



(12) 发明专利

(10) 授权公告号 CN 107075546 B

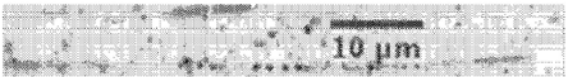
(45) 授权公告日 2021.08.31

(21) 申请号 201580056648.3	(73) 专利权人 哈佛学院董事及会员团体
(22) 申请日 2015.08.19	地址 美国马萨诸塞州
(65) 同一申请的已公布的文献号 申请公布号 CN 107075546 A	(72) 发明人 G·M·丘奇 F·魏格纳尔特 K·U·米尔
(43) 申请公布日 2017.08.18	(74) 专利代理机构 上海专利商标事务所有限公 司 31100
(30) 优先权数据 62/039,341 2014.08.19 US	代理人 陈扬扬 陶启长
(85) PCT国际申请进入国家阶段日 2017.04.18	(51) Int.Cl. C12Q 1/6832 (2018.01) C12Q 1/6874 (2018.01)
(86) PCT国际申请的申请数据 PCT/US2015/045805 2015.08.19	审查员 邵文博
(87) PCT国际申请的公布数据 W02016/028843 EN 2016.02.25	

权利要求书1页 说明书76页
序列表48页 附图21页

(54) 发明名称
用于对核酸探测并作图的RNA-引导的系统

(57) 摘要
使用引导RNA和Cas9蛋白提供对靶核酸的检测、探测、作图和定向测序的方法。提供了检测引导RNA/Cas9复合物与靶核酸的结合的方法,其中引导RNA包含可与探针杂交的3'尾序列。提供了检测引导RNA/Cas9复合物与靶核酸结合的方法,其中物理检测该复合物。



1. 一种检测固定在表面上的靶核酸序列的方法, 包括

(a) 使所述固定在表面上的靶核酸序列与Cas9蛋白和引导核糖核酸(RNA)分子接触, 所述分子包含(i)与所述靶核酸序列互补的引导RNA序列和(ii)条码序列, 使得所述引导RNA分子和所述Cas9蛋白共定位至所述靶核酸序列以形成复合物;

(b) 使包含所述引导RNA分子的复合物与探针分子接触, 所述探针分子包含(i)与所述条码序列互补的探针序列和(ii)可检测标记物, 以将所述探针序列与所述引导RNA序列杂交; 并且

(c) 检测所述可检测标记物以鉴定与所述靶核酸序列杂交的引导RNA序列, 从而检测所述靶核酸序列。

2. 如权利要求1所述的方法, 其中所述探针分子包含多个可检测标记物。

3. 如权利要求1所述的方法, 其中所述靶核酸是双链基因组DNA。

4. 如权利要求1所述的方法, 其中所述靶核酸是染色体脱氧核糖核酸(DNA)。

5. 如权利要求1所述的方法, 其中所述Cas9蛋白是核酸酶无效Cas9。

6. 如权利要求1所述的方法, 其中所述可检测标记物是光学标记物。

7. 如权利要求1所述的方法, 其中所述可检测标记物是发光标记物。

8. 如权利要求1所述的方法, 其中所述可检测标记物是荧光标记物。

9. 如权利要求1所述的方法, 其中所述可检测标记物由二级检测试剂检测。

10. 如权利要求9所述的方法, 其中所述可检测标记物选自下组: 生物素、地高辛、二硝基苯基、荧光素及其组合, 并且所述二级检测试剂是对所述可检测标记物有特异性的金或银包被的抗体。

11. 如权利要求1所述的方法, 其中(b)包括将多个探针序列与所述引导RNA序列的条码序列杂交, 其中所述条码序列的多个子序列与所述多个探针序列杂交。

12. 如权利要求1所述的方法, 其中(b)包括将多个探针序列与所述引导RNA分子的条码序列杂交, 其中所述多个探针序列组装以形成线性核酸结构或高分支核酸结构。

13. 如权利要求1所述的方法, 其中所述探针序列是环状探针序列。

14. 如权利要求13所述的方法, 其中所述环状探针序列经扩增。

15. 如权利要求14所述的方法, 其中所述环状探针序列通过滚环扩增来扩增。

16. 如权利要求1所述的方法, 其中所述探针序列包含第一序列和第二序列, 其中所述第一序列与所述条码序列杂交并且所述第二序列不与所述条码序列杂交。

17. 如权利要求16所述的方法, 其中所述第二序列是环状序列的扩增产物。

18. 如权利要求1所述的方法, 其中所述靶核酸序列在生物样品内。

19. 如权利要求18所述的方法, 其中所述生物样品是细胞或组织。

20. 如权利要求18所述的方法, 其中所述生物样品是固定的样品。

21. 如权利要求1所述的方法, 其中所述Cas9蛋白不切割所述靶核酸序列。

用于对核酸探测并作图的RNA-引导的系统

[0001] 相关申请信息

[0002] 本申请要求于2014年8月19日提交的美国临时专利申请62/039,341号的优先权，其通过引用其全文纳入本文用于所有目的。

背景技术

[0003] 细菌和古细菌CRISPR-Cas系统依赖于与Cas蛋白负荷的短引导RNA以引导入侵外来核酸内存在的互补序列的直接降解。参见Deltcheva, E.等, 通过反式编码小RNA和宿主因子RNA酶III的CRISPR RNA突变 (CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III). *Nature* 471, 602-607 (2011); Gasiunas, G, Barrangou, R., Horvath, P. 和 Siksnys, V. Cas9-crRNA核蛋白复合物介导细菌的获得性免疫的特异性DNA切割 (Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria). *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579-2586 (2012); Jinek, M.等, 获得性细菌免疫中可编程的双RNA-引导的DNA内切核酸酶 (A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity). *Science* 337, 816-821 (2012); Saprunauskas, R.等, 嗜热链球菌CRISPR/Cas系统提供大肠杆菌中的免疫 (The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*). *Nucleic acids research* 39, 9275-9282 (2011); 和 Bhaya, D., Davison, M. 和 Barrangou, R. 细菌和古细菌中的CRISPR-Cas系统: 获得性防御和调节的通用小RNA (CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation). *Annual review of genetics* 45, 273-297 (2011)。最近的酿脓链球菌 (*S. pyogenes*) II型CRISPR系统的体外重建证明与正常反式编码tracrRNA融合的crRNA (“CRISPR RNA”) (“反式-作用CRISPR RNA”) 足够引导Cas9蛋白序列特异性切割匹配cdRNA的靶DNA序列。表达与靶位点同源的gRNA导致Cas9招募以及靶DNA降解。参见H. Deveau等, 对嗜热链球菌中CRISPR编码的抗性的噬菌体响应 (Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*). *Journal of Bacteriology* 190, 1390 (2008年2月)。已知CRISPR/Cas9系统的各种用途。参见W02014/099744、W02013176772、US 8,697,359和Sternberg等, *Nature*, 第507卷, 第62-67页 (2014)。

[0004] 发明概述

[0005] 本发明的方面涉及检测靶核酸序列的方法, 包括使靶核酸序列与引导RNA序列接触, 所述引导RNA序列具有与靶核酸序列互补的部分和Cas9蛋白, 其中引导RNA和Cas9蛋白共定位至靶核酸序列以形成复合物, 并且其中检测该复合物从而检测靶核酸序列。根据一个方面, 离体, 例如, 体外进行该方法, 如在容器内或在基材上。根据一个方面, 引导RNA和Cas9蛋白经制备并分离以用作本发明的体外方法中的试剂。本发明的方法的方面包括探测, 如对核酸, 如DNA的分析型探测或制备型探测、检测、标记、作图、和测序。例如, 本发明涉及探测DNA的方法, 如在单分子水平上, 用于鉴定存在DNA, 探测DNA, 用于亲和纯化该DNA、对

DNA作图,以沿着DNA标注具体重要区域,或产生测序起始位点。

[0006] 根据本文所述的方法,形成包含引导RNA、DNA结合蛋白如Cas9蛋白、和双链DNA靶序列的复合物。根据某些方面,本发明范围内的DNA结合蛋白包括与引导RNA形成复合物的蛋白质,并且引导RNA将复合物引导到双链DNA序列,复合物在该序列处结合至DNA序列。本发明的这一方面可被称为RNA和DNA结合蛋白共定位至双链DNA或与双链DNA共定位。通过这种方式,可施用DNA结合蛋白-引导的RNA复合物来在具体靶DNA序列处形成可检测的复合物,从而检测靶DNA序列的存在。根据某些方面,由于存在可检测标记物,可检测复合物。根据某些方面,复合物可直接标记或间接标记。根据某些方面,可检测标记物可存在于引导RNA、Cas9蛋白或复合物上。

[0007] 根据某些方面,引导RNA的共定位因子可能不是DNA-结合蛋白。可使用试剂来用引导RNA在靶核酸序列处共定位。根据某些方面,引导RNA不需要存在可用于本发明的某些方面的DNA结合蛋白。可能不存在DNA结合蛋白。例如,引导RNA本身可结合至靶核酸序列,并且引导RNA可具有与之接合的标记物或其他功能部分以将标记物或其他功能部分定位在靶核酸序列处或其附近。

[0008] 根据某些方面,可通过检测没有可检测标记物的复合物的结构来检测复合物。复合物的物理结构被探测,与观察荧光或其他视觉或光谱上可检测的部分相反。根据某些方面,可通过检测没有可检测标记物的复合物的物理化学性质,如静电荷来检测复合物。这类方法包括使用纳米孔检测方法、电子显微术、光学显微术、扫描探针显微术、原子力显微术、悬臂检测方法、石英晶体检测方法、效应晶体管检测方法检测复合物,其全部都是本领域技术人员已知的。本领域技术人员将易于想象其他能够检测基于本发明的复合物的结构的方法。

[0009] 根据某些方面,CRISPR Cas9系统内容中的术语“引导RNA”是本领域技术人员已知的并且包括部分,如20个核苷酸部分,其与靶核酸互补。设计引导RNA的方法是本领域技术人员众所周知的。本文所述的方法包括使靶核酸序列与多种引导RNA序列互补,这些序列各自具有与靶核酸序列互补的部分。本文所述的方法包括使队中靶核酸序列与多种相应引导RNA序列互补,其各自具有与相应靶核酸序列互补的部分。

[0010] 根据某些方面,本发明的引导RNA包括与靶核酸互补的部分和与探针序列或可检测标记物互补或可与之互补或者与之结合的3'-尾部分或序列。根据一个方面,3'尾部分提供了特异性功能。3'尾部分可以是模块或含有多个元件,针对相同或针对多种功能。根据一个方面,3'尾部分可与一个或多个或多重探针序列或可检测标记物互补或者与之结合(例如,通过适体机制)。各探针序列或可检测标记物可用起到不同作用(例如,一种序列的作用可以是结合至CY3标记的寡核苷酸并且第二种序列的作用可以是结合至Cy5标记的寡核苷酸)。例如,尾序列可用于将功能性蛋白定位至靶核酸序列。例如,尾序列可结合被引导RNA取代的靶双链的部分。

[0011] 根据一个方面,探针序列包括可检测标记物,并且该探针序列结合至3'尾序列。根据一个方面,探针序列包括多个可检测标记物,并且该探针序列结合至3'尾序列。根据一个方面,探针序列包括可检测标记物,并且该探针序列结合至3'尾序列,并且其中探针序列经扩增。根据一个方面,探针序列结合至3'尾序列,并且其中探针序列经扩增。根据一个方面,引导RNA包括3'尾序列作为与探针或可检测标记物的结合对。根据一个方面,尾序列可在结

合至模板序列时起到引物的作用。尾序列然后延伸以纳入一个或多个可检测标记物,如荧光核苷酸,或一个或多个结合部分,如生物素或地高辛(dig)标记的核苷酸,一个或多个标记物可直接或间接与这些结合部分结合。根据一个方面,滚环扩增可与尾引物序列和滚环扩增模板联用以产生具有多个可检测部分或可检测部分可与之接合的结合部分的滚环多联体产物。通过这种方式,可使用滚环扩增来扩增信号强度。滚环扩增方法是本领域技术人员已知的并且包括Drmanac等,使用对自组装DNA纳米阵列的解链的碱基读数的人基因组测序(Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays),Science,第327卷,第78-81页(2009)。

[0012] 根据一个方面,靶核酸是双链核酸。根据一个方面,靶核酸是双链基因组DNA。根据一个方面,靶核酸是染色体DNA。

[0013] 根据某些方面,引导RNA包括种子区域序列。根据某些方面,引导RNA在引导RNA的非种子区域处包括简并位置或序列或通用碱基。

[0014] 根据一个方面,靶核酸在基材,如平面基材或孔或通道上延长,即,靶核酸在孔或通道内延长。

[0015] 根据一个方面,Cas9蛋白是野生型Cas9、Cas9切口酶或无核酸酶Cas9,如本领域技术人员已知。分离野生型Cas9的方法是本领域技术人员已知的。制备Cas9切口酶的方法是本领域技术人员已知的。制备无核酸酶Cas9的方法是本领域技术人员已知的。

[0016] 根据一个方面,可检测标记物直接或间接结合至Cas9蛋白。根据一个方面,可检测标记物直接或间接结合至引导RNA。根据一个方面,可检测标记物是引导RNA的部分(例如,荧光标记的核苷酸在引导RNA的制备期间纳入(例如,通过体外转录))。根据一个方面,可检测标记物直接或间接结合至该复合物。

[0017] 根据一个方面,提供方法从而通过测序方法来确定靶核酸的序列。根据一个方面,Cas9蛋白是Cas9切口酶,其使靶核酸的链产生切口并且其中从切口开始引物延伸或链延伸,互补链用作模板,从而对靶核酸,如靶核酸的链之一进行测序。相对于其他切口方法,如使用切口内切核酸酶和DNA酶1的优势在于可通过使用引导RNA对产生切口的位置进行编程,并且可靶向多个特定位置。可使用计算机执行方法和软件来鉴定用于合成的感兴趣基因组的部分,对制备gRNA所需的DNA模板进行排序,体外转录引导RNA并且然后在所需位置处执行切口以进行靶向测序。通过沿着模板的引物延伸进行测序的方法是本领域技术人员已知的。使用切口作为引物是本领域技术人员已知的。本发明的这一特征也可用于在延伸产物中包括可检测标记物,从而检测靶核酸,如代替或另外从靶核酸获得序列信息。因此,Cas9蛋白是Cas9切口酶,其在靶核酸的链上形成切口,并且其中引物延伸从切口开始以包括可检测标记物,互补链用作模板,从而检测靶核酸。根据这一方面,一旦标记物纳入延伸产物,gRNA/Cas9不需要与靶DNA保持在一起,因为检测到已经纳入延伸产物的标记物。

[0018] 根据一个方面,分离的引导RNA和分离的Cas9蛋白在合适条件下结合并且然后在反应或复合物形成介质中与靶核酸体外接触。根据一个方面,靶核酸在样品,如核酸样品内。核酸样品可包括多个核酸并且可被称为核酸的复杂混合物。根据某些方面,提供了用于在核酸的复杂混合物内用对靶核酸有特异性的引导RNA鉴定靶核酸的方法。根据一个方面,提供了在核酸的复杂混合物内用对一种或多种或数种靶核酸有特异性的引导RNA鉴定一种或多种或数种靶核酸的方法。在这一方面,提供了用于检测多种靶核酸的多重方法。多种靶

核酸中的各靶核酸可被相应的引导RNA/Cas 9蛋白复合物结合,并且因此能够如本文所述或本领域已知的那样被检测或测序。

[0019] 根据某些方面,提供了用于在核酸的复杂混合物内用对靶核酸有特异性的引导RNA对靶核酸进行亲和纯化的方法。根据一个方面,提供了在核酸的复杂混合物内用对一种或多种或数种靶核酸有特异性的引导RNA亲和纯化一种或多种或数种靶核酸。在这一方面,提供了用于亲和纯化多种靶核酸的多重方法。多种靶核酸中的各靶核酸可被相应引导RNA/Cas 9蛋白复合物结合。根据这些方面,可使用采用本领域技术人员已知的结合对的亲和系统。通过消耗复杂混合物中的其他核酸来纯靶核酸。通过从复杂混合物中消耗靶核酸来纯靶核酸。消耗可通过待消耗的靶标的亲和捕获进行。或者,消耗可通过待消耗的靶标的切割进行。在一些情况中,可能需要从复杂混合物中消耗混合物内具有高丰度的核酸以分析较低丰度的靶标。例如,可能需要从样品中消耗相应的DNA或其他高浓度核酸。

[0020] 根据某些方面,提供了用于在核酸的复杂混合物内用对针对消耗靶向的核酸有特异性的引导RNA消耗靶核酸的方法。根据一个方面,提供了在核酸的复杂混合物内用对一种或多种或数种靶核酸有特异性的引导RNA一种或多种或数种靶核酸。在这一方面,提供了用于消耗多种靶核酸的多重方法。多种针对消耗靶向的核酸中的各靶核酸可被相应引导RNA/Cas 9蛋白复合物结合。

[0021] 根据某些方面,靶核酸在溶液样品内。根据某些方面,靶核酸在基材上。根据某些方面,靶核酸结合至基材。根据某些方面,靶核酸在细胞内。根据一个方面,细胞是真核细胞。根据一个方面,细胞是酵母细胞、植物细胞或动物细胞。根据一个方面,细胞是哺乳动物细胞。在某些实施方式中,哺乳动物细胞是活细胞并且引导RNA或DNA结合蛋白,如Cas9或其他DNA结合蛋白通过电穿孔、载体介导的递送(例如,脂质转染)、微注射和本领域技术人员已知的其他方法。在某些实施方式中,哺乳动物细胞是固定的细胞,其在含有引导RNA和DNA结合蛋白,如Cas9或其他待递送的DNA结合蛋白的溶液中沉浸。通过相似的方式,使用gRNA和DNA结合蛋白在靶核酸序列处共定位可在中期染色体分散上进行。

[0022] 根据一个方面,引导RNA在约10至约500个核苷酸之间。根据一个方面,引导RNA在约20至约100个核苷酸之间。根据一个方面,引导RNA是tracrRNA-crRNA融合体。根据一个方面,tracrRNA和crRNA是分开的物质并且被融合。

[0023] 根据一个方面,DNA是基因组DNA、线粒体DNA、病毒DNA、或外源性DNA。

[0024] 根据一个方面,提供了用于探测包含两种或更多种由其DNA含量表征的不同多核苷酸物质或细胞的混合物的方法,通过选择与样品中的一种或多种多核苷酸物质互补的一种或多种序列,产生包括互补序列的一种或多种gRNA,合并一种或多种gRNA与Cas9,使样品接触gRNA和Cas9,检测gRNA/Cas9与样品中一种或多种多核苷酸物质的结合;基于检测确定样品的细胞或多核苷酸组分的种类。根据一个方面,gRNA和Cas9体外产生或存在于体外。根据一个方面,gRNA和Cas9体外合并。根据一个方面,gRNA体外产生或存在于体外,而Cas9体内产生。根据另一个方面,样品包含多种不同的多核苷酸物质,如许多可存在于数十或数百或数千或数万种不同多核苷酸物质的复杂混合物中。

[0025] 其他应用包括评价靶生物体的种类的方法,包括使用引导RNA和Cas9;评价靶生物体的状态的方法,包括使用引导RNA和Cas9;对DNA分子作图的方法,包括解析结合在DNA分子上的多种Cas9和引导RNA复合物;或在DNA分子中解析等位基因变体的方法,包括使用多

种Cas9和引导RNA复合物和多种探针。这些具体应用各自基于使用本文所述的gRNA/Cas9系统在靶DNA位点处形成复合物并检测gRNA/Cas9复合物来探测DNA的方法。

[0026] DNA分子可以是染色体或染色体外的。Cas9核酸内切酶可以是活性的,或无活性的或部分无活性的。Cas9可与荧光蛋白(如GFP、荧光素酶等)和/或一个或多个亲和标签融合。亲和标签可被一个或多个荧光探针识别。亲和标签可被一个或多个标签识别,其向Cas9添加可测量的属性(例如,电荷或形状)。Cas9可含有一个或多个正交氨基酸。正交氨基酸可向其他分子提供亲和性,如探针、标签、接头。

[0027] 可直接使用一个或多个荧光探针来探测引导RNA。引导RNA可被一个或多个向Cas9添加可测量的属性(例如,电荷或形状)的标签直接探测。引导RNA可含有一个或多个修饰的碱基。修饰的碱基可向其他分子提供亲和性,如探针、标签、接头。

[0028] 生物体可以是原核或真核,单细胞或多细胞。从生物体提取DNA。DNA可以是其天然形式,或者DNA可在表面上或装置中拉伸。DNA可通过通道或纳米孔移位。生物体经固定并可渗透体外合成的Cas9和引导RNA复合物。

[0029] 在某些实施方式中,Cas9和引导RNA在用于靶向DNA之前复合。引导RNA与靶DNA互补。通过测量荧光信号来检测与DNA结合的复合物。通过测量当前信号同时通过纳米孔传感器或靠近纳米孔或纳米间隙传感器移位来检测与DNA结合的复合物。

[0030] 可一次检测DNA上的一个或多个复合物。DNA上的任意两个复合物之间的分辨率可低至1纳米或5纳米或10纳米或高至1000毫米(和之间的任何数字)。检测到特定复合物表示存在特定等位基因。可产生与DNA结合的复合物的图案并用于提供DNA分子的图。可产生与DNA结合的复合物的图案并用于提供生物体的种类和/或状态。根据一个方面,可向活或无活性(即,死)细胞提供引导RNA或Cas9或引导RNA/Cas9复合物。

[0031] 提供了使用Cas9和引导RNA复合物来在DNA分子上产生序列特异性起始位点的方法。可使用聚合酶或连接酶来进行单分子测序。起始位点可靠近基因组变体、重复序列、高变区。

[0032] 提供了使用Cas9和引导RNA复合物来拉下DNA分子的方法,即亲和纯化。能允许拉下的亲和标签结合至Cas9和引导RNA复合物。亲和标签在复合物结合至DNA之前或之后结合至复合物。可从集合中提取结合至一个或多个Cas9和引导RNA复合物的一个特异性或多个特异性靶DNA分子。提取的靶DNA分子可经过测序,如深度测序。

[0033] 在某些实施方式中,在某些条件下,使用引导RNA而不使用Cas9(或其他DNA结合蛋白),并且当针对某些类型的序列靶向时,引导RNA足够形成与DNA靶标的稳定或瞬时接合。

[0034] 在某些实施方式中,gRNA/Cas9共定位复合物提供双链切割,但是gRNA/Cas9共定位复合物保持与靶核酸结合。可使用条件如添加7M尿素开始打断复合物来从靶核酸去除这种gRNA/Cas9共定位(参见Sternberg等,Nature 507:62(2014),通过引用全文纳入本文)。

[0035] 在一些实施方式中,gRNA和Cas9的复合物在结合至靶核酸序列,如DNA之前形成。在一些实施方式中,复合物在Cas9首先与本核酸相互作用之后形成,即Cas9与靶核酸相互作用,然后与引导RNA形成共定位复合物。

[0036] 本发明的某些实施方式的其他特征和优势将在权利要求中以及以下附图和实施方式的说明下更为显而易见。

附图说明

[0037] 结合附图,通过以下示例性实施方式的详述能够更清楚地理解本发明的上述和其他特征和其他优点,其中:

[0038] 图1表示用标记的gRNA/Cas9探测的固定的小鼠细胞的图。

[0039] 图2表示按照图1所示的Cas9探测方案用标记的寡核苷酸探测的固定的小鼠细胞的图。

[0040] 图3表示gRNA/Cas9切割的琼脂糖凝胶和凝胶迁移试验。

[0041] 图4表示用于探测gRNA尾的各种示意图。

[0042] 图5表示侧流试验的图。

[0043] 图6表示用标记的gRNA/Cas9探测的拉伸的DNA的图。

[0044] 图7表示使用gRNA/Cas9探测鉴定基因组重排的图。

[0045] 图8表示使用与gRNA/Cas9接合的折纸(origami)条码鉴定基因组区域的图。

[0046] 图9表示引发从切口位点开始测序的图。

[0047] 图10表示使用CHOPCHOP鉴定gRNA/Cas9靶位点以鉴定Her2的输出图。

[0048] 图11表示PCR组装策略以从寡核苷酸集合制备gRNA模板。

[0049] 图12表示使用gRNA/Cas9探测鉴定基因组融合的图。

[0050] 发明详述

[0051] 本发明的实施方式基于使用DNA结合蛋白和引导RNA在靶核酸处共定位或复合,然后通过检测与复合物结合或接合的可检测部分或通过物理探测复合物本身来检测靶核酸。本领域技术人员易于知道这类DNA结合蛋白(包括RNA-引导的DNA结合蛋白)结合DNA用于各种目的。这种DNA结合蛋白可以是天然产生的。本发明范围内包括的DNA结合蛋白包括由RNA引导的那些,在本文中称为引导RNA。根据该方面,引导RNA和RNA引导的DNA结合蛋白在DNA处形成共定位复合物。根据某些方面,DNA结合蛋白可以是无核酸酶DNA结合蛋白。根据这一方面,无核酸酶DNA结合蛋白可产生自对具有核酸酶活性的DNA结合蛋白的改变或修饰。这类具有核酸酶活性的DNA结合蛋白为本领域技术人员所知,并且包括具有核酸酶活性的天然产生的DNA结合蛋白,例如在II型CRISPR系统中存在的Cas9蛋白。这种Cas9蛋白和II型CRISPR系统在本领域中充分记载。参见Makarova等,Nature Reviews, Microbiology, 第9卷,2011年6月,第467-477页,包括所有补充信息,通过引用全文纳入本文。

[0052] 具有核酸酶活性的示例性DNA结合蛋白的功能是切开或切割双链DNA。这种核酸酶活性可来自具有显示核酸酶活性的一个或多个多肽序列的DNA结合蛋白。这类示例性DNA结合蛋白可具有2个分开的核酸酶结构域,各结构域负责切割或切开双链DNA的特定链。本领域技术人员已知的具有核酸酶活性的示例性多肽序列包括McrA-HNH核酸酶相关结构域和RuvC-样核酸酶结构域。因此,示例性的DNA结合蛋白是天然含有一个或多个McrA-HNH核酸酶相关结构域和RuvC-样核酸酶结构域的那些。根据某些方面,DNA结合蛋白经改变或修饰以使得核酸酶活性灭活。这类改变或修饰包括改变一个或多个氨基酸以使核酸酶活性或核酸酶结构域灭活。这类修饰包括去除显示出核酸酶活性的一个或多个多肽序列,即核酸酶结构域,使得DNA结合蛋白上没有显示出核酸酶活性的一个或多个多肽序列,即核酸酶结构域。基于本发明,灭活核酸酶活性的其他修饰将对于本领域技术人员而言是显而易见的。因此,无核酸酶的DNA结合蛋白包括经修饰以灭活核酸酶活性或去除一个或多个多肽序列以

灭活核酸酶活性的多肽序列。无核酸酶DNA结合蛋白保留了结合至DNA的能力,即使核酸酶活性已经被灭活。因此,DNA结合蛋白包括DNA结合所需的一个或多个多肽序列,但是可缺少显示核酸酶活性的一个或多个或全部核酸酶序列。因此,DNA结合蛋白包括DNA结合所需的一个或多个多肽序列,但是可具有显示灭活的核酸酶活性的一个或多个或全部核酸酶序列。

[0053] 根据一个方面,具有2个或更多个核酸酶结构域的DNA结合蛋白可经修饰或改变以使除了核酸酶结构域之一以外的全部核酸酶结构域灭活。这种经修饰或改变的DNA结合蛋白被称为DNA结合蛋白切口酶,该DNA结合蛋白仅切割或切开双链DNA的一条链。当由RNA引导至DNA时,DNA结合蛋白切口酶被称为RNA引导的DNA结合蛋白切口酶。因此,可用的Cas9蛋白可以是野生型Cas9、Cas9切口酶或无核酸酶Cas9及其同源物和直向同源物。参见Jinek等,Science 337,816-821 (2012),通过引用全文纳入本文。

[0054] 在酿脓链球菌(*S. pyogenes*)中,Cas9通过由蛋白质中的以下2个催化结构域介导的过程在原型间隔子-相邻基序(PAM)上游3bp处生成钝末端双链断裂:切割DNA的互补链的HNH结构域和切割非互补链的RuvC-样结构域。参见Jinek等,Science 337,816-821 (2012),通过引用全文纳入本文。已知Cas9蛋白存在于许多II型CRISPR系统,包括Makarova等,Nature Reviews, Microbiology,卷9,2011年6月,第467-477页的补充信息中所鉴定的以下系统:海沼甲烷球菌(*Methanococcus maripaludis*) C7;白喉棒杆菌(*Corynebacterium diphtheriae*);高效棒杆菌(*Corynebacterium efficiens*) YS-314;谷氨酸棒杆菌(*Corynebacterium glutamicum*) ATCC 13032 Kitasato;谷氨酸棒杆菌(*Corynebacterium glutamicum*) ATCC 13032 Bielefeld;谷氨酸棒杆菌(*Corynebacterium glutamicum*) R;果聚糖棒杆菌(*Corynebacterium kroppenstedtii*) DSM 44385;脓肿分枝杆菌(*Mycobacterium abscessus*) ATCC 19977;皮诺卡氏菌(*Nocardia farcinica*) IFM10152;红串红球菌(*Rhodococcus erythropolis*) PR4;红球菌(*Rhodococcus jostii*) RHA1;混浊红球菌(*Rhodococcus opacus*) B4uid36573;解纤维热酸菌(*Acidothermus cellulolyticus*) 11B;氯酚节杆菌(*Arthrobacter chlorophenolicus*) A6;黄克里布所菌(*Kribbella fiavida*) DSM17836 uid43465;弯曲热单孢菌(*Thermomonospora curvata*) DSM 43183;齿双歧杆菌(*Bifidobacterium dentium*) Bd1;长双歧杆菌(*Bifidobacterium longum*) DJ010A;还原天芥菜碱斯奈克氏菌(*Slackia heliotrinireducens*) DSM 20476;海皮尔氏菌(*Persephonella marina*) EX H1;脆弱拟杆菌(*Bacteroides fragilis*) NCTC 9434;黄褐二氧化碳嗜纤维菌(*Capnocytophaga ochracea*) DSM 7271;嗜冷黄杆菌(*Flavobacterium psychrophilum*) JIP02 86;黏液阿克曼菌(*Akkermansia muciniphila*) ATCC BAA 835;卡氏玫瑰弯菌(*Roseiflexus castenholzii*) DSM 13941;玫瑰弯菌(*Roseiflexus*) RS1;集胞蓝细菌(*Synechocystis*) PCC6803;小迷踪菌(*Elusimicrobium minutum*) Pei191;未培白蚁1组细菌种系Rs D17;产琥珀酸丝状杆菌(*Fibrobacter succinogenes*) S85;蜡状芽孢杆菌(*Bacillus cereus*) ATCC 10987;无害利斯特氏菌(*Listeria innocua*);干酪乳杆菌(*Lactobacillus casei*);鼠李糖乳杆菌(*Lactobacillus rhamnosus*) GG;唾液乳杆菌(*Lactobacillus salivarius*) UCC118;无乳链球菌(*Streptococcus agalactiae*) A909;无乳链球菌(*Streptococcus agalactiae*) NEM316;无乳链球菌(*Streptococcus agalactiae*) 2603;停乳链球菌(*Streptococcus dysgalactiae equisimilis*) GGS 124;马链球菌

(*Streptococcus equi zooepidemicus*) MGCS10565; 鸡链球菌 (*Streptococcus gallolyticus*) UCN34uid46061; 格氏链球菌 (*Streptococcus gordonii*) Challis subst CH1; 变异链球菌 (*Streptococcus mutans*) NN2025uid46353; 变异链球菌 (*Streptococcus mutans*); 酿脓链球菌 (*Streptococcus pyogenes*) M1GAS; 酿脓链球菌 (*Streptococcus pyogenes*) MGAS5005; 酿脓链球菌 (*Streptococcus pyogenes*) MGAS2096; 酿脓链球菌 (*Streptococcus pyogenes*) MGAS9429; 酿脓链球菌 (*Streptococcus pyogenes*) MGAS10270; 酿脓链球菌 (*Streptococcus pyogenes*) MGAS6180; 酿脓链球菌 (*Streptococcus pyogenes*) MGAS315; 酿脓链球菌 (*Streptococcus pyogenes*) SSI-1; 酿脓链球菌 (*Streptococcus pyogenes*) MGAS10750; 酿脓链球菌 (*Streptococcus pyogenes*) NZ131; 嗜热链球菌 (*Streptococcus thermophiles*) CNRZ1066; 嗜热链球菌 (*Streptococcus thermophiles*) LMD-9; 嗜热链球菌 (*Streptococcus thermophiles*) LMG 18311; 肉毒梭菌 (*Clostridium botulinum*) A3 Loch Maree; 肉毒梭菌 (*Clostridium botulinum*) B Eklund 17B; 肉毒梭菌 (*Clostridium botulinum*) Ba4657; 肉毒梭菌 (*Clostridium botulinum*) F Langeland; 解纤维梭菌 (*Clostridium cellulolyticum*) H10; 微小微单胞菌 (*Fingoldia magna*) ATCC 29328; 直肠假杆菌 (*Eubacterium rectale*) ATCC 33656; 鸡败血枝原体 (*Mycoplasma gallisepticum*); 运动枝原体 (*Mycoplasma mobile*) 163K; 穿透枝原体 (*Mycoplasma penetrans*); 关节液枝原体 (*Mycoplasma synoviae*) 53; 念珠状链杆菌 (*Streptobacillus moniliformis*) DSM 12112; 慢生根瘤菌 (*Bradyrhizobium*) BT Ail; 汉堡硝化杆菌 (*Nitrobacter hamburgensis*) X14; 血色红假单胞菌 (*Rhodopseudomonas palustris*) BisB18; 血色红假单胞菌 (*Rhodopseudomonas palustris*) BisB5; 食清洁剂细小棒菌 (*Parvibaculum lavamentivorans*) DS-1; 恒雄芝氏沟鞭玫瑰杆属 (*Dinoroseobacter shibae*) DFL 12; 固氮葡萄糖酸醋酸杆菌 (*Gluconacetobacter diazotrophicus*) Pal 5FAPERJ; 固氮葡萄糖酸醋酸杆菌 (*Gluconacetobacter diazotrophicus*) Pal 5 JGI; 固氮螺菌 (*Azospirillum*) B510 uid46085; 深红红螺菌 (*Rhodospirillum rubrum*) ATCC 11170; 絮凝剂产生菌 (*Diaphorobacter*) TPSY uid29975; 蚯蚓虫肾杆菌 (*Verminephrobacter eiseniae*) EF01-2; 脑膜炎奈瑟球菌 (*Neisseria meningitides*) 053442; 脑膜炎奈瑟球菌 (*Neisseria meningitides*) alpha14; 脑膜炎奈瑟球菌 (*Neisseria meningitides*) Z2491; 需盐脱硫弧菌 (*Desulfovibrio salexigens*) DSM 2638; 空肠弯曲杆菌 (*Campylobacter jejuni doylei*) 26997; 空肠弯曲杆菌 (*Campylobacter jejuni*) 81116; 空肠弯曲杆菌 (*Campylobacter jejuni*); 红嘴鸥弯曲杆菌 (*Campylobacter lari*) RM2100; 肝螺杆菌 (*Helicobacter hepaticus*); 产琥珀酸沃林氏菌 (*Wolinella succinogenes*); 甲苯单胞 (*Tolomonas auensis*) DSM 9187; 交替假单胞菌 (*Pseudoalteromonas atlantica*) T6c; 希瓦氏菌 (*Shewanella pealeana*) ATCC700345; 嗜肺军团菌 (*Legionella pneumophila*) Paris; 产琥珀酸放线杆菌 (*Actinobacillus succinogenes*) 130Z; 多杀巴斯德菌 (*Pasteurella multocida*); 新凶手弗朗西丝氏菌 (*Francisella tularensis novicida*) U112; 土拉热弗朗西丝氏菌全北区亚种 (*Francisella tularensis holarctica*); 土拉热弗朗西丝氏菌 (*Francisella tularensis*) FSC 198; 土拉热弗朗西丝氏菌 (*Francisella tularensis*); 土拉热弗朗西丝氏菌 (*Francisella tularensis*) WY96-3418; 和齿垢密螺旋体 (*Treponema denticola*) ATCC 35405。因此, 本发明的方面设计在II型CRISPR系统中存在的Cas9蛋白, 其

已被赋予无核酸酶,或已经被赋予切口酶,如本文所述。

[0055] 在文献中,本领域技术人员可将Cas9蛋白称为Csn1。酿脓链球菌Cas9蛋白序列如下所示。参见Deltcheva等,Nature 471,602-607 (2011),通过引用全文纳入本文。

[0056] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
[0057] ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFG
[0058] NIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSD
[0059] VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGN
[0060] LIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAI
[0061] LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLEPKYKEIFFDQSKNGYA
[0062] GYIDGGASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELH
[0063] AILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
[0064] VVDKGASASFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAFL
[0065] SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI
[0066] IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
[0067] RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSL
[0068] HEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
[0069] MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDH
[0070] IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
[0071] TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
[0072] KLVSDFRKDFQFYKVVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRK
[0073] MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
[0074] ATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWD PKKYGGFDSPTVA
[0075] YSVLVVAKVEKGSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPK
[0076] YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVE
[0077] QHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGA
[0078] PAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD- (SEQ ID NO:7)

[0079] 靶核酸包括可使用本文所述的共定位复合物来检测的任意核酸序列。靶核酸包括基因。靶核酸可在从单细胞提取的DNA内。靶核酸可以是单染色体提取的DNA。出于本发明的目的,DNA,如双链DNA可包括靶核酸并且共定位复合物可结合至或与DNA共定位在靶核酸处或与之相邻或其附近,并且以一定方式以检测靶核酸。这类靶核酸可包括内源性(或天然产生)核酸和外源性(或外来)核酸。这类靶核酸可以在核酸的混合物中。这类靶核酸可结合至基材。这类靶核酸可使用本领域技术人员已知的方法延长或拉伸。拉伸DNA的方法描述于KH Rasmussen,R Marie,JM Lange,WE Svendsen,A Kristensen,和KU Mir,Lab Chip, 2011,11:1431-44和用于从人染色体提取、操作和拉伸DNA的装置;DLV Bauer,R Marie,KH Rasmussen,A Kristensen,KU Mir,2012 Nucl Acids Res,2012,1-7,DNA连锁维持人中期染色体的结构。

[0080] 可检测标记物或部分是本领域技术人员已知的。本文所用术语“可检测标记物”是指可用于鉴定靶核酸的标记物。可检测标记物使用本领域技术人员已知的方法接合至gRNA或Cas9蛋白。或者,gRNA或Cas9蛋白可包括结合对的一半,而相应结合对的另一半结合至可

检测标记物。通过这种方式,由于结合对的结合,标记物可间接结合至gRNA或Cas9蛋白。合适的结合对或结合力是本领域技术人员已知的并且包括互补核酸序列、生物素-亲和素、生物素-链霉亲和素、NHS-酯等、硫醚连接、静电相互作用、范德华力等(参见例如,Holtke等,美国专利号5,344,757;5,702,888;和5,354,657;Huber等,美国专利号5,198,537;Miyoshi,美国专利号4,849,336;Misiura和Gait,PCT公开号W0 91/17160)。生物素,或其衍生物可用作寡核苷酸标记物(例如,作为靶向部分、可恢复部分和/或可检测标记物),并随后被亲和素/链霉亲和素衍生物(例如,可检测标记的,例如,藻红蛋白-偶联的链霉亲和素),或抗-生物素抗体(例如,可检测标记的抗体)结合。地高辛可以标记物纳入并随后被可检测地标记的抗-地高辛抗体(例如,可检测地标记的抗体,例如,荧光素化的抗-地高辛)结合。氨基烯丙基-dUTP残基可纳入寡核苷酸并随后偶联至N-羟基琥珀酰亚胺(NHS)衍生的荧光染料。一般而言,可将偶联物对的任何一员纳入可恢复部分和/或可检测标记物,限制条件是可检测地标记的偶联物伴侣能被结合,以允许检测。本文中所用的术语抗体指的是任何类别的抗体分子,或其任何子片段,例如Fab。

[0081] 可检测标记物可在尺寸和组成上广泛变化;以下参考文献提供了选择特定实施方式的合适寡核苷酸标签的指南:Brenner,美国专利号5,635,400;Brenner等,Proc.Natl.Acad.Sci.,97:1665;Shoemaker等,(1996)Nature Genetics,14:450;Morris等,EP专利公开号0799897A1;Wallace,美国专利号5,981,179等。

[0082] 将可检测标记物纳入核酸探针的方法是众所周知的。一般而言,可检测标记物(例如,半抗原-或荧光团-偶联的脱氧核糖核苷酸)纳入核酸中,如核酸探针,在聚合或扩增步骤期间,例如,通过PCR、切口平移、随机引物标记、末端转移酶加尾(例如,一个或多个标记物可在引物序列切割之后添加),和其他(参见Ausubel等,1997,《新编分子生物学实验指南》(Current Protocols In Molecular Biology),格林出版和韦利科学公司(Greene Publishing and Wiley-Interscience),纽约)。

[0083] 可使用可检测部分、标记物或报告物来检测本文所述的靶核酸。可以多种方式标记引导RNA或Cas9蛋白,包括直接或间接接合可检测部分,如荧光部分、半抗原、比色部分等。标记物可接合的位置在本文中称为标记物添加位点或可检测部分添加位点,并且可包括标记物能够接合的核苷酸。本领域技术人员能够参考针对核酸或蛋白质标记的参考文献。可检测部分的示例包括各种放射性部分、酶、辅基、荧光标志物、发光标志物、生物发光标志物、金属颗粒、蛋白质-蛋白质结合对、蛋白质-抗体结合对等。荧光部分的示例包括但不限于,黄色荧光蛋白(YFP)、绿色荧光蛋白(GFP)、青色荧光蛋白(CFP)、伞形花内酯、荧光素、荧光素异硫氰酸酯、罗丹明、二氯三嗪胺荧光素、花青、丹磺酰氯、藻蓝蛋白、藻红蛋白等。生物发光标志物的示例包括但不限于萤光素酶(如细菌、萤火虫、叩头虫(click beetle)等)、萤光素、发光蛋白等。有可见检测信号的酶系统示例包括但不限于半乳糖苷酶、葡糖醛酸苷酶(glucorimidase)、磷酸酶、过氧化物酶、乙酰胆碱酯酶等。可鉴定标志物也包括放射活性化合物,如¹²⁵I、³⁵S、¹⁴C、或³H。可鉴定标志物可以从各种来源市售获得。

[0084] 荧光标记物及其与核苷酸和/或寡核苷酸的接合描述于许多综述,包括Haugland,《荧光探针和研究化学品》(Handbook of Fluorescent Probes and Research Chemicals),第9版(分子探针公司(Molecular Probes,Inc.),尤金(Eugene),2002);Keller和Manak,《DNA探针》(DNA Probes),第2版(斯托克顿出版社(Stockton Press),纽

约,1993);Eckstein编,《寡核苷酸和类似物:实践方法》(Oligonucleotides and Analogues:A Practical Approach)(IRL出版社,牛经,1991);和Wetmur,生物化学与分子生物学评论(Critical Reviews in Biochemistry and Molecular Biology),26:227-259(1991)。可应用于本发明的具体方法公开于以下参考文献的样本中:美国专利号4,757,141、5,151,507和5,091,519。在一个方面中,一个或多个荧光染料用作标记的靶序列的标记物,例如,美国专利号5,188,934(4,7-二氯荧光素染料);5,366,860(光谱可分辨若丹明染料);5,847,162(4,7-二氯若丹明染料);4,318,846(醚-取代的荧光素染料);5,800,996(能量转移染料);Lee等,5,066,580(黄嘌呤染料);5,688,648(能量转移染料)等所述。也可用量子点进行标记,如以下专利和专利公开中所示:美国专利号6,322,901、6,576,291、6,423,551、6,251,303、6,319,426、6,426,513、6,444,143、5,990,479、6,207,392,2002/0045045和2003/0017264。本文所用术语“荧光标记物”包括信号转导部分,其通过一个或多个分子的荧光吸收和/或发射性质传递信息。这种荧光性质包括荧光强度、荧光寿命、发射光谱特异性、能量转移等。

[0085] 易于纳入核苷酸和/或寡核苷酸序列的市售荧光核苷酸类似物包括,但不限于Cy3-dCTP、Cy3-dUTP、Cy5-dCTP、Cy5-dUTP(新泽西州皮斯卡塔韦的安玛西亚生物科学公司(Amersham Biosciences,Piscataway,NJ))、荧光素-12-dUTP、四甲基若丹明-6-dUTP、德州红TM-5-dUTP、级联蓝TM-7-dUTP、BODIPYTMFL-14-dUTP、BODIPYTMR-14-dUTP、BODIPYTMTR-14-dUTP、若丹明绿TM-5-dUTP、俄勒冈绿TM488-5-dUTP、德州红TM-12-dUTP、BODIPYTM 630/650-14-dUTP、BODIPYTM650/665-14-dUTP、ALEXA FLUORTM 488-5-dUTP、ALEXA FLUORTM532-5-dUTP、ALEXA FLUORTM 568-5-dUTP、ALEXA FLUORTM594-5-dUTP、ALEXA FLUORTM 546-14-dUTP、荧光素-12-UTP、四甲基若丹明-6-UTP、德州红TM-5-UTP、mCherry、级联蓝TM-7-UTP、BODIPYTM FL-14-UTP、BODIPYTMR-14-UTP、BODIPYTM TR-14-UTP、若丹明绿TM-5-UTP、ALEXA FLUORTM 488-5-UTP、ALEXA FLUORTM 546-14-UTP(俄勒冈州尤金的分子探针有限公司)等。或者,上述荧光团和本文所述的那些可使用例如亚磷酰胺或NHS化学试剂在寡核苷酸合成期间加入。定制合成具有其他荧光团的核苷酸的方案是本领域已知的(参见,Henegariu等,(2000)Nature Biotechnol.18:345)。2-氨基嘌呤是荧光碱基,其可在其合成期间直接纳入寡核苷酸序列中。核酸也可先用例如DAPI、YOYO-1、溴化乙锭、花青染料(例如,SYBR绿)等染色。

[0086] 可用于合成后接合的其他荧光团包括但不限于ALEXA FLUORTM 350、ALEXA FLUORTM 405、ALEXA FLUORTM 430、ALEXA FLUORTM 532、ALEXA FLUORTM 546、ALEXA FLUORTM 568、ALEXA FLUORTM 594、ALEXA FLUORTM 647、BODIPY 493/503、BODIPY FL、BODIPY R6G、BODIPY 530/550、BODIPYTMR、BODIPY 558/568、BODIPY 558/568、BODIPY 564/570、BODIPY 576/589、BODIPY 581/591、BODIPY TR、BODIPY 630/650、BODIPY 650/665、级联蓝、级联黄、丹磺酰、丽丝胺若丹明B、玛丽娜蓝、俄勒冈绿488、俄勒冈绿514、太平洋蓝、太平洋橙、若丹明6G、若丹明绿、若丹明红、四甲基若丹明、德州红(购自俄勒冈州尤金的分子探针公司)、Cy2、Cy3、Cy3.5、Cy5、Cy5.5、Cy7(新泽西州皮斯卡塔韦的安玛西亚生物科学公司)等。也可使用FRET串联荧光团,包括但不限于,PerCP-Cy5.5、PE-Cy5、PE-Cy5.5、PE-Cy7、PE-德州红、APC-Cy7、PE-Alexa染料(610,647,680)、APC-Alexa染料等。

[0087] 也可使用FRET串联荧光团,如PerCP-Cy5.5、PE-Cy5、PE-Cy5.5、PE-Cy7、PE-德州

红、和APC-Cy7;以及PE-Alexa染料(610,647,680)和APC-Alexa染料。

[0088] 金属银或金颗粒也可用于增强来自荧光标记的核苷酸和/或寡核苷酸序列的信号(Lakowicz等,(2003)BioTechniques 34:62)。

[0089] 生物素或其衍生物,也可用作核苷酸和/或寡核苷酸序列上的标记物,并且后续被可检测地标记的亲合素/链霉亲和素衍生物(例如藻红蛋白偶联的链霉亲和素),或可检测地标记的抗生物素抗体结合。生物素/亲和素是配体-配体结合对的示例。抗体/抗原结合对也可与本文所述的方法联用。其他配体-配体结合对或偶联物结合对也是本领域技术人员众所周知的。地高辛可以标记物纳入,并后续被可检测地标记的抗地高辛抗体(例如荧光素化的抗地高辛)结合。氨基烯丙基-dUTP或氨基己基丙烯酸酰胺-dCTP残基可纳入寡核苷酸序列并随后偶联至N-羟基琥珀酰亚胺(NHS)衍生的荧光染料。一般而言,可将偶联物对的任何一员纳入检测寡核苷酸,限制条件是可检测地标记的偶联物伴侣能被结合,以允许检测。本文中所用的术语抗体指的是任何类别的抗体分子,或其任何子片段,例如Fab。

[0090] 其他合适的标记物包括荧光素(FAM,FITC)、地高辛、二硝基酚(DNP)、丹磺酰、生物素、溴脱氧尿苷(BrdU)、六聚组氨酸(6xHis)、磷光体-氨基酸(例如,P-tyr,P-ser,P-thr)等。在一个实施方式中,以下半抗原/抗体对用于检测,其中各抗体用以下可检测标记物衍生化:生物素/ α -生物素、地高辛/ α -地高辛、二硝基酚(DNP)/ α -DNP、5-羧基荧光素(FAM)/ α -FAM。

[0091] 在某些示例性实施方式中,核苷酸和/或寡核苷酸序列可间接标记,尤其是用随后被捕获剂结合的半抗原,例如,如美国专利号5,344,757、5,702,888、5,354,657、5,198,537和4,849,336、PCT公开号W0 91/17160等所述。可使用许多不同的半抗原-捕获剂对。示例性的半抗原包括但不限于生物素、des-生物素和其他衍生物、二硝基酚、丹磺酰、荧光素、CY5、地高辛等。对于生物素,捕获剂可以是亲和素、链霉亲和素或抗体。抗体可用作其他半抗原的捕获剂(许多染料-抗体对是市售的,例如,俄勒冈州尤金的分子探针公司)。

[0092] 根据某些方面,本文所述的可检测部分是光谱可分辨的。针对多种荧光标记物的“光谱可分辨”表示标记物的荧光发射带足够不同,即,足够不重叠,使得可通过标准光检测系统基于由相应标记物生成的荧光信号区分相应标记物接合的分子标签,例如,采用带通滤光片和光电倍增管的系统等,如美国专利号4,230,558、4,811,218等,或Wheelless等,《流式细胞术:设备与数据分析》(Flow Cytometry:Instrumentation and Data Analysis)(学术出版社,纽约,1985)中第21-76页所述的系统所例示。在一个方面中,光谱可分辨的有机染料,如荧光素、若丹明等表示波长发射峰间隔至少20nm,并且在另一个方面中,至少40nm。在另一个方面中,对于镧系化合物、量子点等,光谱可分辨表示波长发射峰间隔至少10nm,并且在另一个方面中,至少15nm。

[0093] 在某些实施方式中,可检测部分可在与电子显微术联用时与常用核酸相比提供更高的可检测性。具有较高可检测性的部分通常为金属和有机金属,如乙酸汞、二甲基亚砷铂、几种金属-二吡啶基复合物(例如,钌-bipy、钇-bipy、铂-bipy)。虽然这些部分中的一些可易于对核酸特异性染色,但也可使用接头来连接这些部分与核酸。在合成期间添加至核苷酸的这类接头是丙烯酰胺-和巯基-修饰的实体,氨基反应基团,以及叠氮化物和炔烃基团,用于进行点击化学。也更容易检测到一些核酸类似物,如 γ -腺苷-三磷酸酯、碘脱氧胞苷-三磷酸酯、和一般金属核苷(参见例如,Dale等,Proc.Nat.Acad.Sci.USA,第70卷,第8

期,第2238-2242页(1973))。在合成期间添加修饰的核苷酸。合成可指代,例如,寡核苷酸的固体支持物合成。在这种情况下,修饰的核酸(其可以是核酸类似物、或用可检测部分修饰的核酸、或具有接合化学接头)可一个接着一个添加至正在固体支持物上形成的核酸片段,通过亚磷酸胺合成是最常见的方法。合成也可指代由聚合酶进行的过程,同时其合成核酸模板的互补链。某些DNA聚合酶能够使用并纳入核酸类似物,或修饰的核酸(用可检测部分或接合化学接头修饰)至互补核酸模板。

[0094] 使用的检测方法将取决于反应性标记物、可恢复标记物和/或可检测标记物中使用的具体可检测标记物。在某些示例性实施方式中,通过本文所述的探针具有与其结合的一个或多个反应性标记物、可恢复标记物、或可检测标记物的包括但不限于间期、早前期、前期、前中期、中期、后期、末期和胞质分裂的细胞周期各阶段期间的靶核酸,如染色体和染色体的亚染色体区域可经选择和/或扫描用于使用显微镜、分光光度计、管发光计或板发光计、x-射线膜、闪烁器、荧光活化的细胞分选(FACS)设备、微流体设备等。

[0095] 本文所用术语“染色体”是指活细胞中携带遗传的基因的支持物,包括DNA、蛋白质、RNA和其他相关因子。本文中使用用于鉴定和编号人基因组染色体的常规国际系统。单个染色体的尺寸将在多染色体基因组内变化并且在基因组间变化。染色体可获自任何物种。染色体可获自成年对象、青少年对象、婴儿对象、未出生对象(例如,来自胎儿,例如,通过产前测试如羊膜穿刺术、绒毛膜绒毛取样、等或直接来自胎儿,例如胎儿手术期间)来自生物样品(例如,生物组、流体或细胞(例如,痰液、血液、血细胞、组织或细针活检样品、尿液、脑脊液、腹膜液、和胸膜液、或来自其的细胞)或来自细胞培养样品(例如,原代细胞、永生细胞、部分永生细胞等)。在某些示例性实施方式中,一种或多种染色体可获自一个或多个属,包括但不限于人属、果蝇属、线虫属、鱼丹属、鲤属、猫属、犬属、羊属、大马哈鱼属、鳟属、牛属、猪属、原鸡属、番茄属、小麦属、稻属、玉蜀黍属、大麦属、芭蕉属、燕麦属、杨属、芸薹属、甘蔗属等。

[0096] 当使用荧光标记的靶向部分或可检测标记物时,可使用荧光光学显微镜来检测并记录采用本领域已知的常规方法的原位杂交的结果。或者,可使用具有图像处理能力的数字(计算机执行)荧光显微镜。2种熟知的用于对具有与之结合的多色标记物的染色体的成像FISH的系统包括多重-FISH(M-FISH)和光谱核型分析(SKY)。参见Schrock等,(1996) *Science* 273:494;Roberts等,(1999) *Genes Chrom. Cancer* 25:241;Fransz等,(2002) *Proc. Natl. Acad. Sci. USA* 99:14584;Bayani等,(2004) *Curr. Protocol. Cell Biol.* 22.5.1-22.5.25;Danilova等,(2008) *Chromosoma* 117:345;美国专利号6,066,459;和FISH TAGTM DNA多色试剂盒说明书(分子探针公司),针对用于喷涂染色体和检测喷涂的染色体的方法的综述。

[0097] 在某些示例性实施方式中,使用带修饰(例如,软件,Chroma 84000滤片组,和增强滤波器)的计算机成像系统,如Applied Imaging Corporation CytoVision系统(加利福尼亚州圣克拉拉的应用成像公司(Applied Imaging Corporation, Santa Clara, Calif.))来检测并记录荧光标记的染色体的图像。其他合适的系统包括计算机成像系统,使用与Zeiss Axiophot显微镜偶联的冷却的CCD相机(Photometrics,装配Kodak KAF 1400 CCD的NU200系列),图像处理如Ried等,(1992) *Proc. Natl. Acad. Sci. USA* 89:1388所述。其他合适的成像和分析系统由Schrock等,同上和Speicher等,同上描述

[0098] 使用本文所述的方法生成的探针的原位杂交方法可在多种生物或临床样品上,在处于任何(或全部)细胞周期(例如,有丝分裂、减数分裂、间期、G0、G1、S和/或G2)的细胞中进行。示例包括全部类型的细胞培养物、动物或植物组织、外周血淋巴细胞、口颊涂片、从未经培养的原代肿瘤制备的涂片标本、癌细胞、骨髓、从活检获得的细胞、或体液(例如,血液、尿液、痰液等)中的细胞、来自羊水的细胞、来自母体血液的细胞(例如,胎儿细胞)、来自睾丸和卵巢的细胞等。样品经制备用于使用常规技术的本发明的试验,其一般取决于获得样品或试样的来源。这些例子不构成对可用于本文所述的方法和/或组合物的样品类型的限制。

[0099] 在某些示例性实施方式中,探针包括差异标记的多个gRNA/Cas9复合物(即,gRNA/Cas9复合物中的至少两个是差异标记的)。多色染色体喷涂的各种方法已经在本领域中描述并且可按照本文提供的指南适用于本发明。这种差异标记的示例(“多色FISH”)包括Schrock等,(1996) Science 273:494,和Speicher等,(1996) Nature Genet. 12:368所述的那些。Schrock等描述了光谱成像方法,其中使用落射荧光滤波器组和计算机软件来检测并区分差异标记的DNA探针,其同时与靶染色体组杂交。Speicher等描述了使用5种荧光团的不同组合在27-色FISH(称为“组合多荧光FISH”)中标记人染色体(或染色体臂)。也可使用其他合适方法(参见例如,Ried等,1992, Proc. Natl. Acad. Sci. USA 89:1388-92)。

[0100] 根据某些方面,Cas9-gRNA复合物用于探测并评价天然双链DNA上的感兴趣区域,而不需要使靶DNA变成单链。使用本领域技术人员已知的方法设计针对感兴趣的靶双链核酸序列有特异性的引导RNA,用Cas9预孵育,然后添加至含有靶DNA的样品。引导RNA和Cas9然后将共定位至靶DNA并与靶DNA形成复合物。

[0101] 基于本发明,本领域技术人员将能易于鉴定或设计引导RNA和Cas9蛋白,其共定位至包括靶核酸的DNA。本领域技术人员还将能够鉴定用于直接或间接结合至引导RNA或Cas9蛋白的可检测部分。DNA包括基因组DNA、线粒体DNA、病毒DNA或外源性DNA。

[0102] 根据一个方面,设计了针对感兴趣的序列有特异性的引导RNA。gRNA与Cas9预孵育,然后向含有靶DNA的样品中添加组合,或者与靶DNA接触。gRNA或Cas9可包括可检测标记物或者该可检测标记物可在复合物形成之后添加。组分混合物将全部以溶液提供或者靶核酸可固定在表面上或存在于细胞或组织内。

[0103] 根据本发明的方面,本文所述的CRISPR Cas9系统具有序列特异性的优势(适当设计)并且靶序列通过gRNA上的17-25个核苷酸间隔子序列“程序化”。

[0104] 当某些序列用于gRNA的“种子”区域中时,Cas9系统显示高效结合;基于这些短序列的基因组中的频繁出现,这可用作基因组作图工具。

[0105] Cas9系统也可经编程以变为作图工具,通过在gRNA的非种子区域处使用简并位置(和/或通用碱基)。

[0106] 为了增加标记的特异性,引导物的聚类可在感兴趣基因座周围结合。

[0107] 可通过RNA的直接固相合成(购自供应商如IDT)或通过固相合成的DNA寡聚物的体外转录来制备引导RNA。

[0108] gRNA可易于从阵列-合成的寡聚物(其在大于所需的约100-200nt的长度上可得)合成,并经扩增使得各引导物的价值非常低,并且使大量gRNA的生成更易于放大。例如,定制阵列公司(Custom Array Inc.)在其设备的一轮运行中可提供90000个适于gRNA生成的

阵列合成的寡聚物。

[0109] 反应动力学是等温的(37℃,可能室温)、快速的、在1分钟内,并且所得的复合物非常稳定并易于用作探针。

[0110] 靶DNA可以是溶液中或固定在表面上的堆积DNA、细胞中的原位DNA、在表面上铺展的染色体上的DNA、和在表面上或纳米通道中拉伸的单个DNA分子。

[0111] 其他工程改造的核酸酶,如寻靶核酸内切酶(HE)、大范围核酸酶、转录活化样效应子核酸酶(TALEN)、锌指核酸酶(ZFN)、原核阿尔古蛋白质(pAgo)、或BurrH-基核酸酶(BuDN)可代替Cas9使用,或与Cas9平行使用。例如,TtAGO对RNA有高亲和性并且对dsDNA有低亲和性。

[0112] 根据某些方面,可通过直接标记Cas9蛋白来检测DNA-结合的Cas9-sgRNA(例如,通过与量子点或有机染料结合的亲和标签)。市售的Cas9蛋白已经含有亲和标签(来自PNABio公司的Cas9包含人流感病毒血凝素(HA);来自新英格兰生物实验室公司的Cas9包含组氨酸(His)标签)。可通过标记gRNA来检测DNA-结合的Cas9-sgRNA,如在gRNA的3'端的尾部分处,并且其中可探测尾部分。例如,荧光部分可结合至尾,不同颜色的探针可结合以产生编码方案,探针可交换以增加代码库。与DNA-PAINT结合,可实现超高分辨率成像。与流体装置结合,可进行EXCHANGE-PAINT,其使得能够在超高分辨率下进行多重化。在该方案中,通过进行试剂交换循环,有限数量的代码可用于对大量基因座的超高分辨率成像,其中在各循环处使用相同颜色但连接至不同DNA PAINT成像器序列。例如,仅2种标记物(包括Cy3B,Atto 655)使用5次,每次与不同成像器序列偶联,具有编码10种gRNA的容量。在每次循环中,gRNA的亚组变得被标记。在循环已经完成之后,通过确定颜色和特定gRNA点亮时的循环数来解码gRNA的种类。

[0113] 为了增加信号强度,用多种荧光团标记的寡聚物可结合至gRNA的尾部分。或者,可进行通过使用结合尾并以锁式探针环化的寡聚物由尾的3'端引发的滚环扩增。也可使用杂交链反应或本领域技术人员已知的其他信号扩增方法。

[0114] 检测方法包括荧光检测方法、电致发光检测方法、生物发光检测方法和比色检测方法。

[0115] 可使用除涉及检测荧光、电致发光、化学发光、生物发光或比色部分或复合物的那些以外的检测方法如使Cas9-sgRNA结合的DNA链通过纳米孔或纳米间隙或纳米通道来确定Cas9-sgRNA的结合位置,使用电子显微术或扫描探针显微术来检测Cas9-sgRNA与表面上延伸/拉伸的DNA的结合位置,使用悬臂、石英晶体微天平、场效应晶体管等检测Cas9-sgRNA与靶DNA的结合。

[0116] 根据某些方面,使用本领域技术人员已知的纳米孔或纳米间隙检测技术或者纳米孔或纳米间隙测序技术来确定靶核酸处的gRNA和Cas9的复合物的存在。简言之,在导电介质中,具有与之结合的gRNA和Cas9的靶核酸在电压差的影响下通过纳米孔。使用离子电流中的界面依赖性变化来区分单个核苷酸和与核酸结合的gRNA/Cas9复合物。可通过这种方式检测gRNA/Cas9复合物的存在。根据一个方面,离子电流中的界面依赖性变化决定靶核酸进入纳米孔或纳米间隙和gRNA/Cas9是否结合至靶核酸以及结合的位置。当线性聚合物的核酸进入纳米孔时,离子电流下降,因为孔中聚合物的物理存在干扰离子流过孔。如果gRNA/Cas9复合物结合至DNA上的具体位置,则当该位置进入孔时,离子流进一步减少,降低

了离子电流。这种电流的降低表明gRNA/Cas9复合物结合至靶核酸。根据其尺寸和物理-化学性质,与靶核酸(即,DNA聚合物)结合的各种类型的结构或复合物将产生离子电流的特征性变化。可差异标记靶向不同位置或等位基因的gRNA-Cas9复合物,使得可通过纳米孔读取来区分它们。

[0117] “纳米孔”表示具有纳米级宽度的洞或通道,如通过膜的平坦表面的洞或通道。可通过多亚基蛋白环,如在脂质双层中形成纳米孔。纳米孔可以氮化硅、石墨烯或这类非生物材料的固态平面表面中的物理洞。一般而言,通道是0.2-25nm宽。本文所用的纳米孔可包括跨膜结构,其可使分子通过膜。纳米孔的示例包括 α -溶血素(金黄色葡萄球菌(*Staphylococcus aureus*))和MspA(耻垢分枝杆菌(*Mycobacterium smegmatis*))。纳米孔的其他示例可在说明纳米孔测序的领域中发现,或在孔形成毒素领域中描述,如 β -PFT Panton-Valentine杀白细胞素S、气单胞菌溶素、和梭菌 ϵ 毒素、 α -PFT溶细胞素A、二元PFT炭疽毒素、或其他如肺炎链球菌溶血素或短杆菌肽。随着纳米孔测序技术的出现,纳米孔在技术和经济上正变得重要。纳米孔测序的方法是本领域已知的,例如,如U.S.P.N.5,795,782中所述,其通过引用纳入本文。简言之,纳米孔检测包括将纳米孔-穿孔的膜浸入电压-导电流体中,如包含例如KCl、NaCl、NiCl、LiCl或本领域技术人员已知的其他离子形成无机化合物的离子溶液。跨膜施加电压,并且从通过纳米孔的离子传导产生电流。当纳米孔与聚合物相互作用如DNA时,根据在任何给定时间移位通过孔的聚合物亚片段的特性调节通过纳米孔的流,如以单体-特异性方式,产生能够鉴定单体或亚片段的电流变化。本发明范围内的纳米孔包括本领域技术人员已知的固态非蛋白纳米孔和本领域技术人员已知的DNA折纸纳米孔。这种纳米孔提供比已知的蛋白质纳米孔更大的纳米孔宽度,其使得能够通过较大的分子用于检测,如带双链靶核酸的Cas9/gRNA复合物,同时仍然足够敏感以检测复合物通过纳米孔时的离子电流变化。

[0118] “纳米孔分析”表示基于聚合物与纳米孔的相互作用确定聚合物组分,如包含gRNA/Cas9复合物的多核苷酸的方法。可通过测量当通过与聚合物相互作用改变开口尺寸时发生的通过纳米孔的离子的导电性变化来实现纳米孔分析。

[0119] 除了纳米孔以外,本申请设想使用作为两个电极之间的间隙的本领域已知的纳米间隙,其中间隙的宽度为约几纳米,如约0.2nm至约25nm或约2nm至约5nm。间隙模拟纳米孔中的开口并且使DNA通过或跨过两个电极之间的间隙。本发明的方面还设想使用纳米通道。电极与DNA通过的纳米通道相邻放置。另外或或者,当复合物经光学标记时,可确定纳米通道中拉伸的DNA聚合物的复合物结合位置。应理解本领域技术人员将易于设想分子或部分鉴定和测序的不同实施方式,其基于通过电场的分子或部分运动并产生代表通过电场的结构的电场变形。

[0120] 根据另一个方面,可使用Cas9切口酶来切开靶标双链核酸并且该切口用作基于聚合酶或连接酶的测序的序列限定的起始位点,从而显示Cas9-sgRNA靶标位点附近的序列信息。可设计引导物文库,其能够对基因组的许多选择部分,例如,外显子,由GWAS信号鉴定的区域,与心脏病、癌症等相关的特定基因进行测序。也可使用引物延伸来标记切口位点,如通过纳入荧光核苷酸,通过连接或通过纳入荧光核苷酸,之后进行连接。如果引物延伸置换已有的链,则可通过例如用寡聚体与之杂交来标记置换的侧翼。如果存在多个切口位点,如当部分引导物含有简并序列时,则多个标记位点可用作作图工具。引物延伸方法为本领域

技术人员已知。

[0121] 本文所述方法的具体应用包括鉴定或诊断或作图方法。通常人类基因组的“暗物质”，如着丝粒重复，如图6A所示。由于高重复性，现有的参考基因组中大多数没有着丝粒DNA序列。使用CRISPR/Cas9的本文所述的单分子作图或测序方法使得以靶向的方式在超高分辨率下通过对重复位点进行作图来对参考基因组进行更完全组装，并且也使得能够从这些位点进行测序。这类方法可用于个人基因组。

[0122] 本发明的方法包括以比目前使用标记的克隆进行的那些更高的效率和分辨率进行对染色体分散上的FISH。标记的gRNA/Cas9复合物是本文所述的示例性方法中的有效FISH探针，因为其使得能够获得更清晰的信号。该方法使得能够对与许多癌症和不孕不育问题相关的染色体断裂点、复合物易位或重排进行更快且更好的鉴定。

[0123] 本发明的方法使得能够进行体外诊断。

[0124] 本发明的方法能够在快速诊断平台中探测在细菌或病毒中编码多重耐药性的基因；在这种情况下，Cas9-sgRNA可用于直接并稳定地靶向dsDNA。

[0125] 本发明也设想以下示例性方法。

[0126] 包括Mg存在的野生型Cas9

[0127] 在一些实施方式中，本发明包括检测、标记、拉下或靶向核酸中的位点的方法，包括：(a) 使核酸与包含具有或保留蛋白酶活性的Cas9蛋白和引导RNA的复合物在下述条件下接触：使复合物结合核酸，切割核酸，但不易于从核酸解离（即，保持接合核酸，将2条切割的链保持在一起），和 (b) 分析步骤 (a) 的产物。在一些实施方式中，这种条件包括存在二价阳离子如 Mg^{2+} 。在一些实施方式中，在步骤 (a) 和步骤 (b) 之间存在至少一个洗涤步骤。在一些实施方式中，所得复合物的至少一种组分被标记或带标签。在一些实施方式中，引导RNA结合与天然PAM位点相邻的序列。在一些实施方式中，引导RNA结合与人工PAM位点相邻的序列。在一些实施方式中，Cas9是Cas9的变化形式。在一些实施方式中，引导RNA是截短的引导RNA。在一些实施方式中，截短的引导物包括引导RNA的仅种子区域，即，与PAM位点相邻的4-7个核苷酸。

[0128] Mg不存在下的野生型Cas9

[0129] 在一些实施方式中，本发明包括检测、标记、拉下或靶向核酸中的位点的方法，包括：(a) 使核酸与包含具有或保留蛋白酶活性的Cas9蛋白和引导RNA的复合物在下述条件下接触：使复合物结合核酸，不切割核酸且不易于从核酸解离，和 (b) 分析步骤 (a) 的产物。在一些实施方式中，这种条件包括不存在二价阳离子如 Mg^{2+} 。在一些实施方式中，在步骤 (a) 和步骤 (b) 之间存在至少一个洗涤步骤。在一些实施方式中，所得复合物的至少一种组分被标记或带标签。在一些实施方式中，引导RNA结合与天然PAM位点相邻的序列。在一些实施方式中，引导RNA结合与人工PAM位点相邻的序列。在一些实施方式中，Cas9是Cas9的变化形式。在一些实施方式中，引导RNA是截短的引导RNA。在一些实施方式中，截短的引导物包括引导RNA的仅种子区域，即，与PAM位点相邻的4-7个核苷酸。

[0130] 无效/死Cas9

[0131] 在一些实施方式中，本发明包括检测、标记、拉下或靶向核酸中的位点的方法，包括：(a) 使核酸与包含无酶活性或核酸酶无效的Cas9蛋白（例如，D10A/H840A dCas9）和引导RNA的复合物在复合物不切割核酸且不易于从核酸解离的条件下接触，和 (b) 分析步骤 (a)

的产物。在一些实施方式中,在步骤(a)和步骤(b)之间存在至少一个洗涤步骤。在一些实施方式中,所得复合物的至少一种组分被标记或带标签。在一些实施方式中,引导RNA结合与天然PAM位点相邻的序列。在一些实施方式中,引导RNA结合与人工PAM位点相邻的序列。在一些实施方式中,Cas9是Cas9的变化形式。在一些实施方式中,引导RNA是截短的引导RNA。在一些实施方式中,截短的引导物包括引导RNA的仅种子区域,即,与PAM位点相邻的4-7个核苷酸。

[0132] 切口酶Cas9

[0133] 在一些实施方式中,本发明包括检测、标记、拉下或靶向核酸中的位点的方法,包括:(a)使核酸与包含Cas9切口酶(即,Cas9蛋白的切口突变体如Cas9 D10A或H840A突变体)和引导RNA的复合物在复合物不易于从核酸解离的条件下接触,和(b)分析步骤(a)的产物。在一些实施方式中,在步骤(a)和步骤(b)之间存在至少一个洗涤步骤。在一些实施方式中,所得复合物的至少一种组分被标记或带标签。在一些实施方式中,引导RNA结合与天然PAM位点相邻的序列。在一些实施方式中,引导RNA结合与人工PAM位点相邻的序列。在一些实施方式中,Cas9是Cas9的变化形式。在一些实施方式中,引导RNA是截短的引导RNA。在一些实施方式中,截短的引导物包括引导RNA的仅种子区域,即,与PAM位点相邻的4-7个核苷酸。

[0134] 没有DNA结合蛋白的情况下使用引导RNA的方法

[0135] 在一些实施方式中,本发明包括检测、标记、拉下或靶向核酸中的位点的方法,包括:(a)使核酸与引导RNA在其不易于从核酸解离的条件下接触,和(b)分析步骤(a)的产物。在一些实施方式中,在步骤(a)和步骤(b)之间存在至少一个洗涤步骤。在一些实施方式中,所得引导RNA-DNA复合物经标记或带标签。在一些实施方式中,标记物或标签在尾上。在一些实施方式中,引导RNA结合与天然PAM位点相邻的序列。在一些实施方式中,引导RNA结合与人工PAM位点相邻的序列。在一些实施方式中,引导RNA不需要结合与PAM位点相邻的序列。在一些实施方式中,引导RNA是截短的引导RNA。在一些实施方式中,截短的引导RNA包括引导RNA的仅种子区域,即,与PAM位点相邻的4-7个核苷酸。在一些实施方式中,靶基因座是重复DNA的区域并且信号被放大。在一些实施方式中,靶核酸在添加引导RNA之前变性。

[0136] 使用RNA的方法

[0137] 在一些实施方式中,本发明包括检测、标记、拉下或靶向核酸中的位点的方法,包括:(a)使核酸与RNA在其结合核酸的条件下接触,和(b)分析步骤(a)的产物。在一些实施方式中,在步骤(a)和步骤(b)之间存在至少一个洗涤步骤。在一些实施方式中,所得RNA-DNA复合物经标记或带标签。在一些实施方式中,标记物或标签在尾上。在一些实施方式中,靶基因座是重复DNA的区域并且信号被放大。

[0138] 双链体去稳定/开口剂和RNA

[0139] 在一些实施方式中,本发明包括检测、标记、拉下或靶向核酸中的位点的方法,包括:(a)使核酸与双链体去稳定/开口剂和引导RNA或RNA在下述条件下接触:复合物形成并结合核酸,且复合物不易于从核酸解离,和(b)分析步骤(a)的产物。在一些实施方式中,在步骤(a)和步骤(b)之间存在至少一个洗涤步骤。在一些实施方式中,所得复合物的至少一种组分被标记或带标签。在一些实施方式中,标记物或标签在尾上。在一些实施方式中,Cas9是Cas9的变化形式。在一些实施方式中,引导RNA是截短的引导RNA。在一些实施方式中,截短的引导物包括引导RNA的仅种子区域,即,与PAM位点相邻的4-7个核苷酸。在一些实

施方式中,靶基因座是重复DNA的区域并且信号被放大。去稳定剂(其包括单链DNA稳定剂)包括解旋酶、复制蛋白A(RPA)、大肠杆菌单链结合蛋白(SSB)、甜菜碱、甜菜碱/甘氨酸、甲酰胺、尿素、DMSO等。双链体开口剂包括引发体蛋白PriA、三链体形成双-PNA、 γ PNA等。

[0140] 体外RNA合成和探测

[0141] 在一些实施方式中,本发明包括检测、标记、拉下或靶向核酸中的位点的方法,包括:(a)在无细胞系统中合成RNA,(b)使核酸与RNA或其他组分接触,形成复合物并且所述复合物不易于从核酸解离,和(c)分析步骤(b)的产物。在一些实施方式中,在步骤(b)和步骤(c)之间存在至少一个洗涤步骤。在一些实施方式中,所得复合物的至少一种组分被标记或带标签。在一些实施方式中,标记物或标签在尾上。在一些实施方式中,Cas9是Cas9的变化形式。在一些实施方式中,引导RNA是截短的引导RNA。在一些实施方式中,截短的引导物包括引导RNA的仅种子区域,即,与PAM位点相邻的4-7个核苷酸。

[0142] 切口和测序

[0143] 在一些实施方式中,本发明包括靶向测序,包括(a)使核酸与包含Cas9蛋白的切口突变体和靶向特定位置的引导RNA在复合物诱导核酸的一条链中的切口的条件下接触,(b)用核苷酸延伸切口的3'端,(c)检测步骤(b)的产物并以一定方式重复步骤(b)以对DNA进行测序。在一些实施方式中,核苷酸经标记。在一些实施方式中,核苷酸在末端磷酸酯处经标记。在一些实施方式中,核苷酸含有超过三个磷酸酯。在一些实施方式中,核酸在碱基处通过可切割连接标记。在一些实施方式中,核苷酸是可逆终止子。在一些实施方式中,在碱基处的标记物提供可逆终止。在一些实施方式中,在糖上3'或2'位处的修饰提供了可逆终止。在一些实施方式中,延伸同时可使用所有4种核苷酸。在一些实施方式中,使用单分子检测方法。在一些实施方式中,以线性串分析单分子。在一些实施方式中,在细胞中原位分析核酸。在一些实施方式中,在细胞中原位分析核酸的情况中,在分析之前固定细胞。在一些实施方式中,原位分析的核酸是DNA分子并且在分析之前去除RNA分子。在一些实施方式中,使用试剂在切口后从核酸去除复合物。在一些实施方式中,在步骤(b)和步骤(c)之间存在至少一个洗涤步骤。在一些实施方式中,Cas9是Cas9的变化形式。在一些实施方式中,引导RNA是截短的引导RNA。在一些实施方式中,截短的引导物包括引导RNA的仅种子区域,即,与PAM位点相邻的4-7个核苷酸。

[0144] 切口和捕获

[0145] 在一些实施方式中,用修饰的核苷酸延伸切口的3'端。在一些实施方式中,修饰的核苷酸是生物素修饰的。在一些实施方式中,使用纳入的生物素来标记靶DNA,例如,通过与自身标记(或变得标记)的链霉亲和素/中性亲和素或抗-生物素抗体的相互作用。在一些实施方式中,使用纳入的生物素,例如,通过与自身结合(或变得结合)生物素包被的捕获材料如磁珠或琼脂糖珠,或表面的链霉亲和素/中性亲和素或抗-生物素抗体的相互作用来捕获靶DNA。在一些实施方式中,使用Cas9/gRNA引导的切口和经修饰以辅助捕获的碱基纳入在溶液中进行反应中分离核酸样品的特定单个或多个部分。在该实施方式中,例如,在切口和纳入生物素化的dUTP之后,产物与链霉亲和素包被的磁珠反应一段时间(例如,1小时),这段时间使基因组的生物素化的部分被珠上的链霉亲和素结合。然后施加磁体并且接合至固相珠的基因组的靶向的部分从上清分离,从而分离靶向的感兴趣区域。弃去上清。在各种程度的洗涤严谨性并去除上清之后,通过本领域已知的方法从珠分离选择的基因组DNA(例

如,通过加热超过90℃)。然后可通过如本领域技术人员已知的下一代测序方法和装置来对捕获的分子进行测序,其可包括选自尺寸选择、抛光、条码化、加尾、文库制备、聚类扩增、集落扩增等的步骤。本文所述的用于选择或富集核酸样品的部分或亚组的切口和纳入方法比引导RNA/Cas9的结合有优势,因为延伸使得能够纳入多个生物素,从而改善了捕获效率。另一个优势是降低的脱靶捕获,因为在靶标变得可捕获之前需要三个步骤:gRNA/cas9结合、切口和延伸。这种选择方法比现有方法更清楚(例如,Sureselect),并且导致更少测序,因为脱靶序列更少。在一些实施方式中,在切口的5'端上进行反应,如连接反应,如连接生物素化的寡核苷酸,作为示例。在一些实施方式中,可分离DNA双链体的正义和反义链。根据这一方面,通过gRNA/Cas9结合或通过纳入生物素化的核苷酸捕获一条链并且从上清收集另一条链。或者,可通过结合至单链结合蛋白、羟基磷灰石或结合至序列特异性寡核苷酸来捕获由gRNA结合置换的双链体的链。

[0146] 切口和直接测序

[0147] 在一些实施方式中,用接合至表面的靶核酸,即DNA进行gRNA引导的切口,或者,在进行切口之后,DNA接合至表面。在一些实施方式中,然后可使用切口的3'端来启动DNA测序。切口的3'端允许进行基于聚合酶的测序,如通过合成的亿明达(Illumina)测序。切口的3'和5'端都支持基于连接的测序,如SOLID(生命技术公司)和通过连接的测序(完整基因组学公司(Complete Genomics Inc))。在一些实施方式中,基因组DNA在长的长度上保留并且在其已经接合至表面之前或之后进行切口。在一些实施方式中,DNA接合至表面并拉伸或延长使得能够分析序列或沿着其长度的特征。这种分析可通过光学、电子、X-射线、或扫描探针显微术来进行。在一些实施方式中,在通过光学方法进行分析的情况中,在设置在表面上的多核苷酸上的测序反应上进行全内反射或渐消波/波导成像。在一些实施方式中,多核苷酸可以线性化,但是不接合至表面。在一些实施方式中,可通过多核苷酸接合在一端并在流中悬空来产生线性化。在一些实施方式中,通过流体力学拖曳力使靶核酸,即DNA基本呈线性。在一些实施方式中,通过设置在纳米缝、纳米通道或纳米沟中的纳米-限制来拉伸DNA。一些实施方式中,线性DNA是基本直的。在一些实施方式中,gRNA引导的切口引导在长线性多核苷酸上的多个选择序列位置处进行通过合成的测序。在该实施方式中,通过纳入在通过分析的测序中可检测的核苷酸从切口延伸聚合酶(例如,9 Degree North或其突变体或者Phi29或其突变体)。这种检测可通过pH(如离子激流测序)。在一些实施方式中,检测是通过核苷酸上的荧光标记物。在一些实施方式中,也被荧光标记的核苷酸也发挥可逆终止子的作用并且因此允许通过本领域已知的方法进行逐步四色测序(如亿明达或Lasergen测序)。纳入的核苷酸可以是闪电终止子(Lasergen),其中光可切割部分的切割是通过UV光。在一些实施方式中,gRNA/Cas9复合物形成之后的测序是出于进行选择性或靶向测序的目的。在其他实施方式中,gRNA含有至少一部分的简并部分,并且测序从基因组上分布的多个起始位点开始并且对于特定基因组不是选择性的。在一些实施方式中,发生切口同时DNA在细胞内。在一些实施方式中,细胞被固定。例如,切口可形成荧光原位测序(FISSEQ)反应的部分,其中对DNA进行测序。在一些实施方式中,使用切口来诱导细胞内基因组DNA中的切口,然后启动与切口相邻的区域的扩增,例如,通过链置换聚合酶如Phi29的分支或滚环扩增。然后在扩增的产物上进行测序。或者,使用单分子测序方法,如TrueSeq(Helicos Bio/SeqLL)直接从切口开始对切口的基因组DNA进行测序。

[0148] 结合和纳米孔分析

[0149] 在一些实施方式中,本发明包括检测核酸中结合位点或序列的位置的方法,包括:(a)使核酸与包含Cas9蛋白和引导RNA的复合物在该复合物不易于从核酸解离的条件下接触;(b)使核酸通过纳米孔或纳米间隙,和(c)分析结合位置。在一些实施方式中,使用单分子检测方法。在一些实施方式中,使用单通道记录。在一些实施方式中,以线性串分析单分子。

[0150] 脱靶结合

[0151] 在一些实施方式中,本发明包括检测引导RNA脱靶结合位点的方法,包括:(a)使核酸与包含Cas9蛋白和引导RNA的复合物在该复合物不易于从核酸解离的条件下接触,(b)确定结合的位置,(c)确定结合的靶向的位置,(d)确定脱靶结合的位置,和(e)通过脱靶结合的位置确定脱靶结合的序列的相同性。在一些实施方式中,使用标记物来提供标志,可通过参考其来确定靶或脱靶结合。在一些实施方式中,标记物包含在核酸上产生物理图的结合试剂。在一些实施方式中,结合试剂可以是以下的一种或多种:非-混杂gRNA、限制性酶、切口酶、寡核苷酸等。在一些实施方式中,使用单分子检测方法。在一些实施方式中,以线性串分析单分子。

[0152] 检测拷贝数

[0153] 在一些实施方式中,本发明包括确定染色体或基因组的区域的拷贝数,包括(a)使靶核酸序列接触引导RNA序列,该引导RNA序列具有与拷贝数待确定的该染色体或基因组区域和Cas9蛋白互补的部分,以及与参考染色体和/或基因组区域和Cas9蛋白互补的部分,(b)获得来自拷贝数待确定的基因组的区域/染色体与参考染色体或基因组区域的信号之比。在一些实施方式中,该方法应用于非整倍性检测。在一些实施方式中,非整倍性是21三体。在一些实施方式中,拷贝数待确定的基因组是LSI21q22.13-q22.2。

[0154] Her2

[0155] 在一些实施方式中,本发明包括确定Her2扩增程度的方法,包括:(a)使靶核酸序列与引导RNA序列接触,该引导RNA序列具有与Her2基因座和Cas9互补的部分,以及与参考基因座和Cas9蛋白互补的部分,和(b)获得来自Her2基因座与参考基因座的信号之比。

[0156] 基因融合

[0157] 在一些实施方式中,本发明包括确定出现基因融合的方法,包括:(a)使靶核酸序列与具有与第一基因组基因座和Cas9蛋白互补的部分的引导RNA探针序列,和具有与第二基因座基因座和Cas9蛋白互补的部分的引导RNA探针序列接触,(b)检测第一和第二基因座之间的共定位事件,其中基因融合是基因组区域之间的任意融合。根据一个方面,共定位产生互相相邻的探针。

[0158] 断裂试验

[0159] 在一些实施方式中,本发明包括确定出现基因融合的方法,包括:(a)使靶核酸序列与具有与基因组基因座(第一基因座)和Cas9蛋白互补的部分的引导RNA序列,和具有与相邻基因座基因座(第二基因座)和Cas9蛋白互补(按照参考)的部分的引导RNA序列接触,(b)确定是否没有检测到第一和第二基因座之间的共定位事件。在一些实施方式中,该方法应用于间变性淋巴瘤激酶,如ALK(参见图12)。在一些实施方式中,该方法应用于ROS1。ROS1是胰岛素受体家族的受体酪氨酸激酶。在一些实施方式中,该试验应用于非小细胞肺癌的

诊断。

[0160] 基因组重排

[0161] 在一些实施方式中,本发明包括检测基因组区域间重排的方法,包括(a)使基因组DNA样品接触多个引导RNA序列,所述引导RNA序列各自具有与基因组区域的特定子区域和Cas9蛋白互补的部分,其中各子区域的gRNA包括使其能够与其他子区域的gRNA区分的编码,和(b)对基因组DNA进行成像并对编码解码并比较编码与参考的顺序,其中在感兴趣基因组区域的大致长度上存在编码的共定位。在一些实施方式中,在尾处编码gRNA。在一些实施方式中,通过使尾的编码部分与解码器分子接触来进行对编码的解码,该解码器分子包含DNA或蛋白质探针。在一些实施方式中,除了确定基因组重排以外,也可区分特定基因组区段的不同等位基因。在一些实施方式中,针对不同等位基因以及针对不同基因组区段使用不同编码。在一些实施方式中,感兴趣区域是基因组的BRCA1和/或BRCA2区域。在一些实施方式中,感兴趣区域是MHC或HLA区域。在一些实施方式中,感兴趣区域是MMR基因MLH1-PMS2和MSH2-EPCAM-MSH6周围的区域,并且可用于诊断或分析遗传性非息肉病性结直肠癌(NHPPC)。在一些实施方式中,针对各基因组区段使用多种gRNA并且这多种引导RNA各自用相同编码标记。

[0162] 列举重复数

[0163] 在一些实施方式中,以线性串分析基因组DNA的链。在一些实施方式中,拉伸该DNA。在一些实施方式中,进行纳米孔/纳米间隙分析。在一些实施方式中,该方法用于对人染色体4和10上含3.3kb-D4Z4重复的基因座上的重复单元的数量进行列举。在一些实施方式中,使用对D4Z4区域的列举来诊断或分析患有面肌营养不良(FSHD)的患者。在一些实施方式中,该方法用于对端粒亚单元重复数量进行列举。在一些实施方式中,该方法用于对着丝粒重复数量进行列举。在一些实施方式中,该方法用于对大随体(major satellite)重复数量进行列举。在一些实施方式中,该方法用于对小随体(minor satellite)重复数量进行列举。

[0164] Cas9/引导RNA原位杂交

[0165] 根据某些方面,用于进行Cas9介导的原位杂交的方法包括以下步骤:使靶核酸序列与具有与感兴趣染色体或基因组区域和Cas9蛋白互补的部分的引导RNA序列接触,其中该引导RNA和Cas9蛋白共定位至靶核酸序列以形成复合物,其中感兴趣的染色体或基因组区域设置在流动池中,其中原位杂交的试剂和洗涤试剂流过感兴趣染色体或基因组区域之上。根据某些方面,使用选自下组的方法来检测复合物的位置,包括荧光、化学发光、电致发光、比色检测等。

[0166] 等位基因-特异性检测

[0167] 在一些实施方式中,本发明包括确定特定等位基因存在的方法,包括(a)使靶核酸序列与具有与待检测的等位基因和Cas9蛋白互补的种子部分(与PAM位点相邻的前4-7个核苷酸)的引导RNA序列接触,和(b)检测与核酸共定位的gRNA/Cas9的存在。

[0168] 结合试验

[0169] 根据某些方面,提供了检测特定序列的诊断方法,包括以下步骤:(a)使靶核酸序列与引导RNA序列接触,该引导RNA序列具有与靶核酸序列和Cas9蛋白互补的部分,(b)在表面上的位置上捕获复合物,(c)通过仅存在于复合物且不存在于靶核酸序列的标记物来检

测该位置上捕获的复合物。根据某些方面,提供了检测特定序列的诊断方法,包括以下步骤:(a)使靶核酸序列与引导RNA序列接触,所述引导RNA序列具有与靶核酸序列和Cas9蛋白互补的部分,(b)在表面上的位置上捕获复合物,(c)通过仅存在于复合物且不存在于cas9/gRNA的标记物来检测该位置上捕获的复合物。在一些实施方式中,作为侧流试验、浸渍条试验、纸微流体试验、点印迹试验、微阵列试验的部分进行该试验。在一些实施方式中,该试验是诊断试验。

[0170] 免疫组化和gRNA/Cas 9原位杂交

[0171] 在一些实施方式中,本发明包括在同一样品上组合免疫组化(IHC)和gRNA/cas9介导的原位杂交(ISH)的方法,包括以下步骤:(a)使靶染色质内的靶核酸序列与引导RNA序列接触,所述引导RNA序列具有与靶核酸序列和Cas9蛋白互补的部分,其中引导RNA和Cas9蛋白共定位至靶核酸序列以形成复合物,(b)使靶染色质与蛋白质结合试剂接触,和(c)检测引导RNA/cas9复合物和蛋白质结合试剂的比较位置。在一些实施方式中,蛋白质结合试剂是抗体。在一些实施方式中,蛋白质结合试剂是适体。在一些实施方式中,引导RNA/Cas9复合物和蛋白质结合试剂被差异标记。在一些实施方式中,IHC试剂和引导RNA/Cas9 ISH一起添加。在一些实施方式中,IHC试剂和引导RNA/Cas9 ISH连续,即一个接着一个添加。在一些实施方式中,因为引导RNA/Cas9 ISH不需要变性步骤,蛋白质和染色质结构保持完成用于待进行的IHC。在一些实施方式中,使用引导RNA/Cas9复合物来分离染色质的特定部分,并且使用分析方法来检测分离的染色质上存在的蛋白质。

[0172] 探测引导RNA尾

[0173] 在一些实施方式中,本发明包括检测靶核酸序列的方法,包括以下步骤:(a)使靶核酸序列与引导RNA序列接触,所述引导RNA序列具有与靶核酸序列和Cas9蛋白互补的部分,其中引导RNA和Cas9蛋白共定位至靶核酸序列以形成复合物,并且其中引导RNA包含3'尾序列并且所述尾与探针序列互补或可作为引物,和(b)检测复合物从而检测靶核酸序列。根据某些方面,尾包含与靠近gRNA结合位置的序列互补的序列。根据某些方面,尾包含与双链体的置换链互补的序列。根据某些方面,双链体的靶链之一被引导RNA隔离,留下其他链开口与其他试剂结合。示例性的试剂可包含单链结合蛋白、可标记的互补寡核苷酸、或与该链互补的尾的部分。根据一个方面,尾包含用于DNA PAINT的停靠点(docking site)或柄(handle)。在一些实施方式中,具有另一条链的Cas9/引导RNA复合物稳定化。在一些实施方式中,结合单链结合蛋白,如RPA或结合寡核苷酸或其类似物/模拟物至置换链,可使具有另一条链的Cas9/引导RNA复合物稳定。在一些实施方式中,稳定化效果是由于存在减少的与天然双链体的再扣紧(re-zipping)的竞争。

[0174] 具有改变的PAM特异性的Cas9

[0175] 在一些实施方式中,本发明包括标记或靶向核酸中位点的方法,包括以下步骤:(a)使核酸与标记的复合物在复合物结合至核酸的条件下接触,所述复合物包含经改变以结合至非规范PAM序列附近的Cas9蛋白和引导RNA,和(b)分析步骤(a)的产物。根据一个方面,在步骤(a)和步骤(b)之间任选地存在至少一个洗涤步骤,并且任选地提供辅助试剂以促进gRNA的结合。

[0176] 通常对于本文所述的实施方式,引导RNA和RNA可包括本领域技术人员已知的修饰的RNA核苷酸。通常对于本文所述的实施方式,引导RNA和RNA可包括RNA/DNA嵌合体或RNA/

PNA嵌合体。在一些实施方式中,在无细胞系统中制备引导RNA或RNA。制备gRNA或RNA的方法包括本领域技术人员已知的体外转录和自动化化学RNA合成方法。在一些实施方式中,gRNA或Cas蛋白或辅助蛋白在细胞系统中表达并且在无细胞系统中纯化。在一些实施方式中,gRNA或RNA在无细胞系统中与Cas蛋白或辅助蛋白复合。

[0177] 以下实施例是本发明的代表。这些实施例并不构成对本发明范围的限制,因为这些和其他等价实施方式将对于本发明、附图和所附权利要求而言是显而易见的。

[0178] 实施例1

[0179] 方案

[0180] gRNA的合成方案:

[0181] 如下进行PCR组装。

[0182] 制备包含以下的反应混合物:

[0183] ●12 μ L的Q5DNA聚合酶2x主混合物 (NEB)

[0184] ●3 μ L的10 μ M T7正向引物

[0185] ●3 μ L的10 μ M条码反向引物

[0186] ●3 μ L的10 μ M Sp.gRNA.sp1i60 (正向)

[0187] ●3 μ L的10 μ M gRNA.end (反向)

[0188] PCR装置中的循环条件:

[0189] 1.98 $^{\circ}$ C持续30秒

[0190] 2.98 $^{\circ}$ C持续10秒

[0191] 3.52 $^{\circ}$ C持续20秒

[0192] 4.72 $^{\circ}$ C持续15秒

[0193] 5.重复步骤2保持29个循环

[0194] 6.72 $^{\circ}$ C持续2分钟

[0195] 7.保持4 $^{\circ}$ C

[0196] 在离心柱 (Zymo) 上纯化DNA。通常产生约1 μ g的dsDNA模板。

[0197] 如下进行体外转录 (IVT):

[0198] 制备包含以下的反应混合物:

[0199] ●5.8 μ l无RNA酶的水

[0200] ●2.5 μ l AmpliScribe T7-Flash 10X反应缓冲液 (亿明达)

[0201] ●1.8 μ l 100mM ATP

[0202] ●1.8 μ l 100mM CTP (+2 μ L Cy3-dUTP)

[0203] ●1.8 μ l 100mM GTP

[0204] ●1.8 μ l 100mM UTP

[0205] ●2 μ l 100mM DTT

[0206] ●0.5 μ l RiboGuard RNA酶抑制物

[0207] ●5 μ l DNA模板

[0208] ●2.0 μ l AmpliScribe T7-Flash酶溶液

[0209] 也可用修饰的NTP合成引导RNA,其中进行以下修饰:

[0210] ●添加1:1摩尔比的待修饰的NTP和经修饰的-NTP (例如,0.9 μ l UTP和0.9 μ l UTP-

Cy3)。

[0211] 在PCR装置中在37℃下孵育2-16小时,然后保持4℃。在离心柱(Zymo)上纯化RNA。通常产生约100μg的RNA(即,条码化gRNA)。

[0212] Cas9-gRNA复合物组装方案:

[0213] 作为参考,1μg在尾上带单个条码的gRNA是25pmol的RNA。Cas9蛋白和gRNA一般以1:1摩尔比混合,以形成Cas9-gRNA复合物。可使用相同反应条件来复合野生型Cas9、切口酶Cas9和核酸酶无效或死Cas9。当使用野生型Cas9并且目标是为了防止切割时,可省去MgCl₂。用于表达的质粒、野生型Cas9、核酸酶无效Cas9和Cas9切口酶购自艾德基因公司(Addgene)。质粒可在合适宿主中表达并且可通过本领域已知的方法纯化蛋白,如在表达蛋白中使用His标签。可使用较高的Cas9与gRNA的比率(例如,3:1)来确保更多的gRNA与Cas9复合。一般通过在37℃下预孵育15分钟来形成活性复合物。反应缓冲液可变化。示例性的缓冲液包括20mM HEPES,100mM NaCl,5mM MgCl₂,0.1mM EDTA,pH 6.5;20mM Tris-HCl,100mM KCl,5mM MgCl₂,5%甘油,1mM DTT,pH 7.5;和1X PBS,5mM MgCl₂,0.5%吐温-20,pH 7.0。活性复合物可立即使用或在4℃下储存数周。

[0214] Cas9-gRNA荧光原位试验:

[0215] 根据该试验,样品已经固定至显微镜载玻片(或盖玻片)。玻片可在科普林氏缸(Coplin jar)内孵育,或在流动腔室或流动池中组装以进一步最小化反应提及并使过程自动化。反应以以下顺序进行并且玻片在科普林氏缸内孵育,除非另外说明(对于流动腔室,括号中提供信息)。用含0.5%吐温-20的1X PBS孵育2分钟(用2倍流动腔室体积洗涤,孵育30秒,在各洗涤之间)。用含0.5%曲通-100的1X PBS孵育5分钟(用2倍流动腔室体积洗涤,孵育2分钟,在各洗涤之间)。用0.1N HCl孵育5分钟(用2倍流动腔室体积洗涤,孵育2分钟,在各洗涤之间)。用含0.5%吐温-20的1X PBS孵育2分钟(用2倍流动腔室体积洗涤,孵育30秒,在各洗涤之间)。在Cas9缓冲液中孵育5分钟(用2倍流动腔室体积洗涤,孵育30秒,在各洗涤之间)。这段时间可用于Cas9-gRNA在37℃下在Cas9缓冲液中预复合,或使冷冻的复合物升温至37℃。一般而言,5μM的gRNA和5μM的Cas9在25μL体积/样品中复合在一起。在Cas9缓冲液中添加25μL Cas9-gRNA,并且在37℃下的湿度箱中孵育4小时。或者,用可移去的橡皮泥密封并在37℃下孵育。通过2次在37℃下在Cas9缓冲液中孵育5分钟(用4倍流动腔室体积洗涤,孵育30秒,在各洗涤之间)进行洗涤。用含0.5%吐温-20的1X PBS孵育2分钟(用2倍流动腔室体积洗涤,孵育30秒,在各洗涤之间)。可选的:如果需要,通过向各样品中添加含0.5%吐温-20的20μL 2X SSC中的1μM寡聚探针,并在湿度箱中孵育15分钟来进行探测。或者,在孵育之前用可移去的橡皮泥密封。通过2次含0.5%吐温-20的1X PBS中孵育2分钟(用4倍流动腔室体积洗涤,孵育30秒,在各洗涤之间)进行洗涤。装载10μL含DAPI的抗褪色显微介质并且用指甲油(例如,AntiFade或VectaShield)密封。玻片准备用于成像或可在黑暗中储存1周。在一些情况中,可省去HCL步骤。

[0216] 流动池

[0217] 可通过使用双面胶带或片(粘合剂研究公司(Adhesive Research)或3M)以制成屏障来制成流动池,其夹在含有感兴趣样品的盖玻片或载玻片与第二盖玻片或载玻片之间。可使用购自Ibidi的用于制造流动池的系统(粘性-载玻片VI^{0.4}或粘性-载玻片I Luer)。在此,上面设置细胞、染色体或DNA的载玻片或盖玻片接合至流动池的粘性部分以在基材顶上

产生流动池。试剂通过手动抽吸至入口区域并芯吸(例如,使用吸水纸)到出口区域流入流动池。或者,试剂通过自动试剂流动和交换系统以驱动流体和多向阀流入。这可通过使用注射泵,压力驱动流和抽气来完成。该自动系统与显微镜或成像设备整合,其中加载流动池。

[0218] DNA的分子梳理

[0219] 男性基因组DNA(普洛麦格公司(Promega)或诺瓦基公司(Novagen))或使用胶柱方法从细胞中提取的DNA被梳理(comb)到包被乙烯基硅烷(7-辛烯基三氯硅烷)的盖玻片上。这通过将盖玻片(例如,22x22mm)浸入含有覆盖大部分(例如,对于22x22盖玻片是1-1.5ml)盖玻片的0.5M MES缓冲溶液中DNA的槽中来实现,使DNA末端结合至表面涂层(一般为1分钟至10分钟),然后从槽中撤回盖玻片。可调整槽中DNA的浓度以得到所需密度的梳理的DNA。例如,高至0.5ng/u1的浓度可得到合理的密度,其中可分辨单个拉伸的DNA分子。然后以恒定速度(例如,300 μ m/s)从DNA溶液中撤回盖玻片,使DNA由于撤回盖玻片时弯液面的力而被拉伸。任选地,DNA然后使用约10-20焦/cm²的紫外辐射能交联到盖玻片上。任选地,如上所述在盖玻片上形成流动池。可通过在梳理过程期间或之后用一种或多种插入染料,如YOYO-1染色来使经梳理的DNA可视化。通常使用5:1至10:1的DNA碱基对与YOYO-1染色比率。

[0220] Cas9/gRNA结合至在盖玻片上预拉伸的DNA

[0221] 夹有经梳理的基因组DNA以制成流动池的盖玻片首先通过用PBS吐温和PBS洗涤来水合。任选地,用Blockaid(生命技术公司)来封闭基材。任选地,用Cas9反应缓冲液来洗涤流动池。Cas9-gRNA在37℃下预复合并且然后添加至流动池并静置孵育30分钟至1小时。通过用缓冲液如反应缓冲液和PBS吐温20和PBS洗涤来终止反应。

[0222] 图6a和6b使用的引导RNA是

[0223] 着丝粒16gRNA,具有尾:

[0224] GACGCCUUCGUUGGAAACGUUUUAGAGCUAGAAAUAGCAAGUUAUUUUUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUTCCTCTACCACCTACATCACTTATACATCTA (SEQ ID NO:8)

[0225] 荧光标记的探针与尾杂交

[0226] 确定流动池保持水合并且在Cas9/gRNA复合物已经形成之后,通过在以下杂交缓冲液中向尾添加DNA探针来对复合物进行成像:200ul甲酰胺,20ul sds,120ul blockaid,40ul of 20xssc,和20ul水。

[0227] 用于探测尾的序列是

[0228] TAGATGTATAA GTGATGTAGGTGGTAGAGGA (SEQ ID NO:9)

[0229] 这可在4℃下静置过夜。根据针对探针-尾杂交体的热稳定性,也可使用其他杂交温度和杂交缓冲液组合物。

[0230] 可在一个或两个末端处用荧光标记物如Atto 647N、Atto 655、Alexa 647进行标记。当使用合适的荧光标记物(例如,Atto 655)时,可通过STORM方法进行超高分辨率成像。

[0231] 探针-尾杂交和成像

[0232] 向成像溶液中添加杂交缓冲液中的DNA PAINT成像探针以得到超高分辨率所需的随机打开和关闭图案。与尾上的序列互补或与结合尾的探针上的序列互补并且在图6A和图6B和相似实验中使用的PAINT序列是

[0233] /5Alex647N/TAGATGTATAAAAAATTTAATAAGGT/3AlexF647N (SEQ ID NO:10)

[0234] 或

[0235] /5ATTO647NN/TAGATGTATAAAAAA/3ATTO647NN/(SEQ ID NO:11)

[0236] 成像使用Nikon Ti-E倒置显微镜上的633、640或647nm激光光线和适于Cy3的过滤块在全内反射(TIRF)模式下完成并且在减薄背照式Andor Ixon X3 EMCCD相机上捕获。

[0237] 超高分辨率成像

[0238] 可通过使用以下成像器链进行DNA PAINT过程来进行如图6b所示的超高分辨率成像:P9成像器:ACCTTATTA。这结合至含有P9柄(停靠序列)的尾或已经结合至于含有P9柄的尾的探针:TAATAAGGT。用于对DNA PAINT进行成像的合适缓冲剂是5mM Tris pH 8,10mM MgCl₂,1mM EDTA pH 8和0.05%吐温-20。可使用Nikon NIS Elements软件持续15分钟至30分钟拍摄影像,并且使用以LabVIEW编写的DNA PAINT图像分析代码来构建超高分辨率图像(参见Jungmann等,Nano Lett.,2010,10(11),第4756-4761页)。

[0239] 在溶液中结合Cas9/gRNA至基因组DNA

[0240] 在缓冲溶液存在下向DNA添加Cas9-gRNA复合物组装(在37℃下预孵育10分钟)并在37℃下孵育1小时。任选地纯化结合Cas9/gRNA的DNA。然后,如果需要DNA拉伸,反应在0.5M MES溶液中稀释并且其上装饰gRNA/cas9复合物的DNA被梳理到乙烯基硅烷盖玻片表面上。这种方法用于图6C。

[0241] 复合的DNA的纯化

[0242] 根据使用的引导RNA的尺寸,可使用多种不同的纯化方法来分离靶DNA/gRNA/Cas9复合物,包括尺寸排阻、亲和纯化(例如,使用结合链霉亲和素珠的可切割接头或脱硫生物素)、或透析膜。

[0243] Cas9切口和靶向测序

[0244] Cas9-gRNA在靶DNA上的特定位置处结合dsDNA模板。靶DNA可以在天然基因组中,并且反应在体内进行。靶DNA可在细胞中或生物体固定至表面并且原位进行反应。在这些情况下,优选保持基因组的空间位置。此外,可提取靶DNA并且离体进行反应。提取的靶DNA也可固定在表面上或流体系统(例如,纳米通道)中并且离体进行反应。

[0245] 使用突变体切口酶(例如,携带D10A或H840A突变体的Cas9)来进行gRNA/Cas9介导的切口反应。参见图9。Cas9的RuvC结构域可被D10A突变灭活并且HNH结构域可被H840A突变灭活。

[0246] 可使用2种切口机制之一,一个在与gRNA碱基配对的顶部链上,另一个在底部链,即置换链上。D10A位置的突变能够切口顶部链而不是底部链。H840A位置的突变能够切口底部链而不是顶部链。通过Cas9(D10A)切口酶的切割出现在原型间隔子邻近基序(PAM)上游3bp处。Cas9(H840A)切割出现在底部链上的互补位置处。与识别在基因组中出现许多次的特定短序列的切口内切核酸酶相比,gRNA在基因组中可能只出现一次的较长识别序列处切口。这可用于靶向基因组中的特定独特序列。可通过以下步骤来进行靶向的测序:选择D10A切口酶或H840A切口酶;选择待测序的区域附近的引导RNA;用选择的gRNA和Cas9突变体进行gRNA/Cas9反应以形成复合物;任选地从靶DNA中去除Cas9/gRNA复合物;添加合适的缓冲液中的聚合酶和荧光标记的可逆终止子核苷酸以进行纳入、洗涤和切割的测序循环。这使得能够对靶向的位置进行测序。

[0247] 在一些实施方式中,可从靶DNA和样品中完全去除Cas9-gRNA复合物。去除试剂可以是去污剂(例如,十二烷基硫酸钠)、有机化合物(例如,尿素、盐酸胍/异硫氰酸胍)、酰胺

(例如,甲酰胺)、蛋白水解酶(例如,蛋白酶)、物理性质(例如,温度)或其组合。

[0248] 如果使用Cas9 (D10A),测序通过PAM基因座并行进至该基因座的下游,对含PAM的链进行测序。如果使用Cas9 (H840A),测序在PAM基因座的上游方向上进行,并且对与NGG PAM序列互补的链进行测序。为了避免测序通过引导RNA序列,可选择D10A突变体。可在感兴趣区域上选择多种引导物。

[0249] 如果进行通过合成的逐步测序,各纳入的核苷酸含有可逆终止子,在去除终止子和标记物并且在各靶向的位置处重复下一个碱基的循环之前检测到在各靶向的位置处纳入单核苷酸的结果(优选通过使用TIRF照射和CCD或CMOS检测器)。

[0250] 各种亿明达SBS试剂盒(例如,SBS试剂盒2)可用于测序,按以下顺序进行试剂添加和成像:通用测序缓冲液;纳入主混合物;通用测序缓冲液;成像靶向的基因座;通用扫描混合物;切割试剂主混合物;切割洗涤混合物。亿明达试剂盒的详细内容可从万维网 support.illumina.com/downloads/hiseq-rapid-sbs-kit-v2-reagent-prep-guide-15058772.html上下载到。通过使用针对核苷酸上的四种染料中的两种的532nm激光和针对染料中另外两种的660nm激光来进行成像。由各激光激发的两种染料各自通过使用特定发射滤器和设计成确定各染料的标识的算法来区分。

[0251] 可使用的多种不同亿明达测序设备之一包括Genome Analyzer IIx。可使用与亿明达流动池固定器以及入口和出口端口相容的流动池迹线(footprint)。或者,包含倒置显微镜,具有高数值孔径物镜,激光,CCD相机,荧光团选择性滤器和基于注射泵或压力驱动的试剂交换系统和加热平台的自制系统。自制系统可适用于除亿明达提供的以外的其他核苷酸/染料组合。

[0252] 如果以实时反应(例如,PacBio或Starlight测序)进行测序,在末端磷酸酯处标记核苷酸,末端磷酸酯一旦掺入了核苷酸就是天然离去基团。在CCD或CMOS相机上连续监测该反应。

[0253] 对拉伸的DNA双链体的靶向测序

[0254] 可从细胞中提取基因组DNA并且保持长的长度(例如,通过在胶柱中进行提取),以保留序列的线性位置。这在确定基因组中序列的组织上有重要应用。基因组组织中的结构变异(SV)可能引起疾病,如本文所述的ALK、BRCA、FSHM和HER2示例所证明。然而,某些SV,如白血病中的Bcr-ABL易位可被药物如Gleevac靶向,因此确定白血病患者是否具有可由这种药物靶向的易位是重要的。

[0255] 优选地,使用例如分子梳理来在拉伸的DNA上进行测序,使得能够观察到靶向的序列的基因组中的线性组织。可通过使用切口突变体Cas9和设计在感兴趣位置处结合的gRNA,进行切口反应,并且然后从该切口处延伸来选择靶向的区域。可在梳理DNA之前在溶液中进行切口反应。或者,可在已经梳理DNA之后使用梳理的DNA顶部的流动池来进行切口反应。测序反应通过用纳入缓冲液水合并预调节DNA来初步进行,然后通过使测序试剂流入流动池。结果是不仅获得了靶标区域的序列,也获得其在基因组中的位置。这何时重要的示例是:当基于gRNA的靶向的测序涉及基因组中作为易位热点的序列,但不知道其易位至基因组中的何处和其与什么序列融合时。

[0256] 原位靶向测序

[0257] 本文所述的方法涉及确定细胞或核内基因组的的部分的空间位置和这种空间位置

如何影响基因调控和基因组功能。可使用对这些序列有特异性的gRNA来靶向特定基因组序列在细胞中的空间位置。在固定的细胞中的基于泥足DNA上产生gRNA/Cas9介导的切口并且在基因组DNA上进行原位荧光基测序。如Lee等,2015: (Nat Protoc.2015,10 (3) :442-58) 所述,细胞在玻璃底平皿上生长或者通过本领域已知的方法来装载组织切片。固定细胞(例如,使用PBS中10%甲醛在25℃下持续15分钟,或100%甲醇在-20℃下持续20分钟)。在含曲通X-100的PBS和仅PBS中进行洗涤。任选地添加尿素以从切口的DNA中去除gRNA/Cas9。任选地,通过使用RNA酶来去除RNA。

[0258] 可直接在细胞内对单分子的基因组DNA进行测序。使用聚合酶在切口位点处纳入核苷酸。优选地,核苷酸含有带荧光团的荧光标记物,其提供高量子产率,如Cy3B或多重标记的核苷酸。可使用本领域已知的各种严谨性的洗涤来减少来自未纳入的核苷酸的信号。通过单分子成像方法来进行成像,但是,因为仅部分细胞接触表面,TIRF成像可能仅获得纳入信号的一小部分。可使用共聚焦显微术或光-层照显微术来获得在细胞和组织内的3D空间上分布的信号。使用多光子或双光子激光显微术来获得信号,尤其是当其深埋时。这类方法也有效降低背景信号。

[0259] 在PBS洗涤之后,用于测序反应的缓冲液用于在添加测序反应混合物之前调整样品。根据使用的引物、成像设置以及是否有温度控制,应用SOLID化学、CycLic化学、SBL化学或通过合成化学的测序如亿明达测序或Lasergen测序来纳入核苷酸以及切割标记物和终止子。在各纳入步骤之后拍摄图像。处理图像以产生间断信号。记录来自各连续循环的图像。在各图像上进行碱基判定并且堆叠碱基判定以在各灶点处提供序列读数。可通过在读数积累过程中仅包含序列来滤去非特异性信号,这在许多已记录的图像上一致。使用Bowtie 1.0或其他比对算法来将读数以参考比对,并且进一步滤去不感兴趣的噪音或脱靶序列。观察测序读数的不同灶点以提供细胞内靶向序列的空间定位的信息。测序,而不是仅标记靶向的位置,具有显示感兴趣区域内序列变体及其空间位置的潜力。

[0260] gRNA/Cas9复合沿核酸聚合物的纳米孔分析

[0261] 制造SiN膜,并且用透射电子显微镜(TEM)来钻出20nm直径纳米孔(Janssen, X.J.A.;Jonsson,M.P.;Plesa,C.;Soni,G.V.;Dekker C.;Dekker,N.H.Nanotechnology 2012,23 (47),475302)。然后用聚二甲基硅氧烷(PDMS)涂布膜来降低电容并改善信噪比。膜装载在流动池中,含有由膜隔开的顶部和底部储器,之后用1M KCl,10mM Tris,1mM EDTA, pH 8填充2个储器。向顶部储器添加核酸聚合物并且在顶部(-ve)和底部(+ve)储器之间施加电压,使得能够电泳驱动核酸聚合物以基本线性的方式通过孔。用包括Axopatch 200B放大器和Digidata 1322A DAQ数字转换器的离子通道记录系统来记录电流。使用Transalyzer Matlab包来分析记录(参见Plesa,C.;Dekker,C.Nanotechnology 2015,26 (8),084003)。针对核酸聚合物进入纳米孔时增加的电流堵塞,和然后gRNA/Cas9结合的区域通过孔的狭窄结构时阻断中的进一步间断升高来分析数据。

[0262] 实施例2

[0263] 产生Cas9-gRNA复合物用于Her2/ErbB2诊断

[0264] Her2(也由国际人类基因组组织命名委员会称为ErbB2)基因序列可获自不同的在线储存库,如NCBI万维网ncbi.nlm.nih.gov/gene/2064。

[0265] 分析Her2序列以沿着dsDNA发现PAM基序。优选的基序是5' -

GGNNNNNNNNNNNNNNNNNGG-3' (SEQ ID NO:12)。5' GG通过T7RNA聚合酶提供了最优RNA合成。3' NGG是由Cas9识别的PAM序列。中间的17个碱基N用于靶向的序列并且提供高特异性。优选外显子组,因为该方法可区分基因变体(即,同种型)。

[0266] 可在因特网上自由获得多个程序以寻找靶序列。图10显示了使用CHOPCHOP时提供的输出的图(参见Tessa G.Montague;Jose M.Cruz;James A.Gagnon;George M.Church;Eivind Valen.(2014).CHOPCHOP:用于基因组编辑的CRISPR/Cas9和TALEN网页工具(CHOPCHOP:a CRISPR/Cas9 and TALEN web tool for genome editing).Nucleic Acids Res.42.W401-W407)。

[0267] 表1中显示了Her2的靶序列的部分列表。该表也包含关于人类基因组上靶标的位置、外显子编号、DNA链(即,正义(+)对比反义(-))、和基因组中其他地方的二级靶标的潜在数量(即,错配,MM)的信息。

[0268]

靶序列	基因组位置	外显子	链	MM
1 GGGCGAGGAGGAGCCCCAGCGG	chr17:39700259	1	-	0 (SEQ ID NO:13)
2 GGTGGCGGAGCATGTCCAGGTGG	chr17:39707039	2	-	0 (SEQ ID NO:14)
3 GGTGGGTCTCGGGACTGGCAGGG	chr17:39707021	2	-	0 (SEQ ID NO:15)
4 GGCAGCCCTGGTAGAGGTGGCGG	chr17:39707054	2	-	0 (SEQ ID NO:16)
5 GGAGGCCCTGTGACAGGGGTGG	chr17:39708472	3	-	0 (SEQ ID NO:17)
6 GGGCCTCCCCAGGAGGCCTGCGG	chr17:39708487	3	+	0 (SEQ ID NO:18)
7 GGCCTCCCCAGGAGGCCTGCGGG	chr17:39708488	3	+	0 (SEQ ID NO:19)
8 GGTGGCTGTGCCCCGCTGCAAGGG	chr17:39710106	6	+	0 (SEQ ID NO:20)
9 GGGCAGTGGCCCCTTGACAGCGGG	chr17:39710116	6	-	0 (SEQ ID NO:21)

[0269]

					NO:21)
				0 (SEQ ID	
10	GGGACAGGCAGTCACACAGCTGG	chr17:39710458	7	-	NO:22)
				0 (SEQ ID	
11	GGTTGTGCAGGGGGCAGACGAGG	chr17:39711962	8	-	NO:23)
				0 (SEQ ID	
12	GGGCATGGAGCACTTGCGAGAGG	chr17:39712335	9	+	NO:24)
				0 (SEQ ID	
13	GGAGCACTTGCGAGAGGTGAGGG	chr17:39712341	9	+	NO:25)
				0 (SEQ ID	
14	GGAGCTGCTCTGGCTGGAGCGGG	chr17:39715307	10	-	NO:26)
				0 (SEQ ID	
15	GGAAGACGCTGAGGTCAGGCAGG	chr17:39715477	11	-	NO:27)
				0 (SEQ ID	
16	GGTGGGTGTTATGGTGGATGAGG	chr17:39715824	12	-	NO:28)
				0 (SEQ ID	
17	GGCTGGGGCTGCGCTCACTGAGG	chr17:39715781	12	+	NO:29)
				0 (SEQ ID	
18	GGTGCGGGTTCCGAAAGAGCTGG	chr17:39715875	12	-	NO:30)
				0 (SEQ ID	
19	GGCTGGGCATCAGCTGGCTGGGG	chr17:39715766	12	+	NO:31)
				0 (SEQ ID	
20	GGGCTGGGCATCAGCTGGCTGGG	chr17:39715765	12	+	NO:32)
				0 (SEQ ID	
21	GGAGGAATGCCGAGTACTGCAGG	chr17:39716410	13	+	NO:33)
				0 (SEQ ID	
22	GGCATTCTCCACGCACTCCTGG	chr17:39716398	13	-	NO:34)
				0 (SEQ ID	
23	GGCTGACACTCAGGGTGGCACGG	chr17:39716552	14	-	NO:35)

[0270]

					0 (SEQ ID
24	GGTCAGGTTTCACACCGCTGGGG	chr17:39717382	15	-	NO:36)
					0 (SEQ ID
25	GGGCAGCGGGCCACGCAGAAGGG	chr17:39717362	15	-	NO:37)
					0 (SEQ ID
26	GGGGCAGCGGGCCACGCAGAAGG	chr17:39717363	15	-	NO:38)
					0 (SEQ ID
27	GGTTGGCATTCTGCTGGTCGTGG	chr17:39723346	17	+	NO:39)
					0 (SEQ ID
28	GGAGAATGTGAAAATTCCAGTGG	chr17:39723932	19	+	NO:40)
					0 (SEQ ID
29	GGGCATCTGCCTGACATCCACGG	chr17:39724776	20	+	NO:41)
					0 (SEQ ID
30	GGATGTGCGGCTCGTACACAGGG	chr17:39725066	21	+	NO:42)
					0 (SEQ ID
31	GGTGAACCGCCGGCGGAGAATGG	chr17:39725355	22	-	NO:43)
					0 (SEQ ID
32	GGTCAGGGATCTCCCGGGCTGGG	chr17:39725759	23	-	NO:44)
					0 (SEQ ID
33	GGATGACCACAAAGCGCTGGGGG	chr17:39726635	24	-	NO:45)
					0 (SEQ ID
34	GGATGATTGACTCTGAATGTCGG	chr17:39726565	24	+	NO:46)
					0 (SEQ ID
35	GGTGTCTGAATTCTCCCGCATGG	chr17:39726605	24	+	NO:47)
					0 (SEQ ID
36	GGACAGAAGAAGCCCTGCTGGGG	chr17:39726920	25	-	NO:48)
					0 (SEQ ID
37	GGGGGACCTGGTGGATGCTGAGG	chr17:39726886	25	+	NO:49)
38	GGACGATGACATGGGGGACCTGG	chr17:39726874	25	+	0 (SEQ ID

[0271]

						NO:50)
					0 (SEQ ID	
39	GGTGGATGCTGAGGAGTATCTGG	chr17:39726895	25	+	NO:51)	
					0 (SEQ ID	
40	GGCACCGCAGCTCATCTACCAGG	chr17:39726981	25	+	NO:52)	
					0 (SEQ ID	
41	GGAGTATCTGGTACCCCAGCAGG	chr17:39726907	25	+	NO:53)	
					0 (SEQ ID	
42	GGACCATGCCCCCAGCGCCCGGG	chr17:39726952	25	-	NO:54)	
					0 (SEQ ID	
43	GGGTGCCAGTGGAGACCTGGGGG	chr17:39727344	26	-	NO:55)	
					0 (SEQ ID	
44	GGCGGTGGGGACCTGACACTAGG	chr17:39727298	26	+	NO:56)	
					0 (SEQ ID	
45	GGGGAGGCTTTGCAGCCCCCTTGG	chr17:39727419	26	-	NO:57)	
					0 (SEQ ID	
46	GGTCCTGGTCCCAGTAATAGAGG	chr17:39727933	27	-	NO:58)	
					0 (SEQ ID	
47	GGGACCAGGACCCACCAGAGCGG	chr17:39727944	27	+	NO:59)	
					0 (SEQ ID	
48	GGTGTCCCTTTGAAGGTGCTGGG	chr17:39727976	27	-	NO:60)	
					0 (SEQ ID	
49	GGGGGCTGGGGCCGAACATCTGG	chr17:39727703	27	-	NO:61)	
					0 (SEQ ID	
50	GGCCCAAGACTCTCTCCCCAGGG	chr17:39727782	27	+	NO:62)	

[0272] 表1.

[0273] 应注意到可易于靶向基因以外的区域。靶向特定区域,如外显子和内含子,其可小于1kb至30kb,数十或数百个千碱基,或几个兆碱基的完整基因。

[0274] 序列靶标用作模板以制备gRNA。这包括添加针对gRNA支架的各靶序列。下文提供了组装模型的图,并且该组装使用采用高保真DNA聚合酶(提供熔融温度(T_m))的PCR进行。简言之,从市售供应商获得或定制合成寡核苷酸。存在2个通用寡核苷酸:Fwd-T7-gRNA,也包括T7RNA聚合酶识别基序的部分的正向PCR引物;和gRNA.split60,通用gRNA支架。另外,存在2个可变寡核苷酸:Sp.gRNA.split60,针对感兴趣靶标有特异性的序列;和Rev-B1-gRNA.18,也包括用于多重链监测的条码化的柄的反向PCR引物。

[0275] 这种设计使得能够低成本合成并组装gRNA,同时通过亚磷酰胺化学合成寡核苷酸的最优序列准确性,并且提供各种类型的条码柄,其是标记的实体可停靠的编码。可以不同规模合成DNA模板,包括在寡核苷酸合成以上,或市售获得。可通过PCR扩增和再扩增模板,并且保持模板,其比从头合成成本更低。

[0276] 图11所示的示例性PCR组装可耗时不到1小时。在PCR之后,通过向模板DNA添加T7RNA聚合酶混合物由体外转录 (IVT) 来合成gRNA。在这种情况下,较长的反应将提供较多的gRNA。可在半个小时内制备用于有限数量用途的快速产生的Cas9-gRNA试剂盒。可用进行16小时的反应产生大量。在IVT之后,通过DNA酶I降解DNA模板并且通过使用市售柱纯化试剂盒或通过乙酸钠存在下的乙醇沉淀来纯化RNA。

[0277] 然后,在镁离子存在下,条码化的gRNA与Cas9蛋白以1:1的比例复合以形成Cas9-gRNA复合物。然后可使用试剂。也可在这一阶段添加检测探针,其是结合至条码以促进条码解码的标记的寡核苷酸,这使得能够在下游进行更快的杂交检测方法。基于预期的信噪比来作出决定。例如,如果预期低噪音(例如,在显微镜载玻片表面上铺展的染色体上监测Her2),则可向样品中简单添加Cas9-gRNA-标记物试剂盒,使得在37℃下进行反应。洗涤未反应的复合物并且可对样品进行成像。试剂盒是探测基因座所需的所有Cas9-gRNA复合物的集合。Cas9-gRNA-标记物试剂盒是所有预标记的Cas9-gRNA的集合。

[0278] 编码Cas9-gRNA用于多重检测的示例。

[0279] 可向gRNA添加一个或多个编码或条码柄(探针的停靠点)。原理是确定与DNA样品结合的Cas9-gRNA复合物的种类。这通过杂交可检测的对gRNA柄有特异性的“解码器”探针来进行。gRNA柄可含有一个或多个条码,其可以各种方式排列,如连续堆叠或互相部分重叠。后者能够降低多重条码的总体长度,其增加了针对给定的gRNA柄长度的多重化容量。

[0280] 多重化容量与条码的数量成比例并且等于条码组合的数量,如 $nCr = \frac{n!}{r!(n-r)!}$,其中n是条码的数量,r是读取条码时的事件数量。通过例如2个条码B1和B2,对于总共3个编码(B1B1,B1B2,B2B2)而言,获得C=1并且P=2。用3个条码,获得10个编码。可用至少5个条码生成126个编码,而10个条码提供总共92378个编码,足够测量外显子组。

[0281] 本文提供的编码方法针对以下。每个基因座可分配一个代码,其使得能够在单一试验中鉴定多个靶标。对于给定的基因座,每个同种型可分配一个编码,其使得能够鉴定变异。对于给定的基因座可分配2个编码,其提供给定靶标的信号种类的共定位,使得靶标种类的置信度增加。这在具有高背景的样品的情况中有利,在这种情况下,试验仅在相同基因座上检测到2种信号时是阳性的。对于给定的基因座,可分配超过2个编码,其也允许共定位,但提供改善鉴定并改善分辨率的特征。例如,可由2组Cas9-gRNA来识别2个基因座,第一个用条码B1B2和B1B3编码,第二个用条码B1B3和B1B4编码。共有的B1将提供在基因座上可易于检测到的信号。可更特异性地聚焦于信号位置并且可检测基因座1的剩余条码B2B3,和基因座2的B3B4。

[0282] 表2提供了5个条码的列表,其从在人类基因座中出现不超过一次的12个核苷酸长的序列的列表生成:

[0283]	条码	序列	
	B1	CGTCGATTACCA	(SEQ ID NO:63)
	B2	CCATACTCGTCG	(SEQ ID NO:64)
[0284]	B3	TCGATAGTACGT	(SEQ ID NO:65)
	B4	CGTACTGAACGA	(SEQ ID NO:66)
	B5	CGAAGTACTACG	(SEQ ID NO:67)
	B6	TCGTTACGACCA	(SEQ ID NO:68)
	B7	CCAGACGAATCG	(SEQ ID NO:69)
	B8	TCGAATAGTCGT	(SEQ ID NO:70)
	B9	CGTAACTAACGA	(SEQ ID NO:71)
	B10	CGAACAATCGTA	(SEQ ID NO:72)

[0285] 表2.

[0286] 12个碱基的条码尺寸足够允许在一定长度和解链温度范围上设计一定范围的探针。例如,10-12聚体允许在某些离子浓度(例如,5mM $MgCl_2$)下更稳定的相互作用,而仍然可使用变性剂(例如,50%甲酰胺)去除。8-9聚体探针允许使用DNA-PAINT技术来进行超高分辨率成像,该技术描述于Jungmann R, **Avendaño** MS, Woehrstein JB, Dai M, Shih WM, Yin P.使用DNA-PAINT和Exchange-PAINT的多重化3D细胞超高分辨率成像(Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT). Nat Methods. 2014年3月;11(3):313-8。

[0287] 条码特异性探针可以是许多类型并且可适应试验。例如,探针可以是与序列互补的寡核苷酸。又例如,探针可由环形ssDNA模板组成,如锁式探针,探针的一个区域与条码杂

交并且一个或多个区域用于第二组探针的辅助杂交。然后可使用滚环扩增来扩增环形探针,其增加了信号强度。又例如,杂交探针是线性的并且含有二级探针的位点,二级探针也含有二级或三级探针的位点,以此类推,使得高支化探针能自组装,其将增加信号强度。虽然大部分本文所述的探针是荧光标记的,在另一个示例中,这些探针接合至允许发色检测的分子,例如,给出仅纳米颗粒的高密度,检测可能是比色的并且在定位至核的特定区域时仍然可见。

[0288] 图7中提供了区位RGB编码的示例,按照表3中所示列出的颜色方案条码化引导RNA尾,表3中列出了gRNA至荧光条码BRCA1重排。R是红色,G是绿色,B是蓝色。以形成预期的连续颜色图案的方式隔离条码。根据该方案,各条码与颜色相关,标识需要生成的3个条码序列,其然后按照颜色编码组合。例如,如果gRNA编码是R,则仅需要向gRNA尾添加条码R,而如果gRNA编码是RGB,则需要在gRNA尾处合并所有3个条码。通过添加其相应荧光探针来检测条码,其显示实际连续颜色图案。任何未预期的图案将被鉴定为基因组重排。

[0289]

名称	位置	区位	折纸代码		
		编码	点 1	点 2	点 3
gRNA.BRCA1.1	5	R	RG	RGB	RG
gRNA.BRCA1.2	177	R	RG	RGB	RB
gRNA.BRCA1.3	317	R	RG	RGB	BG
gRNA.BRCA1.4	460	R	RG	RGB	R
gRNA.BRCA1.5	585	R	RG	RGB	G
gRNA.BRCA1.6	610	R	RG	RGB	B
gRNA.BRCA1.7	730	R	RG	RB	RGB
gRNA.BRCA1.8	738	R	RG	RB	RG
gRNA.BRCA1.9	792	R	RG	RB	BG
gRNA.BRCA1.10	827	R	RG	RB	R
gRNA.BRCA1.11	881	R	RG	RB	G
gRNA.BRCA1.12	906	R	RG	RB	B
gRNA.BRCA1.13	960	R	RG	BG	RGB
gRNA.BRCA1.14	1,048	R	RG	BG	RG
gRNA.BRCA1.15	1,220	R	RG	BG	RB

[0290]

gRNA.BRCA1.16	1,310	R	RG	BG	R
gRNA.BRCA1.17	1,337	R	RG	BG	G
gRNA.BRCA1.18	1,403	R	RG	BG	B
gRNA.BRCA1.19	1,410	R	RG	R	RGB
gRNA.BRCA1.20	1,446	R	RG	R	RG
gRNA.BRCA1.21	1,459	R	RG	R	RB
gRNA.BRCA1.22	1,509	R	RG	R	BG
gRNA.BRCA1.23	1,561	R	RG	R	G
gRNA.BRCA1.24	1,647	R	RG	R	B
gRNA.BRCA1.25	1,667	R	RG	G	RGB
gRNA.BRCA1.26	1,675	R	RG	G	RG
gRNA.BRCA1.27	1,813	R	RG	G	RB
gRNA.BRCA1.28	2,118	R	RG	G	BG
gRNA.BRCA1.29	2,119	R	RG	G	R
gRNA.BRCA1.30	2,120	R	RG	B	RGB
gRNA.BRCA1.31	2,212	R	RG	B	RG
gRNA.BRCA1.32	2,304	R	RG	B	RB
gRNA.BRCA1.33	2,372	R	RG	B	BG
gRNA.BRCA1.34	2,533	R	RB	RGB	RG
gRNA.BRCA1.35	2,589	R	RB	RGB	RB
gRNA.BRCA1.36	2,609	R	RB	RGB	BG
gRNA.BRCA1.37	2,662	R	RB	RGB	R
gRNA.BRCA1.38	2,800	R	RB	RGB	G
gRNA.BRCA1.39	2,844	R	RB	RGB	B
gRNA.BRCA1.40	2,917	R	RB	RG	RGB
gRNA.BRCA1.41	2,951	R	RB	RG	RB
gRNA.BRCA1.42	3,012	R	RB	RG	BG
gRNA.BRCA1.43	3,110	R	RB	RG	R
gRNA.BRCA1.44	3,111	R	RB	RG	G

[0291]

gRNA.BRCA1.45	3,112	R	RB	RG	B
gRNA.BRCA1.46	3,132	R	RB	BG	RGB
gRNA.BRCA1.47	3,333	R	RB	BG	RG
gRNA.BRCA1.48	3,362	R	RB	BG	RB
gRNA.BRCA1.49	3,363	R	RB	BG	R
gRNA.BRCA1.50	3,399	R	RB	BG	G
gRNA.BRCA1.51	3,877	R	RB	BG	B
gRNA.BRCA1.52	3,899	R	RB	R	RGB
gRNA.BRCA1.53	4,168	R	RB	R	RG
gRNA.BRCA1.54	4,174	R	RB	R	RB
gRNA.BRCA1.55	4,200	R	RB	R	BG
gRNA.BRCA1.56	4,401	R	RB	R	G
gRNA.BRCA1.57	4,408	R	RB	R	B
gRNA.BRCA1.58	4,422	R	RB	G	RGB
gRNA.BRCA1.59	4,453	R	RB	G	RG
gRNA.BRCA1.60	4,488	R	RB	G	RB
gRNA.BRCA1.61	4,489	R	RB	G	BG
gRNA.BRCA1.62	4,528	R	RB	G	R
gRNA.BRCA1.63	4,691	R	RB	G	B
gRNA.BRCA1.64	4,752	R	RB	B	RGB
gRNA.BRCA1.65	4,777	R	RB	B	RG
gRNA.BRCA1.66	4,831	R	RB	B	RB
gRNA.BRCA1.67	4,919	R	RB	B	BG
gRNA.BRCA1.68	4,955	R	RB	B	R
gRNA.BRCA1.69	5,016	R	RB	B	G
gRNA.BRCA1.70	5,123	R	BG	RGB	RG
gRNA.BRCA1.71	5,146	R	BG	RGB	RB
gRNA.BRCA1.72	5,191	R	BG	RGB	BG
gRNA.BRCA1.73	5,282	R	BG	RGB	R

[0292]

gRNA.BRCA1.74	5,448	R	BG	RGB	G
gRNA.BRCA1.75	5,473	R	BG	RGB	B
gRNA.BRCA1.76	5,675	R	BG	RG	RGB
gRNA.BRCA1.77	5,774	R	BG	RG	RB
gRNA.BRCA1.78	5,808	R	BG	RG	BG
gRNA.BRCA1.79	5,843	R	BG	RG	R
gRNA.BRCA1.80	5,958	R	BG	RG	G
gRNA.BRCA1.81	6,492	R	BG	RG	B
gRNA.BRCA1.82	6,493	R	BG	RB	RGB
gRNA.BRCA1.83	6,642	R	BG	RB	RG
gRNA.BRCA1.84	6,903	R	BG	RB	BG
gRNA.BRCA1.85	6,996	R	BG	RB	R
gRNA.BRCA1.86	7,027	R	BG	RB	G
gRNA.BRCA1.87	7,347	R	BG	RB	B
gRNA.BRCA1.88	7,431	R	BG	R	RGB
gRNA.BRCA1.89	7,444	R	BG	R	RG
gRNA.BRCA1.90	7,453	R	BG	R	RB
gRNA.BRCA1.91	7,474	R	BG	R	BG
gRNA.BRCA1.92	7,479	R	BG	R	G
gRNA.BRCA1.93	7,489	R	BG	R	B
gRNA.BRCA1.94	7,490	R	BG	G	RGB
gRNA.BRCA1.95	7,510	R	BG	G	RG
gRNA.BRCA1.96	7,511	R	BG	G	RB
gRNA.BRCA1.97	7,720	R	BG	G	BG
gRNA.BRCA1.98	7,721	R	BG	G	R
gRNA.BRCA1.99	7,997	R	BG	B	RGB
gRNA.BRCA1.100	8,022	R	BG	B	RG
gRNA.BRCA1.101	8,200	R	BG	B	RB
gRNA.BRCA1.102	8,258	R	BG	B	BG

[0293]

gRNA.BRCA1.103	8,279	R	BG	B	R
gRNA.BRCA1.104	8,380	R	R	RGB	RG
gRNA.BRCA1.105	8,544	R	R	RGB	RB
gRNA.BRCA1.106	8,556	R	R	RGB	BG
gRNA.BRCA1.107	8,702	G	R	RGB	R
gRNA.BRCA1.108	8,859	G	R	RGB	G
gRNA.BRCA1.109	8,990	G	R	RG	RGB
gRNA.BRCA1.110	9,079	G	R	RG	RB
gRNA.BRCA1.111	9,080	G	R	RG	BG
gRNA.BRCA1.112	9,315	G	R	RG	R
gRNA.BRCA1.113	9,339	G	R	RG	G
gRNA.BRCA1.114	9,407	G	R	RG	B
gRNA.BRCA1.115	9,539	G	R	RB	RGB
gRNA.BRCA1.116	9,814	G	R	RB	RG
gRNA.BRCA1.117	9,835	G	R	RB	BG
gRNA.BRCA1.118	10,035	G	R	RB	R
gRNA.BRCA1.119	10,047	G	R	RB	G
gRNA.BRCA1.120	10,107	G	R	RB	B
gRNA.BRCA1.121	10,152	G	R	BG	RGB
gRNA.BRCA1.122	10,201	G	R	BG	RG
gRNA.BRCA1.123	10,303	G	R	BG	RB
gRNA.BRCA1.124	10,478	G	R	BG	R
gRNA.BRCA1.125	10,545	G	R	BG	G
gRNA.BRCA1.126	10,548	G	R	BG	B
gRNA.BRCA1.127	10,549	G	R	G	RGB
gRNA.BRCA1.128	10,551	G	R	G	RG
gRNA.BRCA1.129	10,760	G	R	G	RB
gRNA.BRCA1.130	10,866	G	R	G	BG
gRNA.BRCA1.131	10,874	G	R	B	RG

[0294]

gRNA.BRCA1.132	10,894	G	R	B	RB
gRNA.BRCA1.133	10,990	G	R	B	BG
gRNA.BRCA1.134	10,997	G	G	RGB	RG
gRNA.BRCA1.135	11,082	G	G	RGB	RB
gRNA.BRCA1.136	11,117	G	G	RGB	BG
gRNA.BRCA1.137	11,137	G	G	RGB	R
gRNA.BRCA1.138	11,151	G	G	RGB	G
gRNA.BRCA1.139	11,172	G	G	RGB	B
gRNA.BRCA1.140	11,602	G	G	RG	RGB
gRNA.BRCA1.141	11,796	G	G	RG	RB
gRNA.BRCA1.142	11,804	G	G	RG	BG
gRNA.BRCA1.143	12,424	G	G	RG	R
gRNA.BRCA1.144	12,737	G	G	RG	G
gRNA.BRCA1.145	12,878	G	G	RB	RGB
gRNA.BRCA1.146	12,922	G	G	RB	RG
gRNA.BRCA1.147	12,977	G	G	RB	BG
gRNA.BRCA1.148	12,998	G	G	RB	R
gRNA.BRCA1.149	13,156	G	G	RB	G
gRNA.BRCA1.150	13,178	G	G	RB	B
gRNA.BRCA1.151	13,219	G	G	BG	RGB
gRNA.BRCA1.152	13,483	G	G	BG	RG
gRNA.BRCA1.153	13,578	G	G	BG	RB
gRNA.BRCA1.154	13,637	G	G	BG	R
gRNA.BRCA1.155	13,741	G	G	BG	G
gRNA.BRCA1.156	13,831	G	G	R	RGB
gRNA.BRCA1.157	13,997	G	G	R	RG
gRNA.BRCA1.158	14,306	G	G	R	RB
gRNA.BRCA1.159	14,476	G	G	R	BG
gRNA.BRCA1.160	14,501	G	G	B	RB

[0295]

gRNA.BRCA1.161	14,522	G	G	B	BG
gRNA.BRCA1.162	14,575	G	B	RGB	RG
gRNA.BRCA1.163	14,587	G	B	RGB	RB
gRNA.BRCA1.164	15,033	G	B	RGB	BG
gRNA.BRCA1.165	15,219	G	B	RG	RGB
gRNA.BRCA1.166	15,222	G	B	RG	RB
gRNA.BRCA1.167	15,475	G	B	RG	BG
gRNA.BRCA1.168	15,522	G	B	RG	R
gRNA.BRCA1.169	15,579	G	B	RB	RGB
gRNA.BRCA1.170	15,818	G	B	RB	RG
gRNA.BRCA1.171	15,901	G	B	RB	BG
gRNA.BRCA1.172	16,111	G	B	RB	R
gRNA.BRCA1.173	16,136	G	B	BG	RGB
gRNA.BRCA1.174	16,448	G	B	BG	RG
gRNA.BRCA1.175	16,595	G	B	BG	RB
gRNA.BRCA1.176	16,688	G	B	BG	R
gRNA.BRCA1.177	16,747	G	B	R	RGB
gRNA.BRCA1.178	16,748	G	B	R	BG
gRNA.BRCA1.179	16,918	G	B	G	RG
gRNA.BRCA1.180	16,974	G	B	G	RB
gRNA.BRCA1.181	17,418	G	B	G	BG
gRNA.BRCA1.182	17,422	G	RGB	RG	RGB
gRNA.BRCA1.183	17,589	G	RGB	RG	RB
gRNA.BRCA1.184	17,669	G	RGB	RG	BG
gRNA.BRCA1.185	18,262	G	RGB	RG	R
gRNA.BRCA1.186	18,269	G	RGB	RG	G
gRNA.BRCA1.187	18,970	G	RGB	RG	B
gRNA.BRCA1.188	18,971	G	RGB	RB	RGB
gRNA.BRCA1.189	19,005	G	RGB	RB	RG

[0296]

gRNA.BRCA1.190	19,026	G	RGB	RB	BG
gRNA.BRCA1.191	19,289	G	RGB	RB	R
gRNA.BRCA1.192	19,392	G	RGB	RB	G
gRNA.BRCA1.193	19,445	G	RGB	RB	B
gRNA.BRCA1.194	19,545	G	RGB	BG	RGB
gRNA.BRCA1.195	19,570	G	RGB	BG	RG
gRNA.BRCA1.196	19,818	G	RGB	BG	RB
gRNA.BRCA1.197	19,819	G	RGB	BG	R
gRNA.BRCA1.198	19,839	G	RGB	BG	G
gRNA.BRCA1.199	20,595	G	RGB	BG	B
gRNA.BRCA1.200	20,810	G	RGB	R	RGB
gRNA.BRCA1.201	20,952	G	RGB	R	RG
gRNA.BRCA1.202	20,988	G	RGB	R	RB
gRNA.BRCA1.203	21,054	G	RGB	R	BG
gRNA.BRCA1.204	21,075	G	RGB	R	G
gRNA.BRCA1.205	21,352	G	RGB	G	RGB
gRNA.BRCA1.206	21,440	G	RGB	G	RG
gRNA.BRCA1.207	21,499	G	RGB	G	RB
gRNA.BRCA1.208	21,506	G	RGB	G	BG
gRNA.BRCA1.209	21,581	G	RGB	G	R
gRNA.BRCA1.210	21,592	G	RGB	B	RGB
gRNA.BRCA1.211	22,366	G	RGB	B	RG
gRNA.BRCA1.212	22,373	G	RGB	B	RB
gRNA.BRCA1.213	22,729	B	RGB	B	BG
gRNA.BRCA1.214	22,760	B	RGB	B	G
gRNA.BRCA1.215	22,824	B	RB	RGB	RG
gRNA.BRCA1.216	22,849	B	RB	RGB	RB
gRNA.BRCA1.217	23,339	B	RB	RGB	BG
gRNA.BRCA1.218	23,739	B	RB	RGB	R

[0297]

gRNA.BRCA1.219	24,253	B	RB	RGB	G
gRNA.BRCA1.220	24,443	B	RB	RGB	B
gRNA.BRCA1.221	24,530	B	RB	RG	RGB
gRNA.BRCA1.222	24,877	B	RB	RG	RB
gRNA.BRCA1.223	24,878	B	RB	RG	BG
gRNA.BRCA1.224	24,927	B	RB	RG	R
gRNA.BRCA1.225	25,105	B	RB	RG	G
gRNA.BRCA1.226	25,350	B	RB	RG	B
gRNA.BRCA1.227	25,368	B	RB	BG	RGB
gRNA.BRCA1.228	25,472	B	RB	BG	RG
gRNA.BRCA1.229	25,525	B	RB	BG	RB
gRNA.BRCA1.230	25,614	B	RB	BG	R
gRNA.BRCA1.231	25,693	B	RB	BG	G
gRNA.BRCA1.232	25,852	B	RB	BG	B
gRNA.BRCA1.233	25,936	B	RB	R	RGB
gRNA.BRCA1.234	25,972	B	RB	R	RG
gRNA.BRCA1.235	26,984	B	RB	R	RB
gRNA.BRCA1.236	27,130	B	RB	R	BG
gRNA.BRCA1.237	27,155	B	RB	R	G
gRNA.BRCA1.238	27,252	B	RB	G	RGB
gRNA.BRCA1.239	27,256	B	RB	G	RG
gRNA.BRCA1.240	27,431	B	RB	G	RB
gRNA.BRCA1.241	27,447	B	RB	G	BG
gRNA.BRCA1.242	27,452	B	RB	G	R
gRNA.BRCA1.243	27,610	B	RB	B	RGB
gRNA.BRCA1.244	27,631	B	RB	B	RG
gRNA.BRCA1.245	27,682	B	RB	B	RB
gRNA.BRCA1.246	27,722	B	RB	B	BG
gRNA.BRCA1.247	28,262	B	RB	B	R

[0298]

gRNA.BRCA1.248	28,287	B	RB	B	G
gRNA.BRCA1.249	28,400	B	BG	RGB	RG
gRNA.BRCA1.250	28,425	B	BG	RGB	RB
gRNA.BRCA1.251	28,464	B	BG	RGB	BG
gRNA.BRCA1.252	28,581	B	BG	RGB	R
gRNA.BRCA1.253	28,817	B	BG	RGB	G
gRNA.BRCA1.254	28,958	B	BG	RGB	B
gRNA.BRCA1.255	28,983	B	BG	RG	RGB
gRNA.BRCA1.256	29,080	B	BG	RG	RB
gRNA.BRCA1.257	29,202	B	BG	RG	BG
gRNA.BRCA1.258	29,516	B	BG	RG	R
gRNA.BRCA1.259	30,531	B	BG	RG	G
gRNA.BRCA1.260	30,964	B	BG	RG	B
gRNA.BRCA1.261	30,989	B	BG	RB	RGB
gRNA.BRCA1.262	31,104	B	BG	RB	RG
gRNA.BRCA1.263	31,374	B	BG	RB	BG
gRNA.BRCA1.264	31,469	B	BG	RB	R
gRNA.BRCA1.265	31,486	B	BG	RB	G
gRNA.BRCA1.266	31,547	B	BG	RB	B
gRNA.BRCA1.267	31,698	B	BG	R	RGB
gRNA.BRCA1.268	32,906	B	BG	R	RG
gRNA.BRCA1.269	32,942	B	BG	R	RB
gRNA.BRCA1.270	32,973	B	BG	R	BG
gRNA.BRCA1.271	33,121	B	BG	R	G
gRNA.BRCA1.272	33,128	B	BG	R	B
gRNA.BRCA1.273	33,335	B	BG	G	RGB
gRNA.BRCA1.274	33,496	B	BG	G	RG
gRNA.BRCA1.275	33,994	B	BG	G	RB
gRNA.BRCA1.276	34,521	B	BG	G	BG

[0299]

gRNA.BRCA1.277	34,751	B	BG	G	R
gRNA.BRCA1.278	34,892	B	BG	B	RGB
gRNA.BRCA1.279	34,917	B	BG	B	RG
gRNA.BRCA1.280	34,988	B	BG	B	RB
gRNA.BRCA1.281	35,188	B	BG	B	BG
gRNA.BRCA1.282	35,232	B	BG	B	R
gRNA.BRCA1.283	35,330	B	R	RGB	RG
gRNA.BRCA1.284	35,366	B	R	RGB	RB
gRNA.BRCA1.285	35,445	B	R	RGB	BG
gRNA.BRCA1.286	35,931	B	R	RGB	R
gRNA.BRCA1.287	36,116	B	R	RGB	G
gRNA.BRCA1.288	36,134	B	R	RGB	B
gRNA.BRCA1.289	36,764	B	R	RG	RGB
gRNA.BRCA1.290	36,884	B	R	RG	RB
gRNA.BRCA1.291	36,905	B	R	RG	BG
gRNA.BRCA1.292	37,050	B	R	RG	R
gRNA.BRCA1.293	37,334	B	R	RG	G
gRNA.BRCA1.294	37,867	B	R	RB	RGB
gRNA.BRCA1.295	38,007	B	R	RB	RG
gRNA.BRCA1.296	38,032	B	R	RB	BG
gRNA.BRCA1.297	38,275	B	R	RB	R
gRNA.BRCA1.298	38,325	B	R	RB	G
gRNA.BRCA1.299	38,470	B	R	RB	B
gRNA.BRCA1.300	38,471	B	R	BG	RGB
gRNA.BRCA1.301	38,472	B	R	BG	RG
gRNA.BRCA1.302	38,825	B	R	BG	RB
gRNA.BRCA1.303	38,842	B	R	BG	R
gRNA.BRCA1.304	39,099	B	R	BG	G
gRNA.BRCA1.305	39,109	B	R	BG	B

[0300]

gRNA.BRCA1.306	39,114	B	R	G	RGB
gRNA.BRCA1.307	39,115	B	R	G	RG
gRNA.BRCA1.308	39,786	B	R	G	RB
gRNA.BRCA1.309	39,801	B	R	G	BG
gRNA.BRCA1.310	39,865	B	R	B	RGB
gRNA.BRCA1.311	40,010	B	R	B	RG
gRNA.BRCA1.312	40,054	B	R	B	RB
gRNA.BRCA1.313	40,221	B	G	RGB	RG
gRNA.BRCA1.314	40,663	B	G	RGB	RB
gRNA.BRCA1.315	40,927	B	G	RGB	BG
gRNA.BRCA1.316	41,058	B	G	RGB	R
gRNA.BRCA1.317	41,059	B	G	RGB	G
gRNA.BRCA1.318	41,341	B	G	RG	RGB
gRNA.BRCA1.319	41,688	B	G	RG	RB
gRNA.BRCA1.320	42,346	RG	G	RG	BG
gRNA.BRCA1.321	42,400	RG	G	RG	R
gRNA.BRCA1.322	42,753	RG	G	RG	G
gRNA.BRCA1.323	42,862	RG	G	RG	B
gRNA.BRCA1.324	42,870	RG	G	RB	RGB
gRNA.BRCA1.325	42,891	RG	G	RB	RG
gRNA.BRCA1.326	42,912	RG	G	RB	BG
gRNA.BRCA1.327	42,943	RG	G	RB	R
gRNA.BRCA1.328	42,952	RG	G	RB	G
gRNA.BRCA1.329	42,964	RG	G	RB	B
gRNA.BRCA1.330	43,471	RG	G	BG	RGB
gRNA.BRCA1.331	43,514	RG	G	BG	RG
gRNA.BRCA1.332	43,594	RG	G	BG	RB
gRNA.BRCA1.333	43,783	RG	G	BG	R
gRNA.BRCA1.334	44,610	RG	G	BG	G

[0301]

gRNA.BRCA1.335	44,956	RG	G	R	RGB
gRNA.BRCA1.336	45,085	RG	G	R	RG
gRNA.BRCA1.337	45,408	RG	G	R	RB
gRNA.BRCA1.338	45,642	RG	G	R	BG
gRNA.BRCA1.339	45,834	RG	G	B	RGB
gRNA.BRCA1.340	45,939	RG	G	B	BG
gRNA.BRCA1.341	46,018	RG	B	RGB	RG
gRNA.BRCA1.342	46,057	RG	B	RGB	RB
gRNA.BRCA1.343	46,159	RG	B	RGB	BG
gRNA.BRCA1.344	46,184	RG	B	RGB	G
gRNA.BRCA1.345	46,866	RG	B	RG	RGB
gRNA.BRCA1.346	47,006	RG	B	RG	RB
gRNA.BRCA1.347	47,161	RG	B	RG	BG
gRNA.BRCA1.348	47,295	RG	B	RG	G
gRNA.BRCA1.349	47,316	RG	B	RB	RGB
gRNA.BRCA1.350	47,624	RG	B	RB	RG
gRNA.BRCA1.351	48,296	RG	B	RB	BG
gRNA.BRCA1.352	48,417	RG	B	RB	R
gRNA.BRCA1.353	48,597	RG	B	BG	RGB
gRNA.BRCA1.354	48,709	RG	B	BG	RG
gRNA.BRCA1.355	48,741	RG	B	BG	RB
gRNA.BRCA1.356	49,338	RG	B	BG	G
gRNA.BRCA1.357	49,509	RG	B	R	RG
gRNA.BRCA1.358	49,850	RG	B	R	BG
gRNA.BRCA1.359	50,036	RG	B	G	RG
gRNA.BRCA1.360	50,332	RG	B	G	RB
gRNA.BRCA1.361	50,553	RG	RGB	RG	RGB
gRNA.BRCA1.362	50,611	RG	RGB	RG	RB
gRNA.BRCA1.363	50,725	RG	RGB	RG	BG

[0302]

gRNA.BRCA1.364	51,015	RG	RGB	RG	R
gRNA.BRCA1.365	51,082	RG	RGB	RG	G
gRNA.BRCA1.366	51,519	RG	RGB	RG	B
gRNA.BRCA1.367	51,797	RG	RGB	RB	RGB
gRNA.BRCA1.368	51,939	RG	RGB	RB	RG
gRNA.BRCA1.369	51,964	RG	RGB	RB	BG
gRNA.BRCA1.370	52,106	RG	RGB	RB	R
gRNA.BRCA1.371	52,142	RG	RGB	RB	G
gRNA.BRCA1.372	52,240	RG	RGB	RB	B
gRNA.BRCA1.373	52,265	RG	RGB	BG	RGB
gRNA.BRCA1.374	52,396	RG	RGB	BG	RG
gRNA.BRCA1.375	52,687	RG	RGB	BG	RB
gRNA.BRCA1.376	52,906	RG	RGB	BG	R
gRNA.BRCA1.377	53,519	RG	RGB	BG	G
gRNA.BRCA1.378	54,362	RG	RGB	BG	B
gRNA.BRCA1.379	54,794	RG	RGB	R	RGB
gRNA.BRCA1.380	54,956	RG	RGB	R	RG
gRNA.BRCA1.381	54,957	RG	RGB	R	RB
gRNA.BRCA1.382	54,976	RG	RGB	R	BG
gRNA.BRCA1.383	55,494	RG	RGB	R	G
gRNA.BRCA1.384	55,623	RG	RGB	G	RGB
gRNA.BRCA1.385	55,661	RG	RGB	G	RG
gRNA.BRCA1.386	55,668	RG	RGB	G	RB
gRNA.BRCA1.387	55,678	RG	RGB	G	BG
gRNA.BRCA1.388	56,191	RG	RGB	G	R
gRNA.BRCA1.389	56,192	RG	RGB	B	RGB
gRNA.BRCA1.390	56,200	RG	RGB	B	RG
gRNA.BRCA1.391	57,155	RG	RGB	B	RB
gRNA.BRCA1.392	57,180	RG	RGB	B	BG

[0303]

gRNA.BRCA1.393	57,839	RG	RGB	B	R
gRNA.BRCA1.394	57,864	RG	RGB	B	G
gRNA.BRCA1.395	58,010	RG	RG	RGB	RG
gRNA.BRCA1.396	58,204	RG	RG	RGB	RB
gRNA.BRCA1.397	58,282	RG	RG	RGB	BG
gRNA.BRCA1.398	58,325	RG	RG	RGB	R
gRNA.BRCA1.399	58,449	RG	RG	RGB	G
gRNA.BRCA1.400	58,474	RG	RG	RGB	B
gRNA.BRCA1.401	58,672	RG	RG	RB	RGB
gRNA.BRCA1.402	58,835	RG	RG	RB	RG
gRNA.BRCA1.403	59,401	RG	RG	RB	BG
gRNA.BRCA1.404	59,800	RG	RG	RB	R
gRNA.BRCA1.405	59,983	RG	RG	RB	G
gRNA.BRCA1.406	60,099	RG	RG	RB	B
gRNA.BRCA1.407	60,227	RG	RG	BG	RGB
gRNA.BRCA1.408	60,231	RG	RG	BG	RG
gRNA.BRCA1.409	60,248	RG	RG	BG	RB
gRNA.BRCA1.410	60,252	RG	RG	BG	R
gRNA.BRCA1.411	61,036	RG	RG	BG	G
gRNA.BRCA1.412	61,740	RG	RG	BG	B
gRNA.BRCA1.413	61,926	RG	RG	R	RGB
gRNA.BRCA1.414	62,214	RG	RG	R	RG
gRNA.BRCA1.415	62,304	RG	RG	R	RB
gRNA.BRCA1.416	62,672	RG	RG	R	BG
gRNA.BRCA1.417	62,820	RG	RG	R	G
gRNA.BRCA1.418	63,273	RG	RG	R	B
gRNA.BRCA1.419	63,414	RG	RG	G	RGB
gRNA.BRCA1.420	63,439	RG	RG	G	RG
gRNA.BRCA1.421	63,505	RG	RG	G	RB

[0304]

gRNA.BRCA1.422	64,068	RG	RG	G	BG
gRNA.BRCA1.423	64,187	RG	RG	G	R
gRNA.BRCA1.424	64,302	RG	RG	G	B
gRNA.BRCA1.425	64,404	RG	RG	B	RGB
gRNA.BRCA1.426	64,411	RG	RG	B	RG
gRNA.BRCA1.427	64,455	RB	RG	B	RB
gRNA.BRCA1.428	64,479	RB	RG	B	BG
gRNA.BRCA1.429	64,500	RB	RG	B	R
gRNA.BRCA1.430	65,444	RB	RG	B	G
gRNA.BRCA1.431	65,692	RB	BG	RGB	RG
gRNA.BRCA1.432	65,759	RB	BG	RGB	RB
gRNA.BRCA1.433	66,292	RB	BG	RGB	BG
gRNA.BRCA1.434	66,442	RB	BG	RGB	R
gRNA.BRCA1.435	66,751	RB	BG	RGB	G
gRNA.BRCA1.436	66,897	RB	BG	RGB	B
gRNA.BRCA1.437	66,922	RB	BG	RG	RGB
gRNA.BRCA1.438	67,204	RB	BG	RG	RB
gRNA.BRCA1.439	67,370	RB	BG	RG	BG
gRNA.BRCA1.440	67,444	RB	BG	RG	R
gRNA.BRCA1.441	67,977	RB	BG	RG	G
gRNA.BRCA1.442	68,830	RB	BG	RG	B
gRNA.BRCA1.443	68,873	RB	BG	RB	RGB
gRNA.BRCA1.444	68,912	RB	BG	RB	RG
gRNA.BRCA1.445	69,130	RB	BG	RB	BG
gRNA.BRCA1.446	69,271	RB	BG	RB	R
gRNA.BRCA1.447	69,436	RB	BG	RB	G
gRNA.BRCA1.448	69,925	RB	BG	RB	B
gRNA.BRCA1.449	70,004	RB	BG	R	RGB
gRNA.BRCA1.450	70,147	RB	BG	R	RG

[0305]

gRNA.BRCA1.451	70,290	RB	BG	R	RB
gRNA.BRCA1.452	70,432	RB	BG	R	BG
gRNA.BRCA1.453	70,467	RB	BG	R	G
gRNA.BRCA1.454	70,754	RB	BG	R	B
gRNA.BRCA1.455	70,800	RB	BG	G	RGB
gRNA.BRCA1.456	70,924	RB	BG	G	RG
gRNA.BRCA1.457	71,133	RB	BG	G	RB
gRNA.BRCA1.458	71,158	RB	BG	G	BG
gRNA.BRCA1.459	71,397	RB	BG	G	R
gRNA.BRCA1.460	71,398	RB	BG	B	RGB
gRNA.BRCA1.461	71,773	RB	BG	B	RG
gRNA.BRCA1.462	71,798	RB	BG	B	RB
gRNA.BRCA1.463	71,915	RB	BG	B	BG
gRNA.BRCA1.464	72,051	RB	BG	B	R
gRNA.BRCA1.465	73,315	RB	BG	B	G
gRNA.BRCA1.466	73,457	RB	R	RGB	RG
gRNA.BRCA1.467	73,797	RB	R	RGB	RB
gRNA.BRCA1.468	73,909	RB	R	RGB	BG
gRNA.BRCA1.469	73,951	RB	R	RGB	R
gRNA.BRCA1.470	74,973	RB	R	RGB	G
gRNA.BRCA1.471	74,998	RB	R	RGB	B
gRNA.BRCA1.472	75,052	RB	R	RG	RGB
gRNA.BRCA1.473	75,139	RB	R	RG	RB
gRNA.BRCA1.474	75,148	RB	R	RG	BG
gRNA.BRCA1.475	75,175	RB	R	RG	R
gRNA.BRCA1.476	75,837	RB	R	RG	G
gRNA.BRCA1.477	75,851	RB	R	RB	RGB
gRNA.BRCA1.478	75,926	RB	R	RB	RG
gRNA.BRCA1.479	75,927	RB	R	RB	BG

[0306]

gRNA.BRCA1.480	76,170	RB	R	RB	R
gRNA.BRCA1.481	76,487	RB	R	RB	G
gRNA.BRCA1.482	77,242	RB	R	RB	B
gRNA.BRCA1.483	77,401	RB	R	BG	RGB
gRNA.BRCA1.484	77,803	RB	R	BG	RG
gRNA.BRCA1.485	77,970	RB	R	BG	RB
gRNA.BRCA1.486	78,230	RB	R	BG	R
gRNA.BRCA1.487	78,231	RB	R	BG	G
gRNA.BRCA1.488	78,277	RB	R	G	RGB
gRNA.BRCA1.489	78,625	RB	R	G	RG
gRNA.BRCA1.490	78,933	RB	R	G	RB
gRNA.BRCA1.491	78,958	RB	R	G	BG
gRNA.BRCA1.492	79,186	RB	R	B	RG
gRNA.BRCA1.493	79,211	RB	R	B	RB
gRNA.BRCA1.494	79,475	RB	G	RGB	RG
gRNA.BRCA1.495	79,600	RB	G	RGB	RB
gRNA.BRCA1.496	79,654	RB	G	RGB	BG
gRNA.BRCA1.497	79,679	RB	G	RGB	R
gRNA.BRCA1.498	79,998	RB	G	RGB	G
gRNA.BRCA1.499	80,198	RB	G	RGB	B
gRNA.BRCA1.500	80,737	RB	G	RG	RGB
gRNA.BRCA1.501	80,877	RB	G	RG	RB
gRNA.BRCA1.502	80,902	RB	G	RG	BG
gRNA.BRCA1.503	80,986	RB	G	RG	R
gRNA.BRCA1.504	81,300	RB	G	RG	G
gRNA.BRCA1.505	81,339	RB	G	RB	RGB
gRNA.BRCA1.506	81,479	RB	G	RB	RG
gRNA.BRCA1.507	81,765	RB	G	RB	BG
gRNA.BRCA1.508	81,907	RB	G	RB	R

[0307]

gRNA.BRCA1.509	81,932	RB	G	RB	G
gRNA.BRCA1.510	82,467	RB	G	RB	B
gRNA.BRCA1.511	82,513	RB	G	BG	RGB
gRNA.BRCA1.512	82,521	RB	G	BG	RG
gRNA.BRCA1.513	82,522	RB	G	BG	RB
gRNA.BRCA1.514	82,537	RB	G	BG	R
gRNA.BRCA1.515	82,729	RB	G	BG	G
gRNA.BRCA1.516	82,919	RB	G	R	RGB
gRNA.BRCA1.517	83,157	RB	G	R	RG
gRNA.BRCA1.518	83,334	RB	G	R	RB
gRNA.BRCA1.519	83,681	RB	G	R	BG
gRNA.BRCA1.520	84,189	RB	G	B	RGB
gRNA.BRCA1.521	84,452	RB	G	B	BG
gRNA.BRCA1.522	84,618	RB	B	RGB	RG
gRNA.BRCA1.523	85,185	RB	B	RGB	RB
gRNA.BRCA1.524	85,276	RB	B	RGB	BG
gRNA.BRCA1.525	85,394	RB	B	RG	RGB
gRNA.BRCA1.526	85,527	RB	B	RG	RB
gRNA.BRCA1.527	85,571	RB	B	RG	BG
gRNA.BRCA1.528	85,596	RB	B	RG	R
gRNA.BRCA1.529	85,683	RB	B	RG	G
gRNA.BRCA1.530	85,936	RB	B	RB	RGB
gRNA.BRCA1.531	85,955	RB	B	RB	RG
gRNA.BRCA1.532	86,055	RB	B	RB	BG
gRNA.BRCA1.533	86,194	GB	B	BG	RGB
gRNA.BRCA1.534	86,303	GB	B	BG	RG
gRNA.BRCA1.535	86,310	GB	B	BG	RB
gRNA.BRCA1.536	86,697	GB	B	BG	R
gRNA.BRCA1.537	86,810	GB	B	R	RGB

[0308]

gRNA.BRCA1.538	86,840	GB	B	R	RG
gRNA.BRCA1.539	86,960	GB	B	R	BG
gRNA.BRCA1.540	87,126	GB	B	G	RG
gRNA.BRCA1.541	87,199	GB	B	G	BG
gRNA.BRCA1.542	87,247	GB	RGB	RG	RGB
gRNA.BRCA1.543	87,355	GB	RGB	RG	RB
gRNA.BRCA1.544	88,350	GB	RGB	RG	BG
gRNA.BRCA1.545	88,419	GB	RGB	RG	R
gRNA.BRCA1.546	88,446	GB	RGB	RG	G
gRNA.BRCA1.547	88,455	GB	RGB	RG	B
gRNA.BRCA1.548	88,543	GB	RGB	RB	RGB
gRNA.BRCA1.549	88,597	GB	RGB	RB	RG
gRNA.BRCA1.550	88,604	GB	RGB	RB	BG
gRNA.BRCA1.551	88,622	GB	RGB	RB	R
gRNA.BRCA1.552	88,720	GB	RGB	RB	G
gRNA.BRCA1.553	88,902	GB	RGB	RB	B
gRNA.BRCA1.554	88,909	GB	RGB	BG	RGB
gRNA.BRCA1.555	88,927	GB	RGB	BG	RG
gRNA.BRCA1.556	88,955	GB	RGB	BG	RB
gRNA.BRCA1.557	88,960	GB	RGB	BG	R
gRNA.BRCA1.558	88,998	GB	RGB	BG	G
gRNA.BRCA1.559	89,029	GB	RGB	BG	B
gRNA.BRCA1.560	89,153	GB	RGB	R	RGB
gRNA.BRCA1.561	89,207	GB	RGB	R	RG
gRNA.BRCA1.562	89,351	GB	RGB	R	RB
gRNA.BRCA1.563	89,515	GB	RGB	R	BG
gRNA.BRCA1.564	90,439	GB	RGB	R	G
gRNA.BRCA1.565	90,440	GB	RGB	G	RGB
gRNA.BRCA1.566	90,441	GB	RGB	G	RG

[0309]

gRNA.BRCA1.567	90,462	GB	RGB	G	RB
gRNA.BRCA1.568	90,700	GB	RGB	G	BG
gRNA.BRCA1.569	90,736	GB	RGB	G	R
gRNA.BRCA1.570	90,914	GB	RGB	G	B
gRNA.BRCA1.571	91,209	GB	RGB	B	RGB
gRNA.BRCA1.572	91,284	GB	RGB	B	RG
gRNA.BRCA1.573	91,323	GB	RGB	B	RB
gRNA.BRCA1.574	91,424	GB	RGB	B	BG
gRNA.BRCA1.575	91,460	GB	RGB	B	R
gRNA.BRCA1.576	91,603	GB	RG	RGB	RG
gRNA.BRCA1.577	91,678	GB	RG	RGB	RB
gRNA.BRCA1.578	91,710	GB	RG	RGB	BG
gRNA.BRCA1.579	92,123	GB	RG	RGB	R
gRNA.BRCA1.580	92,226	GB	RG	RGB	G
gRNA.BRCA1.581	92,269	GB	RG	RGB	B
gRNA.BRCA1.582	92,347	GB	RG	RB	RGB
gRNA.BRCA1.583	92,363	GB	RG	RB	RG
gRNA.BRCA1.584	92,368	GB	RG	RB	BG
gRNA.BRCA1.585	92,384	GB	RG	RB	R
gRNA.BRCA1.586	92,520	GB	RG	RB	G
gRNA.BRCA1.587	92,536	GB	RG	RB	B
gRNA.BRCA1.588	92,540	GB	RG	BG	RGB
gRNA.BRCA1.589	92,541	GB	RG	BG	RG
gRNA.BRCA1.590	92,557	GB	RG	BG	RB
gRNA.BRCA1.591	92,604	GB	RG	BG	R
gRNA.BRCA1.592	92,690	GB	RG	BG	G
gRNA.BRCA1.593	92,820	GB	RG	BG	B
gRNA.BRCA1.594	92,899	GB	RG	R	RGB
gRNA.BRCA1.595	92,914	GB	RG	R	RG

[0310]

gRNA.BRCA1.596	92,935	GB	RG	R	RB
gRNA.BRCA1.597	92,936	GB	RG	R	BG
gRNA.BRCA1.598	92,957	GB	RG	R	G
gRNA.BRCA1.599	92,978	GB	RG	R	B
gRNA.BRCA1.600	92,979	GB	RG	G	RGB
gRNA.BRCA1.601	93,001	GB	RG	G	RG
gRNA.BRCA1.602	93,022	GB	RG	G	RB
gRNA.BRCA1.603	93,029	GB	RG	G	BG
gRNA.BRCA1.604	93,045	GB	RG	G	R
gRNA.BRCA1.605	93,067	GB	RG	B	RGB
gRNA.BRCA1.606	93,088	GB	RG	B	RG
gRNA.BRCA1.607	93,107	GB	RG	B	RB
gRNA.BRCA1.608	93,108	GB	RG	B	BG
gRNA.BRCA1.609	93,123	GB	RG	B	G
gRNA.BRCA1.610	93,134	GB	RB	RGB	RG
gRNA.BRCA1.611	93,140	GB	RB	RGB	RB
gRNA.BRCA1.612	93,141	GB	RB	RGB	BG
gRNA.BRCA1.613	93,173	GB	RB	RGB	R
gRNA.BRCA1.614	93,364	GB	RB	RGB	G
gRNA.BRCA1.615	93,611	GB	RB	RGB	B
gRNA.BRCA1.616	93,612	GB	RB	RG	RGB
gRNA.BRCA1.617	93,774	GB	RB	RG	RB
gRNA.BRCA1.618	93,999	GB	RB	RG	BG
gRNA.BRCA1.619	94,032	GB	RB	RG	R
gRNA.BRCA1.620	94,130	GB	RB	RG	G
gRNA.BRCA1.621	94,341	GB	RB	RG	B
gRNA.BRCA1.622	94,367	GB	RB	BG	RGB
gRNA.BRCA1.623	94,388	GB	RB	BG	RG
gRNA.BRCA1.624	94,406	GB	RB	BG	RB

[0311]

gRNA.BRCA1.625	94,429	GB	RB	BG	R
gRNA.BRCA1.626	94,511	GB	RB	BG	G
gRNA.BRCA1.627	94,577	GB	RB	BG	B
gRNA.BRCA1.628	94,608	GB	RB	R	RGB
gRNA.BRCA1.629	94,644	GB	RB	R	RG
gRNA.BRCA1.630	94,645	GB	RB	R	RB
gRNA.BRCA1.631	94,666	GB	RB	R	BG
gRNA.BRCA1.632	94,674	GB	RB	R	G
gRNA.BRCA1.633	94,712	GB	RB	R	B
gRNA.BRCA1.634	94,713	GB	RB	G	RGB
gRNA.BRCA1.635	94,725	GB	RB	G	RG
gRNA.BRCA1.636	94,803	GB	RB	G	RB
gRNA.BRCA1.637	94,804	GB	RB	G	BG
gRNA.BRCA1.638	94,816	GB	RB	G	R
gRNA.BRCA1.639	94,850	RGB	RB	G	B
gRNA.BRCA1.640	94,904	RGB	RB	B	RGB
gRNA.BRCA1.641	94,905	RGB	RB	B	RG
gRNA.BRCA1.642	94,922	RGB	RB	B	RB
gRNA.BRCA1.643	94,923	RGB	RB	B	BG
gRNA.BRCA1.644	94,965	RGB	RB	B	R
gRNA.BRCA1.645	94,966	RGB	RB	B	G
gRNA.BRCA1.646	94,982	RGB	R	RGB	RG
gRNA.BRCA1.647	94,991	RGB	R	RGB	RB
gRNA.BRCA1.648	95,012	RGB	R	RGB	BG
gRNA.BRCA1.649	95,018	RGB	R	RGB	R
gRNA.BRCA1.650	95,033	RGB	R	RGB	G
gRNA.BRCA1.651	95,038	RGB	R	RG	RGB
gRNA.BRCA1.652	95,039	RGB	R	RG	RB
gRNA.BRCA1.653	95,636	RGB	R	RG	BG

[0312]

gRNA.BRCA1.654	95,714	RGB	R	RG	R
gRNA.BRCA1.655	95,802	RGB	R	RG	G
gRNA.BRCA1.656	95,856	RGB	R	RB	RGB
gRNA.BRCA1.657	95,881	RGB	R	RB	RG
gRNA.BRCA1.658	96,183	RGB	R	RB	BG
gRNA.BRCA1.659	96,409	RGB	R	RB	R
gRNA.BRCA1.660	96,593	RGB	R	RB	G
gRNA.BRCA1.661	96,594	RGB	R	RB	B
gRNA.BRCA1.662	96,879	RGB	R	BG	RGB
gRNA.BRCA1.663	97,267	RGB	R	BG	RG
gRNA.BRCA1.664	97,461	RGB	R	BG	RB
gRNA.BRCA1.665	97,550	RGB	R	BG	R
gRNA.BRCA1.666	97,586	RGB	R	BG	G
gRNA.BRCA1.667	98,126	RGB	R	G	RGB
gRNA.BRCA1.668	98,248	RGB	R	G	RG
gRNA.BRCA1.669	98,443	RGB	R	G	RB
gRNA.BRCA1.670	99,499	RGB	R	G	BG
gRNA.BRCA1.671	99,578	RGB	R	B	RGB
gRNA.BRCA1.672	99,858	RGB	R	B	RB
gRNA.BRCA1.673	99,989	RGB	R	B	BG
gRNA.BRCA1.674	100,077	RGB	G	RGB	RG
gRNA.BRCA1.675	100,078	RGB	G	RGB	RB
gRNA.BRCA1.676	100,252	RGB	G	RGB	BG
gRNA.BRCA1.677	100,253	RGB	G	RGB	R
gRNA.BRCA1.678	100,267	RGB	G	RGB	G
gRNA.BRCA1.679	100,333	RGB	G	RG	RGB
gRNA.BRCA1.680	100,512	RGB	G	RG	RB
gRNA.BRCA1.681	100,551	RGB	G	RG	BG
gRNA.BRCA1.682	100,672	RGB	G	RG	R

[0313]

gRNA.BRCA1.683	100,697	RGB	G	RG	G
gRNA.BRCA1.684	100,848	RGB	G	RG	B
gRNA.BRCA1.685	100,884	RGB	G	RB	RGB
gRNA.BRCA1.686	100,942	RGB	G	RB	RG
gRNA.BRCA1.687	101,036	RGB	G	RB	BG
gRNA.BRCA1.688	101,094	RGB	G	RB	R
gRNA.BRCA1.689	101,186	RGB	G	RB	G
gRNA.BRCA1.690	101,204	RGB	G	RB	B
gRNA.BRCA1.691	101,219	RGB	G	BG	RGB
gRNA.BRCA1.692	101,286	RGB	G	BG	RG
gRNA.BRCA1.693	101,483	RGB	G	BG	RB
gRNA.BRCA1.694	101,488	RGB	G	BG	R
gRNA.BRCA1.695	101,709	RGB	G	BG	G
gRNA.BRCA1.696	101,841	RGB	G	BG	B
gRNA.BRCA1.697	101,844	RGB	G	R	RGB
gRNA.BRCA1.698	101,845	RGB	G	R	RG
gRNA.BRCA1.699	101,846	RGB	G	R	RB
gRNA.BRCA1.700	101,909	RGB	G	R	BG
gRNA.BRCA1.701	101,972	RGB	B	RGB	RG
gRNA.BRCA1.702	102,080	RGB	B	RGB	RB
gRNA.BRCA1.703	102,348	RGB	B	RGB	BG
gRNA.BRCA1.704	102,360	RGB	B	RG	RGB
gRNA.BRCA1.705	102,471	RGB	B	RG	RB
gRNA.BRCA1.706	102,654	RGB	B	RG	BG
gRNA.BRCA1.707	102,655	RGB	B	RB	RGB
gRNA.BRCA1.708	102,823	RGB	B	RB	RG
gRNA.BRCA1.709	102,882	RGB	B	RB	BG
gRNA.BRCA1.710	102,889	RGB	B	RB	G
gRNA.BRCA1.711	102,981	RGB	B	BG	RGB

[0314]

gRNA.BRCA1.712	103,008	RGB	B	BG	RG
gRNA.BRCA1.713	103,018	RGB	B	BG	RB
gRNA.BRCA1.714	103,039	RGB	B	BG	R
gRNA.BRCA1.715	103,059	RGB	B	R	RG
gRNA.BRCA1.716	103,143	RGB	B	R	BG
gRNA.BRCA1.717	103,168	RGB	B	G	RG
gRNA.BRCA1.718	103,438	RGB	RGB	RG	RGB
gRNA.BRCA1.719	103,446	RGB	RGB	RG	RB
gRNA.BRCA1.720	103,459	RGB	RGB	RG	BG
gRNA.BRCA1.721	103,472	RGB	RGB	RG	R
gRNA.BRCA1.722	103,555	RGB	RGB	RG	G
gRNA.BRCA1.723	103,559	RGB	RGB	RB	RGB
gRNA.BRCA1.724	103,606	RGB	RGB	RB	RG
gRNA.BRCA1.725	103,670	RGB	RGB	RB	BG
gRNA.BRCA1.726	103,699	RGB	RGB	RB	R
gRNA.BRCA1.727	103,700	RGB	RGB	RB	G
gRNA.BRCA1.728	103,738	RGB	RGB	RB	B
gRNA.BRCA1.729	103,755	RGB	RGB	BG	RGB
gRNA.BRCA1.730	103,756	RGB	RGB	BG	RG
gRNA.BRCA1.731	103,769	RGB	RGB	BG	RB
gRNA.BRCA1.732	103,790	RGB	RGB	BG	R
gRNA.BRCA1.733	103,811	RGB	RGB	BG	G
gRNA.BRCA1.734	103,824	RGB	RGB	R	RGB
gRNA.BRCA1.735	103,880	RGB	RGB	R	RG
gRNA.BRCA1.736	103,881	RGB	RGB	R	RB
gRNA.BRCA1.737	104,232	RGB	RGB	R	BG
gRNA.BRCA1.738	104,563	RGB	RGB	G	RGB
gRNA.BRCA1.739	104,659	RGB	RGB	G	RG
gRNA.BRCA1.740	105,115	RGB	RGB	G	RB

[0315]

gRNA.BRCA1.741	105,270	RGB	RGB	G	BG
----------------	---------	-----	-----	---	----

[0316] 表3.

[0317] 另一个示例使用折纸编码 (Origami Codes) (参见Nat Chem.2012年10月;4(10):832-9);参见图8。核酸折纸方法是本领域技术人员已知的。按照表3中所列的折纸多色方案来条码化gRNA尾。以形成预期的连续gRNA图案的方式隔离条码。根据该方案,折纸是gRNA尾探针。各折纸多色探针具有针对gRNA的特异性接合,意味着对于741个折纸探针,我们有741个独特的gRNA尾序列。折纸多色探针具有3个荧光标记物的接合位置,其可通过高分辨率或超高分辨率成像,即点1,点2,点3来物理分辨。各点具有与之相关的多个颜色编码。这些编码的组合提供对折纸多色探针的种类。

[0318] 通过添加其相应的荧光标记的折纸多色探针来检测gRNA条码。显示了实际连续gRNA图案。任何未预期的图案将被鉴定为基因组重排。与分辨基因组区域的区位RGB编码示例相比,折纸编码可在单个gRNA水平上分辨,提供对基因组区域或基因座上排列的高得多且详细的鉴定。

[0319] Her2试验

[0320] 所有侵入性乳腺癌都推荐使用原位杂交技术确定Her2状态。对目前批准的临床诊断的现有综述突出了不同临床试验的现有挑战(参考文献:Franchet C,Filleron T,Cayre A,Mounié E,Penault-Llorca F,Jacquemier J,Macgrogan G,Arnould L,Lacroix-Triki M.瞬时-质量荧光原位杂交作为乳腺癌中HER2测试的新工具:比较研究(Instant-quality fluorescence in-situ hybridization as a new tool for HER2 testing in breast cancer:a comparative study).Histopathology.2014年1月;64(2):274-83)。

[0321] 采用现有的试验报道了多种挑战。其一,传统FISH是漫长的过程,其一般需要一天来制备用于杂交的样品(脱石蜡、预处理、胃蛋白酶消化、变性),在12-24小时期间进行漫长的杂交,并且第二天充分洗涤杂交并进行观察。

[0322] 根据本文所述的示例性方面,方法不需要对样品的任何预处理、胃蛋白酶消化、或变性。Cas9-gRNA复合物易于特异性杂交至固定样品中的天然双链DNA。同样也易于洗去非杂交的复合物。因此,可在一天以内进行整个过程。这种时间线与通常进行FISH试验的Her2免疫组化试验一致。此外,编码方法将允许IHC和Cas9-gRNA FISH同时进行。

[0323] 试验之间的差异和精确性也被报道为一个问题。考虑到乳腺癌治疗的高成本,通常长达数年,错误的治疗造成明显的社会和经济影响。根据本文所述的某些方面,以提供冗余性和共定位的方式编码Cas9-gRNA。可与偏差相比区分阳性样品。此外,本文所述的方法允许对样品进行再探测,即,可去除探针,同时Cas9-gRNA仍然结合至靶DNA,并且可在更严谨的条件下进行新一轮探测。

[0324] 现有的FISH试验耗费几百美元(参考文献:万维网ncbi.nlm.nih.gov/pmc/articles/PMC2706184/)。这部分是由于探针所使用的材料,其在一些情况中是插入大载体的非常长的DNA拷贝,具有插入荧光分子(PathVysion)、或昂贵的肽核酸(PNA)探针(pharmDx)、或使用抗体和二抗来检测DNA探针(INFORM)。本文所述的方法将带来与进行FISH试验的常规Her2 ICH测试(约100美元)一致的成本。Cas9是细菌来源的蛋白质并且从廉价细菌系统中高效表达并纯化。其也是市售可得的。寡核苷酸阵列合成器中合成gRNA模板是目前合成DNA的最廉价方式(0.0004美元/碱基,市价)。在阵列合成之后,可使用PCR来扩增寡核苷酸,从而产生可无限廉价再扩增的gRNA模板的集合。体外转录是非常高效的,每个

DNA分子产生500-1000个RNA分子。

[0325] Her2试验将每个细胞的Her2基因座的平均拷贝数报告成染色体17的比率。染色体17的最常见报告物靶向着丝粒区域的 α 随体重复。这是已经较好表征并且含有超过1000个单体重复的区域(参见Waye JS, Willard HF. 来自人染色体17的 α 随体DNA的结构、组织和序列:通过不相等交换进化的证据和人X染色体共有的祖先五聚体重复 (Structure, organization, and sequence of alpha satellite DNA from human chromosome 17: evidence for evolution by unequal crossing-over and an ancestral pentamer repeat shared with the human X chromosome). Mol Cell Biol. 1986年9月; 6(9):3156-65)。重复特别适于本文所述的Cas9-gRNA方法。在这种情况下,只需要几个不同的gRNA。表4标识gRNA靶标的列表。该列表中充分表征的D17Z1基因座中提取,其他CEP17探针基于此。该列表也覆盖了由以下报道的变体:O'Keefe CL, Warburton PE, Matera AG, α 随体DNA变体的寡核苷酸探针可通过FISH区分同源染色体 (Oligonucleotide probes for alpha satellite DNA variants can distinguish homologous chromosomes by FISH). Hum Mol Genet. 1996年11月; 5(11):1793-9。然而,采用本文所述的方法,可使用仅11种gRNA来靶向这些重复中的大部分,因为Sac9一般在其5'末端附近有1-2个错配(表5)。

[0326]

靶标名称	靶序列
CEN17.1	GAGCGCTTTCAGGCCTGTGGTGG (SEQ ID NO:73)
CEN17.2	GAGGGCTTTGAGGCCTGTGGTGG (SEQ ID NO:74)
CEN17.3	GAGGGCTTTGTGGTTTGTGGTGG (SEQ ID NO:75)
CEN17.4	GGAATCTGCAAGTGGATATGTGG (SEQ ID NO:76)
CEN17.5	GTGTTGAAACTCTCTTTTTGTGG (SEQ ID NO:77)
CEN17.6	GTTTCCAATCACTCTTTGTGTGG (SEQ ID NO:78)
CEN17.7	GTTTGGAACACTCTTGTTGTGG (SEQ ID NO:79)
CEN17.8	GTTTTGAAACTCTCTTTCTGTGG (SEQ ID NO:80)
CEN17.9	ACACTGCTCTATCCATAGGAGG (SEQ ID NO:81)
CEN17.10	AGATATTTGGACCGCTCTGAGG (SEQ ID NO:82)

[0327]

	NO:82)
	AGCGCTTTCAGGCCTGTGGTGG (SEQ ID
CEN17.11	NO:83)
	AGGAATGTTCAACTCTGTGAGG (SEQ ID
CEN17.12	NO:84)
	AGGGCTTTGAGGCCTGTGGTGG (SEQ ID
CEN17.13	NO:85)
	AGGGCTTTGTGGTTTGTGGTGG (SEQ ID
CEN17.14	NO:86)
	CATCACAGAGAAGCTTCTGAGG (SEQ ID
CEN17.15	NO:87)
	CTGCATTCAACTCACAGTGTGG (SEQ ID
CEN17.16	NO:88)
	GAAAGGAAAGTTCAACTCGGGG (SEQ
CEN17.17	ID NO:89)
	GAATCTGCAAGTGGATATGTGG (SEQ ID
CEN17.18	NO:90)
	GAATGCAAACATCACGAAGAGG (SEQ
CEN17.19	ID NO:91)
	GCATATTTGGACCTCTTTGAGG (SEQ ID
CEN17.20	NO:92)
	GCTTCTGTTTAGTTCTGTGCGG (SEQ ID
CEN17.21	NO:93)
	GCTTCTGTTTAGTTCTGTGCGG (SEQ ID
CEN17.22	NO:94)
	GGACATTTGGAGGGCTTTGAGG (SEQ ID
CEN17.23	NO:95)
	GGACGTTTGGAGGGCTTTGTGG (SEQ ID
CEN17.24	NO:96)

[0328]

CEN17.25	GGAGATTTGGAGCGCTTTGAGG (SEQ ID NO:97)
CEN17.26	GGATATTTAGGCCTCTCTGAGG (SEQ ID NO:98)
CEN17.27	GGATATTTGGACCACTCTGTGG (SEQ ID NO:99)
CEN17.28	GGATATTTGGACCTCTCTGAGG (SEQ ID NO:100)
CEN17.29	GGGATCATTGCACTCTTTGAGG (SEQ ID NO:101)
CEN17.30	TACTACCATAGGCCTAAAGCGG (SEQ ID NO:102)
CEN17.31	TATTTGTAGAATGTGCAAGTGG (SEQ ID NO:103)
CEN17.32	TCCAAAGACATCTTCGGAGAGG (SEQ ID NO:104)
CEN17.33	TCCAACGAAATCCTCAGAGAGG (SEQ ID NO:105)
CEN17.34	TCCAACGAAATCCTCAGAGAGG (SEQ ID NO:106)
CEN17.35	TCCAACGAAATCCTCAGAGCGG (SEQ ID NO:107)
CEN17.36	TCCAACGAAATCTTCAAAGAGG (SEQ ID NO:108)
CEN17.37	TCCAACGAAATGCTCAGAGAGG (SEQ ID NO:109)
CEN17.38	TCGAACGAAGGACACAGAGTGG (SEQ ID NO:110)
CEN17.39	TCGAACGAAGGCCACAGAGTGG (SEQ

[0329]

	ID NO:111)
	TCTGCAAGTGGACATTTGGAGG (SEQ ID
CEN17.40	NO:112)
	TCTGCAAGTGGACGTTTGGAGG (SEQ ID
CEN17.41	NO:113)
	TGGAGCGCTTTCAGGCCTGTGG (SEQ ID
CEN17.42	NO:114)
	TGGAGGGCTTTGAGGCCTGTGG (SEQ ID
CEN17.43	NO:115)
	TGGAGGGCTTTGTGGTTTGTGG (SEQ ID
CEN17.44	NO:116)
	TGTTGAAACTCTCTTTTTGTGG (SEQ ID
CEN17.45	NO:117)
	TTGTTGTGGAATGTGCAAGTGG (SEQ ID
CEN17.46	NO:118)
	TTTCCAATCACTCTTTGTGTGG (SEQ ID
CEN17.47	NO:119)
	TTTCTGTGGCATCTGCAAGGGG (SEQ ID
CEN17.48	NO:120)
	TTTGGAACACTCTTGTTGTGG (SEQ ID
CEN17.49	NO:121)
	TTTGTGTAGAATCTGCAAGTGG (SEQ ID
CEN17.50	NO:122)
	TTTGTGTGGAATCTGCAAGTGG (SEQ ID
CEN17.51	NO:123)
	TTTTCGTAGTGTCTACAAGTGG (SEQ ID
CEN17.52	NO:124)
	TTTGTGAACTCTCTTTCTGTGG (SEQ ID
CEN17.53	NO:125)

[0330]

TTTTTCCAGAATCTGCAAGTGG (SEQ ID CEN17.54 NO:126)
TTTTTCTAGAATCTGCAAGTGG (SEQ ID CEN17.55 NO:127)
TTTTTGCAGGATCTACAAGTGG (SEQ ID CEN17.56 NO:128)
TTTTTGTACAATCTACAAGTGG (SEQ ID CEN17.57 NO:129)
TTTTTGTAGAAACTGCAAGGGG (SEQ ID CEN17.58 NO:130)
TTTTTGTAGAAACTGCAAGTGG (SEQ ID CEN17.59 NO:131)
TTTTTGTAGGATCTGCAAGTGG (SEQ ID CEN17.60 NO:132)
TTTTTGTGGAATCTGCAAGTGG (SEQ ID CEN17.61 NO:133)

[0331] 表4.

[0332]

gRNA 名称	gRNA 靶序列
Cen17gRNA.1	GAGGGCTTTGAGGCCTGTGG (SEQ ID NO:134)
Cen17gRNA.2	GTGTTGAAACTCTCTTTTGG (SEQ ID NO:135)
Cen17gRNA.3	GACACTGCTCTATCCATAGG (SEQ ID NO:136)
Cen17gRNA.4	GAGGGCTTTGAGGCCTGTGG (SEQ ID NO:137)
Cen17gRNA.5	GGAATCTGCAAGTGGATATG (SEQ ID NO:138)

[0333]

Cen17gRNA.6	GGCATATTTGGACCTCTTTG (SEQ ID NO:139)
Cen17gRNA.7	GTACTACCATAGGCCTAAAG (SEQ ID NO:140)
Cen17gRNA.8	GTCCAACGAAATCCTCAGAG (SEQ ID NO:141)
Cen17gRNA.9	GTCTGCAAGTGGACATTTGG (SEQ ID NO:142)
Cen17gRNA.10	GTTGTTGTGGAATGTGCAAG (SEQ ID NO:143)
Cen17gRNA.11	GTTTGTGTAGAATCTGCAAG (SEQ ID NO:144)

[0334] 表5.

[0335] 另外,本文所述的方法允许筛选等位基因变体,其是个性化医疗中的必要信息,并且可能影响治疗。例如,Her2变体I655V已经涉及降低他莫昔芬用作乳腺癌治疗的功效,其应该被考虑(参见Chang NW,Chen DR,Chen FN,Lin C,Wu CT.HER2密码子655G-等位基因与用他莫昔芬治疗的乳腺癌患者中血浆高密度脂蛋白水平降低相关(HER2 codon 655 G-allele is associated with reductions in plasma high-density lipoprotein levels in breast cancer patients treated with tamoxifen).J Investig Med.2011年12月;59(8):1252-7.doi:10.231/JIM.0b013e3182354923)。2种鉴定的gRNA靶标可用于鉴定变体(靶标1:TCTGACGTCCATCATCTCTGCGG (SEQ ID NO:145)和靶标2GCCAACCACCGCAGAGATGATGG (SEQ ID NO:146))。使用一组4种带有等位基因特异性条码的gRNA(2个/靶标),提供方法来区分异亮氨酸等位基因(ATC)或缬氨酸等位基因(GTC),作为标准Her2试验的部分。

[0336] 应理解可靶向染色体17的其他区域。例如,Top2A位于Her2基因座附近并且已经观察到与Her2共扩增,其也可影响合适治疗的选择。参见Smith K,Houlbrook S,Greenall M,Carmichael J,Harris AL.拓扑异构酶II α 在人原发乳腺癌和乳腺癌细胞系中与erbB2共扩增:与m-AMSA的关系和米托蒽醌敏感性(Topoisomerase II alpha co-amplification with erbB2 in human primary breast cancer and breast cancer cell lines: relationship to m-AMSA and mitoxantrone sensitivity).Oncogene.1993年4月;8(4):933-8。

[0337] 下文描述了用于筛选Her2/CEN17的FISH方案的示例。如果免疫组化和FISH试验都在同一样品上进行,细胞或组织样品在4℃下在含10%甲醛的PBS中持续1小时至16小时来固定在显微镜相容的支持物上(例如,干净的铝硼硅酸盐显微镜载玻片或显微镜盖玻片)。或者,样品在-20℃下在100%甲醇中固定20分钟至1小时(甲醇在保留样品中DNA,同时提取

许多蛋白质上更高效)。在固定至载玻片之前,石蜡包埋的组织切片应该脱石蜡,其一般通过在二甲苯中孵育5分钟,之后2次乙醇洗涤,和一次用水洗涤来完成。固定的样品然后用PBST(含0.5%吐温-20的PBS缓冲液)洗涤一次,然后用含0.5%曲通X-100的PBS孵育5分钟。样品然后在PBST中洗涤并且向含5mM $MgCl_2$ 的PBST中的样品添加Cas9-gRNA-标记物复合物混合物,并在37℃下孵育2小时至16小时(即,混合物同时含有Her2和CEN17以及它们相应的鉴定标签的组)。孵育之后,通过在37℃下在PBST中洗涤3次来去除未结合的Cas9。样品用抗衰减固定试剂和DAPI(例如,含DAPI的ProLong或VectaShield)固定,并密封。使用油浸63X物镜对样品进行成像。或者,如果保持避光并且在4℃下时,样品稳定持续数周。然后计数至少20个肿瘤区域中的核,并且针对它们中的每个记录Her2和CEN17共定位灶点的数量。可按照由2013ASCO/CAP指南所述的现有临床实践来计算并报告Her2与CEN17之比(参见Wolff AC,Hammond ME,Hicks DG,Dowsett M,McShane LM,Allison KH,Allred DC,Bartlett JM,Bilous M,Fitzgibbons P,Hanna W,Jenkins RB,Mangu PB,Paik S,Perez EA,Press MF,Spears PA,Vance GH,Viale G,Hayes DF;美国临床肿瘤学协会;美国病理学学院。乳腺癌中人表皮生长因子受体2测试的推荐:美国临床肿瘤学协会/美国病理学学院临床实践指南更新(Recommendations for human epidermal growth factor receptor 2 testing in breast cancer:American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update).J Clin Oncol.2013年11月1日;31(31):3997-4013)。

[0338] 实施例3

[0339] 图1涉及对小鼠胚胎成纤维细胞核的大随体和小随体,以及端粒区域的原位Cas9-gRNA探测。用荧光标记的UTP合成3种gRNA以靶向小鼠大随体重复(Cy5)、小随体重复(Alexa-488)、和端粒(Cy3)。(图1图例:大随体(A),小随体(B),端粒(C),基因组DNA的DAPI染色(D),和覆盖图(E))。gRNA在按照本文所述的方案添加到PFA-固定的样品之前与Cas9复合。图案匹配这些靶标的预期探测(参见Guenatri M,Bailly D,Maison C,Almouzni G.小鼠着丝粒和臂间随体重复形成不同功能异染色质(Mouse centric and pericentric satellite repeats form distinct functional heterochromatin).J Cell Biol.2004年8月16日;166(4):493-505)。在装配63X/1.40油浸物镜,和LED光源以及针对各荧光通道合适的滤器的Zeiss Axio Observer Z1上拍摄图像。以下图例与图1相关:A,大随体;B,小随体;C,端粒;DNA的DAPI染色;E,覆盖。

[0340] 实施例4

[0341] 图2涉及图1的对照实验。靶向小鼠大随体重复(Cy5)、小随体重复(Alexa-488)、和端粒(Cy3)的常用FISH寡核苷酸,在使用时,使用与图1中相同的Cas9探测方案。除了对DNA的DAPI染色(D)以外,其他荧光信号比图1中捕获的图像更弥漫,并且必须在显微镜上高度对比。(图2图例:大随体(A),小随体(B),端粒(C),基因组DNA的DAPI染色(D),和覆盖图(E))。缺少变性防止这些FISH寡聚体与它们的靶标杂交,这些靶标也在核外聚集。在装配63X/1.40油浸物镜,和LED光源以及针对各荧光通道合适的滤器的Zeiss Axio Observer Z1上拍摄图像。以下图例与图1相关:A,大随体;B,小随体;C,端粒;DNA的DAPI染色;E,覆盖。

[0342] 实施例5

[0343] 图3涉及Cas9-gRNA凝胶移位和切割试验,显示天然Cas9结合和切割活性互相独立

并且依赖于镁离子的存在或缺乏。如下所示制备反应:2pmol的核酸酶活性Cas9 (NEB), 2pmol的合成的Grna (在L22116位置处靶向Lambda DNA), 和0.2pmol的2kb PCR扩增的DNA片段 (来自Lambda DNA L21333-L23332), 在存在5Mm $MgCl_2$ (泳道1) 或不存在镁 (泳道2) 的情况下, 混合并在37℃下孵育1小时。泳道1显示镁存在下的二级切割产物, 而泳道2显示没有切割产物, 但显示由仍然与之结合的复合物导致的小幅上移。分别在与泳道1和2相似的条件下进行泳道3和泳道4, 除了从反应中省去Cas9蛋白以外。泳道3和4中没有观察到切割和偏移。

[0344] 实施例6

[0345] 图4A-4L涉及gRNA尾条码和探测的示例。图4A显示Cas9蛋白 (1) 和复合的gRNA (2), 具有伸出的四核苷酸环 (3) 和伸出的含条码的gRNA尾 (4)。该条码可通过含可检测部分, 或标记物 (6) 的杂交探针 (5) 检测。图4B显示可由超过一种条码编码gRNA尾。在图4B中, 显示了2个条码 (4和7), 其可由它们含有可检测部分或标记物 (6) 的相应杂交探针 (5和8) 检测到。在图4C中, 杂交探针可含有多个可检测部分或标记物 (9) 以放大信号。在图4D中, 可检测部分可能不是可直接检测的 (10) 并且可能需要使用二级检测试剂 (11), 其对可检测部分有特异性亲和性。该二级试剂含有可检测部分或标记物 (12) 用于检测或放大检测到的信号。图4E-4I描述了通过滚环扩增探测gRNA尾的方式。图4E显示Cas9蛋白 (1) 和复合的gRNA (2), 具有突出的四核苷酸环 (3) 和含条码的gRNA尾 (4)。由环状杂交探针 (13) 检测条码, 在探针的一个区域中具有针对条码的亲和性并且扎起另一个区域中具有标记的-探针杂交靶位点 (14)。图4F显示环状探针 (13) 可用作滚环扩增聚合酶 (15) 的模板以延伸gRNA尾 (16)。图4G显示滚环扩增产生定位的放大的杂交探针 (17), 其具有多个紧密定位的标记的-探针杂交靶位点 (18)。图4H显示可通过添加可检测探针 (19) 产生信号放大的gRNA尾探针 (20) 使标记的-探针杂交靶位点可被检测到。图4I显示与图4H相似的滚环扩增的探针 (20), 其被生成并标记 (21) 而环状探针不与gRNA条码杂交, 从而避免在gRNA自身上发生滚环扩增和标记步骤。这种脱离gRNA生成的探针可通过区域 (5) 与图4A (4) 所示的gRNA尾的条码区域杂交。图4J-4L描述了通过核酸自组装探测gRNA尾的方式。图4J显示了以连续方式互相线性组装核酸探针。第一片段包含条码杂交区域 (5), 组装区域 (21) 和标记物 (6) 首先杂交至gRNA尾条码 (图4A, 4)。添加与21或22或23部分互补的标记的组装片段 (22和23) 的混合物并且只要反应中存在部分互补的片段就会自组装。在此仅显示了部分结构。这之前所述的杂交链反应 (HCR) 相似。参见Dirks RM, Pierce NA. 通过杂交链反应引发的扩增 (Triggered amplification by hybridization chain reaction). *Proc Natl Acad Sci U S A*. 2004年10月26日; 101 (43): 15275-8。图4K显示了高分支核酸探针枝状聚合物结构的自组装。第一片段包含条码杂交区域 (5), 组装区域 (24) 和标记物 (6) 首先杂交至gRNA尾条码 (图4A, 4)。添加与24或25或26部分互补的标记的组装片段 (24, 25和26) 的混合物并且只要反应中存在部分互补的片段就会自组装成高分支结构。在此仅显示了部分结构。之前已经描述了这种类型的核酸枝状聚合物。参见Li Y, Tseng YD, Kwon SY, D'Espaux L, Bunch JS, McEuen PL, Luo D. 枝状聚合物-状DNA的受控组装 (Controlled assembly of dendrimer-like DNA). *Nat Mater*. 2004年1月; 3 (1): 38-42。图4L显示了组装核酸探针的分支以线性扩增信号。第一片段包含条码杂交区域 (5), 组装区域 (27) 首先杂交至gRNA尾条码 (图4A, 4)。添加与27或29或30部分互补的标记的组装片段 (29和30) 的混合物并且可自组装。可检测的探针

(28)可杂交至组装片段(27)和(30)的部分,只要反应中存在部分互补的片段。在此仅显示了部分结构。这之前所述的分支DNA(bDNA)相似。参见Collins ML,Irvine B,Tyner D,Fine E,Zayati C,Chang C,Horn T,Ahle D,Detmer J,Shen LP,Kolberg J,Bushnell S,Urdea MS,Ho DD.用于对低于100个分子/ml的核酸靶标进行定量的分支DNA信号放大试验(A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml).Nucleic Acids Res.1997年8月1日;25(15):2979-84。

[0346] 应理解可检测部分(例如,6,9,10,12)可提供多种检测方式,如荧光、化学发光或发色,或共振。应理解杂交探针(例如,5,8,9,14)可由多种核酸组分,如DNA、RNA、修饰的DNA、修饰的RNA制成。

[0347] 实施例7

[0348] 图5A涉及与侧流测试系统的表面接合的Cas9-gRNA复合物。研究中的DNA的群加载到系统中并且侧流将其置换到测试区,靶标-特异性Cas9-gRNA在该区结合至特异性DNA。其余的DNA保持流过试验结束并且被控制区(Control Zone)捕获,而靶DNA仍然在测试区与Cas9-gRNA复合物结合。DNA检测可如下进行:在含有可检测部分的特异性或通用寡核苷酸引物存在下,通过DNA的酶促扩增(例如,等位扩增),通过核酸染色,或将可检测探针杂交至DNA,或通过可检测探针共价接合至DNA。通常用于侧流试验的可检测部分包括金或银纳米颗粒。其他部分如生物素、地高辛、二硝基苯基、荧光素通常以二级检测试剂(例如,针对生物素、地高辛、二硝基苯基或荧光素有特异性的金或银包被的抗体)联用。图5B涉及使用具有不同性质来滞留DNA或蛋白质,如疏水性或电荷的2种材料的侧流系统。例如,硝酸纤维素膜滞留蛋白质比DNA更有利,而尼龙膜比蛋白质更优先结合DNA。在加载待研究的DNA的群之后,Cas9-gRNA复合物结合靶DNA。与DNA结合的Cas9在测试区中被DNA滞留膜滞留,而未结合的Cas9将通过并结合至控制区的蛋白质滞留膜。进行检测以检测带DNA的Cas9-gRNA复合物。以本文所述的方式探测gRNA。或者,Cas9蛋白本身可携带部分(例如,金纳米颗粒、银纳米颗粒、生物素、地高辛、二硝基苯基、荧光素)。图5C涉及侧流系统,其中首先通过分子相互作用(例如,电荷、疏水性、共价相互作用、或亲和相互作用如生物素-链霉亲和素)在测试区的表面上捕获。Cas9-gRNA然后在测试区结合至表面-捕获的靶标DNA。剩余未反应的Cas9-gRNA通过并且在控制区被捕获。进行检测以检测带DNA的Cas9-gRNA复合物。以本文所述的方式探测gRNA。或者,Cas9蛋白本身可携带部分(例如,金纳米颗粒、银纳米颗粒、生物素、地高辛、二硝基苯基、荧光素)。

[0349] 实施例8

[0350] 使用引导RNA/Cas9复合物探测靶DNA

[0351] 体外设计、制备、表达和纯化引导RNA。合成单链DNA(ssDNA)寡核苷酸(IDT或定制阵列芯片)。这些寡聚体含有用于体外RNA合成的T7转录起始位点,和形成待用于探测的3'末端尾的延伸。进行T7RNA合成,纯化RNA。

[0352] 根据一个方面,设计延伸的序列或尾以最小化空间位阻,具有低复杂性和低自由能以实现使用DNA-PAINT的超高分辨率成像(其可实现2.7nm分辨率,其是8个DNA碱基,因此低于Cas9的20bp迹线)。设计的引导RNA具有20个不同的PAINT停靠位点,其长度为9聚体,并含有不同的A、T和C碱基(非G)排列,而PAINT探针是互补序列加荧光团。

[0353] 通过洗去探针并且在成像之间加载新探针,可实现20种探针/荧光团,即,每个荧光团可用20种不同gRNA成像。因此,4种荧光团能够在80个gRNA之间区分。

[0354] 引导RNA结构的一个实施方式是(5'至3') gRNA-UUUUU-PAINT停留(dock)。引导RNA序列是与SpCas9常联用的单个引导RNA的最小长度。

[0355] 对于高分辨率成像,延伸的序列或尾设计成最小化空间位阻,具有低复杂性和低自由能。然而,对于高分辨率成像,使用较长的退火区域。延伸的序列或尾的结构可用作条码,其可用序列特异性寡核苷酸,或锁式探针之后进行滚环扩增来探测。

[0356] SpCas9在大肠杆菌中表达。使用Gibson等位组装来组装并克隆Cas9编码基因到适于在大肠杆菌中表达的质粒中。孵育之后,裂解细胞并纯化蛋白质。

[0357] 通过在没有进一步制备的情况下裂解细胞来获得靶DNA。因此,样品包含来自染色体和染色体外(例如,质粒)来源的DNA的混合物。

[0358] Cas9与gRNA混合并且添加至DNA样品中持续足够的时间,使Cas9、gRNA和靶核酸形成复合物(约15分钟),之后添加检测探针。然后获得成像数据(数秒至数分钟)。

[0359] 对于纳米孔检测,Cas9-gRNA复合物在dsDNA片段上设计的间隔处结合。然后DNA易位通过或靠近纳米孔(或纳米间隙电极)。测量电流变化:dsDNA将在特定电流下运行,而与DNA结合的复合物将部分阻断电流,我们将其记录为电流尖峰。通过随时间分析这些尖峰,可推导靶DNA上Cas9/gRNA的位置并与基于引导RNA设计的预测位置比较。

[0360] 使用上述方法,可一次检测多种靶标以获得关于DNA靶标性质的信息,如着丝粒的重复区域、与这些重复连接的染色体的种类、药物抗性基因的种类(包含于细菌基因组或质粒)、可移动元件的种类(例如,转座子、药物抗性盒)、疾病相关基因的特定等位基因(例如,癌基因,自身免疫,神经变性)等。

[0361] 实施例9

[0362] 使用引导RNA/Cas9复合物探测靶DNA

[0363] 人基因组DNA(诺瓦基公司)在pH 5.5的0.5M MES缓冲液中稀释至约0.2ng/ul,并在乙烯基硅烷包被的基材(巴黎的Genomic Vision公司)上进行分子梳理,按照Michalet等所述(动态分子梳理(Dynamic Molecular Combing),Science 1999)。简言之,这包括将乙烯基硅烷包被的盖玻片浸入含DNA的溶液中,然后以固定速度将其拉出,与Langmuir Blodgett装置相似。这用于通过“弯液面”机制将DNA拉伸到基材上。当DNA完全脱离溶液时,其在10000微焦/cm²下经过UV交联。这在表面上产生大量单向对齐的DNA。如果小心制备DNA,可使用DNA染色如YOYO-1观察到兆碱基长度的DNA。

[0364] 在基材上拉伸的基因组DNA用缓冲液润湿,然后向基材中加入预形成(通过在37°C下预孵育10分钟)的gRNA/cas9(NEB)并允许反应1小时。然后洗去过量复合物。引导RNA设计成具有与着丝粒序列互补的部分,并且引导物的3'端设计成具有与探针序列互补的尾核酸序列。本领域技术人员将易于理解可针对任何需要的基因组序列设计引导RNA。包括引导RNA和尾序列,启动子和终止信号序列(IDT)的序列可用于体外转录系统中以合成具有与来自模板的探针序列互补的尾序列的引导RNA。使用Zymo清洁/浓缩器来纯化具有与探针序列互补的尾序列的引导RNA。然后在添加到拉伸的DNA之前,如上所述用Cas9孵育转录的RNA。

[0365] 用BLOCKAID(英杰公司)处理载玻片。具有与gRNA上的尾的互补性的16nt DNA探针然后与复合物在非严谨条件(4XSSC,50%甲酰胺,Blockaid)下反应。使用的探针在2端用

Atto 657N染料标记(来自英杰公司的定制合成订单)。合成反应在4℃下放置过夜。然后洗涤载玻片以去除过量染料并且在TIRF显微镜上成像。

[0366] 因为DNA靶标仍然是双链的,可在Atto647N成像之前或之后用YOYO-1染料染色。通过使用红色激光和合适滤器检测引导RNA上的标记物,并且通过使用蓝色激光和合适滤器检测DNA染色。通过宽视野TIRF显微镜进行成像。

[0367] 图6A是显示与表面上延长或预拉伸的双链人基因组DNA(线)结合的着丝粒特异性引导RNA(gRNA)/Cas9复合物(点)的结果的图像。图6B是gRNA/Cas9(点)结合至人着丝粒DNA(灰色)的超高分辨率图像。

[0368] 如图6A所示,粗略估计,探针靶向的着丝粒DNA占基因组的约1%。检验多个视场并且通常用沿线相关的探针标记点观察场,如图6A所示。染色的DNA显示载玻片具有许多不显示沿线相关的标记物的拉伸的DNA分子,从而证明本文所述的方法可使用可检测gRNA/Cas9复合物鉴定包括多种核酸的样品内的特定靶核酸。

[0369] CRISPR/CAS9缓冲液包含20mM HEPES,100mM NaCl,5mM MgCl₂,0.1mM EDTA,并且在25℃下pH为6.5。杂交缓冲液包含200ul“纯”甲酰胺,20ul 10%SDS,120ul blockaid,40ul的20xssc和20ul水。使用的Cas9是终浓度为25-30nM的Cas9核酸酶,化脓链球菌(*S. pyogenes*) (NEB)。使用的标记的寡核苷酸(IDT)浓度为约800nM。

[0370] 图6C设计Cas9体外结合试验。全长Lambda DNA(约48kb)用Cas9单独探测或与CRISPR-gRNA复合并在乙烯基硅烷官能化的玻璃表面上拉伸。Cas9然后由藻红蛋白-偶联的抗体(深色)标记并且DNA由YOYO-1(浅色)标记。在装配TIRF模式的100x/1.47油浸物镜的Leica DM1600上获取图像。

[0371] 实施例10

[0372] 图7涉及应用于BRCA1基因座的区位红-绿-蓝(RGB)条码化系统。BRCA1和BRCA2基因内的大重排是女性和男性中的乳腺癌和卵巢癌易感性的重要标志物。重排可提供关于发展癌症的风险,和合适治疗的信息。(Judkins T,Rosenthal E,Arnell C,Burbidge LA,Geary W,Barrus T,Schoenberger J,Trost J,Wenstrup RJ,Roa BB.BRCA1和BRCA2中大重排的临床显著性(Clinical significance of large rearrangements in BRCA1 and BRCA2).Cancer.2012年11月1日;118(21):5210-6.;Liede A,Karlan BY,Narod SA.BRCA1或BRCA2种系突变的男性携带者的癌症风险:文献综述(Cancer risks for male carriers of germline mutations in BRCA1 or BRCA2:a review of the literature).J Clin Oncol.2004年2月15日;22(4):735-42)。这些大重排的范围是非常广泛的,并且一般使用对PCR扩增的DNA的Sanger测序来表征这些重排。沿基因座使用特异性颜色图案(表3,区位编码),我们的gRNA/Cas9探测策略以高密度标记各BRCA基因座的全长(每100bp大约1个探测的位点)。对于在基因座内重排的任何区域,颜色图案将改变,使得能够鉴定这些大重排。我们的方法提供了测序的更简单替代。我们的方法也用于检测其他癌症,如急性骨髓性白血病,发育疾病,如丛综合征(Bushy syndrome),和神经发育疾病,如自闭症中的类似重排。适用于该方法的另一种类别的重排包括免疫细胞受体的VDJ重组。

[0373] 实施例11

[0374] 图8涉及应用于由gRNA/cas9复合物靶向的特定位置的DNA折纸条码。各条码对于gRNA靶向的位置而言是独特的。BRCA1和BRCA2基因易受各种幅度的基因组重排的影响,包

括较小的插入和删除,其仅可通过Sanger测序检测到。使用对各gRNA/Cas9靶位点特定的条码,我们的gRNA/Cas9探测策略以高密度标记各BRCA基因座的全长(每100bp大约1个探测的位点)。该条码采取荧光折纸条码,如前所述(Lin C, Jungmann R, Leifer AM, Li C, Levner D, Church GM, Shih WM, Yin P. 亚微米几何编码的从DNA自组装的荧光条码 (Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA). Nat Chem. 2012年10月; 4 (10): 832-9)。通过对各条码(表3, 折纸编码)进行解码,我们鉴定各gRNA/Cas9及其在基因座上的位置,其提供了关于基因座内插入、删除和重排区域的信息。当决定合适的癌症治疗时,临床医师可考虑这种分辨率水平。我们的方法提供了测序的更简单替代。我们的方法也用于检测其他癌症、发育疾病、和神经发育疾病中的类似重排。适用于该方法的另一种类别重排包括免疫细胞受体的VDJ重组。

[0375] 实施例12

[0376] 图9A-9B涉及从gRNA/Cas9切口的聚合酶延伸。图9A显示了由gRNA和Cas9的D10A突变体切割的靶DNA的顶部链。图9B显示了由gRNA和Cas9的H840A突变体切割的靶DNA的底部链。弯曲的箭头显示从切口的引物延伸方向。可使用引物延伸来标记靶向的区域或启动对靶向的区域的测序。我们的gRNA/Cas9切口和测序方法允许对基因组上的测序起始位点进行精确靶向定位。简言之, gRNA/Cas9切口酶在感兴趣的靶向位置处在DNA链中产生切口。然后由多种方法之一来置换gRNA/Cas9,如去污剂、变性剂或温度。然后加入具有链置换活性的DNA聚合酶,和标记的dNTP,从而从切口位点延伸DNA。大量的感兴趣基因组区域适用于这种gRNA/Cas9切口启动方法。重复区域,如着丝粒DNA本身对于测序和比对而言是困难的,由于它们的大量短重复。基因组重排通常含有较小的突变。另外,许多基因组重排难以表征,并且重排或融合的位置未知。

[0377] 实施例13

[0378] 图10是使用CHOPCHOP来找到PAM位点时提供的输出图。CHOPCHOP是许多在线工具使用算法之一,以发现给定基因座上的gRNA/Cas9靶标和脱靶并对其打分(Tessa G. Montague; Jose M. Cruz; James A. Gagnon; George M. Church; Eivind Valen. (2014). CHOPCHOP: 用于基因组编辑的CRISPR/Cas9和TALEN网页工具 (CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing). Nucleic Acids Res. 42. W401-W407)。该图显示在Her2外显子上搜索靶标的图形输出。小三角形标识Her2外显子上的gRNA/Cas9靶位点。靶序列的列表可经提取,管理 (curate) 以仅保留没有或有少量脱靶的靶标,并且然后用于设计gRNA。这种工具用于寻找其他基因座的gRNA/Cas9靶标。其他工具可用于讯在内含子和外显子上,或基因座以外的靶标。

[0379] 实施例14

[0380] 图11涉及使用高保证聚合酶链反应 (PCR) 对用于体外转录引导RNA的DNA模板的组装。该策略依赖于2个通用寡核苷酸: Fwd-T7-gRNA, 也包括用于体外转录的T7RNA聚合酶识别基序的部分的正向PCR引物; 和gRNA.split60, 通用gRNA支架。另外, 存在2个聚体寡核苷酸: Sp.gRNA.split60, 针对感兴趣靶标有特异性的序列; 和Rev-B1-gRNA.18, 也包括用于多重链监测的条码化的柄的反向PCR引物。也提供了建议的熔融温度(T_m)。这种设计是低成本的, 最小化扩增错误, 并且适用于小规模或大规模寡核苷酸。扩增后, 可通过PCR再扩增模板以生成并保持模板, 其比从头合成成本更低。PCR组装耗时不到1小时。在PCR之后, 通过向模

板DNA添加T7RNA聚合酶混合物由体外转录 (IVT) 来合成gRNA。

[0381] 实施例15

[0382] 图12是Cas9-gRNA靶向ALK易位的示意图。该图显示当已知2个融合区域时,可通过使用含gRNA/Cas9的特异性条码来检测融合存在。ALK基因座易产生染色体内和染色体间重排和倒置。存在7种具有不良临床结构的这类重排 (Solomon B,Varella-Garcia M,Camidge DR. ALK 基因重排: 非小细胞肺癌的分子限定亚组中的新治疗靶标 (ALK gene rearrangements: a new therapeutic target in a molecularly defined subset of non-small cell lung cancer). J Thorac Oncol. 2009年12月;4(12):1450-4)。我们的gRNA/Cas9探测策略使用与各基因座相关的颜色以高密度(大约1个探测位点/100bp)标记特定基因座区域。该策略可用于检测基因融合和倒置。当2个标记的基因座相邻时,它们相应的标记物被检测为共定位信号。在重排的情况中,2个信号不再共定位,并且是非常远的。在倒置的情况中,信号不共定位,但仍然在附近。为了增加置信度,标记第三基因座,其在基因融合中(由于ALK染色体间重排)或ALK倒置中提供了不同共定位组合。检测这种ALK缺陷在鉴定多种癌症,如间变性大细胞淋巴瘤 (ALK-NPM1融合),肺腺癌 (ALK-EML4融合和倒置),和某些小儿成神经细胞瘤中是重要的。存在对某些ALK重排的治疗。几种其他疾病通过基因融合表征并且适用于我们的方法。基因融合靶标的一些示例包括ABL1-BCR、AML1-RUNX1T1、AML1-ETV6、BCL-2-IGH、BCL-2-MLT、C-Myc-IGH、COL1A1-PDGFB、CycD1-IGH、ETV6-TRKC、ETV6-JAK、FLI1-EWS、PAX8-PPARG、PMI-NR1B1、TCR-RRTN2、SS18-SSX。

序列表

<110> 哈佛学院董事及会员团体 (President and Fellows of Harvard College)

<120> 用于对核酸探测并作图的RNA-引导的系统

<130> 010498_00633

<140> PCT/US2015/045805

<141> 2015-08-19

<150> US 62/039,341

<151> 2014-08-19

<160> 146

<170> PatentIn version 3.5

<210> 1

<211> 16

<212> DNA

<213> 人工

[0001]

<220>

<223> 图9A从缺口位点的聚合酶延伸

<220>

<221> misc_feature

<222> (1)..(9)

<223> n是a, c, g, 或t

<220>

<221> misc_feature

<222> (12)..(16)

<223> n是a, c, g, 或t

<400> 1

nnnnnnnnng gnnnnn

16

<210> 2

<211> 16

<212> DNA

<213> 人工

<220>
<223> 图9B从缺口位点的聚合酶延伸

<220>
<221> misc_feature
<222> (1)..(9)
<223> n是a, c, g, 或t

<220>
<221> misc_feature
<222> (12)..(16)
<223> n是a, c, g, 或t

<400> 2
nnnnnnnnnc cnnnnn
16

<210> 3
<211> 24
<212> DNA
<213> 人工

[0002] <220>
<223> Fwd-t7-gRNA

<400> 3
gaaattaata cgactcacta tagg
24

<210> 4
<211> 60
<212> DNA
<213> 人工

<220>
<223> 图11 Sp. gRNA. split60

<220>
<221> misc_feature
<222> (21)..(38)
<223> n是a, c, g, 或t

<400> 4
ttaatacgac tcactatagg nnnnnnnnnn nnnnnnnngt tttagagcta gaaatagcaa

60

<210> 5
<211> 81
<212> DNA
<213> 人工

<220>
<223> gRNA.split60

<400> 5
gttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt
60

ggcacccgagt cggtgctttt t
81

<210> 6
<211> 28
<212> DNA
<213> 人工

[0003]

<220>
<223> 图11 Rev_B1_gRNA. 18

<400> 6
accgagtcgg tgctttttta tacatcta
28

<210> 7
<211> 1368
<212> PRT
<213> 酿脓链球菌 (Streptococcus pyogenes)

<400> 7

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

	35		40		45													
	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu	Ala	Thr	Arg	Leu		
	50						55					60						
	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	Asn	Arg	Ile	Cys		
	65					70					75					80		
	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val	Asp	Asp	Ser		
					85					90					95			
	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu	Asp	Lys	Lys		
				100					105					110				
	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu	Val	Ala	Tyr		
			115					120					125					
[0004]	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys	Leu	Val	Asp		
	130						135					140						
	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala	Leu	Ala	His		
	145					150					155					160		
	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp	Leu	Asn	Pro		
					165					170					175			
	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu	Val	Gln	Thr	Tyr		
				180					185					190				
	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser	Gly	Val	Asp	Ala		
			195					200					205					
	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg	Leu	Glu	Asn		
	210						215					220						
	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu	Phe	Gly	Asn		

	225				230					235					240
	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys	Ser	Asn Phe
					245					250					255
	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr	Tyr Asp
				260					265					270	
	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr	Ala Asp
			275					280					285		
	Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	Ser Asp
		290					295					300			
	Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu	Ser	Ala Ser
	305					310					315				320
[0005]	Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu	Leu Lys
					325					330					335
	Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu	Ile	Phe Phe
				340					345					350	
	Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly	Gly	Ala Ser
			355					360					365		
	Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu	Glu	Lys	Met Asp
		370					375					380			
	Gly	Thr	Glu	Glu	Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp	Leu	Leu Arg
	385					390					395				400
	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His Leu
					405					410					415
	Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro Phe

	420	425	430
	Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile 435 440 445		
	Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp 450 455 460		
	Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu 465 470 475 480		
	Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr 485 490 495		
	Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser 500 505 510		
[0006]	Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys 515 520 525		
	Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln 530 535 540		
	Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr 545 550 555 560		
	Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp 565 570 575		
	Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly 580 585 590		
	Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp 595 600 605		
	Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr		

	610	615	620	
	Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala 625 630 635 640			
	His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr 645 650 655			
	Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp 660 665 670			
	Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe 675 680 685			
	Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe 690 695 700			
[0007]	Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu 705 710 715 720			
	His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly 725 730 735			
	Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly 740 745 750			
	Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln 755 760 765			
	Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile 770 775 780			
	Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro 785 790 795 800			
	Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu			

	805								810				815			
	Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg
				820					825					830		
	Leu	Ser	Asp	Tyr	Asp	Val	Asp	His	Ile	Val	Pro	Gln	Ser	Phe	Leu	Lys
			835					840					845			
	Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	Asn	Arg
		850					855						860			
	Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	Lys	Met	Lys
	865					870					875					880
	Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys
					885					890					895	
[0008]	Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp
				900					905					910		
	Lys	Ala	Gly	Phe	Ile	Lys	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr
			915					920					925			
	Lys	His	Val	Ala	Gln	Ile	Leu	Asp	Ser	Arg	Met	Asn	Thr	Lys	Tyr	Asp
		930					935					940				
	Glu	Asn	Asp	Lys	Leu	Ile	Arg	Glu	Val	Lys	Val	Ile	Thr	Leu	Lys	Ser
	945					950					955					960
	Lys	Leu	Val	Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg
					965					970					975	
	Glu	Ile	Asn	Asn	Tyr	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val
				980					985					990		
	Val	Gly	Thr	Ala	Leu	Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe

	995					1000					1005				
	Val	Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	Ala
	1010						1015					1020			
	Lys	Ser	Glu	Gln	Glu	Ile	Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	Phe
	1025						1030					1035			
	Tyr	Ser	Asn	Ile	Met	Asn	Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	Ala
	1040						1045					1050			
	Asn	Gly	Glu	Ile	Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu
	1055						1060					1065			
	Thr	Gly	Glu	Ile	Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	Val
	1070						1075					1080			
[0009]	Arg	Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr
	1085						1090					1095			
	Glu	Val	Gln	Thr	Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys
	1100						1105					1110			
	Arg	Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	Pro
	1115						1120					1125			
	Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	Val
	1130						1135					1140			
	Leu	Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys
	1145						1150					1155			
	Ser	Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser	Ser
	1160						1165					1170			
	Phe	Glu	Lys	Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys

	1175		1180		1185
	Glu Val Lys Lys Asp Leu	Ile	Ile Lys Leu Pro	Lys	Tyr Ser Leu
	1190	1195		1200	
	Phe Glu Leu Glu Asn Gly	Arg	Lys Arg Met Leu	Ala	Ser Ala Gly
	1205	1210		1215	
	Glu Leu Gln Lys Gly Asn	Glu	Leu Ala Leu Pro	Ser	Lys Tyr Val
	1220	1225		1230	
	Asn Phe Leu Tyr Leu Ala	Ser	His Tyr Glu Lys	Leu	Lys Gly Ser
	1235	1240		1245	
	Pro Glu Asp Asn Glu Gln	Lys	Gln Leu Phe Val	Glu	Gln His Lys
	1250	1255		1260	
[0010]	His Tyr Leu Asp Glu Ile	Ile	Glu Gln Ile Ser	Glu	Phe Ser Lys
	1265	1270		1275	
	Arg Val Ile Leu Ala Asp	Ala	Asn Leu Asp Lys	Val	Leu Ser Ala
	1280	1285		1290	
	Tyr Asn Lys His Arg Asp	Lys	Pro Ile Arg Glu	Gln	Ala Glu Asn
	1295	1300		1305	
	Ile Ile His Leu Phe Thr	Leu	Thr Asn Leu Gly	Ala	Pro Ala Ala
	1310	1315		1320	
	Phe Lys Tyr Phe Asp Thr	Thr	Ile Asp Arg Lys	Arg	Tyr Thr Ser
	1325	1330		1335	
	Thr Lys Glu Val Leu Asp	Ala	Thr Leu Ile His	Gln	Ser Ile Thr
	1340	1345		1350	
	Gly Leu Tyr Glu Thr Arg	Ile	Asp Leu Ser Gln	Leu	Gly Gly Asp

	1355	1360	1365
	<210> 8		
	<211> 130		
	<212> DNA		
	<213> 人工		
	<220>		
	<223> 用于图6a和6b的引导RNA_着丝粒16 gRNA带尾		
	<400> 8		
	gacgccuucg uuggaaacgu uuuagagcua gaaauagcaa guuaaaaauaa ggcuaaguccg		
	60		
	uuaucaacuu gaaaaagugg caccgagucg gugcuuuuut cctctaccac ctacatcact		
	120		
	tatacatcta		
	130		
[0011]	<210> 9		
	<211> 31		
	<212> DNA		
	<213> 人工		
	<220>		
	<223> 针对尾的DNA探针		
	<400> 9		
	tagatgtata agtgatgtag gtggtagagg a		
	31		
	<210> 10		
	<211> 27		
	<212> DNA		
	<213> 人工		
	<220>		
	<223> DNA PAINT成像器探针		
	<400> 10		
	tagatgtata aaaaaattta ataaggt		
	27		

	<210>	11
	<211>	16
	<212>	DNA
	<213>	人工
	<220>	
	<223>	DNA PAINT成像器探针
	<400>	11
		tagatgtata aaaaaa
		16
	<210>	12
	<211>	22
	<212>	DNA
	<213>	人工
	<220>	
	<223>	PAM基序
	<220>	
[0012]	<221>	misc_feature
	<222>	(3)..(20)
	<223>	n是a, c, g, 或t
	<400>	12
		ggnnnnnnnnn nnnnnnnnnn gg
		22
	<210>	13
	<211>	23
	<212>	DNA
	<213>	人工
	<220>	
	<223>	Her2的靶序列
	<400>	13
		gggcgaggag gagccccag cgg
		23
	<210>	14
	<211>	23
	<212>	DNA

<213> 人工

<220>

<223> Her2靶序列

<400> 14

ggtggcggag catgtccagg tgg
23

<210> 15

<211> 23

<212> DNA

<213> 人工

<220>

<223> Her2靶序列

<400> 15

ggtgggtctc gggactggca ggg
23

[0013]

<210> 16

<211> 23

<212> DNA

<213> 人工

<220>

<223> Her2靶序列

<400> 16

ggcagccctg gtagaggtgg cgg
23

<210> 17

<211> 23

<212> DNA

<213> 人工

<220>

<223> Her2靶序列

<400> 17

ggaggcccct gtgacagggg tgg
23

	<210>	18
	<211>	23
	<212>	DNA
	<213>	人工
	<220>	
	<223>	Her2靶序列
	<400>	18
		gggcctcccc aggaggcctg cgg
		23
	<210>	19
	<211>	23
	<212>	DNA
	<213>	人工
	<220>	
	<223>	Her2靶序列
	<400>	19
[0014]		ggcctcccca ggaggcctgc ggg
		23
	<210>	20
	<211>	23
	<212>	DNA
	<213>	人工
	<220>	
	<223>	Her2靶序列
	<400>	20
		ggtggctgtg cccgctgcaa ggg
		23
	<210>	21
	<211>	23
	<212>	DNA
	<213>	人工
	<220>	
	<223>	Her2靶序列

	<p><400> 21 gggcagtggc cccttgcagc ggg 23</p>
	<p><210> 22 <211> 23 <212> DNA <213> 人工</p>
	<p><220> <223> Her2靶序列</p>
	<p><400> 22 gggacaggca gtcacacagc tgg 23</p>
	<p><210> 23 <211> 23 <212> DNA <213> 人工</p>
[0015]	<p><220> <223> Her2靶序列</p>
	<p><400> 23 ggttgtgcag ggggcagacg agg 23</p>
	<p><210> 24 <211> 23 <212> DNA <213> 人工</p>
	<p><220> <223> Her2靶序列</p>
	<p><400> 24 gggcatggag cacttgcgag agg 23</p>
	<p><210> 25 <211> 23 <212> DNA <213> 人工</p>

	<220>
	<223> Her2靶序列
	<400> 25
	ggagcacttg cgagaggtga ggg
	23
	<210> 26
	<211> 23
	<212> DNA
	<213> 人工
	<220>
	<223> Her2靶序列
	<400> 26
	ggagctgctc tggctggagc ggg
	23
[0016]	<210> 27
	<211> 23
	<212> DNA
	<213> 人工
	<220>
	<223> Her2靶序列
	<400> 27
	ggaagacgct gaggtcaggc agg
	23
	<210> 28
	<211> 23
	<212> DNA
	<213> 人工
	<220>
	<223> Her2靶序列
	<400> 28
	ggtgggtgtt atggtggatg agg
	23

<210> 29
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 29
ggctggggct gcgctcactg agg
23

<210> 30
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 30
ggtgcgggtt ccgaaagagc tgg
23

[0017]

<210> 31
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 31
ggctgggcat cagctggctg ggg
23

<210> 32
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 32

gggctgggca tcagctggct ggg
23

<210> 33
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 33
ggaggaatgc cgagtactgc agg
23

<210> 34
<211> 23
<212> DNA
<213> 人工

[0018] <220>
<223> Her2靶序列

<400> 34
ggcattcctc cacgcactcc tgg
23

<210> 35
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 35
ggctgacact cagggtggca cgg
23

<210> 36
<211> 23
<212> DNA
<213> 人工

	<220>	
	<223>	Her2靶序列
	<400>	36
		ggtcagggttt cacaccgctg ggg
		23
	<210>	37
	<211>	23
	<212>	DNA
	<213>	人工
	<220>	
	<223>	Her2靶序列
	<400>	37
		gggcagcggg ccacgcagaa ggg
		23
[0019]	<210>	38
	<211>	23
	<212>	DNA
	<213>	人工
	<220>	
	<223>	Her2靶序列
	<400>	38
		ggggcagcgg gccacgcaga agg
		23
	<210>	39
	<211>	23
	<212>	DNA
	<213>	人工
	<220>	
	<223>	Her2靶序列
	<400>	39
		ggttggcatt ctgctggtcg tgg
		23
	<210>	40

<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 40
ggagaatgtg aaaattccag tgg
23

<210> 41
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 41
gggcatctgc ctgacatcca cgg
23

[0020]

<210> 42
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 42
ggatgtgcgg ctctacaca ggg
23

<210> 43
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 43
ggtgaaccgc cggcggagaa tgg

23

<210> 44
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 44
ggtcagggat ctcccgggct ggg
23

<210> 45
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

[0021] <400> 45
ggatgaccac aaagcgctgg ggg
23

<210> 46
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 46
ggatgattga ctctgaatgt cgg
23

<210> 47
<211> 23
<212> DNA
<213> 人工

<220>

	<223> Her2靶序列
	<400> 47 ggtgtctgaa ttctcccgca tgg 23
	<210> 48 <211> 23 <212> DNA <213> 人工
	<220> <223> Her2靶序列
	<400> 48 ggacagaaga agccctgctg ggg 23
	<210> 49 <211> 23 <212> DNA <213> 人工
[0022]	<220> <223> Her2靶序列
	<400> 49 gggggacctg gtggatgctg agg 23
	<210> 50 <211> 23 <212> DNA <213> 人工
	<220> <223> Her2靶序列
	<400> 50 ggacgatgac atgggggacc tgg 23
	<210> 51 <211> 23

<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 51
ggtggatgct gaggagtatc tgg
23

<210> 52
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 52
ggcaccgcag ctcattctacc agg
23

[0023]

<210> 53
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 53
ggagtattctg gtaccccagc agg
23

<210> 54
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 54
ggaccatgcc cccagcgccc ggg
23

<210> 55
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 55
gggtgccagt ggagacctgg ggg
23

<210> 56
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

[0024] <400> 56
ggcgggtgggg acctgacact agg
23

<210> 57
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 57
ggggaggctt tgcagcccct tgg
23

<210> 58
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 58
ggtcctggtc ccagtaatag agg
23

<210> 59
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 59
gggaccagga cccaccagag cgg
23

<210> 60
<211> 23
<212> DNA
<213> 人工

[0025] <220>
<223> Her2靶序列

<400> 60
ggtgtccctt tgaaggctgct ggg
23

<210> 61
<211> 23
<212> DNA
<213> 人工

<220>
<223> Her2靶序列

<400> 61
gggggctggg gccgaacatc tgg
23

<210> 62
<211> 23
<212> DNA

<213> 人工

<220>

<223> Her2靶序列

<400> 62

ggcccaagac tctctcccca ggg
23

<210> 63

<211> 12

<212> DNA

<213> 人工

<220>

<223> 条码B1

<400> 63

cgtcgattac ca
12

[0026]

<210> 64

<211> 12

<212> DNA

<213> 人工

<220>

<223> 条码B2

<400> 64

ccatactcgt cg
12

<210> 65

<211> 12

<212> DNA

<213> 人工

<220>

<223> 条码B3

<400> 65

tcgatagtac gt
12

	<210> 66
	<211> 12
	<212> DNA
	<213> 人工
	<220>
	<223> 条码B4
	<400> 66
	cgtactgaac ga
	12
	<210> 67
	<211> 12
	<212> DNA
	<213> 人工
	<220>
	<223> 条码B5
	<400> 67
[0027]	cgaagtacta cg
	12
	<210> 68
	<211> 12
	<212> DNA
	<213> 人工
	<220>
	<223> 条码B6
	<400> 68
	tcgttacgac ca
	12
	<210> 69
	<211> 12
	<212> DNA
	<213> 人工
	<220>
	<223> 条码B7

	$\langle 400 \rangle$	69
	ccagacgaat cg	12
	$\langle 210 \rangle$	70
	$\langle 211 \rangle$	12
	$\langle 212 \rangle$	DNA
	$\langle 213 \rangle$	人工
	$\langle 220 \rangle$	
	$\langle 223 \rangle$	条码B8
	$\langle 400 \rangle$	70
	tcgaatagtc gt	12
	$\langle 210 \rangle$	71
	$\langle 211 \rangle$	12
	$\langle 212 \rangle$	DNA
	$\langle 213 \rangle$	人工
[0028]	$\langle 220 \rangle$	
	$\langle 223 \rangle$	条码B9
	$\langle 400 \rangle$	71
	cgtaactaac ga	12
	$\langle 210 \rangle$	72
	$\langle 211 \rangle$	12
	$\langle 212 \rangle$	DNA
	$\langle 213 \rangle$	人工
	$\langle 220 \rangle$	
	$\langle 223 \rangle$	条码B10
	$\langle 400 \rangle$	72
	cgaacaatcg ta	12
	$\langle 210 \rangle$	73
	$\langle 211 \rangle$	23
	$\langle 212 \rangle$	DNA
	$\langle 213 \rangle$	人工

	<220>
	<223> CEN17.1靶标
	<400> 73
	gagcgcctttc aggcctgtgg tgg
	23
	<210> 74
	<211> 23
	<212> DNA
	<213> 人工
	<220>
	<223> CEN17.2靶标
	<400> 74
	gagggcctttg aggcctgtgg tgg
	23
[0029]	<210> 75
	<211> 23
	<212> DNA
	<213> 人工
	<220>
	<223> CEN17.3靶标
	<400> 75
	gagggcctttg tggtttgtgg tgg
	23
	<210> 76
	<211> 23
	<212> DNA
	<213> 人工
	<220>
	<223> CEN17.4靶标
	<400> 76
	ggaatctgca agtggatatg tgg
	23

<210> 77
<211> 23
<212> DNA
<213> 人工

<220>
<223> CEN17.5靶标

<400> 77
gtggttgaaac tctctttttg tgg
23

<210> 78
<211> 23
<212> DNA
<213> 人工

<220>
<223> CEN17.6靶标

<400> 78
gtttccaatc actctttgtg tgg
23

[0030]

<210> 79
<211> 23
<212> DNA
<213> 人工

<220>
<223> CEN17.7靶标

<400> 79
gtttggaaac actcttggtg tgg
23

<210> 80
<211> 23
<212> DNA
<213> 人工

<220>
<223> CENT17.8靶标

<400> 80

gttttgaaac tctctttctg tgg
23

<210> 81
<211> 22
<212> DNA
<213> 人工

<220>
<223> 着丝粒重复靶标

<400> 81
acactgctct atccatagga gg
22

<210> 82
<211> 22
<212> DNA
<213> 人工

[0031] <220>
<223> 着丝粒重复靶标

<400> 82
agatatttgg accgctctga gg
22

<210> 83
<211> 22
<212> DNA
<213> 人工

<220>
<223> CENT17.11靶标

<400> 83
agcgctttca ggcctgtggt gg
22

<210> 84
<211> 22
<212> DNA
<213> 人工

	<220>
	<223> CEN17.12靶标
	<400> 84
	aggaatgttc aactctgtga gg
	22
	<210> 85
	<211> 22
	<212> DNA
	<213> 人工
	<220>
	<223> CENT17.13靶标
	<400> 85
	agggctttga ggcctgtggt gg
	22
[0032]	<210> 86
	<211> 22
	<212> DNA
	<213> 人工
	<220>
	<223> CENT17.14靶标
	<400> 86
	agggctttgt ggtttgtggt gg
	22
	<210> 87
	<211> 22
	<212> DNA
	<213> 人工
	<220>
	<223> CEN17.15靶标
	<400> 87
	catcacagag aagcttctga gg
	22
	<210> 88

<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.16靶标

<400> 88
ctgcattcaa ctcacagtgt gg
22

<210> 89
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.17

<400> 89
gaaaggaaag ttcaactcgg gg
22

[0033]

<210> 90
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.18靶标

<400> 90
gaatctgcaa gtggatatgt gg
22

<210> 91
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.19靶标

<400> 91
gaatgcaaac atcacgaaga gg

22

<210> 92
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.20靶标

<400> 92
gcatatttgg acctctttga gg
22

<210> 93
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.21靶标

[0034] <400> 93
gcttctgttt agttctgtgc gg
22

<210> 94
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.22靶标

<400> 94
gcttctgttt agttctgtgc gg
22

<210> 95
<211> 22
<212> DNA
<213> 人工

<220>

<223> CEN17.23靶标

<400> 95
ggacatttgg agggctttga gg
22

<210> 96
<211> 22
<212> DNA
<213> 人工

<220>
<223> CENT17.24靶标

<400> 96
ggacgtttgg agggctttgt gg
22

<210> 97
<211> 22
<212> DNA
<213> 人工

[0035]

<220>
<223> CEN17.25靶标

<400> 97
ggagatttgg agcgctttga gg
22

<210> 98
<211> 22
<212> DNA
<213> 人工

<220>
<223> CENT17.26靶标

<400> 98
ggatatttag gcctctctga gg
22

<210> 99
<211> 22

<212> DNA
<213> 人工

<220>
<223> CEN17.27靶标

<400> 99
ggatatttgg accactctgt gg
22

<210> 100
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.28靶标

<400> 100
ggatatttgg acctctctga gg
22

[0036]

<210> 101
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.29靶标

<400> 101
gggatcattg cactctttga gg
22

<210> 102
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.30靶标

<400> 102
tactaccata ggcctaaagc gg
22

<210> 103
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.31靶标

<400> 103
tat ttg taga atgtgcaagt gg
22

<210> 104
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.32靶标

[0037] <400> 104
tc caaagaca tcttcggaga gg
22

<210> 105
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.33靶标

<400> 105
tc caacgaaa tcctcagaga gg
22

<210> 106
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.34靶标

<400> 106
tccaacgaaa tcctcagaga gg
22

<210> 107
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.35靶标

<400> 107
tccaacgaaa tcctcagagc gg
22

<210> 108
<211> 22
<212> DNA
<213> 人工

[0038] <220>
<223> CENT17.36靶标

<400> 108
tccaacgaaa tcttcaaaga gg
22

<210> 109
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.37靶标

<400> 109
tccaacgaaa tgctcagaga gg
22

<210> 110
<211> 22
<212> DNA

<213> 人工

<220>

<223> CEN17.38靶标

<400> 110

tcgaacgaag gacacagagt gg
22

<210> 111

<211> 22

<212> DNA

<213> 人工

<220>

<223> CEN17.39靶标

<400> 111

tcgaacgaag gccacagagt gg
22

[0039]

<210> 112

<211> 22

<212> DNA

<213> 人工

<220>

<223> CEN17.40靶标

<400> 112

tctgcaagtg gacatttgga gg
22

<210> 113

<211> 22

<212> DNA

<213> 人工

<220>

<223> CEN17.41靶标

<400> 113

tctgcaagtg gacgtttgga gg
22

	<210>	114
	<211>	22
	<212>	DNA
	<213>	人工
	<220>	
	<223>	CEN17.42靶标
	<400>	114
		tggagcgctt tcaggcctgt gg
		22
	<210>	115
	<211>	22
	<212>	DNA
	<213>	人工
	<220>	
	<223>	CEN17.43靶标
	<400>	115
[0040]		tggagggcctt tgaggcctgt gg
		22
	<210>	116
	<211>	22
	<212>	DNA
	<213>	人工
	<220>	
	<223>	CEN17.44靶标
	<400>	116
		tggagggcctt tgtggtttgt gg
		22
	<210>	117
	<211>	22
	<212>	DNA
	<213>	人工
	<220>	
	<223>	CEN17.45靶标

<400> 117
tgttgaaact ctctttttgt gg
22

<210> 118
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.46靶标

<400> 118
ttgttggtgga atgtgcaagt gg
22

<210> 119
<211> 22
<212> DNA
<213> 人工

[0041] <220>
<223> CEN17.47靶标

<400> 119
tttccaatca ctcttttgtgt gg
22

<210> 120
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.48靶标

<400> 120
tttctgtggc atctgcaagg gg
22

<210> 121
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.49靶标

<400> 121
tttggaaca ctcttggtt gg
22

<210> 122
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.50靶标

<400> 122
tttgtgtaga atctgcaagt gg
22

[0042] <210> 123
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.51靶标

<400> 123
tttgtgtgga atctgcaagt gg
22

<210> 124
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.52靶标

<400> 124
ttttcgtagt gtctacaagt gg
22

<210> 125
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.53靶标

<400> 125
ttttgaaact ctctttctgt gg
22

<210> 126
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.54靶标

<400> 126
tttttccaga atctgcaagt gg
22

[0043]

<210> 127
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.55靶标

<400> 127
tttttctaga atctgcaagt gg
22

<210> 128
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.56靶标

<400> 128

tttttgcagg atctacaagt gg
22

<210> 129
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.57靶标

<400> 129
tttttgtaca atctacaagt gg
22

<210> 130
<211> 22
<212> DNA
<213> 人工

[0044] <220>
<223> CEN17.58靶标

<400> 130
tttttgtaga aactgcaagg gg
22

<210> 131
<211> 22
<212> DNA
<213> 人工

<220>
<223> CEN17.59靶标

<400> 131
tttttgtaga aactgcaagt gg
22

<210> 132
<211> 22
<212> DNA
<213> 人工

	<220>
	<223> CEN17.60靶标
	<400> 132
	tttttgtagg atctgcaagt gg
	22
	<210> 133
	<211> 22
	<212> DNA
	<213> 人工
	<220>
	<223> CEN17.61靶标
	<400> 133
	tttttgtagg atctgcaagt gg
	22
[0045]	<210> 134
	<211> 20
	<212> DNA
	<213> 人工
	<220>
	<223> gRNA靶序列
	<400> 134
	gagggctttg aggcctgtgg
	20
	<210> 135
	<211> 20
	<212> DNA
	<213> 人工
	<220>
	<223> gRNA靶序列
	<400> 135
	gtgttgaaac tctctttttg
	20
	<210> 136

<211> 20
<212> DNA
<213> 人工

<220>
<223> gRNA靶序列

<400> 136
gacactgctc tatccatagg
20

<210> 137
<211> 20
<212> DNA
<213> 人工

<220>
<223> gRNA靶序列

<400> 137
gagggctttg aggcctgtgg
20

[0046]

<210> 138
<211> 20
<212> DNA
<213> 人工

<220>
<223> gRNA靶序列

<400> 138
ggaatctgca agtggatatg
20

<210> 139
<211> 20
<212> DNA
<213> 人工

<220>
<223> gRNA靶序列

<400> 139
ggcatatttg gacctctttg

20

<210> 140
<211> 20
<212> DNA
<213> 人工

<220>
<223> gRNA靶序列

<400> 140
gtactaccat aggcctaaag
20

<210> 141
<211> 20
<212> DNA
<213> 人工

<220>
<223> gRNA靶序列

[0047] <400> 141
gtccaacgaa atcctcagag
20

<210> 142
<211> 20
<212> DNA
<213> 人工

<220>
<223> gRNA靶序列

<400> 142
gtctgcaagt ggacatttgg
20

<210> 143
<211> 20
<212> DNA
<213> 人工

<220>

	<223> gRNA靶序列
	<400> 143 gttggttggtg aatgtgcaag 20
	<210> 144 <211> 20 <212> DNA <213> 人工
	<220> <223> gRNA靶序列
	<400> 144 gtttgtgtag aatctgcaag 20
[0048]	<210> 145 <211> 23 <212> DNA <213> 人工
	<220> <223> gRNA靶标
	<400> 145 tctgacgtcc atcatctctg cgg 23
	<210> 146 <211> 23 <212> DNA <213> 人工
	<220> <223> gRNA靶标
	<400> 146 gccaaccacc gcagagatga tgg 23

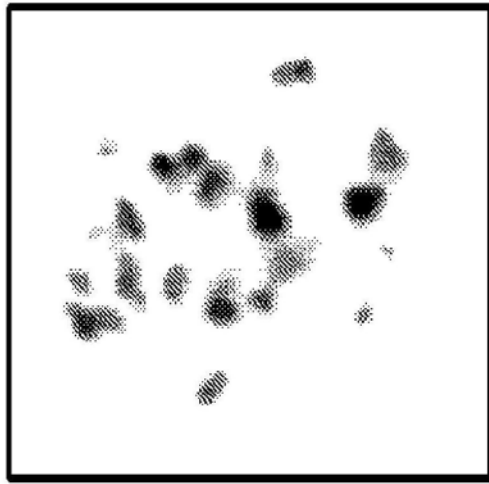


图1A

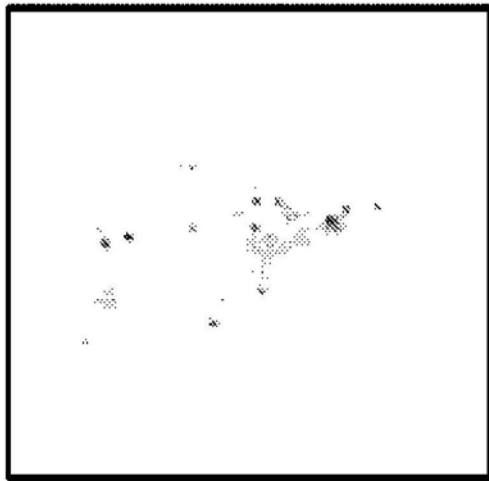


图1B

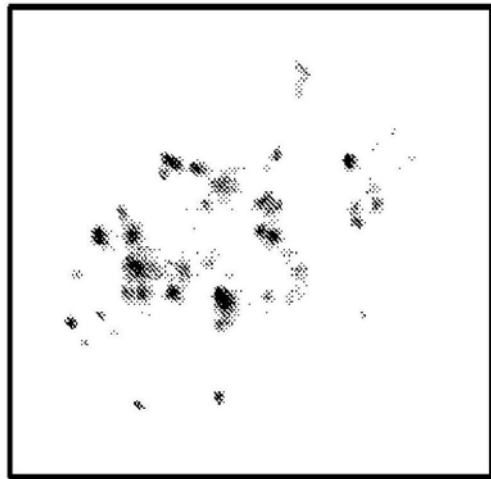


图1C

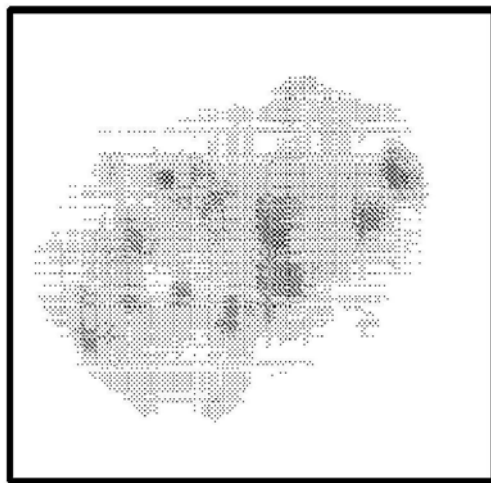


图1D

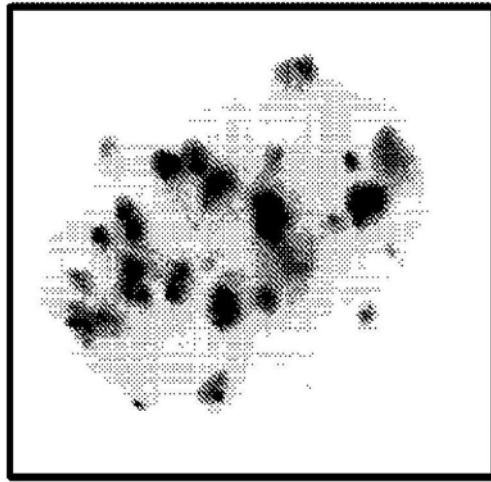


图1E

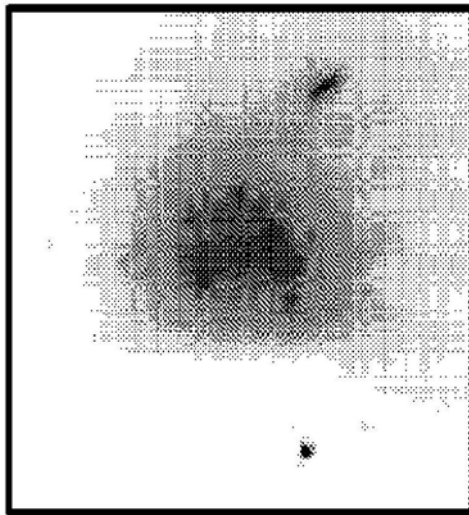


图2A

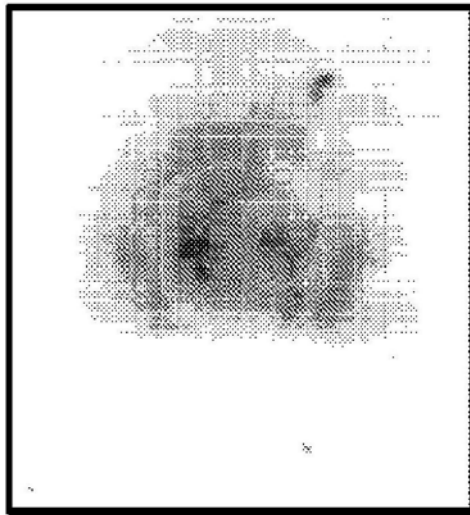


图2B

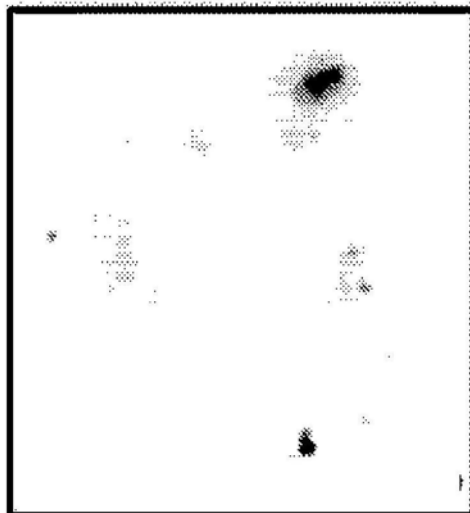


图2C

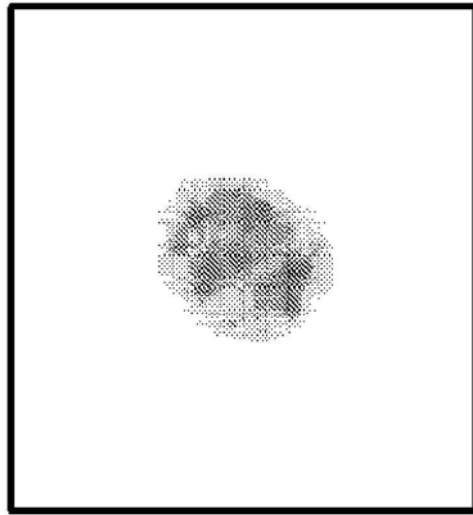


图2D

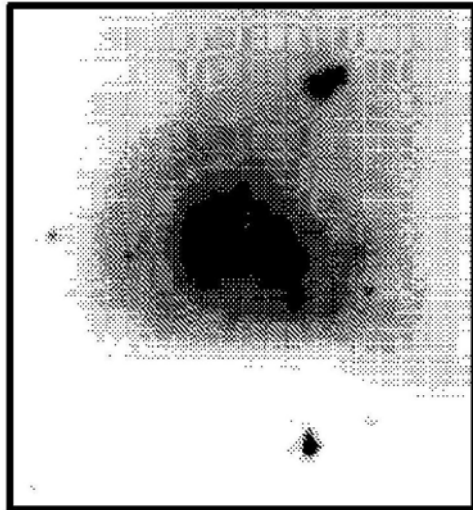


图2E

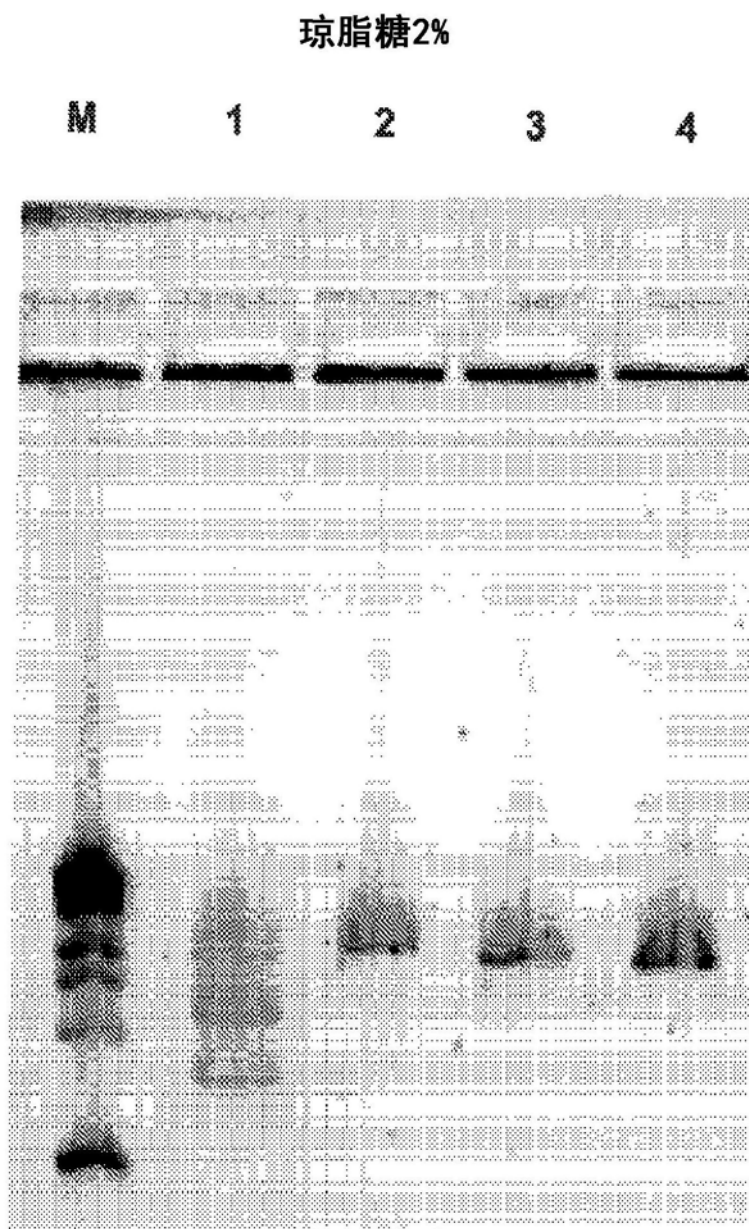


图3

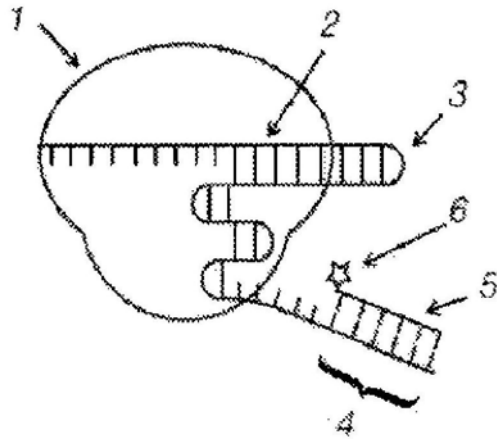


图4A

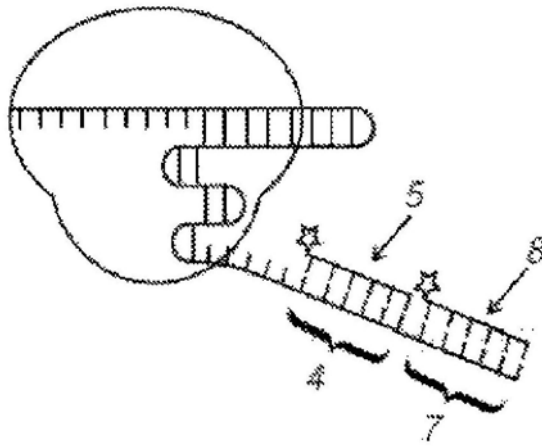


图4B

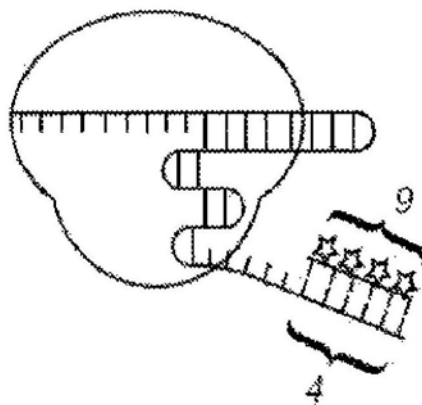


图4C

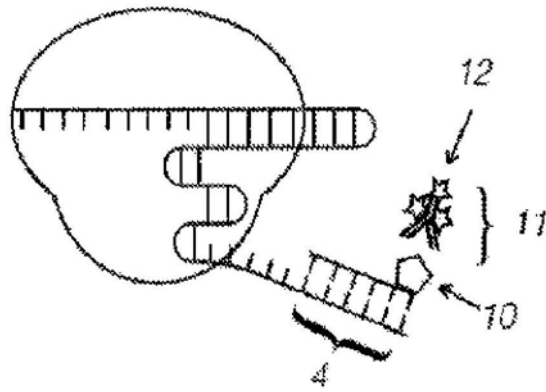


图4D

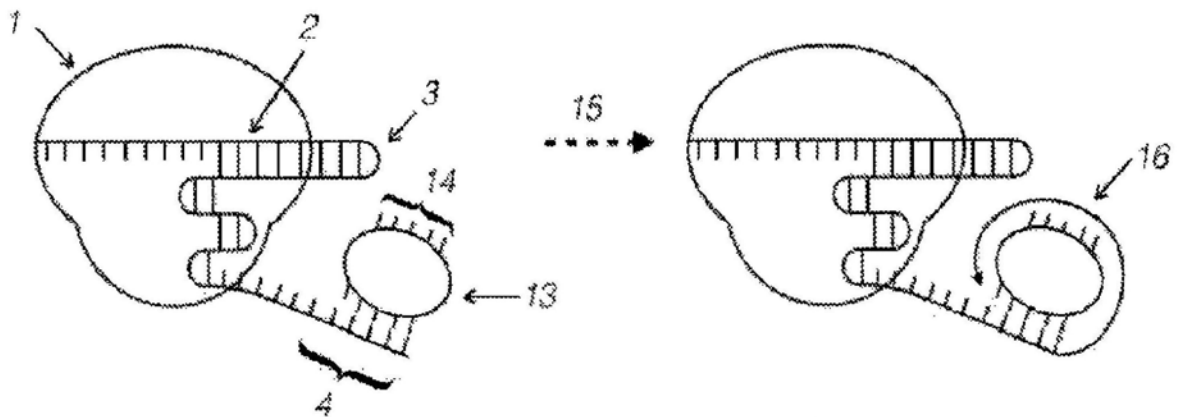


图 4E

图 4F

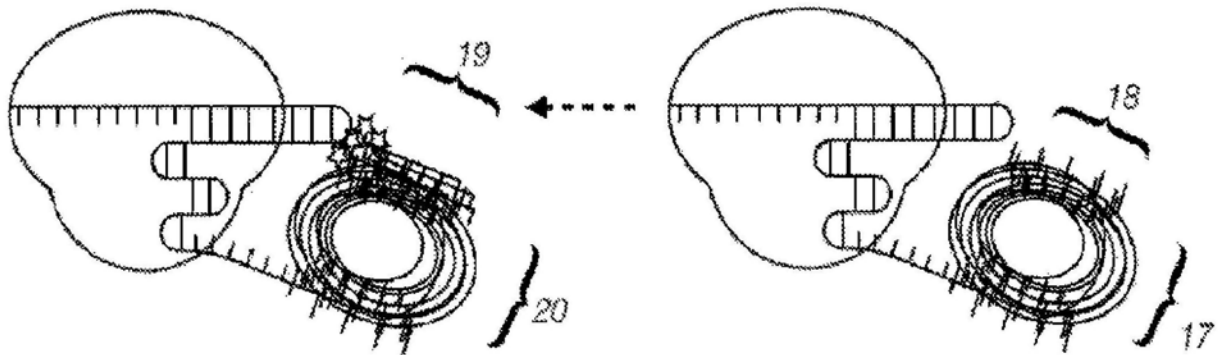


图 4H

图 4G

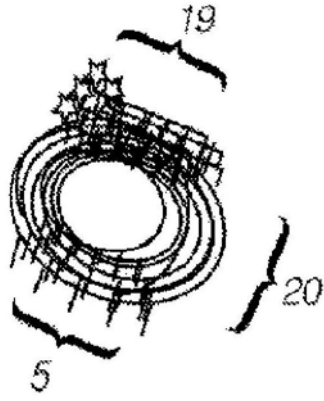


图4I

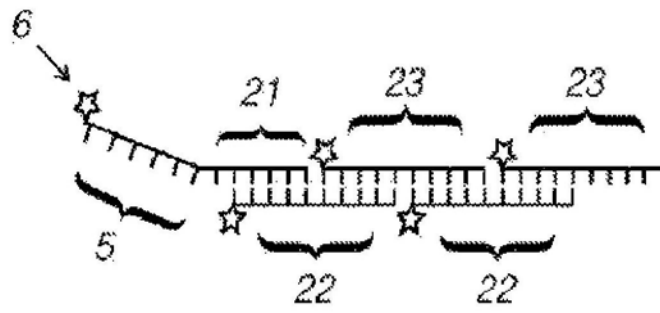


图4J

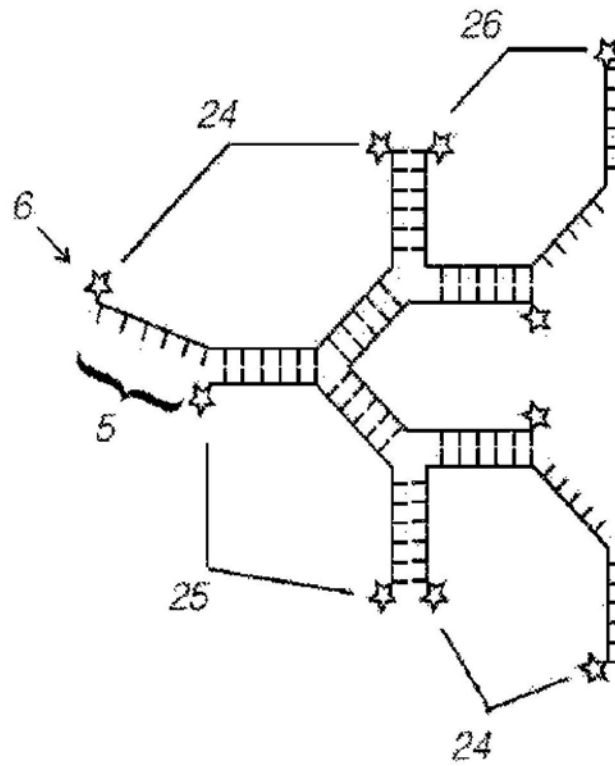


图4K

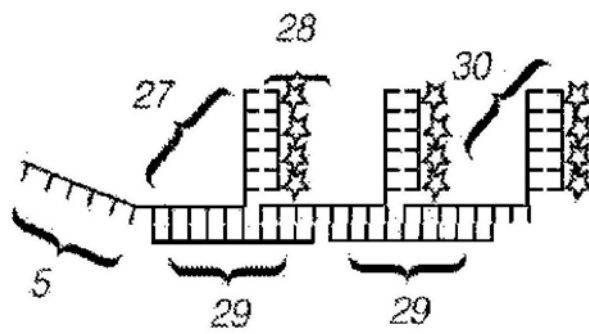


图4L

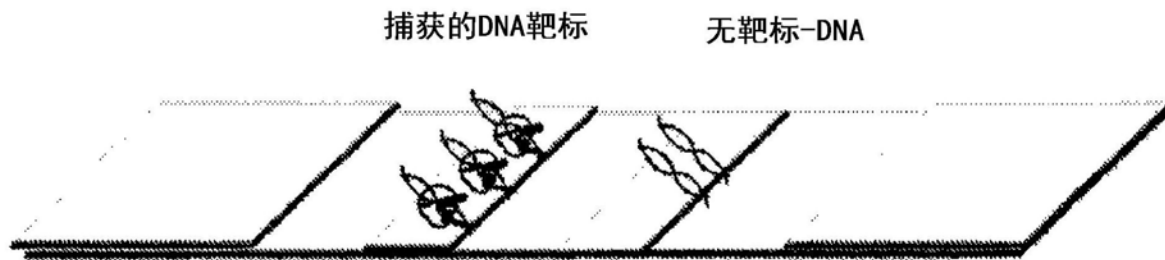
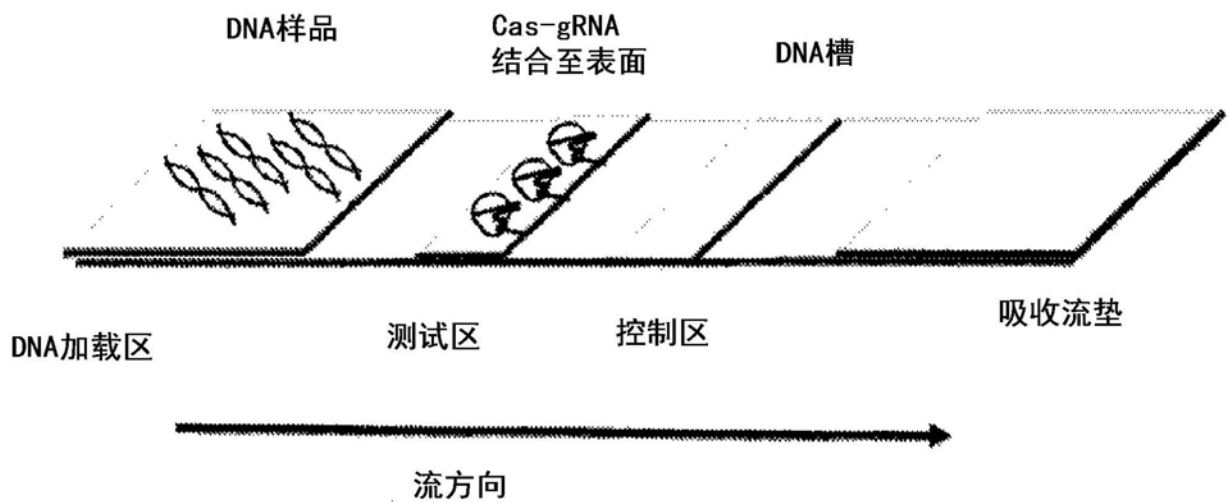


图5A

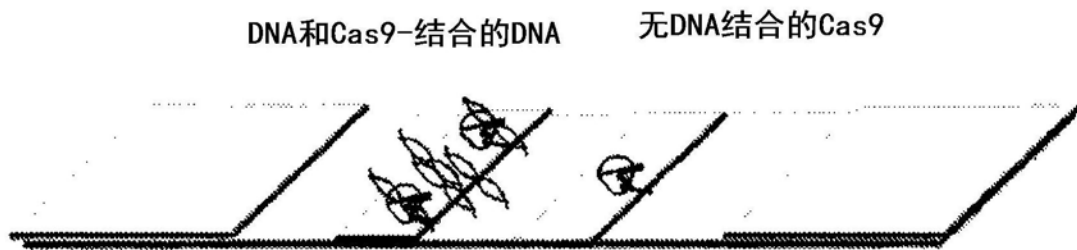
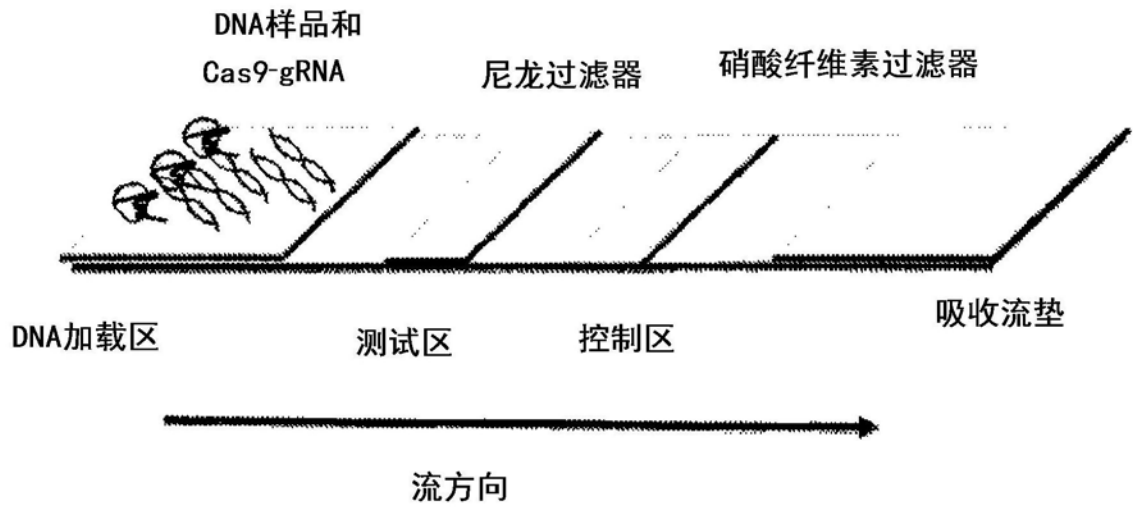


图5B

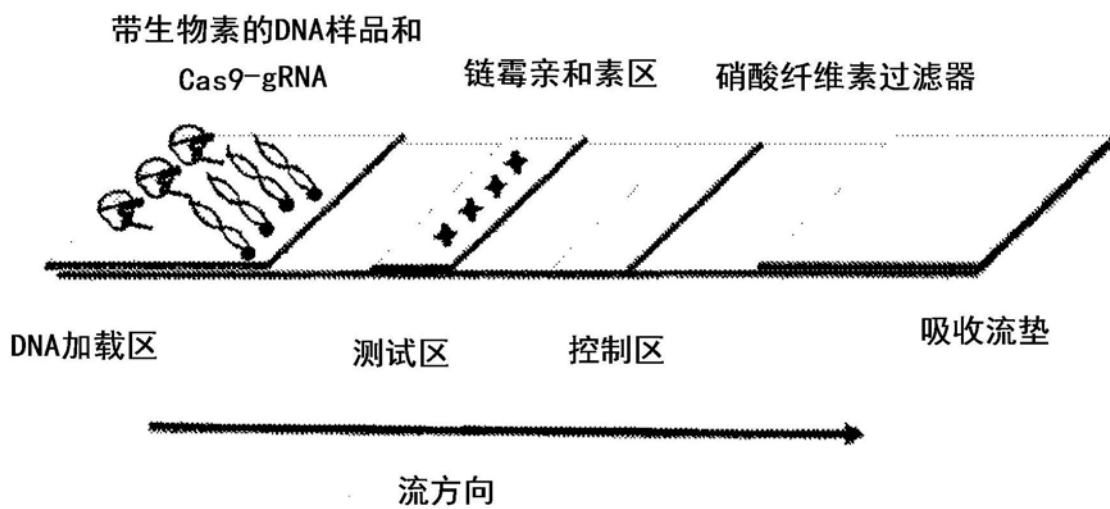


图5C

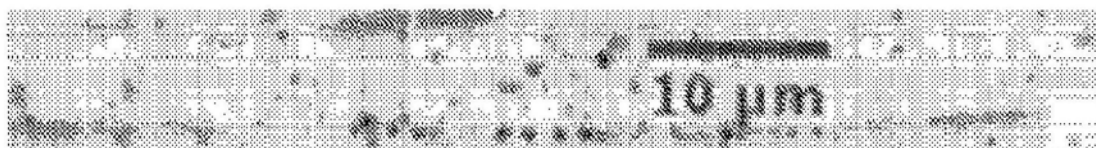


图6A

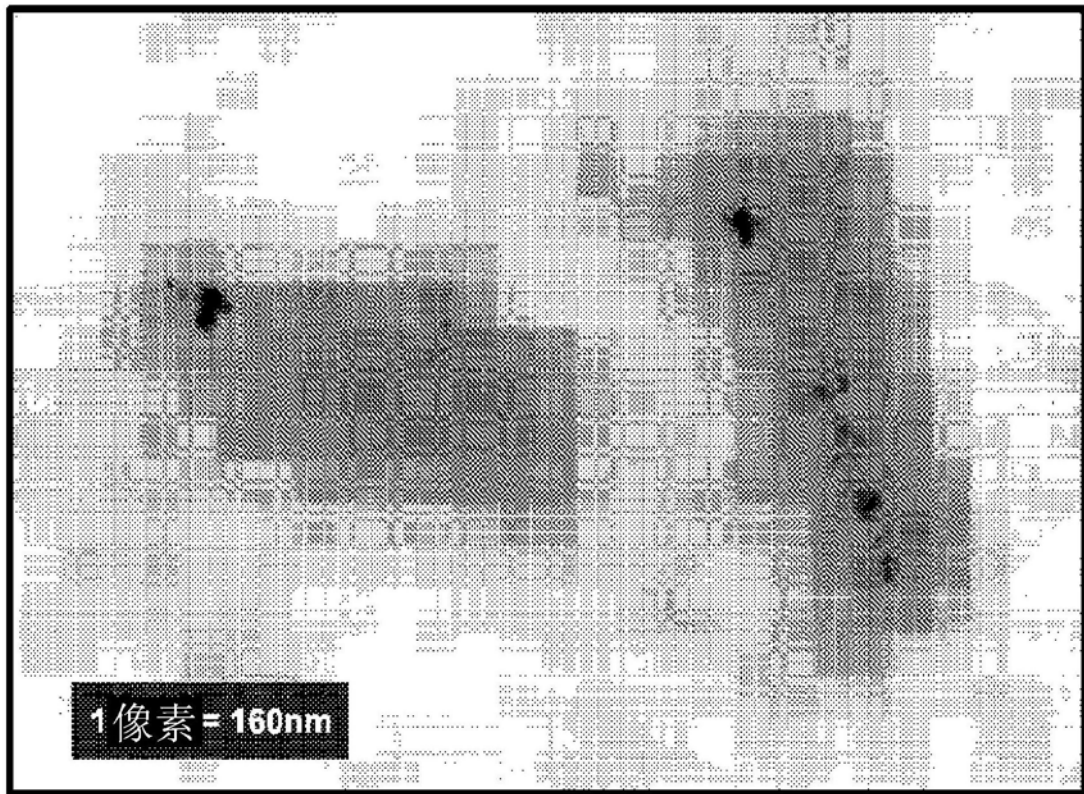


图6B

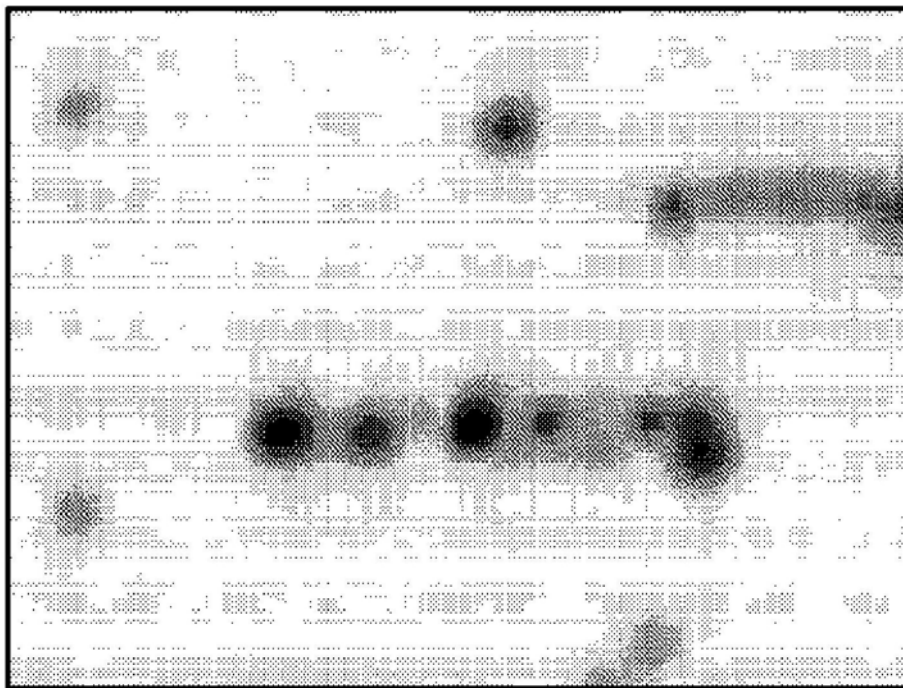


图6C

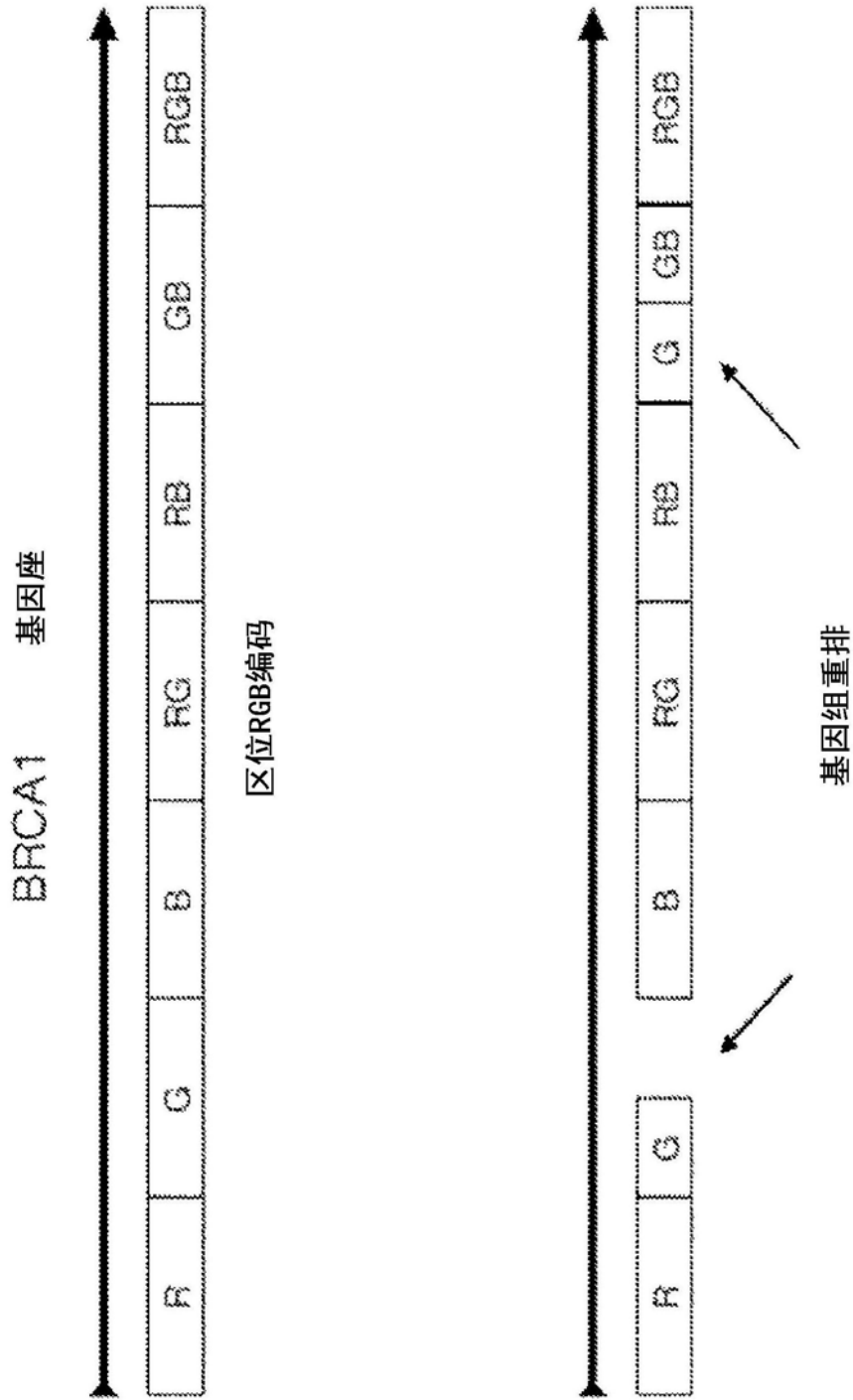


图7

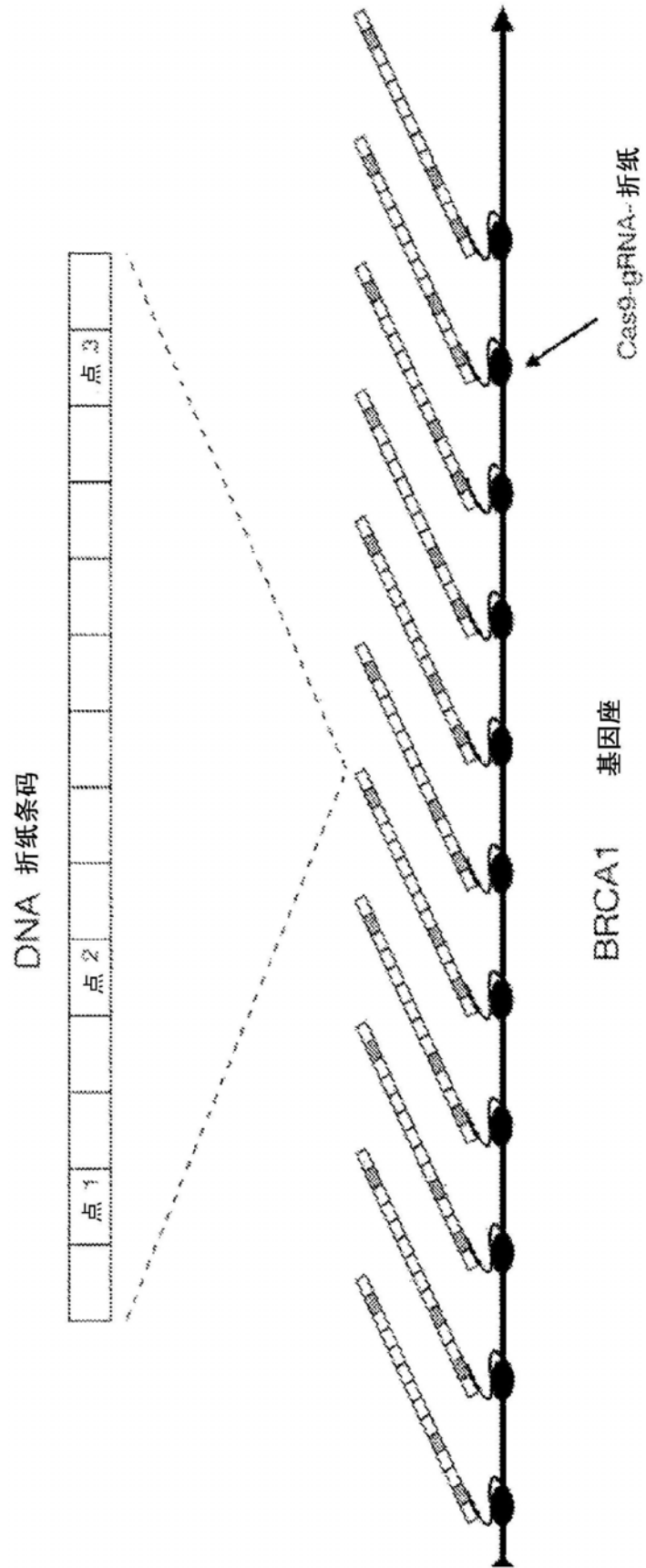


图8



图9A



图9B

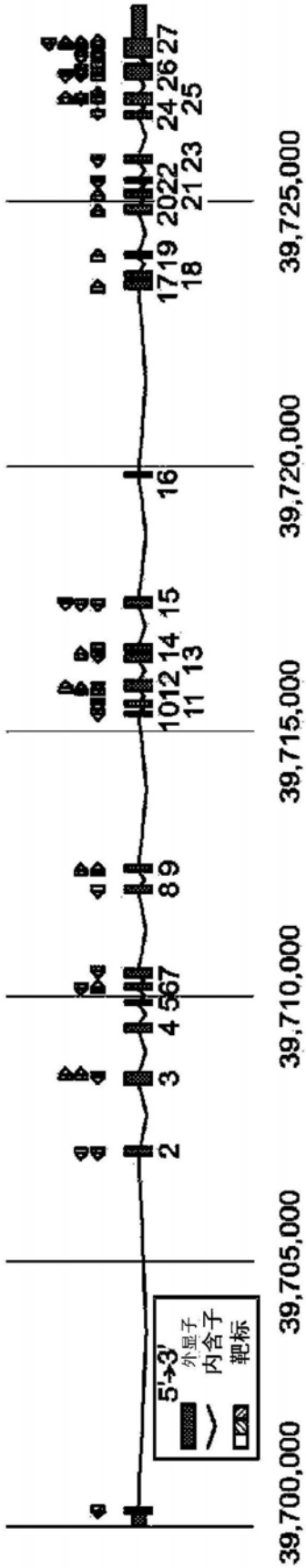


图10

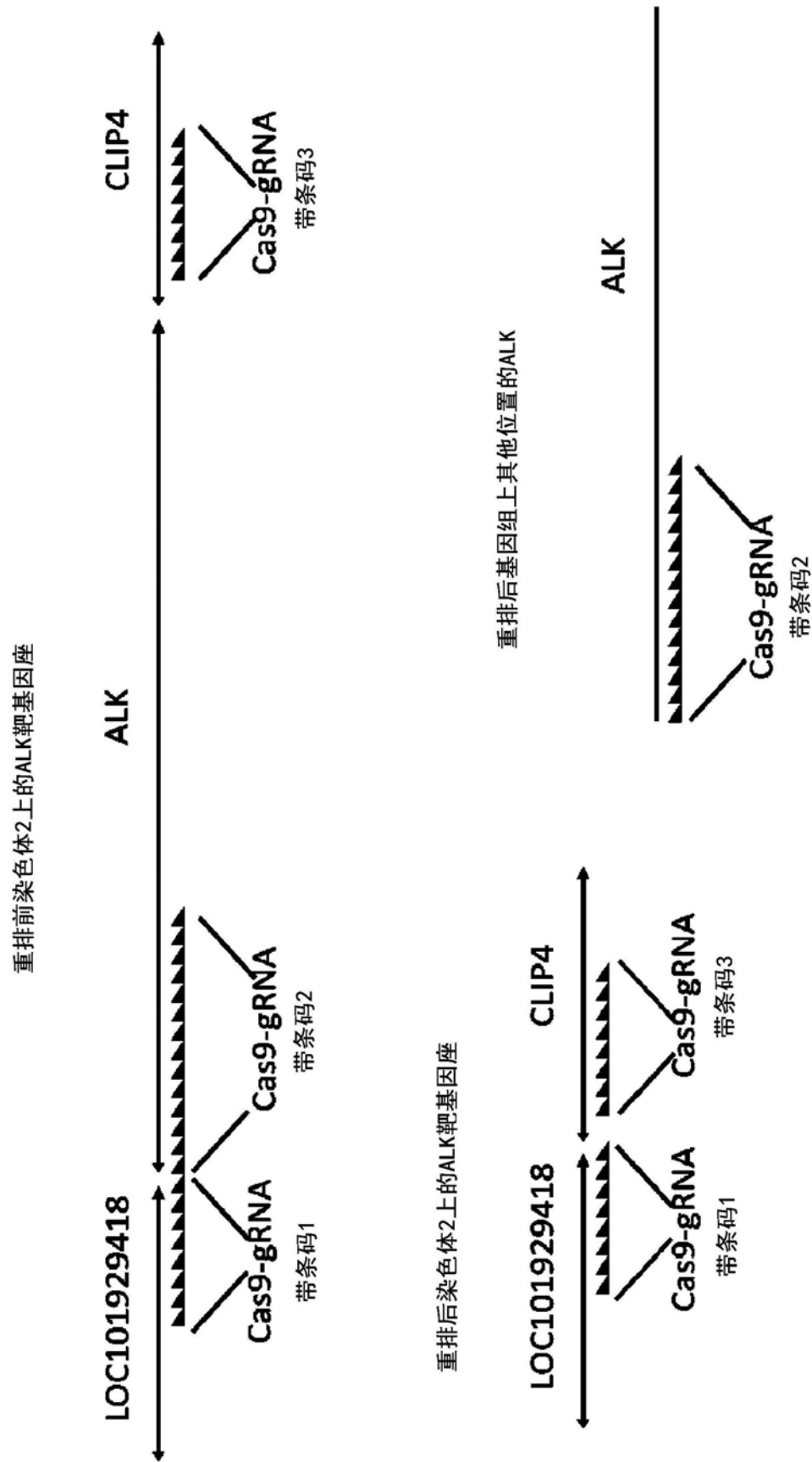


图12