Analysis of Protein–DNA Interactions With the SOLiD™ ChIP-Seq Kit

Introduction

Chromatin immunoprecipitation (ChIP) is a technique for identifying and characterizing protein–DNA interactions involved in gene regulation or chromatin organization. Historically, ChIP reactions were analyzed by PCR or microarrays. While microarrays provide a method for "global" ChIP analysis, direct sequencing of enriched fragments has proven more effective in determining the binding sites of proteins along the genome in an unbiased manner. The massively parallel sequencing throughput, accuracy, and flexibility of the SOLiD™ System make it well suited for ChIP-Seq. The SOLiD™ Chip-Seq Kit provides an end-to-end solution for ChIP-Seq and is optimized for sequencing on the SOLiD™ System.

Analyzing Epigenetic Modifications

Chromatin modifications and regulation of gene expression are among the most intensively studied fields of biology today due to the dynamic epigenetic control observed through various molecular mechanisms, such as DNA methylation and histone modifications (reviewed in [1–4]). Investigating epigenetic mechanisms is crucial for understanding cellular functions including gene transcription, DNA replication and recombination, repair, chromosomal stability, and cell cycle progression [4]. To understand the role of chromatin-associated proteins, it is essential to clarify where these regulatory proteins bind to the chromatin *in vivo*. Various posttranslational modifications such as acetylation, methylation, phosphorylation, and ubiquitination have been identified on

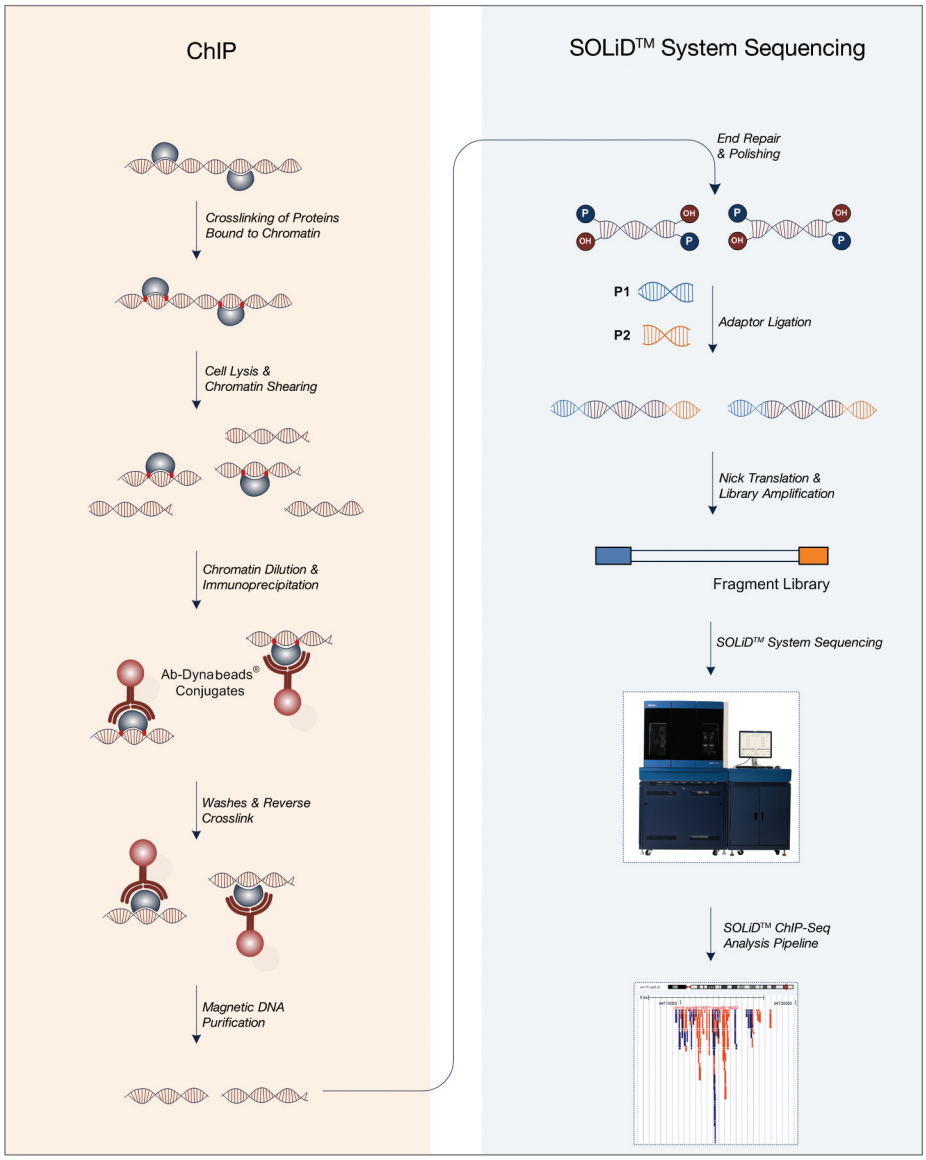

Figure 1. SOLiD™ System ChIP-Seq Workflow.

histones. For example, di- and trimethylation of histone H3 lysine 9 (H3K9me2, H3K9me3) and trimethylation of histone H3 lysine 27 (H3K27me3) generally correlate with the formation of repressive heterochromatin. In contrast, acetylation of histone H3 at lysine 9 and 14 create a more open chromatin environment accessible to transcription factors.

The ChIP assay starts with treating live cells with formaldehyde to crosslink DNA-binding proteins to DNA [5–7]. Subsequently, the cells are lysed, and the chromatin is sheared by sonication to create DNA fragments that are suitable for immunoprecipitation. Specific antibodies are then used to immunoprecipitate the protein of interest, together with the crosslinked DNA. This antibody/DNA complex is isolated typically using Protein A or G agarose, sepharose, or, more recently, magnetic beads. Reversal of the formaldehyde crosslinking with heat permits the recovery and quantitative analysis of the immunoprecipitated DNA. This material is now ready for downstream analyses such as quantitative PCR (qPCR), or genome-wide analyses using promoter-tiling arrays, or massively parallel sequencing. Microarray platforms provide a method for "global" ChIP analysis, but direct sequencing of enriched fragments has proven more effective in providing an unbiased assessment of DNA–protein associations along the genome. Importantly, the SOLiD™ System enables high-throughput sequencing, mapping, and counting of short DNA reads and, in combination with the SOLiD™ ChIP-Seq Kit, can be used for comprehensive, genome-wide mapping, requiring less starting material than other platforms.

There are several challenges associated with genome-wide scanning of protein–DNA interactions that limit its utility for many researchers. These include antibody efficiency, inability to handle smaller numbers of cells or cell equivalents in a single ChIP assay, capture efficiency, time of ChIP DNA preparation, and adequate depth of coverage. The SOLiD™ ChIP-Seq Kit provides a streamlined, optimized assay for the enrichment of chromatin complexes and DNA recovery using magnetic bead capture technology that overcomes many of these challenges.

ChIP-Seq Analysis With the SOLiD™ ChIP-Seq Kit and SOLiD™ System

The SOLiD™ ChIP-Seq Kit provides all reagents needed to perform ChIP with an antibody of interest, and provides a fast, reproducible solution. ChIP sample processing time is typically reduced by one day to just 5 hours, and only 10,000 to 300,000 cells (or equivalent) are needed for library preparation. Consequently, precious samples such as primary cells, stem cells, and biopsied material are used conservatively. The SOLiD™ ChIP-Seq Kit delivers increased sample throughput and reproducibility because it requires smaller volumes, employs magnetic bead-based separation (compatible with multichannel pipetting), and uses an optimized Dynabeads® Protein A/G Mix that works without blocking reagents. In contrast, the standard ChIP assay is a very time-consuming and laborious procedure containing multiple overnight incubations (Table 1).

Table 1. Comparison of ChIP Workflow Timelines.

Workflow Step	SOLiD™ ChIP-Seq Kit	Conventional ChIP
Preclearing	N/A	1–2 hr
Antibody/chromatin incubation	2 hr	Overnight
Bead pulldown	1 hr	2 hr
Washes	30 min (2 buffers)	1–3 hr (4 buffers)
Reverse crosslinking	1.5 hr	Overnight
Proteinase K digestion		2 hr
DNA elution from beads		15–30 min
DNA purification		2 hr to overnight
Average time	5 hr	36–48 hr

In addition to using optimized reagents and workflows, employing the right ChIP antibody is critical for successful ChIP-Seq experiments; therefore, we have qualified several important epigenetic antibodies for use in this assay. Visit www.invitrogen.com/chipantibodies for more information. In cases where a ChIP-qualified antibody is unavailable, there are some factors that may indicate that an antibody will be useful in ChIP assays. The antibody should be specific and well characterized. Ideally, characteristics such as purity, titer (i.e., ELISA), and cross-reactivity (i.e., dot blot), as well as western blot and immunohistochemistry (IHC) or immunoprecipitation (IP) performance will be known. An antibody may have greater success in ChIP if it is affinity-purified, polyclonal (containing a population of antibodies that recognize different epitopes), and recognizes native protein conformations (i.e., qualified by immunoprecipitation). However, even fully characterized antibodies may not be useful in some cases, as the crosslinking step can alter protein epitopes.

The SOLiD™ System's ability to generate hundreds of millions of sequence tags in a single run enables genome-wide ChIP analysis of complex organisms. Sequence tags are mapped to a reference sequence and counted to identify specific regions of protein binding. The throughput of the system provides researchers with the sensitivity and the resolution required to map and accurately characterize the protein–DNA interactions of an entire genome. Additionally, the flexible slide format allows researchers to analyze multiple experimental samples and a control sample in a single run. The SOLiD™ System ChIP-Seq workflow begins with ChIP (Figure 1). DNA derived from the ChIP procedure can range from 100 bp to 1 kb in size and is often limiting in quantity. Using the SOLiD™ ChIP-Seq Kit and the Bioruptor™ sonicator, we have optimized chromatin shearing to produce fragments sized between 100 and 300 bp. The ChIP DNA recovered is used for the construction of the fragment library with the reagents provided in the SOLiD™ ChIP-Seq Kit. The protocol is optimized for low-input samples ranging between 1 and 10 ng. Preparation of a negative control is strongly recommended. Those can include either non-immunoprecipitated, fragmented DNA of similar size range (chromatin input control) or chromatin-immunoprecipitated DNA (precipitated using nonspecific IgG

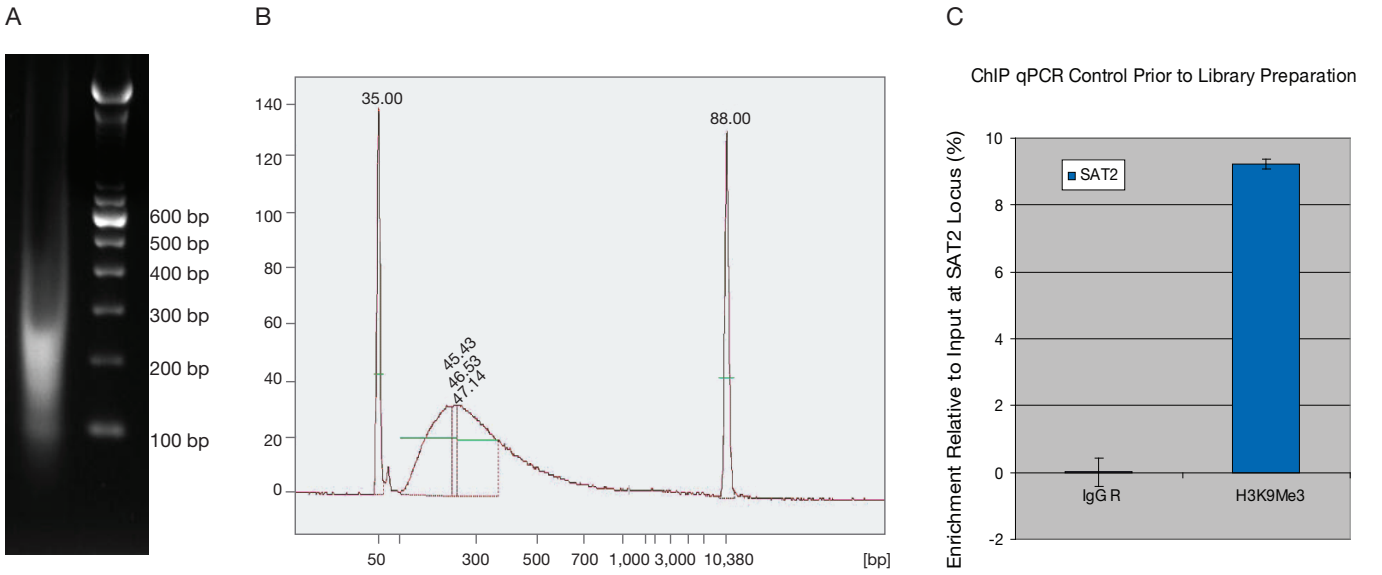

Figure 2. Preparing ChIP DNA and Confirming Enrichment Prior to Library Construction. A. Sheared chromatin from MCF-7 cells was prepared from 1 million cells in 50 µL lysis buffer. Samples were sonicated for 16 cycles of: [30 seconds "ON"/30 seconds "OFF"] with the Bioruptor™ sonicator. Following the SOLiD™ ChIP-Seq Kit protocol, a 10 µL aliquot of chromatin was treated with proteinase K and incubated at 55°C for 20 min on a temperature-controlled mixer (650 rpm). Ten microliters of the chromatin sample was visualized on a 2% E-Gel® precast gel alongside a sample of 100 bp DNA ladder. **B.** When analyzed on the Bioanalzyer 2100 (1 µL sample), the majority of the sheared chromatin was observed to be approximately 255 bp in length. **C.** Sheared chromatin was diluted to 100,000 cells per ChIP, and enrichment of H3K9me3 at the SAT2 locus was analyzed by qPCR.

antiserum to detect differential enrichment). Once these libraries are created, the samples are sequenced on the SOLiD™ System. The short sequence reads from the SOLiD™ System are counted and mapped against genomic sequences using the SOLiD™ System alignment tools such as SOLiD™ BioScope™ Software 1.2 (available through the Applied Biosystems Software Development Community http://info.appliedbiosystems.com/SOLiDsoftwarecommunity) or third-party tools. Data can then be viewed through the University of California, Santa Cruz Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) or third-party tools to identify and quantify the regions of sequence that bind to the protein of interest.

ChIP-Seq Analysis of Histone Lysine Methylation

In this study, we were looking for epigenetic changes in histone H3. Covalent modification of histones is a diverse group of changes and includes acetylation of lysines, methylation of lysines and arginines, phosphorylations of serines and threonines, ADP-ribosylation of glutamic acids, and ubiquitination and sumoylation of lysine residues. The combination of these covalent modifications gives rise to what is known as the "histone code." Histone modifications appear to occur in specific patterns during neoplastic transformation. The histone H3 methylation at lysine 27 (H3K27me3) is associated with epigenetic silencing in mammalian cells and established as a marker of neoplasia in various cancer cell lines [8]. ChIP-Seq analysis was performed on DNA from MCF-7 human breast carcinoma cell lines to identify loci associated with histone H3 Lysine 27 trimethylation (H3K27me3).

Materials and Methods

ChIP DNA was prepared according to the SOLiD™ ChIP-Seq Kit protocol from MCF-7 cells and was precipitated using the Invitrogen™ ChIP-qualified antibody specific to histone H3-K27Me3. Briefly, sheared chromatin from MCF-7 cells was prepared from 1 million cells in 50 µL lysis buffer. Samples were sonicated for 16 cycles of: [30 seconds "ON"/30 seconds "OFF"] with the Bioruptor™ sonicator (Figure 2). Subsequently, chromatin was diluted to 100,000 cells per ChIP according to the protocol. One microgram of antibody (IgG or H3K9me3) and 1 µL of H3K27me3 were used per ChIP experiment. A fraction (10%) of the sonicated chromatin was set aside to be used as an input control. For a positive control, an antibody that consistently binds chromatin-associated proteins under a wide variety of cellular conditions should be selected. For example, enrichment of heterochromatin markers such as H3K9me3 at the satellite repeat locus (SAT2) is consistently observed. Enrichment of H3K9me3 at the SAT2 locus was used as a positive control for the ChIP sample preparation prior to library preparation (Figure 2). Input (non-immunoprecipitated sample) and ChIP samples were quantified using the Invitrogen™ Quant-iT™ dsDNA High Sensitivity Assay Kit. The amount of ChIP DNA recovered depends on many factors, including epitope accessibility and protein binding-site accessibility. The Quant-iT™ dsDNA High Sensitivity Assay Kit and Quant-iT™ dsDNA High Sensitivity Assay Kits for use with the Qubit® Fluorometer facilitate accurate quantitation of most chromatin input samples. In some cases, the chromatin sample may be too dilute to be accurately quantified, even with high-sensitivity assays. In these instances, quantifications based on the chromatin input may be used.

Next, control chromatin input (10 ng) and ChIP DNA (~1 ng) were prepared in parallel using the SOLiD™ ChIP-Seq Kit protocol designed to accommodate the low input of library material and 12 cycles of PCR amplification. Templated bead generation for

each library was performed according to the SOLiD™ System 3 Plus User Guide standard protocol. Each sample was deposited on a quadrant of the slide at a target bead density of 175,000 beads/panel. A duplicate slide was generated and processed in the same way so that the technical reproducibility of the system could be assessed.

High-throughput sequencing was performed using the SOLiD™ System, and analysis of 50-base reads was carried out. All reads were filtered according to high-quality standards (0–5 mismatches in color space), as well as for alignment and unique placement in the human reference sequence.

The short sequence reads were mapped to the human reference (UCSC human genome build hg18) chromosome by chromosome. Subsequently, the output files (.ma files) from both the control and the experimental samples were converted to BED format, and data was analyzed using MACS [9]. The resulting peak files can be further filtered, and the resulting files (containing the peak coordinates as well as other information such as the fold of enrichment) can be converted to GFF format and/or FASTA format. These files can be visualized using common software tools such as the UCSC Genome Browser and MOTIF (Figure 3).

Results and Discussion

A summary of matching statistics from SOLiD™ System sequencing results of input DNA and ChIP DNA is presented in Table 2. (Read lengths and specific mapping tools used may affect matching statistics and should be taken into consideration when comparing different data sets.) Each sample in a quad section generated 26 to 28 million uniquely mapped reads and over 1 billion bases of data. Generally, 10 million mapped reads are required for ChIP sequencing. This massive throughput enables researchers to systematically survey the regions of the genome where proteins contact their DNA targets.

Next, the pattern of H3K27me3 at the RGMA locus located on chromosome 15 was analyzed. The RGMA gene has been shown to be silenced by chromatin modifications such as H3K27me3 [8]. We observed specific enrichment of H3K27me3 as depicted in Figure 3B (visualized via the UCSC Genome Browser). SYBR® dye qPCR primers targeting

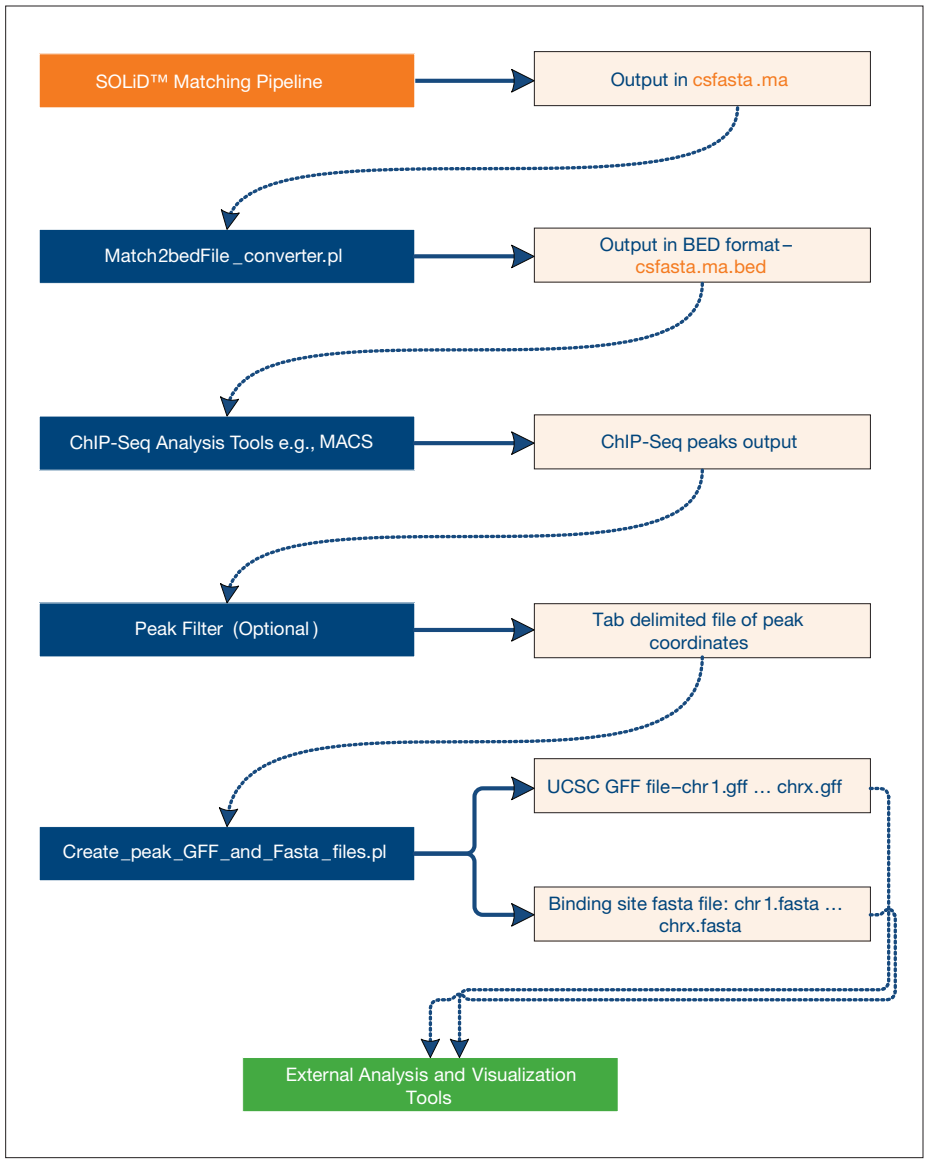

Figure 3. ChIP-Seq Data Analysis Workflow.

Table 2. Summary Statistics for SOLiD™ System Sequencing Run of Libraries Prepared With the SOLiD™ ChIP-Seq Kit and Protocol. Uniquely mapped reads (0–5 mismatches) for SOLiD™ ChIP-Seq (ChIP 1IP 1 ng H3K27me3) and control input (chromatin input 10 mg) libraries are shown and represent those reads mapping to a single, unique location with up to five mismatches in color space. Data can be visualized with a tool such as the UCSC Genome Browser to identify and quantify the regions of sequence that bind to the protein of interest.

Library	Number of Uniquely Mapped Reads	Throughput (Mbp)
Chromatin input 10 ng	28,470,110	1,423.51
ChIP 1IP 1 ng H3K27me3	26,033,473	1,301.67

the two peak regions observed in Figure 3B were designed and used for validation. A ChIP-qPCR experiment from parallel MCF-7 chromatin samples was performed. Consistent with the ChIP-Seq results, we observed enrichment of H3K27me3 by qPCR (Figure 3C).

Increasing evidence demonstrates that the disruption of epigenetic mechanisms (i.e., changes in gene expression not accounted for by changes in DNA) contribute to cellular transformation and cancer progression, including breast cancer [11]. Histone lysine 27 methylation has attracted particular attention as this modification is involved in neoplasia (through the silencing of developmentally important tumor suppressor genes) and is thought be involved in maintaining stem cell pluripotency [12]. We have also observed enrichment of H3-K27 from embryonal carcinoma cell lines from low–cell-number ChIP-Seq analysis, which is consistent with published reports that histone K3-K27 trimethylation plays an important role in maintaining the silent state of HOX genes [13] (data not shown).

Conclusion

ChIP-Seq on the SOLiD™ System has been streamlined with the SOLiD™ ChIP-Seq Kit's seamless, easy workflow that provides robust enrichment from a small amount of starting material. This is important when sensitive and accurate screening of large numbers of samples is necessary, such as in clinical research and analysis. The SOLiD™ System's throughput, dynamic range, and flexibility for multiplexing permits multiple hypothesis-neutral ChIP-Seq analyses to be performed in a single run. The optimized SOLiD™ ChIP-Seq Kit and SOLiD™ System together provide a powerful, streamlined, and complete workflow for studying protein–DNA interactions with unprecedented scope and depth.

References

1. Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* 403:41–45.

2. Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293:1074–1080.

3. Turner BM (2002) Cellular memory and the histone code. *Cell* 111:285–291.

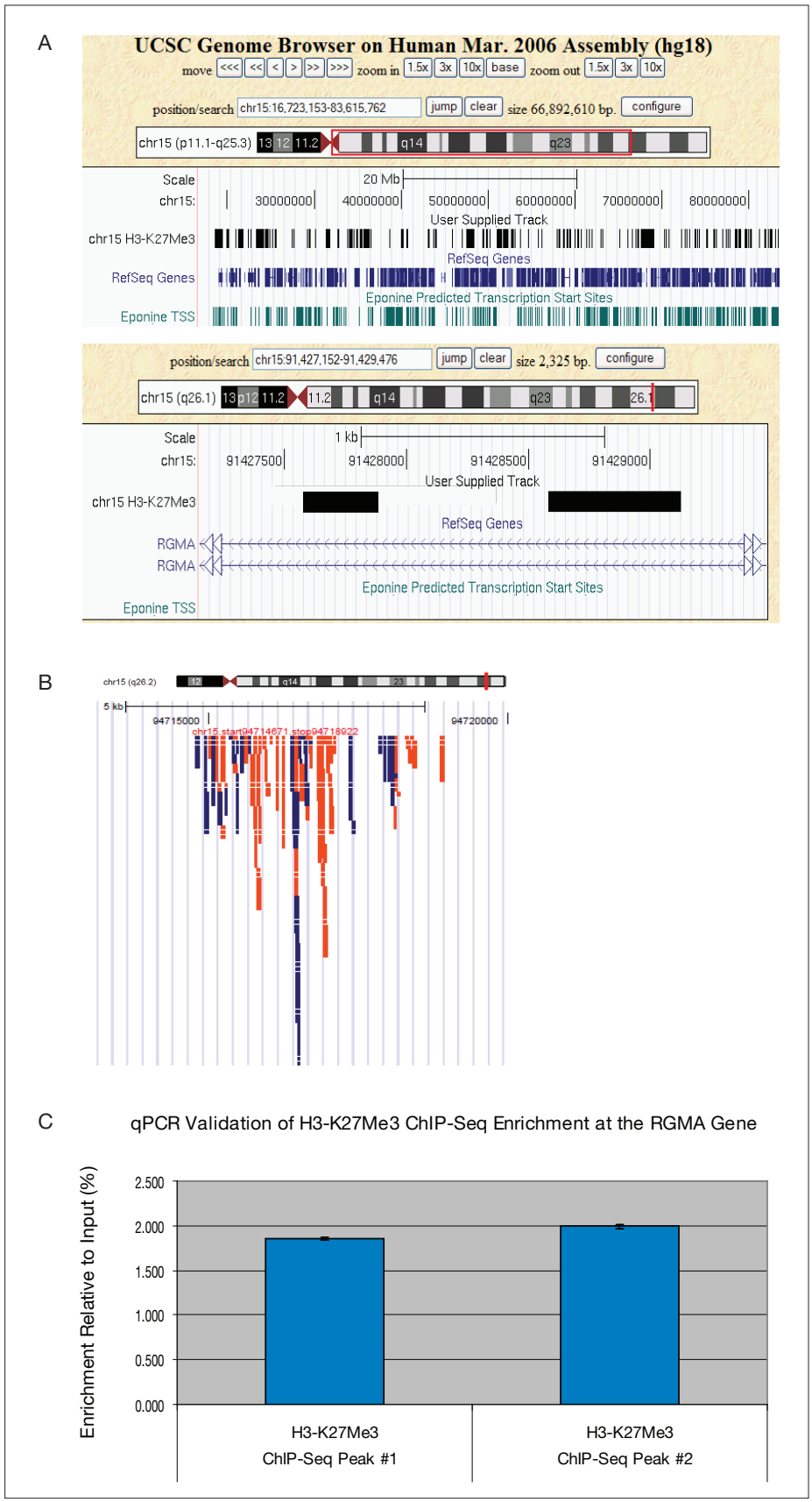

Figure 4. Visualization and Validation of Chip-Seq Data. A. UCSC Genome Browser view of a chromosomal locus near the RGMA gene. SOLiD™ System sequencing reads from H3K27me3 ChIP were mapped and normalized against the input control. **B.** Reads colored in blue are the reads mapped to the positive strand of the reference, whereas the orange-colored reads are mapped to the negative strand of the reference. Recent publications have reported that in mouse embryonic fibroblasts, H3K27me3 is not restricted to the promoter regions of silent genes, but instead generally marks broad localized regions that include silent genes and intergenic regions [10]. **C.** Validation of the ChIP-Seq enrichment of H3K27me3 at the RGMA gene by ChIP-qPCR.

4. Allis CD, Jenuwein T, Reinberg D (eds.), Caparros ML (assoc. ed.) *Epigenetics*. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY, 2006.

5. Orlando V, Strutt H, Paro R (1997) Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods* 11:205–214.

6. Das PM, Ramachandran K, van Wert J, et al. (2004) Chromatin immunoprecipitation assay. *Biotechniques* 6:961–969.

7. Kuo MH, Allis CD (1999) *In vivo* cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* 19:425–433.

8. Kondo Y, Shen L, Cheng AS et al. (2008) Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. *Nat Genet* 40:741–750.

9. Zhang Y, Liu T, Meyer CA et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137.

10. Pauler FM, Sloane MA, Huang R et al. (2009) H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res* 19:221–233.

11. Feinberg AP, Tycko B (2004) The history of cancer epigenetics. *Nat Rev Cancer* 4:143–153.

12. Boyer LA, Plath K, Zeitlinger J et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441:349–353.

13. Cao R, Wang L, Wang H et al. (2002) Role of Histone H3 lysine 27 methylation in polycomb-group silencing. *Science* 298:1039–1043.

Printed in the USA. 04/2010 Publication. CO21048 0410

applied biosystems™ by life technologies™

Headquarters
850 Lincoln Centre Drive | Foster City, CA 94404 USA
Phone 650.638.5800 | Toll Free 800.327.3002
www.appliedbiosystems.com

International Sales
For our office locations please call the division headquarters or refer to our Web site at
www.appliedbiosystems.com/about/offices.cfm