

· 技术方法 ·

Tn5转座酶融合蛋白在CUT&Tag实验中的优化及评价

刘晟宇, 刘晓斌, 朱家富, 苏京, 董志诚, 刘敏*

广州大学生命科学学院, 广东省植物适应性与分子设计重点实验室, 广州市作物基因编辑应用重点实验室,
分子遗传与进化创新研究中心, 广州 510006

摘要 Tn5是一种细菌转座子。经改造的Tn5能够高效地切割DNA, 同时连接上特定的接头序列, 因而广泛应用于高通量二代测序文库构建中。CUT&Tag (Cleavage Under Target & Tagmentation)是一种改进的研究蛋白质与DNA互作的技术, 具有重复性好、信噪比高及操作简便等优点。该技术采用pA (Protein A)或pG (Protein G)与Tn5形成融合蛋白, 定位于特定抗体(用于识别目标蛋白), 利用Tn5的特性, 在目标位点附近打断DNA的同时引入测序接头, 随后提取DNA, 再进行PCR扩增即可获得测序文库。但不同类型的抗体与pA或pG的亲合力不同, 因此限制了部分抗体在CUT&Tag技术中的应用。为克服这一局限, 该文构建了pG与Tn5的融合蛋白表达载体, 通过原核表达及亲和纯化的方式获得pG-Tn5重组蛋白; 并以RNA聚合酶II (Pol II)特异性抗体Pol II Ser5P (小鼠IgG1型抗体和兔IgG型抗体)为例, 在模式植物拟南芥(*Arabidopsis thaliana*)中评估pA-Tn5与pG-Tn5在不同类型抗体的CUT&Tag测序文库构建中的效果。结果表明, IgG1型抗体与pG-Tn5的亲合力更高, 构建的文库质量更好, 而IgG型抗体与2种酶的亲合力相当; 同时, 较低起始量的植物材料也能获得较好的效果, 证明了CUT&Tag的应用优势。该研究优化了CUT&Tag技术, 可为后续CUT&Tag实验中针对不同抗体时Tn5融合蛋白的选择提供参考。

关键词 ChIP, CUT&Tag, RNA聚合酶II, Tn5

刘晟宇, 刘晓斌, 朱家富, 苏京, 董志诚, 刘敏 (2023). Tn5转座酶融合蛋白在CUT&Tag实验中的优化及评价. 植物学报 58, 602–611.

高通量测序技术的发展为生物学研究带来巨大变革, 极大地提升了我们获取基因组结构、基因表达调控和遗传变异等信息的能力。测序前的文库构建工作是获得高质量数据的关键因素之一。在二代测序中, 常规文库构建主要包括DNA或RNA片段化、末端修复、测序接头连接、片段大小筛选和PCR扩增等过程, 步骤较为烦琐。基于Tn5转座酶复合物建立的文库构建方法则具有简便且易操作的优点。Tn5是一种细菌转座子, 包括2个几乎相同且倒置的元件IS50L和IS50R, 其中IS50R编码有活性的转座酶, Tn5通过“剪切-粘贴(cut-and-paste)”的方式催化转座元件的易位, 在切割DNA的同时将转座元件插入靶DNA中 (Lovell et al., 2002)。野生型Tn5的转座效率较低, 经过改造的Tn5则可以高效切割DNA, 同时为DNA两端加上接头, 因此广泛应用于高通量测序文库构建

(Picelli et al., 2014), 如转录组和ATAC-seq (鉴定染色质开放区域) (Buenrostro et al., 2013; Di et al., 2020)。

染色质免疫共沉淀(chromatin immunoprecipitation, ChIP)技术是检测蛋白质与DNA结合的实验技术(王泓力和焦雨铃, 2020)。基于此技术开发的ChIP-seq可以在全基因组范围内检测蛋白质与DNA的相互作用(Schmidt et al., 2009), 但该方法存在背景信号偏高, 以及由于甲醛交联和染色质在免疫沉淀过程中的溶解而产生假阳性的问题(Kharchenko et al., 2008; Kaya-Okur et al., 2019)。Steven Henikoff实验室先后于2017年和2019年发表了CUT&RUN和CUT&Tag两种改进的免疫沉淀技术, 均展现出良好的可重复性以及信噪比(Skene and Henikoff, 2017; Kaya-Okur et al., 2019), 随后他们又对技术进行全

收稿日期: 2022-04-29; 接受日期: 2022-09-07

基金项目: 国家自然科学基金(No.31871289, No.31900463)和广州大学研究生创新能力培养资助计划(No.2021GDJC-M23)

* 通讯作者。E-mail: minl@gzhu.edu.cn

面优化(Thakur and Henikoff, 2018; Brahma and Henikoff, 2019; Meers et al., 2019)。与传统的ChIP-seq相比, CUT&RUN和CUT&Tag具有显著优势。其一是ChIP-seq实验中通过超声打断的方式使DNA片段化, 这种片段化的过程受交联强度、超声设备和条件的限制, 影响了ChIP-seq的重复性和信噪比(Schmidt et al., 2009), 而CUT&RUN和CUT&Tag通过酶切的方式使DNA片段化, 极大地提高了可重复性; 其二是ChIP-seq的实验周期为3–5天, 而CUT&RUN和CUT&Tag最快仅需1天即可完成文库构建, 实验周期极大缩短(Skene and Henikoff, 2017; Kaya-Okur et al., 2019)。此外, CUT&Tag比CUT&RUN更具有优势, 即建库过程中不需要人工加测序接头, 加接头的过程在酶切反应中就已经完成。

CUT&Tag的高信噪比和便捷性在于利用Protein A和Tn5融合表达的pA-Tn5, 直接处理细胞核中的染色体, 将目的蛋白免疫识别、结合DNA纯化及文库构建等简化为一个步骤。Protein A是一种来自金黄色葡萄球菌细胞壁蛋白, 这是一种特异性配体, 可与许多物种IgG的Fc区结合(Lindmark et al., 1983)。因此, 用Protein A和IgG抗体的亲和性使得与其融合表达的Tn5也具备了更高的靶向性, 同时Tn5的转座酶活性可以直接给DNA片段两端添加测序接头。但不同类型的免疫球蛋白与Protein A和Protein G的亲和力具有较大差异(Dancette et al., 1999)。例如, 小鼠(*Mus musculus*)源的IgG1型抗体与Protein A的亲和力较弱, 与Protein G的亲和力较强; 而兔源的IgG型抗体与Protein A和Protein G的亲和力则相当。因此在CUT&Tag实验中, 对于不同的抗体类型应当采用不同融合表达类型的Tn5进行免疫沉淀。

近年来, CUT&Tag已经广泛应用于植物尤其是组蛋白修饰研究中。例如, 利用CUT&Tag在棉花(*Gossypium hirsutum*)中对组蛋白修饰H3K4me3等进行系统分析(Tao et al., 2020)。组蛋白修饰作为一种表观遗传修饰, 在植物生长发育和应对环境胁迫中均发挥重要作用(杨同文和李成伟, 2014; 杜康兮等, 2018; 陈威等, 2019)。因此, 在植物中优化CUT&Tag技术具有重要应用价值。本研究以模式植物拟南芥(*Arabidopsis thaliana*)为实验材料, 利用RNA聚合酶II (Pol II)的Ser5P抗体(特异性识别CTD七肽重复序列中5位丝氨酸磷酸化的RNA聚合酶II,

小鼠IgG1型和兔IgG型)进行CUT&Tag实验, 结果呈现了抗体类型与pA-Tn5和pG-Tn5的亲和力差异, 以及对构建DNA文库效果的影响, 研究结果可为后续CUT&Tag实验中的抗体选择提供参考。

1 材料与方法

1.1 材料

载体: 3× Flag-pA-Tn5-FI和pAG/MNase (Skene and Henikoff, 2017; Kaya-Okur et al., 2019)。

菌株: 大肠杆菌(*Escherichia coli*)菌株DH5α, C3013。

植物材料: 拟南芥(*Arabidopsis thaliana* L.)哥伦比亚生态型(Col-0)。拟南芥种子先用70%乙醇表面消毒30秒, 然后加入次氯酸钠溶液(稀释8倍)消毒10分钟, 无菌水冲洗4–5次后均匀播种于1/2MS培养基上。在4°C放置2天后移至光照培养箱中培养, 培养温度为22°C, 光周期为16小时光照/8小时黑暗, 光照强度为150 μmol·m⁻²·s⁻¹。

接头引物由上海生工合成(HPLC纯化)。Bottom: 5'-CTGTCTCTTATACACATCT-3', 5'磷酸修饰; 10-30: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'; 1031: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'。

PCR扩增引物由广州天一辉远合成(HPLC纯化)。pAG/MNaseRP-F': 5'-AGATGACCATGATTACGCCAACAACTTATAAATTAGTCATCAACGGG-3'; pAG/MNaseRP-R': 5'-CCACCGAATTCTTTATCGTCATCTACGGCTGGCGTCAAC-3'; pA/Tn5RP-F': 5'-TGGCGTAATCATGGTCATCT-3'; pA/Tn5RP-R': 5'-G-ATGACGATAAAGAATTCGGTGG-3'。

抗体: Pol II Ser5P (小鼠IgG1型)购自MBL Life science (Cat No.MABI0603); Pol II Ser5P (兔IgG型)购自Abcam (Cat No.ab5131)。

所有缓冲液及试剂配方见表1。

1.2 pG/Tn5表达载体的构建

以3× Flag-pA-Tn5-FI作为模板, 通过PCR扩增含有Tn5的线性化载体(引物: pA/Tn5RP-F'和pA/Tn5RP-R'); 以pAG/MNase作为模板, 通过PCR扩增Protein G片段(引物: pAG/MNaseRP-F'和pAG/MNaseRP-R'); 用重组酶进行同源重组获得pG/Tn5

表1 缓冲液

Table 1 Buffer used in this study

Buffer	Component	Concentration	Buffer	Component	Concentration
HEGX buffer	NaCl	0.8 mol·L ⁻¹	HBC buffer	Tris-HCl, pH7.5	20 mmol·L ⁻¹
	Glycerol	10%		Sucrose	352 mmol·L ⁻¹
	TritonX-100	0.2%		MgCl ₂	8 mmol·L ⁻¹
	HEPES-KOH, pH7.2	20 mmol·L ⁻¹		TritonX-100	0.08%
	EDTA	1 mmol·L ⁻¹		β-mercaptoethanol	8 mmol·L ⁻¹
Storage buffer	NaCl	0.8 mol·L ⁻¹	10× binding buffer	Glycerol	20%
	Glycerol	10%		HEPES-KOH, pH8.0	20 mmol·L ⁻¹
	Tris-HCl, pH7.5	50 mmol·L ⁻¹		KCl	100 mmol·L ⁻¹
	EDTA	0.2 mmol·L ⁻¹		CaCl ₂	10 mmol·L ⁻¹
	DTT	2 mmol·L ⁻¹		MnCl ₂	10 mmol·L ⁻¹
Annealing buffer	Tris-HCl, pH7.0	5 mmol·L ⁻¹	Blocking buffer	Spermidine	5 mmol·L ⁻¹
	EDTA	0.5 mmol·L ⁻¹		HEPES-KOH, pH7.5	20 mmol·L ⁻¹
	NaCl	50 mmol·L ⁻¹		NaCl	150 mmol·L ⁻¹
5× Tagmentation buffer	TAPS-NaOH, pH8.5	50 mmol·L ⁻¹		Spermidine	0.5 mmol·L ⁻¹
	MgCl ₂	25 mmol·L ⁻¹		BSA	0.1%
Lysis buffer	HEPES-KOH, pH7.2	50 mmol·L ⁻¹	Wash buffer	EDTA	2 mmol·L ⁻¹
	NaCl	150 mmol·L ⁻¹		HEPES-KOH, pH7.5	20 mmol·L ⁻¹
	EDTA	1 mmol·L ⁻¹		NaCl	150 mmol·L ⁻¹
	TritonX-100	1%		Spermidine	0.5 mmol·L ⁻¹
	Glycerol	10%		BSA	0.1%
HBB buffer	β-mercaptoethanol	5 mmol·L ⁻¹	TE buffer	Pepstatin A	1 μg·mL ⁻¹
	Pepstatin A	1 μg·mL ⁻¹		Aprotinin	1 μg·mL ⁻¹
	Aprotinin	1 μg·mL ⁻¹		PMSF	1 mmol·L ⁻¹
	PMSF	1 mmol·L ⁻¹		Tris-HCl, pH8.0	1 mmol·L ⁻¹
	Tris-HCl, pH7.5	25 mmol·L ⁻¹		EDTA	0.1 mmol·L ⁻¹
	Sucrose	0.44 mol·L ⁻¹	Dissolution buffer	NaCl	300 mmol·L ⁻¹
	MgCl ₂	10 mmol·L ⁻¹		EDTA	1 mmol·L ⁻¹
	TritonX-100	0.1%			
	β-mercaptoethanol	10 mmol·L ⁻¹			

EDTA: 乙二胺四乙酸; DTT: 二硫苏糖醇; PMSF: 苯甲基磺酰氟; BSA: 牛血清白蛋白; TAPS-NaOH: 用氢氧化钠调节pH的三羟甲基甘氨酸; HEPES-KOH: 用氢氧化钾调节pH的4-羟乙基哌嗪乙磺酸
EDTA: Ethylenediaminetetraacetic acid; DTT: 1,4-dithiothreitol; PMSF: Phenylmethanesulfonyl fluoride; BSA: Bovine serum albumin; TAPS-NaOH: Trimethylhydroxymethyl amino propane sulfonic acid to regulate pH with sodium hydroxide; HEPES-KOH: 4-hydroxyethyl piperazine ethanesulfonic acid to regulate pH with potassium hydroxide

表达载体，转入大肠杆菌菌株DH5α中扩增并通过测序鉴定。提取质粒并转入菌株C3013中，鉴定后于-80℃冰箱保存备用。

1.3 Tn5、pA-Tn5和pG-Tn5的诱导表达及亲和纯化

根据已发表文献中的方法(Picelli et al., 2014)进行一定调整。取过夜培养的菌液至新的LB培养液中，加入抗生素，18℃培养10–12小时(OD值约为0.6)后加入IPTG诱导表达。随后利用HEGX buffer重悬菌液，超

声破碎细胞后用几丁质凝胶进行亲和纯化，利用intein切除几丁质结合结构域，分别获得Tn5、pA-Tn5和pG-Tn5，用储存缓冲液透析，于-80℃冰箱中保存备用。

1.4 pA-Tn5和pG-Tn5的酶活测定

参考已发表文献中的方法(Kaya-Okur et al., 2019)，将1030、1031和Bottom Oligo用退火缓冲液溶解，配成浓度为200 μmol·L⁻¹的工作液。将200 μmol·L⁻¹的1030、1031 oligo分别与Bottom oligo等体积混合，得

到 1030/Bottom 和 1031/Bottom 双链接头, 将 1030/Bottom 和 1031/Bottom 等体积混合, 得到 $100 \mu\text{mol}\cdot\text{L}^{-1}$ 混合接头。取 pA-Tn5 和 pG-Tn5 各 $10 \mu\text{L}$, 分别加入 $0.78 \mu\text{L}$ $100 \mu\text{mol}\cdot\text{L}^{-1}$ 混合接头, 混匀后室温静置 1 小时, 组装得到 pA-Tn5/接头复合物和 pG-Tn5/接头复合物。取 $3 \mu\text{L}$ pA-Tn5/接头复合物或 pG-Tn5/接头复合物、 $4 \mu\text{L}$ $5\times$ Tagmentation buffer 及 50 ng 拟南芥基因组 DNA (底物), 用 dH_2O 补足至 $20 \mu\text{L}$, 37°C 反应 30 分钟。加 $2 \mu\text{L}$ 0.2% SDS 终止反应, 得到片段化 DNA。采用琼脂糖凝胶电泳观察基因组 DNA 被酶切的情况。

1.5 pA-Tn5和pG-Tn5的CUT&Tag

参考已发表文献中的方法(Kaya-Okur et al., 2019), 收集在 1/2MS 培养基中培养 2 周的拟南芥幼苗地上部分, 用液氮研磨。用 30 mL Lysis buffer 重悬 2 mL 样品, 混匀, 冰上静置 10 分钟。用 1 层 100 目和 1 层 200 目滤网过滤悬液, 过滤液于 4°C 、 $3\ 200\times g$ 离心 20 分钟。去除上清, 用 1 mL HBB buffer 洗涤细胞核沉淀, 然后 4°C 、 $4\ 000\times g$ 离心 5 分钟。去除上清, 用 1 mL HBC 缓冲液悬浮沉淀, 分装成每管 $50 \mu\text{L}$, 然后 4°C 、 $6\ 500\times g$ 、离心 5 分钟。去除上清, 加 $100 \mu\text{L}$ 结合缓冲液重悬沉淀。取 $10 \mu\text{L}$ 磁珠 (BioMag Plus concanavalin A-coated beads, Cat No.86057-10), 用 $5\times$ Binding buffer 洗磁珠, 快速混匀, 置于磁力架上静置 20 秒, 用 $5\times$ Binding buffer 洗涤 2 次, 去除上清。用 $10 \mu\text{L}$ Binding buffer 悬浮磁珠, 加入上述细胞核(重悬于 $100 \mu\text{L}$ Binding buffer)中, 摇晃 10 分钟使磁珠与细胞核充分结合。置于磁力架上静置 1 分钟, 去除上清, 每管加 1 mL Blocking buffer, 室温摇晃 5 分钟, 去除上清。加入配制好的 $200 \mu\text{L}$ Antibody buffer (1% 对应抗体的 Wash buffer), 4°C 孵育 2 小时。去除上清, 用 1 mL Wash buffer 洗涤沉淀, 去除上清, 重复 2 次。加入 $200 \mu\text{L}$ Wash buffer 和 $10 \mu\text{L}$ pA-Tn5/接头复合物或 pG-Tn5/接头复合物, 保证磁珠与液体在充分混匀的情况下室温旋转孵育或者期间手动轻轻震荡混匀 1 小时, 磁力架上静置 1 分钟后吸去上清。再用 1 mL Wash buffer 洗涤, 去除上清, 重复洗涤 5 次。加入 $100 \mu\text{L}$ Tagmentation buffer, 37°C 摇床孵育 1 小时。加入 $4.5 \mu\text{L}$ $0.5 \text{ mol}\cdot\text{L}^{-1}$ EDTA、 $5.5 \mu\text{L}$ 10% SDS 和 $1 \mu\text{L}$ $20 \text{ mg}\cdot\text{mL}^{-1}$ 蛋白酶 K, 55°C 反应 30 分钟; 然后 70°C 孵育 20 分钟, 使蛋白酶 K 失活。每个文库加入

$300 \mu\text{L}$ CTAB, 65°C 反应 15 分钟, 冰上静置 1 分钟。加入 $500 \mu\text{L}$ 酚氯仿异戊醇 PCI ($24:25:1$, v/v/v) 混合液提取 DNA, 手动震荡 2 分钟, 室温静置 1 分钟, 然后 4°C 、 $12\ 000\times g$ 离心 15 分钟。取上清, 加 $400 \mu\text{L}$ 氯仿, 手动震荡 30 秒, 4°C 、 $12\ 000\times g$ 离心 5 分钟。取上清, 加十分之一体积的 $3 \text{ mol}\cdot\text{L}^{-1}$ NaAc、2.5 倍体积的无水乙醇和 $1.5 \mu\text{L}$ Glycoblue, 混匀。 -80°C 孵育至少 2 小时, 然后 4°C 、 $12\ 000\times g$ 离心 15 分钟, 去除上清。每个文库加 1 mL 冰冻的 75% 乙醇, 混匀后 4°C 、 $12\ 000\times g$ 离心 5 分钟, 去除上清, 重复 1 次。吸干液体后风干 5 分钟, 加入 $10\text{--}40 \mu\text{L}$ TE buffer 溶解 DNA。PCR 反应体系: $50 \mu\text{L}$, 反应程序: 72°C 3 分钟, 用于补平缺口, 95°C 30 秒, 随后 95°C 10 秒, 55°C 30 秒, 72°C 40 秒, 14 个循环, 最后 72°C 5 分钟。PCR 产物用 6% 非变性聚丙烯酰胺胶进行电泳, 110 V , 1 小时, 回收大小为 $150\text{--}400 \text{ bp}$ 的胶块。将胶块磨碎, 加入 $300 \mu\text{L}$ Dissolution buffer, 室温摇床过夜。室温、 $7\ 000\times g$ 离心 5 分钟, 缓慢吸取上清液。在上清液中加入 NaAc、无水乙醇和 Glycoblue 沉淀文库, 最后用 $15 \mu\text{L}$ DEPC H_2O 溶解 DNA。用 Qubit QuantIT (Invitrogen) 测定文库浓度, 由上海瀚宇生物公司在 illumina NovaSeq 平台完成二代测序。

1.6 生物信息学分析

首先, 通过 FastQC 软件 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) 查看测序质量, 并用 cutadapt 软件 (Martin, 2011) 去除测序 reads 中的接头序列, 过滤低质量数据。随后, 将过滤后得到的干净 reads 采用软件 STAR (Dobin et al., 2013) 比对至拟南芥基因组 TAIR10, 仅保留唯一比对的 reads。然后, 用 RseqQC 软件 (Wang et al., 2012) 中的 read_distribution.py 工具统计测序 reads 的分布特征, 利用 deepTools (Ramírez et al., 2016) 画出测序 reads 在基因上的平均分布图。最后, 用基因组浏览器 IGV (Thorvaldsdóttir et al., 2013) 查看基因上的 reads 分布特征。

2 结果与讨论

2.1 pG/Tn5表达载体的构建及pA-Tn5和pG-Tn5的诱导表达与亲和纯化

我们采用传统 PCR 扩增方式获得了含有 Tn5 的载体

和pG片段,以3× Flag-pA-Tn5-FI作为模板扩增得到Tn5转座酶基因片段,以pAG/MNase作为模板扩增得到pG DNA片段,随后用同源重组的方式获得表达pG-Tn5重组蛋白的pG/Tn5载体(图1A),通过一代测序的方式验证重组反应是否成功,以及融合蛋白序列是否正确,随后将构建成功的质粒导入表达宿主菌C3013。在宿主菌中诱导表达pG-Tn5和pA-Tn5融合蛋白,利用pG-Tn5和pA-Tn5蛋白上含有的几丁质结合结构域进行亲和纯化,得到高纯度的pA-Tn5(图1B)和pG-Tn5(图1C)。

2.2 pA-Tn5和pG-Tn5具有与Tn5相近的酶活力

我们利用相同浓度的pA-Tn5和pG-Tn5在37℃条件下对大豆(*Glycine max*)威廉82型(W82)基因组DNA

进行切割,并观察切割效果,同时比较Tn5、pA-Tn5以及pG-Tn5的酶活力,发现pA-Tn5和pG-Tn5均具有可观的酶活力,并且与Tn5对照酶活相当(图2),表明获得的重组蛋白pA-Tn5和pG-Tn5能够用于后续实验。

2.3 抗体与酶的亲和力影响CUT&Tag的效果

利用获得的重组蛋白pA-Tn5和pG-Tn5进行CUT&Tag实验,按照酶的种类、抗体种类、酶的用量以及植物材料用量分别设置12个实验组(表2)。其中,实验组1–6针对小鼠IgG1型抗体评估pA-Tn5与pG-Tn5在进行CUT&Tag实验中的效果;实验组7–12针对兔IgG型抗体评估pA-Tn5和pG-Tn5在进行CUT&Tag实验中的效果。实验组1–6和7–12分别进行,每次都设

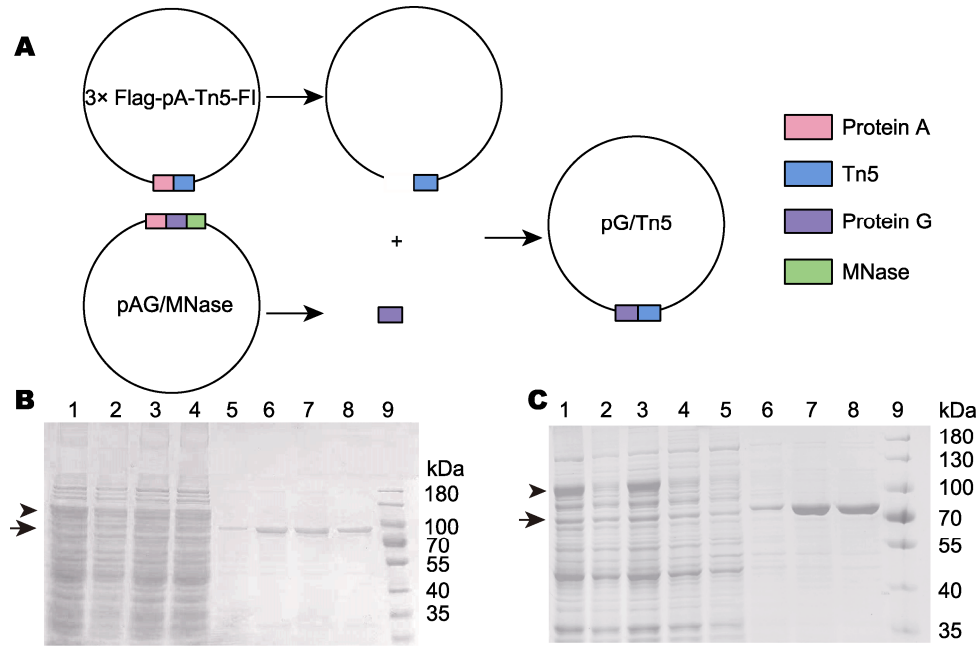


图1 pG/Tn5表达载体的构建及pA-Tn5和pG-Tn5的制备

(A) 利用PCR及同源重组的方式构建pG/Tn5; (B) pA-Tn5利用几丁质亲和和标签的亲和纯化(泳道1–4为纯化前的蛋白混合物, ►表示纯化前带有几丁质结合结构域和intein的pA-Tn5蛋白;泳道5–8为洗脱后纯度较高的pA-Tn5, ➔表示纯化后的蛋白;泳道9: 蛋白分子量标准); (C) pG-Tn5利用几丁质亲和和标签的亲和纯化(泳道1–5为纯化前的蛋白混合物, ►表示纯化前的蛋白;泳道6–8为洗脱后纯度较高的pG-Tn5, ➔表示纯化后的蛋白;泳道9: 蛋白分子量标准)

Figure 1 Construction of the pG/Tn5 expression vector and preparation of pA-Tn5 and pG-Tn5

(A) pG/Tn5 was constructed by PCR and homologous recombination; (B) pA-Tn5 was affinity purified by the chitin-binding domain (lanes 1–4 are the protein mixture before purification, ► represents the protein before purification; lanes 5–8 are pA-Tn5 with high purity after elution, ➔ represents the purified protein; lane 9: Protein molecular weight marker); (C) pG-Tn5 is affinity purified by chitin-binding domain (lanes 1–5 are the protein mixture before purification, ► represents the protein before purification; lanes 6–8 are pG-Tn5 with high purity after elution, ➔ represents the purified protein; lane 9: Protein molecular weight marker)

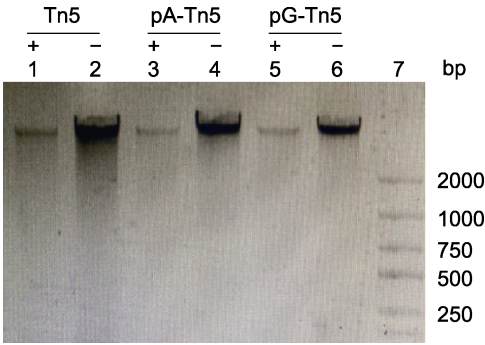


图2 pA-Tn5和pG-Tn5的活力测定
泳道1: 大豆基因组DNA+Tn5; 泳道3: 大豆基因组DNA+pA-Tn5; 泳道5: 大豆基因组DNA+pG-Tn5; 泳道2、4和6: 大豆基因组DNA; 泳道7: Marker

Figure 2 Evaluation of the activities of pA-Tn5 and pG-Tn5
Lane 1: Soybean genomic DNA+Tn5; Lane 3: Soybean genomic DNA+pA-Tn5; Lane 5: Soybean genomic DNA+pG-Tn5; Lanes 2, 4 and 6: Soybean genomic DNA without enzyme; Lane 7: Marker

表2 实验组设计

Table 2 Experimental group design

Group	Enzyme type	Antibody	Dosage of Tn5 (μg)	Dosage of plant material (μL)
1	pA-Tn5	Ser5P Mouse IgG1	0.2	150
2	pA-Tn5	Ser5P Mouse IgG1	1	150
3	pG-Tn5	Ser5P Mouse IgG1	0.2	150
4	pG-Tn5	Ser5P Mouse IgG1	1	150
5	pG-Tn5	IgG	0.2	150
6	pG-Tn5	IgG	1	150
7	pA-Tn5	Ser5P Rabbit IgG	0.5	30
8	pA-Tn5	Ser5P Rabbit IgG	1	150
9	pG-Tn5	Ser5P Rabbit IgG	0.5	30
10	pG-Tn5	Ser5P Rabbit IgG	1	150
11	pA-Tn5	IgG	1	150
12	pG-Tn5	IgG	1	150

置IgG为对照。我们对测序数据进行分析, 由于实验组1的测序量太低将其舍去, 剩余组测序数据中去除低质量数据和测序接头后得到的有效数据量分别为4、0.6、7、10、8.5、0.9、13.6、1、11、0.1和25 M。我们对测序数据的插入片段长度进行分析, 判断Tn5酶的酶切过程是否正常(以实验组4、6、8和10为例, 图3)。结果表明, 插入片段长度以10 bp为1个周期逐级递增, 符合Tn5转座酶识别并切割DNA双螺旋结构

的特点。

接下来我们使用STAR软件将测序数据与拟南芥基因组TAIR10版本进行比对, 比对后的数据用deepTools作图, 得到CUT&Tag测序reads在基因上的平均分布(图4)。结果显示, 在鼠IgG1型抗体中, 使用pG-Tn5的实验组中能够检测到明显的Pol II结合信号, 主要分布于转录起始和终止位置; 而pA-Tn5实验组与对照组IgG的reads分布较为相似(图4A)。结果表明, 相比pA-Tn5, pG-Tn5与IgG1型抗体的亲和性更强。在兔IgG型抗体中, pA-Tn5与pG-Tn5均表现出较好的亲和力, 相比IgG都能检测到Pol II在基因上分布的信号, 且较低起始量材料也能获得较好的效果(图4B)。这表明pA-Tn5和pG-Tn5与IgG型抗体的亲和力相当。同时发现, 实验组5、6以及11、12作为阴性对照却在基因上有一定信号, 推测可能是由于基因区域的染色质状态较为开放, 因此更容易被Tn5转座酶切割。

通过绘制CUT&Tag数据在基因组上不同区域的分布比例, 发现各组数据均呈现相同的分布趋势, 即大部分信号落在外显子编码区上(图5)。

随后, 我们利用基因组浏览器IGV观察并展示了不同实验组CUT&Tag数据在个别基因上的分布, 选取3个基因区段: AT4G14880–AT4G14900区段(图6A)、磷响应蛋白基因EXO (AT4G08950)附近区段(图6B)以及蓝光受体基因CRY1 (AT4G08920)附近区段(图6C)。我们发现其结果与CUT&Tag测序数据在基因上的平均分布结果类似, 在鼠IgG1抗体对比实验中, 其中利用pG-Tn5 (酶量1 μg)的实验组与对照组相比具有更高的信号。但0.2 μg酶量的pG-Tn5中, reads仅分布在少数位置, 说明酶的用量也会影响Tn5的作用效果, 酶量较少可能导致得到的DNA序列复杂度较低。在兔IgG型抗体对比实验中, 相比对照组IgG (Group 11–12), 不同类型酶的CUT&Tag (Group 7–10)中均检测到Pol II在基因上较强的分布信号。

2.4 讨论

随着测序技术的发展, 二代测序DNA和RNA文库构建成为分子生物学的常规实验, 在追求实验过程便捷性的同时, 文库的质量和成本逐渐成为研究者面临的

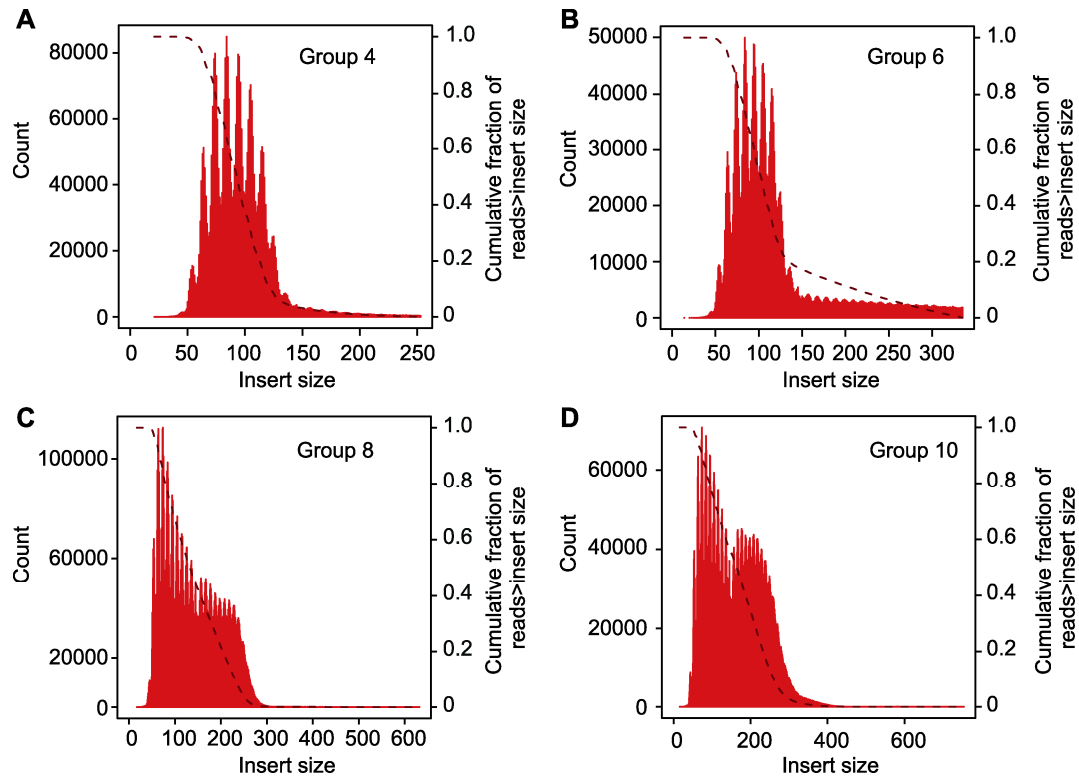


图3 插入片段长度分布
(A) 实验组4插入片段长度分布; (B) 实验组6插入片段长度分布; (C) 实验组8插入片段长度分布; (D) 实验组10插入片段长度分布。Group 4、6、8和10同表2。

Figure 3 Library insert size
(A) Library insert size in experimental group 4; (B) Library insert size in experimental group 6; (C) Library insert size in experimental group 8; (D) Library insert size in experimental group 10. Group 4, 6, 8 and 10 are the same as shown in Table 2.

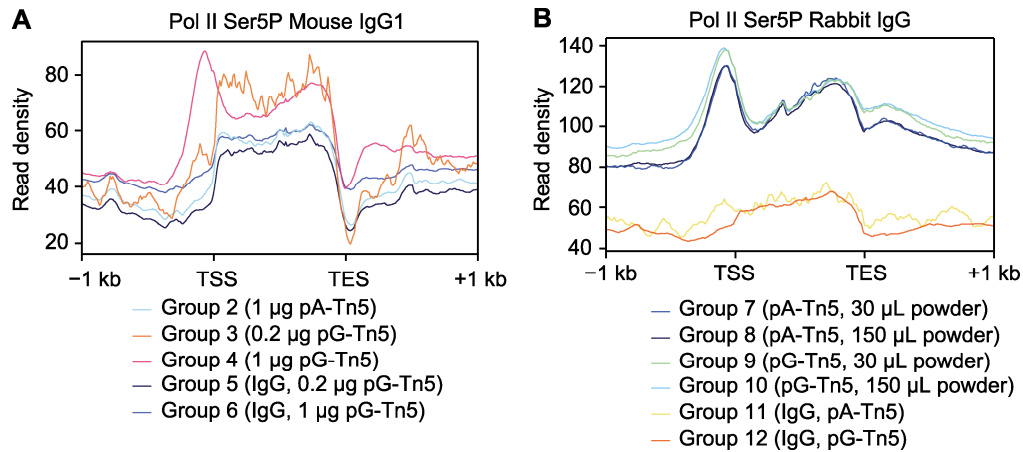


图4 CUT&Tag信号在基因上的平均分布
(A) 小鼠IgG1型抗体的CUT&Tag信号在基因上的平均分布; (B) 兔IgG型抗体的CUT&Tag信号在基因上的平均分布。TSS: 转录起始位点; TES: 转录终止位点

Figure 4 Average distribution of CUT&Tag signals on genes
(A) Average distribution of CUT&Tag signals on genes of mouse IgG1 antibody; (B) Average distribution of CUT&Tag signals on genes of rabbit IgG antibody. TSS: Transcription start site; TES: Transcription end site

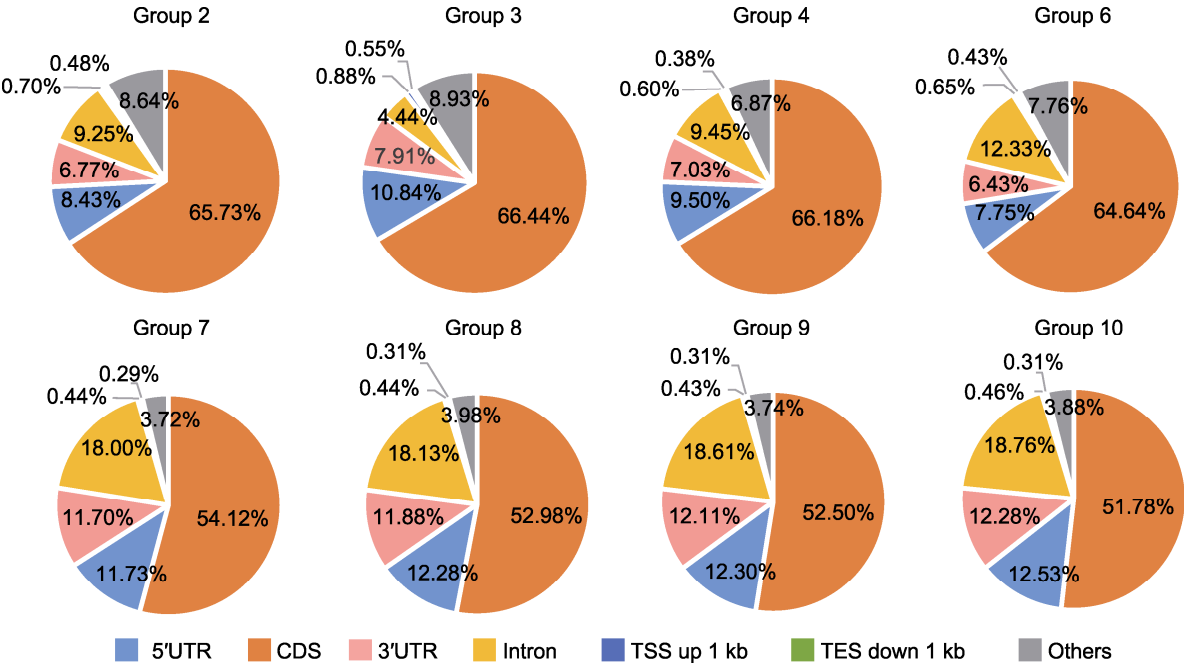


图5 CUT&Tag信号在染色体上的分布
CDS: 蛋白质编码区; 5'UTR: 5'非翻译区; 3'UTR: 3'非翻译区。TSS和TES同图4。Group 2、3、4、6、7、8、9和10同表2。

Figure 5 Distribution of CUT&Tag signals on chromosome
CDS: Coding sequence; 5'UTR: 5' untranslated region; 3'UTR: 3' untranslated region. TSS and TES are the same as shown in Figure 4. Group 2, 3, 4, 6, 7, 8, 9 and 10 are the same as shown in Table 2.

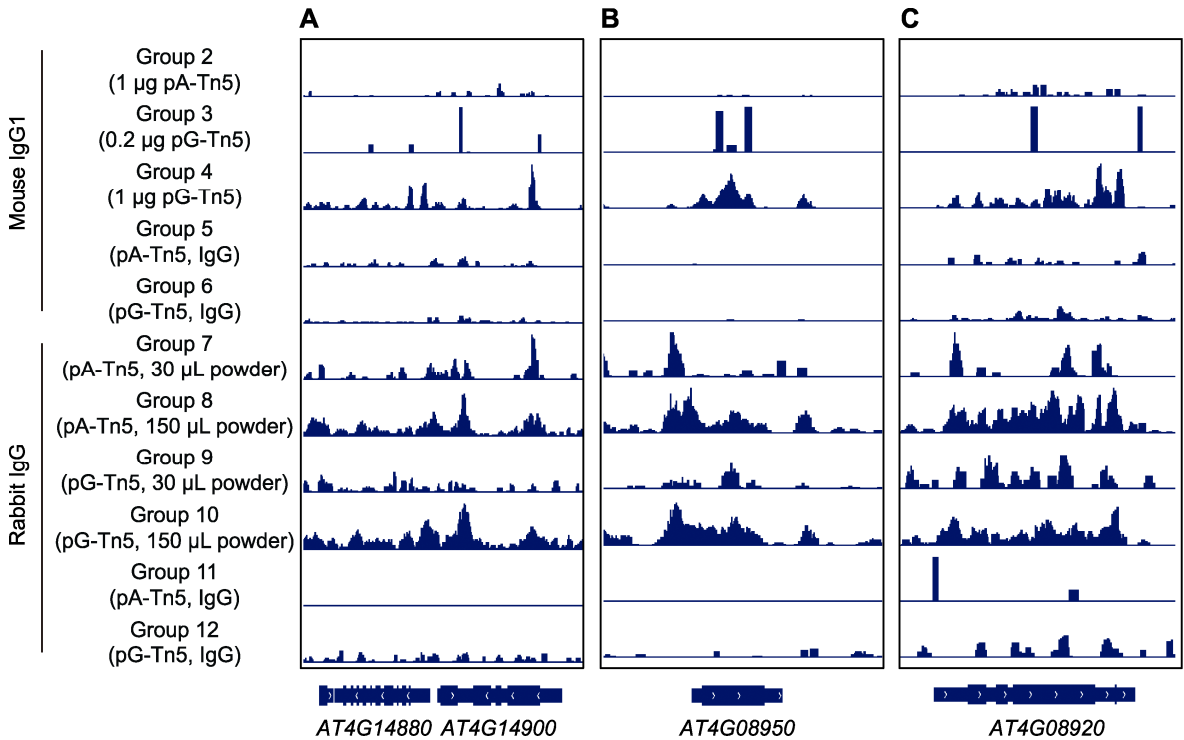


图6 利用IGV展示CUT&Tag信号在个别基因上的分布
(A) 基因AT4G14880-AT4G14900区段的CUT&Tag信号; (B) 磷响应蛋白基因EXO的CUT&Tag信号; (C) 蓝光受体基因CRY1的CUT&Tag信号。从上至下分别对应实验组2-12的实验数据。

Figure 6 Distribution of CUT&Tag signals on individual genes by IGV

(A) CUT&Tag signal of gene *AT4G14880* to *AT4G14900*; (B) CUT&Tag signals of phosphorus responsive protein gene *EXO*; (C) CUT&Tag signals of blue light receptor gene *CRY1*. From top to bottom, it corresponds to the experimental data of experimental groups 2–12.

重要问题。本研究进一步优化现有的CUT&Tag文库构建方法, 在提高文库构建性价比的同时提升文库质量, 为后续测序数据分析提供保障。在最初发表的CUT&Tag技术相关文章中只利用pA-Tn5对组蛋白修饰的抗体进行免疫共沉淀, 而作者Steven Henikoff也曾提出利用pA-Tn5进行CUT&Tag实验时, 为保证其靶向特异性需添加二抗(Kaya-Okur et al., 2019)。本研究构建并表达了融合蛋白pG-Tn5, 通过对比pA-Tn5和pG-Tn5在使用小鼠IgG1型抗体和兔IgG型抗体的CUT&Tag文库构建中的效率, 证明pG-Tn5与IgG1型抗体的亲和性更高, 但pA-Tn5和pG-Tn5与兔IgG型抗体的亲和力相近。因此, 研究者在进行CUT&Tag实验时应根据使用的抗体选择合适的Tn5融合蛋白。尽管目前市面上已有pA-Tn5和pG-Tn5的商业化产品, 但大多售价较高, 不适用于大量文库构建。我们的研究为科研人员提供了一个可行且实惠的备选方案。

在植物研究中, 类似的酶系策略应用研究已有报道。例如, 在拟南芥中, CUT&RUN技术被用于研究起始材料较少的胚乳组织中组蛋白修饰H3K27me3的分布(Zheng and Gehring, 2019)。随后, 基于细胞核提取而进行的CUT&Tag技术在棉花、油菜(*Brassica napus*)以及水稻(*Oryza sativa*)中相继建立。与动物细胞中的结果类似, CUT&Tag具有比ChIP-seq需要的起始材料更少且操作更简便的优点(Tao et al., 2020; Ouyang et al., 2021)。这在我们的研究中也得到证实。这一特点使得CUT&Tag技术为研究植物中更加精细的生物学过程的表现遗传修饰变化, 以及对基因表达的调控(如受精后种子的发育过程)提供了有力保障。

参考文献

- 陈威, 杨颖增, 陈锋, 周文冠, 舒凯 (2019). 表观遗传修饰介导的植物胁迫记忆. 植物学报 54, 779–785.
- 杜康兮, 沈文辉, 董爱武 (2018). 表观遗传调控植物响应非生物胁迫的研究进展. 植物学报 53, 581–593.
- 王泓力, 焦雨铃 (2020). 染色质免疫共沉淀实验方法. 植物学报 55, 475–480.
- 杨同文, 李成伟 (2014). 植物叶片衰老的表现遗传调控. 植物学报 49, 729–737.
- Brahma S, Henikoff S (2019). RSC-associated subnucleosomes define MNase-sensitive promoters in yeast. *Mol Cell* 73, 238–249.
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10, 1213–1218.
- Dancette OP, Taboureau JL, Tournier E, Charcosset C, Blond P (1999). Purification of immunoglobulins G by protein A/G affinity membrane chromatography. *J Chromatogr B: Biomed Sci Appl* 723, 61–68.
- Di L, Fu Y, Sun YS, Li J, Liu L, Yao JC, Wang GB, Wu YL, Lao KQ, Lee RW, Zheng GH, Xu J, Oh J, Wang D, Xie XS, Huang YY, Wang JB (2020). RNA sequencing by direct tagmentation of RNA/DNA hybrids. *Proc Natl Acad Sci USA* 117, 2886–2893.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 10, 1930.
- Kharchenko PV, Tolstorukov MY, Park PJ (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* 26, 1351–1359.
- Lindmark R, Thorén-Tolling K, Sjöquist J (1983). Binding of immunoglobulins to protein A and immunoglobulin levels in mammalian sera. *J Immunol Methods* 62, 1–13.
- Lovell S, Goryshin IY, Reznikoff WR, Rayment I (2002). Two-metal active site binding of a Tn5 transposase synaptic complex. *Nat Struct Biol* 9, 278–281.
- Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17, 10–12.
- Meers MP, Bryson TD, Henikoff JG, Henikoff S (2019). Improved CUT&RUN chromatin profiling tools. *eLife* 8, e46314.
- Ouyang WZ, Zhang XW, Peng Y, Zhang Q, Cao ZL, Li GL, Li XW (2021). Rapid and low-input profiling of histone

- marks in plants using nucleus CUT&Tag. *Front Plant Sci* **12**, 734679.
- Picelli S, Björklund ÅK, Reinius B, Sagasser S, Winberg G, Sandberg R** (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24**, 2033–2040.
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T** (2016). DeepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165.
- Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT** (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240–248.
- Skene PJ, Henikoff S** (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856.
- Tao XY, Feng SL, Zhao T, Guan XY** (2020). Efficient chromatin profiling of H3K4me3 modification in cotton using CUT&Tag. *Plant Methods* **16**, 120.
- Thakur J, Henikoff S** (2018). Unexpected conformational variations of the human centromeric chromatin complex. *Genes Dev* **32**, 20–25.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP** (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–192.
- Wang LG, Wang SQ, Li W** (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185.
- Zheng XY, Gehring M** (2019). Low-input chromatin profiling in *Arabidopsis* endosperm using CUT&RUN. *Plant Reprod* **32**, 63–75.

Optimization and Evaluation of Tn5 Transposase Fusion Protein in CUT&Tag

Shengyu Liu, Xiaobin Liu, Jiafu Zhu, Jing Su, Zhicheng Dong, Min Liu*

Innovative Center of Molecular Genetics and Evolution, Guangzhou Key Laboratory of Crop Gene Editing, Guangdong Provincial Key Laboratory of Plant Adaptation and Molecular Design, School of Life Sciences, Guangzhou University, Guangzhou 510006, China

Abstract Tn5 is a bacterial transposon. The engineered Tn5 can efficiently tag DNA while adding the adapter sequences. Therefore, it has been widely used in the preparation of high-throughput sequencing libraries. Cleavage Under Target & Tagmentation (CUT&Tag) is an improved technology for studying the interaction between protein and DNA, which has the advantages of good repeatability, high signal-to-noise ratio, and easy operation. This technology uses Protein A (pA) or Protein G (pG) and Tn5 to form a fusion protein, which can locate specific antibodies (the antibody is used to identify the target protein) and break the DNA near the target site while introducing sequencing adapters. Then, DNA was extracted, followed by PCR amplification to obtain the sequencing library. However, different types of antibodies have different affinities for pA and pG, thus limiting the application of CUT&Tag for some antibodies. To overcome this limitation, the expression vector of pG-Tn5 was constructed by recombination, and pG-Tn5 recombinant protein was obtained by prokaryotic expression and affinity purification. We used RNA polymerase II (Pol II)-specific antibodies (Pol II Ser5P, mouse IgG1 and rabbit IgG) to compare the efficiency of pA-Tn5 and pG-Tn5 in library preparation of CUT&Tag in *Arabidopsis*. The results showed that the IgG1 antibody had higher affinity for pG-Tn5, and the quality of the constructed library was better when pG-Tn5 was used. The rabbit IgG antibody has comparable affinities to the two enzymes. A lower starting amount of plant material can be applicable in CUT&Tag. This study provides a reference for the selection of Tn5 fusion proteins against different antibodies in future CUT&Tag experiments.

Key words ChIP, CUT&Tag, RNA polymerase II, Tn5

Liu SY, Liu XB, Zhu JF, Su J, Dong ZC, Liu M (2023). Optimization and evaluation of Tn5 transposase fusion protein in CUT&Tag. *Chin Bull Bot* **58**, 602–611.

* Author for correspondence. E-mail: minl@gzhu.edu.cn