Decoding the molecular basis of Iterative Stress Response (ISR)

by

Anmol Aggarwal
under the supervision of
Dr. Gaurav Ahuja

Submitted in partial fulfilment of the requirements for the degree of Master of Technology, Computational Biology

Center for Computational Biology,
Indraprastha Institute of Information Technology - Delhi

Certificate

This is to certify that the thesis titled "**Decoding the molecular basis of Iterative Stress Response (ISR)**" being submitted by Ms. **Anmol Aggarwal** to the Indraprastha Institute of Information Technology Delhi, for the award of the Master of Technology in Computational Biology, is an original research work conducted by her under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted in part or full to any other University or Institute for the award of any degree/diploma.

Sep,2023
Delhi

Dr. Gaurav Ahuja,
Associate Professor,
Department of Computational Biology,
Indraprastha Institute of Information Technology, Delhi,
New Delhi 110020

Declaration

I submit this project entitled "**Decoding the molecular basis of Iterative Stress Response (ISR)**" to the Department of Computational Biology, Indraprastha Institute of Information Technology, Delhi -110020. I declare that this is my original work carried out under the guidance of Dr. Gaurav Ahuja, Associate Professor, Department of Computational Biology at IIIT-Delhi.

Aug,2023
Delhi

Anmol Aggarwal,
M.Tech Student (Dec 2020 - Sep 2023),
Department of Computational Biology,
Indraprastha Institute of Information Technology, Delhi,
New Delhi 110020

Acknowledgement

I would like to express my deepest gratitude to all those who have supported and guided me throughout this research journey, without whom this thesis would not have been possible.

First and foremost, I am immensely thankful to my supervisor, Dr. Gaurav Ahuja, for their unwavering guidance, invaluable insights, and continuous encouragement. His expertise and dedication have been instrumental in shaping the direction of this research.

I extend my thanks to Ms. Aayushi Mittal, my mentor, and a PhD student in our lab. Her assistance in aligning and channelizing my work, timely inputs, and unwavering support in addressing all my doubts, whether trivial or complex, are deeply appreciated.

I would also like to express my gratitude to the faculty members and staff of the Department of Computational Biology at IIIT Delhi. Their constant support has been invaluable throughout our college journey. A special note of appreciation goes to the IT helpdesk for their ongoing assistance in granting access to the college's IT infrastructure.

Lastly, I extend my gratitude to my constants: my parents and siblings, along with all my lab mates and batchmates, for their continuous support throughout my dissertation journey.

Index

List of figures

List of Tables

ABSTRACT

I – Abstract:

Across evolutionary history, every organism has developed the ability to adapt to fluctuations in environmental conditions, thereby striking a balance between efficient growth and survival. In the case of yeast, when exposed to mild stress, it triggers an enhanced tolerance towards subsequent and magnified stresses. This adaptive response showcases yeast's ability to proactively prepare for future challenges, enhancing its survival and adaptability in a dynamic environment. Former studies have demonstrated this phenomenon with varied intensity and exposure to stress, elucidating the genetic and molecular processes that underpin cellular adaptation to elevated temperatures. In an effort to comprehend the genomic expression responses of the budding yeast Saccharomyces cerevisiae, we conducted a novel experiment where thermal stress was given to yeast in a pulsing manner at fixed intervals. We could identify a subpopulation within cells given conventional heat shock with more than 60% proportion in Pulsed samples. A comparative analysis of gene profiling showed that prominent HSPs differentially expressed in cells undergoing thermal stress in a pulsed manner behave exactly opposite in the case of regular thermal-stress population, denoting yeast’s ability to preserve the direction of transcriptome regulation and regulate it’s intensity with recurring stress stimuli.

Introduction

II - Introduction:

II.1 - Yeast as a model organism:

Yeast (*Saccharomyces cerevisiae*) is a quintessential model organism for studying heat stress, owing to its genetic versatility, conserved cellular processes, and practical advantages. The advanced state of yeast's genetic and molecular toolkit has consequently established it as the principal foundation for creating numerous high-throughput methodologies encompassing transcriptome, proteome, and metabolome screenings[5]. It's simple yet manipulable genome enables precise genetic experiments, unveiling molecular pathways in heat stress responses. Yeast growth could be externally controlled, benefiting from its rapid replication rate (90 minutes). Additionally, strains with removed genes, labelled proteins, and comprehensive databases detailing Protein-Protein Interactions (PPI), subcellular positioning, and gene regulation collectively facilitate a swifter comprehension of biological facets. Evolutionary conservation allows findings in yeast to shed light on analogous mechanisms in higher organisms[6]. Its rapid reproduction accelerates evolutionary studies under heat stress. The well-understood heat shock response in yeast forms a robust foundation for investigating stress reactions.

II.2 - Yeast in Thermal-stress studies:

Using yeast as a model organism in thermal-stress studies has proven to be a fruitful endeavour, yielding valuable insights into the cellular response to temperature fluctuations and offering potential implications for understanding human pathology. Studying yeast's response to heat stress holds immense value for understanding human diseases, as it uncovers conserved molecular mechanisms. In recent times, yeast models that are genetically modified have emerged as potent instruments for investigating the molecular foundation of intricate human ailments such as neurodegeneration, specifically targeting intrinsic or foreign proteins associated with the onset of the disease[7].

Table 1: Cross-Species Conservation of HSR Proteins and Associated Human Diseases			
S. cerevisiae	***H. sapiens***	**Function**	**Disease-Associated**
HSP70 Family SSA1, SSA2, SSA3	HSPA1A, HSPA1B, HSPA6	Protein folding and aggregation	Alzheimer's, Parkinson's, and Huntington's diseases
HSP90 Family HSP82	HSP90AA1, HSP90AB1	Protein folding and stabilization	Cancer and neurodegenerative disorders
HSP60 (Chaperonin Family)	HSPD1	Folding newly synthesized protein	Mitochondrial diseases and neurodegenerative disorders
HSF1	HSF1	Activating expression of HSR protein	Cancer, neurodegenerative diseases, and aging-related disorders
HSP27	HSPB1	Protecting cells from stress-induced apoptosis	Charcot-Marie-Tooth disease (CMT2F)
SOD1,SOD2	SOD1, SOD2	Scavenge ROS	Amyotrophic lateral sclerosis (ALS)
CAT(Catalase) CTT1	CAT	Protect cells from H_2O_2	Neurodegenerative diseases and cardiovascular disorders
Ubiquitin UBI4	UBB, UBC	Protein degradation	Parkinson's and Alzheimer's
SSA4	HSPA4	Protein folding and quality control	Kawasaki, neurodegenerative disease; and schizophrenia
Thioredoxins (TRX) TRX2	TXN, TXN2	Maintain cellular redox balance	Alzheimer’s, Parkinson’s, Huntington’s, Brain Stroke, Multiple Sclerosis
HOG1 (MAPK) HOG1	MAPK14 (p38 MAP Kinase)	High Osmolarity Glycerol (HOG) pathway	Inflammatory diseases, cardiovascular disorders, and cancer
MSH2, MSH6	MSH2, MSH6	DNA mismatch repair	High-grade dysplasia
RAD51	RAD51	DNA repair, genomic integrity	Breast and ovarian cancers
CPR1	PPIA, PPIB, PPIC	Histone deacetylation, regulation of meiosis	Inflammatory diseases, viral infections, and cancer
SIS1	DnaJB1	Involved in proteasomal degradation of misfolded cytosolic proteins	Neuromuscular disorders

II.2 - Limitations in utilizing Yeast for Thermotolerance Exploration:

Yeast (*Saccharomyces cerevisiae*) is not an ideal model organism to study heat stress due to its lack of a fully developed thermoregulatory system, limited cellular complexity, and absence of specific stress response pathways found in higher organisms. While yeast has provided valuable insights into heat stress responses, its simplistic nature and differences from more complex organisms pose challenges for directly translating findings to other systems.

S. cerevisiae has undergone significant genetic manipulation and domestication, which may affect the relevance of stress responses to natural systems. Yeast lacks behaviours often associated with animal stress, such as mobility, avoidance, or complex physiological reactions. While many stress response pathways are conserved between yeast and higher organisms, some paths may differ significantly. This can limit the direct translation of findings from yeast stress studies to more complex organisms. For example, research on the unfolded protein response (UPR) in yeast might only partially represent the UPR in mammalian cells[4]. Additionally, Yeast metabolism differs from that of higher organisms. The metabolic characteristics of yeast affect how stressors are sensed and responded to, which can limit the relevance of findings to other organisms. Neurological disorders are often driven by specific pathways and molecular mechanisms unique to higher organisms. Yeast's lack of a nervous system makes it less relevant for testing potential treatments for neurological disorders. Hence, drugs showing promise in yeast-based assays might translate to something other than effective treatments for human neurological conditions.

In conclusion, while yeast may not fully mimic heat stress responses in higher organisms, it remains a valuable tool for understanding fundamental cellular mechanisms. By integrating advanced technologies, conducting comparative studies, and using genetic engineering approaches, researchers can enhance the relevance and applicability of yeast-based heat stress studies.

Literature Review

II - Literature Survey:

In the last 35 years, substantial advancements have occurred in comprehending how yeast cells react to heat stress. This involves the discovery of precise transcription factors accountable for controlling gene activity, as well as the metabolic adjustments that empower cells to withstand extended periods of slight temperature increases without being overwhelmed by the pressure.

II.1 - A Chronological Overview:

In the 1940s and 1950s, early genetic studies in yeast laid the groundwork for understanding its inheritance patterns and the role of genes in cellular processes. In the 1960s, pioneering research by Lee Hartwell and colleagues identified essential genes involved in yeast cell cycle regulation. The first use of Yeast as a model organism to study function conservation dates back to 1985 when the sequence for mammalian *ras* proto-oncogene was expressed in *S. cerevisiae* lacking homologous genes (*RAS1* and *RAS2*) essential for its viability. The restored viability indicated a profound conservation extending beyond the sequence and encompassing intricate biological functions[5]. Later, a seminal study by Lindquist and Craig laid the foundation for understanding the heat shock response in yeast. The researchers demonstrated that yeast cells respond to elevated temperatures by inducing the expression of heat shock proteins (HSPs). They identified the heat shock transcription factor (HSF) responsible for activating HSP gene expression. This work contributed to unravelling the molecular mechanisms underlying the cellular response to thermal stress[6]. For example, HSPs are linked with various cellular roles such as constructing large molecular structures, moving proteins, regulating transcription, orchestrating programmed cell death, and disassembling as well as restructuring aggregates of proteins that have been altered due to stress[7,8,9,10]. In a subsequent investigation, researchers pinpointed the HSP104 gene as the primary contributor to yeast's acquired ability to withstand higher temperatures, which was also crucial for the prolonged lifespan resulting from short-term heat exposure. An association between mitochondrial petite mutations and heat stress was discovered, implying that mitochondria might be vital for extending lifespan through brief heat challenges. The findings propose the potential involvement of RAS genes and mitochondria in the

epigenetic transfer of the decreased mortality advantage facilitated by mild, temporary heat stress[11].

In the 1990s, researchers identified the Unfolded Protein Response (UPR) in yeast. This stress response mechanism is activated when the endoplasmic reticulum (ER) becomes overwhelmed by unfolded or misfolded proteins, triggering a series of events to restore protein homeostasis. During the 2000s, noteworthy progress was made in understanding the oxidative stress response in yeast. Studies revealed how yeast cells manage reactive oxygen species (ROS) and adapt to oxidative stress by activating specific signalling pathways and antioxidant defense mechanisms[12]. Understanding these regulatory mechanisms provided insights into how yeast cells sense and cope with heat stress at a molecular level. Queitsch et al. (2002) explored the role of yeast Hsp90 in mediating evolutionary change, specifically by affecting developmental transitions[13]. It provided a link between heat shock proteins and evolutionary processes. More studies gave insight to how modulating Hsp90's buffering capacity offers a valuable approach to unlocking cryptic genetic variation and gaining insights into the complex interactions among genotypes, environmental factors, and stochastic events that govern phenotypic expression[14,15]. In the 2010s, research focused on the Target of Rapamycin (TOR) pathway's role in yeast stress response. TOR is a central regulator of cell growth and metabolism, and its modulation was found to influence yeast adaptation to various stress conditions, including nutrient deprivation[16].

Over the past decade, as Systems Biology and Computational Modelling gained momentum, researchers started using integrative approaches to study the heat stress response in yeast comprehensively. These studies identified conserved pathways and regulatory networks. They highlighted potential therapeutic targets for human pathologies, suggesting that modulating the expression or activity of specific HSPs might have therapeutic potential in treating protein misfolding diseases[17]. For instance, activating heat shock proteins (HSPs) could protect against protein misfolding and aggregation, a hallmark of neurodegenerative diseases like Alzheimer's, Parkinson's, and Huntington's disease[18,19]. Certain HSPs, including sphingolipidoses, have been explored as potential therapeutic targets for treating lysosomal storage disorders. These disorders are characterized by the accumulation of specific lipids within lysosomes, leading to cellular dysfunction. HSP modulation may enhance the

chaperone machinery and facilitate the proper folding and trafficking of lysosomal enzymes, potentially improving disease phenotypes[20,21,22].

II.2 - Global analysis of the Heat shock response:

Much like the investigation of transcriptional networks in yeast, the advent of DNA microarray technology has bestowed an unprecedented comprehension of the far-reaching consequences of the heat shock response throughout the complete genome. Two comprehensive research endeavours employed microarray methods to unveil shifts in gene expression under diverse stress circumstances, encompassing heat shock, osmotic stress, and nutrient scarcity. Impressively, around 10% of the genome experienced substantial expression alterations during one or more of these stress factors, underscoring the profound impact of stress-triggered transcription on the overall transcriptome[23,24].

The comprehensive analysis of genome-wide studies can discern a multitude of additional insights. It becomes evident that the Heat Shock Response (HSR) exhibits a proportionate relationship with the intensity of the stress imposed. Specifically, when subjected to a temperature shift from 25° to 37°, the HSR demonstrates a prolonged duration and a more pronounced amplitude of change in gene expression in comparison to a shift spanning 29° to 33°. The data indicate that yeast cells can detect gradual changes in temperature stress until they reach a critical threshold point, beyond which the system, particularly Hsf1, becomes maximally activated[25,26]. Notably, a small subset of Hsf1 target genes, such as CUP1 and HSP82, exhibit significantly more robust induction at 39° compared to 37°, strongly suggesting that this threshold lies within the temperature range of 39° to 40°.

In the absence of stress, cells show moderate resistance to various environmental challenges. However, their capacity to endure future insults is significantly enhanced when subjected to mild stress. A key factor contributing to cytoprotection during the Heat Shock Response (HSR) is the disaccharide trehalose[27]. Extensive research in the late 1990s established trehalose as a potent stabilizer of proteins and membranes in diverse biological systems, including yeast[28].

The yeast "chaperome": The introduction of whole genome sequencing has enabled the identification and characterization of the complete chaperones within organisms,

collectively known as the chaperome[29]. This concept includes both chaperones and other Heat Shock Proteins (HSPs) that exist during regular growth circumstances and those whose levels rise or are specifically generated in response to the Heat Shock Response (HSR).

In the past, HSPs of considerable importance were primarily identified by their prevalence during heat shock. This was because their augmented synthesis made them easily detectable through radioactive pulse-chase analysis. Meanwhile, the synthesis of the remaining proteome was concurrently suppressed[30]. These heat shock proteins (HSPs) acquired their names based on their observed molecular sizes, leading to the familiar group of Hsp100, Hsp90, Hsp70, Hsp60, along with the small HSPs that are present universally in eukaryotic cells.

Table 2: Chaperones of the Hsp70 type in the yeast *Saccharomyces cerevisiae*

Gene Name	Allocation	Role	References
Ssa1, Ssa2, Ssa3, Ssa4	Intracellular fluid	Regular folding	Werner-Washburne *et al.* (1987)
Ssb1, Ssb2, Ssz1	Ribosomes	Folding of nascent protein	Nelson *et al.* (1992)
Sse1, Sse2	Intracellular fluid	Binding with substrate	Dragovic *et al.* (2006); Raviol *et al.* (2006); Shaner *et al.* (2006)
Kar2	Endoplasmic Reticulum	Regular folding	Vogel *et al.* (1990)
Lhs1	Endoplasmic Reticulum	Binding with substrate	Baxter *et al.* (1996); Steel *et al.* (2004)
Ssc1, Ecm10	Mitochondria	Post import modification	Craig *et al.* (1989), Baumann *et al.* (2000)
Ssq1	Mitochondria	Iron/Sulphur assembly	Dutkiewicz *et al.* (2003)

Cellular stress response: The cellular stress response is a unifying biological principle. It proposes that in response to abrupt environmental changes, nearly all life forms induce a set of conserved proteins. This response addresses macromolecular

damage, with similar types of cellular damage caused by various environmental stressors leading to the up-regulation of a shared gene set. These environmentally responsive proteins, including molecular chaperones, DNA damage and repair proteins, proteolytic enzymes, and specific metabolic enzymes, mitigate macromolecular damage and stabilize essential cellular functions. Notably, these proteins involved in the cellular stress response exhibit significant conservation across the three super kingdoms of life, underscoring their fundamental importance[31,32].

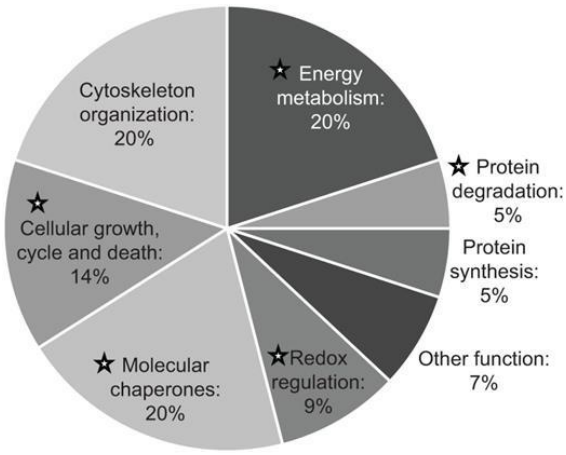


Figure 1: Classification of proteins detected through comparative proteomics using gene ontology[33]

The reaction to Oxidative Stress: Although elevated temperature acts as the main initiator during heat shock, as previously mentioned, a notable subsequent result involves the creation of reactive oxygen species (ROS). All living organisms encounter ROS due to regular aerobic metabolism or exposure to substances that produce radicals[34]. However, like all living organisms, S. cerevisiae possesses efficient antioxidant defense mechanisms. These mechanisms neutralize ROS as they are made and maintain a reduced intracellular redox environment. An oxidative stress situation arises when the production of ROS overwhelms these defense mechanisms, leading to genetic deterioration and disruptions in physiological function. Ultimately, this cascade can culminate in cellular demise. The arsenal of antioxidant defences encompasses a variety of protective enzymes located within distinct subcellular compartments, and their expression can be heightened in response to exposure to ROS. Non-enzymic reasons include small molecules that serve as scavengers for free radicals. Presently, only ascorbic acid and GSH have undergone thorough investigation and characterization within yeast's nonenzymatic defense mechanism.[3]

II.3 - Computational Progress:

Advancements in computational techniques have revolutionized stress research in yeast, leading to groundbreaking insights into cellular responses and molecular mechanisms. These innovations have enabled scientists to probe stress-related phenomena with unprecedented depth and precision, enhancing our understanding of yeast biology and its implications for broader biological systems.

1. Genome-Wide Analysis: Computational tools have allowed researchers to perform genome-wide analyses of yeast under stress conditions. This approach has revealed intricate gene expression patterns, uncovering critical regulatory networks involved in stress response. High-throughput sequencing and bioinformatics pipelines have identified stress-responsive genes, promoters, and regulatory elements.

2. Systems Biology: Integration of computational modelling and experimental data has facilitated the development of comprehensive models of stress responses in yeast. These models simulate the complex interactions between genes, proteins, and metabolites, shedding light on how different components collaborate to mitigate stress. Systems biology approaches have illuminated feedback loops and crosstalk between stress pathways, enhancing our grasp of the holistic cellular response.

3. Network Analysis: Computational techniques have empowered the construction and analysis of molecular interaction networks in yeast. By mapping protein-protein interactions and genetic pathways, researchers have discerned how stress-related proteins coordinate their actions. This network perspective has unveiled previously unrecognized connections and bottlenecks in stress signalling cascades.

4. Structural Biology: Advanced computational methods, such as molecular dynamics simulations and structural modelling, have elucidated the three-dimensional structures of stress-responsive proteins. These insights offer a detailed view of protein conformational changes during stress and how these alterations affect protein function. Understanding structural dynamics has paved the way for rational drug design targeting stress-related proteins.

5. Evolutionary Analysis: Comparative genomics and phylogenetic approaches, driven by computational algorithms, have enabled the study of stress responses across yeast species. By analysing evolutionary conservation and divergence of stress-related genes, researchers have unearthed ancient and lineage-specific stress adaptation strategies, offering insights into the adaptive evolution of stress responses.

6. Machine Learning and AI: Machine learning algorithms have been instrumental in predicting stress-responsive elements, classifying stress types, and identifying novel regulatory motifs in yeast genomes. AI-driven models have accelerated the analysis of large-scale datasets, enabling the extraction of meaningful patterns and predictive insights that guide experimental design.

METHODOLOGY

IV - Methodology:

IV.1 - Experimental design of PULSER Experiment:

The terminology 'Pulser' was adopted due to its correlation with the specific technique employed to induce heat stress in the test sample. The investigation into the transcriptome of yeast cells experiencing thermal pulsation encompassed a comprehensive analysis undertaken through a series of sequential steps:

1. A 50µl inoculum (derived from a glycerol stock) of the yeast strain BY4741 of S. cerevisiae (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) was cultured overnight in 5 ml of autoclaved YPD medium at 30°C with agitation at 150rpm. This is secondary culture.
2. The OD600 reading was approximately 0.7 at the 20-hour mark. 3 replicates were created for both the Test and Control samples by transferring 100µl of the secondary culture into the first row of Deepwell plates. Subsequently, 1.9ml of YPD media was added to each well to achieve a total volume of 2ml for the culture. The initial two columns of the second row in these plates served as blanks, containing 2ml of YPD media.
3. During Cycle 0, the procedure consisted of placing both plates in the incubator at 30°C for 30 minutes. Following this, OD600 measurements were taken using a microplate reader.
4. In the subsequent 15 cycles, the Test plate underwent a 'pulsing' procedure involving alternating temperature conditions of 30°C and 37°C for 15-minute intervals, totalling 30 minutes. In contrast, the Control plate was kept undisturbed at 30°C.
5. RNA was isolated and subsequently forwarded for sequencing analysis.

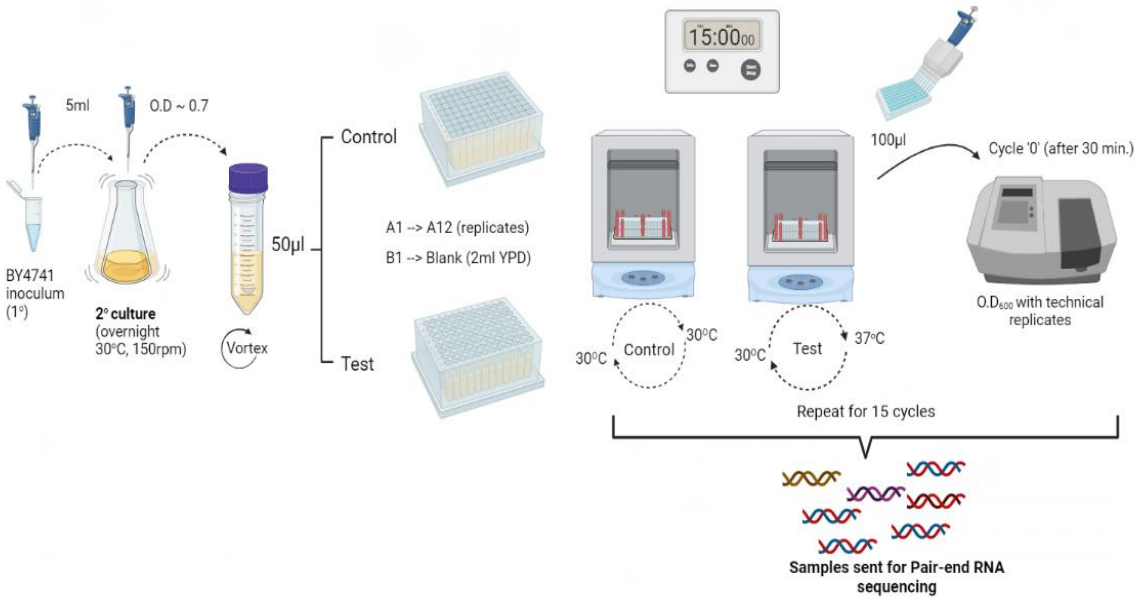


Figure 2: Schematic workflow of PULSER

IV.2 - RNA-seq read alignment and read counting:

Illumina paired-end sequencing generated two fastq files per sample. Bioconductor package, Rsubread, was used to convert fastq files to raw count matrix.

IV.2.1 - The Seed-and-vote mapping paradigm:

This strategy to map the genome employs a collection of overlapping seeds, known as subreads, from each individual read. All seeds collectively influence the optimal position for the read. Subsequently, the algorithm employs conventional alignment methods to elucidate specific mismatches and indel details amid the subreads within the prevailing voting block. The provided illustration depicts the proposed seed-and-vote mapping method through a simplified example.

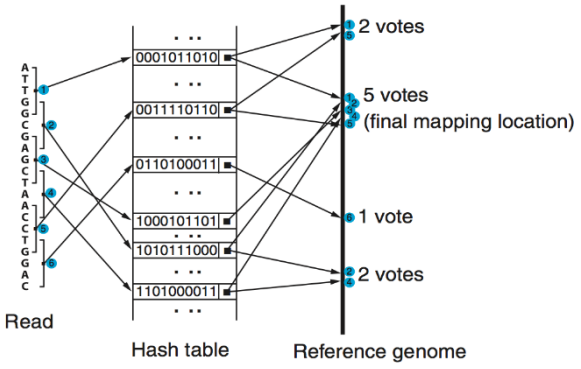


Figure 3: Illustration of seed and vote mapping[36]

For mapping of paired-end reads was achieved, the following equation was used to derive a collection of potential mapping positions for every pair of reads:

$$PE_{score} = w*(V_1 + V_2)$$

V_1 and V_2 represent how many votes each of the two reads in a pair gets. The variable 'w' is set to 1.3 if the two reads are mapped close together as expected, and it's one if they are not. This helps us decide how much weight to give to the votes based on the distance between the mapped positions of the reads.

IV.2.2 - FastQC:

Quality assessments of the post-trim and pre-trim sequencing data were performed using FastQC. It calculates a quality score for each base in every read. This score is usually represented as a Phred score, which quantifies the base-calling accuracy. FastQC plots these quality scores across the length of the reads, helping you identify regions with poor or fluctuating quality.

The Phred score is connected to the base-calling error probability (P) through a logarithmic relationship, which is defined by the following formula:

$$Q = -10 * \log_{10}(P)$$

Here, Q is the Phred score, and P is the base-calling error probability. A higher Phred score corresponds to a lower error probability and better quality.

IV.3 - Data Transformation and Statistical Analysis:

iDEP (integrated Differential Expression and Pathway analysis) was used for a thorough examination of the count matrix of the PULSER experiment.

IV.3.1 - Regularized log (rlog) transformation:

The first step in pre-processing of count data was rlog transformation which stabilized the variance across the range of expression values while preserving the underlying biological variability. After adding a pseudo count value to each count with zero value, the counts were normalized by the library size (total number of reads) of each sample. This step ensured that samples with high sequencing depth didn’t dominate the transformation. The regularized log transformation was applied to the normalized counts.

The rlog-transformed values, denoted as x_{ij}, can be calculated using the following formula:

$$x_{ij} = \log\left(\frac{y_{ij} + \sqrt{y_{ij}^2 + \alpha^2}}{s_j}\right)$$

Where y_{ij} is the normalized count for gene i and sample j, α is a regularization parameter that prevents division by zero. And s_j is a sample-specific size factor (normalized library size) for sample j.

IV.3.2 - Principal Component Analysis (PCA):

Biological samples were projected onto a Principal Component Analysis (PCA) plot to visualize and analysed their relationships and variations in a lower-dimensional subspace derived from the original data. PCA uses eigenvalues and eigenvectors to reduce the dimensionality of data. It calculates the covariance matrix, finds eigenvectors with corresponding eigenvalues, and ranks them by variance. These eigenvectors define a new coordinate system. Data points are projected onto these components, forming a reduced-dimensional plot.

IV.3.3 - Multidimensional scaling (MDS):

MDS was performed on counts data to explore the relationship between samples. As opposed to PCA, MDS is focused on preserving the original relationships or rankings between objects. It does not necessarily aim to capture the maximum variance in the data.

IV.3.4 - Elbow Method:

To determine the optimal number of clusters for clustering analysis, an Elbow plot was constructed. In this approach for determining the value of 'k,' continuous iteration over the range from 'k=1' to 'k=n' is performed. For each 'k' value, we compute the sum of squared distances within clusters, known as the within-cluster sum of squares (WCSS) value.

IV.3.5 - K- means Clustering:

Normalized count data was subjected to K-means clustering to visualize the clusters derived from the Elbow plot. K-means clustering, a commonly used unsupervised learning technique for examining data clusters, begins by randomly selecting an initial

group of centroids. These initial centroids act as reference points for each cluster, and then, through iterative calculations, the algorithm refines the placement of these centroids.

IV.4 - Analysis of genes with differential expression (DEGs):

DESeq2, an R package was used to determine and visualize the significantly differentially expressed genes. It employs a negative binomial distribution model to identify genes that show statistically significant differences in expression levels between different conditions or groups. DESeq2 performs normalization, estimation of variance-mean dependence, and statistical testing to assess differential expression, providing valuable insights into gene regulation and biological processes.

IV.4.1 - Volcano Plot:

The plot displayed the interplay between the statistical significance ($-\log_{10}$ p-value) and the magnitude of fold change for 4668 features on being compared between the two groups (Control and Test). A cutoff of less than 0.1 for the false discovery rate (FDR) and a fold change greater than 2 were utilized to assign distinct color shades to upregulated and downregulated genes in a selective manner.

IV.4.2 - Hierarchical clustering tree:

Enrichment analysis of the DEGs was conducted using the hypergeometric distribution. Due to the presence of duplications in various GO terms, the measurement of inter-term distance was conducted by calculating the percentage of genes shared. This distance metric was then employed to create a hierarchical clustering tree for GO terms.

IV.4.3 - Protein-protein interactions (PPI):

Utilizing the STRINGdb package, iDEP analysed sets of DEGs by employing the STRING API[37] to perform enrichment analysis and retrieve PPI networks. It achieves this by integrating diverse sources of biological data, such as experimental evidence, computational predictions, and curated databases, to construct comprehensive and context-specific PPI networks.

IV.5 - Pathway Analysis:

To uncover the underlying biological processes that are affected by changes in gene expression, pathway analysis was performed using KEGG as gene sets.

IV.5.1 - PGSEA:

PGSEA (Parametric Gene Set Enrichment Analysis) is a statistical method and software tool used for gene set enrichment analysis. PGSEA fits a parametric distribution (Gaussian) to the gene expression values within each gene set and compares the fitted distribution of the gene set with the overall distribution of gene expression values in the dataset. We run PGSEA for pathway analysis on our bulk RNA-seq dataset using the gene set database as KEGG.

IV.6 - Seurat:

Aimed at exploring the molecular dynamics of non-pulsed heat-shocked *S.cerevisiae*, single-cell RNA-seq data of cells given 42℃ thermal shock, mDrop-seq datasets[38] were downloaded from the NCBI Gene Expression Omnibus (GEO). Both Test and Control, consisting of two biological replicates, were integrated using Seurat[39].

IV.6.1 - Linear dimensionality reduction:

The integrated Seurat object (6794 Features x 11646 Cells) was subjected to quality control, normalization, and scaling. A subset of data with 2000 most variable features was subjected to Principal Component Analysis. The expression values of these genes were transformed into principal component scores, and the top principal components were retained based on their explained variance. Further, the original gene expression data was projected onto the selected principal components to reduce its dimensionality. The number of principal components was decided by taking the minimum of the following two metrics -

1. The point at which the principal components account for only 5% of the standard deviation individually, while their cumulative contribution reaches 90% of the standard deviation.
2. The point at which the successive principal components exhibit a percent change in variation of less than 0.1%.

IV.6.2 - Jackstraw Plot:

The Jackstraw Plot was used to assess the significance of principal components (PCs) obtained through linear dimensionality reduction. This involves randomly reshuffling data points for each PC while maintaining the original PC scores. After reshuffling, a

new PCA is conducted to generate alternative principal components. Comparison between permuted and original PCs aims to gauge their statistical difference. The plot displays negative logarithmic p-values for this comparison, with each point representing a principal component.

IV.6.3 -Silhouette Analysis:

The selection of the resolution value yielding the highest Silhouette score and the optimal number of clusters was achieved using the chooseR tool, a framework[40] based on subsampling. The silhouette score *S* for a data point is given by:

$S = (b-a)/ \max(a, b)$

Here, a is the average distance between the data point and other data points in the same cluster (cohesion), b is the smallest average distance between the data point and data points in a different cluster (separation).

IV.6.4 – Clustering:

Clusters of cells were detected at the obtained resolution following silhouette analysis, utilizing a clustering algorithm based on shared nearest neighbor (SNN) modularity optimization. Initially, the algorithm computed k-nearest neighbors and established the SNN graph. Subsequently, the modularity function was iteratively refined using the Louvain algorithm to delineate distinct clusters. The Louvain algorithm, tailored for identifying communities in intricate networks, systematically refines node grouping to enhance modularity, a metric signifying the robustness of community organization.

IV.6.5 - Non-linear dimensionality reduction:

After the clustering step, we employed Uniform Manifold Approximation and Projection (UMAP) as well as t-distributed Stochastic Neighbor Embedding (t-SNE) algorithms for nonlinear dimensionality reduction. These methodologies were employed with the overarching aim of constructing reduced-dimensional embeddings for high-dimensional datasets, thereby facilitating the preservation of local and global similarities among cells within the manifold space.

IV.7 -Similarity Index:

Three analyses involving similarity coefficients were conducted, and the results were depicted using a heatmap to identify resemblances between the mean expression

vectors of cell types from sc-RNA sequencing and those derived from PULSER samples.

IV.7.1 - Correlation:

Pearson correlation coefficient "r" was calculated between two variables, "x" and "y", based on a set of paired data points (x_i, y_i), is as follows:

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})\,(y_i - \overline{y})}{\sqrt{\sum_{\Sigma_i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{\Sigma_i=1}^{n}(y_i - \overline{y})^2}}$$

Where:

- *n* is the number of data points.
- x_i and y_i are the values of the two variables in the i-th data point.
- $\overline{x}$ and $\overline{y}$ are the means (averages) of the respective variables.

IV.7.2 - Covariance:

To assess the direction and strength of the relationship between Cell types and PULSER samples, Covariance between them was determined using the following formula:

$$\text{cov (x, y)} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{N-1}$$

Where:

- N is the number of data points.
- x_i and y_i are the values of the variables in the i-th data point.
- $\overline{x}$ and $\overline{y}$ are the means of the respective variables.

IV.7.3 - Euclidean:

Euclidean distance is a direct measure of geometric distance between two points in the space defined by the features.

$$\text{Euclidean Distance} = \sqrt{\left(x_{A_1} - x_{B_1}\right)^2 + \left(x_{A_2} - x_{B_2}\right)^2 + \ldots\ldots + \left(x_{A_n} - x_{B_n}\right)^2}$$

Where, $x_{A_i}$and x_{B_i} are the values of the i-th feature for samples A and B, respectively. n is the number of features.

IV.8 - MuSiC Deconvolution:

MUlti-Subject SIngle Cell deconvolution (MuSiC)[41] method was employed to deconvolute the cell types of in Bulk RNA-seq data from PULSER experiment using the single-cell RNA-seq (scRNA-seq) data of heat shock stimulated S. cerevisiae at 42℃. The data can be downloaded from - GEO Accession viewer (nih.gov). Beginning with multi-subject single-cell RNA sequencing (scRNA-seq) data, MuSiC functions on the basis that cells within each subject have been categorized into a predefined set of consistent cell types, showcasing consistent characteristics across

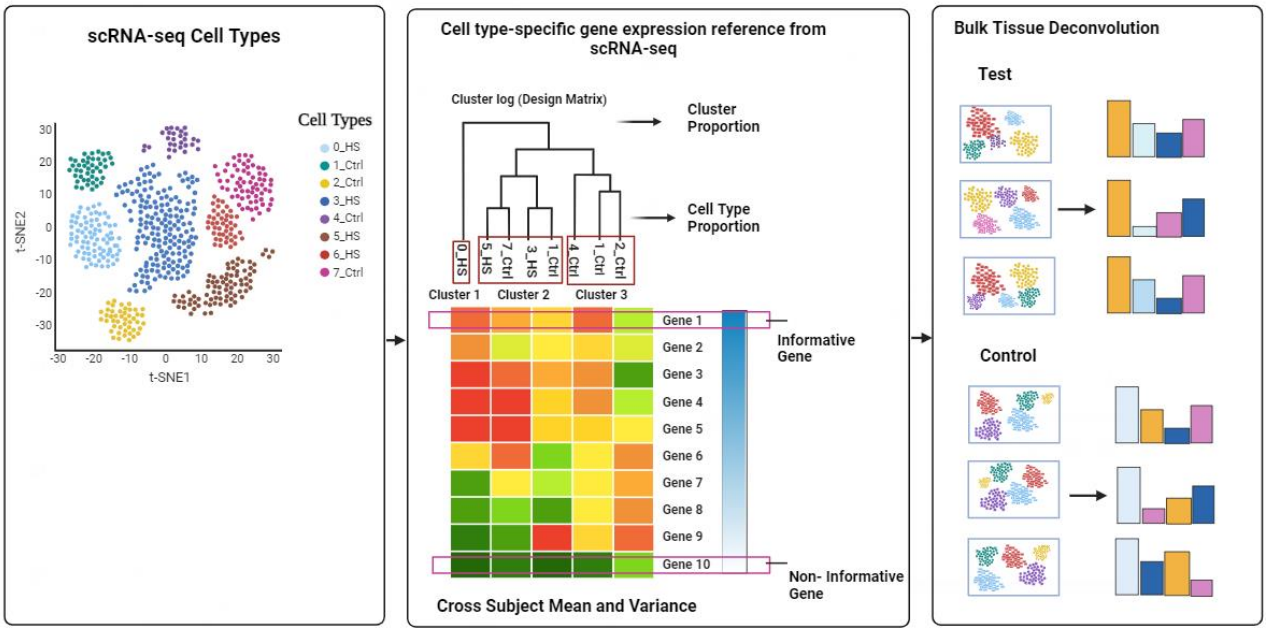


Figure 4: Overview of MuSiC framework.

different subjects. MuSiC's gene weighting mechanism highlights genes that display consistency across subjects, affording greater significance to those with minimal inter-subject variability (termed informative genes). Conversely, genes with higher inter-subject variability (considered less informative) are downplayed.

To mitigate collinearity challenges arising from interrelated cell types within solid tissues, MuSiC employs a tree-guided approach that systematically zooms in on strongly associated cell types. Initially, akin cell types are clustered together into shared clusters, enabling the precise estimation of cluster proportions. Subsequently, this iterative process is applied recursively within each cluster.

IV.9 – Experimental design of Apoptotic Assay:

1. A 50µl inoculum (derived from a glycerol stock) of the yeast strain BY4741 of *S. cerevisiae* (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) was cultured overnight in 5 ml of autoclaved YPD medium at 30°C with agitation at 150rpm. This is secondary culture.
2. The OD600 measurement yielded approximately 0.60 after a 20-hour interval. Subsequently, three Deepwell plates were autoclaved and labelled as Group A, Group B, and Group C, corresponding to distinct thermal conditions of 30°C, 37°C, and 30°C/37°C pulsing, respectively. Each group comprised 95 biological replicates, with a well composition of 95µl media and 5µl culture. Additionally, one well was allocated as a blank control, housing 2ml of YPD media.
3. During Cycle 0, OD600 measurements were captured immediately following the setup.
4. In the subsequent 15 cycles, Group ‘A’ and Group ‘B’ plates were kept in incubator at their respective temperature. However, the Group ‘C’ deep well plate underwent a 'pulsing' procedure involving alternating temperature conditions of 30°C and 37°C for 15-minute intervals, totalling 30 minutes. All these plates were covered with a breath easy membrane to avoid spillage.
5. For a period of 7.5 hours, OD600 was measures for all groups after an interval of 30 minutes.
6. After 15 cycles, all the three groups were further divided into 6 replicates each for Acetic Acid treatment at varied concentrations of 50mM, 100mM and 200mM. The last two rows of setup were kept as blanks for measuring background noise in media. And one column for each group was kept as control, lacking Acetic Acid.
7. Cells were given apoptotic stress for 4 hours and fluorescence of live and dead cells was measured using FDA/PI (15µg/ml) staining solution after an interval of 1 hour incubation for a period of four hours.

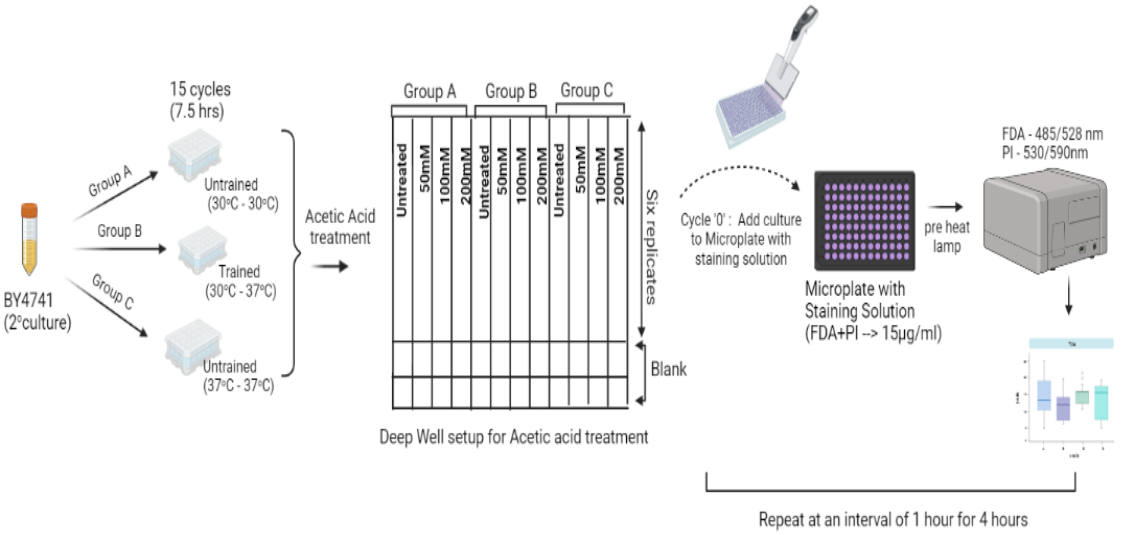


Figure 5: Workflow of Apoptotic Stress Assay

Results and Discussions

V - Results and Discussions:

V.1 – Objective 1: Transcriptome Analysis of thermo-pulsed Yeast cells

The provided files consisted of a read count matrix from the PULSER experiment along with a corresponding design file. Upon uploading the read count data, iDEP accurately identified *Saccharomyces cerevisiae* as the probable species, determined by the quantity of matched gene IDs.

V.1.1 - Pre-processing and EDA

Following Ensembl ID conversion and applying the default filter (requiring a minimum of 0.5 counts per million in at least one sample), the gene pool was reduced to 5067 genes out of the initial 7127. A bar graph depicting the cumulative read counts for each library is produced, revealing limited discrepancies in library sizes. A scatter plot illustrating the relationship between samples demonstrates minor differences among replicates. The distribution of rlog transformed data is visualized through a density plot and a box plot.

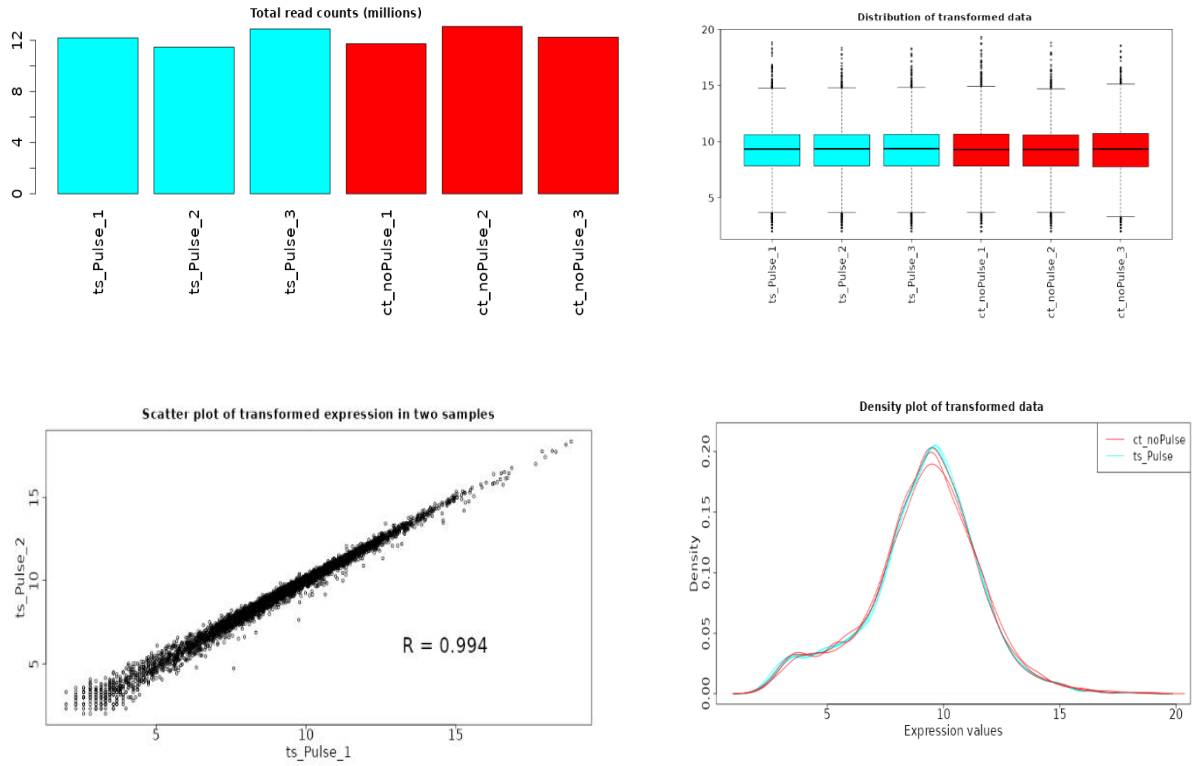


Figure 6: Total read counts per sample and distribution of transformed data.

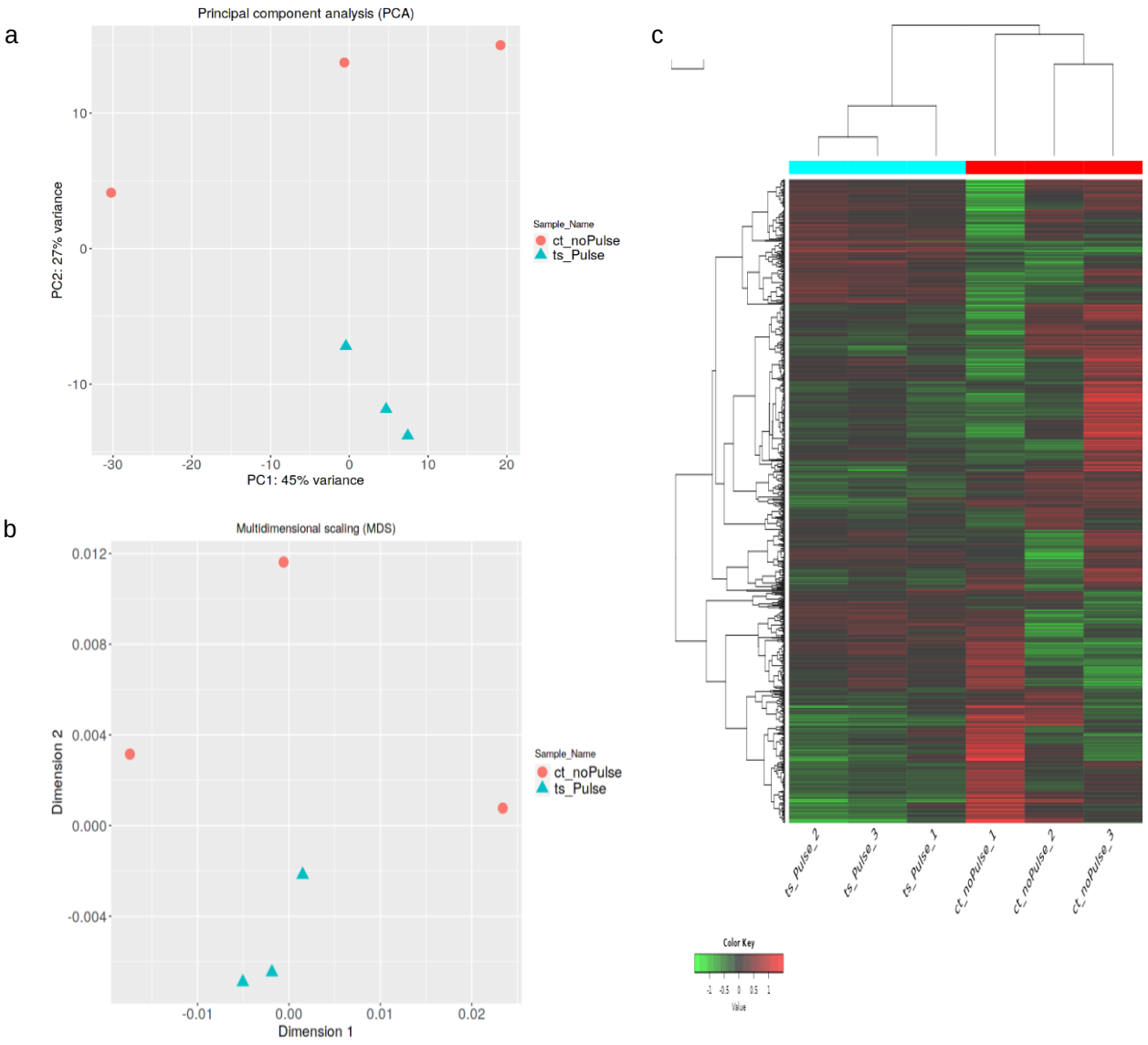


Figure 7: PCA analysis (a), MDS analysis (b), and Hierarchical clustering (c) collectively demonstrate significant dissimilarity among thousands of genes triggered by Thermo-Pulsing. Replicates exhibit minimal diversity.

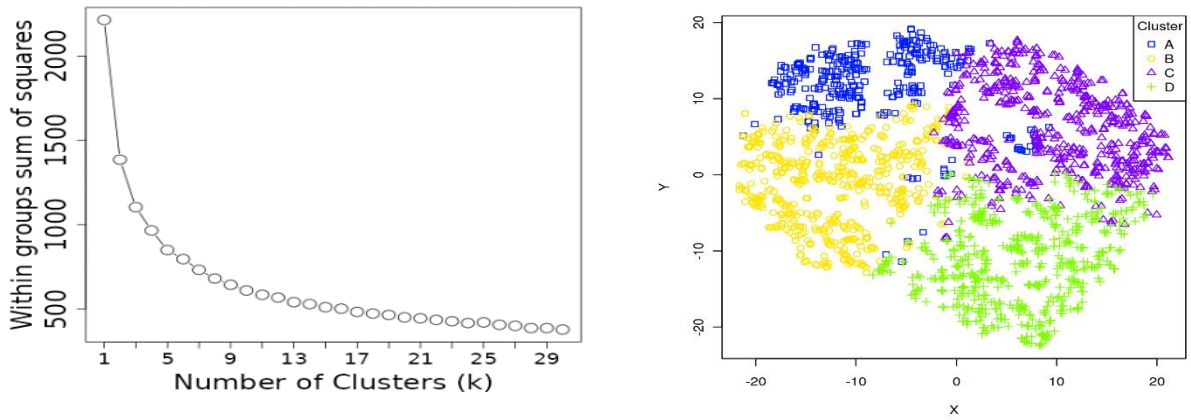


Figure 8: The elbow plot with optimal cluster count as four, and K-means clustering

V.1.2 - Differentially expressed genes (DEGs):

With the DESeq2 package, 10 upregulated and 40 downregulated genes were found employing a criterion threshold of false discovery rate (FDR) < 0.1 and a fold-change exceeding 2. Both the volcano plot and MA plot depict a pronounced transcriptomic response induced by Thermo-pulsing stress.

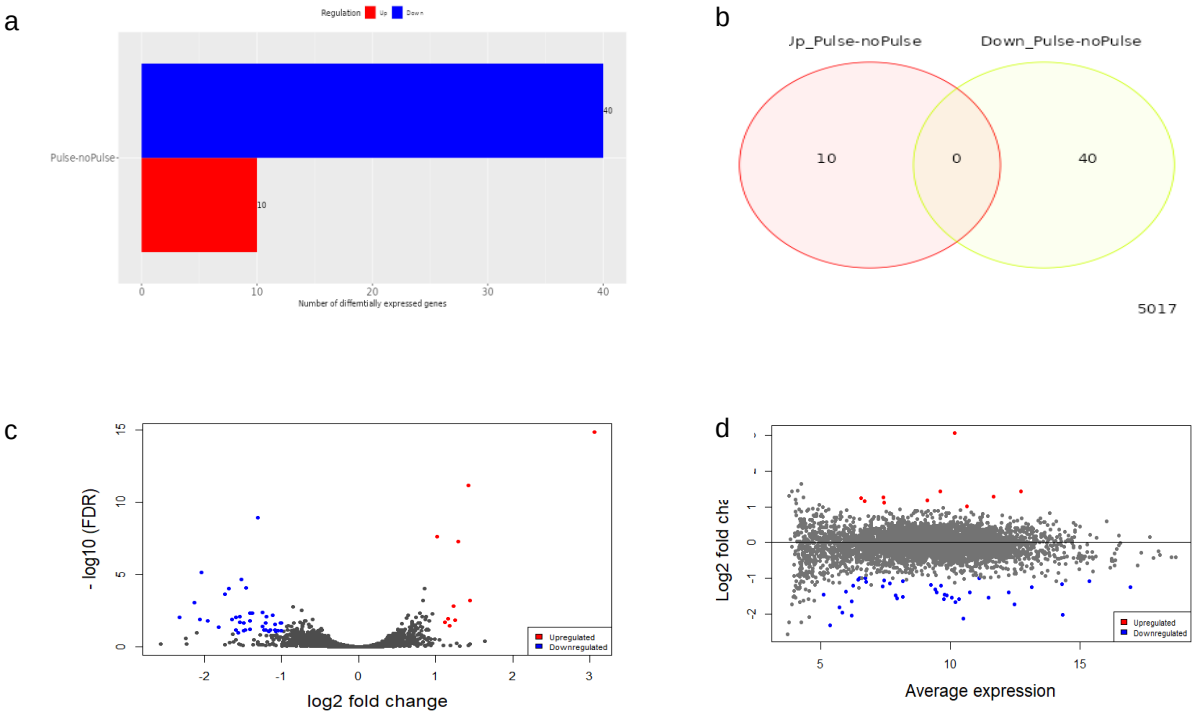


Figure 9: Summary plots for differential expression analysis using DESeq2. a) Barplot showing 10 upregulated and 40 downregulated DEGs with b) no overlap depicted by Venn diagram. (c) Volcano plot and (d) MA plot displaying a peak log fold change value of three

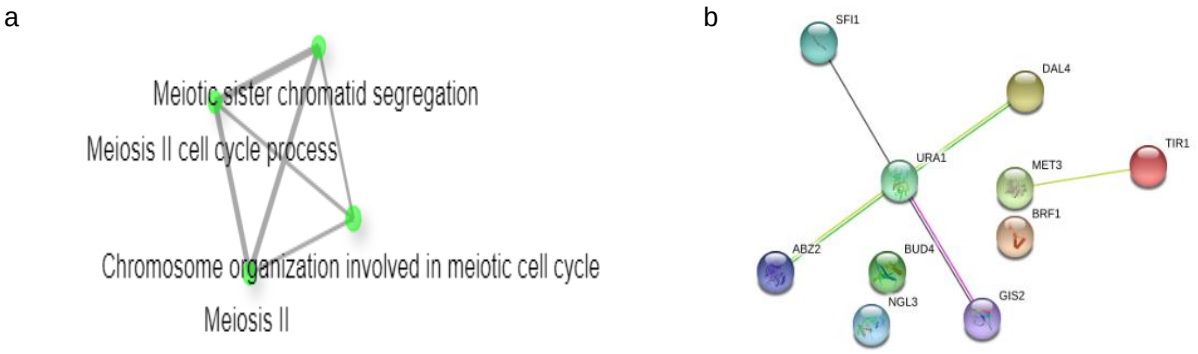


Figure 10: Visualization of enriched GO terms using (a) Network and (b) Protein-protein interactions (PPI) Network

Selecting GO Cellular component within the Gene Ontology framework, it was observed that subjecting *S. cerevisiae* to repeated thermal stress leads to the inhibition of genes responsible for the Meiotic II cell cycle process and the organization of chromosomes during the meiotic cell cycle. Enrichment analysis and the retrieval of Protein-Protein Interaction (PPI) networks unveiled a strongly interconnected network involving URA1, ABZ2, DAL4, and GIS2. These genes are downregulated in heat pulsed yeast culture, implying that the recurring thermal stress may disrupt translation regulation under stress conditions.

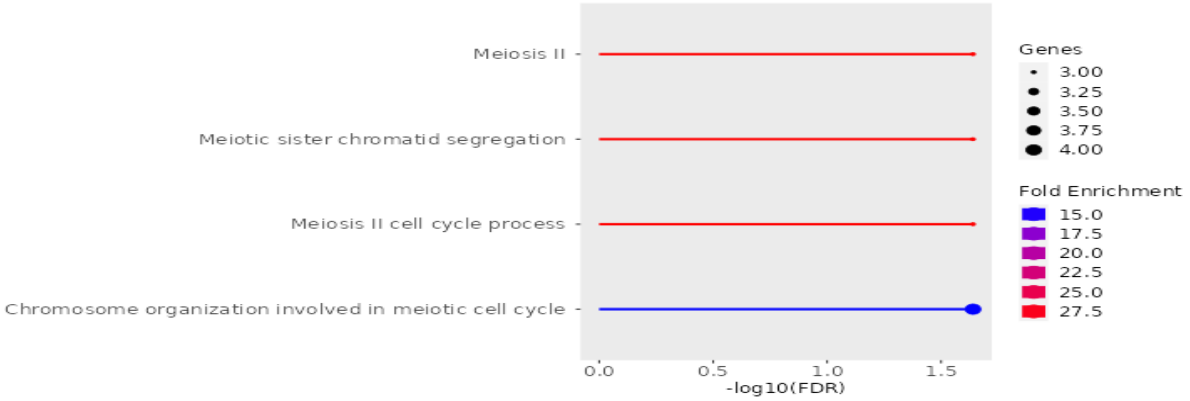


Figure 11: Enrichment plot for DEGs in PULSER experiment

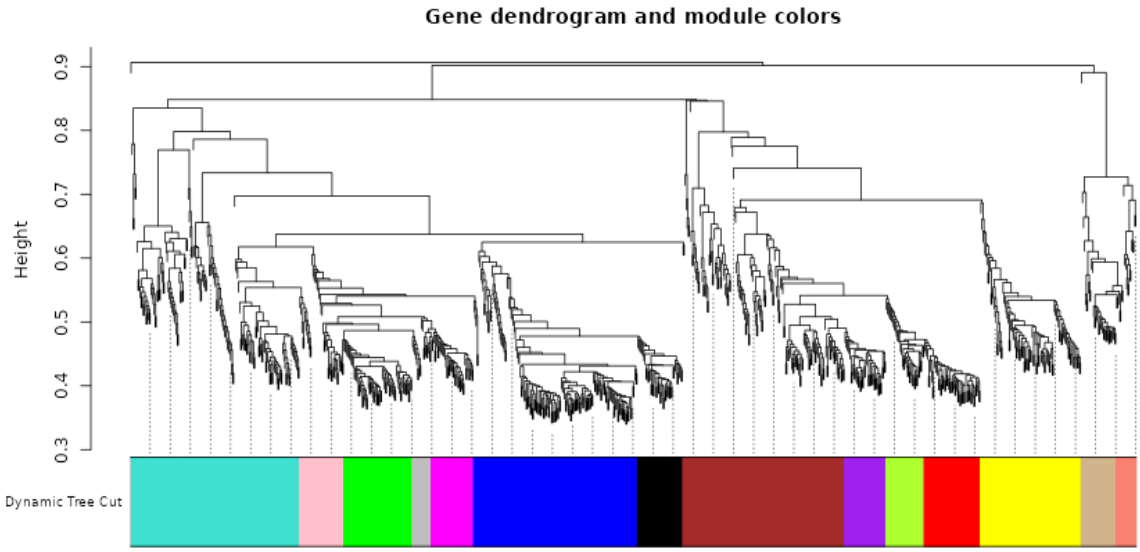


Figure 12: Gene Dendrogram of 1000 most variable genes in PULSER

Table 3: Literature Survey of DEGs and reference of available relevant research					
S.No	Gene	Reference	S.No	Gene	Reference
1	TIR1	https://pubmed.ncbi.nlm.nih.gov/8733242/	26	YEH2	N/A
2	SFI1	https://pubmed.ncbi.nlm.nih.gov/10455233/	27	YOX1	N/A
3	BUD4	https://pubmed.ncbi.nlm.nih.gov/22185758/	28	SNR18	N/A
4	MET3	https://pubmed.ncbi.nlm.nih.gov/15583009/	29	MNT2	N/A
5	BRF1	https://pubmed.ncbi.nlm.nih.gov/31676685/	30	TYS1	N/A
6	GIS2	https://pubmed.ncbi.nlm.nih.gov/30089646/	31	FMP40	N/A
7	DAL4	N/A	32	TRE1	N/A
8	NGL3	N/A	33	CYS3	N/A
9	URA1	N/A	34	IMP1	N/A
10	ABZ2	N/A	35	RUF5-2	N/A
11	SSA4	https://pubmed.ncbi.nlm.nih.gov/12818212/	36	SSL1	N/A
12	RSC4	https://pubmed.ncbi.nlm.nih.gov/20015969/	37	YIG1	N/A
13	RUF5-1	https://pubmed.ncbi.nlm.nih.gov/12853629/	38	RDH54	N/A
14	DSK2	https://pubmed.ncbi.nlm.nih.gov/18199679/	39	DSC3	N/A
15	PEX12	https://mediatum.ub.tum.de/doc/1286993/1286993.pdf	40	SCR1	N/A
16	PSA1	https://pubmed.ncbi.nlm.nih.gov/33439673/	41	SEC9	N/A
17	PHO85	https://pubmed.ncbi.nlm.nih.gov/28637746/	42	BIT2	N/A
18	HSP30	https://pubmed.ncbi.nlm.nih.gov/22654157/	43	SPP2	N/A
19	YDC1	https://pubmed.ncbi.nlm.nih.gov/11694577	44	AIM36	N/A
20	RPR1	https://pubmed.ncbi.nlm.nih.gov/8083243/	45	APL2	N/A
21	FUS3	https://pubmed.ncbi.nlm.nih.gov/29273704/	46	MAM1	N/A
22	TIP1	https://pubmed.ncbi.nlm.nih.gov/12818212/	47	POA1	N/A
23	IML3	N/A	48	GOR1	N/A
24	SRC1	N/A	49	SNR42	N/A
25	MRPS5	N/A	50	LEO1	N/A

V.1.3 – Gene Ontology:

Among the set of 50 Differentially Expressed Genes (DEGs), 4 were classified as non-coding RNAs (ncRNA), 2 were small nucleolar RNAs (snoRNA), and 4 were coding genes with yet undiscovered functions. Notably, genes associated with gene expression mechanisms exhibited significant downregulation within the thermos pulsed population. This downregulation strongly suggests a disruption in cellular processes caused by the heightened temperature—a response intricately woven into the cell's adaptation strategy to thermal stress. Additionally, genes linked to both lipid metabolism and cellular repair pathways displayed downregulation. This indicates a strategic shift in the cell's allocation of resources and energy, redirecting focus away from normal functions towards immediate mechanisms for stress survival. Such adaptation not only fortifies the cell's capacity to endure challenging circumstances but also potentially expedites its recuperation once the stressor subsides.

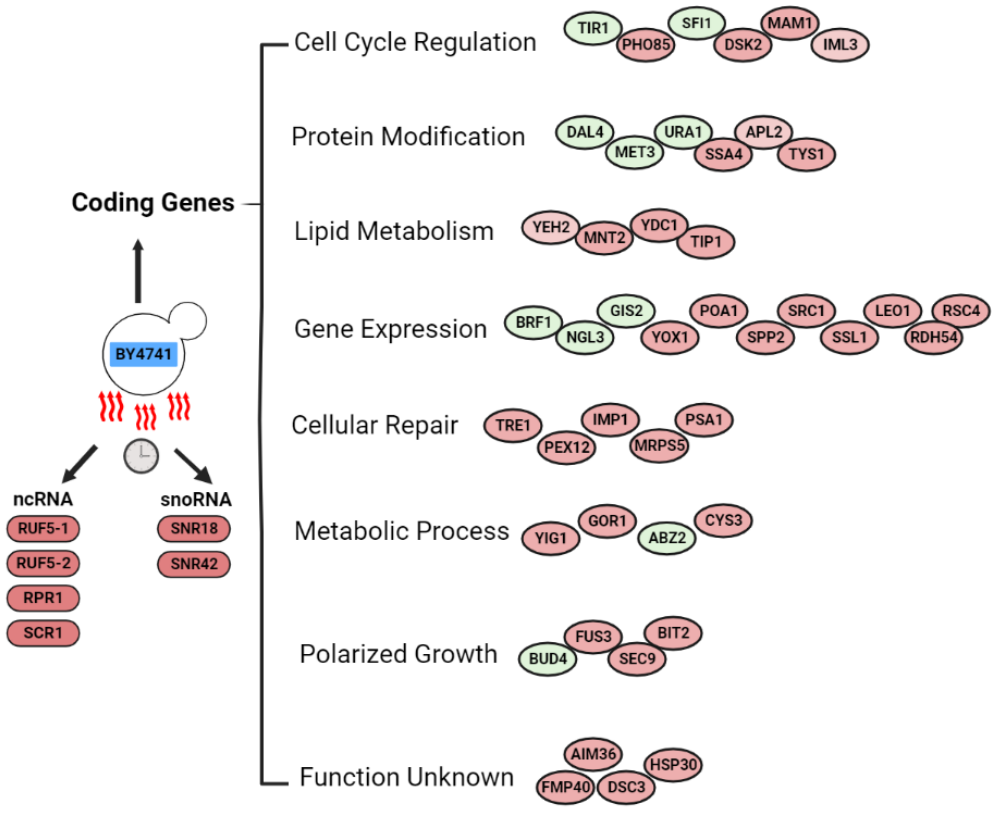


Figure 13: Distribution of molecular pathways of DEGs

V.2 – Objective 2: Deconvolution of Bulk RNA-seq samples using single-cell transcriptomes.

V.2.1 – Seurat Analysis

The fusion of single-cell RNA sequencing (scRNA-seq) data yielded a matrix comprising 6794 attributes and 11646 individual cells. This composite matrix was preserved within the Seurat object as the RNA assay. After subjecting this data to feature engineering, the matrix was subsequently transformed to a dimension of 2000x11646, wherein the top 2000 variable genes were selected. This modified matrix was then designated as the Integrated Assay. Subsequent analysis was conducted specifically on the RNA assay data. Utilizing a Jackstraw plot and generating a heatmap based on the initial 15 principal components (PCs), it became evident that 8 principal components held significant relevance.

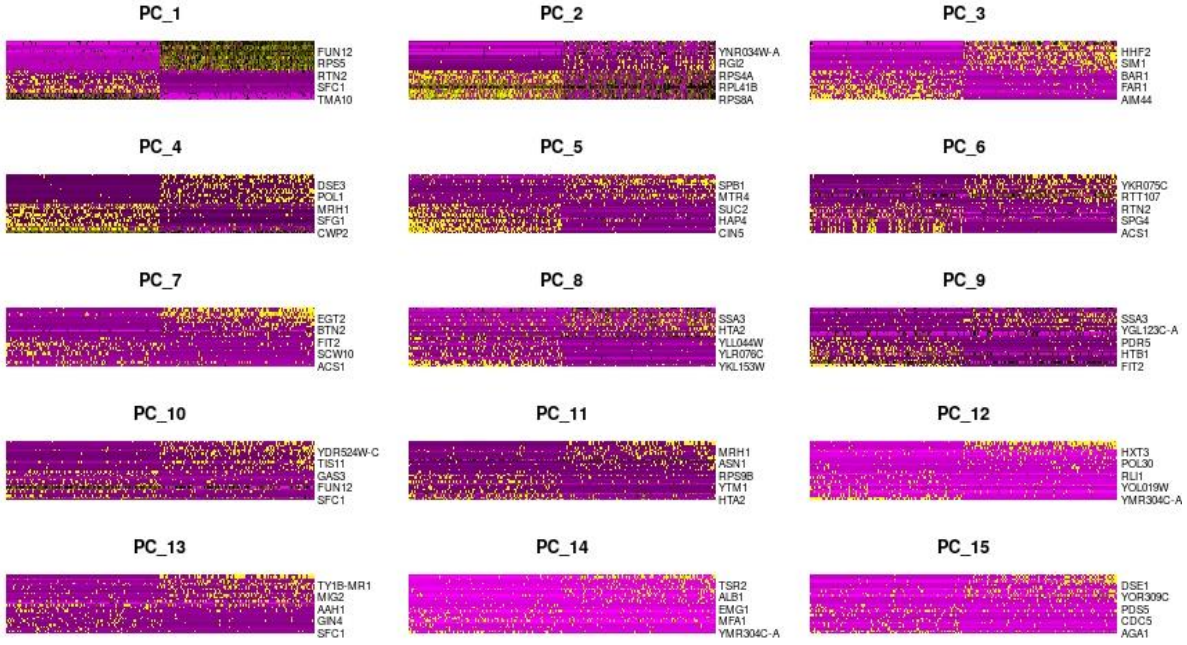


Figure 14: Heatmap for first 15 principal components

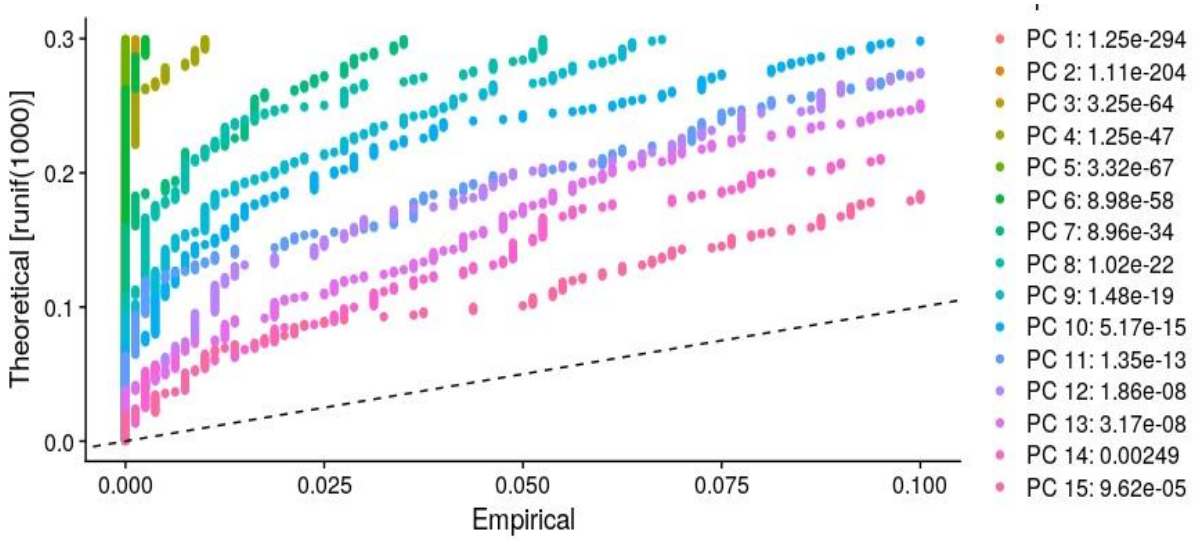


Figure 15: Jackstraw plot for first 15 PCs

The total count of principal components (PCs) was subsequently validated as 7 through examination of an Elbow plot. Silhouette analysis further pinpointed an optimal division into 13 clusters, with an impressive silhouette score of 0.75. The implementation of the FindClusters() function, employing the Louvain algorithm with a resolution of 1.1, facilitated the attainment of the desired optimal clusters. To enhance visualization in two-dimensional space, non-linear dimension reduction techniques, namely t-SNE and UMAP, were harnessed to depict these clusters.

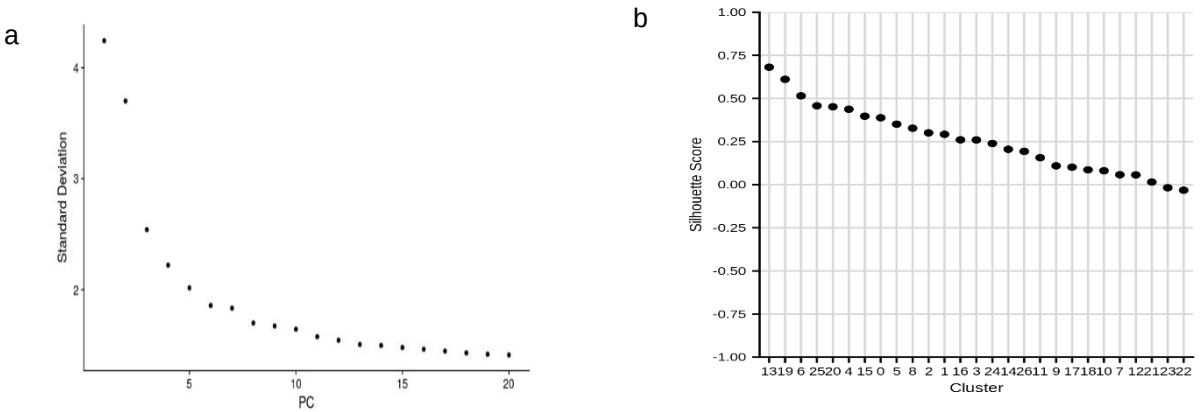


Figure 16: (a) Elbow Plot, (b) Optimum Cluster with Silhouette score

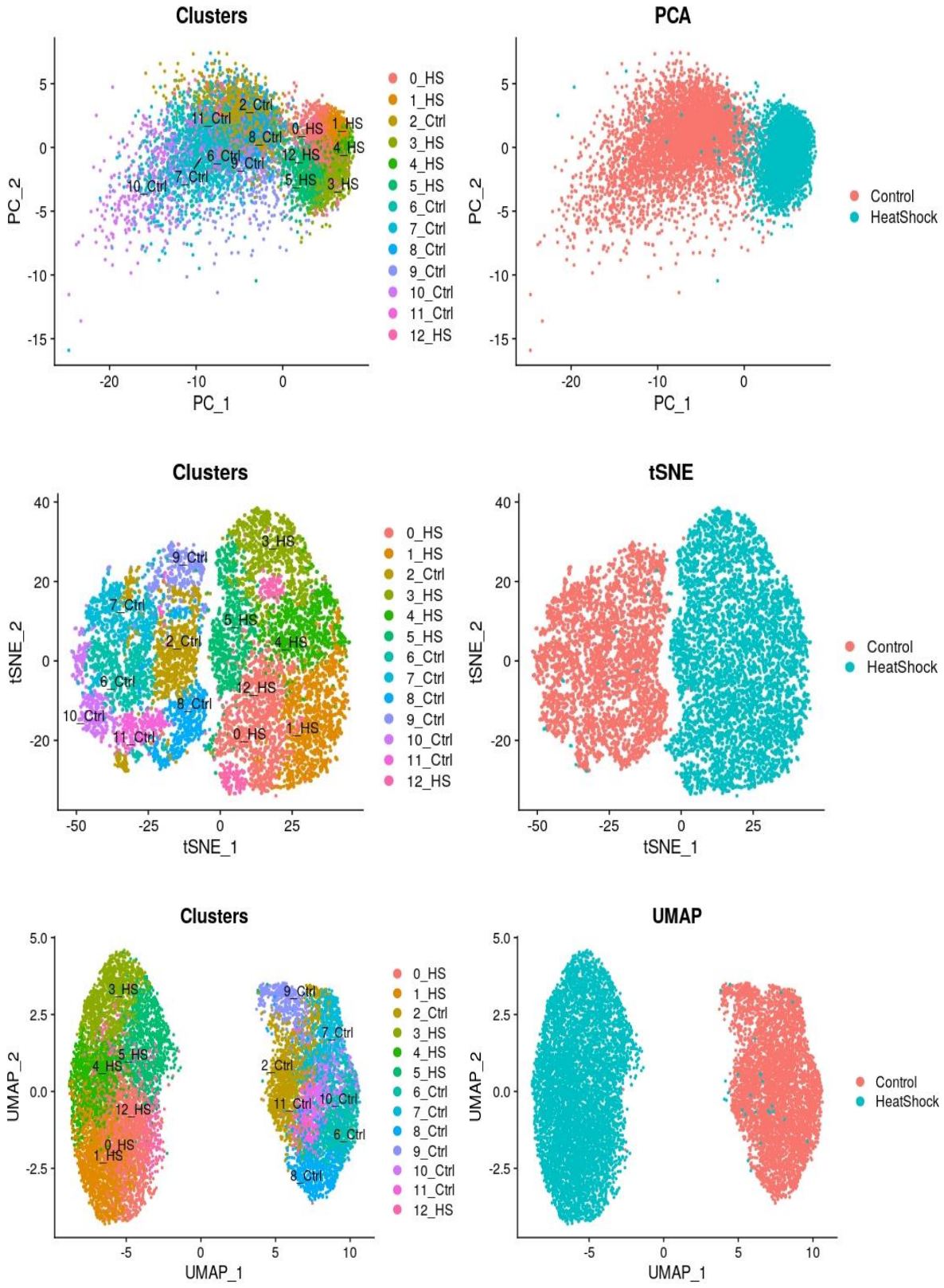


Figure 17: Linear (PCA) and non-linear (tSNE, UMAP) 2-D cluster visualization plots

To elucidate the behavioural dynamics of differentially expressed genes derived from the PULSER experiment within a cellular population subjected to non-pulsed heat shock at 42°C, a comprehensive heatmap visualization was conducted. Notably, this visualization revealed a distinctive pattern wherein specific genes that manifest considerable downregulation in the PULSER experiment's test samples exhibit a marked upregulation in the unpulsed cell population, and conversely, genes that are upregulated in the PULSER experiment appear downregulated in the unpulsed population.

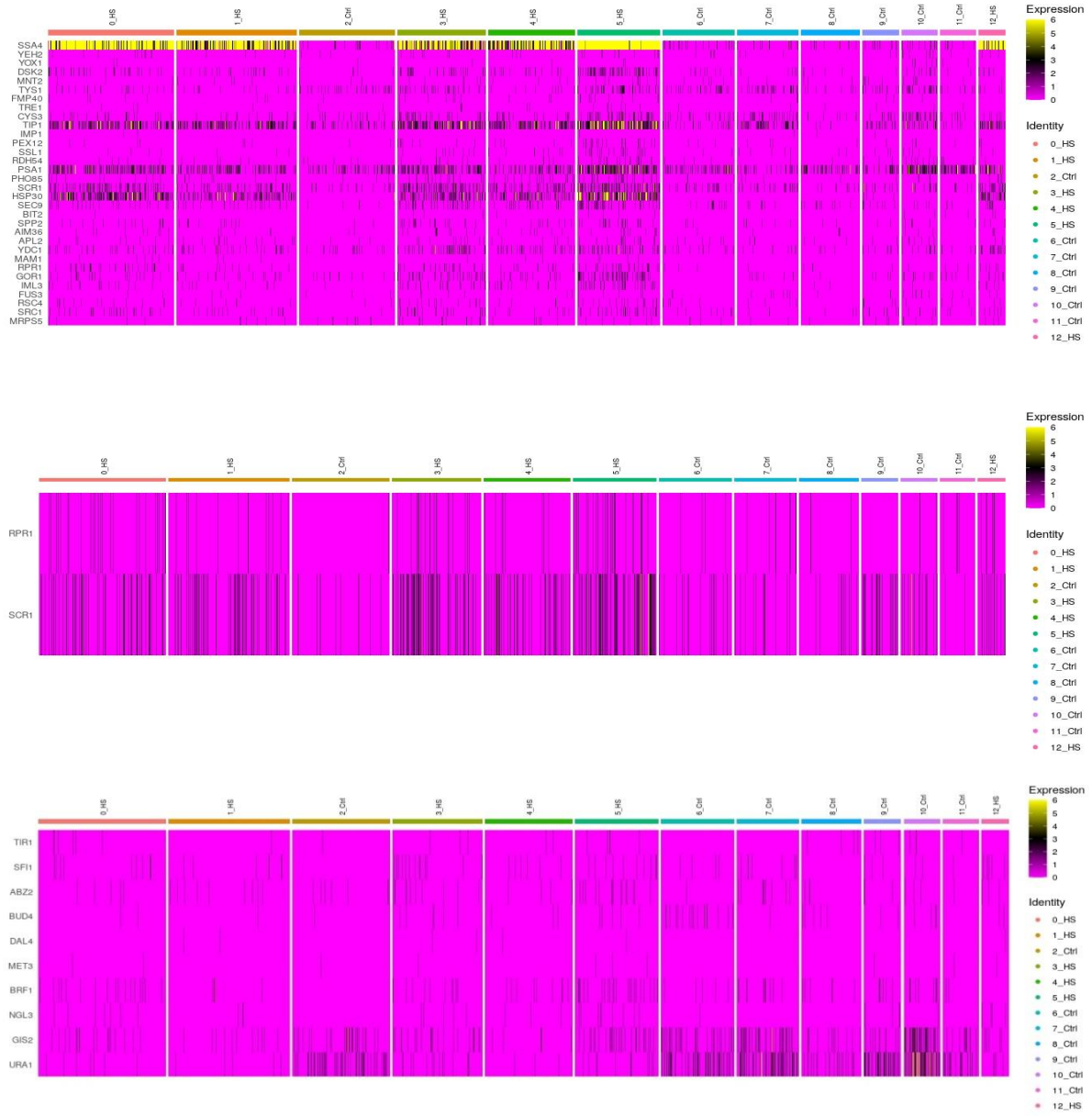


Figure 18: Heatmaps for DEGs (a) Downregulated and (c) Upregulated in PULSER. (b) represents the non-coding genes among the PULSER DEGs, expressed in sc RNA-seq data

Figure 19: Top 10 most variable genes in sc RNA-seq data of heat-shock in *S. cerevisiae*

A prominent instance exemplifying this trend is the gene encoding the HSP70 chaperone SSA4. This gene is widely acknowledged for its heightened induction in response to heat stress, as substantiated both by the heatmap data and the existing scientific literature. Paradoxically, within the context of the PULSER experiment, SSA4 emerges as the most profoundly downregulated gene. This intriguing observation prompts a noteworthy hypothesis: within the realm of recurring heat stress, the yeast species S. cerevisiae appears to strategically fortify its metabolic landscape, resembling that of an unstressed environment. Evidently, this cellular response hints at the retention of a memory mechanism for recurrent stress occurrences.

A visual depiction illustrating the contrast in expression levels of SSA4 between the PULSER dataset and scRNA-seq data substantiated the inversely correlated pattern of SSA4 expression during typical stress conditions and repetitive stress events.

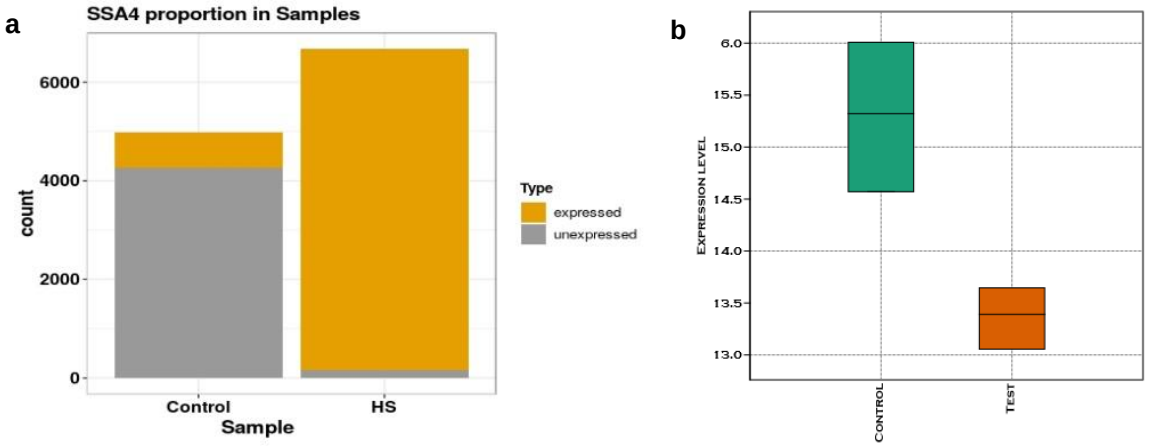
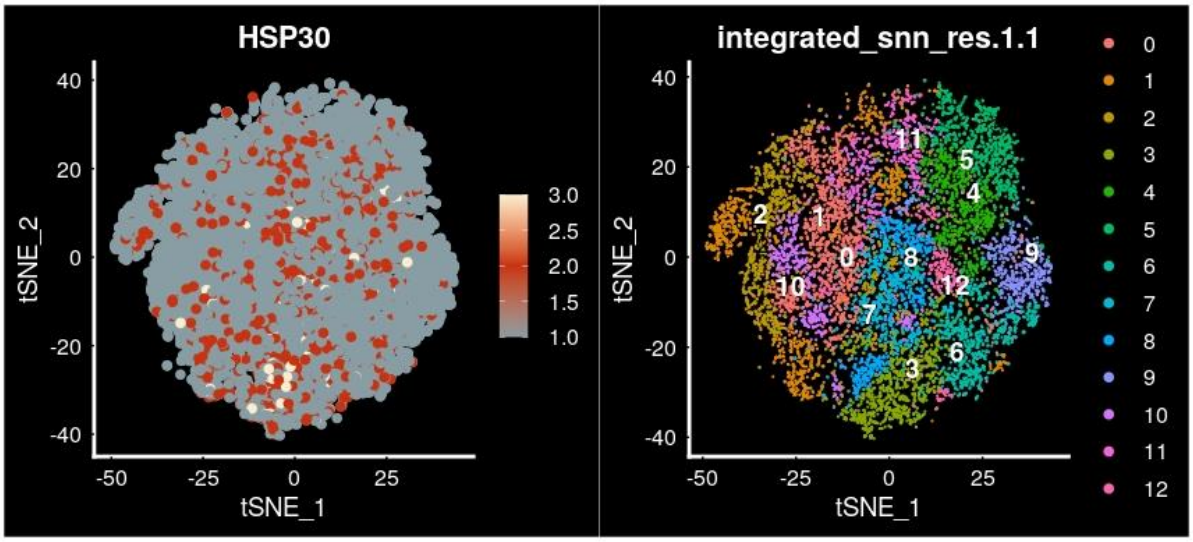


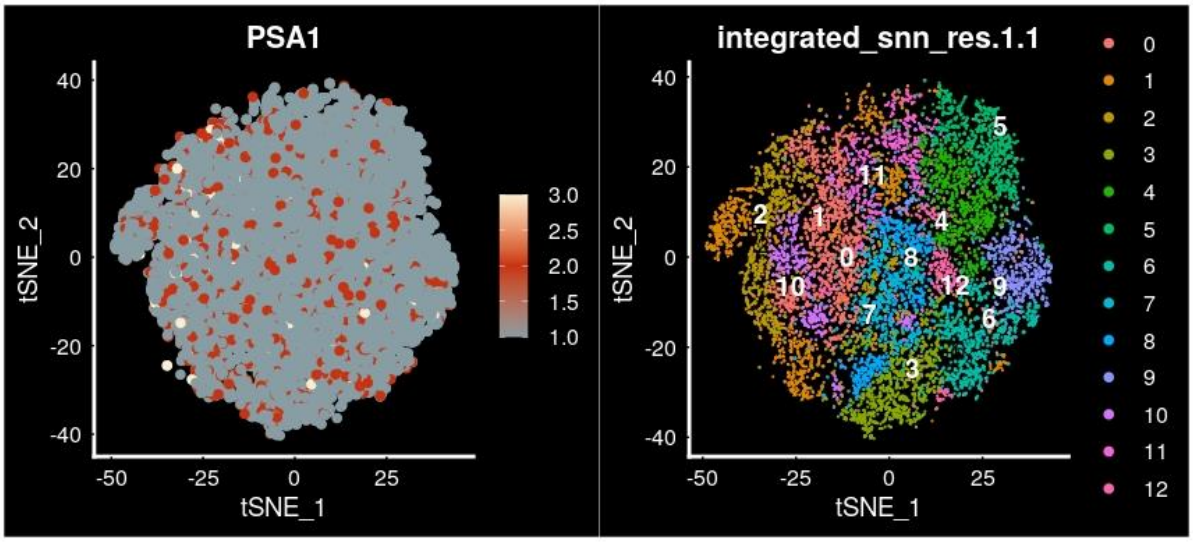
Figure 20: Comparison between expression levels of SSA4 in (a) Normal, (b) Recursive heat stress

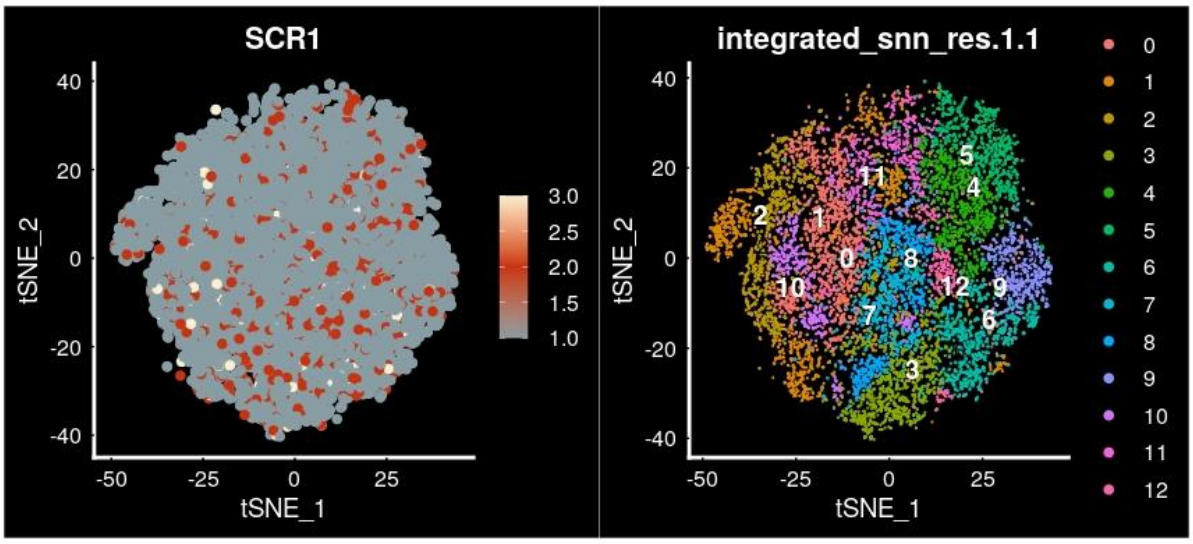
Utilizing Feature Plots, the differentially expressed genes (DEGs) from the PULSER experiment were graphically represented alongside scRNA-seq data to facilitate the visualization of gene-specific expression patterns within distinct clusters. Remarkably, among the genes significantly upregulated during repetitive heat stress, a striking observation emerged: nine out of ten of these genes exhibited marked downregulation in the Test samples under typical stress conditions. This intriguing phenomenon potentially implies that cells cultured under optimal growth circumstances retain an evolutionary memory of stress across multiple generations. Consequently, when subjected to successive rounds of stress, these cells surpass the expression levels of genes associated with normal growth conditions, thereby orchestrating a concerted effort to sustain a uniform metabolomic state.

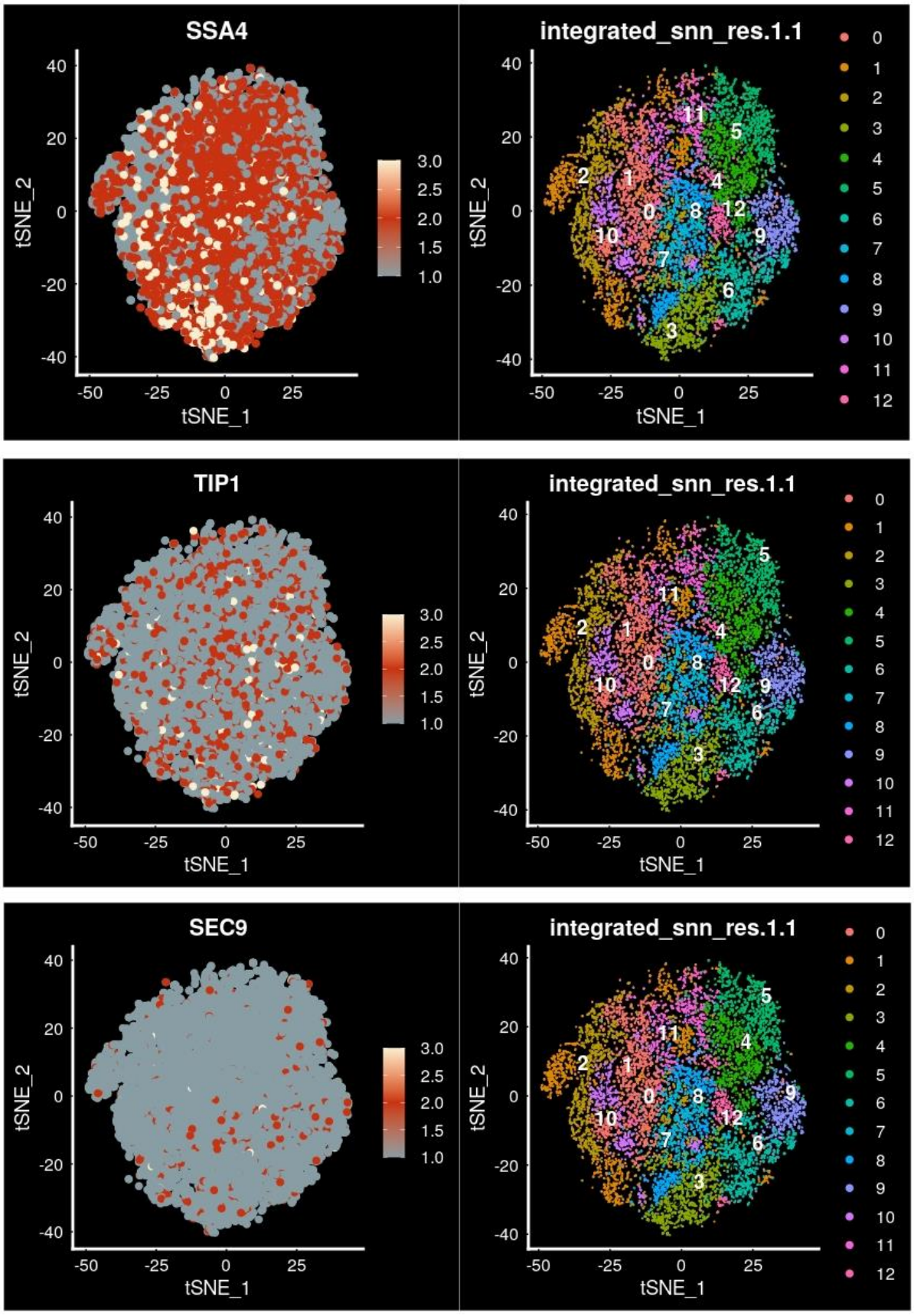
Some of the DEGs from PULSER experiment having profound expression in sc RNA-seq data are shown below. For visualizing all fifty DEGs, please refer to Appendix-I, publicly available at : APPENDIX - I

Similarly, Violin Plots of these significantly expressed DEGs are pasted below. In order to visualize the distribution of expression of all fifty PULSER DEGs, kindly refer to Appendix-II, publicly available at : APPENDIX-II


HSP30
integrated_snn_res.1.1
tSNE_1
tSNE_2
0
1
2
3
4
5
6
7
8
9
10
11
12


PSA1
integrated_snn_res.1.1
tSNE_1
tSNE_2
0
1
2
3
4
5
6
7
8
9
10
11
12


SCR1
integrated_snn_res.1.1
tSNE_1
tSNE_2
0
1
2
3
4
5
6
7
8
9
10
11
12


SSA4
integrated_snn_res.1.1
TIP1
integrated_snn_res.1.1
SEC9
integrated_snn_res.1.1
tSNE_1
tSNE_2
3.0
2.5
2.0
1.5
1.0
0
1
2
3
4
5
6
7
8
9
10
11
12

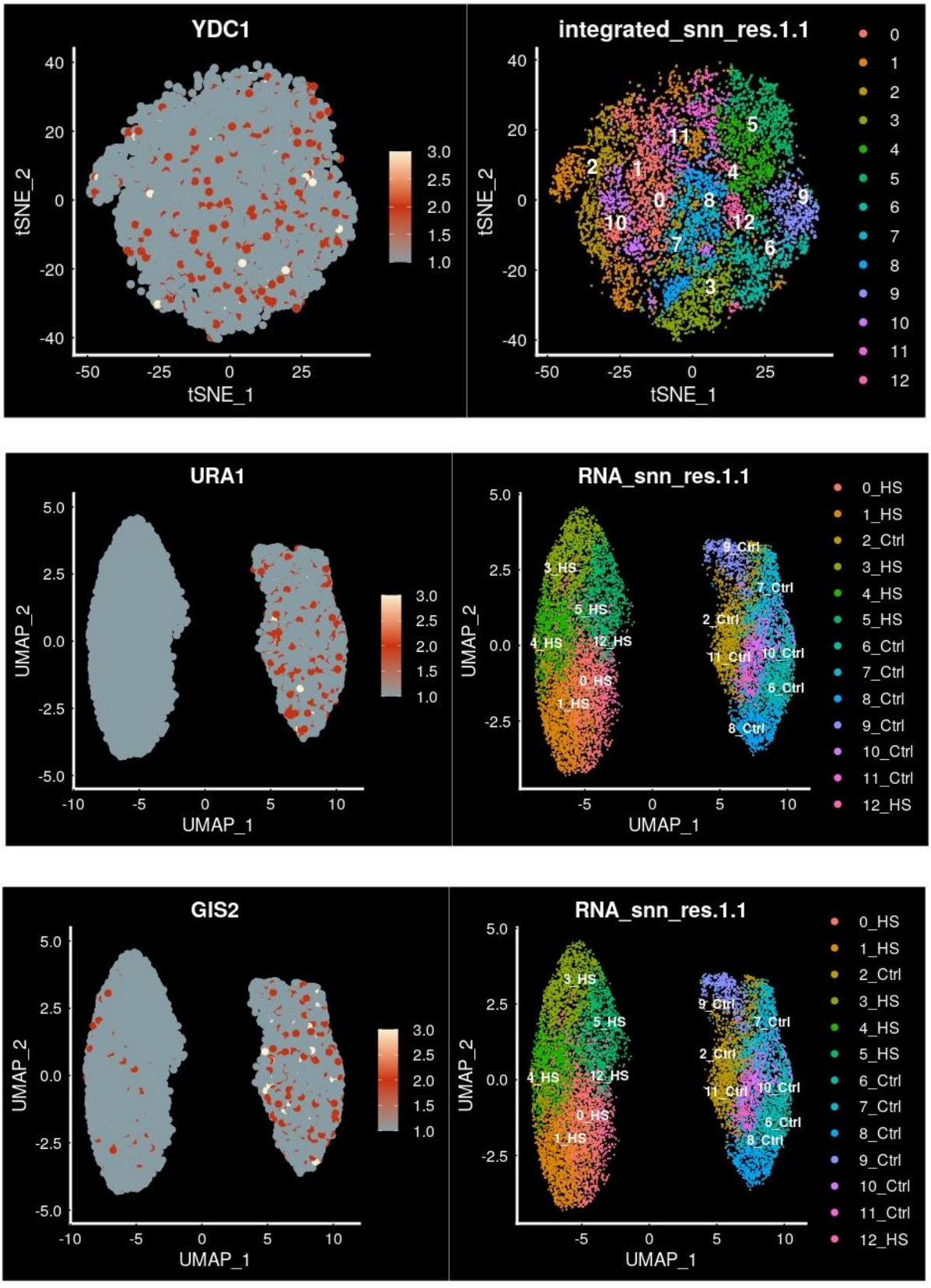
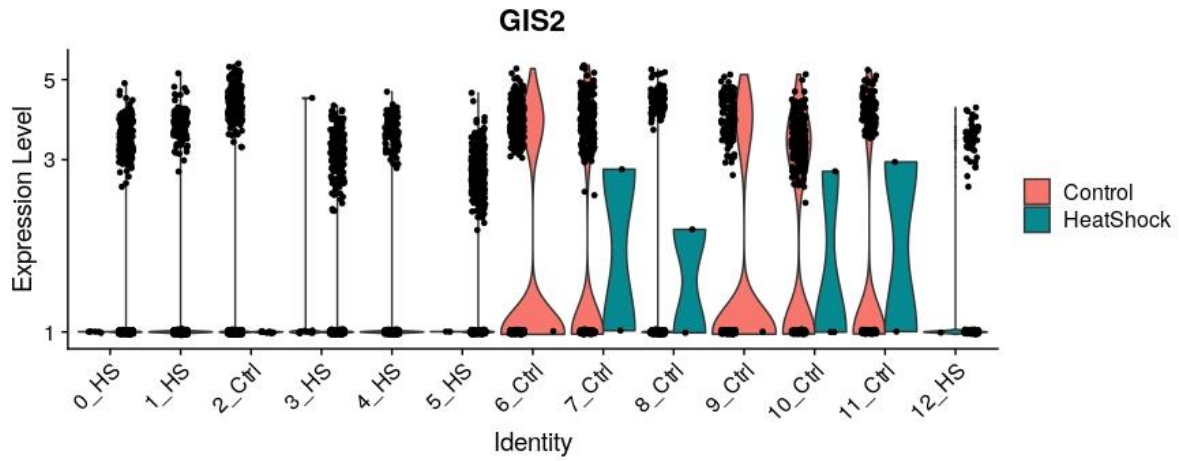
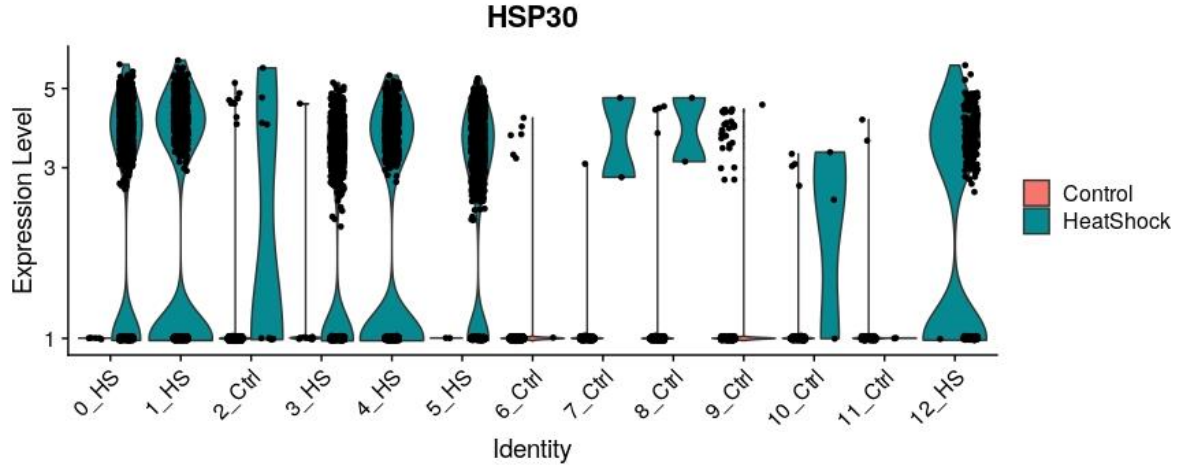
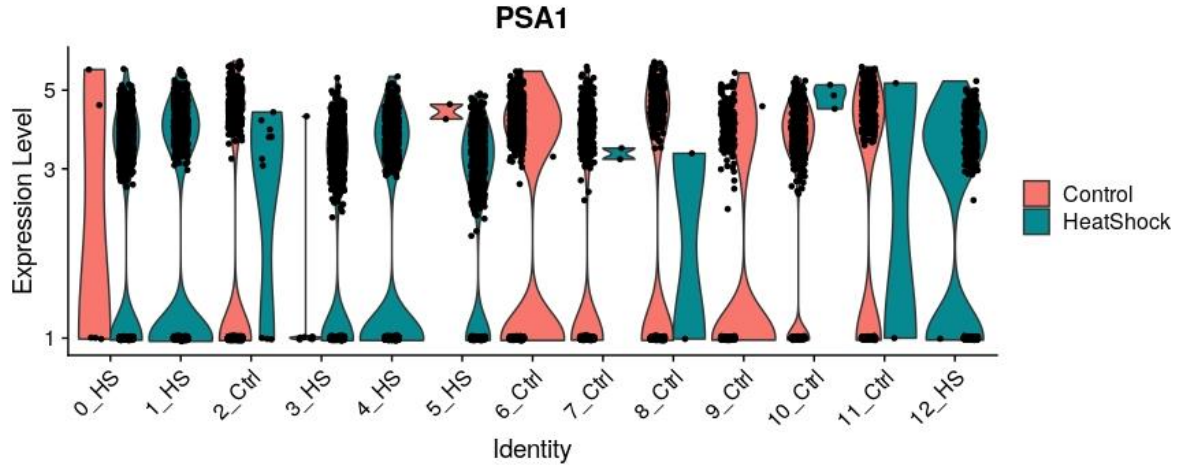
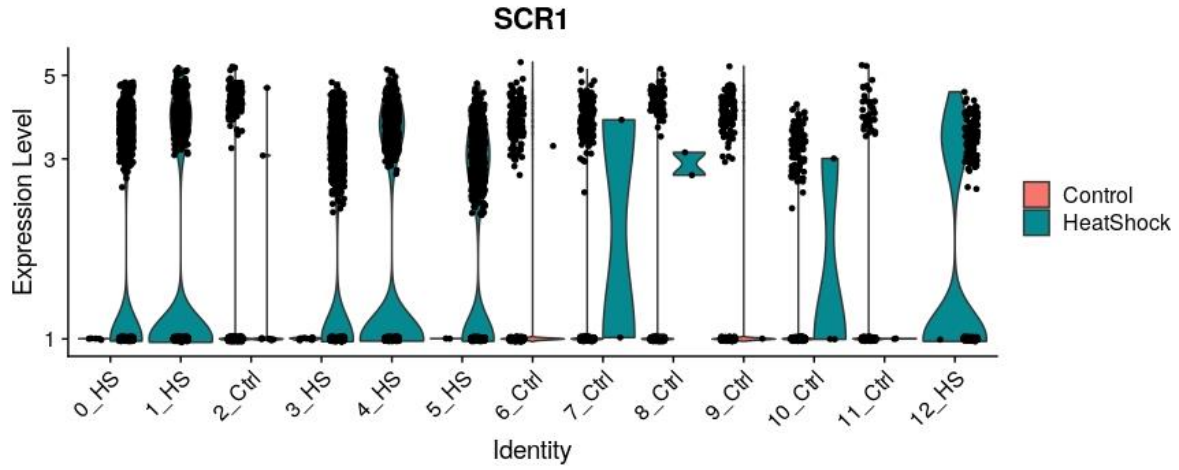


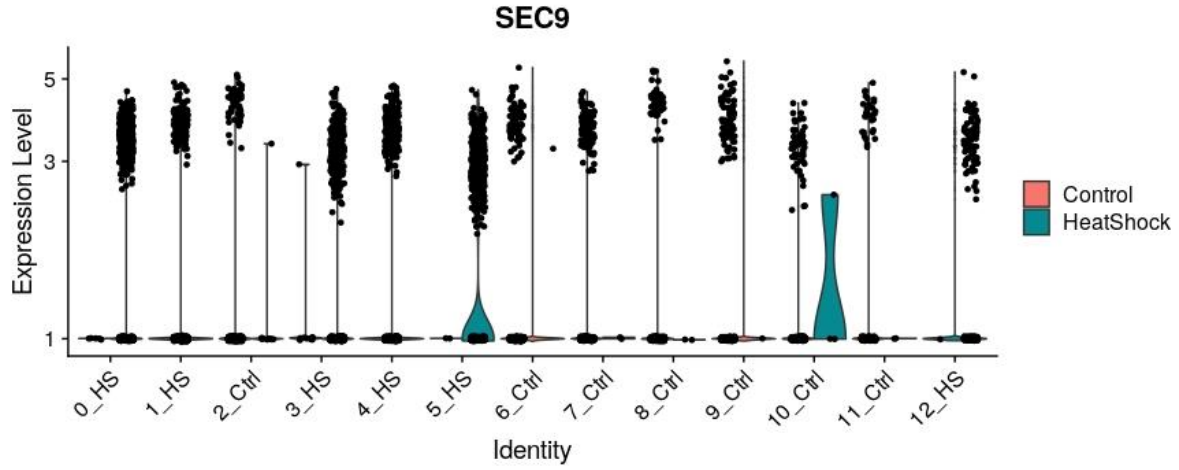
Figure 21: Feature Plots to visualize PULSER DEGs expression in non-recursive stress condition.

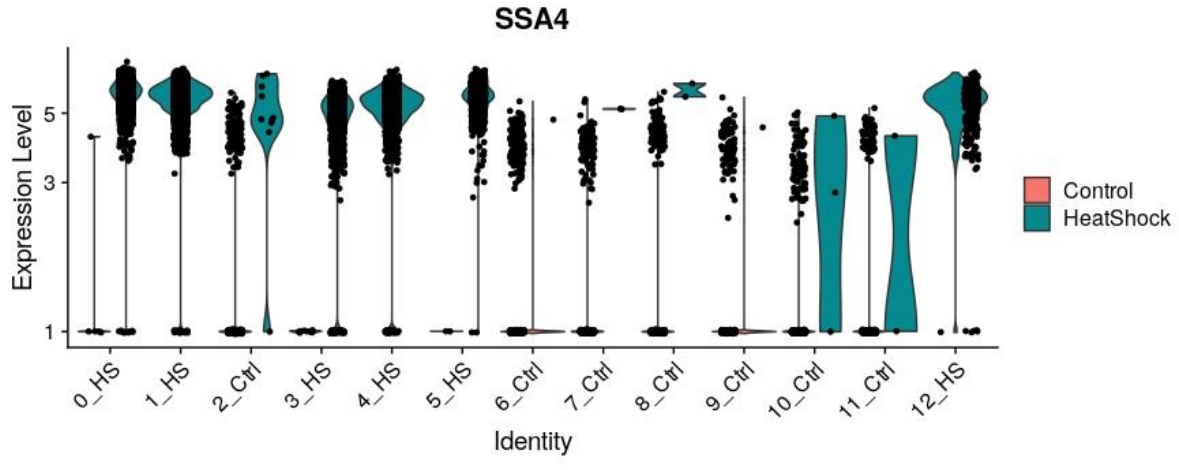

GIS2
Expression Level
5
3
1
0_HS
1_HS
2_Ctrl
3_HS
4_HS
5_HS
6_Ctrl
7_Ctrl
8_Ctrl
9_Ctrl
10_Ctrl
11_Ctrl
12_HS
Identity
Control
HeatShock


HSP30
Expression Level
5
3
1
0_HS
1_HS
2_Ctrl
3_HS
4_HS
5_HS
6_Ctrl
7_Ctrl
8_Ctrl
9_Ctrl
10_Ctrl
11_Ctrl
12_HS
Identity
Control
HeatShock


PSA1
Expression Level
5
3
1
0_HS
1_HS
2_Ctrl
3_HS
4_HS
5_HS
6_Ctrl
7_Ctrl
8_Ctrl
9_Ctrl
10_Ctrl
11_Ctrl
12_HS
Identity
Control
HeatShock


SCR1
Expression Level
5
3
1
0_HS
1_HS
2_Ctrl
3_HS
4_HS
5_HS
6_Ctrl
7_Ctrl
8_Ctrl
9_Ctrl
10_Ctrl
11_Ctrl
12_HS
Identity
Control
HeatShock


SEC9
Expression Level
5
3
1
0_HS
1_HS
2_Ctrl
3_HS
4_HS
5_HS
6_Ctrl
7_Ctrl
8_Ctrl
9_Ctrl
10_Ctrl
11_Ctrl
12_HS
Identity
Control
HeatShock


SSA4
Expression Level
5
3
1
0_HS
1_HS
2_Ctrl
3_HS
4_HS
5_HS
6_Ctrl
7_Ctrl
8_Ctrl
9_Ctrl
10_Ctrl
11_Ctrl
12_HS
Identity
Control
HeatShock

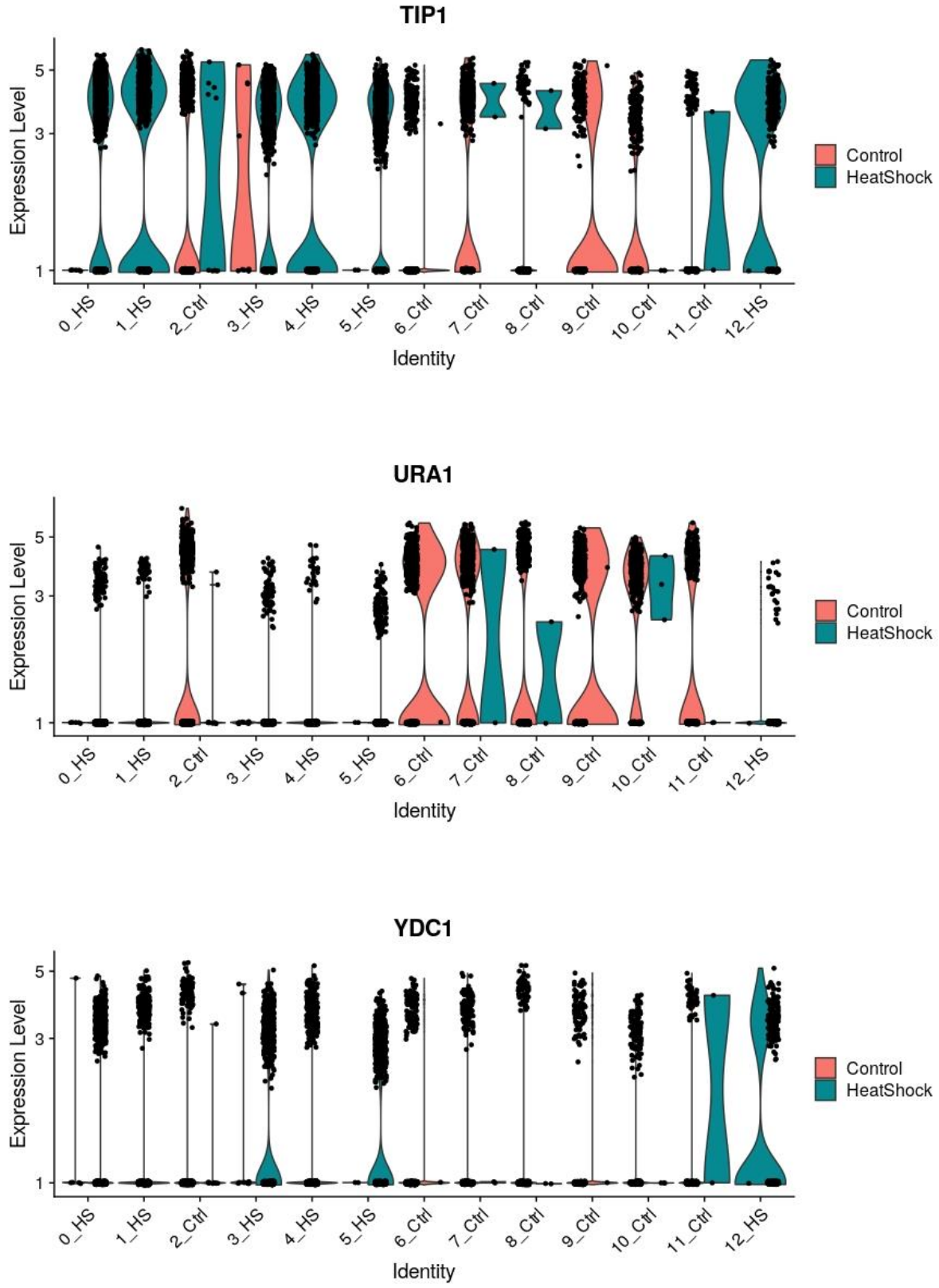


Figure 22: Violin plots to visualize expression of DEGs under non-recursive heat stress.

The plots depicting correlation, covariance, and Euclidean distance didn't reveal any resemblance between the PULSER sample and scRNA-seq data clusters. This disparity could potentially be attributed to the presence of batch effects. Nevertheless, it's noteworthy that the PULSER samples exhibited a strong degree of similarity among each other in terms of the similarity index.

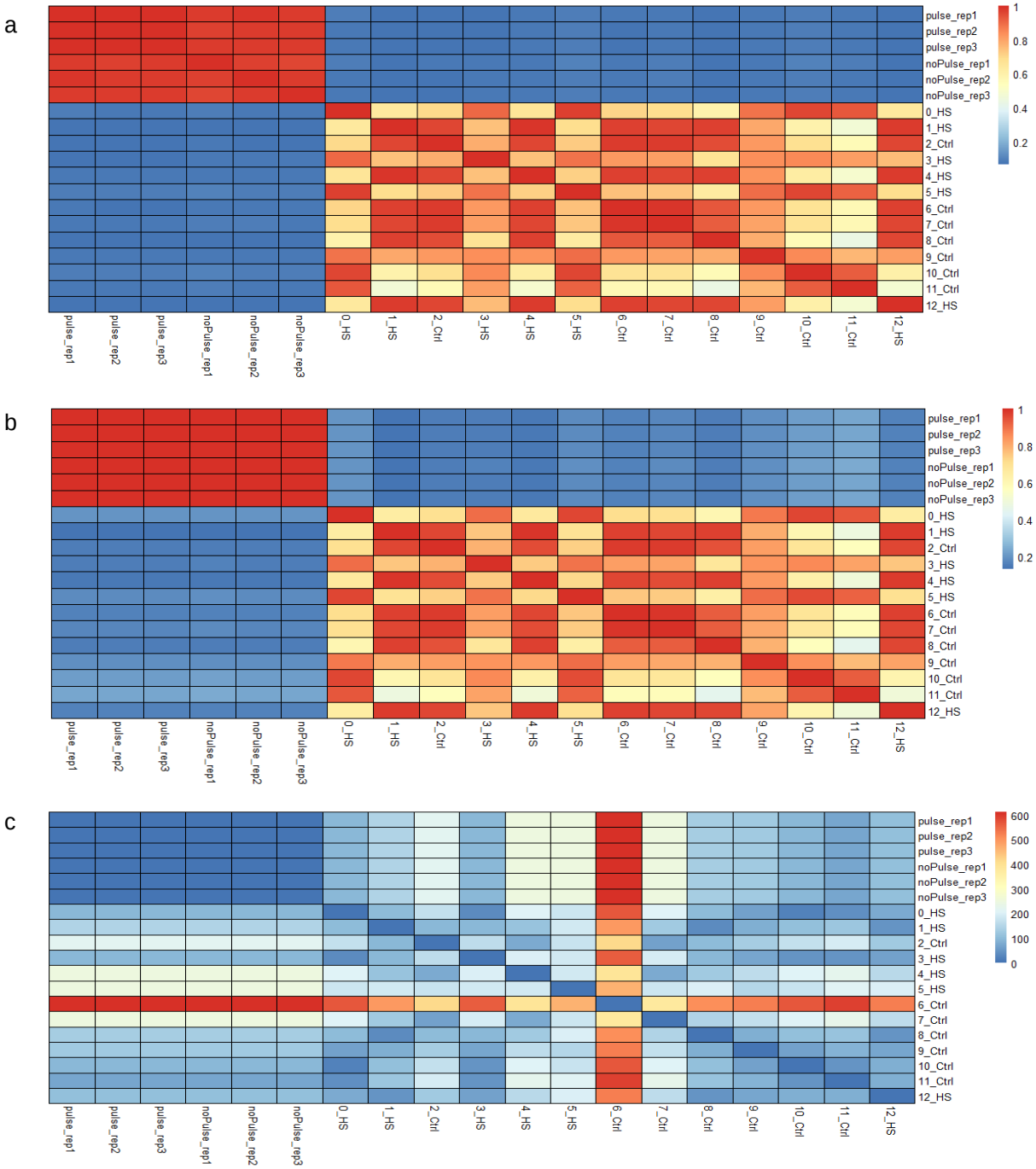


Figure 23: Similarity among clusters and PULSER samples visualized through (a) Correlation, (b) Covariance and (c) Euclidean distance.

V.2.2 – MuSiC Deconvolution:

Hierarchical clustering of clusters obtained from normally heat stressed cells resulted in further clustering them in four groups.

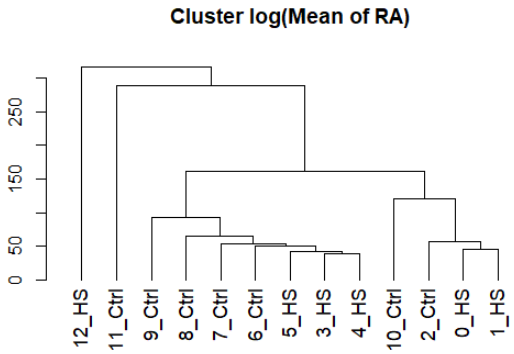


Figure 24: Dendrogram of hierarchical clustering of sc RNA-seq clusters

Deconvolving bulk samples from PULSER experiments revealed that the '2_Ctrl' cell type dominated the composition (~0.6), followed by '8_Ctrl' and '10_Ctrl' types. Strikingly, the subclusters within the 'Control' cell population, closely mirrored the high proportions observed in PULSER samples. Notably, the '8_Ctrl' cell type exhibited a more pronounced presence in cell populations exposed to iterative stress, surpassing even the baseline control population. This finding reinforces the proposition of an evolutionary memory mechanism in yeast, operative within a subset of *S. cerevisiae* populations. This mechanism assists yeast in recognizing and combatting recurrent stress instances.

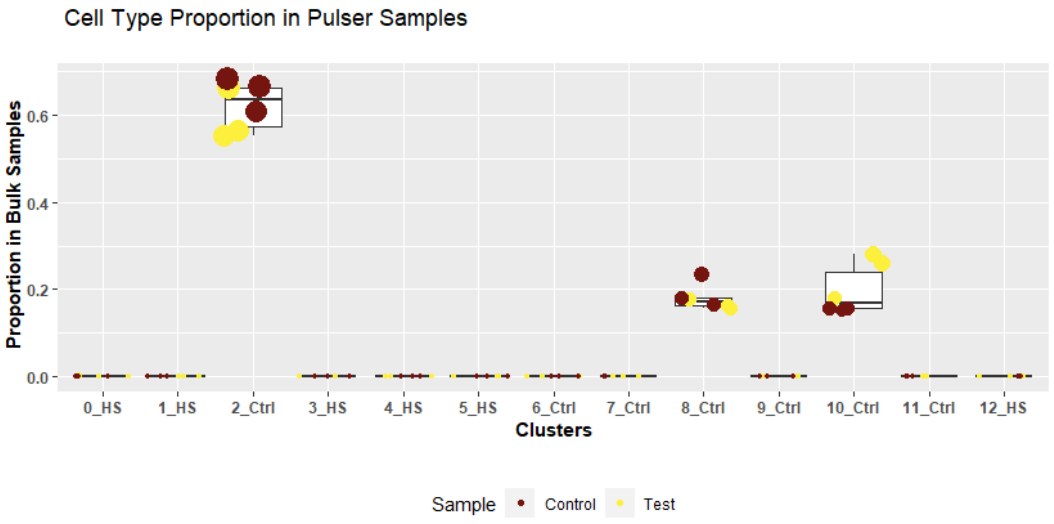


Figure 25: Cell Type proportions of sc RNA-seq data in PULSER samples

V.3 – Objective 3: Functional characterization to elucidate the impact of thermo-pulsing on yeast.

The growth profiles of S. cerevisiae BY4741 indicated a pronounced exponential phase in cells subjected to continuous thermal stress, as opposed to both thermo-pulsed conditions and cells cultivated under optimal growth circumstances. Interestingly, prior studies[42] have suggested a potential inverse relationship between stress survival and growth rate.

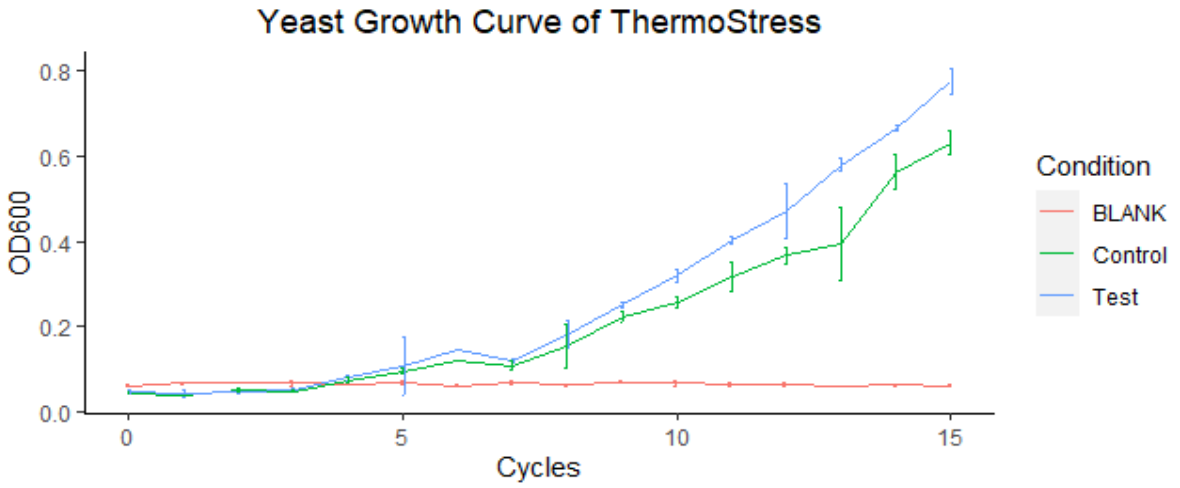


Figure 26: Yeast Growth curve with Pulsed thermal stress of in Test sample

A comparison between the growth curves of pulsed and unpulsed thermal stress indicate a clear acceleration of growth in cells undergoing continuous thermal stress as opposed to recursive thermal stress.

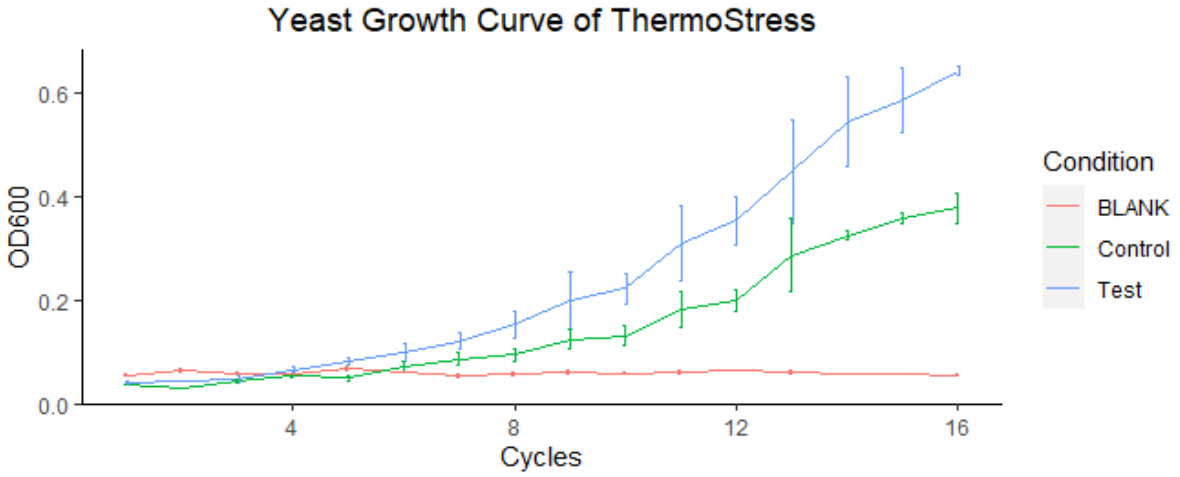


Figure 27: Yeast Growth curve with continuous thermal stress of in Test sample

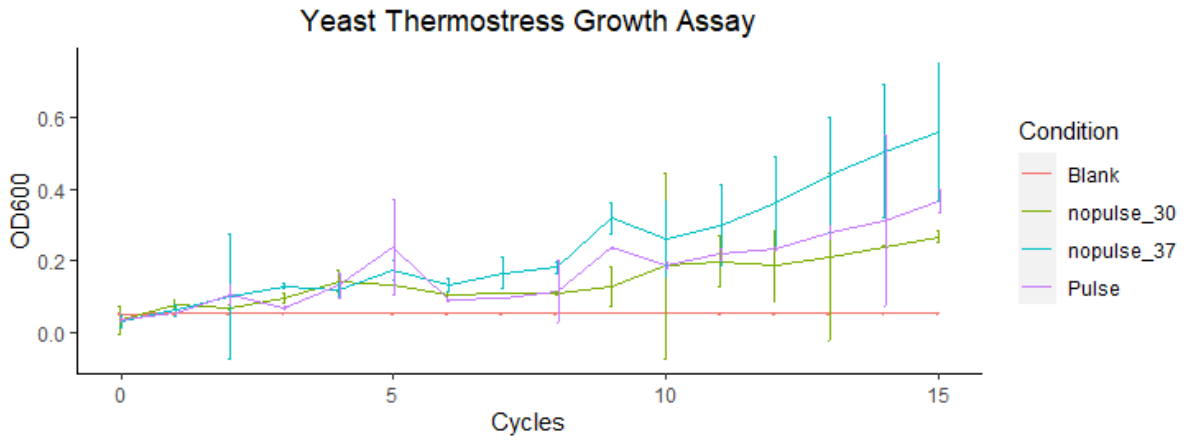


Figure 28: Integrated graph with growth curves for Pulsed, non-pulsed and optimum thermal environment.

Follow-up studies conducted to assess the ability of these cultures to withstand varying concentrations of Acetic Acid (an apoptotic agent) produced a nuanced result concerning the comparative stress resistance capabilities of the non-thermal-pulsed and thermal-pulsed populations.

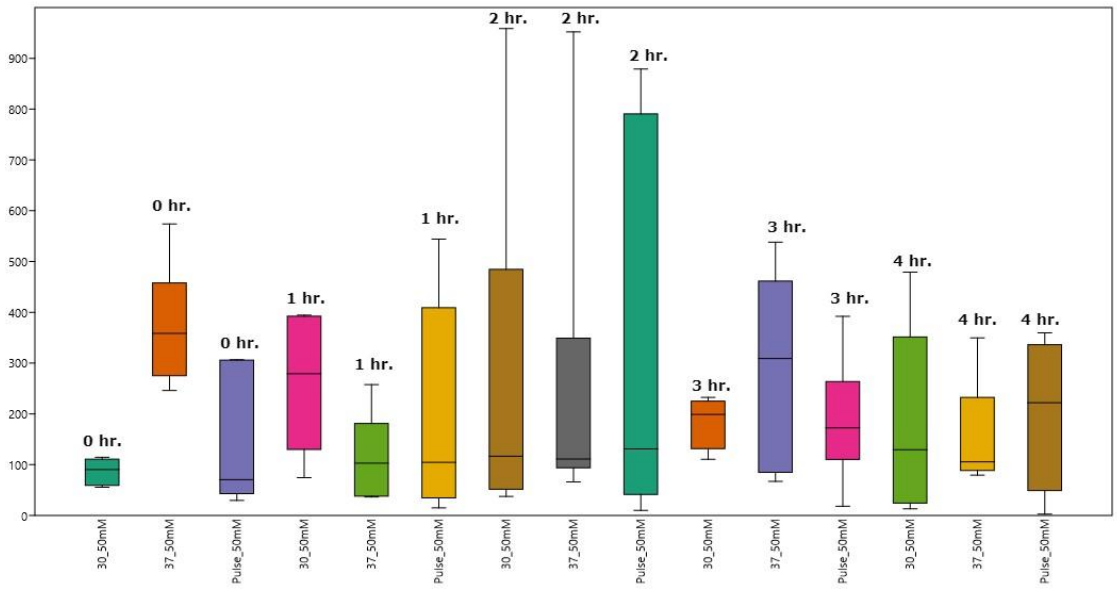


Figure 29: *S. cerevisiae* population of Group A (50 mM Acetic Acid treatment) with FDA fluorescence (485 nm) measured at interval of 1 hour for a period of 4 hours.

One hour after exposure to a 50mM acetic acid treatment, cells subjected to repeated thermal stress displayed the greatest number of viable cells. This pattern persisted in the following hour, indicating the resilience of the intermittent population to the

apoptotic agent. After three hours, cells cultured under constant thermal stress at 37°C exhibited the highest count of viable cells, followed by the intermittent stress group. Nevertheless, after four hours, cells grown at the optimal temperature and those subjected to intermittent stress demonstrated an equal number of viable cells, with the continuously thermal stressed population trailing behind.

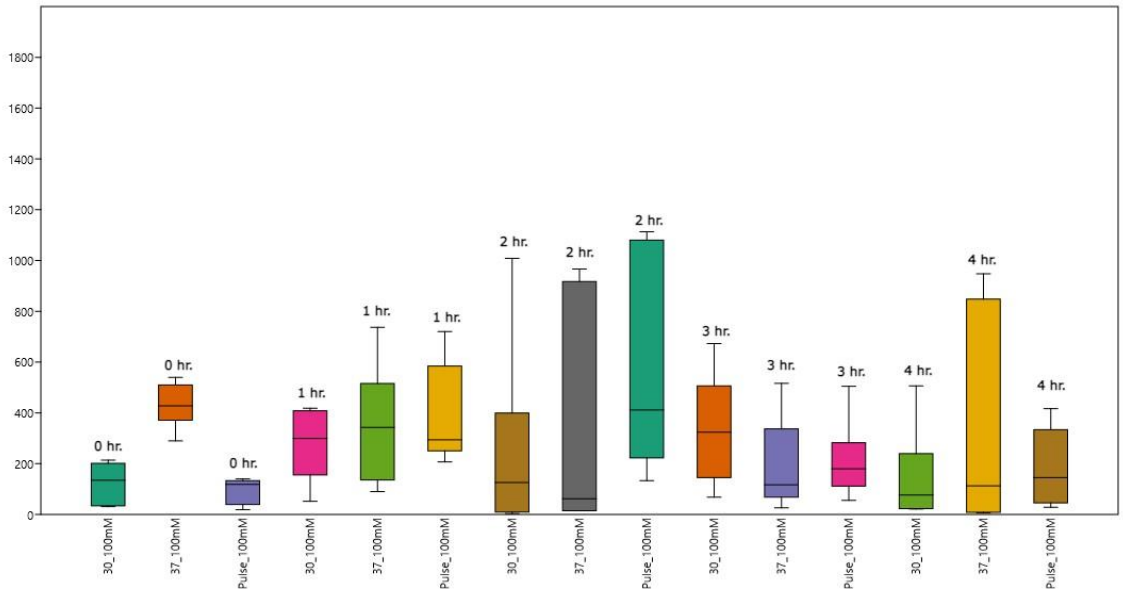


Figure 30: *S. cerevisiae* population of Group B (100 mM Acetic Acid treatment) with FDA fluorescence (485 nm) measured at interval of 1 hour for a period of 4 hours.

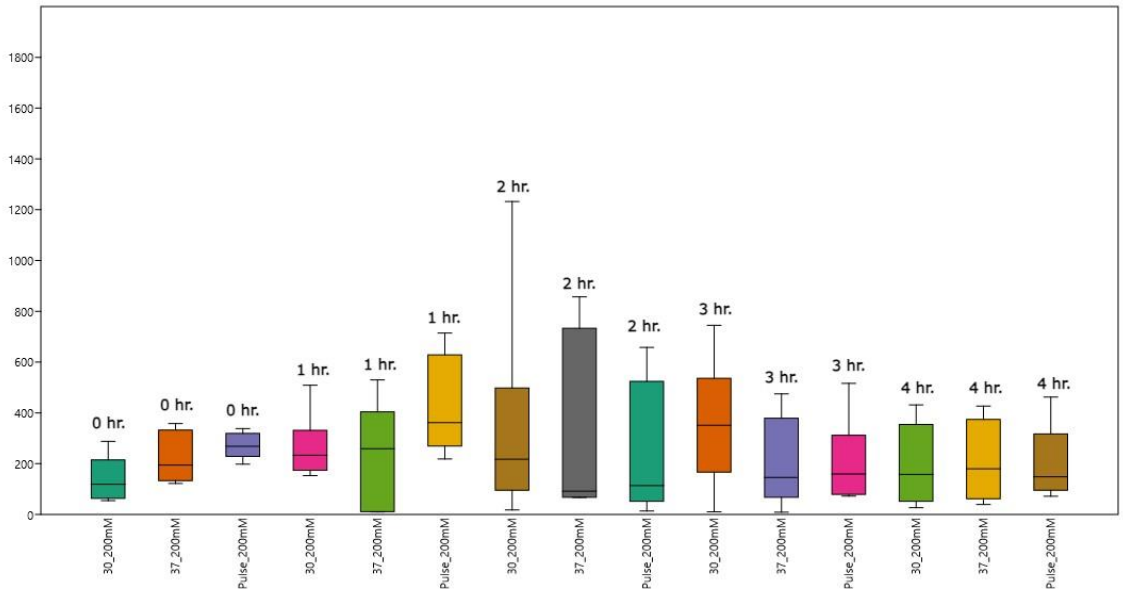


Figure 31: *S. cerevisiae* population of Group C (200 mM Acetic Acid treatment) with FDA fluorescence (485 nm) measured at interval of 1 hour for a period of 4 hours.

The pattern witnessed in the live population subjected to repeated thermal stress persisted in Group B (100 mM Acetic Acid). However, in the case of Group C (200 mM Acetic Acid), the pulsed population displayed the highest number of cells among all the samples after one hour of acetic acid treatment.

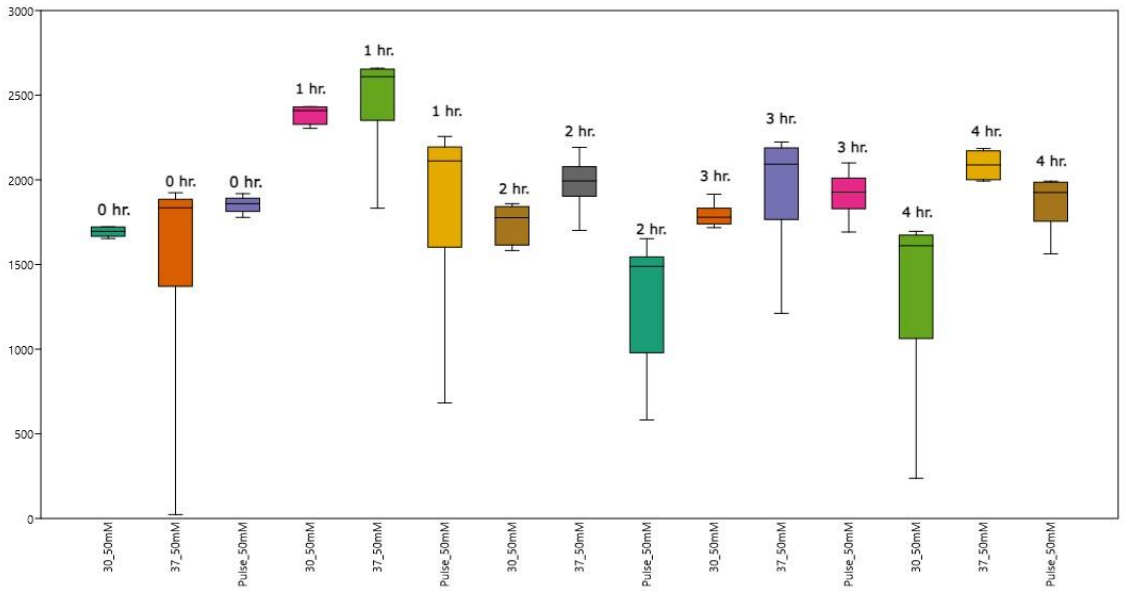


Figure 32: *S. cerevisiae* population of Group A (50 mM Acetic Acid treatment) with PI fluorescence (544 nm) measured at interval of 1 hour for a period of 4 hours.

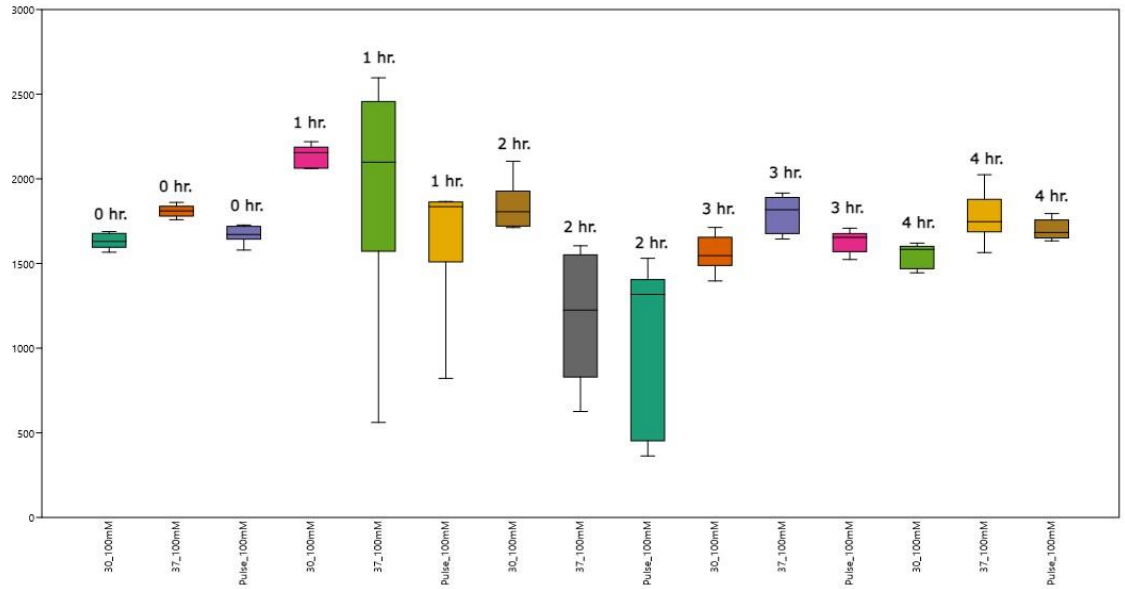


Figure 33: *S. cerevisiae* population of Group B (100 mM Acetic Acid treatment) with PI fluorescence (544 nm) measured at interval of 1 hour for a period of 4 hours.

For both Group A and Group B, one hour after being treated with acetic acid, cells that experienced repeated thermal stress exhibited the highest count of non - viable cells. This trend remained consistent in the subsequent hour, highlighting the endurance of the intermittent population against the apoptotic agent. After three hours, cells cultured at the optimal temperature displayed the greatest number of viable cells, with the intermittent stress group following suit. This identical pattern persisted into the fourth hour as well.

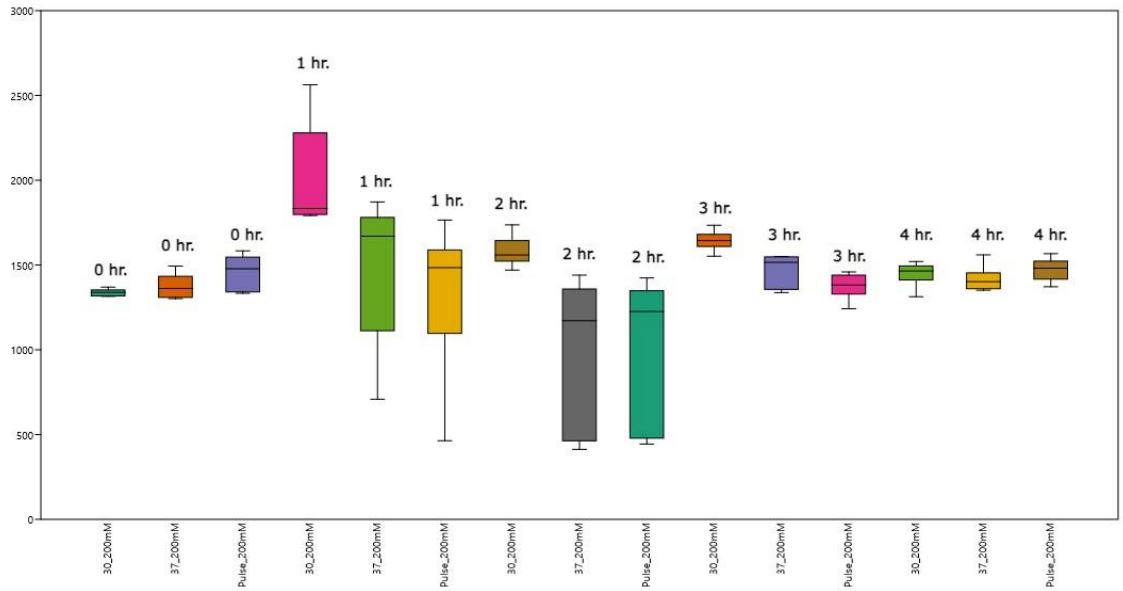


Figure 34: *S. cerevisiae* population of Group C (200 mM Acetic Acid treatment) with PI fluorescence (544 nm) measured at interval of 1 hour for a period of 4 hours.

At 200mM concentration of acetic acid, intermittent population outperformed all samples in terms of robustness to apoptotic stress till three hours of exposure.

Conclusion

VI - Conclusion:

Applying recursive thermal stress to *S. cerevisiae* cells conferred them with increased resilience against apoptotic stress within a defined timeframe, in comparison to cells cultivated under optimal conditions or subjected to non-recursive heat stress. Additionally, it was observed that even within a population of cells grown at an ideal temperature, a subset of cells consistently exhibited characteristics indicative of heat-shock response. This distinct subset of cells exhibited differential gene expression profiles that shared similarities with the PULSER population.

Particularly intriguing was the significant downregulation observed in the expression of heat shock response (HSR) proteins, including SSA4, within the PULSER population. This finding raises questions about how this downregulation might contribute to the enhanced robustness of these cells during instances of apoptotic stress. Through the utilization of single-cell RNA sequencing (scRNA-seq) data obtained from an experiment involving a 42°C heat shock applied to the BY4741 strain, the deconvolution of bulk samples revealed the proportions of different cell types present within the intermittent population that underwent recursive heat stress.

These discoveries shed light on what could be termed as an "evolutionary memory" of stress within *S. cerevisiae* cells. This memory appears to have developed as an adaptive response to the dynamic and frequently changing laboratory environmental conditions experienced by yeast cell lines, including processes like thawing. The technique of employing recursive stress induction in yeast cells, resulting in minimized protein denaturation, holds significant promise for addressing stress-related disorders such as neurodegenerative conditions like Parkinson's.

REFERENCE

VII - References:

1. Mohammadi S, Saberidokht B, Subramaniam S, Grama A. Scope and limitations of yeast as a model organism for studying human tissue-specific pathways. BMC Syst Biol. 2015 Dec 29;9:96. doi: 10.1186/s12918-015-0253-0. PMID: 26714768; PMCID: PMC4696342..
2. Carmona-Gutierrez D, Eisenberg T, Büttner S, Meisinger C, Kroemer G, Madeo F. Apoptosis in yeast: triggers, pathways, subroutines. Cell Death Differ. 2010 May;17(5):763-73. doi: 10.1038/cdd.2009.219. Epub 2010 Jan 15. PMID: 20075938.
3. Pereira C, Bessa C, Soares J, Leão M, Saraiva L. Contribution of yeast models to neurodegeneration research. J Biomed Biotechnol. 2012;2012:941232. doi: 10.1155/2012/941232. Epub 2012 Jul 15. PMID: 22910375; PMCID: PMC3403639.
4. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell. 2000 Dec;11(12):4241-57. doi: 10.1091/mbc.11.12.4241. PMID: 11102521; PMCID: PMC15070.
5. Kataoka T, Powers S, Cameron S, Fasano O, Goldfarb M, Broach J, Wigler M. Functional homology of mammalian and yeast RAS genes. Cell. 1985 Jan;40(1):19-26. doi: 10.1016/0092-8674(85)90304-6. PMID: 2981628.
6. Lindquist S, Craig EA. The heat-shock proteins. Annu Rev Genet. 1988;22:631-77. doi: 10.1146/annurev.ge.22.120188.003215. PMID: 2853609.
7. Hartl FU. Molecular chaperones in cellular protein folding. Nature. 1996 Jun 13;381(6583):571-9. doi: 10.1038/381571a0. PMID: 8637592.
8. Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. Molecular chaperone functions in protein folding and proteostasis. Annu Rev Biochem. 2013;82:323-55. doi: 10.1146/annurev-biochem-060208-092442. PMID: 23746257..
9. Pirkkala L, Nykänen P, Sistonen L. Roles of the heat shock transcription factors in regulation of the heat shock response and beyond. FASEB J. 2001 May;15(7):1118-31. doi: 10.1096/fj00-0294rev. PMID: 11344080.
10. Garrido C, Gurbuxani S, Ravagnan L, Kroemer G. Heat shock proteins: endogenous modulators of apoptotic cell death. Biochem Biophys Res

Commun. 2001 Aug 24;286(3):433-42. doi: 10.1006/bbrc.2001.5427. PMID: 11511077..

11. Shama S, Lai CY, Antoniazzi JM, Jiang JC, Jazwinski SM. Heat stress-induced life span extension in yeast. Exp Cell Res. 1998 Dec 15;245(2):379-88. doi: 10.1006/excr.1998.4279. PMID: 9851879.
12. Guerzoni ME, Lanciotti R, Cocconcelli PS. Alteration in cellular fatty acid composition as a response to salt, acid, oxidative and thermal stresses in Lactobacillus helveticus. Microbiology (Reading). 2001 Aug;147(Pt 8):2255-2264. doi: 10.1099/00221287-147-8-2255. PMID: 11496002.
13. Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. Nature. 2002 Jun 6;417(6889):618-24. doi: 10.1038/nature749. Epub 2002 May 12. PMID: 12050657.
14. Rutherford SL. Between genotype and phenotype: protein chaperones and evolvability. Nat Rev Genet. 2003 Apr;4(4):263-74. doi: 10.1038/nrg1041. PMID: 12671657.
15. Sangster TA, Salathia N, Undurraga S, Milo R, Schellenberg K, Lindquist S, Queitsch C. HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. Proc Natl Acad Sci U S A. 2008 Feb 26;105(8):2963-8. doi: 10.1073/pnas.0712200105. Epub 2008 Feb 19. PMID: 18287065; PMCID: PMC2268568.
16. Rohde JR, Bastidas R, Puria R, Cardenas ME. Nutritional control via Tor signaling in Saccharomyces cerevisiae. Curr Opin Microbiol. 2008 Apr;11(2):153-60. doi: 10.1016/j.mib.2008.02.013. Epub 2008 Apr 8. PMID: 18396450; PMCID: PMC2394285.
17. Lu RC, Tan MS, Wang H, Xie AM, Yu JT, Tan L. Heat shock protein 70 in Alzheimer's disease. Biomed Res Int. 2014;2014:435203. doi: 10.1155/2014/435203. Epub 2014 Nov 6. PMID: 25431764; PMCID: PMC4241292.
18. Moloney TC, Hyland R, O'Toole D, Paucard A, Kirik D, O'Doherty A, Gorman AM, Dowd E. Heat shock protein 70 reduces α-synuclein-induced predegenerative neuronal dystrophy in the α-synuclein viral gene transfer rat model of Parkinson's disease. CNS Neurosci Ther. 2014 Jan;20(1):50-8. doi: 10.1111/cns.12200. Epub 2013 Nov 27. PMID: 24279716; PMCID: PMC6493192.

19. Patterson KR, Ward SM, Combs B, Voss K, Kanaan NM, Morfini G, Brady ST, Gamblin TC, Binder LI. Heat shock protein 70 prevents both tau aggregation and the inhibitory effects of preexisting tau aggregates on fast axonal transport. Biochemistry. 2011 Nov 29;50(47):10300-10. doi: 10.1021/bi2009147. Epub 2011 Nov 8. PMID: 22039833; PMCID: PMC3387688.
20. Stetler RA, Gan Y, Zhang W, Liou AK, Gao Y, Cao G, Chen J. Heat shock proteins: cellular and molecular mechanisms in the central nervous system. Prog Neurobiol. 2010 Oct;92(2):184-211. doi: 10.1016/j.pneurobio.2010.05.002. Epub 2010 Jun 4. PMID: 20685377; PMCID: PMC2939168.
21. Lackie RE, Maciejewski A, Ostapchenko VG, Marques-Lopes J, Choy WY, Duennwald ML, Prado VF, Prado MAM. The Hsp70/Hsp90 Chaperone Machinery in Neurodegenerative Diseases. Front Neurosci. 2017 May 16;11:254. doi: 10.3389/fnins.2017.00254. PMID: 28559789; PMCID: PMC5433227.
22. Moll A, Ramirez LM, Ninov M, Schwarz J, Urlaub H, Zweckstetter M. Hsp multichaperone complex buffers pathologically modified Tau. Nat Commun. 2022 Jun 27;13(1):3668. doi: 10.1038/s41467-022-31396-z. PMID: 35760815; PMCID: PMC9237115.
23. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell. 2000 Dec;11(12):4241-57. doi: 10.1091/mbc.11.12.4241. PMID: 11102521; PMCID: PMC15070.
24. Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, Lee TI, True HL, Lander ES, Young RA. Remodeling of yeast genome expression in response to environmental changes. Mol Biol Cell. 2001 Feb;12(2):323-37. doi: 10.1091/mbc.12.2.323. PMID: 11179418; PMCID: PMC30946.
25. Santoro N, Johansson N, Thiele DJ. Heat shock element architecture is an important determinant in the temperature and transactivation domain requirements for heat shock transcription factor. Mol Cell Biol. 1998 Nov;18(11):6340-52. doi: 10.1128/MCB.18.11.6340. PMID: 9774650; PMCID: PMC109220.
26. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. Genomic expression programs in the response of yeast

cells to environmental changes. Mol Biol Cell. 2000 Dec;11(12):4241-57. doi: 10.1091/mbc.11.12.4241. PMID: 11102521; PMCID: PMC15070.

27. Berry DB, Gasch AP. Stress-activated genomic expression changes serve a preparative role for impending stress in yeast. Mol Biol Cell. 2008 Nov;19(11):4580-7. doi: 10.1091/mbc.e07-07-0680. Epub 2008 Aug 27. PMID: 18753408; PMCID: PMC2575158.
28. Singer MA, Lindquist S. Thermotolerance in Saccharomyces cerevisiae: the Yin and Yang of trehalose. Trends Biotechnol. 1998 Nov;16(11):460-8. doi: 10.1016/s0167-7799(98)01251-7. PMID: 9830154.
29. Morimoto RI. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. Genes Dev. 2008 Jun 1;22(11):1427-38. doi: 10.1101/gad.1657108. PMID: 18519635; PMCID: PMC2732416.
30. Subjeck JR, Sciandra JJ, Johnson RJ. Heat shock proteins and thermotolerance; a comparison of induction kinetics. Br J Radiol. 1982 Aug;55(656):579-84. doi: 10.1259/0007-1285-55-656-579. PMID: 7116088.
31. Kültz D. Evolution of the cellular stress proteome: from monophyletic origin to ubiquitous function. J Exp Biol. 2003 Sep;206(Pt 18):3119-24. doi: 10.1242/jeb.00549. PMID: 12909693.
32. Kültz D. Molecular and evolutionary basis of the cellular stress response. Annu Rev Physiol. 2005;67:225-57. doi: 10.1146/annurev.physiol.67.040403.103635. PMID: 15709958.
33. Wang P, Bouwman FG, Mariman EC. Generally detected proteins in comparative proteomics--a matter of cellular stress response? Proteomics. 2009 Jun;9(11):2955-66. doi: 10.1002/pmic.200800826. PMID: 19415655.
34. Halliwell B. Reactive species and antioxidants. Redox biology is a fundamental theme of aerobic life. Plant Physiol. 2006 Jun;141(2):312-22. doi: 10.1104/pp.106.077073. PMID: 16760481; PMCID: PMC1475431.
35. Morano KA, Grant CM, Moye-Rowley WS. The response to heat shock and oxidative stress in Saccharomyces cerevisiae. Genetics. 2012 Apr;190(4):1157-95. doi: 10.1534/genetics.111.128033. Epub 2011 Dec 29. PMID: 22209905; PMCID: PMC3316637.
36. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013 May 1;41(10):e108.

doi: 10.1093/nar/gkt214. Epub 2013 Apr 4. PMID: 23558742; PMCID: PMC3664803.

37. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015 Jan;43(Database issue):D447-52. doi: 10.1093/nar/gku1003. Epub 2014 Oct 28. PMID: 25352553; PMCID: PMC4383874..
38. Dohn R, Xie B, Back R, Selewa A, Eckart H, Rao RP, Basu A. mDrop-Seq: Massively Parallel Single-Cell RNA-Seq of *Saccharomyces cerevisiae* and *Candida albicans*. Vaccines (Basel). 2021 Dec 27;10(1):30. doi: 10.3390/vaccines10010030. PMID: 35062691; PMCID: PMC8779198.
39. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015 May;33(5):495-502. doi: 10.1038/nbt.3192. Epub 2015 Apr 13. PMID: 25867923; PMCID: PMC4430369.
40. Patterson-Cross RB, Levine AJ, Menon V. Selecting single cell clustering parameter values using subsampling-based robustness metrics. BMC Bioinformatics. 2021 Feb 1;22(1):39. doi: 10.1186/s12859-021-03957-4. PMID: 33522897; PMCID: PMC7852188.
41. Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun. 2019 Jan 22;10(1):380. doi: 10.1038/s41467-018-08023-x. PMID: 30670690; PMCID: PMC6342984.
42. Zakrzewska A, van Eikenhorst G, Burggraaff JE, Vis DJ, Hoefsloot H, Delneri D, Oliver SG, Brul S, Smits GJ. Genome-wide analysis of yeast stress survival and tolerance acquisition to analyze the central trade-off between growth rate and cellular robustness. Mol Biol Cell. 2011 Nov;22(22):4435-46. doi: 10.1091/mbc.E10-08-0721. Epub 2011 Sep 30. PMID: 21965291; PMCID: PMC3216668.