



(12) 发明专利

(10) 授权公告号 CN 107429246 B

(45) 授权公告日 2021.06.01

(21) 申请号 201580065962.8

(22) 申请日 2015.10.30

(65) 同一申请的已公布的文献号
申请公布号 CN 107429246 A

(43) 申请公布日 2017.12.01

(30) 优先权数据

62/073,126 2014.10.31 US

62/166,302 2015.05.26 US

(85) PCT国际申请进入国家阶段日
2017.06.02

(86) PCT国际申请的申请数据
PCT/US2015/058304 2015.10.30

(87) PCT国际申请的公布数据
W02016/070037 EN 2016.05.06

(73) 专利权人 麻省理工学院

地址 美国马萨诸塞州

(72) 发明人 卢冠达 黄兆麟 蔡正姿

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 郑斌 刘振佳

(51) Int.Cl.

C12N 15/10 (2006.01)

审查员 黎舒婷

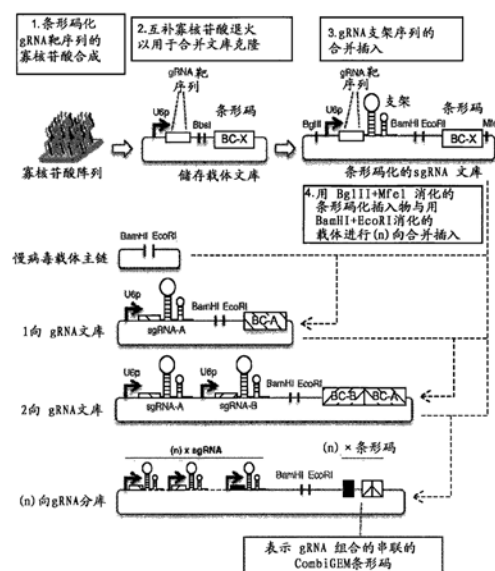
权利要求书3页 说明书59页
序列表64页 附图47页

(54) 发明名称

用于CRISPR的大规模并行组合遗传学

(57) 摘要

本文中描述了能够迅速产生包含CRISPR指导序列和支架序列的遗传元件的高阶组合以及用于迅速鉴定单细胞或合并群体内所编码遗传元件之组合的条形码的方法和组合物。本文中還描述了表观遗传基因的抑制剂的组合物和用于降低细胞增殖和/或治疗癌症的方法。



1. 遗传构建体,其包含

第一DNA元件,所述第一DNA元件包含

CRISPR指导序列和

支架序列;

所述第一DNA元件侧翼的第一相容性末端元件和第二相容性末端元件,其中所述第一和第二相容性末端元件能够与彼此退火;

条形码元件;

所述条形码元件侧翼的第三相容性末端元件和第四相容性末端元件,其中所述第三和第四相容性末端元件能够与彼此退火但不能与所述第一或第二相容性末端元件退火;以及

位于所述第四相容性末端元件和所述第一相容性末端元件之间的隔离位点,其中所述DNA元件、第一相容性末端元件和第二相容性末端元件位于隔离位点的一侧,且所述条形码元件、第三相容性末端元件和第四相容性末端元件位于隔离位点的另一侧。

2. 权利要求1所述的遗传构建体,其还包含所述第一DNA元件上游的启动子元件。

3. 载体,其包含根据权利要求1或2所述的遗传构建体。

4. 遗传构建体,其包含:

多个DNA元件,其中所述多个DNA元件中的每个DNA元件包含CRISPR指导序列和支架序列;

所述多个DNA元件侧翼的第一相容性末端元件和第二相容性末端元件,其中所述第一和第二相容性末端元件能够与彼此退火;

多个条形码元件;

所述多个条形码元件侧翼的第三相容性末端元件和第四相容性末端元件,其中所述第三和第四相容性末端元件能够与彼此退火但不能与所述第一或第二相容性末端元件退火;以及

位于所述多个DNA元件与所述多个条形码元件之间的隔离位点。

5. 载体,其包含

根据权利要求4所述的遗传构建体和

位于每个所述CRISPR指导序列上游的启动子序列。

6. 用于产生组合载体的方法,其包括:

(a) 提供含有第一遗传构建体的载体,所述第一遗传构建体包含:

CRISPR指导序列;

所述CRISPR指导序列侧翼的第二相容性末端元件和第一限制性酶的第一识别位点;

条形码元件;和

所述条形码元件侧翼的第三相容性末端元件和第二限制性酶的第二识别位点;

(b) 在所述第一识别位点处切割所述第一遗传构建体,得到第五相容性末端元件,并且在所述第二识别位点处切割所述载体,得到第六相容性末端元件;

(c) 提供支架元件,所述支架元件包含

支架序列;

包含第一相容性末端元件和第四相容性末端元件的隔离位点;和

所述支架元件侧翼的第七相容性末端元件和第八相容性末端元件,其中所述第七相容

性末端元件能够与所述第五相容性末端元件退火,并且所述第八相容性末端元件能够与所述第六相容性末端元件退火;

(d) 使所述支架元件与经切割的第一遗传构建体退火,其中所述退火发生在所述载体和所述支架元件内能够与彼此退火的相容性末端元件处,并且其中在退火后,所述支架元件整合在所述CRISPR指导序列与所述条形码元件之间,并且其中所述隔离位点位于所述支架序列与所述条形码元件之间,从而产生组合载体。

7. 权利要求6所述的方法,其还包括:

(a) 提供根据权利要求6的组合载体;

(b) 在所述支架元件内的隔离位点处切割所述载体,得到第一相容性末端元件和第四相容性末端元件;

(c) 提供第二遗传构建体,所述第二遗传构建体包含
CRISPR指导序列;

支架序列;

条形码元件;和

所述第二遗传构建体侧翼的第二相容性末端元件和第三相容性末端元件,其中所述第二遗传构建体的第二相容性末端元件能够与所述载体的第一相容性末端元件退火,并且所述第二遗传构建体的第三相容性末端元件能够与所述载体的第四相容性末端元件退火;

(d) 使所述第二遗传构建体与经切割的载体退火,其中所述退火发生在所述第二遗传构建体和所述载体内能够与彼此退火的相容性末端元件处,并且其中在退火后,所述第二遗传构建体整合在所述载体中,产生包含串联的条形码元件和串联的CRISPR指导序列以及支架序列的组合载体。

8. 权利要求7所述的方法,其中所述组合载体还包含所述CRISPR指导序列上游的启动子元件。

9. 权利要求7或8所述的方法,其中所述方法是迭代的。

10. 权利要求6至8中任一项所述的方法,其中所述第一识别位点和所述第二识别位点具有相同的识别位点序列,并且所述第一限制性酶和所述第二限制性酶是相同的限制性酶。

11. 用于产生组合载体的方法,其包括:

(a) 提供载体,所述载体包含:

多个CRISPR指导序列;

条形码元件,其中所述条形码元件位于所述多个CRISPR指导序列的下游;

任选的启动子序列,其位于所述多个CRISPR指导序列中的至少一个之上游;和

多种限制性酶的多个识别位点,其中所述多个识别位点各自位于所述多个CRISPR指导序列之一的下游;

(b) 用所述多种限制性酶中的至少一种在所述多个识别位点的至少一个处切割所述载体,得到第一相容性末端元件和第二相容性末端元件;

(c) 提供第一支架元件,所述第一支架元件包含:

支架序列,

任选的启动子序列,和

所述第一支架元件侧翼的第三相容性末端元件和第四相容性末端元件,其中所述第三相容性末端元件能够与经切割的载体的第一相容性末端元件退火,并且所述第四相容性末端元件能够与经切割的载体的第二相容性末端元件退火;

(d) 使所述第一支架元件与经切割的载体退火,其中所述退火发生在所述第一支架元件和经切割的载体内的相容性末端元件处,并且其中在退火后,所述第一支架元件整合在所述多个CRISPR指导序列之一的下游,从而产生组合载体。

12. 权利要求11所述的方法,其中所述方法是迭代的。

13. 用于产生组合载体的方法,其包括:

(a) 提供载体,所述载体包含:

多个CRISPR指导序列,

条形码元件,其中所述条形码元件位于所述多个CRISPR指导序列的上游;

任选的启动子序列,其位于所述多个CRISPR指导序列的至少一个之上游;和

多种限制性酶的多个识别位点,其中所述多个识别位点各自位于所述多个CRISPR指导序列之一的上游;

(b) 用所述多种限制性酶中的至少一种在所述多个识别位点的至少一个处切割所述载体,得到第一相容性末端元件和第二相容性末端元件;

(c) 提供第一支架元件,所述第一支架元件包含:

任选的支架序列,

启动子序列,和

所述第一支架元件侧翼的第三相容性末端元件和第四相容性末端元件,其中所述第三相容性末端元件能够与经切割的载体的第一相容性末端元件退火,并且所述第四相容性末端元件能够与经切割的载体的第二相容性末端元件退火;

(d) 使所述第一支架元件与经切割的载体退火,其中所述退火发生在所述第一支架元件和经切割的载体内的相容性末端元件处,并且其中在退火后,所述第一支架元件整合在所述多个CRISPR指导序列之一的上游,从而产生组合载体。

14. 权利要求13所述的方法,其中所述方法是迭代的。

用于CRISPR的大规模并行组合遗传学

[0001] 相关申请

[0002] 本申请根据35 U.S.C.119 (e) 要求2014年10月31日提交的美国临时申请号62/073,126和2015年5月26日提交的美国临时申请号62/166,302的权益,其各自通过引用整体并入本文。

[0003] 政府资助

[0004] 本发明是在由国立卫生研究院授予的基金号0D008435的政府支持下完成的。政府对本发明享有一定的权利。

技术领域

[0005] 本发明涉及用于迅速产生包含CRISPR指导序列和支架序列的遗传元件的高阶组合的方法和组合物,以及所述遗传元件的鉴定。本发明还涉及靶向表观遗传基因以抑制细胞增殖的抑制剂的组合物和相关方法。

背景技术

[0006] 成簇规律间隔短回文重复序列(clustered regularly interspaced short palindromic repeats,CRISPR)/Cas系统最初在细菌和古细菌物种中作为抵抗外来遗传物质(例如质粒和噬菌体)的防御机制被发现。天然存在的CRISPR/Cas系统依赖于三种组分的表达:与靶序列互补的指导RNA序列、有助于将第三种组分(即,内切核酸酶)募集到位点的支架RNA。尽管在许多细菌和古细菌物种中,CRISPR/Cas系统用于降解外来遗传物质,但该系统已适用于很多种原核生物和真核生物,并已被用于许多方法,包括基因敲除、诱变和表达激活或抑制(Hsu,等Cell (2014) 157 (6):1262-1278)。在遗传改造的CRISPR/Cas系统中,可通过表达小指导RNA(small guide RNA,sgRNA)来避免对三种独立组分的需求,所述小指导RNA含有用于结合靶序列的CRISPR指导RNA序列和共同模拟由单个指导RNA序列和支架序列形成的结构并且足以将内切核酸酶募集到适当的靶位点的支架RNA二者(Jinek等, Science (2012) 337 (6096):816-821)。

发明内容

[0007] 产生用于表达包含多于一种sgRNA(指导序列和支架序列)的多个CRISPR系统的载体和遗传元件是非常费力的,并且使用传统克隆方法构建的CRISPR系统的文库的复杂度非常有限。本文描述的方法允许产生包含多种sgRNA的载体,每种sgRNA包含CRISPR指导序列和支架序列,以及可检测到的且用作CRISPR指导序列身份之指示物的串联的条形码(barcode)元件。这些方法还提供了高复杂度载体文库的简单且迅速的产生。

[0008] 本发明的一些方面提供了这样的遗传构建体,其包含含有CRISPR指导序列和支架序列的第一DNA元件;位于所述第一DNA元件侧翼的第一相容性末端元件和第二相容性末端元件,其中所述第一和第二相容性末端元件能够与彼此退火;条形码元件;位于所述条形码元件侧翼的第三相容性末端元件和第四相容性末端元件,其中所述第三和第四相容性末端

元件能够与彼此退火但不能与第一或第二相容性末端元件退火;以及位于所述第四相容性末端元件和第一相容性末端元件之间的隔离位点,其中所述DNA元件、第一相容性末端元件和第二相容性末端元件位于隔离位点的一侧,而所述条形码元件、第三相容性末端元件和第四相容性末端元件位于隔离位点的另一侧。

[0009] 在一些实施方案中,所述遗传构建体还包含第一DNA元件上游的启动子元件。

[0010] 一些方面提供了包含本文提供的任何遗传构建体的载体。

[0011] 另一些方面提供了这样的遗传构建体,其包含多个DNA元件,其中所述多个DNA元件的每个DNA元件包含CRISPR指导序列和支架序列;位于多个DNA元件侧翼的第一相容性末端元件和第二相容性末端元件,其中所述第一和第二相容性末端元件能够与彼此退火;多个条形码元件;位于所述多个条形码元件侧翼的第三相容性末端元件和第四相容性末端元件,其中所述第三和第四相容性末端元件能够与彼此退火但不能与第一或第二相容性末端元件退火;以及位于所述多个DNA元件和所述多个条形码元件之间的隔离位点。

[0012] 另一些方面提供了这样的载体,其包含本文中提供的任何遗传构建体和位于每个CRISPR指导序列上游的启动子序列。

[0013] 还有另一些方面提供了用于产生组合载体的方法,其包括(a)提供这样的载体,所述载体含有包含以下的第一遗传构建体:CRISPR指导序列;CRISPR指导序列侧翼的第二相容性末端元件和第一限制性酶的第一识别位点;条形码元件;以及所述条形码元件侧翼的第三相容性末端元件和第二限制性酶的第二识别位点;(b)在第一识别位点处切割第一遗传构建体,得到第五相容性末端元件,并且在第二识别位点处切割所述载体,得到第六相容性末端元件;(c)提供这样的支架元件,其包含:支架序列;包含第一相容性末端元件和第四相容性末端元件的隔离位点;以及所述支架元件侧翼的第七相容性末端元件和第八相容性末端元件,其中所述第七相容性末端元件能够与所述第五相容性末端元件退火,并且所述第八相容性末端元件能够与所述第六相容性末端元件退火;和(d)使所述支架元件与经切割的第一遗传构建体退火,其中所述退火发生在所述载体和支架元件内能够与彼此退火的相容性末端元件处,并且其中在退火后,所述支架元件整合在CRISPR指导序列与条形码元件之间,并且其中所述隔离位点位于支架序列与条形码元件之间,产生组合载体。

[0014] 在一些实施方案中,所述方法还包括(a)提供本文所述的任何组合载体;(b)在所述支架元件内的隔离位点处切割所述载体,得到第一相容性末端元件和第四相容性末端元件;(c)提供这样的第二遗传构建体,其包含:CRISPR指导序列;支架序列;条形码元件;以及所述第二遗传构建体侧翼的第二相容性末端元件和第三相容性末端元件,其中所述第二遗传构建体的第二相容性末端元件能够与所述载体的第一相容性末端元件退火,并且所述第二遗传构建体的第三相容性末端元件能够与所述载体的第四相容性末端元件退火;和(d)将所述第二遗传构建体与经切割的载体退火,其中所述退火发生在所述第二遗传构建体和载体内能够与彼此退火的相容性末端元件处,并且其中在退火后,所述第二遗传构建体整合在所述载体中,产生包含串联的条形码元件和串联的CRISPR指导和支架序列的组合载体。

[0015] 在一些实施方案中,组合载体还包含CRISPR指导序列上游的一个或更多个启动子。在一些实施方案中,所述方法是迭代的(iterative)。在一些实施方案中,所述第一识别位点和第二识别位点具有相同的识别位点序列,并且所述第一限制性酶和第二限制性酶是

相同的限制性酶。

[0016] 本发明的另一些方面提供了这样的遗传构建体,其包含:至少两个CRISPR指导序列;条形码元件;以及位于每个CRISPR指导序列之间和位于所述条形码元件与最接近所述条形码元件的CRISPR指导序列之间的限制性识别位点。

[0017] 另一些方面提供了这样的遗传构建体,其包含:多个DNA元件,其各自包含CRISPR指导序列和支架序列;条形码元件;以及位于多个DNA元件的每个DNA元件上游的启动子序列。在一些实施方案中,所述条形码元件位于遗传构建体的5'末端。在一些实施方案中,所述条形码元件位于遗传构建体的3'末端。

[0018] 另一些方面提供了包含本文所述的任何遗传构建体的载体。

[0019] 还有另一些方面提供了用于产生组合载体的方法,其包括(a)提供这样的载体,其包含:多个CRISPR指导序列;条形码元件,其中所述条形码元件位于多个CRISPR指导序列的下游;位于所述多个CRISPR指导序列的至少一个之上游的任选的启动子序列;以及多种限制性酶的多个识别位点,其中所述多个识别位点各自位于所述多个CRISPR指导序列之一的下游;(b)用所述多种限制性酶的至少一种在所述多个识别位点的至少一个处切割所述载体,得到第一相容性末端元件和第二相容性末端元件;(c)提供第一支架元件,其包含:支架序列,任选的启动子序列,以及所述第一支架元件侧翼的第三相容性末端元件和第四相容性末端元件,其中所述第三相容性末端元件能够与经切割的载体的第一相容性末端退火,并且所述第四相容性末端元件能够与经切割的载体的第二相容性末端元件退火;和(d)使所述第一支架元件与经切割的载体退火,其中所述退火发生在所述第一支架元件和经切割的载体内的相容性末端元件处,并且其中在退火后,所述第一支架元件整合在所述多个CRISPR指导序列之一的下游,从而产生组合载体。在一些实施方案中,所述方法是迭代的。

[0020] 另一些方面提供了用于产生组合载体的方法,其包括(a)提供这样载体,其包含:多个CRISPR指导序列,条形码元件,其中所述条形码元件位于所述多个CRISPR指导序列的上游;位于所述多个CRISPR指导序列的至少一个之上游的任选的启动子序列;以及多种限制性酶的多个识别位点,其中所述多个识别位点各自位于所述多个CRISPR指导序列之一的上游;(b)用所述多种限制性酶的至少一种在所述多个识别位点的至少一个处切割所述载体,得到第一相容性末端元件和第二相容性末端元件;(c)提供这样的第一支架元件,其包含:任选的支架序列,启动子序列以及所述第一支架元件侧翼的第三相容性末端元件和第四相容性末端元件,其中所述第三相容性末端元件能够与经切割载体的第一相容性末端元件退火,并且所述第四相容性末端元件能够与经切割载体的第二相容性末端元件退火;(d)使所述第一支架元件与经切割的载体退火,其中所述退火发生在所述第一支架元件和经切割载体内的相容性末端元件处,并且其中在退火后,所述第一支架元件整合在多个CRISPR指导序列之一的上游,从而产生组合载体。在一些实施方案中,所述方法是迭代的。

[0021] 本发明的一些方面提供了包含靶向两种或更多种表观遗传基因的两种或更多种抑制剂的组合物,所述两种或更多种表观遗传基因选自表2所示的表观遗传基因之组合。在一些实施方案中,两种或更多种抑制剂各自降低或阻止表观遗传基因的表达或降低或阻止由表观遗传基因编码的蛋白质的活性。在一些实施方案中,所述抑制剂各自选自CRISPR指导序列和支架序列;shRNA;和小分子。在一些实施方案中,至少一种所述抑制剂是CRISPR指导序列和支架序列;并且所述组合物还包含或编码Cas9内切核酸酶。在一些实施方案中,

CRISPR指导序列或shRNA由重组表达载体表达。在一些实施方案中,表观遗传基因的组合包含BRD4和KDM4C或BRD4和KDM6B。在一些实施方案中,BRD4的抑制剂是JQ1((6S)-4-(4-氯苯基)-2,3,9-三甲基-6H-噻吩并[3,2-f][1,2,4]三唑并[4,3-a][1,4]二氮杂~~草~~-6-乙酸1,1-二甲基乙酯)。在一些实施方案中,KDM4C的抑制剂是SD70(N-(呋喃-2-基(8-羟基喹啉-7-基)甲基)异丁酰胺)。在一些实施方案中,KDM6B的抑制剂是GSK-J4(3-((6-(4,5-二氢-1H-苯并[d]氮杂~~草~~-3(2H)-基)-2-(吡啶-2-基)嘧啶-4-基)氨基)丙酸乙酯,单盐酸盐)。

[0022] 另一些方面提供了降低细胞增殖的方法,其包括使所述细胞与靶向两种或更多种表观遗传基因的两种或更多种抑制剂的组合接触,所述两种或更多种表观遗传基因选自表2所示表观遗传基因之组合。在一些实施方案中,所述细胞是癌细胞。在一些实施方案中,所述癌细胞是卵巢癌细胞。在一些实施方案中,所述两种或更多种抑制剂各自降低或阻止表观遗传基因的表达或降低或阻止由表观遗传基因编码的蛋白质的活性。在一些实施方案中,所述抑制剂各自选自CRISPR指导序列和支架序列;shRNA;和小分子。在一些实施方案中,至少一种所述抑制剂是CRISPR指导序列和支架序列;并且所述组合物还包含或编码Cas9内切核酸酶。在一些实施方案中,CRISPR指导序列或shRNA由重组表达载体表达。在一些实施方案中,表观遗传基因的组合包含BRD4和KDM4C或BRD4和KDM6B。在一些实施方案中,BRD4的抑制剂是JQ1((6S)-4-(4-氯苯基)-2,3,9-三甲基-6H-噻吩并[3,2-f][1,2,4]三唑并[4,3-a][1,4]二氮杂~~草~~-6-乙酸1,1-二甲基乙酯)。在一些实施方案中,KDM4C的抑制剂是SD70(N-(呋喃-2-基(8-羟基喹啉-7-基)甲基)异丁酰胺)。在一些实施方案中,KDM6B的抑制剂是GSK-J4(3-((6-(4,5-二氢-1H-苯并[d]氮杂~~草~~-3(2H)-基)-2-(吡啶-2-基)嘧啶-4-基)氨基)丙酸乙酯,单盐酸盐)。

[0023] 另一些方面提供了用于在对象中治疗癌症的方法,其包括向对象施用靶向两种或更多种表观遗传基因的两种或更多种抑制剂的组合,所述两种或更多种表观遗传基因选自表2所示表观遗传基因之组合,其中所述两种或更多种抑制剂各自以有效量施用。在一些实施方案中,所述抑制剂各自选自CRISPR指导序列、shRNA和小分子。在一些实施方案中,在组合中施用的两种或更多种抑制剂各自的有效量小于所述抑制剂当不在组合中施用时的有效量。在一些实施方案中,两种或更多种抑制剂各自降低或阻止表观遗传基因的表达或降低或阻止由表观遗传基因编码的蛋白质的活性。在一些实施方案中,所述抑制剂各自选自CRISPR指导序列和支架序列;shRNA;和小分子。在一些实施方案中,至少一种所述抑制剂是CRISPR指导序列和支架序列;并且所述组合物还包含或编码Cas9内切核酸酶。在一些实施方案中,CRISPR指导序列或shRNA由重组表达载体表达。在一些实施方案中,表观遗传基因的组合包含BRD4和KDM4C或BRD4和KDM6B。在一些实施方案中,BRD4的抑制剂是JQ1((6S)-4-(4-氯苯基)-2,3,9-三甲基-6H-噻吩并[3,2-f][1,2,4]三唑并[4,3-a][1,4]二氮杂~~草~~-6-乙酸1,1-二甲基乙酯)。在一些实施方案中,KDM4C的抑制剂是SD70(N-(呋喃-2-基(8-羟基喹啉-7-基)甲基)异丁酰胺)。在一些实施方案中,KDM6B的抑制剂是GSK-J4(3-((6-(4,5-二氢-1H-苯并[d]氮杂~~草~~-3(2H)-基)-2-(吡啶-2-基)嘧啶-4-基)氨基)丙酸乙酯,单盐酸盐)。

[0024] 另一些方面提供了用于鉴定表观遗传基因的抑制剂之组合的方法,所述抑制剂降低细胞增殖,所述方法包括使第一细胞群体和第二细胞群体与两种或更多种CRISPR指导序

列和支架序列的多种组合以及Cas9内切核酸酶接触;培养所述第一细胞群体和第二细胞群体,使得与所述第一细胞群相比,所述第二细胞群体培养更长的持续时间;鉴定所述第一细胞群体中两种或更多种CRISPR指导序列和支架序列的组合以及所述第二细胞群体中两种或更多种CRISPR指导序列和支架序列的组合;将所述第一细胞群体中两种或更多种CRISPR指导序列和支架序列的每种组合的丰度与所述第二细胞群体中两种或更多种CRISPR指导序列和支架序列的每种组合的丰度进行比较;以及将这样的两种或更多种CRISPR指导序列和支架序列的组合鉴定为降低细胞增殖的CRISPR指导序列和支架序列的组合,在第二细胞群体中不存在或丰度降低,但在第一细胞群体中存在或丰度提高。

[0025] 还有另一些方面提供了用于鉴定待抑制以降低细胞增殖的基因之组合的方法,所述方法包括使所述第一细胞群体和第二细胞群体与两种或更多种CRISPR指导序列和支架序列的多种组合以及Cas9内切酶接触;培养所述第一细胞群体和第二细胞群体,使得与第一细胞群体相比,第二细胞群体培养更长的持续时间;鉴定所述第一细胞群体中两种或更多种CRISPR指导序列和支架序列的组合以及所述第二细胞群体中两种或更多种CRISPR指导序列和支架序列的组合;将所述第一细胞群体中的两种或更多种CRISPR指导序列和支架序列的每种组合的丰度与所述第二细胞群体中两种或更多种CRISPR指导序列和支架序列的每种组合的丰度进行比较;并将这样的两种或更多种CRISPR指导序列和支架序列的组合鉴定为待抑制以降低细胞增殖的基因的组合在第二细胞群体中不存在或丰富降低,但在第一细胞群体中存在或丰度提高。

[0026] 参考附图和本发明的详细描述,本发明的这些和另一些方面及其多个实施方案将变得更加明显。

[0027] 本发明的每种限制可包括本发明的多个实施方案。因此,预期涉及任何一个元素或元素组合的本发明的每种限制可包括在本发明的每个方面中。本发明并不将其应用局限于其在以下说明书中阐述的或在附图中所示的结构细节和组件布置。本发明能够具有另一些实施方案并且能够以多种方式实施或执行。

附图说明

[0028] 附图并不旨在按比例绘制。为了清楚起见,并非在每个附图中都标出每个组件。在附图中:

[0029] 图1示出了描述本发明非限制性实施方案的示意图。在步骤1和2中,合成寡核苷酸文库,并将相应的寡核苷酸对一起退火。每个寡核苷酸包含CRISPR指导序列、两个BbsI限制性识别位点和条形码元件。在步骤3中,在一锅(one-pot)连接反应中将寡核苷酸连接到储存载体中,得到含有寡核苷酸的载体。在步骤4中,在BbsI限制性识别位点消化载体以允许插入支架序列以及由BamHI和EcoRI限制性识别位点形成的隔离位点。可以在隔离位点处迭代地消化条形码化指导RNA文库以插入含有CRISPR指导序列、支架序列、隔离位点和条形码元件的另外的元件,从而得到具有串联的条形码元件的复杂指导RNA文库。序列从上至下对应于SEQ ID NO:364-366。

[0030] 图2A和2B示出了描绘本发明的非限制性实施方案的示意图。在图2A的步骤1中,合成寡核苷酸,每个寡核苷酸含有多个CRISPR指导序列和在CRISPR指导序列下游的单个条形码元件。在每个CRISPR指导序列之后存在限制性识别位点(RE)。在步骤2中,在一锅连接反

应中将合并的合成的寡核苷酸连接到目的载体中。如步骤3所示,可使用不同的限制性酶在每个CRISPR指导序列后的每个限制性识别位点处依次消化载体,以允许插入支架元件,并且在一些情况下允许插入驱动下游CRISPR指导序列表达的启动子元件,得到具有单个条形码元件的编码多个CRISPR指导序列和支架序列的条形码化组合指导RNA文库。在图2B的步骤1中,合成寡核苷酸,每个寡核苷酸含有多个CRISPR指导序列和在CRISPR指导序列上游的单个条形码元件。在每个CRISPR指导序列之后存在限制性识别位点(RE)。在步骤2中,在一锅连接反应中将合并的合成的寡核苷酸连接到目的载体中。如步骤3所示,可使用不同的限制性酶在每个CRISPR指导序列后的每个限制性识别位点处顺序地消化载体,以允许插入支架元件,并且在一些情况下允许插入驱动下游CRISPR指导序列表达的启动子元件,得到具有单个条形码元件的编码多个CRISPR指导序列和支架序列的条形码化组合指导RNA文库。

[0031] 图3A-3D示出了高覆盖度的组合gRNA文库的产生以及文库到人细胞的有效地递送。图3A示出了从大肠杆菌(*E. coli*)提取的质粒库(plasmid pool)中单向gRNA文库的条形码读取的累积分布,表明对所有预期组合的全面覆盖。图3B示出了在质粒库和慢病毒感染的OVCAR8-ADR-Cas9细胞库二者中的双向gRNA文库,表明对所有预期组合接近全覆盖。大多数条形码化gRNA组合在每个组合的平均条形码读数的5倍范围内检测到(由阴影区域突出显示并由箭头指示)。图3C示出了质粒库和感染的OVCAR8-ADR-Cas9细胞库内条形码表现(barcode representation)(归一化的条形码计数的 \log_2 值)之间的高相关性,表明双向文库向人细胞中的有效慢病毒递送。图3D示出了在用双向gRNA文库感染后5天培养的OVCAR8-ADR-Cas9细胞中对于两个生物学重复之间的条形码表现的高重复性。 R 是皮尔逊(Pearson)相关系数。

[0032] 图4A-4C示出了使用高通量筛选来鉴定抑制癌细胞增殖的gRNA组合。图4A示出了高通量筛选的示意图,其中OVCAR8-ADR-Cas9细胞用条形码化双向gRNA文库感染并培养15或20天。使用Illumina HiSeq鉴定和量化细胞库内的条形码表现,并在两个库之间进行比较。图4B(右图)示出了双向gRNA组合,发现其按照培养20天的细胞相对于培养15天的细胞的归一化条形码计数之间的 \log_2 比例排列来调节细胞增殖。图4B(左图)示出了与对照gRNA配对的相同gRNA。与对照gRNA对的组合以空心三角形突出显示。在另一生物学重复中证实的gRNA组合的抗增殖作用以空心圆圈突出显示(参见图12)。进一步验证了标记的gRNA组合。图4C示出了调节癌细胞增殖的双向组合的验证。用指定的双向gRNA组合感染OVCAR8-ADR-Cas9细胞群体并培养15天。然后将相等数量的细胞重新铺板并如指示培养另外一段时间。使用MTT测定法测量细胞生存力,并通过吸光度测量值($OD_{570}-OD_{650}$)($n=3$)来表征。数据代表平均值 \pm 标准差(SD)。

[0033] 图5A-5D示出了KDM4C与BRD4或KDM6B与BRD4的组合抑制对人卵巢癌细胞生长的抑制。图5A示出了相对于用表达对照gRNA的慢病毒感染的细胞,用表达所指示的单个或组合gRNA之慢病毒感染的OVCAR8-ADR-Cas9细胞的细胞生存力的倍数变化。将细胞培养15天,然后将相等数量的经感染细胞重新铺板并再培养5天。图5B示出了相对于用表达对照shRNA的慢病毒感染的细胞,用表达所指示的shRNA之慢病毒共感染的OVCAR8-ADR细胞的细胞生存力的倍数变化。将细胞培养9天,然后将相等数量的经感染细胞重新铺板并再培养4天。图5C示出了相对于未接受药物的对照细胞,用指定浓度的SD70和JQ1处理5天的OVCAR8-ADR细胞的细胞生长抑制百分比。对于SD70和JQ1的组合(中间图和右图),还示出了所计算的超过所

预测的Bliss独立模型和HSA模型的过量抑制。图5D示出了相对于未接受药物的对照细胞,用指定浓度的GSK-J4和JQ1处理7天的OVCAR8-ADR细胞的细胞生长抑制百分比。对于GSK-J4和JQ1的组合(中间图和右图),还示出了计算的超过所预测的Bliss独立模型和HSA模型的过量抑制。通过MTT测定法测定细胞生存力。数据代表平均值 \pm SD(对于(图5A), $n=3$;对于(图5B-5D), $n=6$)。星号(* $P<0.05$)和井号(# $P<0.05$)分别表示使指定样品之间和药物处理的样品之间相对于无药物的对照样品之间的显著差异。

[0034] 图6A-6E示出了组合gRNA表达构建体的慢病毒递送提供有效的靶基因抑制。图6A示出了用于在人细胞中测试慢病毒组合gRNA表达构建体的策略的示意图。产生了这样的慢病毒,其含有编码分别在UBC和CMV启动子控制下表达的RFP和GFP的基因,以及靶向RFP(RFP-sg1或RFP-sg2)和GFP(GFP-sg1)序列的gRNA的由串联U6启动子驱动的表达盒。使用慢病毒来感染OVCAR8-ADR或OVCAR8-ADR-Cas9细胞,并使用流式细胞术和荧光显微术评估GFP和RFP表达。图6B示出了评估用编码所指示的gRNA表达构建体之慢病毒感染4天的细胞中GFP和RFP表达的流式细胞术散点图。编码组合gRNA表达构建体的慢病毒降低了OVCAR8-ADR-Cas9细胞中RFP和GFP荧光阳性细胞的百分比,而OVCAR8-ADR细胞中则没有。图6C示出了在用编码所指示的gRNA表达构建体的慢病毒感染后第4天GFP(左栏)和RFP(右栏)阳性细胞的百分比。图6D示出了在用编码所指示的gRNA表达构建体的慢病毒感染后第8天GFP(左栏)和RFP(右栏)阳性细胞的百分比。检测到靶向RFP和GFP的gRNA之间有限的交叉反应性。图6B中的数据代表感染4天的细胞的流式细胞术测量值,而6C和6D的量化表示平均值 \pm 标准差($n=3$)。图6E示出了代表性的荧光显微照片,证明组合gRNA表达构建体在感染后第3天在OVCAR8-ADR-Cas9细胞中有效地抑制了RFP和GFP荧光水平,而在OVCAR8-ADR细胞中则没有。

[0035] 图7A-7C示出了OVCAR8-ADR-Cas9细胞中靶标基因的gRNA的切割效率。图7A示出了这样的汇总表,其显示使用Surveyor测定法在OVCAR8-ADR-Cas9细胞中检测到的插失(indel)百分比。用从筛选文库随机选择的8种不同的gRNA感染细胞8或12天。对于Surveyor测定法,检测到的未切割的和经切割的PCR产物的预期大小以碱基对列出。图7B和7C示出了这样的琼脂糖凝胶,其显示未被感染或被所指示的gRNA感染8或12天的OVCAR8-ADR-Cas9细胞中DNA切割效率的Surveyor测定结果。

[0036] 图8A和8B示出了OVCAR8-ADR-Cas9细胞中双gRNA表达构建体在靶定的基因处的切割效率。图8A示出了以碱基对列出的Surveyor测定法检测到的未切割的和经切割的PCR产物的预期大小(上图)。琼脂糖凝胶示出了使用Surveyor测定法在用所指示的单gRNA或双gRNA表达构建体感染12天的OVCAR8-ADR-Cas9细胞中检测到的插失百分比(下图)。图8B是示出了用载体对照或所指示的单gRNA或双gRNA构建体感染15天的OVCAR8-ADR-Cas9细胞中蛋白质水平的免疫印迹分析。

[0037] 图9A-9C示出了用双gRNA表达构建体感染的单细胞衍生的OVCAR8-ADR-Cas9克隆之靶定等位基因的DNA比对。图9A示出了来自用表达靶向BMI1和PHF8之sgRNA的慢病毒感染的OVCAR8-ADR-Cas9细胞之序列的比对。序列从上到下分别对应于SEQ ID NO:

[0038] 203,204,203,203,204,204,203,205,206,206,203,203,204,207,208,209,204,210,208,208,210,210,203,219,206,206,208,208,220,220,221,222,204,204,223,224,204,204,203,203,204,和204。

[0039] 图9B示出了来自用表达靶向BRD4和KDM4C之gRNA的慢病毒感染的OVCAR8-ADR-Cas9细胞之序列的比对。将OVCAR8-ADR-Cas9细胞用慢病毒感染12天并铺板为单细胞。提取每个单细胞扩增克隆的基因组DNA。通过PCR扩增靶标等位基因,并通过TA克隆插入到TOPO载体中进行Sanger测序。示出了每个克隆的两个等位基因的序列。核苷酸的突变和插入以粗体表示,而缺失则表示为“-”。示出了靶定基因的野生型(WT)序列作为参考,其中20bp的gRNA靶标加下划线,PAM序列以粗斜体显示。序列从上至下分别对应于SEQ ID NO:

[0040] 211,212,211,211,213,213,211,211,214,214,211,215,212,216,211,211,217,218,211,211,212,212,211,211,225,226,211,211,226,227,211,211,212,212,211,228,226,229,211,211,229,226,211,211,230,和226。

[0041] 图9C是显示单基因和双基因编辑的细胞之频率的维恩图。将含有所指示的双gRNA表达构建体的OVCAR8-ADR-Cas9细胞通过FACS铺板为单细胞。用Illumina MiSeq从40个全基因组扩增的单细胞中对靶定的等位基因进行测序。75% (即,30/40) 和80% (即,32/40) 的单细胞分别在靶标BMI1和PHF8基因座处具有至少一个突变等位基因。62.5% (即,25/40) 的单细胞在BMI1和PHF8基因二者中含有至少一个突变等位基因。每个单细胞的两个等位基因的序列示于表6中。观察到通过Sanger测序(图9B)由单细胞来源的克隆和通过Illumina MiSeq由全基因组扩增的单细胞(图9C)测定的类似突变等位基因频率。

[0042] 图10A和10B示出了对于组合gRNA筛选的生物学重复之间的条形码定量的高重复性。图10A示出了这样的散点图,其比较了使用双向gRNA文库感染后培养15天的OVCAR8-ADR-Cas9细胞的两个生物学重复之间的条形码表现(归一化的条形码计数的 \log_2 数)。图10B示出了这样的散点图,其比较了用双向gRNA文库感染后培养20天的OVCAR8-ADR-Cas9细胞的两个生物学重复之间的条形码表现(归一化的条形码计数的 \log_2 数)。R是皮尔逊相关系数。

[0043] 图11示出了在表达构建体内以不同顺序排列的相同gRNA组合之间的条形码定量的一致性倍数变化。对于以不同顺序排列的每个双向gRNA组合(即,sgRNA-A+sgRNA-B和sgRNA-B+sgRNA-A),测定了变异系数(coefficient of variation,CV;定义为培养20天的OVCAR8-ADR-Cas9细胞相对于培养15天的归一化条形码计数之倍数变化的SD/平均值)。在细胞增殖筛选中,超过82%的双向gRNA组合的 $CV < 0.2$,95%的双向gRNA组合的 $CV < 0.4$ 。

[0044] 图12A-12C示出了鉴定抑制癌细胞增殖的gRNA对的组合筛选的生物学重复。图12A示出了用图4B中使用的相同双向gRNA文库感染的OVCAR8-ADR-Cas9细胞的 \log_2 倍数变化。将调控增殖的指导RNA对(右图)及其gRNA+对照对应物(即,基因靶向gRNA+对照gRNA;左图)的组合,按照培养20天的细胞与培养15天的细胞相比的归一化条形码计数的 \log_2 比进行排列。在另一个生物学重复中证实的gRNA组合的抗增殖作用以空心圆圈突出显示(图4B),而与对照gRNA对的组合以空心三角形突出显示。在图4C中进一步验证标记的gRNA组合。图12B示出了这样的散点图,其显示了用双向gRNA文库感染的OVCAR8-ADR-Cas9细胞的两个生物学重复之间,培养20天的细胞相对于培养15天的归一化条形码计数的 \log_2 比。图12C示出了在合并的筛选中gRNA组合的 \log_2 比的频率分布。所示出的 \log_2 比由两个生物学重复的平均值计算。

[0045] 图13示出了合并的筛选和验证数据中单一命中(individual hit)之间的高度一致性。对于每个双向gRNA组合,将从合并的筛选数据(“筛选表型”)获得的培养20天的细胞

相对于培养15天的归一化条形码计数的倍数变化与从单细胞增殖测定(“验证表型”)确定的载体对照相比,相对于其相对细胞生存力作图($R=0.932$)。筛选表型的数据是两个生物学重复的平均值;单一验证表型代表三次独立实验的平均值。 R 是皮尔逊相关系数。

[0046] 图14A-14F示出了OVCAR8-ADR细胞中由shRNA介导的靶标基因的敲低。图14A示出了表达对照shRNA或靶向KDM4C的shRNA的OVCAR8-ADR细胞中KDM4C的相对mRNA水平。图14B示出了表达对照shRNA或靶向BRD4的shRNA的OVCAR8-ADR细胞中BRD4的相对mRNA水平。图14C示出了表达对照shRNA或靶向KDM6B的shRNA的OVCAR8-ADR细胞中KDM6B的相对mRNA水平。通过qRT-PCR对mRNA水平定量,并将其相对于肌动蛋白mRNA水平归一化。数据代表平均值 \pm SD($n=3$)。图14D-14F示出了表达对照shRNA或靶向KDM4C、BRD4或KDM6B的shRNA的OVCAR8-ADR细胞中相对蛋白质水平的Western印迹分析。将测量的蛋白质水平相对于肌动蛋白水平归一化,相对于对照shRNA样品归一化,并绘制为下图中的相对蛋白质水平。星号($*P<0.05$)表示表达基因靶向shRNA相对于对照shRNA的细胞之间mRNA或蛋白质水平的显著差异。

[0047] 图15示出了用于组装条形码化的组合gRNA文库的策略。合成条形码化gRNA寡核苷酸对,使其退火并克隆到合并形式的储存载体中。将具有gRNA支架序列的寡核苷酸插入到合并的储存载体文库中以产生条形码化sgRNA文库。详细的组装步骤如图1所示。使用CombiGEM策略构建组合gRNA文库。将使用BglIII和MfeI消化从sgRNA文库制备的合并的条形码化sgRNA插入物通过BamHI和EcoRI消化在目的载体中产生的相容性突出末端连接。迭代一锅连接产生的具有独特条形码的(n)向gRNA文库对应于在一端串联的gRNA,从而使得能够通过下一代测序追踪在合并的群体中的单个组合成员。

[0048] 图16A-16E示出了在OVCAR8-ADR-Cas9细胞中在靶标基因组基因座处gRNA插失分析的深度测序结果。用指定的sgRNA感染细胞15天,然后进行深度测序。图16A示出了插失频率。图16B示出了移码突变和框内突变的百分比。图16C示出了插失大小的分布。图16D示出了在使用单sgRNA(KDM4C或BRD4)(上图)或双gRNA表达构建体(KDM4和BRD4)(下图)感染15天的OVCAR8-ADR-Cas9细胞中通过靶标基因组基因座的深度测序分析的插失分布。图16E示出了在使用单sgRNA(PHF8或BMI1)(上图)或双gRNA表达构建体(PHF8和BMI1)(下图)感染15天的OVCAR8-ADR-Cas9细胞中通过靶向基因组基因座的深度测序分析的插失分布。

[0049] 图17A-17C示出了混合细胞群体中促增殖性gRNA和抗增殖性gRNA之频率的数学建模。图17A示出了在具有不同分数(即,2%,5%或10%)的初始含抗增殖性gRNA(f_s)和促增殖性(f_f)gRNA的细胞的细胞群体中,抗增殖性gRNA和促增殖性gRNA的相对频率的模拟。相对频率被定义为与初始时间点相比,在给定时间处的条形码丰度。在本实施例中,由于CRISPR-Cas9系统(p)的遗传扰动,具有改变生长速率的细胞的分数设定为1.0(即,100%),抗增殖克隆的倍增时间($T_{\text{倍增},m}$)为48小时。图17B和17C示出了相对于参数p、 $T_{\text{倍增},m}$ 、 f_s 和 f_f 的变化,混合细胞群体中抗增殖性gRNA的模型化相对频率。在图17B和17C各图中,从顶部到底部的线表示 $p=0.2$, $p=.4$, $p=0.6$, $p=0.8$ 和 $p=1.0$ 。在图17A-17C中,促增殖性克隆的倍增时间设定为12小时。具体定义在实施例3中描述。

[0050] 图18示出了单一gRNA组合的合并筛选和验证数据。对于每个gRNA组合,将从合并的筛选数据(“筛选表型”)获得的培养20天的细胞相对于培养15天的归一化条形码计数的倍数变化与从单细胞增殖测定(“验证表型”)确定的载体对照相比,相对于其相对细胞生存

力作图。从与三个对照sgRNA各自配对的相应sgRNA的倍数变化来平均每个单个sgRNA的筛选表型。筛选数据的数据是两个生物学重复的平均值,而单一验证数据代表平均值 \pm SD ($n \geq 3$)。

[0051] 图19示出了靶向KDM4C、KDM6B和BRD4的gRNA的中靶和脱靶 (on-target and off-target) 插失产生速率的测量。每一行代表对应于20bp指导序列(黑色),接着是3bp PAM序列(灰色)的基因组基因座。粗黑色字体的序列代表gRNA的中靶基因组序列。KDM4C-sg1、KDM6B-sg2和BRD4-sg3的每个虚线下面都是使用CRISPR设计工具(Ran,等,Nature Protocols (2013) 8:2281-2308)和CCTop (Stemmer, PLoS One (2015) 10:e0124633)所鉴定的预测外显子脱靶基因组序列。还评估了对BRD4-sg2预测的五个外显子/内含子脱靶位点。加下划线的核苷酸突出显示了脱靶序列与中靶序列之间的差异。从 $\sim 10,000$ 个细胞中对每个基因组基因座进行PCR扩增,并用 > 420 万次读取进行深度测序。*n.d.*表示基因组序列的PCR未能提供用于测序的特异性扩增子。序列从上到下对应于SEQ ID NO:368-393。

[0052] 图20A-20B示出了携带KDM4C和BRD4两者移码突变的OVCAR8-ADR-Cas9细胞的生长降低。图20A描绘了对具有KDM4C和BRD4两者移码突变的单细胞扩增的OVCAR8-ADR-Cas9突变体克隆(即,来源自图9B所示的克隆#3)的细胞生长测定。铺板相等数量的细胞,并培养5天,然后进行MTT测定。数据代表相对于对照OVCAR8-ADR-Cas9细胞的吸光度测量值($OD_{570} - OD_{650}$)的平均值 \pm SD ($n=3$)。星号(* $P < 0.01$)代表对照和突变细胞之间的显著差异。图20B示出了来自图20A的对照和突变细胞中蛋白质水平的免疫印迹分析。

[0053] 图21A-21B示出了用gRNA表达构建体感染的OVCAR8-ADR-Cas9细胞的RNA测序分析。图21A示出了显示由相应的单gRNA或双gRNA靶向的在OVCAR8-ADR-Cas9细胞的每个样品(列)中每个基因转录本(行)之相对表达水平的代表性热图。在热图中包括当与载体对照相比时,在用指定的gRNA感染的OVCAR8-ADR-Cas9细胞中鉴定为显著差异性表达的转录本。值是使用RNA-Seq所测量的 \log_2 转化的FPKM,并且以转录本为中心平均值。基于皮尔逊相关性进行转录本和样品的系统聚类(Hierarchical clustering)。图21B示出了当与载体对照相比,在用指定的gRNA感染的OVCAR8-ADR-Cas9细胞中所鉴定的差异表达基因中,生物学过程的前十富集性基因集(Q值 < 0.05)。与基因集(y轴)相关的差异表达基因(x轴)的子集在表中为灰色阴影。

[0054] 图22A-22B示出了KDM4C与BRD4以及KDM6B与BRD4对另外的癌细胞系的细胞生长的作用。图22A示出了组合gRNA表达构建体在乳腺癌MDA-MB231-Cas9和胰腺癌Bx-PC3-Cas9细胞中有效抑制靶标荧光基因。在具有或不具有靶向RFP和GFP序列的gRNA的串联U6启动子驱动的表达盒的情况下,将含有由组成型启动子表达的RFP和GFP基因的慢病毒载体递送至MDA-MB231-Cas9和Bx-PC3-Cas9细胞,以用于根据流式细胞术进行GFP和RFP表达的分析。详细的策略描述在图6中。编码组合gRNA表达构建体的慢病毒在感染后第4天降低RFP和GFP荧光阳性细胞的百分比。图22B示出了将用表达所指示的单gRNA或组合gRNA的慢病毒感染的MDA-MB231-Cas9和Bx-PC3-Cas9细胞培养14天。然后将相等数量的感染细胞重新铺板并再培养5天。通过MTT测定法测定相对于对照sgRNA的细胞生存力。数据代表生物学重复的平均值 \pm SD ($n=6$)。星号(* $P < 0.05$)表示指示的样品之间的显著差异。这些结果表明,取决于细胞背景,靶向表观遗传基因的组合gRNA可具有可变的表型。

[0055] 发明详述

[0056] 产生用于表达包含多于一个sgRNA(指导序列和支架序列)的多个CRISPR系统的载体和遗传元件是非常费力的,并且使用传统克隆方法构建的CRISPR系统的文库的复杂度非常有限。本文描述的方法产生具有串联条形码以及CRISPR指导序列和支架序列的载体。这些方法对于构建用于组合遗传筛选的大型文库可能是高效的,并且利用了这样的事实:因为指导序列和条形码元件的长度较短,大量寡核苷酸可被容易地复印(print),并且具有靶标特异性确定区的指导序列可被印到这些寡核苷酸上。

[0057] 本文描述的产生CRISPR构建体和载体的大规模并行组合遗传学方法允许迅速产生包含能够靶向宿主细胞核酸的CRISPR系统(CRISPR指导序列和支架序列)之组分的组合遗传构建体组。这些方法还使得能够合并筛选多种组合顺序(例如,可将双向、三向和n向组合合并并同时一起筛选),鉴定给定应用所需的最小组合。组合的遗传构建体组(例如使用本文所述的方法产生的那些)可用于鉴定协同相互作用以调控细胞过程或表型(例如癌细胞生长)的基因和遗传途径。本文中描述了使用本文所述的组合CRISPR构建体所鉴定的表观遗传基因的新组合,当一起抑制时,其具有抗癌效果,例如降低细胞增殖。

[0058] 本公开内容的一些方面涉及遗传构建体、包含遗传构建体的载体、组合载体的载体以及使用大规模并行组合遗传学方法产生组合载体的方法,其可见于PCT公布号W02014/00542,其通过引用整体并入本文。如本文使用的“遗传构建体”是指包含CRISPR指导序列和支架序列以及条形码元件的一个或更多个DNA元件,使得每个DNA元件与条形码元件关联。如本文使用的,特定DNA元件和条形码元件之间的关联意味着特定DNA元件和条形码元件总是包含在相同的遗传构建体内。因此,遗传构建体中特定条形码元件的存在或检测表明所关联的特定DNA元件也存在于相同遗传构建体内。

[0059] 在宿主细胞中,包含CRISPR指导序列和支架序列的DNA元件被转录并形成CRISPR小指导RNA(sgRNA),所述CRISPR小指导RNA用于将内切核酸酶募集到宿主细胞中的特定靶核酸,其可能导致位点特异性CRISPR活性。如本文使用的“CRISPR指导序列”是指与宿主细胞中的靶核酸序列互补的核酸序列。CRISPR指导序列将sgRNA靶向靶核酸序列(也称为靶位点)。与靶核酸互补的CRISPR指导序列的长度可以为15至25个核苷酸、18至22个核苷酸或19至21个核苷酸。在一些实施方案中,与靶核酸互补的CRISPR指导序列的长度为15、16、17、18、19、20、21、22、23、24或25个核苷酸。在一些实施方案中,与靶核酸互补的CRISPR指导序列的长度为20个核苷酸。

[0060] 应理解,如果CRISPR指导序列能够与靶核酸杂交,则CRISPR指导序列与宿主细胞中的靶核酸互补。在一些实施方案中,CRISPR指导序列与靶核酸互补性为至少50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或至少100%(也参见美国专利8,697,359,其对于CRISPR指导序列与靶多核苷酸序列的互补性的教导通过引用并入)。已证明,CRISPR指导序列与靶核酸在靶核酸3'末端附近的错配可能会消除核酸酶切割活性(Upadhyay等,Genes Genome Genetics(2013)3(12):2233-2238)。在一些实施方案中,CRISPR指导序列与靶核酸的3'末端(例如,靶核酸的3'末端的最后5、6、7、8、9或10个核苷酸)的互补性为至少50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%或至少100%。

[0061] CRISPR指导序列可从本领域已知的任何来源获得。例如,CRISPR指导序列可以是宿主细胞的核酸中存在的指定长度的任何核酸序列(例如,基因组核酸和/或基因组外核

酸)。在一些实施方案中,可以将CRISPR指导序列设计并合成为靶向所期望的核酸,例如编码转录因子、信号转导蛋白、转运蛋白等的核酸。在一些实施方案中,将CRISPR指导序列设计并合成为靶向表观遗传基因。例如,可以将CRISPR指导序列设计成靶向表2中所示表观遗传基因的任何组合。在一些实施方案中,CRISPR指导序列包含表1中提供的任何示例性CRISPR指导序列。

[0062] 如本文使用的“支架序列”(也称为tracrRNA),是指将内切核酸酶募集到与互补性CRISPR指导序列结合(杂交)的靶核酸的核酸序列。包含至少一个茎环结构并招募内切核酸酶的任何支架序列均可用于本文所述的遗传元件和载体中。示例性支架序列对于本领域技术人员将是明显的,并且可见于例如Jinek等,Science (2012) 337 (6096):816-821、Ran等,Nature Protocols (2013) 8:2281-2308、PCT申请号W02014/093694和PCT申请号W02013/176772。

[0063] 术语“靶核酸”、“靶位点”和“靶序列”在全文中可互换使用,并且是指宿主细胞中可能被本文所述的CRISPR指导序列靶向的任何核酸序列。靶核酸的下游(3'侧)侧翼为前间区临近基序(protospacer adjacent motif,PAM),其可与内切核酸酶相互作用并且还参与将内切核酸酶活性靶向靶核酸。通常认为靶核酸侧翼的PAM序列取决于内切核酸酶和获得所述内切核酸酶的来源。例如,对于来源于化脓性链球菌(*Streptococcus pyogenes*)的Cas9内切核酸酶,PAM序列是NGG。对于来源于金黄色葡萄球菌(*Staphylococcus aureus*)的Cas9内切核酸酶,PAM序列是NNGRRT。对于来源于脑膜炎奈瑟氏菌(*Neisseria meningitidis*)的Cas9内切核酸酶,PAM序列是NNNNGATT。对于来源于嗜热链球菌(*Streptococcus thermophilus*)的Cas9内切核酸酶,PAM序列是NNAGAA。对于衍生自齿垢密螺旋体(*Treponema denticola*)的Cas9内切核酸酶,PAM序列是NAAAAC。对于Cpf1核酸酶,PAM序列是TTN。

[0064] 在一些实施方案中,CRISPR指导序列和支架序列被表达为不同的转录本。在这样的实施方案中,CRISPR指导序列还包含与支架序列的一部分互补的另外的序列,并且其用作与支架序列结合(杂交)并将内切核酸酶募集到靶核酸。在另一些实施方案中,CRISPR指导序列和支架序列被表达为单一转录本,作为可被称为单指导RNA(sgRNA)的嵌合RNA。sgRNA具有与靶核酸结合(杂交)以及将内切核酸酶募集到靶核酸的双重功能。在这样的实施方案中,支架序列还可包含接头环(linker loop)序列。

[0065] 条形码元件可用作遗传构建体的识别物并且可指示载体或遗传元件中一个或更多个特异性CRISPR指导序列的存在。一组条形码元件的成员具有足够独特的核酸序列,使得每个条形码元件可容易地与该组的其他条形码元件区分开。条形码元件可以是任何长度的核苷酸,但优选长度小于30个核苷酸。在一些实施方案中,条形码元件是6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、25、26、27、28、29或30个或更多个核苷酸。使用检测条形码元件和确定条形码元件或多个条形码元件的核酸序列来确定遗传构建体的相关DNA元件的存在。如本文所述的条形码元件可通过本领域已知的任何方法检测,包括测序或微阵列方法。

[0066] 图1示出了与本发明相关的遗传构建体的非限制性实例的几个示意图。在图1步骤4中,包含CRISPR指导序列(称为“指导序列”)和支架序列(称为“支架”)的DNA元件的侧翼是以“BamHI”指示的第一相容性末端元件和以“BglII”指示的第二相容性末端元件,其能够与

彼此退火。遗传构建体还包含条形码元件(称为“条形码”),其侧翼是以“EcoRI”指示的第三相容性末端元件和以“MfeI”指示的第四相容性末端元件,其能够与彼此退火但是不能与第一相容性末端元件和第二相容性末端元件退火。遗传构建体还包含隔离位点,使得条形码元件位于隔离位点的一侧,而DNA元件位于隔离位点的另一侧。图1还描绘了允许DNA元件表达(转录)的DNA元件上游(相对于DNA元件的5')的启动子元件。虽然图1描绘了DNA元件在条形码元件上游(相对于条形码元件的5'),该排列也可以颠倒。

[0067] 相容性末端可以以对于本领域技术人员而言明显的多种方式产生,并且可以由多种不同的序列组成。如本文使用的“相容性末端元件”是指能够彼此连接或退火的DNA区域。能够彼此结合或退火的相容性末端元件对于本领域技术人员来说是明显的,并且是指在核苷酸序列中彼此互补且因此能够彼此碱基配对的末端元件。在几个非限制性实施方案中,相容性末端元件可由以下组成:具有相容性突出末端的限制性位点、Gibson组装序列或任何其他DNA组装方法的功能元件,包括重组酶、大范围核酸酶、TAL效应物/锌指核酸酶、反式切割核酸酶/DNA酶或整合酶。

[0068] 在一些实施方案中,Gibson组装用于产生相容性突出末端。Gibson组装是指可在单个反应中连接多个DNA片段的等温DNA末端连接技术。该方法在Gibson等,(2009) Nature Methods 6:343-5中进一步描述并通过引用并入。

[0069] 在另一些实施方案中,限制性消化用于产生相容性末端,如图1所示。使用这种方法,两个独特的限制性酶产生相容性突出末端。当这些突出末端连接时,产生不再被任何酶识别的瘢痕(scar)。应理解,可使用产生相容性突出末端的任何限制性酶。在一些非限制性实施方案中,使用例如**BIOBRICKS®**(The BioBricks Foundation)或BglBricks(Anderson等(2010) Journal Of Biological Engineering 4:1)的标准生物学组件(biological parts)以及与这种标准生物学组件相关的酶。标准生物学组件例如**BIOBRICKS®**或BglBricks的用途对于本领域普通技术人员是常规的。应理解,尽管可使用经典的限制性酶(例如I、II或III型限制性酶),但是也可使用其他DNA切割分子。例如,靶标核酸酶可用于特异性靶位点的切割。也可使用大范围核酸酶来使插入DNA元件之干扰的可能性最小化。TALE或ZF核酸酶也可用于靶向长DNA位点以使插入的DNA元件中内部切割的可能性最小化。此外,**TOPO®**克隆可用于完成限制性消化和结合。

[0070] 在一些实施方案中,通过用限制性酶BamHI识别和切割来产生第一相容性末端元件,并通过用限制性酶BglIII识别和切割来产生第二相容性末端元件。在一些实施方案中,通过用限制性酶MfeI识别和切割来产生第三相容性末端元件,并通过用限制性酶EcoRI识别和切割来产生第四相容性末端元件。

[0071] 如本文使用的遗传构建体的“隔离位点”是指允许构建体线性化的区域。应理解,隔离位点是构建体核酸内的位点,在该位点处切割使构建体线性化并且可允许插入另外的遗传元件。在一些实施方案中,隔离位点是限制性酶识别位点。例如,在图1中,隔离位点由分别由BamHI和EcoRI识别位点指示的第一和第四相容性末端元件形成。使用相应的限制性酶(BamHI和EcoRI)切割构建体使构建体线性化,并允许插入另外的遗传构建体。在一些实施方案中,隔离位点由一个识别位点形成。在一些实施方案中,隔离位点由多于一个识别位点形成。

[0072] 本发明的一些方面涉及用于产生包含本文所述的遗传构建体的组合载体的方法。

如图1的步骤3所示,所述方法包括提供含有第一遗传构建体的载体,所述第一遗传构建体包含称为“20bp指导序列”的CRISPR指导序列,其侧翼为由“BglII”指示的第二相容性末端元件,和称为“BbsI”的第一限制性酶的第一识别位点;称为“条形码”的条形码元件,其侧翼为由“MfeI”指示的第三相容性末端元件,以及第二限制性酶的第二识别位点。在一些实施方案中,可通过将包含相容性末端的第一遗传构建体与经切割的载体退火并连接来产生载体,如图1的步骤2所示。在一些实施方案中,例如通过寡核苷酸阵列合成来合成第一遗传构建体。第一遗传构建体可在第一识别位点处切割,得到第五相容性末端元件,并且在第二识别位点处切割,得到第六相容性末端元件。提供了一种支架元件,其包含支架序列和由“BamHI”和“EcoRI”指示的隔离位点,其侧翼是能够与经切割的载体的第五相容性末端元件退火的第七相容性末端元件,和能够与经切割的载体的第六相容性末端元件退火的第八相容性末端元件。使用相容性末端元件将支架元件与载体的经切割的第一遗传构建体退火。在退火后,支架元件整合在CRISPR指导序列与条形码元件之间,且隔离位点位于支架序列与条形码元件之间。

[0073] 应理解,可使用多种不同的酶组合来切割第一和第二识别位点。在一些实施方案中,通过相同的限制性酶识别位于CRISPR指导序列和条形码元件外部的两个识别位点,其产生带有支架元件的相容性末端。在另一些实施方案中,由两种不同的限制性酶识别位于CRISPR指导序列和条形码元件外部的两个限制性位点,其各自产生带有支架元件的相容性末端。

[0074] 本发明的另一些方面涉及组合构建体和用于产生组合构建体的方法。如本文使用的“组合构建体”是指包含多个DNA元件的遗传构建体。如本文使用的,多个DNA元件是指多于一个DNA元件,每个DNA元件包含CRISPR指导序列和支架序列。如图1的步骤5所示,组合构建体的产生可涵盖通过在遗传构建体内的隔离位点切割载体,使包含与本发明相关的第一遗传构建体的载体线性化。可将与本发明相关的第二遗传构建体插入到经切割的载体中,并退火以及与载体连接。如本文使用的“插入物”是指旨在插入经切割载体中的遗传构建体。在一些实施方案中,例如通过PCR或限制性消化,从载体中纯化插入物。可通过插入物内终端相容性末端元件与线性化载体中其相容性组分的退火而将插入物与经切割的载体连接。

[0075] 图1的步骤5的(n)向指导RNA文库描绘了包含多个DNA元件和多个相应的条形码元件的组合后组合构建体。在图1的步骤5中描绘的非限制性实施例中,遗传构建包含四个不同的DNA元件和四个相应的条形码元件。组合构建体还包含位于多个条形码元件与多个DNA元件之间的隔离位点。

[0076] 本文描述的用于产生组合构建体的方法可以是迭代的。例如,图1所描绘的组合载体可在隔离位点处再次被切割,并且可将一个或更多个另外的插入物连接到组合构建体中,同时为另外的插入物保持隔离位点。重要的是,在整个迭代过程中,随着遗传构建体内DNA元件的数量持续提高,与每个DNA元件相关联的独特条形码与其相关的DNA元件保持在相同的遗传构建体内。在一些实施方案中,组合过程重复至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20次或多于20次。在一些实施方案中,该过程重复第n次,其中n可以是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或大于20的数字。

[0077] 应理解,组合构建体可包含任何数量的DNA元件和相关联的条形码元件。在一些实

施方案中,组合构建体包含2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或多于100个DNA元件和相关联的条形码元件。

[0078] 本发明的另一方面涉及包含与单一条形码元件相关联的多于一个CRISPR指导序列的遗传构建体和载体。图2示出了与本发明相关的遗传构建体的非限制性实例的几个示意图。图2A和2B的步骤1示出了包含以下的遗传构建体:称为“20bp指导序列A”、“20bp指导序列B”和“20bp指导序列C”的三个CRISPR指导序列,由“条形码”表示的条形码元件和位于每个CRISPR指导序列之间以及条形码元件与最接近条形码元件的CRISPR指导序列之间的识别位点。在一些实施方案中,条形码元件可位于CRISPR指导序列的下游,如图2A所示。在另一些实施方案中,条形码元件可位于CRISPR指导序列的上游,如图2B所示。在一些实施方案中,位于CRISPR指导序列之间的识别位点和位于条形码元件与最接近条形码元件的CRISPR指导序列之间的识别位点各自是不同限制性酶的不同识别位点。在一些实施方案中,包含至少两个CRISPR指导序列、条形码元件和识别位点的遗传构建体通过本领域已知的任何方法合成,例如通过寡核苷酸阵列合成。

[0079] 包含多个DNA元件和一个条形码元件的遗传构建体也在本发明的范围内。在图2的步骤3中,遗传构建体包含三个DNA元件,其各自包含CRISPR指导序列和支架序列、条形码元件和位于每个DNA元件上游的启动子序列。在一些实施方案中,条形码元件可位于CRISPR指导序列的下游,如图2A所示。在另一些实施方案中,条形码元件可位于CRISPR指导序列的上游,如图2B所示。

[0080] 本发明的一些方面涉及产生包含本文中所述的遗传构建体之组合载体的方法。在一些实施方案中,所述方法包括提供这样的载体,其包含多个CRISPR指导序列和位于多个CRISPR指导序列下游的条形码元件。如图2A和2B的步骤1所示,三个CRISPR指导序列被称为“20bp指导序列A”、“20bp指导序列B”和“20bp指导序列C”,条形码元件由“条形码”表示。载体还包含多种限制性酶的多个识别位点。在图2A的步骤1中,每个识别位点位于CRISPR指导序列的下游并由“RE1”、“RE2”和“RE3”表示。在图2B的步骤1中,每个识别位点位于CRISPR指导序列的上游并由“RE1”、“RE2”和“RE3”表示。在一些实施方案中,载体还包含位于至少一个CRISPR指导序列的上游的启动子序列。至少一个CRISPR指导序列下游的相容性末端元件通过本领域已知的任何方法生成。在一些实施方案中,用限制性酶在至少一个识别位点处切割载体,得到第一相容性末端元件和第二相容性末端元件。提供了支架元件,其包含支架序列、任选的启动子序列和能够与经切割载体的第一相容性末端元件和第二相容性末端元件退火的第三相容性末端元件和第四相容性末端元件。通过将支架元件内的终端相容性末端元件与经切割的载体内的相容性组分的退火而使支架元件与经切割的载体退火。本文所述的方法可以是迭代的,以得到包含多个CRISPR指导序列和支架序列和一个条形码元件的组合载体。

[0081] 应理解,组合载体可包含与一个条形码元件相关联的任何数量的DNA元件。在一些实施方案中,组合构建体包含2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、

47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100或多于100个DNA元件和一个条形码元件。与一个条形码元件相关联的DNA元件的数量可取决于能够被合成的包含CRISPR指导序列、条形码和识别位点的遗传构建体的长度。

[0082] 在本文所述的任何构建体或载体中,可将一个或更多个RNA结构域插入到一个或更多个CRISPR指导序列中。在一些实施方案中,CRISPR指导序列与一个或更多个RNA结构域融合。在一些实施方案中,RNA是非编码RNA或其片段。在这样的实施方案中,RNA结构域可靶向DNA位点。这样的构建体或载体可用于CRISPR展示。

[0083] 本发明的另一些方面涉及用于鉴定遗传构建体或载体内的一个或更多个DNA元件的方法。在组合事件之后,与特定DNA元件相关联的独特的条形码保留在与特定DNA元件相同的遗传构建体内。因此,条形码元件或多个条形码元件的鉴定允许鉴定在相同遗传构建体内所关联的DNA元件或多个DNA元件。在一些实施方案中,条形码元件和/或DNA元件的序列通过测序或通过微阵列分析来确定。应理解,确定DNA序列的任何手段均适用于鉴定一个或更多个条形码元件和相应的DNA元件。重要的是,在组合构建体中,例如图1的步骤5所描绘的,多个条形码元件彼此紧密靠近,这允许通过例如DNA测序的方法同时迅速鉴定多个条形码元件并因此鉴定多个DNA元件。

[0084] 本发明的另一些方面涉及本文所述的包含两种或更多种遗传构建体的文库,其适用于大规模并行组合遗传学方法。如本文使用的,遗传构建体的文库是指两种或更多种遗传构建体的集合。在一些实施方案中,产生这样的遗传构建体文库,其中每个独特的DNA元件在质粒上。可合并该质粒文库以形成载体文库。例如,可通过在载体文库上进行PCR来产生插入物文库。在第一组合事件中,所有载体都可与所有插入物配对,从而产生配对组合的完整的组合集。该配对文库与插入物文库之间的进一步反应可得到产生自单个载体文库的三向、四向或多于四向的文库。组合构建体的文库可用于对表达所述组合构建体文库的宿主细胞进行筛选。在一些实施方案中,组合构建体的文库含有DNA元件或DNA元件与靶向表观遗传基因的CRISPR指导序列(例如表1所示的示例CRISPR指导序列)的组合。

[0085] 应理解,由于组合步骤在体外进行,所以该技术可扩展到可接受DNA的任何宿主细胞或生物体。在一些实施方案中,宿主细胞是细菌细胞。在一些实施方案中,生物体是细菌并且构建体在质粒或噬菌体上携带。在一些实施方案中,宿主细胞是酵母细胞。在另一些实施方案中,生物体是酵母,并且构建体在质粒或穿梭载体上携带。在另一些实施方案中,宿主细胞是哺乳动物细胞,例如人细胞。在这样的实施方案中,本文所述的遗传构建体可在质粒上携带或通过病毒(例如慢病毒或腺病毒)递送。

[0086] 本文中所述的遗传构建体和载体涉及CRISPR系统之组分的表达,所述组分包括CRISPR指导序列和支架序列。其中表达CRISPR系统的宿主细胞可表达一种或更多种另外的CRISPR组分,例如内切核酸酶。在一些实施方案中,宿主细胞还表达内切核酸酶,例如Cas内切核酸酶。在一些实施方案中,Cas内切核酸酶是Cas1、Cas2或Cas9内切核酸酶。在一些实施方案中,宿主细胞表达来源于化脓性链球菌、金黄色葡萄球菌、脑膜炎奈瑟氏菌、嗜热链球菌或齿垢密螺旋体的Cas9内切核酸酶。在一些实施方案中,编码Cas9内切核酸酶的核苷酸序列可以经密码子优化以用于在宿主细胞或生物体中表达。在一些实施方案中,内切核酸

酶是Cas9同源或直向同源的。

[0087] 在一些实施方案中,编码Cas9内切核酸酶的核苷酸序列被进一步修饰以改变蛋白质的活性。在一些实施方案中,Cas9内切核酸酶是无催化活性Cas9。例如,dCas9含有催化活性残基(D10和H840)的突变并且不具有核酸酶活性。作为替代或补充,Cas9内切核酸酶可与另一种蛋白质或其部分融合。在一些实施方案中,dCas9与阻遏物结构域融合,例如KRAB结构域。在一些实施方案中,这样的dCas9融合蛋白与本文所述的构建体一起用于多重基因抑制(例如CRISPR干扰(CRISPRi))。在一些实施方案中,dCas9与激活子结构域(例如VP64或VPR)融合。在一些实施方案中,这样的dCas9融合蛋白与本文所述的构建体一起用于多重基因激活(例如CRISPR激活(CRISPRa))。在一些实施方案中,dCas9与表观遗传调节结构域(例如组蛋白脱甲基酶结构域或组蛋白乙酰转移酶结构域)融合。在一些实施方案中,dCas9与LSD1或p300或其部分融合。在一些实施方案中,dCas9融合用于基于CRISPR的表观遗传调控。在一些实施方案中,dCas9或Cas9与FokI核酸酶结构域融合。在一些实施方案中,与FokI核酸酶结构域融合的Cas9或dCas9用于多重基因编辑。在一些实施方案中,Cas9或dCas9与荧光蛋白(例如GFP、RFP、mCherry等)融合。在一些实施方案中,与荧光蛋白融合的Cas9/dCas9蛋白用于基因组基因座的多重标记和/或可视化。

[0088] 作为替代或补充,内切核酸酶是Cpf1核酸酶。在一些实施方案中,宿主细胞表达来源于普氏菌属(*Provetella* spp.)或弗朗西斯菌属(*Francisella* spp.)的Cpf1核酸酶。在一些实施方案中,编码Cpf1核酸酶的核苷酸序列可以经密码子优化以用于在宿主细胞或生物中表达。

[0089] 本发明涵盖其中可引入DNA的任何细胞类型,包括原核细胞和真核细胞。在一些实施方案中,细胞是细菌细胞,例如埃希氏菌属(*Escherichia* spp.)、链霉菌属(*Streptomyces* spp.)、*Zymonas* spp、醋杆菌属(*Acetobacter* spp.)、柠檬酸杆菌属(*Citrobacter* spp.)、集胞藻属(*Synechocystis* spp.)、根瘤菌属(*Rhizobium* spp.)、梭菌属(*Clostridium* spp.)、棒状杆菌属(*Corynebacterium* spp.)、链球菌属(*Streptococcus* spp.)、黄单胞菌属(*Xanthomonas* spp.)、乳杆菌属(*Lactobacillus* spp.)、乳球菌属(*Lactococcus* spp.)、芽孢杆菌属(*Bacillus* spp.)、产碱菌属(*Alcaligenes* spp.)、假单胞菌属(*Pseudomonas* spp.)、气单胞菌属(*Aeromonas* spp.)、固氮菌属(*Azotobacter* spp.)、丛毛单胞菌属(*Comamonas* spp.)、分枝杆菌属(*Mycobacterium* spp.)、红球菌属(*Rhodococcus* spp.)、葡萄糖酸杆菌属(*Gluconobacter* spp.)、罗尔斯通氏菌属(*Ralstonia* spp.)、酸硫杆菌属(*Acidithiobacillus* spp.)、小月菌属(*Microbunatus* spp.)、地杆菌属(*Geobacter* spp.)、土芽孢杆菌属(*Geobacillus* spp.)、节杆菌属(*Arthrobacter* spp.)、黄杆菌属(*Flavobacterium* spp.)、沙雷氏菌属(*Serratia* spp.)、糖多胞菌属(*Saccharopolyspora* spp.)、栖热菌属(*Thermus* spp.)、寡养单胞菌种(*Stenotrophomonas* spp.)、色杆菌属(*Chromobacterium* spp.)、中华根瘤菌属(*Sinorhizobium* spp.)、糖多胞菌属(*Saccharopolyspora* spp.)、土壤杆菌属(*Agrobacterium* spp.)和泛菌属(*Pantoea* spp.)。细菌细胞可以是革兰阴性细胞例如大肠杆菌细胞或革兰氏阳性细胞,例如芽孢杆菌(*Bacillus*)属的物种。

[0090] 在另一些实施方案中,细胞是真菌细胞例如酵母细胞,例如酵母属(*Saccharomyces* spp.)、裂殖酵母属(*Schizosaccharomyces* spp.)、毕赤酵母属(*Pichia*

spp.)、法夫酵母属(*Phaffia* spp.)、克鲁维酵母属(*Kluyveromyces* spp.)、假丝酵母属(*Candida* spp.)、踝节菌属(*Talaromyces* spp.)、酒香酵母属(*Brettanomyces* spp.)、管囊酵母属(*Pachysolen* spp.)、德巴利酵母属(*Debaryomyces* spp.)、耶氏酵母属(*Yarrowia* spp.)和工业多倍体酵母菌株。优选地,酵母菌株是酿酒酵母菌株。真菌的另一些实例包括曲霉属(*Aspergillus* spp.)、青霉属(*Penicillium* spp.)、镰刀菌属(*Fusarium* spp.)、根霉属(*Rhizopus* spp.)、枝顶孢属(*Acremonium* spp.)、脉孔孢菌属(*Neurospora* spp.)、粪壳菌属(*Sordaria* spp.)、稻瘟菌属(*Magnaporthe* spp.)、异水霉属(*Allomyces* spp.)、黑粉菌属(*Ustilago* spp.)、葡萄孢属(*Botrytis* spp.)、木霉属(*Trichoderma* spp.)。

[0091] 在另一些实施方案中,细胞是藻类细胞、植物细胞、昆虫细胞、啮齿动物细胞或哺乳动物细胞,包括啮齿动物细胞或人细胞(例如,人胚肾细胞(例如,HEK293T细胞)、人皮肤成纤维细胞、人癌细胞例如OVCAR8细胞或OVCAR8-ADR细胞)。在一些实施方案中,细胞是人癌细胞,例如人卵巢癌细胞。

[0092] 本文还提供了包含靶向表观遗传基因之抑制剂的组合物。如本文使用的术语“表观遗传基因”是指影响细胞中另一种分子或过程的表现遗传调控的任何基因。在一些实施方案中,表观遗传基因编码参与表观遗传调控的蛋白质。在一些实施方案中,表观遗传基因编码影响表观遗传调控的核酸,例如RNA(例如,微小RNA)。一般而言,表观遗传是指对分子或过程的不涉及细胞基因组DNA突变的任何改变(Jaenisch和Bird,Nat.Gene.(2003) 33: 245-254)。表观遗传调控涉及细胞中的DNA介导的过程,例如转录、DNA修复和通过包括DNA甲基化、组蛋白修饰、核小体重塑和RNA介导靶向的机制进行的复制(Dawson和Kouzarides, Cell (2012) 150 (1):12-27)。表观遗传基因的非限制性实例包括:

[0093] DNMT1, DNMT3A, DNMT3B, DNMT3L, MBD1, MBD2, CREBBP, EP300, HDAC1, HDAC2, SIRT1, CARM1, EZH1, EZH2, MLL, MLL2, NSD1, PRMT1, PRMT2, PRMT3, PRMT5, PRMT6, PRMT7, SETD2, KDM1A-, KDM1B, KDM2A, KDM2B, KDM3A, KDM3B, KDM4A, KDM4B, KDM4C, KDM5A, KDM5B, KDM5C, KDM5D, KDM6A, KDM6B, PHF2, PHF8, BMI1, BRD1, BRD3, BRD4, ING1, ING2, ING3, ING4, 和ING5。

[0094] 如本文使用的术语“抑制剂”是指降低或阻止表观遗传基因表达或降低或阻止由表观遗传基因编码的蛋白质的活性的任何分子,例如蛋白质、核酸或小分子。在一些实施方案中,与不存在抑制剂组合的情况下表观遗传基因的表达相比,两种或更多种表观遗传基因抑制剂的组合使表观遗传基因的表达降低至少10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%或至少65%。在一些实施方案中,与不存在抑制剂组合的情况下由表观遗传基因编码的蛋白质的活性相比,两种或更多种表观遗传基因抑制剂的组合使由表观遗传基因编码的蛋白质的活性降低至少10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%或至少65%。

[0095] 在一些实施方案中,表观遗传基因抑制剂的组合包含至少2、3、4、5、6、7、8、9或至少10种或更多种表观遗传基因的抑制剂。在一些实施方案中,表观遗传基因抑制剂的组合包含两种表观遗传基因抑制剂。在一些实施方案中,抑制剂的组合抑制2、3、4、5、6、7、8、9、10或更多种表观遗传基因。在一些实施方案中,表观遗传基因抑制剂的组合包含靶向并抑制两种表观遗传基因的两种抑制剂。

[0096] 在一些实施方案中,抑制剂组合中的至少一种抑制剂是直接或间接降低或阻止表观遗传基因表达或降低或阻止由表观遗传基因编码的蛋白质的活性的蛋白质。例如,蛋白

质可以是降低或阻止表观遗传基因表达的阻遏物或由表观遗传基因编码之蛋白质的变构抑制剂。在一些实施方案中,两种或更多种抑制剂是靶向两种或更多种表观遗传基因的蛋白质。在一些实施方案中,两种或更多种抑制剂是靶向表2中所示表观遗传基因的任何组合的蛋白质。

[0097] 在一些实施方案中,抑制剂组合中的至少一种抑制剂是降低或阻止表观遗传基因表达或降低或阻止由表观遗传基因编码的蛋白质之活性的核酸。在一些实施方案中,核酸是CRISPR指导序列,其与支架序列一起将内切核酸酶募集到表观遗传基因。在一些实施方案中,两种或更多种抑制剂是靶向两种或更多种表观遗传基因的CRISPR指导序列。在一些实施方案中,两种或更多种抑制剂是靶向表2中所示表观遗传基因的任何组合的CRISPR指导序列。在一些实施方案中,两种或更多种抑制剂是选自靶向表1中提供的表观遗传基因之示例CRISPR指导序列的CRISPR指导序列。在一些实施方案中,表观遗传基因的组合包含BRD4和KDM4C。在一些实施方案中,表观遗传基因的组合包含BRD4和KDM6B。

[0098] 在一些实施方案中,核酸是通过细胞的RNA干扰(RNAi)途径加工以沉默靶基因的表达(例如,降低mRNA水平和/或蛋白质产生)的shRNA。在一些实施方案中,两种或更多种抑制剂是靶向两种或更多种表观遗传基因的shRNA。在一些实施方案中,两种或更多种抑制剂是靶向表2中所示的表观遗传基因之任何组合的shRNA。在一些实施方案中,两种或更多种抑制剂是选自靶向表4中提供的表观遗传基因之示例shRNA的shRNA。一些实施方案中,表观遗传基因的组合包含BRD4和KDM4C。在一些实施方案中,表观遗传基因的组合包含BRD4和KDM6B。

[0099] 在一些实施方案中,抑制剂组合中的至少一种抑制剂是降低或阻止表观遗传基因表达或降低或阻止由表观遗传基因编码的蛋白质的活性的小分子。在一些实施方案中,两种或更多种抑制剂是靶向两种或更多种表观遗传基因的小分子。在一些实施方案中,两种或更多种抑制剂是靶向表2中所示的表观遗传基因之任何组合的小分子。在一些实施方案中,表观遗传基因的组合包含BRD4和KDM4C。在一些实施方案中,表观遗传学基因的组合包含BRD4和KDM6B。

[0100] 降低或阻止BRD4的表达或降低或阻止由BRD4编码的蛋白质的活性的任何小分子均可能适用于本文所述的组合物和方法。BRD4抑制剂的实例包括但不限于:JQ1((6S)-4-(4-氯苯基)-2,3,9-三甲基-6H-噻吩并[3,2-f][1,2,4]三唑并[4,3-a][1,4]二氮杂~~草~~-6-乙酸1,1-二甲基乙酯)、MS417([(6S)-4-(4-氯苯基)-2,3,9-三甲基-6H-噻吩并[3,2-f][1,2,4]三唑并[4,3-a][1,4]二氮杂~~草~~-6-基]乙酸甲酯),或者RVX-208(2-[4-(羟乙氧基)-3,5-二甲基苯基]-5,7-二甲氧基-4(3H)-喹唑啉酮)。在一些实施方案中,BRD4抑制剂是JQ1。另外的BRD4抑制剂对于本领域技术人员将是明显的,并且可见于例如PCT公布WO 2014/154760A1和Vidler等J.Med.Chem. (2013) 56:8073-8088中。

[0101] 降低或阻止KDM4C(也称为JMJD2)的表达或者降低或阻止由KDM4C编码的蛋白质之活性的任何小分子均可适用于本文所述的组合物和方法。在一些实施方案中,KDM4C抑制剂是SD70(N-(呋喃-2-基(8-羟基喹啉-7-基)甲基)异丁酰胺)或咖啡酸。另外的KDM4C抑制剂对于本领域技术人员将是明显的,并且可见于例如Leurs等,Bioorg.&Med.Chem.Lett. (2012) 22(12):5811-5813和Hamada等,Bioorg.&Med.Chem.Lett. (2009) 19:2852-2855中。

[0102] 降低或阻止KDM6B(也称为JMJD3)的表达或者降低或阻止由KDM6B编码的蛋白质之

活性的任何小分子均可适用于本文所述的组合物和方法。KDM6B抑制剂的实例包括但不限于：GSK-J4 (3-((6-(4,5-二氢-1H-苯并[d]氮杂~~草~~-3(2H)-基)-2-(吡啶-2-基)嘧啶-4-基)氨基)丙酸乙酯,单盐酸盐)、GSK-J1 (N-[2-(2-吡啶基)-6-(1,2,4,5-四氢-3H-3-苯并氮杂~~草~~-3-基)-4-嘧啶基]-β-丙氨酸)和IOX1 (8-羟基-5-喹啉羧酸;8-羟基-5-喹啉羧酸)。在一些实施方案中,KDM6B抑制剂是GSK-J4。另外的KDM4C抑制剂对于本领域技术人员是明显的。

[0103] 两种或更多种抑制剂的组合可包含两种或更多种蛋白质抑制剂、两种或更多种CRISPR指导序列、两种或更多种shRNA或两种或更多种小分子抑制剂。在一些实施方案中,两种或更多种抑制剂是不同类型的抑制剂(例如蛋白质、核酸、小分子)。在一些实施方案中,所述组合包含蛋白质抑制剂和一种或更多种另外的抑制剂(例如,CRISPR指导序列、shRNA和/或小分子抑制剂)。在另一些实施方案中,所述组合包含CRISPR指导序列和一种或更多种另外的抑制剂(例如蛋白质、shRNA和/或小分子抑制剂)。在另一些实施方案中,所述组合包含shRNA和一种或更多种另外的抑制剂(例如,蛋白质、CRISPR指导序列和/或小分子抑制剂)。在另一些实施方案中,所述组合包含小分子抑制剂和一种或更多种另外的抑制剂(例如蛋白质、shRNA和/或shRNA)。

[0104] 本文所述的方法和组合物可用于降低细胞的增殖,所述细胞例如癌细胞或需要降低增殖的其他细胞。在一些实施方案中,使细胞与表观遗传基因的两种或更多种抑制剂的组合(例如,靶向表2中所示的表观遗传基因之抑制剂的组合)接触部分地或完全地降低细胞的增殖。在一些实施方案中,与未与抑制剂组合接触的细胞相比,将细胞与表观遗传基因的两种或更多种抑制剂的组合接触部分地或完全地降低细胞的增殖。在一些实施方案中,与未与抑制剂组合接触的细胞相比,将细胞与两种或更多种表观遗传基因抑制剂的组合接触将细胞的增殖降低至少10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%或至少65%。在一些实施方案中,和与抑制剂组合接触的非癌细胞相比,使细胞与表观遗传基因的两种或更多种抑制剂的组合接触部分地或完全地降低癌细胞的增殖。可通过本领域已知的任何方法(例如使用细胞生存力测定、MTT测定或BrdU细胞增殖测定)来评估和定量细胞增殖。

[0105] 本发明的另一些方面涉及用于治疗对象之癌症的方法和组合物。癌症是以不受控制或控制异常的细胞增殖和其他恶性细胞特性为特征的疾病。如本文使用的术语“癌症”是指本领域已知的任何类型的癌症,包括但不限于:乳腺癌、胆道癌、膀胱癌、脑癌、宫颈癌、绒毛膜癌、结肠癌、子宫内膜癌、食管癌、胃癌、血液肿瘤、T细胞急性淋巴细胞白血病/淋巴瘤、毛细胞白血病、慢性髓细胞性白血病、多发性骨髓瘤、艾滋病相关白血病和成人T细胞白血病/淋巴瘤、上皮内肿瘤、肝癌、肺癌、淋巴瘤、神经母细胞瘤、口腔癌、卵巢癌、胰腺癌、前列腺癌、直肠癌、肉瘤、皮肤癌、睾丸癌、甲状腺癌和肾癌。癌细胞可以是在体内(即,在生物体内)、离体(即,从生物体中取出并在体外保持)或在体外的癌细胞。

[0106] 所述方法包括向对象以有效量施用两种或更多种表观遗传基因抑制剂的组合。在一些实施方案中,所述对象是患有癌症、怀疑患有癌症或有发生癌症风险的对象。在一些实施方案中,所述对象是哺乳动物对象,包括但不限于狗、猫、马、牛、猪、绵羊、山羊、鸡、啮齿动物或灵长类动物。在一些实施方案中,所述对象是人对象,例如患者。人对象可能是小儿(pediatric)对象或成人对象。对象是否被视为有患癌症的“风险”可由技术人员确定。

[0107] 如本文使用的“治疗”包括改善、治愈、防止其变差,减缓发展速度或防止疾病再次

发生(即,为了防止复发)。组合物的有效量是产生治疗效果的组合物的量。例如,在用于治疗对象之癌症的方法中,靶向表观遗传基因之抑制剂组合的有效量是提供抗癌效果的任何量,例如降低或防止癌细胞增殖或对癌细胞具有细胞毒性。在一些实施方案中,与当不存在一种或更多种靶向表观遗传基因的另外抑制剂的情况下施用的抑制剂有效量相比,将一种抑制剂与一种或更多种靶向表观遗传基因的另外的抑制剂同时或组合施用,靶向表观遗传基因的抑制剂的有效量降低。在一些实施方案中,当不存在一种或更多种靶向表观遗传基因的另外抑制剂的情况下施用,靶向表观遗传基因的抑制剂不会降低或阻止癌细胞的增殖。

[0108] 靶向表观遗传基因的抑制剂或靶向表观遗传基因的抑制剂的组合(例如,表2中所示的表观遗传基因的组合)可使用本领域已知的任何方法向对象施用。在一些实施方案中,通过局部、肠内或肠胃外施用途径施用抑制剂。在一些实施方案中,通过静脉内、肌内或皮下施用抑制剂。

[0109] 本文所述的表观遗传基因的任何抑制剂可通过本领域已知的任何方法施用至对象,或递送至细胞或与细胞接触。在一些实施方案中,表观遗传基因的抑制剂通过纳米颗粒、细胞渗透肽、聚合物、脂质体或重组表达载体递送至细胞。在另一些实施方案中,表观遗传基因的抑制剂与一种或更多种纳米颗粒、细胞渗透肽和/或聚合物缀合。在另一些实施方案中,表观遗传基因的抑制剂包含在脂质体内。

[0110] 还提供了用于鉴定降低或防止细胞或细胞群体增殖的表观遗传基因抑制剂之组合的方法。如图4A所示,该方法涉及将两个细胞群体与靶向表观遗传基因的CRISPR指导序列和支架序列(例如,条形码化CRISPR文库)以及Cas9内切核酸酶的组合文库接触。将两个细胞群体培养不同的持续时间。例如,一个细胞群体可培养15天,而另一个细胞群体培养20天。(例如通过测序方法)确定每个细胞群体之两个或更多个CRISPR指导序列和支架序列的组合的身份。例如,可通过对作为CRISPR指导序列的独特识别物的条形码进行测序来鉴定CRISPR指导序列和支架序列。将培养较长持续时间的细胞群体中CRISPR指导序列和支架序列的每种组合的丰度与培养较短时间的细胞群体中CRISPR指导序列和支架序列的每种组合的丰度进行比较。与培养较短时间的细胞群体中CRISPR指导序列的丰度相比,在培养较长持续时间的细胞群体中降低细胞增殖的CRISPR指导序列和支架序列的组合的丰度较低。这些组合被鉴定为降低细胞增殖的表观遗传基因的抑制剂的组合。

[0111] 提供了其他方法来鉴定这样的表观遗传基因的组合,当所述组合被抑制时可降低或防止细胞或细胞群体增殖。如图4A所示,该方法涉及将两个细胞群体与靶向表观遗传基因的CRISPR指导序列和支架序列(例如,条形码化的CRISPR文库)的组合文库以及Cas9内切核酸酶接触。将两个细胞群体培养不同的持续时间。例如,一个细胞群体可培养15天,而另一个细胞群体培养20天。(例如通过测序方法)确定每个细胞群体之两个或更多个CRISPR指导序列和支架序列之组合的身份。例如,可通过对作为CRISPR指导序列的独特识别物的条形码进行测序来鉴定CRISPR指导序列和支架序列。将培养较长持续时间的细胞群体中的CRISPR指导序列和支架序列的每种组合的丰度与培养较短时间的细胞群体中CRISPR指导序列和支架序列的每种组合的丰度进行比较。在培养较长持续时间的细胞群体中降低细胞增殖的CRISPR指导序列和支架序列之组合的丰度较低。这些组合被鉴定为可能被抑制剂靶向以降低或阻止细胞增殖的表观遗传基因的组合。

[0112] 在一些实施方案中,在重组表达载体中表达靶向与本发明相关的表观遗传基因的一种或更多种基因或抑制剂。如本文使用的“载体”可以是这样的多种核酸中的任一种,通过限制和连接(例如,使用CombiGEM方法)或通过重组可向其插入所期望序列,以用于在不同遗传环境之间转运或在宿主细胞(例如,癌细胞)中表达。载体通常由DNA组成,尽管也可用RNA载体。载体包括但不限于:质粒、福斯黏粒(fosmid)、噬菌粒、病毒基因组和人造染色体。在一些实施方案中,载体是慢病毒载体。在一些实施方案中,靶向表观遗传基因的两种或更多种基因或抑制剂在同一重组表达载体上表达。在一些实施方案中,靶向表观遗传基因的两种或更多种基因或抑制剂在两种或更多种重组表达载体上表达。

[0113] 克隆载体是能够在宿主细胞中自主复制或整合在基因组中的那些,其特征还在于具有一个或更多个内切核酸酶限制性位点,在所述位点处载体可以以可确定的方式被切割并将所期望的DNA序列连接到其中,或者具有重组位点,在所述位点处具有相容性末端的插入物可被整合以使得新的重组载体保留其在宿主细胞中复制的能力。在质粒的情况下,随着宿主细胞(例如宿主细菌)内质粒的拷贝数提高,所期望序列的复制可能发生多次,或者在宿主通过有丝分裂繁殖之前,在每个宿主中仅发生一次。在噬菌体的情况下,复制可以在裂解阶段期间主动发生或者在溶原相(lysogenic phase)期间被动地发生。

[0114] 表达载体是可以通过限制和连接或重组将所期望的DNA序列插入其中,使得其有效连接至调控序列并且可被表达为RNA转录本的表达载体。载体还可包含一种或更多种标志物序列,其适用于鉴定已经或尚未用载体转化或转染的细胞。标志物包括例如编码提高或降低对抗生素或其他化合物的抗性 or 敏感性之蛋白质的基因、编码其活性可通过本领域已知的标准测定法检测之酶的基因(例如 β -半乳糖苷酶、萤光素酶或碱性磷酸酶),以及可视地影响转化或转染的细胞、宿主、集落或斑块之表型(例如,绿色荧光蛋白、红色荧光蛋白)的基因。优选的载体是能够自主复制并表达存在于与其有效连接的DNA区段中之结构基因产物的那些载体。

[0115] 如本文使用的,当编码序列和调控序列以这样方式共价连接,以使得编码序列在调控序列的影响或控制下表达或转录时,编码序列和调控序列被称为“有效地”连接。如果期望将编码序列翻译成功能性蛋白质,则如果诱导5'调控序列中的启动子引起编码序列的转录,并且如果两个DNA序列之间连接的性质没有(1)导致引入移码突变,(2)干扰启动子区域指导编码序列转录的能力,或(3)干扰将相应的RNA转录本翻译成蛋白质的能力,则两个DNA序列被称为有效地连接。因此,如果启动子区域能够影响该DNA序列的转录,使得所得到的转录本可翻译成所期望的蛋白质或多肽,则启动子区域被有效地连接到编码序列。

[0116] 当核酸分子在细胞中表达时,多种转录调控序列(例如,启动子/增强子序列)可用于指导其表达。启动子可以是天然启动子,即在其内源环境中的基因启动子,其提供基因表达的正常调控。在一些实施方案中,启动子可以是组成型的,即,启动子不受控制,这允许其相关基因连续转录。还可使用多种条件启动子,例如通过分子存在或不存在控制的启动子。在一些实施方案中,启动子是RNA聚合酶II启动子,例如哺乳动物RNA聚合酶II启动子。在一些实施方案中,启动子是人泛素C启动子(UBCp)。在一些实施方案中,启动子是病毒启动子。在一些实施方案中,启动子是人巨细胞病毒启动子(CMVp)。在一些实施方案中,启动子是RNA聚合酶III启动子。RNA聚合酶III启动子的实例包括但不限于H1启动子、U6启动子、小鼠U6启动子、猪U6启动子。在一些实施方案中,启动子是U6启动子(U6p)。

[0117] 基因表达所需的调控序列的精确性质可在物种或细胞类型之间变化,但一般(必要时)应包括分别涉及转录和翻译起始的5' 非转录和5' 非翻译序列,例如TATA框、加帽序列、CAAT序列等。特别地,这样的5' 非转录调控序列将包括启动子区,所述启动子区包括用于对有效连接的基因进行转录控制的启动子序列。根据需要,调控序列还可包括增强子序列或上游激活子序列。本发明的载体可任选地包括5' 前导或信号序列。适当载体的选择和设计在本领域普通技术人员的能力和判断之内。

[0118] 包含所有表达必需元件的表达载体是市售的并且是本领域技术人员已知的。参见例如Sambrook等,Molecular Cloning:A Laboratory Manual,第4版,Cold Spring Harbor Laboratory Press,2012。通过向细胞引入异源DNA (RNA) 对细胞进行遗传改造。将异源DNA (RNA) 置于转录因子的有效控制下以使得在宿主细胞中表达异源DNA。

[0119] 可使用本领域标准的方法和技术将与本发明相关的核酸分子引入细胞。例如,可通过标准方案(例如转化,包括化学转化和电穿孔、病毒转导、粒子轰击等)引入核酸分子。在一些实施方案中,使用慢病毒来实现病毒转导。也可通过将核酸分子整合在基因组中来实现核酸分子表达。

[0120] 本发明并不限于其在以下说明书中阐述或在附图中示出的结构细节和组件布置。本发明能够具有另一些实施方案并且能够以多种方式被实施或执行。此外,本文使用的措辞和术语是出于描述的目的,不应被视为限制。本文使用“包括”、“包含”或“具有”、“含有”、“涉及”及其变体意在包括其后列出的项目及其等同物以及另外的项目。

[0121] 通过以下实施例进一步说明本发明,其决不应被解释为进一步的限制。在整个申请中引用的所有参考文献(包括参考文献、已授予的专利、公开的专利申请和共同未决的专利申请)的全部内容通过引用明确并入本文,特别是对于本文所引用的教导。

实施例

[0122] 实施例1

[0123] 如图1所示,可使用本文所述的方法制备条形码化CRISPR分子的高复杂度组合文库。在步骤1中,使用基于阵列的寡核苷酸合成(包括针对每个CRISPR指导序列的正向和反向寡核苷酸(例如寡核苷酸F-A和寡核苷酸R-A;寡核苷酸F-B和寡核苷酸R-B))来产生大量的CRISPR指导RNA (gRNA) 文库。合成的寡核苷酸的长度与终产物文库的复杂度无关。然后在步骤2中,分离出gRNA的寡核苷酸对并使其退火。对应每个gRNA的寡核苷酸包含20碱基对的CRISPR指导序列、两个BbsI位点和条形码元件。寡核苷酸还可含有5' 和3' 单链突出末端区域以用于将寡核苷酸连接到储存载体中。在步骤3中,将含有BbsI和MfeI突出末端的经退火的寡聚体合并在一起进行一锅连接反应,以将gRNA文库插入已用BbsI和MfeI消化的AWp28储存载体中。这使得产生AWP28储存载体的文库,每个载体包含CRISPR指导序列、两个BbsI位点和条形码元件。

[0124] 然后在步骤4中,使用BbsI对合并的储存载体文库进行单锅消化(single-pot digestion)以在CRISPR指导序列和条形码元件之间打开载体,从而允许在CRISPR指导序列和条形码元件之间插入gRNA支架元件。由于gRNA支架元件的插入物在支架元件的3' 末端含有由限制性识别位点BamHI和EcoRI形成的隔离位点,所以gRNA储存载体的文库可进行迭代克隆步骤以逐步产生更复杂的n向条形码化文库,其包含可使用Cas9核酸酶或效应物来介

导组合基因敲除、活化或抑制的多个gRNA表达盒,如步骤5所示。简言之,使用与限制性识别位点互补的限制性酶(例如,BamHI和EcoRI)在隔离位点处消化条形码化指导RNA文库。消化允许插入编码CRISPR指导序列、支架序列、隔离位点和条形码元件的另外的片段。因为相同的限制性酶组可用于多轮克隆以构建(n)-向文库,所以该策略的优点之一是(n+1)-向文库不提高所需的限制性酶组。

[0125] 为了构建gRNA成员的(n)-向文库,需要(n+2)轮克隆步骤。该策略的另外一个优点是可使用相同的CRISPR指导库来生成更高阶的复杂文库。

[0126] 实施例2:

[0127] 如图2A所示,可使用本文所述的方法制备条形码化CRISPR分子的复杂组合文库。在步骤1中,在单个寡核苷酸上合成多个CRISPR指导序列和单个条形码元件以产生大量的组合gRNA文库。在每个CRISPR指导序列后都存在限制性识别位点。指导序列和条形码通过不同的限制性酶位点连接在一起。图2A的遗传构建体和载体示出了包含三个CRISPR指导序列和位于CRISPR指导序列下游的条形码元件的示例性寡核苷酸。

[0128] 在步骤2中,将合并的合成的寡核苷酸在单锅组装(single-pot assembly)中连接到目的载体中。目的载体可包含用于驱动至少一个CRISPR指导序列之表达的启动子。如步骤3所示,用不同限制性酶在每个CRISPR指导序列之后的每个限制性识别位点处顺序消化载体,这允许插入支架元件,并且在一些情况下,允许插入驱动下游CRISPR指导序列表达的启动子元件。该方法导致产生编码多个CRISPR指导序列和支架序列以及单个条形码元件的条形码化组合指导RNA文库,以用于使用Cas9核酸酶或效应物进行组合基因敲除、活化或抑制。

[0129] 图2B描绘了产生条形码化CRISPR分子的复杂组合文库的替代策略。在步骤1中,在单个寡核苷酸上合成多个CRISPR指导序列和单个条形码元件以产生大量的组合gRNA文库。在每个CRISPR指导序列之后存在限制性识别位点并且在第一CRISPR指导序列上游存在限制性识别位点以用于插入启动子元件。指导序列和条形码通过不同的限制性酶位点连接在一起。图2B的遗传构建体和载体示出了包含三个CRISPR指导序列和位于CRISPR指导序列上游的条形码元件的示例性寡核苷酸。

[0130] 在步骤2中,在单锅组装中将合并的合成寡核苷酸连接到目的载体中。目的载体可含有用于驱动至少一个CRISPR指导序列的表达的启动子。如步骤3所示,用不同的限制性酶在每个CRISPR指导序列之后的每个限制性识别位点处顺序消化载体,这允许插入支架元件,并且在一些情况下,允许插入驱动下游CRISPR指导序列表达的启动子元件。该方法导致产生编码多个CRISPR指导序列和支架序列以及单个条形码元件的条形码化组合指导RNA文库,以用于使用Cas9核酸酶或效应物进行组合基因敲除、激活或抑制。

[0131] 使用图2A和2B中所描绘的策略,可用(n+1)轮克隆步骤构建(m)个gRNA成员的(n)向-文库。所产生文库的复杂度取决于步骤1中合成的寡核苷酸的长度。因为对于每个CRISPR指导序列以及启动子元件允许插入支架序列的限制性识别位点是不同的,所以提高寡核苷酸的复杂度(即,CRISPR指导序列的数目)也提高了消化步骤所需的限制性酶的数量。

[0132] 实施例3:

[0133] 使用基于CombiGEM的DNA组装方法进行条形码化组合gRNA文库的高效且可扩展的

组装。通过慢病毒将文库递送到人细胞中以便创建遗传超多样化 (ultra-diverse) 细胞群体,其具有可通过合并测定中的条形码测序来跟踪的独特的gRNA组合。该策略(称为CombiGEM-CRISPR)使用简单的一锅克隆步骤来实现高阶组合gRNA文库的可扩展组装,从而简化并加速了组合基因功能的系统分析工作流程。

[0134] 为了创建初始的条形码化sgRNA文库,首先合成编码条形码化gRNA靶序列文库的寡核苷酸对、使其退火,并以相等的比例合并以用于克隆储存载体中U6启动子的下游(图1)。随后,将在单锅连接反应中将gRNA的支架序列插入储存载体文库中。CombiGEM方法应用于高阶组合gRNA文库的可扩展组装(图1)。在条形码化sgRNA构建体中,BamHI和EcoRI位点位于gRNA序列及其条形码之间,而BglIII和MfeI位点位于末端。这些限制性酶位点的策略定位导致条形码与其gRNA序列在酶消化后分离,并且导致代表插入物连接后其各自gRNA的条形码的串联。为了构建单向文库,通过用BglIII和MfeI限制性消化储存载体来制备条形码化sgRNA表达单元的合并插入物,并将其与用BamHI和EcoRI消化的慢病毒目的载体中的相容性末端连接。然后将单向文库充当目的载体用于下一轮条形码化sgRNA表达单元的合并插入,从而产生双向文库,其中表示每个sgRNA的条形码定位于每个慢病毒构建体的一端。可迭代地重复该过程以产生更高阶的条形码化组合gRNA文库。可通过检测串联的条形码来跟踪组合gRNA的身份,其对于每个组合是独特的(例如,通过高通量测序)。

[0135] 为了评估我们的慢病毒组合gRNA表达系统的功能,构建了靶向编码绿色荧光蛋白(GFP)和红色荧光蛋白(RFP)序列的gRNA组合(表1)。通过使用流式细胞术(图6A和6B)和荧光显微术(图6E)测定组合基因扰动表型。将携带双重RFP和GFP报道子的慢病毒与条形码化组合gRNA表达单元一起用于感染稳定表达人密码子优化的Cas9核酸酶(OVCAR8-ADR-Cas9)的人卵巢癌细胞(OVCAR8-ADR)(Honma等,Nat.Med.(2008)14:939-948)(图6A)。预期活性gRNA将靶向编码GFP和RFP的序列,并产生插失以敲除GFP和RFP的表达。观察到GFP和RFP荧光水平的有效抑制,因为在感染后第4天和第8天携带Cas9核酸酶和靶向RFP和GFP二者的gRNA表达单元的细胞中,GFP和RFP双阴性群体是观察到的主要群体(总群体的~83%至97%),相比而言载体对照为<0.7%(图6C和6D)。在表达靶向GFP和/或RFP的gRNA但无Cas9核酸酶的对照细胞系中没有观察到这种抑制(图6B)。证实了基因扰动的特异性,因为仅携带GFP靶向性sgRNA的细胞表现出GFP信号而非RFP信号的损失。类似地,含有RFP靶向性sgRNA的细胞表现出RFP表达的降低,但对GFP表达没有作用(图6E)。这些结果证明了慢病毒载体编码可在单个人细胞内同时抑制多个基因表达的组合gRNA构建体的能力。

[0136] 不同的表观遗传修饰倾向于协同作用以调控基因表达模式(Wang等,Nat.Genet.(2008)40:897-903),并且组合表观遗传调控正在成为有效癌症治疗的有希望的策略(Dawson等,Cell(2012)150:12-27;Juergens等,Cancer Discov.(2011)1:598-607)。使用本文所述的CombiGEM-CRISPR方法和组合物,系统地评估了表观遗传基因扰动对抗癌表型的组合作用。构建了一个文库,其包含靶向一组50个表观遗传基因(每个基因3个sgRNA)的153个条形码化sgRNA和基于GeCK0v2文库的3个对照sgRNA。(Shalem等,Science(2014)343:84-87)(表1)。

[0137] 表1:sgRNA靶序列

[0138]

sgRNA ID	sgRNA 靶序列
GFP-sg1	GGGCGAGGAGCTGTTACCG (SEQ ID NO: 7)
RFP-sg1	CACCCAGACCATGAAGATCA (SEQ ID NO: 8)
RFP-sg2	CCACTTCAAGTGCACATCCG (SEQ ID NO: 9)
Control-sg1	ATCGTTTCCGCTTAACGGCG (SEQ ID NO: 10)
Control-sg2	AAACGGTACGACAGCGTGTG (SEQ ID NO: 11)
Control-sg3	CCATCACCGATCGTGAGCCT (SEQ ID NO: 12)
DNMT1-sg1	CTAGACGTCCATTCACTTCC (SEQ ID NO: 13)
DNMT1-sg2	TTTCCAAACCTCGCACGCCC (SEQ ID NO: 14)
DNMT1-sg3	ACGTAAAGAAGAATTATCCG (SEQ ID NO: 15)
DNMT3A-sg1	CCGCCCCACCTTCCGTGCCG (SEQ ID NO: 16)
DNMT3A-sg2	TGGCGCTCCTCCTTGCCACG (SEQ ID NO: 17)
DNMT3A-sg3	CCGCTCCGCAGCAGAGCTGC (SEQ ID NO: 18)
DNMT3B-sg1	AGAGTCGCGAGCTTGATCTT (SEQ ID NO: 19)
DNMT3B-sg2	ATCCGCACCCCGGAGATCAG (SEQ ID NO: 20)
DNMT3B-sg3	GAAGACTCGATCCTCGTCAA (SEQ ID NO: 21)
DNMT3L-sg1	AGGGATCTGCGCCCCATGTA (SEQ ID NO: 22)
DNMT3L-sg2	ACTCACCTTCTATATTTCGC (SEQ ID NO: 23)
DNMT3L-sg3	CACCAAAATCACGTCCATGC (SEQ ID NO: 24)
MBD1-sg1	TCACCCGTAGGCAACGTCGC (SEQ ID NO: 25)
MBD1-sg2	GTGTCCAGCGACGTTGCCTA (SEQ ID NO: 26)
MBD1-sg3	ACGTTGTGCAAAGACTGTCTG (SEQ ID NO: 27)
MBD2-sg1	CGGCGACTCCGCCATAGAGC (SEQ ID NO: 28)
MBD2-sg2	GGAGCCGGTCCCTTTCCCGT (SEQ ID NO: 29)
MBD2-sg3	AGTCTTGAAAGCGCATGCCA (SEQ ID NO: 30)
CREBBP-sg1	AGCGGCTCTAGTATCAACCC (SEQ ID NO: 31)
CREBBP-sg2	GAATCACATGACGCATTGTC (SEQ ID NO: 32)
CREBBP-sg3	CCCGCAAATGACTGGTCACG (SEQ ID NO: 33)
EP300-sg1	CTGTCAGAATTGCTGCGATC (SEQ ID NO: 34)
EP300-sg2	CTTGGCAAGACTTGCTGAC (SEQ ID NO: 35)
EP300-sg3	TAGTTCCCCTAACCTCAATA (SEQ ID NO: 36)
HDAC1-sg1	ACACCATTCGTAACGTTGCC (SEQ ID NO: 37)
HDAC1-sg2	TCACTCGAGATGCGCTTGTC (SEQ ID NO: 38)
HDAC1-sg3	AGAATGCTGCCGCACGCACC (SEQ ID NO: 39)
HDAC2-sg1	TCCGTAATGTTGCTCGATGT (SEQ ID NO: 40)

[0139]

HDAC2-sg2	TCCAACATCGAGCAACATTA (SEQ ID NO: 41)
HDAC2-sg3	TACAACAGATCGTGTAATGA (SEQ ID NO: 42)
SIRT1-sg1	GTTGACTGTGAAGCTGTACG (SEQ ID NO: 43)
SIRT1-sg2	AACAGGTTGCGGGAATCCAA (SEQ ID NO: 44)
SIRT1-sg3	TACCCAGAACATAGACACGC (SEQ ID NO: 45)
CARM1-sg1	CTCGCCGTTTCGCGTCGCCGA (SEQ ID NO: 46)
CARM1-sg2	CCCGTACTCACGGCTGTAGA (SEQ ID NO: 47)
CARM1-sg3	GGGCCACGTACCGTTGGGTG (SEQ ID NO: 48)
EZH1-sg1	ACAGGCTTCATTGACTGAAC (SEQ ID NO: 49)
EZH1-sg2	AGCTGATCAATAACTATGAT (SEQ ID NO: 50)
EZH1-sg3	CCTCATCTGAGTACTGATTC (SEQ ID NO: 51)
EZH2-sg1	ACACGCTTCCGCCAACAAC (SEQ ID NO: 52)
EZH2-sg2	TGCGACTGAGACAGCTCAAG (SEQ ID NO: 53)
EZH2-sg3	AAAACCTTCATCTCCCATATA (SEQ ID NO: 54)
MLL-sg1	GTACAAATTGTACGACGGAG (SEQ ID NO: 55)
MLL-sg2	GACCCCTCGGCGGTTTATAG (SEQ ID NO: 56)
MLL-sg3	TATATTGCGACCACCAAAC (SEQ ID NO: 57)
MLL2-sg1	CAGAGAGCACACGCCGCAC (SEQ ID NO: 58)
MLL2-sg2	GGAACCGCTGGCAGTCGCGC (SEQ ID NO: 59)
MLL2-sg3	CTCCCGCTGCCCCGTGTAGAC (SEQ ID NO: 60)
NSD1-sg1	CTGGCTCGAGATTTAGCGCA (SEQ ID NO: 61)
NSD1-sg2	AATCTGTTTCATGCGCTTACG (SEQ ID NO: 62)
NSD1-sg3	GATTCCAGTACCAGTACATT (SEQ ID NO: 63)
PRMT1-sg1	CTCACCGTGGTCTAACTTGT (SEQ ID NO: 64)
PRMT1-sg2	GGATGTCATGTCCTCAGCGT (SEQ ID NO: 65)
PRMT1-sg3	TTTGACTCCTACGCACACTT (SEQ ID NO: 66)
PRMT2-sg1	CGTGATGAGTACGACCCCG (SEQ ID NO: 67)
PRMT2-sg2	TCTTCTGTGCACACTATGCG (SEQ ID NO: 68)
PRMT2-sg3	CTGTCCCAGAAGTGAATCGC (SEQ ID NO: 69)
PRMT3-sg1	GCCATGTGCTCGTTAGCGTC (SEQ ID NO: 70)
PRMT3-sg2	GCCTGACGCTAACGAGCACA (SEQ ID NO: 71)
PRMT3-sg3	GAATTCATGTACTCAACTGT (SEQ ID NO: 72)
PRMT5-sg1	CGGAATGCGGGGTCCGAAC (SEQ ID NO: 73)
PRMT5-sg2	CAGCATACAGCTTTATCCGC (SEQ ID NO: 74)
PRMT5-sg3	ATGAACTCCCTCTTGAAACG (SEQ ID NO: 75)
PRMT6-sg1	ATTGTCCGGCGAGGACGTGC (SEQ ID NO: 76)
PRMT6-sg2	CTTCGCCACGCGCTGTCTCA (SEQ ID NO: 77)
PRMT6-sg3	GACGGTACTGGACGTGGGCG (SEQ ID NO: 78)
PRMT7-sg1	CAATCCGACCACGGGGTCTG (SEQ ID NO: 79)
PRMT7-sg2	GAGGTTCAAACCGCCTGCTA (SEQ ID NO: 80)
PRMT7-sg3	TAAAGTCGGCTGGTGACACC (SEQ ID NO: 81)
SETD2-sg1	AGTTCTTCTCGGTGTCCAAA (SEQ ID NO: 82)
SETD2-sg2	GACTATCAGTTCCAGAGATA (SEQ ID NO: 83)
SETD2-sg3	AACTTACGAAGGAAGGTCTT (SEQ ID NO: 84)
KDM1A-sg1	TTACCTTCGCCCCTTGCGC (SEQ ID NO: 85)

[0140]

KDM1A-sg2	CCGGCCCTACTGTCGTGCCT (SEQ ID NO: 86)
KDM1A-sg3	AGAGCCGACTTCCTCATGAC (SEQ ID NO: 87)
KDM1B-sg1	CATACCGCATCGATAAGTCT (SEQ ID NO: 88)
KDM1B-sg2	ATAGCCAAGACTTATCGATG (SEQ ID NO: 89)
KDM1B-sg3	GAACATACCTTCTGTAGTAA (SEQ ID NO: 90)
KDM2A-sg1	ACGCTACTATGAGACCCCAG (SEQ ID NO: 91)
KDM2A-sg2	TATGGCAGGGAGTCGTCGCA (SEQ ID NO: 92)
KDM2A-sg3	GTAACGAATCCTTTCTTCTT (SEQ ID NO: 93)
KDM2B-sg1	CCTCGTTCTCGTCGTATCGC (SEQ ID NO: 94)
KDM2B-sg2	GCGTTACTACGAGACGCCCCG (SEQ ID NO: 95)
KDM2B-sg3	CTTGGTCAAGCGTCCGACTG (SEQ ID NO: 96)
KDM3A-sg1	TAAATGCCGAGAGTGTGCGCT (SEQ ID NO: 97)
KDM3A-sg2	GTCTGTCAAAACCGACTTCC (SEQ ID NO: 98)
KDM3A-sg3	GATACTGCTTGGCTGTACTG (SEQ ID NO: 99)
KDM3B-sg1	TCTTGTATGGGCGCCCCGTG (SEQ ID NO: 100)
KDM3B-sg2	GCCTTGACTGTTACCGGCTC (SEQ ID NO: 101)
KDM3B-sg3	TCCTGAGCCGGTAACAGTCA (SEQ ID NO: 102)
KDM4A-sg1	ACTCCGCACAGTTAAACCA (SEQ ID NO: 103)
KDM4A-sg2	GCGGAACTCTCGAACAGTCA (SEQ ID NO: 104)
KDM4A-sg3	TTCCACTCACTTATCGCTAT (SEQ ID NO: 105)
KDM4B-sg1	CCCCGCGTACTTCTCGCTGT (SEQ ID NO: 106)
KDM4B-sg2	GTATGATGACATCGACGACG (SEQ ID NO: 107)
KDM4B-sg3	TCACCAGGTACTGTACCCCG (SEQ ID NO: 108)
KDM4C-sg1	CCTTTGCAAGACCCGCACGA (SEQ ID NO: 109)
KDM4C-sg2	AGTAGGCTTCGTGTGATCAA (SEQ ID NO: 110)
KDM4C-sg3	GTCTAAAGGAGCCCATCGTG (SEQ ID NO: 111)
KDM5A-sg1	CATGAACCCCAACGTGCTAA (SEQ ID NO: 112)
KDM5A-sg2	CTGGGATTCAAATAACTCGG (SEQ ID NO: 113)
KDM5A-sg3	TCTCTGGTATGAAAGTGCCG (SEQ ID NO: 114)
KDM5B-sg1	GTCCGCGAACTCTTCCCAGC (SEQ ID NO: 115)
KDM5B-sg2	TCGAAGACCGGGCACTCGGG (SEQ ID NO: 116)
KDM5B-sg3	GGACTTATTTTCAGCTTAATA (SEQ ID NO: 117)
KDM5C-sg1	CTTACCGCCATGACACACTT (SEQ ID NO: 118)
KDM5C-sg2	GATAACAATGCGTTCGTAG (SEQ ID NO: 119)
KDM5C-sg3	GGGCTACCCGAGCCCACCGA (SEQ ID NO: 120)
KDM5D-sg1	GATTTACTCCTCGCGTCCAA (SEQ ID NO: 121)
KDM5D-sg2	AAAGACTTACCGCGGGTGGG (SEQ ID NO: 122)
KDM5D-sg3	TAAGGCCCGACATGGAACCG (SEQ ID NO: 123)
KDM6A-sg1	ACTGTAAACTGTAGTACCTC (SEQ ID NO: 124)
KDM6A-sg2	CAGCATTATCTGCATACCAG (SEQ ID NO: 125)
KDM6A-sg3	AGACTATGAGTCTAGTTTAA (SEQ ID NO: 126)
KDM6B-sg1	TACCACAGCGCCCTTCGATA (SEQ ID NO: 127)
KDM6B-sg2	ATCCCCCTCCTCGTAGCGCA (SEQ ID NO: 128)
KDM6B-sg3	CAAAGGCTTCCCGTGCAGCG (SEQ ID NO: 129)
PHF2-sg1	TTCTGCACGGGCTTGACGTC (SEQ ID NO: 130)

[0141]

PHF2-sg2	GACGTCAAGCCCGTGCAGAA (SEQ ID NO: 131)
PHF2-sg3	CAGTGACGTGCGAGAACTACG (SEQ ID NO: 132)
PHF8-sg1	TCTGACGAACGTAGGGCTCC (SEQ ID NO: 133)
PHF8-sg2	GGCTTAGTGAAAAACGCCG (SEQ ID NO: 134)
PHF8-sg3	CCTCGCCATCATTCACTGTG (SEQ ID NO: 135)
BMI1-sg1	AACGTGTATTGTTTCGTTACC (SEQ ID NO: 136)
BMI1-sg2	TCCTACCTTATATTCAGTAG (SEQ ID NO: 137)
BMI1-sg3	AAAGGTTTACCATCAGCAGA (SEQ ID NO: 138)
BRD1-sg1	CACCGTGTCTATAGAGCCG (SEQ ID NO: 139)
BRD1-sg2	CGGCGCGAGGTGGACAGCAT (SEQ ID NO: 140)
BRD1-sg3	CGACTCACCGGCTGCGATCC (SEQ ID NO: 141)
BRD3-sg1	CGACGTGACGTTTGCAGTGA (SEQ ID NO: 142)
BRD3-sg2	CAAAGGTCGGAAGCCGGCTG (SEQ ID NO: 143)
BRD3-sg3	CATCACTGCAAACGTCACGT (SEQ ID NO: 144)
BRD4-sg1	ACTAGCATGTCTGCGGAGAG (SEQ ID NO: 145)
BRD4-sg2	TCTAGTCCATCCCCATTAC (SEQ ID NO: 146)
BRD4-sg3	GGAACAATAAAGAAGCGCT (SEQ ID NO: 147)
ING1-sg1	GAGATCGACGCGAAATACCA (SEQ ID NO: 148)
ING1-sg2	TATAAATCCGCGCCCGAAAG (SEQ ID NO: 149)
ING1-sg3	CCCATAACAGTTATTGCGCT (SEQ ID NO: 150)
ING2-sg1	GCAGCAGCAACTGTACTCGT (SEQ ID NO: 151)
ING2-sg2	GCAGCGACTCCACGCACTCA (SEQ ID NO: 152)
ING2-sg3	GATCTTCAAGAAGACCCCGC (SEQ ID NO: 153)
ING3-sg1	TCTCGCGCATTTCCGTGAAG (SEQ ID NO: 154)
ING3-sg2	CTTCACGGAAATGCGCGAGA (SEQ ID NO: 155)
ING3-sg3	TCGATACTGCATTTGTAATC (SEQ ID NO: 156)
ING4-sg1	GCTGCTCGTGCTCGTTCCAA (SEQ ID NO: 157)
ING4-sg2	CCTAGAAGGCCCGACTCAAA (SEQ ID NO: 158)
ING4-sg3	GGCACTACTCATATACTCAG (SEQ ID NO: 159)
ING5-sg1	GATCTGCTTCAAAGCGCGCC (SEQ ID NO: 160)
ING5-sg2	CTTCCAGCTGATGCGAGAGC (SEQ ID NO: 161)
ING5-sg3	GAAGTTCCTCTGAAGTTCGC (SEQ ID NO: 162)

[0142] 使用qRT-PCR在OVCAR8-ADR细胞中评估了这50种表观遗传基因的表达。使用CombiGEM方法产生双向 (153×153 sgRNA = 23,409个总组合) 合并的条形码化gRNA文库。产生慢病毒库以将文库递送到OVCAR8-ADR-Cas9细胞中。通过聚合酶链式反应 (polymerase chain reaction, PCR) 分离来自合并的细胞群体的基因组DNA以用于无偏差的条形码扩增。使用Illumina HiSeq测序对储存在大肠杆菌以及经感染的人细胞库的质粒库中单个条形码化组合的表示进行定量 (图3A-3D)。在质粒和经感染的细胞库二者中都实现了双向文库的近乎全面覆盖, 每个样品的读取为约2300万次至3400万次 (图3B), 并且观察到条形码化gRNA组合的相对均匀的分布 (图3A和3B)。此外, 在质粒和经感染的细胞库中的条形码表现之间存在高相关性 (图3C), 以及在经感染细胞库的生物学重复中表示的条形码中的高可重复性 (图3D)。因此, CombiGEM-CRISPR可用于有效地组装并将条形码化组合gRNA文库递送到人细胞中。

[0143] 为了证实gRNA在OVCAR8-ADR-Cas9细胞中编辑内源基因的功能, 进行Surveyor测

定以评估来自文库的由gRNA靶定的8个随机选择的位点的切割效率。在感染后第12天观察到插入产生效率为1.9%至26.2% (图7A-7C)。检测到感染后第12天所有gRNA靶向基因座的DNA错配的切割 (图7A和8A)。在我们的双gRNA系统中,在多个基因座处测定同时切割效率,并且在表达单gRNA或双gRNA的细胞中观察到相当的切割水平 (图7A和8A)。还检测到表达单gRNA和双gRNA细胞中靶标蛋白质水平的消耗 (图8B)。这些结果表明复合系统并没有阻碍gRNA的活性。为了将单细胞内由双gRNA指导的双切割事件与整个感染群体区分开,分离了来源于用双gRNA感染的单细胞的克隆,并使用Sanger测序检测了两个靶标基因组基因座中具有插入、缺失或突变的细胞 (图9A和9B)。

[0144] 还通过在靶标基因组基因座上进行深度测序评估了插入产生效率。在不同的gRNA之间观察到产生插入的速率的较大变化 (即,14%至93%;图16A) 和移码突变的较大变化 (即,所有插入中52%至95%;图S5B)。此外,在先前的研究中使用A375黑素瘤细胞验证的gRNA (Shalem等,2014Science 343:84-7) 在OVCAR8-ADR-Cas9细胞中显示出活性降低 (例如,对于NF1-sg4和MED12-sg1sgRNA) 和有差异的插入产生偏好 (例如,NF1-sg1sgRNA) (图16C)。这样的差异可能部分是由于靶基因座处染色质可及性的变化 (Wu等 (2014) Nat.Biotechnol.32:670-6) 和在细胞类型之间可能不同的DNA断裂修复机制 (Ghezraoui等 (2014) Mol Cell 55:829-42)。在gRNA设计优化方面的持续努力 (包括提高靶切割率 (Donesch等 (2014) Nat Biotechnol 32:1262-7) 和使脱靶切割最小化) 应该能够创建更有效的gRNA组,其将提高其对于广泛范围的细胞类型的大规模遗传扰动筛选的适用性。通过复合系统中的gRNA进一步评估了插入产生。深度测序分析检测到在sgRNA或双gRNA系统下表达的相同gRNA的大体相当的插入产生频率和偏好 (图16D)。为了将单细胞内由双gRNA指导的双切割事件与分布在整个群体中的切割事件区分开,分离了来源于用双gRNA构建体感染的单细胞的克隆。检测到在两个靶基因组基因座中具有插入、缺失或突变的细胞 (图9A-9C;表6)。我们的研究结果表明,CombiGEM组合gRNA文库可用于在OVCAR8-ADR-Cas9细胞中产生双重遗传突变体。

[0145] 启动了OVCAR8-ADR-Cas9细胞的合并的组合遗传筛选以鉴定调控癌细胞增殖的gRNA组合。构建数学模型来绘制群体中每个文库成员的丰度的相对变化如何取决于多个参数 (参见下面的方法;图17A和17B)。模拟了含有不同gRNA组合之异源亚群 (heterogeneous subpopulation) 的群体。具体来说,在模拟开始时将总体群体的具体百分比定义为含有抗增殖 (f_s) 和促增殖 (f_p) gRNA组合的亚群。在每个亚群中,在模拟开始时,通过CRISPR-Cas9系统 (p) 突变一部分细胞,导致倍增时间 ($T_{\text{倍增},m}$) 改变。该模型表明在模拟条件下整个细胞群体中具有抗增殖性gRNA的条形码化细胞的表现可被降低约23至97% (即 f_s 和 $f_p = 2, 5$ 或 10% ; $p = 0.2, 0.4, 0.6, 0.8$ 或 1.0 ; $T_{\text{倍增},m} = 36, 48$ 或 60 小时) (图17B)。一般,预计提高突变效率、提高抗增殖细胞的倍增时间、降低促增殖细胞的倍增时间以及提高群体中促增殖组合的百分比 (图17C) 将导致整个群体中抗增殖条形码的较大的条形码降低。

[0146] 在实验筛选中,为了鉴定调节癌细胞增殖的gRNA组合,将用双向组合gRNA文库感染的OVCAR8-ADR-Cas9细胞群体培养15天和20天,然后从细胞中分离基因组DNA以用于整合的条形码之无偏差的扩增和定量 (图4A,10A和10B)。在第20天和第15天的组条形码丰度 (每百万次读取进行归一化) 之间的比较产生 \log_2 (条形码计数比) 值 (图4A,10A和10B)。预期抑制细胞增殖的指导RNA组合会产生负的 \log_2 比,而预期赋予细胞生长优势的那些具有正的

\log_2 比。为了降低变异性,过滤出第15天组中小于约100个绝对读取的组合,并且将每个gRNA对(即,sgRNA-A+sgRNA-B和sgRNA-B+sgRNA-A)的两个潜在排列的 \log_2 比取平均值(图11)。对于两个生物学重复,确定每个gRNA组合的 \log_2 比并排列(图5B和12A)。大多数gRNA组合在第15天和第20天组的条形码表现之间没有显示出显著变化,包括来自GeCKOv2文库(Shalem等(2014) Science 343:84-7)的三种对照gRNA,其不具有人基因组中的中靶基因座作为内部对照。示出了61个gRNA组合在两个生物学重复中发挥可观的抗增殖作用(\log_2 比 <-0.90) (Q值 <0.01 ,表2和图12B)中,这产生了潜在的基因集以进一步探究抑制癌细胞生长的能力。

[0147] 表2:基于合并筛选抑制OVCAR8-ADR细胞增殖的双向sgRNA命中

[0148]

sgRNA-A	sgRNA-B	Log ₂ 比 - 第20天/ 第15天 重复 1	Log ₂ 比 - 第20天/ 第15天 重复 2	Z-得分 重复 1	Z-得分 重复 2	Q 值
BRD4_sg3	MLL_sg3	-2.61	-3.32	-8.45	-10.72	7.01E-33
BMI1_sg2	HDAC2_sg3	-3.91	-1.93	-12.62	-6.29	3.41E-32
BMI1_sg2	KDM1B_sg3	-1.34	-3.71	-4.39	-11.99	1.20E-23
ING3_sg3	BMI1_sg1	-3.68	-1.11	-11.87	-3.64	4.57E-21
BRD4_sg3	KDM6A_sg2	-2.30	-2.20	-7.45	-7.15	1.52E-18
BRD4_sg3	PHF2_sg2	-1.57	-2.88	-5.12	-9.31	4.05E-18
BMI1_sg2	PRMT6_sg1	-2.80	-1.59	-9.05	-5.19	1.19E-17
ING3_sg2	KDM5A_sg2	-2.28	-2.04	-7.40	-6.65	3.48E-17
KDM5B_sg3	MLL_sg3	-3.17	-1.05	-10.23	-3.47	2.61E-16
KDM6A_sg3	KDM6A_sg2	-1.95	-2.26	-6.35	-7.33	2.61E-16
BRD4_sg3	KDM4C_sg1	-1.48	-2.28	-4.84	-7.40	7.88E-13
KDM6B_sg1	KDM3A_sg2	-0.98	-2.70	-3.23	-8.74	3.00E-12
ING3_sg3	KDM6A_sg2	-1.46	-2.10	-4.79	-6.81	2.02E-11
PRMT5_sg3	PRMT5_sg3	-1.54	-1.92	-5.04	-6.26	9.13E-11
BRD4_sg3	BRD4_sg2	-2.45	-1.00	-7.93	-3.30	1.24E-10
BMI1_sg2	NSD1_sg3	-2.19	-1.20	-7.13	-3.94	2.68E-10
BMI1_sg2	MBD2_sg3	-1.44	-1.93	-4.70	-6.27	3.94E-10
BMI1_sg2	KDM1A_sg1	-1.62	-1.40	-5.29	-4.59	5.85E-08
KDM3A_sg2	PRMT5_sg3	-1.28	-1.72	-4.21	-5.62	6.62E-08
BRD4_sg3	EP300_sg3	-1.28	-1.71	-4.22	-5.58	7.28E-08
BMI1_sg2	KDM3A_sg1	-1.47	-1.48	-4.79	-4.85	1.32E-07
BRD4_sg3	KDM6B_sg1	-1.20	-1.64	-3.94	-5.35	5.58E-07

[0149]

PHF8_sg2	KDM1A_sg1	-1.11	-1.69	-3.65	-5.52	8.77E-07
KDM6A_sg3	HDAC2_sg3	-1.03	-1.77	-3.40	-5.77	8.77E-07
BRD4_sg3	KDM6B_sg2	-0.92	-1.86	-3.05	-6.05	1.13E-06
PRMT5_sg3	HDAC1_sg1	-1.26	-1.43	-4.15	-4.68	3.14E-06
PHF8_sg2	PRMT5_sg3	-1.60	-1.08	-5.23	-3.57	3.50E-06
BRD4_sg3	PHF2_sg1	-1.51	-1.13	-4.94	-3.71	5.35E-06
BRD4_sg3	EZH2_sg3	-1.02	-1.57	-3.38	-5.12	8.80E-06
PRMT5_sg3	EZH1_sg1	-0.96	-1.62	-3.19	-5.28	9.53E-06
KDM6A_sg2	KDM5A_sg2	-1.48	-1.04	-4.86	-3.44	1.72E-05
BRD4_sg3	KDM2B_sg2	-1.00	-1.49	-3.30	-4.88	2.59E-05
KDM6B_sg3	PRMT5_sg3	-1.03	-1.43	-3.41	-4.67	3.60E-05
BRD4_sg1	KDM6B_sg1	-1.30	-1.15	-4.27	-3.79	3.74E-05
KDM1A_sg3	PRMT5_sg3	-1.52	-0.91	-4.96	-3.01	4.81E-05
PRMT5_sg3	EP300_sg2	-1.41	-1.00	-4.61	-3.31	5.44E-05
BRD4_sg3	PRMT5_sg2	-1.46	-0.91	-4.79	-3.02	7.59E-05
KDM5D_sg2	MLL_sg3	-1.01	-1.34	-3.35	-4.40	9.28E-05
BRD4_sg3	PRMT5_sg3	-1.03	-1.33	-3.39	-4.35	9.28E-05
PHF2_sg1	PRMT5_sg3	-1.19	-1.13	-3.92	-3.71	1.32E-04
PRMT5_sg2	DNMT1_sg1	-1.28	-1.02	-4.21	-3.37	1.54E-04
KDM6A_sg2	KDM5C_sg1	-1.28	-1.01	-4.19	-3.34	1.73E-04
BMI1_sg2	KDM6A_sg2	-0.96	-1.30	-3.17	-4.25	2.47E-04
KDM1A_sg1	PRMT5_sg3	-0.95	-1.29	-3.15	-4.24	2.68E-04
BRD4_sg2	KDM2B_sg1	-1.06	-1.17	-3.51	-3.85	2.82E-04
KDM4A_sg2	PRMT5_sg3	-1.07	-1.16	-3.52	-3.81	3.18E-04
PRMT6_sg2	PRMT5_sg2	-1.02	-1.13	-3.37	-3.74	6.20E-04
KDM2B_sg3	PRMT5_sg3	-1.13	-1.01	-3.73	-3.34	6.95E-04
KDM6B_sg1	MLL_sg3	-0.92	-1.19	-3.05	-3.90	9.39E-04
BRD4_sg3	BMI1_sg2	-0.94	-1.14	-3.11	-3.76	1.10E-03
KDM5A_sg3	NSD1_sg3	-0.91	-1.14	-3.03	-3.75	1.37E-03
BRD4_sg3	KDM3A_sg2	-0.92	-1.13	-3.04	-3.74	1.37E-03
KDM6B_sg3	PRMT7_sg3	-1.11	-0.90	-3.65	-3.00	1.85E-03
BRD3_sg3	KDM4C_sg1	-0.99	-1.01	-3.27	-3.33	2.13E-03
KDM4C_sg1	PRMT5_sg3	-0.93	-1.06	-3.08	-3.51	2.15E-03
KDM3B_sg1	PRMT5_sg3	-1.00	-0.99	-3.31	-3.28	2.15E-03
PRMT5_sg2	MBD1_sg1	-0.94	-1.04	-3.11	-3.42	2.47E-03
EP300_sg3	MBD1_sg3	-0.92	-1.03	-3.05	-3.41	2.93E-03
ING3_sg1	BRD4_sg3	-0.99	-0.96	-3.28	-3.17	2.93E-03
KDM1A_sg1	HDAC2_sg3	-0.94	-0.95	-3.12	-3.14	4.71E-03
PRMT5_sg2	CARM1_sg1	-0.95	-0.91	-3.14	-3.02	5.78E-03

[0150] 通过使用经慢病毒递送的相应gRNA对,在个体(非合并的)细胞生长测定中评估gRNA对抑制OVCAR8-ADR-Cas9细胞增殖的能力(即在5天内约33%)来验证筛选的命中(图4C)。在8从合并的筛选中收集的数据和单独的验证测定之间存在高度一致性(图13)。总之,本文所述的方法提供了能够对卵巢癌细胞施加抗增殖作用的条形码化组合gRNA的系统筛选的实验流水线(experimental pipeline)。

[0151] 当与其他表观遗传基因靶向gRNA组合使用时,靶向表观遗传基因的许多gRNA比当

与对照gRNA组合使用时表现出更强的抗增殖作用(图4B和12A)。

[0152] 通过深度测序评估gRNA的脱靶活性,其显示了通过对两种gRNA的CRISPR设计和CCTop工具经计算预测的所有外显子脱靶基因组基因座处较低的插失产生率(即,0.15至0.38%) (图19)。总之,建立并验证了一种实验流程以用于系统筛选能够对卵巢癌细胞发挥抗增殖作用的条形码化组合gRNA。

[0153] 通过验证测定(图5A和8)证实gRNA对,靶向KDM4C和BRD4的shRNA对(图5B和14)同时导致癌细胞生长的协同降低。此外,用小分子KDM4C抑制剂SD70 (Jin等,PNAS (2014) 111:9235-9240) 和小分子BRD4抑制剂JQ1 (Asangani等,Nature (2014) 510:278-282) (图5C) 共处理协同地抑制OVCAR8-ADR细胞的增殖。类似地,同时靶向KDM6B和BRD4的gRNA对(图5A和8)和shRNA对(图5B和14)显示出协同作用,以KDM6B/6A抑制剂GSK-J4和JQ1的共处理也是如此(Kruidenier,等,Nature (2012) 488:404-408) (图5D)。通过Bliss独立性(Bliss Ann.Appl.Biol. (1939) 6:585-615) 和最高单一药剂(Borisy等,PNAS (2003) 100:7977-7982) 模型二者证实了这两种成对小分子药物组合之间的协同作用)) (图5C和5D)。

[0154] 本文所述的方法允许鉴定抑制癌细胞增殖的新表观遗传靶基因对以及协同药物疗法的潜在发展。这些方法还扩展了基于CRISPR-Cas9的系统以高通量能力进行系统多重遗传扰动筛选的效用。

[0155] 这些方法还可帮助确定生物学调查的新领域,例如研究观察表型的基础的机制。例如,在用编码靶向KDM4C和BRD4、KDM6B和BRD4的gRNAs的慢病毒感染的细胞群体中评估基因表达模式(图21A)。显著扰动的基因与涉及癌症相关途径(包括TNF α /NF κ B信号传导、p53途径和凋亡)的基因集相关(图21B)。此外,表观遗传扰动的组合作用是复杂的,并且可在不同的细胞类型之间变化(图22A和22B)。

[0156] 方法

[0157] 载体构建

[0158] 使用标准分子克隆技术(包括限制性酶消化、连接、PCR和Gibson组装) 构建载体(表3)。定制寡核苷酸购自Integrated DNA Technologies。将载体构建体转化到大肠杆菌菌株DH5 α 中,并使用50 μ g/ml羧苄青霉素(Teknova) 分离含有构建体的菌落。使用Plasmid Mini或Midi试剂盒(Qiagen) 提取和纯化DNA。使用Genewiz的DNA测序服务验证载体构建体的序列。

[0159] 表3:构建体

[0160]

构建体 ID	设计
pAWp28	pBT264-U6p-{2xBbsI}-sgRNA 支架 -{MfeI}
pAWp28-1	pBT264-U6p-GFP-sg1
pAWp28-2	pBT264-U6p-RFP-sg1
pAWp28-3	pBT264-U6p-RFP-sg2
pAWp28-4	pBT264-U6p-KDM4C-sg1
pAWp28-5	pBT264-U6p-PHF2-sg1
pAWp28-6	pBT264-U6p-KDM6B-sg2
pAWp28-7	pBT264-U6p-PHF2-sg2
pAWp28-8	pBT264-U6p-DNMT1-sg1
pAWp28-9	pBT264-U6p-DNMT3B-sg1
pAWp28-10	pBT264-U6p-PRMT2-sg3
pAWp28-11	pBT264-U6p-HDAC2-sg1
pAWp28-12	pBT264-U6p-ING4-sg1
pAWp28-13	pBT264-U6p-KDM1B-sg3
pAWp28-14	pBT264-U6p-KDM2A-sg3
pAWp28-15	pBT264-U6p-PRMT6-sg1
pAWp28-16	pBT264-U6p-BMI1-sg2
pAWp28-17	pBT264-U6p-PHF8-sg2
pAWp9	pFUGW-UBCp- <i>RFP</i> -CMVp- <i>GFP</i> -{BamHI+EcoRI}
pAWp9-1	pFUGW-UBCp- <i>RFP</i> -CMVp- <i>GFP</i> -U6p-GFP-sg1
pAWp9-2	pFUGW-UBCp- <i>RFP</i> -CMVp- <i>GFP</i> -U6p-RFP-sg1
pAWp9-3	pFUGW-UBCp- <i>RFP</i> -CMVp- <i>GFP</i> -U6p-RFP-sg2
pAWp9-4	pFUGW-UBCp- <i>RFP</i> -CMVp- <i>GFP</i> -U6p-RFP-sg1-U6p-GFP-sg1
pAWp9-5	pFUGW-UBCp- <i>RFP</i> -CMVp- <i>GFP</i> -U6p-RFP-sg2-U6p-GFP-sg1

[0161]

pAWp11	pFUGW-CMVp
pAWp12	pFUGW-CMVp- <i>GFP</i>
pAWp12-1	pFUGW-CMVp- <i>GFP</i> -[U6p-BRD4-sg3]-[U6p-PHF2-sg1]
pAWp12-2	pFUGW-CMVp- <i>GFP</i> -[U6p-BRD4-sg3]-[U6p-KDM6B-sg2]
pAWp12-3	pFUGW-CMVp- <i>GFP</i> -[U6p-BRD4-sg3]-[U6p-KDM4C-sg1]
pAWp12-4	pFUGW-CMVp- <i>GFP</i> -[U6p-BRD4-sg3]-[U6p-PHF2-sg2]
pAWp12-5	pFUGW-CMVp- <i>GFP</i> -[U6p-BRD4-sg3]
pAWp12-6	pFUGW-CMVp- <i>GFP</i> -[U6p-KDM4C-sg1]
pAWp12-7	pFUGW-CMVp- <i>GFP</i> -[U6p-KDM6B-sg2]
pAWp12-8	pFUGW-CMVp- <i>GFP</i> -[U6p-DNMT1-sg1]
pAWp12-9	pFUGW-CMVp- <i>GFP</i> -[U6p-DNMT3B-sg1]
pAWp12-10	pFUGW-CMVp- <i>GFP</i> -[U6p-PRMT2-sg3]
pAWp12-11	pFUGW-CMVp- <i>GFP</i> -[U6p-HDAC2-sg1]
pAWp12-12	pFUGW-CMVp- <i>GFP</i> -[U6p-ING4-sg1]
pAWp12-13	pFUGW-CMVp- <i>GFP</i> -[U6p-KDM1B-sg3]
pAWp12-14	pFUGW-CMVp- <i>GFP</i> -[U6p-KDM2A-sg3]
pAWp12-15	pFUGW-CMVp- <i>GFP</i> -[U6p-PRMT6-sg1]
pAWp12-16	pFUGW-CMVp- <i>GFP</i> -[U6p-BMI1-sg2]-[U6p-PHF8-sg2]
pAWp21	pLKO.1-Control sh
pAWp21-1	pLKO.1-KDM4C-sh1
pAWp21-2	pLKO.1-KDM4C-sh2
pAWp21-3	pLKO.1-KDM6B-sh
pAWp21-4	pLKO.1-BRD4-sh1
pAWp21-5	pLKO.1-BRD4-sh2
pAWp30	pFUGW-EFSp- <i>Cas9-P2A-Zeo</i>

[0162] 为了产生编码靶向特定基因之shRNA的慢病毒载体,合成含有正义和反义序列的寡核苷酸对、使其退火并克隆到AgeI和EcoRI消化的pLKO.1载体29(Addgene plasmid#10879)中。基于siRNA筛选程序(sirna.wi.mit.edu/)设计并构建了shRNA正义和反义序列(表4)。

[0163] 表4:用于个体验证测定的shRNA反义序列

[0164]

shRNA ID	shRNA反义序列
对照-sh	CGAGGGCGACTTAACCTTAGG (SEQ ID NO:163)
KDM4C-sh1	AAATCTTCGTAATCCAAGTAT (SEQ ID NO:164)
KDM4C-sh2	GTAATACCGGGTGTTCGATG (SEQ ID NO:165)
KDM6B-sh	ATTAATCCACACGAGGTCTCC (SEQ ID NO:166)
BRD4-sh1	TATAGTAATCAGGGAGGTTCA (SEQ ID NO:167)
BRD4-sh2	TTTAGACTTGATTGTGCTCAT (SEQ ID NO:168)

[0165] 为了产生编码Cas9蛋白和作为选择标志物的吉欧霉素(Zeocin)的pAWp30慢病毒表达载体,从Addgene质粒#49535扩增EFS启动子和Cas9序列,而吉欧霉素序列使用Phusion DNA聚合酶(New England Biolabs)通过PCR从Addgene质粒#25736扩增。使用Gibson Assembly Master Mix(New England Biolabs)将PCR产物克隆到pAWp11慢病毒载体骨架

中。

[0166] 为了构建含有U6启动子(U6p)-驱动靶向特定基因之sgRNA的表达的储存载体,合成具有20bp sgRNA靶序列的寡核苷酸对,退火并使用T4连接酶(New England Biolabs)将其克隆到BbsI消化的pAWp28载体中。为了构建用于U6p驱动的单个或多个sgRNA之表达的慢病毒载体,用BglIII和MfeI酶(Thermo Scientific)经由对储存载体的消化来制备U6p-sgRNA表达盒,并使用通过用BamHI和EcoRI酶(Thermo Scientific)消化载体产生的相容性黏性末端进行连接来将其分别插入pAWp12载体骨架或单sgRNA表达载体中。为了与双重RFP和GFP荧光蛋白报道子一起表达sgRNA,使用上述相同的策略将U6p驱动的sgRNA表达盒插入pAWp9(替代pAWp12)的慢病毒载体骨架中。通过将独特的BamHI和EcoRI位点引入载体中从pAWp7载体骨架修饰得pAWp9载体,以使得能够插入U6p-sgRNA表达盒。

[0167] 条形码化组合sgRNA文库池的组装

[0168] 合成了含有条形码化sgRNA序列的153个寡核苷酸对阵列(寡核苷酸F-(x)和寡核苷酸R-(x),其中x=1至153)并退火以产生这样双链插入物,其含有20bp sgRNA靶标序列、两个BbsI限制性位点、对每个sgRNA独特的8bp条形码而彼此具有至少两个不同碱基,并且在末端具有5'突出末端。为了产生合并的储存载体文库,将153个退火的插入物以相等的比例混合,并通过其相容性末端的单锅连接反应克隆到pAWp28储存载体(用BbsI和MfeI消化)中。为了构建条形码化sgRNA文库,用经BbsI消化的合并储存载体文库和这样的插入物进行另一个一锅连接反应,其中所述插入物含有sgRNA支架序列、BamHI和EcoRI限制性位点和在其末端的5'突出末端,其通过寡核苷酸对S1和S2的合成和退火制备。合并的储存载体和条形码化sgRNA文库均在Endura感受态细胞(Lucigen)中制备并通过Plasmid Midi试剂盒(Qiagen)纯化。

[0169] 用与上述产生单个和组合sgRNA构建体相同的策略构建具有单个或组合gRNA的合并慢病毒载体文库,不同之处在于使用合并的插入物和载体代替单独的插入物和载体来进行组装。简言之,通过用BglIII和MfeI对合并的储存载体文库进行单锅消化来产生合并的U6p-sgRNA插入物。将目的慢病毒载体(pAWp12)用BamHI和EcoRI消化。将经消化的插入物和载体通过其相容性末端(即,BamHI+BglIII和EcoRI+MfeI)连接以在慢病毒载体中产生合并的单向sgRNA文库(153sgRNA)。用BamHI和EcoRI再次消化单向sgRNA载体文库,并与相同的U6p-sgRNA插入库连接,以组装双向sgRNA文库($153 \times 153 \text{sgRNA} = 23,409$ 个总组合)。在合并组装步骤之后,将sgRNA定位于载体构建体的一端,并将其各自的条形码串接在另一端。在XL10-Gold超感受态细胞(Agilent Technologies)中制备慢病毒sgRNA文库,并通过Plasmid Midi试剂盒(Qiagen)纯化。

[0170] 细胞培养

[0171] 从ATC获得HEK293T细胞,并在37℃、5%CO₂下,在补充有10%热灭活胎牛血清和1X抗生素-抗真菌剂(Life Technologies)的DMEM中培养。OVCAR8-ADR细胞是T.Ochiya(日本国立癌症中心研究所,Japanese National Cancer Center Research Institute,日本)的赠品。确认了OvCAR8-ADR细胞的身份(Genetica DNA Laboratories)。通过用pAWp30载体对OVCAR8-ADR细胞进行慢病毒感染,并在200μg/ml吉欧霉素(Life Technologies)存在下筛选三周,以产生稳定表达Cas9蛋白(OVCAR-ADR-Cas9)的OVCAR8-ADR细胞。在37℃,5%CO₂下,将OVCAR8-ADR和OVCAR8-ADR-Cas9细胞在补充有10%热灭活胎牛血清和1×抗生素-抗

真菌剂的RPMI中培养。对于药物处理,使用SD70 (Xcessbio#M60194)、GSK-J4 (Cayman Chemical#12073) 和/或(+)-JQ1 (Cayman Chemical#11187) 以指定的药物剂量处理OVCAR8-ADR细胞,之后进行细胞生存力测定。

[0172] 慢病毒生产和转导

[0173] 产生慢病毒并以6孔形式在HEK283T细胞中包装。转染前将HEK293T细胞维持在约70%汇合。将FuGENE HD转染试剂(Promega)与0.5 μ g的慢病毒载体、1 μ g的pCMV-dR8.2-dvpr载体和0.5 μ g的pCMV-VSV-G载体混合在100 μ l的OptiMEM培养基(Life Technologies)中,在室温下孵育15分钟,然后添加到细胞培养物中。第二天更换培养基。在转染后48小时和96小时收集含有新产生病毒的上清液,并通过0.45 μ m聚醚砜膜(Pall)过滤。在8 μ g/ml聚凝胺(Sigma)存在下,用500 μ l经过滤的病毒上清液感染250,000个细胞过夜,从而以单个载体构建体转导。对于筛选中使用的合并慢病毒文库生产,使用相同的实验程序扩大慢病毒生产和转导。使用Amicon超离心过滤单元(Millipore)浓缩经过滤的病毒上清液。在8 μ g/ml聚凝胺的存在下,以0.3至0.5的感染倍数感染细胞以确保大多数细胞中的单拷贝整合,其对应于30-40%的感染效率。筛选中使用的细胞总数约为文库大小的300倍,以保持文库覆盖率并降低由于随机慢病毒融合到基因组中的任何乱真效应(spurious effect)。在感染后的第二天更换细胞培养基并培养指定的时间段,之后进行实验。

[0174] 用于条形码排序的样品制备

[0175] 为了从培养细胞中制备样品用于条形码测序,使用DNeasy Blood&Tissue试剂盒(Qiagen)根据制造商的方案提取并制备基因组DNA。对于条形码化sgRNA质粒文库,使用Plasmid Midi试剂盒(Qiagen)提取转化到大肠杆菌中的质粒DNA。使用Quant-iT PicoGreen dsDNA测定试剂盒(Life Technologies)测定DNA浓度。

[0176] 使用Kapa Hotstart Ready Mix(Kapa Biosystems)从质粒/基因组DNA样品中PCR扩增含有表示合并载体和感染细胞文库中各组合之独特条形码的约360bp片段。对于质粒DNA,向25 μ l PCR反应添加1ng的DNA模板。对于基因组DNA,向50 μ l的PCR反应添加800ng的DNA,并且对于每个基因组DNA样品进行共计64个PCR反应以确保被扩增的细胞基因组的数量大于文库大小的100倍。此外,优化PCR参数以确保在指数期间保持PCR扩增步骤以避免PCR偏差。在PCR期间添加Illumina锚定序列和8个碱基对索引条形码以用于多重测序。用于扩增条形码序列的引物对序列是:

[0177] 5'-AATGATACGGCGACCACCGAGATCTACACGGATCCGCAACGGAATTC-3' (SEQ ID NO:1) 和

[0178] 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNGGTTGCGTCAGCAAACACAG-3' (SEQ ID NO:2),其中NNNNNNN表示分配给每个实验样品的特定索引条形码。

[0179] 然后通过1.5%琼脂糖凝胶上运行,基于片段大小来纯化含有条形码序列的PCR产物,并使用QIAquick凝胶提取试剂盒(Qiagen)进一步提取。使用KAPA SYBR Fast qPCR Master Mix(Kapa Biosystems)和Illumina文库定量试剂盒(Kapa Biosystems)通过定量PCR测定PCR产物浓度。用于定量PCR的正向和反向引物分别为

[0180] 5'-AATGATACGGCGACCACCGA-3' (SEQ ID NO:3) 和

[0181] 5'-CAAGCAGAAGACGGCATACGA-3' (SEQ ID NO:4)。然后将来自不同样品的PCR产物以所需的比例合并以用于多重样品测序,并与CombiGEM条形码引物

[0182] (5'-CCACCGAGATCTACACGGATCCGCAACGGAATTC-3' (SEQ ID NO:5)) 和索引条形码引

物

[0183] (5'-GTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACC-3' (SEQ ID NO:6)) 上样到 Illumina HiSeq 系统上。

[0184] 条形码测序数据分析

[0185] 从 Illumina 测序数据中处理每个 sgRNA 组合的条形码读取。对于由索引条形码分类的每个样品,将表示每个组合的条形码读取以每百万次读取进行归一化。作为细胞增殖的测量,以倍数变化计算第20天组相比于第15天组的归一化条形码读取的条形码计数比。促增殖和抗增殖表型的归一化条形码读取分别为 >1 和 <1 ,而没有表型变化导致倍数变化 $=1$ 。过滤掉在第15天组得到小于约100绝对读取的条形码,以提高数据的可靠性。将每个相同 sgRNA 组合的不同可能顺序的倍数变化进行平均,观察到倍数变化的高度一致性(即,超过82%和95%的组合的变异系数(coefficient of variation, CV)分别 <0.2 和 <0.4 (图11)),将计算出的倍数变化进行对数转化以得到 \log_2 比。用相同慢病毒文库之独立感染的两个生物学重复进行筛选。按所有实验条件下的 \log_2 比将组合进行排序。将命中最高的一组(空心圆圈)定义为其 \log_2 比值与两个生物学重复中仅含有对照 sgRNA 的 sgRNA 组合(空心三角形)的平均值相比至少相差三个标准差的那些(图4B、12A和12B)。

[0186] 细胞生存力测定

[0187] 进行MTT比色测定以评估细胞生存力。对于每个96孔,向细胞培养物中添加100 μ l 的MTT (3-(4,5-二甲基噻唑-2-基)-2,5-二苯基四唑溴化物) 溶液(Sigma)。将细胞在37°C和5%CO₂下孵育3小时。活细胞将可溶性MTT盐转化为不溶性蓝色甲臞晶体。在37°C下,将甲臞晶体用100 μ l溶解缓冲液溶解。使用Synergy H1微板读数仪(BioTek) 测量570nm和650nm(参考)的光密度(optical density, OD)的吸光度读数。

[0188] 药物协同作用量化

[0189] Bliss独立性(Bliss independence, BI) 模型(Bliss Ann. Appl. Biol. (1939) 26: 585-615) 和最高单一药剂(Highest Single Agent, HSA) 模型(Borisy等, PNAS (2003) 100: 7977-7982) 是评估药物组合之间的协同作用的常用方法。

[0190] 基于BI模型,由下式得到预期作用($E_{\text{预期}}$):

$$E_{\text{预期}} = E_A + E_B - (E_A \times E_B),$$

[0192] 其中 E_A 是单独药物A在某一浓度下观察到的生长抑制作用, E_B 是单独药物B在某一浓度下观察到的生长抑制作用。 $E_{\text{观察}}$ 是对于药物组合(A+B)(各自的浓度分别与 E_A 和 E_B 相同)观察到的生长抑制作用。每个作用表示为0和1之间的分数抑制。当 $E_{\text{观察}} - E_{\text{预期}} > 0$ 时,两种药物被认为是协同相互作用。

[0193] HSA模型类似于BI模型,不同之处在于根据HSA模型, $E_{\text{预期}}$ 等于组合的单个药剂(E_A 或 E_B)在与药物组合(A+B)中之相同浓度下产生的较大生长抑制作用。

[0194] 为了提高我们标准的严格性,在两种模型中对于至少两种不同浓度的组合,如果 $E_{\text{观察}} - E_{\text{预期}} > 0.1$ (即,超过在图5C和5D中预测的BI和HSA模型中10%的过量抑制),则认为两种药物是协同作用。

[0195] 流式细胞术

[0196] 在感染后第4天和第8天收集细胞。将样品洗涤并重悬于补充有2%胎牛血清的1 \times PBS中。将重悬的细胞通过细胞粗滤器以去除任何细胞团,然后上样到LSRII Fortessa流式

细胞仪 (Becton Dickinson) 上。每个样品至少获得20,000个事件。基于细胞样品选择适当的激光组和过滤器。使用前向散射和侧向散射来鉴定适当的细胞群体。使用制造商的内置软件分析数据。

[0197] 荧光显微镜

[0198] 慢病毒感染后三天,在倒置荧光显微镜 (Zeiss) 下直接观察培养的细胞。使用 Zeiss 内置软件捕获图像。

[0199] 免疫印迹分析

[0200] 将细胞在补充有蛋白酶抑制剂的2×RIPA缓冲液中裂解。使用研磨电动机混合器 (Agros) 将裂解物匀浆30秒,然后在4℃下以15,000rpm离心15分钟。使用BCA测定 (Thermo Scientific) 对上清液进行定量。将蛋白质在99℃下变性5分钟,然后在4-15%聚丙烯酰胺凝胶 (Bio-Rad) 上进行凝胶电泳。在4℃下,将蛋白质在80V下转移至硝酸纤维素膜2小时。使用的一抗是:抗BRD4 (1:2,000, Cell Signaling#13440)、抗KDM4C (1:1,000, Abcam ab85454)、抗KDM6B (1:1000, Abcam ab85392) 和抗β-肌动蛋白 (1:4,000, Abcam ab6276)。使用的二抗是:HRP连接的抗兔IgG (1:2,000, Cell Signaling#7074) 和HRP连接的抗小鼠IgG (1:4,000, Cell Signaling#7076)。将膜通过SuperSignal West Pico化学发光底物 (Thermo Scientific) 显影,并使用ChemiDoc Touch成像系统 (BioRad) 成像。

[0201] 用于基因组修饰的Surveyor测定和测序分析

[0202] 进行Surveyor测定以评估DNA切割效率。根据制造商的方案,使用QuickExtract DNA提取溶液 (Epicenter) 从细胞培养物中提取基因组DNA。使用Phusion DNA聚合酶和表5中列出的引物通过PCR产生携带靶标基因座的扩增子。将约200ng的PCR扩增子进行变性、自退火,并与1.5μl的Surveyor核酸酶 (Transgenomic) 在42℃温育30分钟。然后在2%琼脂糖凝胶上分析样品。使用ImageJ软件对DNA条带强度进行定量,并用下式估计插失发生率 (Ran 等, Nat. Protoc. (2013) 8:2281-2308) :

[0203] $\text{插失}(\%) = 100 \times (1 - ((1 - f_{\text{切割}}))^{\frac{1}{2}})$ 的平方根),

[0204] 其中 $f_{\text{切割}} = (b+c) / (a+b+c)$, 其中a是未切割的PCR扩增子的条带强度,而b和c是每个经切割条带的强度。预期的靶标等位基因的未切割和切割的条带列于图7和8中。

[0205] 进行Sanger测序以分析由组合sgRNA的表达所产生的基因组修饰。将用组合sgRNA构建体感染的细胞培养12天,并通过连续稀释培养物将其在96孔板中重新铺板为单细胞。使用QuickExtract DNA提取溶液培养5至21天后,从分离的单细胞扩增克隆中提取基因组DNA,并如上所述通过PCR制备携带靶标等位基因的扩增子。使用TA克隆试剂盒 (Life Technologies), 根据制造商的方案,将PCR扩增子克隆到TOPO载体中,并使用Sanger测序鉴定核苷酸突变、插入和缺失。

[0206] 表5: Surveyor测定和Sanger测序中使用的PCR引物的列表

[0207]

靶 sgRNA ID	正向引物 (5' 到 3')	反向引物 (5' 到 3')
BMI1-sg2	AGAAATTAAACGGCTACCCTCCA (SEQ ID NO: 169)	GTTGGTACAAAGTGGTGAAGGC (SEQ ID NO: 183)
BRD4-sg2	TCCATAGTGTCTTGAGCACCAC (SEQ ID NO: 170)	ACGTGGCTTCATTGTACATCCT (SEQ ID NO: 184)
BRD4-sg3	CAC TTGCTGATGCCAGTAGGAG (SEQ ID NO: 171)	AAGCACATGCTTCAGGCTAACA (SEQ ID NO: 185)
DNMT1-sg1	GTGAATAGCTTGGAATGTGGG (SEQ ID NO: 172)	TCATCTGCTCTTACGCTTAGCC (SEQ ID NO: 186)
DNMT3B-sg1	GCCACACTCTACATGGGAGC (SEQ ID NO: 173)	CTCGGCAACCTCCATACAT (SEQ ID NO: 187)
HDAC2-sg1	GACTTTTCCATCAGGGACACCT (SEQ ID NO: 174)	AACCATGCACAGAATCCAGATTTA (SEQ ID NO: 188)
ING4-sg1	GGTGGACAAACACATTCGGC (SEQ ID NO: 175)	AAGAGTTCTTGGCGCAGACA (SEQ ID NO: 189)
KDM1B-sg3	CCTATCATTGCCCAAGGAGTC (SEQ ID NO: 176)	TCGTCCAAGTTACAGTCATCACA (SEQ ID NO: 190)
KDM2A-sg3	CTAGGCCTCCGACAGTTGTAAT (SEQ ID NO: 177)	TCCTCTGGTGCACAGAAAAGTC (SEQ ID NO: 191)

[0208]

KDM4C-sg1	AGCCACCCTTGGTTGGTTTT (SEQ ID NO: 178)	TTCTCTCCAGACACTGCCCT (SEQ ID NO: 192)
KDM6B-sg2	GGTAAGGGAAACTCTGGGGC (SEQ ID NO: 179)	GTGCCCAGAACTACTGCCAT (SEQ ID NO: 193)
PHF8-sg2	CTCCCTCCCTTCCTAAGGCT (SEQ ID NO: 180)	GAGGTGAGTCCAGCTTCCC (SEQ ID NO: 194)
PRMT2-sg3	ATTGCCTTAAGTCGACACCTGAT (SEQ ID NO: 181)	CACCTTACAGGCACTGCGTT (SEQ ID NO: 195)
PRMT6-sg1	GACTGTAGAGTTGCCGGAACAG (SEQ ID NO: 182)	CTCCCTCCCTAGAGGCTATGAG (SEQ ID NO: 196)

[0209] 表6:携带BMI1-sg2和PHF8-sg2表达构建体的OVCAR8-ADR-Cas9单细胞中靶标等位基因的序列。黑体的核苷酸表示PAM序列;而加下划线的核苷酸是指碱基对插入或突变。

[0210]

单细胞	sgRNA ID	序列 (5' 到 3')	SEQ ID NO.	插失
1	BMI1-sg2	TACAACCTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTTCAG----- GTCTTGTGAACTTGGACATCACAAATAGGAC	231	-10bp
1	BMI1-sg2	TACAACCTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTTCAG----- GTCTTGTGAACTTGGACATCACAAATAGGAC	231	-10bp
1	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAAC----- TTCAAAGGGGCATGATACACACAAGGGGAAACC AGTGAAGACC	232	-10bp
1	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAAC----- TTCAAAGGGGCATGATACACACAAGGGGAAACC AGTGAAGACC	232	-10bp
2	BMI1-sg2	TACAACCTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATA----- GGTCTGGTCTTGTGAACTTGGACATCACAAATAG GAC	233	-9bp
2	BMI1-sg2	TACAACCTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTTCAGT--- GGTCTGGTCTTGTGAACTTGGACATCACAAATAG GAC	234	-3bp
2	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA- CGCCGTGGATCTTCAAAGGGGCATGATACACAC AAGGGGAAACCAGTGAAGACC	235	-1bp
2	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA-C---- TGGATCTTCAAAGGGGCATGATACACACAAGGG GAAACCAGTGAAGACC	236	-5bp
3	BMI1-sg2	ACTCCAATAATAATTACAAACAAGGAATTTCAAC AGTTTCCTACCTTATATTCTGTGATCTGTGGTCTG GTCTTGTGAACTTGGACATCACAAATAGGAC	237	2mut & +4bp

[0211]

3	BMI1-sg2	ACTCCAATAATAATTACAAACAAGGAATTTCAAC AGTTTCCTACCTTATATTC <u>GTGATCTGTGGTCTG</u> GTCTTGTGAACTTGGACATCACAAATAGGAC	237	2mut & +4bp
3	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAAGTGAAGACC	238	-3bp
3	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAAGTGAAGACC	238	-3bp
4	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
4	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
4	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAAGTGAAGACC	240	WT
4	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAAGTGAAGACC	240	WT
5	BMI1-sg2	ATTTCAACAGTTTCCTACCTTATATACTATACTAT <u>ATATATATATATATACTATATATATAGTGGTCTG</u> GTCTTGTGAACTTGGACATCACAAATAGGAC	241	+30bp
5	BMI1-sg2	ATTTCAACAGTTTCCTACCTTATATACTATACTAT <u>ATATATATATATATACTATATATATAGTGGTCTG</u> GTCTTGTGAACTTGGACATCACAAATAGGAC	241	+30bp
5	PHF8-sg2	CTCTCATGTTTTTCTGGCTTAGTGAAAAAACTTCA <u>CTAAGTTTTTACTTAGTGGATCTTCAAAGGGGCA</u> TGATACACACAAGGGGAAACCAAGTGAAGACC	242	+16bp & 2mut
5	PHF8-sg2	CTCTCATGTTTTTCTGGCTTAGTGAAAAAACTTCA <u>CTAAGTTTTTACTTAGTGGATCTTCAAAGGGGCA</u> TGATACACACAAGGGGAAACCAAGTGAAGACC	242	+16bp & 2mut
6	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTC----- TGGTCTGGTCTTGTGAACTTGGACATCACAAATA GGAC	243	-5bp
6	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTAC----- TTGTGAACTTGGACATCACAAATAGGAC	244	-24bp
6	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- C <u>ATGGATCTTCAAAGGGGCATGATACACACAAG</u> GGGAAACCAAGTGAAGACC	245	-4bp & 1mut
6	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACG--- TGGATCTTCAAAGGGGCATGATACACACAAGGG GAAACCAAGTGAAGACC	246	-3bp

[0212]

7	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCT----- AGTGGTCTGGTCTTGTGAACTTGGACATCACAAA TAGGAC	247	-10bp
7	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCT----- AGTGGTCTGGTCTTGTGAACTTGGACATCACAAA TAGGAC	247	-10bp
7	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAAGTGAAGACC	246	-3bp
7	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAAGTGAAGACC	246	-3bp
8	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCT--- GTGGTCTGGTCTTGTGAACTTGGACATCACAAAT AGGAC	248	1mut & -3bp
8	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCT--- GTGGTCTGGTCTTGTGAACTTGGACATCACAAAT AGGAC	248	1mut & -3bp
8	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAAGTGAAGACC	246	-3bp
8	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAAGTGAAGACC	246	-3bp
9	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
9	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTG----- TGGTCTGGTCTTGTGAACTTGGACATCACAAATA GGAC	243	-5bp
9	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CATGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAAGTGAAGACC	245	-4bp & 1mut
9	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAAGTGAAGACC	246	-3bp
10	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
10	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT

[0213]

10	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACTAAGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	249	3mut
10	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
11	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
11	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTAC----- TAGTGGTCTGGTCTTGTGAACTTGGACATCACAA ATAGGAC	250	-11bp
11	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
11	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
12	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
12	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
12	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
12	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
13	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
13	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
13	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGA----- ATCTTCAAAGGGGCATGATACACACAAGGGGAA ACCAGTGAAGACC	251	-13bp
13	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGA----- ATCTTCAAAGGGGCATGATACACACAAGGGGAA ACCAGTGAAGACC	251	-13bp
14	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT

[0214]

14	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
14	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACG--- TGGATCTTCAAAGGGGCATGATACACACAAGGG GAAACCAGTGAAGACC	246	-3bp
14	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACG--- TGGATCTTCAAAGGGGCATGATACACACAAGGG GAAACCAGTGAAGACC	246	-3bp
15	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
15	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
15	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAAC--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
15	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAAC--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
16	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
16	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCA----- TGGTCTGGTCTTGTGAACTTGGACATCACAAATA GGAC	243	-5bp
16	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAAC--- CATGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	245	-4bp & 1mut
16	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAAC--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
17	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCA--- GTGGTCTGGTCTTGTGAACTTGGACATCACAAAT AGGAC	234	-3bp
17	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCA--- GTGGTCTGGTCTTGTGAACTTGGACATCACAAAT AGGAC	234	-3bp
17	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT

[0215]

17	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
18	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
18	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
18	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
18	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
19	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
19	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCAA----- GTGGTCTGGTCTTGTGAACTTGGACATCACAAAT AGGAC	252	-14bp & 1mut
19	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
19	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CATGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	245	-4bp & 1mut
20	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
20	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	239	WT
20	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAAC--- GTGGATCTTCAAAGGGGCATGATACACACAAGG GGAAACCAGTGAAGACC	253	-3bp
20	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAAC--- GTGGATCTTCAAAGGGGCATGATACACACAAGG GGAAACCAGTGAAGACC	253	-3bp
21	BMI1- sg2	ACAACTCCAATAATAATTACAAACAAGGAATTT AACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	254	+1bp
21	BMI1- sg2	ACAACTCCAATAATAATTACAAACAAGGAATTT AACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACTTGGACATCACAAATAGGAC	254	+1bp

[0216]

21	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
21	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACG--- TGGATCTTCAAAGGGGCATGATACACACAAGGG GAAACCAGTGAAGACC	246	-3bp
22	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTA----- GTGGTCTGGTCTTGTGAACTTGGACATCACAAAT AGGAC	255	-14bp
22	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTC--TAG- GGTCTGGTCTTGTGAACTTGGACATCACAAATAG GAC	256	-3bp
22	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
22	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
23	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCAG----- TTGTGAACTTGGACATCACAAATAGGAC	257	-13bp
23	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATA----- TAGTGGTCTGGTCTTGTGAACTTGGACATCACAA ATAGGAC	258	-5bp
23	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
23	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
24	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTC--- GTGGTCTGGTCTTGTGAACTTGGACATCACAAAT AGGAC	239	-3bp
24	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTA----- GTGGTCTGGTCTTGTGAACTTGGACATCACAAAT AGGAC	255	-14bp
24	PHF8-sg2	TCTGACTCACAATCTCTCATGTTTTTCTGGCTTAG TGAAAAACATGCCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	259	+3bp & 2mut
24	PHF8-sg2	TCTGACTCACAATCTCTCATGTTTTTCTGGCTTAG TGAAAAACATGCCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	259	+3bp & 2mut

[0217]

25	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCA--- GTGGTCTGGTCTTGTGAACCTGGACATCACAAAT AGGAC	239	-3bp
25	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCA--- GTGGTCTGGTCTTGTGAACCTGGACATCACAAAT AGGAC	239	-3bp
25	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
25	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
26	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCAAGTAGTGGTCT GGTCTTGTGAACCTGGACATCACAAATAGGAC	239	WT
26	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCAAGT--- GGTCTGGTCTTGTGAACCTGGACATCACAAATAG GAC	234	-3bp
26	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
26	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
27	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTAT----- TGGTCTGGTCTTGTGAACCTGGACATCACAAATA GGAC	260	-9bp
27	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTAT----- TGGTCTGGTCTTGTGAACCTGGACATCACAAATA GGAC	260	-9bp
27	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
27	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACG--- TGGATCTTCAAAGGGGCATGATACACACAAGGG GAAACCAGTGAAGACC	246	-3bp
28	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACA----- GTGGTCTGGTCTTGTGAACCTGGACATCACAAAT AGGAC	261	-22bp
28	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACA----- GTGGTCTGGTCTTGTGAACCTGGACATCACAAAT AGGAC	261	-22bp

[0218]

28	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
28	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
29	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACCTTGGACATCACAAATAGGAC	239	WT
29	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTC----- TGGTCTGGTCTTGTGAACCTTGGACATCACAAATA GGAC	243	-5bp
29	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CATGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	245	-4bp & 1mut
29	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
30	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACCTTGGACATCACAAATAGGAC	239	WT
30	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACCTTGGACATCACAAATAGGAC	239	WT
30	PHF8-sg2	TCTCATGTTTTTCTGGCTTAGTGAAATCTAAGCTT <u>AGTGAAATCTAAGCCGTGGATCTTCAAAGGGGC</u> ATGATACACACAAGGGGAAACCAGTGAAGACC	262	+17bp & 1mut
30	PHF8-sg2	TCTCATGTTTTTCTGGCTTAGTGAAATCTAAGCTT <u>AGTGAAATCTAAGCCGTGGATCTTCAAAGGGGC</u> ATGATACACACAAGGGGAAACCAGTGAAGACC	262	+17bp & 1mut
31	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCTTATATGGTCTG GTCTTGTGAACCTTGGACATCACAAATAGGAC	263	5mut
31	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTA----- GTGGTCTGGTCTTGTGAACCTTGGACATCACAAAT AGGAC	255	-14bp
31	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp
31	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAA--- CGTGGATCTTCAAAGGGGCATGATACACACAAG GGGAAACCAGTGAAGACC	246	-3bp

[0219]

32	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACCTTGGACATCACAAATAGGAC	239	WT
32	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATA--- TAGTGGTCTGGTCTTGTGAACCTTGGACATCACAA ATAGGAC	264	-3bp & 1mut
32	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
32	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
33	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAG----- AACTTGGACATCACAAATAGGAC	265	-18bp
33	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAG----- AACTTGGACATCACAAATAGGAC	265	-18bp
33	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
33	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
34	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACCTTGGACATCACAAATAGGAC	239	WT
34	BMI1-sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTAGTAGTGGTCT GGTCTTGTGAACCTTGGACATCACAAATAGGAC	239	WT
34	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
34	PHF8-sg2	GTTCTGACTCACAATCTCTCATGTTTTTCTGGCT AGTGAAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	266	+1bp
35	BMI1-sg2	TAATTACAAACAAGGAATTTCAACAGTTTCCTAC CTTATATTCAGTATAATATATTCATTAGTGGTCTG GTCTTGTGAACCTTGGACATCACAAATAGGAC	267	+14bp
35	BMI1-sg2	TAATTACAAACAAGGAATTTCAACAGTTTCCTAC CTTATATTCAGTATAATATATTCATTAGTGGTCTG GTCTTGTGAACCTTGGACATCACAAATAGGAC	267	+14bp
35	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTG----- GATCTTCAAAGGGGCATGATACACACAAGGGGA AACCAGTGAAGACC	268	-13bp

35	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTCTGGCT TAGTG----- GATCTTCAAAGGGGCATGATACACACAAGGGGA AACCAGTGAAGACC	268	-13bp
36	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTC-- TAGTGGTCTGGTCTTGTGAACCTGGACATCACAA ATAGGAC	269	-2bp
36	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTC-- TAGTGGTCTGGTCTTGTGAACCTGGACATCACAA ATAGGAC	269	-2bp
36	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTCTGGCT TAGTGAAAAAACG--- TGGATCTTCAAAGGGGCATGATACACACAAGGG GAAACCAGTGAAGACC	246	-3bp
36	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
37	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTATATAGTGGTCTG GTCTTGTGAACCTGGACATCACAAATAGGAC	270	3mut
37	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCAGTAGTGGTCT GGTCTTGTGAACCTGGACATCACAAATAGGAC	239	WT
37	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTCTGGCT TAGTGAAAAA-C--- TGGATCTTCAAAGGGGCATGATACACACAAGGG GAAACCAGTGAAGACC	236	-5bp
37	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTCTGGCT TAGTGAAAAA-C--- TGGATCTTCAAAGGGGCATGATACACACAAGGG GAAACCAGTGAAGACC	236	-5bp
38	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTC--- GTGGTCTGGTCTTGTGAACCTGGACATCACAAAT AGGAC	239	-3bp
38	BMI1- sg2	TAATTACAAACAAGGAATTTCAACAGTTTCCTAC CTTATATTCAGGTAGTGAATCTGAATAGTGGTCT GGTCTTGTGAACCTGGACATCACAAATAGGAC	271	+14bp
38	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTCTGGCT TAGTGAAAAAAC--- GTGGATCTTCAAAGGGGCATGATACACACAAGG GAAACCAGTGAAGACC	253	-3bp
38	PHF8- sg2	CGTTCTGACTCACAATCTCTCATGTTTTCTGGCT TAGTGAAAAAAC--- GTGGATCTTCAAAGGGGCATGATACACACAAGG GAAACCAGTGAAGACC	253	-3bp
39	BMI1- sg2	TACAACTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTT----- CTGGTCTTGTGAACCTGGACATCACAAATAGGAC	272	-15bp

[0220]

[0221]

39	BMI1-sg2	TACAACCTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATTCA--- GTGGTCTGGTCTTGTGAACCTTGGACATCACAAAT AGGAC	239	-3bp
39	PHF8-sg2	ACAATCTCTCATGTTTTTCTGGCTTAGTGAAICTT <u>CAAAGGGATCTTCAAAAAGGATCTTCAAAGGGGC</u> ATGATACACACAAGGGGAAACCAGTGAAGACC	273	+11bp & 4mut
39	PHF8-sg2	ACAATCTCTCATGTTTTTCTGGCTTAGTGAAICTT <u>CAAAGGGATCTTCAAAAAGGATCTTCAAAGGGGC</u> ATGATACACACAAGGGGAAACCAGTGAAGACC	273	+11bp & 4mut
40	BMI1-sg2	TACAACCTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATT----- GTGAACCTTGGACATCACAAATAGGAC	274	-18bp
40	BMI1-sg2	TACAACCTCCAATAATAATTACAAACAAGGAATTT CAACAGTTTCCTACCTTATATT--- AGTGGTCTGGTCTTGTGAACCTTGGACATCACAA TAGGAC	275	-3bp
40	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT
40	PHF8-sg2	CGTTCTGACTCACAATCTCTCATGTTTTTCTGGCT TAGTGAAAAAACGCCGTGGATCTTCAAAGGGGC ATGATACACACAAGGGGAAACCAGTGAAGACC	240	WT

[0222] RNA提取和定量RT-PCR (qRT-PCR)

[0223] 根据制造商的方案,使用TRIzol Plus RNA纯化试剂盒 (Life Technologies) 从细胞中提取RNA,并用PureLink柱上DNase试剂盒 (Life Technologies) 进行处理。使用NanoDrop分光光度计测定RNA质量和浓度。使用SuperScript III逆转录酶 (Life Technologies)、随机引物Mix (New England Biolabs) 和RNase OUT (Invitrogen) 将RNA样品逆转录。为了评估基因表达水平,使用SYBR FAST qPCR MasterMix (KAPA) 在LightCycler480系统 (Roche) 上进行定量PCR。使用LifeCyler480 SW 1.1内置软件对数据进行量化和分析。使用PrimerBlast (NCBI) 设计和评估PCR引物。引物序列列于表7。

[0224] 表6:在qRT-PCR中使用的PCR引物

[0225]

靶基因ID	正向引物(5'到3')	反向引物(5'到3')
BRD4	GTTGATGTGATTGCCGGCTC (SEQ ID NO: 197)	TTAGGCAGGACCTGTTTCGG (SEQ ID NO: 200)
KDM4C	CGTACGGGTTTCATGCAAGTT (SEQ ID NO: 198)	CGTTTGCTTAAGAGCACCTCC (SEQ ID NO: 201)
KDM6B	CCCCTCACCGCCTATCAGTA (SEQ ID NO: 199)	TCTTGAACAAGTCGGGGTCCG (SEQ ID NO: 202)

[0226] 表8:在用于插失检测的深度测序中使用的PCR引物

[0227]

靶 gRNA ID	靶位 点的 类型	20 bp gRNA 靶向序列 (5'到3')	SEQ ID NO.	正向引物 (5'到3')	SEQ ID NO.	反向引物 (5'到3')	SEQ ID NO.
NF1- sg1	中靶	GTTGT GCTCA GTACT GACTT	276	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- GGTATCTGTGGTT GATGCAGTTTCC	304	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- AGTAGTGAGGCC GCTTATAACC	332
NF1- sg4	中靶	TTTCA GCTTC CAATA AAAAC	277	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- GGTATCTGTGGTT GATGCAGTTTCC	304	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- AGTAGTGAGGCC GCTTATAACC	332
NF2- sg2	中靶	ATTCC ACGGG AAGGA GATCT	278	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- GCACAGAGCTGC TGCTTGGAGTG	305	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- TAACAAGGAGAT GCCCTGGCTGG	333
MED1 2-sg1	中靶	AGGAT TGAAG CTGAC GTTCT	279	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- CGCTTTCCTGCCT CAGGATGAAC	306	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- AGGTCATGAAGG CAAACCTCAGCC	334
PHF8- sg2	中靶	GGCTT AGTGA AAAAA CGCCG	280	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- TTGGAAGAGAAG GATCTGCTGAGG C	307	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- CACCTGTCAAAA GTCCTACTCCGG	335
BMI- sg2	中靶	TCCTA CCTTAT ATTCA GTAG	281	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- GCTACCCTCCACA AAGCACACAC	308	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- CCTGGAGACCAG CAAGTATTGTCC	336
KDM4 C-sg1	中靶	CCTTTG CAAGA CCCGC ACGA	109	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- CCTTCAGAAACA ATGTCCCAAATCG	309	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- GTCCTCTGAACCC CAGCTGTAAG	337
KDM4 C-sg1	脱靶	GCTTT GCCCG AACCG CACGA	282	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- AGCCTTCTGAGA GCGGGCTAG	310	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- CAAACAGAGGCC AAAGGGTGTCCC	338
KDM4 C-sg1	脱靶	CCTAG GCCAG ACCTG CACGA	283	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- GCCTCCTCTCATC CTCTCGCTTC	311	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- ACAGGAGGTCGT GGTGCAGTTCTC	339

KDM4 C-sg1	脱靶	GCTCT GGAAG ACCCG CACCA	284	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- GATTGGCTCCAA GCGGCCATCAAA C	312	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- TTGTGTGAGGAA CGTTGACGCTACC	340
KDM4 C-sg1	脱靶	CCTTAT CAAGA CCCAC ACCA	285	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- TTGAACTCAAGG CTCAGCCAACAG GC	313	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- GAGCGTAGGTCC TCTGCATGGAG	341
KDM6 B-sg2	中靶	ATCCC CCTCCT CGTAG CGCA	286	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- CAACTCAGGCTG GATGCATCGG	314	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- CCACAGAATGAC AGGAACCCATGG	342
KDM6 B-sg2	脱靶	CTGCT CCTCCT CGTAG CGCT	287	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- TTGGTGGCCGCTG AGTGTGTGTAC	315	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- ACTGAGCAGAGC CTAGGAGGCAG	343
KDM6 B-sg2	脱靶	TGCGC CCTCCT CCTAG CGCA	288	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- CAGCATGTTGAC ATAGCGGC	316	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- TGTTGCCAGATCC AGAGGCGTC	344
KDM6 B-sg2	脱靶	CTCCTC CTCCG CGTAG CGCT	289	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- TGAGAGGAGATG AGTCGGGGTC	317	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- AACTGGCCCGAG TAGTCGGAGCAG	345
KDM6 B-sg2	脱靶	CTGCC CCTCCT GGTAG CGCC	290	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- GACGGGTCAAAG CCTCAGGAGAG	318	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- TCCTCAGAGTGTG TGGAAGTGCTGG	346
KDM6 B-sg2	脱靶	AACCA GCTCC TCGTA GCTCA	291	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- TTAGCTGCCCAGC TCACAGCTACC	319	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- AAGAGCTCCTAG GGGAGGATCAG	347
KDM6 B-sg2	脱靶	ACCGC CCTCCT CCTAG CTCA	292	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- GAGCCCCAAGAG CGAGACAA	320	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- TGGCAGGAGCAC AGCCTAAGGA	348
KDM6 B-sg2	脱靶	AGCCC GCTCC TCGTG GGGCA	293	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT- GGCGCTCAGAAG GCTGTGCAG	321	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT- ACACCCGCCTCG GAGATCAACAC	349
BRD4- sg3	中靶	GGGAA CAATA AAGAA	294	ACACTCTTTCCCT ACACGACGCTCTT CCGATCT-	322	GTGACTGGAGTTC AGACGTGTGCTCT TCCGATCT-	350

[0229] 表9:在用于插失检测的全基因组扩增中使用的PCR引物

	靶 sgRNA ID	正向引物 (5' 到 3')	反向引物 (5' 到 3')
[0230]	BMI1-sg2	GTCTCGTGGGCTCGGAGATGTGTAT AAGAGACAGGCTACCCTCCACAAA GCACACAC (SEQ ID NO: 360)	TCGTCGGCAGCGTCAGATGTGTATA AGAGACAGCCTGGAGACCAGCAAGT ATTGTCC (SEQ ID NO: 362)
	PHF8-sg2	GTCTCGTGGGCTCGGAGATGTGTAT AAGAGACAGTTGGAAGAGAAGGAT CTGCTGAGGC (SEQ ID NO: 361)	TCGTCGGCAGCGTCAGATGTGTATA AGAGACAGCACCTGTCAAAAGTCCT ACTCCGG (SEQ ID NO: 363)

[0231] RNA-Seq和数据分析

[0232] 使用TRIzol Plus RNA纯化试剂盒 (Life Technologies) 根据制造商的方案,从细胞中提取RNA并用PureLink柱上DNA酶试剂盒 (Life Technologies) 进行处理。使用NanoDrop分光光度计测定RNA质量和浓度。使用SuperScript III逆转录酶 (Life Technologies)、随机引物Mix (New England Biolabs) 和RNase OUT (Invitrogen) 将RNA样品逆转录。使用Illumina文库制备试剂盒制备 (Library Prep试剂盒) 对文库进行测序,1μg输入量的总RNA开始并遵循制造商的建议进行。PCR扩增后,在2%琼脂糖凝胶 (E-Gel EX, Invitrogen) 上对文库以300+/-25bp的大小进行选择,并在Illumina HiSeq 2000仪器上进行单末端测序。以两个生物学重复进行RNA-Seq实验。

[0233] 使用TopHat2 (Kim等, Genome Biol (2013) 14:R36) 和Bowtie (Langmead等, Genome Biol (2009)) 将cDNA片段的原始单端读取与人转录组 (RefSeq, hg19) 10:R25) 进行比对。使用具有偏差校正选项的Cuffdiff2 (Trapnell等, Nat Biotechnol (2013) 31:46-53) 对样品之间差异表达的基因进行调用,这掩蔽了定位到线粒体和核糖体RNA转录本处的读取。如果基因在至少一个测试条件下达到最少0.1个片段每百万次读取 (FPKM), 绝对log₂倍数变化为至少0.5,且多重假设校正后的P值 (Q值) 小于0.05,则称基因为差异表达。使用MSigDB数据库 (broadinstitute.org/gsea/index.jsp) (Subramanian等, Proc Natl Acad Sci USA (2005) 102:15545-50) 进行基因集富集分析。

[0234] 在混合群体中细胞增殖的数学建模

[0235] 为了估计在我们的合并筛选中不同参数如何影响结果,我们在含有gRNA组合的混合细胞群体中模拟了显示不同生长速率的细胞增殖。在给定时间 (t), 群体中的总细胞计数 (C_N) 由含有不同组合的各个细胞计数的总和表示 (C_i ; 1至N, 其中N是组合的总数), 从而得到

$$C_N(t) = \sum_{i=1}^N C_i(t).$$

[0236] 对于每个单独的gRNA组合, 细胞生长由等式 (1) 表示。基于指数细胞生长模型, 每个gRNA组合的细胞由两个群体组成: 一个由于CRISPR-Cas9系统的基因破坏具有修饰的生长速率 (k_m), 另一个 (未修饰的细胞) 具有野生型生长速率 (k_{wt})。前面的群体被定义为细胞分数, p, 其受到CRISPR-Cas9系统的切割效率的限制。为了简单起见, 我们假设整个测定期间p是恒定的。

$$[0237] \quad C_i(t) = p C_i(t) + (1-p) C_i(t) = p C_0 e^{k_m t} + (1-p) C_0 e^{k_{wt} t} \quad (\text{Eq. 1})$$

[0238] 其中p表示具有修饰的生长速率的突变细胞的分数, C_0 表示携带相同条形码化gRNA组合之细胞的初始数目。根据等式2由细胞的倍增时间 ($T_{\text{倍增}}$) 评估细胞生长速率 (k)。野生型OVCAR8-ADR-Cas9细胞的倍增时间经实验测定为约24小时 (数据未示出)。

$$[0239] \quad k = \ln 2 / T_{\text{倍增}} \quad (\text{Eq. 2})$$

[0240] 为了简化建模,我们将总细胞群体分为具有不同生长表型的三个亚群,如等式(3)所述。

$$[0241] \quad C_N(t) = \sum_{i=1}^N C_i(t) \approx f_{wt} N \times C_0 e^{k_{wt} t} + f_s N \times C_{i,s}(t) + f_f N \times C_{i,f}(t) \quad (\text{Eq. 3})$$

[0242] 其中 $C_{i,慢}(t)$ 和 $C_{i,快}(t)$ 分别表示通过等式1确定的具有抗增殖性gRNA和促增殖性gRNA的细胞的平均生长谱。在实验开始时,表现为野生型或含有抗增殖性gRNA和促增殖性gRNA的总体群体的百分比分别表示为 f_{wt} 、 f_s 和 f_f 。

[0243] 基于这种混合细胞生长模型,我们模拟了在整个群体中促增殖gRNA和抗增殖gRNA之表现的相对频率(relative frequency, R.F)。相对频率被定义为与初始时间点相比,给定时间的条形码丰度(即, $R.F. = \frac{F(t)}{F(t=0)}$, 其中 $F = C_i(t) / C_N(t)$)。库中的组合总数N和初始的细胞数 C_0 不影响相对频率的结果。在用定义参数运行模拟之后,我们分别观察到群体中促增殖性gRNA和抗增殖性gRNA的富集和消耗(图17A)。观察到富集和消耗程度随着初始gRNA组合的不同百分比(即2、5或10%)而变化,其定义为具有抗增殖(f_s)和促增殖(f_f)响应(图17A)。我们通过调节被修饰细胞的倍增时间($T_{倍增,m}$)和具有对于抗增殖性gRNA的被修饰生长速率(p)的细胞的分数来进一步评估抗增殖性gRNA表现的相对频率。假定在整个实验中p保持恒定,在图17B-17C所示的参数范围内,抗增殖克隆在整个细胞群体中的表现可降低约23%至97%。该模型通过将细胞群体分成具有平均生长速率的亚群体而代表了细胞生长动力学的简化版本,并且其不考虑细胞之间的潜在相互作用。基于我们的模型,可利用提高gRNA效率来提高筛选的灵敏度,从而通过提高测定时间来提高具有修饰生长速率的细胞的分数。

[0244] 因此,已经描述了本发明的至少一个实施方案的几个方面,应理解,本领域技术人员将容易进行多种改变、修改和改进。这些改变、修改和改进旨在成为本公开内容的一部分,并且旨在属于本发明的精神和范围内。因此,上述说明和附图仅作为示例。

[0245] 参考文献

[0246] 1.Shalem,O.et al.Genome-scale CRISPR-Cas9 knockout screening in human cells.Science 343,84-7(2014) .

[0247] 2.Wang,T.,Wei,J.J.,Sabatini,D.M.&Lander,E.S.Genetic screens in human cells using the CRISPR-Cas9 system.Science 343,80-4(2014) .

[0248] 3.Zhou,Y.et al.High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells.Nature 509,487-491(2014) .

[0249] 4.Koike-Yusa,H.,Li,Y.,Tan,E.-P.,Velasco-Herrera,M.D.C.&Yusa,K.Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library.Nat.Biotechnol.32,267-73(2014) .

[0250] 5.Gilbert,L.A.et al.Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.Cell 159,647-661(2014) .

[0251] 6.Konermann,S.et al.Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.Nature doi:10.1038/nature14136(2014) .

[0252] 7.Cong,L.et al.Multiplex genome engineering using CRISPR/Cas systems.Science 339,819-23(2013) .

- [0253] 8.Mali,P.et al.RNA-guided human genome engineering via Cas9.Science 339,823-6(2013) .
- [0254] 9.Gilbert,L.A.et al.CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.Cell 154,442-451(2013) .
- [0255] 10.Cheng,A.A.,Ding,H.&Lu,T.K.Enhanced killing of antibiotic-resistant bacteria enabled by massively parallel combinatorial genetics.Proc.Natl.Acad.Sci.111,12462-7(2014) .
- [0256] 11.Honma,K.et al.RPN2 gene confers docetaxel resistance in breast cancer.Nat.Med.14,939-948(2008) .
- [0257] 12.Wang,Z.et al.Combinatorial patterns of histone acetylations and methylations in the human genome.Nat.Genet.40,897-903(2008) .
- [0258] 13.Dawson,M.A.&Kouzaides,T.Cancer epigenetics:From mechanism to therapy.Cell 150,12-27(2012) .
- [0259] 14.Juergens,R.A.et al.Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer.Cancer Discov.1, 598-607(2011) .
- [0260] 15.Jones,P.A.&Baylin,S.B.The Epigenomics of Cancer.Cell 128,683-692(2007) .
- [0261] 16.Yoo,C.B.&Jones,P.A.Epigenetic therapy of cancer:past,present and future.Nat.Rev.Drug Discov.5,37-50(2006) .
- [0262] 17.Jin,C.et al.Chem-seq permits identification of genomic targets of drugs against androgen receptor regulation selected by functional phenotypic screens.Proc.Natl.Acad.Sci.U.S.A.111,9235-40(2014) .
- [0263] 18.Asangani,I.a et al.Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer.Nature 510,278-82(2014) .
- [0264] 19.Kruidenier,L.et al.A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response.Nature 488,404-408(2012) .
- [0265] 20.Bliss,C.I.THE TOXICITY OF POISONS APPLIED JOINTLY1.Ann.Appl.Biol.26,585-615(1939) .
- [0266] 21.Borisy,A.A.et al.Systematic discovery of multicomponent therapeutics.Proc.Natl.Acad.Sci.U.S.A.100,7977-7982(2003) .
- [0267] 22.Pattanayak,V.et al.High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity.Nat.Biotechnol.31, 839-43(2013) .
- [0268] 23.Kuscu,C.,Arslan,S.,Singh,R.,Thorpe,J.&Adli,M.Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease.Nat.Biotechnol.32,677-683(2014) .
- [0269] 24.Wu,X.et al.Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells.Nat.Biotechnol.32,670-676(2014) .

- [0270] 25.Doench,J.G.et al.Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation.Nat.Biotechnol.32,1262-1267 (2014) .
- [0271] 26.Essletzbichler,P.et al.Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line.Genome Res.24,2059-2065 (2014) .
- [0272] 27.Blasco,R.B.et al.Simple and rapid in vivo generation of chromosomal rearrangements using CRISPR/Cas9 technology.Cell Rep.9,1219-1227 (2014) .
- [0273] 28.Choi,P.S.&Meyerson,M.Targeted genomic rearrangements using CRISPR/Cas technology.Nat.Comm.5,3728 (2014) .
- [0274] 29.Moffat,J.et al.A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen.Cell 124,1283-1298 (2006) .
- [0275] 30.Ran,F.A.et al.Genome engineering using the CRISPR-Cas9 system.Nat.Protoc.8,2281-308 (2013) .
- [0276] 31.Kim D,et al. (2013) TopHat2:accurate alignment of transcriptomes in the presence of insertions,deletions and gene fusions.Genome Biol 14:R36.
- [0277] 32.Langmead B,Trapnell C,Pop M,Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.Genome Biol 10:R25.
- [0278] 33.Trapnell C,et al. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq.Nat Biotechnol 31:46-53.
- [0279] 34.Subramanian A,et al. (2005) Gene set enrichment analysis:a knowledge-based approach for interpreting genome-wide expression profiles.Proc Natl Acad Sci USA 102:15545-50.
- [0280] 35.Stemmer M (2015) CCTop:An Intuitive,Flexible and Reliable CRISPR/Cas9 Target Prediction Tool.PLoS One 10(4):e0124633.

[0281] 等同方案

[0282] 尽管本文已经描述和说明了多个本发明的实施方案,但是本领域普通技术人员将容易地想到用于实现本文所述功能和/或获得本文所述结果和/或一个或更多个优点的多种其他手段和/或结构,并且每种这些变化和/或修改均被认为在本文所述的本发明实施方案的范围内。更普遍地,本领域技术人员将容易理解,本文描述的所有参数、尺寸、材料和构型意在示例,而实际参数、尺寸、材料和/或构型将取决于本发明教导所使用的具体应用。本领域技术人员将认识或者能够仅使用常规实验就能确定针对本文描述的具体发明实施方案的许多等同方案。因此,应理解,前述实施方案仅以示例的方式呈现,并且在所附权利要求及其等同方案的范围内,本发明的实施方案可以以不同于具体描述并且要求保护的方式来实践。本公开内容的本发明实施方案涉及本文描述的每个单独特征、系统、物件、材料、试剂盒和/或方法。另外,如果这些特征、系统、物件、材料、试剂盒和/或方法不互相矛盾,则两种或更多种这样的特征、系统、物件、材料、试剂盒和/或方法的任意组合也包括在本公开内容的发明范围内。

[0283] 本文定义和使用的所有定义应理解为统括所限定术语在字典中的定义、通过引用

并入的文献中的定义,和/或其通常含义。

[0284] 本文在说明书和权利要求书中使用的未用数量词限定的名词应理解为意指“至少一个”,除非明确提到相反情况。

[0285] 本文在说明书和权利要求书中使用的短语“和/或”应理解为意指所连接要素的“两者之一或两者”,即在一些情况下要素相结合地存在,而在另一些情况下要素分开存在。用“和/或”列举的多个要素应以相同方式理解,即所连接要素中的“一个或更多个”。除“和/或”从句明确指出的要素之外,其他要素可任选地存在,无论明确指出的那些要素是否相关。因此,作为一个非限制性实例,当与开放式语言例如“包含”联合使用时,表达“A和/或B”在一个实施方案中可指仅有A(任选地包括除B以外的元素);在另一个实施方案中可指仅有B(任选地包括除A以外的元素);而在另一个实施方案中可指A和B二者(任选地包括其他要素)等。

[0286] 本文在说明书和权利要求书中使用的“或”应理解为具有与上文定义的“和/或”相同的含义。例如,当在列表中分列项目时,“或”或者“和/或”应理解为包括性的,即不但包括许多要素或要素列表中的至少一个,而且也包括超过一个,并且任选地包括额外未列出的要素。只有明确表示相反含义的术语,例如“仅一个”或“正好一个”或者当在权利要求中使用“由……组成”时,将指包含许多要素或要素列表中的正好一个要素。一般来说,当前面有其他排他性术语(例如“任一”、“其一”、“仅其一”或“恰好其一”)时,本文使用的术语“或”仅应理解为排他性备选方案(即,“一个或另一个,但并非二者”)。当权利要求中使用“基本上由……组成”时,应具有专利法领域中所使用的通常含义。

[0287] 当提到一个或更多个要素的列表时,本文在说明书和权利要求书中使用的短语“至少一个”应理解为意指选自要素列表中的任意一个或更多个要素的至少一个要素,但未必包括要素列表中明确列出的每一个和每个要素的至少一个,并且不排除要素列表中要素的任意组合。该定义还允许,除词语“至少一个”所指要素列表中明确指出的要素之外,任选地存在的要素,不论其与明确指出的那些要素是否相关。因此,作为一个非限制性实例,“A和B中的至少一个”(或者等同于“A或B中的至少一个”,或者等同于“A和/或B中的至少一个”)在一个实施方案中可指至少一个A、任选地包括超过一个A,但是不存在B(并且任选地包括除B以外的要素);在另一个实施方案中,可指至少一个B、任选地包括超过一个B,但是不存在A(并且任选地包括除A以外的要素);在另一个实施方案中,可指至少一个A、任选地包括超过一个A,以及至少一个B、任选地包括超过一个B(并且任选地包括其他要素)等。

[0288] 应理解的是,除非明确地指出相反情况,否则在本文要求保护的包括多于一个步骤或动作的任何方法中,所述方法的步骤或动作的顺序未必限于所记载之所述方法的步骤或动作的顺序。

[0289] 本文公开的所有参考文献、专利和专利申请均关于其被引用的各自主题通过引用并入,在一些情况下,其可涵盖文件的整个内容。

[0290] 在权利要求中以及前述说明书中,所有过渡性短语(例如“包括”、“包含”、“带有”、“具有”、“含有”、“涉及”、“持有”、“构成”等)应理解为开放性的,即,意指包括但不限于。仅过渡性短语“由……组成”和“本质上由……组成”分别是封闭式或半封闭式过渡性短语,如美国专利局专利审查程序手册(United States Patent Office Manual of Patent Examining Procedures)2111.03部分所述。

	<110> 麻省理工学院	
	<120> 用于CRISPRR的大规模并行组合遗传学	
	<130> M0656.70342W000	
	<140> PCT/US2015/058304	
	<141> 2015-10-30	
	<150> US 62/166,302	
	<151> 2015-05-26	
	<150> US 62/073,126	
	<151> 2014-10-31	
	<160> 387	
	<170> PatentIn version 3.5	
	<210> 1	
	<211> 47	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 1	
	aatgatacgg cgaccaccga gatctacacg gatccgcaac ggaattc	47
	<210> 2	
	<211> 52	
	<212> DNA	
	<213> 人工序列	
[0001]	<220>	
	<223> 合成的多核苷酸	
	<220>	
	<221> misc_feature	
	<222> (25)..(32)	
	<223> n为a, c, g或t	
	<400> 2	
	caagcagaag acggcatacag agatnnnnnn nnggttgcgt cagcaaacac ag	52
	<210> 3	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 3	
	aatgatacgg cgaccaccga	20
	<210> 4	
	<211> 21	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 4	
	caagcagaag acggcatacag a	21
	<210> 5	

[0002]	<211>	34	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	5	
		ccaccgagat ctacacggat ccgcaacgga attc	34
	<210>	6	
	<211>	35	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	6	
		gtggcgtggt gtgcactgtg ttgctgacg caacc	35
	<210>	7	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	7	
		gggcgaggag ctgttcaccg	20
	<210>	8	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	8	
		cacccagacc atgaagatca	20
	<210>	9	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	9	
		ccacttcaag tgcacatccg	20
	<210>	10	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	10	
		atcgtttccg cttacggcg	20
	<210>	11	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	

	<220>		
	<223>	合成的多核苷酸	
	<400>	11	
		aaacggtacg acagcgtgtg	20
	<210>	12	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	12	
		ccatcaccga tcgtgagcct	20
	<210>	13	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	13	
		ctagacgtcc attcacttcc	20
	<210>	14	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
[0003]	<220>		
	<223>	合成的多核苷酸	
	<400>	14	
		tttccaaacc tcgcacgccc	20
	<210>	15	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	15	
		acgtaaagaa gaattatccg	20
	<210>	16	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	16	
		ccgccccacc ttccgtgccg	20
	<210>	17	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	

	<400> 17 tggcgctcct ccttgccacg	20
	<210> 18 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 18 ccgctccgca gcagagctgc	20
	<210> 19 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 19 agagtcgcga gcttgatctt	20
	<210> 20 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
[0004]	<400> 20 atccgcaccc cggagatcag	20
	<210> 21 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 21 gaagactcga tcctcgtcaa	20
	<210> 22 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 22 aggatctgc gcccatgta	20
	<210> 23 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 23 actcaccttc tatatttcgc	20

	<210> 24	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 24	
	caccaaaatc acgtccatgc	20
	<210> 25	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 25	
	tcacccgtag gcaacgtcgc	20
	<210> 26	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 26	
	gtgtccagcg acgttccta	20
[0005]	<210> 27	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 27	
	acgttgtgca aagactgtcg	20
	<210> 28	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 28	
	cggcgactcc gccatagagc	20
	<210> 29	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 29	
	ggagccggtc cctttccgt	20
	<210> 30	

[0006]	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 30		
	agtcttgaaa gcgcattgcc	20	
	<210> 31		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 31		
	agcggctcta gtatcaaccc	20	
	<210> 32		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 32		
	gaatcacatg acgcattgtc	20	
	<210> 33		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 33		
	cccgcaaatg actggtcacg	20	
	<210> 34		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 34		
	ctgtcagaat tgctgcgac	20	
	<210> 35		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 35		
	cttggcaaga cttgcctgac	20	
	<210> 36		
	<211> 20		
	<212> DNA		
	<213> 人工序列		

	<220>		
	<223>	合成的多核苷酸	
	<400>	36	
		tagttccct aacctcaata	20
	<210>	37	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	37	
		acaccattcg taacgttgcc	20
	<210>	38	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	38	
		tcactcgaga tgcgcttgtc	20
	<210>	39	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
[0007]	<220>		
	<223>	合成的多核苷酸	
	<400>	39	
		agaatgctgc cgcacgcacc	20
	<210>	40	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	40	
		tccgtaatgt tgctcgatgt	20
	<210>	41	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	41	
		tccaacatcg agcaacatta	20
	<210>	42	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	

	<400> 42 tacaacagat cgtgtaatga	20
	<210> 43 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 43 gttgactgtg aagctgtacg	20
	<210> 44 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 44 aacaggttgc gggaaatccaa	20
	<210> 45 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
[0008]	<400> 45 taccagaac atagacacgc	20
	<210> 46 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 46 ctgcgcgttc gcgtgcgcga	20
	<210> 47 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 47 cccgtactca cggtgtaga	20
	<210> 48 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 48 gggccacgta ccgttgggtg	20

	<210> 49	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 49	
	acaggcttca ttgactgaac	20
	<210> 50	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 50	
	agctgatcaa taactatgat	20
	<210> 51	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 51	
	cctcatctga gtactgattc	20
[0009]	<210> 52	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 52	
	acacgcttcc gccacaac	20
	<210> 53	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 53	
	tgcgactgag acagctcaag	20
	<210> 54	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 54	
	aaaacttcac ctcccatata	20
	<210> 55	

[0010]	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 55		
	gtacaaattg tacgacggag	20	
	<210> 56		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 56		
	gacccctcgg cggtttatag	20	
	<210> 57		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 57		
	tatattgcga ccaccaaact	20	
[0010]	<210> 58		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 58		
	cagagagcac aacgccgcac	20	
	<210> 59		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 59		
	ggaaccgctg gcagtcgcgc	20	
	<210> 60		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 60		
	ctcccgtgc ccgtgtagac	20	
	<210> 61		
	<211> 20		
	<212> DNA		
	<213> 人工序列		

	<220>		
	<223>	合成的多核苷酸	
	<400>	61	
		ctggctcgag atttagcgca	20
	<210>	62	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	62	
		aatctgtica tgcgcttacg	20
	<210>	63	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	63	
		gattccagta ccagtacatt	20
	<210>	64	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
[0011]	<220>		
	<223>	合成的多核苷酸	
	<400>	64	
		ctcacgtgg tctaacttgt	20
	<210>	65	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	65	
		ggatgtcatg tcctcagcgt	20
	<210>	66	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	66	
		tttgactcct acgcacatt	20
	<210>	67	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	

	<p><400> 67 cgtggatgag tacgaccccg</p>	20
	<p><210> 68 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 68 tcttctgtgc acactatgcg</p>	20
	<p><210> 69 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 69 ctgtcccaga agtgaatcgc</p>	20
	<p><210> 70 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
[0012]	<p><400> 70 gccatgtgct cgttagcgtc</p>	20
	<p><210> 71 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 71 gcctgacgct aacgagcaca</p>	20
	<p><210> 72 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 72 gaattcatgt actcaactgt</p>	20
	<p><210> 73 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 73 cggaatgcgg ggtccgaact</p>	20

	<210> 74	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 74	
	cagcatacag ctttatccgc	20
	<210> 75	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 75	
	atgaactccc tcttgaaacg	20
	<210> 76	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 76	
	attgtccggc gaggacgtgc	20
[0013]	<210> 77	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 77	
	cttcgccacg cgetgtctca	20
	<210> 78	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 78	
	gacggtactg gacgtgggcg	20
	<210> 79	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 79	
	caatccgacc acggggtctg	20
	<210> 80	

[0014]	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	80	
		gaggttcaaa ccgcctgcta	20
	<210>	81	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	81	
		taaagtcggc tggtagacacc	20
	<210>	82	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	82	
		agttcttctc ggtgtccaaa	20
[0014]	<210>	83	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	83	
		gactatcagt tccagagata	20
	<210>	84	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	84	
		aacttacgaa ggaaggcttt	20
	<210>	85	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	85	
		ttaccttcgc ccgcttgccg	20
	<210>	86	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	

	<220>		
	<223>	合成的多核苷酸	
	<400>	86	
		ccggccctac tgcgtgcct	20
	<210>	87	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	87	
		agagccgact tcctcatgac	20
	<210>	88	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	88	
		cataccgcat cgataagtct	20
	<210>	89	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
[0015]	<220>		
	<223>	合成的多核苷酸	
	<400>	89	
		atagccaaga cttatcgatg	20
	<210>	90	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	90	
		gaacatacct tctgtagtaa	20
	<210>	91	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	91	
		acgtactat gagaccccag	20
	<210>	92	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	

	<p><400> 92 tatggcaggg agtcgtcgca</p>	20
	<p><210> 93 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 93 gtaacgaatc ctttcttctt</p>	20
	<p><210> 94 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 94 cctcgttctc gtcgtatcgc</p>	20
	<p><210> 95 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
[0016]	<p><400> 95 gcgttactac gagacgcccg</p>	20
	<p><210> 96 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 96 cttggtaag cgtccgactg</p>	20
	<p><210> 97 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 97 taaatgccga gagtgtcgct</p>	20
	<p><210> 98 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 98 gtctgtcaaa accgacttcc</p>	20

	<210> 99	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 99	
	gatactgctt ggctgtactg	20
	<210> 100	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 100	
	tcttgtatgg gcgccccgtg	20
	<210> 101	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 101	
	gccttgactg ttaccggctc	20
[0017]	<210> 102	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 102	
	tcctgagccg gtaacagtca	20
	<210> 103	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 103	
	actccgcaca gttaaaacca	20
	<210> 104	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 104	
	gcggaactct cgaacagtca	20
	<210> 105	

[0018]	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 105		
	ttccactcac ttatcgtat	20	
	<210> 106		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 106		
	ccccgcgtac ttctcgtgt	20	
	<210> 107		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 107		
	gtatgatgac atcgacgacg	20	
	<210> 108		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 108		
	tcaccaggtg ctgtaccccg	20	
	<210> 109		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 109		
	cctttgcaag acccgcacga	20	
	<210> 110		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 110		
	agtaggcttc gtgtgatcaa	20	
	<210> 111		
	<211> 20		
	<212> DNA		
	<213> 人工序列		

	<220> <223> 合成的多核苷酸	
	<400> 111 gtctaaagga gcccatcgtg	20
	<210> 112 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 112 catgaacccc aacgtgctaa	20
	<210> 113 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 113 ctgggattca aataactcgg	20
	<210> 114 <211> 20 <212> DNA <213> 人工序列	
[0019]	<220> <223> 合成的多核苷酸	
	<400> 114 tctctggtat gaaagtgccg	20
	<210> 115 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 115 gtccgcgaac tcttcccagc	20
	<210> 116 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 116 tcgaagaccg ggcactcggg	20
	<210> 117 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	

	<400> 117 ggacttatatt cagcttaata	20
	<210> 118 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 118 cttaccgcca tgacacactt	20
	<210> 119 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 119 gataaacaat gcgttcgtag	20
	<210> 120 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
[0020]	<400> 120 gggctaccgg agcccaccga	20
	<210> 121 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 121 gatttactec tcggtccaa	20
	<210> 122 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 122 aaagacttac cgcgggtggg	20
	<210> 123 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 123 taaggcccga catggaaccg	20

	<210> 124	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 124	
	actgtaaact gtagtacctc	20
	<210> 125	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 125	
	cagcattatc tgcataccag	20
	<210> 126	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 126	
	agactatgag tctagtttaa	20
[0021]	<210> 127	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 127	
	taccacagcg cccttcgata	20
	<210> 128	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 128	
	atccccctcc tcgtagegca	20
	<210> 129	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 129	
	caaaggcttc ccgtgcagcg	20
	<210> 130	

[0022]	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	130	
		ttctgcacgg gcttgacgtc	20
	<210>	131	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	131	
		gacgtcaagc ccgtgcagaa	20
	<210>	132	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	132	
		cagtgacgtc gagaactacg	20
	<210>	133	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	133	
		tctgacgaac gtagggtcc	20
	<210>	134	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	134	
		ggcttagtga aaaaacgccg	20
	<210>	135	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	135	
		cctcgccatc attcactgtg	20
	<210>	136	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	

	<220>		
	<223>	合成的多核苷酸	
	<400>	136	
		aacgtgtatt gttcgttacc	20
	<210>	137	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	137	
		tcctacctta tattcagtag	20
	<210>	138	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	138	
		aaaggtttac catcagcaga	20
	<210>	139	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
[0023]	<220>		
	<223>	合成的多核苷酸	
	<400>	139	
		caccgtgttc tatagagccg	20
	<210>	140	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	140	
		cggcgcgagg tggacagcat	20
	<210>	141	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	141	
		cgactcaccg gctgcgatcc	20
	<210>	142	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	

	<p><400> 142 cgacgtgacg ttgcagtg</p>	20
	<p><210> 143 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 143 caaaggtcgg aagccgctg</p>	20
	<p><210> 144 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 144 catcactgca aacgtcacgt</p>	20
	<p><210> 145 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
[0024]	<p><400> 145 actagcatgt ctgcggagag</p>	20
	<p><210> 146 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 146 tctagtccat cccccattac</p>	20
	<p><210> 147 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 147 gggaacaata aagaagcgt</p>	20
	<p><210> 148 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 148 gagatcgacg cgaaatacca</p>	20

	<210> 149	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 149	
	tataaatccg cgcccgaag	20
	<210> 150	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 150	
	cccataccag ttattgcgt	20
	<210> 151	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 151	
	gcagcagcaa ctgtactcgt	20
[0025]	<210> 152	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 152	
	gcagcgactc cagcactca	20
	<210> 153	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 153	
	gatcttcaag aagacccgc	20
	<210> 154	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 154	
	tctcgcgat ttccgtgaag	20
	<210> 155	

[0026]	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 155		
	cttcacggaa atgcgcgaga	20	
	<210> 156		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 156		
	tcgatactgc atttgaatc	20	
	<210> 157		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 157		
	gctgctcgtg ctcgttccaa	20	
	<210> 158		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 158		
	cctagaaggc cggactcaaa	20	
	<210> 159		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 159		
	ggcactactc atatactcag	20	
	<210> 160		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 160		
	gatctgcttc aaagcgcgcc	20	
	<210> 161		
	<211> 20		
	<212> DNA		
	<213> 人工序列		

	<220> <223> 合成的多核苷酸	
	<400> 161 cttcagctg atgcgagagc	20
	<210> 162 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 162 gaagttcctc tgaagttcgc	20
	<210> 163 <211> 21 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 163 cgagggcgac ttaaccttag g	21
	<210> 164 <211> 21 <212> DNA <213> 人工序列	
[0027]	<220> <223> 合成的多核苷酸	
	<400> 164 aaatcttcgt aatccaagta t	21
	<210> 165 <211> 21 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 165 gtaataccgg gtgttcgat g	21
	<210> 166 <211> 21 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 166 attaatccac acgaggtctc c	21
	<210> 167 <211> 21 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	

	<400> 167 tatagtaatc agggaggttc a	21
	<210> 168 <211> 21 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 168 tttagacttg attgtgctca t	21
	<210> 169 <211> 23 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 169 agaaattaaa cggtaccct cca	23
	<210> 170 <211> 22 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
[0028]	<400> 170 tccatagtgt cttgagcacc ac	22
	<210> 171 <211> 22 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 171 cacttgctga tgccagtagg ag	22
	<210> 172 <211> 22 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 172 gtgaatagct tggaatgtg gg	22
	<210> 173 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 173 gccacactct acatgggagc	20

	<210> 174	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 174	
	gacttttcca tcagggacac ct	22
	<210> 175	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 175	
	ggttgacaaa cacattcggc	20
	<210> 176	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 176	
	cctatcattg cccaaggag tc	22
[0029]	<210> 177	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 177	
	ctaggcctcc gacagttgta at	22
	<210> 178	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 178	
	agccaccctt ggttggtttt	20
	<210> 179	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 179	
	ggtaaggga actctggggc	20
	<210> 180	

[0030]	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	180	
		ctccctccct tcctaaggct	20
	<210>	181	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	181	
		attgccttaa gtcgacacct gat	23
	<210>	182	
	<211>	22	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	182	
		gactgtagag ttgccggaac ag	22
	<210>	183	
	<211>	22	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	183	
		gttggtagaa agtggtagaag gc	22
	<210>	184	
	<211>	22	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	184	
		acgtggcttc attgtacatc ct	22
	<210>	185	
	<211>	22	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	185	
		aagcacatgc ttcaggctaa ca	22
	<210>	186	
	<211>	22	
	<212>	DNA	
	<213>	人工序列	

	<220>		
	<223>	合成的多核苷酸	
	<400>	186	
		tcatctgctc ttacgcttag cc	22
	<210>	187	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	187	
		ctcggcaacc ctccatacat	20
	<210>	188	
	<211>	24	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	188	
		aacctatgcac agaattccaga tttta	24
	<210>	189	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
[0031]	<220>		
	<223>	合成的多核苷酸	
	<400>	189	
		aagagttctt ggcgcagaca	20
	<210>	190	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	190	
		tcgtccaagt tacagtcac aca	23
	<210>	191	
	<211>	22	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	191	
		tcctctggtg cacagaaaag tc	22
	<210>	192	
	<211>	20	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	

	<p><400> 192 ttctctccag acactgccct</p>	20
	<p><210> 193 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 193 gtgccccagaa ctactgccat</p>	20
	<p><210> 194 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 194 gaggtgagtt ccagcttccc</p>	20
	<p><210> 195 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
[0032]	<p><400> 195 caccttacag gcactgcgtt</p>	20
	<p><210> 196 <211> 22 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 196 ctccctccct agaggctatg ag</p>	22
	<p><210> 197 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 197 gttgatgtga ttgccgctc</p>	20
	<p><210> 198 <211> 20 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 198 cgtacgggtt catgcaagtt</p>	20

	<210> 199	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 199	
	ccccctcaccg cctatcagta	20
	<210> 200	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 200	
	ttaggcagga cctgtttcgg	20
	<210> 201	
	<211> 21	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 201	
	cgtttgctta agagcacctc c	21
[0033]	<210> 202	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 202	
	tcttgaacaa gtcggggtcg	20
	<210> 203	
	<211> 61	
	<212> DNA	
	<213> 人(Homo sapiens)	
	<400> 203	
	tttcaacagt ttcctacctt atatcagtag tggctctggtc ttgtgaactt ggacatcaca	60
	a	61
	<210> 204	
	<211> 62	
	<212> DNA	
	<213> 人	
	<400> 204	
	tgactcacia tctctcatgt ttttctggct tagtgaaaaa acgccgtgga tcttcaaagg	60
	gg	62
	<210> 205	
	<211> 44	
	<212> DNA	

[0034]	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	205	
		tttcaacagt ttcctacctt atatgtgaa cttggacatc acaa	44
	<210>	206	
	<211>	63	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	206	
		tgactcacia tctctcatgt tttctggct tagtgaaaa acggccgtgg atcttcaaag	60
		ggg	63
	<210>	207	
	<211>	59	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	207	
		tgactcacia tctctcatgt tttctggct tagtgaaaa acgtggatct tcaaagggg	59
[0034]	<210>	208	
	<211>	59	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	208	
		tttcaacagt ttcctacctt atattcagtg gtctggcttt gtgaacttgg acatcacia	59
	<210>	209	
	<211>	13	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	209	
		ttcaacatca caa	13
[0034]	<210>	210	
	<211>	59	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	210	
		tgactcacia tctctcatgt tttctggct tagtgaaaa acgtggatct tcaaagggg	59
	<210>	211	
	<211>	60	
	<212>	DNA	
	<213>	人	

	<400> 211 tagacatttg ggaagtttct agtccatccc ccattactgg cagattttctc aatctcgtcc	60
	<210> 212 <211> 59 <212> DNA <213> 人	
	<400> 212 gaaggataat caccttttgca agacccgcac gatgggctcc tttagactcc atgtatgca	59
	<210> 213 <211> 57 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 213 gaaggataat caccttttgca agacccacga tgggctcctt tagactccat gtatgca	57
	<210> 214 <211> 57 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 214 gaaggataat caccttttgca agacccgcga tgggctcctt tagactccat gtatgca	57
[0035]	<210> 215 <211> 52 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 215 tagacatttg ggaagtttct agtccatcct ggcagatttc tcaatctcgt cc	52
	<210> 216 <211> 58 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 216 gaaggataat caccttttgca agacccgcgc atgggctcct ttagactcca tgtatgca	58
	<210> 217 <211> 34 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 217 gaaggataat caccttttgca aatccatgta tgca	34
	<210> 218	

	<211> 50	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 218	
	gaaggataat cacctttgca agatgggctc ctttagactc catgtaigca	50
	<210> 219	
	<211> 49	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 219	
	tttcaacagt ttcttacctt atattcagtt gtgaacttgg acatcacaa	49
	<210> 220	
	<211> 65	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 220	
	tgactcacaa tctctcatgt tttcttggt tagtgaaaa acaaaaccgt ggatcttcaa	60
	agggg	65
[0036]	<210> 221	
	<211> 56	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 221	
	tttcaacagt ttcttacctt attagtggct tggctcttggt aacttggaca tcacaa	56
	<210> 222	
	<211> 52	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 222	
	tttcaacagt ttcttacctt atattctggt ctgtgaact tggacatcac aa	52
	<210> 223	
	<211> 60	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 223	
	tttcaacagt ttcttacctt atattctagt ggtctggtct tgtgaacttg gacatcacaa	60
	<210> 224	
	<211> 55	

<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的多核苷酸	
<400>	224	
	tttcaacagt ttcctacctt ttagtggctt ggtcttgtga acttggacat caciaa	55
<210>	225	
<211>	90	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的多核苷酸	
<400>	225	
	tacaactcca ataataatta caaacaagga atttcaacag tttcctacct tatattcagg	60
	tcttgtgaac ttggacatca caaataggac	90
<210>	226	
<211>	90	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的多核苷酸	
<400>	226	
	cgttctgact cacaatctct catgtttttc tggcttagtg aaaaaacttc aaaggggcat	60
	gatacacaca aggggaaacc agtgaagacc	90
[0037]		
<210>	227	
<211>	91	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的多核苷酸	
<400>	227	
	tacaactcca ataataatta caaacaagga atttcaacag tttcctacct tataggtctg	60
	gtcttgtgaa cttggacatc acaaatagga c	91
<210>	228	
<211>	97	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的多核苷酸	
<400>	228	
	tacaactcca ataataatta caaacaagga atttcaacag tttcctacct tatattcagt	60
	ggtctggctt tgtgaacttg gacatcacia ataggac	97
<210>	229	
<211>	99	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	合成的多核苷酸	
<400>	229	

[0038]	cgttctgact cacaatctct catgtttttc tggcttagtg aaaaacgccg tggatcttca	60
	aaggggcatg atacacacaa ggggaaacca gtgaagacc	99
	<210> 230	
	<211> 95	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 230	
	cgttctgact cacaatctct catgtttttc tggcttagtg aaaaactgga tcttcaaagg	60
	ggcatgatac acacaagggg aaaccagtga agacc	95
	<210> 231	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 231	
	actccaataa taattacaaa caaggaattt caacagtttc ctacattata ttctgtgac	60
	tgtggtctgg tcttgtgaac ttggacatca caaataggac	100
[0039]	<210> 232	
	<211> 97	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 232	
	cgttctgact cacaatctct catgtttttc tggcttagtg aaaaaacgtg gatcttcaaa	60
	ggggcatgat acacacaagg ggaaccagt gaagacc	97
	<210> 233	
	<211> 100	
	<212> DNA	
	<213> 人	
	<400> 233	
	tacaactcca ataataatta caaacaagga attcaacag tticctacct tatattcagt	60
	agtggctctgg tcttgtgaac ttggacatca caaataggac	100
[0040]	<210> 234	
	<211> 100	
	<212> DNA	
	<213> 人	
	<400> 234	
	cgttctgact cacaatctct catgtttttc tggcttagtg aaaaaacgcc gtggatcttc	60
	aaaggggcat gatacacaca aggggaaacc agtgaagacc	100
	<210> 235	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	

	<223> 合成的多核苷酸	
	<400> 235	
	atttcaacag tttcctacct tatatactat actatatata tatatatata ctatatatat	60
	agttggtctgg tcttgtgaac ttggacatca caaataggac	100
	<210> 236	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 236	
	ctctcatgtt tttctggctt agtgaaaaa cttcactaag tttttactta gtggatcttc	60
	aaaggggcat gatacacaca aggggaaacc agtgaagacc	100
	<210> 237	
	<211> 95	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 237	
	tacaactcca ataataatta caaacaagga atttcaacag tttcctacct tatattctgg	60
	tctggtcttg tgaacttgga catcacaaat aggac	95
[0039]	<210> 238	
	<211> 76	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 238	
	tacaactcca ataataatta caaacaagga atttcaacag tttcctactt gtgaacttgg	60
	acatcacaaa taggac	76
	<210> 239	
	<211> 96	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 239	
	cgttctgact cacaatctct catgtttttc tggcttagtg aaaaacatgg atcttcaaag	60
	gggcatgata cacacaaggg gaaaccagtg aagacc	96
	<210> 240	
	<211> 97	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 240	
	cgttctgact cacaatctct catgtttttc tggcttagtg aaaaaacgtg gatcttcaaa	60

	ggggcatgat acacacaagg ggaaccagt gaagacc	97
	<210> 241	
	<211> 90	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 241	
	tacaactcca ataataatta caaacaagga atttcaacag tticctacct agtggctctgg	60
	tcttctgaac ttggacatca caaataggac	90
	<210> 242	
	<211> 97	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 242	
	tacaactcca ataataatta caaacaagga atttcaacag tticctacct tatattctgt	60
	ggctctggctct tgtgaacttg gacatcaca ataggac	97
	<210> 243	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0040]	<223> 合成的多核苷酸	
	<400> 243	
	cgttctgact cacaatctct catgtttttc tggcttagtg aaaaaactaa gtggatcttc	60
	aaaggggcat gatacacaca aggggaaacc agtgaagacc	100
	<210> 244	
	<211> 89	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 244	
	tacaactcca ataataatta caaacaagga atttcaacag tticctacta gttgtctggt	60
	cttgtgaact tggacatcac aaataggac	89
	<210> 245	
	<211> 87	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 245	
	cgttctgact cacaatctct catgtttttc tggcttagtg aatcttcaaa ggggcatgat	60
	acacacaagg ggaaccagt gaagacc	87
	<210> 246	
	<211> 86	

	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 246	
	tacaactcca ataataatta caaacaagga atttcaacag tticcaagtg gtctggtctt	60
	gtgaacttgg acatcacaaa taggac	86
	<210> 247	
	<211> 97	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 247	
	cgttctgact cacaatctct catgtttttc tggcttagtg aaaaaacgtg gatcttcaaa	60
	ggggcatgat acacacaagg ggaaccagt gaagacc	97
	<210> 248	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 248	
	acaactccaa taataattac aaacaaggaa ttccaacagt ttcttacctt atattcagat	60
[0041]	agtggctctgg tcttgtgaac ttggacatca caaataggac	100
	<210> 249	
	<211> 86	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 249	
	tacaactcca ataataatta caaacaagga atttcaacag tticctagtgt gtctggtctt	60
	gtgaacttgg acatcacaaa taggac	86
	<210> 250	
	<211> 97	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 250	
	tacaactcca ataataatta caaacaagga atttcaacag tticctacct tatattctag	60
	ggctctggtct tgtgaacttg gacatcacia ataggac	97
	<210> 251	
	<211> 87	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	

[0042]	<400> 251	
	tacaactcca ataataatta caaacaagga atttcaacag tttctacct tatattcagt	60
	tgtagaacttg gacatcaca ataggac	87
	<210> 252	
	<211> 95	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 252	
	tacaactcca ataataatta caaacaagga atttcaacag tttctacct tatatagtg	60
	tctggctcttg tgaacttgga catcacaaat aggac	95
	<210> 253	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
[0042]	<400> 253	
	tctgactcac aatctctcat gtttttcttg cttagtgaaa aacatgcccc gtggatcttc	60
	aaaggggcat gatacacaca aggggaaacc agtgaagacc	100
	<210> 254	
	<211> 91	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 254	
	tacaactcca ataataatta caaacaagga atttcaacag tttctacct tattggcttg	60
	gtcttgtgaa cttggacatc acaaatagga c	91
	<210> 255	
	<211> 78	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
[0042]	<400> 255	
	tacaactcca ataataatta caaacaagga atttcaacag tggctctggtc ttgtgaactt	60
	ggacatcaca aataggac	78
	<210> 256	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 256	
	tctcatgttt ttctggctta gtgaaatcta agccttagtga aatctaaagcc gtggatcttc	60
	aaaggggcat gatacacaca aggggaaacc agtgaagacc	100

	<210> 257	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 257	
	tacaactcca ataataatta caaacaagga atttcaacag tttcctacct tatattctta	60
	taagggtctgg tcttgtgaac ttggacatca caaataggac	100
	<210> 258	
	<211> 97	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 258	
	tacaactcca ataataatta caaacaagga atttcaacag tttcctacct tatatatagt	60
	ggctctggctct tgtgaacttg gacatcacia ataggac	97
	<210> 259	
	<211> 82	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
[0043]	<400> 259	
	tacaactcca ataataatta caaacaagga atttcaacag tttcctacct tatattcaga	60
	acttggacat cacaatatagg ac	82
	<210> 260	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 260	
	gttctgactc acaatctctc atgtttttct ggcttagtga aaaaaacgcc gtggatcttc	60
	aaaggggcat gatacacaca aggggaaacc agtgaagacc	100
	<210> 261	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 261	
	taattacaaa caaggaattt caacagtctt ctaccittata ttcagtataa tatattcatt	60
	agtggtctgg tcttgtgaac ttggacatca caaataggac	100
	<210> 262	
	<211> 87	
	<212> DNA	

	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 262	
	cgttctgact cacaatctct catgtttttc tggcttagtg gatcttcaaa ggggcatgat	60
	acacacaagg ggaaccagt gaagacc	87
	<210> 263	
	<211> 98	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 263	
	tacaactcca ataataatta caaacaagga atttcaacag tttctacct tatattctag	60
	tggcttggtc ttgtgaactt ggacatcaca aataggac	98
	<210> 264	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 264	
	tacaactcca ataataatta caaacaagga atttcaacag tttctacct tatattatat	60
[0044]	agtggctcgg tcttgtgaac tiggacatca caaataggac	100
	<210> 265	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 265	
	taattacaaa caaggaattt caacagtttc ctaccttata ttcaggtagt gaatctgaat	60
	agtggctcgg tcttgtgaac tiggacatca caaataggac	100
	<210> 266	
	<211> 85	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 266	
	tacaactcca ataataatta caaacaagga atttcaacag tttctacct tctggctctg	60
	tgaacttga catcacaaat aggac	85
	<210> 267	
	<211> 100	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	

[0045]	<400> 267	
	acaatctctc atgtttttct ggcttagtga atcttcaaag ggatcttcaa aaggatcttc	60
	aaaggggcat gatacacaca aggggaaacc agtgaagacc	100
	<210> 268	
	<211> 82	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 268	
	tacaactcca ataataatta caaacaagga atttcaacag tttctacct tatattgga	60
	acttggacat cacaaatagg ac	82
	<210> 269	
	<211> 97	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
[0045]	<400> 269	
	tacaactcca ataataatta caaacaagga atttcaacag tttctacct tatattcagt	60
	ggctctggctct tgtgaacttg gacatcaca ataggac	97
	<210> 270	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 270	
	gttgtgctca gtactgactt	20
	<210> 271	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
[0045]	<400> 271	
	tttcagcttc caataaaaac	20
	<210> 272	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 272	
	attccacggg aaggagatct	20
	<210> 273	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	

	<220> <223> 合成的多核苷酸	
	<400> 273 aggattgaag ctgacgttct	20
	<210> 274 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 274 ggcttagtga aaaaacgccg	20
	<210> 275 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 275 tcctacctta tattcagtag	20
	<210> 276 <211> 20 <212> DNA <213> 人工序列	
[0046]	<220> <223> 合成的多核苷酸	
	<400> 276 gctttgcccg aaccgcacga	20
	<210> 277 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 277 cctagggcag acctgcacga	20
	<210> 278 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 278 gtcttgaag acccgaccca	20
	<210> 279 <211> 20 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	

	<p><400> 279 ccttatcaag acccacacca</p>	20
	<p><210> 280 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 280 atccccctcc tcgtagcgca</p>	20
	<p><210> 281 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 281 ctgctcctcc tcgtagcgct</p>	20
	<p><210> 282 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
[0047]	<p><400> 282 tgcgccctcc tcctagcgca</p>	20
	<p><210> 283 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 283 ctcctcctcc gcgtagcgct</p>	20
	<p><210> 284 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 284 ctgccccctcc tggtagcgcc</p>	20
	<p><210> 285 <211> 20 <212> DNA <213> 人工序列</p>	
	<p><220> <223> 合成的多核苷酸</p>	
	<p><400> 285 aaccagctcc tcgtagctca</p>	20

	<210> 286	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 286	
	accgccctcc tcctagctca	20
	<210> 287	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 287	
	agccccgtcc tcgtggggca	20
	<210> 288	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 288	
	gggaacaata aagaagcgt	20
[0048]	<210> 289	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 289	
	tggaataaaca aagaagcgt	20
	<210> 290	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 290	
	gggaagtata aggaagcgt	20
	<210> 291	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 291	
	gtgagcaata aagcagccct	20
	<210> 292	
	<211> 20	

[0049]	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 292		
	tctagtccat cccccattac	20	
	<210> 293		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 293		
	aatattccat tccccattac	20	
	<210> 294		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 294		
	tgttgtccat acctcattac	20	
	<210> 295		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 295		
	tctaggtcat gcaccattac	20	
	<210> 296		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 296		
	cccattcctt cccccattac	20	
	<210> 297		
	<211> 20		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的多核苷酸		
	<400> 297		
	tccacaccct cccccattac	20	
	<210> 298		
	<211> 59		
	<212> DNA		
	<213> 人工序列		

	<220>		
	<223>	合成的多核苷酸	
	<400>	298	
		acactctttc cctacacgac gctcttccga tctggatatct gtggttgatg cagttttcc	59
	<210>	299	
	<211>	56	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	299	
		acactctttc cctacacgac gctcttccga tctgcacaga gctgctgctt ggagtg	56
	<210>	300	
	<211>	56	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	300	
		acactctttc cctacacgac gctcttccga tctcgctttc ctgcctcagg atgaac	56
	<210>	301	
	<211>	58	
	<212>	DNA	
	<213>	人工序列	
[0050]	<220>		
	<223>	合成的多核苷酸	
	<400>	301	
		acactctttc cctacacgac gctcttccga tctttggaag agaaggatct gctgaggc	58
	<210>	302	
	<211>	56	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	302	
		acactctttc cctacacgac gctcttccga tctgctacc tccacaaagc acacac	56
	<210>	303	
	<211>	58	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	303	
		acactctttc cctacacgac gctcttccga tctccttcag aaacaatgtc ccaaaticg	58
	<210>	304	
	<211>	55	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	

	<400> 304 acactctttc cctacacgac gctcttccga tctagccttt ctgagagcgg gctag	55
	<210> 305 <211> 56 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 305 acactctttc cctacacgac gctcttccga tctgcctcct ctcatcctct cgcttc	56
	<210> 306 <211> 58 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 306 acactctttc cctacacgac gctcttccga tctgattggc tccaagcggc catcaaac	58
	<210> 307 <211> 59 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
[0051]	<400> 307 acactctttc cctacacgac gctcttccga tctttgaact caaggctcag ccaacaggc	59
	<210> 308 <211> 55 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 308 acactctttc cctacacgac gctcttccga tctcaactca ggctggatgc atcgg	55
	<210> 309 <211> 57 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 309 acactctttc cctacacgac gctcttccga tctttggtgg ccgctgagtg tgtgtac	57
	<210> 310 <211> 53 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 310 acactctttc cctacacgac gctcttccga tctcagcatg ttgacatagc ggc	53

	<210> 311	
	<211> 55	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 311	
	acactctttc cctacacgac gctcttccga tcttgagagg agatgagtcg gggtc	55
	<210> 312	
	<211> 56	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 312	
	acactctttc cctacacgac gctcttccga tctgacgggt caaagcctca ggagag	56
	<210> 313	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 313	
	acactctttc cctacacgac gctcttccga tcttttagctg cccagctcac agctacc	57
[0052]	<210> 314	
	<211> 53	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 314	
	acactctttc cctacacgac gctcttccga tctgagcccc aagagcgaga caa	53
	<210> 315	
	<211> 54	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 315	
	acactctttc cctacacgac gctcttccga tctggcgctc agaaggctgt gcag	54
	<210> 316	
	<211> 56	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 316	
	acactctttc cctacacgac gctcttccga tctcctaggt gacactggac ttttgc	56
	<210> 317	
	<211> 58	

	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 317	
	acactctttc cctacacgac gctcttccga tctaggttca cctcaggctg ctcagaag	58
	<210> 318	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 318	
	acactctttc cctacacgac gctcttccga tctcgtctct ctccatgtga gcttgtg	57
	<210> 319	
	<211> 59	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 319	
	acactctttc cctacacgac gctcttccga tctggaaaga tcatctgatc aggcccatc	59
[0053]	<210> 320	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 320	
	acactctttc cctacacgac gctcttccga tctcttgta gggttggagg tctctgg	57
	<210> 321	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 321	
	acactctttc cctacacgac gctcttccga tctaaggccc gtaaaggga agttcag	57
	<210> 322	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 322	
	acactctttc cctacacgac gctcttccga tctggtgaca ggaagctgtc ggaacat	57
	<210> 323	
	<211> 58	
	<212> DNA	
	<213> 人工序列	

	<220>			
	<223>	合成的多核苷酸		
	<400>	323		
		acactctttc cctacacgac gctcttccga tctctcactg tgatctgaca ccaaacac	58	
	<210>	324		
	<211>	58		
	<212>	DNA		
	<213>	人工序列		
	<220>			
	<223>	合成的多核苷酸		
	<400>	324		
		acactctttc cctacacgac gctcttccga tctctagttg ccttcatgcc ttacagac	58	
	<210>	325		
	<211>	55		
	<212>	DNA		
	<213>	人工序列		
	<220>			
	<223>	合成的多核苷酸		
	<400>	325		
		acactctttc cctacacgac gctcttccga tctctgctcc cactccagac taccc	55	
	<210>	326		
	<211>	56		
	<212>	DNA		
	<213>	人工序列		
[0054]	<220>			
	<223>	合成的多核苷酸		
	<400>	326		
		gtgactggag ttacagcgtg tgctcttccg atctagtagt gaggccgctt ataacc	56	
	<210>	327		
	<211>	57		
	<212>	DNA		
	<213>	人工序列		
	<220>			
	<223>	合成的多核苷酸		
	<400>	327		
		gtgactggag ttacagcgtg tgctcttccg atcttaacaa ggagatgccc tggctgg	57	
	<210>	328		
	<211>	57		
	<212>	DNA		
	<213>	人工序列		
	<220>			
	<223>	合成的多核苷酸		
	<400>	328		
		gtgactggag ttacagcgtg tgctcttccg atctaggtca tgaaggcaaa ctcagcc	57	
	<210>	329		
	<211>	58		
	<212>	DNA		
	<213>	人工序列		
	<220>			
	<223>	合成的多核苷酸		

	<400> 329 gtgactggag ttcagacgtg tgctcttccg atctcacctg tcaaaagtcc tactccgg	58
	<210> 330 <211> 58 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 330 gtgactggag ttcagacgtg tgctcttccg atctcctgga gaccagcaag tattgtcc	58
	<210> 331 <211> 57 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 331 gtgactggag ttcagacgtg tgctcttccg atctgtcttc tgaaccccag ctgtaag	57
	<210> 332 <211> 58 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
[0055]	<400> 332 gtgactggag ttcagacgtg tgctcttccg atctcaaaca gaggccaaag ggtgtccc	58
	<210> 333 <211> 58 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 333 gtgactggag ttcagacgtg tgctcttccg atctacagga ggtcgtggtg cagttctc	58
	<210> 334 <211> 59 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 334 gtgactggag ttcagacgtg tgctcttccg atctttgtgt gaggaacgtt gacgctacc	59
	<210> 335 <211> 57 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 335 gtgactggag ttcagacgtg tgctcttccg atctgagcgt aggtcctctg catggag	57

	<210> 336	
	<211> 58	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 336	
	gtgactggag ttcagacgtg tgctcttccg atctccacag aatgacagga acccatgg	58
	<210> 337	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 337	
	gtgactggag ttcagacgtg tgctcttccg atctactgag cagagcctag gaggcag	57
	<210> 338	
	<211> 56	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 338	
	gtgactggag ttcagacgtg tgctcttccg atcttggtgc cagatccaga ggcgtc	56
[0056]	<210> 339	
	<211> 58	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 339	
	gtgactggag ttcagacgtg tgctcttccg atctaactgg cccgagtagt cggagcag	58
	<210> 340	
	<211> 59	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 340	
	gtgactggag ttcagacgtg tgctcttccg atcttcctca gagtgtgtgg aagtgttg	59
	<210> 341	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 341	
	gtgactggag ttcagacgtg tgctcttccg atctaagagc tcctagggga ggatcag	57
	<210> 342	
	<211> 56	

	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 342	
	gtgactggag ttcagacgtg tgctcttccg atcttggcag gagcacagcc taagga	56
	<210> 343	
	<211> 57	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 343	
	gtgactggag ttcagacgtg tgctcttccg atctacaccc gcctcggaga tcaacac	57
	<210> 344	
	<211> 58	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 344	
	gtgactggag ttcagacgtg tgctcttccg atctccaccc ctacatctca ccttgttg	58
[0057]	<210> 345	
	<211> 55	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 345	
	gtgactggag ttcagacgtg tgctcttccg atctgtgagg ttccacgtg ccagc	55
	<210> 346	
	<211> 60	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 346	
	gtgactggag ttcagacgtg tgctcttccg atctccaaca attccaggta tgaaactccc	60
	<210> 347	
	<211> 59	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 347	
	gtgactggag ttcagacgtg tgctcttccg atctctcccc acttgtaggt tcctaattcc	59
	<210> 348	
	<211> 58	
	<212> DNA	
	<213> 人工序列	

	<220>		
	<223>	合成的多核苷酸	
	<400>	348	
		gtgactggag ttcagacgtg tgctcttccg atctgtagag tgcttggatga agaattgtg	58
	<210>	349	
	<211>	57	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	349	
		gtgactggag ttcagacgtg tgctcttccg atcttccaga ctgttgttca gtcctgt	57
	<210>	350	
	<211>	58	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	350	
		gtgactggag ttcagacgtg tgctcttccg atctctctgg atttgccac acctagtc	58
	<210>	351	
	<211>	58	
	<212>	DNA	
	<213>	人工序列	
[0058]	<220>		
	<223>	合成的多核苷酸	
	<400>	351	
		gtgactggag ttcagacgtg tgctcttccg atctgcatgc ttgctttctg aaggtggc	58
	<210>	352	
	<211>	59	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	352	
		gtgactggag ttcagacgtg tgctcttccg atctttccaa gcaagtgagc ttcagacc	59
	<210>	353	
	<211>	54	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	353	
		gtgactggag ttcagacgtg tgctcttccg atctccacc atgacacagg aggg	54
	<210>	354	
	<211>	57	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	

	<400> 354 gtctcgtggg ctcgagatg tgtataagag acaggctacc ctccacaaag cacacac	57
	<210> 355 <211> 59 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 355 gtctcgtggg ctcgagatg tgtataagag acagttggaa gagaaggatc tgctgaggc	59
	<210> 356 <211> 57 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<400> 356 tcgtcggcag cgtcagatgt gtataagaga cagcctggag accagcaagt attgtcc	57
	<210> 357 <211> 57 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
[0059]	<400> 357 tcgtcggcag cgtcagatgt gtataagaga cagcacctgt caaaagtcct actccgg	57
	<210> 358 <211> 59 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<220> <221> misc_feature <222> (6)..(24) <223> n为a, c, g或t	
	<220> <221> misc_feature <222> (51)..(58) <223> n为a, c, g或t	
	<400> 358 caccgnnnnn nnnnnnnnnn nnnngtttgg gtcttcgaga agacctatc nnnnnnnnc	59
	<210> 359 <211> 59 <212> DNA <213> 人工序列	
	<220> <223> 合成的多核苷酸	
	<220> <221> misc_feature <222> (6)..(13)	

	<223> n为a, c, g或t	
	<220>	
	<221> misc_feature	
	<222> (40)..(58)	
	<223> n为a, c, g或t	
	<400> 359	
	aattgnnnnn nnngaataagg tcttctcgaa gacccaaacn nnnnnnnnnn nnnnnnnnc	59
	<210> 360	
	<211> 63	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<220>	
	<221> misc_feature	
	<222> (6)..(24)	
	<223> n为a, c, g或t	
	<220>	
	<221> misc_feature	
	<222> (51)..(58)	
	<223> n为a, c, g或t	
	<400> 360	
	caccgnnnnn nnnnnnnnnn nnnngtttgg gtcttcgaga agacctattc nnnnnnnnca	60
	att	63
[0060]	<210> 361	
	<211> 63	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<220>	
	<221> misc_feature	
	<222> (6)..(13)	
	<223> n为a, c, g或t	
	<220>	
	<221> misc_feature	
	<222> (40)..(58)	
	<223> n为a, c, g或t	
	<400> 361	
	aattgnnnnn nnngaataagg tcttctcgaa gacccaaacn nnnnnnnnnn nnnnnnnncg	60
	gtg	63
	<210> 362	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 362	
	cctttgcaag acccgcacga tgg	23
	<210> 363	
	<211> 23	

	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 363	
	gctttgccccg aaccgcacga gag	23
	<210> 364	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 364	
	cctaggccag acctgcacga tgg	23
	<210> 365	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 365	
	gctctggaag acccgcacca ggg	23
[0061]	<210> 366	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 366	
	ccttatcaag acccacacca gag	23
	<210> 367	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 367	
	atccccctcc tcgtagegca tgg	23
	<210> 368	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的多核苷酸	
	<400> 368	
	tgcgccctcc tcctagegca tgg	23
	<210> 369	
	<211> 23	
	<212> DNA	
	<213> 人工序列	

	<220>		
	<223>	合成的多核苷酸	
	<400>	369	
		ctgccccctcc tggtagcgcc tgg	23
	<210>	370	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	370	
		ctgctcctcc tcgtagcgct ggg	23
	<210>	371	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	371	
		aaccagctcc tcgtagctca ggg	23
	<210>	372	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
[0062]	<220>		
	<223>	合成的多核苷酸	
	<400>	372	
		agccccgctcc tcgtggggca cgg	23
	<210>	373	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	373	
		accgccctcc tcctagctca ggg	23
	<210>	374	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	374	
		ctctctctcc gcgtagcgct tgg	23
	<210>	375	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	

	<p><400> 375 ctcctctctcc tcctaccgca agg</p>	23
	<p><210> 376 <211> 23 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 376 attcctctcc tggtagcgca agg</p>	23
	<p><210> 377 <211> 23 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 377 ctcacctctcc tcctagcaca tag</p>	23
	<p><210> 378 <211> 23 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
[0063]	<p><400> 378 gggaacaata aagaagcgct tgg</p>	23
	<p><210> 379 <211> 23 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 379 tggaaaaaca aagaagagct tgg</p>	23
	<p><210> 380 <211> 23 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 380 gggaagtata aggaagagct cag</p>	23
	<p><210> 381 <211> 23 <212> DNA <213> 人工序列</p> <p><220> <223> 合成的多核苷酸</p>	
	<p><400> 381 gtgagcaata aagcagccct aag</p>	23

[0064]	<210>	382	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	382	
		tctagtccat cccccattac tgg	23
	<210>	383	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	383	
		aatattccat tccccattac tgg	23
	<210>	384	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	384	
		tgttgtccat acctcattac tgg	23
	<210>	385	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	385	
		tctaggtcat gcaccattac tgg	23
	<210>	386	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	386	
		tccacaccct cccccattac tag	23
	<210>	387	
	<211>	23	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	合成的多核苷酸	
	<400>	387	
		cccattcctt cccccattac cag	23

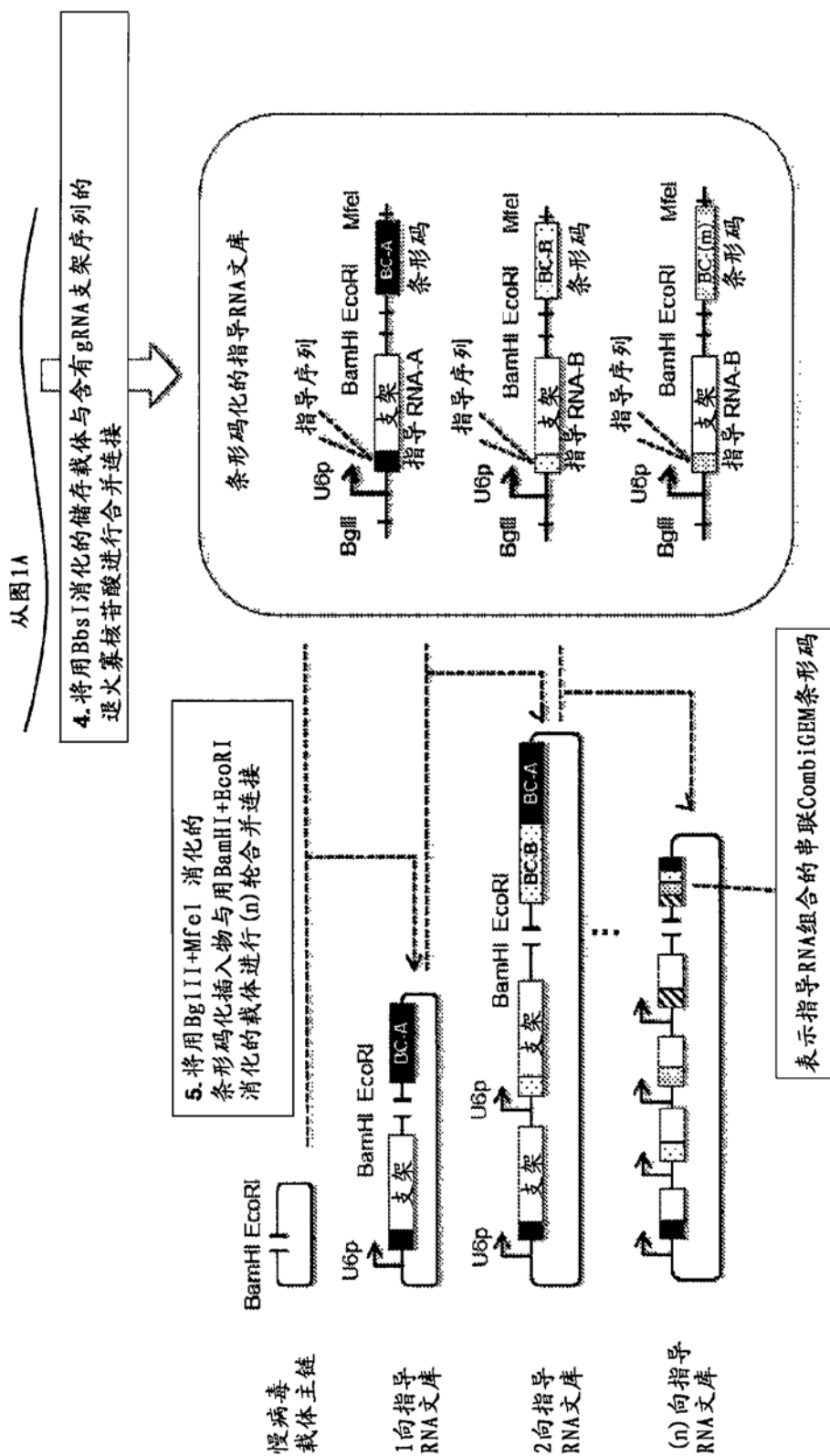


图1A(续)

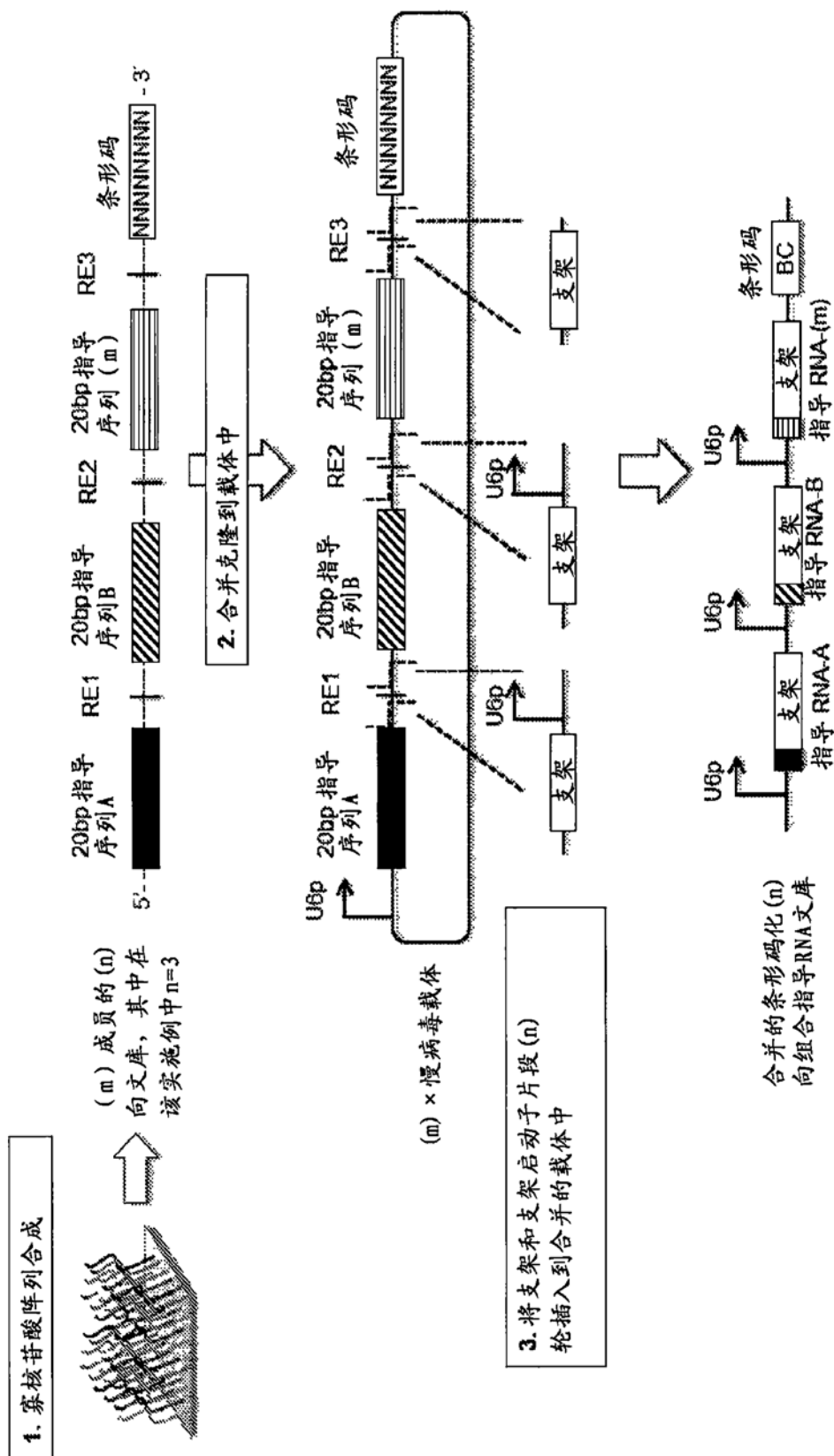


图2A

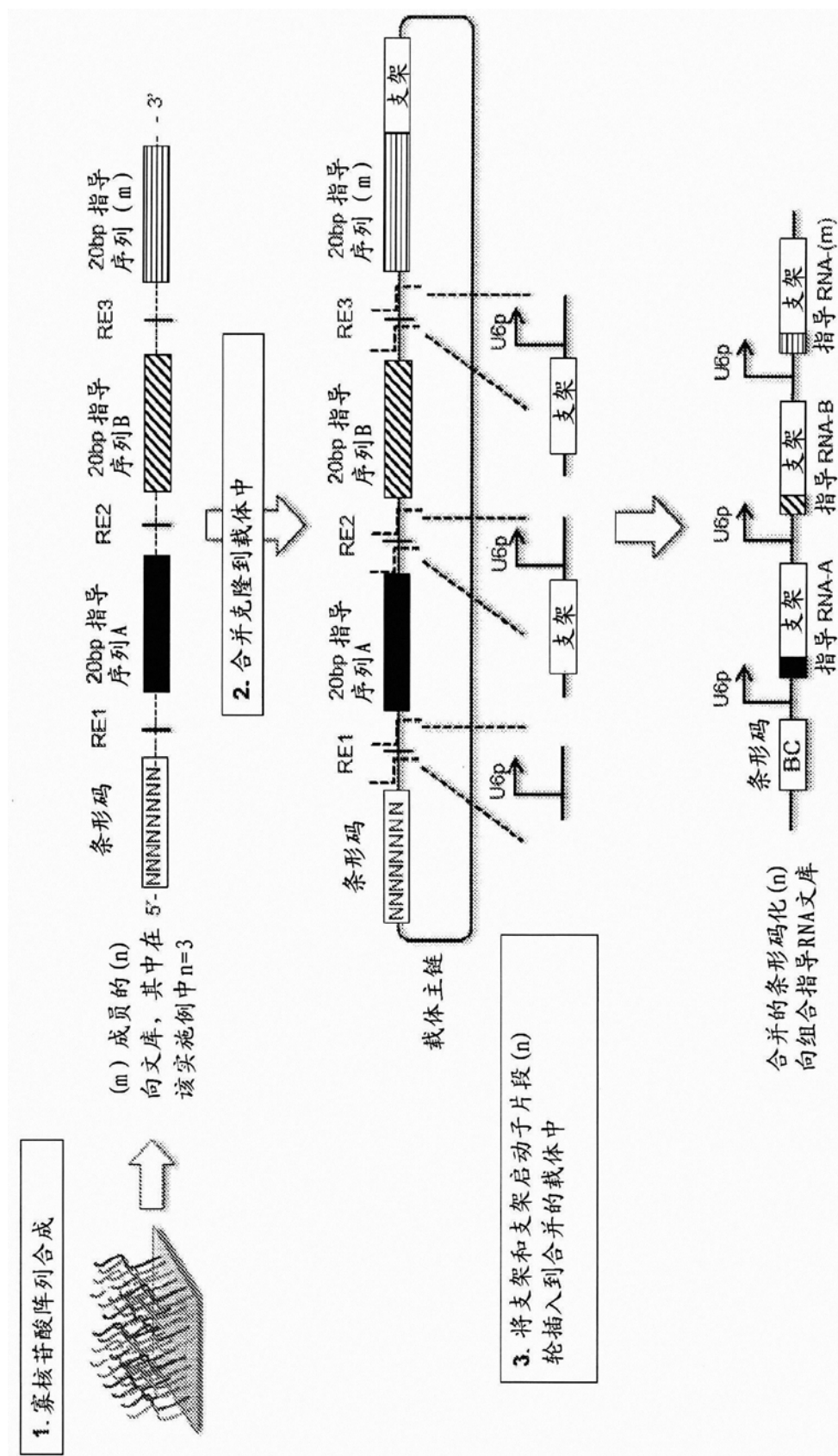


图2B

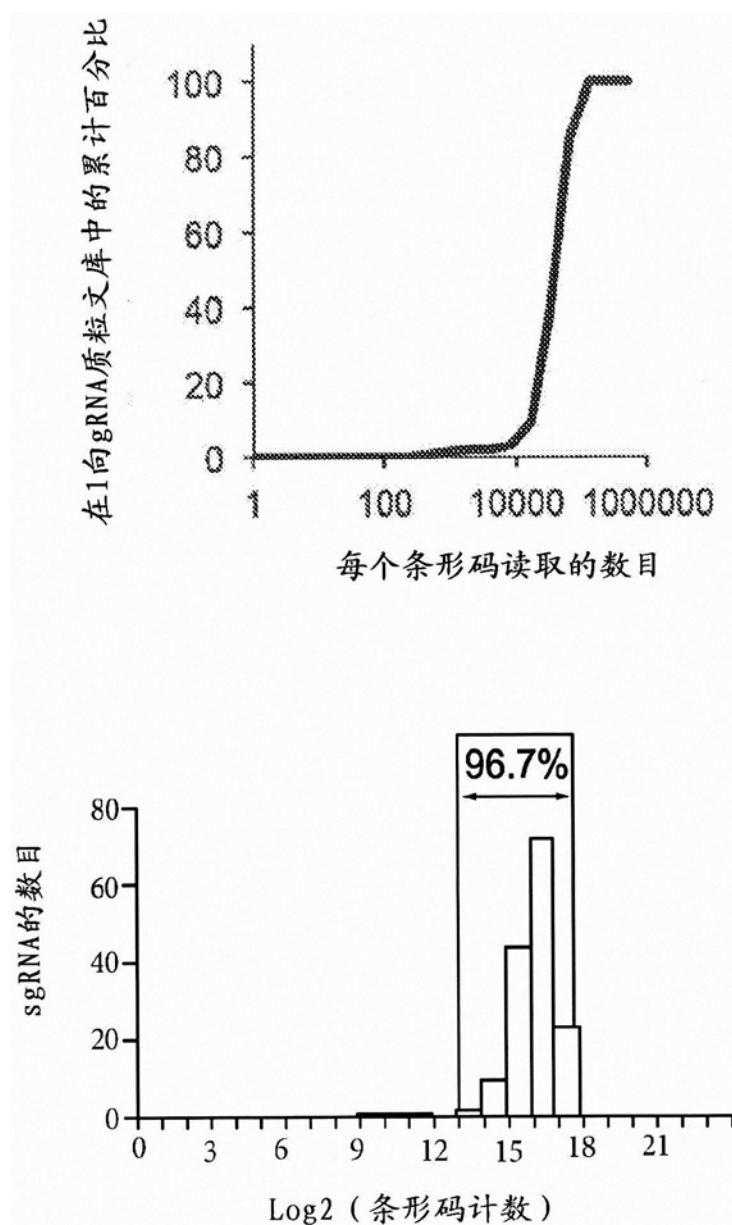


图3A

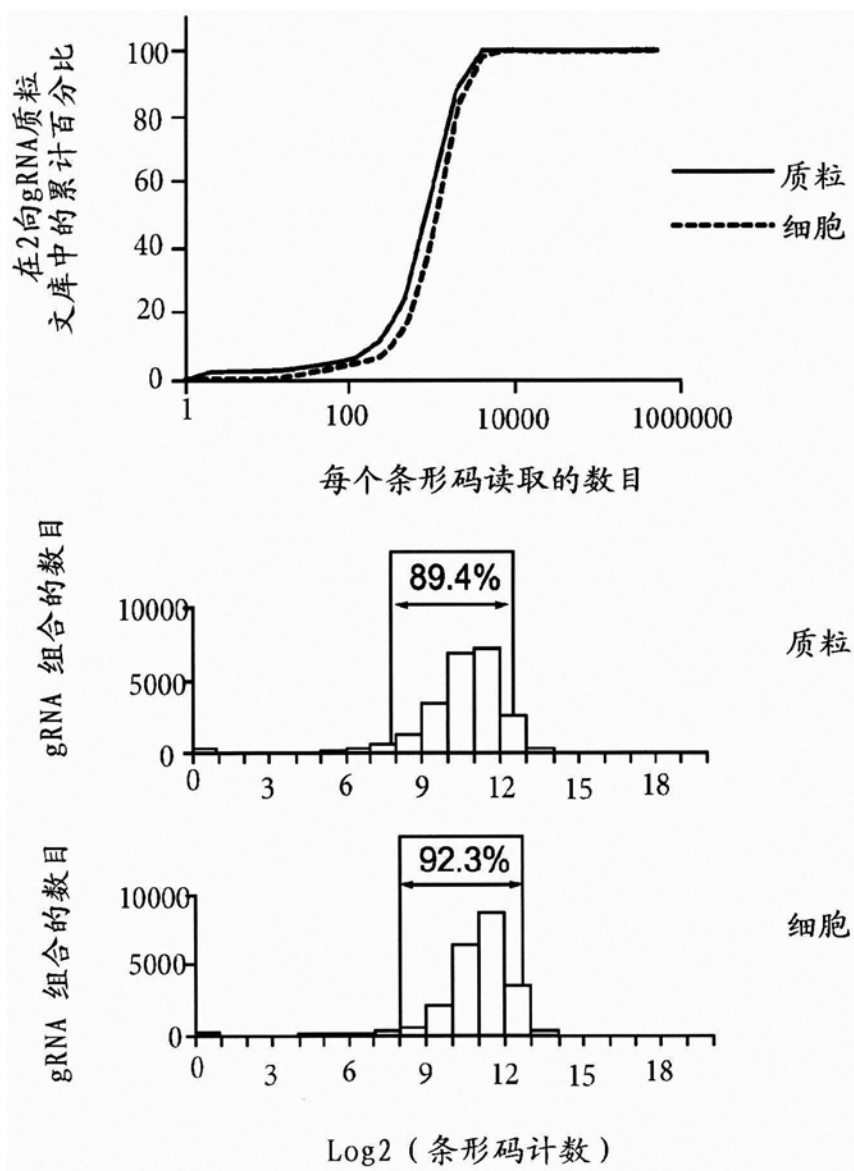


图3B

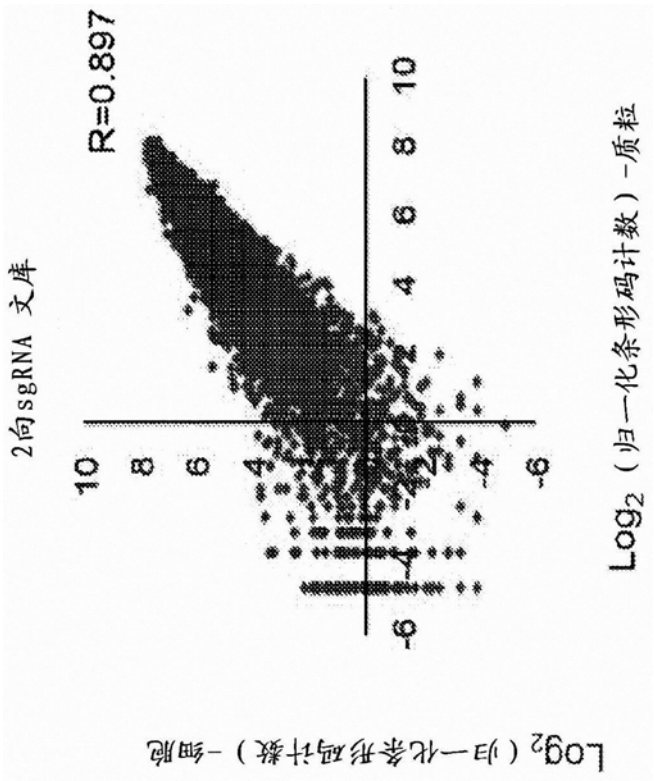


图3C

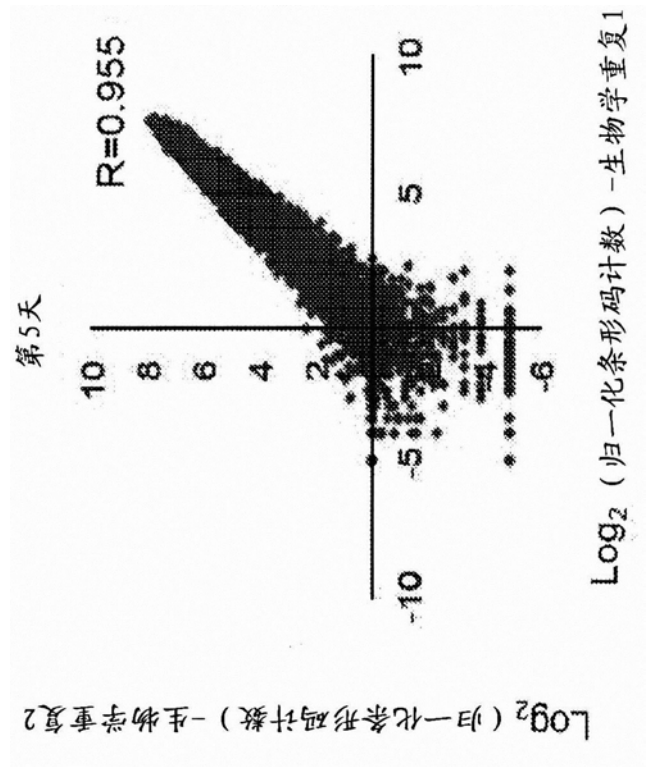


图3D

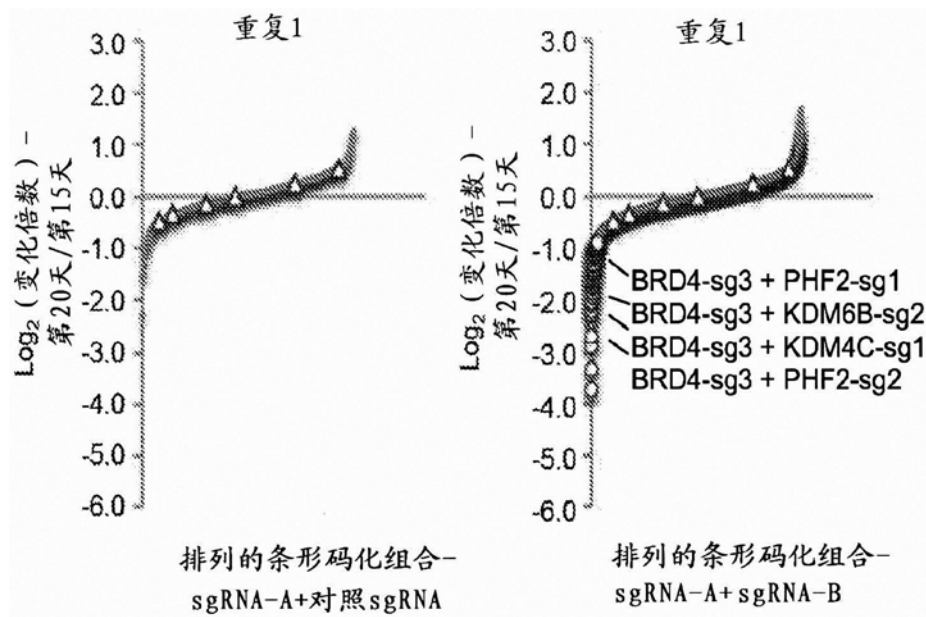


图4B

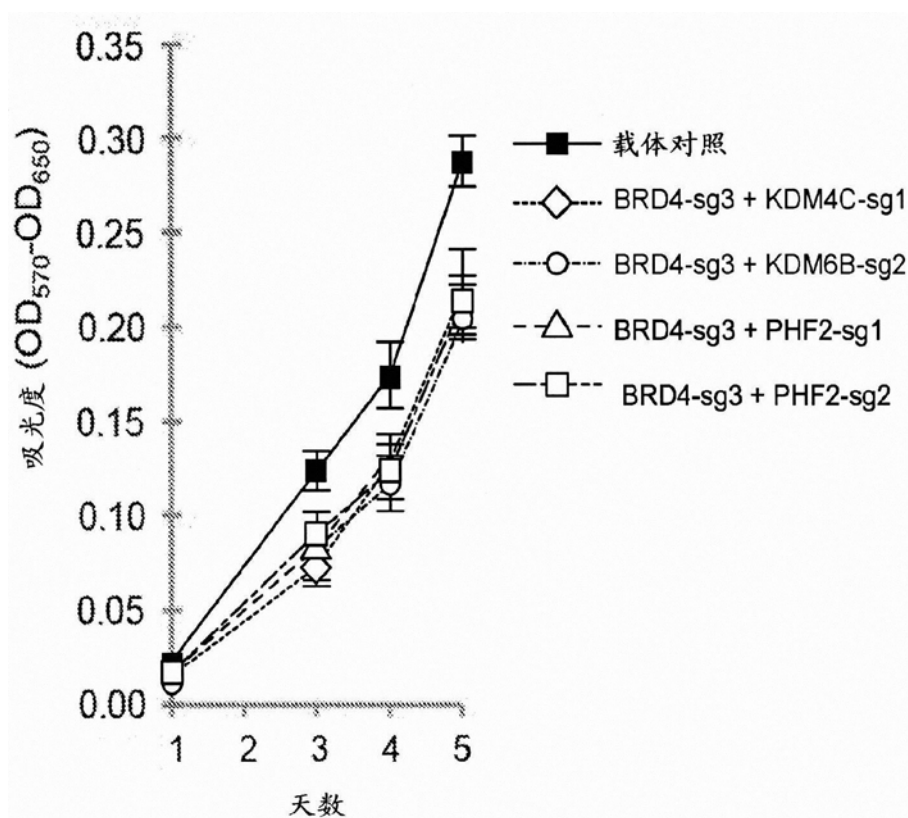


图4C

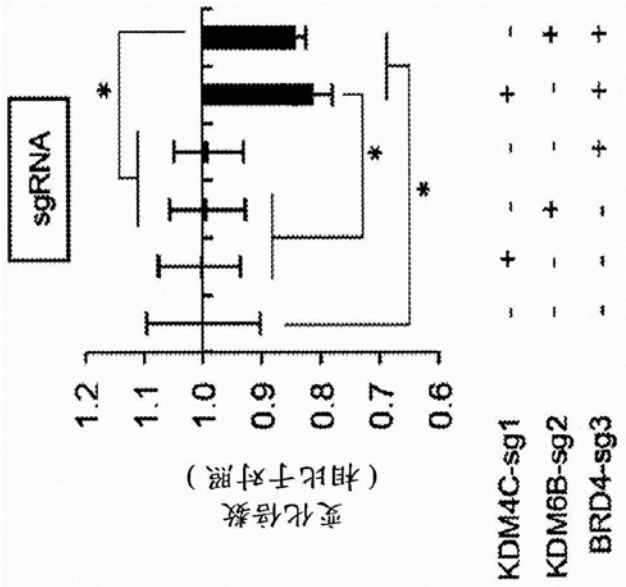


图5A

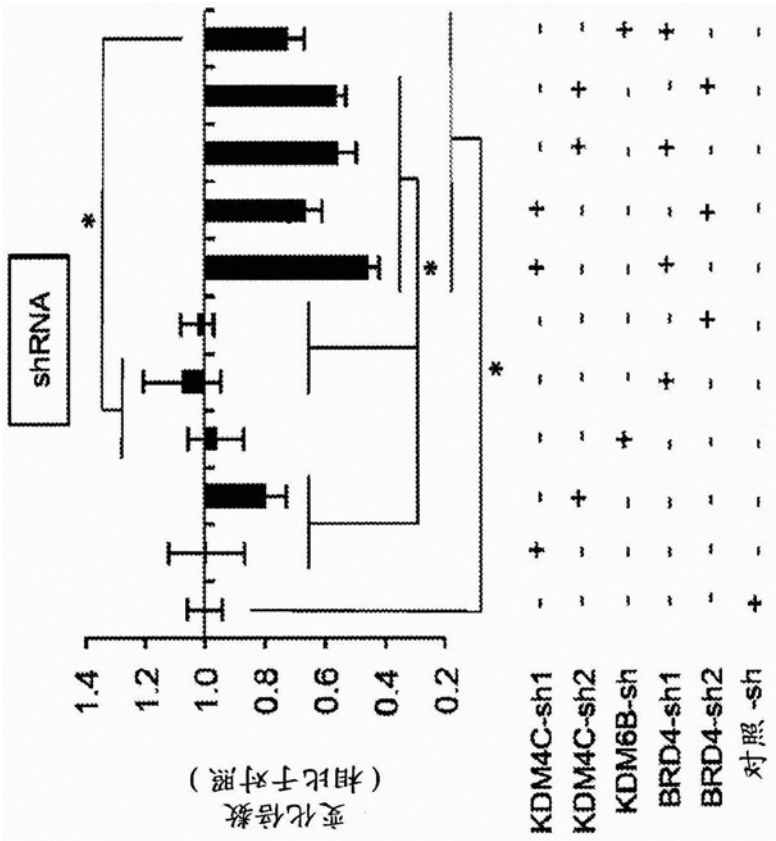


图5B

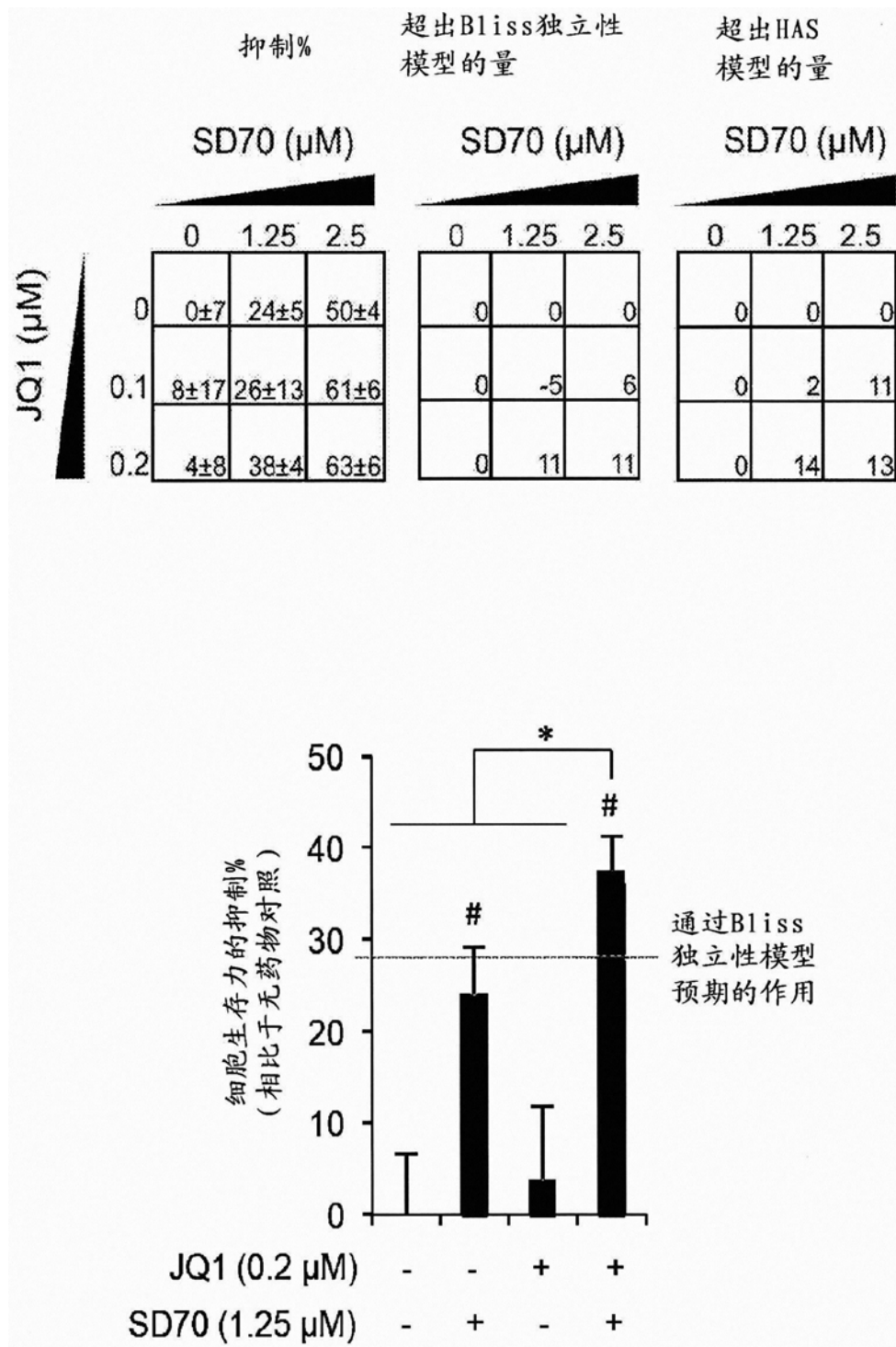


图5C

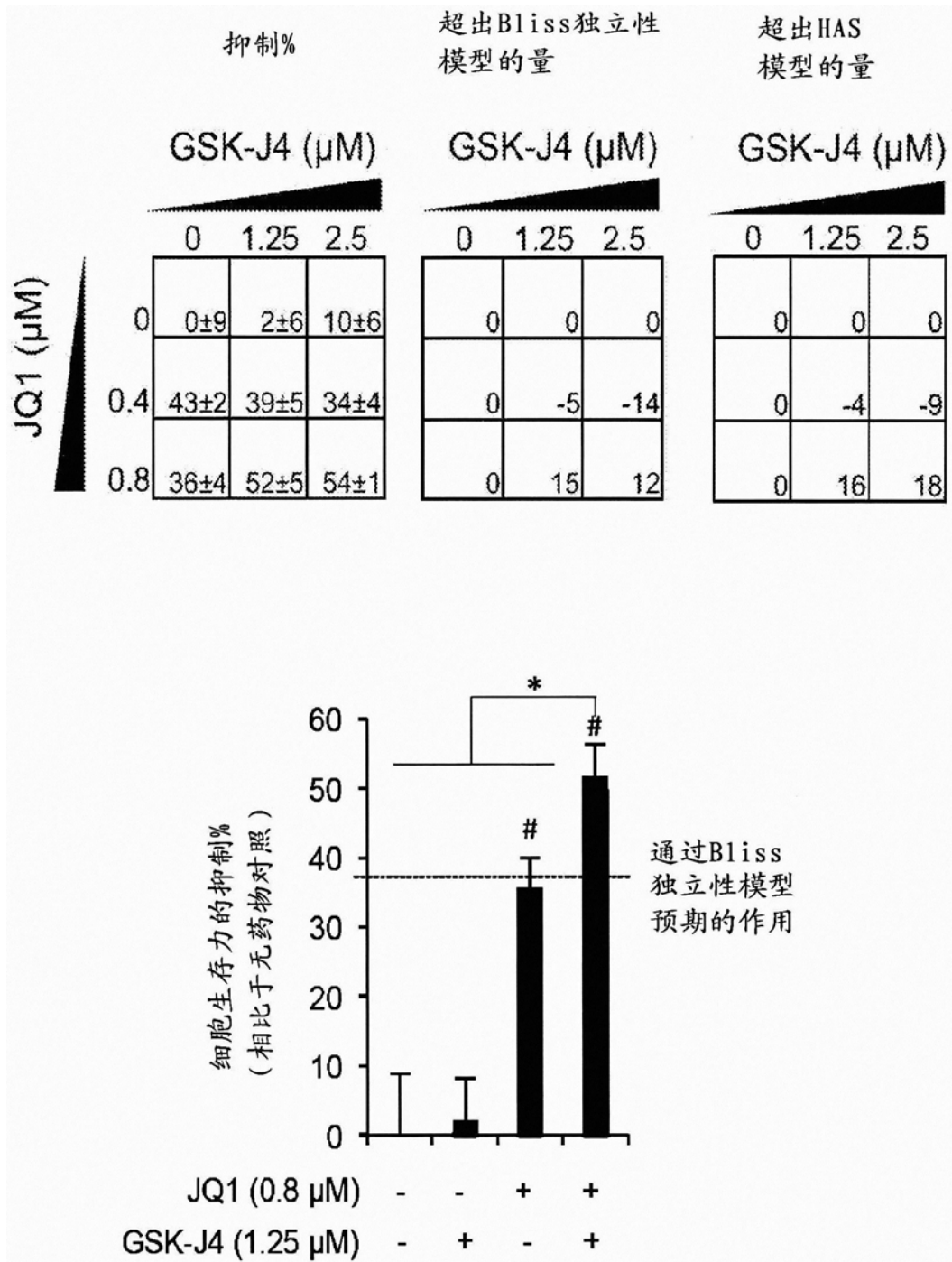


图5D

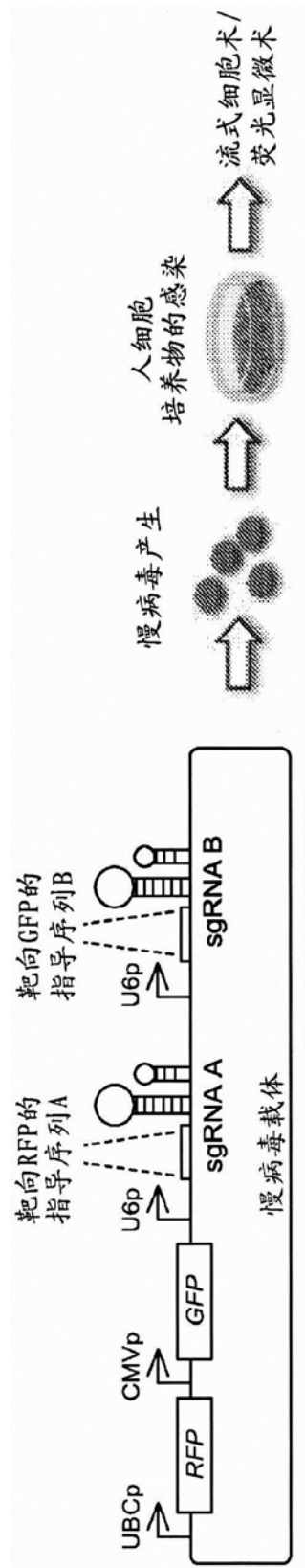


图6A

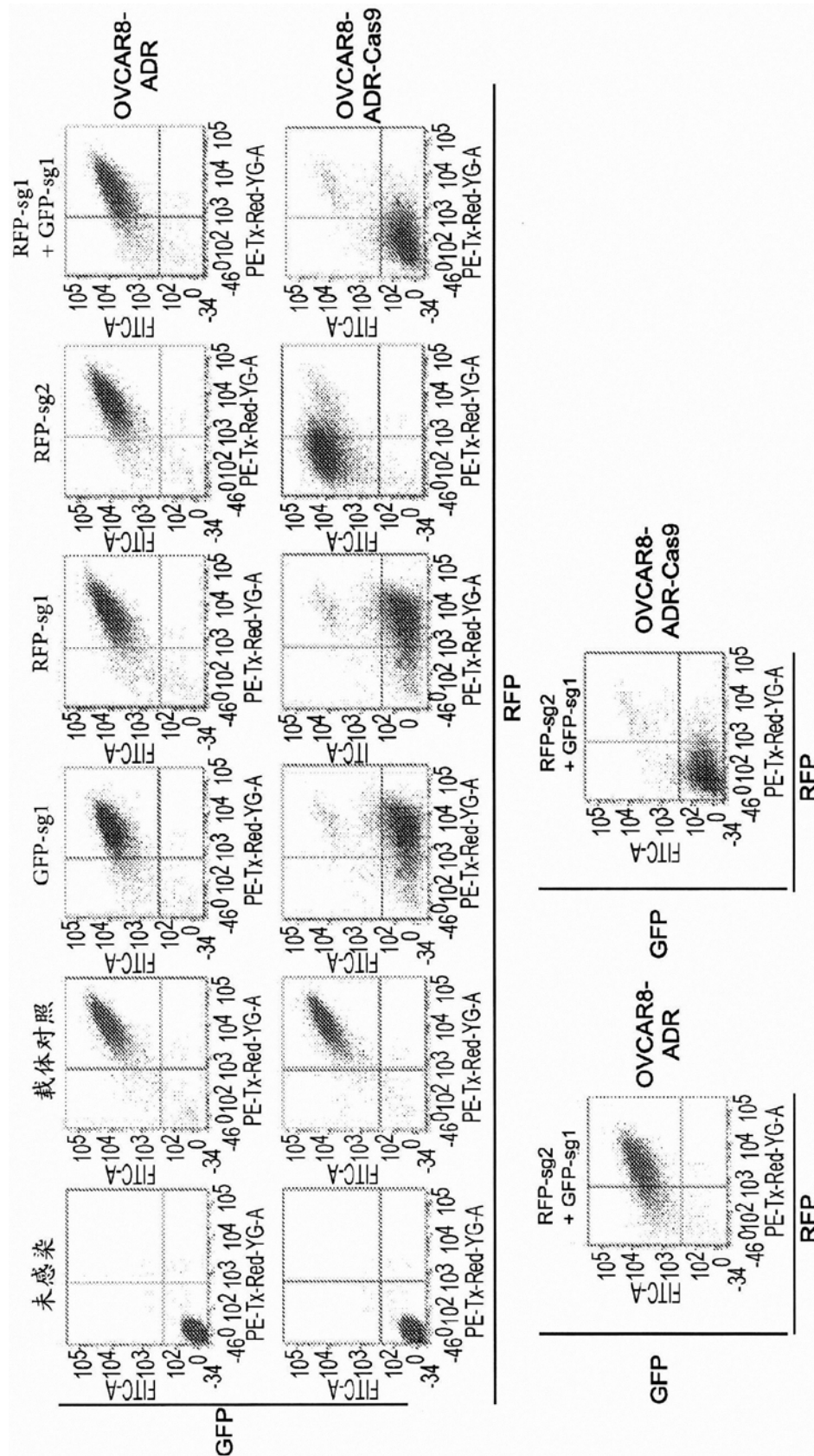


图6B

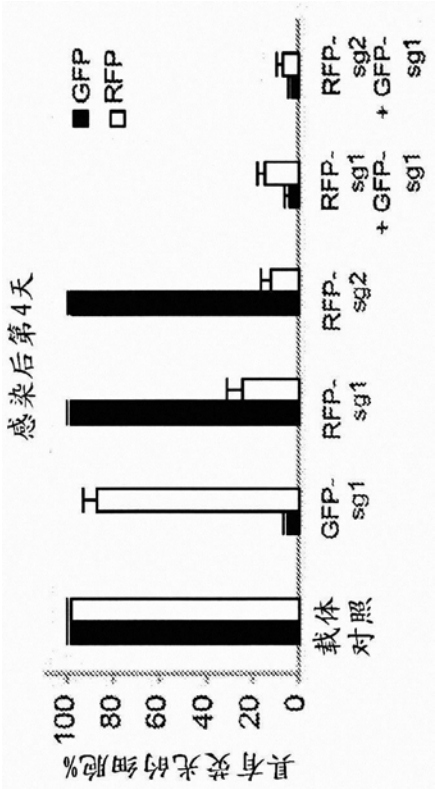


图6C

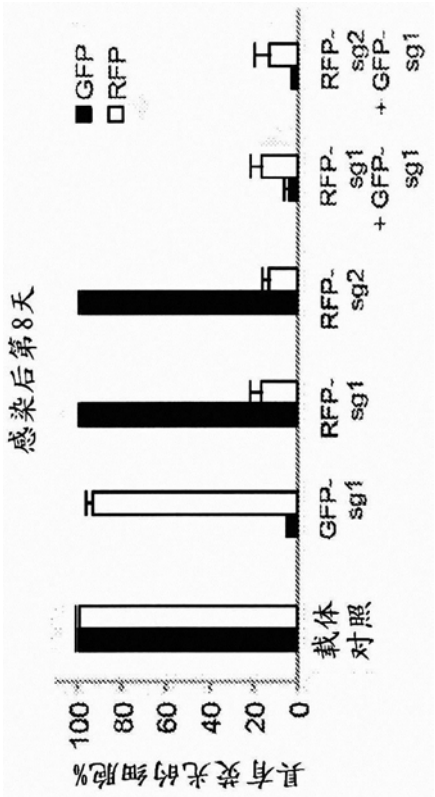


图6D

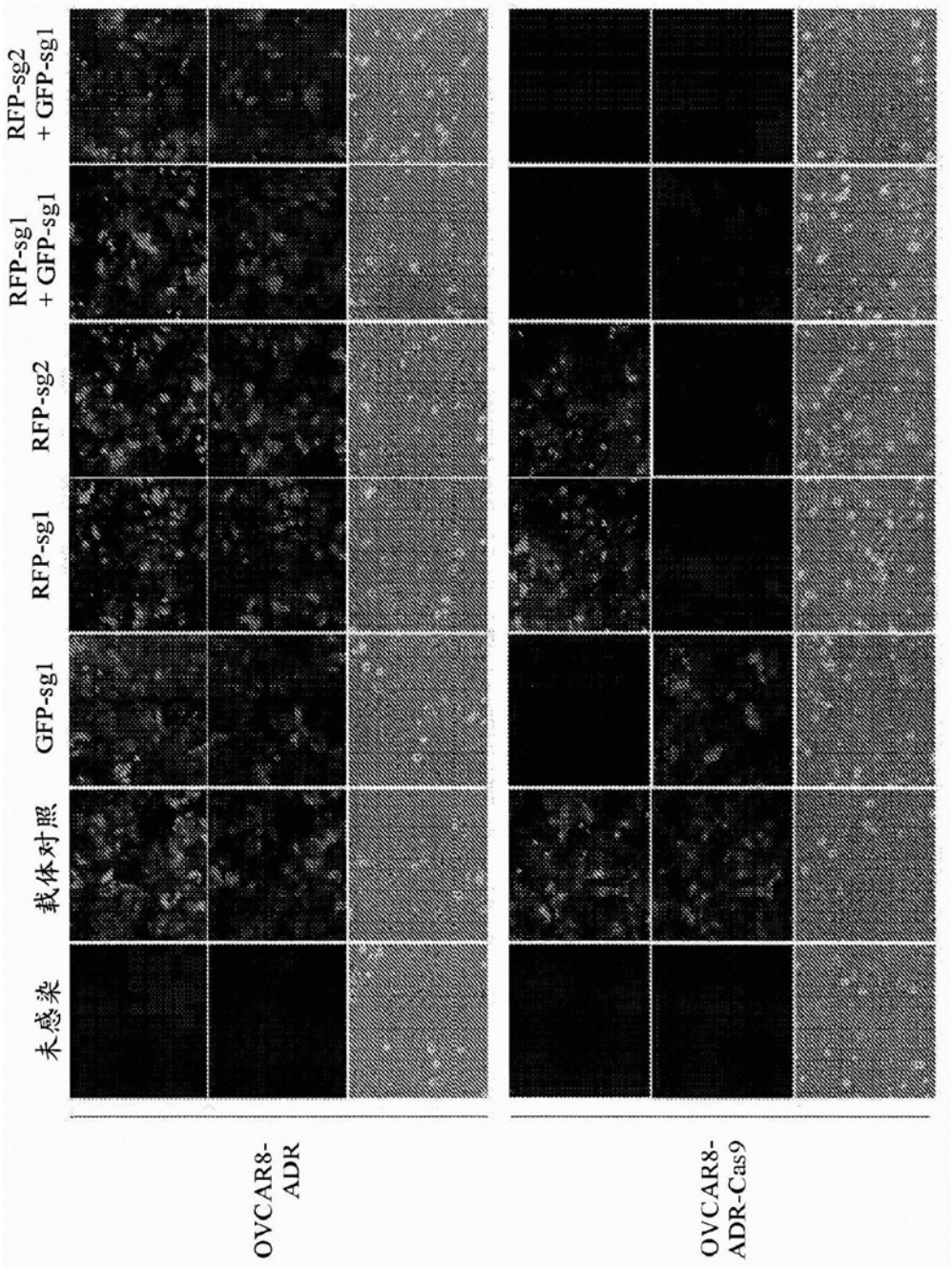


图6E

sgRNA	未切割的条带	经切割的条带
DNMT1-sg1	600	450/150
DNMT3B-sg1	544	400/150
PRMT2-sg3	559	370/170
HDAC2-sg1	871	500/300
ING4-sg1	502	300/200
KDM1B-sg3	499	350/150
KDM2A-sg3	455	300/150
PRMT6-sg1	849	550/300
KDM4C-sg1	419	264/155
KDM6B-sg2	489	273/216
BRD4-sg3	541	341/200

图7A

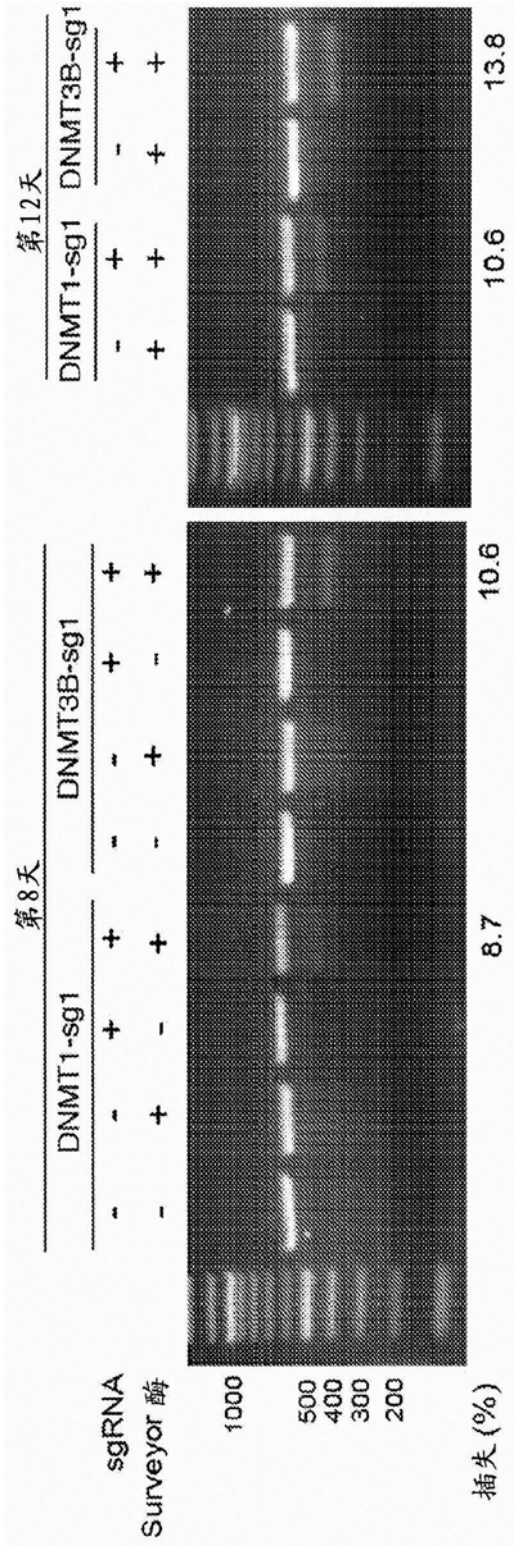


图7B

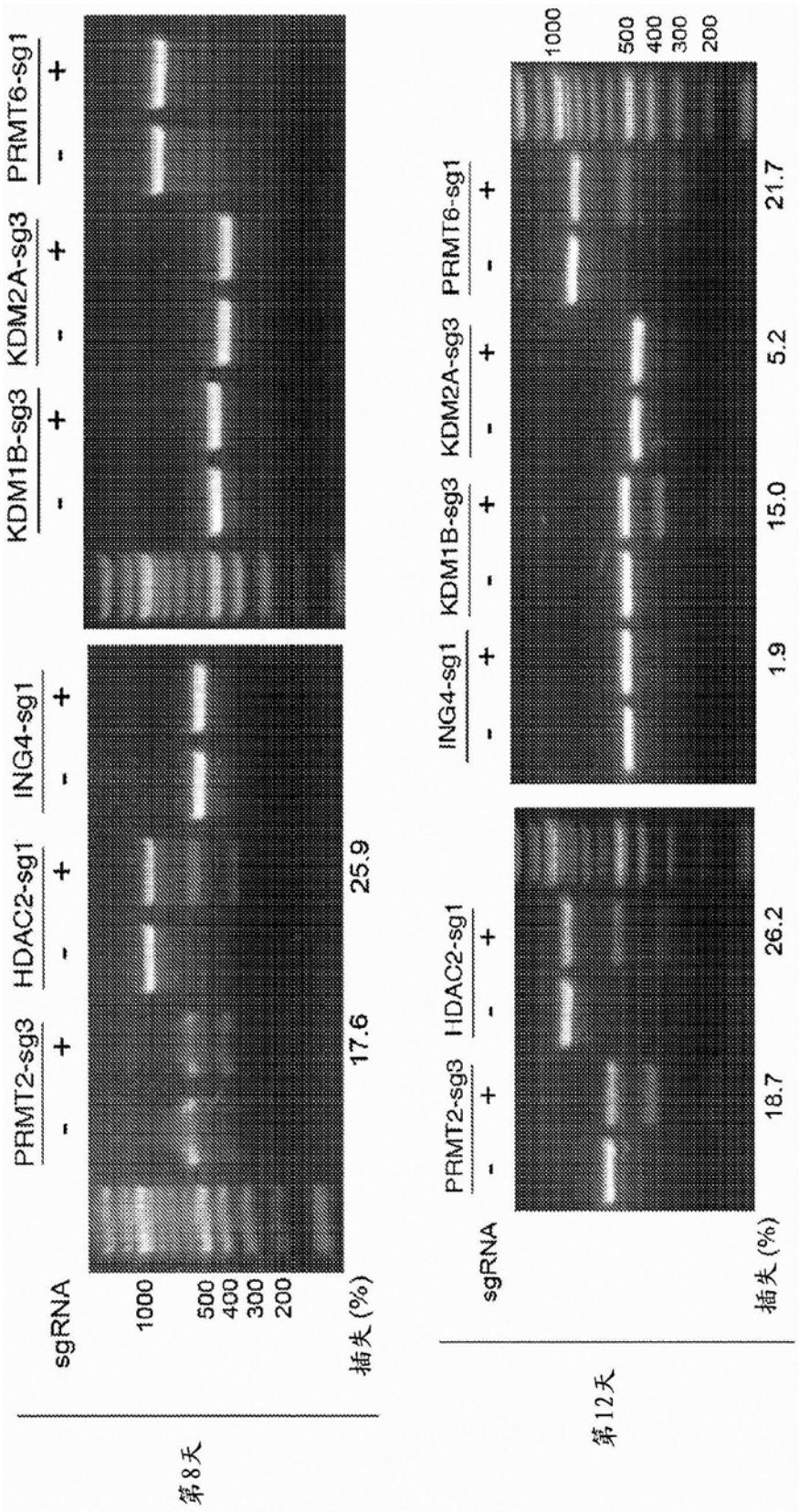


图7C

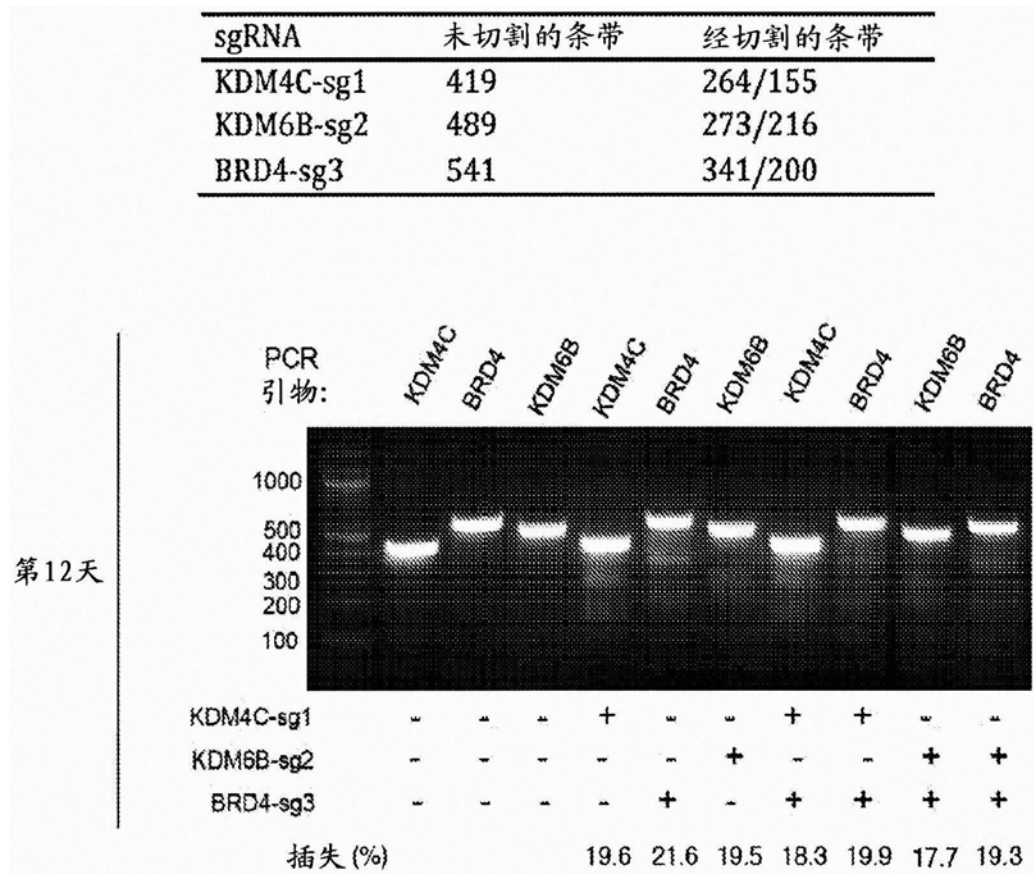


图8A

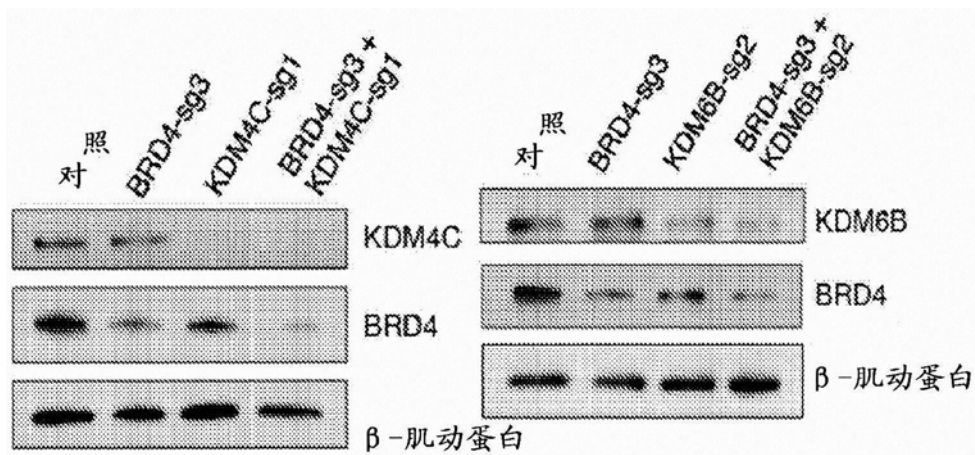


图8B

BMI1-sg2 + PHF8-sg2				
WT	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	BMI1
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	PHF8
单细胞来源 的克隆#1	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	WT
	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	WT
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	WT
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	WT
单细胞来源 的克隆#2	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	WT
	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	-18 bp
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	+1 bp
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	+1 bp
单细胞来源 的克隆#3	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	WT
	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	WT
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	WT
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	-3 bp
单细胞来源 的克隆#4	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	-3 bp
	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	-48 bp
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	WT
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	-3 bp
单细胞来源 的克隆#5	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	-3 bp
	5'-	TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTGTGTAACCTTGGACATCACA	-3'	-3 bp
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	-3 bp
	5'-	TGACTCACAATCTCTCATGTTTTCTGGCTTAGTGAAAAAACGCCGTGGATCTTCAAAGGG	-3'	-3 bp

图9A

BM11-sg2 + PHF8-sg2	
单细胞来源 的克隆#6	<p>5'~ TTTCAACAGTTTCCTACCTTATATTCAGTAGTGGTCTTGTGAACCTTGGACATCACA -3'</p> <p>5'~ TTTCAACAGTTTCCTACCTTATATTCAGTGGTCTTGTGAACCTTGGACATCACA -3'</p> <p>5'~ TGACTCACAATCTCTCATGTTTCTGGCTTAGTGAAAAAACCGCGGTGGATCTTCAAAGGGG -3'</p> <p>5'~ TGACTCACAATCTCTCATGTTTCTGGCTTAGTGAAAAAACCGCGGTGGATCTTCAAAGGGG -3'</p> <p>WT -13 bp</p> <p>+1 bp</p> <p>+1 bp</p> <p>BM11</p> <p>PHF8</p>
单细胞来源 的克隆#7	<p>5'~ TTTCAACAGTTTCCTACCTTATATTCAGTGGTCTTGTGAACCTTGGACATCACA -3'</p> <p>5'~ TTTCAACAGTTTCCTACCTTATATTCAGTGGTCTTGTGAACCTTGGACATCACA -3'</p> <p>5'~ TGACTCACAATCTCTCATGTTTCTGGCTTAGTGAAAAAACCGCGGTGGATCTTCAAAGGGG -3'</p> <p>5'~ TGACTCACAATCTCTCATGTTTCTGGCTTAGTGAAAAAACCGCGGTGGATCTTCAAAGGGG -3'</p> <p>-3 bp</p> <p>-3 bp</p> <p>1 mut.; +3 bp</p> <p>1 mut.; +3 bp</p> <p>BM11</p> <p>PHF8</p>
单细胞来源 的克隆#8	<p>5'~ TTTCAACAGTTTCCTACCTTATATTCAGTGGTCTTGTGAACCTTGGACATCACA -3'</p> <p>5'~ TTTCAACAGTTTCCTACCTTATATTCAGTGGTCTTGTGAACCTTGGACATCACA -3'</p> <p>5'~ TGACTCACAATCTCTCATGTTTCTGGCTTAGTGAAAAAACCGCGGTGGATCTTCAAAGGGG -3'</p> <p>5'~ TGACTCACAATCTCTCATGTTTCTGGCTTAGTGAAAAAACCGCGGTGGATCTTCAAAGGGG -3'</p> <p>-6 bp</p> <p>-10 bp</p> <p>WT</p> <p>WT</p> <p>BM11</p> <p>PHF8</p>
单细胞来源 的克隆#9	<p>5'~ TTTCAACAGTTTCCTACCTTATATTCAGTGGTCTTGTGAACCTTGGACATCACA -3'</p> <p>5'~ TTTCAACAGTTTCCTACCTTATATTCAGTGGTCTTGTGAACCTTGGACATCACA -3'</p> <p>5'~ TGACTCACAATCTCTCATGTTTCTGGCTTAGTGAAAAAACCGCGGTGGATCTTCAAAGGGG -3'</p> <p>5'~ TGACTCACAATCTCTCATGTTTCTGGCTTAGTGAAAAAACCGCGGTGGATCTTCAAAGGGG -3'</p> <p>-2 bp</p> <p>1 mut.; -7 bp</p> <p>WT</p> <p>WT</p> <p>BM11</p> <p>PHF8</p>
单细胞来源 的克隆#10	<p>5'~ TTTCAACAGTTTCCTACCTTATATTCAGTGGTCTTGTGAACCTTGGACATCACA -3'</p> <p>5'~ TTTCAACAGTTTCCTACCTTATATTCAGTGGTCTTGTGAACCTTGGACATCACA -3'</p> <p>5'~ TGACTCACAATCTCTCATGTTTCTGGCTTAGTGAAAAAACCGCGGTGGATCTTCAAAGGGG -3'</p> <p>5'~ TGACTCACAATCTCTCATGTTTCTGGCTTAGTGAAAAAACCGCGGTGGATCTTCAAAGGGG -3'</p> <p>WT</p> <p>WT</p> <p>WT</p> <p>WT</p> <p>BM11</p> <p>PHF8</p>

图9A (续)

BRD4-sg2 + KDM4C-sg1

WT	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCCATTTACTGGCAGATTTCTCAATCTCGTCC-	3'	BRD4
	5'-	GAAGGATAATCACCCTTGGGAAGACCCGACGATGGGCTCCTTTAGACTCCATGATGCA-	3'	KDM4C
单细胞来源 的克隆#1	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCCATTTACTGGCAGATTTCTCAATCTCGTCC-	3'	WT
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCCATTTACTGGCAGATTTCTCAATCTCGTCC-	3'	WT
	5'-	GAAGGATAATCACCCTTGGCAAGACCC--ACGATGGGCTCCCTTTAGACTCCATGATGCA-	3'	-2 bp
	5'-	GAAGGATAATCACCCTTGGCAAGACCC--ACGATGGGCTCCTTTAGACTCCATGATGCA-	3'	-2 bp
单细胞来源 的克隆#2	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCCATTTACTGGCAGATTTCTCAATCTCGTCC-	3'	WT
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCCATTTACTGGCAGATTTCTCAATCTCGTCC-	3'	WT
	5'-	GAAGGATAATCACCCTTGGCAAGACCCG--CGATGGGCTCCCTTTAGACTCCATGATGCA-	3'	-2 bp
	5'-	GAAGGATAATCACCCTTGGCAAGACCCG--CGATGGGCTCCCTTTAGACTCCATGATGCA-	3'	-2 bp
单细胞来源 的克隆#3	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCCATTTACTGGCAGATTTCTCAATCTCGTCC-	3'	WT
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATC-----CTGGCAGATTTCTCAATCTCGTCC-	3'	-8 bp
	5'-	GAAGGATAATCACCCTTGGCAAGACCCGACGATGGGCTCCCTTTAGACTCCATGATGCA-	3'	WT
	5'-	GAAGGATAATCACCCTTGGCAAGACCCGACGATGGGCTCCCTTTAGACTCCATGATGCA-	3'	-1 bp
单细胞来源 的克隆#4	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCCATTTACTGGCAGATTTCTCAATCTCGTCC-	3'	WT
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCCATTTACTGGCAGATTTCTCAATCTCGTCC-	3'	WT
	5'-	GAAGGATAATCACCCTTGGCAAA-----TCCATGATGCA-	3'	1 mut., -25 bp
	5'-	GAAGGATAATCACCCTTGGCA-----GATGGGCTCCTTTAGACTCCATGATGCA-	3'	-10 bp
单细胞来源 的克隆#5	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCCATTTACTGGCAGATTTCTCAATCTCGTCC-	3'	WT
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCCATTTACTGGCAGATTTCTCAATCTCGTCC-	3'	WT
	5'-	GAAGGATAATCACCCTTGGCAAGACCCGACGATGGGCTCCCTTTAGACTCCATGATGCA-	3'	WT
	5'-	GAAGGATAATCACCCTTGGCAAGACCCGACGATGGGCTCCCTTTAGACTCCATGATGCA-	3'	WT

图9B

BRD4-sg2 + KDM4C-sg1

单细胞来源 的克隆#6	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	BRD4
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
单细胞来源 的克隆#7	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	BRD4
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
单细胞来源 的克隆#8	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	BRD4
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
单细胞来源 的克隆#9	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	BRD4
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
单细胞来源 的克隆#10	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	BRD4
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
单细胞来源 的克隆#10	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	BRD4
	5'-	TAGACATTTGGGAAGTTTCTAGTCCATCCCCATTACTGGCAGATTCTCAATCTCGTCC-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	
	5'-	GAAGGATAATCACCCTTTGCAAGACCCCGCAACGATGGGCTCCTTTAGACTCCCATGTATGCA-	3'	

图9B (续)

BMI1-sg2 + PHF8-sg2

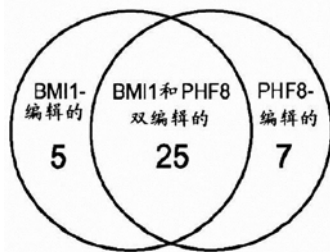


图9C

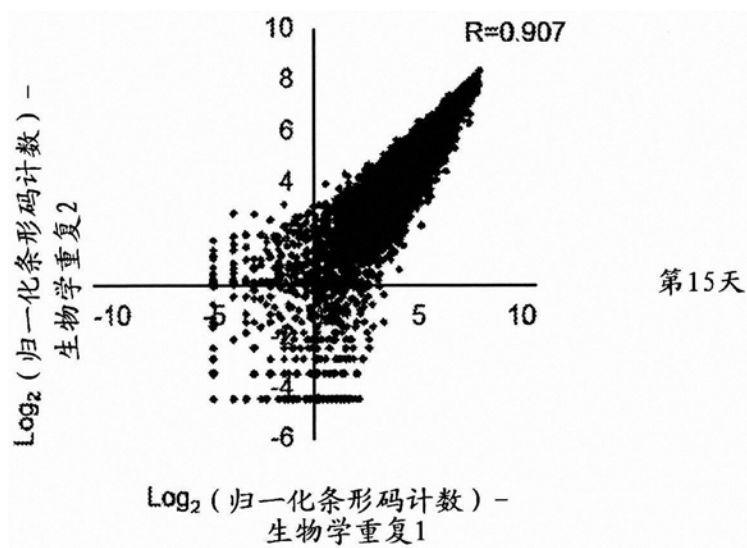


图10A

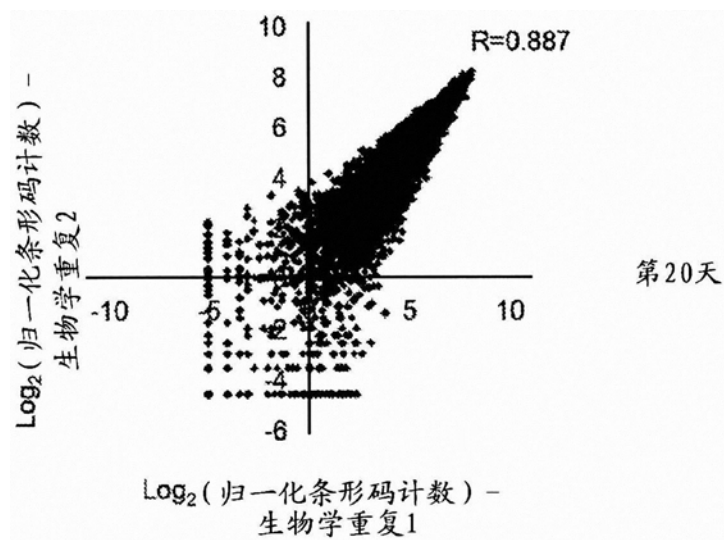


图10B

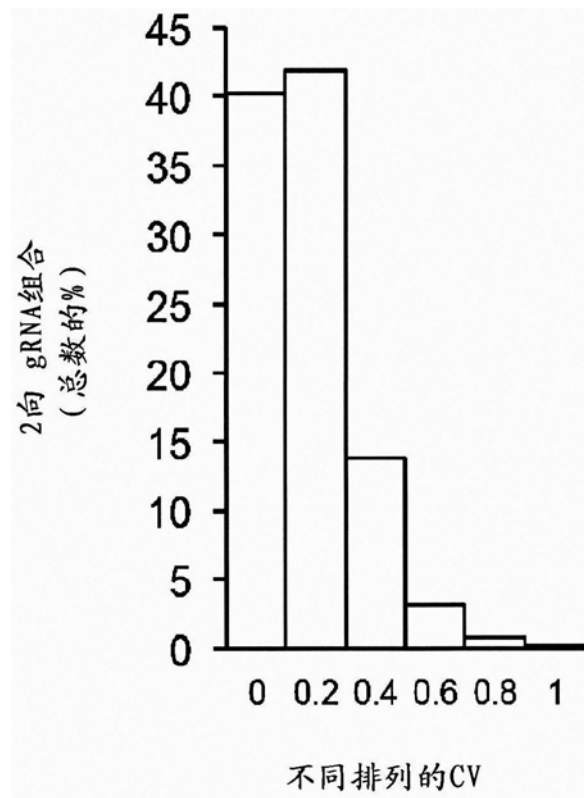


图11

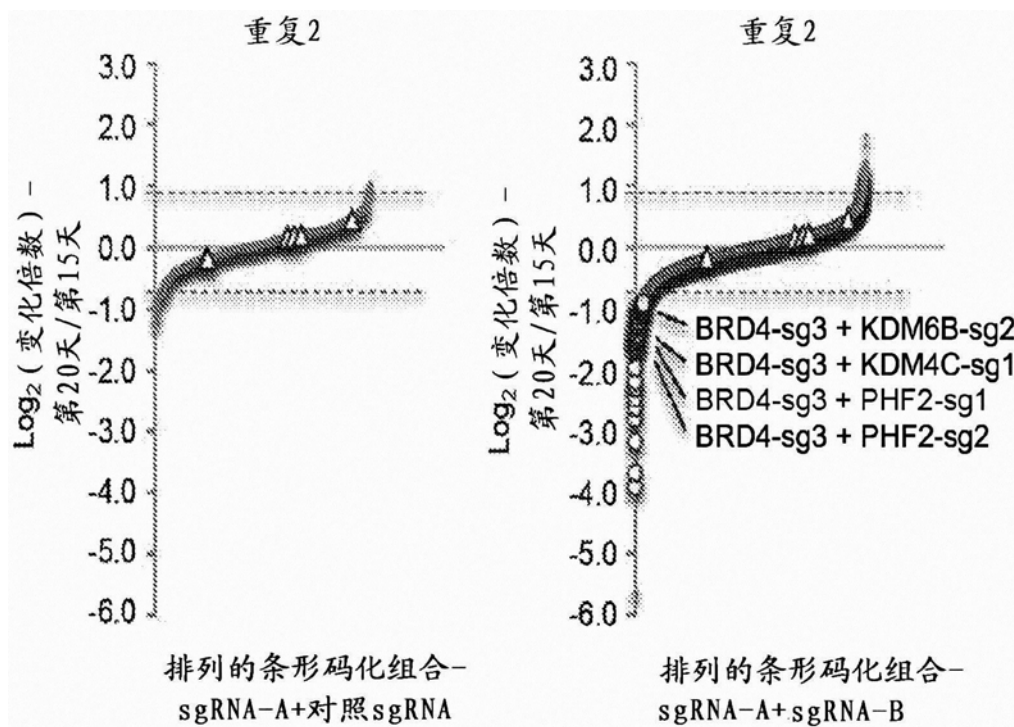


图12A

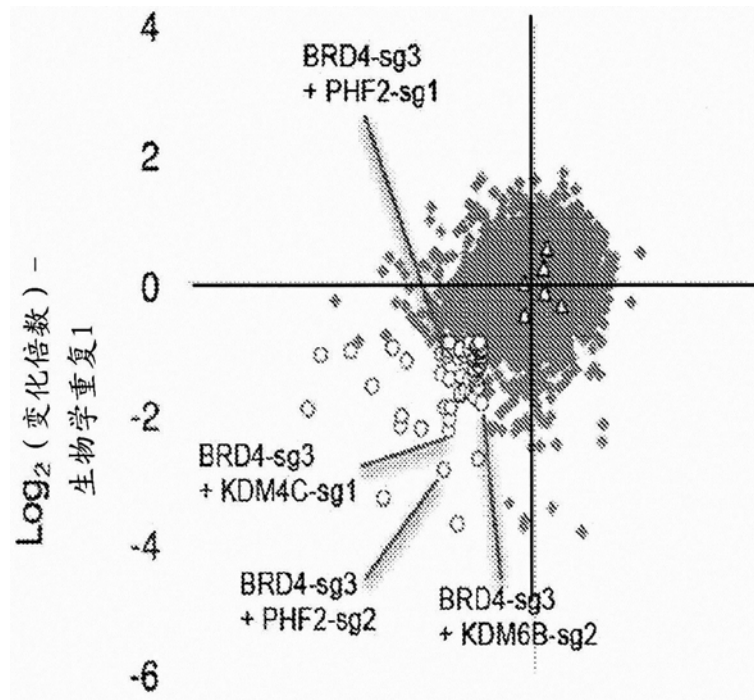


图12B

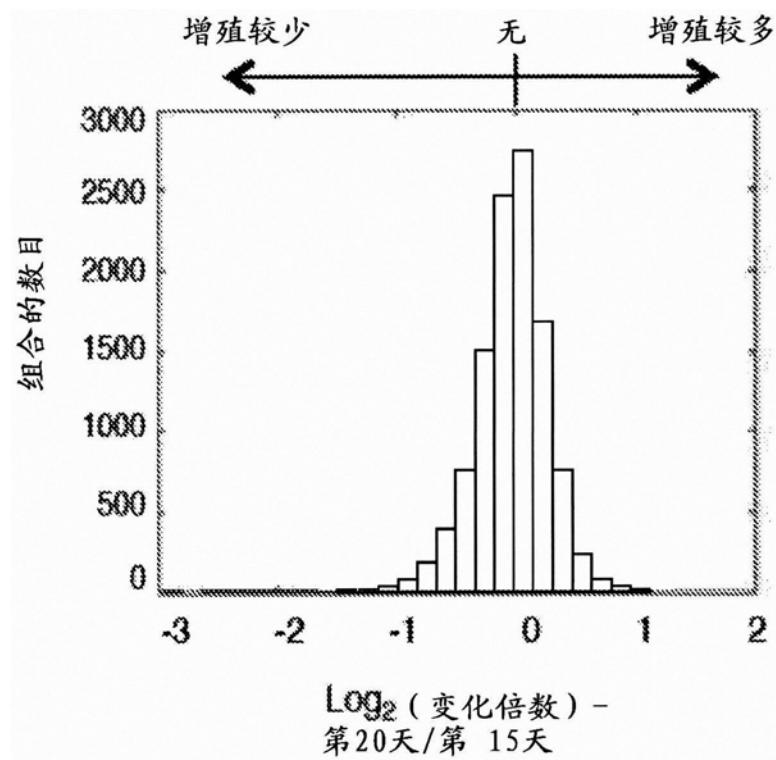


图12C

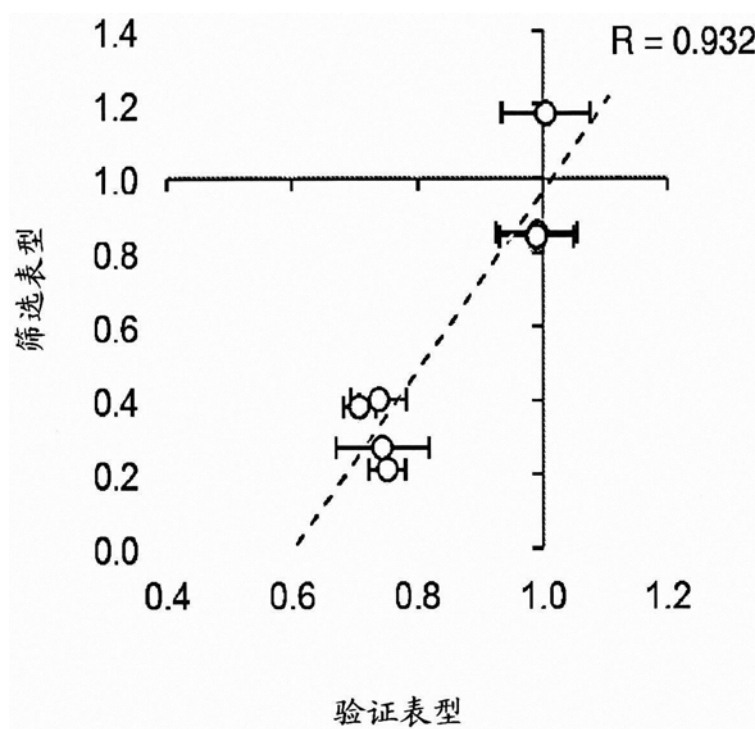


图13

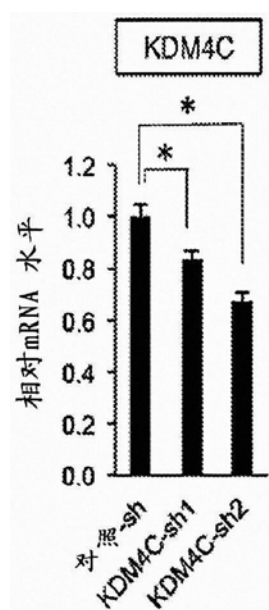


图14A

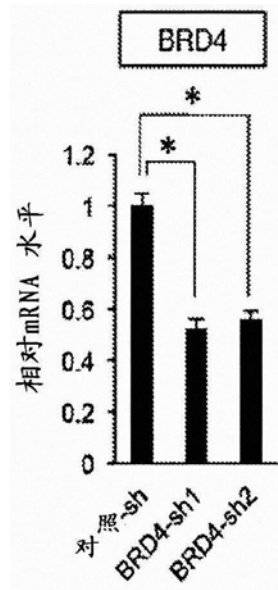


图14B

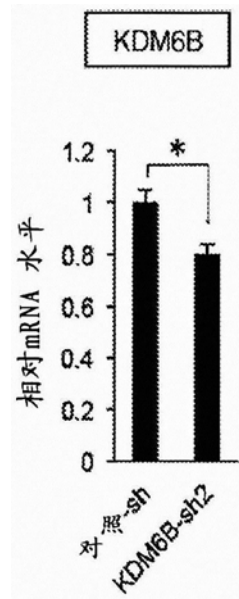
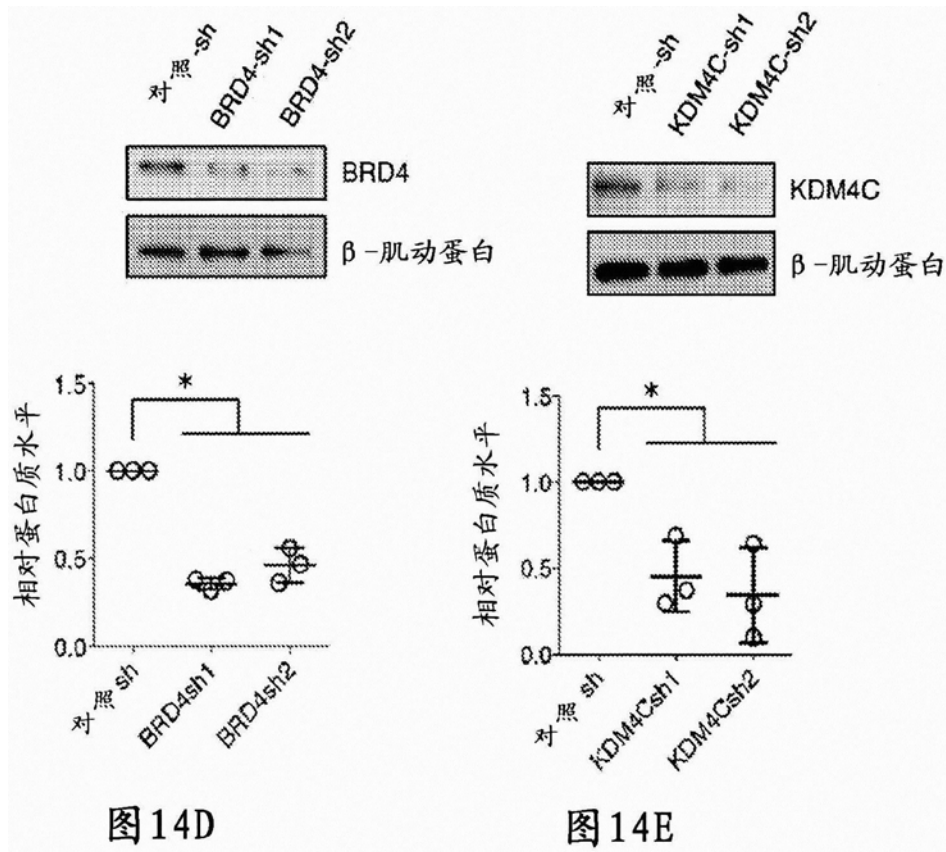


图14C



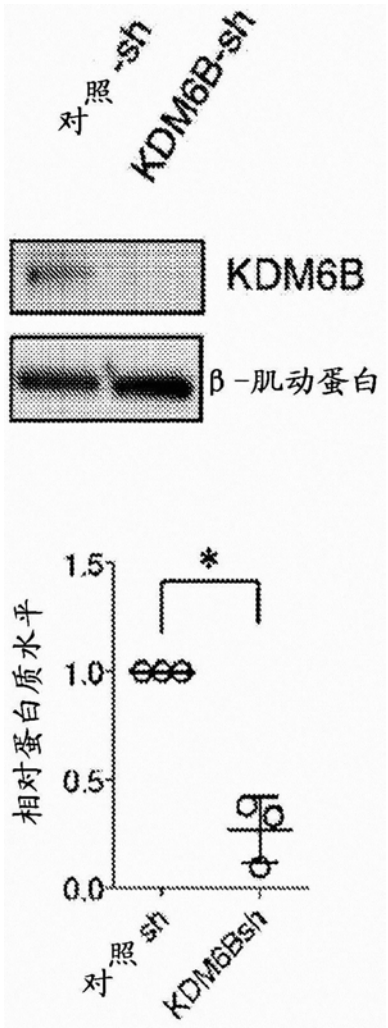


图14F

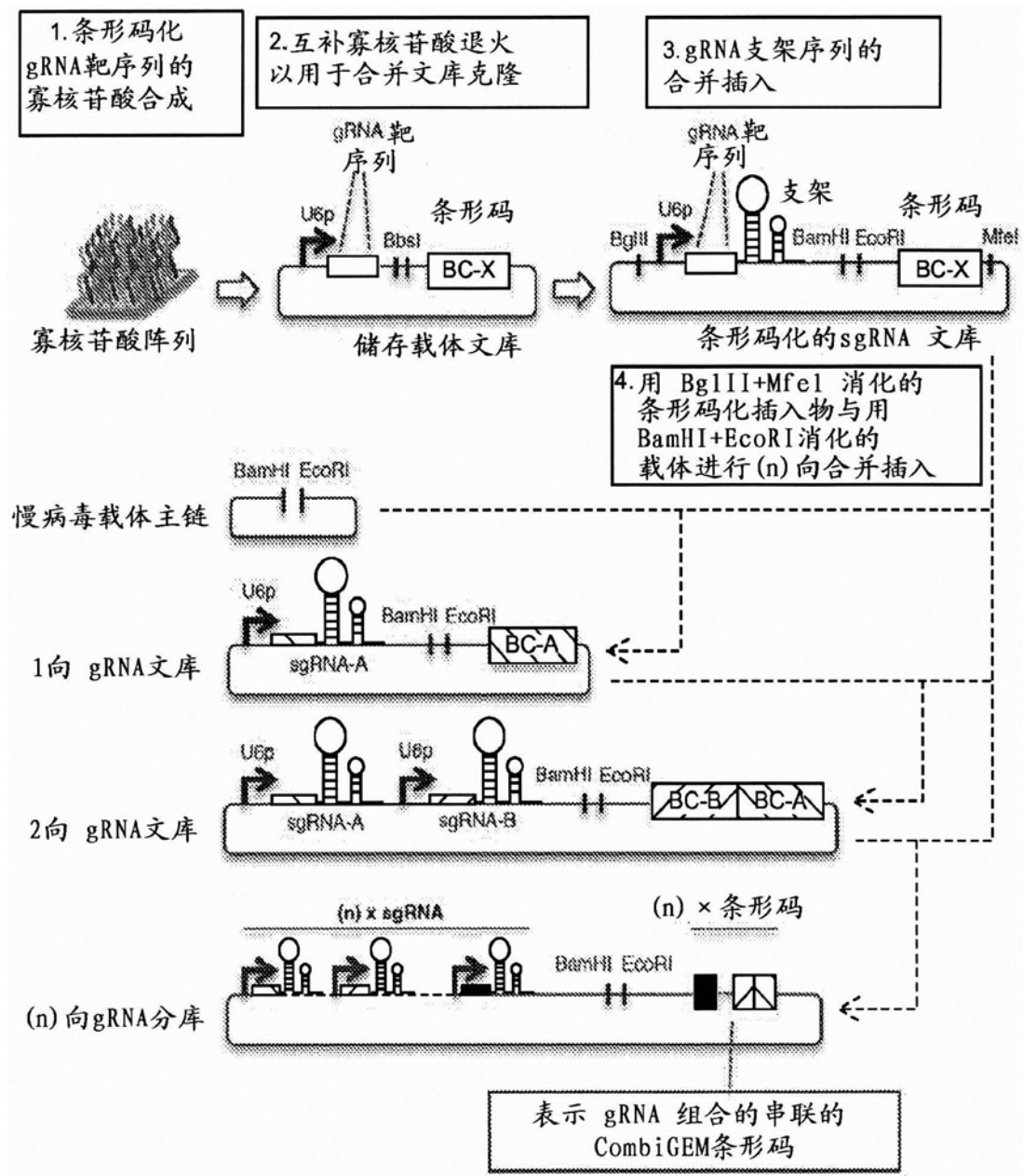


图15

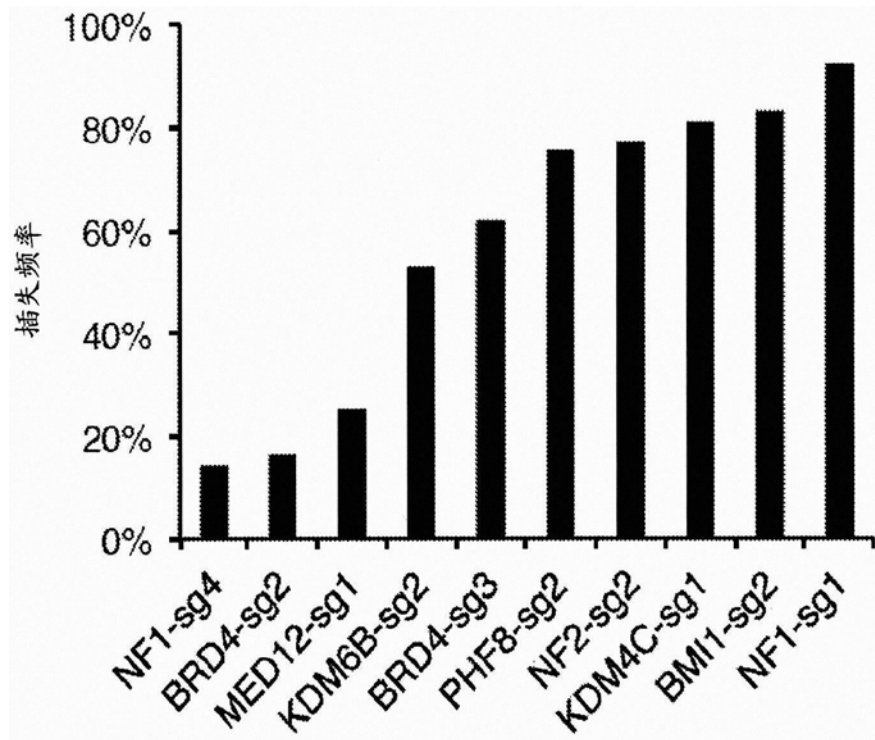


图16A

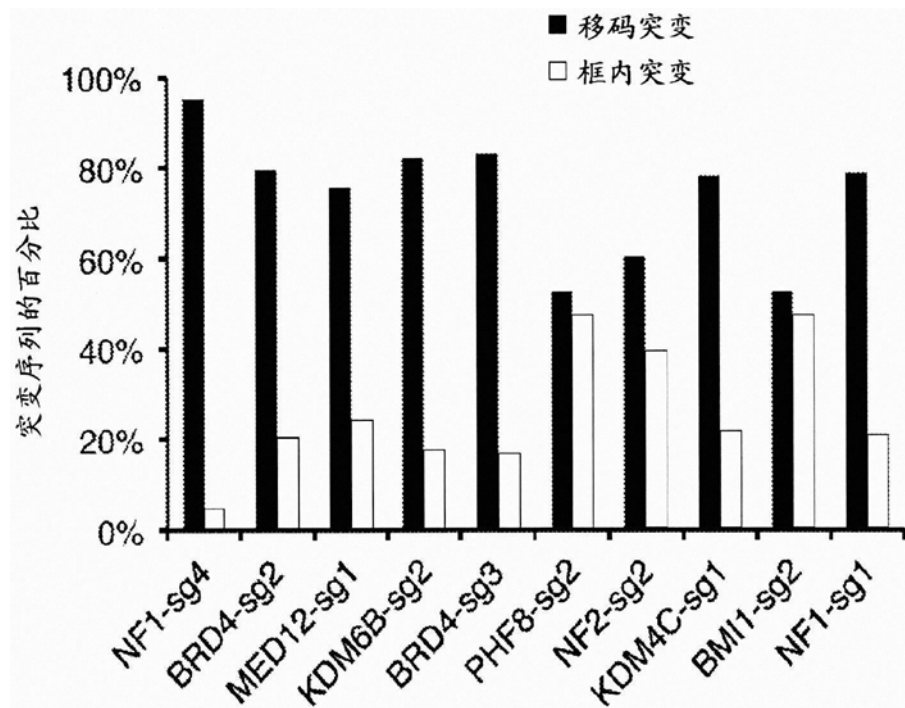


图16B

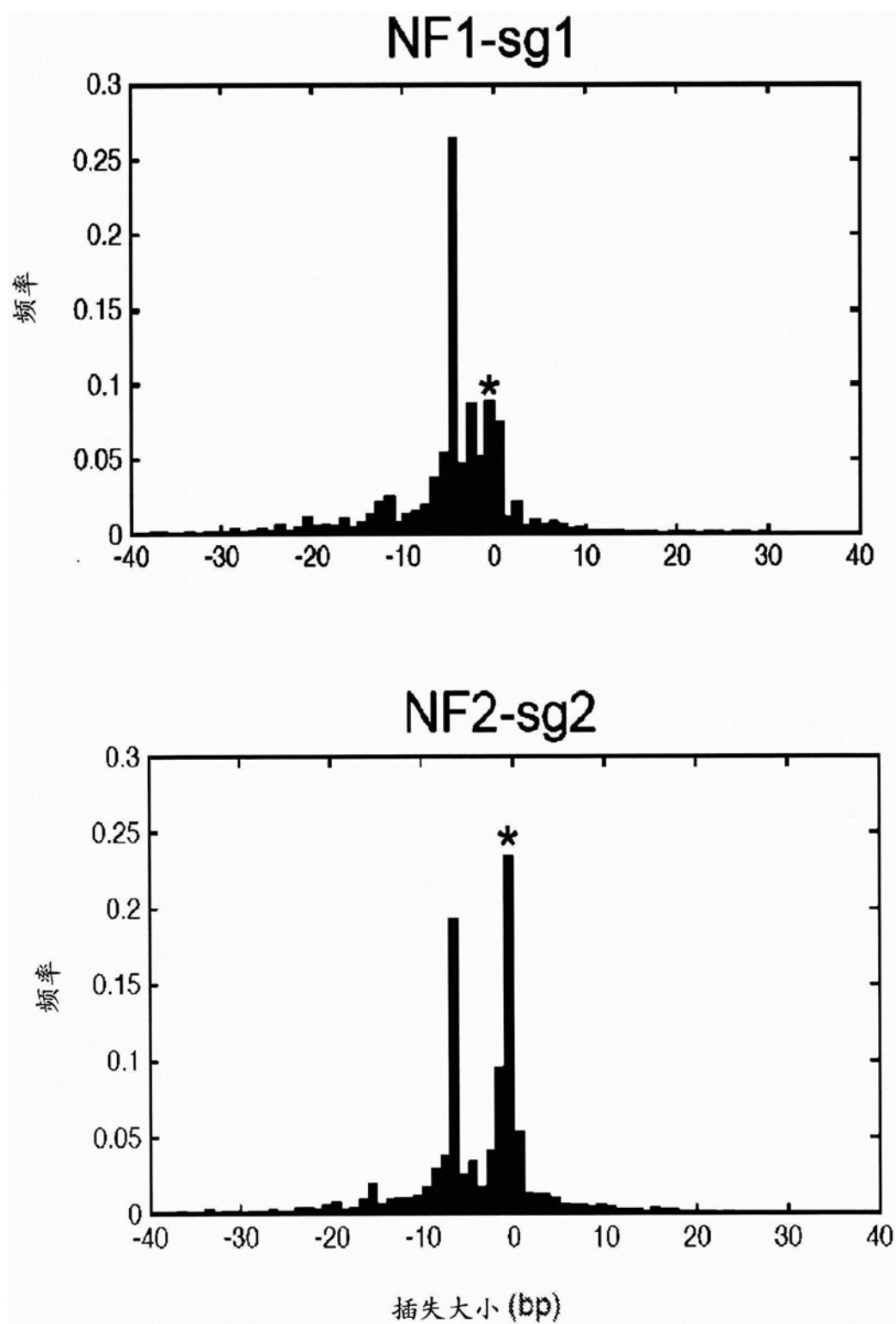


图16C

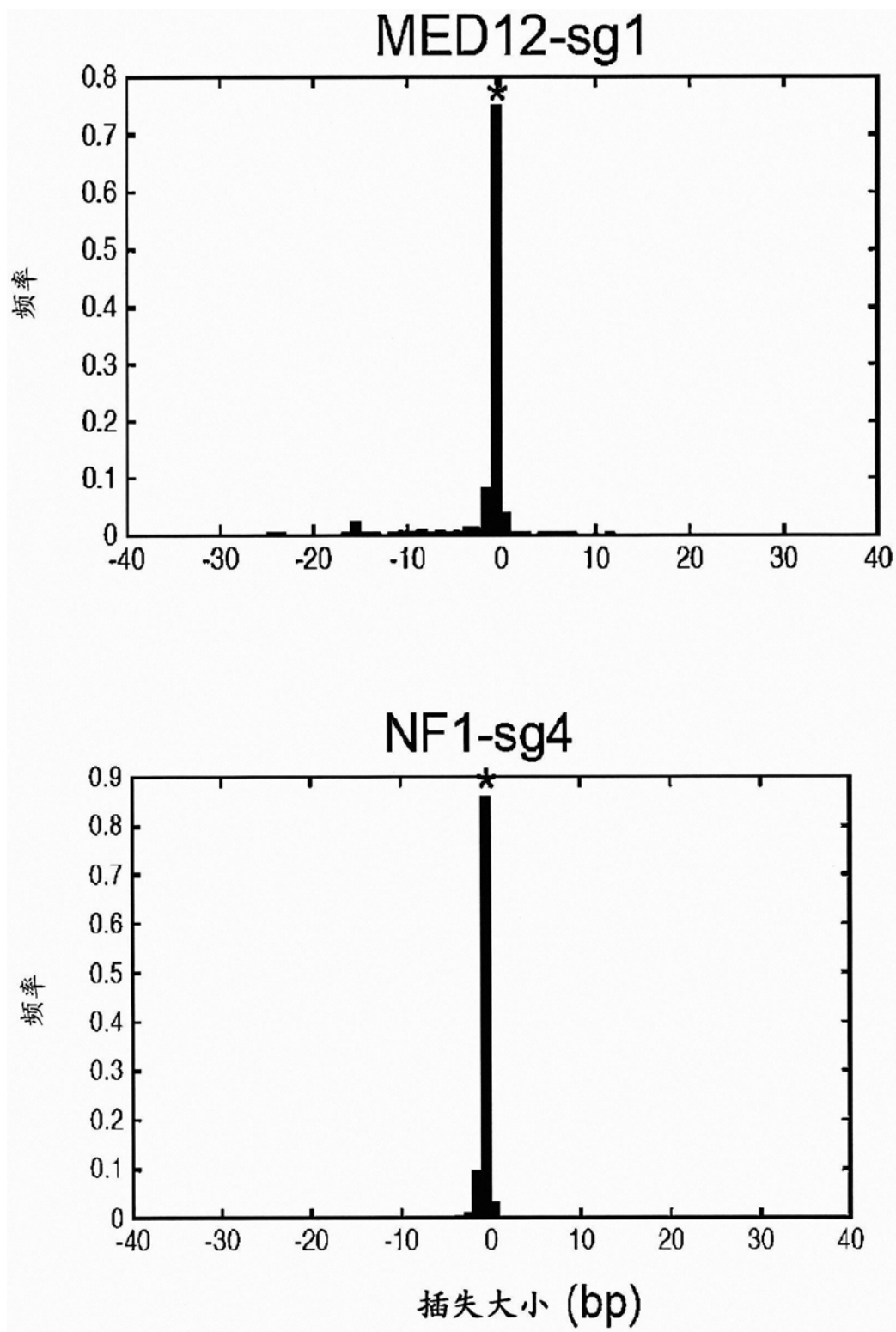


图16C (续)

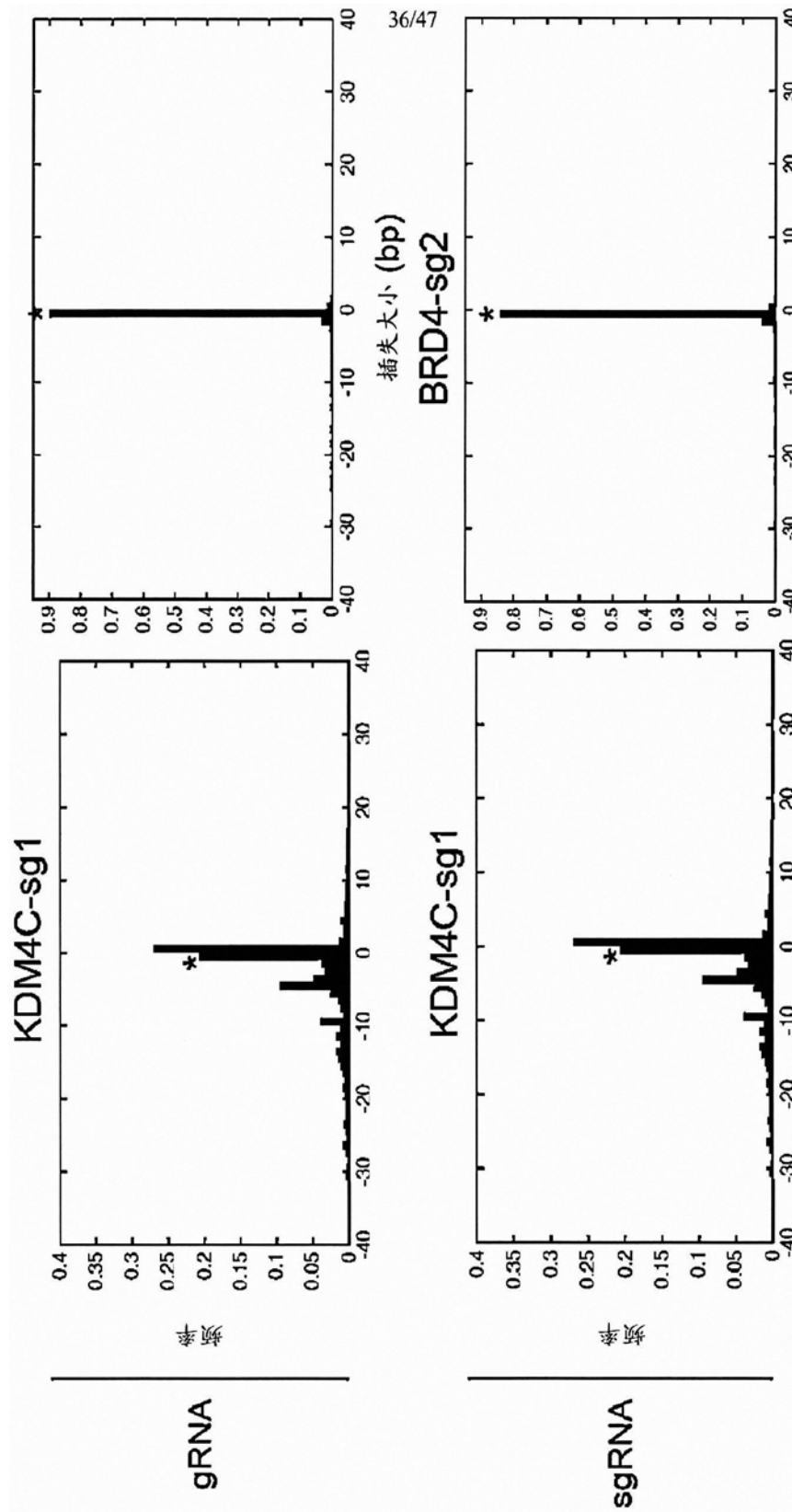


图16D

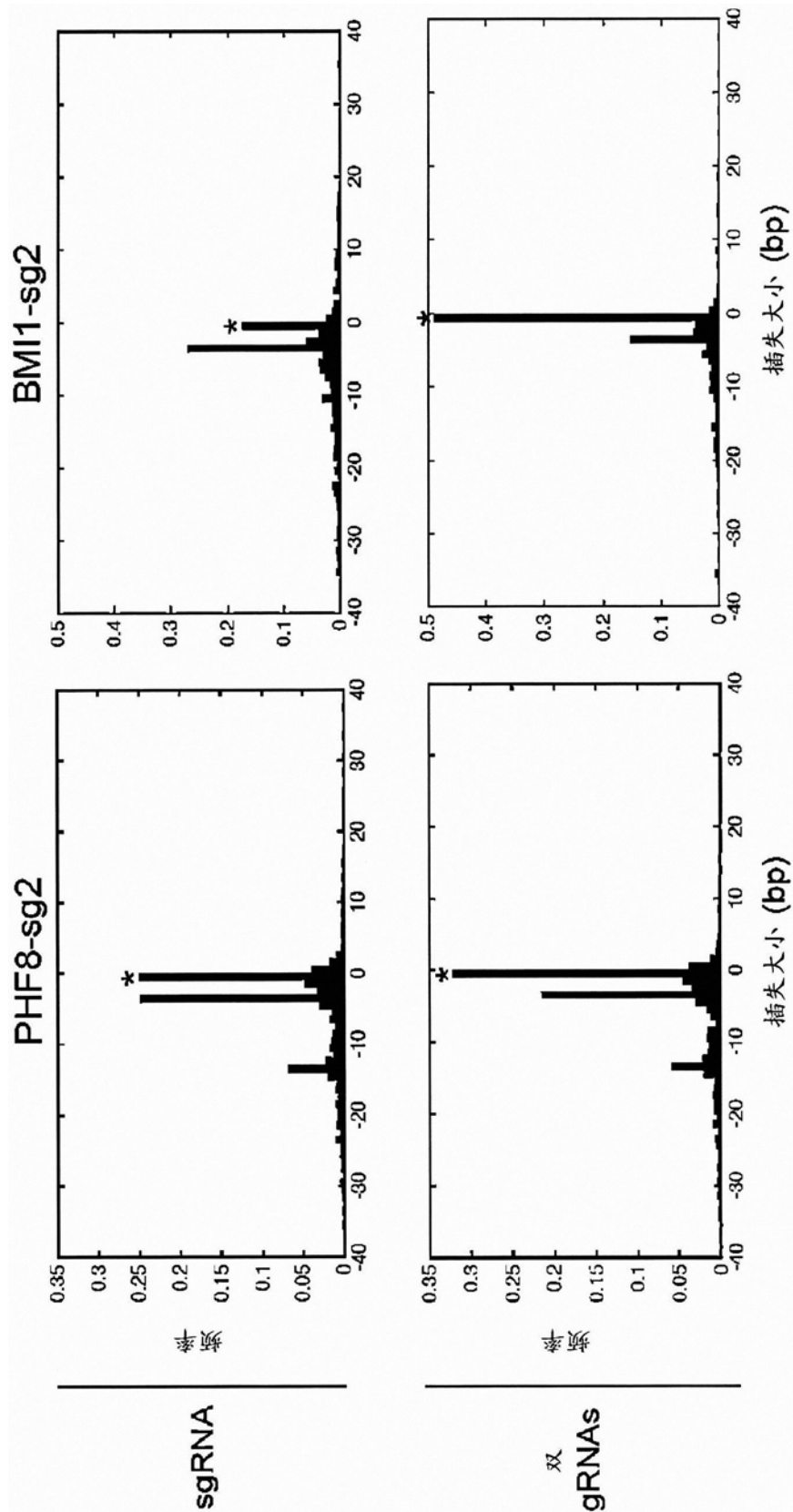


图16E

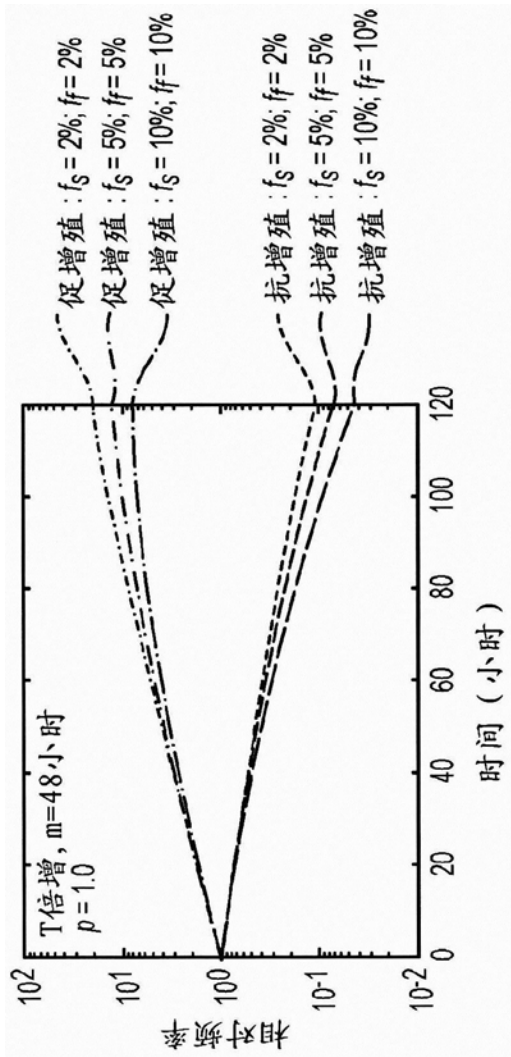


图17A

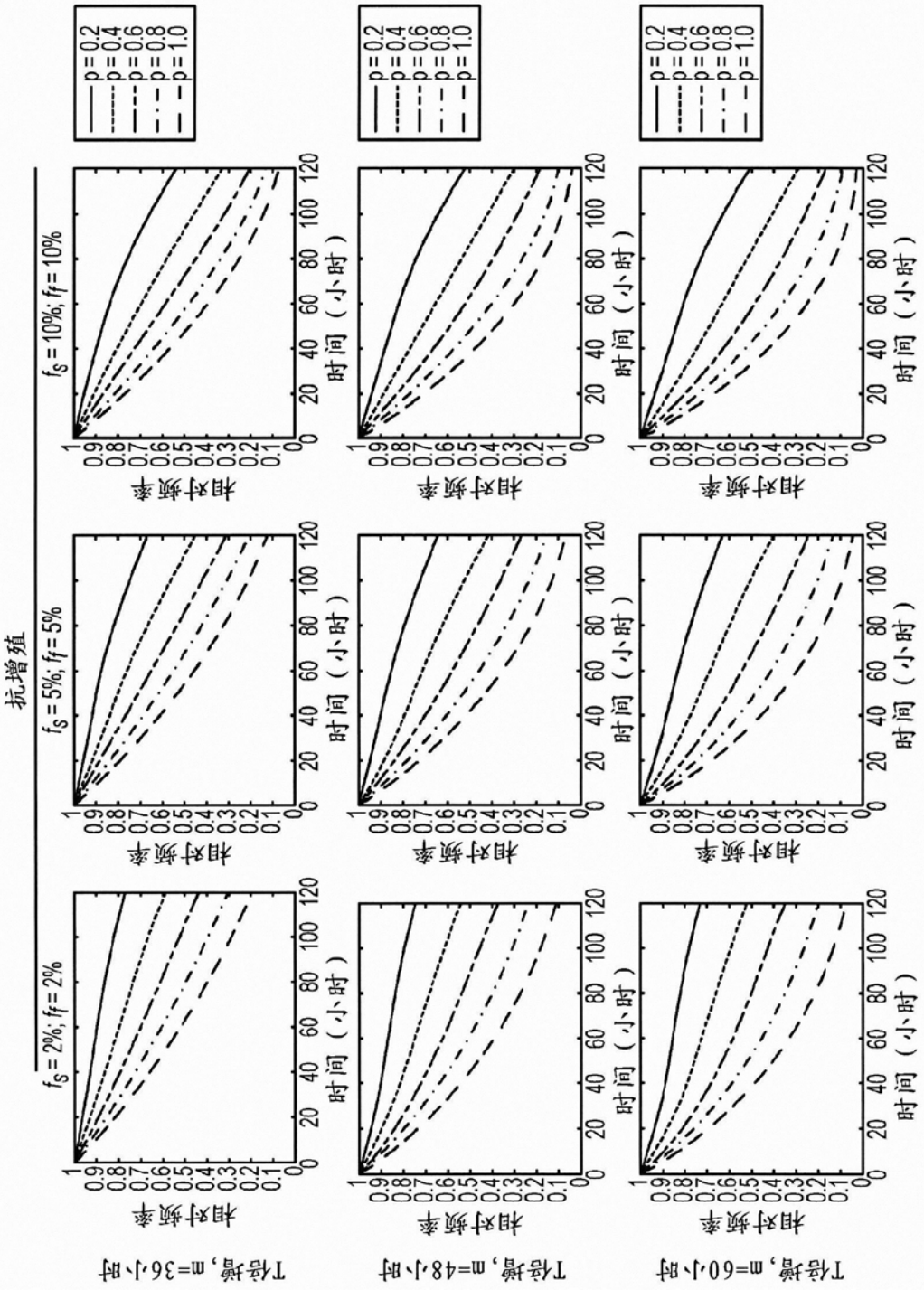


图17B

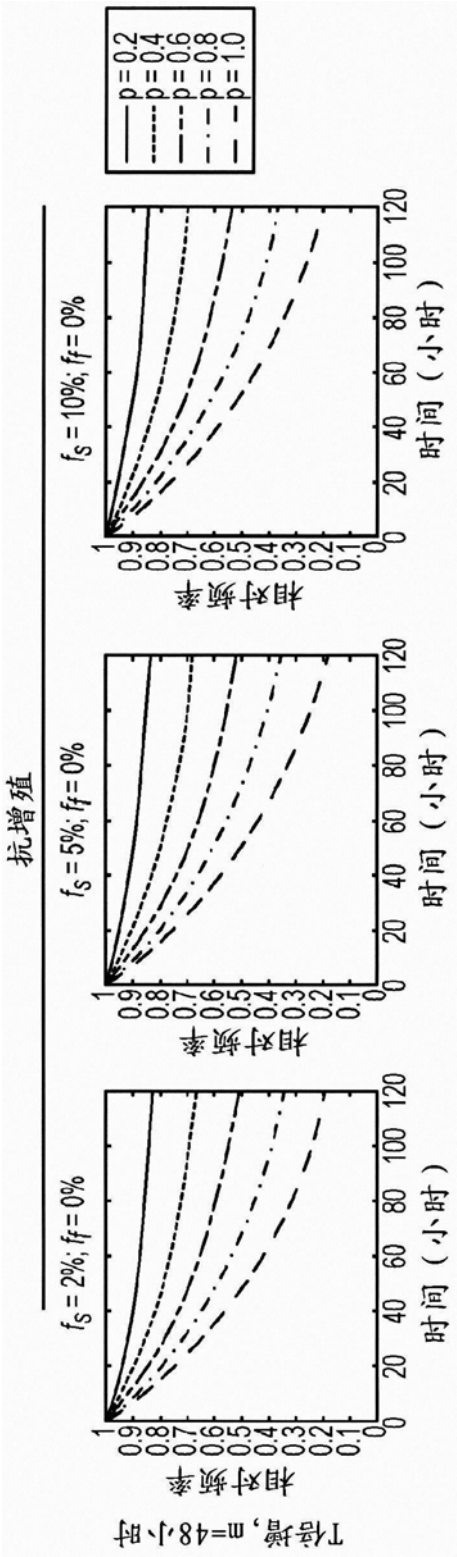


图17C

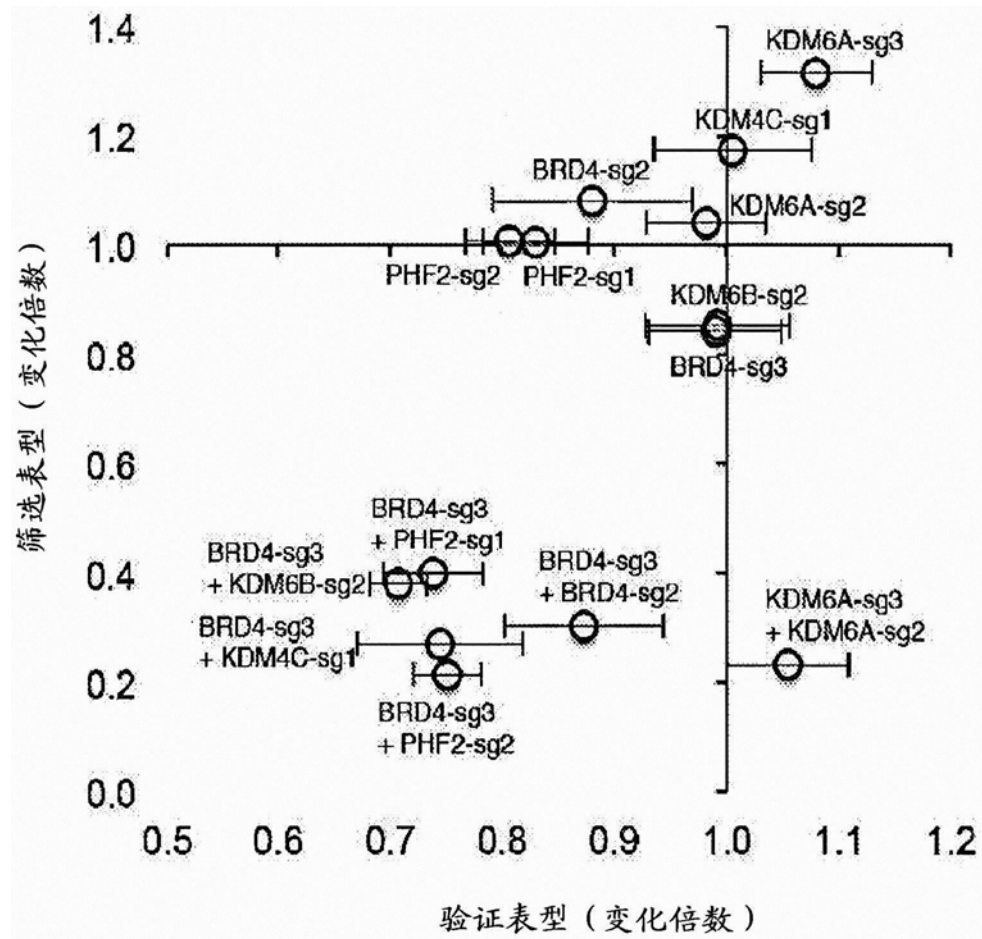


图18

	<u>指导序列</u>	<u>PAM</u>		
KDM4C-sg1	CCTTTGCAAGACCCGCACGA	TGG	80.85%	

	GCTTTGCCCCGAACCGCACGA	GAG	0.29%	
	CCTAGGCCAGACCTGCACGA	TGG	0.23%	
	GCTCTGGAAGACCCGCACCA	GGG	0.21%	外显子的
	CCTTATCAAGACCCACACCA	GAG	0.15%	
KDM6B-sg2	ATCCCCCTCCTCGTAGCGCA	TGG	52.74%	

	TGCGCCCTCCTCCTAGCGCA	TGG	3.74%	
	CTGCCCCCTCCTGGTAGCGCC	TGG	1.01%	
	CTGCTCCTCCTCGTAGCGCT	GGG	0.43%	
	AACCAGCTCCTCGTAGCTCA	GGG	0.32%	
	AGCCCGCTCCTCGTGGGGCA	CGG	0.19%	外显子的
	ACCGCCCTCCTCCTAGCTCA	GGG	0.17%	
	CTCCTCCTCCGCGTAGCGCT	TGG	0.14%	
	CTCCTCCTCCTCCTACCGCA	AGG	n.d.	
	ATTCTCTCCTGGTACCGCA	AGG	n.d.	
	CTCACCTCCTCCTAGCACA	TAG	n.d.	
BRD4-sg3	GGGAACAATAAAGAAGCGCT	TGG	61.97%	

	TGGAAAAACAAAGAAGAGCT	TGG	0.38%	
	GGGAAGTATAAGGAAGAGCT	CAG	0.26%	外显子的
	GTGAGCAATAAAGCAGCCCT	AAG	0.18%	
BRD4-sg2	TCTAGTCCATCCCCATTAC	TGG	16.65%	

	AATATTCCATTCCCCATTAC	TGG	0.33%	
	TGTTGTCCATACCTCATTAC	TGG	0.19%	外显子的
	TCTAGGTCATGCACCATTAC	TGG	0.19%	

	TCCACACCCTCCCCATTAC	TAG	0.25%	内含子的
	CCCATTCCTTCCCCATTAC	CAG	0.21%	

图19

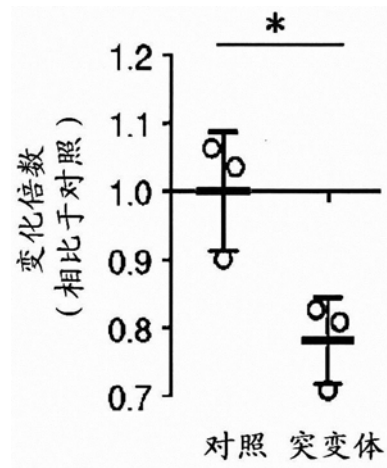


图20A

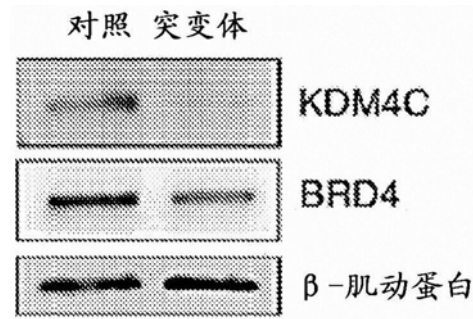


图20B

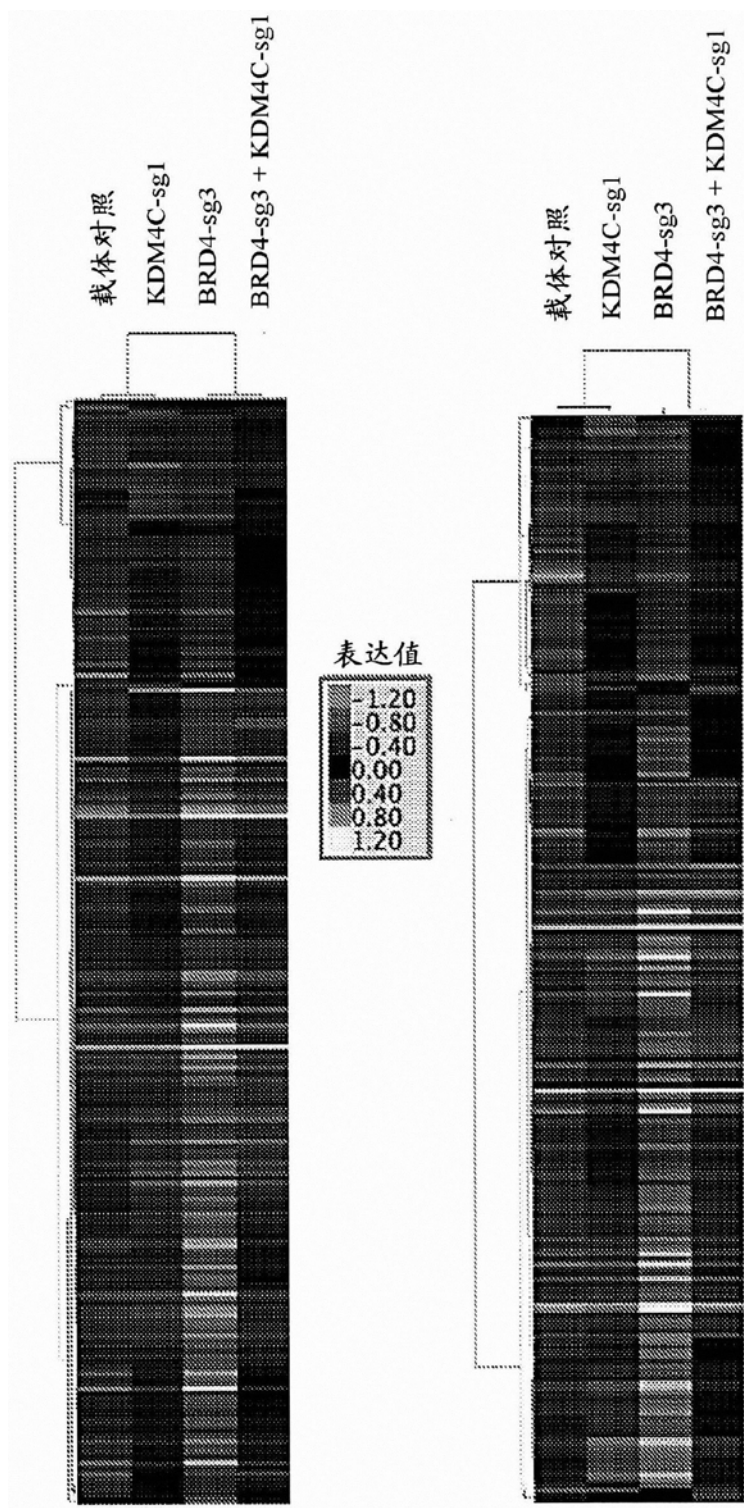


图21A

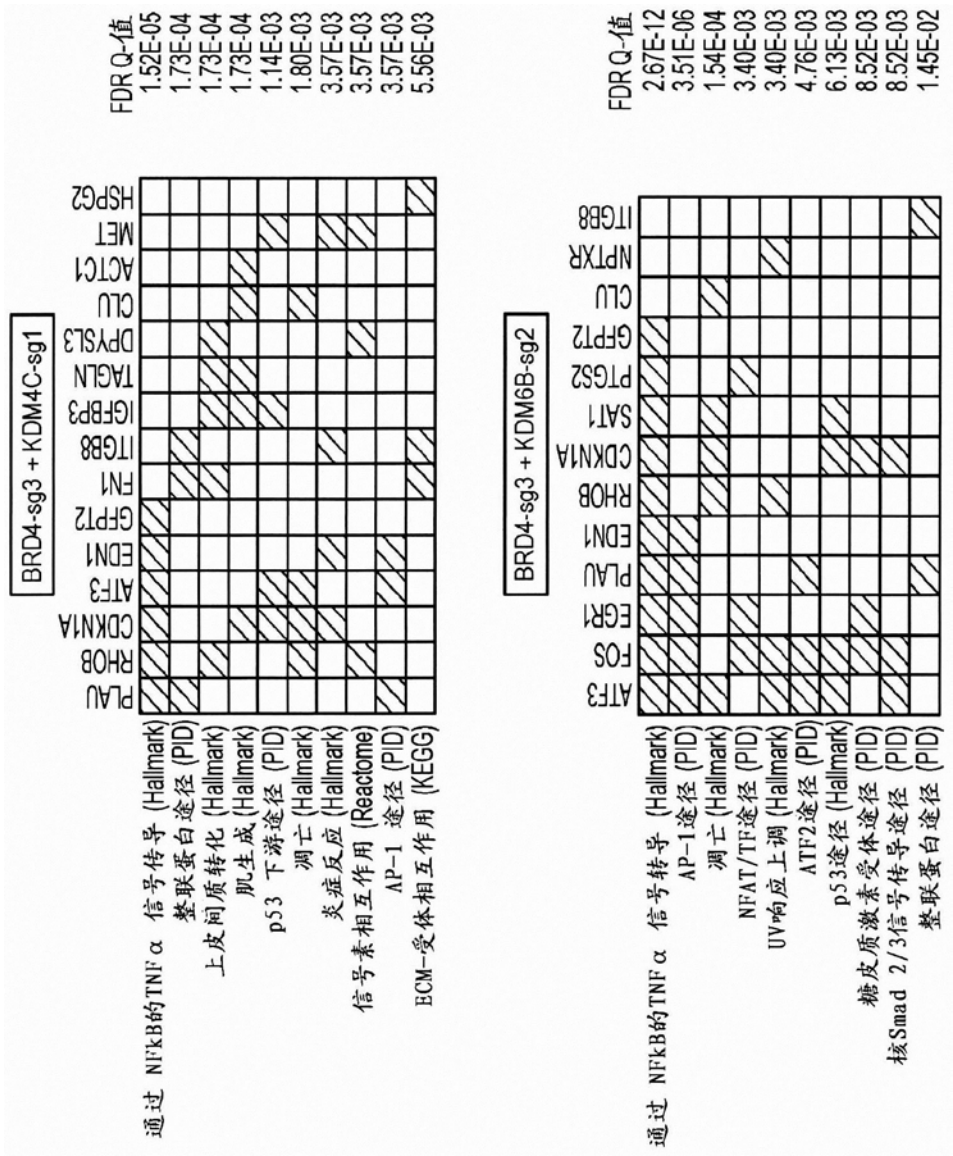


图21B

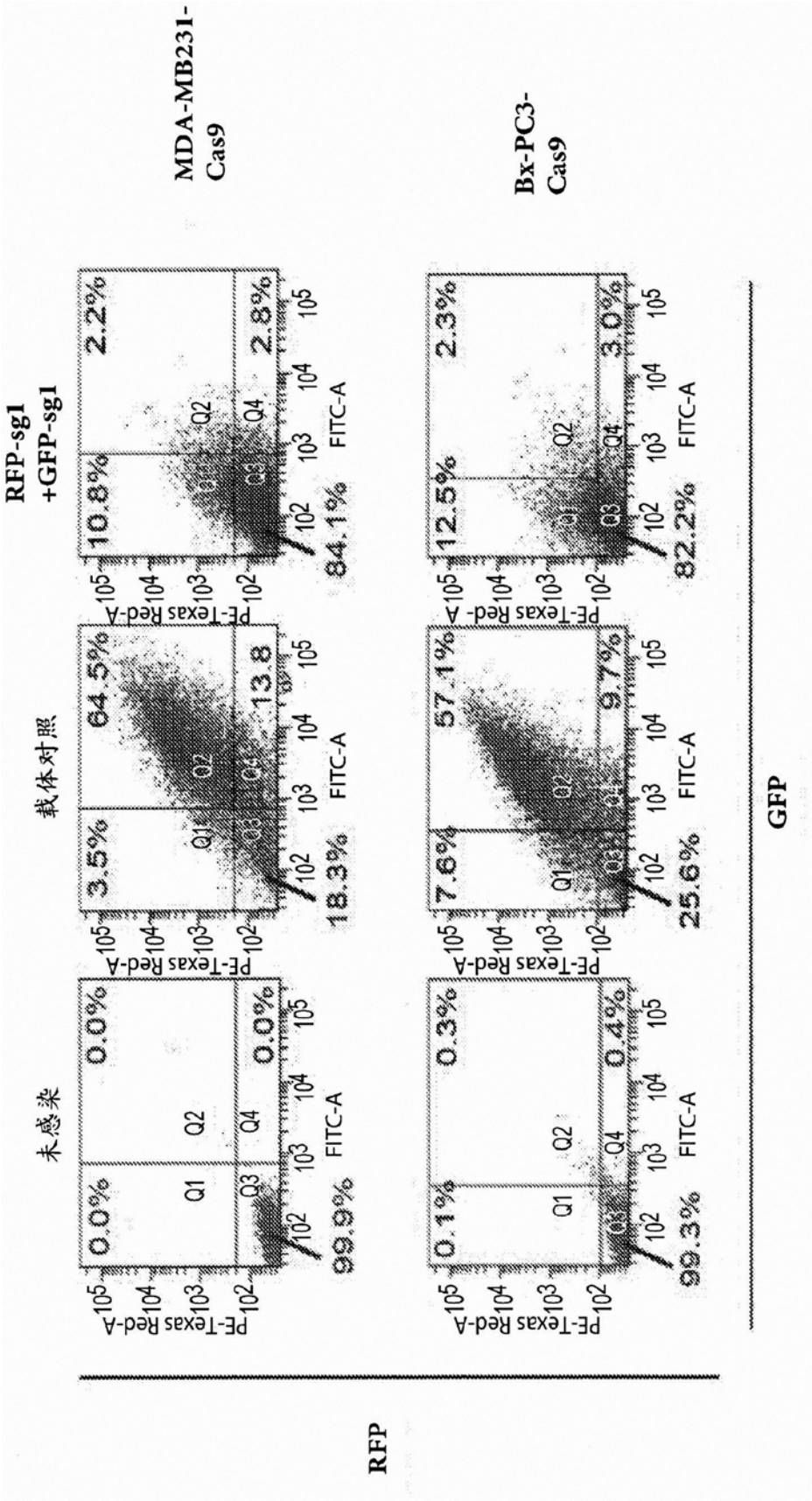


图22A

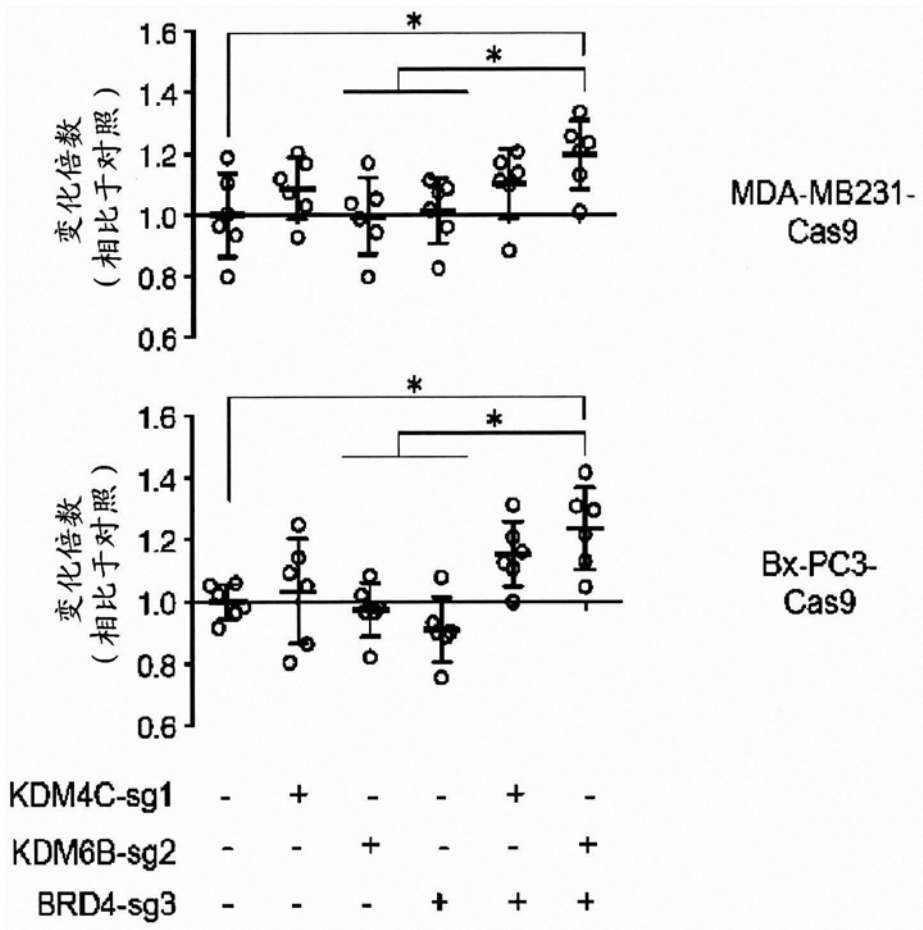


图22B