

Pipeline overview

The pipeline takes alignment bam files (processed from ChIP-seq Mapping Pipeline) from replicates and its control experiment to produce several outputs:

- Signal tracks' depicting control-normalized tag density (bigWig format) are created for each replicate, and for both replicates' reads pooled together.
 - The signal is expressed in two ways: as fold-over control at each position, and as a p-value to reject the null hypothesis that the signal at that location is present in the control.
- Relaxed peak calls (bed and bigBed format) for each replicate and for both replicates' reads pooled together. These peak calls are thresholded to sample enough noise in the experiment for efficient statistical comparison of replicates in subsequent steps. Ergo, many false positives are expected to be present in these peak sets. They should not be interpreted as definitive binding events, but rather intended as inputs for subsequent statistical comparison of replicates.
- **For Histones:**
 - Replicated peak calls (bed and bigBed format) are the set of peak calls from the pooled replicates that are either observed in both replicates, or are observed in two pseudoreplicates of the pool. Pseudoreplicates are peak sets called on half of the pooled reads, chosen at random without replacement.
- **For Transcription Factors and Chromatin Remodelers:**
 - Final peak calls (bed and bigBed format) are the set of peak calls that pass IDR at a threshold of 2%. The conservative set are peaks derived from IDR analysis of biological replicates, whereas the optimal set are the largest set of peaks derived from IDR analysis of biological replicates and pseudoreplicates. Pseudoreplicates are peak sets called on half of the pooled reads, chosen at random without replacement.

Pipeline Restrictions

- The read length should be a minimum of 50 base pairs, though longer read lengths are encouraged.
- The sequencing platform used should be indicated.
- Replicates should match in terms of read length and run type.
- Pipeline files are mapped to either the GRCh38 or mm10 sequences:
<https://www.encodeproject.org/references/ENCSR425FOI/>