

*Weighted Gene Correlation Network Analysis (WGCNA) of Arabidopsis Somatic Embryogenesis (SE) and identification of key gene modules to uncover SE-associated hub genes*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open access

de Silva, K. K. ORCID: <https://orcid.org/0000-0002-8050-5417>, Dunwell, J. M. ORCID: <https://orcid.org/0000-0003-2147-665X> and Wickramasuriya, A. M. ORCID: <https://orcid.org/0000-0003-0249-9733> (2022) Weighted Gene Correlation Network Analysis (WGCNA) of Arabidopsis Somatic Embryogenesis (SE) and identification of key gene modules to uncover SE-associated hub genes. International Journal of Genomics, 2022. 7471063. ISSN 2314-436X doi: <https://doi.org/10.1155/2022/7471063> Available at <https://centaur.reading.ac.uk/106056/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1155/2022/7471063>

Publisher: Hindawi

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

## Research Article

# Weighted Gene Correlation Network Analysis (WGCNA) of *Arabidopsis* Somatic Embryogenesis (SE) and Identification of Key Gene Modules to Uncover SE-Associated Hub Genes

Kithmee K. de Silva <sup>1</sup>, Jim M. Dunwell <sup>2</sup>, and Anushka M. Wickramasuriya <sup>1</sup>

<sup>1</sup>Department of Plant Sciences, Faculty of Science, University of Colombo, Colombo 03, Sri Lanka

<sup>2</sup>School of Agriculture, Policy and Development, University of Reading, Reading RG6 6EU, UK

Correspondence should be addressed to Anushka M. Wickramasuriya; [anushka@pts.cmb.ac.lk](mailto:anushka@pts.cmb.ac.lk)

Received 9 February 2022; Accepted 23 May 2022; Published 4 July 2022

Academic Editor: Marco Gerdol

Copyright © 2022 Kithmee K. de Silva et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Somatic embryogenesis (SE), which occurs naturally in many plant species, serves as a model to elucidate cellular and molecular mechanisms of embryo patterning in plants. Decoding the regulatory landscape of SE is essential for its further application. Hence, the present study was aimed at employing Weighted Gene Correlation Network Analysis (WGCNA) to construct a gene coexpression network (GCN) for *Arabidopsis* SE and then identifying highly correlated gene modules to uncover the hub genes associated with SE that may serve as potential molecular targets. A total of 17,059 genes were filtered from a microarray dataset comprising four stages of SE, i.e., stage I (zygotic embryos), stage II (proliferating tissues at 7 days of induction), stage III (proliferating tissues at 14 days of induction), and stage IV (mature somatic embryos). This included 1,711 transcription factors and 445 *EMBRYO DEFECTIVE* genes. GCN analysis identified a total of 26 gene modules with the module size ranging from 35 to 3,418 genes using a dynamic cut tree algorithm. The module-trait analysis revealed that four, four, seven, and four modules were associated with stages I, II, III, and IV, respectively. Further, we identified a total of 260 hub genes based on the degree of intramodular connectivity. Validation of the hub genes using publicly available expression datasets demonstrated that at least 78 hub genes are potentially associated with embryogenesis; of these, many genes remain functionally uncharacterized thus far. *In silico* promoter analysis of these genes revealed the presence of *cis*-acting regulatory elements, “soybean embryo factor 4 (SEF4) binding site,” and “E-box” of the *napA* storage-protein gene of *Brassica napus*; this suggests that these genes may play important roles in plant embryo development. The present study successfully applied WGCNA to construct a GCN for SE in *Arabidopsis* and identified hub genes involved in the development of somatic embryos. These hub genes could be used as molecular targets to further elucidate the molecular mechanisms underlying SE in plants.

## 1. Introduction

The ability to produce embryos from undifferentiated somatic cells *in vitro* is a unique developmental pathway found within the plant kingdom. Since the first report of somatic embryo induction from callus cells of carrot [1, 2], this developmental pathway based on cellular totipotency has been studied extensively due to its biological and scientific significance; it has been recognized as a model system for studying early plant embryogenesis. Until now, most studies have focused on the mechanism of somatic embryo

development at the morphological level [2–4] or the development of optimized protocols for the generation of somatic embryos from a range of explants [5–8].

Somatic embryogenesis (SE) involves a complex signaling network [9]; transcriptional regulation of a set of genes in response to stress caused by plant growth regulators, nutrients, certain stress conditions, and other signaling elements triggers cellular reprogramming and transformation of somatic cells into embryos [10, 11]. In 2007, Zeng et al. [12] developed the first draft gene regulatory network for early SE employing a set of transcriptionally regulated SE-

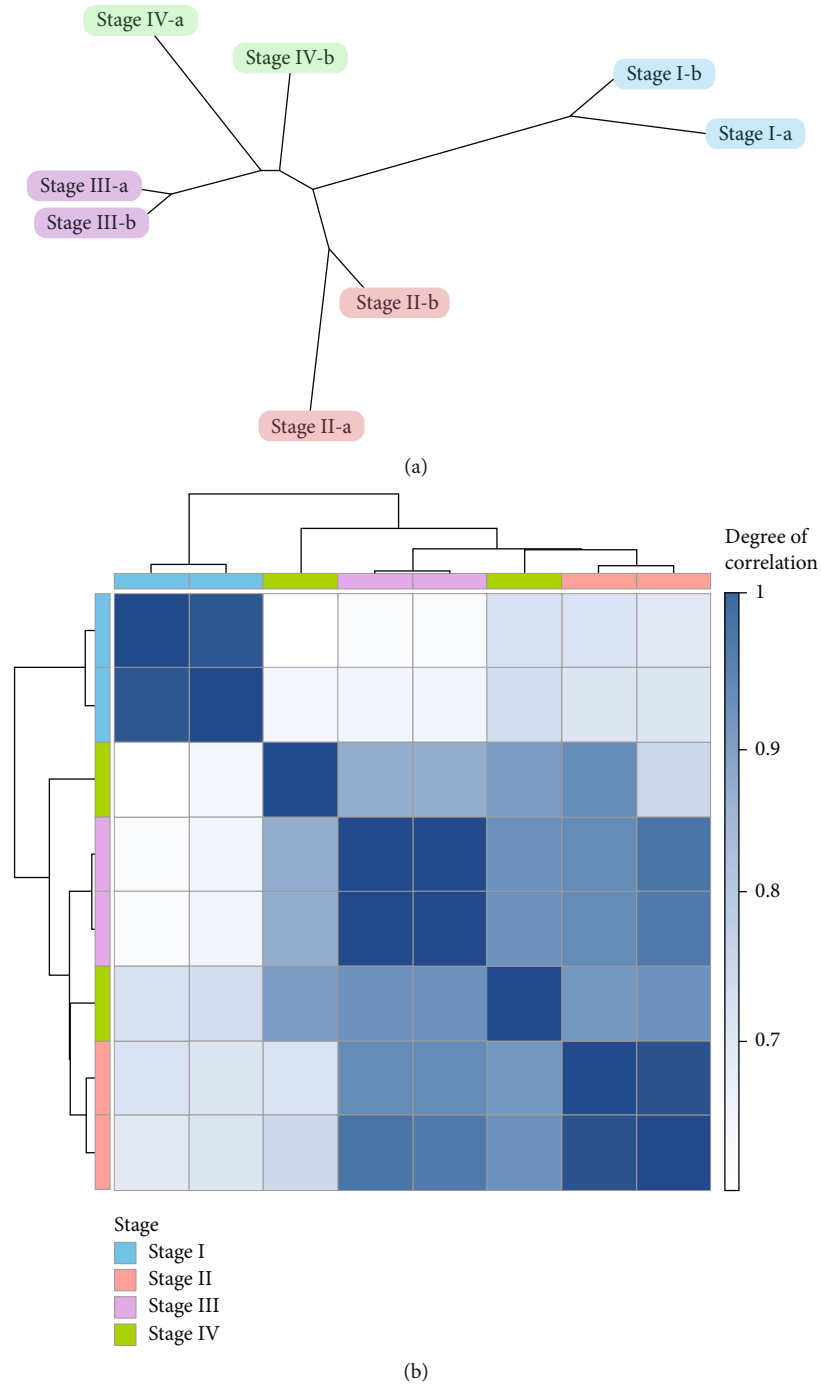


FIGURE 1: Hierarchical clustering of somatic embryo transcriptomes based on their Euclidean distance using average linkage clustering (replicates of each stage are labeled as “a” and “b”). (a) Unrooted hierarchical clustering dendrogram (the length between nodes corresponds to the distance between samples). (b) Hierarchical clustering heatmap visualizing the correlations between the samples.

related genes in cotton. Although a set of genes have been identified as markers for the initiation phase of SE [13, 14], for example, *SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 (SERK1)* [15, 16], *LEAFY COTYLEDON (LEC)* [17–21], *BABY BOOM (BBM)* [22], and *WUSCHEL (WUS)* [18, 23], the current scientific knowledge on the underlying regulatory landscape of SE is limited. The use of transcriptomics has uncovered a large number of differen-

tially expressed genes (DEGs) during SE in many crops, including *Arabidopsis* [24], rice [25], bread wheat [26], cotton [27], maize [28], and coconut [29]. However, the functions of many of these genes in SE are still not understood.

Gene coexpression networks (GCNs) are increasingly used to understand the interactions among a set of transcriptionally regulated genes. There are many types of coexpression networks: signed/unsigned coexpression networks

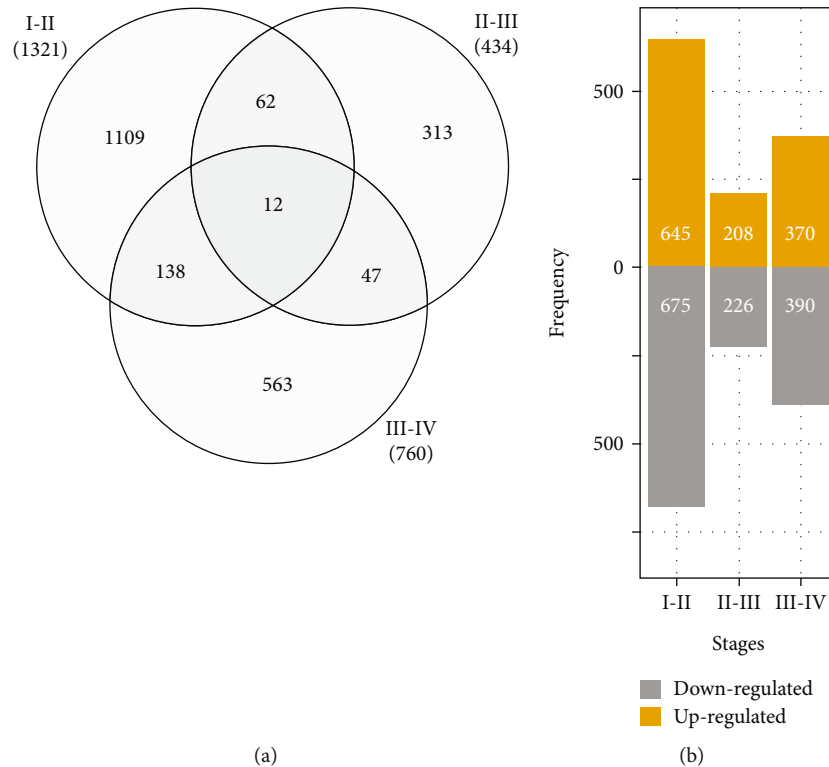


FIGURE 2: Overview of DEGs. (a) Distribution of DEGs between consecutive somatic embryo developmental stages. (b) Number of up- or downregulated DEGs between stages.

and weighted/unweighted coexpression networks [30]. In the present study, we have focused on weighted network construction as it is likely to produce more robust findings than unweighted networks [31]. Weighted Correlation Network Analysis (WGCNA) is one of the most popular clustering packages for GCN analysis [31, 32] and the first tool to be employed to construct GCNs from RNA-sequencing (RNA-seq) data. This coexpression tool is easy to use and can be used to find clusters (modules) of highly correlated genes and to identify biologically relevant associations between phenotypes/sample traits and modules from expression data [30]. Recently, WGCNA has been effectively used to identify stage-specific gene expression clusters associated with key stages of *Arabidopsis* zygotic embryo development [33]. In addition, this approach has been successfully used to discover the regulatory landscape of SE in rice [25] and several other biological pathways in plants [34–36]. Here, we have analyzed a transcriptome dataset covering four somatic embryo developmental stages in *Arabidopsis* using WGCNA to understand better the system-level functionality of the transcriptionally regulated genes in dicot SE.

## 2. Materials and Methods

**2.1. Data Collection and Gene Filtering.** The transcriptome data covering somatic embryo developmental stages of wild-type *Arabidopsis* were retrieved from the National Centre for Biotechnology Information (NCBI) GEO database (GEO accession: GSE48915) [37]. The dataset consisted of four developmental stages (zygotic embryos, proliferating tissues

at 7 days of induction, proliferating tissues at 14 days of induction, and mature somatic embryos) with two replicates for each stage. Subsequently, the genes with variance greater than the second quartile of variance were filtered to eliminate low-expressed or nonvarying genes, and the remaining genes were used in GCN analysis (<https://horvath.genetics.ucla.edu/html/coexpressionnetwork/rpackages/wgcna/faq.html>, accessed on 11 May 2022). In addition, DEGs between consecutive embryonic stages were identified by calculating the fold change (FC) in gene expression through a simple *t*-test. Arbitrary FC cutoff of  $|\log_2 \text{FC}| \geq 2.0$  and *p* value of  $<0.05$  were used to reduce false discoveries.

**2.2. GCN Construction.** “WGCNA” package in R software [32] was employed to identify significant gene modules and hub genes in *Arabidopsis* somatic embryo transcriptomes. A gene coexpression similarity matrix was constructed between the expression profiles of the filtered genes using the Pearson correlation. The similarity matrix was then transformed into an adjacency matrix where each entry encodes the connection strength between each pair of genes (“nodes”). The adjacency matrix defines a measure of node dissimilarity from which the nodes (genes) are clustered into network modules. Consequently, the GCN was developed using the automatic one-step network construction and module detection method with the following parameters:

```
net = blockwiseModules (data, power = 15, corType = "pearson", networkType =
"signed", TOMType = "signed", minModuleSize = 30, mergeCutHeight =
0.25)
```

The soft threshold value (power parameter) was decided by the scale-free topology fit index curve.

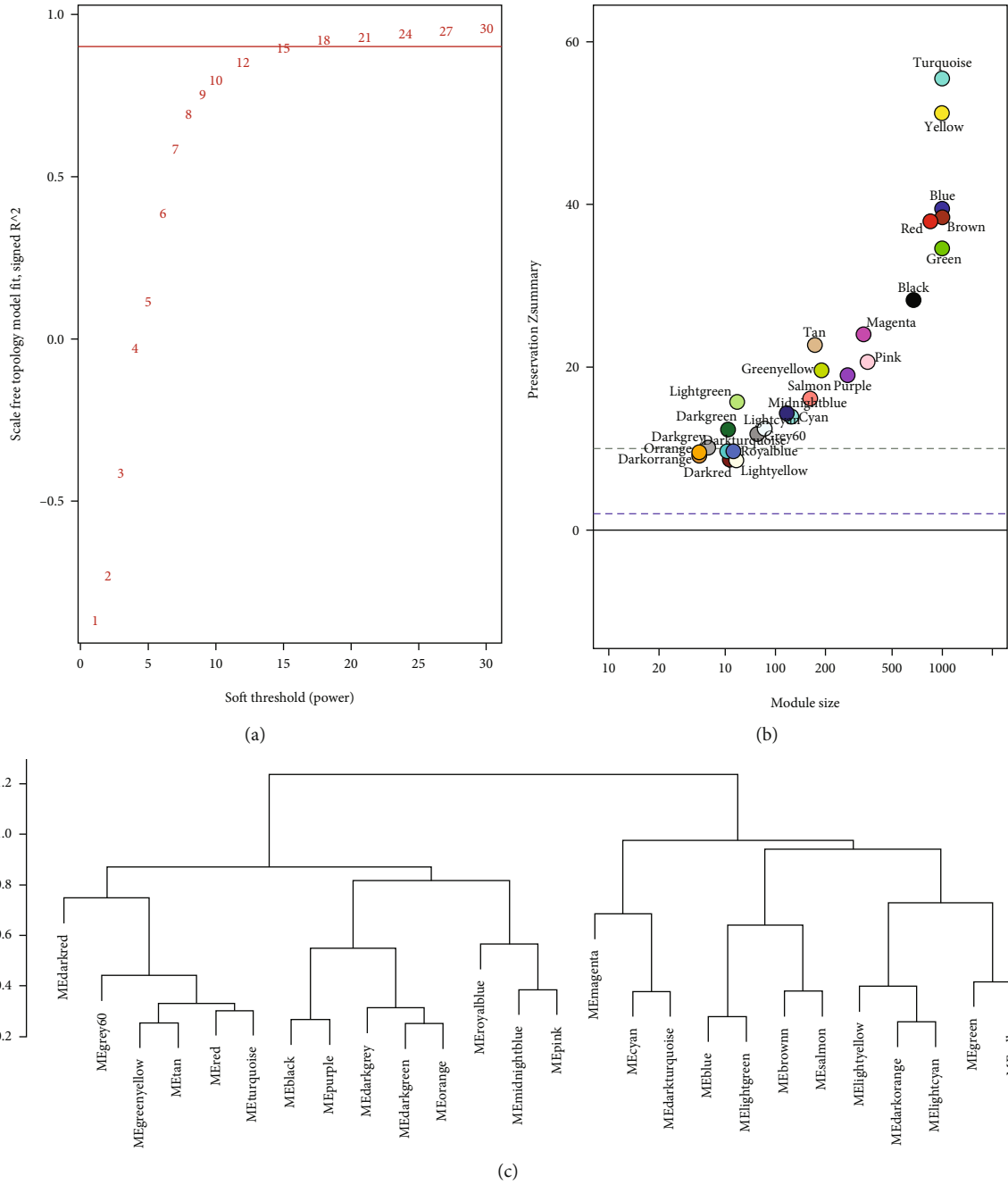


FIGURE 3: Construction of the draft GCN for SE. (a) Network topology for different soft-thresholding powers. (b) Module preservation statistics. (c) Hierarchical clustering dendrogram of MEs.

2.3. *GCN Visualization.* The constructed modular networks were exported to Cytoscape (version 3.7.2) for visualization; gene correlations with  $p$  value  $< 0.05$  were filtered as significant gene correlations and visualized. The modular networks were analyzed by the “network analyzer” tool in Cytoscape for a concise and informative representation of nodes and edges.

2.4. *Validation of Network Modules.* The robustness of the coexpression modules was assessed through module preservation and quality statistics, which were computed using the

modulePreservation function in the WGCNA package [38]. The adjacency matrix of the network was taken as the reference, and the dataset was selected as test data with 200 permutations ( $nPermutations = 200$ ). The stability of the modules was tested through the statistics median rank and  $Z_{summary}$ .

2.5. *Inferring Module-Stage Relationships.* Module-stage relationships of the GCN were evaluated through module eigengenes (MEs). The correlation relationships between the MEs and different somatic embryo developmental stages



(a)

FIGURE 4: Continued.

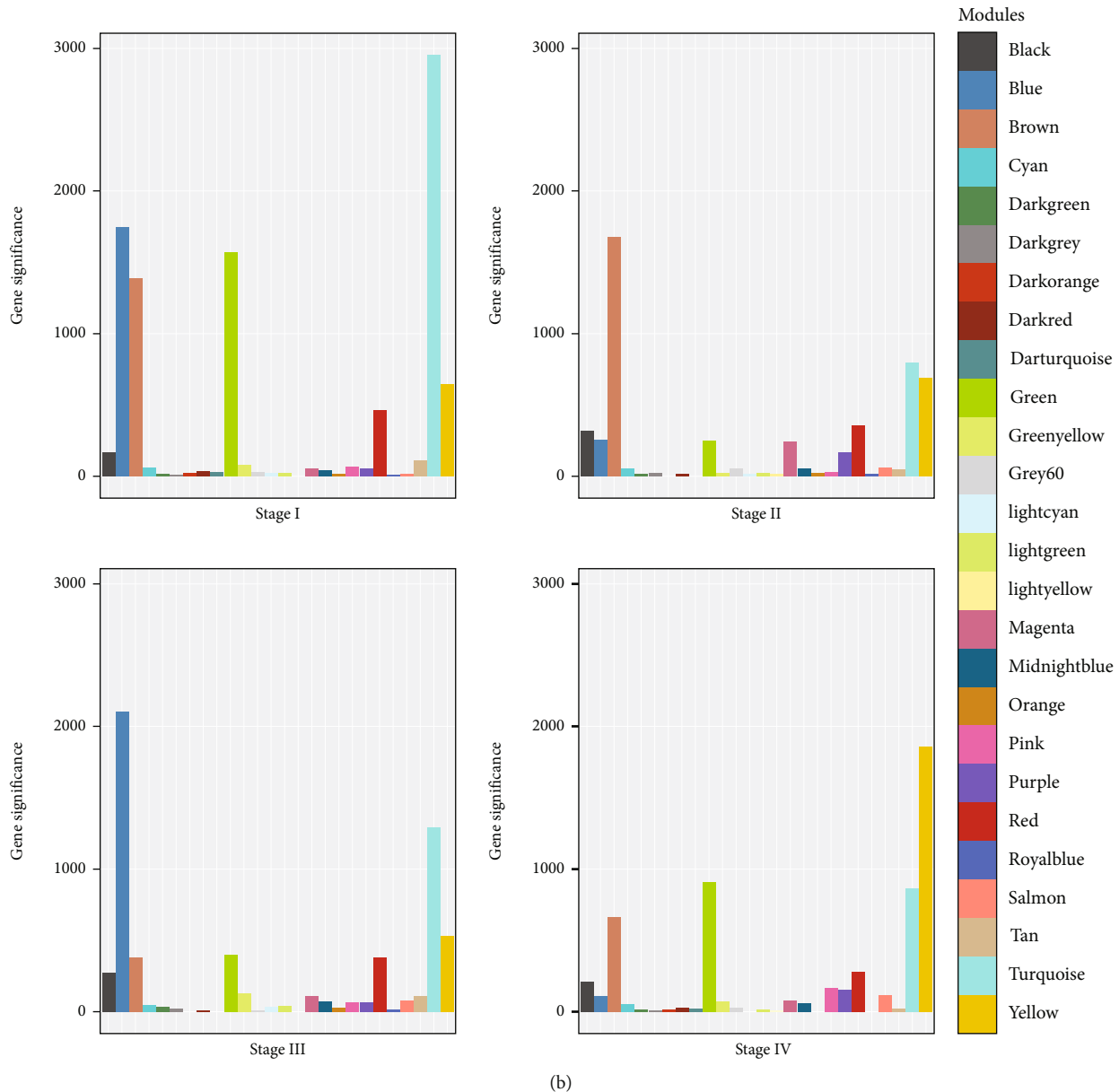


FIGURE 4: Stage-specific gene modules detected by WGCNA. (a) Module-trait relationship heatmap. Each row corresponds to a module, and each column corresponds to a stage. The degree of correlation is illustrated with the colour legend. The numbers in the table correspond to the  $p$  value. Modules that are significantly associated with each somatic embryo development stage ( $|r| > 0.8$  and  $p$  value  $\leq 0.01$ ) are indicated by an asterisk. (b) Gene significance values of coexpression modules related to different somatic embryo developmental stages.

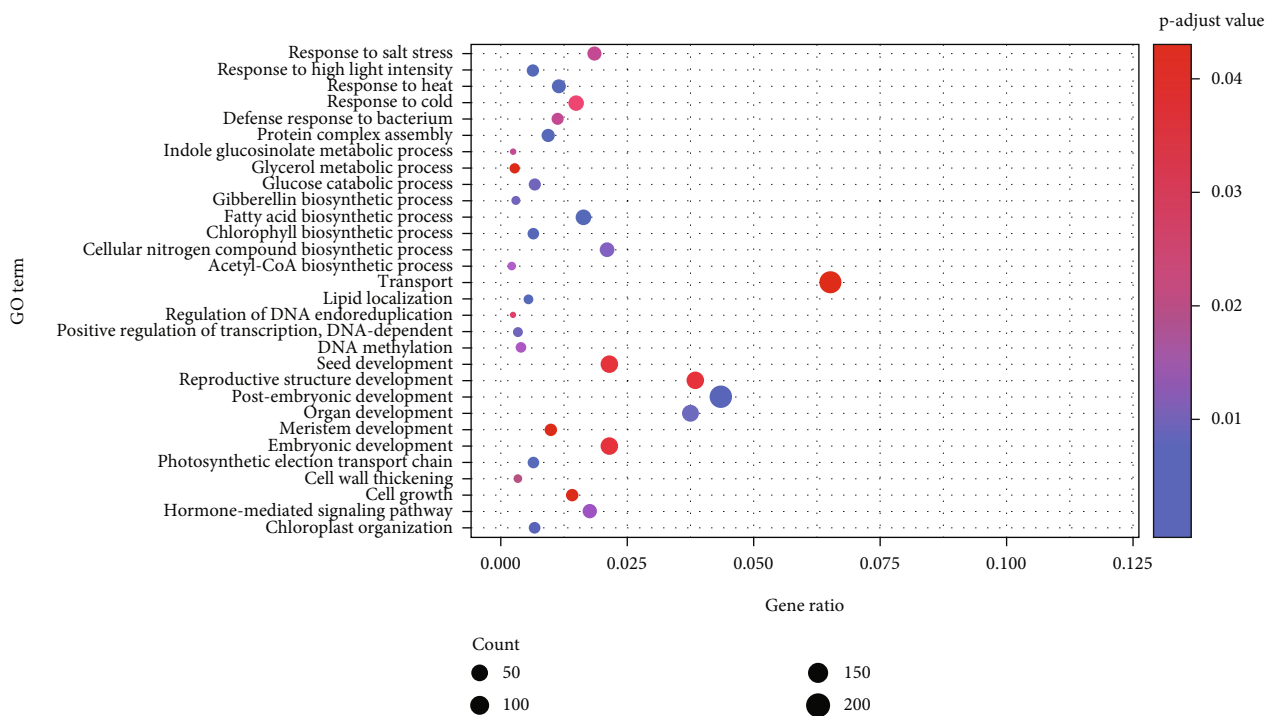
were analyzed and visualized through a heatmap. Gene significance was calculated based on the  $p$  value of the linear regression between the gene expression profile and the associated developmental stage.

**2.6. Functional Enrichment Analysis.** Functional enrichment analysis was performed to detect enriched biological processes in gene modules. Gene Ontology (GO) terms enriched in each module were elucidated using the “singular enrichment analysis” tool provided by agriGO v2.0 [39]. “Arabidopsis genome locus (TAIR10)” was used as the reference, and all other parameters were set as the default for the analysis. Overrepresented GO terms in each network module

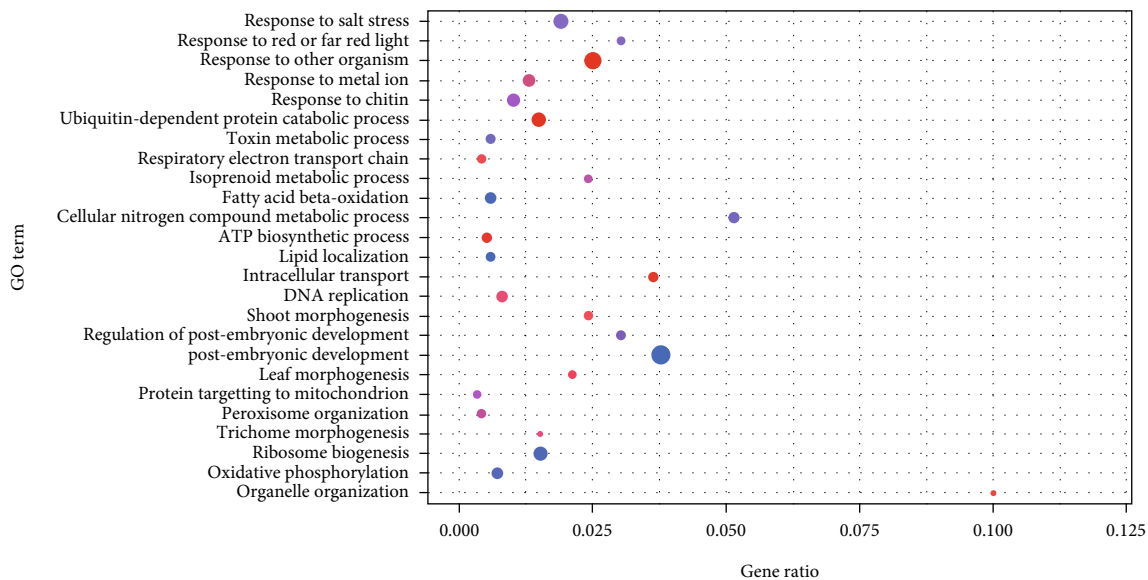
were identified using the hypergeometric test. To further explore the DEGs mapped to each gene module, the distribution of the following genes across modules was studied: SE-related marker genes [40], plant transcription factors (TFs) (<http://plantfdb.cbi.pku.edu.cn/index.php>), EMBRYO DEFECTIVE (*EMB*) genes [41], and gene encoding epigenetic regulators [42, 43].

**2.7. Identification and Validation of Hub Genes.** Genes in each module were arranged based on gene connectivity. The top 10 genes of each module were considered as hub genes. The transcriptome dataset published by Wickramasuriya and Dunwell in 2015 was retrieved from the





(a)



(b)

FIGURE 5: Continued.

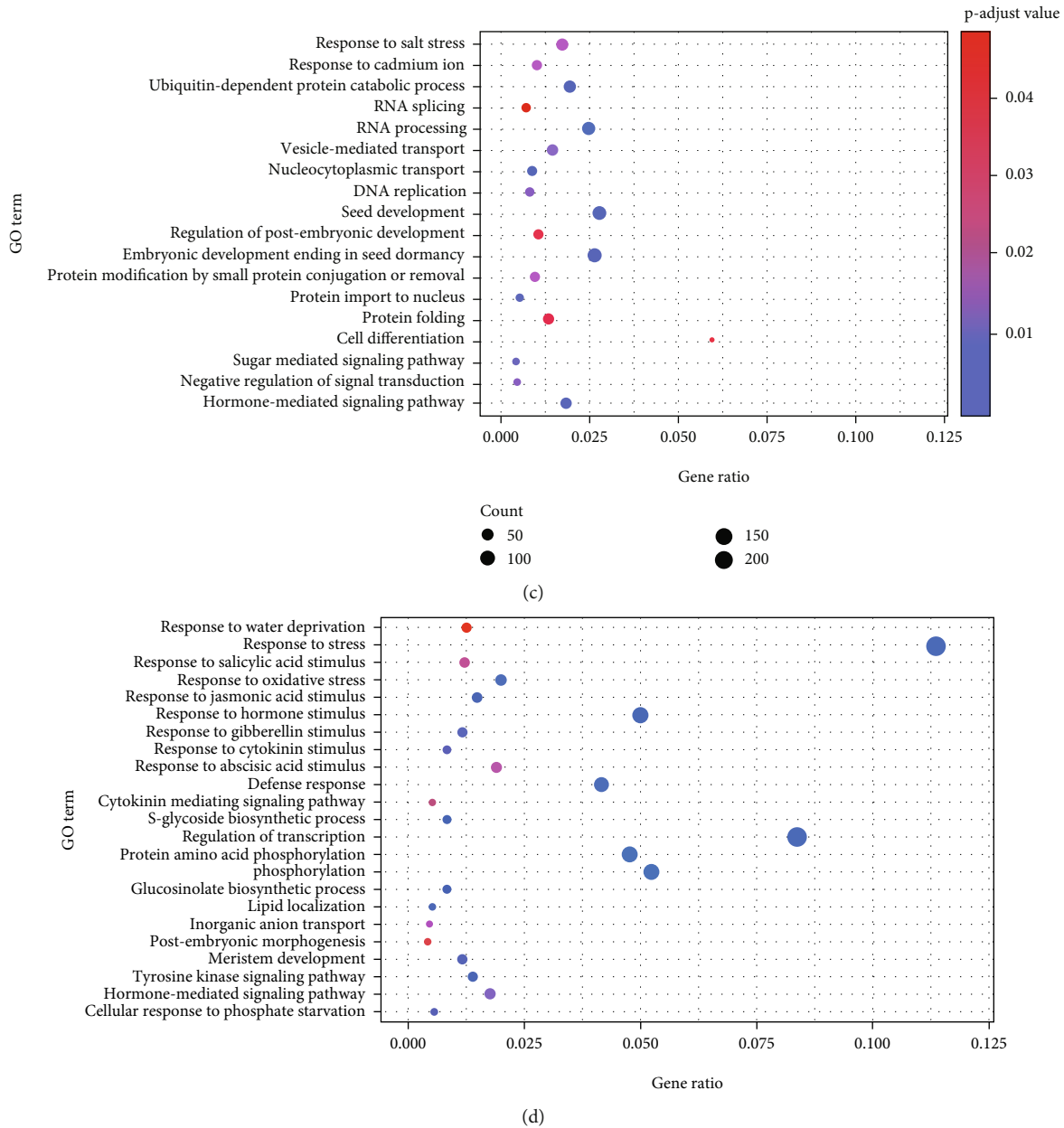


FIGURE 5: Functional enrichment analysis of the stage-specific modules. (a)–(d) represent the significantly enriched GO terms ( $p$  value  $< 0.05$ ) in modules specifically associated with stages I–IV, respectively.

ArrayExpress database (E-MTAB-2403) [24] to study the expression of hub genes during SE.

**2.8. In Silico Analysis of Hub Genes.** The promoter sequences of hub genes (1000 bp upstream from the transcription start site) were retrieved from “The Arabidopsis Information Resource” (TAIR) database and analyzed using the Multiple Em for Motif Elicitation (MEME) tool in the MEME Suite 5.3.3 [44]. The following parameters were used in the analysis: number of motifs: 10; motif site distribution: zero or once per occurrence (ZOOOPS); minimum width: 6; maximum width: 50; and background model: zero-order model of sequences. Further, the biological significance of the predicted MEME motifs was investigated using the Gene Ontol-

ogy for MOTifs (GOMo) version 5.3.3 [45] provided in the MEME Suite. Additionally, the retrieved promoter sequences were searched against the Plant *cis*-acting regulatory DNA elements (PLACE) database to identify overrepresented *cis*-acting regulatory elements (CREs; [46]).

### 3. Results

**3.1. Hierarchical Clustering of Somatic Embryo Transcriptomes.** In the present study, transcriptome datasets generated through microarray experiments were retrieved from the NCBI covering four somatic embryo developmental stages (with two replicates for each stage), referred to herein as stages I (zygotic embryos), II (proliferating tissues

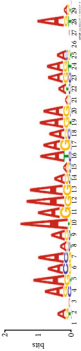
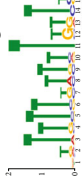

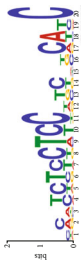
TABLE 1: Top 10 hub genes ordered by the degree of connectivity.

Gene identifier	Degree of connectivity	Gene module	Gene name	Description
AT1G27120	3327	Turquoise	<i>AT1G27120 (GALT4)</i>	Galactosyltransferase family protein
AT5G52820	2853	Blue	<i>NOTCHLESS (NLE)</i>	WD-40 repeat family protein/notchless protein
AT5G56090	2348	Brown	<i>CYTOCHROME C OXIDASE 15 (COX15)</i>	Encodes a homolog of COX15
AT2G43100	2134	Yellow	<i>ISOPROPYLMALATE ISOMERASE 2 (IPMI2)</i>	Isopropylmalate isomerase 2
AT1G71010	1958	Green	<i>FORMS APLOID AND BINUCLEATE CELLS 1C (FAB1C)</i>	Encodes a protein that is predicted to act as a phosphatidylinositol-3P 5-kinase but lacks a FYVE domain
AT2G29890	806	Red	<i>VILLIN 1 (VLN1)</i>	Encodes a ubiquitously expressed villin-like protein
AT2G45600	569	Black	<i>AT2G45600</i>	Alpha/beta-hydrolases superfamily protein
AT1G72400	321	Magenta	<i>AT1G72400</i>	Hypothetical protein
AT3G53980	311	Pink	<i>AT3G53980</i>	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
AT5G65350	243	Purple	<i>HISTONE 3 11 (HTR11)</i>	Histone 3 11
AT5G27560	182	Green-yellow	<i>AT5G27560</i>	DUF1995 domain protein, putative (DUF1995)
AT5G54855	170	Tan	<i>AT5G54855</i>	Pollen Ole e 1 allergen and extensin family protein
AT1G75630	156	Salmon	<i>VACUOLAR H+-PUMPING ATPASE 16 KDA PROTEOLIPID SUBUNIT 4 (AVAP4)</i>	Vacuolar H+-pumping ATPase 16 kD proteolipid (ava-p) mRNA
AT1G74450	107	Cyan	<i>AT1G74450</i>	BPS1-like protein (DUF793)
AT2G23940	107	Midnight-blue	<i>AT2G23940</i>	Transmembrane protein (DUF788)
AT1G30460	82	Light-cyan	<i>CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR 30 (CPSF30)</i>	Encodes AtCPSF30, the 30-KDa subunit of cleavage and polyadenylation specificity factor
AT1G06040	74	Grey60	<i>SALT TOLERANCE (STO)</i>	B-box zinc finger family protein that encodes a salt tolerance protein
AT2G11560	58	Light-green	<i>AT2G11560</i>	Mutator-like transposase/similar to MURA transposase of maize
AT3G55050	50	Dark-green	<i>D-CLADE TYPE 2C PROTEIN PHOSPHATASE 4 (PP2C.D4)</i>	Protein phosphatase 2C family protein
ATCG01070	48	Dark-red	<i>NAD(P)H-QUINONE OXIDOREDUCTASE SUBUNIT 4L (NDHE)</i>	NADH dehydrogenase ND4L
AT3G25950	48	Royal-blue	<i>AT3G25950</i>	TRAM, LAG1, and CLN8 (TLC) lipid-sensing domain containing protein
AT1G65410	39	Dark-turquoise	<i>ATP-BINDING CASSETTE I13 (ABCI13)</i>	Encodes a member of NAP subfamily of transporters
AT5G02310	38	Light-yellow	<i>PROTEOLYSIS 6 (PRT6)</i>	Encodes a component of the N-end rule pathway that targets protein degradation
AT3G61130	37	Dark-grey	<i>GALACTURONOSYLTRANSFERASE 1 (GAUT1)</i>	Encodes a protein with putative galacturonosyltransferase activity
AT3G53350	34	Dark-orange	<i>ROP INTERACTIVE PARTNER 3 (RIP3)</i>	Encodes a microtubule-binding protein
AT5G43490	34	Orange	<i>AT5G43490</i>	Myb-like protein X

at 7 days of induction), III (proliferating tissues at 14 days of induction), and IV (mature somatic embryos). The hierarchical clustering of samples (Figure 1(a)) confirmed that the sample replicates of each stage have a higher degree of correlation with each other than with other developmental stages; sample outliers were not detected in the dataset. The clustering heat-

map clearly distinguished four discrete clusters of related expression patterns corresponding to the stages of somatic embryo development (Figure 1(b)). Further, stage I showed a poor correlation with the other three stages. This suggests that stage I may have a distinct expression profile as compared to other somatic embryo developmental stages.

TABLE 2: Conserved motifs identified in the promoter regions of hub genes using the MEME tool.

Motif	E-value	Motif width	Sites	Significant GO enriched terms (q value<0.05)
<p>NAVAAAAAAAAARARARAAAAAAHMAA</p> <p>Consensus sequence:                      [AGT]A[AG]AA[AC]AAAA[AG]A[AG][AG][AGC]A                      [AG]AA[AG][AG]A[AG]AA[ATC][AC]A[AGT]</p> <p>Consensus logo:  </p>	2.8e-077	29	229	(i) GO:0042023: DNA endoreduplication (ii) GO:0009944: Polarity specification of adaxial/abaxial axis (iii) GO:0009735: Response to cytokinin stimulus (iv) GO:0009744: Response to sucrose stimulus (v) GO:0006468: Protein amino acid phosphorylation (vi) GO:0009965: Leaf morphogenesis
<p>DTTTTITKTTTKITY</p> <p>Consensus sequence:                      [AGT][TG]TTTT[TG]TTTT[TG][TG]T[TCG]</p> <p>Consensus logo:  </p>	6.0e-020	15	245	
<p>NDRRAGDDRRWARRRRARAGAADRRDAG</p> <p>Consensus sequence:                      [AGC][GAT][AG][AG][AT][GA][AGT][ATG][AGT][GA][AG]                      [AT]A[GA][AG][GAC]A[AG]A[GA]A[AG][GAT][AG][AG]                      [GTA][AT][GT]</p> <p>Consensus logo:  </p>	3.6e-025	28	121	(i) GO:0009944: Polarity specification of adaxial/abaxial axis (ii) GO:0048481: Ovule development (iii) GO:0010050: Vegetative phase change (iv) GO:0010051: Xylem and phloem pattern formation (v) GO:0042023: DNA endoreduplication (vi) GO:0035196: Production of miRNAs involved in gene silencing by miRNA (vii) GO:0009740: Gibberellic acid mediated signaling pathway
<p>CMAYCTYTCDDTCHBCATC</p> <p>Consensus sequence:                      [CT][CA]A[TC]CT[CT]C[TAG][TGC][CT][ATC][GTC]C[AT][TC]C</p> <p>Consensus logo:  </p>	5.3e-007	20	44	

**3.2. Filtering of Genes for the GCN Construction and Downstream Analysis.** As recommended by Langfelder and Horvath [32], genes were filtered by the variance for the GCN construction; filtering genes for variance greater than 0.25 quantile identified a total of 17,059 genes (see Table S1). This included 445 *EMB* genes [41], 10 SE marker genes [40], and 1,711 *Arabidopsis* TFs (65.3%).

In addition, DEGs were identified by a pairwise ratio of expression between consecutive stages of development. A total of 2,244 genes were identified by threshold filtering based on  $|\log_2 FC| \geq 2.0$  and  $p$  value  $< 0.05$ . 64 *EMB* genes [41], four SE marker genes [40], and 458 TFs were present within the DEGs identified (see Table S2). A total of 12 genes including the genes *STRESS INDUCED FACTOR 2* (AT1G51850), *LIGHT-HARVESTING-LIKE 3:1* (AT4G17600), *BETA GLUCOSIDASE 28* (AT2G44460), *FERREDOXIN C 1* (AT4G14890), and *PHOTOSYSTEM II SUBUNIT Q* (AT4G05180) were differentially expressed throughout SE (Figure 2(a)). In addition, a considerable number of genes were up- and downregulated during early embryo developmental stages (Figure 2(b)).

**3.3. Construction of GCN.** The expression profiles of the filtered 17,059 genes were used to construct a scale-free gene expression network with a soft threshold of 15 (Figure 3(a)). The dynamic hierarchical clustering approach integrated with the WGCNA pipeline distinguishes groups of genes with coexpression patterns and clusters them into network modules. In total, 26 distinct coexpression gene modules were detected with the module size ranging from 35 to 3,418 genes (Figures 3(b) and 3(c)); each module was assigned with a unique colour. The module comprising most genes was the turquoise (3,418 genes) followed by the blue (2,973 genes) and brown (2,437 genes) (Figure 3(b)). The expression profiles of coexpressed genes clustered in each module were summarized as MEs. Among the filtered genes, 13 genes that failed to fit within a distinct group were assigned to the grey module and removed from the downstream analysis. Module preservation analysis indicated high module preservation, confirming that the modules generated here can also be found in diverse independent datasets (Figure 3(b)). Each module was exported and visualized using Cytoscape.

**3.4. Identification of Stage-Related Modules.** The relationships between the gene modules and different somatic embryo developmental stages were determined by assessing the Pearson correlation coefficient ( $r$ ) between the MEs and developmental stages. Module-trait correlation analyses revealed that multiple modules are related to SE (Figure 4(a)). A total of 18 modules were significantly associated with the somatic embryo developmental stages ( $|r| > 0.8$  and  $p$  value  $\leq 0.01$ ; Figure 4), and these modules were “stage-specific,” i.e., the module was significantly associated with only one particular developmental stage of SE: tan, turquoise, dark-orange, and green to stage I; grey60, magenta, brown, and light-yellow to stage II; green-yellow, dark-gray, dark-green, orange, blue, light-green, and light-cyan to stage III; and pink, dark-turquoise, salmon, and yellow to stage IV.

Gene significance, the correlation between modular gene expression and each stage, is shown in Figure 4(b).

**3.5. Functional Enrichment Analysis of “Stage-Specific” Gene Modules.** GO enrichment analysis performed on “stage-specific” modules showed that the genes in green and turquoise modules which exhibited a significant association with stage I were mainly enriched in the biological processes being involved in postembryonic development, hormone-mediated signaling pathway, biosynthesis pathways (sterol and fatty acids), DNA methylation, and transcription regulation (Figure 5(a)). Genes in brown, light-yellow, and magenta modules, which showed significant association with stage II, were mainly enriched in the biological processes involved in root and shoot development, ATP synthesis, response to the metal ions, and DNA replication (Figure 5(b)), whereas genes in blue and light-cyan modules, which showed significant association with stage III, were enriched for the biological processes involved in transition postembryonic and seed development, hormone- and sugar-mediated signaling pathways, cell differentiation, protein modification, and RNA processing (Figure 5(c)). Moreover, the yellow module, which showed a significant relationship to stage IV, was mainly enriched in biological processes involved in ion transport, postembryonic development, signal transduction, lipid localization, response to oxidative and water stress, as well as response to phytohormones (abscisic acid, gibberellin, cytokinin, and jasmonic acid) (Figure 5(d)).

**3.6. Analysis of Hub Genes.** Hub genes are nodes in a network often hypothesized to be functionally significant due to their high degree of intramodular connectivity. A total of 260 genes (top 10 genes of each module with high connectivity) were identified as potential hub genes; the hub gene with the highest degree of connectivity in each module is given in Table 1 (the complete list of hub genes is given in Table S3). GO enrichment analysis of the hub genes revealed that they are mainly enriched for biological processes such as metabolic processes (mRNA and cellular amino acid), oxidation-reduction, protein folding, and postembryonic development.

Among the hub genes, only 234 genes were functionally annotated; of these, 13 were TFs: *AUXIN RESPONSE FACTOR 9* (ARF9), *FLOWERING BHLH 4* (FBH4), *BASIC HELIX-LOOP-HELIX 39* (BHLH39), *BASIC LEUCINE-ZIPPER 44* (bZIP44), *bZIP19*, *ZIM-LIKE 2* (ZML2), *AT5G60820*, *AT4G01270*, *KANADI 3* (KAN3), *HOMEODOMAIN GLABROUS 4* (HDG4), *CELL DIVISION CYCLE 5* (CDC5), *NAC DOMAIN CONTAINING PROTEIN 80* (NAC080), and *SALT TOLERANCE* (STO)). In addition, five genes encoding transposable elements (i.e., *AT2G11560*, *AT3G33066*, *AT5G32430*, *AT3G42820*, and *AT4G28900*) were identified.

*In silico* analysis of the promoter sequences (1000 bp upstream from the transcription start site) of the hub genes using the MEME tool identified four significant motifs ranging in length from 15 to 29 bp (Table 2). Motifs 1, 2, and 3



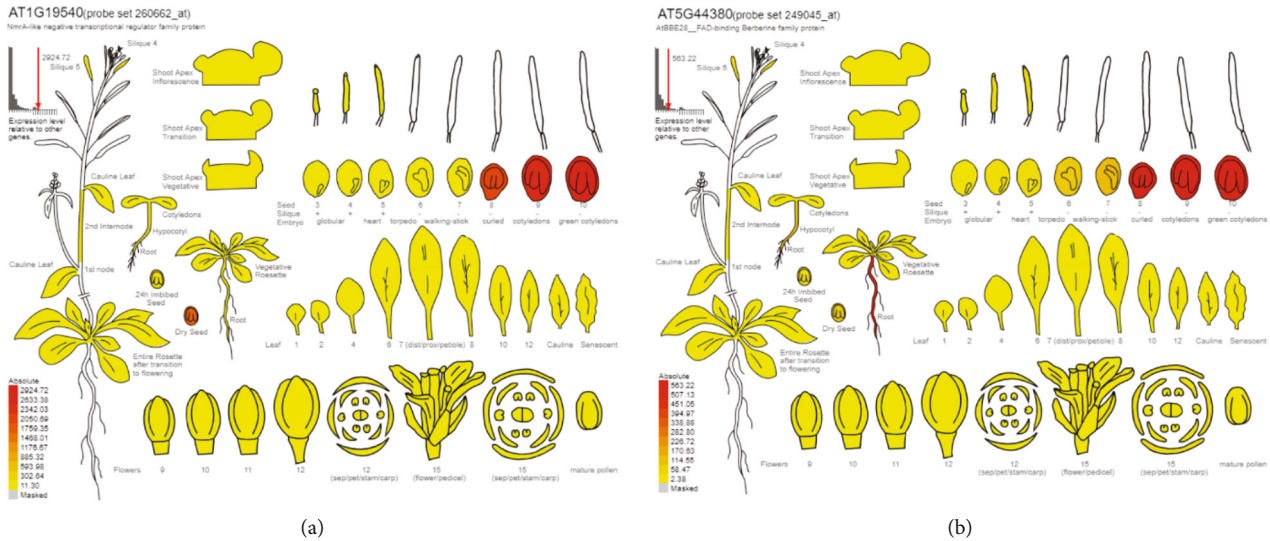


FIGURE 6: Expression patterns of two hub genes, *AT1G19540* (a) and *AT5G44380* (b), when viewed through the *Arabidopsis* eFP browser. The normalized expression value for each gene is colour-coded as indicated by the legend.



FIGURE 7: Transcript abundance extracted from the somatic embryo transcriptome dataset, E-MTAB-2465 for the hub genes. The hub genes significantly upregulated ( $\log_2 FC \geq 2.0$ ) and downregulated ( $\log_2 FC \leq -2.0$ ) in somatic embryonic tissues compared to leaf tissues are indicated with yellow asterisks and red diamonds, respectively. Transcript abundances are shown in fragments per kilobase of transcript per million mapped reads (FPKM).

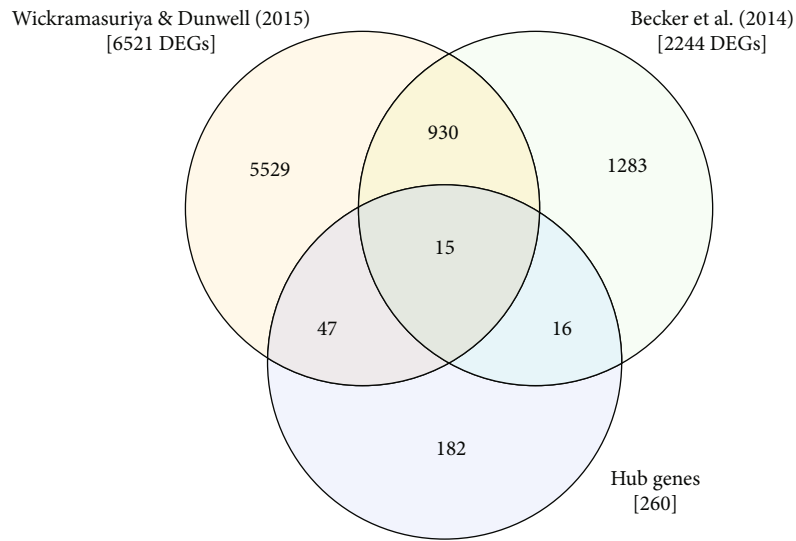


FIGURE 8: Venn diagram indicating the intersection of hub genes and DEGs ( $|\log_2 FC| \geq 2.0$  and  $p$  value  $< 0.05$ ) obtained from E-MTAB-2465 [24] and GSE48915 [37].

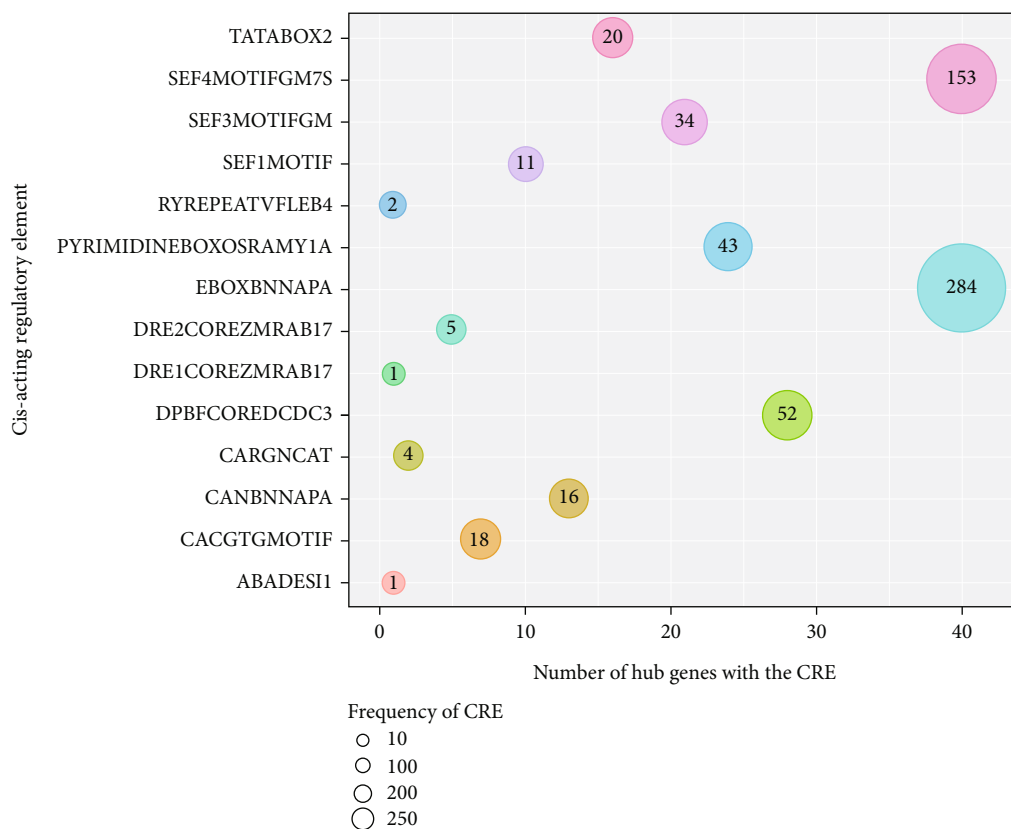


FIGURE 9: The distribution of several important plant CREs present in the promoter regions of functionally uncharacterized hub genes. The number of hub genes that contain the relevant CRE in their promoter region is indicated by the  $x$ -axis. The size of the circle depicts the occurrence of the CREs within the promoter regions of the hub genes (as indicated within the circle).

were detected across 229, 245, and 121 hub genes, respectively. Further analysis of the predicted motifs using the GOMo tool provided in the MEME suite indicated that motifs 1 and 3 may be involved in the DNA endoreplication,

polarity specification of axial/abaxial axis, and hormone-mediated signaling pathways; motifs 1 and 3 seem to function in association to cytokinin and gibberellic acid, respectively.

TABLE 3: Functional roles of several important CREs detected in the functionally uncharacterized hub gene promoter sequences retrieved from the PLACE database [46].

Cis-acting regulatory element	Function*
ABADESI1	“ACGT” motif; transacting factor: TAF-1; responsive to ABA and desiccation. Expressed in seeds late during embryogenesis. Induced by ABA and osmotic stress in vegetative tissues.
CACGTGMOTIF	“CACGTG motif”; essential for expression of beta-phaseolin gene during embryogenesis
CANBNNAPA	Core of “(CA) <sub>n</sub> element” in storage protein genes; embryo- and endosperm-specific transcription of napin (storage protein) gene
CARGNCAT	Noncanonical CARg motif (CC-Wx8-GG); A relevant cis-element for the response to AGL15 (AGAMOUS-like 15) in vivo
DPBFCOREDCDC3	DPBF-1 and 2 (Dc3 promoter-binding factor-1 and 2) binding core sequence; Dc3 expression is embryo-specific and induced by ABA
DRE1COREZMRAB17	“DRE1” core found in maize (Z.M.) rab17 gene promoter; “DRE1” was protected, in vivo footprinting, by a protein in embryos specifically, but in leaves, was protected when was treated with ABA and drought; rab17 is expressed during late embryogenesis and is induced by ABA
DRE2COREZMRAB17	“DRE2”; core sequence in rab17 gene promoter. rab17 is expressed during late embryogenesis and is induced by ABA
EBOXBNNAPA	“E-box” of napA storage-protein gene
PYRIMIDINEBOXOSRAMY1A	Found in the promoter of alpha-amylase (Amy2/32b) gene which is induced in the aleurone layers in response to GA in embryo
RYREPEATVFLEB4	RY repeat motif; quantitative seed expression; binding site of <i>Arabidopsis</i> B3-domain-containing transcription factor FUS3, mediates abscisic acid-induced transcription
SEF1MOTIF	“SEF1 (soybean embryo factor 1)” binding motif; regulates the expression of genes encoding for the beta-conglycinin seed storage proteins
SEF3MOTIFGM	“SEF3 binding site”; regulates the expression of genes encoding for the beta-conglycinin seed storage proteins
SEF4MOTIFGM7S	“SEF4 (soybean embryo factor 4)” binding motif; regulates the expression of genes encoding for the beta-conglycinin seed storage proteins
TATABOX2	“TATA box”; TATA box found in beta-phaseolin promoter which is accurate transcription initiation in the embryo stage

\*Details of PLACE entries were retrieved from the [https://www.dna.affrc.go.jp/place/place\\_seq.shtml](https://www.dna.affrc.go.jp/place/place_seq.shtml) (accessed on 19th May 2022).

TABLE 4: Distribution of SE markers across network modules ordered by the number of interactors.

	Gene identifier	Module	Gene name	No. of interactors
1	AT3G26790	Turquoise	<i>FUSCA3 (FUS3)</i>	3346
2	AT5G13790	Turquoise	<i>AGAMOUS-LIKE 15 (AGL15)</i>	3308
3	AT1G21970	Turquoise	<i>LEAFY COTYLEDON 1 (LEC1)</i>	3297
4	AT5G45980	Turquoise	<i>WUSCHEL RELATED HOMEODOMAIN 8 (WOX8)</i>	3158
5	AT3G24650	Turquoise	<i>ABSCISIC ACID INSENSITIVE 3 (ABI3)</i>	3115
6	AT1G78080	Brown	<i>WOUND INDUCED DEDIFFERENTIATION 1 (WIND1)</i>	2111
7	AT5G57390	Turquoise	<i>AINTEGUMENTA-LIKE 5 (AIL5)</i>	752
8	AT4G37750	Red	<i>AINTEGUMENTA (ANT)</i>	688
9	AT1G63470	Red	<i>AT-HOOK MOTIF NUCLEAR LOCALIZED PROTEIN 5 (AHL5)</i>	522
10	AT5G65510	Purple	<i>AINTEGUMENTA-LIKE 7 (AIL7)</i>	216

3.7. *Validation of Hub Genes.* A comparison of hub genes and DEGs showed that 31 hub genes are differentially expressed in SE (the expression values of differentially expressed hub genes are given in Table S4). Further, expression analysis of these genes using the *Arabidopsis* eFP browser demonstrated that two hub genes, *AT1G19540* (Figure 6(a)) and *AT5G44380* (Figure 6(b)), exhibit a seed-specific pattern of expression.

Moreover, analysis of the expression profiles of hub genes in the *Arabidopsis* somatic embryo transcriptome dataset (E-MTAB-2465) published by Wickramasuriya and Dunwell (2015) revealed that 62 hub genes are differentially expressed in somatic embryonic tissues compared to leaf tissues ( $|\log_2 FC| \geq 2.0$  and  $p$  value  $< 0.05$ ; Figure 7). Of these, 15 genes were identified as DEGs in the present analysis. For instance, *CYSTEINE-RICH TRANSMEMBRANE*



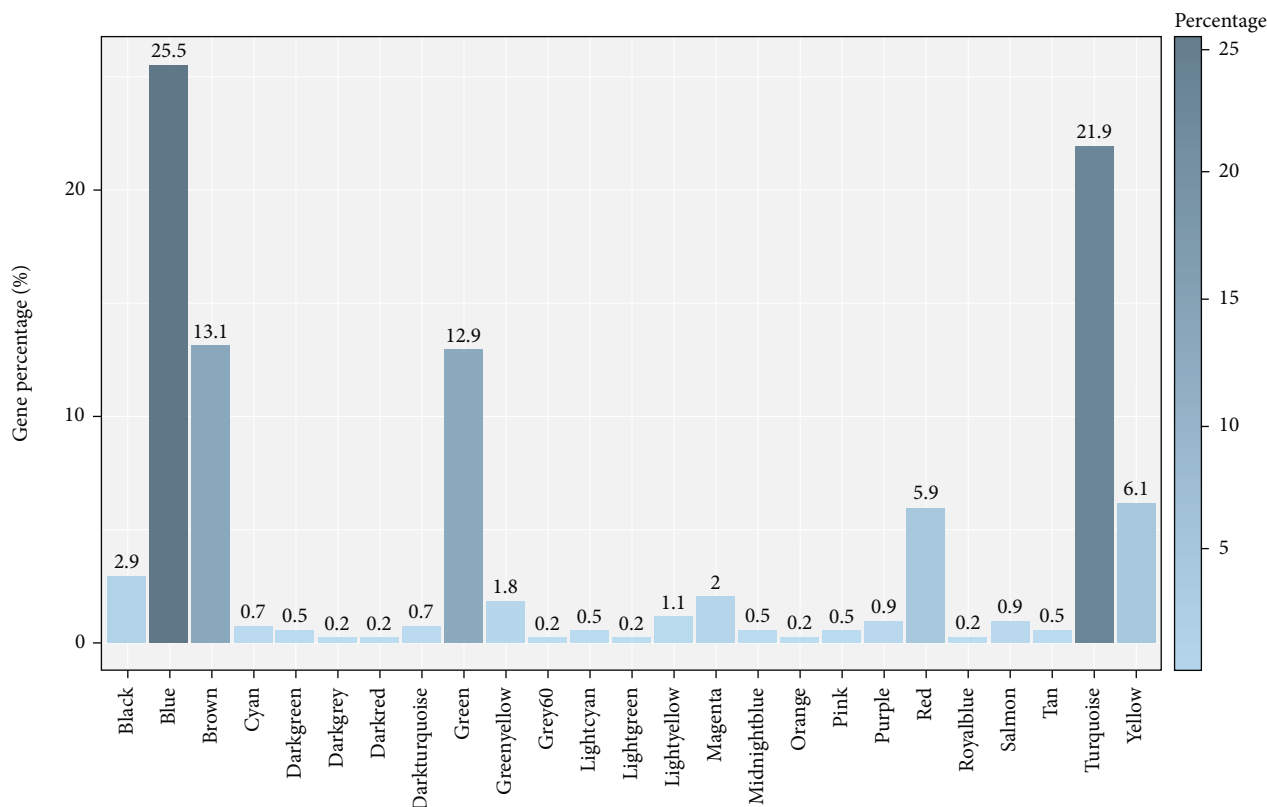


FIGURE 10: Distribution of *EMB* genes across gene modules. The coloured bars represent the ratio between the number of *EMB* genes in each module and the total number of *EMB* genes in the network.

*MODULE 7 (ATHCYSTM7/AT2G33520), HEPTAHELICAL TRANSMEMBRANE PROTEIN2 (AT4G30850), INDOLE-3-ACETIC ACID INDUCIBLE 30 (IAA30/AT3G62100), RPS9C, VASCULATURE COMPLEXITY AND CONNECTIVITY (AT2G32280), AT2G21820, AT2G38900, and AT5G43770* showed a marked expression in somatic embryonic tissues as compared to leaf tissues. Expression analysis using the *Arabidopsis* eFP browser further showed that *AT2G29300, AT2G21820, AT2G38900, AT5G43770, ATHCYSTM7, and AT1G19540* exhibit a seed-specific pattern of gene expression.

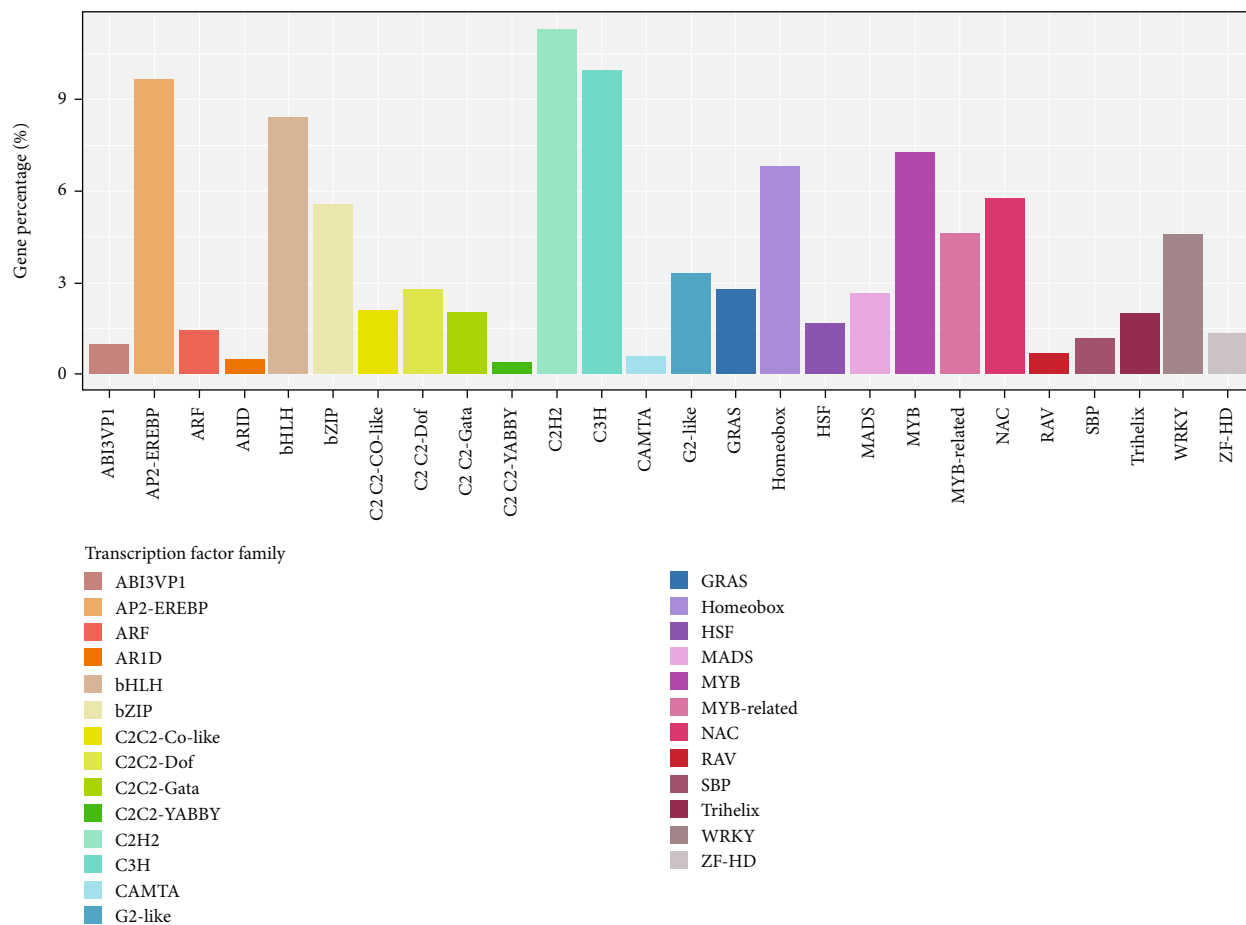
As expected, few hub genes highly expressed in leaf tissues were repressed in somatic embryos indicating the importance of gene regulation in SE (Figure 7); for instance, *CELLULOSE SYNTHASE-LIKE B4 (AT2G32540), CHOLINE/ETHANOLAMINE KINASE 3 (AT4G09760), GLUTAMATE DECARBOXYLASE 2 (AT1G65960), ISOPROPYLMALATE ISOMERASE 2 (AT2G43100), PEROXIREDOXIN Q (PRXQ/AT3G26060), PHOTOSYNTHETIC NDH SUBCOMPLEX L 4 (PnsL4/AT4G39710), PLASTID RIBOSOMAL PROTEIN S20 (AT3G15190), STO (AT1G06040), SINAPOYLGLUCOSE 1 (SNG1/AT2G22990), THYLAKOID RHODANESE-LIKE (TROL/AT4G01050), TONOPLAST INTRINSIC PROTEIN 2 (TIP2/AT3G26520), AT3G50685, AT4G33666, AT5G16010, and AT5G54540* genes showed a marked repression in somatic embryos compared to leaf tissues.

In summary, the present study identified a total of 78 hub genes as potential regulators of SE (Figure 8), including

genes showing marked overexpression as well as repression in SE. Of these, 41 genes have not been functionally annotated thus far. The analysis of the promoter sequences of these uncharacterized hub genes using the PLACE database identified a total of 215 different plant CREs; *ARR1AT, CAATBOX1, CACTFTPPCA1, DOFCOREZM, GATABOX, GT1CONSENSUS, POLLEN1LELAT52, and WRKY71OS* were observed in all 41 functionally uncharacterized potential hub genes. Moreover, several CREs related to embryogenesis were identified (Figure 9). The functions of the predicted CREs are included in Table 3.

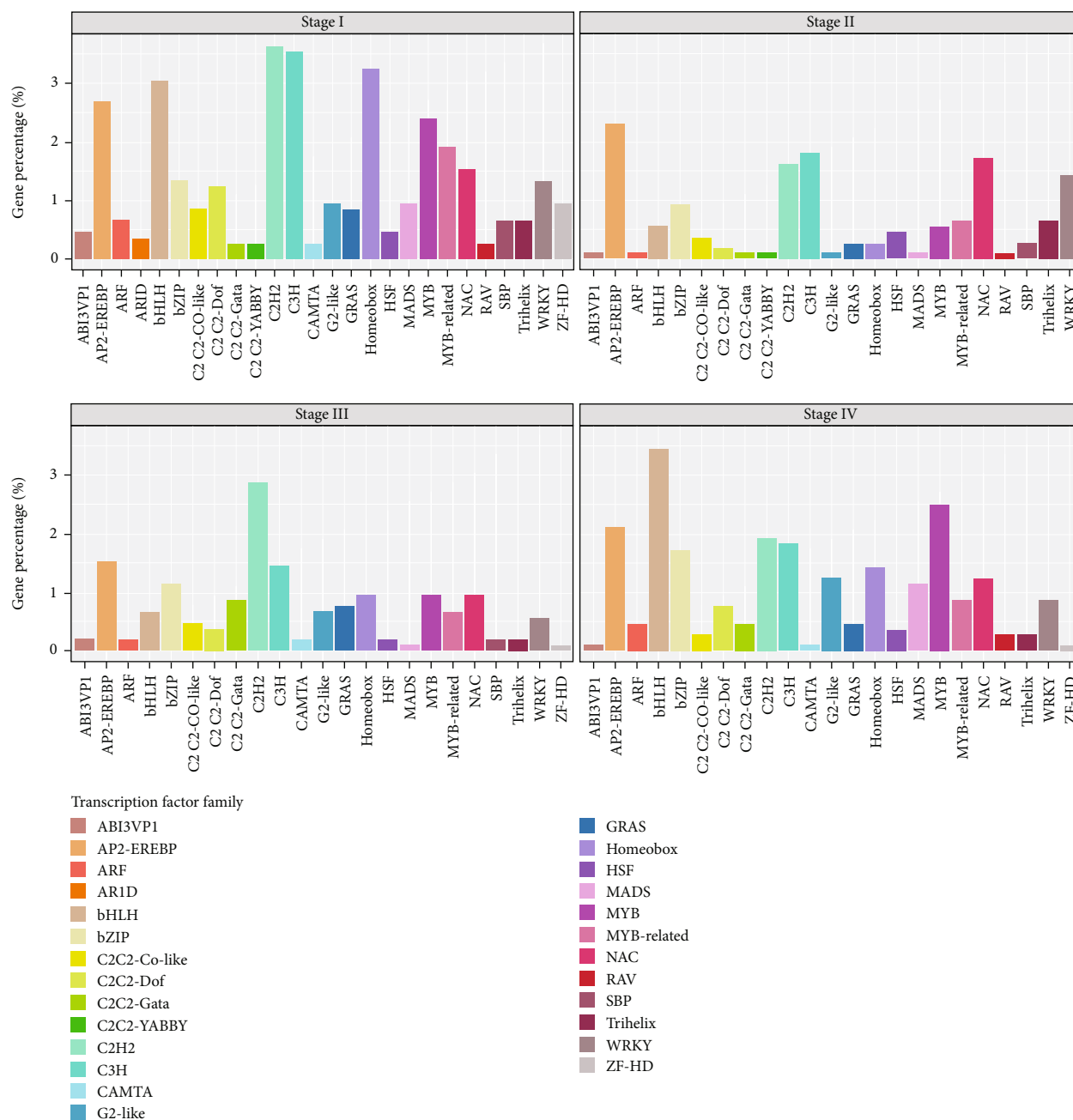
**3.8. Distribution of Embryogenesis-Related Genes across Network Modules.** Further exploration of genes mapped to each network module found that 10 key regulators of SE including *LEC1, FUSCA3 (FUS3), and ABSCISIC ACID INSENSITIVE 3 (ABI3)* are present among the highly connected genes in the network (Table 4); SE-related marker genes, *LEC2, SERK1, WUS, BBM, and WUSCHEL RELATED HOMEBOX 2 (WOX2)* showed low variance in the present dataset and thus were not included in the GCN analysis. We also observed that the majority of previously published *EMB* genes [41] are localized to the blue and turquoise modules, which showed significant association with stage I and stage III, respectively (Figure 10; see Table S5).

In addition, we observed that 1,711 *Arabidopsis* TFs are distributed across all the gene modules except in light-green and royal-blue modules, with the highest number of



(a)

FIGURE 11: Continued.



(b)

FIGURE 11: Distribution of TFs in SE. (a) Overall distribution of TFs. The percentage is calculated as the ratio of TFs belonging to each family with respect to the total number of TFs in the network. (b) Distribution of SE-related TFs across different somatic embryo developmental stages. The percentage is calculated for each stage as the ratio of TFs present in each family with respect to the total number of TFs in the network.

TFs present in the turquoise module (the complete list of TFs included in the GCN is given in Table S6). Notably, AP2/EREBP (APETALA2/ethylene-responsive element binding proteins), bHLH (basic helix-loop-helix), bZIP, C2H2 (Cys2-His2), HB (homeobox), NAC (NAM, ATAF, and CUC), MYB (MYB-domain), C3H, and WRKY TF families were highly represented (Figure 11(a)). Of these, members of AP2/EREBP, bHLH, C2H2, HB, NAC, MYB, and WRKY TF families were involved in early SE

(Figure 11(b)). Interestingly, TFs that are targets of several microRNAs (miRNAs) were also recovered from the GCN (Table S7).

Notably, several gene encoding epigenetic regulators were localized in network modules (Figure 12). This included 14 genes involved in DNA modification, 51 genes involved in histone modification, 34 genes involved in chromatin remodeling, 15 genes encoding polycomb-group proteins, and 55 genes associated with RNA silencing (see

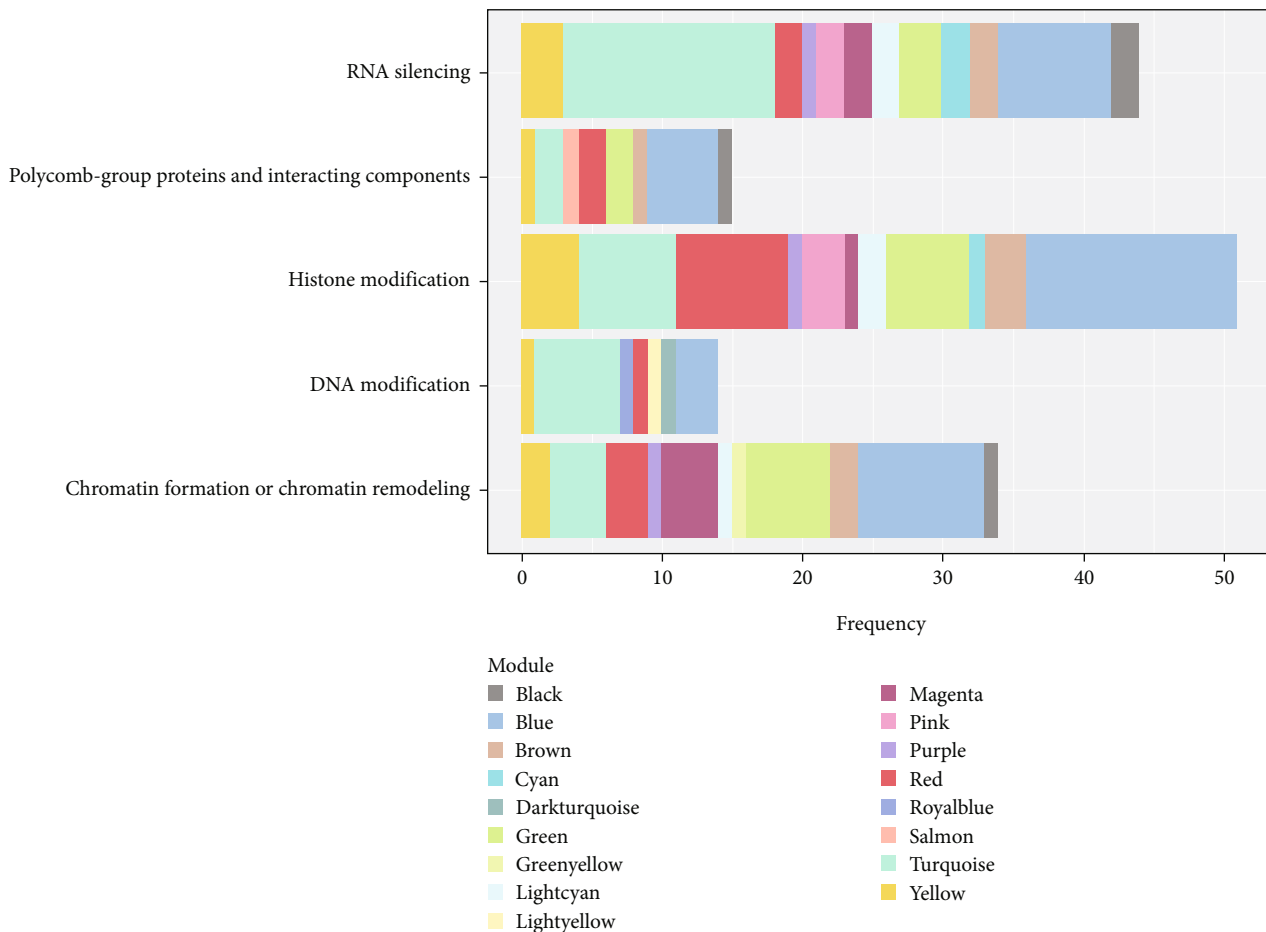


FIGURE 12: Distribution of genes encoding epigenetic regulators across the network modules.

Table S8). Each of these genes directly interacted with numerous modular genes forming a complex network.

#### 4. Discussion

Plant embryogenesis is a meticulous developmental process that requires the regulation of multiple genes. A GCN will serve as a map of statistically significant gene interactions that helps in narrowing down the transcriptome to the potential gene interactions involved in biological processes. Recently, Clercq et al. report an integrated gene regulatory network for *Arabidopsis* covering TFs and target genes [47]. In the present study, WGCNA was employed to explore potential clusters of highly coregulated genes and hub genes associated with SE. Although WGCNA has been previously applied to construct a GCN for *Arabidopsis* zygotic embryogenesis (ZE) [33], to the best of our knowledge, this is the first report on the use of WGCNA to construct a GCN for *Arabidopsis* SE and to explore SE-related network modules and hub genes. The findings of this study provide new insights into the molecular mechanism of SE in plants.

The GCN constructed for SE comprised of 26 network modules: black (674 genes), blue (2,973 genes), brown (2,437 genes), cyan (125 genes), dark-green (52 genes),

dark-grey (39 genes), dark-orange (35 genes), dark-red (54 genes), dark-turquoise (52 genes), green (2,132 genes), green-yellow (189 genes), grey60 (79 genes), light-cyan (86 genes), light-green (59 genes), light-yellow (58 genes), magenta (338 genes), midnight-blue (117 genes), orange (35 genes), pink (357 genes), purple (271 genes), red (853 genes), royal-blue (56 genes), salmon (162 genes), tan (172 genes), turquoise (3,418 genes), and yellow (2,223 genes) modules. Among them, 18 modules showed strong associations with different stages of SE; module-trait relationship analysis revealed that four, four, seven, and four modules were significantly correlated with stages I, II, III, and IV of SE, respectively. This suggests that SE involves complex genetic networks.

Functional enrichment analysis using GO is one of the most widely used bioinformatic methods to classify genes into functionally related groups [48–50]. GO analysis of the coexpressed gene clusters (or network modules) showed that the initial stages of SE were mainly enriched with biological processes such as hormone-mediated signaling, biosynthesis pathways, ATP synthesis, DNA methylation, and replication. Notably, genes involved in lipid transport, post-embryonic development, signal transduction, and seed dormancy were enriched in later stages of SE; this indicates the developmental shift in the maturation phase with the

accumulation of embryo-specific food reserves, a process that aids in withstanding dormancy and postembryonic development [2, 10, 51]. Furthermore, genes related to stress responses (e.g., oxidative and water stress), phytohormones (e.g., cytokinin, abscisic acid, gibberellin, and jasmonic acid), and metabolic processes were enriched in all stages of somatic embryo development studied, from the initiation to maturation stage. These findings further confirmed the importance of cell-cell interactions [52], signaling [9, 13, 53], and transcriptional activation of stress responses [54, 55] during plant SE.

High-degree nodes or the genes with high network connectivity in GCN modules (“hub genes”) may have important biological functions [36, 56–58]; often, they may serve as biological markers. Several studies have successfully employed WGCNA to mine hub genes controlling biological processes [34, 59–62]. The present study reports 260 potential hub genes related to SE based on the degree of connectivity. These genes may play pivotal roles in the regulation of SE. Importantly, 13 TFs encoded by hub genes were identified in the coexpression network. They were *ARF9*, *NAC080*, *ZML2*, *bHLH39*, *KAN3*, *bZIP19*, *bZIP44*, *HDG4*, *FBH4*, *STO*, *CDC5*, *AT5G60820*, and *AT4G01270*; functional roles of many of these genes in the regulation of SE are not reported. Previous studies have reported that *ARF9* represses the expression of its target genes such as *TOPLESS* (*TPL*) and *TPL*-related proteins [63, 64]. Wójcikowska and Gaj observed stable expression of *ARF9* during SE [65]. In addition, *KAN3*, a member of GARP TF family, has also exhibited an embryonic expression pattern.

In addition, *ROOT UV-B SENSITIVE 6* (*RUS6*; *AT5G49820*), which encodes a DUF647 (DOMAIN OF UNKNOWN FUNCTION 647) containing protein, an ankyrin repeat-containing gene designated as *AT5G65860* and a gene that encodes hydroxyproline-*O*-glycosyltransferases (*Hyp-O-GALT*), *GALT4* (*AT1G27120*), was also identified as hub genes in the coexpression network. The members of the *RUS* gene family play diverse roles in plant development [66]. Interestingly, knockout mutants of *RUS6* have shown a strong embryo-lethal phenotype. In *Arabidopsis*, ankyrin repeat-containing proteins have been classified into 16 groups [67], and of these, proteins with only ankyrin repeats have been associated with disease resistance, antioxidation, embryogenesis, and development [68–70]. For instance, T-DNA mutants of the *EMB 506* gene, which encodes a protein containing five ankyrin repeats, have shown defective embryo development at the globular-to-heart stage transition [70]. Moreover, *Hyp-O-GALT* enzymes are responsible for hydroxyproline glycosylation of arabinogalactan proteins, which are known to function in various aspects of plant growth and development including SE [71–73]. Although the hub genes identified in the present study are implicated to function in many plant developmental processes, the functions of many of the hub genes in SE remain to be elucidated. Hence, these genes could be potential targets for functional studies in the future.

Promoter analysis of the functionally uncharacterized hub genes using the PLACE database revealed the overrepresentation of two motifs in many of the promoter regions.

These were EBOXBNNAPA (consensus sequence: CANNTG) and SEF4MOTIFGM7S (consensus sequence: [A/G]TTTTT[A/G]). Of these, EBOXBNNAPA (“E-box” motif) is a CRE found in the regulatory region of the napin gene, *napA* in *Brassica napus* [74]; this gene encodes a storage protein. Moreover, CANNTG provides the binding site for bHLH TFs [75]. bHLH is one of the most frequently represented gene families in DEGs in ZE [76] and SE and is known to have diverse functions in plants [24] including cell proliferation [75]. The recognition sequence of SEF4MOTIFGM7S motif is known to interact with SEF3, a protein expressed in immature soybean seeds that acts as a transcriptional activator of the  $\beta$ -conglycinin  $\alpha$  subunit gene [77]. Hence, the uncharacterized hub genes that showed considerable expression in embryonic tissues are more likely to play a significant role in plant embryo development.

Differential gene expression analysis of hub genes revealed that 78 genes could be considered as potential regulators of SE; of these, 15 genes were differentially expressed in transcriptome datasets derived from two independent studies related to SE [24, 37]. One of the genes identified was *IAA30*, which is a member of one of the families of auxin signaling proteins (*Aux/IAA*; [78]). *iaa30* mutants have displayed significantly impaired SE efficiency, producing fewer somatic embryos per explant [76] and suggesting its role in the initiation phase of SE. Moreover, *IAA30* is a target of two important SE marker genes, *LEC2* and *AGL15* [79, 80]. In addition, two hub genes, *AT1G19540* and *AT5G44380*, showed a marked expression in seed development, suggesting their roles in embryogenesis.

To enhance our understanding of the regulatory mechanism of SE, the distribution of embryogenesis-related genes across the gene modules was examined. Horstman et al. report *LEC1*–*LEC2*–*FUS3*–*BBM*–*ABI3* network to induce SE in *Arabidopsis* [81]. Moreover, Zheng et al. suggest a MADS-domain TF encoding gene, and *AGL15* may associate with *LEC2*, *FUS3*, and *ABI3* during SE [82]. However, a recent study has found that *AGL15* is not essential to promote SE [83]. In the present analysis, 10 key regulators of SE including *LEC1*, *ABI3*, *FUS3*, *AGL15*, and three members of the AINTEGUMENTA-LIKE/PLETHORA (*AIL/PLT*) subfamily (*ANT*, *AIL5*, and *AIL7*) were identified in the coexpression network. Consistent with previous literature, members of the AP2/EREBP, bHLH, bZIP, MYB, HB, WRKY, NAC, C3H, and C2H2 TF families were overrepresented in the GCN [76, 84]. In addition, members of the TF families (i.e., SPB (SQUAMOSA promoter binding protein-like), GRAS (GRAS-domain), trihelix, G2-like, and CAMTA (CALMODULIN BINDING TRANSCRIPTION ACTIVATOR 3)) that are not or to a lesser extent reported to be involved in SE were identified. The members of GRAS, trihelix, and CAMTA families are known to be involved in the regulation of stress responses [47, 85, 86].

Further, it is reported that miRNAs (e.g., miR156, miR159, miR162, miR164, miR166, miR167, miR169, miR168, miR171, miR319, miR393, and miR396) play an important role in SE [87–91]. Consistent with previous studies, several TFs targeted by miRNAs were recovered from the SE-related GCN. This included seven miR156/157 targeting

genes of the SPB TF family, seven miR169 targeting genes of the CCAAT TF family, six miR396 targeting genes of the GRF TF family, five miR166/miR165 targeting genes of the HB TF family, five miR164 targeting genes of the NAC family, and five miR159/miR319 targeting genes of the TCP TF family. These miRNA-targeted TF encoding genes may play a significant role in the regulation of SE responses.

Recent studies have uncovered critical roles of epigenetic modifications in the regulation of SE, in particular, DNA methylation/demethylation [92–94] and histone modifications [91, 95, 96]. Recently, an expression study on *Arabidopsis* embryos at single-cell resolution has provided evidence for distinct expression patterns for many epigenetic regulators across embryonic tissues [97]. Our coexpression network also revealed that many genes encoding epigenetic regulators such as *METHYLTRANSFERASE 1 (MET1)*, *CHROMOMETHYLASE 3 (CMT3)*, *DEMETER (DME)*, *DEMETER-LIKE (DML1,-2)*, histone acetyltransferases (*HISTONE ACETYLTRANSFERASE OF THE CBP FAMILY (HAC1,-4,-5,-12)*), histone deacetylases (i.e., *HISTONE DEACETYLASE (HDA1,-2,-3,-5,-6,-8,-9,-14,-15,-17)*), and histone demethylases (*JUMONJI DOMAIN-CONTAINING PROTEIN 16 (JM14,-16,-22,-27,-29)*) were coexpressed with key genes involved in the regulation of SE.

The present study showed that the WGCNA pipeline could be used to identify biologically relevant modules of SE. However, our analysis has some limitations. The main limitations were the small sample size used in the analysis and the lack of an independent dataset to replicate the findings. Langfelder and Horvath [32] recommend using at least 15 samples to construct robust networks. However, high-quality, clean data could also result in biologically meaningful networks even with <15 samples. Therefore, further experiments are recommended to validate the hub genes discovered in the present study. Furthermore, the GCN built in the present study was based on microarray gene expression data. Although hybridization-based gene expression profiling approaches are high-throughput and relatively inexpensive, they have a number of limitations; most importantly, they provide only an indirect measure of the level of gene expression and can only be used to study the expression levels of genes that the arrays are designed to detect and are subjected to cross-hybridization biases [98]. Given the limitations of this approach, it would be recommended to perform a GCN analysis employing an expression dataset generated through high-throughput transcriptome sequencing (RNA-seq) with an appropriate number of replicates. Unlike microarrays, RNA-seq is not dependent on prior knowledge about the genome sequence and has higher sensitivity to genes expressed either at a low or very high level and also has higher levels of reproducibility than microarrays [99]. Therefore, it could generate a more suitable dataset for GCN analysis.

## 5. Conclusion

In this study, a GCN was successfully constructed for SE employing WGCNA. Gene modules and hub genes related to *Arabidopsis* somatic embryo development were success-

fully mined based on their statistical significance. The findings reported here provide a unique resource to advance the regulation of SE at the molecular level.

## Abbreviations

<i>ABI3</i> :	<i>ABSCISIC ACID INSENSITIVE 3</i>
<i>AGL</i> :	<i>AGAMOUS-LIKE</i>
<i>AIL</i> :	<i>AINTEGUMENTA-LIKE</i>
<i>BBM</i> :	<i>BABY BOOM</i>
bHLH:	Basic helix-loop-helix
<i>bZIP</i> :	<i>BASIC LEUCINE-</i>
<i>C2H2</i> :	<i>Cys2-His2</i>
<i>CRE</i> :	<i>Cis-acting regulatory element</i>
<i>DEG</i> :	<i>Differentially expressed gene</i>
<i>EMB</i> :	<i>EMBRYO-DEFECTIVE</i>
<i>FC</i> :	<i>Fold change</i>
<i>FUS3</i> :	<i>FUSCA3</i>
<i>GCN</i> :	<i>Gene coexpression network</i>
<i>GEO</i> :	<i>Gene Expression Omnibus</i>
<i>GO</i> :	<i>Gene Ontology</i>
<i>HB</i> :	<i>HOMEBOX</i>
<i>IAA30</i> :	<i>INDOLE-3-ACETIC ACID INDUCIBLE 30</i>
<i>JMJ</i> :	<i>JUMONJI DOMAIN-CONTAINING</i>
<i>KAN3</i> :	<i>KANADI 3</i>
<i>LEC</i> :	<i>LEAFY COTYLEDON</i>
<i>ME</i> :	<i>Module eigengene</i>
<i>MEME</i> :	<i>Multiple Em for Motif Elicitation</i>
<i>miRNA</i> :	<i>MicroRNA</i>
<i>PLACE</i> :	<i>Plant cis-acting regulatory DNA elements</i>
<i>r</i> :	<i>Pearson correlation coefficient</i>
<i>RNA-Seq</i> :	<i>RNA-sequencing</i>
<i>RUS</i> :	<i>ROOT UV-B SENSITIVE</i>
<i>SE</i> :	<i>Somatic embryogenesis</i>
<i>STO</i> :	<i>SALT TOLERANCE</i>
<i>TF</i> :	<i>Transcription factor</i>
<i>WGCNA</i> :	<i>Weighted Gene Correlation Network Analysis.</i>

## Data Availability

The datasets used to support the findings of this study are included within the article and within the supplementary information files.

## Disclosure

A preprint has previously been published [100].

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Authors' Contributions

KKD participated in the design of the study, performed the GCN construction and analysis, and drafted the manuscript. JMD helped to interpret data and draft the manuscript. AMW conceived the study, participated in the design of



the study, and helped to draft the manuscript. All authors have read and approved the final manuscript.

## Acknowledgments

The authors gratefully acknowledge the support from the University of Colombo, Sri Lanka.

## Supplementary Materials

*Supplementary 1. Table S1.* The list of variance filtered genes (variance >0.25 quantile).

*Supplementary 2. Table S2.* The list of differentially expressed genes.

*Supplementary 3. Table S3.* The top 10 hub genes of each network module.

*Supplementary 4. Table S4.* Important hub genes based on differential gene expression analysis.

*Supplementary 5. Table S5.* EMB genes in the coexpression network.

*Supplementary 6. Table S6.* TFs in the coexpression network.

*Supplementary 7. Table S7.* The list of microRNA-targeted TFs.

*Supplementary 8. Table S8.* Distribution of gene encoding epigenetic regulators across network modules.

## References

- [1] F. C. Steward, M. O. Mapes, and K. Mears, "Growth and organized development of cultured cells. II. Organization in cultures grown from freely suspended cell," *American Journal of Botany*, vol. 45, no. 10, pp. 705–708, 1958.
- [2] J. L. Zimmerman, "Somatic embryogenesis: a model for early development in higher plants," *The Plant Cell*, vol. 5, no. 10, pp. 1411–1423, 1993.
- [3] A. Capron, S. Chatfield, N. Provart, and T. Berleth, "Embryogenesis: pattern formation from a single cell," *The Arabidopsis Book*, vol. 7, article e0126, 2009.
- [4] S. C. de Vries and D. Weijers, "Plant embryogenesis," *Current Biology*, vol. 27, no. 17, pp. R870–R873, 2017.
- [5] H. Etienne, "Somatic embryogenesis protocol: coffee (*Coffea arabica* L)," in *Protocol for Somatic Embryogenesis in Woody Plants*, pp. 167–179, Springer-Verlag, Berlin/Heidelberg, 2005.
- [6] D. A. Steinmacher, C. R. Clement, and M. P. Guerra, "Somatic embryogenesis from immature peach palm inflorescence explants: towards development of an efficient protocol," *Plant Cell, Tissue and Organ Culture*, vol. 89, no. 1, pp. 15–22, 2007.
- [7] S. Manrique-Trujillo, D. Díaz, R. Reaño, M. Ghislain, and J. Kreuze, "Sweetpotato plant regeneration via an improved somatic embryogenesis protocol," *Scientia Horticulturae*, vol. 161, pp. 95–100, 2013.
- [8] S. Vinoth, P. Gurusaravanan, and N. Jayabalan, "Optimization of somatic embryogenesis protocol in *Lycopersicon esculentum* L. using plant growth regulators and seaweed extracts," *Journal of Applied Phycology*, vol. 26, no. 3, pp. 1527–1537, 2014.
- [9] H. A. Méndez-Hernández, M. Ledezma-Rodríguez, R. N. Avilez-Montalvo et al., "Signaling overview of plant somatic embryogenesis," *Frontiers in Plant Science*, vol. 10, p. 77, 2019.
- [10] P. K. Dantu, U. K. Tomar, and G. Tripathi, "Somatic embryogenesis," in *Cellular and Biochemical Science*, pp. 892–908, IK International House Pvt Ltd, New Delhi, 2010.
- [11] M. A. El-Esawi, "Nonzygotic embryogenesis for plant development," in *Plant Tissue Culture: Propagation, Conservation and Crop Improvement*, M. Anis and N. Ahmad, Eds., pp. 583–598, Springer Singapore, Singapore, 2016.
- [12] F. Zeng, X. Zhang, L. Cheng et al., "A draft gene regulatory network for cellular totipotency reprogramming during plant somatic embryogenesis," *Genomics*, vol. 90, no. 5, pp. 620–628, 2007.
- [13] A. Smertenko and P. Bozhkov, "The life and death signalling underlying cell fate determination during somatic embryogenesis," in *Applied Plant Cell Biology*, pp. 131–178, Springer, Berlin, Heidelberg, 2014.
- [14] J. E. Cetz-Chel and V. M. Loyola-Vargas, "Transcriptome profile of somatic embryogenesis," in *Somatic Embryogenesis: Fundamental Aspects and Applications*, V. M. Loyola-Vargas and N. Ochoa-Alejo, Eds., pp. 39–52, Springer International Publishing, Cham, 2016.
- [15] V. Hecht, J.-P. Vielle-Calzada, M. V. Hartog et al., "The *Arabidopsis somatic embryogenesis receptor kinase 1* gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture," *Plant Physiology*, vol. 127, no. 3, pp. 803–816, 2001.
- [16] X. Yang and X. Zhang, "Regulation of somatic embryogenesis in higher plants," *Critical Reviews in Plant Science*, vol. 29, no. 1, pp. 36–57, 2010.
- [17] M. D. Gaj, S. Zhang, J. J. Harada, and P. G. Lemaux, "Leafy cotyledon genes are essential for induction of somatic embryogenesis of *Arabidopsis*," *Planta*, vol. 222, no. 6, pp. 977–988, 2005.
- [18] M. Ikeda, M. Takahashi, S. Fujiwara, N. Mitsuda, and M. Ohme-Takagi, "Improving the efficiency of adventitious shoot induction and somatic embryogenesis via modification of *WUSCHEL* and *LEAFY COTYLEDON 1*," *Plants*, vol. 9, no. 11, p. 1434, 2020.
- [19] T. Lotan, M. Ohto, K. M. Yee et al., "*Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells," *Cell*, vol. 93, no. 7, pp. 1195–1205, 1998.
- [20] S. L. Stone, L. W. Kwong, K. M. Yee et al., "*LEAFY COTYLEDON2* encodes a B3 domain transcription factor that induces embryo development," *Proceedings of the National Academy of Sciences*, vol. 98, no. 20, pp. 11806–11811, 2001.
- [21] S. L. Stone, S. A. Braybrook, S. L. Paula et al., "*Arabidopsis* LEAFY COTYLEDON2 induces maturation traits and auxin activity: implications for somatic embryogenesis," *Proceedings of the National Academy of Sciences*, vol. 105, no. 8, pp. 3151–3156, 2008.
- [22] K. Boutilier, R. Offringa, V. K. Sharma et al., "Ectopic expression of *BABY BOOM* triggers a conversion from vegetative to embryonic growth," *The Plant Cell*, vol. 14, no. 8, pp. 1737–1749, 2002.
- [23] J. Zuo, Q.-W. Niu, G. Frugis, and N.-H. Chua, "The *WUSCHEL* gene promotes vegetative-to-embryonic

- transition in *Arabidopsis*,” *The Plant Journal*, vol. 30, no. 3, pp. 349–359, 2002.
- [24] A. M. Wickramasuriya and J. M. Dunwell, “Global scale transcriptome analysis of *Arabidopsis* embryogenesis *in vitro*,” *BMC Genomics*, vol. 16, no. 1, p. 301, 2015.
- [25] Y. Indoliya, P. Tiwari, A. S. Chauhan et al., “Decoding regulatory landscape of somatic embryogenesis reveals differential regulatory networks between *japonica* and *indica* rice subspecies,” *Scientific Reports*, vol. 6, no. 1, article 23050, 2016.
- [26] B. Singla, A. K. Tyagi, J. P. Khurana, and P. Khurana, “Analysis of expression profile of selected genes expressed during auxin-induced somatic embryogenesis in leaf base system of wheat (*Triticum aestivum*) and their possible interactions,” *Plant Molecular Biology*, vol. 65, no. 5, pp. 677–692, 2007.
- [27] F. Zeng, X. Zhang, L. Zhu, L. Tu, X. Guo, and Y. Nie, “Isolation and characterization of genes associated to cotton somatic embryogenesis by suppression subtractive hybridization and microarray,” *Plant Molecular Biology*, vol. 60, no. 2, pp. 167–183, 2006.
- [28] S. A. G. D. Salvo, C. N. Hirsch, C. R. Buell, S. M. Kaeppler, and H. F. Kaeppler, “Whole transcriptome profiling of maize during early somatic embryogenesis reveals altered expression of stress factors and embryogenesis-related genes,” *PLoS One*, vol. 9, no. 10, article e111407, 2014.
- [29] M. K. Rajesh, T. P. Fayas, S. Naganeeswaran et al., “De novo assembly and characterization of global transcriptome of coconut palm (*Cocos nucifera* L.) embryogenic calli using Illumina paired-end sequencing,” *Protoplasma*, vol. 253, no. 3, pp. 913–928, 2016.
- [30] S. van Dam, U. Vösa, A. van der Graaf, L. Franke, and J. P. de Magalhães, “Gene co-expression analysis for functional classification and gene–disease predictions,” *Briefings in Bioinformatics*, vol. 19, pp. 575–592, 2017.
- [31] B. Zhang and S. Horvath, “A general framework for weighted gene co-expression network analysis,” *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, p. Article17, 2005.
- [32] P. Langfelder and S. Horvath, “WGCNA: an R package for weighted correlation network analysis,” *BMC Bioinformatics*, vol. 9, no. 1, article 559, 2008.
- [33] P. Gao, D. Xiang, T. D. Quilichini et al., “Gene expression atlas of embryo development in *Arabidopsis*,” *Plant Reproduction*, vol. 32, no. 1, pp. 93–104, 2019.
- [34] R. Shaik and W. Ramakrishna, “Genes and co-expression modules common to drought and bacterial stress responses in *Arabidopsis* and rice,” *PLoS One*, vol. 8, no. 10, article e77261, 2013.
- [35] Y. Tai, C. Liu, S. Yu et al., “Gene co-expression network analysis reveals coordinated regulation of three characteristic secondary biosynthetic pathways in tea plant (*Camellia sinensis*),” *BMC Genomics*, vol. 19, no. 1, p. 616, 2018.
- [36] M. Zhu, H. Xie, X. Wei et al., “WGCNA analysis of salt-responsive core transcriptome identifies novel hub genes in rice,” *Genes*, vol. 10, no. 9, p. 719, 2019.
- [37] M. G. Becker, A. Chan, X. Mao et al., “Vitamin C deficiency improves somatic embryo development through distinct gene regulatory networks in *Arabidopsis*,” *Journal of Experimental Botany*, vol. 65, no. 20, pp. 5903–5918, 2014.
- [38] P. Langfelder, R. Luo, M. C. Oldham, and S. Horvath, “Is my network module preserved and reproducible?,” *PLoS Computational Biology*, vol. 7, no. 1, article e1001057, 2011.
- [39] T. Tian, Y. Liu, H. Yan et al., “agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update,” *Nucleic Acids Research*, vol. 45, no. W1, pp. W122–W129, 2017.
- [40] E. Magnani, J. M. Jiménez-Gómez, L. Soubigou-Taconnat, L. Lepiniec, and E. Fiume, “Profiling the onset of somatic embryogenesis in *Arabidopsis*,” *BMC Genomics*, vol. 18, no. 1, p. 998, 2017.
- [41] D. W. Meinke, “Genome-wide identification of *EMBRYO-DEFECTIVE* (*EMB*) genes required for growth and development in *Arabidopsis*,” *The New Phytologist*, vol. 226, no. 2, pp. 306–325, 2020.
- [42] D. Xiang, P. Venglat, C. Tibiche et al., “Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in *Arabidopsis*,” *Plant Physiology*, vol. 156, no. 1, pp. 346–356, 2011.
- [43] C. S. Pikaard and S. O. Mittelsten, “Epigenetic regulation in plants,” *Cold Spring Harbor Perspectives in Biology*, vol. 6, no. 12, article a019315, 2014.
- [44] T. L. Bailey, M. Boden, F. A. Buske et al., “MEME SUITE: tools for motif discovery and searching,” *Nucleic Acids Research*, vol. 37, suppl\_2, pp. W202–W208, 2009.
- [45] F. A. Buske, M. Bodén, D. C. Bauer, and T. L. Bailey, “Assigning roles to DNA regulatory motifs using comparative genomics,” *Bioinformatics*, vol. 26, no. 7, pp. 860–866, 2010.
- [46] K. Higo, Y. Ugawa, M. Iwamoto, and T. Korenaga, “Plant cis-acting regulatory DNA elements (PLACE) database: 1999,” *Nucleic Acids Research*, vol. 27, no. 1, pp. 297–300, 1999.
- [47] I. De Clercq, J. Van de Velde, X. Luo et al., “Integrative inference of transcriptional networks in *Arabidopsis* yields novel ROS signalling regulators,” *Nature Plants*, vol. 7, no. 4, pp. 500–513, 2021.
- [48] Consortium TGO, “Creating the Gene Ontology resource: design and implementation,” *Genome Research*, vol. 11, no. 8, pp. 1425–1433, 2001.
- [49] M. Ashburner, C. A. Ball, J. A. Blake et al., “Gene Ontology: tool for the unification of biology,” *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000.
- [50] K. Rue-Albrecht, P. A. McGettigan, B. Hernández et al., “GOexpress: an R/Bioconductor package for the identification and visualisation of robust gene ontology signatures through supervised learning of gene expression data,” *BMC Bioinformatics*, vol. 17, no. 1, p. 126, 2016.
- [51] M. A. L. West and J. J. Harada, “Embryogenesis in higher plants: an overview,” *The Plant Cell*, vol. 5, no. 10, pp. 1361–1369, 1993.
- [52] E. G. Williams and G. Maheswaran, “Somatic embryogenesis: factors influencing coordinated behaviour of cells as an embryogenic group,” *Annals of Botany*, vol. 57, no. 4, pp. 443–462, 1986.
- [53] A. Smertenko and P. V. Bozhkov, “Somatic embryogenesis: life and death processes during apical–basal patterning,” *Journal of Experimental Botany*, vol. 65, no. 5, pp. 1343–1360, 2014.
- [54] M. A. Zavattieri, A. M. Frederico, M. Lima, R. Sabino, and B. Arnholdt-Schmitt, “Induction of somatic embryogenesis as an example of stress-related plant reactions,” *Electronic Journal of Biotechnology*, vol. 13, no. 1, pp. 1–9, 2010.
- [55] F. Jin, L. Hu, D. Yuan et al., “Comparative transcriptome analysis between somatic embryos (SEs) and zygotic embryos in cotton: evidence for stress response functions in SE development,” *Plant Biotechnology Journal*, vol. 12, no. 2, pp. 161–173, 2014.



- [56] J. Qiu, Z. Du, Y. Wang et al., “Weighted gene co-expression network analysis reveals modules and hub genes associated with the development of breast cancer,” *Medicine*, vol. 98, no. 6, article e14345, 2019.
- [57] Y. Liu, H.-Y. Gu, J. Zhu, Y.-M. Niu, C. Zhang, and G.-L. Guo, “Identification of hub genes and key pathways associated with bipolar disorder based on weighted gene co-expression network analysis,” *Frontiers in Physiology*, vol. 10, p. 1081, 2019.
- [58] Z. Zhu, Z. Jin, Y. Deng et al., “Co-expression network analysis identifies four hub genes associated with prognosis in soft tissue sarcoma,” *Frontiers in Genetics*, vol. 10, p. 37, 2019.
- [59] J. Du, S. Wang, C. He, B. Zhou, Y.-L. Ruan, and H. Shou, “Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis,” *Journal of Experimental Botany*, vol. 68, no. 8, pp. 1955–1972, 2017.
- [60] X. Zhang, H. Feng, Z. Li et al., “Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis,” *Oncotargets and Therapy*, vol. Volume 11, pp. 6001–6021, 2018.
- [61] Q. Wang, X. Zeng, Q. Song, Y. Sun, Y. Feng, and Y. Lai, “Identification of key genes and modules in response to cadmium stress in different rice varieties and stem nodes by weighted gene co-expression network analysis,” *Scientific Reports*, vol. 10, no. 1, p. 9525, 2020.
- [62] F. Zhang, L. Wang, P. Bai et al., “Identification of regulatory networks and hub genes controlling nitrogen uptake in tea plants [*Camellia sinensis* (L.) O. Kuntze],” *Journal of Agricultural and Food Chemistry*, vol. 68, no. 8, pp. 2445–2456, 2020.
- [63] B. Causier, M. Ashworth, W. Guo, and B. Davies, “The TOPLESS interactome: a framework for gene repression in *Arabidopsis*,” *Plant Physiology*, vol. 158, no. 1, pp. 423–438, 2012.
- [64] B. Gulzar, A. Mujib, M. Q. Malik, R. Sayeed, J. Mamgain, and B. Ejaz, “Genes, proteins and other networks regulating somatic embryogenesis in plants,” *Journal, Genetic Engineering & Biotechnology*, vol. 18, no. 1, p. 31, 2020.
- [65] B. Wójcikowska and M. D. Gaj, “Expression profiling of *AUXIN RESPONSE FACTOR* genes during somatic embryogenesis induction in *Arabidopsis*,” *Plant Cell Reports*, vol. 36, no. 6, pp. 843–858, 2017.
- [66] N. Perry, C. D. Leasure, H. Tong, E. M. Duarte, and Z.-H. He, “RUS6, a DUF647-containing protein, is essential for early embryonic development in *Arabidopsis thaliana*,” *BMC Plant Biology*, vol. 21, no. 1, p. 232, 2021.
- [67] C. Becerra, T. Jahrmann, P. Puigdomènech, and C. M. Vicent, “Ankyrin repeat-containing proteins in *Arabidopsis*: characterization of a novel and abundant group of genes coding ankyrin-transmembrane proteins,” *Gene*, vol. 340, no. 1, pp. 111–121, 2004.
- [68] J. Yan, J. Wang, and H. Zhang, “An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism,” *The Plant Journal*, vol. 29, no. 2, pp. 193–202, 2002.
- [69] H. Zhang, D. C. Scheirer, W. H. Fowle, and H. M. Goodman, “Expression of antisense or sense RNA of an ankyrin repeat-containing gene blocks chloroplast differentiation in *Arabidopsis*,” *The Plant Cell*, vol. 4, no. 12, pp. 1575–1588, 1992.
- [70] S. Albert, B. Despres, J. Guillemot et al., “The *EMB506* gene encodes a novel ankyrin repeat containing protein that is essential for the normal development of *Arabidopsis* embryos,” *The Plant Journal*, vol. 17, no. 2, pp. 169–179, 1999.
- [71] S. Poon, R. L. Heath, and A. E. Clarke, “A chimeric arabinogalactan protein promotes somatic embryogenesis in cotton cell culture,” *Plant Physiology*, vol. 160, no. 2, pp. 684–695, 2012.
- [72] D. Basu, L. Tian, W. Wang et al., “A small multigene hydroxyproline-*O*-galactosyltransferase family functions in arabinogalactan-protein glycosylation, growth and development in *Arabidopsis*,” *BMC Plant Biology*, vol. 15, no. 1, p. 295, 2015.
- [73] S. Duchow, R. I. Dahlke, T. Geske, W. Blaschek, and B. Classen, “Arabinogalactan-proteins stimulate somatic embryogenesis and plant propagation of *Pelargonium sidoides*,” *Carbohydrate Polymers*, vol. 152, pp. 149–155, 2016.
- [74] K. Stålberg, M. Ellerstöm, I. Ezcurra, S. Ablov, and L. Rask, “Disruption of an overlapping E-box/ABRE motif abolished high transcription of the *napa* storage-protein promoter in transgenic *Brassica napus* seeds,” *Planta*, vol. 199, no. 4, pp. 515–519, 1996.
- [75] M. A. Heim, M. Jakoby, M. Werber, C. Martin, B. Weisshaar, and P. C. Bailey, “The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity,” *Molecular Biology and Evolution*, vol. 20, no. 5, pp. 735–747, 2003.
- [76] M. Gliwicka, K. Nowak, S. Balazadeh, B. Mueller-Roeber, and M. D. Gaj, “Extensive modulation of the transcription factor transcriptome during somatic embryogenesis in *Arabidopsis thaliana*,” *PLoS One*, vol. 8, no. 7, article e69261, 2013.
- [77] R. D. Allen, F. Bernier, P. A. Lessard, and R. N. Beachy, “Nuclear factors interact with a soybean beta-conglycinin enhancer,” *The Plant Cell*, vol. 1, no. 6, pp. 623–631, 1989.
- [78] E. Liscum and J. W. Reed, “Genetics of Aux/IAA and ARF action in plant growth and development,” *Plant Molecular Biology*, vol. 49, no. 3/4, pp. 387–400, 2002.
- [79] S. A. Braybrook, S. L. Stone, S. Park et al., “Genes directly regulated by *LEAFY COTYLEDON2* provide insight into the control of embryo maturation and somatic embryogenesis,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 9, pp. 3468–3473, 2006.
- [80] A. M. Wójcik, B. Wójcikowska, and M. D. Gaj, “Current perspectives on the auxin-mediated genetic network that controls the induction of somatic embryogenesis in plants,” *International Journal of Molecular Sciences*, vol. 21, no. 4, p. 1333, 2020.
- [81] A. Horstman, M. Li, I. Heidmann et al., “The *BABY BOOM* transcription factor activates the *LEC1-ABI3-FUS3-LEC2* network to induce somatic embryogenesis,” *Plant Physiology*, vol. 175, no. 2, pp. 848–857, 2017.
- [82] Y. Zheng, N. Ren, H. Wang, A. J. Stromberg, and S. E. Perry, “Global identification of targets of the *Arabidopsis* MADS domain protein *AGAMOUS-Like15*,” *The Plant Cell*, vol. 21, no. 9, pp. 2563–2577, 2009.
- [83] S. Joshi, C. Keller, and S. E. Perry, “The *EAR* motif in the *Arabidopsis* MADS transcription factor *AGAMOUS-like 15* is not necessary to promote somatic embryogenesis,” *Plants*, vol. 10, no. 4, p. 758, 2021.
- [84] K. Nowak and M. D. Gaj, “Transcription factors in the regulation of somatic embryogenesis,” in *Somatic Embryogenesis*:

- Fundamental Aspects and Applications*, V. M. Loyola-Vargas and N. Ochoa-Alejo, Eds., pp. 53–79, Springer International Publishing, Cham, 2016.
- [85] P. Pant, Z. Iqbal, B. K. Pandey, and S. V. Sawant, “Genome-wide comparative and evolutionary analysis of calmodulin-binding transcription activator (CAMTA) family in *Gossypium* species,” *Scientific Reports*, vol. 8, no. 1, p. 5573, 2018.
- [86] R. N. Kaplan-Levy, P. B. Brewer, T. Quon, and D. R. Smyth, “The trihelix family of transcription factors - light, stress and development,” *Trends in Plant Science*, vol. 17, no. 3, pp. 163–171, 2012.
- [87] Z. H. Siddiqui, Z. K. Abbas, M. W. Ansari, and M. N. Khan, “The role of miRNA in somatic embryogenesis,” *Genomics*, vol. 111, no. 5, pp. 1026–1033, 2019.
- [88] A. Alves, D. Cordeiro, S. Correia, and C. Miguel, “Small non-coding RNAs at the crossroads of regulatory pathways controlling somatic embryogenesis in seed plants,” *Plants*, vol. 10, no. 3, p. 504, 2021.
- [89] A. M. Wójcik and M. D. Gaj, “miR393 contributes to the embryogenic transition induced *in vitro* in *Arabidopsis* via the modification of the tissue sensitivity to auxin treatment,” *Planta*, vol. 244, no. 1, pp. 231–243, 2016.
- [90] K. Szyrajew, D. Bielewicz, J. Dolata et al., “MicroRNAs are intensively regulated during induction of somatic embryogenesis in *Arabidopsis*,” *Plant Science*, vol. 8, 2017.
- [91] K. Nowak, J. Morończyk, A. Wójcik, and M. D. Gaj, “AGL15 controls the embryogenic reprogramming of somatic cells in *Arabidopsis* through the histone acetylation-mediated repression of the miRNA biogenesis genes,” *International Journal of Molecular Sciences*, vol. 21, no. 18, p. 6733, 2020.
- [92] X. Chen, X. Xu, X. Shen et al., “Genome-wide investigation of DNA methylation dynamics reveals a critical role of DNA demethylation during the early somatic embryogenesis of *Dimocarpus longan* Lour,” *Tree Physiology*, vol. 40, no. 12, pp. 1807–1826, 2020.
- [93] D. Grzybkowska, K. Nowak, and M. D. Gaj, “Hypermethylation of auxin-responsive motifs in the promoters of the transcription factor genes accompanies the somatic embryogenesis induction in *Arabidopsis*,” *International Journal of Molecular Sciences*, vol. 21, no. 18, p. 6849, 2020.
- [94] L. Ji, S. M. Mathioni, S. Johnson et al., “Genome-wide reinforcement of DNA methylation occurs during somatic embryogenesis in soybean,” *The Plant Cell*, vol. 31, no. 10, pp. 2315–2331, 2019.
- [95] H. Rodríguez-Sanz, J. Moreno-Romero, M.-T. Solís, C. Köhler, M. C. Risueño, and P. S. Testillano, “Changes in histone methylation and acetylation during microspore reprogramming to embryogenesis occur concomitantly with BnHKMT and BnHAT expression and are associated with cell totipotency, proliferation, and differentiation in *Brassica napus*,” *Cytogenetic and Genome Research*, vol. 143, no. 1-3, pp. 209–218, 2014.
- [96] B. Wójcikowska, M. Botor, J. Morończyk et al., “Trichostatin A triggers an embryogenic transition in *Arabidopsis* explants via an auxin-related pathway,” *Frontiers in Plant Science*, vol. 9, p. 1353, 2018.
- [97] P. Kao, M. A. Schon, M. Mosiolek, and M. D. Nodine, “Gene expression variation in *Arabidopsis* embryos at single-nucleus resolution,” *Development*, vol. 148, no. 13, p. dev199589, 2021.
- [98] R. Bumgarner, “Overview of DNA microarrays: types, applications, and their future,” *Current Protocols in Molecular Biology*, vol. 101, no. 1, p. Unit 22.1., 2013.
- [99] S. Zhao, W. P. Fung-Leung, A. Bittner, K. Ngo, and X. Liu, “Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells,” *PLoS One*, vol. 9, no. 1, article e78644, 2014.
- [100] K. K. de Silva, J. M. Dunwell, and A. M. Wickramasuriya, “Weighted Gene Correlation Network Analysis (WGCNA) of *Arabidopsis* somatic embryogenesis (SE) and identification of key gene modules to uncover SE-associated hub genes,” 2022.