

TECHNICAL NOTE

Open Access

A method for generating highly multiplexed ChIP-seq libraries

Ethan Ford[†], Chrysa Nikopoulou[†], Antonis Kokkalis[†] and Dimitris Thanos^{*}

Abstract

Background: The barcoding of next generation sequencing libraries has become an essential part of the experimental design. Barcoding not only allows the sequencing of more than one sample per lane, but also reduces technical bias. However, current barcoding strategies impose significant limitations and/or technical barriers in their implementation for ChIP-sequencing.

Findings: Converting Y-shaped sequencing adapters to double stranded DNA prior to agarose gel size selection reduces adapter dimer contamination and quantitating the number of cycles required for amplification of the library with qPCR prior to library amplification eliminates library over-amplification.

Conclusions: We describe an efficient and cost effective method for making barcoded ChIP-seq libraries for sequencing on the Illumina platform.

Keywords: Chromatin immunoprecipitation, ChIP-seq, Next generation sequencing, Genomics, Epigenetics

Findings

Chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) is a powerful method to identify the genome-wide binding profiles of chromatin-associated proteins and has emerged as one of the most important tools used to study transcriptional regulation, DNA replication, DNA recombination and chromatin structure [1].

Technological advancements in next-generation sequencing throughput have facilitated the production of more sequencing data per lane on the Illumina platform than what is necessary to achieve saturation in a ChIP-seq experiment. Thus, in order to perform experiments in cost effective manner, multiple barcoded ChIP-seq libraries must be pooled together and sequenced in a single lane. In addition, pooling barcoded samples from the same or different experiments reduces technical variability between samples [2]. Thus, barcoding not only reduces costs, it also produces higher confidence and quality data. Unfortunately, current barcoding strategies impose significant limitations and/or technical barriers to their implementation for ChIP-seq analysis.

There are currently two main barcoding strategies for the generation of NGS (next generation sequencing) libraries for sequencing on the Illumina platform. The first embeds the barcode in the adapter oligonucleotide used for library construction so that the first nucleotides sequenced corresponds to the barcode sequence [3]. On the Illumina platform, it is critical that the four bases (A, T, G and C) are represented in roughly equal proportions in the first nucleotides sequenced [4]. Thus, it is necessary to pool libraries in multiples of four so that the nucleotide composition is balanced. However, this causes inflexibility in that only multiples of four samples must be used, and also limits the researcher from pooling libraries in unequal ratios if more reads are required from one sample over another. In addition, the invariant thymidine residue at the 3' end of the oligonucleotide adapters, which is required for ligation to the immunoprecipitated DNA, can have an adverse effect on the quality of the sequences generated. To address these issues it is usually recommended that libraries are spiked with Illumina's PhiX control library and/or sequenced at a lower density [5]. However, these methods reduce sequencing throughput significantly.

Because of the difficulties associated with the incorporation of the barcode at the beginning of the sequencing read, Illumina has implemented a second-read barcoding strategy as a centerpiece of its TruSeq technology. The

* Correspondence: thanos@bioacademy.gr

[†]Equal contributors

Biomedical Research Foundation, Academy of Athens, 4 Soranou Efessiou Street, Athens 11527, Greece

second read strategy circumvents the problems associated with having the barcode embedded at the beginning of the sequencing read, however current adapter design makes it difficult to implement this strategy using DNA from ChIP-seq experiments. Specifically, during library preparation adapter dimers are formed and must be removed from the library before sequencing. Size selection with AMPure XP beads is an efficient method to remove adapter dimers [5], but since ChIP libraries often contain DNA in the size range of 100 to 200 bp, the size difference between the adapter dimers and adapter-ligated immunoprecipitated DNA is difficult to resolve with this method. Size selection by agarose gel electrophoresis, in principle, can resolve the size difference between adapter-dimers and adapter-ligated immunoprecipitated DNA, but the large Y-shaped DNA adapter molecules required for second read barcoding run aberrantly through agarose gels.

Here, we describe a method that bypasses the problems associated with size selection. Specifically, compared to the two methods discussed, we perform five cycles of PCR prior to size selection on an agarose gel to convert the Y-shaped DNA to double-stranded DNA, so that both the adapter-dimers and adapter-ligated immunoprecipitated DNA run true to their size during agarose gel electrophoresis and accurate size selection can be performed. This step also results in a modest amplification of the immunoprecipitated DNA prior to agarose gel size selection, which increases the yield of an otherwise inefficient step in the library preparation protocol. In addition, we also include a method to accurately quantitate the number of cycles required for PCR amplification of the library, which reduces biases caused by over-amplification [6]. Library amplification is performed using Kapa HiFi polymerase, which is more efficient and produces less bias than Phusion

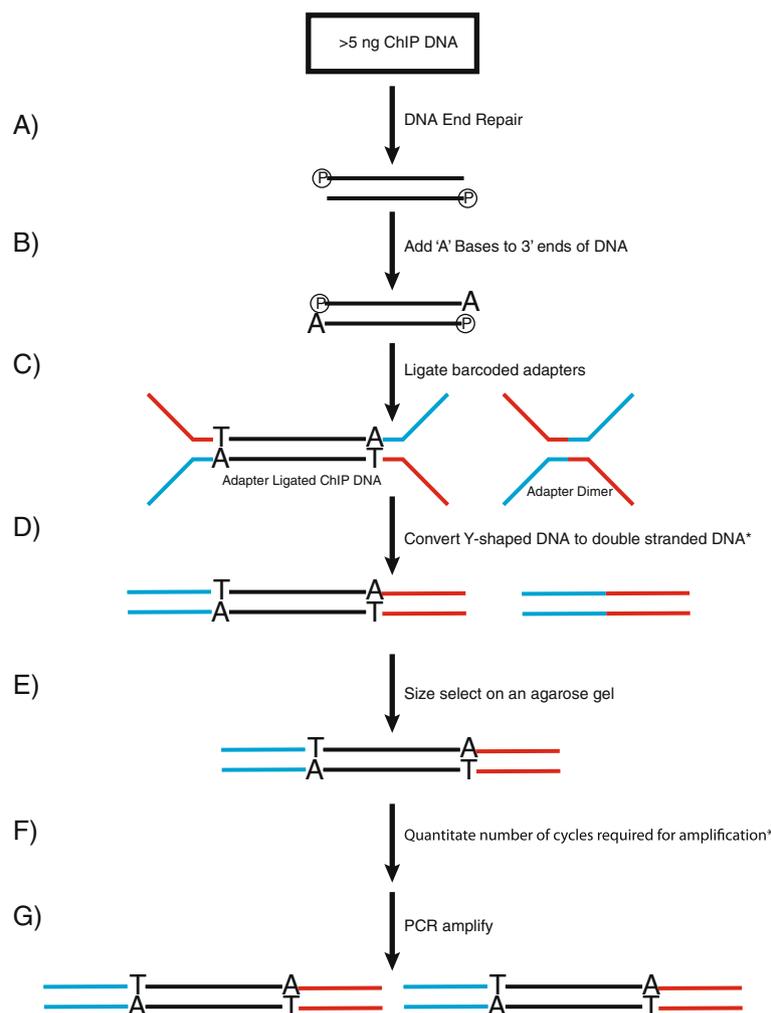
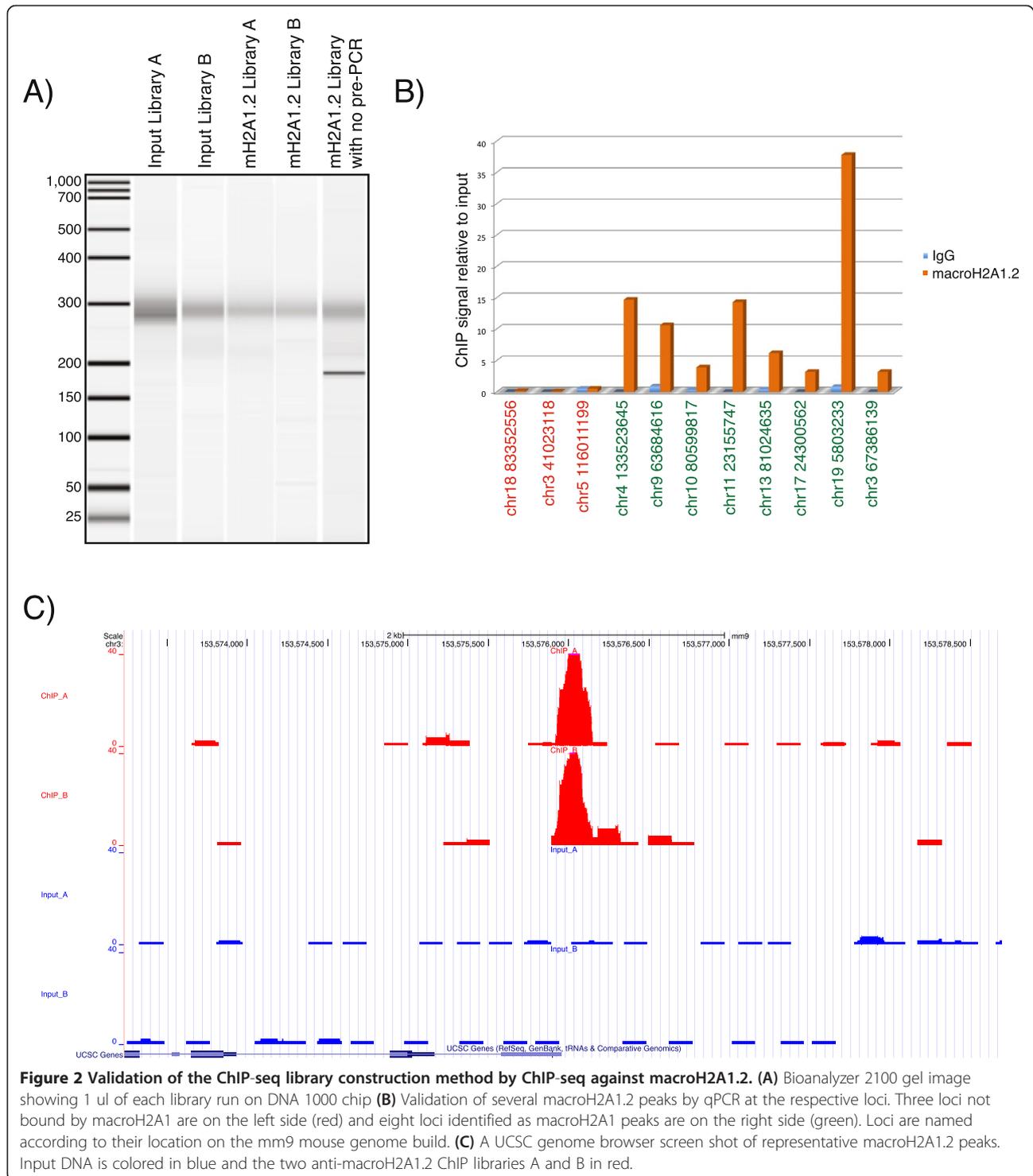


Figure 1 A schematic representation of the method for making barcoded ChIP-seq libraries. An asterisk denotes novel steps in the protocol. (A) End-repair. (B) A-tailing. (C) Adapter ligation. (D) Conversion of Y-shaped DNA to double-stranded DNA. (E) Size selection. (F) Cycle quantitation. (G) PCR amplification.

polymerase [7]. Finally, we have reduced the amount of reagents required for library construction making library construction more affordable, which is especially important as sequencing costs drop.

A schematic step-wise representation of the method is illustrated in Figure 1. Briefly, a minimum of 2 ng of immunoprecipitated DNA is treated with T4 DNA polymerase,

Klenow fragment and T4 polynucleotide kinase to blunt the DNA ends and add 3' phosphates (Figure 1A). Next, single 5' adenine overhangs are generated by incubating the DNA with Klenow exo-minus and dATP (Figure 1B). Barcoded Y-shaped oligonucleotide adapters are ligated to the DNA (Figure 1C). The DNA molecules are then converted to double-stranded DNA with five cycles of PCR



(Figure 1D), so that the DNA molecules run predictably during the subsequent size selection by agarose gel electrophoresis (Figure 1E). The precise number of PCR cycles required for amplification is determined by qPCR (Figure 1F) and finally the library is amplified by PCR accordingly (Figure 1G).

To demonstrate the robustness of our library preparation protocol we performed ChIP against the histone variant macroH2A1 in mouse embryonic fibroblasts (MEFs). MacroH2A1 is a unique histone variant, in that in addition to the canonical histone H2A domain, it also contains a relatively large (30 kD) amino-terminal 'macro' domain. While macroH2A1 is generally associated with gene repression, it is also required for the activation of a subset of genes [8]. The alternative use of a 64 bp or 73 bp mutually exclusive exon results in the production of two distinct proteins, macroH2A1.1 and macroH2A1.2, respectively [9]. While macroH2A1.1 efficiently binds the NAD⁺ metabolite ADP-ribose, macroH2A1.2 does not [10]. We prepared Illumina libraries from immunoprecipitated material using antibodies targeted to macroH2A1.2 as well as input DNA (for a detailed protocol see Additional file 1). Five ChIP-seq libraries were made, two from the macroH2A1.2 immunoprecipitated DNA (macroH2A1.2 library A and macroH2A1.2 library B), two from the input DNA (input library A and input library B) and one from input DNA in which the pre-agarose gel PCR step was omitted. The libraries were amplified with 12 cycles of PCR (including the pre-size selection PCR cycles) and visualized on the Agilent Bioanalyzer (Figure 2A) before being sequenced in one lane of the HiSeq2000. In total, 76 million reads were obtained and demultiplexed with Illumina's Casava data analysis pipeline. 24 million, 15 million, 21 million and 15 million reads for macroH2A1.2 library A, macroH2A1.2 library B, input library A and input library B respectively, were mapped to the mouse genome build mm9 with the Bowtie short read alignment software. MacroH2A1.2 enriched regions were identified with the MACS software package [11] and the data sets were uploaded onto the UCSC genome browser. In total we identified 33 peaks in macroH2A1.2 library A. Every peak identified in macroH2A1.2 library A was also a peak in macroH2A1.2 library B, demonstrating the reproducibility of the library construction protocol. An example peak is shown in Figure 2C as a UCSC genome browser screenshot. A few of the identified peaks were validated by ChIP-qPCR (Figure 2B) further supporting the specificity and the effectiveness of the method. Adapter sequences represented less than 1% of sequences, which highlights the effectiveness and necessity to convert the Y-shaped adapters to double-stranded DNA prior to agarose gel size selection.

In conclusion, we present a robust and economical method for generating barcoded ChIP-seq libraries, which

has substantial advantages over current methodologies. We have successfully pooled between two and eight samples in a single lane of the HiSeq2000 and consider that any number of samples could be pooled without difficulties. The libraries are compatible with Illumina's TruSeq platform and can be demultiplexed by Illumina's Casava data analysis pipeline. Importantly, the two innovations detailed in this method, that is the conversion of the Y-shaped adapter DNA to double-stranded DNA prior to agarose gel size selection and the determination of the precise number of cycles to perform during PCR amplification can also be applied to other next-generation sequencing techniques such as MeDIP-seq.

Additional file

Additional file 1: Detailed protocol. Chromatin immunoprecipitation and multiplexed illumina library preparation protocol.

Abbreviations

ChIP: Chromatin immunoprecipitation; NGS: Next-generation sequencing; PCR: Polymerase chain reaction; MeDIP-seq: Methylated DNA immunoprecipitation.

Competing interests

The authors declare no competing interests.

Authors' contributions

EF developed the method, designed the experiments, analyzed the data and co-wrote the manuscript. CN and AK performed the experiments and co-wrote the paper. DT supervised the project and co-wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank S. Tsiftoglou, A. Banos and M. Lavigne for critical reading of the manuscript.

Role of the funding sources

This work was supported by a FP6 mobility Marie Curie International Incoming Fellowship grant to EF, a Cooperation grant from the Greek Secretariat for Research and Technology (EDGE) and a KMW offsets grant to DT.

Received: 17 February 2014 Accepted: 7 May 2014

Published: 22 May 2014

References

1. Park PJ: ChIP-seq: advantages and challenges of a maturing technology. *Nature reviews. Genetics* 2009, **10**:669–680.
2. Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N, Huentelman MJ: Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods* 2008, **5**:887–893.
3. Lefrançois P, Euskirchen GM, Auerbach RK, Rozowsky J, Gibson T, Yellman CM, Gerstein M, Snyder M: Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing. *BMC Genomics* 2009, **10**:37.
4. Krueger F, Andrews SR, Osborne CS: Large scale loss of data in low-diversity illumina sequencing libraries can be recovered by deferred cluster calling. *PLoS One* 2011, **6**:e16607.
5. Wang L, Si Y, Dedow LK, Shao Y, Liu P, Brutnell TP: A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq. *PLoS One* 2011, **6**:e26426.
6. Mathieu-Daude F, Welsh J, Vogt T, McClelland M: DNA rehybridization during PCR: the 'Cot effect' and its consequences. *Nucleic Acids Res* 1996, **24**:2080–2086.

7. Quail MA, Otto TD, Gu Y, Harris SR, Skelly TF, McQuillan JA, Swerdlow HP, Oyola SO: **Optimal enzymes for amplifying sequencing libraries.** *Nat Methods* 2012, **9**:10–11.
8. Gamble MJ, Frizzell KM, Yang C, Krishnakumar R, Kraus WL: **The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing.** *Genes Dev* 2010, **24**:21–32.
9. Pehrson JR, Costanzi C, Dharia C: **Developmental and tissue expression patterns of histone macroH2A1 subtypes.** *J Cell Biochem* 1997, **65**:107–113.
10. Kustatscher G, Hothorn M, Pugieux C, Scheffzek K, Ladurner AG: **Splicing regulates NAD metabolite binding to histone macroH2A.** *Nat Struct Mol Biol* 2005, **12**:624–625.
11. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based Analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9**:R137.

doi:10.1186/1756-0500-7-312

Cite this article as: Ford et al.: A method for generating highly multiplexed ChIP-seq libraries. *BMC Research Notes* 2014 **7**:312.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

