



(12) 发明专利

(10) 授权公告号 CN 102373288 B

(45) 授权公告日 2013. 12. 11

(21) 申请号 201110391990. 2

(22) 申请日 2011. 11. 30

(73) 专利权人 盛司潼

地址 518057 广东省深圳市南山区高新区科技中二路软件园二期 11 栋 4 楼北 402 室

(72) 发明人 盛司潼

(51) Int. Cl.

C12Q 1/68(2006. 01)

(56) 对比文件

WO 2011/074960 A1, 2011. 06. 23, 权利要求书, 说明书第 20 页第 20 行 - 第 24 页第 27 行、第 24 页第 30 行 - 第 27 页第 24 行, 附图 3、4.

CN 102212612 A, 2011. 10. 12, 权利要求书, 实施例, 附图 1、3.

CN 101278058 A, 2008. 10. 01, 权利要求书, 说明书实施例, 附图 1.

CN 101432438 A, 2009. 05. 13, 权利要求书,

实施例, 附图 1、2.

WO 2010/039991 A2, 2010. 04. 08, 权利要求书, 实施例, 附图 3-5.

WO 01/83696 A2, 2001. 11. 08, 全文.

US 2011/0189677 A, 2011. 08. 04, 全文.

CN 101434988 A, 2009. 05. 20, 全文.

审查员 曲凯

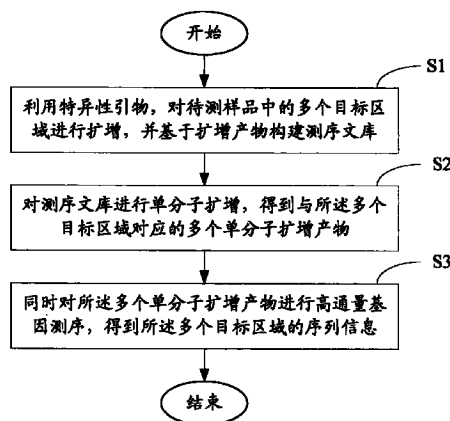
权利要求书2页 说明书16页
序列表6页 附图4页

(54) 发明名称

一种对目标区域进行测序的方法及试剂盒

(57) 摘要

本发明涉及基因工程领域, 提供了一种对目标区域进行测序的方法及试剂盒, 该方法包括以下步骤: A. 利用特异性引物, 对待测样品中的多个目标区域进行扩增, 并基于扩增产物构建测序文库; B. 对测序文库进行单分子扩增, 得到与所述多个目标区域对应的多个单分子扩增产物; C. 同时对所述多个单分子扩增产物进行高通量基因测序, 得到所述多个目标区域的序列信息。该方法及其相应的试剂盒利用高通量测序技术对多个目标区域同时进行区域测序, 能准确得出这些区域的序列信息, 包括已知突变和未知突变位点的变异情况, 检测灵敏度高, 并能进一步同时对大量样品进行检测。



1. 一种对目标区域进行测序的方法,其特征在于,包括以下步骤:

A. 利用特异性引物,对待测样品中的多个目标区域进行扩增,并基于扩增产物构建测序文库;

B. 对测序文库进行单分子扩增,得到与所述多个目标区域对应的多个单分子扩增产物;

C. 同时对所述多个单分子扩增产物进行高通量基因测序,得到所述多个目标区域的序列信息;

所述步骤 A 包括以下步骤:

A1. 利用特异性引物,对待测样品中的多个目标区域进行扩增,得到与所述多个目标区域对应的扩增产物;

A2. 利用接头元件,与所述多个目标区域对应的扩增产物进行连接,得到测序文库;

所述接头元件包括至少一个分叉接头;所述分叉接头是双链核酸分子,包括配对区和分叉区;所述分叉区的两条单链各包含至少一个引物结合位点;所述配对区包括至少一个限制性内切酶识别位点,所述配对区的 3' 末端为平末端或突出末端;所述配对区的 3' 末端为平末端时,平末端最后一个核苷酸为带有双脱氧碱基的核苷酸;

所述接头元件中的至少一个接头包含有第一标签序列,用于在文库构建过程中,对不同待测样品的测序文库进行标记;

每个目标区域对应的特异性引物中,至少有一条引物与该目标区域部分互补,该部分互补的引物的 5' 端包含有第二标签序列,用于在扩增目标区域过程中,对不同待测样品的目标区域扩增产物进行标记。

2. 根据权利要求 1 所述的对目标区域进行测序的方法,其特征在于,所述分叉接头中至少一个分叉接头的配对区的 3' 末端为突出末端,该突出末端为 T。

3. 根据权利要求 1 所述的对目标区域进行测序的方法,其特征在于,所述步骤 A2 包括以下步骤:

A21. 对与所述多个目标区域对应的扩增产物进行片段化,得到片段化产物;

A22. 利用接头元件,与所述片段化产物进行连接,构建测序文库。

4. 根据权利要求 3 所述的对目标区域进行测序的方法,其特征在于,所述步骤 A22 包括以下步骤:

A221. 利用第一接头与片段化产物的两端连接,得第一连接产物;

A222. 环化第一连接产物,得环化产物;

A223. II s 型限制性内切酶酶切环化产物,得酶切产物;

A224. 在酶切产物两端接上第二接头和第三接头,得测序文库。

5. 根据权利要求 3 所述的对目标区域进行测序的方法,其特征在于,所述步骤 A22 包括以下步骤:

A221'. 利用第四接头与片段化产物连接,得第二连接产物;

A222'. II s 型限制性内切酶酶切第二连接产物,得带有第四接头的酶切片段;

A223'. 带有第四接头的酶切片段与第五接头连接,形成测序文库。

6. 根据权利要求 1 至 5 中任一项所述的对目标区域进行测序的方法,其特征在于,所述特异性引物为 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3 和 SEQ ID NO:4。

7. 根据权利要求5所述的对目标区域进行测序的方法,其特征在于,所述第四接头为由 SEQ ID NO:13 和 SEQ ID NO:14 组成的分叉接头,SEQ ID NO:13 中的 5'-NNNN-3' 和 SEQ ID NO:14 中的 5'-NNNN-3' 之间反向互补,所述第五接头为由 SEQ ID NO:15 和 SEQ ID NO:16 组成的分叉接头,SEQ ID NO:15 中的 5'-NNN-3' 和 SEQ ID NO:16 中的 5'-NNN-3' 之间反向互补。

8. 根据权利要求5所述的对目标区域进行测序的方法,其特征在于,步骤B中所述单分子扩增使用的引物为 SEQ ID NO:5 和 SEQ ID NO:6。

9. 一种对目标区域进行测序的试剂盒,其特征在于,包括:

特异性引物,用于对待测样品中的多个目标区域进行扩增;

接头元件,用于与扩增产物结合构建测序文库;

所述接头元件包括至少一个分叉接头;所述分叉接头是从双链核酸分子,包括配对区和分叉区;所述分叉区的两条单链各包含至少一个引物结合位点;所述配对区包括至少一个限制性内切酶识别位点,所述配对区的 3' 末端为平末端或突出末端;所述配对区的 3' 末端为平末端时,平末端最后一个核苷酸为带有双脱氧碱基的核苷酸;

所述接头元件中的至少一个接头包含有第一标签序列,用于在文库构建过程中,对不同待测样品的测序文库进行标记;

每个目标区域对应的特异性引物中,至少有一条引物与该目标区域部分互补,该部分互补的引物的 5' 端包含有第二标签序列,用于在扩增目标区域过程中,对不同待测样品的目标区域扩增产物进行标记。

10. 根据权利要求9所述的对目标区域进行测序的试剂盒,其特征在于,所述分叉接头中至少一个分叉接头的配对区的 3' 末端为突出末端,该突出末端为 T。

11. 根据权利要求9、10所述的对目标区域进行测序的试剂盒,其特征在于,所述特异性引物为 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3 和 SEQ ID NO:4。

12. 根据权利要求9、10所述的对目标区域进行测序的试剂盒,其特征在于,所述接头元件包括第四接头和第五接头,所述第四接头为由 SEQ ID NO:13 和 SEQ ID NO:14 组成的分叉接头,SEQ ID NO:13 中的 5'-NNNN-3' 和 SEQ ID NO:14 中的 5'-NNNN-3' 之间反向互补,所述第五接头为由 SEQ ID NO:15 和 SEQ ID NO:16 组成的分叉接头,SEQ ID NO:15 中的 5'-NNN-3' 和 SEQ ID NO:16 中的 5'-NNN-3' 之间反向互补。

13. 根据权利要求9、10所述的对目标区域进行测序的试剂盒,其特征在于,还包括单分子扩增引物,所述单分子扩增引物为 SEQ ID NO:5 和 SEQ ID NO:6。

一种对目标区域进行测序的方法及试剂盒

技术领域

[0001] 本发明涉及基因工程领域,更具体地说,涉及一种对目标区域进行测序的方法及试剂盒。

背景技术

[0002] 自从人类基因组核苷酸序列草图和许多其他基因组草图问世,以及人类和许多其他生物体基因组草图中大量单核苷酸多态性(SNP)和小的插入/缺失多态性的发现,人们开始对核酸测定产生兴趣。研究表明,某些特定位置的核酸序列变异包含了大量的生物学信息,可能与人类或动植物的某些性状相关,这些信息将对医药和农业等的方方面面产生广泛的影响。举例来说,可以预期的是,这个领域的发展有可能实现个性化医疗。

[0003] 目前,对目标区域(特定位置)进行测序的常见方法为Sanger测序法,通过Sanger测序法能够对目标区域进行区域检测,但是Sanger测序法每次只能对一个样品的某一段区域进行测序,测序成本高,灵敏度低(20%),且不能得出各突变位点发生变异的精确数据。

[0004] 因此,需要一种能同时对待测样品的多个目标区域进行区域检测的新方法,该方法能够准确得到这些区域的序列信息,包括已知突变和未知突变的各突变位点的变异情况,检测灵敏度高,还能进一步同时对大量样品进行检测。

发明内容

[0005] 本发明的目的在于提供一种对目标区域进行测序的方法及试剂盒,能同时对待测样品的多个目标区域进行区域检测,准确得到这些区域的序列信息,包括已知突变和未知突变的各突变位点的变异情况,检测灵敏度高,还能进一步同时对大量样品进行检测。

[0006] 本发明是这样实现的,一种对目标区域进行测序的方法,包括以下步骤:

[0007] A. 利用特异性引物,对待测样品中的多个目标区域进行扩增,并基于扩增产物构建测序文库;

[0008] B. 对测序文库进行单分子扩增,得到与所述多个目标区域对应的多个单分子扩增产物;

[0009] C. 同时对所述多个单分子扩增产物进行高通量基因测序,得到所述多个目标区域的序列信息。

[0010] 其中,所述步骤A包括以下步骤:

[0011] A1. 利用特异性引物,对待测样品中的多个目标区域进行扩增,得到与所述多个目标区域对应的扩增产物;

[0012] A2. 利用接头元件,与所述多个目标区域对应的扩增产物进行连接,得到测序文库;所述接头元件采用平末端接头、突出末端接头、带茎环结构的接头和分叉接头中的至少一种。

[0013] 其中,步骤A2包括以下步骤:

- [0014] A21. 对与所述多个目标区域对应的扩增产物进行片段化,得到片段化产物;
- [0015] A22. 利用接头元件,与片段化产物进行连接,构建测序文库。
- [0016] 步骤 A 所述的目标区域测序文库的长度无特殊限制,优选为 25bp ~ 500bp。更优选为 50bp ~ 200bp,更优选为 70bp ~ 130bp。
- [0017] 其中,所述步骤 A22 包括以下步骤:
- [0018] A221. 利用第一接头与片段化产物的两端连接,得第一连接产物;
- [0019] A222. 环化第一连接产物,得环化产物;
- [0020] A223. II s 型限制性内切酶酶切环化产物,得酶切产物;
- [0021] A224. 在酶切产物两端接上第二接头和第三接头,得测序文库。
- [0022] 其中,所述步骤 A22 包括以下步骤:
- [0023] A221'. 利用第四接头与片段化产物连接,得第二连接产物;
- [0024] A222'. II s 型限制性内切酶酶切第二连接产物,得带有第四接头的酶切片段;
- [0025] A223'. 带有第四接头的酶切片段与第五接头连接,形成测序文库。
- [0026] 其中,步骤 A2 所述接头元件中的至少一个接头包含有第一标签序列,用于在文库构建过程中,对不同待测样品的测序文库进行标记。所述第一标签序列,优选为带有特定碱基序列的核酸分子,其碱基数不限。
- [0027] 进一步的,所述第一标签序列碱基数为 3 ~ 20,更优选为 4 ~ 10。
- [0028] 其中,步骤 A 所述的特异性引物与目标区域完全互补或部分互补。
- [0029] 进一步的,每个目标区域对应的特异性引物中,至少一条引物与该目标区域部分互补,该部分互补的引物的 5' 端包含有第二标签序列,用于在扩增目标区域过程中,对不同待测样品的目标区域扩增产物进行标记。所述第二标签序列,优选为带有特定碱基序列的核酸分子,其碱基数不限。
- [0030] 进一步的,所述第二标签序列碱基数为 3 ~ 20,更优选为 4 ~ 10。
- [0031] 其中,所述步骤 A 中同一样品的各目标区域的扩增,同时进行或部分同时进行或分别独立进行。
- [0032] 其中,步骤 B 所述的单分子扩增的方法为乳液 PCR、桥式 PCR 中的至少一种。
- [0033] 其中,步骤 C 所述的高通量测序技术为基于聚合酶的合成测序法或基于连接酶的连接测序法。
- [0034] 本发明的还提供了一种能够用于本发明的任一种测序方法的试剂盒,本发明是这样实现的,一种对目标区域进行测序的试剂盒,包括:
- [0035] 特异性引物,用于对待测样品中的多个目标区域进行扩增;
- [0036] 接头元件,用于与扩增产物结合构建测序文库。
- [0037] 其中,所述接头元件采用平末端接头、突出末端接头、带茎环结构的接头和分叉接头中的至少一种。
- [0038] 其中,所述接头元件中的至少一个接头包含有第一标签序列,用于在文库构建过程中,对不同待测样品的测序文库进行标记。所述第一标签序列优选为带有特定碱基序列的核酸分子,其碱基数不限,优选为 3 ~ 20。
- [0039] 其中,所述的特异性引物与目标区域完全互补或部分互补。
- [0040] 进一步的,每个目标区域对应的特异性引物中,至少一条引物与该目标区域部分

互补,该部分互补的引物的 5' 端包含有第二标签序列,用于在扩增目标区域过程中,对不同待测样品的目标区域扩增产物进行标记。所述第二标签序列优选为带有特定碱基序列的核酸分子,其碱基数不限,优选为 3 ~ 20,更优选为 4 ~ 10。

[0041] 与现有技术相比,本发明的方法及试剂盒通过高通量测序技术对待测样品的多个目标区域同时进行深度测序,准确得出这些目标区域的序列信息,包括已知突变和未知突变的各突变位点的变异情况,精确得出各样品的各突变位点发生变异的频率,检测灵敏度高,并能进一步对大量样品同时进行多区域测序。

附图说明

[0042] 图 1 是本发明一个实施例中的对目标区域进行测序的方法流程图;

[0043] 图 2 是本发明一个实施例中的单突出末端接头的接头示意图;

[0044] 图 3 是本发明一个实施例中的双突出末端接头的结构示意图;

[0045] 图 4 是本发明一个实施例中的带茎环结构的接头的结构示意图;

[0046] 图 5 是本发明一个实施例中的分叉接头的结构示意图;

[0047] 图 6 是本发明一个实施例中的 T 末端分叉接头的结构示意图;

[0048] 图 7 是本发明另一个实施例中的分叉接头的结构示意图;

[0049] 图 8 是本发明一个实施例中的双脱氧分叉接头的结构示意图;

[0050] 图 9 是本发明一个实施例中利用片段化产物和接头元件构建测序文库的方法流程图;

[0051] 图 10 是本发明另一个实施例中利用片段化产物和接头元件构建测序文库的方法流程图;

[0052] 图 11 是本发明另一个实施例中利用片段化产物和接头元件构建测序文库的方法流程图。

具体实施方式

[0053] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。

[0054] 本发明所述的目标区域,为任意基因组上的任意序列,可根据需要进行选择,包括但不限于基因的内部序列、基因的外部调控区域,所述基因的内部序列,包括但不限于基因的内含子区域、外显子区域、同时含有内含子与外显子的区域。

[0055] 图 1 示出了本发明的一种对目标区域进行测序的方法流程,该方法包括以下步骤:

[0056] S1. 利用特异性引物,对待测样品中的多个目标区域进行扩增,并基于扩增产物构建测序文库;

[0057] S2. 对测序文库进行单分子扩增,得到与所述多个目标区域对应的多个单分子扩增产物;

[0058] S3. 同时对所述多个单分子扩增产物进行高通量基因测序,得到所述多个目标区域的序列信息。

[0059] 本方法通过高通量测序技术对待测样品的目标区域进行深度测序,该方法能同时

对待测样品的多个目标区域进行检测,准确得到这些区域的序列信息,包括已知突变和未知突变的各突变位点的变异情况,检测灵敏度高,能精确得出各样品的各突变位点发生变异的频率。此外,通过控制各目标区域扩增产物的大小,可免去了片段化步骤,减少了实验步骤,提高了实验效率,降低了成本。

[0060] 通过本方法所得的序列信息,可用于各种科学研究,包括但不限于人群序列分析、基因功能研究、蛋白功能研究。

[0061] 需要说明的是:

[0062] 步骤 S1 中所述待测样品为能提取核酸的任意形式的样品,包括但不限于:全血、血清、血浆和组织样品;所述组织样品包括但不限于:石蜡包埋组织、新鲜组织和冰冻切片。

[0063] 步骤 S1 所得测序文库中,存在多种测序文库分子,对测序文库进行单分子扩增,即是指,将测序文库中的多种文库分子,以极微量(甚至单分子)的形式在空间上隔离(但这些文库分子整体上还是属于同一个反应体系),并且在各自的空间内实现扩增,以提升各种分子在后续测序反应中的信号。

[0064] 现有技术中,Sanger 测序技术每次只能对一个样品的某一段区域进行测序,要实现对多个目标区域的测序,只能是通过多次反应来实现。而在本发明中,测序文库中的各个分子经过单分子扩增后,每个测序文库分子均形成单分子拷贝阵列,各单分子拷贝阵列在进行高通量基因测序时处于不同的位置,使得测序引物与单分子拷贝阵列之间的杂交,以及在酶作用下的延伸反应可同时进行,相互之间互不干扰。因此,可以同时大量的(成百万上千万,甚至更多的)单分子拷贝阵列同时进行测序反应,然后通过采集相应的信号,进而获得所需的序列信息,且测序的灵敏度较 Sanger 更高。

[0065] 其中,步骤 S1 所述的特异性引物与目标区域完全互补或部分互补。

[0066] 进一步的,每对针对目标区域的引物中,至少一条引物与目标区域部分互补,且该引物的 5' 端带有第二标签序列。该第二标签序列,用于在扩增目标区域过程中,对不同待测样品的目标区域扩增产物进行标记。

[0067] 该第二标签序列,优选为带有特定序列的核酸分子,其碱基数不限。该第二标签序列的碱基数优选为 3 ~ 20,更优选为 4 ~ 10。

[0068] 此外,所述特异性引物还可带有其它标记物,包括但不限于:生物素标记、多聚组氨酸标记、抗原、抗体,从而使得目标区域扩增产物的纯化极为方便。

[0069] 另外,同一待测样品的不同目标区域的扩增可同时进行或分别独立进行或部分同时进行。在具体的实验过程中,可根据需要选用上述任一种方案进行。

[0070] 如果分别对各目标区域进行分别扩增的话,能够通过测定扩增产物的量来确保步骤 S1 中用于构建目标区域测序文库的目标区域扩增产物的分子数保持一致,不会因为扩增步骤而导致同一待测样品的不同目标区域拷贝数不同,进而影响后续的测序反应结果。

[0071] 当然,当各目标区域的大小相近,GC 含量也相近的时候,通过合理的设计引物,利用多重 PCR 技术,步骤 S1 可对多个目标区域进行同时扩增,且保证各目标区域之间的扩增效率保持基本一致,这样就能够有效的提高实验效率,降低反应的成本。

[0072] 当待测样品有多个时,不同样品的目标区域的扩增必须分别进行。

[0073] 在一个实施例中,步骤 S1 的具体实现过程是:

[0074] S11. 利用特异性引物,对待测样品中的多个目标区域进行扩增,得到与所述多个目标区域对应的扩增产物;

[0075] S12. 利用接头元件,与所述多个目标区域对应的扩增产物进行连接,得到测序文库;所述接头元件采用平末端接头、突出末端接头、带茎环结构的接头和分叉接头中的至少一种。

[0076] 需要说明的是:

[0077] 步骤 S11 中,对同一待测样品中的多个目标区域的扩增,可同时进行或分别独立进行或部分同时进行。可根据实际情况,如:扩增特异性引物的退火温度,扩增的目标区域的大小、GC 含量,扩增的目标区域的数量等,进行相应的调整。不同待测样品的目标区域的扩增必须分别进行。

[0078] 步骤 S12 中,所述接头元件与扩增产物的连接方式,可以采用多种方式实现,包括接头元件与扩增产物直接连接,或对扩增产物进行处理之后再行连接。

[0079] 步骤 S12 中的接头元件,用于构建测序文库,可包括一种或多种接头。

[0080] 其中,步骤 S12 所述接头元件中的至少一个接头包含有第一标签序列,该第一标签序列,用于在文库构建过程中,对不同待测样品的测序文库进行标记。这样,在分别获得目标区域测序文库后,不同的待测样品的目标区域测序文库可以混合在同一个反应体系中,进行单分子扩增反应,进而同时进行高通量测序。提高测序反应的效率,降低了样品检测的成本。

[0081] 该第一标签序列优选为带有特定碱基序列的核酸分子,其碱基数不限。进一步的,所述第一标签序列的碱基数为 3~20,这样,每次至少能够对 4^3 个样品进行同时检测。在综合考虑各种情况后,如:标签的特异性、接头的成本、接头的长度等,第一标签序列的碱基数优选为 4~10。通过第二标签和第一标签的结合,本发明的发明每次能够检测至少 $4^3 \times 4^3$ 个样品,即 4096 个样品。

[0082] 接头的修饰方式有多种,包括但不限于:被生物素化或甲基化,或同时被生物素化和甲基化。在一个实施例中,该接头被生物素化,并与未生物素化的片段化产物连接,生物素的存在有利于构建的测序文库的分离纯化。在另一个实施例中,该接头被甲基化,并与未甲基化的片段化产物连接,然后用仅切割甲基化 DNA 的限制性内切酶消化连接产物,因为只有成功连接的连接产物才能被切割,所以只有成功连接的连接物被切割,从而确保酶切产物的单一性。

[0083] 接头的结构形式也有多种,包括但不限于:平末端接头、突出末端接头、带茎环结构的接头和分叉接头。构建测序文库过程中可以使用一种或多种接头。其中,突出末端接头、带茎环结构的接头和分叉接头均能够有效防止在连接过程中多个接头自连现象的发生。针对接头的上述接头形式,以下将提供多个实施例。

[0084] 在第一实施例中,接头元件采用平末端接头,该接头是双链完全互补的核酸分子。

[0085] 在第二实施例中,接头元件采用突出末端接头,该接头是双链核酸分子,该双链核酸分子至少包括一突出末端。该突出末端的碱基数无具体限制,优选为 1~10 个碱基。根据该双链核酸分子的结构,突出末端接头可分成两类,分别是单突出末端接头、双突出末端接头。

[0086] 如图 2 所示的单突出末端接头,其一端为平末端,另一端为突出末端。其中带有分

叉接头的单突出末端接头能够防止接头自连。为了防止一端为平末端的单突出末端接头自连,可对平末端上的 3' OH 进行修饰(包括但不限于用氨基封闭羟基),或将平末端上的 5' 磷酸基团去除。

[0087] 如图 3 所示的双突出末端接头,其含有两个突出末端,这两个突出末端可在一条核苷酸链上(图 3a)或在不同的核苷酸链上(图 3b)。当这两个突出末端在不同的核酸链上时,他们相互之间不互补,以防在连接时出现接头自连。

[0088] 在第三实施例中,接头元件采用带茎环结构的接头,如图 4 所示。该接头为单链核酸分子,该单链核酸分子包括第一互补配对区 1、茎环区 2 和第二互补配对区 3(图 4a),第一互补配对区 1 能够与第二互补配对区 3 互补配对,且它们形成的互补配对区包括至少一个限制性内切酶识别位点,而通过该酶切识别位点,特定的酶能够将茎环区切开或切除,从而将单链核酸分子变成双链核酸分子,以便于后续的操作。如图 4b 所示,带茎环结构的接头还可带有突出末端 4,该突出末端可位于单链核酸分子的 5' 端或 3' 端。突出末端 4 的存在能够进一步的防止接头自连现象的发生。该突出末端优选为 T。

[0089] 在第四实施例中,接头元件采用分叉接头,如图 5 所示,该接头是双链核酸分子,包括配对区和分叉区,所述分叉区的两条单链各包含至少一个扩增引物结合位点。优选的,所述分叉接头的配对区包括至少一个限制性内切酶识别位点,该酶切识别位点可以在建库过程中酶切形成末端,以便于进行后续的操作。

[0090] 该分叉接头的分叉设计能够在建库过程避免多个接头自连现象的出现;所述分叉区上包含的扩增引物结合位点,可以直接用于结合扩增引物,进行扩增反应。

[0091] 其中,所述分叉接头的分叉区的每条链各含有 N 个核苷酸;优选的, $9 \leq N \leq 30$ 。其中,所述分叉接头的配对区互补配对的核苷酸对数不限;优选的,互补配对的核苷酸对数为 7 ~ 15,更优选的,互补配对的核苷酸对数为 9 ~ 13。

[0092] 其中,所述分叉接头的配对区的 3' 末端为突出末端或平末端。优选的,所述分叉接头的配对区的 3' 末端为突出末端,该突出末端可与所述片段化产物的粘性末端互补配对,提高了连接效率,以利于构建测序文库反应的顺利进行。

[0093] 优选的,所述分叉接头为 T 末端分叉接头,该接头的配对区的 3' 末端为突出末端,且突出末端最后一个碱基为 T;例如图 6 所示的 T 末端分叉接头,图中 N 为 A、T、C、G 碱基中的任一种。

[0094] 优选的,所述分叉接头的配对区的 3' 末端为突出末端,且突出末端的核苷酸包括通用碱基。其中,突出末端的碱基数无特殊限制,优选为 1 ~ 4。例如图 7 所示的分叉接头,图中 N 为 A、T、C、G 碱基中的任一种, X 为通用碱基。

[0095] 优选的,所述分叉接头为双脱氧接头,该接头的配对区的 3' 末端为平末端,且 3' 末端最后一个核苷酸为带有双脱氧碱基的核苷酸;例如图 8 所示的双脱氧分叉接头,图中 N 为 A、T、C、G 碱基中的任一种, dd 表示该 3' 末端最后一个核苷酸为带有双脱氧碱基的胞嘧啶核苷酸。

[0096] 应当说明的是,上述接头元件只是部分实施例,并不用以限制本发明的保护范围。

[0097] 关于步骤 S12 的实现方式:

[0098] 在一个实施例中,步骤 S12 采用接头元件与扩增产物直接连接的方式,构建出测序文库。

[0099] 在另一个实施例中,当目标区域扩增产物较大时或者需要构建的目标区域测序文库较小时,则对扩增产物进行片段化处理之后,片段化产物再与接头元件进行连接,构建测序文库,如图 9 所示,步骤 S12 包括以下步骤:

[0100] S121. 对与所述多个目标区域对应的扩增产物进行片段化,得到片段化产物;

[0101] S122. 利用接头元件,与片段化产物进行连接,构建测序文库。

[0102] 通过片段化步骤 S121 将目标区域扩增产物变成较小的片段化产物,从而有助于对目标区域的进一步深度测序。此外,还可以通过片段化处理,将不同的扩增产物变成长短相似的片段化产物,能够有助于后续的统一测序。

[0103] 需要说明的是:

[0104] 步骤 S121 中,所述片段化目标区域扩增产物的方法有多种,包括但不限于:超声法、喷雾法、化学剪切法和酶切法。可根据实际情况,采用相适应的方法进行实验。

[0105] 所述片段化处理之后还可包括片段化产物的分离纯化以及末端修饰的步骤。根据测序的片段长度需要,对于片段化得到的核酸片段,进行目的核酸片段的分离纯化,分离方法可以采用常用方法,如凝胶电泳、蔗糖梯度或氯化铯梯度沉降、柱层析分离等。根据所使用的片段化方法,对所得的目的核酸片段进一步的末端修饰,包括但不限于:磷酸化或去磷酸化、末端补平和末端加 A,以便于后续的接头元件连接。上述目的核酸片段长短不限,优选为 25bp ~ 500bp,更优选为 30bp ~ 200bp,更优选为 40 ~ 100bp。

[0106] 步骤 S121 所述片段化产物的长度不限,优选为 25bp ~ 500bp,更优选为 30bp ~ 200bp,更优选为 40bp ~ 100bp。在实现对目标区域的测序的前提下,随着目标区域测序文库分子中含有的目标区域片段长度的减短,高通量测序技术的对目标区域的测序深度加深;而测序深度越深,即对目标区域的每一个碱基位置的测序次数越多,测序结果越准确,对样品中的少量突变的检测就越灵敏;这样就能够有效防止因为样品中带有突变的目标区域的比例偏低,而导致该突变的测序信号的绝对值偏低,发生测序结果不准确的现象。

[0107] 为了实现对目标区域测序文库大小的限制,可在步骤 S121 或 S122 之后,对片段化产物或目标区域测序文库进行分离纯化。分离纯化的方法有多种,包括但不限于:凝胶方法、蔗糖梯度或氯化铯梯度沉降和柱层析分离。可根据实际情况,采用相适应的方法进行实验。

[0108] 根据上述接头元件,针对步骤 S122,下面将通过多个实施例和附图对本步骤进行进一步的说明。

[0109] 在本发明的一个实施例中,直接在片段化产物的两端接上接头形成测序文库。

[0110] 所述接头可采用上述的平末端接头、突出末端接头、带茎环结构的接头和分叉接头中的至少一种。

[0111] 在本发明的另一个实施例中,如图 10 所示,步骤 S122 具体可由以下步骤实现:

[0112] S1221. 利用第一接头与片段化产物的两端连接,得第一连接产物;

[0113] S1222. 环化第一连接产物,得环化产物;

[0114] S1223. II s 型限制性内切酶酶切环化产物,得酶切产物;

[0115] S1224. 在酶切产物两端接上第二接头和第三接头,得测序文库。

[0116] 在步骤 S1221 中,所述第一接头可采用平末端接头、突出末端接头、带茎环结构的接头和分叉接头中的一种,所述第一接头包含有 II s 型限制性内切酶酶切识别位点;所述

的 II s 型限制性内切酶为切割位在识别序列之外的限制性内切酶,包括但不限于:Acu I、Alw I、Bbs I、Bbv I、Bcc I、BceA I、BciV I、BfuA I、Bmr I、Bpm I、BpuE I、Bsa I、BseM II、BseR I、Bsg I、BsmA I、BsmB I、BsmF I、BspCN I、BspM I、BspQ I、BtgZ I、Ear I、Eci I、EcoP15 I、Fau I、Fok I、Hga I、Hph I、HpyAV、Mbo II、Mly I、Mme I、Mnl I、NmeAIII、Ple I、Sap I、SfaN I 和 TspDT I,优选为 Acu I、Bsg I、EcoP15 I 或 Mme I。

[0117] 当所述第一接头为分叉接头时,该 II s 型限制性内切酶酶切识别位点位于配对区;当所述第一接头为带茎环结构的接头时,限制性内切酶识别位点与茎环结构之间的距离,较 II s 型限制性内切酶酶切识别位点与茎环结构之间的距离近。

[0118] 若片段化产物经过末端修复酶修复,以及末端加 A 反应,所述第一接头优选为带 T 末端的分叉接头。若片段化产物只是经过末端修复酶修复,将片段化产物的末端补平,则所述第一接头优选双脱氧接头。

[0119] 在步骤 S1222 中,环化第一连接产物有多种实现方式。

[0120] 在本发明的一实施例中,步骤 S1222 包括以下步骤:

[0121] S12221. 利用酶切引物对第一连接产物进行扩增,得扩增产物;

[0122] S12222. 对扩增产物进行酶切,使得扩增产物形成粘性末端,并自身环化成环化产物。

[0123] 所述酶切引物的 3' 端分别与第一连接产物的两个末端部分互补,5' 端均含有限制性内切酶识别位点。经过步骤 S12221 的扩增形成的扩增产物,其两个末端均含有限制性内切酶识别位点,然后在相应的酶的作用下,使扩增产物的两端形成粘性末端,且这两个粘性末端互补,能够进行自身环化。

[0124] 在本发明的另一个实施例中,所述第一接头包含有 2 个酶切识别位点,其中一个为限制性内切酶识别位点,用于使步骤 S1222 形成的第一连接产物的两端在相应的酶的作用下,形成粘性末端,且它们之间互补,能够进行自身环化。另一个为 II s 型限制性内切酶酶切识别位点,用于在步骤 S1223 中,利用识别该酶切位点的酶识别环化产物,进行酶切,进而得到酶切产物。

[0125] 应当说明,以上两个实施例仅为本发明中的两种实现环化第一连接产物的实施方案,对于本发明的保护范围不做任何具体的限制。

[0126] 在步骤 S1223 中,利用能够识别第一接头上的酶切识别位点,并切割环化产物(DNA) 但不切割第一接头的酶进行酶切。所述第一接头上的酶切识别位点包括但不限于:Mme I 酶切识别位点、Acu I 酶切识别位点、Bsg I 酶切识别位点。

[0127] 步骤 S1224 中,所述第二接头可为平末端接头、突出末端接头、带茎环结构的接头和分叉接头中的一种,可带有生物素标记。所述第三接头可为平末端接头、突出末端接头、带茎环结构的接头和分叉接头中的一种。第二接头与第三接头可相同或不同。优选的,所述第二接头和第三接头相同,均为分叉接头。更优选的,所述分叉接头为 T 末端分叉接头或双脱氧分叉接头。

[0128] 需要特别说明的是,步骤 S1222 与 S1223 之间还可包括步骤 S1222A:滚环扩增环化产物,得滚环扩增产物。通过步骤 S1222A,能够保证后续的酶切步骤 S1223 有足够的原材料。

[0129] 又或者,在步骤 S1221 与 S1222 之间还可包括步骤 S1221A:利用扩增引物对第一

连接产物进行扩增,得扩增产物。所述扩增引物分别与第一连接产物两端的接头序列互补。通过步骤 S1221A,能够保证后续的环化步骤有足够的原材料。

[0130] 本方案可避免在步骤 S1222 之后进行滚环扩增,而用步骤 S1222 之前的步骤 S1221A 进行普通的 PCR 扩增代替,可有效减少后续酶切步骤中,II s 型限制性内切酶的用量。

[0131] 其中,所述第一接头优选为分叉接头。

[0132] 其中,所述分叉接头的配对区可包含至少一个酶切识别位点。所述酶切识别位点可为普通的限制性内切酶识别位点,也可为 II s 型限制性内切酶酶切识别位点。

[0133] 其中,步骤 S1221A 所述扩增引物优选为生物素化引物,有利于扩增产物的回收纯化。

[0134] 其中,步骤 S1221A 所述扩增引物带有至少一个特异性酶切识别位点。

[0135] 若扩增引物上所带的特异性酶切识别位点是尿嘧啶碱基,则步骤 S1222 中利用尿嘧啶特异性切除试剂进行酶切,然后再进行连接环化。

[0136] 若扩增引物所带特异性酶切识别位点为限制性内切酶识别位点,则步骤 S1222 中利用相应的限制性内切酶进行酶切,然后再进行连接环化。

[0137] 在本发明的另一个实施例中,如图 11 所示,步骤 S122 具体可由以下步骤实现:

[0138] S1221'. 利用第四接头与片段化产物连接,得第二连接产物;

[0139] S1222'. II s 型限制性内切酶酶切第二连接产物,得带有第四接头的酶切片段;

[0140] S1223'. 带有第四接头的酶切片段与第五接头连接,形成测序文库。

[0141] 需要说明的是:

[0142] 在步骤 S1221' 中,所述第四接头可采用平末端接头、突出末端接头、带茎环结构的接头和分叉接头中的一种,所述第四接头包含有 II s 型限制性内切酶酶切识别位点;当所述第四接头为分叉接头时,该 II s 型限制性内切酶酶切识别位点位于配对区;当所述第四接头为带茎环结构的接头时,限制性内切酶识别位点与茎环结构之间的距离,较 II s 型限制性内切酶酶切识别位点与茎环结构之间的距离近。

[0143] 在步骤 S1222' 中,利用能够识别第四接头上的酶切识别位点,并切割环化产物(DNA) 但不切割第四接头的酶进行酶切。所述第四接头上的酶切识别位点包括但不限于:Mme I 酶切识别位点、Acl I 酶切识别位点、Bsp I 酶切识别位点。若片段化产物经过末端修复酶修复,以及末端加 A 反应,所述第四接头优选为带 T 末端的分叉接头。若片段化产物只是经过末端修复酶修复,则所述第四接头优选双脱氧接头。

[0144] 在步骤 S1223' 中,所述第五接头可为平末端接头、突出末端接头、带茎环结构的接头和分叉接头中的一种,可带有生物素标记。

[0145] 在本发明的另一个实施例中,步骤 S122 具体包括以下步骤:

[0146] S1221". 直接在片段化产物的两端接上带茎环结构的接头,形成带茎环结构的片段化产物;

[0147] S1222". 利用限制性内切酶,将带茎环结构的片段化产物的茎环区切开或切除,从而形成测序文库。

[0148] 本技术方案利用带茎环结构的接头,防止了多个接头自连现象的发生。

[0149] 当通过上述几个方案形成的测序文库的数量过少,不利于后续的单分子扩增步骤

时,还可以进行以下步骤:

[0150] 利用与目标区域测序文库两端的接头部分互补的引物,以所得的测序文库为模版进行扩增,得扩增后的测序文库。该步骤可满足后续实验对目标测序文库的量的要求。

[0151] 其中,步骤 S2 所述的单分子扩增是指对目标区域测序文库中的分子,以极微量(甚至单分子)的形式在空间上隔离(但这些文库分子整体上还是属于同一个反应体系),并且在各自的空间内实现扩增,以提升各种分子在后续测序反应中的信号。所述单分子扩增的方法包括但不限于:乳液 PCR(Emulsion PCR, EPCR)、桥式 PCR。

[0152] 所述 EPCR 最大的特点是可以形成数目庞大的独立反应空间以进行 DNA 扩增。基本过程是在 PCR 反应前,将包含 PCR 所有反应成分的水溶液注入到高速旋转的矿物油表面,水溶液瞬间形成无数个被矿物油包裹的小水滴。这些小水滴就构成了独立的 PCR 反应空间。理想状态下,每个小水滴只含一个 DNA 模板(目标区域测序文库分子)和一个磁珠,磁珠上含有与目标区域测序文库分子的共有序列(由接头元件引入)互补的引物,在 PCR 反应后,磁珠表面就固定有拷贝数目巨大的同来源的 DNA 模板扩增产物。EPCR 的具体步骤可参考 PCR amplification from single DNA molecules on magnetic beads in emulsion: application for high-throughput screening of transcription factor targets, Takaaki Kojima, Yoshiaki Takei, Miharuru Ohtsuka et al, Nucleic Acids Research, 2005, Vol. 33, No. 17; Dual primer emulsion PCR for next generation DNA sequencing, Ming Yan Xu, Anthony D. Aragon, Monica R. Mascarenas et al, BioTechniques 48: 409-412 (May 2010); BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions, Frank Diehl, Meng Li, Yiping He, nature methods, Vol. 3, No. 7, July 2006 等。

[0153] 所述桥式 PCR 的基本原理是,桥式 PCR 的引物被固定在固相载体上,PCR 过程中 PCR 扩增产物会被固定在固相载体上,且 PCR 扩增产物能够与固相载体上的引物互补配对,成桥状,然后互补配对的引物以与其成桥的扩增产物为模板进行扩增。通过控制初始模板加入的量,桥式 PCR 反应完成后,扩增产物在固相载体上以一簇簇的形式存在,且每一簇的扩增产物为同来源的 DNA 模板扩增产物。其具体的原理和实施方案可参考以下文献: CN20061009879. X、US6227604。

[0154] 如前所述,现有技术中, Sanger 测序技术由于自身的技术限制,每次只能对一个样品的某一段区域进行测序。为了一次性实现对样品中多个区域的同时检测,本发明在测序方法上采取高通量基因测序方法。高通量基因测序相对 Sanger 测序法检测序列信息更为方便灵敏,测序文库中的各个分子经过单分子扩增后,每个测序文库分子均形成单分子拷贝阵列,各单分子拷贝阵列在进行高通量基因测序时处于不同的位置,使得测序引物与单分子拷贝阵列之间的杂交,以及在酶作用下的延伸反应可同时进行,相互之间互不干扰。因此,可以同时大量的(成百万上千万,甚至更多的)单分子拷贝阵列同时进行测序反应,然后通过采集相应的信号,进而准确的获得所需的序列信息,且测序的灵敏度较 Sanger 更高。尤其是对多个目标区域的扩增产物进行了片段化处理,相当于对相同序列的目标区域分子的每个碱基的测序次数增加了,能够进一步提高测序的灵敏度。

[0155] 其中,步骤 S3 所述的高通量测序技术包括但不限于:基于聚合酶的合成测序法、基于连接酶的连接测序法。

[0156] 基于聚合酶的合成测序法是基于带可去除标记的核苷酸进行的。在每次合成反应中,每个模板链至多只能延伸一次,基于聚合酶的合成测序法的大致流程如下:

[0157] a. 测序引物通过互补配对结合在单分子扩增产物共有的已知序列上(该单分子扩增产物固定在引物-固相载体复合物上),在 DNA 聚合酶的作用下,以带可去除标记的核苷酸进行单碱基延伸合成反应,收集该次加入核苷酸的标记信号,即可得到与测序引物 3' 最末端碱基互补的单分子扩增产物(固定在引物-固相载体复合物上)的下一位的碱基序列信息。

[0158] b. 切除可去除标记,然后在 DNA 聚合酶的作用下,以带可去除标记的核苷酸继续进行单碱基延伸合成反应,收集加入核苷酸的标记信号,即可得到与测序引物 3' 末端碱基互补的单分子扩增产物的下两位的碱基序列信息。

[0159] 重复 b 步骤,直至不能继续进行合成反应为止,从而获得单分子扩增产物的全部序列信息。

[0160] 基于连接酶的连接测序法均是基于带荧光标记的寡核苷酸探针进行的。其中一种连接酶的连接测序法是基于在特定位置带有荧光标记的寡核苷酸探针进行的,该寡核苷酸探针带有 n 个碱基,从其 5' 端数起的第 a 位带有荧光标记,其中不同的碱基对应特定的荧光标记,因为该寡核苷酸探针的 3' 端进行了特定的修饰,寡核苷酸探针之间不能直接相互连接,每次连接反应,每个单分子扩增产物只能连接一个寡核苷酸探针。该连接测序法的大致流程如下:

[0161] A. 测序引物通过互补配对结合在单分子扩增产物共有的已知序列上(该单分子扩增产物固定在引物-固相载体复合物上)上,利用上述的寡核苷酸探针,在连接酶的作用下,将核酸探针与上述寡核苷酸链连接,然后采集荧光信号,即可得到与单分子扩增产物共有的已知序列的 3' 末端后第 a 位碱基序列信息。

[0162] B. 切除寡核苷酸上的荧光标记,在连接酶的作用下,以上述的寡核苷酸探针为原料,继续进行连接反应,然后采集荧光信号,从而得到单分子扩增产物共有的已知序列的 3' 末端后第 $2a$ 位的碱基序列信息。

[0163] 重复 B 步骤,直至不能继续进行连接反应为止,从而获得单分子扩增产物共有的已知序列的 3' 末端后第 $a, 2a, 3a, 4a, \dots$ 位的碱基序列信息。

[0164] 然后将测序引物及其所连接的寡核苷酸探针从单分子扩增产物上变性洗脱下来,换用与之前的测序引物相比 3' 末端少一个碱基的引物重复上述反应,从而获得单分子扩增产物共有的已知序列的 3' 末端后第 $a-1, 2a-1, 3a-1, 4a-1, \dots$ 位的碱基序列信息。重复此步骤,最后获得单分子扩增产物共有的已知序列的 3' 末端后第 $a-(a-1), 2a-(a-1), 3a-(a-1), 4a-(a-1), \dots$ 位的碱基序列信息,从而获得单链扩增产物的全部序列信息。

[0165] 另一种连接酶的连接测序法同样也是基于带有荧光标记的寡核苷酸探针进行的,该寡核苷酸探针带有 n 个碱基,分为 h ($h \leq n$) 组,同一组寡核苷酸探针的不同荧光标记对应同一特定位置的不同碱基序列,不同组之间的区别在于:不同荧光标记对应的特定位置不同,因为该寡核苷酸探针的 3' 端进行了特定的修饰,寡核苷酸探针之间不能直接相互连接,每次连接反应,每个单分子扩增产物只能连接一个寡核苷酸探针。该连接测序法的大致流程如下:

[0166] a. 测序引物通过互补配对结合在单分子扩增产物共有的已知序列上(该单分子

扩增产物固定在引物-固相载体复合物上)上,利用上述寡核苷酸探针中的一组(荧光标记对应的碱基位置为 $x, x \leq h$),在连接酶的作用下,将核酸探针与上述寡核苷酸链连接,然后采集荧光信号,即可得到与单链扩增产物共有的已知序列的 3' 末端后第 x 位碱基序列信息,将测序引物及其所连接的寡核苷酸探针从单分子扩增产物上变性洗脱下来。

[0167] b. 然后重新将测序引物结合在单分子扩增产物上,换用与 a 步骤不同的寡核苷酸探针组(荧光标记对应的碱基位置为 $y, y \leq h$),在连接酶的作用下,将核酸探针与上述寡核苷酸链连接,然后采集荧光信号,即可得到与单链扩增产物共有的已知序列的 3' 末端后第 y 位碱基序列信息,将测序引物及其所连接的寡核苷酸探针从单分子扩增产物上变性洗脱下来。

[0168] c. 重复步骤 b,直至 h 组寡核苷酸探针均分别进行过一次连接反应,从而获得单分子扩增产物共有的已知序列的 3' 末端后第 1、2、...、 h 位的碱基序列信息。

[0169] 换用与之前的测序引物相比 3' 末端多一个或各通用碱基的引物按上述原理进行反应,能够延长获得的单分子扩增产物共有的已知序列的 3' 末端碱基序列的读长。

[0170] 这种基于连接酶的连接测序法的原理和具体实施方案可参考 CN200710170507.1。

[0171] 以检测 EGFR 基因的外显子 19、21 为例,本发明提出一实施例,设计用于扩增 EGFR 基因外显子 19、21 的特异性引物:19F(SEQ ID NO:1)、19R(SEQ ID NO:2)、21F(SEQ ID NO:3)、21R(SEQ ID NO:4)。

[0172] 一、待测样品 DNA 的提取

[0173] 利用市场上常见的核酸提取试剂盒分别提取全血样品(1至10)、血清样品(11至20)、石蜡组织样品(21至30)的 DNA,并分别做上相应的标记。

[0174] 二、目标区域的扩增

[0175] 同一样品的 EGFR 基因外显子 19、21 的扩增在同一反应体系中进行,反应体系如下:

[0176] 19F(10 μ M) 2 μ L;

[0177] 19R(10 μ M) 2 μ L;

[0178] 21F(10 μ M) 2 μ L;

[0179] 21R(10 μ M) 2 μ L;

[0180] dNTP(各 2.5mM) 4 μ L;

[0181] 待测样品 DNA 20ng;

[0182] Ex Taq(5U/ μ L) 0.25 μ L;

[0183] 10 \times Ex Taq Buffer 5 μ L;

[0184] ddH₂O up to 50 μ L。

[0185] PCR 反应条件如下:

[0186] 95 $^{\circ}$ C 3min;

[0187] 94 $^{\circ}$ C 30s, 58 $^{\circ}$ C 30s, 72 $^{\circ}$ C 30s;重复 25 个循环;

[0188] 72 $^{\circ}$ C 7min。

[0189] 利用 PCR 清洁试剂盒,分别对各样品的扩增产物进行清洁,除去未扩增的引物和 dNTP,回收扩增产物。

[0190] 三、构建目标区域测序文库

[0191] 此步骤可有多种实施方案,本发明的一个实施例中,包括以下两个步骤:

[0192] 1. 扩增产物的片段化

[0193] 利用超声法进行片段化处理。具体操作为:将每个样品的 PCR 清洁回收后的扩增产物(约 50 μ L)加入至 400 μ L 的 TE buffer 中,在 430W 功率条件下超声 4s,间隔 20s,反复 5 次,得片段化产物。利用 1%琼脂糖凝胶对各样品的片段化产物进行分离纯化,切胶回收大小在 40bp 至 100bp 之间的片段化产物。

[0194] 2. 构建测序文库

[0195] 在构建目标区域测序文库之前,需对切胶回收的片段化产物分别进行末端修饰,以便于接头元件的连接。在本实施例中,对切胶回收的片段化产物的末端修饰包括磷酸化、末端补平和末端加 A 反应。

[0196] 具体实现如下:

[0197] 1) 磷酸化以及末端补平反应

[0198] 体系为:

[0199] 切胶回收的片段化产物 20 μ L(约 2000ng);

[0200] 10mM dNTP 1.5 μ L;

[0201] T4DNA 聚合酶 (5U/ μ L) 1 μ L;

[0202] Klenow DNA 聚合酶 0.1 μ L;

[0203] T4 多核苷酸激酶 (10U/ μ L) 0.5 μ L;

[0204] 10m MATP 1.5 μ L;

[0205] 10 \times T4 连接酶缓冲液 10 μ L;

[0206] 加 ddH₂O 至 100 μ L。

[0207] 反应条件为:20 $^{\circ}$ C 孵育 20min。反应结束后利用回收试剂盒进行纯化回收。

[0208] 2) 末端加 A 尾

[0209] 反应体系为:

[0210] 磷酸化以及末端补平后的回收产物 60 μ L(约 1000ng);

[0211] Klenow 缓冲液 (NEB Buffer2) 10 μ L;

[0212] 10mM dATP 2 μ L;

[0213] Klenow 酶 (3' to 5' exo minus, 10U/ μ L) 1 μ L;

[0214] 加 ddH₂O 至 100 μ L。

[0215] 反应条件为:37 $^{\circ}$ C 孵育 30min。反应结束后利用纯化试剂盒纯化回收。

[0216] 3) 连接接头 1

[0217] 本实施例中,采用如图 6 所示的 T 末端分叉接头作为接头 1,同一样品的所使用的 T 末端分叉接头相同,不同样品的 T 末端分叉接头不同,区别在于标签序列不同,不同样品对应的标签序列如下表所示。

[0218]

样本编号	标签序列 (5'→3')	样本编号	标签序列 (5'→3')	样本编号	标签序列 (5'→3')
1	AGCT	11	GACT	21	CGAT
2	AGTC	12	GATC	22	CGTA
3	ATGC	13	GTCA	23	CTGA
4	ATCG	14	GTAC	24	CTAG
5	ACTG	15	GCAT	25	CAGT
6	ACGT	16	GCTA	26	CATG
7	AAGG	17	ACAC	27	TGTC
8	AACC	18	AGAG	28	TCTG
9	TTGG	19	TCTC	29	AGAC
10	TTCC	20	TGTG	30	ACAG

[0219] 在 T4 连接酶的作用下,上述 T 末端分叉接头分别与加 A 尾后纯化回收的产物进行连接,形成带接头 1 的片段。连接体系为:

[0220] 加 A 尾后纯化回收的产物 50 μ L (约 500ng);
 [0221] 接头 1 2 μ L (约 3000ng);
 [0222] 10mM ATP 5 μ L;
 [0223] T4DNA 连接酶 (10U/ μ L) 1 μ L;
 [0224] 10 \times T4 连接酶缓冲液 10 μ L;

[0225] 加 ddH₂O 至 100 μ L。

[0226] 反应条件为:16 $^{\circ}$ C 孵育 4h 以上。反应结束后利用纯化试剂盒纯化回收。

[0227] 4)PCR 扩增带接头 1 的片段

[0228] 扩增体系为:

[0229] 带接头 1 的片段 约 300ng;
 [0230] 10 \times Ex Taq Buffer 100 μ L;
 [0231] Primer F (100 μ M, SEQ ID NO :11) 10 μ L;
 [0232] Primer R (100 μ M, SEQ ID NO :12) 10 μ L;
 [0233] Ex Taq (5U/ μ L) 7.5 μ L;

[0234] ddH₂O up to 1000 μ L。

[0235] PCR 反应条件如下:

[0236] 95 $^{\circ}$ C 3min;

[0237] 94 $^{\circ}$ C 30s, 58 $^{\circ}$ C 30s, 72 $^{\circ}$ C 30s;重复 25 个循环;

[0238] 72 $^{\circ}$ C 7min。

[0239] 利用 PCR 清洁试剂盒,分别对各样品的扩增产物进行清洁,除去未扩增的引物和 dNTP,回收带接头 1 的片段的扩增产物。

[0240] 5)II s 型限制性内切酶酶切

[0241] 利用 Acu I 酶切回收后的带接头 1 的片段的扩增产物,反应体系如下:

- [0242] 回收后的带接头 1 的片段的扩增产物 60 μ L (3 ~ 5 μ g) ;
- [0243] 10 \times NEB Buffer 4 8 μ L ;
- [0244] Acu I (NEB) 2 μ L (10U) ;
- [0245] SAM (3. 2mM) 1 μ L (终浓度 50 μ M) ;
- [0246] ddH₂O up to 80 μ L。
- [0247] 反应条件 :37 $^{\circ}$ C 孵育 1h。反应结束后利用纯化试剂盒纯化回收酶切产物。
- [0248] 6) 连接接头 2
- [0249] 利用如图 7 所示的突出末端分叉接头作为接头 2, 与回收的酶切产物进行连接, 得测序文库, 连接体系如下 :
- [0250] 回收的酶切产物 2 μ g ;
- [0251] 接头 2 1 μ L (10pM) ;
- [0252] 10T4DNA 连接酶 (3U/ μ L) 1 μ L ;
- [0253] 10 \times T4 连接酶缓冲液 2 μ L ;
- [0254] 加 ddH₂O 至 100 μ L。
- [0255] 连接条件, 14 $^{\circ}$ C 孵育 2h。反应结束后利用纯化试剂盒纯化回收, 然后变性形成单链, 得测序文库。
- [0256] 四、对目标区域测序文库进行单分子扩增
- [0257] 测定 (三) 步骤所获得 30 个目标区域测序文库各自的浓度, 然后按同等浓度混合, 然后进行单分子扩增, 得单分子扩增产物, 所述单分子扩增的方法可采用 EPCR 或桥式 PCR。
- [0258] 优选为 EPCR 扩增, 单分子扩增引物优选为 :SEQ ID NO :5 和 SEQ ID NO :6。
- [0259] 五、对单分子扩增产物进行高通量测序
- [0260] 所述测序方法可采用基于合成酶的合成测序法, 也可采用基于连接酶的连接测序法。对测序结果进行生物信息学分析, 即可得到上述 30 个样品的 EGFR 基因的外显子 19 和 21 的序列信息。分析序列结果可知, 除样品 23 的 EGFR 基因的外显子 19 发生突变外 (序列信息见 SEQ ID NO :8, 突变率为 6. 2%), 其它样品的各外显子均为野生型。
- [0261] 应当说明的是, 本实施例只是本发明的一个具体实施例, 对本发明无任何限定作用, 例如 EGFR 基因的特异性引物可用其他的经过合理设计的引物进行替换 ; 除此之外, 本实施例中的多个步骤均可参照前述的方法进行替换, 在此不再赘述。
- [0262] 针对本发明的对目标区域进行测序的方法的灵敏度, 本发明采用下述实施例进行验证。
- [0263] 通过常规设计, 构建分别含有 EGFR 基因外显子 19 野生型序列 (SEQ IDNO :7) 和突变型序列 (SEQ ID NO :8) 的质粒 ; 含有 EGFR 基因外显子 21 野生型序列 (SEQ ID NO :9) 和突变型序列 (SEQ ID NO :10) 的质粒。
- [0264] 然后配制突变率分别为 20%、10%、5%、3%、1%、0% 的质粒混合液。以含有 1000 个拷贝质粒的上述质粒混合液为模板, 然后按参照上述实施例的方法进行检测。检测结果如下表所示 :
- [0265]

EGFR E19						
实际突变率(%)	20.0	10.0	5.0	3.0	1.0	0.0
检测突变率(%)	19.7	10.2	4.7	1.2	0.0	0.0
EGFR E21						
实际突变率(%)	20.0	10.0	5.0	3.0	1.0	0.0
检测突变率(%)	20.3	10.4	4.5	1.0	0.0	0.0

[0266] 结果显示,本发明的对目标区域进行测序的方法的最低检测限为 3%左右,在 5% 及以上的检测结果与实际情况基本一致。

[0267] 一种对目标区域进行测序的试剂盒,包括:

[0268] 特异性引物,用于对待测样品中的多个目标区域进行扩增;

[0269] 接头元件,用于与扩增产物结合构建测序文库。

[0270] 该试剂盒用于目标区域测序,能同时对样本的多个目标区域进行深度测序,准确得出这些目标区域的序列信息,包括已知突变和位置突变的各突变位点的变异情况,精确得出各样品的各突变位点发生变异的频率,检测灵敏度高,能用于目标区域的大规模人群筛查,确定该目标区域的各种基因型的比例。

[0271] 其中,所述接头元件用于构建目标区域测序文库,包括至少一种或多种接头。

[0272] 其中,所述接头元件可采用多种形式,如:平末端接头、突出末端接头、带茎环结构的接头、分叉接头中的至少一种。

[0273] 其中,所述接头可被甲基化或生物素化,或同时被生物素化和甲基化。

[0274] 其中,所述接头元件中的至少一个接头包含有第一标签序列,用于在文库构建过程中,对不同待测样品的测序文库进行标记。所述第一标签序列优选为带有特定碱基序列的核酸分子,其碱基数不限,优选为 3~20,更优选为 4~20。

[0275] 其中,所述的特异性引物与目标区域互补或部分互补。

[0276] 进一步的,每个目标区域对应的特异性引物中,至少一条引物与该目标区域部分互补,该引物的 5' 端包含有第二标签序列,用于在扩增目标区域过程中,对不同待测样品的目标区域扩增产物进行标记。所述第二标签序列优选为带有特定碱基序列的核酸分子,其碱基长度不限。

[0277] 更进一步的,所述第二标签序列的碱基数为 3~20,更优选为 4~10。

[0278] 其中,所述试剂盒还可以包括 PCR 扩增试剂,所述 PCR 扩增试剂包括 PCR 酶、dNTP、PCR 缓冲液、 Mg^{2+} 溶液。

[0279] 其中,所述试剂盒还可包括用于识别接头元件上的酶切识别位点的酶。

[0280] 应当说明的是,通过本发明的方法或试剂盒检测得到的基因序列信息,可用于各种科学研究,包括但不限于人群序列分析、基因功能研究、蛋白功能研究。例如,结合后续进一步的分子生物学试验、临床试验、临床观察及综合数据的统计分析等,从而实现多种科研目的,包括但不限于:某特定人群的某些特定基因的基因型分布、各基因的突变类型与相关生理活动的关系等。

[0281] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

[0001]

说明书核苷酸和氨基酸序列表替换页

SEQUENCE LISTING

- <110> 深圳华因康基因科技有限公司
 <120> 一种对目标区域进行测序的方法及试剂盒
 <160> 16
 <170> PatentIn version 3.3
- <210> 1
 <211> 21
 <212> DNA
 <213> 人工序列
- <400> 1
 ccttaggtgc ggctccacag c 21
- <210> 2
 <211> 22
 <212> DNA
 <213> 人工序列
- <400> 2
 catttaggat gtggagatga gc 22
- <210> 3
 <211> 22
 <212> DNA
 <213> 人工序列
- <400> 3
 cagccataag tcctcgacgt gg 22
- <210> 4
 <211> 24
 <212> DNA
 <213> 人工序列

[0002]

说明书核苷酸和氨基酸序列表替换页

<400> 4	
catcctcccc tgcattggtt aaac	24
<210> 5	
<211> 26	
<212> DNA	
<213> 人工序列	
<400> 5	
tttttcgtc gccaaatgcg atctct	26
<210> 6	
<211> 26	
<212> DNA	
<213> 人工序列	
<400> 6	
tttttagct acgcagcgtc cgctag	26
<210> 7	
<211> 349	
<212> DNA	
<213> 人工序列	
<400> 7	
ccttaggtgc ggctccacag cccagtgct cctcacctc ggggtgcatc gctggtaaca	60
tccaccaga tcaactggca gcatgtggca ccatccaca attgccagt aacgtcttc	120
ttctctctct gcataggga ctctggatcc cagaaggtga gaaagttaa attcccgtc	180
ctatcaagga ataaagagaa gcaacatctc cgaaagccaa caaggaaatc ctcgatgtga	240
gtttctgctt tgctgtgtgg gggccatgg ctctgaacct caggcccacc tttctcatg	300
tctggcagct gctctgctct agaccctgct catctccaca tctaaatg	349
<210> 8	
<211> 334	

[0003]

说明书核苷酸和氨基酸序列表替换页

<212> DNA

<213> 人工序列

<400> 8

ccttaggtgc ggctccacag cccagtgtc cctcaccctc ggggtgcatc gctggtaaca	60
tccaccaga tcactgggca gcatgtggca ccatctcaca attgccagtt aacgtcttcc	120
ttctctctct gtcataggga cctggatcc cagaaggtga gaaagttaa atcecgteg	180
ctatcaaac atctccgaaa gccacaagg gaatcctcga tgtgagttc tgctttgctg	240
tgtgggggtc catggetctg aacctcagge ccaccttfc tcatgtctgg cagetgctct	300
gctctagacc ctgctcatct ccacatccta aatg	334

<210> 9

<211> 374

<212> DNA

<213> 人工序列

<400> 9

cagccataag tctcgcact ggagaggctc agagcctggc atgaacatga ccctgaattc	60
ggatgcagag cttcttccca tgatgatctg tcctcaccag cagggcttct tctgtttcag	120
ggcatgaact acttggagga ccgtcgcctg gtgcaccgcg acctggcagc caggaacgta	180
ctggtgaaaa caccgcagca tgcaagatc acagattttg ggctggccaa actgctgggt	240
gcggaagaga aagaatacca tgcagaagga ggcaaagtaa ggaggtggct ttaggtcagc	300
cagcatttct ctgacaccag ggaccaggct gccttcccac tagctgtatt gtttaacaca	360
tgcaggggag gatg	374

<210> 10

<211> 374

<212> DNA

<213> 人工序列

<400> 10

cagccataag tctcgcact ggagaggctc agagcctggc atgaacatga ccctgaattc	60
ggatgcagag cttcttccca tgatgatctg tcctcaccag cagggcttct tctgtttcag	120
ggcatgaact acttggagga ccgtcgcctg gtgcaccgcg acctggcagc caggaacgta	180
ctggtgaaaa caccgcagca tgcaagatc acagattttg ggctggccaa actgctgggt	240

[0004]

说明书核苷酸和氨基酸序列表替换页

gcggaagaga aagaatacca tgcagaagga ggcaaagtaa ggaggtggct ttaggtcagc 300
 cagcatttcc ctgacaccag ggaccaggct gccttcccac tagctgtatt gttaacaca 360
 tgcagggggag gatg 374

<210> 11
 <211> 20
 <212> DNA
 <213> 人工序列

<400> 11
 ctagcgtacg ctgcgtagct 20

<210> 12
 <211> 20
 <212> DNA
 <213> 人工序列

<400> 12
 cgtcgcaaaa tgcgatctct 20

<210> 13
 <211> 31
 <212> DNA
 <213> 人工序列

<220>
 <221> misc_feature
 <222> (21)..(24)
 <223> n is a, c, g or t

<400> 13
 ctagcgtacg ctgcgtagct nnnnctgaag t 31

<210> 14
 <211> 30

[0005]

说明书核苷酸和氨基酸序列表替换页

<212> DNA

<213> 人工序列

<220>

<221> misc_feature

<222> (7)..(10)

<223> n is a, c, g or t

<400> 14

cttcagnnnn agagatcgca ttggcgacg

30

<210> 15

<211> 35

<212> DNA

<213> 人工序列

<220>

<221> misc_feature

<222> (25)..(27)

<223> n is a, c, g or t

<220>

<221> misc_feature

<222> (34)..(35)

<223> x is a, c, g or t

<400> 15

cgcttcctg cagtcttat gggcnmtcc gacxx

35

<210> 16

<211> 29

<212> DNA

<213> 人工序列

<220>

[0006]

说明书核苷酸和氨基酸序列表替换页

<221> misc_feature

<222> (7)..(9)

<223> n is a, c, g or t

<400> 16

gtcggannna gagaatgagg aaccgggg

29

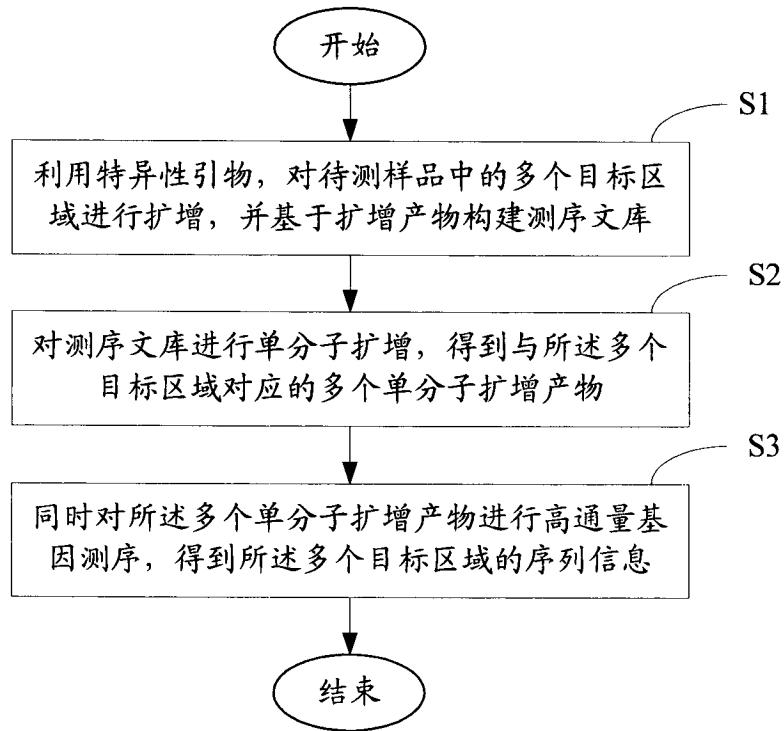


图 1

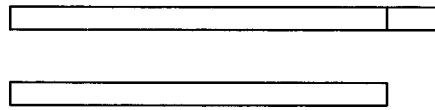


图 2



图 3

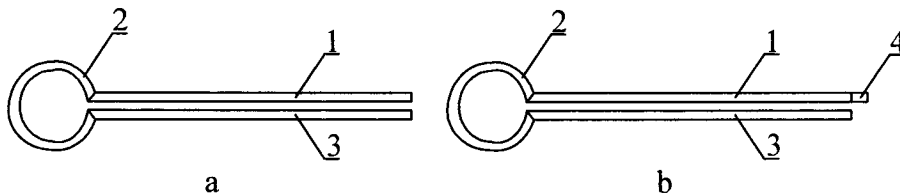


图 4

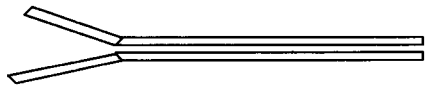


图 5

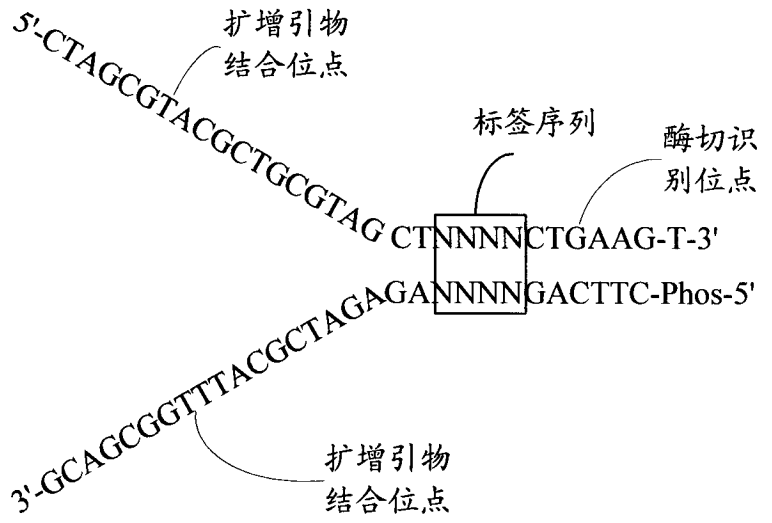


图 6

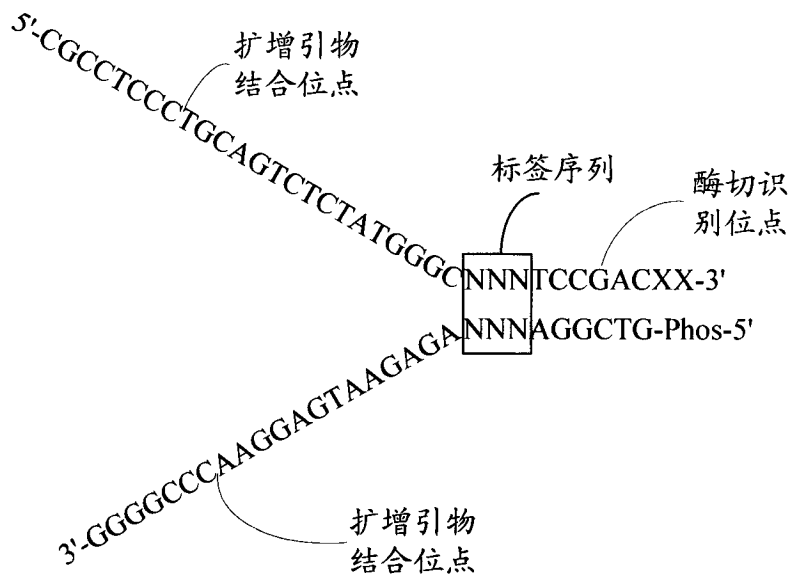


图 7

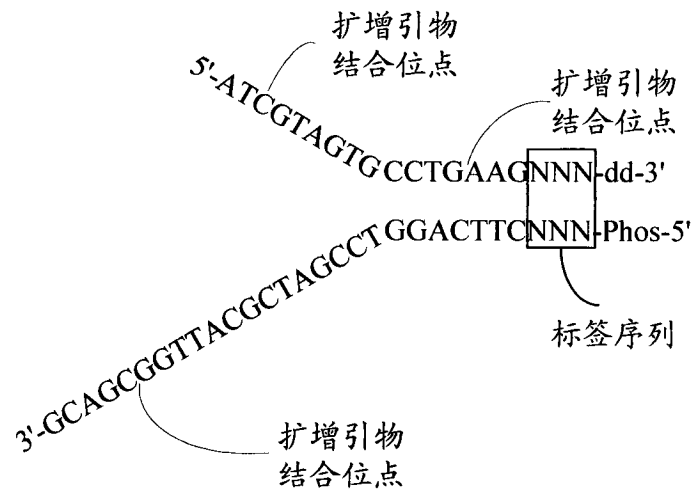


图 8

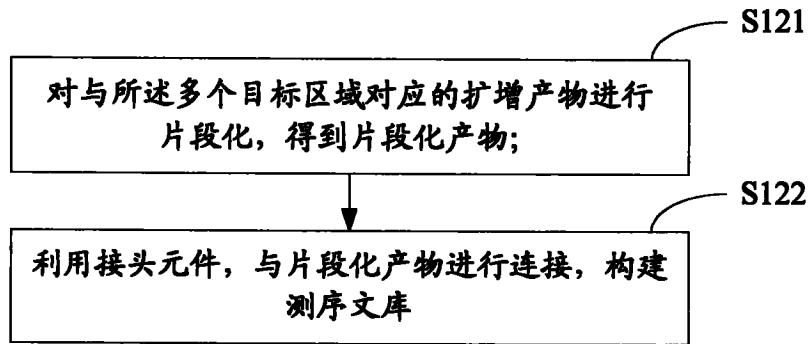


图 9

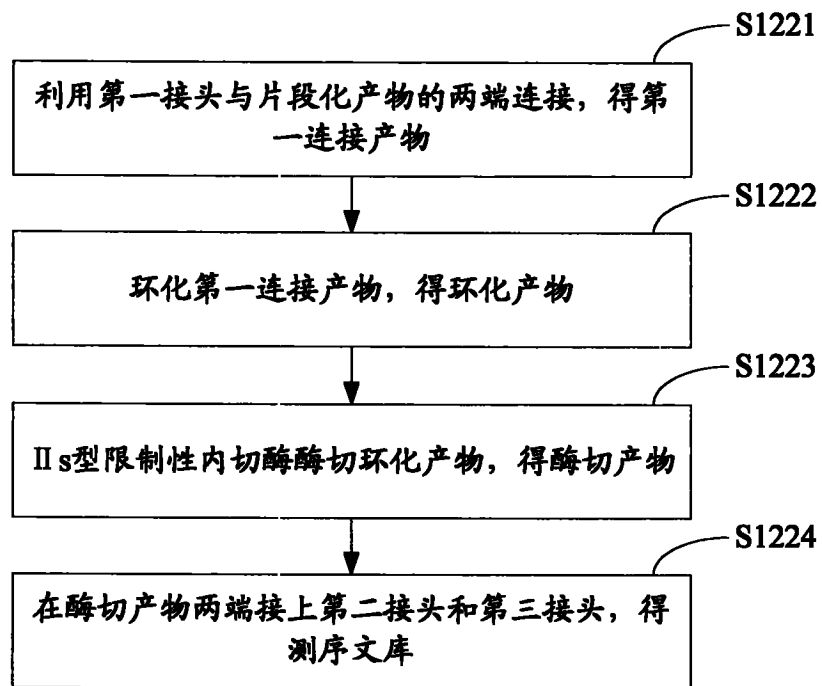


图 10

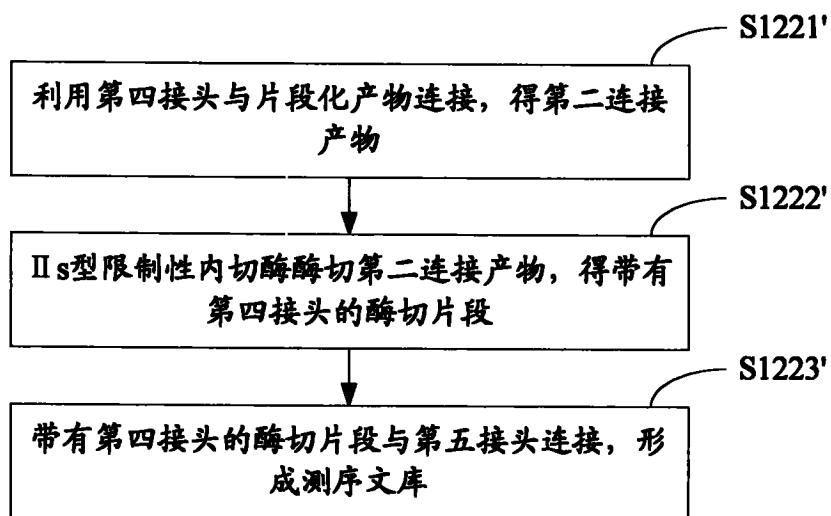


图 11