



(12) 发明专利

(10) 授权公告号 CN 113106145 B

(45) 授权公告日 2024. 10. 11

(21) 申请号 202110396910.6

(22) 申请日 2019.04.02

(65) 同一申请的已公布的文献号
申请公布号 CN 113106145 A

(43) 申请公布日 2021.07.13

(66) 本国优先权数据
PCT/CN2018/081748 2018.04.03 CN

(62) 分案原申请数据
201980002533.4 2019.04.02

(73) 专利权人 广州燃石医学检验所有限公司
地址 510300 广东省广州市广州国际生物
岛螺旋四路7号3栋第六层601房

(72) 发明人 张之宏 郑涛 李冰思 邓望龙
汉雨生

(74) 专利代理机构 华进联合专利商标代理有限公司 44224

专利代理师 章上晓

(51) Int.Cl.
C12Q 1/6806 (2018.01)
C12N 15/10 (2006.01)
C40B 50/06 (2006.01)

(56) 对比文件
W0 2013112923 A1, 2013.08.01
审查员 姚进孝

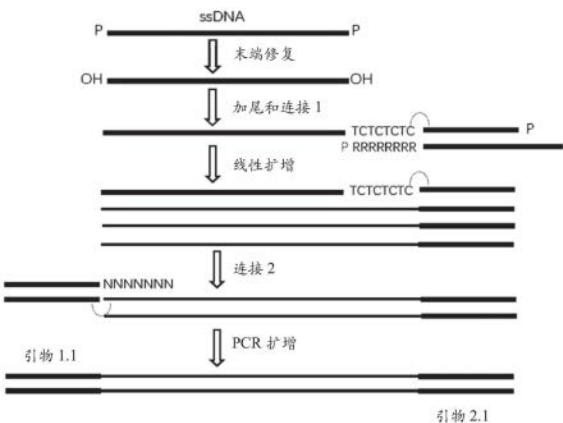
权利要求书2页 说明书42页
序列表6页 附图13页

(54) 发明名称

用于制备核酸文库的组合物和方法

(57) 摘要

本发明涉及用于制备核酸文库的组合物和方法。在多个方面,本发明公开提供了用于制备核酸文库,如用于多核苷酸测序的方法、组合物、反应混合物、试剂盒和系统。在一些实施方式中,制备方法包括加尾反应、用于连接接头的连接反应,和连接反应之间的扩增反应。



1. 用于制备多核苷酸文库的方法,其特征在于,所述方法包括:

a. 在第一加尾反应中,通过不依赖模板的聚合反应将第一尾添加至多个目标多核苷酸中的每一个,其中所述第一加尾反应包括第一接头,所述第一接头包含杂交至所述第一尾的悬突;

b. 在第一连接反应中,将第一接头链连接至所述第一尾;

c. 使包含所述第一接头链的目标多核苷酸扩增,所述扩增包括通过使杂交至所述第一接头链的第一引物延伸和线性扩增,所述扩增的循环次数为5-25次;所述扩增完成后,反应体系在95℃加热2分钟;

d. 在第二加尾反应中,通过不依赖模板的聚合反应将第二尾添加至多个所述扩增的目标多核苷酸中的每一个,其中所述第二加尾反应包括第二接头,所述第二接头包含杂交至所述第二尾的悬突;或者,所述第二加尾反应被省略;

和

e. 在第二连接反应中,将第二接头链连接至所述第二尾;

所述方法包含:

1) 将多核苷酸片段化以产生所述目标多核苷酸;

2) 使所述目标多核苷酸的一端或两端脱磷酸;

3) 使双-链多核苷酸变性为单链多核苷酸以产生所述目标多核苷酸。

2. 根据权利要求1所述的方法,其特征在于,所述方法可选地包含以下的一个或多个:

i) 所述目标多核苷酸包括无细胞多核苷酸,或其扩增产物;

ii) 所述第一加尾反应中目标多核苷酸的量为0.1-500ng;

iii) 所述目标多核苷酸的平均长度为50至600个核苷酸;

iv) 在所述脱磷酸步骤之前处理所述目标多核苷酸以差异修饰甲基化的胞嘧啶或未甲基化的胞嘧啶;

v) 通过聚合酶催化所述不依赖模板的聚合反应;

vi) 所述第一尾包含不同于所述第二尾的序列;

vii) 所述第一尾,所述第二尾或两者由一种或两种类型的核苷酸组成;

viii) 所述尾中的至少一个由从两种类型的核苷酸的池聚合的两种类型的核苷酸组成,其中所述池中的所述两种类型的核苷酸以相同或不同的量存在;

ix) 所述第一接头和所述第二接头包括多核苷酸序列不同的双-链区;

x) 所述第一和/或第二接头的悬突为3'-悬突;

xi) 所述第一和/或第二接头的悬突的长度为6至12个核苷酸;

xii) 所述第一加尾反应和所述第一连接反应在相同反应混合物中发生;

xiii) 所述第二加尾反应和所述第二连接反应在相同反应混合物中发生。

3. 根据权利要求2所述的方法,其特征在于,所述目标多核苷酸包括单链无细胞DNA(cfDNA)。

4. 根据权利要求2所述的方法,其特征在于,所述第一加尾反应中目标多核苷酸的量为1-100ng。

5. 根据权利要求4所述的方法,其特征在于,所述第一加尾反应中目标多核苷酸的量为5-50ng。

6. 根据权利要求2所述的方法, 其特征在于, 所述差异修饰包括用亚硫酸氢盐处理所述目标多核苷酸或基于酶的转化。

7. 根据权利要求2所述的方法, 其特征在于, 所述聚合酶为末端脱氧核苷酸转移酶(TdT)。

8. 根据权利要求1所述的方法, 其特征在于, 所述第一尾和所述第二尾包含相同的序列。

9. 根据权利要求2所述的方法, 其特征在于, 所述第一尾, 所述第二尾或两者选自多聚-A、多聚-C和多聚-C/T。

10. 根据权利要求2所述的方法, 其特征在于, 所述池中的所述两种类型的核苷酸处于9:1, 5:1, 3:1或1:1的比例。

11. 根据权利要求1至10中任一项所述的方法, 其特征在于, 还包括通过使杂交至所述第二接头链的第二引物延伸来扩增包含所述第二接头链的目标多核苷酸。

12. 根据权利要求11所述的方法, 其特征在于, 与所述第一接头链杂交的所述第一引物的序列不同于与所述第二接头杂交的所述第二引物的序列。

13. 根据权利要求11所述的方法, 其特征在于, 使用杂交至所述第二接头链的引物的扩增是指数扩增。

14. 根据权利要求11所述的方法, 其特征在于, 所述方法还包括对包含所述第二引物的扩增的扩增产物测序。

15. 根据权利要求11所述的方法, 其特征在于, 还包括使用第三引物和第四引物的扩增反应, 其中

i) 所述第三引物杂交至所述第一引物的至少一部分的互补序列; 和

ii) 所述第四引物杂交至所述第二引物的至少一部分的互补序列。

16. 根据权利要求15所述的方法, 其特征在于, 所述第三引物的可杂交序列不同于所述第一引物的可杂交序列, 和/或所述第四引物的可杂交序列不同于所述第二引物的可杂交序列。

17. 根据权利要求15所述的方法, 其特征在于, 所述第三引物和所述第四引物的序列是不同的。

18. 根据权利要求15所述的方法, 其特征在于, 所述第三引物, 所述第四引物或两者包含鉴别所述目标多核苷酸的样品来源的索引序列。

19. 根据权利要求15所述的方法, 其特征在于, 所述方法还包括对包含所述第三和第四引物的扩增的扩增产物测序。

20. 根据权利要求15所述的方法, 其特征在于, 所述方法还包括根据所述索引序列将测序读序分组。

21. 根据权利要求15所述的方法, 其特征在于, 所述测序包括相对于参考序列, 检测序列变体或核苷酸甲基化中的差异。

22. 根据权利要求1所述的方法, 其特征在于, 所述第二加尾反应被省略的情况下, 所述第一和/或第二接头具有3'-悬突和5'-悬突两者。

23. 根据权利要求22所述的方法, 其特征在于, 所述第一和/或第二接头的5'-悬突的长度为2至6个核苷酸。

用于制备核酸文库的组合物和方法

[0001] 相关申请

[0002] 本申请为2019年4月2日提交的、发明名称为“用于制备核酸文库的组合物和方法”、申请号为201980002533.4的中国发明专利申请的分案申请。

[0003] 序列表

[0004] 本发明申请包含以ASCII格式电子提交并且以其全部内容作为参考并入本文的序列表。2019年3月28日创建的所述ASCII拷贝的名称为232396-228002_SL.txt并且大小为5,661字节。

发明背景

[0005] 鉴别和分析复杂核酸群体是具有多种应用的活跃的发展领域。通过大规模平行核酸测序(也称为“高通量测序”或者“下一代测序”(NGS))极大地辅助了这些分析。由于如小样本输入以及操作中各个阶段的误差等困难,仍难以检测以相对低丰度存在的核酸物质。这些困难可以在如测试可能的污染物(例如,食品或水中),检测复杂群体中特定细菌的存在(例如,环境测试中)和检测,特别是在早期检测与疾病(例如,感染或癌症)有关的核酸的存在的情况中出现。

发明概述

[0006] 考虑到上述内容,仍需要制备核酸文库的改善的方法。本文所公开的组合物和方法解决了这种需要,并且还提供了其它优势。

[0007] 在一个方面,本发明公开提供了制备多核苷酸文库的方法。在一些实施方式中,所述方法包括(a)在第一加尾反应中,通过不依赖模板的聚合反应将第一尾添加至多个目标多核苷酸中的每一个,其中第一加尾反应包括第一接头,所述第一接头包含杂交至第一尾的悬突;(b)在第一连接反应中,将第一接头链连接至第一尾;(c)通过使杂交至第一接头链的第一引物延伸,使包含第一接头链的目标多核苷酸扩增;(d)在第二加尾反应中,通过不依赖模板的聚合反应将第二尾添加至多个所述扩增的目标多核苷酸中的每一个,其中第二加尾反应包括第二接头,所述第二接头包含杂交至第二尾的悬突;和(e)在第二连接反应中,将第二接头链连接至第二尾。在一些实施方式中,所述方法包括以下步骤中的一个或多个:(a)将多核苷酸片段化以产生所述目标多核苷酸;(b)使所述目标多核苷酸的一端或两端脱磷酸;和(c)使双-链多核苷酸变性为单链多核苷酸以产生所述目标多核苷酸。在一些实施方式中,所述多个目标多核苷酸包括单链DNA。在一些实施方式中,所述目标多核苷酸包括无细胞多核苷酸,或其扩增产物。在一些实施方式中,所述目标多核苷酸包括单链无细胞DNA(cfDNA)。在一些实施方式中,第一加尾反应中目标多核苷酸的量为约0.1-500ng、1-100ng或者5-50ng。在一些实施方式中,所述目标多核苷酸的平均长度为约50至600个核苷酸。在一些实施方式中,在第一连接反应之前处理所述目标多核苷酸以差异修饰甲基化的胞嘧啶或未甲基化的胞嘧啶,如通过用亚硫酸氢盐处理所述目标多核苷酸。在一些实施方式中,通过聚合酶,如末端脱氧核苷酸转移酶(TdT)催化所述不依赖模板的聚合反应。在一

些实施方式中,第一尾包含不同于第二尾的序列。在一些实施方式中,第一尾和第二尾包含相同的序列。在一些实施方式中,第一尾,第二尾或两者由一种或两种类型的核苷酸组成。在一些实施方式中,第一尾,第二尾或两者选自多聚-A、多聚-C和多聚-C/T。在一些实施方式中,所述尾中的至少一个由从两种类型的核苷酸的池聚合的两种类型的核苷酸组成,其中所述池中的两种类型的核苷酸以相同或不同的量存在。在一些实施方式中,所述池中的两种类型的核苷酸处于约9:1,5:1,3:1或1:1的比例。在一些实施方式中,第一接头和第二接头包含多核苷酸序列不同的双-链区。在一些实施方式中,所述扩增包括线性扩增。在一些实施方式中,第一和/或第二接头的悬突为3'-悬突。在一些实施方式中,第一和/或第二接头的悬突的长度为6至12个核苷酸。在一些实施方式中,(i) 第一加尾反应和第一连接反应在相同反应混合物中发生,和/或(ii) 第二加尾反应和第二连接反应在相同反应混合物中发生。

[0008] 在一些实施方式中,所述方法还包括通过使杂交至第二接头链的第二引物延伸来扩增包含所述第二接头链的目标多核苷酸。在一些实施方式中,与第一接头链杂交的第一引物的序列不同于与第二接头杂交的第二引物的序列。在一些实施方式中,使用杂交至第二接头链的引物的扩增是指数扩增。在一些实施方式中,所述方法还包括使用第三引物和第四引物的扩增反应,其中(i) 所述第三引物杂交至所述第一引物的至少一部分的互补序列,和(ii) 所述第四引物杂交至所述第二引物的至少一部分的互补序列。在一些实施方式中,第三引物的可杂交序列不同于第一引物的可杂交序列,和/或第四引物的可杂交序列不同于第二引物的可杂交序列。在一些实施方式中,第三引物和第四引物的序列是不同的。在一些实施方式中,第三引物,第四引物或两者包含鉴别目标多核苷酸的样品来源的索引序列(index sequence)。在一些实施方式中,所述方法还包括对包含第二引物的扩增的扩增产物测序。在一些实施方式中,所述方法还包括对包含第三和第四引物的扩增的扩增产物测序。在一些实施方式中,所述方法还包括根据索引序列将测序读序分组。在一些实施方式中,测序包括相对于参考序列,检测序列变体或核苷酸甲基化中的差异。

[0009] 在一个方面,本发明公开提供了用于在本文所述的一个或多个方法中使用的组合物。

[0010] 在一个方面,本发明公开提供了根据本文所述的任何方法产生的多核苷酸。

[0011] 在一个方面,本发明公开提供了用于制备多核苷酸文库的试剂盒。在一些实施方式中,所述试剂盒包含:(a) 不依赖于模板的聚合酶;(b) 可以通过不依赖于模板的聚合酶聚合的第一核苷酸池;(c) 可以通过不依赖于模板的聚合酶聚合的第二核苷酸池;(d) 包含可杂交至通过所述第一多核苷酸池的聚合所形成的尾的悬突的第一接头;和(e) 包含可杂交至通过所述第二多核苷酸池的聚合所形成的尾的悬突的第二接头,其中所述第二接头包含与所述第一接头不同的序列。在一些实施方式中,所述不依赖于模板的聚合酶是末端脱氧核苷酸转移酶(TdT)。在一些实施方式中,第一池和第二池中的至少一个含有在另一个池中不存在的至少一种类型的核苷酸。在一些实施方式中,第一池和第二池包含相同的一种或多种类型的核苷酸。在一些实施方式中,第一池,第二池或两者由一种或两种类型的核苷酸组成。在一些实施方式中,第一池,第二池或两者选自(i) dATP的池,(ii) dCTP的池,和(iii) dCTP和dTTP的池。在一些实施方式中,第一池和第二池中的至少一个由以相同或不同的量存在的两种类型的核苷酸组成。在一些实施方式中,所述池中的所述两种类型的核苷酸处

于约9:1,5:1,3:1或1:1的比例。在一些实施方式中,所述第一接头和所述第二接头包含多核苷酸序列不同的双-链区。在一些实施方式中,所述第一和/或第二接头的悬突为3'-悬突。在一些实施方式中,所述第一和/或第二接头的悬突的长度为6至12个核苷酸。在一些实施方式中,所述试剂盒还包含在引物延伸反应条件下可杂交至第一接头链的第一引物。在一些实施方式中,所述试剂盒还包含在引物延伸反应条件下可杂交至第二接头链的第二引物。在一些实施方式中,可杂交至第一接头链的第一引物的序列不同于可杂交至第二接头的第二引物的序列。在一些实施方式中,所述试剂盒还包含第三引物和第四引物,其中(i)在引物延伸反应条件下,所述第三引物可杂交至所述第一引物的至少一部分的互补序列,和(ii)在引物延伸反应条件下,所述第四引物可杂交至所述第二引物的至少一部分的互补序列。在一些实施方式中,第三引物的可杂交序列不同于第一引物的可杂交序列,和/或第四引物的可杂交序列不同于第二引物的可杂交序列。在一些实施方式中,第三引物的可杂交序列相对于第一引物的可杂交序列5'杂交,和/或第四引物的可杂交序列相对于第二引物的可杂交序列5'杂交。在一些实施方式中,第三引物和第四引物的序列是不同的。在一些实施方式中,第三引物,第四引物或两者包含鉴别目标多核苷酸的样品来源的索引序列(index sequence)。

[0012] 在本发明用于制备多核苷酸文库的方法的一些实施方式中,所述方法包括(a)在第一加尾反应中,通过不依赖模板的聚合反应将第一尾添加至多个目标多核苷酸中的每一个,其中所述第一加尾反应包括第一接头,所述第一接头包含杂交至所述第一尾的悬突;(b)在第一连接反应中,将所述第一接头链连接至所述第一尾;(c)通过使杂交至所述第一接头链的第一引物延伸,使包含所述第一接头链的目标多核苷酸扩增;和(d)在第二连接反应中,将第二接头链连接至所述扩增的目标多核苷酸。在一些实施方式中,第二连接反应包括,在第二加尾反应中,通过不依赖模板的聚合反应将第二尾添加至多个所述扩增的目标多核苷酸中的每一个。在一些实施方式中,第二加尾反应包括第二接头,所述第二接头包含杂交至第二尾的悬突。在一些实施方式中,在第二连接反应中,将第二接头链连接至第二尾。在一些实施方式中,第二连接反应包括第二接头,所述第二接头包含杂交至所述扩增的目标多核苷酸的悬突。

[0013] 在一些实施方式中,所述方法包括以下步骤中的一个或多个:(a)将多核苷酸片段化以产生所述目标多核苷酸;(b)使所述目标多核苷酸的一端或两端脱磷酸;和(c)使双-链多核苷酸变性为单链多核苷酸以产生所述目标多核苷酸。在一些实施方式中,所述多个目标多核苷酸包括单链DNA。在一些实施方式中,所述目标多核苷酸包括无细胞多核苷酸,或其扩增产物。在一些实施方式中,所述目标多核苷酸包括单链无细胞DNA(cfDNA)。在一些实施方式中,所述第一加尾反应中目标多核苷酸的量为约0.1-500ng、1-100ng或者5-50ng。在一些实施方式中,所述目标多核苷酸的平均长度为约50至600个核苷酸。在一些实施方式中,在步骤(b)之前处理所述目标多核苷酸以差异修饰甲基化的胞嘧啶或未甲基化的胞嘧啶。在一些实施方式中,所述差异修饰包括用亚硫酸氢盐处理所述目标多核苷酸。在一些实施方式中,通过聚合酶催化所述不依赖模板的聚合反应。在一些实施方式中,所述聚合酶是末端脱氧核苷酸转移酶(TdT)。在一些实施方式中,所述第一尾包含不同于所述第二尾的序列。在一些实施方式中,所述第一尾和所述第二尾包含相同的序列。在一些实施方式中,所述第一尾,所述第二尾或两者由一种或两种类型的核苷酸组成。在一些实施方式中,所述第

一尾,所述第二尾或两者选自多聚-A、多聚-C和多聚-C/T。在一些实施方式中,所述尾中的至少一个由从两种类型的核苷酸的池聚合的两种类型的核苷酸组成,其中所述池中的所述两种类型的核苷酸以相同或不同的量存在。在一些实施方式中,所述池中的两种类型的核苷酸处于约9:1,7:1,5:1,3:1或1:1的比例。在一些实施方式中,省略第二加尾反应。在一些实施方式中,第一接头和第二接头包括多核苷酸序列不同的双-链区。在一些实施方式中,所述扩增包括线性扩增。在一些实施方式中,所述第一和/或第二接头的悬突为3'-悬突。在一些实施方式中,第一和/或第二接头具有3'-悬突和5'-悬突两者。在一些实施方式中,第一和/或第二接头的3'-悬突的长度为6至12个核苷酸。在一些实施方式中,第一和/或第二接头的5'-悬突的长度为2至6个核苷酸。在一些实施方式中,(i)所述第一加尾反应和所述第一连接反应在相同反应混合物中发生,和/或(ii)所述第二加尾反应和所述第二连接反应在相同反应混合物中发生。

[0014] 在一些实施方式中,所述方法还包括通过使杂交至所述第二接头链的第二引物延伸来扩增包含所述第二接头链的目标多核苷酸。在一些实施方式中,与第一接头链杂交的第一引物的序列不同于与第二接头杂交的第二引物的序列。在一些实施方式中,使用杂交至第二接头链的引物的扩增是指数扩增。在一些实施方式中,所述方法还包括使用第三引物和第四引物的扩增反应,其中(i)所述第三引物杂交至所述第一引物的至少一部分的互补序列,和(ii)所述第四引物杂交至所述第二引物的至少一部分的互补序列。在一些实施方式中,第三引物的可杂交序列不同于第一引物的可杂交序列,和/或第四引物的可杂交序列不同于第二引物的可杂交序列。在一些实施方式中,第三引物和第四引物的序列是不同的。在一些实施方式中,第三引物,第四引物或两者包含鉴别目标多核苷酸的样品来源的索引序列(index sequence)。在一些实施方式中,所述方法还包括对包含第二引物的扩增的扩增产物测序。在一些实施方式中,所述方法还包括对包含第三和第四引物的扩增的扩增产物测序。在一些实施方式中,所述方法还包括根据索引序列将测序读序分组。

[0015] 在一个方面,本发明公开提供了用于在本文所述的一个或多个方法中使用的组合物。

[0016] 在一个方面,本发明公开提供了根据本文所述的任何方法产生的多核苷酸。

[0017] 在一个方面,本发明公开提供了用于制备多核苷酸文库的试剂盒。在一些实施方式中,所述试剂盒包含:(a)不依赖于模板的聚合酶;(b)可以通过不依赖于模板的聚合酶聚合的第一核苷酸池;(c)可以通过不依赖于模板的聚合酶聚合的第二核苷酸池;(d)第一接头,其包含可杂交至通过所述第一多核苷酸池的聚合所形成的尾的悬突;和(e)第二接头,其包含可杂交至所述扩增的目标多核苷酸的悬突。在一些实施方式中,所述不依赖于模板的聚合酶是末端脱氧核苷酸转移酶(TdT)。在一些实施方式中,第一池和第二池中的至少一个含有在另一个池中不存在的至少一种类型的核苷酸。在一些实施方式中,第一池和第二池包含相同的一种或多种类型的核苷酸。在一些实施方式中,第一池,第二池或两者由一种或两种类型的核苷酸组成。在一些实施方式中,第一池,第二池或两者选自(i) dATP的池,(ii) dCTP的池,和(iii) dCTP和dTTP的池。在一些实施方式中,第一池和第二池中的至少一个由以相同或不同的量存在的两种类型的核苷酸组成。在一些实施方式中,所述池中的两种类型的核苷酸处于约9:1,7:1,5:1,3:1或1:1的比例。在一些实施方式中,第一接头和第二接头包含多核苷酸序列不同的双-链区。在一些实施方式中,所述第一和/或第二接头的

悬突为3'-悬突。在一些实施方式中,第一和/或第二接头具有3'-悬突和5'-悬突两者。在一些实施方式中,第一和/或第二接头的3'-悬突的长度为6至12个核苷酸。在一些实施方式中,第一和/或第二接头的5'-悬突的长度为2至6个核苷酸。在一些实施方式中,所述试剂盒还包含在引物延伸反应条件下可杂交至第一接头链的第一引物。在一些实施方式中,所述试剂盒还包含在引物延伸反应条件下可杂交至第二接头链的第二引物。在一些实施方式中,可杂交至第一接头链的第一引物的序列不同于可杂交至第二接头的第二引物的序列。在一些实施方式中,所述试剂盒还包含第三引物和第四引物,其中(i)在引物延伸反应条件下,所述第三引物可杂交至所述第一引物的至少一部分的互补序列,和(ii)在引物延伸反应条件下,所述第四引物可杂交至所述第二引物的至少一部分的互补序列。在一些实施方式中,第三引物的可杂交序列不同于第一引物的可杂交序列,和/或第四引物的可杂交序列不同于第二引物的可杂交序列。在一些实施方式中,第三引物的可杂交序列相对于第一引物的可杂交序列5'杂交,和/或第四引物的可杂交序列相对于第二引物的可杂交序列5'杂交。在一些实施方式中,第三引物和第四引物的序列是不同的。在一些实施方式中,第三引物,第四引物或两者包含鉴别目标多核苷酸的样品来源的索引序列(index sequence)。

附图说明

[0018] 图1显示了根据实施方式,示例性文库制备方法。图示包括序列CCCTCCTC (SEQ ID NO:1),TTTTTTTTTTTT (SEQ ID NO:2)和AAAAAAAAAAAA (SEQ ID NO:3)。

[0019] 图2显示了根据实施方式,示例性接头。图示以自上而下的顺序包括SEQ ID NO:4-7。

[0020] 图3显示了根据实施方式制备的多核苷酸之间的比较,其包括加尾反应(下图),和作为替代,使用“Y”接头所制备的多核苷酸(上图)。图示以从左至右,然后自上而下的顺序包括SEQ ID NO:8-15。

[0021] 图4显示了示例性毛细管电泳分析图。

[0022] 图5A-C显示了示例性毛细管电泳分析图。

[0023] 图6A-B显示了示例性电泳分析图。

[0024] 图7显示了不同样品之间,12,977个目标CpG位点的甲基化水平。

[0025] 图8A-B显示了示例性毛细管电泳分析图。

[0026] 图9显示了根据本发明的实施方式,示例性文库制备方法。图示包括序列TCTCTCTC和NNNNNNN,其中N为任何碱基。

[0027] 图10显示了根据本发明的实施方式,示例性接头。图示以自上而下的顺序包括SEQ ID NO:4、22、6和23。

[0028] 图11显示了示例性毛细管电泳分析图(图上自上而下的线,10ngλ、5ngλ、2ngλ、1ngλ)。

发明详述

[0029] 除非另外说明,否则本文所公开的一些实施方式的特定步骤的实践使用了免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组和重组DNA的常规技术,这些技术在本领域的技术范围内。参见,例如,Sambrook and Green,Molecular Cloning:A

Laboratory Manual,第4版(2012);the series Current Protocols in Molecular Biology(F.M.Ausubel等人主编);the series Methods In Enzymology(Academic Press, Inc.),PCR 2:A Practical Approach(M.J.MacPherson,B.D.Hames and G.R.Taylor主编.(1995)),Harlow and Lane主编,(1988)Antibodies,A Laboratory Manual,and Culture of Animal Cells:AManual of Basic Technique and Specialized Applications,第6版(R.I.Freshney主编,(2010))。

[0030] 除非上下文明确规定,否则如在说明书和权利要求中所使用的,单数形式的“一个”和“所述”包括复数提及对象。

[0031] 术语“约”或“大约”表示在如本领域的技术人员所确定的具体的值的可接受误差范围内,其将部分基于所述值是如何测量或确定的,即测量系统的限制。例如,“约”可以表示根据本领域中的实践,在一个或大于一个标准偏差之内。作为另外一种选择,“约”可以表示给定值的多至20%,多至10%,多至5%或者多至1%的范围。作为另外一种选择,具体地对于生物系统或过程,该术语可以表示在值的数量级内,优选地,在值的5-倍之内,更优选地,在值的2-倍之内。当在本发明申请和权利要求中描述具体的值时,除非另有说明,否则应接受术语“约”表示在所述具体的值的可接受的误差范围之内。

[0032] 术语“多核苷酸”、“核苷酸”、“核酸”和“寡核苷酸”是可互换使用的。它们表示具有任何长度的核苷酸(脱氧核糖核苷酸或者核糖核苷酸)的多聚形式,或其类似物。多核苷酸可以具有任何立体结构,并且可以发挥任何功能,无论是已知的还是未知的。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、根据连锁分析所限定的基因座(基因座)、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)、核糖酶、cDNA、重组多核苷酸、分枝多核苷酸、质粒、载体、具有任何序列的分离的DNA、具有任何序列的分离的RNA、核酸探针、引物和接头。多核苷酸可以包括一个或多个修饰的核苷酸,如甲基化核苷酸和核苷酸类似物。如果存在,可以在所述聚合物组装之前或之后赋予对核苷酸结构的修饰。可以通过非核苷酸组分中断核苷酸序列。可以在聚合后进一步修饰多核苷酸,如通过用标记组分的缀合。

[0033] 一般地,如应用于多核苷酸的术语“无细胞的”、“循环”和“胞外的”(例如,“无细胞DNA”和“无细胞RNA”)可互换地用于表示在不对如最初所采集的样品(例如,如来自细胞或病毒的提取中的样品)应用溶胞步骤的情况下,可以分离或另外操作的存在于来自对象的样品或其部分中的多核苷酸。因此,即使在采集对象的样品之前,无细胞多核苷酸是非包封的或者“游离于”它们所来源的细胞或病毒的。可以作为细胞死亡(例如,细胞凋亡或坏死)或将多核苷酸脱落,释放到周围体液或循环中的细胞的副产物产生无细胞多核苷酸。因此,无细胞多核苷酸可以分离自血液的非细胞部分(例如,血清或血浆),分离自其它体液(例如,尿)或者分离自其它类型样品的非细胞部分。

[0034] 如本文所使用的,“对象”可以是哺乳动物,如非灵长类(例如,牛、猪、马、猫、狗、大鼠等)或者灵长类(例如,猴或人)。在一些实施方式中,所述对象是人。在一些实施方式中,所述对象是患有或潜在患有在本文中描述了其实例的疾病、病症或病况的哺乳动物(例如,人)。在一些实施方式中,所述对象是具有发展在本文中描述了其实例的疾病、病症或病况的风险的哺乳动物(例如,人)。

[0035] 如本文所使用的术语“扩增”和“扩增的”一般地表示由目标多核苷酸或其部分组

成一个或多个拷贝的任何过程。扩增多核苷酸(例如,DNA和/或RNA)的多种方法是可用的,在本文中描述了它们的一些实例。扩增可以是线性的、指数的或者在多阶段扩增方法中包括线性和指数阶段两者。扩增方法可以包括温度变化,如热变性步骤,或者可以是不需要热变性的等温过程。

[0036] “杂交”是指其中一个或多个多核苷酸反应以形成通过核苷酸残基的碱基之间的氢键稳定的复合物的反应。可以通过沃森-克里克碱基配对、胡格斯丁结合(Hoogstein binding)或者根据碱基互补以任何其它序列特异性方式发生氢键作用。所述复合物可以包括形成双螺旋结构的两条链,形成多链复合物的三条或更多条链、自杂交单链或这些的任意组合。杂交反应可以构成更广泛的方法中的步骤,如PCR的起始或者通过核酸内切酶的多核苷酸的酶促切割。将与第一序列完全互补的或者使用第一序列作为模板,通过聚合酶聚合的第二序列称为与所述第一序列“互补”。如应用于多核苷酸的术语“可杂交的”是指多核苷酸在杂交反应中形成通过核苷酸残基的碱基之间的氢键稳定的复合物的能力。在一些实施方式中,可杂交的核苷酸序列与它所杂交的序列至少约50%、60%、70%、75%、80%、85%、90%、95%或100%互补。在一些实施方式中,可杂交的序列是作为多步过程(例如,连接反应或扩增反应)中的步骤的一部分并且在所述步骤的条件下杂交至一个或多个靶标序列的序列。

[0037] “互补性”是指核酸通过常规沃森-克里克碱基配对或者其它非常规类型与另一核酸序列形成氢键的能力。互补性百分比表示可以与第二核酸序列形成氢键(例如,沃森-克里克碱基配对)的第一核酸序列中的残基的百分比(例如,10个中的5、6、7、8、9或10个,分别为50%、60%、70%、80%、90%和100%互补的)。“完全互补的”是指第一核酸序列的所有邻接残基将与第二核酸序列中相同个数的邻接残基氢键键合。如出于评价互补性百分比的目的,可以通过任何适合的比对算法来测量序列同一性,其包括(但不限于)Needleman-Wunsch算法(参见,例如,任选地以缺省设置,在www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html可用的EMBOSS Needle比对器)、BLAST算法(参见,例如,任选地以缺省设置,在blast.ncbi.nlm.nih.gov/Blast.cgi可用的BLAST比对工具)或者Smith-Waterman算法(参见,例如,任选地以缺省设置,在www.ebi.ac.uk/Tools/psa/emboss_water/nucleotide.html可用的EMBOSS Water比对器)。可以使用所选算法的任何适合的参数,包括缺省参数来评价最优比对。

[0038] 一般地,术语“序列变体”是指相对于一条或多条参考序列,序列中的任何变化。通常,对于参考序列已知的个体的给定群体,序列变体以低于参考序列的频率发生。在一些情况下,所述参考序列是单个已知的参考序列,如单个个体的基因组序列。在一些情况下,所述参考序列是通过多条已知序列的比对所形成的共有序列,如用作参比群体的多个个体的基因组序列或者来自相同个体的多核苷酸的多个测序读序。在一些情况下,所述序列变体在所述群体中以低频率发生(也称为“稀有”序列变体)。例如,所述序列变体可以以约或小于约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,所述序列变体以约或小于约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化。序列变化可以包括单个核苷酸或者多个核苷酸(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。当序列变体包括两个或更多个核苷酸差异时,不同的核苷酸可以

是彼此邻接的或者是间隔的。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)、拷贝数变异 (CNV)、短串联重复序列 (STR)、简单顺序重复 (SSR)、可变数目串联重复序列 (VNTR)、扩增片段长度多态性 (AFLP)、反转座子-基插入多态性、序列特异性扩增多态性和可以作为序列变体检测的表观遗传标记差异 (例如, 甲基化差异)。在一些实施方式中, 序列变体可以表示染色体重排, 其包括 (但不限于) 易位或融合基因。

[0039] 在一个方面, 本发明公开提供了制备多核苷酸文库的方法。在一些实施方式中, 所述方法包括 (a) 在第一加尾反应中, 通过不依赖模板的聚合反应将第一尾添加至多个目标多核苷酸中的每一个, 其中所述第一加尾反应包括第一接头, 所述第一接头包含杂交至所述第一尾的悬突; (b) 在第一连接反应中, 将所述第一接头链连接至所述第一尾; (c) 通过使杂交至所述第一接头链的第一引物延伸, 使包含所述第一接头链的目标多核苷酸扩增; (d) 在第二加尾反应中, 通过不依赖模板的聚合反应将第二尾添加至多个所述扩增的目标多核苷酸中的每一个, 其中所述第二加尾反应包括第二接头, 所述第二接头包含杂交至所述第二尾的悬突; 和 (e) 在第二连接反应中, 将所述第二接头链连接至所述第二尾。

[0040] 在一个方面, 本发明公开提供了制备多核苷酸文库的方法。在一些实施方式中, 所述方法包括 (a) 在第一加尾反应中, 通过不依赖模板的聚合反应将第一尾添加至多个目标多核苷酸中的每一个, 其中所述第一加尾反应包括第一接头, 所述第一接头包含杂交至所述第一尾的悬突; (b) 在第一连接反应中, 将所述第一接头链连接至所述第一尾; (c) 通过使杂交至所述第一接头链的第一引物延伸, 使包含所述第一接头链的目标多核苷酸扩增; 和 (d) 在第二连接反应中, 将第二接头链连接至所述扩增的目标多核苷酸。在这种实施方式中, 在无加尾反应的情况下使用第二接头连接。任选地, 在该方法中, 所述第二连接反应可以包括, 在第二加尾反应中, 通过不依赖模板的聚合反应将第二尾添加至多个所述扩增的目标多核苷酸中的每一个。在一个实施方式中, 第二加尾反应可以包括第二接头, 所述第二接头包含杂交至第二尾的悬突。在一个实施方式中, 在第二连接反应中, 将第二接头链连接至第二尾。在一个实施方式中, 第二连接反应包括第二接头, 所述第二接头包含杂交至所述扩增的目标多核苷酸的悬突。这种实施方式允许后续连接。在一个实施方式中, 第二接头连接可以使用所述接头中的随机碱基的3'悬突以用作辅助连接的碎片 (splinter)。可以将第二接头添加至所述扩增的目标多核苷酸的3'端。将所述接头的3'悬突用作碎片以稳定底物链并且辅助底物链的3'端和磷酸化的相对接头链的5'端之间的连接。

[0041] 在本发明公开所述的方法中有用的多核苷酸可以来源于任何多种样品来源。在一些实施方式中, 所述样品是环境样品, 如天然存在或人工大气、水样、土样、表面拭子或任何其它所关心的样品。在一些实施方式中, 多核苷酸来源于生物样品, 如对象的样品。生物样品的非限制性实例包括组织 (例如, 皮肤、心脏、肺、肾、骨髓、胸、胰脏、肝、肌肉、平滑肌、膀胱、胆囊、结肠、肠、脑、前列腺、食管、甲状腺和肿瘤)、体液 (例如, 血液、血液部份、血清、血浆、唾液、尿、乳汁、胃液和消化液、泪液、精液、阴道液、来源于肿瘤组织的间隙液、眼部液、汗液、粘液、油、腺分泌物、脊髓液、脑脊液、胎水、羊膜水、脐血、腔液、痰液、脓)、大便、拭子或洗液 (例如, 鼻拭子、喉拭子和鼻咽洗液)、活组织检查及其它排泄物或身体组织。在一些实施方式中, 所述样品为血液、血液部份、血浆、血清、唾液、痰液、尿、精液、经阴道液、脑脊髓液或者大便。在一些实施方式中, 所述样品为血液, 如全血或血液部份 (例如, 血清或血浆)。

[0042] 在一些实施方式中,从样品提取多核苷酸,如当要分析的多核苷酸包含在细胞或病毒衣壳内时。当使用提取法时,所选的方法可以部分基于要处理的样品类型。多种提取法是可用的。例如,可以用苯酚、苯酚/氯仿/异戊醇或者类似制剂,包括TRIzol和TriReagent,通过有机提取纯化核酸。在一些实施方式中,处理样品以除去一种或多种组分或使其降解,所述组分如蛋白(例如,通过蛋白酶K处理)或者RNA(例如,通过RNA酶处理),和/或以保留一种或多种组分,如RNA(例如,通过用RNA酶抑制剂处理)。当在提取程序期间或之后一起分离了DNA和RNA两者时,可以使用其它步骤,从而相对于另一个,单独纯化所述之一或两者。还可以产生所提取的核酸的亚组分,例如,按照尺寸、序列或者其它物理或化学特性纯化。除初始核酸分离步骤之外,可以在后续操作之后进行核酸的纯化,从而除去过量或不希望的试剂、反应物或产物。

[0043] 在一些实施方式中,本文所述的方法包括在不进行细胞提取的情况下(例如,无细胞、病毒和/或包含核酸的其它小囊的溶胞步骤),对得自对象样品的无细胞多核苷酸的操作。在一些实施方式中,直接在如所收集的生物样品中操作多核苷酸。在一些实施方式中,将无细胞多核苷酸与样品的其它组分(例如,细胞和/或蛋白)分离,而不进行处理以使包含在可能存在于样品中的细胞中的多核苷酸释放。对于包含细胞的样品,可以处理所述样品以将细胞与所述样品分离。在一些实施方式中,对样品进行离心并分离包含所述无细胞多核苷酸的上清液以用于进一步处理(例如,多核苷酸与其它组分的分离,或者多核苷酸的其它操作)。在一些实施方式中,从初始样品的其它组分(例如,细胞和/或蛋白)中纯化出无细胞多核苷酸。在不进行细胞提取的情况下,用于分离多核苷酸的多种程序是可行的,如通过沉淀或者与基材的非特异性结合,然后清洗所述基材以释放所结合的多核苷酸。

[0044] 分离自样品来源(例如,环境样品或者来自对象的样品)的多核苷酸的起始量可以是不同的,并且在一些情况下可以是较少的。在一些实施方式中,起始多核苷酸的量为约或小于约1000ng、500ng、100ng、50ng、25ng、20ng、15ng、10ng、5ng、4ng、3ng、2ng、1ng、0.5ng、0.1ng或更少。在一些实施方式中,起始多核苷酸的量在约0.1-500ng的范围内,如1-100ng之间或者5-50ng之间。一般地,较少的起始材料提高了从单一处理步骤至下一个步骤回收多核苷酸的重要性。降低样品中多核苷酸参与后续反应的量的处理降低了可以检测稀有多核苷酸(例如,突变)的灵敏度。在一些实施方式中,相对于在先检测方法,本文所公开的方法提高了检测灵敏度。

[0045] 在一些实施方式中,要分析的多核苷酸包括来自样品的多核苷酸的扩增产物。可以特异性地扩增(例如,通过使用靶标-特异性扩增引物)或者非特异性地扩增(例如,通过使用非特异性扩增引物的池)扩增产物。在一些实施方式中,扩增模板包括DNA和/或RNA。在一些实施方式中,要分析的多核苷酸包括作为反转录(RT)反应的一部分,反转录为DNA的RNA。一般地,反转录包括使用目标RNA分子作为产生互补DNA(cDNA)的模板,通过RNA-依赖性DNA聚合酶(也称为“反转录酶”),使杂交至目标RNA的寡核苷酸引物延伸。反转录酶的实例包括(但不限于)反转录病毒反转录酶(例如,莫洛尼小鼠白血病病毒(M-MLV)、禽类成髓细胞瘤病毒(AMV)或者劳氏肉瘤病毒(RSV)反转录酶)、Superscript ITM、Superscript IITM、Superscript IIITM、反转座子反转录酶、乙型肝炎反转录酶、花椰菜花叶病毒反转录酶、细菌反转录酶及其突变体、变体或衍生物。在一些实施方式中,所述反转录酶是热启动反转录酶。

[0046] 在一些实施方式中,所述多核苷酸是已进行片段化的多核苷酸。在一些实施方式中,所述片段具有小于预定长度或者在预定长度范围内的平均长度、中值长度或分级长度分布(例如,占至少50%、60%、70%、80%、90%或以上)。在一些实施方式中,所述预定长度为约或小于约1500、1000、800、600、500、300、200、100或者50个核苷酸长度。在一些实施方式中,所述预定长度范围在10-1000,10-800,10-700,50-600,100-600或者150-400个核苷酸长度之间。在一些实施方式中,所述片段化多核苷酸具有处于预定范围内的平均尺寸(例如,约10至约1,000个核苷酸长度,如10-800,10-700,50-600,100-600或者150-400个核苷酸之间的平均或中值长度;或者小于1500、1000、750、500、400、300、250、100、50个或更少的核苷酸长度的平均或中值长度)。

[0047] 在一些实施方式中,使多核苷酸片段化包括机械片段化、化学片段化和/或加热。在一些实施方式中,通过机械完成所述片段化,其包括对样品多核苷酸进行声学处理。在一些实施方式中,所述片段化包括在适合于一种或多种酶产生核酸断裂(例如,双链断裂)的条件下,使用一种或多种酶处理所述样品多核苷酸。在产生多核苷酸片段中有用的酶的实例包括序列特异性和非序列特异性核酸酶。核酸酶的非限制性实例包括DNA酶I、片段化酶、限制性核酸内切酶、其变体及其组合。例如,在不存在Mg⁺⁺的情况下和在存在Mn⁺⁺的情况下,用DNA酶I消化可以在DNA中引起随机双链断裂。在一些实施方式中,片段化包括使用一种或多种限制性核酸内切酶处理样品多核苷酸。片段化可以产生具有5'悬突、3'悬突、平端或它们的组合的片段。在一些实施方式中,如当片段化包括一种或多种限制性核酸内切酶的使用时,样品多核苷酸的切割留下了具有可预测序列的悬突。可以对片段化的多核苷酸进行片段尺寸选择步骤,如通过琼脂糖凝胶的柱纯化或分离。

[0048] 在一些实施方式中,处理多核苷酸以制备用于后续步骤,如延伸或连接步骤的5'端和/或3'端。在片段化程序之后,多核苷酸端的制备可以是特别有帮助的。多核苷酸端的制备经常称为端“削平”或“修复”。在一些实施方式中,修复多核苷酸的末端以产生平端或者具有5'磷酸化末端的单链片段(例如,使用dNTP、T4 DNA聚合酶、Klenow大片段、T4多核苷酸激酶和ATP)。在一些实施方式中,末端修复包括将腺嘌呤增加至3'末端以产生3'-A悬突(例如,使用dATP、Klenow片段(3'-5'外切)或者Taq聚合酶)。在一些实施方式中,将多核苷酸末端之一或两者脱磷酸,如通过用磷酸酶处理。

[0049] 在一些实施方式中,所述方法包括第一加尾反应,其中通过不依赖模板的聚合反应将第一尾添加至多个目标多核苷酸中的每一个。在一些实施方式中,所述目标多核苷酸是单链的。所述目标多核苷酸可以是天然单链,或者如果尚未如此,则处理成单链。例如,目标RNA可以反转录以形成DNA-RNA杂交分子,然后可以用RNA酶H处理或在存在RNA酶A的情况下热-变性以降解RNA并获得单链cDNA。作为进一步的实例,可以使双链DNA热-变性(例如,通过在约95℃培育),任选地随后快速冷却(例如,在冰上培育)。在一些实施方式中,所述目标多核苷酸包括单链DNA。在一些实施方式中,所述目标多核苷酸包括单链cfDNA。

[0050] 一般地,通过不依赖模板的聚合反应所产生的“尾”是指聚合至对其进行聚合反应的目标多核苷酸的末端的新合成的核苷酸串(string)。所述尾的长度和核苷酸序列将部分基于由其聚合成所述尾的核苷酸的类型(例如,A、T、G和C中的1、2、3或4个)、反应持续时间、所使用的聚合酶和其它试剂的存在(例如,包含在聚合反应期间杂交至第一尾的悬突的接头)。在一些实施方式中,所述尾仅聚合至一个或多个目标多核苷酸的3'端。

[0051] 在一些实施方式中,从由4种类型的DNA碱基(A、T、G和C)组成的池聚合尾,从而所得尾具有包含任何或全部4种碱基的机会。在一些实施方式中,从由碱基A、T、G和C中的任何3种组成的池聚合尾,从而所得尾具有包含任何或全部3种所选择的碱基的机会。在一些实施方式中,从由碱基A、T、G和C中的任何两种类型组成的池(如C/T或A/G)聚合尾,从而所得尾具有包含2种所选择的碱基中的任一种或两种的机会。在一些实施方式中,从由选自A、T、G和C中的1种类型的碱基组成的池聚合尾,从而所得尾由所选类型的碱基组成。在一些实施方式中,所述池由胸腺嘧啶碱基(从而获得多聚-T尾)或者胞嘧啶碱基(从而获得多聚-C尾)组成。通常,所述碱基处于三磷酸盐形式(例如,dATP、dTTP、dGTP和/或dCTP)。当所述池中存在不止一种类型的碱基时,可以通过调整所述池中碱基类型的比例来调节所述尾的组成。在一些实施方式中,所述池中所有类型的碱基以近似相等的量存在,从而任一种类型与任何其它类型的比例为约1:1。在一些实施方式中,所述池中一种类型的碱基与另一种类型的比例为约或大于约2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、15:1或更高。在一些实施方式中,所述池中的一种类型的碱基与另一种类型的比例为约或大于约3:1、5:1或9:1。在一些实施方式中,所述比例为约或大于约9:1。当所述池中存在不止一种类型的核苷酸时,所述尾的序列可以表示为代表所述池中成员的字母的简并序列。例如,“RRR”是指三个嘌呤的序列并且代表序列AAA、AAG、AGA、GAA、AGG、GAG、GGA和GGG;“YYY”是指三个嘧啶的序列并且代表序列TTT、TTC、TCT、CTT、TCC、CCT、CTC和CCC。在这些情况下,一个分子上的尾可以或可以不与另一个相同。然而,可以基于所述池中核苷酸的类型以及它们的相对量来调节所得加尾多核苷酸的池内的可能序列的组以及它们的相对可能性。在包含不止一个加尾反应的实施方式中,可以选择每个反应的条件以产生如就长度、所包括的核苷酸的类型和/或核苷酸的相对量(如果所述池中存在不止一个时)而言,相同或不同的尾。在一些实施方式中,所述方法包括两个加尾反应并且所述尾是相同的。在一些实施方式中,所述方法包括两个加尾反应并且所述尾是不同的。

[0052] 在一些实施方式中,一个或多个步骤包括通过聚合酶的多核苷酸延伸。示例性多核苷酸延伸反应包括反转录、加尾和扩增。多种聚合酶是可用的并且可以适合地选择用于适当类型的多核苷酸延伸反应。在一些实施方式中,所述多核苷酸延伸反应是加尾反应,如不依赖于模板的加尾反应。在一些实施方式中,所述不依赖于模板的加尾反应包括通过不依赖于模板的聚合酶的多核苷酸延伸。一般地,不依赖于模板的聚合酶是能够在不存在与要聚合的序列互补的模板的情况下催化多核苷酸延伸反应的聚合酶。当为了催化反应,不依赖于模板的聚合酶不需要模板的存在时,聚合反应的发生与模板分子是否存在无关,从而不必要求模板的不存在。不依赖于模板的聚合酶的非限制性实例包括末端脱氧核苷酰转移酶(TdT;也称为DNA核苷酸外转移酶(DNTT)或者末端转移酶)、多聚-A聚合酶、RNA-特异性核苷酸转移酶、多聚(U)聚合酶及其突变或修饰形式。在一些实施方式中,所述不依赖于模板的聚合酶为TDT。所述不依赖于模板的聚合酶可以来自任何适合的来源。不依赖于模板的聚合酶的具体非限制性实例包括重组产生的小牛胸腺TDT和大肠杆菌(E.coli)多聚-A聚合酶,以上两者均是可商购的。

[0053] 在一些实施方式中,加尾反应包括接头,所述接头包含杂交至尾的悬突。所述悬突可以在多核苷酸延伸反应期间杂交至尾;然而,在通过不依赖于模板的聚合酶引发的不依赖模板的聚合反应中,这种杂交不否认所述反应状态是不依赖于模板的。具有悬突的接头

包括至少一个单链区(所述悬突)和至少一个双链区(紧邻所述悬突)。接头可以包括位于两端的悬突,并且包括相同或不同的链。例如,可以通过杂交位于较长寡核苷酸中间的短寡核苷酸来形成双-链区。作为另一个实例,两个寡核苷酸可以彼此杂交,从而通过寡核苷酸之一形成位于一端的悬突,并且通过另一个寡核苷酸形成位于另一端的悬突。在一些实施方式中,仅在一端存在悬突,从而另一端以配对的核苷酸结束(也称为“平端”)。还可以通过杂交大于两个寡核苷酸来形成接头,并且所述接头可以包括位于双链区之间的内部单链区(例如,作为在沿长寡核苷酸相隔一个或多个核苷酸的区域杂交至相同长寡核苷酸的两个短寡核苷酸中)。在一些实施方式中,仅在5'或3'端中的任一个上存在单一悬突。在一些实施方式中,所述悬突是3'-悬突。在一些实施方式中,所述接头具有3'悬突和5'悬突两者。不受具体理论的束缚,5'悬突产生了可以防止接头本身上的泄漏加尾反应(leaky tailing reaction)的隐性3'端。5'悬突在另一条链上产生了3'隐性端,其防止了寡核苷酸合成期间,由于3'端化学封端不完全所造成的所述接头上的泄漏加尾反应。

[0054] 一般地,杂交至特定尾的悬突包括被设计以与要聚合的尾互补的序列。在一些实施方式中,设计使所述悬突的整个长度杂交至尾。设计以杂交至尾的序列不需要与尾完全互补;而是,所述悬突仅需要设计成在特定反应条件下,如在加尾反应期间杂交至尾。在一些实施方式中,将所述悬突设计成完全互补的。如果从单一类型的核苷酸池中聚合尾(例如,多聚-A),则设计完全互补的悬突(或其部分)是相对直接的(例如,就多聚-A来说,多聚-T)。

[0055] 如果从两种或更多种类型的多核苷酸的池聚合尾,则单个尾序列可以是不同的,从而与一个单个尾完全互补的接头悬突将不会与另一个完全互补。在一些实施方式中,将单个接头悬突序列设计成与从两种或更多种核苷酸聚合的尾具有最大互补性。例如,可以将C:T的比例为5:1的C和T聚合的尾设计为多聚-G。在该实例中,沿相同长度的多聚-G接头悬突,预期10个核苷酸的尾将具有平均2个错配。作为另外一种选择,接头序列可以表现为含有基于被设计使其杂交至的尾的简并位置所选择的一个或多个(或者全部)简并位置。例如,对于序列“YYY”所表示的尾,可以设计悬突以具有序列“RRR”。当悬突包含一个或多个简并碱基位置时,“接头”代表接头寡核苷酸的池,其中在所述池中代表每个简并位置具有的不同的核苷酸中的每一种。在接头寡核苷酸池中,可以调节悬突中特定核苷酸的相对代表,或者所述池中一条或多条序列的相对量(例如,以对应于尾从中所聚合的核苷酸池中的核苷酸的相对量)。例如,可以从与尾核苷酸互补且处于相应相对量(例如,对于从9:1C:T聚合的尾,9:1的G:A)的核苷酸池聚合形成能够形成悬突的接头链的寡核苷酸。作为另一个实例,设计以杂交至多聚-C/T尾(例如,9:1C:T)的接头可以设计为10个核苷酸长度并且等量包含所有可能的具有单个腺嘌呤的悬突,和任选地每条具有两个腺嘌呤的序列。用于设计杂交至从给定核苷酸池聚合的尾的悬突的其它变化是可能的。

[0056] 在一些实施方式中,选择接头的悬突的长度以控制通过不依赖于模板的聚合酶所产生的尾的长度,特别是在其中聚合酶缺少链置换活性的情况下。在这些实施方式中,当尾杂交至悬突时,接头的双-链区抑制尾的伸长。抑制尾伸长不必要求在伸长反应中所产生的所有尾具有与悬突相同的长度。而是,如果在不依赖模板的聚合反应中所产生的平均尾长度比在不存在接头的情况下所产生的平均尾长度短,则认为尾伸长受接头抑制。在一些实施方式中,接头悬突的长度为约或小于约3、4、5、6、7、8、9、10、11、12、13、14、15、20、25个

或更多个核苷酸。在一些实施方式中,接头悬突的长度在约3-25、5-20或10-15个核苷酸之间。在一些实施方式中,悬突的长度为约6-12个核苷酸。

[0057] 在包含不止一个接头(例如,第一接头和第二接头)的方法中,接头或其任何部分(例如,悬突、双-链区或一些其它序列元件,如引物结合位点)的长度和/或序列可以是相同或不同的。在一些实施方式中,所述方法包括两个加尾反应,其分别包括接头并且所述两个接头具有长度相等和/或序列相同的悬突。在一些实施方式中,所述方法包括两个加尾反应,其分别包括接头并且所述两个接头具有长度不同和/或序列不同的悬突。在一些实施方式中,接头以相对于所述反应中目标多核苷酸的量的约或小于约0.25-倍、0.5-倍、0.75-倍、1-倍、2-倍、3-倍、4-倍、5-倍、10-倍或以上的相对摩尔量存在于加尾反应中。在一些实施方式中,接头以相对于目标多核苷酸约1:1的摩尔比存在于加尾反应中。

[0058] 在一些实施方式中,除与尾杂交的悬突之外,接头包含多个序列元件中的一个或多个。其它序列元件的实例包括(但不限于)一个或多个扩增引物退火序列或其互补序列、一个或多个测序引物退火序列或其互补序列、一个或多个索引序列(例如,与可以用于鉴别所述索引与之相关的目标多核苷酸的来源的特定样品来源或反应有关的一个或多个序列)、多个不同接头或不同接头的亚组所共有的一个或多个公用序列、一个或多个限制性内切酶识别位点、一个或多个探针结合位点(例如,用于连接至测序平台,如用于大规模平行测序的流通池,如Illumina, Inc.所开发的流通池)、一个或多个随机或近-随机序列(例如,在一个或多个位置,从两个或更多个不同的核苷酸组随机选择的一个或多个核苷酸,其中在包含所述随机序列的接头池中代表一个或多个位置选择的所述不同核苷酸中的每一个)及其组合。在一些实施方式中,接头用于纯化它们所连接的目标多核苷酸,例如,通过使用涂覆了包含连接至目标多核苷酸的所述接头(或其部分)的互补序列的寡核苷酸的珠(具体地,便于处理的磁珠)。两个或更多个序列元件可以是彼此不相邻的(例如,通过一个或多个核苷酸分离的),彼此相邻的,部分重叠的或者完全重叠的。例如,扩增引物退火序列也可以用作测序引物退火序列。序列元件可以位于3'端或其附近,位于5'端或其附近或者在接头寡核苷酸的内部。序列元件可以具有任何适合的长度,如长度约或小于约3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50个或更多个核苷酸。接头寡核苷酸可以具有任何适合的长度,其至少足以容纳它们所包括的一个或多个序列元件。在一些实施方式中,接头包括分别独立选择以具有约或小于约10、15、20、25、30、35、40、45、50、55、60、65、70、75个或更多个核苷酸长度的寡核苷酸。在一些实施方式中,接头寡核苷酸在约10至75个核苷酸长度的范围内,如约15至50个核苷酸长度。在一些实施方式中,接头包含长度为约或小于约5、10、15、20、25、30、35、40、45、50、75个或更多个核苷酸的双-链部分。

[0059] 在一些实施方式中,接头包括不是用于多核苷酸延伸(如在不依赖模板的聚合反应期间)的底物的一个或多个3'端。在这些情况下,将所述3'端称为“封端的”。在一些实施方式中,封端的3'端是杂交至不依赖模板的聚合反应期间所形成的尾的悬突的3'端,从而所述3'端在所述反应期间不延伸。多种方法对于形成不可以延伸的3'端可用,其无限制地包括在3'端引入不可以延伸的核苷酸并修饰所述3'端核苷酸以使其不可延伸。在一些实施方式中,所述3'端缺少聚合酶所需的共价连接另一种核苷酸的3'羟基。在一些实施方式中,将封端基团添加至所述接头的末端3'-OH或2'-OH。封端基团的一些非限制性实例包括烷基、非核苷酸接头、磷酸基、硫代磷酸酯基、烷烃-二醇部分和氨基。在一些实施方式中,通过用

氟取代氢或者通过酯、酰胺、硫酸酯或糖苷的形成来修饰所述3'-羟基。在一些实施方式中,用氢替换所述3'-OH基(以形成双脱氧核苷酸)。在一些实施方式中,所述3'端包括磷酸基。

[0060] 在一些实施方式中,如在连接反应中,将接头链连接至尾序列。在一些实施方式中,在与加尾反应相同的反应混合物中发生连接。在一些实施方式中,用于进行连接反应的试剂包含在加尾反应中。在一些实施方式中,在起始或终止加尾之后,将用于进行连接反应的试剂加入至反应混合物。在一些实施方式中,连接受连接酶的影响。多种连接酶是可用的,它们的非限制性实例包括NAD-依赖性连接酶,其包括Taq DNA连接酶、丝状栖热菌(*Thermus filiformis*) DNA连接酶、大肠杆菌(*E. coli*) DNA连接酶、Tth DNA连接酶、水管致黑栖热菌(*Thermus scotoductus*) DNA连接酶(I和II)、热稳定连接酶、Ampligase热稳定DNA连接酶、VanC-型连接酶和9°N DNA连接酶;和ATP-依赖性连接酶,其包括T4 RNA连接酶在内、T4 DNA连接酶、T3 DNA连接酶、T7 DNA连接酶、Pfu DNA连接酶、DNA连接酶I、DNA连接酶III和DNA连接酶IV。

[0061] 在一些实施方式中,处理目标多核苷酸以差异修饰甲基化的胞嘧啶或未甲基化的胞嘧啶。在一些实施方式中,在扩增反应之前,如在包括所述目标多核苷酸的第一连接反应之后但是在后续扩增之前,在所述连接反应期间或者在所述连接反应之前(例如,在对目标多核苷酸加尾之前,或者作为样品制备的一部分)进行区分胞嘧啶甲基化状态的处理。在一些实施方式中,对来自特定来源的目标多核苷酸的一部分进行区分胞嘧啶甲基化状态的处理并且来自相同来源的另一部分(例如,作为来自共同溶液的不同等份)是未处理的,从而可以后续比较处理和未处理的样品。在某些过程中,比较有利于鉴别胞嘧啶甲基化状态,如在鉴别由于处理所产生的序列差异中。用于差异修饰甲基化或未甲基化的胞嘧啶的多种处理方法是可用的。选择性修饰甲基化的胞嘧啶的试剂的实例是TET蛋白家族(例如,TET1、TET2、TET3和CSSC4),其通过羟基化将胞嘧啶核苷酸5-甲基胞嘧啶转化为5-羟甲基胞嘧啶。可以选择性修饰5-羟甲基胞嘧啶,如通过用金属(VI)氧复合物(例如,锰酸盐($Mn(VI)O_4^{2-}$)、高铁酸盐($Fe(VI)O_4^{2-}$)、钨酸盐($Os(VI)O_4^{2-}$)、钌酸盐($Ru(VI)O_4^{2-}$)或者钼酸盐($Mo(VI)O_4^{2-}$))处理。用金属(VI)氧复合物处理使5-羟甲基胞嘧啶(5hmC)残基氧化成5-甲酰胞嘧啶(5fC)残基,其随后可以通过亚硫酸氢盐处理转化为尿嘧啶。在一些实施方式中,差异修饰甲基化的胞嘧啶或者未甲基化的胞嘧啶的处理包括用亚硫酸氢钠(亚硫酸氢盐)处理目标多核苷酸,其使未甲基化的胞嘧啶磺化,但是不能有效磺化甲基化的胞嘧啶。磺化的未甲基化的胞嘧啶易于自发脱氨基,这获得了磺化的尿嘧啶。然后,磺化的尿嘧啶可以在高pH去磺化为尿嘧啶。嘧啶尿嘧啶和胞嘧啶的碱基配对性质在本质上是不同的:在DNA中,尿嘧啶被认为是胸腺嘧啶的等价形式并因此在DNA的杂交或聚合期间与腺嘌呤配对,然而胞嘧啶在DNA的杂交或聚合期间与鸟嘌呤核苷配对。因此,对亚硫酸氢盐处理的DNA进行基因组测序或PCR可以用于区分基因组中未甲基化的胞嘧啶,相对于保持未转化的甲基化的胞嘧啶,其已转化为尿嘧啶。当与其它技术,如微阵列杂交和高通量测序结合时,这些技术能够作为大规模筛选方法。在(例如)US 9,822,394、US 9,115,386和US20150299781中描述了差异修饰和区分甲基化或未甲基化的胞嘧啶的方法的实例,以上专利作为参考并入本文。

[0062] 在一些实施方式中,使包括由于进行第一加尾反应和第一连接反应而连接至第一接头链的第一尾的目标多核苷酸扩增。在一些实施方式中,扩增包括使杂交至在早期连接反应中连接的第一接头链的第一引物延伸。在这些情况下,所述引物包括可杂交至所述连

接的接头链的至少一部分的序列。在一些实施方式中,所述可杂交的序列与它所杂交的序列互补。在一些实施方式中,所述引物杂交至存在于在连接反应期间连接的所有第一接头多核苷酸中的公用序列。在一些实施方式中,所述引物的可杂交部分的长度为约或大于约10、15、20、25、30、35、45、50或更多个核苷酸。通常,引物的可杂交部分包括所述引物的3'端。在一些实施方式中,所述第一引物包括一个或多个其它序列元件。其它序列元件的实例包括(但不限于)一个或多个引物退火序列或其互补序列(例如,测序引物)、一个或多个索引序列(例如,与可以用于鉴别所述索引与之相关的目标多核苷酸的来源的特定样品来源或反应有关的一个或多个序列)、一个或多个限制性内切酶识别位点、一个或多个探针结合位点(例如,用于连接至测序平台,如用于大规模平行测序的流通池,如Illumina, Inc.所开发的流通池)、一个或多个随机或近-随机序列(例如,在一个或多个位置,从两个或更多个不同的核苷酸组随机选择的一个或多个核苷酸,其中在包含所述随机序列的接头池中代表一个或多个位置选择的所述不同核苷酸中的每一个)及其组合。序列元件可以具有任何适合的长度,如长度约或小于约3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50个或更多个核苷酸。

[0063] 多种扩增方法对于扩增包含连接至第一接头链的第一尾的目标多核苷酸是可用的,并且包括指数和非指数(例如,线性)方法两者。在指数扩增中,引物延伸产物用作产生与所述第一尾互补的进一步的引物延伸产物的模板。作为对比,线性扩增反应通常设计以最大程度减少或消除在所述反应期间所形成的其它引物延伸产物的脱模板的(templated off)引物延伸产物的形成。在一些实施方式中,包括连接至第一接头链的第一尾的目标多核苷酸的扩增是线性扩增。扩增的第一步包括引物退火,其中第一引物杂交至连接至所述尾的接头链。如果引物杂交位点包括所述接头的双-链部分,则将首先暴露所述模板链中的杂交位点。可以通过使所述接头的非模板链变性和/或降解来实现杂交位点的暴露。变性可以包括热变性,如加热至约或大于约90℃或95℃一段时间(例如,约或大于约1、2、3、4、5、10或更多分钟)。多种方法对于降解所述接头的非模板链是可用的,并且可以基于要降解的链的组成适当选择。例如,当所述链包括一个或多个RNA碱基时,核糖核酸酶(例如,RNA酶H或RNA酶A)可以用于降解所述非模板链。作为其它实例,当所述接头的非模板链包括一个或多个尿嘧啶碱基时,可以通过添加尿嘧啶-特异性切除试剂(USER)酶来引起降解,所述USER酶是尿嘧啶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。

[0064] 多种用于线性扩增的方法是可用的,并且实例包括等温和非等温方法。在非等温方法中,所述方法包括在不同温度进行的变性和引物延伸步骤。变性释放出在模板上所形成的引物延伸产物,从而露出引物杂交位点以用于和所述引物的另一个拷贝杂交。第一引物的其它拷贝的延伸产生了来自相同模板的另一种引物延伸产物,并且可以通过几个变性和延伸“循环”来重复整个过程。在一些实施方式中,使用非等温方法,并且循环次数为约或至少约2、5、10、15、20、25或更多次。等温线性扩增方法的实例为单引物等温扩增(SPIA)。一般地,SPIA包括具有3'DNA部分和5'RNA部分的复合引物的延伸,所述RNA部分被RNA酶H的降解,所述复合引物的另一个拷贝的退火和通过具有链置换活性的聚合酶的所述复合引物的另一个拷贝的延伸,所有这些可以在相同温度发生。这些及其它扩增反应的进一步描述可见于(例如)US20170362636 A1,其作为参考并入本文。在一些实施方式中,扩增产生了与模

板目标多核苷酸互补的多个单链拷贝,其包括与所述第一尾和所述连接的第一接头链的至少一部分互补的序列。在一些实施方式中,选择扩增条件以产生约或小于约1、2、3、4、5、6、7、8、9、10、15、20、25、30、40、50、100、200、500或更多个目标多核苷酸的拷贝。

[0065] 在一些实施方式中,对使用第一引物的扩增反应的扩增产物进行加尾反应,其称为第二加尾反应。所述第二加尾反应通过不依赖模板的聚合反应将第二尾添加至多个扩增的目标多核苷酸的每一个。如第一加尾反应一样,所述尾的长度和核苷酸序列将部分基于由其聚合成所述尾的核苷酸的类型(例如,A、T、G和C中的1、2、3或4个)、反应持续时间、所使用的聚合酶和其它试剂的存在(例如,包含在聚合反应期间杂交至第二尾的悬突的接头)。对于第二加尾反应,如以上所提供的,有关尾形成和组成的考虑通常同样适用。在一些实施方式中,所述尾仅聚合至一个或多个扩增的目标多核苷酸的3'端。在一些实施方式中,所述第二加尾反应设计以产生具有与第一尾相同或基本相同的序列的尾,或与之互补的序列。例如,可以从仅有腺嘌呤碱基的池形成第一和第二尾,从而形成多聚-A尾。当对与加尾的目标多核苷酸模板互补的扩增产物进行第二加尾反应时,所得的第二-加尾多核苷酸将包括位于一端的多聚-A尾和邻近于第一尾所杂交至的接头链的互补序列的至少一部分的多聚-T尾。作为其它实例,所述第一尾可以是多聚-A尾并且所述第二尾可以是多聚-T尾。当对与所述加尾的目标多核苷酸模板互补的扩增产物进行第二加尾反应时,该实例中的结果将是具有两个多聚-T延伸的多核苷酸,一个来自第一尾,一个来自第二尾。在一些实施方式中,第二加尾反应设计以产生具有与第一尾不同的序列的尾,如通过使用在第一加尾反应中所使用的池中未使用的第二加尾反应的核苷酸池中的一个或多个核苷酸。不同的第一和第二尾的多种组合是可能的。尾组合的非限制性实例包括:(a)一个尾由一种类型的核苷酸组成,并且另一个尾由另一种类型的核苷酸组成;(b)一个尾由一种类型的核苷酸组成,并且另一个尾包括两种或更多种类型的核苷酸或者由其组成;(c)两个尾均包括两种或更多种类型的核苷酸或者由其组成,但是每个尾包含至少一种类型的在另一个尾中不含的核苷酸。在一些实施方式中,所述第一尾,所述第二尾或两者选自多聚-A、多聚-C和多聚-C/T。

[0066] 在一些实施方式中,第二加尾反应包括接头(称为第二接头),所述接头包含杂交至第二尾的悬突。所述悬突可以在多核苷酸延伸反应期间杂交至尾;然而,在通过不依赖于模板的聚合酶引发的不依赖模板的聚合反应中,这种杂交不否认所述反应状态是不依赖于模板的。所述第二接头包括至少一个单链区(所述悬突)和至少一个双链区(紧邻所述悬突)。所述第二接头可以包括位于两端的悬突,并且包括相同或不同的链。例如,可以通过杂交位于较长寡核苷酸中间的短寡核苷酸来形成双-链区。作为另一个实例,两个寡核苷酸可以彼此杂交,从而通过寡核苷酸之一形成位于一端的悬突,并且通过另一个寡核苷酸形成位于另一端的悬突。在一些实施方式中,仅在一端存在悬突,从而另一端以配对的核苷酸结束(也称为“平端”)。还可以通过杂交大于两个寡核苷酸来形成接头,并且所述接头可以包括位于双链区之间的内部单链区(例如,作为在沿长寡核苷酸相隔一个或多个核苷酸的区域杂交至相同长寡核苷酸的两个短寡核苷酸中)。在一些实施方式中,仅在5'或3'端中的一个上存在单一悬突。在一些实施方式中,所述悬突是3'-悬突。在一些实施方式中,所述接头具有3'悬突和5'悬突两者。如果使用第一和第二接头,则两个接头可以具有5'悬突和3'悬突两者。

[0067] 对于第二接头以及在第二加尾反应中其与第二尾的关系,如以上所提供的,与接

头的形成和组成有关的考虑,通常包括其与尾的关系是同样适用的。这些考虑包括(但不限于)悬突长度、悬突序列、核苷酸组成、封端的3'端的任选的使用和除所述悬突之外的一个或多个序列元件的任选的包含。在一些实施方式中,第二接头与第一接头相同。在一些实施方式中,第二接头的至少一部分不同于第一接头。在一些实施方式中,第一和第二接头包括一个或多个相同部分,但是其它部分不同。例如,所述第一和第二接头可以包括相同的引物结合序列,其设计使得在第二接头与扩增的目标多核苷酸连接之后,可以通过杂交至所述相同的引物结合序列或其互补序列的单一引物实现进一步的指数扩增。在一些实施方式中,第一和第二接头两者包括设计用于通过不同引物进行的指数扩增的引物结合序列。

[0068] 在一些实施方式中,将第二接头链连接至第二尾序列,如在连接反应中(称为第二连接反应)。在一些实施方式中,作为第二加尾反应在相同反应混合物中发生连接。在一些实施方式中,用于进行第二连接反应的试剂包括在第二加尾反应中。在一些实施方式中,在起始或终止加第二加尾之后,将用于进行第二连接反应的试剂加入至反应混合物。在一些实施方式中,通过连接酶引起连接,以上提供了连接酶的实例。在一些实施方式中,第二连接反应的产物是一系列多核苷酸,每个多核苷酸从5'至3'包括以下元件:(a)与连接的第一接头链的至少一部分互补的序列,(b)与第一尾互补的序列,(c)与目标多核苷酸互补的序列,(d)第二尾和(e)连接的第二接头链。为简单起见,这些连接产物及其扩增产物将被称为“两-连接的”或“双-连接的”目标多核苷酸,尽管应理解元件(a)可以不包括第一接头的整个连接的接头链,元件(b)是目标多核苷酸的互补拷贝,和元件(e)可以不包括整个连接的接头链(例如,就第二连接产物的扩增产物来说)。当在两-接头目标多核苷酸系列中表示多种不同的目标多核苷酸时,所述系列可以被称为文库。

[0069] 在一些实施方式中,在扩增反应中扩增所述双-连接的目标多核苷酸。在一些实施方式中,所述扩增包括使杂交至所述连接的第二接头链的第二引物延伸。在这些情况下,所述第二引物包括可杂交至所述连接的第二接头链的至少一部分的序列。在一些实施方式中,所述可杂交的序列与它所杂交的序列互补。在一些实施方式中,所述引物杂交至存在于在第二连接反应期间连接的所有第二接头多核苷酸中的公用序列(common sequence)。在一些实施方式中,所述引物的可杂交部分的长度为约或大于约10、15、20、25、30、35、45、50或更多个核苷酸。通常,引物的可杂交部分包括所述引物的3'端。在一些实施方式中,所述第二引物包括一个或多个其它序列元件。其它序列元件的实例包括(但不限于)一个或多个引物退火序列或其互补序列(例如,测序引物)、一个或多个索引序列(例如,与可以用于鉴别所述索引与之相关的目标多核苷酸的来源的特定样品来源或反应有关的一个或多个序列)、一个或多个限制性内切酶识别位点、一个或多个探针结合位点(例如,用于连接至测序平台,如用于大规模平行测序的流通池,如Illumina, Inc.所开发的流通池)、一个或多个随机或近-随机序列(例如,在一个或多个位置,从两个或更多个不同的核苷酸组随机选择的一个或多个核苷酸,其中在包含所述随机序列的接头池中代表一个或多个位置选择的所述不同核苷酸中的每一个)及其组合。序列元件可以具有任何适合的长度,如长度约或小于约3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50个或更多个核苷酸。

[0070] 使用第二引物的扩增可以是指数的或非指数的(例如,线性的)。扩增可以是等温的或非等温的。在一些实施方式中,第二连接反应的产物是基本线性的,并且扩增包括通过第二引物的延伸使所述连接产物成为双链的。在一些实施方式中,第二引物与第一引物相

同,或者包括与第一引物相同的可杂交序列。在一些实施方式中,第二引物不同于第一引物,如对于所述可杂交序列来说。在一些实施方式中,扩增反应包括第二引物和不同于所述第二引物的反向引物。在一些实施方式中,所述反向引物是第一引物(如以上对于第一连接的扩增产物所述的)。在一些实施方式中,所述反向引物杂交至位于所述第一引物杂交位置下游(也称为“巢式”)并且可以任选地包括一个或多个其它序列元件(例如,如上所述的任何一个或多个引物序列元件)的序列。在一些实施方式中,所述反向引物包括第一引物的可杂交序列的全部或部分,和不同于第一引物的一个或多个序列元件(例如,如上所述的任何一个或多个引物序列元件)。扩增的第一步包括引物退火,其中第二引物杂交至连接至第二尾的第二接头链。如果引物杂交位点包括所述第二接头的双-链部分,则将首先暴露所述模板链中的杂交位点。可以通过使所述接头的非模板链变性和/或降解来实现杂交位点的暴露,其示例性方法如上所述。线性扩增方法的非限制性实例如上所述。指数扩增方法的非限制性实例如上所述,并且在下文中更详细地进行了描述。

[0071] 在一些实施方式中,在使用第三引物和第四引物的扩增反应中扩增双-连接的目标多核苷酸,其中(i)所述第三引物杂交至所述第一引物的至少一部分的互补序列,和(ii)所述第四引物杂交至所述第二引物的至少一部分的互补序列。在一些实施方式中,该扩增步骤替代了使用第二引物的扩增步骤,在这种情况下,第三和第四引物类似于如上所述的第二引物和反向引物。在一些实施方式中,除使用第二引物的扩增(其可以或不包括使用反向引物的扩增)之外,使用第三和第四引物进行扩增。在一些实施方式中,第三引物的可杂交序列不同于第一引物的可杂交序列,和/或第四引物的可杂交序列不同于第二引物的可杂交序列。在一些实施方式中,第三引物对于第一引物是巢式的(nested)和/或第四引物对于第二引物是巢式的。

[0072] 在一些实施方式中,第三和/或第四引物的可杂交部分独立地选自约或大于约10、15、20、25、30、35、45、50或更多个核苷酸的长度。通常,引物的杂交部分包括所述引物的3'端。在一些实施方式中,第三和/或第四引物包括一个或多个其它序列元件(例如,如上所述的任何一个或多个引物序列元件)。序列元件可以具有任何适合的长度,如长度约或小于约3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50个或更多个核苷酸。在一些实施方式中,第三引物和第四引物是不同的,如对于总长、序列、可杂交序列的序列、一个或多个序列元件的存在、一个或多个序列元件的长度和一个或多个序列元件的序列中的一个或多个来说。

[0073] 在一些实施方式中,第三引物、第四引物或两者包括索引序列(也称为条形码,或简单地称为“索引”)。一般地,术语“索引”是指使得能够鉴别所述索引所相关的多核苷酸的一些特征的已知的核酸序列。在一些实施方式中,所述要鉴别的多核苷酸的特征是获得所述多核苷酸的来源(例如,样品、样品部份或反应)。在一些实施方式中,索引的长度为约或至少约3、4、5、6、7、8、9、10、11、12、13、14、15个或更多个核苷酸。在一些实施方式中,索引的长度短于10、9、8、7、6、5或4个核苷酸。在一些实施方式中,与一些多核苷酸有关的索引具有不同于与其它多核苷酸有关的索引的长度。一般地,索引具有足够的长度并且包括使得能够基于与之相关的索引鉴别来源的足够不同的序列,特别是从与混合物中来自不同来源的多核苷酸有关的不同的索引中。在一些实施方式中,可以在索引序列中的一个或多个核苷酸的突变、插入或缺失,如1、2、3、4、5、6、7、8、9、10个或更多个核苷酸的突变、插入或缺失之后,准确鉴别索引和与之相关的来源。在一些实施方式中,多个索引中的每个索引在多个,

至少三个核苷酸位置,如至少3、4、5、6、7、8、9、10个或更多个核苷酸位置不同于任何其它索引。可以在来自不同来源的多核苷酸的池中表示多个索引,每个来源包括包含一个或多个索引的多核苷酸,所述索引不同于在所述池中来源于其它来源的多核苷酸中包含的索引。在本文中强调所述索引仅需要在给定实验内是唯一的。因此,相同索引可以用于给在不同实验中进行处理的不同样品加标签。另外,在某些实验中,用户可以使用相同索引给相同实验内的不同样品亚组加标签。例如,可以用相同索引给来源于具有特定表型的个体的所有样品加标签,例如,可以用第一索引给来源于对照(或者野生型)对象的所有样品加标签,同时可以用第二索引(不同于所述第一索引)给患有疾病病况的对象加标签。作为另一实例,可以期望使用不同索引给来源于相同来源的不同样品(例如,随时间所获得的样品,来源于组织内不同位点的样品或者进行不同处理的相同样品的不同等份(例如,使用或不使用亚硫酸氢盐处理))加标签。一旦连接索引,则可以合并包含不同索引的多核苷酸的池以用于进一步处理,如扩增和/或测序。通过测序,所述索引可以用于将来源于相同来源的序列分组,借此将具有一个或多个特定索引的序列与该来源相关联。在一些实施方式中,方法包括基于目标多核苷酸(或其互补序列或衍生物)相连的索引序列,鉴别所述目标多核苷酸所来源的样品。索引的实例及其在鉴别样品来源中的使用可见于US20140121116、US20150087535和US20120071331,以上专利作为参考并入本文。

[0074] 在一些实施方式中,所述方法包括指数扩增步骤。指数扩增包括(例如)包括正反向引物的反应,从而所述正向引物的引物延伸产物用作反向引物的引物延伸的模板,反之亦然。扩增可以是等温的或非等温的。多种用于目标多核苷酸扩增的方法是可用的,并且无限制地包括基于聚合酶链反应(PCR)的方法。可以在所述方法的多个步骤优化有利于通过PCR的靶标序列扩增的条件,并且所述条件取决于所述反应中的成份特性,如靶标类型、靶标浓度、要扩增的序列长度、靶标和/或一个或多个引物的序列、引物长度、引物浓度、所使用的聚合酶、反应体积、一种或多种成份与一种或多种其它成份的比等,它们中的一些或全部可以适当改变。一般地,PCR包括要扩增的靶标的变性(如果是双链)、一个或多个引物与所述靶标的杂交和通过DNA聚合酶的引物延伸的步骤,通过步骤重复(或“循环”)以扩增所述靶标序列。可以对于多种结局,如提高得率,降低假产物的形成和/或提高或降低引物退火的特异性来优化该方法中的步骤。优化方法包括扩增反应中成份的类型或量和/或所述方法中给定步骤的条件,如特定步骤的温度、特定步骤的持续时间和/或循环次数的调整。在一些实施方式中,扩增反应包括至少5、10、15、20、25、30、35、50或更多次循环。在一些实施方式中,扩增反应包括不超过5、10、15、20、25、35、50或更多次循环。循环可以包含任何步骤数,如1、2、3、4、5或更多个步骤。步骤可以包括适合于实现给定步骤目的,包括(但不限于)3'端延伸、引物退火、引物延伸和链变性的任何温度或温度梯度。步骤可以具有任何持续时间,其包括(但不限于)约或小于约1、5、10、15、20、25、30、35、40、45、50、55、60、70、80、90、100、120、180、240、300、360、420、480、540、600或更多秒,包括无限长直至手动中断。在一些实施方式中,在将来自独立样本或等份的目标多核苷酸(例如,双-接头目标多核苷酸)混合之前或之后进行扩增。PCR扩增技术的非限制性实例包括定量PCR(qPCR或实时PCR)、数字PCR和靶标-特异性PCR。

[0075] 用于在PCR中使用的聚合酶的非限制性实例包括热稳定性DNA聚合酶,如嗜热栖热菌(*Thermus thermophilus*) HB8聚合酶;死亡嗜热菌(*Thermus oshimai*)聚合酶;水管致黑

栖热菌 (*Thermus scotoductus*) 聚合酶;嗜热栖热菌 (*Thermus thermophilus*) 聚合酶;水生栖热菌 (*Thermus aquaticus*) 聚合酶 (例如, **AmpliTaq®**FS或Taq (G46D;F667Y);激烈热球菌 (*Pyrococcus furiosus*) 聚合酶;热球菌种 (*Thermococcus* sp.) (9°N-7株) 聚合酶;Tsp聚合酶;Phusion高保真度DNA聚合酶 (ThermoFisher);及其突变体、变体或衍生物。用于一些PCR反应的聚合酶的其它实例包括 (但不限于) DNA聚合酶I、突变的DNA聚合酶I、克列诺片段、克列诺片段 (3'至5'核酸外切酶-)、T4 DNA聚合酶、突变的T4 DNA聚合酶、T7 DNA聚合酶、突变的T7 DNA聚合酶、phi29 DNA聚合酶和突变的phi29 DNA聚合酶。在一些实施方式中,使用了热启动聚合酶。热启动聚合酶是需要热激活的DNA聚合酶的修饰形式。通常,以未激活的状态提供热启动酶。通过热激活,释放修饰或修饰因子,从而产生激活的酶。一些热启动聚合酶可得自多种商品化来源,如Applied Biosystems;Bio-Rad;ThermoFisher;New England Biolabs;Promega;QIAGEN;Roche Applied Science;Sigma-Aldrich;等。

[0076] 在一些实施方式中,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为连接酶链式反应 (LCR) (参见,例如,美国专利No.5,494,810和5,830,711);转录介导的扩增 (TMA) (参见,例如,美国专利No.5,399,491、5,888,779、5,705,365、5,710,029);核酸序列-基扩增 (NASBA) (参见,例如,美国专利No.5,130,238);信号介导的RNA扩增技术 (SMART) (参见,例如,Wharam等人,Nucleic Acids Res.2001,29,e54);链置换扩增 (SDA) (参见,例如,美国专利No.5,455,166);嗜热SDA (参见,例如,美国专利No.5,648,211);滚环扩增 (RCA) (参见,例如,美国专利No.5,854,033);DNA的环介导等温扩增 (LAMP) (参见,例如,美国专利No.6,410,278);蜗牛酶-依赖性扩增 (HDA) (参见,例如,美国专利申请20040058378);基于SPIA的指数扩增方法 (参见,例如,美国专利No.7,094,536);和环形蜗牛酶-依赖性扩增 (cHDA) (例如,美国专利申请20100075384)。

[0077] 在一些实施方式中,方法包括对双-连接的多核苷酸测序。在一些实施方式中,所述方法包括对使用第二引物的扩增产物测序。在一些实施方式中,所述方法包括对使用第三和第四引物的扩增产物测序。多种测序方法是可用的,特别是高通量测序方法。实例无限制地包括通过Illumina (如 **HiSeq®**和 **MiSeq®**的测序系统)、Life Technologies (Ion **Torrent®**、**SOLiD®**等) 生产的测序系统、Roche's 454Life Sciences系统、Pacific Biosciences系统、Oxford Nanopore Technologies的纳米孔测序平台等。在一些实施方式中,测序包括产生具有约或大于约50、75、100、125、150、175、200、250、300或更多个核苷酸长度的读序。在一些实施方式中,测序包括通过合成方法测序,其中当将它们加入至正在生长的引物延伸产物中时,反复鉴别个体核苷酸。焦磷酸测序是通过测定所得合成混合物中测序反应副产物,即焦磷酸盐的存在来鉴别核苷酸掺入的合成方法测序的实例,其实例描述可见于US 6,210,891。根据一些测序方法,将引物/模板/聚合酶复合物固定在基底上并将所述复合物与标记的核苷酸接触。在US20160304954、US7,033,764、US 7,416,844和W02016077602中描述了测序技术的其它非限制性实例。

[0078] 在一些情况下,如本文所述,多种类型的测序反应可以包括多种样品处理单元。样品处理单元可以包括 (但不限于) 基本同时处理多个样品组的多道、多通道、多孔及其它方式。另外,所述样品处理单元可以包括多个样品室以有利于同时处理多个运行 (run)。在一些实施方式中,使用多路测序进行同时测序反应。在一些实施方式中,对多核苷酸测序以同时 (如在单个反应或反应容器中) 产生约或大于约5000、10000、50000、100000、1000000、

5000000、10000000或更多个测序读序。可以对测序反应的全部或部分进行后续数据分析。当多核苷酸与索引序列有关时,数据分析可以包括基于索引序列对序列分组以用于共同分析,和/或与和一个或多个不同索引有关的序列相比较。

[0079] 在一些实施方式中,序列分析包括一个或多个读序与参考序列(例如,对照序列、参考群体的测序数据、来自相同对象的不同组织的测序数据、相同对象在另一时间点的测序数据或者参考基因组)的比较,如通过实施比对。在典型的比对中,紧靠参考中非匹配碱基的测序读序中的碱基表示在该点发生了替换突变。类似地,当一个序列包括紧靠另一序列中的碱基的缺口时,则推断发生了插入或缺失突变(“indel”)。当期望指明一个序列正与另一个序列比对时,则有时将比对称称为成对比对。多重序列比对一般是指两条或更多条序列的比对,包括(例如)通过一系列成对比对。在一些实施方式中,对比对打分包括设定替换和indel概率的值。当比对单个碱基时,匹配或不匹配有助于通过替换概率的比对得分。通过空位罚分从比对得分中扣除indel。空位罚分和替换概率可以基于经验知识或者序列如何突变的在先假定。它们的值影响所得的比对。用于实施比对的算法的实例无限制地包括Smith-Waterman(SW)算法、Needleman-Wunsch(NW)算法、基于Burrows-Wheeler变换(BWT)的算法和hash函数比对器,如Novoalign(Novocraft Technologies;得自www.novocraft.com)、ELAND(Illumina, San Diego, Calif.)、SOAP(得自soap.genomics.org.cn)和Maq(得自maq.sourceforge.net)。实施BWT方法的一个示例性比对程序是得自由Geeknet(Fairfax, Va.)维护的SourceForge网站的Burrows-Wheeler比对器(BWA)。实施Smith-Waterman算法版本的比对程序是MUMmer,得自由Geeknet(Fairfax, Va.)维护的SourceForge网站。比对程序的其它非限制性实例包括:来自Kent Informatics(Santa Cruz, Calif.)的BLAT;SOAP2,来自北京基因组研究所(Beijing, Conn.)或者BGI Americas Corporation(Cambridge, Mass.);核苷酸数据库的有效大规模比对(ELAND)或者序列和变异一致评价(CASAVA)软件的ELANDv2组件(Illumina, San Diego, Calif.);来自Real Time Genomics, Inc. (San Francisco, Calif.)的RTG Investigator;来自Novocraft (Selangor, Malaysia)的Novoalign;Exonerate,欧洲生物信息学研究所(Hinxton, UK)、Clustal Ω ,来自University College Dublin(Dublin, Ireland);和来自University College Dublin(Dublin, Ireland)的ClustalW或ClustalX。

[0080] 在一些实施方式中,对扩增产物测序以相对于参考序列或者在无突变背景中检测序列变异,例如,插入、缺失、替换、重复、移位和/或稀有体细胞突变。在一些实施方式中,所述序列变异与疾病或性状有关。在一些实施方式中,所述序列变异与疾病或性状无关。一般地,对其存在与疾病或性状有关的统计学、生物学和/或功能性证据的序列变异称为“致病性遗传变体(causal genetic variant)”。单一致病性遗传变体可以与不止一种疾病或性状有关。在一些情况下,致病性遗传变体与孟德尔性状、非孟德尔性状或两者有关。致病性遗传变体可以表示为多核苷酸中的变化,如1、2、3、4、5、6、7、8、9、10、20、50个或更多个序列差异(如包含所述致病性遗传变体的多核苷酸和缺少所述致病性遗传变体的多核苷酸之间在相同相对基因组位置的序列差异)。致病性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)、拷贝数变异(CNV)、短串联重复序列(STR)、限制性片段长度多态性(RFLP)、简单序列重复(SSR)、可变数目串联重复序列(VNTR)、随机扩增多态性DNA(RAPD)、扩增片段长度多态性(AFLP)、逆转座子间扩增多态性(IRAP)、长散在和短散

在元件 (LINE/SINE)、长衔接重复 (LTR)、可动元件、逆转座子-微卫星扩增多态性、逆转座子-基插入多态性、序列特异性扩增多态性和可遗传的表观遗传修饰 (例如, DNA 甲基化)。致病性遗传变体可以包括一组密切相关的遗传变体。一些致病性遗传变体可以在 RNA 中作为序列变异施加影响。在该水平, 还通过一种 RNA 的存在或不存在来表示一些致病性遗传变体。一些致病性遗传变体在蛋白中导致产生了序列变化。已报道了一些致病性遗传变体。作为 SNP 的致病性遗传变体的实例为导致镰刀形红细胞贫血症的血红蛋白的 HbS 变异。作为 DIP 的致病性遗传变体的实例为导致囊性纤维化的 CFTR 基因的 δ -F508 突变。作为 CNV 的致病性遗传变体的实例为 21 三体, 其导致了唐氏综合征。作为 STR 的致病性遗传变体的实例为导致亨廷顿病的衔接重复。US2014121116 中描述了致病性遗传变体的其它非限制性实例。

[0081] 致病性遗传变体可以与之相关的疾病和基因靶标的实例包括 (但不限于) 21-羟化酶缺乏症、ABCC8-相关高胰岛素血症、ARSACS、软骨发育不全、色盲、磷酸腺苷脱氨酶 1、伴有神经病变的胼胝体发育不全、尿黑酸尿、 α -1-抗胰蛋白酶缺乏、 α -甘露糖苷贮积症、 α -肌聚糖病、 α -地中海贫血、阿尔茨海默症、血管紧张素 II 受体 I 型、载脂蛋白 E 基因分型、精氨酸琥珀酸尿、天冬氨酰葡萄糖胺尿症、共济失调与维生素 E 缺乏、共济失调毛细血管扩张症、1 型自身免疫性多内分泌腺病综合征、BRCA1 遗传性乳腺癌/卵巢癌、BRCA2 遗传性乳腺癌/卵巢癌、一种或多种其它类型的癌症、巴比二氏综合征、贝斯特氏卵黄状黄斑营养不良、 β -肌聚糖病、 β -地中海贫血、生物素缺乏症、布劳综合征 (Blau Syndrome)、布卢姆综合征 (Bloom Syndrome)、CFTR-相关病症、CLN3-相关神经元蜡样脂褐质沉积症、CLN5-相关神经元蜡样脂褐质沉积症、CLN8-相关神经元蜡样脂褐质沉积症、卡纳万病、肉碱棕榈酰转移酶 IA 缺乏症、肉碱棕榈酰转移酶 II 缺乏症、软骨毛发发育不全、脑海绵状血管畸形、无脉络膜、科恩综合征、先天性白内障、面部畸形和神经病变、糖基化 Ia 的先天性病症、糖基化 Ib 的先天性病症、先天性芬兰肾病、克罗恩氏病、胱氨酸过多症、DFNA 9 (COCH)、糖尿病和听力丧失、早发性原发性肌张力障碍 (DYTI)、赫利茨-皮尔森型交界型大疱性表皮松解症、FANCC-相关范可尼贫血、FGFR1-相关颅缝早闭、FGFR2 相关颅缝早闭、FGFR3-相关颅缝早闭、凝血因子 V 莱顿血栓形成倾向、凝血因子 V R2 突变血栓形成倾向、因子 XI 缺乏、因子 XIII 缺乏、家族性腺瘤性息肉病、家族性自律神经失调、家族性高胆固醇血症 B 型、家族性地中海热、游离唾液酸贮积病、伴有帕金森-17 的额颞叶痴呆、延胡索酸酶缺乏症、GJB2-相关 DFNA 3 非综合征性听力丧失和耳聋、GJB2-相关 DFNB 1 非综合征性听力丧失和耳聋、GNE-相关骨骼肌变性、半乳糖血症、戈谢病、葡萄糖-6-磷酸脱氢酶缺乏症、1 型戊二酸血症、1a 型糖原贮积症、1b 型糖原贮积症、II 型糖原贮积病、III 型糖原贮积病、V 型糖原贮积病、竹叶综合症、HFE-相关遗传性青铜色糖尿病、Halder AIM、血红蛋白 β -地中海贫血、遗传性果糖不耐症、遗传性胰腺炎、遗传性胸腺嘧啶-尿嘧啶尿症、己糖胺酶 A 缺乏症、出汗性外胚层发育不良 2、由胱硫醚 β 合成酶缺乏引起的高胱氨酸尿症、1 型高血钾周期性麻痹、高鸟氨酸血症-高氨血症-高瓜氨酸尿综合征、原发性 I 型高草酸尿症、原发性 2 型高草酸尿症、软骨发育不良、I 型低钾性周期性麻痹、2 型低钾性周期性麻痹、低磷酸酯酶症、婴儿肌病和乳酸性酸中毒 (致命和非致命形式)、异戊酸血症、克拉伯病、LGMD2I、莱伯氏遗传性视神经病、法国-加拿大型 Leigh 综合征、长链 3-羟酰辅酶 A 脱氢酶缺乏症、MLAS、MERRF、MTHFR 缺乏、MTHFR 不耐热变体、MTRNR1-相关性听力丧失和耳聋、MTTS1-相关性听力丧失和耳聋、MYH-相关性息肉病、1A 型枫糖尿病、1B 型枫糖尿病、麦-奥二氏 (McCune-Albright) 综合征、中链酰基辅酶 A 脱氢酶缺乏症、巨脑性白质脑病

伴皮层下囊肿、异染性脑白质营养不良、心肌线粒体病、线粒体DNA-相关性Leigh综合征和NARP、IV型粘脂贮积病、I型粘多糖贮积症、IIIA型粘多糖贮积症、VII型粘多糖贮积症、2型多发性内分泌腺瘤病、肌肉-眼-脑疾病、神经表型线形体肌病、由神经髓磷脂酶缺乏引起的尼-皮二氏病、C1型尼-皮二氏病、奈梅亨断裂综合征(Nijmegen Breakage Syndrome)、PPT1-相关的神经元蜡样质脂褐质沉积症、PROP1-垂体性激素缺乏症、Pallister-Hall综合征、先天性副肌强直症、彭德莱综合征(Pendred Syndrome)、过氧化物酶体双功能酶缺乏症、综合性发育障碍、苯丙氨酸羟化酶缺乏症、纤溶酶原激活物抑制剂I、常染色体隐性多囊肾病、凝血酶原G20210A血栓形成倾向、假维生素D缺乏性佝偻病、致密成骨不全症、波的尼亚型常染色体隐性视网膜色素变性、雷特综合征、1型肢近端型点状软骨发育不良、短链酰基辅酶A脱氢酶缺乏症、施-戴综合征(Shwachman-Diamond Syndrome)、舍格伦-拉松综合征(Sjogren-Larsson Syndrome)、史-伦-奥三氏综合征(Smith-Lemli-Opitz Syndrome)、痉挛性截瘫13、硫酸盐转运蛋白-相关骨软骨发育不良、TFR2-相关遗传性青铜色糖尿病、TPP1-相关神经元蜡样质脂褐质沉积病、致死性发育不良、甲状腺素运载蛋白淀粉样变性、三功能蛋白缺乏症、酪氨酸羟化酶-缺陷DRD、I型酪氨酸血症、威尔逊病、X-连锁型青少年视网膜劈裂症和齐薇格综合征谱。

[0082] 与癌症有关的序列变体的实例包括(但不限于)PIK3CA基因中的序列变体(存在于(例如)结肠直肠癌中;最常见位于外显子9(螺旋域)和外显子20(激酶域)内的两个“热点”区内;可以特异性靶向位置3140);BRAF基因中的序列变体(存在于(例如)恶性黑素瘤中,包括来源于无慢性日光-引起的损害的皮肤的黑色素瘤,特别是导致V600E的错义突变);EGFR基因中的序列变体(存在于(例如)非小细胞肺癌,特别是EGFR外显子18-21内,并且包括外显子19缺失和外显子21L858R点突变);KIT基因中的序列变体(存在于(例如)胃肠道间质瘤(GIST),特别是近膜域(外显子11)、胞外二聚化基序(外显子9)、酪氨酸激酶1(TK1)域(外显子13)以及酪氨酸激酶2(TK2)域和激活环(外显子17))。在一些实施方式中,鉴别了与癌症有关的一个或多个基因中的序列变体。与癌症有关的基因的非限制性实例包括PTEN;ATM;ATR;EGFR;ERBB2;ERBB3;ERBB4;Notch1;Notch2;Notch3;Notch4;AKT;AKT2;AKT3;HIF;HIF1a;HIF3a;Met;HRG;Bcl2;PPAR α ;PPAR γ ;WT1(胚胎性癌肉瘤);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;Bcl2;半胱天冬氨酸蛋白酶家族(9个成员:1、2、3、4、6、7、8、9、12);Kras;和Apc。

[0083] 在一些实施方式中,本发明的方法对于检测以相对低丰度存在的核酸物质具有高灵敏度。在一些实施方式中,所述低丰度物质是污染物(例如,食品或水中),复杂群体中的特定细菌(例如,环境测试中)和与疾病(例如,感染或致病遗传变体)有关的核酸。在一些实施方式中,所述方法检测以约或小于约1/1000、1/5000、1/10000、1/20000或者更低的量存在的核酸物质(例如,参考多核苷酸的突变体形式)。

[0084] 在一些实施方式中,方法还包括检测对象中疾病,如癌症或感染的存在或不存在。与大部分细胞一样,可以通过更新速率来鉴别癌细胞,其中老的细胞死亡并且被新的细胞取代。通常,与给定对象中的脉管系统接触,死细胞可以将DNA或DNA片段释放到血流中。这对于多个疾病阶段期间的癌细胞也是成立的。根据疾病阶段,还可以通过多种致病遗传变体,如拷贝数变异和稀有突变来鉴别癌细胞。使用本文所述的方法和系统,这种现象可以用

于检测对象中癌症的存在或不存在。在一些情况下,在疾病症状或其它标志发生之前检测癌症。可以检测的癌症的类型和数目包括(但不限于)血癌、脑癌、肺癌、皮肤癌、鼻咽癌、肝癌、骨癌、淋巴瘤、胰腺癌、皮肤癌、肠癌、直肠癌、甲状腺癌、膀胱癌、肾癌、口腔癌、胃癌、实体瘤、异质肿瘤、同质肿瘤(homogenous tumors)等。在一些实施方式中,本文所述的系统和方法用于帮助鉴别某些癌症。由本发明公开所述的系统和方法所产生的遗传数据可以使得开业医生能够帮助更好地鉴别特定癌症形式。经常地,癌症在组成和分期(staging)两者方面是异质的。基因谱数据可以使得能够鉴别特定癌症亚型,其在该特定亚型的诊断或治疗中可能是重要的。该信息还可以为对象或开业医生提供有关特定癌症类型预后的提示。通过检测某些致病遗传变体随时间的出现、消失或相对量的变化,可以追踪癌症发展的进展和/或对治疗方案的反应。

[0085] 在一个方面,本发明公开提供了用于在本文所述的方法(包括相对于本发明公开的任何多个其它方面和实施方式)中使用的组合物或者通过本文所述的方法产生的组合物。本发明公开的组合物可以包括任何一个或多个本文所述的成份。在一些实施方式中,组合物包括以下中的一种或多种:可以由其聚合尾的一个或多个核苷酸池、包含杂交至尾的3'悬突的一个或多个接头,用于差异修饰甲基化或未甲基化的胞嘧啶的一种或多种试剂、一种或多种扩增引物、一种或多种测序引物、一种或多种酶(例如,一种或多种聚合酶、反转录酶、连接酶、核糖核酸酶和糖基化酶)、一种或多种缓冲液(例如,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)、用于使用任何这些的试剂,包含任何这些的反应混合物和使用任何这些的说明。在一些实施方式中,提供了根据本文所述的方法所产生的多核苷酸。

[0086] 在一个方面,本发明公开提供了用于在本文所述的方法(包括相对于本发明公开的任何多个其它方面)中使用的反应混合物或者通过本文所述的方法产生的反应混合物。在一些实施方式中,所述反应混合物包含一个或多个本文所述的组合物。

[0087] 在一个方面,本发明公开提供了用于在任何本文所述的方法(包括相对于本发明公开的任何多个其它方面)中使用的试剂盒。在一些实施方式中,所述试剂盒包含一个或多个本文所述的组合物。还可以(无限制地)以任何量和/或组合(如,在相同试剂盒或相同容器中)提供所述试剂盒的成份。在一些实施方式中,试剂盒包括用于根据本发明所述的方法的用途的其它试剂。可以在任何适合的容器,包括(但不限于)试管、小瓶、烧瓶、瓶、安瓿瓶、注射器等中提供试剂盒成份。可以以可以直接在本发明所述的方法中使用的形式,或者以使用之前需要制备的形式,如冷冻干燥试剂的复原的形式提供所述试剂。可以以用于单次使用的等份或者作为从中可以获得(如在一些反应中)多次使用的储料来提供试剂。在一些实施方式中,试剂盒包含:(a)不依赖于模板的聚合酶;(b)可以通过不依赖于模板的聚合酶聚合的第一核苷酸池;(c)可以通过不依赖于模板的聚合酶聚合的第二核苷酸池;(d)第一接头,其包含可杂交至通过所述第一多核苷酸池的聚合所形成的尾的悬突;和(e)第二接头,其包含可杂交至通过所述第二多核苷酸池的聚合所形成的尾的悬突,其中所述第二接头包含与所述第一接头不同的序列。在一些实施方式中,所述试剂盒还包含一种或多种引物。在本文中,包括相对于本发明公开所述的多种方法,公开了聚合酶、核苷酸池、接头和引物的实例。

[0088] 在一个方面,本发明公开提供了用于实施本文所述的方法(包括相对于本发明公

开的任何多个其它方面)的系统,如计算机系统。应理解对于无辅助的人来说,实施本文所公开的方法的一些实施方式中所涉及的计算操作是不现实的,或甚至在大多数情况下,是不可能的。例如,在没有计算设备辅助的情况下,将来自样品的单个30bp的读序对人染色体中的任一个进行定位(mapping)可能需要数年的工作。当然,在低等位基因频率突变的可靠识别(call)需要将数千(例如,至少约10,000)或甚至数百万个读序对一个或多个染色体进行定位的情况下,无辅助的序列分析和比对的困难是复合的。因此,本文所述方法的一些实施方式不能够单独在人脑中进行,或者仅通过铅笔和纸来进行,而是必需使用计算系统,如包括程序化实施一个或多个分析过程的一个或多个处理器的系统。

[0089] 在一些实施方式中,本发明公开提供了包括用于实施多种计算机-执行的操作的程序指令和/或数据(包括数据结构)的有形和/或永久性计算机可读媒体或计算机程序产品。计算机-可读媒体的实例包括(但不限于)半导体存储器装置、磁媒体,如磁盘驱动器、磁带、光学媒体,如CD、磁光媒体,和具体配置以储存和执行程序指令的硬件装置,如只读存储器装置(ROM)和随机存取存储器(RAM)。计算机可读媒体可以直接通过终端用户控制或者所述媒体可以间接通过终端用户控制。直接控制的媒体的实例包括位于用户设备的媒体和/或不与其它实体共享的媒体。间接控制的媒体的实例包括用户通过外部网络和/或通过提供共享资源的服务,如“云”间接可访问的媒体。程序指令的实例包括机器代码,如通过编译器所产生的机器代码,和含有使用翻译器可以由计算机执行的高级代码的文件两者。

[0090] 在一些实施方式中,以电子格式提供了在本文所公开的方法和系统中使用的数据或信息。这种数据或信息的实例包括(但不限于)来源于核酸样品的测序读序、参考序列(包括仅提供或主要提供多态性的参考序列)、在测序读序制备中使用的一个或多个寡核苷酸的序列(包括其部分和/或其互补序列)、识别(calls),如癌症诊断识别、咨询建议、诊断等。如本文所使用的,以电子格式提供的数据或其它信息对于在机器上的存储和机器之间的传输可用。通常,以数字形式提供电子格式的数据并且所述数据可以作为比特和/或字节在多种数据结构、列表、数据库等中储存。所述数据可以以电子、光学等形式体现(embodied)。

[0091] 在一些实施方式中,本文提供了用于产生指示测试样品中多核苷酸序列的输出的计算机程序产品。所述计算机产品可以含有用于实施任何一个或多个上述用于制备多核苷酸文库的方法和任选地确定多核苷酸序列的指令。如所解释的,所述计算机产品可以包括非暂时性和/或有形计算机可读媒体,其具有记录其上的用于使得处理器能够确定所关心的序列的计算机可执行或可编译的逻辑(例如,指令)。在一个实例中,所述计算机产品可以包括计算机可读媒体,其具有记录其上的用于使得处理器能够诊断病况和/或确定所关心的核酸序列的计算机可执行或可编译的逻辑(例如,指令)。

[0092] 在一些实施方式中,使用改造或配置以实施如本文所述的方法的计算机处理系统来实施本文所述的方法(或其部分)。在一个实施方式中,所述系统包括改造或配置用于对多核苷酸测序以获得本文其它处(如相对于本文所述的任何多个方面)所述的序列信息类型的测序装置。在一些实施方式中,所述设备包括用于处理样品的组件,如液体处理机和测序系统,包括用于实施本文所述的任何多种方法的一个或多个步骤(例如,样品处理、多核苷酸纯化和多种反应(例如,加尾反应、连接反应、扩增反应和测序反应))的模块。

[0093] 在一些实施方式中,将序列或其它数据直接或间接输入计算机或存储在计算机可读媒体上。在一个实施方式中,将计算机系统直接连接至读取和/或分析来自样品的核酸序

列的测序装置。通过计算机系统界面提供来自这些工具的序列或其它信息。作为另外一种选择,从序列存储源,如数据库或其它资料档案库提供通过系统处理的序列。一旦对于处理设备可用,则存储装置或大容量存储装置至少暂时缓存或存储核酸序列。另外,存储装置可以存储多种染色体或基因组等的读序计数。存储器还可以存储用于分析序列或定位数据的多种例行程序(routines)和/或程序。在一些实施方式中,所述程序/例行程序包括用于实施统计分析的程序。

[0094] 在一个实例中,用户将多核苷酸样品提供至测序设备中。通过连接至计算机的测序设备收集和/或分析数据。计算机上的软件使得能够进行数据收集和/或分析。数据可以存储、显示(通过显示器或其它类似装置)和/或发送至另一位置。计算机可以连接到因特网,其用于将数据传输至远程用户(例如,医师、科学家或分析员)所使用的手持装置。应理解可以在传输之前存储和/或分析数据。在一些实施方式中,收集原始数据并发送至将分析和/或存储所述数据的远程用户或设备。传输可以通过因特网发生,但是还可以通过卫星或其它连接发生。替代地,数据可以存储在计算机-可读媒体上,并且可以将所述媒体运输至终端用户(例如,通过邮件)。远程用户可以处于相同或不同的地理位置,包括(但不限于)建筑物、城市、州、国家或大陆。

[0095] 在一些实施方式中,所述方法包括收集有关多个多核苷酸序列的数据(例如,读序和/或参考染色体序列)并将数据发送至计算机或其它计算系统。例如,可以将计算机连接至实验室设备,例如,采样设备、核苷酸扩增设备或核苷酸测序设备。然后,计算机可以收集通过实验室装置采集的可应用数据。所述数据可以在任何步骤,例如,在收集的同时实时地,在发送之前,在发送期间或与发送一同,或者在发送之后存储在计算机上。所述数据可以存储在可以从所述计算机提取的计算机-可读媒体上。可以将所收集或存储的数据从计算机传输至远程位置,例如,通过局域网或广域网,如因特网。可以在远程位置对所传输的数据进行多种操作。

[0096] 可以在本文所公开的系统、设备和方法中存储、传输、分析和/或利用的电子格式数据类型如下所示:通过测序核酸所获得的读序、参考基因组或序列、将测试样品识别为受影响、不受影响或未识别的阈值、与所关心的序列有关的医学病况的实际识别、诊断(与识别有关的临床状况)、来源于识别和/或诊断的对进一步测试的建议,来源于识别和/或诊断的治疗和/或监测方案。在一些实施方式中,使用独特的设备在一个或多个位置获得、存储、传输、分析和/或操作这些多种数据类型。处理选择覆盖广泛的选择范围。在所述范围的一端,在处理测试样品的位置存储和使用该信息的全部或大部分,例如,在医生的办公室或其它临床环境。在所述范围的另一端,在一个位置获得样品,在不同的位置处理所述样品并任选地测序,在一个或多个不同位置比对读序并进行识别,并且在其它位置(其可以是获得样品的位置)做出诊断、建议和/或方案。

实施例

[0097] 出于说明本发明的多种实施方式的目的提供了以下实施例,并且这些实施例不表示以任何方式限制本发明。本发明的实施例以及本文所述的方法是本发明的优选实施方式的代表,是示例性的,并且不旨在作为对本发明范围的限制。在如权利要求的范围所定义的本发明的精神内所涵盖的其中的变化和其它用途将是本领域技术人员能够想到的。

[0098] 实施例1:

[0099] NA12878基因组DNA得自Coriell Institute(Coriell Institute,NA12878)。通过Qubit dsDNA HS测定试剂盒(Thermo Fisher Scientific,Q32851)测量浓度,并且在文库制备中所使用的DNA的量为10ng。将DNA底物稀释至50 μ l IDTE缓冲液(IDT,11-05-01-09),并使用聚焦声学超声波剪切仪(Covaris,M220)剪切成约100-600bp的片段。超声处理参数设置如下:峰值入射功率50W,占空比20%,循环破碎系数(cycle per burst)200,持续时间150秒,温度6-8 $^{\circ}$ C。通过LabChip GXII touch 24(Perkin Elmer)确认所剪切的DNA片段的尺寸。

[0100] 如果未提及,则以2至3个技术重复来实施所有实验。

[0101] 使用根据EZ-96DNA甲基化-lightningTM MagPrep(Zymo,D5047)的修改规程进行亚硫酸氢盐转化步骤(BC)。将97.5 μ l Lightning转化试剂和15 μ l所剪切的基因组DNA或cfDNA加入至48-孔板(Thermo Fisher Scientific,AB0648)中。通过上下吸取来混合样品并将其以下列条件在热循环仪中培育:(i)98 $^{\circ}$ C,8分钟;(ii)54 $^{\circ}$ C,60分钟;(iii)4 $^{\circ}$ C储存多至20小时。将BC-处理的DNA样品转移至对于每个孔预加载了450 μ l M-结合缓冲液和7.5 μ l MagBinding珠的96-孔midi-板(Thermo Scientific,AB0859)。将组分彻底混合,并使所述板在室温下静置5分钟。然后,将所述板转移到磁力支架上保持另外5分钟,并除去上清液。用300 μ l M-清洗缓冲液清洗所述珠,并将所述珠与150 μ l L-脱磷酸缓冲液在室温下(20-30 $^{\circ}$ C)培育25分钟。将所述板置于磁力支架3分钟,弃去上清液,然后用300 μ l M-清洗缓冲液清洗所述珠两次。在清洗步骤之后,将所述板转移到55 $^{\circ}$ C的金属加热器(Illumina,SC-60-504,BD-60-601)上并保持30分钟以干燥所述珠,然后加入16 μ l M-洗脱缓冲液,并在55 $^{\circ}$ C培育另外4min。然后,将所述板移至磁力支架并保持1分钟,并且作为用于后续文库制备步骤的模板回收上清液。

[0102] 裂片(splinter)接头MDA1被设计为以具有以9:1的摩尔比随机合成的多个八个G或A。在第一加尾和连接步骤期间,它退火至单链DNA底物的3'端多聚-C/T尾(如图3,下图所示)。图2显示了形成MDA1的寡核苷酸的序列。通过将寡聚ATN-R2-Top和ATN-R2-Bot退火在一起制备了MDA1接头。详细地,将50 μ l每种寡聚核苷酸(100 μ M)混合并在95 $^{\circ}$ C培育10分钟,并使其在含有0.1mM EDTA和50mM NaCl的10mM Tris-HCl中缓慢冷却至室温。通过磷酸基使两个寡聚核苷酸的3'端封端以防止自连接。按照类似的策略,通过ATN-R1-Top和ATN-R1-Bot寡聚核苷酸制备了MDA2接头。图2还显示了形成MDA2的寡核苷酸的序列。形成MDA1、MDA2的寡核苷酸和表示为“锚定引物”的扩增引物的序列如表1所示。

[0103] 表1:

	寡聚核苷酸	序列	注释
[0104]	ATN-R2-Top	AGATCGGAAGAGCACACGTCTGAACTCCA GTCAC (SEQ ID NO: 4)	5'磷酸; 3'磷酸
	ATN-R2-Bot	GTGACTGGAGTTCAGACGTGTGCTCTTCC GATCTRRRRRRRR (SEQ ID NO: 5)	3'磷酸; R (G:A) =9:1 的预混合物
	ATN-R1-Top	AGATCGGAAGAGCGTCGTGTAGGGAAAG AGTGT (SEQ ID NO: 6)	5'磷酸; 3'磷酸
	ATN-R1-Bot	ACACTCTTTCCCTACACGACGCTCTTCCGA TCTTTTTTTTTTTT (SEQ ID NO: 7)	3'磷酸
	LAP (锚定引物)	GTGACTGGAGTTCAGACGTGTGCTCTTCC GATC (SEQ ID NO: 16)	

[0105] 通过将12.5μl DNA样品、1.5μl 10×CutSmart缓冲液(NEB,B7204S)、1μl虾碱性磷酸酶(NEB,M0371L)混合并在37℃培育30分钟来对亚硫酸氢盐转化的DNA片段进行末端-修复。通过在95℃培育5min并在冰上快速冷却,使所述产物进一步变性。

[0106] 然后,在含有预处理的DNA底物、1×CutSmart缓冲液、0.25mM CoCl₂(NEB,B0252S)、0.025mMβ-烟酰胺腺嘌呤二核苷酸(NEB,B9007S)、0.09mM dCTP(Roche,11934520001)、0.01mM dTTP(Roche,11934546001)、1μM MDA1接头、0.5U/μl大肠杆菌(E.coli)连接酶(NEB,M0205L)和0.5U/μl末端脱氧核苷酸转移酶(TdT;NEB,M0315S)的20μl反应体积中进行第一连接反应。将所述反应在37℃培育30分钟,然后在95℃加热2分钟并保持在4℃。

[0107] 在存在1×KAPA HiFi HotStart Uracil+ReadyMix(KAPA,KK2802)和0.91μM锚定引物的情况下使所述连接产物延伸并线性扩增。按照以下热温度谱进行线性扩增反应:(i) 95℃,5分钟;(ii) 98℃,20秒,62℃,30秒,72℃,1分钟,循环15次和(iii) 72℃,5分钟。反应完成后,通过用2.5×AMPure XP珠(Beckman Coulter,A63881)纯化来更换缓冲液并用11.5 μl洗脱缓冲液(10mM Tris-HCl,pH 8.0)洗脱。

[0108] 在含有10μl纯化的DNA产物、1×CutSmart缓冲液、0.25mM CoCl₂(NEB,B0252S)、0.025mMβ-烟酰胺腺嘌呤二核苷酸(NEB,B9007S)、0.1mM dATP(Roche,11934511001)、1μM MDA2、0.5U/μl大肠杆菌(E.coli)连接酶(NEB,M0205L)和0.5U/μl末端脱氧核苷酸转移酶(NEB,M0315S)的20μl反应体积中进行第二连接反应。将所述反应在37℃培育30分钟,然后在95℃加热2分钟并保持在4℃。与包括“Y”接头的连接反应产物(上图)相比,图3(下图)提供了第二连接的示例性产物的图示。

[0109] 在含有20μl上述DNA产物、1×KAPA HiFi缓冲液、dNTP、1μM引物F和引物R以及1U/μl KAPA HiFi聚合酶的50μl反应中进行连接产物的PCR富集。PCR程序如下所示:(i) 95℃,5分钟;(ii) 98℃,20秒,60℃,30秒,72℃,1分钟,循环12次和(iii) 72℃,10分钟。使用Agencourt AMPure XP珠(Beckman Coulter,A63881)纯化所述PCR产物并在18μl EB(10mM Tris-HCl,pH 8.0)中洗脱。引物F的序列为ACACTCTTTCCCTACACGACGCTCTTCCGATCT(SEQ ID NO:17)。引物R的序列为GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC(SEQ ID NO:18)。

[0110] 将15μl纯化的DNA文库(50-200ng/μl)与4μl阻断剂混合物(blocker mix)良好混合,并在热循环仪中按以下条件培育:(i) 95℃,5分钟;(ii) 65℃保持。同时,将10μl杂交缓冲液(13×SSPE;13.5mM EDTA;13×Denhart溶液;0.45%SDS)、0.5μlRNA酶-抑制剂和0.5μl

Agilent SureSelect定制组探针池 (Custom Panel Probe Pool) 在65℃预加热2分钟。然后,将DNA-阻断剂混合物的整个内容物转移至探针混合物,从而使杂交反应在65℃进行16-24小时。

[0111] 图4显示了示例性的毛细管电泳分析图,其显示了在PCR富集之后预捕获文库片段的示例性尺寸分布。预期峰尺寸为200-400bp。将所有文库加载至HT DNA高灵敏度LabChip试剂盒 (Perkin Elmer)。300bp处的最高的曲线显示了当提供有1×MDA1接头时所连接的底物。自上而下的下一条曲线分别代表2×、3×和4×接头。数据表明1×MDA1足以连接接头,并且在这些条件下,连接效率随着MDA1浓度的提高而降低。

[0112] 杂交后,通过用200μl结合缓冲液 (10mM Tris-HCl pH 8.0、0.5mM EDTA、1M NaCl) 清洗四次来调整25μl的抗生蛋白链菌素-缀合的DynaBeads™ (Thermo Fisher Scientific, 65602)。在25℃,在热混合器中以600RPM实施30分钟的DNA捕获。为了除去通过非特异性结合沉淀的 (pulled down) 非靶标DNA,首先用500μl清洗缓冲液1在室温下清洗所述珠1次,然后用清洗缓冲液2 (10mM Tris-HCl pH 8.0, 0.02% Triton X-100) 在65℃清洗3次。然后,将所述珠在20μl的洗脱缓冲液 (10mM Tris-HCl, pH 8.0) 中再悬浮并用作随后索引PCR步骤的模板。

[0113] 对于多路测序,将5μl索引引物 (预混合的i5和i7,分别20μM) 加入含有20μl再悬浮的T1珠和25μl Kapa HiFi hot start ready混合物 (Kapa Biosystem, KK2602) 的50μl反应中。PCR程序如下所示: (i) 98℃, 45秒; (ii) 98℃, 15秒, 60℃, 30秒, 72℃, 1分钟, 循环12次和 (iii) 72℃, 5分钟。将纯化的DNA文库在20μl EB中洗脱并通过Qubit dsDNA HS测定试剂盒定量。索引引物i5的序列为AATGATACGCGACACCGAGATCTACACGTTAGTTCACACTCTTCCCTACACGACG (SEQ ID NO:19; 下划线序列对应于示例性索引序列)。索引引物i7的序列为CAAGCAG AAGACGGCATACGAGATGTGATGCCGTGACTGGAGTTCAGACGTG (SEQ ID NO:20; 下划线序列对应于示例性索引序列)。

[0114] 根据生产商的说明,使用PE150循环运行,在Illumina HiSeq 2500或NovaSeq上对索引PCR步骤产物测序。通过分析性管道 (analytical pipeline) 分离 (de-multiplexed) FASTQ序列,并分析常规文库质量指标。下表2A和2B显示了说明性文库生物信息学QC总结表。

[0115] 表2A:

样品名称	输入DNA (ng)	总PF读序	定位比	插入尺寸
MDA1-1X	10	8,666,046	95.79%	188
MDA1-2X	10	7,577,663	95.87%	187
MDA1-3X	10	8,150,850	96.12%	187
MDA1-4X	10	8,851,169	96.01%	189

[0117] 表2B:

样品名称	覆盖的复	命中%	去重的中	预去重的中值	均一度 (0.2×
------	------	-----	------	--------	-----------

[0119]		杂度		值覆盖度	覆盖度	平均值)
	MDA1-1X	65.26%	64.16%	366	537	96.40%
	MDA1-2X	64.45%	65.43%	323	478	96.30%
	MDA1-3X	59.59%	68.02%	337	537	96.20%
	MDA1-4X	52.65%	67.63%	324	580	96.30%

[0120] 图1提供了示例性文库制备方法的总览图示。通过适当的dNTP,使用TdT进行加尾步骤以在ssDNA片段的3'端产生均聚物或近-均聚物尾。所述均聚物退火至上链(top strand)中含有5'磷酸基的接头的3'悬突。通过连接酶催化的连接反应使ssDNA片段的3'端封闭以防止过度加尾。通过锚定引物竞争掉所述接头的下链(bottom strand),从而暴露出线性扩增过程的起始位点。将所扩增的ssDNA链用作第二轮加尾和连接的模板,然后使其产物扩增。

[0121] 实施例2:

[0122] NA12878基因组DNA得自Coriell Institute(Coriell Institute,NA12878)。通过Qubit dsDNA HS测定试剂盒(Thermo Fisher Scientific,Q32851)测量浓度,并且在文库制备中所使用的DNA的量在2-30ng的范围内。将DNA底物稀释至50 μ l IDTE缓冲液(IDT,11-05-01-09),并使用聚焦声学超声波剪切仪(Covaris,M220)剪切成约100-600bp的片段。超声处理参数设置如下:峰值入射功率50W,占空比20%,循环破碎系数(cycle per burst)200,持续时间150秒,温度6-8 $^{\circ}$ C。通过LabChip GXII touch 24(Perkin Elmer)确认所剪切的DNA片段的尺寸。

[0123] 血浆样品得自人抽血样品。使用QiaAmp循环核酸试剂盒(Qiagen,55114)提取无细胞DNA(cfDNA)。作为NA12878基因组DNA,通过Qubit dsDNA HS测定试剂盒定量cfDNA,但是不对其进行片段化。

[0124] 如果未提及,则以2至3个技术重复来实施所有实验。

[0125] 使用根据EZ-96DNA甲基化-lightningTM MagPrep(Zymo,D5047)的修改规程进行亚硫酸氢盐转化步骤(BC)。将97.5 μ l Lightning转化试剂和15 μ l所剪切的基因组DNA或cfDNA加入至48-孔板(Thermo Fisher Scientific,AB0648)中。通过上下吸取来混合样品并将其以下列条件在热循环仪中培育:(i)98 $^{\circ}$ C,8分钟;(ii)54 $^{\circ}$ C,60分钟;(iii)4 $^{\circ}$ C储存多至20小时。将BC-处理的DNA样品转移至对于每个孔预加载了450 μ l M-结合缓冲液和7.5 μ l MagBinding珠的96-孔midi-板(Thermo Scientific,AB0859)。将组分彻底混合,并使所述板在室温下静置5分钟。然后,将所述板转移到磁力支架上保持另外5分钟,并除去上清液。用300 μ l M-清洗缓冲液清洗所述珠,并将所述珠与150 μ l L-脱磷酸缓冲液在室温下(20-30 $^{\circ}$ C)培育25分钟。将所述板置于磁力支架3分钟,弃去上清液,然后用300 μ l M-清洗缓冲液清洗所述珠两次。在清洗步骤之后,将所述板转移到55 $^{\circ}$ C的金属加热器(Illumina,SC-60-504,BD-60-601)上并保持30分钟以干燥所述珠,然后加入16 μ l M-洗脱缓冲液,并在55 $^{\circ}$ C培育另外4min。然后,将所述板移至磁力支架并保持1分钟,并且作为用于后续文库制备步骤的模板回收上清液。

[0126] 裂片(splinter)接头MDA1设计以具有以9:1的摩尔比随机合成的多个八个G或A。在第一加尾和连接步骤期间,它退火至单链DNA底物的3'端多聚-C/T尾(如图3,下图所示)。图2显示了形成MDA1的寡核苷酸的序列。如实施例1制备MDA1和MDA2接头。形成MDA1、MDA2的

寡核苷酸和表示为“锚定引物”的扩增引物的序列如以上表1所示。

[0127] 通过将12.5 μ l DNA样品、1.5 μ l 10 $\times$ CutSmart缓冲液(NEB,B7204S)、1 μ l虾碱性磷酸酶(NEB,M0371L)混合并在37 $^{\circ}$ C培育30分钟来对亚硫酸氢盐转化的DNA片段进行末端-修复。通过在95 $^{\circ}$ C培育5min并在冰上快速冷却,使所述产物进一步变性。

[0128] 然后,在含有预处理的DNA底物、1 $\times$ CutSmart缓冲液、0.25mM CoCl₂(NEB,B0252S)、0.025mM β -烟酰胺腺嘌呤二核苷酸(NEB,B9007S)、0.09mM dCTP(Roche,11934520001)、0.01mM dTTP(Roche,11934546001)、1 μ M MDA1接头、0.5U/ μ l大肠杆菌(E.coli)连接酶(NEB,M0205L)和0.5U/ μ l末端脱氧核苷酸转移酶(TdT;NEB,M0315S)的20 μ l反应体积中进行第一连接反应。将所述反应在37 $^{\circ}$ C培育30分钟,然后在95 $^{\circ}$ C加热2分钟并保持在4 $^{\circ}$ C。

[0129] 在存在1 $\times$ KAPA HiFi HotStart Uracil+ReadyMix(KAPA,KK2802)和0.91 μ M锚定引物的情况下使所述连接产物延伸并线性扩增。按照以下热温度谱进行线性扩增反应:(i) 95 $^{\circ}$ C,5分钟;(ii) 98 $^{\circ}$ C,20秒,62 $^{\circ}$ C,30秒,72 $^{\circ}$ C,1分钟,循环15次和(iii) 72 $^{\circ}$ C,5分钟。反应完成后,通过用2.5 $\times$ AMPure XP珠(Beckman Coulter,A63881)纯化来更换缓冲液并用11.5 μ l洗脱缓冲液(10mM Tris-HCl,pH 8.0)洗脱。

[0130] 在含有10 μ l纯化的DNA产物、1 $\times$ CutSmart缓冲液、0.25mM CoCl₂(NEB,B0252S)、0.025mM β -烟酰胺腺嘌呤二核苷酸(NEB,B9007S)、0.1mM dATP(Roche,11934511001)、1 μ M MDA2、0.5U/ μ l大肠杆菌(E.coli)连接酶(NEB,M0205L)和0.5U/ μ l末端脱氧核苷酸转移酶(NEB,M0315S)的20 μ l反应体积中进行第二连接反应。将所述反应在37 $^{\circ}$ C培育30分钟,然后在95 $^{\circ}$ C加热2分钟并保持在4 $^{\circ}$ C。与包括“Y”接头的连接反应产物(上图)相比,图3(下图)提供了第二连接的示例性产物的图示。

[0131] 在含有20 μ l上述DNA产物、1 $\times$ KAPA HiFi缓冲液、dNTP、1 μ M引物F和引物R以及1U/ μ l KAPA HiFi聚合酶的50 μ l反应中进行连接产物的PCR富集。PCR程序如下所示:(i) 95 $^{\circ}$ C,5分钟;(ii) 98 $^{\circ}$ C,20秒,60 $^{\circ}$ C,30秒,72 $^{\circ}$ C,1分钟,循环12次和(iii) 72 $^{\circ}$ C,10分钟。使用Agencourt AMPure XP珠(Beckman Coulter,A63881)纯化所述PCR产物并在18 μ l EB(10mM Tris-HCl,pH 8.0)中洗脱。

[0132] 图5A-C显示了示例性的毛细管电泳分析图,其显示了在PCR富集之后预捕获文库片段的示例性尺寸分布。预期峰尺寸为200-400bp。所述预捕获的文库得率随输入的增加而增加。以10ng输入,cfDNA具有比所述剪切的基因组DNA(gDNA)更高的得率。将所有文库加载至HT DNA高灵敏度LabChip试剂盒(Perkin Elmer)。

[0133] 将15 μ l纯化的DNA文库(50-200ng/ μ l)与4 μ l阻断剂混合物良好混合,并在热循环仪中按以下条件培育:(i) 95 $^{\circ}$ C,5分钟;(ii) 65 $^{\circ}$ C保持。同时,将10 μ l杂交缓冲液(13 $\times$ SSPE;13.5mM EDTA;13 $\times$ Denhart溶液;0.45% SDS)、0.5 μ l RNA酶-抑制剂和0.5 μ l Agilent SureSelect定制组探针池(Custom Panel Probe Pool)在65 $^{\circ}$ C预加热2分钟。然后,将DNA-阻断剂混合物的整个内容物转移至探针混合物,从而使杂交反应在65 $^{\circ}$ C进行16-24小时。

[0134] 杂交后,通过用200 μ l结合缓冲液(10mM Tris-HCl pH 8.0、0.5mM EDTA、1MNaCl)清洗四次来调整25 μ l的抗生蛋白链菌素-缀合的DynaBeadsTM(Thermo Fisher Scientific,65602)。在25 $^{\circ}$ C,在热混合器中以600RPM实施30分钟的DNA捕获。为了除去通过非特异性结合沉淀的非靶标DNA,首先用500 μ l清洗缓冲液1(0.15M氯化钠,0.015M柠檬酸

钠,0.1% SDS) 在室温下清洗所述珠1次,然后用清洗缓冲液2(0.015M氯化钠,0.0015M柠檬酸钠,0.1% SDS) 在65℃清洗3次。然后,将所述珠在20 μ l的洗脱缓冲液(10mM Tris-HCl,pH 8.0) 中再悬浮并用作随后索引PCR步骤的模板。

[0135] 对于多路测序,将5 μ l索引引物(预混合的i5和i7,分别20 μ M) 加入含有20 μ l再悬浮的T1珠和25 μ l Kapa HiFi hotstart ready混合物(Kapa Biosystem,KK2602) 的50 μ l反应中。PCR程序如下所示:(i) 95℃,5分钟;(ii) 98℃,20秒,60℃,30秒,72℃,1分钟,循环12次和(iii) 72℃,10分钟。将纯化的DNA文库在20 μ l EB中洗脱并通过Qubit dsDNA HS测定试剂盒定量。

[0136] 根据生产商的说明,使用PE150循环运行,在Illumina HiSeq 2500或NovaSeq上对索引PCR步骤产物测序。通过分析性管道(analytical pipeline) 分离(de-multiplexed) FASTQ序列,并分析常规文库质量指标。下表3A和3B显示了说明性文库生物信息学QC总结表。

[0137] 表3A:

样品名称	样本	输入 DNA (ng)	总 PF 读序	定位比	插入尺寸
2ng-12878-1	NA12878 基因组 DNA	2	1.42E+07	97.34%	184
2ng-12878-2	NA12878 基因组 DNA	2	1.39E+07	97.69%	184
5ng-12878-1	NA12878 基因组 DNA	5	1.36E+07	97.89%	183
5ng-12878-2	NA12878 基因组 DNA	5	1.36E+07	97.70%	184
10ng-12878-1	NA12878 基因组 DNA	10	1.35E+07	97.87%	179
10ng-12878-2	NA12878 基因组 DNA	10	1.35E+07	98.15%	186
30ng-12878-1	NA12878 基因组 DNA	30	1.37E+07	98.24%	194
30ng-12878-2	NA12878 基因组 DNA	30	1.37E+07	98.14%	193
10ng-PLA-1	cfDNA	10	1.56E+07	98.45%	163
10ng-PLA-2	cfDNA	10	1.54E+07	98.50%	163

[0139] 表3B:

样品名称	命中%	去重的中值覆盖度	预去重的中值覆盖度	覆盖均一度 (>0.2×平均值)
2ng-12878-1	79.02%	291	984	95.60%
2ng-12878-2	80.32%	300	985	95.60%
5ng-12878-1	79.59%	472	989	96.20%
5ng-12878-2	80.25%	475	987	96.30%
10ng-12878-1	80.94%	603	992	95.80%
10ng-12878-2	80.77%	600	991	96.40%
30ng-12878-1	80.25%	750	991	96.70%
30ng-12878-2	80.13%	745	989	96.70%
10ng-PLA-1	82.81%	620	991	93.30%
10ng-PLA-2	82.98%	634	990	93.40%

[0141] 实施例3:

[0142] 具有高甲基化水平的SW48基因组DNA购自ATCC(ATCC,CCL231)。通过Qubit dsDNA HS测定试剂盒(Thermo Fisher Scientific,Q32851) 测量浓度。通过REPLI-g微型试剂盒

(Qiagen 150023),在50 μ l下列标准规程(包括在30 $^{\circ}$ C培育16小时)中全基因组扩增(WGA) 10ng SW48基因组DNA。通过100 μ l Ampure XP珠(Beckman Coulter,A63881)纯化扩增材料并在50 μ l IDTE缓冲液(IDT,11-05-01-09)中洗脱。最终的WGA DNA得率为约3 μ g,其甲基化水平为原始SW48的约1/300。以0%、20%、50%、80%和100%的水平,将WGA DNA按比例与原始SW48基因组DNA混合以模拟基因组范围的甲基化水平梯度。使用聚焦声学超声波剪切仪(Covaris,M220),将50ng的每种DNA混合物剪切成约100-600bp的片段。超声处理参数设置如下:峰值入射功率50W,占空比20%,循环破碎系数(cycle per burst) 200,持续时间150秒,温度6-8 $^{\circ}$ C。通过LabChip GXII touch 24(Perkin Elmer)确认所剪切的DNA片段的尺寸。

[0143] 使用根据EZ-96DNA甲基化-lightningTM MagPrep(Zymo,D5047)的修改规程进行亚硫酸氢盐转化步骤(BC)。将97.5 μ l Lightning转化试剂和15 μ l,40ng所剪切的基因组DNA混合物加入至48-孔板(Thermo Fisher Scientific,AB0648)中。通过上下吸取来混合样品并将其以下列条件在热循环仪中培育:(i)98 $^{\circ}$ C,8分钟;(ii)54 $^{\circ}$ C,60分钟;(iii)4 $^{\circ}$ C储存多至20小时。将BC-处理的DNA样品转移至对于每个孔预加载了450 μ l M-结合缓冲液和7.5 μ l MagBinding珠的96-孔midi-板(Thermo Scientific,AB0859)。将组分彻底混合,并使所述板在室温下静置5分钟。然后,将所述板转移到磁力支架上保持另外5分钟,并除去上清液。用300 μ l M-清洗缓冲液清洗所述珠,并将所述珠与150 μ l L-脱磷酸缓冲液在室温下(20-30 $^{\circ}$ C)培育25分钟。将所述板置于磁力支架3分钟,弃去上清液,然后用300 μ l M-清洗缓冲液清洗所述珠两次。在清洗步骤之后,将所述板转移到55 $^{\circ}$ C的金属加热器(Illumina,SC-60-504,BD-60-601)上并保持30分钟以干燥所述珠,然后加入16 μ l M-洗脱缓冲液,并在55 $^{\circ}$ C培育另外4min。然后,将所述板移至磁力支架并保持1分钟,并且作为用于后续文库制备步骤的模板回收上清液。

[0144] 如实施例1制备MDA1和MDA2接头。形成MDA1、MDA2的寡核苷酸和表示为“锚定引物”的扩增引物的序列如以上表1所示。

[0145] 通过将12.5 μ l DNA样品、1.5 μ l 10 $\times$ CutSmart缓冲液(NEB,B7204S)、1 μ l虾碱性磷酸酶(NEB,M0371L)混合并在37 $^{\circ}$ C培育30分钟来对10ng的每种亚硫酸氢盐转化的DNA片段进行末端-修复。通过在95 $^{\circ}$ C培育5min并在冰上快速冷却,使所述产物进一步变性。

[0146] 如实施例1所述,进行第一连接、后续扩增、第二连接和PCR富集。将15 μ l纯化的DNA文库(50-200ng/ μ l)与4 μ l阻断剂混合物良好混合,并在热循环仪中按以下条件培育:(i)95 $^{\circ}$ C,5分钟;(ii)65 $^{\circ}$ C保持。同时,将10 μ l杂交缓冲液(13 $\times$ SSPE;13.5mM EDTA;13 $\times$ Denhart溶液;0.45% SDS)、0.5 μ l RNA酶-抑制剂和0.5 μ l Agilent SureSelect定制组探针池(Custom Panel Probe Pool)在65 $^{\circ}$ C预加热2分钟。然后,将DNA-阻断剂混合物的整个内容物转移至探针混合物,从而使杂交反应在65 $^{\circ}$ C进行16-24小时。

[0147] 图6A显示了示例性的毛细管电泳分析图,其显示了在PCR富集之后预捕获文库片段的尺寸分布。自上而下的曲线对应于图例中自下而上所指明的样品。预期峰尺寸为200-400bp。将所有文库加载至HT DNA高灵敏度LabChip试剂盒(Perkin Elmer)。所有预捕获的文库具有非常类似的得率和插入尺寸,表示所述文库制备方法对于甲基化状态无偏差。

[0148] 如实施例1所述,使用抗生蛋白链菌素-缀合的DynaBeadsTM捕获DNA,洗脱并使用索引引物扩增。图6B显示了示例性的毛细管电泳分析图,其显示了索引PCR之后的捕获后文

库片段的尺寸分布。将所有文库加载至HT DNA高灵敏度LabChip试剂盒(Perkin Elmer)。文库得率随着原始甲基化水平的提高而逐渐降低,其表明了在这些条件下文库制备程序的一般性GC偏差。

[0149] 根据生产商的说明,使用PE150循环运行,在Illumina HiSeq 2500上对索引PCR步骤产物测序。通过分析性管道(analytical pipeline)分离(de-multiplexed)FASTQ序列,并分析常规文库质量指标。下表4A和4B显示了说明性文库生物信息学QC总结表。

[0150] 表4A:

样品名称	%SW48 DNA	%WGA DNA	输入 DNA (ng)	PF 读序	定位比
SW48-1	100	0	10	8.26E+06	99.2%
SW48-2	80	20	10	8.96E+06	99.0%
SW48-3	50	50	10	8.04E+06	98.7%
SW48-4	20	80	10	7.61E+06	97.6%
SW48-5	0	100	10	6.88E+06	97.5%

[0152] 表4B:

样品名称	覆盖的复杂度	命中%	去重的中值覆盖度	预去重的中值覆盖度	均一度 (0.2×平均值)
SW48-1	62%	68.5%	324	502	0.97
SW48-2	65%	64.9%	348	510	0.974
SW48-3	68%	61.1%	288	408	0.97
SW48-4	80%	34.2%	160	194	0.971
SW48-5	81%	33.7%	140	168	0.953

[0154] 基于比对结果和碱基计数,计算每个目标CpG甲基化水平。图7显示了12,977个目标CpG位点的甲基化水平。在SW48-1样品中,这些位点具有>97%的甲基化水平(100%SW48, 0%WGA)。通过不同的WGA样品加标,这些位点的甲基化水平按比例降低并且处于预期范围内。这表明全文库制备和捕获方法可以准确且精确地测量CpG甲基化水平。

[0155] 实施例4:

[0156] NA12878基因组DNA和定制的5%突变基因组DNA参考得自Coriell Institute (Coriell Institute, NA12878)和Horizon Discovery (HD-C669)。通过Qubit dsDNA HS测定试剂盒(Thermo Fisher Scientific, Q32851)测量浓度。将HD-C669与NA12878以1:9的比例按比例混合以预期0.5%的突变等位基因频率(所得混合物命名为“PC1”)。表6A中列出了突变和它们的预期频率。将50ng纯NA12878和0.5%AF混合的DNA底物稀释至50μl IDTE缓冲液(IDT, 11-05-01-09),并使用聚焦声学超声波剪切仪(Covaris, M220)剪切成约100-600bp的片段。超声处理参数设置如下:峰值入射功率50W, 占空比20%, 循环破碎系数(cycle per burst)200, 持续时间150秒, 温度6-8°C。通过LabChip GXII touch 24(Perkin Elmer)确认所剪切的DNA片段的尺寸。通过Qubit dsDNA HS测定试剂盒定量剪切材料以获得作为文库制备输入的10ng。

[0157] 如果未提及,则以2至3个技术重复来实施所有实验。

[0158] 作为参考,使用典型的“Y”接头程序制备文库。将10ng所剪切的基因组DNA在50μl IDTE中的溶液加入至48-孔板(Thermo Fisher Scientific, AB0648)。使用标准KAPA

Hyper制备试剂盒 (KAPA Biosystem, KK8504) 末端修复和连接样品。在连接系统中以0.8 μ M的最终浓度使用图3(上图)中所述的“Y”接头。

[0159] 对于裂片(splinter)接头辅助的文库制备,将10ng剪切的基因组DNA在12.5 μ l IDTE中的溶液加入至48-孔板(Thermo Fisher Scientific, AB0648)中并通过将1.5 μ l 10 $\times$ CutSmart缓冲液(NEB, B7204S)和1 μ l 虾碱性磷酸酶(NEB, M0371L)混合进行末端-修复。将所述混合物在37 $^{\circ}$ C培育30分钟,然后加热至95 $^{\circ}$ C, 5min,接着在冰上快速冷却。如实施例1制备MDA1和MDA2接头。形成MDA1、MDA2的寡核苷酸和表示为“锚定引物”的扩增引物的序列如以上表1所示。如实施例1所述,进行第一连接、后续扩增、第二连接和PCR富集。

[0160] 在含有20 μ l DNA产物、1 $\times$ KAPA HiFi缓冲液、dNTP、1 μ M引物F和引物R以及1U/ μ l KAPA HiFi聚合酶的50 μ l反应中进行使用“Y”接头和裂片接头两者的连接产物的PCR富集。PCR程序如下所示: (i) 95 $^{\circ}$ C, 5分钟; (ii) 98 $^{\circ}$ C, 20秒, 60 $^{\circ}$ C, 30秒, 72 $^{\circ}$ C, 1分钟, 循环12次和 (iii) 72 $^{\circ}$ C, 10分钟。使用Agencourt AMPure XP珠(Beckman Coulter, A63881)纯化所述PCR产物并在18 μ l EB(10mM Tris-HCl, pH 8.0)中洗脱。

[0161] 图8A显示了示例性的毛细管电泳分析图,其显示了在PCR富集之后预捕获文库片段的示例性尺寸分布(上图和下图分别为ELSA-12878-pre和HS-12878-pre。“ELSA”表示裂片接头文库,“HS”表示“Y”接头文库)。预期峰尺寸为200-500bp。将所有文库加载至HT DNA高灵敏度LabChip试剂盒(Perkin Elmer)。

[0162] 将750ng纯化的DNA文库在15 μ l洗脱缓冲液中的溶液与4 μ l阻断剂混合物良好混合,并在热循环仪中按以下条件培育: (i) 95 $^{\circ}$ C, 5分钟; (ii) 65 $^{\circ}$ C保持。同时,将10 μ l杂交缓冲液(13 $\times$ SSPE; 13.5mM EDTA; 13 $\times$ Denhart溶液; 0.45% SDS)、0.5 μ l RNA酶-抑制剂和0.5 μ l Agilent SureSelect定制组探针池(Custom Panel Probe Pool)在65 $^{\circ}$ C预加热2分钟。然后,将DNA-阻断剂混合物的整个内容物转移至探针混合物,从而使杂交反应在65 $^{\circ}$ C进行16-24小时。

[0163] 杂交后,通过用200 μ l结合缓冲液(10mM Tris-HCl pH 8.0、0.5mM EDTA、1M NaCl)清洗四次来调整25 μ l的抗生蛋白链菌素-缀合的DynaBeadsTM (Thermo Fisher Scientific, 65602)。在25 $^{\circ}$ C,在热混合器中以600RPM实施30分钟的DNA捕获。为了除去通过非特异性结合沉淀的非靶标DNA,首先用500 μ l清洗缓冲液1(0.15M氯化钠,0.015M柠檬酸钠,0.1% SDS)在室温下清洗所述珠1次,然后用清洗缓冲液2(0.015M氯化钠,0.0015M柠檬酸钠,0.1% SDS)在65 $^{\circ}$ C清洗3次。然后,将所述珠在20 μ l的洗脱缓冲液(10mM Tris-HCl, pH 8.0)中再悬浮并用作随后索引PCR步骤的模板。

[0164] 对于多路测序,将5 μ l索引引物(预混合的i5和i7,分别20 μ M)加入含有20 μ l再悬浮的T1珠和25 μ l Kapa HiFi hotstart ready混合物(Kapa Biosystem, KK2602)的50 μ l反应中。PCR程序如下所示: (i) 95 $^{\circ}$ C, 5分钟; (ii) 98 $^{\circ}$ C, 20秒, 60 $^{\circ}$ C, 30秒, 72 $^{\circ}$ C, 1分钟, 循环14次和 (iii) 72 $^{\circ}$ C, 10分钟。将纯化的DNA文库在20 μ l EB中洗脱并通过Qubit dsDNA HS测定试剂盒定量。图8B显示了示例性的毛细管电泳分析图,其显示了索引PCR之后的捕获文库片段的示例性尺寸分布(上图和下图分别为ELSA-12878-post和HS-12878-post)。

[0165] 根据生产商的说明,使用PE150循环运行,在Illumina NextSeq上对索引PCR步骤产物测序。通过分析性管道(analytical pipeline)分离(de-multiplexed)FASTQ序列,并分析常规文库质量指标。表5A-D显示了通过Picard HSMetrics所产生的说明性文库生物信

息学QC总结表(“PC1”表示0.5%AF DNA混合物,“12878”表示NA12878基因组DNA)。

[0166] 表5A

样品	诱饵区域 (Bait Territory)	PF 读序	PF 唯一 读序	PCT PF UQ 读序	比对的 PF UQ 碱基	上饵碱基 (On Bait Bases)
ELSA-128 78-1	52,552	10,904,818	6,476,017	0.594	665,757,553	307,386,965
ELSA-128 78-2	52,552	10,769,038	6,107,990	0.567	626,201,560	305,050,477
ELSA-PC 1-1	52,552	10,918,648	6,254,827	0.573	635,301,234	328,222,731
ELSA-PC 1-2	52,552	10,494,670	6,226,391	0.593	634,757,074	316,119,222
HS-12878 -1	52,552	10,184,874	3,285,943	0.323	345,044,568	74,843,333
HS-12878 -2	52,552	10,034,950	3,258,049	0.325	341,880,314	75,197,794
HS-PC1-1	52,552	10,293,830	3,389,731	0.329	355,347,808	90,862,657
样品	诱饵区域	PF 读序	PF 唯一 读序	PCT PF UQ 读序	比对的 PF UQ 碱基	上饵碱基
HS-PC1-2	52,552	9,526,184	2,976,248	0.312	311,924,121	70,668,683

[0168] 表5B

样品	近 饵 碱 基 (Near Bait Bases)	脱 饵 碱 基 (Off Bait Bases)	命中碱 基	PCT 选择 的碱基	PCT 脱饵	上饵的 vs 所选择的
ELSA-128 78-1	73,107,540	285,263,048	307,386,965	0.572	0.428	0.808
ELSA-128 78-2	69,640,725	251,510,358	305,050,477	0.598	0.402	0.814
ELSA-PC 1-1	66,460,387	240,618,116	328,222,731	0.621	0.379	0.832
ELSA-PC 1-2	66,861,856	251,775,996	316,119,222	0.603	0.397	0.825
HS-12878 -1	25,087,954	245,113,281	74,843,333	0.29	0.71	0.749
HS-12878 -2	24,939,238	241,743,282	75,197,794	0.293	0.707	0.751
HS-PC1-1	26,981,562	237,503,589	90,862,657	0.332	0.668	0.771
HS-PC1-2	21,796,096	219,459,342	70,668,683	0.296	0.704	0.764

[0170] 表5C

[0171]

样品	平均诱饵 覆盖度	平均目标 覆盖度	PCT 可用 的上饵碱基	PCT 可用 的命中碱基	富集倍 数	零覆盖目 标 PCT
ELSA-128 78-1	5,849	5,849	0.255	0.255	27,252	0
ELSA-128 78-2	5,805	5,805	0.257	0.257	28,753	0
ELSA-PC 1-1	6,246	6,246	0.274	0.274	30,494	0
ELSA-PC 1-2	6,015	6,015	0.274	0.274	29,395	0
HS-12878 -1	1,424	1,424	0.067	0.067	12,803	0
HS-12878 -2	1,431	1,431	0.068	0.068	12,982	0
HS-PC1-1	1,729	1,729	0.08	0.08	15,092	0
HS-PC1-2	1,345	1,345	0.067	0.067	13,372	0

[0172] 表5D

[0173]

样品	80 倍碱基 罚分 (Fold 80 Base Penalty)	Hs 文库大 小	Hs 罚分 50×	Hs 罚分 100×	At 脱落 (Dropo ut)	GC 脱落 (Dropo ut)
ELSA-128 78-1	1.32	2,134,279	2.88	2.89	1.03	6.07
ELSA-128 78-2	1.32	2,066,386	2.72	2.73	1.06	5.6
ELSA-PC 1-1	1.33	2,227,506	2.59	2.6	1.19	5.27
ELSA-PC 1-2	1.33	2,191,344	2.69	2.7	1.01	5.62
HS-12878 -1	1.09	452,276	5.11	5.22	1.25	0.73
HS-12878 -2	1.09	453,694	5.06	5.16	1.29	0.71
HS-PC1-1	1.1	536,676	4.35	4.43	1.46	0.62
HS-PC1-2	1.09	419,039	4.93	5.03	1.64	0.57

[0174] 分析序列以鉴别突变。表6A-C中列出了所识别的体细胞突变,其比较了裂片和“Y”接头文库之间的表现。裂片接头文库在0.5%AF PC1中具有更好的突变检测灵敏度,但是在NA12878中具有几个推定的假阳性识别。

[0175] 表6A

[0176]

突变	位置	预期的 PC1
ALK:p.F1174L	2:29443695,G>T	0.50%
BRAF:p.V600E	7:140453136,A>T	0.50%
EGFR:p.E746_A750del	7:55242464,AGGAATTAAG AGAAGC (SEQ ID NO : 21) >A	0.50%
EGFR:p.T790M	7:55249071,C>T	0.50%
KRAS:p.G12A	12:25398284,C>G	0.50%
MET:c.3028+1G>T	7:116412044,G>T	0.50%
NRAS:p.Q61H	1:115256528,T>A	0.50%
PIK3CA:p.E545K	3:178936091,G>A	0.50%
EGFR:p.G719S	7:55241707,G>A	1.00%
KRAS:p.G13D	12:25398281,C>T	2.00%
PIK3CA:p.H1047R	3:178952085,A>G	2.00%
KRAS:p.G12S	12:25398285,C>T	
MET:c.3028+1G>A	7:116412044,G>A	
MET:p.D1010Y	7:116412043,G>T	
MET:p.L238fs	7:116339847,GT>G	
RET:c.2136+14C>T	10:43610198,C>T	
突变_计数	-	

[0177] 表6B

[0178]

突变	ELSA-12 878-1	ELSA-12 878-2	ELSA-P C1-1	ELSA-P C1-2
ALK:p.F1174L			0.27%	0.35%
BRAF:p.V600E			0.44%	0.49%
EGFR:p.E746_A750del			0.31%	0.29%
EGFR:p.T790M			0.55%	0.92%
KRAS:p.G12A			0.71%	0.29%
MET:c.3028+1G>T			1.49%	0.65%
NRAS:p.Q61H			0.52%	0.70%
PIK3CA:p.E545K	0.31%		0.27%	0.67%
EGFR:p.G719S			1.17%	0.84%
KRAS:p.G13D			2.14%	1.77%
PIK3CA:p.H1047R			2.08%	1.82%
KRAS:p.G12S		0.20%		
MET:c.3028+1G>A				0.17%
MET:p.D1010Y	0.11%			
MET:p.L238fs			1.76%	1.69%
RET:c.2136+14C>T				
突变_计数	2	1	12	13

[0179] 表6C

[0180]

突变	HS-1287 8-1	HS-1287 8-2	HS-PC1- 1	HS-PC1- 2
----	----------------	----------------	--------------	--------------

[0181]	ALK:p.F1174L			0.64%	0.64%
	BRAF:p.V600E			0.71%	0.55%
	EGFR:p.E746_A750del			0.17%	0.14%
	EGFR:p.T790M			1.12%	0.44%
	KRAS:p.G12A			1.26%	0.60%
	MET:c.3028+1G>T			2.38%	1.62%
	NRAS:p.Q61H				0.18%
	PIK3CA:p.E545K			0.29%	
	EGFR:p.G719S			0.63%	0.95%
	KRAS:p.G13D			1.94%	2.16%
	PIK3CA:p.H1047R			1.89%	3.12%
	KRAS:p.G12S	0.11%			
	MET:c.3028+1G>A				0.76%
	MET:p.D1010Y				
	MET:p.L238fs				
	RET:c.2136+14C>T			0.78%	1.31%
	突变_计数	1	0	11	12

[0182] 实施例5:

[0183] λ DNA购自Promega (Madison, WI, 产品目录号: D1521)。通过Qubit dsDNA HS测定试剂盒 (Thermo Fisher Scientific, Waltham, MA, Q31851) 测量浓度, 并且在文库制备中所使用的DNA的量在1-10ng的范围内。将DNA底物稀释至50 μ l IDTE缓冲液 (Integrated DNA Technologies, Coralville, IA; 11-05-01-09), 并使用聚焦声学超声波剪切仪 (Covaris, Woburn, MA, M220) 剪切成约100-600bp的片段。超声处理参数设置如下: 峰值入射功率50W, 占空比20%, 循环破碎系数 (cycle per burst) 200, 持续时间150秒, 温度6-8 $^{\circ}$ C。通过LabChip GXII touch 24 (Perkin Elmer, Waltham, MA) 确认所剪切的DNA片段的尺寸。

[0184] 使用根据EZ-96DNA甲基化-lightningTM MagPrep (Zymo, Irvine, CA, D5047) 的修改规程进行亚硫酸氢盐转化步骤 (BC)。将97.5 μ l Lightning转化试剂和15 μ l所剪切的基因组DNA加入至48-孔板 (Thermo Fisher Scientific, AB0648) 中。通过上下吸取来混合样品并将其以下列条件在热循环仪中培育: (i) 98 $^{\circ}$ C, 8分钟; (ii) 54 $^{\circ}$ C, 60分钟; (iii) 4 $^{\circ}$ C储存多至20小时。将BC-处理的DNA样品转移至对于每个孔预加载了450 μ l M-结合缓冲液和7.5 μ l MagBinding珠的96-孔midi-板 (Thermo Scientific, AB0859)。将组分彻底混合, 并使所述板在室温下静置5分钟。然后, 将所述板转移到磁力支架上保持另外5分钟, 并除去上清液。用300 μ l M-清洗缓冲液清洗所述珠, 并将所述珠与150 μ l L-脱磺酸缓冲液在室温下 (20-30 $^{\circ}$ C) 培育25分钟。将所述板置于磁力支架3分钟, 弃去上清液, 然后用300 μ l M-清洗缓冲液清洗所述珠两次。在清洗步骤之后, 将所述板转移到55 $^{\circ}$ C的金属加热器 (Illumina, San Diego, CA, SC-60-504, BD-60-601) 上并保持30分钟以干燥所述珠, 然后加入16 μ l M-洗脱缓冲液, 并在55 $^{\circ}$ C培育另外4分钟。然后, 将所述板移至磁力支架并保持1分钟, 并且作为用于后续文库制备步骤的模板回收上清液。

[0185] 将接头MDA1设计为在下链上具有8个碱基的3'悬突和4个碱基的5'悬突。所述3'悬突具有以3:1的摩尔比随机合成的多个八个G或A。所述4个碱基的5'悬突在上链上产生了隐性3'端, 其防止由于上链3'端的不完全封端所造成的漏泄的TdT活性。在第一加尾和连接步骤期间, 所述3'悬突退火至单链DNA底物的3'端多聚-C/T尾 (如图9所示)。图10显示了形成

MDA1的寡核苷酸的序列。通过将寡聚ATN-R2-Top和ATN-R2-Bot退火在一起制备了MDA1接头。详细地,将50 μ l每种寡聚核苷酸(100 μ M)混合并在95 $^{\circ}$ C培育10分钟,并使其在含有0.1mM EDTA和50mM NaCl的10mM Tris-HCl中缓慢冷却至室温。通过磷酸基使两个寡聚核苷酸的3'端封端以防止自连接。

[0186] 将MDA2接头设计以具有多个7个N(以1:1:1:1的摩尔比随机合成的A、T、G或C)。它退火至单链DNA底物的3'端并且加快了第二连接步骤期间MDA2和DNA底物之间的连接(如图9所示)。通过将寡聚ATN-R1-Top和ATN-R1-Bot退火在一起制备了MDA2接头。图10显示了形成MDA2的寡核苷酸的序列。形成MDA1、MDA2的寡核苷酸和表示为“锚定引物”的扩增引物的序列如表7所示。

[0187] 表7

[0188]	寡聚核苷酸	序列	注释
	ATN-R2-Top	AGATCGGAAGAGCACACGTCTGAACTCCAGTCA C (SEQ ID NO: 4)	5'磷酸; 3'磷酸
	ATN-R2-Bot	AGTCGTGACTGGAGTTCAGACGTGTGCTCTTCC GATCTRRRRRRRR (SEQ ID NO: 22)	3'磷酸; R (G:A) =3:1 的预混合物
	ATN-R1-Top	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT (SEQ ID NO: 6)	5'磷酸; 3'磷酸
	ATN-R1-Bot	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNNN (SEQ ID NO: 23)	3'磷酸
	LAP (锚定引物)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC (SEQ ID NO: 16)	

[0189] 通过将12.5 μ l DNA样品、1.5 μ l 10 $\times$ CutSmart缓冲液(NEB,B7204S)、1 μ l虾碱性磷酸酶(New England Biolabs(NEB),Ipswich,MA,M0371L)混合并在37 $^{\circ}$ C培育30分钟来对亚硫酸氢盐转化的DNA片段进行末端-修复。通过在95 $^{\circ}$ C培育5分钟并在冰上快速冷却,使所述产物进一步变性。

[0190] 然后,在含有预处理的DNA底物、1 $\times$ CutSmart缓冲液、0.25mM CoCl₂(NEB,B0252S)、0.025mM β -烟酰胺腺嘌呤二核苷酸(NEB,B9007S)、0.09mM dCTP(Roche,11934520001,由Sigma-Aldrich,St.Louis,MO出售)、0.01mM dTTP(Roche,11934546001)、1 μ M MDA1接头、0.5U/ μ l大肠杆菌(E.coli)连接酶(NEB,M0205L)和0.5U/ μ l末端脱氧核苷酸转移酶(TdT;NEB,M0315S)的20 μ l反应体积中进行第一连接反应。将所述反应在37 $^{\circ}$ C培育30分钟,然后在95 $^{\circ}$ C加热2分钟并保持在4 $^{\circ}$ C。

[0191] 在存在1 $\times$ KAPA HiFi HotStart Uracil+ReadyMix(KAPA,Biosystems,Wilmington,MA,KK2802)和0.91 μ M锚定引物的情况下使所述连接产物延伸并线性扩增。按照以下热温度谱进行线性扩增反应:(i)95 $^{\circ}$ C,5分钟;(ii)98 $^{\circ}$ C,20秒,62 $^{\circ}$ C,30秒,72 $^{\circ}$ C,1分钟,循环15次和(iii)72 $^{\circ}$ C,5分钟。反应完成后,通过用2.5 $\times$ AMPure XP珠(Beckman Coulter,Brea,CA,A63881)纯化来更换缓冲液并用11.5 μ l洗脱缓冲液(EB)(10mM Tris-HCl,pH 8.0)洗脱。

[0192] 在含有10 μ l纯化的DNA产物、1 $\times$ T4 DNA连接酶缓冲液、10%PEG8000、1 μ MMDA1接头和20U/ μ l T4 DNA连接酶(NEB,M0202L)的20 μ l反应体积中进行第二连接反应。将所述反应

在20℃培育30分钟,然后在65℃加热20分钟并保持在4℃。

[0193] 在含有20μl上述DNA产物、1×KAPA HiFi缓冲液、dNTP、1μM引物F和引物R以及1U/μl KAPA HiFi聚合酶的50μl反应中进行连接产物的PCR富集。PCR程序如下所示:(i) 95℃,5分钟;(ii) 98℃,20秒,60℃,30秒,72℃,1分钟,循环8次和(iii) 72℃,10分钟。使用Agencourt AMPure XP珠(Beckman Coulter,A63881)纯化所述PCR产物并在18μl EB(10mM Tris-HCl,pH 8.0)中洗脱。

[0194] 对于多路测序,将5μl索引引物(预混合的i5和i7,分别20μM)加入含有1μl上述纯化的PCR产物和25μl Kapa HiFi hotstart ready混合物(Kapa Biosystem,KK2602)的50μl反应中。PCR程序如下所示:(i) 98℃,45秒;(ii) 98℃,15秒,60℃,30秒,72℃,1分钟,循环6次和(iii) 72℃,5分钟。将纯化的DNA文库在20μl EB中洗脱并通过Qubit dsDNA HS测定试剂盒定量。

[0195] 图11显示了示例性的毛细管电泳分析图,其显示了索引PCR之后的文库片段的大小分布。将所有文库加载至HT DNA高灵敏度LabChip试剂盒(Perkin Elmer)。

[0196] 根据生产商的说明,使用PE150循环运行,在Illumina Novaseq上对索引PCR步骤产物测序。通过分析性管道(analytical pipeline)分离(de-multiplexed)FASTQ序列,并分析常规文库质量指标。下表8显示了说明性文库生物信息学QC总结表。

[0197] 表8A

[0198]	样品名称	样本	输入DNA (ng)	总PF读序	定位比	插入尺寸
	1ng-λ	λ基因组DNA	1	1575300	0.988	157
	2ng-λ	λ基因组DNA	2	1262550	0.989	158
	5ng-λ	λ基因组DNA	5	1276862	0.991	161
	10ng-λ	λ基因组DNA	10	1448128	0.992	168

[0199] 表8B

[0200]	样品名称	去重的中值覆盖度	预去重的中值覆盖度	80 倍碱基罚分
	1ng-λ	3505	4160	1.11
	2ng-λ	2904	3353	1.11
	5ng-λ	2965	3430	1.12
	10ng-λ	3377	3954	1.11

[0201] 图9提供了如上所述的文库制备方法的总览图示。通过适当的dNTP,使用TdT进行加尾步骤以在ssDNA片段的3'端产生均聚物或近-均聚物尾。所述均聚物退火至上链(top strand)中含有5'磷酸基的接头的3'悬突。通过连接酶催化的连接反应使ssDNA片段的3'端封闭以防止过度加尾。通过锚定引物竞争掉所述接头的下链(bottom strand),从而暴露出线性扩增过程的起始位点。将扩增的ssDNA链用作第二轮连接的底物,其中将裂片寡核苷酸用于产生dsDNA片段的短延伸,其允许使用通过T4 DNA连接酶的标准dsDNA连接的后续接头连接。

[0202] 根据上述内容将理解,尽管本文已出于说明的目的描述了本发明的具体实施方式,但是在不背离本发明的精神和范围的情况下可以做出多种改变。

[0203] 在本发明的整个描述期间,参考了多个专利申请和专利公开,它们中的每一个以

其全部内容作为参考并入本文。

[0001] 序列表

[0002] <110> 广州燃石医学检验所有限公司

[0003] <120> 用于制备核酸文库的组合物和方法

[0004] <130> 232396-228002

[0005] <140>

[0006] <141>

[0007] <150> PCT/CN2018/081748

[0008] <151> 2018-04-03

[0009] <160> 23

[0010] <170> PatentIn version 3.5

[0011] <210> 1

[0012] <211> 8

[0013] <212> DNA

[0014] <213> 人工序列

[0015] <220>

[0016] <223> 人工序列的描述:合成的寡核苷酸

[0017] <400> 1

[0018] ccctcctc 8

[0019] <210> 2

[0020] <211> 12

[0021] <212> DNA

[0022] <213> 人工序列

[0023] <220>

[0024] <223> 人工序列的描述:合成的寡核苷酸

[0025] <400> 2

[0026] tttttttttt tt 12

[0027] <210> 3

[0028] <211> 12

[0029] <212> DNA

[0030] <213> 人工序列

[0031] <220>

[0032] <223> 人工序列的描述:合成的寡核苷酸

[0033] <400> 3

[0034] aaaaaaaaaa aa 12

[0035] <210> 4

[0036] <211> 34

[0037] <212> DNA

[0038] <213> 人工序列

- [0039] <220>
[0040] <223> 人工序列的描述:合成的寡核苷酸
[0041] <400> 4
[0042] agatcggaag agcacacgtc tgaactccag tcac 34
[0043] <210> 5
[0044] <211> 42
[0045] <212> DNA
[0046] <213> 人工序列
[0047] <220>
[0048] <223> 人工序列的描述:合成的寡核苷酸
[0049] <400> 5
[0050] gtgactggag ttcagacgtg tgctcttccg atctrrrrrr rr 42
[0051] <210> 6
[0052] <211> 33
[0053] <212> DNA
[0054] <213> 人工序列
[0055] <220>
[0056] <223> 人工序列的描述:合成的寡核苷酸
[0057] <400> 6
[0058] agatcggaag agcgtcgtgt agggaaagag tgt 33
[0059] <210> 7
[0060] <211> 45
[0061] <212> DNA
[0062] <213> 人工序列
[0063] <220>
[0064] <223> 人工序列的描述:合成的寡核苷酸
[0065] <400> 7
[0066] acactctttc cctacacgac gctcttccga tctttttttt ttttt 45
[0067] <210> 8
[0068] <211> 33
[0069] <212> DNA
[0070] <213> 人工序列
[0071] <220>
[0072] <223> 人工序列的描述:合成的寡核苷酸
[0073] <400> 8
[0074] acactctttc cctacacgac gctcttccga tct 33
[0075] <210> 9
[0076] <211> 34
[0077] <212> DNA

- [0078] <213> 人工序列
- [0079] <220>
- [0080] <223> 人工序列的描述:合成的寡核苷酸
- [0081] <400> 9
- [0082] agatcggaag agcacacgtc tgaactccag tcac 34
- [0083] <210> 10
- [0084] <211> 34
- [0085] <212> DNA
- [0086] <213> 人工序列
- [0087] <220>
- [0088] <223> 人工序列的描述:合成的寡核苷酸
- [0089] <400> 10
- [0090] agatcggaag agcacacgtc tgaactccag tcac 34
- [0091] <210> 11
- [0092] <211> 33
- [0093] <212> DNA
- [0094] <213> 人工序列
- [0095] <220>
- [0096] <223> 人工序列的描述:合成的寡核苷酸
- [0097] <400> 11
- [0098] acactctttc cctacacgac gctcttccga tct 33
- [0099] <210> 12
- [0100] <211> 45
- [0101] <212> DNA
- [0102] <213> 人工序列
- [0103] <220>
- [0104] <223> 人工序列的描述:合成的寡核苷酸
- [0105] <400> 12
- [0106] acactctttc cctacacgac gctcttccga tctttttttt ttttt 45
- [0107] <210> 13
- [0108] <211> 42
- [0109] <212> DNA
- [0110] <213> 人工序列
- [0111] <220>
- [0112] <223> 人工序列的描述:合成的寡核苷酸
- [0113] <400> 13
- [0114] ccctcctcag atcggaagag cacacgtctg aactccagtc ac 42
- [0115] <210> 14
- [0116] <211> 45

[0117] <212> DNA
[0118] <213> 人工序列
[0119] <220>
[0120] <223> 人工序列的描述:合成的寡核苷酸
[0121] <400> 14
[0122] aaaaaaaaaa aaagatcgga agagcgtcgt gtagggaaag agtgt 45
[0123] <210> 15
[0124] <211> 42
[0125] <212> DNA
[0126] <213> 人工序列
[0127] <220>
[0128] <223> 人工序列的描述:合成的寡核苷酸
[0129] <400> 15
[0130] gtgactggag ttcagacgtg tgctcttccg atctrrrrrr rr 42
[0131] <210> 16
[0132] <211> 33
[0133] <212> DNA
[0134] <213> 人工序列
[0135] <220>
[0136] <223> 人工序列的描述:合成的引物
[0137] <400> 16
[0138] gtgactggag ttcagacgtg tgctcttccg atc 33
[0139] <210> 17
[0140] <211> 33
[0141] <212> DNA
[0142] <213> 人工序列
[0143] <220>
[0144] <223> 人工序列的描述:合成的引物
[0145] <400> 17
[0146] acactctttc cctacacgac gctcttccga tct 33
[0147] <210> 18
[0148] <211> 33
[0149] <212> DNA
[0150] <213> 人工序列
[0151] <220>
[0152] <223> 人工序列的描述:合成的引物
[0153] <400> 18
[0154] gtgactggag ttcagacgtg tgctcttccg atc 33
[0155] <210> 19

- [0156] <211> 58
[0157] <212> DNA
[0158] <213> 人工序列
[0159] <220>
[0160] <223> 人工序列的描述:合成的引物
[0161] <400> 19
[0162] aatgatacgg cgaccaccga gatctacacg ttagttcaca ctctttccct acacgacg 58
[0163] <210> 20
[0164] <211> 52
[0165] <212> DNA
[0166] <213> 人工序列
[0167] <220>
[0168] <223> 人工序列的描述:合成的引物
[0169] <400> 20
[0170] caagcagaag acggcatacg agatgtgatg ccgtgactgg agttcagacg tg 52
[0171] <210> 21
[0172] <211> 16
[0173] <212> DNA
[0174] <213> 未知
[0175] <220>
[0176] <223> 未知的描述:EGFR 序列
[0177] <400> 21
[0178] aggaattaag agaagc 16
[0179] <210> 22
[0180] <211> 46
[0181] <212> DNA
[0182] <213> 人工序列
[0183] <220>
[0184] <223> 人工序列的描述:合成的寡核苷酸
[0185] <400> 22
[0186] agtcgtgact ggagttcaga cgtgtgctct tccgatctrr rrrrrr 46
[0187] <210> 23
[0188] <211> 40
[0189] <212> DNA
[0190] <213> 人工序列
[0191] <220>
[0192] <223> 人工序列的描述:合成的寡核苷酸
[0193] <220>
[0194] <221> 修改的_碱基

-
- [0195] <222> (34) .. (40)
[0196] <223> a, c, t, or g
[0197] <400> 23
[0198] acactctttc cctacacgac gctcttccga tctnnnnnnn 40

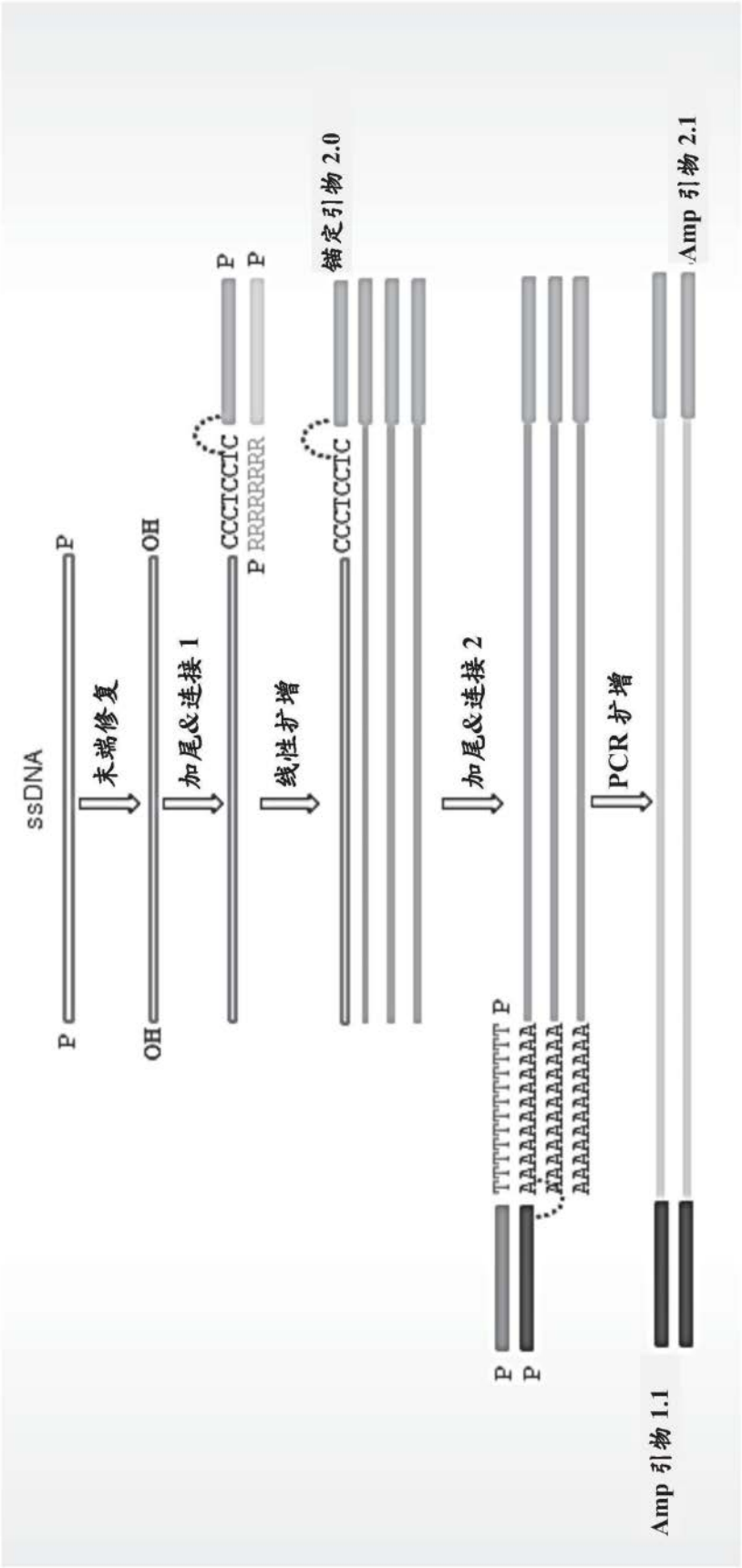


图1

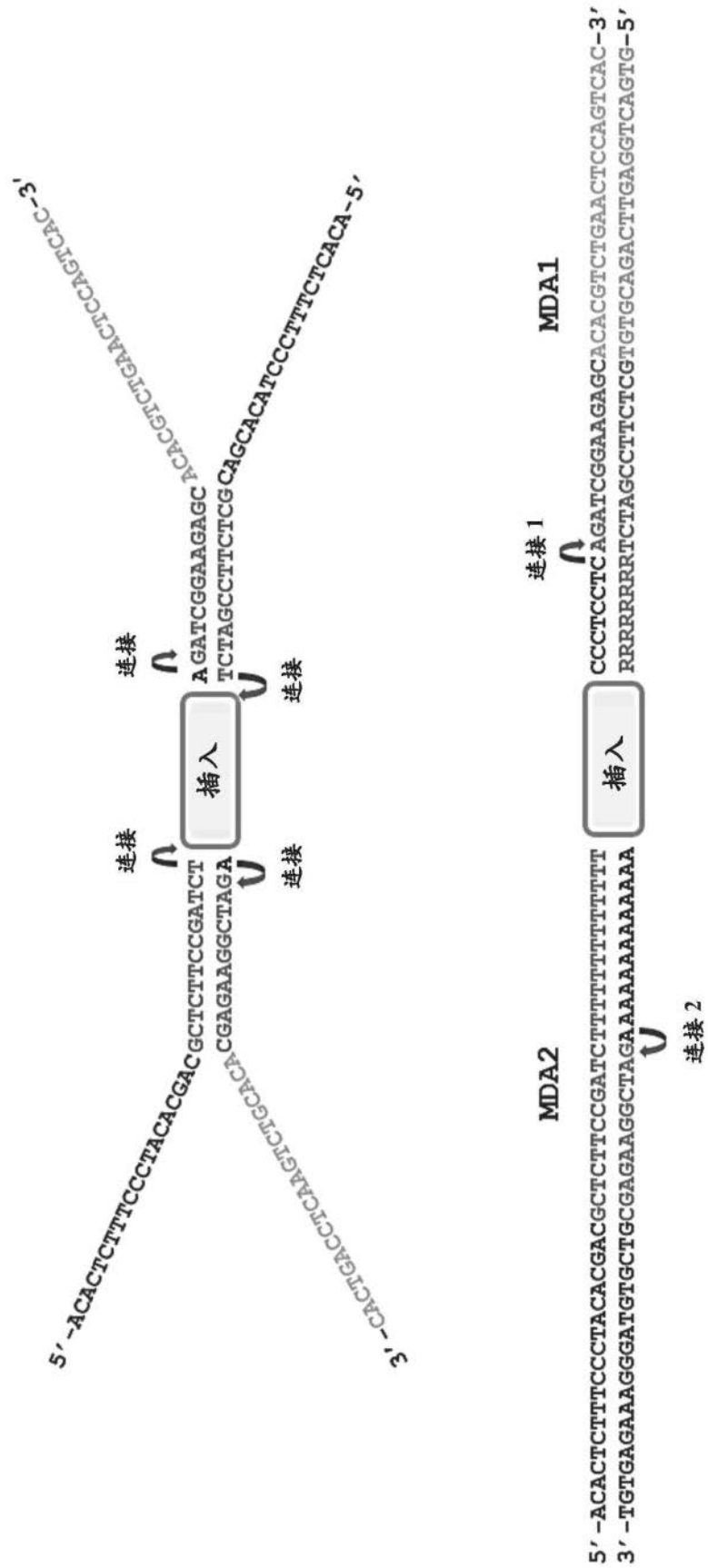


图3

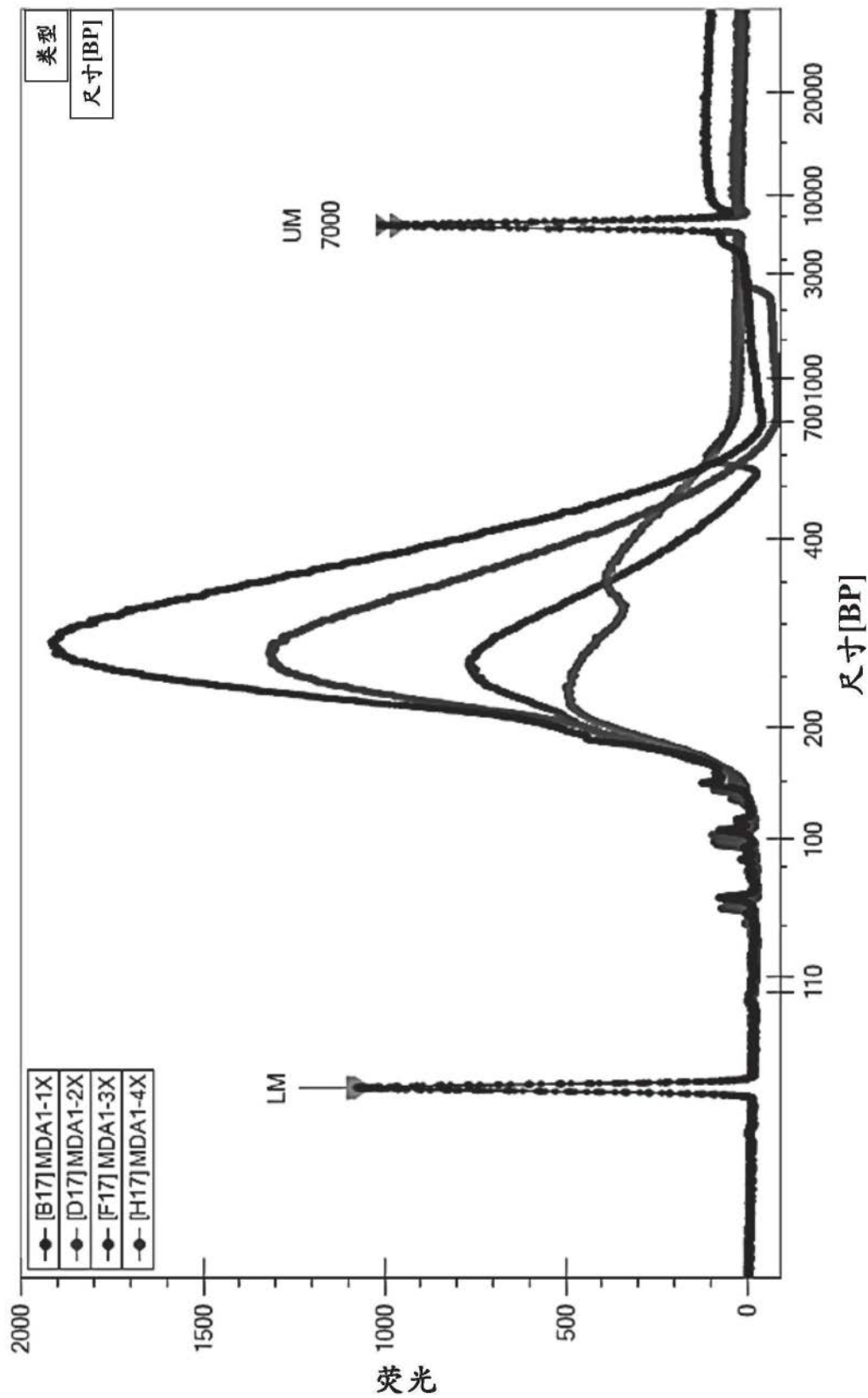


图4

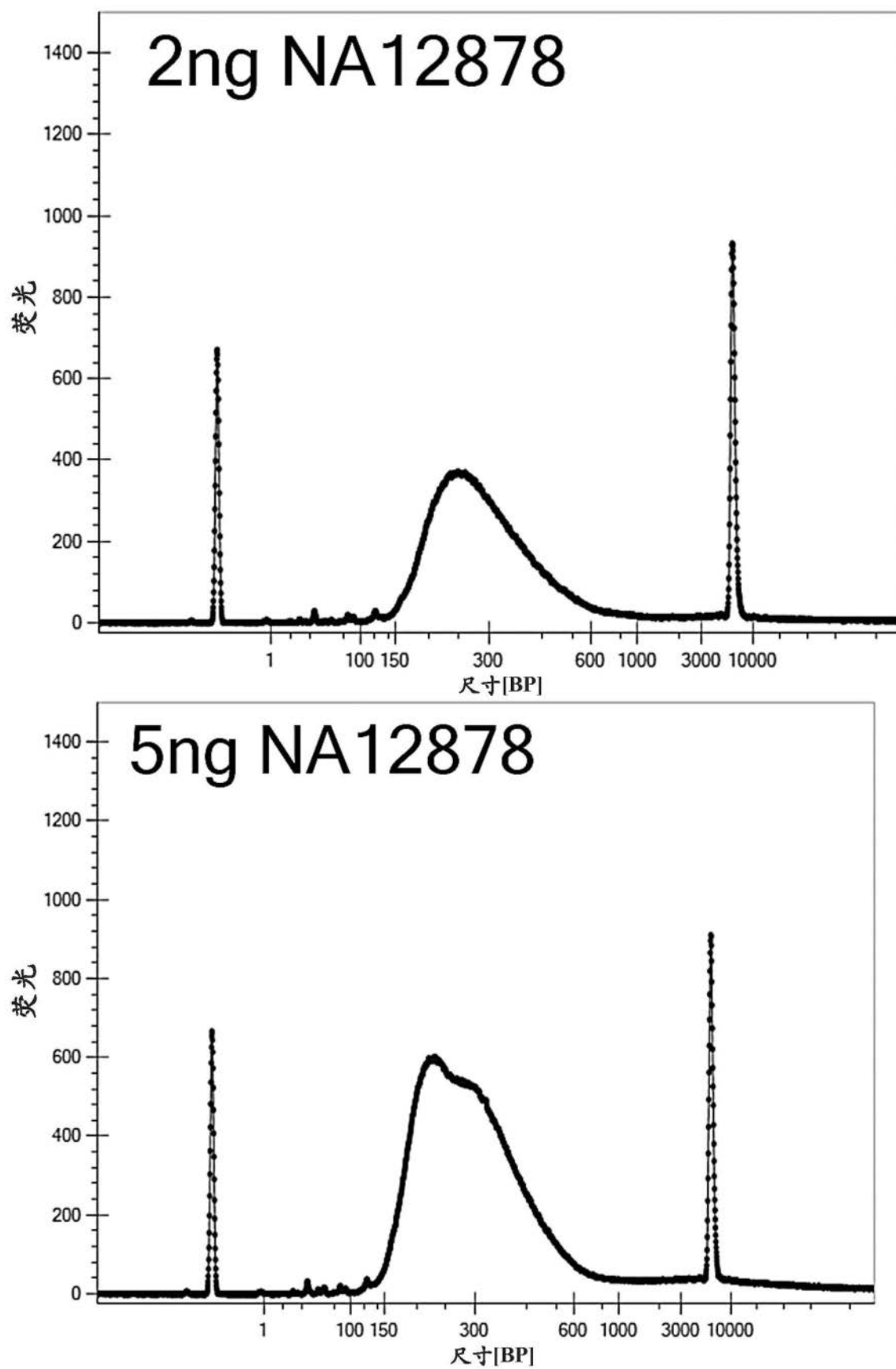


图5A

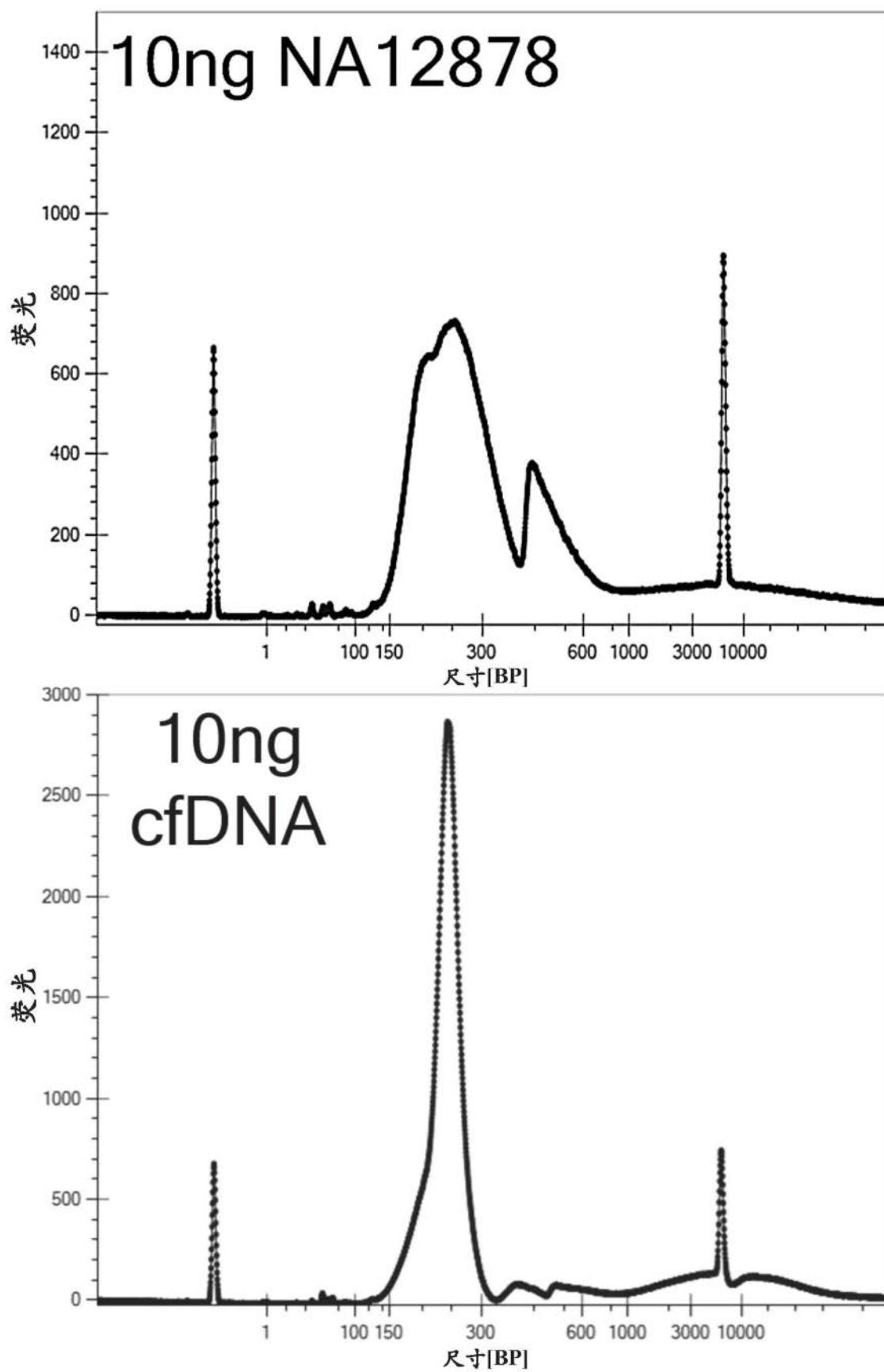


图5B

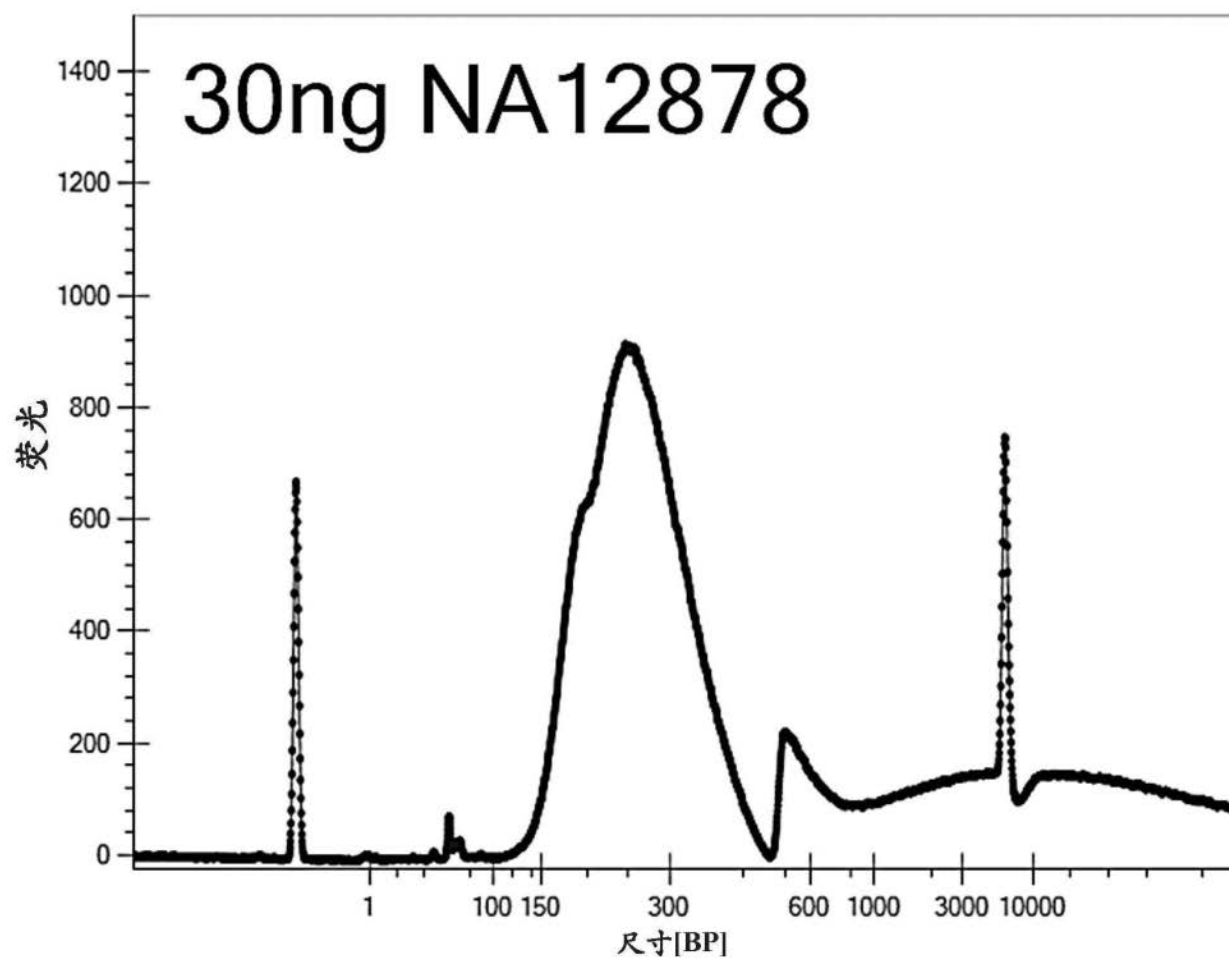


图5C

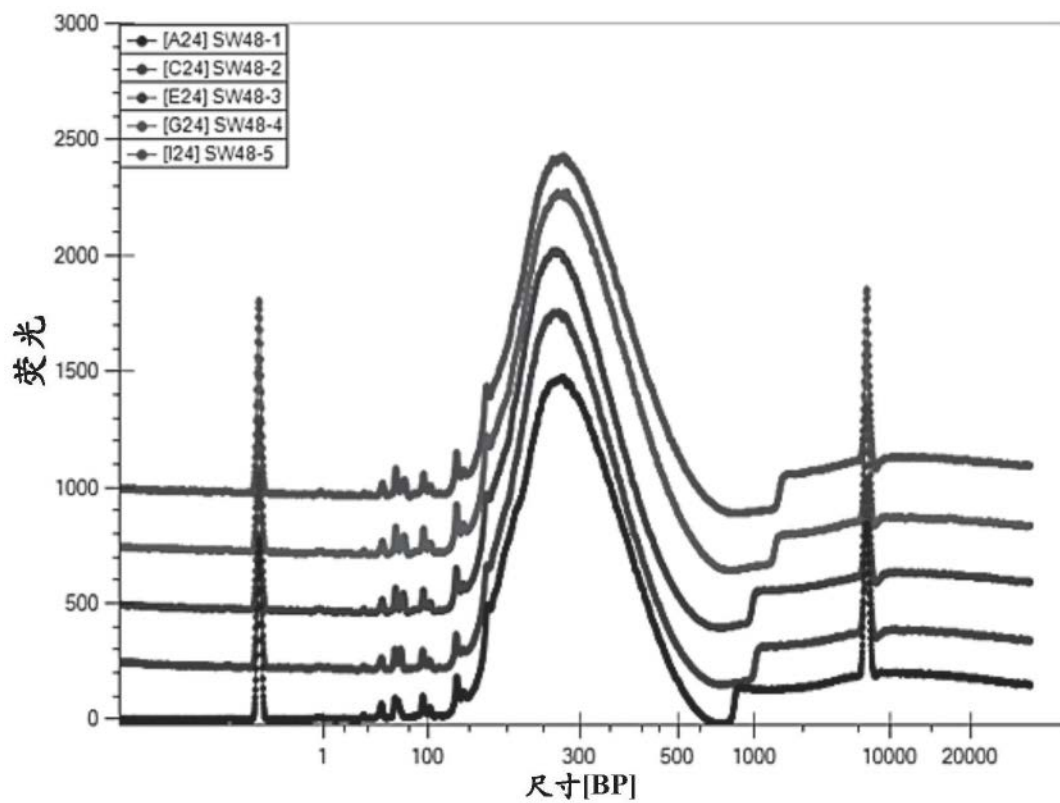


图6A

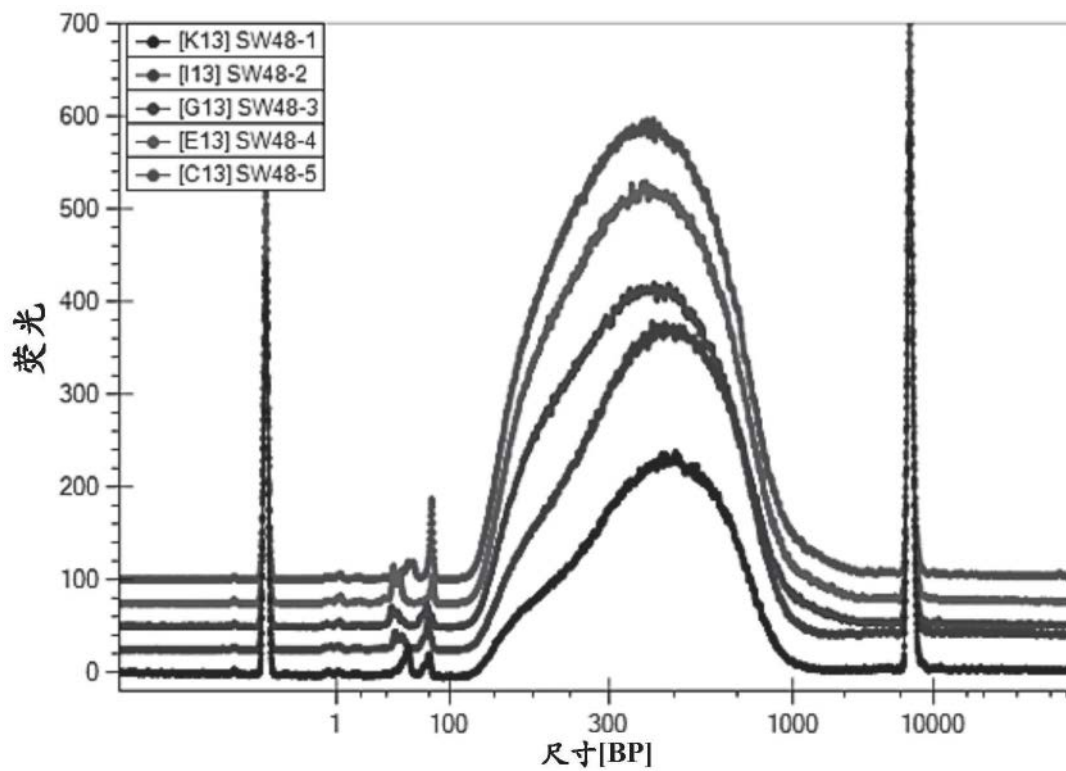


图6B

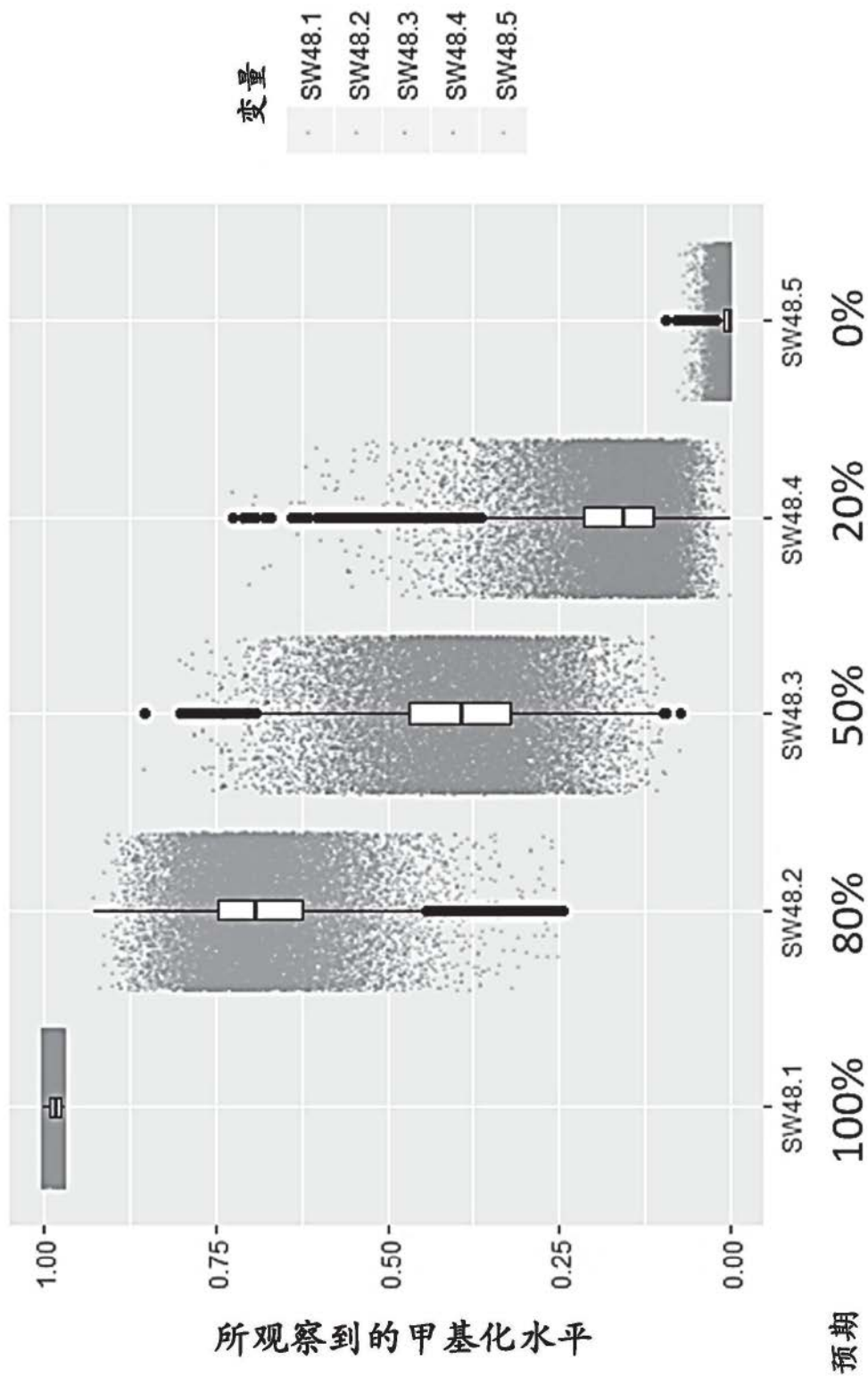


图7

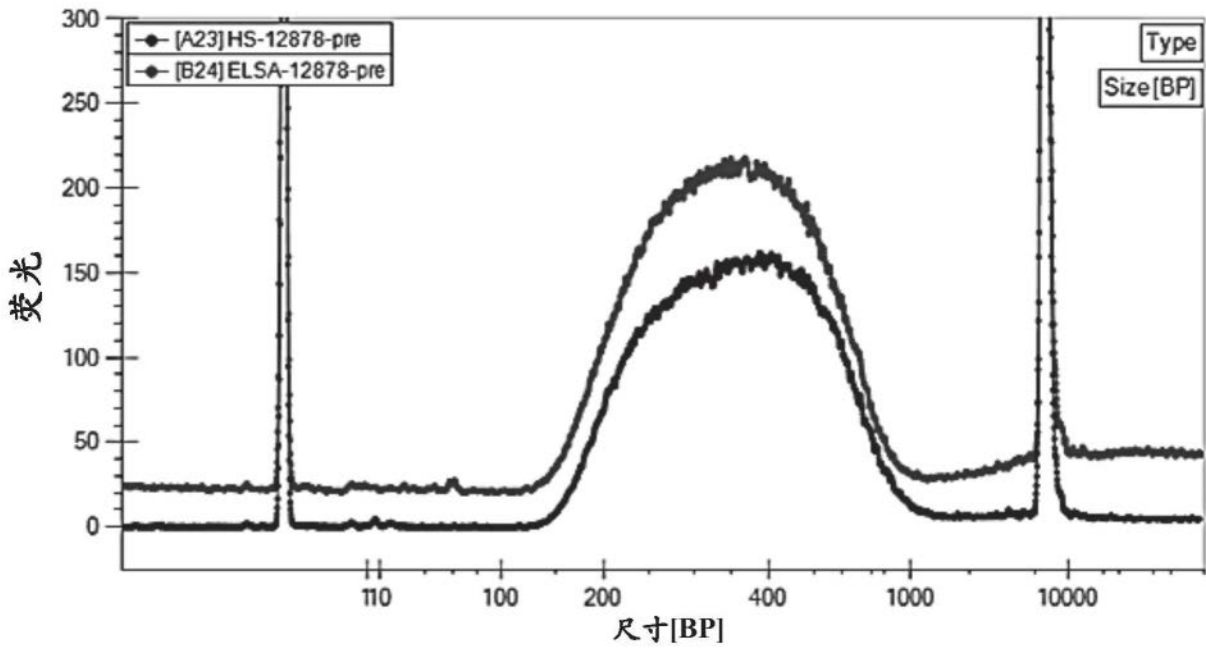


图8A

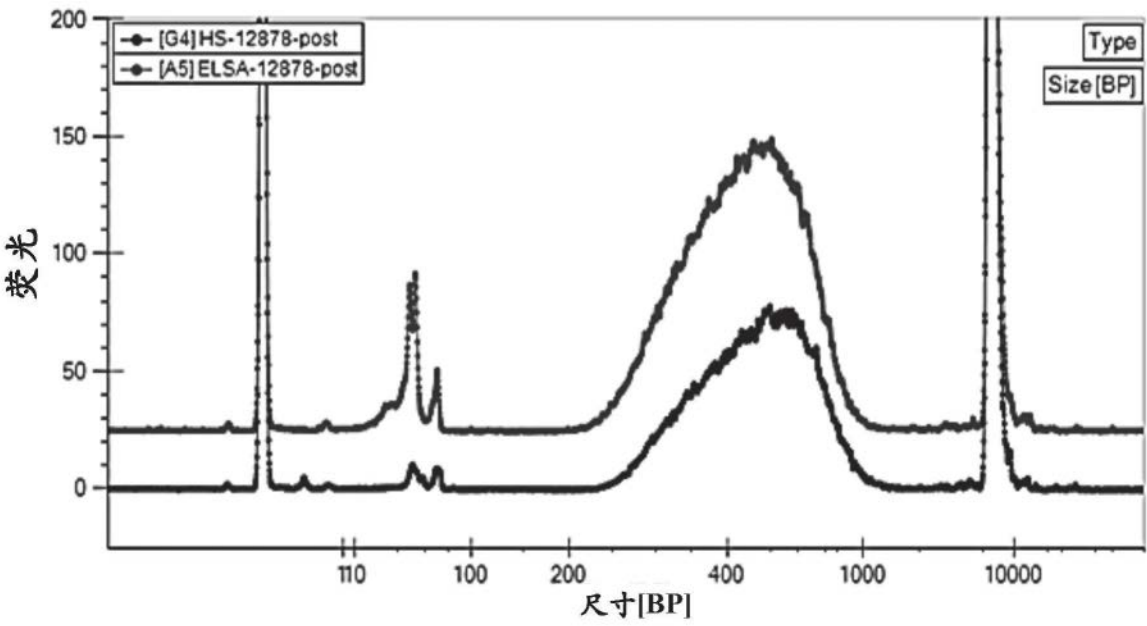


图8B

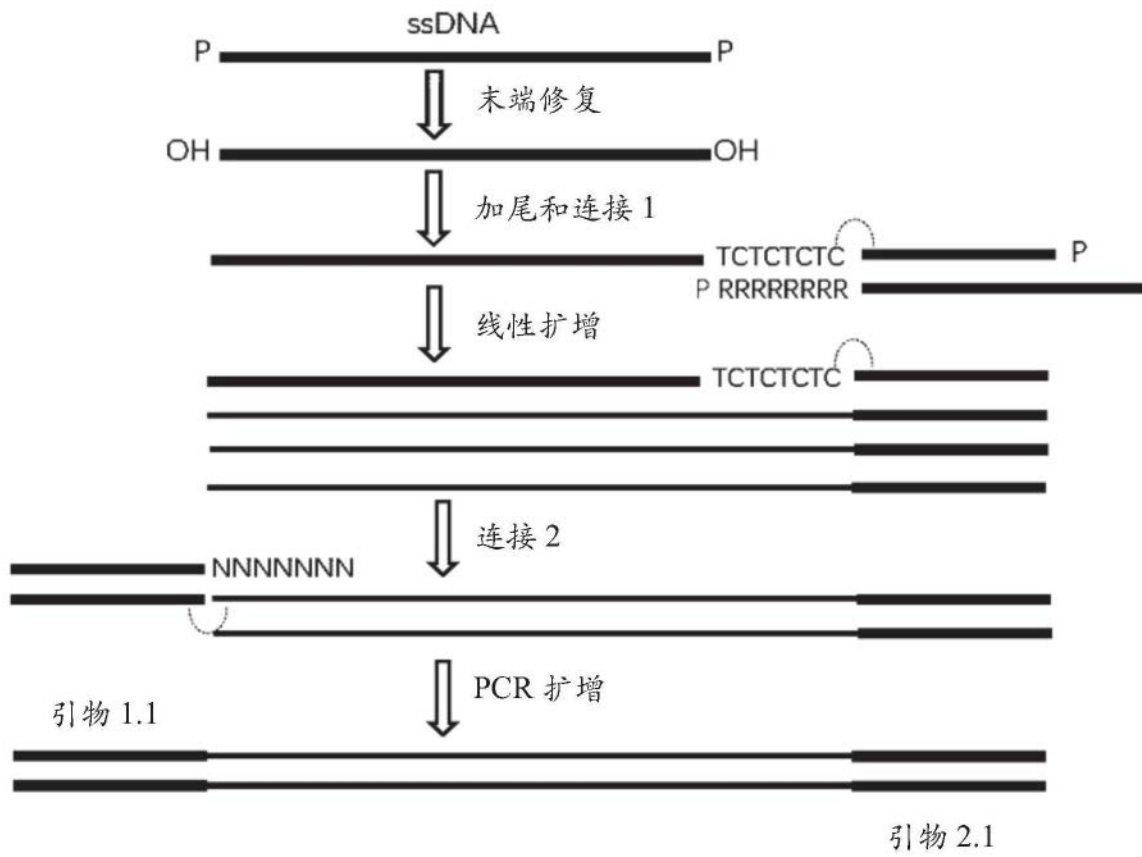


图9

MDA1

5' p A G A T C G G A A G A G C A C A C G T C T G A A C T C C A G T C A C p 3'

5' A G T C G T G A C T G G A G T T C A G A C G T G T G C T C T C C G A T C T R R R R R R R p 3'

MDA2

5' p A G A T C G G A A G A G C G T C G T G T A G G G A A A G A G T G T p 3'

5' A C A C T C T T T C C C T A C A C G A C G C T C T T C C G A T C T N N N N N N N p 3'

图10

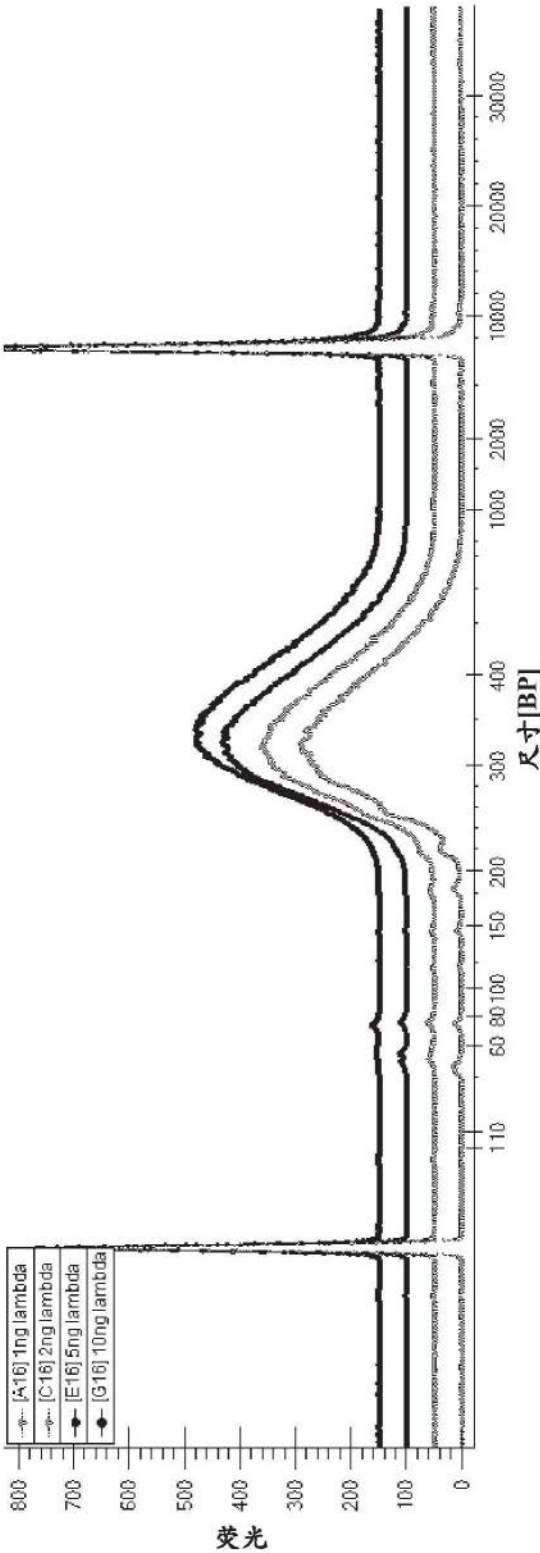


图11