ChIP-Seq Library Prep

Thomas Westerling, PhD., Center for Functional Cancer Epigenetics, Dana Farber Cancer Institute, Boston, MA.

Introduction

ChIP-Seq analysis is valuable for identifying genomic DNA sequences that are associated with immunoprecipitable proteins, such as modified histones or transcription factors, in vivo. A typical ChIP-Seq library prep procedure includes these steps:

1. Cells are treated to elicit desired changes in a gene regulatory network.
2. Soluble chromatin is prepared through a number alternate methods e.g. acoustic shearing, or enzymatic treatment from fixed (typically formaldehyde) or unfixed cells.
3. The soluble fragmented chromatin is immunoprecipitated using an antibody specific for the protein of interest (e.g. a specific transcription factor or a modified histone).
4. The DNA is processed to generate an adapter ligated NGS library.
5. The library is size-selected. Traditionally performed on standard agarose gels.
6. The size-fractionated library is enriched by PCR and size-selected again.
7. The sequences produced are analyzed and aligned to a reference genome in order map the sites of protein occupancy, and changes in occupancy associated with the change in gene expression.

One of the technically challenging and labor-intensive steps is the size selection of DNA fragments during the NGS library construction and is a source of sample loss and sample cross contamination. This tech note is a comparison of two alternative methods for the size-selection step: Pippin Prep™ automated DNA sizing and E-gel® SizeSelect™ gels, for the purpose of determining the suitability of the Pippin Prep system for NGS library construction using limited amounts of ChIP DNA starting material.

Comparison of Pippin Prep and E-gel SizeSelect in ChIP-seq protocols for study of a transcription factor

In this comparison, MCF7 cells were stimulated with estradiol, followed by shearing by sonication and immunopreciptation using antibodies against estrogen receptor alpha (ER). The input DNA (75 ng) (serving as a genomic background control) and DNA recovered from ChIP (3 ng) were processed for NGS library construction, divided in equal parts and subjected to size-selection using each of the two methods. After PCR-amplification, and a second size selection step using the respective methods the quality and yield were checked with agarose gels and Agilent BioAnalyzer® runs.

The DNA size selection in this trial was intended to collect a broad size range (180 bp to 400 bp) able to capture the whole range of DNA fragments present in the ChIP sample, while still excluding primer products and high molecular weight DNA. This is an efficient way to maintain the complexity of your final library resulting in more usable sequencing data. In addition it allows for the visualization of the DNA smear post PCR and reduces the chance of excluding part of your sample due to blinded size selection. Briefly, after PCR 1 µl is visualized on any analytical instrument the desired size range can be determined and programmed into the Pippin Prep. For established ChIP NGS library preparations known default values could be used to save time and effort. For both input DNA and ChIP DNA, the Pippin Prep collected DNA fragments across the entire size range with high yield and tight high- and low-end cutoffs (Figure 1).

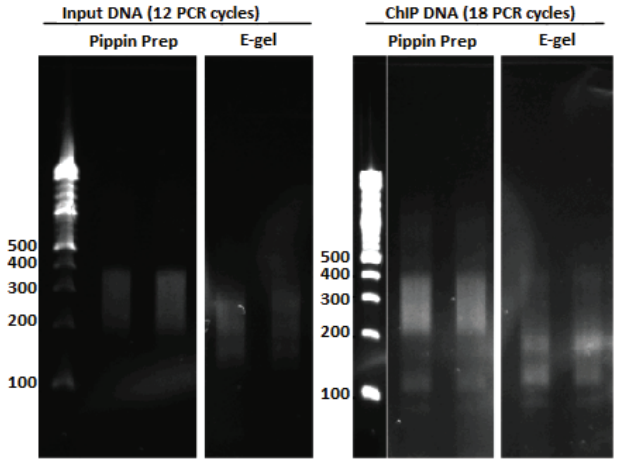
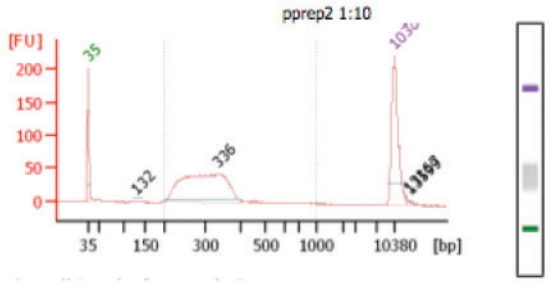


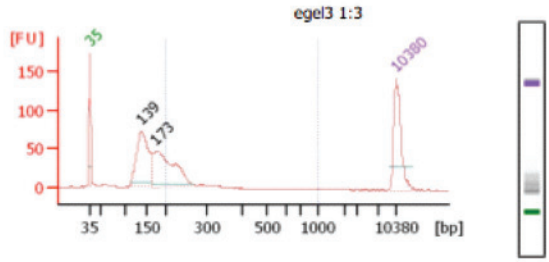
Figure 1. Direct comparison of input and ChIP DNA libraries selected for a 180 bp – 400 bp size range using the Pippin Prep and E-gel systems. Duplicate PCR reactions were loaded side-by-side.

In an effort to collect the full 180 bp to 400 bp size range using the E-gel system, 3 successive fractions were collected while the gel was running (This routine can be employed to create a redundant back up NGS library if the fractions are not pooled). These are visible as 3 bands on the gel. Both input DNA and ChIP DNA were efficiently collected, however the manual involvement and difficulties in accurate assessment of actual size

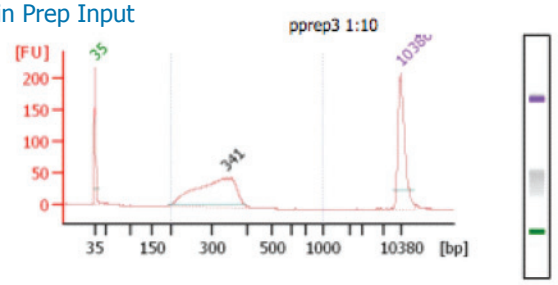
Pippin Prep ChIP



Egel ChIP



Pippin Prep Input



Egel Input

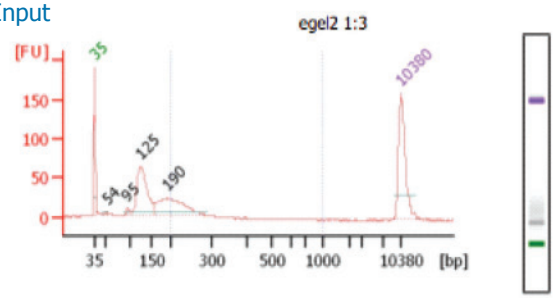


Figure 2. BioAnalyzer profiles confirmed the high yield across the full size range for the libraries size-selected with the Pippin Prep. For the libraries size selected with the E-gel, the profiles show loss of larger fragments and the presence of contaminating primer-dimers.

provides a source of inconsistency that here is seen as a lack of fragments in the upper end of the size range using E-gels (Figure 1). In addition the collection of slightly lower than intended fragment sizes apparently contributed to significant primer product carryover and low yield and quality of the library.

The BioAnalyzer tracings confirm that the Pippin Prep size selection selected more accurate broad size ranges with good yield across the full size range selected (Figure 2, left panels). In contrast, the BioAnalyzer tracing for the E-gel run confirmed a lower than expected size range, with a marked absence of above 300bp fragments in the ChIP sample. In addition, the analysis revealed that smaller primer-dimers and adaptor-dimers were not adequately separated from the amplified ChIP DNA (Figure 2, right panels).

Preservation of ChIP enrichment

Quantitative PCR was used to validate the enrichment for known ER binding sites (Figure 3). The ChIP libraries prepared using the Pippin Prep (blue bars) were found to be enriched 1.5 to 5-fold for specific gene sequences relative to the original input DNA. These enrichment levels are typical of those seen in ChIP procedures. The samples prepared using E-gels had poor agreement to the original sample, with generally lower or negligible enrichment and would not be usable for sequencing.

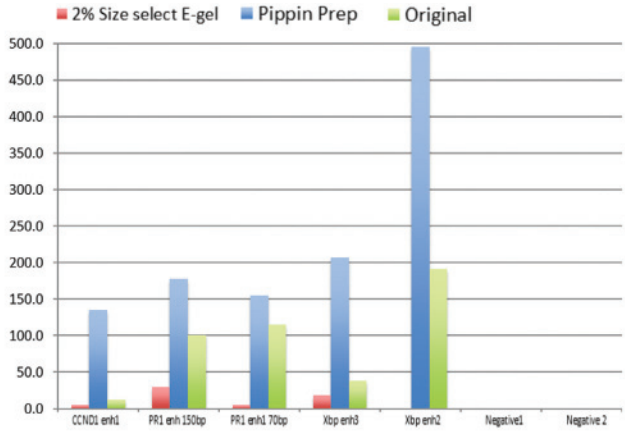


Figure 3. PCR amplified ChIP libraries prepared using the Pippin Prep (blue bars) show good agreement of enrichment for ER binding sites compared to the original starting DNA (green bars).

High quality sequence reads from limited ChIP DNA

The DNA yield of the Pippin Prep size selection enabled a high yield of unique reads to be obtained despite the very limited amounts of ChIP DNA available. In similar independent samples high yields of sequencing data has been generated from <2ng to 60ng of starting ChIP material. A representative figure of such an experiment (5ng and 10ng ChIP and 50ng Input DNA starting material) centered on the ER regulated Cyclin D1 encoding gene is provided. The overall genomic background has a smooth and even profile, whereas low and high binding ER ChiP-seq samples can be clearly

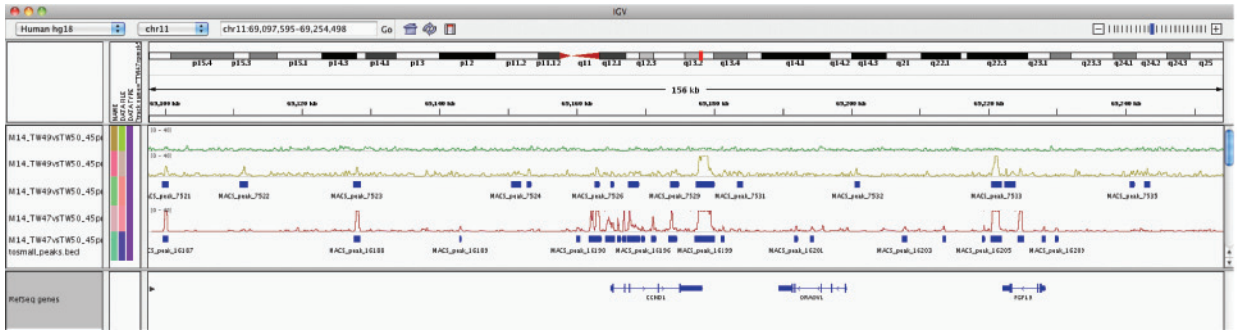


Figure 4. Comparison of low and high binding ER ChIP-seq samples using MACS peak calling software

differentiated (shifted tag count output generated by the MACS peak calling software (Zhang et al., 2008a) visualized in Integrated Genome Browser, Nicol et al., 2009) (Figure 4: Input, 'low' ER, and 'high' ER).

Comparison of Pippin Prep and E-gel SizeSelect in ChIP-seq protocols for study of histone variants

Micrococcal nuclease-digested chromatin from unfixed OV2008 cells was immunoprecipitated using an Ab against histone H3 bearing a dimethylated lysine4 residue (H3K4Me2). ~10ng of DNA recovered from this procedure was ligated to adaptors, size selected using either a Pippin Prep or an E-gel system, and then PCR-amplified and analyzed for quality and yield. The limited amount of DNA obtained from a typical ChIP procedure is difficult to visualize accurately during the size selection step so one must rely on DNA size standards and predictable performance of the system for good results.

The target sizes for mono- and di-nucleosome DNA (including adaptors) were ~180 bp and ~330 bp, respectively. For the Pippin Prep, automatic size collection was set to collect the mono-nucleosome band (with a size range of 165 bp to 210 bp). Figure 5 shows that the manual targeting of the E-gel size selection to 180 bp largely missed the mono-nucleosome band (lane 1), although the di-nucleosome band was recovered (lane 2). The Pippin Prep recovered the targeted mono-nucleosome-sized DNA effectively (lane 3).

In three independent ChIP samples (anti-histone H3K-4Me2) prepared using the Pippin Prep, single-end read runs on the Illumina HiSeq2000 instrument consistently gave close to 90M (90 million) total reads per run, with between 68M and 80M of these reads being uniquely mapped. Downstream genomewide NPS (nucleosome positioning software (Zhang et al., 2008b) analysis identified ~300,000 uniquely positioned nucleosomes.

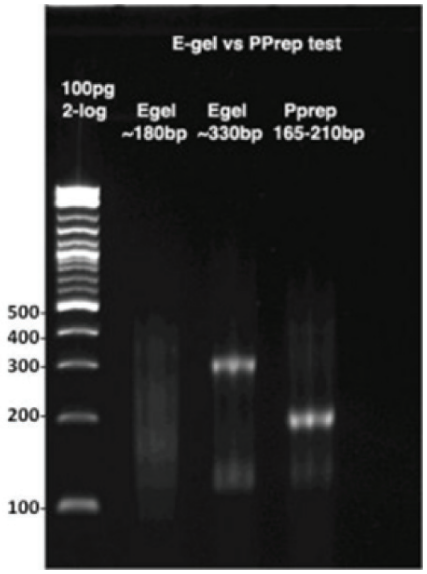


Figure 5. Size selection of a ChIP DNA sample (immunoprecipitated with anti- H3K4Me2) with either the E-gel SizeSelect or Pippin Prep system.

Conclusion

The PippinPrep is an excellent platform for the accurate and reproducible size selection process in a variety of NGS library applications especially low starting material (ng scale) ChIP-seq library generation. It provides a flexible, controllable, programmable and easy to use low labor alternative that removes a significant bottleneck in NGS library creation.

References

Nicol, J. W., Helt, G. A., Blanchard, S. G., Jr., Raja, A., and Loraine, A. E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25, 2730-2731.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008a). Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137.

Zhang, Y., Shin, H., Song, J. S., Lei, Y., and Liu, X. S. (2008b). Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. BMC Genomics 9, 537.

sage science

Sage Science, Inc. Suite 3150 • 500 Cummings Center • Beverly, MA 01915
support@sagescience.com • www.sagescience.com
888.744.2244 (U.S) • 978.922.1832 (outside North America)