

gsGNA: A Universal Pipeline for Mining Plant Stress-Resistance Regulators

Zhangbin Chen

Dali University

Zhenhao Zan

Dali University

Qian Lu

Dali University

Jiaxin Wang

Dali University

Junpeng Zhu

Dali University

Qing Zhao

Dali University

Pinghong Wang

Dali University

Zidan Hu

Dali University

Shujie Fan

Dali University

Xuemei Wei

wxm201@dali.edu.cn

Dali University

Junpeng Zhang

Dali University

Research Article

Keywords: rice, abiotic stress, GRNs, TFs, ensemble strategy, feature fusion

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11125843/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

gsGNA: A Universal Pipeline for Mining Plant Stress-Resistance

Regulators

Zhangbin Chen¹, Zhenhao Zan¹, Qian Lu¹, Jiabin Wang¹, Junpeng Zhu¹, Qing Zhao¹,
Pinghong Wang¹, Zidan Hu¹, Shujie Fan¹, Xuemei Wei^{1,*}, Junpeng Zhang^{1,*}

*Correspondence:

Xuemei Wei

wxm201@dali.edu.cn

Junpeng Zhang

zjp@dali.edu.cn

Full list of author information is available at the end of the article.

Abstract

Background Under global warming, important food crops such as rice are severely impacted by abiotic stresses. However, few studies have explored the mechanisms underlying abiotic stress responses from a systems biology perspective, limiting the identification of breeding targets for developing novel rice varieties with enhanced stress tolerance.

Results Utilizing rice RNA-seq data under aridity, alkaline, and cold stress conditions, we proposed gsGNA (General Stress-resistance Gene Regulatory Network), a universal computational pipeline designed to identify plant stress-resistance regulators.

This pipeline employs six ensemble strategies to integrate the predictive outputs of six machine learning and deep learning algorithms, thereby systematically constructing high-confidence gene regulatory networks (GRNs) of rice under abiotic stress. Comparative analysis demonstrated that the XGBoost ensemble method consistently exhibited stable predictive performance across all three stress conditions. Subsequent network analysis identified 20 core transcription factors (TFs) that are shared among the three stresses. Functional enrichment analysis revealed that these TFs and their target genes cooperatively regulate key biological pathways, including hormone signal transduction, the MAPK signaling pathway, and primary metabolism. Network module analysis indicated that GRNs under different stresses are highly conserved in their overall architecture but exhibit significant divergence at the functional module level, reflecting an organizational pattern characterized by a "conserved core and divergent periphery". Additionally, through a reverse screening strategy based on core pathway target genes, 45 potentially novel regulators were predicted using the GRNs, providing new clues for further dissecting rice abiotic stress responses.

Conclusions By applying the proposed gsGNA pipeline, we systematically mapped the GRN landscape of rice in response to common abiotic stresses, elucidated its network organizational patterns, and identified a set of candidate key TFs. These findings provide a theoretical foundation and a valuable target repository for rice stress-tolerance breeding.

Keywords: rice, abiotic stress, GRNs, TFs, ensemble strategy, feature fusion

Background

Plants in natural environments are frequently subjected to multiple concurrent environmental stresses, including both abiotic and biotic factors, which severely impair their growth, development, and reproduction [1]. Nevertheless, plants are not passive under such conditions; they orchestrate complex gene regulatory networks (GRNs) through various biomolecules, such as hormones, proteins, and metabolites, to mediate stress responses. These responses include the activation of defense-related genes, the regulation of metabolic pathways, and alterations in cellular structures, thereby enabling plants to survive and adapt to adverse environmental changes [2,3]. In systems biology, a GRN represents an interpretable computational model that treats transcription factors (TFs) and target genes as nodes to form network-connected graphs. This model systematically describes the complex interactions between TFs and their target genes, facilitating a comprehensive understanding of the TF-dominated regulatory systems and providing a valuable resource for discovering novel regulators and investigating underlying mechanisms [4].

Rice, as a globally essential food crop, has its yield and quality directly linked to food security. However, the intensification of global climate change has led to frequent occurrences of drought, low temperature and soil salinization, all of which severely threaten normal rice growth and productivity [5,6]. Consequently, there is an urgent need to develop new stress-tolerant rice varieties by investigating resistance mechanisms. Although several studies have explored GRNs and associated TFs in rice under individual abiotic stresses [7–9], systematic investigations of GRNs across

multiple abiotic stresses remain limited. Moreover, traditional experimental approaches are often costly and time intensive, making it difficult to comprehensively reveal the dynamic changes in GRNs under multiple stress conditions [10,11].

The rapid accumulation of omics data - including bulk RNA-seq, single-cell RNA-seq, and Stereo-seq - combined with advances in computational biology, has created new opportunities for studying abiotic stress-responsive GRNs [4]. Classical machine learning algorithms, such as Bayesian networks, random forests, support vector machines and gradient boosting, have been extensively applied to GRN inference [12–14]. Deep learning, a transformative approach capable of learning from large datasets and uncovering complex nonlinear relationships, is particularly well-suited to the challenging task of GRN inference [15]. However, individual algorithms exhibit inherent limitations in GRN inference; for example, tree-based models may overlook weak regulatory signals, mutual information methods may overestimate indirect regulatory relationships, and deep learning models often generalize poorly with small sample sizes. Multi-algorithm fusion strategies can effectively mitigate these methodological biases, reduce false positives, and improve robustness by integrating predictions from diverse methods [16,17]. Existing fusion strategies include simple aggregation, parallel ensemble methods, and serial ensemble techniques like gradient boosting and stacking, providing rich technical options for GRN inference [18]. Nevertheless, ensemble-based GRN methods have predominantly been applied to steady-state tissues in model organisms such as humans, mice, and *Arabidopsis thaliana* [19–21]; studies employing multi-model

ensemble fusion for GRN inference in rice under abiotic stress remain relatively scarce. Furthermore, multi-model ensemble fusion strategies developed for humans, mice, or *Arabidopsis* exhibit strong species specificity and context dependence, and thus cannot be directly transferred to rice abiotic stress GRN inference. This limitation arises partly from predictive distribution shifts caused by cross-species regulatory heterogeneity and partly from the fact that most existing fusion frameworks operate under a supervised learning paradigm that assumes the availability of ample, high-quality annotated data—an assumption that is difficult to reconcile with the scarcity of reliable gold-standard datasets in rice stress biology [22,23].

This study introduces an ensemble learning strategy and proposes gsGNA (General Stress-resistance Gene Regulatory Network), a universal pipeline for identifying plant stress-resistance regulators based on publicly available RNA-seq data from rice leaves subjected to three abiotic stresses: aridity, alkaline, and cold. The pipeline constructs base GRNs using six GRN inference methods, evaluates the performance of six ensemble learning strategies for integrating the feature weights of these base GRNs, and selects the optimal ensemble strategy. We analyze the topological features and reliability of the ensemble GRNs derived from the optimal strategy, use high-confidence ensemble GRNs to identify key TFs and functional modules, and mine potentially novel regulators. This work provides a theoretical foundation and candidate gene resources for enhancing stress tolerance in rice breeding.

Methods

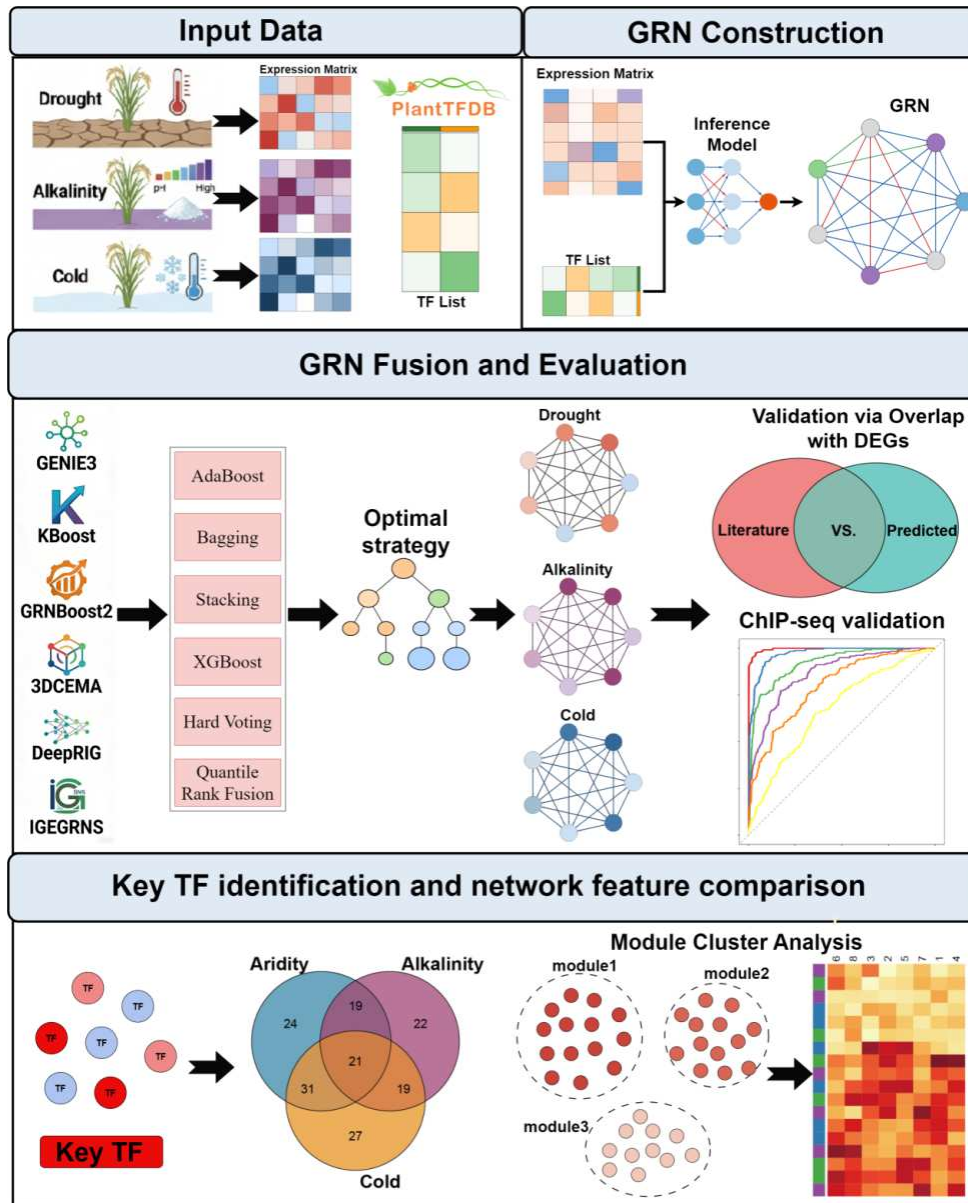
Data collection and processing

To obtain RNA-seq data from rice subjected to aridity, alkaline, and cold stress conditions, we downloaded rice-leaf RNA-seq datasets GSE121303, GSE104928 and GSE112547 from the GEO database. TF data for rice was downloaded from the PlantRegMap/PlantTFDB v5.0 database, and TFs expressed in rice leaf tissues under different stress conditions were subsequently filtered [24]. Rice ChIP-seq data were collected from the ConnecTF database [25] and relevant published literature. For external validation to evaluate the generalization capability of the algorithms, three independent biotic-stress datasets (GSE113553, GSE157400, GSE176497) were collected.

The gsGNA pipeline

We designed a GRN analysis workflow to investigate plant responses to abiotic stress, establishing a generalized pipeline termed gsGNA for identifying plant stress-resistance regulators (Fig. 1). Specifically, gene expression matrices derived from rice leaves subjected to aridity, alkaline, and cold stresses, along with the TF list obtained from the PlantTFDB database, served as input data. Six inference models were employed to construct base GRNs for each stress condition. An ensemble learning strategy was then used to integrate the feature weights from these base GRNs, generating ensemble GRNs. To evaluate and select the optimal ensemble strategy, we utilized ChIP-seq data, transcription factor binding site (TFBS) data, and published differentially expressed gene (DEG) data from rice TF interference experiments. The

133 topological structure and reliability of the GRNs generated by the optimal ensemble
 134 strategy were further analyzed. The resulting high-confidence GRNs were
 135 subsequently used to identify key regulators and functional modules, as well as to
 136 mine potentially novel regulators.



137
 138 **Fig. 1** The universal pipeline gsGNA for identifying plant stress-resistance regulators.

139 Base GRN inference

140 In this study, six GRN inference algorithms were employed as base models: two
 141 implemented in R—GENIE3 [26] and KBoost [27]—and four implemented in

Python—GRNBoost2 [28], 3DCEMA [29], DeepRIG [30] and IGEGRNS [31]. All algorithms received identical inputs: a normalized gene expression matrix and the TF list. To ensure comparability across methods, the raw regulatory weights generated by each algorithm were standardized using Z-score normalization, and only regulatory interaction pairs with $|Z| > 1.96$ (two-tailed test, $P < 0.05$) were retained.

Multi-model ensemble fusion and performance evaluation

To select the optimal ensemble strategy and obtain high-confidence ensemble GRNs, six integration schemes—AdaBoost, Bagging, Stacking, XGBoost, hard voting, and quantile rank fusion—were employed to combine the feature weights from the GRNs inferred by the aforementioned algorithms. The regulatory relationships validated by ChIP-seq data were compared with the regulatory edges predicted by these fusion algorithms, and the Area Under the Receiver Operating Characteristic Curve (AUROC) and the Area Under the Precision-Recall Curve (AUPR) were computed as evaluation metrics. The performance of the ensemble strategies was evaluated using both the raw mean of the metrics and a scenario ranking scoring system. The raw metric mean represented the average of AUROC and AUPR for each ensemble strategy across different scenarios, reflecting the model's absolute inference accuracy. In the scenario ranking scoring system, the six ensemble strategies were ranked by metric value from highest to lowest: the top-ranked strategy received 6 points, each subsequent rank was assigned one point less, and the sixth-ranked strategy received 0 points. The total score for each strategy was obtained by summing the scores across all scenarios. In addition, a random permutation null

model was constructed for significance testing to determine whether the GRN predictions significantly outperformed the random null model, following the procedure described in reference [32]. External validation datasets (GSE113553, GSE157400, GSE176497) were used to test the stability and robustness of the ensemble strategies.

The primary hyperparameter settings for the optimal ensemble strategy, XGBoost, in this study were as follows: max_depth=3, eta=0.15, nrounds=500, early_stopping_rounds=30, subsample=0.8, colsample_bytree=0.8, with a global random seed set to 123. Model evaluation was conducted using stratified 5-fold cross-validation. Complete hyperparameter details for the six ensemble strategies are provided in Supplementary Dataset 1.

Ablation experiments

To evaluate the marginal contribution of each of the six base GRN inference algorithms in the fusion framework, we systematically enumerated all possible algorithm combinations ($2^6 - 1 = 63$ in total). The performance of each combination was assessed separately under alkaline, aridity, and cold stress conditions. TF–target gene interaction pairs validated by ChIP-seq served as the gold standard, and AUROC was employed as the primary evaluation metric.

Evaluation of network reliability and predictive performance

To assess the reliability and predictive performance of the ensemble GRNs, we computed key network topological metrics, including the average degree, global clustering coefficient, average path length, network density, and degree distribution.

For comparative analysis, control random networks were generated based on the core parameters of the ensemble networks to examine their non-random structural properties. TFBS data for rice TFs were downloaded from the RiceTFtarget database [33]. The degree of overlap between these TFBS data and the predicted regulatory pairs was analyzed to further evaluate predictive performance. Additionally, published DEG data from rice TF interference experiments were used to verify the overlap between network-inferred TF–target pairs and the experimental data (Supplementary Dataset 2) [34–36].

Identification of key TFs

The degree centrality for each TF in the GRNs was computed. Using the Poisson distribution as the null model, we performed a right-tailed test to calculate the *P*-value for the degree of each TF and to evaluate statistical significance. Raw *P*-values were adjusted using the Bonferroni correction. TFs with adjusted $P < 0.05$ were designated as key regulators. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed to explore the biological functions of these key TFs under different stress conditions, as well as the enriched pathways of their target genes.

Functional module identification and GO enrichment analysis

Functional modules within the GRNs of rice leaves under distinct abiotic stresses were identified using eight network clustering algorithms available in the miRspongeR R package: FN, MCL, MCODE, Infomap, Prop, Eigen, Louvain and Walktrap [37]. Only modules containing between 10 and 300 genes were retained for

subsequent analysis. A target set of modules was selected based on both module size (number of genes) and the abundance of key TFs within each module. These selected modules were then subjected to GO functional enrichment analysis.

Screening of potentially novel regulators

To identify potentially novel key TFs involved in rice abiotic stress responses, we analyzed the KEGG pathway enrichment results of the target genes regulated by these key TFs. Core plant stress response signaling pathways were extracted, and target genes with well-validated functional roles in these pathways were designated as core target genes (threshold: BH-adjusted q -value < 0.05). The GRNs were then used to perform a reverse screening to identify upstream TFs regulating these core target genes. Following a comprehensive literature review, TFs that were present under multiple stress conditions but lacked previously reported stress-related functions were ultimately selected as novel key regulators.

Results

Rice stress responses exhibit gene transcriptional expression specificity

To explore the transcriptional regulatory characteristics of rice under different abiotic stress conditions, we examined RNA-seq data from aridity, alkaline and cold stress treatments. The analysis revealed that alkaline stress induced the expression of the largest number of genes (37,447 genes, including 1,645 TFs), followed by cold stress (33,209 genes, 1,387 TFs), while aridity stress elicited the fewest (28,829 genes, 1,268 TFs) (Fig. 2a). A total of 22,328 core expressed genes and 1,153 core TFs were

common across all three stresses. Each stress condition also exhibited a large number of specifically expressed genes (alkaline: 8,006; aridity: 4,997; cold: 4,894) and a limited number of stress-specific TFs (alkaline: 198; aridity: 19; cold: 26) (Fig. 2b). These findings indicate that all three abiotic stresses trigger extensive transcriptional responses in the rice genome, characterized by a high degree of shared genes and TFs alongside relatively few stress-specific genes and TFs. This suggests that TF-mediated GRNs under different stress responses comprise both conserved regulatory modules and distinct, stress-adaptive regulatory patterns.

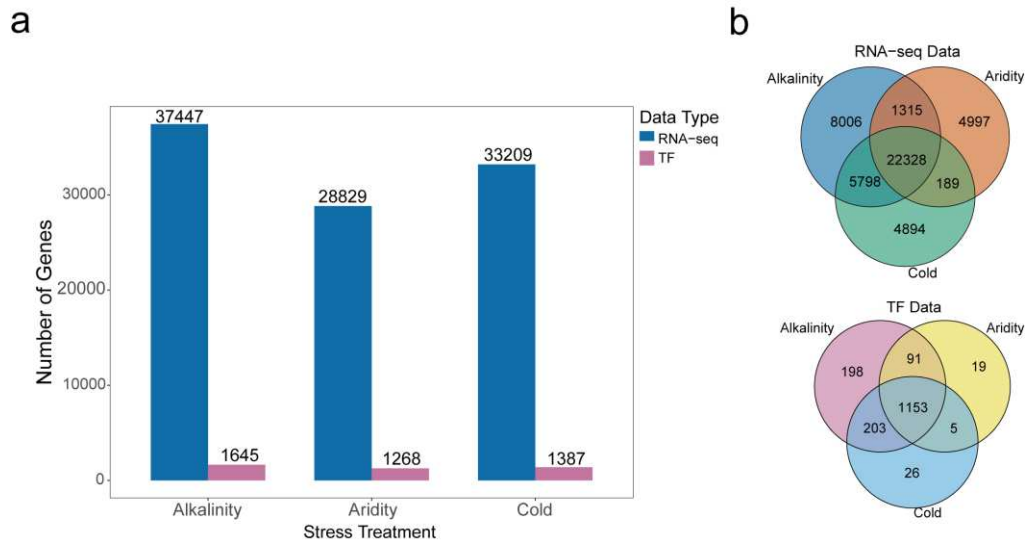


Fig. 2 Comparison of gene and TF expression levels in rice subjected to different abiotic stress conditions. **a** Bar chart showing the total number of expressed genes (depicted by blue bars) and total number of TFs (represented by purple bars) under the three stress conditions: alkaline, aridity and cold. **b** Venn diagrams showing the distribution of general genes and TFs among the three stress conditions.

The XGBoost ensemble strategy significantly improves GRN inference performance under multiple stresses

To address the inherent bias and instability of individual algorithms in GRN

inference, we employed ensemble strategies that utilized the base GRNs generated by six distinct algorithms to obtain high-confidence GRNs for subsequent functional and mechanistic analyses. To select the optimal ensemble strategy, we used ChIP-seq data to validate and compare the performance of six fusion strategies—Stacking, AdaBoost, Bagging, hard voting, quantile rank fusion, and XGBoost—in the context of GRN inference under multiple stress conditions (Supplementary Dataset 3). The cumulative ranking total scores across all scenarios (Fig. 3a) showed that XGBoost attained scores of 994 for AUPR and 944 for AUROC, significantly outperforming all other strategies. In contrast, Stacking yielded the lowest total score. A global average performance evaluation further confirmed that XGBoost simultaneously achieved the highest average AUROC and average AUPR, indicating superior predictive accuracy (Fig. 3b). Meanwhile, the per-stress, per-metric scoring results (Fig. 3c) showed that XGBoost consistently achieved the highest scores across all stress conditions and evaluation metric combinations, whereas the performance of other fusion strategies exhibited marked variability under different stress environments. By integrating the results from both evaluation frameworks—namely, the absolute metric means and the cross-scenario ranking scores—we conclude that the XGBoost ensemble fusion strategy optimally combines predictive accuracy with cross-dataset stability, thereby making it the preferred fusion strategy for the gsGNA pipeline.

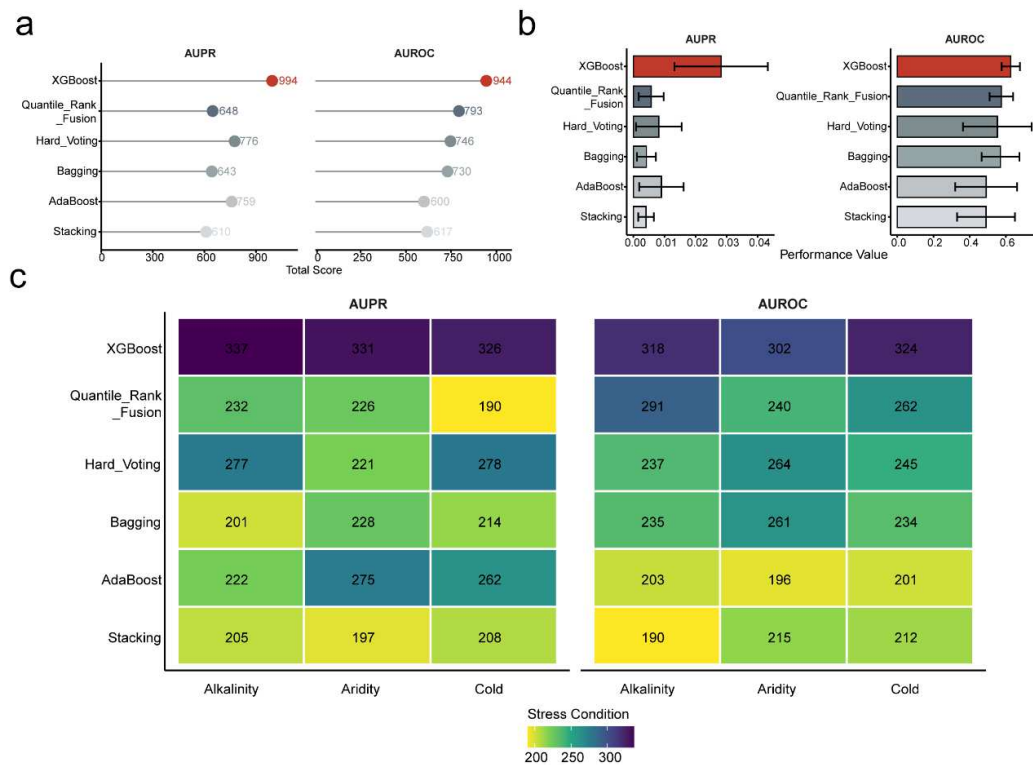


Fig. 3 Comprehensive comparative analysis of six ensemble fusion strategies for multi-stress GRN inference. **a** Cumulative ranking total scores of AUPR and AUROC metrics for each strategy across all network scenarios. **b** Global mean distributions of AUPR and AUROC metrics for each fusion strategy; error bars represent the range of metric variation. **c** Heatmap of cumulative scores under different stresses and metrics; the values in the colored cells represent the total scores for each strategy under the corresponding conditions, with a color gradient from yellow to purple indicating scores from low to high.

To investigate the performance advantage of the ensemble strategy over single GRN inference methods, we calculated AUROC using ChIP-seq datasets as the gold standard, assessed the statistical reliability of the results through permutation null model tests, and performed external validation using three independent datasets. The results showed that the predictive accuracy of the XGBoost ensemble strategy was generally superior to that of most single base algorithms, although it was slightly

279 lower than that of IGEGRNS and KBoost under aridity stress (Fig. 4a). In the
280 independent validation sets for *Magnaporthe oryzae* (MO), *Nephotettix cincticeps*
281 feeding (NC) and rice stripe virus (RSV), XGBoost achieved the highest AUROC
282 values across all datasets, reaching 0.62, 0.90, and 0.94, respectively. Permutation
283 tests indicated that the fusion network reached an extremely significant level ($P <$
284 0.001) under most stress conditions, whereas single algorithms exhibited considerable
285 performance variability and poor generalization (Fig. 4b). These results suggest that
286 the XGBoost ensemble strategy combines higher predictive accuracy with more stable
287 generalization performance across different stress scenarios.

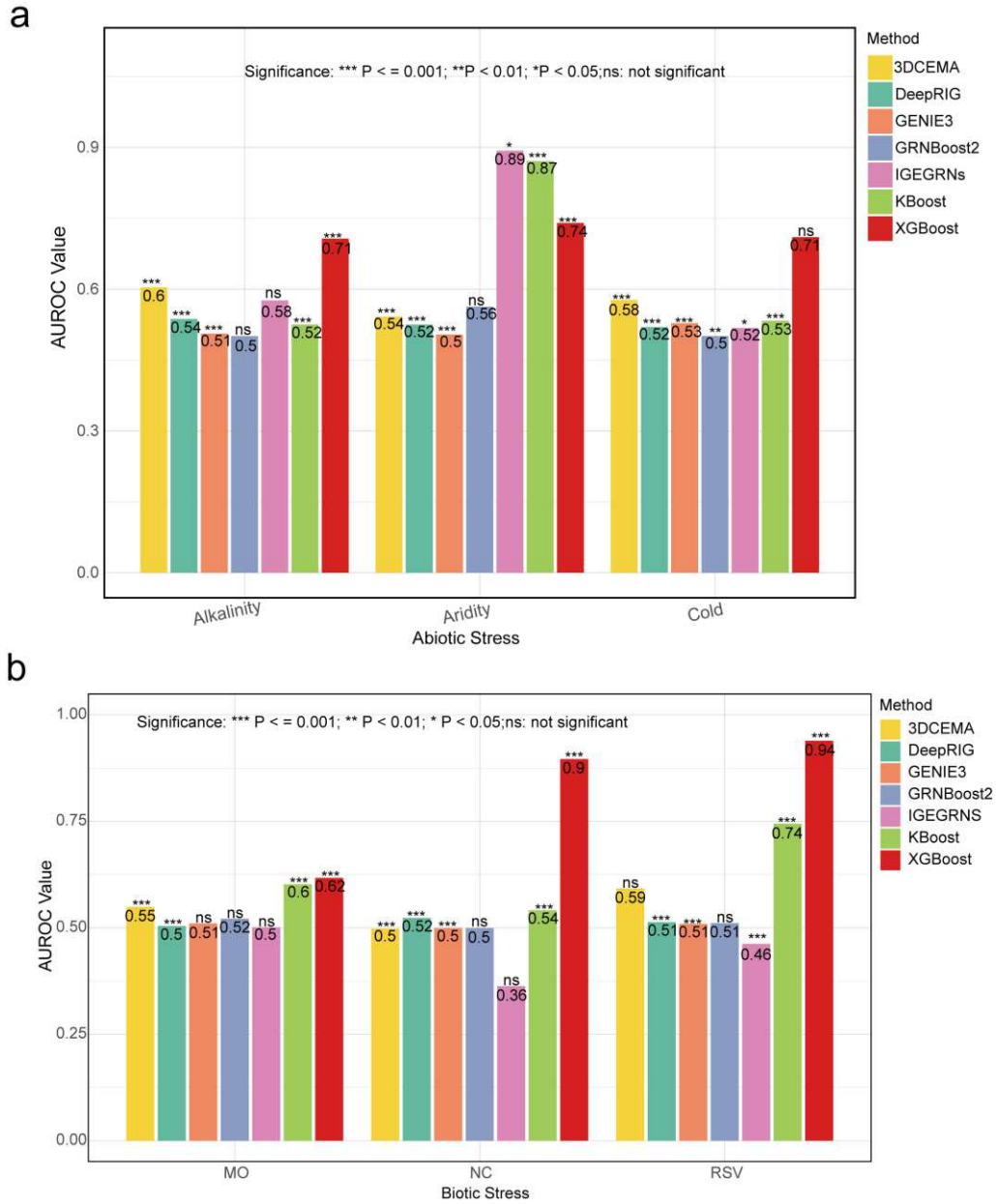


Fig. 4 Evaluation of AUROC for six individual GRN inference algorithms and the XGBoost ensemble fusion strategy on biotic and abiotic stress datasets. **a** Bar chart presents the AUROC values for each algorithm across three abiotic stress conditions. **b** Bar chart displays AUROC values for each algorithm in the three biotic stress external validation sets. Significance P-values were assessed using permutation null model tests: *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, and ns indicating non-significant.

Ablation experiments revealed that the optimal algorithm combination for XGBoost fusion varied depending on the stress type (Table 1 and Supplementary

Dataset 4). Under alkaline stress, the combination of 3DCEMA, GENIE3, GRNBoost2, and IGEGRNS was optimal (AUROC = 0.707); under aridity stress, the combination of 3DCEMA, GRNBoost2, and KBoost was optimal (AUROC = 0.733); and under cold stress, the combination of 3DCEMA, GENIE3, and GRNBoost2 was optimal (AUROC = 0.740). The AUROC values for all three optimal combinations were significantly higher than those of the full six-algorithm combination model (AUROC = 0.663, 0.615, and 0.667, respectively). 3DCEMA and GRNBoost2 were essential components in the optimal combinations across all three stress conditions, constituting the core algorithm set; KBoost contributed significantly only under aridity stress; DeepRIG did not appear in any optimal combination; and IGEGRNS made only a marginal contribution under alkaline stress. These results indicate that the best performance of XGBoost fusion is not achieved by stacking more algorithms but rather by selecting complementary subsets tailored to specific stress conditions.

Table 1 Optimal algorithm combinations of XGBoost and AUROC under the three stresses

Stress type	Optimal algorithm combination	AUROC of	AUROC of six	Improvement
		optimal	base-model	
		combination	ensemble	
3DCEMA + GENIE3				
Alkaline stress	+ GRNBoost2 +	0.707	0.663	+0.044
IGEGRNS				
Aridity stress	3DCEMA +	0.733	0.615	+0.118
GRNBoost2 +				

	KBoost			
	3DCEMA + GENIE3			
Cold stress	0.740	0.667	+0.073	
	+ GRNBoost2			

Ensemble GRNs exhibit typical biological topological features and excellent predictive performance

To validate the authenticity and biological plausibility of the GRNs constructed using the XGBoost ensemble method from the perspective of overall network architecture, we analyzed their global topological features. The log-log degree distribution demonstrated that the GRNs under all three stress conditions followed a power-law, scale-free distribution, characterized by a limited number of highly connected hub genes. Notably, the distribution in the high-node-degree region significantly deviated from that of the random networks (Fig. 5a). Quantitative comparisons of topological metrics revealed that, compared with random networks, the ensemble GRNs under aridity and cold stresses exhibited higher clustering coefficients and shorter average path lengths, displaying typical small-world characteristics. Although the average path length of the ensemble GRN under alkaline stress was slightly higher than that of the random network, its clustering coefficient was significantly higher, and the difference in path length between the two was minimal, still satisfying the criteria for classification as a small-world network [38] (Fig. 5b). These results indicate that the ensemble GRNs are not randomly connected networks but conform to the topological rules inherent in biological regulatory systems, thereby exhibiting reliability and biological rationality.

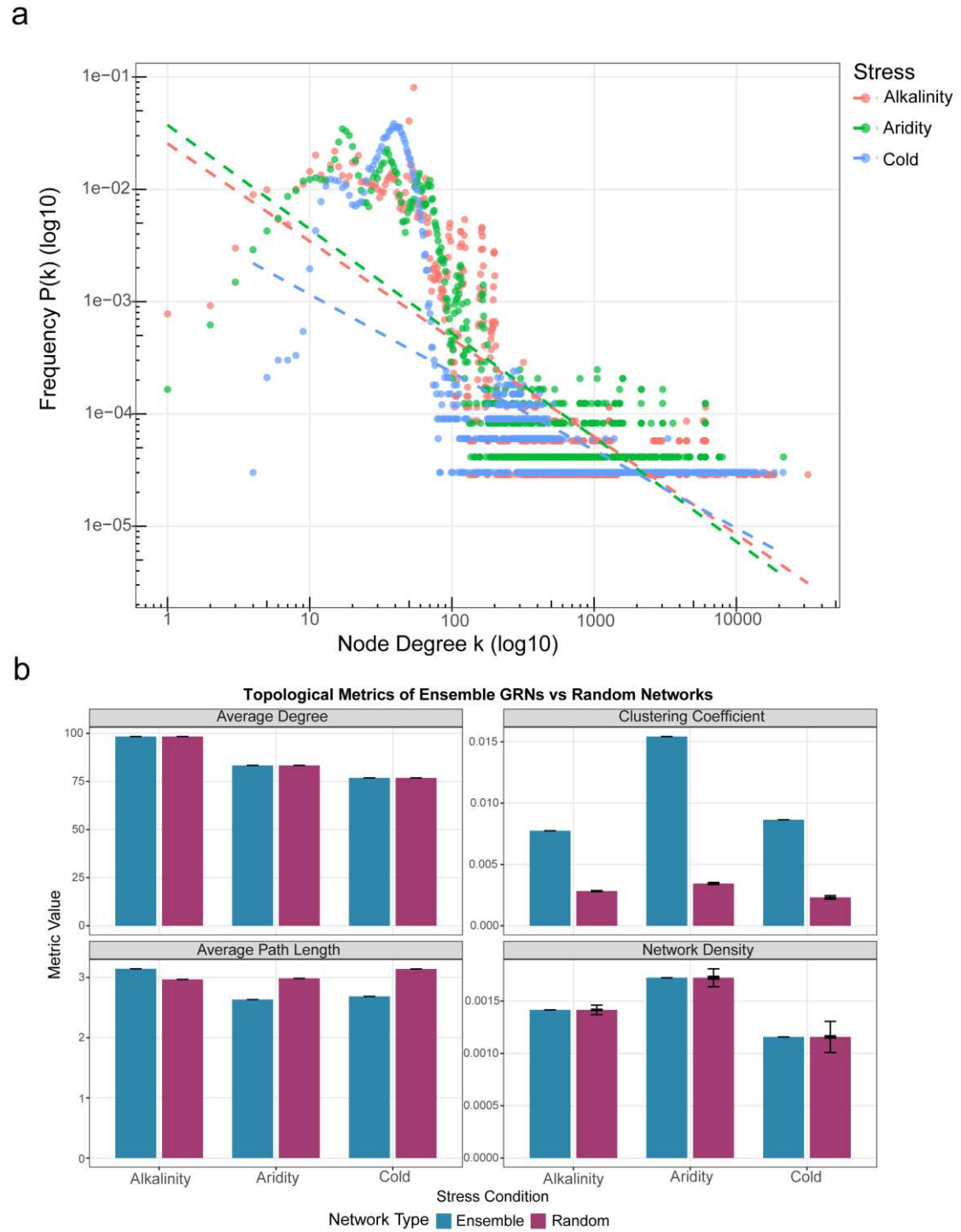


Fig. 5 Topological feature analysis of XGBoost ensemble-fused GRNs. **a** Log-log plots of node degree distribution for the fused GRNs under the three abiotic stresses. The x-axis represents node degree k ($\log_{10}$ -transformed), while the y-axis represents the frequency $P(k)$ of nodes with the corresponding degree ($\log_{10}$ -transformed). Red, green and blue scatter points represent the alkaline, aridity and cold

stress fused networks, respectively, and the dashed lines are power-law fitting curves. **b** Comparative analysis of four key topological metrics (average degree, clustering coefficient, average shortest path length and network density) between the real fused GRN (illustrated by blue bars) and an equal-scale random network (depicted by purple-red bars). Error bars represent the standard error of repeated metric calculations.

To further evaluate the reliability of the ensemble GRNs, we validated them using TFBS data and DEG data from TF interference experiments. The results showed that, under all three stress conditions, a proportion of regulatory relationships in the ensemble GRNs were supported by TFBS data. Specifically, under alkaline stress, a total of 34,126 TF-target regulatory interaction pairs were identified from TFBS data, among which 1,207 TF-target edges overlapped with those in the ensemble GRNs. Under aridity stress, 29,444 unique TFBS regulatory pairs were observed, and 1,259 regulatory relationships in the GRN predictions were corroborated by TFBS data. Under cold stress, 31,152 regulatory relationships were supported by TFBS data, with 1,269 overlapping with the inferences from the XGBoost model. Hypergeometric distribution tests revealed that all three sets of overlapping regulatory interactions were significantly enriched compared to random expectation, confirming the biological relevance of the predicted networks (Fig. 6a). Further evaluation using DEG data from three stress-related TF interference experiments (*OsWRKY1*, *OsZIP62*, *OsWRKY70*) showed that the target genes predicted by XGBoost for *OsWRKY1*, *OsZIP62*, and *OsWRKY70* overlapped with the respective DEG sets by 473, 58 and 70 genes, respectively. Hypergeometric tests indicated that these overlaps

were statistically significant ($P < 0.01$, FDR-corrected) (Fig. 6b). These findings demonstrate that the XGBoost ensemble method can effectively capture TF–target regulatory relationships supported by experimental evidence across multiple abiotic stress conditions. This confirms that the inferred ensemble GRNs are reliable and applicable, and exhibit robust predictive performance, thereby establishing a stable foundation for subsequent identification of core regulators and elucidation of functional mechanisms.

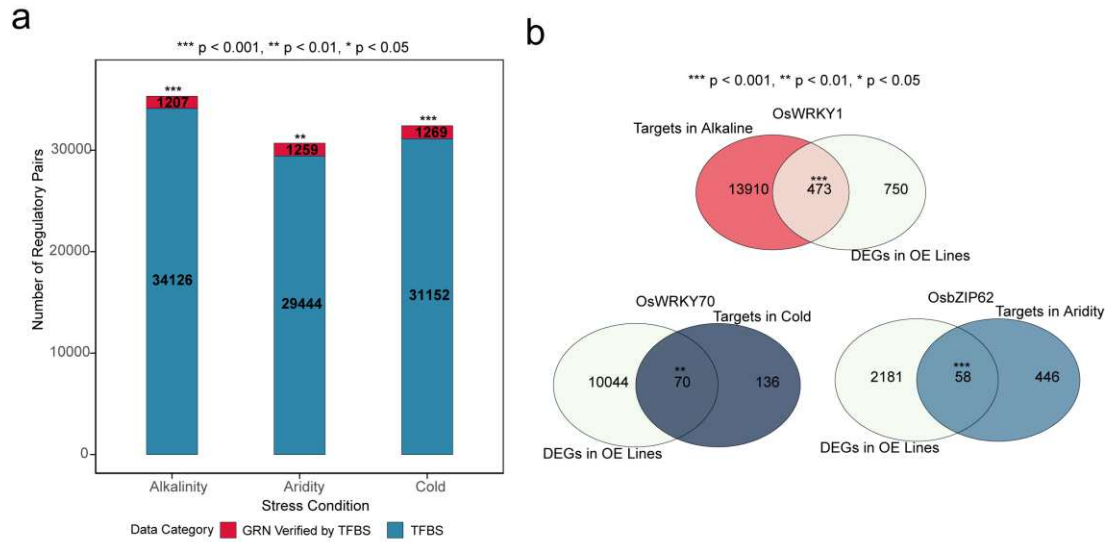


Fig. 6 Evaluation of the reliability of XGBoost ensemble-fused GRNs. **a** The bar chart shows the number of TFBS regulatory pairs obtained by screening TFs under the three abiotic stresses (depicted by blue bars), and the number of regulatory edges predicted by XGBoost and validated by TFBS under each stress (represented by red bars); significance P -values are based on hypergeometric distribution tests. **b** Overlap analysis between target genes and DEGs was conducted, showing the overlap between TF-predicted target genes and DEGs from the corresponding TF interference lines under abiotic stress; significance P -values were calculated based on hypergeometric distribution tests.

Rice possesses stress-specific and shared key TFs under different abiotic stresses

To investigate the mechanisms underlying rice responses to different abiotic stresses, we used the ensemble GRNs to identify key TFs under each stress condition. As shown in Fig. 7a–b, a total of 318 key TFs were identified under alkaline stress, and 290 under aridity stress. Cold stress exhibited the fewest key TFs, with only 202. Each stress condition had unique key TFs (32 specific to alkaline stress, 1 specific to cold stress and 0 specific to aridity stress), indicating that different abiotic stresses activate distinct independent transcriptional regulatory pathways to adapt to the specific damage they induce. Furthermore, several key TFs were shared among stress conditions: 54 were common to alkaline and aridity stresses, 46 to aridity and cold, 41 to alkaline and cold, and 20 core key TFs were shared across all three stresses. A literature review revealed that these shared TFs serve as core regulatory elements in plant responses to multiple abiotic stresses, suggesting the existence of overlapping molecular networks and common pathways in rice abiotic stress responses [39,40].

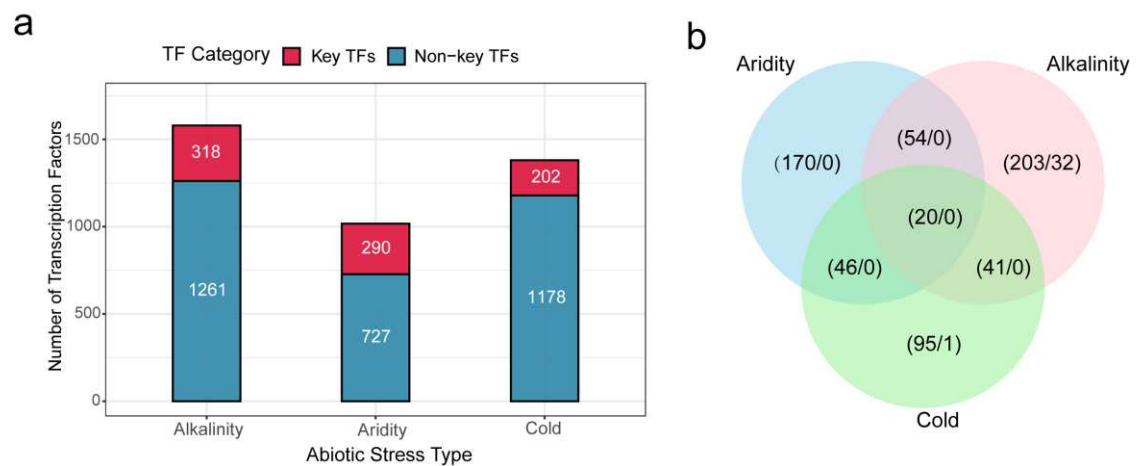


Fig. 7 Identification of key TFs based on ensemble GRNs. **a** Quantitative distribution of key and non-key TFs across different stress types. The bar chart shows the numbers of key TFs (depicted in red) and non-key TFs (depicted in blue) under alkaline, aridity and cold stress conditions. **b** Overlap of key

TFs among the three stress types. The Venn diagram shows both shared and unique key TFs among the three stress conditions. Numbers presented in parentheses follow the format (total key TFs / number of specifically expressed TFs under the corresponding stress); the second value is set to 0 for multi-stress overlapping regions, as such regions do not contain single stress specific TFs.

Ensemble GRNs follow an organizational pattern of "conserved core and divergent periphery"

To systematically compare the heterogeneity of rice GRNs under different abiotic stress conditions, we performed topological structure and functional module analyses on the three stress-responsive ensemble GRNs. Topological analysis revealed that the alkaline stress GRN was large in scale and highly interconnected; the aridity stress network exhibited a "small but dense" structure, whereas the cold stress network displayed a "large but loose" structural characteristic (Fig. 8a).

Functional module identification using multiple algorithms revealed that the Infomap algorithm detected the most functional modules (Fig. 8b). Modules containing 10 to 300 genes were retained, and those with the highest number of key TFs were selected for GO enrichment analysis. The results indicated that all three stress conditions commonly showed enrichment in fundamental stress-response pathways, such as carbohydrate metabolism and secondary metabolism. Under alkaline stress, pathways related to translation and peptide/amide biosynthesis were significantly enriched across multiple modules. Moreover, under cold stress, these same pathways exhibited even higher enrichment intensity. Although aridity stress also showed enrichment in photosynthesis and precursor energy generation, the

411 intensity was markedly lower, with extracellular stimulus response being more
412 prominent. Cold stress displayed dual enrichment characteristics, involving energy
413 metabolism, cellular homeostasis regulation, and stimulus response (Fig. 8c and
414 Supplementary Dataset 5). Functional module enrichment analysis demonstrated that
415 alkaline module 41 (photosynthesis, $q = 6.59 \times 10^{-11}$) and module 27
416 (carbohydrate/secondary metabolism, $q = 2.61 \times 10^{-9}$), cold module 54 (carbohydrate
417 metabolism, $q = 1.34 \times 10^{-8}$) and module 104 (energy generation/translation, $q = 1.34$
418 $\times 10^{-8}$), and aridity module 37 (carbohydrate/stimulus response, $q = 2.02 \times 10^{-7}$) and
419 module 5 (carbohydrate/secondary metabolism, $q = 1.21 \times 10^{-7}$) all had q-values well
420 below 0.05, indicating highly significant enrichment. These represent the core
421 modules with the strongest responses under each stress condition, and can be regarded
422 as key targets for further biological experimental analysis. In summary, the GRNs
423 under the three stress conditions are highly conserved in overall topology but exhibit
424 significant divergence at the module level.

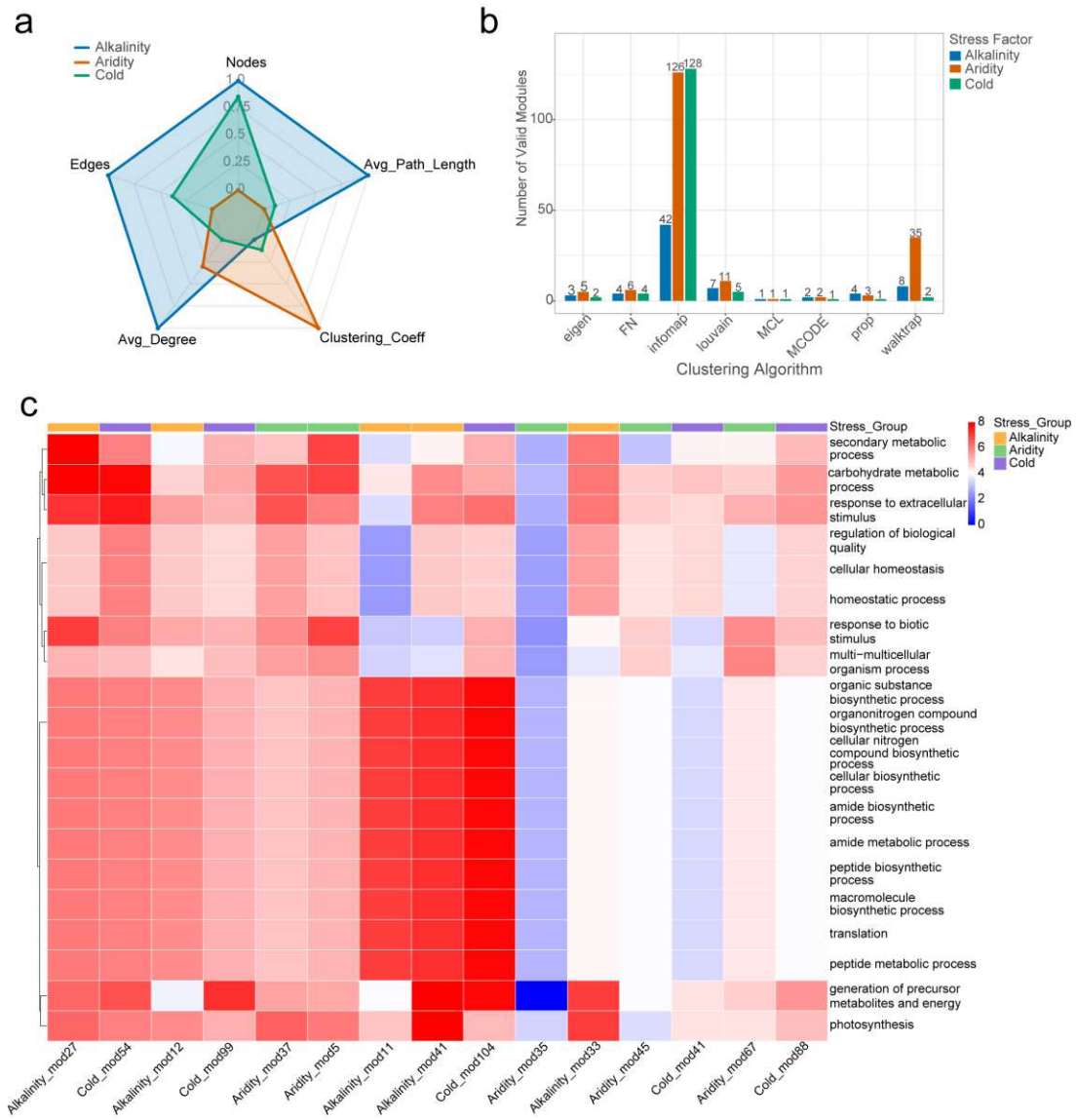


Fig. 8 Module structure and functional enrichment profiles of GRNs in rice under abiotic stresses. **a** Radar chart illustrating key topological parameters of GRNs under alkaline (blue), aridity (orange) and cold (green) stress conditions. Parameters assessed include the number of nodes (Nodes), number of edges (Edges), average degree (Avg_Degree), average path length (Avg_Path_Length), and clustering coefficient (Clustering_Coeff). **b** Number of effective modules identified across different clustering algorithms. The bar chart shows the number of effective modules detected in the GRNs under the three stress conditions using eight clustering algorithms. **c** GO functional enrichment heatmap of network modules. The heatmap presents the GO functional enrichment profiles of modules in the GRNs under

alkaline, aridity and cold stresses; the color of the cells represents enrichment significance, with deeper red indicating higher pathway enrichment and blue indicating no significant enrichment.

Reverse screening via core pathways identifies 45 potentially novel regulators of abiotic stress responses in rice

To further identify potentially novel key TFs involved in rice stress responses, we performed GO and KEGG analyses on both the key TFs under individual stress conditions and the target genes of the 20 shared key TFs across three stress conditions (Supplementary Dataset 6 and Fig. S1). The results showed that the target genes of the shared key TFs were significantly enriched in "plant hormone signal transduction" and the "MAPK signaling pathway". As conserved central hubs of stress responses in eukaryotes, these two pathways play essential roles in sensing and transmitting extracellular stress signals and activating downstream defense responses in plants [41,42]. Based on this enrichment, we selected 11 core target genes enriched in "plant hormone signal transduction" and the "MAPK signaling pathway", including *BWMK1* [43], *MAPK6* and *WRKY53* [44], *EIN2* [45], *ETR2* and *ETR3* [46], *JAR1* [47] (Fig. 9a). Using a "pathway-oriented reverse-screening" strategy, we screened TFs that regulate these core target genes within the ensemble GRNs. In total, 119 TFs targeting nine of the 11 core-pathway target genes (the remaining two had no upstream TF identified in the GRNs) were obtained, of which 90 TFs had complete annotation information. Among these, 45 TFs have been previously reported as typical developmental regulators or demonstrated to participate in stress-response functions (Supplementary Dataset 8), while the remaining 45 TFs have not been documented in prior studies

(Fig. 9b). These uncharacterized TFs may represent potentially novel key regulatory factors for further biological experimental validation.

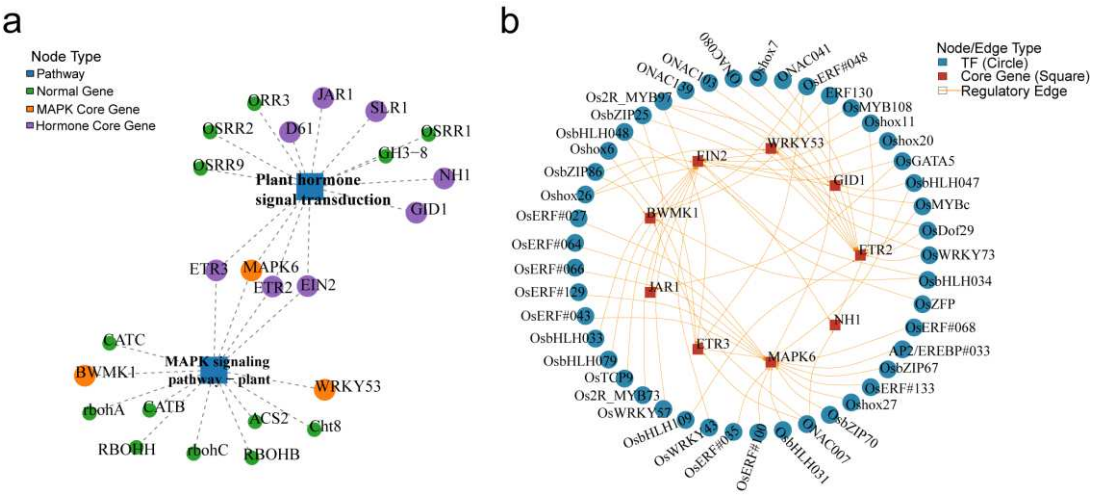


Fig. 9 Visualization of central pathways and core gene networks. **a** Target genes of common key TFs involved in two stress response central pathways (MAPK signaling pathway—plant and Plant hormone signal transduction). Blue squares represent central pathways, orange circles represent MAPK core genes, purple circles represent hormone signaling core genes, and green circles represent other core genes. **b** The regulatory network comprising potential TFs and their target genes (involved in two stress response central pathways). Blue circular nodes represent potential TFs, red square nodes represent pathway core target genes, and orange dashed lines represent regulatory interactions.

Discussion

Existing studies indicate that machine learning or deep learning approaches are suited to distinct application scenarios when analyzing transcriptomic data at different resolutions. Traditional machine learning methods, such as random forest, kernel learning, and gradient boosting, tend to offer greater stability in plant datasets characterized by bulk transcriptomics, limited sample sizes and high noise levels, as

they effectively control overfitting and capture nonlinear regulatory relationships. In contrast, deep learning methods (e.g., graph neural networks, deep autoencoders) exhibit significant advantages in handling single-cell multi-omics data and large-scale, high-dimensional datasets, excelling at modeling complex topological structures and long-range dependencies. However, these methods require substantial data volume and quality, and tend to reach performance saturation in conventional bulk RNA-seq analyses [48,49]. Each of the aforementioned individual algorithms exhibits inherent limitations in GRN inference. Multi-algorithm fusion can compensate for individual biases and reduce false positives; consequently, integrating multiple machine learning methods to construct GRNs has emerged as a mainstream paradigm in systems biology [50,51].

The theoretical foundation for the superiority of ensemble learning over single-algorithm GRN inference lies in the bias-variance-covariance decomposition, i.e., averaging individual errors by combining diverse base learners with low covariance enhances overall performance. Nevertheless, this advantage depends on algorithmic complementarity rather than simple stacking; indiscriminately adding algorithms may introduce noise. Ablation experiments conducted in this study further confirmed this point: the AUROC for the full six-algorithm combination was lower than that of the optimal subset, indicating that algorithms should be "combined on demand" rather than "the more the better" [52,53]. Among the six ensemble learning strategies, XGBoost achieved the best performance. Across three independent biotic stress datasets, XGBoost achieved the highest AUROC values, reaching 0.90 and 0.94

in the NC and RSV datasets, respectively, confirming its robust generalization across diverse stress conditions. This conclusion is also supported by other studies; for example, the MRF-XGBLC framework, which hybridizes XGBoost with random forest and Lasso-Cox, significantly outperformed existing methods on multiple datasets; similarly, BiXGBoost achieved competitive scores in the DREAM4 benchmark [54,55]. Collectively, these reports indicate that an XGBoost-based ensemble strategy, leveraging regularization to counteract overfitting, efficient feature evaluation, and a serial error-correction mechanism, has become a preferred fusion approach in GRN inference, balancing accuracy with robustness.

Reliability assessment of GRNs must incorporate multiple data types for cross-validation, as relying on a single dimension tends to overestimate network quality. Previous studies have widely adopted a four-layer validation framework integrating ChIP-seq/ChIP-chip data, TFBS, DEGs from TF interference experiments, and independent transcriptomic datasets, significantly improving the credibility of inferred regulatory relationships [56]. Integrating TF binding site data with interference data provides evidence for both physical binding and functional effects, while cross-platform consistency testing helps eliminate false positives caused by technical biases [57]. In this study, we calculated AUROC using ChIP-seq data, verified binding preferences using TFBS analysis, validated functional associations using DEGs from stress-related TF interference data, and assessed biological rationality through topological analysis. This forms a comprehensive validation chain that is highly consistent with the evaluation strategies of cutting-edge methods such as

LINGER and MultiGNN, demonstrating that the constructed GRNs possess true biological significance rather than being sets of mathematical associations [58,59].

Biological networks generally exhibit both small-world and scale-free topological features, with degree distributions following a power law, a pattern highly conserved across gene regulatory, protein-protein interaction, and metabolic networks [60]. The small-world property is characterized by a high clustering coefficient and a short average path length, which facilitates rapid and efficient transmission of regulatory signals. The scale-free feature is manifested by the presence of a limited number of hub nodes with extremely high connectivity, while most nodes have low degrees, conferring robustness against random perturbations [61]. In this study, the GRNs constructed under all three stress conditions exhibit power-law decay in degree distribution and show significant differences from random networks in key topological parameters, consistent with the characteristics of authentic biological regulatory networks. Such a structure ensures the stable operation of core stress-response pathways while allowing peripheral modules to flexibly respond to different stresses, thereby providing a structural basis for the "conserved core and divergent periphery" organizational pattern. These findings further confirm that the GRNs inferred in this study possess a high degree of biological authenticity.

Rice responses to aridity, alkalinity, and cold stresses are mediated via mechanisms that integrate both conserved core pathways and stress-specific pathways. Drought mainly involves osmotic adjustment, stomatal closure, and photosynthetic regulation; alkaline stress primarily involves ion homeostasis, reactive oxygen species

(ROS) scavenging, and osmoprotection; while cold stress centers on membrane stability, nucleic acid metabolism, and transcriptional reprogramming [8]. The 20 core TFs and 45 potentially novel TFs identified in this study belong to well-known stress-tolerance families including bZIP, WRKY, ERF, and NAC. Previous studies have demonstrated that bZIP family members participate in abscisic acid (ABA) signaling and responses to drought and salt stress, WRKY TFs regulate broad-spectrum stress tolerance through the MAPK cascade; NAC proteins play key roles in osmotic and ionic stress adaptation; and ERF members integrate hormone and stress signals [62–64]. These factors perform conserved functions across multiple stress conditions; however, their target gene profiles and regulatory intensities exhibit marked divergence under different adverse conditions, which is highly consistent with our proposed model of a "conserved core and divergent periphery." Meanwhile, the newly predicted TFs identified in this study have no previously reported stress-related functions, thereby providing high-value candidates for subsequent experimental validation.

Although our gsGNA pipeline enables robust inference of ensemble GRNs by integrating the outputs from six different GRN inference algorithms, there are still some limitations. Specifically, the networks are constructed exclusively from bulk RNA-seq data, which represent averaged transcriptional states across heterogeneous cell populations and do not capture cell-type-specific regulations. Additionally, the 45 potentially novel regulators and their roles in stress tolerance are currently only computational predictions without validation from genetic or molecular experiments.

Future work incorporating single-cell transcriptomics and targeted experimental validation will help address these limitations.

Conclusion

In this study, we developed gsGNA, a universal pipeline for inferring plant stress-responsive GRNs, characterized by "multi-algorithm fusion and multi-dimensional data evaluation." By integrating six machine learning and deep learning algorithms through an ensemble fusion strategy, we systematically constructed high-confidence GRNs for rice under three abiotic stresses: aridity, alkaline, and cold. Our analysis reveals that rice abiotic stress-responsive GRNs are organized according to a pattern of "conserved core and divergent periphery." We identified 20 core TFs common to all three stresses and discovered 45 potentially novel regulators through a core pathway reverse screening strategy. This work provides a systems perspective for understanding the molecular mechanisms of plant stress adaptation, and presents a reusable computational framework for research on plant stress GRNs.

Supplementary Information

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7 (Figure S1)

Supplementary Material 8

Acknowledgements

We thank the researchers who deposited the RNA-seq datasets (GSE104928, GSE112547, GSE121303, GSE113553, GSE157400, GSE176497) in the Gene Expression Omnibus database. We are grateful to the developers of PlantRegMap/PlantTFDB, RiceTFtarget, ConnecTF, and KOBAS for providing publicly accessible resources.

Authors' contributions

XW, JZhang and ZC conceived and designed the study. JW and JZhu collected and preprocessed the RNA-seq data. ZC performed the GRN inference, algorithm evaluation, topological analysis, module detection, key transcription factor identification, and reverse screening of novel regulators. QZ, PW, ZH and SF participated in data interpretation and validation. ZC drafted the manuscript. JZhang and XW supplied funding for the experiments and offered technical assistance, reviewed, revised and finalized the manuscript, and supervised the project. All authors read and approved the final manuscript.

Funding

This study was supported by the Doctor Research Supporting Fund of Dali University (Grant Number: KYBS2023002), the Joint Special Fund for Basic Research of Local Universities in Yunnan Province (Grant Number: 202301AO070378), and the National Natural Science Foundation of China (Grant Number: 32660173).

Data availability

The RNA-seq datasets analyzed in this study are publicly available from the NCBI Gene Expression Omnibus (GEO) under accession numbers GSE121303 (aridity stress), GSE104928 (alkaline stress), and GSE112547 (cold stress). Three independent biotic-stress datasets (GSE113553, GSE157400, GSE176497) were also obtained from GEO for external validation. Transcription-factor annotation data were retrieved from PlantRegMap/PlantTFDB v5.0. ChIP-seq datasets used for AUROC calculation were compiled from the ConnectTF database and published literature cited in the manuscript. TFBS data were downloaded from RiceTFtarget. Differentially expressed gene datasets from gene overexpression experiments were acquired from previously reported studies. All processed datasets generated in this work are available from the corresponding author upon reasonable request. Source code and analytical scripts for reproducing all results and figures are publicly accessible on GitHub at <https://github.com/CZBA/gsgNA>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors have read and approved the final manuscript and consent to its publication.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Dali University, ² Hongsheng Road, Dali, 671003, China.

Zhangbin Chen, Zhenhao Zan, Qian Lu, Jiaxin Wang, Junpeng Zhu, Qing Zhao, Pinghong Wang, Zidan Hu, Shujie Fan, Xuemei Wei & Junpeng Zhang.

References

1. Pandey P, Ramegowda V, Senthil-Kumar M. Shared and unique responses of plants to multiple individual stresses and stress combinations: physiological and molecular mechanisms. *Front Plant Sci.* 2015;6. <https://doi.org/10.3389/fpls.2015.00723>.
2. Atkinson NJ, Urwin PE. The interaction of plant biotic and abiotic stresses: from genes to the field. *J Exp Bot.* 2012;63:3523–43. <https://doi.org/10.1093/jxb/ers100>.
3. Zhu J-K. Abiotic Stress Signaling and Responses in Plants. *Cell.* 2016;167:313–24.

649 <https://doi.org/10.1016/j.cell.2016.08.029>.

650 4. Badia-i-Mompel P, Wessels L, Müller-Dott S, Trimbou R, Ramirez Flores RO,
651 Argelaguet R, et al. Gene regulatory network inference in the era of single-cell
652 multi-omics. *Nat Rev Genet.* 2023;24:739–54.
653 <https://doi.org/10.1038/s41576-023-00618-5>.

654 5. Radha B, Sunitha NC, Sah RP, T. P. MA, Krishna GK, Umesh DK, et al.
655 Physiological and molecular implications of multiple abiotic stresses on yield and
656 quality of rice. *Front Plant Sci.* 2023;13:996514.
657 <https://doi.org/10.3389/fpls.2022.996514>.

658 6. *Advances in Rice Research for Abiotic Stress Tolerance*. Elsevier; 2019.
659 <https://doi.org/10.1016/C2017-0-01486-6>.

660 7. Todaka D, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K. Toward
661 understanding transcriptional regulatory networks in abiotic stress responses and
662 tolerance in rice. *Rice.* 2012;5:6. <https://doi.org/10.1186/1939-8433-5-6>.

663 8. Gupta C, Ramegowda V, Basu S, Pereira A. Using Network-Based Machine
664 Learning to Predict Transcription Factors Involved in Drought Resistance. *Front*
665 *Genet.* 2021;12:652189. <https://doi.org/10.3389/fgene.2021.652189>.

666 9. Xiang Y, Tang N, Du H, Ye H, Xiong L. Characterization of OsbZIP23 as a Key
667 Player of the Basic Leucine Zipper Transcription Factor Family for Conferring
668 Absciscic Acid Sensitivity and Salinity and Drought Tolerance in Rice. *Plant Physiol.*
669 2008;148:1938–52. <https://doi.org/10.1104/pp.108.128199>.

670 10. Varadharajan V, Rajendran R, Muthuramalingam P, Runthala A, Madhesh V,

671 Swaminathan G, et al. Multi-Omics Approaches Against Abiotic and Biotic Stress—A
672 Review. *Plants*. 2025;14:865. <https://doi.org/10.3390/plants14060865>.

673 11. Roychowdhury R, Das SP, Gupta A, Parihar P, Chandrasekhar K, Sarker U, et al.
674 Multi-Omics Pipeline and Omics-Integration Approach to Decipher Plant's Abiotic
675 Stress Tolerance Responses. *Genes*. 2023;14:1281.
676 <https://doi.org/10.3390/genes14061281>.

677 12. Ajmal HB, Madden MG. Dynamic Bayesian Network Learning to Infer Sparse
678 Models From Time Series Gene Expression Data. *IEEE/ACM Trans Comput Biol*
679 *Bioinform*. 2022;19:2794–805. <https://doi.org/10.1109/TCBB.2021.3092879>.

680 13. Petralia F, Wang P, Yang J, Tu Z. Integrative random forest for gene regulatory
681 network inference. *Bioinformatics*. 2015;31:i197–205.
682 <https://doi.org/10.1093/bioinformatics/btv268>.

683 14. Patel N, Wang JTL. Semi-supervised prediction of gene regulatory networks using
684 machine learning algorithms. *J Biosci*. 2015;40:731–40.
685 <https://doi.org/10.1007/s12038-015-9558-9>.

686 15. Dong J, Li J, Wang F. Deep Learning in Gene Regulatory Network Inference: A
687 Survey. *IEEE/ACM Trans Comput Biol Bioinform*. 2024;21:2089–101.
688 <https://doi.org/10.1109/TCBB.2024.3442536>.

689 16. The DREAM5 Consortium, Marbach D, Costello JC, Küffner R, Vega NM, Prill
690 RJ, et al. Wisdom of crowds for robust gene network inference. *Nat Methods*.
691 2012;9:796–804. <https://doi.org/10.1038/nmeth.2016>.

692 17. Pratapa A, Jaliyal AP, Law JN, Bharadwaj A, Murali TM. Benchmarking

693 algorithms for gene regulatory network inference from single-cell transcriptomic data.
694 Nat Methods. 2020;17:147–54. <https://doi.org/10.1038/s41592-019-0690-6>.

695 18. Mienye ID, Sun Y. A Survey of Ensemble Learning: Concepts, Algorithms,
696 Applications, and Prospects. IEEE Access. 2022;10:99129–49.
697 <https://doi.org/10.1109/ACCESS.2022.3207287>.

698 19. Zou W, Mwesigwa S, Hosseini S-R, Zhao Z. Rank Aggregation Methods and
699 Tools in Genomic Data Analysis. Curr Genomics. 2025;26:329–40.
700 <https://doi.org/10.2174/0113892029320249240830060611>.

701 20. Yang B, Li J, Li X, Liu S. Gene regulatory network inference based on novel
702 ensemble method. Brief Funct Genomics. 2024;23:866–78.
703 <https://doi.org/10.1093/bfgp/ela036>.

704 21. Shen B, Coruzzi G, Shasha D. EnsInfer: a simple ensemble approach to network
705 inference outperforms any single method. BMC Bioinformatics. 2023;24:114.
706 <https://doi.org/10.1186/s12859-023-05231-1>.

707 22. Cochran K, Srivastava D, Shrikumar A, Balsubramani A, Hardison RC, Kundaje
708 A, et al. Domain-adaptive neural networks improve cross-species prediction of
709 transcription factor binding. Genome Res. 2022;32:512–23.
710 <https://doi.org/10.1101/gr.275394.121>.

711 23. Maetschke SR, Madhamshettiwar PB, Davis MJ, Ragan MA. Supervised,
712 semi-supervised and unsupervised inference of gene regulatory networks. Brief
713 Bioinform. 2014;15:195–211. <https://doi.org/10.1093/bib/bbt034>.

714 24. Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward

715 a central hub for transcription factors and regulatory interactions in plants. *Nucleic*
716 *Acids Res.* 2017;45:D1040–5. <https://doi.org/10.1093/nar/gkw982>.

717 25. Brooks MD, Juang C-L, Katari MS, Alvarez JM, Pasquino A, Shih H-J, et al.
718 ConnecTF: A platform to integrate transcription factor–gene interactions and validate
719 regulatory networks. *Plant Physiol.* 2021;185:49–66.
720 <https://doi.org/10.1093/plphys/kiaa012>.

721 26. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring Regulatory Networks
722 from Expression Data Using Tree-Based Methods. *PLoS ONE.* 2010;5:e12776.
723 <https://doi.org/10.1371/journal.pone.0012776>.

724 27. Iglesias-Martinez LF, De Kegel B, Kolch W. KBoost: a new method to infer gene
725 regulatory networks from gene expression data. *Sci Rep.* 2021;11:15461.
726 <https://doi.org/10.1038/s41598-021-94919-6>.

727 28. Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J,
728 et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory
729 networks. *Bioinformatics.* 2019;35:2159–61.
730 <https://doi.org/10.1093/bioinformatics/bty916>.

731 29. Fan Y, Ma X. Gene Regulatory Network Inference using 3D Convolutional Neural
732 Network. *Proc AAAI Conf Artif Intell.* 2021;35:99–106.
733 <https://doi.org/10.1609/aaai.v35i1.16082>.

734 30. Wang J, Chen Y, Zou Q. Inferring gene regulatory network from single-cell
735 transcriptomes with graph autoencoder model. *PLOS Genet.* 2023;19:e1010942.
736 <https://doi.org/10.1371/journal.pgen.1010942>.

737 31. Gan Y, Yu J, Xu G, Yan C, Zou G. Inferring gene regulatory networks from
738 single-cell transcriptomics based on graph embedding. *Bioinformatics*.
739 2024;40:btae291. <https://doi.org/10.1093/bioinformatics/btae291>.

740 32. De S, Pedersen BS, Kechris K. The dilemma of choosing the ideal permutation
741 strategy while estimating statistical significance of genome-wide enrichment. *Brief*
742 *Bioinform*. 2014;15:919–28. <https://doi.org/10.1093/bib/bbt053>.

743 33. Zhang B, Zhu X, Chen Z, Zhang H, Huang J, Huang J. RiceTFtarget: A rice
744 transcription factor–target prediction server based on coexpression and machine
745 learning. *Plant Physiol*. 2023;193:190–4. <https://doi.org/10.1093/plphys/kiad332>.

746 34. Wang T, Huang S, Tian X, Zhou D, Xing J, Liu W, et al. A β -1,3-glucanase
747 antagonizes a phase-separating WRKY repressor to confer saline–alkaline tolerance in
748 rice. *New Phytol*. 2026;251:3313–31. <https://doi.org/10.1111/nph.71368>.

749 35. Li J, Chen Y, Zhang R, Wang R, Wu B, Zhang H, et al. OsWRKY70 Plays
750 Opposite Roles in Blast Resistance and Cold Stress Tolerance in Rice. *Rice*.
751 2024;17:61. <https://doi.org/10.1186/s12284-024-00741-9>.

752 36. Yang S, Xu K, Chen S, Li T, Xia H, Chen L, et al. A stress-responsive bZIP
753 transcription factor OsbZIP62 improves drought and oxidative tolerance in rice. *BMC*
754 *Plant Biol*. 2019;19:260. <https://doi.org/10.1186/s12870-019-1872-1>.

755 37. Zhang J, Liu L, Xu T, Xie Y, Zhao C, Li J, et al. miRspongeR: an R/Bioconductor
756 package for the identification and analysis of miRNA sponge interaction networks and
757 modules. *BMC Bioinformatics*. 2019;20:235.
758 <https://doi.org/10.1186/s12859-019-2861-y>.

759 38. Van Noort V, Snel B, Huynen MA. The yeast coexpression network has a small-
760 world, scale-free architecture and can be explained by a simple model. *EMBO Rep.*
761 2004;5:280–4. <https://doi.org/10.1038/sj.embor.7400090>.

762 39. Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K,
763 et al. Crosstalk between abiotic and biotic stress responses: a current view from the
764 points of convergence in the stress signaling networks. *Curr Opin Plant Biol.*
765 2006;9:436–42. <https://doi.org/10.1016/j.pbi.2006.05.014>.

766 40. Sun L, Huang L, Hong Y, Zhang H, Song F, Li D. Comprehensive Analysis
767 Suggests Overlapping Expression of Rice ONAC Transcription Factors in Abiotic and
768 Biotic Stress Responses. *Int J Mol Sci.* 2015;16:4306–26.
769 <https://doi.org/10.3390/ijms16024306>.

770 41. De Zelicourt A, Colcombet J, Hirt H. The Role of MAPK Modules and ABA
771 during Abiotic Stress Signaling. *Trends Plant Sci.* 2016;21:677–85.
772 <https://doi.org/10.1016/j.tplants.2016.04.004>.

773 42. Colcombet J, Hirt H. *Arabidopsis* MAPKs: a complex signalling network
774 involved in multiple biological processes. *Biochem J.* 2008;413:217–26.
775 <https://doi.org/10.1042/BJ20080625>.

776 43. Hong W, He C, Wang L, Wang D, Joseph LM, Jantasuriyarat C, et al. *BWMK1*
777 Responds to Multiple Environmental Stresses and Plant Hormones. *J Integr Plant Biol.*
778 2007;49:843–51. <https://doi.org/10.1111/j.1744-7909.2007.00505.x>.

779 44. Tian X, He M, Mei E, Zhang B, Tang J, Xu M, et al. WRKY53 integrates classic
780 brassinosteroid signaling and the mitogen-activated protein kinase pathway to

781 regulate rice architecture and seed size. *Plant Cell*. 2021;33:2753–75.
782 <https://doi.org/10.1093/plcell/koab137>.

783 45. Zhou Y, Xiong Q, Yin C, Ma B, Chen S, Zhang J. Ethylene Biosynthesis,
784 Signaling, and Crosstalk with Other Hormones in Rice. *Small Methods*.
785 2020;4:1900278. <https://doi.org/10.1002/smt.201900278>.

786 46. Zhao H, Yin C, Ma B, Chen S, Zhang J. Ethylene signaling in rice and
787 *Arabidopsis*: New regulators and mechanisms. *J Integr Plant Biol*. 2021;63:102–25.
788 <https://doi.org/10.1111/jipb.13028>.

789 47. Wakuta S, Suzuki E, Saburi W, Matsuura H, Nabeta K, Imai R, et al. OsJAR1 and
790 OsJAR2 are jasmonyl-l-isoleucine synthases involved in wound- and
791 pathogen-induced jasmonic acid signalling. *Biochem Biophys Res Commun*.
792 2011;409:634–9. <https://doi.org/10.1016/j.bbrc.2011.05.055>.

793 48. Wani N, Raza K. Integrative approaches to reconstruct regulatory networks from
794 multi-omics data: A review of state-of-the-art methods. *Comput Biol Chem*.
795 2019;83:107120. <https://doi.org/10.1016/j.compbiolchem.2019.107120>.

796 49. Mummadi ST, Islam MK, Busov V, Wei H. Gene regulatory network prediction
797 using machine learning, deep learning, and hybrid approaches. *For Res*. 2025;5:0–0.
798 <https://doi.org/10.48130/forres-0025-0014>.

799 50. Zhao M, He W, Tang J, Zou Q, Guo F. A comprehensive overview and critical
800 evaluation of gene regulatory network inference technologies. *Brief Bioinform*.
801 2021;22:bbab009. <https://doi.org/10.1093/bib/bbab009>.

802 51. Hegde A, Nguyen T, Cheng J. Machine learning methods for gene regulatory

803 network inference. Brief Bioinform. 2025;26:bbaf470.
804 <https://doi.org/10.1093/bib/bbaf470>.

805 52. Dietterich TG. Ensemble Methods in Machine Learning. In: Multiple Classifier
806 Systems. Berlin, Heidelberg: Springer Berlin Heidelberg; 2000. p. 1–15.
807 https://doi.org/10.1007/3-540-45014-9_1.

808 53. Brown G, Wyatt JL, Tiño P. Managing diversity in regression ensembles. J Mach
809 Learn Res. 2005;6:1621–50.

810 54. Zheng R, Li M, Chen X, Wu F-X, Pan Y, Wang J. BiXGBoost: a scalable, flexible
811 boosting-based method for reconstructing gene regulatory networks. Bioinformatics.
812 2019;35:1893–900. <https://doi.org/10.1093/bioinformatics/bty908>.

813 55. Bai L, Xu L, Ji B, Wang S, Peng S. MRF-XGBLC: Large-scale gene regulatory
814 network inference based on multi-model fusion. In: 2024 IEEE International
815 Conference on Bioinformatics and Biomedicine (BIBM). Lisbon, Portugal: IEEE;
816 2024. p. 485–90. <https://doi.org/10.1109/BIBM62325.2024.10822482>.

817 56. Höllbacher B, Balázs K, Heinig M, Uhlenhaut NH. Seq-ing answers: Current data
818 integration approaches to uncover mechanisms of transcriptional regulation. Comput
819 Struct Biotechnol J. 2020;18:1330–41. <https://doi.org/10.1016/j.csbj.2020.05.018>.

820 57. Yang T-H, Wu W-S. Inferring functional transcription factor-gene binding pairs by
821 integrating transcription factor binding data with transcription factor knockout data.
822 BMC Syst Biol. 2013;7 Suppl 6:S13. <https://doi.org/10.1186/1752-0509-7-S6-S13>.

823 58. Yuan Q, Duren Z. Inferring gene regulatory networks from single-cell multiome
824 data using atlas-scale external data. Nat Biotechnol. 2025;43:247–57.

<https://doi.org/10.1038/s41587-024-02182-7>.

59. Liu D, Chen H, Wang J, Wang Y. MultiGNN: A Graph Neural Network Framework for Inferring Gene Regulatory Networks from Single-Cell Multi-Omics Data. *Computation*. 2025;13:124. <https://doi.org/10.3390/computation13050124>.

60. Albert R, Barabási A-L. Statistical mechanics of complex networks. *Rev Mod Phys*. 2002;74:47–97. <https://doi.org/10.1103/RevModPhys.74.47>.

61. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*. 2004;5:101–13. <https://doi.org/10.1038/nrg1272>.

62. Wang X, Liu X, Su Y, Shen H. Rice Responses to Abiotic Stress: Key Proteins and Molecular Mechanisms. *Int J Mol Sci*. 2025;26:896. <https://doi.org/10.3390/ijms26030896>.

63. Li T, Li B, Wang Y, Xu J, Li W, Chen Z, et al. WRKY Transcription Factors in Rice: Key Regulators Orchestrating Development and Stress Resilience. *Plant Cell Environ*. 2025;48:8388–406. <https://doi.org/10.1111/pce.70124>.

64. Xie Z, Nolan TM, Jiang H, Yin Y. AP2/ERF Transcription Factor Regulatory Networks in Hormone and Abiotic Stress Responses in Arabidopsis. *Front Plant Sci*. 2019;10:228. <https://doi.org/10.3389/fpls.2019.00228>.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.xlsx](#)
- [SupplementaryMaterial1.xlsx](#)
- [SupplementaryMaterial2.xlsx](#)
- [SupplementaryMaterial3.xlsx](#)
- [SupplementaryMaterial4.xlsx](#)
- [SupplementaryMaterial5.xlsx](#)
- [SupplementaryMaterial6.xlsx](#)
- [SupplementaryMaterial7.docx](#)
- [SupplementaryMaterial8.xlsx](#)