Integrated analysis of lncRNAs, mRNAs, and TFs to identify network modules underlying diterpenoid biosynthesis in *Salvia miltiorrhiza*

Lin Wang, Peijin Zou, Fang Liu, Rui Liu, Zhu-Yun Yan and Xin Chen

[1] School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China
[2] Key Laboratory of Characteristic Chinese Medicinal Resources in Southwest, Chengdu, Sichuan, China

ABSTRACT

Long non-coding RNAs (lncRNAs) are transcripts of more than 200 nucleotides (nt) in length, with minimal or no protein-coding capacity. Increasing evidence indicates that lncRNAs play important roles in the regulation of gene expression including in the biosynthesis of secondary metabolites. *Salvia miltiorrhiza* Bunge is an important medicinal plant in China. Diterpenoid tanshinones are one of the main active components of *S. miltiorrhiza*. To better understand the role of lncRNAs in regulating diterpenoid biosynthesis in *S. miltiorrhiza*, we integrated analysis of lncRNAs, mRNAs, and transcription factors (TFs) to identify network modules underlying diterpenoid biosynthesis based on transcriptomic data. In transcriptomic data, we obtained 6,651 candidate lncRNAs, 46 diterpenoid biosynthetic pathway genes, and 11 TFs involved in diterpenoid biosynthesis. Combining the co-expression and genomic location analysis, we obtained 23 candidate lncRNA-mRNA/TF pairs that were both co-expressed and co-located. To further observe the expression patterns of these 23 candidate gene pairs, we analyzed the time-series expression of *S. miltiorrhiza* induced by methyl jasmonate (MeJA). The results showed that 19 genes were differentially expressed at least a time-point, and four lncRNAs, two mRNAs, and two TFs formed three lncRNA-mRNA and/or TF network modules. This study revealed the relationship among lncRNAs, mRNAs, and TFs and provided new insight into the regulation of the biosynthetic pathway of *S. miltiorrhiza* diterpenoids.

Subjects Bioinformatics, Genetics, Genomics, Molecular Biology, Plant Science
Keywords LncRNA-mRNA/TF, Diterpenoid pathway, Salvia miltiorrhiza

Submitted 27 October 2022
Accepted 11 April 2023
Published 9 May 2023

Corresponding author
Xin Chen, chenxin@cdutcm.edu.cn

Academic editor
Douglas Domingues

Additional Information and Declarations can be found on page 15

DOI **10.7717/peerj.15332**

INTRODUCTION

Long non-coding RNAs (lncRNAs) are defined as RNA transcripts with little or no potential for protein-coding capacity, and with at least 200 nucleotides (nt) in size (*Ponting, Oliver & Reik, 2009*; *Nagano & Fraser, 2011*; *Palazzo & Koonin, 2020*). They are usually transcribed by RNA polymerase II (*Rinn & Chang, 2020*). Functional analysis of eukaryotic lncRNAs has revealed that they act as molecular scaffolds, guide molecules, molecular sponges and decoys, precursors of microRNAs (miRNAs) and other small RNAs, or as miRNA target mimics (TMs) to regulate gene expression at multiple levels (epigenetic regulation, transcriptional regulation, and post-transcriptional regulation) (*Franco-Zorrilla et al., 2007*;

Mercer, Dinger & Mattick, 2009; *Wang & Chekanova, 2017*; *Rai et al., 2019*). LncRNAs can also act on transcription factors (TFs); for example, lncRNA can be used as a TF-binding site to regulate its expression (*Yu et al., 2019*). LncRNAs are usually expressed at a low level and in a tissue-specific manner (*Liu et al., 2012*; *Palazzo & Koonin, 2020*). The subcellular localization of lncRNAs is the primary determinant of their molecular functions (*Carlevaro-Fita & Johnson, 2019*).

There is much evidence suggesting that lncRNAs play key roles in plant secondary metabolism. For example, mLncR8 putatively regulates terpenoid biosynthesis, and mLncR31 is involved in the biosynthesis of the isoprenoid side chain of ubiquinone and plastoquinone in *Digitalis purpurea* (*Wu et al., 2012a*). LncRNAs might regulate genes in the phenylpropanoid pathway of *Populus tomentosa* (*Zhou et al., 2017*). LncRNAs were also found to be involved in rubber biosynthesis in *Eucommia ulmoides* (*Liu et al., 2018*). LncRNAs were possibly involved in the biosynthesis of different fatty acids and lipid metabolism through post-transcriptional regulation in tree peony seeds (*Yin et al., 2018*). LncRNAs might be involved in the lignin biosynthetic pathway in *Populus* (*Quan et al., 2019*).

Studies have suggested that lncRNAs can act as local regulators, and lncRNA expression is correlated with the expression of nearby genes (*Guil & Esteller, 2012*; *Engreitz et al., 2016*). These correlations are attributed to sequence-specific functions of the mature lncRNA transcript, the transcription or splicing of an RNA, or DNA elements within the lncRNA promoter or gene locus, namely *cis*-regulation (*Guil & Esteller, 2012*; *Engreitz et al., 2016*; *Kopp & Mendell, 2018*).

Co-expressed genes are usually members of the same protein complex or metabolic pathway, and they are functionally controlled by the same transcriptional regulatory program. The genes or proteins within the co-expression network may have the same expression patterns (*Liao et al., 2011*; *Rao & Dixon, 2019*).

The medicinal plant *S. miltiorrhiza* produces a variety of diterpenoids (*Ma et al., 2015*). Tanshinones are the main bioactive compounds of *S. miltiorrhiza*, and they mainly accumulate in the roots of *S. miltiorrhiza* (*Xu et al., 2015*; *Chang et al., 2019*). Another diterpenoid in *S. miltiorrhiza* is the plant hormone gibberellin (GA), which is one of the five classic plant hormones (*Brockdorff, 1998*). In general, the biosynthetic pathway of terpenoids in *S. miltiorrhiza* can be divided into three stages (*Ma et al., 2012*). The first stage leads to the synthesis of the universal isoprene precursor isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) through the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway and/or the mevalonate (MVA) pathway. In the second stage, the intermediate diphosphate precursors, including geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP) are synthesized under the catalysis of isoprenyl diphosphate synthases (IDSs), including geranyl diphosphate synthase (GPPS), farnesyl diphosphate synthase (FPPS), and geranylgeranyl diphosphate synthase (GGPPS). The last stage involves the formation of diverse diterpenoids under the catalysis of terpene synthases (TPSs), such as copalyl diphosphate synthase (CPS) and kaurene synthase (KS) catalyze the formation of miltiradiene (*Kai et al., 2010*; *Kai et al., 2011*; *Lu, 2021*), *ent*-copalyl diphosphate synthase (*ent*-CPS) (*Shimane et al., 2014*) and *ent*-kaurene

synthase (*ent*-KS) are involved in the conversion of GGPP to the tetracyclic hydrocarbon intermediate *ent*-kaurene (*Yamaguchi, 2008*; *Shimane et al., 2014*). Then tanshinones and GAs are formed by cytochrome P450 monooxygenases (P450s) and 2-oxoglutarate-dependent dioxygenases (2ODDs) modification.

Methyl jasmonate (MeJA) is a hormone involved in plant signal transduction, which is considered to play an indispensable role as a second messenger in the induction process leading to the accumulation of secondary metabolites. Therefore, it is often used as an inducer to explore the regulation mechanism of biosynthesis (*Gundlach et al., 1992*; *Wasternack, 2007*). MeJA, an effective elicitor, can enhance the accumulation of tanshinones and phenolic acids by inducing the expression of tanshinone biosynthesis- and phenolic acid biosynthesis-related genes in *S. miltiorrhiza* (*Gao et al., 2009*; *Xiao et al., 2011*; *Liang et al., 2012*; *Luo et al., 2014*). MeJA is used for genome-wide identification and characterization of novel terpenoid biosynthetic genes in *S. miltiorrhiza* (*Ma et al., 2012*).

Previous studies of lncRNAs in *S. miltiorrhiza* (*Li, Shao & Lu, 2015*) showed that 3,044 lncRNAs responded to Ag^{+} solution and yeast extract (YE) in the roots of *S. miltiorrhiza*,15 lncRNAs differentially expressed in leaves under MeJA treatment. Jiang et al. identified the differential expression of natural antisense transcripts (NATs) with polyA tail in different tissues of *S. miltiorrhiza*, and some *cis*-NATs of *SmKSL1* showed a high co-expression relationship and possible participation in tanshinone synthesis (*Jiang et al., 2021*). However, although lncRNAs reportedly play important roles in *S. miltiorrhiza*, the role of lncRNAs in the diterpenoid biosynthetic pathway of *S. miltiorrhiza* remains largely unclear.

Based on transcriptomic data from four varieties of *S. miltiorrhiza*, we carried out among lncRNAs, mRNAs, and TFs expression correlation analysis and genome loci analysis to obtain candidate lncRNA-mRNA/TF pairs, and to construct the co-expression network. To further explore the relationship among the candidate lncRNA-mRNA/TF pairs, we analyzed the time-series expression patterns of the candidate lncRNA-mRNA/TF pairs in MeJA-induced *S. miltiorrhiza*.

MATERIALS & METHODS

Plant materials and MeJA treatment

S. miltiorrhiza seedlings were cultured in the greenhouse under 23 °C−−25 °C. The plants were sprayed with 200 μM MeJA solution as mentioned in a previous report (*Luo et al., 2014*). After being treated with MeJA solution for 6, 12, 24, and 48 h, the treated and the non-treated (0 h) roots were collected and rinsed with water. These roots were dried gently and quickly with absorbent paper. The cleaned roots were immediately frozen in liquid nitrogen and stored at −80 °C until RNA isolation. Three biological replicates were carried out for each experiment.

S. miltiorrhiza genomic and transcriptomic data

The *S. miltiorrhiza* genome data were downloaded from NCBI (National Center for Biotechnology Information) Sequence Read Archive database (SRP051524, SRP051564, SRP028388) (*Xu et al., 2016*) and NCBI (National Center for Biotechnology Information)

BioProject: PRJNA682867 (*Ma et al., 2021*). The transcriptomic data were obtained from our previous study (*Zhou et al., 2021*) with an accession number assigned to PRJNA712174. The transcriptomic data were gathered from four varieties of *S. miltiorrhiza* root tissues during the tanshinone accumulation stage and included Fragments Per Kilobase Million (FPKM) expression value and annotation information.

Pipeline for lncRNA identification

The following four filters were used to shortlist the bona fide lncRNAs from the transcriptomic data: (1) transcripts with annotation information in one of the Nr (Non-RedundantProtein Sequence), Swiss-Prot, and COG/KOG databases were removed; (2) transcripts shorter than 200 nt with an open reading frame (ORF) longer than 100 aa were discarded, found and extracted ORFs on getorf (http://emboss.bioinformatics.nl/cgi-bin/emboss/getorf), selection of ORF length <100 aa by R (version 4.2.1) script; (3) transcripts were searched against the Pfam database (*Punta et al., 2012*) (http://pfam.xfam.org/) by HMMER to remove transcripts possibly containing known protein domains; and (4) the protein-coding potential of each transcript was calculated using PLEK (*Li, Zhang & Zhou, 2014*) and Coding Potential Calculator 2 (CPC2, http://cpc2.cbi.pku.edu.cn) (*Kang et al., 2017*), and these with PLEK and CPC2 scores >0 were discarded. Through the above process, identified lncRNAs were obtained. FPKM value less than 0.05 was used as the standard for low expression levels (*Li et al., 2016*). The transcripts that remained were regarded as candidate lncRNAs.

Characterization and conservation analysis of *S. miltiorrhiza* lncRNAs

To gain more understanding of these lncRNAs in *S. miltiorrhiza*, we compared several different features of the lncRNAs and mRNAs: GC content and transcript length. GC content and transcript length were determined by R (version 4.2.1), and statistical analysis was carried out by Excel.

We aligned the lncRNA sequences identified here with BLAST+ (*Camacho et al., 2009*) (blast−2.11.0+) against the genome sequences of the *Lamiaceae* family: *Salvia splendens*, *Salvia hispanica*, *Mentha longifolia*, *Scutellaria baicalensis*, *Pogostemon cablin*, and *Sesamum indicum,* of which *Salvia splendens* and *Salvia hispanica* both belong to the *Salvia* genus; *Solanaceae* family: *Nicotiana tabacum*, *Brassicaceae* family: *Brassica napus* and *Arabidopsis thaliana*; and *Selaginellaceae* family: *Selaginella moellendorffii*. A cutoff *E*-value <1e−10 was used. The genomes were downloaded from the NCBI databases: GCF_004379255.1 (SspV2), GCF_023119035.1 (UniMelb_Shisp_WGS_1.0), GCA_001642375.2 (Mlong_CMEN585_v), GCA_005771605.1 (ASM577160v1), GCA_023678885.1 (GZUCM_PCab_1.0), GCF_000512975.1 (S_indicum_v1.0), GCF_000715135.1 (Ntab-TN90), GCF_020379485.1 (Da-Ae), GCF_000001735.4 (TAIR10.1), and GCF_000143415.4 (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000143415.4/). The lncRNAs that had coverage of >20% of matched regions were defined as conserved lncRNAs.

Precursors of miRNA and miRNA target prediction in *S. miltiorrhiza* lncRNAs

The candidate lncRNAs may act as the precursors of miRNAs, the *S. miltiorrhiza* miRNAs in miRBase (*Kozomara, Birgaoanu & Griffiths-Jones, 2019*) (Release 22.1, http://www.mirbase.org/) and PmiREN2.0 (*Guo et al., 2022*) (https://pmiren.com/) were aligned to the sequences of the candidate lncRNAs. The secondary structure of lncRNAs was predicted by RNAfold (*Gruber et al., 2008*) (http://rna.tbi.univie.ac.at/). LncRNAs with classical stem-loop hairpins were regarded as the putative precursors of miRNA (*Zhou et al., 2017*). Given that miRNA targets and lncRNAs have highly similar miRNA-binding sites, miRNA can be sequestered by lncRNA (*Paschoal et al., 2017*). Three kinds of prediction software were used to determine the miRNAs targeted to candidate lncRNAs. The first was TAPIR (*Bonnet et al., 2010*) (http://bioinformatics.psb.ugent.be/webtools/tapir); this server offers the possibility to search for plant miRNA targets using a fast and precise algorithm (score ≤ 4, free energy ratio ≥ 0.7). The second was psRobot (*Wu et al., 2012b*) (Version 1.2) with default parameters, which is a widely used online miRNA target prediction tool. The third was psRNATarget (2017 release) with default settings, and psRNATarget was developed to identify plant sRNA targets by (i) analyzing complementary matching between the sRNA sequence and target mRNA sequence using a predefined scoring schema and (ii) by evaluating target site accessibility (*Dai, Zhuang & Zhao, 2018*), targets with an *E* value less than 5.0 were retained as potential miRNA targets.

Subcellular localization of lncRNAs

The subcellular localization of *S. miltiorrhiza* lncRNAs was predicted by lncLocator (*Cao et al., 2018*). LncLocator is an ensemble classifier-based predictor, which adopts both *k-mer* features and high-level abstraction features generated by unsupervised deep models and constructs four classifiers by feeding these two types of features to support vector machine and random forest, respectively. The current lncLocator can predict five subcellular localizations of lncRNAs, namely, cytoplasm, nucleus, cytosol, ribosome, and exosome.

Predicted lncRNAs related to the diterpenoid biosynthesis in *S. miltiorrhiza*

Diterpenoid biosynthetic genes and the genes encoding TFs involved in diterpenoid biosynthesis were screened from the transcriptomic annotation data of *S. miltiorrhiza*. Our research focused on the downstream of the diterpenoid biosynthetic pathway (Fig. S1).

The functions of lncRNA can be performed on diterpenoid biosynthetic genes/TFs in a *cis* manner (*Kopp & Mendell, 2018*), and the lncRNAs and their target genes are considered lncRNA-mRNA/TF pairs. Pearson correlation coefficient (PCC) of the expression of lncRNA and its target gene pair was calculated using the psych (*Revelle, 2022*) package in R, requiring co-expressed gene pairs with $|PCC| \geq 0.4$ and $p \leq 0.05$. The basis for predicting *cis*-regulation was related to the positional relationship of lncRNA genes and coding genes on the genome (*Liu et al., 2019*). It was determined to be *cis*-regulatory if the lncRNA gene was within 100 kilobases (kb) upstream or downstream regions of the target genes used (*Huang et al., 2018*; *Wang et al., 2021*). used. For genomic location analysis, the co-expressed gene pairs were aligned against two *S. miltiorrhiza* genomes (*Xu et al., 2016*;

Ma et al., 2021) by BLAST+ (BLAST−2.11.0+) with a cutoff of *E* value <1e−5. Finally, the candidate lncRNA-mRNA/TF pairs related to the diterpenoid biosynthesis of *S. miltiorrhiza* were obtained.

Construction of the diterpenoid biosynthesis-related lncRNA-mRNA-TF networks and identification of hub genes

Cytoscape 3.2 (*Shannon et al., 2003*) was utilized to draw the putative lncRNA-mRNA-TF co-expression network with $|PCC| \geq 0.8$ ($p \leq 0.05$). Genes with a degree ≥ 10 were considered hub genes.

Candidate lncRNA-mRNA/TF pairs expression profile in MeJA-induced *S. miltiorrhiza*

MeJA is an effective elicitor for the production of diterpenoid tanshinones in *S. miltiorrhiza* (*Luo et al., 2014*). To further explore whether candidate lncRNA-mRNA/TF pairs have a response to the diterpenoid biosynthetic pathway, we performed the quantitative real-time polymerase chain reaction (qRT-PCR) to analyze expression patterns of the detected lncRNA-mRNA/TF pairs under MeJA induction in *S. miltiorrhiza*. Plant materials treated with MeJA dissolving media were used as a control (0 h), and three biological replications were carried out. The 2(-delta-delta CT) method was used for calculating the relative expression levels of genes. ANOVA was calculated using SPSS (Version 23.0, IBM, USA), and $p < 0.05$ and $p < 0.01$ were considered statistically significant. The hub genes were identified by using the cytoHubba (*Chin et al., 2014*) plug-in of Cytoscape 3.2. To increase the sensitivity and specificity, we proposed Maximal Clique Centrality (MCC) to discover featured nodes. Subsequently, we constructed a network module of these hub genes.

RNA extraction and qRT-PCR

Total RNA was extracted from plant tissues using the RNeasy plant kit (PH-01013-B, Foregene, Chengdu, China). RNA degradation and contamination were monitored using 1% RNase-free agarose gel electrophoresis, and the RNA purity was analyzed using a NanoPhotometer™ -N60 ultra-micro spectrophotometer. The reverse transcription reaction was used RT Easy™ II (With gDNase) kit (Version Number: 1.0-1904, Foregene, Chengdu, China) following the instruction manual. qRT-PCR was performed in a 20 μL reaction volume containing primers by Real-Time PCR EasyTM-SYBR Green I kit (Cat. No. QP-01011/01012/01013/01014, Foregene, Chengdu, China) following the instruction manual. qRT-PCR was carried out in triplicate reactions in the LineGene K Plus real-time PCR detection system (Bioer, Hangzhou, China). Primers were designed using the Primer-BLAST (*Ye et al., 2012*). Primer sequences were listed in Table S1. *β-actin* (*Yang et al., 2010*) gene was selected as a reference since it showed stable expression in the *S. miltiorrhiza* tissues analyzed compared to others. The reaction program was as follows: 3 min at 95 °C, 10 s at 95 °C, 30 s at 60 °C, and 40 cycles. The temperature was then gradually increased to produce melting curves for amplification specificity verification. The mean value of three replicates was normalized using *β-actin* as the endogenous control. The 2(-delta-delta CT) method was used for calculating the relative expression of genes.

RESULTS

Identification of candidate lncRNAs

On the basis of the strict lncRNA identification pipeline, a total of 30,347 transcripts with no annotation information were obtained. A total of 21,742 transcripts with length >200 nt and ORF <100 aa were obtained. Subsequently, a total of 21,468 identified lncRNAs were obtained from the intersection of PLEK and CPC2 software prediction results (Fig. 1, Tables S2, S3). Expression transcripts with FPKM $\leq$ 0.05 were filtered, and a total of 6,651 candidate lncRNAs were selected for further analysis (Table S4).

The 21,468 identified lncRNAs ranged from 201 nt to 4,340 nt in length, the majority of which (approximately 68.95%) were 200–400 nt. The mean length was 392 nt, which was lower than the values observed in *S. miltiorrhiza* mRNAs (mean length = 1,515 nt). The GC content of lncRNAs was mainly concentrated at 31.28%–37.28% (accounting for 46.32%), whereas mRNAs were mainly concentrated at 39.28%–41.28% (accounting for 36.00%). The mean GC content of lncRNAs was approximately 36.86%, which was slightly lower than that of mRNA sequences (approximately 43.42%) (Fig. 2, Table S5).

Characterization and conservation of lncRNAs in *S. miltiorrhiza*

Conservation analysis showed that 21.02% and 15.40% of *S. miltiorrhiza* lncRNAs (6,651 candidate lncRNAs) were conserved compared with *Salvia splendens* and *Salvia hispanica* of the same genus, respectively. However, in the same family of the different genera of *Mentha longifolia*, *Scutellaria baicalensis*, *Pogostemon cablin,* and *Sesamum indicum*, 6.54%, 0.96%, 0.83%, and 0.51% of *S. miltiorrhiza* lncRNAs were conserved, respectively. Among plants of different families, only 0.05%, 0.02%, and 0.02% of *S. miltiorrhiza* lncRNAs were conserved in *Nicotiana tabacum*, *Arabidopsis thaliana*, and *Brassica napus*, respectively. There was no match in *Selaginella moellendorffii*. The results of the conservation analysis are presented in Table S6.

LncRNAs might be used as precursors or target mimics of *S. miltiorrhiza* miRNAs

In this study, a total of 17 lncRNAs were identified as potential precursors of 41 *S. miltiorrhiza* miRNAs (Table S7). To improve the prediction accuracy, the intersection of three kinds of software prediction results was selected. A total of 14 lncRNAs were identified as potential targets of 66 *S. miltiorrhiza* miRNAs from 18 families according to PmiREN2.0 (Table S7).

Subcellular localization predictions indicated that most *S. miltiorrhiza* lncRNAs were found in the cytoplasm and nucleus

Subcellular localization of lncRNA is closely related to its function. At present, many studies have indicated that lncRNA contains specific RNA motifs with nuclear localization (*Zhang et al., 2014*). Meanwhile, in the subcellular localization of lncLocator (*Cao et al., 2018*), a total of 3,381 cytoplasms, 3,316 nucleus, 83 exosomes, 51 cytosols, and 20 ribosomes were obtained from 6,651 candidate lncRNAs. Among our candidate lncRNA-mRNA/TF pairs, 13 lncRNAs were found in the nucleus, 9 lncRNAs were found in the cytoplasm, and 1 lncRNA was found in the exosome (Table S8).

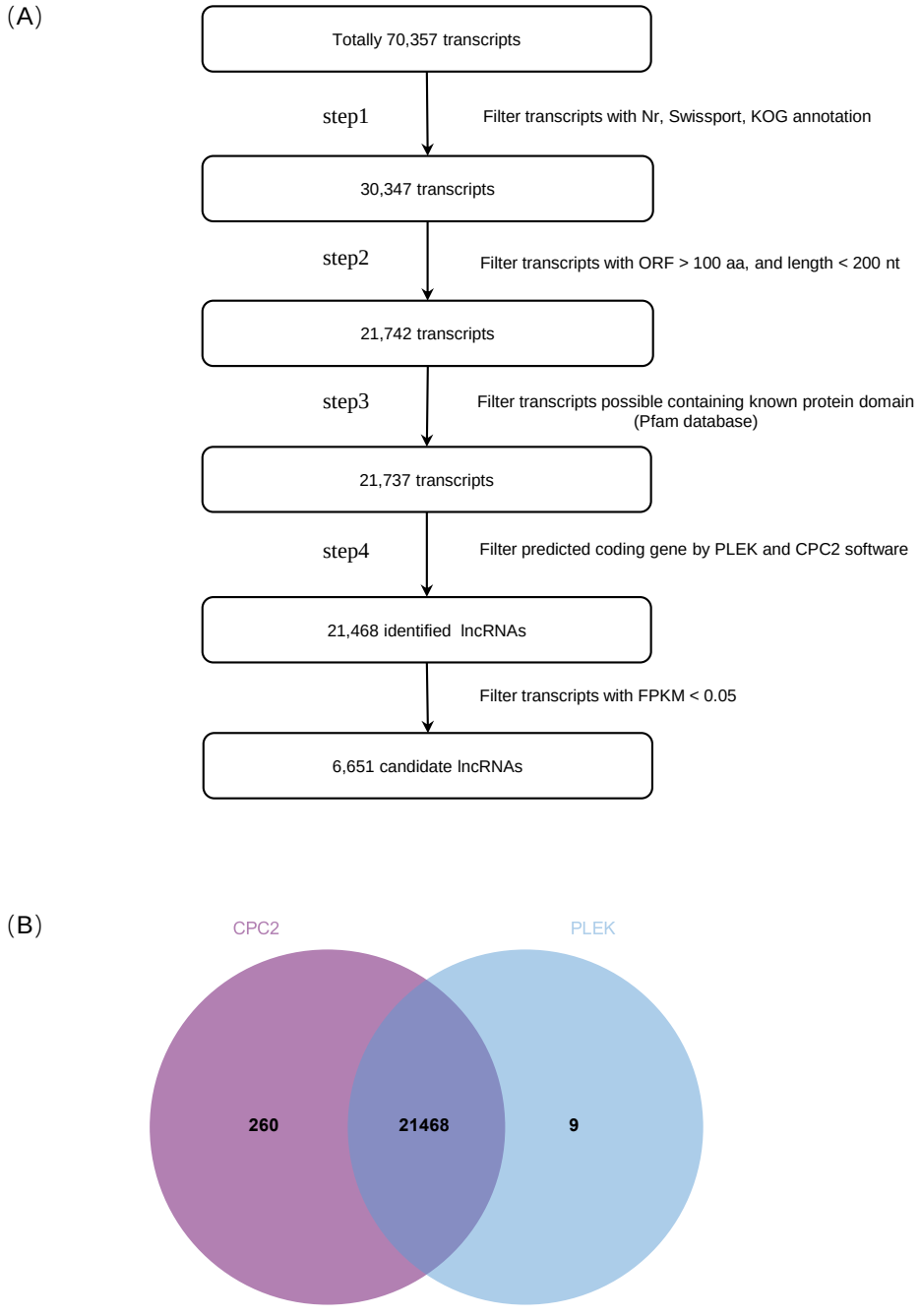

Figure 1 **Identification of lncRNAs in *S. miltiorrhiza*.** (A) Pipeline for lncRNAs identification in *S. miltiorrhiza*. Step1: transcripts with annotation information in one of the Nr, Swiss-Prot, and COG/KOG databases were removed; step2: transcripts shorter than 200 nt with an ORF longer than 100 aa were discarded; step3: transcripts were searched against the Pfam database to remove transcripts possibly containing known protein domains; and step4: the protein-coding potential of each transcript was calculated using PLEK and CPC2, transcripts with PLEK and CPC2 scores >0 were discarded. *nt* nucleotide, *ORF* open reading frame, *aa* amino acid. (B) Venn diagram showing the number of lncRNAs identification results in PLEK and CPC2 software, and their overlap.

Full-size DOI: 10.7717/peerj.15332/fig-1

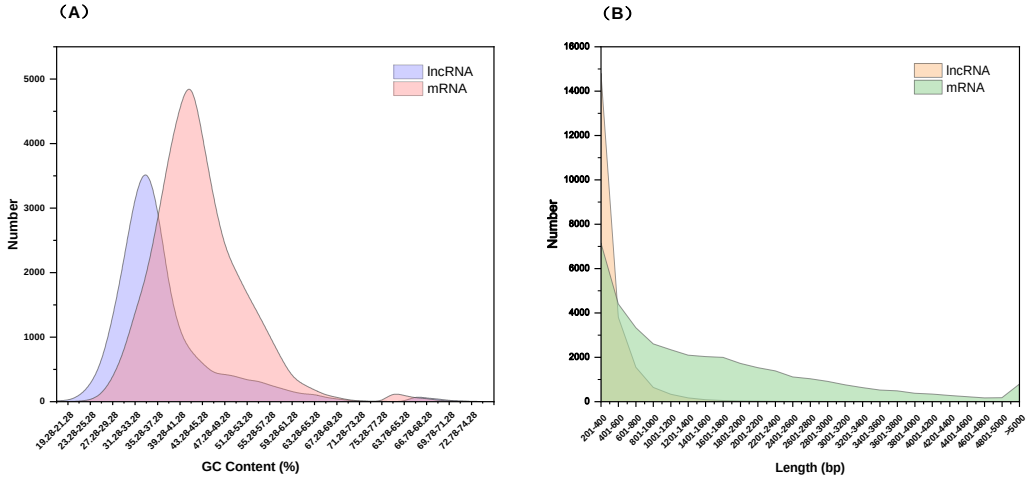

Figure 2 **Characterization of lncRNAs in *S. miltiorrhiza*.** (A) GC% distribution of the identified lncRNAs and mRNAs in *S. miltiorrhiza*. (B) Size distribution of the identified lncRNAs and mRNAs in *S. miltiorrhiza*.

Full-size DOI: 10.7717/peerj.15332/fig-2

Genes and TFs involved in diterpenoid biosynthesis from transcriptomic data

To investigate the potential lncRNAs involved in diterpenoid biosynthesis, we predicted the potential targets of lncRNAs in *cis*-regulatory relationships. In transcriptomic annotation, we obtained 46 diterpenoid biosynthetic genes: GPPS (Van (*Schie et al., 2007*), FPPS, GGPPS, CPS, *ent*-CPS, KS, *ent*-KS, CYP76AH1 (*Guo et al., 2013*; *Ma et al., 2016*), CYP76AK2, CYP76AK3 (*Li et al., 2021*), CYP76AK5 (*Xiangdong & Lizhi, 2017*), *ent*-kaurene oxidase (KO) (*Hedden & Thomas, 2012*), *ent*-kaurenoic acid oxidase (KAO) (*Helliwell et al., 2001*), GA 2-oxidase (GA2ox), GA 3-oxidase (GA3ox), GA 20-oxidase (GA20ox) (*Hedden & Thomas, 2012*; *Du et al., 2015*), and 11 TFs: bHLH148, ERF6, GRAS1, MYB36 and WRKY2 (*Li & Lu, 2014*; *Zhang et al., 2015*; *Li et al., 2015*; *Li et al., 2019*; *Ji et al., 2016*), belonging to five TF families: bHLH, ERF, GRAS, MYB, and WRKY. The list of the above genes is shown in Table S9.

LncRNAs related to the diterpenoid biosynthetic pathway

We detected 6455 lncRNAs that were co-expressed with 46 genes and 11 TFs involved in diterpenoid biosynthesis. In the PCC matrix, we obtained 45,198 correlation gene pairs with $|PCC| \geq 0.4$ ($p \leq 0.05$), of which 47,946 pairs were positively correlated and 5,775 pairs were negatively correlated. At the positional level, combining the results from the two *S. miltiorrhiza* genomes (*Xu et al., 2016*; *Ma et al., 2021*), a total of 180 pairs of lncRNA-mRNA/TF were located adjacent to each other within 100 kb (including ±3 kb), of which 23 pairs were co-located and co-expressed. A total of 23 candidate lncRNA-mRNA/TF pairs were obtained (Table S10).

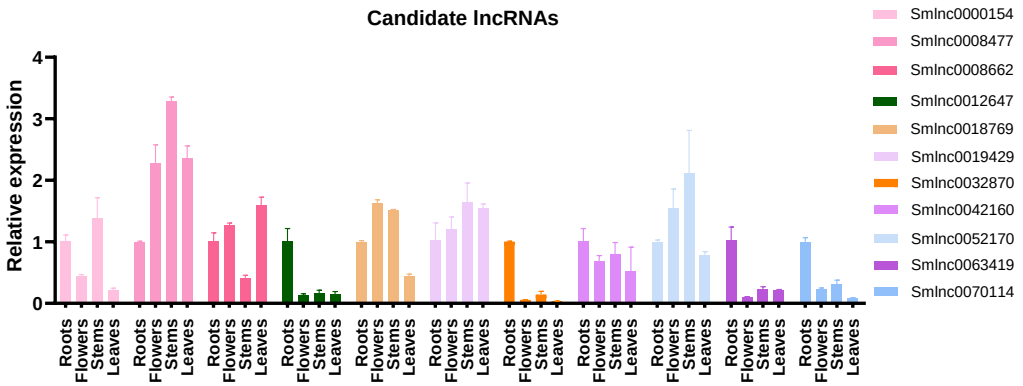

Figure 3 **Tissue specificity of lncRNA expression.** Validation of the expression level of 11 lncRNAs in roots, flowers, stems, and leaves tissues of *S. miltiorrhiza*. Fold changes of lncRNA levels were shown. The level of transcripts in roots was arbitrarily set to 1 and the level in other tissues was given relative to this. Error bars represent the standard deviations of three qRT-PCR replicates.

Full-size DOI: 10.7717/peerj.15332/fig-3

LncRNA-mRNA-TF networks and hub genes

The gene co-expression network was constructed with $|PCC| \geq 0.8$ ($p \leq 0.05$) as the threshold, the visualization is shown in Fig. S2A. We obtained 1,402 gene pairs with $|PCC| \geq 0.8$. We obtained 24 mRNAs with a degree > 10, which were considered hub genes (Table S11). In the 23 gene pairs of lncRNA-mRNA/TF, 8 mRNAs were present in these hub genes. Only two gene pairs *Smlnc0032870-Sm0012648* (*GA2OX*) and *Smlnc0018769-Sm0037093* (*KS2*) existed in the network with high expression correlation of 0.85 and 0.80, respectively. The hub genes *Sm0012648* and *Sm0037093* in these two gene pairs were used to construct the subnetworks (Figs. S2B, S2C).

Tissue specificity of lncRNA expression

Using the qRT-PCR, expression patterns of lncRNAs in the candidate lncRNA-mRNA/TF pairs were analyzed in roots, stems, leaves, and flowers of 2-year-old field greenhouse-grown *S. miltiorrhiza* plants. Of them, 11 lncRNAs were detected in at least one tissue and showed tissue-specific expression. The other lncRNAs were undetected, suggesting that they could be not expressed or expressed at a low level in the tissues analyzed. *Smlnc0012647*, *Smlnc0032870*, *Smlnc0042160*, *Smlnc0063419*, and *Smlnc0070114* exhibited the highest expression in root tissue. *Smlnc0000154*, *Smlnc0008477*, *Smlnc0019429*, and *Smlnc0052170* were more stem-specific. *Smlnc0018769* was expressed mainly in flowers and stems. *Smlnc0008662* showed high expression in roots, flowers, and leaves and low expression in stems (Fig. 3). Thus, lncRNAs showed obvious differential expression in different tissues, which may be related to their regulatory function. These results suggest that the expression of lncRNAs may be limited to specific tissue types or regulated by development in *S. miltiorrhiza*.

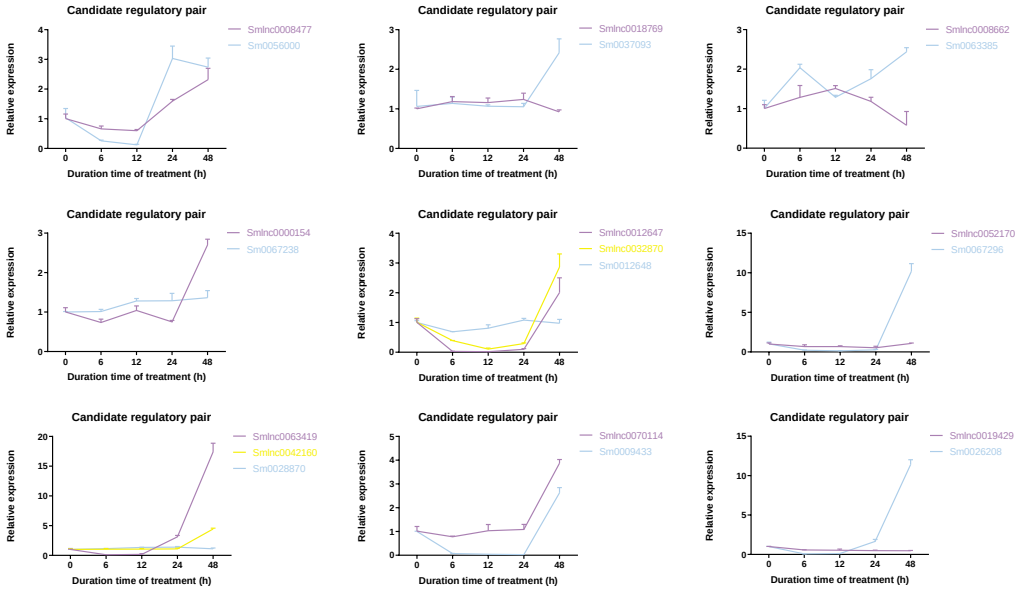

Figure 4 Time-series expression pattern of candidate lncRNA-mRNA/TF pairs in MeJA-induced *S. miltiorrhiza*. Expression of Smlnc0008477, Smlnc0018769, Smlnc0008662, Smlnc0000154, Smlnc0012647, Smlnc0032870, Smlnc0052170, Smlnc0063419, Smlnc0042160, Smlnc0070114, Smlnc0019429 and the target gene: *Sm0056000*, *Sm0037093*, *Sm0063385, Sm0067238*, *Sm0012648*, *Sm0067296*, *Sm0028870*, *Sm0009433*, *Sm0026208* in MeJA-induced *S. miltiorrhiza* at five time points (0 h, 6 h, 12 h, 24 h, and 48 h). Fold changes of expression levels were shown. The relative expression levels were normalized against β-actin levels. The level of lncRNA, mRNA, and TF genes in untreated roots (0 h) was arbitrarily set to 1. Error bars represent the standard deviations of three qRT-PCR replicates.
Full-size DOI: 10.7717/peerj.15332/fig-4

Time-series expression pattern of candidate lncRNA-mRNA/TF pairs induced by MeJA and module detection in *S. miltiorrhiza*

To validate whether the candidate lncRNA-mRNA/TF pairs are related to diterpenoid biosynthesis, we analyzed the time-series expression patterns of candidate gene pairs in response to MeJA treatment. MeJA treatment significantly changed the expression of 19 genes in *S. miltiorrhiza* at least a time-point of MeJA treatment (Fig. S3). As shown in Fig. 4, the expression of most genes showed a downward trend at the 6 h time point, and only the *Smlnc0018769-Sm0037093* pair and *Smlnc0008662-Sm0063385* pair had an upward trend. The *Smlnc0008477-Sm0056000* pair and *Smlnc0012647-Smlnc0032870* pair had the same expression trend. The *Smlnc0042160-Sm0028870* pair had the same expression trends at 6, 12, and 24 h. The *Smlnc0018769-Sm0037093*, *Smlnc0008662-Sm0063385*, and *Smlnc0052170-Sm0067296* pairs had opposite expression trends at 48 h, whereas the *Smlnc0019429-Sm0026208* pair had opposite trends at 24 and 48 h. Interestingly, the two gene pairs in the co-expression network, *Smlnc0032870-Sm0012648* and *Smlnc0018769-Sm0037093*, which were positively correlated, showed complex expression correlation under the induction of MeJA, even showed opposite expression trends after 24 h. The hub genes: *Sm0056000*, *Sm0037097*, *Sm0063385*, and *Sm0067296* in the co-expression network showed significantly differential expression in MeJA-induced *S. miltiorrhiza*.

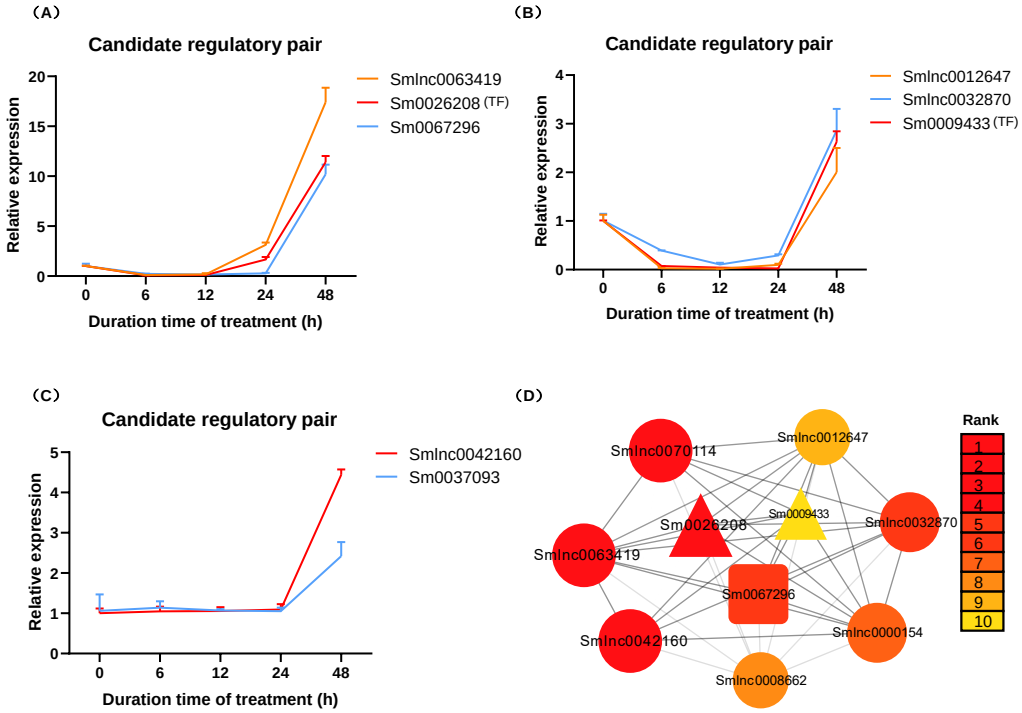

Figure 5 LncRNA-mRNA-TF module, lncRNA-TF module, lncRNA-mRNA module and hub genes in *S. miltiorrhiza*. (A–C) Gene expression pattern analysis: three types of network modules, namely, lncRNA-mRNA-TF module, lncRNA-TF module and lncRNA-mRNA module. (D) Network module analysis of lncRNAs, mRNAs, and TFs by cytoHubba. The circle node represents lncRNA, the square node represents mRNA, and the triangle node represents TF gene. Rank 1 indicates the highest rank, and Rank 10 indicates the lowest rank.

Full-size DOI: 10.7717/peerj.15332/fig-5

The data in Fig. 5A showed that *Smlnc0063419*, *Sm0026208*, and *Sm0067296* had similar expression patterns, which formed the lncRNA-mRNA-TF module. *Sm0026208* and *Sm0067296* were annotated with TF *WRKY2* and mRNA *GA3ox2*, respectively. *Smlnc0012647*, *Smlnc0032870*, and *Sm0009433* had similar expression patterns, as shown in Fig. 5B. *Sm0009433* was annotated with TF *MYB36*. *Smlnc0042160* and *Sm0037093* had similar expression patterns, as shown in Fig. 5C. *Sm0037093* was annotated with mRNA *KS2*. Through the plug-in cytoHubba, we calculated the top 10 hub genes (Table S12 and Fig. 5D). The modules of *Smlnc0063419-Sm0026208-Sm0067296* and *Smlnc0012647-Smlnc0032870-Sm0009433*, in which all genes were present in the top 10 hub genes.

DISCUSSION

In this study, the transcriptomic data of *S. miltiorrhiza* during the tanshinone accumulation stage were used for exploring the lncRNAs and their target transcripts involved in the diterpenoid biosynthetic pathway. A total of 6,651 lncRNAs were obtained by a strict bioinformatic pipeline. We found some common features of lncRNAs, which may be related to their function. Among the candidate lncRNAs, diterpenoid biosynthetic genes/TFs in *S. miltiorrhiza*, we detected 23 candidate lncRNA-mRNA/TF pairs with a *cis*-regulatory

relationship. To further verify their relationship, we used MeJA as an inducer to observe the gene expression of these candidate lncRNA-mRNA/TF pairs with MeJA treatment for 6, 12, 24, and 48 h. Through the expression data of time series, which provided an exploratory method for the role of lncRNA, three lncRNA-mRNA and/or TF network modules were finally obtained.

In the present study, the identified lncRNAs of *S. miltiorrhiza* were found to be shorter in length compared with protein-coding transcripts, which was consistent with the previous reports (*Hao et al., 2015*; *Liu et al., 2018*; *Shen et al., 2018*). The mean GC content of lncRNAs was slightly lower than that of mRNAs, which has also been reported in *Populus tomentosa* (*Zhou et al., 2017*). Conservation analysis showed that plants of the same genus are more conservative than those of different families. A previous study also showed that the majority of lncRNAs have high sequence conservation at the intra-species and sub-species levels (*Deng et al., 2018*). Polymerase C-terminal domain (CTD) modifications [e.g., CTD modification threonine 4 phosphorylation (CTD-T4P)] are on the promoters of lncRNAs, which leads to the decrease in the polymerase pause and the advance of the termination in the whole lncRNA genome. The transcription rate of lncRNAs is very fast, which means that they can quickly act on the regulatory target and respond to the signal. Therefore, transcription accuracy and sequence conservation were low (*Rinn & Chang, 2020*). The lncRNA function was maintained across large evolutionary distances even when the lncRNA sequence substantially diverged (*Ulitsky, 2016*). Another possible explanation for this is that RNA secondary structures may be the units of lncRNA words rather than the primary sequence, and disparate sequences form similar structure–function relationships to transmit symbolic language like hieroglyphics, thereby forming the molecular grammar of lncRNAs (*Rinn & Chang, 2020*).

LncRNAs have been proposed to carry out their functions by *cis* or *trans*, transcriptional regulation *cis*-acting lncRNAs influence the expression and/or chromatin state of nearby genes. We predicted lncRNA-mRNA pairs in the *cis*-acting relationship. The results of our study showed that 45,198 and 180 lncRNA-mRNA/TF pairs were co-expressed ($|PCC| \geq 0.4$) or co-localized, respectively, and 23 lncRNA-mRNA/TF pairs were both co-localized and co-expressed (Table S10). The results indicated that most lncRNAs were not co-expressed with their nearby coding genes and were transcribed independently in *S. miltiorrhiza* (*Liao et al., 2011*). Two pairs were both co-expressed and co-localized in the high correlation co-expression network with $|PCC| \geq 0.8$. Thus, the modes of action of lncRNAs were not limited to local regulators. Another possible explanation is that the strength of the required gene co-expression may depend on the stability or toxicity of the metabolites, and strong co-expression should only be required for unstable monomers (*Obayashi & Kinoshita, 2009*).

Although advances have been made in the miRNA and miRNA target prediction fields, the precision of miRNA target prediction needs to be improved (*Akgül et al., 2022*). To reduce false positives, we used three kinds of prediction software to predict miRNA targets. Although many miRNA databases and prediction software are published for plants, few of them are available (*De Amorim, Pedro & Paschoal, 2022*). This limitation reduced our chance to find miRNAs associated with the lncRNA-mRNA/TF pairs in the diterpenoid

biosynthetic pathway. On the basis of the relationship between miRNAs and lncRNAs, we predicted that 14 lncRNAs were potential targets or TMs of 66 miRNAs in *S. miltiorrhiza*, and no one existed in the candidate lncRNA-mRNA/TF pairs. However, all the results were predicted preliminarily based on bioinformatic analyses and need to be further validated.

Building gene regulatory networks from transcriptomic studies often results in a static view of gene expression, which can make it difficult to disentangle the regulatory pathway structure response to a stimulus. Time-series expression analysis may uncover the temporal transcriptional logic for plant response systems, and provide more accurate predictions for targeted breeding (*Greenham & McClung, 2018*). By studying the time-series expression of some candidate regulatory pairs inducted by MeJA, we observed a more detailed landscape. We found that some lncRNAs were downregulated in the early stage. The expression of the corresponding mRNAs was upregulated in the later stage between the regulatory pairs, and plants' response to the signal had a time delay. For example, in response to vernalization, *COOLAIR* is transiently induced by prolonged cold, reaching a maximum expression level after 2 weeks (*Swiezewski et al., 2009*). Meanwhile, we found that some genes showed different expression patterns within a period, namely, both upregulated and downregulated, this phenomenon may be because gene regulatory networks are inherently complex, with multiple feedback and feedforward loops (*Wils & Kaufmann, 2017*).

Nineteen genes were differentially expressed significantly under MeJA-induced, and some of them were present in hub genes of the co-expression network. These results indicated that these genes might participate in the biosynthesis of secondary terpenoids such as tanshinone and also play an important role in the defense response of *S. miltiorrhiza*. Our results indicated that the expression levels of *Sm0056000* (*CPS*) and *Sm0037093* (*KS*) were upregulated under MeJA induction, which was also observed in the previous report (*Luo et al., 2014*). However, the levels of these two genes *Sm0056000* (*CPS*) and *Sm0037093* (*KS*) started to increase at 12 h and 24 h in our study, respectively. The response of TFs to MeJA was also observed in this study (*Luo et al., 2014*), which was consistent with the TF *WRKY* we studied. For MeJA treatment, the expression of *CYP76AH1* was up-regulated over time and reached a peak at 12 h in the previous study (*Li et al., 2021*), however, after peaking at 6 h, our *Sm0063385* (*CYP76AH1*) expression was down-regulated, and then it began to up-regulate at 12 h and peaked at 48 h. Previous research has suggested that *WRKYs* might regulate the development of bast fiber in response to GA_3 stress in jute (*Corchorus capsularis*) (*Zhang et al., 2020*), the *Smlnc0063419-Sm0026208-Sm0067296* module (*Sm0026208*: TF *WRKY2*, *Sm0067296*: *GA3ox2*) in our study may form a similar response module in *S. miltiorrhiza*.

The tissue specificity and subcellular localization of lncRNAs may suggest their function. We obtained 3 lncRNA-mRNA and/or TF modules: *Smlnc0063419-Sm0026208* (TF)-*Sm0067296*, *Smlnc0012647-Smlnc0032870-Sm0009433* (TF), and *Smlnc0042160-Sm0037093*, among of lncRNAs: *Smlnc0063419*, *Smlnc0012647*, *Smlnc0032870*, and *Smlnc0042160* mainly expressed in roots (Figs. 3 and 5), which was consistent with the place where tanshinones accumulated of *S. miltiorrhiza* (*Chang et al., 2019*). In the prediction of subcellular localization, *Smlnc0012647*, *Smlnc0042160*, and *Smlnc0063419* were predicted to be located in the cytoplasm, this suggests that these predicted cytoplasmic

lncRNAs may interfere with protein post-translational modifications or regulate mRNA export (*Chen, 2016*). Although *Smlnc0012647* and *Smlnc0032870* existed in the same module, *Smlnc0032870* was predicted to be located in the nucleus, a possible explanation for this result may be that some lncRNAs are located in both the nucleus and cytoplasm (*Cabili et al., 2015*).

CONCLUSIONS

This study set out to investigate the possible functions of the lncRNAs of *S. miltiorrhiza* related to diterpenoid biosynthesis, we predicted the potential targets of lncRNAs in *cis*-regulatory relationships, a summarizing figure is shown in Fig. S4. Through a strict bioinformatic pipeline, we identified 6,651 candidate lncRNAs, and obtained three lncRNA-mRNA and/or TF network modules. This study revealed the possible roles of the lncRNAs of *S. miltiorrhiza* related to diterpenoid biosynthesis. These findings indicated that lncRNA is generally complex in regulating mRNA and/or TF. This study provides useful information to deepen our understanding of the function and regulatory mechanisms of lncRNAs in the diterpenoid biosynthetic pathway of *S. miltiorrhiza*.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This study was funded by the National Natural Science Foundation of China (81973416). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
The National Natural Science Foundation of China: 81973416.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Lin Wang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Peijin Zou performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Fang Liu and Rui Liu performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Zhu-Yun Yan conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Xin Chen conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data is available at NCBI: PRJNA712174 and in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.15332#supplemental-information.

REFERENCES

Akgül B, Stadler PF, Hawkins LJ, Hadj-Moussa H, Storey KB, Ergin K, Çetinkaya R, Paschoal AR, Nachtigall PG, Tutar Y, Yousef M, Allmer J. 2022. 44 current challenges in miRNomics. *Methods in Molecular Biology* **2257**:423–438 DOI 10.1007/978-1-0716-1170-8_19.

Bonnet E, He Y, Billiau K, Van de Peer Y. 2010. TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics* **26**:1566–1568 DOI 10.1093/bioinformatics/btq233.

Brockdorff N. 1998. The role of Xist in X-inactivation. *Current Opinion in Genetics & Development* **8**:328–333 DOI 10.1016/S0959-437X(98)80090-7.

Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL, Raj A. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biology* **16**:20 DOI 10.1186/s13059-015-0586-4.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421 DOI 10.1186/1471-2105-10-421.

Cao Z, Pan X, Yang Y, Huang Y, Shen H-B. 2018. The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics* **34**:2185–2194 DOI 10.1093/bioinformatics/bty085.

Carlevaro-Fita J, Johnson R. 2019. Global positioning system: understanding long noncoding RNAs through subcellular localization. *Molecular Cell* **73**:869–883 DOI 10.1016/j.molcel.2019.02.008.

Chang Y, Wang M, Li J, Lu S. 2019. Transcriptomic analysis reveals potential genes involved in tanshinone biosynthesis in *Salvia miltiorrhiza*. *Scientific Reports* **9**:14929 DOI 10.1038/s41598-019-51535-9.

Chen L-L. 2016. Linking long noncoding RNA localization and function. *Trends in Biochemical Sciences* **41**:761–772 DOI 10.1016/j.tibs.2016.07.003.

Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8**:S11 DOI 10.1186/1752-0509-8-S4-S11.

Dai X, Zhuang Z, Zhao PX. 2018. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Research* **46**:W49–W54 DOI 10.1093/nar/gky316.

De Amorim TS, Pedro DLF, Paschoal AR. 2022. MicroRNA databases and tools. *Methods in Molecular Biology* **2257**:131–166 DOI 10.1007/978-1-0716-1170-8_7.

Deng P, Liu S, Nie X, Weining S, Wu L. 2018. Conservation analysis of long non-coding RNAs in plants. *Science China Life Sciences* **61**:190–198 DOI 10.1007/s11427-017-9174-9.

Du Q, Li C, Li D, Lu S. 2015. Genome-wide analysis, molecular cloning and expression profiling reveal tissue-specifically expressed, feedback-regulated, stress-responsive and alternatively spliced novel genes involved in gibberellin metabolism in *Salvia miltiorrhiza*. *BMC Genomics* **16**:1087 DOI 10.1186/s12864-015-2315-5.

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES. 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**:452–455 DOI 10.1038/nature20149.

Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, García JA, Paz-Ares J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genetics* **39**:1033–1037 DOI 10.1038/ng2079.

Gao W, Hillwig ML, Huang L, Cui G, Wang X, Kong J, Yang B, Peters RJ. 2009. A functional genomics approach to Tanshinone biosynthesis provides stereochemical insights. *Organic Letters* **11**:5170–5173 DOI 10.1021/ol902051v.

Greenham K, McClung CR. 2018. Time to build on good design: resolving the temporal dynamics of gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America* **115**:6325–6327 DOI 10.1073/pnas.1807707115.

Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. 2008. The Vienna RNA websuite. *Nucleic Acids Research* **36**:W70–W74 DOI 10.1093/nar/gkn188.

Guil S, Esteller M. 2012. Cis-acting noncoding RNAs: friends and foes. *Nature Structural & Molecular Biology* **19**:1068–1075 DOI 10.1038/nsmb.2428.

Gundlach H, Müller MJ, Kutchan TM, Zenk MH. 1992. Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures. *Proceedings of the National Academy of Sciences of the United States of America* **89**:2389–2393 DOI 10.1073/pnas.89.6.2389.

Guo J, Zhou YJ, Hillwig ML, Shen Y, Yang L, Wang Y, Zhang X, Liu W, Peters RJ, Chen X, Zhao ZK, Huang L. 2013. CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proceedings of the National Academy of Sciences of the United States of America* **110**:12108–12113 DOI 10.1073/pnas.1218061110.

Guo Z, Kuang Z, Zhao Y, Deng Y, He H, Wan M, Tao Y, Wang D, Wei J, Li L, Yang X. 2022. PmiREN2.0: from data annotation to functional exploration of plant microRNAs. *Nucleic Acids Research* **50**:D1475–D1482 DOI 10.1093/nar/gkab811.

Hao Z, Fan C, Cheng T, Su Y, Wei Q, Li G. 2015. Genome-wide identification, characterization and evolutionary analysis of long intergenic noncoding RNAs in *Cucumber*. *PLOS ONE* **10**:e0121800 DOI 10.1371/journal.pone.0121800.

Hedden P, Thomas SG. 2012. Gibberellin biosynthesis and its regulation. *Biochemical Journal* **444**:11–25 DOI 10.1042/BJ20120245.

Helliwell CA, Chandler PM, Poole A, Dennis ES, Peacock WJ. 2001. The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. *Proceedings of the National Academy of Sciences of the United States of America* **98**:2065–2070 DOI 10.1073/pnas.98.4.2065.

Huang L, Dong H, Zhou D, Li M, Liu Y, Zhang F, Feng Y, Yu D, Lin S, Cao J. 2018. Systematic identification of long non-coding RNAs during pollen development and fertilization in *Brassica rapa*. *The Plant Journal* **96**:203–222 DOI 10.1111/tpj.14016.

Ji AJ, Luo HM, Xu ZC, Zhang X, Zhu YJ, Liao BS, Yao H, Song JY, Chen SL. 2016. Genome-wide identification of the AP2/ERF gene family involved in active constituent biosynthesis in *Salvia miltiorrhiza*. *The Plant Genome* **9**(**2**):plantgenome2015.08.0077 DOI 10.3835/plantgenome2015.08.0077.

Jiang M, Chen H, Liu J, Du Q, Lu S, Liu C. 2021. Genome-wide identification and functional characterization of natural antisense transcripts in *Salvia miltiorrhiza*. *Scientific Reports* **11**:4769 DOI 10.1038/s41598-021-83520-6.

Kai G, Liao P, Zhang T, Zhou W, Wang J, Xu H, Liu Y, Zhang L. 2010. Characterization, expression profiling, and functional identification of a gene encoding geranylgeranyl diphosphate synthase from *Salvia miltiorrhiza*. *Biotechnology and Bioprocess Engineering* **15**:236–245 DOI 10.1007/s12257-009-0123-y.

Kai G, Xu H, Zhou C, Liao P, Xiao J, Luo X, You L, Zhang L. 2011. Metabolic engineering tanshinone biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *Metabolic Engineering* **13**:319–327 DOI 10.1016/j.ymben.2011.02.003.

Kang Y-J, Yang D-C, Kong L, Hou M, Meng Y-Q, Wei L, Gao G. 2017. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research* **45**:W12–W16 DOI 10.1093/nar/gkx428.

Kopp F, Mendell JT. 2018. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**:393–407 DOI 10.1016/j.cell.2018.01.011.

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Research* **47**:D155–D162 DOI 10.1093/nar/gky1141.

Li A, Zhang J, Zhou Z. 2014. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved *k*-mer scheme. *BMC Bioinformatics* **15**:311 DOI 10.1186/1471-2105-15-311.

Li B, Li J, Chai Y, Huang Y, Li L, Wang D, Wang Z. 2021. Targeted mutagenesis of CYP76AK2 and CYP76AK3 in *Salvia miltiorrhiza* reveals their roles in tanshinones biosynthetic pathway. *International Journal of Biological Macromolecules* **189**:455–463 DOI 10.1016/j.ijbiomac.2021.08.112.

Li C, Li D, Shao F, Lu S. 2015. Molecular cloning and expression analysis of *WRKY* transcription factor genes in *Salvia miltiorrhiza*. *BMC Genomics* **16**:200 DOI 10.1186/s12864-015-1411-x.

Li C, Lu S. 2014. Genome-wide characterization and comparative analysis of R2R3-MYB transcription factors shows the complexity of MYB-associated regulatory networks in *Salvia miltiorrhiza*. *BMC Genomics* **15**:277 DOI 10.1186/1471-2164-15-277.

Li D, Shao F, Lu S. 2015. Identification and characterization of mRNA-like noncoding RNAs in *Salvia miltiorrhiza*. *Planta* **241**:1131–1143 DOI 10.1007/s00425-015-2246-z.

Li S, Yamada M, Han X, Ohler U, Benfey PN. 2016. High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. *Developmental Cell* **39**:508–522 DOI 10.1016/j.devcel.2016.10.012.

Li W, Bai Z, Pei T, Yang D, Mao R, Zhang B, Liu C, Liang Z. 2019. SmGRAS1 and SmGRAS2 regulate the biosynthesis of Tanshinones and phenolic acids in *Salvia miltiorrhiza*. *Frontiers in Plant Science* **10**:1367 DOI 10.3389/fpls.2019.01367.

Liang Z-S, Yang D-F, Liang X, Zhang Y-J, Liu Y, Liu F-H. 2012. Roles of reactive oxygen species in methyl jasmonate and nitric oxide-induced tanshinone production in *Salvia miltiorrhiza* hairy roots. *Plant Cell Reports* **31**:873–883 DOI 10.1007/s00299-011-1208-6.

Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, Zhao G, Luo H, Bu D, Zhao H, Skogerbø G, Wu Z, Zhao Y. 2011. Large-scale prediction of long non-coding RNA functions in a coding–non-coding gene co-expression network. *Nucleic Acids Research* **39**:3864–3878 DOI 10.1093/nar/gkq1348.

Liu H, Lu Y, Wang J, Hu J, Wuyun T. 2018. Genome-wide screening of long non-coding RNAs involved in rubber biosynthesis in *Eucommia ulmoides*. *Journal of Integrative Plant Biology* **60**:1070–1082 DOI 10.1111/jipb.12693.

Liu H, Wang R, Mao B, Zhao B, Wang J. 2019. Identification of lncRNAs involved in rice ovule development and female gametophyte abortion by genome-wide screening and functional analysis. *BMC Genomics* **20**:90 DOI 10.1186/s12864-019-5442-6.

Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua N-H. 2012. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *The Plant Cell* **24**:4333–4345 DOI 10.1105/tpc.112.102855.

Lu S. 2021. Biosynthesis and regulatory mechanisms of bioactive compounds in *Salvia miltiorrhiza*, a model system for medicinal plant biology. *Critical Reviews in Plant Sciences* **40**:243–283 DOI 10.1080/07352689.2021.1935719.

Luo H, Zhu Y, Song J, Xu L, Sun C, Zhang X, Xu Y, He L, Sun W, Xu H, Wang B, Li X, Li C, Liu J, Chen S. 2014. Transcriptional data mining of *Salvia miltiorrhiza* in response to methyl jasmonate to examine the mechanism of bioactive compound biosynthesis and regulation. *Physiologia Plantarum* **152**:241–255 DOI 10.1111/ppl.12193.

Ma X-H, Ma Y, Tang J-F, He Y-L, Liu Y-C, Ma X-J, Shen Y, Cui G-H, Lin H-X, Rong Q-X, Guo J, Huang L-Q. 2015. The biosynthetic pathways of tanshinones and phenolic acids in *Salvia miltiorrhiza*. *Molecules* **20**:16235–16254 DOI 10.3390/molecules200916235.

Ma Y, Cui G, Chen T, Ma X, Wang R, Jin B, Yang J, Kang L, Tang J, Lai C, Wang Y, Zhao Y, Shen Y, Zeng W, Peters RJ, Qi X, Guo J, Huang L. 2021. Expansion within the CYP71D subfamily drives the heterocyclization of tanshinones synthesis in *Salvia miltiorrhiza*. *Nature Communications* **12**:685 DOI 10.1038/s41467-021-20959-1.

Ma Y, Ma X-H, Meng F-Y, Zhan Z-L, Guo J, Huang L-Q. 2016. RNA interference targeting CYP76AH1 in hairy roots of *Salvia miltiorrhiza* reveals its key role in the biosynthetic pathway of tanshinones. *Biochemical and Biophysical Research Communications* **477**:155–160 DOI 10.1016/j.bbrc.2016.06.036.

Ma Y, Yuan L, Wu B, Li X, Chen S, Lu S. 2012. Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in *Salvia miltiorrhiza*. *Journal of Experimental Botany* **63**:2809–2823 DOI 10.1093/jxb/err466.

Mercer TR, Dinger ME, Mattick JS. 2009. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics* **10**:155–159 DOI 10.1038/nrg2521.

Nagano T, Fraser P. 2011. No-nonsense functions for long noncoding RNAs. *Cell* **145**:178–181 DOI 10.1016/j.cell.2011.03.014.

Obayashi T, Kinoshita K. 2009. Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. *DNA Research* **16**:249–260 DOI 10.1093/dnares/dsp016.

Palazzo AF, Koonin EV. 2020. Functional long non-coding RNAs evolve from junk transcripts. *Cell* **183**:1151–1161 DOI 10.1016/j.cell.2020.09.047.

Paschoal AR, Lozada-Chávez I, Domingues DS, Stadler PF. 2017. ceRNAs in plants: computational approaches and associated challenges for target mimic research. *Briefings in Bioinformatics* **19**:1273–1289 DOI 10.1093/bib/bbx058.

Ponting CP, Oliver PL, Reik W. 2009. Evolution and functions of long noncoding RNAs. *Cell* **136**:629–641 DOI 10.1016/j.cell.2009.02.006.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Research* **40**:D290–D301 DOI 10.1093/nar/gkr1065.

Quan M, Du Q, Xiao L, Lu W, Wang L, Xie J, Song Y, Xu B, Zhang D. 2019. Genetic architecture underlying the lignin biosynthesis pathway involves noncoding RNAs and transcription factors for growth and wood properties in *Populus*. *Plant Biotechnology Journal* **17**:302–315 DOI 10.1111/pbi.12978.

Rai MI, Alam M, Lightfoot DA, Gurha P, Afzal AJ. 2019. Classification and experimental identification of plant long non-coding RNAs. *Genomics* **111**:997–1005 DOI 10.1016/j.ygeno.2018.04.014.

Rao X, Dixon RA. 2019. Co-expression networks for plant biology: why and how. *Acta Biochimica Et Biophysica Sinica* **51**:981–988 DOI 10.1093/abbs/gmz080.

Revelle W. 2022. psych: procedures for psychological, psychometric, and personality research. R package version 2.2.5. Evanston: Northwestern University. *Available at https://cran.r-project.org/web/packages/psych/index.html*.

Rinn JL, Chang HY. 2020. Long noncoding RNAs: molecular modalities to organismal functions. *Annual Review of Biochemistry* **89**:283–308 DOI 10.1146/annurev-biochem-062917-012708.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**:2498–2504 DOI 10.1101/gr.1239303.

Shen E, Zhu X, Hua S, Chen H, Ye C, Zhou L, Liu Q, Zhu Q-H, Fan L, Chen X. 2018. Genome-wide identification of oil biosynthesis-related long non-coding RNAs in allopolyploid *Brassica napus*. *BMC Genomics* **19**:745 DOI 10.1186/s12864-018-5117-8.

Shimane M, Ueno Y, Morisaki K, Oogami S, Natsume M, Hayashi K, Nozaki H, Kawaide H. 2014. Molecular evolution of the substrate specificity of *ent*-kaurene synthases to adapt to gibberellin biosynthesis in land plants. *Biochemical Journal* **462**:539–546 DOI 10.1042/BJ20140134.

Swiezewski S, Liu F, Magusin A, Dean C. 2009. Cold-induced silencing by long antisense transcripts of an arabidopsis polycomb target. *Nature* **462**:799–802 DOI 10.1038/nature08618.

Ulitsky I. 2016. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nature Reviews Genetics* **17**:601–614 DOI 10.1038/nrg.2016.85.

Van Schie CCN, Ament K, Schmidt A, Lange T, Haring MA, Schuurink RC. 2007. Geranyl diphosphate synthase is required for biosynthesis of gibberellins. *The Plant Journal* **52**:752–762 DOI 10.1111/j.1365-313X.2007.03273.x.

Wang H-Lv, Chekanova JA. 2017. Long noncoding RNAs in plants. *Advances in Experimental Medicine and Biology* **1008**:133–154 DOI 10.1007/978-981-10-5203-3_5.

Wang Y, Zhang H, Li Q, Jin J, Chen H, Zou Y, Huang X, Ding Y. 2021. Genome-wide identification of lncRNAs involved in fertility transition in the photo-thermosensitive genic male sterile rice line Wuxiang S. *Frontiers in Plant Science* **11**:580050 DOI 10.3389/fpls.2020.580050.

Wasternack C. 2007. Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Annals of Botany* **100**:681–697 DOI 10.1093/aob/mcm079.

Wils CR, Kaufmann K. 2017. Gene-regulatory networks controlling inflorescence and flower development in Arabidopsis thaliana. *Biochimica Et Biophysica Acta (BBA)—Gene Regulatory Mechanisms* **1860**:95–105 DOI 10.1016/j.bbagrm.2016.07.014.

Wu B, Li Y, Yan H, Ma Y, Luo H, Yuan L, Chen S, Lu S. 2012a. Comprehensive transcriptome analysis reveals novel genes involved in cardiac glycoside biosynthesis and mlncRNAs associated with secondary metabolism and stress response in *Digitalis purpurea*. *BMC Genomics* **13**:15 DOI 10.1186/1471-2164-13-15.

Wu H-J, Ma Y-K, Chen T, Wang M, Wang X-J. 2012b. PsRobot: a web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Research* **40**:W22–W28 DOI 10.1093/nar/gks554.

Xiangdong PU, Lizhi W. 2017. Cloning and bioinformatic analysis of CYP76AK5 related to Tanshinone biosynt hesis in *Salvia miltiorrhiza*. *Modern Chinese Medicine* **19**:1106–1110 DOI 10.13313/j.issn.1673-4890.2017.8.011.

Xiao Y, Zhang L, Gao S, Saechao S, Di P, Chen J, Chen W. 2011. The c4 h, tat, hppr and hppd genes prompted engineering of rosmarinic acid biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *PLOS ONE* **6**:e29713 DOI 10.1371/journal.pone.0029713.

Xu H, Song J, Luo H, Zhang Y, Li Q, Zhu Y, Xu J, Li Y, Song C, Wang B, Sun W, Shen G, Zhang X, Qian J, Ji A, Xu Z, Luo X, He L, Li C, Sun C, Yan H, Cui G, Li X, Li X, Wei J, Liu J, Wang Y, Hayward A, Nelson D, Ning Z, Peters RJ, Qi X, Chen S. 2016. Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Molecular Plant* **9**:949–952 DOI 10.1016/j.molp.2016.03.010.

Xu Z, Peters RJ, Weirather J, Luo H, Liao B, Zhang X, Zhu Y, Ji A, Zhang B, Hu S, Au KF, Song J, Chen S. 2015. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *The Plant Journal* **82**:951–961 DOI 10.1111/tpj.12865.

Yamaguchi S. 2008. Gibberellin metabolism and its regulation. *Annual Review of Plant Biology* **59**:225–251 DOI 10.1146/annurev.arplant.59.032607.092804.

Yang Y, Hou S, Cui G, Chen S, Wei J, Huang L. 2010. Characterization of reference genes for quantitative real-time PCR analysis in various tissues of *Salvia miltiorrhiza*. *Molecular Biology Reports* **37**:507–513 DOI 10.1007/s11033-009-9703-3.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**:134 DOI 10.1186/1471-2105-13-134.

Yin D-D, Li S-S, Shu Q-Y, Gu Z-Y, Wu Q, Feng C-Y, Xu W-Z, Wang L-S. 2018. Identification of microRNAs and long non-coding RNAs involved in fatty acid biosynthesis in tree peony seeds. *Gene* **666**:72–82 DOI 10.1016/j.gene.2018.05.011.

Yu T, Tzeng DTW, Li R, Chen J, Zhong S, Fu D, Zhu B, Luo Y, Zhu H. 2019. Genome-wide identification of long non-coding RNA targets of the tomato MADS box transcription factor RIN and function analysis. *Annals of Botany* **123**:469–482 DOI 10.1093/aob/mcy178.

Zhang B, Gunawardane L, Niazi F, Jahanbani F, Chen X, Valadkhan S. 2014. A novel RNA motif mediates the strict nuclear localization of a long noncoding RNA. *Molecular and Cellular Biology* **34**:2318–2329 DOI 10.1128/MCB.01673-13.

Zhang L, Wan X, Xu Y, Niyitanga S, Qi J, Zhang L. 2020. De novo assembly of transcriptome and genome-wide identification reveal GA3 stress-responsive WRKY transcription factors involved in fiber formation in jute (*Corchorus capsularis*). *BMC Plant Biology* **20**:403 DOI 10.1186/s12870-020-02617-8.

Zhang X, Luo H, Xu Z, Zhu Y, Ji A, Song J, Chen S. 2015. Genome-wide characterisation and analysis of bHLH transcription factors related to tanshinone biosynthesis in *Salvia miltiorrhiza*. *Scientific Reports* **5**:11244 DOI 10.1038/srep11244.

Zhou D, Du Q, Chen J, Wang Q, Zhang D. 2017. Identification and allelic dissection uncover roles of lncRNAs in secondary growth of Populus tomentosa. *DNA Research* **24**:473–486 DOI 10.1093/dnares/dsx018.

Zhou J, Liu R, Shuai M, Yan Z-Y, Chen X. 2021. Comparative transcriptome analyses of different *Salvia miltiorrhiza* varieties during the accumulation of tanshinones. *PeerJ* **9**:e12300 DOI 10.7717/peerj.12300.