Tissue-specific transcriptional biomarkers in medicinal plants: Application of large-scale meta-analysis and computational systems biology

Ahmad Tahmasebi, Esmaeil Ebrahimie, Hassan Pakniyat, Mansour Ebrahimi, Manijeh Mohammadi-Dehcheshmeh

PII: S0378-1119(19)30004-6
DOI: https://doi.org/10.1016/j.gene.2018.12.056
Reference: GENE 43501

To appear in: *Gene*

Received date: 29 March 2018
Revised date: 1 December 2018
Accepted date: 27 December 2018

Please cite this article as: Ahmad Tahmasebi, Esmaeil Ebrahimie, Hassan Pakniyat, Mansour Ebrahimi, Manijeh Mohammadi-Dehcheshmeh , Tissue-specific transcriptional biomarkers in medicinal plants: Application of large-scale meta-analysis and computational systems biology. Gene (2019), https://doi.org/10.1016/j.gene.2018.12.056

Tissue-specific transcriptional biomarkers in medicinal plants: application of large-scale meta-analysis and computational systems biology

Ahmad Tahmasebi[1], Esmaeil Ebrahimie[2,3,4,5*], Hassan Pakniyat[1], Mansour Ebrahimi[6], Manijeh Mohammadi-Dehcheshmeh[3,7]

[1]Department of Crop Production and Plant Breeding, College of Agriculture, Shiraz University, Shiraz 7144165186, Iran

[2]Adelaide Medical School, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide 5005, Australia

[3]Institute of Biotechnology, Shiraz University, Shiraz 7144165186, Iran

[4]Division of Information Technology, Engineering and the Environment, School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide 5005, Australia

[5]School of Biological Sciences, Faculty of Science and Engineering, Flinders University, Adelaide 5005, Australia

[6]Department of Biology, Qom University, Qom 3713166779, Iran

[7]Australian Centre for Antimicrobial Resistance Ecology, School of Animal and Veterinary Sciences, The University of Adelaide, Adelaide 5005, Australia

*** Corresponding:** esmaeil.ebrahimie@adelaide.edu.au

Abstract

Biosynthesis of secondary metabolites in plant is a complex process, regulate by many genes and influence by several factors. In recent years, the next-generation sequencing (NGS) technology and advanced statistical analysis such as meta-analysis and computational systems biology have provided novel opportunities to overcome biological complexity. Here, we performed a meta-analysis and multi-step integrated approach of publicly available transcriptome datasets of twelve economically significant medicinal plants to identify differentially expressed genes (DEGs) between shoot and root tissues to identify the key molecular features which may be effective in the biosynthesis of secondary metabolites. Meta-analysis identified a total of 880 genes with differential expression between two tissues. Functional enrichment and KEGG pathway analysis indicated that the functions of those DEGs are highly associated with the developmental process, starch metabolic process, response to stimulus, porphyrin and chlorophyll metabolism, biosynthesis of secondary metabolites and phenylalanine metabolism. In addition, systems biology analysis of the DEGs was applied to find protein–protein interaction network and discovery of significant modules. The detected modules were associated with hormone signal transduction, transcription repressor activity, response to light stimulus and epigenetic processes. Finally, analysis was extended to search for putative miRNAs that are associated with DEGs. A total of 31 miRNAs were detected which belonged to 16 conserved families. The present study provides a comprehensive view to better understand the tissue-specific expression of genes and mechanisms involved in secondary metabolites synthesis and may provide candidate genes for future researches to improve yield of secondary metabolites.

Keywords: Medicinal plant, Shoot and root tissues, Secondary metabolites, Transcriptome data, Meta-Analysis

Abbreviations

ROM: ratio of means
FC: fold change
GO: gene ontology
DEG: differentially expressed gene
qPCR: quantitative real-time RT-PCR
PPI: protein–protein interaction
CHS: chalcone synthase
C3′H: cytochrome p450 reductase
CYP: cytochrome P450
PAL: phenylalanine ammonia-lyase
HMGR: 3-hydroxy-3-methylglutaryl-coenzyme A reductase
SK1: shikimate kinase 1

1. Introduction

The biochemical diversity of plants is a valuable resource for improvement and discovery of new medicines. The value of herbal medicines comes from their pharmaceutical properties that is linked to secondary metabolites. Originally, secondary metabolites play a significant role in response to a wide variety of environmental changes in plants (Hussain et al., 2012). Understanding the molecular basis of secondary metabolite synthesis and employing this knowledge to improve the production of secondary metabolites has been an important goal for the producers. Biosynthesis and accumulation of secondary metabolite are often tissue-specific, and related genes of enzymes and regulators also show organ- or tissue-specific expression patterns (Upadhyay et al., 2014; Garg et al., 2015). The accumulation of anthocyanins and flavonols in vegetative organs under environmental stresses (Ferreyra et al., 2012), biosynthesis of tropane alkaloid scopolamine in roots and different tissues of *Atropa belladonna* and *Hyoscyamus muticus* (Ziegler and Facchini, 2008), biosynthesis of flavonoid pigments and volatile terpene in floral tissues and the monoterpene production in peppermint glandular (Croteau et al., 2000) are typical examples of organ-based production of secondary metabolites. Genes with tissue-specific expression have significant impacts on the status of these metabolites. Variation in the gene expression patterns across developmental stages, tissue types and various physiological or environmental conditions plays a central role in understanding of gene regulation in medicinal plants.

As mentioned above, the ability to identify genes and changes in their expression that correlate with synthesis of secondary metabolites is a fundamental aspect for utilization of transcriptomics in medicinal plants. The advancement of sequencing techniques for transcriptomic profiling has been the important core of new developments to overcome biological complexity. The advent of next-generation sequencing has opened new avenues in biology. Recently, RNA-seq is a potentially powerful and effective tool for expression profiling, especially in non-model plants for where the reference genome is not available (Wang et al., 2009). Due to the extensive application of RNA-seq technology, the amounts of data in this area is increasing exponentially, in recent years. With the considerable amount of deposited transcriptome data for the various plant species, it is necessary to generalize the concept from different studies to provide a more comprehensive and precise knowledge.

In general, to identify genes with potential importance, researchers have compared the global gene expression profiles under different conditions. Such analyses usually generate a list of genes that are differentially expressed, making it difficult to determine the genes that play key role (Rhodes et al., 2004).

One limitation in the field of medicinal plants is that most of the information have been obtained from experiments under the confined conditions. Meta-analysis is a powerful strategy that integrates transcriptomics data to identify core gene sets that regulate the complex traits (Frierson et al., 2002; Sharifi et al., 2018). Consequently, this creates an opportunity to reuse the data and discover critical components of the biological processes, particularly those associated to secondary metabolism. Whereas meta-analysis on transcriptome data, especially microarray data, has been widely applied in human and animal genomes, limited number of similar studies have been conducted in plants (Shaar-Moshe et al., 2015). However, these holistic methods for the study of medicinal plants are not yet used (Pathania and Acharya, 2016).

We hypothesized that identification of differential gene expression between tissues will give an overall picture about the genes involved in secondary metabolites production. Accordingly, meta-analysis of tissue-specific RNA-seq data and systems biology analysis were used to identify genes and important transcriptional programs that may play a determinant role in biosynthesis of secondary metabolites. We also confirmed findings of meta-analysis by tissue-specific expression of some selected genes for *Echinacea purpurea*, one of the important medicinal plants by RT-qPCR.

2. Materials and Methods

2.1. Data collection and pre-processing

Transcriptomic sequences and gene expression data of different tissues of twelve medicinal plants (Table 1) were obtained from the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) databases and Medicinal Plant Genomics Resource (MPGR, http://medicinalplantgenomics.msu.edu/) (Góngora-Castillo et al., 2012). All expressed transcripts were also annotated by performing BLASTX (an E-value cutoff ≤ 1E-6) searches using the TAIR (http://Arabidopsis.org) database. The expression level (log_2 transformed FPKM values) of each tissue was estimated then the dataset was grouped into a root class and shoot class for the twelve plants.

2.2. Meta-analysis of expression dataset

A meta-analysis was carried out on the integrated dataset to find the DEGs between the root and shoot tissues. First, to reduce number of statistical tests and control of false positives, 20% of genes that have low expression levels and last variance were excluded. A comparison between two classes for each species designed and moderated t-statistic with 1000 random permutaions conducted to determine the significance of each comparison for each gene. The adjusted p-value (FDR <0.05) (Benjamini and Hochberg, 1995) were considered significant and used for further analysis. To combine p-values and detection of differentially expressed genes between two tissue types, Fisher's method was used. The log ratio of means (ROM), which is the natural log of the ratio (Hu et al., 2009), was applied to measure the gene expression values. ROM were calculated by following formula:

$$y_{gn} = ln\left[\frac{\bar{r}_{gr}}{\bar{r}_{gs}}\right]$$

where y_{gn}, $\bar{r}_{gr}$ and $\bar{r}_{gs}$ represent the ROM, mean expression level in root and shoot for each gene in dataset, respectively. The preprocessing of the data, analyses and graph generation were performed with the various Bioconductor (http://www.bioconductor.org) packages; MetaDE, pheatmap, pathview.

2.3. Functional annotation and pathway enrichment analysis

GO enrichment analysis was conducted on significant DEGs obtained from meta-analysis using agriGO (Du et al., 2010). The functional characterization was annotated based on GO terms for biological process, molecular function and cellular component under a significance threshold of FDR <0.05. To identify important pathways of the DEGs, a pathway enrichment analysis also was carried out using the David database (http://david.abcc.ncifcrf.gov/).

2.4. Protein-protein interactions (PPI) and network construction

The protein-protein interactions analysis was performed to assess the possible interaction among DEGs in protein level. The BioGRID (http://thebiogrid.org/) was employed for PPI network. The ClusterONE plugin (Nepusz et al., 2012) (http://apps.cytoscape.org/apps/clusterone) in Cytoscape software was used to statistically distinct modules. The biological functions of gene sets corresponding to modules was tested using the ClueGO plugin (Bindea et al., 2009) with the cutoff of FDR <0.05. The Cytoscape software was utilized to visualize and integrate interaction networks.

2.5. Prediction of potential miRNAs

The sequences of DEGs corresponding to *Arabidopsis thaliana* were obtained from the Phytozome database (http://www.phytozome.net/). Then these sequences were used as queries to identify potential small RNA. We used psRNATarget server (http://plantgrn.noble.org/psRNATarget/) with default parameters except for, maximum expectation that was set to 2. The published miRNAs of *Arabidopsis thaliana* were used to find the miRNAs.
To further functional insight about the central regulators of identified miRNAs, common pathway regulators including small molecule, transcription factors, miRNAs and treatments were constructed using Pathway Studio 11 (http://www.elsevier.com/online-tools/pathway-studio) (Pashaiasl et al., 2016). The list of miRNAs was used to determine common pathway regulators based on several types of interactions such as direct regulation, expression, promoter binding and regulations.

2.6. Plant materials and quantitative real-time PCR (RT-qPCR)

To validate the meta-analysis results, five DEGs were selected and quantitative real-time PCR (RT-qPCR) was performed for *Echinacea purpurea* tissues. Seeds of *E. purpurea* were planted in glasshouse conditions. The shoot and root tissues from four-month-old plants were sampled. For RT-qPCR, total RNA was extracted from the tissues using the Denazist kit following the manufacturer's instructions. RT-qPCR was carried out using SYBR Green reagent. The relative expression levels of each gene were normalized to the internal reference gene *GAPDH*. Sequences of the primers are listed in Table S 6.

3. Results and Discussions

To study and compare tissue-specific profiling of genes in herbal medicines, gene expression data was collected from different tissues of different species of medicinal plant. Before performing the meta-analysis, PCA analysis was applied to all data sets and evaluated patterns of clustering among tissues. The first and second principal components explained 63% of the total variation in the data set (Fig S 1). PCA revealed that the majority of homologous tissues of different species were grouped into the same category. Accordingly, this observation confirms that many gene expression patterns in the developmental programs are conserved (Brawand et al., 2011).

3.1. Identification of DEGs by meta-analysis

At the first step of the analytical pipeline for determining DEGs, a moderated t-statistic with a threshold (FDR < 0.01) was performed for shoot versus root group in each species. The number of significant DEGs varied from 9 to 362 (Table S 1). The meta-analyses (using Fisher method) of the all dataset identified 880 genes differentially expressed (FDR < 0.01) among plant parts (Table S 2). The meta-analysis recognized genes which were not identified through the individual studies, demonstrating the extra ability of this approach to identify effective genes. Four genes, including *HPR, ATPC1 RCA* and *GUN5* revealed the highest fold change (|ROM| >3) (Fig 1). These genes are mostly involved during photosynthesis in chloroplast stroma. *HPR* is a coding gene that performs in the glycolate pathway of photorespiration (Wingler et al., 1998). The *ATPC1* is a protein coding that regulates ATPase activity and is required for starch accumulation (Dal Bosco et al., 2004). The *RCA* encodes a chloroplast protein that functions as a molecular chaperone with an important role in grain yield and regulating JA signaling (Attaran et al., 2014; Sun et al., 2016). *GUN5* is a mediators of the tetrapyrrole-dependent signaling pathway that involve in chlorophyll biosynthesis (Maruta et al., 2015) and abscisic acid (ABA) signaling (Voigt et al., 2010). Previous studies have shown a link between ABA and JA, suggesting that ABA may participate in cross-talk with JA signal transduction in biosynthesis of secondary metabolites. (Pan et al., 2010; Li et al., 2015). These results show that the expression of genes involved in photosynthesis and metabolite biosynthesis might be coordinated and regulated by phytohormone-associated genes.

3.4. Functional enrichment and KEGG pathway analysis of DEGs

To get more insight into the overall trends of biological functions in studied tissues, gene ontology (GO) based enrichment tests was implemented by searching of DEGs against the AgriGO database with an E-value threshold of 10^{-3}. GO analysis showed that genes are enriched in 12 biological process categories include developmental process, multicellular organismal process, pigment biosynthetic process, cellular component organization, multi-organism process, cellular process, reproductive process, metabolic process, establishment of localization, response to stimulus, starch metabolic process and cellular component biogenesis (Fig 2). Previous report on *Bacopa monnieri* has suggested that differential expression levels of genes are linked to secondary metabolite pathways in the root and shoot tissues (Jeena et al., 2017). It has been also suggested that synthetic modification of these metabolites may occur by different signaling cascades (Jeena et al., 2017). For the molecular function, major GO categories were the transporter activity and catalytic activity (Fig 2). The genes associated with ABC transporters were represented in the transporter activity category. These transporters are localized in most cell membranes and vacuolar (Ramilowski et al., 2013) and also are proposed to have a significant role in membrane transport of phytochemicals in the plant body (Yazaki, 2006). As described above, the DEGs were remarkably enriched in metabolic process. This analysis shows that tissue type conspicuously is decisive on metabolic composition. Furthermore, it has been demonstrated that DEGs may handle specific functions in certain tissue that determine the tissue-specialized processes.

In addition, the GO cellular component showed that the DEGs mainly are in the cytoplasm, chloroplast, thylakoid, plastid stroma, membrane, peroxisome, mitochondrion, endoplasmic reticulum (Fig 2). The biosynthesis of secondary metabolites localizes to different subcellular compartments. The enzymes are involved in monoterpenoid

indole alkaloids, for example, are present in the cytosol, chloroplast, endoplasmic reticulum and vacuole of *Catharanthus roseus* cells (Ziegler and Facchini, 2008; Nascimento and Fett-Neto, 2010) .

With the aim to identify the significant pathways, KEGG pathway enrichment analysis of the DEGs was performed based on a p-value cut-off 0.05 that allowed us to define pathways that depend on the tissue. KEGG analysis revealed that biosynthesis of secondary metabolites, porphyrin and chlorophyll metabolism, biosynthesis of antibiotics, flavonoid biosynthesis, phenylalanine metabolism, carbon metabolism and fatty acid metabolism are significantly enriched (Table 2). It has been reported that most of genes related to terpenoid and phenylpropanoid pathways are highly expressed in root than in shoot of *Valeriana fauriei* (Park et al., 2016). Interestingly, in the biosynthesis of secondary metabolites pathway, genes encoding OPC-8:0 CoA ligase 1 (*OPCL1*), acyl-CoA oxidase 1 (*ACX1*) and allene oxide synthase (*AOS*) were identified, and all of these enzymes involved in jasmonic acid biosynthesis. The phytohormone JA play a major role in many aspects of plant growth, developmental stages and reprogramming of plant secondary metabolism (De Geyter et al., 2012). Application of JA as elicitors dramatically increases the level of secondary metabolites in various plants. Currently, this strategy has been applied for enhancing the productivity of these valuable metabolites (De Geyter et al., 2012; Hussain et al., 2012).

In addition, the pathway analysis indicated that phenylalanine metabolism and flavonoid biosynthesis were significant. The amino acid phenylalanine (Phe) plays a critical role as interconnectivity of primary and secondary metabolism in plants (Pascual et al., 2016). This aromatic amino acid has closely linked to shikimate pathway, acts as precursors for a wide range of secondary metabolites and leads carbon from photosynthesis to the biosynthesis of phenylpropanoids (Tzin and Galili, 2010). The metabolism of Phe is also a source of anthocyanins, flavonoids, phenylpropanoids, lignans, condensed tannins, lignin, stilbenes and the plant hormone salicylate (Pascual et al., 2016).

Flavonoids are important type of plant secondary metabolite. In meta-analysis, 7 DEGs were detected that encode enzymes involved in flavonoid biosynthesis (Fig 3). Among the enzymes, chalcone synthase [EC 2.3.1.74] (CHS), flavanone 3-hydroxylase [EC:1.14.11.9] (F3H) and cytochrome P450 reductase [EC:1.14.13.36] (C3′H) were found. CHS convert 4-coumaroyl-CoA and malonyl-CoA to naringenin chalcone that the first step in the flavonoid biosynthesis. F3H catalyze the naringenin to substance dihydroflavonol. The majority of F3H genes have been obtained from some herbal medicines, including *Ginkgo biloba* (Shen et al., 2006), *Carthamus tinctorius* (Tu et al., 2016) and *Camellia sinensis* (Singh et al., 2008). The genes expression of CHS and F3H is influenced in response to environmental stress and exposure to JA which increased accumulation of flavonoid (Dao et al., 2011; Tu et al., 2016). Cytochrome P450 reductase belongs to the cytochrome P450 superfamily (CYP) that catalyzes 3-hydroxylation of p-coumaroyl shikimate to monolignol lignin precursor. In plants, CYP enzymes are involved in the biosynthesis or catabolism of all phytohormones and performs majority of the oxidation processes of secondary metabolites (Ramilowski et al., 2013). CYP also plays pivotal roles in synthesizing red colour in fruit, flowers and epidermal tissues (Tanaka and Brugliera, 2013). Currently, engineering of CYP-derived metabolites has been successful usage in plants and microorganisms (Renault et al., 2014).

3.5. Protein–protein interaction (PPI) network analysis and modules detection

For understanding of the global biological context of the detected DEGs from meta-analysis at the protein level, PPI networks were first constructed based on proteins encoded by DEGs that were obtained from BioGrid databases. PPI network had 2344 nodes and 4328 edges (Fig 4A). Among the nodes, *HHP4, COP1, SUMO1, PHYB, GAI* and *UBQ3* were identified with the highest degree that were the most significant nodes of the network. Interestingly, it was found that these nodes had significant association with response to stimulus.

The functional enrichment analysis led us to the major pathways. In other words, to get more detailed information, we used a network modules analysis to investigate the core molecular networks that may function in regulation biosynthesis of secondary metabolisms. Accordingly, clusterONE algorithm (Nepusz et al., 2012) was performed to identify the modules. The gene ontology (GO) enrichment analysis on the four top significant modules was performed based on the biological process and KEGG pathways (Table S 3). The modules varied in size, ranging from 13 to 76 nodes.

Module 1 that contained 76 nodes (Fig 4B) was enriched in hormone signal transduction and response to gibberellin (Table S 3). The plant hormone signaling, including gibberellin signaling, involves in various phenomena during plant growth and development (Santner and Estelle, 2009). It is well documented that a strong link between hormone signaling and production of secondary metabolites are often triggered by ethylene, abscisic acid, jasmonate and salicylic acid (Zhao et al., 2005; Nascimento and Fett-Neto, 2010); for example, taxol production in *Taxus spp* cell culture can be induced by ABA (Luo et al., 2001). In *C. roseus* cell culture, indole alkaloids are stimulated by salicylic acid (Zhao et al., 2000). Also, in *Echinacea purpurea* hairy root culture, treatment with gibberellic acid induces production of cichoric acid, caftaric acid, and chlorogenic acid (Abbasi et al., 2012). Similarly, gibberellic acid significantly increases accumulation of phenolic acids and phenylalanine ammonia lyase (PAL) enzyme activity in the *Salvia miltiorrhiza* (Liang et al., 2013).

The module 2 with 29 nodes (Fig 4C) has of 3 known genes from the transcription repressor activity (Table S 3). These genes (such as *TIFY7*, and *JAZ*) are related to transcriptional repressors that interact with several TFs to regulate aspect of the signalling pathway. The findings have demonstrated that JAZ regulates secondary metabolite production (De Geyter et al., 2012). For instance, in tobacco (*Nicotiana tabacum*), JAZ proteins as repressors of JA-mediated activation involve in nicotine biosynthesis. Other processes were associated with response to brassinosteroid. Brassinolides are steroid hormones that interacts with phenolic compounds such as flavonoids in response to environmental stresses (Bartwal et al., 2013). As reported, the exogenous brassinolides application increased activities of enzymes related to secondary metabolism such as PAL and SKDH in tomato (*Solanum lycopersicum* L.) (Ahammed et al., 2013). This module also contains *COI1* gene, which acts as a receptor in jasmonate signaling pathway (Liao et al., 2015). The *COI1* gene is required to start the transcription of JA-responsive genes and inducing biosynthesis of secondary metabolites (Brader et al., 2001).

Module 3 consists of 26 nodes (Fig 4D) and was mainly enriched in response to light stimulus, circadian rhythm and photoperiodism (Table S 3). The circadian rhythm involves in regulation of central plant activities like biotic and abiotic stress responses, growth, flowering time and metabolism (Greenham and McClung, 2015). The studies have shown that hormones biosynthesis and signalling-pathway genes are regulated by the circadian clock and light intensity (Covington et al., 2008) which influence expression of genes encoding enzymes responsible for secondary

metabolite biosynthesis (Khan et al., 2010). In *Arabidopsis thaliana*, variation in the glucosinolate structure has been associated with altered circadian clock transcriptional patterns (Kerwin et al., 2011). Also, for callus culture of *Zingiber officinale*, production of gingerol and zingiberene was increased in response to light (Anasori and Asghari, 2009).

Most notably, module 4 includes 17 nodes (Fig 4E) composed of 3 genes (*SCE1*, *SIZ1* and *SUMO1*). These genes are associated with the protein sumoylation (Table S 3). It is known that small ubiquitin-related modifier (*SUMO*) plays a critical role in the posttranslational modification of proteins, chromatin-related processes and regulation of hormone signalling, abiotic and biotic responses in plants (Park et al., 2011). The recent study addressed that *SIZ1* play key role in signaling pathways and responses to environmental stresses. The observations revealed that *siz1* mutation evokes ABA hyperinduction of several ABA-responsive genes, indicates that *SIZ1* has a negatively role in regulates ABA signaling (Lois et al., 2003; Miura and Hasegawa, 2010). *SUMO1* and *SCE1* are essential for development because knockout of *SUMO1* or *SCE1* cause embryo stage arrest (Elrouby, 2015). In addition, *SUMO* can affect the production of the secondary metabolite. This phenomenon is demonstrated among fungus such as *Aspergillus nidulans*, where the deletion of gene encoding *SUMO*, caused a dramatic increase in the production of asperthecin (Szewczyk et al., 2008). This result indicates that tissue type may regulate the production of the secondary metabolite through epigenetic changes such as histone modification.

3.6. Identification of potential miRNAs

MiRNAs are a class of small non-coding RNA molecules which play a key role in regulation of various biological processes by post-transcriptional regulation of gene expression. Additionally, many miRNA families are highly conserved across the plant kingdom (Zhang et al., 2006). The identification of putative miRNAs associated with DEGs, could prove a clue to clarify how miRNAs and target genes regulate tissue-specific biological functions in medicinal plants. Here, to find putative miRNAs associated with DEGs, the computational algorithm psRNATarget with high specificity based on the highly stringent penalty score (≤2) was employed. A total of 31 miRNAs were detected which belonged to 16 conserved families. The list of the miRNAs is shown in Table S 4. Among the detected miRNAs, the miR169 and mir399 families comprised the highest frequency with 7 and 6 members, respectively. A large proportion of these miRNAs, miR169, miR2936, miR774, miR395, miR5648-3p, miR773, miR830, miR854 and miR5641, are associated with a wide range of process, such as development and stress responses. In Arabidopsis, miR169 regulates stress-induced flowering by targeting the AtNF-YA transcription factor (Xu et al., 2013). Similarly, miR399 was reported to be involved in flowering time (Kim et al., 2011). *Dof* genes in the cucumber that play important roles during growth, development, and environmental stresses were targeted by miR5658 (Wen et al., 2016). MiR2936 and miR774 reportedly targets F-box protein encoding genes, whose degradation of cellular proteins, during floral transition and seed development (Jain et al., 2007; Merchan et al., 2009). MiR395 regulate sulfate concentration by targeting *APS* genes (Ai et al., 2016). MiR5648-3p expresses in late meristematic, elongation, and maturation zones and controls root growth (Breakfield et al., 2012). Additionally, miR773, miR830 and mir854 are involved in resistance to plant pathogens, response to temperature and drought stress, respectively (Lee et al., 2010; Li et al., 2010; Zhou et al., 2010). In tomato, miR5641 had important function

in signal transduction pathway and targeted seven genes such as DRP, ABF, LEA, ERF and SBP (Candar-Cakir et al., 2016). Because of the functions and importance of miRNAs in controlling biosynthetic pathways, recently, strategies have led to use them to improve secondary metabolites production. In this study, we also found four miRNA families, namely, miR393, miR5021, miR414 and miR408 that potentially regulate secondary metabolite synthesis in plants. Of these, miR414 and miR5021 have roles in terpenoid backbone biosynthesis in *Mentha* spp and *C. roseus*, respectably (Pani and Mahapatra, 2013; Singh et al., 2016). Moreover, it has been shown that in *Stevia rebaudiana*, genes of the steviol glycoside biosynthesis pathway may be governed by miR414 (Guleria and Yadav, 2011). MiR393 is one of the conserved miRNA families in plants that targets auxin receptors. This microRNA changes the production of secondary metabolites by repressing auxin signaling. It has been shown that miR393 overexpression in *Arabidopsis thaliana* increased levels of glucosinolate (Robert-Seilaniantz et al., 2011). MiR408 plays a central role in response to abiotic stress and its over-expression leads to reduced levels of reactive oxygen species (Ma et al., 2015). It has been found that, miR408 possibly targets a gene encoding reticuline oxidase-like protein that involve in the benzylisoquinoline alkaloids pathway (Boke et al., 2015).

A network analysis was performed to identify upstream regulators of miRNAs related to DEGs (Fig 5 and Table S 5). The regulatory network was noteworthy enriched with several transcription factors, miRNAs, treatments and small molecules. At small molecules, a high level of interaction is found between miRNAs and hormones, heavy metal, carbohydrates, etc (Fig 5C). It shows that auxin, ABA and MeJA are the major regulators of this miRNA network. Moreover, small molecules such as Cu^{2+}, Pi and sulfur have a positive effect on the regulation and expression of miR399 and miR395; however, some small molecules, such as ROS have a negative effect on miR393 expression. Previous studies demonstrated that higher concentrations of Cu^{2+} promoted the accumulation of betacyanins in *Amaranthus caudatus* seedlings (Obrenovic, 1990). In *C. roseus*, higher concentrations of phosphate induced alkaloid production through affecting on activities of phenylalanine ammonia-lyase (PAL) and tryptophan decarboxylase (TDC) enzymes (Knobloch and Berlin, 1983). Interactions between the different treatments and miRNAs are also noticeable, namely for starvation exposure, drought, salt and oxidative stress (Fig 5D). They involve in level of metabolites by coordinating the hormones and miRNAs (Gao et al., 2016). In addition, upstream microRNAs, including miR166, miR397, miR162 and miR169, in network are indicated involvement of these miRNAs in the several biological processes related to plant development and responses to stresses (Fig 5B). Recently, in different rose cultivars, miR166 has been implied as a negative regulator for flavonoids or carotenoids accumulation, resulting in white flowers (Kim et al., 2012).

In the network, the 3 proteins consisting of AGO1, PHR1 and MYB2, were also found that regulating miRNAs (Fig 5A). AGO proteins are crucial component of the silencing complexes that in many organisms are conserved (Mallory and Vaucheret, 2010). These proteins associate with miRNAs induces translational repression. The investigations indicated that AGO proteins related to HYL1 (a nuclear dsRNA binding protein) and could modulate secondary metabolic processes (Bulgakov and Avramenko, 2015). The MYB2 and PHR1 are a classification of plant transcription factors which were identified as the regulators of miR399. Transcription factors act as regulatory major fraction of miRNAs (Martinez and Walhout, 2009). Because importance of TFs in biological mechanisms, researches on determine functions of them in synthesis of metabolites have been promoted. Recently, MYB

transcription factors has been confirmed as regulators of phenylpropanoid-derived compounds synthesis in plants (Liu et al., 2015). MYB2 is an R2R3 MYB DNA binding domain that participate in primary and secondary metabolism, response to biotic and abiotic stresses, hormone signal transduction and determination of cell fate (Dubos et al., 2010). Transcription factor PHR1 has been recognized as a major regulator of metabolic changes, especially for phosphorus limitation. The concentrations of majority of glucosinolates, sinapic acid, fructose, sucrose and caffeic acid were reported to be reduced in the phr1 mutant grown in P-limiting conditions (Pant et al., 2015).

3.7. Validation of the meta-analysis

To verify the meta-analysis results, RT-qPCR was performed for expression of five selected DEGs in root and shoot tissues of *Echinacea purpurea*. *E. purpurea* is an economically important species with a wide range of therapeutic properties that is used in both modem and traditional medicinal plant. Echinacea products are the fourth most selling medicinal herb in Europe, and the sixth most selling in the United States, with an annual turnover of nearly €140 mill in Europe (Miller and Yu, 2004; Blumenthal et al., 2015). The presence of 216 different pharmaceutically active compounds in *E. purpurea* has been confirmed (Murch et al., 2006).

The relative abundance of genes (*HMGR1*, *JAZ1*, *CYP98A3*, *WRKY2* and *SK1*) that were differentially expressed in the meta-analysis were monitored. Overall, in most of the genes, RT-qPCR analysis correspond well with meta-analysis results, although, for *WRKY2* no statistically significant difference in expression levels was observed between the two tissues (Fig 6). The expression of these genes was greatly affected by the tissue type. The RT-qPCR results indicated that expression levels of *JAZ1* and *CYP98A3* genes in root were higher than shoot. In contrast, expressions levels of *HMGR1* and *SK1* genes were lower in root than shoot. *JAZ1* is involved in JA signaling and plays a regulatory function in biosynthesis of the phenolic compounds. The *CYP98A3* is involved in phenylpropanoid metabolism (Nair et al., 2002). In addition, it was found that the expression of the *HMGR* correlates with alkaloid levels in *C. roseus* (Ayora-Talavera et al., 2002). There are reports documenting the presence of major active phenylpropanoid compounds such as caftaric acid present in the root of *E. purpurea*, contrasting with higher concentration of alkamides, a class of proto-alkaloids, accumulate in shoot (Miller and Yu, 2004). This can explain different concentration of these compounds in different parts of *E. purpurea*.

4. Conclusion

Biosynthesis of secondary metabolites in medicinal plants are subjected to tissue-specific transcriptional regulation. Here, by integrating transcriptomic data of medicinal plants, meta-analysis as well as computational systems biology analysis, we monitored the transcriptomic changes between root and shoot in medicinal plants. Meta-analysis identified 880 tissue-associated genes that some of them have key roles in biosynthesis of secondary metabolites. These genes can be potential new molecular biomarkers for assessment of medicinal products in plants. Moreover, the PPI network analysis highlighted modules that affected by tissue type. These modules were associated with hormone signal transduction, transcription repressor activity, response to light stimulus and epigenetic processes. Additionally, we validated four biomarkers in *E. purpurea*. The biomarkers associated with tissue identified in this study can serve as useful resource for future breeding and genetic engineering programs in medicinal plants.

Author Contributions

AT, HP and EE had the idea, performed the data collection and analysis. AT, MM and EE performed the experiment. M.E. and M.M. contributed on the manuscript preparation and data interpretation.

Funding

This research was supported by the Shiraz University.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

References

Abbasi, B.H., Stiles, A.R., Saxena, P.K. and Liu, C.-Z., 2012. Gibberellic acid increases secondary metabolite production in *Echinacea purpurea* hairy roots. Biotechnol App. Biochem 168, 2057-2066.

Ahammed, G., Zhou, Y., Xia, X., Mao, W., Shi, K. and Yu, J., 2013. Brassinosteroid regulates secondary metabolism in tomato towards enhanced tolerance to phenanthrene. Biologia Plantarum 57, 154-158.

Ai, Q., Liang, G., Zhang, H. and Yu, D., 2016. Control of sulfate concentration by miR395-targeted APS genes in *Arabidopsis thaliana*. Plant Diversity.

Anasori, P. and Asghari, G., 2009. Effects of light and differentiation on gingerol and zingiberene production in callus culture of *Zingiber officinale* Rosc. Res Pharm Sci 3, 59-63.

Attaran, E., Major, I.T., Cruz, J.A., Rosa, B.A., Koo, A.J., Chen, J., Kramer, D.M., He, S.Y. and Howe, G.A., 2014. Temporal dynamics of growth and photosynthesis suppression in response to jasmonate signaling. Plant Physiol 165, 1302-1314.

Ayora-Talavera, T., Chappell, J., Lozoya-Gloria, E. and Loyola-Vargas, V.M., 2002. Overexpression in *Catharanthus roseus* hairy roots of a truncated hamster 3-hydroxy-3-methylglutaryl-CoA reductase gene. Biotechnol App. Biochem 97, 135-145.

Bartwal, A., Mall, R., Lohani, P., Guru, S. and Arora, S., 2013. Role of secondary metabolites and brassinosteroids in plant defense against environmental stresses. J Plant Growth Regul 32, 216-232.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B, 289-300.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z. and Galon, J., 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25, 1091-1093.

Blumenthal, M., Lindstrom, A., Lynch, M. and Rea, P., 2015. Herb sales continue growth: up 3.3% in 2010. , HerbalGram. pp. 64-67.

Boke, H., Ozhuner, E., Turktas, M., Parmaksiz, I., Ozcan, S. and Unver, T., 2015. Regulation of the alkaloid biosynthesis by miRNA in opium poppy. Plant Biotechnol J 13, 409-420.

Brader, G., Tas, E. and Palva, E.T., 2001. Jasmonate-dependent induction of indole glucosinolates in arabidopsis by culture filtrates of the nonspecific pathogenerwinia carotovora. Plant Physiol 126, 849-860.

Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A. and Kircher, M., 2011. The evolution of gene expression levels in mammalian organs. Nature 478, 343-348.

Breakfield, N.W., Corcoran, D.L., Petricka, J.J., Shen, J., Sae-Seaw, J., Rubio-Somoza, I., Weigel, D., Ohler, U. and Benfey, P.N., 2012. High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis. Genome Res 22, 163-176.

Bulgakov, V.P. and Avramenko, T.V., 2015. New opportunities for the regulation of secondary metabolism in plants: focus on microRNAs. Biotechnol Lett 37, 1719-1727.

Candar-Cakir, B., Arican, E. and Zhang, B., 2016. Small RNA and degradome deep sequencing reveals drought-and tissue-specific micrornas and their important roles in drought-sensitive and drought-tolerant tomato genotypes. Plant Biotechnol J 1, 277–283.

Covington, M.F., Maloof, J.N., Straume, M., Kay, S.A. and Harmer, S.L., 2008. Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. Genome Biol 9, 1.

Croteau, R., Kutchan, T.M. and Lewis, N.G., 2000. Natural products (secondary metabolites). Biochem Mo. Biol Plants 24, 1250-1319.

Dal Bosco, C., Lezhneva, L., Biehl, A., Leister, D., Strotmann, H., Wanner, G. and Meurer, J., 2004. Inactivation of the chloroplast ATP synthase γ subunit results in high non-photochemical fluorescence quenching and altered nuclear gene expression in *Arabidopsis thaliana*. J Biol Chem 279, 1060-1069.

Dao, T., Linthorst, H. and Verpoorte, R., 2011. Chalcone synthase and its functions in plant resistance. Phytochem Rev 10, 397-412.

De Geyter, N., Gholami, A., Goormachtig, S. and Goossens, A., 2012. Transcriptional machineries in jasmonate-elicited plant secondary metabolism. Trends Plant Sci 17, 349-359.

Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z., 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res, gkq310.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C. and Lepiniec, L., 2010. MYB transcription factors in Arabidopsis. Trends Plant Sci 15, 573-581.

Elrouby, N., 2015. Analysis of Small Ubiquitin-Like Modifier (SUMO) Targets Reflects the Essential Nature of Protein SUMOylation and Provides Insight to Elucidate the Role of SUMO in Plant Development. Plant Physiol 169, 1006-1017.

Ferreyra, M.L.F., Rius, S.P. and Casati, P., 2012. Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front Plant Sci 3.

Frierson, H.F., El-Naggar, A.K., Welsh, J.B., Sapinoso, L.M., Su, A.I., Cheng, J., Saku, T., Moskaluk, C.A. and Hampton, G.M., 2002. Large scale molecular analysis identifies genes with altered expression in salivary adenoid cystic carcinoma. Am J Pathol 161, 1315-1323.

Gao, F., Wang, N., Li, H., Liu, J., Fu, C., Xiao, Z., Wei, C., Lu, X., Feng, J. and Zhou, Y., 2016. Identification of drought-responsive microRNAs and their targets in Ammopiptanthus mongolicus by using high-throughput sequencing. Sci Rep 6.

Garg, A., Agrawal, L., Misra, R.C., Sharma, S. and Ghosh, S., 2015. Andrographis paniculata transcriptome provides molecular insights into tissue-specific accumulation of medicinal diterpenes. BMC Genomics 16, 659.

Góngora-Castillo, E., Childs, K.L., Fedewa, G., Hamilton, J.P., Liscombe, D.K., Magallanes-Lundback, M., Mandadi, K.K., Nims, E., Runguphan, W. and Vaillancourt, B., 2012. Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. PLoS ONE 7, e52506.

Greenham, K. and McClung, C.R., 2015. Integrating circadian dynamics with physiological processes in plants. Nat Rev Genet 16, 598-610.

Guleria, P. and Yadav, S.K., 2011. Identification of miR414 and expression analysis of conserved miRNAs from *Stevia rebaudiana*. Genomics Proteomics Bioinformatics 9, 211-217.

Hu, P., Greenwood, C.M. and Beyene, J., 2009. Using the ratio of means as the effect size measure in combining results of microarray experiments. BMC Syst Biol 3, 1.

Hussain, M.S., Fareed, S., Ansari, S., Rahman, M.A., Ahmad, I.Z. and Saeed, M., 2012. Current approaches toward production of secondary plant metabolites. J Pharm Bioallied Sci 4, 10.

Jain, M., Nijhawan, A., Arora, R., Agarwal, P., Ray, S., Sharma, P., Kapoor, S., Tyagi, A.K. and Khurana, J.P., 2007. F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. Plant Physiol 143, 1467-1483.

Jeena, G.S., Fatima, S., Tripathi, P., Upadhyay, S. and Shukla, R.K., 2017. Comparative transcriptome analysis of shoot and root tissue of *Bacopa monnieri* identifies potential genes related to triterpenoid saponin biosynthesis. BMC Genomics 18, 490.

Kerwin, R.E., Jimenez-Gomez, J.M., Fulop, D., Harmer, S.L., Maloof, J.N. and Kliebenstein, D.J., 2011. Network quantitative trait loci mapping of circadian clock outputs identifies metabolic pathway-to-clock linkages in Arabidopsis. The Plant Cell 23, 471-485.

Khan, S., Rowe, S.C. and Harmon, F.G., 2010. Coordination of the maize transcriptome by a conserved circadian clock. BMC Plant Biol 10, 1.

Kim, J., Park, J.H., Lim, C.J., Lim, J.Y., Ryu, J.-Y., Lee, B.-W., Choi, J.-P., Kim, W.B., Lee, H.Y. and Choi, Y., 2012. Small RNA and transcriptome deep sequencing proffers insight into floral gene regulation in Rosa cultivars. BMC Genomics 13, 1.

Kim, W., Ahn, H.J., Chiou, T.-J. and Ahn, J.H., 2011. The role of the miR399-PHO2 module in the regulation of flowering time in response to different ambient temperatures in *Arabidopsis thaliana*. Mol Cells 32, 83-88.

Knobloch, K. and Berlin, J., 1983. Influence of phosphate on the formation of the indole alkaloids and phenolic compounds in cell suspension cultures of *Catharanthus roseus* I. Comparison of enzyme activities and product accumulation. Plant cell, tissue and organ culture 2, 333-340.

Lee, H., Yoo, S.J., Lee, J.H., Kim, W., Yoo, S.K., Fitzgerald, H., Carrington, J.C. and Ahn, J.H., 2010. Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in Arabidopsis. Nucleic Acids Res, gkp1240.

Li, D., Li, L., Luo, Z., Mou, W., Mao, L. and Ying, T., 2015. Comparative transcriptome analysis reveals the influence of abscisic acid on the metabolism of pigments, ascorbic acid and folic acid during strawberry fruit ripening. PLoS ONE 10, e0130037.

Li, Y., Zhang, Q., Zhang, J., Wu, L., Qi, Y. and Zhou, J.-M., 2010. Identification of microRNAs involved in pathogen-associated molecular pattern-triggered plant innate immunity. Plant Physiol 152, 2222-2231.

Liang, Z., Ma, Y., Xu, T., Cui, B., Liu, Y., Guo, Z. and Yang, D., 2013. Effects of abscisic acid, gibberellin, ethylene and their interactions on production of phenolic acids in *Salvia miltiorrhiza* bunge hairy roots. PLoS ONE 8, e72806.

Liao, Y., Wei, J., Xu, Y. and Zhang, Z., 2015. Cloning, expression and characterization of COI1 gene (AsCOI1) from *Aquilaria sinensis* (Lour.) Gilg. Acta Pharmaceutica Sinica B 5, 473-481.

Liu, J., Osbourn, A. and Ma, P., 2015. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. Mol Plant 8, 689-708.

Lois, L.M., Lima, C.D. and Chua, N.-H., 2003. Small ubiquitin-like modifier modulates abscisic acid signaling in Arabidopsis. The Plant Cell 15, 1347-1359.

Luo, J., Liu, L. and Wu, C., 2001. Enhancement of paclitaxel production by abscisic acid in cell suspension cultures of *Taxus chinensis*. Biotechnol Lett 23, 1345-1348.

Ma, C., Burd, S. and Lers, A., 2015. miR408 is involved in abiotic stress responses in Arabidopsis. The Plant Journal 84, 169-187.

Mallory, A. and Vaucheret, H., 2010. Form, function, and regulation of ARGONAUTE proteins. The Plant Cell 22, 3879-3889.

Martinez, N.J. and Walhout, A.J., 2009. The interplay between transcription factors and microRNAs in genome-scale regulatory networks. Bioessays 31, 435-445.

Maruta, T., Miyazaki, N., Nosaka, R., Tanaka, H., Padilla-Chacon, D., Otori, K., Kimura, A., Tanabe, N., Yoshimura, K. and Tamoi, M., 2015. A gain-of-function mutation of plastidic invertase alters nuclear gene expression with sucrose treatment partially via GENOMES UNCOUPLED1-mediated signaling. New Phytol 206, 1013-1023.

Merchan, F., Boualem, A., Crespi, M. and Frugier, F., 2009. Plant polycistronic precursors containing non-homologous microRNAs target transcripts encoding functionally related proteins. Genome Biol 10, 1.

Miller, S.C. and Yu, H.-c., 2004. Echinacea: the genus Echinacea, Crc press.

Miura, K. and Hasegawa, P.M., 2010. Sumoylation and other ubiquitin-like post-translational modifications in plants. Trends Cell Biol 20, 223-232.

Murch, S.J., Peiris, S.E., Shi, W.L., Zobayed, S. and Saxena, P.K., 2006. Genetic diversity in seed populations of *Echinacea purpurea* controls the capacity for regeneration, route of morphogenesis and phytochemical composition. Plant Cell Reports 25, 522-532.

Nair, R.B., Xia, Q., Kartha, C.J., Kurylo, E., Hirji, R.N., Datla, R. and Selvaraj, G., 2002. Arabidopsis CYP98A3 mediating aromatic 3-hydroxylation. Developmental regulation of the gene, and expression in yeast. Plant Physiol 130, 210-220.

Nascimento, N.C.d. and Fett-Neto, A.G., 2010. Plant secondary metabolism and challenges in modifying its operation: an overview. Methods Mol Biol, 1-13.

Nepusz, T., Yu, H. and Paccanaro, A., 2012. Detecting overlapping protein complexes in protein-protein interaction networks. Nature methods 9, 471-472.

Obrenovic, S., 1990. Effect of Cu (II) D-penicillamine on phytochrome-mediated etacyanin formation in *Amaranthus caudatus*. Plant Physiol Biochem 28, 639-646.

Pan, Q., Chen, Y., Wang, Q., Yuan, F., Xing, S., Tian, Y., Zhao, J., Sun, X. and Tang, K., 2010. Effect of plant growth regulators on the biosynthesis of vinblastine, vindoline and catharanthine in *Catharanthus roseus*. Plant Growth Regul 60, 133-141.

Pani, A. and Mahapatra, R.K., 2013. Computational identification of microRNAs and their targets in *Catharanthus roseus* expressed sequence tags. Genomics data 1, 2-6.

Pant, B.D., Pant, P., Erban, A., Huhman, D., Kopka, J. and SCHEIBLE, W.R., 2015. Identification of primary and secondary metabolites with phosphorus status-dependent abundance in Arabidopsis, and of the transcription factor PHR1 as a major regulator of metabolic changes during phosphorus limitation. Plant Cell Environ 38, 172-187.

Park, B.S., Song, J.T. and Seo, H.S., 2011. Arabidopsis nitrate reductase activity is stimulated by the E3 SUMO ligase AtSIZ1. Nat Commun 2, 400.

Park, Y.J., Li, X., Noh, S.J., Kim, J.K., Lim, S.S., Park, N.I., Kim, S., Kim, Y.B., Kim, Y.O. and Lee, S.W., 2016. Transcriptome and metabolome analysis in shoot and root of *Valeriana fauriei*. BMC Genomics 17, 303.

Pascual, M.B., El-Azaz, J., Fernando, N., Cañas, R.A., Avila, C. and Cánovas, F.M., 2016. Biosynthesis and Metabolic Fate of Phenylalanine in Conifers. Front Plant Sci 7.

Pashaiasl, M., Ebrahimi, M. and Ebrahimie, E., 2016. Identification of the key regulating genes of diminished ovarian reserve (DOR) by network and gene ontology analysis. Mol Biol Rep 43, 923-937.

Pathania, S. and Acharya, V., 2016. Computational Analysis of "-omics" Data to Identify Transcription Factors Regulating Secondary Metabolism in *Rauvolfia serpentina*. Plant Mol Biol Report 34, 283-302.

Ramilowski, J.A., Sawai, S., Seki, H., Mochida, K., Yoshida, T., Sakurai, T., Muranaka, T., Saito, K. and Daub, C.O., 2013. *Glycyrrhiza uralensis* transcriptome landscape and study of phytochemicals. Plant Cell Physiol 54, 697-710.

Renault, H., Bassard, J.-E., Hamberger, B. and Werck-Reichhart, D., 2014. Cytochrome P450-mediated metabolic engineering: current progress and future challenges. Curr Opin Plant Biol 19, 27-34.

Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A. and Chinnaiyan, A.M., 2004. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci U S A 101, 9309-9314.

Robert-Seilaniantz, A., MacLean, D., Jikumaru, Y., Hill, L., Yamaguchi, S., Kamiya, Y. and Jones, J.D., 2011. The microRNA miR393 re-directs secondary metabolite biosynthesis away from camalexin and towards glucosinolates. The Plant Journal 67, 218-231.

Santner, A. and Estelle, M., 2009. Recent advances and emerging trends in plant hormone signalling. Nature 459, 1071-1078.

Shaar-Moshe, L., Hübner, S. and Peleg, Z., 2015. Identification of conserved drought-adaptive genes using a cross-species meta-analysis approach. BMC Plant Biol 15, 1.

Sharifi, S., Pakdel, A., Ebrahimi, M., Reecy, J.M., Farsani, S.F. and Ebrahimie, E., 2018. Integration of machine learning and meta-analysis identifies the transcriptomic bio-signature of mastitis disease in cattle. PLoS ONE 13, e0191227.

Shen, G., Pang, Y., Wu, W., Deng, Z., Zhao, L., Cao, Y., Sun, X. and Tang, K., 2006. Cloning and characterization of a flavanone 3-hydroxylase gene from *Ginkgo biloba*. Biosci Rep 26, 19-29.

Singh, K., Rani, A., Kumar, S., Sood, P., Mahajan, M., Yadav, S.K., Singh, B. and Ahuja, P.S., 2008. An early gene of the flavonoid pathway, flavanone 3-hydroxylase, exhibits a positive relationship with the concentration of catechins in tea (*Camellia sinensis*). Tree Physiol 28, 1349-1356.

Singh, N., Srivastava, S., Shasany, A.K. and Sharma, A., 2016. Identification of miRNAs and their targets involved in the secondary metabolic pathways of Mentha spp. Comp Biol Chem 64, 154-162.

Sun, Q., Zhang, Y., Chen, B., Jia, B., Zhang, Z.L., Cui, M., Kan, X., Shi, H.B., Deng, D.X. and Yin, Z.T., 2016. Expression quantitative trait loci analysis of the Rubisco activase gene in maize. Photosynthetica, 1-10.

Szewczyk, E., Chiang, Y.-M., Oakley, C.E., Davidson, A.D., Wang, C.C. and Oakley, B.R., 2008. Identification and characterization of the asperthecin gene cluster of *Aspergillus nidulans*. Appl Environ Microbiol 74, 7607-7612.

Tanaka, Y. and Brugliera, F., 2013. Flower colour and cytochromes P450. Phil. Trans. R. Soc. B 368, 20120432.

Tu, Y., Liu, F., Guo, D., Fan, L., Zhu, Z., Xue, Y., Gao, Y. and Guo, M., 2016. Molecular characterization of flavanone 3-hydroxylase gene and flavonoid accumulation in two chemotyped safflower lines in response to methyl jasmonate stimulation. BMC Plant Biol 16, 1.

Tzin, V. and Galili, G., 2010. New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. Mol Plant 3, 956-972.

Upadhyay, S., Phukan, U.J., Mishra, S. and Shukla, R.K., 2014. De novo leaf and root transcriptome analysis identified novel genes involved in steroidal sapogenin biosynthesis in Asparagus racemosus. BMC Genomics 15, 746.

Voigt, C., Oster, U., Börnke, F., Jahns, P., Dietz, K.J., Leister, D. and Kleine, T., 2010. In-depth analysis of the distinctive effects of norflurazon implies that tetrapyrrole biosynthesis, organellar gene expression and ABA cooperate in the GUN-type of plastid signalling. Physiol Plant 138, 503-519.

Wang, Z., Gerstein, M. and Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57-63.

Wen, C.-l., Cheng, Q., Zhao, L., Mao, A., Yang, J., Yu, S., Weng, Y. and Xu, Y., 2016. Identification and characterisation of Dof transcription factors in the cucumber genome. Sci Rep 6.

Wingler, A., von Schaewen, A., Leegood, R.C., Lea, P.J. and Quick, W.P., 1998. Regulation of leaf senescence by cytokinin, sugars, and light effects on NADH-dependent hydroxypyruvate reductase. Plant Physiol 116, 329-335.

Xu, M.Y., Zhang, L., Li, W.W., Hu, X.L., Wang, M.-B., Fan, Y.L., Zhang, C.Y. and Wang, L., 2013. Stress-induced early flowering is mediated by miR169 in *Arabidopsis thaliana*. J Exp Bot, ert353.

Yazaki, K., 2006. ABC transporters involved in the transport of plant secondary metabolites. FEBS Lett 580, 1183-1191.

Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P. and Anderson, T.A., 2006. Conservation and divergence of plant microRNA genes. The Plant Journal 46, 243-259.

Zhao, J., Davis, L.C. and Verpoorte, R., 2005. Elicitor signal transduction leading to production of plant secondary metabolites. Biotechnol Adv 23, 283-333.

Zhao, J., Zhu, W.-H., Hu, Q. and He, X.-W., 2000. Improved alkaloid production in *Catharanthus roseus* suspension cell cultures by various chemicals. Biotechnol Lett 22, 1221-1226.

Zhou, L., Liu, Y., Liu, Z., Kong, D., Duan, M. and Luo, L., 2010. Genome-wide identification and analysis of drought-responsive microRNAs in *Oryza sativa*. J Exp Bot 61, 4157-4168.

Ziegler, J. and Facchini, P.J., 2008. Alkaloid biosynthesis: metabolism and trafficking. Annu Rev Plant Biol 59, 735-769.

Figure legends

Fig 1. Heat map for top 30 differentially expressed genes (DEGs) that represents expression patterns in different tissues. The heatmap plot was visualized using the gplots package of R (http://www.r-project.org/).

Fig 2. GO enrichment of differentially expressed genes (DEGs): biological process (BP), molecular functions (MF) and cellular components (CC). Number of DEGs enriched in each GO term is shown.

Fig 3. Pathway map of flavonoid biosynthesis (Map:00941). Boxes coded violet are EC number of the enzymes encoded by differentially expressed genes (DEGs).

Fig 4. The constructed protein-protein interaction (PPI) and modules network of differentially expressed genes (DEGs). The modules identified from the PPI network using ClusterONE. (**A**) Global network involving the total DEGs (The red nodes indicate the identified hub proteins). (**B**) Module 1 (**C**) module 2 (**D**) module 3 and (**E**) module 4.

Fig 5. The regulatory network of identified potential miRNAs. (**A-D**) Protein, miRNA, small molecule and treatment sub-networks, respectively.

Fig 6. Validation of meta-analysis results for 5 selected genes in *Echinacea purpurea* root and shoot.

RT-qPCR was carried out to obtain expression levels of genes. The expression levels were normalized by GAPDH as reference gene. To evaluate statistical significance between two tissues, t-test was performed on data. Results are represented as the mean ± SD (n = 3); (* $P < 0.05$, ** $P < 0.01$; ns, $P > 0.05$). HMGR1, 3-hydroxy-3-methylglutaryl coa reductase 1; JAZ1, jasmonate-zim-domain protein 1; SK1, shikimate kinase 1; CYP98A3, cytochrome P450 98A3.

Table 1 List of 12 medicinal species used for transcriptome meta-analysis in this study.

Name of Species	Family	Samples (Shoot:Root)	Active constituent(s)
Camptotheca acuminata	Cornineae	11:2	Terpenoid, Indole Alkaloid, Camptothecin
Cannabis sativa	Nyssaceae	14:3	Cannabigerol, Cannabinol
Catharanthus roseus	Apocynaceae	4:6	Indole Alkaloid
Dioscorea villosa	Dioscoreaceae	9:3	Diosgenin, Dioscin
Echinacea purpurea	Asteraceae	17:3	Phenolic
Glycyrrhiza uralensis	Legume	1:3	Terpenoids, Flavonoids
Ginkgo biloba	Ginkgoaceae	7:2	Flavonoid, Ginkgolides
Hoodia gordonii	Apocynaceae	3:3	Steroidal Glycoside
Hypericum perforatum	Hypericaceae	15:2	Dianthrone Glycosides
Rauvolfia serpentina	Apocynaceae	5:2	Antiarrhythmic Alkaloid Ajmaline
Panax quinquefolius	Araliaceae	12:3	Triterpene Saponins
Valeriana officinalis	Caprifoliaceae	7:3	Acevaltrate, Actinidine, Hesperidin

Table 2 The KEGG pathway enrichment of the total differentially expressed genes (DEGs).

Pathway	Gene count[a]	Percentage	q-value
Biosynthesis of secondary metabolites	118	13.4	4.10E-10
Porphyrin and chlorophyll metabolism	13	1.47	7.60E-04
Biosynthesis of antibiotics	46	5.23	1.00E-02
Flavonoid biosynthesis	7	0.79	2.60E-02
Phenylalanine metabolism	8	0.90	1.80E-02
Carbon metabolism	29	3.29	4.30E-03
Fatty acid metabolism	10	1.14	2.20E-02

[a] The number and percentage of genes in each pathway.

Research Highlights

- ✓ Tissue-specific transcriptomic signature of genes involved in secondary metabolites synthesis
- ✓ Meta-analysis identified 880 genes with differential expression between roots and shoots
- ✓ 31 miRNAs were predicted to control the differential transcriptomic signature between shoots and roots

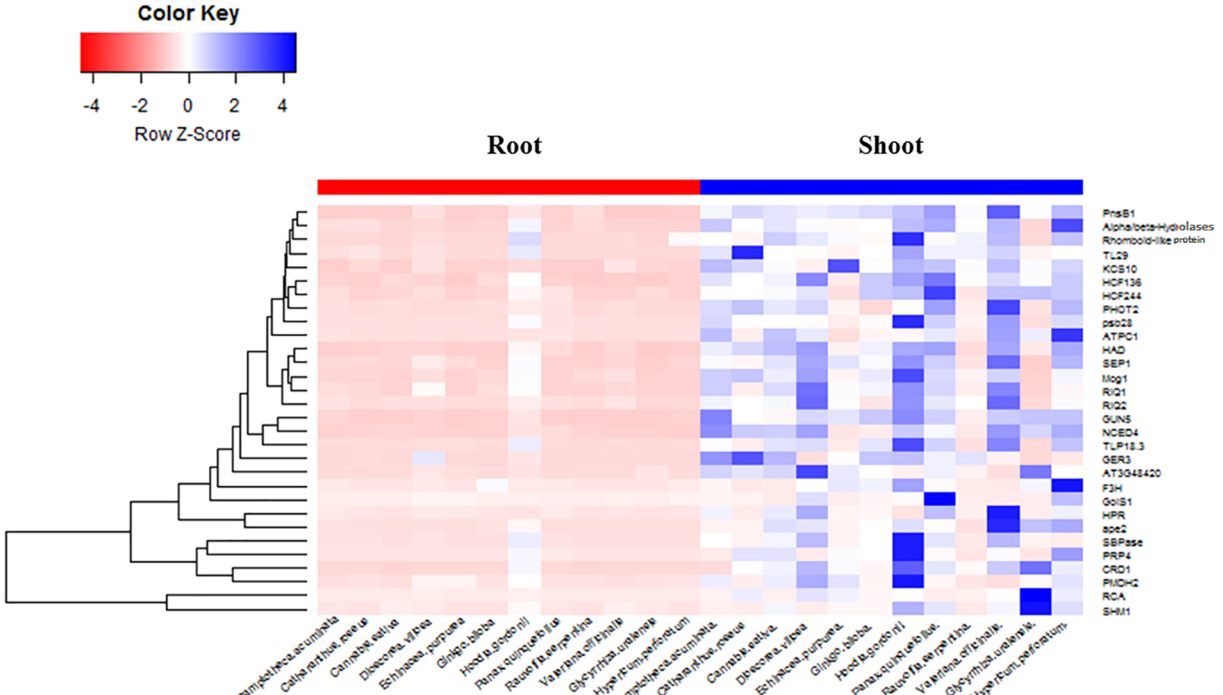
Color Key
-4
-2
0
2
4
Row Z-Score
Root
Shoot
PnsB1
Alphabeta-Hydrolases
Rhomboid-like protein
TLP9
KOS10
HCF136
HCF244
PHOT2
psb28
ATPC1
HAD
SEP1
Mog1
RIQ1
RIQ2
GUN5
NCED4
TLP18.3
GER3
AT3G48420
F3H
GolS1
HPR
ape2
SBPase
PRP4
CRD1
PMDH2
RCA
SHM1
Camptotheca.acuminata.
Catharanthus.roseus
Cannabis.sativa.
Dioscorea.villosa
Echinacea.purpurea.
Ginkgo.biloba.
Hoodia.gordonii
Panax.quinquefolius.
Rauvolfia.serpentina.
Valeriana.officinalis.
Glycyrrhiza.uralensis.
Hypericum.perforatum.

Figure 1

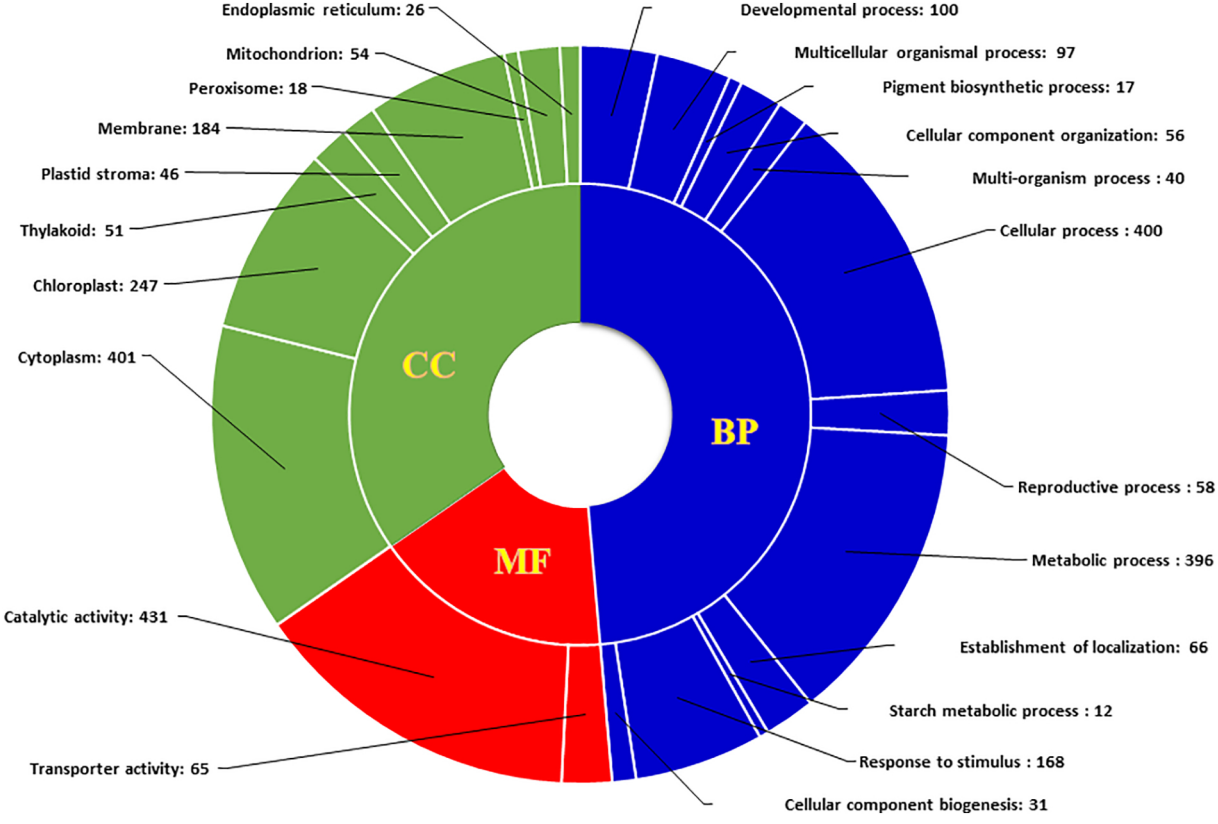
Endoplasmic reticulum: 26
Mitochondrion: 54
Peroxisome: 18
Membrane: 184
Plastid stroma: 46
Thylakoid: 51
Chloroplast: 247
Cytoplasm: 401
CC
BP
MF
Developmental process: 100
Multicellular organismal process: 97
Pigment biosynthetic process: 17
Cellular component organization: 56
Multi-organism process : 40
Cellular process : 400
Reproductive process : 58
Metabolic process : 396
Establishment of localization: 66
Starch metabolic process : 12
Response to stimulus : 168
Cellular component biogenesis: 31
Catalytic activity: 431
Transporter activity: 65

Figure 2

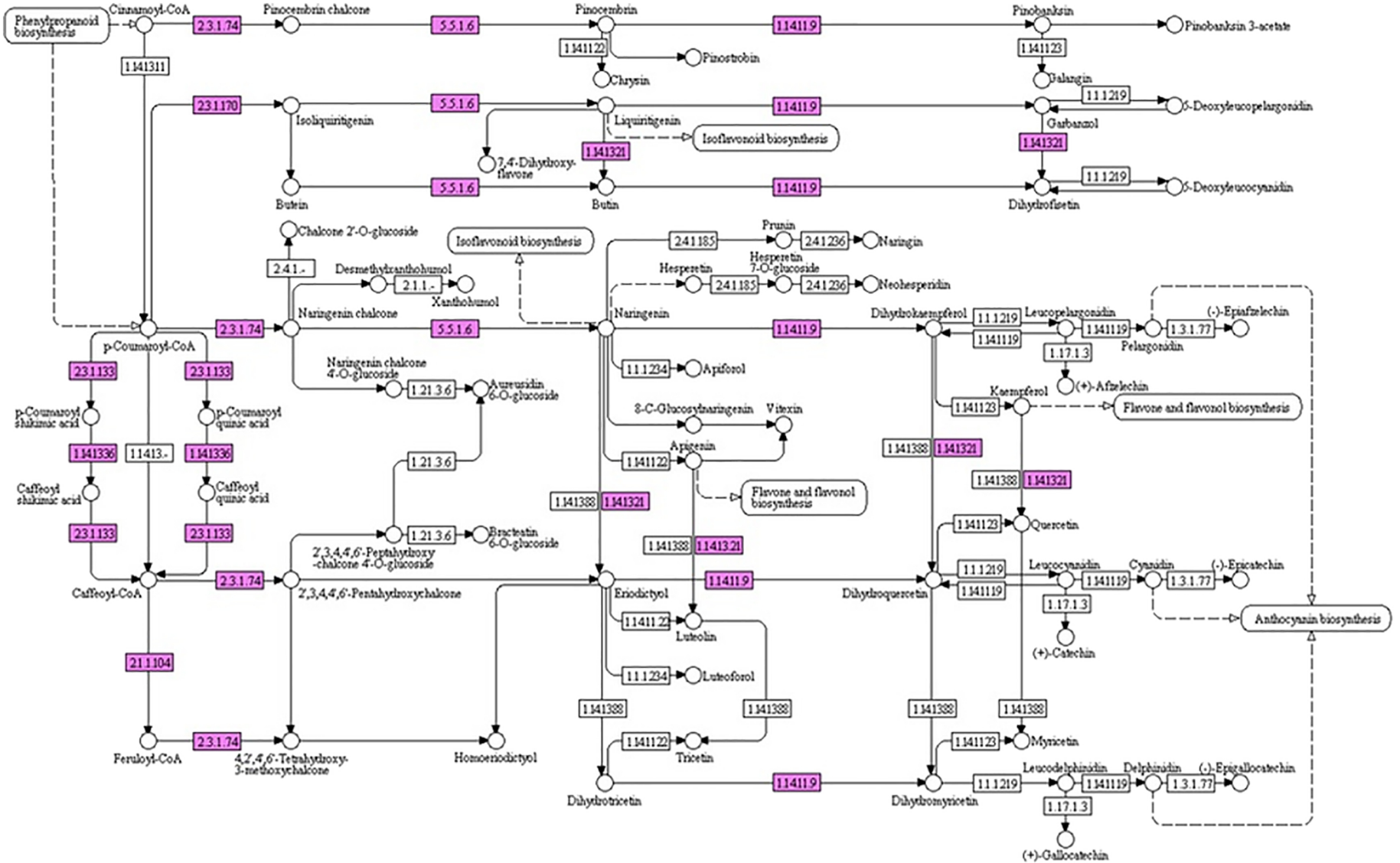
Phenylpropanoid biosynthesis
Cinnamoyl-CoA
2.3.1.74
Pinocembrin chalcone
5.5.1.6
Pinocembrin
1.14.11.9
Pinobanksin
Pinobanksin 3-acetate
1.14.11.22
Pinostrobin
Chrysin
1.14.11.23
Galangin
1.14.11.9
5-Deoxyleucopelargonidin
Garbanzol
1.14.13.21
Dihydrofisetin
5-Deoxyleucocyanidin
1.14.13.11
Isoliquiritigenin
Liquiritigenin
Isoflavonoid biosynthesis
7,4'-Dihydroxy-flavone
Butein
Butin
Chalcone 2'-O-glucoside
2.4.1.-
Desmethylxanthohumol
2.1.1.-
Xanthohumol
Naringenin chalcone
Naringenin
2.4.1.85
Prunin
2.4.1.236
Naringin
Hesperetin
Hesperetin 7-O-glucoside
Neohesperidin
Dihydrokaempferol
1.14.11.19
Leucopelargonidin
1.3.1.77
(-)-Epiafzelechin
Pelargonidin
1.17.1.3
(+)-Afzelechin
p-Coumaroyl-CoA
2.3.1.133
p-Coumaroyl shikimic acid
p-Coumaroyl quinic acid
Naringenin chalcone 4'-O-glucoside
1.21.3.6
Aureusidin 6-O-glucoside
1.14.13.-
1.14.13.64
Apiforol
8-C-Glucosylnaringenin
Vitexin
Apigenin
Kaempferol
Flavone and flavonol biosynthesis
Caffeoyl shikimic acid
Caffeoyl quinic acid
Bracteatin 6-O-glucoside
2',3,4,4',6'-Pentahydroxy-chalcone 4'-O-glucoside
Caffeoyl-CoA
2',3,4,4',6'-Pentahydroxychalcone
Eriodictyol
Luteolin
Luteoforol
Dihydroquercetin
Quercetin
Leucocyanidin
Cyanidin
(-)-Epicatechin
(+)-Catechin
Anthocyanin biosynthesis
2.1.1.104
Feruloyl-CoA
4,2',4',6'-Tetrahydroxy-3-methoxychalcone
Homoeriodictyol
Tricetin
Dihydrotricetin
Myricetin
Dihydromyricetin
Leucodelphinidin
Delphinidin
(-)-Epigallocatechin
(+)-Gallocatechin

Figure 3

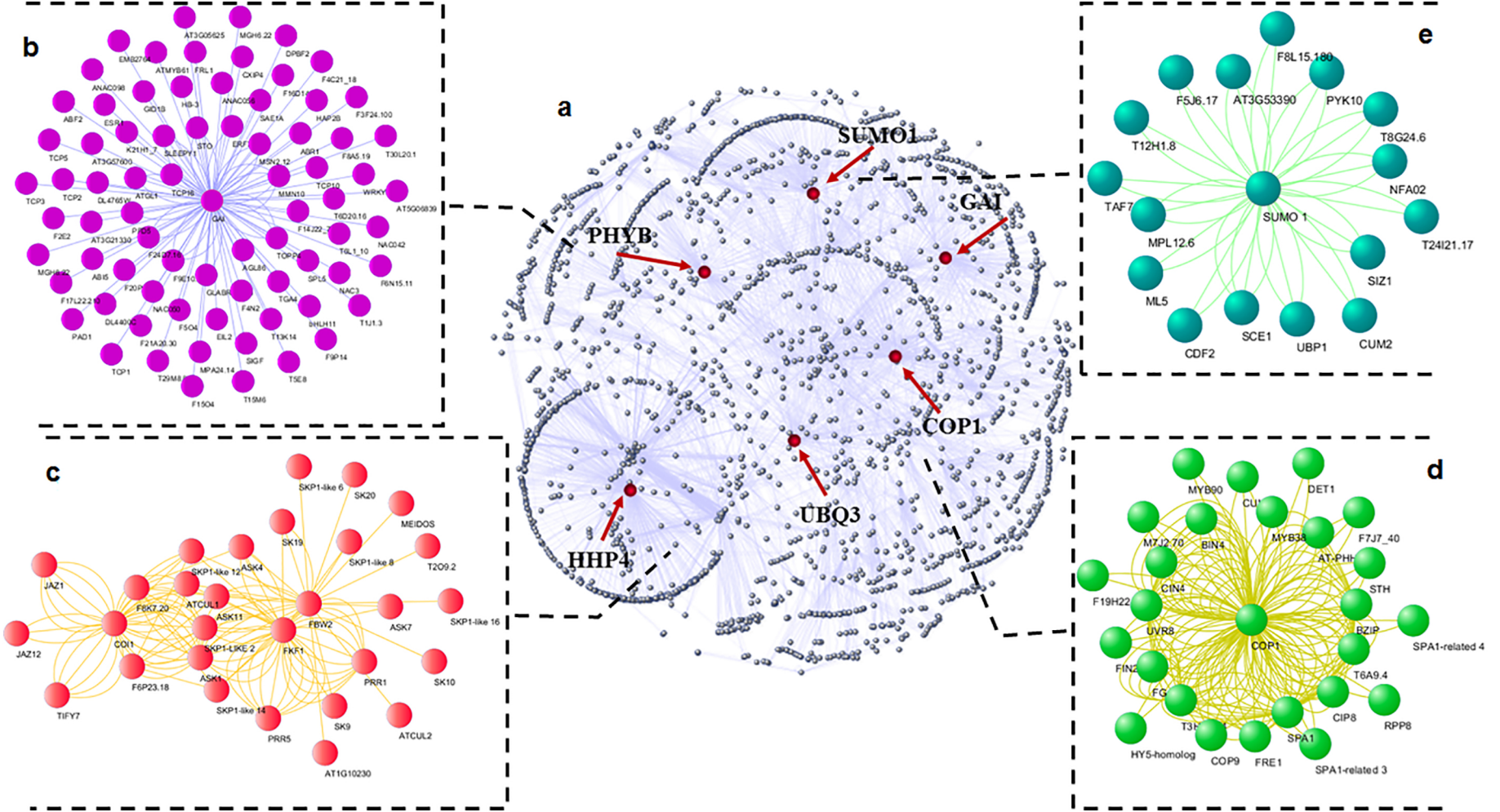
a
SUMO1
GAI
PHYB
COP1
UBQ3
HHP4
b
c
d
e
SUMO 1
COP1
SPA1-related 4
SPA1-related 3

Figure 4

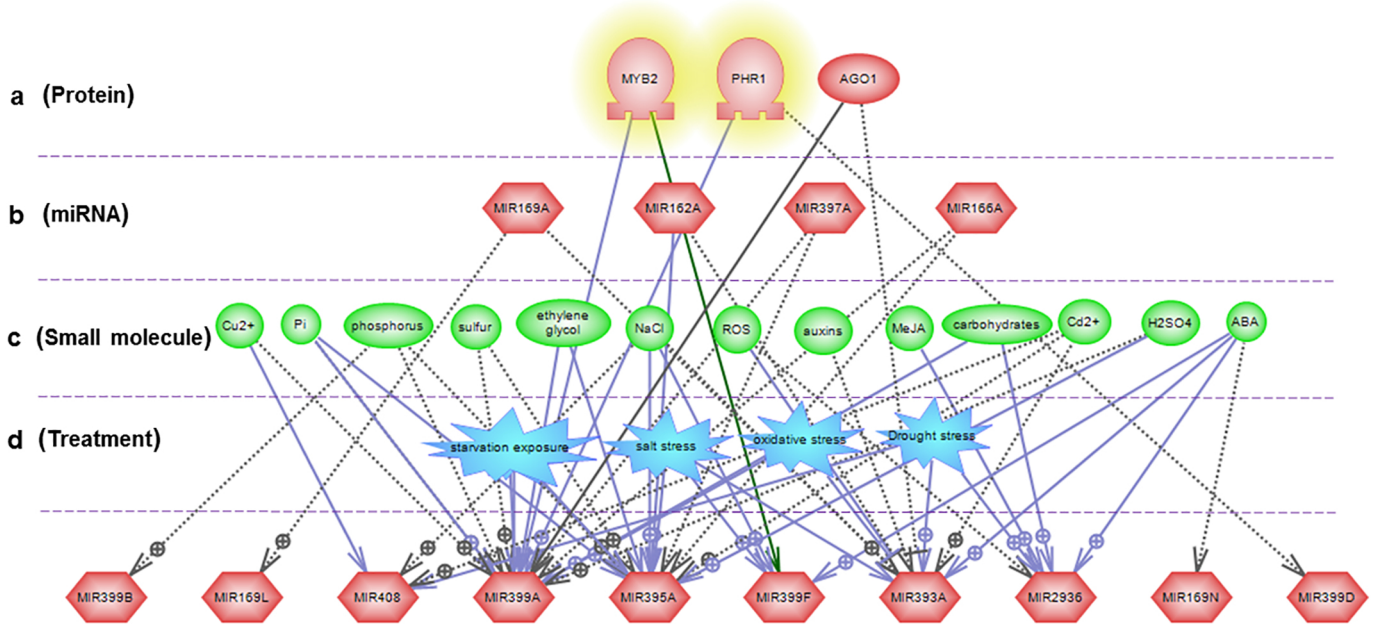
a (Protein)
b (miRNA)
c (Small molecule)
d (Treatment)
MYB2
PHR1
AGO1
MIR169A
MIR162A
MIR397A
MIR166A
Cu2+
Pi
phosphorus
sulfur
ethylene glycol
NaCl
ROS
auxins
MeJA
carbohydrates
Cd2+
H2SO4
ABA
starvation exposure
salt stress
oxidative stress
Drought stress
MIR399B
MIR169L
MIR408
MIR399A
MIR395A
MIR399F
MIR393A
MIR2936
MIR169N
MIR399D

Figure 5

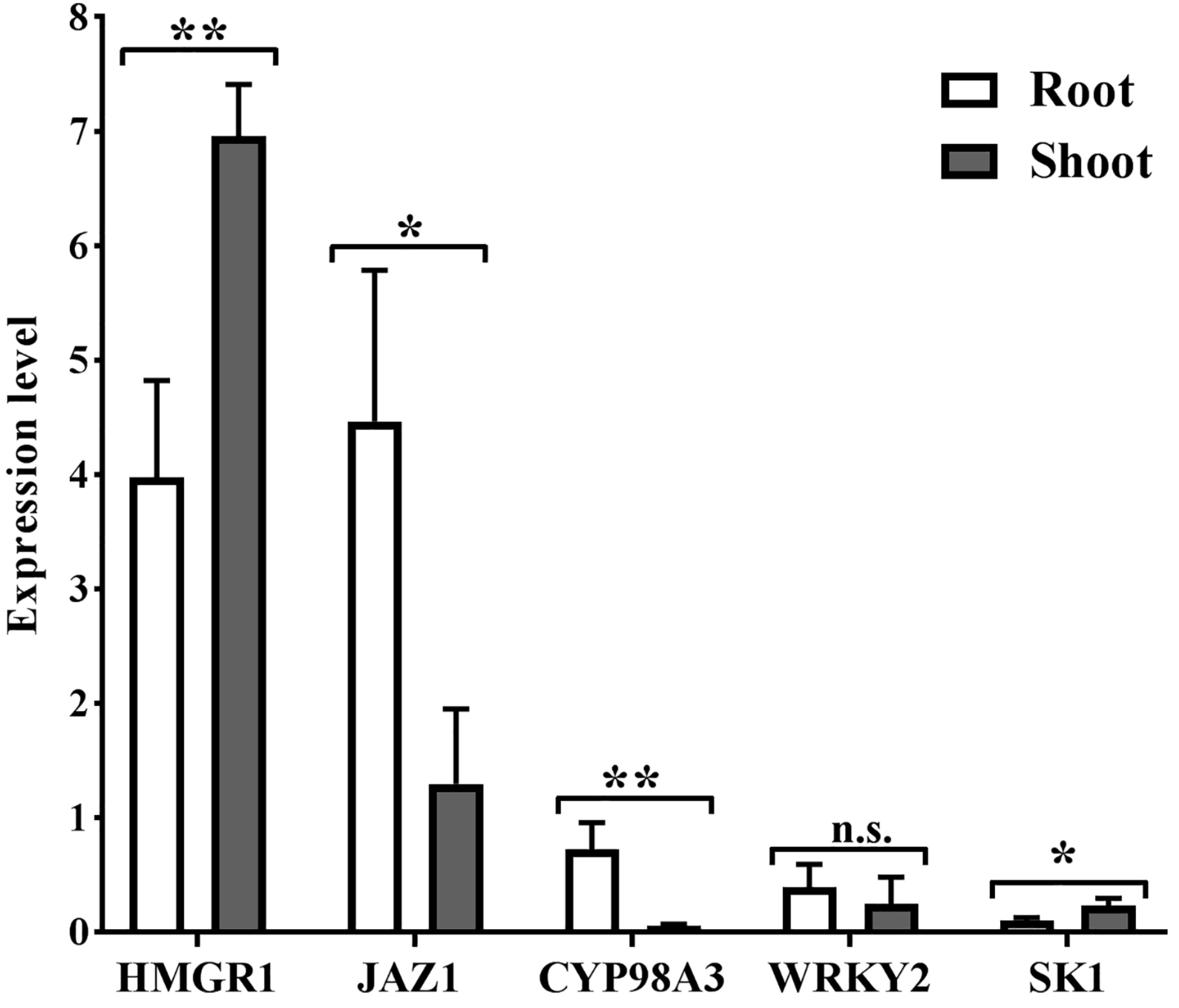
Root
Shoot
Expression level
0
1
2
3
4
5
6
7
8
**
*
**
n.s.
*
HMGR1
JAZ1
CYP98A3
WRKY2
SK1

Figure 6

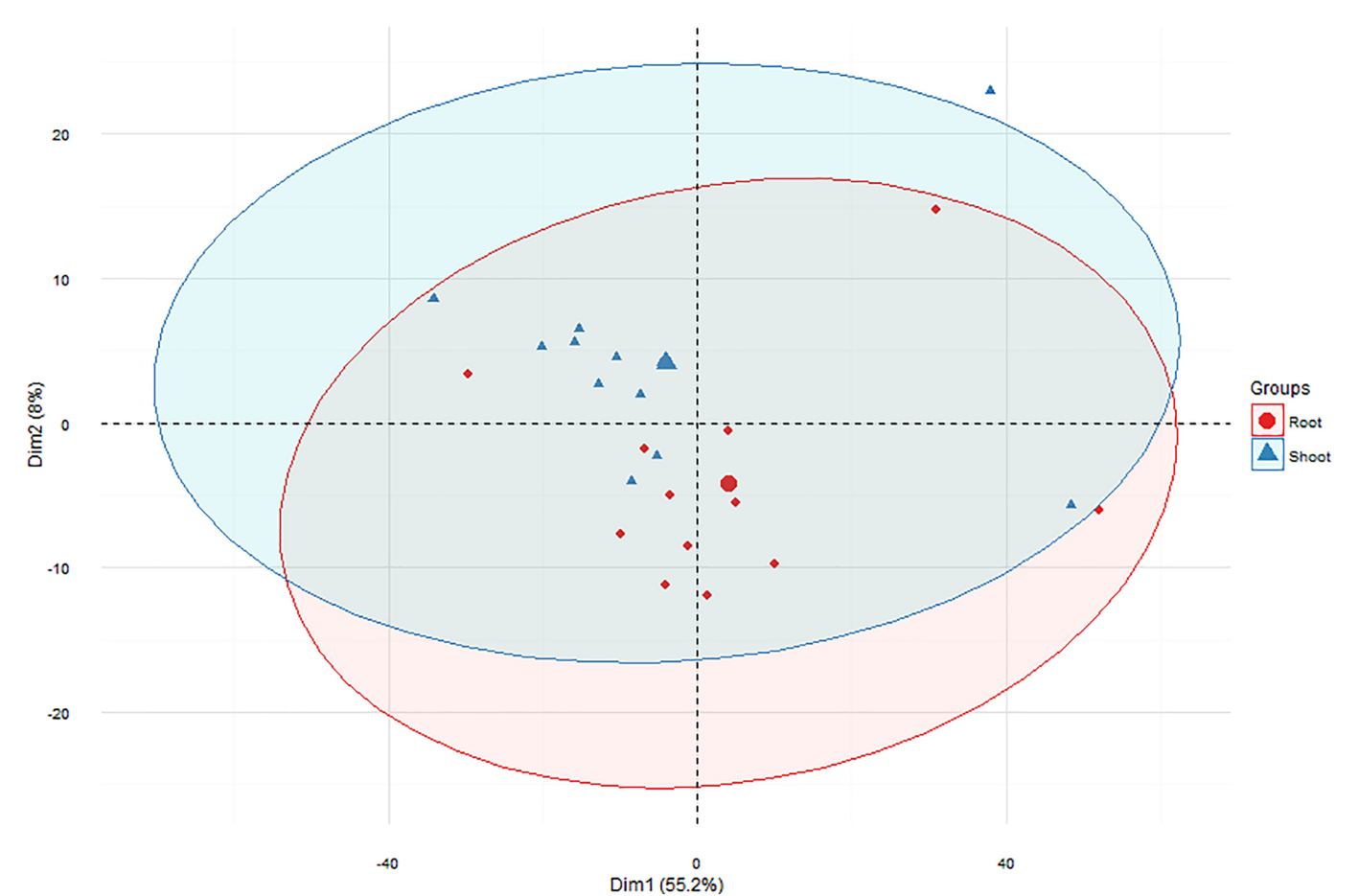

Figure 7