Harnessing Template Switching Technology for Preparation of Low-Input Ligation-Free Sequencing Libraries

Rashmi Rajendra, Nathalie Bolduc, Simon Lee, Craig Betts*, Suvarna Gandlur & Andrew Farmer[1]

[1]Corresponding Author: Andrew_Farmer@clontech.com
*Current Affiliation: Cellular Research Inc., Palo Alto, CA

Abstract

Preparation of sequencing libraries from minute amounts of DNA can be extremely challenging. Most methods rely on inefficient adapter ligation and require an input of 5–10 ng of high-quality, double-stranded DNA. Template switching has gained renewed attention as an alternative to ligation for NGS library production due to its single-step adapter addition and inherent sensitivity for amplifying picogram quantities of material. However, until now, template switching has been limited to use only with RNA samples. Here, we report a modified template switching reaction that accepts DNA instead of RNA as a template. This new technology accommodates either double-stranded or single-stranded DNA templates, thus making it very amenable to a wide range of applications.

We have created a single-tube workflow allowing the use of DNA template switching for sequencing on Illumina® platforms. The workflow can be completed in 4 hours and works with 100 pg to 10 ng of double-stranded or single-stranded DNA. To validate the performance of the technology, we applied it to chromatin immunoprecipitation (ChIP) sequencing (ChIP-seq). We used histone H3 trimethyl lysine 4 (H3K4me3) pull-down on HEK293T cells, and found that over 85% of the peaks identified with the DNA template switching method overlap with peaks identified from data generated by the ENCODE project. Furthermore, we found 86% overlap between the peaks identified when comparing H3K4me3 pull down experiments performed with 1 x 10^6 cells and 1 x 10^4 cells. Inputs containing as little as 500 pg yielded a non-redundant rate >80%, meeting standards established by the ENCODE consortium. These data illustrate that DNA template switching technology provides a robust and reliable tool for preparation of sequencing libraries from challenging samples such as ChIP DNA, particularly at low inputs.

Introduction

The ChIP Elute Kit is a fast and very simple tool to replace the slow and tedious DNA recovery steps performed at the end of ChIP experiments. It is compatible with any upstream ChIP protocol, and efficiently isolates, purifies, and concentrates ChIP DNA in less than one hour. It produces ssDNA that is directly compatible with the DNA SMART™ ChIP-Seq Kit that uses an adaptation of SMART® template switching technology for use with DNA. High-quality libraries can be generated using these two products in an efficient, time-saving manner.

Materials and Methods

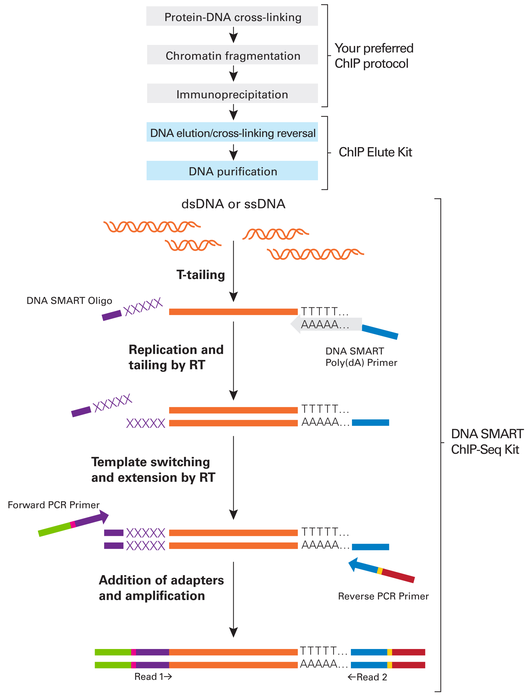
ChIP was performed on HEK 293T cells with ChIP-grade antibodies according to standard methods. After washes, DNA–protein complexes immobilized to Protein A/G agarose beads were processed using the traditional method or the ChIP Elute Kit. For the traditional method, DNA–protein complexes were eluted from the beads in SDS buffer (1% SDS, 0.1 M $NaHCO_3$) and cross-links were reversed in the presence of 250 mM NaCl at 65°C overnight, followed by a two-hour Proteinase K treatment. The DNA was purified and concentrated with a MN NucleoSpin Gel and PCR Clean-Up kit (using Buffer NTB). Samples processed with the ChIP Elute Kit were handled according to the kit protocol (15 min Protinase K treatment).

ChIP-seq libraries were generated using the DNA SMART ChIP-Seq Kit according to the user manual with size selection using Option 1 or Option 4. Sequencing was carried out on Illumina MiSeq® or HiSeq® 2500 instruments. All runs were performed with paired-end sequencing; the Custom Read2 Seq Primer from the DNA SMART ChIP-Seq Kit was used for some runs.

Mapping of reads (unpaired) to the human genome (hg19) was performed using Bowtie2 with default settings (plus trimming of the first three 5′ nucleotides of the reads obtained with the Read Primer 1). Uniquely mapping reads were selected and the SAM files were sorted and converted to BAM files using SAMTOOLS. Peaks were identified using MACS version 1.4 (default settings except the p-value cutoff set at 1 x 10^{-7}) with a non-enriched (total chromatin input) sample as a control. For quantitative peak overlap comparison to data generated by the ENCODE consortium (293 cells, anti-H3K4me3 antibody, U. Washington), raw data were downloaded as fastq files from http://genome.ucsc.edu/ENCODE/downloads.html and analyzed similarly to the data generated with the DNA SMART ChIP-Seq Kit (Figure 5, Panel C). Peaks were visualized using the UCSC genome browser (along with tracks displaying anti-CTCF peaks for 293 cells; U. Washington. Figure 2, Panel A; and Figure 3, Panel A).

03.15 US (633753)

1 Workflow overview

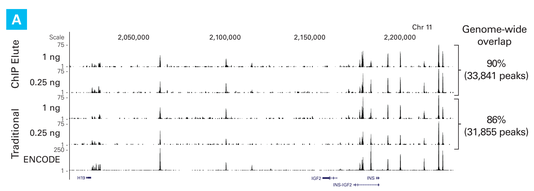


ChIP Elute and DNA SMART workflow. The ChIP Elute Kit dramatically decreases the time for cross-linking reversal and DNA elution compared to traditional methods. While traditional methods may take up to overnight, the ChIP Elute method takes only one hour. The DNA SMART ChIP-Seq Kit uses adapted SMART technology to add Illumina adaptors to ssDNA or dsDNA in a ligation-free manner. After addition of a T-tail to the DNA templates, the SMARTScribe™ Reverse Transcriptase (RT) copies the DNA strand using a proprietary DNA SMART Poly(dA) Primer. When the SMARTScribe RT reaches the 5′ end of the template, its terminal transferase activity adds a few additional nucleotides to the newly synthesized DNA. This allows the DNA SMART Oligonucleotide to bind and provide an extended template for the RT. Finally, the ChIP-seq library is amplified using PCR primers containing Illumina indexes, and is then purified and size selected.

2 Fast DNA elution and cross-linking reversal

Table I: Sequence Metrics from Different Cross-Linking Reversal Methods

ChIP antibody	CTCF			
Input amount (ng)	1		0.25	
No. of PCR cycles	13		16	
Elution/cross-linking reversal method	ChIP Elute Kit	Traditional	ChIP Elute Kit	Traditional
Library yield (nM)	21.4	16.4	37.3	28.1
Mapped to genome (%)	89.7	90.2	86.4	87.3
Mapped uniquely to genome (%)	71.6	73.2	70.2	71.1
Useful reads (uniquely mapped, non-duplicates; %)	66.0	67.2	58.1	59.0
Non-redundant rate	0.92	0.96	0.83	0.83
Peaks identified	37,610	36,924	39,525	37,134

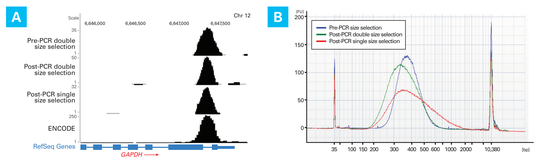


Libraries generated from ChIP DNA eluted with the ChIP Elute Kit or traditional methods are comparable. Libraries were created from ChIP DNA eluted with either the ChIP Elute Kit or traditional methods. Sequence analysis was performed using 12–16 million reads per sample (**Table I**). The number and size of peaks was very similar between the two methods and was well matched to ENCODE data (**Panel A**). There was high overlap in the number of peaks identified using different ChIP DNA input amounts, and the overlap was also similar between the two elution methods (90% for 1 ng inputs, 89% for 0.25 ng inputs).

3 Improved library complexity by eliminating pre-PCR size selection

Table II: Sequencing Metrics Comparing Pre- and Post-PCR Size Selection

ChIP antibody	CTCF		
Size selection	pre-PCR	post-PCR	post-PCR (single selection)
Library yield (nM)	10.9	17.2	26.9
No. uniquely mapping reads (millions)	5.1	5.9	4.3
Non-redundant rate	0.75	0.85	0.89
No. of peaks identified with 3.85 M reads (uniquely mapped, non-duplicates)	32,827	34,011	33,398
Number of overlapping peaks	28,469 (87%)		
		28,924 (85%)	

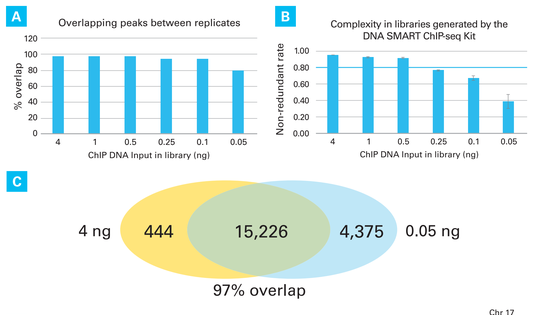


Performing library size selection after PCR results in better yield and complexity. Libraries were generated from 200 pg of the same input ChIP DNA, eluted with the traditional method, with size selection before or after library amplification (16 cycles of PCR). Size selection (**Panel B**) was performed either in two steps (eliminating both small and large inserts) or in a single selection (only removing small inserts such as primer dimers). Both library complexity (non-redundant rate) and yield were improved with post-PCR size selection. The location and shape of the peaks identified using post-PCR size selection still matched reported ENCODE data (**Panel A**), while library quality was improved. The number of overlapping peaks was between 85–87% across all libraries (**Table II**).

4 Excellent ChIP-seq library complexity and reproducibility from low inputs

Table III: Sequencing Metrics from Various Amounts of Input ChIP DNA

ChIP antibody	H3K4me3					
Input amount (ng)	4	1	0.5	0.25	0.1	0.05
No. of PCR cycles	12	13	14	15	17	18
Library yield (nM)	44.5	19.2	32.2	12.0	24.3	14.3
Mapped to genome (%)	91.9	91.0	91.0	91.1	90.7	90.2
Mapped uniquely to genome (%)	72.8	72.5	72.3	72.7	72.1	71.8
Useful reads (uniquely mapped, non-duplicates; %)	68.2	64.4	64.3	50.3	45.0	23.8
Peaks identified	16,738	16,811	16,366	17,277	16,584	19,601

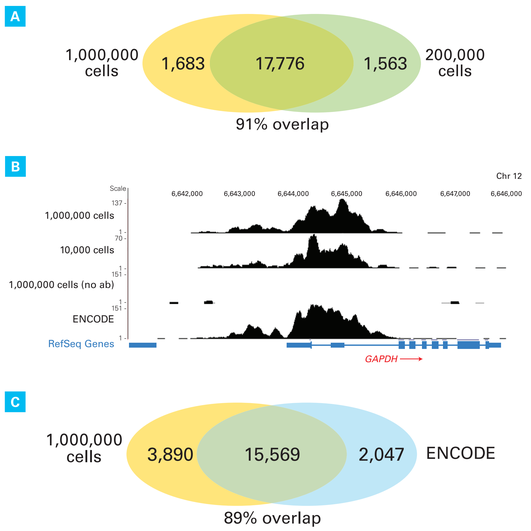


ChIP-seq library complexity and reproducibility is maintained across input amounts. ChIP-seq libraries were generated from different amounts of the same starting material using the traditional elution method. Sequence analysis was performed using 19–24 million reads per sample (**Table III**). The reproducibility ranged from 80–98% (**Panel A**). The non-redundant rate (normalized for 10 million uniquely mapped reads) was well above the standard recommended by the ENCODE project (0.8; blue line) for inputs ≥0.5 ng (**Panel B**; error bars indicate the standard deviation of two technical replicates). 94–97% of the peaks identified in the library generated from 4 ng of ChIP DNA were also identified in lower-input libraries (**Panel C**). The quality of the data (the shape and location of the peaks) was similar across input levels, and was well matched to ENCODE data, even for as little as 50 pg of input DNA (**Panel D**).

5 Excellent sensitivity from ChIP experiments using low cell numbers

Table IV: Sequencing Metrics from Specified Numbers of Cells

ChIP antibody	H3K4me3			
Input amount (cells)	1,000,000	200,000	50,000	10,000
No. of PCR cycles	15	18	18	18
Library yield (nM)	86.7	101	44.6	20.5
Mapped to genome (%)	92.7	88.6	84.3	75.8
Mapped uniquely to genome (%)	79.0	74.8	70.4	59.6
Useful reads (uniquely mapped, non-duplicates; %)	66.8	63.5	49.9	34.0
Non-redundant rate	0.85	0.85	0.71	0.57
Peaks identified	19,459	19,339	18,549	22,564



The DNA SMART ChIP-Seq Kit generates high-quality libraries from low cell number ChIP experiments. ChIP was performed with an anti-H3K4me3 antibody using various cells inputs and ChIP DNA was eluted with the ChIP Elute Kit. The entire amount of DNA obtained was used for library preparation. Sequence analysis was performed using 14–19 million reads per sample. Mapping statistics were very good across all input levels (**Table IV**). Peaks identified from ChIP experiments using few cells showed at least 86% overlap with peaks identified in the 1,000,000 cell experiment (**Panel A**). In addition, the peaks were of similar shape across cell inputs and matched the peaks obtained by the ENCODE project (**Panel B**). The number of peaks identified from the 1,000,000 cell experiment was very similar to those identified by the ENCODE data, indicating that the DNA SMART ChIP-seq libraries are robust and unbiased (**Panel C**).

Conclusions

The DNA SMART ChIP-Seq Kit generates sequencing libraries for Illumina platforms that maintain the complexity of the ChIP input sample, even when used with low input amounts. This sensitive, ligation-independent method expands the range of input samples that can be used for ChIP-seq experiments. It is particularly well-suited for ssDNA like that generated by the fast and efficient ChIP Elute Kit.

- **Efficient workflows**—The ChIP Elute Kit provides a fast DNA elution and cross-linking reversal method, and the DNA SMART ChIP Seq Kit adds Illumina indexes and adaptors at the same time as library amplification
- **Compatible with low-input samples**—Libraries can be generated from as little as 100 pg of fragmented input DNA
- **Preservation of library complexity**—Excellent non-redundant rate from as little as 250–500 pg input DNA or 50,000–200,000 cells (evaluated with ChIP DNA generated using anti-H3K4me3 or anti-CTCF antibodies)

To download this poster visit:
www.clontech.com/ABRF2015-ChIP-Seq