

(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
2025 年 1 月 9 日 (09.01.2025)



(10) 国际公布号
WO 2025/007252 A1

(51) 国际专利分类号:

C12Q 1/6869 (2018.01)

(21) 国际申请号:

PCT/CN2023/105577

(22) 国际申请日:

2023 年 7 月 3 日 (03.07.2023)

(25) 申请语言:

中文

(26) 公布语言:

中文

(71) 申请人: 深圳华大智造科技股份有限公司
(**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: 罗银玲 (**LUO, Yinling**); 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)。徐崇钧 (**XU, Chongjun**); 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)。龚梅花 (**GONG, Meihua**); 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)。李计广 (**LI, Jiguang**); 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)。欧阳凯 (**OUYANG, Kai**); 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: 北京法胜知识产权代理有限公司 (**FASHENG INTELLECTUAL PROPERTY COMPANY, LTD.**); 中国北京市海淀区悦秀路 99 号 1 单元 4 层 401, Beijing 100085 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI,

GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))。

(54) **Title:** REPEATED SEQUENCING METHOD

(54) 发明名称: 重复测序的方法

(57) **Abstract:** Provided is a repeated sequencing method, comprising: on the basis of a first sequencing primer, performing first sequencing on a sample to obtain a first sequencing sequence based on a first sequencing synthetic strand; bringing an elution reagent into contact with the first sequencing synthetic strand so as to remove the first sequencing synthetic strand; on the basis of the first sequencing primer, performing second sequencing on the sample to obtain a second sequencing sequence based on a second sequencing synthetic strand; and comparing the first sequencing sequence with the second sequencing sequence to determine sequence information of the sample.

(57) 摘要: 提供了一种重复测序方法, 包括: 基于第一测序引物, 对样本进行第一测序, 以获得基于第一测序合成链的第一测序序列; 使洗脱试剂与所述第一测序合成链接触, 以去除所述第一测序合成链; 基于所述第一测序引物, 对所述样本进行第二测序, 以获得基于第二测序合成链的第二测序序列; 和将所述第一测序序列和所述第二测序序列进行比对, 以确定所述样本的序列信息。

WO 2025/007252 A1

重复测序的方法

技术领域

本申请涉及生物测序技术领域，具体涉及一种重复测序的方法。

背景技术

近年来，随着基因测序行业的发展，基因测序的技术突飞猛进，从一代的 Sanger 测序法、二代边合成边测序（sequencing by synthesis, SBS）法到三代单分子测序法，这些测序技术在测序过程中，由于测序试剂、测序样本和测序仪器和外部环境等多种原因，不可避免地会出现测序错误。特别是二代测序技术，其基于 PCR 原理，在边合成边测序的过程中由于内部、外部环境，难免出现测序错误，导致测序准确度降低。

为了提高测序准确度，各个测序平台提出了相关的方案，如 Illumina 近期公布了“Chemistry X”的新的测序准确度高的化学方法技术，包括新的染料、连接剂、阻断剂、聚合酶；Singular 公司发布了“Sequencing Engine”（测序引擎）技术；塞纳生物结合了 Fluorogenic 荧光发生测序化学技术以及 ECC（Error-Correction Code）纠错编码技术，以提高测序准确度。然而，这些平台的优化多侧重生化反应和信号采集，如提高荧光强度、提高聚合酶的保真性、提高洗脱的效率或者提高光学系统的精度，这使得测序成本增加，并且对测序平台的依赖性较高。

由此，亟待提供一种适用于多平台的简单易行且成本较低的高准确度测序方法。

发明内容

为此，本申请的实施例提供了一种重复测序方法。

本申请第一方面实施例提出了重复测序方法，包括：基于第一测序引物，对样本进行第一测序，以获得基于第一测序合成链的第一测序序列；使洗脱试剂与所述第一测序合成链接触，以去除所述第一测序合成链；基于所述第一测序引物，对所述样本进行第二测序，以获得基于第二测序合成链的第二测序序列；和将所述第一测序序列和所述第二测序序列进行比对，以确定所述样本的序列信息。

在一些实施例中，所述基于第一测序引物，对样本进行第一测序或第二测序包括：使所述第一测序引物与所述样本退火结合并进行多个碱基测序循环，其中每个循环完成一个碱基的信号检测，由此获得所述第一测序合成链或所述第二测序合成链的光学信号，并基于所述第一测序合成链或所述第二测序合成链的光学信号获得所述第一测序序列或所述第二测序序列。

在一些实施例中，基于第一测序引物，对样本进行第一测序还包括：将所述样本固定于用于测序的固相载体。在一些实施例中，所述固相载体为珠粒或芯片。

在一些实施例中，所述洗脱试剂为以下的一种或多种：DNA 变性剂和/或核酸外切酶，其中所述 DNA 变性剂包括：甲醇、乙醇、尿素、甲酰胺和氢氧化钠，所述核酸外切酶包括：蛇毒磷酸二酯酶、大肠杆菌核酸外切酶 I、大肠杆菌核酸外切酶 II、大肠杆菌核酸外切酶 III、脾脏磷酸二酯酶和嗜酸乳杆菌核酸酶。在一些实施例中，所述 DNA 变性剂为甲酰胺和/或氢氧化钠，所述核酸外切酶为大肠杆菌核酸外切酶 III。

在一些实施例中，所述甲酰胺的浓度为 30%-100%。在一些实施例中，所述甲酰胺的浓度为 40-100%。

在一些实施例中，所述大肠杆菌核酸外切酶 III 的浓度为 1-50 U/μL。在一些实施例中，所述大肠杆菌核

酸外切酶 III 的浓度为 1-10 U/ μ L。

在一些实施例中，所述洗脱试剂的洗脱时间为 0-20 min。在一些实施例中，所述洗脱试剂的洗脱时间为 5-15min。在一些实施例中，所述洗脱试剂的洗脱时间为 5-10min。

在一些实施例中，所述洗脱试剂的洗脱温度为 10-60°C。在一些实施例中，所述洗脱试剂的洗脱温度为 20-50°C。在一些实施例中，所述洗脱试剂的洗脱温度为 25-45°C。

在一些实施例中，将所述第一测序序列和所述第二测序序列进行比对，以确定所述样本的序列信息，包括：将所述第一测序序列和所述第二测序序列的对应位置的碱基进行比对；响应于所述第一测序序列和所述第二测序序列的对应位置的碱基相同，将所述相同碱基确定为所述位置处的碱基类型；响应于所述第一测序序列和所述第二测序序列的对应位置的碱基不同，通过统计学数据确定所述位置处的碱基类型。

在一些实施例中，所述统计学数据为 Q 值。

在一些实施例中，所述通过统计学数据确定所述位置处的碱基类型具体包括：比较所述第一测序序列和所述第二测序序列的对应位置的碱基的所述 Q 值，基于所述第一测序序列的所述位置处的碱基 x 的所述 Q 值大于所述第二测序序列的所述位置处的碱基 y 的所述 Q 值，将所述碱基 x 确定为所述位置处的碱基类型。

在一些实施例中，所述重复测序的方法还包括：重复基于所述第一测序引物的测序过程，包括：对所述样本再进行 m 次测序，以获得基于 m 个测序合成链的 m 个测序序列；使所述洗脱试剂分别与所述 m 个或 m-1 个测序合成链接触，以去除所述 m 个或 m-1 个测序合成链；和将所述第一测序序列、所述第二测序序列和所述 m 个测序序列进行比对，以确定所述样本的序列信息。

在一些实施例中，m 为大于 0 的正整数。在一些实施例中， $1 \leq m \leq 10$ 。

在一些实施例中，所述将所述第一测序序列、所述第二测序序列和所述 m 个测序序列进行比对，以确定所述样本的序列信息包括：将所述第一测序序列、所述第二测序序列和所述 m 个测序序列的对应位置的碱基进行比对；响应于所述第一测序序列、所述第二测序序列和所述 m 个测序序列的对应位置的碱基相同，将所述相同碱基确定为所述位置处的碱基类型；响应于所述第一测序序列、所述第二测序序列和所述 m 个测序序列的对应位置的碱基不同，将 m+2 个碱基中出现频率最高的碱基确定为所述位置处的碱基类型，或者响应于所述第一测序序列、所述第二测序序列和所述 m 个测序序列的对应位置的碱基不同，则通过统计学数据确定所述位置处的碱基类型。

在一些实施例中，所述重复测序的方法还包括：响应于所述第一测序序列、所述第二测序序列和所述 m 个测序序列的对应位置的碱基不同，且 m+2 个碱基中存在出现频率相同的碱基类型，则通过统计学数据确定所述位置处的碱基类型。在一些实施例中，所述统计学数据为 Q 值。

在一些实施例中，所述第一测序引物是正向测序引物或反向测序引物。

本申请另一方面实施例还提出了一种双端重复测序的方法，包括：基于第一测序引物，根据如本申请上述任一实施例中所述的重复测序方法对样本进行第一端重复测序，以获得所述样本的第一端序列信息；基于第二测序引物，对样本进行第二端测序，以获得所述样本的第二端序列信息；和分析所述第一端序列信息和所述第二端序列信息，以确定所述样本的序列信息，其中所述第一测序引物为正向测序引物或反向测序引物中的一个，所述第二测序引物为正向测序引物或反向测序引物中的另一个。

本申请另一方面实施例还提出了一种双端重复测序的方法，包括：基于第一测序引物，根据如本申请上述任一实施例所述的重复测序方法对样本进行第一端重复测序，以获得所述样本的第一端序列信息；基于第二测序引物，根据如本申请上述任一实施例所述的重复测序方法对样本进行第二端重复测序，以获得所述样本的第二端序列信息；和分析所述第一端序列信息和所述第二端序列信息，以确定所述样本的序列信息，其中所述第一测序引物为正向测序引物，所述第一端序列信息为正读序列信息，并且所述第二测序引物为反向测序引物，所述第二端序列信息为反读序列信息。

在一些实施例中，基于所述第二测序引物对所述样本进行第二端重复测序包括：生成所述样本的互补链，并基于所述互补链进行所述第二端重复测序。

在一些实施例中，在生成所述样本的互补链之后，所述方法还包括：将所述样本移除，并基于所述互补链进行第二端测序或所述第二端重复测序。

在一些实施例中，所述样本为 DNA 纳米球，所述互补链为 MDA 链。

在一些实施例中，所述样本中含有条形码 1 和可选地条形码 2，所述方法还包括：分别对所述条形码 1 和所述条形码 2 进行测序，以获得所述条形码 1 和所述条形码 2 的序列信息；和根据所述条形码 1 和所述条形码 2 的序列信息，筛选出所述样本的序列信息。

本申请另一方面实施例还提出了一种重复测序体系，包括：测序试剂和洗脱试剂，其中所述洗脱试剂为以下的一种或多种：DNA 变性剂和/或核酸外切酶，其中所述 DNA 变性剂包括：甲醇、乙醇、尿素、甲酰胺、氢氧化钠，所述核酸外切酶包括：蛇毒磷酸二酯酶、大肠杆菌核酸外切酶 I、大肠杆菌核酸外切酶 II、大肠杆菌核酸外切酶 III、脾脏磷酸二酯酶和嗜酸乳杆菌核酸酶。在一些实施例中，所述 DNA 变性剂为甲酰胺和/或氢氧化钠，所述核酸外切酶为大肠杆菌核酸外切酶 III。

在一些实施例中，所述测序试剂包括第一测序引物、反应酶和 dNTP。在一些实施例中，所述 dNTP 上连接荧光基团以用于碱基报告。在一些实施例中，所述测序试剂还包括第二测序引物。在一些实施例中，所述测序试剂还包括第一条形码测序引物和可选地第二条形码测序引物。

在一些实施例中，所述反应酶包括 DNA 聚合酶和可选地 DNA 连接酶。

本申请另一方面实施例还提出了一种测序试剂盒，包括如本申请上述任一实施例所述的重复测序体系。

本申请的技术方案实现了如下技术效果：

本申请实施例中的重复测序方法及其应用，从测序逻辑上对测序准确度进行了提升，通过对同一核苷酸序列模板进行多次测序，得到多条测序序列，然后利用生物信息分析对所得的多条测序序列进行互相校正，实现了对序列错误的有效检出，有效降低了测序过程中由于内部、外部环境等引起的测序错误对测序结果产生的影响，显著提高了测序准确度；同时，无需进行多次独立的重复测序过程，而仅需在单次实验中针对同一核苷酸序列模板进行多次测序操作，因此具有操作简单且成本较低的优点。此外，此方法适用于多个测序平台，能够广泛用于多平台的测序准确度的提升。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为根据本申请实施例的重复测序方法示意图；

图 2 为根据本申请实施例的另一重复测序方法示意图。

具体实施方式

下面结合具体实施方式对本发明进行进一步的详细描述，给出的实施例仅为了阐明本发明，并非限制本发明的范围。以下提供的实施例可作为本技术领域普通技术人员进行进一步改进的指南，并不以任何方式构成对本发明的限制。

本申请是基于发明人的以下认识做出的：

在相关技术中，二代测序平台均使用 SBS 方法，它们的测序过程相似，包括将反应液通过压力注入到芯片，发生聚合反应，采集光信号，然后进行洗脱，进行下一个碱基测序。这样的基于 PCR 原理的过程受外环境、仪器本身、测序试剂、样本等多种因素影响，比如容易出现碱基错配，因此无法避免会出现测序错误，且随着测序读长的累积，测序错误概率增加，导致测序片段末端的错误率偏高。更准确地读取碱基是测序发展的重要发展方向。目前相关平台的优化多侧重生化反应和信号采集上，如提高荧光强度、提高聚合酶的保真性、提高洗脱的效率或者提高光学系统的精度，这些改进使得测序成本增加，且对专用测序试剂、测序平台的依赖性增加。此外，相关技术中，往往针对同一样本进行多次独立的重复测序来增加数据量以增加测序深度，由此提高样本序列还原的准确性。然而，这样的多次独立的重复测序需要较大的样本投入量，并且测序试剂的消耗量大，测序成本高昂。

基于此，发明人经多次实验和测试，开发了一种重复测序的方法，该方法在测序逻辑上对测序方案进行了优化，通过在单次测序中对同一核苷酸序列模板进行重复测序，得到多条测序序列，然后对此多条测序序列进行互相校正，有效提高了测序准确度。

在本申请实施例中，通过对相同 DNA 模板进行第二次或者第三次测序，能够有效降低如上提到的测序错误的影响因素，比如基于多次测序，第一次测序错误的碱基在第二次或者第三次测序中能测序正确，或者第二次测序错误的碱基在第一次或者第三次测序中能测序正确，能够有效完成错误碱基的识别和校正，即基于对同一 DNA 模板进行两次或两次以上测序，同一位置处出现两次或两次以上相同结果的碱基序列判定为真实的正确的序列，由此有效提高了测序准确度。同时，相较于多次独立的测序过程，这样的基于单次测序的重复检测方法操作简单且成本较低。

本申请实施例中，“样本”、“待测序列”、“待测样本”、“待测的核苷酸序列模板”是指待测核酸样品。在一些实施例中，该样本可以为应用于各个测序平台的待测文库，例如单链的 DNB、双链 DNA 文库等。在一些实施例中，样本中包含能够与样本测序引物结合的引物结合序列。在一些实施中，样本中还包含能够与条形码测序引物结合的条形码引物结合序列，以基于条形码（barcode 或 index）的测序结果确定正向/反向序列，或实现不同样本间的区分。在一些实施例中，样本中还包含能与固相载体结合的序列，以将该样本固定于固相载体上。

本申请实施例中，“固相载体”是指能够将待测核酸样本固定于其上以便后续测序的固相介质。在一些实施例中，固相载体上固定有能够识别并用于待测核酸样本结合的序列，以将待测核酸样本固定于其上。在另一些实施例中，固相载体上具有预制大小的凹槽，以通过与待测核酸样本的大小匹配以将其固定在其上。在另一些实施例中，固相载体上附接有化学基团，以通过化学键、范德华力等与待测核酸样本连接以将其固定在其上。在一些实施例中，固相载体可以是珠粒（bead）、芯片等。

在一些实施例中，固相载体可以是芯片。在一些实施例中，该芯片用于固定常规双链 DNA 文库或单链 DNA 文库。在一些实施例中，该芯片还用于基于该双链 DNA 文库或单链 DNA 文库所制备的 DNA 纳米球 (DNB)。在待测核酸样本为 DNB 的情况下，可以借助于例如探针技术将 DNB 固定在该芯片上，并基于 cPAL(组合探针锚定连接法)或 CPAS(联合探针锚定聚合技术)对测序芯片进行测序，从而获得 DNA 纳米球的核酸信息。具体地，例如，可以利用四种不同颜色标记的探针读取碱基，然后 DNA 连接酶将四种不同颜色标记的探针结合到模板的相应碱基上，通过对荧光基团的成像来判断碱基类型。在一些实施例中，DNA 纳米球通过网状小孔或者六甲基二硅氮烷固定在芯片上。

本申请实施例中，“重复测序”是指在单个独立的测序过程中，针对同一待测样本，通过洗脱测序过程中生成的合成链并重新启动边合成边测序过程，以进行多次重复的测序过程，例如进行第一测序、第二测序或更多次重复的测序。在一些实施例中，还可以对同一样本在两次测序的基础上再进行 m 次测序，以获得基于 m 个测序合成链的 m 个测序序列；使用洗脱试剂分别将该 m 个或 m-1 个测序合成链去除；并将第一测序序列、第二测序序列和该 m 个测序序列进行比对，以确定样本的序列信息，其中 m 可以为大于 0 的任一正整数，例如可以为 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15 或更多次重复测序。可以理解的是，两次或三次重复测序有效提高了结果的准确性，而随着重复测序次数的增加，测序结果的准确率也会在一定程度上有所提高，重复测序的次数可以根据用户的实际需求确定。此外，可以理解的是，在多次（两次或两次以上）重复测序时，最后一次生成的测序合成链可以选择性的去除。

本申请实施例中，“测序引物”可以包括样本测序引物，如用于正向测序 (Forward strand) 的正向样本测序引物和用于反向测序 (Reverse strand) 的反向样本测序引物，即第一测序引物和/或第二测序引物。在一些实施例中，“测序引物”还可以包括条形码测序引物，如用于正向测序 (Forward strand) 的正向条形码测序引物和用于反向测序 (Reverse strand) 的反向条形码测序引物，即条形码 1 测序引物和所述条形码 2 测序引物。

本申请实施例中，“基于测序合成链获得测序序列”是指基于边合成边测序的原理：通过向反应体系中同时添加 DNA 聚合酶、引物和带有碱基特异荧光标记的 4 种 dNTP，这些 dNTP 的 3' -OH 被化学方法所保护，因而每次只能添加一个 dNTP，这就确保了在测序过程中，一次只会被添加一个碱基。同时在 dNTP 被添加到合成链上后，所有未使用的游离 dNTP 和 DNA 聚合酶会被洗脱掉。通过光学设备完成荧光信号的记录，最后利用计算机分析将光学信号转化为测序碱基。这样荧光信号记录完成后，再加入化学试剂淬灭荧光信号并去除 dNTP 3' -OH 保护基团，以便能进行下一轮的测序反应（也就是下一个碱基的读取）。因此，可以理解的是，“基于测序合成链获得测序序列”是指在对样本进行测序时生成了测序合成链，基于合成该合成链时所释放的特异性光学信号（荧光信号），通过光学信号读取与转换，即可完成碱基的读取，从而获得测序序列。

本申请实施例中，“合成链的洗脱”是指使用洗脱试剂去除测序过程中产生的合成链。在一些实施例中，洗脱试剂可以为使 DNA 双链的碱基对间的氢键断裂，从而使双链变成单链的试剂。在一些实施例中，洗脱试剂可以为 DNA 变性剂和/或核酸外切酶，其中 DNA 变性剂可以为有机或无机试剂，其能够破坏双螺旋结构；核酸外切酶通过将 DNA 分子链的末端顺次水解磷酸二酯键，而使该 DNA 分子链水解为单核苷酸。在一些实施例中，DNA 变性剂可以为：甲醇、乙醇、尿素、甲酰胺、氢氧化钠等。在一些实施例中，核酸外切酶可以为：蛇毒磷酸二酯酶、大肠杆菌核酸外切酶 I、大肠杆菌核酸外切酶 II、大肠

杆菌核酸外切酶 III、脾脏磷酸二酯酶和嗜酸乳杆菌核酸酶（*Lac-tobacillus acidophilus*）等。在一些实施例中，核酸外切酶可以包括蛇毒磷酸二酯酶、大肠杆菌核酸外切酶 II 和/或大肠杆菌核酸外切酶 III。

本申请实施例中，DNA 变性剂例如甲酰胺或氢氧化钠的使用浓度可以为 30%-100%，优选 40-100%。在一些实施例中，DNA 变性剂的使用浓度为 50%或 100%。在一些实施例中，核酸外切酶的使用浓度可以为 1-50 U/ μ L，优选为 1-10 U/ μ L。在一些实施例中，外切酶的使用浓度可以为 4-8 U/ μ L 及其范围内的任一浓度值，例如 4 U/ μ L、5 U/ μ L、6 U/ μ L、7 U/ μ L、8 U/ μ L 等。在一些实施例中，洗脱试剂的洗脱时间可以为 0-20 min，优选为 5-15min，更优选为 5-10min，例如可以为 5 min、6 min、7 min、8 min、9 min、10 min。在一些实施例中，洗脱试剂的洗脱温度可以为 10-60°C，优选为 20-50°C，更优选为 25-45°C 间的任一数值，例如可以为 25、26、27、28、29、30、31、32、33、34、35、36、37、37.5、38、38.5、39、39.5、40、40.5、41、42、43、44、45 或其间的任一数值。

本申请实施例中，以文库量为 10 fmol 每个通道（一张载片 4 个通道，相应的测序体系量为单通道体积 40 μ L）的测序为例，可根据反应的温度和时间对甲酰胺的使用浓度进行调整（甲酰胺的浓度母液为 100%）。在一些实施例中，在洗脱时间为 2 min-20 min、洗脱温度为 25°C-40°C 的情况下，甲酰胺的浓度可以为 25%-100%之间的任一数值。在一些实施例中，同样以文库量为 10 fmol 每个通道（一张载片 4 个通道，相应的测序体系量为单通道体积 40 μ L）的测序为例，基于大肠杆菌核酸外切酶 III 的母液浓度为 100U/ μ L，反应条件可以为大肠杆菌终浓度为 1U/ μ L-2U/ μ L、反应温度为 25-37°C、反应时间为 2-10min。

本申请实施例中，重复测序获得的序列之间的比对（如第一测序序列和第二测序序列间的比对）是指将针对同一样本经多次重复测序测出的多条序列进行比较，该比对可以基于同一参考序列的坐标信息确定重复测序获得的各个序列的各个碱基的位置，并对同一位置处的各个碱基进行统计，以分析该位置处的碱基的真实类型。在另一些实施例中，该比对可以在重复测序获得的序列之间直接进行，而无需参考序列提供坐标信息。

本申请实施例中，基于重复测序获得的测序序列的比对来确定样本的序列信息。具体而言，在重复测序次数为 2 的情况下，通过将重复测序获得的两个序列的同一位置（也为对应位置）处的碱基进行统计，响应于第一测序序列和第二测序序列的对应位置的碱基相同，则将该相同碱基确定为该位置处的碱基类型；响应于第一测序序列和第二测序序列的对应位置的碱基不同，则通过统计学数据确定该位置处的碱基类型。在一些实施例中，该统计学数据为 Q 值（Q-score/Q phred，即碱基质量值，其体现测得的碱基的质量值分数，一般情况下 Q 值越大，表示识别错误的可能性越小，可信度就越高）。在一些实施例中，比较第一测序序列和第二测序序列的对应位置的碱基的 Q 值，根据 Q 值权重法，基于第一测序序列的该位置处的碱基 x 的 Q 值大于第二测序序列的位置处的碱基 y 的 Q 值，则将碱基 x 确定为该位置处的碱基类型。

在一些实施例中，在重复测序次数为 2+m 的情况下（其中 m 为大于 0 的正整数），上述重复测序序列之间的比对包括：将第一测序序列、第二测序序列和 m 个测序序列进行比对，以确定样本的序列信息。在一些实施例中，该过程具体包括：将第一测序序列、第二测序序列和 m 个测序序列的对应位置的碱基进行比对；响应于第一测序序列、第二测序序列和 m 个测序序列的对应位置的碱基相同，则将该相同碱基确定为该位置处的碱基类型；响应于第一测序序列、第二测序序列和 m 个测序序列的对应位置的碱基

不同，统计 $m+2$ 个碱基的频率，将 $m+2$ 个碱基中出现频率最高的碱基确定为该位置处的碱基类型。进一步地，响应于第一测序序列、第二测序序列和 m 个测序序列的对应位置的碱基不同，且 $m+2$ 个碱基中存在出现频率相同的碱基类型，则通过统计学数据确定该位置处的碱基类型。在一些实施例中，统计学数据为 Q 值，即，将 Q 值较大的碱基类型确定为该位置处的碱基类型。可以理解的是，通过对相同位置多次测序进行错误检测和校正，能够有效减少由测序内外环境因素导致的测序错误。

本申请实施例中，上述重复测序的方法可以应用于单端测序（Single-end sequencing）和/或双端测序（Pair-end sequencing）。在一些实施例中，基于测序引物是正向测序引物或反向测序引物，上述重复测序的方法可以应用于单端测序，此单端测序可以为正向测序或反向测序。可以理解的是，本申请实施例提出的重复测序方法在应用于单端测序的情况下，不会影响另一端的测序，也就是说，在进行单端的重复测序后，可以正常进行另一端的测序。在一些实施例中，另一端的测序可以为单次测序，也可以根据本申请上述任一实施例所述的重复测序方法，基于另一端测序引物，对另一端进行重复测序。

由此，本申请实施例提出了一种双端测序方法。在一些实施例中，在另一端为正常测序的情况下（即单次测序），该双端测序方法包括：基于第二测序引物，对样本进行第二端测序，以获得该样本的第二端序列信息；和分析通过单端重复测序获得的第一端序列信息和该第二端序列信息，以确定样本的序列信息，其中第一测序引物为正向测序引物或反向测序引物中的一个，第二测序引物为正向测序引物或反向测序引物中的另一个。可以理解的是，通过单端的重复测序和另一端的正常测序，能够有效提高测序数据的准确性。

本申请实施例还提出了另一种双端测序方法。在另一些实施例中，在另一端为重复测序的情况下，该双端测序方法包括：基于第一测序引物，根据本申请上述任一实施例所述的重复测序方法对样本进行第一端重复测序，以获得该样本的第一端序列信息；基于第二测序引物，同样根据本申请上述任一实施例中所述的重复测序方法对样本进行第二端重复测序，以获得该样本的第二端序列信息；和分析第一端序列信息和第二端序列信息，以确定所述样本的序列信息，其中第一测序引物为正向测序引物，第一端序列信息为正读序列信息，并且第二测序引物为反向测序引物，第二端序列信息为反读序列信息。可以理解的是，本申请实施例提出的双端测序方法包括了两个单端的重复测序，并基于重复测序的多个序列确定碱基序列，由此有效提高了测序结果的准确性。

可以理解的是，本申请实施例中，另一端测序是指相对于先进行的一个单端而言的。换言之，在先进行的单端测序为正向测序的情况下，该另一端测序是指反向测序；在先进行的单端测序为反向测序的情况下，该另一端测序是指正向测序。

在一些实施例中，对样本进行反向测序还可以包括：生成该样本的互补链（即第二链），并基于该互补链进行反向测序。在一些实施例中，在生成样本的互补链后，还包括将样本移除，以基于该互补链进行反向测序。

可以理解的是，当样本为 DNB 时，该互补链（第二链）的生成基于 DNA 聚合酶的多重置换扩增（multiple displacement amplification, MDA），由此该互补链也可称为 MDA 链。

在本申请实施例中，在进行单端测序时，该方法还包括：对样本中的条形码序列进行测序，其中该条形码序列具有样本特异性，以区分不同样本。在一些实施例中，还可以对条形码序列进行重复测序。

在本申请实施例中，在进行双端测序时，该方法还包括：分别对样本中包含的条形码 1 和条形码 2

进行测序，以获得条形码 1 和所述条形码 2 的序列信息；和根据条形码 1 和根据条形码 2 的序列信息，筛选出样本的序列信息。在一些实施例中，条形码 1 和条形码 2 具有方向特异性。可以理解的是，通过结合条形码 1 和条形码 2 的序列信息，能够区分样本的序列信息的方向，即确定某序列为正向读长序列或反向读长序列。因此在一些实施例中，样本的序列信息可以包括第一端序列信息（正读序列信息）和

5 第二端序列信息（反读序列信息）。

本申请实施例中，测序数据的后期统计和分析可以以分析软件、脚本、命令行等实现，只要保证能够完成本申请实施例中的序列比对、Q 值统计、碱基确定等步骤即可。

可以理解的是，本申请实施例中提出的重复测序方法适用于多个测序平台，其既可以应用于以滚环扩增（Rolling Circle Replication）为原理的 DNB 模板的测序，也可应用于以桥式扩增（Bridge PCR）为原理的线性 DNA 模板的测序，基于模板和/或模板的反向互补链，并对测序过程中生成的测序合成链进行洗脱，有效提升了多个平台的序列产出的准确度。

10

需要说明的是，本申请实施例中提出的双端测序方法，在另一端测序也为重复测序的情况下，该另一端的重复测序的操作和数据处理也与本申请上方任一实施例中提出的单端的重复测序方法相同，此处不再赘述。

本申请实施例中的重复测序方法，从测序逻辑上对测序准确度进行了提升，通过对同一核苷酸序列模板进行多次测序，得到多条测序序列，然后利用生物信息分析对所得的多条测序序列进行互相校正，实现了对序列错误的有效检出，有效降低了测序过程中由于内部、外部环境等引起的测序错误对测序结果产生的影响，显著提高了测序准确度；同时，无需进行多次独立的重复测序过程，而仅需在单次实验中针对同一核苷酸序列模板进行多次测序操作，因此具有操作简单且成本较低的优点。此外，此方法适用于多个测序平台，能够广泛用于多平台的测序准确度的提升。

15 20

本申请实施例还提出了一种重复测序体系，包括：测序试剂和洗脱试剂，其中洗脱试剂为以下的一种或多种：DNA 变性剂和/或核酸外切酶，其中 DNA 变性剂包括：甲醇、乙醇、尿素、甲酰胺、氢氧化钠；核酸外切酶包括：蛇毒磷酸二酯酶、大肠杆菌核酸外切酶 I、大肠杆菌核酸外切酶 II、大肠杆菌核酸外切酶 III、脾脏磷酸二酯酶和嗜酸乳杆菌核酸酶。在一些实施例中，DNA 变性剂为甲酰胺和/或氢氧化钠，核酸外切酶为大肠杆菌核酸外切酶 III。

25

在一些实施例中，测序试剂包括测序引物，例如正向测序引物和/或反向测序引物（对应于第一测序引物和/或第二测序引物）、反应酶和 dNTP。在一些实施例中，该 dNTP 上附接荧光基团以用于碱基报告，以基于不同的荧光颜色读取碱基类型，从而获得样本的测序序列。在一些实施例中，测序试剂中还包括条形码测序引物，以用于进行单端或双端测序中的条形码信息的读取。

在一些实施例中，测序试剂中的反应酶用于合成链的扩增等，该反应酶可以包括 DNA 聚合酶和可选地 DNA 连接酶。

30

本申请实施例还提出了一种测序试剂盒，包括如上任一实施例提出的重复测序体系。

需要说明的是，上方对重复测序方法的实施例的解释说明也适用于上述实施例中的重复测序体系和测序试剂盒，在此不再赘述。

35

下述实施例中的实验方法，如无特殊说明，均为常规方法，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。下述实施例中所用的材料、试剂等，如无特殊说明，均可从商业途径得到。

如无特殊说明，以下实施例中的定量试验，均设置三次重复实验，结果取平均值。

实施例 1

图 1 为根据本申请实施例 1 的重复测序方法示意图。如图 1 所示，该方法可以包括 i. 生化实验方案
5 和 ii. 生物信息分析方案，其中生化实验方案包括：

a. 将待测的核苷酸序列模板固定在芯片上，通入测序引物进行退火，然后进行第一次测序，检测并记录第一次测序中的多轮 SBS 反应的荧光信息，通过荧光信息获得待测核苷酸模板第一次的碱基序列信息。

10 b. 通入能使 DNA 双链变性的有机溶剂或者具有 3'端外切功能的酶，对如上(a)中描述的第一次测序链进行去除，再次通入测序引物进行退火，进行第二次测序，检测并记录第二次测序中的多轮 SBS 反应的荧光信息，通过荧光信息获得待测核苷酸模板第二次的碱基序列信息。

c. 通入能使 DNA 双链变性的有机溶剂或者具有 3'端外切功能的酶，对如上(b)中描述的第二次测序链进行去除，再次通入测序引物进行退火，进行第三次测序，检测并记录第三次测序中的多轮 SBS 反应的荧光信息，通过荧光信息获得待测核苷酸模板第三次的碱基序列信息。

15 d. 重复进行更多次：去除测序链→测序引物退火→测序，从而获得相同核苷酸序列模板的多次测序碱基序列信息。

生物信息分析方案包括以下步骤：对每个核苷酸序列模板的三次测序所得到的序列进行生物信息分析：首先，同一比对位点三次测序读取到的都是相同碱基，将此三次相同碱基识别为该位点正确的碱基；同一比对位点两次测序碱基相同、一次碱基不同，将此两次相同的碱基识别为该位点正确的碱基；同一
20 比对位点三次测序碱基都不同，则进行进一步校正，具体为：采取 Q 值权重法将 Q 值最大者确定为该位点正确的碱基。例如三次测序中，测得碱基有两次相同一次不同，则采取少数服从多数法；如三次测序中，测得的碱基都不同，则采取 Q 值权重法取 Q 值最大者。最后将读取正确的碱基和校正后碱基再拼接成完整序列，以此来提高测序序列的准确性。

本实施例中的重复测序方法，通过对同一模板进行两次或两次以上测序（测序-去除测序链-再测序），
25 并对两次或两次以上测序结果进行相互比对校正分析，有效提高了序列准确度。

实施例 2

图 2 为根据本申请实施例 2 的另一重复测序方法示意图。如图 2 所示，该方法可以应用于相关的基于桥式扩增的测序技术，包括 i. 生化实验方案和 ii. 生物信息分析方案，其中生化实验方案包括：

1. 将待测样本的 DNA 分子固定在布满接头的流动槽（Flowcell）上。
- 30 2. 以流动槽上的接头为引物进行多轮桥式扩增，形成 Cluster，即待测核苷酸序列模板。
3. 通入测序引物进行退火，接着进行第一次测序的多轮 SBS 反应（每一轮 SBS 反应测得一个碱基），检测并记录第一次测序中的多轮 SBS 反应的荧光信息，通过荧光信息获得第一次测序碱基序列信息。
4. 使用 NaOH 碱溶液将第一次测序的测序生成链去除。
5. 再次通入测序引物进行退火，接着进行第二次测序的多轮 SBS 反应（每一轮 SBS 反应测得一个
35 碱基），检测并记录第二次测序中的多轮 SBS 反应的荧光信息，通过荧光信息获得第二次测序碱基序列

信息。

6. 重复步骤 4 和 5, 检测并记录第三次测序中的多轮 SBS 反应的荧光信息 (每一轮 SBS 反应测得一个碱基), 通过荧光信息获得第三次测序的碱基序列信息。

生物信息分析方案包括以下步骤: 对每个核苷酸序列模板的三次测序所得到的序列进行生物信息分析: 首先, 同一比对位点三次测序读取到的都是相同碱基, 将此三次相同碱基识别为该位点正确的碱基; 同一比对位点两次测序碱基相同、一次碱基不同, 将此两次相同的碱基识别为该位点正确的碱基; 同一比对位点三次测序碱基都不同, 则进行进一步校正, 具体为: 采取 Q 值权重法将 Q 值最大者确定为该位点正确的碱基例如三次测序中, 测得碱基有两次相同一次不同, 则采取少数服从多数法; 如三次测序中, 测得的碱基都不同, 则采取 Q 值权重法取 Q 值最大者。最后将读取正确的碱基和校正后碱基再拼接成完整序列, 以此来提高测序序列的准确性。

本实施例中的基于桥式扩增的重复测序方法还可选地包括对二链进行反向测序, 以获得双端测序结果, 该方法可以包括: 合成每个核苷酸序列模板的互补链 (即第二链); 将每个核苷酸序列模板移除; 基于该互补链进行反向测序。其生物信息分析方案与上述单端测序的分析方案相同。

本实施例中的基于桥式扩增的重复测序方法, 通过对同一模板进行两次或两次以上测序 (测序-去除测序链-再测序), 并对两次或两次以上测序结果进行相互比对校正分析, 有效提高了序列准确度。

实施例 3

采用 E.coli PCR free 样本 (MGI), 利用 MGISEQ-2000RS 高通量测序试剂盒 (MGI, 货号 1000012552) 在 MGISEQ-2000 平台上, 按照试剂盒和测序平台的使用说明, 以该样本的核苷酸序列为模板进行三次 SE100 测序 (即单端 100bp 的重复测序):

1) 首先将该 E.coli PCR free 样本的核苷酸序列模板固定在芯片上, 通入上方测序试剂盒中的 CPAS AD153 测序引物工作液进行退火, 然后进行第一次 SE100 测序, 检测并记录第一次测序中的多轮 SBS 反应的荧光信息 (其中每轮 SBS 反应测得一个碱基), 通过荧光信息获得核苷酸模板第一次的碱基序列信息。

2) 通入 40 μ L 浓度为 100% 的甲酰胺在 40 $^{\circ}$ C 温度条件反应 8 min (为单通道 (lane) 用量), 以对如上 1) 中描述的第一次测序链 (即测序合成链) 进行去除, 再次通入 CPAS AD153 测序引物工作液进行退火, 进行第二次 SE100 测序, 检测并记录第二次测序中的多轮 SBS 反应的荧光信息 (其中每轮 SBS 反应测得一个碱基), 通过荧光信息获得核苷酸模板的第二次的碱基序列信息。

3) 再次通入 40 μ L 浓度为 100% 的甲酰胺在 40 $^{\circ}$ C 温度条件反应 8min, 对如上 2) 中描述的第二次测序链进行去除, 再次通入 CPAS AD153 测序引物工作液进行退火, 进行第三次 SE100 测序, 检测并记录第三次测序中的多轮 SBS 反应的荧光信息 (其中每轮 SBS 反应测得一个碱基), 通过荧光信息获得核苷酸模板第三次的碱基序列信息。

4) 对三次测得的碱基序列信息进行分析, 以获得高准确度的碱基序列。

表 1 展示了步骤 1) 所得的第一次 SE100 测序的结果 (即 1st); 表 2 展示了步骤 2) 和 3) 所得的第二次 (2nd) 和第三次 SE100 测序 (3rd) 的结果。由表 1 和 2 可见, 第一次测序的质量指标 Q30 为 91.03%, 经过第一遍甲酰胺反应去除测序链, 第二次测序的质量指标 Q30 为 83.46%, 再次经过第二遍甲酰胺反应去除测序链, 第三次测序的质量指标 Q30 为 81.07%。可见使用甲酰胺去除测序链, 会对 DNB 模板产

生一定的损伤，导致第二第三次的测序质量有所下降，但数据整体上仍保持较高的质量，即三轮测序均保持着较高的数据质量产出。

表1：第一次SE100测序结果

items	1 st
CycleNumber	100
Q30(%)	91.03

表2：第二次和第三次SE100测序结果

items	2 nd	3 rd
CycleNumber	100	100
Q30(%)	83.46	81.07

5 使用 BWA MEM 比对算法 (<https://doi.org/10.48550/arXiv.1303.3997>)，对以上三次 SE100 测序所得到的序列进行互相校正：首先，同一比对位点三次测序读取到的都是相同碱基，将此三次相同碱基识别为该位点正确的碱基；同一比对位点两次测序碱基相同、一次碱基不同，将此两次相同的碱基识别为该位点正确的碱基；同一比对位点三次测序碱基都不同，则进行进一步校正：采取 Q 值权重法将 Q 值最大者确定为该位点正确的碱基。

10 表 3 为本实施例的第一次 SE100 测序的生物信息分析结果；表 4 为本实施例的三次 SE100 测序进行互相校正的结果。

表3：第一次SE100测序的生物信息分析结果-单次测序

items	1 st
ReadNum	478076152
BaseNum	47807615200
N%	0
GC%	50.92
Q20%	96.93
Q30%	91.03
Mapping rate(%)	99.94
Mismatch rate(%)	0.48
Mismatch rate !N(%)	0.48

表4：三次SE100测序的生物信息分析结果-三次测序互相校正

items	Triple
ReadNum	278825671
BaseNum	27882567100
N%	0

GC%	49.94
Q20%	99.74
Q30%	97.37
Mapping rate(%)	99.98
Mismatch rate(%)	0.12
Mismatch rate !N(%)	0.12

由表 3 和表 4 分别所示的第一次 SE100 的单次测序分析结果和三次 SE100 测序互相校正后的结果的比较可见, 三次测序后, 数据量指标 BaseNum 呈现出下降趋势, 第一次测序时的 BaseNum 约为 47.8Gb, 经过甲酰胺去除测序链, 第二次和第三次测序数据量有所减少, 最后在三次测序互相校正中用到的 BaseNum 约为 27.9Gb, 可见即使经由三次重复测序, 本申请实施例的重复测序方法所产生的总数据量仍

5 体积较小, 易于后期存储和处理。

同时, 三次测序后, 错误率指标 Mismatch rate !N(%)呈现出下降趋势, 第一次测序时的 Mismatch rate !N(%)为 0.48%。三次测序进行互相校正后的 Mismatch rate !N(%)为 0.12%, 与第一次测序的错误率相比下降了 75%; 此外, 相较于单次测序, 三次测序整合分析后的数据质量指标 Q20 和 Q30 值明显提升, 且比对率 Mapping rate(%)也有所增加, 说明本申请实施例的重复测序方法能够保证测序结果的质量,

10 并有效提升测序结果的准确性。

实施例 4

采用 E.coli PCR free 样本(MGI), 利用 MGISEQ-2000RS 高通量测序试剂盒(MGI, 货号 1000012551) 在 MGISEQ-2000 平台上, 按照试剂盒和测序平台的使用说明, 以该样本的核苷酸序列为模板进行三次 SE50 测序 (即单端 50bp 的重复测序):

1) 首先将该 E.coli PCR free 样本的核苷酸序列模板固定在芯片上, 通入 CPAS AD153 测序引物工作液进行退火, 然后进行第一次 SE50 测序, 检测并记录第一次测序中的多轮 SBS 反应的荧光信息 (其中每轮 SBS 反应测得一个碱基), 通过荧光信息获得核苷酸模板第一次的碱基序列信息。

2) 通入 40μL 浓度为 100%的甲酰胺在 40℃温度条件反应 8 min, 以对如上 1)中描述的第一次测序链 (即测序合成链) 进行去除, 再次通入 CPAS AD153 测序引物工作液进行退火, 进行第二次 SE50 测序, 检测并记录第二次测序中的多轮 SBS 反应的荧光信息 (其中每轮 SBS 反应测得一个碱基), 通过荧光信息获得核苷酸模板的第二次的碱基序列信息。

3) 再次通入 40μL 浓度为 100%的甲酰胺在 40℃温度条件反应 8min, 对如上 2)中描述的第二次测序链进行去除, 再次通入 CPAS AD153 测序引物工作液进行退火, 进行第三次 SE50 测序, 检测并记录第三次测序中的多轮 SBS 反应的荧光信息 (其中每轮 SBS 反应测得一个碱基), 通过荧光信息获得核苷酸模板第三次的碱基序列信息。

4) 对三次测得的碱基序列信息进行分析, 以获得高准确度的碱基序列。

表 5 展示了步骤 1)所得的第一次 SE50 测序的结果 (即 1st); 表 6 展示了步骤 2)和 3)所得的第二次 (2nd) 和第三次 SE50 测序 (3rd) 的结果。由表 5 和 6 可见, 第一次测序的质量指标 Q30 为 95.34%, 经过第一遍甲酰胺反应去除测序链, 第二次测序的质量指标 Q30 为 93.82%, 再次经过第二遍甲酰胺反应去除测序链, 第三次测序的质量指标 Q30 为 93.71%, 可见使用甲酰胺去除测序链, 会对 DNB 模板产

生一定的损伤，导致第二第三次的测序质量有微小下降，但数据整体上仍保持较高的质量。此外，与实施例 3 中的 SE100 的测序结果相比，本实施例中 SE50 的测序读长较短，其整体的 Q30 有所提升，且第二和第三次的测序质量下降幅度并不明显，即三轮测序均保持着极高的数据质量产出。

表5：第一次SE50测序结果

items	1 st
CycleNumber	50
Q30(%)	95.34

5

表6：第二次和第三次SE50测序结果

items	2 nd	3 rd
CycleNumber	50	50
Q30(%)	93.82	93.71

使用 BWA MEM 比对算法，对以上三次 SE50 测序所得到的序列进行互相校正：首先，同一比对位点三次测序读取到的都是相同碱基，将此三次相同碱基识别为该位点正确的碱基；同一比对位点两次测序碱基相同、一次碱基不同，将此两次相同的碱基识别为该位点正确的碱基；同一比对位点三次测序碱基都不同，则进行进一步校正：采取 Q 值权重法将 Q 值最大者确定为该位点正确的碱基。

10

表 7 为本实施例的第一次 SE50 测序的生物信息分析结果；表 8 为本实施例的三次 SE50 测序进行互相校正的结果。

表7：第一次SE50测序的生物信息分析结果-单次测序

items	1 st
ReadNum	558209259
BaseNum	27910462950
N%	0.08
GC%	50.92
Q20%	98.51
Q30%	95.34
Mapping rate(%)	99.12
ErrorRate!N(%)	0.2

表8：三次SE50测序的生物信息分析结果-三次测序互相校正

items	Triple
ReadNum	522475407
BaseNum	26123770350
N%	0
GC%	50.81
Q20%	99.86

Q30%	99.17
Mapping rate(%)	99.94
ErrorRate!N(%)	0.06

由表 7 和表 8 分别所示的第一次 SE50 的单次测序分析结果和三次 SE50 测序互相校正后的结果的比较可见，三次测序后，数据量指标 BaseNum 呈现出下降趋势，第一次测序时的 BaseNum 约为 27.9Gb，经过甲酰胺去除测序链，第二次和第三次测序数据量稍有减少。最后在三次测序互相校正中用到的 BaseNum 约为 26.1Gb，可见即使经由三次重复测序，本申请实施例的重复测序方法所产生的总数据量仍

同时，三次测序后，错误率指标 Mismatch rate !N(%)呈现出下降趋势，第一次测序时的 Mismatch rate !N(%)为 0.2%。三次测序进行互相校正后的 Mismatch rate !N(%)为 0.06%，与第一次测序的错误率相比下降了 60%；此外，相较于单次测序，三次测序整合分析后的数据质量指标 Q20 和 Q30 值明显提升，且比对率 Mapping rate(%)也有所增加，说明本申请实施例的重复测序方法能够保证测序结果的质量，并

实施例 5

采用 E.coli PCR free (MGI) 和 E.coli PCR 样本 (MGI, 货号: 1000005033)，利用 BGISEQ-500RS 高通量测序试剂盒 (MGI, 货号 1000005485) 在 BGISEQ-500 平台上，按照试剂盒和测序平台的使用说明，以两样本的核苷酸序列为模板分别进行三次 SE50 测序 (即单端 50bp 的重复测序)：

1) 首先将该 E.coli PCR free 和 E.coli PCR 这两个样本的核苷酸序列模板固定在芯片上，通入 CPAS AD153 测序引物工作液进行退火，然后进行第一次 SE50 测序，检测并记录第一次测序中的多轮 SBS 反应的荧光信息 (其中每轮 SBS 反应测得一个碱基)，通过荧光信息获得两核苷酸模板第一次的碱基序列信息。

2) 通入 60 μ L 浓度为 4 U/ μ L 的 Exonuclease III 在 37 $^{\circ}$ C 温度条件反应 10min，以对如上 1) 中描述的第一次测序链 (即测序合成链) 进行去除，再次通入 CPAS AD153 测序引物工作液进行退火，进行第二次 SE50 测序，检测并记录第二次测序中的多轮 SBS 反应的荧光信息 (其中每轮 SBS 反应测得一个碱基)，通过荧光信息获得两核苷酸模板的第二次的碱基序列信息。

3) 再次通入 60 μ L 浓度为 4 U/ μ L 的 Exonuclease III 在 37 $^{\circ}$ C 温度条件反应 10min，对如上 2) 中描述的第二次测序链进行去除，再次通入 CPAS AD153 测序引物工作液进行退火，进行第三次 SE50 测序，检测并记录第三次测序中的多轮 SBS 反应的荧光信息 (其中每轮 SBS 反应测得一个碱基)，通过荧光信息获得两核苷酸模板第三次的碱基序列信息。

4) 对两样本的三次测得的碱基序列信息进行分析，以获得高准确度的碱基序列。

表 9 展示了步骤 1) 所得的两样本的第一次 SE50 测序的结果 (即 1st)；表 10 展示了步骤 2) 和 3) 所得的两样本的第二次 (2nd) 和第三次 SE50 测序 (3rd) 的结果。由表 9 可见，对于第一次测序的质量指标 Q30，PCR free 样本为 92.75%；PCR 样本为 92.71%。由表 10 可见，经过第一遍 Exonuclease III 反应去除测序链，对于第二次测序的质量指标 Q30，PCR free 样本为 86.33%；PCR 样本为 85.42%，再次经过第二遍 Exonuclease III 反应去除测序链，对于第三次测序的质量指标 Q30，PCR free 样本为 82.82%；PCR 样本为 82.35%，可见使用 Exonuclease III 去除测序链，会对 DNB 模板产生一定的损伤，导致第二第三

次的测序质量有所下降，但数据整体上仍保持较高的质量，即三轮测序均保持着较高的数据质量产出。

表9：第一次SE50测序结果

items	1 st	
	PCR free	PCR
CycleNumber	50	50
Q30	92.75	92.71

表10：第二次和第三次SE50测序结果

items	2 nd		3 rd	
	PCR free	PCR	PCR free	PCR
CycleNumber	50	50	50	50
Q30	86.33	85.42	82.82	82.35

使用 BWA MEM 比对算法，对以上两样本的三次 SE50 测序所得到的序列分别进行互相校正：首先，
5 同一比对位点三次测序读取到的都是相同碱基，将此三次相同碱基识别为该位点正确的碱基；同一比对
位点两次测序碱基相同、一次碱基不同，将此两次相同的碱基识别为该位点正确的碱基；同一比对位点
三次测序碱基都不同，则进行进一步校正：采取 Q 值权重法将 Q 值最大者确定为该位点正确的碱基。

表 11 为本实施例的两样本的第一次 SE50 测序的生物信息分析结果；表 12 为本实施例的两样本的
三次 SE50 测序进行互相校正的结果。

10

表11：第一次SE50测序的生物信息分析结果-单次测序

Category	1 st	
	PCR free	PCR
ReadNum	217649356	221734445
BaseNum	10882467800	11086722250
N%	0.1	0.09
GC%	50.68	50.93
Q20%	98.64	98.65
Q30%	92.75	92.71
Mapping rate(%)	99.56	99.58
ErrorRate(%)	0.08	0.09
ErrorRate !N(%)	0.04	0.06

表12：三次SE50测序的生物信息分析结果-三次测序互相校正

Category	Triple	
	PCR free	PCR
ReadNum	135180765	132040207

BaseNum	6759038250	6602010350
N%	0	0
GC%	49.59	49.79
Q20%	99.94	99.93
Q30%	99.32	99.26
Mapping rate(%)	73.62	74.62
ErrorRate(%)	0.03	0.05
ErrorRate !N(%)	0.03	0.05

由表 11 和表 12 分别所示的两样本的第一次 SE50 的单次测序分析结果和三次 SE50 测序互相校正后的结果的比较可见，三次测序后，两样本的数据量指标 BaseNum 均呈现出下降趋势，第一次测序时的 BaseNum，PCR free 文库约为 10.9Gb；PCR 文库约为 11.1Gb。经过 Exonuclease III 去除测序链，第二次

5 和第三次测序数据量有所减少。最后在三次测序互相校正中用到的 BaseNum，PCR free 文库约为 6.8Gb；PCR 文库约为 6.6Gb，可见即使经由三次重复测序，本申请实施例的重复测序方法所产生的总数据量仍体积较小，易于后期存储和处理。

同时，三次测序后，错误率指标 Mismatch rate !N(%)呈现出下降趋势，即对于第一次测序时的 Mismatch rate !N(%)，PCR free 文库为 0.04%；PCR 文库为 0.06%。而三次测序进行互相校正后的 Mismatch rate !N(%)，PCR free 文库为 0.03%，与第一次测序的错误率相比下降了 25%；PCR 文库为 0.05%，与第一次测序的错误率相比下降了 17%。此外，相较于单次测序，三次测序整合分析后的两样本的数据质量指标 Q20 和 Q30 值均有明显提升，说明本申请实施例的重复测序方法能够保证测序结果的质量，并有效提升测序结果的准确性。

实施例 6

15 采用 E.coli PCR 样本（MGI），利用 MGISEQ-2000RS 高通量测序试剂盒（MGI，货号 1000012551）在 MGISEQ-2000 平台上，按照试剂盒和测序平台的使用说明，以该样本的核苷酸序列为模板进行一次 SE50 测序（即单端 50bp 的测序），并在去除测序合成链后，再进行一次 PE50 测序（即双端各 50bp 的测序）：

1) 首先将该 E.coli PCR 样本的核苷酸序列模板固定在芯片上，通入 CPAS AD153 测序引物工作液进行退火，按照 MGISEQ-2000 的操作说明书，进行一次 SE50 测序，检测并记录此次 SE50 测序中的多轮 SBS 反应的荧光信息（其中每轮 SBS 反应测得一个碱基），通过荧光信息获得核苷酸模板此次 SE50 20 测序的碱基序列信息。

2) 通入 40μL 浓度为 50%的甲酰胺在 40℃温度条件反应 8 min，以对如上 1)中描述的 SE50 的测序链（即测序合成链）进行去除。

3) 进行 PE50 测序：通入测序试剂盒中的第一链 Read1 的测序引物进行退火，按照测序平台的操作说明书进行第一链 Read1 的 50 bp 测序，检测并记录 Read1 测序中的多轮 SBS 反应的荧光信息，通过 25 荧光信息获得第一链核苷酸模板的 50 个碱基序列信息；然后通入试剂盒中具备链置换功能的高保真聚合酶和 dNTP 底物，在此聚合酶的作用下合成第二链的测序模板（即第一链的互补链），再通入测序试剂盒中的第二链 Read2 的测序引物进行退火，接着按照测序平台的操作说明书进行第二链 Read2 的 50bp 测

序，检测并记录 Read2 测序中的多轮 SBS 反应的荧光信息，通过荧光信息获得第二链核苷酸模板的 50 个碱基序列信息。

4) 对测得的碱基序列信息进行分析，以获得碱基序列。

5 表 13 展示了步骤 1)所得的单次 SE50 测序的结果（即 1st）；表 14 展示了步骤 3)所得的 PE50 测序的结果。

表13：单次SE50测序结果

items	1 st
CycleNumber	50
Q30(%)	96.1

表14：PE50测序结果

items	PE50
CycleNumber	100
Q30(%) Read1	96.37
Q30(%) Read2	96.0
Q30(%) Total	96.19

10 由表 13 和 14 可见，第一次单次 SE 测序的质量指标 Q30 为 96.1%，经过第一遍甲酰胺反应去除测序链，对于后续的 PE50 测序的质量指标 Q30，Read1 为 96.37%，Read2 为 96.0%，Total 为 96.19%，可见使用甲酰胺去除 SE50 的测序链，再进行 PE50 测序，未显示出甲酰胺对 DNB 模板具有损伤，数据整体上保持着较高的质量。

使用 BWA MEM 比对算法，对以上 SE50 和 PE50 测序结果进行生物信息分析，数据统计如表 15 和表 16 所示。

表15：SE50测序的生物信息分析结果

items	1 st
ReadNum	555606616
BaseNum	27780330800
N%	0
GC%	50.83
Q20%	98.85
Q30%	96.1
Mapping rate(%)	99.7
ErrorRate!N(%)	0.15

15 表16：PE50测序的生物信息分析结果

items	PE50-Read1	PE50-Read2	PE50-Total
ReadNum	529249007	529249007	1058498014

BaseNum	26462450350	26462450350	52924900700
N%	0	0	0
GC%	50.78	50.82	50.8
Q20%	98.98	98.9	98.94
Q30%	96.37	96	96.19
Mapping rate(%)	99.94	99.91	99.93
ErrorRate!N(%)	0.16	0.18	0.17

由表 15 和表 16 分别所示的 SE50 测序的生物信息分析结果和其后续的 PE50 测序的生物信息分析结果可见，数据量指标 BaseNum 呈现出下降趋势，第一次 SE50 测序时的 BaseNum 约为 27.78Gb，经过甲酰胺去除测序链，第二次 PE50 测序时 PE50-Read1 的数据量与 SE50 相当，为 26.46Gb。

同时，比较 SE 测序和 PE 测序结果发现，错误率指标 Mismatch rate !N(%)、数据质量指标 Q20 和 Q30 值以及比对率 Mapping rate(%)指标相当，均指示了较高的数据质量，说明根据本申请实施例提出的重复测序方法进行 SE50 测序并去除测序链之后，后续可正常进行 PE 测序，此前进行的单次或多次单端测序不会影响后续的双端测序，因此本申请实施例中提出的重复测序方法能够与正常的单端/双端测序灵活结合，并能够获得高质量和高准确性的测序结果。

实施例 7

采用 E.coli PCR 样本 (MGI, 货号: 1000005033)，利用 MGISEQ-2000RS 高通量测序试剂盒 (MGI, 货号 1000012551) 在 MGISEQ-2000 平台上，按照试剂盒和测序平台的使用说明，以该样本的第二链 (Reverse Strand) 为模板进行四次单端的 50bp 测序 (即单端 50bp 的重复测序)：

1) 首先将该 E.coli PCR 样本的核苷酸序列模板固定在芯片上，通入试剂盒中具备链置换功能的高保真聚合酶和 dNTP 底物，在此聚合酶的作用下合成第二链的测序模板 (即第一链的互补链)。

2) 通入测序试剂盒中的第二链的测序引物 (CPAS AD153 测序引物 2 工作液) 进行退火，接着按照测序平台的操作说明书进行第二链的第一次 50bp 测序，检测并记录测序中的多轮 SBS 反应的荧光信息，通过荧光信息获得第二链核苷酸模板的第一次的 50 个碱基序列信息。

3) 通入 40μL 浓度为 50% 的甲酰胺在 40℃ 温度条件反应 8min，以对如上 2) 中描述的第一次 50 个碱基的测序链 (即测序合成链) 进行去除。

4) 再次通入 CPAS AD153 测序引物 2 工作液进行退火，进行第二次 50 个碱基的测序，检测并记录第二次测序中的多轮 SBS 反应的荧光信息，通过荧光信息获得以第二链核苷酸模板的第二次 50 个碱基序列信息。

5) 重复步骤 3) 和 4)，检测并记录第三次 50bp 测序中的多轮 SBS 反应的荧光信息，通过荧光信息获得第二链核苷酸模板的第三次 50 个碱基序列信息。

6) 重复步骤 3) 和 4)，检测并记录第四次 50bp 测序中的多轮 SBS 反应的荧光信息，通过荧光信息获得第二链核苷酸模板的第四次 50 个碱基序列信息。

7) 对四次测得的碱基序列信息进行分析，以获得高准确度的碱基序列。

表 17 展示了步骤 2) 所得的第一次 50bp 测序的结果 (即 1st)；表 18 展示了步骤 4) 和 5) 所得的第二次 (2nd)、第三次 (3rd) 和第四次 (4th) 50bp 测序的结果。

表17：第一次50bp测序结果

items	1 st
CycleNumber	50
Q30(%)	94.7

表18：第二次、第三次和第四次的50bp测序结果

items	2 nd	3 rd	4 th
CycleNumber	50	50	50
Q30(%)	94.7	94.79	93.18

由表 17 和 18 可见，第一次 50bp 测序的质量指标 Q30 为 94.7%，经过第一遍甲酰胺反应去除测序链，第二次 50bp 测序的质量指标 Q30 为 94.7%，经过第二遍甲酰胺反应去除测序链，第三次 50bp 测序的质量指标 Q30 为 94.79%，再次经过第三遍甲酰胺反应去除测序链，第四次 50bp 测序的质量指标 Q30 为 93.18%。可见使用甲酰胺去除测序链以进行多次重复测序时，第二第三次的测序质量上均没有显示出明显的下降幅度，说明本实施例的方法能够有效用于测序链的洗脱并能够有效保证测序质量；此外，甲酰胺去除测序链会对 DNB 模板产生一定的损伤，导致第四次的测序质量有微小下降，但数据整体上仍保持较高的质量，即四轮测序均保持着极高的数据质量产出。

使用 BWA MEM 比对算法，对以上四次 50bp 的二链测序所得到的序列进行互相校正，包括三次测序结果间的互相校正(Triple)以及四次测序结果间的互相校正（Quadruple），具体如下：

三次测序互相校正流程（取第一、第三、第四次测序数据进行分析）：首先，同一比对位点三次测序读取到的都是相同碱基，将此三次相同的碱基识别为该位点正确的碱基；同一比对位点两次测序读取的碱基相同、一次不同，将此两次读取的相同的碱基识别为该位点正确的碱基；同一比对位点三次测序读取的碱基都不同，则采取 Q 值权重法将 Q 值最大者作为该位点正确的碱基。

四次测序互相校正流程：首先，同一比对位点四次测序读取到的都是相同碱基，将此四次相同的碱基识别为该位点正确的碱基；同一比对位点三次测序读取的碱基相同、一次不同，将此两次读取的相同的碱基识别为该位点正确的碱基；同一比对位点两次测序读取的碱基相同、两次不同，或者四次读取的碱基都不同，则进行进一步校正：采取 Q 值权重法取 Q 值最大者成为该位点正确的碱基。

表19：第一次50bp测序的生物信息分析结果-单次测序

items	1 st
ReadNum	537471491
BaseNum	26873574550
N%	0.01
GC%	50.97
Q20%	98.5
Q30%	94.7
Mapping rate(%)	99.4
ErrorRate!N(%)	0.18

表20：四次50bp测序的生物信息分析结果-三次测序互相校正和四次测序互相校正

items	Triple	Quadruple
ReadNum	509664359	507233297
BaseNum	25483217950	25361664850
N%	0	0
GC%	50.82	50.82
Q20%	99.73	99.84
Q30%	98.41	98.84
Mapping rate(%)	99.82	99.88
ErrorRate!N(%)	0.08	0.07

表 19 为本实施例的第一次 50bp 测序的生物信息分析结果；表 20 为本实施例的三次 50bp 测序及四次 50bp 测序进行互相校正的结果。由表 19 和表 20 所示的结果可见，四次二链测序后，数据量指标 BaseNum 呈现出下降趋势，第二链第一次 50bp 测序时的 BaseNum 约为 26.87Gb，经过甲酰胺去除测序链，第二次、第三次和第四次 50bp 测序数据量稍有减少。最后在三次 50bp 测序互相校正中用到的 BaseNum 约为 25.48Gb，在四次 50bp 测序互相校正中用到的 BaseNum 约为 25.36Gb，可见即使经由三次重复测序，本申请实施例的重复测序方法所产生的总数据量仍体积较小，易于后期存储和处理。

同时，四次二链测序后，错误率指标 Mismatch rate !N(%)呈现出下降趋势，第二链第一次 50bp 测序时的 Mismatch rate !N(%)为 0.18%。三次 50bp 测序进行互相校正后的 Mismatch rate !N(%)为 0.08%，与第一次 50bp 测序的错误率相比下降了 55.6%；四次 50bp 测序进行互相校正后的 Mismatch rate !N(%)为 0.07%，与第一次 50bp 测序的错误率相比下降了 61.1%，说明本申请实施例的重复测序方法能够保证测序结果的质量，并有效提升测序结果的准确性。

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。在本公开的描述中，“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必须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、替换和变型。

权利要求书

1. 一种重复测序方法，包括：
基于第一测序引物，对样本进行第一测序，以获得基于第一测序合成链的第一测序序列；
使洗脱试剂与所述第一测序合成链接触，以去除所述第一测序合成链；
基于所述第一测序引物，对所述样本进行第二测序，以获得基于第二测序合成链的第二测序序列；
和
将所述第一测序序列和所述第二测序序列进行比对，以确定所述样本的序列信息。
2. 根据权利要求1所述的方法，其中所述基于第一测序引物，对样本进行第一测序或第二测序包括：
使所述第一测序引物与所述样本退火结合并进行多个碱基测序循环，其中每个循环完成一个碱基的信号检测，由此获得所述第一测序合成链或所述第二测序合成链的光学信号，并基于所述第一测序合成链或所述第二测序合成链的光学信号获得所述第一测序序列或所述第二测序序列。
3. 根据权利要求1或2所述的方法，其中基于第一测序引物，对样本进行第一测序还包括：
将所述样本固定于用于测序的固相载体，可选地，所述固相载体为珠粒或芯片。
4. 根据权利要求1所述的方法，其中所述洗脱试剂为以下的一种或多种：
DNA 变性剂和/或核酸外切酶，其中所述 DNA 变性剂包括：甲醇、乙醇、尿素、甲酰胺和氢氧化钠，所述核酸外切酶包括：蛇毒磷酸二酯酶、大肠杆菌核酸外切酶 I、大肠杆菌核酸外切酶 II、大肠杆菌核酸外切酶 III、脾脏磷酸二酯酶和嗜酸乳杆菌核酸酶，
优选地，所述 DNA 变性剂为甲酰胺和/或氢氧化钠，所述核酸外切酶为大肠杆菌核酸外切酶 III。
5. 根据权利要求3所述的方法，其中所述甲酰胺的浓度为 30%-100%，优选 40-100%；所述大肠杆菌核酸外切酶 III 的浓度为 1-50 U/ μ L，优选为 1-10 U/ μ L，
可选地，所述洗脱试剂的洗脱时间为 0-20 min，优选为 5-15min，更优选为 5-10min，
可选地，所述洗脱试剂的洗脱温度为 10-60°C，优选为 20-50°C，更优选为 25-45°C。
6. 根据权利要求1所述的方法，其中将所述第一测序序列和所述第二测序序列进行比对，以确定所述样本的序列信息，包括：
将所述第一测序序列和所述第二测序序列的对应位置的碱基进行比对；
响应于所述第一测序序列和所述第二测序序列的对应位置的碱基相同，将所述相同碱基确定为所述位置处的碱基类型；
响应于所述第一测序序列和所述第二测序序列的对应位置的碱基不同，通过统计学数据确定所述位置处的碱基类型，
可选地，所述统计学数据为 Q 值。

7. 根据权利要求6所述的方法，其中所述通过统计学数据确定所述位置处的碱基类型具体包括：
比较所述第一测序序列和所述第二测序序列的对应位置的碱基的所述Q值，基于所述第一测序序列的所述位置处的碱基x的所述Q值大于所述第二测序序列的所述位置处的碱基y的所述Q值，将所述碱基x确定为所述位置处的碱基类型。
8. 根据权利要求1至7中任一项所述的方法，还包括：重复基于所述第一测序引物的测序过程，包括：
对所述样本再进行m次测序，以获得基于m个测序合成链的m个测序序列；
使所述洗脱试剂分别与所述m个或m-1个测序合成链接触，以去除所述m个或m-1个测序合成链；
和
将所述第一测序序列、所述第二测序序列和所述m个测序序列进行比对，以确定所述样本的序列信息。
9. 根据权利要求8所述的方法，其中m为大于0的正整数，优选地， $1 \leq m \leq 10$ 。
10. 根据权利要求8或9所述的方法，其中所述将所述第一测序序列、所述第二测序序列和所述m个测序序列进行比对，以确定所述样本的序列信息包括：
将所述第一测序序列、所述第二测序序列和所述m个测序序列的对应位置的碱基进行比对；
响应于所述第一测序序列、所述第二测序序列和所述m个测序序列的对应位置的碱基相同，将所述相同碱基确定为所述位置处的碱基类型；
响应于所述第一测序序列、所述第二测序序列和所述m个测序序列的对应位置的碱基不同，将m+2个碱基中出现频率最高的碱基确定为所述位置处的碱基类型，或者响应于所述第一测序序列、所述第二测序序列和所述m个测序序列的对应位置的碱基不同，则通过统计学数据确定所述位置处的碱基类型。
11. 根据权利要求10所述的方法，还包括：
响应于所述第一测序序列、所述第二测序序列和所述m个测序序列的对应位置的碱基不同，且m+2个碱基中存在出现频率相同的碱基类型，则通过统计学数据确定所述位置处的碱基类型，
可选地，所述统计学数据为Q值。
12. 根据权利要求1至11中任一项所述的方法，其中所述第一测序引物是正向测序引物或反向测序引物。
13. 一种双端重复测序的方法，包括：
基于第一测序引物，根据权利要求1至12中任一项所述的重复测序方法对样本进行第一端重复测序，以获得所述样本的第一端序列信息；

基于第二测序引物，对样本进行第二端测序，以获得所述样本的第二端序列信息；和
分析所述第一端序列信息和所述第二端序列信息，以确定所述样本的序列信息，
其中所述第一测序引物为正向测序引物或反向测序引物中的一个，所述第二测序引物为正向测序引物或反向测序引物中的另一个。

14. 一种双端重复测序的方法，包括：

基于第一测序引物，根据权利要求1至12中任一项所述的重复测序方法对样本进行第一端重复测序，以获得所述样本的第一端序列信息；

基于第二测序引物，根据权利要求1至12中任一项所述的重复测序方法对样本进行第二端重复测序，以获得所述样本的第二端序列信息；和

分析所述第一端序列信息和所述第二端序列信息，以确定所述样本的序列信息，

其中所述第一测序引物为正向测序引物，所述第一端序列信息为正读序列信息，并且
所述第二测序引物为反向测序引物，所述第二端序列信息为反读序列信息。

15. 根据权利要求14所述的方法，其中基于所述第二测序引物对所述样本进行第二端重复测序包括：
生成所述样本的互补链，并基于所述互补链进行所述第二端重复测序，
可选地，所述样本为DNA纳米球，所述互补链为MDA链。

16. 根据权利要求15所述的方法，其中在生成所述样本的互补链之后，所述方法还包括：
将所述样本移除，并基于所述互补链进行所述第二端重复测序。

17. 根据权利要求16所述的方法，其中所述样本中含有条形码1和可选地条形码2，所述方法还包括：

分别对所述条形码1和所述条形码2进行测序，以获得所述条形码1和所述条形码2的序列信息；
和

根据所述条形码1和所述条形码2的序列信息，筛选出所述样本的序列信息。

18. 一种重复测序体系，包括：测序试剂和洗脱试剂，其中所述洗脱试剂为以下的一种或多种：
DNA变性剂和/或核酸外切酶，其中所述DNA变性剂包括：甲醇、乙醇、尿素、甲酰胺、氢氧化钠，
所述核酸外切酶包括：蛇毒磷酸二酯酶、大肠杆菌核酸外切酶I、大肠杆菌核酸外切酶II、大肠杆菌核酸外切酶III、脾脏磷酸二酯酶和嗜酸乳杆菌核酸酶，

优选地，所述DNA变性剂为甲酰胺和/或氢氧化钠，所述核酸外切酶为大肠杆菌核酸外切酶III。

19. 根据权利要求18所述的体系，其中所述测序试剂包括第一测序引物、反应酶和dNTP，
优选地，所述dNTP上附接荧光基团以用于碱基报告，
可选地，所述测序试剂还包括第二测序引物，

可选地，所述测序试剂还包括第一条形码测序引物和可选地第二条形码测序引物。

20. 根据权利要求 19 所述的体系，其中所述反应酶包括 DNA 聚合酶和可选地 DNA 连接酶。

21. 一种测序试剂盒，包括如权利要求 18 至 20 中任一项所述的重复测序体系。

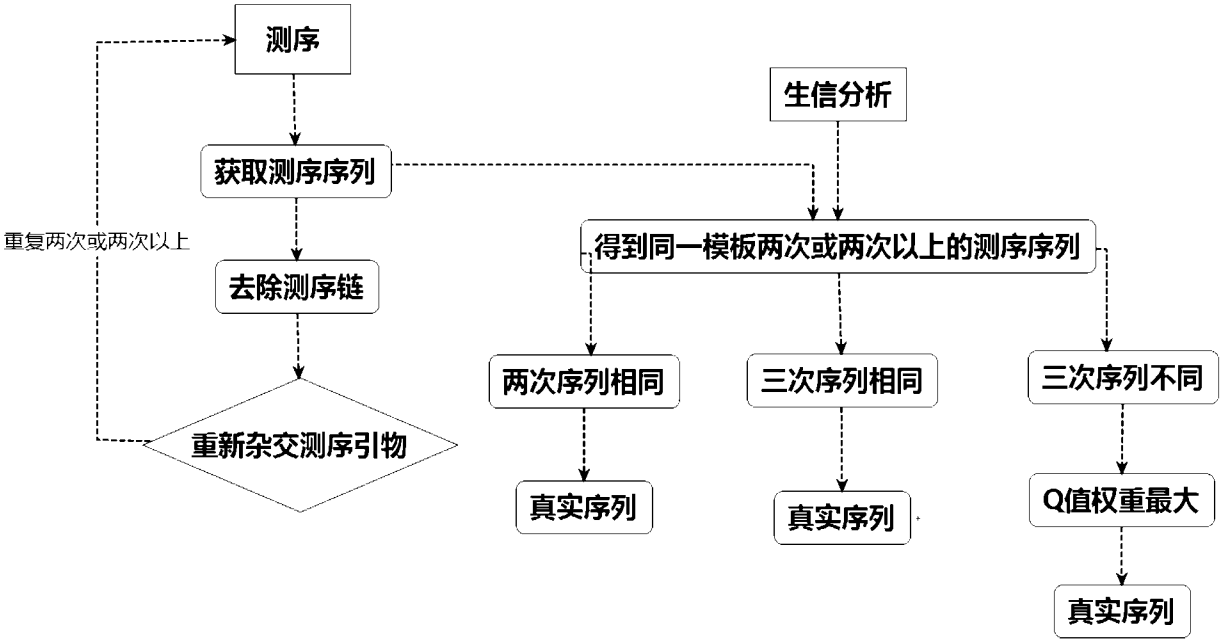


图1

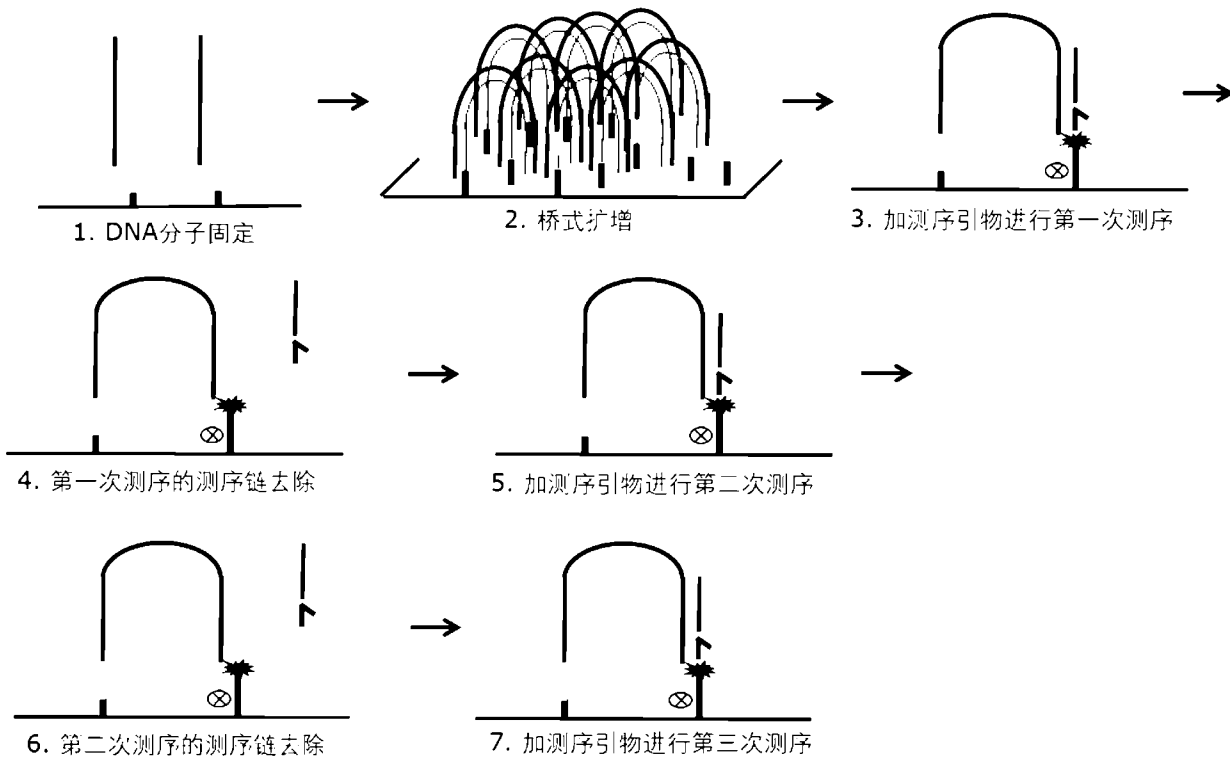


图2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/105577

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/6869(2018.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: 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, DWPI, WPABS, WPABSC, CJFD, AUABS, TWMED, ILABS, HKABS, MOABS, CNTXT, ENTXT, ENTXTC, WOTXT, USTXT, VCN, VEN, GBTXT, EPTXT, CATXT, CNKI, PubMed, EMBL, Web of Science, GoogleScholar: 测序, 重复, 二次, 再次, Q值, sequencing, duplication, repeat+, Q-score, Q phred

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101575639 A (WUXI AGENE BIOLOGICAL INFORMATION TECHNOLOGY CO., LTD.) 11 November 2009 (2009-11-11) claims and embodiments	1-21
A	US 2002012930 A1 (454 LIFE SCIENCES CORP.) 31 January 2002 (2002-01-31) claims and embodiments	1-21
A	WO 2020010495 A1 (EGI TECH (SHEN ZHEN) CO., LTD.) 16 January 2020 (2020-01-16) claims and embodiments	1-21
A	CN 113293205 A (GENEMIND BIOSCIENCES CO., LTD.) 24 August 2021 (2021-08-24) claims and embodiments	1-21
A	WO 2021142769 A1 (EGI TECH (SHEN ZHEN) CO., LTD.) 22 July 2021 (2021-07-22) claims and embodiments	1-21

☐ Further documents are listed in the continuation of Box C.☒ 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

“D” document cited by the applicant in the international application

“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

15 March 2024

Date of mailing of the international search report

29 March 2024

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/105577

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	101575639	A	11 November 2009	None	
US	2002012930	A1	31 January 2002	US 7244559 B2	17 July 2007
				WO 02077287 A1	03 October 2002
				ATE 468915 T1	15 June 2010
				US 2003148344 A1	07 August 2003
				US 7264929 B2	04 September 2007
				CA 2441603 A1	03 October 2002
				CA 2441603 C	10 January 2012
				US 2003068629 A1	10 April 2003
				AU 2002247390 A1	08 October 2002
				AU 2002247390 B2	25 January 2007
				AU 2002247390 B9	22 March 2007
				DE 60236502 D1	08 July 2010
				JP 2005505748 A	24 February 2005
				JP 4354184 B2	28 October 2009
				EP 1381693 A1	21 January 2004
				EP 1381693 A4	21 December 2005
				EP 1381693 B1	26 May 2010
				US 2003100102 A1	29 May 2003
				US 7335762 B2	26 February 2008
				US 2007092872 A1	26 April 2007
				ES 2348435 T3	07 December 2010
WO	2020010495	A1	16 January 2020	None	
CN	113293205	A	24 August 2021	None	
WO	2021142769	A1	22 July 2021	CA 3164884 A1	22 July 2021
				US 2022349001 A1	03 November 2022
				AU 2020423361 A1	15 September 2022
				JP 2023510424 A	13 March 2023
				EP 4092135 A1	23 November 2022
				EP 4092135 A4	18 January 2023

国际检索报告		国际申请号 PCT/CN2023/105577
A. 主题的分类 C12Q 1/6869(2018.01)i 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		
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) IPC: C12Q 包含在检索领域中的除最低限度文献以外的检索文献 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)) CNMED, CNABS, DWPI, WPABS, WPABSC, CJFD, AUABS, TWMED, ILABS, HKABS, MOABS, CNTXT, ENTXT, ENTXTC, WOTXT, USTXT, VCN, VEN, GBTXT, EPTXT, CATXT, CNKI, PubMed, EMBL, Web of Science, GoogleScholar: 测序, 重复, 二次, 再次, Q值, sequencing, duplication, repeat+, Q-score, Q phred		
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 101575639 A (无锡艾吉因生物信息技术有限公司) 2009年11月11日 (2009 - 11 - 11) 权利要求书和实施例	1-21
A	US 2002012930 A1 (454 LIFE SCIENCES CORP) 2002年1月31日 (2002 - 01 - 31) 权利要求书和实施例	1-21
A	WO 2020010495 A1 (EGI TECH SHEN ZHEN CO LTD) 2020年1月16日 (2020 - 01 - 16) 权利要求书和实施例	1-21
A	CN 113293205 A (深圳市真迈生物科技有限公司) 2021年8月24日 (2021 - 08 - 24) 权利要求书和实施例	1-21
A	WO 2021142769 A1 (MGI TECH CO LTD) 2021年7月22日 (2021 - 07 - 22) 权利要求书和实施例	1-21
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 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 “D” 申请人在国际申请中引证的文件 “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 “&” 同族专利的文件		
国际检索实际完成的日期 2024年3月15日		国际检索报告邮寄日期 2024年3月29日
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088		受权官员 张艳霞 电话号码 (+86) 010-62089438

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/105577

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	101575639	A	2009年11月11日	无			
US	2002012930	A1	2002年1月31日	US	7244559	B2	2007年7月17日
				WO	02077287	A1	2002年10月3日
				ATE	468915	T1	2010年6月15日
				US	2003148344	A1	2003年8月7日
				US	7264929	B2	2007年9月4日
				CA	2441603	A1	2002年10月3日
				CA	2441603	C	2012年1月10日
				US	2003068629	A1	2003年4月10日
				AU	2002247390	A1	2002年10月8日
				AU	2002247390	B2	2007年1月25日
				AU	2002247390	B9	2007年3月22日
				DE	60236502	D1	2010年7月8日
				JP	2005505748	A	2005年2月24日
				JP	4354184	B2	2009年10月28日
				EP	1381693	A1	2004年1月21日
				EP	1381693	A4	2005年12月21日
				EP	1381693	B1	2010年5月26日
				US	2003100102	A1	2003年5月29日
				US	7335762	B2	2008年2月26日
				US	2007092872	A1	2007年4月26日
				ES	2348435	T3	2010年12月7日
WO	2020010495	A1	2020年1月16日	无			
CN	113293205	A	2021年8月24日	无			
WO	2021142769	A1	2021年7月22日	CA	3164884	A1	2021年7月22日
				US	2022349001	A1	2022年11月3日
				AU	2020423361	A1	2022年9月15日
				JP	2023510424	A	2023年3月13日
				EP	4092135	A1	2022年11月23日
				EP	4092135	A4	2023年1月18日