October 25, 2017

CSHL PFB Project Proposal

Goal: Using ChIP-Seq data for two different conditions (or time points) assess whether a specific chromatin peak shows any difference in width/spread between the two samples.

Proposed steps:

Select .bed files for two different time points (T=0, T=2) for a particular ChIP (a specific transcription factor or histone modification e.g. H3K27ac or RNAPolII)

File format: <chr> <start> <end> < peak identifier or other info >

1. Select only chr1 data for T=0 and T=2.

Create a ‘for loop’ and only work with lists starting with ‘chr1\s’

1. Using import os, ‘bedtools –intersect’ run a script on unix to determine which of the ChIP-Seq peaks for T=0 overlaps with T=2. Another option is to run the program pybedtools within python to extract the intersecting genomic regions. Perform the comman with both ‘wb’ and ‘-wa’ options which will output the original entry in T=2 and for each overlap and the original entry in T=0 for each overlap.

<http://bedtools.readthedocs.io/en/latest/content/tools/intersect.html>

https://daler.github.io/pybedtools/intersections.html

1. Count how many instances of overlap there are between the two datasets for Chromosome 1.
2. From the wb\_intersect\_output and the wa\_intersect output, perform a calculation ‘<end> - <start>’ to determine the length of each intersected peak for T=0 and T=2.

\*Some peak start and ends will not match up between the two datasets. Will need to

1. Set up an ‘if statement’ within a ‘for loop’ to iterate out those peak lengths that are larger in T=2 than in T=0.
2. Write these files into a new file in .bed format

<chr> <start> <end> < X >

1 54 62

1 34 62

1 20 40

1 40 72