

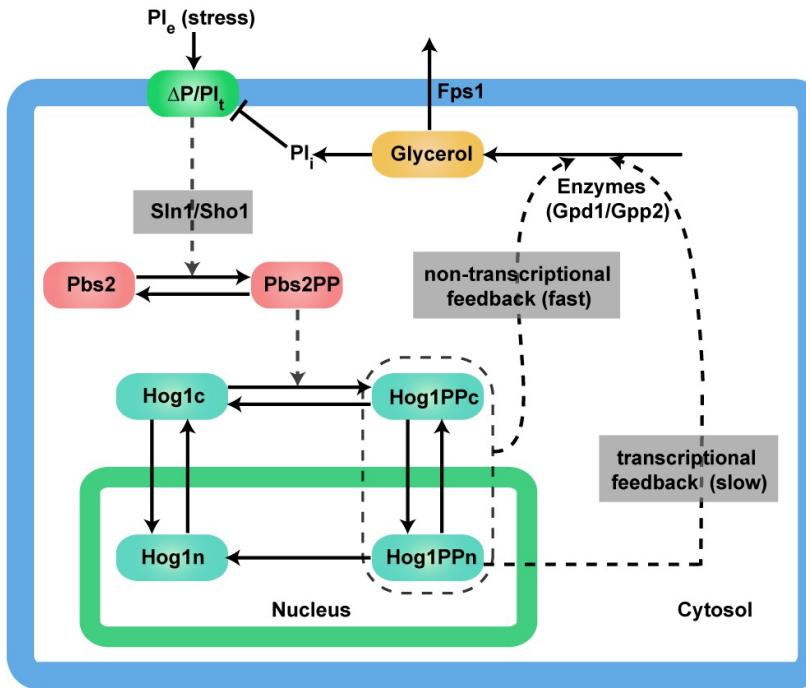


# BMS5022 Topic 2

Jaskirat Kaur Sandhu

27667421

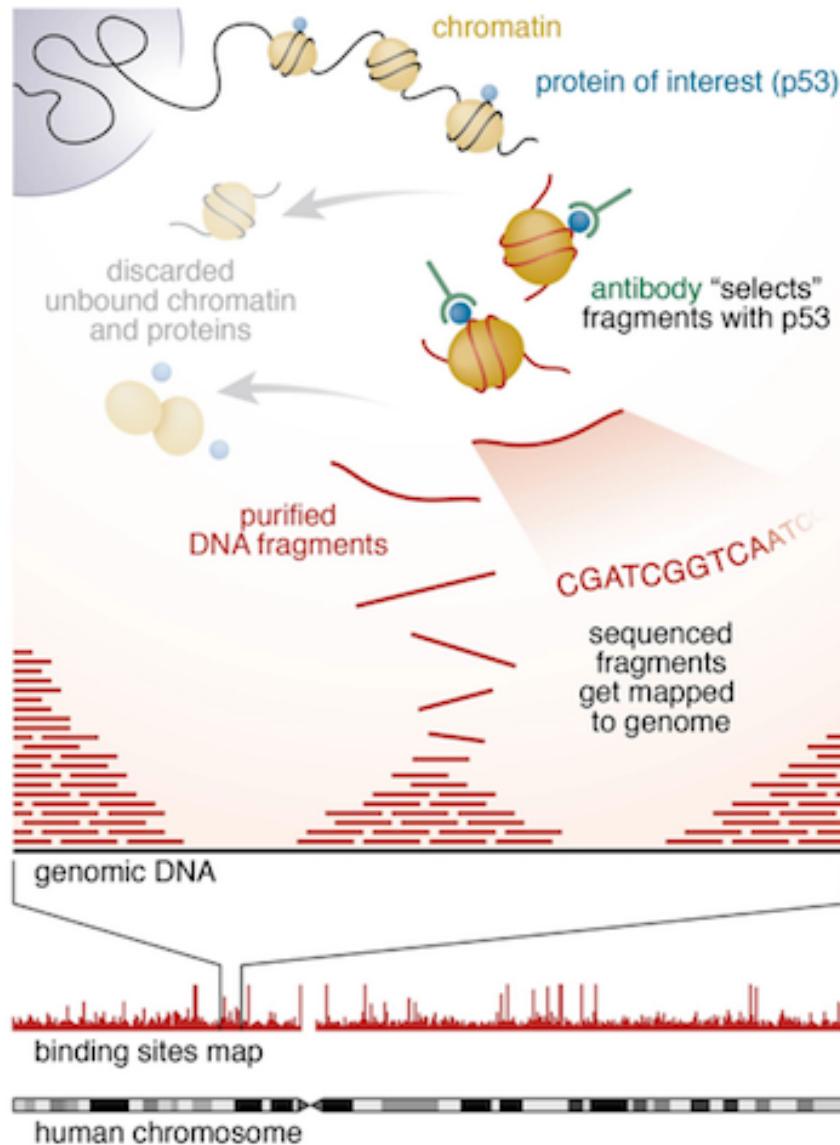
## Aim of topic 2 workshop



Overall Aim: Linking Hog1 Chromatin binding with gene expression in budding yeast

Analyse RNA-seq data that traces the transcriptional response to high osmotic stress to determine if the enriched peaks identified by Hog1 binding correlate with genes upregulated over time (0,5,15,30 mins) after high salt treatment.

# Chromatin immunoprecipitation sequencing (ChIP-seq)



Crosslink protein of interest and DNA

Chromatin is fragmented into pieces of about 150 to 500 bp.

Chromatin immunoprecipitation is carried out using a specific antibody against the protein of interest

Cross-links are reversed and the immunoprecipitated (IP) DNA is purified

The IP DNA fragments are then sequenced using NGS

The enriched regions of DNA or peaks are then identified



# METHODOLOGY

## Dataset used for analysis

### ChIPSeq dataset

*Research Paper: Hog1 Controls Global Reallocation of RNA Pol II upon Osmotic Shock in *Saccharomyces cerevisiae* (Cook & O'Shea, 2012)*

YEPD (control group)

KCl (treatment group in which potassium chloride was added to cell culture to induce osmotic stress)

- SRR500622: Hog1IP\_wildtype\_YEPD
- SRR500623: Hog1IP\_wildtype\_KCl
- SRR500630: mockIP\_wildtype\_YEPD
- SRR500631: mockIP\_wildtype\_KCl
- SRR500626: Sko1IP\_wildtype\_YEPD
- SRR500627: Sko1IP\_wildtype\_KCl
- SRR500628: Hot1IP\_wildtype\_YEPD
- SRR500629: Hot1IP\_wildtype\_KCl

### RNASeq dataset

*Beilharz lab*

Dataset containing the transcriptional expression profile change of the W303 budding yeast (*Saccharomyces cerevisiae*) strain to high osmotic stress at different timepoints:

- 0 minutes
- 5 minutes
- 15 minutes
- 30 minutes

## Quality Control: FASTQC

FastQC provides a simple way to do some quality control checks on raw sequence datasets from high throughput sequencing (FastQ files)



## Alignment to Genome: Bowtie2

- Bowtie2 is used to align sequencing reads to long reference sequences
- Bowtie2 outputs alignments in an unsorted SAM format



## Filter BAM: Samtools

- The alignment files need to be filtered so that it only contains uniquely mapped reads in order to increase confidence in site discovery and improve reproducibility
- This involves:
  1. Change alignment file format from SAM to BAM
  2. Sort BAM file by genomic coordinates
  3. Filter to keep only uniquely mapped reads and remove unmapped reads

## Peak Calling: MACS2

- Peak calling is a method used to identify areas in the genome that have been enriched with aligned reads from performing ChIP-seq
- MACS2 is used to identify transcription factor binding sites
- There are 6 output files generated, which can be used for further analysis



## Visualization & Peak Annotation: IGV, ChipSeeker (R)

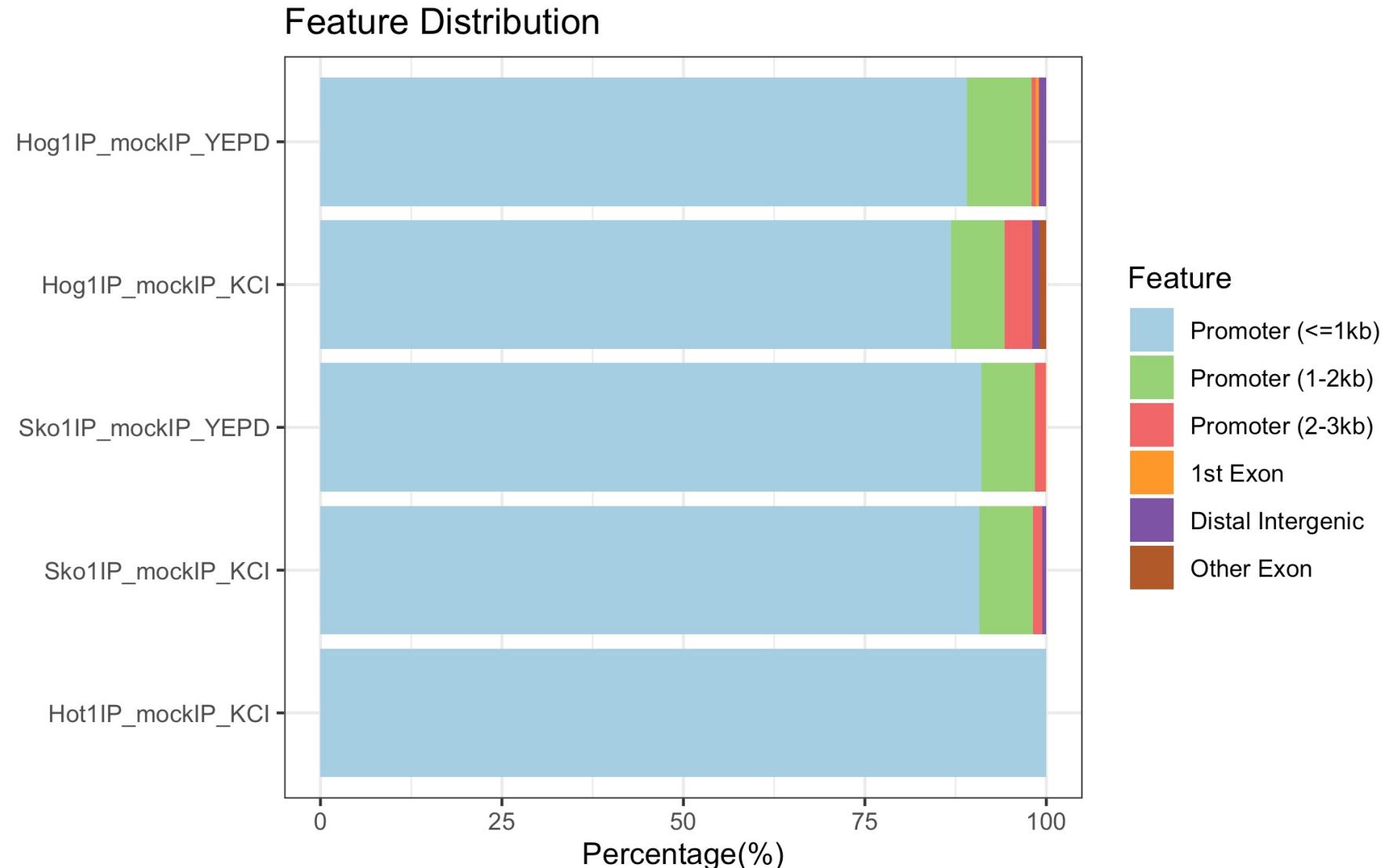
### Qualitative assessment using IGV (Integrative Genomics Viewer)

- This is used to assess the quality of your alignment
- IGV is a high-performance, easy-to-use, interactive tool for the visual exploration of genomic data

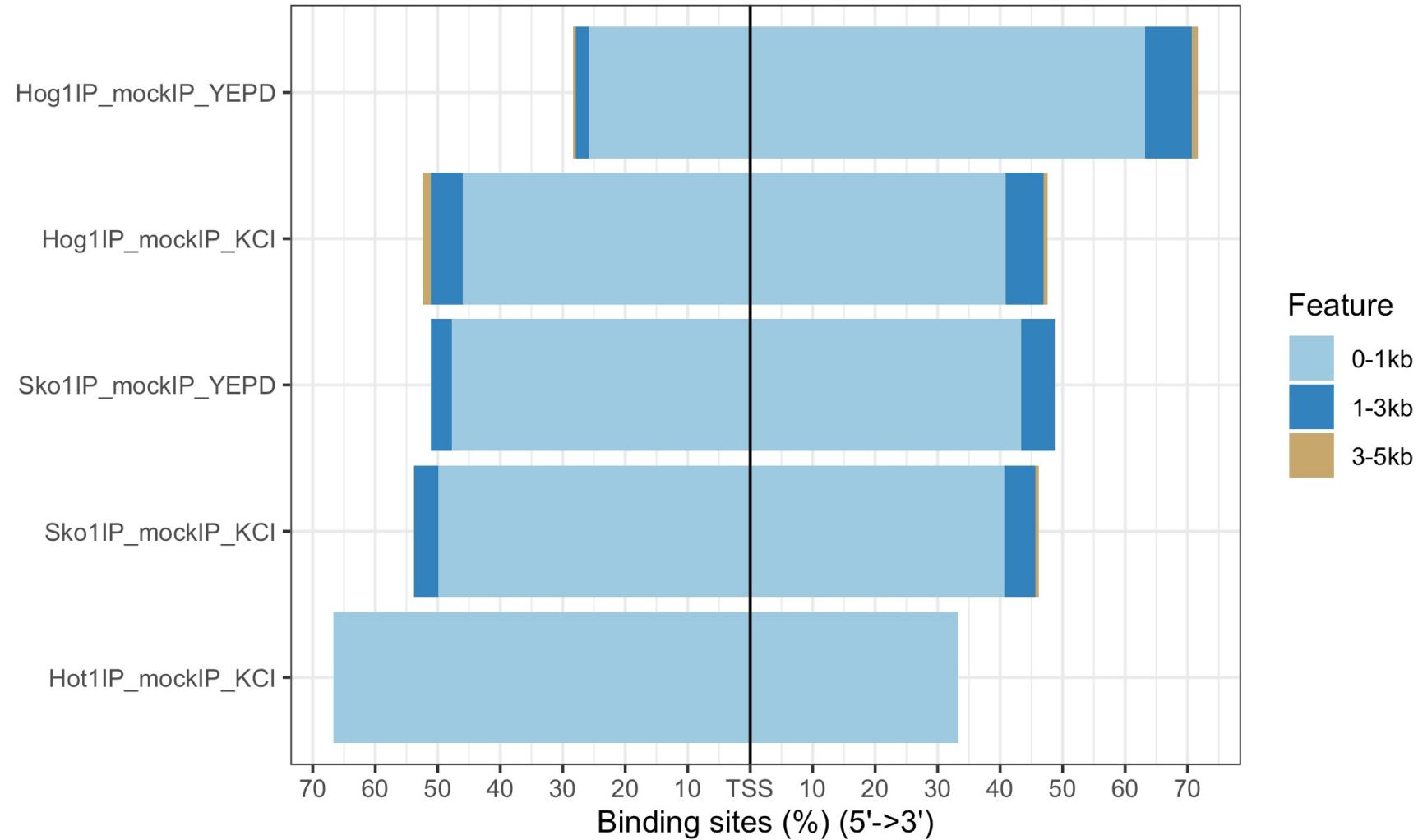
### Peak Annotation (ChIPseeker)

- ChIPseeker is a R Bioconductor package for annotating peaks
- It also has various visualization functions to assess peak coverage over chromosomes and profiles of peaks binding to TSS regions

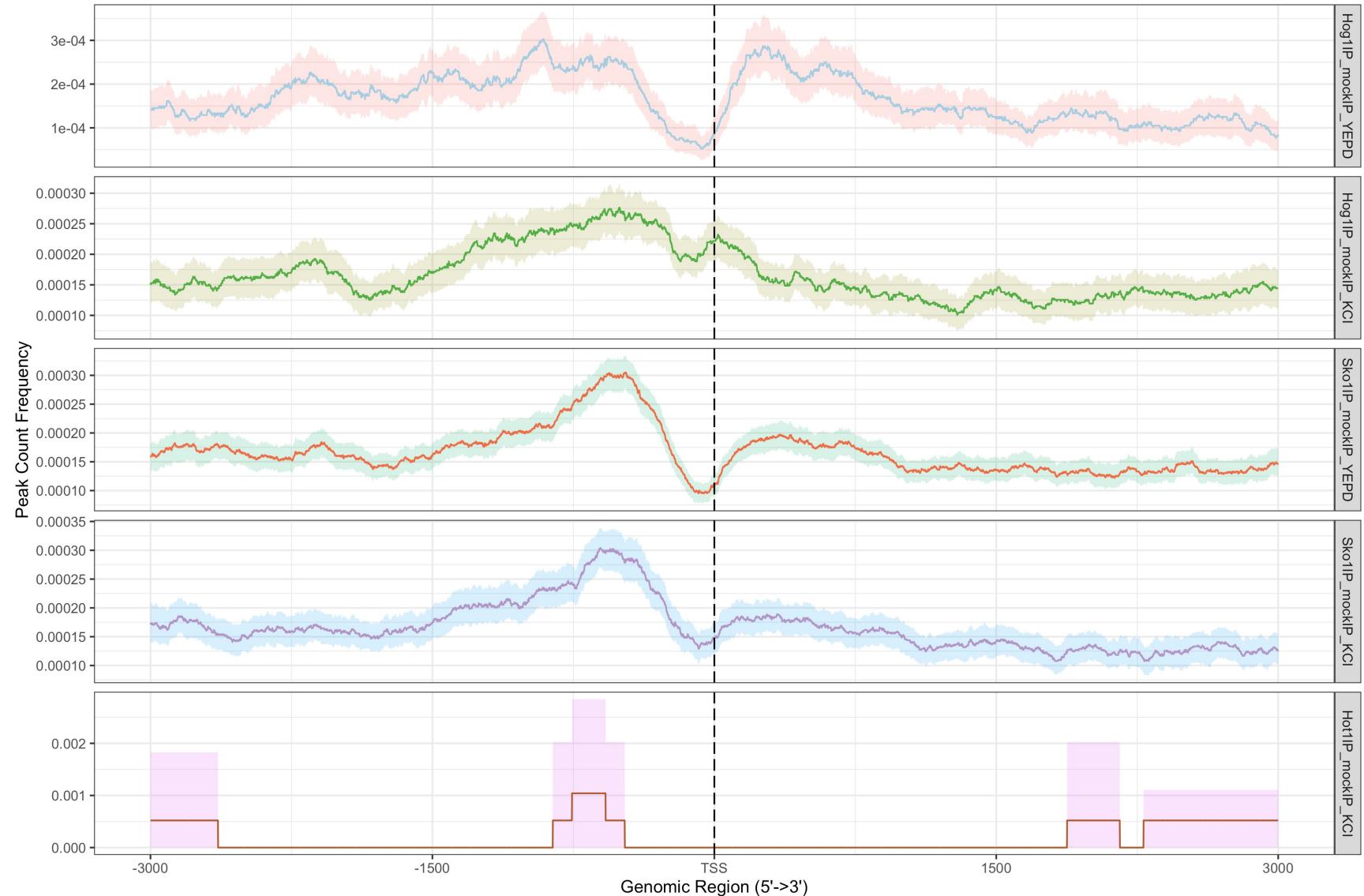
# ANALYSIS & RESULTS



## Distribution of transcription factor-binding loci relative to TSS

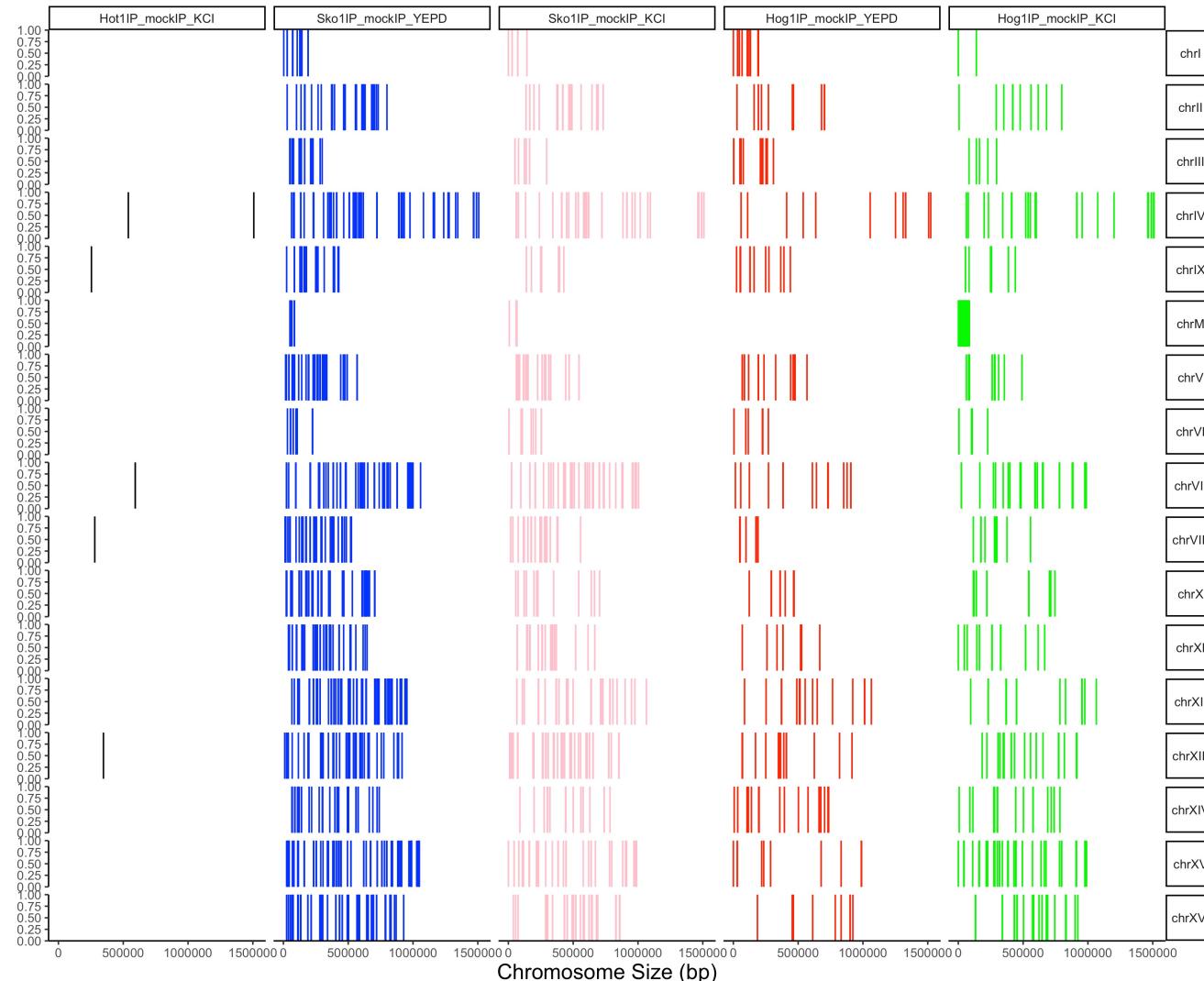


# Analysis of Chipseq data



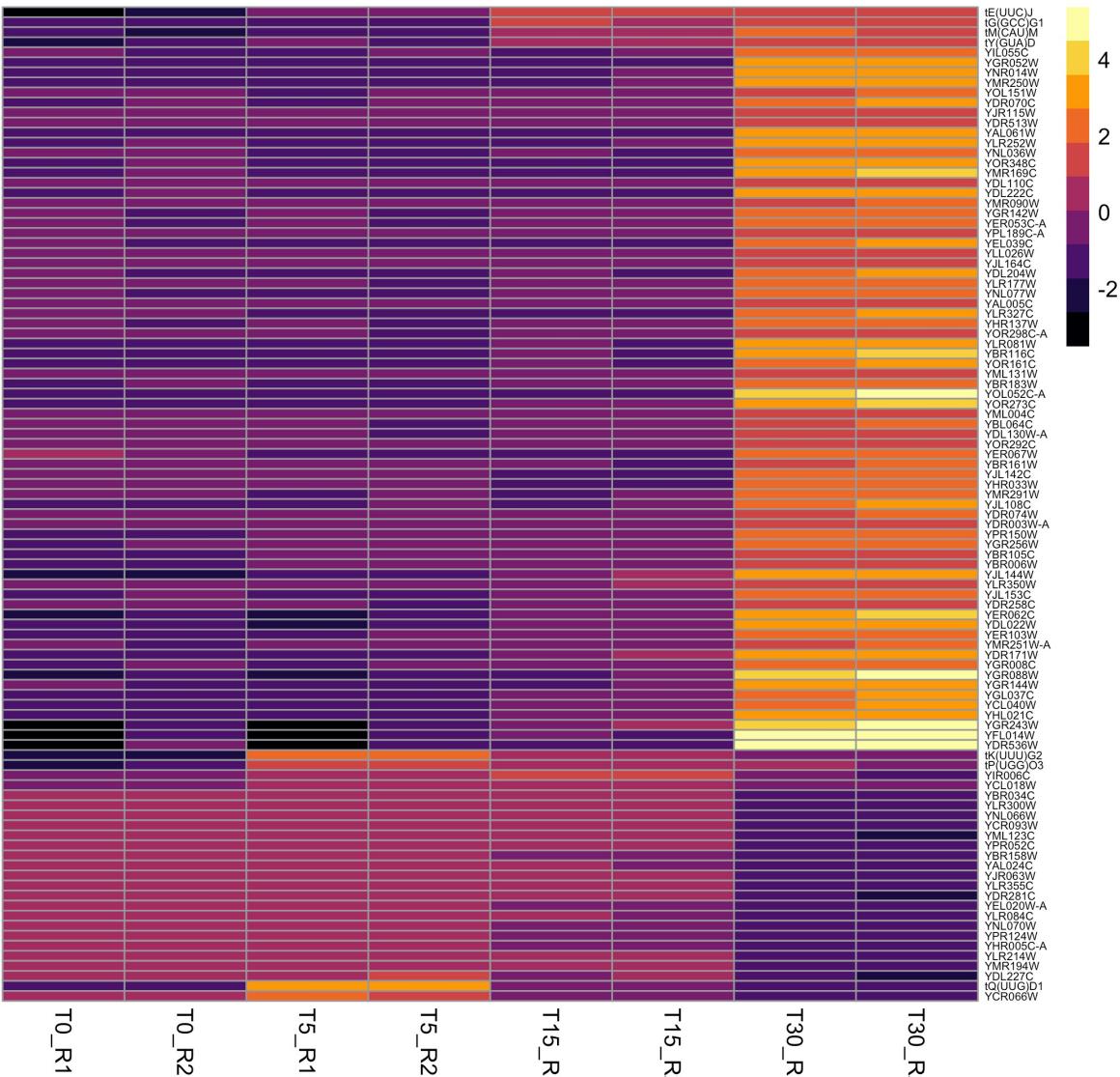
# Analysis of Chipseq data

ChIP Peaks over Chromosomes

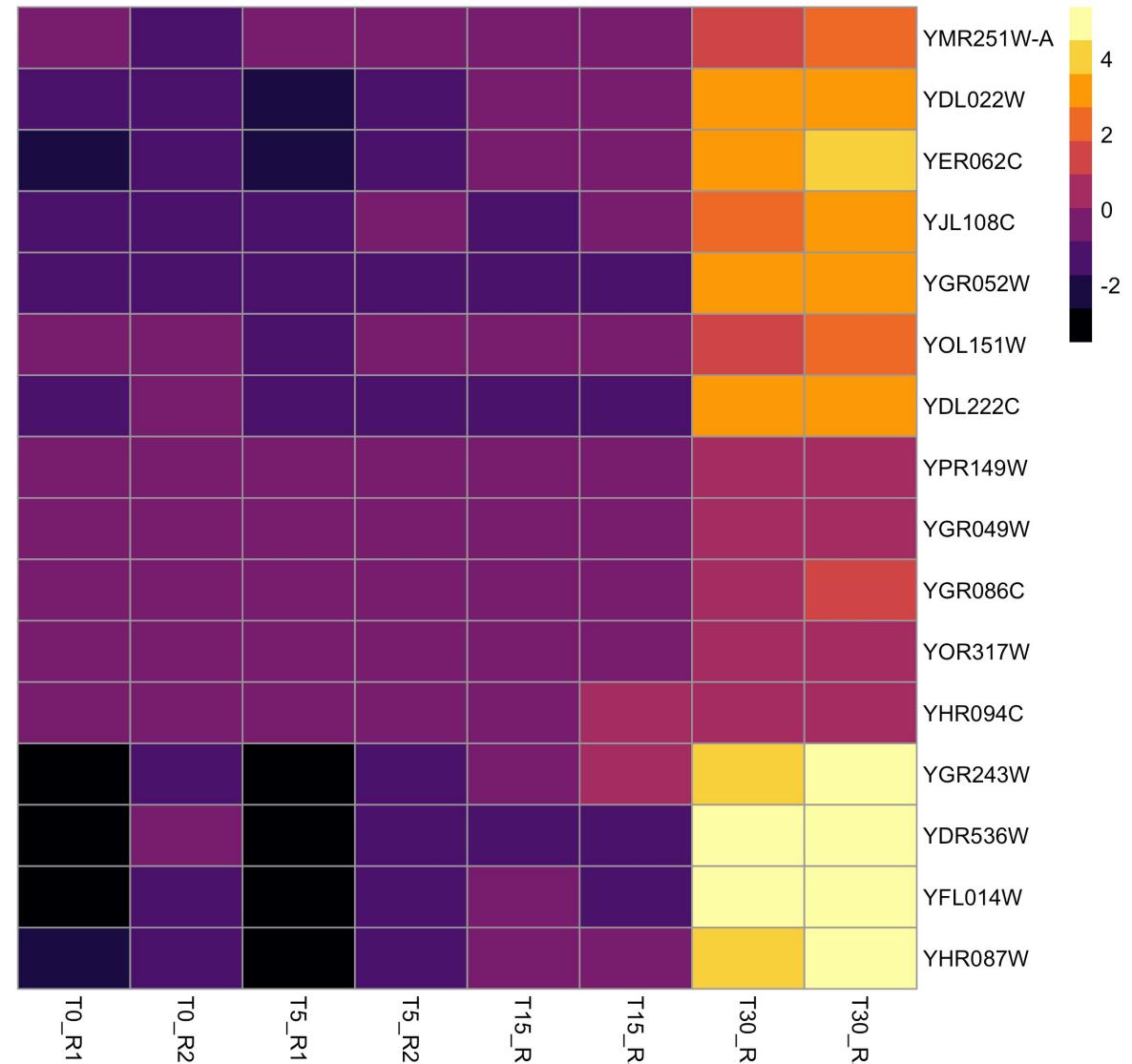


# Expression profile changes

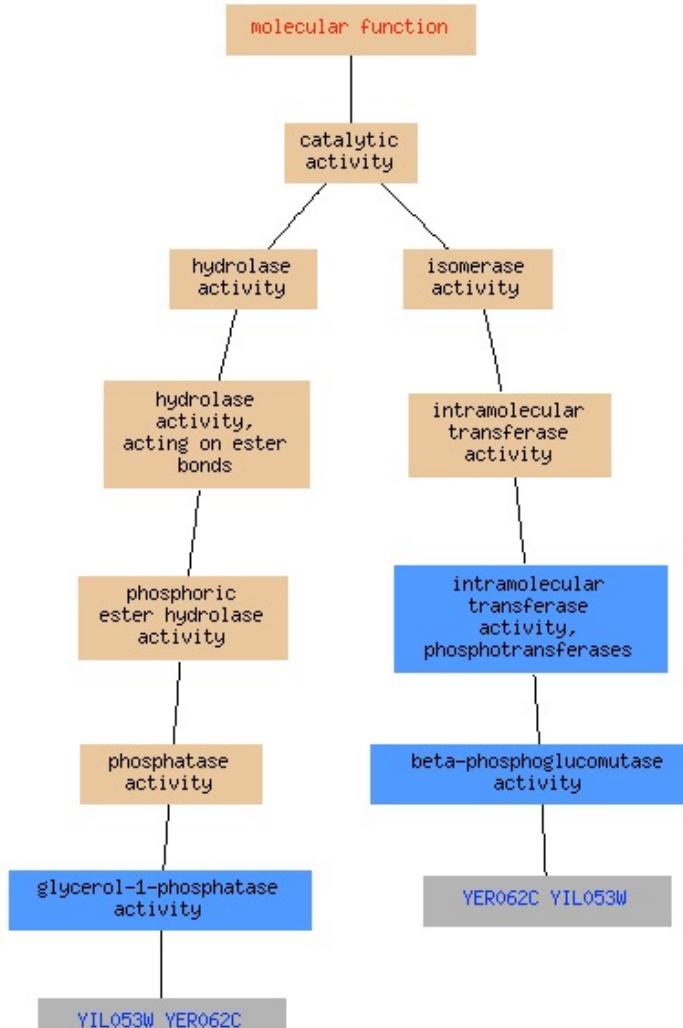
**Heatmap of top 100 significant genes**



**Heatmap of genes from the gene list**

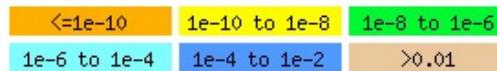


# Functional Analysis: Molecular function

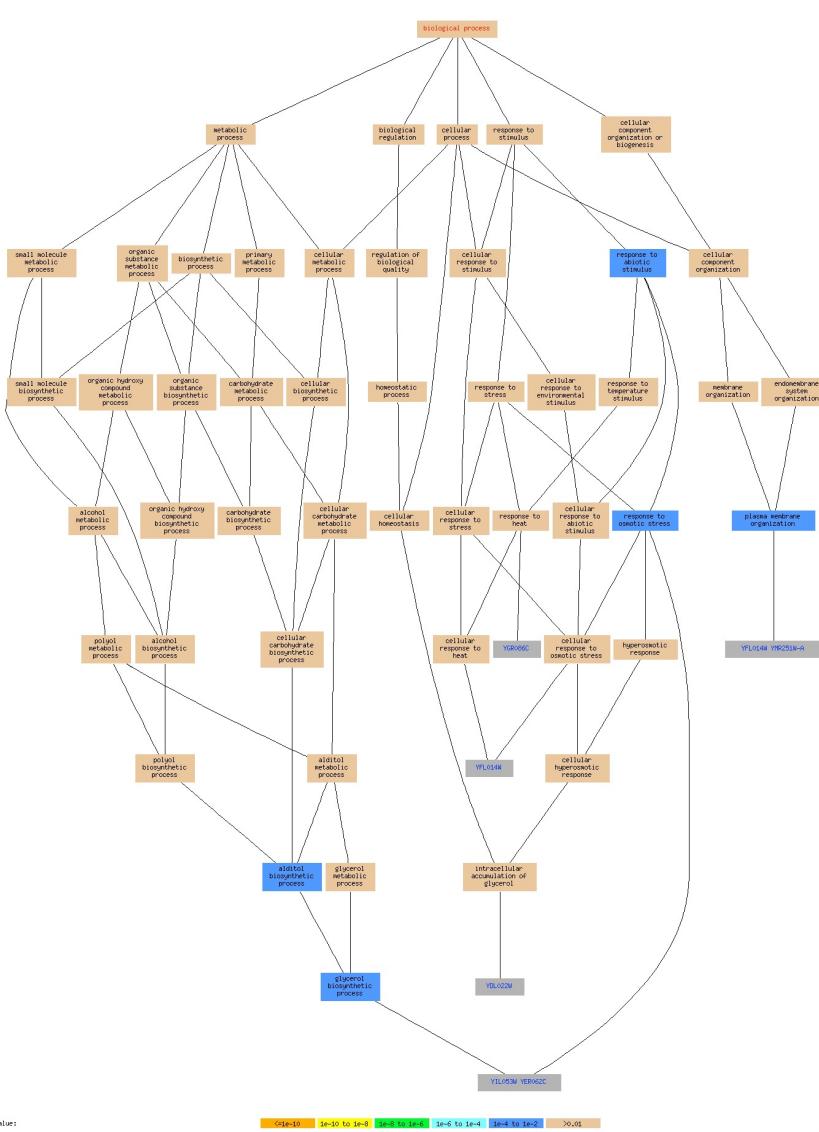


Gene Ontology term	Cluster frequency	Genome frequency	Corrected P-value	FDR	False Positives	Genes annotated to the term
glycerol-1-phosphatase activity	2 of 11 genes, 18.2%	3 of 7166 genes, 0.0%	0.00017	0.00%	0.00	YIL053W, YER062C
beta-phosphoglucomutase activity	2 of 11 genes, 18.2%	4 of 7166 genes, 0.1%	0.00035	1.00%	0.02	YER062C, YIL053W
intramolecular transferase activity, phosphotransferases	2 of 11 genes, 18.2%	16 of 7166 genes, 0.2%	0.00711	2.67%	0.08	YIL053W, YER062C

pvalue:



# Functional Analysis: Biological Process



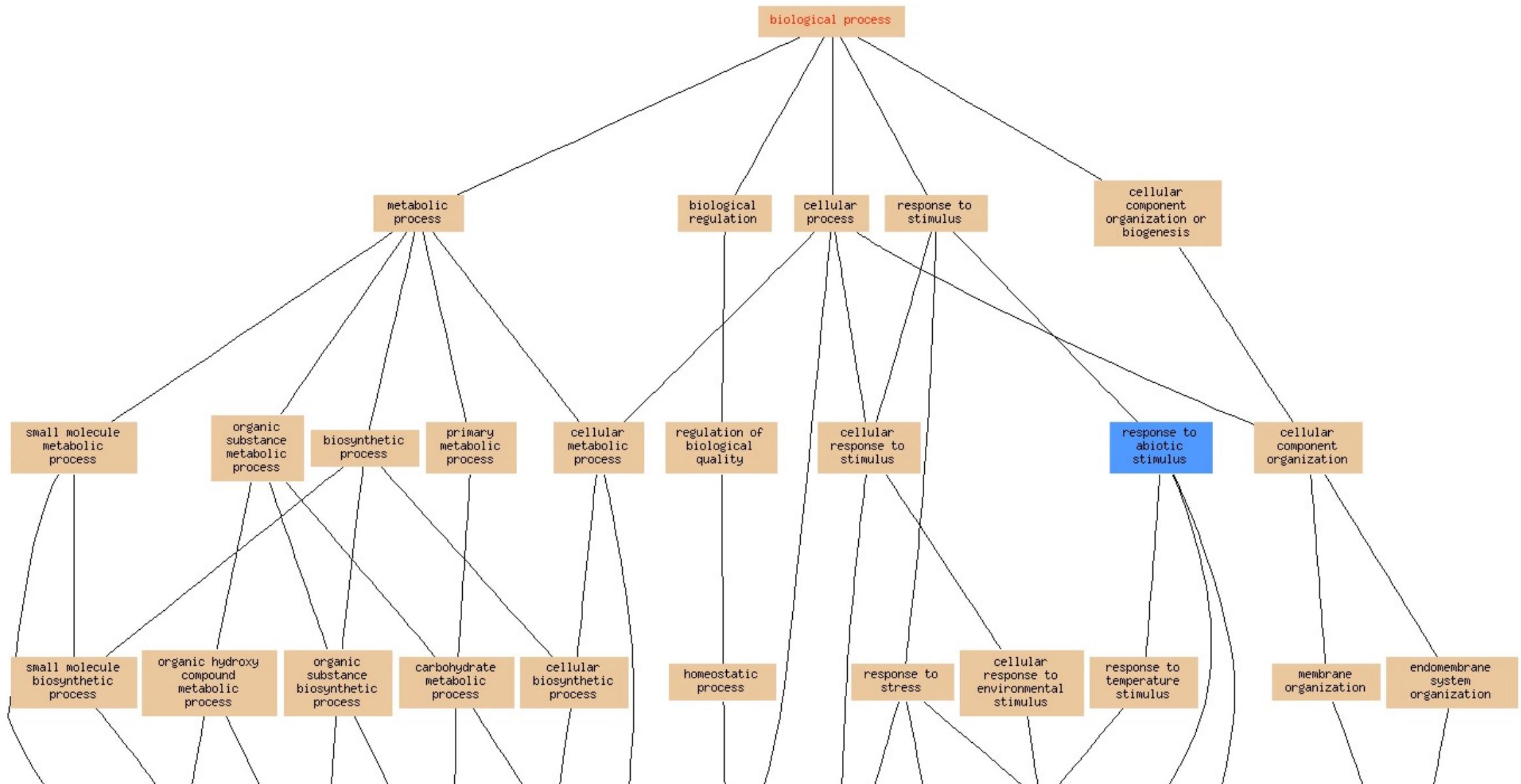
Gene Ontology term	Cluster frequency	Genome frequency	Corrected P-value	FDR	False Positives	Genes annotated to the term
response to abiotic stimulus	5 of 11 genes, 45.5%	171 of 7166 genes, 2.4%	0.00022	0.00%	0.00	YDL022W, YIL053W, YFL014W, YGR086C, YER062C
response to osmotic stress	4 of 11 genes, 36.4%	73 of 7166 genes, 1.0%	0.00022	0.00%	0.00	YDL022W, YFL014W, YER062C, YIL053W
glycerol biosynthetic process	2 of 11 genes, 18.2%	3 of 7166 genes, 0.0%	0.00047	0.00%	0.00	YIL053W, YER062C
alditol biosynthetic process	2 of 11 genes, 18.2%	3 of 7166 genes, 0.0%	0.00047	0.00%	0.00	YER062C, YIL053W
plasma membrane organization	2 of 11 genes, 18.2%	10 of 7166 genes, 0.1%	0.00708	2.80%	0.14	YFL014W, YMR251W-A
glycerol metabolic process	2 of 11 genes, 18.2%	14 of 7166 genes, 0.2%	0.01428	6.67%	0.40	YER062C, YIL053W
alditol metabolic process	2 of 11 genes, 18.2%	19 of 7166 genes, 0.3%	0.02672	8.00%	0.56	YIL053W, YER062C
polyol biosynthetic process	2 of 11 genes, 18.2%	19 of 7166 genes, 0.3%	0.02672	7.00%	0.56	YER062C, YIL053W
alcohol metabolic process	3 of 11 genes, 27.3%	106 of 7166 genes, 1.5%	0.03523	6.89%	0.62	YIL053W, YER062C, YOL151W

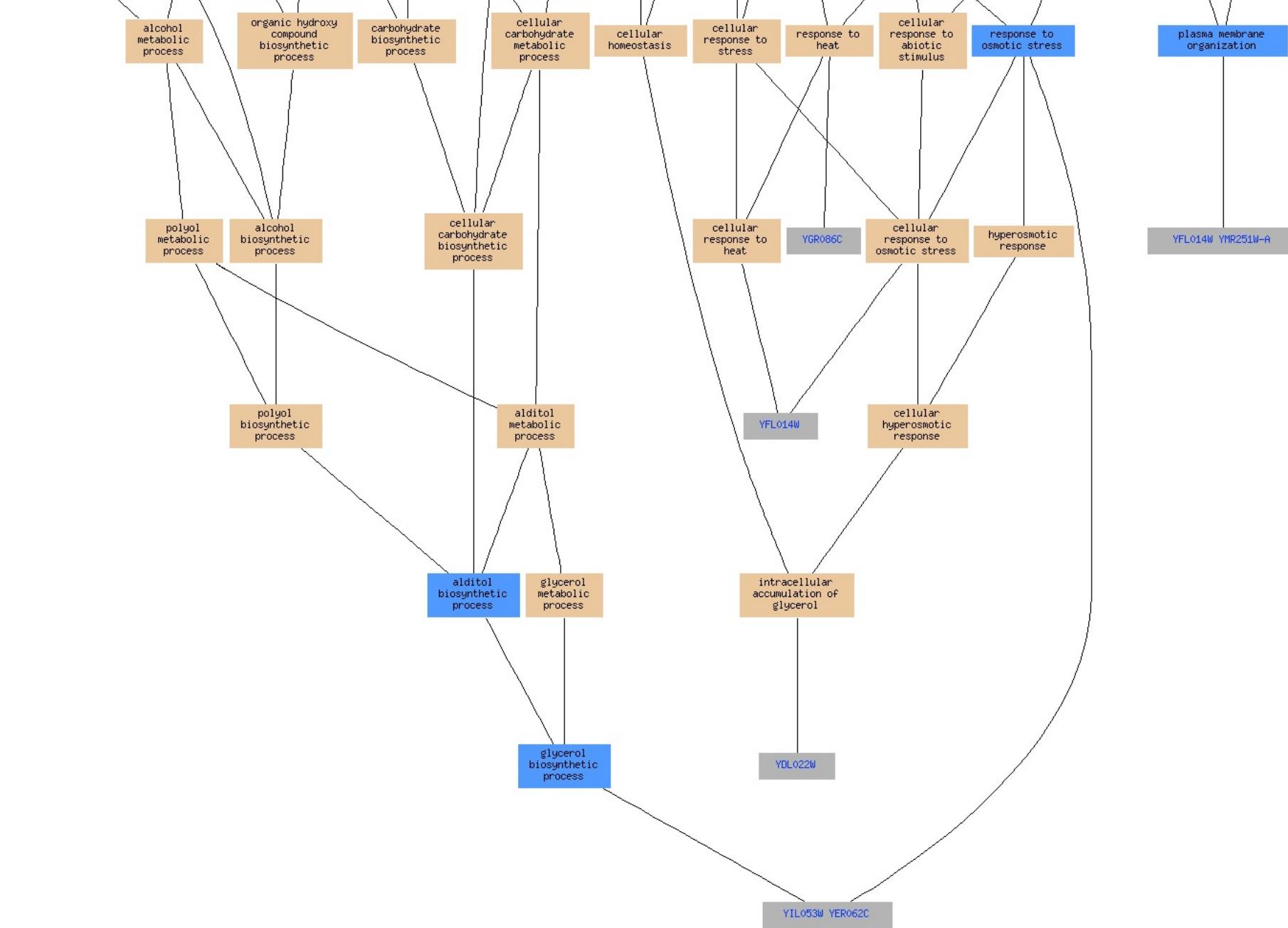
Introduction

Methodology

Analysis & Results

Conclusion

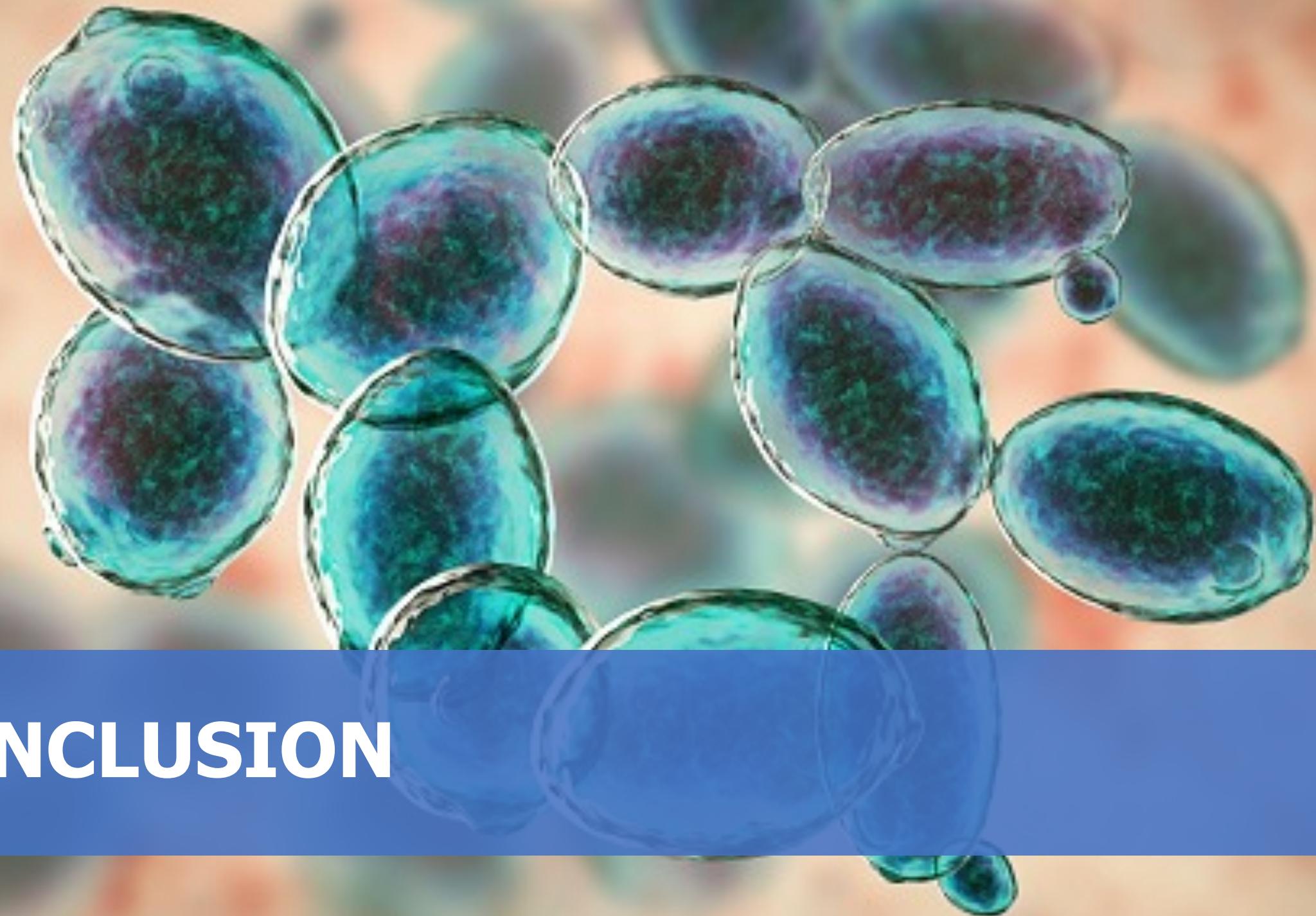




value:

<=1e-10 1e-10 to 1e-8 1e-8 to 1e-6 1e-6 to 1e-4 1e-4 to 1e-2 >0.01

# CONCLUSION



## Give an overall conclusion(s) that summarises the outcomes of your analysis

### Hog1 Genes:

- YDL022W
- YDR536W
- YER062C
- YFL014W
- YGR052W
- YGR243W
- YHR087W
- YIL053W
- YMR251W-A
- YOL151W
- YGR086C

### Sko1 Genes:

- YDL022W
- YGR243W
- YMR251W-A
- YOL151W
- YGR086C

- Hog1 induces a differential expression change in the 11 genes from the gene-list provided, as can be seen from the heatmap
- These genes had enriched peaks in the high salt treatment dataset
- Hog1 is involved in osmotic stress as can be seen from the (biological process and molecular function) GO analysis conducted on the differentially expressed genes from the annotation of the enriched peaks
- Sko1 is involved in the process of osmotic stress however not as much as Hog1 as only 5 genes were identified in the gene-list and ChipSEQ analysis
- These genes had enriched peaks in the control group

# REFERENCES

<https://yeastgenome.org/goTermFinder>

Hog1 Controls Global Reallocation of RNA Pol II upon Osmotic Shock in *Saccharomyces cerevisiae*  
Kristen E. Cook and Erin K. O'Shea

Ssk1p-Independent Activation of Ssk2p Plays an Important Role in the Osmotic Stress Response in  
*Saccharomyces cerevisiae*: Alternative Activation of Ssk2p in Osmotic Stress  
Hui Zhi, Leihan Tang, Yiji Xia, Jianhua Zhang