Recommended Standards for ChIP-Seq Experiments
Version 1.0
November 5, 2008

1. Matched sequenced reference sample
   Deeply sequenced reference samples such as Input DNA exhibit "peaks" in the promoter regions of many known genes. In order to identify regions of transcription factor binding in a ChIP-Seq experiment it is necessary to have sequenced a relevant reference sample in the same cell-line (and same cellular condition) in order to show that the region is enriched in the ChIP sample compared to the control. For the case of ChIP-Seq experiments for different transcription factors, only one control sample need be produced for each cell-line. The control (i.e. Input DNA) should be sequenced to at least the same depth as the greatest depth of any matching individual transcription factor ChIP-Seq experiment.

2. Depth of Sequencing
   The required depth of sequencing needed will vary depending on the nature of the binding of the transcription factor (number of binding sites, domain vs. point source binding, efficiency of the antibody). Saturation for some transcription factors/chromatin modifications might require exceptionally deep sequencing to achieve probable saturation of all biologically relevant sites. We recognize that for many factors, we will not have a large sample of previously known biologically relevant sites at the time of the ChIP-Seq measurement. One of three criteria should therefore be used for depth of sequencing:
   a) When the number of targets begins to saturate (i.e. approach the asymptotic amount of identifiable targets). The criteria should be ≥ 95% of the extrapolated total number of targets (for HeLa-S3 Pol II 12 million mapped reads yields greater than 95% of the approximately 30,000 extrapolated total targets).
   b) If the total number of targets does not approach saturation, one should detect 99% of targets that show at least 2-fold enrichment over control with 90% of the data.
   c) If either a) or b) are not satisfied, then at least 12 million mapped reads should be sequenced for human or 3-4 million mapped reads for worm or fly. Since individual lanes of data for Solexa/Illumina ChIP-Seq samples typically generate 3-4 million mapped reads, human would correspond to 3-4 lanes and worm or fly would correspond to one lane.

3. Number and Reproducibility of Biological Replicas
   At least two biological replicates are necessary in order ensure that the experiment is reproducible. It does not seem that there will be a significant gain in information beyond two biological replicates, when they are in reasonable agreement. The data from replicates can then analyzed and reported as the required independent target site lists. We propose to require the following agreement between biological replicas (target lists identified using common number of mapped reads):
   a) The number of mapped reads from different replicas should be within a factor of two of each other.
   b) Length of target lists should be within a factor of two of each other.
   c) Either of the following options
      I. Intersect top fraction (40%) of target list from one replica against the entire other target list and require a threshold amount of overlap (90%) and repeat for the reciprocal.

II. Target lists scored using all available reads share more than 75% of targets in common.

These standards have been recommended following an analysis by the Snyder lab of deeply sequenced ChIP-Seq datasets for HeLa-S3 Pol II. These recommendations may be revisited as more datasets are available.