



(12) 发明专利

(10) 授权公告号 CN 108368545 B

(45) 授权公告日 2022.05.17

(21) 申请号 201680072543.1

(72) 发明人 翁莉 林盛榕 邓凌锋

(22) 申请日 2016.10.07

(74) 专利代理机构 北京安信方达知识产权代理有限公司 11262

(65) 同一申请的已公布的文献号

申请公布号 CN 108368545 A

专利代理人 贺淑东

(43) 申请公布日 2018.08.03

(51) Int.CI.

C12Q 1/6853 (2018.01)

(30) 优先权数据

C12Q 1/6858 (2018.01)

62/239,690 2015.10.09 US

G16B 25/00 (2019.01)

(85) PCT国际申请进入国家阶段日

G16B 25/20 (2019.01)

2018.06.11

(56) 对比文件

(86) PCT国际申请的申请数据

US 6143495 A, 2000.11.07

PCT/US2016/056126 2016.10.07

US 2010075384 A1, 2010.03.25

(87) PCT国际申请的公布数据

CN 103717752 A, 2014.04.09

W02017/062863 EN 2017.04.13

审查员 孙谦

(73) 专利权人 安可济控股有限公司

权利要求书2页 说明书58页 附图11页

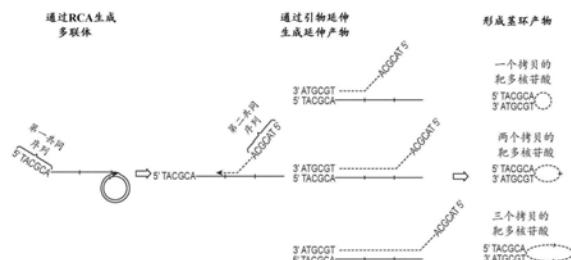
地址 开曼群岛大开曼岛

(54) 发明名称

用于富集扩增产物的方法及组合物

(57) 摘要

在一些方面,本公开内容提供了用于富集包含至少两个或更多个拷贝的靶多核苷酸的多联体的扩增子或扩增产物的方法。在一些实施方案中,方法包括对包含至少两个或更多个拷贝的靶多核苷酸的扩增子进行测序。在一些实施方案中,所述靶多核苷酸包含由染色体重排产生的序列,该染色体重排包括但不限于点突变、单核苷酸多态性、插入、缺失和包含融合基因的易位。在一些方面,本公开内容提供了在所述方法中有用的组合物和反应混合物。



1. 一种用于富集包含至少两个或更多个拷贝的靶多核苷酸的多联体的扩增子的方法，所述方法包括：

(a) 通过采用第一引物进行多个循环的扩增生成包含来自环状靶多核苷酸的单链多核苷酸的多联体，其中所述多个循环的扩增的每个循环包括变性温度下变性、在退火温度下的引物退火以及在延伸温度下的引物延伸；并且其中所述第一引物包含通过序列互补性与所述靶多核苷酸特异性杂交的第一3' 端和不通过序列互补性与所述靶多核苷酸特异性杂交的包含第一共同序列的第一5' 端；

(b) 通过延伸第二引物生成包含一个或多个拷贝的所述靶多核苷酸的多种延伸产物，所述第二引物包含通过序列互补性与所述多联体特异性杂交的第二3' 端和不通过序列互补性与所述多联体特异性杂交的包含第二共同序列的第二5' 端，其中所述第一共同序列和所述第二共同序列各自在5' 端包含至少10个连续核苷酸并且当最佳比对时至少90% 相同；以及

(c) 在生成多个扩增子的条件下扩增步骤(b) 的所述多种延伸产物，其中富集包含至少2个或更多个拷贝的所述靶多核苷酸的扩增子。

2. 如权利要求1所述的方法，其中所述第一共同序列和所述第二共同序列是相同的。

3. 如权利要求1所述的方法，其中在步骤(c) 的所述扩增中包括第三引物的引物延伸，其中所述第三引物包含通过序列互补性与所述第一共同序列或所述第二共同序列特异性杂交的序列。

4. 如权利要求1所述的方法，其中 (c) 的所述扩增步骤产生的具有两个或更多个拷贝的所述靶多核苷酸的扩增子的百分比大于具有少于两个拷贝的所述靶多核苷酸的扩增子的百分比。

5. 如权利要求4所述的方法，其中具有两个或更多个拷贝的所述靶多核苷酸的扩增子的百分比至少为90%。

6. 如权利要求4所述的方法，其中具有两个或更多个拷贝的所述靶多核苷酸的扩增子的百分比至少为80%。

7. 如权利要求4所述的方法，其中具有两个或更多个拷贝的所述靶多核苷酸的扩增子的百分比至少为60%。

8. 如权利要求3所述的方法，其中所述延伸产物形成茎环结构，所述茎环结构包含 (i) 所述第一共同序列与所述第二共同序列的互补体之间的分子内杂交，或 (ii) 所述第二共同序列与所述第一共同序列的互补体之间的分子内杂交。

9. 如权利要求8所述的方法，其中所述茎环结构的形成通过在退火步骤的温度保持在所述第三引物的解链温度±5°C 的情况下进行步骤(c) 的扩增而实现。

10. 如权利要求8所述的方法，其中所述茎环结构的形成通过在退火步骤保持在低于70°C 的温度的情况下进行步骤(c) 的扩增而实现。

11. 如权利要求8所述的方法，其中所述茎环结构包含至少9个碱基对的分子内杂交。

12. 如权利要求8所述的方法，其中所述茎环结构包含至少15个碱基对的分子内杂交。

13. 如权利要求8所述的方法，其中所述茎环结构包含至少20个碱基对的分子内杂交。

14. 如权利要求8所述的方法，其中所述茎环结构包含至少25个碱基对的分子内杂交。

15. 如权利要求8所述的方法，其中所述茎环结构包含至少30个碱基对的分子内杂交。

16. 如权利要求1所述的方法,其中步骤(b)包括不超过6个循环的所述第二引物的延伸。
17. 如权利要求1所述的方法,其中步骤(b)包括不超过8个循环的所述第二引物的延伸。
18. 如权利要求1所述的方法,其中步骤(b)包括不超过10个循环的所述第二引物的延伸。
19. 如权利要求3所述的方法,其中所述第一共同序列、所述第二共同序列和所述第三引物的杂交序列均具有在彼此±5°C内的解链温度(T_m)。
20. 如权利要求1所述的方法,其中所述环状靶多核苷酸是环化的无细胞DNA。
21. 如权利要求1所述的方法,其中所述环状靶多核苷酸是基因组DNA的环化片段。
22. 如权利要求1所述的方法,其中所述环状靶多核苷酸包含由染色体重排产生的序列。
23. 如权利要求22所述的方法,其中所述染色体重排是缺失、重复、倒位和易位中的至少一种。
24. 如权利要求1或22所述的方法,其中沿所述靶多核苷酸从5'至3'对应于(i)与所述第一3'端互补的序列、(ii)与所述第二3'端相同的序列以及(iii)(i)与(ii)之间的间插序列的所述靶多核苷酸的序列部分的组合长度为75个或更少的核苷酸。
25. 如权利要求1所述的方法,其中至少50%的多联体包含长度为至少75个核苷酸的靶多核苷酸。
26. 如权利要求1所述的方法,其中所述环状靶多核苷酸为单链。
27. 如权利要求1所述的方法,其进一步包括对步骤(c)中产生的所述多个扩增子进行测序。
28. 如权利要求27所述的方法,其中所述测序在没有相对于包含仅一个拷贝的所述靶多核苷酸的扩增子选择性地纯化包含两个或更多个拷贝的所述靶多核苷酸的扩增子的情况下进行。
29. 如权利要求1所述的方法,其进一步包括对步骤(c)中产生的所述多个扩增子中包含两个或更多个拷贝的所述靶多核苷酸的扩增子进行纯化。
30. 如权利要求29所述的方法,其进一步包括对经纯化的扩增子进行测序。
31. 如权利要求1所述的方法,其中多种不同的靶多核苷酸在同一反应混合物中扩增。

用于富集扩增产物的方法及组合物

[0001] 交叉引用

[0002] 本申请要求于2015年10月9日提供的美国临时申请号62/239,690的权益,该临时申请通过引用并入本文。

[0003] 发明背景

[0004] 大规模平行核酸测序的出现已使得对复杂群体内的序列变异的鉴定变得可行。一种利用具有链置换能力的聚合酶的扩增方法——滚环扩增(RCA),已作为对用于制备核酸以供测序分析的聚合酶链反应(PCR)程序的有用替代和补充而出现。RCA涉及通过将核苷酸连续添加至与环状多核苷酸模板(如环状DNA模板)退火的引物来生长具有重复序列的多核苷酸。该延伸过程可多次覆盖环状多核苷酸模板的全长,从而导致形成该模板的重复序列或者所谓的多联体。多联体也可作为模板来生成进一步的扩增产物。然而,该延伸只会进行直到达到线性多联体的末端。随着生长多核苷酸链的前端遇到DNA的双链部分,该生长链替代模板的现有链。结果通常是形成各种长度的由可变数目的模板序列重复组成的双链DNA。在常规的RCA方法中,与包含靶序列的多个重复的较长多联体相比,较短多联体经常不成比例地扩增。因此,对较长多联体的后续分析可能会更加困难。

[0005] 由于常用技术的固有误差频率比群体内许多实际序列变异的频率更大,大规模平行测序具有显著的局限性。例如,在标准的高通量测序中已报道了0.1-1%的误差率。当变体频率低,如等于或低于该误差率时,对罕见序列变体的检测具有高的假阳性率。

[0006] 由于多种原因,检测罕见序列变体的能力非常关键。例如,检测罕见特征性序列可用于鉴定和区分有害环境污染物如细菌分类群的存在。表征细菌分类群的一种常见方法是鉴定高度保守序列如rRNA序列的差异。然而,迄今为止典型的基于测序的方法面临与给定样品中不同基因组的绝对数量和成员之间的同源性程度相关的挑战,从而为本已繁琐的程序呈现出复杂的问题。

[0007] 用于检测序列变异的现有技术在检测融合基因变异和染色体重排方面特别低效。通常,与重排基因融合的‘配偶体’基因是未知的,这使得检测更具挑战性。如果没有观察到接合点,融合基因也可能难以检测。

发明内容

[0008] 鉴于上述情况,需要检测罕见序列变异,特别是罕见序列变化和基因融合事件的替代并且/或者稳健的方法和组合物。本公开内容的组合物和方法满足了该需求,并且也提供了另外的益处。特别是,本公开内容的各个方面提供了可用于大规模平行测序方法的包含多个拷贝的靶多核苷酸的扩增子。使用包含多个拷贝的靶多核苷酸的扩增子,可不止一次对靶多核苷酸进行测序,从而减少对罕见序列变体和融合基因进行测序时的错误。

[0009] 在一方面,公开了一种用于富集包含至少两个或更多个拷贝的靶多核苷酸的多联体的扩增子的方法。该方法包括:(a)通过延伸第一引物生成包含来自环状靶多核苷酸的单链多核苷酸的多联体,该第一引物包含通过序列互补性与靶多核苷酸特异性杂交的第一3'端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5'端;(b)通

过延伸第二引物生成包含一个或多个拷贝的靶多核苷酸的多种延伸产物,该第二引物包含通过序列互补性与多联体特异性杂交的第二3'端和不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5'端,其中第一共同序列和第二共同序列各自在5'端包含至少10个连续核苷酸并且当最佳比对时至少90%相同;以及(c)在生成多个扩增子的条件下对步骤(b)的多种延伸产物进行扩增,其中富集包含至少2个或更多个拷贝的靶多核苷酸的扩增子。在一些实施方案中,步骤(a)通过具有链置换活性的聚合酶而实现。在一些实施方案中,所述第一共同序列和所述第二共同序列是相同的。在一些实施方案中,步骤(c)的扩增包括第三引物的引物延伸,其中该第三引物包含通过序列互补性与第一共同序列或第二共同序列特异性杂交的序列。在一些实施方案中,(c)的扩增步骤产生的具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比大于具有少于两个拷贝的靶多核苷酸的扩增子的百分比。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为90%。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为80%。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为60%。在一些实施方案中,所述延伸产物形成茎环结构,该茎环结构包含(i)第一共同序列与第二共同序列的互补体之间的分子内杂交,或(ii)第二共同序列与第一共同序列的互补体之间的分子内杂交。在一些实施方案中,茎环产物的形成通过在退火步骤的温度保持在第三引物的解链温度±5°C的情况下进行步骤(c)的扩增而实现。在一些实施方案中,茎环产物的形成通过在退火步骤保持在低于70°C的温度的情况下进行步骤(c)的扩增而实现。在一些实施方案中,该茎环结构包含至少9个碱基对的分子内杂交。在一些实施方案中,该茎环结构包含至少15个碱基对的分子内杂交。在一些实施方案中,该茎环结构包含至少20个碱基对的分子内杂交。在一些实施方案中,该茎环结构包含至少25个碱基对的分子内杂交。在一些实施方案中,该茎环结构包含至少30个碱基对的分子内杂交。在一些实施方案中,步骤(b)包括不超过6个循环的第二引物延伸。在一些实施方案中,步骤(b)包括不超过8个循环的第二引物延伸。在一些实施方案中,步骤(b)包括不超过10个循环的第二引物延伸。在一些实施方案中,所述第一共同序列、所述第二共同序列和所述第三引物的杂交序列均具有在彼此±5°C内的解链温度(Tm)。在一些实施方案中,该环状靶多核苷酸为环化的无细胞DNA。在一些实施方案中,该环状靶多核苷酸为基因组DNA的环化片段。在一些实施方案中,该环状靶多核苷酸包含由染色体重排产生的序列。在一些实施方案中,染色体重排是缺失、重复、倒位和易位中的至少一种。在一些实施方案中,沿靶多核苷酸从5'至3'对应于(i)与第一3'端互补的序列、(ii)与第二3'端相同的序列以及(iii)(i)与(ii)之间的间插序列的所述靶多核苷酸的序列部分的组合长度为75个或更少的核苷酸。在一些实施方案中,至少50%的多联体包含长度为至少75个核苷酸的靶多核苷酸。在一些实施方案中,环状靶多核苷酸为单链。在一些实施方案中,该方法进一步包括对步骤(c)中产生的多个扩增子进行测序。在一些实施方案中,所述测序在没有相对于包含仅一个拷贝的靶多核苷酸的扩增子选择性地纯化包含两个或更多个拷贝的靶多核苷酸的扩增子的情况下进行。在一些实施方案中,该方法进一步包括对步骤(c)中产生的多个扩增子中包含两个或更多个拷贝的靶多核苷酸的扩增子进行纯化。在一些实施方案中,该方法进一步包括对经纯化的扩增子进行测序。在一些实施方案中,多种不同的靶多核苷酸在同一反应混合物中扩增。

[0010] 在另一方面,公开了一种用于富集包含至少两个或更多个拷贝的靶多核苷酸的多

联体的扩增子的反应混合物。在一个实施方案中,该反应混合物包含:(a)环状靶多核苷酸,(b)第一引物,其包含通过序列互补性与靶多核苷酸特异性杂交的第一3'端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5'端;以及(c)第二引物,其包含通过序列互补性与多联体特异性杂交的第二3'端和不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5'端,其中第一共同序列和第二共同序列各自在5'端包含至少10个连续核苷酸并且当最佳比对时至少90%相同,并且该多联体为第一引物的延伸产物。在一些实施方案中,所述第一共同序列和所述第二共同序列是相同的。在一些实施方案中,该反应混合物容纳在容器中。在一些实施方案中,该容器为孔、板、管、腔室、流动池或芯片。在一些实施方案中,该反应混合物进一步包含具有通过序列互补性与第一共同序列或第二共同序列特异性杂交的序列的第三引物。在一些实施方案中,所述第一共同序列、所述第二共同序列和所述第三引物的杂交序列均具有在彼此±5℃内的解链温度(T_m)。在一些实施方案中,所述第一共同序列和所述第二共同序列各自包含至少15个核苷酸。在一些实施方案中,该环状靶多核苷酸为环化的无细胞DNA。在一些实施方案中,该环状靶多核苷酸为基因组DNA的环化片段。在一些实施方案中,该环状靶多核苷酸包含由染色体重排产生的序列。在一些实施方案中,该染色体重排是缺失、重复、倒位和易位中的至少一种。在一些实施方案中,沿靶多核苷酸从5'至3'对应于(i)与第一3'端互补的序列、(ii)与第二3'端相同的序列以及(iii)(i)与(ii)之间的间插序列的所述靶多核苷酸的序列部分的组合长度为75或更少的个核苷酸。在一些实施方案中,该环状靶多核苷酸为单链。

[0011] 在另一方面,公开了一种用于富集包含至少两个或更多个拷贝的靶多核苷酸的多联体的扩增子的试剂盒。在一个实施方案中,该试剂盒包含:(a)第一引物,其包含通过序列互补性与靶多核苷酸特异性杂交的第一3'端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5'端;(b)第二引物,其包含通过序列互补性与多联体特异性杂交的第二3'端和不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5'端,其中第一共同序列和第二共同序列各自在5'端包含至少10个连续核苷酸并且当最佳比对时至少90%相同,并且该多联体为第一引物的延伸产物;以及(c)第三引物,其具有通过序列互补性与第一共同序列或第二共同序列特异性杂交的序列。在一些实施方案中,所述第一共同序列和所述第二共同序列是相同的。在一些实施方案中,所述第一共同序列、所述第二共同序列和所述第三引物的杂交序列均具有在彼此±5℃内的解链温度(T_m)。在一些实施方案中,沿靶多核苷酸从5'至3'对应于(i)与第一3'端互补的序列、(ii)与第二3'端相同的序列以及(iii)(i)与(ii)之间的间插序列的所述靶多核苷酸的序列部分的组合长度为75个或更少的核苷酸。

[0012] 在另一方面,公开了一种用于设计用于富集包含至少两个或更多个拷贝的靶多核苷酸的多联体的扩增子的引物的系统。在一个实施方案中,该系统包含:(a)计算机,其被配置为接收客户请求以设计用于扩增指定的靶序列的引物;(b)包含代码的计算机可读介质,该代码在由一个或多个处理器执行时设计至少三种用于扩增靶序列的引物,其中所述至少三种引物包含:(i)第一引物,其包含通过序列互补性与靶多核苷酸特异性杂交的第一3'端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5'端;(ii)第二引物,其包含通过序列互补性与多联体特异性杂交的第二3'端和不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5'端,其中所述第一共同序列和所述第二共同序

列各自在5' 端包含至少10个连续核苷酸并且当最佳比对时至少90%相同，并且该多联体为第一引物的延伸产物；以及(iii)第三引物，其具有通过序列互补性与第一共同序列或第二共同序列特异性杂交的序列；以及(c)向接收者发送报告的报告生成器，其中该报告包含所述至少三种引物的序列。在一些实施方案中，所述第一共同序列和所述第二共同序列是相同的。在一些实施方案中，所述第一共同序列、所述第二共同序列和所述第三引物的杂交序列均具有在彼此±5°C内的解链温度(Tm)。在一些实施方案中，沿靶多核苷酸从5' 至3' 对应于(i)与第一3' 端互补的序列、(ii)与第二3' 端相同的序列以及(iii) (i)与(ii)之间的间插序列的所述靶多核苷酸的序列部分的组合长度为75个或更少的核苷酸。

[0013] 在一方面，本公开内容提供了一种进行滚环扩增的方法。该方法包括(a)提供包含靶多核苷酸的环状多核苷酸；(b)使扩增反应混合物经受多个循环的滚环扩增以生成包含多联体的多种扩增产物，其中该扩增反应混合物包含(i)具有链置换活性的聚合酶，(ii)环状多核苷酸以及(iii)引物；且其中多个循环的滚环扩增中的每个循环均包含在变性温度下的变性、在退火温度下的引物退火以及在延伸温度下持续给定延伸时间段的引物延伸，从而生成多种扩增产物；且其中所生成的多种扩增产物的特征在于与通过利用变性和引物退火条件相当但延伸时间段相当于所述多个循环的延伸时间段之和的一个扩增循环生成的多种扩增产物相比，其含有更高比例的具有至少两个拷贝的靶多核苷酸的多联体。

[0014] 在一方面，本公开内容提供了一种提高由滚环扩增生成的具有至少两个拷贝的靶多核苷酸的多联体的比例的方法。该方法包括(a)提供包含靶多核苷酸的环状多核苷酸；(b)使扩增反应混合物经受多个循环的滚环扩增以生成包含多联体的多种扩增产物，其中该扩增反应混合物包含(i)具有链置换活性的聚合酶，(ii)环状多核苷酸，以及(iii)引物；且其中多个循环的滚环扩增中的每个循环均包含在变性温度下的变性、在退火温度下的引物退火以及在延伸温度下持续给定延伸时间段的引物延伸，从而生成多种扩增产物；从而提高具有至少两个拷贝的靶多核苷酸的多联体的比例。在一些实施方案中，与通过利用变性和引物退火条件相当但延伸时间段相当于所述多个循环的延伸时间段之和的一个扩增循环生成的多种扩增产物相比，所述多种扩增产物中具有至少两个拷贝的靶多核苷酸的多联体的比例增加。

[0015] 在一些实施方案中，所述聚合酶选自：BsU DNA聚合酶、Vent聚合酶、Bst DNA聚合酶、phi29DNA聚合酶、PyroPhage 3173聚合酶、其任何变体及其任何片段。

[0016] 在一些实施方案中，当反应混合物中使用的环状多核苷酸包含人无细胞DNA(cfDNA)时，所述多种扩增产物展现出约180个碱基对的平均片段长度。在一些实施方案中，当反应混合物中使用的环状多核苷酸包含人无细胞DNA(cfDNA)时，所述多种扩增产物展现出约170个碱基对的中值片段长度。在一些实施方案中，当反应混合物中使用的环状多核苷酸包含人无细胞DNA(cfDNA)时，所述多种扩增产物展现出约40个碱基至约450个碱基的片段长度分布。在一些实施方案中，当反应混合物中使用的环状多核苷酸包含人无细胞DNA(cfDNA)时，所述多种扩增产物展现出约100个碱基至约200个碱基的片段长度分布。

[0017] 在一些实施方案中，具有至少两个拷贝的靶多核苷酸的多联体的比例增加至少约1%。在一些实施方案中，所述方法进一步包括在多个循环的滚环扩增中的至少一个循环之后用聚合酶补充反应混合物。

[0018] 在一些实施方案中，所述环状多核苷酸具有约40个碱基至约500个碱基的长度。在

一些实施方案中,该环状多核苷酸包含无细胞DNA (cfDNA)。在一些实施方案中,该环状多核苷酸包含基因组DNA的片段。在一些实施方案中,该环状多核苷酸包含由染色体重排产生的序列。在一些实施方案中,该染色体重排是缺失、重复、倒位和易位中的至少一种。在一些实施方案中,该环状多核苷酸为双链。在一些实施方案中,该环状多核苷酸为单链。

[0019] 在一些实施方案中,多种不同的环状多核苷酸在扩增反应混合物中扩增。在一些实施方案中,所述多个循环包括至少两个循环。在一些实施方案中,所述多个循环中的每个循环均包含 (i) 在约75°C至约95°C的变性温度下持续约5秒至约60秒的变性, (ii) 在约45°C至约65°C的退火温度下持续约5秒至约60秒的引物退火,以及 (iii) 在约65°C至约75°C的延伸温度下持续约30秒至约10分钟的延伸时间段的引物延伸。在一些实施方案中,所述多个循环中的每个循环均包含 (i) 在约80°C的变性温度下持续约15秒至约30秒的变性, (ii) 在约50°C的退火温度下持续约15秒至约45秒的引物退火,以及 (iii) 在约70°C的延伸温度下续约3分钟至约10分钟的延伸时间段的引物延伸。

[0020] 在一些实施方案中,所述引物包含能够与环状多核苷酸的不同区域随机杂交并引发该不同区域以供引物延伸的随机序列。在一些实施方案中,该引物包含能够以序列特异性方式与环状多核苷酸的区域杂交并引发该区域以供引物延伸的基因特异性序列。在一些实施方案中,该引物包括第一引物,该第一引物包含 (i) 通过序列互补性与环状多核苷酸特异性杂交的第一3' 端,和 (ii) 不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5' 端,其中在多个循环的滚环扩增期间,通过使用环状多核苷酸作为模板延伸所述第一引物来生成包含单链多核苷酸的多联体。在一些实施方案中,该引物包括第二引物,该第二引物包含 (i) 通过序列互补性与包含单链多核苷酸的多联体特异性杂交的第二3' 端,和 (ii) 不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5' 端,其中在多个循环的滚环扩增期间,通过使用该多联体作为模板延伸所述第二引物来生成包含一个或多个拷贝的靶多核苷酸的多种延伸产物。在一些实施方案中,所述第一共同序列和所述第二共同序列各自在5' 端包含至少10个连续核苷酸,并且当最佳比对时至少90% 相同。在一些实施方案中,所述第一共同序列和所述第二共同序列是相同的。

[0021] 在一些实施方案中,所述方法进一步包括在生成多个扩增子的条件下对多种延伸产物进行扩增,其中富集包含至少2个或更多个拷贝的靶多核苷酸的扩增子。在一些实施方案中,扩增包括第三引物的引物延伸,其中该第三引物包含通过序列互补性与第一共同序列或第二共同序列特异性杂交的序列。在一些实施方案中,扩增产生的具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比大于具有少于两个拷贝的靶多核苷酸的扩增子的百分比。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为5%。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为10%。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为20%。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为30%。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为40%。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为60%。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为80%。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为90%。

[0022] 在一些实施方案中,所述多种延伸产物形成茎环结构,该茎环结构包含(i)第一共同序列与第二共同序列的互补体之间的分子内杂交,或(ii)第二共同序列与第一共同序列的互补体之间的分子内杂交。在一些实施方案中,该茎环结构的形成通过在退火步骤的温度保持在第三引物的解链温度±5°C的情况下进行扩增而实现。在一些实施方案中,该茎环结构的形成通过在退火步骤保持在低于约70°C的温度下的情况进行扩增而实现。在一些实施方案中,该茎环结构包含至少9个碱基对的分子内杂交。在一些实施方案中,该茎环结构包含至少15个碱基对的分子内杂交。在一些实施方案中,该茎环结构包含至少20个碱基对的分子内杂交。在一些实施方案中,该茎环结构包含至少25个碱基对的分子内杂交。在一些实施方案中,该茎环结构包含至少30个碱基对的分子内杂交。在一些实施方案中,所述第一共同序列、所述第二共同序列和所述第三引物的杂交序列均具有在彼此±5°C内的解链温度(T_m)。

[0023] 在一些实施方案中,沿靶多核苷酸从5'至3'对应于(i)与第一3'端互补的序列、(ii)与第二3'端相同的序列以及(iii)(i)与(ii)之间的间插序列的所述靶多核苷酸的序列部分的组合长度为75个或更少的核苷酸。

[0024] 在一些实施方案中,所述方法进一步包括对包含多联体的多种扩增产物进行测序。在一些实施方案中,所述测序在没有选择性地将具有至少两个拷贝的靶多核苷酸的多联体与包含少于两个拷贝的靶多核苷酸的多联体相分离的情况下进行。

[0025] 在一些实施方案中,所述方法进一步包括将包含至少两个拷贝的靶多核苷酸的多联体与包含少于两个拷贝的靶多核苷酸的多联体相分离。在一些实施方案中,该方法进一步包括对包含至少两个拷贝的靶多核苷酸的多联体进行测序。

[0026] 援引并入

[0027] 本说明书中提及的所有出版物、专利和专利申请均通过引用并入本文,其程度如同特别地和单独地说明每个单独的出版物、专利或专利申请通过引用而并入。

附图说明

[0028] 本发明的新颖特征在所附的权利要求书中进行了具体阐述。通过参考以下对利用了本发明原理的说明性实施方案加以阐述的详细描述及其附图(图),将会获得对本发明的特征和优点的更好的了解,在附图中:

[0029] 图1图示了根据一个实施方案的茎环产物的形成。

[0030] 图2图示了根据一个实施方案的背靠背(B2B)引物设计,其中正向和反向引物被设计为具有用于RCA扩增的相邻5'端。

[0031] 图3图示了根据一个实施方案的使用背靠背引物构建测序文库的方法。

[0032] 图4图示了根据一个实施方案的用于富集扩增子的方法。

[0033] 图5示出了通过一个循环的滚环扩增和多个循环的滚环扩增生成的扩增产物的大小分布,由琼脂糖凝胶获得。

[0034] 图6呈现的表格示出了具有不同数目重复的扩增产物之间的比率。

[0035] 图7图示了包含一个循环和多个循环的示例性滚环扩增反应中各个重复元件的大小分布。

[0036] 图8示出了通过一个循环的滚环扩增和多个循环的滚环扩增生成的测序靶标的片

段大小分布。

[0037] 图9呈现的表格图示了根据一个实施方案,HD664、ELM4/ALK融合DNA样品与野生型参考DNA样品以不同比例的混合。

[0038] 图10图示了根据一个实施方案的融合等位基因检测。

[0039] 图11图示了根据一个实施方案在多重反应中可检测到多种靶多核苷酸序列。

具体实施方式

[0040] 除非另有说明,否则本文公开的一些方法的实施采用了本领域技术范围内的免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组学和重组DNA的常规技术。参见,例如,Sambrook和Green,Molecular Cloning:A Laboratory Manual,第四版(2012);Current Protocols in Molecular Biology系列(F.M.Ausubel等编著);Methods In Enzymology系列(Academic Press, Inc.),PCR 2:A Practical Approach(M.J.MacPherson, B.D.Hames和G.R.Taylor编著(1995));Harlow和Lane编著(1988)Antibodies,A Laboratory Manual, and Culture of Animal Cells:A Manual of Basic Technique and Specialized Applications,第6版(R.I.Freshney编著(2010))。

[0041] 术语“约”或“大约”意指如本领域普通技术人员所确定的特定值在可接受的误差范围内,其部分地取决于该值是如何测量或确定的,即,测量系统的限制。例如,“约”可以是指根据本领域的实践,在1个或大于1个标准差内。或者,“约”可以是指给定值的高达20%、高达10%、高达5%或高达1%的范围。或者,尤其是对于生物系统或过程,该术语可以是指在数值的一个数量级内,优选地在5倍以内,更优选地在2倍以内。在本申请和权利要求书中描述特定值时,除另有说明外,否则术语“约”应该被认为是指在特定值的可接受的误差范围内。

[0042] 术语“多核苷酸”、“核酸”和“寡核苷酸”是可以互换使用的。它们是指任意长度的核苷酸(脱氧核糖核苷酸或核糖核苷酸)或其类似物的聚合形式。多核苷酸可以具有任意三维结构,并且可以行使任何已知的或未知的功能。以下是多核苷酸的非限制性实例:基因或基因片段的编码或非编码区,通过连锁分析确定的基因座(座位)、外显子、内含子、信使RNA(mRNA)、转运RNA(tRNA)、核糖体RNA(rRNA)、短干扰RNA(siRNA)、短发夹RNA(shRNA)、微小RNA(miRNA)、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任意序列的分离的DNA、任意序列的分离的RNA、核酸探针和引物。多核苷酸可以包含一个或多个修饰的核苷酸,例如甲基化核苷酸和核苷酸类似物。如果存在,对核苷酸结构的修饰可以在聚合物组装之前或之后赋予。核苷酸序列可以被非核苷酸组分打断。多核苷酸可以在聚合后被进一步修饰,例如,通过与标记组分偶联。

[0043] 术语“靶多核苷酸”是指具有靶序列的核酸分子起始群体中的核酸分子或多核苷酸,该靶序列的存在、量和/或核苷酸序列或其中一个或多个的改变希望加以确定。靶多核苷酸可以是较大多核苷酸的一部分(例如待扩增、待测序或有待以其他方式分析的部分),或可用于指包含靶序列的较大多核苷酸。通常,术语“靶序列”是指在核酸单链上的核酸序列。靶序列可以是基因的一部分,调节序列,基因组DNA,cDNA,融合基因,包括mRNA、miRNA、rRNA在内的RNA,等等。靶序列可以是来自样品的靶序列或次级靶标如扩增反应的产物。

[0044] 通常,术语“序列变体”是指序列中相对于一个或多个参考序列的任何变异。一般

而言,对于参考序列已知的个体的给定群体,序列变体以比参考序列更低的频率发生。在一些情况下,参考序列为单个已知参考序列,如单个个体的基因组序列。在一些情况下,参考序列是通过对多个已知序列,如作为参考群体的多个个体的基因组序列,或来自相同个体的多核苷酸的多个测序读取进行比对而形成的共有序列。在一些情况下,序列变体在群体中以低频率发生(也称为“罕见”序列变体)。例如,序列变体可以以约为或低于约5%、4%、3%、2%、1.5%、1%、0.75%、0.5%、0.25%、0.1%、0.075%、0.05%、0.04%、0.03%、0.02%、0.01%、0.005%、0.001%或更低的频率发生。在一些情况下,序列变体以约为或低于约0.1%的频率发生。序列变体可以是相对于参考序列的任何变异。序列变异可以由一个核苷酸或多个核苷酸(例如2、3、4、5、6、7、8、9、10个或更多个核苷酸)的变化、插入或缺失组成。当序列变体包含两个或更多个核苷酸差异时,不同的核苷酸可以是彼此相邻近的,或不连续的。序列变体类型的非限制性实例包括单核苷酸多态性(SNP)、缺失/插入多态性(DIP)、拷贝数变体(CNV)、短串联重复(STR)、简单序列重复(SSR)、可变数目串联重复(VNTR)、扩增片段长度多态性(AFLP)、基于反转录转座子的插入多态性、序列特异性扩增多态性和可检测为序列变体的表观遗传标记的差异(例如,甲基化差异)。在一些实施方案中,序列变体可以指染色体重排,包括但不限于易位或融合基因。

[0045] 如本文所用的,术语“多联体”通常指包含连续多核苷酸的连接产物或扩增产物,该连续多核苷酸含有多个拷贝的靶多核苷酸序列(例如,至少2、3、4、5、6、7、8、9或10个拷贝的靶序列;在一些情况下,至少2个拷贝)。在一些情况下,多联体包含串联连接的多个拷贝的靶多核苷酸序列。在一些情况下,附加的多核苷酸序列散布在多个拷贝的靶多核苷酸序列之间。

[0046] 如本文所用的,术语“杂交”和“退火”通常是指这样的反应,在该反应中,一个或多个多核苷酸发生反应以形成复合体,该复合体通过在核苷酸残基的碱基之间的氢键键合而得到稳定。该氢键键合可以通过Watson Crick碱基配对、Hoogstein结合或以任意其他序列特异性方式而发生。该复合体可包含形成双链体结构的两条链、形成多链复合体的三条或更多条链、自杂交的单链或其任意组合。杂交反应可以构成更广泛的过程中的步骤,例如PCR的起始,或核酶对多核苷酸的酶切。可以通过与第二序列的核苷酸残基的碱基发生氢键键合而稳定化的第一序列被称为可与第二序列“杂交”。在这种情况下,第二序列也可称为可与第一序列杂交。

[0047] 如本文所用的,术语“互补体”、“互补”和“互补性”通常是指与给定序列完全互补且可杂交的序列。在一些情况下,如果给定区域上的碱基序列能够与其结合配偶体的碱基序列互补地结合,使得例如形成A-T、A-U、G-C和G-U碱基对,则与该给定核酸杂交的序列被称为给定分子的“互补体”或“反向互补体”。通常,可与第二序列杂交的第一序列与该第二序列特异性或选择性地杂交,使得在杂交反应期间,与第二序列或第二序列组的杂交相对于与非靶序列的杂交是优选的(例如,在给定的一组条件,如本领域常用的严格条件下更加热力学稳定)。一般而言,可杂交序列在其各自的全长或部分长度上具有一定程度的序列互补性,如25%-100%的互补性,包括至少25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%和100%的序列互补性。序列同一性,例如为了评估互补性百分比,可以通过任何合适的比对算法进行测量,包括但不限于Needleman-Wunsch算法(参见,例如,EMBOSS Needle比对器,

可从 www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html 获得, 任选地具有默认设置)、BLAST 算法(参见, 例如, BLAST 比对工具, 可从 blast.ncbi.nlm.nih.gov/Blast.cgi 获得, 任选地具有默认设置) 或者 Smith-Waterman 算法(参见, 例如, EMBOSS Water 比对器, 可从 www.ebi.ac.uk/Tools/psa/emboss_water/nucleotide.html 获得, 任选地具有默认设置)。最优算法可以使用所选定算法的任意合适的参数(包括默认参数)进行评估。

[0048] 如本文所用的, 术语“延伸产物”通常指这种反应的产物, 其中在该反应中, 核苷酸引物通过核苷酸的共价添加而延伸。在一些情况下, 核苷酸掺入可以由模板引导。在一些情况下, 核苷酸掺入可以在没有模板的情况下进行。在一些情况下, 延伸产物是诸如来自 PCR 扩增、滚环扩增 (RCA) 或等温扩增的扩增产物。

[0049] 如本文所用的, 术语“扩增”通常是指由靶多核苷酸或其部分形成一个或多个拷贝的任意过程。多种扩增多核苷酸(例如DNA和/或RNA)的方法是可用的, 本文描述了这些方法的一些实例。扩增可以是线性的, 指数式的, 或在多阶段扩增过程中涉及线性和指数阶段。扩增方法可包括温度的改变, 例如热变性步骤, 或者可以是不需要热变性的等温过程。

[0050] 如本文所用的, 术语“茎环产物”和“茎环结构”通常指其中在部分多核苷酸之间发生分子内杂交的多核苷酸二级结构。当单个多核苷酸链的两个区域杂交形成可被称为“茎”的双链部分和可被称为“环”的未配对单链环时, 可形成茎环。茎可具有任意可变长度的碱基对, 并且沿着茎的碱基配对可以被参与该茎的一个或两个部分上的一个或多个未配对碱基的间隙内部中断。环可以具有任意可变长度的未配对碱基。在一些情况下, 环的长度为至少3个碱基。在一些情况下, 形成“茎”的两个区域完全互补。在一些情况下, 形成“茎”的两个区域部分互补。在一些情况下, 单个多核苷酸可包含一个茎环结构。在一些情况下, 单个多核苷酸可包含多于一个茎环结构。茎环结构的茎部分可以以没有突出端的双链部分、包含5'突出端的单链部分、包含3'突出端的单链部分, 或从5'端和3'端两者延伸的单链部分终止。

[0051] 本公开内容提供了可用于生成包含两个或更多个拷贝的靶多核苷酸的扩增子的方法和组合物。在一些实施方案中, 该方法对检测罕见序列变体和融合基因是有用的。在一些实施方案中, 基因融合在没有预先知晓配偶体基因的情况下被检测到, 并且可用于诸如在无细胞DNA或基因组DNA样品中筛选基因重排事件。本公开内容的各个方面提供了可以与大规模平行测序方法一起使用的包含两个或更多个拷贝的靶多核苷酸的扩增子。

[0052] 在一方面, 本公开内容提供了提高由滚环扩增生成的具有至少两个拷贝的靶多核苷酸的多联体的比例的方法。该方法包括 (a) 提供包含靶多核苷酸的环状多核苷酸, (b) 使扩增反应混合物经受多个循环的滚环扩增以生成包含多联体的多种扩增产物。该反应混合物可包含 (i) 具有链置换活性的聚合酶, (ii) 包含靶多核苷酸的环状多核苷酸, 以及 (iii) 引物。多个循环的滚环扩增中的每个循环可包含在变性温度下的变性、在退火温度下的引物退火和在延伸温度下持续给定延伸时间段的引物延伸。多个循环的滚环扩增可产生比例增加的具有至少两个拷贝的靶多核苷酸的多联体的多种扩增产物。所生成的多种扩增产物的特征可以在于与通过利用变性和引物退火条件相当但延伸时间段相当于所述多个循环的延伸时间段之和的一个扩增循环生成的多种扩增产物相比, 其含有更高比例的具有至少两个拷贝的靶多核苷酸的多联体。

[0053] 具有链置换活性的聚合酶(例如, 具有链置换活性的DNA聚合酶)可以促进滚环扩

增。在本方法中有用的多种聚合酶是可用的,其非限制性实例包括Bst DNA聚合酶,大片段;Bsu DNA聚合酶,大片段;Deep Vent_RTM DNA聚合酶;Deep Vent_RTM(exo-)DNA聚合酶;Klenow片段(3' -5' exo-);DNA聚合酶I,大片段;M-MuLV逆转录酶;phi29DNA聚合酶;PyroPhage 3173聚合酶;Vent_R[®] DNA聚合酶;以及 Vent_R[®] (exo-) DNA聚合酶。

[0054] 用于进行滚环扩增的扩增反应混合物可包含用于引物延伸反应的必需试剂,包括但不限于模板(例如,环状多核苷酸)、一种或多种引物、dNTP和缓冲液组分。一个扩增循环可包含(i)在变性温度下的变性,其中双链模板被转化为单链多核苷酸,(ii)在退火温度下的引物退火,其中引物与单链多核苷酸杂交,以及(iii)在延伸温度下持续给定延伸时间段的引物延伸,其中使用单链多核苷酸作为模板延伸与单链多核苷酸杂交的引物。链置换聚合酶在RCA中尤其有用,因为置换允许聚合酶不止一次地围绕环状模板持续进行聚合,从而生成与环状模板互补的序列的多联体串联拷贝。使用环状多核苷酸作为模板,引物延伸可以在该模板上持续,从而生成包含多个拷贝的环状多核苷酸序列(例如,多联体)的扩增产物。在一些实施方案中,通过使扩增反应混合物经受多个循环的滚环扩增来生成多种扩增产物。

[0055] 在一些实施方案中,所述多个循环包括至少2个循环(例如,至少3、4、5、6、7、8、9或10个循环)。RCA的多个循环可导致多个线性多联体由环状模板形成。在变性期间,终止第一多联体从环状模板的延伸。通过重复引物结合和延伸,可以经多个循环从环状模板生成多个多联体。在一些实施方案中,使用三个温度阶段——用于变性的第一温度阶段、用于引物结合的第二温度阶段和用于引物延伸的第三温度阶段。在一些实施方案中,选择高于引物结合温度的引物延伸温度以使引物延伸期间的引物结合最小化。如在扩增反应混合物中包含反向引物的情况下,在引物延伸期间使引物结合最小化可减少较短扩增产物的形成并减少短片段的偏倚扩增,因为引物不太可能在扩增产物形成时与之杂交。在扩增产物形成时与之杂交的引物也可参与引物延伸,但可能导致小片段的优先扩增,因为在延伸期间,小环倾向于在给定时间段内比大片段生成更多拷贝的重复单元和更多引物结合位点。在一些实施方案中,选择用于引物延伸的温度可以比选择用于引物退火的温度高至少5°C(例如,至少6°C、7°C、8°C、9°C、10°C、11°C、12°C、13°C、14°C、15°C或更多)。选择用于引物延伸的温度可以比选择用于引物退火的温度高约1°C至20°C(例如,约2°C至18°C、约4°C至15°C或约5°C至10°C)。适用于非等温RCA的温度范围可取决于所用的聚合酶的性质。

[0056] 所述多个循环中的每个循环可包含在至少约70°C、71°C、72°C、73°C、74°C、75°C、76°C、77°C、78°C、79°C、80°C、81°C、82°C、83°C、84°C、85°C、86°C、87°C、88°C、89°C、90°C、91°C、92°C、93°C、94°C或95°C的变性温度下的变性。所述多个循环中的每个循环可包含在约70°C、71°C、72°C、73°C、74°C、75°C、76°C、77°C、78°C、79°C、80°C、81°C、82°C、83°C、84°C、85°C、86°C、87°C、88°C、89°C、90°C、91°C、92°C、93°C、94°C或95°C的变性温度下的变性。所述多个循环中的每个循环可包含在至多约70°C、71°C、72°C、73°C、74°C、75°C、76°C、77°C、78°C、79°C、80°C、81°C、82°C、83°C、84°C、85°C、86°C、87°C、88°C、89°C、90°C、91°C、92°C、93°C、94°C或95°C的变性温度下的变性。所述多个循环中的每个循环可包含在约70°C至约100°C、约70°C至约95°C、约70°C至约90°C、约70°C至约85°C、约70°C至约80°C或约70°C至约75°C的变性温度下的变性。所述多个循环中的每个循环可包含在约70°C至约100°C、约75°C至约100°C、约80°C至约100°C、约85°C至约100°C、约90°C至约100°C或约95°C至约100°C的变

性温度下的变性。所述多个循环中的每个循环可包含在变性温度下持续至少约5秒、10秒、15秒、20秒、25秒、30秒、35秒、40秒、45秒、50秒、55秒或60秒的变性。所述多个循环中的每个循环可包含在变性温度下持续约5秒、10秒、15秒、20秒、25秒、30秒、35秒、40秒、45秒、50秒、55秒或60秒的变性。所述多个循环中的每个循环可包含在变性温度下持续至多约5秒、10秒、15秒、20秒、25秒、30秒、35秒、40秒、45秒、50秒、55秒或60秒的变性。所述多个循环中的每个循环可包含在变性温度下持续约5秒至60秒、5秒至55秒、5秒至50秒、5秒至45秒、5秒至40秒、5秒至35秒、5秒至30秒、5秒至25秒、5秒至20秒、5秒至15秒或5秒至10秒的变性。所述多个循环中的每个循环可包含在变性温度下持续约5秒至60秒、10秒至60秒、15秒至60秒、20秒至60秒、25秒至60秒、30秒至60秒、35秒至60秒、40秒至60秒、45秒至60秒、50秒至60秒或55秒至60秒的变性。

[0057] 所述多个循环中的每个循环可包含在至少约45°C、46°C、47°C、48°C、49°C、50°C、51°C、52°C、53°C、54°C、55°C、56°C、57°C、58°C、59°C、60°C、61°C、62°C、63°C、64°C或65°C的退火温度下的引物退火。所述多个循环中的每个循环可包含在约45°C、46°C、47°C、48°C、49°C、50°C、51°C、52°C、53°C、54°C、55°C、56°C、57°C、58°C、59°C、60°C、61°C、62°C、63°C、64°C或65°C的退火温度下的引物退火。所述多个循环中的每个循环可包含在至多约45°C、46°C、47°C、48°C、49°C、50°C、51°C、52°C、53°C、54°C、55°C、56°C、57°C、58°C、59°C、60°C、61°C、62°C、63°C、64°C或65°C的退火温度下的引物退火。所述多个循环中的每个循环可包含在约45°C至约65°C、约45°C至约60°C、约45°C至约55°C或约45°C至约50°C的退火温度下的引物退火。所述多个循环中的每个循环可包含在约45°C至约65°C、约50°C至约65°C、约55°C至约65°C或约60°C至约65°C的退火温度下的引物退火。所述多个循环中的每个循环可包含在退火温度下持续至少约5秒、10秒、15秒、20秒、25秒、30秒、35秒、40秒、45秒、50秒、55秒或60秒的引物退火。所述多个循环中的每个循环可包含在退火温度下持续约5秒、10秒、15秒、20秒、25秒、30秒、35秒、40秒、45秒、50秒、55秒或60秒的引物退火。所述多个循环中的每个循环可包含在退火温度下持续至多约5秒、10秒、15秒、20秒、25秒、30秒、35秒、40秒、45秒、50秒、55秒或60秒的引物退火。所述多个循环中的每个循环可包含在退火温度下持续约5秒至60秒、5秒至55秒、5秒至50秒、5秒至45秒、5秒至40秒、5秒至35秒、5秒至30秒、5秒至25秒、5秒至20秒、5秒至15秒或5秒至10秒。所述多个循环中的每个循环可包含在退火温度下持续约5秒至60秒、10秒至60秒、15秒至60秒、20秒至60秒、25秒至60秒、30秒至60秒、35秒至60秒、40秒至60秒、45秒至60秒、50秒至60秒或55秒至60秒的引物退火。

[0058] 所述多个循环中的每个循环可包含在至少约65°C、66°C、67°C、68°C、69°C、70°C、71°C、72°C、73°C、74°C或75°C的延伸温度下的引物延伸。多个循环中的每个循环可包含在约65°C、66°C、67°C、68°C、69°C、70°C、71°C、72°C、73°C、74°C或75°C的延伸温度下的引物延伸。所述多个循环中的每个循环可包含在至多约65°C、66°C、67°C、68°C、69°C、70°C、71°C、72°C、73°C、74°C或75°C的延伸温度下的引物延伸。所述多个循环中的每个循环可包含在约65°C至约75°C或约65°C至约70°C的延伸温度下的引物延伸。所述多个循环中的每个循环可包含在约65°C至约75°C或约70°C至约75°C的延伸温度下的引物延伸。所述多个循环中的每个循环可包含在延伸温度下持续至少约30秒、1分钟、2分钟、3分钟、4分钟、5分钟、6分钟、7分钟、8分钟、9分钟或10分钟的延伸时间段的引物延伸。所述多个循环中的每个循环可包含在延伸温度下持续约30秒、1分钟、2分钟、3分钟、4分钟、5分钟、6分钟、7分钟、8分钟、9分钟

或10分钟的延伸时间段的引物延伸。所述多个循环中的每个循环可包含在延伸温度下持续至多约30秒、1分钟、2分钟、3分钟、4分钟、5分钟、6分钟、7分钟、8分钟、9分钟或10分钟的延伸时间段的引物延伸。所述多个循环中的每个循环可包含在延伸温度下持续约30秒至10分钟、30秒至9分钟、30秒至8分钟、30秒至7分钟、30秒至6分钟、30秒至5分钟、30秒至4分钟、30秒至3分钟、30秒至2分钟或30秒至1分钟的延伸时间段的引物延伸。所述多个循环中的每个循环可包含在延伸温度下持续约30秒至10分钟、1分钟至10分钟、2分钟至10分钟、3分钟至10分钟、4分钟至10分钟、5分钟至10分钟、6分钟至10分钟、7分钟至10分钟、8分钟至10分钟或9分钟至10分钟的延伸时间段的引物延伸。可以选择延伸时间段的长度来优化扩增产物产率。选择延伸时间段时所考虑的一些因素包括但不限于环状多核苷酸(例如,环状模板)的大小、环状多核苷酸序列的GC含量以及扩增产物中二级结构的形成。例如,对于产生相似产量的扩增产物,与较短环状多核苷酸相比,较长环状多核苷酸可能使用更长的延伸时间。对于进一步的实例,对于产生相似产量的扩增产物,与具有较低GC含量但长度相当的环状多核苷酸相比,具有较高GC含量的环状多核苷酸可能使用更长的延伸时间。

[0059] 在一些实施方案中,所述多个循环中的每个循环包含(i)在约75°C至约95°C的变性温度下持续约5秒至约60秒的变性,(ii)在约45°C至约65°C的退火温度下持续约5秒至约60秒的引物退火,以及(iii)在约65°C至约75°C的延伸温度下持续约30秒至约10分钟的延伸时间段的引物延伸。在一些实施方案中,所述多个循环中的每个循环包含(i)在约80°C的变性温度下持续约15秒至约30秒的变性,(ii)在约50°C的退火温度下持续约15秒至约45秒的引物退火,以及(iii)在约70°C的延伸温度下续约3分钟至约10分钟的延伸时间段的引物延伸。在一些实施方案中,所述多个循环中的每个循环包含(i)在约80°C的变性温度下持续约20秒的变性,(ii)在约50°C的退火温度下持续约30秒的引物退火,以及(iii)在约70°C的延伸温度下持续约6分钟的延伸时间段的引物延伸。

[0060] 在一些实施方案中,在所述多个循环的滚环扩增中的任意循环处为反应混合物补充聚合酶。在一些实施方案中,在所述多个循环的滚环扩增中的至少两个循环处为反应混合物补充聚合酶。在一些情况下,由于聚合酶活性的热失活,可能需要补充扩增反应混合物。一些聚合酶在升高的温度下热失活。热失活的温度可取决于所使用的聚合酶。根据选择用于扩增的聚合酶和针对变性温度、引物退火温度和引物延伸温度中的任一个所选择的温度,可任选地在至少一个扩增循环之后补充聚合酶。在一些实施方案中,在每隔一个循环之后,为反应混合物补充聚合酶。在各种实施方案中,例如由扩增产物的产率所决定的,根据需要为反应混合物补充聚合酶。

[0061] 在一些实施方案中,使用本文公开的方法生成的多种扩增产物的特征在于与利用变性和引物退火条件相当但延伸时间段相当于所述多个循环的延伸时间段之和的一个扩增循环生成的多种扩增产物相比,其含有更高比例的具有至少两个拷贝(例如,至少三个、四个或五个拷贝)的靶多核苷酸的多联体。

[0062] 相对于具有较少拷贝靶多核苷酸的扩增产物,具有较多拷贝靶多核苷酸的多联体或扩增产物可具有更大的片段大小。确定扩增产物或多联体中靶多核苷酸的拷贝数可使用多种方法来进行,例如通过琼脂糖凝胶、大小排阻色谱法或下一代测序进行分析。如使用任何合适的方法(例如,琼脂糖凝胶、大小排阻色谱法或下一代测序)所确定的,与包含一个循环的滚环扩增相比,使用如本文公开的包含多个循环滚环扩增的方法所生成的多联体或扩

增产物可具有比例增加的具有至少两个拷贝的靶多核苷酸的多联体。在一些实施方案中，具有至少两个拷贝的靶多核苷酸的多联体的比例增加至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、25%或50%。

[0063] 在一些实施方案中，所述多种扩增产物展现出至少约100、110、120、130、140、150、160、170、180、190、200、210、220、230、240或250个碱基对的平均片段长度。在一些实施方案中，所述多种扩增产物展现出约100、110、120、130、140、150、160、170、180、190、200、210、220、230、240或250个碱基对的平均片段长度。在一些实施方案中，所述多种扩增产物展现出至多约100、110、120、130、140、150、160、170、180、190、200、210、220、230、240或250个碱基对的平均片段长度。在一些实施方案中，所述多种扩增产物展现出约150个碱基对的平均片段长度。在一些实施方案中，所述多种扩增产物展现出约160个碱基对的平均片段长度。在一些实施方案中，所述多种扩增产物展现出约170个碱基对的平均片段长度。在一些实施方案中，所述多种扩增产物展现出约180个碱基对的平均片段长度。在一些实施方案中，所述多种扩增产物展现出约190个碱基对的平均片段长度。在一些实施方案中，所述多种扩增产物展现出约200个碱基对的平均片段长度。

[0064] 在一些实施方案中，所述多种扩增产物展现出至少约100、110、120、130、140、150、160、170、180、190、200、210、220、230、240或250个碱基对的中值片段长度。在一些实施方案中，所述多种扩增产物展现出约100、110、120、130、140、150、160、170、180、190、200、210、220、230、240或250个碱基对的中值片段长度。在一些实施方案中，所述多种扩增产物展现出至多约100、110、120、130、140、150、160、170、180、190、200、210、220、230、240或250个碱基对的中值片段长度。在一些实施方案中，所述多种扩增产物展现出约150个碱基对的中值片段长度。在一些实施方案中，所述多种扩增产物展现出约160个碱基对的中值片段长度。在一些实施方案中，所述多种扩增产物展现出约170个碱基对的中值片段长度。在一些实施方案中，所述多种扩增产物展现出约180个碱基对的中值片段长度。在一些实施方案中，所述多种扩增产物展现出约190个碱基对的中值片段长度。在一些实施方案中，所述多种扩增产物展现出约200个碱基对的中值片段长度。

[0065] 在一些实施方案中，期望分析无细胞多核苷酸，例如无细胞DNA。如本文其他各处进一步描述的无细胞DNA (cfDNA) 可以是循环肿瘤DNA或循环胎儿DNA。在一些情况下，无细胞多核苷酸可包括无细胞RNA。例如，无细胞DNA可通过使用诸如连接酶等酶的连接而环化并经由本方法来扩增。在一些实施方案中，当反应混合物中所使用的环状多核苷酸包含cfDNA时，所述多种扩增产物展现出约40个碱基至约450个碱基、40个碱基至约400个碱基、40个碱基至约350个碱基、40个碱基至约300个碱基、40个碱基至约250个碱基、40个碱基至约200个碱基、40个碱基至约150个碱基、40个碱基至约100个碱基或40个碱基至约50个碱基的片段长度分布。在一些实施方案中，当反应混合物中所使用的环状多核苷酸包含cfDNA时，所述多种扩增产物展现出约40个碱基至约450个碱基、50个碱基至约450个碱基、100个碱基至约450个碱基、150个碱基至约450个碱基、200个碱基至约450个碱基、250个碱基至约450个碱基、300个碱基至约450个碱基、350个碱基至约450个碱基或400个碱基至约450个碱基的片段长度分布。在一些实施方案中，当反应混合物中所使用的环状多核苷酸包含cfDNA时，所述多种扩增产物展现出约100个碱基至约200个碱基、110个碱基至约200个碱基、120个碱基至约200个碱基、130个碱基至约200个碱基、140个碱基至约200个碱基、150个碱基至

约200个碱基、160个碱基至约200个碱基、170个碱基至约200个碱基、180个碱基至约200个碱基或190个碱基至约200个碱基的片段长度分布。在一些实施方案中,当反应混合物中所使用的环状多核苷酸包含cfDNA时,所述多种扩增产物展现出约100个碱基至约200个碱基、100个碱基至约190个碱基、100个碱基至约180个碱基、100个碱基至约170个碱基、100个碱基至约160个碱基、100个碱基至约150个碱基、100个碱基至约140个碱基、100个碱基至约130个碱基、100个碱基至约120个碱基或100个碱基至约110个碱基的片段长度分布。

[0066] 在一些实施方案中,扩增反应混合物中的引物包含随机序列。在一些实施方案中,扩增反应混合物中的引物包含基因特异性序列。在一些实施方案中,引物包含针对多个基因(例如,至少2、3、4、5、6、7、8、9或10个基因)的基因特异性序列。在一些实施方案中,所述引物包括第一引物,该第一引物包含(i)通过序列互补性与环状多核苷酸特异性杂交的第一3'端和(ii)不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5'端。包含单链多核苷酸的多联体可在多个循环的滚环扩增期间通过使用环状多核苷酸作为模板延伸第一引物而生成。所述引物可包括第二引物,该第二引物包含(i)通过序列互补性与包含单链多核苷酸的多联体特异性杂交的第二3'端和(ii)不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5'端。包含一个或多个拷贝的靶多核苷酸的多种延伸产物可在多个循环的滚环扩增期间通过使用多联体作为模板延伸第二引物而生成。如本文进一步描述的,使用这样的第一引物和第二引物进行扩增的方法可用于富集包含至少两个或更多个拷贝的靶多核苷酸的多联体的扩增子。

[0067] 在一方面,本公开内容提供了一种用于富集包含至少两个或更多个拷贝的靶多核苷酸的多联体的扩增子的方法。在一个实施方案中,该方法包括(a)通过延伸第一引物生成包含来自环状靶多核苷酸的单链多核苷酸的多联体,该第一引物包含通过序列互补性与靶多核苷酸特异性杂交的第一3'端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5'端,(b)通过延伸第二引物生成包含一个或多个拷贝的靶多核苷酸的多种延伸产物,该第二引物包含通过序列互补性与多联体特异性杂交的第二3'端和不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5'端,其中第一共同序列和第二共同序列各自在5'端包含至少10个连续核苷酸并且当最佳比对时至少90%相同,以及(c)在生成多个扩增子的条件下对步骤(b)的多种延伸产物进行扩增,其中富集包含至少2个或更多个拷贝的靶多核苷酸的扩增子。

[0068] 在一些实施方案中,从环状靶多核苷酸生成包含单链多核苷酸的多联体包括第一引物的延伸。引物延伸可以通过包括但不限于热循环反应和等温反应的扩增反应完成。在一些实施方案中,热循环反应涉及例如变性、引物结合和引物延伸的若干个循环。在一些实施方案中,生成本方法的多联体通过聚合酶而实现。在本方法中有用的各种聚合酶是可用的,其非限制性实例在本文中提供。在一些实施方案中,用于实现多联体生成的聚合酶具有链置换活性。在一些实施方案中,从环状靶多核苷酸生成包含单链多核苷酸的多联体包括等温滚环扩增(RCA)。链置换聚合酶在RCA中尤其有用,因为置换允许聚合酶不止一次地围绕环状模板持续进行聚合,从而生成与环状模板互补的序列的多联体串联拷贝。在一些实施方案中,生成包含单链多核苷酸的多联体包括非等温RCA。非等温RCA可包括至少2个温度阶段(例如,至少2、3、4个或更多个温度阶段)的至少两个循环(例如,至少2、3、4、5、6、7、8、9、10个或更多个循环)。例如,第一温度阶段可适用于引物结合和延伸,而第二温度阶段可

适用于使双链多核苷酸变性。在一些实施方案中,非等温RCA包括至少两个温度阶段的2至35个循环(例如,3至30个、4至20个、5至15个或6至10个循环)。在非等温RCA期间循环通过包括适用于使双链多核苷酸变性的第二温度阶段在内的温度阶段可导致由环状模板形成多个线性多联体。在变性期间,终止第一多联体从环状模板的延伸。通过重复引物结合和延伸,可以经若干个循环从环状模板生成多个多联体。在一些实施方案中,使用三个温度阶段-用于引物退火的第一温度阶段、用于引物延伸的第二温度阶段和用于使双链多核苷酸变性的第三温度阶段。在一些实施方案中,选择高于引物结合温度的引物延伸温度以使引物延伸期间的引物结合最小化。如在扩增反应混合物中包含反向引物的情况下,在引物延伸期间使引物结合最小化可减少较短扩增产物的形成并减少短片段的偏倚扩增,因为引物不太可能在扩增产物形成时与之杂交。在扩增产物形成时与之杂交的引物也可参与引物延伸,但可能导致小片段优先扩增,因为在延伸期间,小环倾向于在给定时间段内比大片段生成更多拷贝的重复单元和更多引物结合位点。在一些实施方案中,选择用于引物延伸的温度可比选择用于引物退火的温度高至少5°C(例如,至少6°C、7°C、8°C、9°C、10°C、11°C、12°C、13°C、14°C、15°C或更多)。选择用于引物延伸的温度可比选择用于引物退火的温度高约1°C至20°C(例如,约2°C至18°C、约4°C至15°C或约5°C至10°C)。在一些实施方案中,非等温RCA可包含第四温度阶段。第四温度阶段可例如适合于在延伸产物中形成二级结构,例如形成茎环结构。适合于非等温RCA的温度范围可取决于所使用的聚合酶的性质。

[0069] 在一些实施方案中,RCA(等温或非等温)可包含沿线性模板的引物延伸反应,如沿着由沿环状模板引物延伸产生的线性多联体的反向引物延伸。例如,第二引物可与包含作为第一引物的延伸产物而生成的线性多联体模板的模板杂交,并且在引物延伸阶段第二引物的引物延伸可生成线性双链多核苷酸,该线性双链多核苷酸的链可进一步作为通过附加拷贝的第一和第二引物进行延伸的模板。

[0070] 在一些实施方案中,生成包含一个或多个拷贝的靶多核苷酸的多种延伸产物包括延伸与作为第一引物的延伸产物而产生的线性多联体模板杂交的第二引物。在一些实施方案中,在非等温RCA期间,多种延伸引物的生成可与线性多联体从环状模板的的生成同时进行。引物延伸和扩增的方法通常有利于短片段扩增,而不是长片段扩增。根据一些实施方案,可优化生成延伸产物的引物延伸反应以减少有利于较短产物的偏倚,并由此增加较长产物(如包含两个或更多个拷贝的靶多核苷酸的产物)的比例。本公开内容考虑了实现该目的的各种方式,其可以单独使用或组合使用。实现该目的的一种方式是通过限制引物延伸循环的数目,使得短片段不被优先扩增,或与长度相同但缺乏发夹结构的模板相比以减少的频率扩增。在一些实施方案中,生成延伸产物包括不超过15个循环(例如,不超过10、8、6个或更少的循环)的第二引物延伸。在一些实施方案中,生成延伸产物包括2至15个循环的第二引物延伸。在一些实施方案中,生成延伸产物包括2至10个循环的第二引物延伸。在一些实施方案中,第二引物的延伸与第一引物的延伸同时发生。

[0071] 在一些实施方案中,本文中与环状多核苷酸可互换使用的可用于生成多联体的环状靶多核苷酸由线性靶多核苷酸形成。在一些实施方案中,环状靶多核苷酸或环状多核苷酸为单链。在一些实施方案中,环状靶多核苷酸或环状多核苷酸为双链。环状靶多核苷酸或环状多核苷酸可以具有任意长度。在一些实施方案中,环状靶多核苷酸或环状多核苷酸的长度为约25、50、75、100、150、200、250、300、400、500、600、700或800个核苷酸。在一些实施

方案中,环状靶多核苷酸或环状多核苷酸的长度为25-1000个核苷酸。在一些实施方案中,环状靶多核苷酸或环状多核苷酸的长度为50-500个核苷酸。在一些实施方案中,环状靶多核苷酸或环状多核苷酸的长度为75-250个核苷酸。在一些实施方案中,环状靶多核苷酸或环状多核苷酸的长度为100-200个核苷酸。环状靶多核苷酸或环状多核苷酸可包含染色体或基因片段。在一些实施方案中,环状靶多核苷酸或环状多核苷酸包含基因产物,包括但不限于miRNA、rRNA、tRNA和mRNA。在一些实施方案中,环状靶多核苷酸或环状多核苷酸包含由点突变、SNP、插入或缺失产生的序列。在一些实施方案中,环状靶多核苷酸或环状多核苷酸包含由染色体重排产生的序列。染色体重排可以是一个或多个倒位;一个或多个缺失;一个或多个重复;一个或多个易位;或其组合。在一些实施方案中,包含一个或多个易位的环状靶多核苷酸或环状多核苷酸包含融合基因的融合点或融合接头。在一些实施方案中,环状靶多核苷酸或环状多核苷酸包含倒位、缺失、重复和易位中的至少一种。

[0072] 在一种或多种本方法中用于生成多联体的第一引物可包含通过序列互补性与靶多核苷酸特异性杂交的第一3'端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5'端。第一引物可以是任意合适的长度,如至少5、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90或100个核苷酸,其任何部分均可与引物所杂交的相应靶序列互补(例如,至少5、10、15、20、25、30、35、40、45或50个核苷酸)。通过序列互补性与靶序列特异性杂交的第一引物的第一3'端可以是任意合适的长度,如至少10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90或95个核苷酸长。在一些实施方案中,第一引物包含含有随机核苷酸序列的第一3'端,该随机核苷酸序列随机杂交并引发环状靶多核苷酸或环状多核苷酸的各个随机区域以供引物延伸,每个随机序列均通过序列互补性与相应互补序列特异性杂交。当引物包含随机3'端序列时,靶序列是通过引物延伸而扩增的序列。一般而言,3'端包含3'末端核苷酸。第一5'端(具有第一引物的第一共同序列且不通过序列互补性与靶多核苷酸特异性杂交)可以是任意合适的长度,如至少10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90或95个核苷酸长。第一共同序列可以是任意合适的长度。在一些实施方案中,第一引物的第一共同序列为至少10个核苷酸长(例如,至少15、20、25、30个或更多个核苷酸长)。通常,5'端是指相对于3'端处于5'的多核苷酸部分。在一些实施方案中,5'端包含5'末端核苷酸。

[0073] 包含通过序列互补性与多联体特异性杂交的第二3'端和不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5'端的第二引物可用于通过引物延伸生成包含一个或多个拷贝的靶多核苷酸的多种延伸产物。用于生成多种延伸产物的第二引物可以具有任意合适的长度,如约为或至少为5、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90或100个核苷酸,其任何部分均可与引物所杂交的相应靶序列互补(例如,至少5、10、15、20、25、30、35、40、45、50个核苷酸)。通过序列互补性与多联体特异性杂交的第二引物的第二3'端可以是任意合适的长度,如至少10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90或95个核苷酸长。在一些实施方案中,第二引物包含含有随机核苷酸序列的第二3'端,该随机核苷酸序列随机杂交并引发多联体的各个随机区域以供引物延伸。不通过序列互补性与多联体特异性杂交的、包含第二引物的第二共同序列的第二5'端可以是任意合适的长度,如至少10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90或95个核苷酸长。第二共同序列可以是任意合适的长度。在一些实施方案中,第二引物的第二共同序列的

长度为至少10个核苷酸(例如,至少15、20、25、30个或更多个核苷酸长)。通常,5' 端是指相对于3' 端在5' 的多核苷酸部分。在一些实施方案中,5' 端包含5' 末端核苷酸。在一些实施方案中,第二共同序列与第一共同序列至少80%相同(例如,至少90%、95%或100%相同),使得第二共同序列可在合适的反应条件(例如,扩增反应中的一个或多个步骤,如引物杂交步骤和/或引物延伸步骤)下与第一共同序列的互补体杂交。在一些实施方案中,第一共同序列和第二共同序列是相同的。

[0074] 通常,不与靶标特异性杂交的共同序列被设计成在3' 端与靶多核苷酸杂交的条件(例如,扩增反应中的一个或多个步骤)下不与靶多核苷酸杂交。在一些实施方案中,共同序列被设计成与沿着靶多核苷酸在相对于3' 端杂交处的3' 的序列、靶多核苷酸内任意处的序列,或样品(例如,生物体如细菌、病毒、植物或动物的所有基因组序列,包括人类基因组DNA序列)中的多核苷酸的任意基团具有小于75%、50%、25%、10%或更低的互补性。在某些实施方案中,第一共同序列和第二共同序列各自在5' 端包含至少10个(例如,至少15、20、25、30、40、50个或更多个)连续核苷酸,并且当最佳比对时至少90%相同。在某些实施方案中,第一共同序列和第二共同序列在5' 端包含至少5、10、15、20、25或30个连续核苷酸,并且当最佳比对时至少70%相同(例如,至少80%、90%、95%或100%相同)。

[0075] 在一些实施方案中,本方法的延伸产物形成茎环结构,该茎环结构包含(i)第一共同序列与第二共同序列的互补体之间的分子内杂交,和/或(ii)第二共同序列与第一共同序列的互补体之间的分子内杂交。延伸产物可在非等温RCA期间,例如在具有适于形成茎环结构的温度的第四温度阶段期间形成茎环结构。在一些实施方案中,茎环结构在RCA之后的后续扩增反应期间形成。茎环结构的形成可取决于双链茎区和单链环区的稳定性。茎的稳定性可取决于其长度、错配数目和碱基组成。茎环结构的稳定性还取决于环长度。没有二级结构的大环可能不稳定,并且短于三个碱基长的环可能在空间上是不可能的。在一些实施方案中,具有较长茎部分的茎环结构可比具有相同环和较短茎的茎环结构更稳定。在一些情况下,具有较长环的茎环结构可能不如具有相同茎和较短环的茎环结构稳定。

[0076] 生成多联体的方法的说明性实施方案示于图1中。通过滚环扩增(RCA)延伸第一引物,该第一引物包含通过序列互补性与靶多核苷酸特异性杂交的第一3' 端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列5' -TACGCA-3' 的第一5' 端。接下来,生成包含一个或多个拷贝的靶多核苷酸的多种延伸产物包括延伸第二引物,该第二引物包含通过序列互补性与多联体特异性杂交的第二3' 端和不通过序列互补性与多联体特异性杂交的包含第二共同序列5' -TACGCA-3' (与第一共同序列相同)的第二5' 端。由于第二共同序列与第一共同序列的互补体之间的分子内杂交,延伸产物可形成茎环结构。茎环结构可包含可变拷贝数的靶多核苷酸。碱基对中茎的长度可以变化。在一些实施方案中,茎的长度为至少6、9、15、20、25、30个或更多个碱基对。在一些实施方案中,该茎环结构包含5至30个碱基对的分子内杂交。在一些实施方案中,该茎环结构包含10至20个碱基对的分子内杂交。表1中提供了可用作第一或第二共同序列的序列的非限制性实例。

[0077] 表1.候选共同序列的非限制性实例

候选共同序列	序列	
共同_001	CCATCTAATTCAACAAGAATTGGGACAAC	
共同_002	ACATGGGTGGTGGTATAGCGCTTGC	
共同_003	CAATTACATCTTATTATTAAACG	
共同_004	AGCTCGTTAGTGAACCGTCAGATC	
共同_005	GAGTCACTTAAAATTGTATACAC	
共同_006	CAAGGCTGTTAGAGAGATAATTGGA	
共同_007	GTGAGTGATGGTTGAGGTAGTGTGGAG	
共同_008	AGCTGGACATCACCTCCCACAACG	
共同_009	CTCTGAATACTTCAACAAGTTAC	
共同_010	AATATACCTCTATACTTTAACGTC	
共同_011	GATGAAGCCCTGAAAGACGCGCAG	
共同_012	GCATCAATGCAGAAGCTGATCTCA	
共同_013	GACGGCATCGCAGCTGGATACAC	
共同_014	CTTAGCATGTCCGTGGGGTTGAAT	
共同_015	GAGCGGATAACAATTACACAGG	
[0078]	共同_016	CGGTAGGTATTGATTGTAATTCTG
	共同_017	CCCAGTCACGACGTTGTAAAACG
	共同_018	AGCGGATAACAATTACACAGG
	共同_019	CCCTTGAACCTCCTCGTTGACC
	共同_020	CCCTTGAACCTCCTCGTTGACC
	共同_021	CAGCGGGCTGCTAAAGCGCATGC
	共同_022	CTACAAACTCTCCTGTTAGTTAG
	共同_023	CGGGTGTGTTGTTGTTGTTGTTG
	共同_024	CGGGTGTGTTGTTGTTGTTGTTG
	共同_025	CGGGTGTGTTGTTGTTGTTGTTG
	共同_026	CGGGTGTGTTGTTGTTGTTGTTG
	共同_027	CGGGTGTGTTGTTGTTGTTGTTG
	共同_028	CGGGTGTGTTGTTGTTGTTGTTG
	共同_029	CGGGTGTGTTGTTGTTGTTGTTG
	共同_030	CGGGTGTGTTGTTGTTGTTGTTG
	共同_031	CGGGTGTGTTGTTGTTGTTGTTG
[0079]	共同_032	CGGGTGTGTTGTTGTTGTTGTTG
	共同_033	CGGGTGTGTTGTTGTTGTTGTTG
	共同_034	CGGGTGTGTTGTTGTTGTTGTTG
	共同_035	CGGGTGTGTTGTTGTTGTTGTTG
	共同_036	CGGGTGTGTTGTTGTTGTTGTTG
	共同_037	CGGGTGTGTTGTTGTTGTTGTTG
	共同_038	CGGGTGTGTTGTTGTTGTTGTTG
	共同_039	CGGGTGTGTTGTTGTTGTTGTTG
	共同_040	CGGGTGTGTTGTTGTTGTTGTTG
	共同_041	CGGGTGTGTTGTTGTTGTTGTTG
	共同_042	CGGGTGTGTTGTTGTTGTTGTTG
	共同_043	CGGGTGTGTTGTTGTTGTTGTTG

[0080] 参与分子内杂交的碱基对的数目可取决于第一共同序列和第二共同序列的连续核苷酸的数目或第二共同序列与第一共同序列的互补体的连续核苷酸的数目。参与分子内杂交的碱基对的数目还可取决于第一共同序列与第二共同序列的同一性百分比，其中同一性百分比是指当第一共同序列和第二共同序列最佳比对时第一共同序列与第二共同序列之间相同碱基的百分比。在一些实施方案中，第一共同序列和第二共同序列各自在5'端包含至少10个连续核苷酸，并且当最佳比对时至少90%相同。在一些实施方案中，第一共同序列和第二共同序列各自在5'端包含至少5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45或50个连续核苷酸。在一些实施方案中，第一共同序列和第二共同序列各自在5'端包含5至25个连续核苷酸。在一些实施方案中，当最佳比对时，第一共同序列和第二共同序列至少60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%或99%相同。在一些实施方案中，当最佳比对时，第一共同序列和第二共同序列60%至100%相同。在一些实施方案中，第一共同序列和第二共同序列在5'端包含5至25个连续核苷酸，并且当最佳比对时60%至100%相同。

[0081] 对第一引物和第二引物的多种延伸产物进行扩增可包括第三引物的引物延伸。在一些实施方案中，第三引物包含通过序列互补性与第一共同序列和/或第二共同序列特异性杂交的序列。用于核酸扩增的第三引物可以具有任意合适的长度，如至少5、6、7、8、9、10、

15、20、25、30、35、40、45、50、55、60、65、70、75、80、90或100个核苷酸,其任何部分或全部可与相应靶序列互补(例如,约5、10、15、20、25、30、35、40、45、50个核苷酸或更多)。第三引物可包含区段,该区段包含一个或多个扩增引物退火序列或其互补体;一个或多个测序引物退火序列或其互补体;一个或多个条形码序列;一个或多个在多种不同引物之间共有的共同序列;一个或多个限制酶识别位点;一个或多个探针结合位点或测序衔接子(例如,用于衔接至测序平台,如用于大规模平行测序的流动池);一个或多个随机或接近随机的序列(例如,从在一个或多个位置处的一组两个或更多个不同核苷酸中随机选择的一个或多个核苷酸);以及它们的组合。扩增引物退火序列也可用作测序引物退火序列。

[0082] 根据茎环结构的稳定性,该茎环结构可以以可变的效率扩增。通常,在包含相同长度的茎和可变长度的环的多个茎环结构中,包含较长环的茎环结构热力学较不稳定,作为模板更容易接近,因此可以更有效地被扩增。因此,在一些实施方案中,对两侧是可杂交的共同序列的靶序列的扩增富集了包含至少2个或更多个拷贝的靶序列的扩增子。由第三引物的引物延伸产生的扩增子可包含可变拷贝数的靶多核苷酸。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为50% (例如,至少60%、70%、80%、90%或更大)。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比为10% - 100% (例如,20% - 90%、30% - 80%或40% - 60%)。

[0083] 在实施本方法时,茎环产物的形成和包含至少2个或更多个拷贝的靶多核苷酸的扩增子的富集可通过指定第一共同序列、第二共同序列和第三引物序列的杂交序列的解链温度以及调节进行扩增的温度中的一项或多项来优化。对于其中引物延伸产物的扩增包含与第一共同序列和/或第二共同序列特异性杂交的第三引物的引物延伸的实施方案,第三引物的引物结合效率可取决于第一共同序列、第二共同序列和第三引物的杂交序列的解链温度中的一个或多个。在一些实施方案中,第一共同序列、第二共同序列和第三引物的杂交序列均具有在彼此±15°C (例如,±10°C、±5°C或±1°C) 内的解链温度 (T_m)。在一些实施方案中,第一共同序列、第二共同序列和第三引物的杂交序列均具有在彼此±5°C 内的解链温度 (T_m)。一般而言,T_m通常代表50% 的由参考序列(其实际上可能是较大多核苷酸内的子序列)及其互补序列组成的寡核苷酸杂交(或分离)时的温度。T_m可基于本领域可用的标准计算、算法或测量。用于测量T_m的示例性工具OligoAnalyzer由Integrated DNA Technologies使得在www.idtdna.com/calc/analyzer下可用,该工具可设置为使用默认参数。其他类似的工具是可用的。

[0084] 进行扩增的温度也可影响长茎环产物和短茎环产物的引物结合和延伸的效率。在某些实施方案中,茎环结构的形成可通过在退火步骤的温度保持在第三引物的解链温度±15°C 内(例如,±10°C、±5°C或±1°C内)的情况下进行(c)的扩增步骤而实现。在一些实施方案中,茎环产物的形成通过在退火步骤保持在低于75°C (例如,低于70°C、65°C、60°C或更低)的温度的情况下进行(c)的扩增步骤而实现。在一些实施方案中,茎环产物的形成通过在退火步骤保持在55°C - 75°C (例如,60°C - 70°C) 的温度的情况下进行(c)的扩增步骤而实现。

[0085] 在一些实施方案中,所述环状靶多核苷酸或环状多核苷酸为环化的无细胞多核苷酸(例如,无细胞DNA、cDNA或RNA)。在一些实施方案中,所述环状靶多核苷酸或环状多核苷酸为基因组DNA的环化片段。在一些实施方案中,所述环状靶多核苷酸或环状多核苷酸包含

由染色体重排产生的序列。在某些实施方案中,该染色体重排是缺失、重复、倒位和易位中的至少一种。在一些实施方案中,本方法的环状靶多核苷酸或环状多核苷酸为单链。在一些实施方案中,本方法的环状靶多核苷酸或环状多核苷酸为双链。在某些实施方案中,沿靶多核苷酸从5'至3'对应于(i)与第一3'端互补的序列、(ii)与第二3'端相同的序列以及(iii)(i)与(ii)之间的间插序列的靶多核苷酸序列部分的组合长度为75个或更少的核苷酸。在一些实施方案中,序列部分的组合长度为60个或更少的核苷酸。在一些实施方案中,序列部分的组合长度为50个或更少的核苷酸。在一些实施方案中,序列部分的组合长度为40个或更少的核苷酸。在一些情况下,序列部分的组合长度为30个或更少的核苷酸。

[0086] 在一个说明性实施方案中,第一引物和第二引物如图2所示排列。简单起见,第一和第二引物的相对杂交位置相对于靶多核苷酸的单链进行说明。然而,如下所述,一个引物与包含靶序列的链杂交,而另一引物与包含靶序列的互补体的链杂交。第一引物即正向引物(F引物)的第一3'端通过序列互补性与靶多核苷酸特异性杂交,而第一5'端未与靶多核苷酸特异性杂交。第二引物即反向引物(R引物)的第二3'端通过序列互补性与靶多核苷酸的互补体特异性杂交,而第二5'端未与靶多核苷酸的互补体特异性杂交。鉴于正向引物(F引物)和反向引物(R引物)相对于靶序列单体的取向,这种排列可被称为“背靠背”(B2B)或“倒置”引物。与传统的头对头设计相比,这种引物设计具有减少的引物足迹(一对引物所跨越的总距离)。在图2中,沿靶多核苷酸从5'至3'对应于(i)与第一3'端互补的序列、(ii)与第二3'端相同的序列以及(i)与(ii)之间的间插序列的靶多核苷酸序列部分的组合长度为约30-100个核苷酸(例如,40-80或50-70个核苷酸)。该组合长度也被称为“引物足迹”。在一些实施方案中,引物足迹的长度小于100个核苷酸(例如,长度小于90、80、70、60、50个或更少的核苷酸)。在一些实施方案中,可使用具有背靠背排列的第一引物和第二引物扩增包含点突变、indel(插入/缺失)或基因融合的环状靶多核苷酸或环状多核苷酸。这种引物对减小的引物足迹允许对围绕靶序列的各种片段化事件进行扩增,因为在B2B引物之间发生接合(例如,融合接合)的可能性小于在典型扩增反应(彼此面对,跨越靶序列)中发现的引物排列中发生该接合的可能性。

[0087] 在一个说明性实施方案中,如图3所示构建测序文库。线性DNA分子首先被环化形成RCA的模板。在5'端具有常规测序衔接子的背靠背引物与靶分子结合,而具有链置换活性的聚合酶在RCA期间扩增该靶标。可以将该文库测序,或在测序以检测序列变体(例如点突变、SNP和融合基因)之前通过PCR扩增将该文库进一步扩增。

[0088] 在一些实施方案中,可从样品中的多种靶多核苷酸生成多个多联体。样品可包含一种或多种靶序列。每种靶序列可具有一个或多个相应多联体。对应于独特靶多核苷酸的每个多联体可包含可变拷贝数的靶多核苷酸。在一些实施方案中,可以优化本方法以生成可变长度的多联体。多联体长度的可变性可由靶多核苷酸长度和/或每个多联体中靶多核苷酸的拷贝数的变化而导致。在一些实施方案中,至少50%(例如,至少60%、70%、80%、90%或更多)的多联体包含长度为至少75个核苷酸(例如,长度为至少100、150、200个或更多个核苷酸)的靶多核苷酸。在一些实施方案中,至少80%的多联体包含长度为至少75个核苷酸的靶多核苷酸。在一些实施方案中,至少60%的多联体包含长度为至少100个核苷酸的靶多核苷酸。在一些实施方案中,至少50%的多联体包含长度为至少150个核苷酸的靶多核苷酸。

[0089] 在一些实施方案中,本公开内容的方法包括对步骤(c)中产生的多个扩增子进行测序。在一些实施方案中,所述测序在没有相对于包含仅一个拷贝的靶多核苷酸的扩增子选择性地纯化包含两个或更多个拷贝的靶多核苷酸的扩增子的情况下进行。在一些实施方案中,本公开内容的方法包括对步骤(c)产生的多个扩增子中包含两个或更多个拷贝的靶多核苷酸的扩增子进行纯化。在一些实施方案中,对本方法的经纯化的扩增子进行测序。在某些实施方案中,本公开内容的方法包括在同一反应混合物中扩增多种不同的靶多核苷酸。所述多种靶多核苷酸的组分可以具有不同的长度。在一些实施方案中,所述靶多核苷酸的长度为30个核苷酸至1000个核苷酸(例如,长度为50-600、75-500、100-400或200-300个核苷酸)。在一些实施方案中,通过在单一反应混合物中连接而将所述靶多核苷酸环化。

[0090] 在一个说明性实施方案中,如图4所示对包含无细胞多核苷酸混合物的核酸样品进行扩增。混合物中的多核苷酸(例如,单链DNA,“ssDNA”)可被环化形成环状靶多核苷酸或环状多核苷酸。在非等温RCA的第一温度阶段(如55°C)中的引物结合和在非等温RCA的第二温度阶段(如70°C)中对一种或多种第一引物的引物延伸生成多联体混合物。选择用于引物延伸的温度高于选择用于引物结合的温度可以使引物延伸期间额外的引物结合最小化。每个靶多核苷酸可具有一个或多个相应多联体。对应于独特靶多核苷酸的每个多联体可包含可变拷贝数的靶多核苷酸。根据图示,一个或多个第一引物包含通过序列互补性与靶多核苷酸特异性杂交的第一3'端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5'端。在非等温扩增的后续循环中,多种第二延伸产物的生成与多联体的生成同时发生,该延伸产物由在第二温度阶段中一种或多种第二引物的引物延伸而得到,该一种或多种第二引物已与在第一温度阶段中作为第一引物的延伸产物生成的线性多联体模板杂交,并且由于在先前循环中的第三温度阶段(如94°C)期间的变性而不与环状模板杂交。第二引物的新杂交位点经环状模板周围的进展聚合酶的置换而暴露。所述一种或多种第二引物包含通过序列互补性与多联体特异性杂交的第二3'端和不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5'端。延伸产物可包含不同拷贝数的靶多核苷酸。在混合样品中,靶多核苷酸可具有不同的长度,并且所得到的延伸产物也可具有不同的长度。可以由第一共同序列与第二共同序列的互补体之间的分子内杂交,或由第一共同序列的互补体与(第二)共同序列之间的分子内杂交而形成不同大小的茎环结构。茎环结构可以在扩增的一个或多个阶段(例如,退火、延伸或用于茎环形成的第四温度阶段(例如58°C))期间形成。茎环结构还可在RCA之后的后续扩增反应期间形成。延伸产物可作为扩增反应的模板以生成扩增子,并且茎环结构的稳定性可影响引物结合和延伸。在后续扩增反应期间,相比于其中茎环结构具有较小环的延伸产物,可优先富集包含更长靶多核苷酸序列或更多拷贝的延伸产物。在一些实施方案中,富集包含两个或更多个拷贝的靶多核苷酸的扩增子。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比至少为50%(例如,至少60%、70%、80%、90%或更大)。在一些实施方案中,具有两个或更多个拷贝的靶多核苷酸的扩增子的百分比为10%-100%(例如,20%-90%、30%-80%或40%-60%)。在一些实施方案中,富集包含较长靶多核苷酸的扩增子。在一些实施方案中,至少50%(例如,至少60%、70%、80%、90%或更多)的多联体包含长度为至少75个核苷酸(例如,长度为至少100、150、200个或更多个核苷酸)的靶多核苷酸。在一些实施方案中,至少80%的多联体包含长度为至少75个核苷酸的靶多核苷酸。在一些实施方案中,至少60%的

多联体包含长度为至少100个核苷酸的靶多核苷酸。在一些实施方案中,至少50%的多联体包含长度为至少150个核苷酸的靶多核苷酸。

[0091] 在另一方面,本公开内容提供了用于执行根据本公开内容的方法中的方法的反应混合物。反应混合物可包含如本文中关于各个方面和方法中的任一个所描述的各种组分中的一种或多种。在一些实施方案中,本公开内容提供了用于富集包含至少两个或更多个拷贝的靶多核苷酸的多联体的扩增子的反应混合物。在一个实施方案中,该反应混合物包含:(a)环状靶多核苷酸,(b)第一引物,其包含通过序列互补性与靶多核苷酸特异性杂交的第一3'端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5'端;以及(c)第二引物,其包含通过序列互补性与多联体特异性杂交的第二3'端和不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5'端,其中第一共同序列和第二共同序列各自在5'端包含至少10个连续核苷酸并且当最佳比对时至少90%相同。

[0092] 在一些实施方案中,本公开内容的反应混合物容纳在容器内。每种组分可以被封装到不同的容器中,或者在交叉反应性和保质期允许的情况下,可在容器中提供组分的组合。容器可以是孔、板、管、腔室、流动池或芯片。

[0093] 在一些实施方案中,反应混合物包含具有通过序列互补性与第一共同序列或第二共同序列特异性杂交的序列的第三引物。在一些实施方案中,第三引物可用于扩增多种延伸产物。用于核酸扩增的第三引物可具有任意合适的长度,如至少5、6、7、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90或100个核苷酸,其任何部分或全部可与相应靶序列互补(例如,约5、10、15、20、25、30、35、40、45、50个核苷酸或更多)。第三引物可包含区段,该区段包含一个或多个扩增引物退火序列或其互补体;一个或多个测序引物退火序列或其互补体;一个或多个条形码序列;一个或多个在多种不同引物之间共有的共同序列;一个或多个限制酶识别位点;一个或多个探针结合位点或测序衔接子(例如,用于附接至测序平台,如用于大规模平行测序的流动池);一个或多个随机或接近随机的序列(例如,从一个或多个位置处的一组两个或更多个不同核苷酸中随机选择的一个或多个核苷酸);以及它们的组合。扩增引物退火序列也可以用作测序引物退火序列。

[0094] 在某些实施方案中,延伸产物可形成茎环产物,并且可通过优化第一共同序列、第二共同序列和第三引物的第三杂交序列的性质,例如通过优化它们的解链温度来优化从第三引物的引物延伸的扩增子产率。在一些实施方案中,第一共同序列、第二共同序列和第三引物的杂交序列均具有在彼此±15°C内的解链温度(T_m)。在一些实施方案中,第一共同序列、第二共同序列和第三引物的杂交序列均具有在彼此±10°C内的解链温度(T_m)。在一些实施方案中,第一共同序列、第二共同序列和第三引物的杂交序列均具有在彼此±5°C内的解链温度(T_m)。在一些实施方案中,第一共同序列、第二共同序列和第三引物的杂交序列均具有在彼此±1°C内的解链温度(T_m)。

[0095] 在一些实施方案中,本公开内容的反应混合物包含环化的无细胞DNA作为环状靶多核苷酸或环状多核苷酸。在一些实施方案中,本公开内容的反应混合物包含基因组DNA的环化片段作为环状靶多核苷酸或环状多核苷酸。在一些实施方案中,该环状靶多核苷酸或环状多核苷酸包含由染色体重排产生的序列。在某些实施方案中,该染色体重排是缺失、重复、倒位和易位中的至少一种。在一些实施方案中,本方法的环状靶多核苷酸或环状多核苷酸为单链。在一些实施方案中,本方法的环状靶多核苷酸或环状多核苷酸为双链。

[0096] 在一些实施方案中,本公开内容的反应混合物包含沿靶多核苷酸从5'至3'对应于(i)与第一3'端互补的序列、(ii)与第二3'端相同的序列以及(iii)(i)与(ii)之间的间插序列的靶多核苷酸序列部分的组合长度为75个或更少的核苷酸。在一些实施方案中,靶多核苷酸序列部分的组合长度为60个或更少的核苷酸。在一些实施方案中,靶多核苷酸序列部分的组合长度为50个或更少的核苷酸。在一些实施方案中,靶多核苷酸序列部分的组合长度为40个或更少的核苷酸。在一些实施方案中,靶多核苷酸序列部分的组合长度为30个或更少的核苷酸。

[0097] 在包括本公开内容的方法和反应混合物在内的本文所述各个方面的一些实施方案中,通过连接线性靶多核苷酸形成环状靶多核苷酸或环状多核苷酸。由线性靶多核苷酸形成的环状靶多核苷酸可包含待表征的序列,例如罕见序列变体或融合基因。在一些实施方案中,线性靶多核苷酸为单链。在其他实施方案中,线性靶多核苷酸为双链。靶多核苷酸的非限制性实例包括DNA、RNA、cDNA、dsDNA、ssDNA、质粒DNA、粘粒DNA、染色体DNA、基因组DNA、病毒DNA、细菌DNA、mtDNA(线粒体DNA)、mRNA、rRNA、tRNA、nRNA、siRNA、snRNA、snoRNA、scRNA、微小RNA、dsRNA、核酶、核糖开关和病毒RNA(例如,逆转录病毒RNA)。

[0098] 在各个方面中的任何方面的一些实施方案中,环状靶多核苷酸或环状多核苷酸包含无细胞多核苷酸,包括但不限于无细胞DNA或无细胞RNA(cfDNA或cfRNA)。在一些实施方案中,无细胞多核苷酸是循环肿瘤DNA或循环肿瘤RNA(ctDNA或ctRNA)。在一些实施方案中,无细胞多核苷酸包括胎儿DNA或胎儿RNA。在一些实施方案中,无细胞多核苷酸是来源于细胞但并非直接从细胞来源如组织样品获得的多核苷酸。可发源出无细胞多核苷酸的来源的非限制性实例是正常细胞和组织、异常细胞和组织(例如,癌细胞或组织)、胎儿细胞和组织,以及病原体。存在于非细胞来源中的无细胞多核苷酸可由细胞死亡(例如,凋亡或坏死)或细胞脱落产生。可以使用对无细胞多核苷酸的序列分析来表征无细胞DNA所源自的细胞或细胞群,诸如肿瘤细胞(例如,在癌症检测中)、胎儿细胞(例如在产前诊断中)、来自移植组织的细胞(例如,在移植失败的早期检测中)或病原体(例如,细菌或病毒)。

[0099] 本公开内容的实施方案可使用任何无细胞多核苷酸。无细胞多核苷酸可从受试者如任何动物或活生物体获得。受试者的非限制性实例是哺乳动物,诸如人、非人灵长类动物、啮齿动物如小鼠和大鼠、狗、猫、猪、羊、兔等。在一些实施方案中,受试者是健康的,因此从该受试者获得的无细胞多核苷酸可能不包含与疾病或病症相关的序列变体。在一些实施方案中,受试者疑似患有疾病或病症,因此从该受试者获得的无细胞多核苷酸可能包含与疾病或病症相关的序列变体。在一些实施方案中,受试者怀孕,因此从该受试者获得的无细胞多核苷酸包括胎儿多核苷酸。

[0100] 无细胞多核苷酸可从各种非细胞来源获得。可获得无细胞多核苷酸的非细胞来源的非限制性实例是血清、血浆、血液、汗液、唾液、尿液、粪便、精液、粘膜排泄物、脊髓液、羊水和淋巴液。用于收集可获得无细胞多核苷酸的非细胞来源的样品的各种方法是可用的。在一些实施方案中,从受试者获得可获得无细胞多核苷酸的非细胞来源的样品。在一些实施方案中,通过静脉穿刺获得样品。在一些实施方案中,通过抽吸获得样品。

[0101] 各种方法和商用试剂盒可用于从样品获得无细胞多核苷酸,如无细胞DNA。用于提取和分离无细胞多核苷酸(包括无细胞DNA)的方法和试剂盒的实例是苯酚/氯仿提取、苯酚/氯仿/异戊醇(PCI)-糖原提取、NaI(碘化钠)提取、胍树脂提取、带有载体RNA的QIAmp

DNA Blood Midi试剂盒、ChargeSwitch血清试剂盒、ZR血清DNA试剂盒、Qiagen QubitTM dsDNA HS Assay试剂盒、AgilentTM DNA 1000试剂盒、TruSeqTM测序文库制备以及Puregene DNA纯化系统血液试剂盒。

[0102] 无细胞多核苷酸(包括无细胞DNA)可以通过分隔步骤从体液中提取和分离,该分隔步骤中无细胞多核苷酸与体液的细胞和其他非可溶性组分相分离。分隔技术的实例是离心和过滤。在一些实施方案中,没有首先将细胞与无细胞多核苷酸相分隔,而是首先进行裂解。在一些实施方案中,通过选择性沉淀来分隔完整细胞的基因组DNA。无细胞多核苷酸(包括DNA)可保持可溶性,并可与不溶性基因组DNA相分离并被提取。根据一些程序,在添加缓冲液以及不同试剂盒特定的其他洗涤步骤后,可以使用异丙醇沉淀来沉淀DNA。可以使用进一步的清理步骤,如基于二氧化硅的柱来去除污染物或盐。通用步骤可针对特定应用进行优化。例如,可以在整个反应期间添加非特异性批量载体多核苷酸以优化该程序的某些方面,如产率。

[0103] 在本文公开的各个方面中的任何方面的一些实施方案中,环状靶多核苷酸或环状多核苷酸包含基因组DNA。在一些实施方案中,环状靶多核苷酸或环状多核苷酸来源于基因组DNA。基因组DNA可使用可用的各种方法和商用试剂盒(如Qiagen DNeasy Tissue试剂盒)从细胞样品获得。可使用先前在本文其他各处描述的任何提取、分离和纯化方法从样品获得并纯化基因组DNA。提取技术的其他非限制性实例包括:(1)有机萃取然后乙醇沉淀,例如,使用苯酚/氯仿有机试剂(Ausubel等人,1993),使用或不使用自动核酸提取器,例如可从Applied Biosystems(Foster city,Calif)获得的341DNA型提取器;(2)固定相吸附法(美国专利号5,234,809;Walsh等人,1991);和(3)盐诱导核酸沉淀法(Miller等人,1988),该沉淀方法一般被称作“盐析”法。核酸分离和/或纯化的另一个实例包括使用核酸能够特异性或非特异性结合的磁性颗粒,然后使用磁体分离珠子,并洗涤和从珠子中洗脱核酸(参见,例如,美国专利号5,705,628)。例如,可使用固相可逆固定(SPRI)珠(Agencourt AMPure XP)将核酸分离和纯化。在一些实施方案中,上述分离方法之前可先进行酶消化步骤以帮助从样品中去除不需要的蛋白质,例如用蛋白酶K或其他类似的蛋白酶进行消化。如果需要,可向裂解缓冲液中添加RNase抑制剂。对于特定的细胞或样品类型,可能需要在方案中增加蛋白质变性/消化步骤。纯化方法可以针对分离DNA、RNA或此两者。当DNA和RNA在提取程序过程中或之后被一起分离时,可使用进一步的步骤来与另一种分开地纯化一种或两者。也可生成提取的核酸的亚级分,例如,根据大小、序列或其他物理或化学特性进行纯化。除了初始核酸分离步骤,核酸的纯化还可以在所公开的方法的任意步骤之后进行,例如用于去除过量的或不需要的试剂、反应物或产物。多种用来确定样品中的核酸量和/或核酸纯度的方法是可用的,例如通过吸光度(例如,在260nm、280nm处的光吸收,和其比值)和标记物的检测(例如,荧光染料和嵌入剂,例如SYBR绿、SYBR蓝、DAPI、碘化丙啶、Hoechst染色剂、SYBR金、溴化乙锭)。

[0104] 在一些实施方案中,环状靶多核苷酸或环状多核苷酸包含片段化无细胞DNA或片段化基因组DNA。多种方法可用于对多核苷酸进行片段化,包括但不限于化学方法、酶促方法和机械方法,如超声处理、剪切和与限制酶接触。在一些实施方案中,无细胞DNA片段的长度大致均匀。在一些实施方案中,无细胞DNA片段的长度并非大致均匀。在一些实施方案中,无细胞DNA片段具有约50至约1000个核苷酸的平均长度。在一些实施方案中,无细胞DNA片

段具有约50至约500个核苷酸的平均长度。在一些实施方案中,无细胞DNA片段具有约50至约250个核苷酸的平均长度。在一些实施方案中,无细胞DNA片段具有约50至约200个核苷酸的平均长度。在一些实施方案中,无细胞DNA片段具有约50至约100个核苷酸的平均长度。在一些实施方案中,无细胞DNA片段具有约40至约1000个核苷酸的平均长度。在一些实施方案中,无细胞DNA片段具有约40至约500个核苷酸的平均长度。在一些实施方案中,无细胞DNA片段具有约40至约250个核苷酸的平均长度。在一些实施方案中,无细胞DNA片段具有约40至约200个核苷酸的平均长度。在一些实施方案中,无细胞DNA片段具有约40至约100个核苷酸的平均长度。在一些实施方案中,基因组DNA被片段化为较短长度的多核苷酸。在一些实施方案中,基因组DNA片段的长度大致相同。在一些实施方案中,基因组DNA片段的长度几乎不同。在一些实施方案中,基因组DNA片段具有约50至约100个核苷酸的平均长度。在一些实施方案中,基因组DNA片段具有约50至250个核苷酸的平均长度。在一些实施方案中,基因组DNA片段具有约50至500个核苷酸的平均长度。在一些实施方案中,基因组DNA片段具有约50至750个核苷酸的平均长度。在一些实施方案中,基因组DNA片段具有约100至1000个核苷酸的平均长度。

[0105] 可通过各种方法由线性靶多核苷酸形成环状靶多核苷酸或环状多核苷酸。在一些实施方案中,通过末端联接将单个线性靶多核苷酸环化。在一些实施方案中,将第一线性靶多核苷酸联接至第二线性靶多核苷酸,然后将第一靶多核苷酸的未联接端联接至第二线性靶多核苷酸的未联接端以形成包含第一和第二靶多核苷酸的环状靶多核苷酸或环状多核苷酸。待环化的多核苷酸可以是单链或双链的。当期望单链环时,多核苷酸可以是如最初分离的单链多核苷酸,或者可被处理以使该多核苷酸成为单链(例如,通过变性)。在一些实施方案中,用于使多核苷酸环化的方法涉及酶,如使用连接酶(例如, RNA连接酶或DNA连接酶)。可用于将线性靶多核苷酸连接成环状靶多核苷酸或环状多核苷酸的酶的非限制性实例是ATP依赖性双链多核苷酸连接酶、NAD⁺依赖性DNA或RNA连接酶和单链多核苷酸连接酶。连接酶的非限制性实例是CircLigase I和CircLigase II (Epicentre, Madison, WI)、大肠杆菌DNA连接酶、丝状栖热菌DNA连接酶、Tth DNA连接酶、水管致黑栖热菌DNA连接酶(I型和II型)、T3DNA连接酶、T4DNA连接酶、T4RNA连接酶、T7DNA连接酶、Taq连接酶、Ampligase (Epicentre[®] Technologies Corp.)、VanC-型连接酶、9°N DNA连接酶、Tsp DNA连接酶、DNA连接酶I型、DNA连接酶III型、DNA连接酶IV型、Sso7-T3DNA连接酶、Sso7-T4DNA连接酶、Sso7-T7DNA连接酶、Sso7-Taq DNA连接酶、Sso7-大肠杆菌DNA连接酶、Sso7-Ampligase DNA连接酶,以及热稳定连接酶。连接酶可以是野生型、突变同工型和基因工程化变体。连接反应可含有缓冲组分、小分子连接增强剂和其他反应组分。在一些实施方案中,调节多核苷酸和酶的浓度以促进分子间连接而不是分子内连接。在一些实施方案中,调节反应温度和反应时间或反应的时间长度。反应温度和时间也可调节。在一些实施方案中,使用60°C来促进分子内环的形成。在一些实施方案中,反应时间为12-16小时。反应条件可以是所选择的酶的制造商所规定的条件。在一些实施方案中,联接多核苷酸的末端以形成环状多核苷酸(直接联接至其自身或联接至一个或多个其他多核苷酸例如,包含两个靶多核苷酸的环状靶多核苷酸或环状多核苷酸)产生具有连接序列的接合。在一些实施方案中,可以包括外切核酸酶步骤以在环化反应后消化任何未连接的核酸。也就是说,闭合环不含游离5'或3'端,因此引入5'或3'外切核酸酶不会消化闭合环但会消化未连接的组分。这尤其可用于多路系统

中。

[0106] 环化之后,反应产物可在扩增或测序之前进行纯化以提高可参与后续步骤的环化多核苷酸的相对浓度或纯度(例如,通过环状多核苷酸的分离或反应中一种或多种其他分子的去除)。例如,可处理环化反应或其组分以去除单链(未环化的)多核苷酸,例如通过外切核酸酶处理。作为进一步的实例,环化反应或其部分可进行大小排阻色谱法,借此保留及丢弃小试剂,或在单独的体积中保留并释放环化产物。多种用于清理连接反应的试剂盒是可用的,例如由Zymo Research制造的Zymo 寡核苷酸纯化试剂盒所提供的试剂盒。在一些实施方案中,纯化包括用于去除或降解在环化反应中使用的连接酶和/或将环化多核苷酸从该连接酶中纯化的处理。在一些实施方案中,用于降解连接酶的处理包括用蛋白酶如蛋白酶K进行的处理。蛋白酶K处理可遵循制造商的方案或标准方案(例如,如Sambrook和Green, Molecular Cloning: A Laboratory Manual, 第4版(2012)所提供的)。蛋白酶处理之后还可进行提取和沉淀。在一个实例中,环化多核苷酸如下纯化:在0.1% SDS和20mM EDTA的存在下进行蛋白酶K(Qiagen)处理,用1:1苯酚/氯仿和氯仿抽提,并用乙醇或异丙醇沉淀。在一些实施方案中,沉淀在乙醇中进行。

[0107] 本公开内容的一些实施方案包括引物延伸和扩增反应,如生成多联体、生成多种延伸产物,以及对多种延伸产物进行扩增中的一种或多种。引物延伸反应可涉及温度变化(热循环)或恒温(等温)。在一些实施方案中,引物延伸反应包括聚合酶链反应(PCR)。PCR涉及通过多个阶段的变性、引物对与相反链退火,以及用于使靶序列拷贝数指数增加的引物延伸而进行的循环,这些阶段中的至少一些通常在不同的反应温度下发生。PCR扩增技术的非限制实例是定量PCR(qPCR或实时PCR)、逆转录PCR(RT-PCR)、数字PCR(dPCR或dePCR)、靶标特异性PCR以及定量逆转录PCR(qRT-PCR)。可用于PCR的聚合酶的实例为热稳定聚合酶,包括但不限于嗜热栖热菌HB8;突变Thermus oshimai;水管致黑栖热菌;嗜热栖热菌1B21;嗜热栖热菌GK24;水生栖热菌聚合酶(**AmpliTaq®** FS或Taq(G46D;F667Y),Taq(G46D;F667Y;E6811)和Taq(G46D;F667Y;T664N;R660G));激烈火球菌聚合酶;Thermococcus gorgonarius聚合酶;火球菌属物种GB-D聚合酶;栖热球菌属(菌株9°N-7)聚合酶;嗜热脂肪芽孢杆菌聚合酶;Tsp聚合酶;ThermalAce™聚合酶(Invitrogen);黄栖热菌聚合酶;Thermus litoralis聚合酶;栖热菌属Z05聚合酶;δZ05聚合酶(例如,δZ05Gold DNA聚合酶);以及其突变体、变体其衍生物。可用于PCR的聚合酶的其他实例是非热稳定聚合酶,包括但不限于DNA聚合酶I;突变体DNA聚合酶I,包括但不限于Klenow片段和Klenow片段(3'至5'外切核酸酶(-));T4DNA聚合酶;突变体T4DNA聚合酶;T7DNA聚合酶;突变体T7DNA聚合酶;phi29DNA聚合酶;以及突变体phi29DNA聚合酶。在一些实施方案中,使用热启动聚合酶。热启动聚合酶是需要热激活的DNA聚合酶的修饰形式。这样的聚合酶可用于例如进一步提高敏感度、特异性和产率;并且/或者进一步改善低拷贝靶标扩增。通常,热启动酶以非活性状态提供。热活化后,释放修饰物或改性剂,从而生成活性酶。许多热启动聚合酶可从各种商业来源获得,如Applied Biosystems;Bio-Rad;eEnzyme LLC;Eppendorf North America;Finnzymes Oy;GeneChoice, Inc.;Invitrogen;Jena Bioscience GmbH;MIDSCI;Minerva Biolabs GmbH;New England Biolabs;Novagen;Promega;QIAGEN;Roche Applied Science;Sigma-Aldrich;Stratagene;Takara Mirus Bio;USB Corp.;Yorkshire Bioscience Ltd;等等。

[0108] 在一些实施方案中,引物延伸和扩增反应包括等温反应。等温扩增技术的非限制

性实例是连接酶链反应 (LCR) (例如, 美国专利号5,494,810和5,830,711); 转录介导的扩增 (TMA) (例如, 美国专利号5,399,491-5,888,779-5,705,365-5,710,029); 基于核酸序列的扩增 (NASBA) (例如, Malek等人, 美国专利号5,130,238); 信号介导的RNA扩增技术 (SMART) (例如, Wharam等人, Nucleic Acids Res. 2001, 29, e54); 链置换扩增 (SDA) (例如, 美国专利号5,455,166); 嗜热SDA (Spargo等人, Mol Cell Probes 1996, 10:247-256; 欧洲专利号0684315); 滚环扩增 (RCA) (例如, Lizardi, "Rolling Circle Replication Reporter Systems," 美国专利号5,854,033); 环介导的DNA等温扩增 (LAMP) (例如, Notomi等人, "Process for Synthesizing Nucleic Acid," 美国专利号6,410,278); 解旋酶依赖性扩增 (HDA) (例如, 美国专利申请US 20040058378); 单引物等温扩增 (SPIA) (例如, WO2001020035 和美国专利号6,251,639); 以及环状解旋酶依赖性扩增 (cHDA) (例如, 美国专利申请US.10/594,095)。

[0109] 在一些实施方案中, 引物延伸反应通过具有链置换活性的聚合酶实现如对于RCA。在一些实施方案中, 等温扩增包括滚环扩增 (RCA)。RCA反应混合物可包含一种或多种引物、具有链置换活性的聚合酶, 和dNTP。链置换是指在合成期间置换下游DNA的能力。具有链置换活性的聚合酶可具有不同程度的链置换活性。在一些实施方案中, 聚合酶可具有弱链置换活性或没有链置换活性。在一些实施方案中, 聚合酶可具有强链置换活性。在一些实施方案中, 具有链置换活性的聚合酶可在不同反应温度下具有不同水平的链置换活性。在一些实施方案中, 聚合酶可在中等温度例如20°C-37°C下显示链置换活性。在一些实施方案中, 聚合酶可在升高的温度例如65°C下显示链置换活性。可调节反应温度以有利于具有链置换活性的聚合酶的活性水平。在一些实施方案中, 反应温度为至少20°C、25°C、30°C、35°C、40°C、45°C、50°C、55°C、60°C、65°C、70°C、75°C、80°C、85°C或90°C。在一些实施方案中, 反应温度为20°C至80°C。在一些实施方案中, 反应温度为20°C至70°C。在一些实施方案中, 反应温度为20°C至60°C。在一些实施方案中, 反应温度为20°C至50°C。在一些实施方案中, 可以在不同阶段中循环通过不同反应温度以提高或降低聚合酶的链置换活性。具有链置换活性的聚合酶的非限制性实例是Bst DNA聚合酶, 大片段; Bsu DNA聚合酶, 大片段; Deep Vent_RTM DNA聚合酶; Deep Vent_RTM (exo-) DNA聚合酶; Klenow片段 (3' - 5' exo-); DNA聚合酶I, 大片段; M-MuLV逆转录酶; phi29DNA聚合酶; PyroPhage 3173聚合酶; Vent_R[®] DNA聚合酶; 以及 Vent_R[®] (exo-) DNA聚合酶。

[0110] 作为扩增反应(包括热循环方法、等温方法和这些的组合)的产物生成的多联体可包含两个或更多个拷贝的靶多核苷酸。多联体可包含约2、3、4、5、6、7、8、9、10个或更多个拷贝的靶多核苷酸。在一些实施方案中, 多联体作为引物延伸反应的产物从多种靶多核苷酸生成, 其中所述多种靶多核苷酸的组分的长度不均匀且包含多个序列。

[0111] 在本公开内容的各个方面中的任何方面的一些实施方案中, 引物可包含一个或多个部分。例如, 引物可包含一个或多个扩增引物退火序列或其互补体; 一个或多个测序引物退火序列或其互补体; 一个或多个条形码序列; 一个或多个在多种不同引物之间共有的共同序列; 一个或多个限制酶识别位点; 一个或多个探针结合位点或测序衔接子(例如, 用于衔接至测序平台, 如用于大规模平行测序的流动池); 一个或多个随机或接近随机的序列(例如, 从一个或多个位置处的一组两个或更多个不同核苷酸中随机选择的一个或多个核

昔酸,其中在一个或多个位置处选择的不同核昔酸中的每一个均表现在包含随机序列的引物池中);以及它们的组合。在一些实施方案中,引物如第三引物包含测序衔接子元件(在本文中也被称为衔接子),该元件通常指在多核昔酸的5'和/或3'端掺入以促进多核昔酸测序反应中的一个或多个步骤的寡核昔酸。在一些实施方案中,使用测序衔接子将包含测序衔接子的多核昔酸结合至流动池以供下一代测序。下一代测序方法的非限制性实例是单分子实时测序、离子半导体测序、焦磷酸测序、合成测序、连接测序和链终止。用于流动池附接的测序衔接子可包括与下一代测序系统(例如,454测序、Ion Torrent Proton或PGM和Illumina X10)兼容的任何合适序列。用于下一代测序方法的测序衔接子的非限制性实例包括适合与Illumina测序系统一起使用的P5和P7衔接子;TruSeq通用衔接子;和TruSeq索引衔接子。在一些实施方案中,可使用测序衔接子来例如通过扩增如聚合酶链反应(PCR)来富集包含衔接子序列的多核昔酸。测序衔接子可进一步包含条形码序列和/或样品索引序列。

[0112] 在某些其他实施方案中,引物如第三引物包含条形码序列。条形码序列是指允许鉴定与条形码相关联的多核昔酸的一些特征的已知核酸序列。条形码各自可具有5至35个核昔酸、6至30个核昔酸或8至20个核昔酸的长度。在一些实施方案中,条形码的长度为至少3、4、5、6、7、8、9、10、11、12、13、14或15个核昔酸。在一些实施方案中,条形码的长度小于6个核昔酸。在一些实施方案中,与一些靶多核昔酸相关联的条形码可具有不同于与其他靶多核昔酸相关联的条形码的长度。一组内的条形码的解链温度可以在彼此的±10°C内、彼此的±5°C内或彼此的±2°C内。条形码可以是最小交叉杂交组的成员。例如,这样的组中的每个成员的核昔酸序列可以与该组中的每个其他成员的核昔酸序列明显不同,以致于在中等或严格杂交条件下,没有成员可以与任何其他成员的互补体形成稳定的双链体。最小交叉杂交组的每个成员的核昔酸序列可以与每个其他成员的核昔酸序列有至少两个核昔酸不同。一些条形码技术描述于Winzeler等人,(1999) *Science* 285:901;Brenner (2000) *Genome Biol.* 1:1;Kumar等人,(2001) *Nature Rev.* 2:302;Giaever等人,(2004) *Proc.Natl.Acad.Sci.USA* 101:793;Eason等人,(2004) *Proc.Natl.Acad.Sci.USA* 101:1104;以及Brenner (2004) *Genome Biol.* 5:240,其中的每一个均通过引用以其全部内容并入本文。

[0113] 本公开内容的某些实施方案包括对多个扩增子进行测序。多种测序方法可用于对多个扩增子进行测序。在一些实施方案中,使用高通量测序方法。可使用的测序方法的非限制性实例包括Illumina制造的测序系统(诸如 **HiSeq®** 和 **MiSeq®** 的测序系统)、Life Technologies制造的测序系统(Ion Torrent®、SOLiD®等)、Roche的454Life Sciences系统、Pacific Biosciences系统等。在一些实施方案中,测序包括使用 **HiSeq®** 和 **MiSeq®** 系统产生长度约为或多于约50、75、100、125、150、175、200、250、300个或更多个核昔酸的读取。在一些实施方案中,测序包括合成测序过程,其中随着单独的核昔酸被添加至生长的引物延伸产物上,该核昔酸被迭代地鉴定。焦磷酸测序是合成测序法的一个实例,其通过分析所产生的合成混合物中测序反应副产物即焦磷酸的存在而鉴定核昔酸的掺入。特别是,引物/模板/聚合酶复合体与一种类型的核昔酸接触。如果该核昔酸掺入,则聚合反应切割三磷酸链的α和β磷酸之间的三磷酸核昔,从而释放焦磷酸。然后使用化学荧光

酶报告系统鉴定所释放的焦磷酸的存在,该系统将含有AMP的焦磷酸转化为ATP,之后用萤光素酶测量ATP以生成可测量的光信号。当检测到光时,碱基已掺入,当未检测到光时,碱基未掺入。在适当的洗涤步骤后,使各种碱基周期性地与该复合体接触,以连续地鉴定模板序列中的后续碱基。参见,例如,美国专利号6,210,891。

[0114] 在一些实施方案中,对扩增子进行测序以相对于参考序列或在无突变的背景下检测序列变体,例如,倒位、缺失、重复、易位和罕见体细胞突变。在一些实施方案中,该序列变体与疾病相关。在一些实施方案中,该序列变体与疾病不相关。通常,与疾病或性状的关联性具有统计学证据、生物学证据和/或功能证据的序列变体被称为“因果遗传变体”。单个因果遗传变体可与多于一种疾病或性状相关。在一些情况下,因果遗传变体可与孟德尔性状、非孟德尔性状或与两者相关。因果遗传变体可表现为多核苷酸变异,如1、2、3、4、5、6、7、8、9、10、20、50个或更多个序列差异(如包含因果遗传变体的多核苷酸与在相同的相对基因组位置处缺乏因果遗传变体的多核苷酸之间的序列差异)。因果遗传变体类型的非限制性实例包括单核苷酸多态性(SNP)、缺失/插入多态性(DIP)、拷贝数变体(CNV)、短串联重复(STR)、限制性片段长度多态性(RFLP)、简单序列重复(SSR)、可变数目串联重复(VNTR)、随机扩增多态DNA(RAPD)、扩增片段长度多态性(AFLP)、逆转录转座子间扩增多态性(IRAP)、长散在元件和散在元件(LINE/SINE)、长串联重复(LTR)、移动元件、逆转录转座子微卫星扩增多态性、基于逆转录转座子的插入多态性、序列特异性扩增多态性和可遗传表观遗传修饰(例如,DNA甲基化)。因果遗传变体也可能是一组密切相关的因果遗传变体。一些因果遗传变体可能施加与RNA多核苷酸中的序列变异一样的影响。在这个水平下,一些因果遗传变体还可由存在或不存在RNA多核苷酸物质来指示。另外,一些因果遗传变体导致蛋白质多肽的序列变异。已经报道了许多因果遗传变体。为SNP的因果遗传变体的实例是导致镰状细胞性贫血的血红蛋白Hb S变体。为DIP的因果遗传变体的实例是导致囊性纤维化的CFTR基因的δ508突变。为CNV的因果遗传变体的实例是21三体,其导致唐氏综合症。为STR的因果遗传变体的实例是导致亨廷顿病的串联重复。W02014015084中描述了因果遗传变体的其他非限制性实例。W02015089333中描述了用于鉴定罕见序列变体的方法的其他非限制性实例。

[0115] 在本公开内容的各个方面中的任何方面的某些实施方案中,在测序之前对扩增子进行纯化。可通过各种方法对扩增子进行纯化。可将扩增子纯化以去除过量或不需要的试剂、反应物或产物。扩增子可进一步通过大小、序列或其他物理或化学特性进行纯化。在一些实施方案中,扩增子可经受大小排阻色谱法,由此截留并丢弃仅包含一个拷贝的靶多核苷酸的扩增子和/或较小试剂(例如,引物),或截留包含两个或更多个拷贝的靶多核苷酸的扩增子并释放在单独的体积中。在一些实施方案中,扩增子可以经受从凝胶上切下片段和凝胶过滤(例如,用于富集长度大于约300、400、500个或更多个核苷酸的片段);以及用于通过微调结合缓冲液浓度进行大小选择的SPRI珠(Agencourt AMPure XP)。例如,可以在与DNA片段混合过程中使用0.6x结合缓冲液来优先结合大于约500个碱基对(bp)的DNA片段。在一些实施方案中,尤其是在已用B2B引物进行了扩增时,对扩增产物进行处理,以根据大小过滤所得到的扩增子,从而减少和/或去除包含多联体的混合物中的单体数目。这可以使用如本文其他各处描述的任何纯化技术来完成。

[0116] 本文提供的公开内容的实施方案可用于富集包含与一种或多种癌症相关的各种序列变体的扩增子。可用于本公开内容的方法的具有肿瘤学意义的合适靶序列包括但不限

于TP53基因、ALK基因、KRAS基因、PIK3CA基因、BRAF基因、EGFR基因和KIT基因的改变。可特异性扩增和/或针对序列变体进行特异性分析的靶序列可以是癌症相关基因的全部部分。在一些实施方案中,在TP53基因中鉴定出一个或多个序列变体。TP53是人类癌症中最常突变的基因之一,例如,在45%的卵巢癌、43%的大肠癌和42%的上呼吸消化道癌症中均发现TP53突变(参见,例如M.Olivier等人,TP53Mutations in Human Cancers:Origins,Consequences, and Clinical Use.Cold Spring Harb Perspect Biol.2010年1月;2(1))。TP53突变状态的表征可帮助临床诊断,提供预后价值并影响对癌症患者的治疗。例如,TP53突变可用作来源于胶质细胞的CNS肿瘤患者的预后不良的预测因子,并且是慢性淋巴细胞白血病患者中快速疾病进展的预测因子(参见,例如McLendon RE等人,Cancer.2005年10月15日;104(8):1693-9;Dicker F等人,Leukemia.2009年1月;23(1):117-24)。序列变异可发生在基因的任何地方。因此,在本文中可能评估全部或部分TP53基因。换言之,如本文其他各处所述的,当使用靶标特异性组分(例如,靶特异性引物)时,可使用多个TP53特异性序列,例如以便扩增和检测跨该基因的片段,而不仅仅是可用于选定靶标的一个或多个选定的子序列(如突变“热点”)。或者,可以设计靶标特异性引物,使其在一个或多个选定子序列(如也被术语“热点”涵盖在内的与一类受试者之间突变率增加相关的核苷酸或核苷酸区域)的上游或下游杂交。可以设计跨这种子序列的标准引物,并且/或者可以设计在这种子序列的上游或下游杂交的B2B引物。

[0117] 在一些实施方案中,在ALK基因的全部或部分中鉴定一个或多个序列变体。已报道在多达7%的肺肿瘤中存在ALK融合,其中一些与EGFR酪氨酸激酶抑制剂(TKI)抗性相关(参见,例如Shaw等人,J Clin Oncol.2009年9月10日;27(26):4247-4253)。截至2013年,已经在对ALK酪氨酸激酶抑制剂(TKI)具有继发耐药性的患者中发现了跨整个ALK酪氨酸激酶结构域的几个不同的点突变(Katayama R2012Sci Transl Med.2012年2月8日;4(120))。因此,ALK基因中的突变检测可用于辅助癌症疗法决策。

[0118] 在一些实施方案中,在KRAS基因的全部或部分中鉴定一个或多个序列变体。已报道约15-25%的肺腺癌患者和40%的结直肠癌患者携带肿瘤相关的KRAS突变(参见,例如Neuman 2009,Pathol Res Pract.2009;205(12):858-62)。大多数突变位于KRAS基因的密码子12、13和61处。这些突变激活KRAS信号传导途径,触发肿瘤细胞的生长和增殖。一些研究表明携带KRAS突变的肿瘤患者不大可能从单独的抗-EGFR抗体疗法或与化疗联合的抗-EGFR抗体疗法获益(参见,例如Amado等人,2008J Clin Oncol.2008年4月1日;26(10):1626-34,Bokemeyer等人,2009J Clin Oncol.2009年2月10日;27(5):663-71)。可用于鉴定序列变异的序列变异的一个特定“热点”位于该基因的位置35。KRAS序列变体的鉴定可用于治疗选择,例如用于患有结直肠癌的受试者的治疗选择。

[0119] 在一些实施方案中,在PIK3CA基因的全部或部分中鉴定一个或多个序列变体。在各种类型癌症,例如在10-30%的结直肠癌中经常发现PIK3CA的体细胞突变(参见,例如Samuels等人,2004Science.2004年4月23日;304(5670):554)。这些突变最常位于外显子9(螺旋结构域)和外显子20(激酶结构域)内的两个“热点”区域内,可专门靶向这两个“热点”区域进行扩增和/或分析以供序列变体检测。还可以专门靶向位置3140。

[0120] 在一些实施方案中,在BRAF基因的全部或部分中鉴定一个或多个序列变体。已报道全部恶性黑素瘤中近50%都携带BRAF的体细胞突变(参见,例如Maldonado等人,J Natl

Cancer Inst. 2003年12月17日; 95 (24) :1878-90。在所有黑素瘤亚型中均发现BRAF突变,但在来源于没有慢性日光诱导损伤的皮肤的黑素瘤中最常见。黑素瘤中最常见的BRAF突变是错义突变V600E,其将位置600处的缬氨酸置换为谷氨酰胺。BRAF V600E突变与BRAF抑制剂治疗的临床益处相关。BRAF突变的检测可用于黑素瘤治疗选择和针对靶向疗法的抗性研究。

[0121] 在一些实施方案中,在EGFR基因的全部或部分中鉴定一个或多个序列变体。EGFR突变经常与非小细胞肺癌相关(在美国约为10%,在东亚为35%;参见例如Pao等人,Proc Natl Acad Sci US A. 2004年9月7日;101 (36) :13306-11)。这些突变通常发生在EGFR外显子18-21内,并且通常是杂合的。这些突变中约90%是外显子19缺失或外显子21L858R点突变。

[0122] 在一些实施方案中,在KIT基因的全部或部分中鉴定一个或多个序列变体。已报道近85%的胃肠道间质瘤(GIST)携带KIT突变(参见,例如Heinrich等人,2003 J Clin Oncol. 2003年12月1日;21 (23) :4342-9)。大多数KIT突变见于在近膜域(外显子11,70%)、细胞外二聚化基序(外显子9,10-15%)、酪氨酸激酶I(TKI)结构域(外显子13,1-3%)和酪氨酸激酶2(TK2)结构域和活化环(外显子17,1-3%)中。通常在靶向疗法伊马替尼后并且患者对该疗法发展出耐药性后鉴定继发性KIT突变。

[0123] 与癌症相关的基因(可根据本文所述的方法来分析其全部或一部分的序列变体)的其他非限制性实例包括但不限于PTEN;ATM;ATR;EGFR;ERBB2;ERBB3;ERBB4;Notch1;Notch2;Notch3;Notch4;AKT;AKT2;AKT3;HIF;HIF1a;HIF3a;Met;HRG;Bc12;PPAR α ;PPAR γ ;WT1(Wilms肿瘤);FGF受体家族成员(5个成员:1、2、3、4、5);CDKN2a;APC;RB(成视网膜细胞瘤);MEN1;VHL;BRCA1;BRCA2;AR;(雄激素受体);TSG101;IGF;IGF受体;Igf1(4个变体);Igf2(3个变体);Igf 1受体;Igf 2受体;Bax;Bc12;半胱天冬酶家族(9个成员:1、2、3、4、6、7、8、9、12);Kras;以及Apc。本文其他各处提供了其他实例。可基于根据本文公开的方法对一个或多个序列变体进行判别而诊断的癌症的实例包括但不限于棘皮瘤、腺泡细胞癌、听神经瘤、肢端着色斑性黑素瘤、顶端螺旋瘤、急性嗜酸细胞白血病、急性淋巴母细胞性白血病、急性巨核母细胞性白血病、急性单核细胞性白血病、成熟的急性成髓细胞性白血病、急性髓样树突细胞白血病、急性髓样白血病、急性早幼粒细胞性白血病、釉质瘤、腺瘤、腺样囊性癌、腺瘤、牙源性腺瘤样瘤、肾上腺皮质癌、成人T-细胞白血病、侵袭性NK-细胞白血病、AIDS相关癌症、AIDS相关淋巴瘤、软组织腺泡状肉瘤、成釉细胞纤维瘤、肛门癌、间变性大细胞淋巴瘤、甲状腺未分化癌、血管免疫母细胞性T细胞淋巴瘤、血管肌脂瘤、血管肉瘤、阑尾癌、星形细胞瘤、非典型性畸胎样横纹肌样瘤、基底细胞癌、基底细胞样癌、B-细胞白血病、B细胞淋巴瘤、Bellini管癌、胆道癌、膀胱癌、母细胞瘤、骨癌、骨瘤、脑干胶质瘤、脑瘤、乳腺癌、Brenner瘤、支气管瘤、细支气管肺泡癌、棕色瘤、伯基特淋巴瘤、原发灶不明的癌症、类癌瘤、癌、原位癌、阴茎癌、原发灶不明癌、癌肉瘤、Castleman病、中枢神经系统胚胎瘤、小脑星形细胞瘤、大脑星形细胞瘤、宫颈癌、胆管细胞癌、软骨瘤、软骨肉瘤、脊索瘤、绒毛膜癌、脉络丛乳头状瘤、慢性淋巴细胞白血病、慢性单核细胞性白血病、慢性髓性白血病、慢性骨髓增生性疾病、慢性嗜中性粒细胞白血病、明细胞肿瘤、结肠癌、结直肠癌、颅咽管瘤、皮肤T细胞淋巴瘤、Degas病、隆凸性皮肤纤维肉瘤、皮样囊肿、结缔组织增生性小圆细胞瘤、弥漫性大B细胞淋巴瘤、胚胎发育不良神经上皮瘤、胚胎癌、内胚窦瘤、子宫内膜癌、子宫内膜子宫

癌、子宫内膜样肿瘤、肠病相关的T细胞淋巴瘤、室管膜母细胞瘤、室管膜瘤、上皮样肉瘤、红白血病、食管癌、鼻腔神经胶质瘤、尤因家族肿瘤、尤因家族肉瘤、尤因肉瘤、颅外生殖细胞瘤、性腺外生殖细胞瘤、肝外胆管癌、非乳腺性佩吉特病、输卵管癌、胎中胎、纤维瘤、纤维肉瘤、滤泡性淋巴瘤、滤泡性甲状腺癌、胆囊癌、神经节胶质瘤、神经节瘤、胃癌、胃淋巴瘤、胃肠癌、胃肠类癌瘤、胃肠间质瘤、生殖细胞肿瘤、生殖细胞瘤、妊娠性绒毛膜癌、妊娠滋养细胞瘤、骨巨细胞瘤、多形性胶质母细胞瘤、胶质瘤、大脑胶质瘤病、血管球瘤、胰高血糖素瘤、成性腺细胞瘤、粒层细胞瘤、毛细胞白血病、头颈癌、心脏癌、血管母细胞瘤、血管外皮细胞瘤、血管肉瘤、恶性血液肿瘤、肝细胞癌、肝脾T细胞淋巴瘤、遗传性乳腺癌-卵巢癌综合征、霍奇金淋巴瘤、霍奇金氏淋巴瘤、下咽癌、下丘脑胶质瘤、炎性乳腺癌、眼内黑素瘤、胰岛细胞癌、胰岛细胞瘤、幼年型粒单核细胞白血病、卡波西肉瘤、卡波西氏肉瘤、肾癌、Klatskin瘤、Krukenberg瘤、喉癌、恶性雀斑样痣黑素瘤、白血病、唇癌和口腔癌、脂肪肉瘤、肺癌、黄体瘤、淋巴管瘤、淋巴管肉瘤、淋巴上皮瘤、淋巴样白血病、淋巴瘤、巨球蛋白血症、恶性纤维组织细胞瘤、恶性纤维组织细胞瘤、骨恶性纤维组织细胞瘤、恶性胶质瘤、恶性间皮瘤、恶性外周神经鞘瘤、恶性棒状瘤、恶性特里顿瘤、MALT淋巴瘤、套细胞淋巴瘤、肥大细胞淋巴瘤、纵隔生殖细胞瘤、纵隔肿瘤、甲状腺髓样癌、髓母细胞瘤、髓上皮瘤、黑素瘤、脑膜瘤、Merkel细胞癌、间皮瘤、原发灶隐匿转移性鳞状颈癌、转移性尿路上皮癌、混合Mullerian瘤、单核细胞性白血病、口腔癌、粘液瘤、多发性内分泌瘤综合征、多发性骨髓瘤、蕈样肉芽肿病、骨髓增生异常疾病、骨髓增生异常综合征、髓样白血病、髓样肉瘤、骨髓增生性疾病、粘液瘤、鼻腔癌、鼻咽癌、鼻咽癌、赘生物、神经鞘瘤、神经母细胞瘤、神经纤维瘤、神经瘤、结节型黑素瘤、非霍奇金淋巴瘤、非黑素瘤皮肤癌、非小细胞肺癌、眼肿瘤、少突星形细胞瘤、少突神经胶质瘤、嗜酸粒细胞腺瘤、视神经鞘脑膜瘤、口癌、口咽癌、骨肉瘤、卵巢癌、卵巢上皮癌、卵巢生殖细胞瘤、卵巢低恶性潜能肿瘤、乳腺Paget病、Pancoast瘤、胰腺癌、乳头状甲状腺癌、乳头状瘤病、副神经节瘤、副鼻窦癌、甲状腺旁腺癌、阴茎癌、血管周上皮样细胞瘤、咽癌、嗜铬细胞瘤、中度分化的松果体实质瘤、成松果体细胞瘤、垂体细胞瘤、垂体腺瘤、垂体瘤、浆细胞瘤、胸膜肺母细胞瘤、多胚瘤、前体T成淋巴细胞淋巴瘤、原发性中枢神经系统淋巴瘤、原发性渗出性淋巴瘤、原发性肝细胞癌、原发性肝癌、原发性腹膜癌、原始神经外胚层瘤、前列腺癌、腹膜假粘液瘤、直肠癌、肾细胞癌、涉及染色体15上NUT基因的呼吸道癌、视网膜母细胞瘤、横纹肌瘤、横纹肌肉瘤、Richter转化、骶尾部畸胎瘤、唾腺癌、肉瘤、施万鞘瘤、神经鞘瘤病(Schwannomatosis)、皮脂腺癌、继发性肿瘤、精原细胞瘤、浆液性肿瘤、Sertoli-Leydig细胞瘤、性索间质瘤、Sezary综合征、Signet细胞癌、皮肤癌、小蓝圆细胞瘤、小细胞癌、小细胞肺癌、小细胞淋巴瘤、小肠癌、软组织肉瘤、生长抑素瘤、煤烟疣、脊髓瘤、脊髓肿瘤、脾边缘区淋巴瘤、鳞状细胞癌、胃癌、表浅扩散性黑素瘤、幕上原始神经外胚层瘤、表面上皮-间质瘤、滑膜肉瘤、T细胞急性淋巴母细胞性白血病、T细胞大颗粒淋巴细胞白血病、T细胞白血病、T细胞淋巴瘤、T细胞幼淋巴细胞白血病、畸胎瘤、晚期淋巴癌、睾丸癌、泡膜细胞瘤、喉癌、胸腺癌、胸腺瘤、甲状腺癌、肾盂和输尿管移行细胞癌、移行细胞癌、膀胱癌、尿道癌、泌尿生殖系统肿瘤、子宫肉瘤、葡萄膜黑素瘤、阴道癌、Verner Morrison综合征、疣状癌、视通路胶质瘤、外阴癌、瓦尔登斯特伦巨球蛋白血症、Warthin瘤、维尔姆斯(Wilms)瘤，以及它们的组合。

[0124] 表2中提供了与癌症相关的基因(可根据本文所述的方法来分析其全部或一部

分(例如,启动子区域、内含子、外显子等)的序列变体)的其他非限制性实例。

[0125] 表2

基因	描述
ABCC6	ATP-结合盒, 亚家族 C (CFTR/MRP), 成员 6
ABI1	abl-相互作用物 1
ABL1	c-abl 癌基因 1, 非受体酪氨酸激酶
ABL2	v-abl Abelson 鼠白血病病毒癌基因同源物 2
ACSL3	酰基辅酶 A 合成酶长链家族成员 3
ACSL6	酰基辅酶 A 合成酶长链家族成员 6
AFF1	AF4/FMR2 家族, 成员 1
AFF3	AF4/FMR2 家族, 成员 3
AFF4	AF4/FMR2 家族, 成员 4
AIP	芳烃受体相互作用蛋白
AKAP9	激酶 (PRKA) 锚蛋白 (yotiao) 9
AKT1	v-akt 鼠胸腺瘤病毒癌基因同源物 1
AKT2	v-akt 鼠胸腺瘤病毒癌基因同源物 2
AKT3	v-akt 鼠胸腺瘤病毒癌基因同源物 3 (蛋白激酶 B, γ)
ALDH2	醛脱氢酶 2 家族 (线粒体)
ALK	间变性淋巴瘤受体酪氨酸激酶
APC	腺瘤性结肠息肉
AR	雄激素受体
ARHGAP26	Rho GTP 酶激活蛋白 26
ARHGEF12	Rho 鸟嘌呤核苷酸交换因子 (GEF) 12
ARID1A	富含 AT 的相互作用结构域 1A (SWI1-样)
ARID1B	富含 AT 的相互作用结构域 1B (SWI1-样)
ARID2	富含 AT 的相互作用结构域 2 (ARID, RFX-样)
ARID3A	富含 AT 的相互作用结构域 3A (BRIGHT-样)

[0126]

[0127]	ARID3B	富含 AT 的相互作用结构域 3B (BRIGHT-样)
	ARID4A	富含 AT 的相互作用结构域 4A (RBP1-样)
	ARID4B	富含 AT 的相互作用结构域 4B (RBP1-样)
	ARID5A	富含 AT 的相互作用结构域 5A (MRF1-样)
	ARID5B	富含 AT 的相互作用结构域 5B (MRF1-样)
	ARNT	芳烃受体核易位体
	ASPSCR1	软组织腺泡状肉瘤染色体区域, 候选物 1
	ASXL1	附加的性梳样 1 (果蝇)
	ATF1	激活转录因子 1
	ATIC	5-氨基咪唑-4-甲酰胺核糖核苷酸甲酰基转移酶/IMP 环化水解酶
	ATM	突变的毛细血管扩张性共济失调
	ATR	毛细血管扩张性共济失调和 Rad3 相关
	ATRX	α -地中海贫血/智力低下综合征 X-连锁
	AURKA	极光激酶 A
	AXIN2	轴蛋白 2
	BAP1	BRCA1 相关蛋白-1 (遍在蛋白羧基末端水解酶)
	BCL10	B-细胞 CLL/淋巴瘤 10
	BCL11A	B-细胞 CLL/淋巴瘤 11A (锌指蛋白)
	BCL11B	B-细胞 CLL/淋巴瘤 11B (锌指蛋白)
	BCL2	B-细胞 CLL/淋巴瘤 2
	BCL3	B-细胞 CLL/淋巴瘤 3
	BCL6	B-细胞 CLL/淋巴瘤 6
	BCL7A	B-细胞 CLL/淋巴瘤 7A
	BCL9	B-细胞 CLL/淋巴瘤 9
	BCOR	BCL6 辅阻遏剂
	BCR	断点簇区
	BIRC3	含有杆状病毒 IAP 重复序列的 3
	BLID	含有 BH3-样基序, 细胞死亡诱导物
	BLM	Bloom 综合征, RecQ 解旋酶-样
	BMPR1A	骨形态发生蛋白受体, IA 型
	BRAF	v-raf 鼠肉瘤病毒癌基因同源物 B1
	BRCA1	乳腺癌 1, 发病初期
	BRCA2	乳腺癌 2, 发病初期
	BRD3	含有溴结构域的 3
	BRD4	含有溴结构域的 4
	BRIP1	BRCA1 相互作用蛋白 C-末端解旋酶 1
	BTG1	B-细胞易位基因 1, 抗增殖

[0128]

BUB1B	不受苯丙咪唑 1 同源物 β (酵母) 抑制的出芽
C15orf55	染色体 15 开放阅读框 55
CANT1	钙激活的核苷酸酶 1
CARD11	半胱天冬酶募集的结构域家族, 成员 11
CARS	半胱酰胺-tRNA 合成酶
CASC5	癌症敏感性候选物 5
CBFA2T3	核心结合因子, runt 结构域, α 亚基 2; 易位到, 3
CBFB	核心结合因子, β 亚基
CBL	Cbl 原癌基因, E3 遍在蛋白连接酶
CBLB	Cbl 原癌基因, E3 遍在蛋白连接酶 B
CBLC	Cbl 原癌基因, E3 遍在蛋白连接酶 C
CCDC6	含有卷曲螺旋结构域的 6
CCNB1IP1	细胞周期蛋白 B1 相互作用蛋白 1, E3 遍在蛋白连接酶
CCND1	细胞周期蛋白 D1
CCND2	细胞周期蛋白 D2
CCND3	细胞周期蛋白 D3
CCNE1	细胞周期蛋白 E1
CD274	CD274 分子
CD74	CD74 分子, 主要组织相容性复合体, II 类不变链
CD79A	CD79a 分子, 免疫球蛋白相关 α
CD79B	CD79b 分子, 免疫球蛋白相关 β
CDC73	细胞分裂周期 73, Paf1/RNA 聚合酶 II 复合体组分, 同源物 (酿酒酵母)
CDH1	钙黏着蛋白 1, 1 型, E-钙黏着蛋白 (上皮)
CDH11	钙黏着蛋白 11, 2 型, OB-钙黏着蛋白 (成骨细胞)
CDH6	钙黏着蛋白 6, 2 型, K-钙黏着蛋白 (胎儿肾脏)
CDK12	细胞周期蛋白依赖性激酶 12
CDK2AP2	细胞周期蛋白依赖性激酶 2 相关蛋白 2
CDK4	细胞周期蛋白依赖性激酶 4
CDK6	细胞周期蛋白依赖性激酶 6
CDK8	细胞周期蛋白依赖性激酶 8
CDKN1A	细胞周期蛋白依赖性激酶抑制剂 1A (p21, Cip1)
CDKN1B	细胞周期蛋白依赖性激酶抑制剂 1B (p27, Kip1)
CDKN2A	细胞周期蛋白依赖性激酶抑制剂 2A
CDKN2B	细胞周期蛋白依赖性激酶抑制剂 2B (p15, 抑制 CDK4)
CDKN2C	细胞周期蛋白依赖性激酶抑制剂 2C (p18, 抑制 CDK4)
CDKN2D	细胞周期蛋白依赖性激酶抑制剂 2D (p19, 抑制 CDK4)
CDX2	尾型同源框 2

[0129]	CEBPA	CCAAT/增强子结合蛋白 (C/EBP), α
	CHCHD7	含有卷曲-螺旋-螺旋-卷曲-螺旋-螺旋结构域的 7
	CHD5	染色质域解旋酶 DNA 结合蛋白 5
	CHD6	染色质域解旋酶 DNA 结合蛋白 6
	CHEK1	检查点激酶 1
	CHEK2	检查点激酶 2
	CHIC2	富含半胱氨酸的疏水结构域 2
	CHN1	嵌蛋白 (嵌合蛋白) 1
	CIC	Capicua 同源物 (果蝇)
	CIITA	II 类, 主要组织相容性复合体, 反式激活因子
	CLP1	CLP1, 切割和聚腺苷酸化因子 I 亚基, 同源物 (酿酒酵母)
	CLTC	网格蛋白, 重链 (Hc)
	CLTCL1	网格蛋白, 重链样 1
	CNBP	CCHC 型锌指, 核酸结合蛋白
	CNTRL	centriolin
	COL1A1	胶原蛋白, I 型, $\alpha 1$
	COX6C	细胞色素 c 氧化酶亚基 VIc
	CREB1	cAMP 响应元件结合蛋白 1
	CREB3L1	cAMP 响应元件结合蛋白 3-样 1
	CREB3L2	cAMP 响应元件结合蛋白 3-样 2
	CREBBP	CREB 结合蛋白
	CRKL	v-crk 肉瘤病毒 CT10 癌基因同源物 (鸟) -样
	CRLF2	细胞因子受体-样因子 2
	CRTC1	CREB 调节转录辅激活物 1
	CRTC3	CREB 调节转录辅激活物 3
	CSF1R	集落刺激因子 1 受体
	CTNNB1	联蛋白 (钙黏着蛋白相关蛋白), $\beta 1$, 88kDa
	CXCR7	趋化因子 (C-X-C 基序) 受体 7
	CYLD	圆柱瘤 (头巾样瘤综合征)
	CYP1B1	细胞色素 P450, 家族 1, 亚家族 B, 多肽 1
	DAXX	死亡结构域相关蛋白
	DDB2	损伤-特异性 DNA 结合蛋白 2, 48kDa
	DDIT3	DNA-损伤-诱导型转录物 3
	DDX10	DEAD (Asp-Glu-Ala-Asp) 框多肽 10
	DDX5	DEAD (Asp-Glu-Ala-Asp) 框解旋酶 5
	DDX6	DEAD (Asp-Glu-Ala-Asp) 框解旋酶 6
	DEK	DEK 癌基因

[0130]	DICER1	切酶 1, 核糖核酸酶 III 型
	DNMT3A	DNA (胞嘧啶-5-)-甲基转移酶 3 α
	DUX4	双同源框 4
	EBF1	早期 B-细胞因子 1
	EGFR	表皮生长因子受体
	EIF4A2	真核生物转译起始因子 4A2
	ELAC2	elaC 同源物 2 (大肠杆菌)
	ELF4	E74-样因子 4 (ets 结构域转录因子)
	ELK4	ELK4, ETS-结构域蛋白 (SRF 辅助蛋白 1)
	ELL	延伸因子 RNA 聚合酶 II
	ELN	弹性蛋白
	EML4	棘皮动物微管相关蛋白样 4
	EP300	E1A 结合蛋白 p300
	EPCAM	上皮细胞粘附分子
	EPHA10	EPH 受体 A10
	EPHA3	EPH 受体 A3
	EPHA5	EPH 受体 A5
	EPHA6	EPH 受体 A6
	EPHB6	EPH 受体 B6
	EPS15	表皮生长因子受体途径底物 15
	ERBB2	v-erb-b2 成红细胞白血病病毒癌基因同源物 2, 神经/胶质母细胞瘤衍生的癌基因同源物 (鸟)
	ERBB3	v-erb-b2 成红细胞白血病病毒癌基因同源物 3 (鸟)
	ERBB4	v-erb-a 成红细胞白血病病毒癌基因同源物 4 (鸟)
	ERC1	ELKS/RAB6-相互作用/CAST 家族成员 1
	ERCC1	切除修复交叉-互补啮齿动物修复缺陷, 互补群 1 (包括重叠的反义序列)
	ERCC2	切除修复交叉-互补啮齿动物修复缺陷, 互补群 2
	ERCC3	切除修复交叉-互补啮齿动物修复缺陷, 互补群 3
	ERCC4	切除修复交叉-互补啮齿动物修复缺陷, 互补群 4
	ERCC5	切除修复交叉-互补啮齿动物修复缺陷, 互补群 5
	ERG	v-ets 成红细胞增多症病毒 E26 癌基因同源物 (鸟)
	ETV1	ets 变体 1
	ETV4	ets 变体 4
	ETV5	ets 变体 5
	ETV6	ets 变体 6
	EWSR1	尤因肉瘤断点区 1
	EXT1	exostosin 1

	EXT2	exostosin 2
	EZH2	zeste 同源物 2 的增强子 (果蝇)
	FAM123B	具有序列相似性的家族 123B
	FAM22A	具有序列相似性的家族 22, 成员 A
	FAM22B	具有序列相似性的家族 22, 成员 B
	FAM46C	具有序列相似性的家族 46, 成员 C
	FANCA	范科尼贫血, 互补群 A
	FANCC	范科尼贫血, 互补群 C
	FANCD2	范科尼贫血, 互补群 D2
	FANCE	范科尼贫血, 互补群 E
	FANCF	范科尼贫血, 互补群 F
	FANCG	范科尼贫血, 互补群 G
	FAS	Fas (TNF 受体超家族, 成员 6)
	FBXO11	F-框蛋白 11
[0131]	FBXW7	含有 F-框和 WD 重复序列结构域的 7, E3 遍在蛋白蛋白质连接酶
	FCGR2B	IgG 的 Fc 片段, 低亲和力 IIb, 受体 (CD32)
	FCRL4	Fc 受体-样 4
	FEV	FEV (ETS 癌基因家族)
	FGF23	成纤维细胞生长因子 23
	FGFR1	成纤维细胞生长因子受体 1
	FGFR1OP	FGFR1 癌基因配偶体
	FGFR2	成纤维细胞生长因子受体 2
	FGFR3	成纤维细胞生长因子受体 3
	FGFR4	成纤维细胞生长因子受体 4
	FH	延胡索酸水合酶
	FHIT	脆性组氨酸三联体
	FHL1	四个半 LIM 结构域 1
	FIP1L1	FIP1 样 1 (酿酒酵母)
	FKBP1B	FK506 结合蛋白 1B, 12.6 kDa
	FKBP9	FK506 结合蛋白 9, 63 kDa
	FLCN	雌酮
	FLI1	弗罗德白血病病毒整合 1
	FLT1	未表征的蛋白质 LOC145788
	FLT3	fms- 相关酪氨酸激酶 3
	FLT4	fms- 相关酪氨酸激酶 4
	FNBP1	形成蛋白结合蛋白 1
	FOLR1	叶酸受体 1 (成人)

[0132]

FOXC1	叉头框 C1
FOXL2	叉头框 L2
FOXO1	叉头框 O1
FOXO3	叉头框 O3
FOXO4	叉头框 O4
FOXP1	叉头框 P1
FSTL3	促滤泡素抑制剂-样 3 (分泌的糖蛋白)
FUBP1	远上游元件 (FUSE) 结合蛋白 1
FUS	在肉瘤中融合
GALNT3	UDP-N-乙酰基- α -D-半乳糖胺: 多肽 N-乙酰半乳糖胺基转移酶 3 (GalNAc-T3)
GAS7	生长停滞-特异性 7
GATA1	GATA 结合蛋白 1 (球蛋白转录因子 1)
GATA2	GATA 结合蛋白 2
GATA3	GATA 结合蛋白 3
GLMN	肾小球蛋白, FKBP 相关蛋白
GMPS	鸟嘌呤单磷酸合成酶
GNA11	鸟嘌呤核苷酸结合蛋白 (G 蛋白), $\alpha 11$ (Gq 类)
GNAQ	鸟嘌呤核苷酸结合蛋白 (G 蛋白), q 多肽
GNAS	GNAS 复合体基因座
GOLGA5	高尔基体蛋白 A5
GOPC	含有高尔基-相关 PDZ 和卷曲螺旋基序
GPC3	磷脂酰肌醇蛋白聚糖 3
GPHN	桥尾蛋白
GSTM1	谷胱甘肽 S-转移酶 mu 1
GUCY1A2	鸟苷酸环化酶 1, 可溶的, $\alpha 2$
HECW1	含有 HECT、C2 和 WW 结构域的 E3 遍在蛋白蛋白质连接酶 1
HERPUD1	同型半胱氨酸-诱导型, 内质网应激-诱导型, 泛素样结构域成员 1
HEY1	与 YRPW 基序 1 有关的多毛/分裂增强物
HIP1	亨廷顿相互作用蛋白 1
HIST1H4I	组蛋白簇 1, H4i
HLF	肝白血病因子
HMGA1	高移动性组 AT-钩 1
HMGA2	高移动性组 AT-钩 2
HMGN2P46	高移动性组核小体结合结构域 2 假基因 46
HNF1A	HNF1 同源框 A

[0133]	HNRNPA2B1	核不均一核糖核蛋白 A2/B1
	HOOK3	钩同源物 3 (果蝇)
	HOXA11	同源框 A11
	HOXA13	同源框 A13
	HOXA9	同源框 A9
	HOXC11	同源框 C11
	HOXC13	同源框 C13
	HOXD11	同源框 D11
	HOXD13	同源框 D13
	HRAS	v-Ha-ras Harvey 大鼠肉瘤病毒癌基因同源物
	HSD17B3	羟基类固醇 (17- β) 脱氢酶 3
	HSD3B2	羟基- δ -5-类固醇脱氢酶, 3 β -和类固醇 δ -异构酶 2
	HSP90AA1	热休克蛋白 90kDa α (胞质), A 类成员 1
	HSP90AB1	热休克蛋白 90kDa α (胞质), B 类成员 1
	IDH1	异柠檬酸脱氢酶 1 (NADP+), 可溶的
	IDH2	异柠檬酸脱氢酶 2 (NADP+), 线粒体的
	IGF1R	胰岛素样生长因子 1 受体
	IKBKE	B 细胞中 κ 轻链多肽基因增强子抑制剂, 激酶 ε
	IKZF1	IKAROS 家族锌指 1 (Ikaros)
	IL2	白细胞介素 2
	IL21R	白细胞介素 21 受体
	IL6ST	白细胞介素 6 信号转导 (gp130, 制癌蛋白 M 受体)
	IL7R	白细胞介素 7 受体
	IRF4	干扰素调节因子 4
	ITK	IL2-诱导型 T-细胞激酶
	JAK1	Janus 激酶 1
	JAK2	Janus 激酶 2
	JAK3	Janus 激酶 3
	JAZF1	JAZF 锌指 1
	JUN	jun 原癌基因
	KAT6A	K (赖氨酸) 乙酰转移酶 6A
	KAT6B	K (赖氨酸) 乙酰转移酶 6B
	KDM5A	赖氨酸 (K) -特异性脱甲基酶 5A
	KDM5C	赖氨酸 (K) -特异性脱甲基酶 5C
	KDM6A	赖氨酸 (K) -特异性脱甲基酶 6A
	KDR	激酶插入结构域受体 (III 型受体酪氨酸激酶)
	KDSR	3-酮二氢鞘氨酸还原酶
	KEAP1	kelch-样 ECH-相关蛋白 1

	KIAA1549	KIAA1549
	KIF1B	驱动蛋白家族成员 1B
	KIT	v-kit Hardy-Zuckerman 4 猫肉瘤病毒癌基因同源物
	KL	klotho
	KLF6	Kruppel-样因子 6
	KLK2	激肽释放酶相关肽酶 2
	KRAS	v-Ki-ras2 Kirsten 大鼠肉瘤病毒癌基因同源物
	KRT17	角蛋白 17
	KTN1	移动结合蛋白 1 (驱动蛋白受体)
	LASP1	LIM 和 SH3 蛋白质 1
	LCK	淋巴细胞-特异性蛋白酪氨酸激酶
	LCP1	淋巴细胞胞质蛋白 1 (L-丝束蛋白)
	LHFP	脂肪瘤 HMGIC 融合配偶体
	LIFR	白血病抑制因子受体 α
[0134]	LMO1	仅 LIM 结构域 1 (斜方蛋白 1 (rhombotin-1))
	LMO2	仅 LIM 结构域 2 (斜方蛋白-样 1)
	LPP	脂肪瘤中含有 LIM 结构域的优先易位配偶体
	LRP5	低密度脂蛋白受体-相关蛋白 5
	LTBP2	潜在转化生长因子 β 结合蛋白 2
	LTBP3	潜在转化生长因子 β 结合蛋白 3
	LYL1	淋巴母细胞性白血病衍生的序列 1
	MAD2L1BP	MAD2L1 结合蛋白
	MAF	v-maf 肌腱膜纤维肉瘤癌基因同源物 (鸟)
	MAFB	v-maf 肌腱膜纤维肉瘤癌基因同源物 B (鸟)
	MALAT1	转移相关肺腺癌转录物 1 (非蛋白质编码)
	MALT1	粘膜相关淋巴组织淋巴瘤易位基因 1
	MAML2	mastermind-样 2 (果蝇)
	MAP2K1	促分裂原活化蛋白激酶激酶 1
	MAP2K2	促分裂原活化蛋白激酶激酶 2
	MAP2K4	促分裂原活化蛋白激酶激酶 4
	MAP3K1	促分裂原活化蛋白激酶激酶激酶 1, E3 遍在蛋白连接酶
	MAP3K8	促分裂原活化蛋白激酶激酶激酶 8
	MAX	MYC 相关因子 X
	MC1R	黑皮质素 1 受体 (α 促黑素细胞激素受体)
	MCL1	髓细胞白血病序列 1 (BCL2-相关)
	MDM2	Mdm2, p53 E3 遍在蛋白连接酶同源物 (小鼠)
	MDM4	Mdm4 p53 结合蛋白同源物 (小鼠)
	MDS2	骨髓增生异常综合征 2 易位相关

[0135]	MECOM	MDS1 和 EVI1 复合体基因座
	MED12	中介体复合物亚基 12
	MEN1	多发性内分泌瘤 I
	MET	met 原癌基因 (肝细胞生长因子受体)
	MITF	小眼-相关转录因子
	MKL1	巨核母细胞白血病 (易位) 1
	MLF1	髓样白血病因子 1
	MLH1	mutL 同源物 1, 结肠癌, 非息肉病性 2 型 (大肠杆菌)
	MLL	髓样/淋巴样或混合谱系白血病 (trithorax 同源物, 果蝇)
	MLL2	髓样/淋巴样或混合谱系白血病 2
	MLL3	髓样/淋巴样或混合谱系白血病 3
	MLLT1	髓样/淋巴样或混合谱系白血病 (trithorax 同源物, 果蝇); 易位到, 1
	MLLT10	髓样/淋巴样或混合谱系白血病 (trithorax 同源物, 果蝇); 易位到, 10
	MLLT11	髓样/淋巴样或混合谱系白血病 (trithorax 同源物, 果蝇); 易位到, 11
	MLLT3	髓样/淋巴样或混合谱系白血病 (trithorax 同源物, 果蝇); 易位到, 3
	MLLT4	髓样/淋巴样或混合谱系白血病 (trithorax 同源物, 果蝇); 易位到, 4
	MLLT6	髓样/淋巴样或混合谱系白血病 (trithorax 同源物, 果蝇); 易位到, 6
	MN1	脑膜瘤 (平衡易位破坏) 1
	MNX1	运动神经元和胰腺同源框 1
	MPL	骨髓增生性白血病病毒癌基因
	MRE11A	MRE11 减数分裂重组 11 同源物 A (酿酒酵母)
	MSH2	mutS 同源物 2, 结肠癌, 非息肉病 1 型 (大肠杆菌)
	MSH6	mutS 同源物 6 (大肠杆菌)
	MSI2	musashi 同源物 2 (果蝇)
	MSN	膜突蛋白
	MTCP1	成熟 T 细胞增殖 1
	MTCP1NB	成熟 T 细胞增殖 1 相邻物
	MTOR	雷帕霉素的机制靶标 (丝氨酸/苏氨酸激酶)
	MTUS2	微管相关肿瘤抑制剂候选物 2
	MUC1	粘蛋白 1, 细胞表面相关
	MUTYH	mutY 同源物 (大肠杆菌)
	MYB	v-myb 成髓细胞增多症病毒癌基因同源物 (鸟)

MYC	v-myc 髓细胞瘤病病毒癌基因同源物 (鸟)
MYCL1	v-myc 髓细胞瘤病病毒癌基因同源物 1, 肺癌衍生 (鸟)
MYCN	v-myc 髓细胞瘤病病毒相关癌基因, 成神经细胞瘤衍生 (鸟)
MYD88	髓样分化初级应答基因 (88)
MYH11	肌球蛋白, 重链 11, 平滑肌
MYH9	肌球蛋白, 重链 9, 非肌肉
MYOC	肌纤蛋白, 小梁网诱导型糖皮质激素反应
NACA	新生多肽相关复合体 α 亚基
NBN	断蛋白
NCKIPSD	具有 SH3 结构域的 NCK 相互作用蛋白
NCOA1	核受体辅激活物 1
NCOA2	核受体辅激活物 2
NCOA4	核受体辅激活物 4
NDRG1	N-myc 下游调节的 1
NF1	神经纤维瘤蛋白 1
NF2	神经纤维瘤蛋白 2 (膜突样蛋白)
NFE2L2	核因子 (红细胞系-衍生 2) -样 2
NFIB	核因子 I/B
NFKB2	B 细胞中 κ 轻链多肽基因增强子的核因子 2 (p49/p100)
NIN	ninein (GSK3B 相互作用蛋白)
NKX2-1	NK2 同源框 1
NONO	含有非-POU 结合域, 八聚体-结合
NOTCH1	切迹 1
NOTCH2	切迹 2
NOTCH3	切迹 3
NOTCH4	切迹 4
NPM1	核磷蛋白 (核仁磷蛋白 B23, 核基质蛋白)
NR4A3	核受体亚家族 4, A 组, 成员 3
NRAS	成神经细胞瘤 RAS 病毒 (v-ras) 癌基因同源物
NSD1	核受体结合 SET 结构域蛋白质 1
NTRK1	神经营养酪氨酸激酶, 受体, 1 型
NTRK2	神经营养酪氨酸激酶, 受体, 2 型
NTRK3	神经营养酪氨酸激酶, 受体, 3 型
NUMA1	核有丝分裂器蛋白 1
NUP214	核孔蛋白 214kDa
NUP98	核孔蛋白 98kDa
OLIG2	少突胶质细胞谱系转录因子 2

[0136]

OMD	骨调蛋白
OPTN	视神经蛋白
P2RY8	嘌呤受体 P2Y, G-蛋白偶联, 8
PAFAH1B2	血小板-激活因子乙酰水解酶 1b, 催化亚基 2 (30kDa)
PAK7	p21 蛋白质 (Cdc42/Rac) -活化激酶 7
PALB2	BRCA2 的配偶体和定位器
PALLD	Palladin, 细胞骨架相关蛋白
PATZ1	含有 POZ (BTB) 和 AT 钩的锌指 1
PAX2	配对框 2
PAX3	配对框 3
PAX5	配对框 5
PAX6	配对框 6
PAX7	配对框 7
PAX8	配对框 8
PBRM1	polybromo 1
PBX1	前-B-细胞白血病同源框 1
PCM1	中心粒外周物质 1
PCSK7	蛋白质转换酶枯草杆菌蛋白酶/kexin 7 型
PDCD1LG2	程序性细胞死亡 1 配体 2
PDE4DIP	磷酸二酯酶 4D 相互作用蛋白
PDGFB	血小板源生长因子 β 多肽
PDGFRA	血小板源生长因子, α 多肽
PDGFRB	血小板源生长因子, β 多肽
PDPK1	3-磷酸肌醇依赖性蛋白激酶-1
PER1	周期性同源物 1 (果蝇)
PHF6	PHD 指蛋白 6
PHOX2B	配对-样同源框 2b
PICALM	磷脂酰肌醇结合网格蛋白装配蛋白
PIK3CA	磷脂酰肌醇-3-激酶, 催化性, α 多肽
PIK3R1	磷脂酰肌醇-3-激酶, 调节亚基 1 (α)
PIM1	pim-1 癌基因
PLAG1	多形腺瘤基因 1
PLK1	polo-样激酶 1
PML	前髓细胞性白血病
PMS1	PMS1 减数分裂后分离增加 1 (酿酒酵母)
PMS2	PMS2 减数分裂后分离增加 2 (酿酒酵母)
POU2AF1	POU 2 类相关因子 1
POU5F1	POU 5 类同源框 1

[0137]

[0138]	PPARG	过氧化物酶体增殖物激活受体 γ
	PPP2R1A	蛋白质磷酸酶 2, 调节亚基, α
	PRCC	乳头状肾细胞癌 (易位相关)
	PRDM1	含有 PR 结构域的 1, 具有 ZNF 结构域
	PRDM16	含有 PR 结构域的 16
	PRF1	穿孔素 1 (孔形成蛋白质)
	PRKAR1A	蛋白激酶, cAMP-依赖性, 调节性, I 型, α (组织特异性消灭基因 1)
	PRKDC	蛋白激酶, DNA-激活, 催化多肽
	PRRX1	配对相关同源框 1
	PSIP1	PC4 和 SFRS1 相互作用蛋白 1
	PTCH1	修补的 1
	PTEN	磷酸酶和张力蛋白同源物
	PTK2	PTK2 蛋白酪氨酸激酶 2
	PTK2B	PTK2B 蛋白酪氨酸激酶 2 β
	PTPN11	蛋白酪氨酸磷酸酶, 非受体 11 型
	PTPRD	蛋白酪氨酸磷酸酶, 受体 D 型
	RABEP1	Rab 接触蛋白, RAB GTP 酶结合效应蛋白 1
	RAD51B	RAD51 同源物 B (酿酒酵母)
	RAF1	v-raf-1 鼠白血病病毒癌基因同源物 1
	RALGDS	ral 鸟嘌呤核苷酸解离刺激物
	RANBP17	RAN 结合蛋白 17
	RAP1GDS1	RAP1, GTP-GDP 解离刺激物 1
	RARA	视黄酸受体, α
	RB1	视网膜母细胞瘤 1
	RBM15	RNA 结合基序蛋白质 15
	RECQL4	RecQ 蛋白质-样 4
	REL	v-rel 网状内皮增殖病毒癌基因同源物 (鸟)
	RET	Ret 原癌基因
	RHOH	ras 同源物家族成员 H
	RICTOR	MTOR 的 RPTOR 独立伴随物, 复合体 2
	RMI2	RMI2, RecQ 介导的基因组不稳定性 2, 同源物 (酿酒酵母)
	RNASEL	核糖核酸酶 L (2,5-寡异腺苷酸合成酶-依赖性)
	ROS1	c-ros 癌基因 1, 受体酪氨酸激酶
	RPL22	核糖体蛋白质 L22
	RPN1	核糖体结合糖蛋白 I
	RPTOR	MTOR 的调节相关蛋白, 复合体 1

[0139]	RRM1	核糖核苷酸还原酶 M1
	RUNX1	runt-相关转录因子 1
	RUNX1T1	runt-相关转录因子 1；易位到，1 (细胞周期蛋白 D-相关)
	SARDH	肌氨酸脱氢酶
	SBDS	Shwachman-Bodian-Diamond 综合征
	SDHAF2	琥珀酸脱氢酶复合体装配因子 2
	SDHB	琥珀酸脱氢酶复合体，亚基 B，铁硫 (Ip)
	SDHC	琥珀酸脱氢酶复合体，亚基 C，膜内在蛋白质，15kDa
	SDHD	琥珀酸脱氢酶复合体，亚基 D，膜内在蛋白质
	SEPT5	隔蛋白 5
	SEPT6	隔蛋白 6
	SEPT9	隔蛋白 9
	SET	SET 核癌基因
	SETD2	含有 SET 结构域的 2
	SETDB1	SET 结构域，分叉 1
	SF3B1	剪接因子 3b，亚基 1，155kDa
	SF3B2	剪接因子 3b，亚基 2，145kDa
	SFPQ	剪接因子，富含脯氨酸/谷氨酰胺
	SH3GL1	SH3-结构域 GRB2-样 1
	SLC45A3	溶质载体家族 45，成员 3
	SMAD2	SMAD 家族成员 2
	SMAD3	SMAD 家族成员 3
	SMAD4	SMAD 家族成员 4
	SMARCA4	SWI/SNF 相关，基质相关，染色质的肌动蛋白依赖性调节物，亚家族 a，成员 4
	SMARCB1	SWI/SNF 相关，基质相关，染色质的肌动蛋白依赖性调节物，亚家族 b，成员 1
	SMO	平滑的，卷曲家族受体
	SNX29	分选微管连接蛋白 29
	SOCS1	细胞因子信号传导抑制蛋白 1
	SOX2	SRY (性别决定区域 Y) - 框 2
	SPECC1	具有钙调理蛋白同源性和卷曲螺旋结构域 1 的精子抗原
	SPEN	Spn 同源物，转录调节因子 (果蝇)
	SRC	v-src 肉瘤 (Schmidt-Ruppin A-2) 病毒癌基因同源物 (鸟)
	SRD5A2	类固醇-5- α -还原酶， α 多肽 2 (3-氧化-5 α -类固醇 δ 4-脱氢酶 α 2)
	SRGAP3	SLIT-ROBO Rho GTP 酶激活蛋白 3
	SRSF2	富含丝氨酸/精氨酸的剪接因子 2

[0140]	SRSF	富含丝氨酸/精氨酸的剪接因子 3
	SS18	滑膜肉瘤易位, 染色体 18
	SS18L1	在染色体 18-样 1 上的滑膜肉瘤易位基因
	SSX1	滑膜肉瘤, X 断点 1
	SSX2	滑膜肉瘤, X 断点 2
	SSX4	滑膜肉瘤, X 断点 4
	STAT3	信号转导和转录激活因子 3 (急性时相反应因子)
	STIL	SCL/TAL1 中断基因座
	STK11	丝氨酸/苏氨酸激酶 11
	STX11	突触融合蛋白 11
	STXBP2	突触融合蛋白结合蛋白 2
	SUFU	融合同源物的抑制蛋白 (果蝇)
	SUZ12	zeste 12 同源物的抑制蛋白 (果蝇)
	SYK	脾酪氨酸激酶
	TAF15	TAF15 RNA 聚合酶 II, TATA 框结合蛋白 (TBP) - 相关因子, 68kDa
	TAL1	T-细胞急性淋巴细胞白血病 1
	TAL2	T-细胞急性淋巴细胞白血病 2
	TCEA1	转录延伸因子 A (SII), 1
	TCEA1P2	转录延伸因子 A (SII), 1 假基因 2
	TCF12	转录因子 12
	TCF3	转录因子 3 (E2A 免疫球蛋白增强子结合因子 E12/E47)
	TCF4	转录因子 4
	TCF7L2	转录因子 7-样 2 (T-细胞特异性, HMG-框)
	TCL1A	T-细胞白血病/淋巴瘤 1A
	TCL6	T-细胞白血病/淋巴瘤 6 (非蛋白质编码)
	TERT	端粒酶逆转录酶
	TET1	tet 甲基胞嘧啶双加氧酶 1
	TET2	tet 甲基胞嘧啶双加氧酶 2
	TFE3	与 IGHM 增强子 3 结合的转录因子
	TFEB	转录因子 EB
	TFG	TRK-融合基因
	TFPT	TCF3 (E2A) 融合配偶体 (儿童期白血病)
	TFRC	运铁蛋白受体 (p90, CD71)
	TGFBR2	转化生长因子, β 受体 II (70/80kDa)
	THRAP3	甲状腺激素受体相关蛋白 3
	TLX1	T-细胞白血病同源框 1
	TLX3	T-细胞白血病同源框 3

[0141]	TMEM127	跨膜蛋白 127
	TMPRSS2	跨膜蛋白, 丝氨酸 2
	TNFAIP3	肿瘤坏死因子, α -诱导蛋白 3
	TNFRSF14	肿瘤坏死因子受体超家族, 成员 14
	TNFRSF17	肿瘤坏死因子受体超家族, 成员 17
	TOP1	拓扑异构酶 (DNA) I
	TOP2A	拓扑异构酶 (DNA) II α 170kDa
	TP53	肿瘤蛋白 p53
	TPM3	原肌球蛋白 3
	TPM4	原肌球蛋白 4
	TPR	易位的启动子区域, 核篮蛋白
	TRIM24	含有三重基序的 24
	TRIM27	含有三重基序的 27
	TRIM33	含有三重基序的 33
	TRIP11	甲状腺激素受体相互作用物 11
	TSC1	结节性硬化 1
	TSC2	结节性硬化 2
	TSHR	促甲状腺激素受体
	TTL	微管蛋白酪氨酸连接酶
	TYK2	酪氨酸激酶 2
	U2AF1	U2 小核 RNA 辅因子 1
	UNC13D	unc-13 同源物 D (秀丽线虫)
	USP6	遍在蛋白特异性肽酶 6 (Tre-2 癌基因)
	UTY	遍在转录的三十四肽重复序列基因, Y-连锁
	VHL	von Hippel-Lindau 肿瘤抑制蛋白, E3 遍在蛋白连接酶
	VTI1A	通过与 t-SNARE 同源物 1A 相互作用的囊泡运输 (酵母)
	WAS	Wiskott-Aldrich 综合征 (湿疹-血小板减少症)
	WDR36	WD 重复结构域 36
	WHSC1	Wolf-Hirschhorn 综合征候选物 1
	WHSC1L1	Wolf-Hirschhorn 综合征候选物 1-样 1
	WIF1	WNT 抑制因子 1
	WRN	Werner 综合征, RecQ 解旋酶-样
	WT1	Wilms 肿瘤 1
	XPA	着色性干皮病, 互补群 A
	XPC	着色性干皮病, 互补群 C
	XPO1	核输出蛋白 1 (CRM1 同源物, 酵母)
	YWHAE	酪氨酸 3-单加氧酶/色氨酸 5-单加氧酶活化蛋白, ϵ 多肽
	ZBTB16	含有锌指和 BTB 结构域的 16

[0142]	ZMYM2	锌指, MYM 型 2
	ZNF331	锌指蛋白 331
	ZNF384	锌指蛋白 384
	ZNF521	锌指蛋白 521
	ZNF668	锌指蛋白 668
	ZRSR2	锌指 (CCCH 型), 富含 RNA-结合基序和丝氨酸/精氨酸的 2

[0143] 在一方面,本公开内容提供了用于富集包含至少两个或更多个拷贝的靶多核苷酸的多联体的扩增子的试剂盒。试剂盒可包含本文关于各个方面中的任何方面以任何组合方式公开的一个或多个元件。在一些实施方案中,该试剂盒包含: (a) 第一引物,其包含通过序列互补性与靶多核苷酸特异性杂交的第一3' 端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5' 端; (b) 第二引物,其包含通过序列互补性与多联体特异性杂交的第二3' 端和不通过序列互补性与多联体特异性杂交的包含第二共同序列的第二5' 端,其中第一共同序列和第二共同序列各自在5' 端包含至少10个连续核苷酸并且当最佳比对时至少90% 相同;以及 (c) 第三引物,其具有通过序列互补性与第一共同序列或第二共同序列特异性杂交的序列。试剂盒中的试剂和其他材料可以容纳在任何合适的容器中,并且可以是直接可用的形式,或者需要与试剂盒中的其他试剂或使用者提供的试剂进行组合(例如,浓缩的组合物的稀释或冻干组合物的重建)。试剂盒可提供缓冲液,其非限制性实例包括碳酸钠缓冲液、碳酸氢钠缓冲液、硼酸盐缓冲液、Tris缓冲液、MOPS缓冲液、HEPES缓冲液及其组合。试剂盒可包含对照样品,例如纯化的DNA,用作阳性对照或定量标准。在一些实施方案中,该试剂盒包含一种或多种用于扩增多核苷酸的酶,如逆转录酶和聚合酶中的一种或多种。在需要的情况下,本试剂盒可进一步包含一个或多个可检测标志物,以实现例对扩增产物积累的监测,如实时监测。上文描述了可检测标志物的非限制性实例,并且其包括在扩增步骤期间优先或排外地结合双链DNA的染料,如SYBR绿色染料或BEB0染料。在一些实施方案中,该试剂盒包含探针寡核苷酸,该探针寡核苷酸包含荧光团和淬灭剂以检测扩增反应的进展或产物。在一些实施方案中,该试剂盒包含根据本文公开的一种或多种方法的试剂盒使用说明书。在一些实施方案中,第一共同序列和第二共同序列是相同的。在一些实施方案中,第一共同序列、第二共同序列和第三引物的杂交序列均具有在彼此±5°C内的解链温度 (Tm)。在一些实施方案中,沿靶多核苷酸从5' 至3' 对应于 (i) 与第一3' 端互补的序列、(ii) 与第二3' 端相同的序列以及 (iii) (i) 与 (ii) 之间的间插序列的靶多核苷酸序列部分的组合长度为75个或更少的核苷酸。

[0144] 在一方面,本公开内容提供了用于设计用于富集包含至少两个或更多个拷贝的靶多核苷酸的多联体的扩增子的引物的系统。关于本公开内容的各个方面中的任何方面,引物可包含本文所述的任何特征。在一些实施方案中,该系统包含: (a) 计算机,其被配置用于接收客户请求以设计用于扩增指定靶序列的引物; (b) 包含代码的计算机可读介质,该代码在由一个或多个处理器执行时设计至少三种用于扩增靶序列的引物,其中所述至少三种引物包含: (i) 第一引物,其包含通过序列互补性与靶多核苷酸特异性杂交的第一3' 端和不通过序列互补性与靶多核苷酸特异性杂交的包含第一共同序列的第一5' 端; (ii) 第二引物,其包含通过序列互补性与多联体特异性杂交的第二3' 端和不通过序列互补性与多联体特

异性杂交的包含第二共同序列的第二5'端,其中第一共同序列和第二共同序列各自在5'端包含至少10个连续核苷酸并且当最佳比对时至少90%相同,并且该多联体为第一引物的延伸产物;以及(iii)第三引物,其具有通过序列互补性与第一共同序列或第二共同序列特异性杂交的序列;以及(c)向接收者发送报告的报告生成器,其中该报告包含所述至少三种引物的序列。在一些实施方案中,第一共同序列和第二共同序列是相同的。在一些实施方案中,第一共同序列、第二共同序列和第三引物的杂交序列均具有在彼此±5°C内的解链温度(Tm)。在一些实施方案中,沿靶多核苷酸从5'至3'对应于(i)与第一3'端互补的序列、(ii)与第二3'端相同的序列以及(iii)(i)与(ii)之间的间插序列的靶多核苷酸序列部分的组合长度为75个或更少的核苷酸。

[0145] 在一些实施方案中,所述计算机包含一个或多个处理器。处理器可以与一个或多个控制器、计算单元和/或计算机系统的其他单元相关联,或者根据需要植入围件中。如果在软件中实现,则例程可存储在任何计算机可读存储器中,诸如存储在RAM、ROM、闪速存储器、磁盘、激光盘或其他存储介质中。同样地,该软件可经由任何已知的传送方法而传送至计算设备,所述传送方法例如包括通过诸如电话线、因特网、无线连接等通信信道,或者经由诸如计算机可读盘、闪存驱动器等可移动介质。各个步骤可实现为各个块、操作、工具、模块或技术,而所述各个块、操作、工具、模块和技术继而可在硬件、固件、软件或其组合中实现。当在硬件中实现时,所述块、操作、技术等之中的一些或全部块、操作、技术可例如在定制集成电路(IC)、专用集成电路(ASIC)、现场可编程逻辑阵列(FPGA)、可编程逻辑阵列(PLA)等中实现。在一些实施方案中,所述计算机被配置用于接收客户请求以设计用于扩增指定靶序列(其也可以由客户提供)的引物。所述计算机可直接(例如,通过输入设备如由客户操作的键盘、鼠标或触摸屏,或通过用户输入客户请求)或间接地(例如,通过有线或无线连接,包括经互联网)接收客户请求。

[0146] 在一些实施方案中,所述系统包含向接受者发送报告的报告生成器,其中报告包含所述至少三种引物的序列。该报告生成器可响应于客户请求而自动发送报告。或者,该报告生成器可以响应于来自操作者的指令而发送报告。可以使用任何合适的通信介质将报告传送给本地或远程位置的接收者。例如,该通信介质可以是网络连接、无线连接或互联网连接。报告可通过这样的网络或连接(或任何其他合适的传送信息的手段,包括但不限于邮寄体检报告,诸如打印输出)传送,以供接收和/或由接收者查阅。接收者可以是但不限于客户或电子系统(例如,一个或多个计算机,和/或一个或多个服务器)。在一些实施方案中,报告生成器将报告发送给接收者的设备,诸如个人计算机、电话、平板电脑或其他设备。该报告可以被在线查看、保存在接收者的设备上或打印。

[0147] 在一方面,本公开内容提供了一种包含代码的计算机可读介质,该代码一旦由一个或多个处理器执行,就实现根据任意本文公开方法的方法。计算机可读介质可以采取许多形式,包括但不限于有形存储介质、载波介质或物理传输介质。非易失性存储介质例如包括光盘或磁盘(诸如任何计算机中的任何存储设备)等,诸如可用于实现计算步骤、处理步骤的存储介质等。易失性存储介质包括动态存储器,例如计算机的主存储器。有形传输介质包括同轴线缆、铜线和光纤,包括构成计算机系统内的总线的导线。载波传输介质可以采取电信号或电磁信号或者声波或光波的形式,诸如在射频(RF)和红外(IR)数据通信期间所生成的电信号或电磁信号或者声波或光波。因此,计算机可读介质的常见形式包括,例如:软

盘、柔性盘、硬盘、磁带、任何其他磁介质、CD-ROM、DVD或DVD-ROM、任何其他光学介质、穿孔卡片纸带、任何具有孔洞图案的其他物理存储介质、RAM、PROM和EPROM、FLASH-EPROM、任何其他存储器芯片或盒、载波传输数据或指令、传输此类载波的缆线或链路,或者任何可让计算机从中读取编程代码和/或数据的其他介质。这些计算机可读介质的形式中的许多形式可参与向处理器传送一个或多个序列的一个或多个指令以供执行。

[0148] 实施例

[0149] 下列实施例是为了说明本发明的各种实施方案而给出的,并非意在以任何方式限制本发明。这些实施例以及其中描述的方法目前均代表优选的实施方案,是示例性的,并且不应作为对本发明范围的限制。本领域技术人员将会想到由权利要求范围限定的本发明精神内所包含的其变化和其他用途。

[0150] 实施例1:来自一个RCA扩增循环和多个RCA扩增循环的产物的比较

[0151] 将基因组DNA超声处理至约180bp的平均片段大小。用0.9x Ampure珠将片段化DNA纯化以去除小于100bp的片段。然后将超声处理的基因组DNA连接以形成环状靶多核苷酸。为了连接,12 μ l经纯化的DNA片段(>10ng)通过在95°C下加热30秒并在冰上冷却2分钟而变性。然后,将含有2 μ l 10x CircLigase缓冲液、4 μ l 5M甜菜碱、1 μ l 50mM MnCl₂和1 μ l CircLigase II的8 μ l连接混合物添加至变性的DNA样品,并且将该反应在60°C下温育至少12小时。在连接过程结束时,通过外切核酸酶处理步骤去除剩余的线性单链DNA分子。对于外切核酸酶处理,将连接产物在80°C下加热45秒,然后添加1 μ l外切核酸酶混合物(ExoI 20U/ μ l:ExoIII 100U/ μ l,比例为1:2)。将样品在热循环仪上于37°C下温育30分钟,然后在80°C下温育20分钟。外切核酸酶处理后,向每个管添加1 μ l 50mM EDTA。

[0152] 使环状靶多核苷酸经受一个RCA扩增循环或多重RCA扩增。对于一个RCA扩增循环和多个RCA扩增循环,均使用10ng环化DNA样品作为起始材料。对于每个反应,向每10ng DNA样品添加0.34 μ L 1M Tris-HCl(pH9.2)、1 μ l 100mM MgSO₄、2.78 μ l 180mM (NH₄)₂SO₄、0.75 μ L dNTP混合物(各25mM)、0.5 μ l 10%吐温20、1.20 μ l 1M KC1、2 μ l 10 μ M背靠背正向和反向引物、18.28 μ l水。将反应在80°C下加热1分钟,并在63°C下温育5分钟,然后冷却至4°C。接下来,向每个反应添加15个单位的Bst 2.0热启动DNA聚合酶。对于一个RCA扩增循环,将反应在63°C下温育2小时。对于多个RCA扩增循环,将反应在具有以下程序的热循环仪中温育:8个循环的60°C持续30秒;70°C持续4.5分钟;94°C持续20秒;和58°C持续10秒。在每两个循环结束时,添加15个单位的Bst 2.0热启动DNA聚合酶。

[0153] 根据制造商关于剩余洗涤步骤的说明,通过添加50 μ l Ampure珠来纯化所有扩增产物。对于洗脱,将55 μ l洗脱缓冲液添加至每个管,并将珠在65°C下温育5分钟。短暂旋转后,将管返回到磁体。从每个反应回收约50 μ l洗脱产物。

[0154] 对于来自一个RCA循环的扩增产物的衔接子附接,每50 μ l洗脱液与5.7 μ l 10x AccuPrime缓冲液、1 μ l 25 μ M衔接子引物(其与在先前扩增反应中使用的引物的3'端处的共同序列互补)和2个单位的AccuPrime HiFi Taq聚合酶混合。通过使用以下PCR程序进行扩增来附接衔接子:95°C持续2分钟;30个循环的95°C持续30秒、60°C持续30秒、72°C持续2.5分钟;以及在72°C下持续7分钟的最终延伸。对于来自多个RCA循环的扩增产物,使用Illumina的KAPA hyper prep试剂盒来附接测序衔接子。通过琼脂糖凝胶或下一代测序对PCR扩增的文库产物进行分析。为了对测序数据进行生物信息学分析,从HiSeq运行获得

FASTQ文件。将FASTQ文件与包含靶序列的参考文件比对。基于测序数据计算插入物大小。

[0155] 如图5的琼脂糖凝胶中定性示出的,与一个RCA循环相比,利用B2B引物和温度循环的扩增产生包含更多多核苷酸拷贝的扩增子,以及具有更长靶多核苷酸的扩增子。在图6中,该表提供了通过琼脂糖凝胶的强度分析进行的对包含一个重复(约150bp大小)、两个重复(约300bp)或3个重复(约450bp)的产物之比的半定量比较。与一个RCA循环相比,多个RCA循环减少了只有1个重复的产物的相对量。

[0156] 扩增产物的测序分析还表明,与一个RCA循环相比,使用多个RCA循环时较大DNA片段的比例增加。图7以读取计数示出了片段大小分布,而图8以分子百分比示出了所观察到的分子大小的分布。

[0157] 实施例2:使用具有茎结构的引物或不具有茎结构的引物进行多个RCA循环产生的产物的比较

[0158] 将基因组DNA超声处理至约150bp的平均片段大小。用0.9x Ampure珠将片段化DNA纯化以去除小于100bp的片段。然后将超声处理的基因组DNA连接以形成环状靶多核苷酸。为了连接,12 μ l经纯化的DNA片段(>10ng)通过在95°C下加热30秒并在冰上冷却2分钟而变性。然后,将含有2 μ l 10x CircLigase缓冲液、4 μ l 5M甜菜碱、1 μ l 50mM MnCl₂和1 μ l CircLigase II的8 μ l连接混合物添加至变性的DNA样品,并且将该反应在60°C下温育至少12小时。在连接过程结束时,通过外切核酸酶处理步骤去除剩余的线性单链DNA分子。对于外切核酸酶处理,将连接产物在80°C下加热45秒,然后添加1 μ l外切核酸酶混合物(ExoI 20U/ μ l:ExoIII 100U/ μ l,比例为1:2)。将样品在热循环仪上于37°C下温育30分钟,然后在80°C下温育20分钟。外切核酸酶处理后,向每个管添加1 μ l 50mM EDTA。

[0159] 使环状靶多核苷酸经受能够形成19聚体茎结构的引物或设计成无茎结构的引物的多个RCA扩增循环。能够形成茎结构的示例性常见序列包括表1中提供的那些序列及其任何片段。

[0160] 对于每个反应,向每10ng DNA样品添加0.34 μ L 1M Tris-HCl(pH9.2)、1 μ l 100mM MgSO₄、2.78 μ l 180mM (NH₄)₂SO₄、0.75 μ L dNTP混合物(各25mM)、0.5 μ l 10%吐温20、1.20 μ l 1M KCl、2 μ l 10 μ M背靠背正向和反向引物、18.28 μ l水。将反应在80°C下加热1分钟,并在63°C下温育5分钟,然后冷却至4°C。接下来,向每个反应添加15个单位的Bst 2.0热启动DNA聚合酶。将反应在具有以下程序的热循环仪中温育:8个循环的60°C持续30秒;70°C持续4.5分钟;94°C持续20秒;和58°C持续10秒。在每两个循环结束时,添加15个单位的Bst 2.0热启动DNA聚合酶。

[0161] 根据制造商关于剩余洗涤步骤的说明,通过添加50 μ l Ampure珠来纯化所有扩增产物。对于洗脱,将55 μ l洗脱缓冲液添加至每个管,并将珠在65°C下温育5分钟。短暂旋转后,将管返回到磁体。从每个反应回收约50 μ l洗脱产物。

[0162] 使用Illumina的KAPA hyper prep试剂盒来附接测序衔接子。通过琼脂糖凝胶分析PCR扩增的文库产物,并进一步收集在大小范围550bp-1000bp内的产物以供测序。通过测序来分析所得的扩增产物。为了对测序数据进行生物信息学分析,从HiSeq运行获得FASTQ文件。将FASTQ文件与包含靶序列的参考文件比对。基于测序数据计算插入物大小。

[0163] 如表3所示,利用包含茎结构的B2B引物和温度循环进行扩增产生了含有更多多核苷酸拷贝的扩增子。表3列出了包含多于一个重复的测序读取的百分比。与无茎引物相比,

使用具有19个碱基茎的引物的扩增显著增加了具有多于一个重复的读取的百分比。

[0164] 表3

用于RCA的引物	具有重复的读取的百分比
19聚体茎的引物	64.62%
无茎的引物	31.49%

[0166] 实施例3:从混合基因组DNA样品检测低频融合等位基因

[0167] 在许多癌症类型中观察到染色体重排。本实施例描述了用于使用环化DNA分子和背靠背(B2B)引物设计来检测融合等位基因的方法。该方法能够在没有预先知晓‘配偶体’基因的情况下进行DNA样品的融合检测，并且可应用于筛选无细胞DNA或基因组DNA样品中的基因重排事件。

[0168] 将来自含有50% EML4/ALK融合等位基因的EML4/ALK DNA标准参照物(HD664 Horizon Diagnostics)的基因组DNA和参照基因组DNA超声处理至约150bp的平均片段大小。用0.9x Ampure珠将片段化DNA纯化以去除小于100bp的片段。为了连接,12μl经纯化的DNA片段(>10ng)通过在95°C下加热30秒并在冰上冷却2分钟而变性。然后,将含有2μl 10x CircLigase缓冲液、4μl 5M甜菜碱、1μl 50mM MnCl₂和1μl CircLigase II的8μl连接混合物添加至变性的DNA样品,并且将该反应在60°C下温育至少12小时。在连接过程结束时,通过外切核酸酶处理步骤去除剩余的线性单链DNA分子。对于外切核酸酶处理,将连接产物在80°C下加热45秒,然后添加1μl外切核酸酶混合物(ExoI 20U/μl:ExoIII 100U/μl,比例为1:2)并在热循环仪上于37°C下温育30分钟,然后在80°C下温育20分钟。外切核酸酶处理后,向每个管添加1μl 50mM EDTA。

[0169] 然后将超声处理的基因组DNA连接以形成环状靶多核苷酸。通过qPCR对两个环化DNA样品、HD664和参照基因组DNA进行定量,之后混合在一起以获得以浓度计2.5%、0.5%、0.05%和0%的融合等位基因(图9)。对于滚环扩增,使用10ng每种混合DNA样品作为起始材料。对于每个反应,向每10ng DNA样品添加0.34μL 1M Tris-HCl (pH9.2)、1μl 100mM MgSO₄、2.78μl 180mM (NH4) 2SO₄、0.75μL dNTP混合物(各25mM)、0.5μl 10%吐温20、1.20μl 1M KC1、被特别设计用于靶向ALK/EML4融合区域的2μl 10μM正向和反向引物(表4中提供的引物序列)、18.28μl水。将反应在80°C下加热1分钟,并在63°C下温育5分钟,之后冷却至4°C。向每个反应添加15个单位的Bst 2.0热启动DNA聚合酶,然后将每个反应在63°C下温育2小时。

[0170] 表4. 用于ALK融合检测的正向和反向引物序列

寡核苷酸名称	寡核苷酸序列
HD664_ALK_F	CCTTGGCACCCGAGAATTCCATTGAGGGATGGCACCATAT
HD664_ALK_R	GTTCAGAGTTCTACAGTCCGACGATCGGGACAGGATAATAGGAG CTAACCA

[0172] 根据制造商关于剩余洗涤步骤的说明,通过添加50μl Ampure珠来纯化所有扩增产物。对于洗脱,将55μl洗脱缓冲液添加至每个管,并将珠在65°C下温育5分钟。短暂旋转后,将管返回到磁体。从每个反应回收约50μl洗脱产物。将每50μl洗脱液与5.7μl 10x AccuPrime缓冲液、1μl 25uM每种Illumina测序文库衔接子引物,和2个单位的AccuPrime HiFi Taq聚合酶混合。通过扩增的衔接子衔接使用以下PCR程序:95°C持续2分钟;25个循环

的95℃持续30秒、60℃持续30秒、72℃持续2.5分钟；以及在72℃下持续7分钟的最终延伸。通过琼脂糖凝胶分析PCR产物，并收集在大小范围550bp-1000bp内的产物以供测序。

[0173] 为了对测序数据进行生物信息学分析，从HiSeq运行获得FASTQ文件。将FASTQ文件与包含靶序列的参考文件比对。如图10所示，在样品中发现了具有2.5%、0.5%或0.05%的融合等位基因掺入的融合等位基因，但在样品中未发现具有0%融合等位基因掺入的融合等位基因。

[0174] 实施例4：在利用B2B引物的多重RCA中的靶标检测

[0175] 在多个RCA和利用B2B引物的扩增中检测样品中多种靶多核苷酸的成分靶多核苷酸。将从人血清样品 (H6914, Sigma) 提取的20ng对照cfDNA重悬于共12μl的Tris pH 8缓冲液中，并通过在95℃下加热30秒并在冰上冷却2分钟而变性。然后，将含有2μl 10x CircLigase缓冲液、4μl 5M甜菜碱、1μl 50mM MnCl₂和1μl CircLigase II的8μl连接混合物添加至变性的DNA样品，并且将反应在60℃下温育至少12小时。在连接过程结束时，通过外切核酸酶处理步骤去除剩余的线性单链DNA分子。对于外切核酸酶处理，将连接产物在80℃下加热45秒，然后添加1μl外切核酸酶混合物 (ExoI 20U/μl:ExoIII 100U/μl，比例为1:2)。将样品在热循环仪上于37℃下温育30分钟，然后在80℃下温育20分钟。外切核酸酶处理后，向每个管添加1μl 50mM EDTA。

[0176] 使环状靶多核苷酸经受滚环扩增以及随后用B2B引物进行的扩增。表5中提供了B2B引物的实例。

[0177] 表5. 背靠背B2B引物的实例

基 因 名 称	B2B 正向引物 名 称	正向引物序列	B2B 反向引物 名 称	反向引物序列
[0178]	BRAF	GTTCAGAGTTCT ACAGTCCGACGA TCCAGTTGAAC AGTTGTCTGGAT C	BRAF-BX1b	CCTTGGCACCCG AGAATTCCAAAA CTGATGGGACCC ACTCC
	CYP2	GTTCAGAGTTCT ACAGTCCGACGA TCTTCCCACATAC ATTGATTATTC	CYP2c19-BXd	CCTTGGCACCCG AGAATTCCATGG GAAAATTATTGC ATATCTAAGAG
	EGFR	GTTCAGAGTTCT ACAGTCCGACGA TCCTTCTCACCT TCTGGGATCC	EGFR-BX1b	CCTTGGCACCCG AGAATTCCAAAA TTCCCGTCGCTAT CAAG
	EGFR	GTTCAGAGTTCT ACAGTCCGACGA TCCCACATACGTA GGCTTCCTG	EGFR-BX2b	CCTTGGCACCCG AGAATTCCAATG GCCAGCGTGGAC AAC
	EGFR	GTTCAGAGTTCT ACAGTCCGACGA TCGACATAGTCC AGGAGGCAGC	EGFR-BX3b	CCTTGGCACCCG AGAATTCCATGTC CGGAAACACAAA GAC
	EGFR	GTTCAGAGTTCT ACAGTCCGACGA TCAAGCGACGGT CCTCCAAG	EGFR-BX4b	CCTTGGCACCCG AGAATTCCATGG CAGCCAGGAACG TAC
	EGFR	GTTCAGAGTTCT ACAGTCCGACGA TCAGTACGTTCCCT GGCTGCC	EGFR-BX5b	CCTTGGCACCCG AGAATTCCAAC ACCGCAGCATGT CAAG
	EGFR	GTTCAGAGTTCT ACAGTCCGACGA	EGFR-BX6b	CCTTGGCACCCG AGAATTCCAAG

[0179]

		TCATCCACTTGAT AGGCACCTG		TGGATGGCATTG GAATC
EGFR	EGFR-BX7a	GTTCAGAGTTCT ACAGTCCGACGA TCTCTCGCTGGC AGGGATTG	EGFR-BX7b	CCTTGGCACCCG AGAATTCCACCT GGAGAAAGGAGA ACGC
EGFR	EGFR-BX9a	GTTCAGAGTTCT ACAGTCCGACGA TCAACTTGGGC GAATCTGC	EGFR-BX9b	CCTTGGCACCCG AGAATTCCAAGT TCCGTGAGTTGAT CATCG
EGFR	EGFR-BX10a	GTTCAGAGTTCT ACAGTCCGACGA TCTTGGAGTCTGT AGGACTTGGC	EGFR-BX10b	CCTTGGCACCCG AGAATTCCAACCT CTACCGTGCCCTG ATG
EGFR	EGFR-BX11a	GTTCAGAGTTCT ACAGTCCGACGA TCCTGCTGTGGG ATGAGGTACTC	EGFR-BX11b	CCTTGGCACCCG AGAATTCCACAC AGCAGGGCTTCTT CAG
EGFR	EGFR-BX12a	GTTCAGAGTTCT ACAGTCCGACGA TCCATGGAATGC TTGTACCACATC	EGFR-BX12b	CCTTGGCACCCG AGAATTCCACAT GGCAACTTCTCT GTTC
EGFR	EGFR-BX4c	GTTCAGAGTTCT ACAGTCCGACGA TCCTGGCAGCCA GGAACGTACT	EGFR-BX4d	CCTTGGCACCCG AGAATTCCACGA CGGTCCCTCCAAGT AGTTC
EGFR	EGFR-BX5c	GTTCAGAGTTCT ACAGTCCGACGA TCACACCGCAGC ATGTCAAGATC	EGFR-BX5d	CCTTGGCACCCG AGAATTCCAAGT ACGTTCCCTGGCTG CCAG
EGFR	EGFR_hot1-B Xc	GTTCAGAGTTCT ACAGTCCGACGA TCGCACGGTGTA TAAGGTAAGGTC C	EGFR_hot1-BX d	CCTTGGCACCCG AGAATTCCAGAA TTCAGTTCCATTC AAGATCC
KRAS	KRAS-BX1a	GTTCAGAGTTCT ACAGTCCGACGA TCAAGAGTCCT TGACGATACAGC	KRAS-BX1b	CCTTGGCACCCG AGAATTCCATCTT GCCTACGCCACC AG
KRAS	KRAS_c181-B Xc	GTTCAGAGTTCT ACAGTCCGACGA TCTCGAGAATAT CCAAGAGACAGG	KRAS_c181-B Xd	CCTTGGCACCCG AGAATTCCAAGA GGAGTACAGTGC AATGAGG
PIK3CA	PIK3CA-BX1a	GTTCAGAGTTCT ACAGTCCGACGA TCGCTTGAGCT GTTCTTGTCAATT	PIK3CA-BX1b	CCTTGGCACCCG AGAATTCCAAAA GCAATTCTACAC GAGATCC
PIK3CA	PIK3CA-BX2a	GTTCAGAGTTCT ACAGTCCGACGA TCTTAATTGTGT GGAAGATCCAAT	PIK3CA-BX2b	CCTTGGCACCCG AGAATTCCAATT AAACAGCATGCA TTGAACGT

[0180]

		C		
[0180]	PIK3CA	PIK3CA-BX1c	GTTCAGAGTTCT ACAGTCCGACGA TCCCTCTCTCTGA AATCACTGAGC	PIK3CA-BX1d CCTTGGCACCCG AGAATTCCAGAG GATCTCGTAG AAATTGC
	PTEN	PTEN-BX1a	GTTCAGAGTTCT ACAGTCCGACGA TCTGTTCTGCTA ACGATCTCTTG	PTEN-BX1b CCTTGGCACCCG AGAATTCCAAGG AGATATCAAGAG GATGGATTG
	PTEN	PTEN-BX2a	GTTCAGAGTTCT ACAGTCCGACGA TCCAGGAAATCC CATAGCAATAAT G	PTEN-BX2b CCTTGGCACCCG AGAATTCCATCCT GCAGAAAGACTT GAAGG
	PTEN	PTEN-BX3a	GTTCAGAGTTCT ACAGTCCGACGA TCGCTTGAATCC AAAAACCTTAAA AC	PTEN-BX3b CCTTGGCACCCG AGAATTCCAGGA TCAAAGCATAA AAACCATTAC
	PTEN	PTEN-BX3c	GTTCAGAGTTCT ACAGTCCGACGA TCGGATTCAAAG CATAAAAACCAT TAC	PTEN-BX3d CCTTGGCACCCG AGAATTCCAGCTT TGAATCCAAAAA CCTTAAAAC
	PTEN	PTEN-BX15c	GTTCAGAGTTCT ACAGTCCGACGA TCGATGTTAGTG ACAATGAACCTG ATC	PTEN-BX15d CCTTGGCACCCG AGAATTCCATGG TGTTACAGAAGTT GAAC
	TP53	TP53-BX1a	GTTCAGAGTTCT ACAGTCCGACGA TCCCTGACTCAG ACTGACATTCTC C	TP53-BX1b CCTTGGCACCCG AGAATTCCACAG GCCCTCTGTCTT GAAC
	TP53	TP53-BX2a	GTTCAGAGTTCT ACAGTCCGACGA TCATGTTCCGAG AGCTGAATGAG	TP53-BX2b CCTTGGCACCCG AGAATTCCAGAA CATCTCGAAGCG CTCAC
	TP53	TP53-BX3a	GTTCAGAGTTCT ACAGTCCGACGA TCTTAAAGGACC AGACCAGCTTC	TP53-BX3b CCTTGGCACCCG AGAATTCCATTAT GGTATAAGTTGG TGTTCTGAAG
	TP53	TP53-BX3c	GTTCAGAGTTCT ACAGTCCGACGA TCTTATGGTATA AGTTGGTGTCT GAAG	TP53-BX3d CCTTGGCACCCG AGAATTCCATTA AAGGACCAGACC AGCTTTC
	TP53	TP53-BX16c	GTTCAGAGTTCT ACAGTCCGACGA TCGCTCGACGCT AGGATCTGAC	TP53-BX16d CCTTGGCACCCG AGAATTCCACTCT GAGTCAGGAAAC ATTTTCAG

[0181] 对于每个反应,向每10ng DNA样品添加0.34μL 1M Tris-HCl (pH9.2)、1μL 100mM MgSO₄、2.78μL 180mM (NH₄)₂SO₄、0.75μL dNTP混合物(各25mM)、0.5μL 10%吐温20、1.20μL 1M KCl、2μL 10μM背靠背正向和反向引物(图11的靶标列表中列出的8个配对引物靶向特异

性靶标的混合物)、18.28 μ l水。将反应在80℃下加热1分钟，并在63℃下温育5分钟，然后冷却至4℃。接下来，向每个反应添加15个单位的Bst 2.0热启动DNA聚合酶。对于用B2B引物进行的等温扩增，将该反应在63℃下温育2小时。

[0182] 根据制造商关于剩余洗涤步骤的说明，通过添加50 μ l Ampure珠来纯化所有扩增产物。对于洗脱，将55 μ l洗脱缓冲液添加至每个管，并将珠在65℃下温育5分钟。短暂旋转后，将管返回到磁体。从每个反应回收约50 μ l洗脱产物。

[0183] 针对衔接子附接，将每50 μ l洗脱液与5.7 μ l 10x AccuPrime缓冲液、1 μ l 25uM衔接子引物(其与在等温和B2B扩增中使用的引物的3'端处的共同序列互补)和2个单位的AccuPrime HiFi Taq聚合酶混合。通过使用以下PCR程序进行扩增来附接衔接子：95℃持续2分钟；30个循环的95℃持续30秒、60℃持续30秒、72℃持续2.5分钟；以及在72℃下持续7分钟的最终延伸。通过琼脂糖凝胶对PCR产物进行分析，并进一步收集在大小范围550bp-1000bp内的产物以供测序。通过测序对所得的扩增产物进行分析。如图11所示的序列分析结果表明在多重反应中可检测到多种靶多核苷酸序列。

[0184] 尽管本文中已经示出并描述了本发明的优选实施方案，但对于本领域技术人员显而易见的是，这些实施方案仅以示例的方式提供。本领域技术人员在不脱离本发明的情况下现将想到多种变化、改变和替代。应当理解，本文中所述的本发明实施方案的各种替代方案均可用于实施本发明。旨在由以下权利要求来限定本发明的范围，并由此涵盖这些权利要求范围内的方法和结构及其等同物。

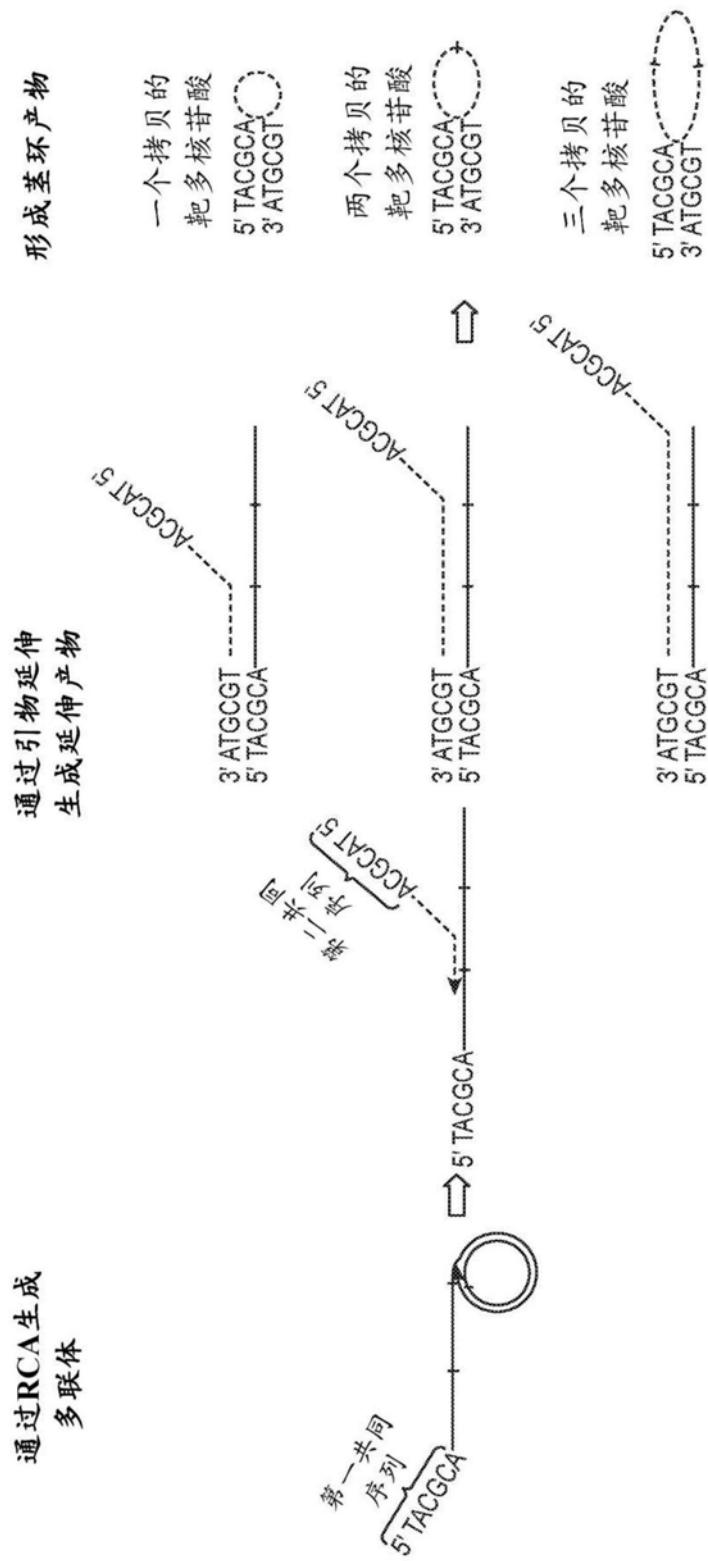


图1

• X: 点突变或 InDel (插入/缺失)

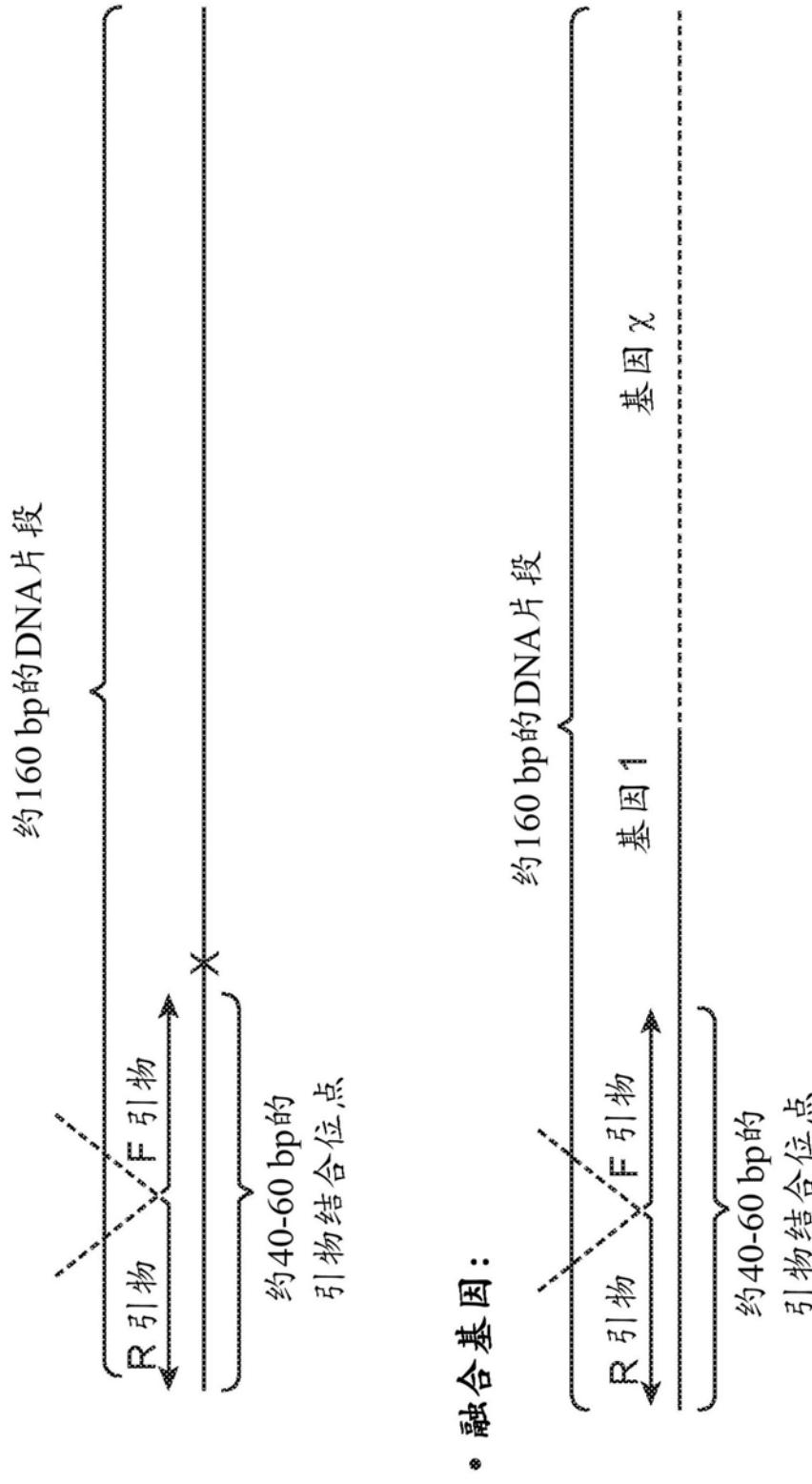


图2

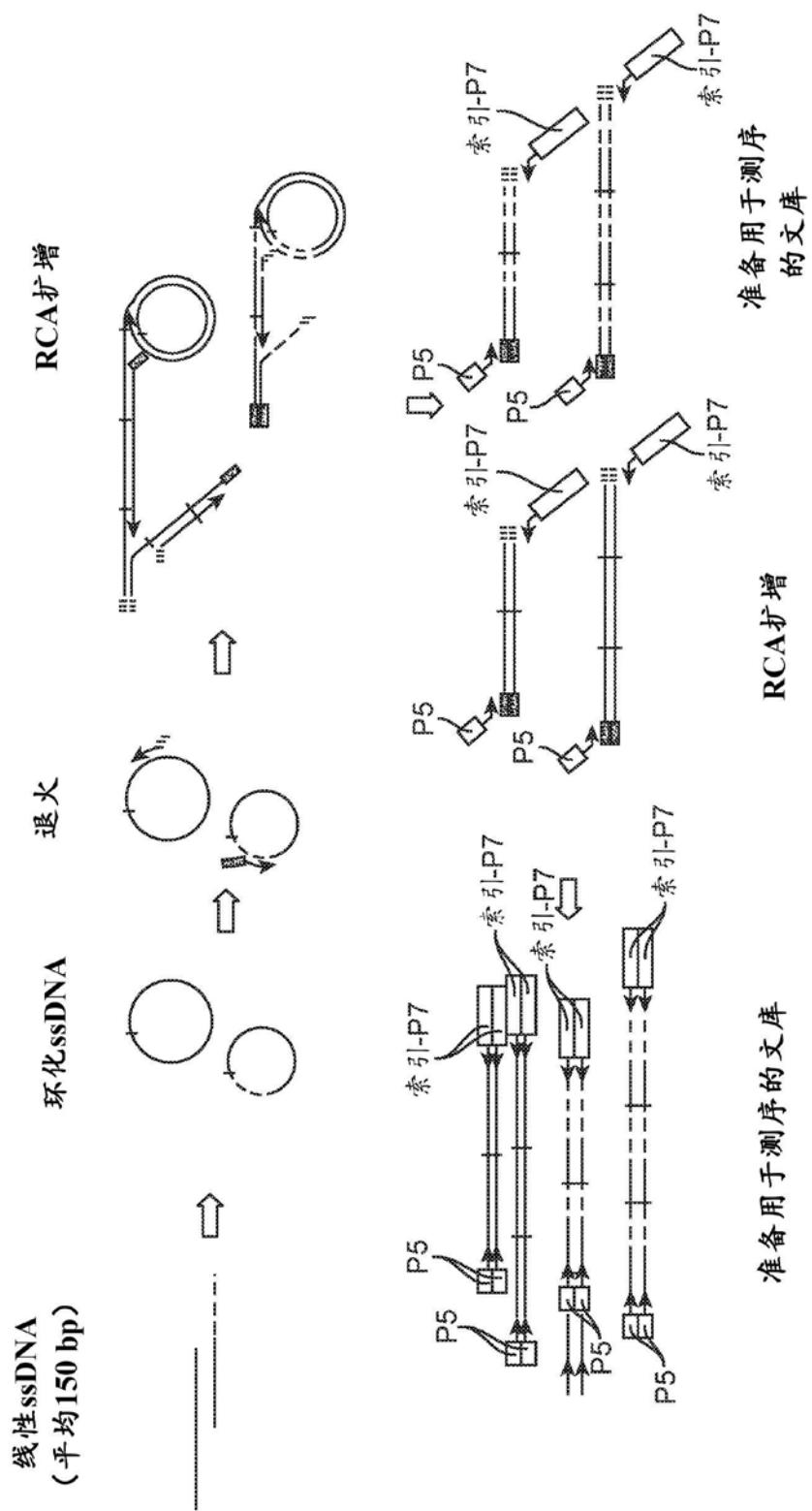


图3

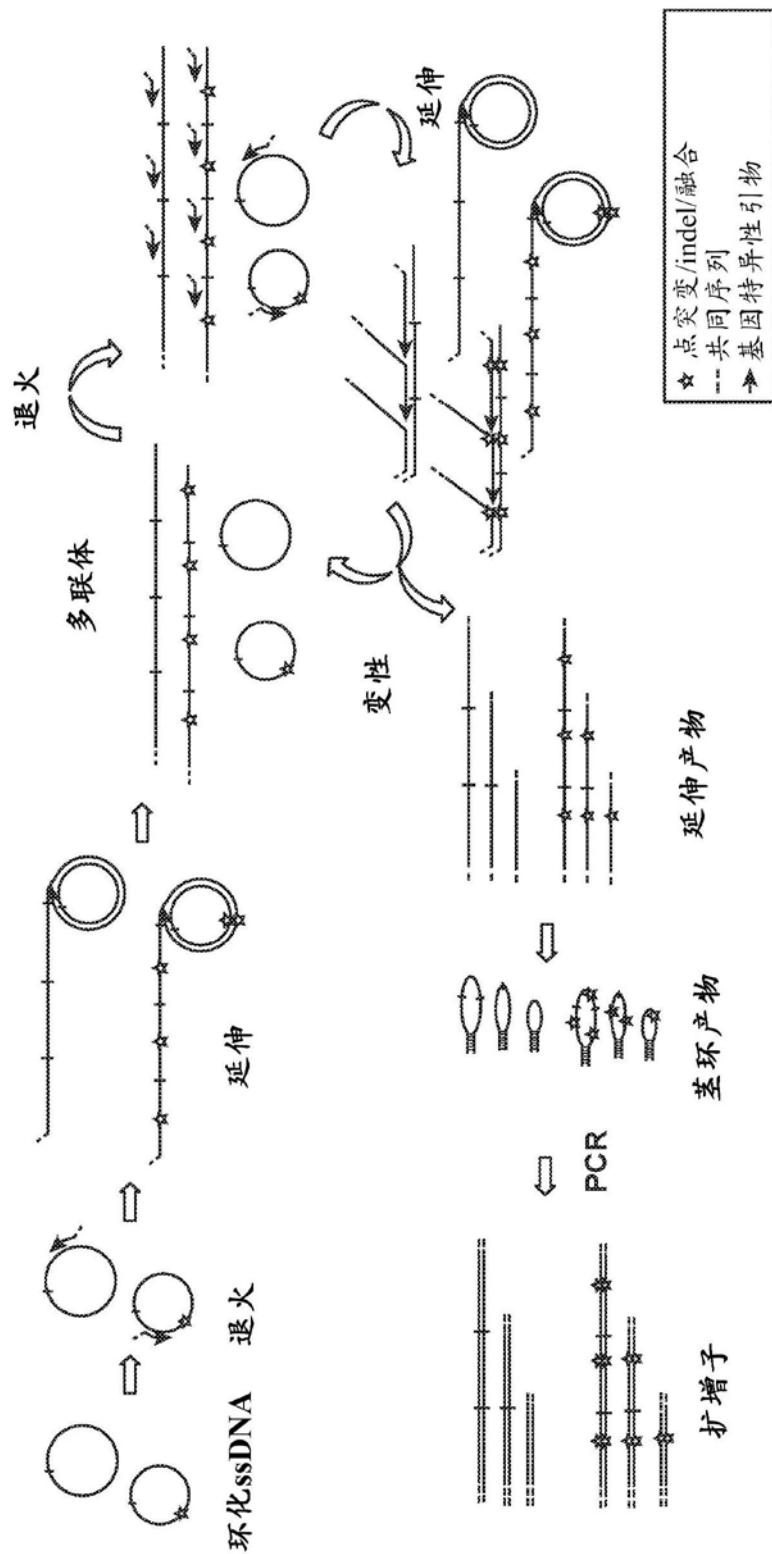


图4

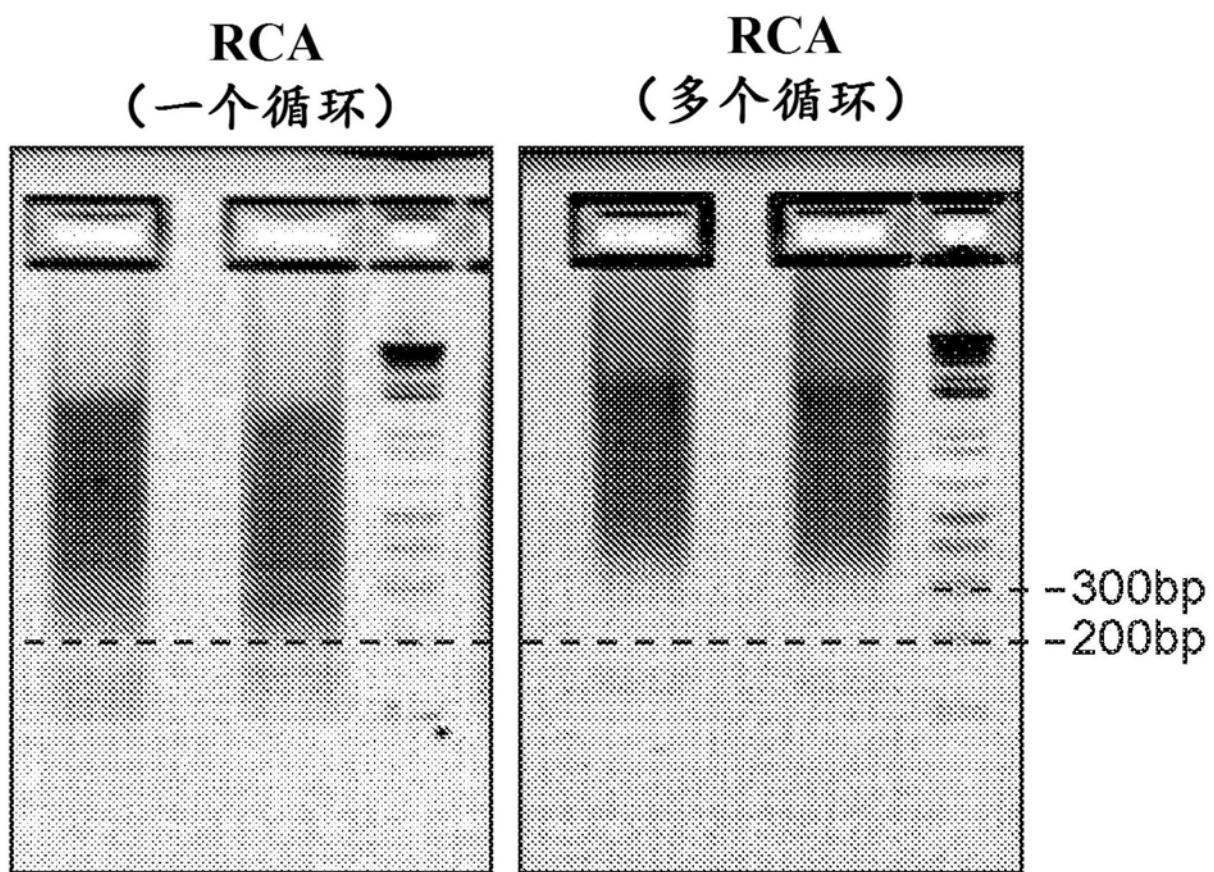


图5

质量比	RCA (一个循环)	RCA (多个循环)
2个重复: 1个重复	3.59	33.46
3个重复: 1个重复	4.97	40.68

图6

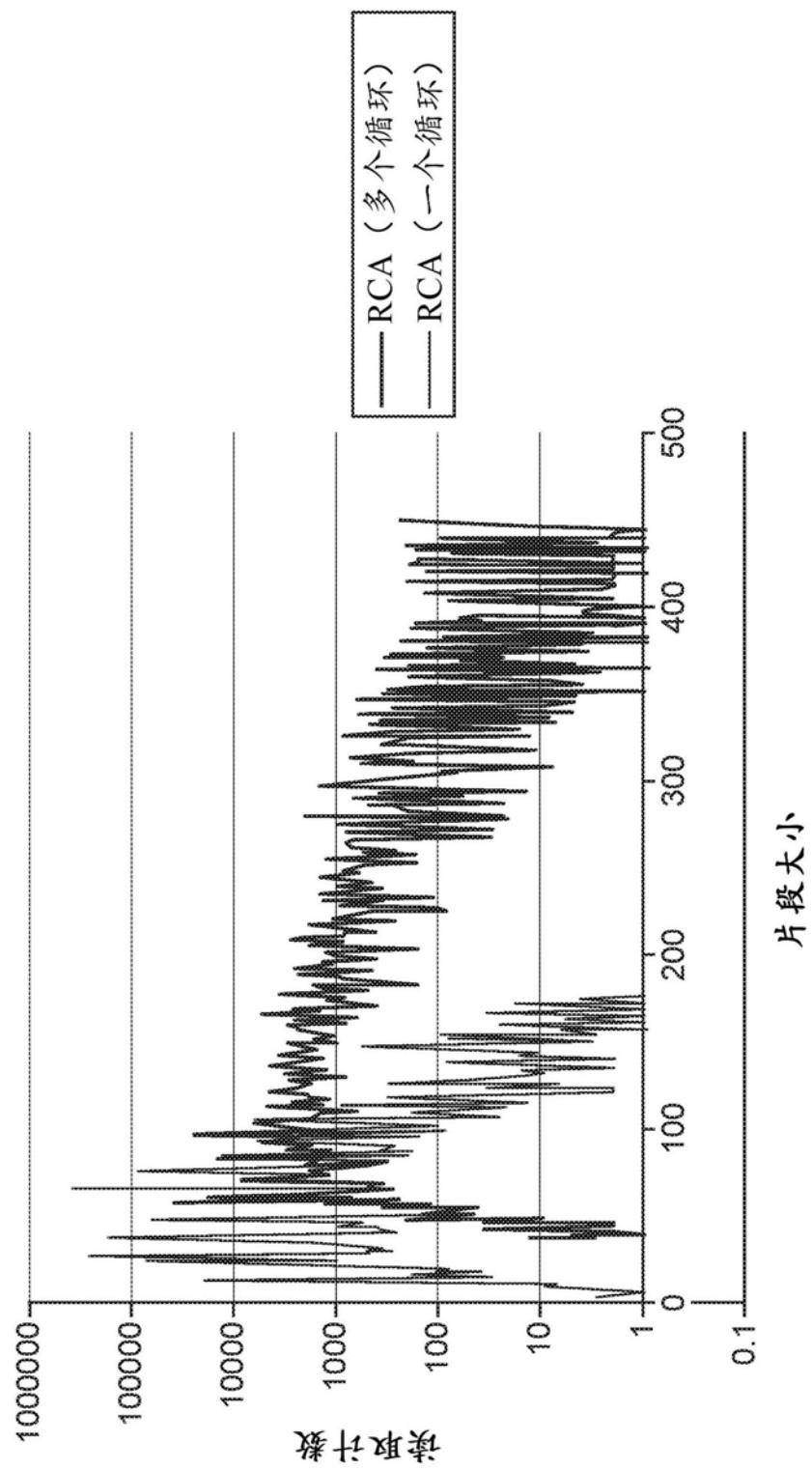


图7

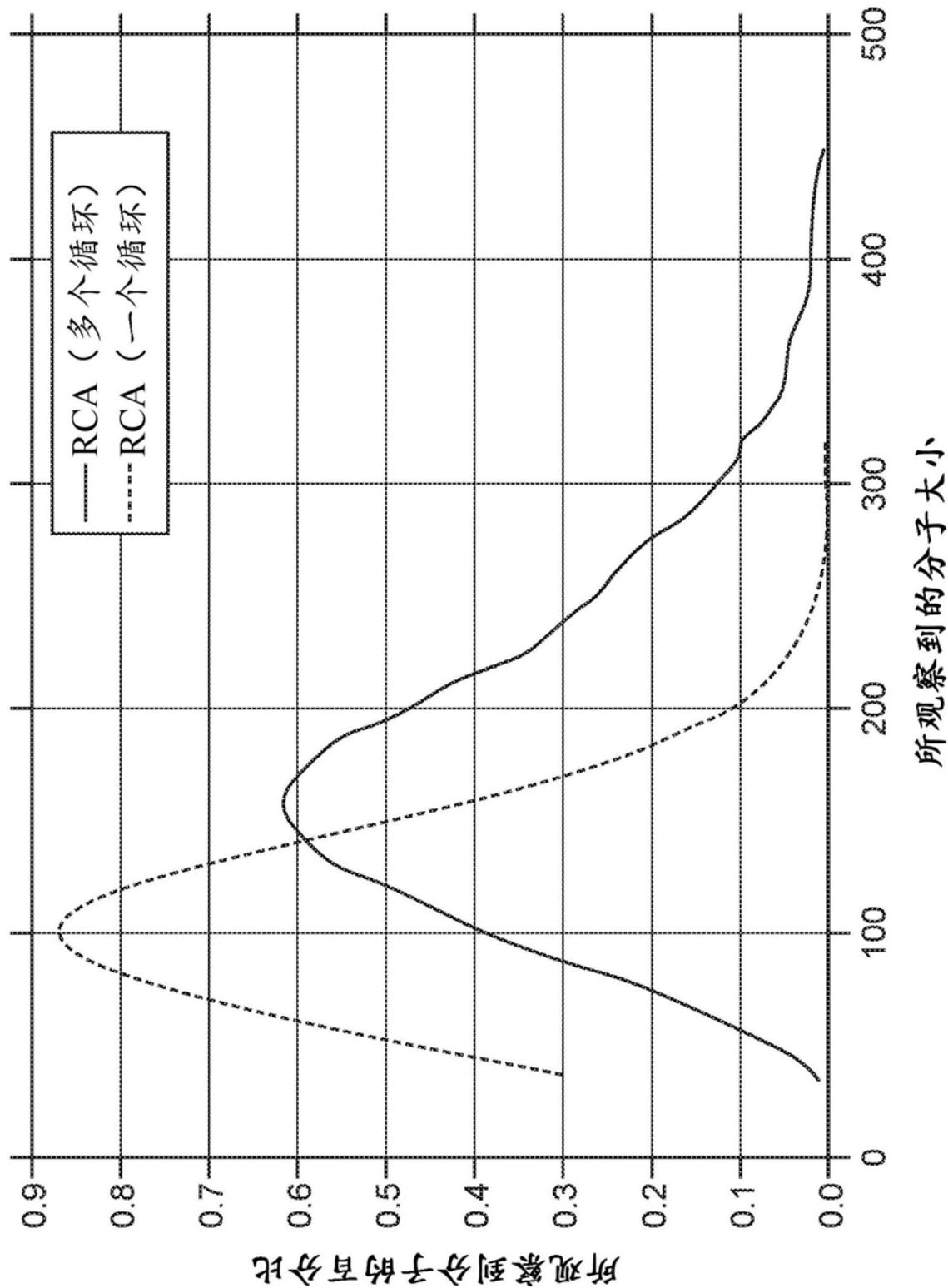
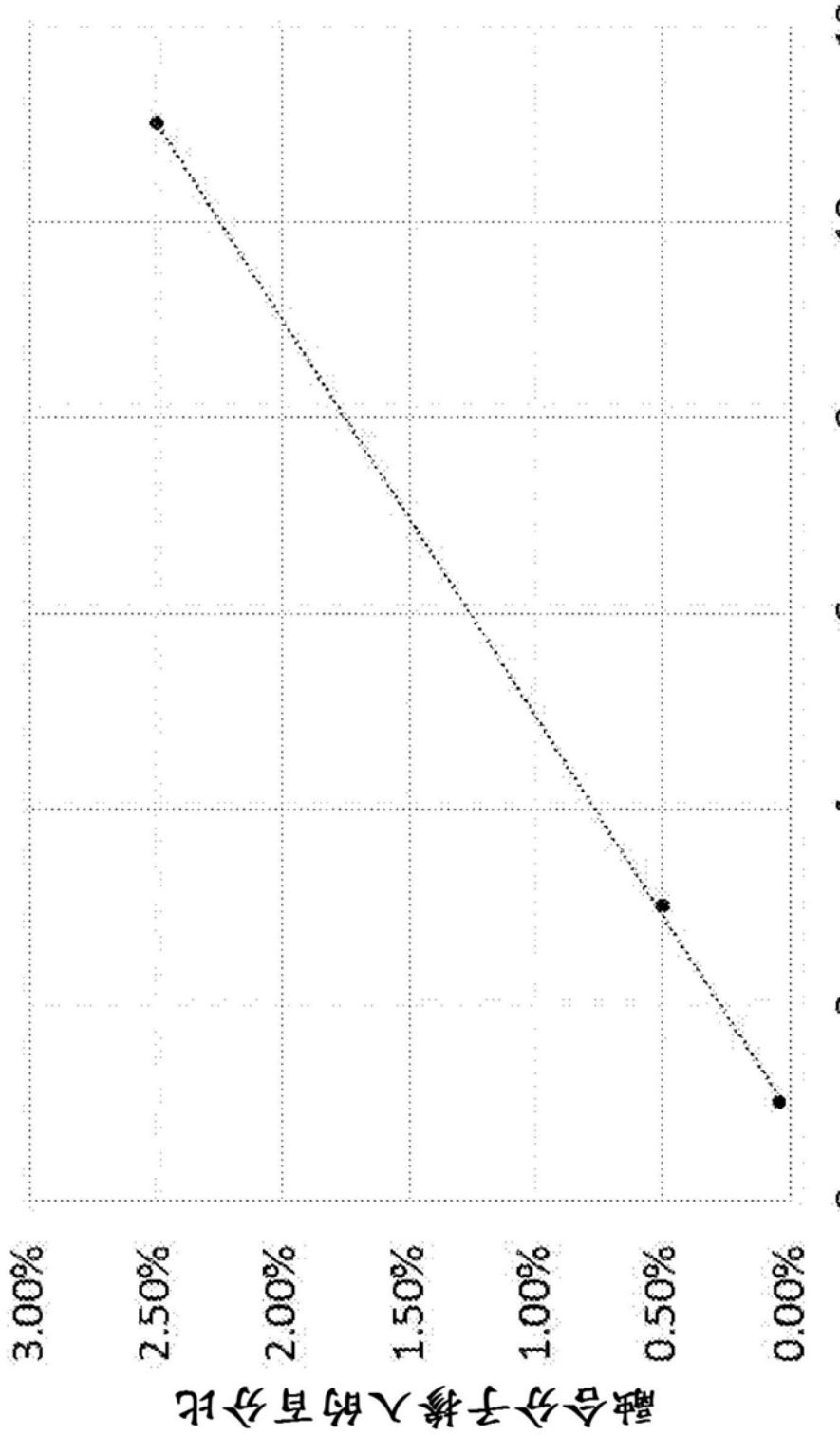


图8

插入融合等位基因的频率	2.50%	0.50%	0.12%	0.01	0.00
ng*融合DNA (具有50%融合等位基因的HD664 ^a)	0.50		0.12		
ng*对照DNA (野生型)	9.50		9.88	9.99	10.00
* ng环化DNA					

aHD664是指50% EML4/ALK DNA参考标准

图9



所观察到的融合分子的数目

图10

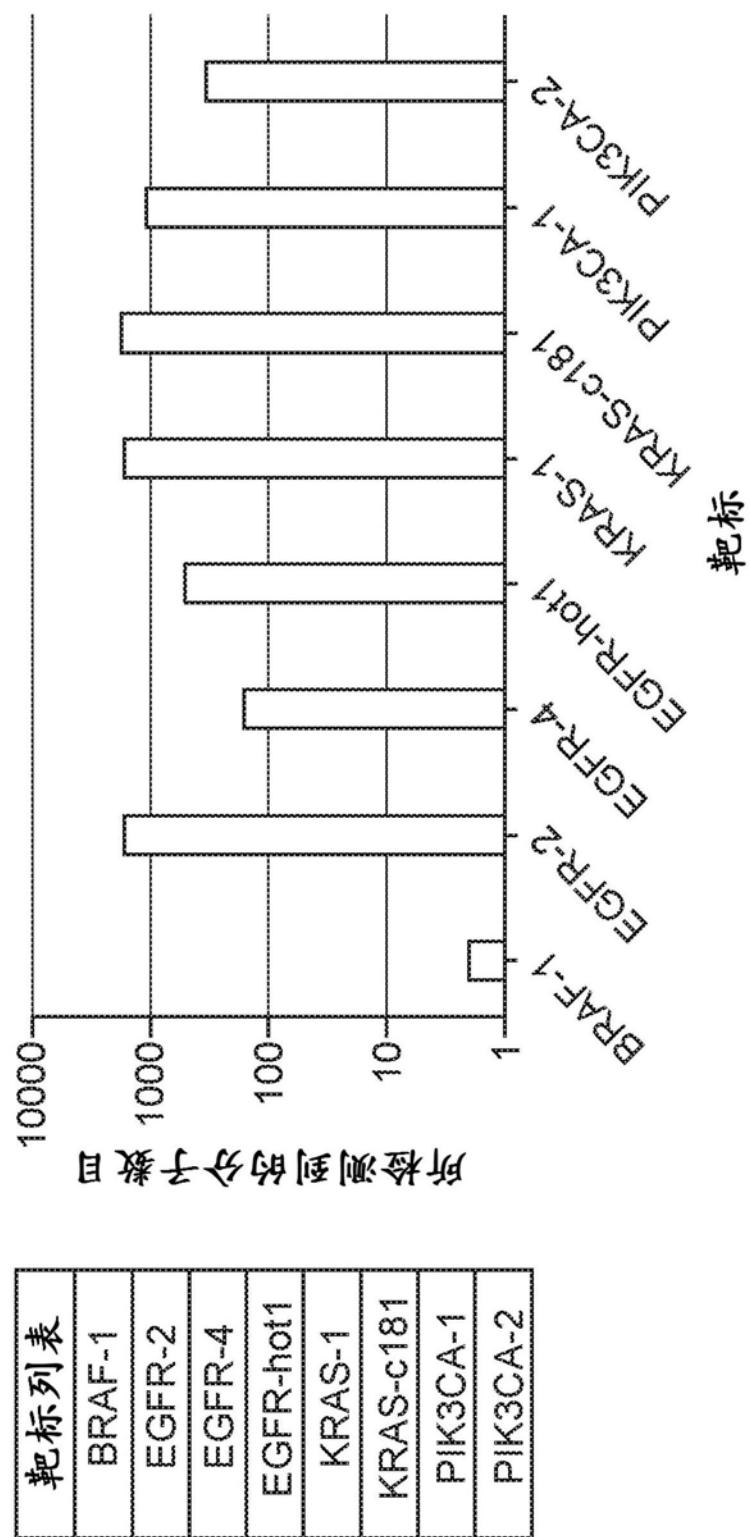


图11