

(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 10 月 27 日 (27.10.2016)



(10) 国际公布号
WO 2016/169431 A1

- (51) 国际专利分类号:
C40B 50/06 (2006.01) C12N 15/10 (2006.01)
C40B 40/06 (2006.01)
- (21) 国际申请号: PCT/CN2016/079278
- (22) 国际申请日: 2016 年 4 月 14 日 (14.04.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510187820.0 2015 年 4 月 20 日 (20.04.2015) 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)。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)。林霖 (LIN, Lin); 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 DNA LIBRARY

(54) 发明名称: 一种长片段 DNA 文库构建方法

(57) Abstract: Disclosed is a method for constructing a long fragment DNA library, comprising the following steps: 1) breaking a long fragment DNA into target fragments of 3-10 kb by transposase, then amplifying the target fragments, and obtaining target fragment amplification products containing dUTP; 2) amplifying the dUTP in the products by removing the target fragments, fragmenting the target fragments secondarily into DNA short fragments of 300-1200 bp; 3) connecting both ends of the DNA short fragments with sequencing linker single chains A and sequencing linker single chains B respectively; and obtaining connecting sequencing linker products; and 4) PCR amplifying the connecting sequencing linker products, to obtain amplification products.

(57) 摘要: 提供了一种长片段 DNA 文库构建方法, 包括如下步骤: 1) 通过转座酶使长片段 DNA 断开为大小为 3-10kb 的目的片段, 再扩增所述目的片段, 得到含有 dUTP 的目的片段扩增产物; 2) 通过去除所述目的片段扩增产物中的 dUTP, 使所述目的片段二次片段化为 300-1200bp DNA 短片段; 3) 在所述 DNA 短片段两端分别连接测序接头单链 A 和测序接头单链 B; 得到连接测序接头产物; 4) PCR 扩增所述连接测序接头产物, 得到扩增产物。



WO 2016/169431 A1

一种长片段 DNA 文库构建方法

技术领域

本发明涉及生物技术领域，尤其涉及一种长片段 DNA 文库构建方法。

5 背景技术

长片段读取技术 (Long Fragment Read, LFR)，是美国 Complete Genomics 公司提出的 DNA 文库构建测序技术 (Methods and compositions for *long fragment read* sequencing, United States Patent 8,592,150)。其通过对基因组样品进行 384 孔板物理分隔的方法，将来自父本与母本的 DNA 长片段进行分离并分别加入不同的标签序列构建文库，在测序完成后可以完全定相基因组，确认突变位点是否在同一亲本染色体上。文库构建过程中对分离至孔板中的 DNA 长片段进行 MDA 扩增，在扩增过程中，掺入 dUTP 或其它 dNTP 类似物。后续通过内切酶的作用去除 dNTP 类似物并使用缺口平移的方法将扩增产物打断成较小的片段，然后使用带方向性的接头通过“连接接头 1-延伸反应-连接接头 2”的多步反应实现接头 A 的连接操作。完成以上步骤后进行环化并按照 CG 建库的后续流程连接测序接头 B 并完成文库构建。长片段读取技术可以有效改善测序的准确性并且降低起始 DNA 的投入量。

为满足文库构建实验的需求，分入 384 孔板中的 DNA 片段需要先进行扩增处理，CG 公司使用多重链置换扩增 (Multiple Displace Amplification, MDA) 的方法对投入 384 孔板每个单孔中的 DNA 长片段进行扩增。而 MDA 方法容易形成非特异扩增，其在扩增过程中被置换的单链会与新的随机引物结合而形成高级复杂结构影响后续反应。并且，接头 A 连接的多步酶反应操作复杂繁琐。

25 发明公开

本发明的一个目的是提供一种长片段 DNA 文库构建方法。

本发明提供的方法，包括如下步骤：

- 1) 对长片段 DNA 依次进行转座酶断裂、引入 dUTP 扩增和去除 dUTP，得到断裂片段；
- 30 2) 将带有不同标签的测序接头单链 A 和与其部分互补的带有不同标签的测序接头单链 B 以单链的形式分别加入含有所述断裂片段的体系中反

应，使所述断裂片段两端连接测序接头，通过测序接头单链 A 和测序接头单链 B 中标签序列的排列组合，使每份所述断裂片段对应的测序接头均相互区别，得到连接不同测序接头的产物；

所述带有不同标签的测序接头单链 A 和所述带有不同标签的测序接头单链 B 退火可成所述测序接头；

3) 以所述连接测序接头后的产物为模板，以与所述测序接头匹配的引物，进行 PCR 扩增，得到的 PCR 扩增产物为连接不同测序接头的 PCR 扩增产物；

4) 用所述连接不同测序接头的 PCR 扩增产物构建文库，即得到长片段 DNA 文库。

所述构建文库（具体见实施例）为将所述连接测序接头产物的 PCR 扩增产物依次进行消化 dUTP、双链环化、EcoP15 酶切、末端补平、去磷酸化、第二接头连接、第二接头连接产物扩增、分离得到单链，单链环化、即得到长片段 DNA 文库。

上述方法中，所述步骤 1) 的方法包括如下步骤：

(1) 用转座酶将所述长片段 DNA 进行断裂、并在断裂后的片段两端均连上扩增接头，得到连接扩增接头的产物；

(2) 以所述连接扩增接头的产物为模板，以与所述扩增接头匹配的 DNA 为引物，进行第一次 PCR 扩增，得到第一次 PCR 扩增产物；所述第一次 PCR 扩增过程中引入 dUTP；

上述 PCR 扩增的反应体系（没有模板）：2x buffer 291.2uL, Primer B (20uM) 8.48uL, Primer C (20uM) 8.48uL, dNTP (各 25mM) 20.17uL, dUTP(4mM) 5.04uL, Pfu turbo Cx 聚合酶 7.68uL, 20%（体积百分比）Triton X-100 水溶液 20.8uL, 10%（体积百分比）Tween20 水溶液 20.8uL, 加入无核酶水至总体积 560uL。

(3) 去除所述第一次 PCR 扩增产物中的 dUTP，得到缺口，进行缺口平移，并在 3' 末端加 A，得到的产物为 1) 中所述断裂片段。

上述方法中，所述转座酶为包埋扩增接头的转座酶；

和/或，所述扩增接头为 1 种或 2 种，所述扩增接头由转座酶识别单链 DNA 分子和与其部分反向互补的单链 DNA 分子形成；

和/或，相邻两个所述包埋扩增接头的转座酶的作用位置之间的片段大小为 3-10kb。

和/或，(2)中与所述扩增接头匹配的 DNA 为引物为连接在所述断裂后的片段两端的扩增接头中与所述转座酶识别单链 DNA 分子反向互补的单链 DNA 分子中，除所述转座酶识别单链 DNA 分子反向互补序列外，剩余部分序列形成的引物对。

上述方法中，

所述扩增接头为接头 1 和接头 2，

所述扩增接头为接头 1 和接头 2，所述接头 1 为由转座酶识别单链 DNA 分子 A 和与其部分反向互补的单链 DNA 分子 B 组成，所述接头 2 为由所述转座酶识别单链 DNA 分子 A 和与其部分反向互补的单链 DNA 分子 C 组成；

所述与所述扩增接头匹配 DNA 分子作为引物由引物 B 和引物 C 组成，所述引物 B 为所述单链 DNA 分子 B 中除去与所述转座酶识别单链 DNA 分子 A 互补部分外的其余序列；所述引物 C 为所述单链 DNA 分子 C 中除去与所述转座酶识别单链 DNA 分子 A 互补部分外的其余序列；

和/或，去除所述第一次 PCR 扩增产物中的 dUTP 的方法包括如下：用尿嘧啶 DNA 糖基化酶和人源脱嘌呤脱嘧啶核酸内切酶对所述第一次 PCR 扩增产物进行酶切反应，得到酶切产物；再用聚合酶 I、Taq 聚合酶和 dATP 对所述酶切产物进行聚合反应，得到 300-1200bp 大小的 DNA 片段。

上述酶切反应体系（不含有模板）：UDG（2U/uL）14.56 uL，APE 1（10U/uL）2.9uL，加无核酶水将总体积补充至 560uL。

上述聚合反应体系（不含有模板）：聚合酶 I（10U/uL）2.86uL，Taq 聚合酶（5U/uL）5.7uL，dATP（100mM）50.4uL，加无核酶水将总体积补充至 560uL。

上述方法中，所述标签序列是由 n 个碱基经过排列组合得到，所述碱基为 A、G、C、T 中至少一种，n 大于等于 8。

上述方法中，所述带有不同标签的测序接头单链 A 从 5' 至 3' 方向依次包括片段甲、片段乙、标签序列和片段丙；

所述带有不同标签的测序接头单链 B 从 5' 至 3' 方向依次包括与所述片段丙反向互补的片段、标签序列、片段丁和与所述片段甲反向互补的

片段；

所述步骤 3) 中，所述引物中的一条引物与所述测序接头单链 A 中的片段乙匹配（相同或反向互补），所述引物中的另一条引物与单链 B 中的片段丁匹配（反向互补或相同）。

- 5 上述方法中，所述带有不同标签的测序接头单链 A 的条数为小于等于 72，且各条单链标签序列不同；

所述带有不同标签的测序接头单链 B 的条数为小于等于 72，且各条单链标签序列不同；

所述标签序列中 n 大于等于 8 且小于 15。

- 10 上述方法中，步骤 2) 中，所述将带有标签的测序接头单链 A 和与其互补的带有不同标签的测序接头单链 B 以单链的形式分别加入含有所述断裂片段的体系中反应包括如下步骤：

- 2) -A, 将 72 条所述带有不同标签序列的测序接头单链 A 分别加入含有所述断裂片段体系的芯片的纵向 72 列中；所述测序接头单链 A 反应缓冲液不含 T4 连接酶；
- 15

2) -B, 将 72 条所述带有不同标签序列的测序接头单链 B 分别加入 2) -A 处理后的芯片的横向 72 行中，反应，得到连接测序接头产物。所述测序接头单链 B 反应缓冲液含有 T4 连接酶；

- 上述方法中，步骤 3) 中，所述引物中的一条引物与所述测序接头单链 A 中的片段乙相同或反向互补；
- 20

所述引物中的另一条引物与所述测序接头单链 B 中的片段丁的片段反向互补或相同。

所述 PCR 扩增是将 5184 孔板上各孔中所述连接不同测序接头的产物混合一起进行扩增；

- 25 上述方法中，所述方法的步骤 1) 的 (2)、(3) 和步骤 2) 均在 5184 孔芯片中进行。所述在所述断裂片段被分成若干份的情况下，是指将所述断裂片段分装在 5184 孔芯片的各孔中，分装可在步骤 1) 或步骤 2) 中进行。

- 上述方法中，所述长片段 DNA 为大于 100Kb 的片段。具体为大小为 30 400kb 的片段；

所述转座酶为 Tn5 转座酶；本实施例中，包埋扩增接头的转座酶为 Vazyme 公司产品，命名为 TruePrep mini DNA Sample Prep Kit (S302-01-B) 转座酶试剂盒，最佳使用浓度为 100 倍稀释转座酶。包埋扩增接头的转座酶和长片段 DNA 的量见如下反应体系中，上述反应的反应体系包括所述单链 DNA 分子 B、所述单链 DNA 分子 C、反应缓冲液、dNTP、dUTP 和聚合酶；
5 2uL 100 倍稀释后包埋扩增接头的转座酶、2 uL 5x buffer (S302-01, Vazyme 公司)、7ng 人基因组 DNA (长度为大于 100kb)，加水补充至 10 uL。

上述方法中，所述方法的步骤 1) 的 (2)、(3) 和步骤 2) 中均采用 wafergen MSND 移液平台将各种物质加入芯片的孔中。
10

上述方法中，在所述用 wafergen MSND 移液平台将各种物质加入芯片的孔中后，还包括如下步骤：离心沉降各物质。

上述方法中，步骤 1) 的 (1) 和 (2) 中，还包括如下步骤：用变性试剂消化所述转座酶；所述变性试剂具体采用 0.1-1%SDS 溶液；所述消化的条件具体为 25℃ 静置 10 分钟。
15

所述转座酶识别单链 DNA 分子 A 的核苷酸序列为序列表中序列 5；
所述单链 DNA 分子 B 的核苷酸序列为序列表中序列 6；
所述单链 DNA 分子 C 的核苷酸序列为序列表中序列 7；
所述引物 B 的核苷酸序列为序列表中序列 12；
20 所述引物 C 的核苷酸序列为序列表中序列 13；
所述测序接头单链 A 的核苷酸序列为序列表中序列 1 或序列 8；
所述测序接头单链 B 的核苷酸序列为序列表中序列 2 或序列 9；
所述单链引物甲（一条引物）的核苷酸序列为序列表中序列 3 或序列 10；
25 所述单链引物乙（另一条引物）的核苷酸序列为序列表中序列 4 或序列 11。

上述方法中，步骤 1) 的 (1) 和 (2) 中，还包括如下步骤：用变性试剂消化所述转座酶；所述变性试剂具体采用 0.1-1%SDS 溶液；所述消化的条件具体为 25℃ 静置 10 分钟。上述变性试剂为含有 SDS 溶液，具体由
30 11.2uL 1%SDS (质量/体积百分比) 和 548.8uL 2x buffer (MP01137,

Complete Genomics) 组成。

上述方法中，步骤 1) 的 (1) 中，所述转座酶断裂反应的条件为 55℃ 5 分钟；

步骤 1) 的 (2) 中，所述扩增的退火温度为 68℃，退火时间为 18min；

5 步骤 1) 的 (3) 中，所述酶切反应条件为 37℃ 2 小时，再 65℃ 15 分钟；所述聚合反应条件为 23℃ 1 小时，65℃ 30 分钟；

步骤 2) 的 2) -B 中，所述反应条件为 20℃ 2 小时；

步骤 3) 中，所述 PCR 扩增的退火温度为 68℃，退火时间为 10min。

由上述方法制备得到的长片段 DNA 文库也是本发明保护的范围。

10 本发明的另一个目的是提供一种用于长片段 DNA 文库构建的连接测序接头产物制备方法。

本发明提供的方法，包括如下步骤：为上述方法中的步骤 1) -2) 。

上述的方法在构建长片段 DNA 文库中的应用也是本发明保护的范围。

15 本发明的变性试剂可以为 NTbuffer、SDS 或其他可以使转座酶变性并从 DNA 长片段中脱落的试剂。

附图说明

图 1 为转座酶打断基因组片段示意图。

图 2 为 wafergen MSND 移液平台芯片示意图。

20 图 3 为转座酶在与基因组 DNA 结合示意图。

图 4 为转座酶脱落后加入与转座酶插入的接头 1/2 相匹配的引物序列进行扩增。

图 5 为 dUTP 在 UDG 酶及 APE1 酶的作用下被从 PCR 产物中切除且在产物上留下 1bp 的缺口。

25 图 6 为依靠 dUTP 的消化所产生的缺口平移。

图 7 为第一接头的各自带有标签序列的两条单链的加入方式示意图。

图 8 为第一接头的两条单链与插入片段连接方式示意图。

图 9 为第一接头连接完成后 PCR 扩增反应示意图。

图 10 为 wafergen MSND 两种喷液方式示意图。

30 图 11 为转座酶片段打断后 PCR 产物电泳胶图。

图 12 为芯片建库过程各步骤产物电泳图。

图 13 为 EcoP15 内切酶切割产物。

图 14 为单链环化后电泳图。

图 15 为各孔中的数据量及对应频数的直方图。

图 16 测序深度分布曲线图。

5 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。

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。

本发明通过利用转座酶插入特异序列，借助该特异序列在带有 5184 孔的芯片中进行 PCR，最终实现文库构建。

下述实施例中借助 wafergen 公司的 MSND 移液平台及带有 5184 孔的芯片，每次可以对样品或反应液进行不同方式的喷液，具体举例如下表 1：

表 1 为不同方式的喷液

分液程序	分 液 体 积	应用
单样品分液流程	35nL	同一样品分液至 5184 孔中
72 样品 35nL 分液 流程	35nL	单个样品分至芯片的一个 纵列
72 样品 50nL 分液 流程	50nL	单个样品分至芯片的一个 横排

实施例 1、构建长片段 DNA 文库(适配 Complete Genomics 测序平台)

一、通过转座酶使长片段 DNA 断开为大小为 3-10kb 的目的片段

图 1 为转座酶打断基因组片段示意图，包埋有接头 1 和接头 2 的转座酶，在与基因组 DNA 混合后，随机与基因组中的位置结合，通过控制转座酶的使用量，控制相邻两个转座酶作用位置之间的片段大小约为 3~10kb 之间。

图 2 为 wafergen MSND 移液平台芯片示意图。该芯片横向共 72 行，纵向共 72 列，总微孔数量为 5184 个。

转座酶包埋 1 种或 2 种接头都可以。

本实施例采用的包埋扩增接头的转座酶包埋有两种接头，为将转座酶

识别单链 DNA 分子 A 和与 A 部分反向互补的单链 DNA 分子 B 退火得到的接头、转座酶识别单链 DNA 分子 A 和与 A 部分反向互补的单链 DNA 分子 C 退火得到的接头。这两种接头按各 100uM 的浓度等体积混合，然后与转座酶混合并进行包埋得到。本实施例采用的包埋扩增接头的转座酶为 Vazyme 公司产品 Tn5 转座酶，命名为 TruePrep mini DNA Sample Prep Kit (S302-01-B) 转座酶试剂盒。

上述序列信息如下：

转座酶识别单链 DNA 分子 A: 5' -CTGTCTCTTATACACATCT-3' (序列 5)

单链 DNA 分子 B: 5' -TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3' (序列 6)

单链 DNA 分子 C: 5' -GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3' (序列 7)

在离心管中，控制包埋有接头序列的 Tn5 转座酶的使用量对基因组 DNA 进行作用，使两个相邻的转座酶作用位点在同一条 DNA 片段相距较长距离（约 3-10K）。由于转座酶本身的特性，在作用完成后并不会立刻脱离 DNA，而是仍然存在于作用位点，连接住被打断的长片段 DNA，故此时整条 DNA 长片段仍然保持原有长度而非断开状态。

将上述转座酶作用产物稀释至一定浓度（使 DNA 长片段间可以实现充分分离），然后利用 wafergen 的微量移液平台转移到芯片之中，实现 DNA 长片段的物理分隔。该芯片带有 5184 个微孔，即最多可将单个 DNA 样品分离成 5184 份。后续步骤加入的试剂均使用该移液平台进行。

1、摸索转座酶使用浓度

利用转座酶介导 PCR 反应，获得实验需求的片段长度 3-10kb。

将包埋扩增接头的转座酶分别稀释 50 倍、100 倍和 150 倍三个梯度。

在 PCR 反应管中加入如下反应体系：5x buffer 2uL (S302-01, Vazyme 公司)，人基因组 DNA（从离体血液细胞中提取得到基因组 DNA；7ng，大于 100kb）6uL，转座酶（不同稀释倍数）2uL。

将上述加有不同稀释倍数转座酶的反应体系混合后分别放入 PCR 仪中，55 摄氏度 5 分钟反应，然后冷却至常温，反应混合物加入终浓度 0.1%

的变性剂 SDS 混合，在室温下静置 10 分钟，得到大小为 3-10kb 的目的片段。然后将大小为 3-10kb 的目的片段稀释至终浓度约 8pg/uL。

2) PCR 反应

将上述 1) 得到的大小为 3-10kb 的目的片段加入如下 PCR 反应体系
5 进行 PCR 扩增，得到目的片段扩增产物。

用于扩增的引物为连接在所述断裂后的片段两端的扩增接头中与所述转座酶识别单链 DNA 分子反向互补的单链 DNA 分子中，除与所述转座酶识别单链 DNA 分子反向互补序列外，剩余部分序列形成的引物对，具体如下：

10 引物 B: 5' -TCGTCGGCAGCGTC-3' (序列 12)

引物 C: 5' -GTCTCGTGGGCTCGG-3' (序列 13)

上述 100uL PCR 反应体系包括：大小为 3-10kb 的目的片段 (8pg/uL) 5uL、2x buffer 50uL、引物 B (20uM) 0.5uL、引物 C (20uM) 0.5uL、DNA 聚合酶 0.5uL，补水至 100uL。

15 使用表 2 所示的 PCR 程序进行扩增。

表 2 为 PCR 程序

反应温度	反应时间	循环数量
72℃	10min	
94℃	3min	
94℃	30sec	18cycles
68℃	18min	
10℃	On hold	

不同稀释倍数的转座酶结果如下：

100 倍稀释转座酶条件下目的片段扩增产物的结果所示，可以看出，得到 3-10kb 的扩增产物，为所需目的片段。

20 50 倍稀释转座酶条件下和 150 倍稀释转座酶条件下均未得到 3-10kb 的目的片段扩增产物。

可以看出，最优转座酶稀释倍数为 100 倍稀释。

2、最优转座酶稀释浓度片段化基因组 DNA 目的片段

采用上述 1 摸索的最优 100 倍稀释转座酶对基因组 DNA 进行处理，使

基因组 DNA 约 3-10K 左右距离间加入转座酶包埋带有的 PCR 接头序列，并以此接头序列为起点进行 PCR 扩增，具体如下：

1)、转座酶片段化 DNA

将包埋扩增接头的转座酶和人基因组 DNA 按照如下反应体系进行反应，
5 具体如下：

上述反应体系如下：2uL 100 倍稀释后包埋扩增接头的转座酶、2 uL 5x buffer (S302-01, Vazyme 公司)、7ng 人基因组 DNA (长度为 400kb)，加水补充至 10 uL。

反应体系可按照比例扩大或缩小。

10 上述反应为：反应体系放入 PCR 仪中，55 摄氏度 5 分钟反应，然后冷却至常温，得到反应产物，将其稀释至终浓度约 8pg/uL，得到稀释后的反应产物（每 3-10KD 插入转座酶的基因组 DNA）。

使用 wafergen MSND 移液平台单样品分液流程将稀释后的片段化产物按照 35nL/孔分入 5184 孔板各孔中（孔的溶剂为 200nL）；再对芯片进行
15 离心（4000rpm, 5min），使所喷液体沉降至芯片底部，得到含有反应产物的 5184 孔板。

图 3 为转座酶在与基因组 DNA 结合并插入接头 1/2 后，并不会立刻脱离而是附着在其作用位点中。在反应溶液中加入一定浓度的 SDS，使得转座酶从 DNA 中脱落。

20 2)、消化转座酶

使用 wafergen MSND 移液平台单样品分液流程将变性试剂加入上述 1) 得到的含有反应产物的 5184 孔板各孔中，再对该芯片进行离心（时间 4000rpm, 5min），使所喷液体沉降至芯片底部；离心完毕后室温下静置 10 分钟左右，用于消化转座酶，使其脱离基因组 DNA 并实现基因组 DNA 的片
25 段化，得到 3-10KD 目的片段，含有其的芯片为含有 3-10KD 目的片段的 5184 孔芯片。

本操作可使用其它变性试剂及浓度，不限于 SDS。

上述变性试剂由 11.2uL 1%SDS（质量/体积百分比）和 548.8uL 2x buffer 组成；

30 3)、扩增目的片段得到目的片段扩增产物

图 4 为在完成转座酶的脱落后，在 PCR 体系中加入引物 B 和引物 C 进行扩增。同时，在 PCR 体系中投入一定比例的 dUTP（4%，dUTP:dATP+dTTP+dCTP+dGTP），使扩增产物中掺入一定比例的 dUTP。

使用设计好的引物，对上一步所得的产物进行 PCR。在 PCR 体系之中掺入 dUTP（dUTP：dNTP=4：100），使最终产物中部分的 dTTP 被 dUTP 代替，具体如下：

使用 wafergen MSND 移液平台单样品分液流程将下述 PCR 反应体系加入步骤 2) 得到的含有 3-10KD 目的片段的 5184 孔芯片各孔中，之后离心使所喷液体沉降，再将芯片板放入 PCR 仪中进行 PCR 扩增反应，得到含有 dUTP 的目的片段扩增产物的 5184 孔芯片。

在上述反应体系中加入一定比例的 dUTP，使得在扩增中一部分的 dTTP 被 dUTP 替代，从而使扩增产物在后期可通过 dUTP 的去除而被再次打断。

上述 PCR 扩增反应体系：2x buffer 291.2uL，单链 DNA 分子 B(20uM) 8.48uL，单链 DNA 分子 C (20uM) 8.48uL，dNTP（各 25mM）20.17uL，dUTP(4mM) 5.04uL，Pfu turbo Cx 聚合酶 7.68uL，20%（体积百分比）Triton X-100 水溶液 20.8uL，10%（体积百分比）Tween20 水溶液 20.8uL，加入无核酶水至总体积 560uL。

上述 PCR 扩增反应程序如表 3 所示：

表 3 为 PCR 扩增反应程序

反应温度	反应时间	循环数量
72℃	10min	18cycles
94℃	3min	
94℃	30sec	
68℃	18min	
10℃	On hold	

图 11 为转座酶片段打断后 PCR 产物电泳胶图。

反应完毕后，将芯片放置于 37℃ 金属浴中静置 15.5 分钟蒸干多余液体。

二、3-10KD 目的片段二次片段化为 300-1200bp DNA 短片段

1) 去除 dUTP 形成单链缺口

使用具有单独去除 dUTP 活性的酶 (UDG、USER 等), 对上步 PCR 产物作用, 去除其中的 dUTP 及该位置的糖骨架, 得到中间带有一些 1 个碱基长度的缺口的 DNA 双链产物 (图 5), 具体如下:

- 使用尿嘧啶 DNA 糖基化酶(UDG)及人源脱嘌呤脱嘧啶核酸内切酶(APE
5 1) 消化扩增中引入的 dUTP。

具体如下:

- 按照 wafergen MSND 移液平台单样品分液流程将酶切反应体系加入上述一得到的含有 dUTP 的目的片段扩增产物的 5184 孔芯片的各孔中, 之后离心使所喷液体沉降 (4000rpm, 5min); 再将芯片放入专用 PCR 仪中,
10 进行酶切反应, 得到含有酶切产物的 5184 孔芯片。

上述酶切反应体系: UDG (2U/uL) 14.56 uL, APE 1 (10U/uL) 2.9uL, 加无核酶水将总体积补充至 560uL。

上述酶切反应条件为, 37℃ 2 小时, 65℃15 分钟, 之后恢复至常温放置。

- 15 2) 聚合反应完成断裂

- 缺口平移打断 DNA, 并在末端加 A。在反应体系中加入 DNA 聚合酶 I 和 Taq 酶。在 23 摄氏度下, 利用 DNA 聚合酶 I 的作用活性 (5' -3' 外切酶活性和 5' -3' 聚合酶活性), 对上一步所得产物进行打断操作。DNA 聚合酶 I 结合于去除 dUTP 后留下的缺口处, 然后从 5' -3' 方向切除 DNA
20 碱基, 并从 5' -3' 方向进行 DNA 碱基的聚合, 实现缺口从 5' -3' 方向的平移。当 DNA 双链的两个缺口因为平移而相遇时, 该 DNA 双链便会断开成更小的片段。同时在体系中加入的 Taq 聚合酶, 在 65 摄氏度下对 DNA 双链进行补平, 并在 3' 末端加入一个 dATP 碱基。而在该温度下, DNA 聚合酶 I 因变性而失去活性。

- 25 图 6 为依靠 dUTP 的消化所产生的缺口, 后来投入到反应体系中的 DNA 聚合酶 I 结合在该缺口位置发挥其 5' -3' 方向切割及 5' -3' 方向合成的作用, 使得该缺口沿 5' -3' 方向平移。双链中位于正反两条链上的缺口同时进行平移并相遇, 最终在相遇位置使整个双链断开。而同时加入到体系中的 Taq 酶, 在 65℃ 的温度下发挥其活性, 使得产物的 3' 末端形成突出一个
30 个 A 碱基。

利用聚合酶缺口平移的特性，依靠上述去除 dUTP 后的缺口进行缺口平移实现 DNA 片段打断，同时在反应体系中加入 Taq 聚合酶，在反应中使缺口平移后的 DNA 产物的 3' 末端加入 dATP，具体如下：

按照 wafergen MSND 移液平台单样品分液流程将聚合反应体系喷液至
5 上述 1) 含有酶切产物的 5184 孔芯片各孔中，之后离心使所喷液体沉降（4000rpm，5min）；再将芯片放入专用 PCR 仪中，进行聚合反应，得到含有 300-1200bp 大小的 DNA 片段 5184 孔芯片。

上述聚合反应体系：聚合酶 I（10U/uL）2.86uL，Taq 聚合酶（5U/uL）5.7uL，dATP（100mM）50.4uL，加无核酶水将总体积补充至 560uL。

10 上述反应条件为，23℃ 1 小时，65℃ 30 分钟，之后恢复至室温。

三、连接测序接头

在上述二得到的 300-1200bp 大小的 DNA 片段两端连接带有部分测序引物的测序接头，测序接头由带有不同标签的测序接头单链 A 和带有不同标签的测序接头单链 B 退火形成，为了对芯片上每一个孔都可以在测序的过程中进行区分，设计在芯片中纵向的每一列（对应测序接头单链 A）加入
15 编号为 1-72 的带有不同标签测序接头单链 A，在芯片中横向的每一行（对应测序接头单链 B）加入编号为 1-72 的带有不同标签测序接头单链 B，这样通过两个单独的标签序列便形成 72x72 的矩阵，使得每一个孔都获得特有的双标签序列组合。

20 测序接头单链 A

CGAAGCACTCAAA**GGTCGCCAGCCCTATGGC**NNNNNNNNNNCAGTACG**TCAGCAGT***T（序列 1）

其中，从 5' 至 3' 方向所有的粗体如下：

片段甲 **CGAAGCACTC** 与测序接头单链 B3' 端的 **GAGTGCTTCG** 反向互补；

25 片段乙 **GGTCGCCAGCCCTATGGC** 与序列 3 所示的 PCR 引物第一链相同，且仅将序列 3 的 U 替换为 T；

NNNNNNNNNN 为 10 个碱基组成标签序列；

片段丙 **TCAGCAGT** 与测序接头单链 B5' 端的 **ACTGCTGA** 反向互补；

30 *代表的是硫代磷酸，即在最后一个碱基合成时，使用的是硫代磷酸的核酸。

测序接头单链 B

/5Phos/**ACTGCTGAGTCGAG**ANNNNNNNNN**CTGACAAGGTCGCCAGCCCTGAGTGCT**
TCGAA/3' AmMO/ (序列 2)

其中, 从 5' 至 3' 方向所有的粗体如下:

5 **ACTGCTGA** 与测序接头单链 A3' 端的片段丙 **TCAGCAGT** 反向互补;

NNNNNNNNNN 为 10 个碱基组成标签序列;

片段丁 CTGACAAGGTCGCCAGCCCT 与序列 4 所示的 PCR 引物第二链反向互补, 且仅将序列 4 的 U 替换为 T;

GAGTGCTTCG 与测序接头单链 A5' 端的片段甲 **CGAAGCACTC** 反向互补;

10 图 7 为第一测序接头的各自带有标签序列的两条单链的加入方式示意图。为了使芯片中每一个微孔均带有特有的标签序列以进行相互区分, 采用两个标签序列组成一个组合的形式。在实验中第一接头的两条单链分别加入。其中第一链的分别 72 种标签序列的单链 (测序接头单链 A), 以纵向形式加入, 即每一纵列加入的第一链的标签序列相同。第二链的分别 72
15 种标签序列的单链 (测序接头单链 B), 以横向形式加入, 即每一横行加入的第二链的标签序列相同。这样最终形成 72x72 中标签序列的矩阵, 每一个孔获得唯一的两两标签序列组合。

为了使用两个单独的标签序列实现 72x72 种标签序列组合, 在连接接头的步骤中, 带有标签序列的测序接头单链 A 以纵向方式加入, 带有标签
20 序列的测序接头单链 B 以横向方式加入, 具体如下:

1、测序接头单链 A 的添加

配制测序接头单链 A 测序接头单链 A 反应液如下: 10x buffer (25% (体积比) PEG8000、500mM Tris-HCl、10mM ATP、100mM MgCl_2) 252uL, 加无核酶水将总体积补充至 957uL;

25 1) 按照 wafergen MSND 移液平台 “72 样品 35nL 分液流程” 要求将上述测序接头单链 A 反应液以每孔 11.96uL 加入到上述二得到的含有 300-1200bp 大小的 DNA 片段 5184 孔芯片的纵向 72 列孔中;

2) 再将 72 种不同的测序接头单链 A 测序接头单链 A (10uM) 以每孔 2.14uL 的量按照 wafergen MSND 移液平台 “72 样品 35nL 分液流程” 加入
30 上述 1) 处理后的 5184 孔板的纵向 72 列孔中, 离心沉降液体, 备用,

得到添加测序接头单链 A 测序接头单链 A 的 5184 孔板。

2、测序接头单链 B 的添加

配制测序接头单链 B 测序接头单链 B 反应液： 10xbuffer (25% (体积比) PEG8000、500mM Tris-HCl、10mM ATP、100mM MgCl_2) 296.4uL、T4
5 连接酶 (600U/uL, Enzymatics) 125.97uL, 加无核酶水将总体积补充至 1186uL。

1) 按照 wafergen MSND 移液平台 “72 样品 50nL 分液流程” 要求将上述测序接头单链 B 测序接头单链 B 反应液以每孔 13.95uL 的体积加入到上述 1 得到的添加测序接头单链 A 的 5184 孔板的横向 72 行孔中；

10 2) 再将 72 种不同的测序接头单链 B (10uM) 以每孔 1.65uL 的量按照 wafergen MSND 移液平台 “72 样品 50nL 分液流程” 加入上述 1) 处理后的 5184 孔板的横向 72 行孔中，离心沉降液体，备用，得到添加测序接头单链 B 的 5184 孔板。

将上述添加测序接头单链 B 的 5184 孔板置于 PCR 仪中反应，反应条件为 20℃ 2 小时，得到含有连接测序接头产物的 5184 孔板。
15

图 8 为第一接头(测序接头)的两条单链与插入片段连接方式示意图。

图 10 为 wafergen MSND 两种喷液方式示意图。左图为 “72 样品 35nL 分液流程” 配合测序接头单链 A 的加液方式，使得第一链的分别 72 种标签序列的单链 (测序接头单链 A)，实现纵向形式加入。右图为 “72 样品
20 50nL 分液流程” 配合第一接头第二链的加液方式，使得第二链的分别 72 种标签序列的单链 (测序接头单链 B)，实现横向形式加入。

四、PCR 扩增

图 9 为第一接头 (测序接头) 连接完成后 PCR 扩增反应示意图。

将上述三获得的含有连接测序接头产物的 5184 孔板所有孔的产物均
25 通过离心的方式取出并混合于一离心管之中，然后进行 PCR 反应扩增，扩增过程会将接头不匹配的部分重新变成匹配的双链。由于产物同一条单链两端是带有不同标签序列的两条接头单链，所以即使 PCR 后接头的形状进行改变，但不会影响其标签序列组合。也因此，各个孔的产物混合在一起，也不会影响最后对每一个标签序列的组合的拆分，具体如下：

30 将上述三获得的含有连接测序接头产物的 5184 孔板所有孔的产物离

心收集于 1.5mL 离心管中，并使用 1 倍 AmpureXP 磁珠进行纯化，回收溶于 100uL TE 溶液中纯化产物，得到纯化后连接测序接头产物，对其进行 PCR 扩增反应，得到连接测序接头产物的 PCR 扩增产物。

上述 PCR 扩增采用的引物如下：

5 PCR 引物第一链（上游引物）：GGUCGCCAGCCCUATGGC（序列 3）

PCR 引物第二链（下游引物）：AGGGCUGGCGACCUTGTCAG（序列 4）

上述 PCR 扩增采用的反应体系如下：连接测序接头产物 100uL，2x PfuCx buffer 275uL，PCR 引物第一链(20uM)14uL,PCR 引物第二链(20uM) 14uL，PfuCx 聚合酶 11 uL，加无核酶水补充至 550uL。

10 上述 PCR 扩增的反应程序如下表 4。

表 4 为扩增反应程序

反应温度	反应时间	循环数量
95℃	5min	7cycles
95℃	30sec	
56℃	30sec	
72℃	4min	
68℃	10min	
10℃	0n hold	

将连接测序接头产物的 PCR 扩增产物使用 1 倍 AmpureXP 磁珠进行纯化，纯化产物溶于 60uL TE 溶液中。

上述各步骤反应产物电泳检测如下图 12 所示，1、长片段 PCR 扩增产物（步骤一的 2 的 3）得到的扩增产物）；2、长片段 PCR 阴性扩增产物（步骤一的 2 的 3）得到的扩增产物，模板用水替代基因组 DNA）；3、去除 dUTP 产物（步骤二的 1））；4、去除 dUTP 反应阴性产物（步骤二的 1），含有 dUTP 的目的片段扩增产物替换为水）；5、缺口平移末端加 A 产物(步骤二的 2) 聚合反应产物)；6、缺口平移末端加 A 阴性产物(步骤二的 2) 聚合反应中的酶切产物替换为水)；7、接头连接产物 1（步骤三的 2 的连接测序接头产物)；8 接头连接产物 2(步骤三的 2 的连接测序接头产物)；9、PCR 产物 1（步骤四的 PCR 扩增产物）；10、PCR 产物 2（步骤四的 PCR 扩增产物）；11、PCR 阴性产物；M2、100bp 分子量标记。从图 12 可以看

出各步骤得到的目标产物。

五、消化扩增产物的 dUTP

上述连接测序接头产物的 PCR 扩增产物需要经过消化扩增引物中带有
5 的 dUTP，形成粘性末端，从而在后续实验中实现自身环化。dUTP 消化
反应如下，10x Taq buffer 11uL (RM00059, Complete Genomics), User
enzyme (RM00017, Complete Genomics), 纯化后 PCR 产物 60uL，加无核
酶水补充至 110uL，37℃ 反应 1 小时，得到反应产物。

六、双链环化

上述五得到的反应产物加入 10xTAbuffer (RM06601, Complete
10 Genomics) 180uL，加无核酶水至 1810uL，混合物平均分成 4 管，于水浴
60℃ 反应 30 分钟后，转至常温水浴 20 分钟。

预先配制反应混合物，无核酶水 98uL，20x Circ mix (MP01134,
Complete, Genomics) 100uL，连接酶 (L603-HC-1500, Enzymatics) 2uL，
混合后平均分成 4 份加入上述反应产物中混合，室温反应 1 小时以环化，
15 得到环化产物。

对每一管反应产物，使用 AmpureXP 磁珠纯化回收，产物回收溶于 70uL
TE。

七、消化未环化 DNA

环化反应中存在相当一部分未环化 DNA，为了不影响后续实验结果，
20 需要对未环化的 DNA 进行消化。消化反应如下，对于每管上述六得到的纯
化环化产物，加入 9x PS mix (MP01154, Complete Genomics) 8.9uL, Plasmid-Safe (RM02046, Complete Genomics) 10.4uL，加无核酶
水补充至 80uL，混合后于 37℃ 反应 1 小时。反应产物使用 AmpureXP 磁珠
纯化后溶于 40uL TE 溶液，得到消化后环化的产物。

25 八、使用 EcoP15 酶切并筛选目的片段

环化后的产物，最初加入的两端接头连接在一起，此处使用 EcoP15
酶切，使环化产物由接头两端的 EcoP15 限制酶切位点处向插入片段内部
两端各切下约 27bp，后续对酶切下来的产物进行片段筛选，获得中间带接
头序列，两端带有酶切下来的插入片段的产物，具体如下：

30 酶切消化体系配制，上述七得到的消化后环化的产物 37uL，5xEcoP 15

mix 3 (MP01149, Complete Genomics) 72uL, EcoP 15 (RM00063, Complete Genomics) 10.8uL, 加无核酶水补充至 360uL, 于 37℃ 反应 16 小时。

配制 PEG32 磁珠(按体积比混合, AmpureXP 磁珠:0.5%Tween=100:1, 下同)

- 5 对于上述 360uL 酶反应产物, 加入 252uL PEG32 磁珠结合, 结合产物保留上清, 并加入 PEG32 磁珠 184uL, 混合结合后回收磁珠并溶于 52uL TE 回收液(带有 0.001% Tween20, 体积比), 得到酶切产物。

图 13 为 EcoP15 内切酶切割, 片段回收后产物。产物片段在 140bp 左右。

10 九、末端补平

对 EcoP15 酶切产物进行末端补平, 以连接后续第二接头。

- 配制末端补平反应体系, 取上述八得到的酶切产物 44uL, 10x NEB buffer 2 (New England biolabs) 5.4uL, 25mM dNTP 0.8uL, 10mg/mL BSA 0.4uL, T4 DNA polymerase (M0203-com, New England Biolabs), 12℃
15 反应 20 分钟, 得到末端补平产物。

反应产物使用 PEG32 磁珠进行纯化并溶于 48uL TE。

十、去磷酸化

- 对上述九得到的末端补平产物进行去磷酸化处理, 以配合后续第二接头连接。配制反应体系如下, 10x NEB buffer 2 (New England Biolabs)
20 5.75uL, Fast AP (EF0651, Fermentas) 5.75uL, 加入末端补平纯化后产物 46uL, 于 37℃ 反应 45 分钟。反应产物使用 75uL PEG32 磁珠进行纯化, 并溶于 42uL TE 溶液, 得到去磷酸化产物。

十一、第二接头连接

- 第二接头进行定向连接, 需要经过 4 步酶反应, 先将上述十得到的
25 去磷酸化产物进行 3' 末端接头序列的引入, 再次进行末端补平并末端引入 dATP 后, 连接引入 5' 末端接头序列, 最后使用一段寡核苷酸序列替代其中一段序列并使用连接酶连接。

接头序列

3' 末端接头

- 30 /phos/ GTCTCCAGTCGAAGCCCGACG/3ddC/

/3ddC/AGAGGTCAGCTTCG

5' 末端接头

TTGGCCTCCGACT/3dT-Q/

/ddC/TGCTGGCGAACCGGAGGCTGA/5phos/

5

序列替换的两条序列

寡核苷酸序列 1

/52bio/TCCTAAGACCGCTTGGCCTCCGACT

寡核苷酸序列 2

10 /5phos/AGACAAGCTCGAGCTCGAGCGATCGGGCTTCGACTGGAGAC

PCR 引物序列

PCR 引物 1

/52bio/TCCTAAGACCGCTTGGCCTCCGACT

15 PCR 引物 2

/5phos/AGACAAGCTCGAGCTCGAGCGATCGGGCTTCGACTGGAGAC

第二接头 3' 末端序列引入

20 配制反应混合液, 3xHB (MP01139, Complete Genomics) 24.7uL, 无核酶水 1.9uL, T4 连接酶 (600U/uL, Enzymatics) 1.9uL, 二接头 3' 端接头 5.6uL (9uM), 加入去磷酸化的产物 40uL。混合物于 14℃ 反应 2 小时, 反应完成使用 63uL PEG32 磁珠进行纯化, 并溶于 42uL TE 溶液。

末端补平加 A

25 配制反应混合液, 5x klex NTA mix (MP01150, Complete Genomics) 10.7uL, klenow (RM00066, Complete Genomics) 1.1uL, 无核酶水 1.5uL, 加入上步骤产物 40uL。混合物于 37℃ 反应 1 小时。反应完成使用 69uL PEG32 磁珠进行纯化, 并溶于 42uL TE 溶液。

第二接头 3' 末端序列引入

30 配制反应混合液, 3xHB 24.7uL, 无核酶水 1.9uL, T4 连接酶 1.9uL, 二接头 5' 端接头 5.6uL (9uM), 加入上步骤产物 42uL。混合物于 14℃

反应 2 小时，反应完成使用 63uL PEG32 磁珠进行纯化，并溶于 42uL TE 溶液。

序列替换

- 配制 Oligo 反应液，10x Taq buffer (RM00059, Complete Genomics) 8uL, 100mM ATP (MP01164, Complete Genomics) 0.8uL, 25mM dNTP 0.32uL, 寡核苷酸序列 1 (20mM) 1uL, 寡核苷酸序列 2 (20mM) 1uL, 加入无核酶水至 32uL。

配制酶反应液，10x Taq buffer 0.4uL, T4 连接酶 4.8uL, Taq 聚合酶 4.8uL。

- 实验中将 32uL Oligo 反应液加入到 40uL 上述纯化后 DNA 产物中，然后加入 8uL 酶反应液，混合后于 37℃ 中孵育 20 分钟。反应完成使用 80uL PEG32 磁珠进行纯化，并溶于 47uL TE 溶液。纯化产物进行定量检测。

得到第二接头连接产物。

十二、第二接头连接产物扩增

- 往一 1.5mL 离心管中，加入 2xPfuCx mix3 275uL, PfuCx 聚合酶 11uL, 第二接头 PCR 引物 1 (20uM) 13.75uL, 第二接头 PCR 引物 2 (20uM) 13.75uL, 定量后的上述十一制备的第二接头连接产物 60ng, 加无核酶水补充至 550uL。反应物混合后平均分至 4 个 PCR 管中进行 PCR 扩增反应，得到扩增产物。

- 扩增反应程序如下表 5。

表 5 为扩增反应程序

反应温度	反应时间	循环数量
95℃	3min	18cycles
95℃	30sec	
60℃	30sec	
72℃	4min	
68℃	10min	
0.1℃/s 降温至 4℃0n hold		

反应完成使用 550uL PEG32 磁珠进行纯化，并溶于 85uL TE 溶液。对纯化产物进行定量检测，取 600ng 进行后续单链环化步骤。

十三、单链环化

配制 1xBWB/tween 混合液，往离心管中加入 1xBWB (MP01111, Complete Genomics), 0.5% Tween20 20uL, 充分混合。

清洗磁珠。取 streptavidin beads (MP01162, Complete Genomics) 120uL, 放置于在磁力架上使磁珠充分吸附后, 去除上清液。使用 1xBBB (MP01110, Complete Genomics) 600uL 清洗两次, 然后加入 120uL 的 1xBBB 使磁珠重新悬浮, 然后加入 1% 体积的 0.5% Tween20。

使用清洗后的磁珠对产物进行单链分离。取上述十二得到的扩增产物 600ng, 加入 TE 补充至 60uL, 加入 4xBBB (MP01145, Complete Genomics) 20uL, 30uL 清洗后的 streptavidin beads, 充分混合后静置 15 分钟, 然后放置于磁力架上使磁珠充分吸附, 使用配制好的 1xBWB/tween 混合液清洗 2 次。清洗完成后加入 75uL 0.1M NaOH 溶液, 混合放置 2 分钟后, 放置于磁力架上使磁珠充分吸附, 回收上清液。最后往回收的上清液中加入 0.3M MOPS acid (MP01165, Complete Genomics) 37.5uL, 混合均匀。

上述步骤结束后获得 112.5uL 单链产物, 产物进入单链环化步骤。预先配制以下反应液:

引物混合物, 取 Bridge Oligo (20uM) 20uL, 加入 43uL 无核酶水, 混合均匀待用;

酶反应混合物, 取无核酶水 135.3uL, 加入 10xTA buffer (RM06601, Complete Genomics) 35uL, 100mM ATP 3.5uL, 连接酶 (600U/uL) 1.2uL, 混合均匀待用。

往上述 112.5uL 产物中, 加入引物混合物 63uL, 酶反应混合物 175uL, 混合均匀后, 置于 37℃ 孵育 1.5 小时, 得到单链环化产物。

图 14 为单链环化后电泳图。

十四、消化未环化单链产物

往上述十三得到的单链环化产物中, 加入无核酶水 1.5uL, 10xTA buffer 3.7uL, Exo I (M0293L, New England Biolabs) 11.1uL, Exo III (M0206L, New England Biolabs) 3.7uL, 充分混合后, 置于 37℃ 反应 30 分钟。反应完成后加入 500mM EDTA 15.4uL, 充分混合以终止反应。反应混合物加入 500uL PEG32 磁珠进行纯化, 最终溶解于 16uL TE 溶液, 得到长

片段文库。

文库制备过程结束。

实施例 2、文库构建方法(适配 illumina 测序平台)

- 一、通过转座酶使长片段 DNA 断开为大小为 3-10kb 的目的片段
- 5 与实施例 1 的步骤相同。
- 二、目的片段 3-10KD 目的片段二次片段化为 300-1200bp DNA 短片段
- 与实施例 1 的步骤相同

三、连接测序接头

- 在上述二得到的 300-1200bp 大小的 DNA 片段上一步反应产物的两端
- 10 连接带有部分测序引物的测序接头，该步骤的接头双链分别带有不同的标签序列，为了对芯片上每一个孔都可以在测序的过程中进行区分，设计在芯片中纵向的每一列（对应测序接头第一条链）加入编号为 1-72 的标签序列，在芯片中横向的每一行（对应接头第二条链）加入编号为 1-72 的标签序列。这样通过两个单独的标签序列便形成 72x72 的矩阵，使得每一个孔都获得特有的双标签序列组合。在上一步反应产物两端连接带有部分
 - 15 测序引物序列的接头。

测序接头单链 A

AATGATACGGCGACCACCGAGATCTACACNNNNNNNACACTCTTTCCCTACACGACGCTCTT
CCGATC*T（序列 8）

- 20 从 5' 至 3' 方向依次由 Flow cell 接头 P5（斜体）、10 个碱基组成标签序列（N）和与 Read 1 测序接头互补的单链 DNA 分子(加粗，*代表的是硫代磷酸，即在最后一个碱基合成时，使用的是硫代磷酸的核酸)组成。

测序接头单链 B

- /5Phos/*GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNNNATCTCGTATGC*
25 *CGTCTTCTGCTTG* /3' AmMO/（序列 9）

从 5' 至 3' 方向依次由 Read 2 测序接头（加粗）、10 个碱基组成标签序列（N）和与 Flow cell 接头 P7 互补的单链 DNA 分子(斜体，Phos 即磷酸修饰，AmMo 即氨基修饰)组成；72 条 3' 端标签引物的 10 个碱基组成随机片段均不相同。

- 30 图 7 为第一接头的各自带有标签序列的两条单链的加入方式示意图。

为了使芯片中每一个微孔均带有特有的标签序列以进行相互区分，采用两个标签序列组成一个组合的形式。在实验中第一接头的两条单链分别加入。其中第一链的分别 72 种标签序列的单链，以纵向形式加入，即每一纵列加入的第一链的标签序列相同。第二链的分别 72 种标签序列的单链，以
5 横向形式加入，即每一横行加入的第二链的标签序列相同。这样最终形成 72x72 中标签序列的矩阵，每一个孔获得唯一的两两标签序列组合。

为了使用两个单独的标签序列实现 72x72 种标签序列组合，在连接接头的步骤中，第一个带有标签序列的接头单链以纵向方式加入，第二个带有标签序列的接头单链以横向方式加入，具体如下：

10 1、测序接头单链 A 的添加

配制测序接头单链 A 反应液如下：10x buffer（25%（体积比）PEG8000、500mM Tris-HCl、10mM ATP、100mM MgCl_2 ）252uL，加无核酶水将总体积补充至 957uL；

15 1）按照 wafergen MSND 移液平台“72 样品 35nL 分液流程”要求将上述测序接头单链 A 反应液以每孔 11.96uL 加入到上述二得到的含有 300-1200bp 大小的 DNA 片段 5184 孔芯片的纵向 72 列孔中；

2）再将 72 种不同的测序接头单链 A（10uM）以每孔 2.14uL 的量按照 wafergen MSND 移液平台“72 样品 35nL 分液流程”加入上述 1）处理后的 5184 孔板的纵向 72 列孔中，离心沉降液体，备用，得到添加测序接
20 头单链 A 的 5184 孔板。

2、测序接头单链 B 的添加

配制测序接头单链 B 反应液：10xbuffer（25%（体积比）PEG8000、500mM Tris-HCl、10mM ATP、100mM MgCl_2 ）296.4uL、T4 连接酶（600U/uL，Enzymatics）125.97uL，加无核酶水将总体积补充至 1186uL。

25 1）按照 wafergen MSND 移液平台“72 样品 50nL 分液流程”要求将上述测序接头单链 B 反应液以每孔 13.95uL 的体积加入到上述 1 得到的添加测序接头单链 A 的 5184 孔板的横向 72 行孔中；

2）再将 72 种不同的测序接头单链 B（10uM）以每孔 1.65uL 的量按照 wafergen MSND 移液平台“72 样品 50nL 分液流程”加入上述 1）处理
30 后的 5184 孔板的横向 72 行孔中，离心沉降液体，备用，得到添加测序接

头单链 B 的 5184 孔板。

将上述添加测序接头单链 B 的 5184 孔板置于 PCR 仪中反应，反应条件为 20℃ 2 小时，得到含有连接测序接头产物的 5184 孔板。

图 8 为第一接头的两条单链与插入片段连接方式示意图。

5 图 10 为 wafergen MSND 两种喷液方式示意图。左图为“72 样品 35nL 分液流程”配合第一接头第一链的加液方式，使得第一链的分别 72 种标签序列的单链，实现纵向形式加入。右图为“72 样品 50nL 分液流程”配合第一接头第二链的加液方式，使得第二链的分别 72 种标签序列的单链，实现横向形式加入。

10 四、PCR 扩增

图 9 为第一接头连接完成后 PCR 扩增反应示意图。

将上述三获得的含有连接测序接头产物的 5184 孔板所有孔的产物均通过离心的方式取出并混合于一离心管之中，然后进行 PCR 反应扩增，扩增过程会将接头不匹配的部分重新变成匹配的双链。由于产物同一条单链
15 两端是带有不同的两条接头单链/标签序列的，所以即使 PCR 后接头的形状进行改变，但不会影响其标签序列组合。也因此，各个孔的产物混合在一起，也不会影响最后对每一个标签序列的组合的拆分，具体如下：

将上述三获得的含有连接测序接头产物的 5184 孔板所有孔的产物离心收集于 1.5mL 离心管中，并使用 1 倍 AmpureXP 磁珠进行纯化，回收溶
20 于 100uL TE 溶液中纯化产物，得到纯化后连接测序接头产物，对其进行 PCR 扩增反应，得到连接测序接头产物的 PCR 扩增产物。

上述 PCR 扩增采用的引物如下：

PCR 引物第一链（上游引物）：AATGATACGGCGACCACCGAGATCT（序列
10）

25 PCR 引物第二链（下游引物）：CAAGCAGAAGACGGCATACGAGAT（序列 11）

上述 PCR 扩增采用的反应体系如下：连接测序接头产物 100uL，2x PfuCx buffer 275uL，PCR 引物第一链(20uM)14uL，PCR 引物第二链(20uM) 14uL，PfuCx 聚合酶 11 uL，加无核酶水补充至 550uL。

上述 PCR 扩增的反应程序如下表 6。

30 表 6 为扩增反应程序

反应温度	反应时间	循环数量
95℃	5min	7cycles
95℃	30sec	
56℃	30sec	
72℃	4min	
68℃	10min	
10℃	0n hold	

将连接测序接头产物的 PCR 扩增产物使用 1 倍 AmpureXP 磁珠进行纯化，纯化产物溶于 60uL TE 溶液中。

上述各步骤反应产物电泳检测，得到目标产物。

至此文库制备结束。

5 实施例 3、文库测序结果分析

一、测序

使用实施例 2 制备的 DNA 长片段文库在 illumina 平台上进行测序，测序深度 40X。

二、比对

10 使用序列比对软件 SOAP aligner 2.20 (LiR, LiY, Kristiansen K, et al, SOAP: short oligonucleotide alignment program. Bioinformatics 2008, 24(5):713-714; LiR, YuC, LiY, et al, SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 2009, 25(15):1966-1967;

15 <http://soap.genomics.org.cn/soapaligner.html>) 把读取片段(reads) 比对人类参考基因组 Reference hg19 (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/>), 比对存在多重结果时只选取一个比对结果 (-r 1)。

三、按照标签组合信息，确定每个孔对应的 reads。

20 四、统计各孔中的数据量及对应频数，绘制直方图（图 15）。

数据统计如下：

单位：Mb

均值：36.9000 标准差：13.55647

最小值：0.3168 最大值：123.2000 中位数：36.7100

由上可知，本方法可以将所有 5184 个孔进行成功标记。

实施例 4：不同文库构建方法的测序结果对比

5 一、文库构建和测序

使用 CG 公司的多重链置换扩增 (Multiple Displace Amplification, MDA) 的长片段 DNA 文库构建方法 (Methods and compositions for *long fragment read* sequencing, United States Patent 8,592,150) 制备 DNA 长片段文库,然后在 Complete Genomics 平台上进行测序,测序深度 80X。

10 使用实施例 1 制备 DNA 长片段文库,然后在 Complete Genomics 平台上进行测序,测序深度 60X。

二、比对

使用序列比对软件 SOAP aligner 2.20 (LiR, LiY, Kristiansen K, et al, SOAP: short oligonucleotide alignment program. Bioinformatics 2008, 24(5):713-714; LiR, YuC, LiY, et al, SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 2009, 25(15):1966-1967;

http://soap.genomics.org.cn/soapaligner.html) 把测序后得到读取片段 (reads) 分别比对到人类参考基因组 Reference hg19 (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/), 比对存在多重比对时只选取一个比对结果 (-r 1)。

三、使用 SOAP aligner 2.20 中的 soap.coverage 计算 Reads 在 Reference 上的单点覆盖度 (-cvg)。

四、绘制测序深度分布曲线图 (图 16)。

25 由图 16 可知,本方法的扩增较 MDA 扩增方法的均一性更好,更接近于常规测序方法的测序深度分布。

工业应用

本发明在 LFR 文库构建的原理基础上,取代原 DNA 扩增使用的 MDA 扩增方式以及接头 A 的连接方式,同时使用 wafergen 微量移液平台的芯片取代 384 孔板进行试验。本发明使用转座酶插入特定序列,并利用该特定

序列对每个单孔中的 DNA 长片段进行 PCR 扩增，扩增过程中为常规的“变性-退火-延伸”的方式，避免了 MDA 扩增形成复杂结构以及非特异扩增的问题。本发明中创造性地将两个独立的标签序列设计于双链接头的两个单链之中，然后以单链的形式分别加入，在连接的时候同时将两个标签序列
5 连接到插入片段的两端。由于连接反应在常温下进行，本发明设计的引物在常温反应体系中会先进行退火，然后在连接酶的作用下，与插入片段的两个末端进行连接。在连接反应之前先将打断的插入片段 3' 末端延伸一个 dATP，在连接反应中通过末端 A-T 配对的方式使接头的两个单链在插入片段的两端进行定向连接。

10 同时，本发明与 wafergen 微量移液平台相结合，将 DNA 长片段分离进入带有 5184 个细孔的芯片之中，单样品分离孔的数量为 384 孔板的 10 倍以上，分离效果更充分，由此获得更好的突变位点定位及组装效果。

本发明的实验证明，本发明通过与微孔芯片相结合的方式，提高了基因组中同源的长片段的分离概率，与 Complete Genomics 公司的 384 孔板
15 分离方式相比，芯片中带有的 5184 孔分离效果更好。384 孔板的分离，约有 5% 的同源长片段会被分入同一孔中而无法区分，而通过 5184 孔芯片的分离，该概率将进一步下降从而提升分析效率。

本发明使用长片段 PCR 的方式，对分离后每一个孔中的 DNA 片段进行扩增，取代了原方法中使用 MDA 扩增，因此避免了因 MDA 扩增导致的局部
20 区域形成多级复杂结构的情况。在数据中，这些复杂结构所产生的数据因为无法被识别而去除，限制了所得的有效数据比例。使用长片段 PCR 扩增的方式，获得的数据比对效率更好。

接头连接过程中采用两接头单链分别带有标签序列并分别加入的方式，在连接过程中边进行退火边实现接头连接。相比较原方法中先进行接头连接-延伸-连接另一端接头的做法，本方法一步反应即可实现两端不同
25 接头的连接并加入两端不同标签序列的组合，大大减少了操作的复杂度及操作时间。

权利要求

1、一种长片段 DNA 文库构建方法，包括如下步骤：

1) 对长片段 DNA 依次进行转座酶断裂、引入 dUTP 扩增和去除 dUTP，
5 得到断裂片段；

2) 将带有不同标签的测序接头单链 A 和与其部分互补的带有不同标签的测序接头单链 B 以单链的形式分别加入含有所述断裂片段的体系中反应，使所述断裂片段两端连接测序接头，通过测序接头单链 A 和测序接头单链 B 中标签序列的排列组合，使每份所述断裂片段对应的测序接头均相互区别，得到连接不同测序接头的产物；
10

所述带有不同标签的测序接头单链 A 和所述带有不同标签的测序接头单链 B 退火可成所述测序接头；

3) 以所述连接测序接头后的产物为模板，以与所述测序接头匹配的 DNA 为引物，进行 PCR 扩增，得到的 PCR 扩增产物为连接不同测序接头的 PCR 扩增产物；
15

4) 用所述连接不同测序接头的 PCR 扩增产物构建文库，即得到长片段 DNA 文库。

2、根据权利要求 1 所述的方法，其特征在于：

所述步骤 1) 的方法包括如下步骤：

(1) 用转座酶将所述长片段 DNA 进行断裂、并在断裂后的片段两端连上扩增接头，得到连接扩增接头的产物；(2) 以所述连接扩增接头的产物为模板，以与所述扩增接头匹配的 DNA 为引物，进行第一次 PCR 扩增，得到第一次 PCR 扩增产物；所述第一次 PCR 扩增过程中引入 dUTP；
20

(3) 去除所述第一次 PCR 扩增产物中的 dUTP，得到缺口，进行缺口
25 平移，并在 3' 末端加 A，得到的产物为 1) 中所述断裂片段。

3、根据权利要求 1 或 2 所述的方法，其特征在于：

步骤 1) 中，所述转座酶为包埋扩增接头的转座酶；

和/或，所述扩增接头为 1 种或 2 种，所述扩增接头由转座酶识别单链 DNA 分子和与其部分反向互补的单链 DNA 分子形成；

30 和/或，相邻两个所述包埋扩增接头的转座酶的作用位置之间的片段

大小为 3-10kb。

4、根据权利要求 2 或 3 所述的方法，其特征在于：

所述扩增接头为接头 1 和接头 2，所述接头 1 为由转座酶识别单链 DNA 分子 A 和与其部分反向互补的单链 DNA 分子 B 组成，所述接头 2 为由所述
5 转座酶识别单链 DNA 分子 A 和与其部分反向互补的单链 DNA 分子 C 组成；

(2)中与所述扩增接头匹配 DNA 作为的引物由引物 B 和引物 C 组成，
所述引物 B 为所述单链 DNA 分子 B 中除去与所述转座酶识别单链 DNA 分子 A 互补部分外的其余序列；所述引物 C 为所述单链 DNA 分子 C 中除去与所
述转座酶识别单链 DNA 分子 A 互补部分外的其余序列；

10 和/或，去除所述第一次 PCR 扩增产物中的 dUTP 的方法包括如下：用
尿嘧啶 DNA 糖基化酶和人源脱嘌呤脱嘧啶核酸内切酶对所述第一次 PCR 扩
增产物进行酶切反应，得到酶切产物；再用聚合酶 I、Taq 聚合酶和 dATP
对所述酶切产物进行聚合反应，得到 300-1200bp 大小的 DNA 片段。

5、根据权利要求 1-4 中任一所述的方法，其特征在于：

15 所述标签序列是由 n 个碱基经过排列组合得到，所述碱基为 A、G、C、
T 中至少一种，n 大于等于 8。

6、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，所述带有不同标签的测序接头单链 A 从 5' 至 3' 方向
依次包括片段甲、片段乙、标签序列和片段丙；

20 所述带有不同标签的测序接头单链 B 从 5' 至 3' 方向依次包括与所
述片段丙反向互补的片段、标签序列、片段丁和与所述片段甲反向互补的
片段；

所述步骤 3) 中，所述引物中的一条引物与所述测序接头单链 A 中的
片段乙匹配，所述引物中的另一条引物与单链 B 中的片段丁匹配。

25 7、根据权利要求 1-6 中任一所述的方法，其特征在于：

所述带有不同标签的测序接头单链 A 的条数为小于等于 72，且各条
单链标签序列不同；

所述带有不同标签的测序接头单链 B 的条数为小于等于 72，且各条
单链标签序列不同；

30 所述标签序列中 n 大于等于 8 且小于 15。

8、根据权利要求 1-7 中任一所述的方法，其特征在于：

步骤 2) 中，所述将带有标签的测序接头单链 A 和与其部分互补的带有不同标签的测序接头单链 B 以单链的形式分别加入含有所述断裂片段的体系中反应包括如下步骤：

5 2) -A，将 72 条所述带有不同标签序列的测序接头单链 A 分别加入含有所述断裂片段体系的芯片的纵向 72 列中；

2) -B，将 72 条所述带有不同标签序列的测序接头单链 B 分别加入 2) -A 处理后的芯片的横向 72 行中，反应，得到连接测序接头产物。

9、根据权利要求 1-8 中任一所述的方法，其特征在于：

10 步骤 3) 中，所述引物中的一条引物与所述测序接头单链 A 中的片段乙相同或反向互补；

所述引物中的另一条引物与所述测序接头单链 B 中的片段丁的片段互补反向互补或相同。

10、根据权利要求 2-9 中任一所述的方法，其特征在于：

15 所述方法的步骤 1) 的 (2)、(3) 和步骤 2) 均在 5184 孔芯片中进行。

11、根据权利要求 1-10 中任一所述的方法，其特征在于：

所述长片段 DNA 为大于 100Kb 的片段，具体为大小为 400kb 的片段；
所述转座酶为 Tn5 转座酶。

20 12、根据权利要求 2-11 中任一所述的方法，其特征在于：

所述方法的步骤 1) 的 (2)、(3) 和步骤 2) 中均采用 wafergen MSND 移液平台将各种物质加入芯片的孔中。

25 13、根据权利要求 12 所述的方法，其特征在于：在所述用 wafergen MSND 移液平台将各种物质加入芯片的孔中后，还包括如下步骤：离心沉降各物质。

14、根据权利要求 2-11 中任一所述的方法，其特征在于：

步骤 1) 的 (1) 和 (2) 中，还包括如下步骤：用变性试剂消化所述转座酶；所述变性试剂具体采用 0.1-1%SDS 溶液；所述消化的条件具体为 25℃ 静置 10 分钟。

30 15、根据权利要求 4-14 中任一所述的方法，其特征在于：

- 步骤 1) 的 (1) 中, 所述转座酶断裂反应的条件为 55℃ 5 分钟;
- 步骤 1) 的 (2) 中, 所述扩增的退火温度为 68℃, 退火时间为 18min;
- 步骤 1) 的 (3) 中, 所述酶切反应条件为 37℃ 2 小时, 再 65℃ 15 分钟; 所述聚合反应条件为 23℃ 1 小时, 65℃ 30 分钟;
- 5 步骤 2) 的 2) -B 中, 所述反应条件为 20℃ 2 小时;
- 步骤 3) 中, 所述 PCR 扩增的退火温度为 68℃, 退火时间为 10min。
- 16、由权利要求 1-15 所述方法制备得到的长片段 DNA 文库。
- 17、一种用于长片段 DNA 文库构建的连接测序接头产物制备方法, 包括如下步骤: 为权利要求 1-16 中任一所述方法中的步骤 1) -2) 。
- 10 18、权利要求 17 所述的方法在构建长片段 DNA 文库中的应用。

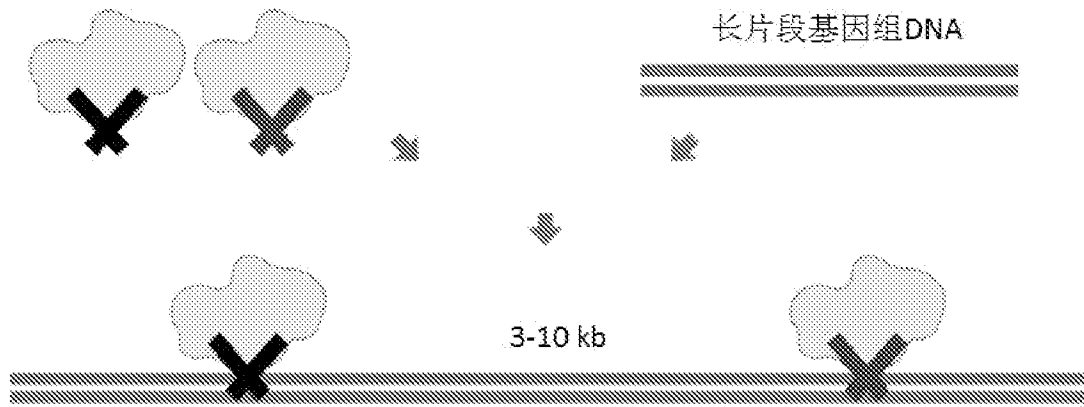


图 1

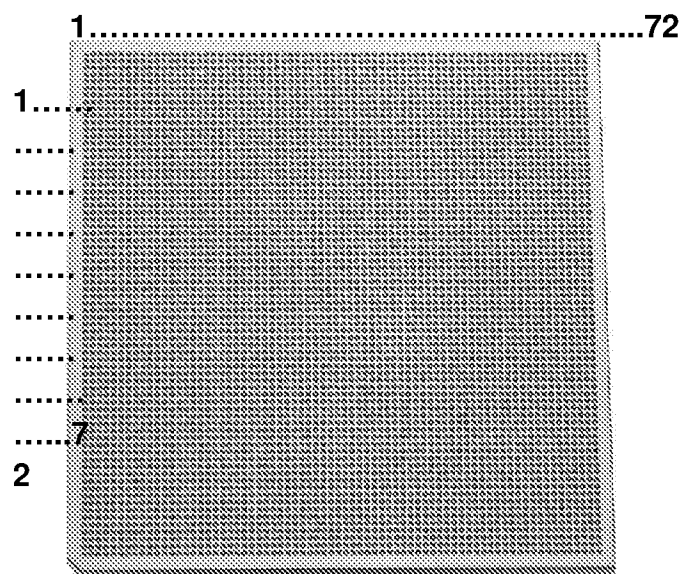


图 2

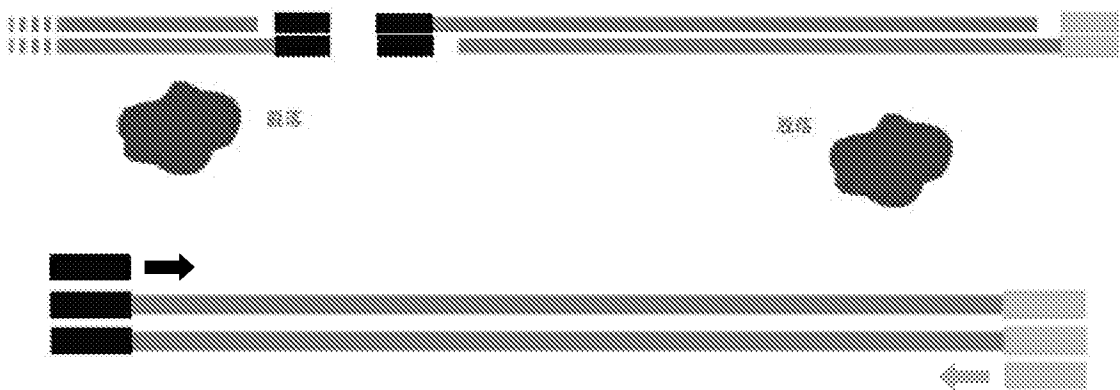


图 3



图 4



图 5

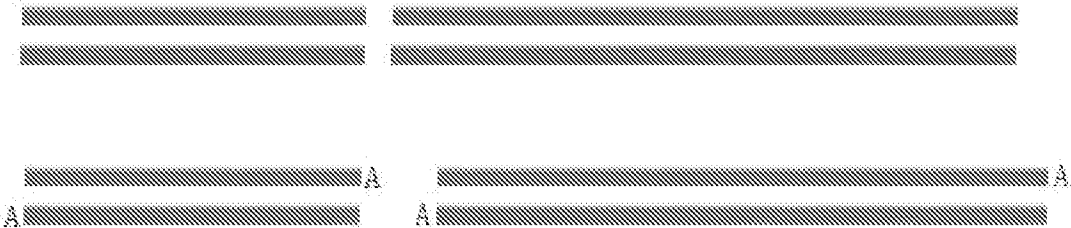
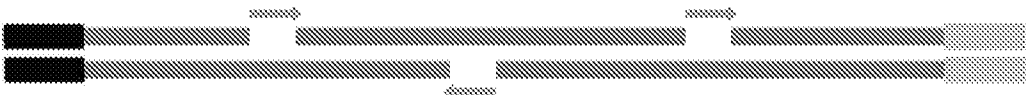


图 6

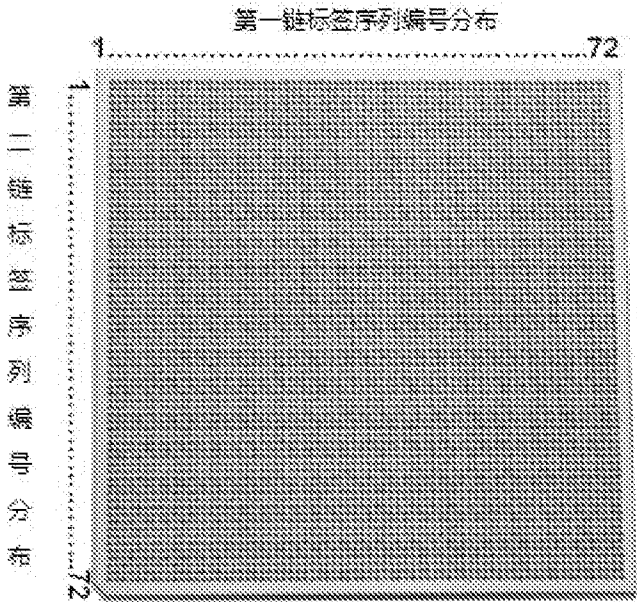


图 7



图 8

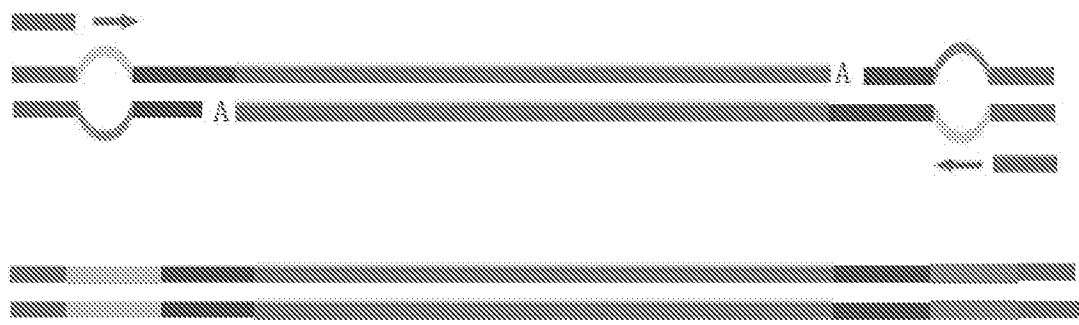
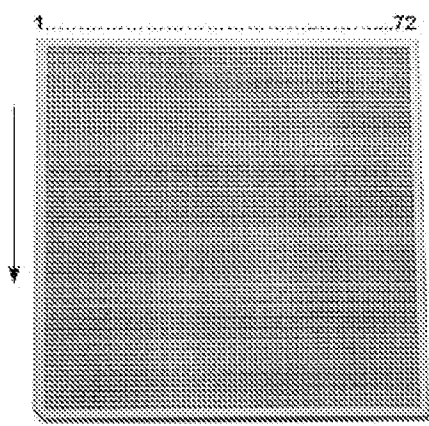
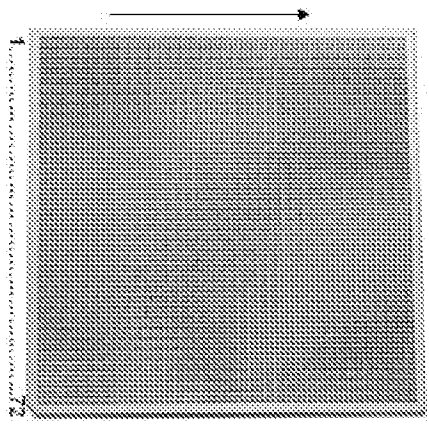


图 9



72 样品 35nL 分液流程



72 样品 50nL 分液流程

图 10

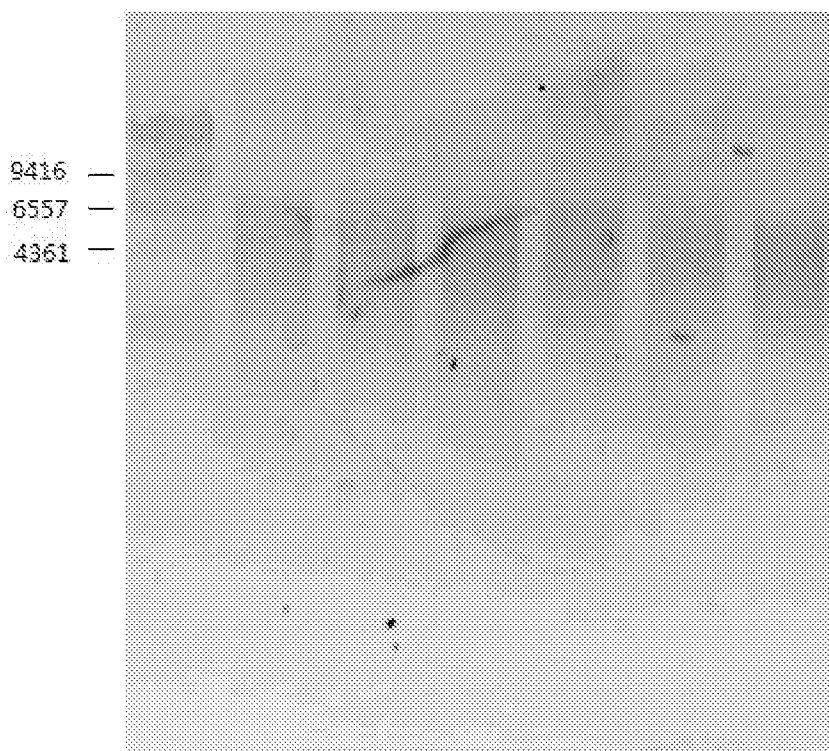


图 11

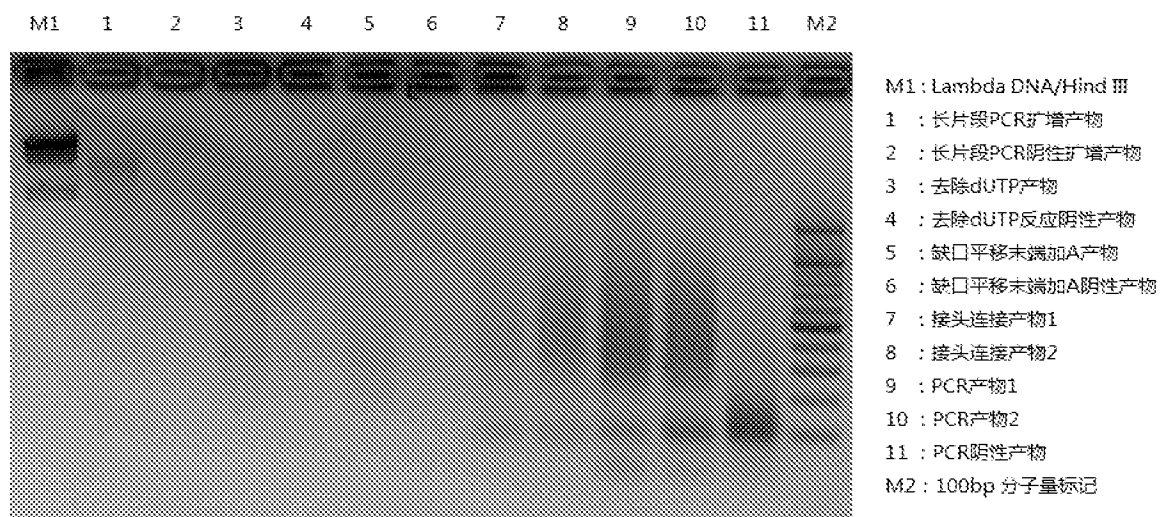


图 12

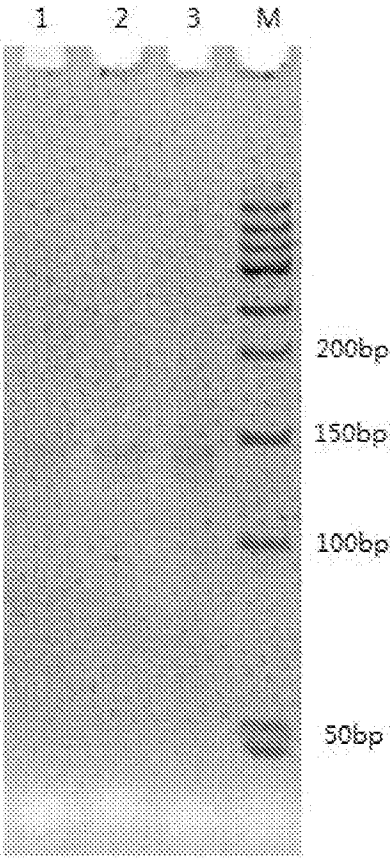


图 13

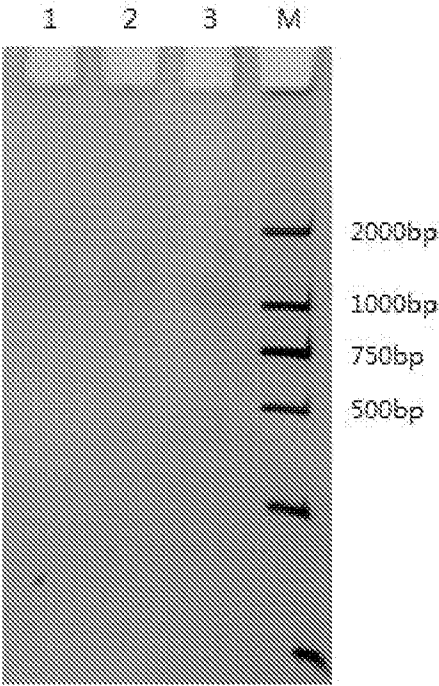


图 14

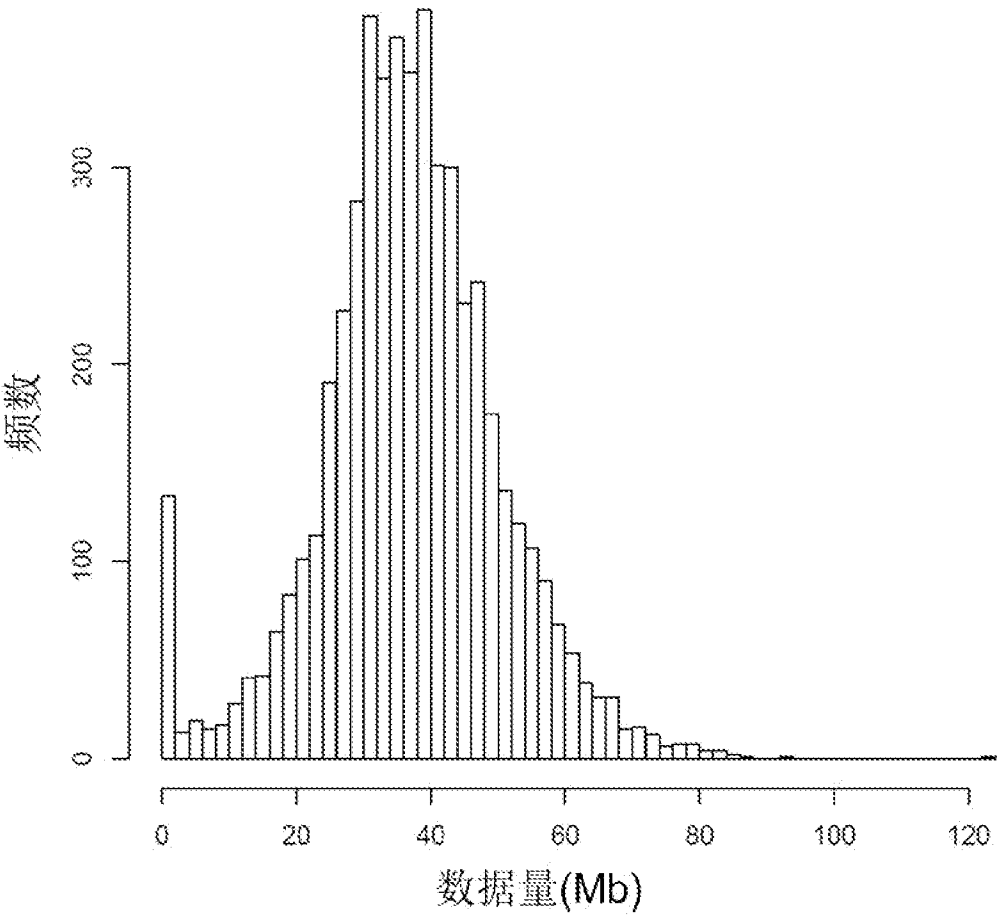


图 15

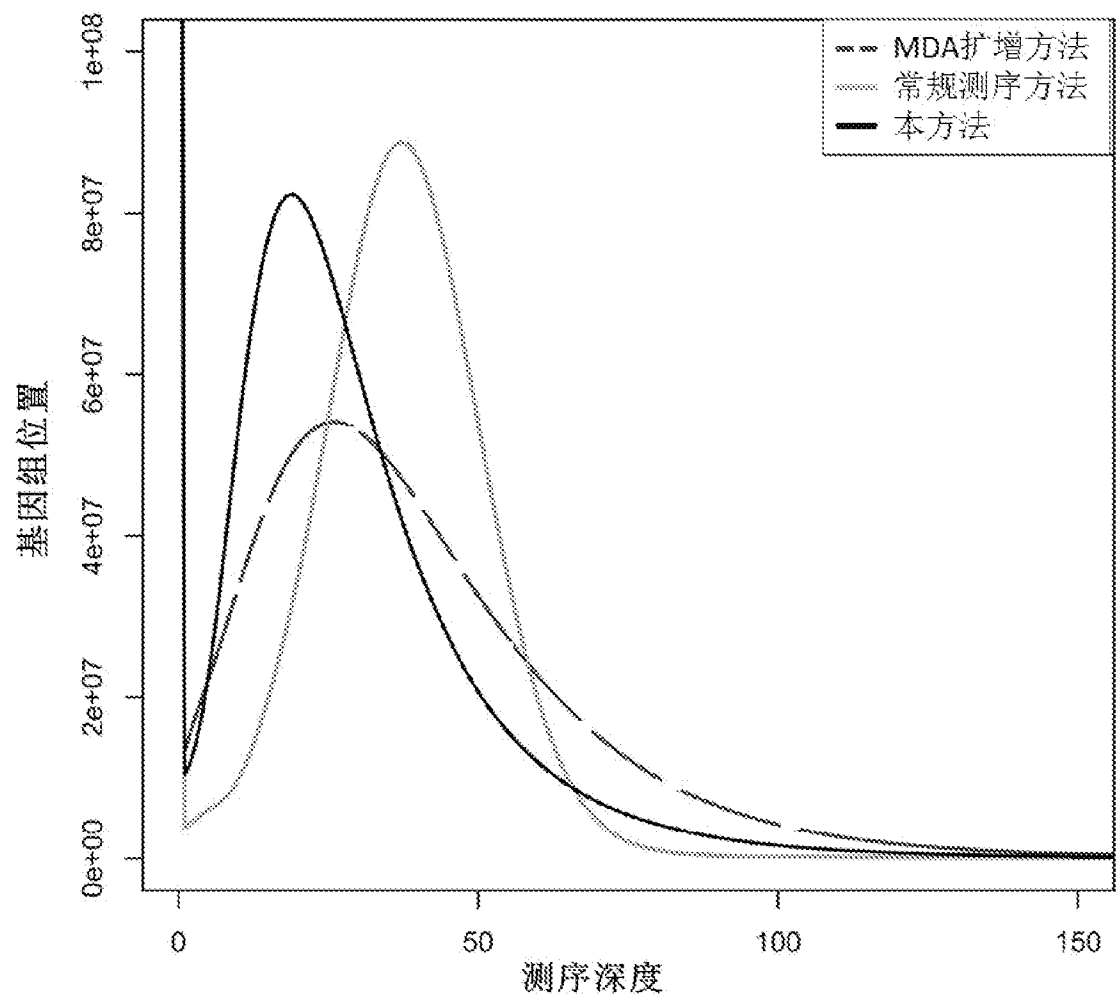


图 16

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/079278

A. CLASSIFICATION OF SUBJECT MATTER

C40B 50/06 (2006.01) i; C40B 40/06 (2006.01) i; C12N 15/10 (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)

C40B; C12N

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; CNKI; DWPI; ISI Web of knowledge and keywords: polymerase chain reaction, transposase, DNA, transferase, library, dutp, PCR, etc.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 102703426 A (WUJIANG HUIJIE BIOTECHNOLOGY CO., LTD.), 03 October 2012 (03.10.2012) see the whole document	1-18
A	CN 102264914 A (EPICENTRE TECHNOLOGIES CORP.), 30 November 2011 (30.11.2011) see the whole document	1-18
A	CN 103443338 A (UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION), 11 December 2013 (11.12.2013) see the whole document	1-18

☐ 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 13 July 2016 (13.07.2016)	Date of mailing of the international search report 20 July 2016 (20.07.2016)
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 JIA, Tao Telephone No.: (86-10) 62411993

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2016/079278

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102703426 A	03 October 2012	None	
CN 102264914 A	30 November 2011	EP 2787565 B1	16 September 2015
		EP 2376517 B1	16 January 2013
		DK 2376517 T3	11 February 2013
		ES 2515065 T3	29 October 2014
		PT 2787565 E	17 November 2015
		EP 2963709 A1	06 January 2016
		EP 2508529 A1	10 October 2012
		EP 2664678 B1	08 October 2014
		JP 2012506704 A	22 March 2012
		DK 2787565 T3	12 October 2015
		EP 2376517 A4	27 June 2012
		CN 102264914 B	12 August 2015
		EP 2787565 A1	08 October 2014
		ES 2403756 T3	21 May 2013
		HK 1163107 A1	19 July 2013
		WO 2010048605 A1	29 April 2010
		EP 2376517 A1	19 October 2011
		HK 1169659 A1	29 November 2013
		CN 105018548 A	04 November 2015
		EP 2508529 B1	28 August 2013
		ES 2550237 T3	05 November 2015
		EP 2664678 A1	20 November 2013
		CA 2750054 A1	29 April 2010
		JP 5794572 B2	14 October 2015
CN 103443338 A	11 December 2013	EP 2670894 A2	11 December 2013
		EP 2670894 A4	22 October 2014
		WO 2012106546 A3	14 November 2013
		WO 2012106546 A2	09 August 2012
		US 2013203605 A1	08 August 2013
		AU 2012212148 A1	15 August 2013
		JP 2014506788 A	20 March 2014
		CA 2826131 A1	09 August 2012

国际检索报告

国际申请号

PCT/CN2016/079278

A. 主题的分类

C40B 50/06(2006.01)i; C40B 40/06(2006.01)i; C12N 15/10(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)

C40B; C12N

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;CNKI;DWPI;ISI Web of knowledge和关键词:转座酶, 文库, 聚合酶链反应, transposase, DNA, transferase, library, dutp, PCR, 等

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 102703426 A (吴江汇杰生物科技有限公司) 2012年 10月 3日 (2012 - 10 - 03) 见全文	1-18
A	CN 102264914 A (阿霖震中科技公司) 2011年 11月 30日 (2011 - 11 - 30) 见全文	1-18
A	CN 103443338 A (华盛顿大学商业化中心) 2013年 12月 11日 (2013 - 12 - 11) 见全文	1-18

☐ 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 7月 13日

国际检索报告邮寄日期

2016年 7月 20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

传真号 (86-10)62019451

授权官员

贾涛

电话号码 (86-10)62411993

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/079278

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	102703426	A	2012年 10月 3日	无			
CN	102264914	A	2011年 11月 30日	EP	2787565	B1	2015年 9月 16日
				EP	2376517	B1	2013年 1月 16日
				DK	2376517	T3	2013年 2月 11日
				ES	2515065	T3	2014年 10月 29日
				PT	2787565	E	2015年 11月 17日
				EP	2963709	A1	2016年 1月 6日
				EP	2508529	A1	2012年 10月 10日
				EP	2664678	B1	2014年 10月 8日
				JP	2012506704	A	2012年 3月 22日
				DK	2787565	T3	2015年 10月 12日
				EP	2376517	A4	2012年 6月 27日
				CN	102264914	B	2015年 8月 12日
				EP	2787565	A1	2014年 10月 8日
				ES	2403756	T3	2013年 5月 21日
				HK	1163107	A1	2013年 7月 19日
				WO	2010048605	A1	2010年 4月 29日
				EP	2376517	A1	2011年 10月 19日
				HK	1169659	A1	2013年 11月 29日
				CN	105018548	A	2015年 11月 4日
				EP	2508529	B1	2013年 8月 28日
				ES	2550237	T3	2015年 11月 5日
				EP	2664678	A1	2013年 11月 20日
				CA	2750054	A1	2010年 4月 29日
				JP	5794572	B2	2015年 10月 14日
CN	103443338	A	2013年 12月 11日	EP	2670894	A2	2013年 12月 11日
				EP	2670894	A4	2014年 10月 22日
				WO	2012106546	A3	2013年 11月 14日
				WO	2012106546	A2	2012年 8月 9日
				US	2013203605	A1	2013年 8月 8日
				AU	2012212148	A1	2013年 8月 15日
				JP	2014506788	A	2014年 3月 20日
				CA	2826131	A1	2012年 8月 9日