Exploiting plant transcriptomic databases: Resources, tools, and approaches

Peng Ken Lim¹, Xinghai Zheng¹, Jong Ching Goh¹ and Marek Mutwil¹,*
¹School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Singapore
*Correspondence: Marek Mutwil (mutwil@ntu.edu.sg)
https://doi.org/10.1016/j.xplc.2022.100323

ABSTRACT

There are now more than 300,000 RNA sequencing samples available, stemming from thousands of experiments capturing gene expression in organs, tissues, developmental stages, and experimental treatments for hundreds of plant species. The expression data have great value, as they can be re-analyzed by others to ask and answer questions that go beyond the aims of the study that generated the data. Because gene expression provides essential clues to where and when a gene is active, the data provide powerful tools for predicting gene function, and comparative analyses allow us to study plant evolution from a new perspective. This review describes how we can gain new knowledge from gene expression profiles, expression specificities, co-expression networks, differential gene expression, and experiment correlation. We also introduce and demonstrate databases that provide user-friendly access to these tools.

Key words: gene expression, comparative transcriptomics, gene function, co-expression, differential expression, databases

Lim P.K., Zheng X., Goh J.C., and Mutwil M. (2022). Exploiting plant transcriptomic databases: Resources, tools, and approaches. Plant Comm. 3, 100323.

INTRODUCTION

Plants are a vital crop source of food and feed and have tremendous ecological significance because of their monumental roles as primary producers in the terrestrial ecosystem (Friend, 2010). Furthermore, they are a valuable library of medicinal secondary metabolites that can be adapted for therapeutic use (Hussain et al., 2012; Rao and Ravishankar, 2002). Because plants are so central to our survival and well-being, we need a greater understanding of plant genes, which will allow us to improve desirable traits (Nowicka et al., 2018; Niazian, 2019). For example, the study of plant genes involved in secondary metabolism can elucidate biosynthetic pathways of medicinal compounds, which can be transferred to suitable expression hosts for large-scale production and modification (Paddon et al., 2013; Cravens et al., 2019; Sabzehzari et al., 2020). However, the characterization of plant gene function is a slow process, typically limited to a handful of model plants. This is mainly because of the high redundancy, complexity, and large size of plant genomes, lack of genetic transformation options, long generation time, and difficulty in cultivation (Kondou et al., 2010; Bolle et al., 2013; Xuan et al., 2021). In addition, many novel and unique genes discovered in non-model plants cannot be functionally annotated by sequence homology, as many plant genes are unique and do not show sequence similarity to characterized genes (Hamilton and Buell, 2012). Furthermore, functional annotation of plant genes with characterized homologs can still be difficult if they belong to large and functionally diverse gene families (He and Zhang, 2005). Consequently, many bioinformatic approaches have been developed that use genomic, transcriptomic, metabolomic, and phenotypic data to address the difficulties in gene function prediction (Radivojac et al., 2013; Rhee and Mutwil, 2014; Hansen et al., 2018).

Most gene function prediction approaches are based on the principle of guilt by association, which states that genes with similar characteristics (e.g., sequence or expression) are likely to have the same function (Rhee and Mutwil, 2014). For example, two genes that show similar gene expression profiles across different organs and tissues are likely to be part of the same protein complex (Persson et al., 2005) or biosynthetic pathway (Delli-Ponti et al., 2021) or to be targeted to the same subcellular compartment (Ryngajilo et al., 2011). As publicly available gene expression data are growing at a near-exponential rate, many online databases with diverse functionalities and plant species have emerged as valuable contributions to the scientific community. Here, we provide a guide to the various concepts and databases that can be used for gene function prediction and as electronic evidence to support experimental results (Figure 1; Table 1). Because of the large number of online gene expression databases, we cover...
only those that offer unique approaches for studying gene function.

**Studying gene function with gene expression profiles**

Because most genes are expected to be expressed only when needed, gene expression profiles are useful in predicting gene function. For example, a gene expressed specifically in pollen is likely to be important for pollen function (Bernal et al., 2008), whereas another gene strongly expressed during heat stress is likely to be necessary for survival during heat. Numerous studies and databases comprise customized expression compendia for answering different questions (Table 1). For example, GENEVESTIGATOR’s anatomy compendium contains 10 562 samples from 127 different anatomical parts of *Arabidopsis thaliana* (Hruz et al., 2008), showing the average expression values of each gene in diverse organs, tissues, and treatments (Table 1). The expression data in the ePlant database are also arranged into multiple compendia that capture various (abiotic) stresses.

To demonstrate how gene expression profile analyses can reveal gene function, we will use *Agamous* (AG, AT4G18960), a floral homeotic gene that specifies floral meristem and carpel and stamen identity (Yanofsky et al., 1990). The ePlant database (https://bar.utoronto.ca/eplant/) (Waese et al., 2017) showcases expression patterns of a single gene as electronic fluorescent pictographs. Each gene expression profile generated is based on a high-quality transcriptome dataset that has been thoughtfully selected from one or two studies. The Developmental Map of AG shows strong expression in stamens and carpels at flowering stages 12 and 15 of *A. thaliana*, which is in line with the function of AG (Figure 2A). The advantage of this expression visualization is that the pictures provide an easy-to-understand anatomical context for the different samples. On the other hand, boxplots and bar charts are typically used to indicate the average/median expression of a single or a few genes over multiple samples. CoNekT-Plants (https://conekt.sbs.ntu.edu.sg/) (Proost and Mutwil, 2018) uses bar charts to indicate the mean expression, and the points above and below the bar indicate the maximum and minimum expression in a given organ, respectively (Figure 2B). The boxplots in GENEVESTIGATOR depict the mean expression, and whiskers indicate the distribution of the data with standard error (Figure 2C). Heatmaps are two-dimensional (or even three-dimensional; Fernandez-Pozo et al., 2020) grids suited to visualizing gene expression of multiple genes and conditions simultaneously. Typically, genes are arranged in rows, and organs, tissues, cell types, and treatments are shown in columns. The cells in the heatmap usually indicate the average expression of a gene in a given condition through the use of colors, numbers, or both. Although heatmaps cannot directly show the distribution of the data as bar charts and boxplots, they enable the plotting of dense, data-rich figures. For example, the AG expression profile in the GENEVESTIGATOR Condition Search in the Anatomy function uses a heatmap to show its expression profile in 127 different anatomical parts (Figure 2D). In addition, the tool shows the number of samples, and the users can also click on a specific organ, which will give more details on the experiments from which the sample data were obtained.

**Single-cell RNA-seq**

Single-cell RNA sequencing (scRNA-seq) allows researchers to discover complex and novel cell populations and track the
## Databases Notable onboard functionalities/statistical methods

### Gene expression profile analyses

| Databases                  | Notable onboard functionalities/statistical methods                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ePlant by BAR              | gene expression profiles as eFP heatmaps  
|                            | each eFP is generated from gene expression data from a few high-quality studies  
|                            | scRNA-seq gene expression data for *A. thaliana* roots                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Expression Atlas           | gene expression profiles as heatmaps  
|                            | scRNA-seq gene expression data are available for *A. thaliana*, *Oryza sativa*, *S. lycopersicum*, and *Zea mays*                                                                                                                                                                                                                                                                                                                                                                                                                           |
| GENEVESTIGATOR             | gene expression profiles in the form of boxplots, scatterplots, or heatmaps  
|                            | conditions in gene expression profiles are amalgamations of similar samples across multiple studies, with the exception of perturbation conditions  
|                            | enables gene expression profiles to be generated from user-defined datasets                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| CoNekT-Plants              | gene expression profiles as bar charts and heat maps  
|                            | conditions in gene expression profiles are amalgamations of similar samples across multiple studies  
|                            | gene expression profiles for most plants predominantly feature only anatomical conditions  
|                            | cross-species comparative gene expression profile analyses available between user-defined gene sets or within gene families                                                                                                                                                                                                                                                                                                                                                           |

### Co-expression analyses

| Databases                  | Notable onboard functionalities/statistical methods                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Expression Angler by BAR   | hosts thematic co-expression data from well-curated samples of relevant studies  
|                            | uses PCC as a co-expression metric                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| ATTED-II                   | hosts comprehensive co-expression data amalgamated from both RNA-seq and microarray samples  
|                            | hosts thematic co-expression data  
|                            | uses MR as a co-expression metric  
|                            | enables network visualization of gene co-expression neighborhoods                                                                                                                                                                                                                                                                                                                                                                                                               |
| CoNekT-Plants              | hosts comprehensive co-expression data from RNA-seq samples  
|                            | uses HRR as a co-expression metric  
|                            | enables GO enrichment analyses on co-expression clusters  
|                            | enables comparative network analyses between species based on gene family information  
|                            | enables network visualization of gene co-expression neighborhoods and clusters                                                                                                                                                                                                                                                                                                                                                                                                     |

### DGE analyses

| Databases                  | Notable onboard functionalities/statistical methods                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| AtCAST                     | enables users to search for control/treatment DGE comparisons based on single- or multi-gene query  
|                            | features onboard functionalities to extract DEGs from DGE comparisons based on user-defined fold change and p value  
|                            | control and treatment samples in comparisons are predefined                                                                                                                                                                                                                                                                                                                                                                                                               |
| Expression Atlas           | enables users to search for control/treatment DGE comparisons based on single- or multi-gene query  
|                            | features onboard functionalities to extract DEGs from DGE comparisons based on user-defined fold change and p value  
|                            | control and treatment samples in comparisons are predefined                                                                                                                                                                                                                                                                                                                                                                                                               |
| GENEVESTIGATOR             | enables users to search for control/treatment DGE comparisons based on single- or multi-gene query  
|                            | features onboard functionalities to extract DEGs from DGE comparisons based on user-defined fold change and p value  
|                            | control and treatment samples in comparisons are user defined                                                                                                                                                                                                                                                                                                                                                                                                               |

### Gene expression-specificity/stability analyses

| Databases                  | Notable onboard functionalities/statistical methods                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Rice Expression Database   | enables housekeeping and specific genes to be identified based on the stability of expression across different conditions  
|                            | stability is calculated using \( \tau \) value                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| CoNekT-Plants              | features tool to identify specific genes in tissue of choice  
|                            | specificity of gene expression in tissue is calculated from SPM, Tau, and entropy  
|                            | enables cross- or within-species comparison of different tissue-specific gene sets based on gene family information                                                                                                                                                                                                                                                                                                                                                   |
| GENEVESTIGATOR             | features tool to identify housekeeping genes based on the stability of expression across a set of user-defined conditions  
|                            | stability is calculated based on expression variance score                                                                                                                                                                                                                                                                                                                                                                                                               |

### Experiment correlation analyses

| Databases                  | Notable onboard functionalities/statistical methods                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| AtCAST                     | enables users to search for correlated experiments based on an experimental condition or user-uploaded gene expression data  
|                            | relationships among experiments can be visualized as a network                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

---

*Table 1. Summary of the featured databases (Continued on next page)*
developmental pathways of distinct cell types (Hwang et al., 2018). Several databases containing scRNA-seq data exist, such as ePlant, Root Cell Atlas (http://wanglab.sippe.ac.cn/rootatlas; Zhang et al., 2019), The Plant scRNA-Seq Browser (https://www.zmbp-resources.uni-tuebingen.de/timmermans/ plant-single-cell-browser; Ma et al., 2020), and the EBI Expression Atlas (https://www.ebi.ac.uk/gxa/sc/home) (Kapushesky et al., 2010). scRNA-seq datasets can be visualized as t-Distributed Stochastic Neighbor Embedding (tSNE) (Kobak and Berens, 2019) or Uniform Manifold Approximation and Projection (UMAP) (Becht et al., 2018) plots, which enable the identification of subpopulations of transcriptionally similar cells (Shulse et al., 2019). The subpopulations within the data can be identified by clustering (e.g., k value in the EBI Expression Atlas) or from other metadata (inferred cell type, genotype).

To demonstrate how single-cell gene expression profiles can reveal gene function, we used CSLD3 (AT3G03050), a gene that functions to initiate root hair development (Hu et al., 2018), as a query in the Single Cell eFP tool of ePlant to generate a tSNE plot (Figure 2E). From the tSNE plot, the expression of CSLD3 is highest in root hair cells relative to other root cell types, which confirms its specialized function in root hair cells. Furthermore, within the different subpopulations of the root hair cells, CSLD3 has the highest expression in root hair cells at the early stage of differentiation, which is in line with the gene’s function in root hair initiation (Figure 2E) (Bernal et al., 2008, p. 2). This insight could not have been gained from typical expression analysis of roots and highlights the utility of scRNA-seq and the need for this type of data for other tissues, organs, and species.

### Table 1. Continued

| Databases                  | Notable onboard functionalities/statistical methods                                      |
|----------------------------|-----------------------------------------------------------------------------------------|
| ATCAST                     | hosts species data for A. thaliana only                                                 |
| Sasaki et al., 2011        | database: http://atbpsm.yokohama-cu.ac.jp/cgi/atcast/home                               |
| ATTED-II                   | hosts species data for A. thaliana, Brassica rapa, G. max, M. truncatula, O. sativa, V. vinifera, S. lycopersicum, Z. mays, and Populus trichocarpa |
| Obayashi et al., 2009      | database: https://atted.jp/                                                              |
| CoNekT-Plants              | hosts species data for A. trichopoda, A. thaliana, Chlamydomonas reinhardtii, Cyanophora paradoxa, G. biloba, Marchantia polymorpha, O. sativa, Physcomitrella patens, Picea abies, Selaginella moellendorfii, V. vinifera, S. lycopersicum, and Z. mays |
| Proost and Mutwil, 2018    | database: https://www.zmbp-resources.uni-tuebingen.de/timmermans/                       |
| ePlant by BAR              | hosts species data for A. thaliana and P. trichocarpa                                    |
| Waese et al., 2017         | database: http://bar.utoronto.ca/                                                       |
| Expression Angler by BAR   | hosts species data for A. thaliana                                                     |
| Toufighi et al., 2005      | database: http://bar.utoronto.ca/                                                       |
| Expression Atlas           | hosts species data for A. thaliana, Brachypodium distachyon, G. max, H. vulgare, M. truncatula, O. sativa, Sorgum bicolor, S. lycopersicum, S. tuberosum, T. aestivum, V. vinifera, and Z. mays |
| Kapushesky et al., 2010    | database: https://www.ebi.ac.uk/gxa/home                                                 |
| GENEVESTIGATOR             | hosts species data for A. thaliana, Brassica oleracea, G. max, H. vulgare, M. truncatula, Nicotiana tabacum, O. sativa, P. patens, S. bicolor, S. lycopersicum, T. aestivum, and Z. mays |
| Hruz et al., 2008          | client download: https://genevestigator.com/gv/start/start.jsp                          |
| Rice Expression Database   | hosts species data for O. sativa only                                                    |
| Xia et al., 2017           | database: http://expression.ic4r.org/                                                    |

### Additional information

DGE, differential gene expression; eFP, electronic pictograph; PCC, Pearson’s correlation coefficient; MR, mutual rank; HRR, highest reciprocal rank; GO, Gene Ontology; DEGs, differentially expressed genes; SPM, specificity measure.

---

**Exploiting plant transcriptomic databases**

Expression of multiple genes can be compared within and across species, typically as two-dimensional heatmaps in which rows and columns are configured to address a specific question. The comparative aspect of heatmaps can be used to identify functionally equivalent genes across species. For example, the CoNekT-Plants comparative heatmap viewer for the orthogroup containing AG presents genes from multiple species in rows and their expression values in different organs in columns (Figure 3A). As expected, AG shows predominantly flower-specific expression in A. thaliana, whereas STK, AGL1, and SHP2 show seed-specific expression, consistent with their function in seed development (Ehlers et al., 2016; Paolo et al., 2021). Ginkgo biloba contains one ortholog of AG (Gb_16301), which is also specifically expressed in the strobilus, a reproductive organ from the Flower category in CoNekT-Plants. Interestingly, Amborella trichopoda contains two orthologs of AG, of which only AMTTR_s00021p00254030 shows flower-specific expression (Figure 3A). Thus, based on the comparative heatmap, AMTTR_s00021p00254030, but not AMTTR_s00071p00193200, is functionally equivalent to AG from A. thaliana.

Gene expression information can be combined with phylogenetic trees to better understand gene function and evolution. Navigator
Figure 2. Methods to visualize gene expression profiles, demonstrated using AG (At4G18960).

(A) ePlant: Plant Viewer eFP. Each organ is colored to indicate gene expression level; yellow and red colors indicate low and high transcripts per million (TPM) values, respectively.

(legend continued on next page)
Plant Communications

Exploiting plant transcriptomic databases

Users have control over visualization and functional control on this page, enabling them to adjust the amount of descriptive information (e.g., full gene names, Kyoto Encyclopedia of Genes and Genomes [KEGG] annotations) and to select options for ranking co-expressed genes based on different transcriptome datasets. The top 10 genes for A. thaliana include SEP1/2/3 and AP3, which are also involved in flower development (Hugouvieux et al., 2018; Jetha et al., 2014; Parenicová et al., 2003). By contrast, CoNekT-Plants allows users to download the co-expression gene list as a text file for offline viewing by clicking a button in the Coexpression Networks section (Supplemental Figure 1B, red circle) on the gene page. Expression Angler from BAR (Toufighi et al., 2005) offers a choice of multiple gene expression datasets and an interactive interface for adjusting the r-value threshold (http://bar.utoronto.ca/ExpressionAngler/?agi_id=AT4G18960&match_count=25&active_view=Developmental%20Map). In addition to the typical list of genes, the tool also displays an expression profile heatmap, enabling users to further evaluate the expression similarity (Supplemental Figure 1C).

Different strategies for generating co-expression data

The databases use different strategies for generating co-expression data. These differences include the selection of the transcriptome dataset, gene expression quantification and normalization, batch correction, sample balancing, and statistical metrics used to measure co-expression, which further add to the degrees of freedom in which co-expression analyses can be implemented. Currently, there is a lack of consensus on the best strategy, as different approaches are suited to answering different biological questions (reviewed in Rao and Dixon, 2019; Serin et al., 2016; Usadel et al., 2009), and this translates into poor agreement among the databases.

Expression Angler uses Pearson’s correlation coefficient (PCC) to measure the linear correlation between gene pairs across samples. However, PCC can be sensitive to outliers, and false correlations between otherwise uncorrelated genes can be inferred if they have a high expression value in a single sample (Rao and Dixon, 2019; Usadel et al., 2009). To mitigate sensitivity to outliers (Obayashi and Kinoshita, 2009), ATTED-II and CoNekT-Plants rank genes based on PCC scores to generate directional rankings between gene pairs. Because directional rankings within a gene pair are likely to be asymmetrical (e.g., gene B ranks highest within gene A’s list, whereas gene A ranks second within gene B’s list), ATTED-II takes the geometric mean of the two to produce a mutual rank (MR) score. CoNekT-Plants uses the highest reciprocal rank (HRR), which retains the higher of the two ranks within each gene pair to measure co-expression (Mutwil et al., 2010).

CoNekT-Plants uses a condition-independent approach in which the transcriptome dataset consists of all available RNA sequencing (RNA-seq) samples that passed quality control. By

(B) CoNekT-Plants. The samples and gene expression level are shown on the x and y axis, respectively. The bars and points indicate the mean and maximum/minimum expression.

(C) GENEVESTIGATOR developmental atlas. Points indicate mean values, and whiskers indicate standard errors.

(D) GENEVESTIGATOR: anatomy search tool. The different organs, tissues, and cell types have been arranged into logical hierarchies. The gene expression levels are indicated by a heatmap, where white and dark beige colors indicate low and high expression, respectively.

(E) Visualization of scRNA-seq data. Single-cell RNA-seq of A. thaliana cellulose synthase-like D2 (AT3G03050) visualized in ePlant. Each point depicts a single cell, and the low and high expression value of the gene are represented by yellow and red colors, respectively.

6 Plant Communications 3, 100323, July 11 2022 © 2022 The Author(s).
contrast, Expression Angler utilizes a condition-dependent approach and hosts co-expression data built from different thematic transcriptome datasets. For example, the Developmental Map dataset consists of carefully selected microarray experiments from *A. thaliana* samples at different stages of development (Nakabayashi et al., 2005; Schmid et al., 2005), and the Abiotic Stress dataset contains *A. thaliana* data capturing different abiotic stresses (Kilian et al., 2007). ATTED-II offers the

![Figure 3. Comparative gene expression analyses.](image)

(A) Comparative expression heatmap of AG. Rows correspond to genes, and columns represent organs. Low and high expression values are represented by green and red cells, respectively. Black cells indicate missing data. Each row has been scaled to have one as the highest value.

(B) ePlant Navigator viewer. The cladogram captures phylogenetic relationships among genes, and the bars depict sequence and expression similarities of the genes to the query (AG).

(C) Phylogenetic tree of the AG orthogroup. The different species are represented by color-coded leaves (species without expression data are black). The expression levels in the organs are visualized as a heatmap, where low and high expressions are indicated by yellow and dark blue colors, respectively. The different clades and their expression are indicated by colored boxes. The species are indicated by gene names: Zm (*Zea mays*), LOC (*Oryza sativa*), Bradi (*Brachypodium distachyon*), Zosma (*Zosteria marina*), GS (*Vitis vinifera*), Sol (*Solanum lycopersicum*), AM (*Amborella trichopoda*), MA (*Picea abies*), and Gb (*Ginkgo biloba*).
Plant Communications

best of both worlds, as it caters to both approaches; it aggregates both RNA-seq and microarray samples to generate a Universal co-expression dataset, and it also features co-expression data from thematic condition-dependent datasets (e.g., Tissue dataset, Hormones dataset).

These methodological and statistical differences may produce different results when the same gene is queried in the three databases. As an example of this discrepancy, the gene lists of the top 50 genes co-expressed with AG according to the three databases (CoNekT-Plants-Plants, ATTED-II; Universal dataset, Expression Angler; Developmental Map dataset) share only two common genes (data not shown). These two genes are SEP1 (AT5G15800) and SEP2 (AT3G02310). This relatively poor agreement between co-expression networks has been observed before (Jaccard index of ~0.1 between different versions of A. thaliana networks; Obayashi et al., 2018). Therefore, although co-expression clearly works, it is likely that different databases will return different (to a degree) co-expression lists and networks. Consequently, we advise comparing results from multiple databases.

Visualization and analyses using co-expression networks

A co-expression network typically features genes as nodes and co-expression relationships between genes as edges between nodes (Usadel et al., 2009). The topology of the co-expression network enables powerful analyses based on network theory to extract additional insights. For example, genes with more connections (higher degree) tend to be more essential and produce more severe phenotypes when knocked out (Mutwil et al., 2010). Betweenness centrality (the degree to which nodes are shortest-path connectors) has also been used to identify hub genes that are highly connected in the co-expression network and functionally important to biological processes (van Dam et al., 2018). Owing to the hierarchical organization of gene-to-gene relationships inherent in complex biological systems (e.g., regulator-regulatee, upstream-downstream effector relationships), co-expression networks have a heterogeneous topology, with regions of highly interconnected genes, called clusters or modules (Usadel et al., 2009; Ruprecht et al., 2016). These clusters have been shown to represent genes in important biological processes (van Dam et al., 2018). The visualization of co-expression networks offers a more realistic and informative representation of co-expression relationships than gene lists, as networks display all pairwise relationships (Jupiter and VanBuren, 2008).

Visualization of a whole co-expression network is often impractical, as the networks often contain tens of thousands of nodes and several orders of magnitude more edges. ATTED-II and CoNekT-Plants allow visualizations of co-expression neighborhoods (genes closely connected to GOIs) and also map more functional data onto the networks by modifying the styles of nodes and edges. CoNekT-Plants enables users to view neighborhoods via the CoexpressionNetworks section of gene pages, which shows co-expression relationships among all genes co-expressed with AG (Figure 4A). In CoNekT-Plants, node color and shape depict orthogroups, and edges represent co-expression relationships (HRR cutoff ≤100). These networks are dynamic and can be adjusted to highlight important parts of the network; the colors and shapes of nodes can be changed to represent orthogroups or Pfam domains, and edge colors can be toggled to indicate HRR scores. The neighborhood networks in ATTED-II are built from multiple neighborhood levels of AG and indicate protein-protein interactions with red dotted lines (Figure 4B).

Identification of biological pathways with co-expression networks

Co-expression networks can be used to reveal known and novel pathways by identifying groups of highly connected genes (clusters/modules). For example, CoNekT-Plants partitions co-expression networks into co-expression modules with the heuristic cluster chiseling algorithm (HCCA) (Mutwil et al., 2010). The functions of these modules can be predicted by over-representation (enrichment) analyses of biological functions captured by, e.g., Gene Ontology (GO; Ashburner et al., 2000) or MapMan terms (Thimm et al., 2004). For example, AG belongs to cluster 69 of A. thaliana (https://conekt.sbs.ntu.edu.sg/cluster/view/453; links to cluster pages can be accessed via gene pages), which is enriched for genes involved in flower developmental processes (e.g., stamen development, petal development) and auxin response (Figure 4A). This agrees with the known function of AG as a transcription factor that specifies floral meristem, carpel, and stamen identity and highlights the link between auxin homeostasis and flower development (Yamaguchi et al., 2017). The identity of other co-expression clusters involved in stamen development across multiple species can be revealed by, e.g., clicking on the ”GO:0048443 (Stamen development)” link on the cluster page or using the “Find enriched clusters” tool (https://conekt.sbs.ntu.edu.sg/search/enriched/clusters), with which one can find co-expression clusters involved in any process captured by GO. Thus, co-expression offers a powerful hypothesis-generating tool for predicting gene function and identifying new genes involved in the biological process of interest.

Comparative co-expression network analyses

Similarly, to conserve gene expression profiles across multiple species (Figure 3), regions of co-expression networks can be conserved even across vast evolutionary distances (Mutwil et al., 2011; Ferrari and Mutwil, 2019). Comparative co-expression network analyses enable the transfer of functional knowledge from model to non-model species (Movahedi et al., 2012; Sibout et al., 2017) and the study of biological pathway evolution (Ruprecht et al., 2017; Ferrari et al., 2020). Furthermore, because conserved co-expression relationships tend to be functionally significant (Hansen et al., 2014), a comparative analysis can remove potentially irrelevant co-expression relationships.

Using AG, we demonstrate how functional information can be transferred between species. AG is a MADS domain transcription factor, and CoNekT-Plants identified 413 other genes in the same gene family (orthogroup) across 13 different Viridiplantae (https://conekt.sbs.ntu.edu.sg/family/view/12). CoNekT-Plants uses orthogroups to score similarities between clusters, as similar clusters are expected to contain a high number of identical
orthogroups. The “Expression context conservation (ECC)” page for AG (https://conekt.sbs.ntu.edu.sg/sequence/view/46111) lists all genes in the AG orthogroup with similar co-expression neighborhoods https://conekt.sbs.ntu.edu.sg/ecc/graph/46111/2). This list can be sorted by the ECC value (Jaccard index orthogroup similarity of cluster pairs) to suggest functionally equivalent genes across species or duplicated gene modules within species (Delli-Ponti et al., 2021; Ruprecht et al., 2016). In the cluster page for cluster 69 of A. thaliana (https://conekt.sbs.ntu.edu.sg/cluster/view/453), the Similar Clusters section allows one to visually compare clusters in a network view. A comparison between cluster 69 from A. thaliana and cluster 51 from V. vinifera with a Jaccard index of 0.038 (https://conekt.sbs.ntu.edu.sg/graph_comparison/cluster/453/6929/2) identified genes encoding transcription factors involved in floral development (Cheng et al., 2009; Gross et al., 2018), such as MYB21 (AT3G27810), MYB24 (AT5G40350), MYB57 (AT3G01530), and CRC (AT1G69180) (Figure 4C). Because the grape cluster contains a similar set of genes to the A. thaliana cluster, it is likely that the genes in the grape cluster are also involved in flower development.

Predicting gene function with differential gene expression analysis

A classical treatment/control experiment can reveal hundreds of differentially expressed genes (DEGs) and can identify genes important for the plant response to the treatment. For example, genes that are upregulated during cold treatment may be necessary for surviving freezing temperatures (To et al., 2011, p. 6). Thus, identifying these DEGs can rapidly dissect the mechanism of freezing tolerance.

Expression Atlas (Supplemental Figure 2A) can identify experimental conditions in which a GOI is significantly (indicated by p value) up- or downregulated (indicated by log-fold change). AG is significantly upregulated in the comparison of the clf28 mutant versus wild type, consistent with the fact that CLF28 is necessary for AG repression (Križek et al., 2006). Users can also click on clf28 mutant versus wild type in Shoot to view the experimental information and identify all DEGs specific to this comparison. Similarly, AtCAST (Sasaki et al., 2011) revealed that AG is downregulated (negative signal ratio) in its mutant background (ag-12) and in the lfy-12 mutant, which supports the activation of AG by LFY (Supplemental Figure 2B) (Busch et al., 1999).

In addition, Expression Atlas allows the users to specify a set of genes (e.g., members of a biological pathway), which are then matched to the precalculated DEGs from all experiments in the database (Supplemental Figure 2C). Thus, users can rapidly identify experimental conditions that cause the largest expression changes in their genes of interest. The database...
Plant Communications

ranks the matches by the observed/expected ratio, i.e., the ratio between the observed intersection of the user gene set with the differentially expressed genes (DEGs) in an experiment and the intersection expected by chance. Although the above tools use predefined control/treatment experimental groups, GENVESTIGATOR allows users to define the control-group and treatment-group experiments and reveals the corresponding upregulated and downregulated gene sets (Supplemental Figure 2D).

Identifying housekeeping reference genes

Genes expressed at relatively constant levels are often involved in maintaining basal cellular functions that are essential for the existence of a cell (Gutierrez et al., 2008; Lin et al., 2014). These genes, termed reference genes or housekeeping genes, are often used as normalization controls in transcript quantification assays such as qRT-PCR (e.g., GAPDH or ACTB). However, an over-reliance on an all-purpose group of commonly used housekeeping genes for normalization may not be ideal because their expression, although stable across some conditions, is not universally stable across all conditions (Kozera and Rapacz, 2013; Joseph et al., 2018). Thus, it is vital to identify housekeeping genes that are specific to the experimental context to address this issue.

To that end, the RefGenes tool of GENVESTIGATOR enables the identification of context-specific housekeeping genes with the highest expression stability across a user-defined set of conditions (Supplemental Figure 3A). Users can also filter candidate reference genes based on their expression level, which is essential when choosing reference genes for qRT-PCR (Kozera and Rapacz, 2013). Similarly, Rice Expression Database (Xia et al., 2017) also allows users to extract housekeeping genes using \( r \) values which range from 0 to 1 (Supplemental Figure 3B).

Predicting gene function by specificity analyses

On the opposite end of the spectrum from housekeeping genes, specific (selective) genes are expressed exclusively in a particular organ, tissue, developmental stage, or treatment. Because organ-specific genes are likely to be important for the organ’s function (e.g., flower-specific AG is important for flower development; Figure 2) (Julca et al., 2021), extracting and identifying specifically expressed genes is a powerful method for predicting gene function.

CoNekT-Plants uses the specificity measure (SPM; ranges between zero and one, where one indicates that the gene is exclusively expressed in the tissue), Tau (high values indicate that a profile is specified in a tissue), and entropy (indicates how much a profile fluctuates across all tissues, and genes with very specific or very stable expression have low entropy) to identify specific genes. Users can query a certain condition or tissue type in the Specific Profiles tool, which will provide a gene list with information about the SPM score, entropy score, and Tau score of specific genes. Users can also compare two lists of specifically expressed genes within or across species in the Compare Specificities tool (https://conekt.sbs.ntu.edu.sg/specificity_comparison/). For example, by selecting Arabidopsis thaliana/Tissue Specificity/Flower/SPM 0.85 and Solanum lycopersicum/Tissue Specificity/Flower/SPM 0.85 and clicking Compare Specificity, one obtains 107 orthogroups specifically expressed in the flowers of the two species (Supplemental Figure 4A). The list of the orthogroups contains A. thaliana AG, its four orthologs from tomato, and, e.g., MYB57, which is essential for stamen development (Cheng et al., 2009, p. 21).

Similar to CoNekT-Plants, GENVESTIGATOR contains a Gene Search tool that can help to identify specific genes. This tool offers a greater degree of freedom in interrogating the database for specific genes, as users can define a set of experiments as background and choose a subset of experiments in which to identify specifically expressed genes (Supplemental Figure 4B).

Experiment correlation tools

When the gene expression profiles of two independent experiments are highly correlated, we can infer that the treatments, genotypes, organs, or other experimental variables are related. Thus, the genome-wide comparison of gene expression patterns of experiments can uncover novel biological relationships. At-CAST (Sasaki et al., 2011) collects publicly available A. thaliana gene expression data and provides a list of experiments that are positively and negatively correlated to the query experiment. To mitigate the noise between gene expression data from different studies, the tool only compares the expression profiles of genes with significant expression changes (termed “module” according to the authors). Here, an RNA-seq experiment on the AG mutant (ag-12) is strongly correlated with other samples from LFY, UFO, and AP3 mutants (Supplemental Figure 5A), consistent with the shared functions of these genes in flower development (Ng and Yanofsky, 2001).

Users can also upload their RNA-seq or microarray data to screen the compendium for similar experiments. The system also provides a graph-based network to visualize the relationship of experiments in the table (Supplemental Figure 5B). Furthermore, users can click the More Info button behind each experiment in the table to view its module genes (Supplemental Figure 5C).

CONCLUDING REMARKS

In order to uncover the functions of plant genes, a multitude of analytical methods have been developed to interrogate the rapidly growing publicly available gene expression data. To make the outcome of these methods readily available to non-bioinformaticians, the bioinformatics community has produced various plant gene expression databases. In this review, we have discussed the principles behind the different gene expression analyses and demonstrated the databases that offer these analyses. Thus, the ubiquity and usefulness of gene expression data and the plethora of online resources provide powerful gene function prediction tools that plant biologists should incorporate into their research.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at Plant Communications Online.
Exploiting plant transcriptomic databases

FUNDING
This review was supported by a Singapore Ministry of Education tier-two grant (MOE2018-T2-2-053) and a Singapore Food Agency grant (SFS_RND_SUPF_001_05).

ACKNOWLEDGMENTS
No conflict of interest is declared.

Received: January 10, 2022
Revised: March 3, 2022
Accepted: April 6, 2022
Published: April 9, 2022

REFERENCES
Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat. Genet. 25:25–29. https://doi.org/10.1038/75556.

Becth, E., McInnes, L., Healy, J., et al. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. https://doi.org/10.1038/nbt.4314.

Bernal, A.J., Yoo, C.M., Mutwil, M., Jensen, J.K., Hou, G., Blaukopf, C., Sörensen, I., Biancofiori, E.B., Scheller, H.V., and Willats, W.G.T. (2008). Functional analysis of the cellulose synthase-like genes CSLD1, CSLD2, and CSLD4 in tip-growing arabidopsis cells. Plant Physiol. 148:1238–1253. https://doi.org/10.1104/pp.108.121939.

Bolle, C., Huep, G., Kleinbolting, N., Haberer, G., Mayer, K., Leister, D., and Weishhaar, B. (2013). GABI-DUPLIO: a collection of double mutants to overcome genetic redundancy in Arabidopsis thaliana. Plant J. 75:157–171. https://doi.org/10.1111/tjpl.12197.

Brown, D.M., Zeef, L.A.H., Ellis, J., Goodacre, R., and Turner, S.R. (2005). Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. Plant Cell 17:2281–2295. https://doi.org/10.1105/tpc.105.031542.

Busch, M.A., Bombles, K., and Weigel, D. (1999). Activation of a floral homeotic gene in Arabidopsis. Science 285:585–587. https://doi.org/10.1126/science.285.5427.585.

Caputi, L., Franke, J., Farrow, S.C., Chung, K., Payne, R.M.E., Nguyen, T.-D., Dang, T.-T.T., Carqueijeiro, I.S.T., Koudounas, K., Bernonville, T.D., et al. (2018). Missing enzymes in the biosynthesis of the anticancer drug vinbliniste in Madagascar periwinkle. Science 360:1235–1239. https://doi.org/10.1126/science.aat4100.

Cheng, H., Song, S., Xiao, L., Soo, H.M., Cheng, Z., Xie, D., and Peng, J. (2009). Gibberellin acts through jasmonate to control the expression of MYB21, MYB24, and MYB57 to promote stamen filament growth in Arabidopsis. PLoS Genet. 5:e1000440. https://doi.org/10.1371/journal.pgen.1000440.

Cravens, A., Payne, J., and Smolke, C.D. (2019). Synthetic biology strategies for microbial biosynthesis of plant natural products. Nat. Commun. 10:2142. https://doi.org/10.1038/s41467-019-09848-w.

Delli-Ponti, R., Shivhare, D., and Mutwil, M. (2021). Using gene expression to study specialized metabolism as a practical guide. Front. Plant Sci. 11:2074. https://doi.org/10.3389/fpls.2020.625035.

Ehlers, K., Bhide, A.S., Tekleyohans, D.G., Wittkop, B., Snowdon, R.J., and Becker, A. (2016). The MADS box genes ABS, SHP1, and SHP2 are essential for the coordination of cell divisions in ovule and seed coat development and for endosperm formation in Arabidopsis thaliana. PLoS One 11:e0165075. https://doi.org/10.1371/journal.pone.0165075.

Fernandez-Pozo, N., Haas, F.B., Meyberg, R., Ullrich, K.K., Hess, M., Perroud, P.-F., Hanke, S., Kratz, V., Powell, A.F., Vesty, E.F., et al. (2020). PEATmoss (Physcomitrella Expression Atlas Tool): a unified gene expression atlas for the model plant Physcomitrella patens. Plant J. 102:165–177. https://doi.org/10.1111/tpj.14607.

Ferrari, C., and Mutwil, M. (2019). Gene expression analysis of Cyanophora paradoxa reveals conserved abiotic stress responses between basal algae and flowering plants. New Phytol. https://doi.org/10.1111/nph.16257.

Ferrari, C., Shivhare, D., Hansen, B.O., Pasha, A., Esteban, E., Provart, N.J., Krangler, F., Fernie, A.R., Tohge, T., and Mutwil, M. (2020). Expression atlas of selaginella moellendorfii provides insights into the evolution of vasculature, secondary metabolism, and roots. Plant Cell, tpc.00780. https://doi.org/10.1105/tpc.19.00780.

Friend, A.D. (2010). Terrestrial plant production and climate change. J. Exp. Bot. 61:1293–1309. https://doi.org/10.1093/jxb/erq019.

Gao, Q., Xu, S., Zhu, X., Wang, L., Yang, Z., and Zhao, X. (2018). Genome-wide identification and characterization of the RIO atypical kinase family in plants. Genes Genom. 40:669–683. https://doi.org/10.1007/s13258-018-0658-4.

Gross, T., Broholm, S., and Becker, A. (2018). CRABS CLAW acts as a bifunctional transcription factor in flower development. Front. Plant Sci. 9:835. https://doi.org/10.3389/fpls.2018.00835.

Gutierrez, L., Mauriat, M., Pelloux, J., Bellini, C., and Van Wuytswinkel, O. (2008). Towards a systematic validation of references in real-time rt-PCR. Plant Cell 20:1734–1735. https://doi.org/10.1105/tpc.108.059774.

Hamilton, J.P., and Buell, C.R. (2012). Advances in plant genome sequencing. Plant J. 70:177–190. https://doi.org/10.1111/j.1365-313X.2012.04894.x.

Hansen, B.O., Meyer, E.H., Ferrari, C., Vaid, N., Movahedi, S., Vandepoele, K., Nikoloszi, Z., and Mutwil, M. (2018). Ensemble gene function prediction database reveals genes important for complex I formation in Arabidopsis thaliana. New Phytol. 217:1521–1534. https://doi.org/10.1111/nph.14921.

Hansen, B.O., Vaid, N., Musialak-Lange, M., Janowski, M., and Mutwil, M. (2014). Elucidating gene function and function evolution through comparison of coexpression networks of plants. Front. Plant Sci. 5. Article AUG. https://doi.org/10.3389/fpls.2014.00394.

He, B., Shi, P., Lv, Y., Gao, Z., and Chen, G. (2020). Gene coexpression network analysis reveals the role of SRS genes in senescence leaf of maize (Zea mays L.). J. Genet. 99:3.

He, X., and Zhang, J. (2005). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157–1164. https://doi.org/10.1534/genetics.104.037051.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv. Bioinformatics 2008:420747. https://doi.org/10.1155/2008/420747.

Hu, H., Zhang, R., Dong, S., Li, Y., Fan, C., Wang, Y., Xia, T., Chen, P., Wang, L., Feng, S., et al. (2018). AICS(LD3 and GhCSLD3 mediate root growth and cell elongation downstream of the ethylene response pathway in Arabidopsis. J. Exp. Bot. 69:1065–1080. https://doi.org/10.1039/jb/erx470.

Hugouvieux, V., Silva, C.S., Jourdain, A., Stigliani, A., Charras, Q., Conn, V., Conn, S.J., Carles, C.C., Parcy, F., and Zubieten, C. (2018). Tetramerization of MADS family transcription factors in flower development. Front. Plant Sci. 9:835. https://doi.org/10.3389/fpls.2018.00835.

Hussain, M.S., Fareed, S., Ansari, S., Rahman, M.A., Ahmad, I.Z., and Saeed, M. (2012). Current approaches toward production of

Plant Communications

Plant Communications 3, 100323, July 11 2022 © 2022 The Author(s). 11
Plant Communications

secondary plant metabolites. J. Pharm. Bioallied Sci. 4:10–20. https://doi.org/10.4103/0975-7406.92725.

Hwang, B., Lee, J.H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp. Mol. Med. 50:1–14. https://doi.org/10.1038/s12276-018-0071-8.

Jetha, K., Theißen, G., and Melzer, R. (2014). Arabidopsis SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. Nucleic Acids Res. 42:10927–10942. https://doi.org/10.1093/nar/gku755.

Joseph, J.T., Poolakkalody, N.J., and Shah, J.M. (2018). Plant reference genes for development and stress response studies. J. Biol. 43:173–187.

Julca, I., Ferrari, C., Flores-Tornero, M., Proost, S., Lindner, A.-C., Hackenberg, D., Steinbachová, L., Michaelidis, C., Gomes Pereira, S., Misra, C.S., et al. (2021). Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants. Nat. Plants 7:1143–1159. https://doi.org/10.1038/s41477-021-00958-2.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Parkinson, H., and Hartzel, A. (2010). Gene expression atlas at the European bioinformatics institute. Nucleic Acids Res. 38:D690–D698. Database issue. https://doi.org/10.1093/nar/gp936.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., Kapushesky, M., Emam, I., Holloway, E., Kurnosov, P., Zorin, A., and Emam, I., Holloway, E., Kurnosov, P., Zorin, A., et al. (2012). PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. Plant Cell 23:895–910. https://doi.org/10.1105/tpc.111.083667.

Mutwil, M., Usadel, B., Schütte, M., Loraine, A., Ebenhoh, O., and Persson, S. (2010). Assembly of an interactive correlation network for the Arabidopsis genome using a novel Heuristic Clustering Algorithm. Plant Physiol. 152:29–43. https://doi.org/10.1104/pp.109.145318.

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., and Nambara, E. (2005). Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed. Plant J. 41:697–709. https://doi.org/10.1111/j.1365-313X.2005.02337.x.

Ng, M., and Yanofsky, M.F. (2001). Activation of the Arabidopsis B class homeotic genes by APETALA1. Plant Cell 13:739–753. https://doi.org/10.1105/tpc.13.4.739.

Niazian, M. (2019). Application of genetics and biotechnology for improving medicinal plants. Planta 249:953–973. https://doi.org/10.1007/s00425-019-03099-1.

Obayashi, T., Aoki, Y., Tadaka, S., Kagaya, Y., and Kinoshita, K. (2018). ATTED-II in 2018: a plant coexpression database based on investigation of the statistical property of the mutual rank index. Plant Cell Physiol. 59:e3. https://doi.org/10.1093/pcp/pcx191.

Obayashi, T., Hayashi, S., Saeki, M., Ohta, H., and Kinoshita, K. (2009). ATTED-II provides coexpressed gene networks for Arabidopsis. Nucleic Acids Res. 37:D987–D991. https://doi.org/10.1093/nar/gkn807.

Obayashi, T., and Kinoshita, K. (2009). Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. DNA Res. 16:249–260. https://doi.org/10.1093/dnares/dsp016.

Oliver, S. (2000). Guilt-by-association goes global. Nature 403:601–603. https://doi.org/10.1038/35011165.

Paddock, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. Nature. https://doi.org/10.1038/nature12051.

Paolino, D., Rotasparti, L., Schnitter, A., Masiero, S., Colombo, L., and Mizzotti, C. (2021). The arabidopsis MADS-domain transcription factor SEEDSTICK controls seed size via direct activation of E2Fa. Plants (Basel, Switzerland) 10:192. https://doi.org/10.3390/plants10020192.

Parencová, L., de Folter, S., Kieffer, M., Horner, D.S., Favalli, C., Busscher, J., Cook, H.E., Ingram, R.M., Kater, M.M., Davies, B., et al. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell 15:1538–1551. https://doi.org/10.1105/tpc.011544.

Persson, S., Wei, H., Milne, J., Page, G.P., and Somerville, C.R. (2005). Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proc. Natl. Acad. Sci. U S A 102:8633–8638. https://doi.org/10.1073/pnas.0503392102.

Proost, S., and Mutwil, M. (2018). CoNeKt: an open-source framework for comparative genomic and transcriptomic network analyses. Nucleic Acids Res. 46:W133–W140. https://doi.org/10.1093/nar/gky336.

Radivojac, P., Clark, W.T., Oron, T.R., Schnoes, A.M., Wittkop, T., Sokolov, A., Gram, K., Funk, C., Verspoor, K., Ben-Hur, A., et al. (2013). A large-scale evaluation of computational protein function
Exploiting plant transcriptomic databases

prediction. Nat. Methods 10:221–227. https://doi.org/10.1038/nmeth.2340.

Rao, S.R., and Ravishankar, G.A. (2002). Plant cell cultures: chemical factories of secondary metabolites. Biotechnol. Adv. 20:101–153. https://doi.org/10.1016/s0734-9750(02)00007-1.

Rao, X., and Dixon, R.A. (2019). Coexpression networks for plant biology: why and how. Acta Biochim. Biophys. Sinica 51:981–988. https://doi.org/10.1093/abbs/gmz080.

Rhee, S.Y., and Mutwil, M. (2014). Towards revealing the functions of all genes in plants. Trends Plant Sci. 19:212–221. https://doi.org/10.1016/j.tplants.2013.10.006.

Ruprecht, C., Mendrinna, A., Tohge, T., Sampathkumar, A., Klie, S., Fernie, A.R., Nikoloski, Z., Persson, S., and Mutwil, M. (2016). Famnet: a framework to identify modular elements driving pathway expansion in plants. Plant Physiol. 170:1878–1894. https://doi.org/10.1104/pp.15.01281.

Ruprecht, C., Vaid, N., Proost, S., Persson, S., and Mutwil, M. (2017). Beyond genomics: studying evolution with coexpression networks. Trends Plant Sci. 22: https://doi.org/10.1016/j.tplants.2016.12.011.

Rynagillo, M., Childs, L., Lohse, M., Giorgi, F.M., Lude, A., Selbig, J., and Usadel, B. (2011). SLocX: predicting subcellular localization of arabidopsis proteins leveraging gene expression data. Front. Plant Sci. 2:43. https://doi.org/10.3389/fpls.2011.00043.

Sabzezhari, M., Zeinali, M., and Naghavi, M.R. (2020). Alternative sources and metabolic engineering of Taxol: advances and future perspectives. Biotechnol. Adv. 43:107569. https://doi.org/10.1016/j.biotechadv.2020.107569.

Sasaki, E., Takahashi, C., Asami, T., and Shimada, Y. (2011). AtCAST, a tool for exploring gene expression similarities among DNA microarray experiments using networks. Plant Cell Physiol. 52:169–180. https://doi.org/10.1093/pcp/pcq185.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Virgone, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of Arabidopsis thaliana development. Nat. Genet. 37:501–506. https://doi.org/10.1038/ng1543.

Serin, E.A.R., Nijveen, H., Hilhorst, H.W.M., and Ligterink, W. (2016). Learning from coexpression networks: possibilities and challenges. Front. Plant Sci. 7:444. https://doi.org/10.3389/fpls.2016.00444.

Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G.M., Zhu, Y., O’Malley, R.C., Brady, S.M., et al. (2019). High-Throughput single-cell transcriptome profiling of plant cell types. Cell Rep. 27:2241–2247.e4. https://doi.org/10.1016/j.celrep.2019.04.054.

Sibout, R., Proost, S., Hansen, B.O., Vaid, N., Giorgi, F.M., Ho-Yue-Kuang, S., Legée, F., Cézart, L., Bouchabké-Coussa, O., Souhait, C., et al. (2017). Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in Brachypodium distachyon. New Phytol. 215:1009–1025. https://doi.org/10.1111/nph.14635.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y., and Stitt, M. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 37:914–939. https://doi.org/10.1111/j.1365-313X.2004.02016.x.

To, T.K., Nakaminami, K., Kim, J.-M., Morosawa, T., I shida, J., Tanaka, M., Yokoyama, S., Shinozaki, K., and Seki, M. (2011). Arabidopsis HDA6 is required for freezing tolerance. Biochem. Biophys. Res. Commun. 406:414–419. https://doi.org/10.1016/j.bbrc.2011.02.058.

Toufighi, K., Brady, S.M., Austin, R., Ly, E., and Provat, N.J. (2005). The botany array resource: E-Northerns, expression angling, and promoter-arrays. Plant Journal: Cell Mol. Biol. 43:153–163. https://doi.org/10.1111/j.1365-313X.2005.02437.x.

Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S., and Provat, N.J. (2009). Coexpression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ. 32:1633–1651. https://doi.org/10.1111/j.1365-3040.2009.02040.x.

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J.P. (2018). Gene coexpression analysis for functional classification and gene-disease predictions. Brief. Bioinform. 19:575–592. https://doi.org/10.1093/bib/bbw139.

Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., Cumming, M., Kelley, L.A., Sternberg, M.J., Krishnakumar, V., et al. (2017). ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology[OPEN]. The Plant Cell 29:1806–1821. https://doi.org/10.1105/tpc.17.00073.

Xia, L., Zou, D., Sang, J., Xu, X., Yin, H., Li, M., Wu, S., Hu, S., Hao, L., and Zhang, Z. (2017). Rice Expression Database (RED): an integrated RNA-Seq-derived gene expression database for rice. J. Genet. Genomics 44:235–241. https://doi.org/10.1016/j.jgg.2017.05.003.

Xuan, L., Zhang, J., Lu, W., Gluza, P., Ebert, B., Kotake, T., Lu, M., Zhang, Y., Clausen, M.H., Johnson, K.L., et al. (2021). A pipeline towards the biochemical characterization of the arabidopsis GT14 family. Int. J. Mol. Sci. 22:1360. https://doi.org/10.3390/ijms22031360.

Yamaguchi, N., Huang, J., Xu, Y., Tano, K., and Ito, T. (2017). Fine-tuning of auxin homeostasis governs the transition from floral stem cell maintenance to gynoecium formation. Nat. Commun. 8:1125. https://doi.org/10.1038/s41467-017-01252-6.

Yanovsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., and Meyerowitz, E.M. (1990). The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. Nature 346:35–39. https://doi.org/10.1038/346035a0.

Zhang, T.Q., Xu, Z.G., Shang, G.D., and Wang, J.W. (2019). A single-cell RNA sequencing profiles the developmental landscape of Arabidopsis root. Mol Plant 12 (5):648–660. https://doi.org/10.1016/j.molp.2019.04.004.