

(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
2016 年 8 月 11 日 (11.08.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/124069 A1

(51) 国际专利分类号:

C12Q 1/68 (2006.01) C40B 50/06 (2006.01)
C12N 15/10 (2006.01) C40B 60/14 (2006.01)
C12M 1/00 (2006.01)

(21) 国际申请号: PCT/CN2016/070789

(22) 国际申请日: 2016 年 1 月 13 日 (13.01.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:

201510058616.9 2015 年 2 月 4 日 (04.02.2015) CN

(71) 申请人: 深圳华大基因研究院 (BGI SHENZHEN)

[CN/CN]; 中国广东省深圳盐田区北山工业区综合楼, Guangdong 518083 (CN)。深圳华大基因科技有限公司 (BGI SHENZHEN CO., LIMITED) [CN/CN]; 中国广东省深圳盐田区北山路 146 号北山工业区综合楼 11F-3, Guangdong 518083 (CN)。

(72) 发明人: 王欧 (WANG, Ou); 中国广东省深圳盐田区北山工业区综合楼, Guangdong 518083 (CN)。

程小芳 (CHENG, Xiaofang); 中国广东省深圳盐田区北山工业区综合楼, Guangdong 518083 (CN)。邹良英 (ZOU, Liangying); 中国广东省深圳盐田区北山工业区综合楼, Guangdong 518083 (CN)。常灿坤 (CHANG, Cankun); 中国广东省深圳盐田区北山工业区综合楼, Guangdong 518083 (CN)。蒋慧 (JIANG, Hui); 中国广东省深圳盐田区北山工业区综合楼, Guangdong 518083 (CN)。章文蔚 (ZHANG,

Wenwei); 中国广东省深圳盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(74) 代理人: 北京纪凯知识产权代理有限公司 (JEEKAI & PARTNERS); 中国北京市丰台区广安路 9 号 5 号楼 15A 层, Beijing 100055 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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: METHOD FOR CONSTRUCTING LONG FRAGMENT SEQUENCING LIBRARY

(54) 发明名称: 一种构建长片段测序文库的方法

(57) Abstract: Provided in the present invention is a method for constructing a long fragment sequencing library. The method comprises the following steps: 1) preparing a 5184-well plate containing a single-stranded DNA; 2) placing the single-stranded DNA in equal amounts into each well on the 5184-well plate; 3) placing an entire genome amplification reaction system into each well of the 5184-well plate processed in step 2) and reacting the same; 4) placing a fragmented reaction solution into each well of the 5184-well plate processed by the step 3) and reacting the same; 5) adding different tag sequences to nucleic acid molecules in each well of the 5184-well plate processed in step 4) to obtain the sequencing library.

(57) 摘要: 本发明提供了一种构建长片段测序文库的方法。本发明提供的方法包括如下步骤: 1) 制备含有单链 DNA 的 5184 孔板; 2) 将所述单链 DNA 等量分装到 5184 孔板每个孔中; 3) 将全基因组扩增反应体系分装到步骤 2) 处理后的 5184 孔板每个孔中, 反应; 4) 将片段化反应液分装到步骤 3) 处理的 5184 孔板每个孔中, 反应; 5) 将步骤 4) 处理后的 5184 孔板每个孔中的核酸分子添加不同标签序列, 即得到测序文库。



WO 2016/124069 A1

一种构建长片段测序文库的方法

技术领域

本发明涉及生物技术领域，尤其涉及一种构建长片段测序文库的方法。

5 背景技术

新一代高通量并行测序仪使基因组学领域发生了革命性变化，基因测序所需成本和时间大大降低，同时一些新的应用，诸如宏基因组测序，基因结构变异和基因表达分析也被开发了出来。尽管如此，目前其他物种基因组的组装，离达到人类基因组项目设定的标准的，常规的基因组还有非常长的距离。受限于测序原理和技术瓶颈，现有的高通量并行测序仪的测序读长分布从几百碱基到几千碱基不等，而需要进行研究的基因组或基因组元件的或可多达几亿个碱基，这要求生物信息学家将这些片段化的信息重新还原成生物体本身的长片段染色体信息，即对测序生成的短片段进行组装。

事实上，基因组的组装效果受到很多方面的制约，除了制定的测序策略，测序数据质量，以及所使用的组装软件等人为因素外，待研究的基因组的自身特性也会影响基因组组装的效果，其中较为重要的两个因素是基因组中的重复区域和基因组的杂合程度。由于现有测序手段的读长较短，无法跨越重复区域而造成拼接失败。而基因组的杂合程度过高则会导致组装软件将同源染色体单独组装出来，从而造成组装的基因组偏离于基因组的真实情况。在不改变现有测序仪读长的情况下，这些因素无法通过改进组装拼接算法完全消除。此外，潜在的测序错误，以及文库构建过程中扩增导致的偏向性及错误都会对组装效果产生负面影响。

因此越来越多的研究者试图通过在实验设计方面进行改进从而提升基因组拼装效果。针对基因组重复区域，传统的方法一般通过增加文库跳跃长度来辅助组装，如构建不同长度的 Mate pair jump 文库或 Fosmid 文库，用于跨过不同大小的基因组重复区域。另外还可以采用混合测序类型的方法，如使用 PacBio 测序仪长读长序列生成长脚手架序列，然后利用 illumina 测序仪的短读长进行错误修正，从而达到较好的组装效果。

而对于高杂合度引起的组装问题，可以通过单倍型组装定相（Phasing）来解决。即通过实验手段将单倍体信息从多倍体信息中分离出来，从而使单倍体型被完整的组装出来。目前已经有一些研究使用不同的研究手段获得了样品的单倍体型信息，这些手段包括：1. 通过对样品以及样品的父本进行全基因组测序进而获得样品的单倍体型信息。2. 通过使用 Fosmid 测序方法进行单倍体型测序。3. 在细胞分裂中期，使用显微操作技术将染色体分离并测序，进而获得单倍体型信息。4. 通过临近随机连接法进行单倍体测序。

然而，以上几种单倍型定相方法均具有一定的局限性：1. 同时对父代和子代样本进行测序，然后根据基因型进行单倍型定相的方法要求同时拥有父本和

母本的样品，这大大限制了它的使用范围，并且该方法无法对 De novo 突变进行检测。2. 使用 Fosmid 测序的方法需要至少一周的文库制备时间，包含大量建库实验，因此该方法需要微克级的样品作为起始。无法针对临床少量样品进行分析。3. 染色体分离方法需要有复杂的专业显微操作设备，同时要求实验人员的操作水平非常高。4. 受限于实验原理，临近连接法检测到的突变体有限，只能检测到 80%左右的 SNV，不能满足临床分析的需要。因此，为了应对个性化医疗的需要，急需一种高准确度，高覆盖度，低成本，低起始量，实验条件相对简便的单倍体型测序技术。

针对这一情况，研究人员研发了使用短读序列组装成长读长片段，从而进行单倍体型组装的方法。代表技术有 Complete Genomics 公司的长片段读取 (Long Fragment Read, 简称 LFR 技术, 下同) 技术, 以及 illumina 公司旗下的 Truseq synthetic long read 产品。在实验原理上, 这两者与 Fosmid 单倍型分型方法相似, 都是通过将 DNA 随机分离到不同物理分隔中, 以实现将不同来源的同源染色体分离的效果。而与 Fosmid 测序不同的是, 这两者的文库构建时间均有明显缩短, 并且无需大量操作, 单倍型组装效果均可达到 Haplotype N50 500kb 左右。特别值得一提的是, LFR 技术仅需要 100pg 左右的 DNA, 即 10-20 个细胞的基因组作为起始即可完成文库构建, 并可以覆盖 92%以上的 SNV 位点, 其正确率可以达到 99.99999%, 相比同样基于连接测序的全基因组测序方法的正确率提高了 10 倍。

也应承认, 目前的短读组装技术仍存在一定不足之处。首先, 该类型方法目前多使用 384 孔板作为物理分隔, 理论上讲每个分隔内包含 10%-20%的基因组, 仍然是较高的单分隔基因组容量, 组装上仍存在一定难度。其次, illumina 公司的 Truseq synthetic long read 产品在起始步骤将 DNA 打断至 8-10kb 的片段, 该方法对具有大重复区域的基因组的组装效果不理想。而 Complete Genomics 的 LFR 技术起始量低的样本进行建库, 需要使用多重链置换反应 (Multiple Displacement Amplification, MDA) 对样品进行大量扩增, 该过程会带来一些扩增偏好性。

发明公开

本发明的一个目的是提供一种构建长片段测序文库的方法。

本发明提供的方法, 包括如下步骤:

1) 制备含有单链 DNA 的 5184 孔板;

所述制备含有单链 DNA 的 5184 孔板的方法为如下 A) 或 B) :

A) 先提取待测样品基因组 DNA, 变性, 得到单链 DNA, 再将所述单链 DNA 分子分装到 5184 孔板的每个孔中, 得到含有单链 DNA 的 5184 孔板;

B) 先将待测样品分装到 5184 孔板的每个孔中, 再裂解变性, 得到含有单链 DNA 的 5184 孔板;

所述单链 DNA 的片段长度不小于 100Kb;

2) 将全基因组扩增反应体系分装到所述含有单链 DNA 的 5184 孔板的每个孔中, 进行全基因组扩增反应, 得到含有全基因组扩增产物的 5184 孔板;

3) 将片段化反应体系分装到所述含有全基因组扩增产物的 5184 孔板的每个孔中, 进行片段化反应, 得到片段化产物, 含有所述片段化产物的孔板即为含有片段化产物的 5184 孔板;

所述片段化产物的长度为 200-1500bp;

4) 对所述含有片段化产物的 5184 孔板每个孔中片段化产物添加不同标签序列, 即得到测序文库。

上述方法中, 所述含有单链 DNA 的 5184 孔板中的单链 DNA 总量满足 5184 孔板中每个孔内出现来自父母双方相同基因组位置的 (两条有相互交叉的) 同源染色体片段的概率小于 1%;

所述含有单链 DNA 的 5184 孔板中的单链 DNA 总量具体为 10-500 个细胞的 DNA 量。

上述方法中, 所述 5184 孔板每孔的容积为 190 纳升至 350 纳升, 所述 5184 孔板每孔的容积具体为 200 纳升至 350 纳升。

上述方法中, 步骤 1) 中, 所述变性或所述裂解变性采用的试剂为碱变性试剂;

所述变性的反应条件为 25 度孵育 2 分钟;

所述裂解变性的反应条件为 85 摄氏度孵育 2 分钟;

所述提取待测样品基因组 DNA 的方法采用透析法或碱裂解法或琼脂糖包埋法。

上述方法中, 步骤 2) 中, 所述全基因组扩增反应体系中含有由 8 个碱基组成的随机引物。

上述方法中, 在步骤 1) 和步骤 2) 之间, 还包括如下步骤: 将由 8 个碱基组成的随机引物分装到所述含有单链 DNA 的 5184 孔板中。

上述方法中, 步骤 2) 中, 所述全基因组扩增反应体系为多重链置换聚合酶扩增反应体系;

所述全基因组扩增反应为 30℃孵育 1 小时, 再 65℃孵育 5 分钟;

或所述全基因组扩增反应为 37℃孵育 45 分钟, 再 65℃孵育 5 分钟。

上述方法中, 步骤 3) 中, 所述片段化反应体系为转座酶反应体系;

所述转座酶反应体系包括转座酶反应缓冲液和包埋接头的转座酶;

所述片段化反应的条件为 55℃孵育 10 分钟;

所述接头为接头 1 和/或接头 2。

所述接头 1 和接头 2 为序列不同的接头。

上述方法中, 步骤 4) 中, 所述添加不同标签序列的方法包括如下步骤: 将 72 种 3' 端标签引物、聚合酶链式反应体系和 72 种 5' 端标签引物分装到所述含有片段化产物的 5184 孔板各孔中, 进行聚合酶链式反应, 得到原始测序文库;

所述 3' 端标签引物从 5' 至 3' 方向依次由测序接头 A、8 个碱基组成随机片段和与所述接头 1 互补的单链 DNA 分子组成；

所述 5' 端标签引物从 5' 至 3' 方向依次由测序接头 B、8 个碱基组成随机片段和与所述接头 2 互补的单链 DNA 分子组成；

5 所述 72 条 3' 端标签引物和 72 条 5' 端标签引物的 8 个碱基组成随机片段均不相同；

所述聚合酶链式反应体系含有引物 1 和引物 2；

所述引物 1 与所述测序接头 A 相同或互补；

所述引物 2 与所述测序接头 B 互补或相同。

10 所述测序接头 A 和测序接头 B 为序列不同的接头。

上述方法中，所述 72 种 3' 端标签引物的分装方式为：将所述 72 种 3' 端标签引物分装到所述含有片段化产物的 5184 孔板的 72 列的每个孔中；

所述 72 种 5' 端标签引物的分装方式为：将所述 72 种 5' 端标签引物分装到所述含有片段化产物的 5184 孔板的 72 行的每个孔中；

15 所述聚合酶链式反应体系的分装方式为：将所述聚合酶链式反应体系分装到所述含有片段化产物的 5184 孔板的每个孔中。

上述方法中，所述聚合酶链式反应的退火温度为 60℃ 30 秒。

上述方法中，所述分装的方法均为：先将各待分装物质分装到 384 孔板中的分装模式孔中，再将所述分装模式孔中的物质用纳升级分液器分装到所述
20 5184 孔板各孔中；

所述分装的量均为各孔等量分装。

上述方法中，所述 384 孔板的分装模式孔符合 4×2 排列，且长边为 2 短边为 4，所述分装模式孔的孔数范围 8-384 孔，所述分装模式孔具体为 24 孔或 72 孔。

25 上述方法中，在步骤 4) 中得到原始测序文库后，还包括如下步骤：将所述原始测序文库进行文库片段分选，得到长度在 200-1100bp 范围内的片段选择文库，即为长片段测序文库。

上述方法中，所述 200-1100bp 片段选择文库为 250-550 bp 片段选择文库和 550-1000 bp 片段选择文库；

30 所述分选的方法为磁珠纯化或琼脂糖电泳纯化；

上述方法中，所述待测样品为真核细胞或混合物微生物细胞，所述真核细胞来源于人类血液细胞；

所述混合物微生物细胞来源于含有混合微生物的粪便或含有混合微生物土壤。

35 由上述方法制备的长片段测序文库也是本发明保护的范围。

上述方法在长片段 DNA 分子测序中的应用也是本发明保护的范围；

所述长片段 DNA 分子不小于 100Kb；

上述方法在混合微生物纯化单菌落中的应用也是本发明保护的范围；
所述混合微生物具体来源于粪便或土壤。

本发明的另一个目的是提供一种制备长片段测序文库的成套产品。

本发明提供的产品，包括 384 孔板、上述 5184 孔板和纳升级分液器。

5 本发明的第三个目的是提供一种上述定制的 5184 孔板。

本发明提供的定制的 5184 孔板，其特征在于：所述孔板每孔容积为 200-350 纳升。

所述变性试剂为碱变性试剂，配方如实施例中的表 1；

10 所述 384 孔板的样品分装模式符合 4*2 且长边为 2 短边为 4，分装孔数范围 8-384 孔，优选 24 孔或 72 孔。

所述包埋接头为接头 1 和/或接头 2；可根据需要自行设计。本实施例的接头 1 和接头 2 的设计是针对 illumina 测序仪应用而设计的：

接头 1 和接头 2 序列

接头 1 序列及结构：

15 5' -GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG-3'
3' -TCTACACATATTCTCTGTC-5'

接头 2 序列及结构：

5' -GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG-3'
3' -TCTACACATATTCTCTGTC-5'

20 若更换其他的测序平台，可以根据下方设计原则，更换相应的序列。接头 1 和接头 2 的序列的设计原则为：

1. 设计的接头 1 和接头 2 是由两条链组成，一条长链一条短链。

2. 长链端的 3' 末尾 19bp 序列固定为 AGATGTGTATAAGAGACAG（5' 至 3' 方向）。

25 3. 长链的 5' 末端可根据不同的测序平台更换为不同序列。如针对 Life Technology 公司的 Ion Proton 测序仪，接头 1 长链变更为 CCTCTCTATGGGACGTCGGTGATAGATGTGTATAAGAGACAG（5' 至 3' 方向）；接头 2 的长链变更为：CCATCTCATCCCTGCGTGTCTCCGACTCAGAGATGTGTATAAGAGACAG（5' 至 3' 方向）。

30 4. 短链为一条 19bp 的寡聚核苷酸链，序列为 CTGTCTCTTATACACATCT（5' 至 3' 方向），它与 2 中所述长链端的 3' 末尾 19bp 序列为反向互补关系。

5. 应当注意到，接头 1 与接头 2 的序列可以相同，也可以不相同。

6. 针对不同的测序仪的接头 1 和接头 2，需要按照 PCR 引物设计原则，有针对性的设计不同的引物 1、2、3 和 4。

35 附图说明

图 1 为长片段测序技术原理示意图。

图 2 为长片段测序文库构建方法流程图。

图 3 为本方法建库时间与 illumina 常规文库建库所需时间对比。

图 4 为本方法建库时间与主流单倍型文库建库所需时间对比。

图 5 为高分子量 DNA 提取电泳结果示意图。

图 6 为给定的分隔数量下，细胞数量与父本母本片段交叠概率关系图。

5 图 7 为 384 孔储液来源板加液位置示意图 1。

图 8 为 384 孔储液来源板加液位置示意图 2。

图 9 为未进行片段选择文库的 Agilent 2100 bioanalyzer 检测结果图。

图 10 为片段选择文库（250-550bp）的 Agilent 2100 bioanalyzer 检测结果图。

10 图 11 为片段选择文库（550-1000bp）的 Agilent 2100 bioanalyzer 检测结果图。

图 12 为长片段 100kb 拼接结果示意图（请将英文替换为中文）。

实施发明的最佳方式

下述实施例中所使用的实验方法如无特殊说明，均为常规方法。

15 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。

实施例1、长片段测序文库的构建

一、待测序单链DNA等量分装到5184孔板中每个孔中

1、长片段基因组DNA的提取

提取待测样本基因组 DNA，且使提取的基因组 DNA 片段不小于 100kb；

20 具体提取方法如下：

使用 Agilent RecoverEase 透析法 DNA 提取试剂盒，提取人血液中的白细胞（也可以为培养细胞，组织样品则需进行研磨匀浆处理。）基因组 DNA，提取后使用脉冲场凝胶电泳检测其片段长度分布，检测条件为：6V/cm，50-90 秒转换范围，运行时间 20 小时。结果如图 5 中第三泳道，看出提取片段长度分布主带大于 100kb，片段跨度为 50kb-800kb。

2、长片段基因组 DNA 双链变为单链 DNA

1)、反应起始用量

30 本例以人类基因组为例，人类基因组包含 6Gb 单倍体型信息，经计算采用 150 皮克，即 20 个人类细胞所含 DNA 质量的 DNA 为最终的反应起始用量。该用量保证了在所有物理分隔中，同一分隔内出现来自父母双方相同基因组位置的（两条有相互交叉的）同源染色体片段的概率小于 1%。待测样品的来源（物种）不同，DNA 用量也不同，范围为 10-500 个细胞的 DNA 量。

设细胞数量 Cells= n，物理分隔数量（孔数）Wells = w，细胞为二倍体基因组；

35 定义每一个孔中存在的基因组数量 $\lambda = 4n/w$ ；

可推出同一孔中出现来自父本和母本相同基因组位置的同源染色体片段的概率公式：

$$P = \sum_{x=2}^{\infty} p(x), \quad p(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

X 为单碱基位点的测序深度；

由于孔中长片段来自父本或母本的概率都为 50%，因此最终每个孔中长片段位于基因组相同位置，且分别来自父本和母本的概率： $P*50\%$ 。

5 由此公式可以推出，在同一孔中出现来自父本和母本的同源染色体片段的概率只与投入的细胞数量和物理分隔数量有关，而与投入片段长度和基因组大小无关。

在给定的分隔数量下，可以生成细胞数量与概率的分布图(图 6)。按照该分布图，可以推算出在某概率下，需要投入的细胞数量。

10 对于本发明为了满足同一分隔内出现来自父母双方相同基因组位置的同源染色体片段的概率小于 1%，则 5184 孔板中 DNA 加入量为 10-500 个细胞的 DNA 量。

2)、双链变单链

15 将 DNA 的浓度调整至 664.56 pg/ μ L，使用剪口移液吸头吸取 9.57 μ L DNA 样品置于 1.5mL 离心管中备用。随后将 20 μ L 碱变性试剂原液（配方如表 1 所示）稀释 10 倍至 200 μ L，取 172.3 μ L 稀释后液体加入到上述 1.5mL 离心管中，轻弹管壁混匀。室温静止两分钟（变性）后将反应体系转移至于冰上，注意此处处理时间不得超过 5 分钟，否则会对 DNA 样品造成损害。该步骤的目的是破坏其 DNA 分子双链间的氢键，使其成为单链 DNA。

20 表 1 为碱变性试剂配方

碱变性试剂	
试剂	储液
氢氧化钾	8960mg
500mM EDTA	2000uL
总体积	加水至 100mL

3、单链 DNA 分装到 5184 孔板中

1)、随机引物的添加

25 取出 13.40 μ L 步骤二的 2 得到的样品，加入到 49.78 μ L 的 1mM 浓度的 8 碱基随机引物中，无需混匀。室温静止两分钟。随后向样品中加入 496.78 μ L 无核酸酶的水，使总体积达到 560 μ L。

2)、将步骤 1 中处理好的样品，使用阔口移液器吸头分装至 384 孔板中的 24 个孔（具体位置如图 7 蓝色标记所示，24 孔为第 1 列和第 2 列的前 12 孔，样品分装模式符合 4*2 且长边为 2 短边为 4，分装孔数范围 8-384 孔，优选 24

孔，但不限于。此处样品分装模式为 3 组精密排列的 4*2 的孔，每孔体积 22 微升（每孔 DNA 含量为 18.41 皮克）。

- 3)、使用 WaferGenMultisampleNanodispenser 仪器（分液器），将步骤 2 中准备好 384 孔板中的样品，使用“LFR 35nl Single Sample dispensing.seq”程序等量分装到定制化的 5184 孔板的各孔中，每个孔里面加 35nl 体积，每个孔中 DNA 含量为 0.0293 皮克（1 个人白细胞细胞核内 DNA 含量为 6.5 皮克，5184 个孔中的细胞总数为 23 个），得到含有 DNA 单链的 5184 孔板。

定制化的 5184 孔板为适合分液器 5184 孔板，且每孔容积为 200nl 的孔板。定制化的 5184 孔板每孔容积范围在 200nl-350nl，也可用普通标准 5184 孔板，视不同来源的 DNA 起始量而定。

二、全基因组扩增反应

- 1、上述一得到的含有 DNA 单链的 5184 孔板使用封口膜封好，将容器放入 Eppendorf 5810 孔板离心机，3220 ×g，离心 5 分钟。室温放置备用（为了使溶液放在管底部）。
- 2、在 1.5mL 离心管中，依据表 2 加入配置好的多重链置换聚合酶扩增缓冲反应液（全基因组扩增反应体系），之后进行震荡混匀，离心后将该反应液分装到 384 孔板中的 24 孔，具体位置如图 7 所示，每孔体积 22 微升。

表 2 为多重链置换聚合酶扩增缓冲反应液

试剂	体积（微升）
ddH ₂ O	421.92
10x Phi Buffer（Enzymatics #B7020）	110.21
10% Pluronic F68	1.18
25mM dNTPs	11.85
Phi29（Enzymatics #P7020-LC-L）	14.81

- 3、小心移除容器表面的封膜，使用 WaferGenMultisampleNanodispenser 仪器，将步骤 2 中 384 孔板 24 孔中的样品，使用“LFR 35nl Single Sample dispensing.seq”程序等量分装到经过上述 1) 处理后的含有 DNA 单链的 5184 孔板各孔中。

- 4、加液完成后，将容器使用封口膜封好，将容器放入 Eppendorf 5810 孔板离心机，3220 ×g，离心 5 分钟。随后将其放入适配的温浴设备中，30℃ 孵育 1 小时（也可以 37℃ 孵育 45 分），再 65℃ 孵育 5 分钟。最后冷却至室温备用，得到全基因组扩增产物，得到含有全基因组扩增产物的 5184 孔板。

三、片段化

将上述二得到的全基因组扩增产物进行片段化，使其大小为 200-1500bp，采用包埋接头 1 和接头 2 的转座酶进行片段化；

- 1、在 1.5mL 离心管中，依据表 3 说明配置好的包埋接头 1 和接头 2 的转座酶反应体系。之后上下颠倒二十次混匀，短暂离心后将该反应体系分装到 384

孔板的 24 孔中，具体位置如图 7 所示，每孔体积 22 微升。

接头 1 序列及结构：

5' -GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG-3'
3' -TCTACACATATTCTCTGTC-5'

5 接头 2 序列及结构：

5' -GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACAG-3'
3' -TCTACACATATTCTCTGTC-5'

灰色背景内碱基通过碱基互补配对原则，以氢键作用相互结合在一起，保证两条序列组成一个特定的接头。

10 表 3 为用于片段化的转座酶反应液

试剂	体积（微升）
5 x Tagment Buffer (Vazyme # TD108-02)	336.00
Tagment Enzyme Mix* (Vazyme # TD108-02)	224.00

2、小心移除容器表面的封膜，之后使用

15 WaferGenMultisampleNanodispenser 仪器，将步骤 1 中准备好的 384 孔板 24 孔中的样品，使用“LFR 35nl Single Sample dispensing.seq”程序等量分装到上述二得到的含有全基因组扩增产物的 5184 孔板各孔中。

3、加液完成后，将该容器使用封口膜封好，将容器放入 Eppendorf 5810 孔板离心机，3220 ×g，离心 5 分钟。随后将其放入适配的温浴设备中，55℃ 孵育 10 分钟，之后冷却至室温备用。

20 4、依据表 4 配置好转座酶中和反应液，之后震荡混匀，短暂离心后将该反应液通过装到 384 孔板的 24 孔中，具体位置如图 7 所示，每孔体积 14 微升，再将 384 孔板 24 孔中的样品，使用“LFR 35nl Single Sample dispensing.seq”程序等量分装到上述 3 得到的 5184 孔板各孔中，得到 200-1500bp 的片段化产物，即得到含有片段化产物 5184 孔板。

25 片段化产物通过 Agilent Bioanalyzer 2100 进行毛细管电泳确定片段大小分布，如图 10 和图 11，可以看出大小为 200-1500bp。

表 4 为转座酶中和反应液

试剂	体积（微升）
5X NT buffer (Vazyme # TD108-02)	823.2
ddH2O	164.64

四、加标签序列得到测序文库

1、加标签序列用于区分不同样本

30 1)、取出已经分装在 96 孔板内的 72 种带有标签序列的 3' 端引物 (2.5 μmol/L，序列如表 5 所示)，使用八道移液器取 0.56 μL，加入到 384 孔

板的 72 孔中, 如图 8 绿色区域所示。将 384 孔板置于混匀仪上震荡混匀, 随后将孔板放入 Eppendorf 5810 孔板离心机, $3220 \times g$, 离心 5 分钟。检查无气泡后室温静置备用。

- 3' 端标签引物从 5' 至 3' 方向依次由测序接头 A、8 个碱基组成随机片段
5 和与接头 1 互补的单链 DNA 分子组成; 72 条 3' 端标签引物的 8 个碱基组成随机片段均不相同;

表 5 3' 端标签序列引物

名称	序列 5' -3'	序列编号
N701	<u>CAAGCAGAAGACGGCATACGAGAT</u> <i>TCGCC</i> TTAGTCTCGTGGGCTCGG	序列 1
N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGG	序列 2
N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGG	序列 3
N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG	序列 4
N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG	序列 5
N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG	序列 6
N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGG	序列 7
N708	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGG	序列 8
N709	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGG	序列 9
N710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGG	序列 10
N711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGG	序列 11
N712	CAAGCAGAAGACGGCATACGAGATTCTCTACGTCTCGTGGGCTCGG	序列 12
N713	CAAGCAGAAGACGGCATACGAGATAAGCAATGGTCTCGTGGGCTCGG	序列 13
N714	CAAGCAGAAGACGGCATACGAGATAATCCGAAGTCTCGTGGGCTCGG	序列 14
N715	CAAGCAGAAGACGGCATACGAGATAATGATGAGTCTCGTGGGCTCGG	序列 15
N716	CAAGCAGAAGACGGCATACGAGATACAGGAGCGTCTCGTGGGCTCGG	序列 16
N717	CAAGCAGAAGACGGCATACGAGATACCGAGCTGTCTCGTGGGCTCGG	序列 17
N718	CAAGCAGAAGACGGCATACGAGATACCTGTTGGTCTCGTGGGCTCGG	序列 18
N719	CAAGCAGAAGACGGCATACGAGATACCTGAAGTCTCGTGGGCTCGG	序列 19
N720	CAAGCAGAAGACGGCATACGAGATACTACGTGGTCTCGTGGGCTCGG	序列 20
N721	CAAGCAGAAGACGGCATACGAGATACTCTTACGTCTCGTGGGCTCGG	序列 21
N722	CAAGCAGAAGACGGCATACGAGATAGAAGGTAGTCTCGTGGGCTCGG	序列 22
N723	CAAGCAGAAGACGGCATACGAGATAGAGACTTGTCTCGTGGGCTCGG	序列 23
N724	CAAGCAGAAGACGGCATACGAGATAGATCTCTGTCTCGTGGGCTCGG	序列 24
N725	CAAGCAGAAGACGGCATACGAGATAGGTTTCATGTCTCGTGGGCTCGG	序列 25
N726	CAAGCAGAAGACGGCATACGAGATAGTCTGGTGTCTCGTGGGCTCGG	序列 26
N727	CAAGCAGAAGACGGCATACGAGATAGTTATAGGTCTCGTGGGCTCGG	序列 27
N728	CAAGCAGAAGACGGCATACGAGATAGTTCCGCGTCTCGTGGGCTCGG	序列 28

N729	CAAGCAGAAGACGGCATAACGAGATATAACTAGGTCTCGTGGGCTCGG	序列 29
N730	CAAGCAGAAGACGGCATAACGAGATATATAAGAGTCTCGTGGGCTCGG	序列 30
N731	CAAGCAGAAGACGGCATAACGAGATATCGATTCTCGTGGGCTCGG	序列 31
N732	CAAGCAGAAGACGGCATAACGAGATATCTTATTGTCTCGTGGGCTCGG	序列 32
N733	CAAGCAGAAGACGGCATAACGAGATATGGCATACTCGTGGGCTCGG	序列 33
N734	CAAGCAGAAGACGGCATAACGAGATATTAGAATGTCTCGTGGGCTCGG	序列 34
N735	CAAGCAGAAGACGGCATAACGAGATCAACATTAGTCTCGTGGGCTCGG	序列 35
N736	CAAGCAGAAGACGGCATAACGAGATCAAGTAACGTCTCGTGGGCTCGG	序列 36
N737	CAAGCAGAAGACGGCATAACGAGATCAGTGAATGTCTCGTGGGCTCGG	序列 37
N738	CAAGCAGAAGACGGCATAACGAGATCATATGATGTCTCGTGGGCTCGG	序列 38
N739	CAAGCAGAAGACGGCATAACGAGATCATTAAAGCGTCTCGTGGGCTCGG	序列 39
N740	CAAGCAGAAGACGGCATAACGAGATCCATATCCGTCTCGTGGGCTCGG	序列 40
N741	CAAGCAGAAGACGGCATAACGAGATCCATCAAGGTCTCGTGGGCTCGG	序列 41
N742	CAAGCAGAAGACGGCATAACGAGATCCGATCTTGTCTCGTGGGCTCGG	序列 42
N743	CAAGCAGAAGACGGCATAACGAGATCCGGTTAAGTCTCGTGGGCTCGG	序列 43
N744	CAAGCAGAAGACGGCATAACGAGATCGACTTAGGTCTCGTGGGCTCGG	序列 44
N745	CAAGCAGAAGACGGCATAACGAGATCGCGAATAGTCTCGTGGGCTCGG	序列 45
N746	CAAGCAGAAGACGGCATAACGAGATCGTGCTTCGTCTCGTGGGCTCGG	序列 46
N747	CAAGCAGAAGACGGCATAACGAGATCTACTGGAGTCTCGTGGGCTCGG	序列 47
N748	CAAGCAGAAGACGGCATAACGAGATCTAGACAAGTCTCGTGGGCTCGG	序列 48
N749	CAAGCAGAAGACGGCATAACGAGATTGTGCAACGTCTCGTGGGCTCGG	序列 49
N750	CAAGCAGAAGACGGCATAACGAGATCGATAGATGTCTCGTGGGCTCGG	序列 50
N751	CAAGCAGAAGACGGCATAACGAGATAACAGTAAGTCTCGTGGGCTCGG	序列 51
N752	CAAGCAGAAGACGGCATAACGAGATCCGCGTGTGTCTCGTGGGCTCGG	序列 52
N753	CAAGCAGAAGACGGCATAACGAGATTCTGGATAGTCTCGTGGGCTCGG	序列 53
N754	CAAGCAGAAGACGGCATAACGAGATTATTCCTAGTCTCGTGGGCTCGG	序列 54
N755	CAAGCAGAAGACGGCATAACGAGATTACGTTCTCGTGGGCTCGG	序列 55
N756	CAAGCAGAAGACGGCATAACGAGATAACGCAATGTCTCGTGGGCTCGG	序列 56
N757	CAAGCAGAAGACGGCATAACGAGATGCTTACGAGTCTCGTGGGCTCGG	序列 57
N758	CAAGCAGAAGACGGCATAACGAGATTACTTCGCGTCTCGTGGGCTCGG	序列 58
N759	CAAGCAGAAGACGGCATAACGAGATCGCAGTCCGTCTCGTGGGCTCGG	序列 59
N760	CAAGCAGAAGACGGCATAACGAGATCAATGCTCGTCTCGTGGGCTCGG	序列 60
N761	CAAGCAGAAGACGGCATAACGAGATCACGGCGAGTCTCGTGGGCTCGG	序列 61
N762	CAAGCAGAAGACGGCATAACGAGATCGCCGTGGTCTCGTGGGCTCGG	序列 62
N763	CAAGCAGAAGACGGCATAACGAGATGCATCCTTGTCTCGTGGGCTCGG	序列 63
N764	CAAGCAGAAGACGGCATAACGAGATGCCATTGCGTCTCGTGGGCTCGG	序列 64

N765	CAAGCAGAAGACGGC ATACGAGATGAGAATACGTCTCGTGGGCTCGG	序列 65
N766	CAAGCAGAAGACGGC ATACGAGATGTAATGACGTCTCGTGGGCTCGG	序列 66
N767	CAAGCAGAAGACGGC ATACGAGATGCTTGATGTCTCGTGGGCTCGG	序列 67
N768	CAAGCAGAAGACGGC ATACGAGATAGTATACCGTCTCGTGGGCTCGG	序列 68
N769	CAAGCAGAAGACGGC ATACGAGATGCACGCAAGTCTCGTGGGCTCGG	序列 69
N770	CAAGCAGAAGACGGC ATACGAGATCCGTGCGAGTCTCGTGGGCTCGG	序列 70
N771	CAAGCAGAAGACGGC ATACGAGATATGCCTGCGTCTCGTGGGCTCGG	序列 71
N772	CAAGCAGAAGACGGC ATACGAGATTCGCTGGCGTCTCGTGGGCTCGG	序列 72

上表中，下划线为测序接头 A、斜体为 8 个碱基组成随机片段，加粗部分为与所述接头 A 互补的单链 DNA 分子。

2、小心移除容器表面的封膜，随后使用

WaferGenMultisampleNanodispenser 仪器，将上述 1 得到的 384 孔板中的样品，

5 使用“LFR 35nl 72 Sample dispensing.seq”程序等量分装到上述三得到的含有片段化产物 5184 孔板 72 列的每个孔中。

3、加液完成后，将该容器使用封口膜封好，将容器放入 Eppendorf 5810 孔板离心机，3220 ×g，离心 5 分钟。室温静置 10 分钟，备用。

4、依据表 6 配置好聚合酶链式反应缓冲液，表 6 中 Primer1 和 2 的序列见表 7。之后震荡混匀，短暂离心后将该反应液分装到 384 孔板中，具体位置如图 8 绿色区域所示所示，每孔体积 15.6 微升；使用“LFR 35nl Single Sample dispensing.seq”程序等量分装到经上述 3 处理得到的 5184 孔板各孔中。

表 6 聚合酶链式反应缓冲液

试剂	体积（微升）
5 x TruePrep Amplify Buffer	865.83
TruePrep Amplify Enzyme(1 U/ μ l)	93.4185
dNTP(10mM)	93.4185
Primer1 100 μ M	15.9495
Primer2 100 μ M	15.9495
ddH2O	22.785

表 7 Primer1 和 Primer2 引物序列

名称	序列 5'-3'
Primer1	AATGATACGGCGACCACCGA
Primer2	CAAGCAGAAGACGGC ATACGA

15 引物 1 与测序接头 A 相同或互补；

引物 2 与测序接头 B 互补或相同。

5、取出已经分装在 96 孔板内的 72 种带有标签序列的 5' 端引物（2.5 μ mol/L，序列如表 8 所示），使用八道移液器取 0.56 μ L，加入到准备好

的 384 孔板中。将 384 孔板置于混匀仪上震荡混匀，随后将孔板放入 Eppendorf 5810 孔板离心机， $3220 \times g$ ，离心 5 分钟。检查无气泡后室温静置备用。

5' 端标签引物从 5' 至 3' 方向依次由测序接头 B、8 个碱基组成随机片段和与接头 B 互补的单链 DNA 分子组成；

5 72 条 3' 端标签引物和 72 条 5' 端标签引物的 8 个碱基组成随机片段均不相同；

表 8 5' 端标签序列引物

名称	序列 5' -3'	序列编号
N501	<u>AATGATACGGCGACCACCGAGATCTACAC</u> <i>TAGATCGCTCGTCGGCAGCGTC</i>	序列 73
N502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTC	序列 74
N503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC	序列 75
N504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTC	序列 76
N505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGTCGTCGGCAGCGTC	序列 77
N506	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC	序列 78
N507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTC	序列 79
N508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTC	序列 80
N509	AATGATACGGCGACCACCGAGATCTACACCTAGCGCTTCGTCGGCAGCGTC	序列 81
N510	AATGATACGGCGACCACCGAGATCTACACCTCACAGGTCGTCGGCAGCGTC	序列 82
N511	AATGATACGGCGACCACCGAGATCTACACCTTAGTTGTCGTCGGCAGCGTC	序列 83
N512	AATGATACGGCGACCACCGAGATCTACACCTTCTATTCGTCGGCAGCGTC	序列 84
N513	AATGATACGGCGACCACCGAGATCTACACCTTGTAGTTCGTCGGCAGCGTC	序列 85
N514	AATGATACGGCGACCACCGAGATCTACACGAACCATCTCGTCGGCAGCGTC	序列 86
N515	AATGATACGGCGACCACCGAGATCTACACGAATGTGGTCGTCGGCAGCGTC	序列 87
N516	AATGATACGGCGACCACCGAGATCTACACGACCAAGATCGTCGGCAGCGTC	序列 88
N517	AATGATACGGCGACCACCGAGATCTACACGATCCTCGTCGTCGGCAGCGTC	序列 89
N518	AATGATACGGCGACCACCGAGATCTACACGATGGACTTCGTCGGCAGCGTC	序列 90
N519	AATGATACGGCGACCACCGAGATCTACACGATTAGTGTCGTCGGCAGCGTC	序列 91
N520	AATGATACGGCGACCACCGAGATCTACACGCGCCTTATCGTCGGCAGCGTC	序列 92
N521	AATGATACGGCGACCACCGAGATCTACACGGAACAGTTCGTCGGCAGCGTC	序列 93
N522	AATGATACGGCGACCACCGAGATCTACACGGAGGAAGTCGTCGGCAGCGTC	序列 94
N523	AATGATACGGCGACCACCGAGATCTACACGGAGTCGCTCGTCGGCAGCGTC	序列 95
N524	AATGATACGGCGACCACCGAGATCTACACGGCCTGTATCGTCGGCAGCGTC	序列 96
N525	AATGATACGGCGACCACCGAGATCTACACGGCTTAACTCGTCGGCAGCGTC	序列 97
N526	AATGATACGGCGACCACCGAGATCTACACGGTAATTATCGTCGGCAGCGTC	序列 98
N527	AATGATACGGCGACCACCGAGATCTACACGGTGTTATTCGTCGGCAGCGTC	序列 99
N528	AATGATACGGCGACCACCGAGATCTACACGTCCTACGTCGTCGGCAGCGTC	序列 100

N529	AATGATACGGCGACCACCGAGATCTACACGTCGAGAGTCGTCGGCAGCGTC	序列 101
N530	AATGATACGGCGACCACCGAGATCTACACGTCGTAGTCGTCGGCAGCGTC	序列 102
N531	AATGATACGGCGACCACCGAGATCTACACGTAAACCTTCGTCGGCAGCGTC	序列 103
N532	AATGATACGGCGACCACCGAGATCTACACGTTGCAACTCGTCGGCAGCGTC	序列 104
N533	AATGATACGGCGACCACCGAGATCTACACTAATTGAGTCGTCGGCAGCGTC	序列 105
N534	AATGATACGGCGACCACCGAGATCTACACTAGACTTGTCGTCGGCAGCGTC	序列 106
N535	AATGATACGGCGACCACCGAGATCTACACTAGGTTGTTTCGTCGGCAGCGTC	序列 107
N536	AATGATACGGCGACCACCGAGATCTACACTATGGTAGTCGTCGGCAGCGTC	序列 108
N537	AATGATACGGCGACCACCGAGATCTACACTATGTGTCTCGTCGGCAGCGTC	序列 109
N538	AATGATACGGCGACCACCGAGATCTACACTATTATCTTCGTCGGCAGCGTC	序列 110
N539	AATGATACGGCGACCACCGAGATCTACACTCACC GCGTCGTCGGCAGCGTC	序列 111
N540	AATGATACGGCGACCACCGAGATCTACACTCATAGTATCGTCGGCAGCGTC	序列 112
N541	AATGATACGGCGACCACCGAGATCTACACTCCAACAATCGTCGGCAGCGTC	序列 113
N542	AATGATACGGCGACCACCGAGATCTACACTCCTCACTTCGTCGGCAGCGTC	序列 114
N543	AATGATACGGCGACCACCGAGATCTACACTCGGCGATTCGTCGGCAGCGTC	序列 115
N544	AATGATACGGCGACCACCGAGATCTACACTCTATAAGTCGTCGGCAGCGTC	序列 116
N545	AATGATACGGCGACCACCGAGATCTACACTCTCATGGTCGTCGGCAGCGTC	序列 117
N546	AATGATACGGCGACCACCGAGATCTACACTGAGGTGATCGTCGGCAGCGTC	序列 118
N547	AATGATACGGCGACCACCGAGATCTACACTGCAAGGTTTCGTCGGCAGCGTC	序列 119
N548	AATGATACGGCGACCACCGAGATCTACACTGGAGTATTCGTCGGCAGCGTC	序列 120
N549	AATGATACGGCGACCACCGAGATCTACACGCGAGGCCTCGTCGGCAGCGTC	序列 121
N550	AATGATACGGCGACCACCGAGATCTACACTGCGCGCCTCGTCGGCAGCGTC	序列 122
N551	AATGATACGGCGACCACCGAGATCTACACAGGTGGCGTCGTCGGCAGCGTC	序列 123
N552	AATGATACGGCGACCACCGAGATCTACACGCCGCATGTCGTCGGCAGCGTC	序列 124
N553	AATGATACGGCGACCACCGAGATCTACACCTGTTGCCTCGTCGGCAGCGTC	序列 125
N554	AATGATACGGCGACCACCGAGATCTACACTGATACCGTCGTCGGCAGCGTC	序列 126
N555	AATGATACGGCGACCACCGAGATCTACACATTGGCCGTCGTCGGCAGCGTC	序列 127
N556	AATGATACGGCGACCACCGAGATCTACACGGACGGCTTCGTCGGCAGCGTC	序列 128
N557	AATGATACGGCGACCACCGAGATCTACACCACTCTGTTTCGTCGGCAGCGTC	序列 129
N558	AATGATACGGCGACCACCGAGATCTACACGGCTGCGTTTCGTCGGCAGCGTC	序列 130
N559	AATGATACGGCGACCACCGAGATCTACACGTCAGCTCTCGTCGGCAGCGTC	序列 131
N560	AATGATACGGCGACCACCGAGATCTACACAGCCATCATCGTCGGCAGCGTC	序列 132
N561	AATGATACGGCGACCACCGAGATCTACACATGATTCATCGTCGGCAGCGTC	序列 133
N562	AATGATACGGCGACCACCGAGATCTACACGTCTGTTCATCGTCGGCAGCGTC	序列 134
N563	AATGATACGGCGACCACCGAGATCTACACACGACCACTCGTCGGCAGCGTC	序列 135
N564	AATGATACGGCGACCACCGAGATCTACACCTCCACGCTCGTCGGCAGCGTC	序列 136

N565	AATGATACGGCGACCACCGAGATCTACACGCGGAAGTTCGTCGGCAGCGTC	序列 137
N566	AATGATACGGCGACCACCGAGATCTACACGTACATGTTCGTCGGCAGCGTC	序列 138
N567	AATGATACGGCGACCACCGAGATCTACACTTAGCCGGTCGTCGGCAGCGTC	序列 139
N568	AATGATACGGCGACCACCGAGATCTACACCAGGATCGTCGTCGGCAGCGTC	序列 140
N569	AATGATACGGCGACCACCGAGATCTACACATATCGTCTCGTCGGCAGCGTC	序列 141
N570	AATGATACGGCGACCACCGAGATCTACACTGGCCAGGTCGTCGGCAGCGTC	序列 142
N571	AATGATACGGCGACCACCGAGATCTACACTAGAGAGCTCGTCGGCAGCGTC	序列 143
N572	AATGATACGGCGACCACCGAGATCTACACGACACGCTTCGTCGGCAGCGTC	序列 144

上表中，下划线为测序接头 B、斜体为 8 个碱基组成随机片段，加粗部分为与接头 B 互补的单链 DNA 分子。

6、小心移除容器表面的封膜，随后使用

WaferGenMultisampleNanodispenser 仪器，将步骤 5 中准备好的 384 孔板中

5 的样品，使用“LFR 50nl 72 Sample dispensing.seq”程序等量分装到经上述 4 处理得到的 5184 孔板的 72 行的每个孔中。

7、将经上述 6 处理得到的 5184 孔板按照表 9 所示的 PCR 程序进行聚合酶链式扩增。

表 9 聚合酶链式扩增反应程序

热盖设置为 105℃		
步骤	孵育温度	孵育时间
1	72℃	3 分钟
2	98℃	30 秒
3	98℃	10 秒
4	60℃	30 秒
5	72℃	3 分钟
6	返回步骤 2，重复 14 次	
7	72℃	5 分钟
8	4℃	保持

10

8、将经上述 7 处理后的 5184 孔板中所有孔的扩增产物混合，得到测序文库（原始）。

将测序文库（原始）用 2%琼脂糖电泳，回收电泳产物，实现测序文库（原始）分选，得到 250-550 bp 片段选择文库 1 和 550-1000 bp 片段选择文库 2。

15 将测序文库（原始）、片段选择文库 1、片段选择文库 2 使用 Agilent Bioanalyzer 进行检测，测定其文库片段分布情况。

图 9 为原始文库结果，可以看出，片段为 200-1500bp；

图 10 为片段选择文库 1 结果，可以看出，片段为 250-550bp；

图 11 为片段选择文库 2 结果，可以看出，片段为 550-1000bp。

用实时荧光定量 PCR 检测文库中的有效分子浓度标准 qPCR 浓度（可以使用 Kapa Biosystems 公司的 KK4824 KAPA Library Quantification Kits for illumina platform 进行 qPCR 检测。该试剂盒包含该检测的所有试剂及引物。qPCR 程序如下表 10）），结果如表 11。

5

表 10 qPCR 使用程序

步骤	温度（目的）	时间	循环数
1	95 度（变性 DNA，热启动）	5 分钟	不循环
2	95 度（变性 DNA）	30 秒	35 个循环
3	60 度（引物结合，延伸，收集荧光信号）	45 秒	

表 11 文库中的有效分子浓度

编号	标准 qPCR 浓度	
图 10 文库	37.61	nmol/L
图 11 文库	17.20	nmol/L

检测合格（合格标准为 qPCR 浓度大于 5nmol/L）后进行 illuminaHiseq 2000 或者 Hiseq 2500 测序。

10

用 SOAP2 软件进行拼接测序结果，结果为拼接得到 100KB 片段，说明本发明方法正确，可以测序大小为 100KB 的长片段。

拼接分析如下：

方法：（1）使用 soap2 把 reads 比对 Ref hg19，比对时只取 unique 的比对结果（-r 0）。

15

（2）使用 soap.coverage 计算 reads 在 reference 上单点覆盖度（物理覆盖度 -phy）。

（3）计算每个 well 上覆盖的区块(block)的长度和 block 之间的间隔(gap)的长度。注：这里 gap 容忍度为 500K，即长度小于 500K 的不认定为 gap。把三种插入长度的文库合并起来计算。

20

结果：

长片段长度分布直方图如图 12 所示。

中位数： 45,510 bp

平均值： 195,457 bp

标准差： 551,263 bp

25

最大值：147,859,018 bp。

实施例2、长片段测序文库用于分离纯化单菌

一、长片段基因组DNA的提取

提取混合微生物基因组 DNA，且使提取的基因组 DNA 片段不小于 100kb（100-800kb），同时变性基因组 DNA 片段；

30

具体提取方法如下：

1、分离菌落

1) 取 0.2 克含复杂微生物群落的鸡粪置于 1.5 毫升离心管中，用 1000 微升磷酸盐缓冲液进行悬浮，涡旋振荡 5 分钟至充分混匀。

2) 将上述样品在高速离心机上以 10000 转每分钟的转速离心 5 分钟，丢弃上清液后加入 800 微升磷酸盐缓冲液，涡旋振荡 3 分钟至充分混匀。

3) 将上述样品在离心机上以 2000 转每分钟的转速离心 3 分钟，然后小心吸取上清到一个新的 1.5 毫升离心管中，然后重复本步骤 2 次；

4) 将上述样品在高速离心机上以 10000 转每分钟的转速离心 5 分钟，丢弃上清液后加入 800 微升磷酸盐缓冲液，涡旋振荡 3 分钟至充分混匀，然后重复本步骤 2 次。

5) 使用孔径 20 微米的醋酸纤维滤膜对前述样品进行过滤，收集滤液到一个新的 1.5 毫升离心管中，然后在高速离心机上以 10000 转每分钟的转速离心 5 分钟，丢弃上清；得到微生物混合物；

2、细胞固定

在上述 1 得到的微生物混合物中加入 800 微升用无菌水配制的 4%多聚甲醛溶液，涡旋振荡 3 分钟至充分混匀后置于 4 摄氏度静置过夜，得到固定微生物细胞。

将上述固定微生物细胞短暂漩涡振荡至混匀，用红细胞计数板对其中细胞浓度进行计数，然后用无菌水稀释至 30 细胞/微升后备用，得到稀释后微生物细胞。

3、裂解微生物得到长片段核酸分子

1) 将 20 微升碱变性试剂原液（见表 1 内容）稀释 30 倍至 600 微升，并分装至 384 孔板中，具体位置如图 7 中蓝色标记所示，每孔加入体积 22 微升。

2) 使用 WaferGenMultisampleNanodispenser 仪器，将步骤 1) 中准备好的 384 孔板中的样品，使用“LFR 35nl Single Sample dispensing.seq”程序加入到定制化的 5184 孔板中。每孔容积为 350 纳升。

3) 加液完成后，将容器使用封口膜封好，将容器放入 Eppendorf 5810 孔板离心机用 4000 转每分钟的转速离心 5 分钟，然后小心移去封口膜，金属浴加热容器至 85 摄氏度保持 15 分钟，使容器中的水全部蒸发，然后取出容器待其自然冷却至室温；

4) 将上述 2 得到的稀释后微生物细胞分装至 384 孔板的 24 孔中，具体位置如图 7 中蓝色标记所示，每孔加入体积 22 微升。

5) 使用 WaferGenMultisampleNanodispenser 仪器，将步骤 4) 中准备好的 384 孔板中的样品，使用“LFR 35nl Single Sample dispensing.seq”程序加入到步骤 3) 所述的容器中。

6) 加液完成后，将容器使用封口膜封好，将容器放入 Eppendorf 5810 孔板离心机用 4000 转每分钟的转速离心 5 分钟。随后将其放入适配的温浴设备中，

85 摄氏度孵育 2 分钟，将容器放入 Eppendorf 5810 孔板离心机用 4000 转每分钟的转速离心 2 分钟，再将容器放入 85 摄氏度孵育 2 分钟。

7) 将上述容器放入 Eppendorf 5810 孔板离心机用 4000 转每分钟的转速离心 5 分钟，置于室温备用，得到含有单链 DNA 分子样本，含有其的孔板为含有

5 单链 DNA 分子样本的 5184 孔板。

二、全基因组扩增反应

与实施例 1 的二方法相同。

三、片段化

与实施例 1 的三方法相同。

10 四、加标签序列，区分不同样本

与实施例 1 的六方法相同，得到测序文库（原始）。

分选得到 500bp-600bp 片段选择文库。

用实时荧光定量 PCR 检测文库中的有效分子浓度。检测合格后进行 illuminaHiseq 2000 或者 Hiseq 2500 测序。

15 结果：得到已知粪便中的一个细菌基因组的组装。

工业应用

本方法通过采用一种新型的高通量微量移液平台进行前期建库实验。该平台可以进行 35 纳升的移液操作，从而将多重链置换反应的反应体积控制在 100 纳升以下，有证据显示降低反应体积可以显著改善多重链置换反应的扩增偏好性。同时本方法将物理分隔数目提升到了 5184 孔，在建库 DNA 起始量保持不变的情况下，该容器每个物理分隔中所包含的 DNA 仅为原有 384 分隔的产品的十分之一左右，即 1%左右的基因组，这无疑降低了每个分隔中 denovo 组装的难度。同时，物理分隔数目的增加可以降低来自父母双方相同基因组位置的同源染色体片段在相同孔内出现的概率至 0.5%-1%，这对于定相长度的提升有非常大的帮助，从而最终可以得到更好的单倍体型组装效果。（图 1，长片段测序技术原理示意图。）

25 本方法通过将多重链置换扩增技术，转座酶片段化技术与纳升级微量移液平台进行优化整合，建立了一套可用于低样品起始量的单倍体型文库构建流程。（图 2 为长片段测序文库构建方法流程图。）

30 本方法操作相对简单，反应步骤较少，文库构建时间只需 10 小时，其中手工操作部分仅为 3.5 小时。作为对比，传统的 illumina 测序文库构建需要 2 天时间，手工操作部分需要 5 小时左右。而 Fosmid 构建单倍体型文库技术需要 8 天时间才能完成文库构建，而 illumina 商业化的 Truseq synthetic long reads 试剂盒需要三天时间完成建库，其中手工操作时间需 6-8 小时。对比以上两种建库方式，本方法在报告周期方面具有一定优势。（图 3 为本方法建库时间与 illumina 常规文库建库所需时间对比。图 4 为本方法建库时间与主流单倍型文库建库所需时间对比。）

除以上特点之外，本方法仅需一台纳升级微量移液平台，无需其他仪器或者自动化移液装置的辅助即可完成全部实验流程。相比之下，染色体中期分离进行单倍体型构建技术需要的显微操作装置，Complete Genomics 公司的 LFR 技术需要三台大型自动化移液装置，本方法无疑具有更好的可行性。

5 应用方面，本方法对 DNA 样品的需求量较少，最低可以低至 150 皮克（等效于 25 个人类细胞中所含有的 DNA 量）。传统的 Fosmid 构建单倍体型文库技术需要 8 微克 DNA 进行建库，illuminaTruseq synthetic long reads 试剂盒则需要 500ng DNA 作为起始。起始量的降低使得本方法的应用场景得到了极大拓展，可以用于游离肿瘤细胞，游离胎儿细胞或胚胎细胞的单倍型测序。

10 本方法采用了较温和的提取方法，可以分离出较长的 DNA 片段，不小于 100kb。而长片段的优势在于可以更好的解决基因组中重复区域的组装，同时也可以提升在杂合区域的组装表现。因此可以将本方法应用于动物，植物以及微生物的基因组 de novo 组装文库的构建。

除以上两种应用场景之外，当以一定浓度的细胞悬液作为起始材料时，本方法可以实现单细胞分离的效果。再辅以一定的细胞观测以及裂解手段，本方法可以实现高通量单细胞测序文库构建，每次运行理论上可以平行构建约 1000 个单细胞测序文库。因此可以将该方法用于复杂环境中的菌群研究，进而可以更深入的了解复杂环境中菌群组成。更重要的是，使用本方法可以获得一些无法分离培养的细菌的基因组信息，这无疑对于宏基因组学研究具有非常重大的意义。作为延伸，还可以将本方法应用于单细胞组学研究当中，可作为前期高通量，自动化单细胞扩增文库的构建方法。

无论是 Fosmid 方法，Truseq synthetic long reads 试剂盒，还是 Complete Genomics 公司的 LFR 技术都是基于分隔稀释法的原理进行单倍型组装，即通过物理分隔的方法降低同源染色体片段出现在同一分隔中的概率，从而对杂合 SNP 进行定相（Phasing），其分隔数一般不超过 384。可以通过同一样品进行重复实验的方法增加分隔数，但这意味着建库和测序成本成倍增加。

25 总之，本发明的方法是通过使用纳升级微量移液装置以及定制的 5184 孔容器，可以有效的提升物理分隔数，进一步降低同源染色体片段出现在同一分隔中的概率，从而实现更好的单倍型组装效果。该定制容器相较原有容器的单位孔容积更大，可以容纳 200nl-350nl 体积的试剂（原有容积为 100nl）。该定制化容器可以容纳 5 次加样，因此可以在该定制化 5184 孔容器中完成全部建库反应。

也应当注意到，操作人员可以将该 5184 孔容器进行再次分隔处理，即可以在一个容器中同时处理多个样品，如 2×2597 ，或 3×1728 等模式，从而达到增加通量，降低成本的目的。

35 另外还应当注意到，本方法还可以通过新增加一个维度标签的方式，引入虚拟分隔，从而进一步提升分隔数目。如在 DNA 转座酶片段化过程中引入两个组

合标签，标记第一重物理分隔的位置，之后在聚合酶链式扩增过程中，再引入两个组合标签，标记第二重虚拟分隔的位置，从而最终可以实现 5184×5184 （包括但不限于此，可以是 12×5184 ， 24×5184 或者更多）物理分隔的效果。

权利要求

- 1、一种构建长片段测序文库的方法，包括如下步骤：
 - 1) 制备含有单链 DNA 的 5184 孔板；
- 5 所述制备含有单链 DNA 的 5184 孔板的方法为如下 A) 或 B)：
 - A) 先提取待测样品基因组 DNA，变性，得到单链 DNA，再将所述单链 DNA 分子分装到 5184 孔板的每个孔中，得到含有单链 DNA 的 5184 孔板；
 - B) 先将待测样品分装到 5184 孔板的每个孔中，再裂解变性，得到含有单链 DNA 的 5184 孔板；
- 10 所述单链 DNA 的片段长度不小于 100Kb；
 - 2) 将全基因组扩增反应体系分装到所述含有单链 DNA 的 5184 孔板的每个孔中，进行全基因组扩增反应，得到含有全基因组扩增产物的 5184 孔板；
 - 3) 将片段化反应体系分装到所述含有全基因组扩增产物的 5184 孔板的每个孔中，进行片段化反应，得到片段化产物，含有所述片段化产物的孔板即为
- 15 含有片段化产物的 5184 孔板；
 - 所述片段化产物的长度为 200-1500bp；
 - 4) 对所述含有片段化产物的 5184 孔板每个孔中片段化产物添加不同标签序列，即得到测序文库。
- 20 2、根据权利要求 1 所述的方法，其特征在于：
 - 所述含有单链 DNA 的 5184 孔板中的单链 DNA 总量满足 5184 孔板中每个孔内出现来自父母双方相同基因组位置的同源染色体片段的概率小于 1%；
 - 所述含有单链 DNA 的 5184 孔板中的单链 DNA 总量具体为 10-500 个细胞的 DNA 量。
- 25 3、根据权利要求 1 或 2 所述方法，其特征在于：所述 5184 孔板每孔的容积为 190 纳升至 350 纳升，所述 5184 孔板每孔的容积具体为 200 纳升至 350 纳升。
- 4、根据权利要求 1-3 中任一所述的方法，其特征在于：
 - 步骤 1) 中，所述变性或所述裂解变性采用的试剂为碱变性试剂；
 - 所述变性的反应条件为 25 度孵育 2 分钟；
- 30 所述裂解变性的反应条件为 85 摄氏度孵育 2 分钟；
 - 所述提取待测样品基因组 DNA 的方法采用透析法或碱裂解法或琼脂糖包埋法。
- 5、根据权利要求 1-4 中任一所述的方法，其特征在于：
 - 步骤 2) 中，所述全基因组扩增反应体系中含有由 8 个碱基组成的随机引物。
- 35 6、根据权利要求 1-5 中任一所述的方法，其特征在于：
 - 在步骤 1) 和步骤 2) 之间，还包括如下步骤：将由 8 个碱基组成的随机引物分装到所述含有单链 DNA 的 5184 孔板中。

- 7、根据权利要求 1-6 中任一所述的方法，其特征在于：
步骤 2) 中，所述全基因组扩增反应体系为多重链置换聚合酶扩增反应体系；
所述全基因组扩增反应为 30℃ 孵育 1 小时，再 65℃ 孵育 5 分钟；
或所述全基因组扩增反应为 37℃ 孵育 45 分钟，再 65℃ 孵育 5 分钟。
- 5 8、根据权利要求 1-7 中任一所述的方法，其特征在于：
步骤 3) 中，所述片段化反应体系为转座酶反应体系；
所述转座酶反应体系包括转座酶反应缓冲液和包埋接头的转座酶；
所述片段化反应的条件为 55℃ 孵育 10 分钟；
所述接头为接头 1 和/或接头 2。
- 10 9、根据权利要求 1-8 任一所述的方法，其特征在于：
步骤 4) 中，所述添加不同标签序列的方法包括如下步骤：将 72 种 3' 端标签引物、聚合酶链式反应体系和 72 种 5' 端标签引物分装到所述含有片段化产物的 5184 孔板各孔中，进行聚合酶链式反应，得到原始测序文库；
所述 3' 端标签引物从 5' 至 3' 方向依次由测序接头 A、8 个碱基组成随机
15 片段和与所述接头 1 互补的单链 DNA 分子组成；
所述 5' 端标签引物从 5' 至 3' 方向依次由测序接头 B、8 个碱基组成随机
片段和与所述接头 2 互补的单链 DNA 分子组成；
所述 72 条 3' 端标签引物和 72 条 5' 端标签引物的 8 个碱基组成随机片段
均不相同；
- 20 所述聚合酶链式反应体系含有引物 1 和引物 2；
所述引物 1 与所述测序接头 A 相同或互补；
所述引物 2 与所述测序接头 B 互补或相同。
- 10、根据权利要求 9 所述的方法，其特征在于：
所述 72 种 3' 端标签引物的分装方式为：将所述 72 种 3' 端标签引物分装
25 到所述含有片段化产物的 5184 孔板的 72 列的每个孔中；
所述 72 种 5' 端标签引物的分装方式为：将所述 72 种 5' 端标签引物分装
到所述含有片段化产物的 5184 孔板的 72 行的每个孔中；
所述聚合酶链式反应体系的分装方式为：将所述聚合酶链式反应体系分装
到所述含有片段化产物的 5184 孔板的每个孔中。
- 30 11、根据权利要求 9 或 10 所述的方法，其特征在于：
所述聚合酶链式反应的退火温度为 60℃ 30 秒。
- 12、根据权利要求 1-11 任一所述的方法，其特征在于：
所述分装的方法均为：先将各待分装物质分装到 384 孔板中的分装模式孔
中，再将所述分装模式孔中的物质用纳升级分液器分装到所述 5184 孔板各孔中；
35 所述分装的量均为各孔等量分装。
- 13、根据权利要求 12 所述的方法，其特征在于：
所述 384 孔板的分装模式孔符合 4×2 排列，且长边为 2 短边为 4，所述分

装模式孔的孔数范围 8-384 孔，所述分装模式孔具体为 24 孔或 72 孔。

14、根据权利要求 9-13 中任一所述的方法，其特征在于：

在步骤 4) 中得到原始测序文库后，还包括如下步骤：将所述原始测序文库进行文库片段分选，得到长度在 200-1100bp 范围内的片段选择文库，即为长片段测序文库。

15、根据权利要求 14 所述的方法，其特征在于：所述 200-1100bp 片段选择文库为 250-550 bp 片段选择文库和 550-1000 bp 片段选择文库；

所述分选的方法为磁珠纯化或琼脂糖电泳纯化。

16、根据权利要求 1-15 任一所述的方法，其特征在于：

10 所述待测样品为真核细胞或混合物微生物细胞，所述真核细胞来源于人类血液细胞；

所述混合物微生物细胞来源于含有混合微生物的粪便或含有混合微生物土壤。

17、由权利要求 1-16 中任一所述方法制备的长片段测序文库。

15 18、权利要求 1-16 中任一所述方法在长片段 DNA 分子测序中的应用；

所述长片段 DNA 分子不小于 100Kb；

或权利要求 1-16 中任一所述方法在混合微生物纯化单菌落中的应用；

所述混合微生物具体来源于粪便或土壤。

19、一种制备长片段测序文库的成套产品，包括 384 孔板、权利要求 1-16
20 中任一所述方法中的所述 5184 孔板和纳升级分液器。

20、一种权利要求 1-16 中任一所述方法中的定制的 5184 孔板，其特征在于：所述孔板每孔容积为 200-350 纳升。

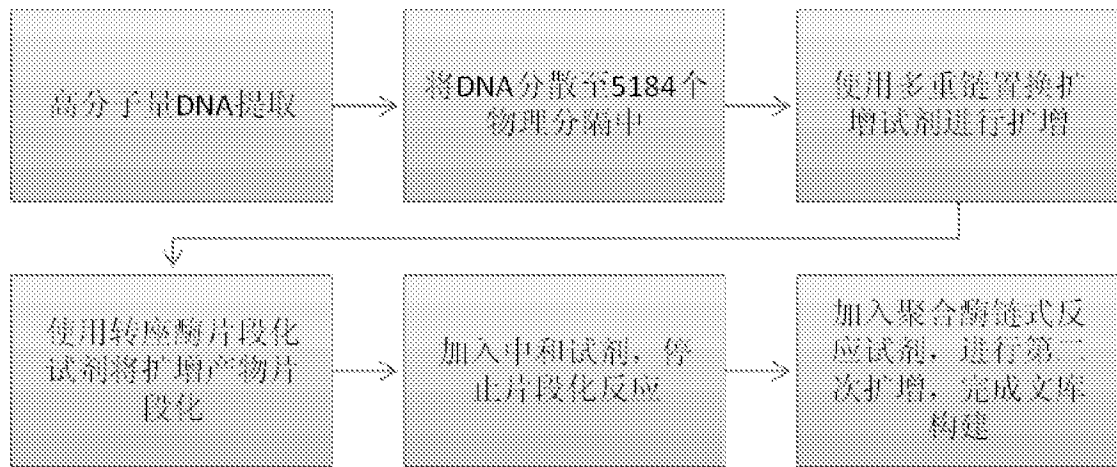


图 2

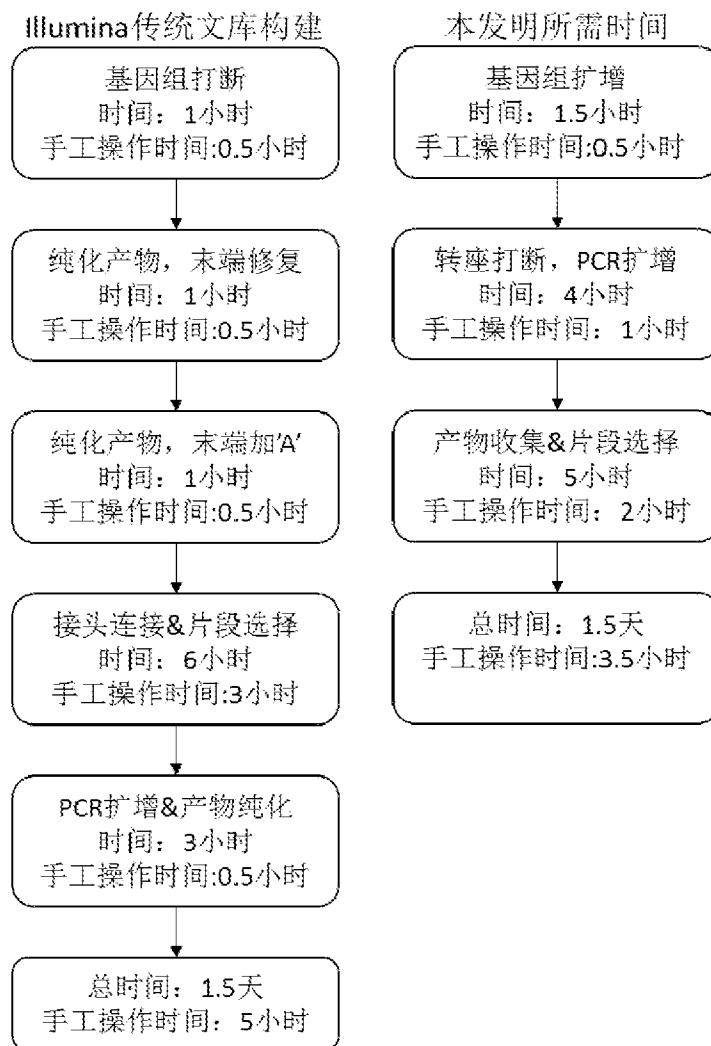


图 3

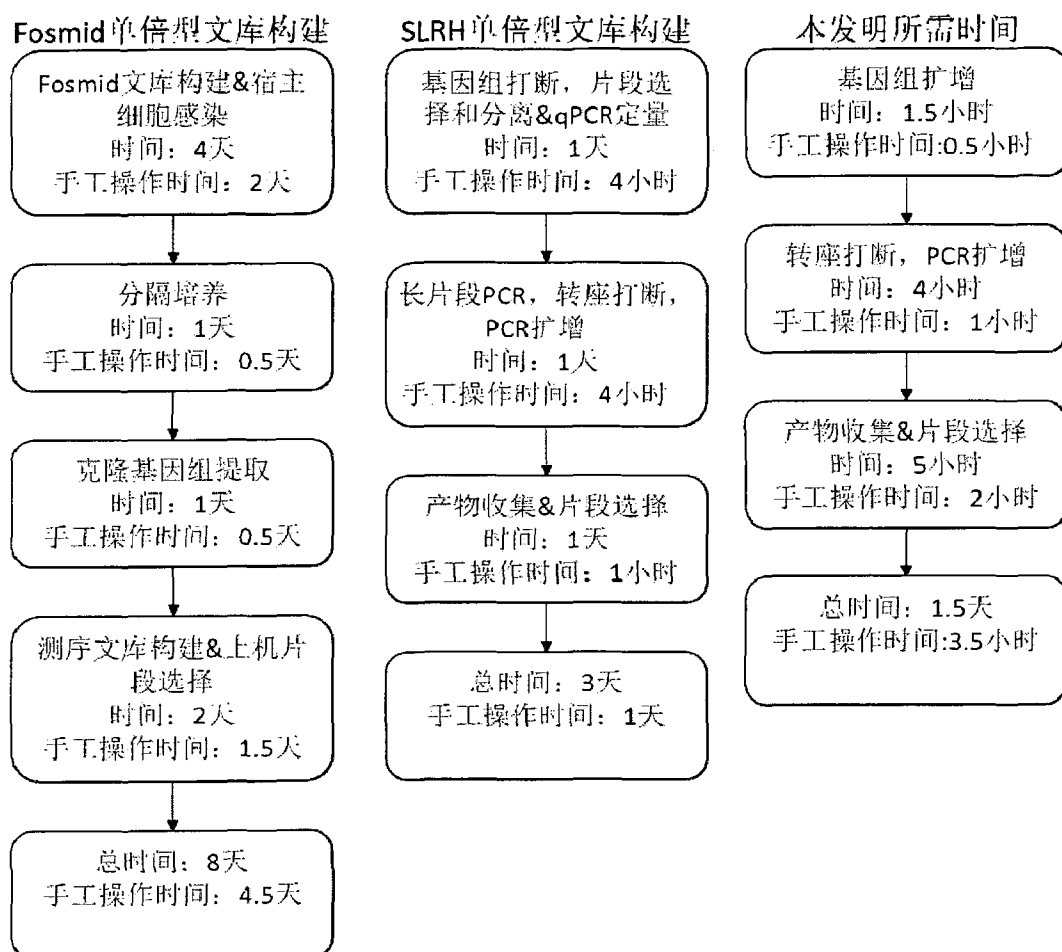


图 4

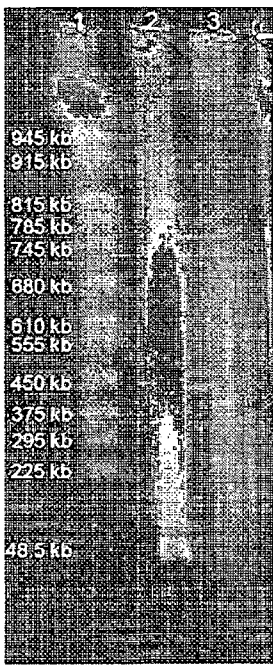


图 5

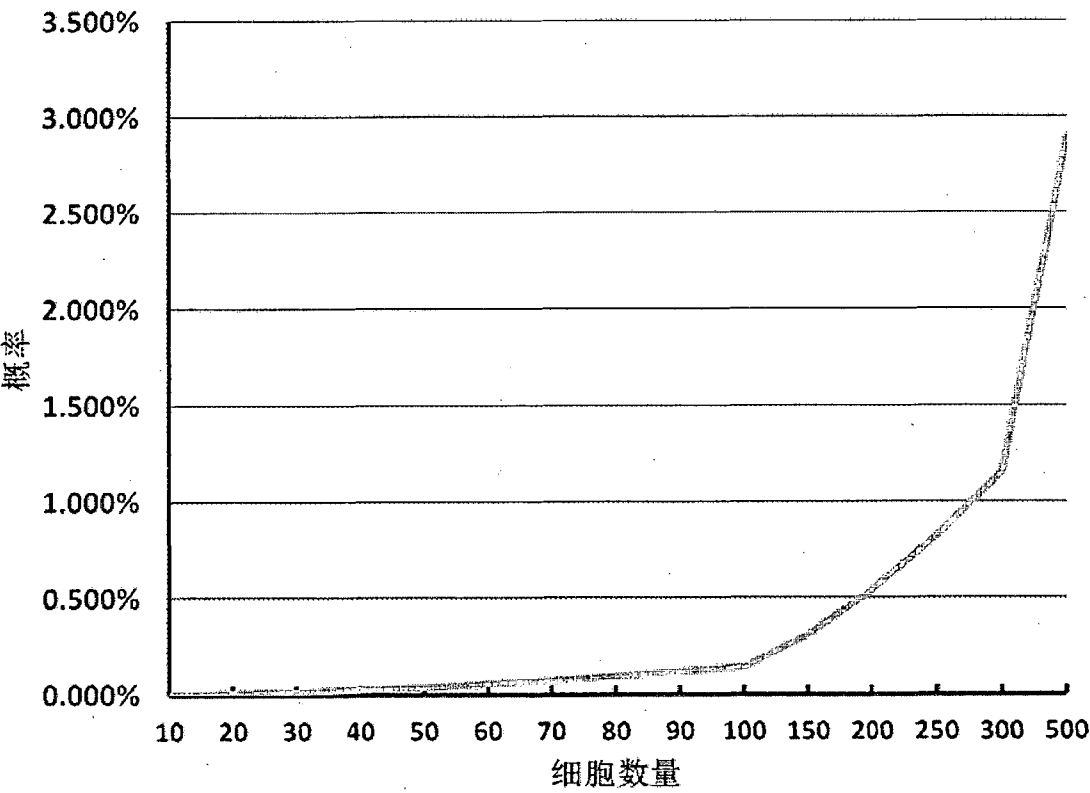


图 6

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	33	50	65	81	97	113	129	145	161	177	193	209	225	241	257	273	289	305	321	337	353	369	385	401
B	34	50	66	82	98	114	130	146	162	178	194	210	226	242	258	274	290	306	322	338	354	370	386	402
C	35	51	67	83	99	115	131	147	163	179	195	211	227	243	259	275	291	307	323	339	355	371	387	403
D	36	52	68	84	100	116	132	148	164	180	196	212	228	244	260	276	292	308	324	340	356	372	388	404
E	37	53	69	85	101	117	133	149	165	181	197	213	229	245	261	277	293	309	325	341	357	373	389	405
F	38	54	70	86	102	118	134	150	166	182	198	214	230	246	262	278	294	310	326	342	358	374	390	406
G	39	55	71	87	103	119	135	151	167	183	199	215	231	247	263	279	295	311	327	343	359	375	391	407
H	40	56	72	88	104	120	136	152	168	184	200	216	232	248	264	280	296	312	328	344	360	376	392	408
I	41	57	73	89	105	121	137	153	169	185	201	217	233	249	265	281	297	313	329	345	361	377	393	409
J	42	58	74	90	106	122	138	154	170	186	202	218	234	250	266	282	298	314	330	346	362	378	394	410
K	43	59	75	91	107	123	139	155	171	187	203	219	235	251	267	283	299	315	331	347	363	379	395	411
L	44	60	76	92	108	124	140	156	172	188	204	220	236	252	268	284	300	316	332	348	364	380	396	412
M	45	61	77	93	109	125	141	157	173	189	205	221	237	253	269	285	301	317	333	349	365	381	397	413
N	46	62	78	94	110	126	142	158	174	190	206	222	238	254	270	286	302	318	334	350	366	382	398	414
O	47	63	79	95	111	127	143	159	175	191	207	223	239	255	271	287	303	319	335	351	367	383	399	415
P	48	64	80	96	112	128	144	160	176	192	208	224	240	256	272	288	304	320	336	352	368	384	400	416

图 7

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	17	33	49	65	81	97	113	129	145	161	177	193	209	225	241	257	273	289	305	321	337	353	369	385
B	18	34	50	66	82	98	114	130	146	162	178	194	210	226	242	258	274	290	306	322	338	354	370	386
C	19	35	51	67	83	99	115	131	147	163	179	195	211	227	243	259	275	291	307	323	339	355	371	387
D	20	36	52	68	84	100	116	132	148	164	180	196	212	228	244	260	276	292	308	324	340	356	372	388
E	21	37	53	69	85	101	117	133	149	165	181	197	213	229	245	261	277	293	309	325	341	357	373	389
F	22	38	54	70	86	102	118	134	150	166	182	198	214	230	246	262	278	294	310	326	342	358	374	390
G	23	39	55	71	87	103	119	135	151	167	183	199	215	231	247	263	279	295	311	327	343	359	375	391
H	24	40	56	72	88	104	120	136	152	168	184	200	216	232	248	264	280	296	312	328	344	360	376	392
I	25	41	57	73	89	105	121	137	153	169	185	201	217	233	249	265	281	297	313	329	345	361	377	393
J	26	42	58	74	90	106	122	138	154	170	186	202	218	234	250	266	282	298	314	330	346	362	378	394
K	27	43	59	75	91	107	123	139	155	171	187	203	219	235	251	267	283	299	315	331	347	363	379	395
L	28	44	60	76	92	108	124	140	156	172	188	204	220	236	252	268	284	300	316	332	348	364	380	396
M	29	45	61	77	93	109	125	141	157	173	189	205	221	237	253	269	285	301	317	333	349	365	381	397
N	30	46	62	78	94	110	126	142	158	174	190	206	222	238	254	270	286	302	318	334	350	366	382	398
O	31	47	63	79	95	111	127	143	159	175	191	207	223	239	255	271	287	303	319	335	351	367	383	399
P	32	48	64	80	96	112	128	144	160	176	192	208	224	240	256	272	288	304	320	336	352	368	384	400

图 8

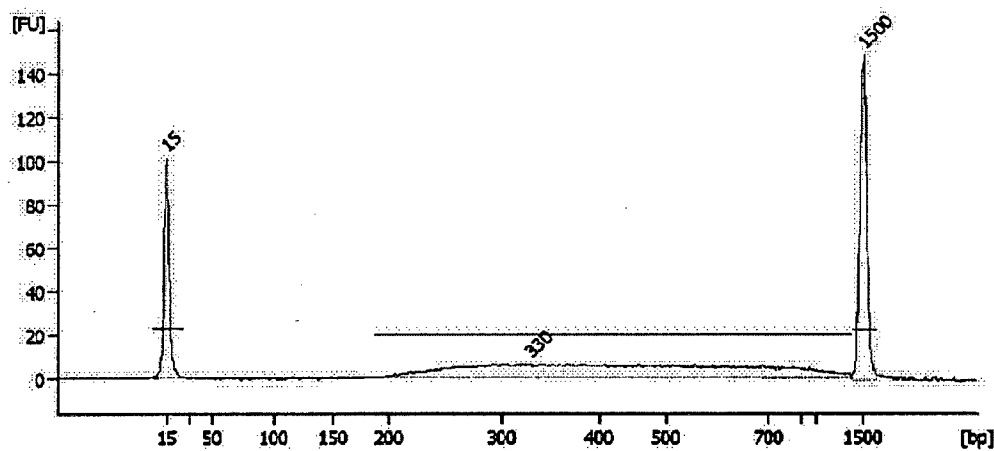


图 9

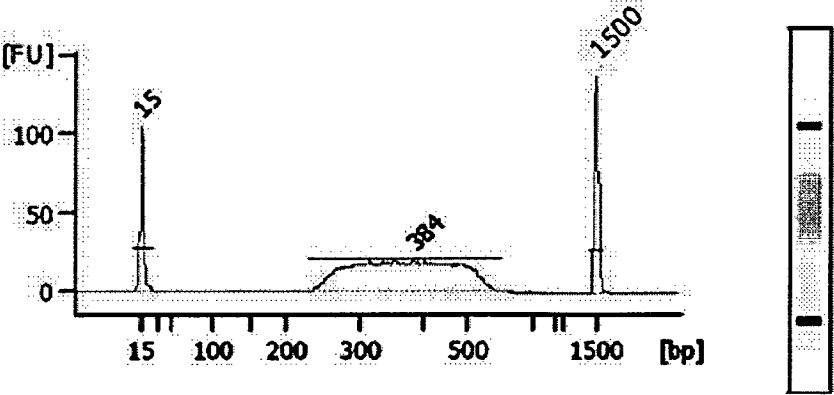


图 10

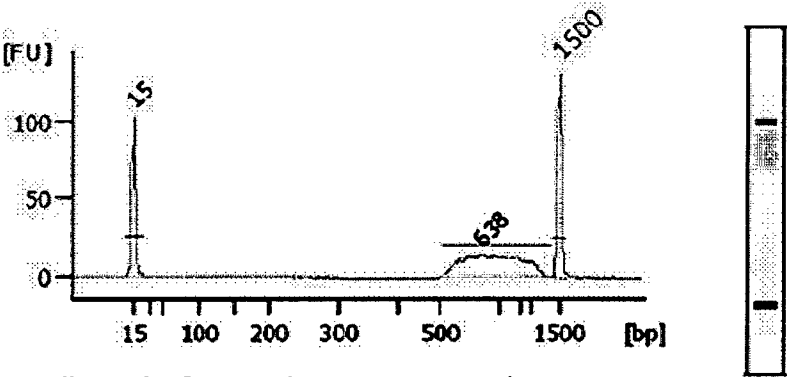


图 11

长片段长度分布柱形图

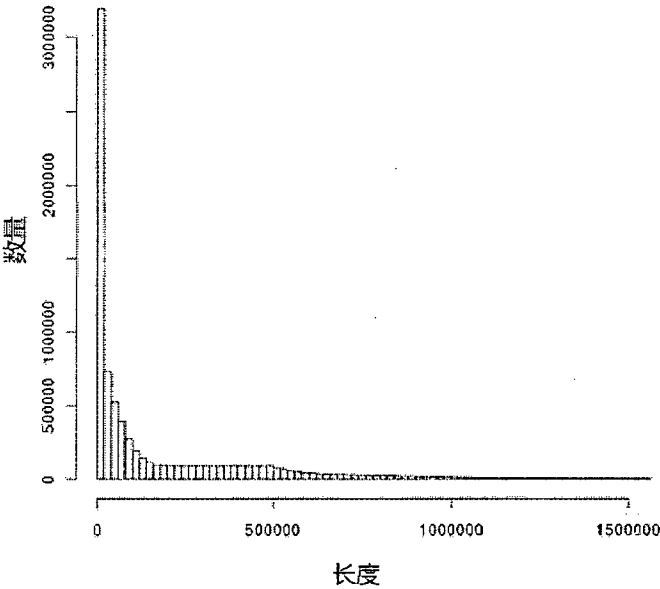


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/070789

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 filed or furnished:
 - a. (means)
 - ☐ on paper
 - ☒ in electronic form
 - b. (time)
 - ☒ in the international application as filed
 - ☐ together with the international application in electronic form
 - ☐ subsequently to this Authority for the purposes of search
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 in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/070789

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (2006.01) i; C12N 15/10 (2006.01) i; C12M 1/00 (2006.01) i; C40B 50/06 (2006.01) i; C40B 60/14 (2006.01) 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)

C12Q; C12N; C12M; C40B

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)

CNABS; DWPI; SIPOABS; CNKI; Pubmed; Google scholar: sequencing, long fragment, 5184 pore plate, separate, whole genome amplification, ploid, well plate?, long fragment sequencing library, libraries, LFR, genomics, amplif+, label, tag, adaptor, construct+, prepar+, search based on sequence SEQ ID NOs: 1-144

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 101932729 A (COMPLETE GENOMICS INCORPORATED), 29 December 2010 (29.12.2010), see the whole document	1-20
A	CN 102864498 A (TIANJIN INSTITUTE OF INDUSTRIAL BIOTECHNOLOGY), 09 January 2013 (09.01.2013), see the whole document	1-20
A	WO 2014145820 A2 (COMPLETE GENOMICS INC.), 18 September 2014 (18.09.2014), see the whole document	1-20

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

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 06 April 2016 (06.04.2016)	Date of mailing of the international search report 19 April 2016 (19.04.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer LI, Kangqi Telephone No.: (86-10) 62411034

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2016/070789

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101932729 A	29 December 2010	AU 2008335362 B2	25 July 2013
		HK 1187078 A1	22 January 2016
		CA 2707901 C	15 September 2015
		EP 2610351 A1	03 July 2013
		EP 2565279 A1	06 March 2013
		EP 2227563 A2	15 September 2010
		CA 2707901 A1	18 June 2009
		HK 1176095 A1	16 October 2015
		JP 2011520420 A	21 July 2011
		SG 193776 A1	30 October 2013
		EP 2610351 B1	08 July 2015
		JP 5362738 B2	11 December 2013
		AU 2008335362 A1	18 June 2009
		EP 2227563 B1	06 June 2012
		CN 103290106 A	11 September 2013
		EP 2565279 B1	03 December 2014
		CN 103290106 B	29 July 2015
		CN 101932729 B	27 March 2013
CN 102864498 A	09 January 2013	CN 102864498 B	16 July 2014
WO 2014145820 A2	18 September 2014	US 2016046985 A1	18 February 2016
		WO 2014145820 A3	27 November 2014
		CA 2902882 A1	18 September 2014
		US 2014323316 A1	30 October 2014
		AU 2014232515 A1	03 September 2015
		CN 105189308 A	23 December 2015
		EP 2969847 A2	20 January 2016

国际检索报告

国际申请号

PCT/CN2016/070789

A. 主题的分类

C12Q 1/68(2006.01)i; C12N 15/10(2006.01)i; C12M 1/00(2006.01)i; C40B 50/06(2006.01)i; C40B 60/14(2006.01)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)

C12Q; C12N; C12M; C40B

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))

CNABS;DWPI;SIPOABS;CNKI;Pubmed;Google scholar: 测序, 文库, 构建, 长片段, 长片断, 5184孔板, 分隔, 接头, 标签, 全基因组扩增, 倍体, well plate?, long fragment sequencing library, libraries, LFR, genomics, amplif+, label, tag, adaptor, construct+, prepar+, 基于序列SEQ ID NOs:1-144的检索

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
A	CN 101932729 A (考利达基因组股份有限公司) 2010年 12月 29日 (2010 - 12 - 29) 参见全文	1-20
A	CN 102864498 A (天津工业生物技术研究所) 2013年 1月 9日 (2013 - 01 - 09) 参见全文	1-20
A	WO 2014145820 A2 (COMPLETE GENOMICS INC) 2014年 9月 18日 (2014 - 09 - 18) 参见全文	1-20

☐ 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

“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

2016年 4月 6日

国际检索报告邮寄日期

2016年 4月 19日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员

李康琦

传真号 (86-10)62019451

电话号码 (86-10)62411034

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对要求保护的发明必要的核苷酸和/或氨基酸序列，国际检索是在下列基础上进行的：

a. (提交提供)

☐ 纸件形式

☒ 电子形式

b. (提交时间)

☒ 含在申请提交时的国际申请中

☐ 以电子形式与国际申请一起提交

☐ 为检索之用随后提交本单位

2. ☐ 另外，在提交/提供了多个版本或副本的序列表的情况下，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申请中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/CN2016/070789

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	101932729	A	2010年 12月 29日	AU	2008335362	B2	2013年 7月 25日
				HK	1187078	A1	2016年 1月 22日
				CA	2707901	C	2015年 9月 15日
				EP	2610351	A1	2013年 7月 3日
				EP	2565279	A1	2013年 3月 6日
				EP	2227563	A2	2010年 9月 15日
				CA	2707901	A1	2009年 6月 18日
				HK	1176095	A1	2015年 10月 16日
				JP	2011520420	A	2011年 7月 21日
				SG	193776	A1	2013年 10月 30日
				EP	2610351	B1	2015年 7月 8日
				JP	5362738	B2	2013年 12月 11日
				AU	2008335362	A1	2009年 6月 18日
				EP	2227563	B1	2012年 6月 6日
				CN	103290106	A	2013年 9月 11日
				EP	2565279	B1	2014年 12月 3日
				CN	103290106	B	2015年 7月 29日
				CN	101932729	B	2013年 3月 27日
CN	102864498	A	2013年 1月 9日	CN	102864498	B	2014年 7月 16日
WO	2014145820	A2	2014年 9月 18日	US	2016046985	A1	2016年 2月 18日
				WO	2014145820	A3	2014年 11月 27日
				CA	2902882	A1	2014年 9月 18日
				US	2014323316	A1	2014年 10月 30日
				AU	2014232515	A1	2015年 9月 3日
				CN	105189308	A	2015年 12月 23日
				EP	2969847	A2	2016年 1月 20日

表 PCT/ISA/210 (同族专利附件) (2009年7月)