



# (12) 发明专利申请

(10) 申请公布号 CN 115698337 A

(43) 申请公布日 2023. 02. 03

(21) 申请号 202180041375.0

(22) 申请日 2021.06.01

(30) 优先权数据

63/036064 2020.06.08 US

(85) PCT国际申请进入国家阶段日

2022.12.08

(86) PCT国际申请的申请数据

PCT/EP2021/064652 2021.06.01

(87) PCT国际申请的公布数据

WO2021/249825 EN 2021.12.16

(71) 申请人 豪夫迈·罗氏有限公司

地址 瑞士巴塞尔

(72) 发明人 A·洛夫乔伊 M·洛伊泽

(74) 专利代理机构 中国专利代理(香港)有限公司  
72001

专利代理师 权陆军 张华

(51) Int.Cl.

C12Q 1/6883 (2006.01)

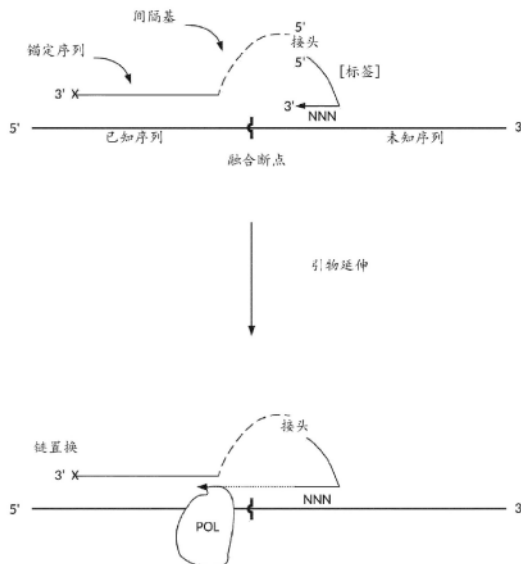
权利要求书4页 说明书41页 附图2页

## (54) 发明名称

用于检测基因组中的结构重排的方法和组合物

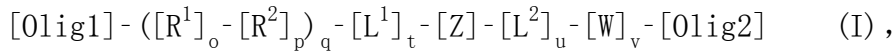
## (57) 摘要

本文公开了用于使用锁核酸引物来检测涉及未知融合配偶体的基因融合的组合物、试剂盒和方法。在一些实施方案中,所述组合物包含化合物,所述化合物包含至少两个通过5'至5'键合直接或间接地连接的核苷酸序列。在一些实施方案中,所述化合物进一步包含间隔部分和/或切割部分。



1. 一种检测核酸样品中的基因融合的方法,所述方法包括

(a) 使所述核酸样品与 (i) 具有聚合酶活性和链置换活性的核酸聚合酶和 (ii) 具有式 (I) 的化合物接触:



其中

o为0或1;

p为0或1;

q为0或1;

t为0、1或2;

u为0、1或2;

v为0或1;

R<sup>1</sup>为具有约1个与约24个之间核苷酸的寡核苷酸;

R<sup>2</sup>为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子;

L<sup>1</sup>和L<sup>2</sup>独立地为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子,并且任选地包含一个或多个羰基基团;

Z为选自以下的部分:三唑、二氢吡嗪、磷酸酯连接基、酰胺连接基、硫醚连接基、异噻唑啉、脞、脲醚和氯-s-三嗪连接基;

W为具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和、脂肪族或芳香族基团,任选地包含一个或多个选自O、N、S的杂原子,前提是W包含至少一个可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团;

Olig1为包含约1个与约30个之间核苷酸并且包含能够与已知融合配偶体杂交的锚定序列的寡核苷酸,并且其中Olig1具有不可延伸的3'端;并且

Olig2为包含约1个与约30个之间核苷酸并且包含可延伸的3'端的寡核苷酸;以及

(b) 用所述核酸聚合酶延伸Olig2的3'端,其中延伸产物包含未知融合配偶体的一部分、已知融合配偶体的一部分和融合断点的拷贝,由此形成所述基因融合的第一链拷贝。

2. 根据权利要求1所述的方法,其进一步包括形成所述基因融合的双链拷贝的文库;其中所述文库的形成包括:将衔接子连接到基因融合的拷贝,其中衔接子包含条形码和引物结合位点。

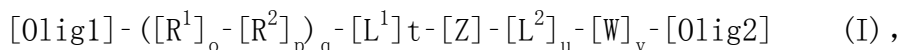
3. 根据权利要求1所述的方法,其进一步包括通过包括以下的方法扩增所述基因融合的拷贝:

(a) 将包含所述基因融合的拷贝的所述样品分成多个反应体积;其中每个反应体积包含能够与拷贝链和所述拷贝链的互补序列杂交的正向和反向扩增引物,以及第一可检测地标记的探针;

(b) 进行扩增反应,其中所述反应包括用所述探针进行检测的步骤;

(c) 确定已检测到所述探针的反应体积的数量,由此检测所述基因融合。

4. 一种具有式 (I) 的化合物,



其中

o为0或1;

p为0或1;

q为0或1;

t为0、1或2;

u为0、1或2;

v为0或1;

R<sup>1</sup>为具有约1个与约24个之间核苷酸的寡核苷酸;

R<sup>2</sup>为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子;

L<sup>1</sup>和L<sup>2</sup>独立地为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子,并且任选地包含一个或多个羰基基团;

Z为选自以下的部分:三唑、二氢吡嗪、磷酸酯连接基、酰胺连接基、硫醚连接基、异噁唑啉、脞、脲和氯-s-三嗪连接基;

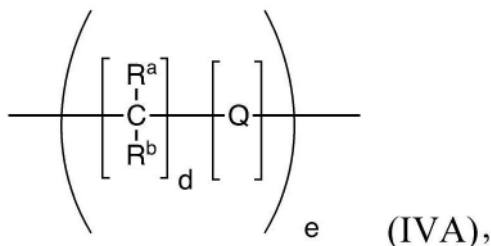
W为具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和、脂肪族或芳香族基团,任选地包含一个或多个选自O、N、S的杂原子,前提是W包含至少一个可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团;

Olig1为具有约1个与约30个之间核苷酸的寡核苷酸,并且其中Olig1具有不可延伸的3'端;并且

Olig2为具有约1个与约30个之间核苷酸的寡核苷酸,并且其中Olig2具有可延伸的3'端。

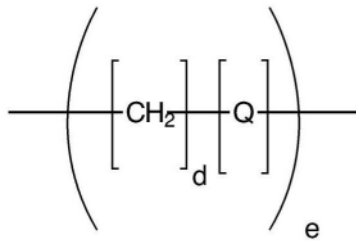
5. 根据权利要求4所述的化合物,其中R<sup>2</sup>包含具有2个与约32个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子,并且任选地包含一个或多个羰基基团。

6. 根据权利要求4所述的化合物,其中R<sup>2</sup>包含具有式(IVA)的结构的部分:



其中d和e为各自独立地在1至32的范围内的整数;Q为键、O、S、N(R<sup>c</sup>) (R<sup>d</sup>)或季胺(N<sup>+</sup>H(R<sup>c</sup>)(R<sup>d</sup>));R<sup>a</sup>和R<sup>b</sup>独立地为H、C<sub>1</sub>-C<sub>4</sub>烷基基团、F、Cl或N(R<sup>c</sup>) (R<sup>d</sup>);并且R<sup>c</sup>和R<sup>d</sup>独立地为CH<sub>3</sub>或H。

7. 根据权利要求4所述的化合物,其中R<sup>2</sup>包含具有式(IVB)的结构的部分:



(IVB)

其中d和e为各自独立地在1至32的范围内的整数;Q为键、O、S或N( $R_c$ )( $R_d$ );并且 $R_c$ 和 $R_d$ 独立地为 $\text{CH}_3$ 或H。

8. 根据权利要求4所述的化合物,其中 $L^1$ 或 $L^2$ 中的至少一者包含具有1个与约4个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子,并且任选地包含一个或多个羰基基团。

9. 根据权利要求4所述的化合物,其中 $o+p=1$ ,并且q为1。

10. 根据权利要求4所述的化合物,其中o为0并且p和q二者为1, $R^1$ 包含至少一个PEG基团,并且 $L^1$ 包含至少一个羰基部分。

11. 根据权利要求4所述的化合物,其中Olig2包含条形码。

12. 根据权利要求4所述的化合物,其中Olig2包含通用引物结合位点。

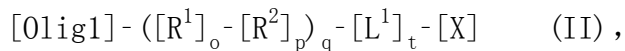
13. 根据权利要求4所述的化合物,其中v为0,并且Olig2包含切割位点,所述切割位点包含至少一个含尿嘧啶的核苷酸。

14. 根据权利要求4所述的化合物,其中Olig2包含随机核苷酸序列。

15. 一种检测已知融合配偶体和未知融合配偶体之间的基因融合的试剂盒,所述试剂盒包括根据权利要求30至63中任一项所述的化合物以及聚合酶。

16. 一种试剂盒,其包括:

(a) 具有式(II)的第一化合物:



其中

o为0或1;

p为0或1;

q为1或2;

t为0、1或2;

$R^1$ 为具有1个与约24个之间核苷酸的寡核苷酸;

$R^2$ 为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子;

$L^1$ 为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子,并且任选地包含一个或多个羰基基团;

X为二苯并环辛炔、反式环辛烯、炔烃、烯炔、叠氮化物、四嗪、马来酰亚胺、N-羟基琥珀酰亚胺、硫醇、1,3-硝酮、醛、酮、肼、羟胺、氨基基团或亚磷酰胺;并且

Olig1为具有约1个与约30个之间核苷酸的寡核苷酸;

(b) 具有式(III)的第二化合物:

$[Y]-[L^2]_u-[W]_v-[Olig2]$  (III),

其中

u为0、1或2;

v为0或1;

Y为二苯并环辛炔、反式环辛烯、炔烃、烯烃、叠氮化物、四嗪、马来酰亚胺、N-羟基琥珀酰亚胺、硫醇、1,3-硝酮、醛、酮、肼、羟胺、氨基基团或亚磷酰胺;

$L^2$ 为具有1个与16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子,并且任选地包含一个或多个羰基基团;

W为具有1个与12个之间碳原子的经取代或未取代的、饱和或不饱和、脂肪族或芳香族基团,任选地包含一个或多个选自O、N、S的杂原子,前提是W包含可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团;并且

Olig2为具有约1个与约30个之间核苷酸的寡核苷酸。

17. 根据权利要求4至14中任一项所述的化合物或根据权利要求15至15所述的试剂盒在对核酸分子进行测序中的用途。

## 用于检测基因组中的结构重排的方法和组合物

### 技术领域

[0001] 本公开涉及基因组学领域。更具体地，本公开涉及检测基因组重排的领域。

### 背景技术

[0002] 基因融合在癌症中很常见。一些基因融合是癌症驱动突变，针对其已经开发了靶向疗法。检测基因融合的能力可有助于检测和诊断癌症，跟踪肿瘤负荷随时间的变化，以及为癌症患者确定最佳的个体化治疗。检测基因组重排的传统方法涉及繁琐的多步骤程序，诸如单倍型融合PCR和连接单倍型，参见Turner等人，(2008) Long range, high throughput haplotype determination via haplotype fusion PCR and ligation haplotyping, Nucl. Acids Res. 36:e82。最近的下一代基于测序的技术能够识别各种基因融合。然而，这需要大量的测序来捕获和验证足够数量的融合序列。这种方法的成本和复杂性使其不适合临床使用。

[0003] 对于一些基因，检测基因融合会因多种融合配偶体的出现而变得更加复杂。例如，神经生长原肌球蛋白受体激酶基因 (NTRK 1、2和3) 可以与任意数量的N末端 (5' -) 配偶体融合，参见Solomon等人 (2019) Identifying patients with NTRK fusion cancer, Ann. Oncol. Nov; 30 Suppl 8: viii16-viii22。由于存在针对活化NTRK的有效疗法，因此用于识别具有NTRK基因融合的合格患者的经济有效的临床试验至关重要。类似地，成纤维细胞生长因子受体基因 (FGFR 2和3) 可以与任意数量的C末端 (3' -) 配偶体融合，产生组成型活性受体激酶蛋白，参见Facchinetti等人 (2020) Facts and New Hopes on Selective FGFR Inhibitors in Solid Tumors, Clin. Cancer Res. 2020 Feb 15; 26 (4) : 764-774。随着多种FGFR激酶抑制剂正在开发中，需要用来识别各种肿瘤类型中具有FGFR基因融合的合格患者的实践临床试验。

### 发明内容

[0004] 基于上述，需要以较低的成本利用较少测序来鉴定基因融合，以增加患者获得可能挽救生命的疗法的机会。

[0005] 本公开涉及用于检测核酸样品中的一种或多种基因融合的组合物、试剂盒和方法。在一些实施例中，本公开提供一种或多种各自具有式 (I) 的化合物：

[0006]  $[Olig1] - ([R^1]_o - [R^2]_p)_q - [L^1]_t - [Z] - [L^2]_u - [W]_v - [Olig2]$  (I)，

[0007] 其中

[0008] o为0或1；

[0009] p为0或1；

[0010] q为0或1；

[0011] t为0、1或2；

[0012] u为0、1或2；

[0013] v为0或1；

[0014]  $R^1$ 为具有约1个与约24个之间核苷酸的寡核苷酸；

[0015]  $R^2$ 为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代；

[0016]  $L^1$ 和 $L^2$ 独立地为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地包含一个或多个选自O、N或S的杂原子，并且任选地包含一个或多个羰基基团；

[0017] Z为选自以下的部分：三唑、二氢吡嗪、磷酸酯键合、酰胺键合、硫醚键合、异噁唑啉、脞、脞醚和氯-s-三嗪键合；

[0018] W为具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和、脂肪族或芳香族基团，任选地被一个或多个选自O、N、S的杂原子取代，前提是W包含至少一个可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团；

[0019] Olig1为包含约1个与约30个之间核苷酸的寡核苷酸；并且

[0020] Olig2为包含约1个与约30个之间核苷酸的寡核苷酸。

[0021] 在一些实施例中，式(I)化合物可用于促进基因融合的检测。在这方面，本公开还涉及使用一种或多种式(I)化合物检测基因融合的方法。在一些实施例中，式(I)化合物促进基因融合的捕获，其中一种融合配偶体用于样品(例如组织学样品、细胞学样品等)中的一种或多种基因融合的检测、扩增和/或测序。本文进一步描述了本公开的这些和其他方面。

[0022] 在本公开的第一方面是一种检测核酸样品中的基因融合的方法，所述方法包括：  
(a)使样品与聚合酶(例如具有聚合酶活性和链置换活性的核酸聚合酶)以及与具有式(I)的化合物接触：

[0023]  $[Olig1] - ([R^1]_o - [R^2]_p)_q - [L^1]_t - [Z] - [L^2]_u - [W]_v - [Olig2]$  (I)，

[0024] 其中

[0025] o为0或1；

[0026] p为0或1；

[0027] q为0或1；

[0028] t为0、1或2；

[0029] u为0、1或2；

[0030] v为0或1；

[0031]  $R^1$ 为具有约1个与约24个之间核苷酸的寡核苷酸；

[0032]  $R^2$ 为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代；

[0033]  $L^1$ 和 $L^2$ 独立地为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地包含一个或多个选自O、N或S的杂原子，并且任选地包含一个或多个羰基基团；

[0034] Z为选自以下的部分：三唑、二氢吡嗪、磷酸酯键合、酰胺键合、硫醚键合、异噁唑啉、脞、脞醚和氯-s-三嗪键合；

[0035] W为具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和、脂肪族或芳香族基团，任选地被一个或多个选自O、N、S的杂原子取代，前提是W包含至少一个可光切

割的、可酶促切割的、可化学切割的、或pH敏感的基团；

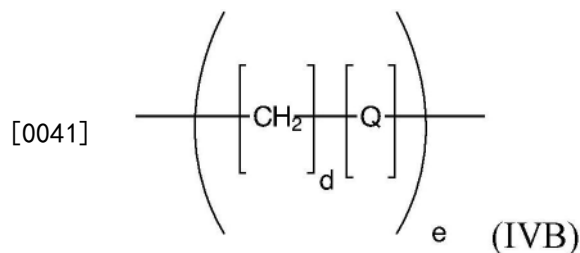
[0036] Olig1为包含约1个与约30个之间核苷酸并且包含能够与已知融合配偶体杂交的锚定序列的寡核苷酸，并且其中Olig1具有不可延伸的3'端；并且

[0037] Olig2为包含约1个与约30个之间核苷酸并且包含可延伸的3'端的寡核苷酸；以及

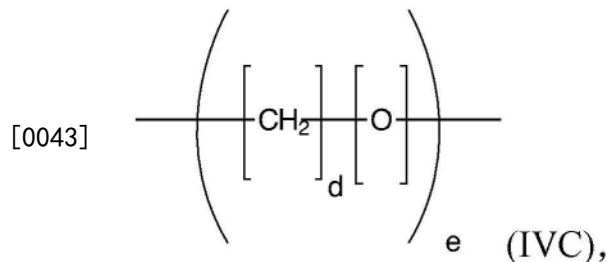
[0038] (b)用聚合酶延伸具有式(I)的化合物的Olig2的3'端，从而产生延伸产物。在一些实施例中，延伸产物包含未知融合配偶体的一部分、已知融合配偶体的一部分和融合断点的拷贝，由此形成基因融合的第一链拷贝。

[0039] 在一些实施例中，Olig2包含随机序列。在一些实施例中，随机序列包含2个与20个之间核苷酸。

[0040] 在一些实施例中， $o+p=1$ ，并且 $q$ 为1。在一些实施例中， $R^2$ 包含具有式(IVB)结构的部分：



[0042] 其中 $d$ 和 $e$ 为各自独立地在1至32的范围内的整数； $Q$ 为键、O、S或N( $R_c$ )( $R_d$ )；并且 $R_c$ 和 $R_d$ 独立地为 $CH_3$ 或H。在一些实施例中， $R^2$ 包含具有式(IVC)结构的部分：



[0044] 其中 $d$ 和 $e$ 为各自独立地在1至32的范围内的整数。在一些实施例中， $d$ 和 $e$ 在1至16的范围内。在一些实施例中， $d$ 和 $e$ 在2至8的范围内。在一些实施例中，方法进一步包括通过复制第一链拷贝形成基因融合的第二链拷贝，由此形成基因融合的双链拷贝。

[0045] 在一些实施例中， $R^1$ 包含约2个与约9个之间核苷酸。在一些实施例中， $R^1$ 包含4个与8个之间核苷酸。

[0046] 在一些实施例中， $v$ 为1。在一些实施例中，方法进一步包括切割连接引物的可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团。

[0047] 在一些实施例中， $v$ 为0并且Olig2包含切割位点，所述切割位点包括含尿嘧啶的核苷酸。

[0048] 在一些实施例中， $L^1$ 或 $L^2$ 中的至少一个包含具有1个与约4个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在一些实施例中，脂肪族基团是直链的。在一些实施例中，脂肪族基团是直链的且未取代的。在一些实施例中，脂肪族基团是直链的且未取代的并且包含一个羰基基团。在一些实施例中，脂肪族基团是直链的且经取代的。

[0049] 在一些实施例中，方法进一步包括对基因融合的拷贝进行测序。在一些实施例中，



方法进一步包括形成基因融合的双链拷贝的文库。在一些实施例中,形成文库包括:将衔接子连接至基因融合的拷贝,其中衔接子包含条形码和引物结合位点。在一些实施例中,方法进一步包括通过通用扩增来扩增形成的文库的至少一部分。在一些实施例中,方法进一步包括对形成的文库的至少一部分进行测序。在一些实施例中,条形码包含唯一分子条形码(unique molecular barcode,UID),并且测序包括通过UID将文库核酸序列分组为家族,确定每个家族的共有序列读段,以及将共有序列读段与参考基因组进行比对,从而确定基因融合的序列。

[0050] 在一些实施例中,方法进一步包括通过包括以下的方法扩增拷贝链:(a)将包含拷贝链的样品分成多个反应体积;其中每个反应体积包含能够与拷贝链和拷贝链的互补链杂交的正向和反向扩增引物,以及第一可检测标记的探针;(b)进行扩增反应,其中所述反应包括用所述探针进行检测的步骤;(c)确定已检测到探针的反应体积的数量,从而检测基因融合。在一些实施例中,反应体积是液滴。在一些实施例中,可检测标记包含荧光团和猝灭剂的组合。

[0051] 在一些实施例中,通过使样品与两种或更多种具有式(I)的化合物接触来检测样品中的多重融合。在一些实施例中,两种或更多种式(I)化合物中的每一种的Olig1能够与选自以下的基因杂交:ALK、PPARG、BRAF、EGFR、FGFR1、FGFR2、FGFR3、MET、NRG1、NTRK1、NTRK2、NTRK3、RET、ROS1、AXL、PDGFRA、PDGFB、ABL1、ABL2、AKT1、AKT2、AKT3、ARHGAP26、BRD3、BRD4、CRLF2、CSF1R、EPOR、ERBB2、ERBB4、ERG、ESR1、ESRRA、ETV1、ETV4、ETV5、ETV6、EWSR1、FGR、IL2RB、INSR、JAK1、JAK2、JAK3、KIT、MAML2、MAST1、MAST2、MSMB、MUSK、MYB、MYC、NOTCH1、NOTCH2、NUMBL、NUT、PDGFRB、PIK3CA、PKN1、PRKCA、PRKCB、PTK2B、RAF1、RARA、RELA、RSP02、RSP03、SYK、TERT、TFE3、TFEB、THADA、TMPRSS2、TSLP、TY、BCL2、BCL6、BCR、CAMTA1、CBFB、CCNB3、CCND1、CIC、CRFL2、DUSP22、EPCI、FOXO1、FUS、GLI1、GLIS2、HMGA2、JAZF1、KMT2A、MALT1、MEAF6、MECOM、MKL1、MKL2、MTB、NCOA2、NUP214、NUP98、PAX5、PDGFB、PICALM、PLAG1、RBM15、RUNX1、RUNX1T1、SS18、STAT6、TAF15、TAL1、TCF12、TCF3、TFG、TYK2、USP6、YWHAE、AR、BRCA1、BRCA2、CDKN2A、ERBB4、FLT3、KRAS、MDM4、MYBL1、NF1、NOTCH4、NUTM1、PRKACA、PRKACB、PTEN、RAD51B和RBI。

[0052] 在本公开的第二方面是一种具有式(I)的化合物,

[0053]  $[Olig1] - ([R^1]_o - [R^2]_p)_q - [L^1]_t - [Z] - [L^2]_u - [W]_v - [Olig2]$  (I),

[0054] 其中

[0055] o为0或1;

[0056] p为0或1;

[0057] q为0或1;

[0058] t为0、1或2;

[0059] u为0、1或2;

[0060] v为0或1;

[0061]  $R^1$ 为具有约1个与约24个之间核苷酸的寡核苷酸;

[0062]  $R^2$ 为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地被一个或多个选自O、N或S的杂原子取代;

[0063]  $L^1$ 和 $L^2$ 独立地为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱

和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子,并且任选地包含一个或多个羰基基团;

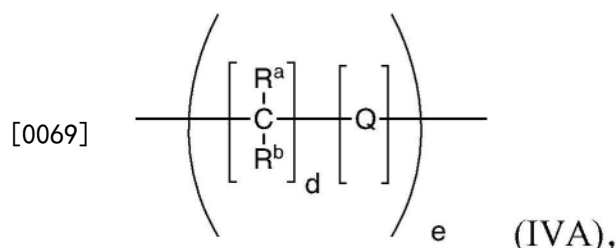
[0064] Z为选自以下的部分:三唑、二氢吡嗪、磷酸酯键合、酰胺键合、硫醚键合、异噁唑啉、脞、脲醚和氯-s-三嗪键合;

[0065] W为具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和、脂肪族或芳香族基团,任选地被一个或多个选自O、N、S的杂原子取代,前提是W包含至少一个可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团;

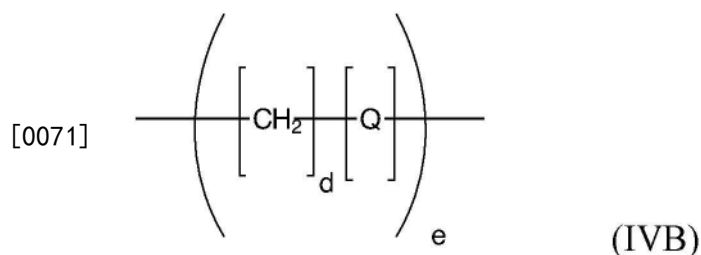
[0066] Olig1为具有约1个与约30个之间核苷酸的寡核苷酸,并且其中Olig1具有不可延伸的3'端;并且

[0067] Olig2为具有约1个与约30个之间核苷酸的寡核苷酸,并且其中Olig2具有可延伸的3'端。

[0068] 在一些实施例中, $R^2$ 包含具有2个与约32个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地被一个或多个选自O、N或S的杂原子取代,并且任选地包含一个或多个羰基基团。在一些实施例中, $R^2$ 包含具有式(IVA)结构的部分:



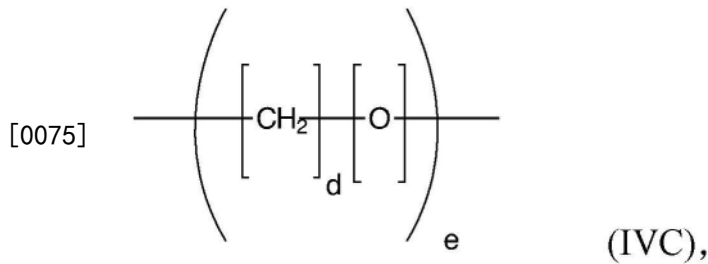
[0070] 其中d和e为各自独立地在1至32的范围内的整数;Q为键、O、S、N( $R^c$ ) ( $R^d$ )或季胺( $N^+H$  ( $R^c$ ) ( $R^d$ )); $R^a$ 和 $R^b$ 独立地为H、 $C_1$ - $C_4$ 烷基基团、F、Cl或N( $R^c$ ) ( $R^d$ );并且 $R^c$ 和 $R^d$ 独立地为 $CH_3$ 或H。在一些实施例中,d为2或3;并且其中e为在1和12之间的范围内的整数。在一些实施例中, $R^2$ 包含具有式(IVB)结构的部分:



[0072] 其中d和e为各自独立地在1至32的范围内的整数;Q为键、O、S或N( $R_c$ ) ( $R_d$ );并且 $R_c$ 和 $R_d$ 独立地为 $CH_3$ 或H。在一些实施例中,d为2或3;并且其中e为在1和12之间的范围内的整数。在一些实施例中,d为2或3;并且其中e为在1和8之间的范围内的整数。

[0073] 在一些实施例中, $L^1$ 或 $L^2$ 中的至少一个包含具有1个与约4个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地被一个或多个选自O、N或S的杂原子取代,并且任选地包含一个或多个羰基基团。

[0074] 在一些实施例中, $o+p=1$ ,并且 $q=1$ 。在一些实施例中, $R^1$ 包含约1个与约16个之间核苷酸。在一些实施例中, $R^1$ 包含约2个与约9个之间核苷酸。在一些实施例中, $R^2$ 包含具有式(IVC)结构的部分:



[0076] 其中d和e为各自独立地在1至32的范围内的整数。在一些实施例中,d为2或3;并且其中e为在1和12之间的范围内的整数。在一些实施例中,d为2或3;并且其中e为在1和8之间的范围内的整数。在一些实施例中,d为2;并且其中e为在1和12之间的范围内的整数。在一些实施例中,d为2;并且其中e为在1和8之间的范围内的整数。在一些实施例中,d为2;并且其中e为在2和6之间的范围内的整数。

[0077] 在一些实施例中,o为0并且p和q二者为1,R<sup>1</sup>包含至少一个PEG基团,并且L<sup>1</sup>包含至少一个羰基部分。在一些实施例中,o为0并且p和q二者为1,R<sup>1</sup>包含至少两个PEG基团,并且L<sup>1</sup>包含至少一个羰基部分。在一些实施例中,o为0并且p和q二者为1,R<sup>1</sup>包含至少三个PEG基团,并且L<sup>1</sup>包含至少一个羰基部分。在一些实施例中,o为0并且p和q二者为1,R<sup>1</sup>包含至少四个PEG基团,并且L<sup>1</sup>包含至少一个羰基部分。在一些实施例中,o为0并且p和q二者为1,R<sup>1</sup>包含至少六个PEG基团,并且L<sup>1</sup>包含至少一个羰基部分。在一些实施例中,o为0并且p和q二者为1,R<sup>1</sup>包含至少八个PEG基团,并且L<sup>1</sup>包含至少一个羰基部分。在一些实施例中,o为0并且p和q二者为1,R<sup>1</sup>包含至少十二个PEG基团,并且L<sup>1</sup>包含至少一个羰基部分。

[0078] 在一些实施例中,0lig2包含条形码。在一些实施例中,条形码是唯一分子条形码(UID)、样品条形码和识别标签中的一种或多种。在一些实施例中,0lig2包含通用引物结合位点。在一些实施例中,v为0并且0lig2包含切割位点,所述切割位点包括含尿嘧啶的核苷酸。在一些实施例中,0lig2包含随机核苷酸序列。

[0079] 在一些实施例中,0lig1的至少一部分包含能够与选自由以下项组成的组的基因杂交的核苷酸序列:ALK、PPARG、BRAF、EGFR、FGFR1、FGFR2、FGFR3、MET、NRG1、NTRK1、NTRK2、NTRK3、RET、ROS1、AXL、PDGFRA、PDGFB、ABL1、ABL2、AKT1、AKT2、AKT3、ARHGAP26、BRD3、BRD4、CRLF2、CSF1R、EPOR、ERBB2、ERBB4、ERG、ESR1、ESRRA、ETV1、ETV4、ETV5、ETV6、EWSR1、FGR、IL2RB、INSR、JAK1、JAK2、JAK3、KIT、MAML2、MAST1、MAST2、MSMB、MUSK、MYB、MYC、NOTCH1、NOTCH2、NUMBL、NUT、PDGFRB、PIK3CA、PKN1、PRKCA、PRKCB、PTK2B、RAF1、RARA、RELA、RSPO2、RSPO3、SYK、TERT、TFE3、TFEB、THADA、TMPRSS2、TSLP、TY、BCL2、BCL6、BCR、CAMTA1、CBFB、CCNB3、CCND1、CIC、CRFL2、DUSP22、EPC1、FOXO1、FUS、GLI1、GLIS2、HMGA2、JAZF1、KMT2A、MALT1、MEAF6、MECOM、MKL1、MKL2、MTB、NCOA2、NUP214、NUP98、PAX5、PDGFB、PICALM、PLAG1、RBM15、RUNX1、RUNX1T1、SS18、STAT6、TAF15、TAL1、TCF12、TCF3、TFG、TYK2、USP6、YWHAE、AR、BRCA1、BRCA2、CDKN2A、ERBB4、FLT3、KRAS、MDM4、MYBL1、NF1、NOTCH4、NUTM1、PRKACA、PRKACB、PTEN、RAD51B和RBI1。

[0080] 在一些实施例中,0lig1是不可延伸的。在一些实施例中,0lig2是可延伸的。在一些实施例中,0lig1包含1个与约10个之间核苷酸。在一些实施例中,0lig2包含1个与约10个之间核苷酸。

[0081] 在一些实施例中,基团-([R<sup>1</sup>]<sub>o</sub>-[R<sup>2</sup>]<sub>p</sub>)<sub>q</sub>-的大小在约15埃至约400埃之间的范围内。

在一些实施例中,基团- $([R^1]_o-[R^2]_p)_q$ -的大小在约15埃至约200埃之间的范围内。在一些实施例中,基团- $([R^1]_o-[R^2]_p)_q$ -的大小在约15埃至约100埃之间的范围内。在一些实施例中,基团- $([R^1]_o-[R^2]_p)_q$ -的大小在约15埃至约50埃之间的范围内。在一些实施例中,基团- $([R^1]_o-[R^2]_p)_q$ -的大小在约20埃至约45埃之间的范围内。在一些实施例中,基团- $([R^1]_o-[R^2]_p)_q$ -的大小在约20埃至约40埃之间的范围内。

[0082] 在本公开的第三方面是一种用于检测基因融合,诸如用于检测已知融合配偶体和未知融合配偶体之间的基因融合的试剂盒,其中所述试剂盒包含(a) DNA聚合酶;(b) 具有式(I)的化合物,

[0083]  $[Olig1]-([R^1]_o-[R^2]_p)_q-[L^1]_t-[Z]-[L^2]_u-[W]_v-[Olig2]$  (I),

[0084] 其中

[0085] o为0或1;

[0086] p为0或1;

[0087] q为0或1;

[0088] t为0、1或2;

[0089] u为0、1或2;

[0090] v为0或1;

[0091]  $R^1$ 为具有约1个与约24个之间核苷酸的寡核苷酸;

[0092]  $R^2$ 为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地被一个或多个选自O、N或S的杂原子取代;

[0093]  $L^1$ 和 $L^2$ 独立地为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子,并且任选地包含一个或多个羰基基团;

[0094] Z为选自以下的部分:三唑、二氢吡嗪、磷酸酯键合、酰胺键合、硫醚键合、异噻唑啉、脞、脲醚和氯-s-三嗪键合;

[0095] W为具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和、脂肪族或芳香族基团,任选地被一个或多个选自O、N、S的杂原子取代,前提是W包含至少一个可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团;

[0096] Olig1为具有约1个与约30个之间核苷酸的寡核苷酸,并且其中Olig1具有不可延伸的3'端;并且

[0097] Olig2为具有约1个与约30个之间核苷酸的寡核苷酸,并且其中Olig2具有可延伸的3'端。

[0098] 在一些实施例中,试剂盒进一步包含正向扩增引物和反向扩增引物。在一些实施例中,Olig2包含至少一个含尿嘧啶的核苷酸,并且其中试剂盒进一步包含尿嘧啶-N-DNA糖基化酶(UNG)。在一些实施例中,DNA聚合酶是逆转录酶并且试剂盒进一步包含热稳定的DNA依赖性DNA聚合酶。

[0099] 在一些实施例中,Olig1的至少一部分包含能够与选自由以下项组成的组的基因杂交的核苷酸序列:ALK、PPARG、BRAF、EGFR、FGFR1、FGFR2、FGFR3、MET、NRG1、NTRK1、NTRK2、NTRK3、RET、ROS1、AXL、PDGFRA、PDGFB、ABL1、ABL2、AKT1、AKT2、AKT3、ARHGAP26、BRD3、BRD4、CRLF2、CSF1R、EPOR、ERBB2、ERBB4、ERG、ESR1、ESRRA、ETV1、ETV4、ETV5、ETV6、EWSR1、FGR、

IL2RB、INSR、JAK1、JAK2、JAK3、KIT、MAMIL2、MAST1、MAST2、MSMB、MUSK、MYB、MYC、NOTCH1、NOTCH2、NUMBL、NUT、PDGFRB、PIK3CA、PKN1、PRKCA、PRKCB、PTK2B、RAF1、RARA、RELA、RSPO2、RSPO3、SYK、TERT、TFE3、TFEB、THADA、TMPRSS2、TSLP、TY、BCL2、BCL6、BCR、CAMTA1、CBFB、CCNB3、CCND1、CIC、CRFL2、DUSP22、EPCI、FOXO1、FUS、GLI1、GLIS2、HMGA2、JAZF1、KMT2A、MALT1、MEAF6、MECOM、MKL1、MKL2、MTB、NCOA2、NUP214、NUP98、PAX5、PDGFB、PICALM、PLAG1、RBM15、RUNX1、RUNX1T1、SS18、STAT6、TAF15、TAL1、TCF12、TCF3、TFG、TYK2、USP6、YWHAE、AR、BRCA1、BRCA2、CDKN2A、ERB84、FLT3、KRAS、MDM4、MYBL1、NF1、NOTCH4、NUTM1、PRKACA、PRKACB、PTEN、RAD51B和RB1。

[0100] 在本公开的第四方面是包含具有式 (I) 的化合物的反应容器，

[0101]  $[Olig1] - ([R^1]_o - [R^2]_p)_q - [L^1]_t - [Z] - [L^2]_u - [W]_v - [Olig2]$  (I)，

[0102] 其中

[0103] o为0或1；

[0104] p为0或1；

[0105] q为0或1；

[0106] t为0、1或2；

[0107] u为0、1或2；

[0108] v为0或1；

[0109]  $R^1$ 为具有约1个与约24个之间核苷酸的寡核苷酸；

[0110]  $R^2$ 为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代；

[0111]  $L^1$ 和 $L^2$ 独立地为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地包含一个或多个选自O、N或S的杂原子，并且任选地包含一个或多个羰基基团；

[0112] Z为选自以下的部分：三唑、二氢吡嗪、磷酸酯键合、酰胺键合、硫醚键合、异噁唑啉、脞、脲和氯-s-三嗪键合；

[0113] W为具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和、脂肪族或芳香族基团，任选地被一个或多个选自O、N、S的杂原子取代，前提是W包含至少一个可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团；

[0114] Olig1为具有约1个与约30个之间核苷酸的寡核苷酸，并且其中Olig1具有不可延伸的3'端；并且

[0115] Olig2为具有约1个与约30个之间核苷酸的寡核苷酸，并且其中Olig2具有可延伸的3'端。

[0116] 在一些实施例中，反应容器包含至少一种聚合酶。在一些实施例中，至少一种聚合酶是DNA聚合酶。在一些实施例中，反应容器进一步包含至少一种缓冲液。在一些实施例中，反应容器进一步包含至少一种辅因子。在一些实施例中，反应容器进一步包含dNTP。

[0117] 在本公开的第五方面是：(a) 具有式 (II) 的化合物：

[0118]  $[Olig1] - ([R^1]_o - [R^2]_p)_q - [L^1]_t - [X]$  (II)，

[0119] 其中

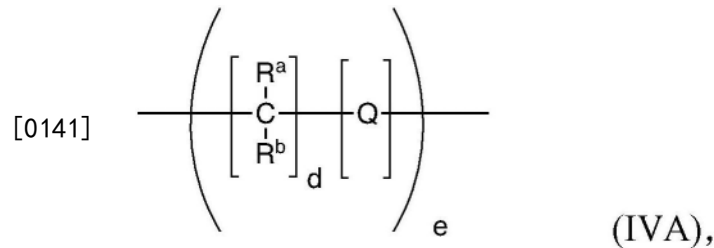
[0120] o为0或1；

- [0121] p为0或1；
- [0122] q为1或2；
- [0123] t为0、1或2；
- [0124] R<sup>1</sup>为具有1个与约24个之间核苷酸的寡核苷酸；
- [0125] R<sup>2</sup>为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代；
- [0126] L<sup>1</sup>为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团；
- [0127] X为二苯并环辛炔、反式环辛烯、炔烃、烯炔、叠氮化物、四嗪、马来酰亚胺、N-羟基琥珀酰亚胺、硫醇、1,3-硝酮、醛、酮、肼、羟胺、氨基基团或亚磷酰胺；并且
- [0128] 0lig1为具有约1个与约30个之间核苷酸的寡核苷酸；以及
- [0129] (b) 具有式(III)的化合物：
- [0130] [Y]-[L<sup>2</sup>]<sub>u</sub>-[W]<sub>v</sub>-[0lig2] (III)，
- [0131] 其中
- [0132] u为0、1或2；
- [0133] v为0或1；
- [0134] Y为二苯并环辛炔、反式环辛烯、炔烃、烯炔、叠氮化物、四嗪、马来酰亚胺、N-羟基琥珀酰亚胺、硫醇、1,3-硝酮、醛、酮、肼、羟胺、氨基基团或亚磷酰胺；
- [0135] L<sup>2</sup>为具有1个与16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地包含一个或多个选自O、N或S的杂原子，并且任选地包含一个或多个羰基基团；
- [0136] W为具有1个与12个之间碳原子的经取代或未取代的、饱和或不饱和脂肪族或芳香族基团，任选地被一个或多个选自O、N、S的杂原子取代，前提是W包含至少一个可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团；并且
- [0137] 0lig2为具有约1个与约30个之间核苷酸的寡核苷酸。
- [0138] 在一些实施例中，0lig1包含不可延伸的3'端；并且其中0lig2包含可延伸的3'端。在一些实施例中，0lig1包含1个与约10个之间核苷酸。在一些实施例中，0lig2包含1个与约10个之间核苷酸。在一些实施例中，0lig1的至少一部分能够与选自由以下项组成的组的基因杂交：ALK、PPARG、BRAF、EGFR、FGFR1、FGFR2、FGFR3、MET、NRG1、NTRK1、NTRK2、NTRK3、RET、ROS1、AXL、PDGFRA、PDGFB、ABL1、ABL2、AKT1、AKT2、AKT3、ARHGAP26、BRD3、BRD4、CRLF2、CSF1R、EPOR、ERBB2、ERBB4、ERG、ESR1、ESRRA、ETV1、ETV4、ETV5、ETV6、EWSR1、FGR、IL2RB、INSR、JAK1、JAK2、JAK3、KIT、MAML2、MAST1、MAST2、MSMB、MUSK、MYB、MYC、NOTCH1、NOTCH2、NUMBL、NUT、PDGFRB、PIK3CA、PKN1、PRKCA、PRKCB、PTK2B、RAF1、RARA、RELA、RSPO2、RSPO3、SYK、TERT、TFE3、TFEB、THADA、TMPRSS2、TSLP、TY、BCL2、BCL6、BCR、CAMTA1、CBFB、CCNB3、CCND1、CIC、CRFL2、DUSP22、EPCI、FOXO1、FUS、GLI1、GLIS2、HMGA2、JAZF1、KMT2A、MALT1、MEAF6、MECOM、MKL1、MKL2、MTB、NCOA2、NUP214、NUP98、PAX5、PDGFB、PICALM、PLAG1、RBM15、RUNX1、RUNX1T1、SS18、STAT6、TAF15、TAL1、TCF12、TCF3、TFG、TYK2、USP6、YWHAE、AR、BRCA1、BRCA2、CDKN2A、ERBB4、FLT3、KRAS、MDM4、MYBL1、NF1、NOTCH4、NUTM1、PRKACA、PRKACB、PTEN、

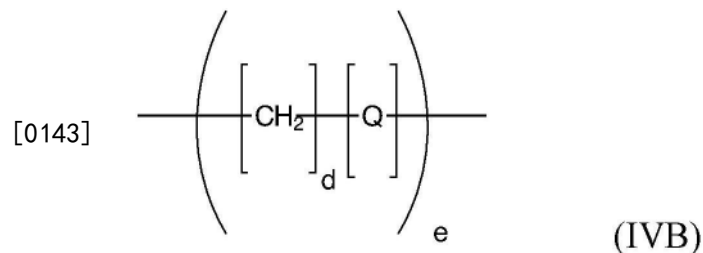
RAD51B和RB1。

[0139] 在一些实施例中，X或Y中的一个包含炔烃部分；并且X或Y中的另一个包含叠氮化物部分。在一些实施例中，炔烃部分为DBC0。在一些实施例中，X或Y中的一个包含马来酰亚胺部分；并且X或Y中的另一个包含硫醇部分。在一些实施例中，X或Y中的一个包含烯烃部分；并且X或Y中的另一个包含四嗪部分。在一些实施例中，包含氨基部分，并且其中试剂盒进一步包含s-三氯三嗪。

[0140] 在一些实施例中，R<sup>2</sup>包含具有2个与约32个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在一些实施例中，R<sup>2</sup>包含具有式(IVA)的部分：

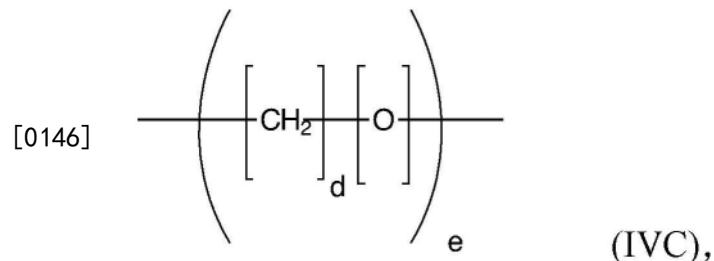


[0142] 其中d和e为各自独立地在1至32的范围内的整数；Q为键、O、S、N(R<sup>c</sup>) (R<sup>d</sup>) 或季胺(N<sup>+</sup>H(R<sup>c</sup>)) (R<sup>d</sup>)；R<sup>a</sup>和R<sup>b</sup>独立地为H、C<sub>1</sub>-C<sub>4</sub>烷基基团、F、Cl或N(R<sup>c</sup>) (R<sup>d</sup>)；并且R<sup>c</sup>和R<sup>d</sup>独立地为CH<sub>3</sub>或H。在一些实施例中，d为2；并且e为在1至约12的范围内的整数。在一些实施例中，R<sup>2</sup>包含具有式(IVB)的部分：



[0144] 其中d和e为各自独立地在1至32的范围内的整数；Q为键、O、S或N(R<sub>c</sub>) (R<sub>d</sub>)；并且R<sub>c</sub>和R<sub>d</sub>独立地为CH<sub>3</sub>或H。在一些实施例中，d为2；并且e为在1至约12的范围内的整数。在一些实施例中，d为2；并且e为在1至约6的范围内的整数。

[0145] 在一些实施例中，o+p=1，并且q=1。在一些实施例中，R<sup>1</sup>包含约2个与约9个之间核苷酸。在一些实施例中，R<sup>2</sup>包含具有式(IVC)的部分：



[0147] 其中d和e为各自独立地在1至32的范围内的整数。在一些实施例中，d为2；并且e为在1至约12的范围内的整数。在一些实施例中，d为2；并且e为在1至约6的范围内的整数。在一些实施例中，o为0且p和q二者为1，并且L包含至少一个PEG基团。

[0148] 在一些实施例中，01ig2包含条形码。在一些实施例中，条形码是唯一分子条形码

(UID)、样品条形码和识别标签中的一种或多种。在一些实施例中,0lig2包含通用引物结合位点。在一些实施例中,v为0并且0lig2包含切割位点,所述切割位点包括含尿嘧啶的核苷酸。

[0149] 在一些实施例中,试剂盒进一步包含聚合酶。在一些实施例中,聚合酶是DNA聚合酶。在一些实施例中,所述试剂盒进一步包含含有至少一种基因融合物的核酸样品。在一些实施例中,试剂盒进一步包含稀释剂。

## 附图说明

[0150] 图1是说明式(I)化合物的退火和延伸步骤的图。

[0151] 图2是说明链置换和链切割释放包含基因融合序列的拷贝链的步骤的图。

## 具体实施方式

### [0152] 概述

[0153] 本公开涉及有助于检测包含一种或多种靶标核酸的样品中的结构基因组重排的组合物和试剂盒。本公开还涉及利用基于扩增子的方法检测结构基因组重排,更特别地基因融合的方法。在一些实施例中,本文所述的方法利用一种或多种式(I)化合物来扩增其中一种融合配偶体未知的基因融合。在一些实施例中,用一种或多种式(I)化合物进行扩增有利于在有或没有测序步骤的情况下检测基因融合。在使用测序步骤的那些实施例中,这种测序需要最小的测序深度。

### [0154] 定义

[0155] 除非另有定义,否则本文所用的科学技术术语具有如本领域的普通技术人员通常理解的相同意义。参见,Sambrook等人.,Molecular Cloning,A Laboratory Manual,第4版,冷泉港实验室出版社(2012)。

[0156] 还应该理解的是,除非指明是相反情况,否则在本文所要求保护的包括一个以上步骤或动作的任何方法中,所述方法的所述步骤或动作的顺序不必限于表述所述方法的所述步骤或动作的所述顺序。

[0157] 如本文所用,除非上下文另有明确指示,否则单数形式“一(a/an)”和“该/所述”包括复数个指代物。同样,除非上下文另有明确指示,否则词语“或”旨在包括“和”。术语“包括”定义为包容性,如“包括A或B”是指包括A、B或A和B。

[0158] 如本文在说明书和权利要求书中所用,“或”应理解为与上文定义的“和/或”具有相同的含义。例如,在分隔列表中的项目时,“或”或“和/或”应解释为具有包容性,即包含若干元素或元素列表中的至少一个元素,但也包含一个以上元素,以及任选地包含额外的未列出的项目。只有指明与之相反的术语,如“只有一个”或“恰好一个”,或者在权利要求中使用的“由...组成”,将指包含若干元素或元素列表中的恰好一个元素。一般来说,本文使用的术语“或者”只有在前面有“或”、“其中之一”、“只有一个”或“恰好一个”等排他性术语时,才应解释为表示排他性的替代选择(即“一个或另一个,但不是两个”)。在权利要求书中使用的“基本上由...组成”应具有在专利法领域使用的普通含义。

[0159] 如本文所用,“包括”、“包含”、“具有”等术语可互换使用,且含义相同。类似地,“包括”、“包含”、“具有”等可互换使用并且具有相同的含义。具体而言,每个术语的定义都与普



通美国专利法对“包括”的定义一致,因此每个术语都可理解为一个开放性术语,其含义为“至少以下”,并且也可释为不排除额外的特征、限制、方面等。因此,例如“具有组件a、b和c的装置”是指所述装置至少包括组件a、b和c。同样,短语:“涉及步骤a、b和c的方法”是指所述方法至少包括步骤a、b和c。此外,尽管本文可以特定的顺序概述步骤和过程,但是本领域技术人员将认识到,所述顺序步骤和过程可能会有所不同。

[0160] 如本文在说明书和权利要求书中所用,就一个或多个元素的列表而言,短语“至少一个”应理解为选自元素列表中任何一个或多个元素的至少一个元素,但不一定包括元素列表中具体列出的每个元素中的至少一个,也不排除元素列表中的任何元素组合。除了在短语“至少一个”所涉及的元素列表中具体确定的元素之外,该定义还允许其他元素任选地存在,无论这些元素与具体确定的元素相关与否。因此,作为一个非限制性实例,“A和B中的至少一个”(或者等效地,“A或B中的至少一个”,或者等效地,“A和/或B中的至少一个”)在一个实施例中可以指至少一个任选地包括一个以上的A,但没有B(以及选择性地包括B以外的元素);在另一个实施例中,指至少一个选择性地包括一个以上的B,但没有A(以及选择性地包括A以外的元素);在又一个实施例中,指至少一个选择性地包括一个以上的A,以及至少一个选择性地包括一个以上的B(以及选择性地包括其他元素)等。

[0161] 如本文所用,术语“衔接子”是指核苷酸序列,可将其加入另一序列中以便赋予该另一序列以另外的元件和性质。额外的元件包括但不限于:条形码、引物结合位点、捕获部分、标签、二级结构。

[0162] 如本文所用,术语“脂肪族的”是指直链或支链的烃链,其可以是饱和的或单不饱和或多不饱和的。不饱和脂肪族基团包含一个或多个双键和/或三键。烃链的分支可以包括直链以及非芳族环状元素。除非另有说明,否则烃链可以具有任何长度,并且可以包含任何数目的支链。主链和支链都可以进一步包含杂原子,例如B、N、O、P、S、Se或Si。

[0163] 如本文所用,术语“条形码”是指可被检测和鉴定的核酸序列。条形码通常可以为2个以上且最长可达约50个核苷酸。条形码被设计成与群体中的其他条形码具有至少最小数量的差异。条形码对于样品中的每个分子可以为唯一的,或对样品是唯一的,并且由样品中的多个分子共享。术语“多重标识符”、“MID”或“样品条形码”是指识别样品或样品来源的条形码。就此而言,来自单一来源或样品的所有或基本上所有的MID条形码化的多核苷酸将共享相同序列的MID;而来自不同来源或样品的所有或基本上所有(例如,至少90%或99%)的MID条形码化的多核苷酸将具有不同的MID条形码序列。可以将来自具有不同MID的不同来源的多核苷酸进行混合并进行并行测序,同时保持MID条形码中编码的样品信息。术语“唯一分子标识符”或“UID”是指识别与其附接的多核苷酸的条形码。通常,UID条形码化的多核苷酸混合物中的所有或基本上所有(例如,至少90%或99%)的UID条形码是唯一的。条形码也可以用作工作流程的一部分的“识别标签”。例如,源自RNA的DNA分子(例如,cDNA)可以凭借仅在cDNA合成期间连接到cDNA的标签而与源自基因组DNA的具有相同序列的DNA分子区分开来。此类条形码可称为“RNA识别标签”或简称为“识别标签”。

[0164] 如本文所用,术语“ctDNA”是指从原发性肿瘤细胞、血液循环系统中的循环肿瘤细胞和坏死或凋亡的肿瘤细胞释放到外周血的游离DNA,或其任何组合。

[0165] 如本文所用,术语“DNA聚合酶”是指从脱氧核苷酸执行模板导向合成多核苷酸的酶。DNA聚合酶包括原核Pol I、Pol II、Pol III、Pol IV和Pol V,真核DNA聚合酶,古细菌

DNA聚合酶、端粒酶和反转录酶。术语“热稳定的聚合酶”是指在通过聚合酶链式反应(PCR)进行的核酸指数扩增中有用的酶,因为该酶是耐热的。热稳定的酶当在高温下经过实现双链核酸变性所需要的时间后,保留足够的活性以实现随后的多核苷酸延伸反应,并且不会不可逆变性(失活)。

[0166] 在一些实施例中,来自物种热球菌属(*Thermococcus*)、火球菌属(*Pyrococcus*)、硫化叶菌产甲烷球菌(*Sulfolobus Methanococcus*)和其他古细菌B聚合酶的热稳定的聚合酶。在一些情况下,核酸(例如DNA或RNA)聚合酶可以是经修饰的天然存在的A型聚合酶。本公开的进一步的实施例通常涉及一种方法,其中例如在引物延伸、端部修饰(例如末端转移酶、降解或补平)或扩增反应中,经修饰的A型聚合酶可以选自以下任何属的任何种:亚栖热菌属(*Meiothermus*)、热袍菌门(*Thermotoga*)或嗜热菌属(*Thermomicrobium*)。本公开的另一实施例通常从属于一种方法,其中例如在引物延伸、端部修饰(例如末端转移酶、降解或补平)或扩增反应中,聚合酶可以从以下任一种分离栖热水生菌(*Thermus aquaticus*) (Taq)、嗜热栖热菌(*Thermus thermophilus*)、嗜钙质热菌(*Thermus caldophilus*)或丝状栖热菌(*Thermus filiformis*)。本公开的另外的实施例通常涵盖一种方法,其中例如在引物延伸、末端修饰(例如末端转移酶、降解或补平)或扩增反应中,修饰的A型聚合酶可以从以下分离:嗜热脂肪芽孢杆菌、嗜热球形杆菌(*Sphaerobacter thermophilus*)、嗜热链球菌(*Dictyoglomus thermophilum*)或大肠杆菌(*Escherichia coli*)。在另一实施例中,本公开通常涉及一种方法,其中例如在引物延伸、端部修饰(例如末端转移酶、降解或补平)或扩增反应中,修饰的A型聚合酶可以为突变Taq-E507K聚合酶。本公开的另一实施例通常从属于一种方法,其中热稳定的聚合酶可用于引起靶标核酸的扩增。

[0167] 如本文所用,术语“富集”是指增加多个分子中靶分子的相对量。富集可以增加靶分子的相对量,直至完全或几乎完全排除非靶分子。靶核酸富集的示例包括线性杂交捕获、扩增、指数扩增(PCR)和引物延伸靶富集(PETE),参见例如美国申请序列号14/910,237、15/228,806、15/648,146和国际申请序列号PCT/EP2018/085727。

[0168] 如本文所用,术语“基因融合”是指与参考基因组相比基因组序列的变化,其包含其中一个基因的一部分与另一序列融合的易位。一些基因融合导致功能性融合mRNA。这些基因融合的子集进一步产生功能性融合蛋白。基因融合具有参考编码融合蛋白的mRNA而指定的5'-配偶体和3'-配偶体。5'-融合配偶体编码蛋白质的N-末端部分,而3'-融合配偶体编码蛋白质的C-末端部分。

[0169] 如本文所用,术语“杂原子”是指包括硼(B)、氧(O)、氮(N)、硫(S)、磷(P)和硅(Si)。在一些实施例中,“杂环”可包含一个或多个杂原子。在其他实施例中,脂肪族基团可以包含一个或多个杂原子或被一个或多个杂原子取代。

[0170] 如本文所用,术语“核酸”或“多核苷酸”是指单链形式或双链形式的脱氧核糖核酸(DNA)或核糖核酸(RNA)及其聚合物。除非特别限定,否则该术语涵盖包含天然核苷酸的已知类似物的核酸,该天然核苷酸具有与参考核酸相似的结合特性并且以与天然存在的核苷酸相似的方式进行代谢。除非另外指出,否则特定的核酸序列还隐含地涵盖其保守修饰的变体(例如,简并密码子替换)、等位基因、同源基因序、SNP和互补序列,以及明确指出的序列。

[0171] 如本文所用,术语“寡核苷酸”是指核苷酸或核苷单体单元的寡聚物,其中该寡聚

物任选地包括非核苷酸单体单元和/或在寡聚物的内部和/或外部位置连接的其他化学基团。寡聚物可以是天然的或合成的,并且可以包括天然存在的寡核苷酸,或包括具有非天然存在(或修饰的)碱基、糖部分、磷酸二酯-类似物键和/或替代性单体单元手性和异构结构的核苷的寡聚物(例如5'-键至2'-键,L-核苷, $\alpha$ -异头物核苷、 $\beta$ -异构体核苷、锁核酸(LNA)、肽核酸(PNA))。

[0172] 如本文所用,术语“引物”是指与单链模板核酸分子的特定区结合并经由聚合酶介导的酶促反应启动核酸合成的寡核苷酸。通常,引物包括少于约100个核苷酸,且优选包括少于约30个核苷酸。靶标特异性引物在杂交条件下与靶标多核苷酸特异性杂交。此类杂交条件可包括但不限于在等温扩增缓冲液中(20mM的Tris-HCl,10mM的 $(\text{NH}_4)_2\text{SO}_4$ 、50mM的KCl、2mM的 $\text{MgSO}_4$ 、0.1%的TWEEN®20、25°C下pH为8.8)在约40°C个与约70°C的温度下进行杂交。除了靶标结合区域外,引物可以具有附加区域,通常位于5'-部分。附加区域可以包括通用引物结合位点或条形码。为了发生指数扩增,引物必须是向内的,即与靶标核酸的相反链杂交,且3'端彼此面对。这种扩增引物的取向有时称为“正确取向”。进一步地,为了发生指数扩增,引物在相距于彼此合适的距离内与靶核酸杂交。在标准PCR条件下,与相距超过2000个碱基对的相反链杂交的引物不会产生足够量的产物。对于cfDNA样品,典型的片段大小相距175个碱基对,因此与相距超过175个碱基对的相反链杂交的引物通常不会产生扩增产物。

[0173] 如本文所用,术语“参考基因组”和“参考基因组序列”是指向公众发布并由国家生物技术信息中心(NCBI)定期更新的整个人类基因组序列(“基因组版本”),目前版本GRCh38。参考基因组可通过染色体位置和序列进行搜索,以能够比较来自单个样品的序列并鉴定样品中的任何序列变化。

[0174] 如本文所用,术语“重排的基因组”是指当与参考基因组相比时包含一个或多个重排的基因组。可以理解,重排的基因组还在不参与重排的其他基因座处包含非重排序列。重排的基因组中的此类基因座具有与相对应的参考基因组基因座相同的序列。术语“重排的基因组序列”是指重排的基因组中的重排序列。

[0175] 如本文所用,术语“读取深度”或“测序深度”指序列已被测序的次数(测序的深度)。例如,可以通过比对多个测序运行结果并在一定大小(例如100bp)的非重叠窗口中计数读取的起始位置来确定读取深度。可以使用本领域已知的方法基于读取深度来确定拷贝数变化。例如,使用Yoon等人(Genome Research 2009September;19(9):1586-1592)、Xie等人(BMC Bioinformatics 2009Mar.6;10:80)或Medvedev等人(Nature Methods 2009November;6(11Suppl)):S13-20。

[0176] 如本文所用,术语“样品”是指包括核酸分子的任何生物样品,通常包括DNA或RNA。样品可以是组织、细胞或其提取物,或者可以是核酸分子的纯化样品。术语“样品”是指任何含有或假定含有靶核酸的组合物。使用术语“样品”并不一定意味着存在于样品中的核酸分子中存在靶标序列。该样品可以为包括从个体分离的组织或液体的样本,例如,皮肤、血浆、血清、脊髓液、淋巴液、滑液、尿液、泪液、血液细胞、器官和肿瘤,也指从取自个体的细胞建立的体外培养物的样品,包括福尔马林固定石蜡包埋组织(FFPET)和自其分离的核酸。样品也可包括不含细胞的材料,诸如含有无细胞DNA(cfDNA)或循环肿瘤DNA(ctDNA)的不含细胞的血液级分(fraction)。样品可以从非人类受试者或从环境中收集。

[0177] 在一些实施例中，“样品”为“代表性样品”。在一些实施例中，代表性样品是指一种样品(或样品的子集)，其准确反映整体组成并且因此该样品是对整个群体的无偏指示。通常，这意指代表性样品或其一部分中不同类型的细胞及其相对比例或百分比基本准确地反映或模拟了整个组织标本(通常是实体瘤或其部分)内这些细胞类型的相对比例或百分比。采样是获得对象的一部分以供后续分析的操作。代表性样品以可以获得对所研究对象的合理密切了解的方式产生。相比之下，传统的随机采样方法一般不会产生“代表性样品”。虽然从较大样品中选择较小的个体子样品可能会根据所选区域产生偏差，但对较大样品(例如整个肿瘤或淋巴结)进行均质化会导致空间分离的元素在整个样品中均匀分散。

[0178] 如本文所用，术语“测序”或“DNA测序”是指用于确定DNA寡核苷酸中核苷酸碱基、腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶的顺序的生化方法。如本文中所使用的术语，测序可以包括但不限于平行测序或本领域技术人员已知的任何其他测序方法，例如链终止法、快速DNA测序法、游走点分析(wandering-spot analysis)、Maxam-Gilbert测序、染料终止剂测序，或使用任何其他现代的自动化DNA测序仪器。

[0179] 如本文所用，术语“靶标”或“靶标核酸”是指样品中的目标核酸。样品可能包含多个靶标以及每个靶标的多个拷贝。

[0180] 如本文所用，术语“通用引物”是指可以与通用引物结合位点杂交的引物。通用引物结合位点可以是通常以非靶标特异性方式添加到靶标序列的天然或人工序列。

[0181] 主题组合物

[0182] 在本公开的一个方面是式(I)化合物(在本文中也称为“连接的引物”)：

[0183]  $[Olig1] - ([R^1]_o - [R^2]_p)_q - [L^1]_t - [Z] - [L^2]_u - [W]_v - [Olig2]$  (I)，

[0184] 其中

[0185] o为0或1；

[0186] p为0或1；

[0187] q为0或1；

[0188] t为0、1或2；

[0189] u为0、1或2；

[0190] v为0或1；

[0191]  $R^1$ 为具有约1个与约24个之间核苷酸的寡核苷酸；

[0192]  $R^2$ 为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代；

[0193]  $L^1$ 和 $L^2$ 独立地为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地包含一个或多个选自O、N或S的杂原子，并且任选地包含一个或多个羰基基团；

[0194] Z为选自以下的部分：三唑、二氢吡嗪、磷酸酯键合、酰胺键合、硫醚键合、异噻唑啉、脞、脲和氯-s-三嗪键合；

[0195] W为具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和、脂肪族或芳香族基团，任选地被一个或多个选自O、N、S的杂原子取代，前提是W包含至少一个可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团；

[0196] Olig1为包含约1个与约30个之间核苷酸的寡核苷酸；并且

[0197] Olig2为包含约1个与约30个之间核苷酸的寡核苷酸。

[0198] 当一个基团被描述为“经取代或未取代的”，如果经取代时，取代基可以选自一个或多个所示取代基。如果没有指示取代基，则是指所指示的“取代的”基团可以被一个或多个基团单独地并且独立地选自由以下项组成的组的基团取代：烷基、烯基、炔基、环烷基、环烯基、环炔基、芳基、杂芳基、杂脂环基、芳烷基、杂芳烷基、(杂脂环基)烷基、羟基、保护的羟基、烷氧基、芳氧基、酰基、巯基、烷硫基、芳硫基、氰基、氰酸酯、卤素、硫代羰基、0-氨基甲酰基、N-氨基甲酰基、0-硫代氨基甲酰基、N-硫代氨基甲酰基、C-酰胺基、N-酰胺基、S-磺酰胺基、N-磺酰胺基、C-羧基、保护的C-羧基、0-羧基、异氰硫基、氰硫基、异硫氰硫基、硝基、甲硅烷基、硫基、亚磺酰基、磺酰基、卤代烷基、卤代烷氧基、三卤代甲磺酰基、三卤代甲磺酰基、醚、氨基(例如单取代的氨基或二取代的氨基)及其保护的衍生物。上述任何基团可包括一个或多个杂原子，包括O、N或S。例如，当部分被烷基取代时，该烷基可包含选自O、N或S的杂原子(例如-(CH<sub>2</sub>-CH<sub>2</sub>-O-CH<sub>2</sub>-CH<sub>3</sub>))。

[0199] 在一些实施例中，Olig1包含约1个与约24个之间核苷酸。在另一些实施例中，Olig1包含约1个与约20个之间核苷酸。在另一些实施例中，Olig1包含约1个与约16个之间核苷酸。在又一些实施例中，Olig1包含约1个与约12个之间核苷酸。在又一些实施例中，Olig1包含约2个与约16个之间核苷酸。在又一些实施例中，Olig1包含约2个与约12个之间核苷酸。在又一些实施例中，Olig1包含约3个与约12个之间核苷酸。在又一些实施例中，Olig1包含约4个与约12个之间核苷酸。在又一些实施例中，Olig1包含约3个与约8个之间核苷酸。在又一些实施例中，Olig1包含约4个与约8个之间核苷酸。

[0200] 在一些实施例中，Olig1具有不可延伸的3'端。在一些实施例中，3'-端是不可延伸的，因为存在终止子化学结构，包括例如双脱氧核苷酸、2'-磷酸核苷酸，如美国专利号8,163,487中所述，或任何其他3'-O-封闭的可逆终止子和3'未封闭的可逆终止子，如例如美国专利申请公开号2014/0242579或J., 等人, Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides, P.N.A.S. 2008105 (27) 9145-9150中所述。

[0201] 在一些实施例中，Olig1包含能够与靶标序列杂交的锚定序列。换言之，Olig1的至少一部分能够与靶标核酸序列杂交。在一些实施例中，靶标核酸序列是已知融合配偶体。融合配偶体的非限制性实例包括ALK、PPARG、BRAF、EGFR、FGFR1、FGFR2、FGFR3、MET、NRG1、NTRK1、NTRK2、NTRK3、RET、ROS1、AXL、PDGFRA、PDGFB、ABL1、ABL2、AKT1、AKT2、AKT3、ARHGAP26、BRD3、BRD4、CRLF2、CSF1R、EPOR、ERBB2、ERBB4、ERG、ESR1、ESRRA、ETV1、ETV4、ETV5、ETV6、EWSR1、FGR、IL2RB、INSR、JAK1、JAK2、JAK3、KIT、MAML2、MAST1、MAST2、MSMB、MUSK、MYB、MYC、NOTCH1、NOTCH2、NUMBL、NUT、PDGFRB、PIK3CA、PKN1、PRKCA、PRKCB、PTK2B、RAF1、RARA、RELA、RSP02、RSP03、SYK、TERT、TFE3、TFEB、THADA、TMPRSS2、TSLP、TY、BCL2、BCL6、BCR、CAMTA1、CBFB、CCNB3、CCND1、CIC、CRFL2、DUSP22、EPCI、FOXO1、FUS、GLI1、GLIS2、HMGA2、JAZF1、KMT2A、MALT1、MEAF6、MECOM、MKL1、MKL2、MTB、NCOA2、NUP214、NUP98、PAX5、PDGFB、PICALM、PLAG1、RBM15、RUNX1、RUNX1T1、SS18、STAT6、TAF15、TAL1、TCF12、TCF3、TFG、TYK2、USP6、YWHAE、AR、BRCA1、BRCA2、CDKN2A、ERBB4、FLT3、KRAS、MDM4、MYBL1、NF1、NOTCH4、NUTM1、PRKACA、PRKACB、PTEN、RAD51B和RB1。

[0202] 在一些实施例中，Olig1的至少一部分与靶标序列完全互补。在另一些实施例中，

Olig1与靶标序列仅部分互补。在任何一种情况下，Olig1在用于引物退火的合适反应条件下与已知融合配偶体序列形成稳定的杂交体，所述条件为例如在包含20mM Tris-HCl、10mM  $(\text{NH}_4)_2\text{SO}_4$ 、50mM KCl、2mM  $\text{MgSO}_4$ 、0.1% TWEEN®20, pH 8.8的缓冲液中在25°C下，或在包含10mM Tris-HCl、50mM KCl、1.5mM  $\text{MgCl}_2$ , pH 8.3的缓冲液中在25°C下。

[0203] 在一些实施例中，Olig2包含约1个与约24个之间核苷酸。在另一些实施例中，Olig2包含约1个与约16个之间核苷酸。在又一些实施例中，Olig1包含约1个与约12个之间核苷酸。在又一些实施例中，Olig1包含约2个与约9个之间核苷酸。在一些实施例中，Olig2包含不可延伸的3'端。

[0204] 在一些实施例中，Olig2包含随机序列“(N)n”)。在一些实施例中，随机序列的长度可以为3个、4个、5个、6个、7个、8个或10个或更多个核苷酸。为了选择随机序列的适当长度，本领域技术人员将寻求具有能够在用于锚定序列杂交的条件下形成稳定杂交体的解链温度( $T_m$ )的序列。在另一些实施例中，Olig2包含单个重复核苷酸，例如聚T寡核苷酸。在一些实施例中，Olig2延伸穿过融合断点以形成包含上游融合配偶体的一部分、融合断点和下游融合配偶体的一部分的拷贝链。在一些实施例中，拷贝链用于进一步分析，例如通过扩增和(或)测序。

[0205] 在一些实施例中，Olig2的一部分不能与靶标序列杂交。在一些实施例中，Olig2的5'-部分可以包括诸如以下的元件：通用引物结合位点、平台特异性测序引物结合位点、条形码(样品条形码或分子条形码)或用户设计的其他标签序列。在一些实施例中，如本文进一步解释的，标签将RNA起始材料与DNA起始材料区分开来。

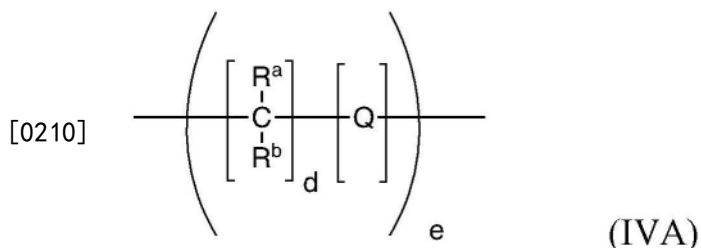
[0206] 如上所述，在一些实施例中， $R^1$ 可以为具有约1个与约16个之间核苷酸的寡核苷酸。在另一些实施例中， $R^1$ 包括具有约1个与约12个之间核苷酸的寡核苷酸。在又一些实施例中， $R^1$ 包括具有约1个与约8个之间核苷酸的寡核苷酸。在另一些实施例中， $R^1$ 的分子量为约350g/mol至约5200g/mol。在另一些实施例中， $R^1$ 的分子量为约650g/mol至约300g/mol。

[0207] 如上所述，在一些实施例中， $R^2$ 可以为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在另一些实施例中， $R^2$ 包含具有2个与约32个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在又一些实施例中， $R^2$ 包含具有2个与约28个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。

[0208] 在进一步的实施例中， $R^2$ 可以为具有2个与约24个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在更进一步的实施例中， $R^2$ 包含具有2个与约20个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在更进一步的实施例中， $R^2$ 包含具有2个与约6个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在更进一步的实施例中， $R^2$ 包含具有2个与约12个之间碳原子的经取代或未

取代的、饱和或不饱和、直链或环状脂肪族基团,任选地被一个或多个选自O、N或S的杂原子取代。在一些实施例中,一个或多个羰基基团可以为酮、酰胺或羧基。在另一些实施例中, $R^2$ 不包含羰基基团。

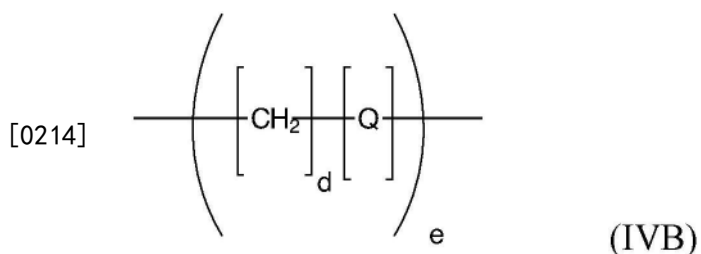
[0209] 在一些实施例中, $R^2$ 包含具有式 (IVA) 的部分:



[0211] 其中d和e为各自独立地在1至32的范围内的整数;Q为键、O、S、N( $R^c$ ) ( $R^d$ ) 或季胺( $N^+$  H( $R^c$ ) ( $R^d$ )); $R^a$ 和 $R^b$ 独立地为H、 $C_1$ - $C_4$ 烷基基团、F、Cl或N( $R^c$ ) ( $R^d$ );并且 $R^c$ 和 $R^d$ 独立地为 $CH_3$ 或H。

[0212] 在一些实施例中,d和e为各自独立地在2至18的范围内的整数。在一些实施例中,e在1至10的范围内的。在另一些实施例中,e在1至8的范围内的。在又一些实施例中,e在2至6的范围内的。在又一些实施例中,e在2至4的范围内的。在一些实施例中,d为在1至8的范围内的整数,并且e为在2至16的范围内的整数。在另一些实施例中,d为在2至8的范围内的整数,并且e为在2至12的范围内的整数。在另一些实施例中,d为2或3,并且e为在2至12的范围内的整数。在另一些实施例中,d为2或3,并且e为在2至8的范围内的整数。在另一些实施例中,d为2或3,并且e为在2至6的范围内的整数。在另一些实施例中,d为2或3,并且e为在2至4的范围内的整数。在一些实施例中, $R^a$ 或 $R^b$ 中的至少一个为 $-CH_3$ 。

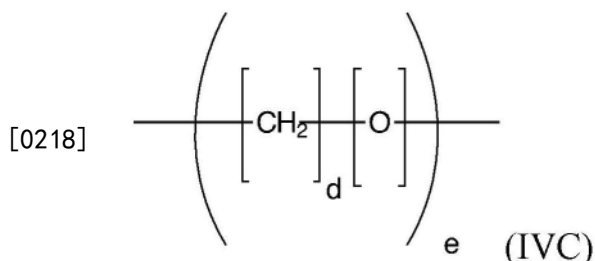
[0213] 在一些实施例中, $R^2$ 包含具有式 (IVB) 的部分:



[0215] 其中d和e为各自独立地在1至32的范围内的整数;Q为键、O、S或N( $R^c$ ) ( $R^d$ );并且 $R^c$ 和 $R^d$ 独立地为 $CH_3$ 或H。

[0216] 在一些实施例中,e在1至10的范围内的。在另一些实施例中,e在1至8的范围内的。在又一些实施例中,e在2至6的范围内的。在又一些实施例中,e在2至4的范围内的。在另一些实施例中,Q为O。在一些实施例中,d为在1至8的范围内的整数,并且e为在2至16的范围内的整数。在另一些实施例中,d为在2至8的范围内的整数,并且e为在2至12的范围内的整数。在另一些实施例中,d为2或3,并且e为在2至12的范围内的整数。在另一些实施例中,d为2或3,并且e为在2至8的范围内的整数。在另一些实施例中,d为2或3,并且e为在2至6的范围内的整数。在另一些实施例中,d为2或3,并且e为在2至4的范围内的整数。

[0217] 在一些实施例中, $R^2$ 包含具有式 (IVC) 的部分:



[0219] 其中d和e为各自独立地在1至32的范围内的整数。在一些实施例中，e在1至10的范围内。在另一些实施例中，e在1至8的范围内。在又一些实施例中，e在2至6的范围内。在又一些实施例中，e在2至4的范围内。在一些实施例中，d在1至4的范围内，并且e在1至约8的范围内。在一些实施例中，d为在1至8的范围内的整数，并且e为在2至约16的范围内的整数。在另一些实施例中，d为在2至8的范围内的整数，并且e为在2至约12的范围内的整数。在另一些实施例中，d为2或3，并且e为在2至12的范围内的整数。在另一些实施例中，d为2或3，并且e为在2至8的范围内的整数。在另一些实施例中，d为2或3，并且e为在2至6的范围内的整数。在另一些实施例中，d为2或3，并且e为在2至4的范围内的整数。

[0220] 在一些实施例中，R<sup>2</sup>包含增溶基团。在一些实施例中，增溶基团是聚乙二醇(PEG)基团或聚丙二醇基团。在又一些实施例中，接头包含约2个与约8个之间PEG基团或聚丙二醇基团。在又一些实施例中，接头包含约6个PEG基团或聚丙二醇基团。在又一些实施例中，接头包含约4个PEG基团聚丙二醇基团。在又一些实施例中，接头包含2个PEG基团或聚丙二醇基团。

[0221] 如上所述，在一些实施例中，L<sup>1</sup>可以为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在另一些实施例中，L<sup>1</sup>包含具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在又一些实施例中，L<sup>1</sup>包含具有1个与约8个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。

[0222] 在进一步的实施例中，L<sup>1</sup>包含具有1个与约6个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在更进一步的实施例中，L<sup>1</sup>包含具有1个与约4个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在一些实施例中，基团L<sup>1</sup>可以包含一个或多个增溶基团，例如PEG基团。在一些实施例中，羰基基团选自酮、酰胺和羧基。在一些实施例中，基团L<sup>1</sup>包含酮。在一些实施例中，基团L<sup>1</sup>包含酰胺。

[0223] 在一些实施例中，基团 - ([R<sup>1</sup>]<sub>o</sub> - [R<sup>2</sup>]<sub>p</sub>)<sub>q</sub> - 的长度在约15埃至约1000埃之间的范围内。在另一些实施例中，基团 - ([R<sup>1</sup>]<sub>o</sub> - [R<sup>2</sup>]<sub>p</sub>)<sub>q</sub> - 的长度在约15埃至约500埃之间的范围内。在又一些实施例中，基团 - ([R<sup>1</sup>]<sub>o</sub> - [R<sup>2</sup>]<sub>p</sub>)<sub>q</sub> - 的长度在约15埃至约400埃之间的范围内。在又一些实施例中，基团 - ([R<sup>1</sup>]<sub>o</sub> - [R<sup>2</sup>]<sub>p</sub>)<sub>q</sub> - 的长度在约15埃至约300埃之间的范围内。在又一些实施例中，基团 - ([R<sup>1</sup>]<sub>o</sub> - [R<sup>2</sup>]<sub>p</sub>)<sub>q</sub> - 的长度在约15埃至约250埃之间的范围内。在又一些实施例中，基



团- $([R^1]_o - [R^2]_p)_q$ -的长度在约15埃至约200埃之间的范围内。在又一些实施例中,基团- $([R^1]_o - [R^2]_p)_q$ -的长度在约15埃至约150埃之间的范围内。在又一些实施例中,基团- $([R^1]_o - [R^2]_p)_q$ -的长度在约15埃至约100埃之间的范围内。在又一些实施例中,基团- $([R^1]_o - [R^2]_p)_q$ -的长度在约15埃至约50埃之间的范围内。在又一些实施例中,基团- $([R^1]_o - [R^2]_p)_q$ -的长度在约20埃至约40埃之间的范围内。

[0224] 在一些实施例中, $o+p=1$ ,并且 $q$ 为1。在另一些实施例中, $o$ 为1, $p$ 为0,并且 $q$ 为1。在又一些实施例中, $o$ 为0, $p$ 为1,并且 $q$ 为1。在又一些实施例中, $o$ 为0, $p$ 为1,并且 $q$ 为2。

[0225] 在一些实施例中, $o$ 为1, $p$ 为0,并且 $q$ 为1,并且 $R^1$ 包含约1个与约12个之间核苷酸。在一些实施例中, $o$ 为1, $p$ 为0,并且 $q$ 为1,并且 $R^1$ 包含约1个与约8个之间核苷酸。在一些实施例中, $o$ 为1, $p$ 为0,并且 $q$ 为1,并且 $R^1$ 包含约1个与约6个之间核苷酸。

[0226] 在一些实施例中, $o$ 为0, $p$ 和 $q$ 二者为1,并且 $R^2$ 包含增溶基团。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含至少一个PEG基团。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含至少4个PEG基团。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含至少6个PEG基团。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含至少8个PEG基团。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含至少10个PEG基团。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含至少12个PEG基团。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含至少16个PEG基团。

[0227] 在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团,并且其中 $e$ 在1至16的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团,并且其中 $e$ 在1至12的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团,并且其中 $e$ 在1至8的范围内。

[0228] 在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2或3,并且其中 $e$ 在1至16的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2或3,并且其中 $e$ 在1至12的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2或3,并且其中 $e$ 在1至10的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2或3,并且其中 $e$ 在1至8的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2或3,并且其中 $e$ 在1至6的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2或3,并且其中 $e$ 在1至4的范围内。

[0229] 在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2,并且其中 $e$ 在1至12的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2,并且其中 $e$ 在1至10的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2,并且其中 $e$ 在1至8的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2,并且其中 $e$ 在1至6的范围内。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVB)的基团, $d$ 为2,并且其中 $e$ 在1至4的范围内。

[0230] 在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVC)的基团。在一些实施例中, $o$ 为0且 $p$ 和 $q$ 二者为1,并且 $R^2$ 包含具有式(IVC)的基团,并且其中 $e$ 在1至16的范围

内。在一些实施例中， $o$ 为0且 $p$ 和 $q$ 二者为1，并且 $R^2$ 包含具有式(IVC)的基团，并且其中 $e$ 在1至12的范围内。在一些实施例中， $o$ 为0且 $p$ 和 $q$ 二者为1，并且 $R^2$ 包含具有式(IVC)的基团，并且其中 $e$ 在1至8的范围内。在一些实施例中， $o$ 为0且 $p$ 和 $q$ 二者为1，并且 $R^2$ 包含具有式(IVC)的基团，并且其中 $e$ 在1至4的范围内。在一些实施例中， $o$ 为0且 $p$ 和 $q$ 二者为1，并且 $R^2$ 包含具有式(IVC)的基团， $d$ 为2或3，并且其中 $e$ 在1至16的范围内。在一些实施例中， $o$ 为0且 $p$ 和 $q$ 二者为1，并且 $R^2$ 包含具有式(IVC)的基团， $d$ 为2或3，并且其中 $e$ 在1至12的范围内。在一些实施例中， $o$ 为0且 $p$ 和 $q$ 二者为1，并且 $R^2$ 包含具有式(IVC)的基团， $d$ 为2或3，并且其中 $e$ 在1至10的范围内。在一些实施例中， $o$ 为0且 $p$ 和 $q$ 二者为1，并且 $R^2$ 包含具有式(IVC)的基团， $d$ 为2或3，并且其中 $e$ 在1至8的范围内。在一些实施例中， $o$ 为0且 $p$ 和 $q$ 二者为1，并且 $R^2$ 包含具有式(IVC)的基团， $d$ 为2或3，并且其中 $e$ 在1至6的范围内。在一些实施例中， $o$ 为0且 $p$ 和 $q$ 二者为1，并且 $R^2$ 包含具有式(IVC)的基团， $d$ 为2或3，并且其中 $e$ 在1至4的范围内。

[0231] 如上所述，在一些实施例中， $L^2$ 可以为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在另一些实施例中， $L^2$ 包含具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在又一些实施例中， $L^2$ 包含具有1个与约8个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。

[0232] 在进一步的实施例中， $L^2$ 包含具有1个与约6个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在更进一步的实施例中， $L^2$ 包含具有1个与约4个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团，任选地被一个或多个选自O、N或S的杂原子取代，并且任选地包含一个或多个羰基基团。在一些实施例中，基团 $L^2$ 可以包含一个或多个增溶基团，例如PEG基团。在一些实施例中，羰基基团选自酮、酰胺和羧基。在一些实施例中，基团 $L^2$ 包含酮。在一些实施例中，基团 $L^2$ 包含酰胺。

[0233] 在一些实施例中，式(I)化合物包含用于切割式(I)化合物的切割位点。在一些实施例中，切割位点位于 $Olig_2$ 内。在这些实施例中， $v$ 为0并且不存在W基团。在这些实施例中， $Olig_2$ 可以包含例如至少一个含尿嘧啶的核苷酸。在一些实施例中，可以通过添加尿嘧啶-N-DNA糖基化酶(UNG)任选地在伯胺的存在下切割含尿嘧啶的核苷酸，如美国专利号8,669,061中所述。在一些实施例中，通过糖基化酶和核酸内切酶的组合，例如通过尿嘧啶DNA糖基化酶(UDG)和DNA糖基化酶-裂解酶核酸内切酶VIII的混合物进行切割。

[0234] 在另一些实施例中，切割位点位于 $Olig_2$ 外部，诸如在基团W中。在一些实施例中，如上所述，W包含具有1个与约12个之间碳原子的经取代或未取代的、饱和或不饱和脂肪族或芳香族基团，任选地被一个或多个选自O、N、S的杂原子取代，前提是W包含可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团。在另一些实施例中，W包含具有1个与约8个之间碳原子的经取代或未取代的、饱和或不饱和脂肪族或芳香族基团，任选地被一个或多个选自O、N、S的杂原子取代，前提是W包含可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团。

[0235] 在又一些实施例中,W包含具有1个与约6个之间碳原子的经取代或未取代的、饱和或不饱和脂肪族或芳香族基团,任选地被一个或多个选自O、N、S的杂原子取代,前提是W包含可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团。在进一步的实施例中,W包含具有1个与约4个之间碳原子的经取代或未取代的、饱和或不饱和脂肪族或芳香族基团,任选地被一个或多个选自O、N、S的杂原子取代,前提是W包含可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团。

[0236] 在一些实施例中,W包含至少一个可光切割部分。在一些实施例中,可光切割部分可以在暴露于波长在约200nm至约400nm (UV) 或约400nm至约800nm (可见) 之间的电磁辐射源时即被切割。合适的可光切割部分的实例包括但不限于芳基羰基甲基基团(例如4-乙酰基-2-硝基苄基、二甲基苯甲酰甲基(DMP)); 2-(烷氧基甲基)-5-甲基- $\alpha$ -氯苯乙酮、2,5-二甲基苯甲酰基环氧乙烷、苯偶姻基团(例如3',5'-二甲氧基苯偶姻(DMB))、邻硝基苄基基团(例如1-(2-硝基苯基)乙基(NPE)、1-(甲氧基甲基)-2-硝基苯、4,5-二甲氧基-2-硝基苄基(DMNB)、 $\alpha$ -羧基硝基苄基( $\alpha$ -CNB)); 邻硝基-2-苯乙基氧基羰基(例如1-(2-硝基苯基)乙基氧基羰基和2-硝基-2-苯乙基衍生物); 邻硝基苯胺(例如酰化5-溴-7-硝基二氢吲哚); 香豆素-4-基-甲基基团(例如7-甲氧基香豆素衍生物); 9-取代的咕吨和芳甲基基团(例如邻羟基芳甲基)。

[0237] 在一些实施例中,至少一个可光切割部分可以在暴露于波长为在约700nm至约1000nm之间的电磁辐射源时即被切割。合适的可近红外光切割基团包括花菁基团,包括C4二烷基胺取代的七甲川菁。

[0238] 在一些实施例中,W包含至少一个可化学切割部分。在一些实施例中,可化学切割部分是基团,该基团可以被不同的化学反应物(包括还原剂),或通过诱导的pH变化(例如,基团在小于pH 7下切割)化学切割。可化学切割部分的非限制性实例包括基于二硫键的基团;重氮苯基团(例如2-(2-烷氧基-4-羟基-苯偶氮基);苯甲酸支架;基于酯键的基团;和酸敏感基团(例如二烷氧基二苯基硅烷基团或酰基脲基团)。认为亲电切割的基团(例如对烷氧基苄基酯和对烷氧基苄基酰胺)被质子切割并且包括对酸敏感的切割。

[0239] 在一些实施例中,W包含至少一个可酶促切割部分。在一些实施例中,可酶促切割部分可以被例如胰蛋白酶可切割基团和V8蛋白酶可切割基团切割。在一些实施例中,可酶促切割部分可以被USER酶、尿嘧啶-N-糖基化酶、RNA酶A、 $\beta$ -葡糖醛酸酶、 $\beta$ -半乳糖苷酶或TEV-蛋白酶之一酶促切割。

[0240] 本公开的另一面是一种具有式(II)的化合物:

[0241]  $[Olig1] - ([R1]_o - [R2]_p)_q - [L1]_t - [X]$  (II),

[0242] 其中

[0243] o为0或1;

[0244] p为0或1;

[0245] q为1或2;

[0246] t为0、1或2;

[0247]  $R^1$ 为具有约1个与约24个之间核苷酸的寡核苷酸;

[0248]  $R^2$ 为具有2个与约48个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地被一个或多个选自O、N或S的杂原子取代;

[0249]  $L^1$ 为具有1个与约16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地被一个或多个选自O、N或S的杂原子取代,并且任选地包含一个或多个羰基基团;

[0250] X为二苯并环辛炔、反式环辛烯、炔烃、烯炔、叠氮化物、四嗪、马来酰亚胺、N-羟基琥珀酰亚胺、硫醇、1,3-硝酮、醛、酮、肼、羟胺、氨基基团或亚磷酰胺;并且

[0251] Olig1为包含约1个与约30个之间核苷酸的寡核苷酸。

[0252] 在一些实施例中,Olig1包含约1个与约24个之间核苷酸。在另一些实施例中,Olig1包含约1个与约16个之间核苷酸。在一些实施例中,Olig1具有不可延伸的3'端。

[0253] 在一些实施例中,Olig1包含能够与已知融合配偶体杂交的锚定序列。融合配偶体的非限制性实例包括ALK、PPARG、BRAF、EGFR、FGFR1、FGFR2、FGFR3、MET、NRG1、NTRK1、NTRK2、NTRK3、RET、ROS1、AXL、PDGFRA、PDGFB、ABL1、ABL2、AKT1、AKT2、AKT3、ARHGAP26、BRD3、BRD4、CRLF2、CSF1R、EPOR、ERBB2、ERBB4、ERG、ESR1、ESRRA、ETV1、ETV4、ETV5、ETV6、EWSR1、FGR、IL2RB、INSR、JAK1、JAK2、JAK3、KIT、MAML2、MAST1、MAST2、MSMB、MUSK、MYB、MYC、NOTCH1、NOTCH2、NUMBL、NUT、PDGFRB、PIK3CA、PKN1、PRKCA、PRKCB、PTK2B、RAF1、RARA、RELA、RSP02、RSP03、SYK、TERT、TFE3、TFEB、THADA、TMPRSS2、TSLP、TY、BCL2、BCL6、BCR、CAMTA1、CBFB、CCNB3、CCND1、CIC、CRFL2、DUSP22、EPC1、FOXO1、FUS、GLI1、GLIS2、HMGA2、JAZF1、KMT2A、MALT1、MEAF6、MECOM、MKL1、MKL2、MTB、NCOA2、NUP214、NUP98、PAX5、PDGFB、PICALM、PLAG1、RBM15、RUNX1、RUNX1T1、SS18、STAT6、TAF15、TAL1、TCF12、TCF3、TFG、TYK2、USP6、YWHAE、AR、BRCA1、BRCA2、CDKN2A、ERBB4、FLT3、KRAS、MDM4、MYBL1、NF1、NOTCH4、NUTM1、PRKACA、PRKACB、PTEN、RAD51B和RBI1。

[0254] 本公开的另一方面是一种具有式(II)的化合物:

[0255]  $[Y] - [L^2]_u - [W]_v - [Olig2]$  (III),

[0256] 其中

[0257] u为0、1或2;

[0258] v为0或1;

[0259] Y为二苯并环辛炔、反式环辛烯、炔烃、烯炔、叠氮化物、四嗪、马来酰亚胺、N-羟基琥珀酰亚胺、硫醇、1,3-硝酮、醛、酮、肼、羟胺、氨基基团或亚磷酰胺;

[0260]  $L^2$ 为具有1个与16个之间碳原子的经取代或未取代的、饱和或不饱和、直链或环状脂肪族基团,任选地包含一个或多个选自O、N或S的杂原子,并且任选地包含一个或多个羰基基团;

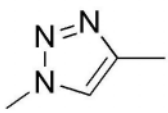
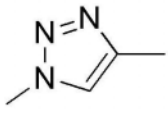
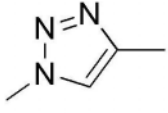
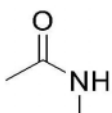
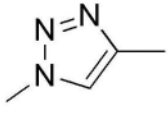
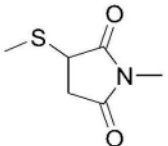
[0261] W为具有1个与12个之间碳原子的经取代或未取代的、饱和或不饱和脂肪族或芳香族基团,任选地被一个或多个选自O、N、S的杂原子取代,前提是W包含至少一个可光切割的、可酶促切割的、可化学切割的、或pH敏感的基团;并且

[0262] Olig2为包含约1个与约30个之间核苷酸的寡核苷酸。

[0263] 在一些实施例中,Olig2包含约1个与约24个之间核苷酸。在另一些实施例中,Olig2包含约1个与约16个之间核苷酸。在又一些实施例中,Olig2包含约1个与约12个之间核苷酸。在一些实施例中,Olig2包含不可延伸的3'端。在一些实施例中,Olig2包含随机序列。在另一些实施例中,Olig2包含单个重复核苷酸,例如聚T寡核苷酸。

[0264] 式(I)、(II)和(III)化合物的制备

[0265] 本领域技术人员将理解式 (II) 和式 (III) 的化合物可以彼此反应以形成具有式 (I) 的化合物。在一些实施例中, 式 (I) 的基团 Z 分别由式 (II) 和 (III) 的 X 和 Y 基团确定。表 2 列出了式 (II) 和 (III) 的 X 和 Y 基团以及所形成的具有式 (I) 的化合物的基团 Z。

[0266]	<u>X</u>	<u>Y</u>	<u>Z</u>
	炔烃	叠氮化物	三唑 
	叠氮化物	炔烃	三唑 
	二芳基环辛炔 (“DBCO”)	叠氮化物	三唑 
	烯烃	四嗪	二氢哒嗪
	反式环辛烯 (“TCO”)	四嗪	二氢哒嗪
	马来酰亚胺	硫醇	酰胺键合 
[0267]	DBCO	1,3-硝酮	异噁唑啉
	醛或酮	肼	脞
	醛或酮	羟基胺	脞醚
	叠氮化物	DBCO	三唑 
	四嗪	TCO	二氢哒嗪
	硫醇	马来酰亚胺	硫醚键合 
	1,3-硝酮	DBCO	异噁唑啉
	肼	醛或酮	脞
	羟基胺	醛或酮	脞醚

	四嗪	烯烃	二氢吡嗪
[0268]	氨基 (在 s-三氯三嗪的存在下)	氨基 (在 s-三氯三嗪的存在下)	氯-s-三嗪键合

[0269] 在一些实施例中,式(II)和(III)的基团0lig1和0lig2根据本领域普通技术人员已知的方法制备。在一些实施例中,基团0lig1和0lig2是使用固相合成技术采用亚磷酸胺化学合成的(参见例如Protocols for Oligonucleotides and Analogs,Agrawal,S.,编辑,Humana Press,Totowa,N.J.,1993,通过引用整体并入本文)。其他合成0lig1、0lig2和/或式(II)和(III)化合物的方法描述于美国专利号5,955,591、6,057,431、8,889,843和6,124,445中;以及美国专利公开号2008/0119645和2003/0153743中,其公开内容通过引用整体并入本文。

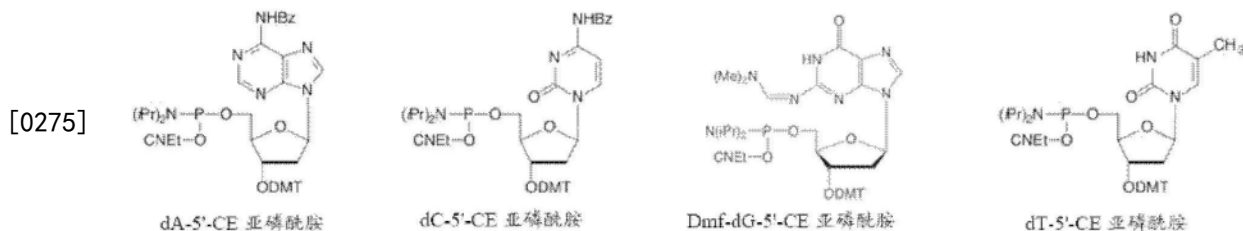
[0270] 在一些实施例中,这种方法的第一步是使用本领域已知的标准方法和程序,通常通过接头,将含有受保护的5'-羟基的第一单体或更高级亚基连接至固体支持物。参见例如,Oligonucleotides and Analogues A Practical Approach,Ekstein,F.Ed.,IRL Press,N.Y,1991。然后处理支持物结合的单体或更高级的第一合成子以去除5'-保护基团。在一些实施例中,这通过用酸处理来实现。在一些实施例中,固体支持物结合的单体然后与亚磷酸胺反应以形成亚磷酸酯键合。在一些实施例中,含亚磷酸酯的化合物被氧化以产生具有所期望核苷酸间键合的化合物。在一些实施例中,氧化剂的选择将决定亚磷酸酯键合是否将被氧化成例如磷酸三酯、硫代磷酸三酯或二硫代磷酸三酯键合。

[0271] 在一些实施例中,加帽步骤在亚磷酸三酯、硫代亚磷酸三酯或二硫代亚磷酸三酯氧化之前或之后进行。在一些实施例中,加帽步骤涉及将“帽”部分连接到在给定偶联循环中未反应的寡核苷酸链。在一些实施例中,帽部分与不参与偶联循环的寡核苷酸的末端部分反应,但不与确实参与的寡核苷酸反应,此外,它本身不与偶联试剂反应。

[0272] 用酸进一步处理氧化的寡聚物会去除5'-羟基保护基团,从而将固体支持物结合的寡聚物转化为进一步的化合物,该化合物可以随后反应以开始下一合成迭代。重复该过程直到产生所期望长度的寡聚物。

[0273] 在一些实施例中,式(II)和(III)的化合物可以反应形成式(I)化合物。在这些实施例中,可在式(II)化合物和这些式(II)化合物之间形成5'至5'键合。在一些实施例中,具有式(II)的化合物例如使用上述程序在3'至5'方向上合成。这种合成可以使用3'亚酰胺进行。

[0274] 式(III)化合物也可以类似方式合成,但使用5'亚酰胺代替3'亚酰胺。5'亚酰胺的非限制性实例如下所述。以这种方式,式(III)化合物可以在5'至3'方向合成。在一些实施例中,式(II)和(III)的化合物可以通过磷酸酯键合连接。



[0276] 在一些实施例中,式(II)和式(III)的化合物可以使用“点击化学”彼此反应。“点

击化学”是一种化学原理,其由Sharpless和Meldal的研究组独立地定义,描述了定制以通过将小单元连接在一起而快速可靠地生成物质的化学过程。“点击化学”已应用于一组可靠并且自主的有机反应(Kolb,H.C.;Finn,M.G.;Sharpless,K.B.Angew.Chem.Int.Ed.2001,40,2004-2021)。例如,将铜催化的叠氮化物-炔烃[3+2]环加成反应鉴定为水中高度可靠的分子连接(Rostovtsev,V.V.等人,Angew.Chem.Int.Ed.2002,41,2596-2599),其已经用于增强多种类型的生物分子相互作用的研究(Wang,Q.等人,J.Am.Chem.Soc.2003,125,3192-3193;Speers,A.E.等人,J.Am.Chem.Soc.2003,125,4686-4687;Link,A.J.;Tirrell,D.A.J.Am.Chem.Soc.2003,125,11164-11165;Deiters,A.等人,J.Am.Chem.Soc.2003,125,11782-11783)。此外,还已出现在有机合成(Lee,L.V.等人,J.Am.Chem.Soc.2003,125,9588-9589)、药物发现(Kolb,H.C.;Sharpless,K.B. Drug Disc.Today 2003,8,1128-1137;Lewis,W.G.等人,Angew.Chem.Int.Ed.2002,41,1053-1057)以及表面功能化(Meng,J.-C.等人,Angew.Chem.Int.Ed.2004,43,1255-1260;Fazio,F.等人,J.Am.Chem.Soc.2002,124,14397-14402;Collman,J.P.等人,Langmuir 2004,ASAP,in press;Lummerstorfer,T.;Hoffmann,H.J.Phys.Chem.B 2004,in press)方面的应用。

[0277] 在一些实施例中,首先修饰式(II)化合物的前体以引入能够参与“点击化学”反应的反应性官能团对的第一成员。同样,在一些实施例中,修饰式(III)化合物的前体以引入能够参与“点击化学”反应的反应性官能团对的第二成员。在一些实施例中,能够参与“点击化学”反应的反应性官能团对的第一和第二成员在表1中确定。在一些实施例中,“点击化学”反应由引入的试剂催化。在一些实施例中,引入的试剂是Cu<sup>+</sup>。

[0278] 表1:反应性官能团对的第一和第二成员。

[0279]

一对点击缀合物的第一个成员上的反应性官能团	一对点击缀合物的第二个成员上的反应性官能团
炔烃	叠氮化物
叠氮化物	炔烃
二芳基环辛炔 (“DBCO”)	叠氮化物
烯烃	四嗪
反式环辛烯 (“TCO”)	四嗪
马来酰亚胺	硫醇
DBCO	1,3-硝酮
醛或酮	肼
醛或酮	羟基胺
叠氮化物	DBCO
四嗪	TCO
硫醇	马来酰亚胺
1,3-硝酮	DBCO
肼	醛或酮
羟基胺	醛或酮
四嗪	烯烃

[0280] 仅举例来说,可以修饰式(II)化合物的前体以引入伯卤素。随后,可引入叠氮化钠,其与伯卤素反应使得式(II)化合物的前体转化为叠氮化物。在一些实施例中,式(II)化合物的前体与包含伯卤素的亚酰胺直接或通过接头间接反应。合适的亚酰胺的非限制性实例如下所示:

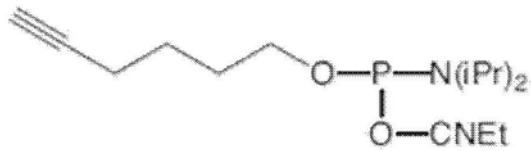
[0281]



5'-溴己基亚磷酸胺

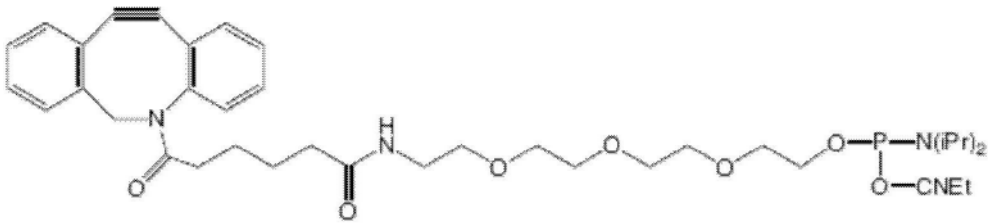
[0282] 再次举例来说,可以修饰(例如用亚酰胺)式(III)化合物的前体以引入与式(II)的叠氮化物反应的部分,诸如包含烷基的部分。下文提供了合适的亚酰胺的非限制性实例:





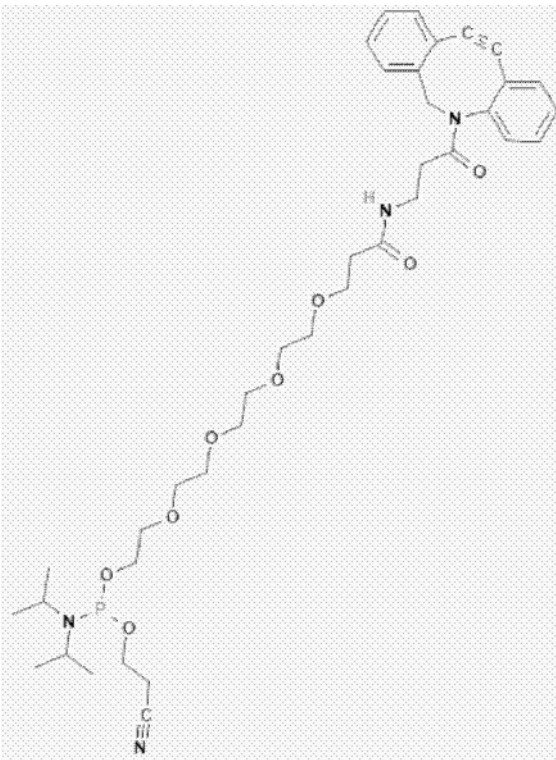
5'-己炔基亚磷酰胺

[0283]



5'-DBCO-TEG

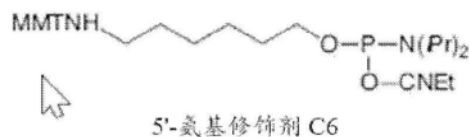
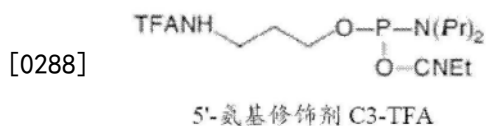
[0284] 另一种合适的试剂是DBCO-PEG-亚磷酰胺,例如DBCO-PEG4-亚磷酰胺:



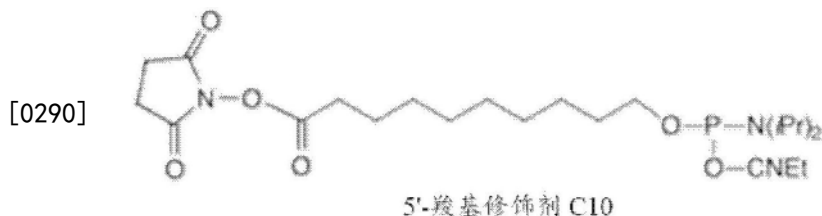
[0285]

[0286] 然后允许所得的式 (II) 和 (III) 的化合物,其各自带有能够参与“点击化学”反应的反应性基团的成员,彼此反应以形成5'至5'键合。在上面提供的实例中,叠氮化物和炔烃将反应形成三唑键合。

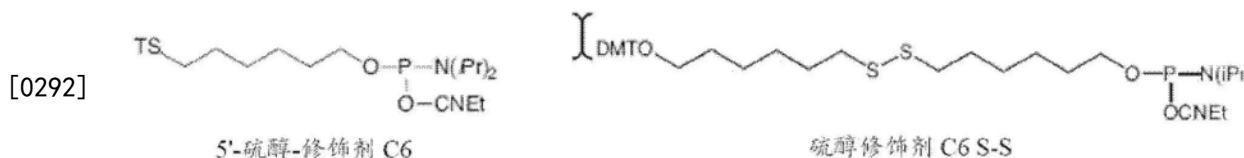
[0287] 在一些实施例中,式 (II) 和 (III) 的化合物可以各自包含促进在化合物之间形成酰胺键合的反应性基团(分别为X和Y)。为了实现这一点,在一些实施例中,式 (II) 和 (III) 中的每一个的化合物的前体可以与分别引入基团X和Y的试剂反应。在这些实施例中,具有式 (II) 的化合物的前体在5'端用氨基部分修饰。例如,可以将亚酰胺引入具有式 (II) 的化合物的前体,其中亚酰胺包含末端氨基部分。此类亚酰胺试剂的非限制性实例包括以下:



[0289] 类似地,具有式(III)的化合物的前体也可以在5'端修饰以终止于羧基。例如,可以将亚酰胺引入具有式(III)的化合物的前体,其中亚酰胺包含末端羧基部分。此类亚酰胺试剂的非限制性例子是:



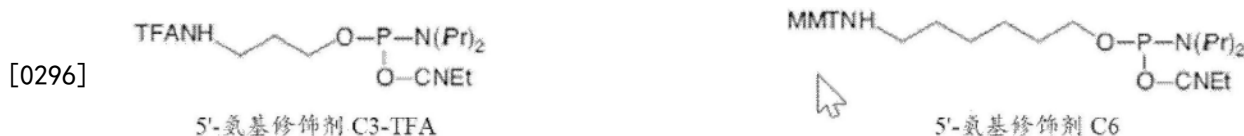
[0291] 在一些实施例中,式(II)和(III)的化合物可以各自包含促进在化合物之间形成硫醚键合的反应性基团(分别为X和Y)。为了实现这一点,在一些实施例中,式(II)和(III)中的每一个的化合物的前体可以与分别引入基团X和Y的试剂反应。在这些实施例中,具有式(II)的化合物的前体在5'端用硫醇部分修饰。例如,可以将亚酰胺引入具有式(II)的化合物的前体,其中亚酰胺包含末端硫醇部分。此类亚酰胺试剂的非限制性实例包括以下:



[0293] 具有式(III)的化合物也可以在5'端修饰以终止于马来酰亚胺基团。例如,可以将亚酰胺引入具有式(III)的化合物的前体,其中亚酰胺包含末端马来酰亚胺部分。此类亚酰胺试剂的非限制性例子是:

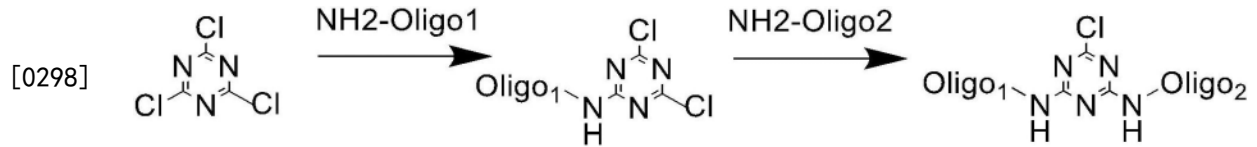


[0295] 在一些实施例中,式(II)和(III)的化合物可以各自包含促进在化合物之间形成三嗪键合的反应性基团(分别为X和Y)。为了实现这一点,在一些实施例中,式(II)和(III)中的每一个的化合物的前体可以与分别引入基团X和Y的试剂反应。在一些实施例中,三嗪键合是氯-s-三嗪键合。在这些实施例中,具有式(II)的化合物的前体在5'端用氨基部分修饰。同样,具有式(III)的化合物的前体在5'端用氨基部分修饰。用于引入此类5'氨基基团的合适的亚酰胺的非限制性实例如下所述:

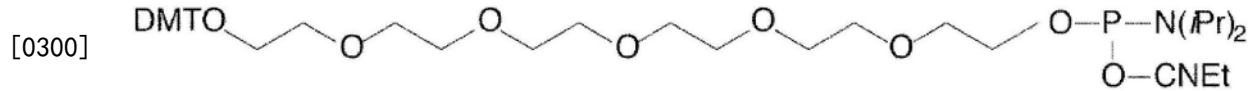


[0297] 在对式(II)化合物的前体和式(III)化合物的前体二者进行修饰之后,所形成的式(II)和(III)的化合物然后与偶联剂反应。在一些实施例中,偶联试剂为s-三氯三嗪。该

反应如下图示：

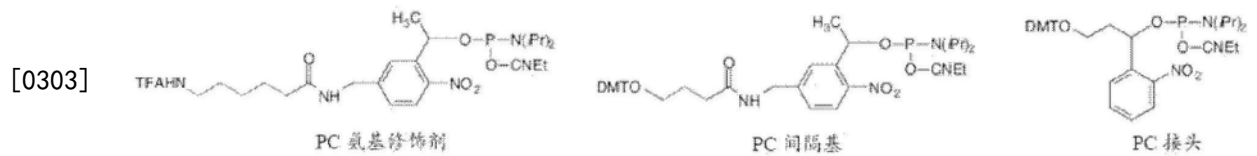


[0299] 在一些实施例中，式 (II) 或 (III) 化合物的任何前体可以反应以引入接头或间隔基，诸如基于PEG的接头或间隔基。引入基于PEG的接头或间隔基的合适试剂的非限制性实例如下所述：



[0301] 用于将基于PEG的接头或间隔基掺入式 (II) 和/或 (III) 化合物的前体中的其他试剂和方法描述于美国专利公开号2006/0063147中，其公开内容通过引用整体并入本文。

[0302] 在一些实施例中，式 (II) 或 (III) 化合物的任何前体可以反应以引入接头或间隔基，诸如包含可切割基团的接头或间隔基。合适试剂的非限制性实例如下所述：



#### [0304] 试剂盒

[0305] 本公开的另一方面是试剂盒，诸如包含一种或多种式 (I) 化合物的试剂盒。在一些实施例中，试剂盒包含一种或多种式 (I) 化合物和聚合酶。在一些实施例中，聚合酶是DNA聚合酶。在一些实施例中，DNA聚合酶是热稳定的DNA依赖性DNA聚合酶。试剂盒可进一步包含扩增引物。在一些实施例中，试剂盒进一步包含正向引物和/或反向引物中的至少一种。在一些实施例中，试剂盒包含能够与第一寡核苷酸的拷贝杂交的正向引物和能够与第二寡核苷酸杂交的反向引物。在另一些实施例中，试剂盒包含能够与第一寡核苷酸杂交的正向引物和能够与第二寡核苷酸的拷贝杂交的反向引物。

[0306] 在另一些实施例中，试剂盒可包含式 (I)、(II) 或 (III) 化合物中的一种或多种和一种或多种缓冲液。在一些实施例中，试剂盒包含一种或多种式 (I) 化合物和主混合物。在一些实施例中，主混合物包含酶、缓冲液、辅因子 (例如MgCl<sub>2</sub>或MgSO<sub>4</sub>)、水和dNTP中的两种或更多种。在一些实施例中，主混合物还包含模板DNA。

[0307] 在另一些实施例中，试剂盒可包含式 (II) 化合物和式 (III) 化合物。在一些实施例中，式 (II) 化合物包含第一反应性基团，其能够与式 (III) 化合物的第二反应性基团反应。

[0308] 在一些实施例中，第一反应性基团包含炔烃部分；并且第二反应性基团包含叠氮化物部分。在一些实施例中，炔烃部分为DBCO。在一些实施例中，第一反应性基团包含马来酰亚胺部分；并且第二反应性基团包含硫醇部分。在一些实施例中，第一反应性基团包含烯炔部分并且第二反应性基团包含四嗪部分。在一些实施例中，第一和第二反应性基团二者是氨基部分，并且其中试剂盒进一步包含s-三氯三嗪。

[0309] 在一些实施例中，式 (I)、(II) 和/或 (III) 的任何化合物可以与一种或多种另外的组分一起包含在反应容器中。如本文所用，术语“反应容器”通常是指可在其中根据本教导发生反应的任何容器、腔室、装置或组件。在一些实施例中，反应容器包括dPCR芯片的孔。在

一些实施例中，dPCR芯片可以包括例如蚀刻有纳米级或更小的反应孔的硅衬底。在一些实施例中，dPCR芯片具有低热质量。例如，芯片可以由不储存热能的薄、高传导材料构成。在一些实施例中，dPCR芯片具有约 $50\text{mm}^2$ 至约 $150\text{mm}^2$ 的表面积。在一些实施例中，dPCR芯片具有约 $100\text{mm}^2$ 的表面积。限制表面积可以允许在熔体分析过程中芯片的加热更加均匀，并减少熔体固化分析中的运行间变化，减少熔体曲线生成中的误差，并增加分析中熔解曲线的辨别力。其他dPCR芯片描述在PCT公开号W0/2016/133783中，其公开内容通过引用整体并入本文。

#### [0310] 方法

[0311] 本公开的另一方面是一种检测一种或多种基因融合的方法，其中一种融合配偶体是未知的。在一些实施例中，该方法利用一种或多种式(I)化合物。在一些实施例中，该方法进一步包括扩增核酸和/或形成经扩增的核酸的文库。在一些实施例中，该方法进一步包括对经扩增的核酸的文库进行测序，从而检测样品中的一种或多种基因组重排。本文描述了方法的这些和其他步骤。

[0312] 基因融合在癌症中很常见。用于基因融合的临床测试能够检测和诊断癌症，随着时间跟踪肿瘤负荷，以及为癌症患者开发个体化治疗方案。特别有用的是检测基因融合的基于血液的方法。基于血液的方法获取患者的无细胞核酸(cfDNA和cfRNA)，其包含循环肿瘤核酸(ctDNA和ctRNA)。虽然基于血液的测试比活检的侵入性更小，但主要困难是检测与正常的非肿瘤来源核酸混合的极少量的肿瘤来源核酸。多种商业上可用的测试能够检测ctDNA中的突变，包括单核苷酸变异(SNV)、拷贝数变异(CNV)和基因融合(例如，AVENIO ctDNA测试试剂盒，Roche Sequencing Solutions, Pleasanton, Cal.)

[0313] 对于一些癌症相关的基因融合，检测ctDNA中的融合产物会因多种融合配偶体的出现而变得更加复杂。具有混杂融合的肿瘤相关基因包括许多实例，诸如NTRK 1、NTRK 2和NTRK 3，以及FGFR 2和FGFR 3。

#### [0314] 样品

[0315] 本公开的方法利用包含一种或多种核酸(包括一种或多种靶标核酸)的样品。在一些实施例中，该样品获自受试者或患者。在一些实施例中，该样品可包括例如通过活检而获自该受试者或患者的固体组织或实体肿瘤的片段。所述样品还可包括体液(例如尿液、痰、血清、血浆或淋巴、唾液、痰、汗液、泪液、脑脊液、羊水、滑液、心包液、腹膜液、胸膜液、囊液、胆汁、胃液、肠液或粪便样品)。样品可以包括全血或其中可能存在正常细胞或肿瘤细胞的血液级分。在一些实施例中，该样品，特别是液体样品可包含无细胞材料，诸如无细胞DNA或RNA，包括无细胞胎儿DNA或胎儿RNA的无细胞肿瘤DNA或肿瘤RNA。在一些实施例中，该样品是无细胞样品，例如，存在无细胞肿瘤DNA或肿瘤RNA或无细胞胎儿DNA或胎儿RNA的无细胞血源性样品。在另一些实施例中，样品是培养样品，例如，培养物或者含有或疑似含有来源于培养物中的细胞的核酸的培养物上清液。

[0316] 在一些实施例中，样品为代表性样品。在一些实施例中，代表性样品由肿瘤样品、淋巴结样品、血液样品和/或均质化的其他组织样品(单独或一起)制备。“均质化”是指一种过程(诸如机械过程和/或生化过程)，由此使生物样品达到使得样品的所有部分在组成上相等的状态。代表性样品(如本文所定义)可以通过去除已经均质化的样品的一部分来制备。将均质化样品(“匀浆物”)充分混合，使得去除样品的一部分(等分试样)基本上不会改

变剩余样品的整体组成,并且去除的等分试样的组分与剩余样品的组分基本相同。在本公开中,“均质化”通常将保持样品内大多数细胞的完整性,例如均质过程的结果是样品中至少50%的细胞不会破裂或溶解。在另一些实施例中,均质化将保持样品中至少80%的细胞的完整性。在另一些实施例中,均质化将保持样品中至少85%的细胞的完整性。在另一些实施例中,均质化将保持样品中至少90%的细胞的完整性。在另一些实施例中,均质化将保持样品中至少95%的细胞的完整性。在另一些实施例中,均质化将保持样品中至少96%的细胞的完整性。在另一些实施例中,均质化将保持样品中至少97%的细胞的完整性。在另一些实施例中,均质化将保持样品中至少98%的细胞的完整性。在另一些实施例中,均质化将保持样品中至少99%的细胞的完整性。在另一些实施例中,均质化将保持样品中至少99.9%的细胞的完整性。匀浆可以基本上分解成单个细胞(或细胞簇),并且所得的一种或多种匀浆基本上是均质的(由相似的元素组成或由其构成,或者整体上是均匀的)。

[0317] 在一些实施例中,输入样品包含来源于肿瘤样品、淋巴结样品、血液样品或其任何组合的细胞的代表性样品。在一些实施例中,输入样品来源于人类患者或哺乳动物受试者,其(i)被诊断患有癌症,(ii)疑似患有癌症,(iii)有患癌症的风险;(iv)有癌症复发或重现的风险;和/或(v)疑似有癌症复发。在另一些实施例中,输入样品来源于健康人类患者或哺乳动物受试者。产生代表性样品和/或制备用于下游处理的代表性样品的其他方法在PCT申请号PCT/US19/62857中描述,其部内容通过引用整体并入本文。

#### [0318] 靶标核酸

[0319] 靶标核酸是样品中可能存在的目标核酸。每个靶标的特征在于其核酸序列。本公开能够检测一种或多种RNA或DNA靶标。在一些实施例中,DNA靶标核酸是参与融合事件的基因或基因片段(包括外显子和内含子)或融合断点所在的基因间区域。RNA靶标核酸是融合产生的基因或编码序列的转录物或转录物的一部分。在一些实施例中,靶标核酸包括生物标志物,即基因,该基因的变体(诸如基因融合)与疾病或病症相关。例如,靶核酸可以选自于2015年9月10日递交的美国专利申请序列号14/774,518中描述的疾病相关标志物组合。此类组合可作为AVENIO ctDNA分析试剂盒(Roche Sequencing Solutions,Pleasanton, Ca1.)获得。

[0320] 特别感兴趣的是已知在肿瘤中经历基因融合的靶标基因。例如,已知ALK、RET、ROS、FGFR2、FGFR3和NTRK1会发生融合,导致异常活跃的激酶表型。已知或预期进行与癌症相关的融合的其他基因包括ALK、PPARG、BRAF、EGFR、FGFR1、FGFR2、FGFR3、MET、NRG1、NTRK1、NTRK2、NTRK3、RET、ROS1、AXL、PDGFRA、PDGFB、ABL1、ABL2、AKT1、AKT2、AKT3、ARHGAP26、BRD3、BRD4、CRLF2、CSF1R、EPOR、ERBB2、ERBB4、ERG、ESR1、ESRRA、ETV1、ETV4、ETV5、ETV6、EWSR1、FGR、IL2RB、INSR、JAK1、JAK2、JAK3、KIT、MAML2、MAST1、MAST2、MSMB、MUSK、MYB、MYC、NOTCH1、NOTCH2、NUMBL、NUT、PDGFRB、PIK3CA、PKN1、PRKCA、PRKCB、PTK2B、RAF1、RARA、RELA、RSP02、RSP03、SYK、TERT、TFE3、TFEB、THADA、TMPRSS2、TSLP、TY、BCL2、BCL6、BCR、CAMTA1、CBFB、CCNB3、CCND1、CIC、CRFL2、DUSP22、EPC1、FOXO1、FUS、GLI1、GLIS2、HMGA2、JAZF1、KMT2A、MALT1、MEAF6、MECOM、MKL1、MKL2、MTB、NCOA2、NUP214、NUP98、PAX5、PDGFB、PICALM、PLAG1、RBM15、RUNX1、RUNX1T1、SS18、STAT6、TAF15、TAL1、TCF12、TCF3、TFG、TYK2、USP6、YWHAE、AR、BRCA1、BRCA2、CDKN2A、ERBB4、FLT3、KRAS、MDM4、MYBL1、NF1、NOTCH4、NUTM1、PRKACA、PRKACB、PTEN、RAD51B和RB1。

[0321] 在一些实施例中,靶标核酸是RNA(包括mRNA)。在此类实施例中,延伸式(I)化合物的DNA聚合酶是逆转录酶。在其他实施例中,靶核酸是DNA,包括细胞DNA或无细胞DNA(cfDNA),包括循环肿瘤DNA(ctDNA)和无细胞胎儿DNA。在此类实施例中,延伸式(I)化合物的DNA聚合酶是任何DNA聚合酶,例如任何B族DNA聚合酶。靶标核酸可以以短形式或长形式存在。在一些实施例中,较长的靶核酸通过如下所述的酶促或物理处理而片段化。在一些实施例中,靶核酸是天然片段化的,例如,包括循环细胞游离DNA(cfDNA)或化学降解的DNA,诸如在化学保存的或古老样品中发现的一种。在一些实施例中,ctDNA或cfDNA来源于代表性样品(参见PCT申请号PCT/US19/62857,其公开内容通过引用整体并入本文)。

#### [0322] DNA分离

[0323] 在一些实施例中,本公开的方法包括分离核酸的步骤。通常,可以使用任何产生含有DNA、RNA或DNA和RNA的混合物的分离核酸的核酸提取方法。可以使用基于溶液或基于固相的核酸提取技术从组织、细胞、液体活检样品(包括血液或血浆样品)提取基因组DNA或细胞RNA或DNA和RNA的混合物。核酸提取可包含基于洗涤剂的细胞裂解、核蛋白质变性,以及任选地去除污染物。从保藏样品中提取核酸还可以包括脱蜡步骤。基于溶液的核酸提取方法可以包括盐析法、或者有机溶剂或离液剂法。固相核酸提取方法可以包括但不限于二氧化硅树脂法、阴离子交换法或磁性玻璃颗粒和顺磁珠(KAPA纯珠、罗氏测序解决方案公司,普莱森顿,加州)或AMPure珠(贝克曼库尔特,布雷亚市,加州)。

[0324] 典型的提取方法包含裂解样品中存在的组织材料和细胞。从裂解的细胞中释放的核酸可以与存在于溶液或柱或膜中的固体支持物(珠或颗粒)结合,其中核酸可以经历一个或多个洗涤步骤以从样品中去除包括蛋白质、脂质及其片段在内的污染物。最后,结合的核酸可以从固体支持物、柱或膜中释放,并存储在相应的缓冲液中直到准备进一步处理。因为必须分离DNA和RNA,所以不可以使用核酸酶,在纯化过程中应当注意抑制任何核酸酶活性。

[0325] 在一些实施例中,核酸分离利用如2019年10月14日提交的PCT/EP2019/077714和2018年11月13日提交的PCT/EP2018/081049中所述的加速电泳(ETP)。ETP利用具有圆形电极排列的装置,其中核酸在前导电质和尾随电解质之间迁移并浓缩。圆形配置允许将核酸浓缩在装置中心收集的非常小的体积中。ETP的使用对于在大体积中含有少量无细胞核酸的血浆样品特别有利。

[0326] 在一些实施例中,输入DNA或输入RNA需要进行片段化。在此类实施例中,RNA可以通过热和例如镁的金属离子组合来片段化。在一些实施例中,在镁存在下将样品加热至85°-94°C持续1-6分钟。(KAPA RNA HyperPrep试剂盒,KAPA生物,威尔明顿,马萨诸塞州)。DNA可以通过以下进行片段化:物理手段(例如超声),使用可商业获得的仪器(Covaris, Woburn, 马萨诸塞州)或酶的手段(KAPA片段化酶试剂盒,KAPA生物)。

[0327] 在一些实施例中,DNA修复酶靶向分离的核酸中的受损碱基。在一些实施例中,样品核酸是来自保藏样品的部分受损DNA,例如福尔马林固定石蜡包埋(FFPET)样品。碱基的脱氨作用和氧化作用会导致测序过程中错误的碱基读段。在一些实施例,用尿嘧啶N-DNA糖基化酶(UNG/UDG)和/或8-氧代鸟嘌呤DNA糖基化酶处理受损DNA。

[0328] 本公开的方法适用于多种不同类型的核酸。在一些实施例中,本公开的方法利用分离的DNA(即,通过RNA酶消化与RNA分离的DNA)。在一些实施例中,本公开的方法利用分离的RNA(即,通过DNA酶消化与DNA分离的RNA)。在另一些实施例中,本公开的方法利用DNA和

RNA的混合物(即,未用核酸酶处理的分离的核酸)。

[0329] 富集

[0330] 在一些实施例中,本公开的方法进一步包括靶标富集步骤。在一些实施例中,该方法利用寡核苷酸探针(例如,捕获探针)库。在一些实施例中,通过差减法进行富集,在这种情况下,捕获探针能够与包括核糖体RNA(rRNA)或大量表达的基因(例如珠蛋白)的大量不需要的序列杂交。在差减法的情况下,不需要的序列被捕获探针捕获,从靶标核酸的溶液中去掉并丢弃。去除可以通过利用具有结合部分的捕获探针来完成,该结合部分可以被捕获在固体支持物上。在另一些实施例中,通过保留进行富集,在这种情况下,捕获探针能够与一种或多种靶标序列(即融合配偶体基因的已知序列)杂交。在一些实施例中,靶标序列与基因特异性捕获探针杂交并从溶液中去掉,例如,利用具有可捕获在固体支持物上的结合部分的捕获探针。保留捕获的靶标-探针杂交体,而丢弃包含非靶标序列的溶液的其余部分。

[0331] 为了富集,捕获探针可以在溶液中游离或固定在固体支持物上。探针还可以包含结合部分(例如,生物素)并且能够捕获在固体支持物(例如,含有支撑材料的抗生物素蛋白或链霉抗生物素蛋白)上。

[0332] 使样品或靶标富集样品与连接的引物诸如与式(I)化合物接触

[0333] 参考图1(底部)和图2,本公开提供了一种通过使样品与连接的引物(诸如式(I)的那些中的任一种)接触来检测基因融合的方法。在一些实施例中,连接的引物包含第一寡核苷酸序列(例如,式(I)的“Olig1”),该序列直接或通过键合(例如,式(I)的基团“Z”)间接与第二寡核苷酸序列(例如,“式(I)的Olig2”)偶联。在一些实施例中,并且如图1所示,连接的引物包含第一寡核苷酸序列(左侧,式(I)的“Olig1”),该序列包含能够与已知的5'-融合配偶体杂交的锚定序列。连接的引物还包含“间隔基”(例如式(I)的基团“-([R<sup>1</sup>]<sub>o</sub>-[R<sup>2</sup>]<sub>p</sub>)<sub>q</sub>-”)。第二寡核苷酸(右侧,式(I)的“Olig2”)包含随机序列(“NNN”)和可延伸的3'-端。

[0334] 如图1(底部)所示,样品与具有聚合酶活性和链置换活性(“POL”)的核酸聚合酶接触。在一些实施例中,样品中的核酸是DNA并且使用DNA依赖性DNA聚合酶,例如具有链置换活性的任何B家族聚合酶。在一些实施例中,样品中的核酸是RNA并且使用逆转录酶。

[0335] 在一些实施例中,样品中的核酸是DNA和RNA的混合物。可以使用2019年8月19日提交的美国临时申请序列号62/888963“Single tube preparation of DNA and RNA for sequencing,”中描述的方法处理这样的样品,以在单个试管中靶向DNA和RNA,所述申请通过引用并入本文。简而言之,所述方法包括在DNA起始材料不反应条件下,用具有识别RNA起始材料的标签的第一引物形成cDNA。形成cDNA后,靶标cDNA与靶标DNA一起通过不包括第一个引物的通用扩增引物组进行扩增和检测。源自RNA的最终产物与源自DNA的最终产物的区别在于通过第一引物引入的RNA特异性标签(“RNA识别标签”)的存在。在一些实施例中,第二寡核苷酸(例如式(I)的“Olig2”)的5'-部分包含RNA识别标签。

[0336] 在一些实施例中,聚合酶延伸第二寡核苷酸(例如式(I)的“Olig2”)的3'端,同时置换与已知基因融合配偶体的已知序列杂交的第一寡核苷酸(例如式(I)的“Olig1”)的锚定序列。(图1,底部)。在一些实施例中,称为第一拷贝链的延伸产物包含3'-融合配偶体的一部分和5'-融合配偶体的一部分的拷贝,由此形成基因融合的第一链拷贝。

[0337] 在一些实施例中,第一拷贝链被拷贝以形成第二拷贝,由此形成基因融合的双链

拷贝。在一些实施例中,与已知融合配偶体中的序列互补的引物可用于形成第二拷贝链。在一些实施例中,该引物也是扩增引物。在一些实施例中,该引物在5'-部分中包含选自以下的一种或多种附加特征:样品条形码、分子条形码、通用引物结合位点和测序平台特异性引物结合位点。

[0338] 在一些实施例中,希望从第一拷贝链中去除第一寡核苷酸(例如式(I)的“Olig1”)。在一些实施例中,第一和第二寡核苷酸(例如式(I)的“Olig1”和“Olig2”)之间的基团(例如式(I)的基团“W”)包含可切割部分。在一些实施例中,可切割接头选自可光切割的、可酶促切割的、可化学切割的或pH敏感的基团。在包括可光切割部分的那些实施例中,可通过引入具有特定波长的辐射(例如,具有在约400nm与约800nm之间的波长范围的辐射)来切割可光切割部分。在包括可酶促切割基团的那些实施例中,可酶促切割基团可以被USER酶、尿嘧啶-N-糖基化酶、RNA酶A、 $\beta$ -葡糖醛酸酶、 $\beta$ -半乳糖苷酶或TEV-蛋白酶之一切割。在包括可化学切割基团的那些实施例中,可通过引入适当的亲电试剂和/或亲核试剂来切割可化学切割基团。

[0339] 在一些实施例中,式(I)化合物不包含基团“W”(其中 $v=0$ )并且可切割部分包含在“Olig2”内。在一些实施例中,“Olig2”包含由一个或多个含尿嘧啶核苷酸组成的切割位点。在一些实施例中,通过使反应混合物与尿嘧啶-N-DNA糖基化酶(UNG)接触来切割包含含尿嘧啶核苷酸的链(例如,第一拷贝链),任选地在伯胺的存在下进行,如美国专利号8,669,061中所述。UNG识别存在于单链或双链DNA中的尿嘧啶,并切割尿嘧啶碱基和脱氧核糖之间的N-糖苷键,留下脱碱基位点。参见美国专利号6,713,294,其公开内容通过引用整体并入本文)。

[0340] 在一些实施例中,通过糖基化酶和核酸内切酶的组合进行切割,例如,尿嘧啶DNA糖基化酶(UDG)和DNA糖基化酶-裂解酶核酸内切酶VIII的混合物。切割切割位点将第一拷贝链与第一寡核苷酸(例如式(I)的“Olig1”)和接头结构(图2,底部)分开。在一些实施例中,切割发生在形成第二拷贝链之前。

[0341] 在一些实施例中,对基因融合的第一拷贝链或双链拷贝进行测序。在一些实施例中,在测序之前,对基因融合的第一拷贝链或双链拷贝在测序之前进行扩增。如本文所述,扩增可包括基因特异性引物、特异性引物或通用引物。可以将通用引物结合位点引入连接的引物或用于形成第二拷贝链的引物的第二寡核苷酸(例如式(I)的“Olig2”)的5-部分中。

[0342] 在一些实施例中,该方法是多路复用的,这意味着该方法靶向已知参与基因融合事件的多种基因。在此类实施例中,提供了包含两种或更多种式(I)化合物的反应混合物,其中两种或更多种式(I)化合物中的每一种具有对已知参与基因融合的特异性基因的锚定序列。例如,相同的反应混合物可以包含两种或更多种式(I)化合物,所述化合物具有靶向以下一种或多种的锚定序列:ALK、PPARG、BRAF、EGFR、FGFR1、FGFR2、FGFR3、MET、NRG1、NTRK1、NTRK2、NTRK3、RET、ROS1、AXL、PDGFRA、PDGFB、ABL1、ABL2、AKT1、AKT2、AKT3、ARHGAP26、BRD3、BRD4、CRLF2、CSF1R、EPOR、ERBB2、ERBB4、ERG、ESR1、ESRRA、ETV1、ETV4、ETV5、ETV6、EWSR1、FGR、IL2RB、INSR、JAK1、JAK2、JAK3、KIT、MAML2、MAST1、MAST2、MSMB、MUSK、MYB、MYC、NOTCH1、NOTCH2、NUMBL、NUT、PDGFRB、PIK3CA、PKN1、PRKCA、PRKCB、PTK2B、RAF1、RARA、RELA、RSPO2、RSPO3、SYK、TERT、TFE3、TFEB、THADA、TMPRSS2、TSLP、TY、BCL2、BCL6、BCR、CAMTA1、CBFB、CCNB3、CCND1、CIC、CRFL2、DUSP22、EPCI、FOXO1、FUS、GLI1、GLIS2、



HMGA2、JAZF1、KMT2A、MALT1、MEAF6、MECOM、MKL1、MKL2、MTB、NCOA2、NUP214、NUP98、PAX5、PDGFB、PICALM、PLAG1、RBM15、RUNX1、RUNX1T1、SS18、STAT6、TAF15、TAL1、TCF12、TCF3、TFG、TYK2、USP6、YWHAE、AR、BRCA1、BRCA2、CDKN2A、ERBB4、FLT3、KRAS、MDM4、MYBL1、NF1、NOTCH4、NUTM1、PRKACA、PRKACB、PTEN、RAD51B和RB1。

[0343] 在一些实施例中,连接的引物经设计为适应短输入核酸。例如,包括循环肿瘤DNA(ctDNA)在内的无细胞DNA平均长度为175bp。在这样的实施例中,连接的引物的长度可以不超过175个碱基。

#### [0344] 扩增

[0345] 在一些实施例中,本公开包括扩增步骤。如图2(底部)所示,形成的拷贝链可以通过线性或指数扩增进行拷贝和扩增。扩增可以是等温的或涉及热循环。在一些实施例中,扩增是指数的并且涉及PCR。在一些实施例中,至少一种基因特异性引物(例如能够与已知融合配偶体杂交的引物)用于扩增。在一些实施例中,连接的引物的5'-部分包含用于扩增中使用的第二引物的引物结合位点。在另一些实施例中,将通用引物结合位点添加到待扩增的核酸中。在一些实施例中,可以通过连接包含通用引物结合位点的衔接子来添加通用引物结合位点。在另一些实施例中,通过延伸具有包含通用引物结合位点的5'-尾的基因特异性引物来添加通用引物结合位点。具有相同通用引物结合位点的所有核酸可以方便地用相同的引物组和在相同的条件下进行扩增。使用通用引物的扩增循环的数量可以较低,但也可以为约10个、约20个或高达约30个或更多个循环,这取决于后续步骤所需的产物量。由于使用通用引物的PCR降低了序列偏倚,因此无需为了避免扩增偏倚而限制扩增循环数。

#### [0346] 引物

[0347] 在一些实施例中,本公开涉及利用正向引物和反向引物的扩增步骤。正向引物和反向引物中的一者或两者可以是靶特异性的。靶特异性引物包含对靶标核酸特异的(即至少部分互补并与之形成稳定杂交体)的至少3'-部分。如果存在另外的序列,诸如条形码或通用引物结合位点,它们通常位于该引物的5'部分。

[0348] 在一些实施例中,为了扩增如图2(底部)所示形成的拷贝链,可以使用对融合断点上游的已知基因序列特异的第一引物。在一些实施例中,第二引物对第二连接的寡核苷酸中存在的标签序列或任何其他工程化序列是特异的。

[0349] 在一些实施例中,第一和第二特异性引物在引物的5'-部分中包含通用引物结合位点。在一轮或多轮特异性扩增后,进行通用扩增。

#### [0350] 文库

[0351] 在一些实施例中,本公开是针对如本文所述的融合特异性核酸进行富集的核酸文库。该文库包含两侧连接有如下所述的衔接子序列的双链核酸分子。文库中的核酸可以包含元件,诸如存在于衔接子序列中的条形码和通用引物结合位点,如下文所述。在一些实施例中,额外的元件存在于衔接子中并通过衔接子连接添加到文库的核酸中。在其他实施例中,一些或所有额外的元件存在于扩增引物中并且在衔接子连接之前通过引物的延伸添加到文库的核酸中。

[0352] 在一些实施例中,文库是在使用如本文所述的连接的引物进行融合检测之前由样品中的所有核酸形成的。在该实施例中,衔接子分子被添加到样品中的所有核酸中。利用连接的引物检测融合的方法使用文库分子作为起始材料。在一些实施例中,通用扩增(使用与

位于衔接子中的引物结合位点杂交的通用引物)发生在利用连接的引物进行融合特异性扩增之前。通用扩增增加了用于使用如本文所述进行的利用连接的引物进行融合特异性扩增的起始材料的量。

[0353] 在一些实施例中,文库分子包括包含唯一分子条形码的衔接子。对文库进行测序包括确定条形码化文库核酸的序列,通过唯一分子条形码将序列分组为家族,以及确定每个家族的共有序列读数,从而检测基因融合。

#### [0354] 衔接子

[0355] 在一些实施例中,本公开利用衔接子核酸。衔接子可以通过平端连接或粘性末端连接添加到核酸中。在一些实施例中,可以通过单链连接添加衔接子。在一些实施例中,通过使用在引物的5'-部分具有衔接子序列的平铺引物进行扩增来添加衔接子。可用于通过连接或扩增添加衔接子的方法和组合物描述于例如美国专利号9476095、9260753、8822150、8563478、7741463、8182989和8053192中,其公开内容通过引用整体并入本文。

[0356] 在一些实施例中,衔接子分子为在体外合成的人工序列。在另一些实施例中,衔接子分子为体外合成的天然存在的序列。在又一些实施例中,衔接子分子为分离的天然存在的分子或分离的非天然存在的分子。

[0357] 在通过连接添加衔接子的情况下,衔接子寡核苷酸可以在末端具有突出端或平端以与靶核酸连接。在一些实施例中,衔接子包括平端,靶核酸的平端连接可以施加到所述平端。靶标核酸可以是平端的或可以通过酶处理(例如,“端修复”)而被赋予平端。在另一些实施例中,平端的DNA经历A加尾,其中单个A核苷酸被添加到一个或两个平端的3'端。本文所描述的衔接子被制成具有从平端延伸的单个T核苷酸以促进核酸和衔接子之间的连接。用于执行衔接子连接的可商购试剂盒包括AVENIO ctDNA文库制备试剂盒、或KAPA HyperPrep和HyperPlus试剂盒(Roche Sequencing Solutions,Pleasanton,Cal.)。在一些实施例中,衔接子连接的DNA可以从过量的衔接子和未连接的DNA分离。

[0358] 衔接子可以进一步包括诸如通用引物结合位点(包括测序引物结合位点)、条形码序列(包括样品条形码(SID)或独特的分子条形码或标识符(UID或UMI))的特征。在一些实施例中,衔接子包括所有上述特征,而在其他实施例中,一些特征是在衔接子连接后通过延伸含有上述一些元件的加尾的引物而添加的。

[0359] 衔接子还可包括捕获部分。捕获部分可以是能够与另一个捕获分子特异性相互作用的任何部分。捕获部分-捕获分子对包括抗生物素蛋白(链霉抗生物素蛋白)-生物素、抗原-抗体、磁性(顺磁性)颗粒-磁体或寡核苷酸-互补寡核苷酸。捕获分子可以与固体支持物结合,使得其上存在捕获部分的任何核酸被捕获在固体支持物上并与样品或反应混合物的其余部分分离。在一些实施例中,捕获分子包括用于第二捕获分子的捕获部分。例如,衔接子中的捕获部分可以是与捕获寡核苷酸互补的核酸序列。捕获寡核苷酸可以被生物素化,从而可以在链霉抗生物素蛋白磁珠上捕获经衔接的核酸-捕获寡核苷酸杂合物。

[0360] 在一些实施例中,通过捕获捕获部分和将衔接子连接的靶核酸与样品中未连接的核酸分离来富集衔接子连接的核酸。

[0361] 在一些实施例中,衔接子的茎部分包括提高捕获寡核苷酸解链温度的修饰核苷酸,例如,5-甲基胞嘧啶、2,6-二氨基嘌呤、5-羟基丁炔基-2'-脱氧尿苷、8-氮杂-7-脱氮鸟苷、核糖核苷酸、2'-O-甲基核糖核苷酸或锁核酸。在另一方面,捕获寡核苷酸被修饰以抑制

通过核酸酶(例如硫代磷酸核苷酸)进行的消化。

[0362] 在一些实施例中,通过衔接子的连接或通过用加尾引物扩增,将衔接子序列添加到如图2(底部)所示形成的拷贝链中。衔接子可以添加到包含图2所示拷贝链的单链或双链分子中。

#### [0363] 条形码

[0364] 在一些实施例中,本公开利用条形码。检测单个分子通常需要分子条形码,诸如美国专利号7,393,665、8,168,385、8,481,292、8,685,678和8,722,368中所描述的。唯一分子条形码是短人工序列,其通常在体外操作的最初步骤中添加到患者样品中的每个分子上。所述条形码标记了分子及其子代。所述唯一分子条形码(UID)有多种用途。条形码允许跟踪样品中的每个单个核酸分子,以评估例如患者的血液中循环肿瘤DNA(ctDNA)分子的存在和数量,以便在不进行活检的情况下检测和监测癌症(Newman, A., 等人., (2014) An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage, Nature Medicine doi:10.1038/nm.3519)。

[0365] 条形码可以是在样品被混合(多重化)的情况下用于鉴定样品来源的多重样品ID(MID)。条形码也可以作为唯一的分子ID(UID),用于鉴定每个原始分子及其子代。条形码也可以是UID和MID的组合。在一些实施例中,将单个条形码用作UID和MID。在一些实施例中,每个条形码包括预定义序列。在其他实施例中,条形码包括随机序列。在本公开的一些实施例中,条形码的长度在约4-20个碱基之间,从而将96个与384个之间不同的衔接子添加到人类基因组样品中,每个衔接子具有不同的相同条形码对。普通技术人员会认识到条形码的数量取决于样品的复杂性(即,唯一靶标分子的预期数量),并且将能够为每个实验创建合适数量的条形码。

[0366] 唯一分子条形码也可用于分子计数和纠正测序错误。单个靶分子的整个子代都用相同的条形码标记,并形成条形码家族。不被带条形码家族的所有成员共享的序列变异被作为伪像丢弃而不是真突变。条形码还可用于位置去重(positional deduplication)和靶标量化,因为整个家族代表原始样品中的单个分子(Newman, A., 等人., (2016) Integrated digital error suppression for improved detection of circulating tumor DNA, Nature Biotechnology 34:547)。

[0367] 在一些实施例中,多个衔接子或含有条形码的引物中的UID数量可能超过多个核酸中的核酸数量。在一些实施例中,多个核酸中的核酸数量超过多个衔接子中的UID的数量。

#### [0368] 纯化

[0369] 在一些实施例中,本公开包括中间纯化步骤。例如,任何未使用寡核苷酸(诸如过量引物和过量衔接子)被去除,例如通过选自凝胶电泳、亲和层析和尺寸排阻层析的尺寸选择法。在一些实施例中,可以使用来自贝克曼库尔特(布雷亚市,加州)的固相可逆固定化(SPRI)来执行尺寸选择。在一些实施例中,捕获部分用于从未连接的核酸捕获并分离衔接子连接的核酸或从指数扩增产物捕获并分离过量的引物。在一些实施例中,使用形成包围待去除的寡核苷酸的封闭环状结构的特异性捕获核酸去除包括未使用的引物或衔接子的过量寡核苷酸,如2020年5月8日提交的美国申请序列号63/021875“Removal of excess oligonucleotides from a reaction mixture”中所述。

### [0370] 测序

[0371] 在一些实施例中,可以对基因融合序列的拷贝链、双链拷贝和包括基因融合序列的核酸文库或其扩增子进行核酸测序。可以根据本领域普通技术人员已知的任何方法进行测序。在一些实施例中,测序方法包括Sanger测序和染料终止测序,以及下一代测序技术诸如焦磷酸测序、纳米孔测序、基于微孔的测序、纳米球测序、MPSS、SOLiD、Illumina、Ion Torrent、Starlite、SMRT、tSMS、合成测序、连接测序、质谱测序、聚合酶测序、RNA聚合酶(RNAP)测序、基于显微镜的测序、微流控Sanger测序、基于显微镜的测序、RNAP测序、隧道电流DNA测序和体外病毒测序。参见W02014144478、W02015058093、W02014106076和W02013068528,它们各自通过引用全文并入本文。

[0372] 在一些实施例中,测序可以通过许多不同的方法进行,诸如通过采用合成测序技术。根据现有技术的合成测序被定义为任何测序方法,其监测在测序反应期间掺入特定脱氧核苷-三磷酸后副产物的产生(Hyman,1988,Anal.Biochem.174:423-436;Rhonaghi等人,1998,Science 281:363-365)。合成反应测序的一个突出实施例是焦磷酸测序方法。在这种情况下,核苷酸掺入过程中焦磷酸盐的产生由导致化学发光信号产生的酶促级联监测。454基因组测序系统(罗氏应用科学目录号04 760 085 001)是合成测序的实例,它基于焦磷酸测序技术。如产品文献中所述,对于在454 GS20或454 FLX仪器上进行测序,平均基因组DNA片段大小分别在200或600bp范围内。

[0373] 在一些实施例中,合成反应测序可以可替代地基于测序反应的终止染料类型。在这种情况下,掺入的染料脱氧核苷三磷酸(ddNTP)结构单元包含可检测标记,其优选为防止新生DNA链进一步延伸的荧光标记。然后在将ddNTP结构单元掺入模板/引物延伸杂交体中后,例如通过使用包含3'-5'核酸外切酶或校对活性的DNA聚合酶去除和检测标记。

[0374] 在一些实施例中,使用下一代测序方法例如Illumina,Inc.提供的方法("Illumina测序方法")进行测序。不希望受任何特定理论的束缚,Illumina下一代测序技术使用克隆扩增和合成测序(SBS)化学来实现快速、准确的测序。该过程同时鉴定DNA碱基,同时将它们掺入到核酸链中。每个碱基在添加到生长链时都会发出独特的荧光信号,用于确定DNA序列的顺序。

[0375] 在一些实施例中,测序方法是利用纳米孔的高通量单分子测序方法。在一些实施例中,如本文所述形成的核酸和核酸文库通过涉及穿过生物纳米孔(参见US10337060,其公开内容通过引用整体并入本文)或固态纳米孔(参见US10288599,US20180038001、US10364507,其公开内容通过引用整体并入本文)的方法进行测序。在其他实施例中,测序涉及将标签穿过纳米孔。(参见US8461854,其公开内容通过引用整体并入本文)或利用纳米孔的任何其他目前存在或未来的DNA测序技术。

[0376] 在另一些实施例中,通过其他合适的高通量单分子测序技术进行测序。包括依诺米那(Illumina)HiSeq平台(Illumina,San Diego,Cal.)、离子激流(Ion Torrent)平台(Life Technologies,Grand Island,NY)、利用单分子实时(SMRT)的太平洋生物科学(Pacific BioSciences)平台(Pacific Biosciences,Menlo Park,Cal.)或任何其他现有或未来DNA测序技术,该技术涉及或不涉及通过合成进行测序。

[0377] 测序步骤可利用平台特异性测序引物。可以将这些引物的结合位点引入扩增步骤中使用的扩增引物的5'-部分。如果条形码分子文库中不存在引物位点,则可以执行引入此

类结合位点的额外短扩增步骤。

[0378] 在一些实施例中,测序步骤涉及序列分析。在一些实施例中,该分析包括序列比对步骤。在一些实施例中,比对用于从多个序列(例如,具有相同条形码(UID)的多个序列)中确定共有序列。在一些实施例中,条形码(UID)用于从具有相同条形码(UID)的多个序列中确定共有序列。在其他实施例中,使用条形码(UID)来消除伪像,即,存在于一些但并非全部具有相同条形码(UID)的序列中的变异。源自PCR误差或测序误差的此类伪像可以被消除。

[0379] 在一些实施例中,通过定量样品中每个条形码(UID)的序列的相对数量,可以定量样品中的每个序列的数量。每个UID代表原始样品中的单个分子,且计数与每个序列变体相关的不同UID可以确定每个序列在原始样品中的比例。本领域技术人员将能够确定为确定共有序列所必需的序列读出的数量。在一些实施例中,为了准确的定量结果,每个UID(“序列深度”)都需要读取相关数量。在一些实施例中,期望的深度是每个UID 5-50次读取。

[0380] 在一些实施例中,测序步骤进一步包括通过共有序列确定进行错误校正的步骤。通过合成本文公开的有缺口的环状模板的环状链进行测序能够进行迭代或重复测序。通过对每个核苷酸或整个序列或序列的一部分建立共有序列判读,对相同核苷酸位置的多次读段能够纠正测序错误。从每个位置的确定共有碱基中获得核酸链的最终序列。在一些实施例中,核酸的共有序列从通过比较互补链的序列或通过比较互补链的共有序列获得。在一些实施例中,本公开在测序步骤之后包括序列读段比对的步骤和生成共有序列的步骤。在一些实施例中,共识是美国专利8535882中描述的简单多数共识。在另一些实施例中,共有序列由以下文献中描述的偏序对齐(POA)方法确定:Lee等人(2002)“Multiple sequence alignment using partial order graphs,”*Bioinformatics*,18(3):452-464以及Parker和Lee(2003)“Pairwise partial order alignment as a supergraph problem-aligning alignments revealed,”*J.Bioinformatics Computational Biol.*,11:1-18。基于用于确定共有序列的迭代读数的数量,该序列可以在很大程度上没有或基本上没有错误。

#### [0381] 无测序

[0382] 在一些实施例中,无需测序即可检测基因融合序列的拷贝链、双链拷贝和包括基因融合序列的核酸文库或其扩增子。检测可以通过扩增完成,包括通过终点聚合酶链式反应(PCR)、定量PCR(qPCR)或数字PCR(dPCR),包括数字微滴PCR(ddPCR)。在一些实施例中,基因融合的检测是定量的,诸如通过qPCR和dPCR实现的检测类型。在另一些实施例中,基因融合的检测是定性的,即,读出是通过凝胶电泳、毛细管电泳、质谱或检测具有特征大小或特征分子量的核酸的另一种方法的融合特异性扩增产物的存在或不存在。

#### [0383] 聚合酶链式反应

[0384] 在一些实施例中,根据本公开的基因融合特异性扩增通过包括数字微滴PCR(ddPCR)在内的数字PCR(dPCR)执行。

[0385] 数字PCR是一种定量扩增核酸的方法,描述在例如美国专利号9,347,095中,该专利的公开内容通过引用并入本文。该过程涉及将样品分成反应体积,使得每个体积包含一个或较少的靶核酸拷贝。在一些实施例中,分区的反应体积是水性液滴。

[0386] 在一些实施例中,分区中的靶标核酸是拷贝链。在另一些实施例中,分区中的靶标核酸是基因融合序列的双链拷贝。每个分区进一步包含扩增引物,即,能够支持靶标核酸的指数扩增的正向引物和反向引物。在一些实施例中,正向和反向引物能够与已知的融合序

列和第二寡核苷酸的5'-序列杂交(图1)。

[0387] 每个数字PCR反应体积进一步包含能够与正向引物和反向引物的扩增子杂交的可检测地标记的探针。在一些实施例中,探针能够与已知融合序列杂交。在一些实施例中,探针被设计为避免与野生型非融合基因序列结合。

[0388] 可检测地标记的探针可以用荧光团的组合进行标记,并且可以用具有5'-3'-外切核酸酶活性的核酸聚合酶进行指数扩增。

[0389] 在一些实施例中,本公开的方法包括用正向引物和反向引物进行扩增反应,其中该反应包括用探针检测扩增子的步骤,并确定已检测到探针的反应体积的数量,从而检测样品中基因融合的存在。

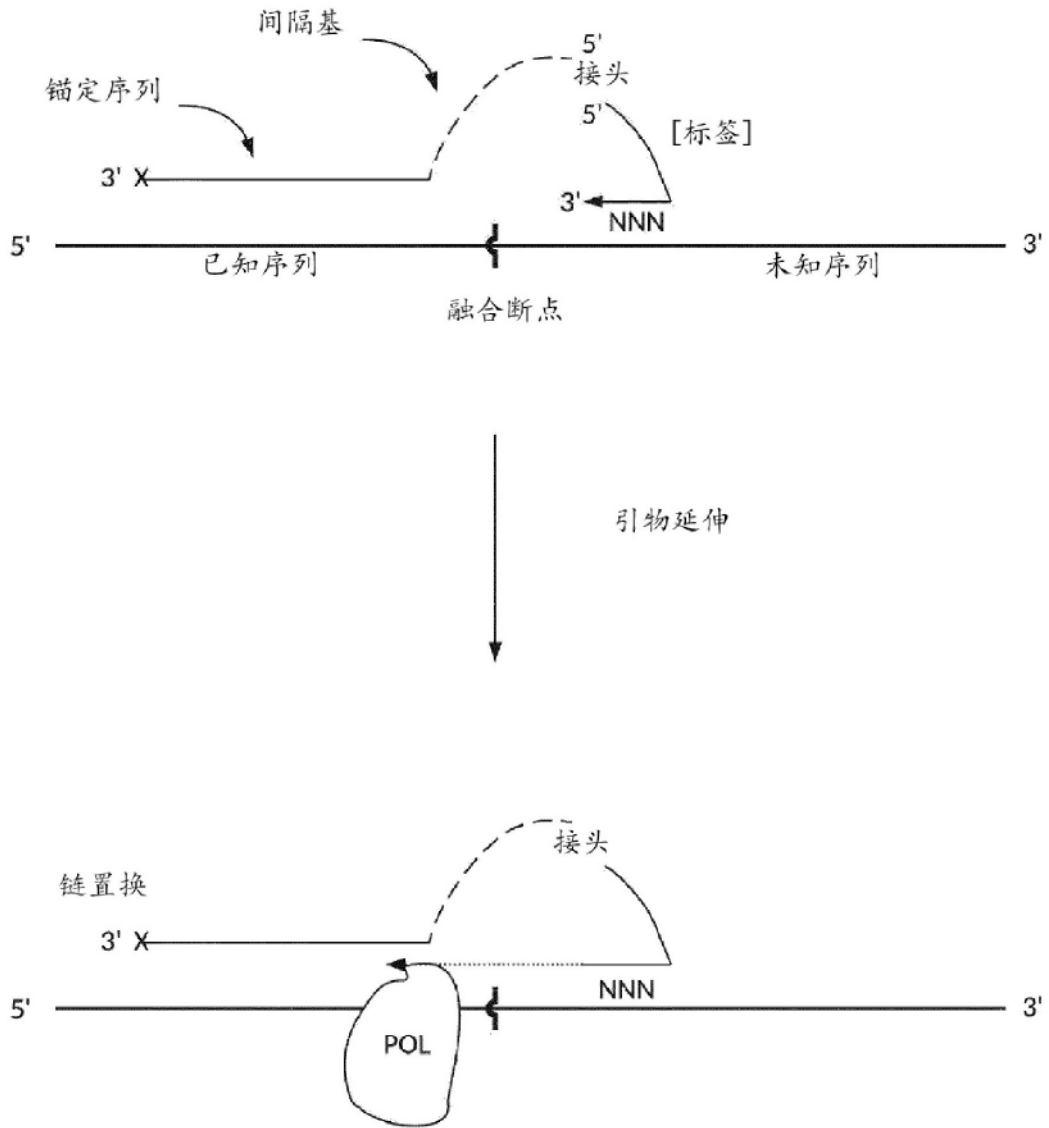


图 1

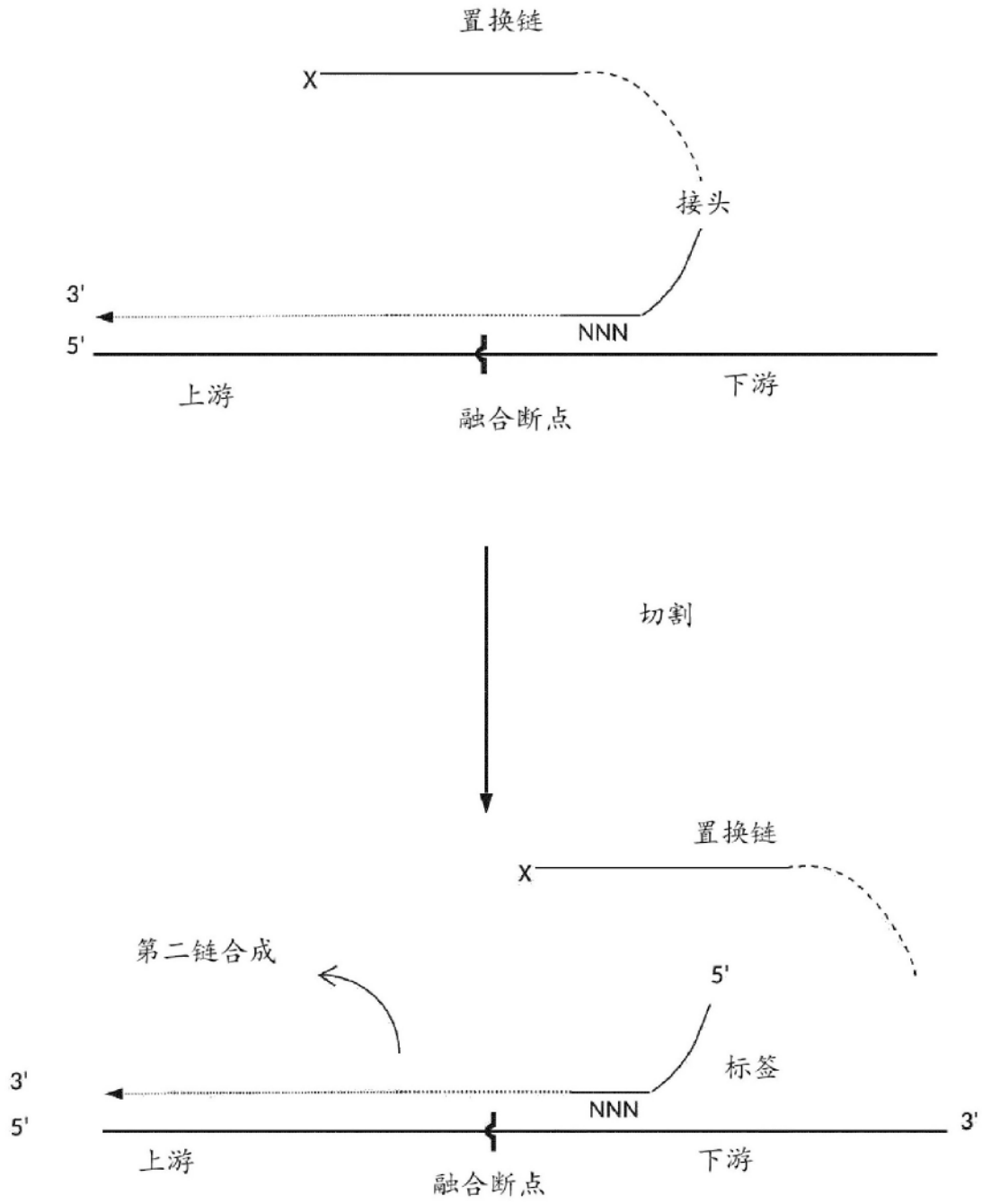


图 2