

高畅,任鹏程,李明格,等. Tn5 转座子及其在生物技术领域的应用 [J]. 中国实验动物学报, 2022, 30(8): 1087-1094.  
Gao C, Ren PC, Li MG, et al. Tn5 transposon and its applications in biotechnology [J]. Acta Lab Anim Sci Sin, 2022, 30(8): 1087-1094.

Doi:10.3969/j.issn.1005-4847.2022.08.010

# Tn5 转座子及其在生物技术领域的应用

高畅<sup>1#</sup>, 任鹏程<sup>1#</sup>, 李明格<sup>2</sup>, 吕赫喆<sup>2</sup>, 胡仪<sup>1\*</sup>

(1. 北京师范大学生命科学学院, 北京 100875; 2. 中国科学院动物研究所, 北京 100101)

**【摘要】** 随着二代测序技术的快速发展, 测序及其相关技术愈发成熟并向高精度和高通量的方向深入。传统的测序方式对样本量要求很高, 而许多珍贵样本难以满足要求, 因此能够降低样本量需求的技术亟待开发。而 Tn5 转座酶在文库构建中可以在片段化 DNA 的同时将接头序列加入到靶序列两侧, 大大缩减了实验流程, 同时也降低了对样本量的需求。因此, Tn5 转座酶在生物技术领域的应用被逐渐开发, 例如 CUT&Tag、stLFR 和 TRACE-seq 等技术就是将 Tn5 转座酶应用于表观测序、长片段测序和转录组测序等领域的研究成果。简单的实验流程和极低的成本使其备受青睐。本文中, 我们对 Tn5 转座子及其在生物技术领域的应用进行了综述, 以期方便大家理解 Tn5 转座子及其相关技术, 为未来 Tn5 转座子相关技术的改进和推广提供参考。

**【关键词】** 二代测序; Tn5 转座子; 生物技术

**【中图分类号】** Q95-33 **【文献标识码】** A **【文章编号】** 1005-4847 (2022) 08-1087-08

## Tn5 transposon and its applications in biotechnology

GAO Chang<sup>1#</sup>, REN Pengcheng<sup>1#</sup>, LI Mingge<sup>2</sup>, LYU Hezhe<sup>2</sup>, HU Yi<sup>1\*</sup>

(1. College of Life Sciences, Beijing Normal University, Beijing 100875, China. 2. Institute of Zoology, Chinese Academy of Sciences, Beijing 100101)

Corresponding author: HU Yi. E-mail: yi.hu@bnu.edu.cn

**【Abstract】** With the rapid advancement of next-generation sequencing (NGS), sequencing and its related technologies are becoming more mature and providing higher accuracy and throughput. Traditional method of sequencing require a large sample quantity that is difficult to obtain for some samples. Thus, techniques with a lower demand on sample quantity are in urgent need. In library construction, the Tn5 transposase cuts DNA and adds an adapter sequence to both sides of the target sequence, which requires fewer steps and a lower sample quantity. Therefore, applications of Tn5 transposase have been gradually developed in biotechnology. For example, CUT&Tag, stLFR, and TRACE-seq are techniques that use Tn5 transposase in epigenomic sequencing, long fragment reading, and transcriptome sequencing. The simple steps and very low cost make the techniques popular. In this review, we summarize the Tn5 transposon and its applications in biotechnology to make it understood more easily and provide a reference for improvement and promotion in the future.

**【Keywords】** next-generation sequencing (NGS); Tn5 transposon; biotechnology

Conflicts of Interest: The authors declare no conflict of interest.

[基金项目] 国家自然科学基金 (32070401)。

Funded by National Natural Science Foundation of China (32070401).

[作者简介] 高畅 (2000—), 女, 在读本科生。Email: cherrelle1107@163.com;

任鹏程 (2001—), 男, 在读本科生。Email: tutu\_pcr@163.com。

#共同第一作者

[通信作者] 胡仪 (1985—), 女, 博士, 副教授, 博士生导师, 研究方向: 微生物生态学。Email: yi.hu@bnu.edu.cn

# 1 Tn5 转座子及转座机制

## 1.1 Tn5 转座子

Tn5 是一种复合型转座子,其中心序列可表达 3 种抗生素( $\text{kan}^r$ 、 $\text{str}^r$ 、 $\text{bleo}^r$ ),两侧为高度同源的倒置序列 IS50R 和 IS50L<sup>[1]</sup>(图 1A)。IS50R 可以表达产生转座酶(transposase, Tnp)和抑制转座酶活性的抑制蛋白(inhibitor, Inh), IS50L 表达产生两个非活性蛋白<sup>[2-3]</sup>。每个 IS50 序列都位于两个 19 bp 的倒置末端(外末端 outside end, OE 和内末端 inside end, IE)之间。倒置末端为转座酶的作用位点,但 OE 和 IE 在序列上略有不同<sup>[1,4]</sup>。后续的研究发现,转座的发生只需其中的关键成分,即转座酶和两个 19 bp 的倒置末端,内部可以包含任何足够长的 DNA 片段(图 1B)。并且,这两个 19 bp 的末端可以是倒置的 OE 和 IE、两个 OE、两个 IE,也可以是两个马赛克末端(mosaic end, ME)。实际上,ME 是一种人工构建的末端,其序列与 OE 存在 3 个碱基的不同,且具有比 OE 和 IE 更高的转座速率<sup>[5]</sup>。

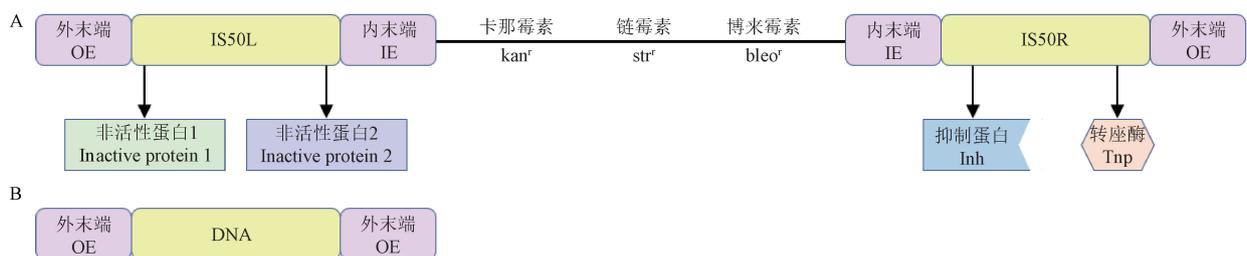
## 1.2 转座酶结构

由 IS50R 编码的 Tn5 转座酶可以分成 3 个结构域:与转座子末端 DNA 结合的 N 端结构域、具有催化活性的中部结构域和负责蛋白质结合的 C 端结构域。N 端结构域的二级结构仅包含  $\alpha$ -螺旋和转角,用于与转座子末端序列的识别和结合;C 端结构域也仅有  $\alpha$ -螺旋和转角,可以与转座酶结合促进转座过程,也可以同 Inh 结合形成二聚体抑制转座<sup>[6]</sup>。而转座酶的活性位点则位于由约 300 个氨基酸残基

组成的中部结构域内,该结构域内含有一个“核酸酶类 H 模体”起催化作用,且该模体在所有转座酶和逆转录病毒整合酶中都起类似的作用<sup>[6-8]</sup>。Tn5 转座酶的“核酸酶类 H 模体”活性中心内有 D97、D188、E326 作为保守的氨基酸残基组成 DDE 模体,对转座酶的活性有至关重要的作用。该 DDE 模体与两个  $\text{Mn}^{2+}$  结合,两个  $\text{Mn}^{2+}$  分别与 OE 末端的 3'-OH 和 5'-P 配位作为催化活性中心<sup>[6]</sup>。此外,在 DDE 模体的谷氨酸残基后还存在一对保守的赖氨酸残基(E $\times\times\times$ K $\times\times$ K),其中,第二个赖氨酸残基与 Y319、R322 及 DDE 模体中的 E326 组成了高度保守的 YREK 模体,该模体使得非转移链的 5'-P 发生位移,为靶 DNA 的结合留下了空间<sup>[9]</sup>。

## 1.3 Tn5 转座机制

Tn5 转座需要多个步骤,主要包括转座酶二聚体的形成、供体 DNA 的切割、靶 DNA 的捕获与链转移、链转移复合物的脱离 4 个步骤(见图 2)。首先,在缓冲液中加入  $\text{Mg}^{2+}$ ,两个转座酶分子的 N 端 26~65 位氨基酸残基及活性中心的几个氨基酸残基与 Tn5 转座子末端序列特异性结合,构成两个转座酶-DNA 复合物<sup>[6,10]</sup>。紧接着,这两个复合物利用转座酶 C 端  $\alpha$  螺旋交叉,发生剪刀状二聚化,形成一个转座酶二聚体<sup>[6-7]</sup>。在形成转座酶二聚体后,转座酶启动 DNA 的切割,并且,一侧的转座酶催化另一侧末端的反应,呈现出交叉作用的现象<sup>[11]</sup>。整个催化反应大致包括 3 个步骤:首先,转座酶活化的水分子与 DNA 链反应,使 OE 末端暴露出游离的 3'-OH;其在接下来的反应中与互补链形成发夹结构,



注:A:Tn5 的天然结构:Tn5 转座子由编码 3 种抗生素( $\text{kan}^r$ 、 $\text{str}^r$ 、 $\text{bleo}^r$ )的中心序列和两侧的倒置序列(IS50L 和 IS50R)组成,每个 IS50 都位于一个外末端(OE)和一个内末端(IE)之间,IS50L 编码两个非活性蛋白,IS50R 编码转座酶(Tnp)和转座酶抑制蛋白(Inh);B:Tn5 转座子的简化结构:任意足够长的 DNA 片段在其两侧加上外末端(OE)或者其它末端(IE、ME)都构成一个转座子。

图 1 Tn5 转座子结构

Note. A. Traditional transposon Tn5. It contains two inverted versions of IS50 (IS50L and IS50R) bracketing three antibiotic resistance genes ( $\text{kan}^r$ ,  $\text{str}^r$ ,  $\text{bleo}^r$ ), and the IS50 elements are defined by the outside end (OE) and inside end (IE). IS50L encodes two inactive proteins, while IS50R encodes transposase (Tnp) and an inhibitor of transposition (Inh). B. Simplified transposon Tn5. Two outside ends (OE) or other ends (IE, ME) brackets sufficient DNA.

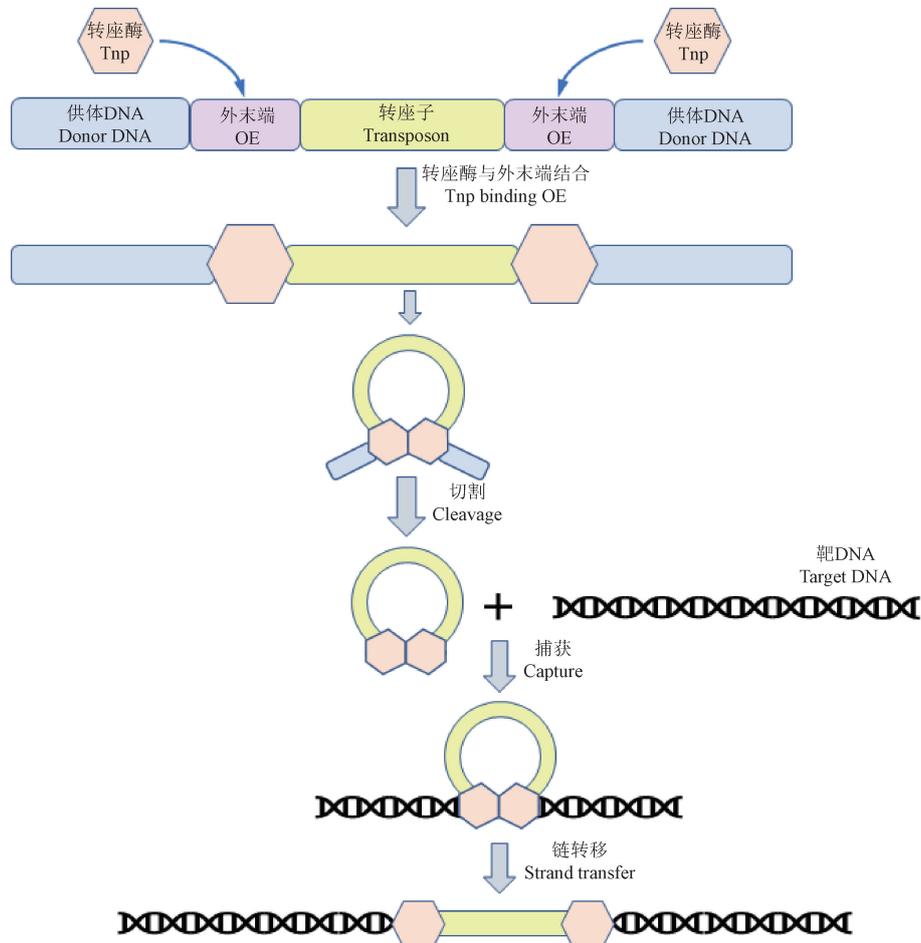
Figure 1 Structure of transposon Tn5

使转座子 DNA 从供体 DNA 骨架中脱离;最后,另一活化的水分子将该发夹结构打开,释放出 3'-OH 为链转移做准备<sup>[12]</sup>。在整个催化反应完成后,转座酶二聚体捕获靶 DNA 并进行链转移。许多研究表明, Tn5 的链转移插入位点在大范围内是随机的,但是仍然存在优选的 9 bp 插入位点<sup>[13]</sup>。并且,研究者们发现了非转移链的 5'-P 会导致 5' 末端在空间上的运动,进而使得 DNA 构象发生改变,更易于与靶 DNA 结合,并将 3'-OH 暴露在一个空腔中使其更易发生链转移<sup>[9]</sup>。当转座酶二聚体与靶 DNA 结合后,转座子 DNA 的 3'-OH 分别在 9 bp 靶位点两侧进行亲核攻击完成链转移,并在转座子两端各形成一个 9 bp 的粘性末端。据推测,在体内实现链转移后,链转移复合物可能被宿主的某些功能释放,随后两个粘性末端被补平,形成 9 bp 的正向重复序列。而在体外时, Tn5 链转移复合物十分稳定<sup>[14]</sup>。

## 2 Tn5 建库方法

自人类基因组计划以来, DNA 测序技术在生命科学领域的研究不断深入。二代测序技术的发展使大规模并行 DNA 测序成为可能<sup>[15]</sup>。二代测序技术的流程步骤主要包括样品提取、文库构建与测序及最后的数据处理与分析。其中,文库构建在很大程度上直接影响到数据分析的结果,是整个技术的核心。传统的文库构建流程通常包括 DNA 片段化(机械法或酶促法)、末端补齐、测序接头的连接和 PCR 扩增,在 RNA-seq 中还需要首先合成 cDNA 文库。尽管 barcode 接头的连接可以同时分析多个样本,但是对于开发一种样本起始量低、实验周期短、成本低的文库构建方法的需求越来越强烈<sup>[16]</sup>。

基于 Tn5 转座酶的文库构建方法大大缩减了传



注: Tn5 的转座机制: 两个 Tn5 转座酶结合 OE, 形成复合体并发生二聚化。随后, 通过亲核反应使转座子 DNA 脱离供体 DNA。最终, Tn5 二聚体捕获靶 DNA, 完成链转移。

图 2 Tn5 的转座机制

Note. Mechanism of Tn5 transposition. Two Tn5 transposases bind to the OEs and then dimerize forming the synaptic complex. Then, Nucleophilic reactions occur resulting in cleavage of transposon-donor DNA bonds. Finally, the synaptic complex captures the target DNA and strand transfer occurs.

Figure 2 Mechanism of Tn5 transposition

统方案中文库构建的步骤、时间和成本。在此方法中, Tn5 转座酶的高活性衍生物的运用可以在短时间内将转座序列插入靶 DNA 中, 直接引发双链 DNA 的片段化, 同时将接头序列加至片段的两端<sup>[17]</sup>。其流程大致为, 首先将测序接头和 OE 序列整合为包被接头, 与转座酶在体外孵育, 形成结合了 DNA 的转座酶二聚体; 随后该复合体与待测序 DNA 共孵育并将其打断, 形成可测序的 DNA 片段; 最后加入酶将复合体插入时产生的切口补平, 并通过 PCR 扩增完成文库的构建。此外, 在 PCR 扩增时还可以引入 barcode 以实现大规模并行的 DNA 测序。

但是 PCR 扩增易产生扩增偏差和错误<sup>[18]</sup>, 这在细胞异质性研究的单细胞测序中是极为不利的。因此, 一种利用 Tn5 转座子插入启动子进行线性扩增的方法 (linear amplification via transposon insertion, LIANTI) 应运而生。该方法利用 Tn5 转座子将单细胞基因组 DNA 随机打断, 同时每个片段上添加 T7 启动子, 再用 DNA 聚合酶补齐片段两端的单链 T7 启动子, 最后进行体外转录。随后, 再通过逆转录、RNA 酶消化、cDNA 二链的合成, 最终形成 LIANTI 文库进行测序<sup>[18]</sup>。并且, 若在 LIANTI 转座子和引物中添加组合 barcode, 也可以实现单细胞的高通量测序。此外, 虽然 Tn5 偏好于包含鸟苷 (G) 和胞苷 (C) 的插入序列, 导致插入间隔不够均匀, 进而导致测序时的覆盖范围不够均匀, 但是 Tn5 转座酶突变体 Tn5-059 可以有效解决这个问题<sup>[19]</sup>。同时, 多种高活性 Tn5 转座酶的高效生产进一步降低了文库构建的时间和成本<sup>[20-22]</sup>。这些优势都使 Tn5 转座酶在建库领域备受青睐。

### 3 Tn5 建库在表观测序中的应用

#### 3.1 基于 Tn5 为全基因组亚硫酸氢盐测序建库

表观遗传是指基因序列不变时, 通过一些表观修饰影响基因表达情况进而影响表型的现象。胞嘧啶 5-C 位置的酶促甲基化是一种广泛存在的表观遗传修饰, 在基因调控方面有着重要作用<sup>[23]</sup>。全基因组亚硫酸氢盐测序 (whole genome bisulfite sequencing, WGBS) 是研究胞嘧啶甲基化分辨率较高且较为全面的一种方法<sup>[24]</sup>。该方法中, 亚硫酸氢钠被用来处理基因组 DNA, 在 5-甲基胞嘧啶保持非反应性的条件下将单链 DNA 中的胞嘧啶残基转化为尿嘧啶残基, 随后用胞嘧啶甲基化处理的引物扩增并测序, 序列中剩余的所有胞嘧啶残基代表基因

组中先前甲基化的胞嘧啶<sup>[25]</sup>。随着大规模并行测序成本的下降, WGBS 也越来越普及。但是 WGBS 有一个关键的限制因素是需要较大量的 DNA 样本, 这对于许多实验是难以实现的<sup>[24, 26]</sup>。据此, Adey 等<sup>[27]</sup>开发了 Tn5Mc-seq, 该方法将 Tn5 转座系统用于 WGBS 来快速制备亚硫酸氢盐测序文库, 先将传统的 Tn5 建库中与 Tn5 结合的接头中的胞嘧啶进行甲基化修饰, 将基因组 DNA 片段化后再与亚硫酸氢盐反应。转座酶的使用使 DNA 的片段化、末端的补齐和接头的连接简化为单个反应且将 DNA 更有效地转化为测序兼容材料。因此其减少建库过程中对 DNA 样本量的损耗, 大大降低了 WGBS 所需要的基因组 DNA 起始量<sup>[27]</sup>。

#### 3.2 基于 Tn5 对蛋白质与染色体的相互作用测定

染色质和蛋白质的相互作用在表观调控中也起到非常重要的作用, 因此研究这种相互作用的众多方法相继诞生, 染色质免疫沉淀测序 (chromatin immunoprecipitation sequencing, ChIP-seq) 就是一个十分典型的例子<sup>[28]</sup>。但 ChIP-seq 需要甲醛来介导 DNA 与蛋白质的交联, 往往具有低信号、高背景且样本需求量大等缺点<sup>[28]</sup>。随后出现的 CUT&RUN 技术就是在此基础上的改进, 它的大致思路为在非交联细胞中通过连续使用特定抗体, 将蛋白 A-微球菌核酸酶复合物 (pA-MNase) 导向蛋白质结合位点, 随后通过钙离子的释放来激活核酸酶, 将所要的片段释放到上清液中, 最后进行文库的制备和测序等<sup>[29]</sup>。该方法可以将背景水平大大降低, 也可以在一定程度上降低对样本的需求量, 仅在 100~1000 个细胞中就可以产生高质量的数据, 但是必须经过 DNA 末端的补齐和接头的连接才能制备测序文库, 这会增加整个过程的时间、成本和工作量。紧接着出现的 CUT&Tag 技术是对 CUT&RUN 技术的改进, 研究人员将 MNase 替换为 Tn5 转座酶并使用  $Mg^{2+}$  在特定时期激活 pA-Tn5 融合蛋白组成的转座体, 在获取靶 DNA 的同时直接插入测序接头, 进一步简化了实验步骤, 降低了时间成本<sup>[30]</sup>。在实际操作中, CUT&RUN 和 CUT&Tag 技术都需要将细胞与磁珠结合, 由于磁珠的悬浮性较差等因素, 酶的效率往往会下降, 于是研究者们又对磁珠进行改良, 有效改善了这一问题, 提高了检测的灵敏度<sup>[31]</sup>。最近, 研究者在原有的 CUT&Tag 技术上进行创新, 开发出了 multi-CUT&Tag, 可以同时映射同一细胞中的多个染色体蛋白质, 直接检测不同蛋白质在同一细

胞中的共定位<sup>[32]</sup>。与此同时,与 CUT&Tag 技术一样基于抗体将 Tn5 靶向目的位点的技术还有 ACT-seq<sup>[33]</sup>、CoBATCH<sup>[34]</sup>和 ChIL-seq<sup>[35]</sup>等,在单细胞表面测序方面都具有深远意义和广阔的发展前景。

通常情况下,染色质可分为松弛型和紧密型两种状态,紧密型染色质往往不易于与蛋白质等调控因子结合,可及性较弱;而松弛型染色质更容易被调控因子等识别并结合,可及性较强,因此染色质可及性的研究也是基因组研究的一大重点。以转座酶为载体研究染色质可及性的技术(assay for transposase accessible chromatin with high-throughput sequencing, ATAC-seq)在这一领域有着极为重要的地位。该技术利用 Tn5 转座酶优先结合 DNA 中未受核小体保护的区域的特点,将测序接头引入染色质的松弛区域进行测序,以测定全基因组的染色质可及性<sup>[36]</sup>。相比以往染色质可及性的测定方法,ATAC-seq 中 Tn5 的使用使 DNA 的切割和接头的插入同时进行,简化了实验流程,从而大大降低了对样本量的要求。但同时 ATAC-seq 也存在一些问题,例如,由于 Tn5 转座酶在插入测序接头时方向随机,一个片段两端的接头可能相同,导致仅有一半的片段可被测序<sup>[37]</sup>。此外,当片段过大时,也难以进行体外的 PCR 扩增<sup>[37]</sup>。

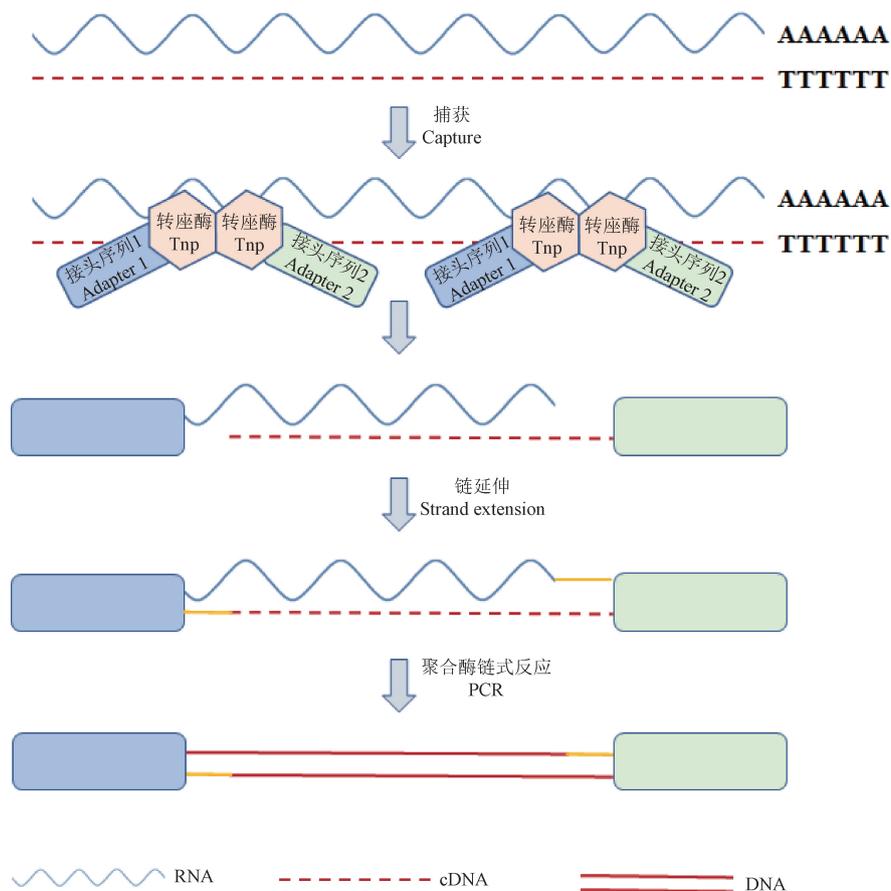
近年来,研究者们对 ATAC-seq 技术进行了多种改进。THS-ATAC-seq 技术是利用类似 LIANTI 技术的方法将 T7 启动子引入染色质可及位点并进行体外转录,随后对转录本进行建库测序。并且,为了解决 ATAC-seq 中片段两端接头相同导致测序效率降低的问题,研究者们还设计了由两套转座体复合物二聚体先后进行转座以分别加入两端测序接头的方法<sup>[37]</sup>。此外,Single Cell ATAC-seq 技术提供了单细胞染色质可及性的测定方案,使异质性细胞染色质可及性的差异得以检测<sup>[38-40]</sup>。Omni-ATAC-seq 技术在转座反应中使用 PBS 以增加信噪比,可以消除线粒体的干扰,提高染色质可及性数据的质量<sup>[40-41]</sup>。UMI-ATAC-seq 技术通过把带有标准 TruSeq 测序接头的 UMI (unique molecular identifiers)应用到 ATAC-seq 中,将 PCR 导致的重复序列与固有的重复序列区分开,能更准确地量化染色质可及性<sup>[42]</sup>。近期,新开发的 ATAC-MS 技术将 ATAC-seq 与质谱技术相结合,使开放染色质结合蛋白得到更精确的检测,这相比之前在 ATAC-seq 后通过数据分析猜测蛋白质结合位点要更加精准<sup>[43]</sup>。

## 4 Tn5 建库在长片段测序中的应用

二代测序技术虽然实现了人们对基因的进一步研究,但是它只能有效读取短片段的特点使得最初的长片段测序往往需要先长片段打断,测序后进行拼接。这种方法不仅复杂繁琐,而且给基因组中同源序列的辨别和染色体结构变异的鉴定带来困难。Peters 等<sup>[44]</sup>利用稀释法将长片段相互分离后打断,并在打断后的每个短片段上添加 barcode 和测序接头,在测序后利用 barcode 区分不同的长片段并完成拼接。但将混合的长片段稀释为单条片段操作复杂,且后续处理成本高。2017 年, stLFR (single-tube long fragment read) 的开发解决了这一困难。研究人员利用 Tn5 在被 SDS 去除之前都一直与 DNA 片段结合且不脱落这一特性<sup>[45]</sup>,将 Tn5 转座酶先与长片段 DNA 结合后,再与表面含有不同 barcode 的磁珠共同孵育<sup>[46]</sup>。由于转座酶上结合的特殊 DNA 片段可以与 barcode 侧翼链杂交并连接,不同的长片段会附着在不同的磁珠上。随后加入 SDS 可打断长片段并将磁珠表面的 barcode 转座到每一个短片段中,最后通过 PCR 加入测序接头即可进行常规测序<sup>[47]</sup>。该方法弥补了二代测序的不足,相比三代测序成本更低,准确性更高,具有广阔的应用前景。

## 5 Tn5 建库在转录组测序中的应用

在常规转录组建库中,往往需要先富集目标 RNA 再将其片段化,随后进行一链和二链 cDNA 合成、接头连接和 PCR 扩增等。尽管 Tn5 转座酶建库方法的普及简化了其中的步骤,但是其仍然无法满足对微量 RNA 样本方便、快捷地进行转录组测序及分析的需求。2020 年, Tn5 转座酶被证实可以直接作用于 DNA/RNA 杂交链,基于此,研究者们开发了一种转录组建库新方法,并命名为 TRACE-seq (transposase-assisted RNA/DNA hybrids co-tagmentation) (见图 3),该方法可以简化实验流程,降低建库成本及样本的输入量<sup>[48-49]</sup>。TRACE-seq 具有广阔的应用前景,它有望用于获取与复杂的宿主背景混合而无法体外培养的低滴度病毒的全基因组<sup>[50]</sup>。近期,TRACE-seq 也被证明具有获取综合信息的潜力,可用于表征复杂的宿主与微生物组之间的相互作用<sup>[51]</sup>。



注:基于 TRACE-seq 的建库流程,在 RNA 逆转录后,带有接头序列的 Tn5 转座酶直接结合 DNA/RNA 杂交链,完成链的延伸补齐后利用 DNA 聚合酶进行扩增完成建库。

图 3 TRACE-seq 建库方法

Note. Workflow of the library preparation strategy based on TRACE-seq. After RNA reverse transcription, the DNA/RNA hybrid is tagged by Tn5 transposase with adapters directly. DNA polymerase then amplifies the cDNA into a sequencing library after end extension.

Figure 3 Library preparation based on TRACE-seq

## 6 小结与展望

Tn5 转座子独特的转座机制使其可以将文库构建中 DNA 的片段化、末端的补齐和接头的连接 3 个步骤简化为单一且快速的反应,这大大简化了文库的构建流程,同时降低了对样本量的需求,这在疾病的研究中具有深远的意义和广阔的前景。此外,LIANTI 的应用降低了由 PCR 扩增带来的扩增偏差和错误,使得文库中 DNA 的扩增更加保守。在长片段测序中,Tn5 转座酶的应用弥补了二代测序短读长的缺点,相较于三代测序其成本更低,准确性更高,这使得一些染色体疾病的检测变得更加便捷。

除此之外,Tn5 转座酶还被应用在多个领域中。例如 BAT Hi-C 技术利用 Tn5 转座酶进行染色体 3D 构象的捕获<sup>[52]</sup>;FACT-seq 技术利用 Tn5 转座酶开

发出有效分析福尔马林固定石蜡包埋组织中组蛋白修饰的方法<sup>[53]</sup>;SorTn-seq 技术利用 Tn5 转座酶高通量筛选细菌基因表达的调节因子<sup>[54]</sup>。Tn5 转座酶在各种技术中的应用显著降低了对样本量的要求,这使许多稀缺的样本也能够满足各种测序的要求。同时,Tn5 转座酶与 barcode 等技术的结合提高了单细胞测序的可行性。目前,Tn5 转座酶在单细胞测序领域的研究还相当稀缺,对相关技术的开发和改进还有待研究。但 Tn5 转座酶所具有的单细胞测序潜力在细胞分化、肿瘤异质性的研究中备受关注。虽然目前 Tn5 转座酶的使用并未普及,但是随着技术研究的不断深入,Tn5 转座酶有望打开单细胞测序等领域的新局面。

### 参考文献 (References)

[1] Rothstein SJ, Jorgensen RA, Yin JC, et al. Genetic organization

- of Tn5 [J]. Cold Spring Harb Symp Quant Biol, 1981, 45: 99–105.
- [ 2 ] Isberg RR, Lazaar AL, Syvanen M. Regulation of Tn5 by the right-repeat proteins; control at the level of the transposition reaction [J]. Cell, 1982, 30(3): 883–892.
- [ 3 ] Lowe JB, Berg DE. A product of the Tn5 transposase gene inhibits transposition [J]. Genetics, 1983, 103(4): 605–615.
- [ 4 ] Rothstein SJ, Jorgensen RA, Postle K, et al. The inverted repeats of Tn5 are functionally different [J]. Cell, 1980, 19(3): 795–805.
- [ 5 ] Zhou M, Bhasin A, Reznikoff WS. Molecular genetic analysis of transposase-end DNA sequence recognition; cooperativity of three adjacent base-pairs in specific interaction with a mutant Tn5 transposase [J]. J Mol Biol, 1998, 276(5): 913–925.
- [ 6 ] Steiniger-White M, Rayment I, Reznikoff WS. Structure/function insights into Tn5 transposition [J]. Curr Opin Struct Biol, 2004, 14(1): 50–57.
- [ 7 ] Davies DR, Mahnke Braam L, Reznikoff WS, et al. The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-Å resolution [J]. J Biol Chem, 1999, 274(17): 11904–11913.
- [ 8 ] Rice P, Mizuuchi K. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration [J]. Cell, 1995, 82(2): 209–220.
- [ 9 ] Klenchin VA, Czyz A, Goryshin IY, et al. Phosphate coordination and movement of DNA in the Tn5 synaptic complex; role of the (R)YREK motif [J]. Nucleic Acids Res, 2008, 36(18): 5855–5862.
- [ 10 ] Zhou M, Reznikoff WS. Tn5 transposase mutants that alter DNA binding specificity [J]. J Mol Biol, 1997, 271(3): 362–373.
- [ 11 ] Naumann TA, Reznikoff WS. Trans catalysis in Tn5 transposition [J]. Proc Natl Acad Sci U S A, 2000, 97(16): 8944–8949.
- [ 12 ] Bhasin A, Goryshin IY, Reznikoff WS. Hairpin formation in Tn5 transposition [J]. J Biol Chem, 1999, 274(52): 37021–37029.
- [ 13 ] Goryshin IY, Miller JA, Kil YV, et al. Tn5/IS50 target recognition [J]. Proc Natl Acad Sci U S A, 1998, 95(18): 10716–10721.
- [ 14 ] Goryshin IY, Reznikoff WS. Tn5 *in vitro* transposition [J]. J Biol Chem, 1998, 273(13): 7367–7374.
- [ 15 ] Shendure J, Ji H. Next-generation DNA sequencing [J]. Nat Biotechnol, 2008, 26(10): 1135–1145.
- [ 16 ] Craig DW, Pearson JV, Szelinger S, et al. Identification of genetic variants using bar-coded multiplexed sequencing [J]. Nat Methods, 2008, 5(10): 887–893.
- [ 17 ] Adey A, Morrison HG, Asan, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition [J]. Genome Biol, 2010, 11(12): R119.
- [ 18 ] Chen C, Xing D, Tan L, et al. Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI) [J]. Science, 2017, 356(6334): 189–194.
- [ 19 ] Kia A, Gloeckner C, Osothprarop T, et al. Improved genome sequencing using an engineered transposase [J]. BMC Biotechnol, 2017, 17(1): 6.
- [ 20 ] Picelli S, Björklund AK, Reinius B, et al. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects [J]. Genome Res, 2014, 24(12): 2033–2040.
- [ 21 ] Hennig BP, Velten L, Racke I, et al. Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol [J]. G3 (Bethesda), 2018, 8(1): 79–89.
- [ 22 ] Xu T, Xiao M, Yu L. Method for efficient soluble expression and purification of recombinant hyperactive Tn5 transposase [J]. Protein Expr Purif, 2021, 183: 105866.
- [ 23 ] Paulsen M, Ferguson-Smith AC. DNA methylation in genomic imprinting, development, and disease [J]. J Pathol, 2001, 195(1): 97–110.
- [ 24 ] Cokus SJ, Feng S, Zhang X, et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning [J]. Nature, 2008, 452(7184): 215–219.
- [ 25 ] Clark SJ, Harrison J, Paul CL, et al. High sensitivity mapping of methylated cytosines [J]. Nucleic Acids Res, 1994, 22(15): 2990–2997.
- [ 26 ] Lister R, O’Malley RC, Tonti-Filippini J, et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis* [J]. Cell, 2008, 133(3): 523–536.
- [ 27 ] Adey A, Shendure J. Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing [J]. Genome Res, 2012, 22(6): 1139–1143.
- [ 28 ] Park PJ. ChIP-seq: advantages and challenges of a maturing technology [J]. Nat Rev Genet, 2009, 10(10): 669–680.
- [ 29 ] Skene PJ, Henikoff JG, Henikoff S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers [J]. Nat Protoc, 2018, 13(5): 1006–1019.
- [ 30 ] Kaya-Okur HS, Wu SJ, Codomo CA, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells [J]. Nat Commun, 2019, 10(1): 1930.
- [ 31 ] Fujiwara Y, Tanno Y, Sugishita H, et al. Preparation of optimized concanavalin A-conjugated Dynabeads<sup>®</sup> magnetic beads for CUT&Tag [J]. PLoS One, 2021, 16(11): e0259846.
- [ 32 ] Gopalan S, Fazio TG. Multi-CUT&Tag to simultaneously profile multiple chromatin factors [J]. STAR Protoc, 2022, 3(1): 101100.
- [ 33 ] Carter B, Ku WL, Kang JY, et al. Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq) [J]. Nat Commun, 2019, 10(1): 3747.
- [ 34 ] Wang Q, Xiong H, Ai S, et al. CoBATCH for high-throughput single-cell epigenomic profiling [J]. Mol Cell, 2019, 76(1): 206–216.
- [ 35 ] Harada A, Maehara K, Handa T, et al. A chromatin integration labelling method enables epigenomic profiling with lower input

- [J]. *Nat Cell Biol*, 2019, 21(2): 287–296.
- [36] Buenrostro JD, Wu B, Chang HY, et al. ATAC-seq: A method for assaying chromatin accessibility genome-wide [J]. *Curr Protoc Mol Biol*, 2015, 109: 21.29.1–21.29.9.
- [37] Sos BC, Fung HL, Gao DR, et al. Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay [J]. *Genome Biol*, 2016, 17: 20.
- [38] Chen X, Miragaia RJ, Natarajan KN, et al. A rapid and robust method for single cell chromatin accessibility profiling [J]. *Nat Commun*, 2018, 9(1): 5345.
- [39] Xu W, Wen Y, Liang Y, et al. A plate-based single-cell ATAC-seq workflow for fast and robust profiling of chromatin accessibility [J]. *Nat Protoc*, 2021, 16(8): 4084–4107.
- [40] Sun Y, Miao N, Sun T. Detect accessible chromatin using ATAC-sequencing, from principle to applications [J]. *Hereditas*, 2019, 156: 29.
- [41] Corces MR, Trevino AE, Hamilton EG, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues [J]. *Nat Methods*, 2017, 14(10): 959–962.
- [42] Zhu T, Liao K, Zhou R, et al. ATAC-seq with unique molecular identifiers improves quantification and footprinting [J]. *Commun Biol*, 2020, 3(1): 675.
- [43] Zhang H, Qin Z, Yue X, et al. Proteome-wide profiling of transcriptional machinery on accessible chromatin with biotinylated transposons [J]. *Sci Adv*, 2021, 7(43): eabh1022.
- [44] Peters BA, Kermani BG, Sparks AB, et al. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells [J]. *Nature*, 2012, 487(7406): 190–195.
- [45] Amini S, Pushkarev D, Christiansen L, et al. Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing [J]. *Nat Genet*, 2014, 46(12): 1343–1349.
- [46] Zhang F, Christiansen L, Thomas J, et al. Haplotype phasing of whole human genomes using bead-based barcode partitioning in a single tube [J]. *Nat Biotechnol*, 2017, 35(9): 852–857.
- [47] Wang O, Chin R, Cheng X, et al. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and *de novo* assembly [J]. *Genome Res*, 2019, 29(5): 798–808.
- [48] Di L, Fu Y, Sun Y, et al. RNA sequencing by direct tagmentation of RNA/DNA hybrids [J]. *Proc Natl Acad Sci U S A*, 2020, 117(6): 2886–2893.
- [49] Lu B, Dong L, Yi D, et al. Transposase-assisted tagmentation of RNA/DNA hybrid duplexes [J]. *Elife*, 2020, 9: e54919.
- [50] Zhang Z, Liu D, Wang D, et al. Library preparation based on transposase assisted RNA/DNA hybrid co-tagmentation for next-generation sequencing of human noroviruses [J]. *Viruses*, 2021, 13(1): 65.
- [51] Lu B, Yan Y, Dong L, et al. Integrated characterization of SARS-CoV-2 genome, microbiome, antibiotic resistance and host response from single throat swabs [J]. *Cell Discov*, 2021, 7(1): 19.
- [52] Huang J, Jiang Y, Zheng H, et al. BAT Hi-C maps global chromatin interactions in an efficient and economical way [J]. *Methods*, 2020, 170: 38–47.
- [53] Zhao L, Xing P, Polavarapu VK, et al. FACT-seq: profiling histone modifications in formalin-fixed paraffin-embedded samples with low cell numbers [J]. *Nucleic Acids Res*, 2021, 49(21): e125.
- [54] Smith LM, Jackson SA, Gardner PP, et al. SorTn-seq: a high-throughput functional genomics approach to discovering regulators of bacterial gene expression [J]. *Nat Protoc*, 2021, 16(9): 4382–4418.

[收稿日期] 2022-05-12