10、循环10-15、循环15-20、循环20-25、循环25-30、循环30-35或循环35-40)后,引物的退火温度可重新计算并提高。在扩增的初始循环后,引物的5'一半可从每个感兴趣的基因座并入产物中;因此可基于每个引物的5'一半和3'一半两者的序列重新计算 T_M 。

[0289] 进行本文公开的方法的一个或多个反应可包括使用一个或多个引物。如本文所用的,引物包含足够互补以与模板多核苷酸杂交的双链、单链或部分单链的多核苷酸。在与模板多核苷酸结合之前,引物可为单链DNA。在一些实施方案中,引物最初含有双链序列。引物位点包括引物杂交的模板区域。在一些实施方案中,引物能够作为模板指导的核酸合成的起始点。例如,当存在四种不同的核苷酸和聚合试剂或酶如DNA或RNA聚合酶或逆转录酶时,引物可启动模板指导的核酸合成。引物对包含2个引物:具有与模板序列的5'端杂交的5'上游区的第一引物,以及具有与模板序列的3'端的互补体杂交的3'下游区的第二引物。引物组包含两个或更多个引物:具有与模板序列或多个模板序列的5'端杂交的5'上游区的第一引物或第一多个引物,以及具有与模板序列或多个模板序列的3'端的互补体杂交的3'下游区的第二引物或第二多个引物。在一些实施方案中,引物包含目标特异性序列。在一些实施方案中,引物包含样品条码序列。在一些实施方案中,引物包含通用引发序列。在一些实施方案中,引物包含PCR引发序列。在一些实施方案中,引物包含用于启动多核苷酸扩增的PCR引发序列。(Dieffenbach,PCR Primer:A Laboratory Manual,第二版(Cold Spring

Harbor Press, New York (2003))。通用引物结合位点或序列允许通用引物与多核苷酸和/或扩增子附接。通用引物是本领域公知的,并且包括但不限于-47F (M13F)、 α MF、AOX3'、AOX5'、BGHr、CMV-30、CMV-50、CVMf、LACrmt、 λ gt10F、 λ gt10R、 λ gt11F、 λ gt11R、M13 rev、M13Forward (-20)、M13Reverse、male、p10SEQPpQE、pA-120、pet4、pGAP Forward、pGLRVpr3、pGLpr2R、pKLAC14、pQEFS、pQERS、pucU1、pucU2、reversA、seqIREStam、seqIRESzpet、seqori、seqPCR、seqIRES-、seqIRES+、seqpSecTag、seqpSecTag+、seqretro+PSI、SP6、T3-prom、T7-prom和T7-termInv。如本文所用的,附接可指共价相互作用和非共价相互作用两者或之一。通用引物与通用引物结合位点的附接可用于多核苷酸和/或扩增子的扩增、检测和/或测序。通用引物结合位点可包含至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000个核苷酸或碱基对。在另一个实例中,通用引物结合位点包含至少约1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500或10000个核苷酸或碱基对。在一些实施方案中,通用引物结合位点包含1-10、10-20、10-30或10-100个核苷酸或碱基对。在一些实施方案中,通用引物结合位点包含约1-90、1-80、1-70、1-60、1-50、1-40、1-30、1-20、1-10、2-90、2-80、2-70、2-60、2-50、2-40、2-30、2-20、2-10、1-900、1-800、1-700、1-600、1-500、1-400、1-300、1-200、1-100、2-900、2-800、2-700、2-600、2-500、2-400、2-300、2-200、2-100、5-90、5-80、5-70、5-60、5-50、5-40、5-30、5-20、5-10、10-90、10-80、10-70、10-60、10-50、10-40、10-30、10-20、10-10、5-900、5-800、5-700、5-600、5-500、5-400、5-300、5-200、5-100、10-900、10-800、10-700、10-600、10-500、10-400、10-300、10-200、10-100、25-900、25-800、25-700、25-600、25-500、25-400、25-300、25-200、25-100、100-1000、100-900、100-800、100-700、100-600、100-500、100-400、100-300、100-200、200-1000、200-900、200-800、200-700、200-600、200-500、200-400、200-300、300-1000、300-900、300-800、300-700、300-600、300-500、300-400、400-1000、400-900、400-800、400-700、400-600、400-500、500-1000、500-900、500-800、500-700、500-600、600-1000、600-900、600-800、600-700、700-1000、700-900、700-800、800-1000、800-900或900-1000个核苷酸或碱基对。

[0290] 引物可具有与其在合成引物延伸产物中的应用相匹配的长度。引物可为长度为8-200个核苷酸的多核苷酸。引物的长度可取决于模板多核苷酸的序列和模板基因座。例如,可优化引物或引物组的长度和/或解链温度(T_M)。在一些情况下,引物的长度可为约、大于约或小于约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59或60个核苷酸。在一些实施方案中,引物的长度为约8-100个核苷酸,例如10-75、15-60、15-40、18-30、20-40、21-50、22-45、25-40、7-9、12-15、15-20、15-25、15-30、15-45、15-50、15-55、15-60、20-25、20-30、20-35、20-45、20-50、20-55或20-60个核苷酸的长度和之间的任何长度。在一些实施方案中,引物的长度为最多约10、12、15、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、55、60、65、70、75、80、85、90、95或100个核苷酸。

[0291] 通常,在指数扩增反应中可使用一对或多对引物;引物对的一个引物可为正向引物且引物对的一个引物可为反向引物。在一些实施方案中,在指数扩增反应中可使用第一对引物;第一对的一个引物可为与第一模板多核苷酸分子的序列互补的正向引物,且第一

对的一个引物可为与第一模板多核苷酸分子的第二序列互补的反向引物,且第一模板基因座可位于第一序列与第二序列之间。在一些实施方案中,在扩增反应中可使用第二对引物;第二对的一个引物可为与第二目标多核苷酸分子的第一序列互补的正向引物,且第二对的一个引物可为与第二目标多核苷酸分子的第二序列互补的反向引物,且第二目标基因座可位于第一序列与第二序列之间。在一些实施方案中,第二目标基因座包含可变轻链抗体序列。在一些实施方案中,在扩增反应中可使用第三对引物;第三对的一个引物可为与第三模板多核苷酸分子的第一序列互补的正向引物,且第三对的一个引物可为与第三模板多核苷酸分子的第二序列互补的反向引物,且第三模板基因座可位于第一序列与第二序列之间。

[0292] 所述一个或多个引物可与多个模板多核苷酸的至少一部分退火。所述一个或多个引物可与所述多个模板多核苷酸的3'端和/或5'端退火。所述一个或多个引物可与所述多个模板多核苷酸的内部区域退火。该内部区域可以距所述多个模板多核苷酸的3'端或5'端至少约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、100、150、200、220、230、240、250、260、270、280、290、300、310、320、330、340、350、360、370、380、390、400、410、420、430、440、450、460、470、480、490、500、510、520、530、540、550、560、570、580、590、600、650、700、750、800、850、900或1000个核苷酸。所述一个或多个引物可包含固定的引物组。所述一个或多个引物可包含至少一个或多个定制引物。所述一个或多个引物可包含至少一个或多个控制引物。所述一个或多个引物可包含至少一个或多个管家基因引物。所述一个或多个引物可包含通用引物。该通用引物可与通用引物结合位点退火。在一些实施方案中,所述一个或多个定制引物与SBC、目标特异性区域、其互补体或其任意组合退火。所述一个或多个引物可包含通用引物。可设计所述一个或多个引物以扩增或进行引物延伸、逆转录、线性延伸、非指数扩增、指数扩增、PCR或一个或多个目标或模板多核苷酸的任何其他扩增方法。

[0293] 目标特异性区域可包含至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、100、150、200、220、230、240、250、260、270、280、290、300、310、320、330、340、350、360、370、380、390、400、410、420、430、440、450、460、470、480、490、500、510、520、530、540、550、560、570、580、590、600、650、700、750、800、850、900或1000个核苷酸或碱基对。在另一个实例中,目标特异性区域包含至少约1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500或10000个核苷酸或碱基对。在一些实施方案中,目标特异性区域包含约5-10、10-15、10-20、10-30、15-30、10-75、15-60、15-40、18-30、20-40、21-50、22-45、25-40、7-9、12-15、15-20、15-25、15-30、15-45、15-50、15-55、15-60、20-25、20-30、20-35、20-45、20-50、20-55、20-60、2-900、2-800、2-700、2-600、2-500、2-400、2-300、2-200、2-100、25-900、25-800、25-700、25-600、25-500、25-400、25-300、25-200、25-100、100-1000、100-900、100-800、100-700、100-600、100-500、100-400、100-300、100-200、200-1000、200-900、200-800、200-700、200-600、200-500、200-400、200-300、300-1000、300-900、300-800、300-700、300-600、300-500、300-400、400-1000、400-900、400-800、400-700、400-600、400-500、500-1000、500-900、500-800、500-700、500-600、600-1000、600-900、600-800、600-700、700-1000、700-900、700-800、800-1000、800-900或900-1000个核苷酸或碱基对。

[0294] 可根据已知参数设计引物,以便避免二级结构和自身杂交。在一些实施方案中,不同的引物对可在大约相同的温度下退火和解链,例如,在另一个引物对的1°C、2°C、3°C、4°C、5°C、6°C、7°C、8°C、9°C或10°C内。在一些实施方案中,多个引物中的一个或多个引物可在大约相同的温度下退火和解链,例如,在多个引物中另一个引物的1°C、2°C、3°C、4°C、5°C、6°C、7°C、8°C、9°C或10°C内。在一些实施方案中,多个引物中的一个或多个引物可在与前述多个引物中另一个引物不同的温度下退火和解链。

[0295] 用于本文所述方法的一个或多个步骤的多个引物可包括多个引物,其包含约、最多约或至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、50,000,000、100,000,000个不同引物。例如,多个引物中的每个引物可包含不同的目标或模板特异性区域或序列。

逆转录

[0296] 在一些情况下,通过逆转录从RNA制备目标多核苷酸。在一些情况下,通过引物延伸,诸如使用聚合酶,从DNA制备目标多核苷酸。

[0297] 本文所述的方法可在耦合的逆转录-PCR(逆转录-PCR)中使用。例如,逆转录和PCR可在两个不同的步骤中进行。首先,可使用多核苷酸dT引物、序列特异性引物、通用引物或本文所述的任何引物合成样品mRNA的cDNA拷贝。

[0298] 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。

[0299] 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°C至约75°C下、在约37°C至约50°C下、在约37°C至约55°C下、在约37°C至约60°C下、在约55°C至约75°C下、在约55°C至约60°C下、在约37°C下或在约60°C下进行。在一些情况下,使用将3个或更多个非模板末端核苷酸转移到转录产物末端的逆转录酶。

[0300] 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液滴中进行。

[0301] 逆转录反应可在范围为5 μ L至100 μ L或10 μ L至20 μ L反应体积的体积内进行。在小液滴中,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体积的小液滴中进行。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液滴中进行。在一些情况下,PCR反应在具有约或小于1nL体积的小液滴中进行。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液滴中进行。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的体积的小液滴中进行。在一些情况下,逆转录反应和PCR反应在不同的小液滴中进行。在一些情况下,逆转录反应和PCR反应在多个小液滴中进行,每个小液

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反应在多个小液滴中进行,每个小液滴具有约或小于1nL的体积。

[0302] 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液滴内进行,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液滴内进行。在一些情况下,第一PCR反应在具有约或小于1nL的体积的第一小液滴内进行,且第二PCR反应在具有约或小于1nL的体积的第二小液滴内进行。

[0303] 在一些情况下,第一PCR反应和第二PCR反应在多个小液滴中进行,每个小液滴具有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反应在多个小液滴中进行,每个小液滴具有约或小于1nL的体积。

[0304] 可使用一种或多种逆转录引物将目标多核苷酸如RNA逆转录为cDNA。该一种或多种逆转录引物可包含与RNA的区域如恒定区(例如,重链或轻链恒定区,或mRNA的聚A尾)互补的区域。在一些实施方案中,逆转录引物可包括第一逆转录引物和第二逆转录引物,第一逆转录引物具有与第一RNA的恒定区互补的区域,第二逆转录引物具有与第二RNA的恒定区互补的区域。在一些实施方案中,逆转录引物可包括第一逆转录引物和一个或多个逆转录引物,第一逆转录引物具有与第一RNA的恒定区互补的区域,所述一个或多个逆转录引物分别具有与一个或多个RNA的恒定区互补的区域。

[0305] 在一些实施方案中,逆转录引物不包含条码。

[0306] 逆转录引物可进一步包含不与RNA的区域互补的区域。在一些实施方案中,不与RNA的区域互补的区域在与RNA互补的引物区域的5'侧。在一些实施方案中,不与RNA的区域互补的区域在与RNA互补的引物区域的3'侧。在一些实施方案中,不与RNA的区域互补的区域为5'突出端区域。在一些实施方案中,不与RNA的区域互补的区域包含用于扩增和/或测序反应的引发位点。通过使用本文所述的一种或多种引物,使用本领域已知的合适试剂逆转录RNA分子。

[0307] 在进行RNA分子的逆转录反应后,产生的cDNA分子可用分子条码和容器条码进行条码化,并通过一个或多个PCR反应进行扩增,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。

[0308] 在一些实施方案中,多个正向/反向引物包括一个或多个正向/反向引物,其中多个正向/反向引物中的每个正向/反向引物包含与cDNA或条码化cDNA的V区段的一个或多个上游或下游区域互补的区域。例如,多个正向/反向引物包括包含与cDNA或条码化cDNA的V区段的上游或下游区域互补的区域的正向/反向引物,以及包含与cDNA或条码化cDNA的V区段的一个或多个其他上游或下游区域互补的区域的正向/反向引物。例如,多个正向/反向引物包括包含与cDNA或条码化cDNA的V区段的第一和/或第二上游或下游区域互补的区域的正向/反向引物,以及包含与cDNA或条码化cDNA的V区段的第二上游或下游区域互补的区域的第二正向/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正向/反向引物。

第二正向/反向引物,包含与cDNA或条码化cDNA的V区段的第二上游或下游区域互补的区域的第二正向/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,等等。多个正向/反向引物中的引物可用来与样品中的细胞(诸如免疫B细胞或T细胞)表达的所有V区段的所有可能的上游或下游区域退火。

[0309] 在一些实施方案中,多个正向/反向引物包括一个或多个正向/反向引物,其中多个正向/反向引物中的每个正向/反向引物包含与cDNA或条码化cDNA的C区段的一个或多个上游或下游区域互补的区域。例如,多个正向/反向引物包括包含与cDNA或条码化cDNA的C区段的上游或下游区域互补的区域的正向/反向引物,以及包含与cDNA或条码化cDNA的C区段的一个或多个其他上游或下游区域互补的区域的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,等等。多个正向/反向引物中的引物可用来与样品中的细胞(诸如免疫B细胞或T细胞)表达的所有C区段的所有可能的上游或下游区域退火。

[0310] 在一些实施方案中,多个正向/反向引物包括一个或多个正向/反向引物,其中多个正向/反向引物中的每个正向/反向引物包含与条码化cDNA的分子条码的一个或多个上游或下游区域互补的区域。例如,多个正向/反向引物包括包含与条码化cDNA的分子条码的上游或下游区域互补的区域的正向/反向引物,以及包含与条码化cDNA的分子条码的一个或多个其他上游或下游区域互补的区域的一个或多个其他正向/反向引物。例如,多个正向/反向引物包括包含与条码化cDNA的分子条码的第一和/或第二上游或下游区域互补的区域的第二正向/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,包含与条码化cDNA的分子条码的第二上游或下游区域互补的区域的第二正向/反向引物,以及包含与条码化cDNA的分子条码的第三上游或下游区域互补的区域的第三正向/反向引物等。多个正向/反向引物可用来与样品中的细胞(诸如免疫B细胞或T细胞)表达的所有分子条码的所有可能的上游或下游区域退火。

[0311] 在一些实施方案中,多个正向/反向引物包括一个或多个正向/反向引物,其中多个正向/反向引物中的每个正向/反向引物包含与条码化cDNA的容器条码的一个或多个上游或下游区域互补的区域。例如,多个正向/反向引物包括包含与条码化cDNA的容器条码的上游或下游区域互补的区域的正向/反向引物,以及包含与条码化cDNA的容器条码的一个或多个其他上游或下游区域互补的区域的一个或多个其他正向/反向引物。例如,多个正向/反向引物包括包含与条码化cDNA的容器条码的第一和/或第二上游或下游区域互补的区域的第二正向/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第二正向/反向引物。

引物,包含与条码化cDNA的容器条码的第二上游或下游区域互补的区域的第二正向/反向引物,以及包含与条码化cDNA的容器条码的第三上游或下游区域互补的区域的第三正向/反向引物等。多个正向/反向引物中的引物可用来与样品中的细胞(诸如免疫B细胞或T细胞)表达的所有容器条码的所有可能的上游或下游区域退火。

[0312] 多个正向/反向引物中的正向/反向引物进一步包含不与RNA的区域互补的区域。在一些实施方案中,不与RNA的区域互补的区域在与RNA互补的正向/反向引物区域(即,V区段的上游或下游区域)的5'侧。在一些实施方案中,不与RNA的区域互补的区域在与RNA互补的正向/反向引物区域的3'侧。在一些实施方案中,不与RNA的区域互补的区域为5'突出端区域。在一些实施方案中,不与RNA的区域互补的区域包含用于扩增和/或第二测序反应的引发位点。在一些实施方案中,不与RNA的区域互补的区域包含用于扩增和/或第三测序反应的引发位点。在一些实施方案中,不与RNA的区域互补的区域包含用于第二和第三测序反应的引发位点。在一些实施方案中,用于第二和第三测序反应的引发位点的序列是相同的。通过使用如本文所述的一个或多个正向/反向引物和反向引物,使用本领域已知的合适试剂扩增cDNA分子。在一些实施方案中,区域互补于RNA的区域,诸如恒定区或mRNA的聚A尾。

扩增

[0313] 含有目标多核苷酸的样品可包含mRNA或其片段,其可进行扩增。在一些情况下,mRNA或其片段的平均长度可为小于约100、200、300、400、500或800个碱基对,或小于约5、10、20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190或200个核苷酸,或小于约1、2、5、10、20、30、40、50、60、70、80、90、100千碱基。在一些情况下,从相对较短的模板,诸如含有约40、45、50、55、60、65、70、75、80、85、90、95或100个碱基的模板的样品扩增目标序列。

[0314] 扩增反应可包含一种或多种添加剂。在一些情况下,所述一种或多种添加剂为二甲亚砜(DMSO)、甘油、甜菜碱(单)水合物(N,N,N-三甲基甘氨酸=[羧甲基]三甲基铵)、海藻糖、7-脱氮-2'-脱氧鸟苷三磷酸(dC7GTP或7-脱氮-2'-dGTP)、BSA(牛血清蛋白)、甲酰胺、四甲基氯化铵(TMAC)、其他四烷基铵衍生物(例如,四乙基氯化铵(TEA-Cl)和四丙基氯化铵(TPrA-Cl)、非离子型去污剂(例如,Triton X-100、吐温20、Nonidet P-40(NP-40))或PREXCEL-Q。在一些情况下,扩增反应包含0、1、2、3、4、5、6、7、8、9或10种不同添加剂。在其他情况下,扩增反应包含至少0、1、2、3、4、5、6、7、8、9或10种不同添加剂。

[0315] 可在包含于反应体积(例如,小液滴)中的样品上进行热循环反应。小液滴可为多分散的或优选单分散的,通过搅拌、超声处理或穿过T通道接合点以微流体方式或本领域熟知的其他手段生成。密度可超过20,000个小液滴/40 μ l(1nL小液滴)、200,000个小液滴/40 μ l(100pL小液滴)。在热循环期间,小液滴可保持完整。在热循环期间,小液滴可以大于约10,000个小液滴/ μ L、100,000个小液滴/ μ L、200,000个小液滴/ μ L、300,000个小液滴/ μ L、400,000个小液滴/ μ L、500,000个小液滴/ μ L、600,000个小液滴/ μ L、700,000个小液滴/ μ L、800,000个小液滴/ μ L、900,000个小液滴/ μ L或1,000,000个小液滴/ μ L的密度保持完整。在其他情况下,两个或更多个小液滴在热循环期间不聚结。在其他情况下,大于100个或大于1,000个小液滴在热循环期间不聚结。

[0316] 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、T4 DNA聚合酶、Taq聚合酶、Pfu DNA聚合酶、

Vent DNA聚合酶、噬菌体29、REDTaq™、基因组DNA聚合酶或测序酶。在一些情况下,使用热稳定的DNA聚合酶。还可进行热启动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。

[0317] 可通过本领域已知的任何手段进行目标核酸的扩增。可通过聚合酶链反应(PCR)或等温DNA扩增来扩增目标核酸。可使用的PCR技术的实例包括但不限于定量PCR、定量荧光PCR(QF-PCR)、多重荧光PCR(MF-PCR)、实时PCR(逆转录-PCR)、单细胞PCR、限制性片段长度多态性PCR(PCR-RFLP)、PCR-RFLP/逆转录-PCR-RFLP、热启动PCR、巢式PCR、原位聚合酶克隆PCR、原位滚环扩增(RCA)、数字PCR(dPCR)、小液滴数字PCR(ddPCR)、桥式PCR、皮升PCR和乳液PCR。其他合适的扩增方法包括连接酶链反应(LCR)、转录扩增、分子倒置探针(MIP)PCR、自动维持序列复制、目标多核苷酸序列的选择性扩增、共有序列引物聚合酶链反应(CP-PCR)、任意引物聚合酶链反应(AP-PCR)、简并多核苷酸引物PCR(DOP-PCR)和基于核酸的序列扩增(NABSA)。本文可使用的其他扩增方法包括美国专利号5,242,794、5,494,810、4,988,617和6,582,938中描述的方法,以及包括Q β 复制酶介导的RNA扩增。扩增可为恒温扩增,例如,恒温线性扩增

[0318] 在一些实施方案中,扩增不在固体支持体上发生。在一些实施方案中,扩增不在小液滴中的固体支持体上发生。在一些实施方案中,当不在小液滴中扩增时,扩增在固体支持体上发生。

[0319] 扩增反应可包含一种或多种添加剂。在一些实施方案中,所述一种或多种添加剂为二甲亚砜(DMSO)、甘油、甜菜碱(单)水合物(N,N,N-三甲基甘氨酸=[羧甲基]三甲基铵)、海藻糖、7-脱氮-2'-脱氧鸟苷三磷酸(dC7GTP或7-脱氮-2'-dGTP)、BSA(牛血清蛋白)、甲酰胺、四甲基氯化铵(TMAC)其他四烷基铵衍生物(例如,四乙基氯化铵(TEA-Cl)和四丙基氯化铵(TPrA-Cl)、非离子型去污剂(例如,Triton X-100、吐温20、Nonidet P-40(NP-40))或PREXCEL-Q。在一些实施方案中,扩增反应可包含0、1、2、3、4、5、6、7、8、9或10种不同添加剂。在其他情况下,扩增反应可包含至少0、1、2、3、4、5、6、7、8、9或10种不同添加剂。

测序

[0320] 在进行本文所述的一个或多个方法或方法步骤后,可对生成的多核苷酸文库进行测序。

[0321] 可通过本领域已知的任何测序方法进行测序。在一些实施方案中,可进行高通量测序。合适的下一代测序技术包括454Life Sciences平台(Roche,Branford,CT)(Margulies等人,Nature,437,376-380(2005));Illumina的基因组分析仪、GoldenGate甲基化分析或Infinium甲基化分析,即Infinium人类甲基化27K BeadArray或VeraCode GoldenGate甲基化分析(Illumina,San Diego,CA;Bibkova等人,Genome Res.16,383-393(2006);和美国专利号6,306,597、7,598,035、7,232,656),或DNA连接测序,SOLiD系统(Applied Biosystems/Life Technologies;美国专利号6,797,470、7,083,917、7,166,434、7,320,865、7,332,285、7,364,858和7,429,453);或Helicos True单分子DNA测序技术

(Harris等人, Science, 320, 106-109 (2008); 以及美国专利号7,037,687、7,645,596、7,169,560和7,769,400)、Pacific Biosciences的单分子、实时 (SMRT™) 技术和测序 (Soni等人, Clin.Chem. 53, 1996-2001 (2007))。这些系统允许对从样品中分离的许多多核苷酸进行多重平行测序 (Dear, Brief Funct. Genomic Proteomic, 1 (4), 397-416 (2003) 和 McCaughan 等人, J. Pathol., 220, 297-306 (2010))。在一些实施方案中, 通过染料修饰的探针的连接测序、焦磷酸测序或单分子测序对多核苷酸进行测序。对多核苷酸序列的测定可通过如下测序方法进行, 诸如 Helioscope™ 单分子测序、纳米孔DNA测序、Lynx Therapeutics 大规模平行签名测序 (MPSS)、454焦磷酸测序、单分子实时 (RNAP) 测序、Illumina (Solexa) 测序、SOLiD测序、Ion Torrent™、离子半导体测序、单分子SMRT™测序、聚合酶克隆测序、DNA纳米球测序和 VisiGen Biotechnologies 方法。或者, 测定多核苷酸的序列可采用测序平台, 包括但不限于基因组分析仪 IIX、HiSeq, 以及 Illumina 提供的 MiSeq, 单分子实时 (SMRT™) 技术, 诸如 Pacific Biosciences (California) 提供的 PacBio RS 系统以及 Solexa 测序仪、真正单分子测序 (tSMS™) 技术, 诸如 Helicos Inc. (Cambridge, MA) 提供的 HeliScope™ 测序仪。测序可包括 MiSeq 测序。测序可包括 HiSeq 测序。在一些实施方案中, 测定多核苷酸的序列包括配对末端测序、纳米孔测序、高通量测序、鸟枪法测序、染料终止子测序、多引物DNA测序、引物步移、Sanger 双脱氧测序、Maxim-Gilbert 测序、焦磷酸测序、真正单分子测序或其任意组合。或者, 可通过电子显微术或化学敏感性场效应晶体管 (chemFET) 阵列测定多核苷酸的序列。

[0322] 方法可进一步包括对文库中的一个或多个多核苷酸进行测序。方法可进一步包括将文库中的一个或多个多核苷酸序列、序列读取、扩增子序列或扩增子集序列彼此进行比对。

[0323] 如本文所用的, 比对包括将测试序列如序列读取与一个或多个其他测试序列、参考序列或其组合进行比较。在一些实施方案中, 比对可用于从多个序列或比对的序列确定共有序列。在一些实施方案中, 比对包括从各自具有相同分子条码或容器条码的多个序列确定共有序列。在一些实施方案中, 为了比较目的而比对的序列的长度为参考序列长度的至少30%、至少40%、至少50%、至少60%、至少70%、至少80%、至少90%或至少95%。可通过公知的方法, 例如, 使用数学算法完成对两个或更多个序列的实际比较。这样的数学算法的非限制性实例描述于 Karlin, S. 和 Altschul, S., Proc. Natl. Acad. Sci. USA, 90-5873-5877 (1993) 中。这样的算法并入 NBLAST 和 XBLAST 程序 (版本 2.0) 中, 如 Altschul, S. 等人, Nucleic Acids Res., 25:3389-3402 (1997) 中所述。当利用 BLAST 和 Gapped BLAST 程序时, 可使用相应程序 (例如, NBLAST) 的任何相关参数。例如, 用于序列比较的参数可设置为分数 = 100, 字长 = 12, 或可变化 (例如, $w = 5$ 或 $w = 20$)。其他实例包括 Myers 和 Miller, CABIOS (1989) 的算法、ADVANCE、ADAM、BLAT 和 FASTA。在一些实施方案中, 可使用例如 GCG 软件包中的 GAP 程序 (Accelrys, Cambridge, UK) 完成两个氨基酸序列之间的百分比同一性。

[0324] 测序可包括对多核苷酸的至少约 10、20、30、40、50、60、70、80、90、100 个或更多个核苷酸或碱基对进行测序。在一些实施方案中, 测序包括对多核苷酸的至少约 200、300、400、500、600、700、800、900、1000 个或更多个核苷酸或碱基对进行测序。在其他情况下, 测序包括对多核苷酸的至少约 1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000 个或更多个核苷酸或碱基对进行测序。

[0325] 测序可包括每次运行至少约200、300、400、500、600、700、800、900、1000个或更多个测序读取。如本文所用的，序列读取包括从通过测序技术生成的序列或数据流确定的核苷酸序列。在一些实施方案中，测序包括每次运行对至少约1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000个或更多个测序读取进行测序。测序可包括每次运行多于、少于或等于约1,000,000,000个测序读取。测序可包括每次运行多于、少于或等于约200,000,000个读取。

[0326] 在一些实施方案中，用于确定共有序列的序列读取的数目为约2-1000个序列读取。例如，用于确定共有序列的序列读取的数目可为约2-900、2-800、2-700、2-600、2-500、2-400、2-300、2-200、2-100、25-900、25-800、25-700、25-600、25-500、25-400、25-300、25-200、25-100、100-1000、100-900、100-800、100-700、100-600、100-500、100-400、100-300、100-200、200-1000、200-900、200-800、200-700、200-600、200-500、200-400、200-300、300-1000、300-900、300-800、300-700、300-600、300-500、300-400、400-1000、400-900、400-800、400-700、400-600、400-500、500-1000、500-900、500-800、500-700、500-600、600-1000、600-900、600-800、600-700、700-1000、700-900、700-800、800-1000、800-900或900-1000个序列读取。在一些实施方案中，用于确定共有序列的序列读取的数目为至少约1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、55,000、60,000、65,000、70,000、75,000、80,000、85,000、90,000、95000、100,000、150,000、200,000、250,000、300,000、350,000、400,000、450,000、500,000、550,000、600,000、650,000、700,000、750,000、800,000、850,000、900,000、950,000、1,000,000、50,000,000或100,000,000个读取。在一些实施方案中，用于确定共有序列的序列读取的数目为至多约1000、1500、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、25,000、30,000、35,000、40,000、45,000、50,000、55,000、60,000、65,000、70,000、75,000、80,000、85,000、90,000、95000、100,000、150,000、200,000、250,000、300,000、350,000、400,000、450,000、500,000、550,000、600,000、650,000、700,000、750,000、800,000、850,000、900,000、950,000、1,000,000、50,000,000或100,000,000个读取。

[0327] 方法可包括对错读进行测序。方法可包括确定错读的数目，诸如用于确定反应条件或设计引物序列。比较在一种或多种第一条件或条件组下生成的错读的数目可用于确定优选的条件或条件组。例如，可在PCR反应期间在高盐浓度下进行第一方法，并且可在PCR反应期间在低盐浓度下进行第二方法，其中除了盐浓度差异之外，第一和第二方法基本相同地进行。如果第一方法导致更高数目的错读，诸如对特定目标多核苷酸序列或引物的更高数目的错读，则可确定对于该特定目标多核苷酸序列或引物优选更低的盐反应条件。

酶

[0328] 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。

[0329] 在一些实施方案中，衔接子与多核苷酸的附接包括使用一种或多种连接酶。连接酶的实例包括但不限于DNA连接酶，诸如DNA连接酶I、DNA连接酶III、DNA连接酶IV和T4 DNA连接酶，以及RNA连接酶，诸如T4 RNA连接酶I和T4 RNA连接酶II。

[0330] 本文所公开的方法和试剂盒可进一步包括使用一种或多种逆转录酶。在一些实施方案中,该逆转录酶为HIV-1逆转录酶、M-MLV逆转录酶、AMV逆转录酶和端粒酶逆转录酶。在一些实施方案中,该逆转录酶为M-MLV逆转录酶。

[0331] 在一些实施方案中,本文所公开的方法和试剂盒包括使用一种或多种蛋白酶。

[0332] 在一些实施方案中,本文所公开的方法和试剂盒包括使用一种或多种聚合酶。聚合酶的实例包括但不限于DNA聚合酶和RNA聚合酶。在一些实施方案中,该DNA聚合酶为DNA聚合酶I、DNA聚合酶II、DNA聚合酶III全酶和DNA聚合酶IV。可商购获得的DNA聚合酶包括但不限于Bst 2.0DNA聚合酶、Bst 2.0WarmStart™DNA聚合酶、Bst DNA聚合酶、硫化叶菌DNA聚合酶IV、Taq DNA聚合酶、9°N™m DNA聚合酶、Deep VentR™(exo-)DNA聚合酶、Deep VentR™DNA聚合酶、Hemo KlenTaq™、LongAmp® Taq DNA聚合酶、OneTaq® DNA聚合酶、Phusion® DNA聚合酶、Q5™高保真DNA聚合酶、Therminator™γ DNA聚合酶、Therminator™DNA聚合酶、Therminator™II DNA聚合酶、Therminator™III DNA聚合酶、VentR® DNA聚合酶、VentR®(exo-)DNA聚合酶、Bsu DNA聚合酶、phi29 DNA聚合酶、T4 DNA聚合酶、T7 DNA聚合酶、末端转移酶、Titanium®Taq聚合酶、KAPA Taq DNA聚合酶和KAPA Taq热启动DNA聚合酶。

[0333] 在一些实施方案中,所述聚合酶为RNA聚合酶,诸如RNA聚合酶I、RNA聚合酶II、RNA聚合酶III、大肠杆菌聚(A)聚合酶、phi6 RNA聚合酶(RdRP)、聚(U)聚合酶、SP6 RNA聚合酶和T7 RNA聚合酶。

人造抗原呈递细胞(aAPC)和APC报道分子

[0334] 在本文中也被称为工程化APC(eAPC)的人造抗原呈递细胞(aAPC)可指不表达内源性MHC蛋白质的APC细胞(在一些实施方案中是指基于K562细胞系的细胞)。在一些实施方案中,aAPC细胞用与探讨中的TCR的来源生物体的MHC基因相匹配的MHC基因来转染。aAPC可进一步用编码感兴趣的肽抗原的表达载体来转染或用靶抗原的文库来脉冲处理。

[0335] 在一些实施方案中,额外地将aAPC工程化为含有报道基因构建体,所述报道基因构建体可在被T细胞识别抗原后表达报道分子(例如,萤光素酶、荧光蛋白、表面蛋白质)。在一些方面,在被T细胞识别和/或与T细胞接合后表达报道分子的APC(和aAPC)被认为对T细胞具有反应性。这些报道基因的表达可以是检测aAPC与T细胞结合的方法,从而允许鉴别感兴趣TCR的靶抗原。在一些实施方案中,靶抗原不是已知的。在一些实施方案中,TCR包含具有已知可变区序列的TCR。例如,本文所述的方法可包括用感兴趣的已知可变区序列鉴别TCR的未知靶抗原。

[0336] 在一些实施方案中,检测通过MHC-肽复合体与TCR的结合而变得激活的aAPC可导致报道分子的表达。在一些实施方案中,报道分子为荧光蛋白,诸如萤光素酶、红色荧光蛋白、绿色荧光蛋白、黄色荧光蛋白、绿色荧光蛋白衍生物或任何工程化荧光蛋白。在进一步的实施方案中,荧光报道分子的检测可使用荧光技术进行。例如,可使用荧光读板器、流式细胞术或荧光显微术来测量荧光蛋白表达。在一些实施方案中,可使用荧光激活细胞分选仪(FACS)基于荧光报道分子的表达来分选激活的APC。

[0337] 在一些实施方案中,检测通过MHC-肽复合体与TCR结合而变得激活的aAPC可导致编码细胞表面标志物的报道基因的表达。APC可在接合TCR后上调或下调各种细胞表面标志

物。在一些实施方案中,细胞表面蛋白质如CD80、CD86、MHC I、MHC II、CD11c、CD11b、CD8 α 、OX40-L、ICOS-1或CD40的表达水平可在MHC-肽复合体与TCR结合之后变化(例如,增加或减少)。在一些实施方案中,细胞表面报道分子的检测可使用诸如免疫组织化学、通过流式细胞术的荧光染色和定量或用cDNA阵列或mRNA定量测定基因表达变化等技术来进行。在一些实施方案中,可使用磁性激活细胞分选基于细胞表面报道分子的表达来分离激活的APC。

[0338] 在一些实施方案中,检测通过MHC-肽复合体与TCR结合而激活的aAPC可导致编码分泌因子如IL6、IL-12、IFN α 、IL-23、IL-1、TNF或IL-10的报道基因的表达。在进一步的实施方案中,可通过mRNA定量、cDNA阵列或通过诸如酶联免疫吸附测定(ELISA)或酶联免疫斑点法(ELISPOT)等测定对表达蛋白质的定量来检测这些分泌因子。

[0339] 在一些实施方案中,报道基因在aAPC中的表达可通过工程化T细胞在T细胞活化期间结合信号而触发。例如,呈递感兴趣抗原的APC可被工程化为表达合成Notch(synNotch)受体。在一些实施方案中,synNotch受体为具有外部结构域和内部结构域的跨细胞蛋白质。在一些实施方案中,synNotch可在MHC-肽TCR复合体形成的同时与靶抗原结合。在其他实施方案中,synNotch可在形成MHC-肽TCR复合体之后与靶抗原结合。在synNotch与靶抗原结合后,受体的内部结构域被切割而产生转录激活域,该转录激活域迁移至细胞核并促进报道基因被APC的表达。在一些方面,只有在靶T细胞被激活(例如,通过TCR与合适的MHC-肽复合体接合)时,由APC表达的合成Notch受体方能识别由所述T细胞表达的蛋白质配体和/或抗原。

[0340] 在一些实施方案中,由APC和/或工程化/人造APC表达的报道分子系统包括信号传导系统,所述信号传导系统包含能够被一种或多种在工程化T细胞识别其同源抗原(例如,识别APC上呈递的MHC-肽复合体)时分泌的效应剂(例如,酶)切割的组分。在一些实施方案中,该系统是基于FRET的检测系统。示例性系统可在国际专利申请公开号W02015143558中找到。在一些方面,aAPC与工程化T细胞的接合导致T细胞的激活和/或酶、细胞因子和/或趋化因子的释放。在一些方面,这些酶、细胞因子和/或趋化因子代表细胞毒性信号和/或效应器功能。在一些方面,效应剂包括但不限于粒酶、穿孔素和颗粒溶素。在一些实施方案中,在T细胞受体与MHC-表位复合体接触和特异性识别后,工程化T细胞被激活并产生含有包括粒酶蛋白酶和穿孔素在内的多种效应剂的颗粒。在一些方面,激活的工程化T细胞的颗粒被运输到细胞表面并释放到邻近aAPC的细胞间环境中。在一些方面,穿孔素形成穿过报道细胞的细胞膜的孔,从而允许粒酶和其他效应剂进入报道细胞(例如,aAPC)的胞质溶胶。在一些实施方案中,该信号传导系统和/或报道分子构建体可包含猝灭的荧光蛋白,该猝灭的荧光蛋白可通过对肽连接的酶促切割(例如,粒酶切割)而去猝灭,其释放荧光部分,从而生成信号。在一些方面,报道分子构建体包含基于FRET的荧光蛋白信号传导系统,该系统在被由工程化T细胞递送至aAPC的效应剂切割(例如,酶促切割)时生成荧光信号。在一个方面,报道分子构建体包含被粒酶B(GzmB)可切割的连接体分隔的基因编码的CFP-YFP融合FRET对。

[0341] 在一些实施方案中,用于测量APC激活的系统包含含有合成报道分子如合成Notch受体的工程化APC。在一些实施方案中,合成报道分子为包含细胞内结构域的受体。细胞内结构域可包含转录激活蛋白。例如,该细胞内结构域可包含GAL4-VP64结构域。在一些实施方案中,合成报道分子在工程化APC中组成型地表达。

[0342] 在一些实施方案中,用于测量APC激活的系统包含工程化T细胞,该工程化T细胞包

含编码处于诱导型启动子如NFAT启动子控制下的细胞表面表达蛋白质配体的基因。合成报道分子,诸如合成Notch受体,可特异性结合工程化T细胞的细胞表面上的细胞表面表达蛋白质配体。

[0343] 在一些实施方案中,用于测量APC激活的系统包含在APC内的激活序列下游的荧光报道分子。例如,激活序列可以为GAL4-UAS。在本文所述的用于测量APC激活的系统中,在非刺激性条件下,T细胞在其表面上不表达蛋白质配体。在接触并随后用呈递在T细胞上表达的TCR的同源MHC-肽复合体的APC刺激后,T细胞的报道分子变得活跃并且细胞表面配体得以表达。例如,该蛋白质配体与APC上的合成Notch受体的细胞外结构域相互作用,从而导致细胞内GAL4-VP64结构域在APC内被切割。切割后,GAL4-VP64结构域转移至细胞核并激活在GAL4-UAS序列下游的荧光报道分子。

鉴别TCR抗原靶标

[0344] 可使用表达载体将感兴趣的天然配对TCR链转染至T细胞中。在一些实施方案中,将T细胞系工程化为缺乏内源性TCR表达。在一些方面,示例性T细胞系包括Jurkat细胞系。在一些方面,将表达感兴趣TCR的T细胞与(上述的)aAPC一起培养,该aAPC在匹配衍生出感兴趣TCR的宿主生物体的MHC分子上呈递抗原文库。在一些实施方案中,如上所述,T细胞通过经由TCR接合MHC-肽复合体来结合aAPC,并且导致APC表达报道基因。可使用荧光激活细胞分选(FACS)或磁性激活细胞分选(MACS)分离激活的APC。在一些实施方案中,可通过用酸和/或反相HPLC(RP-HPLC)处理从MHC分子洗脱出抗原肽。在进一步的实施方案中,可通过质谱法对抗原肽进行测序或分析。该方法允许快速且同时地筛选针对感兴趣TCR的较大靶抗原组,从而允许准确鉴别TCR的靶抗原。

[0345] 在一些实施方案中,可将抗原集合体去卷积。例如,表达来自受试者的TCR的工程化T细胞可与包含表达多个抗原靶标的工程化APC的工程化APC集合体接触。例如,表达来自受试者的TCR的工程化T细胞可与包含表达第一抗原靶标的工程化APC和表达第二抗原靶标的工程化APC的工程化APC集合体接触。例如,表达来自受试者的TCR的工程化T细胞可与包含表达第一抗原靶标的工程化APC和表达第二抗原靶标的工程化APC的第一工程化APC集合体接触,并且表达该TCR的工程化T细胞可与包含表达第三抗原靶标的工程化APC和表达第四抗原靶标的工程化APC的第二工程化APC集合体接触。

[0346] 例如,表达来自受试者的第一TCR的工程化T细胞可与包含表达第一抗原靶标的工程化APC和表达第二抗原靶标的工程化APC的工程化APC集合体接触,并且表达来自受试者的第二TCR的工程化T细胞可与包含表达第一抗原靶标的工程化APC和表达第二抗原靶标的工程化APC的工程化APC集合体接触。

[0347] 例如,表达来自受试者的第一TCR的工程化T细胞可与包含表达第一抗原靶标的工程化APC和表达第二抗原靶标的工程化APC的工程化APC集合体接触,并且表达来自受试者的第二TCR的工程化T细胞可与包含表达第一抗原靶标的工程化APC和表达第二抗原靶标的工程化APC的另一个工程化APC集合体接触。

[0348] 例如,表达来自受试者的第一TCR的工程化T细胞可与包含表达第一抗原靶标的工程化APC和表达第二抗原靶标的工程化APC的第一工程化APC集合体接触,并且表达第一TCR的工程化T细胞可与包含表达第三抗原靶标的工程化APC和表达第四抗原靶标的工程化APC的第二工程化APC集合体接触;并且表达来自受试者的第二TCR的工程化T细胞可以与包含

表达第一抗原靶标的工程化APC和表达第二抗原靶标的工程化APC的另一个第一工程化APC集合体接触,并且表达第二TCR的工程化T细胞可以与包含表达第三抗原靶标的工程化APC和表达第四抗原靶标的工程化APC的另一个第二工程化APC集合体接触。

[0349] 可将含有激活的APC的容器如孔中的APC重新接种到2个或更多个容器如孔中,使得所述两个或更多个容器含有来自APC所源自的容器的APC子集。例如,可将含有激活的APC的容器如孔中的APC重新接种到2个或更多个容器如孔中,使得所述两个或更多个容器比APC所源自的容器含有更少数量的APC。在一些实施方案中,可将含有激活的APC的容器如孔中的APC重新接种到2个或更多个容器如孔中,使得所述两个或更多个容器含有单个APC。

[0350] 然后可通过表达来自受试者的一个或多个TCR的工程化T细胞来测定重新接种的APC的激活。然后可以鉴别被激活的重新接种APC的抗原。例如,然后可对编码被激活的重新接种APC中的抗原的基因进行扩增和测序。

[0351] 在一些实施方案中,可使用下一代测序来鉴别抗原。例如,可将各自表达 m 个抗原中的一个的APC群体的子集合体接种到 n 数目的孔中,使得在任何孔中均表达 n 数目的抗原,其中 n 小于 m 。例如,可平衡子集合体的大小和孔的数目,使得多个孔将含有表达 m 个抗原中的每一个的APC。然后可用表达感兴趣TCR的工程化T细胞来测定每个孔。然后可收集含有激活的APC的孔。然后可鉴别所收集APC的抗原。例如,可以诸如通过下一代测序方法对编码在所收集APC中表达的抗原的DNA进行扩增并随后进行测序。可将含有激活的APC的第一孔中的抗原组与含有激活的APC的第二孔中的抗原组进行比较。例如,可将含有激活的APC的孔中的抗原组与含有激活的APC的每个其他孔中的抗原组进行比较。可将从含有激活APC的孔中的第一抗原组鉴别的且与另一个孔中的第二抗原组中的鉴别抗原相同的抗原鉴别为TCR的靶抗原。例如,可将从含有激活APC的孔中的抗原组鉴别的且与每个其他孔中的抗原组中的鉴别抗原相同的抗原鉴别为TCR的靶抗原。

鉴别抗原特异性TCR

[0352] 本公开内容提供了鉴别对感兴趣抗原具有反应性的人源TCR的方法,从而允许发现治疗相关的TCR。在一些实施方案中,感兴趣的抗原可以代表疾病病况如癌症。在一些实施方案中,可以将肿瘤浸润淋巴细胞的各种TCR转染到表达CD4和CD8共受体的T细胞系中。在一些实施方案中,将T细胞系工程化为缺乏内源性TCR表达。在一些方面,示例性T细胞系包括Jurkat细胞系。在一些实施方案中,可以从产生多个T细胞群体的多个患者的T细胞分离TCR。表达各种感兴趣的TCR的T细胞可以与呈递感兴趣的靶抗原的APC一起培养。在一些实施方案中,aAPC由多个aAPC群体制备,其中每个群体可用MHC分子转染以匹配存在于共培养物中的每个T细胞群体,从而全部呈递相同的靶抗原。在一些方面,分离并分析诱导aAPC中报道基因的表达的TCR表达T细胞,以确定对感兴趣抗原具有反应性的TCR。

[0353] 在一些实施方案中,对代表疾病病况的靶抗原具有反应性的经鉴别的TCR在T细胞中被工程化并用作免疫疗法。在某些实施方案中,可将该T细胞免疫疗法施用于有需要的受试者以治疗该病况。

治疗方法

[0354] 本公开内容提供了用针对天然配对TCR链的经鉴别靶抗原的疗法治疗有需要的受试者的方法。在一些实施方案中,有需要的受试者可以是人类。受试者可患有表达靶抗原的疾病。例如,该疾病可以是癌症,包括B细胞淋巴瘤、急性淋巴母细胞白血病(ALL)、慢性淋巴

细胞白血病、急性髓样白血病、肾上腺皮质癌 (adrenocortical carcinoma)、肾上腺皮质癌 (adrenal cortex cancer)、AIDS相关癌症、肛门癌、阑尾癌、星形细胞瘤、非典型畸胎样/横纹肌样瘤、基底细胞癌、肝外胆管癌、膀胱癌、骨癌 (包括尤因肉瘤和骨肉瘤和恶性纤维组织细胞瘤)、脑肿瘤、乳腺癌、伯基特淋巴瘤、类癌瘤 (胃肠道类癌瘤)、原发灶不明的癌、中枢神经系统淋巴瘤、原发性宫颈癌、胆管上皮癌、慢性淋巴细胞白血病 (CLL)、慢性骨髓性白血病 (CML)、慢性骨髓增殖性肿瘤、结直肠癌、皮肤T细胞淋巴瘤、导管原位癌 (DCIS)、子宫内膜癌、食管癌、尤因肉瘤、性腺外生殖细胞肿瘤、眼癌、眼内黑素瘤、视网膜母细胞瘤、输卵管癌、骨恶性纤维组织细胞瘤,和骨肉瘤、胆囊癌、胃癌、胃肠道类癌瘤、胃肠道间质瘤 (GIST)、性腺外、卵巢、睾丸的生殖细胞肿瘤、妊娠滋养细胞疾病、神经胶质瘤、多毛细胞白血病、头颈癌、肝细胞癌 (肝癌)、朗格汉斯细胞组织细胞增多症、霍奇金淋巴瘤、下咽癌、眼内黑素瘤、胰岛细胞瘤、胰腺神经内分泌瘤、卡波西肉瘤、肾朗格汉斯细胞组织细胞增多症、喉癌、白血病、唇癌和口腔癌、肝癌 (原发性肝癌)、肺癌、淋巴瘤、瓦尔登斯特伦巨球蛋白血症、男性乳腺癌、骨恶性纤维组织细胞瘤和骨肉瘤、黑素瘤、眼内黑素瘤、Merkel细胞癌、恶性间皮瘤、原发灶隐匿转移性颈部鳞状癌、口腔癌、多发性骨髓瘤/浆细胞肿瘤、蕈样肉芽肿病、骨髓增生异常综合征、骨髓增生异常/骨髓增殖性肿瘤和慢性骨髓增生性肿瘤、慢性髓性白血病 (CML)、急性髓样白血病 (AML)、鼻腔癌和鼻窦癌、鼻咽癌、神经母细胞瘤、非霍奇金淋巴瘤、非小细胞肺癌、口腔癌、唇癌和口腔癌以及口咽癌、骨肉瘤和骨恶性纤维组织细胞瘤、卵巢癌、胰腺癌和胰腺神经内分泌肿瘤 (胰岛细胞瘤)、副神经节瘤、副鼻窦癌和鼻腔癌、甲状旁腺癌、阴茎癌、咽癌、嗜铬细胞瘤、垂体瘤、浆细胞肿瘤/多发性骨髓瘤、妊娠期乳腺癌、原发性中枢神经系统 (CNS) 淋巴瘤、原发性腹膜癌、前列腺癌、直肠癌、肾细胞癌 (肾癌)、视网膜母细胞瘤、唾液腺癌、肉瘤、尤因肉瘤、卡波西肉瘤、骨肉瘤、横纹肌肉瘤、子宫肉瘤、Sézary综合征、皮肤癌、小细胞肺癌、小肠癌、软组织肉瘤、鳞状细胞癌、原发灶隐匿转移性颈部鳞状癌、胃癌、T细胞淋巴瘤、皮肤癌、睾丸癌、喉癌、胸腺瘤和胸腺癌、甲状腺癌、肾盂和输尿管移行细胞癌、输尿管和肾盂移行细胞癌、尿道癌、子宫癌、子宫内膜和子宫肉瘤、阴道癌、外阴癌、瓦尔登斯特伦巨球蛋白血症或维尔姆斯瘤 (wilms tumor)。

[0355] 在一些实施方案中,靶抗原可以与任何癌症相关联。在一些实施方案中,可从患者身上采取样品并使用上述方法针对TCR进行筛选。使用本公开内容中所提供的方法,可通过工程化T细胞的反应性来鉴别靶抗原。在某些实施方案中,然后可以用针对靶抗原的治疗剂治疗患者。例如,可向患者施用有效剂量的靶向所鉴别抗原的小分子、抗体、肽或蛋白质生物制剂或免疫疗法。在一些实施方案中,该免疫疗法可包括工程化为表达对靶抗原具有特异性的TCR的T细胞。

[0356] 在一些实施方案中,针对靶抗原的治疗剂可以为TCR或TCR样嵌合抗原受体 (CAR)。在一些方面,TCR样CAR特异性结合和/或识别在MHC的环境中的蛋白质表位。示例性的抗原受体 (包括CAR) 以及用于将这类受体工程化并引入细胞的方法包括在例如国际专利申请公开号W0200014257、W02013126726、W02012/129514、W02014031687、W02013/166321、W02013/071154、W02013/123061、美国专利申请公开号US2002131960、US2013287748、US20130149337、美国专利号6,451,995、7,446,190、8,252,592、8,339,645、8,398,282、7,446,179、6,410,319、7,070,995、7,265,209、7,354,762、7,446,191、8,324,353和8,479,118,以及欧洲专利申请号EP2537416中所述的抗原受体和方法,和/或由Sadelain等人,

Cancer Discov. 2013年4月;3(4):388-398; Davila等人, (2013) PLoS ONE 8(4):e61338; Turtle等人, Curr. Opin. Immunol., 2012年10月;24(5):633-39; Wu等人, Cancer, 2012年3月18日(2):160-75所述的抗原受体和方法。CAR的实例包括如任何前述出版物, 诸如W02014031687、US 8,339,645、US 7,446,179、US 2013/0149337、美国专利号:7,446,190、美国专利号:8,389,282、Kochenderfer等人, 2013, Nature Reviews Clinical Oncology, 10,267-276(2013); Wang等人(2012) J. Immunother. 35(9):689-701; 和Brentjens等人, Sci Transl Med. 20135(177)所公开的CAR。还参见W02014031687、US 8,339,645、US 7,446,179、US 2013/0149337、美国专利号:7,446,190和美国专利号8,389,282。嵌合受体如TCR样CAR通常包括细胞外抗原结合域, 如抗体分子的一部分, 该细胞外抗原结合域通常是抗体的可变重(VH)链区和/或可变轻(VL)链区, 例如scFv和/或scTv抗体片段。还提供涉及将特异性结合和/或识别通常由TCR识别的表位(如MHC环境中的肽表位)的抗体片段(如TCR样scFv和/或scTv)工程化的方法, 以及/或者用于生成和/或工程化具有某些scFv(如这类TCR样scFv和/或scTv)的特征(例如含有其一个或多个CDR)的TCR的方法。

[0357] 在一些实施方案中, TCR样嵌合抗原受体(CAR)可包含细胞内结构域(其包含T细胞受体的细胞内结构域)和细胞外部分(其包含抗体的TCR样抗原结合部分, 例如, 抗体的TCR样sFv和/或scTv)。通常, 嵌合受体(例如, TCR样CAR)包含在抗原结合部分与跨膜结构域之间的连接体或间隔域。在一些实施方案中, 连接体或间隔区来源于免疫球蛋白铰链区。在一些实施方案中, 除了编码包含T细胞受体的细胞内结构域的细胞内结构域和包含抗体的抗原结合部分(例如抗体的sFv和/或scTv)的细胞外部分之外还编码TCR样CAR构建体(在一些实施方案中为TCR样CAR载体)的核酸可包含启动子。例如, 启动子可以是含有骨髓增生性肉瘤病毒增强子修饰的MoMuLV LTR的U3区域的合成启动子。在其他实施方案中, 启动子可以是EF1a启动子或EF1启动子。在一些实施方案中, TCR样CAR可包含细胞内结构域(其包含T细胞受体的共刺激结构域和细胞内结构域)和细胞外部分(其包含抗体, 诸如本文提供的一种或多种抗体, 该抗体可以是单链抗体或其片段)。在一些实施方案中, 抗体是或包含sV、scFv和/或scTv。在一些实施方案中, TCR样CAR可进一步包含另外的细胞外部分, 诸如间隔区和/或铰链区。

[0358] 在一些方面, 富含甘氨酸和丝氨酸(和/或苏氨酸)的连接体包括至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的这些氨基酸。在一些实施方案中, 它们包括至少或约50%、55%、60%、70%或75%的甘氨酸、丝氨酸和/或苏氨酸。在一些实施方案中, 连接体基本上完全由甘氨酸、丝氨酸和/或苏氨酸组成。连接体的长度通常为约5至约50个氨基酸, 通常为10或约10至30或约30个氨基酸, 例如10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个氨基酸, 并且在一些实例中长度为10至25个氨基酸。示例性连接体包括具有不同数目的序列GGGG(4GS)或GGGS(3GS)的重复, 诸如这种序列的2、3、4至5次重复的连接体。示例性连接体包括具有GGGGSGGGSGGGG或由之组成的连接体。示例性连接体进一步包括具有序列GSTSGSGKPGSGEGSTKG或由之组成的连接体。

[0359] 在一些实施方案中, 细胞内信号传导结构域包括共刺激受体如CD28、CD137(4-1BB)、OX40和/或ICOS的细胞内信号传导结构域。在一些实施方案中, TCR样CAR包括调节TCR复合体的初级激活的初级细胞质信号传导序列。以刺激方式起作用的初级细胞质信号传导

序列可含有被称为基于免疫受体酪氨酸的激活基序或ITAM的信号传导基序。含有初级细胞质信号传导序列的ITAM的实例包括衍生自TCR δ 、FcR γ 、FcR β 、CD3 γ 、CD3 δ 、CD3 ϵ 、CD8、CD22、CD79a、CD79b和CD66d的ITAM。在一些实施方案中,TCR样CAR中的细胞质信号传导分子含有细胞质信号传导结构域、其部分或衍生自CD3 δ 的序列。

[0360] 在一些实施方案中,所述方法包括过继性细胞疗法,由此将表达所提供的抗体、TCR样CAR和/或TCR或其片段的基因工程化细胞施用于受试者。这样的施用可以以靶向的方式促进细胞的激活(例如,T细胞激活),使得疾病或病况的细胞靶向破坏。

[0361] 因此,所提供的方法和用途包括用于过继性细胞疗法的方法和用途。在一些实施方案中,所述方法包括将所述细胞或含有所述细胞的组合物施用于受试者、组织或细胞,诸如存在疾病、病况或病症的风险或疑似患有所述疾病、病况或病症的受试者、组织或细胞。在一些实施方案中,例如通过过继性细胞疗法,诸如过继性T细胞疗法,将细胞、群体和组合物施用于患有待治疗的特定疾病或病况的受试者。在一些实施方案中,将细胞或组合物施用于受试者,诸如患有该疾病或病况或存在该疾病或病况的风险的受试者。在一些方面,所述方法由此治疗,例如减轻疾病或病况的一种或多种症状。

[0362] 用于施用细胞以供过继性细胞疗法的方法是已知的并且可以与所提供的方法和组合物结合使用。例如,过继性T细胞治疗方法在例如Gruenberg等人的美国专利申请公开号2003/0170238;Rosenberg的美国专利号4,690,915;Rosenberg(2011)Nat Rev Clin Oncol.8(10):577-85)中有述。参见例如,Themeli等人,(2013)Nat Biotechnol.31(10):928-933;Tsukahara等人,(2013)Biochem Biophys Res Commun 438(1):84-9;Davila等人,(2013)PLoS ONE 8(4):e61338。

[0363] 在一些实施方案中,细胞疗法(例如,过继性细胞疗法,例如过继性T细胞疗法)通过自体转移来实施,其中从接受该细胞疗法的受试者或来源于该受试者的样品分离和/或以其他方式制备细胞。因此,在一些方面,细胞来源于需要治疗的受试者(例如,患者),并且将分离和处理后的细胞施用于同一受试者。

[0364] 在一些实施方案中,细胞疗法(例如,过继性细胞疗法,例如过继性T细胞疗法)通过同种异体转移来实施,其中从除了将要接受或最终接受该细胞疗法的受试者(例如,第一受试者)之外的受试者分离和/或以其他方式制备细胞。在这样的实施方案中,然后将细胞施用于相同物种的不同受试者,例如第二受试者。在一些实施方案中,第一受试者和第二受试者在遗传学上是相同的。在一些实施方案中,第一受试者和第二受试者在遗传学上是相似的。在一些实施方案中,第二受试者表达与第一受试者相同的HLA类别或超类型。

[0365] 在一些实施方案中,所述细胞、细胞群体或组合物所施用至的受试者为灵长类动物,如人类。在一些实施方案中,该灵长类动物为猴或类人猿。受试者可以是雄性的或雌性的,并且可以为任何适合的年龄,包括婴儿、少年、青少年、成人和老年受试者。在一些实施方案中,受试者是非灵长类哺乳动物,如啮齿动物。在一些实例中,患者或受试者是用于疾病、过继性细胞疗法和/或用于评估毒性结果如细胞因子释放综合征(CRS)的经过验证的动物模型。

附加试剂

[0366] 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。

[0367] 固体支持体可包含实际上任何不溶性或固体材料,并且通常选择不溶于水的固体支持体材料组合物。例如,固体支持体可包含以下成分或基本由以下成分组成:硅胶、玻璃(例如,可控孔玻璃(CPG))、尼龙、**Sephadex®**、**Sepharose®**、纤维素、金属表面(例如,钢、金、银、铝、硅和铜)、磁性材料、塑料材料(例如,聚乙烯、聚丙烯、聚酰胺、聚酯、聚偏氟乙烯(PVDF))等。根据实施方案使用的珠子的实例可包含允许珠子与核酸分子相互作用的亲和部分。固相(例如,珠子)可包含结合对的成员(例如,抗生物素蛋白、链霉亲和素或其衍生物)。例如,该珠子可为链霉亲和素包被的珠子,并且用于固定在珠子上的核酸分子可包含生物素部分。在一些情况下,每个多核苷酸分子可包含两个亲和部分,如生物素,来进一步稳定多核苷酸。珠子可包含用于固定核酸的附加特征,或可用于下游筛选或选择过程的附加特征。例如,珠子可包含结合部分、荧光标记或荧光猝灭剂。在一些情况下,珠子可为磁性的。在一些情况下,固体支持体为珠子。珠子的实例包括但不限于链霉亲和素珠子、琼脂糖珠子、磁珠、**Dynabeads®**、**MACS®**微珠、抗体偶联的珠子(例如,抗免疫球蛋白微珠)、蛋白A偶联的珠子、蛋白G偶联珠子、蛋白A/G偶联的珠子、蛋白L偶联的珠子、多核苷酸dT偶联的珠子、二氧化硅珠子、二氧化硅样珠子、抗生物素微珠、抗荧光染料微珠和BcMag™羧基末端磁珠。珠子或粒子可为可膨胀的(例如,聚合物珠子,诸如Wang树脂)或不可膨胀的(例如,CPG)。在一些实施方案中,固相基本上为亲水的。在一些实施方案中,固相(例如,珠子)基本上为疏水的。在一些实施方案中,固相包含结合对的成员(例如,抗生物素蛋白、链霉亲和素或其衍生物),并且基本上为疏水的或基本上为亲水的。在一些实施方案中,固相包含结合对的成员(例如,抗生物素蛋白、链霉亲和素或其衍生物),并且具有每mg固体支持体大于约1350皮摩尔游离捕获剂(例如,游离生物素)的结合能力。在一些实施方案中,包含结合对成员的固相的结合能力为每mg固体支持体大于800、900、1000、1100、1200、1250、1300、1350、1400、1450、1500、1600、1800、2000皮摩尔的游离捕获剂。适用于本发明的珠子的其他实例为金胶体或珠子,诸如聚苯乙烯珠或二氧化硅珠。可使用基本上任何珠半径。珠子的实例可包括半径范围为150纳米至10微米的珠子。还可使用其他大小。

[0368] 本文所公开的方法和试剂盒可包括使用一种或多种缓冲液。缓冲液的实例包括但不限于洗涤缓冲液、连接缓冲液、杂交缓冲液、扩增缓冲液和逆转录缓冲液。在一些实施方案中,杂交缓冲液为可商购获得的缓冲液,诸如TMAC Hyb溶液、SSPE杂交溶液和ECONO™杂交缓冲液。本文所公开的缓冲液可包含一种或多种去污剂。

[0369] 本文所公开的方法和试剂盒可包括使用一种或多种载体。载体可增强或改善本文所公开的一种或多种反应(例如,连接反应、逆转录、扩增、杂交)的效率。载体可降低或防止分子或其任何产物(例如,多核苷酸和/或扩增子)的非特异性损失。例如,载体可通过吸附至表面而降低多核苷酸的非特异性损失。载体可降低多核苷酸对表面或基底(例如,器皿、Eppendorf管、移液管尖端)的亲和力。或者,载体可提高多核苷酸对表面或基底(例如,珠子、阵列、玻璃、载片、芯片)的亲和力。载体可保护多核苷酸免于降解。例如,载体可保护RNA分子不受核糖核酸酶作用。或者,载体可保护DNA分子不受DNA酶作用。载体的实例包括但不限于多核苷酸如DNA和/或RNA,或多肽。DNA载体的实例包括质粒、载体、聚腺苷酸化DNA和DNA多核苷酸。RNA载体的实例包括聚腺苷酸化RNA、噬菌体RNA、噬菌体MS2 RNA、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。在一些实施方案中，载体来自细菌、酵母或病毒。例如，载体可为来源于细菌、酵母或病毒的多核苷酸或多肽。例如，载体为来自枯草芽孢杆菌 (*Bacillus subtilis*) 的蛋白质。在另一个实例中，载体为来自大肠杆菌 (*Escherichia coli*) 的多核苷酸。或者，载体为来自哺乳动物 (例如，人类、小鼠、山羊、大鼠、牛、绵羊、猪、狗或兔)、禽、两栖动物或爬行动物的多核苷酸或肽。

[0370] 本文所公开的方法和试剂盒可包括使用一种或多种对照剂。对照剂可包括对照多核苷酸、非活性酶、非特异性竞争剂。或者，对照剂包括亮杂交、亮探针对照、核酸模板、掺加对照、PCR扩增对照。PCR扩增对照可为阳性对照。在其他情况下，PCR扩增控制物为阴性对照。核酸模板对照可为已知浓度。对照剂可包含一种或多种标记。

[0371] 掺加对照可为向反应或样品中添加的模板。例如，掺加模板可添加到扩增反应中。掺加模板可在第一扩增循环后的任何时间添加到扩增反应中。在一些实施方案中，掺加模板在循环数2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45或50后添加到扩增反应中。掺加模板可在最终扩增循环前的任何时间添加到扩增反应中。掺加模板可包含一种或多种核苷酸或核酸碱基对。掺加模板可包含DNA、RNA或其任意组合。掺加模板可包含一种或多种标记。

[0372] 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都可用所公开方法的任何具体实施方案或实施方案的组合来进行，并且具体涉及每个这样的组合，并且应被认为是已公开的。

[0373] 虽然本文中已经示出并描述了本文所述的一些实施方案，但这些实施方案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，本文中所述的实施方案的各种替代方案可用于实施本文所述的方法。

[0374] 除非另有解释，本文使用的所有技术和科学术语具有与本发明所属领域的普通技术人员通常所理解的相同的含义。下列参考文献含有可用于本文的方法和组合物的实施方案：The Merck Manual of Diagnosis and Therapy, 第18版, Merck Research Laboratories出版, 2006 (ISBN 0-9119102); Benjamin Lewin, Genes IX, Jones&Bartlett Publishing出版, 2007 (ISBN-13:9780763740634); Kendrew等人(编), The Encyclopedia of Mol. Biology, Blackwell Science Ltd.出版, 1994 (ISBN 0-632-02182-9); 以及Robert A. Meyers(编), Mol. Biology and Biotechnology: a Comprehensive Desk Reference, VCH Publishers出版, Inc., 1995 (ISBN 1-56081-569-8)。

[0375] 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)。

实施例

[0376] 包括以下实施例, 以便进一步描述本公开内容的一些方面, 其不应被用于限制本发明的范围。

实施例1-BCR和TCR的高通量鉴别

[0377] 肿瘤浸润淋巴细胞 (TIL) 对于抗癌免疫应答是至关重要的, 但由于其不可预测的丰度、表型和功能, 研究具有挑战性。本发明人开发了在不需细胞分选、刺激或培养的情况下进行深度 TIL 表征的方法。基于乳液的单细胞条码化方法从数百万个输入细胞中的淋巴细胞中捕获天然配对的 B 细胞和 T 细胞受体 (BCR 和 TCR) 序列。与先前的方法相比, 所回收的可变区是全长的, 并且可以伴有额外的 mRNA 和蛋白质靶标。在处理来自卵巢腺癌的 400,000 个未分选细胞之前, 用来自健康人血液中的 3 百万个 B 细胞和来自 HIV 患者的 350,000 个 B 细胞验证了该方法, 回收来自超过 11,000 个 TIL 的配对 BCR 和 TCR。我们的研究结果代表了任何配对 BCR 或 TCR 组库的最深采样, 以及在乳液中从单细胞同时进行 RNA 测序和蛋白质测量的首次演示。

实施例2-单细胞的条码化

[0378] 免疫组库的测序在基础免疫学、自身免疫性疾病、传染性疾病和肿瘤学中具有广泛的应用。许多研究已经探讨了循环血液中的 BCR 和 TCR 多样性, 但对 TIL 的免疫受体的兴趣日益增加, 该免疫受体的功能与癌症生长或退化高度相关, 但可变且通常未得到表征。达到 TIL 的更好理解的关键步骤是其 BCR 和 TCR 的回收和功能表征, 因为这可允许鉴别新的肿瘤相关抗原。迫切需要肿瘤抗原来开发癌症疫苗、了解检查点抑制剂的作用以及在实体瘤中推进嵌合抗原受体 T 细胞 (CAR-T) 疗法。然而, 尽管数十年来在免疫测序方面取得了技术进步, 但研究均未在不进行使所观察的组库受限并产生偏差的步骤——体外培养或细胞分选的情况下从异质样品如肿瘤中回收全长、天然配对的 BCR (重链和轻链) 和 TCR (α 和 β 链)。由于原代未培养的免疫细胞可含有比体外刺激的细胞少 100 倍的受体 RNA, 所以技术挑战特别大。为了允许对来自复杂异质样品的天然配对 BCR 和 TCR 进行综合分析, 开发了基于微流体乳液的方法以供大量单细胞的平行分离和 DNA 条码化 (图 4A-4D)。在单个约 65 皮升的乳液

小液滴中每小时分离高达一百万个细胞。细胞在小液滴内裂解,靶mRNA用靶特异性引物进行逆转录,并且两步式DNA条码化过程将分子特异性条码和小液滴特异性条码二者附接至cDNA。在随后的回收和下一代测序后,双重条码化策略允许将序列读取簇集到其来源的分子和细胞二者中。这允许对错误和扩增偏差进行大量校正、在mRNA水平和细胞水平下进行克隆计数、进行重链同种型测定,并且重要的是,同时以极高的通量回收BCR和TCR的全长、天然配对的V(D)J序列。

实施例3-从健康血液样品大规模回收B细胞 V_HV_L 对

[0379] 首先开发了该技术,并用通过阴性珠子富集从健康志愿者的外周血分离的3百万个B细胞来评估配对能力和通量。将乳液分成六个单独的级分,其并行处理且在测序之前没有再混合。以每个小液滴0.2个细胞装载乳液,得到约90%被占据的小液滴含有单细胞的泊松期望值(Poisson expectation),这与乳液小液滴的观察一致(图5A)。在破乳和另外的文库处理步骤之后,用Illumina MiSeq进行配对末端325+300bp测序。为了处理测序数据,将小液滴和分子条码一起用于从每个原始mRNA分子收集PCR复制读取,并且确定每个mRNA的共有序列,仅保留由至少两个读取构建的mRNA序列。将正向和反向读取拼接以生成包含5' UTR、完整V(D)J序列和足以用于同种型测定的恒定区的全长产物。重排的免疫球蛋白重链和轻链序列用IMGT High-VQuest和/或IgBLAST进行注释。

[0380] 所得数据集含有324,988个与至少一个重链(V_H)和一个轻链(V_L)mRNA相关联的小液滴条码,如通过重链簇集分析所估计的,存在229,869个不同的 V_H 克隆谱系。由于该原始数据集包括来自多细胞以及单细胞小液滴的数据,因此通过过滤掉与非一致的重链或轻链V(D)J序列连接的小液滴条码来富集来自单细胞小液滴的数据(图5B)。该步骤因典型免疫组库的高度多样性而变得可能,其中来自两个随机细胞的 V_H 或 V_L mRNA将几乎从不匹配。所得到的富集数据集包含259,368个 V_HV_L 小液滴条码,并含有182,745个 V_H 克隆谱系,充分地代表迄今为止配对免疫组库的最深采样。由于克隆相关细胞应该在其 V_HV_L 配对中显示一致性,因此直接估计通过鉴别克隆扩增的发生率的 V_HV_L 配对的精确度。鉴别出2,604个 V_H 克隆,其在超过一个乳液级分中以高置信度被观察到为克隆扩增的细胞。整个级分中 V_L 序列与这些 V_H 克隆配对的一致性非常高,并且显示出96.1%的配对精确度,从而允许在259,368个 V_HV_L 对的整个过滤数据集中具有高置信度。交叉级分的 V_H 和 V_L 序列不变地与每个级分中的不同小液滴和分子条码相关联,因此不代表文库交叉污染。该分析可能低估配对精确度,因为已知一些B细胞表达多个轻链。259,368个过滤的 V_HV_L 小液滴条码或 V_HV_L 对”中的75.0%含有IgM和/或IgD(图5C),考虑到原始B细胞的典型IgM⁺IgD⁺表型,IgM和IgD不出意料地经常一起观察到。还发现较低但仍大量的IgA(18.3%)和IgG(6.6%) V_HV_L 对。所有 V_H 同种型与Ig κ 或Ig λ 以约3:2的比例进行配对。在182,745个 V_H 克隆谱系中,以两种方式评估克隆扩增:与克隆相关的小液滴条码的数目和跨乳液级分克隆的观察结果。在多个小液滴条码中观察到的克隆可以反映克隆扩增或多条码小液滴,在条码向小液滴中的初始 $\lambda=1$ 泊松分散的情况下,其预计存在于约37%的小液滴中。然而,由>8个小液滴条码代表的任何克隆均可能被真正扩增(单个小液滴中的泊松概率<10⁻⁶)。虽然在超过一个级分中观察到总体6.0%的克隆,但对于在超过8个小液滴条码中观察到的克隆(总体为0.7%),其中的99%都在超过一个级分中观察到。100个最常见的克隆(各自30-137个小液滴条码,图5D)全部在六个级分中的至少五个中观察到。因此,条码计数和独立级分分析的组合允许在非扩增克隆的巨大背景中

检测罕见的扩增谱系。然而值得注意的是,即使是最丰富的扩增克隆也仅存在于不到千分之一的细胞,这例证了人类外周免疫组库的巨大的多样性。

[0381] 配对内每个 V_H 和 V_L 链的捕获mRNA的数目作为表达水平的估计值(图5E)。通常每个小液滴条码捕获小于10个重链(平均2.0个)和轻链(平均4.0个)mRNA,观察到每个细胞具有数十至数百个捕获的重链和轻链mRNA的较小的小液滴条码群体,该细胞几乎完全来自表达IgG和IgA的细胞。有趣的是,配对内 V_H 和 V_L 突变的程度在每个同种型(例如,IgG的 V_H 与 V_L)内和同种型(例如,IgG与IgM)之间均强烈相关。此外,IgG和IgA对在其 V_H 和 V_L 链中几乎基本全部突变,而IgM和IgD对大部分显示很少的 V_H 或 V_L 突变。这些结果与B细胞激活导致IgM和IgD类别转换成IgG或IgA的机制、增加的免疫球蛋白表达和影响细胞中的重链基因座和轻链基因座的体细胞超突变一致。除了高度突变的 V_H 链倾向于与高度突变的 V_L 链配对的这一观察以外,该方法还能够从静息的人B细胞组库生成大量全长、天然配对的BCR。

实施例4-从HIV elite控制器回收已知的低频率 V_HV_L 对

[0382] 作为测定的配对灵敏度和准确度的进一步验证,对样品进行处理,其中几个罕见(10,000个细胞中<1个细胞)的天然 V_HV_L 配对是已经公开已知的。获得来自HIV elite控制器患者的外周B细胞,近年来针对展示HIV中和活性的抗体大量筛选了该HIV elite控制器的记忆B细胞。处理350,000个B细胞以生成总共38,620个过滤的 V_HV_L 对。有趣的是,该个体显示出比先前的健康样品(图6A)或典型健康外周B细胞组库更大的IgG比例。将来自该数据集的 V_H 序列与来自该个体(包括PGT121)的所有报道的广谱中和抗体(bNAb)进行比较,发现八个接近或相同的 V_H 序列,表明该bNAb家族代表少于0.03%的循环B细胞。至关重要的是,与这些重链配对的所有轻链均属于所预期且类似罕见的bNAb谱系,其展示出与先前报道相同的Ig λ -V3-21/J3重排和标志性三联密码子插入,这支持我们方法的高精确度和高灵敏度。此外,在来自该个体的所有已知和新生成的PGT121样 V_HV_L 对的系统树上(图6B), V_H 和 V_L 树显示出惊人相似的拓扑结构,其中配对的 V_H 和 V_L 序列占据镜样位置,这可能反映了共享系统发生史。本文发现的变体对很符合这个规则。有趣的是,两个公开的抗体PGT122和PGT123表现为例外;没有发现对这两种配对的支持,反而发现了PGT122 V_H :PGT123 V_L 样配对和PGT123 V_H :PGT122 V_L 样配对,对应于原始报告中的未经证实的配对。合成了编码8个新型PGT样 V_HV_L 对的完整V(D)J区的DNA,将抗体表达为完全IgG并测试其中和多个HIV假毒株的能力(图6C)。该抗体表达良好,并且均显示出对该病毒的强中和活性,证明了我们的方法在从相关生物样品快速生成天然配对的功能性抗体变体方面的实用性。

实施例5-来自肿瘤浸润淋巴细胞的B细胞和T细胞受体对

[0383] 在针对配对受体的高通量回收验证的乳液条码化之后,从肿瘤直接回收免疫受体。取得蛋白酶解离的切除的卵巢腺癌样品,并使400,000个未分选的细胞进入乳液中。对样品单独等分试样的CD3/CD19染色表明材料中相当数目的浸润B细胞(约5%)和浸润T细胞(约20%)。乳液中的单细胞分散与纯化的细胞相似,尽管可见一些有限的凝集,并且考虑到样品的细胞类型异质性,小液滴内的细胞大小和形状的广泛变化符合预期(图7A)。

[0384] 将靶向T细胞受体 α 和 β 链的恒定区的引物一起与之前使用的BCR引物一起使用,并且随后的测序和严格过滤回收了数千个与BCR或TCR产物连接的小液滴条码。为了评估单细胞精确度,对小液滴条码内的四个靶基因座(V_H 、 V_L 、 $V\alpha$ 、 $V\beta$)的所有可能组合进行计数(图7B)。

[0385] 绝大多数(97.9%)的具有超过一个靶链的小液滴条码包含生物学上预期的BCR $V_H^+V_L$ 或TCR $V\alpha^+V\beta$ 配对,其中仅含有2.1%的混合BCR-TCR组合。由于产物的条码化相对于靶链是无偏倚的,因此该结果允许所得到的6,056个BCR V_HV_L 和5,217个TCR $V\alpha V\beta$ 对具有高置信度。尽管BCR和其他同种型全都存在IgG(仅IgE<0.05%),但BCR与其他同种型相比显示出明显的IgG优势(>80%)(图7C)。 κ 和 λ 轻链以与外周血数据集相似的比例存在。

[0386] 与外周血类似,在BCR同种型与 V_H 和 V_L 链二者的突变水平之间观察到相关性,其中IgG和IgA对显示出比IgD和IgM更多的 V_H 和 V_L 突变,并且在每个同种型内的 V_H 与 V_L 之间显示出突变的一般相关性(图7D)。有趣的是,尽管IgD、IgM和IgA对在肿瘤与外周血数据集之间显示出非常相似的突变分布(图5F),但肿瘤IgG级分还含有相当大比例的在外周血中未观察到的较少突变序列或无突变序列。对于TCR和含有IgD的BCR,每个小液滴条码所观察到的捕获mRNA的数目与来自外周血的BCR结果相似(图7E,每个小液滴条码几乎都<10)。形成鲜明对比的是,肿瘤衍生的IgM、IgA和IgG对显示出提高10至100倍的平均表达水平,其中数百或数千个靶mRNA被捕获在许多小液滴条码中。然后评估了所捕获的TIL-TCR和BCR组库的多样性(图7F)。在5,217个总TCR对中观察到2,423个不同的TCR β 克隆。七个克隆以>1%的频率存在,其中频率最高的克隆代表所有小液滴条码中的16.9%。在6,056个总BCR对中观察到1,518个不同的重链克隆,其中15个克隆的频率>1%但均未>5%。虽然这代表比健康的外周BCR组库(其中克隆均未以大于0.06%的频率存在)明显更受限的多样性,但在肿瘤样品中存在如此多的类别转换的、突变的和高度表达的克隆表明需要深度和敏感的采样方法来进行TIL表征。这些方法允许同时从B细胞和T细胞二者中快速找回大量的TIL免疫受体对,而无需事先分选或外源地激活所确定的TIL群体。

实施例6-捕获其他感兴趣的表型标志物

[0387] 通过小液滴条码化的受体链配对潜在地允许捕获除了免疫受体之外的另外的靶标。为了探讨这种可能性,通过磁珠富集将健康T细胞分离为 $CD4^+$ 群体和 $CD8^+$ 群体,并将每种类型的20,000个细胞输入至具有靶向TCR α 和 β 链以及 $CD4$ mRNA和 $CD8$ mRNA的引物的单独乳液运行中。测序后,含有来自 $CD4^+$ 分离细胞的TCR $V\alpha$ 和 $V\beta$ (-TCR对)的3,861个小液滴条码中的47.0%与 $CD4$ mRNA相联系,而仅0.3%与 $CD8$ mRNA相联系。相比之下,来自 $CD8^+$ 分离细胞的2,235个TCR对的50.6%与 $CD8$ mRNA相关联,而仅有0.6%与 $CD4$ mRNA相关联。这表明与先前的报道相似,基于mRNA的途径用于细胞表型的特异性较高但灵敏度有限。相比之下,诸如细胞表面受体等蛋白质通常以远多于其编码mRNA的数目(每个细胞1,000-100,000个蛋白质)存在,这可能使得它们更容易检测以及可能与细胞表型更直接地相关。为了测量每个细胞上的靶蛋白质水平,将定制的寡核苷酸DNA标记与抗人 $CD4$ 和 $CD8$ 抗体缀合,并将标记的抗体与未分离的 $CD4^+$ 和 $CD8^+$ T细胞混合物一起温育,然后将30,000个细胞输入乳液中(图8A)。该DNA标记带有抗体特异性序列标签和分子条码以及与扩增的小液滴条码的序列互补性,从而允许乳液小液滴条码化和分子计数与针对mRNA类似地进行。同时靶向DNA标记以及TCR、 $CD4$ 和 $CD8$ mRNA。测序和过滤后,3,682个小液滴条码被鉴别为具有高置信度的TCR $V\alpha V\beta$ 。与先前的实验一致,基于mRNA,可将大约一半(52%)的TCR对分配为 $CD4$ 或 $CD8$ 状态(图8B)。然而,基于蛋白质状态,可将超过95%的小液滴条码分配为 $CD4$ 或 $CD8$,其中对于每个小液滴的平均分子计数, $CD4/8$ 蛋白质(平均值20.5)明显高于 $CD4/8$ mRNA(平均值1.0)。mRNA与蛋白质信号之间的一致性很高(图8C):既给出mRNA判定又给出蛋白质判定的小液滴中的

96.0%是一致的。在一些罕见的情况下,检测到CD4蛋白质和CD8蛋白质二者,这可能是含有两个或更多个小液滴的结果。乳液条码化首次允许将单细胞免疫受体与感兴趣的mRNA和蛋白质标志物直接联系,所有这些均以高通量进行。有理由将这种途径应用于具有扩展的免疫肿瘤相关标志物组如抗PD-1和抗CTLA-4的TIL。

实施例7-人类样品

[0388] 在个人基因组计划(Personal Genome Project)的批准下收集了用于健康组库验证的血液样品。用于HIV bNAb实验的PBMC从供体17中获得,供体17是来自IAVI Protocol G队列的HIV-1感染供体。根据由卢旺达国家伦理委员会(the Republic of Rwanda National Ethics Committee)、埃默里大学机构审查委员会(the Emory University Institutional Review Board)、赞比亚大学研究伦理委员会(the University of Zambia Research Ethics Committee)、查林十字研究伦理委员会(the Charing Cross Research Ethics Committee)、UVRI科学与伦理委员会(the UVRI Science and Ethics Committee)、新南威尔士大学研究伦理委员会(the University of New South Wales Research Ethics Committee)、圣文森特医院和东悉尼地区卫生服务机构(St.Vincent's Hospital and Eastern Sydney Area Health Service)、肯雅塔国家医院伦理研究委员会(Kenyatta National Hospital Ethics and Research Committee)、开普敦大学研究伦理委员会(University of Cape Town Research Ethics Committee)、国际机构审查委员会(the International Institutional Review Board)、玛希隆大学伦理委员会(the Mahidol University Ethics Committee)、沃尔特里德陆军研究所(WRAIR)机构审查委员会(the Walter Reed Army Institute of Research(WRAIR) Institutional Review Board)和象牙海岸—国家生命科学和健康伦理”委员会(CNESVS)(the Ivory Coast Comite—National d'Ethique des Sciences de la Vie et de la Sante”(CNESVS))批准的临床方案,在书面知情同意的情况下收集所有人类HIV样品。根据IRB批准的方案,在知情同意的情况下从Conversant Biologics获得来自单个供体的冷冻保存的、解离切除的卵巢腺癌。

实施例8-细胞制备

[0389] 为了研究3百万个健康B细胞,将50mL血液抽取到具有肝素钠(BD)的真空采血管CPT细胞制备管中,以 $1800 \times g$ 离心20min,在细胞制备缓冲液(补充有2%胎牛血清和2mM EDTA的1x PBS)中洗涤两次,使用 $200 \times g$ 旋转来去除血小板,并将所得PBMC在RPMI-1640培养基(Life Technologies)+20%胎牛血清+10%DMSO中以 -80°C 冷冻保存,直至需要。在生成乳液之前,将PBMC解冻,在细胞制备缓冲液中洗涤两次并计数。根据制造商的说明书使用基于阴性选择的人类B细胞富集试剂盒(Stem Cell Technologies)来分离B细胞。使细胞通过 $20\mu\text{m}$ 细胞过滤器并在细胞制备缓冲液中稀释至 6.2×10^6 个细胞/mL(3百万B细胞实验)或 3.1×10^6 个细胞/mL(PGT供体和卵巢肿瘤实验)。

实施例9-乳液中的免疫受体条码化

[0390] 乳液生成平台由三个通过单个空气压缩机驱动的Mitos P-泵(Dolomite Microfluidics)组成,每个泵均具有Mitos流速传感器,以允许两个水相和一个亲氟油连续相计算机控制地流入亲氟涂覆的石英白云石Small 2-Reagent芯片。一个水性入口通道含有所需密度的细胞以产生期望的单位小液滴细胞占据水平,而第二水性通道含有裂解和反

应混合物,该混合物由AbPair™反应缓冲液和寡核苷酸(www.abvitro.com/catalog AV2070-1S和AV2080-1S)、5单位/ μL 的基于MuMLV的逆转录酶(Thermo Scientific)以及0.1单位/ μL 的Herculase II PCR聚合酶组成。使用100 μL Hamilton Microliter注射器以两次各约100 μL 的LR混合物的注射来过度装载100 μL 内径PEEK管样品环。使用100 μL Hamilton Gastight注射器将约110 μL 细胞悬浮液装载到约100 μL 、内径为0.2mm的FEP管环中。通过在伴有来自芯片中两个油通道的油流动的同时使水相以相同流速聚焦流动喷射穿过两试剂(2-Reagent)芯片而形成乳液。将离开芯片出口通道的乳液滴入在冷却块(cold block)上的0.2-mL PCR带管(Eppendorf)中,之后通过移液从管的底部移除过量的油,添加40 μL 覆盖溶液(25mM Na-EDTA, pH 8.0)并将管转移至标准热循环仪以供转录物标记反应。在45-min的逆转录(RT)步骤期间, RNA在42 $^{\circ}\text{C}$ 下用靶特异性RT引物进行逆转录,其中基于模板转换添加了含有随机分子条码的通用衔接子序列。在达到室温后,使乳液经历40次热循环(每个循环:82 $^{\circ}\text{C}$ 持续10 $^{\circ}\text{C}$, 65 $^{\circ}\text{C}$ 持续25秒)以进行小液滴条码模板的PCR扩增,将其在初始裂解和反应混合物中稀释至30,000cp/ μL ,从而在最终混合物中生成15,000cp/ μL 或每约65p1小液滴约1cp的浓度。小液滴条码的一端包含Illumina读取2(-P7")引物位点,而另一端匹配通用衔接子寡核苷酸的共同序列。因此,在PCR期间,模板转换的cDNA可以与扩增的小液滴条码链退火,并通过重叠延伸而剪接以产生含有靶标、分子条码和小液滴条码序列的全长产物。

实施例10-破乳、清理、下游PCR、合并和测序

[0391] 在热循环之后,通过移液移除覆盖溶液,并将40 μL 破乳溶液(1:1FC-40:全氟辛醇)与15 μL 裂解物澄清溶液(12.5 μL Qiagen蛋白酶,2.5 μL 0.5M Na-EDTA, pH 8.0)一起添加。颠倒10次以破坏乳液后,将混合物在50 $^{\circ}\text{C}$ 下温育15分钟并在95 $^{\circ}\text{C}$ 下温育3分钟以将蛋白酶灭活。在以15,000 $\times g$ 离心1min以分离水相后,将回收的物质严格纯化以除去寡核苷酸、试剂和过量的小液滴条码PCR产物。由于全长产物因RT引物的5'生物素化而含有生物素,因此可以通过在链霉亲和素珠子上进行清理而有效地将其与过量的小液滴条码PCR产物分离,从而尽可能减少下游PCR重组伪迹(在重叠延伸方法中的常见问题)。首先使用AMPure XP珠子(Agencourt),利用制造商的说明书以1:1的比例纯化第一产物,随后使用链霉亲和素珠子(New England Biolabs)同样利用制造商的说明书进行清理,然后在95 $^{\circ}\text{C}$ 的去离子水中洗脱,接着用AMPure XP珠子以1:1的比例进行第二次清理。然后将产物输入靶标富集PCR,其中对B细胞或T细胞受体靶标的恒定区具有特异性的引物与对小液滴条码序列的通用末端具有特异性的引物一起使用。根据制造商的说明书,该反向引物还含有用于MiSeq仪器上的多路测序的六碱基索引条码。因此,仅全长的、小液滴条码化的靶序列在该步骤中得以扩增。首先使用Q5 Hot Start聚合酶(New England Biolabs)在制造商推荐的条件下将所有靶标一起扩增7个循环:98 $^{\circ}\text{C}$ 10秒;64 $^{\circ}\text{C}$ 20秒;72 $^{\circ}\text{C}$ 15秒(包括在反应开始时进行的2分钟98 $^{\circ}\text{C}$ 聚合酶激活步骤)。随后以1.5:1的珠子:PCR比例进行AMPure XP清理。然后使用与之前相同的热循环条件进行分别针对每个链(V_H 、 V_L 、 $V\alpha$ 、 $V\beta$)的第二次七个循环,随后进行AMPure XP清理。然后进行具有相同热循环条件和5-15个循环(取决于通过qPCR判断的产率)的最终PCR以添加全长Illumina测序衔接子并生成足够的物质以供TapeStation D1000(Agilent)定量。然后合并文库并在V3 2x 300bp MiSeq平台(Illumina)上进行测序。

实施例11-对MiSeq平台的修改

[0392] BCR或TCR的完整可变V(D)J区的重建需要拼接两个配对末端Illumina读取。为了

改善这一过程,将2x 300bp试剂盒的正向读取延伸至325bp。由于免疫受体文库在恒定区引物位点具有有限的多样性,因此使用10%phiX掺加来减轻文库多样性有限的问题。

实施例12-对读取进行生物信息学处理的概述

[0393] 使用围绕pRESTO软件包(0.4版)建立的定制流程来处理Illumina MiSeq读取,以从每个小液滴生成mRNA分子的全长共有序列,用IgBLAST和/或IMGT/HighV-QUEST对其进行注释,并进一步比对、过滤并用自定义脚本和Change-O软件包处理以生成统计数据。

实施例13-读取处理和注释,同种型分配。

[0394] 利用pRESTO和Change-O软件包,用定制流程进行原始读取处理、V(D)J注释和克隆分配。简言之,对原始Illumina配对末端325+300bp读取进行质量控制、引物修剪,并由引物位点的模糊匹配来鉴别小液滴特异性条码(DB)和分子特异性条码(MB)。DB和MB一起唯一地指定起源分子,并且该唯一分子标识符(UMI)被用于将来自相同分子的一致PCR复制读取(最少两个)分组以生成每个mRNA序列的共有序列。通过经引物修剪的序列内的已知同种型特异性恒定区的模糊匹配来验证同型特异性引发。使用IgBLAST确定V(D)J种系区段和重排结构,并在适当的情况下用IMGT/HighV-QUEST来验证,通过Change-O和自定义脚本进行解析。通过如在Change-O中实施的在具有匹配的IGHV基因、IGHJ基因和接合长度的功能性V(D)J序列组内的单联簇集来分配克隆。对于3百万个循环细胞数据集,将使用对称化转换/颠换模型的加权克隆内距离4.0用作克隆内的最近邻距离截止值。

实施例14-小液滴免疫受体内容物过滤和配对保真度计算

[0395] 可以用这种条码化方法以以下两种方式评估从小液滴回收B细胞序列的精确度:使用小液滴内mRNA序列一致性,以及通过交叉级分配对一致性。在每个小液滴内,每个基因座捕获多个mRNA;从一个细胞表达的V(D)J序列应该一致。使用多重比对V(D)J区段的2%序列多样性(平均成对核苷酸差异 $\pi_{55} < 0.02$)的截止值来限定序列一致性,将每个小液滴存在超过一个生产性VDJ和一个生产性VJ序列生物信息学地标记为推定的免疫受体包含或多细胞占据。针对每个等位基因排除的小液滴构建重链和轻链共有序列,并将其用于克隆限定和交叉级分配对分析。对于3百万个循环B细胞数据集,每个 V_H 谱系与数据集中 $>20,000$ 个轻链克隆中的一个(理想情况下)相关联。在具有 $V_H V_L$ 对的259,368个免疫基因座排除的小液滴中,10,870个VDJ重链基因重排簇集存在于6个物理分隔的乳液级分的至少两个中。这些簇集代表扩增的谱系或相同VJ外显子的独立但相似的重排。当两个复制中的VDJ重排与一致VJ重排配对时,两个实验均独立地产生真阳性(对于2,604个具有更罕见重排的克隆,35,922个可能成对比较中有33,157个)。因此,每个复制的精确度为96.1% ($0.923^{0.5}$)。

实施例15-HIV bNAb候选序列的发现

[0396] 通过针对与已知bNAb HCDR3s、VDJ序列和使用tblastx、MUSCLE和PhyML从文献中选出的供体17谱系的相似性甄选我们的38,620个 $V_H V_L$ 对来发现针对HIV的新的天然配对的广泛中和抗体(BNAb),随后手动检查全V(D)J氨基酸序列的系统树以选择散布有已知bNAb序列的抗体候选物。

实施例16-HIV bNAb蛋白的表达和纯化

[0397] 合成抗体序列并将其克隆到先前描述的重链和轻链载体中。根据制造商的方案使用293fectin(Invitrogen)将重链和轻链质粒共转染(1:1比例)在293FreeStyle细胞中。转染后四天收获抗体上清液并通过蛋白-A亲和层析进行纯化。在用于进一步测定之前将纯化

的抗体缓冲液交换为PBS。

实施例17-假病毒生产和中和测定

[0398] 通过用HIV-1 Env表达质粒和Env缺陷的基因组骨架质粒(pSG3 Δ Env)转染293T细胞来生成假病毒。转染后72hr收获假病毒用于中和测定。使用单轮复制假病毒测定和TZM-B1靶细胞评估中和活性。简言之,将TZM-B1细胞接种在96孔平底板中。向该板添加假病毒,其在37°C下与抗体的连续稀释一起预温育1hr。一旦裂解并添加Bright-Glo萤光素酶底物(Promega),在感染后的72hr对萤光素酶报道基因表达进行定量。为了确定IC₅₀值,通过非线性回归拟合剂量-响应曲线。

实施例18-卵巢肿瘤靶链的鉴别

[0399] 在乳液中从卵巢解离的肿瘤组织同时捕获BCR和TCR后,使用如先前所述的分子和小液滴条码来过滤读取,但随后寻找四种可能的靶链类型(BCR V_H、BCR V_L、TCR V_α、TCR V_β)中每一种的存在。如果靶链由至少两个mRNA支持且每个mRNA至少有两个测序读取,则保留靶链。进一步分析仅含有BCR V_HV_L对或TCR V_αV_β对的所有小液滴条码。基于不同的CDR3氨基酸序列判定BCR重链和TCRβ链克隆。

实施例19-通过DNA标记的抗体染色在乳液中进行蛋白质检测

[0400] 单链的200bp DNA寡核苷酸被设计为含有独特的5bp抗原ID序列,并用5'氨基基团进行修饰(Integrated DNA Technologies)。根据制造商的方案,使用Thunder-Link试剂盒(Innova Biosciences)将小鼠单克隆抗人CD4(BioLegend,#300516)和CD8a(BioLegend,#301018)抗体与DNA寡核苷酸标签缀合。对于乳液前的细胞标记,将两百万个来自外周血的阴性选择的T细胞在400μL细胞缓冲液+2mM EDTA+0.05%叠氮化钠中稀释。将单链鲑鱼精子DNA添加至细胞,以使最终浓度为200μg/mL,并将细胞在室温下旋转五分钟。将CD4和CD8a DNA标记的抗体(各自至最终浓度为5nM)的混合物添加至细胞并在室温下温育30分钟。用细胞缓冲液+2mM EDTA+0.05%叠氮化钠+200μg/mL单链鲑鱼精子DNA将细胞洗涤三次。在进入乳液分析之前,将细胞重悬于细胞缓冲液+0.05%叠氮化钠中。将30,000个细胞用于乳液测序。

[0401] 实施例20-鉴别TCR抗原靶标

[0402] 在一些情况下,将天然配对的TCR链克隆到编码TCR链的表达载体和编码报道基因(诸如萤光素酶或在被APC激活后表达的荧光蛋白)的DNA构建体中。用编码感兴趣的配对TCR链的表达载体来转染工程化为不表达内源性TCR链的T淋巴细胞细胞系如Jurkat细胞。CD4或CD8共受体由T淋巴细胞细胞系表达。首先通过与同基因APC或肿瘤细胞(其为原代细胞或MHC匹配的细胞系)一起温育来筛选TCR。TCR反应性由报道基因或细胞因子的测量(如通过ELISPOT测量IFN γ)确定,并鉴别感兴趣的候选TCR。

[0403] 将感兴趣的候选的天然配对TCR链(在一些情况下,使用如上所述的初始筛选鉴别的感兴趣的天然配对TCR链)克隆到表达载体中。用编码感兴趣TCR的表达载体转染T淋巴细胞。在一些情况下,T淋巴细胞可能缺乏内源性TCR表达。单独地,通过用与衍生出TCR链的生物体匹配的MHC基因对不表达内源性MHC受体的APC细胞系(如K562细胞系)进行转染而生成人造抗原呈递细胞(aAPC)。aAPC也工程化有报道基因,如萤光素酶或荧光蛋白。将抗原表达载体的文库转染到抗原呈递细胞中。编码抗原的序列可以是感兴趣的肽、小基因构建体或整个cDNA编码序列的确切序列。或者,直接用感兴趣肽的文库对APC进行脉冲处理。

[0404] 将表达感兴趣的天然配对TCR链的T细胞与呈递抗原文库的aAPC共培养。T淋巴细胞通过TCR与MHC-肽复合体的结合而与aAPC的接合导致T细胞的激活和aAPC报道分子的激活。例如,TCR与MHC-肽复合体结合导致荧光素酶或荧光蛋白的aAPC表达。通过分选技术如荧光激活细胞分选(FACS)分离激活的aAPC。通过标准技术如用弱酸和反相HPLC(RP-HPLC)处理将肽与MHC蛋白质解离。进行质谱分析或测序以鉴别天然配对TCR链的抗原靶标。

[0405] 虽然本文中已经示出并描述了本发明的优选实施方案,但对本领域技术人员显而易见的是,这些实施方案仅以示例的方式提供。本领域技术人员在不脱离本发明的情况下现将想到多种变化、改变和替代。应当理解,本文中所述的本发明实施方案的各种替代方案可用于实施本发明。旨在以下权利要求限定本发明的范围,并由此涵盖这些权利要求范围内的方法和结构及其等同项。

测量肿瘤反应性

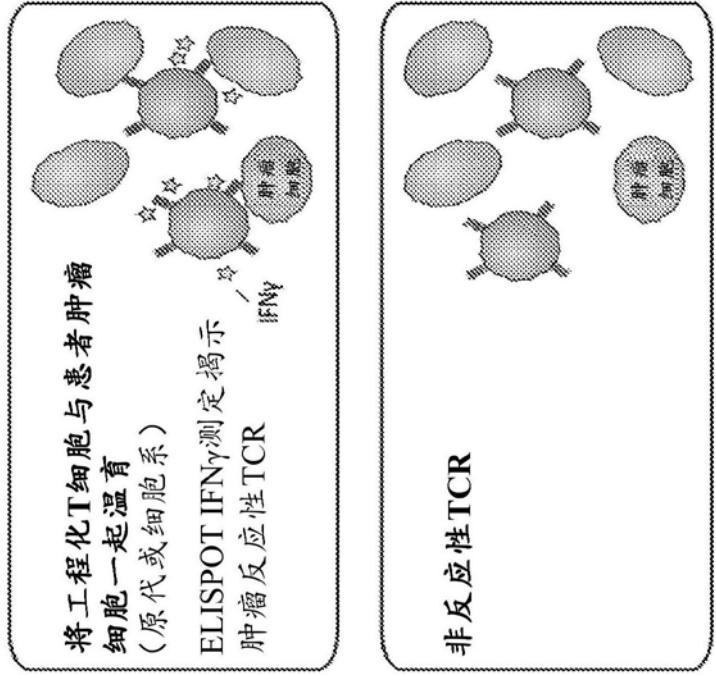


图1

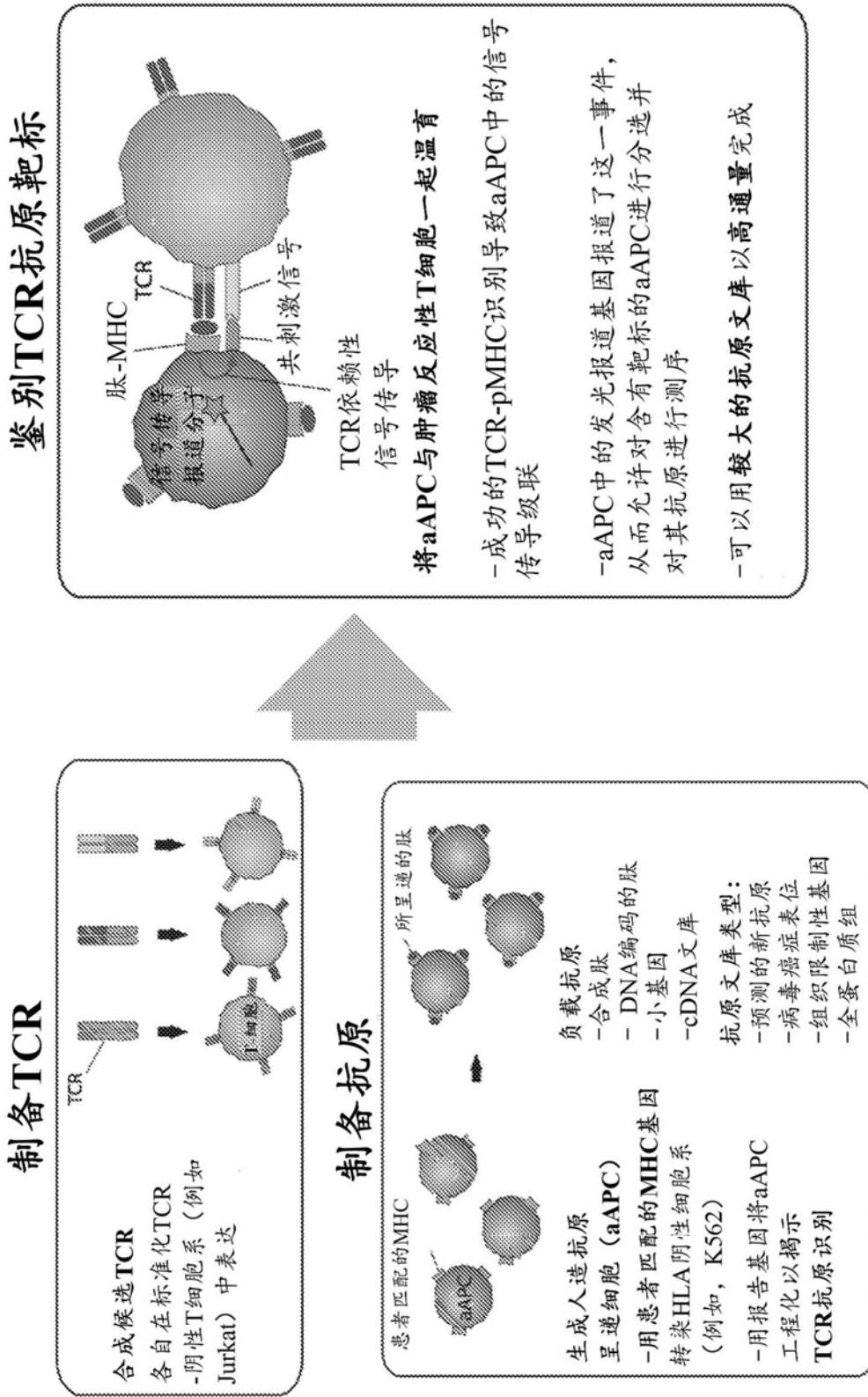
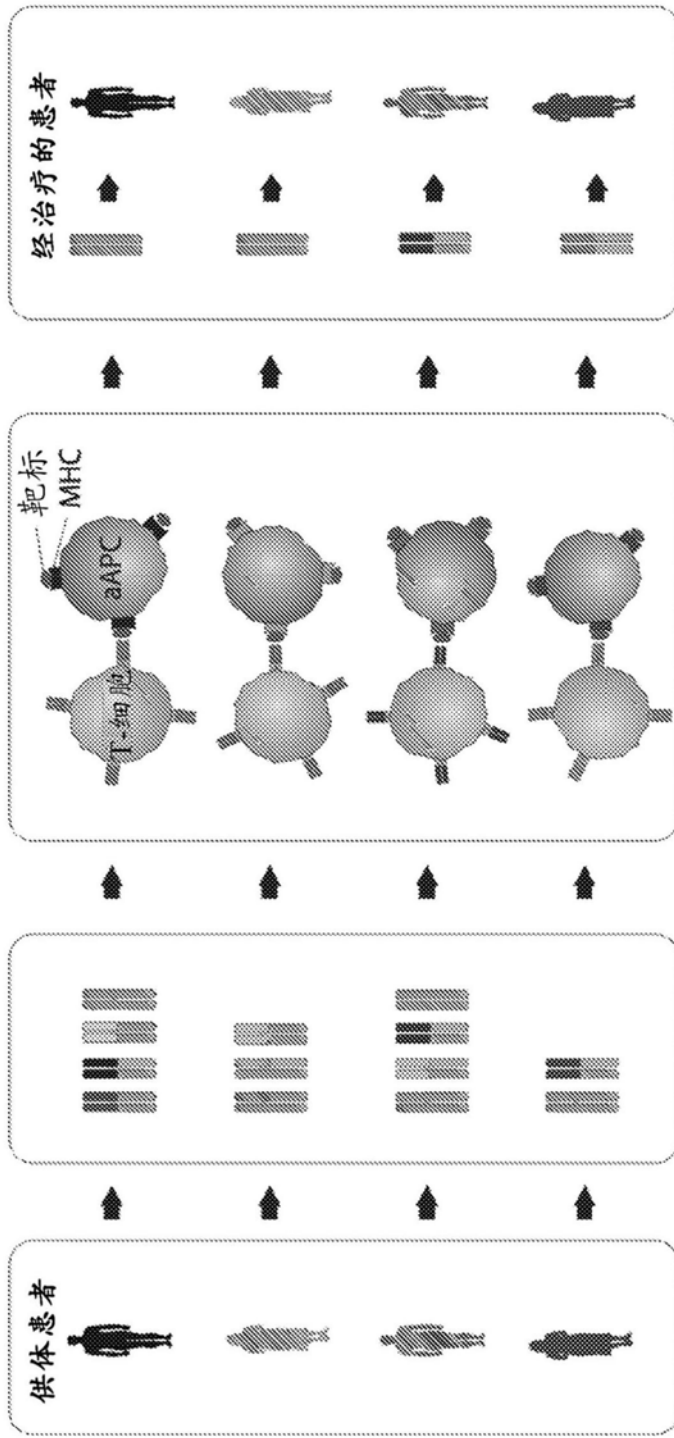


图2



具有针对高值抗原的广泛
群体覆盖的TCR免疫疗法

生成N-TIL-T细胞系和
匹配的HLA aAPC
鉴别针对选定抗原（例如，
ESO-NY-1）的阳性N-TIL-T

TIL TCR的Ab对发现
（共表达PD-1、LAG-3、
TIM-3）

从HLA代表性
群体开始

图3

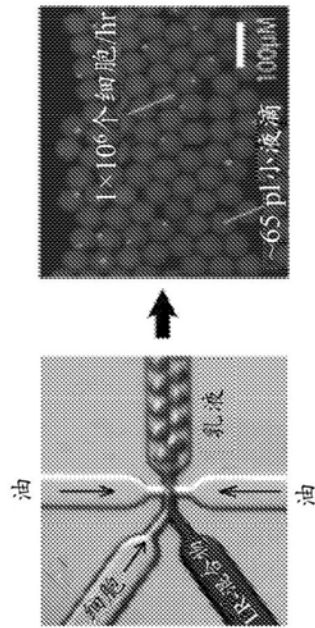


图4A

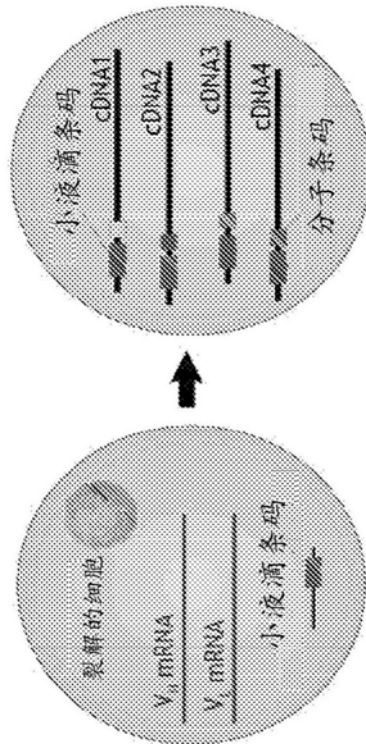


图4B

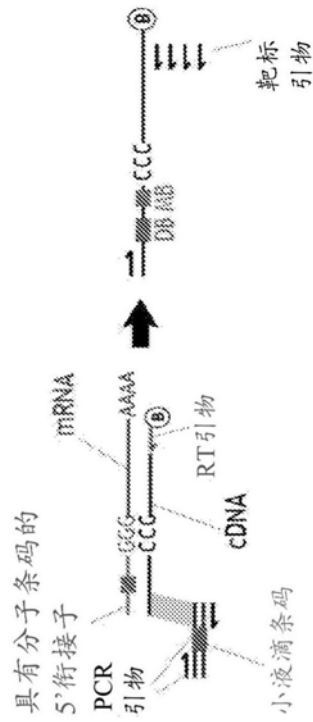


图4C

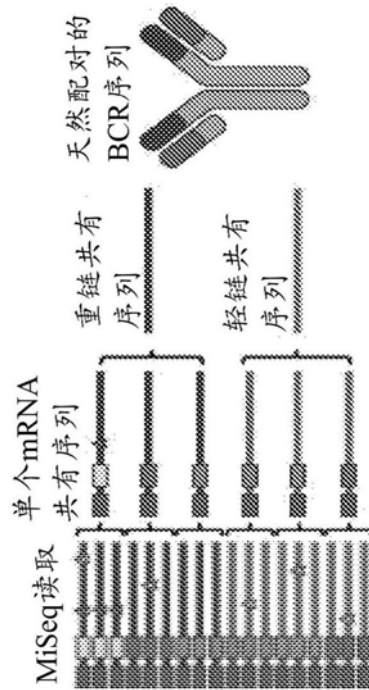


图4D

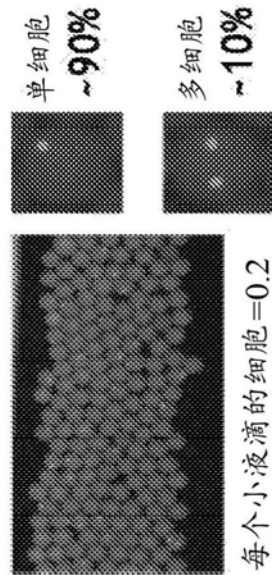


图5A



图5B

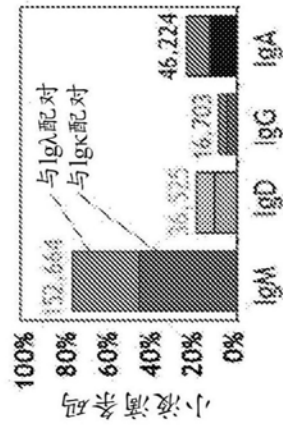


图5C

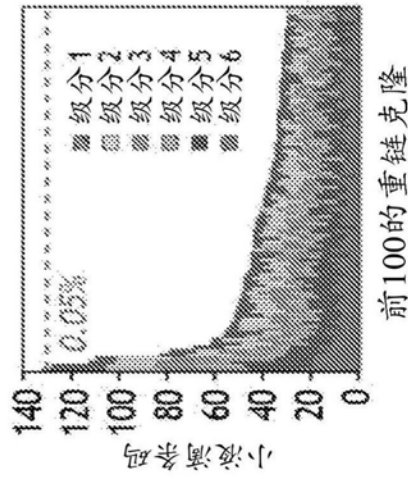


图5D

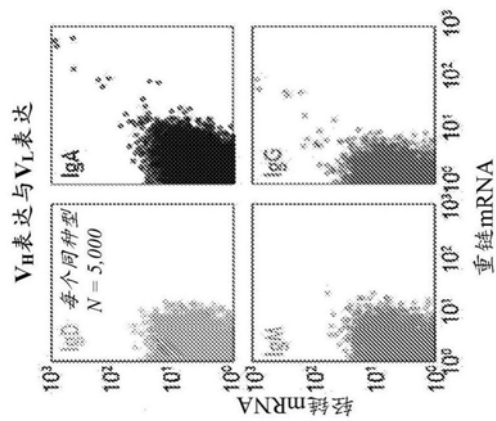


图5E

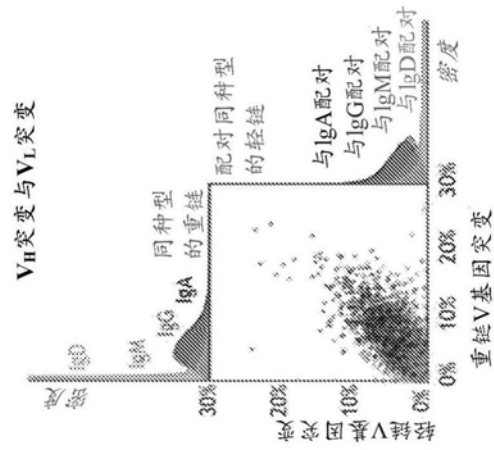


图5F

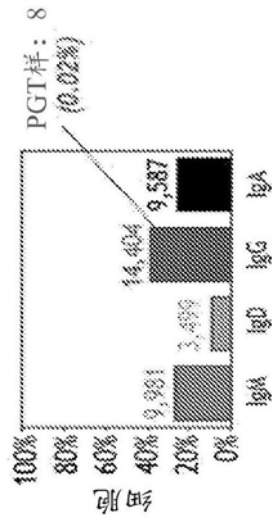


图6A

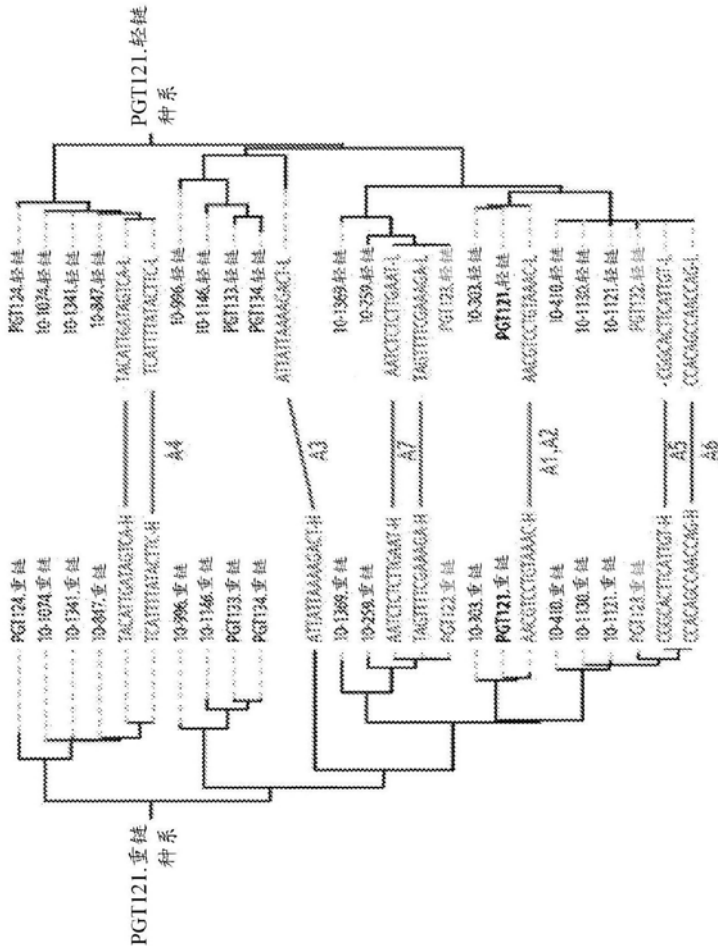
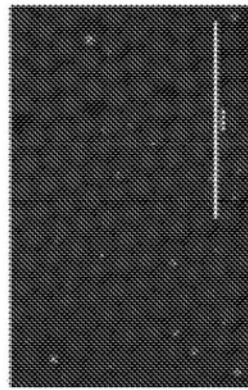


图6B

| WT | 抗体 | | | | | | | |
|----------|--------|-------|-------|-------|-------|-------|-------|-------|
| | PGT121 | A1 | A2 | A3 | A4 | A5 | A6 | A7 |
| BG505 | 0.037 | 0.019 | 0.020 | 0.051 | > 10 | > 10 | 0.051 | 0.004 |
| 92RH020 | 0.001 | 0.023 | 0.020 | 0.047 | 0.064 | > 10 | 0.015 | 0.005 |
| IAVI C22 | 0.001 | 0.005 | 0.009 | 0.015 | 0.053 | 0.433 | 0.014 | 0.003 |
| 94UG103 | 0.400 | 0.363 | 0.255 | 0.561 | > 10 | > 10 | > 10 | 0.033 |
| 92TH021 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 |
| BG505 | 0.052 | 0.015 | 0.018 | > 10 | > 10 | > 10 | 0.015 | 0.017 |
| 92RH020 | 0.014 | 0.009 | 0.018 | > 10 | > 10 | > 10 | > 10 | > 10 |
| IAVI C22 | 0.299 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 |
| 94UG103 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 |
| 92TH021 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 | > 10 |

图6C



来自卵巢肿瘤的
400,000个细胞

图7A

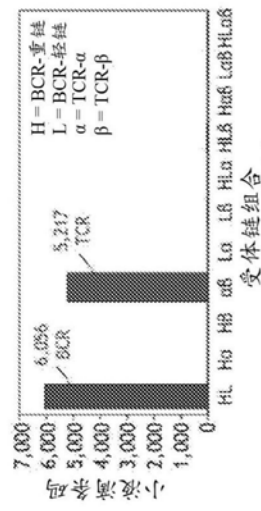


图7B

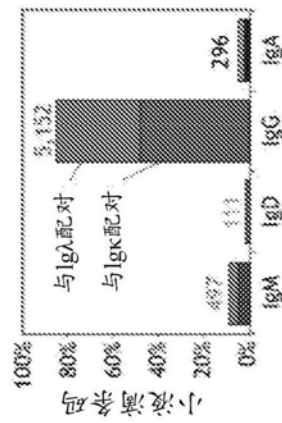


图7C

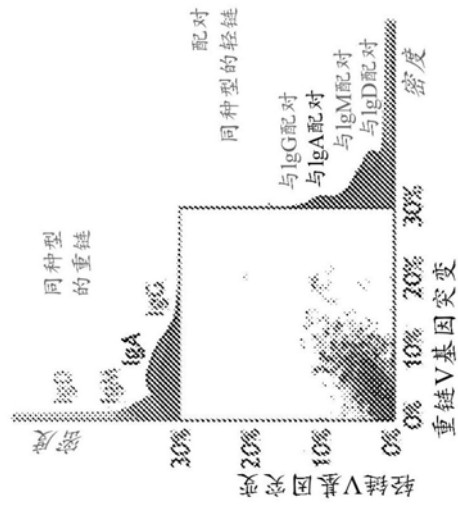


图7D

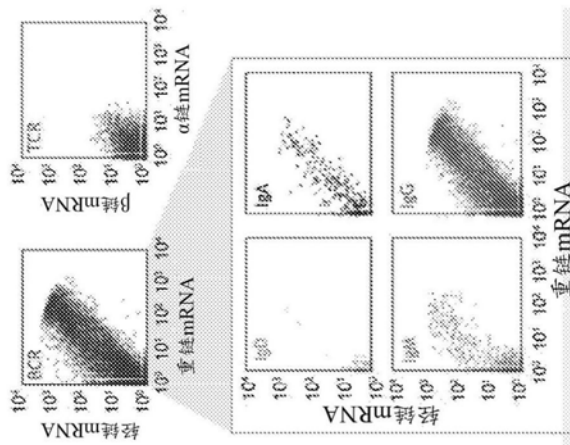


图7E

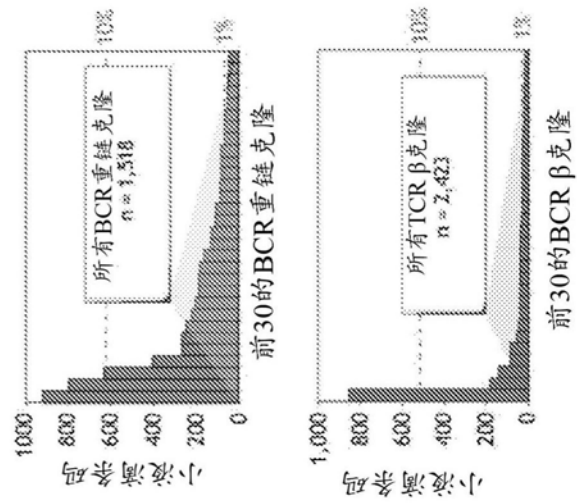


图7F

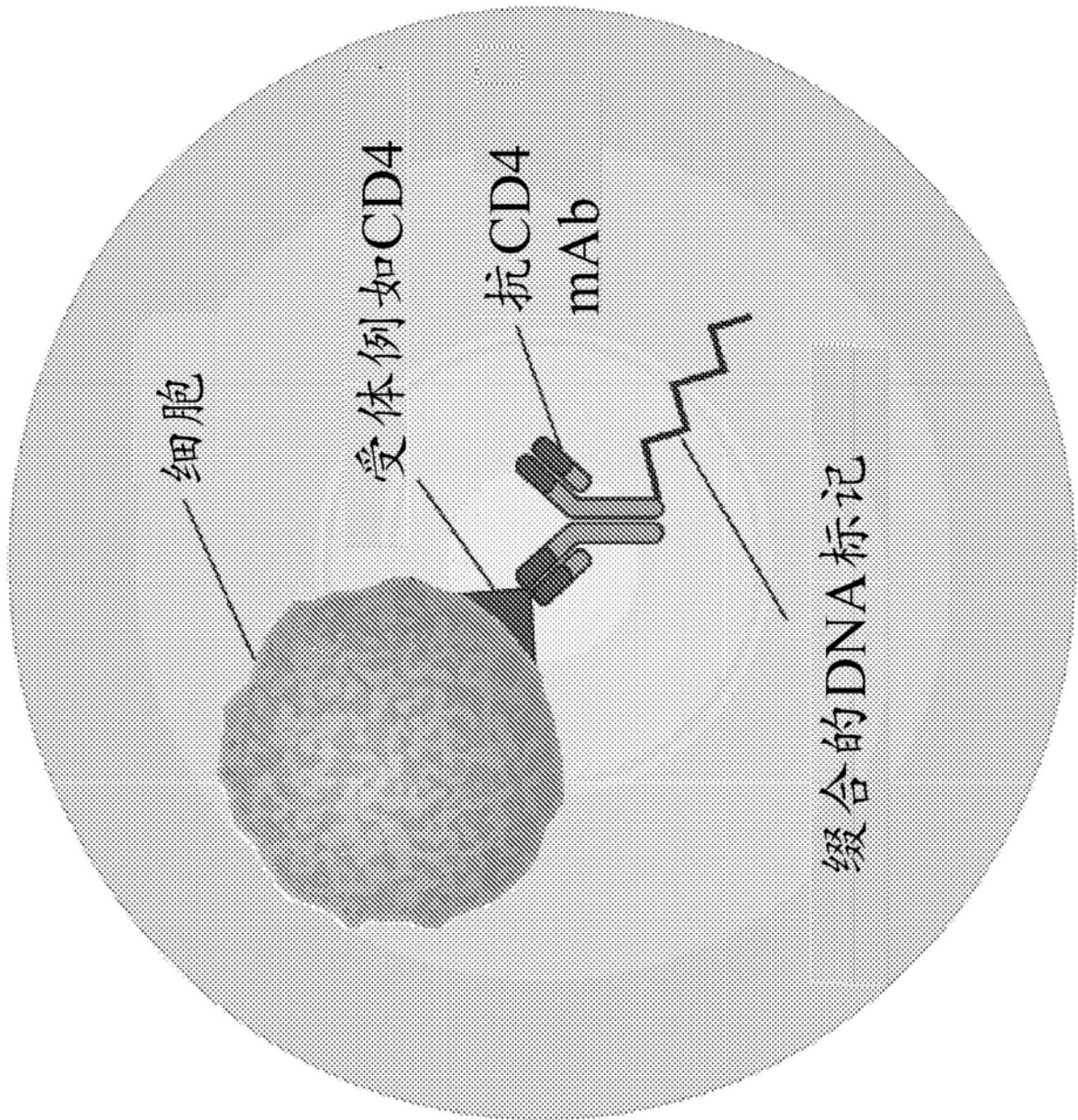


图8A

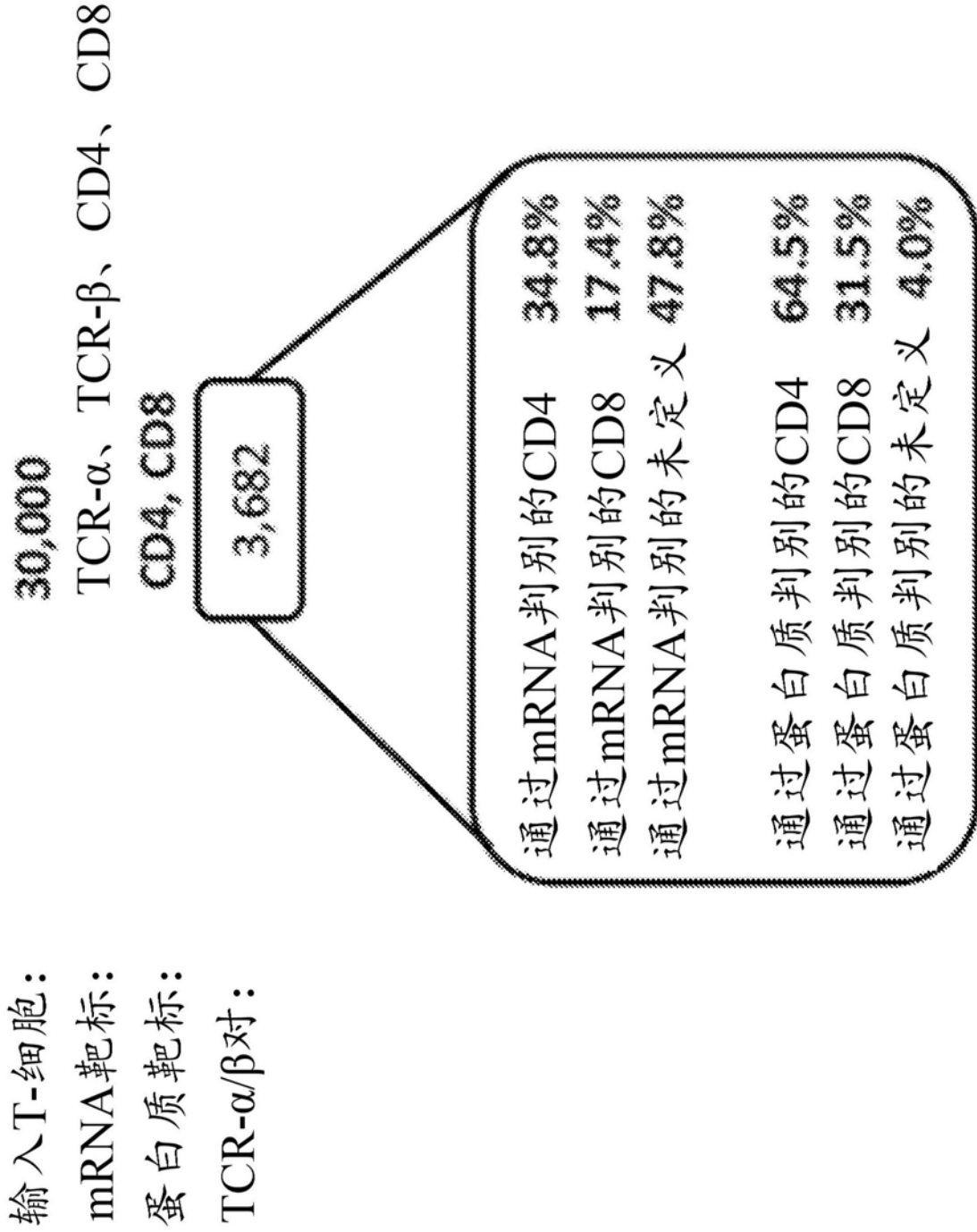


图8B

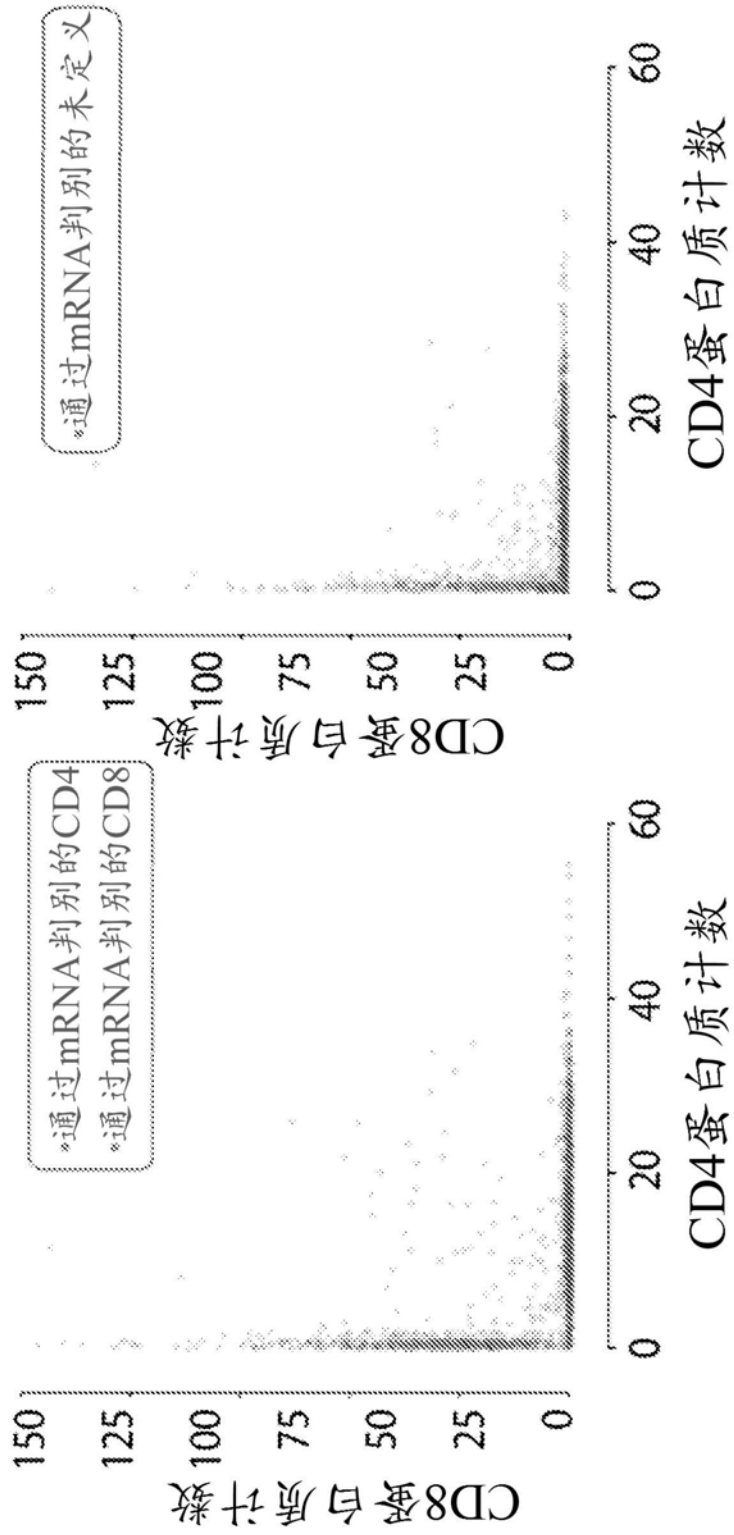


图8C

| | | 蛋白质 | | | |
|---|-----|-------|-------|------|--------|
| | | CD4 | CD8 | 未定义 | 未定义 |
| 3,682 TCR V _α V _β 小液滴条码 | CD4 | 33.0% | 1.4% | 0.4% | 34.8% |
| | CD8 | 0.7% | 16.6% | 0.1% | 17.4% |
| | 未定义 | 30.7% | 13.5% | 3.6% | 47.8% |
| | 总计 | 64.5% | 31.5% | 4.0% | 100.0% |

mRNA

图8D

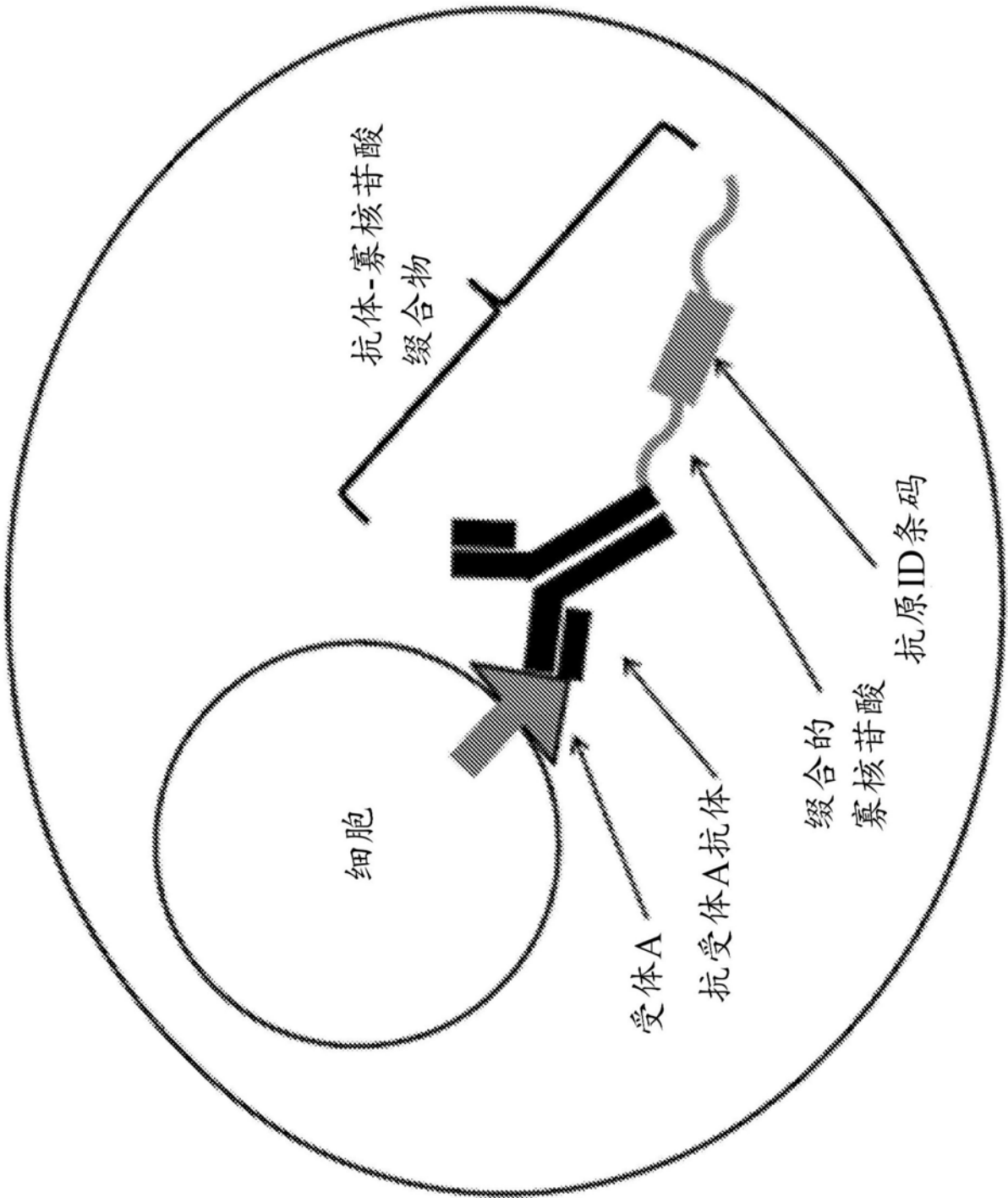


图9A

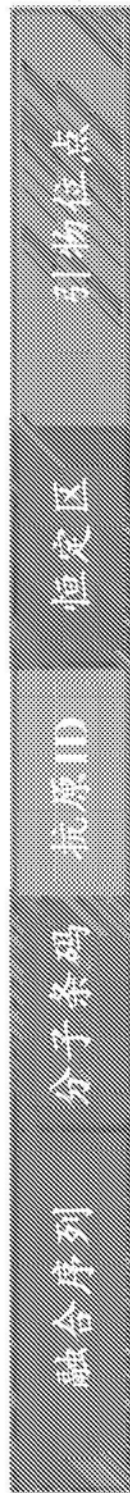


图9B

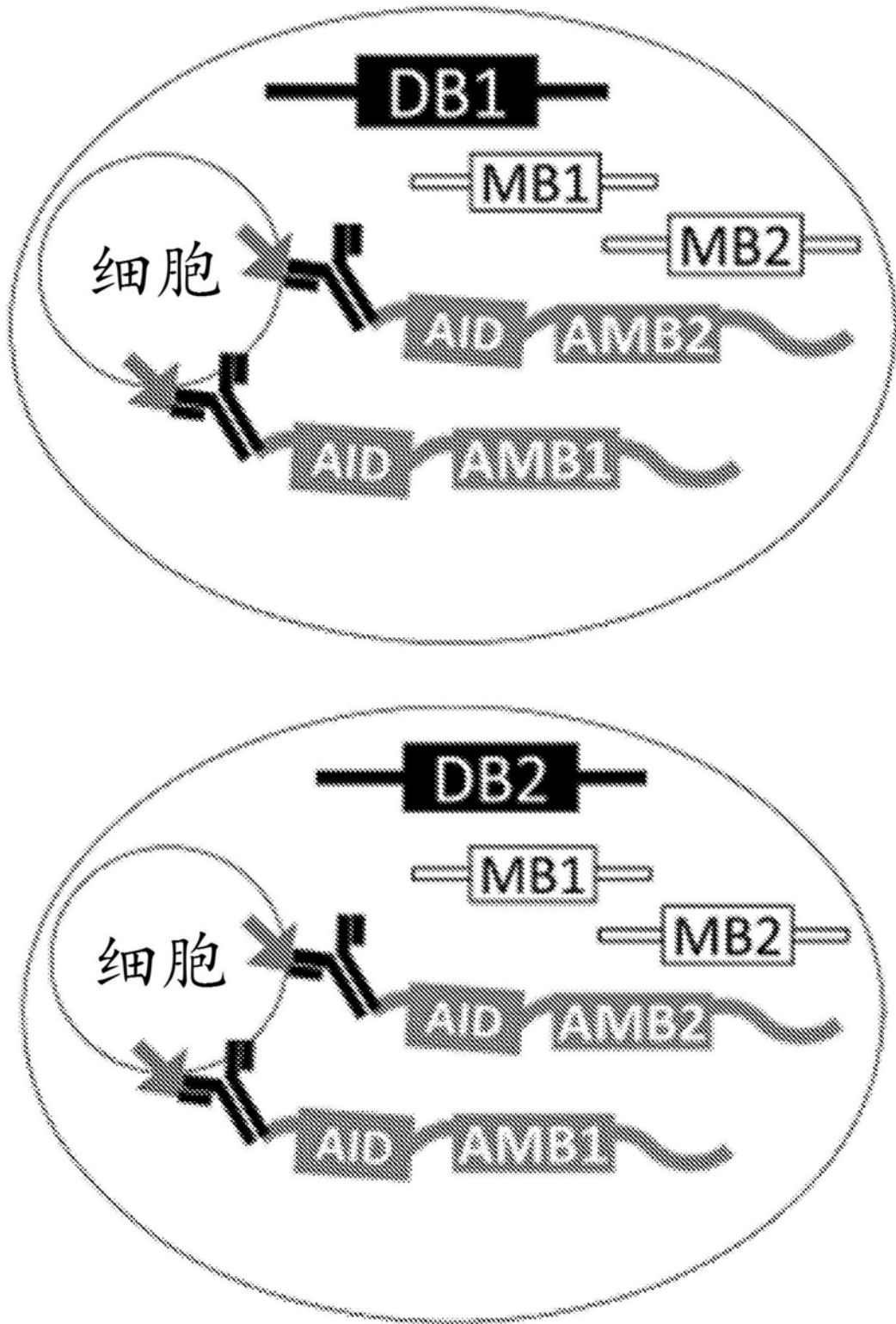


图10A

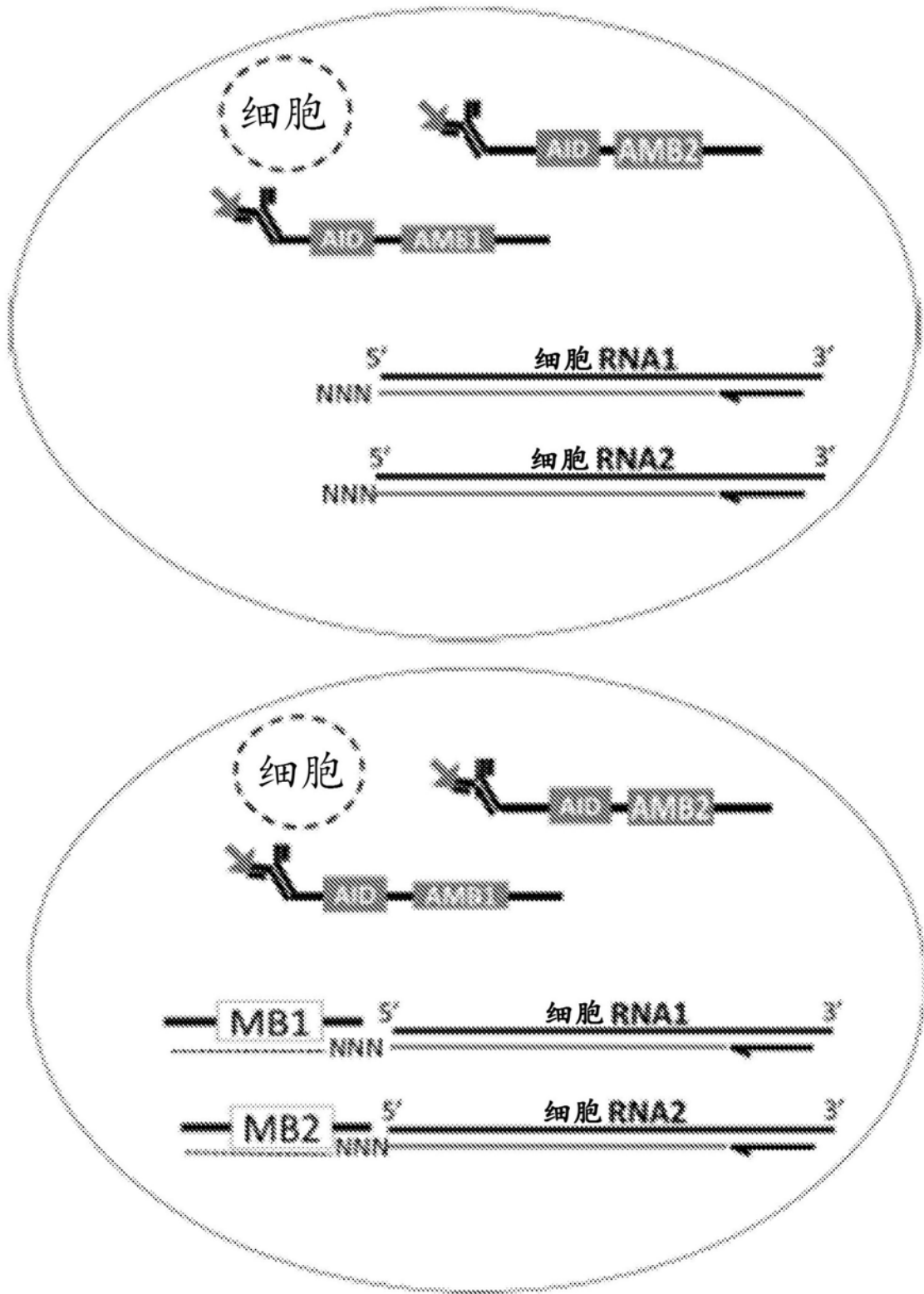


图10B

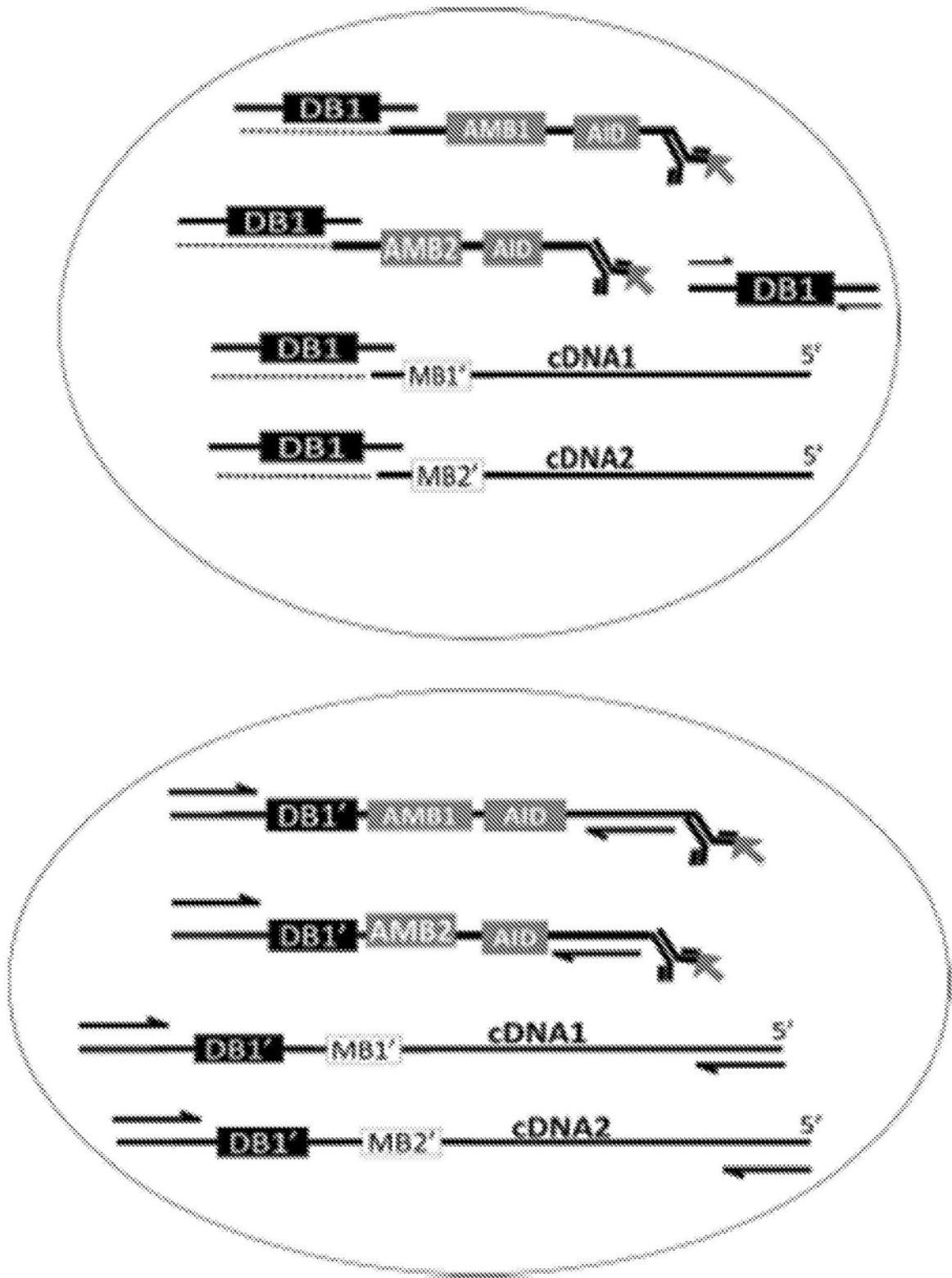


图10C

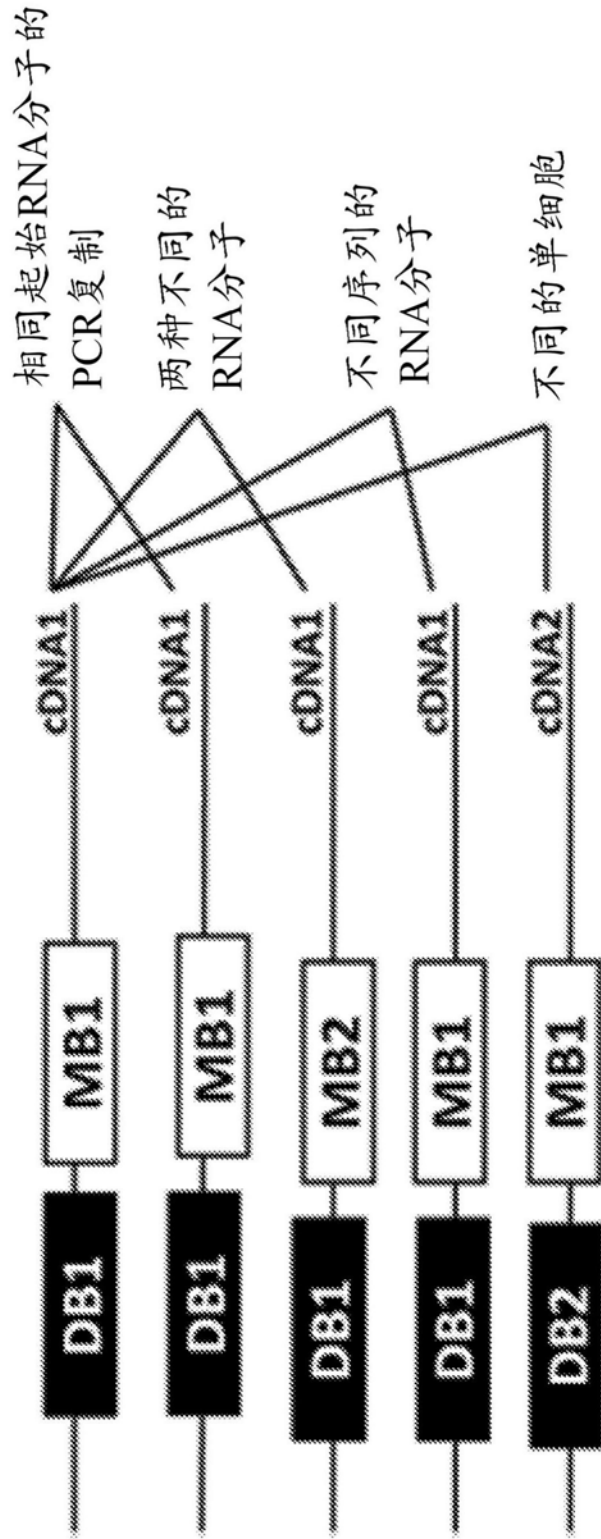


图10D

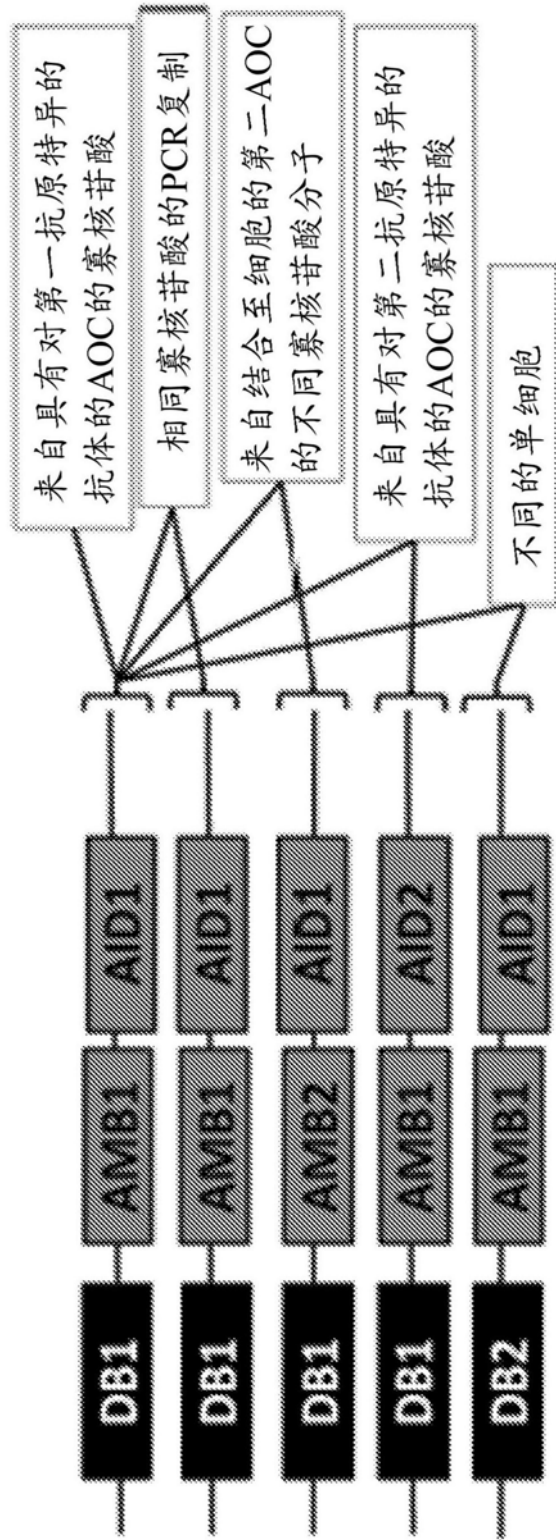


图10E