

## (12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织  
国际局

(43) 国际公布日  
2020年10月1日 (01.10.2020)



(10) 国际公布号  
**WO 2020/191521 A1**

- (51) 国际专利分类号:  
*C12N 15/11* (2006.01) *C12Q 1/6809* (2018.01)  
*C40B 50/06* (2006.01) *C12Q 1/6886* (2018.01)  
*C12Q 1/6806* (2018.01)
- (21) 国际申请号: PCT/CN2019/079235
- (22) 国际申请日: 2019年3月22日 (22.03.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 深圳华大智造科技有限公司 (MGI TECH CO., LTD.) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。
- (72) 发明人: 杨林(YANG, Lin); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 夏军(XIA, Jun); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 张艳艳(ZHANG, Yanyan); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 吕硕(LV, Shuo); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 陈芳(CHEN, Fang); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 梅智颖(MEI, Zhiying); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 陈大洋(CHEN, Dayang); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 蒋慧(JIANG, Hui); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。
- (74) 代理人: 北京清亦华知识产权代理事务所(普通合伙) (TSINGYIHUA INTELLECTUAL PROPERTY LLC); 中国北京市海淀区清华园清华大学照澜院商业楼301室, Beijing 100084 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

### 本国际公布:

- 包括国际检索报告(条约第21条(3))。
- 包括说明书序列表部分(细则5.2(a))。

(54) **Title:** NUCLEOTIDE SEQUENCE, AND METHOD FOR CONSTRUCTING RNA TARGET AREA SEQUENCING LIBRARY AND APPLICATION THEREOF

(54) 发明名称: 核酸序列、RNA目标区域测序文库的构建方法及应用

(57) **Abstract:** Provided are a nucleotide sequence, and a method for constructing an RNA target area sequencing library and an application thereof. The nucleotide sequence comprises a first general sequencing sequence located at terminus 5' and a random primer sequence located at terminus 3' which are connected to each other. The random primer sequence contains 6-10 random nucleotides. The sequence is used to perform random reverse transcription on RNA to obtain a first chain of cDNA, and then the first general sequencing sequence on the nucleotide sequence and upstream and downstream specific primers are used to perform two-step nested PCR on the first chain of cDNA obtained after the reverse transcription, to obtain a high-throughput targeted sequencing library.

(57) 摘要: 提供了一种核酸序列以及RNA目标区域测序文库的构建方法及应用。该核酸序列包含相连的位于5'端的第一测序通用序列和位于3'端的随机引物序列, 随机引物序列含有6~10个随机核苷酸。利用该序列对RNA进行随机的逆转录, 得到cDNA的第一链, 然后通过该核酸序列上的第一测序通用序列以及上下游特异性引物对逆转录得到的cDNA第一条链进行2步巢式PCR, 得到高通量靶向测序文库。



WO 2020/191521 A1

## 核酸序列、RNA 目标区域测序文库的构建方法及应用

### 技术领域

本发明涉及基因检测领域，具体涉及一种核酸序列以及 RNA 目标区域测序文库的构建方法及应用。

5

### 背景技术

RNA测序（RNAseq）技术在基因表达研究中得到了广泛的应用，可对表达基因进行无偏见的采样。相比定量PCR可以对更广泛的基因同步分析，相比芯片方法RNAseq测序结果准确性也更高。然而，真核细胞的表达谱具有转录本数量众多、可变剪切情况复杂，且表达量高低差异大的特点。RNAseq测序结果分散在整个基因组中，不同基因测序深度差异显著，对转录水平偏低的转录本往往测序深度不足，为进行特定基因的可变剪切体拼接及定量研究造成阻碍。

10

自20世纪80年代早期第一个融合基因(BCR/ABL1)被检测证实，融合基因已成为多个癌种的重要标志，例如：BCR/ABL1用于检测慢性骨髓性白血病；EML4/ALK用于检测恶性肺腺癌；TMPRSS2/ERG用于检测前列腺癌，等等。21世纪开始，随着靶向药物的兴起，癌症治疗药物，尤其非小细胞肺癌(NSCLC)的治疗药物，也开始使用融合基因所在的代谢通路。例如，2011年克唑替尼获得FDA许可用于ALK和ROS1基因融合的治疗；2015年艾乐替尼获得FDA许可用于ALK基因融合的治疗；索拉非尼和舒尼替尼获得RET基因融合的治疗许可。在恶性肿瘤精准检测及个性化治疗概念被提出的今天，为了避免药物禁忌和耐受引起无效治疗，耽误宝贵的治疗时间，对于融合基因的有效特异检测成为临床医师给药的重要依据。早期的融合基因检测是在传统组织活检的基础上，采用荧光原位杂交、染色体条带分析以及RT-PCR分析技术来确认患者的肿瘤细胞是否发生了特定基因的基因融合。但是这些早期的细胞学及分子生物学技术对于融合基因的检测存在以下缺陷：1)检测时间久；2)假阳性率偏高；3)检测技术要求高，需要熟练有经验的检测人员，不利于标准化；4)在融合基因基因型细胞在癌组织中含量少的环境下，无法有效检出；5)无法对多个融合基因进行检测。

15

20

25

随着高通量技术的发展，越来越多的人采用RNAseq对融合基因进行检测，但因为真核生物转录本数据众多，全转录本的RNAseq造成很多不必要的浪费，因此针对融合基因的检测还有待进一步改进。

### 发明内容

本发明旨在至少在一定程度上解决相关技术中的技术问题之一。为此，本发明的一个目的在于提出一种核酸序列以及利用这种核酸序列对 RNA 目标区域构建测序文库的方法及应用。

30

利用目标区域捕获技术，能够富集所感兴趣的基因或基因组区段，从而可以实现定向测序。将目标区域捕获技术与RNAseq结合，即对RNAseq的测序文库先进行序列捕获实验富集感兴趣的转录本，再进行测序，实现定向RNAseq或RNA捕获测序（CaptureSeq）。这一方法可以提高目标转录本的测序深度，进行灵敏的基因发现、有效的转录本组装和准确的基因表达定量。该技术结合应用目标区域捕获技术与RNAseq技术，可以称作RNA捕获测序技术。应用RNA捕获测序技术对样本中特定基因的表达进行研究。该方法既RNAseq相比其他表达谱研究方法的优点，又可以进行有目的性的、多样本的研究。尤其在对于表达水平中低等的转录本研究中，可以更好地拼接全长转录本、发现新基因以及定量分析。

35

本发明用带有一段5'端带有测序通用序列，3'端带有6-10nt随机引物序列的核酸序列对RNA样本进行逆转录，获得cDNA的一条链，然后对所获得的cDNA的一条链进行巢氏PCR多重扩增，所获得的多重PCR产物中带有完整的测序接头，可以直接进行高通量测序。从而通过测序获得RNA样本的信息。根

据需要，这些RNA样本可以是生物样品中的总的RNA、mRNA、以及cfRNA（血浆游离DNA）等。

具体而言，本发明提供了如下技术方案：

根据本发明的第一方面，本发明提供了一种核酸序列，包含：第一测序通用序列和随机序列，所述第一测序通用序列和所述随机引物序列相连，所述第一测序通用序列位于所述核酸序列的 5' 端，所述  
5 随机引物序列位于所述核酸序列的 3' 端，其中，所述随机引物序列含有 6~10 个随机核苷酸。本发明提供的核酸序列至少包括两部分，一部分是位于 5' 端的第一测序通用序列，另一部分是位于 3' 端的随机引物序列，其中第一测序通用序列和随机引物序列可以通过 3' -5' 磷酸二酯键直接相连。第一测序通用序列可以是任何一种测序平台的任何一种测序接头序列（部分或者全部序列），也可以是任意一种已知的固定序列，用于后续的 PCR 扩增过程中。随机引物序列含有 6~10 个随机核苷酸，例如可以是 8 个随机  
10 核苷酸。随机核苷酸是指核苷酸上的碱基可以是 A、T、C、G 中的任意一种。该随机引物序列用于以 RNA 样本为模板，用于与 RNA 样本的结合。本发明提供的核酸序列可以用作逆转录反应的引物，即以 RNA 样本为模板，通过逆转录获得 cDNA 的一条链，获得的产物用作后续的 PCR 扩增。

根据本发明的实施例，以上所述核酸序列可以进一步包含如下附加技术特征：

在本发明的一些实施例中，所述随机引物序列含有 8 个随机核苷酸。

15 在本发明的一些实施例中，所述核酸序列进一步包含：第一分子标签序列，所述第一分子标签序列位于所述第一测序通用序列和所述随机引物序列之间；互补序列，所述互补序列位于所述第一分子标签序列和所述随机引物序列之间，所述互补序列与所述第一测序通用序列的部分序列互补。在所述第一测序通用序列和随机引物序列之间添加有分子标签序列，通过大量不同的分子标签序列对原始的 RNA 模板个数进行计数，通过后续对分子标签的统计来追溯原始模板的个数，从而可以实现对于 RNA 模板的  
20 定量研究。同时，为了保证分子标签序列不和随机引物序列互相发生作用，可以对该分子标签序列进行“包裹”，即在随机引物序列和分子标签序列之间插入一段互补序列，该互补序列能够和第一测序通用序列上的 5' 末端的部分序列反向互补。从而使得核酸序列的 5' 端的序列能够和该互补序列退火形成颈环结构，从而保证分子标签被包裹在内部，防止其与随机引物序列发生非特异性的结合或者扩增。

在本发明的一些实施例中，所述第一分子标签序列含有 8~20 个随机核苷酸。第一分子标签序列含有  
25 8-20bp 的随机碱基，可以产生  $4^{8-20}$  不一样的分子标签序列，由于开始 RNA 模板数量一定，当分子标签序列远远大于 RNA 模板数时，每一种 RNA 模板都会加上了一种唯一的分子标签，通过大量不同的分子标签对原始的 RNA 模板个数进行标记，通过后续对分子标签种类的多少统计来追溯原始模板的个数。

在本发明的一些实施例中，所述互补序列与所述第一测序通用序列的 5' 端的部分序列互补。

根据本发明的第二方面，本发明提供了一种 RNA 目标区域测序文库的构建方法，包括：基于 RNA  
30 样本，利用逆转录引物和逆转录酶进行逆转录反应，获得 cDNA 的第一条链，其中所述逆转录引物为本发明第一方面任一实施例所述的核酸序列；基于所述 cDNA 的第一条链，利用上游特异性引物和第一通用引物进行第一轮 PCR 扩增，获得第一扩增产物，所述第一通用引物与所述第一测序通用序列的部分序列重叠，所述上游特异性引物的结合位点位于所述目标区域的上游；基于所述第一扩增产物，利用下游特异性引物、第二通用引物和标签引物进行第二轮巢氏 PCR 扩增，获得所述目标区域测序文库，所述下  
35 游特异性引物的 5' 端含有部分或者全部第二通用测序序列，所述第二通用引物的 5' 端含有磷酸基团，所述第二通用引物和所述下游特异性引物的部分或者全部序列重叠，所述下游特异性引物的结合位点位于所述目标区域的上游，所述标签引物与所述第一通用引物的部分序列重叠；所述下游特异性引物 3' 端的结合位点位于所述上游特异性引物的下游。

核酸序列上的随机引物序列能够随机锚定在所有 RNA 分子上，从而合成 cDNA 的一条链，然后利用

测序通用序列以及上游特异性引物和下游特异性引物可以富集目标区域。表现为：利用多个上游特异性引物和第一通用引物，以cDNA的一条链为模板，进行第一轮PCR扩增，利用多个下游特异性引物、第二通用引物和标签引物进行第二轮巢式PCR扩增，从而能够实现对多个目标区域进行捕获并通过高通量测序技术对目标区域进行分析，一次完成多个基因的检测分析。该方法能够尽可能多的利用cDNA的模板，

5 从而提高了原始RNA分子的利用效率。而且该方法可以对融合基因进行检测，也可以对未知基因的融合情况进行检测。

根据本发明的实施例，以上所述的构建方法可以进一步附加如下技术特征：

在本发明的一些实施例中，所述RNA为cfRNA（血浆游离DNA）、高度降解的RNA或者长度为200~300bp的RNA。

10 在本发明的一些实施例中，所述长度为200~300bp的RNA通过利用Mg<sup>2+</sup>对长片段RNA样本孵育获得。对于总的RNA或者mRNA样本，可以利用Mg<sup>2+</sup>孵育，实现RNA的片段化，从而用于RNA目标区域测序文库的构建。

在本发明的一些实施例中，所述标签引物中含有第二分子标签序列，用于区分不同样本。所述标签引物的3'端和所述第一通用引物的5'端部分序列重叠（或者也可以称为一致，在本文中当表示两个核

15 酸序列对应核酸重叠时，即表示的两个核酸序列对应酸碱基一致），所述标签引物中间有一段分子标签序列（为了同上述核酸序列中的第一分子标签序列相区分，这里称为第二分子标签序列）。

在本发明的一些实施例中，所述第二分子标签序列为8~12个核苷酸。这些分子标签序列是一些测序平台上的固定的序列，每个测序平台有自己一套分子标签序列，例如可以是barcode序列或者index序列，这些分子标签序列的长度可以为10个核苷酸。

20 在本发明的一些实施例中，所述构建方法进一步包括：可以在设计上游特异性引物或者下游特异性引物的过程中，可以在基因组的至少一个内含子上设计，这里含有内含子的引物，采用对RNA样本进行扩增的相同的方式来使用，作为质检引物，来质检RNA样本是否在扩增过程中被DNA污染。这些用作质检引物的上游特异性引物或者下游特异性引物可以针对目标区域设计，也可以针对非目标区域设计。

根据本发明的第三方面，本发明提供了一种对目标基因变异检测的方法，包括：

25 采用本发明第二方面任一实施例所述的构建方法获得目标区域文库；

基于所述目标区域文库，进行测序获得测序数据；

将所述测序数据与参考基因组进行比对，确定所述目标基因的变异结果。

在本发明的一些实施例中，所述参考基因组为人类基因组hg19。

在本发明的一些实施例中，所述变异包括基因融合、SNV、基因缺失、基因插入突变中的至少一种。

30 在本发明的一些实施例中，所述变异为基因融合。

在本发明的一些实施例中，所述基因融合为EML4-ALK基因融合。

根据本发明的第四方面，本发明提供了一种EML4-ALK融合基因的检测方法，包括：基于RNA样本，利用逆转录引物和逆转录酶进行逆转录反应，获得cDNA的第一条链，其中所述逆转录引物为SEQ ID NO:1或者SEQ ID NO:2中的一种；基于所述cDNA的第一条链，利用上游特异性引物和第一通用引物进行第一轮PCR扩增，获得第一扩增产物，所述第一通用引物为SEQ ID NO:3，所述上游特异性引物为选自SEQ ID NO:6~SEQ ID NO:30中的至少一种；基于所述第一扩增产物，利用下游特异性引物、第二通用引物和标签引物进行第二轮巢式PCR扩增，获得所述目标区域测序文库，所述第二通用引物为SEQ ID NO:4，所述标签引物为SEQ ID NO:5，所述下游特异性引物为SEQ ID NO:31~SEQ ID NO:55中的至少一种；将所述目标区域测序文库进行测序，获得测序数据，基于所述测序数据与参考基因组的对

35

比结果，确定所述 EML4-ALK 基因的融合结果。

## 附图说明

图 1 是根据本发明的实施例提供的核酸序列的结构示意图。

5 图 2 是根据本发明的实施例提供的带有分子标签序列和互补序列的核酸序列的结构示意图。

图 3 是根据本发明的实施例提供的利用一些核酸序列进行 RNA 目标区域测序文库构建的方法的实验流程图。

图 4 是根据本发明的实施例提供的利用一些核酸序列进行进行 RNA 目标区域测序文库构建的方法的实验流程图。

10 图 5 是根据本发明的实施例提供的对高度降解的 RNA 或者 cfRNA 构建测序文库的实验流程图。

图 6 是根据本发明的实施例提供的针对融合基因检测的引物设计示意图。

## 具体实施方式

下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，旨在用于解释本发明，而不能理解为对本发明的限制。

同时，为了方便本领域技术人员的理解，对本发明的某些术语进行解释和说明，需要说明的是，这些解释和说明，仅用来帮助对于本发明技术方案的理解，而不应当看做是对本发明保护范围的限制。

本文中，如无特殊说明，当提到两个核酸序列相连时，是指通过 3'-5' 磷酸二酯键相连。

20 本文中，“上游特异性引物”是指结合位点位于目标区域上游的特异性引物。

本文中，“下游特异性引物”也是指结合位点位于目标区域上游的特异性引物，只是相比较于上游特异性引物而言，下游特异性引物的结合位点位于上游特异性引物的下游。

需要说明的是，下游特异性引物的结合位点位于上游特异性引物的下游，不仅仅包括：下游特异性引物的所有结合位点均位于上游特异性引物的下游的情况，还包括：下游特异性引物的部分结合位点和上游特异性引物的部分结合位点重叠的情况，只要下游特异性引物的最末端结合位点位于上游特异性引物的最末端结合位点的下游即可。

RNA 捕获测序技术有天然的优势，其能够有目的性的捕获感兴趣的 RNA 片段进行测序分析，为疾病的诊断提供依据。测序后的分析步骤包括比对、拼接、去除非目标序列、新基因（外显子）发现、定量等。

30 为此，根据本发明的一个方面，本发明提供了一种核酸序列。在至少一些实施方式中，本发明提供的核酸序列为常规引物结构，该核酸序列包括位于 3' 端的 6-10nt 随机引物序列和位于 5' 端的第一测序通用序列，如图 1 所示。利用 3' 端的为随机引物序列能够与 RNA 进行随机互补配对，从而以提取好的总 RNA 或者 mRNA 为模板，实现 cDNA 的第一链合成。在至少一些实施方式中，3' 端的随机序列的长度 6-10nt，优先采用 8nt。5' 端第一测序通用序列既可以是任何一种测序平台的任何一种测序接头序列（部分或全部序列），也可以是一任意固定序列。利用该核酸序列，可以对 RNA 样本进行定性研究。

35 在至少一些实施方式中，本发明还提供了另一种核酸序列，该核酸序列与上述核酸序列不同的是，除了第一测序通用序列和随机引物序列之外，还带有第一分子标签序列和互补序列，如图 2 所示（其中为了方便观察第一测序通用序列 5' 端能够与该互补序列反向互补或者也可以称为反向配对的序列，将这段能够与互补序列配对的序列以与第一测序通用序列不同的线条示出，图 3 也采用相同的策略）。该

核酸序列 3' 端为 6-10nt 的随机引物序列, 5' 端除第一测序通用序列外还有一段 8-20nt 随机的第一分子标签序列。同时为了保证分子标签序列不和模板引物之间的产生互相作用, 在设计上对该分子标签序列进行“包裹”, 在随机引物序列和分子标签序列之间插入一段互补序列, 该互补序列和 5' 末端的第一测序通用序列的部分序列反向互补。正常情况下 5' 端的序列和该互补序列能够退火形成颈环结构, 保证分子标签被包裹在内部而防止进行非特异性的结合或者扩增。利用该核酸序列可以对 RNA 样本进行定性和准确的定量研究。

本发明还提供了一种 RNA 目标区域测序文库的构建方法, 包括: 基于 RNA 样本, 利用逆转录引物和逆转录酶进行逆转录反应, 获得 cDNA 的第一条链, 其中所述逆转录引物为本发明第一方面任一实施例所述的核酸序列; 基于所述 cDNA 的第一条链, 利用上游特异性引物和第一通用引物进行第一轮 PCR 扩增, 获得第一扩增产物, 所述第一通用引物与所述测序通用序列的部分序列重叠, 所述上游特异性引物的结合位点位于所述目标区域的上游; 基于所述第一扩增产物, 利用下游特异性引物、标签引物和第二通用引物进行第二轮巢氏 PCR 扩增, 获得所述目标区域测序文库, 所述第二通用引物的 5' 端含有磷酸基团, 所述第二通用引物和所述下游特异性引物的部分序列重叠, 所述标签引物与所述第一通用引物的部分序列重叠, 所述下游特异性引物的结合位点位于所述目标区域的上游; 所述下游特异性引物的结合位点位于所述上游特异性引物的下游。在至少一些实施方式中, 所述标签引物的 3 端和第一通用引物的 5' 端部分序列重叠, 所述标签引物中间有一段分子标签序列 (称为第二分子标签序列), 用于区分不同的样本。

为了方便理解, 参照图3、图4或者图5所示, 对上述RNA目标区域测序文库的构建方法进行说明。利用上述核酸序列, 以RNA为模板, 进行逆转录获得cDNA的第一条链: 首先, 对总RNA或者mRNA用  $\text{mg}^{2+}$  孵育进行RNA片段化, 不同的  $\text{mg}^{2+}$  离子浓度和温度以及孵育时间得到不一样的片段大小的分子, 优先地打断到200-300bp。当然, 对于已经片段化的RNA或者高度降解的RNA如cfRNA, 则不需要进行RNA片段化, 可以直接以这些片段化的RNA或者高度降解的RNA如cfRNA为模板, 通过逆转录获得cDNA的第一条链。在至少一些实施方式中, 可以RNA为模板在逆转录引物、逆转录酶的作用下孵育30分钟得到cDNA的第一条链, 将得到的第一条链进行纯化, 将多余的引物和离子除去。本文中“cDNA的第一条链”指的是以RNA为模板, 经过逆转录合成的与RNA互补的那条链。该逆转录引物即上文所提到的核酸序列。

然后利用上游特异性引物和第一通用引物进行第一轮PCR扩增, 获得第一扩增产物。上游特异性引物是需要研究的目标区域, 例如目标基因或目标位点的上游区域的基因序列。以融合基因检测如EML4-ALK为例, 上游特异性引物应该在EML4上; 第一通用引物和逆转录引物的5' 端的第一测序通用序列的部分序列一致, 第一轮PCR得到的产物上带有上游特异性引物序列和第一通用序列。扩增得到的产物经过纯化去掉多余的上游特异性引物和离子。

然后用下游特异性引物、第二通用引物以及标签引物进行第二轮巢式扩增, 下游特异性引物由5' 端的部分测序接头序列 (该部分测序接头序列即与第二通用引物的一部分序列重叠或一致) 和基因特异性序列组成, 基因特异性序列在上游特异性引物的下游, 两者可以重叠或相隔。对于融合基因, 上下游特异性引物都设置在上游基因上 (如EML4-ALK, 上下游特异性引物序列都应该在EML4上)。其中, 标签引物用来区分不同的样本, 待到测序文库建好后, 可以等量混合进行上机测序。第二通用引物的5' 端含有磷酸基团, 通过第二通用引物可以使得经过第二轮扩增得到的产物的5' 端带有磷酸基团, 从而避免专门进行磷酸化的过高成本。同时, 下游特异性引物的结合位点位于上游特异性引物的下游, 在进行扩增的同时, 还是二轮特异性筛选, 使得扩增结果更加特异。

为防止DNA污染影响基因表达定量, 可以设计质检引物用于质检, 质检引物跨外显子和内含子或内

含子上设计。如图6所示。正常情况下，DNA质检引物无法扩增出产物。当污染了DNA时，能够扩增出产物。

5 利用上述 RNA 目标区域测序文库的构建方法，对 RNA 进行测序，然后通过参考基因组进行比对，从而可以获得 RNA 的信息。在构建测序文库的过程中，具有极高的模板利用效率，RNA 用带有随机引物序列的逆转录引物进行逆转录。逆转录后的 cDNA 直接用于巢式 PCR，该 PCR 所需要的模板长度小，引物占位低，极大的提高了模板的利用效率。而且采用单方向的巢式引物结合通用引物进行两轮 PCR 扩增，扩增得到的产物一端为固定端，一段为随机端，该方案可以对目标基因、目标位置上发生的任意未知融合突变进行检测。例如，对于未知融合，比如要研究 EML1 号外显子上发生的所有融合，那么在 EML1 号外显子上设计上下游特异性引物，由于本发明对 RNA 样本扩增的方式为一端是固定序列，另外一端的引物结合的序列是加在随机锚定到 RNA 上的第一测序通用序列上，因此可以检测和 EML1 号外显子发生任何随机融合的情况。

10 同时采用第一分子标签序列对逆转录后的 cDNA 进行标记，可以准确地对原始的 mRNA 分子的绝对量进行计数，并且通过分子标签结合信息分析算法可以对突变进行校正，提高对 RNA 突变检测的准确性。而且，整个 RNA 靶向建库实验流程只有三个步骤，逆转录、第一轮 PCR、第二轮 PCR，相对于常规的 RNAseq 以及基于探针杂交的捕获技术，其极大地简化了实验的流程。

15 下面将结合实施例对本发明的方案进行解释。本领域技术人员将会理解，下面的实施例仅用于说明本发明，而不应视为限定本发明的范围。实施例中未注明具体技术或条件的，按照本领域内的文献所描述的技术或条件或者按照产品说明书进行。所用试剂或仪器未注明生产厂商者，均为可以通过市购获得的常规产品。

20

**实施例1 EML4-ALK融合基因检测**

引物设计：在EML4的外显子上设计上下游引物共计20对，在内含子上设计一对上下游内参引物，引物序列如表2，形成一个EML4-ALK检测panel，该panel可以对发生在EML4基因上所有可能的融合基因进行检测；

25

用上述panel对阳性样本（指的是已经确定发生EML4-ALK融合的RNA标准品）和阴性样本（阴性标准品是指未发生EML4-ALK融合的正常人的RNA标准品）进行文库制备，每个重复4次，文库制备完成后进行高通量测序，得到的数据用于基因融合分析，逆转录酶采用TAKARA公司的 MMLV Reverse Transcriptase, GPR，货号639574。

表1 实施例中所用到的核酸序列

核酸	序列
逆转录引物	(SEQ ID NO:1) GACCGCTTGGCCTCCGACTTNNNNNNNN*
带有第一分子标签序列的逆转录引物	(SEQ ID NO:2) GCAGCCTAGCGACCGCTTGGCCTCCGACTTNNNNNNNN*GGTCGCTAGGCTGCNNNNNNNN*
第一通用引物	(SEQ ID NO:3) TTGTCTTCCTAAGACCGCTTGGCCTCCGACTT
第二通用引物	(SEQ ID NO:4) /5Phos/GAACGACATGGCTACGATCCGACTT
标签引物	(SEQ ID NO:5) TGTGAGCCAAGGAGTTGNNNNNNNNNN <sup>#</sup> TTGTCTTCCTAAGACCGCTTGGCCTCCGACTT

其中，\*N为随机碱基，#N为第二分子标签序列，该第二分子标签序列也可以称为样本标签序列。其中N为碱基A、T、C或者G。

表2 上游特异性引物

引物 (Primer) 编号	序列
EML4-E02-01-USP	(SEQ ID NO:6) CCGGCAGTCTCGATGATAGTA
EML4-E02-02-USP	(SEQ ID NO:7) ATGTTTTGAGGCGTCTTGCAATC
EML4-E03-USP	(SEQ ID NO:8) AAGTCATACCAGTGCTGTCTCAA
EML4-E04-USP	(SEQ ID NO:9) CAACCTCTCCAAATACACAGAC
EML4-E05-USP	(SEQ ID NO:10) GTCGAAAATACCTTCAACACCCAA
EML4-E07-USP	(SEQ ID NO:11) CTATGATGACATCAGAACGGAAGT
EML4-E08-USP	(SEQ ID NO:12) GAGAAGTCTCAGCGACTACCTG
EML4-E10-USP	(SEQ ID NO:13) GTCAGAGTGTGGGATTCTGTTAC
EML4-E11-USP	(SEQ ID NO:14) GATGACTCCAATGAGCATATGC
EML4-E12-USP	(SEQ ID NO:15) TACCATAATTACATGCGGTAAATCTCAT
EML4-E13-USP	(SEQ ID NO:16) GAGACTCAGGTGGAGTCATG
EML4-E14-USP	(SEQ ID NO:17) AGACAGAAAATAATTCTGTGGGATCA
EML4-E15-USP	(SEQ ID NO:18) CACGAAACTTTATTTTACGAGGAACAT
EML4-E16-USP	(SEQ ID NO:19) GTGTGCCTGTGGAAGTCAATG
EML4-E17-USP	(SEQ ID NO:20) GATTTTCATCCAAGTGGCACAGT
EML4-E18-USP	(SEQ ID NO:21) CACAGACGGGAATGAACAGCT
EML4-E19-USP	(SEQ ID NO:22) CTCTATGTAGTCTCTGAAAATGGAAGA
EML4-E20-USP	(SEQ ID NO:23) CCCAGACAACAAGTATATAATGTCTA
EML4-E21-USP	(SEQ ID NO:24) CACAATAGAAAGGTGATAGCTGTT
EML4-E22-01-USP	(SEQ ID NO:25) GTGGAACTTGTGGAAAAGTTATC
EML4-E22-02-USP	(SEQ ID NO:26) AGACAGCTGAAGAGGAAAGTAG
GAPDH-DNA-E1-USP	(SEQ ID NO:27) GACACCATGGGGAAGGTGAA
GAPDH-RNA-E2-USP	(SEQ ID NO:28) GTGGATATTGTTGCCATCAATGACC
GAPDH-RNA-E3-USP	(SEQ ID NO:29) TGAGAACGGGAAGCTTGTCA
GAPDH-RNA-E4-USP	(SEQ ID NO:30) TGAGTACGTCGTGGAGTCCA

5 上游引物池由10μM的上述引物等浓度混合而成。其中表2中编号代表：以EML4-E02-01-USP为例，该编号中EML4代表基因，E02代表第二号外显子，01代表第一号外显子的第一对引物，USP代表上游特异性引物等。相应地，表3中DSP代表下游特异性引物。

表3 下游特异性引物

引物 (Primer) 编号	序列
EML4-E02-01-DSP	(SEQ ID NO:31) CGACATGGCTACGATCCGACTTCTGCTGCAAGTACTTCTGATGTTCAA

	GAT
EML4-E02-02-DSP	(SEQ ID NO:32) CGACATGGCTACGATCCGACTTCTGAAGATCATGTGGCCTCAGT
EML4-E03-DSP	(SEQ ID NO:33) CGACATGGCTACGATCCGACTTAAGAAACTCTTTCATCTGCTGCTA
EML4-E04-DSP	(SEQ ID NO:34) CGACATGGCTACGATCCGACTTCAGAAAGCAAGAATGCTACTCC
EML4-E05-DSP	(SEQ ID NO:35) CGACATGGCTACGATCCGACTTCCAAAAGTTACCAAACTGCAGACA
EML4-E07-DSP	(SEQ ID NO:36) CGACATGGCTACGATCCGACTTCCCTGAGAAGCTCAAACCTGGAGT
EML4-E08-DSP	(SEQ ID NO:37) CGACATGGCTACGATCCGACTTGCCATACAGACTGTGTGAAATG
EML4-E10-DSP	(SEQ ID NO:38) CGACATGGCTACGATCCGACTTCTATCCACACTGCAGATTATTGGAC
EML4-E11-DSP	(SEQ ID NO:39) CGACATGGCTACGATCCGACTTTACTGTATGGGACTGGCAGAGG
EML4-E12-DSP	(SEQ ID NO:40) CGACATGGCTACGATCCGACTTTGGAGCGGCAATTCACCTAAC
EML4-E13-DSP	(SEQ ID NO:41) CGACATGGCTACGATCCGACTTTTATATGGAGCAAACTACTGTAGAG C
EML4-E14-DSP	(SEQ ID NO:42) CGACATGGCTACGATCCGACTTGATCTGAATCCTGAAAGAGAAATAGA G
EML4-E15-DSP	(SEQ ID NO:43) CGACATGGCTACGATCCGACTTGATGGCTTCAAATAGAAGTACAG
EML4-E16-DSP	(SEQ ID NO:44) CGACATGGCTACGATCCGACTTGAACACAGGCTGGAATGGAC
EML4-E17-DSP	(SEQ ID NO:45) CGACATGGCTACGATCCGACTTATAGGAACGCACTCAGGCAG
EML4-E18-DSP	(SEQ ID NO:46) CGACATGGCTACGATCCGACTTCTGTGATGCGCTACTCAATAG
EML4-E19-DSP	(SEQ ID NO:47) CGACATGGCTACGATCCGACTTTAGCAGATATGGAAGGTGCACT
EML4-E20-DSP	(SEQ ID NO:48) CGACATGGCTACGATCCGACTTCTCGGGAGACTATGAAATATTGTACT
EML4-E21-DSP	(SEQ ID NO:49) CGACATGGCTACGATCCGACTTCGATGACTTTTGTAAGTCCATCTG

EML4-E22-01-DSP	(SEQ ID NO:50) CGACATGGCTACGATCCGACTTTTTGCCTCAGAATGAGACTGTAG
EML4-E22-02-DSP	(SEQ ID NO:51) CGACATGGCTACGATCCGACTTAATAAGCAGTTCTCCCACACTTC
GAPDH-DNA-E1-DSP	(SEQ ID NO:52) CGACATGGCTACGATCCGACTTGAGTCAACGGGTGAGTTCGC
GAPDH-RNA-E2-DSP	(SEQ ID NO:53) CGACATGGCTACGATCCGACTTCCTTCATTGACCTCAACTACATG
GAPDH-RNA-E3-DSP	(SEQ ID NO:54) CGACATGGCTACGATCCGACTTCCCATCACCATCTTCCAGG
GAPDH-RNA-E4-DSP	(SEQ ID NO:55) CGACATGGCTACGATCCGACTTTGGCGTCTTCACCACCAT

下游引物池有10 $\mu$ M上述引物等浓度混合而成。其中各下游特异性引物的编号与上述上游特异性引物的编号的含义一致。

其测定包括如下步骤:

5

### 1、RNA反转

在200 $\mu$ lPCR管子配置如下反应体系，并在95摄氏度下反应10分钟，然后再冰上放置5分钟。

10 $\times$ 1st cDNA 合成缓冲液	1 $\mu$ l
RNA (1-1000ng)	1 $\mu$ l
N8随机引物 (100 $\mu$ M)	1 $\mu$ l
RNA酶抑制剂	1 $\mu$ l
DEPC水	6 $\mu$ l
共	10 $\mu$ l

### 2、在上述反应体系中加入如下试剂

上一步反应物	10 $\mu$ l
10 $\times$ 1st cDNA 合成缓冲液	1 $\mu$ l
M-MLV酶	1 $\mu$ l
RNA酶抑制剂	1 $\mu$ l
DEPC水	7 $\mu$ l
共	20 $\mu$ l

37度30分钟，45度10分钟，55度10分钟，65度10分钟

得到的产物加入30ul XPure beads进行纯化(Beckman公司agencourt AMPure XP磁珠, 货号A63881),

10 得到的产物溶于20ulTE中。

### 3、第一轮PCR

采用QIAGEN Multiplex PCR Kit, 货号Cat No./ID: 206143, 进行PCR

在200ulPCR管子配置如下试剂

上一步反应物	20 $\mu$ l
2X PCR反应酶	25 $\mu$ l

上游特异性引物池	2.5 $\mu$ l
第一通用引物	2.5 $\mu$ l
共	50 $\mu$ l

按照以下程序进行第一轮PCR

95度	2min	1个循环
95度	10s	15个循环
62度	2min	
72度	30s	
72度	5min	1个循环

得到的PCR产物中加入20 $\mu$ l的XP beads进行纯化（beackman公司agencourt AMPure XP磁珠，货号A63881），得到的产物溶解于20ulTE中；

#### 4、第二轮PCR

- 5 采用QIAGEN Multiplex PCR Kit，货号Cat No./ID: 206143，进行PCR  
在200ulPCR管子配置如下试剂

上一步反应物	20 $\mu$ l
2X PCR反应酶	25 $\mu$ l
下游特异性引物池	1 $\mu$ l
第一通用引物	2 $\mu$ l
标签引物	2 $\mu$ l
共	50 $\mu$ l

按照以下程序进行第一轮PCR

95度	2min	1个循环
95度	10s	15个循环
62度	2min	
72度	30s	
72度	5min	1个循环

得到的PCR产物中加入20ul的XP beads进行纯化（beackman公司agencourt AMPure XP磁珠，货号A63881），得到的产物溶解于20ulTE中，采用agilent2100对产物的大小进行检测，采用qubit 2.0对溶度进行文库质检。

10

#### 5、上机测序

得到的所有产物进行标准化，进行等量混合，混合得到的文库进行平行测序，测序平台BGISEQ-500，测序类型PE100。

#### 6、数据分析

15

分析步骤包括比对、拼接、去除非目标序列；  
统计测序的基本性能参数，包括比对率等；  
统计目标基因捕获的效率；  
统计EML4基因上各个引物扩增得到的序列比对到目标区域的深度；  
分析EML4和ALK发生融合的情况。

20

表4 测序下机数据基本参数统计结果:

样本编号	原始数据	过滤后数据	Q30	比对率	目标区域比例	平均深度	覆盖度
阳性样本 1	394791	389441	92.51%	95.99%	96.71%	12032	100.00%
阳性样本 2	475672	432607	93.03%	94.99%	97.10%	13363	100.00%
阳性样本 3	586062	551179	92.65%	95.98%	96.82%	17029	100.00%
阳性样本 4	513325	486684	92.99%	93.97%	97.30%	15032	100.00%
阴性样本 1	351036	329568	93.06%	93.98%	96.92%	10189	100.00%
阴性样本 2	390874	369312	92.93%	94.98%	97.18%	11414	100.00%
阴性样本 3	441027	427441	92.41%	95.97%	97.23%	13207	100.00%
阴性样本 4	482207	439282	92.48%	93.99%	97.11%	13572	100.00%

其中, 阳性样本1~阳性样本4分别代表来自同一个人的阳性样本; 阴性样本1~阴性样本4分别代表同一个人的阴性样本。

5 从表4给出的结果可以看出, 按照上述方法, 无论是对阳性样本进行测序, 还是对阴性样本进行测序, 所获得的测序数据在0.5M (原始数据) 左右, 捕获率 $97\pm 1\%$  (目标区域比例), 平均reads深度 $14000\pm 3000$ 乘, 覆盖度100%。表4的结果说明, 利用上述设计的引物, 采用本发明提供的方法能够覆盖需要检测的区域, 各项参数优良。

表5 各引物扩增得到的比对到目标位置的reads数

编号	阴性样本 reads (平均值)	阳性样本 reads (平均值)	编号	阴性样本 reads (平均值)	阳性样本 reads (平均值)
EML4-E02-01	13454	15674	EML4-E15	7135	9055
EML4-E02-02	18449	23153	EML4-E16	18490	20930
EML4-E03	13176	14994	EML4-E17	6017	7780
EML4-E04	14498	17152	EML4-E18	9674	11309
EML4-E05	6542	7341	EML4-E19	12571	14356
EML4-E07	4039	5174	EML4-E20	20419	24319
EML4-E08	10764	13649	EML4-E21	3295	4047
EML4-E10	21406	25644	EML4-E22-01	2184	2435
EML4-E11	23861	26485	EML4-E22-02	23194	28853
EML4-E12	20165	24379	GAPDH-DNA-E1	31	38
EML4-E13	10219	12805	GAPDH-RNA-E2	18238	22742
EML4-E13-ALK	0	2756	GAPDH-RNA-E3	11282	14396
EML4-E14	16634	18464	GAPDH-RNA-E4	7099	8185
<b><i>EML4-E14-ALK</i></b>	0	4568			

10 考虑到在阳性样本中融合会占到一定的比例, 也就在阳性样本中是会得到正常的 EML4 转录的

mRNA，也会存在异常转录的 mRNA（EML4-ALK），一般通过异常 mRNA 的比例来判断是否发生融合，发生融合的比例大于 0.1%就认为是存在异常。

从表5给出的数据来看，阴性样本未检测到EML-ALK融合，阳性样本有两条引物检测到目标基因融合（EML-E13-ALK/ EML-E14-ALK），融合比例分别为EML-E13-ALK 21.5%(2756/12805=21.5%)，  
5 EML-E14-ALK 24.7%(4568/18464=24.7%)。由此，采用本发明的方法能够检测到基因融合。

在本发明的描述中，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本发明的描述中，“多个”的含义是至少两个，例如两个，三个等，除非另有明确  
10 具体的限定。

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不必针对的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外，在不  
15 相互矛盾的情况下，本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

尽管上面已经示出和描述了本发明的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本发明的限制，本领域的普通技术人员在本发明的范围内可以对上述实施例进行变化、修改、替换和  
20 变型。

## 权利要求书

1、一种核酸序列，其特征在于，包含：

5 第一测序通用序列和随机引物序列，所述第一测序通用序列和所述随机引物序列相连，所述第一测序通用序列位于所述核酸序列的 5' 端，所述随机引物序列位于所述核酸序列的 3' 端，

其中，所述随机引物序列含有 6~10 个随机核苷酸。

2、根据权利要求 1 所述的核酸序列，其特征在于，所述随机引物序列含有 8 个随机核苷酸。

3、根据权利要求 1 所述的核酸序列，其特征在于，进一步包含：

10 第一分子标签序列，所述第一分子标签序列位于所述第一测序通用序列和所述随机引物序列之间；  
互补序列，所述互补序列位于所述第一分子标签序列和所述随机引物序列之间，所述互补序列与所述第一测序通用序列的部分序列互补。

4、根据权利要求 3 所述的核酸序列，所述第一分子标签序列含有 8~20 个随机核苷酸。

5、根据权利要求 3 所述的核酸序列，其特征在于，所述互补序列与所述第一测序通用序列的 5' 端的部分序列互补。

15 6、一种 RNA 目标区域测序文库的构建方法，其特征在于，包括：

基于 RNA 样本，利用逆转录引物和逆转录酶进行逆转录反应，获得 cDNA 的第一条链，其中所述逆转录引物为权利要求 1~5 中任一项所述的核酸序列；

20 基于所述 cDNA 的第一条链，利用上游特异性引物和第一通用引物进行第一轮 PCR 扩增，获得第一扩增产物，所述第一通用引物与所述第一测序通用序列的部分序列重叠，所述上游特异性引物的结合位点位于所述目标区域的上游；

25 基于所述第一扩增产物，利用下游特异性引物、第二通用引物和标签引物进行第二轮巢氏 PCR 扩增，获得所述目标区域测序文库，所述下游特异性引物的 5' 端含有部分或者全部第二通用测序序列，所述第二通用引物中含有第二测序通用序列，所述第二通用引物的 5' 端含有磷酸基团，所述第二通用引物和所述下游特异性引物的部分或者全部序列重叠，所述下游特异性引物 3' 端的结合位点位于所述目标区域的上游，所述标签引物与所述第一通用引物的部分序列重叠；

其中，所述下游特异性引物的结合位点位于所述上游特异性引物的下游。

7、根据权利要求 6 所述的构建方法，其特征在于，所述 RNA 为 cfRNA、高度降解的 RNA 或者长度为 200~300bp 的 RNA。

30 8、根据权利要求 7 所述的构建方法，其特征在于，所述长度为 200~300bp 的 RNA 通过利用  $Mg^{2+}$  对长片段 RNA 样本孵育获得。

9、根据权利要求 6 所述的构建方法，其特征在于，所述标签引物中含有第二分子标签序列，用于区分不同样本。

10、根据权利要求 9 所述的构建方法，其特征在于，所述第二分子标签序列为 8~12 个核苷酸。

11、根据权利要求 6 所述的构建方法，其特征在于，进一步包括：

35 所述上游特异性引物和/或所述下游特异性引物中含有至少有一个设计在内含子，基于所述至少有一个设计在内含子上的所述上游特异性引物和/或所述下游特异性引物进行所述第一轮 PCR 扩增和/或第二轮 PCR 扩增，检测所述 RNA 样本是否被 DNA 污染。

12、一种对目标区域基因变异检测的方法，其特征在于，包括：

采用权利要求 6~11 中任一项所述的构建方法获得目标区域测序文库；

基于所述目标区域测序文库，测序获得测序数据；

将所述测序数据与参考基因组进行比对，确定所述目标区域基因变异的结果。

13、根据权利要求 12 所述的方法，其特征在于，所述参考基因组为人类基因组 hg19。

5 14、根据权利要求 12 所述的方法，其特征在于，所述变异包括基因融合、SNV、基因缺失、基因插入突变中的至少一种；优选为基因融合。

15、根据权利要求 14 所述的方法，其特征在于，所述基因融合为 EML4-ALK 基因融合。

16、一种 EML4-ALK 融合基因的检测方法，其特征在于，包括：

基于 RNA 样本，利用逆转录引物和逆转录酶进行逆转录反应，获得 cDNA 的第一条链，其中所述逆转录引物为 SEQ ID NO:1 或者 SEQ ID NO:2 中的一种；

10 基于所述 cDNA 的第一条链，利用上游特异性引物和第一通用引物进行第一轮 PCR 扩增，获得第一扩增产物，所述第一通用引物为 SEQ ID NO:3，所述上游特异性引物为选自 SEQ ID NO: 6~SEQ ID NO:30 中的至少一种；

15 基于所述第一扩增产物，利用下游特异性引物、第二通用引物和标签引物进行第二轮巢氏 PCR 扩增，获得所述目标区域测序文库，所述第二通用引物为 SEQ ID NO:4，所述标签引物为 SEQ ID NO:5，所述下游特异性引物为 SEQ ID NO:31 ~SEQ ID NO:55 中的至少一种；

将所述目标区域测序文库进行测序，获得测序数据，基于所述测序数据与参考基因组的对比结果，确定所述 EML4-ALK 基因的融合结果。

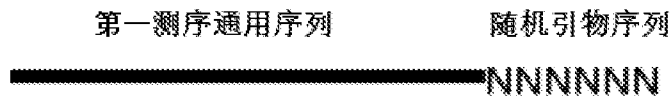


图 1

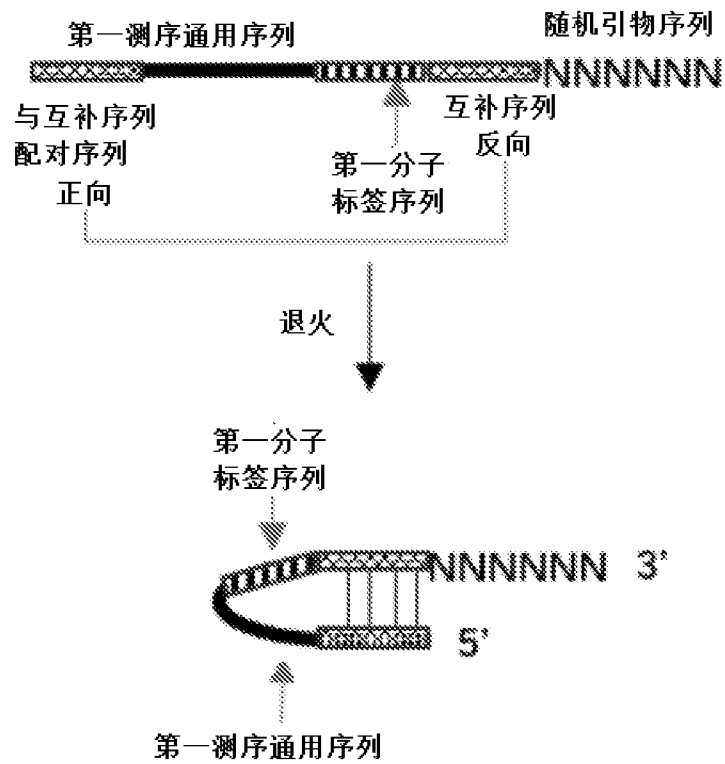


图 2

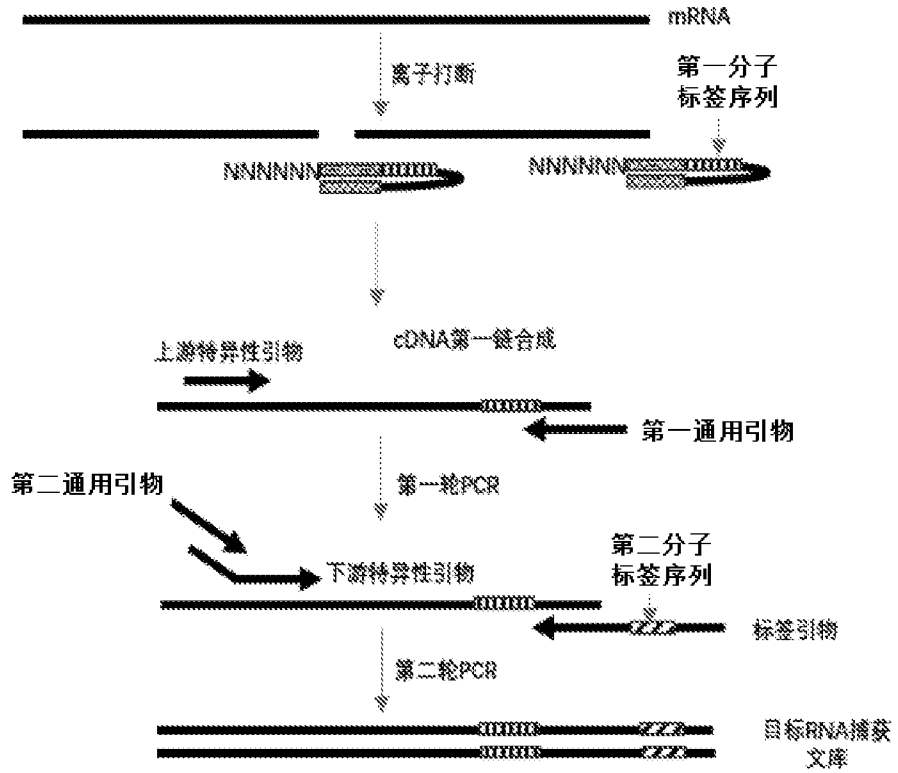


图 3

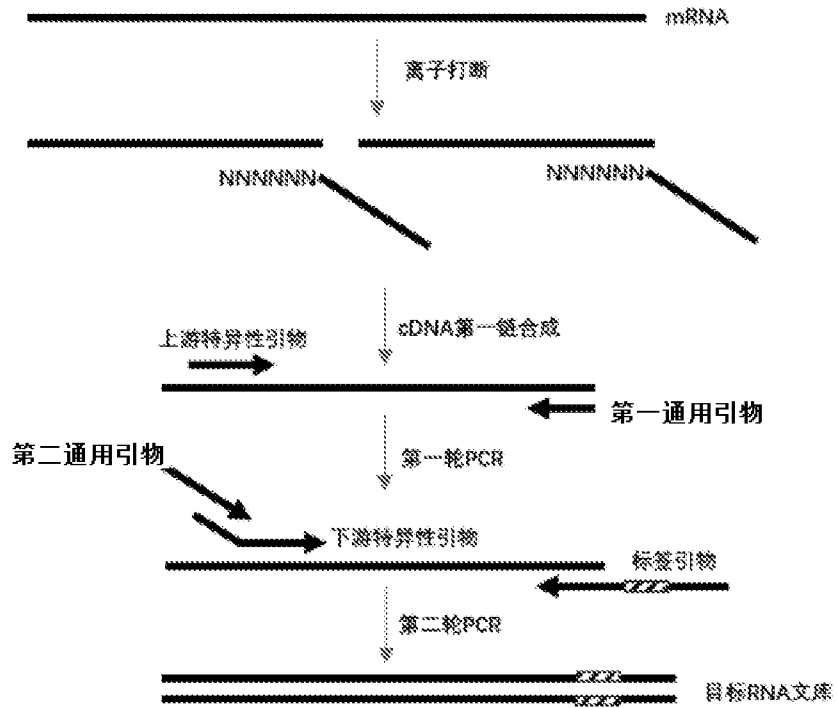


图 4

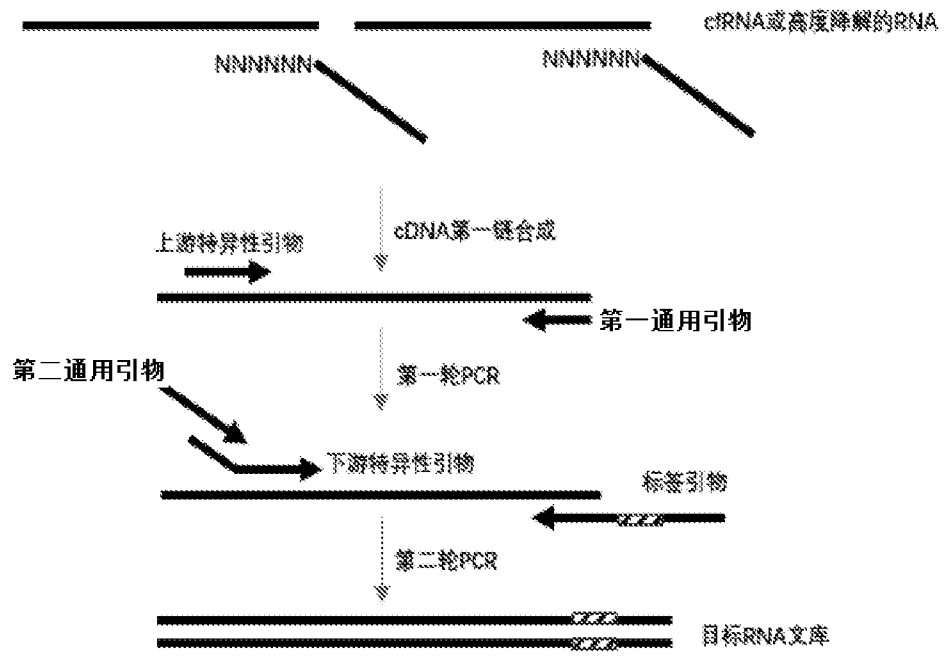


图 5

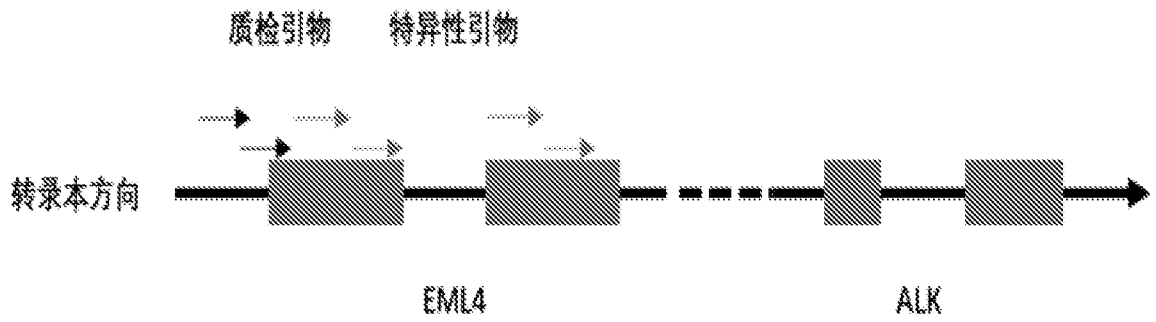


图 6

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/079235

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
C12N 15/11(2006.01)i; C40B 50/06(2006.01)i; C12Q 1/6806(2018.01)i; C12Q 1/6809(2018.01)i; C12Q 1/6886(2018.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) C12N C40B C12Q		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Databases: CNTXT; CNABS; CPRSABS; WOTXT; EPTXT; USTXT; DWPI; SIPOABS; CNKI: 对申请人和发明人的名称检索, cDNA文库, RNA测序, 转录组, 测序, 接头, 逆转录引物, 随机引物, 随机序列, 测序引物, 分子条形码, cDNA library, RNA seq, transcript+, seq, adaptor, reverse-transcription, random, primer, seq+, RNA, cDNA library, transcript+, UMI, EMBL4, ALK; GenBank+EMBL+DDBJ+中国专利序列数据库, GenBank+EMBL+DDBJ+NATIONAL SEQUENCE DATABASE OF CHINESE PATENT: 对SEQ ID NO: 1-3, 6-55进行序列检索, search for SEQ ID NO: 1-3, 6-55		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 106715714 A (BGI COLLEGE) 24 May 2017 (2017-05-24) see the primer design section of the embodiment	1-2
X	CN 105154567 A (SHANGHAI JIAO TONG UNIVERSITY) 16 December 2015 (2015-12-16) see description, paragraphs 40-42	1-5
Y	CN 105154567 A (SHANGHAI JIAO TONG UNIVERSITY) 16 December 2015 (2015-12-16) see description, paragraphs 40-42	6-16
Y	CN 107385042 A (GUANGZHOU FOREVERGEN HEALTH TECHNOLOGY CO., LTD.; GUANGZHOU FOREVERGEN MEDICAL LABORATORY CO., LTD.) 24 November 2017 (2017-11-24) see claims 1-9, and abstract, drawings	6-16
X	CN 107502607 A (ZHEJIANG UNIVERSITY) 22 December 2017 (2017-12-22) see claim 1	1-2
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search <b>19 August 2019</b>		Date of mailing of the international search report <b>09 October 2019</b>
Name and mailing address of the ISA/CN <b>China National Intellectual Property Administration No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088 China</b> Facsimile No. (86-10)62019451		Authorized officer  Telephone No.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/079235

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 107502607 A (ZHEJIANG UNIVERSITY) 22 December 2017 (2017-12-22) see claim 4	6-16
X	CN 105714383 A (BGI SHENZHEN; BGI SHENZHEN CO., LTD.) 29 June 2016 (2016-06-29) see SEQ ID NO. 4	1-2
X	CN 109097467 A (JIANGSU SUPERBIO TECHNOLOGY NANJING CO., LTD.) 28 December 2018 (2018-12-28) see description, paragraph 32	1-2
X	CN 102732598 A (CHEN, Xianfeng) 17 October 2012 (2012-10-17) see claim 1, and figure 1	1-2
X	CN 107075561 A (BGI SHENZHEN CO., LIMITED) 18 August 2017 (2017-08-18) see claims 1-3	1-2
X	CN 108588176 A (HUNAN YEARTH BIOTECHNOLOGY CO., LTD.) 28 September 2018 (2018-09-28) see description, paragraphs 17-19, and abstract, drawings	1-2
X	CN 109295500 A (CAPITALBIO CORPORATION) 01 February 2019 (2019-02-01) see claims 1-2	1-2
Y	CN 106929504 A (ANNOROAD GENE TECHNOLOGY BEIJING CO., LTD.) 07 July 2017 (2017-07-07) see claims 1-4	6-16
Y	VENDRELL, J.A. et al. "Detection of Known and Novel ALK Fusion Transcripts in Lung Cancer Patients Using Next-Generation Sequencing Approaches" <i>Scientific Reports</i> , Vol. 1, No. (7), 02 October 2017 (2017-10-02), ISSN: 2045-2322, pages 1-11, see abstract, figure 1, and page 3, first paragraph	6-16

**Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
  - a.  forming part of the international application as filed:
    - in the form of an Annex C/ST.25 text file.
    - on paper or in the form of an image file.
  - b.  furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
  - c.  furnished subsequent to the international filing date for the purposes of international search only:
    - in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
    - on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).
2.  In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2019/079235**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106715714	A	24 May 2017	WO	2016058173	A1	21 April 2016
				EP	3208344	A1	23 August 2017
				US	2017275616	A1	28 September 2017
				HK	1237002	A0	06 April 2018
CN	105154567	A	16 December 2015	None			
CN	107385042	A	24 November 2017	None			
CN	107502607	A	22 December 2017	None			
CN	105714383	A	29 June 2016	HK	1221710	A	16 August 2016
CN	109097467	A	28 December 2018	None			
CN	102732598	A	17 October 2012	CN	102732598	B	01 March 2017
CN	107075561	A	18 August 2017	WO	2016058121	A1	21 April 2016
				EP	32085343	A1	23 August 2017
				US	2018251813	A1	06 September 2018
				HK	1240283	A0	18 May 2018
CN	108588176	A	28 September 2018	None			
CN	109295500	A	01 February 2019	None			
CN	106929504	A	07 July 2017	None			

国际检索报告

国际申请号

PCT/CN2019/079235

<p><b>A. 主题的分类</b></p> <p>C12N 15/11(2006.01)i; C40B 50/06(2006.01)i; C12Q 1/6806(2018.01)i; C12Q 1/6809(2018.01)i; C12Q 1/6886(2018.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																							
<p><b>B. 检索领域</b></p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>C12N C40B C12Q</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>数据库: CNTXT;CNABS;CPRSABS;WOTXT;EPTXT;USTXT;DWPI;SIPOABS;CNKI: 对申请人和发明人的名称检索, cDNA文库, RNA测序, 转录组, 测序, 接头, 逆转录引物, 随机引物, 随机序列, 测序引物, 分子条形码, cDNA library, RNA seq, transcript+, seq, adaptor, reverse-transcription, random, primer, seq+, RNA, cDNA library, transcript+, UMI, EMBL4, ALK; GenBank+EMBL+DDBJ+中国专利序列数据库: 对SEQ ID NO:1-3、6-55进行序列检索</p>																							
<p><b>C. 相关文件</b></p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 106715714 A (深圳华大基因研究院) 2017年 5月 24日 (2017 - 05 - 24) 参见实施例引物设计部分</td> <td>1-2</td> </tr> <tr> <td>X</td> <td>CN 105154567 A (上海交通大学) 2015年 12月 16日 (2015 - 12 - 16) 参见说明书第40-42段</td> <td>1-5</td> </tr> <tr> <td>Y</td> <td>CN 105154567 A (上海交通大学) 2015年 12月 16日 (2015 - 12 - 16) 参见说明书第40-42段</td> <td>6-16</td> </tr> <tr> <td>Y</td> <td>CN 107385042 A (广州永诺健康科技有限公司 广州永诺医学检验所有限公司) 2017年 11月 24日 (2017 - 11 - 24) 参见权利要求1-9及摘要附图</td> <td>6-16</td> </tr> <tr> <td>X</td> <td>CN 107502607 A (浙江大学) 2017年 12月 22日 (2017 - 12 - 22) 参见权利要求1</td> <td>1-2</td> </tr> <tr> <td>Y</td> <td>CN 107502607 A (浙江大学) 2017年 12月 22日 (2017 - 12 - 22) 参见权利要求4</td> <td>6-16</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 106715714 A (深圳华大基因研究院) 2017年 5月 24日 (2017 - 05 - 24) 参见实施例引物设计部分	1-2	X	CN 105154567 A (上海交通大学) 2015年 12月 16日 (2015 - 12 - 16) 参见说明书第40-42段	1-5	Y	CN 105154567 A (上海交通大学) 2015年 12月 16日 (2015 - 12 - 16) 参见说明书第40-42段	6-16	Y	CN 107385042 A (广州永诺健康科技有限公司 广州永诺医学检验所有限公司) 2017年 11月 24日 (2017 - 11 - 24) 参见权利要求1-9及摘要附图	6-16	X	CN 107502607 A (浙江大学) 2017年 12月 22日 (2017 - 12 - 22) 参见权利要求1	1-2	Y	CN 107502607 A (浙江大学) 2017年 12月 22日 (2017 - 12 - 22) 参见权利要求4	6-16
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
X	CN 106715714 A (深圳华大基因研究院) 2017年 5月 24日 (2017 - 05 - 24) 参见实施例引物设计部分	1-2																					
X	CN 105154567 A (上海交通大学) 2015年 12月 16日 (2015 - 12 - 16) 参见说明书第40-42段	1-5																					
Y	CN 105154567 A (上海交通大学) 2015年 12月 16日 (2015 - 12 - 16) 参见说明书第40-42段	6-16																					
Y	CN 107385042 A (广州永诺健康科技有限公司 广州永诺医学检验所有限公司) 2017年 11月 24日 (2017 - 11 - 24) 参见权利要求1-9及摘要附图	6-16																					
X	CN 107502607 A (浙江大学) 2017年 12月 22日 (2017 - 12 - 22) 参见权利要求1	1-2																					
Y	CN 107502607 A (浙江大学) 2017年 12月 22日 (2017 - 12 - 22) 参见权利要求4	6-16																					
<p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 引用文件的具体类型:                      “A” 认为不特别相关的表示了现有技术一般状态的文件                      “E” 在国际申请日的当天或之后公布的在先申请或专利                      “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)                      “O” 涉及口头公开、使用、展览或其他方式公开的文件                      “P” 公布日先于国际申请日但迟于所要求的优先权日的文件                      “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件                      “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性                      “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性                      “&amp;” 同族专利的文件</p>																							
<p>国际检索实际完成的日期</p> <p>2019年 8月 19日</p>		<p>国际检索报告邮寄日期</p> <p>2019年 10月 9日</p>																					
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>		<p>授权官员</p> <p>张丹丹</p> <p>电话号码 62411308</p>																					

C. 相关文件		
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 105714383 A (深圳华大基因研究院 深圳华大基因科技有限公司) 2016年 6月 29日 (2016 - 06 - 29) 参见SEQ ID NO. 4	1-2
X	CN 109097467 A (江苏苏博生物医学科技南京有限公司) 2018年 12月 28日 (2018 - 12 - 28) 参见说明书第32段	1-2
X	CN 102732598 A (陈先锋) 2012年 10月 17日 (2012 - 10 - 17) 参见权利要求1及说明书附图图1	1-2
X	CN 107075561 A (深圳华大基因科技有限公司) 2017年 8月 18日 (2017 - 08 - 18) 参见权利要求1-3	1-2
X	CN 108588176 A (湖南大地同年生物科技有限公司) 2018年 9月 28日 (2018 - 09 - 28) 参见说明书第17-19段及摘要附图	1-2
X	CN 109295500 A (博奥生物集团有限公司) 2019年 2月 1日 (2019 - 02 - 01) 参见权利要求1-2	1-2
Y	CN 106929504 A (安诺优达基因科技北京有限公司) 2017年 7月 7日 (2017 - 07 - 07) 参见权利要求1-4	6-16
Y	VENDRELL Julie A. et al. "Detection of known and novel ALK fusion transcripts in lung cancer patients using next-generation sequencing approaches" 《Scientific Reports》, 第1卷, 第7期, 2017年 10月 2日 (2017 - 10 - 02), ISSN: 2045-2322, 第1-11页, 参见摘要、图1及第3页首段	6-16

## 第1栏 核苷酸和/或氨基酸序列(续第1页第1. c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列, 国际检索是基于下列序列列表进行的:
- a.  作为国际申请的一部分提交的:
- 附件C/ST. 25文本文件形式
  - 纸件或图形文件形式
- b.  根据细则13之三. 1(a) 仅为国际检索目的以附件C/ST. 25文本文件形式与国际申请同时提交的:
- c.  仅为国际检索目的在国际申请日之后提交的:
- 附件C/ST. 25文本文件形式(细则13之三. 1(a))
  - 纸件或图形文件形式(细则13之三. 1(b)和行政规程第713段)
2.  另外, 在提交/提供了多个版本或副本的序列列表的情况下, 提供了关于随后提交的或附加的副本中的信息与申请时提交的作为申请一部分的序列列表的信息相同或未超出申请时提交的申请中的信息范围(如适用)的所需声明。
3. 补充意见:

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2019/079235

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	106715714	A	2017年 5月 24日	WO	2016058173	A1	2016年 4月 21日
				EP	3208344	A1	2017年 8月 23日
				US	2017275616	A1	2017年 9月 28日
				HK	1237002	A0	2018年 4月 6日
CN	105154567	A	2015年 12月 16日	无			
CN	107385042	A	2017年 11月 24日	无			
CN	107502607	A	2017年 12月 22日	无			
CN	105714383	A	2016年 6月 29日	HK	1221710	A	2016年 8月 16日
CN	109097467	A	2018年 12月 28日	无			
CN	102732598	A	2012年 10月 17日	CN	102732598	B	2017年 3月 1日
CN	107075561	A	2017年 8月 18日	WO	2016058121	A1	2016年 4月 21日
				EP	32085343	A1	2017年 8月 23日
				US	2018251813	A1	2018年 9月 6日
				HK	1240283	A0	2018年 5月 18日
CN	108588176	A	2018年 9月 28日	无			
CN	109295500	A	2019年 2月 1日	无			
CN	106929504	A	2017年 7月 7日	无			

表 PCT/ISA/210 (同族专利附件) (2015年1月)