

## Analysis of ER treatment in MCF7 Cell Line

### Background

The estrogen receptor- $\alpha$  (ER $\alpha$ ) is a transcription factor which plays a critical role in controlling cell proliferation and tumorigenesis. By recruiting various cofactors to estrogen response elements (EREs), ER $\alpha$  effectively induces or represses gene transcription.

The provided data consists of the mRNA expression profiles of MCF7 cells treated with or without estrogen (E2) treatment under negative controls siRNA, BRD4 siRNA or JQ1 treatment, in duplicates.

This analysis will be focusing on the two duplicates of the E2 treatment under negative control siRNA and trying to discover the E2 treatment's potential uses.

### Method

Two replicates of Control siRNA Veh and two replicates of Control siRNA E2 were used in the analysis. And they can be found at the following link (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55922>).

FastQC, a read quality report tool, version 0.72 from Galaxy was used for quality control. In total, two replicates of Control siRNA Veh and two replicates of Control siRNA E2 were fed into the FastQC separately. And default settings were used for all six replicates.

I used Trim Galore! (version: 0.6.3) from Galaxy to trim the replicates separately. Default settings were used. Using the default settings, bases that had a Phred quality score less than 20 were trimmed.

Read mapping was performed using HISAT2(version: 2.1.0, samtools version: 1.9) from Galaxy. Replicates were processed as single end reads and were aligned against hg38 as the reference genome. All the parameters were set as default.

I assembled and counted the transcript using featureCounts (version: 1.6.4), a tool that measures gene expression in RNA-Seq experiments from SAM or BAM files. All four mapped replicate BAM files were put into featureCounts. Build-in index was used to fit the reference. All the parameters were set as default. Four tabular files, which had the counts of all the features for four individual replicates were created in this step. The tabular files were used in the following analyses.

Plot graphs were created, and the gene differential analysis were performed in RStudio (Version 1.2.5033) using a customized R script (attached in the email). Pheatmap was used to make heatmap, RColorBrewer was in charge of color palettes for heatmaps, BiocManager was used for managing Bioconductor packages, org.Hs.eg.db was used for converting gene names and ENSEMBL ids in mapping, DESeq2 was utilized for visualization and differential expression analysis. The tabular files obtained from the previous step was imported into RStudio for plot making and differential analysis. The significant gene list was selected base on the threshold that having FDR-adjusted P value <0.05 and fold-change >1.5 for up-regulated genes or fold-change <-1.5 for down-regulated genes (log2FC > 0.58 for up-regulated genes and log2FC < -0.58 for down-regulated genes).

Enrichment analysis was performed using Enrichr, a web-based enrichment analysis tool ([amp.pharm.mssm.edu/Enrichr/](http://amp.pharm.mssm.edu/Enrichr/)). The significant genes were used in the enrichment analysis. GO molecular function and GO Biological Processes were both under the category of

Ontologies, and KEGG Human was under pathway category. The bar chart was ordered by p-value in descending order.

## Summary

### Differential Expression Analysis

The differential analysis showed that Control siRNA Veh and Control siRNA E2 were expressed significantly differently from each other. The replicates of the same treatment were expressed much more closely to each other compared to the replicates of the other treatment (shown in Figure 1 and Figure 2). There were 202 up-regulated and 522 down-regulated significant genes selected base on the customized threshold. In the top 50 differentially expressed genes (based on the p-value), there were more highly expressed genes in the E2 treatment samples than in the Veh control group (shown in Figure 3).

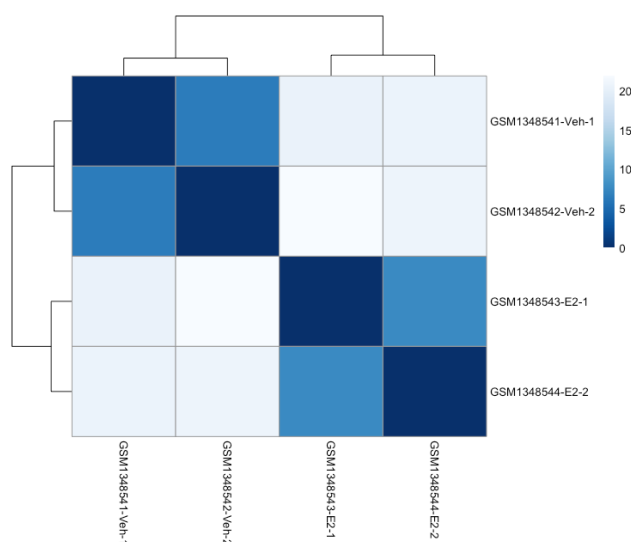


Figure 1 Sample Heat Map

The distance is closer within the same treatment, and the distance is farther within the different treatment.

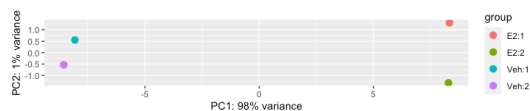


Figure 2 PCA Plot of Treatments

First principal components separated Veh group and E2 group well, meaning the genes expressed in two treatments are highly different.

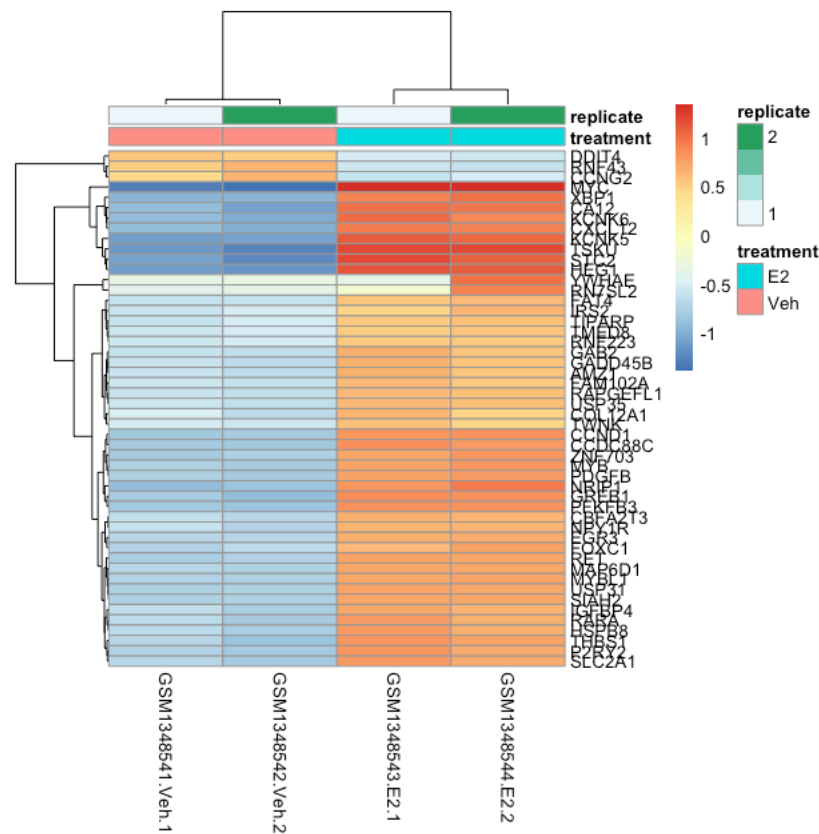


Figure 3 Gene Expression Heat Map

Top 50 most differentially expressed genes (based on p-value) in two treatments are presented.

## Gene Enrichment Analysis

A gene enrichment analysis was performed in order to find out the biological processes, molecular functions, and pathways the treatment affected. The 724 significant genes, mentioned previously, were used in this step.

## Pathway Analysis

Pathways in cancer and colorectal cancer were the most affected pathways according to KEGG Pathways (shown in Figure 4). The genes that were closely related to the pathways are shown in Figure 5. Genes that are highlighted in Figure 6, including MAX, CCND1 and MYC and other genes, such as AMA3, CRKL and JAK2, were down-regulated and led to less proliferation according to the pathway in cancer. In the same pathway, up-regulated TGFB2 and TGFB3 increased the sustained angiogenesis and along with up-regulated TGFBR2 contributed to reduced insensitivity to anti-growth signals. Many down-regulated significant genes including MYC, CCND1, FOS, TGFA, GADD45G and GADD45B led to less proliferation and uncontrolled proliferation/genomic instability according to the colorectal cancer pathway (shown in Figure 8).

KEGG 2019 Human

Bar Graph Table Clustergram  

Click the bars to sort. Now sorted by p-value ranking.

SVG PNG JPG

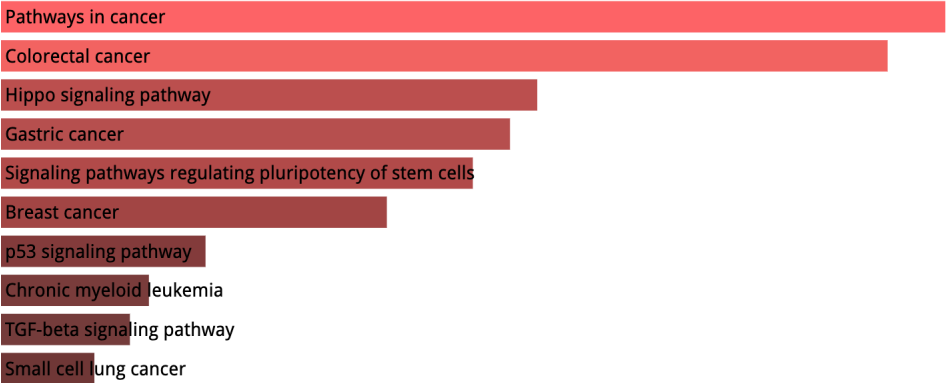


Figure 4 KEGG Pathways Results

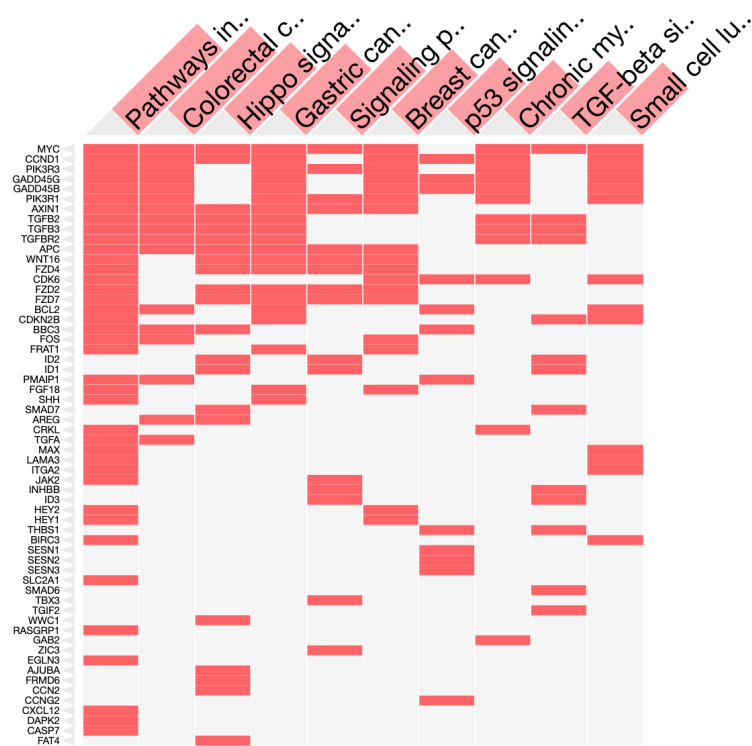


Figure 5 Genes Related to the Pathways



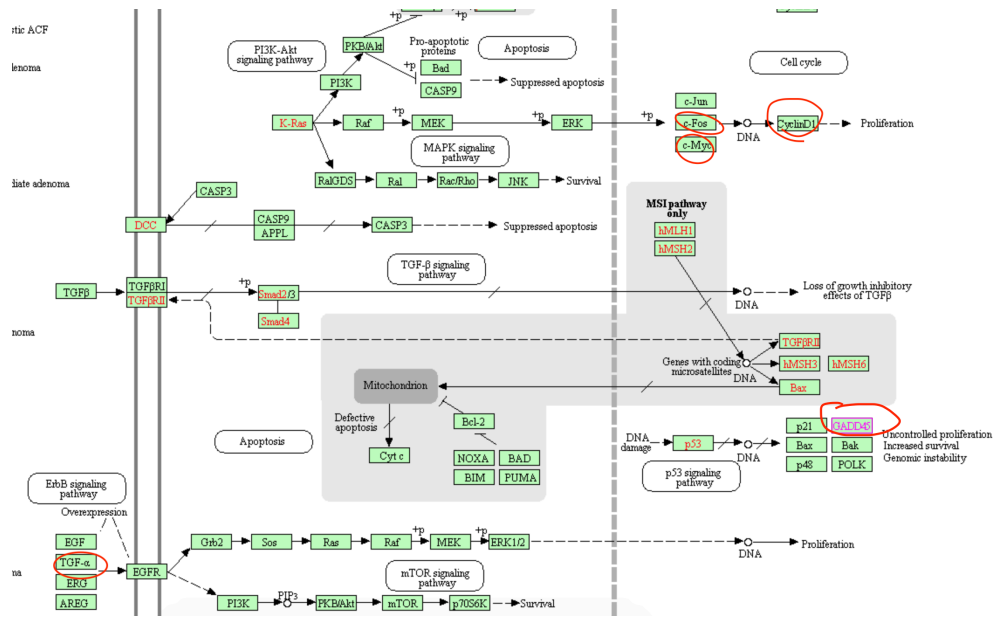


Figure 8 Colorectal Cancer Pathway

## Biological Process

Negative regulation of transcription, DNA-templated (GO:0045892), negative regulation of transcription from RNA polymerase II promoter (GO:0000122) and regulation of cell proliferation (GO:0042127) were the four most significant GO biological processes the genes pointed to (shown in Figure 9). Down-regulated genes including MYC, SHH and PDGFB were present in these three biological processes (shown in Figure 10).

### GO Biological Process 2018

Bar Graph

Table

Clustergram

Settings Information

Click the bars to sort. Now sorted by p-value ranking.

SVG PNG JPG

negative regulation of transcription, DNA-templated (GO:0045892)	
negative regulation of transcription from RNA polymerase II promoter (GO:0000122)	
regulation of cell proliferation (GO:0042127)	
positive regulation of transcription, DNA-templated (GO:0045893)	
ribosome biogenesis (GO:0042254)	
regulation of transcription, DNA-templated (GO:0006355)	
regulation of transcription from RNA polymerase II promoter (GO:0006357)	
positive regulation of gene expression (GO:0010628)	
atrioventricular valve morphogenesis (GO:0003181)	
positive regulation of transcription from RNA polymerase II promoter (GO:0045944)	

Figure 9 Top GO Biological Processes (based on p-value)

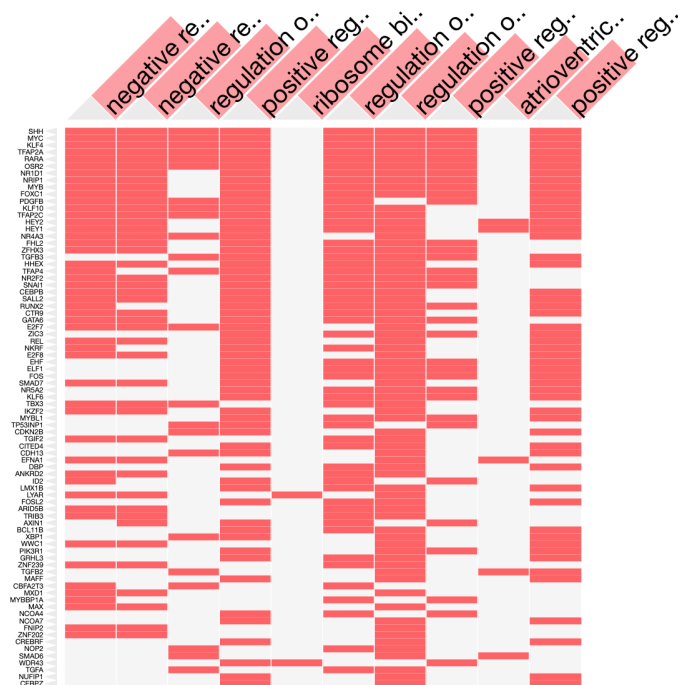


Figure 10 Genes Related to the Biological Processes

## Molecular Function

Transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding (GO:0000982) and RNA polymerase II regulatory region sequence-specific DNA binding (GO:0000977) were the most significant molecular functions found by the analysis (shown in Figure 11). Down-regulated genes including MYC, MAX and KLF10 were present in the molecular functions (shown in Figure 12).

## GO Molecular Function 2018

Bar Graph

Table

Clustergram

Settings

SVG PNG JPG

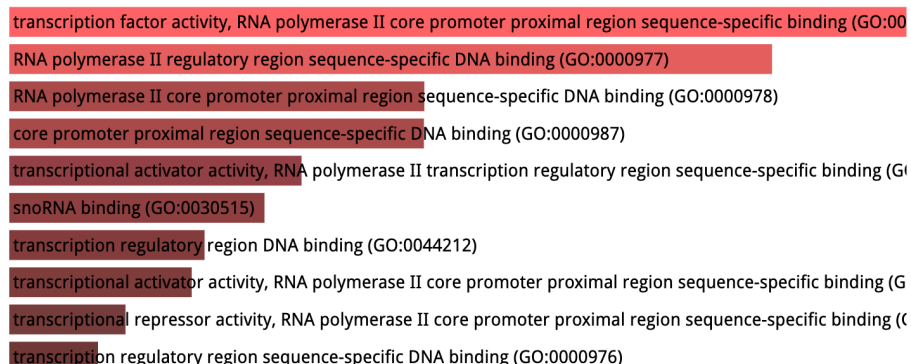


Figure 11 Top GO Molecular Functions (based on p-value)

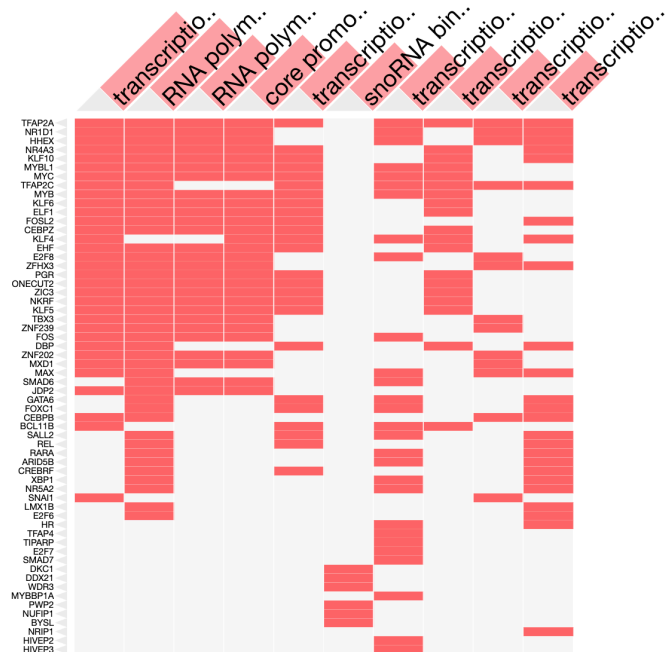


Figure 12 Genes Related to the Molecular Functions

## Conclusion

To conclude, the E2 treatment had a noteworthy impact on many genes in MCF7 cell line. Many significantly affected genes were in charge of cell proliferation according to the pathway analysis, the biological process analysis, and the molecular function analysis. The down-regulated genes showed that the E2 treatment could repress cell proliferation and related functions. Therefore, E2 treatment could have potential uses in cancer treatment.