

Research Article

Open Access



CrossMark

- <sup>1</sup> Bioinformatics Department, Agriculture Genetic Engineering Research Institute, Agriculture Research Center, Giza, Egypt.
- <sup>2</sup> Biotechnology/Biomolecular Chemistry Department, Faculty of Science, Cairo University, Egypt.
- <sup>3</sup> Wheat Res. Dept., Field Crops Res. Inst., ARC, Giza, Egypt.
- <sup>4</sup> Genome Mapping Department, Molecular Genetics and Genome Mapping Laboratory, Agricultural Genetic Engineering Research Institute, Giza, Egypt.

\* To whom correspondence should be addressed: saifeldeenmib99@gmail.com

**Editor:** Aladdin Hamwih, *International Center for Agricultural Research in the Dry Areas (ICARDA), Giza, Egypt.*

**Reviewer(s):**  
Ayed M. Al-Abdallat, *aculty of Agriculture, The University of Jordan, Jordan.*

Tawfiq Istanbuli, *International Center for Agricultural Research in the Dry Areas (ICARDA), Beirut, Lebanon*

**Received:** October 20, 2023

**Accepted:** December 20, 2023

**Published:** December 26, 2023

**Citation:** Ibrahim MS, Ibrahim SM. Unlocking the Genetic Basis of Abiotic Stress Tolerance in Wheat: Insights from Differential Expression Analysis and Machine Learning. 2023 Dec 26;6:bs202307

**Copyright:** © 2023 Ibrahim and Ibrahim. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and supplementary materials.

**Funding:** The authors have no support or funding to report.

**Competing interests:** The authors declare that they have no competing interests.

## Unlocking the Genetic Basis of Abiotic Stress Tolerance in Wheat: Insights from Differential Expression Analysis and Machine Learning

Saifeldeen M. Ibrahim <sup>\*1</sup> , Manar S. Ibrahim <sup>1</sup> , Radwa Khaled<sup>2</sup> , Ahmed Fawzy Elkot <sup>3</sup> , Shafik D. Ibrahim<sup>4</sup> 

### Abstract

Abiotic stresses such as heat and cold temperatures, salinity, and drought are threatening global food security by affecting crop quality and reproductivity. Wheat is the most essential staple crop in the world, its complex genome is the main barrier to finding valuable genes responsive to different stresses. Thus, in our study we conducted differential RNA-seq analysis to identify Differentially Expressed Genes (DEGs) involved in 4 different stresses such as drought, heat, freeze resistance, and water-deficit stress, then applied two machine learning models; the "Extra-tree regressor" and LIME algorithms to accurately predict and select the highly significant genes. Our findings identified a set of 36 significant genes, many of which play important roles in various molecular functions, cellular components, and biological processes related to the response or resistance to abiotic stress in wheat. For example, Hsp101b is a member of the heat shock protein family, which protects cells against stress by stabilizing proteins. BADH, an enzyme involved in the synthesis of stress hormones, is important for the plant's response to different stresses. AGL14 is a member of the AGL protein family, which regulates gene expression and is involved in the plant's response to drought, cold, and salinity stresses. This study demonstrates the prospects of the integration of bioinformatics tools as well as machine learning models to assess the genes responsible for wheat stress resistance, genes' regulatory networks, and their functions in order to save time and cost to improve wheat productivity.

**Keywords:** Wheat, Abiotic stress, Differential Gene Expression, Machine Learning.

### Introduction

Wheat (*Triticum spp.*) is one the most strategically important crops for high proportions of the world population, supplying merely 55% of carbohydrates and 20% of dietary proteins of the world's consumed food [1; 2]. Moreover, it has high nutritional value as it is a rich source of vitamins and minerals [3]. In terms of economic importance, wheat is the third most widely grown crop in the world after rice and maize, the annual production of wheat reached approximately 778.6 million tons in the 2021-2022 season [4; 5]. In addition to being a major staple food, wheat is also used in the production of various other products, such as flour, pasta, bread, and cereals [6]. It is a versatile grain that can be used in a wide variety of food products, making it an important commodity in the global food industry.

Climate change endangers plant productivity by increasing the intensity and extent of numerous abiotic stresses, such as heat, salinity, and drought [7]. For instance, drought stress decreased wheat yields globally by 32%, it is estimated that more than half of the world's cultivated area will experience water scarcity by 2050. Wheat yields are also being reduced by 40% due to soil salinity contamination [8; 9]. Plants suffer from abiotic stresses differently, to withstand this threat plants have developed a variety of biochemical, physiological, and metabolic responses. For instance, numerous stress-responsive genes are activated, which are involved in producing many proteins that aid in activating and adjusting the physiological and biochemical pathways in stress tolerance [10].

Understanding the molecular pathways and mechanisms underlying stress tolerance in crops through genomics, transcriptomics, and proteomics techniques pave the way for the identification of genetic biomarker hallmarks involved with high-stress tolerance and the recognition of the processes behind adaptation to stresses [11].

The contributions of the plant proteome have become more relevant for understanding gene activity and networks in response to an external stimulus [7]. Micro-arrays and RNA-seq techniques are dramatically evolving transcriptomics studies by providing a massive amount of data about the genes involved in stress response, downstream signaling, and the synthesis of stress response molecules at a specific time point and under particular conditions in plants in order to create crop types that can endure biotic and abiotic stress and produce a greater yield [12]. The comparative transcriptome study between cultivars that are resistant to drought and those that are vulnerable identifies possible genes and processes of adaptation to drought stress [13].

Several CIPK genes are elevated in rice, especially under drought-stressed situations, according to earlier comparative analysis investigations [14]. Le et al. demonstrated a large number of kinase-encoding genes that are drought-inducible, including PP2C proteins, hormone-signaling-related proteins, MAP and CIPK kinases, that were thought to be important in the control of the drought response in soybean leaves through microarray analysis. While the over-expression of the ANAC019, ANAC055, and ANAC072 genes in *Arabidopsis* provided the first evidence of the roles of NAC TFs in the enhancement of drought tolerance in plants [15].

The crop genomics field cannot simply interpret molecular complex phenotypes, due to the large, diverse, and heterogeneous data-sets, so leveraging strong data mining algorithms and bioinformatics techniques to anticipate and interpret these phenotypes is crucial [16]. The recent advances in Machine learning Algorithms (MLA) integrated with omics data analysis helped in the recognition, categorization, measurement, and early prediction of plant stress responses in addition to the extensive metabolic description for the target plant species [17]. Machine learning has been applied to a variety of fields, including medicine, engineering, biology, and genomics, it has the potential to revolutionize the way we analyze and interpret gene expression data. One application of machine learning in gene expression analysis is the identification of genes that are differentially expressed between different samples or conditions. By analyzing gene expression data in the context of other genomic data, such as DNA sequence data or protein-protein interaction data, machine learning algorithms can identify patterns and relationships that can provide insights into gene function and the underlying biological processes [18; 19].

The MLAs have recently introduced successful models for the agriculture field. Osco et al. [20] successfully applied the Artificial Neural Networks (ANN) to differentiate between the hyper-spectral response of water-stressed lettuce from the non-

stressed group with an accuracy reached 93%. Another study found that it was possible to evaluate water stress in winter wheat crops over time in connection to other factors including disease and nitrogen accumulation by employing continuous wavelet analysis, Fisher's linear discrimination analysis, and support vector machines [21].

Many studies and research projects have used machine learning to identify genes that are differentially expressed between different samples or conditions. In a study of the molecular markers of drought stress in wheat, Priya et al. [22] used ML algorithms to identify differentially expressed genes that were associated with drought tolerance. They found that the genes identified by the ML analysis were significantly enriched for functions related to drought tolerance, highlighting the potential of ML for identifying key genes involved in stress responses. Another study used differential gene expression analysis and ML to identify genes associated with rust resistance in wheat [23]. The genes identified by the ML analysis were significantly enriched for functions related to rust resistance, and they concluded that the use of bioinformatics approaches, including ML, can be an effective way to identify genes associated with stress responses in wheat. Another study that applied machine learning to gene expression data in wheat used a decision tree algorithm to identify genes that were differentially expressed between different varieties of wheat and to predict their potential functions [24]. Thanmalagan et al. [25] applied machine learning to gene expression data from rice (*Oryza sativa*) to predict gene function and identify pathways that were important for drought tolerance. They used a machine learning algorithm to identify patterns in the gene expression data and predict gene function based on those patterns. Also, they found that the machine learning algorithm was able to accurately predict gene function and identify pathways that were involved in drought tolerance, including those related to stress response and carbohydrate metabolism. Based on the previous investigations, the current study aimed to use both approaches (DE analysis and ML algorithms) to identify the molecular markers related to the most significant genes associated with abiotic stress.

## Methodology

### Data gathering

The Gene Expression Omnibus (GEO) was utilized to retrieve four distinct wheat dataset experiments related to various abiotic stresses using the terms Such as drought and heat stress resistance, freeze resistance, and water deficit response. The experiment name of each accession is supplied for all GSE accessions in Table 1.

### Differential gene expression analysis

The statistical tool GEO2R [26] was used to analyse the raw four gene expression data sets of wheat under different abiotic stresses using the R/Bioconductor and Limma packages. Each sample has been divided into two categories (control and treatment). The differentially expressed genes (DEGs) were then

**Table 1.** Information for the four GEO datasets for wheat.

accession	experiment name	abbreviation
GSE45563	Drought, heat and combined stress in durum wheat	DHSW
GSE14697	Freeze resistance basis of winter wheat mutant lines	FRBW
GSE48620	Long-term growth under elevated CO <sub>2</sub> differentially suppresses biotic stress genes	GUCO
GSE45262	Transcriptomics of water-deficit Stress Responses in TAM Wheat Cultivars	TAMW

visualised using R packages by creating a volcano plot highlighting all significant genes and a heatmap of the top differentially expressed genes based on each group in all samples using  $\log_2FC$  (fold change)  $\geq 1$  and an adjusted  $p$ -value of 0.05 as the DEG threshold. Whereas up-regulated DEGs were considered when the  $\log_2FC$  (fold change)  $\geq 1$ , down-regulated DEGs were considered when the  $\log_2FC$  (fold change)  $\leq -1$ . The volcano plot displays statistical significance ( $P$  value) in accordance with the rate of fold change. It makes it possible to quickly visualise genes with significant fold changes. These genes may be the most important in terms of biology. After reading the data and filtering it according to the adjusted  $p$ -value and  $\log_2FC$ , this figure was created using the R package (ggplot2) (blue: down-regulated, red: upregulated). Following that, a heatmap of the top differentially expressed genes in the RNA-seq data set was created using the pheatmap package after filtering the data. Then, a Venn diagram was used to show all of the genes that were shared by all four wheat experiments using the interactiVenn website [27].

#### Protein-protein interaction and functional enrichment analyses

To assess the link between genes associated with wheat gene expression, the STRING database was used for protein-protein interaction analysis including functional and physical interactions [28]. For each GSE dataset, all information regarding GO, annotated keywords, and protein domains has been presented in Table 2.

#### Machine Learning Model

The "Extra-tree regressor" and "local interpretable model-agnostic explanation" algorithms were used in this study to identify the significant genes related to the response or resistance to abiotic stress in wheat. The "Extra-tree regressor" is a machine learning model that utilizes decision trees to make predictions. Decision trees are constructed by considering the characteristics of a given dataset and dividing it into smaller subsets based on the values of certain features. The "Extra-tree regressor" model differs from traditional decision tree models in that it uses random thresholds for feature selection, rather than using the mean value of each feature [29]. This allows the model to capture non-linear relationships in the data, making it well-suited for the analysis of gene expression data. The "local interpretable model-agnostic explanation" algorithm, or LIME, is a technique for explaining the predictions made by machine learning models.

It works by approximating the complex, non-linear relationships learned by the model with a simpler, interpretable model that is specific to a particular prediction [30]. This allows researchers to understand the factors that contributed to a particular prediction, and can be useful for identifying the underlying mechanisms behind the model's results.

## Results and Discussion

### Identification of DEGs

Identification of molecular mechanisms, biological processes, and cellular components for both up- and down-regulated genes through the different environmental stresses is an essential step toward enhancing wheat survival rates. The GSE14697 dataset, which highlights the genes that might confer and maintain freeze resistance in winter wheat, had the highest number of expressed genes (14,484 genes), according to our analyses. This dataset also found multiple cold-responsive (Cor)/late-embryogenesis-abundant (Lea) genes; the proteins of accumulating COR/LEA genes are thought to promote and sustain the development of freezing tolerance [31] [32]. On the other hand, the GSE45563 dataset had the fewest expressed genes, with just 1,297.

The volcano plot, shown in Figure 1, illustrates which genes are up-regulated (red) and which are down-regulated (blue) in *Triticum aestivum* samples. The fold change ( $\log_2FC$ ) is represented by the horizontal axis, while the adjusted  $p$ -values are represented by the vertical axis. The GSE14697 dataset produced 6,987 up-regulated and 7,497 down-regulated DEGs, according to the volcano plot, whereas the GSE45563 dataset produced 540 up-regulated and 757 down-regulated DEGs while, GSE48620 produced 2,693 up-regulated and 3,164 down-regulated DEGs, and GSE45262, which examined the drought-responsive genes in wheat, produced 1,221 up-regulated and 803 down-regulated DEGs.

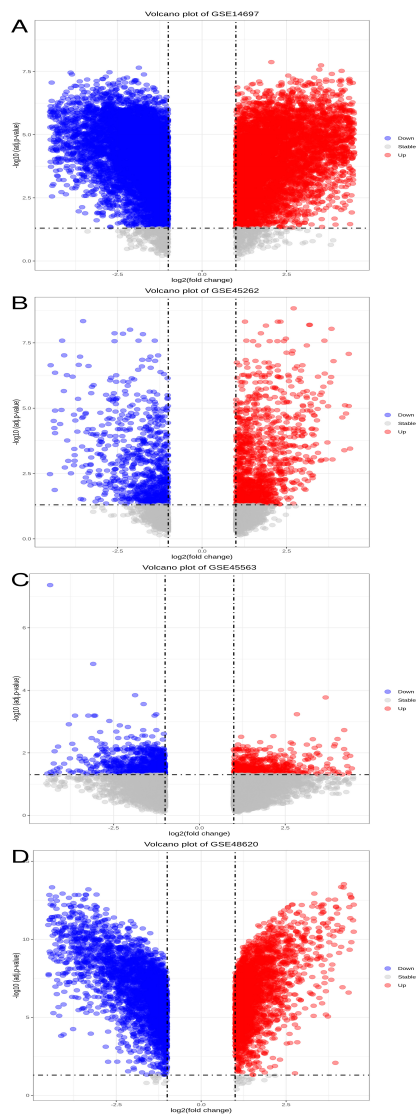
For a more in-depth analysis of the DEGs, a heatmap was generated for the top DE genes based on the adjusted  $p$ -value ranking. The heatmap divides the samples into two groups: control (blue) and treatment (red). Red cells indicate high gene expression (upregulated), whereas blue cells indicate low gene expression (downregulated). Lighter tones and white are used to depict genes with stable expression levels. The samples and genes were reordered using dendrogram hierarchical clustering, as shown in Figure 2. This heatmap gives a deeper insight into the changes in

**Table 2.** The STRING database table contains information on GO, annotated keywords, and protein domains, for each GSE dataset.

GSE	Category	ID	Description	FDR
GSE14697	Cellular Component	GO:0110165	Cellular anatomical entity	6.24E-18
GSE45262	Cellular Component	GO:0110165	Cellular anatomical entity	2.88E-12
	Molecular Function	GO:0004363	Glutathione synthase activity	0.003
		GO:0043295	Glutathione binding	0.0464
	Cellular Component	GO:0110165	Cellular anatomical entity	3.54E-12
	Subcellular localization	GOCC:0017109	Glutamate-cysteine ligase complex	0.003
		GOCC:0036087	Glutathione synthase complex	0.003
	Annotated Keywords	KW-0317	Glutathione biosynthesis	0.00026
		KW-0460	Magnesium	0.019
		PF03199	Eukaryotic glutathione synthase	0.0036
GSE45563	Protein Domains	PF03917	Eukaryotic glutathione synthase, ATP binding domain	0.0036
		IPR004887	Glutathione synthase, substrate-binding domain	0.0034
		IPR005615	Glutathione synthase	0.0034
		IPR014042	Glutathione synthase, alpha-helical	0.0034
	Protein Domains and Features	IPR014049	Glutathione synthase, N-terminal, eukaryotic	0.0034
		IPR014709	Glutathione synthase, C-terminal, eukaryotic	0.0034
		IPR037013	Glutathione synthase, substrate-binding domain superfamily	0.0034
		IPR016185	Pre-ATP-grasp domain superfamily	0.004
	Molecular Function	GO:0005200	Structural constituent of cytoskeleton	0.0293
	Cellular Component	GO:0110165	Cellular anatomical entity	1.76E-38
		KW-0342	GTP-binding	0.00015
		KW-0963	Cytoplasm	0.00015
		KW-0206	Cytoskeleton	0.0062
		KW-0493	Microtubule	0.0073
		KW-1015	Disulfide bond	0.0076
	Annotated Keywords	KW-0326	Glycosidase	0.0241
		KW-0732	Signal	0.0241
		KW-0325	Glycoprotein	0.0279
		KW-0624	Polysaccharide degradation	0.0355
		KW-0547	Nucleotide-binding	0.0426
		KW-0597	Phosphoprotein	0.0426
GSE48620		KW-0809	Transit peptide	0.0426
	Protein Domains	PF03953	Tubulin C-terminal domain	0.0035
		PF00091	Tubulin/FtsZ family, GTPase domain	0.0079
		IPR000217	Tubulin	0.00016
		IPR002453	Beta tubulin	0.00016
		IPR003008	Tubulin/FtsZ, GTPase domain	0.00016
		IPR008280	Tubulin/FtsZ, C-terminal	0.00016
	Protein Domains and Features	IPR013838	Beta tubulin, autoregulation binding site	0.00016
		IPR017975	Tubulin, conserved site	0.00016
		IPR018316	Tubulin/FtsZ, 2-layer sandwich domain	0.00016
		IPR023123	Tubulin, C-terminal	0.00016
		IPR037103	Tubulin/FtsZ, C-terminal domain superfamily	0.00016
		IPR036525	Tubulin/FtsZ, GTPase domain superfamily	0.00017



**Figure 1.** Visualization of DEGs volcano plots. The representations are as follows: x-axis,  $\log_2FC$ ; y-axis,  $-\log_{10}$  of an adjusted  $p$ -value.



gene expression at the cellular level and shows which genes are most affected by abiotic stress. This can help to identify key genetic markers for the development of improved stress tolerance wheat varieties.

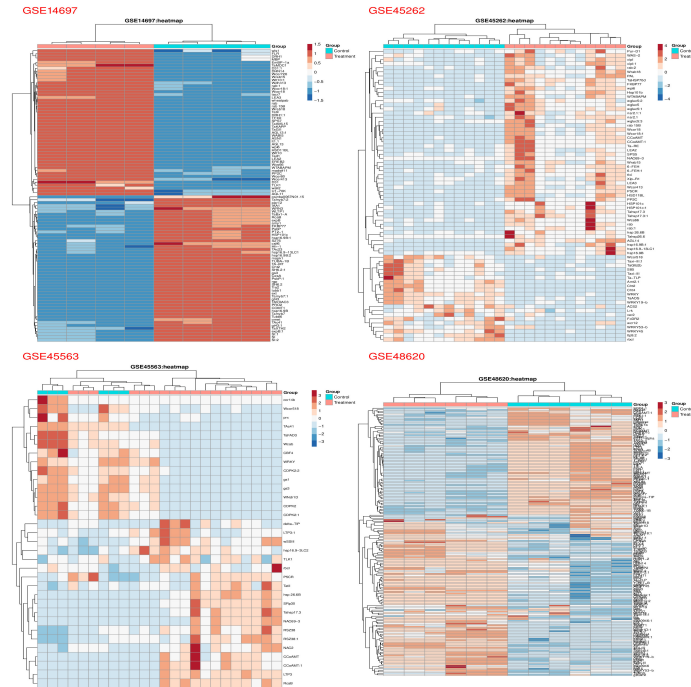
Figure (3) illustrates the Venn diagrams of the DEGs between the four integrated GEO data sets. The numbers within each circle represent the number of differently expressed genes across the various comparisons. The overlapping numbers refer to the common DEGs shared between the experiments, whilst the non-overlapping numbers refer to genes that are unique to each GSE. The GSE14697 and GSE45262 shared 19 genes, the GSE45262 and GSE45563 shared 9 genes, and the GSE14697 and GSE45563 shared 5 genes. All the common and shared genes are listed in the Table 4. It is believed that all of these genes are related to abiotic stress, as they play key roles in the

response to various types of stress, including drought and salt stress.

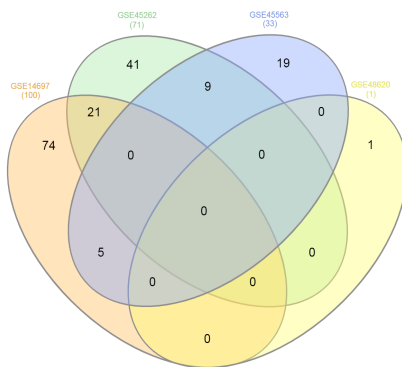
One example is Wrab15, a member of the WRKY transcription factor family, which plays a key role in the response to various types of stress, including abiotic stress. Wrab15 has been shown to be upregulated in response to drought and salt stress in wheat [33]. Another example is CCoAMT, a copper-containing amine oxidase that is involved in the detoxification of reactive oxygen species (ROS) generated during stress conditions [34]. Tat1 is a member of the aquaporin family, which plays a role in the transport of water and other small molecules across cell membranes [35]. Rab is a member of the Rab GTPase family, which regulates vesicle transport and is involved in the response to stress [36]. TLK1 is a member of the protein kinase family, which plays a role in the regulation of gene expression and cell division [37]. Wcor18 and LEA are members of the LEA (late embryogenesis abundant) protein family, which is involved in the protection of cells against stress conditions [38]. NAC693 is a member of the NAC transcription factor family, which plays a role in the regulation of gene expression and is involved in the response to stress [39]. P5CR is a member of the pyrroline-5-carboxylate reductase enzyme family, which is involved in the synthesis of Proline, a compound that plays a role in the protection of cells against stress [40]. GS3 is a member of the grain softness protein family, which is involved in the development of grain hardness and is upregulated in response to abiotic stress in wheat [41]. HSP101c, Hsp16.9B, Hsp16.913LC1, Tahsp17.3 and Hsp 26.6B are members of the heat shock protein family, which plays a role in the protection of cells against stress by stabilizing proteins and preventing their aggregation [42]. Tac41 is a member of the TAC (transcription activator-like) protein family, which plays a role in the regulation of gene expression and is involved in the response to stress [43]. Rbcl is a member of the ribulose-1,5-bisphosphate carboxylase/oxygenase enzyme family, which is involved in photosynthesis and is upregulated in response to abiotic stress. [44]. PP2C is a member of the protein phosphatase 2C enzyme family, which plays a role in the regulation of protein phosphorylation and is involved in the response to stress [45]. Wcor413 is a member of the COR protein family, which is involved in the protection of cells against cold stress [46]. FKBP77 is a member of the FK506-binding protein family, which is involved in the regulation of protein folding and is upregulated in response to abiotic stress in wheat [47].

However, a deeper understanding of the mechanisms by which these genes function and interact with each other is still needed to develop effective strategies for improving the stress tolerance of wheat. This requires further research, including functional characterization and genetic manipulation of these genes, in order to fully understand the underlying mechanisms and to apply this knowledge to improve wheat breeding and crop management practices. Additionally, it will be important to investigate the interactions between these genes and the broader genetic and molecular networks that govern the plant's response to stress.

**Figure 2.** Heatmap of top DE genes data.



**Figure 3.** Venn diagram of DEGs.



**Protein Protein Interaction (PPI) network**

Proteins have greater diversity as biomarkers, exhibit a more direct and dynamic reaction, and have excellent application prospects in cultivar screening. In this study, we utilized the STRING tool to create protein-protein interaction (PPI) networks and to make Gene Ontology (GO) annotation that spans cellular components, biological processes, and molecular functions of our studied genes that contribute to the plants response to stress. Four modules were identified in this constructed network, which consists of 90 nodes and 38 edges. The most significant module was GSE48620, which had 40 nodes and 30 edges and a PPI enrichment *p-value* of 6.95e-5, this module has some functions ex) protein domain, and features. GSE45563, the smallest module,

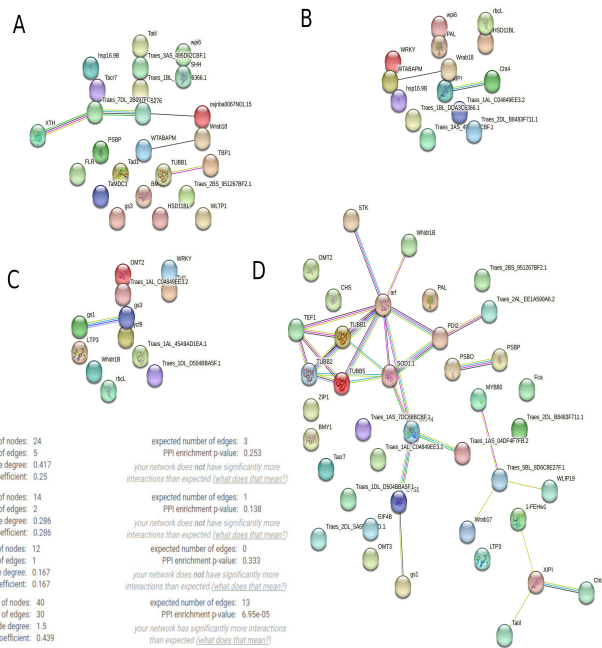
had only 12 nodes and 1 edge and a PPI enrichment *p-value* of 0.333. (Figure 4) and this module has some functions ex) protein domains and features, molecular function (a structural constituent of the cytoskeleton), cellular components, and annotated keywords.

**Machine learning model**

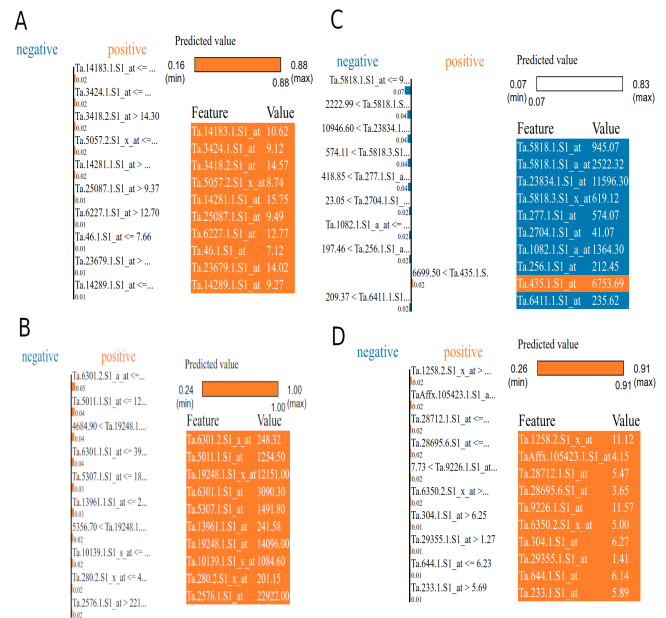
In this study, the Extra-tree regressor and LIME algorithms were employed to examine gene expression data in wheat in order to identify key genes involved in the plant’s response or resistance to abiotic stress. The results of this analysis revealed a total of 40 markers, 36 genes, and 36 gene titles, as shown in Table 3. Each GSE accession consisted of ten markers, with each marker displaying a positive or negative value. These values represent the prediction probabilities of negative and positive outcomes achieved from the classifiers and indicate the extent to which the corresponding genes are affected by the specific abiotic stress of each GSE. Blue represents negative values, while orange represents positive values (as shown in Figure 5).

The model generated a set of genes that respond to biotic stress as the most significant, including hypothetical LOC89+3077, probable light-induced protein, gsf5, PRO3, Tad1, ETT-L alpha, rab2, imidazoleglycerolphosphate dehydratase, PP2C, triticain beta, U2AF small subunit, immature spike ubiquitin-conjugating enzyme 2, Ta-RE, rab, delta tonoplast intrinsic protein TIP2;3, Hsp101b, BADH, AGL14, CDPK2, WNdr1D, cold acclimation protein COR413-TM1, gs3, ZF001, ethylene receptor-like protein, Mlo2, elongation factor, SH6.2, GS2, ribosomal protein L11, metallothionein, PR4, CPK2B, ald1 myosin, protein H2A,

**Figure 4.** PPI networks of DEGs: (a) GSE14697, (b) GSE45262, (c) GSE45563, (d) GSE48620



**Figure 5.** Machine learning results: (a) GSE14697, (b) GSE45563, (c) GSE45262, (d) GSE48620



and Waox1a.

It is important to note that these genes interact with each other and with genes from other families to coordinate the plant's response to stress. Further research is needed to fully understand the mechanisms by which these genes function and interact with one another in order to develop effective strategies for improving the stress tolerance of wheat. Additionally, more research is needed to explore the impact of these identified genes on the plant's overall growth and productivity and to validate the results of this study using other data sets and methods.

These genes are involved in a variety of molecular functions, cellular components, and biological processes that are related to the response or resistance to abiotic stress in wheat. For example, BADH is an enzyme involved in the synthesis of stress hormones, which are important for the plant's response to stress [48]. AGL14 is a member of the AGL (AGAMOUS-like) protein family, which plays a role in the regulation of gene expression and is involved in the plant's response to stress [49]. rab2 is a member of the Rab family of GTPases, which play a role in the regulation of vesicle trafficking and are involved in the plant's response to stress [50]. GS2 is a member of the glutamine synthetase family, which is involved in the synthesis of amino acids and is upregulated in response to abiotic stress in wheat [51].

Additionally, heat shock proteins, such as Hsp101b, play an important role in the plant's response to stress by protecting cellular proteins from damage [52]. Similarly, aquaporins, such as Ta-RE, are integral membrane proteins that play a role

in the transport of water and other small molecules and are involved in the plant's response to drought stress [53]. Furthermore, transcription factors, such as CDPK2, are responsible for the regulation of gene expression and play a critical role in the plant's response to stress by modulating the expression of stress-responsive genes [54].

Interestingly, the transcription factor CDPK2 may interact with the gene GS2, which is involved in the synthesis of amino acids, to modulate its expression in response to abiotic stress. This highlights the complexity of the plant's response to stress and the need for further research to fully understand the mechanisms by which these genes function and interact with each other in order to develop effective strategies for improving the stress tolerance of wheat.

Overall, the results of this study provide a valuable starting point for further research on the genetic basis of the response or resistance to abiotic stress in wheat. By identifying key genes that play a role in this process, researchers can focus their efforts on understanding the mechanisms by which these genes function and interact with each other in order to develop effective strategies for improving the stress tolerance of wheat. This can ultimately lead to the development of new wheat varieties that are more resilient to abiotic stress, which is crucial for ensuring the food security of the world's population.

**Table 3.** The results of the machine learning model represent the most significant genes with their corresponding values.

Accession	Marker	P/N	Value	Gene.Symbol	Gene.title
GSE14697	Ta.14183.1.S1_at	positive	10.62	LOC543077	hypothetical LOC543077
	Ta.3424.1.S1_at	positive	9.12	LOC543347	probable light-induced protein
	Ta.3418.2.S1_at	positive	14.57	gstf5	glutathione transferase F5
	Ta.5057.2.S1_x_at	positive	8.74	PRO3	profilin
	Ta.14281.1.S1_at	positive	15.75	Tad1	defensin
	Ta.25087.1.S1_at	positive	9.49	ETT-L.alpha	ETTIN-like auxin response factor
	Ta.6227.1.S1_at	positive	12.77	rab2	small GTP-binding protein
	Ta.46.1.S1_at	positive	7.12	LOC543224	imidazoleglycerolphosphate dehydratase
	Ta.23679.1.S1_at	positive	14.02	PP2C	protein phosphatase 2C
	Ta.14289.1.S1_at	positive	9.27	LOC100037638	triticain beta
GSE45262	Ta.5818.1.S1_at	negative	945.07	LOC780664	U2AF small subunit
	Ta.5818.1.S1_a_at	negative	2522.32	LOC780664	U2AF small subunit
	Ta.23834.1.S1_at	negative	11596.3	LOC780682	immature spike ubiquitin-conjugating enzyme 2
	Ta.5818.3.S1_x_at	negative	619.12	LOC780664	U2AF small subunit
	Ta.277.1.S1_at	negative	574.07	Ta-RE	pullulanase
	Ta.2704.1.S1_at	negative	41.07	rab	rab protein
	Ta.1082.1.S1_a_at	negative	1364.3	LOC100037645	delta tonoplast intrinsic protein TIP2:3
	Ta.256.1.S1_at	negative	212.45	Hsp101b	heat shock protein 101
	Ta.435.1.S1_at	positive	6753.69	BADH	betaine-aldehyde dehydrogenase
	Ta.6411.1.S1_at	negative	235.62	AGL14	MADS-box transcription factor TaAGL14
GSE45563	Ta.6301.2.S1_a_at	positive	248.32	CDPK2	calcium-dependent protein kinase 2
	Ta.501.1.1.S1_at	positive	1254.5	Wnrd1D	protein kinase
	Ta.19248.1.S1_x_at	positive	12151	LOC543089	cold acclimation protein COR413-TM1
	Ta.6301.1.S1_at	positive	3090.3	CDPK2	calcium-dependent protein kinase 2
	Ta.5307.1.S1_at	positive	1491.8	gs3	glutathione synthetase
	Ta.13961.1.S1_at	positive	241.58	ZF001	GATA-type zinc finger protein
	Ta.19248.1.S1_at	positive	14096	LOC543089	cold acclimation protein COR413-TM1
	Ta.10139.1.S1_s_at	positive	1084.6	LOC543446	ethylene receptor-like protein
	Ta.280.2.S1_x_at	positive	201.15	Mlo2	seven transmembrane-spanning protein
	Ta.2576.1.S1_at	positive	22922	LOC542923	elongation factor
GSE48620	Ta.1258.2.S1_x_at	positive	11.12	SH6.2	S-adenosyl-L-homocysteine hydrolase
	TaAfx.105423.1.S1_at	positive	4.15	GS2	plastid glutamine synthetase isoform GS2a
	Ta.28712.1.S1_at	positive	5.47	LOC606335	ribosomal protein L11
	Ta.28695.6.S1_at	positive	3.65	LOC542898	metallothionein
	Ta.9226.1.S1_at	positive	11.57	PR4	pathogenesis-related protein 4
	Ta.6350.2.S1_x_at	positive	5	CPK2B	calcium-dependent protein kinase
	Ta.304.1.S1_at	positive	6.27	ald	ald protein
	Ta.29355.1.S1_at	positive	1.41	LOC542906	1 myosin
	Ta.644.1.S1_at	positive	6.14	LOC543185	protein H2A
	Ta.233.1.S1_at	positive	5.89	Waox1a	alternative oxidase

**Table 4.** The common genes that were shared between the GSEs.

GSE14697 and GSE45262	GSE45262 and GSE45563	GSE14697 and GSE45563
Wrab15	CCoAMT	Tat1l
rab:1	CCoAMT:1	TLK1
Wcor18:1	NAC693	RcaB
Wcor18	P5CR	gs3
LEA3	Tahsp17.3	TAc41
rab.rab.15B	hsp.26.6B	
Wrab18	Wcor518	
wpi6	WRKY	
HSD11BL	rbcl	
LEA2		
Wes66		
WTABAPM		
PP2C		
Wcor413		
FKBP77		
HSP101c		
hsp16.9B:1		
hsp16.913LC1		
Ch2.hsp16.9B		

## Conclusion

In conclusion, our study has discovered a group of key genes that are connected to the response or resistance to abiotic stress in wheat by leveraging differential gene expression and machine learning models. The response or resilience of wheat plants to abiotic stress situations is greatly influenced by these genes. The discovered genes also come from a variety of gene families, including heat shock proteins, aquaporins, protein kinases, and protein oxidases. To coordinate the plant's response to stress, these genes interact with one another and other genes. To completely understand the mechanisms by which these genes operate and interact with one another, additional study is required in order to create effective strategies for improving the stress tolerance of wheat.

## Reference

- FAOSTAT. FAO Stat. Database. 2019. Food and Agriculture Organization of the United Nations Rome, Italy; 2019.
- Mourad AM, Alomari DZ, Alqudah AM, Sallam A, Salem KF. Recent advances in wheat (*Triticum* spp.) breeding. *Advances in Plant Breeding Strategies: Cereals: Volume 5*. 2019:559-93.
- Poole N, Donovan J, Erenstein O. Agri-nutrition research: revisiting the contribution of maize and wheat to human nutrition and health. *Food Policy*. 2021;100:101976.
- Food, of the United Nations AO. Food Outlook: Biannual report on global food markets. Food and Agriculture Organization of the United Nations. 2021. Accessed December 22, 2021. Available from: <http://www.fao.org/3/ca5162en/CA5162EN.pdf>.
- Golfam R, Kiarostami K, Lohrasebi T, Hasrak S, Razavi K. A review of drought stress on wheat (*Triticum aestivum* L.) starch. *Farming and Management*. 2021;6(1):47-57.
- Olakanmi SJ, Jayas DS, Paliwal J. Implications of Blending Pulse and Wheat Flours on Rheology and Quality Characteristics of Baked Goods: A Review. *Foods*. 2022;11(20):3287.
- Halder T, Choudhary M, Liu H, Chen Y, Yan G, Siddique KH. Wheat proteomics for abiotic stress tolerance and root system architecture: current status and future prospects. *Proteomes*. 2022;10(2):17.
- Mahmud AA, Upadhyay SK, Srivastava AK, Bhojiya AA. Biofertilizers: A Nexus between soil fertility and crop productivity under abiotic stress. *Current Research in Environmental Sustainability*. 2021;3:100063.
- El-Esawi MA, Alayafi AA. Overexpression of rice Rab7 gene improves drought and heat tolerance and increases grain yield in rice (*Oryza sativa* L.). *Genes*. 2019;10(1):56.



10. Yaqoob U, Jan N, Raman PV, Siddique KH, John R. Crosstalk between brassinosteroid signaling, ROS signaling and phenylpropanoid pathway during abiotic stress in plants: Does it exist? *Plant Stress*. 2022;100075.
11. Rakkammal K, Priya A, Pandian S, Maharajan T, Rathinapriya P, Satish L, et al. Conventional and Omics Approaches for Understanding the Abiotic Stress Response in Cereal Crops An Updated Overview. *Plants*. 2022;11(21):2852.
12. Mashabela MD, Piater LA, Steenkamp PA, Dubery IA, Tugizimana F, Mhlongo MI. Comparative metabolite profiling of wheat cultivars (*Triticum aestivum*) reveals signatory markers for resistance and susceptibility to stripe rust and aluminium (Al<sup>3+</sup>) toxicity. *Metabolites*. 2022;12(2):98.
13. Kaur B, Sandhu KS, Kamal R, Kaur K, Singh J, Röder MS, et al. Omics for the improvement of abiotic, biotic, and agronomic traits in major cereal crops: applications, challenges, and prospects. *Plants*. 2021;10(10):1989.
14. Moumeni A, Satoh K, Kondoh H, Asano T, Hosaka A, Venuprasad R, et al. Comparative analysis of root transcriptome profiles of two pairs of drought-tolerant and susceptible rice near-isogenic lines under different drought stress. *BMC plant biology*. 2011;11:1-17.
15. Tran LSP, Nakashima K, Sakuma Y, Osakabe Y, Qin F, Simpson SD, et al. Co-expression of the stress-inducible zinc finger homeodomain ZFHD1 and NAC transcription factors enhances expression of the ERD1 gene in *Arabidopsis*. *The Plant Journal*. 2007;49(1):46-63.
16. Rico-Chávez AK, Franco JA, Fernandez-Jaramillo AA, Contreras-Medina LM, Guevara-González RG, Hernandez-Escobedo Q. Machine learning for plant stress modeling: A perspective towards hormesis management. *Plants*. 2022;11(7):970.
17. Zenda T, Liu S, Dong A, Duan H. Advances in cereal crop genomics for resilience under climate change. *Life*. 2021;11(6):502.
18. van IJzendoorn DG, Szuhai K, Briaire-de Bruijn IH, Kostine M, Kuijjer ML, Bovée JV. Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. *PLoS computational biology*. 2019;15(2):e1006826.
19. Hanczar B, Zehraoui F, Issa T, Arles M. Biological interpretation of deep neural network for phenotype prediction based on gene expression. *BMC bioinformatics*. 2020;21(1):1-18.
20. Osco LP, Ramos APM, Moriya ÉAS, Bavaresco LG, Lima BCD, Estrabis N, et al. Modeling hyperspectral response of water-stress induced lettuce plants using artificial neural networks. *Remote Sensing*. 2019;11(23):2797.
21. Huang W, Lu J, Ye H, Kong W, Mortimer AH, Shi Y. Quantitative identification of crop disease and nitrogen-water stress in winter wheat using continuous wavelet analysis. *International Journal of Agricultural and Biological Engineering*. 2018;11(2):145-52.
22. Priya N, Amuthavalli A. Machine Learning Approaches to Predict the Abiotic and Biotic Stress Tolerance Genes in Plants-A Survey. *Machine Learning*;7(11):2020.
23. Sousa ICd, Nascimento M, Silva GN, Nascimento ACC, Cruz CD, Almeida DPd, et al. Genomic prediction of leaf rust resistance to Arabica coffee using machine learning algorithms. *Scientia Agricola*. 2020;78.
24. N'Diaye A, Byrns B, Cory AT, Nilsen KT, Walkowiak S, Sharpe A, et al. Machine learning analyses of methylation profiles uncovers tissue-specific gene expression patterns in wheat. *The Plant Genome*. 2020;13(2):e20027.
25. Thanmalagan RR, Roy A, Jayaprakash A, Lakshmi P. Comprehensive meta-analysis and machine learning approaches identified the role of novel drought specific genes in *Oryza sativa*. *Plant Gene*. 2022;32:100382.
26. Illimoottil M. Analyzing the Differential Expression of OPTN during Herpes Simplex Virus-2 Infection.
27. Heberle GVD, SFRTGPMR H ; Meirelles. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. 2022.
28. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research*. 2021;49(D1):D605-12.
29. Rambabu M, Ramakrishna N, Polamarasetty PK. Prediction and Analysis of Household Energy Consumption by Machine Learning Algorithms in Energy Management. In: *E3S Web of Conferences*. vol. 350. EDP Sciences; 2022. p. 02002.
30. Shi S, Du Y, Fan W. Kernel-based LIME with feature dependency sampling. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE; 2021. p. 9143-8.
31. Kobayashi F, Takumi S, Kume S, Ishibashi M, Ohno R, Murai K, et al. Regulation by Vrn-1/Fr-1 chromosomal intervals of CBF-mediated Cor/Lea gene expression and freezing tolerance in common wheat. *Journal of Experimental Botany*. 2005;56(413):887-95.



32. Sutton F, Chen DG, Ge X, Kenefick D. Cbfgenes of the Fr-A2 allele are differentially regulated between long-term cold acclimated crown tissue of freeze-resistant and-susceptible, winter wheat mutant lines. *BMC Plant Biology*. 2009;9(1):1-9.
33. Budak H, Hussain B, Khan Z, Ozturk NZ, Ullah N. From genetics to functional genomics: improvement in drought signaling and tolerance in wheat. *Frontiers in plant Science*. 2015;6:1012.
34. Kong Q, Mostafa HH, Yang W, Wang J, Nuerawuti M, Wang Y, et al. Comparative transcriptome profiling reveals that brassinosteroid-mediated lignification plays an important role in garlic adaption to salt stress. *Plant Physiology and Biochemistry*. 2021;158:34-42.
35. Kosova K, Urban MO, Vıtamvas P, Prasıl IT. Plant Abiotic Stress Proteomics: An Insight into Plant Stress Response at Proteome Level. In: *Handbook of Plant and Crop Stress, Fourth Edition*. CRC Press; 2019. p. 207-30.
36. Tsukuba T, Yamaguchi Y, Kadowaki T. Large Rab GTPases: Novel Membrane Trafficking Regulators with a Calcium Sensor and Functional Domains. *International Journal of Molecular Sciences*. 2021;22(14):7691.
37. Khalil MI, Singh V, King J, De Benedetti A. TLK1-mediated MK5-S354 phosphorylation drives prostate cancer cell motility and may signify distinct pathologies. *Molecular Oncology*. 2022.
38. Gharechahi J, Sharifi G, Komatsu S, Salekdeh GH. Proteomic analysis of crop plants under low temperature: A review of cold responsive proteins. *Agricultural Proteomics Volume 2*. 2016:97-127.
39. Sun H, Hu M, Li J, Chen L, Li M, Zhang S, et al. Comprehensive analysis of NAC transcription factors uncovers their roles during fiber development and stress response in cotton. *BMC plant biology*. 2018;18(1):1-15.
40. Garg G, Neha P. Plant transcription factors networking of pyrroline-5-carboxylate (p5c) enzyme under stress condition: A review. *Plant Archives*. 2019;19(2):562-9.
41. Biswal AK, Shamim M, Cruzado K, Soriano G, Ghatak A, Toleco M, et al. Role of biotechnology in rice production. In: *Rice production worldwide*. Springer; 2017. p. 487-547.
42. Muthusamy SK, Dalal M, Chinnusamy V, Bansal KC. Genome-wide identification and analysis of biotic and abiotic stress regulation of small heat shock protein (HSP20) family genes in bread wheat. *Journal of plant physiology*. 2017;211:100-13.
43. Liu W, Yuan JS, Stewart Jr CN. Advanced genetic tools for plant biotechnology. *Nature Reviews Genetics*. 2013;14(11):781-93.
44. Chen JH, Tang M, Jin XQ, Li H, Chen LS, Wang QL, et al. Regulation of Calvin–Benson cycle enzymes under high temperature stress. *aBIOTECH*. 2022:1-13.
45. Chen Y, Zhang JB, Wei N, Liu ZH, Li Y, Zheng Y, et al. A type-2C protein phosphatase (GhDRP1) participates in cotton (*Gossypium hirsutum*) response to drought stress. *Plant Molecular Biology*. 2021;107(6):499-517.
46. Lyu JI, Ramekar R, Kim JM, Hung NN, Seo JS, Kim JB, et al. Unraveling the complexity of faba bean (*Vicia faba* L.) transcriptome to reveal cold-stress-responsive genes using long-read isoform sequencing technology. *Scientific reports*. 2021;11(1):1-13.
47. Wang Y, Zhang X, Liu Y, Liu C, Gao J. Wheat grain softness protein gene family: Structure, function and regulation. *Frontiers in Plant Science*. 2019;10:818.
48. Ozturk M, Turkyilmaz Unal B, Garcıa-Caparros P, Khurshed A, Gul A, Hasanuzzaman M. Osmoregulation and its actions during the drought stress in plants. *Physiologia Plantarum*. 2021;172(2):1321-35.
49. Zhou B, Wang J, Lou H, Wang H, Xu Q. Comparative transcriptome analysis of dioecious, unisexual floral development in *Ribes diacanthum* pall. *Gene*. 2019;699:43-53.
50. Tripathy MK, Deswal R, Sopory SK. Plant RABs: role in development and in abiotic and biotic stress responses. *Current Genomics*. 2021;22(1):26-40.
51. Li Z, Zhu M, Huang J, Jiang S, Xu S, Zhang Z, et al. Genome-Wide Comprehensive Analysis of the Nitrogen Metabolism Toolbox Reveals Its Evolution and Abiotic Stress Responsiveness in Rice (*Oryza sativa* L.). *International Journal of Molecular Sciences*. 2022;24(1):288.
52. Katiyar-Agarwal S, Agarwal M, Grover A. Heat-tolerant basmati rice engineered by over-expression of hsp101. *Plant molecular biology*. 2003;51:677-86.
53. Liu H, Yang L, Xin M, Ma F, Liu J. Gene-wide analysis of aquaporin gene family in *Malus domestica* and heterologous expression of the gene MpPIP2; 1 confers drought and salinity tolerance in *Arabidopsis thaliana*. *International Journal of Molecular Sciences*. 2019;20(15):3710.
54. Zhao L, Li Y, Li Y, Chen W, Yao J, Fang S, et al. Systematical Characterization of the Cotton Di19 Gene Family and the Role of GhDi19-3 and GhDi19-4 as Two Negative Regulators in Response to Salt Stress. *Antioxidants*. 2022;11(11):2225.