Interaktionen und Modifikation von RNA und Proteinen

Prof. Sonja Prohaska

Computational EvoDevo
Universität Leipzig

SS 2018
ChIP-seq

Chromatin Immuno-Precipitation followed by sequencing is a next-generation sequencing (NGS) technology that retrieves genomic positions of protein - DNA interaction.

- **Transcription factors**
  - transcription factors
e.g. estrogen receptor, homeodomain proteins, etc.
  - other protein-based DNA and transcription regulators
e.g. CTCF, nuclear membrane attachment proteins, etc.

- **Histone modifications**
  - modified nucleosomes/histones
e.g. acetylation, methylation, etc. on specific sites of histones such as H3K4me3, H3K27ac, etc.

Different protocols during experiment and data analysis.
ChIP-seq Conceptual Lab Workflow

▶ ChIPseq (Chromatin Immuno-Precipitation followed by sequencing)
▶ proteins are cross-liked to DNA in vivo
▶ chromatin is isolated
▶ sonicate to obtain chromatin/DNA fragments in length range 200-300nt
▶ immunoprecipitate with specific antibodies for target protein (modification)
▶ purify immunocomplexes (remove chromatin fragments not bound by the antibody)
▶ reverse cross-linking
▶ purify DNA from chromatin fragments
▶ get fragments sequenced
ChIP-seq Conceptual Lab Workflow

1. Cross-link cells with formaldehyde. Isolate genomic DNA and sonicate to shear chromatin.

2. Add an antibody specific to the protein of interest.

3. Perform immunoprecipitation to isolate DNA bound by the factor of interest. Reverse cross-links and purify isolated DNA.
ChIP-seq Analysis Workflow

Shown are the steps at which bioinformatics is needed.
ChIP-seq Analysis (general)

- **Sequencing**
  steps shown here are based on sequencing with Illumina

- **Read mapping**
  aligning reads against the genome (“genome alignment”)

- **Peak calling**
  calculate read density
  compare to/normalize with control/background
  commonly ‘input DNA’ or ‘IgG’
  derive peaks, i.e. regions significantly enriched in reads

- compile list of enriched regions

- **data visualization** e.g. as genome browser track
ChIP-seq Analysis for Transcription Factors (TFs)

(0) 5' ends of fragments are sequenced

(1) Short reads are aligned

(2) Short reads are aligned

(3a) Distribution of tags is computed

(3b) Profile is generated from combined tags

Reference genome

Peak identification can be performed on either profile

Figure 5 | Strand-specific profiles at enriched sites.
ChIP-seq Analysis for Transcription Factors (TFs)

(0) a ChIP-seq experiment returns sequence fragments (200-300nt)

(1) **Sequencing**: with Illumina 5’-ends (35-50nt) are sequenced → **reads**

(2) **read mapping** (alignment of reads to the reference genome)
   5’-ends of plus (blue boxes) and minus (red boxes) strands are mapped
   and read distributions are calculated

(3) **peak calling** identification of peaks (against a background)
   
   a) the peak on the plus strand and the peak on the minus strand
      is called → the peaks flank the target binding region
   
   b) after conceptual extension of the reads a single peak is called
      → the peaks cover the target binding regions

▶ target binding region (100-200nt) vs. target binding site (4-12nt)
▶ apply a **motif discovery** tool to derive the binding motif
   from the set of binding regions
References
