

# (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7月9日 (09.07.2020)



(10) 国际公布号  
**WO 2020/140693 A1**

- (51) 国际专利分类号:  
*C12Q 1/6806* (2018.01)
- (21) 国际申请号: PCT/CN2019/123891
- (22) 国际申请日: 2019年12月9日 (09.12.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
201910002408.5 2019年1月2日 (02.01.2019) CN  
201910897689.5 2019年9月20日 (20.09.2019) CN
- (71) 申请人: 上海臻迪基因科技有限公司 (SHANGHAI ZENISIGHT LTD.) [CN/CN]; 中国上海市浦东新区高科东路777弄10号2613室, Shanghai 201210 (CN)。
- (72) 发明人: 郭志伟 (GUO, Zhiwei); 中国上海市浦东新区高科东路777弄10号2613室, Shanghai 201210 (CN)。 李英辉 (LI, Yinghui); 中国上海市浦东新区高科东路777弄10号2613室, Shanghai 201210 (CN)。 陈倩 (CHEN, Qian); 中国上海市浦东新区高科东路777弄10号2613室, Shanghai 201210 (CN)。 胡荣君 (HU, Rongjun); 中国上海市浦东新区高科东路777弄10号2613室, Shanghai 201210 (CN)。
- (74) 代理人: 北京卓岚智财知识产权代理事务所 (特殊普通合伙) (BEIJING ZHUOLANIP INTELLECTUAL PROPERTY AGENT FIRM); 中国上海市房山区长阳昊天北大街48号金桥大厦C座304, Shanghai 102445 (CN)。

(54) Title: GENE TARGET REGION ENRICHMENT METHOD AND KIT

(54) 发明名称: 基因目标区域富集方法及试剂盒

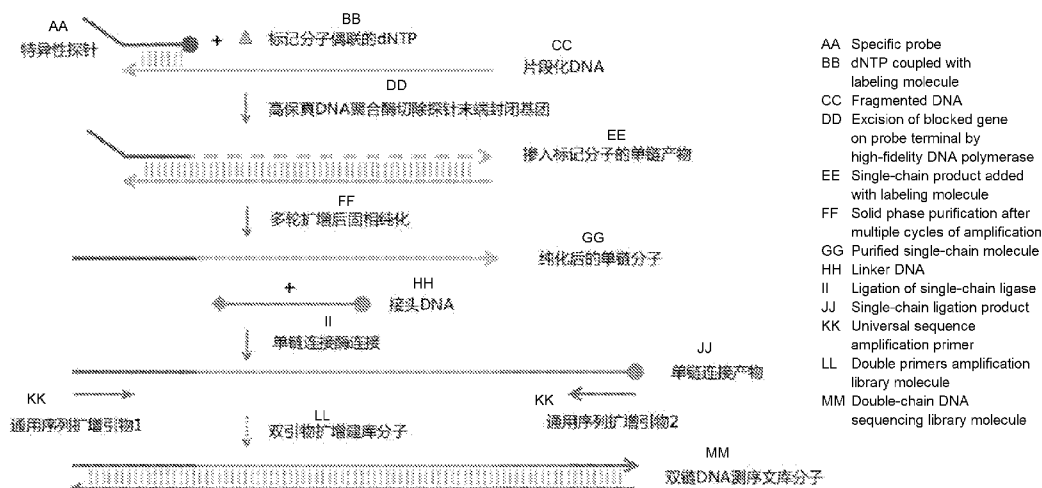


图1

(57) Abstract: Provided is a gene target region enrichment method and a kit. The method comprises (1) amplifying fragmented DNA comprising a target region by means of a specific probe so as to obtain a captured extension product, wherein the specific probe comprises a sequence complementary to the target region of the fragmented DNA, and the 3' terminal nucleotide of the specific probe is modified to prevent a ligation reaction at the 3' terminal of the specific probe; and (2) linking the 3' terminal of the captured extension product obtained in step (1) to linker DNA to obtain a ligation product.

(57) 摘要: 提供了一种基因目标区域富集方法及试剂盒。所述方法包括: (1)通过特异性探针扩增包括目标区域的片段化DNA, 以提供捕获延伸产物, 所述的特异性探针包括与所述片段化DNA的目标区域互补的序列, 所述特异性探针的3'末端核苷酸被修饰, 用于阻止所述特异性探针的3'末端发生连接反应; (2)将步骤(1)所提供的捕获延伸产物的3'末端连接接头DNA, 以提供连接产物。

(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))。

## 基因目标区域富集方法及试剂盒

### 技术领域

本发明涉及生物技术领域，特别是涉及一种基因目标区域富集方法及试剂盒。

### 背景技术

基因测序技术自上世纪七十年代问世以来已经历了近半个世纪，1985年PCR技术的横空出世推动了整个分子生物学领域的发展。下一代测序技术（NGS）具有准确、灵敏、通量高的优点，随着测序成本的不断降低，其应用范围不断扩大，但其应用也受到了需求多样化且费时费力的文库构建这一步骤的制约。在对临床样本比如血浆样本的建库过程中，DNA的存在形式通常是短片段的，受损的，单链或部分双链的，对于这些存在形式的尤其是片段小于200bp的DNA来说，现有的PCR技术并不能做到很好的捕获和富集。

对于微小DNA片段的富集，现有技术仍主要使用传统的PCR建库方法或先加接头连接再扩增的方法比如杂交捕获法。但对于前者来说，由于需要双端引物，极大限制了适合扩增的片段的长度，且存在扩增中的偏好性导致产物的高不均一性，以及指数扩增累积的错误导致后续测序结果不准确的问题；而对于后者，虽然对待富集片段长度的要求没有PCR高，但需要先进行连接反应，而连接的效率通常只有20%~50%，导致捕获效率低下，还存在因连接困难容易丢失稀有分子的问题。为了解决NGS的建库问题，近期发展起来的技术还包括诸如分子倒置探针、多重PCR等。分子倒置探针与杂交捕获技术相比，特异性较好，但其口袋状探针设计复杂，也不适用于微小DNA片段的富集。多重PCR技术适合大规模样本，应用最为广泛，但要么引物设计要求极高且扩增子均一性差，要么扩增产物均一性较好但对起始样本浓度要求很高，同样不适用于起始浓度低的微小DNA片段的富集。这些现有技术构建文库时通常需要双端引物，为了去除接头二聚体污染必然引入纯化步骤，这导致小片段的双链DNA、受损伤的双链DNA和单链DNA分子的信息丢失。然而，在一些转录活跃的基因组区域恰恰是这些存在形式的DNA。综上所述，现有技术对片段化的，尤其是200bp以下的微小DNA进行富集的需求严重未能满足，而找到高效富集片段化DNA目标区域的方法，突破连接效率对富集效果的瓶颈，抑制非目的连接产物的产生，最大程度的捕获稀有分子并保持产物的均一性，是本发明要解决的主要技术问题。

### 发明内容

鉴于以上所述现有技术的缺点，本发明的目的在于提供一种基因目标区域富集方法，用于解决现有技术中的问题。

为实现上述目的及其他相关目的，本发明提供一种基因目标区域富集方法，包括：

(1) 通过特异性探针扩增包括目标区域的片段化 DNA，以提供捕获延伸产物，所述的特异性探针包括与所述片段化 DNA 的目标区域互补的序列，所述特异性探针的 3'末端核苷酸被修饰，用于阻止所述特异性探针的 3'末端发生连接反应；

(2) 将步骤 (1) 所提供的捕获延伸产物的 3'端连接接头 DNA，以提供连接产物。

本发明另一方面提供一种用于富集片段化 DNA 目标区域的试剂盒，包括适用于所述的基因目标区域富集方法的特异性探针和接头 DNA。

## **附图说明**

图 1 是本发明实施例中对于目标区域富集过程的流程示意图。

图 2 是本发明实施例所构建的文库分子示意图。

图 3 是本发明实施例中校准品的构建示意图。

图 4 是本发明实施例中单特异性检测体系的流程示意图。

## **具体实施方式**

本发明发明人经过大量探索性研究，提供了一种基因目标区域富集方法，所述基因目标区域富集方法操作简单、结果可靠，尤其对于短片段核酸具有良好的富集效果，在此基础上完成了本发明。

本发明第一方面提供一种基因目标区域富集方法，包括：

(1) 通过特异性探针扩增包括目标区域的片段化 DNA，以提供捕获延伸产物，所述的特异性探针包括与所述片段化 DNA 的目标区域互补的序列，所述特异性探针的 3'末端核苷酸被修饰，用于阻止所述特异性探针的 3'末端发生连接反应；

(2) 将步骤 (1) 所提供的捕获延伸产物的 3'端连接接头 DNA，以提供连接产物。

本发明所提供的基因目标区富集方法中，可以包括：通过特异性探针扩增包括目标区域的片段化 DNA，以提供捕获延伸产物。反应时，所述特异性探针可以在捕获目标区域互补序列的基础上延伸，获得捕获延伸产物 (Capture-Extension Product, CEP)。

所述基因目标区域富集方法中，通过特异性探针扩增包括目标区域的片段化 DNA 的反应中，包括目标区域的片段化 DNA 可以是一个，也可以是多个。通常来说，特异性探针与包括目标区域的片段化 DNA 是一一对应的，即反应体系中特异性探针的数量可以是一种，也可以多种。此步骤所述的延伸指的是对样本的整个预扩增步骤，包括单轮或多轮变性、退火、延伸步骤的循环。优选的实施例中，此步骤实施多轮循环，比如 2~100、2~10、10~20、20~30、30~40、40~60、60~80、80~100 个循环，有效地增加包含目标区域的分子数。

所述基因目标区域富集方法中，所述片段化 DNA 可以是双链 DNA、单链 DNA 和 cDNA

等, 所述 cDNA 通常可以由 RNA 反转录获得。对于双链 DNA 来说, 特异性探针可以包括与所述片段化 DNA 其中一链的目标区域互补的序列。所以本发明的富集方法也适用于片段化 RNA, 本领域技术人员可以将 RNA 反转录成 cDNA 即可通过本发明所提供的富集方法进行后续操作。所述片段化 DNA 的长度可以是 25~200bps、25~40bps、40~60bps、60~80bps、80~100bps、100~120bps、120~140bps、140~160bps、160~180bps、或 180~200bps。

所述基因目标区域富集方法中, 所述步骤(1)的扩增体系中可以包括特异性探针、DNA 聚合酶和 dNTP。通过特异性探针扩增包括目标区域的片段化 DNA 的反应通常可以在 DNA 聚合酶存在的条件下进行, 3'端封闭修饰的探针在高保真聚合酶的作用下结合模板后, 封闭基团被切除, 探针被活化, 从而可以对目标序列进行有效延伸。所述 DNA 聚合酶可以具有 3'-5'外切酶活性, 从而可以切除结合模板后的探针 3'端的取代基团, 使探针能够延模板延伸, 优选可以为高保真 DNA 聚合酶, 用于进一步提高产物的扩增效率和纯度。所述 DNA 聚合酶也可以是普通的 DNA 聚合酶(即对应不具有 3'-5'外切酶活性)。所述步骤(1)的扩增体系中还包含活性物质, 所述活性物质可以用于切除结合目标区域后所述特异性探针的 3'末端修饰基团, 还可以在捕获延伸体系中与 DNA 聚合酶(例如, 普通的 DNA 聚合酶)组合, 以提升扩增体系的效率, 所述活性物质优选为核酸酶。通过特异性探针扩增包括目标区域的片段化 DNA 的反应通常可以在 dNTP 存在的条件下进行, 所述 dNTP 通常可以是偶联标记分子的 dNTP, 所述偶联标记分子可以是包括但不限于生物素等, 所述 dNTP 可以是包括但不限于 dCTP、dATP 等。所述 dNTP 还可以偶联有标记分子, 所述标记分子可以是生物素等, 所述标记分子通常可以用于捕获延伸产物的纯化。

所述基因目标区域富集方法中, 所述的特异性探针包括与所述片段化 DNA 的目标区域互补的序列, 从而可以实现对片段化 DNA 的目标区域的特异性扩增, 本领域技术人员可选择合适的片段化 DNA 的目标区域, 并根据片段化 DNA 的目标区域, 设计合适的互补的序列。所述特异性探针可以是 3'末端核苷酸被修饰的特异性探针, 用于阻止所述特异性探针的 3'末端与其他基团发生连接反应, 从而可以避免游离探针的自连或连接其它非目的分子, 例如, 非目的分子可以是接头 DNA 等。本领域技术人员可选择合适的修饰基团以实现特异性探针 3'末端的修饰, 例如, 所修饰的基团可以取代特异性探针 3'末端核苷酸上的天然基团(例如, 羟基等), 以阻止所述特异性探针 3'端发生连接反应, 所使用的修饰基团通常可以是封闭基团。探针通过互补序列与模板上的目标区域结合后, 3'端的修饰基团可以被酶切除, 使得探针被活化, 从而可以对目标序列进行有效延伸。特异性探针的 3'末端修饰基团可以是包括但不限于氢原子、C3 Spacer 基团、C6 Spacer 基团、磷酸基团(PO<sub>4</sub>)、氨基(NH<sub>2</sub>)等。不同取代

基团的选择对于探针的捕获效果有明显差异，在本发明一优选实施例中，将探针 3'端羟基取代为 C3 Spacer 得到的效果最佳，与其它取代基团相比具有明显优势。所述特异性探针还包括通用序列，所述通用序列通常能被测序系统识别，从而可以将后续提供的连接产物通过测序系统进行测序，例如，对于 Ion Torrent 测序系统，所述通用序列可以是对应的 P1 序列。在本发明一优选实施方式中，所述特异性探针 3'端尾部区域的碱基可以包含错配，错配的可以是探针 3'端最后一个碱基，也可以是接近 3'端的碱基；错配的可以是一个碱基，也可以是复数个。该错配不仅不影响探针的特异性和结合效率，还更有利于提高高保真 DNA 聚合酶的切割效率和保真度。

所述基因目标区域富集方法中，所述步骤（1）中，还可以包括纯化捕获延伸产物。本领域技术人员可选择合适的方法对捕获延伸产物进行纯化，例如，捕获延伸产物的纯化方法可以是包括磁珠纯化等。在本发明一具体实施例中，可以通过标记分子对捕获延伸产物进行纯化，纯化过程可以使用亲和素或者链霉亲和素包被的磁珠的固相纯化方法等。

本发明所提供的基因目标区富集方法中，还可以包括：将步骤（1）所提供的捕获延伸产物的 3'端连接接头 DNA，以提供连接产物。

所述基因目标区域富集方法中，将捕获延伸产物的 3'端连接接头 DNA 的连接体系中可以包括单链连接酶，所述单链连接酶使接头 DNA 的 5'端连接到所述捕获延伸产物的 3'端，所述单链连接酶优选为 T4 RNA 连接酶或热稳定性 RNA 连接酶等。

所述基因目标区域富集方法中，所述接头 DNA 可以是 5'末端核苷酸被修饰的、且在步骤（2）的反应温度下为单链结构，从而可以在单链连接酶的催化下，通过捕获延伸产物的 3'端羟基与接头 DNA 5'末端的修饰基团生成共价键，得到连接产物。在本发明一具体实施方式中，接头 DNA 的 5'末端核苷酸（例如，具体可以为的 5'末端核苷酸的 5 位羟基）被磷酸基团取代，从而形成磷酸化修饰，在单链连接酶的催化下，捕获延伸产物的 3'端羟基与接头 DNA 磷酸化的 5'末端生成共价键，得到连接产物。在本发明另一具体实施方式中，接头 DNA 的 5'末端核苷酸（例如，具体可以为的 5'末端核苷酸的 5 位羟基）被腺苷基团取代，从而形成腺苷酸化修饰，在 5'App DNA/RNA 热稳定连接酶的催化下，接头 DNA 的 5'末端可以与捕获延伸产物的 3'端发生连接反应。在本发明另一具体实施方式中，接头 DNA 的 5'末端核苷酸（例如，具体可以为的 5'末端核苷酸的 5 位羟基）被磷酸基团取代，从而形成磷酸化修饰，在热稳定 RNA 连接酶的催化下，接头 DNA 的 5'末端可以与捕获延伸产物的 3'端发生连接反应。所述接头 DNA 也可以是 5'端区域具有黏性末端的部分双链结构，其 5'端区域的黏性末端具有单链性质，从而可以通过如上所述的方法对 5'端进行修饰，在合适的单链连接酶催化下，

可以与捕获延伸产物连接。所述接头 DNA 还可以包括通用序列、样本标签序列、分子标签序列等中的一种或多种的组合，这些序列通常可以被测序系统识别的，从而可以对制备获得的连接产物进行测序。例如，对于 Ion Torrent 测序系统，所述通用序列可以为 A 序列，所述样本标签序列可以是 barcode，与其对应的分子标签序列等。引入 barcode 有利于后续的生信分析中对不同来源的样本进行区别，从而实现在单次反应中同时进行多个样本的测序以提高通量。引入分子标签有利于后续的生信分析中识别不同分子，以便进一步对后续扩增中产生的变异进行识别。这些能被测序系统识别的序列的应用，使本发明完成建库后的文库能够通过高通量测序平台进行测序，以提供各种后续研究和临床应用所需要的信息。

在本发明一优选实施例中，所述接头 DNA 的 3'端羟基也被其他基团取代，以阻止所述接头 DNA 3'端的连接反应，避免其自连或连接其它非目的分子，比如来自步骤（1）的游离探针。

所述基因目标区域富集方法中，所述步骤（2）中，还可以包括纯化连接产物。本领域技术人员可选择合适的方法对连接产物进行纯化，例如，连接产物的纯化方法可以是包括但不限于硅胶柱纯化、加热处理等。

本发明所提供的基因目标区域富集方法，可以用于基因检测。通过扩增后的连接产物进一步进行基因检测的方法对于本领域技术人员来说应该是已知的。在本发明一具体实施例中，所述目标区域富集的方法可以应用于高通量测序中。在本发明另一具体实施例中，本发明目标区域富集的方法可以应用于基因序列检测中，例如，所述的目标区域包括：序列发生变异的位点，更具体可以是单碱基突变位点区域，碱基缺失位点区域，碱基插入位点区域和融合突变位点区域等。在本发明另一具体实施例中，本发明目标区域富集的方法可以应用于基因修饰状态的检测中，例如，所述的目标区域包括可能存在甲基化位点的区域，再例如，所述的 DNA 片段经过了重亚硫酸盐的处理。

本发明所提供的基因目标区域富集方法，还可以包括：（3）扩增步骤（2）所提供的连接产物。通常可以通过 DNA 聚合酶和 PCR 扩增引物对连接产物进行 PCR 扩增，通过扩增连接产物，可以进一步富集含有目标区域的 DNA 的产物。本领域技术人员可选择合适的方法和体系对步骤（2）所提供的连接产物进行扩增，例如，PCR 扩增引物具有与所述特异性探针的通用序列和/或所述接头 DNA 的通用序列相配合，具体的，所述的 PCR 扩增引物具有与所述特异性探针的通用序列和所述接头 DNA 的通用序列至少部分互补的序列。

本发明所提供的基因目标区域富集方法，还可以包括：（4）对扩增后的连接产物进行测序，以提供目标区域的测序结果。在对连接产物进行 PCR 扩增之后，可以通过识别扩增产物

中的通用序列对扩增产物进行测序，获得目标区域的测序结果。本领域技术人员可选择合适的方法和体系对扩增后的连接产物进行测序，例如，基于 Ion Torrent 平台的 P1 序列和 A 序列进行测序，获得目标区域的测序结果。需要说明的是，本发明的富集方法可使步骤（2）得到的连接产物包含任意测序通用序列，比如可以是 Ion Torrent 测序平台，也可以是 Illumina 测序平台，当然也可以是其它测序平台的通用序列。

本发明所提供的基因目标区域富集方法，还可以包括：（5）用检测引物 1、检测引物 2 和探针 3 检测步骤（2）提供的连接产物，以非二代测序而是 PCR 检测的手段快速提供目标区域的检测结果。所述的检测引物 1、检测引物 2 和探针 3 中，至少其一包含基因特异性序列，也即是说，对于不同的基因目标区域，三种引物/探针有关特异性序列的组合可以是单特异性、双特异性或三特异性。在本发明一个实施例中，所述检测引物 1 包含基因特异性序列，检测引物 2 和探针 3 包含通用序列，或检测引物 1 和检测引物 2 包含基因特异性序列；或检测引物 1 和探针 3 包含基因特异性序列，或检测引物 1、检测引物 2 和探针 3 均包含基因特异性序列。在一些实施例中，探针 3 还可以包含标记分子，比如可以是荧光分子，且探针 3 的序列与检测引物 1 或 2 不互补。

本发明第二方面提供一种用于富集片段化 DNA 目标区域的试剂盒，包括适用于本发明第一方面所提供的基因目标区域富集方法的特异性探针和接头 DNA。所述特异性探针和接头 DNA 的结构在本发明第一方面已经详细描述，在此不作赘述。

本发明所提供的试剂盒中，还可以包括以下组分的一种或多种：RNA 连接酶、偶联标记分子的 dNTP、DNA 聚合酶、核酸酶等。其中，RNA 连接酶可以是热稳定性 RNA 连接酶、T4 RNA 连接酶、或 5'App DNA/RNA 热稳定连接酶等；dNTP 可以是 dCTP，也可以是 dATP。所述 DNA 聚合酶可以是具有 3'-5'外切酶活性的 DNA 聚合酶，优选的，所述 DNA 聚合酶选用高保真聚合酶。

本发明所提供的试剂盒中，还可以包括用于 PCR 扩增的正向引物和反向引物，它们通常与特异性探针的通用序列和所述接头 DNA 的通用序列相配合，具体可以具有与所述特异性探针的通用序列和所述接头 DNA 的通用序列至少部分互补的序列。

本发明所提供的试剂盒中，还可以包括用于 PCR 检测的检测引物 1、检测引物 2 和探针 3，三者中至少其一含有基因特异性序列，具体的可以是单特异性、双特异性或三特异性的引物/探针组合。优选的情况下，仅检测引物 1 包含基因特异性序列，检测流程请参考图 4。

在本发明一优选实施例中，通过本发明的方法或试剂盒进行文库构建后，文库分子（例如，步骤（2）所提供的连接产物，结构可以如图 2 所示）在结构上依次包含以下序列：5'端



测序通用序列、基因特异性探针序列、富集到的目标区域序列、接头序列、barcode、分子标签序列、3'端测序通用序列，如图2所示。其中富集到的目标区域包含富集前样本DNA的序列信息，该部分序列的特点是：5'端在基因组上的位置是固定的，由特异性探针决定；而3'端位置不固定，由建库初始的DNA片段化状态决定。因此，在富集后的数据分析中，该序列的3'端在基因组上的位置也可以起到分子标签的作用。分子标签可用于区分不同分子，提高检测的灵敏度与准确性。

本发明的有益效果在于，首先，在连接前先对所有片段化DNA样本的目标区域通过捕获延伸的手段进行预扩增，避免在连接阶段因连接酶连接效率不足所致的原始目标分子的损失或漏检，特别是小片段和稀有分子；预扩增阶段的延伸反应属于线性扩增，没有PCR扩增的偏好性，也不会累积PCR扩增所引入的错误，与常规PCR建库技术相比，产物均一性好且后续的测序结果更准确；其次，预扩增阶段，只需为每个目标基因设计一条长约30bp的单链探针，避开了为短片段如cfDNA和更短片段如ctDNA设计双端引物的困难，既提高了建库成功率还提升了建库便利性；而封闭探针3'端，可以阻断探针与模板之外的非目的连接，有效降低游离探针引起的背景噪声，辅以偶联标记分子的dNTP及其纯化系统，进一步提高目的产物纯度，使样本DNA分子达到理论最高转化率；再次，将接头DNA的5'端修饰后，在单链连接酶催化下能很好地连接到延伸产物的3'端但不会连接到封闭修饰的特异性探针3'端，同时封闭接头3'端避免接头的自连和错连，提高目的连接产物在终产物中的比例。最后，富集目标区域的过程中，或在探针和接头上设置好测序通用序列、样本标签序列和分子标签序列，达到建库完成后直接测序和数据分析的目的；或用本发明独创的单、双或三特异性引物/探针体系对目标区域进行快速检测。综上所述，本发明的富集方法及试剂盒，操作简单，结果可靠，针对片段长度小于200bp的DNA使用，可以最大程度的减少原始分子尤其是稀有分子的损失，最高效地富集目标分子；且该富集方法及试剂盒保真性极高，通过后续生信手段可使在测序过程中产生的碱基错误得到有效纠正，得到理论上可达到的最高测序准确率。

以下通过特定的具体实例说明本发明的实施方式，本领域技术人员可由本说明书所揭露的内容轻易地了解本发明的其他优点与功效。本发明还可以通过另外不同的具体实施方式加以实施或应用，本说明书中的各项细节也可以基于不同观点与应用，在没有背离本发明的精神下进行各种修饰或改变。

在进一步描述本发明具体实施方式之前，应理解，本发明的保护范围不局限于下述特定的具体实施方案；还应当理解，本发明实施例中使用的术语是为了描述特定的具体实施方案，

而不是为了限制本发明的保护范围；在本发明说明书和权利要求书中，除非文中另外明确指出，单数形式“一个”、“一”和“这个”包括复数形式。

当实施例给出数值范围时，应理解，除非本发明另有说明，每个数值范围的两个端点以及两个端点之间任何一个数值均可选用。除非另外定义，本发明中使用的所有技术和科学术语与本技术领域技术人员通常理解的意义相同。除实施例中使用的具体方法、设备、材料外，根据本技术领域的技术人员对现有技术的掌握及本发明的记载，还可以使用与本发明实施例中所述的方法、设备、材料相似或等同的现有技术的任何方法、设备和材料来实现本发明。

除非另外说明，本发明中所公开的实验方法、检测方法、制备方法均采用本技术领域常规的分子生物学、生物化学、染色质结构和分析、分析化学、细胞培养、重组 DNA 技术及相关领域的常规技术。这些技术在现有文献中已有完善说明，具体可参见 Sambrook 等 MOLECULAR CLONING : A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001 ; Ausubel 等, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates ; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego ; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998 ; METHODS IN ENZYMOLOGY, Vol. 304, Chromatin (P.M. Wassarman and A.P. Wolffe, eds.), Academic Press, San Diego, 1999 ; 和 METHODS IN MOLECULAR BIOLOGY, Vol. 119, Chromatin Protocols (P. B. Becker, ed.) Humana Press, Totowa, 1999 等。

## 实施例 1

实施例 1 中所用寡核苷酸序列如表 1 所示：

表 1

名称	SEQ ID NO.	序列 (5'→3')
Probe1	1	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGA ACGTACTGGTGAA-OH
Probe2	2	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGA ACGTACTGGTGAA-C3 Spacer
Probe2p	3	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGA ACGTACTGGTGAA-PO4

Probe2c	4	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGA</i> <i>ACGTACTGGTGAA-C6 Spacer</i>
Probe2n	5	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGA</i> <i>ACGTACTGGTGAA-NH<sub>2</sub></i>
Probe2d	6	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGA</i> <i>ACGTACTGGTGAA-DDC</i>
Probe3	7	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGA</i> <i>ACGTACTGGTGAT-C3 Spacer</i>
Probe4	8	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGA</i> <i>ACGTACTGrGTGAT-C3 Spacer</i>
ABar-X1	9	PO <sub>4</sub> - <i>GGATCCNNNNNNNCAGCTTGGACTGAGTCGGAGACACGCAGGG</i> <i>ATGAGATGG</i>
ABar-X2	10	PO <sub>4</sub> - <i>GGATCCNNNNNNNGTTCTCCTTACTGAGTCGGAGACACGCAGGG</i> <i>ATGAGATGG -C3 Spacer</i>
ABar-X3	11	5' App- <i>GGATCCNNNNNNNTTACCTTAGCTGAGTCGGAGACACGCAGGGA</i> <i>TGAGATGG-C3 Spacer</i>
EF-1	12	<i>GATCACAGATTTTGGGC</i>
ER-1	13	<i>TTTGCCTCCTTCTGC</i>
EM-1	14	<i>AACTGCTGGGTGCGGA</i>
AF	15	<i>GTCTCAGCCTCTCTATGGGCAGTCGGTGAT</i>
AR	16	<i>CCATCTCATCCCTGCGTGTCTCCGACTCAG</i>

此实施例中的探针包含测序通用序列，即斜体标示的碱基，此处包含有 Ion Torrent 测序系统中的 P1 序列。表中所有探针的特异性序列都针对 EGFR 基因 21 外显子的 L858R 突变。其中，探针 1 (SEQ ID NO.1) 3'端无修饰，探针 2、3 和 4 (SEQ ID NO.2, SEQ ID NO.7, SEQ ID NO.8) 3'端羟基被替换为 C3 Spacer，探针 2p (SEQ ID NO.3) 3'端羟基被替换为磷酸基，探针 2c (SEQ ID NO.4) 3'端羟基被替换为 C6 Spacer，探针 2n (SEQ ID NO.5) 3'端羟基被替换为磷酸基，探针 2d (SEQ ID NO.6) 3'端羟基被替换为双脱氧胞嘧啶 (DDC)；其中，探针 3 (SEQ ID NO.7) 的最后一个碱基为错配碱基，探针 4 (SEQ ID NO.8) 中靠近 3'端的

G 被替换为 RNA 碱基 rG。

接头 ABar-X1/2/3 (SEQ ID NO.9, SEQ ID NO.10, SEQ ID NO.11) 中包含测序通用序列, 即斜体标示的碱基, 此处为 Ion Torrent 测序系统中的 A 序列; 下划线部分序列为 barcode, 在本发明不同的样本中可以被替换成不同的 barcodes, 此部分的变化在本发明中不再累述。“N”部分为分子标签序列, 其中 NNNNNN 为随机序列, 用以标记同一样本中不同的延伸产物分子。接头 ABar-X1 和 ABar-X2 (SEQ ID NO.9, SEQ ID NO.10) 的 5'端为磷酸化修饰, 接头 ABar-X3 (SEQ ID NO.11) 的 5'端为腺苷酸化修饰; ABar-X2 和 ABar-X3 (SEQ ID NO.10, SEQ ID NO.11) 的 3'端的羟基被替换为 C3 Spacer。

EF-1 和 ER-1 (SEQ ID NO.12, SEQ ID NO.13) 是用于 PCR 检测 EGFR 目的序列的前后引物, EM-1 (SEQ ID NO.14) 为 MGB 探针。AF 和 AR (SEQ ID NO.15, SEQ ID NO.16) 分别是用来扩增文库的前后引物, 与文库中的通用序列, 即上述探针中的 P1 序列和接头中的 A 序列匹配。

### 主要试剂与材料

血浆游离 DNA 提取试剂盒购自 Qiagen; 细胞 DNA 提取试剂盒与纯化单链 DNA 的试剂盒为 RNA Clean Kit 购自 Tiangen Biotech; 热稳定 RNA 连接酶购自美国 Epicentre 公司, 高保真 DNA 聚合酶反应体系、Biotin-dCTP 及用于纯化的 MyOne Strptavidin C1 试剂购自 Invitrogen, DNA 聚合酶、高保真 DNA 聚合酶和 PCR 反应试剂盒购自 TOYOBO; RNase H2 热稳定性核酸酶购自 IDT。Q5 高保真 DNA 聚合酶、T4 RNA 连接酶与 5' App DNA/RNA 热稳定连接酶购自 NEB; Agencourt AMPure 磁珠购自 Beckman; 相关过程中定量检测的校准品按照分子克隆常规方法构建, 具体构建组成见图 3 所示。

### 步骤 1、样本准备

用抽提试剂盒提取健康人血浆样本中的 cfDNA 与 NCI-H1975 细胞株(该细胞株为 EGFR 基因 21 外显子 L858R 突变)中的 DNA, 通过 Qubit 荧光定量仪对 DNA 样本定量, 将 NCI-H1975 细胞株的 DNA 打断成 160bp 左右的片段, 按照 10%, 1%, 0.1%, 0.01%的比例掺入到健康人的 cfDNA 样本中, 同时以健康人 cfDNA 作为空白对照 (QC)。

### 步骤 2、捕获延伸

按表 2、表 3、表 4 分别配制捕获延伸反应体系。

表 2

组分	体积 (μL)	终浓度
----	---------	-----

2X 高保真 DNA 聚合酶 PCR Master Mix (DNA 聚合酶具有 3'-5'外切酶活性); 或, 2X DNA 聚合酶 PCR Master Mix (DNA 聚合酶不具有 3'-5'外切酶活性)	25	1×
特异性探针 (10uM)	0.5	100 nM
所有类型 cfDNA 样本	20	20000 拷贝
H <sub>2</sub> O	4.5	/
Total	50	/

表 3

组分	体积 (μL)	终浓度
2X DNA 聚合酶 PCR Master Mix (DNA 聚合酶不具有 3'-5'外切酶活性)	25	1×
RNAse H2 热稳定性核酸酶	0.5	/
特异性探针 (10uM)	0.5	100 nM
所有类型 cfDNA 样本	20	20000 拷贝
H <sub>2</sub> O	4	/
Total	50	/

表 4

组分	体积 (μL)	终浓度
2X 高保真 DNA 聚合酶 PCR Master Mix (DNA 聚合酶具有 3'-5'外切酶活性)	25	1×
特异性探针 (10uM)	0.5	100 nM
Biotin-dCTP (0.4mM)	2.5	20uM
所有类型 cfDNA 样本	20	20000 拷贝
H <sub>2</sub> O	2	/
Total	50	/

捕获延伸的 PCR 程序如表 5 所示:

表 5

循环数	温度	时间
1 cycle	98°C	60s
1/20/40 cycles	98°C	10 S

	67°C	20 S
	72°C	20 S
1 cycle	72°C	5min

捕获延伸程序完成后用 RNA Clean Kit 试剂盒对延伸产物进行硅胶柱纯化，用 60ul 洗脱缓冲液洗脱。（本发明中后续所指硅胶柱纯化，都与此步骤相同。）对于表 4 的延伸产物，在经过硅胶柱纯化后用链霉亲和素包被的磁珠再次纯化，并最终溶解于 60ul 洗脱缓冲液中。

### 步骤 3、延伸效率检测

按照表6配制延伸产物的PCR检测体系。

表 6

组分	用量 (μL)	终浓度
2×Taqman Mix	25	1×
ER-1 (10uM)	1.5	300 nM
AF (10uM)	1.5	300 nM
EM-1 (10uM)	0.5	100 nM
H <sub>2</sub> O	17.5	/
延伸产物/校准品	4	/
总体积	50	/

检测延伸效率的 PCR 程序如表 7。

表 7

循环数	温度	时间
1 cycle	95°C	3 min
45 cycles	95°C	10s
	56°C	25s
	72°C	20s

### 步骤 4、单链连接

根据步骤 3 的结果，选择步骤 2 得到的部分延伸产物分别按表 8、表 9、表 10 配制连接反应体系。

表 8

组分	体积 (μL)	终浓度	连接条件
CircLigase II 10X Reaction Buffer	4	1X	60 度 1 小时
50 mM MnCl <sub>2</sub>	2	2.5 mM	

5 M Betaine	8	1 M	
热稳定 RNA 连接酶(100 U/μL)	2	5 U/μL	
接头 ABar-X1/ABar-X2 (200 nM)	2	10 nM	
延伸产物	8		
50% PEG	14	17.5%	
Total	40	/	

表 9

组分	体积 (μL)	终浓度	连接条件
10X 连接 Buffer	4	1X	16 度 4 小时
T4 RNA 连接酶	2	5 U/μL	
接头 ABar-X2 (200 nM)	2	10 nM	
延伸产物	8	/	
50% PEG	14	17.5%	
Total	40	/	

表 10

组分	体积 (μL)	终浓度	连接条件
10X NE Buffer 1	4	1X	65 度 1 小时
50 mM MnCl <sub>2</sub>	2	2.5 mM	
H <sub>2</sub> O	8	/	
5' App DNA/RNA 热稳定连接酶 (100 U/μL)	2	5 U/μL	
接头 ABar-X3 (200 nM)	2	10 nM	
延伸产物	8		
50% PEG	14	17.5%	
Total	40	/	

得到连接产物后的处理如表 11 所示。

表 11

处理 1	处理 2	处理 3
不作任何处理 (对照)	95°C 孵育 10min	硅胶柱纯化

### 步骤 5、连接效率检测

按照表 12 配制 PCR 检测体系对步骤 4 的连接效率进行检测。

表 12

试剂	用量 (μL)	终浓度
2 X Taqman Mix	12.5	1×
AF (10uM)	0.75	300 nM
AR (10uM)	0.75	300 nM
EM-1(10uM)	0.25	100 nM
H <sub>2</sub> O	8.75	/
校准品/连接产物	2	/
Total	25	/

检测连接效率的 PCR 程序如表 13 所示。

表 13.

循环数	温度	时间
1 cycle	95°C	4 min
45 cycles	95°C	10s
	60°C	30s
	72°C	20s

### 步骤 6、文库扩增

按表 14 记载的反应体系和程序对步骤 4 得到的连接产物进行 PCR 扩增。

表14

组分		体积 (μL)
连接产物		30
5× Q5 Buffer		10
dNTP(10mM each)		1
Q5高保真DNA聚合酶(2 U/μL)		1
AF (10μM)		1.5
AR (10μM)		1.5
ddH <sub>2</sub> O		5
总体积		50
温度(°C)	时间(秒)	循环数
98	30	1
98	10	15
72	30	
72	300	1

反应结束后，取 80μL Agencourt AMPure 磁珠对扩增后的文库进行纯化，将纯化产物溶于



30ul 洗脱缓冲液中洗脱，至此，待上机测序的文库构建完成。

## 步骤 7、测序与数据分析

将制备好的文库在 Ion Proton 测序仪上进行测序，相关操作包括油包水 PCR、文库富集、芯片加样和上机测序，具体操作流程详见 Ion PITM Hi-Q™ OT2 200 Kit 说明书和 Ion PITM Hi-Q™ Sequencing 200 Kit 说明书。

## 实验结果及分析

结果中捕获延伸倍数、捕获效率及连接效率的计算方式如下：

捕获延伸倍率 = 捕获后 EGFR 基因输出拷贝数/捕获前 EGFR 基因输入拷贝数

捕获效率 = (捕获延伸倍数-1) / 捕获延伸循环数 × 100%

连接效率 = (连接产物输出拷贝数/延伸产物输入拷贝数) × 100%

本实施例中所有过程数据（如捕获效率，连接效率）的分析都来自 QC 样本。

## 结果 1、不同类型的聚合酶对不同探针的捕获延伸结果的影响。

表 15. 不同类型聚合酶的捕获延伸测试结果（QC 样本）

探针	DNA 聚合酶类型	核酸酶	循环数	捕获后 EGFR 输出拷贝数	捕获前 EGFR 输入拷贝数	捕获延伸倍率	捕获效率%
Probe1	高保真聚合酶	无	40	482000	20000	24.1	57.75%
Probe2	高保真聚合酶	无	40	566000	20000	28.3	68.25%
Probe2p	高保真聚合酶	无	40	98400	20000	4.92	9.80%
Probe2c	高保真聚合酶	无	40	428000	20000	21.4	51%
Probe2n	高保真聚合酶	无	40	64800	20000	3.24	5.6%
Probe2d	高保真聚合酶	无	40	71200	20000	3.56	6.4%
Probe3	高保真聚合酶	无	40	524000	20000	26.2	63.00%
Probe4	高保真聚合酶	无	40	482000	20000	24.1	57.75%
Probe1	普通聚合酶	无	40	512000	20000	25.6	61.50%
Probe2	普通聚合酶	无	40	24000	20000	1.2	0.50%
Probe2p	普通聚合酶	无	40	21000	20000	1.05	0.12%
Probe2c	高保真聚合酶	无	40	21000	20000	1.1	0.25%
Probe2n	高保真聚合酶	无	40	23000	20000	1.15	0.38%
Probe2d	高保真聚合酶	无	40	24000	20000	1.20	0.5%
Probe3	普通聚合酶	无	40	22000	20000	1.1	0.25%
Probe4	普通聚合酶	无	40	26000	20000	1.3	0.75%

Probe1	普通聚合酶	有	40	496000	20000	24.8	59.50%
Probe2	普通聚合酶	有	40	26000	20000	1.3	0.75%
Probe3	普通聚合酶	有	40	22000	20000	1.1	0.25%
Probe4	普通聚合酶	有	40	506000	20000	25.3	60.75%

表 15 中所涉及的捕获延伸反应体系参见表 2 和表 3，其中，高保真聚合酶为 2X 高保真 DNA 聚合酶 PCR Master Mix（DNA 聚合酶具有 3'-5'外切酶活性），普通聚合酶为 2X DNA 聚合酶 PCR Master Mix（DNA 聚合酶不具有 3'-5'外切酶活性），所使用的 cfDNA 样本为 QC 样本，延伸效率检测的方法参见步骤 3。

从测试结果可以看出：1) 3'端封闭修饰的探针在高保真聚合酶的作用下结合模板后，封闭基团被切除，探针被活化，从而可以对目标序列进行有效延伸；2) 热稳定性核酸酶与普通聚合酶联用，针对 RNA 基团修饰的探针也能起到活化效果，增强了本发明富集方法的适用性；3) 不同取代基团的选择对于探针的捕获效果有明显差异，将探针 3'端羟基取代为 C3 Spacer 得到的效果最佳，与其它取代基团相比具有明显优势。故此，本实施例从结果 2 开始的数据主要是针对将 3'端羟基取代为 C3 Spacer 的探针得到的结果数据。

## 结果 2、不同延伸循环数对部分探针延伸结果的影响。

表 16. 不同延伸循环数的测试结果（QC 样本）

探针	DNA 聚合酶类型	核酸酶	循环数	捕获后 EGFR 输出拷贝数	捕获前 EGFR 输入拷贝数	捕获延伸倍率	捕获效率%
Probe1	高保真酶	无	40	482000	20000	24.1	57.75%
Probe2	高保真酶	无	40	566000	20000	28.3	68.25%
Probe3	高保真酶	无	40	524000	20000	26.2	63.00%
Probe4	高保真酶	无	40	482000	20000	24.1	57.75%
Probe1	高保真酶	无	20	276000	20000	13.8	64.00%
Probe2	高保真酶	无	20	244000	20000	12.2	56.00%
Probe3	高保真酶	无	20	234000	20000	11.7	53.50%
Probe4	高保真酶	无	20	248000	20000	12.4	57.00%
Probe1	高保真酶	无	1	34800	20000	1.74	74.00%
Probe2	高保真酶	无	1	34400	20000	1.72	72.00%
Probe3	高保真酶	无	1	36400	20000	1.82	82.00%
Probe4	高保真酶	无	1	31400	20000	1.57	57.00%

表 16 中所涉及的捕获延伸反应体系参见表 2，所使用的 cfDNA 样本为 QC 样本，延伸效

率检测的方法参见步骤 3。

测试结果表明，通过循环数的增加可以有效提高捕获延伸倍率。

### 结果 3、dNTP 是否偶联标记分子对捕获延伸结果的影响。

表 17. 捕获延伸体系中掺入 Biotin-dCTP 的测试结果（QC 样本）

探针	DNA 聚合酶类型	Biotin-dCTP 掺入的纯化体系	循环数	biotin 标记 EGFR 输出拷贝数	捕获前 EGFR 输入拷贝数	捕获延伸倍率	捕获效率%
Probe1	高保真酶	无	40	0	20000	0	/
Probe2	高保真酶	无	40	0	20000	0	/
Probe3	高保真酶	无	40	0	20000	0	/
Probe4	高保真酶	无	40	0	20000	0	/
Probe1	高保真酶	有	40	212000	20000	10.6	24.00%
Probe2	高保真酶	有	40	183000	20000	9.15	20.38%
Probe3	高保真酶	有	40	198000	20000	9.9	22.25%
Probe4	高保真酶	有	40	187000	20000	9.35	20.88%

表 17 中所涉及的捕获延伸反应体系参见表 2 和表 4，所使用的 cfDNA 样本为 QC 样本，延伸效率检测的方法参见步骤 3。

从结果可以看出，将偶联了 Biotin 的 dCTP 掺入捕获延伸体系，可有效标记延伸产物，提高捕获效率。

### 结果 4、dNTP 是否偶联标记分子对连接结果的影响。

表 18. 捕获延伸体系中掺入 Biotin-dCTP 后连接反应的测试结果（QC 样本）

探针	Biotin-dCTP 掺入的纯化体系	接头	连接酶	连接后纯化	延伸输入分子拷贝数	连接输出分子拷贝数	连接效率
Probe1	无	ABar-X2	热稳定 RNA 连接酶	硅胶柱	77120	84	0.11%
Probe2	无	ABar-X2	热稳定 RNA 连接酶	硅胶柱	90560	1452	1.60%
Probe3	无	ABar-X2	热稳定 RNA 连接酶	硅胶柱	83840	1878	2.24%
Probe1	有	ABar-X2	热稳定 RNA 连接酶	硅胶柱	28267	328	1.16%

Probe2	有	ABar-X2	热稳定 RNA 连接酶	硅胶柱	24400	13257	54.33%
Probe3	有	ABar-X2	热稳定 RNA 连接酶	硅胶柱	26400	17328	65.64%

表 18 中所涉及连接反应体系参见表 8，其中，延伸产物为根据表 2 或表 4 中的体系制备获得的延伸产物，其中，各体系均使用高保真聚合酶，所述高保真聚合酶为 2X 高保真 DNA 聚合酶 PCR Master Mix（DNA 聚合酶具有 3'-5' 外切酶活性），所使用的 cfDNA 样本为 QC 样本，连接效率检测的方法参见步骤 5。

结果表明，使用将 3' 端羟基替换为 C3 Spacer 的探针，配合使用偶联了 biotin 的 dCTP 和针对 biotin 的纯化系统，能获得最好的连接效果；而未经修饰的探针或者 dNTP 未结合 biotin 的反应体系，几乎无法连接。

#### 结果 5、不同 5' 端修饰的探针和不同连接酶对连接效率的影响。

表 19. 连接体系中不同 5' 端修饰的探针和不同连接酶的测试结果（QC 样本）

探针	Biotin-dCTP 掺入的纯化体系	接头	连接酶	连接后纯化	延伸输入分子拷贝数	连接输出分子拷贝数	连接效率
Probe2	有	ABar-X2	热稳定 RNA 连接酶	硅胶柱	24400	13257	54.33%
Probe3	有	ABar-X2	热稳定 RNA 连接酶	硅胶柱	26400	17328	65.64%
Probe2	有	ABar-X2	T4 RNA	硅胶柱	24400	2264	9.28%
Probe3	有	ABar-X2	T4 RNA	硅胶柱	26400	2828	10.71%
Probe2	有	ABar-X3	5'AppDNA/RNA 连接酶	硅胶柱	24400	2073	8.50%
Probe3	有	ABar-X3	5'AppDNA/RNA 连接酶	硅胶柱	26400	2154	8.16%

表 19 中所涉及连接反应体系参见表 8、表 9 和表 10，其中，延伸产物为根据表 4 中的体系制备获得的延伸产物，所使用的 cfDNA 样本为 QC 样本，连接效率检测的方法参见步骤 5。

结果表明，使用 5' 端磷酸化修饰的探针配合热稳定 RNA 连接酶能显著提高连接效果。

#### 结果 6、不同 3' 端修饰的接头 DNA 对连接效率的影响。

表 20. 连接体系中不同 3' 端修饰的接头 DNA 的连接测试结果（QC 样本）

探针	Biotin-dCTP 掺入的纯化体系	接头	连接酶	连接后 纯化	输入分子 拷贝数	连接分子 拷贝数	连接效率
Probe2	有	ABar-X2	热稳定 RNA 连接 酶	硅胶柱	24400	13257	54.33%
Probe3	有	ABar-X2	热稳定 RNA 连接 酶	硅胶柱	26400	17328	65.64%
Probe2	有	ABar-X1	热稳定 RNA 连接 酶	硅胶柱	24400	4321	17.71%
Probe3	有	ABar-X1	热稳定 RNA 连接 酶	硅胶柱	26400	3875	14.68%

表 20 中所涉及连接反应体系参见表 8，其中，延伸产物为根据表 4 中的体系制备获得的延伸产物，所使用的 cfDNA 样本为 QC 样本，连接效率检测的方法参见步骤 5。

结果表明，使用 3'端封闭修饰的接头 DNA 降低了自连与错连，显著提高了连接效率。

### 结果 7、连接后的不同处理对产物检出率的影响。

表 21. 连接后不同处理的产物检出率结果（QC 样本）

探针	Biotin-dCTP 掺入的纯化体系	接头	连接酶	连接后 处理	输入分子 拷贝数	连接分子 拷贝数	产物检出率
Probe2	有	ABar-X2	热稳定 RNA 连接 酶	硅胶柱 纯化	24400	13257	54.33%
Probe3	有	ABar-X2	热稳定 RNA 连接 酶	硅胶柱 纯化	26400	17328	65.64%
Probe2	有	ABar-X2	热稳定 RNA 连接 酶	加热	24400	10567	43.31%
Probe3	有	ABar-X2	热稳定 RNA 连接	加热	26400	12385	46.91%

			酶				
Probe2	有	ABar-X2	热稳定 RNA 连接 酶	无处理	24400	2324	9.52%
Probe3	有	ABar-X2	热稳定 RNA 连接 酶	无处理	26400	2875	10.89%

表 20 中所涉及连接反应体系参见表 8，其中，延伸产物为根据表 4 中的体系制备获得的延伸产物，所使用的 cfDNA 样本为 QC 样本，连接后处理的具体参数参见表 11，连接效率检测的方法参见步骤 5。

结果表明，连接反应后，硅胶柱纯化或加热处理都可以有效去除或降低连接产物中残留的热稳定 RNA 连接酶的酶活，从而降低连接产物的背景噪音，提高产物检出率。

#### 结果 8、不同反应条件下样本的下机数据。

测序分析的条件设置见表 22，测序方法具体参见步骤 7，对应表 22 中条件编号的测序分析结果见表 23。cfDNA 中突变 DNA 的不同掺入率条件下的检出结果见表 24。

表 22. 测序分析的条件设置

条件编号	探针	DNA 聚合酶类型	核酸酶	循环数	Biotin-dCTP 掺入的纯化体系	接头	连接酶	连接后处理
1	Probe1	高保真酶	无	40	有	ABar-X2	热稳定 RNA 连接酶	硅胶柱
2	Probe2	高保真酶	无	40	有	ABar-X2	热稳定 RNA 连接酶	硅胶柱
3	Probe2p	高保真酶	无	40	有	ABar-X2	热稳定 RNA 连接酶	硅胶柱
4	Probe3	高保真酶	无	40	有	ABar-X2	热稳定 RNA 连接酶	硅胶柱
5	Probe4	普通聚合酶	有	40	有	ABar-X2	热稳定 RNA 连接酶	硅胶柱
6	Probe3	高保真酶	无	40	无	ABar-X2	热稳定 RNA 连接酶	硅胶柱
7	Probe3	高保真酶	无	40	有	ABar-X1	热稳定 RNA	硅胶柱

							连接酶	
8	Probe3	高保真酶	无	40	有	ABar-X2	T4RNA	硅胶柱
9	Probe3	高保真酶	无	40	有	ABar-X3	5'AppDNA/RNA 连接酶	硅胶柱
10	Probe3	高保真酶	无	40	有	ABar-X2	热稳定 RNA 连接酶	加热
11	Probe3	高保真酶	无	40	有	ABar-X2	热稳定 RNA 连接酶	无处理

表 22 中所涉及 PCR 扩增获得的连接产物是根据表 3、表 4 中的体系制备获得的延伸产物进一步根据表 8~表 10 的连接反应体系制备获得的连接产物，连接后处理的具体参数参见表 11。表 23 所涉的测序分析结果实验中，所使用的 cfDNA 样本为 QC 样本。在表 24 所涉及的 cfDNA 中突变 DNA 的不同掺入率条件实验中，对于同一条件编号的技术方案，仅使用的 cfDNA 样本中 NCI-H1975 细胞株的 DNA 的掺入量有所不同，其他实验步骤均与表 22 中相同。

表 23. 测序分析结果

条件编号	1	2	3	4	5	6	7	8	9	10	11
总 reads 数	367842	245740	284632	345320	299538	327782	243698	288673	296643	308712	312156
比对到 hg19 上的 reads 数	184	211600	2343	327080	275605	1052	30292	126641	70809	270555	35336
比对率	0.05%	86.11%	0.82%	94.72%	92.01%	0.32%	12.43%	43.87%	23.87%	87.64%	11.32%
目标区域 reads 数	61	125072	1039	192654	144252	402	14004	67436	34965	149671	16247
目标区域 reads 比例	33.40%	59.11%	44.37%	58.90%	52.34%	38.20%	46.23%	53.25%	49.38%	55.32%	45.98%

表 24. cfDNA 中突变 DNA 的不同掺入率条件下的检出结果

条件编号	1	2	3	4	5	6	7	8	9	10	11
未掺入	0	0	0	0	0	0	0	0	0	0	0
0.01%掺入	0	0.012%	0	0.021%	0.009%	0	0	0	0	0.019%	0
0.1%掺入	0	0.13%	0	0.12%	0.14%	0	0.08%	0.11%	0.09%	0.12%	0.08%
1%掺入	0	1.15%	0.48%	1.08%	0.83%	0	1.12%	1.13%	1.11%	1.18%	1.13%
10%掺入	2.30%	10.25%	9.53%	10.22%	9.89%	5.21%	10.13%	10.23%	9.54%	10.22%	10.15%

由表 23 和表 24 可见, 编号为 2, 4, 5, 10 的反应条件下, 测序比对率高, 检出灵敏度高, 可以在突变掺入率低至 0.01% 的 cfDNA 样本中检测出突变。结果表明, 探针 3' 羟基的封闭修饰至关重要, 可以防止探针自连或与接头相连, 大幅降低背景, 而结合模板后的探针被切除修饰基团后还能有效延伸; 偶联了标记分子的 dNTP 及其纯化系统, 能大幅提高建库产物纯度, 对本发明的技术方案提供了有效补充。另外, 热稳定性 RNA 连接酶的选用, 接头 DNA 3' 端的封闭, 以及连接后的处理, 也都能起到提高建库产物纯度和检测灵敏度的作用, 分别/共同组成了本发明的技术方案。

## 实施例 2

实施例 2 中所用寡核苷酸序列如表 25 所示。

表 25

名称	SEQ ID NO:	序列 (5'→3')
Probe5	17	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGAT</i> TAGTTGGATGGGATTATTT-C3 Spacer
Probe6	18	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGAT</i> TTTTTTTTGGGAGTTTAAATAAAGATTA -C3 Spacer
Probe7	19	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGAT</i> TTTAAATAGAGTTAGTTTTAGTTTTT-C3 Spacer
ABar-X2	10	<i>PO<sub>4</sub>-GGATCCNNNNNNGTTCTCCTTA</i> <i>CTGAGTCGGAGACACGCAGGGATGAGATGG</i> -C3 Spacer
AF	15	GTCTCAGCCTCTCTATGGGCAGTCGGTGAT
AR	16	CCATCTCATCCCTGCGTGTCTCCGACTCAG

此实施例中的探针包含通用序列, 即斜体标示的碱基, 此处为 Ion Torrent 测序系统中的 P1 序列。本实施例所用的 DNA 样本为重亚硫酸盐处理后的 DNA 样本。其中 Probe5 (SEQ ID NO17) 针对基因 SEPT9 的一个甲基化区域, Probe6 (SEQ ID NO18) 针对基因 NDRG4 的一个甲基化区域, Probe7 (SEQ ID NO19) 针对基因 BMP3 的一个甲基化区域。

接头 ABar-X2 (SEQ ID NO10) 中包含通用序列, 即斜体标示的碱基, 此处为 Ion Torrent 测序系统中的 A 序列; 下划线部分的序列为 barcode, 在不同的样本中可以被替换成不同的 barcodes; “N” 部分为分子标签序列; 其中, NNNNNN 为随机序列, 用以标记同一样本中不同的延伸产物分子; 接头 ABar-X2 的 5' 端为磷酸化修饰。



AF (SEQ ID NO15) 和 AR (SEQ ID NO16) 分别是用来扩增文库的 PCR 前、后引物，与文库中的通用序列，即上述 P1 序列和 A 序列匹配。

### 主要试剂与材料

重亚硫酸盐转化试剂盒购自 Promega，Agencourt AMPure 磁珠磁珠购自贝克曼。其他主要试剂与实施例 1 相同。

### 步骤 1、样本准备

抽提健康人血浆样本与肠癌组织标本 5 例，通过 Qubit 荧光定量仪对 DNA 样本定量。将肠癌组织标本 DNA 打断成 160bp 左右的片段，按照 10%和 5%的比例掺入到健康人血浆样本中，分别构建 5%掺入率和 10%掺入率的样本，同时以健康人血浆样本作为空白对照。使用 Promega 的重亚硫酸盐转化试剂盒对健康人血浆样本进行转化。转化后的 DNA 定量后配制成 1000 copies/ $\mu$ L 的浓度。

### 步骤 2、捕获延伸

按表 26 配制捕获延伸反应体系。

表 26

组分	体积 ( $\mu$ L)	终浓度
2X 高保真 DNA 聚合酶 PCR Master Mix (DNA 聚合酶具有 3'-5'外切酶活性)	25	1 $\times$
Probe 5 (10 $\mu$ M)	0.5	100 nM
Probe 6 (10 $\mu$ M)	0.5	100 nM
Probe 7 (10 $\mu$ M)	0.5	100 nM
cfDNA 样本	20	20000 拷贝
H <sub>2</sub> O	3.5	/
Total	50	/

捕获延伸 PCR 程序如表 27 所示。

表 27

循环数	温度	时间
1 cycle	98 $^{\circ}$ C	60s
40 cycles	98 $^{\circ}$ C	10 S
	62 $^{\circ}$ C	20 S
	72 $^{\circ}$ C	20 S
1 cycle	72 $^{\circ}$ C	5min

捕获延伸完成后对产物进行硅胶柱纯化，用 60 $\mu$ l 洗脱缓冲液洗脱。表 4 的捕获延伸产物

在硅胶柱纯化后，用链霉亲和素包被的磁珠进一步纯化，并最终溶解于 60ul 洗脱缓冲液中。

### 步骤 3、单链连接

按表 28 配制的连接反应体系和条件进行连接反应。

表 28

组分	体积 (μL)	终浓度	连接条件
10X 连接酶缓冲液	4	1X	60 度 1 小时
50 mM MnCl <sub>2</sub>	2	2.5 mM	
5 M Betaine	8	1 M	
热稳定 RNA 连接酶(100 U/μL)	2	5 U/μL	
ABar-X2(200 nM)	2	10 nM	
延伸产物	8		
50%PEG	14	17.5%	
Total	40	/	

连接后，用硅胶柱纯化连接产物，最后用 50ul 洗脱缓冲液洗脱。

### 步骤 4、文库扩增

按表 29 的反应体系和条件对连接产物进行 PCR 扩增。

表 29

组分		体积
连接产物		30μL
5× Q5 Buffer		10μL
dNTP(10mM each)		1μL
Q5高保真DNA聚合酶(2 U/μL)		1μL
AF (10μM)		1.5μL
AR (10μM)		1.5μL
ddH <sub>2</sub> O		5μL
总体积		50μL
温度(°C)	时间(秒)	循环数

98	30	1
98	10	12
72	30	
72	300	1

反应结束后，取 80uL 的 Agencourt AMPure 磁珠对 PCR 产物进行纯化，再将纯化产物溶于 30uL 洗脱缓冲液洗脱中。至此，待上机测序的文库构建完成。

### 步骤 5、测序与数据分析

将制备好的文库在 Ion Proton 上测序，操作步骤包括油包水 PCR、文库富集、芯片加样和上机测序，具体操作流程详见 Ion PITM Hi-QTM OT2 200 Kit 说明书和 Ion PITM Hi-QTM Sequencing 200 Kit 说明书。

测序结果分析见表 30 与表 31。

表 30. 测序结果 1

样本类型	总 reads 数 (M)	目标区域 reads 数 (M)	总比对率	目标区域 reads 总特异性
空白对照	1.261	0.448	74.35%	47.84%
5%掺入率	1.376	0.491	79.78%	44.74%
10%掺入率	1.508	0.513	75.42%	45.10%

表 31. 不同肠癌组织标本 DNA 掺入率下各基因甲基化位点占比

基因名称	基因上的甲基化位点	空白样本	5%掺入率	10%掺入率
SEPT9	chr17: 75369623	1.89%	4.03%	5.82%
	chr17: 75369630	1.52%	5.04%	6.18%
	chr17: 75369657	1.35%	4.97%	5.55%
	chr17: 75369663	0.65%	3.67%	5.50%
BMP3	chr4: 81952061	3.01%	3.10%	6.63%
	chr4: 81952065	3.72%	4.22%	6.86%
	chr4: 81952078	2.48%	2.69%	5.99%
	chr4: 81952099	2.72%	3.57%	5.67%
NDRG4	chr16: 58497230	1.66%	2.92%	4.74%
	chr16: 58497236	1.41%	3.69%	4.19%
	chr16: 58497239	1.43%	3.00%	4.67%
	chr16: 58497251	3.17%	5.81%	5.89%

	chr16: 58497265	3.15%	5.88%	6.48%
--	-----------------	-------	-------	-------

由表 30、表 31 可见，掺入 5%、10% 肠癌组织 DNA 的样本与空白样本相比，甲基化程度有明显不同，10% 掺入样本的甲基化程度最高，空白样本的甲基化程度最低，与预期相符。结果表明，使用本发明的富集方法，对经重亚硫酸盐处理的 DNA 片段化样本的目标区域也能很好的进行富集，并应用于基因甲基化程度以及与甲基化程度高度相关的疾病的分析。

**实施例 3**

实施例 3 和 4 中所用寡核苷酸序列如表 32 所示：

表 32

名称	SEQ ID NO.	序列 (5'→3')
F-SRY	20	TTCCAGGAGGCACAGAAATTAC
R-SRY	21	CTTCCGACGAGGTCGATACT
MGB-SRY	22	CATGCACAGAGAGAAATACCCGAAT
F858	23	CAGATTTTGGGCG
ER1	13	TTTGCCCTCCTTCTGC
EM1	14	AACTGCTGGGTGCGGA
F787	24	GCGTGATGAGT
R787	25	GCGTGGACAACCCCCAC
MGB787	26	CACGGTGGAGGTGAGGC
Uni-R	27	CCATCTCATCCCTGCGT
Uni-MGB	28	TCCGACTCAGTAAGGAGAACGA
Probe 8	29	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCAATATTGTCTTT GTGTTCCCGGACATAGTCCT-C3 Spacer
Probe 9	30	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCTATGGCCATTCTT CCAGGAGGCACAGAAATTACT-C3 Spacer
ABar-X2	10	PO4- GGATCCNNNNNGTTCTCCTTACTGAGTCGGAGACACGCAGGG ATGAGATGG -C3 Spacer

主要试剂与材料：

与实施例 1 相同，qPCR 所用校准品按照分子克隆常规方法构建。

**步骤 1、样本准备**

非小细胞肺癌细胞株 NCI-H1975 同时含有 EGFR L858R 位点的突变和 Q787 位点的 SNP。使用 XX 试剂盒 提取非小细胞肺癌细胞株 NCI-H1975 与正常人白细胞的基因组 DNA，超声打断基因组 DNA 到 100-300bp 的分子量范围。通过 Qubit 荧光定量仪定量后，按照 1%，0.1%，0.03%，0.01%，0% 的比例将 NCI-H1975 的基因组 DNA 掺入到正常人白细胞的基因组 DNA 中，并通过本发明实施例 1 中的步骤 2 到步骤 6 对样品进行捕获延伸、连接和扩库准备。

**步骤 2、不同特异性引物/探针组合的 qPCR 检测**

EGFR 基因 L858R 位点三特异性、双特异性、单特异性检测体系分别如表 33-35 所示。

表 33. L858R 位点三特异性检测体系

试剂	用量 (μL)	终浓度
2×Taqman Mix	10	1×
F858 (10 μM)	0.6	300 nM
R858 (10 μM)	0.6	300 nM
MGB858 (10μM)	0.2	100 nM
H <sub>2</sub> O	6.6	/
校准品或待测样本	2	/
Total	20	/

表 34. L858R 位点双特异性检测体系:

试剂	用量 (μL)	终浓度
2×Taqman Mix	10	1×
F858 (10 μM)	0.6	300 nM
Uni-R (10 μM)	0.6	300 nM
MGB858 (10μM)	0.2	100 nM
H <sub>2</sub> O	6.6	/
校准品或待测样本	2	/
Total	20	/

表 35. L858R 位点单特异性检测体系:

试剂	用量 (μL)	终浓度
2×Taqman Mix	10	1×
F858 (10 μM)	0.6	300 nM
Uni-R (10 μM)	0.6	300 nM
Uni-MGB (10μM)	0.2	100 nM
H <sub>2</sub> O	6.6	/
校准品或待测样本	2	/
Total	20	/

EGFR 基因 Q787Q 位点三特异性、双特异性、单特异性检测体系分别如表 36-38 所示。

表 36. Q787Q 位点三特异性检测体系:

试剂	用量 (μL)	终浓度
2×Taqman Mix	10	1×
F787 (10 μM)	0.6	300 nM
R787 (10 μM)	0.6	300 nM
MGB787 (10μM)	0.2	100 nM
H <sub>2</sub> O	6.6	/
校准品或待测样本	2	/
Total	20	/

表 37. Q787Q 位点双特异性检测体系:

试剂	用量 (μL)	终浓度
----	---------	-----

2×Taqman Mix	10	1×
F787 (10 μM)	0.6	300 nM
Uni-R (10 μM)	0.6	300 nM
MGB787 (10μM)	0.2	100 nM
H <sub>2</sub> O	6.6	/
校准品或待测样本	2	/
Total	20	/

表 38. Q787Q 位点单特异性检测体系:

试剂	用量 (μL)	终浓度
2×Taqman Mix	10	1×
F787 (10 μM)	0.6	300 nM
Uni-R (10 μM)	0.6	300 nM
Uni-MGB (10μM)	0.2	100 nM
H <sub>2</sub> O	6.6	/
校准品或待测样本	2	/
Total	20	/

qPCR 检测程序如表 39 所示。

表 39.

循环数	温度	时间
1 cycle	95°C	4 min
40 cycles	95°C	10 s
	61°C	30 s

## 实验结果及分析

**结果 1.** 不同特异性引物/探针组合 L858R 位点检测结果分别如表 40-42 所示。其中三特异性、双特异性、单特异性检测体系分别对应表 33-35 的反应体系。

表 40. L858R 位点三特异性检测结果 ( $\Delta CT = ABS(\text{mean } Q_i - Q_5)$ )

掺入率	CT1	CT2	CT 均值	$\Delta CT$
1%	26.13	26.49	26.31	11.715
0.1%	30.35	30.56	30.46	7.570
0.03%	33.43	33.85	33.64	4.385
0.01%	37.34	36.52	36.93	1.095
0	38.52	37.53	38.03	0

表 41. L858R 位点双特异性检测结果

掺入率	CT1	CT2	CT 均值	$\Delta CT$
1%	24.12	24.14	24.13	13.040
0.1%	28.73	28.24	28.49	8.685

0.03%	30.82	31.32	31.07	6.100
0.01%	33.43	34.18	33.81	3.365
0	37.32	37.02	37.17	0

表 42. L858R 位点单特异性检测结果

掺入率	CT1	CT2	CT 均值	$\Delta$ CT
1%	23.64	23.52	23.58	12.850
0.1%	27.38	27.02	27.20	9.230
0.03%	29.24	29.87	29.56	6.875
0.01%	32.29	31.58	31.94	4.495
0	36.24	36.62	36.43	0

**结果 2.** 不同特异性引物/探针组合 Q787Q 位点检测结果分别如表 43-45 所示。其中三特异性、双特异性、单特异性检测体系分别对应表 36-38 的反应体系。

表 43. Q787Q 位点三特异性检测结果

掺入率	CT1	CT2	CT 均值	$\Delta$ CT
1%	26.09	25.85	25.97	11.055
0.1%	30.52	30.82	30.67	6.355
0.03%	34.07	33.59	33.83	3.195
0.01%	36.82	36.43	36.63	0.400
0	37.14	36.91	37.03	0

表 44. Q787Q 位点双特异性检测结果

掺入率	CT1	CT2	CT 均值	$\Delta$ CT
1%	25.45	25.48	25.47	10.820
0.1%	29.01	29.48	29.25	7.040
0.03%	31.77	32.11	31.94	4.345
0.01%	34.12	34.98	34.55	1.735
0	36.05	36.52	36.29	0

表 45. Q787Q 位点单特异性检测结果

掺入率	CT1	CT2	CT 均值	$\Delta$ CT
1%	24.44	24.17	24.31	11.950
0.1%	27.84	28.19	28.02	8.240
0.03%	30.45	31.01	30.73	5.525
0.01%	32.98	32.45	32.72	3.540
0	36.03	36.48	36.26	0

**结论:** 针对不同掺入率的参考品, 双特异性体系及单特异性这两种引物/探针的体系,

与三特异性体系相比,  $\Delta CT$  值更大, 具有更高的有效分辨率, 如果以  $\Delta CT=3$  作为 cut off, 针对 L858R 位点, 单特异性体系、双特异性体系与三特异性体系的检出限分别是 0.01%, 0.01% 与 0.03%; 针对 Q787 位点, 检出限分别是 0.01%, 0.03% 与 0.01%。由此可见, 在掺入率低至 0.03% 的情况下, 单特异性体系、双特异性体系仍具有可靠的检出率。

#### 实施例 4.

本实施例所用寡核苷酸序列如表 32 所示。主要试剂与材料: 与实施例 3 相同。

#### 步骤 1、样本准备

用血浆游离核酸提取试剂盒分别提取男性和女性的血浆游离 DNA, 用 Qubit 荧光定量仪定量后, 按照 1%, 0.1%, 0.03%, 0.01%, 0% 的比例将男性 cfDNA 掺入到女性 cfDNA 样本中, 同时以女性 cfDNA 作为空白对照, 并通过本发明实施例 1 中的步骤 2 到步骤 6 对样品进行捕获延伸、连接和扩库准备。

#### 步骤 2、不同特异性引物/探针组合的 qPCR 检测

男性特有 SRY 基因三特异性、双特异性、单特异性检测体系分别如表 46-48 所示。

表 46. 男性特有 SRY 基因三特异性检测体系:

试剂	用量 ( $\mu\text{L}$ )	终浓度
2 $\times$ Taqman Mix	10	1 $\times$
F-SRY (10 $\mu\text{M}$ )	0.6	300 nM
R-SRY (10 $\mu\text{M}$ )	0.6	300 nM
MGB-SRY (10 $\mu\text{M}$ )	0.2	100 nM
H <sub>2</sub> O	6.6	/
校准品或待测样本	2	/
Total	20	/

表 47. 男性特有 SRY 基因二特异性检测体系:

试剂	用量 ( $\mu\text{L}$ )	终浓度
2 $\times$ Taqman Mix	10	1 $\times$
F-SRY (10 $\mu\text{M}$ )	0.6	300 nM
Uni-R (10 $\mu\text{M}$ )	0.6	300 nM
MGB-SRY (10 $\mu\text{M}$ )	0.2	100 nM
H <sub>2</sub> O	6.6	/
校准品或待测样本	2	/
Total	20	/

表 48. 男性特有 SRY 基因单特异性检测体系:

试剂	用量 ( $\mu\text{L}$ )	终浓度
2 $\times$ Taqman Mix	10	1 $\times$
F-SRY (10 $\mu\text{M}$ )	0.6	300 nM
Uni-R (10 $\mu\text{M}$ )	0.6	300 nM
Uni-MGB (10 $\mu\text{M}$ )	0.2	100 nM
H <sub>2</sub> O	6.6	/



校准品或待测样本	2	/
Total	20	/

qPCR 检测程序如表 49 所示。

表 49.

循环数	温度	时间
1 cycle	95°C	4 min
40 cycles	95°C	10 s
	61°C	30 s

### 实验结果及分析

**结果：**不同特异性引物/探针组合男性特有 SRY 基因检测结果分别如表 50-52 所示。其中三特异性、双特异性、单特异性检测体系分别对应表 46-48 的反应体系。

表 50. 男性特有 SRY 基因三特异性检测结果 ( $\Delta CT = ABS(\text{mean } Q_i - Q_5)$ )

掺入率	CT1	CT2	CT 均值	$\Delta CT$
1%	31.23	32.08	31.66	13.345
0.1%	35.71	35.94	35.83	9.175
0.03%	37.85	38.02	37.94	7.065
0.01%	39.25	39.52	39.39	5.615
0	ND	ND	45.00	0

表 51. 男性特有 SRY 基因双特异性检测结果

掺入率	CT1	CT2	CT 均值	$\Delta CT$
1%	30.55	30.99	30.77	14.230
0.1%	34.23	34.98	34.61	10.395
0.03%	36.32	37.94	37.13	7.870
0.01%	39.48	39.12	39.30	5.700
0	ND	ND	45.00	0

表 52. 男性特有 SRY 基因单特异性检测结果

掺入率	CT1	CT2	CT 均值	$\Delta CT$
1%	29.38	29.68	29.53	15.470
0.1%	33.28	33.89	33.59	11.415
0.03%	35.45	35.82	35.64	9.365
0.01%	38.21	37.89	38.05	6.950
0	ND	ND	45.00	0

**结论:** SRY 作为仅在 Y 染色体上的基因, 三种方法检测的灵敏度都较高。如果以  $\Delta CT=3$  作为 cut off, 单、双、三特异性体系都能达到 0.01% 的检出灵敏度。但是从具体的 CT 值上来判断, 单特异性体系的检出率更高。

综上所述, 本发明有效克服了现有技术中的种种缺点而具高度产业利用价值。

上述实施例仅例示性说明本发明的原理及其功效, 而非用于限制本发明。任何熟悉此技术的人士皆可在不违背本发明的精神及范畴下, 对上述实施例进行修饰或改变。因此, 举凡所属技术领域中具有通常知识者在未脱离本发明所揭示的精神与技术思想下所完成的一切等效修饰或改变, 仍应由本发明的权利要求所涵盖。

## 权利要求

1、一种基因目标区域富集方法，包括：

(1) 通过特异性探针扩增包括目标区域的片段化 DNA，以提供捕获延伸产物，所述的特异性探针包括与所述片段化 DNA 的目标区域互补的序列，所述特异性探针的 3' 末端核苷酸被修饰，用于阻止所述特异性探针的 3' 末端发生连接反应；

(2) 将步骤 (1) 所提供的捕获延伸产物的 3' 端连接接头 DNA，以提供连接产物。

2、如权利要求 1 所述的基因目标区域富集方法，其特征在于，所述步骤 (1) 中，所述片段化 DNA 包含双链 DNA、单链 DNA 和 cDNA，所述片段化 DNA 的长度为 25~200bp；

和/或，所述步骤 (1) 的扩增体系中包括特异性探针、DNA 聚合酶和 dNTP。

3、如权利要求 2 所述的基因目标区域富集方法，其特征在于，所述 DNA 聚合酶具有 3'-5' 外切酶活性；

和/或，所述 dNTP 还偶联有标记分子，所述标记分子优选为生物素。

4、如权利要求 1 所述的基因目标区域富集方法，其特征在于，所述步骤 (1) 的扩增体系中还包括用于切除结合目标区域后所述特异性探针的 3' 末端修饰基团的活性物质，所述活性物质优选为核酸酶；

和/或，所述特异性探针还包括能被测序系统识别的通用序列；

和/或，所述特异性探针的 3' 末端核苷酸的 3 位羟基被取代；

和/或，所述特异性探针的 3' 末端的取代基团选自氢原子、C3 Spacer 基团、C6 Spacer 基团、磷酸基团或氨基基团；

和/或，所述特异性探针 3' 端尾部区域包含错配碱基。

5、如权利要求 1 所述的基因目标区域富集方法，其特征在于，所述步骤 (2) 的连接体系中包括单链连接酶，所述单链连接酶优选为 T4 RNA 连接酶或热稳定性 RNA 连接酶；

和/或，所述接头 DNA 的 5' 末端核苷酸被修饰、且在步骤 (2) 的反应温度下为单链结构，优选的，接头 DNA 的 5' 末端核苷酸被磷酸基团或腺苷基团取代；

和/或，所述接头 DNA 为 5' 端区域具有黏性末端的部分双链结构；

和/或，所述接头 DNA 包括能够被测序系统识别的通用序列、样本标签序列、分子标签序列中的一种或多种的组合。

- 6、如权利要求 1 所述的基因目标区域富集方法，其特征在于，所述步骤（1）中，还包括纯化捕获延伸产物，捕获延伸产物的纯化方法优选选自磁珠纯化；  
和/或，所述步骤（2）中，还包括纯化连接产物，连接产物的纯化方法优选选自硅胶柱纯化和/或加热处理。
- 7、如权利要求 1 所述的基因目标区域富集方法，其特征在于，还包括：  
（3）扩增步骤（2）所提供的连接产物，优选的，所述步骤（3）中，PCR 扩增引物具有与所述特异性探针的通用序列和/或所述接头 DNA 的通用序列相配合。
- 8、如权利要求 7 所述的基因目标区域富集方法，其特征在于，还包括：  
（4）对扩增后的连接产物进行测序，以提供目标区域的测序结果。
- 9、如权利要求 1 所述的基因目标区域富集方法，还包括：  
（5）用检测引物 1、检测引物 2 和探针 3 检测步骤（2）提供的连接产物，其特征在于，所述检测引物 1、检测引物 2 和探针 3 中，至少其一包含基因特异性序列；  
优选的，所述检测引物 1 包含基因特异性序列，所述检测引物 2 和探针 3 包含通用序列；  
和/或，所述检测引物 2 和/或探针 3 也包含基因特异性序列；  
优选的，所述探针 3 包含标记分子，且所述探针 3 的序列与检测引物 1 或 2 不互补。
- 10、如权利要求 1 所述的基因目标区域富集方法，其特征在于，所述基因目标区域富集方法用于基因检测。
- 11、一种用于富集片段化 DNA 目标区域的试剂盒，包括适用于如权利要求 1~9 任一权利要求所述的基因目标区域富集方法的特异性探针和接头 DNA。
- 12、如权利要求 11 所述的试剂盒，其特征在于，还包括以下的一种或多种：  
RNA 连接酶，偶联标记分子的 dNTP，DNA 聚合酶，核酸酶；  
和/或，还包括正向引物和反向引物，所述正向引物和反向引物具有与所述特异性探针的通用序列和所述接头 DNA 的通用序列中至少部分互补的序列。
- 13、如权利要求 12 所述的试剂盒，其特征在于，  
还包括检测引物 1、检测引物 2 和探针 3，三者中至少其一含有基因特异性序列。

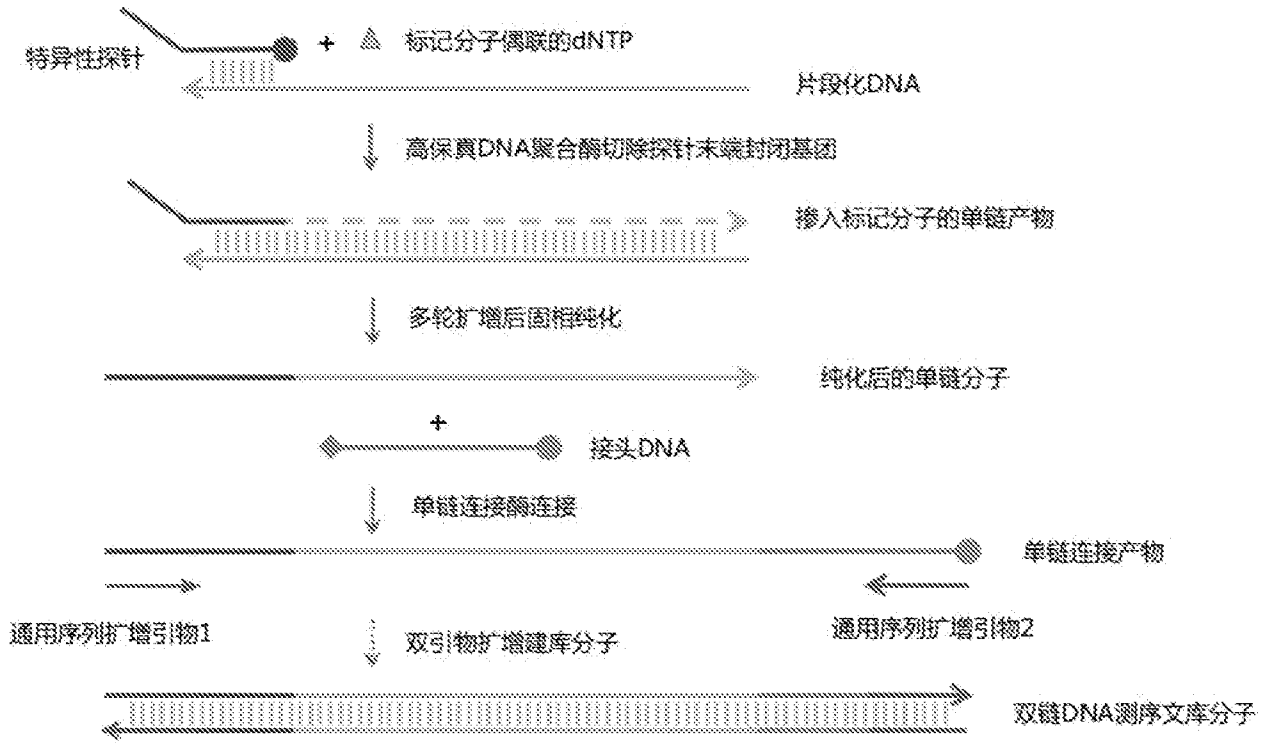


图 1



图 2

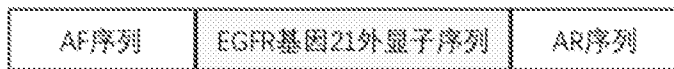


图 3

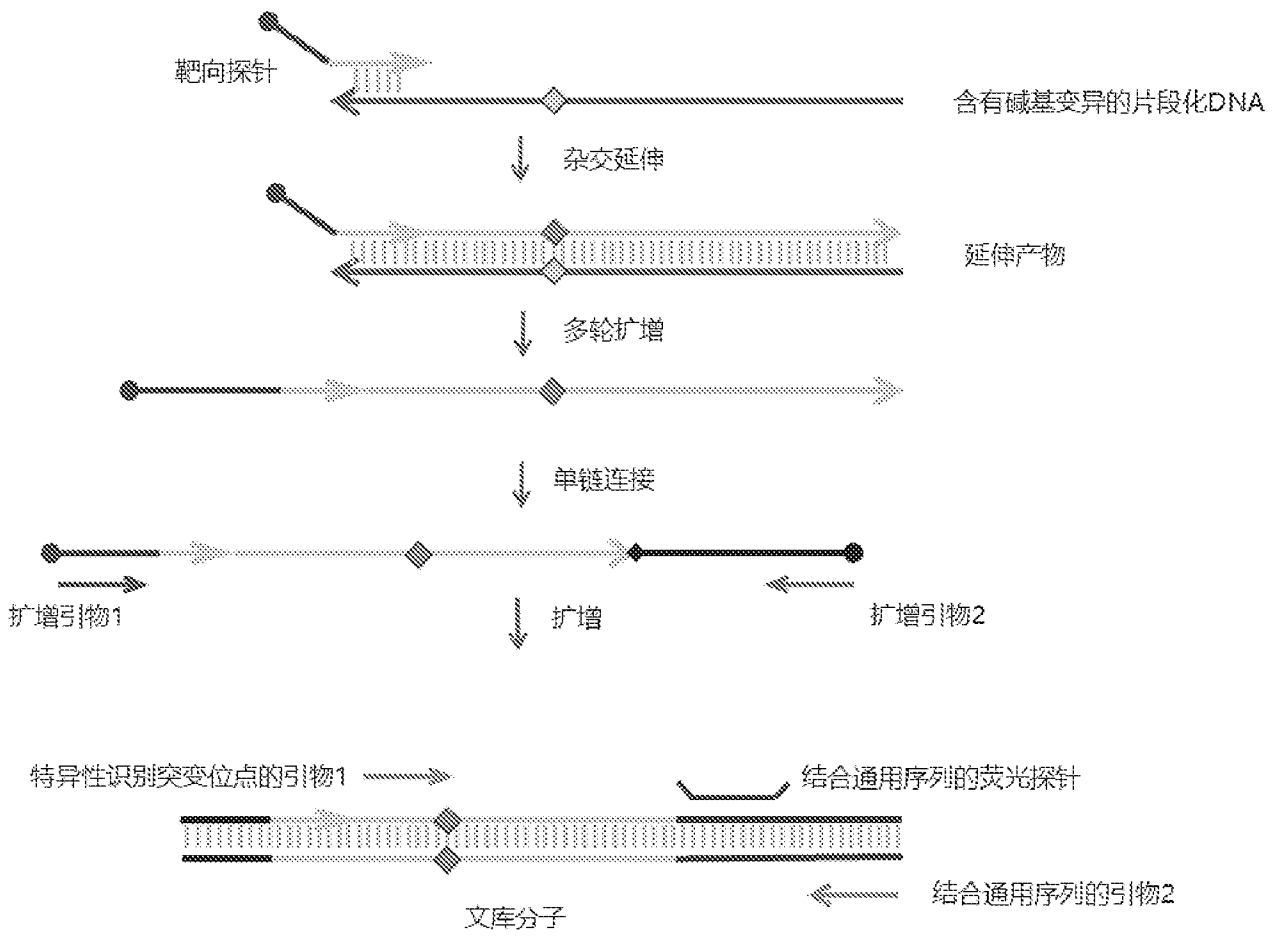


图 4

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/123891

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
C12Q 1/6806(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)		
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)		
CNMED, CNABS, CPEA, DWPI, SIPOABS, AUABS, TWMED, ILABS, TWABS, HKABS, MOABS, SGABS, CNKI, NCBI, EMBL, GoogleScholar: 富集, 捕获延伸, 探针, 扩增 capture-extension product CEP, Specific probe hybridization capture, Probe capture based enrichment		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 107922970 A (F.HOFFMANN-LA ROCHE AG) 17 April 2018 (2018-04-17) see entire document, especially the abstract	1-13
A	WO 2018028001 A1 (CHENGDU NUOEN BIOLOG TECHNOLOGY CO., LTD.) 15 February 2018 (2018-02-15) see entire document, especially the abstract	1-13
A	Johannes Dapprich et al. "The Next Generation of Target Capture Technologies-large DNA Fragment Enrichment and Sequencing Determines Regional Genomic Variation of High Complexity" <i>BMC Genomics</i> , Vol. 17, 09 July 2016 (2016-07-09), the abstract, and figure 1	1-13
<input 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		Date of mailing of the international search report
12 February 2020		06 March 2020
Name and mailing address of the ISA/CN		Authorized officer
China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088 China		
Facsimile No. (86-10)62019451		Telephone No.

**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/CN2019/123891**

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107922970	A	17 April 2018	WO	2017021449	A1	09 February 2017
				EP	3332024	A1	13 June 2018
				US	2017037459	A1	09 February 2017
				JP	2018521675	A	09 August 2018
				CA	2994601	A1	09 February 2017
<hr/>							
WO	2018028001	A1	15 February 2018	CN	106282161	A	04 January 2017
<hr/>							

国际检索报告

国际申请号

PCT/CN2019/123891

<p><b>A. 主题的分类</b></p> <p>C12Q 1/6806(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>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>CNMED, CNABS, CPEA, DWPI, SIPOABS, AUABS, TWMED, ILABS, TWABS, HKABS, MOABS, SGABS, CNKI, NCBI, EMBL, GoogleScholar:富集, 捕获延伸, 探针, 扩增capture-extension product CEP, Specific probe hybridization capture, Probe capture based enrichment</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>A</td> <td>CN 107922970 A (豪夫迈 罗氏有限公司) 2018年 4月 17日 (2018 - 04 - 17) 参见全文, 特别是摘要部分</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>WO 2018028001 A1 (CHENGDU NUOEN BIOLOG TECH CO LTD) 2018年 2月 15日 (2018 - 02 - 15) 参见全文, 特别是摘要部分</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>Johannes Dapprich等. "The next generation of target capture technologies-large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity" BMC Genomics, 第17卷, 2016年 7月 9日 (2016 - 07 - 09), 摘要, 图1</td> <td>1-13</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	A	CN 107922970 A (豪夫迈 罗氏有限公司) 2018年 4月 17日 (2018 - 04 - 17) 参见全文, 特别是摘要部分	1-13	A	WO 2018028001 A1 (CHENGDU NUOEN BIOLOG TECH CO LTD) 2018年 2月 15日 (2018 - 02 - 15) 参见全文, 特别是摘要部分	1-13	A	Johannes Dapprich等. "The next generation of target capture technologies-large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity" BMC Genomics, 第17卷, 2016年 7月 9日 (2016 - 07 - 09), 摘要, 图1	1-13
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求												
A	CN 107922970 A (豪夫迈 罗氏有限公司) 2018年 4月 17日 (2018 - 04 - 17) 参见全文, 特别是摘要部分	1-13												
A	WO 2018028001 A1 (CHENGDU NUOEN BIOLOG TECH CO LTD) 2018年 2月 15日 (2018 - 02 - 15) 参见全文, 特别是摘要部分	1-13												
A	Johannes Dapprich等. "The next generation of target capture technologies-large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity" BMC Genomics, 第17卷, 2016年 7月 9日 (2016 - 07 - 09), 摘要, 图1	1-13												
<p><input 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:</p> <p>"A" 认为不特别相关的表示了现有技术一般状态的文件</p> <p>"E" 在国际申请日的当天或之后公布的在先申请或专利</p> <p>"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)</p> <p>"O" 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>"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</p> <p>"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</p> <p>"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</p> <p>"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</p> <p>"&amp;" 同族专利的文件</p>														
<p>国际检索实际完成的日期</p> <p>2020年 2月 12日</p>		<p>国际检索报告邮寄日期</p> <p>2020年 3月 6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>张艳霞</p> <p>电话号码 62089438</p>												

## 第I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/CN2019/123891

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	107922970	A	2018年 4月 17日	WO	2017021449	A1	2017年 2月 9日
				EP	3332024	A1	2018年 6月 13日
				US	2017037459	A1	2017年 2月 9日
				JP	2018521675	A	2018年 8月 9日
				CA	2994601	A1	2017年 2月 9日
WO	2018028001	A1	2018年 2月 15日	CN	106282161	A	2017年 1月 4日