REMC Standards and Guidelines for ChIP-seq

v1.0

I. Introduction

Next generation sequence based profiling of histone modifications (ChIP-seq) involves enrichment of chromatin fragments with modification-state-specific histone antibodies. This step is followed by deep sequencing of the enriched DNA and read alignment. The aligned reads enable derivation of density maps that describe the genome-wide distribution of the histone modification under study.

II. ChIP assay and QC guidelines.

Cultured cells should be immediately fixed with formaldehyde upon harvest to minimize non-physiologic alterations in chromatin landscape. Tissues may either be fixed when collected, or flash frozen and then subject to fixation upon thawing, immediately prior to sonication. Chromatin fragments from sonication should be evaluated by agarose gel or Agilent Bioanalyzer for all samples to ensure tight size range of ~200-600 bp. Sonication of each new cross linked cell or tissue preparation must be independently optimized to achieve a consistent size range. Primary antibodies may be either monoclonal or polyclonal, and should be validated by (i) Western blot to verify specific recognition of histone, and (ii) dot blots against an array of modified histone tail peptides to verify specificity for the targeted modification. When antibodies are commercially sourced, the vendor may supply the validation data, provided it is derived from the actual lot in use.

III. Library construction protocol.

Library production should proceed per manufacturer's protocols, using minimal PCR amplification (recommended range of ~18 cycles with Illumina procedures). Adapters for Illumina sequencing may be either single plex or multiplex ("barcoded"); in the latter case, it is important to establish that adapters are essentially uncontaminated by other adapter sequences.

IV. ChIP-seq Sequence Experiment QC Metrics.

The following QC metrics should be determined and monitored to ensure that REMC ChIP-seq datasets are of high quality.

1. Read length and sequencing depth

Our general recommendation is for a sequencing depth of at least 20 million aligned reads per replicate, at a 36 base read length, but different sequencing depths could be made for different marks based on distribution of coverage for that particular mark. Ultimately, the objective is to employ a depth of coverage sufficient to discover most of the enriched regions that could be discovered with an arbitrarily great sequencing depth.

2. Fraction aligned reads, duplicate reads. At present, there are no specific recommendations for these parameters. However, it is evident that these metrics can give important insights into the progress of a Chip-Seq experiment, helping one to detect, for instance, primer dimer artifacts (leading to low fraction aligned reads), contamination of human cell lines by mouse cell lines (same effect), or potential PCR artifacts (leading to a low fraction of unique reads). However, it is essential to take into account the expected complexity of the library (which varies from mark to mark) and the sequencing depth before setting a standard for percent duplicate reads, since over sequencing of a good library can produce perfectly good data but with a reduced fraction of unique reads.

3. Concordance between replicate datasets. ChIP-Seq datasets should be replicated at the level of the biological sample, wherever possible. An exception can be made in the case of a tissue type of high clinical significance which is nonetheless difficult to obtain. The replicated datasets should be compared to ascertain consistency following criteria appropriate to this project.

4. Fraction of reads in enriched intervals, and other criteria.

The Roadmap Epigenomics project has an ongoing effort to define and reduce to practice other appropriate data based QA criteria. When appropriate, these should be incorporated into our documentation of standards.

5. Use of controls.

To ascertain the background distribution of fragment abundances in the input material, it is essential to sequence a library derived from the chromatin preparation.