

CUT&Tag 产物回收和建库方法的优化

韦晔^{1,2}, 李科², 卢大儒¹, 朱化星²

1. 复旦大学生命科学学院遗传工程国家重点实验室, 上海 1200438

2. 吴江近岸蛋白质科技有限公司, 上海 215200

摘要: 新兴的染色质靶向切割和标签化(cleavage under target and tagment, CUT&Tag)技术利用转座酶在目标蛋白结合的 DNA 附近进行切割并对切割下的 DNA 片段进行标签化, 通过后续的二代测序可以快速鉴定蛋白质–DNA 相互作用, 极大的简化了染色质免疫共沉淀测序(chromatin immunoprecipitation sequencing, ChIP-seq)的实验过程。CUT&Tag 中转座酶完成标签化后需要 DNA 回收或其他后处理才能进行建库 PCR, 不同的回收方法对 CUT&Tag 结果有着显著的影响。通过建立生物素化转座体–链霉亲和素磁珠体系(streptavidin beads recovery CUT&Tag, srCUT&Tag), 可以快速便捷地完成 CUT&Tag 的产物回收。本文在 K562 细胞中展开 H3K4me3、RNA 聚合酶 II (RNA polymerase II, RNAPII)、转录因子 CTCF 和 HMGA1 的 CUT&Tag 实验, 并利用现有的乙醇沉淀、片段分选(solid-phase reversible immobilization, SPRI)磁珠回收和直接 PCR 法, 以及本研究建立的 srCUT&Tag 方法对产物进行回收。结果表明, 从整体上看, SPRI 磁珠回收和 srCUT&Tag 方法有着较高的回收效率, 而乙醇沉淀法则回收效率低下。在全部 4 种 CUT&Tag 产物回收过程中, SPRI 磁珠回收均会损失大部分小于 150 bp 的产物片段。在 CTCF 和 HMGA1 CUT&Tag 产物的回收中, 直接 PCR 法则损失了大部分大于 300 bp 的片段并与其他回收方法的结果有较大的差别。因此, srCUT&Tag 能够比其他三种回收方法提供更多更完整的测序信息。综上所述, 新建立 srCUT&Tag 回收方法相比现有的 CUT&Tag 产物回收方法能提高建库效率并得到更好的数据质量, 为表观遗传学研究提供了更好的技术选择。

关键词: CUT&Tag; DNA 回收方法; H3K4me3; RNAPII; CTCF; HMGA1

Optimization of CUT&Tag product recovery and library construction method

Ye Wei^{1,2}, Ke Li², Daru Lu¹, Huaxing Zhu²

1. State Key Laboratory of Genetic Engineering School of Life Science, Fudan University, Shanghai 200438, China

2. Novoprotein Scientific Inc., Wujiang 215200, China

Abstract: The emerging cleavage under target and tagment (CUT&Tag) technology uses Tn5 transposase to cleavage

收稿日期: 2021-01-15; 修回日期: 2021-03-04

基金项目: 国家自然科学基金面上项目(编号: 81372706)和上海市科技创新行动计划项目(编号: 17JC1400902)资助[Supported by the National Natural Science Foundation of China (No. 81372706), and Shanghai Science and Technology Innovation Action Plan (No. 17JC1400902)]

作者简介: 韦晔, 在读硕士研究生; 专业方向: 生物工程。E-mail: 17210700130@fudan.edu.cn

通讯作者: 卢大儒, 博士, 教授, 研究方向: 遗传学。E-mail: drlu@fudan.edu.cn

朱化星, 博士, 高级工程师, 研究方向: 遗传学。E-mail: zhuhuaxing@novoprotein.com.cn

DOI: 10.16288/j.ycz.21-016

网络出版时间: 2021/3/15 15:30:30

URI: <https://kns.cnki.net/kcms/detail/11.1913.r.20210311.1727.002.html>

near the DNA binding site of target protein and study the generated DNA fragments by the next-generation sequencing. It can quickly identify protein-DNA interactions, which greatly simplifies the experimental process of ChIP-Seq. After CUT&Tag tagment reaction, DNA recovery or other post-processing is required to perform library construction PCR. Different recovery methods have significant impact respectively. By establishing Streptavidin beads recovery CUT&Tag(srCUT&Tag), we can quickly and conveniently complete the product recovery of CUT&Tag. We carried out CUT&Tag assay of H3K4me3, RNA Polymerase II (RNA polymerase II, RNAPII), transcription factor CTCF and HMGA1 in K562 cells with different recovery methods, including ethanol precipitation, fragment separation magnetic beads (SPRI) Magnetic bead recovery, direct PCR method, as well as our srCUT&Tag recovery method. The results show that among the CUT&Tag results of four different targets, the SPRI magnetic bead recovery and our srCUT&Tag methods have higher recovery efficiency than the direct PCR method and ethanol precipitation method. All CUT&Tag results showed that the recovery of SPRI magnetic beads would lose most of the product fragments less than 150 bp. In the recovery of CTCF and HMGA1, direct PCR lost most of the fragments larger than 300 bp and has significant difference from result of other recovery method. This enables srCUT&Tag to provide more real and higher-resolution information than other recovery method. In summary, the newly established srCUT&Tag recovery method can improve the efficiency of CUT&Tag library construction and obtain better data quality compared with the existing CUT&Tag product recovery method, providing a better technical choice for epigenetics research.

Keywords: CUT&Tag; DNA recovery method; H3K4me3; RNAPII; CTCF; HMGA1

染色质免疫共沉淀测序(chromatin immunoprecipitation sequencing, ChIP-seq)是经典的蛋白质-染色质相互作用研究方法^[1,2]。传统的染色质免疫共沉淀操作繁琐,背景较高,需要很大的测序量,实验可控性差。这些限制了其在低起始细胞量上的使用^[3]。CUT&Tag 是近年兴起的一种新型蛋白质-染色质相互作用研究方法^[4-6],该方法利用 Protein A/Protein G 融合的 Tn5 转座子,通过 Protein A/G-抗体相互作用将转座体栓系在靶抗体周围的区域,从而实现目标特异性的片段化反应。相比于传统的 ChIP-seq, CUT&Tag 操作相对简单,无需甲醛交联,信噪比较高,所需测序量少,且适用于极低起始量和单细胞的应用场景^[7,8]。CUT&Tag 的主要流程包括:(1)将细胞固定在 ConA 磁珠上;(2)利用表面活性剂 digitonin 透化细胞,先后加入一抗和二抗进行孵育;(3)洗去多余的抗体后,加入融合 ProteinA/G 的转座体使其与靶位点附近的抗体结合;(4)洗去多余的转座体,加入镁离子,在 37℃ 的条件下进行片段化反应;(5)加入 SDS 终止反应,通过乙醇沉淀或其他方法将酶切产生的 DNA 片段;(6)PCR 建库及二代测序^[9]。在 CUT&Tag 实验中,高效的 CUT&Tag 产物

DNA 回收对获取良好的 CUT&Tag 结果至关重要,本文主要针对 CUT&Tag 产物 DNA 提取方法进行讨论。

由于 CUT&Tag 所使用的转座体需要用 SDS 失活,体系中存在的 SDS 则会干扰后续的 PCR 反应,因此必须进行 DNA 的提取或利用其他方法消除 SDS 的干扰。早期的 CUT&Tag 实验方法中, CUT&Tag 产物 DNA 提取方法主要包括酚氯仿抽提-乙醇沉淀(后文简称乙醇沉淀)或片段分选磁珠(固相可逆固定磁珠,即 solid-phase reversible immobilization, SPRI, 后文简称 SPRI 磁珠)磁珠回收。酚氯仿抽提-乙醇沉淀是 DNA 回收的传统方法,其优点是原理简单,且没有明显的偏好性。缺陷则是操作相对繁琐,给回收操作带来了显著困难。特别是在低细胞起始量的 CUT&Tag 实验中,很容易造成产物 DNA 损失,增加了 CUT&Tag 实验的难度。SPRI 磁珠回收相对于乙醇沉淀耗时较短,回收效率高,操作难度低。但是目前市售的 SPRI 磁珠是为进行二代测序文库构建的片段分选而设计,不适用于回收 150 bp 以下的小片段。因此, SPRI 磁珠回收所得到的文库可能具有偏好性,有损失相应信息的风险。离心柱回收是另一种常见的 DNA 回收方法并且被广泛用于 CUT&Tag

1.2 仪器设备

细胞超净台、细胞培养箱、台式低温离心机购自、PCR 仪(Veriti 96)、紫外分光光度计(Nanodrop)为 Thermo Fisher 公司产品;涡旋振荡仪(Vertex-5)、旋转混合仪(QB-228)购自海门其林贝尔仪器制造有限公司;金属浴加热器(TU-100)及电热恒温培养箱(DHP-9052)购自上海一恒科学仪器公司。

1.3 细胞收集

K562 细胞复苏后于 10 mL 含有 10% FBS 的 RPMI 1640 培养基中,放入 37℃、5%CO₂ 浓度的细胞培养箱中。隔天按照 1:2 进行传代培养,细胞浓度稳定在 $0.5 \times 10^6 \sim 1.0 \times 10^6$ /mL。以保持实验细胞状态的稳定。室温下 600×g 离心收集新鲜细胞并计数,取所需数量的细胞于新的 1.5 mL EP 管中,室温下 600×g 离心 5 min,小心吸去上清;加入等体积 Wash Buffer (Novoprotein, N259)清洗一遍后重悬,根据需要使得每个反应的细胞浓度为 $10^2 \sim 10^5$ /100 μL。

1.4 CUT&Tag 实验

CUT&Tag 实验按照 NovoNGS[®] CUT&Tag 2.0 High-Sensitivity Kit (for Illumina[®])(N259)说明书上的标准流程进行。孵育一抗时一抗稀释比为 1:100,孵育二抗时二抗稀释比为 1:200,每个反应转座体用量为 1 μL 最终的标签化步骤反应体积为 100 μL。

1.5 DNA 的回收和建库

(a)对于乙醇沉淀回收的样品,向每份 CUT&Tag 产物中加入 3 μL 10% SDS, 10 μL 0.5 mol/L EDTA, 和 2.5 μL 20 mg/mL Proteinase K, 涡旋振荡混匀后, 50℃消化 1 h; 每管样品中加入 300 μL 酚氯仿混合物, 涡旋振荡充分混匀后移至锁相管中, 16000×g 室温离心 5 min。向每管样品中加入 100 μL 氯仿, 涡旋振荡充分混匀后, 16000×g 室温离心 5 min, 吸取水相至新的 1.5 mL EP 管, 加入 1 mL 100%乙醇, 用移液器吹打混匀; 将样品管置于-20℃静置 1 h, 4℃离心 20 min, 离心产物小心地倒出液体并在纸巾上沥干, 向每管样品中分别加入 1 mL 80%乙醇漂洗, 4℃下 16000×g 离心 1 min, 小心地倒出液体并在纸巾上沥去水珠后自然干燥。干燥后, 向每管样品中

加入 35 μL TE-RA-Buffer, 吹打重悬溶解 DNA 沉淀。

(b)对于 SPRI 磁珠回收的样品, 向每份 CUT&Tag 产物中加入 2 μL 10% SDS, 55℃加热 10 min, 然后向每管中加入 200 μL 在室温平衡后的磁珠, 吹打混匀, 室温静置 5 min, 磁珠结合后置于磁力架上静置 5 min, 移去上清, 80%乙醇洗涤 2 遍, 室温晾干。35 μL TE buffer 洗脱, 回收后的 CUT&Tag 产物在-20℃下储存或直接进行下一步 PCR 扩增。

(c)对于 srCUT&Tag 回收的样品, 向每份 CUT&Tag 产物中加入 2 μL 10% SDS, 55℃加热 10 min, 然后向每管中加入 1 μL 洗涤后按照 1:1 比例复溶的链霉亲和素磁珠, 吹打混匀, 室温静置 20 min, 磁珠结合后置于磁力架上静置 2 min, 移去上清, 1×Wash buffer 洗 2 遍, 室温晾干。35 μL TE buffer 复溶, 回收后的 CUT&Tag 产物在-20℃下储存或直接进行下一步 PCR 扩增。

(d)对于直接 PCR 的样品, 在 CUT&Tag 实验结束后移去片段化溶液上清, 利用 100 μL TAPs Wash buffer (10 mmol/L TAPs, 0.2 mmol/L EDTA)清洗样品, 加入 5 μL 0.2% SDS, 70℃加热 1 h。失活后的体系加入 15 μL 0.67% Triton X-100 中和 SDS。直接进行下一步 PCR 扩增。

每组回收后的 CUT&Tag 产物按照 NovoNGS[®] CUT&Tag 2.0 High-Sensitivity Kit (for Illumina[®])的流程建库并进行二代测序。

1.6 生物信息学分析

用 FastQC 和 MulitiQC 检查文库的质量; 用 Cutadapt 进行接头序列的去除和数据清洗; 用 Bowtie2 进行比对。用 Samtools 进行过滤和去重; 用 Macs2 进行识别峰; 用 IGV 进行数据可视化。用 Deeptools 绘制插入片段大小图, 不同实验组数据相关性图。

2 结果与分析

2.1 srCUT&Tag 回收体系中链霉亲和素用量的选择

为了确定回收使用链霉亲和素磁珠的用量和回收后直接 PCR 的可行性, 使用细胞数为 1×10^4 的

K562 细胞进行 H3K4me3 的 CUT&Tag 实验。不同的平行实验组分别用 0.2 μL , 0.5 μL , 1 μL , 2 μL , 3 μL , 4 μL 洗涤后的 NEB streptavidin magnetic beads 和英莱顿的链霉亲和素磁珠进行回收。产物按照 NovoNGS[®] CUT&Tag 2.0 High-Sensitivity Kit (for Illumina[®]) 建库的流程进行 14 个循环的 PCR, 利用琼脂糖凝胶电泳检验回收的片段分布(图 2)。对于 NEB 的磁珠, 0.2~4.0 μL 均可得到预期结果类似的 PCR 条带。而对于浓度稍高的英莱盾的磁珠, 当链霉亲和素磁珠用量为 0.1~2.0 μL 时, 所得到的 PCR 产物大小分布与预期结果类似。当磁珠用量大于 2 μL 时, PCR 条带的扩增水平降低; 4 μL 磁珠的回收产物则无法扩增出 PCR 条带。实验结果说明当链霉亲和素磁珠浓度较大时, 有可能会抑制后续的 PCR 扩增, 从而对建库过程造成干扰。在本文后续的研究中, 选择 1 μL 的英莱盾链霉亲和素磁珠作为 srCUT&Tag 所使用的链霉亲和素磁珠用量。

2.2 不同回收方法效率的比较

利用生物素化的 pAG-Tn5, 对于 10^4 K562 细胞分别进行 H3K4me3、RNAPII、CTCF 和 HGAM1 的 CUT&Tag 实验。所得到的产物分别用链霉亲和素磁珠(srCUT&Tag)、SPRI 磁珠、乙醇沉淀和直接 PCR 法进行回收或后处理后建库, 文库以等体积混匀以后在同一块芯片上进行二代测序, 其中 H3K4me3 和 RNAPII 进行了三重复平行实验, 而 CTCF 和 HMGA1 则进行了双重复实验。得到的数据在清洗

后进行回帖比对。统计所有实验组产出的 unique reads 和 total mapping reads 数量, 结果显示, 在 H3K4me3 和 RNAPII 的 CUT&Tag 实验结果中, srCUT&Tag 在回收 unique reads 产量和去除重复后的 total mapping reads 数上仅仅略低于最高的 SPRI 磁珠回收(图 3)。而在 CTCF 和 HGAM1 的 CUT&Tag 结果中, srCUT&Tag 回收则产生最多的 unique reads 和 total mapping reads 数。直接 PCR 法和乙醇沉淀法贡献的数据量都相对较低, 在 4 种不同的 CUT&Tag 结果中直接法 PCR 得到的 total mapping read 数相当于产出数据量最高方法的 43%~60%, 而乙醇沉淀法产出的数据量则相当于产出数据量最高方法的 24%~30%。整体而言, srCUT&Tag 在 4 种回收方法中都保持着稳定的高回收效率。而若采用回收效率较低的乙醇沉淀法, 则有损失超过 2/3 原始产物信息的风险。

2.3 不同回收方法产出 CUT&Tag 文库片段大小和文库信息的差异

利用 deeptool 中的 collectinsertsizemetrics 统计不同实验组插入片段大小分布。比较使用不同回收方法所产生的文库中插入片段的大小分布(图 4A), 可以发现不同回收方法所保留的底物样本的片段大小分布并不相同。在 4 种不同的回收方法中, srCUT&Tag 和乙醇沉淀法产出的文库片段大小分布相似。对于小于 150 bp 的片段, SPRI 磁珠法的回收率均低于其他几种回收方法。对 H3K4me3 的

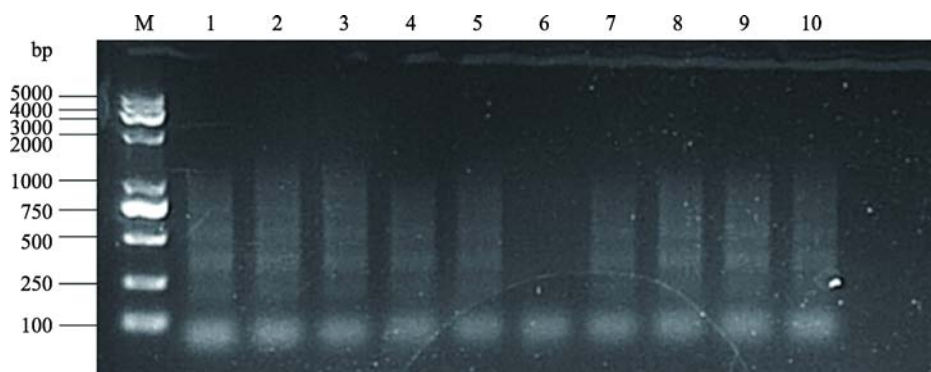


图 2 不同链霉亲和素磁珠及不同用量回收效果测试

Fig. 2 Measurement of CUT&Tag PCR products amplified from DNA fragments recovered using different Streptavidin magnetic beads or its volume

M: DNA ladder 2000 plus; 1: 4 μL NEB 磁珠; 2: 2 μL NEB 磁珠; 3: 1 μL NEB 磁珠; 4: 0.5 μL NEB 磁珠; 5: 0.2 μL NEB 磁珠; 6: 4 μL 英莱盾磁珠; 7: 2 μL 英莱盾磁珠; 8: 1 μL 英莱盾磁珠; 9: 0.5 μL 英莱盾磁珠; 10: 0.2 μL 英莱盾磁珠;

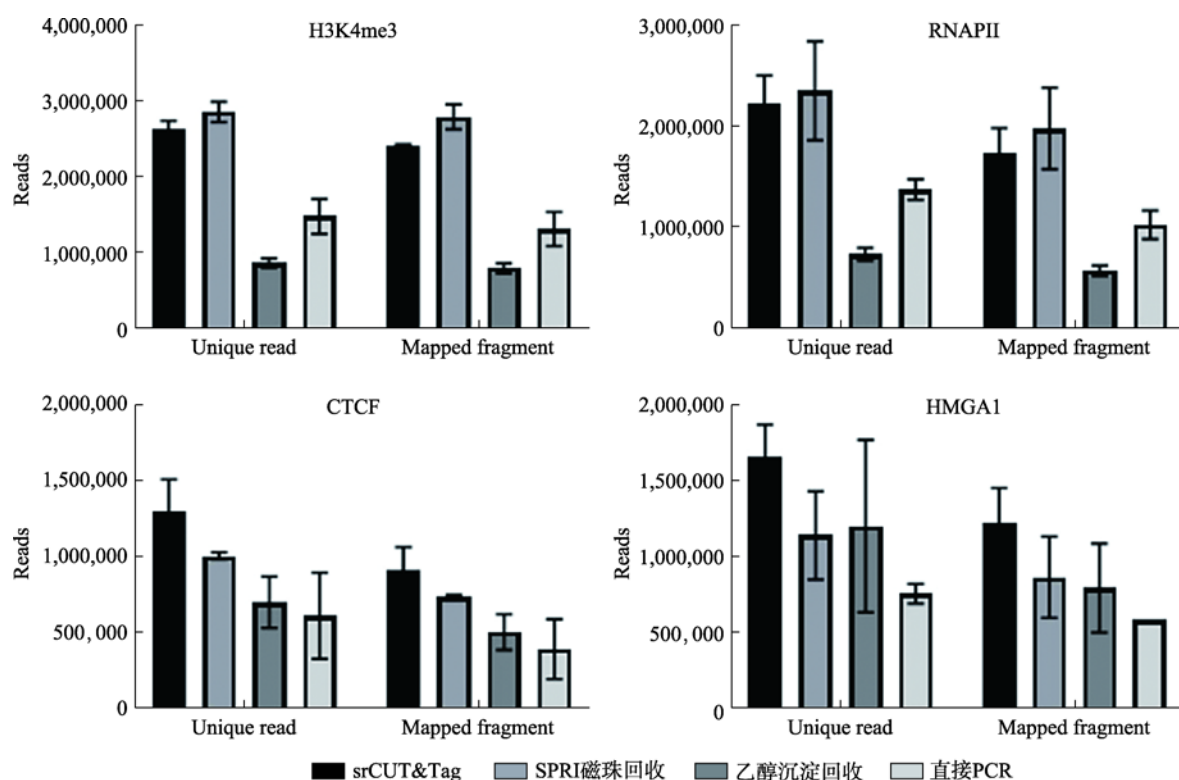


图 3 不同回收方法文库产出规模比较

Fig. 3 Comparison of library output scale of different recovery methods

CUT&Tag 结果进行比较,使用 srCUT&Tag、乙醇沉淀和直接 PCR 法产出的文库中小于 150 bp 的片段在整体文库中的平均占比分别为 23.07%、15.81%和 22.16%,而 SPRI 磁珠所产出的文库中小于 150 bp 片段的占比仅为 10.80%。对转录因子 CTCF 的 CUT&Tag 结果进行比较, srCUT&Tag、乙醇沉淀和直接 PCR 法所产出的文库中小于 150 bp 的片段在整体文库中的平均占比分别为 39.98%, 32.40%和 62.03%,而 SPRI 磁珠回收文库中小于 150 bp 片段的占比则仅为 20.43% (图 4B)。结合 2.2 中所展示的不同回收方法文库产量,在 H3K4me3 的 CUT&Tag 结果中, srCUT&Tag 产出的小于 150 bp 片段的 reads 数为 SPRI 磁珠回收法的 1.75 倍;而在 CTCF 的 CUT&Tag 结果中, srCUT&Tag 产出的小于 150 bp 片段的 reads 数则达到 SPRI 磁珠回收法的 2.4 倍。由于 CTCF 和 HMGA1 文库中含有更多的小于 150 bp 的片段,一定程度上解释了 SPRI 磁珠在 CTCF 和 HMGA1 样品的回收效率现象。此外,在 CTCF 和 HMGA1 的 CUT&Tag 结果中,直接 PCR

法产生的大片段含量明显低于使用其他回收方法所产出的文库。其文库中几乎不含有大于 300 bp 的大片段,且直接 PCR 法也影响了 CTCF 和 HMGA1 文库的数据产量,低于链霉亲和素磁珠回收法产出的文库数据产量。

为了探究不同回收方法产生文库的差异,对 H3K4me3, RNAPII 部分平行组数据和 CTCF 和 HMGA1 全部的平行组的数据结果进行关联性分析 (图 5A)。在 H3K4me3, RNAPII 和 CTCF 的 CUT&Tag 实验中, srCUT&Tag, SPRI 磁珠和乙醇沉淀法均有较好的相关性 (>0.9),而在 HMGA1 的 CUT&Tag 结果中,所有回收方法平行组和相互间的相关性都较低,但是 srCUT&Tag 方法平行组间仍然有 0.7 的相关性系数,仅次于乙醇沉淀法平行组间的 0.71。这证明了 srCUT&Tag 具有和传统 SPRI 磁珠和乙醇沉淀回收相当的可重复性。直接 PCR 法在 H3K4me3 实验中具有较高的可重复性,并且不同回收方法的信号和信噪比相仿。但是在其余的 3 组条件 CUT&Tag 中均展现出较低的平行组内相关性。同时,

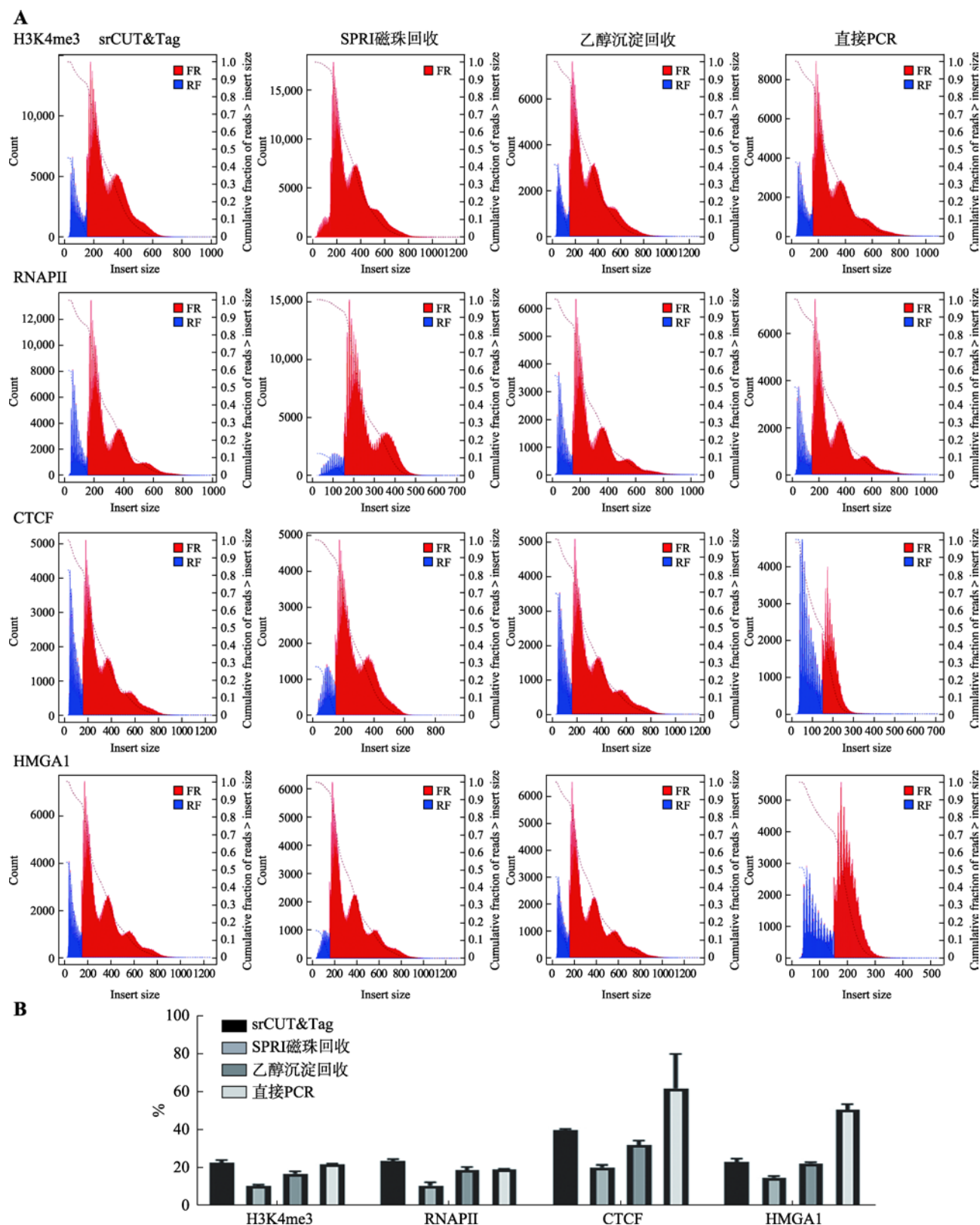


图 4 不同回收方法文库片段大小分布

Fig. 4 Size distribution of library fragments by different recovery methods

A: 4 种不同回收方法文库片段的大小分布; B: 不同回收方法中小于 150 bp 片段的占比

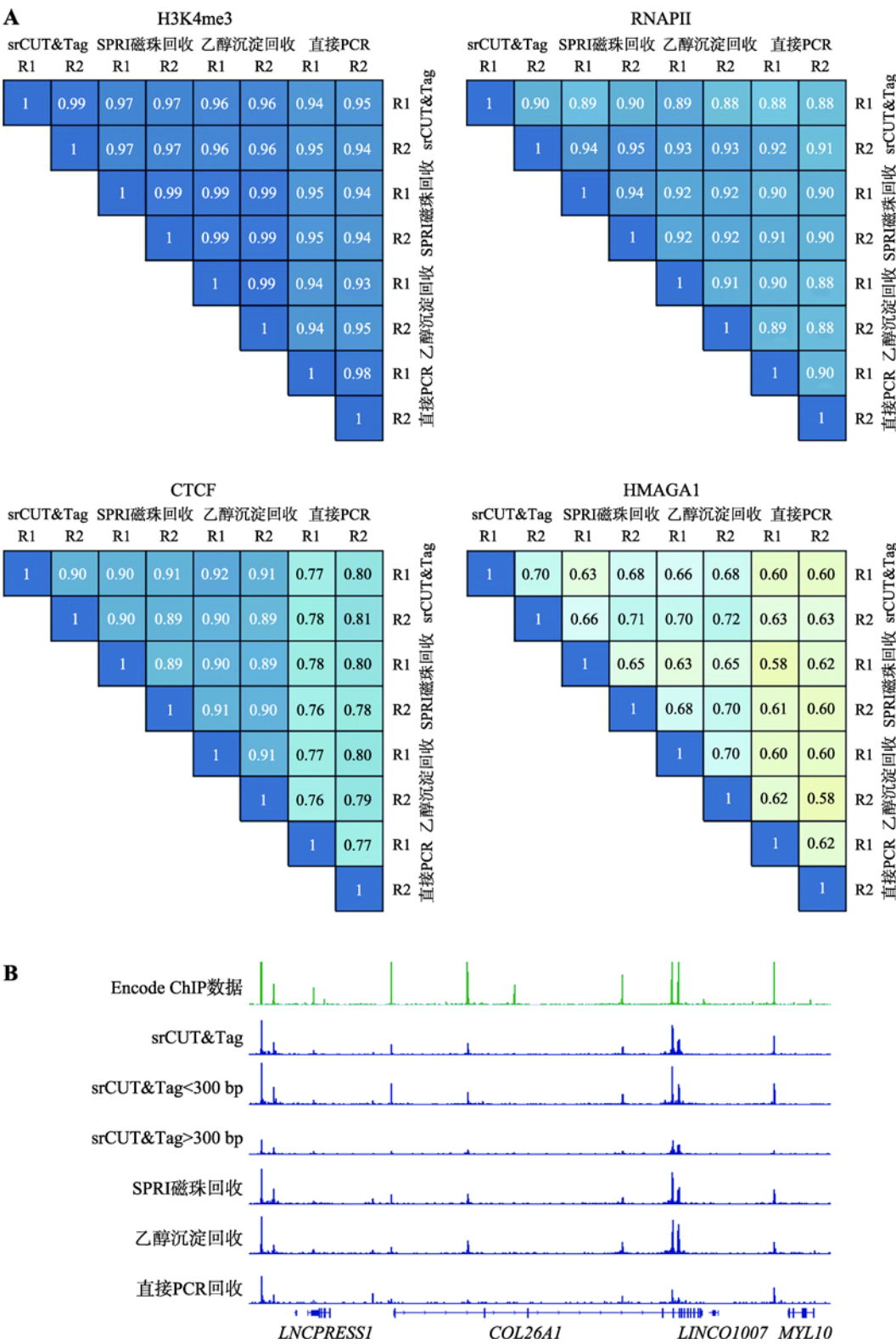


图 5 不同回收方法文库的相关性

Fig. 5 Correlation of CUT&Tag generated by different recovery methods

A: 4 种不同 CUT&Tag 产物中不同回收方法产生结果之间的相关性; B: CTCF CUT&Tag 结果中, 直接 PCR 法和其他三种方法的结果存在明显的差异。

在 H3K4me3 以外的其他 3 类 CUT&Tag 结果中,直接 PCR 法与其他 3 种方法间的相关性系数均较低,这意味着直接 PCR 法与 srCUT&Tag 及传统回收方法所得到的结果均存在明显差异。在相关性系数较低的 CTCF 的 CUT&Tag 结果中,可以明显看到直接 PCR 法产出的信号分布与其他回收方法的结果不同,包括在多个位置的信号强度较弱或完全缺失(图 5B)。通过将 srCUT&Tag 的测序结果拆分为大于 300 bp 和小于 300 bp 的部分,可以看到,这些峰高的差异并非是简单的由上文所提到的直接 PCR 法中

大于 300 bp 片段损失所带来,这暗示了在直接 PCR 法中 TritonX-100 对扩增酶的抑制可能造成了更复杂的带有偏好性的扩增,并对 CUT&Tag 的结果造成了显著的影响。

相关性分析显示,SPRI 磁珠与 srCUT&Tag 和乙醇沉淀的结果差别并不大。为了进一步研究小片段损失对 CUT&Tag 结果可能的影响,选取 RNAPII R1 实验组的 CUT&Tag 结果,将 srCUT&Tag 的测序结果拆分为插入片段小于 150 bp 和大于 150 bp 两部分,比较不同大小的插入片段产生的峰高的差异(图 6)。

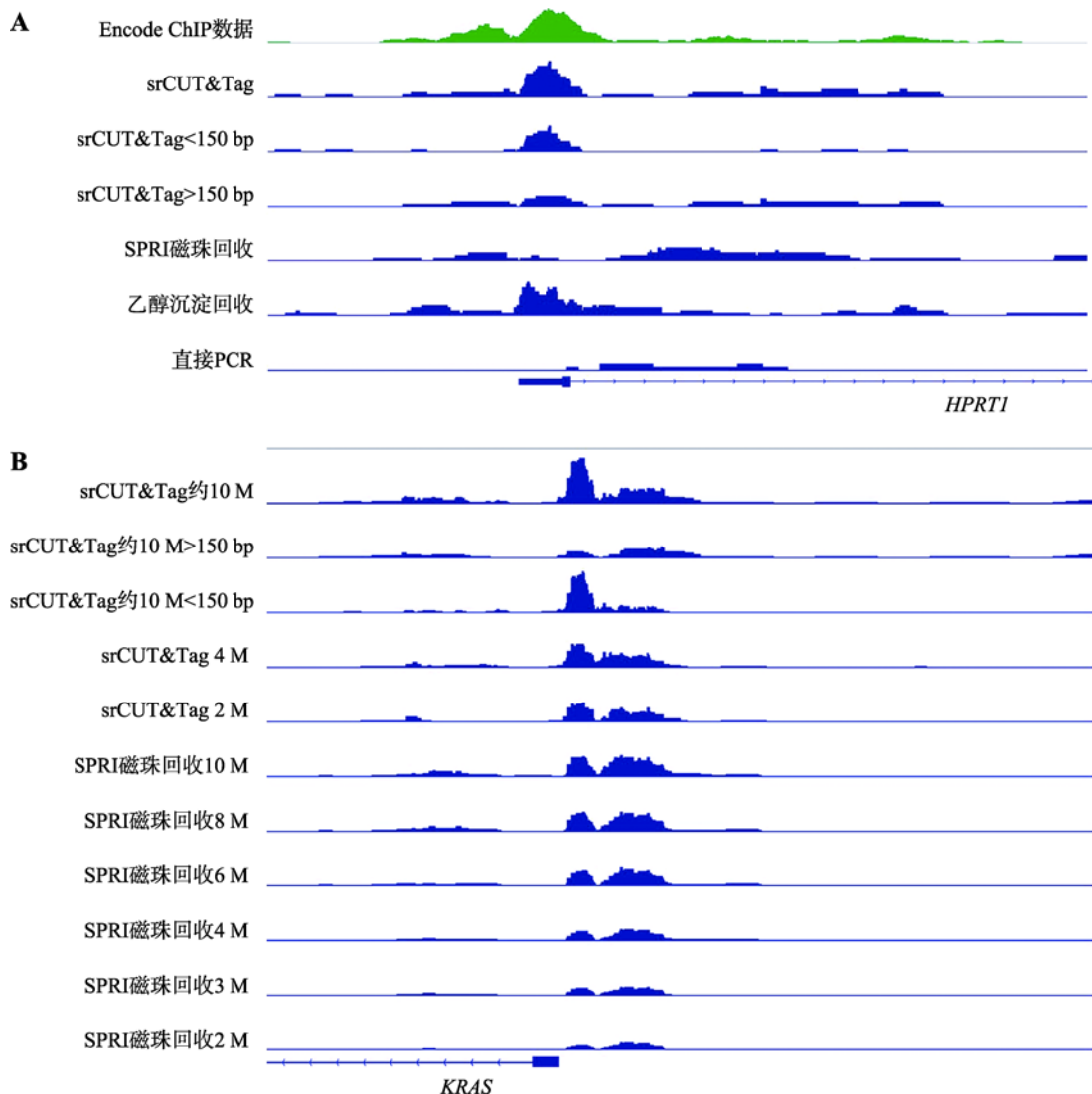


图 6 片段选择性对 CUT&Tag 结果的影响

Fig. 6 The effect of fragment selectivity on CUT&Tag results

A: 小于 150bp 的片段贡献了 RNAPII CUT&Tag 中 *HPRT1* 启动子位点上的主要部分; B: 小于 150bp 的片段贡献了 RNAPII CUT&Tag 中 *KRAS* 启动子位点上的主要部分, srCUT&Tag 和 SPRI 磁珠的差异在数据量较小时尤其明显。

在次黄嘌呤腺苷酸转移酶基因 *HPRT1* 位点上, 小于 150 bp 的插入片段贡献了启动子峰的主要部分。控制不同回收方法投入的起始 unique reads 数为 5 M, 在 *HPRT1* 位点上, 对小于 150 bp 的片段保留效果较好的链霉亲和素磁珠回收法、乙醇沉淀法产生结果一致的峰图, 而对小于 150 bp 的片段保留效果较差的 SPRI 磁珠回收法则没有信号(图 6A)。同样的现象也可以在 *KRAS* 基因位点上观察到。当投入数据量从 10 M 下降到 2 M 时, SPRI 磁珠文库在 *KRAS* 基因启动子上的信号急速衰减, 但是链霉亲和素磁珠回收法产生的文库受投入数据量下降的影响较小, 在 2 M 数据量投入时仍有与 SPRI 磁珠 8~10 M 回收结果类似的效果。显示了在起始数据量较低的情况下, 小于 150 bp 的片段回收率差距带来的信号差异尤其明显(图 6B)。

综上所述, 比较使用 4 种不同的回收方法产出的 CUT&Tag 结果, 基于 srCUT&Tag 的链霉亲和素磁珠回收法对于不同抗体都有较高的数据产量, 而 SPRI 磁珠法均会损失一部分小于 150 bp 的小片段, 直接 PCR 法则在转录因子 CTCF 和 HMGA1 的回收中损失了大部分大于 300 bp 的片段, 从而导致一部分结合位点信息的损失, 乙醇沉淀法的回收效率相对较低。因此, 总体上而言, 基于 srCUT&Tag 的链霉亲和素磁珠法是对 CUT&Tag 结果信息保留效率最高的回收方法。

2.4 srCUT&Tag 有利于低起始细胞量的 CTCF CUT&Tag

考虑到片段选择性在低数据量水平上差异更加显著, 推测对于极低的起始细胞量, 且当目的位为丰度相对较低的转录因子时, 不同回收方法贡献的数据质量可能有更大的差异。利用生物素化的 pAG-Tn5, 分别在 200、500、1000 个 K562 细胞水平上进行 CTCF 的 CUT&Tag 实验, 产物利用 srCUT&Tag、SPRI 磁珠、乙醇沉淀酚氯仿抽提和直接 PCR 回收后建库测序。与 10⁴ 细胞的回收结果类似, 在 200~1000 细胞水平上, 产生的文库从大到小排列为: srCUT&Tag>SPRI 磁珠法>直接 PCR 法>乙醇沉淀法(表 1)。其中链霉亲和素磁珠回收文库在去除重复后的 mapping reads 数比 SPI 磁珠文库高 15%~23%, 比乙醇沉淀文库高 108%~211%, 比直接 PCR 文库高 118%~128%。在链霉亲和素磁珠法产出的 CUT&Tag 结果中, macs2 软件在 200, 500, 1000 细胞水平上分别识别到 8301、13389 和 16269 个峰, 多于在相同的细胞数量上, SPRI 磁珠法产出的峰值数(7630、11875 和 14145), 乙醇沉淀法产出的峰值数(2949、5749 和 9752)和直接 PCR 产出的峰值数(5992、9780 和 11028)。

在基因组上选取一段具有代表性的区域, 通过 IGV 的可视化获取不同回收方法得到的覆盖度图(图 7A)。在 500 细胞和 1000 细胞水平上, srCUT&Tag

表 1 低细胞数量时不同回收方法的回收效率比较

Table 1 Comparison of recovery efficiency of different recovery methods in few cells condition

回收方法	细胞数量	起始 unique read	去重后的 mapping reads	峰值数
srCUT&Tag	200.00	6384518	2436658	8301
srCUT&Tag	500.00	15961295	4628092	13389
srCUT&Tag	1000.00	29650679	6318990	16269
SPRI 磁珠回收	200.00	5407687	2100856	7630
SPRI 磁珠回收	500.00	13168069	3751782	11875
SPRI 磁珠回收	1000.00	23021799	5238500	14145
乙醇沉淀回收	200.00	1580168	783550	2497
乙醇沉淀回收	500.00	4243044	1707974	5749
乙醇沉淀回收	1000.00	7608218	3012370	9752
直接 PCR	200.00	2809188	1116464	5992
直接 PCR	500.00	7461906	2027234	9870
直接 PCR	1000.00	13058335	3446298	11028

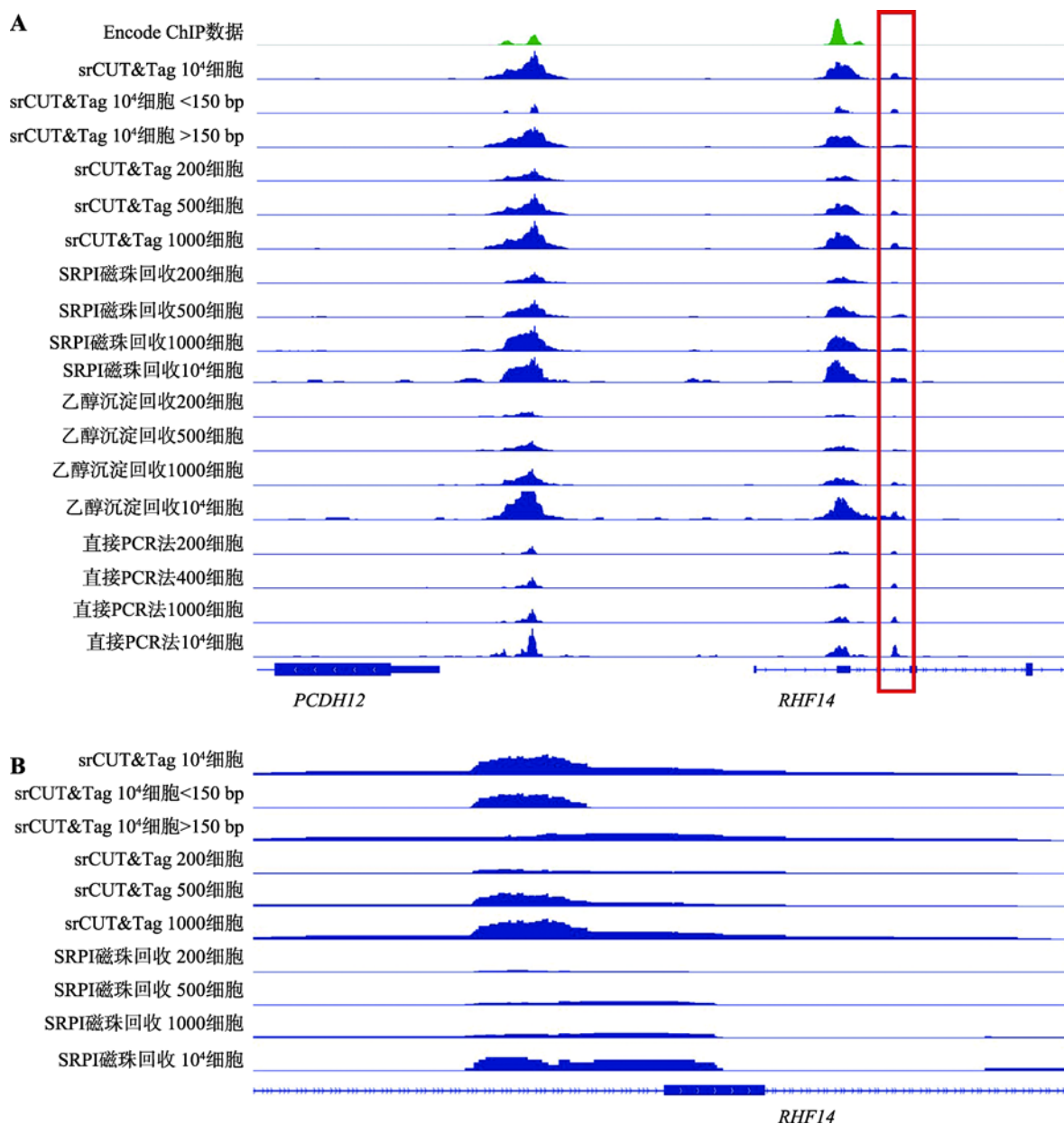


图 7 不同回收方法在少量细胞上的对比

Fig. 7 Comparison of different recovery methods in few cells condition

A: 200~1000 细胞水平上的不同 CTCF 回收结果; B: 在 *RNF14* 基因侧翼的弱结合位点, SRPI 磁珠回收相比于 srCUT&Tag 出现了明显的信号缺失。

和 SPRI 磁珠法均能产生与 10⁵ 细胞水平相类似的 CUT&Tag 产物峰图。而直接 PCR 法和乙醇沉淀法在 500 细胞产生的覆盖图谱峰高则远低于 10⁵ 细胞的结果, 1000 细胞的覆盖图谱整体上同样弱于 10⁵ 细胞产生的图谱结果。另一方面, 与前文所述的结果类似, 在 *RNF14* 基因主峰侧翼的较弱信号位点上, 低细胞起始量状态下的 SPRI 磁珠回收几乎损失了

全部的信号。与 10⁴ 细胞的 SPRI 磁珠回收结果相比, 可以看出 SPRI 磁珠法损失小于 150 bp 片段的问题在细胞数量更少的情况下更加显著, 从而相比于 srCUT&Tag 提供的结果显现出明显的劣势(图 7B)。总体来看 srCUT&Tag 在起始细胞量较少的情况下可以产出更完整和丰富的 CUT&Tag 文库信息, 适用于极低起始细胞量的 CTCF CUT&Tag 实验。

表 2 srCUT&Tag 与其他 CUT&Tag 产物回收方法的比较

Table 2 Comparison between srCUT&Tag and other DNA recovery method

回收方法	耗时	偏好性	单个反应成本(¥)	回收效率
酚氯仿抽提-乙醇沉淀	2 h~1 d	无明显选择性	—	较低
SPRI 磁珠回收	0.5 h	损失<150bp 片段	10	高
srCUT&Tag	0.5 h	无明显选择性	0.5~2.5	高
直接 PCR 法	1 h	部分情况下损失>300bp 片段, 并造成信号损失	—	中等
离心柱回收	1 h	无明显选择性	>30	高

3 讨论

Henikoff在其早期关于CUT&Tag方法的文献报道中主要使用2×SPRI磁珠回收,随后将酚氯仿抽提乙醇沉淀作为CUT&Tag产物的标准回收方法加以推广,而在近期的文献中,则主要推荐其基于Triton中和的PCR直接扩增法^[9]。针对现有的CUT&Tag回收方法各自存在的缺陷,本研究提出了一种新的srCUT&Tag体系:利用含有生物素化接头的pAG-Tn5转座酶进行DNA片段化和生物素化,产物DNA通过链霉亲和素磁珠回收。在K562细胞中使用4种不同抗体开展CUT&Tag实验,srCUT&Tag的链霉亲和素回收法仅对H3K4me3和RNAPII的回收效率略低于传统的SPRI磁珠回收效率,对CTCF和HMGA1的回收效率则高于SPRI磁珠回收法,对于4种抗体其回收效率都远高于乙醇沉淀回收法和直接PCR法。并且srCUT&Tag回收的DNA片段大小分布中与乙醇沉淀回收法相似,不存在2XSPRI磁珠丢失小片段或者直接PCR扩增法丢失部分大片段的问题,更有利于CUT&Tag产物完整信息的保留(表2)。

本研究发现,使用不同的抗体所产生的CUT&Tag文库有着不同的DNA片段大小分布,因此可能导致不同回收方法搭配不同抗体产出的CUT&Tag文库有不同的表现。以链霉亲和素磁珠回收法产出的文库作为标准,H3K4me3和RNAPII的CUT&Tag文库中,小片段含量本身并不高,特别是在H3K4me3的文库中,不同建库方法产生的片段差异并不大。150 bp左右的长度是单个核小体DNA的大小,而在CTCF和HMGA1文库中,由于转录因子结合区域本身有着比组蛋白周围更广阔的核小体耗尽区域,因此转录因子的CUT&Tag文库中,含有更多小于单

个核小体大小的150 bp以下片段。更多小片段的产生不仅凸显了SPRI磁珠在片段选择性上的差异,也造成直接PCR法的扩增偏好。在加入了Triton X-100后,扩增酶的活性在一定程度上受到了影响,从而使得PCR扩增大片能力下降,使得整体的文库大小向小片段的方向偏移。而在CTCF文库中,直接PCR法与srCUT&Tag回收法的差异不仅仅来源于300 bp以上大片段回收效率的低下;将文库中小于300 bp的片段分离后,在相同文库投入数据量的情况下,直接PCR法仍然存在许多不同大小片段的缺失。

综上所述,与传统建库方法相比较,srCUT&Tag回收法兼顾了现有回收方法中酚氯仿抽提乙醇沉淀回收法没有回收偏好性,SPRI磁珠回收效率高和直接PCR法一管式操作便捷的优势,提高了实验的便捷性,避免了更繁琐的实验流程。基于srCUT&Tag系统的链霉亲和素磁珠在10⁴~200细胞水平上都有较高的回收效率,并且在细胞量较低的情况下仍能够提供较好的回收效果,其CUT&Tag文库产量和数据信息优于现有的酚氯仿抽提乙醇沉淀回收、SPRI磁珠回收和直接PCR法。本研究建立了基于srCUT&Tag的链霉亲和素磁珠回收系统,并比较了该回收方法与现有的主流回收方法,对各自的优缺点进行了系统的分析,在保证数据完整性的基础上进一步优化了CUT&Tag的易用性。

参考文献(References):

[1] Kidder BL, Hu GQ, Zhao KJ. ChIP-Seq: technical considerations for obtaining high-quality data. *Nat Immunol*, 2011, 12(10): 918–922. [\[DOI\]](#)
[2] Policastro RA, Zentner GE. Enzymatic methods for genome-wide profiling of protein binding sites. *Brief*

- Funct Genomics*, 2018, 17(2): 138–145. [DOI]
- [3] Ludwig CH, Bintu L. Mapping chromatin modifications at the single cell level. *Development*, 2019, 146(12): dev170217. [DOI]
- [4] Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun*, 2019, 10(1): 1930. [DOI]
- [5] Douse CH, Tchasovnikarova IA, Timms RT, Protasio AV, Seczynska M, Prigozhin DM, Albecka A, Wagstaff J, Williamson JC, Freund SMV, Lehner PJ, Modis Y. TASOR is a pseudo-PARP that directs HUSH complex assembly and epigenetic transposon control. *Nat Commun*, 2020, 11(1): 4940. [DOI]
- [6] Ye BY, Yang GG, Li YM, Zhang CY, Wang QW, Yu GY. ZNF143 in chromatin looping and gene regulation. *Front Genet*, 2020, 11: 338. [DOI]
- [7] Janssens DH, Meers MP, Wu SJ, Babaeva E, Meshinchi S, Sarthy JF, Ahmad K, Henikoff S. Automated CUT&Tag profiling of chromatin heterogeneity in mixed-lineage leukemia. *bioRxiv*, 2020. [DOI]
- [8] Henikoff S, Henikoff JG, Kaya-Okur HS, Ahmad K. Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *eLife*, 2020, 9: e63274. [DOI]
- [9] Kaya-Okur HS, Janssens DH, Henikoff JG, Ahmad K, Henikoff S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc*, 2020, 15(10): 3264–3283. [DOI]
- [10] Skene PJ, Henikoff JG, Henikoff S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc*, 2018, 13(5): 1006–1019. [DOI]
- [11] Yu FL, Sankaran VG, Yuan GC. CUT&RUNTools 2.0: A pipeline for single-cell and bulk-level CUT&RUN and CUT&Tag data analysis. *bioRxiv*, 2021. [DOI]
- [12] Tao XY, Feng SL, Zhao T, Guan XY. Efficient chromatin profiling of H3K4me3 modification in cotton using CUT&Tag. *Plant Methods*, 2020, 16: 120. [DOI]

(责任编辑: 孟卓贤)