ChIP-seq Data Analysis Pipeline Development: Investigating TNFα Signaling and Epigenomic Changes in Breast Cancer Cells

Investigator: Harini Mudradi

Abstract

This report details the development of a comprehensive ChIP-seq data analysis pipeline focused on investigating TNFα signaling and epigenomic changes in breast cancer cells. Using Illumina data from Bioproject PRJNA255509, we applied a stepwise approach involving data download, quality control, alignment, and peak calling to process ChIP-seq data. The pipeline starts with the acquisition of raw sequence data, followed by quality checks with FastQC, alignment to the hg19 reference genome using BWA, and conversion of SAM to BAM format using SAMtools. Peaks were called with MACS2, identifying regions enriched for transcription factor binding. Subsequent peak annotation was performed using ChIPseeker in R, enabling the identification of genomic regions associated with TNFα signaling. Visualization tools, including genomic distribution pie charts, were employed to aid in the interpretation of results. This pipeline not only facilitated the identification of key regulatory regions in breast cancer cells but also honed our ability to process and interpret complex epigenomic datasets.

FlowChart

Download Data (SRA Toolkit) -> Quality Control (FastQC) -> Alignment (BWA) -> Convert and Sort (SAMtools) -> Peak Calling (MACS2) -> Output Peaks

Protocol

1. Download SRR1635445.fastq(ER-Chip) and SRR1635437.fastq(Input) from NCBI and h19.fa from UCSC Browser
2. Index the Ref Genome(hg19) using bwa-index
3. Fastqc for quality control (base quality and adapter content passed the quality check) 4.
Align Er-Chip and Input fastq files to the Ref Genome using bwa mem
5. Convert Sam files into Bam and then index Bam files using samtools
6. call peaks, specifying one sample as the treatment (ER ChIP) and the other as the control (input) using MACS2 tool
7. The peaks will be listed in the output file (ER_chip_peaks.narrowPeak)

1	18318	18650	ER_ChIP_peak_1	130	.	7.922	16.2942	13.0614	219
1	136143	136403	ER_ChIP_peak_2	150	.	10.03	18.3754	15.0942	129
1	327101	327258	ER_ChIP_peak_3	61	.	6.03093	9.12972	6.12304	38
1	440849	441013	ER_ChIP_peak_4	90	.	6.79075	12.156	9.03588	115
1	541870	542035	ER_ChIP_peak_5	47	.	4.94543	7.69448	4.75051	85
1	781101	781290	ER_ChIP_peak_6	65	.	6.33474	9.62006	6.58945	77
1	1000079	1000223	ER_ChIP_peak_7	44	.	5.27895	7.3967	4.46476	52
1	1008031	1008242	ER_ChIP_peak_8	70	.	5.76146	10.0495	7.00799	107
1	1009035	1009680	ER_ChIP_peak_9	1607	.	35.4647	165.606	160.791	159
1	1014786	1015093	ER_ChIP_peak_10	368	.	17.9484	40.4967	36.8515	159
1	1015479	1015899	ER_ChIP_peak_11	233	.	13.1974	26.7766	23.3325	279
1	1315897	1316109	ER_ChIP_peak_12	77	.	6.86264	10.7848	7.7134	132
1	1368445	1368778	ER_ChIP_peak_13	198	.	10.0825	23.2367	19.8559	121

Fig 1: List of peaks (ER_ChIP_peaks.narrowPeak file)

Visualization

1. Using output(.narrowPeak file) from macs2_output Directory and ChIPseeker in R, Annotation of peaks is done and saved in ER_Chip_peaks_annotation.txt file

seqnames	start	end	width	strand	V4	V5	V6	V7	V8	V9	V10
annotation	geneChr	geneStart	geneEnd	geneLength	geneStrand	geneId	transcriptId				
distanceToTSS	ENSEMBL	SYMBOL	GENENAME								
chr1	18319	18650	332	*	ER_ChIP_peak_1	130	.	7.922	16.2942	13.0614	219
Promoter (<=1kb)	1	15796	18061	2266	2	653635	uc009vjd.2	-258	NA		
WASH7P	WASP family homolog 7, pseudogene										
chr1	136144	136403	260	*	ER_ChIP_peak_2	150	.	10.03	18.3754	15.0942	129
3' UTR	1	134773	140566	5794	2	729737	uc021oeg.2	4163	NA	LOC729737	
uncharacterized LOC729737											
chr1	327102	327258	157	*	ER_ChIP_peak_3	61	.	6.03093	9.12972	6.12304	38
Promoter (2-3kb)	1	324288	325896	1609	1	100133331	uc021oeh.1	2814			
NA	NA	NA									
chr1	440850	441013	164	*	ER_ChIP_peak_4	90	.	6.79075	12.156	9.03588	115
Distal Intergenic	1	367659	368597	939	1	729759	uc010nxu.2	73191	ENSG00000284733	OR4F29	
olfactory receptor family 4 subfamily F member 29											
chr1	541871	542035	165	*	ER_ChIP_peak_5	47	.	4.94543	7.69448	4.75051	85
Distal Intergenic	1	621096	622034	939	2	729759	uc010nxv.2	79999	ENSG00000284733	OR4F29	
olfactory receptor family 4 subfamily F member 29											

Fig 2: Annotated peaks

2. PieChart showing genomic distribution was done using plotAnnoPie(peakAnno) command on ChIPseeker.

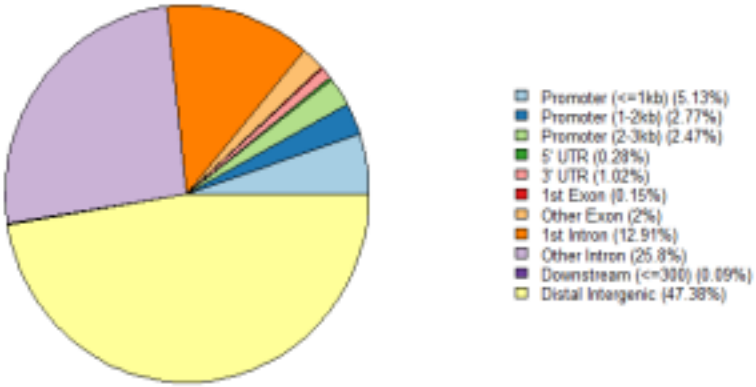

Fig 3: Pie Chart Visualization of genomic distribution

Discussion

This study was performed to investigate the role of TNFα signaling in breast cancer cells by identifying key genomic regions associated with epigenomic changes. TNFα is a critical cytokine involved in inflammation and cancer progression, making it essential to understand its regulatory impact on gene expression. Using ChIP-seq, I aimed to uncover transcription factor binding sites and other regulatory elements that drive TNFα-mediated responses in breast cancer cells.

The ChIP-seq data analysis pipeline developed in this study enabled the identification of peaks, which represent potential binding sites for transcription factors or histone modifications associated with TNFα signaling. By annotating these peaks and visualizing their genomic distribution, we gained insight into the specific genomic regions—such as promoters, enhancers, and intronic regions—most affected by TNFα signaling. This understanding provides a foundation for further exploration into how these regions influence gene expression patterns in breast cancer.

With this information, future studies could focus on validating these regulatory regions through functional assays, such as CRISPR-based gene editing, to determine their role in TNFα-driven cancer progression. Additionally, therapeutic targets could be identified by pinpointing genes regulated by TNFα in critical pathways, opening doors for targeted treatments that could modulate TNFα signaling in breast cancer.

References

1. Aggarwal, B. B. (2003). Signalling pathways of the TNF superfamily: a double-edged sword. *Nature Reviews Immunology*, 3(9), 745-756. DOI: 10.1038/nri1185.
2. Lee, J., & Kim, Y. (2017). TNF-α in cancer: a potential therapeutic target. *Clinical Cancer Research*, 23(12), 3220-3226. DOI: 10.1158/1078-0432.CCR-16-2160.
3. Yu, G., et al. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 31(14), 2382-2383. DOI: 10.1093/bioinformatics/btv095.
4. Li, S., et al. (2014). The role of epigenetics in breast cancer: a review. *Journal of Cellular Biochemistry*, 115(7), 1117-1125. DOI: 10.1002/jcb.24700.
5. Zhang, Y., et al. (2015). CRISPR/Cas9 for gene editing: a new era of genome editing. *Nature Reviews Molecular Cell Biology*, 16(8), 457-460. DOI: 10.1038/nrm.2015.12.