ChIP Sequencing Service For Studying Protein-DNA Interactions

Introduction
Chromatin-immunoprecipitation (ChIP) followed by next generation sequencing of the immuno-precipitated DNA is a powerful tool for the investigation of protein-DNA interactions. ChIP-Seq studies are mainly performed to increase our understanding of transcription factor biology and histone modifications. In the “ChIP” part, chromatin is isolated from cells or tissues and fragmented. Antibodies against chromatin-associated proteins are used to enrich for specific chromatin fragments. In the “Seq” part the isolated DNA is sequenced and aligned to a reference genome to determine specific DNA-protein binding loci at a nucleotide level. Advanced bioinformatics analyses help to gain insights into common binding motifs as well as the protein function by analyzing affected gene regulatory networks. As a plethora of similar methods have emerged for functional experiments (e.g. DNase-Seq, Ribo-Seq, PAR-CLIP, etc.), downstream sequencing and analysis follow a similar pattern. Therefore, feel free to inquire about a sequencing and bioinformatics solution in case you plan any such project.

Figure 1. Typical ChIP-Seq workflow: DNA-protein complexes are stabilized by crosslinking, while unprotected DNA is fragmented and remaining DNA-protein complexes of interest are immunoprecipitated with specific antibodies. Furthermore, DNA is isolated, purified and then subjected to NextGen sequencing followed by thorough bioinformatics analysis.

Microsynth Competences and Services
Experimental Design
Please consider the importance of biological replicates as well as control/reference samples (e.g. by an unrelated IgG) along with your specific antibodies to allow an accurate and meaningful analysis – both are highly advised by Microsynth.

Sequencing
DNA purified by the customer is prepared following the Illumina TruSeq ChIP sample preparation protocol. The DNA libraries are then sequenced and the resulting reads are demultiplexed and trimmed of Illumina sequencing adapter residuals. The required sequencing depth

Figure 2. Microsynth’s workflow for ChIP-Seq projects. It lists all possible input and output points.
depends strongly on the genome size, peak types, library complexity and aim of the study. As a rule of thumb, 10 million uniquely mapped reads per replicate should be aimed for human genomes given point-source peaks.

Bioinformatics Analysis
First, the sequenced reads are checked and filtered for their quality. Second, the sequence data is mapped to the reference genome (e.g. hg19, mm10, danRer7, etc.). Third, the analysis software HOMER is employed as a core module to detect DNA-binding sites (“peak-finding”) and motifs suitable for your DNA binding proteins (point-, broad-, or mixed sources such as transcription factors, certain chromatin marks or RNA polymerase II, respectively). An in-depth annotation (proximity to genes, gene ontology, etc.) is provided, in case a thoroughly annotated reference genome (e.g. human, mouse, zebrafish, etc.) is available.

Analysis Output
The most important results of the ChIP-Seq analysis are presented in form of an HTML document which allows user-friendly navigation through the assessment of the experiment, the annotation of the peaks, the binding motifs, their co-occurrence and the pathway analysis. Furthermore, numerous additional useful analysis results are provided along with the sequencing data in the FASTQ format. Some selected examples for further downstream analysis are shown in Figures 3-4 and Table 1.

![Figure 3](image3.png)

**Figure 3.** For the pathway analysis, the three gene-ontology domains (cellular component, biological process and molecular function) are each displayed in a network graph as a preview for the most significant ontology terms. Here a detail of the “cellular component” graph is presented.

![Figure 4](image4.png)

**Figure 4.** Among the enrichment analysis results, de novo motifs are detected and visualized. Significance is calculated and provided along with GO-terms and cross-links to further information (not shown).

Table 1: A selected excerpt of the comprehensive peak annotation is displayed. Additional information such as cross references, categorical data and motif patterns for each peak position is provided in the annotation table (as supported by the reference genome annotation).

<table>
<thead>
<tr>
<th>PeakID</th>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Strand</th>
<th>Peak Score</th>
<th>Focus Region/Region Size</th>
<th>Gene Name</th>
<th>Annotation</th>
<th>Distance to TSS</th>
<th>Nearest PromoterID</th>
<th>Gene Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr17-17</td>
<td>chr17</td>
<td>72200743</td>
<td>72303943</td>
<td>+</td>
<td>398.8</td>
<td>0.98</td>
<td>KIF19</td>
<td>intron</td>
<td>118508</td>
<td>NM_153209</td>
<td>kinesin family</td>
</tr>
<tr>
<td>chr17-17</td>
<td>chr17</td>
<td>23942474</td>
<td>23942674</td>
<td>+</td>
<td>244.2</td>
<td>0.968</td>
<td>IPOSPI</td>
<td>intergenic</td>
<td>41917</td>
<td>NR_103742</td>
<td>Importin 5 pseuogogene 1</td>
</tr>
</tbody>
</table>

Related Topics
- microRNA Sequencing at Microsynth
- RNA Sequencing at Microsynth

Further Reading