Manasa Rapuru
BF 528
05/12/23

Individual Project
Roles Chosen: Programmer from Project Two and Biologist from Project Four

Introduction:

The technology of focus in project three was on ChIPseq Analysis. ChIP-seq analysis, short for Chromatin Immunoprecipitation Sequencing, uses protein-DNA complexes to target and identify DNA binding sites across the genome that interact with the protein of interest.

The authors of this study used in project three, RUNX1 Contributes to Higher-Order Chromatin Organization and Gene Regulation in Breast Cancer Cells, focus on observing the relationship between RUNX1-mediated transcription and the organization of the genome through CHIP-Seq analysis. Runx1 functions by altering chromatin structure by incorporating chromatin modifier enzymes. The tasks I have chosen to perform for this individual project are the roles of the analyst and biologist from project three.

Methods:

Analyst:

The analyst in this project was responsible for doing correlation analysis using bigWig files and performing further visualization tasks. The tools used in this section come from DeepTools: a suite of tools developed to facilitate the analysis of high throughput sequencing data including ChIP-seq (deepTools: Tools for Exploring Deep Sequencing Data — deepTools 3.5.0 Documentation, n.d.). Using DeepTools, the tasks include: to create a heatmap that shows the clustered correlation metrics between samples, coverage tracks and signal coverage plot across TSS and TSS of all hg19 genes.

This task requires using different utilities in the DeepTools module. The first utility that is used is bamCoverage. The purpose of this utility is to make a coverage track file in the format of a bigwig file using BAM as input. A bigwig file was made for 4 sorted BAM files. The sorted bam files for first replicate input, second replicate input and second replicate IP sample were already provided in "/projectnb/bf528/project_3_chipseq/provided_files/" directory. Sorted file for first IP sample was obtained from the data curator's files.

The next step is to use the multiBigWig Summary utility. The purpose of this utility is to compute the average of the genomic scores found in all of the input bigwig files. The output of the multiwig utility is then put into the plot correlation utility to make a correlation plot using the Persaon correlation. The next step is to calculate the scores of each genomic region. To specifically get the scores for genes that are around and between TSS and TTS I used a bed file

that marked these regions in the input arguments. The bed file was also manually generated by going to UCSC Table Browser and inputting the settings that were provided in the instructions to list the TSS/TTS locations for every gene in the hg19 reference. Lastly, plot Profile was utilized to make the visual.

Biologist:

The role of the biologist in this project was to draw conclusions on whether the transcription factor is likely directly involved in regulation of a selected gene. The steps to these were as follows. First the results of DESeq were downloaded using the GEO accession number GSE75070. Next, using the annotated peaks file generated by the programmer, a stacked bar plot was generated to calculate the proportion of differentially expressed genes that were proportionate. Lastly, a heat map of HI-C data. Additionally, it was the role of the analyst to load the bigwig files for the input and the IP samples into IGV and assess if the results for the binding of RUNX in the genes of NEAT1 and MALAT1 are similar to the ones reported in the paper.

Results:

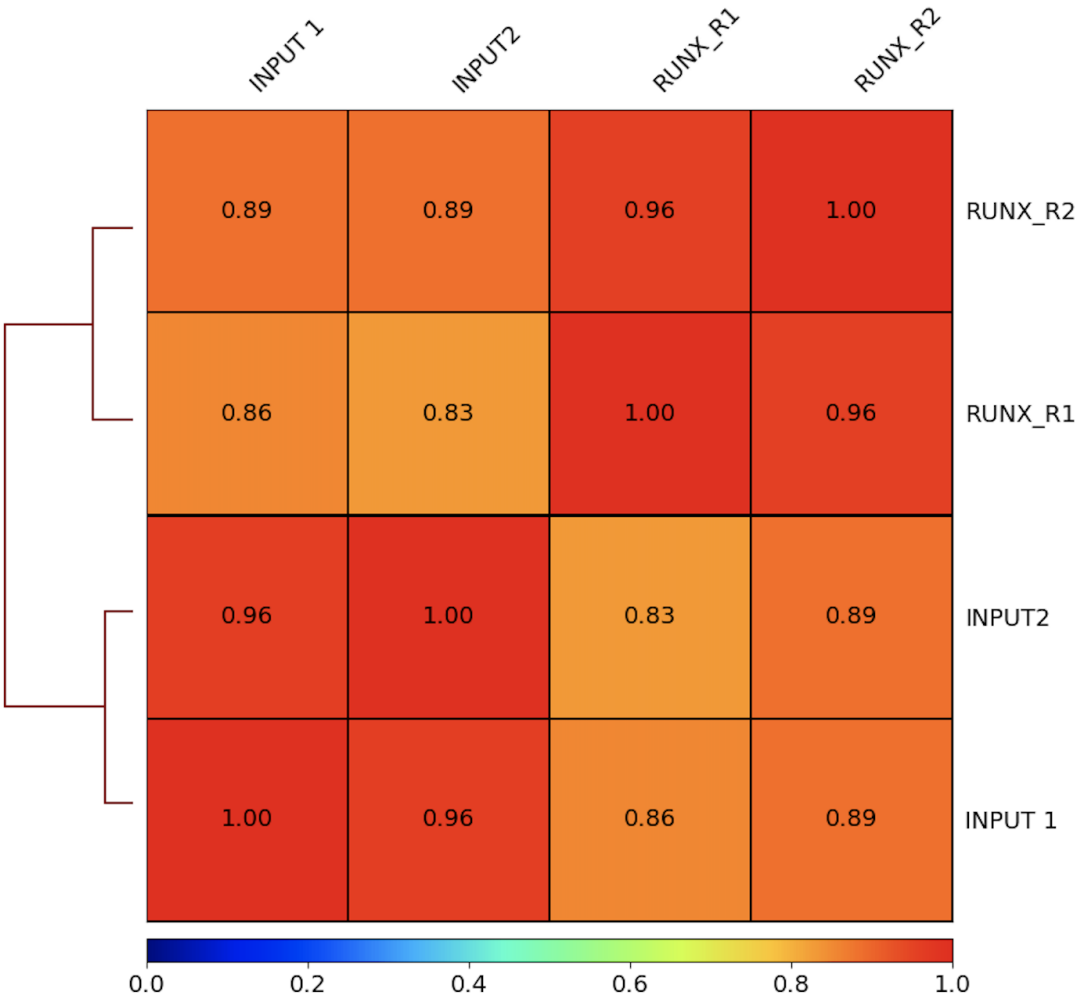

Figure 1: This heat map of cluster correlation with input 1 and 2 and IP samples 1 and 2. As seen here the input samples have higher correlation values between each other in comparison to correlations with the IP samples. The same goes for the IP samples having higher correlation values with each other than they do with just the input samples.

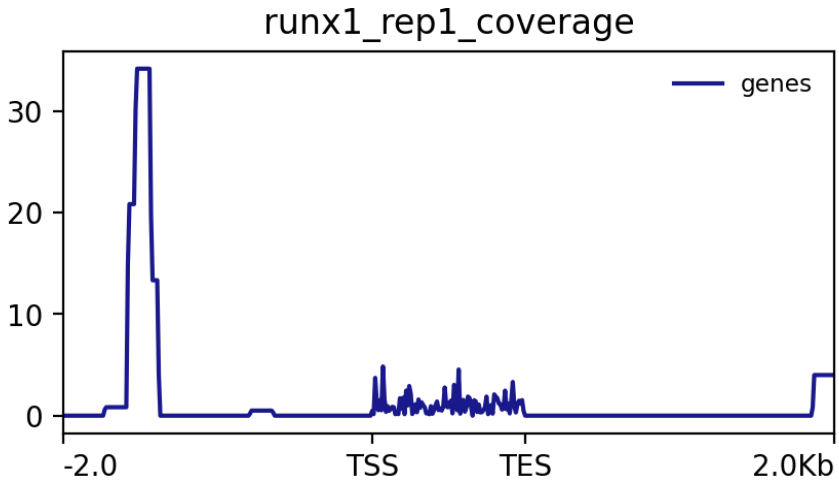

Figure 2: This plot indicates the signal coverage of Runx1 across the h19 genes with TSS and TTS references in IP replicate number one. It is seen in this graph that the highest amount of enrichment is found before the TSS region on this span of 2kb up and downstream of TSS and TES region. Highest signal of RUNX1 is seen upstream of TSS

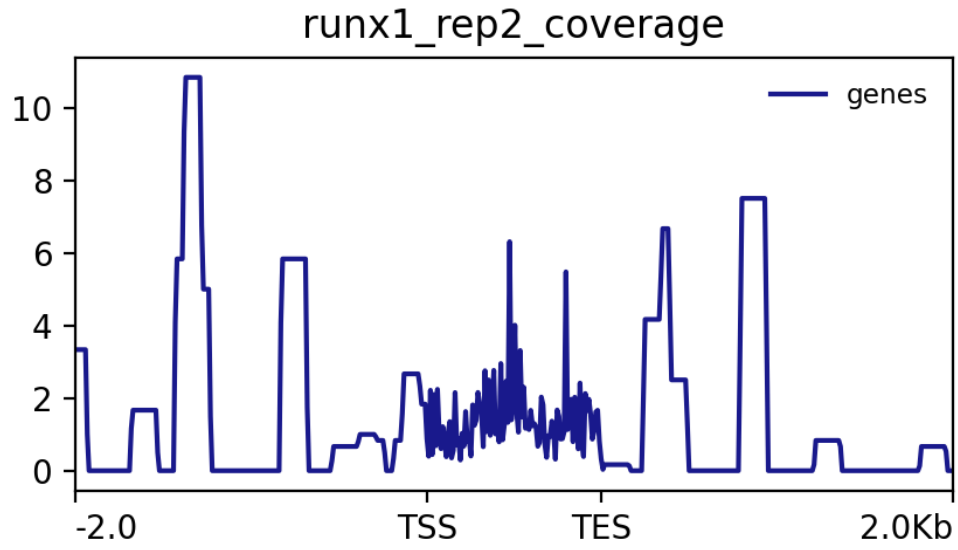

Figure 3: This plot indicates the signal coverage of Runx1 replicate two across the h19 genes with TSS and TTS. Here again highest signal is found upstream of TSS

Figure 4: Shows the percentage of Bound and non bound differentially expressed genes that have a peak within plus or minus of TSS. This figure is to replicate figure 2f from the paper. It is to illustrate how many up regulated and down regulated differentially expressed genes are bound with RUNX1. Figure 2F in from the paper illustrates the results the authors found. After comparing the data that was generated with our group data with the data generated in the paper, it is seen that our data has yielded more of up regulated and down regulated genes (129, and 84 respectively) than what the study found(59, and 48 respectively).

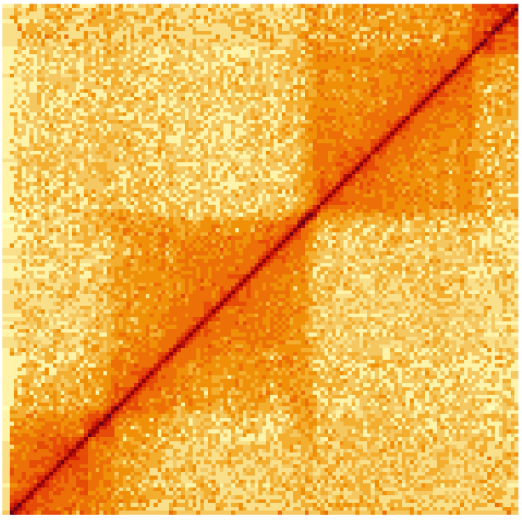

Figure 5:Illustrates HI-C data in the first 5.3 megabases of Chromosome 10

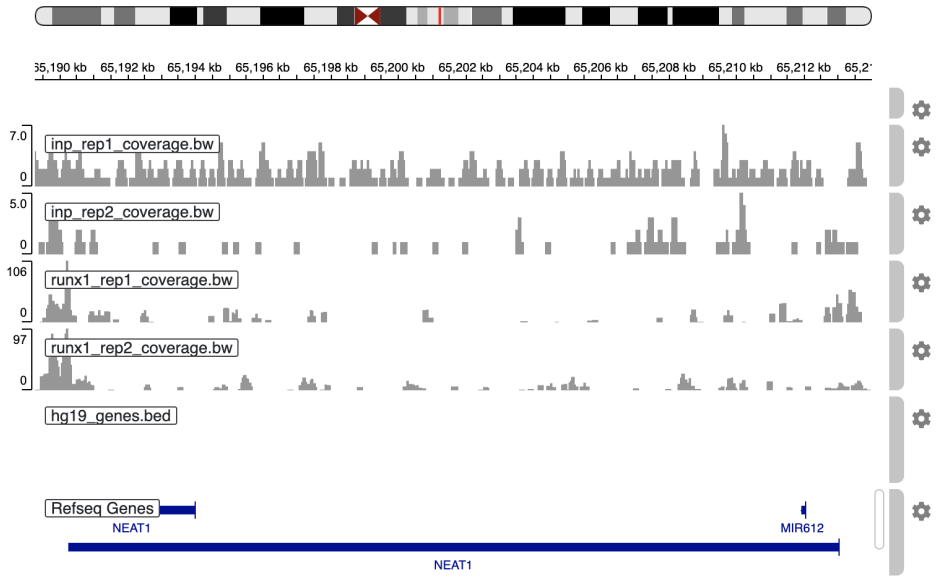

Figure 6: Visualization of NEAT1 TSS1 IGV in all four samples. The overall results seen here seem to be fairly consistent with the results presented in figure 2E of the paper.

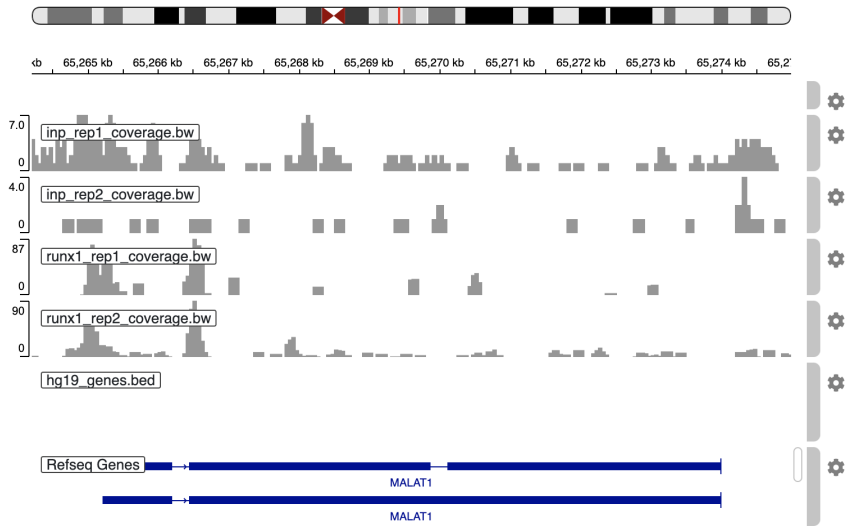

Figure 7: Visualization of MALAT1 TSS1 IGV in all four samples. The overall results seen here seem to be fairly consistent with the results presented in figure 2D of the paper.

Discussion:

The roles of the analyst and the biologist in this ChIP sequence experiment are key to understanding the relationship between RUNX1-mediated transcription and the organization of the genome through CHIP-Seq analysis. Figures 2 and 3 show that in a section of the genome between 2kb upstream and downstream of TSS and TES, the highest enrichment of RUNX1 is seen before TSS providing evidence RUNX1 works with transcription factors and is needed in gene expression. Figures 4, 5, 6 and 7 serve to show examples of RUNX1 binding on to promoters of MALAT1 and NEAT. Overall the results obtained from both of these roles show evidence that RUNX1 plays a role in the regulation of certain genes of interest.

References:

Barutcu et al. RUNX1 Contributes to Higher-Order Chromatin Organization and Gene Regulation in Breast Cancer Cells. Biochimica et Biophysica Acta 1859 (11): 1389–97. PMID: PMID: 27514584

deepTools: tools for exploring deep sequencing data — deepTools 3.5.0 documentation. (n.d.). https://deeptools.readthedocs.io/en/develop/“