



Chromatin immunoprecipitation sequencing (ChIP-Seq) on the SOLiD™ system

Chromatin immunoprecipitation (ChIP) is a technique for identifying and characterizing elements in protein-DNA interactions involved in gene regulation or chromatin organization. Microarray platforms provide a method for 'global' ChIP analysis, but direct sequencing of enriched fragments has proven a more effective means for determining locations of DNA-binding proteins along the genome in an unbiased manner. The massively parallel sequencing capacity, high accuracy and flexibility of the SOLiD™ system make it well suited for ChIP-sequencing (ChIP-Seq) applications.

Sensitive ChIP-Seq analysis using the SOLiD system

The SOLiD System's ability to generate over 400 million sequence tags (35–50-bp sequence reads) in a single run enables whole-genome ChIP analysis of complex organisms. Sequence tags are mapped to a reference sequence and counted to identify specific regions of protein binding. The ultra-high throughput of the system provides researchers with the sensitivity and the statistical resolving power required to map and accurately characterize the protein-DNA interactions of an entire genome. Additionally, the incorporation of barcodes allows researchers to cost-effectively analyze multiple experimental samples and a control sample in a single run.

ChIP-Seq analysis with the SOLiD System begins with a traditional ChIP procedure (Fig. 1). DNA is cross-linked *in vivo* to DNA-binding proteins with formaldehyde and mechanically sheared using sonication. The DNA-protein complex is then precipitated with an antibody that is specific to the DNA-binding protein. The quality of this antibody is critical to the success of ChIP-Seq protocols, as it determines the level of enrichment over background that is obtained. The DNA is released by reversing the cross-link to the protein, and the protein is digested. The size and concentration of the resulting ChIP DNA fragments determine the approach that is taken to process samples for SOLiD fragment library construction and subsequent sequencing.

Typically, DNA derived from the ChIP procedure can range from 100 bp to 2 kb and is often limiting in quantity. Therefore, we recommend using the low-input DNA protocol for generating the fragment library.

We suggest preparing a negative control that consists of either non-immunoprecipitated fragmented DNA of similar size range, or DNA that has been chromatin-immunoprecipitated using nonspecific IgG

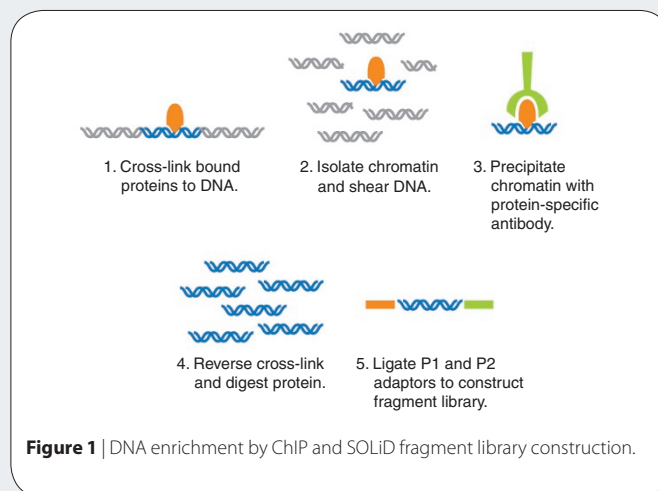


Figure 1 | DNA enrichment by ChIP and SOLiD fragment library construction.

antisera, to detect differential enrichment. Once SOLiD ChIP-Seq and negative control libraries are created, the samples are sequenced on the SOLiD System. The short sequence reads from the SOLiD System are mapped against genomic sequences using the SOLiD System alignment tools available through the Applied Biosystems software development community (<http://info.appliedbiosystems.com/solidsoftwarecommunity/>) or third-party tools compatible with SOLiD sequencing data. Data can then be visualized with a tool such as the University of California, Santa Cruz (UCSC) Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgGateway/>) to identify and quantify the regions of sequence that bind to the protein of interest.

SOLiD system ChIP-Seq analysis of FOXA3 protein

In collaboration with the laboratory of Claes Wadelius at Uppsala University, we performed ChIP-Seq analysis using ChIP DNA isolated from hepatic cell lines to identify loci involved in interactions with the FOXA3 or HNF3 γ protein. This hepatocyte nuclear factor, a member of the forkhead class of DNA-binding proteins, activates transcription

Anjali Shah

Applied Biosystems, Inc. (Part of Life Technologies), Foster City, California, USA.
Correspondence should be addressed to A.S. (anjali.shah@appliedbiosystems.com).

APPLICATION NOTES

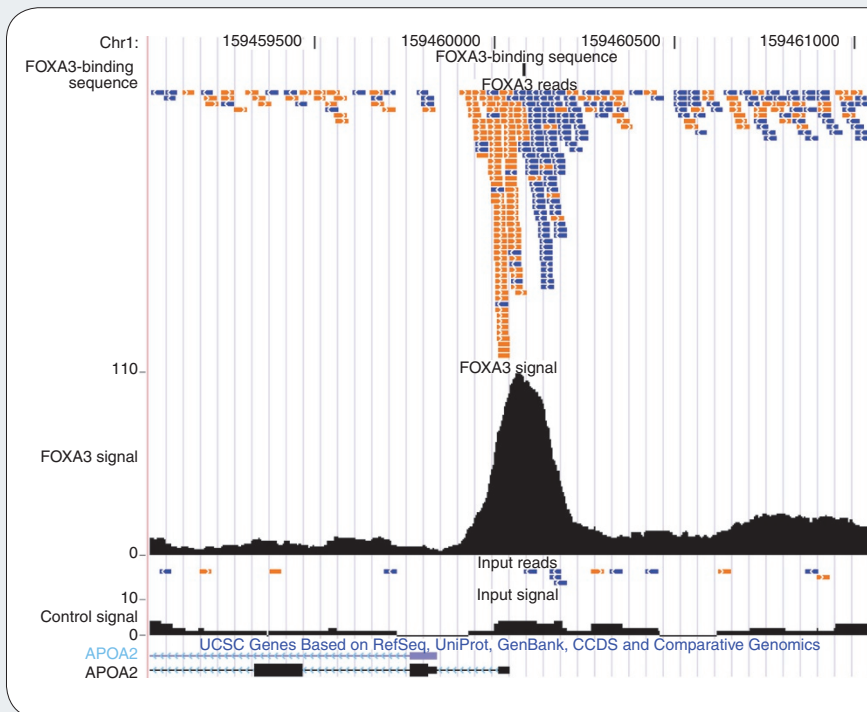


Figure 2 | Evidence for FOXA3 protein–DNA interaction at the *BCAS1* locus binding to the *APOA2* promoter. Graphic representation of alignment of reads derived from size-selected SOLiD ChIP-Seq libraries (orange) and from control input sample (blue) along chromosome 20 is shown. Illustrated in the UCSC Genome Browser are the mapped forward (orange) and reverse (blue) reads and the region of overlap (black). The peak is just upstream of the first exon located and spans the predicted FOXA3-binding motif.

of liver-specific genes *ALB* (encoding albumin) and *APOA2* (encoding apolipoprotein A-II) and also interacts with DNA and histones as a pioneer factor that opens chromatin during development.

We prepared ChIP DNA, as previously described¹, from HepG2 cells using a commercially available antibody (Santa Cruz Biotechnology) specific to the FOXA3 protein. After immunoprecipitating DNA (0.5 µg), we size-selected and purified it using gel electrophoresis to isolate DNA fragments of three different size ranges: 150–200 bp, 200–250 bp and 250–300 bp. Sheared, non-immunoprecipitated hepatic cell line DNA (4 µg), similarly subjected to size selection and purification of 250–300-bp fragments, served as the negative control sample. We performed P1 and P2 adaptor ligation, and all subsequent steps on all four samples, according to SOLiD System 2.0 fragment library preparation: lower-input DNA protocol. Information about this protocol is available at <http://solid.appliedbiosystems.com>. Templated bead generation for each library was performed according to SOLiD System 2.0 User Guide standard protocols. Each sample was deposited on one quadrant of the slide at a target bead density of 60,000–70,000 beads per panel. We generated two slides and processed them similarly to assess the reproducibility of the system.

High-throughput sequencing was performed using the SOLiD System and analysis of 35-bp reads was carried out. All reads were filtered for high-quality reads as well as for alignment and unique placement in the human reference sequence.

In these experiments, the data from one quadrant generated more than enough reads to map all signals in the genome. We extended these reads *in silico* to represent the selected ChIP fragment sizes (in base pairs). We used the following process to qualify peaks: (i) We calculated a cutoff for the overlap signal using the binomial distribu-

tion as a model under the hypothesis that the reads were randomly placed onto the genome. (ii) FOXA3 peaks with a much higher signal in the control sample were filtered from the analysis. (iii) A peak was required to have a significant signal of forward reads upstream or a reverse signal downstream from its center (**Fig. 2**).

We visualized putative FOXA3-binding regions using the UCSC Genome Browser. We detected putative FOXA3-binding sequences in the enriched peak regions using the BCRANK package in R/Bioconductor (<http://www.bioconductor.org/packages/release/bioc/html/BCRANK.html>).

Hypothesis-neutral ChIP-Seq analysis

As expected with an enriched sample, the resulting uniquely mapped reads covered about 10% of the human genome on average. The mapping of reads has unique starting points indicating that the analysis has good genomic representation and minimal amplification bias. Based on these data, 5–10 million uniquely mapped reads is sufficient to map all protein-binding regions in a complex genome such as human.

Feature	SOLiD System (ChIP-Seq)	Microarray (ChIP-chip)
Resolution	>400 million sequence tags per run	~6.5 million oligonucleotides per array
Genome coverage	Unlimited: entire genome can be sequenced hypothesis free	Limited by probe design
Specificity	No cross-hybridization risks; identifies unique sequence tags	Cross-hybridization risks between closely related elements
Sample multiplexing	Yes	No

© 2009 Nature America, Inc. All rights reserved.

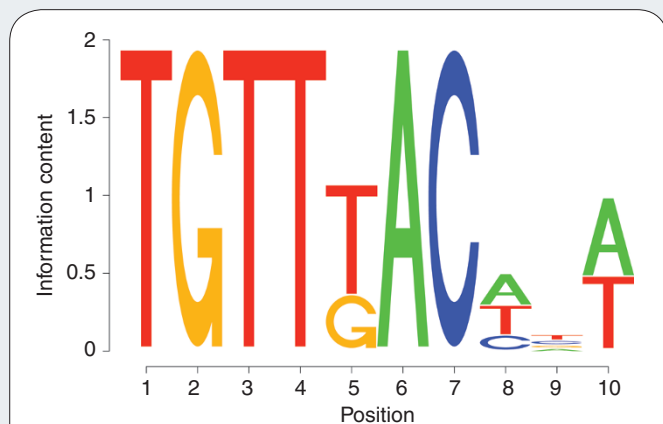


Figure 3 | Putative consensus sequence for FOXA3 transcription factor generated using BCRANK. Transcription factor identified by ChIP-Seq on the SOLiD System. Information content is plotted as a function of nucleotide position. Sequence logo image was created using an R package called seqLogo (<http://bioconductor.org/packages/2.2/bioc/html/seqLogo.html>).

All sequence information was used based on the high correlation values between different replicates, ranging from 0.70–0.83. For peak determination, we used a cutoff value of 21 overlapping fragments at a Bonferroni-corrected $P < 0.01$ based on the simulation model. Using these stringent criteria, >4,000 peaks for FOXA3 were detected either at promoters of protein-coding genes or at a distance from genes that was consistent with FOXA3 interaction with *cis*-regulatory

elements. An example of a FOXA3 peak detected by our peak-detection strategy is illustrated in **Figure 2**. This peak, located within the *APOA2* promoter region—which had previously been shown to bind FOXA3—contains 110 overlapping fragments, well above the determined threshold for overlapping fragments. Based on these data, we established a consensus FOXA3-binding motif (**Fig. 3**), which closely resembles the previously characterized FOXA3-binding sequence. This motif was centered at the peak maxima, including that of the *APOA2* promoter.

The SOLiD System provides a high level of throughput and sensitivity that cannot be achieved with current hybridization technologies (**Table 1**). The SOLiD System's ability to generate over 400 million sequence tags, and to take advantage of multiplexing capabilities, allows multiple hypothesis-neutral ChIP-Seq analyses to be performed in a single run. These system attributes, along with the high degree of accuracy, allow for the determination of regulatory networks in various cellular and pathological states.

ACKNOWLEDGMENTS

The data presented in this application note were obtained in collaboration with C. Wadelius, Department of Genetics and Pathology, Uppsala University, Sweden.

1. Rada-Iglesias, A. *et al.* Binding sites for metabolic disease related transcription factors inferred at base pair resolution by chromatin immunoprecipitation and genomic microarrays. *Hum. Mol. Genet.* **14**, 3435–3447 (2005).

This article was submitted to *Nature Methods* by a commercial organization and has not been peer reviewed. *Nature Methods* takes no responsibility for the accuracy or otherwise of the information provided.