

(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
2021 年 12 月 9 日 (09.12.2021)



(10) 国际公布号
WO 2021/244557 A1

(51) 国际专利分类号:
C12Q 1/6806 (2018.01) **C12Q 1/6869** (2018.01)

(21) 国际申请号: PCT/CN2021/097800

(22) 国际申请日: 2021 年 6 月 2 日 (02.06.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202010495334.6 2020 年 6 月 3 日 (03.06.2020) CN
202010506791.0 2020 年 6 月 5 日 (05.06.2020) CN

(71) 申请人: 中国海洋大学(OCEAN UNIVERSITY OF CHINA) [CN/CN]; 中国山东省青岛市崂山区松岭路 238 号, Shandong 266100 (CN)。

(72) 发明人: 施威扬(SHI, Weiyang); 中国山东省青岛市崂山区松岭路 238 号, Shandong 266100 (CN)。

(74) 代理人: 上海巛石知识产权代理事务所(普通合伙)(SHANGHAI DIANSHI PARTNERS, P.C.); 中国上海市浦东新区盛夏路 608 号 2 号楼 103 室/蒋舫玮, Shanghai 201210 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL,

PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, 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, 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)。

根据细则 4.17 的声明:

— 发明人资格(细则 4.17(iv))

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 ANALYZING TARGET NUCLEIC ACID FROM CELL

(54) 发明名称: 分析来自细胞的目标核酸的方法

(57) Abstract: The present application relates to a method for analyzing target nucleic acid from a cell, comprising: 1) providing the following discrete partition: target nucleic acid derived from a single cell and added with an oligonucleotide adaptor sequence, and a solid support with at least one oligonucleotide tag attached, wherein each oligonucleotide tag comprises a first strand and a second strand, the first strand comprises a barcode sequence and a hybridization sequence located at the 3'-end of the barcode sequence, the second strand comprises a first portion complementary to the hybridization sequence of the first strand and a second portion complementary to the oligonucleotide adaptor sequence attached to the target nucleic acid, and the first strand and the second strand form a partially double-stranded structure, or the second strand and the target nucleic acid attached form a partially double-stranded structure; and 2) in the discrete partition, the oligonucleotide tag is linked to the target nucleic acid attached, thereby producing barcoded target nucleic acid.

(57) 摘要: 本申请涉及一种分析来自细胞的目标核酸的方法, 包括 1) 提供如下离散分区: 源于单个细胞并加了寡核苷酸衔接子序列的目标核酸, 以及附接有至少一个寡核苷酸标签的固体支持物, 其中每个寡核苷酸标签包含第一链以及第二链, 第一链包含条码序列以及位于条码序列 3' 端的杂交序列, 第二链包含与第一链的杂交序列互补的第一部分以及与附接至目标核酸的寡核苷酸衔接子序列互补的第二部分, 且第一链与第二链形成部分双链的结构或者所述第二链与所述经附接的目标核酸形成部分双链的结构; 2) 在离散分区中, 使寡核苷酸标签与经附接的目标核酸连接, 从而产生条码化的目标核酸。

WO 2021/244557 A1

分析来自细胞的目标核酸的方法

技术领域

本申请涉及生物医药领域，具体的涉及一种分析来自细胞的目标核酸的方法及相关制剂。

背景技术

目前，核酸测序技术已经经历了快速和巨大的进步，测序技术产生大量的序列数据，可用于对基因组和基因组区域的研究阐释，并且提供广泛应用于常规生物学研究和诊断信息。基因组测序可以用来获得各种各样的生物学背景的信息，包括诊断学、预后、生物技术和法医生物学。测序包括 Maxam-Gilbert 测序和链终止法或从头测序法（包括鸟枪法测序和桥式 PCR），或新一代方法，其包括聚合酶克隆测序、454 焦磷酸测序、Illumina 测序、SOLiD 测序、Ion Torrent 半导体测序、HeliScope 单分子测序、[image] 测序，等等。对于大多数测序应用，样品如核酸样品在引入测序仪之前先进行处理。

传统对于基因组或转录组表达研究方法，通常是在多细胞水平进行的。因此，最终得到的信号值是多个细胞的平均，丢失了细胞异质性的信息。例如，目前通过直接测序分析细胞的 mRNA 含量时依靠对从包含数百万个细胞的组织样品中获得的大量 mRNA 进行分析，这意味着当在大量 mRNA 中分析基因表达时，单细胞中呈现的很多功能信息将丢失或变得模糊；此外，也不能按总体平均值观察诸如细胞周期的动态过程。类似地，只有单独分析细胞，才能研究复杂组织（例如，大脑）中的某些细胞类型。

目前没有合适的细胞表面标志物来用于分离单细胞以进行研究，并且即使存在合适的细胞表面标志物，少量的单细胞仍不足以捕获基因表达中自然变异的范围。因此需要一种分析方法可用于分析大量的单细胞中的基因信息。

发明内容

本申请提供了一种分析来自细胞的目标核酸的方法，所述方法包括：

a) 提供包含下述的离散分区：

i. 源于单个细胞的目标核酸，其中至少部分所述目标核酸被添加了寡核苷酸衔接子序列而成为经附接的目标核酸；以及

ii. 附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3'端的杂交序列，所述第二链

包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第一链与所述第二链形成部分双链的结构或者所述第二链与所述经附接的目标核酸形成部分双链的结构；

b) 在所述离散分区中，使所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸。

在某些实施方式中，所述寡核苷酸标签可释放地附接至所述固体支持物。

在某些实施方式中，包括从所述固体支持物上释放所述至少一个寡核苷酸标签，并在 b) 中使经释放的所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸。

在某些实施方式中，所述寡核苷酸标签通过其第一链的 5' 端直接或间接附接至所述固体支持物。

在某些实施方式中，所述离散分区中还包含连接酶，且所述连接酶使所述寡核苷酸标签与所述经附接的目标核酸连接。

在某些实施方式中，所述连接酶包括 T4 连接酶。

在某些实施方式中，在所述条码化的目标核酸中，所述目标核酸序列位于所述条码序列的 3' 端。

在某些实施方式中，所述固体支持物为珠粒。

在某些实施方式中，所述珠粒为磁性珠粒。

在某些实施方式中，所述离散分区为孔或微滴。

在某些实施方式中，所述条码序列包含细胞条码序列，且附接至同一个固体支持物上的各寡核苷酸标签所包含的细胞条码序列相同。

在某些实施方式中，所述细胞条码序列包含由连接子序列间隔开的至少 2 个细胞条码区段。

在某些实施方式中，a) 包括将所述源于单个细胞的目标核酸与所述附接有至少一个寡核苷酸标签的固体支持物共分配至所述离散分区中。

在某些实施方式中，b) 包括使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。

在某些实施方式中，b) 包括使所述寡核苷酸标签的第二链的所述第二部分与附接至所述目标核酸的所述寡核苷酸衔接子杂交，以及使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。

在某些实施方式中，所述经附接的目标核酸中包含独特分子鉴别区。

在某些实施方式中，所述独特分子鉴别区位于所述寡核苷酸衔接子序列与所述目标核酸序列之间。

在某些实施方式中，所述寡核苷酸标签还包含扩增引物识别区。

在某些实施方式中，所述扩增引物识别区为通用扩增引物识别区。

在某些实施方式中，所述方法进一步包括：

c) 获得所述条码化的目标核酸的表征结果；以及

d) 至少部分基于 c) 中获得的所述表征结果中存在相同的所述细胞条码序列而将所述目标核酸的序列识别为源于所述单个细胞。

在某些实施方式中，所述方法进一步包括，在 b) 之后并且在 c) 之前，从所述离散分区中释放所述条码化的目标核酸。

在某些实施方式中，c) 包括对所述条码化的目标核酸进行测序，从而获得所述表征结果。

在某些实施方式中，所述方法进一步包括由所述条码化的目标核酸的序列组装所述单个细胞的基因组的至少一部分的连续核酸序列。

在某些实施方式中，基于所述单个细胞的所述基因组的至少一部分的所述核酸序列来表征所述单个细胞。

在某些实施方式中，每个所述离散分区至多包括源自单个细胞的所述目标核酸。

在某些实施方式中，所述方法进一步包括至少部分基于所述独特分子鉴别区的存在将所述条码化的目标核酸中的单个核酸序列鉴别为源于所述目标核酸中的给定核酸。

在某些实施方式中，所述目标核酸包括外源核酸，所述外源核酸包括与蛋白、脂类和/或小分子化合物连接的外源核酸，所述蛋白、脂类和/或小分子化合物能够与细胞内的靶分子结合。

在某些实施方式中，所述方法进一步包括基于所述独特分子鉴别区的存在确定所述目标核酸中给定核酸的量。

在某些实施方式中，包括在 a) 之前对所述细胞进行预处理。

在某些实施方式中，所述预处理包括固定所述细胞。

在某些实施方式中，使用固定剂对所述细胞进行固定，所述固定剂选自下组中的一种或多种：甲醛、多聚甲醛、甲醇、乙醇、丙酮、戊二醛、钨酸和重铬酸钾。

在某些实施方式中，所述预处理包括使所述细胞的细胞核被暴露。

在某些实施方式中，所述预处理包括使用去垢剂处理所述细胞，所述去垢剂包括 Triton、Tween, SDS, NP-40 和/或 digitonin。

在某些实施方式中，所述目标核酸包括选自下组的一种或多种：DNA、RNA 和 cDNA。

在某些实施方式中，其进一步包括，在 b) 之后并且在 c) 之前，对所述条码化的目标核酸进行扩增。

在某些实施方式中，包括在 b) 之后并且在 c) 之前，从所述离散分区中释放所述条码化的目标核酸，且所述扩增在所述条码化的目标核酸从所述离散分区中释放后进行。

在某些实施方式中，所述扩增中使用扩增引物，且所述扩增引物中包含随机引导序列。

在某些实施方式中，所述随机引导序列为随机六聚体。

在某些实施方式中，所述扩增包括使所述随机引导序列与所述条码化的目标核酸至少部分杂交并且以模板定向的方式延伸所述随机引导序列。

在某些实施方式中，包括使至少一部分所述目标核酸从所述离散分区中的所述单个细胞中释放到细胞外，并在 b) 中使经释放的所述目标核酸与所述寡核苷酸标签连接，从而产生条码化的目标核酸。

在某些实施方式中，包括使至少一部分从所述固体支持物释放的所述寡核苷酸标签进入所述单个细胞中，并在 b) 中与所述目标核酸连接，从而产生条码化的目标核酸。

在某些实施方式中，包括使用微流控装置将所述源于单个细胞的目标核酸与所述附接有至少一个寡核苷酸标签的固体支持物共分配至所述离散分区中。

在某些实施方式中，所述离散分区为微滴，且所述微流控装置为微滴发生器。

在某些实施方式中，所述微流控装置包括第一输入通道和第二输入通道，它们在与输出通道流体连接的接合处汇合。

在某些实施方式中，所述方法还包括将包含所述目标核酸的样品引入所述第一输入通道，且将附接有至少一个寡核苷酸标签的所述固体支持物引入所述第二输入通道，从而在所述输出通道中生成所述样品与所述固体支持物的混合物。

在某些实施方式中，所述输出通道与第三输入通道在接合处流体连接。

在某些实施方式中，还包括将油引入所述第三输入通道，使得形成油包水乳液内的水性小滴作为所述离散分区。

在某些实施方式中，每个所述离散分区中至多包含来自单个细胞的所述目标核酸。

在某些实施方式中，所述第一输入通道和所述第二输入通道彼此之间形成基本上垂直的角度。

在某些实施方式中，所述目标核酸包括源自所述单个细胞中 RNA 的 cDNA。

在某些实施方式中，所述 RNA 包括 mRNA。

在某些实施方式中，包括在 a) 之前对所述 RNA 进行反转录，并产生所述经附接的目标核酸。

在某些实施方式中，述反转录中使用反转录引物，所述反转录引物以 5' 至 3' 的方向包含所述寡核苷酸衔接子序列以及 polyT 序列。

在某些实施方式中，所述反转录包括使所述 polyT 序列与所述 RNA 杂交并且以模板定向的方式延伸所述 polyT 序列。

在某些实施方式中，所述目标核酸包括源自所述单个细胞的 DNA。

在某些实施方式中，所述 DNA 包括基因组 DNA、开放染色质 DNA、蛋白质结合的 DNA 区域和/或与蛋白、脂类和/或小分子化合物连接的外源核酸，所述蛋白、脂类和/或小分子化合物能够与细胞内的靶分子结合。

在某些实施方式中，包括在 a) 之前对源自单个细胞的所述 DNA 进行片段化。

在某些实施方式中，在所述片段化之后或者在所述片段化的过程中产生所述经附接的目标核酸。

在某些实施方式中，所述片段化包括使用超声断裂，而后在经断裂的所述 DNA 上添加包含所述寡核苷酸衔接子的序列，从而获得所述经附接的目标核酸。

在某些实施方式中，所述片段化包括使用 DNA 内切酶、DNA 外切酶打断，而后在经断裂的所述 DNA 上添加包含所述寡核苷酸衔接子的序列，从而获得所述经附接的目标核酸。

在某些实施方式中，所述片段化包括使用转座酶-核酸复合物将包含所述寡核苷酸衔接子的序列整合到所述 DNA 中，并释放所述转座酶以获得所述经附接的目标核酸。

在某些实施方式中，所述转座酶-核酸复合物包含转座酶以及转座子末端核酸分子，其中所述转座子末端核酸分子包含所述寡核苷酸衔接子序列。

在某些实施方式中，所述转座酶包括 Tn5。

在某些实施方式中，所述 DNA 包括与蛋白质结合的 DNA 区域，且所述转座酶-核酸复合物中还包含直接或间接识别所述蛋白质的部分。

在某些实施方式中，所述直接或间接识别所述蛋白质的部分包括下组中的一种或多种：特异性结合所述蛋白质的抗体和蛋白质 A 或蛋白质 G。

另一方面，本申请还提供了一种组合物，其包含：多个固体支持物，每个所述固体支持物上附接有至少一个寡核苷酸标签，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3' 端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与所述待测核酸中的序列互补的第二部分，且所述第一

链与所述第二链形成部分双链的结构或者所述第二链与所述经附接的目标核酸形成部分双链的结构；所述寡核苷酸标签的条码序列包含共同条码结构域和可变结构域，所述共同条码结构域在附接于同一个固体支持物的寡核苷酸标签中是相同的，且所述共同条码结构域在所述多个固体支持物中的两个或更多个固体支持物之间是不同的。

另一方面，本申请还提供了用于分析来自细胞的目标核酸的试剂盒，其包含本申请所述的组合物。

在某些实施方式中，所述试剂盒包括转座酶。

在某些实施方式中，所述试剂盒进一步包含核酸扩增剂，逆转录剂，固定剂，通透剂，连接剂和裂解剂中的至少一种。

一种扩增来自细胞的目标核酸的方法，所述方法包括：

a) 提供包含下述的离散分区：i. 源于单个细胞的目标核酸，其中至少部分所述目标核酸被添加了寡核苷酸衔接子序列而成为经附接的目标核酸；以及 ii. 附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3' 端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第一链与所述第二链形成部分双链的结构或者所述第二链与所述经附接的目标核酸形成部分双链的结构；

b) 在所述离散分区中，使所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸；以及

c) 对所述条码化的目标核酸进行扩增。

在某些实施方式中，所述寡核苷酸标签可释放地附接至所述固体支持物。

在某些实施方式中，包括从所述固体支持物上释放所述至少一个寡核苷酸标签，并在 b) 中使经释放的所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸。

在某些实施方式中，所述寡核苷酸标签通过其第一链的 5' 端直接或间接附接至所述固体支持物。

在某些实施方式中，所述离散分区中还包含连接酶，且所述连接酶使所述寡核苷酸标签与所述经附接的目标核酸连接。

在某些实施方式中，所述连接酶包括 T4 连接酶。

在某些实施方式中，在所述条码化的目标核酸中，所述目标核酸序列位于所述条码序列的 3' 端。

在某些实施方式中，所述固体支持物为珠粒。

在某些实施方式中，所述离散分区为孔或微滴。

在某些实施方式中，所述条码序列包含细胞条码序列，且附接至同一个固体支持物上的各寡核苷酸标签所包含的细胞条码序列相同。

在某些实施方式中，所述细胞条码序列包含由连接子序列间隔开的至少 2 个细胞条码区段。

在某些实施方式中，a) 包括将所述源于单个细胞的目标核酸与所述附接有至少一个寡核苷酸标签的固体支持物共分配至所述离散分区中。

在某些实施方式中，b) 包括使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。

在某些实施方式中，b) 包括使所述寡核苷酸标签的第二链的所述第二部分与附接至所述目标核酸的所述寡核苷酸衔接子杂交，以及使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。

在某些实施方式中，所述经附接的目标核酸中包含独特分子鉴别区。

在某些实施方式中，所述独特分子鉴别区位于所述寡核苷酸衔接子序列与所述目标核酸序列之间。

在某些实施方式中，所述寡核苷酸标签还包含扩增引物识别区。

在某些实施方式中，所述扩增引物识别区为通用扩增引物识别区。

在某些实施方式中，包括在 b) 之后并且在 c) 之前，从所述离散分区中释放所述条码化的目标核酸，且所述扩增在所述条码化的目标核酸从所述离散分区中释放后进行。

在某些实施方式中，所述扩增中使用扩增引物，且所述扩增引物中包含随机引导序列。

在某些实施方式中，所述随机引导序列为随机六聚体。

在某些实施方式中，所述扩增包括使所述随机引导序列与所述条码化的目标核酸至少部分杂交并且以模板定向的方式延伸所述随机引导序列。

另一方面，本申请还提供了一种对来自细胞的目标核酸进行测序的方法，所述方法包括：

a) 提供包含下述的离散分区：i. 源于单个细胞的目标核酸，其中至少部分所述目标核酸被添加了寡核苷酸衔接子序列而成为经附接的目标核酸；以及 ii. 附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3' 端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，

且所述第一链与所述第二链形成部分双链的结构或者所述第二链与所述经附接的目标核酸形成部分双链的结构；

b) 在所述离散分区中，使所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸；以及

c) 对所述条码化的目标核酸进行测序。

在某些实施方式中，所述寡核苷酸标签可释放地附接至所述固体支持物。

在某些实施方式中，包括从所述固体支持物上释放所述至少一个寡核苷酸标签，并在 b) 中使经释放的所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸。

在某些实施方式中，所述寡核苷酸标签通过其第一链的 5' 端直接或间接附接至所述固体支持物。

在某些实施方式中，所述离散分区中还包含连接酶，且所述连接酶使所述寡核苷酸标签与所述经附接的目标核酸连接。

在某些实施方式中，所述连接酶包括 T4 连接酶或 T7 连接酶。

在某些实施方式中，在所述条码化的目标核酸中，所述目标核酸序列位于所述条码序列的 3' 端。

在某些实施方式中，所述固体支持物为珠粒。

在某些实施方式中，所述离散分区为孔或微滴。

在某些实施方式中，所述条码序列包含细胞条码序列，且附接至同一个固体支持物上的各寡核苷酸标签所包含的细胞条码序列相同。

在某些实施方式中，所述细胞条码序列包含由连接子序列间隔开的至少 2 个细胞条码区段。

在某些实施方式中，a) 包括将所述源于单个细胞的目标核酸与所述附接有至少一个寡核苷酸标签的固体支持物共分配至所述离散分区中。

在某些实施方式中，b) 包括使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。

在某些实施方式中，b) 包括使所述寡核苷酸标签的第二链的所述第二部分与附接至所述目标核酸的所述寡核苷酸衔接子杂交，以及使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。

在某些实施方式中，所述经附接的目标核酸中包含独特分子鉴别区。

在某些实施方式中，所述独特分子鉴别区位于所述寡核苷酸衔接子序列与所述目标核酸

序列之间。

在某些实施方式中，所述寡核苷酸标签还包含扩增引物识别区。

在某些实施方式中，所述扩增引物识别区为通用扩增引物识别区。

在某些实施方式中，进一步包括由所述条码化的目标核酸的序列组装所述单个细胞的基因组的至少一部分的连续核酸序列。

在某些实施方式中，基于所述单个细胞的所述基因组的至少一部分的所述核酸序列来表征所述单个细胞。

在某些实施方式中，每个所述离散分区至多包括源自单个细胞的所述目标核酸。

在某些实施方式中，进一步包括至少部分基于所述独特分子鉴别区的存在将所述条码化的目标核酸中的单个核酸序列鉴别为源于所述目标核酸中的给定核酸。

在某些实施方式中，进一步包括基于所述独特分子鉴别区的存在确定所述目标核酸中给定核酸的量。本领域技术人员能够从下文的详细描述中容易地洞察到本申请的其它方面和优势。下文的详细描述中仅显示和描述了本申请的示例性实施方式。如本领域技术人员将认识到的，本申请的内容使得本领域技术人员能够对所公开的具体实施方式进行改动而不脱离本申请所涉及发明的精神和范围。相应地，本申请的附图和说明书中的描述仅仅是示例性的，而非为限制性的。

附图说明

本申请所涉及的发明的具体特征如所附权利要求书所显示。通过参考下文中详细描述示例性实施方式和附图能够更好地理解本申请所涉及发明的特点和优势。对附图简要说明书如下：

图 1 显示了本申请中 PCR 法生成适用于非转录组分析的核苷酸标签的示意图。

图 2 显示了本申请中 T4 连接酶法生成适用于非转录组分析的核苷酸标签的示意图。

图 3 显示了本申请中 PCR 法生成适用于转录组分析的核苷酸标签的示意图。

图 4 显示了本申请中 T4 连接酶法生成适用非转录组分析的核苷酸标签的示意图。

图 5 显示了本申请中 Tn5 转座反应介导的人 293T 细胞 ATAC 测序结果的片段长度分布图。

图 6A 和 6B 显示了本申请中 Tn5 转座反应介导的人 293T 细胞 ATAC 测序结果的信号富集转录起始位点 (TSS) 图。

图 7 显示了本申请中 Tn5 转座反应介导的人 293T 细胞 ATAC 测序结果的不同类型序列比例图。

图 8 显示了本申请中的微流控芯片示意图。

图 9 显示了本申请中 ATAC 测序结果依据每种 barcode 中的 read 数目得到的堆积曲线图。

图 10 显示了本申请中 ATAC 测序结果单个细胞中 unique mapped reads 数目分布图。

图 11 显示了本申请中细胞的 ATAC 数据在基因区域的分布图。

图 12 显示了本申请中单细胞的 ATAC 信号相关性分析结果图。

图 13 显示了本申请中 Cut tag 文库片段分布结果。

图 14 显示了本申请中 Cut tag 片段在转录起始位点分布位置结果图。

图 15 显示了本申请中 Cut tag 片段在基因组中分布的比例结果图。

图 16 显示了本申请中单细胞 Cut tag 结果分布结果。

图 17 显示了本申请中根据单细胞转录组清楚区分混合细胞的单个细胞的结果图。

图 18 显示了本申请中每一细胞中检测的转录本及基因数目分布结果。

图 19 显示了本申请中根据单细胞基因组清楚区分混合细胞的单个细胞的结果图。

图 20 显示了本申请中单细胞测序在每个细胞和每个基因组位点有不同的覆盖程度结果图。

图 21 显示了本申请中根据单细胞 DNA 修饰清楚区分混合细胞的单个细胞的结果图。

图 22 显示了本申请中每一细胞中检测的甲基化修饰分布结果。

图 23 显示了本申请中每一细胞中检测的 5hmC 修饰分布结果。

图 24 显示了本申请中根据转录组和 ATAC 均可以很好地区分混合细胞中的单个细胞的结果图。

图 25 显示了本申请中根据转录组和 cut tag 均可以很好地区分混合细胞中的单个细胞的结果图。

图 26 显示了本申请中同一细胞的转录组和甲基化组均可以很好地与基因模型以及已知的甲基化位点进行匹配的结果图。

图 27 显示了本申请中一种空间点阵芯片示意图。

图 28 显示了本申请中切片 HE 染色与空间点阵芯片叠加后基因数目结果图。

具体实施方式

以下由特定的具体实施例说明本申请发明的实施方式，熟悉此技术的人士可由本说明书所公开的内容容易地了解本申请发明的其他优点及效果。

术语定义

在本申请中，术语“测序”通常是指获取核酸分子序列信息的技术。例如分析特定 DNA 片段的碱基序列（例如，腺嘌呤（A）、胸腺嘧啶（T）、胞嘧啶（C）与鸟嘌呤（G）的排列方式等）；测序方法可以包括 Sanger 双脱氧链终止法（Chain Termination Method），焦磷酸测序法，以及新一代测序的 Illumina，Life Technologies 和 Roche 等使用的“合成并行测序”或“连接测序”平台等，华大智造/Complete Genomics 的测序仪；通常还可以包括纳米孔测序方法，例如牛津纳米孔技术公司开发的方法，PacBio 的三代测序仪，或基于电子检测的方法，例如 Life Technologies 推出的离子激流技术（Ion Torrent technology）等。

在本申请中，术语“表征结果”通常是指通过测序或其他基因组和/或蛋白质组学等生物学分析方法获得的核酸及其他相关分子的信息描述。例如可以包括全基因组测序的序列信息、可接近染色质序列及分布信息、核酸序列与其结合因子的结合信息、致病基因突变信息、单核苷酸多态性（SNP）、核苷酸甲基化、转录组组信息（例如基因表达水平的时间或空间变化）等。

在本申请中，术语“蛋白质 A”通常是指一种细胞来源的可以结合不同物种来源的抗体重链保守区的蛋白（即抗体的识别蛋白）。例如，能与人及多种哺乳动物血清 IgG 分子中的 Fc 片段结合，其中的哺乳动物可以包括猪、狗、兔、人、猴、鼠、小鼠及牛等；蛋白质 A 与 IgG 结合的亚类主要可以包括 IgG1、IgG2 和 IgG4；蛋白质 A 除了与 IgG 结合外，还能与血清中的 IgM 和 IgA 结合。例如，蛋白质 A 可以包括来自金黄色葡萄球菌的蛋白质 A（SPA），SPA 是细胞壁抗原的主要成分，几乎 90% 以上的金黄色葡萄球菌菌株含有这种成分，但不同的菌株含量差别悬殊。利用蛋白质 A 能够与抗体结合的功能可以通过形成目标蛋白—抗体—蛋白质 A 复合体从而对目标蛋白进行定位和/或分析。

在本申请中，术语“固体支持物”通常是指适用于或可被修改以适用于附接本文描述的寡核苷酸标签、条码序列、引物等的任何材料。例如，固体支持物包括位于表面中的孔或凹陷的阵列，这些可使用多种技术进行制造，例如光刻法、冲压技术、成型技术和微蚀技术；固体支持物的组成和几何形状可以依据其用途而改变，例如，固体支持物可以是平面结构（例如载玻片、芯片、微芯片和/或阵列等）；例如，固体支持物或其表面还可以是非平面的，例如管或容器的内表面或外表面；例如，固体支持物还可以包括微球或珠粒。

在本申请种，“珠粒（beads）”或“微球（microspheres）”或“粒子（particulates）”通常是指小的离散粒子。适合的珠粒组合物包括但不限于：塑料、陶瓷、玻璃、聚苯乙烯、甲基苯乙烯、丙烯酸聚合物、顺磁材料、氧化钛溶胶、碳石墨、二氧化钛、乳胶或交联葡聚糖（诸如

琼脂糖)、纤维素、尼龙、交联胶束和铁氟龙,并且本文概述的用于固体支持物的任何其他材料全部可以使用,可以参考费雪尔邦斯实验室(Bangs Laboratories,Fishers Ind.)的微球检测指南(Microsphere DetectionGuide);在某些实施例中,微球可以是磁性微球或珠粒。

在本申请中,术语“独特分子鉴别区”也可以称作“分子条形码”、“分子标记”、“唯一标识符(UID)”、“唯一分子标识符(UMI)”等,通常是指为同一样品的每个原始核苷酸片段连接上的一段独一无二的序列编码。其通常可以设计为完全随机的核苷酸链(例如NNNNNNN)、部分简并的核苷酸链(例如NNNRNYN)或指定核苷酸链(例如,模板分子有限时);当被引入核酸分子中时,例如在第一链cDNA合成期间,可通过直接计数在扩增后测序的唯一性分子标识符(UMI)来校正随后的扩增偏倚。可根据本领域已知的方式进行UMI的设计、并入和应用,例如,通过WO2012/142213、Islam等人的(Nat.Methods)(2014)11:163-166,以及Kivioja,T等人的(Nat.Methods)(2012)9:72-74的公开示例的,所述文献通过引用的方式以其全部并入本文中。

在本申请中,术语“扩增引物识别区”通常是指一段能够与扩增所述目标核酸的引物序列互补杂交的核苷酸序列。所述引物与其结合能够引发核苷酸延伸、连接和/或合成,例如在聚合酶链反应作用下实现目标核酸拷贝数增多(即扩增),在一些实施方式中也包括寡核苷酸标签、分子唯一标识符等序列的扩增。

在本申请中,术语“离散分区”通常是指包含待分析目的物质的相互之间独立的空间单元。例如微滴或孔;例如,将目标核酸的样品与附接有寡核苷酸标签的固体支持物共分配形成的微滴;在一些实施方式中,所述离散分区中还可以包含其他根据不同的需求而分配其他物质,例如染料、乳化剂、表面活性剂、稳定剂、聚合物、适体、还原剂、引发剂、生物素标记物、荧光团、缓冲液、酸性溶液、碱性溶液、光敏感的酶、pH敏感的酶、水性缓冲液、去污剂、离子型去污剂、非离子型去污剂等等。

在本申请中,术语“可释放地附接”通常是指寡核苷酸标签与固体支持物之间的连接方式是可释放的、可切割的或可逆的或者可破坏、可消除的。例如,寡核苷酸标签与固体支持物的连接包含不稳定的键,例如,化学、热或光敏感的键,例如,二硫键、UV敏感的键等,通过相应的处理破坏这些不稳定的键从而实现可释放的附接;例如,寡核苷酸标签与固体支持物的连接包含可以被核酸酶识别的特定碱基,例如dU,可以通过UNG酶的作用切割所述连接;例如,寡核苷酸标签与固体支持物的连接包含核酸内切酶识别序列,可以通过核酸酶的作用切割所述连接;例如,所述固体支持物是可降解的,在施以降解条件时通过固体支持物的降解释放所述寡核苷酸标签,实现可释放的附接等。

在本申请中，术语“连接子”通常是指一段将各个功能性序列连接在一起的核苷酸序列也可以包括将寡核苷酸标签连接至固体支持物的分子序列（核酸、多肽或其他化学连接结构等）其中所述的功能性序列可以包括细胞条码区段、条码序列、扩增引物识别区、测序引物识别区、唯一分子识别符等，在某些实施方式中，该核苷酸可以是一段固定的核苷酸序列，在某些实施方式中，所述连接子还可以包含化学修饰。

在本申请中，术语“随机引导序列”通常是指一段在每一位置处可以呈现四重简并的随机引物。随机引导序列与目标核酸的相应区域（包括目标核酸自身序列及其附接的其他核苷酸序列）识别结合从而实现核苷酸序列的合成和/或扩增。

在本申请中，术语“条码序列”通常是指一段能够标识目标核酸的核苷酸序列或其衍生或修饰形式。

在本申请中，术语“细胞条码序列”通常是指可用于识别目标核酸样品来源的核苷酸序列。其中来源可以是，例如来自同一个细胞或不同的细胞。在核酸样品衍生自多种来源的情况下，可采用不同的细胞条码序列对每个来源中的核酸进行标记，使得样品的来源能被识别。条码（通常还称为索引、标签等）是本领域技术人员所熟知的，可使用任何合适的条形码或条形码组，例如在 US2013/0274117 的公开中所述的细胞条码序列。

在本申请中，术语“细胞条码区段”通常是指组成细胞条码序列的条码核苷酸单元，N 个所述细胞条码区段可以通过 PCR 或 DNA 连接酶的作用形成细胞条码区段。N 可以大于或等于 1，使得形成的细胞条码序列足以识别衍生自多种来源的每个核酸样品的细胞来源。

在本申请中，术语“寡核苷酸衔接子”通常是指附接与目标核酸并且包含能够与所述寡核苷酸标签互补杂交序列的一段核苷酸序列。该核苷酸序列可以是部分双链结构，例如可以具有与寡核苷酸标签杂交的突出序列；在某些实施方式中，寡核苷酸衔接子还可以包含转座酶（例如 Tn5 转座酶）结合序列；在某些实施方式中，寡核苷酸衔接子还可以包含扩增引物识别序列；在某些实施方式中，寡核苷酸衔接子还可以包含反转录引物序列。

在本申请中，术语“条码化的目标核酸”通常是指至少附接了细胞条码序列的目标核酸。

在本申请中，术语“共同条码结构域”通常是指用于识别目标核酸来源的条码序列。附接于同一个固体支持物的寡核苷酸标签中包含的共同条码结构域是相同的，附接于不同的固体支持物的寡核苷酸标签中包含的共同条码结构域相互之间是不同的，在某些实施方式中，释放自同一个固体支持物的寡核苷酸标签与来源于一个细胞的目标核酸连接，可以通过所述共同条码结构域识别其细胞来源。

在本申请中，术语“可变结构域”通常是指共同条码结构域之外的根据不同的需要设置的核苷酸序列。例如，连接子序列，扩增引物识别序列、测序引物识别序列等。

在本申请中，术语“转座酶-核酸复合物”通常是指转座酶与包含所述寡核苷酸衔接子的序列形成的复合物。转座酶通常是指一种能够与转座子末端结合并通过剪切、粘贴机制或复制性转座机制催化其向基因组其他部分移动的酶。转座子通常是指一段能够在基因组中自由跳跃的核苷酸片段，是由 Barbara McClintock 在二十世纪四十年代后期研究玉米遗传机制时提出的，之后的其他研究小组描述了转座的分子基础，例如，McClintock 发现染色体片段能够改变位置，从一条染色体跳到另一条染色体。这些转座子的重新定位能够改变其他基因的表达，例如在玉米中转座能够引起颜色变化，在细菌等其他生物中，能够引起抗生素耐药性在人类进化的过程中。转座酶-核酸复合物中可以包含 2 个分别结合了寡核苷酸衔接子的转座酶形成的二聚体，其中的 2 个转座酶可以是相同的转座酶也可以是不同的，其分别结合的寡核苷酸衔接子可以是相同的也可以是不同的。

在本申请中，术语“Tn5”通常是指 Tn5 转座酶，它是核糖核酸酶（RNase）超家族的成员。Tn5 能够在希瓦氏菌和大肠埃希氏菌中发现。Tn5 可以包括天然存在的 Tn5 转座酶及其各种活性突变形式；Tn5 与大多数其他转座酶一样含有 DDE 基序，DDE 基序是催化转座子转移的活性位点。据报道称 DDE 基序能够与二价金属离子（例如镁和锰）协调作用，在催化反应起到重要作用。转座酶 Tn5 可能通过 DDE 区域发生突变而使得转座活性升高，并催化转座子的移动。例如，其中 326 位的谷氨酸转化为天冬氨酸，而 97 和 188 位的两个天冬氨酸转化为谷氨酸（基于 GenBank 登录号 YP_001446289 的氨基酸序列的氨基酸编号）等。

在本申请中，术语“微流控装置”通常是指能够实现微流控的设备或系统。其中，微流控通常是指一种精确控制和操控微尺度流体的技术，尤其特指亚微米结构的技术，“微”通常是指微小的容量或体积（例如纳升，皮升，飞升级别）。微流控技术已广泛应用于多领域，例如生物医药领域，例如，分子生物学方法中的酶分析（如葡萄糖和乳酸分析）、DNA 分析（如聚合酶链式反应和高通量测序）、蛋白质组学分析等。微流控装置主体结构可以包括与其连接的简易贮存器，从装置外来源、歧管、流体流动单元（例如，致动器、泵、压缩机）等递送流体的流体管道，以及将微流体分配递送到随后的处理操作、仪器或部件的流体导管等。

在本申请中，术语“杂交”、“可杂交的”或者“互补的”通常是指在合适的温度和溶液离子强度的体外和/或体内条件下，核酸（例如 RNA，DNA）包含的核苷酸序列能够使其特异性地非共价结合（即形成 Watson-Crick 碱基对和/或 G/U 碱基对）至另一个核酸序列。Watson-

Crick 碱基配对包括：腺嘌呤/腺苷（A）与胸苷/胸腺嘧啶（T）配对，A 与尿嘧啶/尿苷（U）配对，鸟嘌呤/鸟苷（G）配对与胞嘧啶/胞苷（C）配对。在某些实施方式中，两个 RNA 分子（例如，dsRNA）之间的杂交，或者 DNA 分子与 RNA 分子的杂交（例如，当 DNA 靶核酸碱基与引导 RNA 配对时等），G 也可以与 U 碱基配对。杂交需要两个核酸包含互补序列，但是不能排除碱基之间可能错配。适用于两种核酸之间杂交的条件取决于核酸的长度和互补程度，这是本领域众所周知的。两个核苷酸序列之间的互补程度越大，具有这些互补序列的核酸的杂交体的解链温度（ T_m ）的值越大。

在本申请中，术语“读长”即 reads，通常是指核苷酸测序中一个反应获得的测序序列。Reads 可以是一段短的测序片段，是测序仪单次测序所得到的碱基序列数据，不同的测序仪器，reads 长度可以是不一样的。

发明详述

一方面，本申请提供一种分析来自细胞的目标核酸的方法，所述方法包括：

a) 提供包含下述的离散分区：

i. 源于单个细胞的目标核酸，其中至少部分所述目标核酸被添加了寡核苷酸衔接子序列而成为经附接的目标核酸；以及

ii. 附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3' 端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第一链与所述第二链形成部分双链的结构；

或者 ii. 附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3' 端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第二链与所述经附接的目标核酸形成部分双链的结构；

b) 在所述离散分区中，使所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸。

例如，其进一步包括：

c) 获得所述条码化的目标核酸的表征结果；以及

d) 至少部分基于 c) 中获得的所述表征结果中存在相同的所述细胞条码序列而将所述目标核酸的序列识别为源于所述单个细胞。

另一方面，本申请还提供了一种扩增来自细胞的目标核酸的方法，所述方法包括：

a) 提供包含下述的离散分区：i. 源于单个细胞的目标核酸，其中至少部分所述目标核酸被添加了寡核苷酸衔接子序列而成为经附接的目标核酸；以及 ii. 附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3'端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第一链与所述第二链形成部分双链的结构；或者 ii. 的步骤可以为附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3'端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第二链与所述经附接的目标核酸形成部分双链的结构；

b) 在所述离散分区中，使所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸；以及

c) 对所述条码化的目标核酸进行扩增。

另一方面，本申请还提供了一种对来自细胞的目标核酸进行测序的方法，所述方法包括：

a) 提供包含下述的离散分区：i. 源于单个细胞的目标核酸，其中至少部分所述目标核酸被添加了寡核苷酸衔接子序列而成为经附接的目标核酸；以及 ii. 附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3'端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第一链与所述第二链形成部分双链的结构；或者 ii. 的步骤可以为附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3'端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第二链与所述经附接的目标核酸形成部分双链的结构；

b) 在所述离散分区中，使所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸；以及

c) 对所述条码化的目标核酸进行测序。

例如，本申请中的所述寡核苷酸标签可以包含第一链以及第二链，所述第一链和所述第二链可以同时提供或者分别提供。在本申请中，当所述第一链和所述第二链同时提供时，所述

第一链可以与所述第二链形成部分双链的结构；当所述第一链和所述第二链分别提供时，所述第二链可以与所述经附接的目标核酸形成部分双链的结构。

条码化的目标核酸

在本申请中，条码化的目标核酸通过所述寡核苷酸标签与所述经附接的目标核酸连接生成。例如，使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。例如，使所述寡核苷酸标签的第二链的所述第二部分与附接至所述目标核酸的所述寡核苷酸衔接子杂交，以及使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。关于所述杂交，适用于两种核酸之间杂交的条件取决于核酸的长度和互补程度，这是本领域众所周知的。两个核苷酸序列之间的互补程度越大，具有这些互补序列的核酸的杂交体的解链温度（ T_m ）的值越大。

例如，所述寡核苷酸标签的第二链的所述第二部分的长度足以使其与其互补序列（附接至所述目标核酸的所述寡核苷酸衔接子序列或其部分序列）形成双链结构。

例如，所述第二链的所述第二部分的长度可以是 1 个核苷酸或更多，2 个核苷酸或更多，3 个核苷酸或更多，5 个核苷酸或更多，8 个核苷酸或更多，10 个核苷酸或更多，12 个核苷酸或更多，15 个核苷酸或更多，20 个核苷酸或更多，22 个核苷酸或更多，25 个核苷酸或更多或 30 个核苷酸或者更多。

例如，所述杂交不排除碱基之间可能错配。例如，所述第二链的所述第一部分或所述第二链的所述第二部分的序列不必与其杂交序列的序列有 100% 互补性。例如，可以是 60% 或更多，65% 或更多，70% 或更多，75% 或更多，80% 或更多，85% 或更多，90% 或更多，95% 或更多，98% 或更多，99% 或更多，99.5% 或更多地互补性。其余的非互补核苷酸可以与互补核苷酸成簇或散布，并且不需要彼此或与互补核苷酸相邻。例如，多核苷酸可在一个或多个区段上杂交，使得在杂交事件中不涉及中间或相邻区段（例如，形成发夹结构，“凸起”等）。

例如，使用连接反应将寡核苷酸标签与所述经附接的目标核酸连接。该连接可包括通过催化磷酸二酯键的形成将两个核酸区段接合在一起，例如所述寡核苷酸标签的第一链的所述杂交序列和所述附接至所述目标核酸的所述寡核苷酸衔接子。连接反应可包括 DNA 连接酶，诸如大肠杆菌 DNA 连接酶、T4 DNA 连接酶、T7 DNA 连接酶、哺乳动物连接酶（例如，DNA 连接酶 I、DNA 连接酶 III、DNA 连接酶 IV）、热稳定连接酶等。T4 DNA 连接酶可以连接含有 DNA、寡核苷酸、RNA 和 RNA-DNA 杂合体的区段。连接反应可以不包括 DNA 连接酶，而是采用替代物如拓扑异构酶。采用高浓度的 DNA 连接酶且包含 PEG 可实现快速连接。为

了选择连接反应的有利温度，可以考虑 DNA 连接酶的最适温度（例如可以是 37°C）以及待连接的 DNA 的解链温度。可将目标核酸和条形码化的固体支持物悬浮在合适的缓冲液中以使可能影响连接的离子作用最小化。

例如，其包括使至少一部分所述目标核酸从所述离散分区中的所述单个细胞中释放到细胞外，并在 b) 中使经释放的所述目标核酸与所述寡核苷酸标签连接，从而产生条码化的目标核酸。例如，所述使至少一部分所述目标核酸从所述离散分区中的所述单个细胞中释放到细胞外可以包括将细胞与溶解试剂接触，以释放离散分区内的细胞的内容物。所述溶解剂可以包括生物活性试剂，例如用于溶解不同细胞类型（例如革兰氏阳性(gram positive)或阴性细菌、植物、酵母、哺乳动物等）的溶解酶，例如溶菌酶、无色肽酶、溶葡萄球菌酶、硫葡萄糖苷酶白芥子（kitalase）、溶壁酶（lyticase）以及其他可商购的溶解酶。例如，还可使用基于表面活性剂的溶解溶液来溶解细胞，例如，溶解溶液可包括非离子表面活性剂，诸如 TritonX-100 和吐温 (Tween)20。例如，溶解溶液可包括离子表面活性剂，诸如十二烷基肌氨酸钠和十二烷基硫酸钠(SDS)。例如，还可采用可使用的其他方法（诸如电穿孔、热、声或机械细胞破坏）的溶解方法。

例如，所述使至少一部分所述目标核酸从所述离散分区中的所述单个细胞中释放到细胞外可以包括至少 5%、至少 10%、至少 15%、至少 20%、至少 25%、至少 30%、至少 35%、至少 40%、至少 45%的所述目标核酸从所述离散分区中的所述单个细胞中释放到细胞外。

例如，其包括使至少一部分从所述固体支持物释放的所述寡核苷酸标签进入所述单个细胞中，并在 b) 中与所述目标核酸连接，从而产生条码化的目标核酸。

例如，所述使至少一部分从所述固体支持物释放的所述寡核苷酸标签进入所述单个细胞中可以包括至少 25%、至少 30%、至少 35%、至少 40%、至少 50%、至少 55%、至少 60%、至少 70%、至少 75%、至少 75%的所述寡核苷酸标签进入所述单个细胞中。

例如，所述寡核苷酸标签可释放地附接至所述固体支持物。例如，可释放地、可切割地或可逆地附接至所述的固体支持物的寡核苷酸标签包括通过寡核苷酸标签分子与固体支持物之间的联接的切割/破坏而被释放或可释放的寡核苷酸标签，或通过固体支持物自身的降解而被释放的寡核苷酸标签，从而使寡核苷酸标签能够被其他试剂接近或可接近，或包括这两者。

例如，与固体支持物前体连接的 acrydite 部分、与固体支持物前体连接的另一物质或前体本身包含不稳定的键，例如，化学、热或光敏感的键，例如，二硫键、UV 敏感的键等。所述不稳定的键可以在将物质（例如寡核苷酸标签）可逆地连接（共价连接）至固体支持物。例如，热不稳定的键可包括基于核酸杂交的附接（例如，当寡核苷酸与附接至固体支持物的

互补序列杂交时),使得杂合体的热解链从固体支持物(或珠粒)释放寡核苷酸,例如,含有寡核苷酸标签的序列。此外,向凝胶固体支持物添加多种类型的不稳定键可导致能够响应于不同刺激的固体支持物的产生。每种类型的不稳定键可以对相关的刺激(例如,化学刺激、光、温度等)敏感,使得可通过施加合适的刺激来控制通过每种不稳定键附接至固体支持物的物质的释放。例如,通过凝胶珠子的活化官能团,可在凝胶固体支持物形成后将包含不稳定键的另一物质连接至凝胶固体支持物。可提供可释放地附接至固体支持物或以其他方式布置在离散分区中的试剂(带有关联的可激活的基团),使得一旦递送至期望的一组试剂(例如,通过共同分配),可激活的基团可以与期望的试剂反应。这类可激活的基团包括笼蔽基团、可去除的阻断或保护基团,例如,光不稳定基团、热不稳定基团,或可化学去除的基团。除热可切割的键、二硫键和 UV 敏感的键之外,可与前体或固体支持物偶联的不稳定键的其他非限制性实例还包括酯联接(例如,可用酸、碱或羟胺切割的)、邻二醇联接(例如,可通过高碘酸钠切割的)、Diels-Alder 联接(例如,可通过热切割的)、砜联接(例如,可通过碱切割的)、甲硅烷基醚联接(例如,可通过酸切割的)、糖苷联接(例如,可通过淀粉酶切割的)、肽联接(例如,可通过蛋白酶切割的)或磷酸二酯联接(例如,可通过核酸酶(DNA 酶)切割的)。

例如,所述寡核苷酸标签通过其第一链的 5'端直接或间接附接至所述固体支持物。例如,包括从所述固体支持物上释放所述至少一个寡核苷酸标签,并在 b) 中使经释放的所述寡核苷酸标签与所述经附接的目标核酸连接,从而产生条码化的目标核酸。

例如,在所述条码化的目标核酸中,所述目标核酸序列位于所述条码序列的 3'端。例如,所述目标核酸可以直接与所述条码序列的 3'端连接;例如,所述目标核酸不直接与所述条码序列的 3'端连接,所述目标核酸与所述条码序列之间可以存在任意其他核苷酸序列。

例如,在 b) 之后并且在 c) 之前,对所述条码化的目标核酸进行扩增。例如,在 b) 之后并且在 c) 之前,从所述离散分区中释放所述条码化的目标核酸,且所述扩增在所述条码化的目标核酸从所述离散分区中释放后进行。例如,在所述条码化的目标核酸从所述离散分区中释放后,可以进行进一步的化学或酶反应修饰,例如所述修饰可以包括 bisulfite conversion、5hmc conversion 等,之后再进行扩增。

例如,所述扩增中使用扩增引物。例如,所述扩增还可以包括对所述条码化的目标核酸进行进一步的修饰,使得其在另一侧也有固定的序列可以用于进行 PCR 扩增,例如,所述修饰可以包括反转录链转换、第二链合成、可以是末端转移酶(terminal transferase)反应,以及连接上第二种接头(adaptor)。

例如，所述扩增引物还可以包含通用引物。

例如，所述扩增中使用扩增引物，且所述扩增引物可以包含随机引导序列。所述随机引导序列包括在每一位位置处可以呈现四重简并的随机引物。例如，随机引物包括本领域中已知的具有各种随机序列长度的任何核酸引物。例如，随机引物可以包括长度为 3、4、5、6、7、8、10、11、12、13、14、15、16、17、18、19、20 个或更多个核苷酸的随机序列。例如，多个随机引物可以包括具有不同长度的随机引物。例如，多个随机引物可以包括具有相等长度的随机引物。例如，多个随机物可以包括长度为约 5 至约 18 个核苷酸的随机序列。例如，多个随机物包括随机六聚体。所述随机六聚体，可商购获得，并且广泛地用于扩增反应，例如多重置换扩增（MDA），例如，REPLI-g 全基因组扩增试剂盒(QIAGEN, Valencia, CA)为例。任何适合长度的随机引物可以用于本申请所述的方法和组合物中。

例如，所述扩增包括使所述随机引导序列与所述条码化的目标核酸至少部分杂交并且以模板定向的方式延伸所述随机引导序列。

寡核苷酸标签

在本申请中，所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3'端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第一链与所述第二链形成部分双链的结构。或者，在本申请中，所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3'端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第二链与所述经附接的目标核酸形成部分双链的结构。

关于所述杂交序列，适用于两种核酸之间杂交的条件取决于核酸的长度和互补程度，这是本领域众所周知的。两个核苷酸序列之间的互补程度越大，具有这些互补序列的核酸的杂交体的解链温度（ T_m ）的值越大。

例如，所述第二链的所述第一部分或所述第二链的所述第二部分的长度足以使其与其互补序列（例如，所述第一链中位于所述条码序列 3'端的杂交序列，例如，附接至所述目标核酸的所述寡核苷酸衔接子序列或其部分序列）形成双链结构。

例如，所述第二链的所述第一部分或所述第二链的所述第二部分的长度可以是 1 个核苷酸或更多，2 个核苷酸或更多，3 个核苷酸或更多，5 个核苷酸或更多，8 个核苷酸或更多，10 个核苷酸或更多，12 个核苷酸或更多，15 个核苷酸或更多，20 个核苷酸或更多，22 个核苷酸或更多，25 个核苷酸或更多或 30 个核苷酸或者更多。

例如，所述第二链的所述第一部分与所述第二链的所述第二部分的序列的长度可以相同，也可以不相同。

例如，所述双链结构不排除碱基之间可能错配。例如，所述第二链的所述第一部分或所述第二链的所述第二部分的序列不必与其杂交序列的序列有 100% 互补性。例如，可以是 60% 或更多，65% 或更多，70% 或更多，75% 或更多，80% 或更多，85% 或更多，90% 或更多，95% 或更多，98% 或更多，99% 或更多，99.5% 或更多地互补性。其余的非互补核苷酸可以与互补核苷酸成簇或散布，并且不需要彼此或与互补核苷酸相邻。例如，多核苷酸可在一个或多个区段上杂交，使得在杂交事件中不涉及中间或相邻区段（例如，形成发夹结构，“凸起”等）。

例如，附接至同一个固体支持物上的寡核苷酸标签的所述第二部分可以是相同的。

例如，附接至同一个固体支持物上的寡核苷酸标签的所述第二部分可以是不同的。例如，所述附接至同一个固体支持物上的各寡核苷酸标签的所述第二部分可以包括 1 种或以上的核苷酸序列，例如，所述第二部分的序列可以为 2 种或以上，例如，3 种或以上，例如，4 种或以上，例如，5 种或以上，例如，6 种或以上，例如，7 种或以上，例如，8 种或以上，例如，9 种或以上，例如，10 种或以上，例如，11 种或以上，例如，12 种或以上，例如，13 种或以上，例如，14 种或以上，例如，15 种或以上，从而使得所述附接至同一个固体支持物上的寡核苷酸标签能够与相应的 1 种或以上的所述经附接的目标核酸连接。

例如，附接至同一个固体支持物上并且含有同一种所述第二部分的所述寡核苷酸标签的数量可以是 1 个或以上，例如，50 个或以上，100 个或以上，500 个或以上，1000 个或以上，1500 个或以上，2000 个或以上，3000 个或以上，5000 个或以上，8000 个或以上，10000 个或以上，12000 个或以上，15000 个或以上，18000 个或以上，20000 个或以上，22000 个或以上，25000 个或以上，28000 个或以上，30000 个或以上，35000 个或以上，40000 个或以上，45000 个或以上，50000 个或以上。

例如，附接至同一个固体支持物上的含有不同第二部分的所述寡核苷酸标签的数量可以根据需要设置为不同比例，从而与相应的所述经附接的目标核酸连接。

例如，所述条码序列包含细胞条码序列，且附接至同一个固体支持物上的各寡核苷酸标签所包含的细胞条码序列相同。

例如，附接至同一个固体支持物上的寡核苷酸标签可以包括 1 个或更多个寡核苷酸标签，例如，50 个或更多，100 个或更多，500 个或更多，1000 个或更多，1500 个或更多，2000 个或更多，3000 个或更多，5000 个或更多，8000 个或更多，10000 个或更多，12000 个或更多，

15000 个或更多, 18000 个或更多, 20000 个或更多, 22000 个或更多, 25000 个或更多, 28000 个或更多, 30000 个或更多, 35000 个或更多, 40000 个或更多, 45000 个或更多, 50000 个或更多, 55000 个或更多, 60000 个或更多, 65000 个或更多, 70000 个或更多, 75000 个或更多, 80000 个或更多, 85000 个或更多, 90000 个或更多, 95000 个或更多, 100000 个或更多, 110000 个或更多, 120000 个或更多, 这些寡核苷酸标签的细胞条码序列是相同的, 并且其所述第二链的所述第二部分的序列可以是 1 种或以上, 例如, 所述第二部分的序列为 2 种或以上, 3 种或以上, 4 种或以上, 5 种或以上, 6 种或以上, 7 种或以上, 8 种或以上, 9 种或以上, 10 种或以上, 11 种或以上, 12 种或以上, 13 种或以上, 14 种或以上, 15 种或以上, 16 种或以上, 17 种或以上, 18 种或以上, 19 种或以上, 20 种或以上。

例如, 附接至不同的固体支持物上的寡核苷酸标签组所包含的细胞条码序列相互之间不同, 所述寡核苷酸标签组可以是附接至同一个固体支持物上的所有寡核苷酸标签的组合。

例如, 所述细胞条码序列包含至少 2 个细胞条码区段。例如, 所述细胞条码区段为 4 或更多个核苷酸 (nt), 例如, 5 或更多, 例如, 10 或更多, 12 或更多, 15 或更多, 18 或更多, 20 或更多, 21 或更多, 22 或更多, 23 或更多, 24 或更多, 25 或更多, 26 或更多, 27 或更多, 28 或更多, 29 或更多, 30 或更多, 31 或更多, 32 或更多, 33 或更多, 34 或更多, 或 35 或更多。

例如, 所述例如, 所述细胞条码序列包含至少 2 个细胞条码区段, 至少 3 个细胞条码区段, 至少 4 个细胞条码区段, 至少 5 个细胞条码区段, 至少 6 个细胞条码区段, 至少 7 个细胞条码区段, 至少 8 个细胞条码区段, 所述细胞条码区段按照在所述寡核苷酸标签中自 5' 端至 3' 端的顺序编码为细胞条码区段 1, 细胞条码区段 2, 细胞条码区段 3, 细胞条码区段 4, 细胞条码区段 5.....细胞条码区段 n。例如, 所述至少 2 个细胞条码区段可以通过 PCR 或 DNA 连接酶形成所述细胞条码序列。

例如, 可以通过如下方法生成所述细胞条码序列:

1) 将所述至少 1 个所述固体支持物分成至少 2 个初级等分试样, 例如, 至少 8 个等分, 至少 16 个等分, 至少 24 个等分, 至少 32 个等分, 至少 40 个等分, 至少 48 个等分, 至少 56 个等分, 至少 64 个等分, 至少 72 个等分, 至少 80 个等分, 至少 88 个等分, 至少 96 个等分;

2) 向每个所述初级等分试样提供至少 1 个细胞条码区段 1, 例如, 至少 1000 个细胞条码区段 1, 例如, 至少 10000 个细胞条码区段 1, 例如, 至少 100000 个细胞条码区段 1, 例如, 至少 1000000 个细胞条码区段 1, 例如, 至少 10000000 个细胞条码区段 1, 每个等分中

的细胞条码区段 1 与另外任一等分试样中的细胞条码区段 1 的序列和/或长度互相不同；

3) 使每个所述初级等分试样中的至少 1 个固体支持物与细胞条码区段 1 直接或间接连接，每个固体支持物连接至少一个细胞条码区段 1；

4) 合并所述至少 2 个初级等分试样，将合并的初级等分试样分成至少 2 个二级等分试样，例如，至少 8 个等分，至少 16 个等分，至少 24 个等分，至少 32 个等分，至少 40 个等分，至少 48 个等分，至少 56 个等分，至少 64 个等分，至少 72 个等分，至少 80 个等分，至少 88 个等分，至少 96 个等分；

5) 向每个所述二级等分试样提供至少 1 个细胞条码区段 2 或其互补序列，例如，至少 1000 个细胞条码区段 2 或其互补序列，例如，至少 10000 个细胞条码区段 2 或其互补序列，例如，至少 100000 个细胞条码区段 2 或其互补序列，例如，至少 1000000 个细胞条码区段 2 或其互补序列，例如，至少 10000000 个细胞条码区段 2 或其互补序列，每个等分试样中的细胞条码区段 2 或其互补序列与另外任一等分试样中的细胞条码区段 2 或其互补序列的序列和/或长度互相不同；

6) 使每个所述二级等分试样中的至少 1 个连接至固体支持物的细胞条码区段 1 与细胞条码区段 2 直接或间接连接。

例如，可以重复步骤 4) -6)，重复次数可以为 n ， n 可以为 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 或以上，以连接细胞条码区段 3, 细胞条码区段 4, 细胞条码区段 5.....细胞条码区段 n ，以产生足以针对每个细胞的独特序列的细胞条码，使第一细胞中的目标核酸可以具有独特序列的第一细胞条码，第二细胞中的目标核酸可以具有独特序列的第二细胞条码，第二细胞中的目标核酸可以具有独特序列的第二细胞条码，依此类推。

例如，在所述 b) 之后并且在 c) 之前，从所述离散分区中释放所述条码化的目标核酸。

例如，进一步进行 c)：对所述条码化的目标核酸进行测序，从而获得所述表征结果。

例如，所述表征结果可以包括所述条码化的目标核酸的核苷酸序列信息，例如包括细胞条码核苷酸序列信息，目标核酸的核苷酸序列信息、UMI 序列信息。

例如，由所述条码化的目标核酸的序列组装所述单个细胞的基因组的至少一部分的连续核酸序列。

例如，基于所述单个细胞的所述基因组的至少一部分的所述核酸序列来表征所述单个细胞。

例如，所述寡核苷酸标签还包含连接子序列 1，所述细胞条码区段 1 的 5'端可以通过连接子序列 1 连接至固体支持物。所述连接子序列 1 可以包含 acrydite 修饰，光切割修饰，S-S

修饰, dU 碱基修饰等序列, 可以通过各种方法断开, 将寡核苷酸标签释放。

例如, 所述寡核苷酸标签还包含其他功能序列, 所述其他功能序列可以位于所述细胞条码区段 1 和所述连接子序列 1 之间, 例如, 完全或部分的功能序列 (例如, 引物序列 (例如, 通用引物序列、靶向引物序列、随机引物序列) 识别区、引物退火序列、附接序列、测序引物识别区、扩增引物识别区 (例如, 通用扩增引物识别区) 等, 以用于随后的处理。

例如, 所述随后的处理包括扩增。例如, 所述扩增可以包括 PCR 扩增 (例如, Taq DNA 聚合酶扩增、Super Taq DNA 聚合酶扩增、LA Taq DNA 聚合酶扩增、Pfu DNA 聚合酶扩增、Phusion DNA 聚合酶扩增、KOD DNA 聚合酶扩增等)、等温扩增 (例如, 可以包括环介导的等温扩增 (LAMP)、解旋酶依赖性扩增 (HDA)、重组酶聚合酶扩增 (RPA)、链置换扩增 (SDA)、基于核酸序列的扩增 (NASBA)、转录介导扩增 (TMA) 等)、T7 启动子线性扩增、简并寡核苷酸引物 PCR 扩增 (DOP-PCR)、多重置换扩增 (MDA)、多次退火环状循环扩增技术 (MALBAC) 等。

例如, 所述细胞条码也可以不包含连接子, 所述细胞条码可以是一段单独的、由其他方法合成的核酸序列。

例如, 所述通用引物序列可以包括 P5 或其他合适的引物。通用引物 (例如, P5) 还可与测序装置相兼容, 例如能够附接至测序装置内的流动池。例如, 这类通用引物序列可提供约束在测序装置中流动池表面的寡核苷酸的互补序列, 以使得条码化的目标核酸序列能够固定到该表面上以供测序。

例如, 扩增引物序列, 用于进行扩增或复制过程 (例如, 沿着目标核酸序列使引物延伸) 的引物序列, 以便产生扩增的条形码化目标核酸序列。

例如, 测序引物序列, 所得的扩增靶序列将包含这样的引物, 并且容易地转移至测序系统中。例如, 当采用 Illumina 测序系统对扩增的靶标进行测序时, 所述测序引物序列可以包含 R1 引物序列、R2 引物序列。

例如, 所述寡核苷酸标签可以包含 T7 启动子序列。例如, 所述 T7 启动子序列包含如 SEQ ID NO: 1 所示的核苷酸序列 (TAATACGACTCACTATAG)。

例如, 所述寡核苷酸标签可以包含与 SEQ ID NO: 6-9 中的任一项具有至少 70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99% 或 100% 同一性的区域。

例如, 所述的所述核苷酸衔接子序列可以包含 P5 序列。例如, 所述的核苷酸衔接子序列

包含 P7 序列。

例如，所述细胞条码区段 1 和所述连接子序列 1 之间可以包含上述多种功能序列中的任何序列或其组合。例如，这些寡核苷酸可包括以下的任一个或多个：P5、R1 和 R2 序列、不可切割的 5'acrydite-P5、可切割的 5'acrydite-SS-P5、R1c、测序引物、读取引物、通用引物、P5_U、通用读取引物和/或任意这些引物的结合位点。

例如，所述细胞条码序列包含由连接子序列间隔开的至少 2 个细胞条码区段。

例如，细胞条码区段 1 的 3'端具有连接子序列 2，细胞条码区段 2 的 5'端、3'端分别具有连接子序列 3 和 4，细胞条码区段 3 的 5'端、3'端分别具有连接子序列 5 和 6，细胞条码区段 4 的 5'端、3'端分别具有连接子序列 7 和 8，以此类推，细胞条码区段 n 的 5'端、3'端分别具有连接子序列 2n-1 和 2n；连接子序列 2 与连接子序列 3 能够至少部分互补配对形成双链结构，连接子序列 4 与连接子序列 5 能够至少部分互补配对形成双链结构，连接子序列 6 与连接子序列 7 能够至少部分互补配对形成双链结构，以此类推，以启动细胞条码区段 1、细胞条码区段 2、细胞条码区段 3、细胞条码区段 4.....细胞条码区段 n 的连接。

例如，使用连接反应将各细胞条码区段连接形成寡核苷酸标签。该连接可包括通过催化磷酸二酯键的形成将两个核酸区段接合在一起，例如细胞条码区段 1 和前文所述的功能序列，例如，连接子序列 2 和细胞条码区段 2，连接子序列 3 和细胞条码区段 3，连接子序列 4 和细胞条码区段 4，连接子序列 5 和细胞条码区段 5，连接子序列 6 和细胞条码区段 6，以此类推。连接反应可包括 DNA 连接酶，诸如大肠杆菌 DNA 连接酶、T4 DNA 连接酶、哺乳动物连接酶（例如，DNA 连接酶 I、DNA 连接酶 III、DNA 连接酶 IV）、热稳定连接酶等。T4 DNA 连接酶可以连接含有 DNA、寡核苷酸、RNA 和 RNA-DNA 杂合体的区段。连接反应可以不包括 DNA 连接酶，而是采用替代物如拓扑异构酶。采用高浓度的 DNA 连接酶且包含 PEG 可实现快速连接。为了选择连接反应的有利温度，可以考虑 DNA 连接酶的最适温度（例如可以是 37°C）以及待连接的 DNA 的解链温度。可将样品和条形码化的固体支持物悬浮在缓冲液中以使可能影响连接的离子作用最小化。

例如，连接酶生成寡核苷酸标签的条件下，每轮提供的细胞条码区段可以包含如下结构：细胞条码区段以及位于细胞条码区段 3'端的连接子序列为双链结构，位于细胞条码区段 5'端的连接子序列为突出的单链结构，通过其与前一轮细胞条码区段 5'端连接子序列至少部分互补配对形成双链结构。

例如，使用连接反应将各细胞条码区段连接形成寡核苷酸标签的实例可以如图 2 或图 4 所示。

例如，通过聚合酶链式反应（PCR）将各细胞条码区段连接形成寡核苷酸标签。例如，所述聚合酶链式反应可以通过如下任意一种或多种聚合酶：Taq DNA 聚合酶、Super Taq DNA 聚合酶、LA Taq DNA 聚合酶、UlltraPF DNA 聚合酶、Tth DNA 聚合酶、Pfu DNA 聚合酶、VentR DNA 聚合酶、Phusion DNA 聚合酶、KOD DNA 聚合酶、Iproof DNA 聚合酶。例如，所述聚合酶链式反应中还可以包括使得上述聚合酶保持活性的缓冲液、金属离子；例如，所述聚合酶链式反应中还可以包括 dNTP 和或其修饰衍生物。

例如，在聚合酶链式反应（PCR）生成寡核苷酸标签的条件下，每轮提供细胞条码区段的互补序列，所述互补序列为单链结构，5'端和3'端各自具有单链结构的连接子序列，其中，5'端的连接子序列能够与前一轮连接的细胞条码区段3'端的连接子序列至少部分互补配对形成双链结构，3'端的连接子序列能够与后一轮连接的细胞条码区段5'端的连接子序列至少部分互补配对形成双链结构。

例如，通过聚合酶链式反应（PCR）将各细胞条码区段连接形成寡核苷酸标签实例可以如图1或图3所示。

经附接的目标核酸

在本申请中，所述目标核酸包括选自下组的一种或多种：DNA、RNA 和 cDNA。例如，所述目标核酸包括源自所述单个细胞中 RNA 的 cDNA。例如，所述 RNA 包括 mRNA。

在本申请中，所述目标核酸被添加了寡核苷酸衔接子序列而成为经附接的目标核酸。例如，所述寡核苷酸衔接子序列位于所述目标核酸的5'端。

例如，所述寡核苷酸衔接子序列可以包含与所述寡核苷酸标签中的所述第二链的所述第二部分互补的核苷酸序列 L，所述核苷酸序列 L 的长度可以与所述寡核苷酸标签中的所述第二链的所述第二部分长度相同，也可以不同；例如，所述核苷酸序列 L 的长度可以是 1 个核苷酸或更多，2 个核苷酸或更多，3 个核苷酸或更多，5 个核苷酸或更多，8 个核苷酸或更多，10 个核苷酸或更多，12 个核苷酸或更多，15 个核苷酸或更多，20 个核苷酸或更多，22 个核苷酸或更多，25 个核苷酸或更多或 30 个核苷酸或者更多。

例如，所述核苷酸序列 L 可以与所述寡核苷酸标签中的所述第二链的所述第二部分互补配对形成双链结构。例如，所述双链结构不能排除碱基之间可能错配。例如，所述核苷酸序列 L 的序列不必与所述寡核苷酸标签中的所述第二链的所述第二部分的序列有 100% 互补性。例如，可以是 60% 或更多，65% 或更多，70% 或更多，75% 或更多，80% 或更多，85% 或更多，90% 或更多，95% 或更多，98% 或更多，99% 或更多，99.5% 或更多地互补性。其余的非互补核苷酸可以与互补核苷酸成簇或散布，并且不需要彼此或与互补核苷酸相邻。例如，多

核苷酸可在一个或多个区段上杂交，使得在杂交事件中不涉及中间或相邻区段（例如，形成发夹结构，“凸起”等）。

例如，所述核苷酸衔接子序列包含转座子末端序列。例如，所述转座子末端序列是 Tn5 或修饰的 Tn5 转座子末端序列。例如，转座子末端序列是 Mu 转座子末端序列。例如，所述 Tn5 或修饰的 Tn5 转座子末端序列或 Mu 转座子末端序列可以包含 15 至 25 个核苷酸，例如，16 个核苷酸，17 个核苷酸，18 个核苷酸，19 个核苷酸，20 个核苷酸，21 个核苷酸，22 个核苷酸，23 个核苷酸，24 个核苷酸。

例如，Tn5 嵌合端序列 A14 (Tn5MEA) 和/或 Tn5 嵌合端序列 B15 (Tn5MEB) (包括下面阐述的互补的非转移序列 (NTS)) 可作为所述的转座子末端序列。

Tn5MEA: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'; (SEQ ID NO: 2)

Tn5MEB: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'; (SEQ ID NO: 3)

Tn5NTS: 5'-CTGTCTCTTATACACATCT-3'。(SEQ ID NO: 4)

例如，在本申请所述的方法中的步骤 a) 之前对所述 RNA 进行反转录，并产生所述经附接的目标核酸。例如，所述反转录中使用第一链合成引物在每一个 mRNA 样品中由 mRNA 合成 cDNA 的第一链。例如，第一链合成引物是包括寡聚 dT 引物。例如，所述反转录中使用第一链合成引物可以为反转录引物，所述反转录引物以 5' 至 3' 的方向包含所述寡核苷酸衔接子序列以及 polyT 序列。例如，所述反转录包括使所述 polyT 序列与所述 RNA 杂交并且以模板定向的方式延伸所述 polyT 序列。例如，第一链合成引物是随机物。例如，第一链合成引物是寡聚 dT 引物和随机物的混合物。例如，所述方法还包括将模板转换型寡核苷酸引物(TSO 引物)连同寡聚 dT 引物和随机物的混合物一起并入。例如，cDNA 的第二链是利用 TSO 引物进行合成。例如，cDNA 的第二链是利用与 cDNA 的第一链互补的第二扩增引物进行合成，该第一链扩展超过 mRNA 模板，从而包括互补的 TSO 链。

例如，所述目标核酸包括源自所述单个细胞的 DNA。例如，所述 DNA 包括基因组 DNA、

例如，所述 DNA 包括基因组 DNA、开放染色质 DNA、蛋白质结合的 DNA 区域和/或与蛋白、脂类和/或小分子化合物连接的外源核酸，所述蛋白、脂类和/或小分子化合物能够与细胞内的靶分子结合。例如，所述蛋白可以包括抗体、抗原。例如，所述靶分子可以包括细胞内待分析的目标核酸序列。例如，在本申请所述的方法中的步骤 a) 之前对源自单个细胞的所述 DNA 进行片段化。例如，DNA 片段化可包括将 DNA 链分离或破坏成小片或区段。例如，可采用多种方法对 DNA 进行片段化，例如，在 DNA 片段化后再附接所述寡核苷酸衔接子的序列（在此条件下附接的寡核苷酸衔接子的序列不包含所述转座子末端序列），包括限

制性消化或产生剪切力的多种方法。例如，限制性消化可利用限制性内切酶，以通过对两条链的平端切割或通过不均匀切割以产生粘端而在 DNA 序列中制造切口。例如，剪切力介导的 DNA 链破坏可以包括超声处理、声剪切、针剪切、移液或雾化。超声处理是一种类型的流体动力学剪切，其使 DNA 序列暴露于短期的剪切力，这可产生约 700bp 的片段大小。声剪切向碗形换能器内的 DNA 样品施加高频声能。针剪切通过使 DNA 穿过小直径的针来产生剪切力，以将 DNA 物理地撕裂成较小的区段。雾化力可通过使 DNA 通过喷雾器单元的小孔而产生，在该单元中从离开该单元的微细雾沫中收集所得的 DNA 片段。通常，这些片段可为约 200 至约 100000 个碱基之间的任何长度。例如，片段将为约 200bp 至约 500bp，约 500bp 至约 1kb，约 1kb 至约 10kb，或约 5kb 至约 50kb，或约 10kb 至约 30kb，例如，为约 15kb 至约 25kb。例如，较大遗传组分的片段化可通过任何方便可得的方法中进行，例如包括可商购获得的基于剪切的片段化系统（例如，Covaris 片段化系统）、大小靶向片段化系统（例如 Blue Pippin（Sage Sciences））、酶片段化方法（例如，DNA 内切酶、DNA 外切酶）等等。例如，所述片段化包括使用超声断裂，而后在经断裂的所述 DNA 上添加包含所述寡核苷酸衔接子的序列，从而获得所述经附接的目标核酸。

例如，在所述片段化之后或者在所述片段化的过程中产生所述经附接的目标核酸。例如，所述片段化包括使用转座酶-核酸复合物将包含所述寡核苷酸衔接子的序列整合到所述 DNA 中，并释放所述转座酶以获得所述经附接的目标核酸。

例如，所述转座酶包括金黄色葡萄球菌 Tn5（Colegio 等人，《细菌学杂志》(J.Bacteriol.)，183:2384-8，2001；Kirby C 等人，《分子微生物学》(Mol.Microbiol.)，43:173-86，2002），泰乐菌素(Ty1)（Devine 和 Boeke，《核酸研究》(Nucleic Acids Res.)，22:3765-72，1994 和国际公开案 WO 95/23875），转座子 Tn7（Craig,NL，《科学》(Science.)271:1512，1996；Craig,NL，在《微生物学和免疫学当前论题》(Curr Top Microbiol Immunol.)中的综述，204:27-48，1996），Tn/O 和 IS10（KlecknerN 等人，《微生物学和免疫学当前论题》(Curr Top Microbiol Immunol.)，204:49-82，1996），水手转座酶(Mariner transposase)（Lampe DJ 等人，《欧洲分子生物学组织杂志》(EMBO J.)，15:5470-9，1996），Tel（Plasterk R H，《微生物学和免疫学当前论题》(Curr.Topics Microbiol.Immunol.)，204:125-43，1996），P 因子（Gloor,GB，《分子生物学方法》(Methods Mol.Biol.)，260: 97-114，2004），Tn3（Ichikawa 和 Ohtsubo，《生物化学杂志》(J Biol.Chem.)265:18829-32，1990），细菌插入序列（Ohtsubo 和 Sekine，《微生物学和免疫学当前论题》(Curr.Top.Microbiol.Immunol.)204:1-26，1996），逆转录病毒（Brown 等人，《美国国家科学院院刊》(Proc Natl Acad Sci USA)，86:2525-9，1989）和酵母逆转录转座子（Boeke 和

Corces, 《微生物学年度评论》(Annu Rev Microbiol.), 43:403-34, 1989), 以及 IS5、Tn10、Tn903、IS911 和转座酶家族酶的工程化形式 (Zhang 等人, (2009)《公共科学图书馆遗传学》(PLoS Genet.)5:e1000689. 电子出版于 2009 年 10 月 16 日; Wilson C. 等人(2007)《微生物学方法杂志》(J.Microbiol.Methods)71:332-5)。

例如, 所述转座酶-核酸复合物包含转座酶以及转座子末端核酸分子, 其中所述转座子末端核酸分子包含所述寡核苷酸衔接子序列。

例如, 转座酶是 Mu 转座酶。例如, 所述转座酶为 Tn5 转座酶或 Tn10 转座酶。所述 Tn5 转座酶选自全长 Tn5 转座酶、Tn5 转座酶的部分功能域、Tn5 转座酶突变体。所述 Tn10 转座酶选自全长 Tn10 转座酶、Tn10 转座酶的部分功能域、Tn10 转座酶突变体。例如, 所述 Tn5 转座酶突变体可以选自: R30Q, K40Q, Y41H, T47P, E54K/V, M56A, R62Q, D97A, E110K, D188A, Y319A, R322A/K/Q, E326A, K330A/R, K333A, R342A, E344A, E345K, N348A, L372P, S438A, K439A, S445A, G462D, A466D。

例如, 所述两个转座酶分子可以结合相同或者不同的双链 DNA 转座子, 使得插入位点被 1 种或 2 种 DNA 所标记。例如, 所述两个转座酶分子 (例如 Tn5 及包含点突变超活性 T 年或其他类型的转座酶) 可以和一个所述寡核苷酸衔接子序列, 以及另外一个标准的转座子 DNA 序列组装成杂合的转座复合物, 或者只使用上述双链结构 2 形成单一的 Tn5 转复合物。所述标准的转座子 DNA 序列可以包含扩增引物序列和/或测序引物序列。

例如, 所述 DNA 可以包括与蛋白质结合的 DNA 区域, 且所述转座酶-核酸复合物中还包含直接或间接识别所述蛋白质的部分。例如, 所述直接或间接识别所述蛋白质的部分可以包括金黄色葡萄球菌蛋白质 A (ProteinA)、链球菌蛋白质 G (ProteinG)、链球菌蛋白质 L (ProteinL) 或其他具有结合抗体功能的蛋白类似物。例如, 所述直接或间接识别所述蛋白质的部分还可以包括特异性结合所述蛋白质的抗体。例如, 所述金黄色葡萄球菌蛋白质 A (ProteinA)、链球菌蛋白质 G (ProteinG)、链球菌蛋白质 L (ProteinL) 或其他具有结合抗体功能的蛋白类似物各自能够结合所述特异性结合所述蛋白质的抗体。

例如, 所述转座酶与所述金黄色葡萄球菌蛋白质 A (ProteinA)、链球菌蛋白质 G (ProteinG)、链球菌蛋白质 L (ProteinL) 或其他具有结合抗体功能的蛋白类似物形成融合蛋白。

例如, 所述融合蛋白与所述特异性结合所述蛋白质的抗体结合形成复合物, 之后靶向所述蛋白质。

例如, 所述特异性结合所述蛋白质的抗体与所述蛋白质结合, 之后所述融合蛋白与所述抗体结合从而靶向所述蛋白质。

例如，所述寡核苷酸衔接子序列还可以包含抗体识别序列，所述抗体识别序列用于识别/对映追踪不同的所述抗体。所述抗体识别序列可以通过类似随机引物的方式产生。

例如，所述经附接的目标核酸中包含独特分子鉴别区。所述独特分子鉴别区（UMI）是指附接到多个核酸分子中的每一个上的唯一性核酸序列。例如，当被并入在核酸分子中时，UMI 可用于通过直接计数在扩增后测序的独特分子鉴别区（UMI）来校正随后的扩增偏倚。例如，包括至少部分基于所述独特分子鉴别区的存在将所述条码化的目标核酸中的单个核酸序列鉴别为源于所述目标核酸中的给定核酸。例如，包括基于所述独特分子鉴别区的存在确定所述目标核酸中给定核酸的量可以如本领域已知的方式进行 UMI 的设计、并入和应用，例如，通过 WO 2012/142213、Islam 等人的《自然-方法学》(Nat.Methods)(2014)11:163-166，以及 Kivioja,T 等人的《自然-方法学》(Nat.Methods)(2012)9:72-74 的公开所示的，所述每个文献都通过引用的方式以其全部并入本文中。例如，所述独特分子鉴别区位于所述寡核苷酸衔接子序列与所述目标核酸序列之间。

例如，所述目标核酸还可以包括外源核酸，所述外源核酸包括与蛋白、脂类和/或小分子化合物连接的外源核酸，所述蛋白、脂类和/或小分子化合物能够与细胞内的靶分子结合。例如，所述蛋白可以包括抗体、抗原。例如，所述靶分子可以包括细胞内待分析的目标核酸序列。

例如，本文所述的转座反应和方法是批量执行的，然后将生物颗粒(例如，来自单细胞的细胞核/细胞/染色质)分配，使得多个离散分区被生物颗粒(例如，细胞、细胞核、染色质或细胞珠)单独占据。例如，可以将多个生物颗粒分配到多个离散分区中，使得多个离散分区中的离散分区包括单个生物颗粒。

固体支持物

在本申请中，所述固体支持物可以包括珠粒。例如，珠粒可以是多孔的、无孔的和/或其组合。例如珠粒可以是固体的、半固体的、半流体的、流体的和/或其组合。例如，珠粒可以是可溶解的、可破坏的和/或可降解的。例如，珠粒可以是不可降解的。例如，珠粒可以是凝胶珠粒。凝胶珠粒可以是水凝胶珠粒。凝胶珠粒可以由分子前体形成，例如聚合物或单体物质。半固体珠粒可以是脂质体珠粒。固体珠粒可包含金属，包括氧化铁、金和银。例如，珠粒可以是二氧化硅珠粒。例如，所述珠粒为磁性珠粒。例如，珠粒可以是刚性的。例如，珠粒可以是柔性的和/或可压缩的。

例如，珠粒可具有任何合适的形状。例如，珠粒的形状可以包括但不限于球形、非球形、椭圆形、长圆形、无定形、圆形、圆柱形及其变形形式。

例如,珠粒可具有均匀尺寸或不均匀尺寸。例如,珠粒的直径可以是至少约 10nm、100nm、500nm、1 μ m、5 μ m、10 μ m、20 μ m、30 μ m、40 μ m、50 μ m、60 μ m、70 μ m、80 μ m、90 μ m、100 μ m、250 μ m、500 μ m、1mm 或更大。例如,珠粒的直径可小于约 10nm、100nm、500nm、1 μ m、5 μ m、10 μ m、20 μ m、30 μ m、40 μ m、50 μ m、60 μ m、70 μ m、80 μ m、90 μ m、100 μ m、250 μ m、500 μ m、1mm 或更小。例如,珠粒的直径可以在约 40-75 μ m、30-75 μ m、20-75 μ m、40-85 μ m、40-95 μ m、20-100 μ m、10-100 μ m、1-100 μ m、20-250 μ m 或 20-500 μ m 的范围内。

例如,珠粒可以以具有相对单分散尺寸分布的珠粒群体或多个珠粒的方式提供。在需要在离散分区内提供相对一致量的试剂的情况下,保持相对一致的珠粒特性(例如尺寸)可有助于整体一致性。特别地,本文所述的珠粒可具有其横截面尺寸的变异系数小于 50%、小于 40%、小于 30%、小于 20%,并且例如小于 15%、小于 10%、小于 5%或更小的尺寸分布。

例如,珠粒可包含天然和/或合成材料。例如,珠粒可包含天然聚合物、合成聚合物或天然和合成聚合物。天然聚合物可以包括蛋白质和糖,例如脱氧核糖核酸、橡胶、纤维素、淀粉(例如,直链淀粉、支链淀粉)、蛋白质、酶、多糖、丝、聚羟基链烷酸酯、壳聚糖、葡聚糖、胶原、角叉菜胶、卵叶车前子、阿拉伯胶、琼脂、明胶、虫胶、梧桐树胶、黄原胶、玉米糖胶、瓜尔胶、刺梧桐树胶、琼脂糖、海藻酸、藻酸盐或其天然聚合物。合成聚合物可以包括丙烯酸类、尼龙、硅氧烷、氨纶、粘胶人造丝、多元羧酸、聚乙酸乙烯酯、聚丙烯酰胺、聚丙烯酸酯、聚乙二醇、聚氨酯、聚乳酸、二氧化硅、聚苯乙烯、聚丙烯腈、聚丁二烯、聚碳酸酯、聚乙烯、聚对苯二甲酸乙二醇酯、聚三氟氯乙烯、聚环氧乙烷、聚对苯二甲酸乙二醇酯、聚异丁烯、聚甲基丙烯酸甲酯、聚甲醛、聚丙烯、聚苯乙烯、聚四氟乙烯、聚乙烯醇、聚氯乙烯、聚偏二氯乙烯、聚偏二氟乙烯、聚氟乙烯和/或其组合(例如,共聚物)。珠粒也可以由除聚合物之外的材料形成,例如脂质、胶束、陶瓷、玻璃陶瓷、材料复合物、金属、其他无机材料等。

例如,珠粒可含有分子前体(例如,单体或聚合物),其可通过分子前体的聚合形成聚合物网络。例如,前体可以是已经聚合的物质,其能够通过例如化学交联进行进一步的聚合。例如,前体可包含丙烯酰胺或甲基丙烯酰胺单体、低聚物或聚合物中的一种或多种。例如,珠粒可包含预聚物,其是能够进一步聚合的低聚物。例如,可以使用预聚物制备聚氨酯珠粒。例如,珠粒可含有可进一步聚合在一起的单独聚合物。例如,可以通过不同前体的聚合产生珠粒,使得它们包含混合聚合物、共聚物和/或嵌段共聚物。例如,珠粒可在聚合物前体(例如,单体、低聚物、线性聚合物)、核酸分子(例如,寡核苷酸)、引物和其他实体之间包含共价键或离子键。例如,共价键可以是碳-碳键、硫醚键或碳-杂原子键。

例如，交联可以是永久的或可逆的，这取决于所用的特定交联剂。可逆交联可允许聚合物在适当条件下线性化或解离。例如，可逆交联还可以允许结合物质可逆地附接于珠粒表面。例如，交联剂可形成二硫键。例如，形成二硫键的化学交联剂可以是胱胺或改性的胱胺。

例如，二硫键可以在掺入珠粒的分子前体单元（例如，单体、低聚物或线性聚合物）或前体与核酸分子（例如，寡核苷酸）之间形成。例如，胱胺（包括改性的胱胺）是包含二硫键的有机试剂，其可以用作珠粒的单独单体或聚合物前体之间的交联剂。聚丙烯酰胺可以在胱胺或包含胱胺（例如，改性的胱胺）的物质存在下聚合，以产生包含二硫键的聚丙烯酰胺凝胶珠粒（例如，包含可化学还原的交联剂的可化学降解的珠粒）。二硫键可以允许在珠粒暴露于还原剂时使珠粒降解或溶解。

例如，壳聚糖（线性多糖聚合物）可以通过亲水链与戊二醛交联以形成珠粒。壳聚糖聚合物的交联可以通过由热、压力、pH 变化和/或辐射引发的化学反应来实现。

例如，珠粒可以是琼脂糖、聚烯酰胺、PEG 等各种单体聚合而成的单一或混合单体的大分子，或是几丁质、玻尿酸、葡聚糖等大分子凝胶，使用微流控液滴平台，在液滴中聚合为大小均一的凝胶珠粒。

例如，珠粒可包含 **acrydite** 部分，其在某些方面可用于将一个或多个核酸分子（例如，条形码序列、条形码化核酸分子、条形码化寡核苷酸、引物或其他寡核苷酸）附接到珠粒。例如，**acrydite** 部分可以指由 **acrydite** 与一种或多种物质的反应，例如 **acrydite** 与其他单体和交联剂在聚合反应期间的反应所产生的 **acrydite** 类似物。可以修饰 **acrydite** 部分以与待附接的物质形成化学键，例如核酸分子（例如条形码序列、条形码化核酸分子、条形码化寡核苷酸、引物或其他寡核苷酸）。**acrydite** 部分可以用能够形成二硫键的硫醇基团改性，或者可以用已经包含二硫键的基团改性。硫醇或二硫化物（通过二硫化物交换）可以用作待附接物质的锚点，或者 **acrydite** 部分的另一部分可以用于附接。例如，附接可以是可逆的，使得当二硫键断裂时（例如，在还原剂存在下），附接的物质从珠粒中释放出来。在其他情况下，**acrydite** 部分可包含可用于附接的反应性羟基。除了二硫键之外，还可以包括其他的释放方式，例如 UV 光促释放，或者可以用酶释放

离散分区和微流控装置

本申请提供了用于将固体支持物（例如珠粒）与样品共分配的装置，例如，用于共同分配样品组分和珠粒至同一离散分区。例如，将所述源于单个细胞的目标核酸与所述附接有至少一个寡核苷酸标签的固体支持物共分配至所述离散分区中。

例如，该装置可以由任何合适的材料形成。例如，装置可由选自下组的材料形成：熔融

二氧化硅、钠钙玻璃、硼硅酸盐玻璃、聚(甲基丙烯酸甲酯)PMMA、PDMS、蓝宝石、硅、锗、环烯烃共聚物、聚乙烯、聚丙烯、聚丙烯酸酯、聚碳酸酯、塑料、热固性塑料、水凝胶、热塑性塑料、纸、弹性体及其组合。

例如，所述离散分区可以包括孔或微滴。例如，将所述源于单个细胞的目标核酸与所述附接有至少一个寡核苷酸标签的固体支持物共分配至所述孔或微滴中。例如，所述孔可以包括细胞培养板的上样孔或者其他任何能够与所述装置配合并适于共分配的容器孔。例如，所述离散分区为微滴。例如，其中每个所述离散分区至多包括源自单个细胞的所述目标核酸。例如，所述目标核酸位于单个细胞或细胞核中。例如，使用微流控装置将所述源于单个细胞的目标核酸与所述附接有至少一个寡核苷酸标签的固体支持物共分配至所述离散分区中。

例如，离散分区（例如，液滴或孔）包含单细胞并根据本申请所述的方法进行处理。例如，离散分区包含单细胞和/或单细胞核。可以根据本申请所述的方法分配和处理单细胞和/或单细胞核。例如，单细胞核可以是细胞的组成部分。例如，离散分区包含来自单细胞或单细胞核的染色质（例如，单染色体或基因组的其他部分），并且根据本申请所述的方法进行分配和处理。

例如，所述离散分区中还包含连接酶，且所述连接酶使所述寡核苷酸标签与所述经附接的目标核酸连接。所述离散分区中包含但不限于连接酶，还可以包含其他需要的酶。例如，DNA 聚合酶、DNA 内切酶、DNA 外切酶、末端转移酶以及能够使所述寡核苷酸标签从所述固体支持物释放的光敏感酶活 pH 敏感的酶。所述连接酶包括 T4 连接酶，但不限于 T4 连接酶，例如还可以包括大肠杆菌 DNA 连接酶、T4 DNA 连接酶、T7 DNA 连接酶、哺乳动物连接酶（例如，DNA 连接酶 I、DNA 连接酶 III、DNA 连接酶 IV）、热稳定连接酶等。

例如，以包含流体流动通道的方式形成所述装置。可以使用任何合适的通道。例如，装置包含一个或多个流体输入通道（例如，入口通道）和一个或多个流体出口通道。例如，流体通道的内径可以为约 10 μ m、20 μ m、30 μ m、40 μ m、50 μ m、60 μ m、65 μ m、70 μ m、75 μ m、80 μ m、85 μ m、90 μ m、100 μ m、125 μ m 或 150 μ m。例如，流体通道的内径可以大于 10 μ m、20 μ m、30 μ m、40 μ m、50 μ m、60 μ m、65 μ m、70 μ m、75 μ m、80 μ m、85 μ m、90 μ m、100 μ m、125 μ m、150 μ m 或更大。例如，流体通道的内径可以小于约 10 μ m、20 μ m、30 μ m、40 μ m、50 μ m、60 μ m、65 μ m、70 μ m、75 μ m、80 μ m、85 μ m、90 μ m、100 μ m、125 μ m 或 150 μ m。流体通道内的体积流速可以是本领域已知的任何流速。

例如，所述微流控装置为微滴发生器。例如，可以使用微流控装置通过形成同时包含附接有至少一个寡核苷酸标签的固体支持物和样品的水性小液滴而使附接有至少一个寡核苷酸

标签的固体支持物与样品（例如，包含目标核酸样品）组合。所述水性小液滴作为离散分区。该水性小液滴可以是被油相包围的水性核心，例如，油包水乳液内的水性小液滴。该水性小液滴可含有一个或多个附接有至少一个寡核苷酸标签的固体支持物、样品、扩增试剂和还原剂。例如，该水性小液滴可包含以下的一种或多种：水、无核酸酶的水、附接有至少一个寡核苷酸标签的固体支持物、乙腈、固体支持物、凝胶固体支持物、聚合物前体、聚合物单体、聚丙烯酰胺单体、丙烯酰胺单体、可降解的交联剂、不可降解的交联剂、二硫键、acrydite 部分、PCR 试剂、细胞、细胞核、叶绿体、线粒体、核糖体、引物、聚合酶、条形码、多核苷酸、寡核苷酸、DNA、RNA、肽多核苷酸、互补 DNA(cDNA)、双链 DNA(dsDNA)、单链 DNA(ssDNA)、质粒 DNA、粘粒 DNA、染色体 DNA、基因组 DNA、叶绿体 DNA、线粒体 DNA、核糖体 RNA、病毒 DNA、细菌 DNA、mtDNA(线粒体 DNA)、mRNA、rRNA、tRNA、nRNA、siRNA、snRNA、snoRNA、scaRNA、微 RNA、dsRNA、探针、染料、有机物、乳化剂、表面活性剂、稳定剂、聚合物、适体、还原剂、引发剂、生物素标记物、荧光团、缓冲液、酸性溶液、碱性溶液、光敏感的酶、pH 敏感的酶、水性缓冲液、油、盐、去污剂、离子型去污剂、非离子型去污剂，等等。总之，该水性小液滴的组成将根据特定的处理需求而改变。

水性小液滴可以具有均匀的大小或不均匀的大小。例如，水性小液滴的直径可以为约 1 μ m、5 μ m、10 μ m、20 μ m、30 μ m、40 μ m、45 μ m、50 μ m、60 μ m、65 μ m、70 μ m、75 μ m、80 μ m、90 μ m、100 μ m、250 μ m、500 μ m 或 1mm。例如，流体小液滴可以具有至少约 1 μ m、5 μ m、10 μ m、20 μ m、30 μ m、40 μ m、45 μ m、50 μ m、60 μ m、65 μ m、70 μ m、75 μ m、80 μ m、90 μ m、100 μ m、250 μ m、500 μ m、1mm 或更大的直径。例如，流体小液滴可以具有小于约 1 μ m、5 μ m、10 μ m、20 μ m、30 μ m、40 μ m、45 μ m、50 μ m、60 μ m、65 μ m、70 μ m、75 μ m、80 μ m、90 μ m、100 μ m、250 μ m、500 μ m 或 1mm 的直径。例如，流体小液滴可以具有在约 40-75 μ m、30-75 μ m、20-75 μ m、40-85 μ m、40-95 μ m、20-100 μ m、10-100 μ m、1-100 μ m、20-250 μ m 或 20-500 μ m 的范围内的直径。

如上文所述，所述微流控装置（例如，小液滴发生器）可用于将样品与固体支持物（例如，条形码化附接有至少一个寡核苷酸标签的固体支持物的文库）以及（在需要的情况下）能够降解固体支持物的试剂（例如，如果固体支持物以二硫键连接，则是还原剂）组合。例如，可向与第一流体交叉点（例如，第一流体接合处）流体连接的第一流体输入通道提供样品（例如，核酸样）。可以向同样与第一流体交叉点流体连接的第二流体输入通道提供预形成的固体支持物（例如，附接有至少一个寡核苷酸标签的固体支持物，例如可降解的固体支持物），其中第一流体输入通道与第二流体输入通道在该第一流体交叉点交汇。样品和附接有至

少一个寡核苷酸标签的固体支持物可以在第一流体交叉点混合以形成混合物(例如,水性混合物)。例如,可向第四流体输入通道提供还原剂(或其他需要的试剂,例如表面活性剂、稳定剂、聚合物、适体、引发剂、生物素标记物、荧光团、缓冲液、酸性溶液、碱性溶液、光敏感的酶、pH 敏感的酶、水性缓冲液等),该第四流体输入通道同样与第一流体交叉点流体连接,并且与第一和第二流体输入通道在第一流体交叉点交汇。然后,还原剂可以与附接有至少一个寡核苷酸标签的固体支持物和样品在第一流体交叉点混合。例如,还可以在进入微流控装置之前将还原剂(或其他需要的试剂,例如表面活性剂、稳定剂、聚合物、适体、引发剂、生物素标记物、荧光团、缓冲液、酸性溶液、碱性溶液、光敏感的酶、pH 敏感的酶、水性缓冲液等)与样品和/或附接有至少一个寡核苷酸标签的固体支持物预混合,使得通过第一流体输入通道向微流控装置提供样品和/或通过第二流体输入通道向微流控装置提供附接有至少一个寡核苷酸标签的固体支持物。

例如,包含目标核酸的样品和附接有至少一个寡核苷酸标签的固体支持物混合物可以通过与第一流体交叉点(并与构成第一流体交叉点的任何流体通道)流体连接的第一出口通道离开第一流体交叉点。可以向与第一出口通道流体连接的第二流体交叉点(例如,第二流体接合处)提供混合物。例如,油(或其他合适的不混溶的)流体可以从与第二流体交叉点(并与构成该交叉点的任何流体通道)流体连接且在第二流体交叉点与第一出口通道交汇的一个或多个单独的流体输入通道进入第二流体交叉点。例如,可以在与第二流体交叉点(并与第一出口通道)流体连接且在第二流体交叉点与第一出口通道以及彼此交汇的一个或两个单独的流体输入通道中提供油(或其他合适的不混溶的流体)。油以及样品与附接有至少一个寡核苷酸标签的固体支持物的混合物可以在第二流体交叉点混合。形成的水性小液滴可在油内被运送通过从第二流体交叉点离开的第二流体出口通道。例如,形成的水性小液滴还可从第一流体交叉点离开第二出口通道的流体小液滴可被分配到孔中以供进一步处理。

例如,还可以控制包含目标核酸的样品相对于附接有至少一个寡核苷酸标签的固体支持物的占有率。这种控制如美国专利申请公开号 20150292988 中描述,其全部公开内容为了所有目的通过引用以全文并入本文。通常,将包含目标核酸的样品形成小液滴,使得至少 50%、60%、70%、80%、90%或更多的小液滴含有不超过一个附接有至少一个寡核苷酸标签的固体支持物。另外,使得至少 50%、60%、70%、80%、90%或更多的包含目标核酸的样品形成小液滴包含恰好一个附接有至少一个寡核苷酸标签的固体支持物。

例如,可以在混合物进入微流控装置中之前将样品与包含任何其他试剂(例如,样品扩增所需的扩增剂、还原剂等)的附接有至少一个寡核苷酸标签的固体支持物(例如,可降解

的固体支持物)预混合以产生水性反应混合物。在水性混合物进入流体装置时,该混合物可从第一流体输入通道流动并进入流体交叉点。例如,油相可以从同样与流体交叉点流体连接的第二流体输入通道(例如,与第一流体输入通道垂直或基本垂直的流体通道)进入流体交叉点。该水性混合物和油可以在流体交叉点混合,使得油包水乳液(例如,固体支持物-水-油乳液)形成。该乳液可包含在连续油相中的多个水性小液滴(例如,包含水性反应混合物的小液滴)。例如,每个水性小液滴可包含单个固体支持物(例如,附接至一组相同的条形码的凝胶固体支持物)、样品的等份(例如来自一个细胞的目标核酸)以及任何其他试剂(例如,还原剂、样品扩增所需的试剂等)的等份。例如,流体小液滴可包含多个附接有至少一个寡核苷酸标签的固体支持物。在小液滴形成时,小液滴可通过连续油相被运送通过离开流体交叉点的流体出口通道。离开出口通道的流体小液滴可被分配到孔中以供进一步处理。

在可在进入微流控装置之前将还原剂添加至样品或者可在第一流体交叉点添加还原剂的情况下,在第二流体交叉点形成的流体小液滴可含有还原剂。在这种情况下,当小液滴穿过离开第二流体交叉点的出口通道行进时,还原剂可降解或溶解流体小液滴内含有的固体支持物。

例如,微流控装置可含有平行的三个离散的流体交叉点。流体小液滴可以在这三个流体交叉点的任一处形成。样品和附接有至少一个寡核苷酸标签的固体支持物可以在这三个流体交叉点的任一个内混合。还原剂(或其他任和需要的试剂,例如通透剂、扩增剂、使寡核苷酸标签自固体支持物释放的切割剂)可以在这三个流体交叉点的任一个处添加。油可以在这三个流体交叉点的任一处添加。

例如,所述微流控装置包括第一输入通道和第二输入通道,它们在与输出通道流体连接的接合处汇合。例如,出口通道可以与第三输入通道在接合处流体连接。

例如,所述方法还包括将包含所述目标核酸的样品引入所述第一输入通道,且将附接有至少一个寡核苷酸标签的所述固体支持物引入所述第二输入通道,从而在所述输出通道中生成所述样品与所述附接有至少一个寡核苷酸标签的固体支持物的混合物。

例如,还可以包含第四输入通道并且其可以与第三输入通道和出口通道在接合处相交。例如,微流控装置可包含第一、第二和第三输入通道,其中第三输入通道与第一输入通道、第二输入通道或者第一输入通道与第二输入通道的接合处相交。例如,所述输出通道与第三输入通道在接合处流体连接。例如,所述第一输入通道和所述第二输入通道彼此之间形成基本上垂直的角度。

例如,还包括将油引入所述第三输入通道,使得形成油包水乳液内的水性小滴作为所述

离散分区。例如，每个所述离散分区中至多包含来自单个细胞的所述目标核酸。

本申请的方法、组合物、装置和试剂盒可与任何合适的油一起使用。例如，油可用于产生微滴。例如，该油可以包括氟化油、硅油、矿物油、植物油及其组合。

例如，微流控装置内的水性流体也可含有醇。例如，醇可以是甘油、乙醇、甲醇、异丙醇、戊醇、乙烷、丙烷、丁烷、戊烷、己烷及其组合。该醇可以以约 5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19% 或 20%(v/v) 存在于水性流体内。例如，该醇可以以至少约 5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20% 或更高(v/v) 的浓度存在于水性流体内。例如，该醇可以以小于约 5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19% 或 20%(v/v) 存在于水性流体内。

例如，所述油也可含有表面活性剂以稳定乳液。例如，表面活性剂可以是含氟表面活性剂、Krytox 润滑剂、Krytox FSH、工程化的流体、HFE-7500、硅酮化合物、含 PEG 的硅化合物，如 bis krytoxpeg(BKP)。该表面活性剂可以以约 0.1%、0.5%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%、2%、5% 或 10%(w/w) 存在。例如，该表面活性剂可以以至少约 0.1%、0.5%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%、2%、5%、10%(w/w) 或更高的浓度存在。例如，该表面活性剂可以以小于约 0.1%、0.5%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%、2%、5% 或 10%(w/w) 存在。

例如，可向油中添加加速剂和/或引发剂。例如，加速剂可以是四甲基乙二胺(TMEDA 或 TEMED)。例如，引发剂可以是过硫酸铵或钙离子。该加速剂可以以约 0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9% 或 2%(v/v) 存在。例如，该加速剂可以以至少约 0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9% 或 2%(v/v) 或更高的浓度存在。例如，该加速剂可以以小于约 0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9% 或 2%(v/v) 存在。

细胞和样品

在本申请中，所述细胞为任何生物体的细胞。所述生物体的细胞可以是体外细胞（例如，已建立的培养细胞系），可以是离体细胞（来自个体的培养细胞，原代细胞）。细胞可以是体内细胞（生物个体中的细胞），例如来自各种组织中的细胞。

例如所述生物体细胞可以包括动物细胞、植物细胞、微生物细胞。例如所述植物细胞可以包括拟南芥细胞，还可以包括农业作物的细胞，例如小麦，玉米，水稻，高粱，小米，大豆等植物体细胞；所述植物细胞还可以包括水果和坚果植物的细胞，例如产生杏，橙子，柠檬，苹果，李子，梨，杏仁、核桃等的植物体。例如所述植物细胞可以是来源于植物体任意部位的细胞，例如，是根细胞，叶细胞，木质部细胞，韧皮部细胞，形成层细胞，顶端分生组织细胞，薄壁组织细胞。

例如，所述微生物细胞可以包括细菌（例如大肠杆菌，古细菌）、真菌（例如酵母）、放线菌、立克次氏体、支原体、衣原体、螺旋体细胞等。

例如，所述动物细胞可以包括无脊椎动物（例如果蝇、线虫、涡虫等）细胞、脊椎动物（例如斑马鱼、鸡、哺乳动物）细胞。

例如，所述哺乳动物细胞可以包括小鼠、大鼠、兔子、猪、狗、猫、猴子、人类等。

例如，所述动物细胞可以包括来自生物体任何组织的细胞，例如干细胞、诱导性多能干（iPS）细胞、生殖细胞（例如卵母细胞，卵子细胞，精子细胞等），成体干细胞，体细胞（例如成纤维细胞，造血细胞，心肌细胞，神经元，肌肉细胞，骨细胞，肝细胞，胰腺细胞，上皮细胞，免疫细胞以及来源于肺、脾、肾、胃、大肠、小肠等器官或组织的任何细胞）以及胚胎的体外或体内任何阶段的细胞等。

例如，所述细胞可以是来自生物体液中的细胞。例如所述生物体的体液可以包括脑脊液、房水、淋巴液、消化液（例如唾液、胃液、小肠液、胆汁等）、乳汁、血液、尿液、汗液、泪液、粪便、呼吸道分泌物、生殖器官分泌物（例如精液、宫颈黏液）等。

所述样品包括所述细胞和/或由其获得的细胞核。

例如，所述样品可以包括所述生物体的核酸分子。所述核酸分子可以通过所属领域技术人员已知的分离核酸分子的技术手段从任意生物体分离提取的，包括 DNA 和 RNA。例如所述核酸分子提取自上述的生物体细胞或生物体的体液。

例如，所述目标核酸可以包括来自上述任何细胞中的核酸。例如，单个细胞中的核酸。

例如，所述目标核酸可以来自于单个细胞的多核苷酸，例如，双链 DNA。例如所述双链 DNA 可以包括基因组 DNA，例如，编码 DNA 和非编码 DNA；例如，开放染色质区域 DNA，蛋白结合处 DNA，线粒体 DNA 和叶绿体 DNA，例如所述多核苷酸可以包括 RNA，例如核糖体 RNA，mRNA。

例如，该目标核酸还可以是来自于福尔马林固定石蜡包埋的（Formalin-Fixed and Paraffin-Embedded, FFPE）含有细胞的样本。

例如，所述目标核酸还可以包括生物体基因组中含有 SNP 位点的序列，甲基化、羟甲基化修饰的核苷酸序列。

例如，还可以对所述细胞进行预处理。例如，所述预处理还包括使所述细胞的细胞核被暴露。例如，可以通过裂解缓冲液和浓蔗糖溶液处理从而暴露细胞核。

例如，所述细胞和/或由其暴露（获得）的所述细胞核可以被包裹在合适的基质中形成微球，所述微球作为样品进行反应。

例如，所述预处理包括固定所述细胞和/或由其暴露（获得）的所述细胞核。例如，使用固定剂对所述细胞进行固定，所述固定剂选自下组中的一种或多种：甲醛、多聚甲醛、甲醇、乙醇、丙酮、戊二醛、锇酸和重铬酸钾。

其中所述预处理包括使用去垢剂处理所述细胞或细胞核，所述去垢剂包括 Triton、NP-40 和/或 digitonin。

例如，所述预处理还可以包括去除线粒体、叶绿体、核糖体等细胞器。

例如，可将细胞与溶解试剂一起分配，以释放离散分区内的细胞的内容物。例如在通过额外通道将细胞引入微滴产生区的同时，或在即将将细胞引入微滴产生区时使溶解剂与细胞悬浮液接触。溶解剂可以包括生物活性试剂，例如用于溶解不同细胞类型（例如革兰氏阳性(gram positive)或阴性细菌、植物、酵母、哺乳动物等）的溶解酶，例如溶菌酶、无色肽酶、溶葡萄球菌酶、硫葡萄糖苷酶白芥子（kitalase）、溶壁酶（lyticase）以及其他可商购的溶解酶。例如还可以将其他溶解剂与细胞共分配以使得细胞的内容物释放至离散分区中。例如，可使用基于表面活性剂的溶解溶液来溶解细胞，例如，溶解溶液可包括非离子表面活性剂，诸如 TritonX-100 和吐温(Tween)20。例如，溶解溶液可包括离子表面活性剂，诸如十二烷基肌氨酸钠和十二烷基硫酸钠(SDS)。例如，还可采用可使用的其他方法(诸如电穿孔、热、声或机械细胞破坏)的溶解方法。

组合物和试剂盒

本申请还提供了一种组合物，其包含：多个固体支持物，每个所述固体支持物上附接有至少一个寡核苷酸标签，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3'端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与所述待测核酸中的序列互补的第二部分，且所述第一链与所述第二链形成部分双链的结构或者所述第二链与所述经附接的目标核酸形成部分双链的结构；所述寡核苷酸标签的条码序列包含共同条码结构域和可变结构域，所述共同条码结构域在附接于同一个固体支持物的寡核苷酸标签中是相同的，且所述共同条码结构域在所述多个固体支持

物中的两个或更多个固体支持物之间是不同的。本申请还提供了用于分析来自细胞的目标核酸的试剂盒，其包含本申请所述的组合物。例如，所述试剂盒还可以包括转座酶。例如，所述试剂盒进一步包含核酸扩增剂，逆转录剂，固定剂，通透剂，连接剂和裂解剂中的至少一种。

不欲被任何理论所限，下文中的实施例仅仅是为了阐释本申请的方法和用途等，而不用来限制本申请发明的范围。

实施例

实施例 1 检测开放染色质区域 (ATAC)

(1) 制备包含条码序列的核苷酸标签，其被固定于固相支持物上。

该核苷酸标签有两条链，形成部分双链结构 1，如下所示：

链 I：固相支持物~衔接序列——条码序列 (barcode) ——杂交序列 (固定序列，与链 II 的互补部分杂交)，其中条码序列 (barcode) 为 (barcode-linker) n 大于等于 1。

具体实例：Bead-acrydite-S-S-ACACTCTTTCCCTACACGACGCTCTTCCGATCT(read1, SEQ ID NO: 6)-barcode-ATCCACGTGCTTGAG (SEQ ID NO: 12)

链 II：杂交序列 (固定序列，与链 I 中固定 DNA 序列杂交) ——与转座子复合物链 I 的 5' 端互补的序列

具体实例：CGAATGCTCTGGCCTCTCAAGCACGTGGAT (SEQ ID NO: 9)

固体支持物是聚丙烯酰胺微球，其通过微流控设备制备，将丙烯酰胺：Bis 混合物、以及 acrydite-DNA 引物、APS 诱发剂在微流控装置中混合成为液滴，其中含有 TEMED 催化剂，液滴会自发聚合成为凝胶微球，之后微球按照 barcode 合成方式加标签。

在连接反应中，溶液中含有 10mM DTT，S-S 键可以被还原从而释放引物。

(2) 制备转座子复合物，组装含有 DNA 部分双链序列的 Tn5 转座子。

其中含有的一个 DNA 序列为 A 链和 B 链退火形成双链结构 2。

链 A：磷酸基团——与链 II 中核酸分子的链 I 或链 II 中固定 DNA 序列中至少部分互补的序列——(UMI)——Tn5 转座酶结合的序列

具体实例：AGGCCAGAGCATTCGNNNNNNNAGATGTGTATAAGAGACAG (SEQ ID NO: 5)

链 B：Tn5 转座酶结合的序列 (与链 A 中的转座子蛋白 (Tn5) 结合的序列互补的序列) ——磷酸基团

具体实例：p-CTGTCTCTTATACACATCT (SEQ ID NO: 4)

其中，A 链中的 UMI 不是必须的；(1) 和 (2) 中的序列中可以含有修饰碱基，如 5mC。

Tn5 转座复合物是二聚体，两个 Tn5 蛋白可以结合相同或者不同的部分双链 DNA 转座子，使得插入位点被 1 种或 2 种 DNA 所标记；Tn5 蛋白（可以包含点突变超活性或其他类型的转座酶）可以和以上的双链结构 2，以及另外一个标准的转座子 DNA 组装成杂合的转座复合物，或者只使用上述双链结构 2 形成单一的 Tn5 转复合物。

(3) 制备样品。可以是非固定的细胞或细胞核，甲醛（或其他固定剂）固定的细胞或细胞核，非固定或固定的组织切片等。其中，固定或非固定样品用包含有去垢剂（Triton，NP-40 或 Digitonin 等）的缓冲液处理，还可以包括裂解细胞（非固定样品）得到细胞核的中间步骤，去垢剂裂解或通透细胞和细胞核，使得 Tn5 酶可以进入细胞核作用。典型的通透剂溶液可以包括 Tris，蔗糖，氯化钠，去垢剂。

(4) 转座反应。对上述处理好的样品加上包含 2 价金属离（例如，镁离子）的 Tn5 酶缓冲液，加入组装好的 Tn5 转座复合物，进行 ATAC 转座反应（37°C，30 分钟-2 小时）。即该反应体系包括：细胞或细胞核或组织；Tn5 转座复合物；缓冲液。反应完后，用缓冲液对样品进行洗涤，去除未反应的 Tn5 酶。

(5) 连接反应。加入 T4 DNA 连接酶反应缓冲液，连接步骤 (1) 中的核苷酸标签，T4 DNA 连接酶，核苷酸标签，在适当温度（4°C-37°C）下进行连接反应 20 分钟以上。

反应体系包括：细胞或细胞核或组织（转座反应后的）；T4 DNA 连接酶，核苷酸标签，反应后在连接反应体系中加入过量的游离和核苷酸标签互补序列，封闭多余未反应的核苷酸标签。

(6) 提取细胞中的 DNA。对于非固定样品，直接加入裂解液后用 DNA 提取试剂盒，磁珠等方法纯化；对于固定样品，加入蛋白酶 K 反应缓冲液，蛋白酶 K，在 55-65°C 进行解交联后纯化 DNA。

对于纯化的 DNA，1) 如果是使用杂合 Tn5 进行，产物双侧都有 PCR 扩增序列，可以直接扩增，获得测序文库。

2) 如果是使用单一的 Tn5 进行，那么 DNA 产物只有一侧有 PCR 引物，我们需要对这个 DNA 进行打断和连接，在另一侧加入扩增引物，这可以使用单一的 Tn5 酶进行，也可以使用超声或酶打断，然后末端加 A，连接头，最终获得测序文库。

按照上述步骤，以人 293T 细胞为例，取新鲜细胞，制备细胞核，用杂合 Tn5 进行 ATAC 反应，然后连接一个 Illumina 测序文库 P5 端（read1）侧的扩增序列的接头，构建文库，用 read1 引物和杂合 Tn5 中另一个 DNA 片段的 read2 引物扩增产物，最后进行分析。具体步骤

如下:

A. Tn5 转座子 (transposome)

使如下序列退火形成双链:

10uM Top1 5'p- AGGCCAGAGCATTCGNNNNNNNAGATGTGTATAAGAGACAG
(SEQ ID NO: 5) (链 A)

10uM Top2 GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG (SEQ ID NO:
3) (链 A)

20uM Bottom 5'p-CTGTCTCTTATACACATCT (SEQ ID NO: 4) (链 B)

之后与 10uM 的 Tn5 酶 (购自 Epicenter 公司) 室温孵育, 组装成 10uM 浓度的 Tn5 转座子, Top1/Bottom 双链与 Tn5 形成的转座子为 p-Tn5, Top2/Bottom 双链与 Tn5 形成的转座子为 Tn5-B。

B. 珠子制备

1) Bead 上附接的序列如下所示: Bead-S-S-PCR adaptor-barcode1-linker1-barcode2-linker2-barcode3-ligation linker

其中, PCR 接头 (adaptor) 序列为 ACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 6), 连接序列 1 (Linker1) 为 CGACTCACTACAGGG (SEQ ID NO: 7), 连接序列 2 (Linker2) 序列为 TCGGTGACACGATCG (SEQ ID NO: 8), Ligation linker 序列为 ATCCACGTGCTTGAG (SEQ ID NO: 12)。Barcode1=96 种 5bp 碱基序列, Barcode2=96 种 5bp 碱基序列, Barcode3=96 种 5bp 碱基序列。

2) 合成 3x96 种序列

1. PCR handle-96xbarcode1-linker1, 合成 96 个此序列的反向互补序列;
2. linker1-96xbarcode2-linker2, 合成 96 个此序列的反向互补序列;
3. linker2-96xbarcode3-ligation linker, 合成 96 个此序列的反向互补序列。

3) 微球合成:

合成如下氨基序列: 5'amine-S-S-ACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 6) 以及 30um 羧基修饰微球 (知益, KBsphere®, www.kbspheres.com/productshow.asp?id=903)。

偶联反应: 微球+50mM EDC+100uM 氨基序列 (SEQ ID NO: 6), 将氨基序列和羧基微球偶联, 获得如下结构: bead-S-S- ACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 6)

4) 附接标签

将合成的微球均分到 96 孔板, 分别加入 PCR handle-96xbarcode1-linker1, 进行第一轮加标签 (barcoding) 反应。反应体系及过程如下所示: 10ul 微球 + 2ul BstI 缓冲液 + 1ul 10uM dNTP + 1ul 100uM PCR handle-96xbarcode1-linker1, 之后 95℃ 保持 5min, 60℃ 保持 20min; 之后再加入 1ul BstI+5ul H₂O, 60℃ 保持 60min。

第一轮加标签反应完成后收集所有微球, 混合, 95℃ 反应 5min 去除互补链, 洗涤, 得到第一轮加标签 (96xbarcode1-linker1) 的微球。然后把微球均分到 96 孔板中, 加入 linker1-96xbarcode2-linker2 (第二轮)、linker2-96xbarcode3-ligation linker (第三轮), 按照第一轮体系方法进行第二, 三轮加标签反应, 最后得到带有 3 重 barcode 的单链微球, 微球洗涤之后, 跟互补序列 CGAATGCTCTGGCCTCTCAAGCACGTGGAT (SEQ ID NO:9) 退火形成部分双链结构, 最终获得如下附接部分双链结构的微球:

Bead-S-S-ACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 6) -barcode1-CGACTCACTACAGGG-barcode2- TCGGTGACACGATCG (SEQ ID NO: 8) -barcode3-ATCCACGTGCTTGAG (SEQ ID NO: 12)

3'-

TAGGTGCACGAACCTCTCCGGTCTCGTAAGC-5' (SEQ ID NO:9 的反向排列)

C. ATAC 实验

将人 293T 细胞系重悬在裂解液 (10 mM Tris-Cl, pH 7.4; 10 mM NaCl; 3 mM MgCl₂; 0.01% NP-40) 中裂解细胞, 获得细胞核。

取 10 万细胞核与步骤 (1) 获得的 p-Tn5、Tn5-B 进行反应, 反应体系如下所示:

25ul 2xTD Buffer (Illumina) + 2.5ul 10uM p-Tn5 + 2.5ul 10uM Tn5-B + 20ul 细胞核 (10 万个), 37℃ 反应 30min, PBS 洗涤细胞核。

D. 高通量标记

使用如图 8 所示的微流控芯片进行细胞标记, 微球通道 (Bead channel): 100um, 细胞核通道 (Nuclei Channel): 50um。

准备以下溶液:

细胞核溶液 1ml (100 细胞核/ul 浓度), 包括: 200ul 10xT4 DNA ligase Buffer, 10ul T4 DNA ligase, 10ul 1M DTT, 780ul 细胞核/水。

bead 溶液 (100 bead/ul 浓度): Bead in PBS。

细胞核溶液、bead 溶液、油 (FC40 氟碳油, 含有 1% 表面活性剂 FluoroSurfactant, Ran

Biotech) 在微流控芯片上形成 120um 直径的液滴 (drop collection), 37°C 连接 1 小时。

E. 建库

在步骤 D 的液滴中加入等体积全氟辛醇破碎液滴, 离心, 吸取水相, 使用 Qiagen DNA purification kit 纯化水相中的 DNA, 用如下反应体系扩增 DNA 获得最终测序文库: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TrueSeqD501, 1ul 10uM 引物 N701, 1ul Taq, 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 30sec, 18 个循环。

引物 TrueSeqD501 序列:
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
(SEQ ID NO: 10)

引物 N701 序列:
CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

以上文库投入约 500 细胞, 每个细胞测序 10 万个 PE150 reads, 总数据量为 15G。

使用 esATAC 软件对数据进行分析 and 质控, 把所有测序数据合并在一起分析。结果如图 5-7 所示, 测序片段大小呈现 ATAC 典型的核小体梯度 (见图 5), 信号富集转录起始位点 (TSS) 呈现典型的 ATAC 信号 (见图 6A 和 6B), 峰 (peaks) 与已知开放区域高度重合 (见图 7), 其中, 总峰数值为 11898, 落入 DNaseI 超敏位点合集的峰比例 (Peaks overlapped with union DHS ratio) 为 74.0%, 落入黑名单的峰比例 (Peaks overlapped with blacklist ratio) 为 0.5%, FRiP (Fraction of reads in peaks, 落入峰域的读长分数) 为 99.8%。这些结果表明该方法可以精确检测细胞中由 Tn5 介导的 ATAC 插入产物。

F 单细胞数据拆分与分析

对上述测序数据, 首先使用 Dropseq pipeline 识别细胞核酸标签 (cell barcode), 在 read1 中, 1-45bp 为 barcode 位置, 根据 96x96x96 种 barcode 组合可能, 统计每种 barcode 中的 read 数目, 画出堆积曲线, 确定文库中存在的有效细胞数目约为 400, 如图 9 所示。

ii. 获得每个细胞中的 unique mapped reads 数目, 其分布如图 10 所示, 即每一个细胞平均得到的 ATAC reads 数目中位值在 10000 左右, 优于 Bing Ren 的约 2-3000 的平均值。

iii. 单细胞 ATAC 的结果, 把每个 read 比对到基因组上位置 (由 dropseq 程序流程获得) 信息加载到 IGV 基因组浏览器中可视化, 得到如图 11 所示结果, 图中下部为 45 个单细胞的 ATAC 数据在基因区域的分布, 图中部为 45 个单个细胞的 ATAC 数据加和在一起的结果显示, 其跟图上部大量细胞 (约 1 万细胞) 的 ATAC 模式高度相似, 处于基因转录起始位点。

iv. 单细胞相关性分析，通过 R 语言包中计算 Pearson Correlation 的函数得到如图 12 所示的结果，图中颜色的加深表示细胞之间的相关性越高，由图中的显示可以看到单细胞的 ATAC 信号呈现高相关性，表明了由该方法获得的单细胞数据的真实准确性。

实施例 2 检测 DNA 和蛋白质相互作用

(1) CUT&Tag 是最新的研究 DNA 和蛋白相互作用的方法，代替传统的 ChIP-seq 方法，其原理是用一个蛋白质 A（一种细胞来源的可以结合不同物种来源的抗体重链保守区的蛋白）跟 Tn5 形成融合的蛋白，通过蛋白质 A 与抗体的结合，把 Tn5 酶靶向到抗体结合的目标蛋白上，通过 Tn5 酶的转座活性，把 DNA 片段直接插入到目标蛋白结合的 DNA 区域，对这个产物进行扩增测序，从而直接得到蛋白的结合位置信息。因此，CUT&Tag 的分子产物是和 ATAC 一样的，不同的是 ATAC 中 Tn5 酶插入位点在开放染色质区域，CUT&Tag 中 Tn5 插入位点是目的蛋白周边，因此可以使用跟实施例 1 中 ATAC 类似的方法来标记这一产物，使用的 DNA 转座子跟 ATAC 类似，同样也可以组装一种或杂合的 Tn5 转座复合物。不同的步骤在于：用蛋白质 A 或者 G-Tn5 融合蛋白来组装 Tn5 转座复合物；为了区分多种抗体，DNA 转座子上面除了 ATAC Tn5 的序列外，还可以在不同位置包含抗体识别码，用于区分多种抗体。

(2) 制备样品：可以是非固定的细胞或细胞核，甲醛（或其他固定剂）固定的细胞或细胞核，非固定或固定的组织切片等。其中，固定或非固定样品用包含有去垢剂（Triton，NP-40 或 Digitonin 等）的缓冲液处理，还可以包括裂解细胞（非固定样品）得到细胞核的中间步骤，去垢剂裂解或通透细胞和细胞核，使得 Tn5 酶可以进入细胞核作用。

(3) 抗体结合。用血清 BSA 等对样品进行封闭，然后加入一抗与目标蛋白结合，洗涤去除多余的一抗抗体。可以进一步用抗一抗的二抗继续结合样品（该步骤不是必须的），增加蛋白质 A/G 的结合位点，放大信号。如果要同时检测 2 种蛋白质的相互作用，可以将一抗和蛋白质 A/G-Tn5 融合蛋白先结合为复合物，每种抗体结合的蛋白质 A/G-Tn5 融合蛋白上的 DNA 带有不同的抗体识别码。同时将 2 个或多个一抗-蛋白质 A/G-Tn5 融合蛋白复合物直接跟细胞/组织结合，一步把 Tn5 带到目标蛋白周围。

(4) 转座反应。用蛋白质 A-Tn5 融合蛋白（一抗-蛋白质 A-Tn5 融合蛋白复合物）结合样品，洗涤多余的酶，然后样品中加入含二价离子的 Tn5 反应液，进行转座反应，（37°C，30 分钟-2 小时）。

(5) 按照实施例 1 中的方式进行连接反应及后续处理，构建文库，测序。

具体步骤如下：

(1) 制备包含条码序列的核苷酸标签，其被固定于固相支持物上。

该核苷酸标签有两条链，形成部分双链结构 1，如下所示：

链 I：固相支持物~附接序列——条码序列（barcode）——杂交序列（固定序列，与链 II 的互补部分杂交），其中条码序列（barcode）为（barcode-linker） n 大于等于 1。

具体实例：Bead-acrydite-S-S-ACACTCTTTCCCTACACGACGCTCTTCCGATCT(read1, SEQ ID NO: 6)-barcode-ATCCACGTGCTTGAG（SEQ ID NO: 12）

链 II：杂交序列（固定序列，与链 I 中固定 DNA 序列杂交）——与转座子复合物链 I 的 5'端互补的序列

具体实例：CGAATGCTCTGGCCTCTCAAGCACGTGGAT（SEQ ID NO:9）

固体支持物是聚丙烯酰胺微球，其通过微流控设备制备，将丙烯酰胺：Bis 混合物、以及 acrydite-DNA 引物、APS 诱发剂在微流控装置中混合成为液滴，其中含有 TEMED 催化剂，液滴会自发聚合成为凝胶微球，之后微球按照 barcode 合成方式加标签。

在连接反应中，溶液中含有 10mM DTT，S-S 键可以被还原从而释放引物。

（2）制备转座子复合物，组装含有 DNA 部分双链序列的 pA-Tn5 转座子。

其中含有的一个 DNA 序列为 A 链和 B 链退火形成双链结构 2。

链 A：磷酸基团——与链 II 中核酸分子的链 I 或链 II 中固定 DNA 序列中至少部分互补的序列——（UMI）——Tn5 转座酶结合的序列

具体实例：AGGCCAGAGCATTCGNNNNNNNAGATGTGTATAAGAGACAG（SEQ ID NO: 5）

链 B：Tn5 转座酶结合的序列（与链 A 中的转座子蛋白（Tn5）结合的序列互补的序列）——磷酸基团

具体实例：p-CTGTCTCTTATACACATCT（SEQ ID NO: 4）

其中，A 链中的 UMI 不是必须的；（1）和（2）中的序列中可以含有修饰碱基，如 5mC。

Tn5 转座复合物是二聚体，两个 pA-Tn5 蛋白可以结合相同或者不同的部分双链 DNA 转座子，使得插入位点被 1 种或 2 种 DNA 所标记；pA-Tn5 蛋白（可以包含点突变超活性或其他类型的转座酶）可以和以上的双链结构 2，以及另外一个标准的转座子 DNA 组装成杂合的转座复合物，或者只使用上述双链结构 2 形成单一的 Tn5 转复合物。

具体操作

等摩尔浓度的 pA-Tn5 蛋白和退火好的双链引物混合后在室温放置 1 小时以上，形成功能转座子复合物。

（3）制备样品。可以是非固定的细胞或细胞核，甲醛（或其他固定剂）固定的细胞或细

胞核，非固定或固定的组织切片等。其中，固定或非固定样品用包含有去垢剂（Triton，NP-40 或 Digitonin 等）的缓冲液处理，还可以包括裂解细胞（非固定样品）得到细胞核的中间步骤，去垢剂裂解或通透细胞和细胞核，使得抗体及 pA-Tn5 酶可以进入细胞核作用。典型的通透剂溶液可以包括 Tris，蔗糖，氯化钠，去垢剂。

针对目标蛋白的抗体和样本孵育，使得抗体特异结合在目标蛋白上，洗涤去除未结合的抗体。然后用 pA-Tn5 转座子和样本孵育，使得 pA-Tn5 蛋白结合在抗体上，从而定位到目标蛋白附近。

（4）转座反应。对上述处理好的样品加上包含 2 价金属离子（例如，镁离子）的 Tn5 酶缓冲液，加入转座反应（37°C，30 分钟-2 小时）。即该反应体系包括：细胞或细胞核或组织；缓冲液。反应完后，用缓冲液对样品进行洗涤，去除未反应的试剂。

（5）连接反应。加入 T4 DNA 连接酶反应缓冲液，连接步骤（1）中的核苷酸标签，T4 DNA 连接酶，核苷酸标签，在适当温度（4°C-37°C）下进行连接反应 20 分钟以上。

反应体系包括：细胞或细胞核或组织（转座反应后的）；T4 DNA 连接酶，核苷酸标签，反应后在连接反应体系中加入过量的游离和核苷酸标签互补序列，封闭多余未反应的核苷酸标签。

（6）提取细胞中的 DNA。对于非固定样品，直接加入裂解液后用 DNA 提取试剂盒，磁珠等方法纯化；对于固定样品，加入蛋白酶 K 反应缓冲液，蛋白酶 K，在 55-65°C 进行解交联后纯化 DNA。

对于纯化的 DNA，1）如果是使用杂合 Tn5 进行，产物双侧都有 PCR 扩增序列，可以直接扩增，获得测序文库。

2）如果是使用单一的 Tn5 进行，那么 DNA 产物只有一侧有 PCR 引物，我们需要对这个 DNA 进行打断和连接，在另一侧加入扩增引物，这可以使用单一的 Tn5 酶进行，也可以使用超声或酶打断，然后末端加 A，连接头，最终获得测序文库。

按照上述步骤，以人 293T 细胞为例，取新鲜细胞，制备细胞核，用杂合 pA-Tn5 进行 CUT Tag 反应，然后连接一个 Illumina 测序文库 P5 端（read1）侧的扩增序列的接头，构建文库，用 read1 引物和杂合 pA-Tn5 中另一个 DNA 片段的 read2 引物扩增产物，最后进行分析。具体步骤如下：

A. pA-Tn5 转座子（transposome）

使如下序列退火形成双链：

10uM Top1 5'-p- AGGCCAGAGCATTCGNNNNNNNAGATGTGTATAAGAGACAG

(SEQ ID NO: 5) (链 A)

10uM Top2 GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG (SEQ ID NO: 3) (链 A)

20uM Bottom 5'p-CTGTCTCTTATACACATCT (SEQ ID NO: 4) (链 B)

之后与 10uM 的 pA-Tn5 酶 (购买自 Vazyme) 室温孵育, 组装成 10uM 浓度的 pA-Tn5 转座子, Top1/ Bottom 双链与 Tn5 形成的转座子为 p-pA-Tn5, Top2/ Bottom 双链与 Tn5 形成的转座子为 pA-Tn5-B。

B. 细胞标记微球制备

1) Bead 上附接的序列如下所示: Bead-S-S-PCR adaptor-barcode1-linker1-barcode2-linker2-barcode3-ligation linker

其中, PCR 接头 (adaptor) 序列为 ACACTCTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 6), 连接序列 1 (Linker1) 为 CGACTCACTACAGGG (SEQ ID NO: 7), 连接序列 2 (Linker2) 序列为 TCGGTGACACGATCG (SEQ ID NO: 8), Ligation linker 序列为 ATCCACGTGCTTGAG (SEQ ID NO: 12)。Barcode1=96 种 5bp 碱基序列, Barcode2=96 种 5bp 碱基序列, Barcode1=96 种 5bp 碱基序列。

2) 合成 3x96 种序列

1. PCR handle-96xbarcode1-linker1, 合成 96 个此序列的反向互补序列;
2. linker1-96xbarcode2-linker2, 合成 96 个此序列的反向互补序列;
3. linker2-96xbarcode3-ligation linker, 合成 96 个此序列的反向互补序列。

3) 微球合成:

合成如下氨基序列: 5'amine-S-S-ACACTCTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 6) 以及 30um 羧基修饰微球 (知益, KBsphere®, www.kbspheres.com/productshow.asp?id=903)。

偶联反应: 微球+50mM EDC+100uM 氨基序列 (SEQ ID NO: 6), 将氨基序列和羧基微球偶联, 获得如下结构: bead-S-S- ACACTCTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 6)

4) 附接标签

将合成的微球均分到 96 孔板, 分别加入 PCR handle-96xbarcode1-linker1, 进行第一轮加标签 (barcoding) 反应。反应体系及过程如下所示: 10ul 微球 + 2ul BstI 缓冲液 + 1ul 10uM dNTP + 1ul 100uM PCR handle-96xbarcode1-linker1, 之后 95℃ 保持 5min, 60℃ 保持 20min;

之后再加入 1ul BstI+5ul H₂O, 60℃保持 60min。

第一轮加标签反应完成后收集所有微球, 混合, 95℃反应 5min 去除互补链, 洗涤, 得到第一轮加标签 (96xbarcode1-linker1) 的微球。然后把微球均分到 96 孔板中, 加入 linker1-96xbarcode2-linker2 (第二轮)、linker2-96xbarcode3-ligation linker (第三轮), 按照第一轮的体系方法进行第二, 三轮加标签反应, 最后得到带有 3 重 barcode 的单链微球, 微球洗涤之后, 跟互补序列 CGAATGCTCTGGCCTCTCAAGCACGTGGAT (SEQ ID NO:9) 退火形成部分双链结构, 最终获得如下附接部分双链结构的微球:

Bead-S-S-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-barcode1-
CGACTCACTACAGGG-barcode2- TCGGTGACACGATCG -barcode3- ATCCACGTGCTTGAG
3'-
TAGGTGCACGAACCTCTCCGGTCTCGTAAGC-5' (SEQ ID NO:9 的反向排列)

C. ATAC 实验

将人 293T 细胞系重悬在裂解液 (10 mM Tris-Cl, pH 7.4; 10 mM NaCl; 3 mM MgCl₂; 0.01% NP-40) 中裂解细胞, 获得细胞核。

取 10 万细胞核与目标蛋白抗体进行孵育, 例如抗组蛋白 H3K4me3 的抗体(Abcam 公司), 结合条件如下 0.05% Digitonin, 20 mM HEPES, pH 7.5, 300 mM NaCl, 0.5 mM Spermidine, 1X Protease inhibitor (Roche) buffer 中, 抗体浓度 1ug/100ul, 在室温下结合 1hr 或者 4 摄氏度过夜结合。

用 0.05% Digitonin, 20 mM HEPES, pH 7.5, 300 mM NaCl, 0.5 mM Spermidine, 1X Protease inhibitor (Roche) buffer 洗涤样品 3 次。

在样品中加入 1ug/100ul 的 pA-Tn5 转座子复合物, buffer 条件如上, 室温孵育 1hr, 用此 buffer 洗涤样品 3 次。

在 buffer 中加入 MgCl₂ 至镁离子浓度 20mM, 37℃ 1hr 进行转座反应, 在此过程中, pA-Tn5 将切割其结合位置相邻的 DNA 并插入其上的 DNA 序列。

反应后用 PBS 洗涤细胞核。

D. 高通量标记

使用如图 8 所示的微流控芯片进行细胞标记, 微球通道 (Bead channel): 100um, 细胞核通道 (Nuclei Channel): 50um。

准备以下溶液:

细胞核溶液 1ml (100 细胞核/ul 浓度), 包括: 200ul 10xT4 DNA ligase Buffer, 10ul T4

DNA ligase, 10ul 1M DTT, 780ul 细胞核/水。

bead 溶液 (100 bead/ul 浓度): Bead in PBS。

细胞核溶液、bead 溶液、油 (FC40 氟碳油, 含有 1% 表面活性剂 FluoroSurfactant, Ran Biotech) 在微流控芯片上形成 120um 直径的液滴 (drop collection), 37℃ 连接 1 小时。

E. 建库

在步骤 D 的液滴中加入等体积全氟辛醇破碎液滴, 离心, 吸取水相, 使用 Qiagen DNA purification kit 纯化水相中的 DNA, 用如下反应体系扩增 DNA 获得最终测序文库: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TrueSeqD501, 1ul 10uM 引物 N701, 1ul Taq, 94℃ 2min, 94℃ 30sec, 55℃ 30sec, 72℃ 30sec, 18 个循环。

引 物 TrueSeqD501 序 列 :
AATGATACGGCGACCAACGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
(SEQ ID NO: 10)

引 物 N701 序 列 :
CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

以上文库投入约 500 细胞, 每个细胞测序 10 万个 PE150 reads, 总数据量为 15G。

抗体使用 Abcam 公司 rabbit-anti-H3K4me3, 图 13 显示的是, Cut tag 文库片段分布结果。图 14 显示的是 Cut tag 片段在转录起始位点分布位置结果图。图 15 显示的是 Cut tag 片段在基因组中分布的比例。图 16 显示的是单细胞 Cut tag 结果分布结果, 单细胞数据叠加后呈现典型的 H3K4me3 组蛋白修饰分布特征, 与多细胞样品实验结果高度相似, 表明了由该方法获得的单细胞数据的真实准确性。

实施例 3 检测细胞或细胞核中转录组

(1) 制备反转录引物。5'端磷酸化的并且与核苷酸标签互补的序列-UMI 分子计数序列-polyT 序列。核苷酸标签的制备同实施例 1 的方式。RT 引物 AGGCCAGAGCATTCGNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 13);

(2) 制备样品。是非固定的细胞或细胞核, 甲醛 (或其他固定剂) 固定的细胞或细胞核, 非固定或固定的组织切片等

固定或非固定样品用包含有去垢剂 (Triton, NP-40, Digitonin, etc.) 的缓冲液处理, 可能包括裂解细胞 (非固定样品) 得到细胞核的中间步骤, 去垢剂裂解或通透细胞和细胞核, 使得

酶等分子生物学试剂可以进入细胞或细胞核。

(3) 反转录。利用步骤(1)的反转录引物,提供反转录酶反应体系,加入链转化模板,对样品进行细胞内反转录反应,反应后细胞/核仍然是独立完整的形态。反应体系及条件如下:细胞/组织,反转录酶缓冲液, RNA 酶抑制剂, dNTP, TSO 链转换引物, 反转录引物; 50-55°C, 5 分钟, 4°C+反转录酶, 42°C。洗涤去除引物和酶体系,对细胞或组织进行核苷酸标签连接反应。结束后加入引物中和多余引物。

(4) 样品后续处理。纯化 mRNA/cDNA: 非固定组织直接纯化 mRNA/cDNA, 固定组织解交联后纯化 mRNA/cDNA; 对 mRNA/cDNA 进行 PCR 扩增 cDNA, 获得 cDNA 文库, 将 cDNA 文库用 Tn5 或者其他 DNA 打断方法构建成测序文库。

具体步骤如下:

制备包含条码序列的核苷酸标签,其被固定于固相支持物上,步骤同上述实施例。

制备转座子复合物,步骤同上述实施例。

制备样品。细胞核: 在 10 mM Tris-Cl, pH 7.4; 10 mM NaCl; 3 mM MgCl₂; 0.01% NP-40 buffer 将组织匀浆,裂解细胞,500g 5min 离心,用 buffer 重悬一次,500g 5min 离心,重悬在上述 buffer 中。

逆转录

设置如下反应,各组分终浓度如下: 1000/ul 细胞核, 1x RT Buffer, 1uM dNTP, 1uM 上述逆转录引物, 1u/ul RNase 酶抑制剂, 1uM TSO 引物 引物序列 (5' - AAGCAGTGGTATCAACGCAGAGTACATrGrGrG (SEQ ID NO: 14) -3', 其中 3 末端的 G 可以是 rG, rG 表示核糖鸟嘌呤, 1unit/ul RT 酶 (Superscript II reverse transcriptase); 反应条件: 50°C 5min, 4°C 5min, 42°C 60min, 用 PBS 洗涤细胞核, 500g 5min 离心洗涤 2 次, 去除未反应的酶和引物。

高通量标记

使用如图 8 所示的微流控芯片进行细胞标记,微球通道 (Bead channel): 100um, 细胞核通道 (Nuclei Channel): 50um。

准备以下溶液:

细胞核溶液 1ml (100 细胞核/ul 浓度), 包括: 200ul 10xT4 DNA ligase Buffer, 10ul T4 DNA ligase, 10ul 1M DTT, 780ul 细胞核/水。

bead 溶液 (100 bead/ul 浓度): Bead in PBS。

细胞核溶液、bead 溶液、油 (FC40 氟碳油, 含有 1% 表面活性剂 FluoroSurfactant, Ran

Biotech) 在微流控芯片上形成 120um 直径的液滴 (drop collection), 37°C 连接 1 小时。

建库

在高通量标记的液滴中加入等体积全氟辛醇破碎液滴, 离心, 吸取水相, 使用 Qiagen DNA purification kit 纯化水相中的 cDNA/mRNA 复合物。

用如下反应体系扩增 DNA 获得最终测序文库: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10uM 引物 ISPCR(AAGCAGTGGTATCAACGCAGAGT (SEQ ID NO: 15)), 1ul Taq, 94°C 2min, 94°C 30sec, 60°C 30sec, 72°C 3min, 18 个循环。用 AMPure XP 磁珠 1:1 体积纯化扩增后 cDNA, 用 Qubit 定量。

测序文库打断

1ng cDNA, 10ul 2xTD Buffer (Illumina Nextera kit), 1ul Nextera enzyme (Illumina Nextera), 20ul 反应体系, 55°C 7min, 加入 5ul Tn5 stop buffer (Nextera kit)。

文库扩增

25ul 以上反应体系, 1ul 10uM 引物 TruseqD501, 1ul 10uM 引物 Nextera N701 引物, 1ul Taq 酶。72°C 5min, 94°C 2min, 94°C 30sec, 60°C 30sec, 72°C 3sec, 18 个循环。用 AMPure XP 磁珠 1:1 体积纯化文库。

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

以上文库投入约 500 细胞, 每个细胞测序 10 万个 PE150 reads, 总数据量为 15G。

两种混合细胞 293T (人), 3T3 (鼠) 后进行单细胞转录组实验, 根据细胞条形码 (cell barcode) 回帖。图 17 显示的是, 单细胞结果清楚区分 2 种细胞的单个细胞。图 18 显示的是每一细胞中检测的转录本及基因数目分布结果。本申请的方法可以用于单细胞的转录组检测。类似地, 可以通过本申请的方法对两种混合细胞 293T (人), 3T3 (鼠) 后进行单细胞基因组实验。根据测得序列比对到人或鼠基因组的比例, 图 19 显示的是, 单细胞结果清楚区分 2 种细胞的单个细胞, 混合纯表示可以从混合的细胞中分离出纯净的人或鼠的来源, 仅有很少部分细胞匹配冲突。图 20 显示的是, 单个人细胞, 基因组覆盖度, 依据染色体排列, 呈现单细胞测序在每个细胞和每个基因组位点有不同的覆盖程度。本申请的方法可以用于单细胞的基因组检测。本申请的方法还可以通过基因组和转录组检测, 用于区分混合细胞中各种细胞的单细胞。

实施例 4 检测细胞中 DNA 序列、数量

(1) 核苷酸标签的制备同实施例 1 的方式。唯一的差别是 5'amine-S-S-

ACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 6) 这一和微球偶联的序列中, 所有的 C 碱基被替换为 5mC 修饰的碱基。

(2) 制备样品。样品: 固定的单细胞或细胞核。样品处理方式: 细胞或细胞核用一定浓度的 SDS 和/或其他去垢剂, 在加热条件下处理一定时间, 把结合在 DNA 上面的蛋白去除掉, 但是并不解开交联, 因此 DNA 还固定在细胞结构中。

(3) 转座反应。对上述处理好的样品加上包含 2 价金属离子例如 (镁离子) 的 Tn5 酶缓冲液, 加入组装好的 Tn5 酶, 对基因组进行转座反应 (37°C, 30 分钟-2 小时)。体系: 细胞/核, Tn5 缓冲液, Tn5 转座复合物, 37°C。之后用缓冲液对样品进行洗涤, 去除未反应的 Tn5 酶。

(4) 连接反应。加入 T4 DNA 连接酶反应缓冲液, 连接步骤 (1) 中的核苷酸标签, T4 DNA 连接酶, 核苷酸标签, 在适当温度 (4°C-37°C) 下进行连接反应 20 分钟以上。

反应体系包括: 细胞或细胞核或组织 (转座反应后的); T4 DNA 连接酶, 核苷酸标签, T4 DNA 连接酶。反应后在连接反应体系中加入过量的游离和核苷酸标签互补序列, 封闭多余未反应的核苷酸标签。

(5) 获取 DNA。加入蛋白酶 K 反应缓冲液, 蛋白酶 K, 在 55-65°C 进行解交联后纯化 DNA, 从而得到标记的全基因组 DNA。后续进行如下处理:

1) 对 DNA 进行直接测序, 得到全基因组序列信息, 包括基因组不同区域的拷贝数信息 (CNV), 或者是点突变信息 (SNV)。

2) 对 DNA 进行 5mC 检测, 例如用亚硫酸氢盐转化法 (Bisulfite conversion) 或者 NEB 酶转化法 (enzymatic conversion) (NEB), 或基于 MspI 酶切的还原亚硫酸氢盐测序 (reduced bisulfite sequence) 等方法检测基因组上 5mC 信息。在对修饰 C 进行转化时, 连接引物则设计为抵御转化的碱基或修饰碱基, 从而保证扩增。

3) 对 DNA 进行 5hmC 检测, 利用 beta-galactose transferase 对 5hmC 位点进行修饰, 用下游方法进行 5hmC 检测。

4) 对其他 DNA 修饰碱基的检测。

具体步骤如下:

样品处理

细胞用 4% 甲醛在 1xPBS 中室温固定 10min, 加入 Glycine 溶液到终浓度 0.1M 室温终止 5min, 用 PBS 洗涤细胞 2 次, 500g 5min 离心沉淀, 固定细胞可在 -80°C 或 -20°C 保存; 细胞在室温融化, 加入 10mM Tris 0.2% SDS 溶液, 42°C 处理 10min; PBS 溶液洗涤 3 次; 取 10 万

细胞核与以上实施例 ATAC 实验中制备的获得的 p-Tn5、Tn5-B 进行反应，反应体系如下所示：

25ul 2xTD Buffer (Illumina) + 2.5ul 10uM p-Tn5 + 2.5ul 10uM Tn5-B + 20ul 细胞核 (10 万个)，37℃反应 30min，PBS 洗涤细胞核。

高通量标记

使用如图 8 所示的微流控芯片进行细胞标记，微球通道 (Bead channel)：100um，细胞核通道 (Nuclei Channel)：50um。

准备以下溶液：

细胞核溶液 1ml (100 细胞核/ul 浓度)，包括：200ul 10xT4 DNA ligase Buffer，10ul T4 DNA ligase，10ul 1M DTT，780ul 细胞核/水。

bead 溶液 (100 bead/ul 浓度)：Bead in PBS。

细胞核溶液、bead 溶液、油 (FC40 氟碳油，含有 1% 表面活性剂 FluoroSurfactant, Ran Biotech) 在微流控芯片上形成 120um 直径的液滴 (drop collection)，37℃连接 1 小时。

基因组建库

高通量标记的液滴中加入等体积全氟辛醇破碎液滴，离心，吸取水相，在水相加入以下试剂到终浓度 1% SDS, Proteinease K 20ug/ml，55℃反应 2hr；使用 Qiagen DNA purification kit 纯化水相中的 DNA，用如下反应体系扩增 DNA 获得最终测序文库：36ul DNA 模板，10ul 5xPCR Buffer，1ul 10mM dNTP，1ul 10uM 引物 TrueseqD501，1ul 10uM 引物 N701，1ul Taq，94℃ 2min，94℃ 30sec，55℃ 30sec，72℃ 30sec，18 个循环。

引物 TrueseqD501 序列：
AATGATACGGCGACCAACGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
(SEQ ID NO: 10)

引物 N701 序列：
CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

单细胞基因组分析

5mC 甲基化测序

对于以上获得的 DNA，首先用 EpiTect Fast Bisulfite Conversion Kit 或 NEB Enzymatic Methylation conversion kit 等转化试剂盒对上述获得的基因组 DNA 进行转化，例如以 Qiagen kit 为例，参照说明书配置 bisulfite conversion 试剂。

上述 DNA, 85ul Bisulfite solution, 35ul DNA protection Buffer, H₂O, 总体积 140ul。

95℃ 5min, 60℃ 10min, 95℃ 5min, 60℃ 10min, 20℃ 保持。参照说明书中步骤柱纯化转化后的 DNA。

DNA 扩增

用如下反应体系扩增 DNA 获得最终测序文库: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TrueSeqD501, 1ul 10uM 引物 N701, 1ul Taq, 94℃ 2min, 94℃ 30sec, 55℃ 30sec, 72℃ 30sec, 18 个循环。

引 物 TrueSeqD501 序 列 :
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
(SEQ ID NO: 10)

引 物 N701 序 列 :
CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

单细胞甲基化分析

5hmC 甲基化测序

使用赛默飞 EpiJET 5-hmC Enrichment Kit 对回收的 DNA 进行 5hmC 富集, 然后建库测序。

回收 DNA, 12.5 μL 4X Enzyme Reaction Buffer, 10ul 5-hmC Modifying Enzyme, 加水至 50ul, 30℃ 反应 1hr。磁珠 1:1 体积纯化 DNA。

40ul 洗脱样本, 10ul 10xbiotin conjugation buffer, 50ul biotin reagent, 50℃ 5min, 加 100ul elution buffer 中止反应, 然后试剂盒柱纯化 DNA。

DNA 扩增

用如下反应体系扩增 DNA 获得最终测序文库: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TrueSeqD501, 1ul 10uM 引物 N701, 1ul Taq, 94℃ 2min, 94℃ 30sec, 55℃ 30sec, 72℃ 30sec, 18 个循环。

引 物 TrueSeqD501 序 列 :
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
(SEQ ID NO: 10)

引 物 N701 序 列 :
CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

两种混合细胞 293T (人), 3T3 (鼠) 后进行单细胞甲基化实验。图 21 显示的是, 根据测得序列比对到人或鼠基因组的比例, 单细胞结果清楚区分 2 种细胞的单个细胞。图 22 显示的是, 单细胞甲基化分布结果, 单细胞数据叠加后与多细胞样品实验结果高度相似, 表明了由该方法获得的单细胞甲基化数据的真实准确性。

本申请的方法还可以用于单细胞 5hmC 测序。如图 23 显示的是, 单细胞的 5hmC 修饰位点的分布结果。本申请获得的单细胞 5hmC 修饰数据具有真实准确性。

实施例 5 同时检测来源于同一细胞的转录组和 ATAC

使用含有相同 5' 末端连接序列的 dT 引物和 Tn5 酶。制备细胞核, 然后对细胞进行 RT (逆转录) 反应, 洗涤去除 RT 反应体系后再进行 Tn5 ATAC 反应, 之后细胞中的 mRNA 和 ATAC 同时被标记。然后进行跟微球上面释放的引物的连接。回收 ATAC 的 DNA 以及 RT 的 mRNA/cDNA 混合物。

对这一混合物用连接接头上的通用引物以及 Tn5 跟 cDNA 特异的引物分别扩增 Tn5 文库和 cDNA 文库, 建库测序。

具体步骤如下:

将人 293T 细胞系重悬在裂解液 (10 mM Tris-Cl, pH 7.4; 10 mM NaCl; 3 mM MgCl₂; 0.01% NP-40) 中裂解细胞, 获得细胞核。

取 10 万细胞核与本申请实施例中获得的 p-Tn5、Tn5-B 进行反应, 反应体系如下所示:

25ul 2xTD Buffer (Illumina), 2.5ul 10uM p-Tn5, 2.5ul 10uM Tn5-B, 20ul 细胞核 (10 万个), 37°C 反应 30min, PBS 洗涤细胞核。

以上获得的细胞核进行如下 RT 反应

1000/ul 细胞核, 1x RT Buffer, 1uM dNTP, 1uM 上述逆转录引物, 1u/ul RNase 酶抑制剂, 1uM TSO 引物 引物序列 (5' -AAGCAGTGGTATCAACGCAGAGTACATrGrGrG (SEQ ID NO: 14) -3', 其中 3 末端的 G 可以是 rG, rG 表示核糖鸟嘌呤, 1unit/ul RT 酶 (Superscript II reverse transcriptase); 反应条件: 50°C 5min, 4°C 5min, 42°C 60min, 用 PBS 洗涤细胞核, 500g 5min 离心洗涤 2 次, 去除未反应的酶和引物。

高通量标记

使用如图 8 所示的微流控芯片进行细胞标记, 微球通道 (Bead channel): 100um, 细胞核通道 (Nuclei Channel): 50um。

准备以下溶液·

细胞核溶液 1ml (100 细胞核/ul 浓度), 包括: 200ul 10xT4 DNA ligase Buffer, 10ul T4 DNA ligase, 10ul 1M DTT, 780ul 细胞核/水。

bead 溶液 (100 bead/ul 浓度): Bead in PBS。

细胞核溶液、bead 溶液、油 (FC40 氟碳油, 含有 1% 表面活性剂 FluoroSurfactant, Ran Biotech) 在微流控芯片上形成 120um 直径的液滴 (drop collection), 37℃ 连接 1 小时。

建库

在高通量标记的液滴中加入等体积全氟辛醇破碎液滴, 离心, 吸取水相, 使用 Qiagen DNA purification kit 纯化水相中的 ATAC DNA 以及 mRNA/cDNA。

文库扩增, 同时扩增 ATAC DNA 以及 mRNA/cDNA

用如下反应体系扩增 DNA 获得最终测序文库: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10uM 引物 N701, 1ul 10mM ISPCR 引物, 1ul Taq, 72℃ 5min, 94℃ 2min, 94℃ 30sec, 55℃ 30sec, 72℃ 3min, 12 个循环。

引 物 TruseqD501 序 列 :
AATGATACGGCGACCAACGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
(SEQ ID NO: 10)

引 物 N701 序 列 :
CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

ISPCR 引物: AAGCAGTGGTATCAACGCAGAGT (SEQ ID NO: 15)

纯化以上混合文库, AMPure bead 1:1 纯化定量。

ATAC 文库扩增

1ng 上述 DNA, DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10uM 引物 N701, 1ul Taq, 94℃ 2min, 94℃ 30sec, 55℃ 30sec, 72℃ 30sec, 18 个循环。AMPure 1:1 纯化文库, 定量, 测序。

cDNA 扩增和测序文库构建

1ng 上述 DNA, DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10mM ISPCR 引物, 1ul Taq 94℃ 2min, 94℃ 30sec, 55℃ 30sec, 72℃ 30sec, 18 个循环。AMPure 1:1 纯化文库, 定量, 测序。

测序文库打断

1ng cDNA, 10ul 2xTD Buffer (Illumina Nextera kit), 1ul Nextera enzyme (Illumina Nextera), 20ul 反应体系, 55℃ 7min。加入 5ul Tn5 stop buffer (Nextera kit)。

文库扩增

25ul 以上反应体系, 1ul 10uM 引物 TrueSeqD501, 1ul 10uM 引物 Nextera N701 引物, 1ul Taq enzyme, 72°C 5min, 94°C 2min, 94°C 30sec, 60°C 30sec, 72°C 3sec, 18 个循环。

用 AMPure XP 磁珠, 1:1 体积纯化文库。

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

同时分析转录组和 ATAC。

本申请用于同时检测同一细胞的转录组和 ATAC, 图 24 显示的是, 根据转录组和 ATAC 基因组均可以很好地区分 2 种细胞中的单个细胞。本申请的方法用于同时检测同一细胞的转录组和 ATAC 具有准确性。

实施例 6 同时检测来源于同一细胞的转录组和 CUT&Tag

具体步骤如下:

将人 293T 细胞系重悬在裂解液(10 mM Tris - Cl, pH 7.4; 10 mM NaCl; 3 mM MgCl₂; 0.01% NP-40) 中裂解细胞, 获得细胞核。

取 10 万细胞核与目标蛋白抗体进行孵育, 例如抗组蛋白 H3K4me3 的抗体(Abcam 公司), 结合条件如下 0.05% Digitonin, 20 mM HEPES, pH 7.5, 300 mM NaCl, 0.5 mM Spermidine, 1X Protease inhibitor (Roche) buffer 中, 抗体浓度 1ug/100ul, 在室温下结合 1hr 或者 4°C 过夜结合。

用 0.05% Digitonin, 20 mM HEPES, pH 7.5, 300 mM NaCl, 0.5 mM Spermidine, 1X Protease inhibitor (Roche) buffer 洗涤样品 3 次。

在样品中加入 1ug/100ul 的 pA-Tn5 转座子复合物, buffer 条件如上, 室温孵育 1hr, 用此 buffer 洗涤样品 3 次。

在 buffer 中加入 MgCl₂ 至镁离子浓度 20mM, 37°C 1hr 进行转座反应, 在此过程中, pA-Tn5 将切割其结合位置相邻的 DNA 并插入其上的 DNA 序列。

反应后用 PBS 洗涤细胞核。

以上获得的细胞核进行如下 RT 反应 各组分终浓度如下: 1000/ul 细胞核, 1x RT Buffer, 1uM dNTP, 1uM 上述逆转录引物, 1u/ul RNase 酶抑制剂, 1uM TSO 引物 引物序列 (5' - AAGCAGTGGTATCAACGCAGAGTACATrGrGrG (SEQ ID NO: 14) -3' , 其中 3 末端的 G 可以是 rG, rG 表示核糖鸟嘌呤, 1unit/ul RT 酶 (Superscript II reverse transcriptase); 反应条件: 50°C 5min, 4°C 5min, 42°C 60min, 用 PBS 洗涤细胞核, 500g 5min 离心洗涤 2 次, 去除未反应的酶和引物。

高通量标记

使用如图 8 所示的微流控芯片进行细胞标记，微球通道（Bead channel）：100um，细胞核通道（Nuclei Channel）：50um。

准备以下溶液：

细胞核溶液 1ml（100 细胞核/ul 浓度），包括：200ul 10xT4 DNA ligase Buffer，10ul T4 DNA ligase，10ul 1M DTT，780ul 细胞核/水。

bead 溶液（100 bead/ul 浓度）：Bead in PBS。

细胞核溶液、bead 溶液、油（FC40 氟碳油，含有 1% 表面活性剂 FluoroSurfactant, Ran Biotech）在微流控芯片上形成 120um 直径的液滴（drop collection），37℃ 连接 1 小时。

建库

在高通量标记的液滴中加入等体积全氟辛醇破碎液滴，离心，吸取水相，使用 Qiagen DNA purification kit 纯化水相中的 ATAC DNA 以及 mRNA/cDNA

文库扩增，同时扩增 ATAC DNA 以及 mRNA/cDNA

用如下反应体系扩增 DNA 获得最终测序文库：36ul DNA 模板，10ul 5xPCR Buffer，1ul 10mM dNTP，1ul 10uM 引物 TrueSeqD501，1ul 10uM 引物 N701，1ul 10mM ISPCR 引物，1ul Taq，72℃ 5min，94℃ 2min，94℃ 30sec，55℃ 30sec，72℃ 3min，12 个循环。

引物 TrueSeqD501 序列：
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
(SEQ ID NO: 10)

引物 N701 序列：
CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

ISPCR 引物：AAGCAGTGGTATCAACGCAGAGT (SEQ ID NO: 15)

纯化以上混合文库，AMPure bead 1:1 纯化定量。

CUT Tag 文库扩增

1ng 上述 DNA

DNA 模板，10ul 5xPCR Buffer，1ul 10mM dNTP，1ul 10uM 引物 TrueSeqD501，1ul 10uM 引物 N701，1ul Taq 94℃ 2min，94℃ 30sec，55℃ 30sec，72℃ 30sec，18 个循环

AMPure 1:1 纯化文库，定量，测序。

cDNA 扩增和测序文库构建

1ng 上述 DNA，DNA 模板，10ul 5xPCR Buffer，1ul 10mM dNTP，1ul 10uM 引物

TrueSeqD501,, 1ul 10mM ISPCR 引物, 1ul Taq 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 30sec, 18 个循环。

AMPure 1:1 纯化文库, 定量, 测序。

测序文库打断

1ng cDNA, 10ul 2xTD Buffer (Illumina Nextera kit), 1ul Nextera enzyme (Illumina Nextera), 20ul 反应体系, 55°C 7min。加入 5ul Tn5 stop buffer (Nextera kit)。

文库扩增

25ul 以上反应体系, 1ul 10uM 引物 TrueSeqD501, 1ul 10uM 引物 Nextera N701 引物, 1ul Taq enzyme, 72°C 5min, 94°C 2min, 94°C 30sec, 60°C 30sec, 72°C 3sec, 18 个循环。

用 AMPure XP 磁珠, 1:1 体积纯化文库。

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

同时分析转录组 cut tag。

本申请用于同时检测同一细胞的转录组和 cut tag, 图 25 显示的是, 根据转录组和 cut tag 组均可以很好地区分 2 种细胞中的单个细胞。本申请的方法用于同时检测同一细胞的转录组和 cut tag 具有准确性。

实施例 7 同时检测来源于同一细胞的转录组和基因组

与单纯检测基因组 DNA 一样处理样品, 先对细胞核进行分离 (strip), 然后进行 Tn5 转座反应, 再进行 RT (逆转录) 反应, 之后按照实施例 5 的方式进行处理。

具体步骤如下:

样品处理

细胞用 4%甲醛在 1xPBS 中室温固定 10min, 加入 Glycine 溶液到终浓度 0.1M 室温终止 5min, 用 PBS 洗涤细胞 2 次, 500g 5min 离心沉淀, 固定细胞可在 -80°C 或 -20°C 保存; 细胞在室温融化, 加入 10mM Tris 0.2% SDS 溶液, 42°C 处理 10min; PBS 溶液洗涤 3 次; 取 10 万细胞核与以上实施例 ATAC 实验中制备的获得的 p-Tn5、Tn5-B 进行反应, 反应体系如下所示:

25ul 2xTD Buffer (Illumina) + 2.5ul 10uM p-Tn5 + 2.5ul 10uM Tn5-B + 20ul 细胞核 (10 万个), 37°C 反应 30min, PBS 洗涤细胞核。

以上获得的细胞核进行如下 RT 反应

1000/ul 细胞核, 1x RT Buffer, 1uM dNTP, 1uM 上述逆转录引物, 1u/ul RNase 酶抑制剂, 1uM TSO 引物 引物序列 (5' -AAGCAGTGGTATCAACGCAGAGTACATrGrGrG (SEQ ID

NO: 14) -3' , 其中 3 末端的 G 可以是 rG, rG 表示核糖鸟嘌呤, 1unit/ul RT 酶 (Superscript II reverse transcriptase); 反应条件: 50℃ 5min, 4℃ 5min, 42℃ 60min, 用 PBS 洗涤细胞核, 500g 5min 离心洗涤 2 次, 去除未反应的酶和引物。

高通量标记

使用如图 8 所示的微流控芯片进行细胞标记, 微球通道 (Bead channel): 100um, 细胞核通道 (Nuclei Channel): 50um。

准备以下溶液:

细胞核溶液 1ml (100 细胞核/ul 浓度), 包括: 200ul 10xT4 DNA ligase Buffer, 10ul T4 DNA ligase, 10ul 1M DTT, 780ul 细胞核/水。

bead 溶液 (100 bead/ul 浓度): Bead in PBS。

细胞核溶液、bead 溶液、油 (FC40 氟碳油, 含有 1% 表面活性剂 FluoroSurfactant, Ran Biotech) 在微流控芯片上形成 120um 直径的液滴 (drop collection), 37℃ 连接 1 小时。

建库

在高通量标记的液滴中加入等体积全氟辛醇破碎液滴, 离心, 吸取水相, 获取 DNA。加入蛋白酶 K 反应缓冲液, 蛋白酶 K, 在 55-65℃ 进行解交联后纯化 DNA, 从而得到标记的全基因组 DNA。后续进行如下处理:

使用 Qiagen DNA purification kit 纯化水相中的基因组 DNA 以及 mRNA/cDNA。

文库扩增, 同时扩增 ATAC DNA 以及 mRNA/cDNA

用如下反应体系扩增 DNA 获得最终测序文库: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TrueSeqD501, 1ul 10uM 引物 N701, 1ul 10mM ISPCR 引物, 1ul Taq, 72℃ 5min, 94℃ 2min, 94℃ 30sec, 55℃ 30sec, 72℃ 3min, 12 个循环。

引物 TrueSeqD501 序列:
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
(SEQ ID NO: 10)

引物 N701 序列:
CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

ISPCR 引物: AAGCAGTGGTATCAACGCAGAGT (SEQ ID NO: 15)

纯化以上混合文库, AMPure bead 1:1 纯化定量。

基因组文库扩增

1ng 上述 DNA, DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物

TruSeqD501, 1ul 10uM 引物 N701, 1ul Taq, 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 30sec, 18 个循环。AMPure 1:1 纯化文库, 定量, 测序。

cDNA 扩增和测序文库构建

1ng 上述 DNA, DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruSeqD501, 1ul 10mM ISPCR 引物, 1ul Taq 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 30sec, 18 个循环。AMPure 1:1 纯化文库, 定量, 测序。

测序文库打断

1ng cDNA, 10ul 2xTD Buffer (Illumina Nextera kit), 1ul Nextera enzyme (Illumina Nextera), 20ul 反应体系, 55°C 7min。加入 5ul Tn5 stop buffer (Nextera kit)。

文库扩增

25ul 以上反应体系, 1ul 10uM 引物 TruSeqD501, 1ul 10uM 引物 Nextera N701 引物, 1ul Taq enzyme, 72°C 5min, 94°C 2min, 94°C 30sec, 60°C 30sec, 72°C 3sec, 18 个循环。

用 AMPure XP 磁珠, 1:1 体积纯化文库。

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

同时分析转录组和基因组

本申请用于同时检测同一细胞的转录组和基因组可以很好地区分 2 种细胞中的单个细胞。在同时检测同一细胞的转录组和 cut tag 方面具有准确性。

实施例 8 同时检测来源于同一细胞的转录组和 DNA 修饰

具体步骤如下:

样品处理

细胞用 4%甲醛在 1xPBS 中室温固定 10min, 加入 Glycine 溶液到终浓度 0.1M 室温终止 5min, 用 PBS 洗涤细胞 2 次, 500g 5min 离心沉淀, 固定细胞可在 -80°C 或 -20°C 保存; 细胞在室温融化, 加入 10mM Tris 0.2% SDS 溶液, 42°C 处理 10min; PBS 溶液洗涤 3 次; 取 10 万细胞核与以上实施例 ATAC 实验中制备的获得的 p-Tn5、Tn5-B 进行反应, 反应体系如下所示:

25ul 2xTD Buffer (Illumina) + 2.5ul 10uM p-Tn5 + 2.5ul 10uM Tn5-B + 20ul 细胞核 (10 万个), 37°C 反应 30min, PBS 洗涤细胞核。

以上获得的细胞核进行如下 RT 反应

1000/ul 细胞核, 1x RT Buffer, 1uM dNTP, 1uM 上述逆转录引物, 1u/ul RNase 酶抑制剂, 1uM TSO 引物 引物序列 (5' -AAGCAGTGGTATCAACGCAGAGTACATrGrGrG (SEQ ID

NO: 14) -3' , 其中 3 末端的 G 可以是 rG, rG 表示核糖鸟嘌呤, 1unit/ul RT 酶 (Superscript II reverse transcriptase); 反应条件: 50°C 5min, 4°C 5min, 42°C 60min, 用 PBS 洗涤细胞核, 500g 5min 离心洗涤 2 次, 去除未反应的酶和引物。

高通量标记

使用如图 8 所示的微流控芯片进行细胞标记, 微球通道 (Bead channel): 100um, 细胞核通道 (Nuclei Channel): 50um。

准备以下溶液:

细胞核溶液 1ml (100 细胞核/ul 浓度), 包括: 200ul 10xT4 DNA ligase Buffer, 10ul T4 DNA ligase, 10ul 1M DTT, 780ul 细胞核/水。

bead 溶液 (100 bead/ul 浓度): Bead in PBS。

细胞核溶液、bead 溶液、油 (FC40 氟碳油, 含有 1% 表面活性剂 FluoroSurfactant, Ran Biotech) 在微流控芯片上形成 120um 直径的液滴 (drop collection), 37°C 连接 1 小时。

建库

在高通量标记的液滴中加入等体积全氟辛醇破碎液滴, 离心, 吸取水相, 获取 DNA。加入蛋白酶 K 反应缓冲液, 蛋白酶 K, 在 55-65°C 进行解交联。

使用 Qiagen DNA purification kit 纯化水相中的 DNA 以及 mRNA/cDNA

将以上文库分成 2 份, 分别进行甲基化测序和转录组测序

对于甲基化文库, 进行 Bisulfite 测序或者 5hmC 测序

甲基化测序

对于以上获得的 DNA, 首先用 EpiTect Fast Bisulfite Conversion Kit 或 NEB Enzymatic Methylation conversion kit 等转化试剂盒对上述获得的基因组 DNA 进行转化, 例如以 Qiagen kit 为例, 参照说明书配置 bisulfite conversion 试剂。

上述 DNA, 85ul Bisulfite solution, 35ul DNA protection Buffer, H₂O, 总体积 140ul。

95°C 5min, 60°C 10min, 95°C 5min, 60°C 10min, 20°C 保持。参照说明书中步骤柱纯化转化后的 DNA。

DNA 扩增

用如下反应体系扩增 DNA 获得最终测序文库: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10uM 引物 N701, 1ul Taq, 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 30sec, 18 个循环。

引 物 TruseqD501 序 列 :

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

(SEQ ID NO: 10)

引 物 N701 序 列 :

CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

5hmC 甲基化测序, 使用赛默飞 EpiJET 5-hmC Enrichment Kit 对回收的 DNA 进行 5hmC 富集, 然后建库测序。

回收 DNA, 12.5 μ L 4X Enzyme Reaction Buffer, 10ul 5-hmC Modifying Enzyme, 加水至 50ul, 30°C 反应 1hr。磁珠 1:1 体积纯化 DNA。

40ul 洗脱样本, 10ul 10xbiotin conjugation buffer, 50ul biotin reagent, 50°C 5min, 加 100ul elution buffer 中止反应, 然后试剂盒柱纯化 DNA。

DNA 扩增

用如下反应体系扩增 DNA 获得最终测序文库: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10uM 引物 N701, 1ul Taq, 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 30sec, 18 个循环。

引 物 TruseqD501 序 列 :

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

(SEQ ID NO: 10)

引 物 N701 序 列 :

CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

DNA 扩增

用如下反应体系扩增 DNA 获得最终测序文库: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10uM 引物 N701, 1ul Taq, 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 30sec, 18 个循环。

引 物 TruseqD501 序 列 :

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

(SEQ ID NO: 10)

引 物 N701 序 列 :

CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

cDNA 扩增和测序文库构建

上述回收的 DNA 及 cDNA/mRNA, DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10mM ISPCR 引物, 1ul Taq 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 30sec, 18 个循环。

AMPure 1:1 纯化文库, 定量, 测序。

测序文库打断

1ng cDNA, 10ul 2xTD Buffer (Illumina Nextera kit), 1ul Nextera enzyme (Illumina Nextera), 20ul 反应体系, 55°C 7min。加入 5ul Tn5 stop buffer (Nextera kit)。

文库扩增

25ul 以上反应体系, 1ul 10uM 引物 TruseqD501, 1ul 10uM 引物 Nextera N701 引物, 1ul Taq enzyme, 72°C 5min, 94°C 2min, 94°C 30sec, 60°C 30sec, 72°C 3sec, 18 个循环。

用 AMPure XP 磁珠, 1:1 体积纯化文库。

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

同时分析转录组及甲基化。

本申请用于同时检测同一细胞的转录组和甲基化, 图 26 显示的是, 同一细胞的转录组和甲基化组均可以很好地与基因模型以及已知的甲基化位点进行匹配。本申请的方法用于同时检测同一细胞的转录组和甲基化具有准确性。

实施例 9 空间多组学技术平台

(1) 空间点阵芯片: 芯片上是固定间隔的 DNA oligo 簇, 结构如下:

Slide-Surface (玻片表面) - 释放 linker-PCR adaptor-barcode-连接臂

芯片跟互补的单链杂交, 将 oligo 点阵变成如下结构:

Slide-Surface (玻片表面) - 释放 linker-PCR adaptor-barcode-连接臂

连接臂---互补链

空间点阵用 microarray 原位合成方法 (Affymetrix, NimbleGene) 或者其他方法合成, 包括从已有的 array 上面用 PCR 方法转移, 用顺序标记法延伸等。

(2) 组织切片制备: 将非固定组织冰冻切片贴在盖玻片上, 加 1%甲醛, 固定组织, 洗涤。

(3) 通透处理: 用含有去垢剂的缓冲液处理组织。

(4) 在组织上方加入反转录反应 mix, 使用带有 5' 磷酸修饰, 5' 延伸可以跟芯片上的

oligo 互补的反转录引物进行原位 RT（反转录）反应。

（5）洗掉反转录反应体系，在 slide 上加带有 5' 磷酸修饰的 Tn5 酶，进行原位 ATAC 反应。

（6）洗掉 ATAC 反应体系，在组织上方加入 DNA 连接酶 buffer，DNA 连接酶，然后把组织贴到 DNA oligo 点阵上，二者紧密接触。DNA oligo 从载玻片上释放，转移到组织切片上进行连接反应，标记 cDNA 和 Tn5 产物。

（7）反应结束后，对组织进行成像。

（8）终止反应，用蛋白酶消化组织，回收 DNA，按照前述实施例的方式对 cDNA 和 ATAC DNA 建库测序。

具体步骤如下：

使用 Affymetrix 公司技术，在玻璃/硅机制上合成 100x100，大小为 5um，间隔 5um 的引物点阵，总面积为 1cm x 1cm，共一万个 DNA oligo 点阵，图 27 显示的是一种空间点阵芯片，DNA 点阵可以有规律排布的 dT 引物阵列，并且和 FAM-AAAAAAAAAAAAAAAAAAAAAAAAA（SEQ ID NO: 17）引物杂交。具体点阵 DNA 序列为：

S-S-ACACTCTTTCCCTACACGACGCTCT（SEQ ID NO: 16）-NNNNNNNN-ATCCACGTGCTTGAG（SEQ ID NO: 12）

点阵 DNA 序列中 NNNNNNNN 为 8bp 的特定引物序列，点阵上每一个点对应一个特定的 8bp 序列。

以上玻片上方加入 CGAATGCTCTGGCCTCTCAAGCACGTGGAT（SEQ ID NO: 9）引物，在 1M NaCl, 10mM Tris 溶液中，和玻璃室温杂交 1hr，使得点阵上的引物退火为部分双链引物。

将 OCT 包埋的组织用冰冻切片机切片，贴附到多聚赖氨酸表面处理的玻片上。

用 1% 甲醛固定组织，室温 10 分钟，用 PBS 清洗玻片。

玻片上组织用裂解液处理（10 mM Tris-Cl, pH 7.4; 10 mM NaCl; 3 mM MgCl₂; 0.01% NP-40），室温 5min。

用本申请实施例中获得 p-Tn5、Tn5-B 对玻片进行反应，反应体系如下所示：

25ul 2xTD Buffer (Illumina), 2.5ul 10uM p-Tn5, 2.5ul 10uM Tn5-B, 20ul 细胞核（10 万个），37℃ 反应 30min，PBS 洗涤切片。

对切片进行 RT 反应

1000/ul 细胞核，1x RT Buffer, 1uM dNTP, 1uM 上述逆转录引物，1u/ul RNase 酶抑制

剂, 1uM TSO 引物 引物序列 (5' -AAGCAGTGGTATCAACGCAGAGTACATrGrGrG (SEQ ID NO: 14) -3' , 其中 3 末端的 G 可以是 rG, rG 表示核糖鸟嘌呤, 1unit/ul RT 酶 (Superscript II reverse transcriptase); 反应条件: 50°C 5min, 4°C 5min, 42°C 60min, 用 PBS 洗涤切片。

将反应后组织和合成的引物点阵玻片接触, 加入 1xT4 ligase buffer, 1unit/ul T4 DNA 连接酶, 使得玻片上部分双链形式的 adaptor 与组织切片上的 RT 产物及 AATC 产物进行连接反应。

回收 cDNA 及 ATACDNA。在切片上方加入蛋白酶 K 反应缓冲液, 蛋白酶 K, 在 55-65°C 进行解交联后纯化 DNA, 然后用 Qiagen kit 纯化获得基因组 DNA 以及逆转录的 mRNA/cDNA 使用 Qiagen DNA purification kit 纯化水相中的 ATAC DNA 以及 mRNA/cDNA。

文库扩增, 同时扩增 ATAC DNA 以及 mRNA/cDNA

用如下反应体系扩增 DNA 及 cDNA: 36ul DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10uM 引物 N701, 1ul 10mM ISPCR 引物, 1ul Taq, 72°C 5min, 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 3min, 12 个循环。

引物 TruseqD501 序列:
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
(SEQ ID NO: 10)

引物 N701 序列:
CAAGCAGAAGACGGCATACGAGATATCGGCTAGTCTCGTGGGCTCGG (SEQ ID NO: 11)

ISPCR 引物: AAGCAGTGGTATCAACGCAGAGT (SEQ ID NO: 15)

纯化以上混合文库, AMPure bead 1:1 纯化定量。

ATAC 文库扩增

1ng 上述 DNA, DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10uM 引物 N701, 1ul Taq, 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 30sec, 18 个循环。AMPure 1:1 纯化文库, 定量, 测序。

cDNA 扩增和测序文库构建

1ng 上述 DNA, DNA 模板, 10ul 5xPCR Buffer, 1ul 10mM dNTP, 1ul 10uM 引物 TruseqD501, 1ul 10mM ISPCR 引物, 1ul Taq 94°C 2min, 94°C 30sec, 55°C 30sec, 72°C 30sec, 18 个循环。AMPure 1:1 纯化文库, 定量, 测序。

测序文库打断

1ng cDNA, 10ul 2xTD Buffer (Illumina Nextera kit), 1ul Nextera enzyme (Illumina Nextera),

20ul 反应体系，55℃ 7min。加入 5ul Tn5 stop buffer (Nextera kit)。

文库扩增

25ul 以上反应体系，1ul 10uM 引物 TruseqD501，1ul 10uM 引物 Nextera N701 引物，1ul Taq enzyme，72℃ 5min, 94℃ 2min, 94℃ 30sec, 60℃ 30sec, 72℃ 3sec, 18 个循环。

用 AMPure XP 磁珠，1:1 体积纯化文库。

Illumina Novaseq 每个细胞测 10 万个 PE150 读长 (reads)。

同时分析转录组和基因组。

图 28 显示的是，切片 HE 染色与空间点阵芯片叠加，每个圆点的颜色深浅度表示测量获得的基因数目。本申请的方法可以用于空间多组学技术平台的研究。

前述详细说明是以解释和举例的方式提供的，并非要限制所附权利要求的范围。目前本申请所列举的实施方式的多种变化对本领域普通技术人员来说是显而易见的，且保留在所附的权利要求和其等同方案的范围内。

权 利 要 求 书

1. 一种分析来自细胞的目标核酸的方法，所述方法包括：
 - a) 提供包含下述的离散分区：
 - i. 源于单个细胞的目标核酸，其中至少部分所述目标核酸被添加了寡核苷酸衔接子序列而成为经附接的目标核酸；以及
 - ii. 附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3'端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第一链与所述第二链形成部分双链的结构或者所述第二链与所述经附接的目标核酸形成部分双链的结构；
 - b) 在所述离散分区中，使所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸。
2. 根据权利要求 1 所述的方法，其中所述寡核苷酸标签可释放地附接至所述固体支持物。
3. 根据权利要求 1-2 中任一项所述的方法，其包括从所述固体支持物上释放所述至少一个寡核苷酸标签，并在 b) 中使经释放的所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸。
4. 根据权利要求 1-3 中任一项所述的方法，其中所述寡核苷酸标签通过其第一链的 5'端直接或间接附接至所述固体支持物。
5. 根据权利要求 1-4 中任一项所述的方法，其中所述离散分区中还包含连接酶，且所述连接酶使所述寡核苷酸标签与所述经附接的目标核酸连接。
6. 根据权利要求 5 所述的方法，其中所述连接酶包括 T4 连接酶。
7. 根据权利要求 1-6 中任一项所述的方法，其中在所述条码化的目标核酸中，所述目标核酸序列位于所述条码序列的 3'端。
8. 根据权利要求 1-7 中任一项所述的方法，其中所述固体支持物为珠粒。
9. 根据权利要求 8 所述的方法，其中所述珠粒为磁性珠粒。
10. 根据权利要求 1-9 中任一项所述的方法，其中所述离散分区为孔或微滴。
11. 根据权利要求 1-10 中任一项所述的方法，其中所述条码序列包含细胞条码序列，且附接至同一个固体支持物上的各寡核苷酸标签所包含的细胞条码序列相同。
12. 根据权利要求 11 所述的方法，其中所述细胞条码序列包含由连接子序列间隔开的至少 2 个细胞条码区段。
13. 根据权利要求 1-12 中任一项所述的方法，其中 a) 包括将所述源于单个细胞的目标核酸与所

述附接有至少一个寡核苷酸标签的固体支持物共分配至所述离散分区中。

14. 根据权利要求 1-13 中任一项所述的方法，其中 b) 包括使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。
15. 根据权利要求 1-14 中任一项所述的方法，其中 b) 包括使所述寡核苷酸标签的第二链的所述第二部分与附接至所述目标核酸的所述寡核苷酸衔接子杂交，以及使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。
16. 根据权利要求 1-15 中任一项所述的方法，其中所述经附接的目标核酸中包含独特分子鉴别区。
17. 根据权利要求 16 所述的方法，其中所述独特分子鉴别区位于所述寡核苷酸衔接子序列与所述目标核酸序列之间。
18. 根据权利要求 1-17 中任一项所述的方法，其中所述寡核苷酸标签还包含扩增引物识别区。
19. 根据权利要求 18 所述的方法，其中所述扩增引物识别区为通用扩增引物识别区。
20. 根据权利要求 1-19 中任一项所述的方法，其进一步包括：
 - c) 获得所述条码化的目标核酸的表征结果；以及
 - d) 至少部分基于 c) 中获得的所述表征结果中存在相同的所述细胞条码序列而将所述目标核酸的序列识别为源于所述单个细胞。
21. 根据权利要求 20 所述的方法，其进一步包括，在 b) 之后并且在 c) 之前，从所述离散分区中释放所述条码化的目标核酸。
22. 根据权利要求 20-21 中任一项所述的方法，其中 c) 包括对所述条码化的目标核酸进行测序，从而获得所述表征结果。
23. 根据权利要求 20-22 中任一项所述的方法，其进一步包括由所述条码化的目标核酸的序列组装所述单个细胞的基因组的至少一部分的连续核酸序列。
24. 根据权利要求 23 所述的方法，其中基于所述单个细胞的所述基因组的至少一部分的所述核酸序列来表征所述单个细胞。
25. 根据权利要求 1-24 中任一项所述的方法，其中每个所述离散分区至多包括源自单个细胞的所述目标核酸。
26. 根据权利要求 20-25 中任一项所述的方法，其进一步包括至少部分基于所述独特分子鉴别区的存在将所述条码化的目标核酸中的单个核酸序列鉴别为源于所述目标核酸中的给定核酸。
27. 根据权利要求 20-26 中任一项所述的方法，所述目标核酸包括外源核酸，所述外源核酸包括

与蛋白、脂类和/或小分子化合物连接的外源核酸，所述蛋白、脂类和/或小分子化合物能够与细胞内的靶分子结合。

28. 根据权利要求 27 所述的方法，其进一步包括基于所述独特分子鉴别区的存在确定所述目标核酸中给定核酸的量。
29. 根据权利要求 1-28 中任一项所述的方法，其包括在 a) 之前对所述细胞进行预处理。
30. 根据权利要求 29 所述的方法，其中所述预处理包括固定所述细胞。
31. 根据权利要求 30 所述的方法，其中使用固定剂对所述细胞进行固定，所述固定剂选自下组中的一种或多种：甲醛、多聚甲醛、甲醇、乙醇、丙酮、戊二醛、锇酸和重铬酸钾。
32. 根据权利要求 29-31 中任一项所述的方法，其中所述预处理包括使所述细胞的细胞核被暴露。
33. 根据权利要求 29-32 中任一项所述的方法，其中所述预处理包括使用去垢剂处理所述细胞，所述去垢剂包括 Triton、NP-40 和/或 digitonin。
34. 根据权利要求 1-33 中任一项所述的方法，其中所述目标核酸包括选自下组的一种或多种：DNA、RNA 和 cDNA。
35. 根据权利要求 20-34 中任一项所述的方法，其进一步包括，在 b) 之后并且在 c) 之前，对所述条码化的目标核酸进行扩增。
36. 根据权利要求 35 所述的方法，其包括在 b) 之后并且在 c) 之前，从所述离散分区中释放所述条码化的目标核酸，且所述扩增在所述条码化的目标核酸从所述离散分区中释放后进行。
37. 根据权利要求 35-36 中任一项所述的方法，其中所述扩增中使用扩增引物，且所述扩增引物中包含随机引导序列。
38. 根据权利要求 37 所述的方法，其中所述随机引导序列为随机六聚体。
39. 根据权利要求 35-38 中任一项所述的方法，其中所述扩增包括使所述随机引导序列与所述条码化的目标核酸至少部分杂交并且以模板定向的方式延伸所述随机引导序列。
40. 根据权利要求 1-39 中任一项所述的方法，其包括使至少一部分所述目标核酸从所述离散分区中的所述单个细胞中释放到细胞外，并在 b) 中使经释放的所述目标核酸与所述寡核苷酸标签连接，从而产生条码化的目标核酸。
41. 根据权利要求 1-40 中任一项所述的方法，其包括使至少一部分从所述固体支持物释放的所述寡核苷酸标签进入所述单个细胞中，并在 b) 中与所述目标核酸连接，从而产生条码化的目标核酸。
42. 根据权利要求 1-41 中任一项所述的方法，其包括使用微流控装置将所述源于单个细胞的目标核酸与所述附接有至少一个寡核苷酸标签的固体支持物共分配至所述离散分区中。

43. 根据权利要求 42 所述的方法，其中所述离散分区为微滴，且所述微流控装置为微滴发生器。
44. 根据权利要求 42-43 中任一项所述的方法，其中所述微流控装置包括第一输入通道和第二输入通道，它们在与输出通道流体连接的接合处汇合。
45. 根据权利要求 44 所述的方法，其中所述方法还包括将包含所述目标核酸的样品引入所述第一输入通道，且将附接有至少一个寡核苷酸标签的所述固体支持物引入所述第二输入通道，从而在所述输出通道中生成所述样品与所述固体支持物的混合物。
46. 根据权利要求 45 所述的方法，其中所述输出通道与第三输入通道在接合处流体连接。
47. 根据权利要求 46 所述的方法，其还包括将油引入所述第三输入通道，使得形成油包水乳液内的水性小滴作为所述离散分区。
48. 根据权利要求 47 所述的方法，其中每个所述离散分区中至多包含来自单个细胞的所述目标核酸。
49. 根据权利要求 44-48 中任一项所述的方法，其中所述第一输入通道和所述第二输入通道彼此之间形成基本上垂直的角度。
50. 根据权利要求 1-49 中任一项所述的方法，其中所述目标核酸包括源自所述单个细胞中 RNA 的 cDNA。
51. 根据权利要求 50 所述的方法，其中所述 RNA 包括 mRNA。
52. 根据权利要求 34-51 中任一项所述的方法，其包括在 a) 之前对所述 RNA 进行反转录，并产生所述经附接的目标核酸。
53. 根据权利要求 52 所述的方法，其中所述反转录中使用反转录引物，所述反转录引物以 5' 至 3' 的方向包含所述寡核苷酸衔接子序列以及 polyT 序列。
54. 根据权利要求 53 所述的方法，其中所述反转录包括使所述 polyT 序列与所述 RNA 杂交并且以模板定向的方式延伸所述 polyT 序列。
55. 根据权利要求 1-54 中任一项所述的方法，其中所述目标核酸包括源自所述单个细胞的 DNA。
56. 根据权利要求 55 所述的方法，其中所述 DNA 包括基因组 DNA、开放染色质 DNA、蛋白质结合的 DNA 区域和/或与蛋白、脂类和/或小分子化合物连接的外源核酸，所述蛋白、脂类和/或小分子化合物能够与细胞内的靶分子结合。
57. 根据权利要求 56 所述的方法，其包括在 a) 之前对源自单个细胞的所述 DNA 进行片段化。
58. 根据权利要求 57 所述的方法，其中在所述片段化之后或者在所述片段化的过程中产生所述经附接的目标核酸。
59. 根据权利要求 57-58 中任一项所述的方法，其中所述片段化包括使用超声断裂，而后在经断

裂的所述 DNA 上添加包含所述寡核苷酸衔接子的序列，从而获得所述经附接的目标核酸。

60. 根据权利要求 57-59 中任一项所述的方法，其中所述片段化包括使用 DNA 内切酶、外切酶打断，而后在经断裂的所述 DNA 上添加包含所述寡核苷酸衔接子的序列，从而获得所述经附接的目标核酸。
61. 根据权利要求 57-60 中任一项所述的方法，其中所述片段化包括使用转座酶-核酸复合物将包含所述寡核苷酸衔接子的序列整合到所述 DNA 中，并释放所述转座酶以获得所述经附接的目标核酸。
62. 根据权利要求 61 所述的方法，其中所述转座酶-核酸复合物包含转座酶以及转座子末端核酸分子，其中所述转座子末端核酸分子包含所述寡核苷酸衔接子序列。
63. 根据权利要求 61-62 中任一项所述的方法，其中所述转座酶包括 Tn5。
64. 根据权利要求 61-63 中任一项所述的方法，其中所述 DNA 包括与蛋白质结合的 DNA 区域，且所述转座酶-核酸复合物中还包含直接或间接识别所述蛋白质的部分。
65. 根据权利要求 64 所述的方法，其中所述直接或间接识别所述蛋白质的部分包括下组中的一种或多种：特异性结合所述蛋白质的抗体和蛋白质 A 或蛋白质 G。
66. 一种组合物，其包含：多个固体支持物，每个所述固体支持物上附接有至少一个寡核苷酸标签，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3' 端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与所述待测核酸中的序列互补的第二部分，且所述第一链与所述第二链形成部分双链的结构或者所述第二链与所述经附接的目标核酸形成部分双链的结构；所述寡核苷酸标签的条码序列包含共同条码结构域和可变结构域，所述共同条码结构域在附接于同一个固体支持物的寡核苷酸标签中是相同的，且所述共同条码结构域在所述多个固体支持物中的两个或更多个固体支持物之间是不同的。
67. 一种用于分析来自细胞的目标核酸的试剂盒，其包含权利要求 66 所述的组合物。
68. 根据权利要求 67 所述的试剂盒，其包括转座酶。
69. 根据权利要求 67-68 中任一项所述的试剂盒，其进一步包含核酸扩增剂，逆转录剂，固定剂，通透剂，连接剂和裂解剂中的至少一种。
70. 一种扩增来自细胞的目标核酸的方法，所述方法包括：
 - a) 提供包含下述的离散分区：i. 源于单个细胞的目标核酸，其中至少部分所述目标核酸被添加了寡核苷酸衔接子序列而成为经附接的目标核酸；以及 ii. 附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列

以及位于所述条码序列 3'端的杂交序列,所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分,且所述第一链与所述第二链形成部分双链的结构或者所述第二链与所述经附接的目标核酸形成部分双链的结构;

b) 在所述离散分区中,使所述寡核苷酸标签与所述经附接的目标核酸连接,从而产生条码化的目标核酸;以及

c) 对所述条码化的目标核酸进行扩增。

71. 根据权利要求 70 所述的方法,其中所述寡核苷酸标签可释放地附接至所述固体支持物。
72. 根据权利要求 71 所述的方法,其包括从所述固体支持物上释放所述至少一个寡核苷酸标签,并在 b) 中使经释放的所述寡核苷酸标签与所述经附接的目标核酸连接,从而产生条码化的目标核酸。
73. 根据权利要求 70-72 中任一项所述的方法,其中所述寡核苷酸标签通过其第一链的 5'端直接或间接附接至所述固体支持物。
74. 根据权利要求 70-73 中任一项所述的方法,其中所述离散分区中还包含连接酶,且所述连接酶使所述寡核苷酸标签与所述经附接的目标核酸连接。
75. 根据权利要求 74 所述的方法,其中所述连接酶包括 T4 连接酶。
76. 根据权利要求 70-75 中任一项所述的方法,其中在所述条码化的目标核酸中,所述目标核酸序列位于所述条码序列的 3'端。
77. 根据权利要求 70-76 中任一项所述的方法,其中所述固体支持物为珠粒。
78. 根据权利要求 70-77 中任一项所述的方法,其中所述离散分区为孔或微滴。
79. 根据权利要求 70-78 中任一项所述的方法,其中所述条码序列包含细胞条码序列,且附接至同一个固体支持物上的各寡核苷酸标签所包含的细胞条码序列相同。
80. 根据权利要求 79 所述的方法,其中所述细胞条码序列包含由连接子序列间隔开的至少 2 个细胞条码区段。
81. 根据权利要求 70-80 中任一项所述的方法,其中 a) 包括将所述源于单个细胞的目标核酸与所述附接有至少一个寡核苷酸标签的固体支持物共分配至所述离散分区中。
82. 根据权利要求 70-81 中任一项所述的方法,其中 b) 包括使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接,从而产生所述条码化的目标核酸。
83. 根据权利要求 70-82 中任一项所述的方法,其中 b) 包括使所述寡核苷酸标签的第二链的所述

第二部分与附接至所述目标核酸的所述寡核苷酸衔接子杂交，以及使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。

84. 根据权利要求 70-83 中任一项所述的方法，其中所述经附接的目标核酸中包含独特分子鉴别区。
85. 根据权利要求 84 所述的方法，其中所述独特分子鉴别区位于所述寡核苷酸衔接子序列与所述目标核酸序列之间。
86. 根据权利要求 70-85 中任一项所述的方法，其中所述寡核苷酸标签还包含扩增引物识别区。
87. 根据权利要求 86 所述的方法，其中所述扩增引物识别区为通用扩增引物识别区。
88. 根据权利要求 87 所述的方法，其包括在 b) 之后并且在 c) 之前，从所述离散分区中释放所述条码化的目标核酸，且所述扩增在所述条码化的目标核酸从所述离散分区中释放后进行。
89. 根据权利要求 70-88 中任一项所述的方法，其中所述扩增中使用扩增引物，且所述扩增引物中包含随机引导序列。
90. 根据权利要求 89 所述的方法，其中所述随机引导序列为随机六聚体。
91. 根据权利要求 70-90 中任一项所述的方法，其中所述扩增包括使所述随机引导序列与所述条码化的目标核酸至少部分杂交并且以模板定向的方式延伸所述随机引导序列。
92. 一种对来自细胞的目标核酸进行测序的方法，所述方法包括：
 - a) 提供包含下述的离散分区：i. 源于单个细胞的目标核酸，其中至少部分所述目标核酸被添加了寡核苷酸衔接子序列而成为经附接的目标核酸；以及 ii. 附接有至少一个寡核苷酸标签的固体支持物，其中每个所述寡核苷酸标签包含第一链以及第二链，所述第一链包含条码序列以及位于所述条码序列 3' 端的杂交序列，所述第二链包含与所述第一链的所述杂交序列互补的第一部分以及与附接至所述目标核酸的所述寡核苷酸衔接子序列互补的第二部分，且所述第一链与所述第二链形成部分双链的结构或者所述第二链与所述经附接的目标核酸形成部分双链的结构；
 - b) 在所述离散分区中，使所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的目标核酸；以及
 - c) 对所述条码化的目标核酸进行测序。
93. 根据权利要求 92 所述的方法，其中所述寡核苷酸标签可释放地附接至所述固体支持物。
94. 根据权利要求 93 所述的方法，其包括从所述固体支持物上释放所述至少一个寡核苷酸标签，并在 b) 中使经释放的所述寡核苷酸标签与所述经附接的目标核酸连接，从而产生条码化的

目标核酸。

95. 根据权利要求 92-94 中任一项所述的方法，其中所述寡核苷酸标签通过其第一链的 5'端直接或间接附接至所述固体支持物。
96. 根据权利要求 92-95 中任一项所述的方法，其中所述离散分区中还包含连接酶，且所述连接酶使所述寡核苷酸标签与所述经附接的目标核酸连接。
97. 根据权利要求 96 所述的方法，其中所述连接酶包括 T4 连接酶或 T7 连接酶。
98. 根据权利要求 92-97 中任一项所述的方法，其中在所述条码化的目标核酸中，所述目标核酸序列位于所述条码序列的 3'端。
99. 根据权利要求 92-98 中任一项所述的方法，其中所述固体支持物为珠粒。
100. 根据权利要求 92-99 中任一项所述的方法，其中所述离散分区为孔或微滴。
101. 根据权利要求 92-100 中任一项所述的方法，其中所述条码序列包含细胞条码序列，且附接至同一个固体支持物上的各寡核苷酸标签所包含的细胞条码序列相同。
102. 根据权利要求 101 所述的方法，其中所述细胞条码序列包含由连接子序列间隔开的至少 2 个细胞条码区段。
103. 根据权利要求 92-102 中任一项所述的方法，其中 a) 包括将所述源于单个细胞的目标核酸与所述附接有至少一个寡核苷酸标签的固体支持物共分配至所述离散分区中。
104. 根据权利要求 92-103 中任一项所述的方法，其中 b) 包括使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。
105. 根据权利要求 92-104 中任一项所述的方法，其中 b) 包括使所述寡核苷酸标签的第二链的所述第二部分与附接至所述目标核酸的所述寡核苷酸衔接子杂交，以及使所述寡核苷酸标签的第一链的所述杂交序列与附接至所述目标核酸的所述寡核苷酸衔接子连接，从而产生所述条码化的目标核酸。
106. 根据权利要求 92-105 中任一项所述的方法，其中所述经附接的目标核酸中包含独特分子鉴别区。
107. 根据权利要求 106 所述的方法，其中所述独特分子鉴别区位于所述寡核苷酸衔接子序列与所述目标核酸序列之间。
108. 根据权利要求 92-107 中任一项所述的方法，其中所述寡核苷酸标签还包含扩增引物识别区。
109. 根据权利要求 108 所述的方法，其中所述扩增引物识别区为通用扩增引物识别区。
110. 根据权利要求 92-109 中任一项所述的方法，其进一步包括由所述条码化的目标核酸的序列组

装所述单个细胞的基因组的至少一部分的连续核酸序列。

111. 根据权利要求 110 所述的方法，其中基于所述单个细胞的所述基因组的至少一部分的所述核酸序列来表征所述单个细胞。
112. 根据权利要求 92-111 中任一项所述的方法，其中每个所述离散分区至多包括源自单个细胞的所述目标核酸。
113. 根据权利要求 92-112 中任一项所述的方法，其进一步包括至少部分基于所述独特分子鉴别区的存在将所述条码化的目标核酸中的单个核酸序列鉴别为源于所述目标核酸中的给定核酸。
114. 根据权利要求 113 所述的方法，其进一步包括基于所述独特分子鉴别区的存在确定所述目标核酸中给定核酸的量。

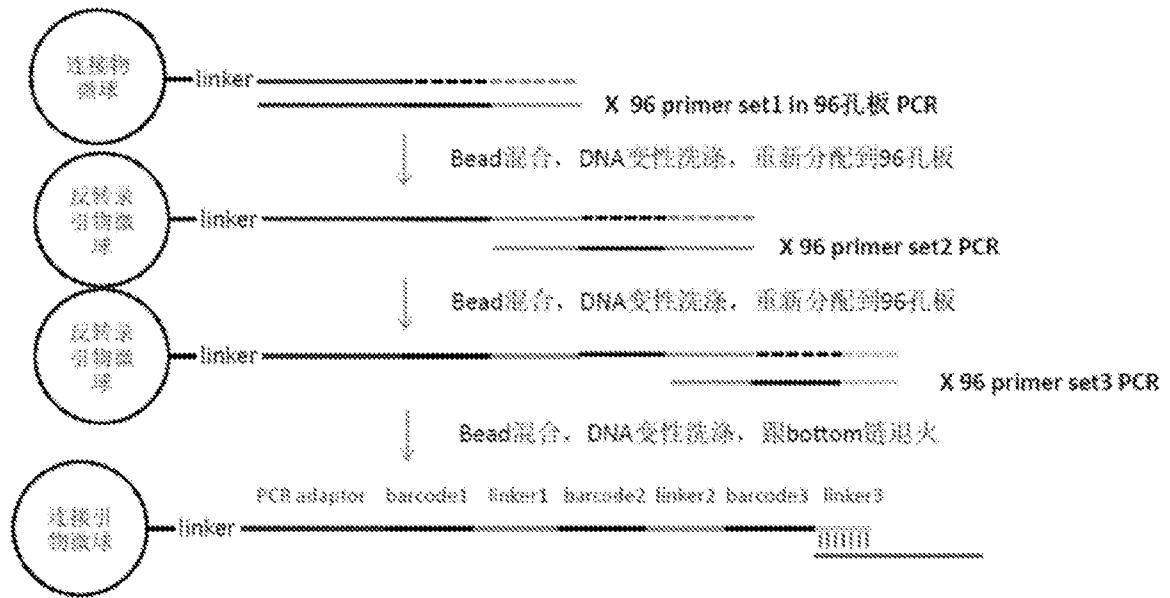


图 1

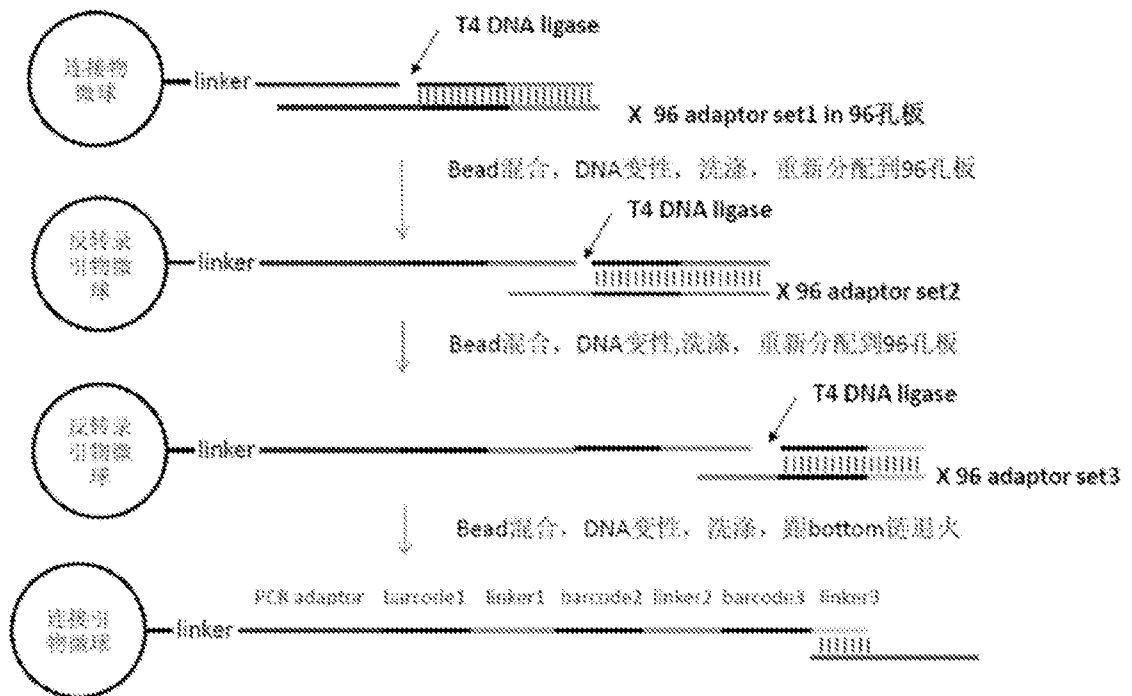


图 2

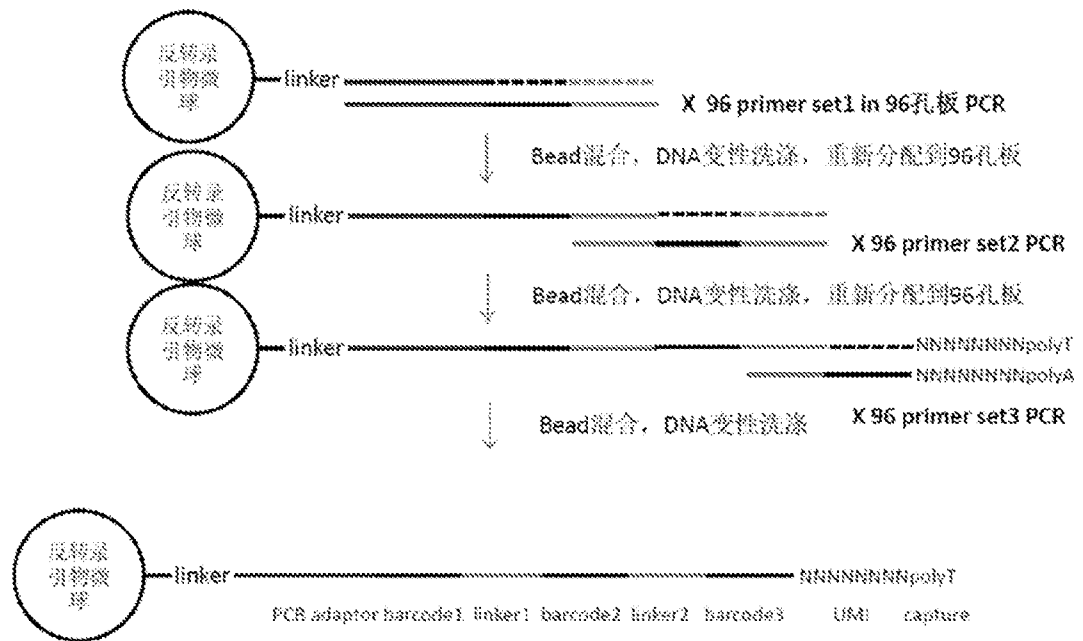


图 3

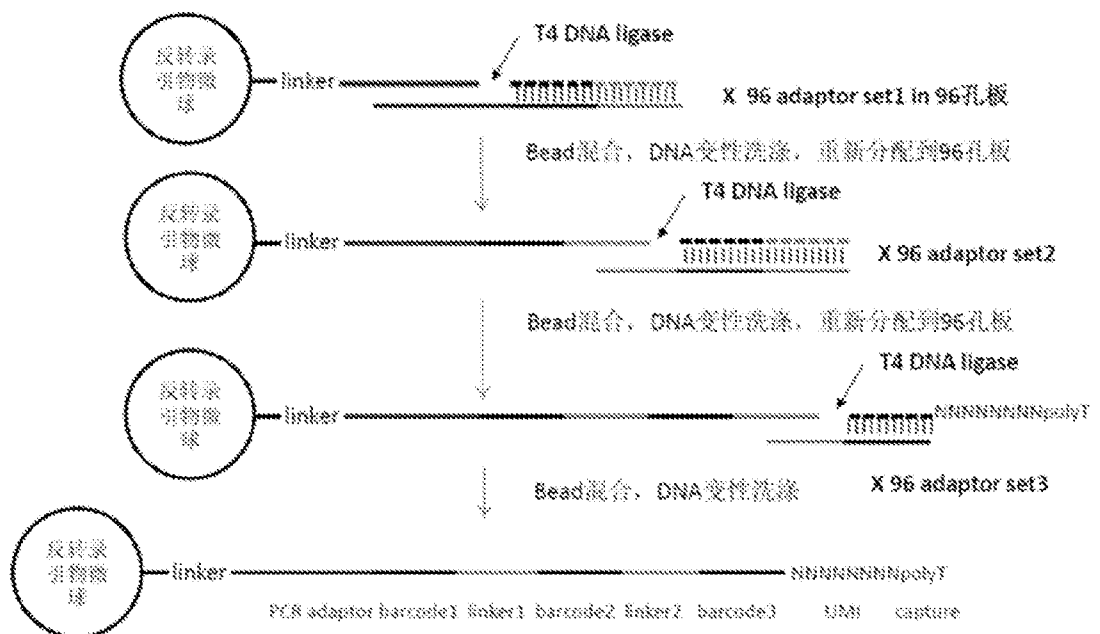


图 4

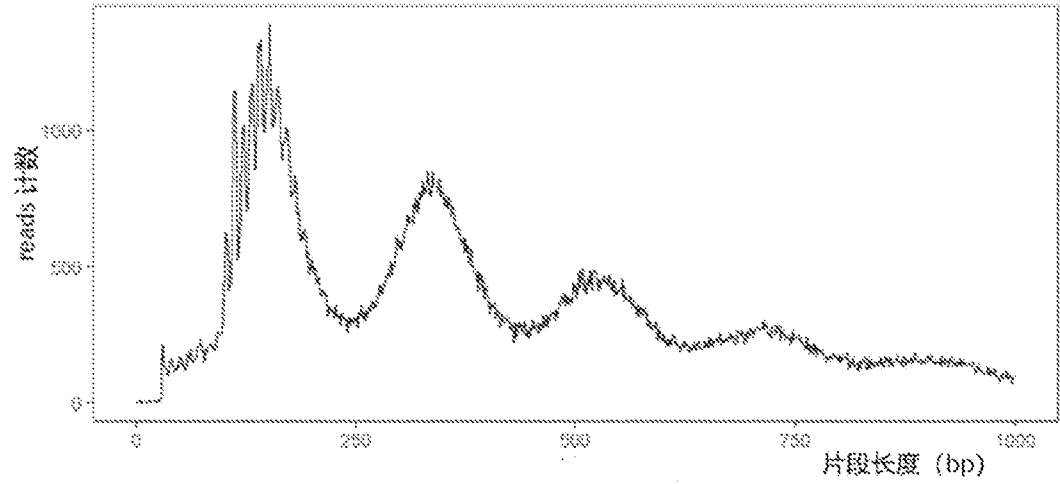


图 5

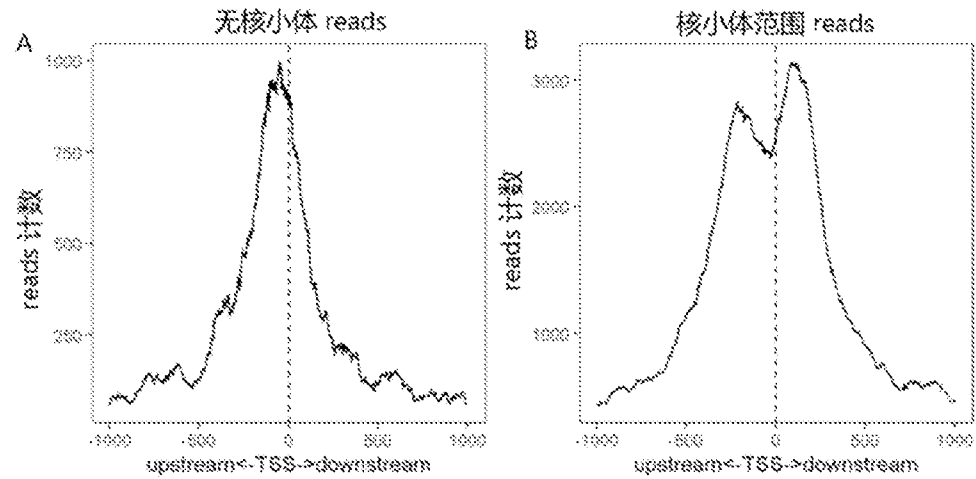


图 6

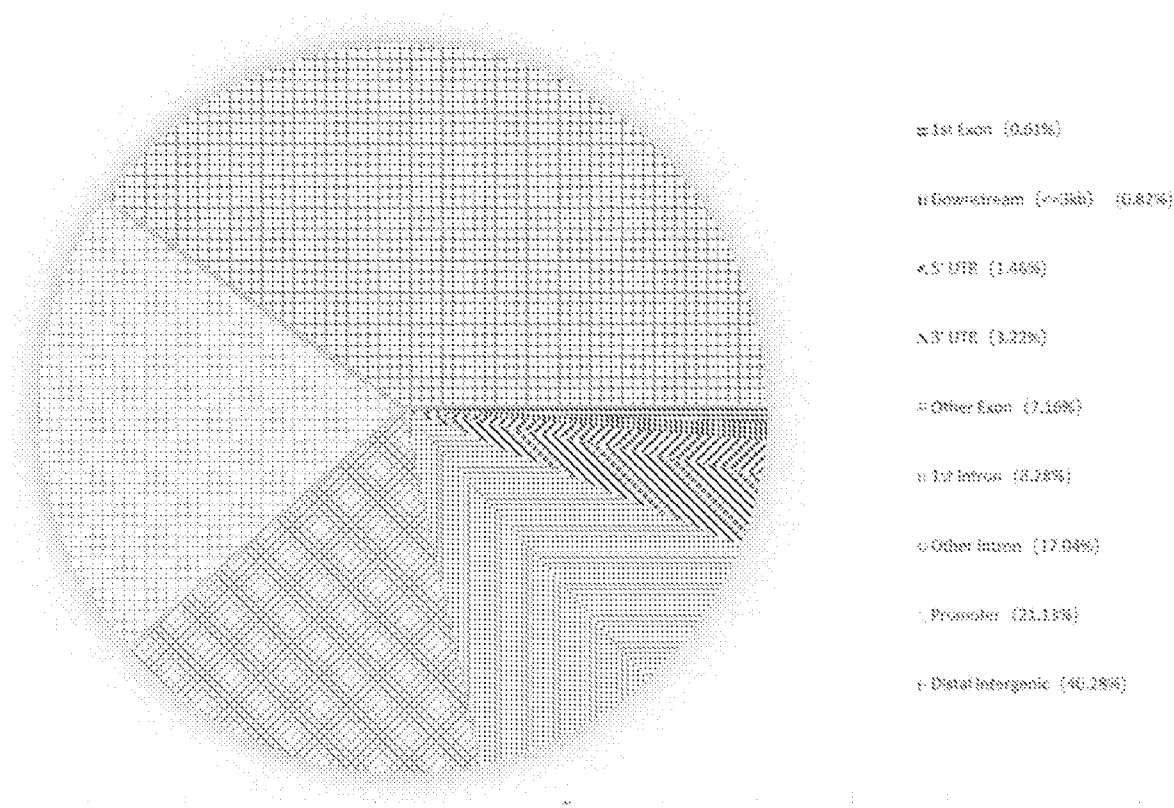


图 7

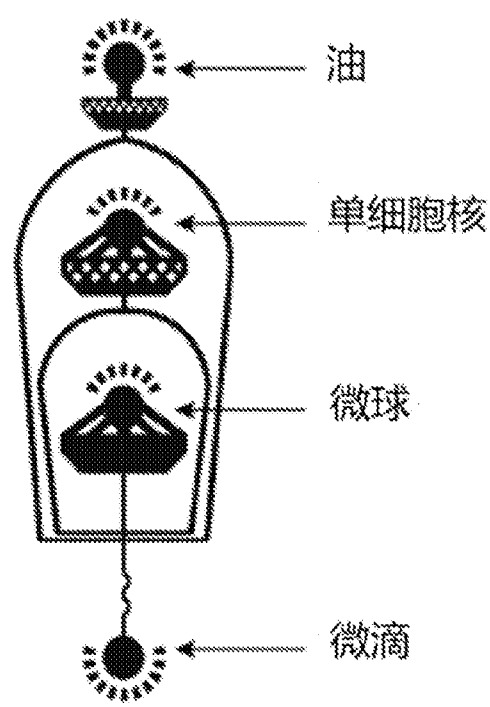


图 8

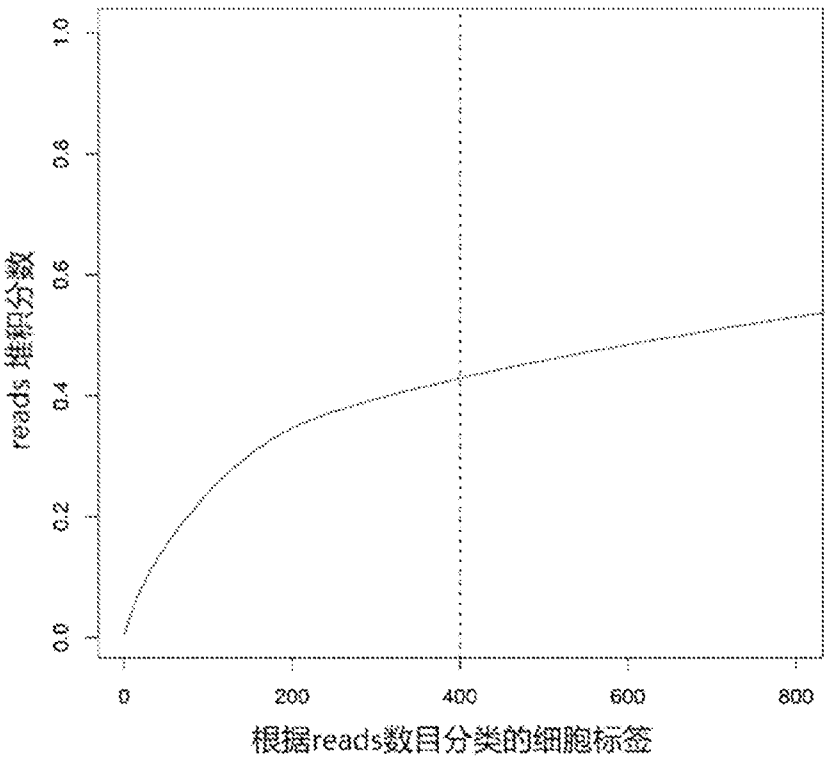


图 9

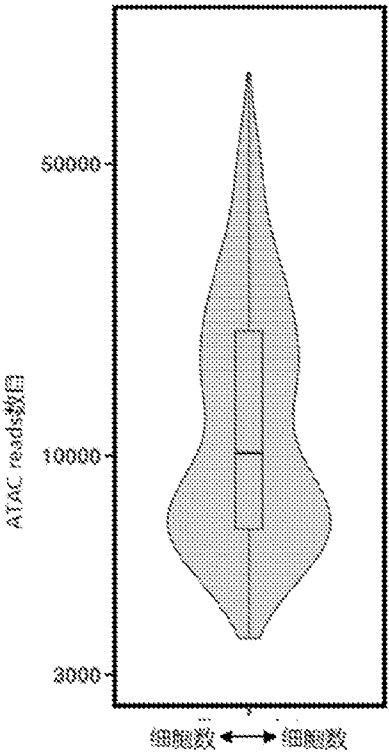


图 10

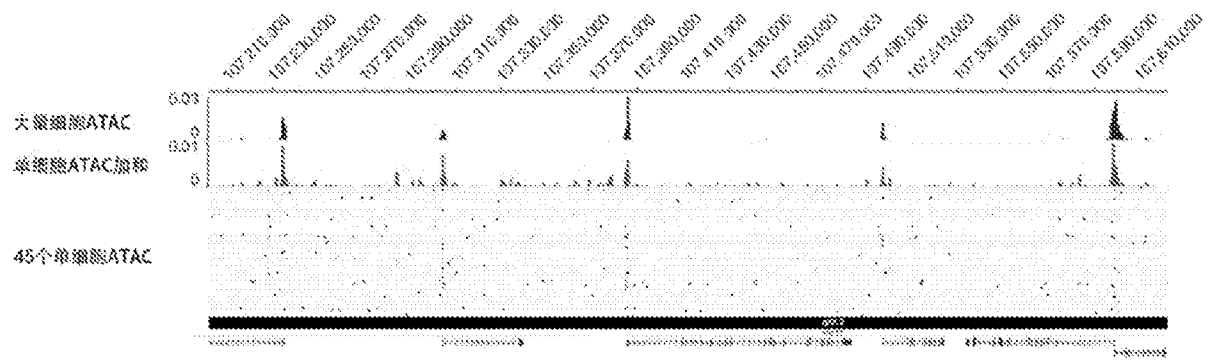


图 11

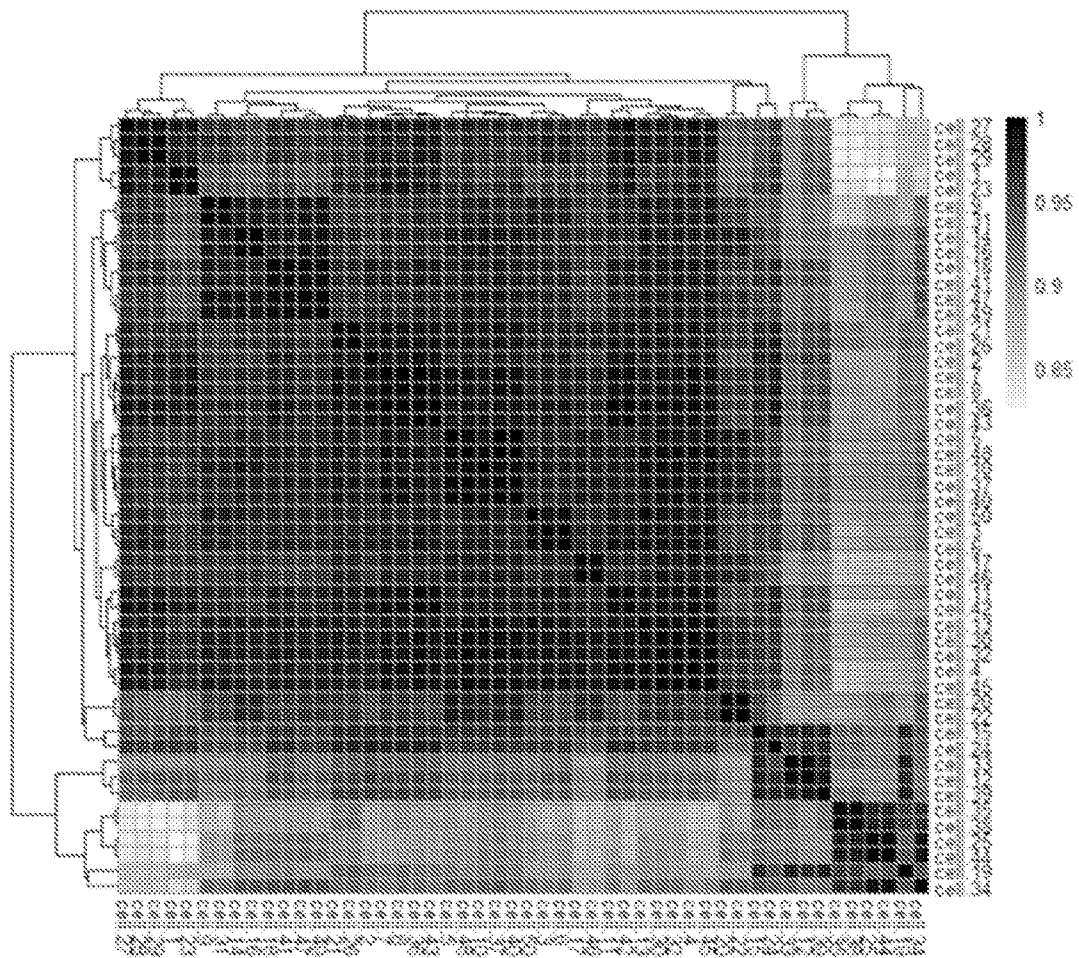


图 12

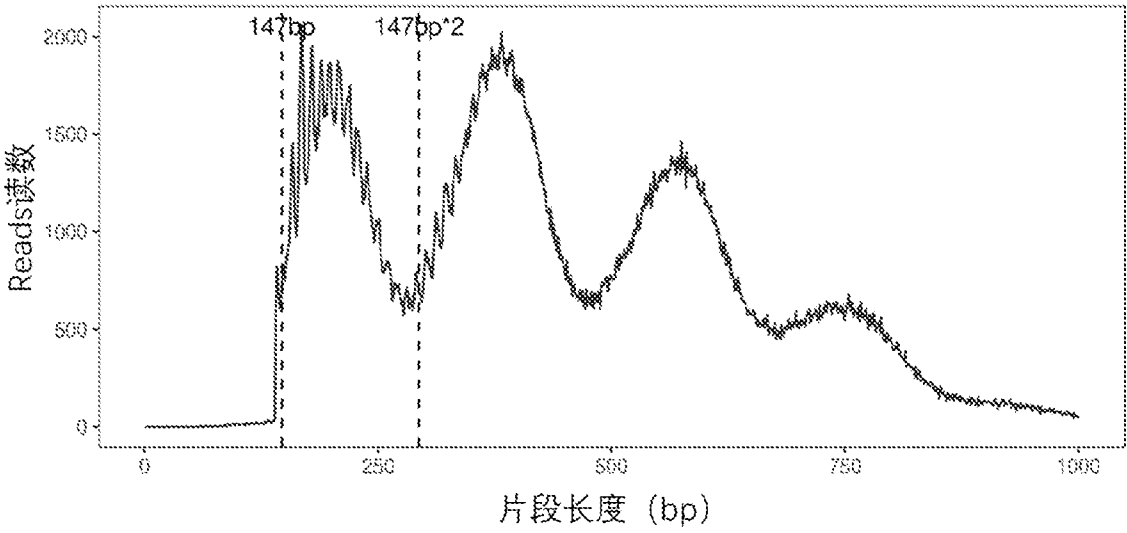


图 13

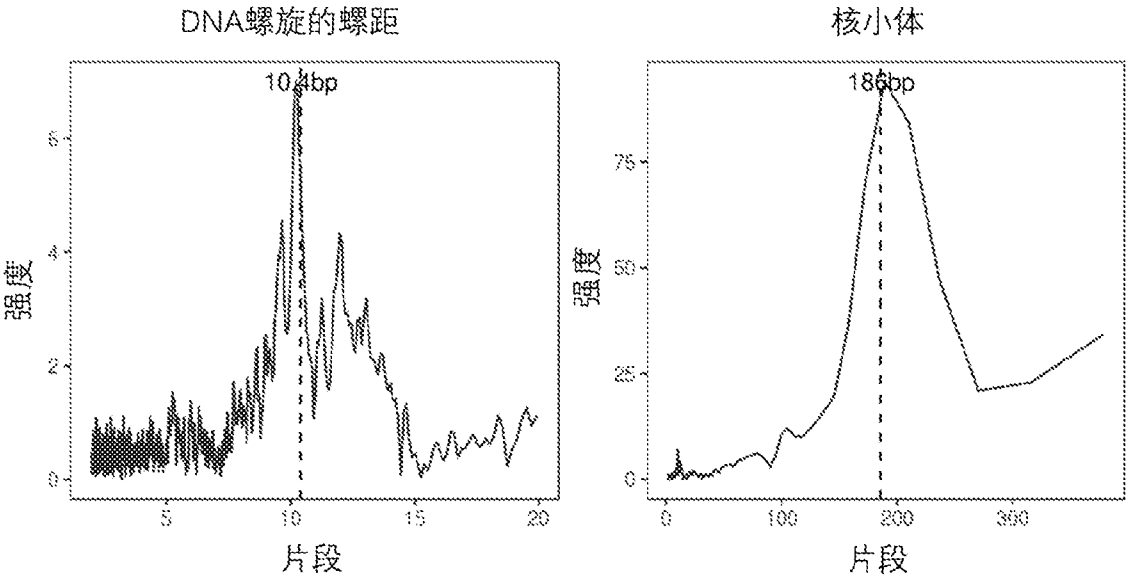


图 14

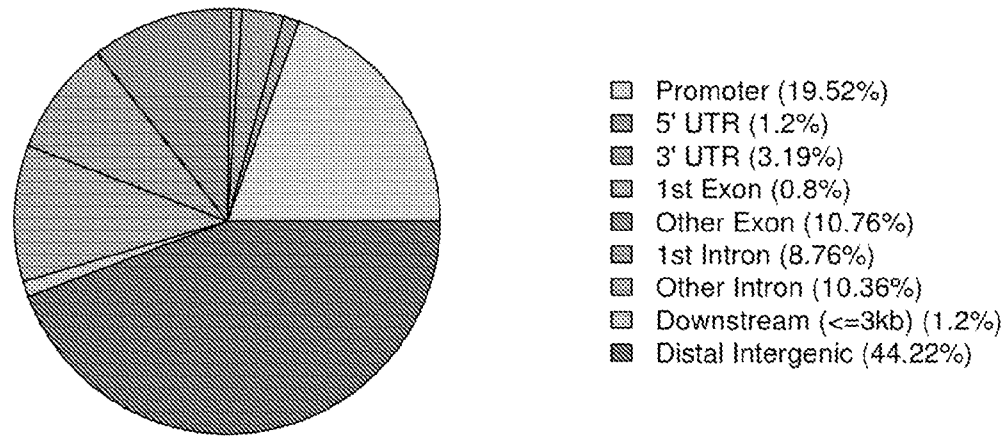


图 15

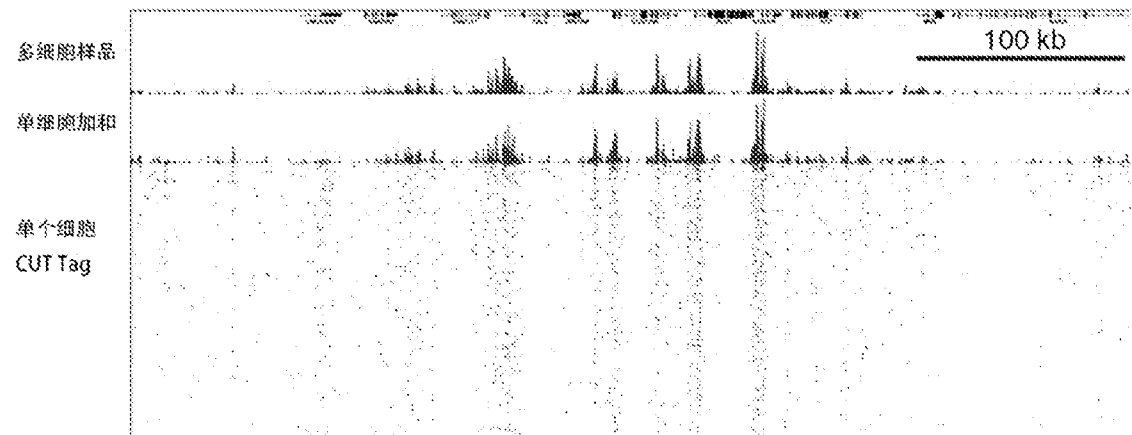


图 16

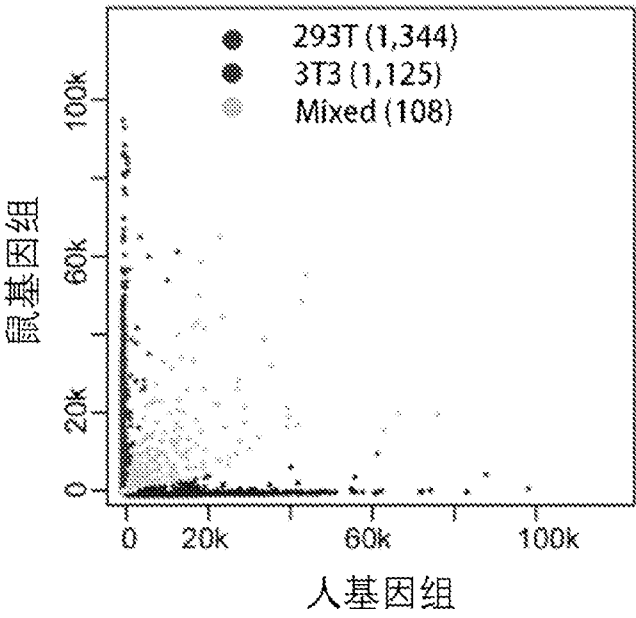


图 17

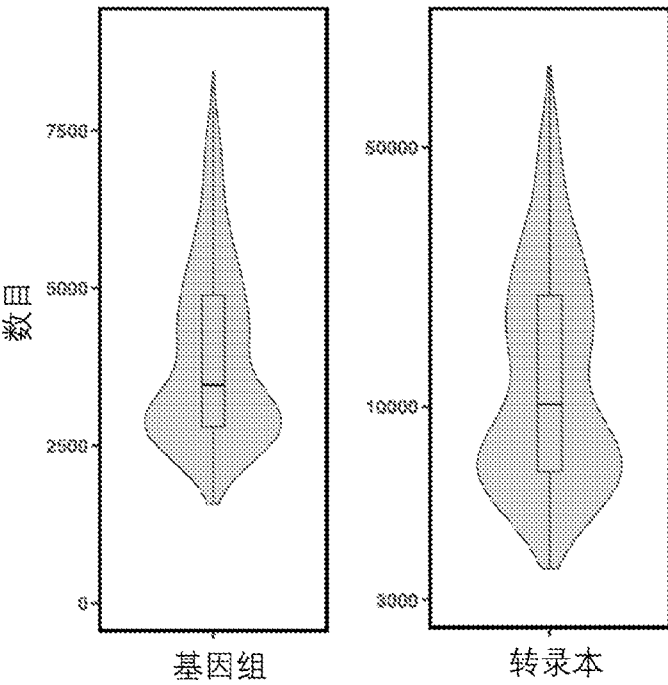


图 18

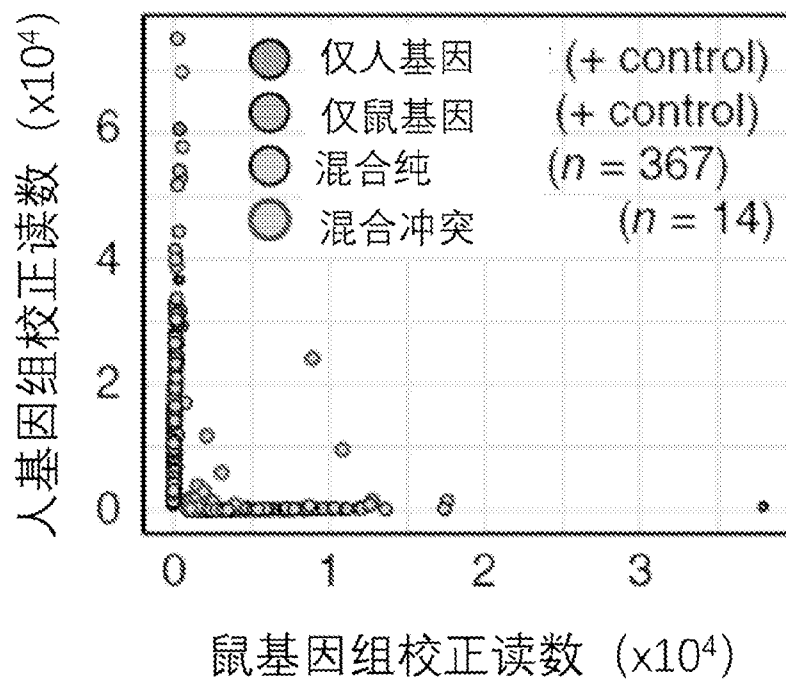


图 19

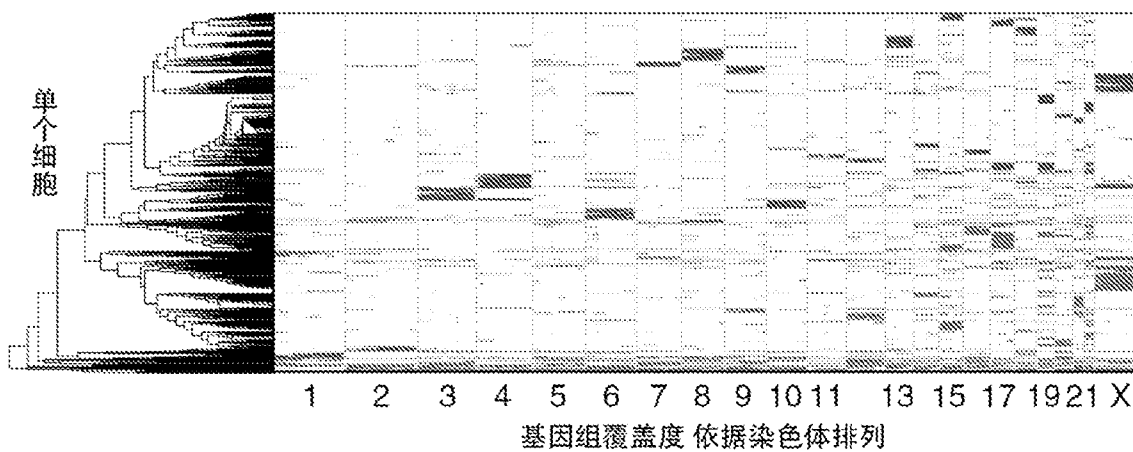


图 20

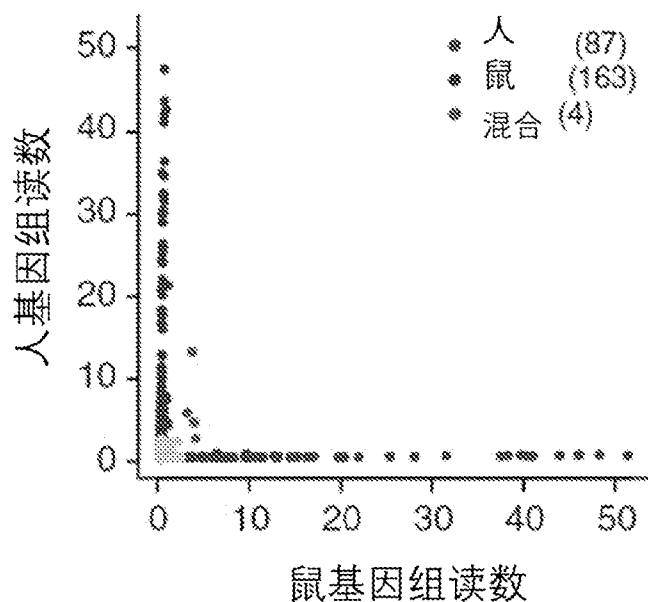


图 21

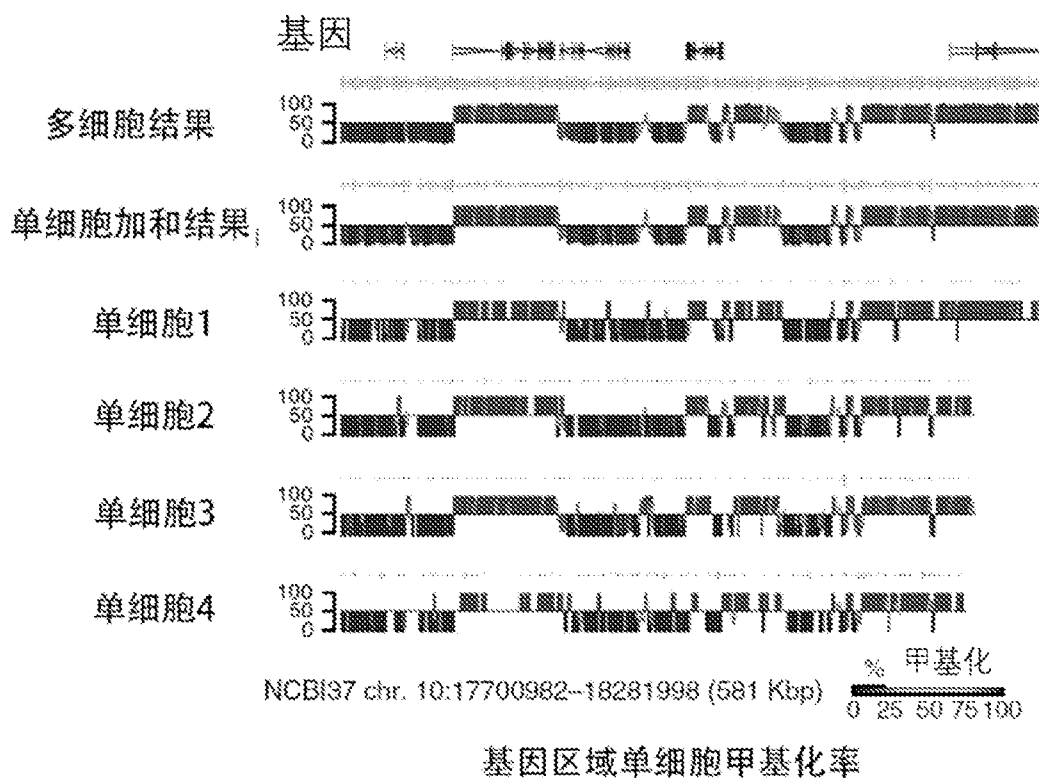
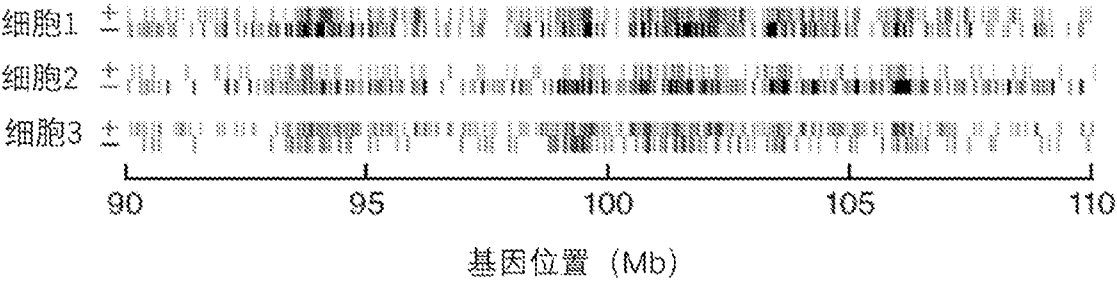


图 22



单细胞5hmC修饰位点，不同单细胞信号部分重叠

图 23

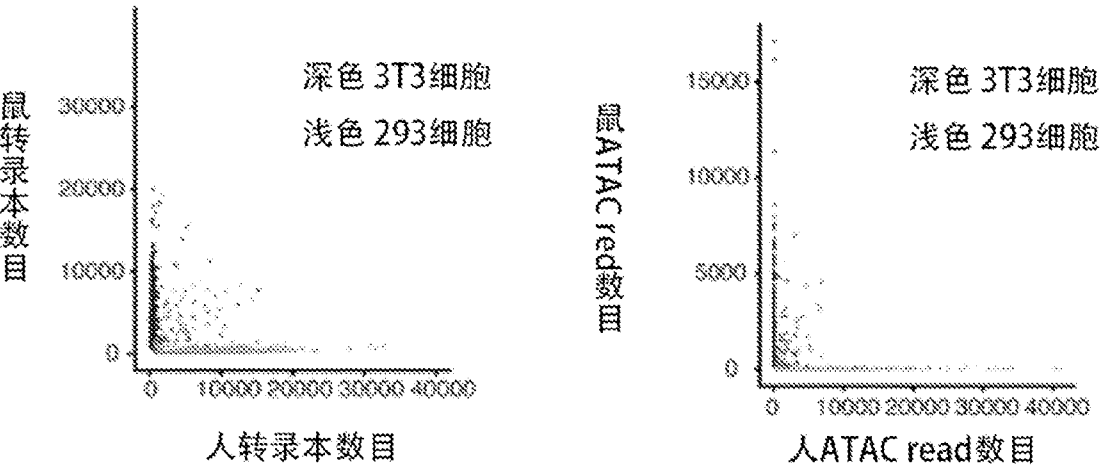


图 24

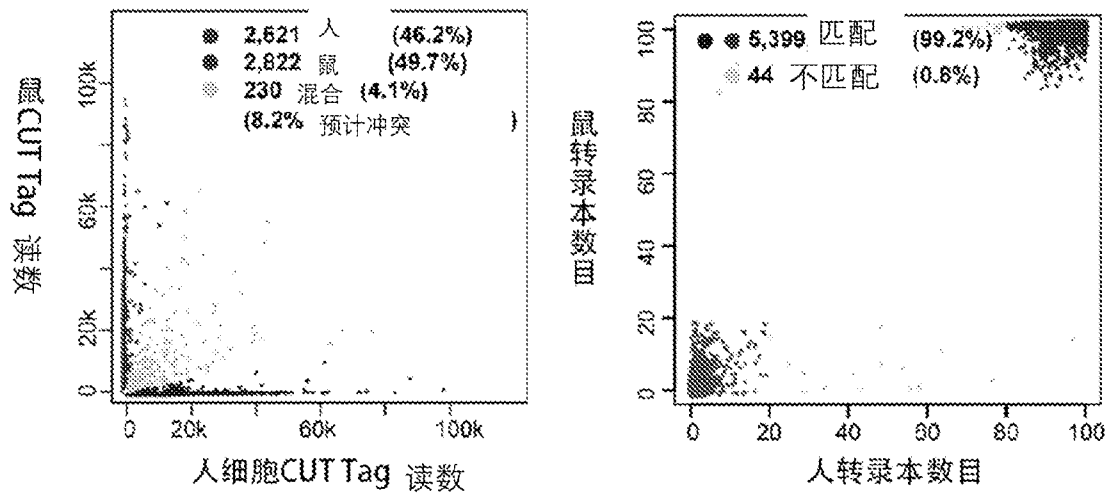


图 25

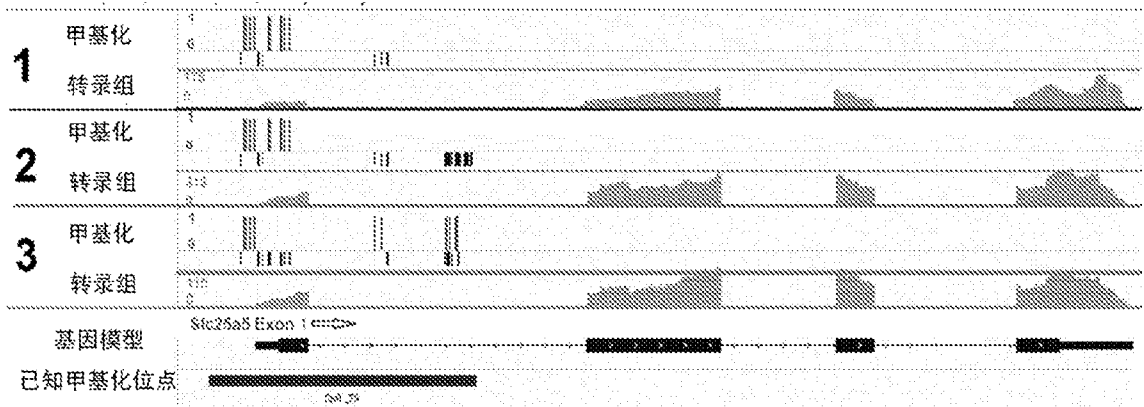


图 26

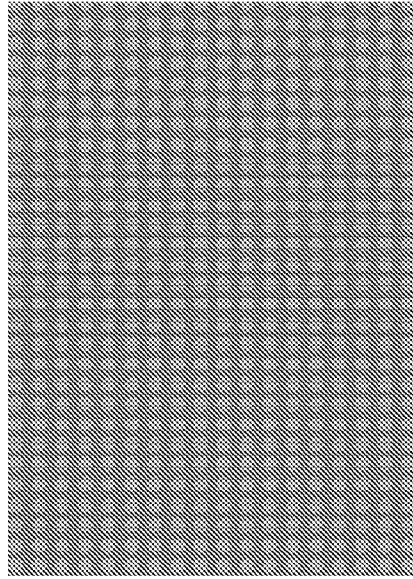


图 27

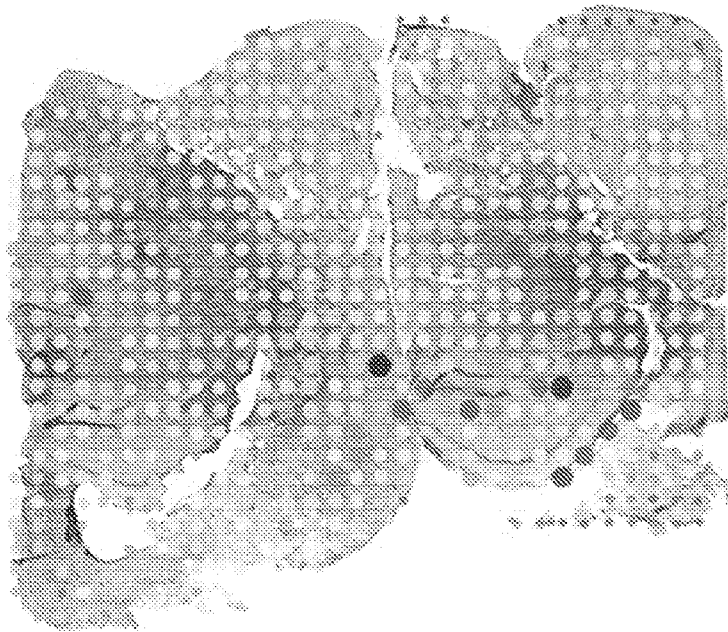


图 28

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/097800

A. CLASSIFICATION OF SUBJECT MATTER C12Q 1/6806(2018.01)i; 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) 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, SIPOABS, DWPI, CNTXT, WOTXT, EPTXT, USTXT, CNKI, 万方, WEB OF SCIENCE, PUBMED, ELSEVIER SCIENCE, Genbank, China Patents Biological Sequence Search System: 单细胞, 样品, 核苷酸, 第一, 第二, 序列, 靶, 目标, 标签, 衔接, 条码, 固体, 待测, T4, 固相, 条码, 探针, 引物, 杂交, 互补, 珠, 孔, 滴, single cell, nuclear acid, first, second, primer, probe, hybrid+, complement+, target, barcode, sample, sequence search for SEQ ID Nos: 1-17				
C. DOCUMENTS CONSIDERED TO BE RELEVANT				
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
Y	CN 108350497 A (ILLUMINA INC.) 31 July 2018 (2018-07-31) claims 1-20	1-114		
Y	US 9834814 B2 (AGILENT TECHNOLOGIES INC.) 05 December 2017 (2017-12-05) abstract, claims 1-3, 19	1-114		
Y	WO 2019165318 A9 (10X GENOMICS, INC.) 26 September 2019 (2019-09-26) abstract, description, paragraphs 5-14	1-114		
Y	CN 107735497 A (SINGULAR BIO INC.) 23 February 2018 (2018-02-23) claims 1-6, figure 65	1-114		
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.				
<table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top; border: none;"> * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed </td> <td style="width: 50%; vertical-align: top; border: none;"> "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 </td> </tr> </table>			* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family			
Date of the actual completion of the international search 02 September 2021		Date of mailing of the international search report 08 September 2021		
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088 China Facsimile No. (86-10)62019451		Authorized officer Telephone No.		

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/097800

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

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

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/097800

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108350497	A	31 July 2018	SG	10201912283 R	A	27 February 2020
				WO	2017040306	A1	09 March 2017
				JP	2020189846	A	26 November 2020
				CA	2997035	A1	09 March 2017
				AU	2020201685	A1	26 March 2020
				EP	3341494	A1	04 July 2018
				AU	2016316773	B2	30 January 2020
				JP	2018526026	A	13 September 2018
				HK	1250757	A1	11 January 2019
				JP	6743150	B2	19 August 2020
				AU	2016316773	A1	22 March 2018
US	9834814	B2	05 December 2017	US	2015148239	A1	28 May 2015
WO	2019165318	A9	26 September 2019	SG	11202008080 R	A	29 September 2020
				WO	2021041974	A1	04 March 2021
				US	2020239874	A1	30 July 2020
				WO	2019165318	A1	29 August 2019
				US	2021040551	A1	11 February 2021
				CN	112074610	A	11 December 2020
				EP	3755812	A1	30 December 2020
CN	107735497	A	23 February 2018	AU	2016219943	A1	12 October 2017
				US	2018023124	A1	25 January 2018
				KR	20170116157	A	18 October 2017
				WO	2016134191	A1	25 August 2016

国际检索报告

国际申请号

PCT/CN2021/097800

A. 主题的分类

C12Q 1/6806(2018.01)i; 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)

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, SIPOABS, DWPI, CNTXT, WOTXT, EPTXT, USTXT, CNKI, 万方, WEB OF SCIENCE, PUBMED, ELSEVIER SCIENCE, Genbank, 中国专利生物序列检索系统: 单细胞, 样品, 核苷酸, 第一, 第二, 序列, 靶, 目标, 标签, 衔接, 条码, 固体, 待测, T4, 固相, 条码, 探针, 引物, 杂交, 互补, 珠, 孔, 滴, single cell, nuclear acid, first, second, primer, probe, hybrid+, complement+, target, barcode, sample, 对SEQ ID Nos: 1-17的序列检索

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
Y	CN 108350497 A (ILLUMINA公司) 2018年 7月 31日 (2018 - 07 - 31) 权利要求1-20	1-114
Y	US 9834814 B2 (AGILENT TECHNOLOGIES INC) 2017年 12月 5日 (2017 - 12 - 05) 摘要, 权利要求1-3, 19	1-114
Y	WO 2019165318 A9 (10X GENOMICS INC) 2019年 9月 26日 (2019 - 09 - 26) 摘要, 说明书第5-14段	1-114
Y	CN 107735497 A (卓异生物公司) 2018年 2月 23日 (2018 - 02 - 23) 权利要求1-6, 图65	1-114

☐ 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

2021年 9月 2日

国际检索报告邮寄日期

2021年 9月 8日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国 北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

杜娟

电话号码 01062411561

第I栏

核苷酸和/或氨基酸序列(续第1页第1. c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列, 国际检索是基于下列序列表进行的:

a. ☒ 作为国际申请的一部分提交的:

☒ 附件C/ST. 25文本文件形式

☐ 纸件或图形文件形式

b. ☐ 根据细则13之三. 1(a) 仅为国际检索目的以附件C/ST. 25文本文件形式与国际申请同时提交的:

c. ☐ 仅为国际检索目的在国际申请日之后提交的:

☐ 附件C/ST. 25文本文件形式(细则13之三. 1(a))

☐ 纸件或图形文件形式(细则13之三. 1(b)和行政规程第713段)

2. ☐ 另外, 在提交/提供了多个版本或副本的序列表的情况下, 提供了关于随后提交的或附加的副本中的信息与申请时提交的作为申请一部分的序列表的信息相同或未超出申请时提交的申请中的信息范围(如适用)的所需声明。

3. 补充意见:

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/097800

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	108350497	A	2018年 7月 31日	SG	10201912283R	A	2020年 2月 27日
				WO	2017040306	A1	2017年 3月 9日
				JP	2020189846	A	2020年 11月 26日
				CA	2997035	A1	2017年 3月 9日
				AU	2020201685	A1	2020年 3月 26日
				EP	3341494	A1	2018年 7月 4日
				AU	2016316773	B2	2020年 1月 30日
				JP	2018526026	A	2018年 9月 13日
				HK	1250757	A1	2019年 1月 11日
				JP	6743150	B2	2020年 8月 19日
				AU	2016316773	A1	2018年 3月 22日
US	9834814	B2	2017年 12月 5日	US	2015148239	A1	2015年 5月 28日
WO	2019165318	A9	2019年 9月 26日	SG	11202008080R	A	2020年 9月 29日
				WO	2021041974	A1	2021年 3月 4日
				US	2020239874	A1	2020年 7月 30日
				WO	2019165318	A1	2019年 8月 29日
				US	2021040551	A1	2021年 2月 11日
				CN	112074610	A	2020年 12月 11日
				EP	3755812	A1	2020年 12月 30日
CN	107735497	A	2018年 2月 23日	AU	2016219943	A1	2017年 10月 12日
				US	2018023124	A1	2018年 1月 25日
				KR	20170116157	A	2017年 10月 18日
				WO	2016134191	A1	2016年 8月 25日