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 antiserum, 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 HNF3γ protein. This hepatocyte nuclear factor, a member of the forkhead class of DNA-binding proteins, activates transcription...
APPLICATION NOTES

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.

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