



(12)发明专利申请

(10)申请公布号 CN 109642230 A

(43)申请公布日 2019.04.16

(21)申请号 201780048963.0

(74)专利代理机构 北京东方亿思知识产权代理有限公司 11258

(22)申请日 2017.08.14

代理人 肖善强

(30)优先权数据

62/375,789 2016.08.16 US

(51)Int.Cl.

C12N 15/10(2006.01)

(85)PCT国际申请进入国家阶段日

C12Q 1/6816(2018.01)

2019.02.02

C12N 9/22(2006.01)

(86)PCT国际申请的申请数据

PCT/US2017/046821 2017.08.14

(87)PCT国际申请的公布数据

W02018/035062 EN 2018.02.22

(71)申请人 加利福尼亚大学董事会

地址 美国加利福尼亚州

(72)发明人 艾米丽·D·克劳福德

埃里克·D·卓乌

约瑟夫·L·德里西

权利要求书2页 说明书20页

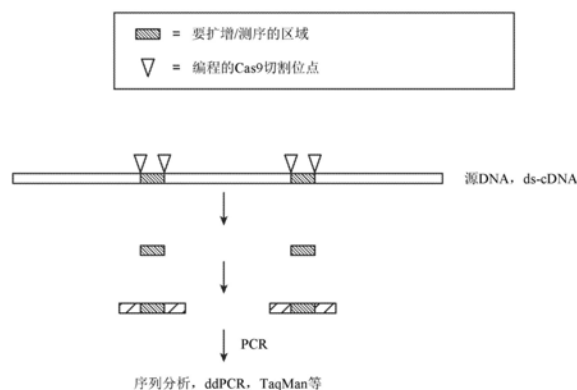
序列列表14页 附图4页

(54)发明名称

通过杂交找到低丰度序列(FLASH)的方法

(57)摘要

提供了一种样品分析方法。在一些实施例中,所述方法包括:(a)使用靶向目标序列的多种重编程的核酸指导的内切核酸酶消化混合核酸样品以产生消化样品,其中所述消化样品中至少一些片段包含:(i)目标序列和(ii)至少一个通过内切核酸酶切割产生的可连接末端;(b)富集含有所述目标序列的片段;以及(c)分析所述富集的片段。还提供了用于执行所述方法的试剂盒。



1. 一种样品分析方法,其包括:

(a) 使用靶向目标序列的多种重编程的核酸指导的内切核酸酶消化末端封端的混合核酸样品以产生消化样品,其中所述消化样品中至少一些片段包含:(i) 目标序列和(ii) 至少一个通过内切核酸酶切割产生的可连接末端;

(b) 富集含有所述目标序列的片段;以及

(c) 分析所述富集的片段。

2. 根据权利要求1所述的方法,其中所述方法包括将衔接子连接至步骤(a)中所述通过内切核酸酶切割产生的可连接末端。

3. 根据权利要求2所述的方法,其中所述衔接子包含捕获部分,并且所述富集是通过将所述捕获部分与载体结合并洗去未结合的核酸来进行。

4. 根据权利要求中任一项所述的方法,其中(a)中所述内切核酸酶消化产生在限定大小范围内的片段。

5. 根据权利要求4所述的方法,其中所述富集是通过选择所述片段的大小来进行。

6. 根据权利要求中任一项所述的方法,其中(a)中所述内切核酸酶消化产生两端均可连接的片段。

7. 根据权利要求中任一项所述的方法,其中(a)中消化产生在两端具有可连接的内切核酸酶切割位点的片段,并且所述方法包括将衔接子连接到所述片段的两端。

8. 根据权利要求7所述的方法,其中所述富集是通过使用与所述衔接子或其补体杂交的引物扩增所连接的片段来进行。

9. 根据权利要求8所述的方法,其中所述衔接子含有分子索引器。

10. 根据权利要求中任一项所述的方法,其中所述方法包括将编索引的衔接子连接到所述片段的两端,其中所述索引是一系列随机核苷酸,其长度足以使每个片段有较高概率接受不同索引;对所连接的片段测序以产生序列读段;然后对与所述序列读段中的目标序列相联的分子索引器序列的数量进行计数,从而提供所述核酸样品中所述目标序列的拷贝数的估计值。

11. 根据权利要求中任一项所述的方法,其中所述方法包括:

(a) 单独地:

(I) 使用靶向目标序列的多种重编程的核酸指导的内切核酸酶消化所述混合核酸样品的第一部分,以产生第一消化样品,其中所述消化样品中的至少一些片段包含:(i) 第一目标序列和(ii) 至少一个通过内切核酸酶切割产生的可连接末端;

(II) 使用靶向目标序列的多种重编程的核酸指导的内切核酸酶消化所述混合核酸样品的第二部分,以产生第二消化样品,其中所述消化样品中的至少一些片段包含:(i) 第二目标序列和(ii) 至少一个通过内切核酸酶切割产生的可连接末端;

其中所述第一消化样品中的至少一些片段与所述第二消化样品中的至少一些片段重叠;

(b) 富集含有所述目标序列的片段;

(c) 对所富集的序列进行测序以产生多个序列读段;以及

(d) 组装重叠的序列读段,从而获得第一目标序列和第二目标序列的重叠群。

12. 根据权利要求中任一项所述的方法,其中所述通过内切核酸酶处理产生的可

连接末端是平末端。

13. 根据以上权利要求中任一项所述的方法,其中所述混合核酸样品包含来自至少两种生物体的DNA。

14. 根据权利要求13所述的方法,其中所述至少两种生物体包含哺乳动物和病原体。

15. 根据权利要求14所述的方法,其中所述病原体是病毒、细菌或真菌。

16. 根据权利要求15所述的方法,其中所述至少两种生物体包含哺乳动物和微生物菌群。

17. 根据以上权利要求中任一项所述的方法,其中所述混合样品包含野生型和突变体DNA。

18. 根据以上权利要求中任一项所述的方法,其中所述混合样品是环境样品、来自犯罪现场的样品或考古样品。

19. 根据以上权利要求中任一项所述的方法,其中所述混合样品是由临床样品构成。

20. 根据权利要求19所述的方法,其中所述临床样品是体液或排泄物。

21. 根据权利要求20所述的方法,其中所述临床样品是血液、痰或粪便。

22. 根据权利要求20所述的方法,其中所述临床样品是肿瘤活检样品。

23. 根据以上权利要求中任一项所述的方法,其中(c)中的所述分析是通过测序进行。

24. 根据以上权利要求中任一项所述的方法,其中所述内切核酸酶是cas9或Argonaught、其直向同源物或其变体。

25. 根据以上权利要求中任一项所述的方法,其中所述样品通过至少2种重编程的核酸指导的内切核酸酶消化。

26. 根据以上权利要求中任一项所述的方法,其中在步骤(a)之前,所述方法包括用磷酸酶处理所述混合核酸样品,从而使所述样品中核酸的5'端不可连接。

27. 一种试剂盒,其包含

多种重编程的核酸指导的内切核酸酶;和
磷酸酶或其它末端封端试剂。

28. 根据权利要求27所述的试剂盒,还包含衔接子。

29. 根据权利要求27所述的试剂盒,还包含PCR引物,其中所述PCR引物的3'端具有与所述衔接子中的序列相同或互补的序列。

通过杂交找到低丰度序列 (FLASH) 的方法

[0001] 交叉引用

[0002] 本申请主张2016年8月16日提交的美国临时申请序列号62/375,789的权益,该申请通过引用并入文中。

背景技术

[0003] 目前用于富集复杂核酸文库中的低丰度序列的方法通常涉及多重PCR或与标记的寡核苷酸的杂交。这两种方法效率低,难以实施,优化成本高,并且在给定样品中可以富集的序列数量有限。

[0004] 一直需要用于富集核酸样品中低丰度序列的新颖方法。

发明内容

[0005] 本文描述的是称为通过杂交找到低丰度序列或“FLASH”的方法,一种在测序或其它分子计数应用前使用序列特异性核酸酶(例如CRISPR/Cas9)来切割DNA文库或其它样品中的特定目标位点的技术。在一些实施中,然后将新暴露的DNA末端自由连接到允许他们被扩增的特定衔接序列。在这些实施例中,仅使用对衔接子特异的一对引物的单个PCR步骤因此可以以完全可编程的方式扩增数百、数千或可能数百万种不同的序列。在一些情况下,为了减少非靶向分子的测序,可以封端DNA来源中DNA分子的末端,例如用磷酸酶处理或在核酸酶消化之前使用另一种方法来封端任何已经可及的DNA末端。

[0006] 在一些实施例中,所述方法可以包括:(a)用多种重编程的核酸指导的内切核酸酶消化末端封端的(例如,磷酸酶处理的)混合核酸样品,所述内切核酸酶靶向目标序列以产生消化的样品,其中消化的样品中的至少一些片段包含:(i)目标序列和(ii)至少一个通过内切核酸酶切割产生的可连接末端;(b)富集含有目标序列的片段;以及(c)分析富集的片段。

[0007] 还提供了用于执行所述方法的试剂盒。

附图说明

[0008] 当结合附图阅读时,将最好地理解以下详细描述的一些方面。需要强调的是,根据惯例,附图的各种特征未按比例绘制。相反,为了清楚起见,各种特征的尺寸被任意扩大或缩小。附图中包括以下图:

[0009] 图1显示了FLASH方法的一些原理。

[0010] 图2显示了不同的FLASH池(由空心 and 实心三角形表示)如何用于将具有重叠片段的所需DNA片段(实心条)片段化。可以使用不同的FLASH池(由空心 and 实心三角形表示)来将具有重叠片段(散列条)的所需DNA片段(实心条)片段化。重叠区域可以靶向SNP、易位或侧翼重复序列的区域。这些重叠片段可以在文库制备之前组合,然后用于组装测序读数。

[0011] 图3图A显示来自i)甲氧西林敏感性金黄色葡萄球菌(MSSA)的培养分离物,ii)耐甲氧西林金黄色葡萄球菌(MRSA)和iii)直接来自MRSA肺炎患者的BAL液的美cA丰度。来自

MRSA肺炎患者的BAL液中可检测的mecA随着FLASH而增加。面板B显示FLASH允许检测来自MRSA和铜绿假单胞菌肺炎患者的BAL液中的氟喹诺酮抗性突变GyrA S84L。

[0012] 图4描述了该方法的另一种形式,其中封端的衔接子与片段化DNA连接,防止聚合酶扩增,除非靶向切口酶(例如具有单个活性位点突变D10A的Cas9)产生允许置换衔接子的封端链的单链切割。结果是仅扩增含有两个切口酶位点的插入物。

[0013] 定义

[0014] 在更详细地描述示例性实施例之前,阐述以下定义以说明和定义说明书中使用的术语的含义和范围。

[0015] 数字范围包括定义范围的数字。除非另有说明,否则核酸以5'至3'方向从左向右书写;氨基酸序列分别以氨基至羧基方向从左向右书写。

[0016] 除非另外定义,否则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。Singleton等人,DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY,第2版,John Wiley和Sons,New York(1994),以及Hale&Markham,THE HARPER COLLINS DICTIONARY OF BIOLOGY,Harper Perennial,N.Y.(1991)向技术人员提供本文使用的许多术语的一般含义。尽管如此,为了清楚和便于参考,下面定义了某些术语。

[0017] 必须注意的是,如本文和所附权利要求中所使用的,单数形式“一(a)”,“一(an)”和“所述”包括复数指示物,除非上下文另有明确说明。例如,术语“引物”是指一种或多种引物,即单一引物和多种引物。还应注意,权利要求可撰写为排除任何任选元素。因此,此陈述旨在作为使用与权利要求元素的叙述相关的“单独”,“仅”等排他性术语或使用“否定”限制的先行基础。

[0018] 如本文所用的术语“样品”涉及材料或材料混合物,通常但不一定是液体形式,其含有一种或多种目标分析物。本文使用的核酸样品可能是复合的,因为他们含有多种含有序列的不同分子。由来自哺乳动物(例如小鼠或人)的mRNA制备的基因组DNA和cDNA是复合样品的类型。复合样品可具有超过 10^4 、 10^5 、 10^6 或超过 10^7 种不同的核酸分子。DNA靶标可能来自任何来源,如基因组DNA、cDNA(来自RNA)或人工DNA构建体。文中可以使用任何含有核酸的样品,例如由组织培养细胞制成的基因组DNA、组织样品、FFPE样品、临床、环境或其它类型的样品。

[0019] 如本文所用的术语“核酸样品”表示含有核酸的样品。本文使用的核酸样品可以是复合的,因为他们含有多种含有序列的不同分子。来自哺乳动物(例如小鼠或人)的基因组DNA、RNA(和由其制备的cDNA)是复合样品的类型。复合样品可具有超过 10^4 、 10^5 、 10^6 或超过 10^7 种不同的核酸分子。靶分子可能来自任何来源,例如基因组DNA,或人工DNA构建体。文中可以使用任何含有核酸的样品,例如由组织培养细胞制成的基因组DNA或组织样品。

[0020] 如本文所用,术语“混合物”是指散布而不以任何特定顺序的元素的组合。混合物是不均匀的,并且在空间上不能分离成其不同的组分。元素混合物的实例包括溶解在相同水溶液中的许多不同元素和在随机位置(即,没有特定顺序)附着于固体载体的许多不同元素。混合物不可寻址。为了举例说明,如本领域公知的,空间上分离的表面结合的多核苷酸阵列不是表面结合的多核苷酸的混合物,因为表面结合的多核苷酸的种类在空间上是不同的并且阵列是可寻址的。

[0021] 术语“核苷酸”旨在包括那些不仅含有已知嘌呤和嘧啶碱基,还含有已经被修饰的其它杂环碱基的部分。这些修饰包括甲基化嘌呤或嘧啶、酰化嘌呤或嘧啶、烷基化核糖或其它杂环。此外,术语“核苷酸”包括那些含有半抗原或荧光标记的部分,并且不仅可以含有常规的核糖和脱氧核糖,还可以含有其它糖。修饰的核苷或核苷酸还包括对糖部分的修饰,例如,其中一个或多个羟基被卤素原子或脂族基团替换,被官能化为醚、胺等。

[0022] 术语“核酸”和“多核苷酸”在本文中可互换使用以描述任何长度的聚合物,例如,大于约2个碱基、大于约10个碱基、大于约100个碱基、大于约500个碱基、大于约1000个碱基、至多约10,000个或更多个由核苷酸(例如脱氧核糖核苷酸或核糖核苷酸)组成的碱基,并且可以通过酶促或合成产生(例如,如美国专利号5,948,902和其中引用的参考文献中所述的肽核酸或PNA),其可以以与两种天然存在的核酸类似的序列特异性方式与天然存在的核酸杂交,例如,可以参与Watson-Crick碱基配对相互作用。天然存在的核苷酸包括鸟嘌呤、胞嘧啶、腺嘌呤、胸腺嘧啶、尿嘧啶(分别为G、C、A、T和U)。DNA和RNA分别具有脱氧核糖和核糖糖骨架,而PNA的骨架由重复的通过肽键连接的N-(2-氨基乙基)-甘氨酸单元组成。在PNA中,各种嘌呤和嘧啶碱基通过亚甲基羰基键与骨架连接。通常称为不可及RNA的锁定核酸(LNA)是修饰的RNA核苷酸。LNA核苷酸的核糖部分用连接2'氧和4'碳的额外桥修饰。该桥将核糖“锁定”在3'-内(北)构象中,这通常在A型双链体中发现。只要需要,LNA核苷酸可以与寡核苷酸中的DNA或RNA残基混合。术语“非结构化核酸”或“UNA”是含有彼此结合从而降低稳定性的非天然核苷酸的核酸。例如,非结构化核酸可含有G'残基和C'残基,其中这些残基对应于G和C的非天然存在形式,即类似物,它们彼此碱基配对,稳定性降低,但保留分别与天然存在的C和G残基碱基配对的能力。非结构化核酸描述于US20050233340中,其通过引用并入本文以用于UNA的公开。

[0023] 如本文所用的术语“寡核苷酸”表示核苷酸的单链多聚体,其长度为约2至200个核苷酸,最多500个核苷酸。寡核苷酸可以是合成的或可以通过酶促制备,并且在一些实施例中,长度为30至150个核苷酸。寡核苷酸可含有核糖核苷酸单体(即,可以是寡核糖核苷酸)和/或脱氧核糖核苷酸单体。例如,寡核苷酸的长度可以是10至20、21至30、31至40、41至50、51至60、61至70、71至80、80至100、100至150或150至200个核苷酸。

[0024] “引物”是指天然或合成的寡核苷酸,其在与多核苷酸模板形成双链体时能够用作核酸合成的起始点并且从其3'端沿着模板延伸,使得形成了扩展双链体。在延伸过程中添加的核苷酸序列由模板多核苷酸的序列决定。通常引物通过DNA聚合酶延伸。引物通常具有与其在引物延伸产物的合成中的用途相容的长度,并且通常长度为8至100个核苷酸,例如10至75、15至60、15至40、18至30、20至40、21至50、22至45、25至40等,更典型地长度为18至40、20至35、21至30个核苷酸,以及所述范围之间的任何长度。典型的引物的长度可以为10至50个核苷酸,例如15至45、18至40、20至30、21至25等,以及所述范围之间的任何长度。在一些实施例中,引物的长度通常不超过约10、12、15、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、55、60、65或70个核苷酸。因此,“引物”与模板互补,并通过氢键结合或与模板杂交形成复合物,以产生用于通过聚合酶引发合成的引物/模板复合物,其通过在DNA合成过程中添加在其3'端连接的与模板互补的共价键合的碱基而延伸。

[0025] 术语“杂交(hybridization)”或“杂交(hybridizes)”是指核酸链在正常杂交条件下与第二互补核酸链退火并形成稳定双链体(同源双链体或异源双链体),并且在相同的正

常杂交条件下,不与不相关的核酸分子形成稳定双链体。通过在杂交反应中使两条互补核酸链退火来完成双链体的形成。通过调节发生杂交反应的杂交条件(通常称为杂交严格性),可以使杂交反应具有高度特异性,使得两条核酸链之间的杂交不会形成稳定的双链体,例如,在正常严格条件下保留双链区域的双链体,除非两条核酸链在特定序列中含有基本上或完全互补的一定数量的核苷酸。对于任何给定的杂交反应,容易确定“正常杂交或正常严格条件”。参见例如,Ausubel等人,《现代分子生物学实验指南(Current Protocols in Molecular Biology)》,John Wiley&Sons,Inc.,New York,或Sambrook等人,《分子克隆:实验室手册》,冷泉港实验室出版社(Molecular Cloning:A Laboratory Manual,Cold Spring Harbor Laboratory Press)。如本文所用,术语“杂交(hybridizing)”或“杂交(hybridization)”是指核酸链通过碱基配对与互补链结合的任何过程。

[0026] 如果两个序列在中度至高度严格杂交和洗涤条件下彼此特异性杂交,则认为核酸与参考核酸序列“可选择性杂交”。中度和高度严格杂交条件是已知的(参见例如,Ausubel等人,《精编分子生物学实验指南(Short Protocols in Molecular Biology)》,第3版,Wiley&Sons 1995和Sambrook等人,《分子克隆:实验室手册》(Molecular Cloning:A Laboratory Manual),第3版,2001纽约冷泉港实验室出版社)。高度严格条件的一个实例包括在约42°C下在50%甲酰胺、5X SSC、5X Denhardt溶液、0.5%SDS和100ug/ml变性载体DNA中杂交,然后在室温下在2X SSC和0.5%SDS中洗涤两次,并且在0.1×SSC和0.5%SDS中在42°C下再洗涤两次。

[0027] 如本文所用,术语“双链体”或“双链体化”描述了两个碱基配对的互补多核苷酸,即杂交在一起。

[0028] 如本文所用的术语“扩增”是指合成与模板核酸的一条链或两条链互补的核酸分子的过程。扩增核酸分子可包括使模板核酸变性,在低于引物的解链温度的温度下使引物与模板核酸退火,并通过酶促从引物延长以产生扩增产物。变性,退火和延长步骤各自可进行一次或多次。在某些情况下,变性,退火和延长步骤进行多次,使得扩增产物的量通常以指数方式增加,尽管本方法不需要指数扩增。扩增通常需要存在脱氧核糖核苷三磷酸,DNA聚合酶和适当的缓冲液和/或辅因子以获得聚合酶的最佳活性。术语“扩增产物”是指核酸序列,其由本文定义的扩增方法产生。

[0029] 术语“确定”、“测量”、“评价”、“评估”、“测定”和“分析”在本文中可互换使用,以指代任何形式的测量,并且包括确定元素是否存在。这些术语包括定量和/或定性测定。评估可以是相对的或绝对的。“评估存在”包括确定存在物的量,以及确定其是否存在。

[0030] 术语“使用”具有其常规含义,并且因此是指采用例如投入使用的方法或组合物来达到目的。例如,如果使用程序来创建文件,则执行程序以生成文件,该文件通常是程序的输出。在另一个实例中,如果使用计算机文件,则通常访问,读取,并且使用存储在文件中的信息来达到目的。类似地,如果使用唯一标识符,例如条形码,则通常读取唯一标识符以识别例如与唯一标识符相关联的对象或文件。

[0031] 如本文所用,术语“基因组区域”是指基因组的区域,例如动物或植物基因组,例如人、猴、大鼠、鱼或昆虫或植物的基因组。

[0032] 如本文所用,术语“基因组序列”是指在基因组中发生的序列。

[0033] 如本文所用,术语“基因组片段”是指基因组的区域,例如动物或植物基因组,例如

人、猴、大鼠、鱼或昆虫或植物的基因组。基因组片段可以是整个染色体,或染色体的片段。基因组片段可以是衔接子连接的(在这种情况下,其具有与片段的一端或两端,或至少与分子的5'端连接的衔接子),或者可以不是衔接子连接的。

[0034] 如本文所用,术语“连接”是指第一DNA分子的5'端的末端核苷酸与第二DNA分子的3'端的末端核苷酸的酶催化接合。

[0035] “多个”包含至少2个成员。在某些情况下,多个可具有至少2个、至少5个、至少10个、至少100个、至少1000个、至少10,000个、至少100,000个、至少 10^6 个、至少 10^7 个、至少 10^8 个或至少 10^9 个或更多个成员。

[0036] 如果两个核酸是“互补的”,则他们在高严格条件下相互杂交。术语“完全互补”用于描述双链体,其中一个核酸的每个碱基与另一个核酸中的互补核苷酸碱基配对。在许多情况下,互补的两个序列具有至少10个,例如至少12或15个具有互补性的核苷酸。

[0037] 如本文所用的术语“链”是指由通过共价键共价连接在一起的核苷酸构成的核酸,例如磷酸二酯键。在细胞中,DNA通常以双链形式存在,因此,具有两条互补的核酸链,在本文中称为“顶部”和“底部”链。在某些情况下,染色体区域的互补链可以称为“正”和“负”链、“第一”和“第二”链、“编码”和“非编码”链、“Watson”和“Crick”链或“有义”和“反义”链。将链分配为顶部或底部链是任意的,并不暗示任何特定的取向、功能或结构。几个示例性哺乳动物染色体区域(例如,BAC、组体、染色体等)的第一链的核苷酸序列是已知的,并且可以在例如NCBI的Genbank数据库中找到。

[0038] 如本文所用,术语“测序”是指通过其获得多核苷酸的至少10个连续核苷酸的身份(例如,至少20个、至少50个、至少100个或至少200个或更多个连续核苷酸的身份)的方法。

[0039] 术语“新一代测序”是指依诺米那(Illumina)、生命科技(Life Technologies)、太平洋生物科学(Pacific Biosciences)和罗氏(Roche)等目前采用的所谓的平行化合成测序或连接测序平台。下一代测序方法也可包括纳米孔测序方法或基于电子检测的方法,例如由生命科技商业化的离子激流(Ion Torrent)技术。

[0040] 如本文所用,术语“延伸”是指通过添加核苷酸使用聚合酶来延伸引物。如果延伸与核酸退火的引物,则核酸作为延伸反应的模板。

[0041] 如本文所用,术语“条形码序列”、“分子条形码”或“索引”是指用于(a)鉴定和/或追踪反应中多核苷酸来源和/或(b)计数初始分子测序的次数(例如,在样品中基本上每个分子用不同序列标记,然后扩增样品的情况下)的独特核苷酸序列。条形码序列可以位于寡核苷酸的5'端,3'端或中间,或5'端和3'端。条形码序列的大小和组成可能有很大变化;以下参考文献提供了用于选择适合于特定实施例的条形码序列组的指导:Brenner,第5,635,400号美国专利;Brenner等人,Proc.Natl.Acad.Sci.,97:1665-1670(2000);Shoemaker等人,Nature Genetics,14:450-456(1996);Morris等人,欧洲专利出版物0799897A1;Wallace,第5,981,179号美国专利等。在特定实施例中,条形码序列的长度可以为4至36个核苷酸,或6至30个核苷酸,或8至20个核苷酸。

[0042] 如本文所用,术语“PCR试剂”是指在模板上进行聚合酶链式反应(PCR)所需的所有试剂。如本领域所知,PCR试剂基本上包括第一引物、第二引物、热稳定聚合酶和核苷酸。取决于所用的聚合酶,也可以存在离子(例如 Mg^{2+})。PCR试剂可任选地含有可从其扩增靶序列的模板。

[0043] 术语“相邻”是指小于核苷酸的最长维度的距离。术语“相邻连接”是指两个核苷酸在链上彼此紧邻,并没有插入核苷酸。

[0044] 在有尾引物或具有5'尾的引物的上下文中,术语“有尾”是指在其5'端具有区域(例如,至少12-50个核苷酸的区域)的引物,该5'端不与引物的3'端杂交到相同的靶标。

[0045] 术语“可区分序列”是指彼此不同的序列。

[0046] 如本文所用的术语“靶核酸”是指研究中的目标多核苷酸。

[0047] 术语“靶核酸分子”是指与其它靶核酸分子一起或不一起存在于组合物中的单个分子。分离的靶核酸分子是指存在于不含其它靶核酸分子的组合物中的单个分子。

[0048] 在两个或更多个可变核酸序列的上下文中,术语“可变”是指两个或更多个相对于彼此具有不同核苷酸序列的核酸。换句话说,如果群体的多核苷酸具有可变序列,则群体的多核苷酸分子的核苷酸序列在不同分子之间变化。术语“可变”不应理解为要求群体中的每个分子具有与群体中的其它分子不同的序列。

[0049] 术语“衔接子”是指可以通过连接酶介导的反应与双链DNA分子的至少一条链接合的核酸。在一个实施例中,衔接子可以是Y衔接子。显而易见的是,衔接子的一端可以被设计成与通过内切核酸酶切割产生的突出端相容,例如,其可以具有平末端或5'T突出端。术语“衔接子”是指至少部分双链的分子。衔接子的长度可以是40至150个碱基,例如50至120个碱基,尽管可以设想在该范围之外的衔接子。

[0050] 术语“Y-衔接子”是指包含以下的衔接子:双链区和单链区,其中相对序列是不互补的。双链区的末端可以与靶分子例如基因组DNA的双链片段接合,例如通过连接。已经连接到Y衔接子的衔接子标记的双链DNA的每条链被不对称标记,因为其在一端具有Y-衔接子的一条链的序列而在另一端具有Y-衔接子的另一条链的序列。在两端已经与Y-衔接子接合的核酸分子的扩增产生不对称标记的核酸,即具有含有一个标签序列的5'端和具有另一个标签序列的3'端的核酸。

[0051] 术语“通用衔接子”是指与研究中的核酸分子的两端连接的衔接子。在某些实施例中,通用衔接子可以是Y-衔接子。在两端已经连接到Y-衔接子的核酸分子的扩增产生不对称标记的核酸,即具有含有一个标签序列的5'端和具有另一个标签序列的3'端的核酸。

[0052] 如本文所用,术语“衔接子标记的”是指已被衔接子标记的核酸。衔接子可以接合到核酸分子的5'端和/或3'端。

[0053] 如本文所用的术语“标记的DNA”是指具有添加的衔接子序列的DNA分子,即合成来源的“标签”。可以通过连接添加(即“附加”)衔接子序列。

[0054] 如本文所用,术语“单独消化”是指两种或更多种不同的切割反应。

[0055] 如本文所用,术语“核酸指导的内切核酸酶”是指DNA和RNA指导的内切核酸酶,其包括Argonaut和II型CRISPR/基于Cas的系统,该系统由两种组分组成:切割靶DNA的核酸酶(例如,Cas9内切核酸酶或其变体或直向同源物)和将核酸酶靶向靶DNA中的特定位点的指导RNA(gRNA)。参见,例如,Hsu等人(Nature Biotechnology 2013 31:827-832)。

[0056] 如本文所用,术语“部分”是指样品的一部分(例如,等分试样)。

[0057] 如本文所用,术语“确定的位点”是指选定的序列。

[0058] 如本文所用,术语“选择性扩增”是指仅选择的序列被扩增的扩增反应(例如,PCR反应)。

[0059] 在某些情况下,本文所述方法中使用的寡核苷酸可以使用参考基因组区域设计,即具有已知核苷酸序列的基因组区域,例如其序列例如保存在NCBI的Genbank数据库或其它数据库中的染色体区域。这种寡核苷酸可用于使用含有测试基因组的样品的测定中,其中测试基因组含有寡核苷酸的结合位点。

[0060] 如本文所用,术语“生物素部分”是指包括包括生物素或生物素类似物的亲和剂,例如脱硫生物素、氧代生物素、2-亚氨基生物素、二氨基生物素、生物素亚砷、生物胞素等。生物素部分与链霉抗生物素蛋白结合,亲和力为至少10⁻⁸M。生物素亲和剂还可以包括接头,例如-LC-生物素、-LC-LC-生物素、-SLC-生物素或-PEG_n-生物素,其中n是3-12。

[0061] 术语“亲和标签”和“捕获部分”是指能够:a)非共价特异性地彼此结合或b)选择性地彼此反应以形成共价键的部分。非共价特异性地彼此结合的合适亲和标签和捕获剂对的实例有很多,包括但不限于:链霉抗生物素蛋白/抗生物素蛋白、洋地黄毒苷/抗洋地黄毒苷抗体、荧光素/抗荧光素抗体,尽管许多其它是已知的。选择性地彼此反应以形成共价键的化学选择性反应性基团的实例有很多,并包括:胺和活性酯例如NHS酯、硫醇和马来酰亚胺或碘乙酰胺),以及可通过Click化学彼此反应的基团,例如叠氮化物和炔烃基团。含有可用于本文的亲和标签的核糖核苷酸可从许多来源商购获得。

[0062] 术语“末端封端”是指已被修饰以使其不能用于连接的末端。尽管可以使用其它方法,但可以通过用磷酸酶处理来末端封端核酸。

具体实施方式

[0063] 在一些实施例中,该方法可以包括使用多种重编程的核酸指导的内切核酸酶消化混合核酸样品(即,含有来自至少两个来源的DNA的核酸样品,其中来自一个来源的DNA可以占样品中总DNA的小于10%、小于5%、小于2%或小于1%),所述内切核酸酶靶向目标序列(即一组Cas9内切核酸酶,Argonaut,其直向同源物或变体,其具有将内切核酸酶靶向靶核酸(例如靶基因组)中的预定靶位点的合成指导RNA或DNA)。该步骤产生在样品中包含核酸片段的消化样品,其中消化样品中的至少一些片段包含:(i)目标序列和(ii)至少一个通过内切核酸酶切割产生的可连接末端(即,含有5'磷酸和3'羟基的末端,其末端是平的或具有有限的突出端)。一些核酸指导的内切核酸酶(包括Cas9)产生平末端,而其它可产生限定的突出端。消化后,可以富集含有目的序列的片段,例如,使用大小选择,通过将包含捕获部分(例如,生物素部分)的衔接子连接到由内切核酸酶产生的可连接末端,将捕获部分结合到载体(例如,链霉抗生物素蛋白载体),洗去未结合的核酸,或通过衔接子(例如“Y-衔接子”)连接到片段的末端,并使用与Y衔接子中的序列结合或互补的引物通过PCR扩增连接的片段。可以通过任何合适的方法分析富集的片段,例如测序。

[0064] 对于Cas9,指导RNA可以由两个分子组成,即一个与靶杂交并提供序列特异性的RNA(“crRNA”)和一个能够与crRNA杂交的RNA(“tracrRNA”)。或者,指导RNA可以是含有crRNA和tracrRNA序列的单个分子(即sgRNA)。Cas9蛋白与野生型Cas9蛋白(例如,化脓性链球菌(*Streptococcus pyogenes*)Cas9蛋白)具有至少60%同一性(例如,至少70%、至少80%、或90%同一性、至少95%同一性或至少98%同一性或至少99%同一性)。Cas9蛋白可具有野生型Cas9蛋白的所有功能,或仅具有一种或一些功能,包括结合活性和核酸酶活性。

[0065] 为了使Cas9成功结合DNA,基因组DNA中的靶序列应该与gRNA序列互补,并且必须

紧接着是正确的原型间隔区相邻基序或“PAM”序列。PAM序列存在于DNA靶序列中但不存在于gRNA序列中。其后是PAM序列的具有正确靶序列的任何DNA序列将被Cas9结合。PAM序列根据产生Cas9的细菌物种而不同。最广泛使用的II型CRISPR系统源自化脓性链球菌，PAM序列是位于gRNA识别序列的紧邻3'端的NGG。来自示例性细菌物种的II型CRISPR系统的PAM序列包括：化脓性链球菌(NGG)、脑膜炎奈瑟氏球菌(*Neisseria meningitidis*) (NNNGATT)、嗜热链球菌(*Streptococcus thermophilus*) (NNAGAA)和齿垢密螺旋体(*Treponema denticola*) (NAAAAC)。对于一些其它序列特异性核酸酶，例如Argonats，不需要PAM位点来结合和切割靶DNA。

[0066] 显而易见的是，该反应可以在体外进行，即在无细胞环境中使用分离的核酸(例如分离的DNA)进行。混合样品可以从任何来源收集，包括任何生物、有机材料或含核酸的物质，包括但不限于植物、动物(例如爬行动物、哺乳动物、昆虫、蠕虫、鱼等)、组织样品、细菌、真菌(例如酵母)、噬菌体、病毒、尸体组织、考古/古代样品等。在某些实施例中，该方法中使用的基因组DNA可以源自哺乳动物，其中在某些实施例中，哺乳动物是人类。

[0067] 例如，该方法中使用的指导RNA可以经设计使得他们指导内切核酸酶与靶基因组中的预定切割位点的结合。在某些情况下，可以选择切割位点以释放含有未知序列区域或含有SNP、核苷酸插入、核苷酸缺失、重排等的区域的片段。由于许多生物的基因组分离方法和核苷酸序列(包括许多细菌、真菌、植物和动物，例如哺乳动物，例如人、灵长类动物和啮齿动物，例如小鼠和大鼠)是已知的，因此设计用于本方法的指导RNA应该在本领域技术人员的技能范围内。例如，Cas9-gRNA复合物可以被编程为与任何序列结合，条件是该序列具有PAM基序。理论上，Cas9-gRNA复合物可切割基因组DNA以产生30-50bp的片段。然而，实际上，切割位点之间的最小间隔可以例如在50-900bp的范围内。在一些实施例中，sgRNA或crRNA可以是简并序列以靶向相对保守的区域。

[0068] 该方法可以使用一组至少2种、至少5种、至少10种、至少100种、至少1,000种、至少10,000种、至少50,000种或至少100,000种或更多种不同的指导RNA/DNA，其各自与一个或多个基因组中不同的，预定义的位点互补。相邻位点之间的距离可以根据所需的应用而极大地变化。在一些实施例中，相邻位点之间的距离可以在100bp至200kb的范围内，并且在特定实施例中，可以选择位点以释放在限定的大小范围内的片段，例如100bp至2kb，例如，200bp至1kb，或更大(例如，500bp至20kb)用于纳米孔和/或PacBio测序。在某些情况下，可以选择指导RNA/DNA以释放具有适合于大小选择的大小(例如，小于1kb或小于500bp)的片段。在这些实施例中，片段的长度可小于1kb或长度大于约1kb，并且可通过选择片段的大小来完成富集。

[0069] 在一些实施例中，内切核酸酶的消化可以产生具有两个末端的片段，两个末端都是可连接的。在这些实施例中，该方法包括将衔接子连接到片段的两端，从而允许使用与衔接子或其补体杂交的单对引物通过PCR富集连接的片段。在这些实施例中，片段还可以在它们连接到衔接子之前或之后通过大小富集。显而易见的是，所用的衔接子应与内切核酸酶产生的末端相容。在一些实施例中，与片段连接的衔接子的末端可以是平端的。在其它实施例中，与片段连接的衔接子的末端可以具有与内切核酸酶产生的突出端互补的突出端。在进一步的实施例中，在连接到T尾衔接子之前，平末端的片段可以是A尾(例如，使用Taq聚合酶)。如上所述，在一些实施例中，衔接子可以是Y-衔接子，并且因此，连接产物的每条链可

以被不对称地标记,因为其在一端具有Y-衔接子的一条链的序列,并且在另一端具有Y-衔接子的另一条链的序列,其中添加的序列不相同或互补。在两端已经与Y-衔接子接合的核酸分子的扩增产生不对称标记的核酸,即具有含有一个标签序列的5'端和具有另一个标签序列的3'端的核酸。

[0070] 在一些实施例中,衔接子可含有索引序列,例如,在分子之间变化的序列,例如随机序列,从而允许对分子进行计数。在这些实施例中,该方法可包括将编索引的衔接子连接到片段的两端,扩增连接的片段,对连接的片段测序以产生序列读段,然后对与序列读段中的目标序列相联的分子索引器序列的数量进行计数,从而提供核酸样品中目标序列的拷贝数的估计值。这允许用户评估原始样品中不同序列分子的相对量。

[0071] 该方法可用于获得重叠序列读段,从而允许重叠群的组装。在这些实施例中,该方法可以包括:(a)单独(例如,在单独的容器中):(I)使用靶向目标序列的第一多种重编程的核酸指导的内切核酸酶消化混合核酸样品的第一部分,以产生第一消化样品,其中消化样品中的至少一些片段包含:(i)第一目标序列和(ii)至少一个通过内切核酸酶切割产生的可连接末端;(II)使用靶向目标序列的第二多种重编程的核酸指导的内切核酸酶消化混合核酸样品的第二部分,以产生第二消化样品,其中消化样品中的至少一些片段包含:(i)第二目标序列和(ii)至少一个通过内切核酸酶切割产生的可连接末端。在该方法中,可以设计每次消化中的内切核酸酶,使得第一消化样品中的至少一些片段与第二消化样品中的至少一些片段重叠,例如,重叠至少50、至少100或至少200个碱基)。可以如上所述富集含有目标序列的片段(例如,通过连接衔接子并使用与衔接子中的序列或其补体杂交的引物通过PCR扩增片段),然后测序以产生多个序列读段。在测序之后,该方法可以包括组装任何重叠序列读段,从而获得包含至少部分第一目标序列和至少部分第二目标序列的序列的重叠群。

[0072] 显而易见的是,用于扩增的衔接子和/或引物可以与在其中使用引物延伸的任何下一代测序平台中的使用相容,例如,依诺米那的可逆终止子方法、罗氏的焦磷酸测序方法(454)、生命科技的通过连接的测序(SOLiD平台)、生命科技的Ion Torrent平台或太平洋生物科学'荧光碱基切割方法。这些方法的实例描述于以下参考文献中:Margulies等人(Nature 2005 437:376-80);Ronaghi等人(Analytical Biochemistry 1996 242:84-9);Shendure(Science 2005 309:1728);Imelfort等人(Brief Bioinform.2009 10:609-18);Fox等人(Methods Mol Biol.2009;553:79-108);Appleby等人(Methods Mol Biol.2009;513:19-39)English(PLoS One.2012 7:e47768)和Morozova(Genomics.2008 92:255-64),为了方法和方法的特定步骤的一般描述,包括每个步骤的所有起始产物,试剂和最终产物,将他们通过引用并入。实际上,如果产物在固体载体上扩增(例如,使用依诺米那流动池),那么扩增子可以在基质上就地测序。测序步骤可以使用任何方便的下一代测序方法完成,并且可以产生至少10,000、至少50,000、至少100,000、至少500,000、至少1M、至少10M、至少100M或至少1B个序列读段。在许多情况下,读数是双端读数。

[0073] 在任何实施例中,在消化之前,该方法可以包括对混合核酸样品进行末端封端,使得核酸片段的末端不可用于连接。例如,在上述任何实施例中,在消化之前,该方法可以包括用磷酸酶(例如,CIP、SAP、PLAP或SEAP)处理混合核酸样品,从而从样品的核酸中除去5'磷酸基团并使样品中核酸的5'端不可连接。替代的封端方法也是可能的,包括发夹衔接子

的连接,含有化学阻断基团的衔接子的连接,缺乏5'磷酸的衔接子的连接,封端基团的化学添加,修饰核苷酸的酶介导的添加,酶介导的一个或多个核苷酸的添加,产生粘性末端突出端,其与特定衔接子的未来连接不相容,或任何其它阻止功能性衔接子的有效下游连接的方法。

[0074] 在一些实施例中,所分析的核酸可以源自单一来源(例如,来自不同位点或单个受试者中的时间过程),而在其它实施例中,核酸样品可以是来自多个不同来源中提取的核酸文库(例如,来自不同受试者的核酸文库),其中“多个”是指两个或更多个。因此,在某些实施例中,核酸样品可含有来自2个或更多个来源、3个或更多个来源、5个或更多个来源、10个或更多个来源、50个或更多个来源、100个或更多个来源、500个或更多个来源、1000个或更多个来源、5000个或更多个来源、最多并包括约10,000个或更多个来源的核酸。分子条形码可以允许来自不同来源的序列在分析后进行区分。

[0075] 在一些实施例中,DNA样品可以与封端的衔接子连接以防止扩增。在扩增之前,可以用靶向切口酶(例如Cas9D10A)处理连接的样品。切口产生将允许靶向序列的延伸和仅在与这些序列附接的衔接子中置换封端基团,允许切口分子的PCR扩增,而不能扩增非靶向分子。封端基团的实例包括但不限于脱氧尿嘧啶(dU)、反向脱氧胸苷(dT)、RNA碱基、脱氧异胞嘧啶(isodC)、脱氧异鸟嘌呤(isodG)、1',2'-双脱氧核苷酸(ddNTP)、间隔基、无碱基位点等。该方法的实施例在图4中示出。该方法可包括将衔接子连接至片段的末端,其中衔接子含有封端基团(即,导致聚合酶停滞的基团),使用核酸指导的切口内切核酸酶(例如,Cas9D10A)切割靶片段,使带切口的链延伸以置换含有封端基团的链,然后使用与衔接子杂交的通用引物扩增靶序列。

[0076] 在一些实施例中,核酸样品可包含来自至少两种生物体的DNA,例如哺乳动物和病原体(例如,其中病原体可以是病毒、细菌或真菌),哺乳动物和微生物菌群,或微生物菌群中的两种微生物,其中来自一种生物体的DNA的丰度可以是另一种的至少10倍、至少50倍、或至少100倍、至少500倍或至少1,000倍。在一些实施例中,核酸样品可含有来自相同生物(例如癌症患者)的野生型和突变体DNA。在一些实施例中,混合样品可以是环境样品、来自犯罪现场的样品或考古样品。在一些实施例中,混合样品由临床样品制成,例如来自疑似已被病原体感染的患者。临床样品可以是下面列出的体液或排泄物。在一些实施例中,临床样品可以是肿瘤活检样品。从各种样品(例如临床、法医和环境样品)中提取总DNA和RNA的方法是本领域熟知的。样品包括但不限于皮肤拭子、皮肤活检、唾液、牙拭子、牙齿刮片、面颊拭子、咽拭子、痰、内脏样本、粪便、尿液、阴道、宫颈、宫颈内膜、子宫内膜、鼻拭子、肺、器官活检和组织活检。样品也可以是体液。感兴趣的体液包括但不限于羊水、房水、玻璃体液、血液(例如全血、分馏血液、血浆、血清等)、母乳、脑脊液(CSF)、耳垢(耳垢)、乳糜、边缘、内淋巴、外淋巴、粪便、胃酸、胃液、淋巴液、粘液(包括鼻腔引流和痰)、心包液、腹膜液、胸膜液、脓液、大黄、唾液、皮脂(皮油)、精液、痰液、汗液、滑液、泪液、呕吐物和尿液。在特定实施例中,样品可以是临床样品,例如从患者收集的样品。在某些情况下,这些样品中的DNA可以高度片段化成,例如平均大小为10bp至5kb,例如20bp至200bp,并且在某些情况下可以使用本文所述的方法进行片段化。从这些样品中提取总DNA的方法是众所周知的。

[0077] 上述方法可用于分析各种诊断、药物发现和研究应用中的样品。上述方法可用于分析生物样品。如本文所用,术语“生物样品”是指从生物或从生物的成分(例如细胞)获得

的样品。样品可以是任何生物组织或流体。在一些情况下,样品将是“临床样品”,其是源自患者的样品。这些样品包括但不限于痰液、血液、血细胞(例如白细胞)、组织或细针活检样品、尿液、腹膜液和胸膜液、或来自其的细胞。生物样品还可包括组织切片,例如用于组织学目的的冷冻切片。本方法还可用于确定水、污水、空气样品、食品,包括动物、蔬菜、种子等、土壤样品、植物样品、微生物培养样品、细胞培养样品、组织培养样品中微生物的特性,以及用于人类医学、兽医学、农业、食品科学、生物威胁和工业微生物学等。本方法允许因不需要培养微生物而鉴定难以培养微生物。因此,本方法提供了对样品中微生物的快速检测,不需要等待培养微生物的时间。

[0078] 在一些实施例中,该方法可用于从临床样品中鉴定微生物病原体。在这些实施例中,内切核酸酶可以靶向来自多种不同病原体(例如,至少10种或至少100种不同病原体)的序列,在不知道哪种病原体负责感染下,可以对富集的核酸进行测序,并且可以将序列与来自已知病原体(例如细菌、东方和病毒病原体)的序列相比较,并且如果发现匹配,则可以将受试者诊断为被该病原体感染。在一些实施例中,核酸酶可以用简并识别序列编程,所述简并识别序列代表编码目标分子途径中涉及的一个或多个特定基因的已知保守基序。以这种方式,可以同时来自几个物种的同源基因富集和测序。在另一个实施例中,该方法可用于宿主相关群落(“微生物菌群”)的深度测序,其优点在于该方法能够对样品中的所有微生物DNA进行测序,而不仅仅是DNA的子集,例如,仅核糖体DNA。可以使用本方法,组合物和试剂盒鉴定的微生物包括但不限于:多种革兰氏(+)细菌、多种革兰氏(-)细菌、肠杆菌科(Enterobacteriaceae)中的多种细菌、肠球菌属(Enterococcus)中的多种细菌、葡萄球菌属(Staphylococcus)中的多种细菌、以及弯曲杆菌属(Campylobacter)、大肠杆菌(E.coli)、各种菌株的大肠杆菌例如,K12-MG1655、CFT073、O157:H7EDL933、O157:H7VT2-Sakai等、肺炎链球菌(*Streptococcus pneumoniae*)、铜绿假单胞菌(*Pseudomonas aeruginosa*)、金黄色葡萄球菌(*Staphylococcus aureus*)、凝固酶阴性葡萄球菌、多种念珠菌属(*Candida*),包括白色念珠菌(*C.albicans*)、热带念珠菌(*C.tropicalis*)、都柏林念珠菌(*C.dublinsiensis*)、维斯念珠菌(*C.viswanathii*)、近平滑念珠菌(*C.parapsilosis*)、克雷白氏杆菌(*Klebsiella pneumoniae*)、多种分枝杆菌属(*Mycobacterium*),例如结核分枝杆菌(*M.tuberculosis*)、牛分枝杆菌(*M.bovis*)、牛分枝杆菌BCG、瘰疬分枝杆菌(*M.scrofulaceum*)、堪萨斯分枝杆菌(*M.kansasii*)、龟分枝杆菌(*M.chelonae*)、戈氏分枝杆菌(*M.gordonae*)、溃疡分枝杆菌(*M.ulcerans*)、日内瓦分枝杆菌(*M.genavense*)、*M.xenoi*、猿分枝杆菌(*M.simiae*)、偶发分枝杆菌(*M.fortuitum*)、海鱼分枝杆菌(*M.malmoense*)、隐藏分枝杆菌(*M.celatum*)、嗜血分枝杆菌(*M.haemophilum* and)和非洲分枝杆菌(*M.africanum*)、利斯特氏菌属(*Listeria*)、布鲁氏菌属(*Brucella*)、耶尔森氏菌属(*Yersinia*)等中的多种细菌。因此,本方法能够将微生物鉴定到微生物的属、种、亚种、菌株或变体的水平。

[0079] 疾病状态可能表现出新微生物的存在,正常微生物的不存在,或微生物比例的改变。疾病状态也可具有与正常状态基本相似的微生物群体,但由于环境或宿主遗传因素而具有不同的微生物功能或对微生物的不同宿主反应。例如,最近的研究已经确定,宿主与其微生物群之间的正常平衡的破坏(通常表现为微生物失衡)与许多病症和疾病相关并且可能导致许多病症和疾病。这些尤其包括克罗恩病、溃疡性结肠炎、肥胖症、哮喘、过敏症、代

谢综合征、糖尿病、牛皮癣、湿疹、酒渣鼻、特应性皮炎、胃肠道反流病、胃肠道癌症、细菌性阴道病、神经发育性疾病如自闭症谱系障碍、以及多种感染。例如,在克罗恩病中,拟杆菌、真细菌和蛋白链球菌属的浓度增加,而双歧杆菌数量减少(Linskens等人,Scand J Gastroenterol Suppl.2001;(234):29-40);在溃疡性结肠炎中,兼性厌氧菌的数量增加。在这些炎性肠病中,这种微生物失衡导致增加的免疫刺激和增强的粘膜通透性(Sartor, Proc Natl Acad Sci USA.2008年10月28日;105(43):16413-4)。在肥胖受试者中,相对于瘦人而言,拟杆菌的相对比例已显示出降低(Ley等人,Nature.2006年12月21日;444(7122):1022-3),以及也已经讨论了微生物失衡与糖尿病发展的可能联系(Cani等人, Pathol Biol (巴黎).2008年7月;56(5):305-9)。在皮肤中,已经在传染性和非传染性疾病和障碍中提出了土着微生物群在健康和疾病中的作用,例如特应性皮炎、湿疹、红斑痤疮、牛皮癣和痤疮(Holland等人Br.J.Dermatol.96:623-626;Thomsen等人 Arch.Dermatol.116:1031-1034;Till等人Br.J.Dermatol.142:885-892;Paulino等人 J.Clin.Microbiol.44:2933-2941)。此外,常驻微生物群也可能因皮肤屏障受损而致病(Roth和James Annu Rev Microbiol.1988;42:441-64)。细菌性阴道病是由天然存在的阴道微生物群的失衡引起的。虽然正常的阴道微生物群以乳杆菌为主,但在2级(中间)细菌性阴道病中,除乳杆菌外,还存在加德纳菌(*Gardnerella*)和动物溃疡伯格菌(*Mobiluncus* spp.)。在3级(细菌性阴道病)中,加德纳菌和动物溃疡伯格菌、以及乳杆菌很少或不存在(Hay等人,Br.Med.J.,308,295-298,1994)。鉴定哪些微生物在样品中可以帮助诊断和治疗这些疾病。

[0080] 此外,该方法可用于检测RNA病毒或逆转录病毒,例如呼肠孤病毒、轮状病毒、肠病毒、鼻病毒、肝病毒、心脏病毒、口疮病毒、脊髓灰质炎病毒、双埃柯病毒、马鼻病毒、嗜病毒、捷申病毒、柯萨奇病毒、诺瓦克病毒、风疹病毒、甲病毒、淋巴细胞性脉络丛脑膜炎病毒、登革热病毒、丙型肝炎病毒、黄热病毒、甲型流感病毒、乙型流感病毒、丙型流感病毒、贫病毒、首戈托病毒、麻疹病毒、腮腺炎病毒、呼吸道合胞病毒、牛瘟病毒、犬瘟热病毒、加利福尼亚脑炎病毒、汉坦病毒、狂犬病病毒、埃博拉病毒、马尔堡病毒、冠状病毒、星状病毒、博纳病毒、动脉病毒、马动脉炎病毒、戊型肝炎病毒、逆转录病毒(例如,HIV-1和HIV-2)以及乙型肝炎病毒。

[0081] 该方法还可用于检测微生物DNA或RNA序列并同时检测人DNA或RNA序列,从而允许在同一实验中研究病原体和宿主。

[0082] 该方法也可用于基因分型。在这种情况下,其将靶向病原体或其它物种内的特定位点,这些位点可用于区分特定菌株或其它亚群。在感染了多个亚群或检测微生物菌群或环境样品中的多个亚群的情况下,施用或未施用唯一分子标识符(UMI)的该方法可用于评价不同亚群的相对量。

[0083] 在一些实施例中,该方法可用于确定核酸酶的靶序列,例如限制性内切核酸酶、归巢内切核酸酶和可编程核酸酶,例如Cas9或Cpf1。例如,为了鉴定特定Cas9指导RNA组合的切割位点,可以用Cas9和指导RNA处理用封端的衔接子合成的基因组DNA文库,以允许选择性扩增含有切割位点的序列。

[0084] 该方法的这些和其它用途将是显而易见的。

[0085] 试剂盒

[0086] 本公开还提供了用于实施如上所述的本方法的试剂盒。在某些实施例中,本试剂盒可以包含:a)多种重编程的核酸指导的内切核酸酶;以及b)磷酸酶或其它封端试剂。这样的试剂盒可任选地含有衔接子和PCR引物,其中PCR引物的3'端具有与衔接子中的序列相同或互补的序列。以上描述了该试剂盒的组分的进一步细节。试剂盒还可以含有上文和下文所述的可以在该方法中使用的其它试剂,例如连接酶、聚合酶等,这取决于该方法将如何实施。

[0087] 除了上述组分之外,本试剂盒还包括使用试剂盒的组分来实施本方法的说明书。用于实施本方法的说明书通常记录在合适的记录介质上。例如,说明书可以印刷在基材上,例如纸或塑料等。因此,说明书可以作为包装插页存在于试剂盒中,存在于试剂盒的容器或其组件的标签中(即与包装或分包装相关联)等。在其它实施例中,说明书作为存在于合适的计算机可读存储介质上的电子存储数据文件存在,例如CD-ROM、软盘等。在又其它实施例中,实际说明书不存在于试剂盒中,而是提供用于例如通过互联网从远程源获得说明书的手段。该实施例的实例是包括网址的试剂盒,在该网址中可以查看说明书和/或可以从该网址下载说明书。与说明书一样,这种用于获得说明书的手段记录在合适的基材上。

[0088] 为了进一步说明本发明,给出以下具体实施例,应理解,提供他们是为了说明本发明,并且不应以任何方式解释为限制其范围。

[0089] 实施例

[0090] 实施例1A.一种样品分析方法,其包括:(a)使用多种重编程的核酸指导的内切核酸酶消化末端封端的混合核酸样品以产生消化样品,所述内切核酸酶靶向目标序列,其中消化样品中至少一些片段包含:(i)目标序列和(ii)至少一个通过内切核酸酶切割产生的可连接末端;(b)富集含有目标序列的片段;以及(c)分析富集的片段。

[0091] 实施例1B.一种样品分析方法,其包括:

[0092] (a)使用多种重编程的核酸指导的内切核酸酶消化磷酸酶处理的混合核酸样品以产生消化样品,所述内切核酸酶靶向目标序列,其中消化样品中至少一些片段包含:(i)目标序列和(ii)至少一个通过内切核酸酶切割产生的可连接末端;(b)富集含有目标序列的片段;以及(c)分析富集的片段。

[0093] 实施例2.根据实施例1A或B所述的方法,其中所述方法包括将衔接子连接至步骤(a)中通过内切核酸酶切割产生的可连接末端。

[0094] 实施例3.根据实施例2所述的方法,其中所述衔接子包含捕获部分,并且所述富集通过将捕获部分与载体结合并洗去未结合的核酸来进行。

[0095] 实施例4.根据以上任一实施例所述的方法,其中(a)的内切核酸酶消化产生在限定大小范围内的片段。

[0096] 实施例5.根据实施例4所述的方法,其中所述富集通过选择片段的大小来进行。

[0097] 实施例6.根据以上任一实施例所述的方法,其中(a)的内切核酸酶消化产生两端均可连接的片段。

[0098] 实施例7.根据以上任一实施例所述的方法,其中(a)的消化产生在两端具有可连接的内切核酸酶切割位点的片段,并且所述方法包括将衔接子连接到片段的两端。

[0099] 实施例8.根据实施例7所述的方法,其中所述富集通过使用与衔接子或其补体杂交的引物扩增连接的片段来进行。

- [0100] 实施例9.根据实施例8所述的方法,其中所述衔接子含有分子索引器。
- [0101] 实施例10.根据以上任一实施例所述的方法,其中所述方法包括将编索引的衔接子连接到片段的两端,其中所述索引是一系列随机核苷酸,其长度足以使每个片段有较高概率接受不同索引;对连接的片段测序以产生序列读段;然后对与序列读段中的目标序列相联的分子索引器序列的数量进行计数,从而提供核酸样品中目标序列的拷贝数的估计值。
- [0102] 实施例11.根据以上任一实施例所述的方法,其中所述方法包括:
- [0103] (a) 单独地:
- [0104] (I) 使用靶向目标序列的多种重编程的核酸指导的内切核酸酶消化混合核酸样品的第一部分,以产生第一消化样品,其中消化样品中的至少一些片段包含:(i) 第一目标序列和(ii) 至少一个通过内切核酸酶切割产生的可连接末端;
- [0105] (II) 使用靶向目标序列的多种重编程的核酸指导的内切核酸酶消化混合核酸样品的第二部分,以产生第二消化样品,其中消化样品中的至少一些片段包含:(i) 第二目标序列和(ii) 至少一个通过内切核酸酶切割产生的可连接末端;
- [0106] 其中第一消化样品中的至少一些片段与第二消化样品中的至少一些片段重叠;
- [0107] (b) 富集含有目标序列的片段;
- [0108] (c) 对富集的序列进行测序以产生多个序列读段;以及
- [0109] (d) 组装重叠序列读段,从而获得第一目标序列和第二目标序列的重叠群。
- [0110] 实施例12.根据以上任一实施例所述的方法,其中通过内切核酸酶处理产生的可连接末端是平末端。
- [0111] 实施例13.根据以上任一实施例所述的方法,其中混合核酸样品包含来自至少两种生物的DNA。
- [0112] 实施例14.根据实施例13所述的方法,其中至少两种生物包含哺乳动物和病原体。
- [0113] 实施例15.根据实施例14所述的方法,其中所述病原体是病毒、细菌或真菌。
- [0114] 实施例16.根据实施例15所述的方法,其中所述至少两种生物包含哺乳动物和微生物菌群。
- [0115] 实施例17.根据以上任一实施例所述的方法,其中所述混合样品包含野生型和突变体DNA。
- [0116] 实施例18.根据以上任一实施例所述的方法,其中所述混合样品是环境样品、来自犯罪现场的样品或考古样品。
- [0117] 实施例19.根据以上任一实施例所述的方法,其中所述混合样品由临床样品构成。
- [0118] 实施例20.根据实施例19所述的方法,其中所述临床样品是体液或排泄物。
- [0119] 实施例21.根据实施例20所述的方法,其中所述临床样品是血液、痰或粪便。
- [0120] 实施例22.根据实施例20所述的方法,其中所述临床样品是肿瘤活检样品。
- [0121] 实施例23.根据以上任一实施例所述的方法,其中(c)中的分析通过测序进行。
- [0122] 实施例24.根据以上任一实施例所述的方法,其中所述内切核酸酶是cas9或Argonaught、其直向同源物或其变体。
- [0123] 实施例25.根据以上任一实施例所述的方法,其中所述样品通过至少2种重编程的核酸指导的内切核酸酶消化。

[0124] 实施例26. 根据以上任一实施例所述的方法, 其中在步骤 (a) 之前, 所述方法包括用磷酸酶处理混合核酸样品, 从而使样品中核酸的5'端不可连接。

[0125] 实施例27A. 一种包含多种重编程的核酸指导的内切核酸酶; 和磷酸酶的试剂盒。

[0126] 实施例27B. 一种包含多种重编程的核酸指导的内切核酸酶; 和磷酸酶或其它末端封端试剂的试剂盒。

[0127] 实施例28. 根据实施例27A或B所述的试剂盒, 还包含衔接子。

[0128] 实施例29. 根据实施例27A或B所述的试剂盒, 还包含PCR引物, 其中所述PCR引物的3'端具有与衔接子中的序列相同或互补的序列。

[0129] 实例

[0130] 给出以下实例是为了说明本发明的各种实施例, 并不意味着以任何方式限制本发明。本实例以及本文所述的方法目前代表优选实施例, 是示例性的, 并不意图作为对本发明范围的限制。本领域技术人员将想到其中的变化和包含在由权利要求的范围限定的本发明的精神内的其它用途。

[0131] 使用FLASH靶向抗生素抗性基因, 证明相对于源自患者样品的随机片段化cDNA文库, 如mecA的序列的富集超过10倍。

[0132] CAS9-DGRNA复合物的设计与构建

[0133] 在本研究中选择了13种能够在金黄葡萄球菌中赋予抗生素抗性的基因用于FLASH。九种 (mecA、Qnr、qacB、MFS、mdeA、NorA、NorB/C、KCTC、和PA01) 是质粒携带基因, 其单独存在指示抗性; 另外四种 (促旋酶A、促旋酶B、parC和parE) 是金黄葡萄球菌内源性基因, 已知其特定点突变会引起喹诺酮抗性。对于质粒携带基因, 在基因两端附近选择指导RNA位点, 并且在基因内约每隔200-300bp。对于具有点突变的基因, 侧接每个突变选择两个位点, 其距离将产生200-300bp的切割产物。指导RNA的列表及其在基因内的位置可以在下表1中找到。

[0134] 表1:

[0135]

基因	位置	crRNA 序列
GyrA	111	CAAAACAGCATAGCTCTAAAACGATGTTTCGTGACGGTTTAAACTAT AGTGAGTCGTATTA (SEQ ID NO:1)
GyrA	334	CAAAACAGCATAGCTCTAAAACCTTGAACCAAAGTTACCTTGGCTAT AGTGAGTCGTATTA (SEQ ID NO:2)
GyrB	1132	CAAAACAGCATAGCTCTAAAACCAAGTCGCACGTACAGTGGTCTA TAGTGAGTCGTATTA (SEQ ID NO:3)
GyrB	1451	CAAAACAGCATAGCTCTAAAACAATGCTGTGATCATTGACGCTAT AGTGAGTCGTATTA (SEQ ID NO:4)
GyrB	1837	CAAAACAGCATAGCTCTAAAACCTTCAATCGCATCTTCAAGTCTAT

[0136]

		AGTGAGTCGTATTA (SEQ ID NO:5)
GyrB	248	CAAAACAGCATAGCTCTAAAACCGTCCGTTATCCGTTACTTTCTATA GTGAGTCGTATTA (SEQ ID NO:6)
GyrB	45	CAAAACAGCATAGCTCTAAAACAGCACCATAATTATCCGTGTCTAT AGTGAGTCGTATTA (SEQ ID NO:7)
GyrB	536	CAAAACAGCATAGCTCTAAAACCTCTGCTTTAAAACGAATGACCTAT AGTGAGTCGTATTA (SEQ ID NO:8)
GyrB	917	CAAAACAGCATAGCTCTAAAACCTAACTATTTAAGACACGCGTCTAT AGTGAGTCGTATTA (SEQ ID NO:9)
KCTC	1129	CAAAACAGCATAGCTCTAAAACCAAGAATATGATACGCCGCCCTAT AGTGAGTCGTATTA (SEQ ID NO:10)
KCTC	19	CAAAACAGCATAGCTCTAAAACCTTACAGGCGTACTGCTGCTGCTAT AGTGAGTCGTATTA (SEQ ID NO:11)
KCTC	327	CAAAACAGCATAGCTCTAAAACGCGCAGAAGTACTGGCCGCACTA TAGTGAGTCGTATTA (SEQ ID NO:12)
KCTC	521	CAAAACAGCATAGCTCTAAAACATGCTGGCATTGGCGTACAGCTAT AGTGAGTCGTATTA (SEQ ID NO:13)
KCTC	807	CAAAACAGCATAGCTCTAAAACCTGCCGAGATTTTATCCTCTAT AGTGAGTCGTATTA (SEQ ID NO:14)
mdeA	1343	CAAAACAGCATAGCTCTAAAACGCGTGATAACCATGCAACATCTAT AGTGAGTCGTATTA (SEQ ID NO:15)
mdeA	381	CAAAACAGCATAGCTCTAAAACACCACATGAACGCGGTAAATCTA TAGTGAGTCGTATTA (SEQ ID NO:16)
mdeA	44	CAAAACAGCATAGCTCTAAAACCCCATCAGCATAGTAATAATCTAT AGTGAGTCGTATTA (SEQ ID NO:17)
mdeA	629	CAAAACAGCATAGCTCTAAAACCTACAATAACCCGCCGAAACCCTAT AGTGAGTCGTATTA (SEQ ID NO:18)
mdeA	989	CAAAACAGCATAGCTCTAAAACACTAAAGCAAGATTTCTAGGCTAT AGTGAGTCGTATTA (SEQ ID NO:19)
mecA	1156	CAAAACAGCATAGCTCTAAAACAAGATAAAAAAGAACCTCTGCTA TAGTGAGTCGTATTA (SEQ ID NO:20)
mecA	1418	CAAAACAGCATAGCTCTAAAACAATTCGAGTGCTACTCTAGCCTAT AGTGAGTCGTATTA (SEQ ID NO:21)
mecA	1668	CAAAACAGCATAGCTCTAAAACAACCTTTGTTTTTCGTGTCTTCTAT AGTGAGTCGTATTA (SEQ ID NO:22)
mecA	1882	CAAAACAGCATAGCTCTAAAACCTCATCATGTTTGGATTATCTCTATA GTGAGTCGTATTA (SEQ ID NO:23)
mecA	2090	CAAAACAGCATAGCTCTAAAACCTTAAATACAATTTCTTCATTCTATA GTGAGTCGTATTA (SEQ ID NO:24)
mecA	215	CAAAACAGCATAGCTCTAAAACAACTATTATATATTTTATCTATA GTGAGTCGTATTA (SEQ ID NO:25)
mecA	28	CAAAACAGCATAGCTCTAAAACCTTATTTAATAGTTGTAGTCTATA GTGAGTCGTATTA (SEQ ID NO:26)
mecA	475	CAAAACAGCATAGCTCTAAAACCATGTTTTCGGTCTAAAATTCTAT AGTGAGTCGTATTA (SEQ ID NO:27)
mecA	735	CAAAACAGCATAGCTCTAAAACCTAGAAAAAGCGACTTCACACTA TAGTGAGTCGTATTA (SEQ ID NO:28)

[0137]

mecA	95	CAAAACAGCATAGCTCTAAAACTCTTTTTCTTTTTCTCTATCTATA GTGAGTCGTATTA (SEQ ID NO:29)
MFS	1294	CAAAACAGCATAGCTCTAAAACTTGCGATGTACTTTCATTGACTAT AGTGAGTCGTATTA (SEQ ID NO:30)
MFS	305	CAAAACAGCATAGCTCTAAACGTAAGTATTATCGGCATCATCTATA GTGAGTCGTATTA (SEQ ID NO:31)
MFS	588	CAAAACAGCATAGCTCTAAACGCGCTGCTTATACCGCTATACTAT AGTGAGTCGTATTA (SEQ ID NO:32)
MFS	75	CAAAACAGCATAGCTCTAAAACTAACAAAGCACAGCTACAAAATA TAGTGAGTCGTATTA (SEQ ID NO:33)
MFS	894	CAAAACAGCATAGCTCTAAACAACCGAAAACAAAGGCAAGTCT ATAGTGAGTCGTATTA (SEQ ID NO:34)
NorA	1100	CAAAACAGCATAGCTCTAAACGCTAATGAAACACCTATAGCCTAT AGTGAGTCGTATTA (SEQ ID NO:35)
NorA	410	CAAAACAGCATAGCTCTAAACGAATTGATAATCGCTGACATCTAT AGTGAGTCGTATTA (SEQ ID NO:36)
NorA	748	CAAAACAGCATAGCTCTAAACCAAATATACCGCCACCCGTA AGTGAGTCGTATTA (SEQ ID NO:37)
NorA	76	CAAAACAGCATAGCTCTAAACGCTTGCCTGTTTATTAACTAT AGTGAGTCGTATTA (SEQ ID NO:38)
NorB/C	1360	CAAAACAGCATAGCTCTAAACAATGATGATAAACGTGTCAACTAT AGTGAGTCGTATTA (SEQ ID NO:39)
NorB/C	17	CAAAACAGCATAGCTCTAAACCCGCGATACGTTTCATTCATCTAT AGTGAGTCGTATTA (SEQ ID NO:40)
NorB/C	355	CAAAACAGCATAGCTCTAAAACTCCACATTGGCCATTATGAACTAT AGTGAGTCGTATTA (SEQ ID NO:41)
NorB/C	650	CAAAACAGCATAGCTCTAAAACTTAGTAATGACAACGTTTAACTAT AGTGAGTCGTATTA (SEQ ID NO:42)
NorB/C	995	CAAAACAGCATAGCTCTAAACAATAACATTGGTCGCTTAGACTAT AGTGAGTCGTATTA (SEQ ID NO:43)
PAO1	1166	CAAAACAGCATAGCTCTAAAACTGCTGCTCCAGGCCGCTGAGCTA TAGTGAGTCGTATTA (SEQ ID NO:44)
PAO1	16	CAAAACAGCATAGCTCTAAACGATCCCTGCCTGTGCGGCCTAT AGTGAGTCGTATTA (SEQ ID NO:45)
PAO1	260	CAAAACAGCATAGCTCTAAACATCTCGAACAGGGTCTCCGGCTA TAGTGAGTCGTATTA (SEQ ID NO:46)
PAO1	489	CAAAACAGCATAGCTCTAAACCTGGCGGTAGTAGTCGCGGACTA TAGTGAGTCGTATTA (SEQ ID NO:47)
PAO1	708	CAAAACAGCATAGCTCTAAAACTAGCGGGCGGTCGTCCTTGCCTAT AGTGAGTCGTATTA (SEQ ID NO:48)
PAO1	954	CAAAACAGCATAGCTCTAAACATGGCGCTGCAACCGCACAGCTA TAGTGAGTCGTATTA (SEQ ID NO:49)
parC	123	CAAAACAGCATAGCTCTAAACGTACAACGTCGTATTTATACTATA GTGAGTCGTATTA (SEQ ID NO:50)
parC	342	CAAAACAGCATAGCTCTAAACCCAGCTGCAATGCGTTACACCTAT AGTGAGTCGTATTA (SEQ ID NO:51)
parE	1158	CAAAACAGCATAGCTCTAAAACTGCTTGTGTGCTTTAATCGCTAT

[0138]

		AGTGAGTCGTATTA (SEQ ID NO:52)
parE	1448	CAAAACAGCATAGCTCTAAAACGCCCGATTGTGTGGATAATCTAT AGTGAGTCGTATTA (SEQ ID NO:53)
qacB	1007	CAAAACAGCATAGCTCTAAAACAATCGCGCCGCTAATCCGGGCTAT AGTGAGTCGTATTA (SEQ ID NO:54)
qacB	1232	CAAAACAGCATAGCTCTAAAACAGGTCATACATAGACTCTTCCTAT AGTGAGTCGTATTA (SEQ ID NO:55)
qacB	1464	CAAAACAGCATAGCTCTAAAACACCTACTAAAGCAGTTGCTACTAT AGTGAGTCGTATTA (SEQ ID NO:56)
qacB	152	CAAAACAGCATAGCTCTAAAACGAAGGCTCTAACTCTTACCTAT AGTGAGTCGTATTA (SEQ ID NO:57)
qacB	356	CAAAACAGCATAGCTCTAAAACGCAATACCAAGTAAAATCGCTAT AGTGAGTCGTATTA (SEQ ID NO:58)
qacB	618	CAAAACAGCATAGCTCTAAAACCGAGTGAGACTTTTCTTTTGCTAT AGTGAGTCGTATTA (SEQ ID NO:59)
Qnr	241	CAAAACAGCATAGCTCTAAAACAGCGCGCTCACATTCCTGAACTAT AGTGAGTCGTATTA (SEQ ID NO:60)
Qnr	346	CAAAACAGCATAGCTCTAAAACGTACCTGAGTACCCATCCA ACTAT AGTGAGTCGTATTA (SEQ ID NO:61)
Qnr	44	CAAAACAGCATAGCTCTAAAACGTGAATCTGTTTCTGTCAATCTAT AGTGAGTCGTATTA (SEQ ID NO:62)
Qnr	635	CAAAACAGCATAGCTCTAAAACATGACAGCGATGCCAAGACGCTA TAGTGAGTCGTATTA (SEQ ID NO:63)

[0139] 由于降低了成本,因此选择双指导RNA而不是单指导RNA。对于每个选择的靶标,设计了60mer crRNA模板寡核苷酸,其包括18个碱基的T7转录起始位点,靶向20mer和crRNA的剩余22个碱基;购买该序列的反向互补(5'-CAAAACAGCATAGCTCTAAAACNNNNNNNNNNNNNNNNNNNNNNNNCTATAGTGAGTCGTATTA-3') (SEQ ID NO:64)。所有寡核苷酸均购自IDT(Integrated DNA Technologies, Coralville, IA, USA)。汇集所有crRNA用于转录。使用定制的T7RNA聚合酶(RNAP)进行转录。在每个100 μ L反应中,将400ng DNA模板与T7RNAP(最终浓度为8ng/ μ L),缓冲液(最终浓度为40mM Tris pH8.0、20mM MgCl₂、5mM DTT和2mM亚精胺)和NTP(ThermoFisher Scientific, Waltham, MA, USA)(最终浓度为各1mM ATP、CTP、GTP和UTP)混合,并在37 $^{\circ}$ C下温育4小时。典型的产量为10至50 μ g RNA。用Zymo RNA Clean&Concentrator-5试剂盒(Zymo Research, Irvine, CA, USA)纯化crRNA。

[0140] 合成tracrRNA获自IDT。通过以等摩尔浓度混合两种组分,将混合物在95度下加热1分钟,然后在操作台上将其缓慢冷却至室温,将crRNA文库与tracrRNA复合。将得到的dgRNA等分,保存在-80 $^{\circ}$ C,解冻后仅使用一次。

[0141] 如Gu和Crawford等人所述,表达并纯化用mRuby2标记的Cas9。2016[1]。

[0142] FLASH和深度测序培养物和患者样品

[0143] 使用标准程序从患者培养分离物(样品31和33)制备基因组DNA。使用用于低核酸含量样品的NuGEN Ovation v.2试剂盒(NuGEN, San Carlos, CA, USA)由患者支气管肺泡灌洗(BAL)液(样品288和1002)制备cDNA,如[2,3]中所述。将来自每个样品的10纳克DNA用小牛肠磷酸酶(CIP)在37 $^{\circ}$ C处理30分钟,以使存在的所有5'DNA末端去磷酸化。将样品自旋柱纯化以除去CIP。通过在1X Cas9活性缓冲液(最终浓度为50mM Tris pH 8.0、100mM NaCl、

10mM MgCl₂和1mM TCEP) 中将浓度各40μM的RNA文库与Cas9混合,并在37℃温育10分钟来形成RNP复合物。然后将Cas9-指导RNA复合物加入到CIP处理的DNA的1X Cas9活性缓冲液中,最终浓度为10μM RNP,总体积为20μL。在37℃温育2小时后,用蛋白酶K在37℃下处理样品15分钟,然后在95℃煮沸5分钟。然后用ampure珠以1:1.4的比率清洗样品。将他们在20μL水中洗脱。

[0144] 作为对照,根据制造商的说明书,用NEBNext片段化酶处理另10ng各DNA样品25分钟。

[0145] 根据制造商的说明书,使用NEBNext Ultra II文库制备试剂盒将Cas9切割的DNA和片段化酶切割的DNA样品转化为依诺米那测序文库,并用双索引条形码引物扩增。

[0146] 用ampure珠以1:0.9的比率纯化扩增样品,然后通过Qubit定量并以相等的浓度合并。生物分析仪高灵敏度DNA试剂盒迹线在约140bp处显示大峰,表明衔接子二聚体。因此,样品用ampure珠以1:0.8的比率再次纯化,并使用KAPA文库扩增试剂盒和依诺米那通用P5和P7引物通过6次以上的PCR循环进行扩增。重复的生物分析仪迹线显示,已经去除了140bp处的峰,并且保留了以200-300bp为中心的宽峰。通过ddPCR (Bio-Rad, Hercules, CA, USA) 定量文库。使用125bp双端测序在依诺米那HiSeq 2500仪器中对两个泳道的部分进行测序。

[0147] 数据使用PriceSeqFilter[4]过滤,然后通过内部病原体发现管道(如Gu和Crawford等人2016[1]所述)。

[0148] 结果

[0149] 图3描绘了用FLASH富集靶向抗生素抗性基因。在没有FLASH的情况下,mecA在培养分离物中很容易检测到,但在由患者样品制备的cDNA中难以检测到。FLASH将mecA覆盖率提高了12倍。

[0150] 讨论

[0151] FLASH方法不限于文中描述的方案。在其它情况下,除了衔接子连接,可以通过大小选择分离释放的DNA部分,然后在分离后扩增或连接。在已知特定目标基因中仅一个可切割序列的情况下,序列特异性核酸酶消化后可以连接附着于标记(例如生物素)的化学标记的寡核苷酸,以捕获和纯化所需序列。然后可以使用随机片段化,然后进行另一个衔接子连接步骤来纯化和扩增含有原始序列的DNA分子。

[0152] 为了向该技术添加定量方面,衔接子可以包括UMI(唯一分子标识符)序列。UMI是与衔接子或索引引物上的条形码相邻的简并序列(例如NNNNNNNNNNNN)。如果置于索引引物上,在索引PCR的第一个循环中,样品中的各单个分子将用含有相同条形码但不同的唯一UMI的引物扩增。在随后的循环中,将扩增UMI序列。当对文库进行测序时,对每个测序片段计数的UMI数量表明核酸酶处理后原始样品中存在的该片段的拷贝数。

[0153] 此外,可以用简并序列构建指导RNA,以靶向可能含有未知物种(或其基因组未测序的已知物种)的样品中的序列的保守区域。例如,这将允许剖析如人类微生物菌群样品的宏基因组样品中存在的代谢能力。

[0154] 参考文献

[0155] 1. Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED, Retallack H等人,通过杂交耗竭高丰度序列(DASH):在测序文库和分子计数应用中使用Cas9去除不必要的高丰度物种(Depletion of Abundant Sequences by Hybridization (DASH):using Cas9to

remove unwanted high-abundance species in sequencing libraries and molecular counting applications.),《基因组生物学(Genome Biol.)》2016;17:1-13。

[0156] 2.Wilson MR,Naccache SN,Samayoa E,Biagtan M,Bashir H,Yu G等人,通过新一代测序对神经钩端螺旋体病进行有指导意义的诊断(Actionable Diagnosis of Neuroleptospirosis by Next-Generation Sequencing.),《新英格兰医学杂志(N.Engl.J.Med.)》2014;370:2408-17。

[0157] 3.Wilson MR,Shanbhag NM,Reid MJ,Singhal NS,Gelfand JM,Sample HA等人,利用宏基因组深度测序诊断巴氏阿米巴原虫性脑炎(Diagnosing Balamuthia mandrillaris Encephalitis With Metagenomic Deep Sequencing.),《神经学年刊(Ann.Neurol.)》2015;78:722-30。

[0158] 4.Ruby JG,Bellare P,Derisi JL.,PRICE:用于靶向组装(宏)基因组序列数据各组件的软件(PRICE:software for the targeted assembly of components of(Meta)genomic sequence data.),G3(马里兰州贝塞斯塔(Bethesda Md.)),2013;3:865-80。

序列表

<110> Crawford, Emily D

Chow, Eric D

Derisi, Joseph L

<120> 通过杂交找到低丰度序列(FLASH)的方法

<130> UCSF-546W0

<150> US 62/375,789

<151> 2016-08-16

<160> 64

<170> PatentIn 3.5版

<210> 1

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 1

caaaacagca tagctctaaa acgatgttcg tgacggttta aactatagtg agtcgtatta 60

<210> 2

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 2

caaaacagca tagctctaaa acttgaacca aagttacctt ggctatagtg agtcgtatta 60

<210> 3

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 3

caaaacagca tagctctaaa accaagtcgc acgtacagtg gtctatagtg agtcgtatta 60

<210> 4

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 4

caaaacagca tagctctaaa acaatgctgt gatcatttga cgctatagtg agtcgtatta 60

<210> 5

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 5

caaaacagca tagctctaaa accttcaatc gcattctcaa gtctatagtg agtcgtatta 60

<210> 6

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 6

caaaacagca tagctctaaa accgtccggt atccgttact ttctatagtg agtcgtatta 60

<210> 7

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 7

caaaacagca tagctctaaa acagcacat aattatccgt gtctatagtg agtcgtatta 60

<210> 8

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 8

caaaacagca tagctctaaa actctgcttt aaaacgaatg acctatagtg agtcgtatta 60

<210> 9

<211> 60

<212> DNA

<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 9
caaaacagca tagctctaaa actaactatt taagacacgc gtctatagtg agtcgtatta 60
<210> 10
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 10
caaaacagca tagctctaaa accaagaata tgatagcgcg ccctatagtg agtcgtatta 60
<210> 11
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 11
caaaacagca tagctctaaa acttacaggc gtactgctgc tgctatagtg agtcgtatta 60
<210> 12
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 12
caaaacagca tagctctaaa acgcgagaa gtactggccg cactatagtg agtcgtatta 60
<210> 13
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 13
caaaacagca tagctctaaa acatgctggc attggcgtac agctatagtg agtcgtatta 60
<210> 14
<211> 60

<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 14
caaaacagca tagctctaaa acctgccgca gatttttatc ctctatagtg agtcgtatta 60
<210> 15
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 15
caaaacagca tagctctaaa acgcgtgata accatgcaac atctatagtg agtcgtatta 60
<210> 16
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 16
caaaacagca tagctctaaa acaccacatg aacgcggtaa atctatagtg agtcgtatta 60
<210> 17
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 17
caaaacagca tagctctaaa accccatcag catagtaata atctatagtg agtcgtatta 60
<210> 18
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 18
caaaacagca tagctctaaa actacaataa cccgccgaaa ccctatagtg agtcgtatta 60
<210> 19

<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 19
caaaacagca tagctctaaa aactaaagc aagatttcta ggctatagtg agtcgtatta 60
<210> 20
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 20
caaaacagca tagctctaaa acaagataaa aaagaacctc tgctatagtg agtcgtatta 60
<210> 21
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 21
caaaacagca tagctctaaa acaattcgag tgctactcta gcctatagtg agtcgtatta 60
<210> 22
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 22
caaaacagca tagctctaaa acaactttgt ttttcgtgtc ttctatagtg agtcgtatta 60
<210> 23
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 23
caaaacagca tagctctaaa actcatcatg tttggattat ctctatagtg agtcgtatta 60

<210> 24
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 24
caaaacagca tagctctaaa acttaaatac aatttettca ttctatagtg agtcgtatta 60
<210> 25
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 25
caaaacagca tagctctaaa acaaactatt atatattttt atctatagtg agtcgtatta 60
<210> 26
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 26
caaaacagca tagctctaaa accttatttt aatagttgta gtctatagtg agtcgtatta 60
<210> 27
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 27
caaaacagca tagctctaaa accattgttt cggctctaaaa ttctatagtg agtcgtatta 60
<210> 28
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 28

caaaacagca tagctctaaa acctagaaaa agcgacttca cactatagtg agtcgtatta 60
<210> 29
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 29

caaaacagca tagctctaaa actctttttt ctttttctct atctatagtg agtcgtatta 60
<210> 30
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 30

caaaacagca tagctctaaa acttgcgatg tactttcatt gactatagtg agtcgtatta 60
<210> 31
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 31

caaaacagca tagctctaaa acgtagtatt tatcgcatc atctatagtg agtcgtatta 60
<210> 32
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 32

caaaacagca tagctctaaa acgcgctgct tataccgcta tactatagtg agtcgtatta 60
<210> 33
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸

<400> 33
caaaacagca tagctctaaa actaacaagc acagctacaa aactatagtg agtcgtatta 60
<210> 34
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 34
caaaacagca tagctctaaa acaaccgaaa acaaaggcaa gtctatagtg agtcgtatta 60
<210> 35
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 35
caaaacagca tagctctaaa acgctaataa aacacctata gcctatagtg agtcgtatta 60
<210> 36
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 36
caaaacagca tagctctaaa acgaattgat aatcgctgac atctatagtg agtcgtatta 60
<210> 37
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 37
caaaacagca tagctctaaa accaaatata cgcaccaccg tactatagtg agtcgtatta 60
<210> 38
<211> 60
<212> DNA
<213> 人工序列
<220>

<223> 合成寡核苷酸

<400> 38

caaaacagca tagctctaaa acgtcttgcc tgtttattta aactatagtg agtcgtatta 60

<210> 39

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 39

caaaacagca tagctctaaa acaatgatga taaacgtgtc aactatagtg agtcgtatta 60

<210> 40

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 40

caaaacagca tagctctaaa acccgcgata cgtttcattc atctatagtg agtcgtatta 60

<210> 41

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 41

caaaacagca tagctctaaa actccacatt ggccattatg aactatagtg agtcgtatta 60

<210> 42

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 42

caaaacagca tagctctaaa acttagtaat gacaacgttt aactatagtg agtcgtatta 60

<210> 43

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 43

caaaacagca tagctctaaa acaataacat tggtcgctta gactatagtg agtcgtatta 60

<210> 44

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 44

caaaacagca tagctctaaa actgctgctc caggccgctg agctatagtg agtcgtatta 60

<210> 45

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 45

caaaacagca tagctctaaa acgattcccc tgctgtgcg gcctatagtg agtcgtatta 60

<210> 46

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 46

caaaacagca tagctctaaa acatctcgaa cagggtctcc ggctatagtg agtcgtatta 60

<210> 47

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 47

caaaacagca tagctctaaa acctggcggg agtagtcgcg gactatagtg agtcgtatta 60

<210> 48

<211> 60

<212> DNA

<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 48
caaaacagca tagctctaaa actagcgggc ggtcgtcctt gcctatagtg agtcgtatta 60
<210> 49
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 49
caaaacagca tagctctaaa acatggegct gcaaccgcac agctatagtg agtcgtatta 60
<210> 50
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 50
caaaacagca tagctctaaa acgtacaacg tcgtatttta tactatagtg agtcgtatta 60
<210> 51
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 51
caaaacagca tagctctaaa acccagctgc aatgcgttac acctatagtg agtcgtatta 60
<210> 52
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 52
caaaacagca tagctctaaa actgcttggtt gtgctttaat cgctatagtg agtcgtatta 60
<210> 53
<211> 60

<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 53
caaaacagca tagctctaaa acgccccgat tgtgtggata atctatagtg agtcgtatta 60
<210> 54
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 54
caaaacagca tagctctaaa acaatcgcgc cgctaaccg ggctatagtg agtcgtatta 60
<210> 55
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 55
caaaacagca tagctctaaa acaggtcata catagactct tcctatagtg agtcgtatta 60
<210> 56
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 56
caaaacagca tagctctaaa acacctacta aagcagttgc tactatagtg agtcgtatta 60
<210> 57
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 57
caaaacagca tagctctaaa acgaaggctc taactctctt acctatagtg agtcgtatta 60
<210> 58

<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 58
caaaacagca tagctctaaa acgcaatacc aagtaaaaat cgctatagtg agtcgtatta 60
<210> 59
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 59
caaaacagca tagctctaaa accgagtgag acttttcttt tgctatagtg agtcgtatta 60
<210> 60
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 60
caaaacagca tagctctaaa acagcgcgct cacattcctg aactatagtg agtcgtatta 60
<210> 61
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 61
caaaacagca tagctctaaa acgtacctga gtacccatcc aactatagtg agtcgtatta 60
<210> 62
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成寡核苷酸
<400> 62
caaaacagca tagctctaaa acgtgaatct gtttctgtca atctatagtg agtcgtatta 60

<210> 63

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<400> 63

caaaacagca tagctctaaa acatgacagc gatgccaaga cgctatagtg agtcgtatta 60

<210> 64

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成寡核苷酸

<220>

<221> misc_feature

<222> (23) .. (42)

<223> n是a、c、g或t

<400> 64

caaaacagca tagctctaaa acnnnnnnnn nnnnnnnnn nnctatagtg agtcgtatta 60

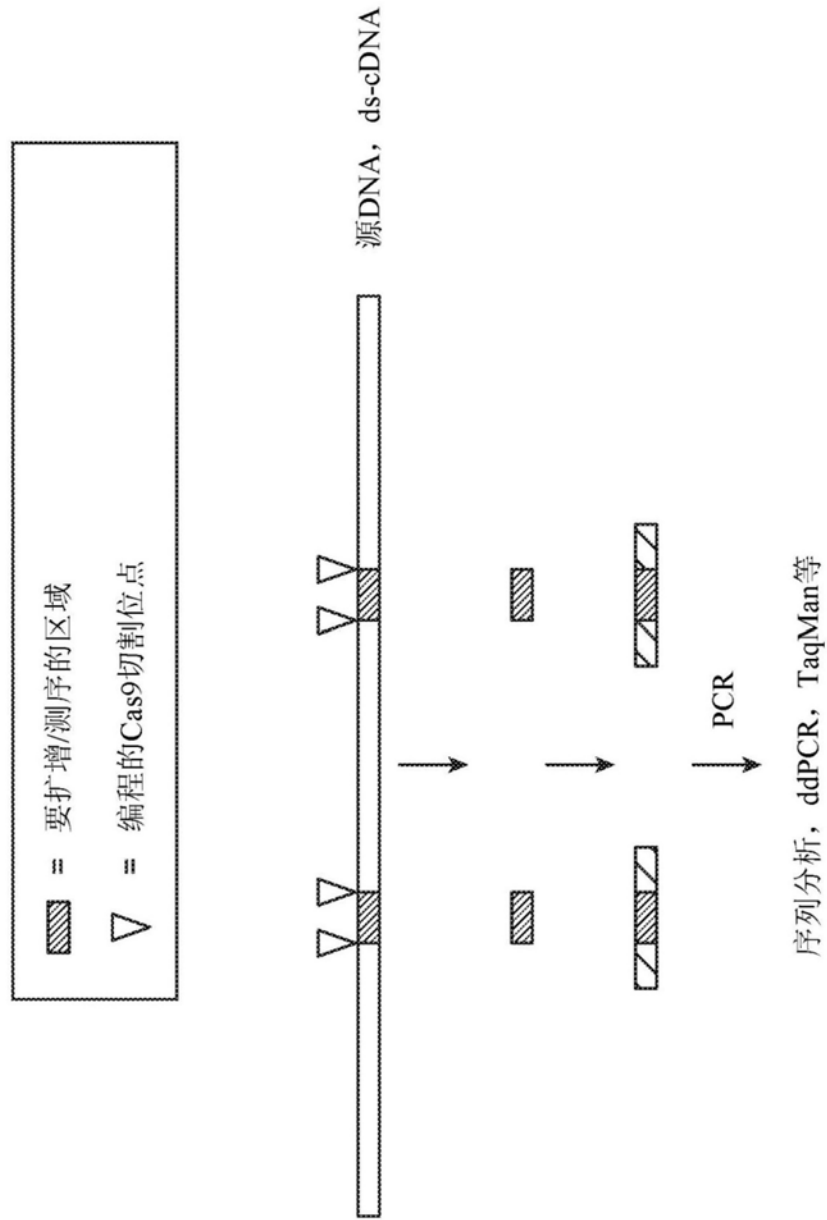


图1

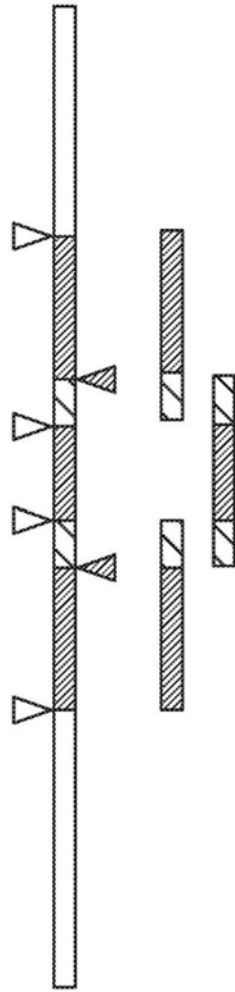


图2

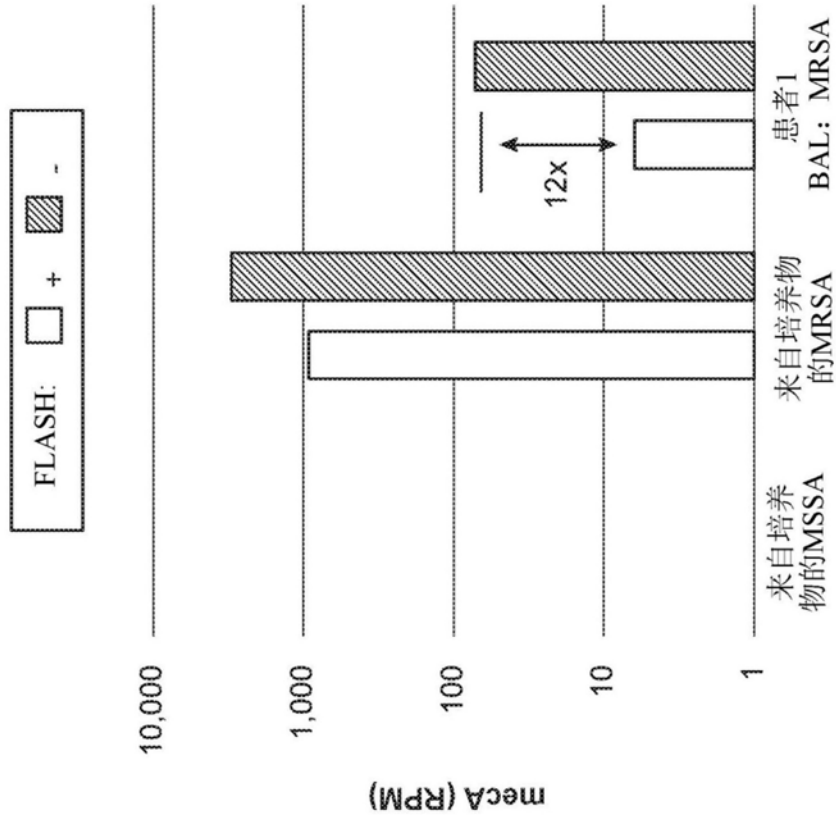


图3A

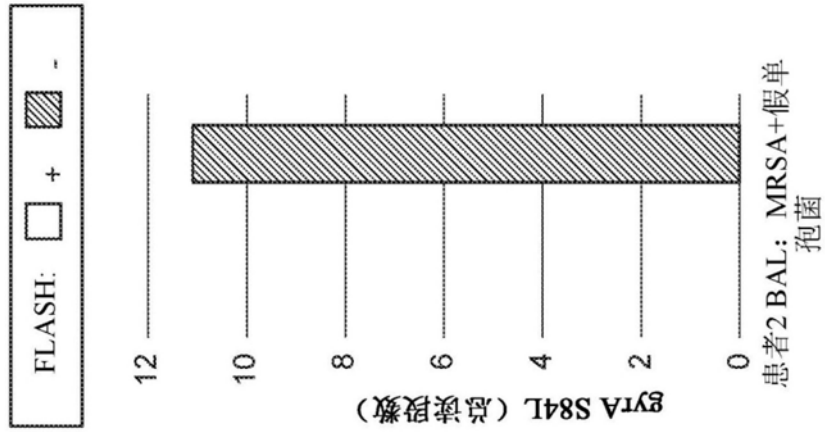


图3B

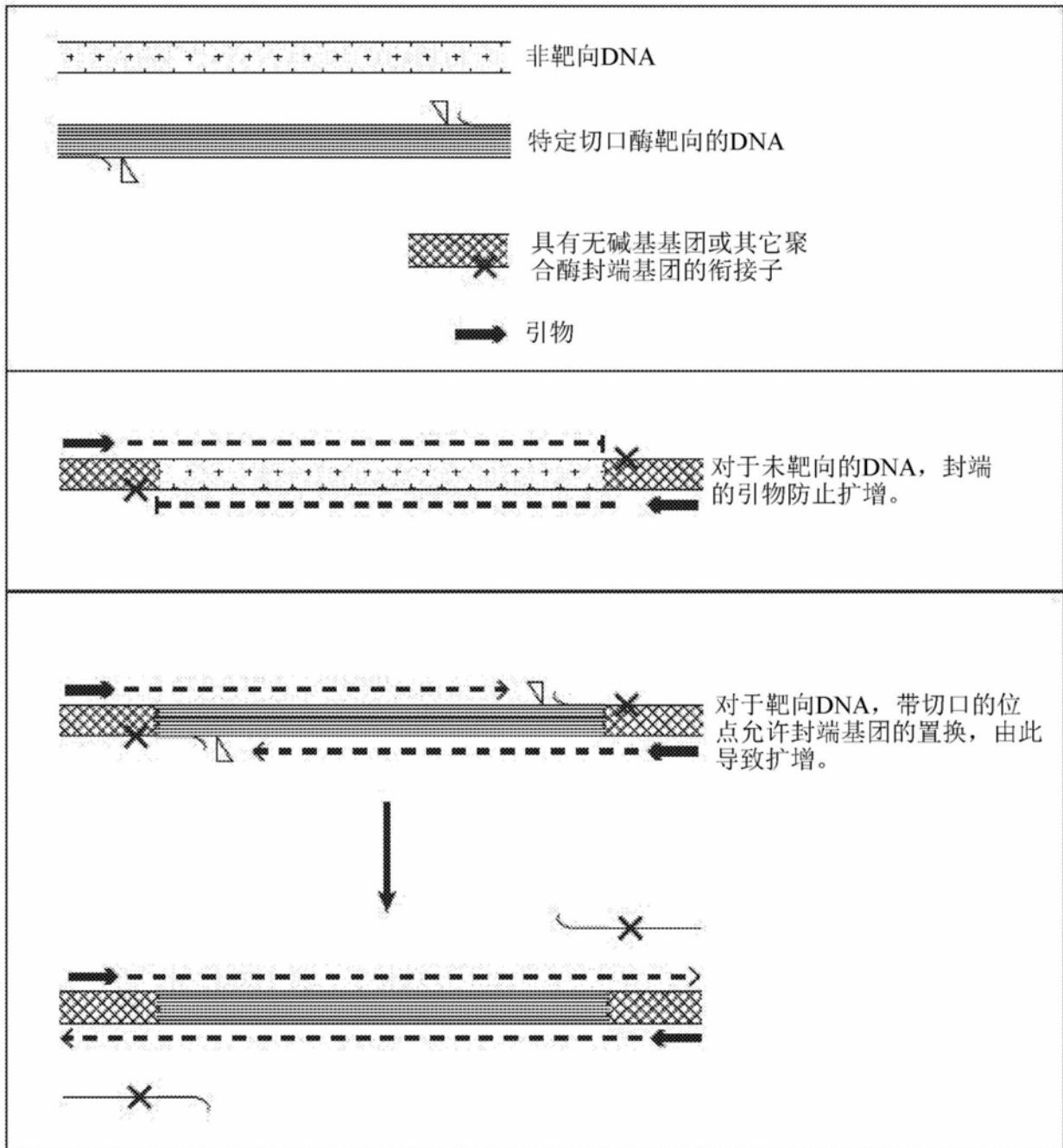


图4