**BIOM262: ChIP-Seq workshop – Part 3**

We will need interactive shell, so please start by typing:

***qsub -I -l nodes=1:ppn=2 -l walltime=3:00:00 -q hotel***

Here we will explore IGVtools, use HOMER for studying NDRs (nucleosome depleted regions), and follow a tutorial for bedtools.

**IGVtools:**

Creating tdf files:

We will need to sort the bam files as well as index them.

We will use data from Javasky & Shamir *et al*. The bam files for chromosome 12 are located at

~/biom262\_2019/Module\_5/workshop\_3/data/

To sort the files, create the ‘sorted’ directory and run:

***for f in <path>/data/\*bam; do fname=`basename $f -chr12.bam`; samtools sort -o*** *<path****>/sorted/$fname.sorted.bam $f; done***

***for f in*** *<path>/****sorted/\*bam; do samtools index $f; done***

and create the ‘tdfs’ directory followed by running:

***for f in*** *<path>/* ***sorted /\*bam; do fname=`basename $f .sorted.bam`;***

**igvtools count -e 164 $f** *<path>****/tdfs/$fname.tdf hg19; done***

Copy the tdfs to your computer. Before loading to IGV, make sure you have selected in view -> preferences -> tracks the “normalize coverage data”. (details from here:

<https://software.broadinstitute.org/software/igv/Preferences>)

*Select to normalize tracks containing coverage data in .tdf files that were created using igvtools. This normalization option multiplies each value by [1,000,000 / (totalReadCount)].*

**Plotting the NDR signals**

The data from Javasky & Shamir *et al*. is derived from human cells, so we first should add to HOMER the needed genome:

**perl** <path; mine was at ~/software/homer/>/configureHomer.pl **-install hg19**

The bam files for chromosome 12 are located at

~/biom262\_2019/Module\_5/workshop\_3/data/

Since Homer will want to access the files directly, make sure you copy the data directory to your account in a meaningful directory:

*cp -R ~/biom262\_2019/Module\_5/workshop\_3/data <location in your directory>*

(Note, if the above does not work due to limited space, look for large data sets in your account you no longer need and can remove. Usually It’s *.sam* files)

Now, let’s make a tag-directories for all files (as we did before):

***for f in*** *<path to your bam files>****/\*.bam; do fname=`basename $f -chr12.bam`; makeTagDirectory <****path to where your tag directories will be****>/$fname -genome hg19***

***-checkGC $f; done***

Once we have the tag-directories, we can use the annotatePeaks.pl command

***annotatePeaks.pl tss hg19 -size 4000 -hist 10 -d*** *<*path to the tagdirs>/\*k27\* > <path>/h3k27ac-hist.txt

and

***annotatePeaks.pl tss hg19 -size 4000 -hist 10 -d*** *<*path to the tagdirs>/\*k4\* > <path>/h3k4me3-hist.txt

Open the files using a spreadsheet program (R, Excel or similar). As we discussed, the first column gives the distance offsets from the TSS followed by columns corresponding to the ‘coverage’ (calculated by extending tags by their estimated ChIP-fragment length; analogous to the profiles made for the Genome Browser), ‘+ Tags’, and ‘- Tags’ (density of 5’ and 3’ aligned tags; these are independent of fragment length) for each experiment.

Plot the coverage of mitosis and interphase H3K27ac and H3K4me3 – do you see a difference in the way they are organized around the TSS?

**BEDtools:**

We will follow the tutorial by Aaron Quinlan

<http://quinlanlab.org/tutorials/bedtools/bedtools.html>

You can browse the different sub-commands here <https://bedtools.readthedocs.io/en/latest/content/bedtools-suite.html>