ChIP-seq annotation and visualization

Global Objective

Given a set of ChIP-seq peaks annotate them in order to find associated genes, genomic categories and functional terms.

Data set

For this practical session, the ChIP-seq data and peaks related to following publication will be used: “GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility.”, Theodorou *et al* (http://www.ncbi.nlm.nih.gov/pubmed/23172872).

1. Plot peak agglomeration around TSS

Using the Galaxy tool “makeTSSDist” we will plot the agglomerative ChIP-seq enrichment around the TSS. One simple solution is to plot the distribution of the distance between peaks and TSSs (relative distance between a peak summit and the TSS of the closest annotated genes).

Procedure

1. Select [siGATA_ER_E2_r3_MACS_peaks] peaks
2. use the default parameters and run the software

Question: discus the distribution properties (e.g. modes, mean).

2. Relate peaks to genes

In this step, we will try to associate gene names to peaks that are close to known gene annotations. Using the Galaxy tool “AnnotatePeaks” annotate the peaks for 2 replicates (WT). Extract associated genes and compare the results.

Question: How many genes are in common between the replicates?

Without filtering the peaks, the number of artifactual peaks can be relatively high. Filter the peaks using a cutoff on the FDR (<5%) and the fold entichement (>4) and perform again the annotation.

Question: How many genes does remain after this filtering step?

3. Visualize ChIP enrichment around a given feature

Using the “deepTools heatmapper” we will try to visualise the local enrichment around the TSS for all known genes. Before drawing the heatmap we need to prepare the data by computing a summary matrix of the local ChIP enrichment using “deepTools computeMatrix”.

Procedure

1. Download the required annotation file (here all UCSC annotated genes) from the UCSC genome browser (Table Browser, UCSC Genes as bed file).
2. Use the obtained annotation file and the previously computed bigWig H3K4me1 profile as an input to “computeMatrix”.
3. Use 'reference point’ as the output option and ‘beginning of region’ as this reference point (TSS).
4. Using the “heatmapper”, load the obtained matrix data and fill the desired options to plot the heat map.

4. Relate peaks to GO terms

For that specific step we will use the GREAT annotation tools. Connect to GREAT web server (http://great.stanford.edu) and perform a GO annotation for the ESR1 peaks. Alternatively GREAT can be launch directly from UCSC web server (using Table browser Custom track and by selecting send to GREAT).

** Procedure**

1. Connect the GREAT (http://great.stanford.edu) web server
2. select the genome assembly version (hg19)
3. upload or paste the peaks obtained previously in BED format

4. use the whole genome as background and run the software

Question Examine the enriched functional categories.

5. Integrative ChIP-seq analysis of regulatory elements

In this part, we will use the ReMap ReMap (http://tagc.univ-mrs.fr/remap/index.php) software to compare the peaks obtained in the peaks calling tutorial to an extensive regulatory catalog of 8 million transcription factor binding sites.

** Procedure**

1. Connect the ReMap (http://tagc.univ-mrs.fr/remap/index.php) web server
2. Go to the Annotation Tool
3. upload or paste the peaks in BED format (select BED format in the data format selector)
4. Add your email and run the software with default parameters

Question: What are the TFs associated to your peaks?