Video 3: Evaluating the quality of your ChIP-seq data
How to evaluate a ChIP-seq run?

• Key Quality Control Steps:
  – Complexity
  – Control sample
  – Measurement of global ChIP enrichment
  – Cross-correlation

Resource

ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia
Complexity
Complexity

- Non-redundant fraction

\[ \text{NRF} = \frac{\text{#unique start positions of uniquely mappable reads}}{\text{#uniquely mappable reads}} \]

- Decreases with sequencing depth
- Recommend > 0.8 for 10 million uniquely mapped reads
Control Sample

- Two types of control sample:
  - “Input” DNA
    - Sonicated DNA fragment w/o ChIP
  - A “mock” ChIP
    - ChIP using non-specific IgG
    - More closely mimics ChIP experiment
    - Must recover enough DNA to gain complexity of the library
Measurement of global ChIP enrichment

• A useful and simple first-cut metric for the success of the immunoprecipitation

• FRiP (fraction of reads in peaks)
  – Calculates the fraction of all mapped reads that fall into peak regions identified by a peak-calling algorithm.

• 1% threshold is a general guideline

• FRiP is very useful for comparing results obtained
  – with the same antibody across cell lines
  – with different antibodies against the same factor

• Affected by algorithm used for peak-calling
Cross-correlation analysis

Highly enriched ChIP

Poorly enriched ChIP
Cross-correlation analysis

- Sequence tag density accumulates on “+” and “-” strands centered around the binding site.
- It is computed as the Pearson correlation between the Crick strand and the Watson strand, after shifting Watson by k base pairs.
Cross-correlation analysis

• (Normalized Strand Coefficient) NSC and (Relative Strand Ratio) RSC are strong metrics for assessing signal-to-noise ratios in a ChIP-seq experiment.

• Consider re-do ChIP if NSC < 1.05 and RSC < 0.8

\[
NSC = \frac{cc(\text{fragment length})}{\text{min}(cc)}
\]

\[
RSC = \frac{cc(\text{fragment length}) - \text{min}(cc)}{cc(\text{read length}) - \text{min}(cc)}
\]