Application Note

Constructing ChIP-Seq libraries from ultra-low DNA inputs

Authors

Jennifer Pavlica
Applications Manager

Rachel W. Kasinskas
Director, Support & Applications

Roche Sequencing & Life Science
Wilmington, MA, USA

Christopher D. Scharer
Assistant Professor, Department of Microbiology and Immunology

Jeremy M. Boss
Professor and Chair of the Department of Microbiology and Immunology

Emory University
Atlanta, GA, USA

Date of first publication: 2016

KAPA HyperPrep: A streamlined solution for the construction of ChIP-Seq libraries from picogram amounts of DNA

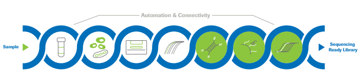
ChIP-Seq is a powerful tool for identifying the binding sites of DNA-associated proteins and covalent histone modifications. The KAPA HyperPrep Kit offers a streamlined and effective solution for the construction of ChIP-Seq NGS libraries from ultra-low DNA inputs. The results shown here demonstrate successful library construction down to 10 pg of input material, opening the door to the interrogation of scarce samples, such as primary cells.

Introduction

ChIP-Seq, chromatin immunoprecipitation followed by next-generation sequencing (NGS), is a valuable method enabling the study of DNA interactions with proteins, such as transcription factors and histone modifications. In ChIP-Seq workflows, proteins bound to DNA fragments of interest are enriched through formaldehyde crosslinking and targeted antibody selection, commonly known as immunoprecipitation. The enriched DNA is then purified and used as input into NGS library construction in preparation for sequencing.

One challenge of ChIP-Seq—depending on the cell population and protein of interest—is the ultra-low yield resulting from immunoprecipitation, which is in some cases less than 100 pg. This issue is often compounded by the fact that ChIP-Seq DNA inputs can be broad and bimodal in size, necessitating size selection, resulting in additional sample loss.

The KAPA HyperPrep Kit offers a streamlined solution for studying these ultra-low-input samples, enabling library construction with inputs as low as 10 pg. The single-tube workflow, along with highly optimized reaction chemistries, converts an increased percentage of input DNA into sequenceable adapter-ligated molecules. Additionally, the use of the KAPA HiFi HotStart Polymerase—engineered for high-efficiency and high-fidelity library amplification—results in improved sequence coverage and reduced bias.



Methods and experimental design

Using Raji, a human Burkitt's lymphoma cell line, six parallel immunoprecipitations were performed in which 10^7 cells were crosslinked with 1% formaldehyde for 10 minutes and sonicated to target an average size of 200 – 600 bp. Immunoprecipitation was performed overnight at 4°C with 5 µg of antibody specific for the histone modification H3K4^{me3} prebound to Protein G Dynal® beads (ThermoFisher). Enrichments for each individual immunoprecipitation were determined by qPCR, using primer sets targeting the following loci: positive control GAPDH (where H3K4^{me3} is expected) and negative control TSH2B (where H3K4^{me3} is not expected). All immunoprecipitated DNA was pooled for downstream manipulations and quantified using a Qubit™ fluorometer (ThermoFisher).

When working with ultra-low inputs, optimization of the library construction process is critical for success. Using the pooled material, libraries were constructed using the KAPA HyperPrep Kit with DNA inputs of 1 ng, 100 pg and 10 pg. A 1 ng input amount was used as a control, as this falls within the validated input range of the kit. Adapter titrations were performed for each DNA input, and the assessed stock concentrations are outlined in Table 1. Post ligation yields were measured by qPCR using the KAPA Library Quantification Kit. The numbers of amplification cycles were chosen to target a 10 nM final library concentration, taking into account the anticipated material loss during post-amplification size selection. The number of amplification cycles utilized were 11, 16 and 19 cycles for 1 ng, 100 pg and 10 pg, respectively.

Table 1. Adapter stock concentrations used for the adapter:input ratio titration for different input amounts

	1 ng	100 pg	10 pg
Adapter Input 1	300 nM*	30 nM	15 nM
Adapter Input 2	600 nM	150 nM	60 nM
Adapter Input 3	1.5 µM	300 nM	150 nM
Adapter Input 4	NA	600 nM	600 nM

*For 1 ng inputs, 300 nM is the validated adapter stock concentration, as recommended in the **KAPA HyperPrep Kit Technical Data Sheet** (KR0961 – v6.17 or later). NA=not assesed.

Optimal adapter concentrations were identified as those which maximized the post-amplification yield without exhibiting adapter dimerization greater than 10% on a molar basis. In comparison to other DNA-Seq applications, a greater threshold for adapter dimerization was allowed post-amplification due to the need for an AMPure® XP (Beckman Coulter) 0.6 – 0.8X dual-SPRI® size selection prior to sequencing.

Libraries with optimal adapter concentrations were replicated to provide reproducible data. Distributions of both post-amplification and size-selected libraries were assessed with the Bioanalzyer High-Sensitivity DNA Kit (Agilent), and concentrations were measured using the KAPA Library Quantification Kit. All six libraries were pooled and sequenced using 2 x 50 bp chemistry on a HiSeq® 2500. Data was analyzed using Bowtie and HOMER.

Results and discussion

Library yields post-amplification, but prior to size selection, are shown in Figure 1 for the adapter titration experiment. For all DNA inputs, library yield increased with increasing adapter concentration. When viewing these results in conjunction with the post-amplification Bioanalyzer traces in Figure 2, the following were identified as optimal adapter stock concentrations: 1.5 µM, 600 nM and 150 nM for 1 ng, 100 pg and 10 pg, respectively. As a note, at 10 pg of input, the adapter concentration that resulted in the highest final library yield also resulted in an increased level of adapter-dimer. At this input amount, the selected adapter concentration reflects a compromise between yield and the formation of adapter-dimers, and the presence of 9% adapter-dimer was considered acceptable.

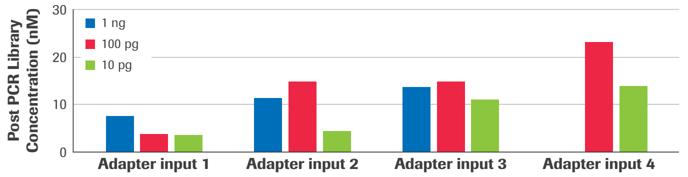


Figure 1. Post-amplification library concentrations (prior to size selection), as assessed by qPCR with the KAPA Library Quantification Kit, for 1 ng, 100 pg and 10 pg inputs.

Bioanalyzer traces for libraries that have undergone dual-SPRI size selection are shown in Figure 3. This illustrates the effectiveness of the dual-SPRI strategy for size distribution tuning, as well as its ability to eliminate adapter-dimer.

Relevant sequencing metrics for duplicate libraries are shown in Table 2. With the exception of the 10 pg_2 sample, mapping rates exceeded 90%. As expected, unique mapping rates decreased significantly with lower DNA input from 1 ng to 10 pg.

Table 2. Relevant sequencing metrics for duplicate libraries

Sample	Total reads (M)	Mapped	Uniquely mapped	Peaks called
1 ng_1	26.8	94.7%	89.6%	25273
1 ng_2	26.3	94.5%	89.8%	27273
100 pg_1	29.2	93.8%	61.7%	27467
100 pg_2	26.0	93.4%	63.9%	26561
10 pg_1	28.2	92.3%	6.8%	19202
10 pg_2	28.4	85.6%	2.3%	21529

Despite the decrease in unique mapping rates, there are strong correlations between the ultra-low input libraries and the 1 ng_1 control, as shown in Figure 4. Pearson correlations were 0.991 and 0.934 for 100 pg_1 and 10 pg_1 samples, respectively. Additionally, 93.0% and 61.1% of the peaks—indicating locations of H3K4^{me3} modified histones—overlapped with the 1 ng control for the 100 pg and 10 pg samples, respectively (Figure 5). These data indicate a moderate to high degree of agreement with the control sample, even at 1% of the minimum validated DNA input for the KAPA HyperPrep Kit.

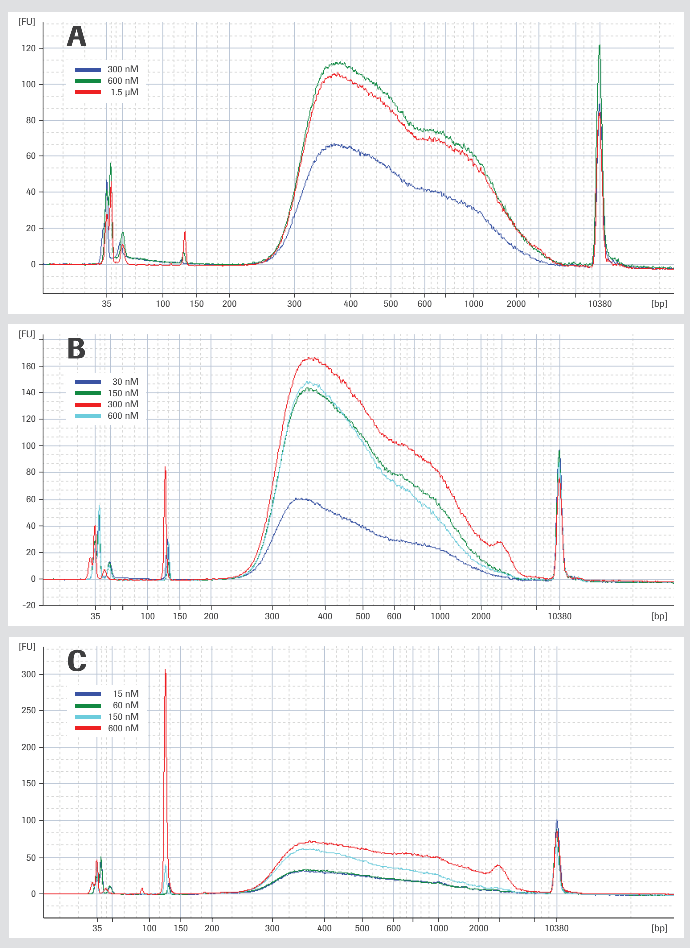


Figure 2. Post-amplification library size distributions (prior to size selection), as assessed by Bioanalyzer for 1 ng inputs (A), 100 pg inputs (B), and 10 pg inputs (C). Adapter input concentrations are labeled with their respective colors.

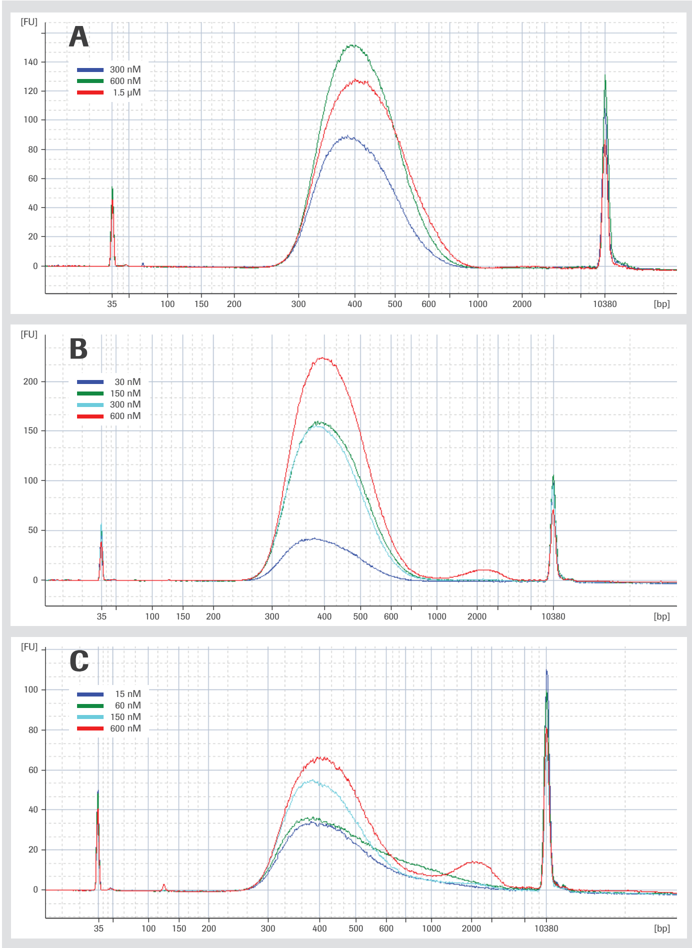


Figure 3. Library size distributions (after size selection), as assessed by Bioanalyzer for 1 ng inputs (A), 100 pg inputs (B), and 10 pg inputs (C). Adapter input concentrations are labeled with their respective colors.

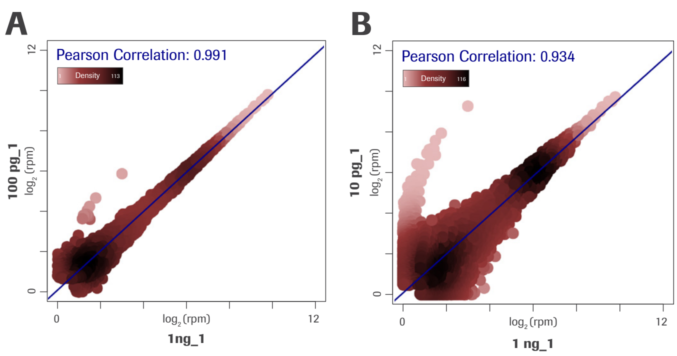


Figure 4. Correlation plots for 100 pg and 1 ng (A) and 10 pg and 1 ng (B). Pearson correlations were 0.991 and 0.934 for 100 pg_1 and 10 pg_1 samples, respectively relative to 1 ng input.

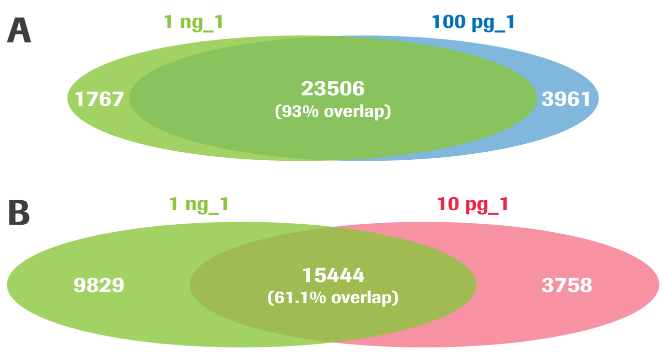


Figure 5. Peak overlap Venn diagrams as compared to 1 ng for 100 pg (A) and 10 pg (B). Data shows 93.0% and 61.1% of the peaks indicating locations of H3K4me3 modified histones overlapped with the 1 ng control for the 100 pg **(A)** and 10 pg **(B)** samples.

Figure 6 shows high correlations between replicates of 1 ng and 100 pg libraries, 0.994 and 0.988, respectively. The correlation of the 10 pg replicates was moderate at 0.734. Further, Figure 7 shows the peak coverage in reads per million (rpm) for ActB, a common housekeeping gene that was chosen as a proof-of-principle locus for illustrative purposes. The locations and shapes of the peaks were similar for all inputs down to 10 pg, with the exception of the 10 pg_2 sample, which had a lower representation of histone H3K4^{me3}. Both the library QC, as well as the sequencing data, indicate library construction was successful with inputs as low as 10 pg. However, increased variability was observed with 10 pg, even in the library QC data, as seen in Figure 8.

Conclusions

The KAPA HyperPrep Kit offers a streamlined and efficient solution for the construction of ultra-low input ChIP-Seq libraries. For 100 pg inputs, library QC metrics and sequencing data indicated successful and reproducible library construction, correlating well with 1 ng control data. For 10 pg inputs, library QC metrics showed successful library construction, but with a higher degree of variability. Sequence data indicated moderate peak overlap and a high correlation to 1 ng control data for one library. However, a replicate library showed lesser agreement, again indicating increased variability.

This work demonstrates that the optimization of library construction is critical to success with ultra-low inputs. For users interested in library construction with inputs less than 100 pg, recommendations for further optimization include assessing several DNA input amounts using a range of adapter concentrations to better identify the lower limit of the workflow that maintains reproducible performance.

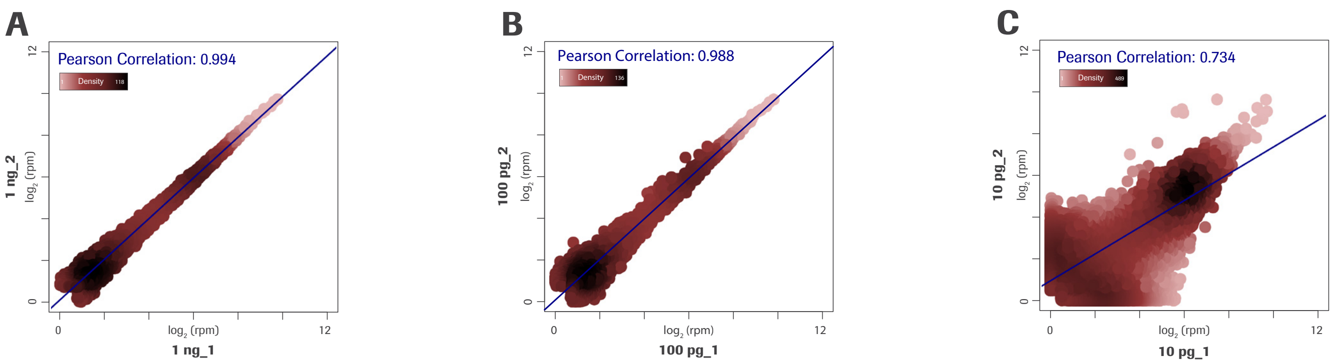


Figure 6. Correlation plots for 1 ng replicates (A), 100 pg replicates (B), and 10 pg replicates (C). High correlations between replicates of 1 ng **(A)** and 100 pg **(B)** libraries were observed with moderate correlation between the 10 pg **(C)** replicates.

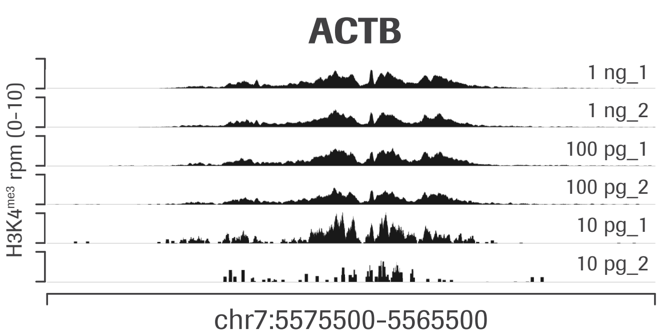


Figure 7. A schematic of the bound sites within the ActB locus for the replicates of 1 ng, 100 pg and 10 pg inputs. ACTB=housekeeping gene.

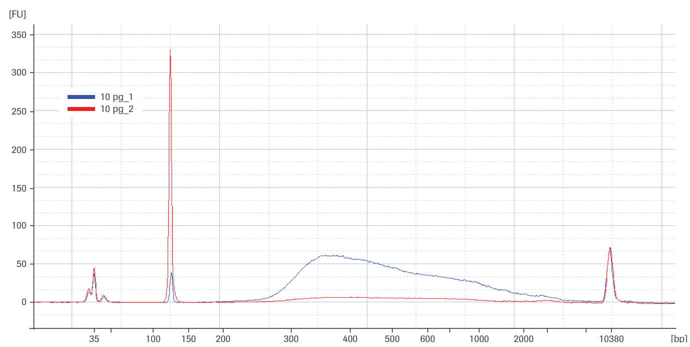


Figure 8. Bioanalyzer traces for the 10 pg replicates post-amplification (prior to size selection) show variability between replicates at 10 pg inputs.

Published by:

Roche Sequencing Solutions, Inc.
4300 Hacienda Drive
Pleasanton, CA 94588

sequencing.roche.com