Short Read Workshop – ChIP-sequencing

- Author: Jessica Huynh-Westfall (jessica.westfall@colorado.edu)

Homework

In the workshop we used the analysis tool MACS2 to peak call on BACH1 transcription factor ChIP-seq sample. In class we used a control sample and experimental sample to call peaks using MACS2. As we briefly mentioned, we can also use the --broad flag to call broad peaks. Let's look at some other options and ideas.

1. There is another dataset for GABPA transcription factor. We have already processed the data and provided you with the bam file. Make a script to use MACS2 to peak call GABPA. There is a background ChIP-seq control sample that you can also use.
2. Try different flags with your script a. Call peaks without background control b. Call peaks and clean up the output with Blacklist c. Call peaks with --broad flag
3. After you have all the different types of peak calls, open up X2go and IGV. View the raw bam file (this file is large and may be slow to load) and your peak call files in one viewer.
4. Look around the genome (note this is subsample and only chr21) and ask yourself the following: a. Did your peak without the background control call more/less peaks for their respective TF dataset? b. Do you see any regions that would be considered a repetitive region (ie., on the Blacklist)? c. Do the broad peaks look different than narrow peaks for the majority of calls? How are they different?
5. In the datafile directory, we have included a motif bed file for GABPA (../sread/data_files/day9/motifs) from HOCOMOCO. Load that into IGV. Did your peak calling script do a good job of calling peaks based on the motif bed? Hint: How long is the GABPA motif? Do you expect every motif to have a peak associated with it? Do you expect all peaks to have a motif overlapping? What are the p-adjusted values of peaks called without GABPA canonical motifs?