## Chromatin Immunoprecipitation-based Sequencing (ChIP-Seq) on the SOLiD™ System

### Introduction

Chromatin immunoprecipitation (ChIP) is a technique (Figure 1) for identifying and characterizing elements in protein-DNA interactions involved in gene regulation or chromatin organization. Historically, ChIP reactions were analyzed using IP Western blotting methods and more recently by microarray technologies also known as ChIP on Chip. Microarray platforms provide a method for "global" CHiP analysis but their probe design is hypothesis-driven and limited to known or predicted protein-interacting regions of the genome. This limitation has been overcome with massively parallel sequencing on the SOLiD™ System which supports hypothesis-neutral ChIP analysis, or ChIP-Seq (Table 1).

| Feature | **SOLiD System** (ChIP-seq) | Microarray (ChIP-chip) |
|---|---|---|
| Resolution | > 85 million sequence tags per run | ~ 6.5 million oligonucleotides per array |
| Genome Coverage | UNLIMITED: Entire genome can be sequenced hypothesis-free | limited |
| Specificity | No cross-hybridization risks.Identify unique sequence tags | Cross-hybridization risks |
| Sample Multiplexing | YES | no |

Table 1: ChIP Analysis features for the SOLiD System and Microarrays

### SOLiD™ System ChIP-Seq Analysis

The SOLiD ™ System's ability to generate over 240 million sequence tags (35 bp read length) in a single run enables whole genome ChIP analysis of complex organisms. Sequence tags are then counted and mapped to a reference sequence 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 to accurately characterize the protein-DNA interactions of an entire genome. Additionally, the flexible slide format allows researchers to analyze both normal and treated/compromised samples in a single run.

ChIP analysis with the SOLiD™ System (Figure 1) begins with a traditional chromatin immunoprecipitation procedure. DNA is cross-linked to DNA-binding proteins with formaldehyde. The DNA-protein complex is then isolated by shearing the DNA into fragments and 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 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 determines the approach that is taken to process this sample for SOLiD™ fragment library construction and

subsequent sequencing.

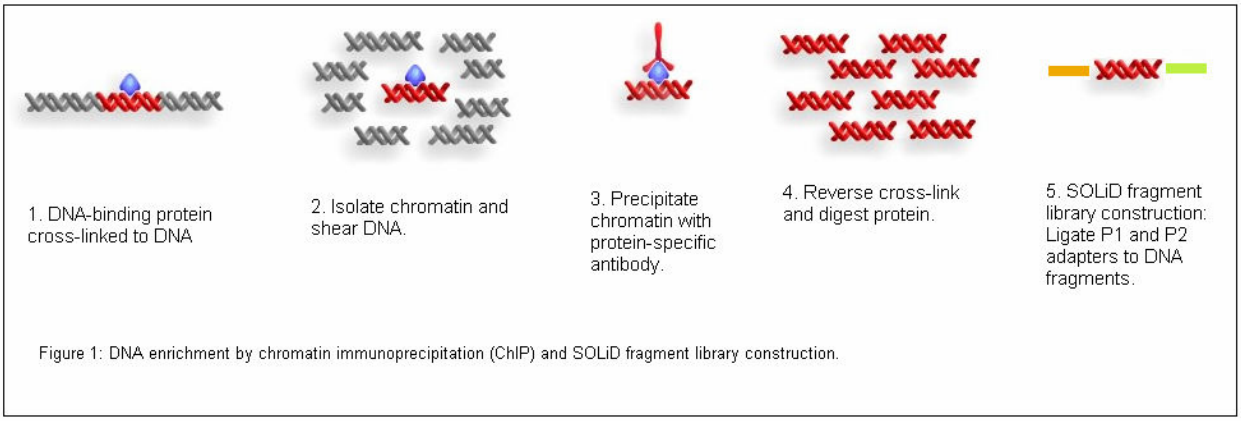


Figure 1: DNA enrichment by chromatin immunoprecipitation (ChIP) and SOLiD fragment library construction.

Typically, DNA derived from ChIP procedure can range from 100 bp to 2 kb in size and is often limiting in quantity (20-500 ng). Therefore, modifications to the standard SOLiD Lower-input DNA-Fragment Library preparation protocol are used to create the ChIP -Seq library. (Figure 2).

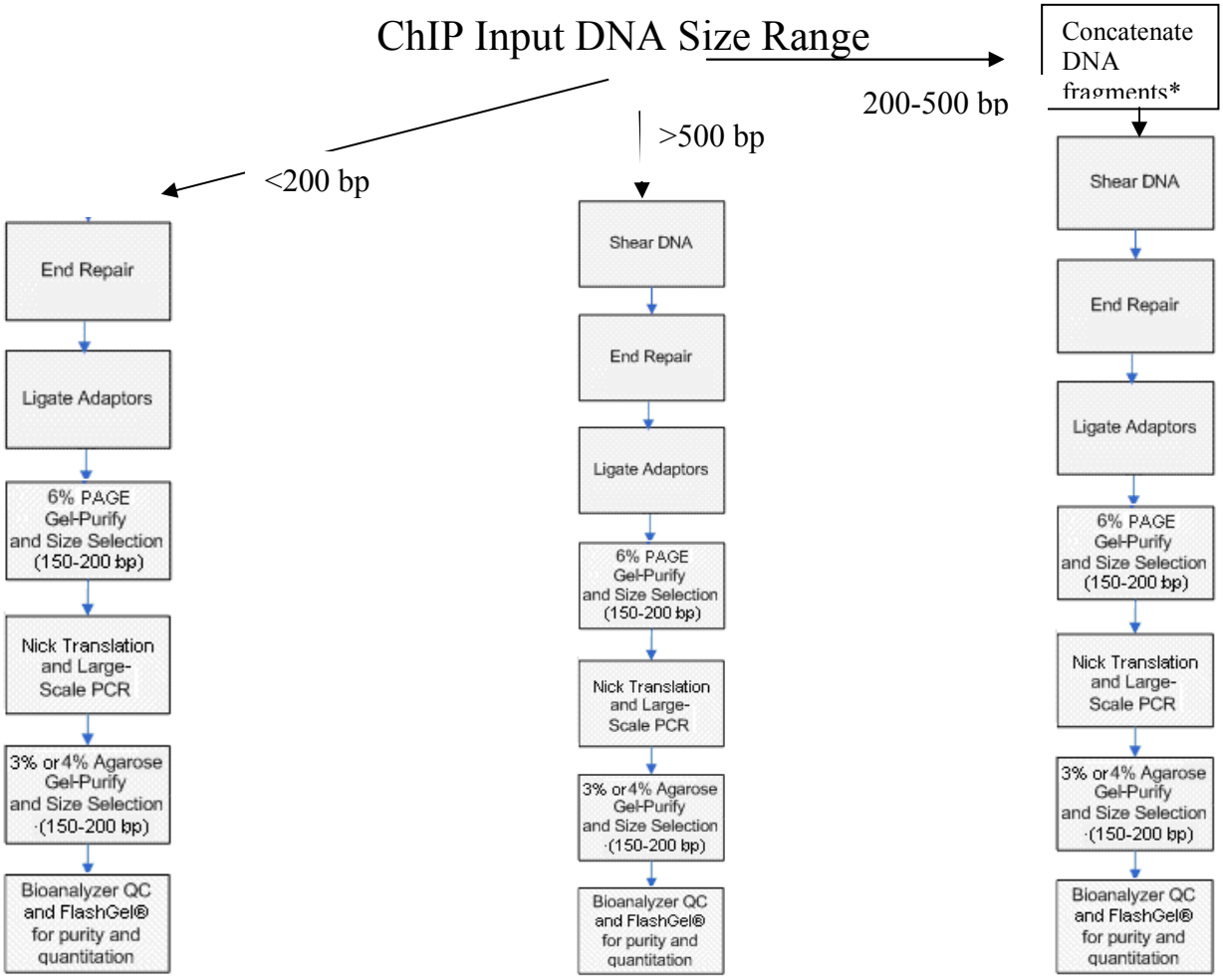

Figure 2. ChIP -Seq SOLiD™ fragment library preparation based on CHiP input DNA size ranges.

Preparation of a negative control consisting non-immunoprecipitated fragmented DNA of similar size range is required for detection of differential enrichment. Once these 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 SOLiD™ system alignment tools available through the Applied Biosystems Software Devlopment Community (http://info.appliedbiosystems.com/solidsoftwarecommunity) or third-party tools compatible with color space, and can be visualized with a tool such as the University of California, Santa Cruz Genome Browser in order to identify and quantify the regions of sequence that bind to the protein of interest.

Once you have determined which path to take, above, you are going to ultimately follow the low input protocol. Following adaptor ligation the library molecules are subjected to size selection on 6% DNA retardation PAGE gel, followed by in-gel PCR.

Depending on the input available, sometimes it is hard to see any DNA on the gel and in such Cases one has to make blind cuts. If such a situation arises, one should cut at least 3 size pieces (approx 150 to 175bp, 175 to 200bp and 200 to 230 bp). For sequencing, the best amplified library with the least number of PCR cycles should be used.