BF528
Final report
Project 3 Analysis and Biologist
Anna McNiff

ChIPseq analysis of the human transcription factor Runx1

1. Introduction

In "RUNX1 contributes to higher-order chromatin organization and gene regulation in breast cancer cells" by Barutcu et al, the authors use a combination of RNA Sequencing, Chromatin Immunoprecipitation Sequencing(Chip-Seq) and High-throughput chromosome conformation capture technique (Hi-C) to investigate the role of runt-related transcription factor 1 (RUNX1). Both expression of wild-type and mutated RUNX1 has been linked to different leukemias. Mutations resulting in the inactivation of RUNX1 have been found in patients with myelodysplastic syndrome (MDS) and cytogenetically normal acute myeloid leukemia (AML). Wild-type un-mutated RUNX1 expression has also been shown to play a role in myeloid leukemogenesis(Schibler et al, 2013). Additionally RUNX1 has been demonstrated to be both down- and up-regulated in breast cancer (Barutcu et al, 2016). In this study the authors expanded upon their previous findings to investigate the impact of RUNX1 expression and knockout on the chromatin structure of the surrounding region.

2. Methods

2.1 Pearson Correlation Matrix and Heatmap Generation

The bamCoverage utility in DeepTools (version 3.5.1) was used to generate a bigwig file from sorted, indexed .bam format sample files. This was done with the 'normalizeUsing CPM' parameter to normalize Counts Per Million mapped reads. In addition a bin size of 10 and the 'extend reads' parameters were used, in accordance with the DeepTools documentation's recommendations for Chip-Seq analysis. A compressed matrix file of the processed bigwig files was generated from the multiBigWigSummary utility. This matrix was used to create a clustered heatmap of the Pearson correlation values between all the samples using the multiBigwigSummary and plotCorrelation functionalities of the DeepTools module(Figure 1).

2.2 ChiP-Seq Signal Expression

A BED file for the hg19 build of the human genome was obtained from the UCSC genome browser. The bigWig files for each IP sample and the hg19 genes BED file was run under the computeMatrix utility in DeepTools. For each replicate, the computeMatrix tool was run in scale-regions mode including 2kb windows up and downstream of the Transcription Start Site(TSS)

and triplex target DNA sites(TTS).ChiP-Seq signal coverage for each replicate was plotted with default parameters using the plotProfile utility(Figures 2-5).

2.3 Identification of Differentially Expressed Genes

The DESeq2 data was filtered in R using the current versions (as of 5/10/2023) of the dplyr, tidyr, ggplot2 and tibble modules. Filtering thresholds were in accordance with those reported in the original study. Genes with an adjusted p-value of greater than .01 were pruned from the dataset and removed from consideration in further analysis. Genes were identified as up-or down regulated based upon their log2 fold change, with up-regulation defined as genes exhibiting a log2 fold change of greater than 1, and down-regulation defined as genes exhibiting a log2 fold change of less than -1.

2.4 Visualization in the IGV

The bigwig files generated in 2.1 were uploaded to the Integrative Genomics Viewer (IGV) for visualization. Additionally, a file containing peaks annotated as present in both replicates was uploaded. Peaks were annotated as present in both replicates if, after quality control and filtering against a blacklist they intersected. Intersection was established using the default parameters of the Bedtools intersect module.

2.5 Heatmap

The processed HI-C data was downloaded from the GEO Accession GSE75070. The data set was cropped to specifically focus on chromosome 10. The data was log base 2 transformed, written to a matrix and mapped as a heatmap using the heatmap utility in R with default parameters.

3. Results

3.1 Pearson Correlation Heatmap

Pearson correlation of samples and replicates identified the highest correlation between samples of the same experimental cell line. This is in line with our expectations and the results reported in the original study by Barutcu et al.

3.2 ChiP-Seq Signal Profiles

The Runx1 knockdown replicates exhibit similar ChiP-Seq signal profiles. The control inp replicates, however, exhibit significantly different profiles. This could be attributable to mishandling in data processing of the control inp replicates in prior stages of analysis.

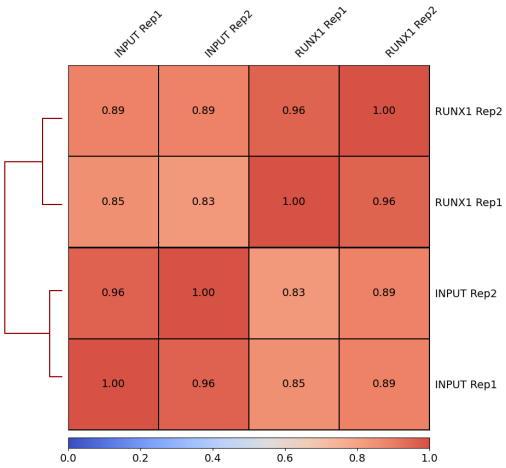
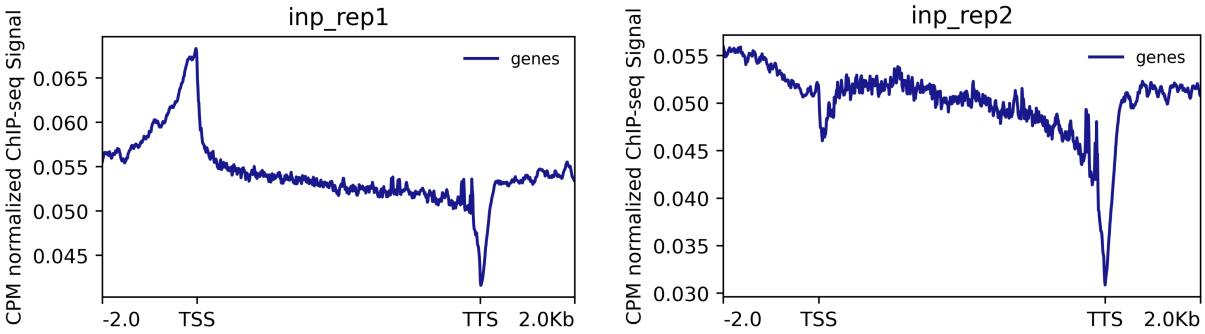
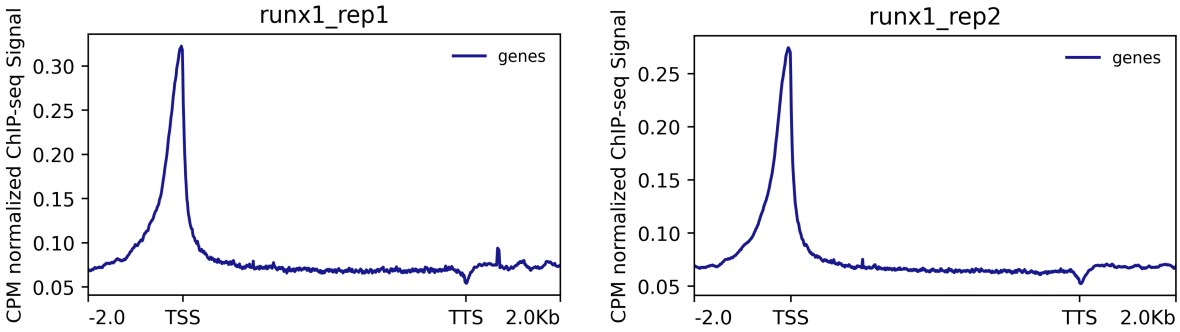


Figure 1. Correlation Plot of Samples and Replicates



Figures 2 and 3. 2. ChiP-Seq Signal Profile control inp Replicates 1 and 2, respectively.



Figures 4 and 5. 4: ChiP-Seq Signal Profile Runx1 Replicates 1 and 2, respectively.

3.3 Differential Gene Expression

Filtering the DESeq2 file, 3562 genes passed filters and 687 were identified as up regulated and 466 as down regulated. 627 genes were found to have a distance to TSS +/- 5kb. Of these genes, 29 up

regulated and 14 downregulated genes which were also present in theDESeq2 results were identified. 6.86% of genes in this set were annotated as bound and the remainder were considered unbound.

Filtering for distance to TSS +/- 20kb, 246 genes were identified. Of these, 11 upregulated and 9 downregulated genes which were also present in the DESeq2 results. 8.13% of genes in this set were annotated as bound and the remainder were considered unbound. These proportions represent an under-estimation of the true numbers of bound genes, as only the provided subset of DESeq2 data was used.

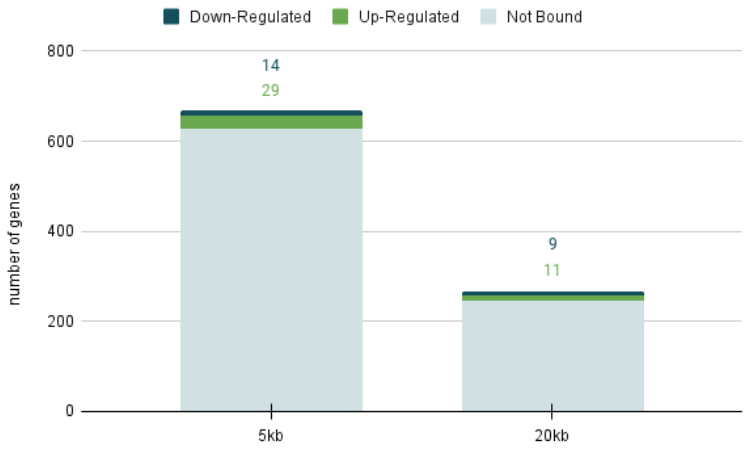


Figure 6. Barchart comparison of Bound and Unbound genes within 5kb and 20kb of the TSS. In this figure genes annotated as Down or Up regulated are bound by RUNX1.

3.4 Heatmap of Hi-C interactions.

Figure 7 depicts a heatmap of the pairwise interaction matrix generated from the Hi-C data as described in 2.5 for a portion of the p arm of chromosome 10. Direct comparisons to the original study are complicated by the cropping and rotating transformations performed on the figure in Barutcu et al. In this heatmap the x and y axes are positions in chromosome 10, with each bin spanning approximately 4000 base pairs. The heatmap itself depicts the interactions between these locuses as measured by the Hi-C data. We are able to observe several regions with strong Hi-C interactions across chromosome 10.

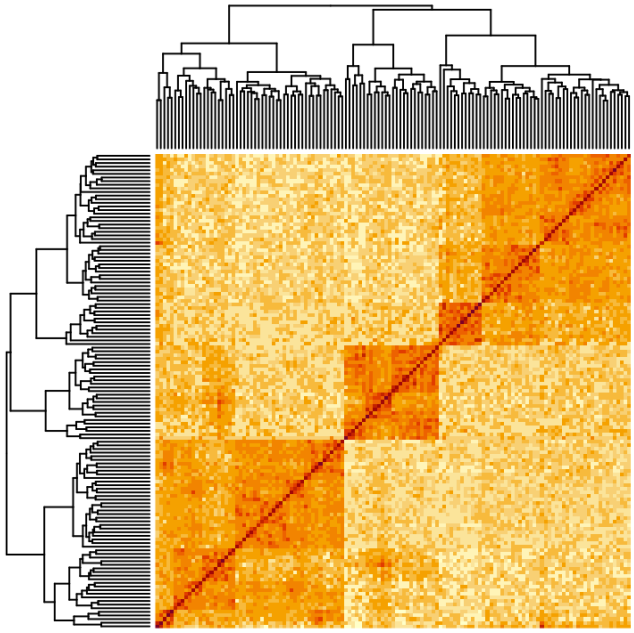


Figure 7. Heatmap of the pairwise interaction matrix from the Hi-C data

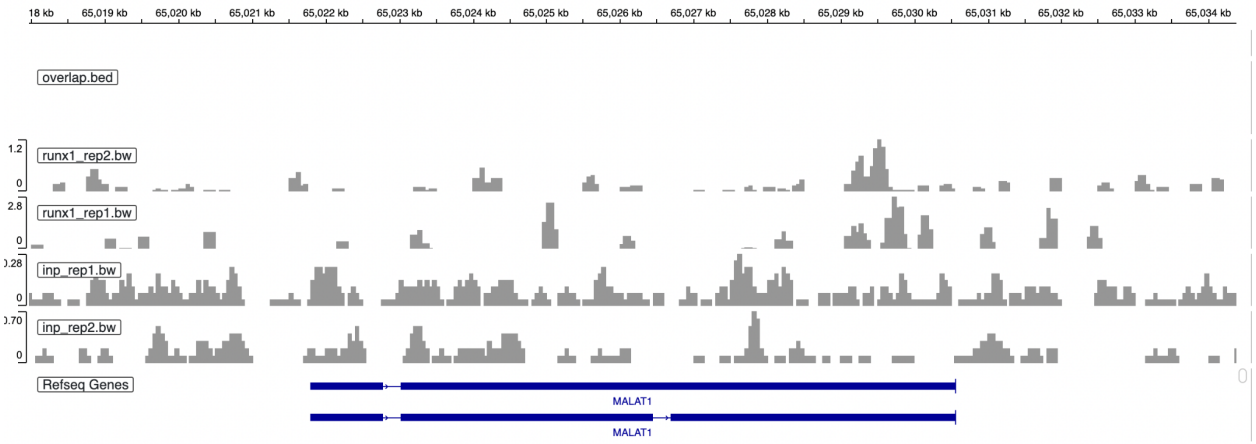


Figure 8. Chip-Seq binding in IGV genome browser around MALAT1.

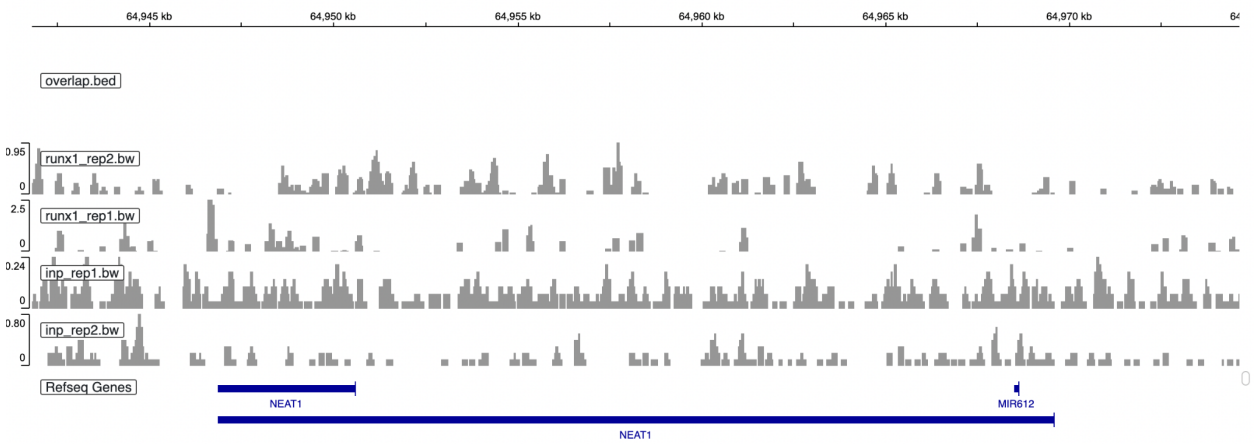


Figure 9. Chip-Seq binding in IGV genome browser around NEAT1.

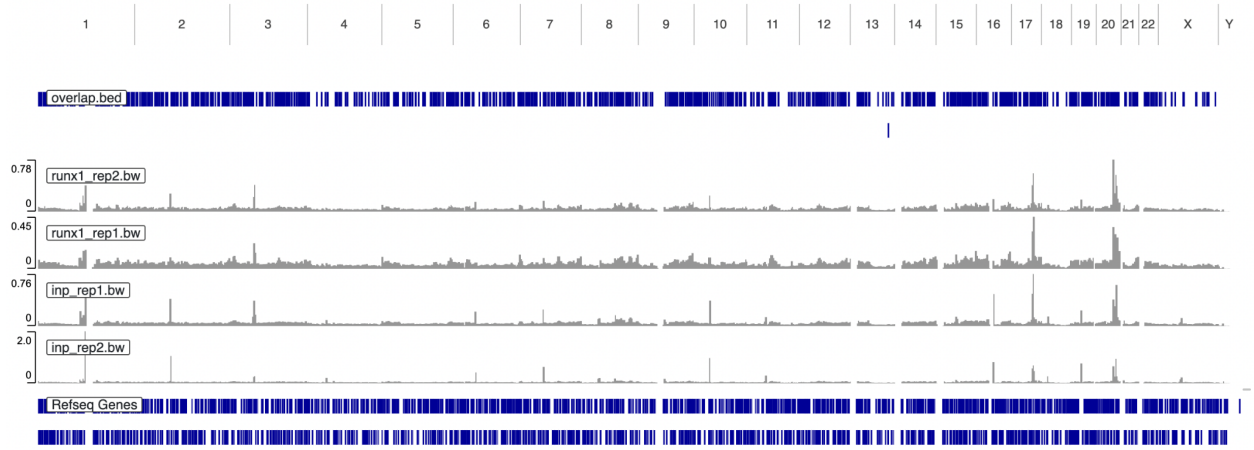


Figure 10. Chip-Seq binding in IGV genome browser across the entire genome. The Overlap.bed track at the top of the figure represents peaks annotated as present in both replicates of both cell lines.

3.5 Chip-Seq expression in IGV genome browser

Chip-Seq binding was visualized in the IGV genome browser. In proximity to the MALAT1 gene the control inp replicates show higher levels of expression than the runx1 replicates, in line with results reported by Barutcu et al(Figure 8). In addition, in proximity to the NEAT1 gene the control inp replicates show higher levels of expression than the runx1 knockdown replicates, also in line with results reported in the original studyl(Figure 9). This in conjecture with other findings indicates that RUNX1 binds at the promoter sequences of these two genes and positively affects their transcription and subsequent expression. Both of these genes have been shown to be related to extracellular matrix components, which are affiliated with the tumor microenvironment(Nallanthighal et al, 2019). Figure 10 displays Chip-Seq expression and peaks annotated as present in both replicates of both experimental cell lines across the entire genome.

4. Discussion

In this study Chip-Seq expression in cell lines with reduced RUNX1 was compared with control cell lines. This study demonstrated that preventing the expression of RUNX1 resulted in reduced Chip-Seq expression in the surrounding region, supporting Barutcu et al's hypothesis that transcription factor RUNX1 plays an important role in the regulation of this region. This was further supported by the comparative analysis of RNA-seq data which identified genes up and down regulated during RUNX1 expression and knockout.

The RUNX1 transcription factor has been previously shown to be associated both with tumor suppression and proliferation. This study implemented several methods of bioinformatic analysis to investigate alterations in RNA expression and RUNX1 binding caused by RUNX1 deactivation. The CHip-Seq analysis performed in this study indicates that RUNX1 impacts chromatin interactions across the genome, with particular impact on chromosome 10. These RUNX1-mediated interactions are consequential in the regulation of gene transcription and expression. Manifold regulatory-affiliated chromatin interactions have been shown to be impaired or deactivated in cancer cells, further underscoring the relationship between transcriptional regulation and oncogenesis. Additionally, visualization of Chip-Seq binding in the IGV provided confirmation of RUNX1 binding at the promoter sequences of MALAT1 and NEAT1 further demonstrating the affiliation of RUNX1 with the mediation of cancer proliferation.

Resources:

Barutcu, A. R., Hong, D., Lajoie, B. R., McCord, R. P., van Wijnen, A. J., Lian, J. B., ... & Stein, G. S. (2016). RUNX1 contributes to higher-order chromatin organization and gene regulation in breast cancer cells. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1859(11), 1389-1397.

Goyama, S., Schibler, J., Cunningham, L., Zhang, Y., Rao, Y., Nishimoto, N., ... & Mulloy, J. C. (2013). Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. The Journal of clinical investigation, 123(9), 3876-3888.

Nallanthighal, S., Heiserman, J. P., & Cheon, D. J. (2019). The role of the extracellular matrix in cancer stemness. *Frontiers in cell and developmental biology*, *7*, 86.