Clontech Laboratories, Inc.
A Takara Bio Company

Harnessing Template Switching Technology for ChIP-Seq Applications

Nathalie Bolduc, Simon Lee, Craig Betts, Andrew A. Farmer[1]

[1]Corresponding Author
Clontech Laboratories, Inc., 1290 Terra Bella Ave., Mountain View, CA 94043

Abstract

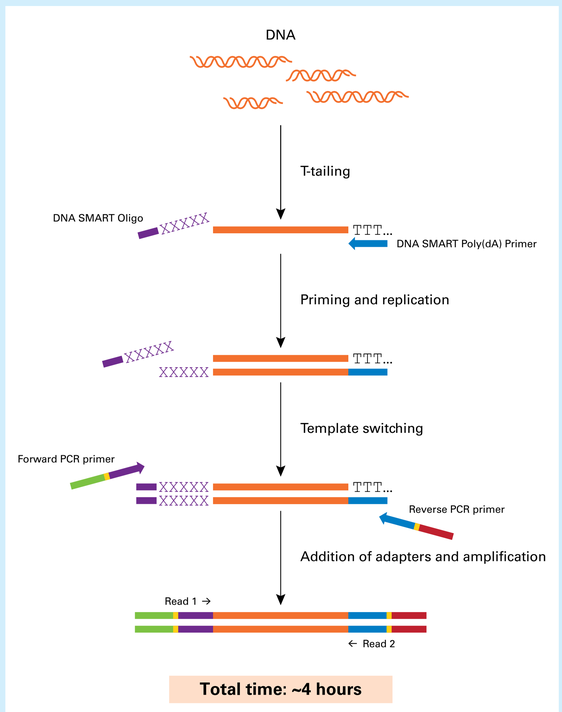
Preparation of sequencing libraries from DNA generated by chromatin immunoprecipitation (ChIP) is inherently challenging due to the minute quantities of DNA immunoprecipitated. Additionally, several protocols that allow for faster and more sensitive ChIP DNA elution and crosslinking reversal, such as the Chelex method, are prohibited due to the requirement of double-stranded (ds) DNA for traditional adapter ligation. We have overcome both of these limitations by adapting template switching technology to ChIP-seq library production.

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 on RNA samples. Here, we report a modified template switching reaction that accepts DNA instead of RNA as a template. This new technology accommodates either single-stranded (ss) DNA or dsDNA templates, including ChIP DNA samples eluted in the presence of Chelex resin, thus making it very amenable to applications such as ChIP-seq. We have assembled a kit allowing the use of DNA template switching for sequencing on Illumina® platforms. The kit features an easy, single-tube workflow that can be completed in 4 hours and is designed to work with between 100 pg and 10 ng of dsDNA or ssDNA. To validate the technology for ChIP-seq, we used 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 over 86% overlap between the peaks identified when comparing H3K4me3 pull-down 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%, thus meeting the standards established by the ENCODE consortium. These data indicate that our DNA template switching technology provides a robust and reliable tool for ChIP-seq applications, particularly at low inputs.

Introduction

The DNA SMART™ ChIP-Seq Kit is based on an adaptation of SMART® template switching technology for use with DNA. 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, it is then purified and size selected. Library generation can be completed in approximately 4 hours.

1 DNA SMART ChIP-Seq Technology



Total time: ~4 hours

Materials and Methods

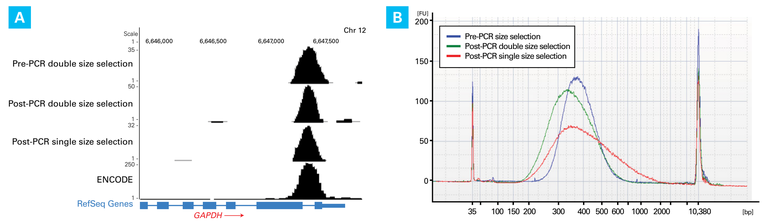
For ChIP assays, HEK 293T cells were grown to 80% confluence and fixed with 1% formaldehyde for 10 minutes. ChIP was performed with ChIP-grade anti-H3K4me3 or anti-CTCF antibodies according to standard methods (chromatin shearing by sonication with Bioruptor Pico; Diagenode). DNA was purified with a Macherey-Nagel NucleoSpin Gel and PCR Clean-Up kit. Sequencing libraries were generated using the DNA SMART ChIP-Seq Kit and size selection was performed with AMPure XP beads (Beckman Coulter) using Option 1 or Option 4 as described in the DNA SMART ChIP-Seq Kit User Manual. Sequencing was carried out on Illumina MiSeq® or HiSeq® 2500 instruments. All runs were 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 was set at 1 x 10^{-7}). Raw data generated by the ENCODE consortium 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. Reads and peaks were visualized using IGV or the UCSC genome browser.

2 Improved library complexity by eliminating all pre-PCR size selection and clean-up

Table I: 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	5,119,363	5,908,181	4,302,276
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	28,039		28,039
		28,469	
		28,924	

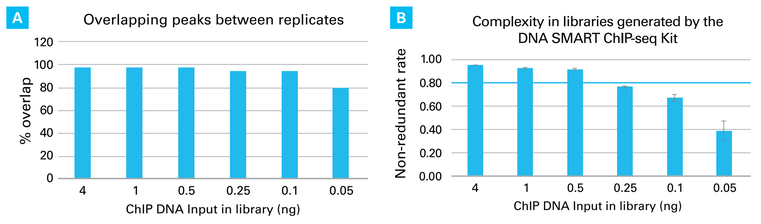


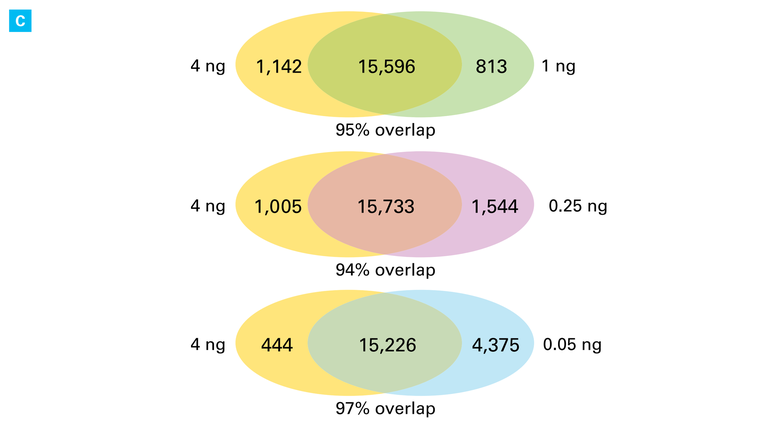
Performing library size selection after PCR results in better yield and complexity. ChIP-seq libraries were generated from 200 pg of the same input ChIP DNA with size selection before or after library amplification (16 cycles of PCR). Library size selection (**Panel B**) was performed either in two steps (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 (**Table I**).

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

Table II: Sequencing metrics from various amounts of input 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
Total no. reads (millions)	19.94	21.63	21.64	20.28	18.97	24.53
% reads mapped	91.9	91.0	91.0	91.1	90.7	90.2
No. of uniquely mapped reads (millions)	14.52	15.68	15.64	14.74	13.67	17.56
No. of unique reads without duplicates (millions)	13.61	13.93	13.91	10.20	8.53	5.83
% useful reads (uniquely mapped, non-duplicates)	68.2	64.4	64.3	50.3	45.0	23.8
No. of peaks identified	16,738	16,811	16,366	17,277	16,584	19,601



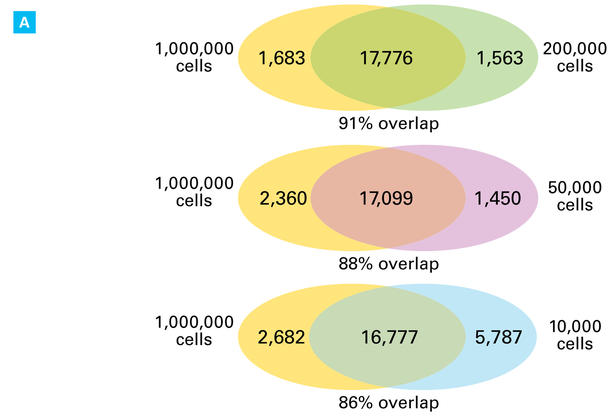


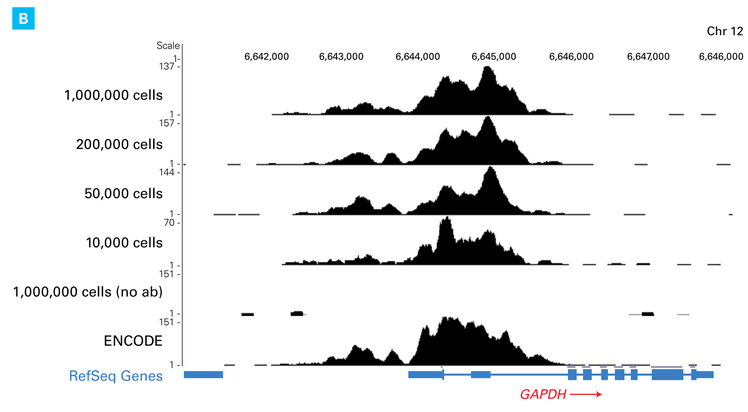
ChIP-seq library complexity and reproducibility is maintained across input amounts. ChIP-seq libraries were generated from different amounts of the same starting material (**Table II**). The reproducibility (i.e., the overlap in the number of peaks identified between replicates from the same input amount) 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) 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 ChIP DNA were also identified in lower input libraries (**Panel C**). The quality of the data (as indicated by 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 input DNA (**Panel D**).

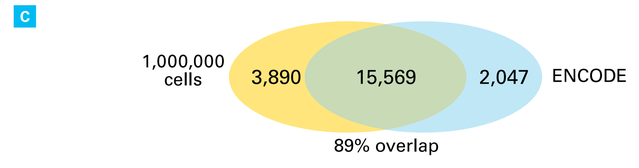
4 Excellent sensitivity from ChIP experiments using low cell numbers

Table III: Sequencing metrics from total DNA from specified numbers of cells

ChIP Antibody	H3K4me3			
Input amount (**millions of cells**)	1	0.2	0.05	0.01
No. of PCR cycles	15	18	18	18
Library yield (nM)	86.7	101	44.6	20.5
Total no. reads (millions)	14.92	17.24	18.45	17.97
% reads mapped	92.7	86.6	84.3	75.8
No. of uniquely mapped reads (millions)	11.79	12.90	12.99	10.71
No. of unique reads without duplicates (millions)	9.97	10.95	9.20	6.11
% useful reads (uniquely mapped, non-duplicates)	66.81	63.50	49.87	34.00
Non-redundant rate	0.85	0.85	0.71	0.57
No. of 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. The entire amount of DNA obtained was used for library preparation. Mapping statistics were very good across all input levels (**Table III**). 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 peaks identified from the 1,000,000 cell experiment highly overlapped with those from the ENCODE data (293 cells, anti-H3K4me3 antibody, U. Washington), 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, low amounts of starting material, or other samples that are incompatible with ligation-based library preparation methods. The efficient single-tube protocol, combined with post-PCR library size selection, allows for the generation of ChIP-seq libraries from low-input DNA samples in about 4 hours.

- **Single-tube workflow**—Illumina indexes are added and the library is amplified in a simple protocol; only a single, final library clean-up step is necessary
- **ssDNA or dsDNA input**—Input DNA is denatured prior to amplification, allowing for the use of either ssDNA or dsDNA
- **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)