ChIP-Seq-seq Data Delivery Specifications

Output data info

- FASTQ.GZ files containing raw sequences.

  Reads will be provided with adapter sequences masked. No quality clipping is provided with raw reads delivery.

ChIP-Seq bioinformatics analysis

- Alignments in BAM format.
- A list of called peaks with information on peak's position, length, height and fold enrichment.
- Peak annotation, *i.e.* association of ChIP regions with functionally important genomic features (chromosome, promoters, gene bodies, or exons). *Provided only for human hg19, mouse mm9, fly dm3 and worm ce6 reference.*
- Comparative ChIP-seq analysis, *i.e.* identification of sites that are differentially bound between two sample groups (conditions or tissues).
- BedGraph format files allowing for data visualization in a Genome Browser such as UCSC.
- Metrics describing overall experiment quality computed from the BAM file (library insert size, read distribution along the genome, % of aligned reads, *etc.*).
- A REPORT file describing the library preparation and analysis flow.

FAQ

Do reads contain adapters?

Unless differently agreed, reads are provided with masking of adapters read-through. When a minimum of 5bp read-through is found with respect to sample-specific (barcode included) adapters, bases are masked with N character. Thus, read length is maintained to its original size. No quality clipping is applied on raw reads delivery, while regularly used in our standard bioinformatic pipelines.

Are reads quality trimmed?

Delivered raw data are not quality trimmed. However, our internal analysis pipelines always rely on a quality trimming step which will be described in the delivery REPORT.

Which is the optimal sequencing depth in ChIP-seq experiments?

An important consideration in experimental design is the minimum number of sequenced reads required to obtain statistically significant results. The number of produced reads, *i.e.* the required sequencing depth, depends on the nature of the mark and the state of the cell in each experiment. However, you can find some good guidelines here (https://bit.ly/2SafMrp). Jung *et al.* observed that sufficient depth is often reached at <20 million reads for fly, while for human they suggest 40-50 million reads as a practical minimum for most marks.

How to treat duplicated reads in ChIP-seq experiments?

The most "politically correct" solution is the MACS2's one. If the read length parameter is set to zero, MACS2 detects read length automatically and proceeds to filter out duplicate reads. By default, it calculates the maximum number of duplicate reads in a single position warranted by the sequencing depth and removes redundant reads in excess of this number. Alternatively, you can select to keep only one read, or all duplicates.