

(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 2 月 4 日 (04.02.2021)



(10) 国际公布号
WO 2021/017883 A1

(51) 国际专利分类号:
C07K 19/00 (2006.01) *C12Q 1/6806* (2018.01)
C12N 9/12 (2006.01) *C12N 15/10* (2006.01)

(21) 国际申请号: PCT/CN2020/102513

(22) 国际申请日: 2020 年 7 月 17 日 (17.07.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910709933.0 2019年8月1日 (01.08.2019) CN

(71) 申请人: 北京大学 (PEKING UNIVERSITY) [CN/CN]; 中国北京市海淀区颐和园路 5 号, Beijing 100080 (CN)。

(72) 发明人: 何爱彬 (HE, Aibin); 中国北京市海淀区颐和园路 5 号, Beijing 100080 (CN)。余先红 (YU, Xianhong); 中国北京市海淀区颐和园路 5 号, Beijing 100080 (CN)。艾珊珊 (AI,

Shanshan); 中国北京市海淀区颐和园路 5 号, Beijing 100080 (CN)。

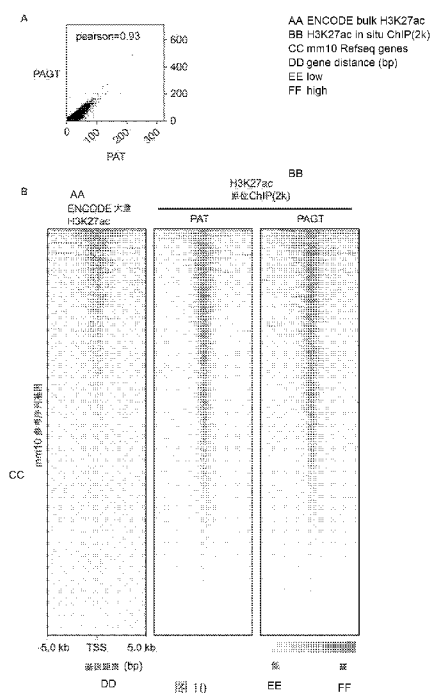
(74) 代理人: 北京超凡宏宇专利代理事务所 (特殊普通合伙) (CHOFN INTELLECTUAL PROPERTY); 中国北京市海淀区北四环西路 68 号左岸工社 1215-1218 室, Beijing 100080 (CN)。

(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,

(54) Title: FUSION PROTEIN FOR PREPARING SINGLE-CELL CHIP-SEQ LIBRARY AND APPLICATION THEREOF

(54) 发明名称: 一种用于单细胞ChIP-seq文库制备的融合蛋白及其应用



(57) Abstract: A fusion protein for preparing a single-cell ChIP-seq library and application thereof. The fusion protein comprises a Tn5 transposase and an Fc binding protein. A kit comprises said fusion protein and other auxiliary detection reagents. The method uses said fusion protein or kit to perform a ChIP-seq detection method. The fusion protein, kit, and method can improve library construction efficiency during a ChIP-seq detection process and reduce a library background, thereby improving the accuracy of the ChIP-seq detection method and simplifying the experimental flow of ChIP-seq.

(57) 摘要: 一种用于单细胞ChIP-seq文库制备的融合蛋白及其应用, 其中, 融合蛋白包括Tn5转座酶和Fc结合蛋白, 试剂盒包括前述融合蛋白以及其他辅助检测试剂, 方法使用前述融合蛋白或试剂盒进行ChIP-seq检测方法。其融合蛋白、试剂盒和方法在ChIP-seq检测过程中能够提高建库效率, 降低文库背景, 从而提高ChIP-seq检测方法的准确性并简化ChIP-seq的实验流程。

NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

- 包括国际检索报告(条约第21条(3))。
- 包括说明书序列表部分(细则5.2(a))。

一种用于单细胞 ChIP-seq 文库制备的融合蛋白及其应用

相关申请的交叉引用

本申请要求于 2019 年 08 月 01 日提交中国专利局的申请号为 2019107099330、名称为“一种用于单细胞 ChIP-seq 文库制备的融合蛋白及其应用”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本公开属于生物技术领域，具体而言，涉及一种用于单细胞 ChIP-seq 文库制备的融合蛋白及其应用。

背景技术

随着基因测序的完成以及后基因组时代的到来，表观遗传学已经成为生物领域的研究热点。表观遗传学 (epigenetics) 主要研究在基因核苷酸序列不发生改变的情况下，DNA 及有关蛋白分子发生的可遗传修饰，这些修饰可被细胞“记忆”并在随后的细胞分裂过程中保留下来，其研究方向包括：一是基因转录水平选择性表达的调控，二是基因转录后的调控。目前，表观遗传学的热点主要集中在基因转录水平选择性表达调控上，特别是转录因子与 DNA 的相互作用、DNA 的甲基化和组蛋白修饰等。

染色质免疫共沉淀 (Chromatin Immunoprecipitation, ChIP) 也称结合位点分析法，其一般流程主要包括：(1) 用甲醛将 DNA 和结合在所述 DNA 上的蛋白交联，染色体分离并打碎成一定大小的片段；

(2) 用特异性抗体免疫沉淀并富集目标蛋白与 DNA 交联的复合物；(3) 采用低 pH 值条件反交联，释放 DNA 片段；(4) 通过对 DNA 片段的纯化与检测，获得 DNA 与蛋白质相互作用的序列信息。由于 ChIP 技术能够研究蛋白质与 DNA 的相互作用，其被广泛应用表观遗传学领域，用于研究转录因子与 DNA 的相互作用、DNA 甲基化和组蛋白修饰等。

随着新一代测序技术的发展，在 ChIP 技术的基础上发展出一种可在全基因组范围内研究蛋白质与 DNA 相互作用的技术——染色质免疫共沉淀-测序 (即，ChIP-seq)。ChIP-seq 技术包括染色质免疫共沉淀和高通量测序两部分，其中，先通过染色质免疫共沉淀技术特异性地富集目的蛋白结合的 DNA，然后构建测序文库，采用新一代测序技术对富集得到的 DNA 片段进行高通量测序，最后通过将获得的数百万条序列标签精确定位到基因组上，从而获得全基因组范围内与目标蛋白相互作用的 DNA 区段信息。

目前，一种优化的并与本技术相关的一种 ChIP-seq 技术叫做“CUT&RUN (Cleavage Under Targets & Release Using Nuclease)”。所述 CUT&RUN 将 Protein A 与核酸酶 MNase (Micrococcal Nuclease) 融合，通过 Protein A 能特异与免疫球蛋白 G 结合的特性，Protein A 将 MNase 引入抗体 (识别特异转录因子或者组蛋白修饰的抗体) 结合的位点，通过 MNase 的核酸内切和外切酶活性，Protein A-MNase 能特异将抗体结合的 DNA 两端切开，被切开的 DNA 片段从细胞核中释放出来，对这些 DNA 片段进行建库和测序分析，能在全基因组水平绘制特定蛋白与 DNA 相互作用图谱。

其具体技术方案如图 1 所示：Con A beads 和细胞结合，加入识别特异转录因子的抗体，孵育一定时间，让抗体与转录因子充分结合，加入 Protein A-MNase 融合蛋白，Protein A-MNase 能特异结合抗体结合的位点，然后加入 Ca^{2+} 离子激活 MNase 核酸酶的活性，Protein A-MNase 能特异将抗体结合的 DNA 的两端切开，终止 MNase 的反应之后，这些被切开的 DNA 片段从细胞核中释放出来，之后对这些 DNA 片段进行建库和二代测序分析，根据测序结果可以在全基因组水平绘制特定转录因子的结合图谱。

然而，目前通用的 ChIP-seq 技术存在以下缺陷：

1、建库效率低，导致做少量细胞时文库信息丢失严重

对 Protein A-MNase 切割后释放出来的 DNA 片段建库时，需要采用传统的 TruSeq 的建库策略，建库时 Adaptor 和 DNA 片段连接效率低。尤其以少量细胞为起始材料时，DNA 片段在建库过程中丢失太多，最终导致文库主要信息丢失，实验失败几率高，也无法得到全基因组水平上的蛋白与 DNA 相互作用图谱。

2、文库背景较高

Protein A-MNase 切割反应被终止后，所有小的 DNA 片段将从细胞核中释放出来，它们没有与转录因子或蛋白质结合。这些小片段，在建库过程中会被 Adaptor 连接，最后通过 PCR 扩增，形成高背景，降低了文库复杂度 (complexity)，导致实验失败。

3、现有的 ChIP-seq 技术无法对组织切片等进行原位检测，破坏其空间分辨率。

有鉴于此，本领域亟待提出一种能够有效提高文库建立的效率、降低文库背景以及能够实现原位检测的 ChIP-seq 技术。

发明内容

本公开首先涉及一种用于染色质免疫共沉淀-测序 (ChIP-seq) 建库的融合蛋白，所述的融合蛋白是

同源二聚体，其单体包括如下结构：

- (1) 第一功能嵌段：Tn5 转座酶突变型；
- (2) 第二功能嵌段：金黄色葡萄球菌 A 蛋白 (Protein A)、链球菌 G 蛋白 (Protein G) 或 Protein A 和 Protein G 的组合；
- (3) 各个功能嵌段的连接结构 (Linker)，蛋白纯化标签。

在一种或多种实施方式中，所述的 Tn5 转座酶突变型为在氨基酸序列如 SEQ ID NO.1 所示的野生型 Tn5 转座酶的氨基酸结构上进行如下点突变：E54K、L372P 获得的 Tn5 转座酶突变型；

所述的 Linker 的氨基酸序列如 SEQ ID NO.3 所示；

所述的蛋白纯化标签为可通过亲和层析纯化的蛋白纯化标签。

在一种或多种实施方式中，所述的 Tn5 转座酶突变型为在野生型 Tn5 转座酶 (氨基酸序列如 SEQ ID NO.1 所示) 的氨基酸结构上进行如下点突变：E54K、L372P，所述的 Tn5 转座酶突变型的氨基酸序列如 SEQ ID NO.2 所示。

在一种或多种实施方式中，通过氨基酸序列如 SEQ ID NO.3 所示的 Linker 连接所述融合蛋白的各个功能嵌段；

在一种或多种实施方式中，所述的蛋白纯化标签为：-HIS 标签、GST 标签、MBP 标签、SUMO 标签、NusA 标签、或直接利用 Protein A 标签和 IgG 蛋白的特异亲和性，用 IgG 磁珠进行纯化所述融合蛋白；

在一种或多种实施方式中，所述的第二功能嵌段连接在第一功能嵌段的氨基酸序列的 N 端 (氮端)；

在一种或多种实施方式中，所述的融合蛋白的单体为 PAT 或 PAGT，PAT 的氨基酸序列如 SEQ ID NO.4 所示，PAGT 的氨基酸序列如 SEQ ID NO.5 所示。

本公开还涉及所述的融合蛋白在制备特定生物学样本的高通量测序文库中的应用，所述的生物学样本包括但不限于：未经交联或固定/经过交联或固定培养细胞样本、新鲜组织/经过交联或固定的组织样本，所述的高通量测序文库为染色质免疫共沉淀-测序 (ChIP-seq) 文库。

本公开还涉及使用所述的融合蛋白制备目标生物学样本的高通量测序文库的方法所述的制备特定生物学样本的高通量测序文库中的方法为：

- (1) 收集并处理目标生物学样本，获得单细胞悬液；
- (2) 用结合缓冲液重悬细胞，加入适量的特定蛋白结合抗体，使抗体充分结合后清洗除去未结合的抗体；
- (3) 重悬细胞，加入所述融合蛋白，使融合蛋白与抗体充分结合后除去未结合的融合蛋白；
- (4) 激活所述融合蛋白，充分反应获得被所述融合蛋白识别并切割的带有标记的 DNA 片段；
- (5) 加入终止缓冲液终止反应，用纯化试剂盒进行纯化 DNA 片段；
- (6) 加入引物进行 PCR 扩增完成建库，

所述的生物学样本包括但不限于：未经交联或固定/经过交联或固定培养细胞样本、新鲜组织/经过交联或固定的组织样本；

所述的高通量测序文库为染色质免疫共沉淀-测序 (ChIP-seq) 文库。

本公开还涉及使用所述的融合蛋白制备目标生物学样本的高通量测序文库的方法所述的制备特定生物学样本的高通量测序文库中的方法为：

- (1) 收集并处理目标生物学样本，获得单细胞悬液；
- (2) 用结合缓冲液重悬细胞，加入适量的特定蛋白结合抗体，使抗体充分结合后清洗除去未结合的抗体；
- (3) 重悬细胞，加入所述融合蛋白，使融合蛋白与抗体充分结合后除去未结合的融合蛋白；
- (4) 激活所述融合蛋白，充分反应获得被所述融合蛋白识别并切割的带有标记的 DNA 片段；
- (5) 加入终止缓冲液终止反应，用纯化试剂盒进行纯化 DNA 片段；
- (6) 加入引物进行 PCR 扩增完成建库。

在一种或多种实施方式中，所述的生物学样本包括但不限于：未经交联或固定/经过交联或固定培养细胞样本、新鲜组织/经过交联或固定的组织样本；

所述的高通量测序文库为染色质免疫共沉淀-测序 (ChIP-seq) 文库。

本公开还涉及一种用于构建高通量测序文库的试剂盒，所述的试剂盒包括：所述的融合蛋白及常规的建库过程所使用的样本洗涤缓冲液、结合缓冲液、激活缓冲液和建库引物等，所述的高通量测序文库为染色质免疫共沉淀-测序 (ChIP-seq) 文库。

与现有技术相比，本公开的有益效果为：

(1) 本公开所述融合蛋白、试剂盒和方法在 ChIP-seq 检测过程中能够提高建库效率，降低文库背景，从而提高 ChIP-seq 检测方法的准确性并简化 ChIP-seq 的实验流程。

(2) 本公开所述方法对 Tn5 转座酶和 Fc 结合蛋白的突变体、表达体系、纯化手段、抗体、融合蛋白的孵育条件以及切割反应的条件进行优化,提高抗体、融合蛋白的结合效率以及 DNA 切割的效率和准确性,从而降低文库的背景和分辨率;

(3) 本公开所述方法能够对组织切片、细胞涂片和细胞爬片进行原位检测,不需要裂解细胞核以及超声打断染色质,保留原样本的空间分辨率。

(4) 本公开所述方法能够实现多器官多组织的单细胞 ChIP-seq,可以同时平行处理多种不同组织来源,或者多种不同病理状态下的细胞。

(5) 本公开所述方法能够快速地完成单细胞 ChIP-seq 的制备,从实验开始,单日便可完成。

(6) 本公开所述方法能够实现高通量的单细胞 ChIP-seq,能够一次获得成上万个单细胞的 ChIP-seq 文库。

(7) 本公开所述方法不需要使用特殊定制的高成本的单细胞文库测序方法,使用特殊设计的 Mosaic Truseq 两步 PCR 的文库制备方法,得到的文库可以使用 illumina 标准的测序方法和平台进行测序。与特殊定制的单细胞测序方法相比,一方面, Mosaic Truseq 文库只需要 1/15 的测序成本,另一方面,实验人员不需要精通专业的测序知识,即可进行大规模的单细胞测序。

附图说明

为了更清楚地说明本公开实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本公开的某些实施例,因此不应被看作是对范围的限定,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他相关的附图。

图 1、ProteinA-MNase 在 ChIP-seq 中的作用原理。

图 2、PAT 表达载体构建图谱。

图 3、高纯 PAT 纯化结果。

图 4、PAT 与 IgG 结合验证实验。

图 5、Protein A 嵌段在 Tn5 嵌段的 N 端插入型融合蛋白 (PA2-Tn5) 产生高活性的 PAT。

图 6、自然状态 (A) 和交联状态 (B) 下,原位 ChIP-seq (In situ-ChIP) 和超声 ChIP-seq (Sonicate ChIP) 方法比较, H3K4me3 在 *Pou5f1* 基因启动子区域, H3K27ac 在 *Nanog* 基因增强子区域信号分布 (IGV track 示意图)。

图 7、自然状态 (A) 和交联状态 (B) 下,原位 ChIP-seq (In situ-ChIP) 和超声 ChIP-seq (Sonicate ChIP) 方法比较, H3K4me3 (A) 和 H3K27ac (B) 在启动子区域的信号分布 (热图), 其中 TSS 表示转录起始位点。

图 8、时空特异的原位 ChIP。

图 9、PAGT、PAT 与传统 ChIP 的比较结果, H3K27ac 在 *Pou5f1* 基因启动子区域的信号分布 (IGV track 显示图)。

图 10、PAGT、PAT 与传统 ChIP 的比较结果, H3K27ac 在 *Pou5f1* 基因启动子区域的信号分布 (热图)。

图 11、高通量单细胞的 ChIP-seq 流程图。

图 12、Mosaic Truseq 两步 PCR 文库制备流程图。

具体实施方式

为使本公开实施例的目的、技术方案和优点更加清楚,下面将对本公开实施例中的技术方案进行清楚、完整地描述。实施例中未注明具体条件者,按照常规条件或制造商建议的条件进行。所用试剂或仪器未注明生产厂商者,均为可以通过市售购买获得的常规产品。

下述实施例中所涉及的融合蛋白 PAT (ProteinA-Tn5) 或融合蛋白 PAGT (ProteinA-proteinG-Tn5), 所用洗涤缓冲液 1、洗涤缓冲液 2、结合缓冲液、反应缓冲液和终止缓冲液的配方如下所示:

洗涤缓冲液 1: 50mM HEPES、150mM NaCl 和 0.5mM 亚精胺;

洗涤缓冲液 2: 50mM HEPES、150mM NaCl、0.5mM 亚精胺和 0.01% (w.t.) 毛地黄皂苷;

结合缓冲液: 50mM Hepes、150mM NaCl、0.5 mM 亚精胺、0.01% (w.t.) 毛地黄皂苷和 5mM EDTA;

反应 (激活) 缓冲液: 25mM Mg²⁺和 50mM 三羟甲基甲基胺基丙磺酸;

终止缓冲液: 100mM EDTA 和 1% (w.t.) SDS。

下述实施例 5~6 融合蛋白为 Protein A-Tn5, 所用洗涤缓冲液 A、洗涤缓冲液 B、洗涤缓冲液 C、抗体孵育缓冲液等的配方如下所示 (高通量单细胞专用):

洗涤缓冲液 A: 20mM HEPES、150mM NaCl 和 0.5mM 亚精胺;

洗涤缓冲液 B: 20mM HEPES、150mM NaCl、0.5mM 亚精胺, 0.01% (w.t.) 毛地黄皂苷;

洗涤缓冲液 C: 20mM HEPES、150mM NaCl、0.5mM 亚精胺, 0.01% (w.t.) 毛地黄皂苷和 0.1% Triton

X-100;

激活缓冲液: 20mM HEPES、10mM KCl、1mM CaCl₂、1mM MnCl₂

抗体孵育缓冲液: 20mM HEPES、150mM NaCl、0.5mM 亚精胺, 0.01%(w.t.) 毛地黄皂苷、0.1% Triton X-100 和 2 mM EDTA;

单细胞反应缓冲液: 25mM Mg²⁺、50mM 三羟甲基甲胺基丙磺酸和 0.01% (w.t.) 毛地黄皂苷;

单细胞终止缓冲液: 40mM EDTA。

单细胞分选缓冲液: 2% BSA/PBS+2mM EDTA

单细胞裂解缓冲液: 10 mM Tris-HCl pH 8.5, 0.05 % SDS, 0.1 mg/ml 蛋白酶 K。

实施例 1、高活性 ProteinA-Tn5 (PAT 蛋白) 的设计、表达及纯化

本实施例提供一种超高活性 Tn5 ((突变型 Tn5 提升酶活) 及 ProteinA-Tn5 融合蛋白的设计及纯化, 本方法成本低, 蛋白生产效率高, 所述方法包括以下步骤:

1、ProteinA-Tn5 融合蛋白 (PAT) 的设计

1) 将突变型 Tn5E54K,L372P (以下简称 Tn5) 的核酸序列克隆到 pET28 表达载体, 如图 2 所示。

2) 连接上 2xProtein A 的 IgG 识别结构域。

3) 通过 Sanger 测序, 确保编码区的真实性。

相比野生型 Tn5, PAT 中的突变型的 Tn5 具有更强的 adaptors 结合能力, 切割活性提高十倍以上, 切割活性的提高有利于其应用到各种实验和技术当中。

2、PAT 蛋白的纯化, 所述方法包括以下步骤:

1) 将测序正确的由上述步骤 1 获得的 PAT 表达质粒转入 BL21(ED3)表达菌中。

2) 在 10 ml LB 培养中添加终浓度为 1 µg/ml 的卡拉霉素, 并且接种一个表达 PAT 的单克隆, 37°C, 220 rpm, 过夜培养。

3) 将过夜培养的菌液 (10 ml) 直接按 1/100 接种到新鲜 LB 培养基中, 37°C, 220 rpm。

4) 约培养 3h 左右, 在冰水中冷却, 冷却约 15 min 后添加终浓度为 0.2 mM 的 IPTG。

5) 23°C, 100 rpm, 诱导 5h, 4°C, 4500 rpm 离心 10 min 收集菌体。

6) 新鲜的菌体沉淀用 30ml 预冷的 PBS 重悬, 4°C, 4500 rpm 离心 10 min, 收集菌体。

7) 用 20ml HGX 缓冲液 (cocktail+PMSF) 重悬菌体, 充分混匀。

8) 超声波破碎: 超声波处理条件: 4s on, 8s off, 150 瓦特, 工时 5min。

9) 4°C, 12000 rpm 离心 30 min, 留上清液, 并用 0.22µm 滤膜过滤, 除去少量菌体。

10) 在 20ml 上清液中加入 100 µl 10% 的 PEI, 充分混匀沉淀细菌 DNA, 在冰上静置 10 分钟。

11) 4°C, 12000 rpm 离心 10 min, 留上清液, 并用 0.22µm 滤膜过滤。

12) 滤液中的 PAT 蛋白通过亲和层析纯化, 并用截留分子量为 10 kD 的透析袋透析含有高浓度盐的 PAT 蛋白, 4°C, 12h。

13) 透析完后, 用截留分量为 30kD 的超滤管浓缩 PAT, 4°C, 4000 rpm 离心至体积为 500 µl。

14) 加入 500 µl 的 100%甘油, 轻轻混匀, -20°C 保存。

对纯化好的蛋白进行电泳分析, 结果如图 3 所示 (我们利用 BSA 为标准品来定量纯化好的蛋白浓度。定量之前, 先取 10 µl 纯化好的蛋白进行 15 倍稀释, 然后将稀释好的蛋白分别取 1 µl 和 2 µl 与 BSA 标准品进行比较从而定量。) 获得质量好酶活高的 PAT 融合蛋白为高通量单细胞测序的相关应用奠定了基础, 我们使用的一步亲和层析纯化方式 (PEI 沉淀细菌 DNA) 是最理想的选择。利用该亲和层析纯化 PAT 的优势是:

1 步骤简便, 可操作性强, 所需时间短;

2 产量高;

3 蛋白纯化费用少;

4 可放大纯化规模。

我们纯化的 PAT 无细菌 DNA 残留, 一步纯化的纯度达到 95%以上, PAT 的产量可以达到 14 mg/L 培养基, 足够进行 100 万个单细胞文库的制备。

实施例 2、PAT 融合蛋白各个功能嵌段的优化

一、linker 序列的优化筛选

选择特异的 linker(-GGSDDDKEF-)序列连接 PAT 的不同功能片段, 能够保护 ProteinA 和 Tn5 的天然构象, 从而获得活性更好的融合蛋白。

连接不同 linker 的融合蛋白的方法和活性验证的步骤如下:

1、在 Tn5 的原有序列上连上 2xProtein A (IgG 识别序列) 时, 通过 PCR 引入不同的 linker 序列 (见表 1, 共 10 种 linker 序列)。

表 1 不同 linker 序列

Linker 序号	Linker 类型	Linker 序列
1	柔性	(GGGGS)
2	柔性	(GGGGGG)
3	柔性	(GGGGGGGG)
3	刚性	(EAAAK)
4	柔性	(GGSDDDKEF)
5	柔性	(DDDKEF)
6	柔性	(GSGGSGDDDKEF)
7	柔性	(SSG)
8	刚性	(PAPAP)
9	刚性	(AAAKEAAKA)
10	刚性	A(EAAAK) ₄ ALEA(EAAAK) ₄ A
11	柔性	(LGGGGSGGGSGGGGSAAA)
12	刚性	(LAAA)

2、按照实施例 1 所述的方法获得通过不同 linker 连接的 PAT，随后对其进行功能鉴定，方法如下：

- 1) 取 10 μ l IgG beads，用 200 μ l HGX 缓冲液洗涤三次，4°C，2000g 离心 3 min。
- 2) 加入 6 μ g Tn5（没有融合 Protein A）或 PAT（融合 Protein A），放大到 100 μ l 体系，4°C，孵育 1h。
- 3) 用 200 μ l HGX 缓冲液漂洗三次，4°C，2000g 离心 3 min，弃掉上清。
- 4) 用 20 μ l 0.5 M HAc(pH 3.4) 洗脱 Tn5 或 PAT，用 1M NaOH 调节 pH 至 7.2。
- 5) 4%-7.5% SDS-PAGE 检测，

我们进行了一组的 IP（蛋白质免疫沉淀）对照试验，检测结果如图 4 所示，每组做了两个平行，当没有融合 Protein A 的 Fc 结构域时，Tn5 不能被 IgG beads 特异捕获，Tn5 蛋白最后全出现在废液中 (post-SN)，当 Tn5 融合 Protein A 的 Fc 结构域时并以 linker(-GGSDDDKEF-)序列连接时，PAT 全部被 IgG beads 捕获，证明 Protein A 可以高效(效率为 98%)的识别 IgG。

通过添加的 linker 筛查证明，linker(-GGSDDDKEF-)序列可以保护 Protein A 和 Tn5 的天然构象，不影响 Tn5 的酶活，并且不影响 Protein A 蛋白对抗体的识别，如图 4 的结果所示。（见下方结果描述，IP1 和 IP2 指的是一个实验的两个平行，我们最后所使用的 linker 序列在表 1 中是编号 4）。

通过上述筛选，最后确定在保护 Protein A 和 Tn5 的天然构象以及维持 Tn5 超高酶活方面有着特殊优势。

二、Protein A 和 Tn5 嵌段的融合方式对 PAT 活性的影响试验

所述试验方法包括以下步骤：

- 1) 在 Tn5 的 N 端融合两个来源于金黄色酿脓葡萄球菌 Protein A 蛋白的 IgG 识别结构域，并命名为 pA2-Tn5。
- 2) 在 Tn5 的 C 端融合两个来源于金黄色酿脓葡萄球菌 Protein A 蛋白的 IgG 识别结构域，并命名为 Tn5-pA2。
- 3) 将纯化完的 pA2-Tn5 和 Tn5-pA2 装配上反应接头（连上切割 adaptors）。
- 4) 取 500ng 的小鼠基因组作为反应底物，用不同浓度的 pA2-Tn5 和 Tn5-pA2，55°C 反应 10 min。
- 5) 用 stopping 缓冲液终止反应，55°C 反应 5 min，经 1.5% 琼脂糖凝胶电泳检测，如图 5 所示。

通过实验比较，如图 5 的结果所示，随着加入的 PAT 量的增加，酶切产生的 DNA 片段越来越小，并且当 Protein A 的编码序列添加在 Tn5 的 N 端时 (pA2-Tn5)，纯化出的 Tn5 酶活更高，是 C 端融合 Protein A (Tn5-pA2) 酶活的 16.7 倍，我们将酶活更高的 pA2-Tn5 统一简称为 PAT。实施例 3、自然状态和细胞交联 ChIP-seq 方法验证 PAT 活性

- 1、PAT ChIP-seq 方法一（细胞不交联，处于自然状态），包括以下步骤：

1) 收集约 1,000,000 个体外培养的胚胎干细胞, 用 PBS 洗 2 次, 离心收集细胞后用洗涤缓冲液 1 洗 3 次。

2) 用结合缓冲液重悬细胞, 加入适量的抗体, 所述抗体为 H3K4me3 抗体 (Millipore, 04-745, Lot: 2872328), 在 4℃ 孵育 30min, 使抗体与目标蛋白充分结合。

3) 用洗涤缓冲液 2 洗涤细胞 3 次, 除去多余没有结合的抗体。

4) 用洗涤缓冲液 2 重悬细胞, 之后加入 PAT 融合蛋白, 在 4℃ 孵育 30min, 使得融合蛋白与抗体充分结合。

5) 用洗涤缓冲液 2 洗涤细胞 3 次以除去多余的 PAT 融合蛋白。

6) 加入反应缓冲液, 激活 PAT 融合蛋白的活性, 为降低反应背景, 在 4℃ 下进行反应, 反应 30min。

7) 加入终止缓冲液终止反应, 反应终止后, 用 QIAGEN DNA 纯化试剂盒进行纯化。

8) 纯化好的 DNA 直接用 NEB Nextera index 引物进行 PCR 扩增完成建库。

9) 将完成建库的文库用于二代测序。

检测上述方法和下述对比例 1 (基于普通超声波的 H3K4me3 Sonicate ChIP-seq) 所述方法的文库背景, 具体检测结果详见图 6-A 和图 7-A。

根据图 6-A (Rep1 和 Rep2 是两个实验重复) 的实验结果可知, 使用所述 PAT 融合蛋白的 ChIP-seq 能够获得与对比例 1 所述超声 ChIP-Seq 基本相同的阳性信号, 同时, 其背景信号 (如图 6-A 所示的阴影部分) 要明显低于对比例 1。

根据图 7-A 所示结果可知, 与对比例 1 相比, 使用所述 PAT 融合蛋白的 ChIP-seq 获得的 H3K4me3 信号在 TSS 附近更为集中, 表现出更高的信噪比, 而对比例 1 的信号在启动子区域比较分散。

2、PAT ChIP-seq 方法二 (细胞交联), 包括以下步骤:

1) 收集 1,000,000 个胚胎干细胞, 用 1% 甲醛室温交联 3-10min, 甘氨酸中和后用 PBS 洗 3 次。

2) 用含 0.3% SDS 的低渗溶液重悬细胞, 37℃ 孵育 30min 使染色质充分打开。

3) 离心去掉上清。

4) 用结合缓冲液洗涤细胞 1 次, 然后用结合缓冲液重悬细胞并加入抗体, 所述抗体为 H3K27ac 抗体, 在 4℃ 孵育 30min 使抗体与蛋白充分结合。

5) 洗涤缓冲液 2 洗涤细胞 3 次, 除去多余没有结合的抗体。

6) 用洗涤缓冲液 2 重悬细胞, 之后加入 PAT 融合蛋白, 4℃ 孵育 30min, 使得融合蛋白与抗体充分结合。

7) 洗涤缓冲液 2 洗涤细胞 3 次以除去多余的没有结合的 PAT 融合蛋白。

8) 加入反应缓冲液, 激活 PAT 融合蛋白的活性, 为降低反应背景, 在 4℃ 下进行反应, 反应 30min。

9) 加入终止缓冲液终止反应, 反应终止后, 用 QIAGEN DNA 纯化试剂盒进行纯化。

10) 纯化好的 DNA 直接用 NEB Nextera index 引物进行 PCR 扩增完成建库。

11) 将完成建库的文库用于二代测序。

检测上述方法和下述对比例 2 (基于普通超声波的 H3K27ac Sonicate ChIP-seq) 所述方法的文库背景, 具体检测结果详见图 6-B 和图 7-B。

根据图 6-B (Rep1 和 Rep2 是两个实验重复) 的实验结果可知, 使用所述 PAT 融合蛋白的 ChIP-seq 能够获得与对比例 2 所述超声 ChIP-Seq 基本相同的阳性信号, 同时, 其背景信号 (如图 6-B 所示的阴影部分) 要明显低于对比例 2。

根据图 7-B 所示结果可知, 与对比例 2 相比, 使用所述 PAT 融合蛋白的 ChIP-seq 获得的 H3K27ac 信号在 TSS 附近更为集中, 表现出更高的信噪比, 而对比例 2 的信号在启动子区域比较弱。

结果显示: 将自然状态下 H3K4Me3 In situ-ChIP-seq, 以及交联状态下 H3K27ac In situ-ChIP-seq 分别与对应的传统 Sonication ChIP-seq 比较, 可以得出 PAT 构建出的文库质量显然都要比传统的方法好, 信噪比更强。

3、PAT ChIP-seq 方法三 (原位组织切片), 包括以下步骤:

1) PBS 洗涤组织切片 3 次, 然后用洗涤缓冲液 1 洗涤切片 3 次。

2) 用结合缓冲液洗涤切片一次, 用结合缓冲液覆盖组织切片, 加入抗体, 所述抗体为 H3K4Me3 抗体, 在 4℃ 孵育 1h 使抗体与蛋白充分结合。

3) 洗涤缓冲液 2 洗涤切片 3 次, 除去多余没有结合的抗体。

4) 用洗涤缓冲液 2 覆盖组织切片, 之后加入 PAT 融合蛋白, 4℃ 孵育 30min, 使得融合蛋白与抗体充分结合。

5) 洗涤缓冲液 2 洗涤切片 3 次以除去多余的 PAT 融合蛋白。

6) 加入反应缓冲液, 激活 PAT 融合蛋白的活性, 为降低反应背景, 在 4℃ 下进行反应, 反应 30min。

7) 加入终止缓冲液终止反应, 反应终止后直接加入 NEB Nextera index 引物和等温 DNA 聚合酶 (如

Phi29 DNA 聚合酶)在一定温度下进行 PCR 扩增建库。

8) PCR 完成后,将 DNA 洗脱下来,用 AMP beads 进行纯化完成建库。

9) 建库完成的文库直接用于二代测序。

结果如图 8 所示,我们对早期 E7.75 的胚胎进行组织切片(图 8-A),利用本公开的原位 ChIP-seq(图 8-B),从空间位置上探索了重要转录因子周围的 H3K4me3 在胚胎发育早期(原肠运动)的动态变化(图 8-C)。这项技术的实现有利于我们对细胞或组织本身进行直接操作,屏蔽了外界条件造成的背景干扰,真正能够做到探求细胞内最真实的生命调控过程。

实施例 4、PAGT 融合蛋白的构建及功能验证

本实施例提供一种 PAT 融合蛋白的改进方案,由于 Protein G 对来源于小鼠的 IgG 有极高的亲和力,因此,我们进一步开发出 PAGT 融合蛋白,在 PAT 中进一步融合 Protein G 蛋白的 IgG 识别结构域,丰富了抗体来源的多样性:

一、PAGT 蛋白的构建

在 Tn5 的 N 端连上 Protein A 和 Protein G 的 IgG 识别结构域(单体顺序连接两个蛋白,从 N 端到 C 端顺序依次是 2xProtein A, Protein G, Tn5, 然后构成同源二聚体,各个功能嵌段的 linker 和 PAT 的一致,linker 序列都为-GGSDDDDKEF-),具体的连接质粒的构建及蛋白纯化方法同实施例 1。

二、PAGT 蛋白的功能验证

利用 PAT 和 PAGT 进行少量细胞的 ChIP,按照上述实施例 3 所述的 ChIP-seq 检测方法的步骤对所述 PAGT 蛋白的功能进行验证(如图 9-10 所示)。

结果显示:通过 H3K27ac (abcam 4729, GR3216173-1) ChIP-seq 的实验比较, PAGT 与 PAT 构建出的文库质量显然都要比传统的方法好,信噪比更强,并且 PAGT 和 PAT 构建出的文库质量没有明显差异。

实施例 5、PAT 融合蛋白进行高通量 ChIP-seq 检测方法(细胞不交联,处于自然状态)

包括以下步骤:

1) 收集约 200,000 个体外培养的胚胎干细胞,用 PBS 洗 2 次,离心收集细胞后用洗涤缓冲液 A 洗 1 次。

2) 用 1ml 洗涤缓冲液 A 重悬,加入 10 μ l 经激活缓冲液激活的 ConA beads,室温反应 10min,回收细胞,弃掉上清。

3) 用 100 μ l 洗涤缓冲液 C 重悬细胞,加入适量的抗体(一般 0.5 μ g/100 μ l),所述抗体为 H3K27ac 抗体(abcam 4729, GR3216173-1,该抗体须全程添加去乙酰化酶抑制剂),在 4 $^{\circ}$ C 孵育 4h,使抗体与目的蛋白充分结合。

4) 用洗涤缓冲液 B 洗涤细胞 2 次,除去多余没有结合的抗体。

5) 用 1% BSA/PBS 液重悬细胞。

6) 准备一个 96 孔板,用洗涤缓冲液 C 配制终浓度为 3 μ g/ml T5/T7 barcoded PAT,每一个孔中都加入 100 μ l 如图 11。

7) 结合有 ConA beads 的细胞经 FACS 分选进上述 96 孔板,每孔 2000 个细胞,在 4 $^{\circ}$ C 孵育 1h,使得 PAT 与抗体充分结合。

8) 弃掉每孔的 PAT,用 180 μ l 洗涤缓冲液 C 洗涤细胞 2 次以除去多余的融合蛋白。

9) 加入 10 μ l 单细胞反应缓冲液,激活融合蛋白的活性,为降低反应背景,在 25 $^{\circ}$ C 下进行反应,反应 60min。

10) 加入 10 μ l 单细胞终止缓冲液终止反应,室温 15min。

11) 反应终止后,每孔加入 20 μ l 单细胞分选缓冲液,合并所有的孔,经 DAPI 冰上染色 15 min。

12) 经 30 μ m 的过滤筛过滤,除掉细胞团块,回收细胞。

13) 用适量的单细胞分选缓冲液重悬细胞。

14) 再准备一个 96 孔板,每孔加入 4 μ l 单细胞裂解缓冲液,结合有 ConA beads 的细胞经 FACS 分选进上述 96 孔板,每孔 20 个细胞。

15) 于 65 $^{\circ}$ C 反应 6h,反应完全后于 85 $^{\circ}$ C, 15 min 失活蛋白酶 K。

16) 每孔加入 1 μ l 1.8% tritonX-100, 37 $^{\circ}$ C 反应 60 min。

17) 采用 Mosaic Truseq 建库方法进行原管建库(Mosaic Truseq 建库流程如图 12 所示)

实施例 6、PAT 融合蛋白进行高通量 ChIP-seq 检测方法(细胞交联)

包括以下步骤:

1) 收集 200,000 个胚胎干细胞,1% 甲醛室温交联 3min,甘氨酸中和后用 PBS 洗 3 次。

2) 用 200 μ l 含 0.3% SDS 的低渗溶液重悬细胞,于 62 $^{\circ}$ C 孵育 10min 使染色质充分打开。

3) 加入 20 μ l 20% Triton X-100, 37 $^{\circ}$ C, 孵育 60min。

4) 加入 1 ml 洗涤缓冲液 A, 加入 10 μ l 经激活缓冲液激活的 ConA beads, 室温反应 10min, 回收细胞, 弃掉上清。

5) 用 100 μ l 洗涤缓冲液 C 重悬细胞, 加入适量的抗体 (一般 0.5 μ g/100 μ l), 所述抗体为 H3K27ac 抗体 (该抗体须全程添加去乙酰化酶抑制剂), 在 4 $^{\circ}$ C 孵育 4h, 使抗体与目的蛋白充分结合。

6) 用洗涤缓冲液 B 洗涤细胞 2 次, 除去多余没有结合的抗体。

7) 用 1% BSA/PBS 液重悬细胞。

8) 准备一个 96 孔板, 用洗涤缓冲液 C 配制终浓度为 3 μ g/ml T5/T7 barcoded PAT, 每一个孔中都加入 100 μ l 如图 11。 ,

9) 结合有 ConA beads 的细胞经 FACS 分选进上述 96 孔板, 每孔 2000 个细胞, 在 4 $^{\circ}$ C 孵育 1h, 使得 PAT 与抗体充分结合。

10) 弃掉每孔的 PAT, 用 180 μ l 洗涤缓冲液 C 洗涤细胞 2 次以除去多余的融合蛋白。

11) 加入 10 μ l 单细胞反应缓冲液, 激活融合蛋白的活性, 为降低反应背景, 在 37 $^{\circ}$ C 下进行反应, 反应 60min。

12) 加入 10 μ l 单细胞终止缓冲液终止反应, 室温 15min。

13) 反应终止后, 每孔加入 20 μ l 单细胞分选缓冲液, 合并所有的孔, 经 DAPI 冰上染色 15 min。

14) 经 30 μ m 的过滤筛过滤, 除掉细胞团块, 回收细胞。

15) 用适量的单细胞分选缓冲液重悬细胞。

16) 再准备一个 96 孔板, 每孔加入 4 μ l 单细胞裂解缓冲液, 结合有 ConA beads 的细胞经 FACS 分选进上述 96 孔板, 每孔 20 个细胞。

17) 于 65 $^{\circ}$ C 反应 6h, 反应完全后于 85 $^{\circ}$ C, 15 min 失活蛋白酶 K。

18) 每孔加入 1 μ l 1.8% tritonX-100, 37 $^{\circ}$ C 反应 60 min。

19) 采用 Mosaic Truseq 建库方法进行原管建库。

实施例 7、单细胞 Mosaic Truseq 两步 PCR 文库制备方法

本实施例提供一种新的单细胞 Mosaic Truseq 两步 PCR 的文库制备方法, 得到的单细胞文库可以使用 illumina 标准测序方法和平台。与特殊定制的单细胞测序方法相比, Mosaic Truseq 文库只需要 1/15 的测序成本。所述 Mosaic Truseq 建库方法包括以下步骤 (如图 12 所示):

1) 按照如下体系配制 PCR 反应体系:

细胞裂解物	5 μ l
5X Q5 reaction buffer	10 μ l
5X Q5 high GC enhancer	10 μ l
10 mM (each) dNTP	1 μ l
20 mg/ml BSA	0.5 μ l
Primer F1st/R1st Mix(50 uM)	0.5 μ l
1 mM MgCl ₂	23 μ l
Q5 polymerase	0.3 μ l

引物序列如下:

F1st—ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGTCGGCAGCGTCTCCACGC (SEQ ID NO.6)

R1st--GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTCGTGGGCTCGGCTGTCCCTGT (SEQ ID NO.7)

扩增程序为: 72 $^{\circ}$ C 5 min, 98 $^{\circ}$ C 30 s, 13 个循环(98 $^{\circ}$ C 10 s + 63 $^{\circ}$ C 30 s + 72 $^{\circ}$ C 1 min), 72 $^{\circ}$ C 延伸 5 min.

2) 每孔加入 0.5 μ l 20 U/ μ l Exo I (NEB M0293S), 37 $^{\circ}$ C 孵育 30min, 72 $^{\circ}$ C 孵育 20 min 失活 Exo I。

3) 每孔加入 1 μ l 25uM 5'index (Truseq P5 index) 和 1 μ l 25uM 3'index (Truseq P7 index), 8 μ l 反应混合液。

上一轮产物	50 μ l
5X Q5 reaction buffer	2 μ l
5X Q5 high GC enhancer	2 μ l

10 mM (each) dNTP	0.5 μ l
Primer (5' 和 3' index)	各 1 μ l
1 mM MgCl ₂	3.4 μ l
Q5 polymerase	0.1 μ l

扩增程序为: 98°C 30 s, 5 个循环(98°C 10 s + 63°C 30 s + 72°C 1 min), 72°C 延伸 5 min.

4) PCR 进行完之后, 混合 96 孔板的 96 孔, 通过柱纯化, 60 μ l elution buffer 洗脱 DNA。

5) 用 1.0 x Ampure XP-beads 纯化一次, 用 (0.5+0.5) x Ampure XP-beads 做一次 size selection, 最后用 20 μ l H₂O 溶解 DNA。

6) illumina 测序平台双端 150bp 测序。

不仅仅限于少量细胞, 本公开的高通量单细胞 ChIP-seq 技术适用于所有的细胞类型(培养的细胞或组织消化的细胞等均可); 适用于甲醛交联或自然状态下的细胞; 适用于组蛋白修饰, DNA 结合蛋白和转录因子。总之, PAT 和 PAGT 同时构建了一套完整的选择方案: 不受抗体来源限制, 不受实验样品来源限制, 不受实验条件限制, 可以很方便的制作出高质量的 ChIP-seq 文库。

对比例 1&2、基于普通超声波的 ChIP-seq

对比例 1、基于普通超声波的 H3K4me3 ChIP-seq (H3K4me3 sonication ChIP)

对比例 2、基于普通超声波的 H3K27ac ChIP-seq (H3K27ac sonication ChIP)

操作方法相同, 包括如下步骤:

1) 收集 1,000,000 个体外培养的细胞, 用甲醛室温交联 10min, 用甘氨酸中和后用 PBS 洗涤细胞 3 次, 然后用液氮速冻。

2) 用 1ml 低渗溶液(加蛋白酶抑制剂)将细胞悬起, 在冰上孵育 15min。

3) 用玻璃 Dounce pestle B 将细胞混 10-20 次使细胞进一步分散, 并将细胞破碎。

4) 4°C, 3000rpm 离心 5min 以收集细胞核。

5) 用 100 μ l 细胞核裂解液(含 1% SDS)重悬细胞核, 轻轻混匀, 4°C 孵育 30min 以充分裂解细胞核。

6) 孵育后, 简单离心, 然后加 ChIP 稀释缓冲液稀释 SDS 浓度到 0.3%, 轻轻混匀。

7) 超声(Q800R2): 时间: 6min; 程序: 15s on, 45s off; 能量: 600 瓦特。

8) 超声后, 加入 ChIP 稀释缓冲液, 混匀后 4°C, 20,000 离心 20min, 将上清转移至新的 EP 管中。

9) 准备 70 μ l Protein G, 用 1% BSA/PBS 洗 3 次, 然后分 20 μ l protein G 磁珠加入到步骤 8) 到上清中, 4°C 孵育 1 个小时, 以去除非一些特异结合的蛋白。分别加入 5 μ g H3K4me3 和 5 μ g H3K27ac 抗体到染色体上清中, 同时加入 1ml 1% BSA/PBS 到剩余的 50 μ l protein G 磁珠中, 4°C 过夜。

10) 1h 后, 用磁力架收集 protein G 磁珠, 转移上清到一个新的 EP 管中, 加入抗体, 4°C 孵育过夜。

11) 第二天, 将封闭过夜的 protein G 磁珠转移到染色质-抗体的混合液中, 4°C 孵育 4h。

12) 洗 protein G 磁珠。

13) 用磁力架收集磁珠, 然后进行如下洗涤过程。

a. 4°C, 用高盐洗涤缓冲液洗涤 protein G 磁珠 1 次, 5min/次。

b. 4°C, 用低盐洗涤缓冲液溶液洗涤 protein G 磁珠 3 次, 5min/次。

c. 4°C, 用 1ml 1 $\times$ TE 溶液简单洗涤磁珠 1 次。

d. 4°C, 用 800 μ l 1 $\times$ TE 溶液简单洗涤磁珠 1 次, 然后转移到一个新的 EP 管中。

14) 去除残留的液体, 然后加入 110 μ l ChIP 洗脱缓冲液洗脱磁珠上的 DNA, 将 ChIP 洗脱缓冲液和磁珠在 70°C 孵育过夜。

15) 第三天, 用磁力架将上清转移到一个新的 EP 管中, 然后加入 100 μ l TE 溶液再次洗脱磁珠上的 DNA, 将第二次洗脱的 TE 溶液和第一次的洗脱 ChIP elution buffer 结合在一起, 加入 3 μ l 10mg/ml 蛋白酶 K, 55°C 孵育 6-8h。

16) 第四天, 用 QIAGEN DNA 纯化试剂盒纯化 ChIPed DNA。

17) 纯化完的 DNA 用“NEBNext Ultra DNA library Prep Kit for Illumina”试剂盒进行常规建库, 常规建库过程需要 1 天。

18) 将建库好的文库进行二代测序。

试剂配方:

ChIP 洗脱缓冲液: 50mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS;

ChIP 稀释缓冲液: 0.01% SDS, 1% Triton X-100, 2mM EDTA, 20mM Tris-HCl (pH 7.5), 150mM NaCl;

低盐洗涤缓冲液：0.1%脱氧胆酸钠，1% Triton X-100，2mM EDTA，50mM HEPES (pH7.5)，150mM NaCl；

高盐洗涤缓冲液：0.1% 脱氧胆酸钠，1% TritonX-100，2mM EDTA，50mM HEPES (pH7.5)，500mM NaCl；

核裂解缓冲液：1% SDS，10mM EDTA，50mM Tris-HCl(pH8.0)。

最后应说明的是：以上各实施例仅用做帮助本领域技术人员理解本公开的实质，不作为对本公开保护范围的限定。

工业实用性

(1) 本公开所述融合蛋白、试剂盒和方法在 ChIP-seq 检测过程中能够提高建库效率，降低文库背景，从而提高 ChIP-seq 检测方法的准确性并简化 ChIP-seq 的实验流程。

(2) 本公开所述方法对 Tn5 转座酶和 Fc 结合蛋白的突变体、表达体系、纯化手段、抗体、融合蛋白的孵育条件以及切割反应的条件进行优化，提高抗体、融合蛋白的结合效率以及 DNA 切割的效率和准确性，从而降低文库的背景和分辨率；

(3) 本公开所述方法能够对组织切片、细胞涂片和细胞爬片进行原位检测，不需要裂解细胞核以及超声打断染色质，保留原样本的空间分辨率。

(4) 本公开所述方法能够实现多器官多组织的单细胞 ChIP-seq，可以同时平行处理多种不同组织来源，或者多种不同病理状态下的细胞。

(5) 本公开所述方法能够快捷的完成单细胞 ChIP-seq 的制备，从实验开始，单日便可完成。

(6) 本公开所述方法能够实现高通量的单细胞 ChIP-seq，能够一次获得成上万个单细胞的 ChIP-seq 文库。

(7) 本公开所述方法不需要使用特殊定制的高成本的单细胞文库测序方法，使用特殊设计的 Mosaic Truseq 两步 PCR 的文库制备方法，得到的文库可以使用 illumina 标准的测序方法和平台进行测序。与特殊定制的单细胞测序方法相比，一方面，Mosaic Truseq 文库只需要 1/15 的测序成本，另一方面，实验人员不需要精通专业的测序知识，即可进行大规模的单细胞测序。

权利要求书

1、一种用于染色质免疫共沉淀-测序（ChIP-seq）建库的融合蛋白，所述的融合蛋白是同源二聚体，其单体包括如下结构：

（1）第一功能嵌段：Tn5 转座酶突变型；

（2）第二功能嵌段：金黄色葡萄球菌 A 蛋白（Protein A）、链球菌 G 蛋白（Protein G）或 Protein A 和 Protein G 的组合；

（3）各个功能嵌段的连接结构（Linker），蛋白纯化标签。

2、根据权利要求 1 所述的融合蛋白，其特征在于，

所述的 Tn5 转座酶突变型为在氨基酸序列如 SEQ ID NO.1 所示的野生型 Tn5 转座酶的氨基酸结构上进行如下点突变：E54K、L372P 获得的 Tn5 转座酶突变型；

所述的 Linker 的氨基酸序列如 SEQ ID NO.3 所示；

所述的蛋白纯化标签为可通过亲和层析纯化的蛋白纯化标签。

3、根据权利要求 1 所述的融合蛋白，其特征在于，所述的 Tn5 转座酶突变型为在野生型 Tn5 转座酶，其氨基酸序列如 SEQ ID NO.1 所示，的氨基酸结构上进行如下点突变：E54K、L372P，所述的 Tn5 转座酶突变型的氨基酸序列如 SEQ ID NO.2 所示。

4、根据权利要求 3 所述的融合蛋白，其特征在于，通过氨基酸序列如 SEQ ID NO.3 所示的 Linker 连接所述融合蛋白的各个功能嵌段。

5、根据权利要求 3 所述的融合蛋白，其特征在于，所述的蛋白纯化标签为：-HIS 标签、GST 标签、MBP 标签、SUMO 标签、NusA 标签、或直接利用 Protein A 标签和 IgG 蛋白的特异亲和性，用 IgG 磁珠进行纯化所述融合蛋白。

6、根据权利要求 4 所述的融合蛋白，其特征在于，所述的蛋白纯化标签为：-HIS 标签、GST 标签、MBP 标签、SUMO 标签、NusA 标签、或直接利用 Protein A 标签和 IgG 蛋白的特异亲和性，用 IgG 磁珠进行纯化所述融合蛋白。

7、根据权利要求 1-6 中任一所述的融合蛋白，其特征在于，所述的第二功能嵌段连接在第一功能嵌段的氨基酸序列的 N 端（氮端）。

8、根据权利要求 1-7 中任一所述的融合蛋白，其特征在于，所述的融合蛋白的单体为 PAT 或 PAGT，

所述的 PAT 的氨基酸序列如 SEQ ID NO.4 所示，

所述的 PAGT 的氨基酸序列如 SEQ ID NO.5 所示。

9、权利要求 1-8 任一所述的融合蛋白在制备目标生物学样本的高通量测序文库中的应用，所述的生物学样本包括但不限于：未经交联或固定/经过交联或固定培养细胞样本、新鲜组织/经过交联或固定的组织样本；

所述的高通量测序文库为染色质免疫共沉淀-测序（ChIP-seq）文库。

10、使用权利要求 1-8 任一所述的融合蛋白制备目标生物学样本的高通量测序文库的方法，所述方法包括如下步骤：

（1）收集并处理目标生物学样本，获得单细胞悬液；

（2）用结合缓冲液重悬细胞，加入适量的特定蛋白结合抗体，使抗体充分结合后清洗除去未结合的抗体；

（3）重悬细胞，加入所述融合蛋白，使融合蛋白与抗体充分结合后除去未结合的融合蛋白；

（4）激活所述融合蛋白，充分反应获得被所述融合蛋白识别并切割的带有标记的 DNA 片段；

（5）加入终止缓冲液终止反应，用纯化试剂盒进行纯化 DNA 片段；

（6）加入引物进行 PCR 扩增完成建库；

所述的生物学样本包括但不限于：未经交联或固定/经过交联或固定培养细胞样本、新鲜组织/经过交联或固定的组织样本；

所述的高通量测序文库未染色质免疫共沉淀-测序（ChIP-seq）文库。

11、使用权利要求 1-8 任一所述的融合蛋白制备目标生物学样本的高通量测序文库的方法，所述方法包括如下步骤：

（1）收集并处理目标生物学样本，获得单细胞悬液；

（2）用结合缓冲液重悬细胞，加入适量的特定蛋白结合抗体，使抗体充分结合后清洗除去未结合的抗体；

（3）重悬细胞，加入所述融合蛋白，使融合蛋白与抗体充分结合后除去未结合的融合蛋白；

（4）激活所述融合蛋白，充分反应获得被所述融合蛋白识别并切割的带有标记的 DNA 片段；

（5）加入终止缓冲液终止反应，用纯化试剂盒进行纯化 DNA 片段；

（6）加入引物进行 PCR 扩增完成建库。

12、根据权利要求 11 所述的方法，其特征在于，所述的生物学样本包括但不限于：未经交联或固定/经过交联或固定培养细胞样本、新鲜组织/经过交联或固定的组织样本；

所述的高通量测序文库未染色质免疫共沉淀-测序（ChIP-seq）文库。

13、一种用于构建高通量测序文库的试剂盒，所述的试剂盒包括：

（1）权利要求 1-8 任一所述的融合蛋白；

（2）常规的建库过程所使用的样本洗涤缓冲液、结合缓冲液、激活缓冲液和建库引物等；

所述的高通量测序文库为染色质免疫共沉淀-测序（ChIP-seq）文库。

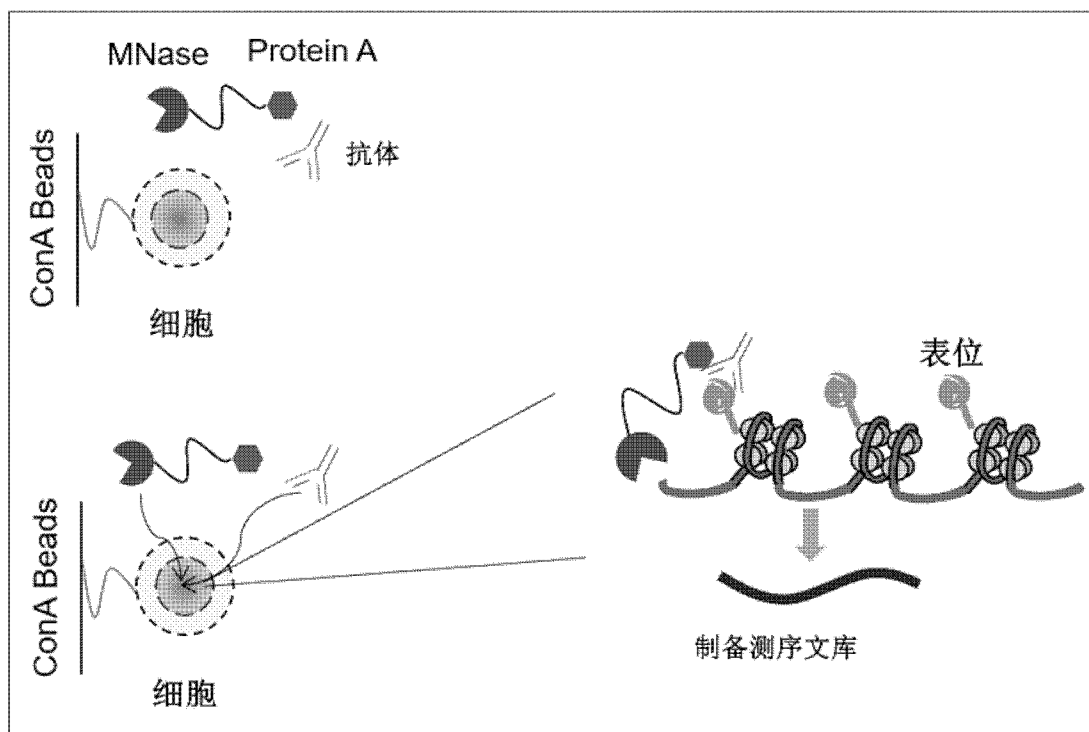


图 1

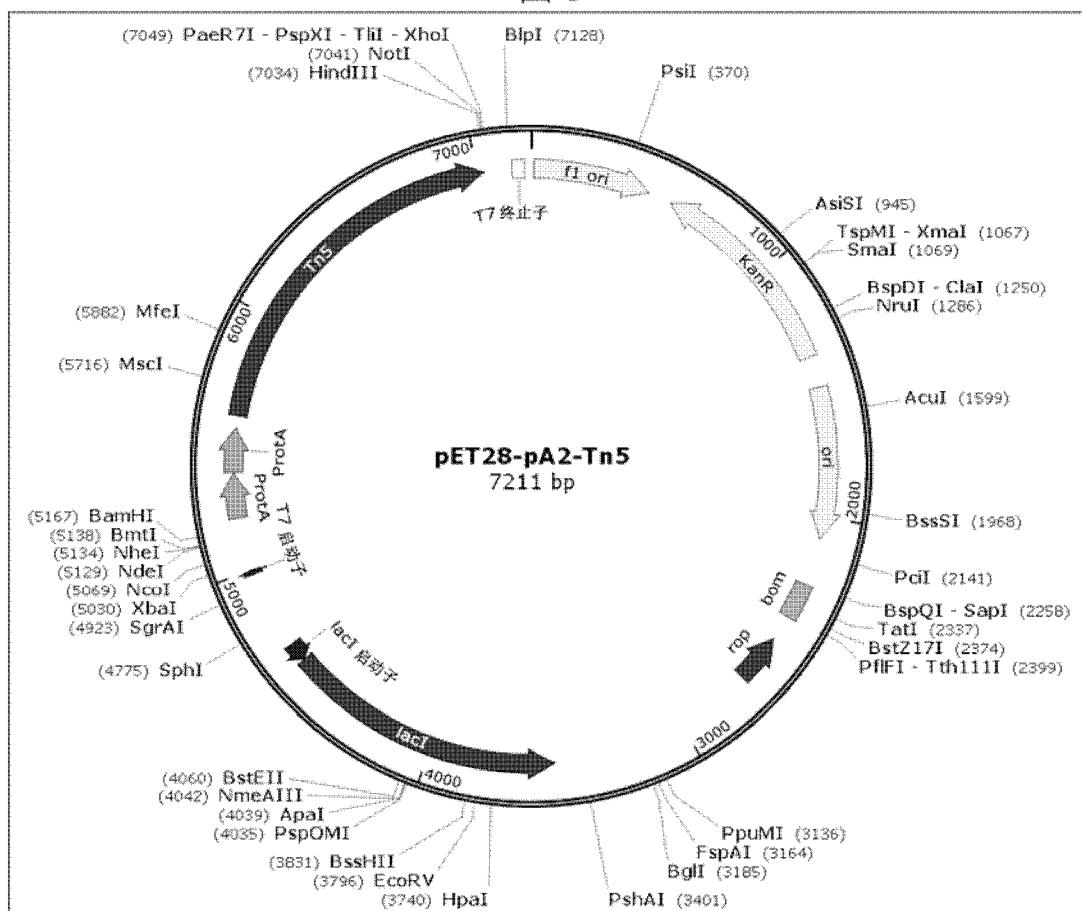


图 2

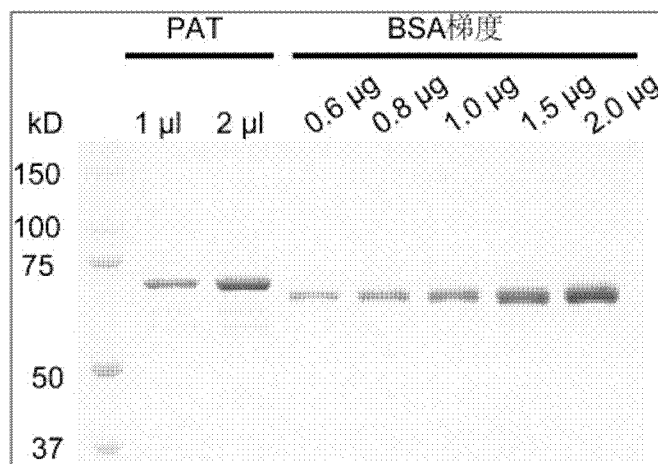


图 3

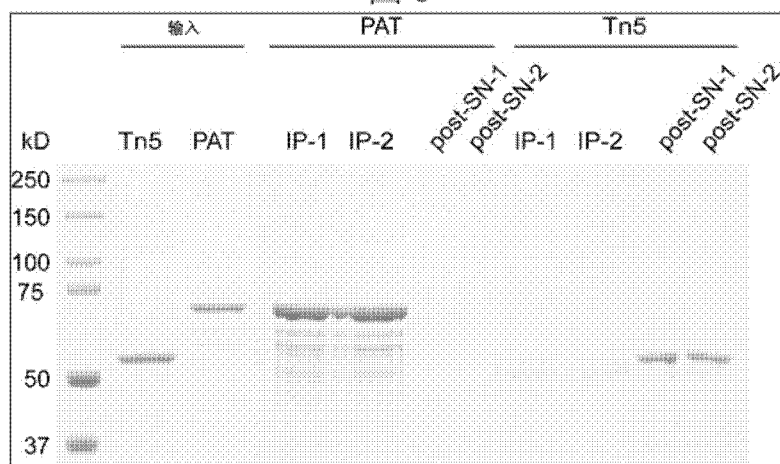


图 4

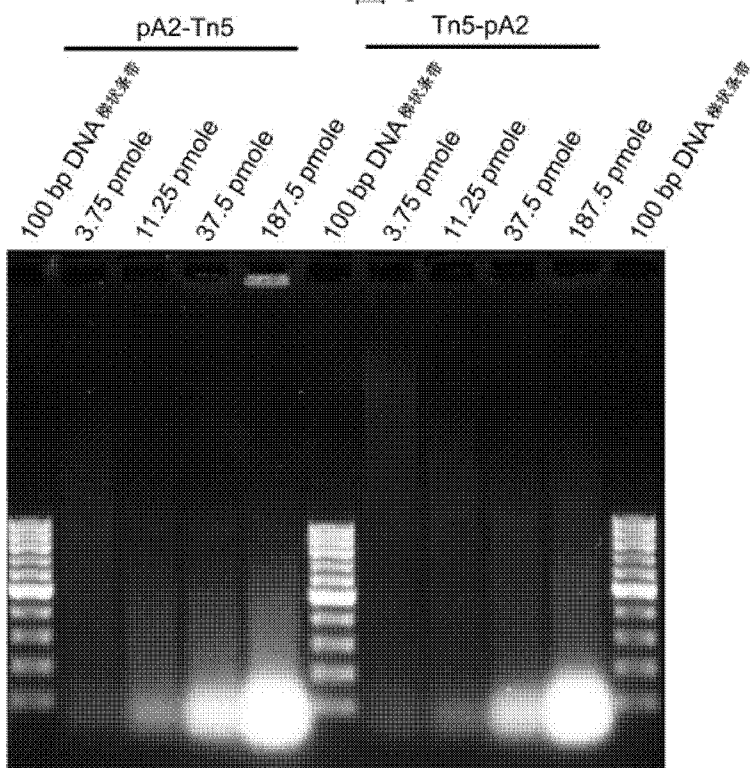
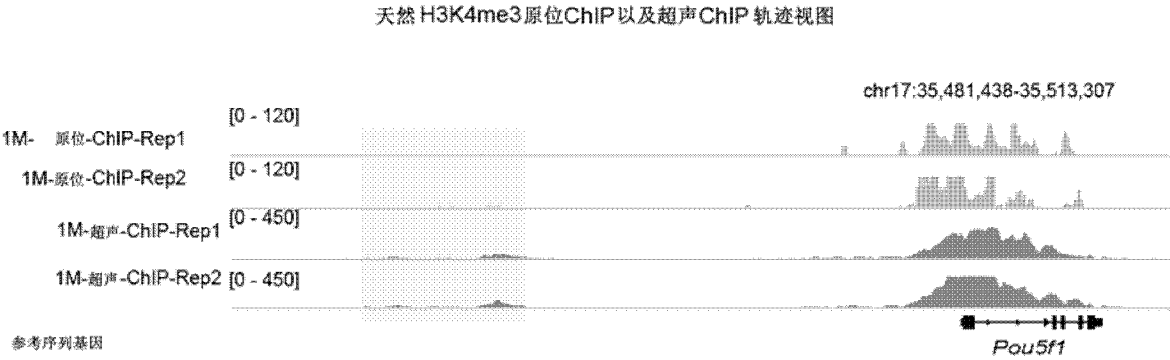


图 5

A



B

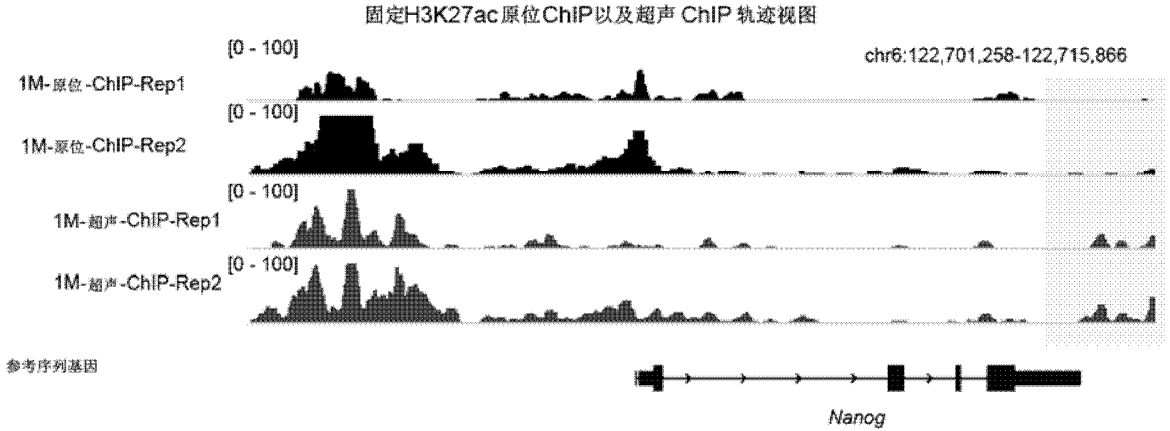


图 6

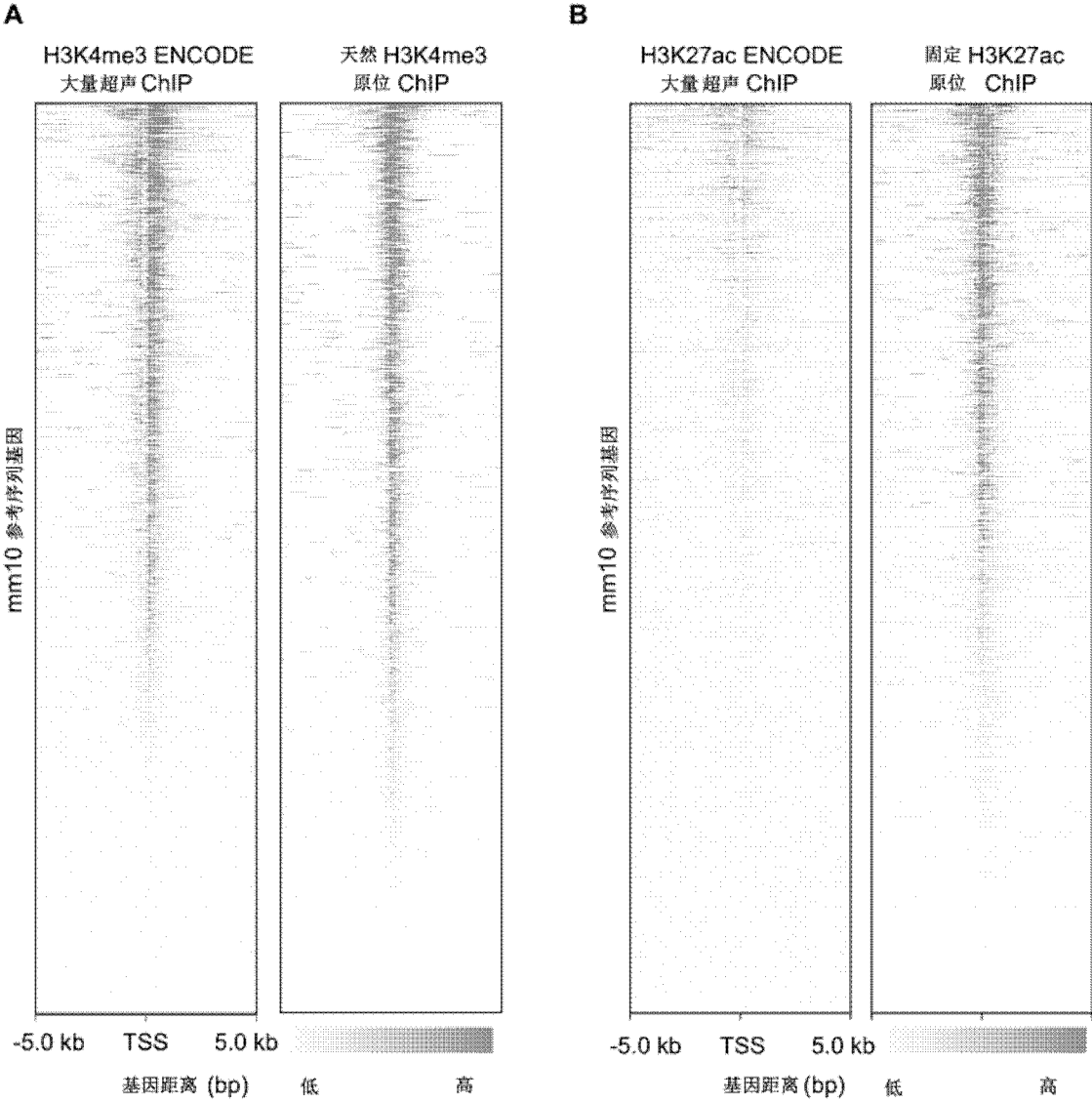


图 7

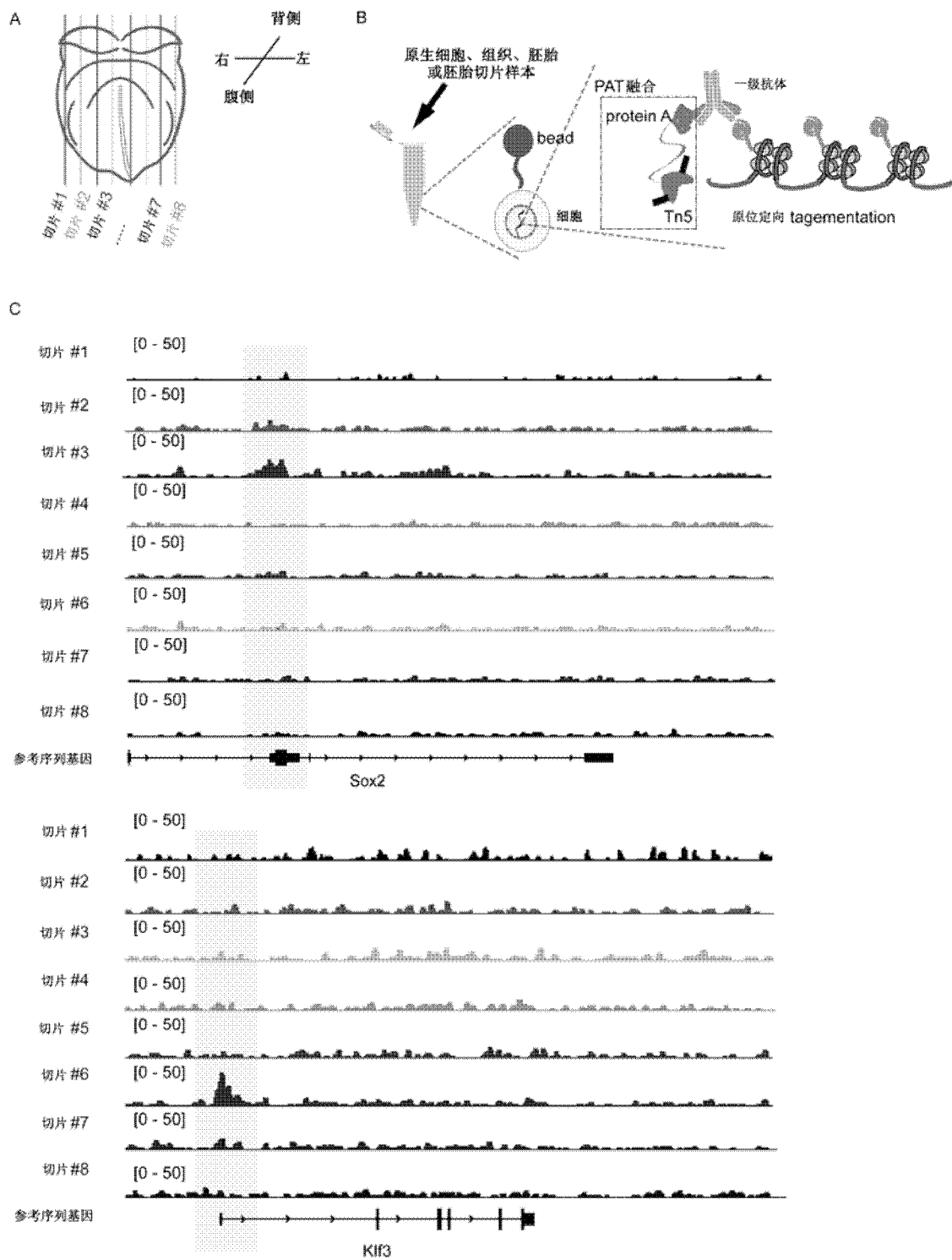


图 8

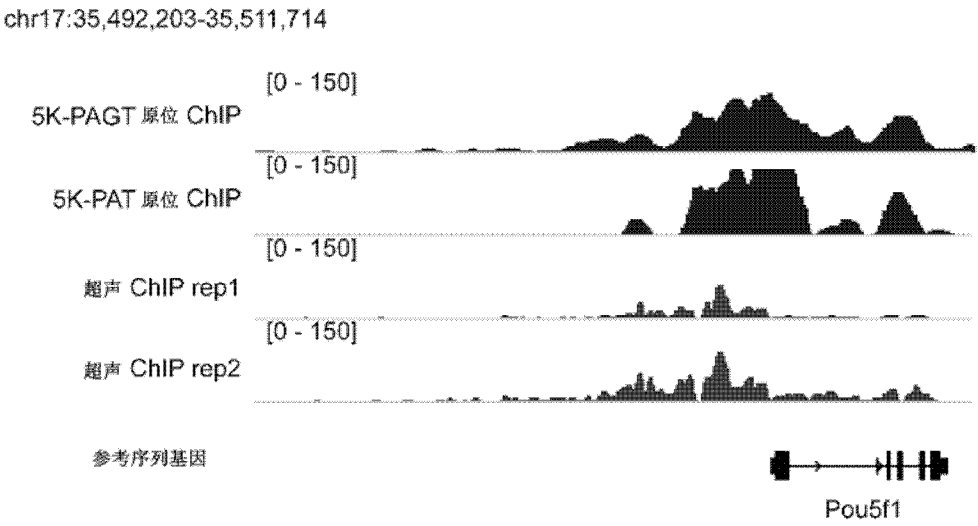


图 9

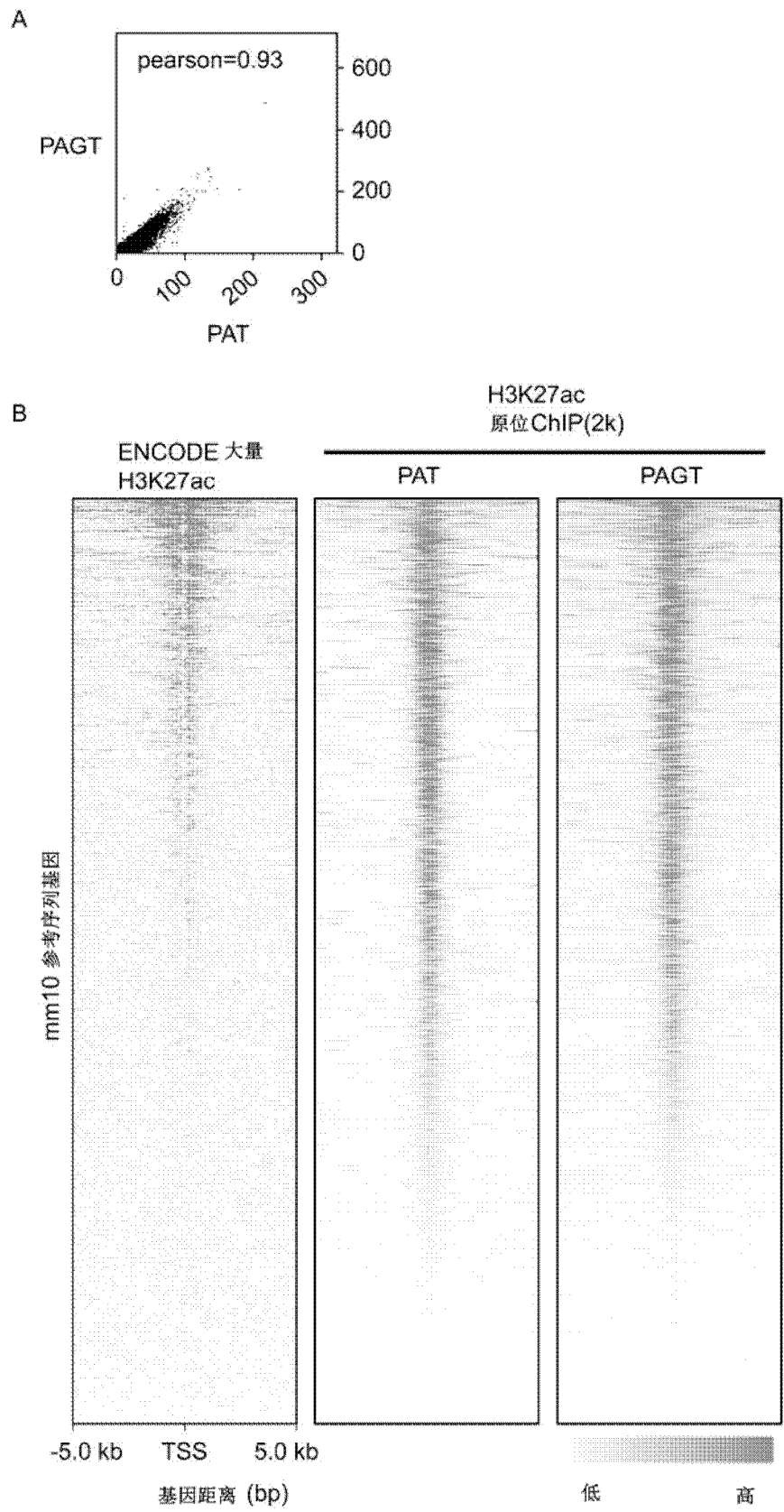


图 10

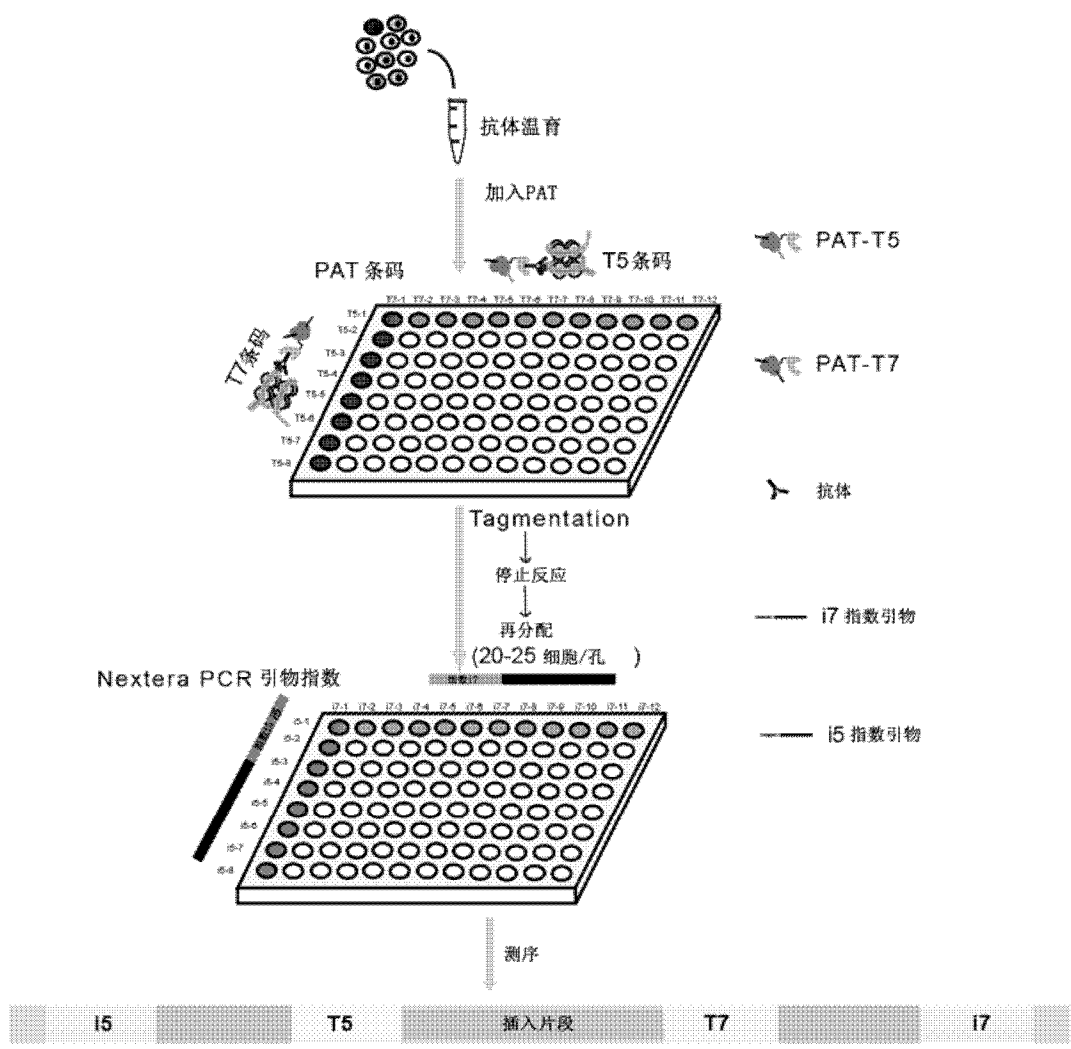


图 11

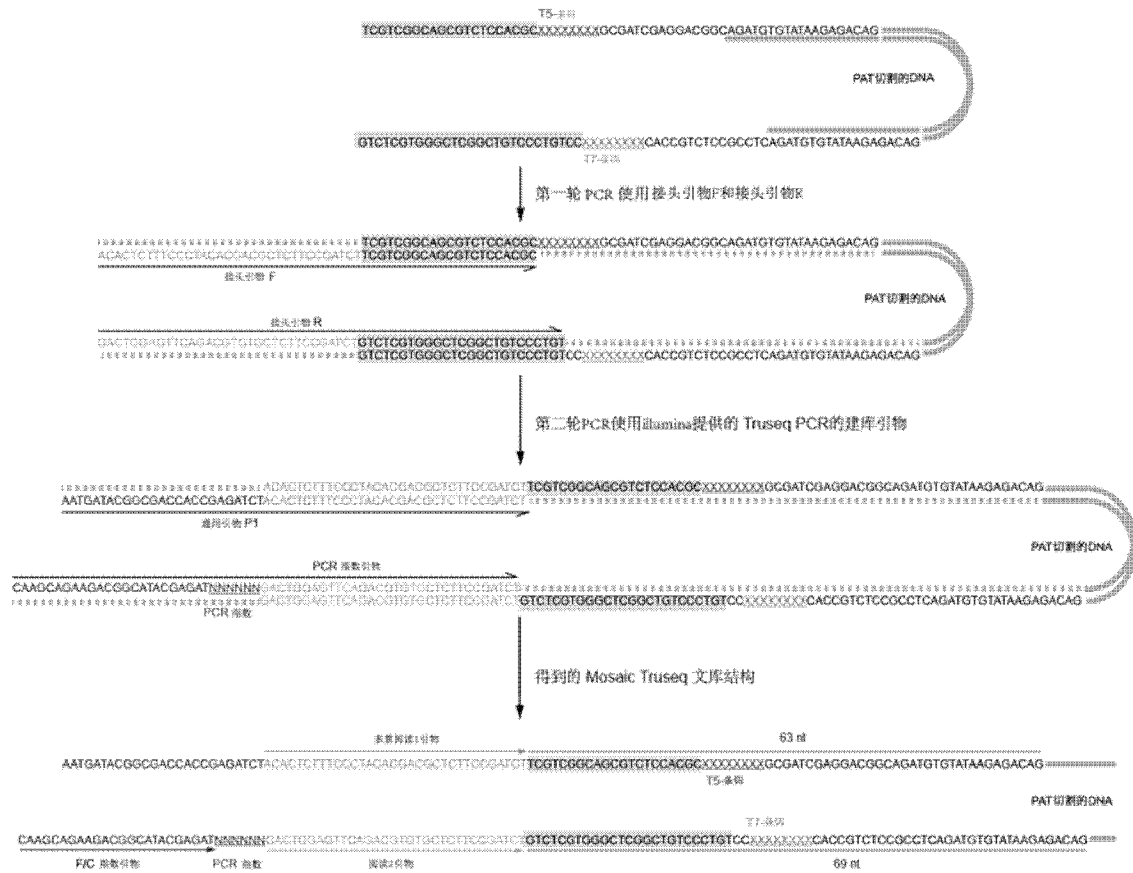


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/102513

A. CLASSIFICATION OF SUBJECT MATTER

C07K 19/00(2006.01)i; C12N 9/12(2006.01)i; C12Q 1/6806(2018.01)i; C12N 15/10(2006.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07K; C12N; 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)

CNABS, SIPOABS, DWPI, CNTXT, EPTXT, WOTXT, USTXT, CNKI, Web of ISI, NCBI, baidu, pubmed, patents, 中国专利序列检索系统, 北京大学/何爱彬, 染色质免疫共沉淀-测序, ChIP-seq, 融合蛋白, Tn5, protein A, 转座酶, A蛋白, SEQ ID NOs: 1-5

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108219006 A (SHANGHAI SINOBIO BIOTECHNOLOGY CO., LTD.) 29 June 2018 (2018-06-29) claims 1, 2, 8-15, 17, 28-30, description page 3 paragraph 29, page 10 paragraph 111	1-13
X	CN 108285494 A (PEKING UNIVERSITY) 17 July 2018 (2018-07-17) claims 1-10	1-13
PX	CN 110372799 A (PEKING UNIVERSITY) 25 October 2019 (2019-10-25) claims 1-7	1-13
A	ROSSI, M.J. et al. "Simplified ChIP-exo assays." <i>Nature Communications</i> , Vol. 9, 20 July 2018 (2018-07-20), article 2842, pages 1-13	1-13
A	SCHMIDL, C. et al. "ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors." <i>Nat Methods</i> , Vol. 12, No. 10, 31 October 2015 (2015-10-31), pp. 963-965	1-13



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"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

14 September 2020

Date of mailing of the international search report

09 October 2020

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/CN2020/102513

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KAYA-OKUR, H.S. et al. "CUT&Tag for efficient epigenomic profiling of small samples and single cells." <i>Nature Communications</i> , Vol. 10, 29 April 2019 (2019-04-29), article 1930, pages 1-10	1-13

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/102513

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/CN2020/102513

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108219006	A	29 June 2018	WO	2019184044	A1	03 October 2019
CN	108285494	A	17 July 2018	WO	2019153851	A1	15 August 2019
CN	110372799	A	25 October 2019	CN	110372799	B	09 June 2020

国际检索报告

国际申请号

PCT/CN2020/102513

A. 主题的分类

C07K 19/00(2006.01)i; C12N 9/12(2006.01)i; C12Q 1/6806(2018.01)i; C12N 15/10(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

C07K; C12N; 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))

CNABS, SIPOABS, DWPI, CNTXT, EPTXT, WOTXT, USTXT, CNKI, Web of ISI, NCBI, baidu, pubmed, patentics, 中国专利序列检索系统, 北京大学/何爱彬, 染色质免疫共沉淀-测序, ChIP-seq, 融合蛋白, Tn5, protein A, 转座酶, A蛋白, SEQ ID NOs: 1-5

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 108219006 A (上海欣百诺生物科技有限公司) 2018年 6月 29日 (2018 - 06 - 29) 权利要求1、2、8-15、17、28-30, 说明书第3页第29段、第10页第111段	1-13
X	CN 108285494 A (北京大学) 2018年 7月 17日 (2018 - 07 - 17) 权利要求1-10	1-13
PX	CN 110372799 A (北京大学) 2019年 10月 25日 (2019 - 10 - 25) 权利要求1-7	1-13
A	ROSSI, M. J. 等. "Simplified ChIP-exo assays." NATURE COMMUNICATIONS., 第9卷, 2018年 7月 20日 (2018 - 07 - 20), 第2842篇第1-13页	1-13
A	SCHMIDL, C. 等. "ChIPmentation: fast, robust, low-input ChIP-seq for hi- stones and transcription factors." Nat Methods., 第12卷, 第10期, 2015年 10月 31日 (2015 - 10 - 31), 第963-965页	1-13

☒ 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

2020年 9月 14日

国际检索报告邮寄日期

2020年 10月 9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

陈皓

电话号码 86-(10)-53962068

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
A	KAYA-OKUR, H.S. 等. "CUT&Tag for efficient epigenomic profiling of small samples and single cells." NATURE COMMUNICATIONS., 第10卷, 2019年 4月 29日 (2019 - 04 - 29), 第1930篇第1-10页	1-13

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/CN2020/102513

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	108219006	A	2018年 6月 29日	W0	2019184044	A1	2019年 10月 3日
CN	108285494	A	2018年 7月 17日	W0	2019153851	A1	2019年 8月 15日
CN	110372799	A	2019年 10月 25日	CN	110372799	B	2020年 6月 9日