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

Lin Wang[1,2], Peijin Zou[1,2], Fang Liu[1,2], Rui Liu[1,2], Zhu-Yun Yan[1,2] and Xin Chen[1,2]

[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

Corresponding Author:
Xin Chen[1,2]

Email address: chenxin@cdutcm.edu.cn

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.

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,

39 2009; Nagano & Fraser, 2011; Palazzo & Koonin, 2020). They are usually transcribed by RNA
40 polymerase II (Rinn & Chang, 2020). Functional analysis of eukaryotic lncRNAs has revealed
41 that they act as molecular scaffolds, guide molecules, molecular sponges and decoys, precursors
42 of microRNAs (miRNAs) and other small RNAs, or as miRNA target mimics (TMs) to regulate
43 gene expression at multiple levels (epigenetic regulation, transcriptional regulation, and post-
44 transcriptional regulation) (Franco-Zorrilla et al., 2007; Mercer, Dinger & Mattick, 2009; Wang
45 & Chekanova, 2017; Rai et al., 2019). LncRNAs can also act on transcription factors (TFs); for
46 example, lncRNA can be used as a TF-binding site to regulate its expression (Yu et al., 2019).
47 LncRNAs are usually expressed at a low level and in a tissue-specific manner (Liu et al., 2012;
48 Palazzo & Koonin, 2020). The subcellular localization of lncRNAs is the primary determinant of
49 their molecular functions (Carlevaro-Fita & Johnson, 2019).
50
51 There is much evidence suggesting that lncRNAs play key roles in plant secondary metabolism.
52 For example, mLncR8 putatively regulates terpenoid biosynthesis, and mLncR31 is involved in
53 the biosynthesis of the isoprenoid side chain of ubiquinone and plastoquinone in *Digitalis*
54 *purpurea* (Wu et al., 2012a). LncRNAs might regulate genes in the phenylpropanoid pathway of
55 *Populus tomentosa* (Zhou et al., 2017). LncRNAs were also found to be involved in rubber
56 biosynthesis in *Eucommia ulmoides* (Liu et al., 2018). LncRNAs were possibly involved in the
57 biosynthesis of different fatty acids and lipid metabolism through post-transcriptional regulation
58 in tree peony seeds (Yin et al., 2018). LncRNAs might be involved in the lignin biosynthetic
59 pathway in *Populus* (Quan et al., 2019).
60
61 Studies have suggested that lncRNAs can act as local regulators, and lncRNA expression is
62 correlated with the expression of nearby genes (Guil & Esteller, 2012; Engreitz et al., 2016).
63 These correlations are attributed to sequence-specific functions of the mature lncRNA transcript,
64 the transcription or splicing of an RNA, or DNA elements within the lncRNA promoter or gene
65 locus, namely *cis*-regulation (Guil & Esteller, 2012; Engreitz et al., 2016; Kopp & Mendell,
66 2018).
67
68 Co-expressed genes are usually members of the same protein complex or metabolic pathway, and
69 they are functionally controlled by the same transcriptional regulatory program. The genes or
70 proteins within the co-expression network may have the same expression patterns (Liao et al.,
71 2011; Rao & Dixon, 2019).
72
73 The medicinal plant *S. miltiorrhiza* produces a variety of diterpenoids (Ma et al., 2015).
74 Tanshinones are the main bioactive compounds of *S. miltiorrhiza*, and they mainly accumulate in
75 the roots of *S. miltiorrhiza* (Xu et al., 2015; Chang et al., 2019). Another diterpenoid in *S.*
76 *miltiorrhiza* is the plant hormone gibberellin (GA), which is one of the five classic plant
77 hormones (Brockdorff, 1998). In general, the biosynthetic pathway of terpenoids in *S.*
78 *miltiorrhiza* can be divided into three stages (Ma et al., 2012). The first stage leads to the

79 synthesis of the universal isoprene precursor isopentenyl diphosphate (IPP) and its isomer
80 dimethylallyl diphosphate (DMAPP) through the 2-C-methyl-D-erythritol 4-phosphate (MEP)
81 pathway and/or the mevalonate (MVA) pathway. In the second stage, the intermediate
82 diphosphate precursors, including geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and
83 geranylgeranyl diphosphate (GGPP) are synthesized under the catalysis of isoprenyl diphosphate
84 synthases (IDSs), including geranyl diphosphate synthase (GPPS), farnesyl diphosphate synthase
85 (FPPS), and geranylgeranyl diphosphate synthase (GGPPS). The last stage involves the
86 formation of diverse diterpenoids under the catalysis of terpene synthases (TPSs), such as
87 copalyl diphosphate synthase (CPS) and kaurene synthase (KS) catalyze the formation of
88 miltiradiene (Kai et al., 2010, 2011; Lu, 2021), *ent*-copalyl diphosphate synthase (*ent*-CPS)
89 (Shimane et al., 2014) and *ent*-kaurene synthase (*ent*-KS) are involved in the conversion of
90 GGPP to the tetracyclic hydrocarbon intermediate *ent*-kaurene (Yamaguchi, 2008; Shimane et
91 al., 2014). Then tanshinones and GAs are formed by cytochrome P450 monooxygenases (P450s)
92 and 2-oxoglutarate-dependent dioxygenases (2ODDs) modification.

93

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

103

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

112

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

119

Materials & Methods

Plant materials and MeJA treatment

122 *S. miltiorrhiza* seedlings were cultured in the greenhouse under 23°C–25°C. The plants were
123 sprayed with 200 μM MeJA solution as mentioned in a previous report (Luo et al., 2014). After
124 being treated with MeJA solution for 6, 12, 24, and 48 h, the treated and the non-treated (0 h)
125 roots were collected and rinsed with water. These roots were dried gently and quickly with
126 absorbent paper. The cleaned roots were immediately frozen in liquid nitrogen and stored at
127 −80°C until RNA isolation. Three biological replicates were carried out for each experiment.
128

S. miltiorrhiza genomic and transcriptomic data

130 The *S. miltiorrhiza* genome data were downloaded from the *S. miltiorrhiza* Genomics Database
131 (http://www.ndctcm.org/shujukujieshao/2015-04-23/27.html) (Xu et al., 2016) and NCBI
132 (National Center for Biotechnology Information) BioProject: PRJNA682867 (Ma et al., 2021).
133 The transcriptomic data were obtained from our previous study (Zhou et al., 2021) with an
134 accession number assigned to PRJNA712174. The transcriptomic data were gathered from four
135 varieties of *S. miltiorrhiza* root tissues during the tanshinone accumulation stage and included
136 Fragments Per Kilobase Million (FPKM) expression value and annotation information.
137

Pipeline for lncRNA identification

139 The following four filters were used to shortlist the bona fide lncRNAs from the transcriptomic
140 data: (1) transcripts with annotation information in one of the Nr (Non-RedundantProtein
141 Sequence), Swiss-Prot, and COG/KOG databases were removed; (2) transcripts shorter than 200
142 nt with an open reading frame (ORF) longer than 100 aa were discarded, found and extracted
143 ORFs on getorf (http://emboss.bioinformatics.nl/cgi-bin/emboss/getorf), selection of ORF length
144 < 100 aa by R (version 4.2.1) script; (3) transcripts were searched against the Pfam database
145 (Punta et al., 2012) (http://pfam.xfam.org/) by HMMER to remove transcripts possibly
146 containing known protein domains; and (4) the protein-coding potential of each transcript was
147 calculated using PLEK (Li, Zhang & Zhou, 2014) and Coding Potential Calculator 2 (CPC2,
148 http://cpc2.cbi.pku.edu.cn) (Kang et al., 2017), and these with PLEK and CPC2 scores > 0 were
149 discarded. Through the above process, identified lncRNAs were obtained. FPKM value less than
150 0.05 was used as the standard for low expression levels (Li et al., 2016). The transcripts that
151 remained were regarded as candidate lncRNAs.
152

Characterization and conservation analysis of *S. miltiorrhiza* lncRNAs

154 To gain more understanding of these lncRNAs in *S. miltiorrhiza*, we compared several different
155 features of the lncRNAs and mRNAs: GC content and transcript length. GC content and
156 transcript length were determined by R (version 4.2.1), and statistical analysis was carried out by
157 Excel.
158

159 We aligned the lncRNA sequences identified here with BLAST+ (Camacho et al., 2009) (blast-
160 2.11.0+) against the genome sequences of the *Lamiaceae* family: *Salvia splendens*, *Salvia*
161 *hispanica*, *Mentha longifolia*, *Scutellaria baicalensis*, *Pogostemon cablin*, and *Sesamum*
162 *indicum,* of which *Salvia splendens* and *Salvia hispanica* both belong to the *Salvia* genus;
163 *Solanaceae* family: *Nicotiana tabacum*, *Brassicaceae* family: *Brassica napus* and *Arabidopsis*
164 *thaliana*; and *Selaginellaceae* family: *Selaginella moellendorffii*. A cutoff E-value < 1e-10 was
165 used. The genomes were downloaded from the NCBI databases: GCF_004379255.1 (SspV2),
166 GCF_023119035.1 (UniMelb_Shisp_WGS_1.0), GCA_001642375.2 (Mlong_CMEN585_v),
167 GCA_005771605.1 (ASM577160v1), GCA_023678885.1 (GZUCM_PCab_1.0),
168 GCF_000512975.1 (S_indicum_v1.0), GCF_000715135.1 (Ntab-TN90), GCF_020379485.1
169 (Da-Ae), GCF_000001735.4 (TAIR10.1), and GCF_000143415.4 (v1.0). The lncRNAs that had
170 coverage of > 20% of matched regions were defined as conserved lncRNAs.
171

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

174 The candidate lncRNAs may act as the precursors of miRNAs, the *S. miltiorrhiza* miRNAs in
175 miRBase (Kozomara, Birgaoanu & Griffiths-Jones, 2019) (Release 22.1,
176 http://www.mirbase.org/) and PmiREN2.0 (Guo et al., 2022) (https://pmiren.com/) were aligned
177 to the sequences of the candidate lncRNAs. The secondary structure of lncRNAs was predicted
178 by RNAfold (Gruber et al., 2008) (http://rna.tbi.univie.ac.at/). LncRNAs with classical stem-loop
179 hairpins were regarded as the putative precursors of miRNA (Zhou et al., 2017). Given that
180 miRNA targets and lncRNAs have highly similar miRNA-binding sites, miRNA can be
181 sequestered by lncRNA (Paschoal et al., 2017). Three kinds of prediction software were used to
182 determine the miRNAs targeted to candidate lncRNAs. The first was TAPIR (Bonnet et al.,
183 2010) (http://bioinformatics.psb.ugent.be/webtools/tapir); this server offers the possibility to
184 search for plant miRNA targets using a fast and precise algorithm (score ≤ 4, free energy ratio ≥
185 0.7). The second was psRobot (Wu et al., 2012b) (Version 1.2) with default parameters, which is
186 a widely used online miRNA target prediction tool. The third was psRNATarget (2017 release)
187 with default settings, and psRNATarget was developed to identify plant sRNA targets by (i)
188 analyzing complementary matching between the sRNA sequence and target mRNA sequence
189 using a predefined scoring schema and (ii) by evaluating target site accessibility (Dai, Zhuang &
190 Zhao, 2018), targets with an E value less than 5.0 were retained as potential miRNA targets.
191

Subcellular localization of lncRNAs

193 The subcellular localization of *S. miltiorrhiza* lncRNAs was predicted by lncLocator (Cao et al.,
194 2018). LncLocator is an ensemble classifier-based predictor, which adopts both *k-mer* features
195 and high-level abstraction features generated by unsupervised deep models and constructs four
196 classifiers by feeding these two types of features to support vector machine and random forest,
197 respectively. The current lncLocator can predict five subcellular localizations of lncRNAs,
198 namely, cytoplasm, nucleus, cytosol, ribosome, and exosome.

199

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

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

205

206 The functions of lncRNA can be performed on diterpenoid biosynthetic genes/TFs in a *cis*
207 manner (Kopp & Mendell, 2018), and the lncRNAs and their target genes are considered
208 lncRNA-mRNA/TF pairs. Pearson correlation coefficient (PCC) of the expression of lncRNA
209 and its target gene pair was calculated using the psych (Revelle, 2022) package in R, requiring
210 co-expressed gene pairs with $|PCC| \geq 0.4$ and $p \leq 0.05$. The basis for predicting *cis*-regulation
211 was related to the positional relationship of lncRNA genes and coding genes on the genome (Liu
212 et al., 2019). It was determined to be *cis*-regulatory if the lncRNA gene was within 100 kilobases
213 (kb) upstream or downstream regions of the target genes used (Huang et al., 2018; Wang et al.,
214 2021). used. For genomic location analysis, the co-expressed gene pairs were aligned against two
215 *S. miltiorrhiza* genomes (Xu et al., 2016; Ma et al., 2021) by BLAST+ (BLAST-2.11.0+) with a
216 cutoff of E value < 1e-5. Finally, the candidate lncRNA-mRNA/TF pairs related to the
217 diterpenoid biosynthesis of *S. miltiorrhiza* were obtained.

218

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

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

224

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

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

237 Clique Centrality (MCC) to discover featured nodes. Subsequently, we constructed a network
238 module of these hub genes.
239

RNA extraction and qRT-PCR

241 Total RNA was extracted from plant tissues using the RNeasy plant kit (No.PH-01013-B,
242 Foregene, Chengdu, China). RNA degradation and contamination were monitored using 1%
243 RNase-free agarose gel electrophoresis, and the RNA purity was analyzed using a
244 NanoPhotometer^{TM}-N60 ultra-micro spectrophotometer. The reverse transcription reaction was
245 used RT Easy^{TM} II (With gDNase) kit (Version Number: 1.0-1904, Foregene, Chengdu, China)
246 following the instruction manual. qRT-PCR was performed in a 20 μL reaction volume
247 containing primers by Real-Time PCR EasyTM-SYBR Green I kit (Cat. No. QP-
248 01011/01012/01013/01014, Foregene, Chengdu, China) following the instruction manual. qRT-
249 PCR was carried out in triplicate reactions in the LineGene K Plus real-time PCR detection
250 system (Bioer, Hangzhou, China). Primers were designed using the Primer-BLAST (Ye et al.,
251 2012). Primer sequences were listed in Table S1. *β-actin* (Yang et al., 2010) gene was selected as
252 a reference since it showed stable expression in the *S. miltiorrhiza* tissues analyzed compared to
253 others. The reaction program was as follows: 3 min at 95°C, 10 s at 95°C, 30 s at 60°C, and 40
254 cycles. The temperature was then gradually increased to produce melting curves for
255 amplification specificity verification. The mean value of three replicates was normalized using *β-*
256 *actin* as the endogenous control. The 2(-delta-delta CT) method was used for calculating the
257 relative expression of genes.
258

Results

Identification of candidate lncRNAs

261 On the basis of the strict lncRNA identification pipeline, a total of 30,347 transcripts with no
262 annotation information were obtained. A total of 21,742 transcripts with length > 200 nt and
263 ORF < 100 aa were obtained. Subsequently, a total of 21,468 identified lncRNAs were obtained
264 from the intersection of PLEK and CPC2 software prediction results (Fig. 1, Table S2, Table
265 S3). Expression transcripts with FPKM ≤ 0.05 were filtered, and a total of 6,651 candidate
266 lncRNAs were selected for further analysis (Table S4).
267

268 The 21,468 identified lncRNAs ranged from 201 nt to 4,340 nt in length, the majority of which
269 (approximately 68.95%) were 200–400 nt. The mean length was 392 nt, which was lower than
270 the values observed in *S. miltiorrhiza* mRNAs (mean length = 1,515 nt). The GC content of
271 lncRNAs was mainly concentrated at 31.28%–37.28% (accounting for 46.32%), whereas
272 mRNAs were mainly concentrated at 39.28%–41.28% (accounting for 36.00%). The mean GC
273 content of lncRNAs was approximately 36.86%, which was slightly lower than that of mRNA
274 sequences (approximately 43.42%) (Fig. 2, Table S5).
275

Characterization and conservation of lncRNAs in *S. miltiorrhiza*

277 Conservation analysis showed that 21.02% and 15.40% of *S. miltiorrhiza* lncRNAs (6,651
278 candidate lncRNAs) were conserved compared with *Salvia splendens* and *Salvia hispanica* of the
279 same genus, respectively. However, in the same family of the different genera of *Mentha*
280 *longifolia*, *Scutellaria baicalensis*, *Pogostemon cablin,* and *Sesamum indicum*, 6.54%, 0.96%,
281 0.83%, and 0.51% of *S. miltiorrhiza* lncRNAs were conserved, respectively. Among plants of
282 different families, only 0.05%, 0.02%, and 0.02% of *S. miltiorrhiza* lncRNAs were conserved in
283 *Nicotiana tabacum*, *Arabidopsis thaliana*, and *Brassica napus*, respectively. There was no match
284 in *Selaginella moellendorffii*. The results of the conservation analysis are presented in Table S6.
285

286 **LncRNAs might be used as precursors or target mimics of *S.***
287 ***miltiorrhiza* miRNAs**
288 In this study, a total of 17 lncRNAs were identified as potential precursors of 41 *S. miltiorrhiza*
289 miRNAs (Table S7). To improve the prediction accuracy, the intersection of three kinds of
290 software prediction results was selected. A total of 14 lncRNAs were identified as potential
291 targets of 66 *S. miltiorrhiza* miRNAs from 18 families according to PmiREN2.0 (Table S7).
292

293 **Subcellular localization predictions indicated that most *S. miltiorrhiza***
294 **lncRNAs were found in the cytoplasm and nucleus**
295 Subcellular localization of lncRNA is closely related to its function. At present, many studies
296 have indicated that lncRNA contains specific RNA motifs with nuclear localization (Zhang et al.,
297 2014). Meanwhile, in the subcellular localization of lncLocator (Cao et al., 2018), a total of
298 3,381 cytoplasms, 3,316 nucleus, 83 exosomes, 51 cytosols, and 20 ribosomes were obtained
299 from 6,651 candidate lncRNAs. Among our candidate lncRNA-mRNA/TF pairs, 13 lncRNAs
300 were found in the nucleus, 9 lncRNAs were found in the cytoplasm, and 1 lncRNA was found in
301 the exosome (Table. S8).
302

303 **Genes and TFs involved in diterpenoid biosynthesis from**
304 **transcriptomic data**
305 To investigate the potential lncRNAs involved in diterpenoid biosynthesis, we predicted the
306 potential targets of lncRNAs in *cis*-regulatory relationships. In transcriptomic annotation, we
307 obtained 46 diterpenoid biosynthetic genes: GPPS (Van Schie et al., 2007), FPPS, GGPPS, CPS,
308 *ent*-CPS, KS, *ent*-KS, CYP76AH1 (Guo et al., 2013; Ma et al., 2016), CYP76AK2, CYP76AK3
309 (Li et al., 2021), CYP76AK5 (PU Xiangdong & WANG Lizhi, 2017), *ent*-kaurene oxidase (KO)
310 (Hedden & Thomas, 2012), *ent*-kaurenoic acid oxidase (KAO) (Helliwell et al., 2001), GA 2-
311 oxidase (GA2ox), GA 3-oxidase (GA3ox), GA 20-oxidase (GA20ox) (Hedden & Thomas, 2012;
312 Du et al., 2015), and 11 TFs: bHLH148, ERF6, GRAS1, MYB36 and WRKY2 (Li & Lu, 2014;
313 Zhang et al., 2015; Li et al., 2015, 2019; Ji et al., 2016), belonging to five TF families: bHLH,
314 ERF, GRAS, MYB, and WRKY. The list of the above genes is shown in Table S9.
315

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).

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 (Fig. S2B, Fig. 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*.

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

356 of most genes showed a downward trend at the 6 h time point, and only the *Smlnc0018769-*
357 *Sm0037093* pair and *Smlnc0008662-Sm0063385* pair had an upward trend. The *Smlnc0008477-*
358 *Sm0056000* pair and *Smlnc0012647-Smlnc0032870* pair had the same expression trend. The
359 *Smlnc0042160-Sm0028870* pair had the same expression trends at 6, 12, and 24 h. The
360 *Smlnc0018769-Sm0037093*, *Smlnc0008662-Sm0063385*, and *Smlnc0052170-Sm0067296* pairs
361 had opposite expression trends at 48 h, whereas the *Smlnc0019429-Sm0026208* pair had opposite
362 trends at 24 and 48 h. Interestingly, the two gene pairs in the co-expression network,
363 *Smlnc0032870-Sm0012648* and *Smlnc0018769-Sm0037093*, which were positively correlated,
364 showed complex expression correlation under the induction of MeJA, even showed opposite
365 expression trends after 24 h. The hub genes: *Sm0056000*, *Sm0037097*, *Sm0063385*, and
366 *Sm0067296* in the co-expression network showed significantly differential expression in MeJA-
367 induced *S. miltiorrhiza*.
368
369 The data in Fig. 5A showed that *Smlnc0063419*, *Sm0026208*, and *Sm0067296* had similar
370 expression patterns, which formed the lncRNA-mRNA-TF module. *Sm0026208* and *Sm0067296*
371 were annotated with TF *WRKY2* and mRNA *GA3ox2*, respectively. *Smlnc0012647*,
372 *Smlnc0032870*, and *Sm0009433* had similar expression patterns, as shown in Fig. 5B.
373 *Sm0009433* was annotated with TF *MYB36*. *Smlnc0042160* and *Sm0037093* had similar
374 expression patterns, as shown in Fig. 5C. *Sm0037093* was annotated with mRNA *KS2*. Through
375 the plug-in cytoHubba, we calculated the top 10 hub genes (Table S12 and Fig. 5D). The
376 modules of *Smlnc0063419-Sm0026208-Sm0067296* and *Smlnc0012647-Smlnc0032870-*
377 *Sm0009433*, in which all genes were present in the top 10 hub genes.
378

Discussion
380 In this study, the transcriptomic data of *S. miltiorrhiza* during the tanshinone accumulation stage
381 were used for exploring the lncRNAs and their target transcripts involved in the diterpenoid
382 biosynthetic pathway. A total of 6,651 lncRNAs were obtained by a strict bioinformatic pipeline.
383 We found some common features of lncRNAs, which may be related to their function. Among
384 the candidate lncRNAs, diterpenoid biosynthetic genes/TFs in *S. miltiorrhiza*, we detected 23
385 candidate lncRNA-mRNA/TF pairs with a *cis*-regulatory relationship. To further verify their
386 relationship, we used MeJA as an inducer to observe the gene expression of these candidate
387 lncRNA-mRNA/TF pairs with MeJA treatment for 6, 12, 24, and 48 h. Through the expression
388 data of time series, which provided an exploratory method for the role of lncRNA, three
389 lncRNA-mRNA and/or TF network modules were finally obtained.
390
391 In the present study, the identified lncRNAs of *S. miltiorrhiza* were found to be shorter in length
392 compared with protein-coding transcripts, which was consistent with the previous reports (Hao et
393 al., 2015; Liu et al., 2018; Shen et al., 2018). The mean GC content of lncRNAs was slightly
394 lower than that of mRNAs, which has also been reported in *Populus tomentosa* (Zhou et al.,
395 2017). Conservation analysis showed that plants of the same genus are more conservative than

396   those of different families. A previous study also showed that the majority of lncRNAs have high
397   sequence conservation at the intra-species and sub-species levels (Deng et al., 2018). Polymerase
398   C-terminal domain (CTD) modifications [e.g., CTD modification threonine 4 phosphorylation
399   (CTD-T4P)] are on the promoters of lncRNAs, which leads to the decrease in the polymerase
400   pause and the advance of the termination in the whole lncRNA genome. The transcription rate of
401   lncRNAs is very fast, which means that they can quickly act on the regulatory target and respond
402   to the signal. Therefore, transcription accuracy and sequence conservation were low (Rinn &
403   Chang, 2020). The lncRNA function was maintained across large evolutionary distances even
404   when the lncRNA sequence substantially diverged (Ulitsky, 2016). Another possible explanation
405   for this is that RNA secondary structures may be the units of lncRNA words rather than the
406   primary sequence, and disparate sequences form similar structure–function relationships to
407   transmit symbolic language like hieroglyphics, thereby forming the molecular grammar of
408   lncRNAs (Rinn & Chang, 2020).
409
410   LncRNAs have been proposed to carry out their functions by *cis* or *trans*, transcriptional
411   regulation *cis*-acting lncRNAs influence the expression and/or chromatin state of nearby genes.
412   We predicted lncRNA-mRNA pairs in the *cis*-acting relationship. The results of our study
413   showed that 45,198 and 180 lncRNA-mRNA/TF pairs were co-expressed ($|PCC| \geq 0.4$) or co-
414   localized, respectively, and 23 lncRNA-mRNA/TF pairs were both co-localized and co-
415   expressed (Table S10). The results indicated that most lncRNAs were not co-expressed with their
416   nearby coding genes and were transcribed independently in *S. miltiorrhiza* (Liao et al., 2011).
417   Two pairs were both co-expressed and co-localized in the high correlation co-expression network
418   with $|PCC| \geq 0.8$. Thus, the modes of action of lncRNAs were not limited to local regulators.
419   Another possible explanation is that the strength of the required gene co-expression may depend
420   on the stability or toxicity of the metabolites, and strong co-expression should only be required
421   for unstable monomers (Obayashi & Kinoshita, 2009).
422
423   Although advances have been made in the miRNA and miRNA target prediction fields, the
424   precision of miRNA target prediction needs to be improved (Akgül et al., 2022). To reduce false
425   positives, we used three kinds of prediction software to predict miRNA targets. Although many
426   miRNA databases and prediction software are published for plants, few of them are available (de
427   Amorim, Pedro & Paschoal, 2022). This limitation reduced our chance to find miRNAs
428   associated with the lncRNA-mRNA/TF pairs in the diterpenoid biosynthetic pathway. On the
429   basis of the relationship between miRNAs and lncRNAs, we predicted that 14 lncRNAs were
430   potential targets or TMs of 66 miRNAs in *S. miltiorrhiza*, and no one existed in the candidate
431   lncRNA-mRNA/TF pairs. However, all the results were predicted preliminarily based on
432   bioinformatic analyses and need to be further validated.
433
434   Building gene regulatory networks from transcriptomic studies often results in a static view of
435   gene expression, which can make it difficult to disentangle the regulatory pathway structure

436   response to a stimulus. Time-series expression analysis may uncover the temporal transcriptional
437   logic for plant response systems, and provide more accurate predictions for targeted breeding
438   (Greenham & McClung, 2018). By studying the time-series expression of some candidate
439   regulatory pairs inducted by MeJA, we observed a more detailed landscape. We found that some
440   lncRNAs were downregulated in the early stage. The expression of the corresponding mRNAs
441   was upregulated in the later stage between the regulatory pairs, and plants' response to the signal
442   had a time delay. For example, in response to vernalization, *COOLAIR* is transiently induced by
443   prolonged cold, reaching a maximum expression level after 2 weeks (Swiezewski et al., 2009).
444   Meanwhile, we found that some genes showed different expression patterns within a period,
445   namely, both upregulated and downregulated, this phenomenon may be because gene regulatory
446   networks are inherently complex, with multiple feedback and feedforward loops (Wils &
447   Kaufmann, 2017).
448
449   Nineteen genes were differentially expressed significantly under MeJA-induced, and some of
450   them were present in hub genes of the co-expression network. These results indicated that these
451   genes might participate in the biosynthesis of secondary terpenoids such as tanshinone and also
452   play an important role in the defense response of *S. miltiorrhiza*. Our results indicated that the
453   expression levels of *Sm0056000* (*CPS*) and *Sm0037093* (*KS*) were upregulated under MeJA
454   induction, which was also observed in the previous report (Luo et al., 2014). However, the levels
455   of these two genes *Sm0056000* (*CPS*) and *Sm0037093* (*KS*) started to increase at 12 h and 24 h in
456   our study, respectively. The response of TFs to MeJA was also observed in this study (Luo et al.,
457   2014), which was consistent with the TF *WRKY* we studied. For MeJA treatment, the expression
458   of *CYP76AH1* was up-regulated over time and reached a peak at 12 h in the previous study (Li et
459   al., 2021), however, after peaking at 6h, our *Sm0063385* (*CYP76AH1*) expression was down-
460   regulated, and then it began to up-regulate at 12h and peaked at 48 h. Previous research has
461   suggested that *WRKYs* might regulate the development of bast fiber in response to GA$_3$ stress in
462   jute (*Corchorus capsularis*) (Zhang et al., 2020), the *Smlnc0063419-Sm0026208-Sm0067296*
463   module (*Sm0026208*: TF *WRKY2*, *Sm0067296*: *GA3ox2*) in our study may form a similar
464   response module in *S. miltiorrhiza*.
465
466   The tissue specificity and subcellular localization of lncRNAs may suggest their function. We
467   obtained 3 lncRNA-mRNA and/or TF modules: *Smlnc0063419-Sm0026208* (TF)-*Sm0067296*,
468   *Smlnc0012647-Smlnc0032870-Sm0009433* (TF), and *Smlnc0042160-Sm0037093*, among of
469   lncRNAs: *Smlnc0063419*, *Smlnc0012647*, *Smlnc0032870*, and *Smlnc0042160* mainly expressed
470   in roots (Fig. 3 and Fig. 5), which was consistent with the place where tanshinones accumulated
471   of *S. miltiorrhiza* (Chang et al., 2019). In the prediction of subcellular localization,
472   *Smlnc0012647*, *Smlnc0042160*, and *Smlnc0063419* were predicted to be located in the
473   cytoplasm, this suggests that these predicted cytoplasmic lncRNAs may interfere with protein
474   post-translational modifications or regulate mRNA export (Chen, 2016). Although
475   *Smlnc0012647* and *Smlnc0032870* existed in the same module, *Smlnc0032870* was predicted to

476 be located in the nucleus, a possible explanation for this result may be that some lncRNAs are
477 located in both the nucleus and cytoplasm (Cabili et al., 2015).
478

Conclusions

480 This study set out to investigate the possible functions of the lncRNAs of *S. miltiorrhiza* related
481 to diterpenoid biosynthesis, we predicted the potential targets of lncRNAs in *cis*-regulatory
482 relationships, a summarizing figure is shown in Fig. S4. Through a strict bioinformatic pipeline,
483 we identified 6,651 candidate lncRNAs, and obtained three lncRNA-mRNA and/or TF network
484 modules. This study revealed the possible roles of the lncRNAs of *S. miltiorrhiza* related to
485 diterpenoid biosynthesis. These findings indicated that lncRNA is generally complex in
486 regulating mRNA and/or TF. This study provides useful information to deepen our
487 understanding of the function and regulatory mechanisms of lncRNAs in the diterpenoid
488 biosynthetic pathway of *S. miltiorrhiza*.
489

References

491 Akgül B, Stadler PF, Hawkins LJ, Hadj-Moussa H, Storey KB, Ergin K, Çetinkaya R, Paschoal
492   AR, Nachtigall PG, Tutar Y, Yousef M, Allmer J. 2022. 44 Current Challenges in
493   miRNomics. In: 423–438. DOI: 10.1007/978-1-0716-1170-8_19.
494 de Amorim TS, Pedro DLF, Paschoal AR. 2022. MicroRNA Databases and Tools. In: 131–166.
495   DOI: 10.1007/978-1-0716-1170-8_7.
496 Bonnet E, He Y, Billiau K, van de Peer Y. 2010. TAPIR, a web server for the prediction of plant
497   microRNA targets, including target mimics. *Bioinformatics* 26:1566–1568. DOI:
498   10.1093/bioinformatics/btq233.
499 Brockdorff N. 1998. The role of Xist in X-inactivation. *Current Opinion in Genetics &*
500   *Development* 8:328–333. DOI: 10.1016/S0959-437X(98)80090-7.
501 Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL,
502   Raj A. 2015. Localization and abundance analysis of human lncRNAs at single-cell and
503   single-molecule resolution. *Genome Biology* 16:20. DOI: 10.1186/s13059-015-0586-4.
504 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
505   BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. DOI: 10.1186/1471-
506   2105-10-421.
507 Cao Z, Pan X, Yang Y, Huang Y, Shen H-B. 2018. The lncLocator: a subcellular localization
508   predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics*
509   34:2185–2194. DOI: 10.1093/bioinformatics/bty085.
510 Carlevaro-Fita J, Johnson R. 2019. Global Positioning System: Understanding Long Noncoding
511   RNAs through Subcellular Localization. *Molecular Cell* 73:869–883. DOI:
512   10.1016/j.molcel.2019.02.008.
513 Chang Y, Wang M, Li J, Lu S. 2019. Transcriptomic analysis reveals potential genes involved in
514   tanshinone biosynthesis in *Salvia miltiorrhiza*. *Scientific Reports* 9:14929. DOI:
515   10.1038/s41598-019-51535-9.

516 Chen L-L. 2016. Linking Long Noncoding RNA Localization and Function. *Trends in*
517 *Biochemical Sciences* 41:761–772. DOI: 10.1016/j.tibs.2016.07.003.

518 Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. 2014. cytoHubba: identifying hub
519 objects and sub-networks from complex interactome. *BMC Systems Biology* 8:S11. DOI:
520 10.1186/1752-0509-8-S4-S11.

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

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

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

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

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

535 Gao W, Hillwig ML, Huang L, Cui G, Wang X, Kong J, Yang B, Peters RJ. 2009. A Functional
536 Genomics Approach to Tanshinone Biosynthesis Provides Stereochemical Insights. *Organic*
537 *Letters* 11:5170–5173. DOI: 10.1021/ol902051v.

538 Greenham K, McClung CR. 2018. Time to build on good design: Resolving the temporal
539 dynamics of gene regulatory networks. *Proceedings of the National Academy of Sciences*
540 115:6325–6327. DOI: 10.1073/pnas.1807707115.

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

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

545 Gundlach H, Müller MJ, Kutchan TM, Zenk MH. 1992. Jasmonic acid is a signal transducer in
546 elicitor-induced plant cell cultures. *Proceedings of the National Academy of Sciences*
547 89:2389–2393. DOI: 10.1073/pnas.89.6.2389.

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

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

555   Hao Z, Fan C, Cheng T, Su Y, Wei Q, Li G. 2015. Genome-Wide Identification,
556   Characterization and Evolutionary Analysis of Long Intergenic Noncoding RNAs in
557   *Cucumber*. *PLOS ONE* 10:e0121800. DOI: 10.1371/journal.pone.0121800.
558   Hedden P, Thomas SG. 2012. Gibberellin biosynthesis and its regulation. *Biochemical Journal*
559   444:11–25. DOI: 10.1042/BJ20120245.
560   Helliwell CA, Chandler PM, Poole A, Dennis ES, Peacock WJ. 2001. The CYP88A cytochrome
561   P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis
562   pathway. *Proceedings of the National Academy of Sciences* 98:2065–2070. DOI:
563   10.1073/pnas.98.4.2065.
564   Huang L, Dong H, Zhou D, Li M, Liu Y, Zhang F, Feng Y, Yu D, Lin S, Cao J. 2018.
565   Systematic identification of long non-coding RNAs during pollen development and
566   fertilization in *Brassica rapa*. *The Plant Journal* 96:203–222. DOI: 10.1111/tpj.14016.
567   Ji AJ, Luo HM, Xu ZC, Zhang X, Zhu YJ, Liao BS, Yao H, Song JY, Chen SL. 2016.
568   Genome-Wide Identification of the AP2/ERF Gene Family Involved in Active Constituent
569   Biosynthesis in *Salvia miltiorrhiza*. *The Plant Genome* 9. DOI:
570   10.3835/plantgenome2015.08.0077.
571   Jiang M, Chen H, Liu J, Du Q, Lu S, Liu C. 2021. Genome-wide identification and functional
572   characterization of natural antisense transcripts in *Salvia miltiorrhiza*. *Scientific Reports*
573   11:4769. DOI: 10.1038/s41598-021-83520-6.
574   Kai G, Liao P, Zhang T, Zhou W, Wang J, Xu H, Liu Y, Zhang L. 2010. Characterization,
575   expression profiling, and functional identification of a gene encoding geranylgeranyl
576   diphosphate synthase from *Salvia miltiorrhiza*. *Biotechnology and Bioprocess Engineering*
577   15:236–245. DOI: 10.1007/s12257-009-0123-y.
578   Kai G, Xu H, Zhou C, Liao P, Xiao J, Luo X, You L, Zhang L. 2011. Metabolic engineering
579   tanshinone biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *Metabolic*
580   *Engineering* 13:319–327. DOI: 10.1016/j.ymben.2011.02.003.
581   Kang Y-J, Yang D-C, Kong L, Hou M, Meng Y-Q, Wei L, Gao G. 2017. CPC2: a fast and
582   accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids*
583   *Research* 45:W12–W16. DOI: 10.1093/nar/gkx428.
584   Kopp F, Mendell JT. 2018. Functional Classification and Experimental Dissection of Long
585   Noncoding RNAs. *Cell* 172:393–407. DOI: 10.1016/j.cell.2018.01.011.
586   Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to
587   function. *Nucleic Acids Research* 47:D155–D162. DOI: 10.1093/nar/gky1141.
588   Li W, Bai Z, Pei T, Yang D, Mao R, Zhang B, Liu C, Liang Z. 2019. SmGRAS1 and SmGRAS2
589   Regulate the Biosynthesis of Tanshinones and Phenolic Acids in *Salvia miltiorrhiza*.
590   *Frontiers in Plant Science* 10. DOI: 10.3389/fpls.2019.01367.
591   Li B, Li J, Chai Y, Huang Y, Li L, Wang D, Wang Z. 2021. Targeted mutagenesis of
592   CYP76AK2 and CYP76AK3 in *Salvia miltiorrhiza* reveals their roles in tanshinones
593   biosynthetic pathway. *International Journal of Biological Macromolecules* 189:455–463.
594   DOI: 10.1016/j.ijbiomac.2021.08.112.

595   Li C, Li D, Shao F, Lu S. 2015. Molecular cloning and expression analysis of *WRKY*
596   transcription factor genes in *Salvia miltiorrhiza*. *BMC Genomics* 16:200. DOI:
597   10.1186/s12864-015-1411-x.
598   Li C, Lu S. 2014. Genome-wide characterization and comparative analysis of R2R3-MYB
599   transcription factors shows the complexity of MYB-associated regulatory networks in
600   *Salvia miltiorrhiza*. *BMC Genomics* 15:277. DOI: 10.1186/1471-2164-15-277.
601   Li D, Shao F, Lu S. 2015. Identification and characterization of mRNA-like noncoding RNAs in
602   *Salvia miltiorrhiza*. *Planta* 241:1131–1143. DOI: 10.1007/s00425-015-2246-z.
603   Li S, Yamada M, Han X, Ohler U, Benfey PN. 2016. High-Resolution Expression Map of the
604   *Arabidopsis* Root Reveals Alternative Splicing and lincRNA Regulation. *Developmental*
605   *Cell* 39:508–522. DOI: 10.1016/j.devcel.2016.10.012.
606   Li A, Zhang J, Zhou Z. 2014. PLEK: a tool for predicting long non-coding RNAs and messenger
607   RNAs based on an improved *k*-mer scheme. *BMC Bioinformatics* 15:311. DOI:
608   10.1186/1471-2105-15-311.
609   Liang Z-S, Yang D-F, Liang X, Zhang Y-J, Liu Y, Liu F-H. 2012. Roles of reactive oxygen
610   species in methyl jasmonate and nitric oxide-induced tanshinone production in *Salvia*
611   *miltiorrhiza* hairy roots. *Plant Cell Reports* 31:873–883. DOI: 10.1007/s00299-011-1208-6.
612   Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, Zhao G, Luo H, Bu D, Zhao H, Skogerbø G,
613   Wu Z, Zhao Y. 2011. Large-scale prediction of long non-coding RNA functions in a
614   coding–non-coding gene co-expression network. *Nucleic Acids Research* 39:3864–3878.
615   DOI: 10.1093/nar/gkq1348.
616   Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua N-H. 2012. Genome-
617   Wide Analysis Uncovers Regulation of Long Intergenic Noncoding RNAs in *Arabidopsis*.
618   *The Plant Cell* 24:4333–4345. DOI: 10.1105/tpc.112.102855.
619   Liu H, Lu Y, Wang J, Hu J, Wuyun T. 2018. Genome-wide screening of long non-coding RNAs
620   involved in rubber biosynthesis in *Eucommia ulmoides*. *Journal of Integrative Plant*
621   *Biology* 60:1070–1082. DOI: 10.1111/jipb.12693.
622   Liu H, Wang R, Mao B, Zhao B, Wang J. 2019. Identification of lncRNAs involved in rice ovule
623   development and female gametophyte abortion by genome-wide screening and functional
624   analysis. *BMC Genomics* 20:90. DOI: 10.1186/s12864-019-5442-6.
625   Lu S. 2021. Biosynthesis and Regulatory Mechanisms of Bioactive Compounds in *Salvia*
626   *miltiorrhiza* , a Model System for Medicinal Plant Biology. *Critical Reviews in Plant*
627   *Sciences* 40:243–283. DOI: 10.1080/07352689.2021.1935719.
628   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,
629   Liu J, Chen S. 2014. Transcriptional data mining of *Salvia miltiorrhiza* in response to
630   methyl jasmonate to examine the mechanism of bioactive compound biosynthesis and
631   regulation. *Physiologia Plantarum* 152:241–255. DOI: 10.1111/ppl.12193.
632   Ma Y, Cui G, Chen T, Ma X, Wang R, Jin B, Yang J, Kang L, Tang J, Lai C, Wang Y, Zhao Y,
633   Shen Y, Zeng W, Peters RJ, Qi X, Guo J, Huang L. 2021. Expansion within the CYP71D

634   subfamily drives the heterocyclization of tanshinones synthesis in *Salvia miltiorrhiza*.
635   *Nature Communications* 12:685. DOI: 10.1038/s41467-021-20959-1.

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

640   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,
641   Guo J, Huang L-Q. 2015. The Biosynthetic Pathways of Tanshinones and Phenolic Acids in
642   *Salvia miltiorrhiza*. *Molecules* 20:16235–16254. DOI: 10.3390/molecules200916235.

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

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

648   Nagano T, Fraser P. 2011. No-Nonsense Functions for Long Noncoding RNAs. *Cell* 145:178–
649   181. DOI: 10.1016/j.cell.2011.03.014.

650   Obayashi T, Kinoshita K. 2009. Rank of Correlation Coefficient as a Comparable Measure for
651   Biological Significance of Gene Coexpression. *DNA Research* 16:249–260. DOI:
652   10.1093/dnares/dsp016.

653   Palazzo AF, Koonin E V. 2020. Functional Long Non-coding RNAs Evolve from Junk
654   Transcripts. *Cell* 183:1151–1161. DOI: 10.1016/j.cell.2020.09.047.

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

658   Ponting CP, Oliver PL, Reik W. 2009. Evolution and Functions of Long Noncoding RNAs. *Cell*
659   136:629–641. DOI: 10.1016/j.cell.2009.02.006.

660   PU Xiangdong, WANG Lizhi. 2017. Cloning and Bioinformatic Analysisof CYP76AK5 Related
661   to Tanshinone Biosynt hesis in *Salvia miltiorrhiza*. *Modern Chinese Medicine* 19:1106–
662   1110. DOI: 10.13313/j.issn.1673-4890.2017.8.011.

663   Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G,
664   Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD. 2012.
665   The Pfam protein families database. *Nucleic Acids Research* 40:D290–D301. DOI:
666   10.1093/nar/gkr1065.

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

671   Rai MI, Alam M, Lightfoot DA, Gurha P, Afzal AJ. 2019. Classification and experimental
672   identification of plant long non-coding RNAs. *Genomics* 111:997–1005. DOI:
673   10.1016/j.ygeno.2018.04.014.

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

676 Revelle W. 2022. psych: Procedures for Psychological, Psychometric, and Personality Research.

677 Rinn JL, Chang HY. 2020. Long Noncoding RNAs: Molecular Modalities to Organismal
678   Functions. *Annual Review of Biochemistry* 89:283–308. DOI: 10.1146/annurev-biochem-
679   062917-012708.

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

683 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B,
684   Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular
685   Interaction Networks. *Genome Research* 13:2498–2504. DOI: 10.1101/gr.1239303.

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

689 Shimane M, Ueno Y, Morisaki K, Oogami S, Natsume M, Hayashi K, Nozaki H, Kawaide H.
690   2014. Molecular evolution of the substrate specificity of *ent*-kaurene synthases to adapt to
691   gibberellin biosynthesis in land plants. *Biochemical Journal* 462:539–546. DOI:
692   10.1042/BJ20140134.

693 Swiezewski S, Liu F, Magusin A, Dean C. 2009. Cold-induced silencing by long antisense
694   transcripts of an Arabidopsis Polycomb target. *Nature* 462:799–802. DOI:
695   10.1038/nature08618.

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

698 Wang H-L v., Chekanova JA. 2017. Long Noncoding RNAs in Plants. In: 133–154. DOI:
699   10.1007/978-981-10-5203-3_5.

700 Wang Y, Zhang H, Li Q, Jin J, Chen H, Zou Y, Huang X, Ding Y. 2021. Genome-Wide
701   Identification of lncRNAs Involved in Fertility Transition in the Photo-Thermosensitive
702   Genic Male Sterile Rice Line Wuxiang S. *Frontiers in Plant Science* 11. DOI:
703   10.3389/fpls.2020.580050.

704 Wasternack C. 2007. Jasmonates: An Update on Biosynthesis, Signal Transduction and Action in
705   Plant Stress Response, Growth and Development. *Annals of Botany* 100:681–697. DOI:
706   10.1093/aob/mcm079.

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

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

714 Wu H-J, Ma Y-K, Chen T, Wang M, Wang X-J. 2012b. PsRobot: a web-based plant small RNA
715   meta-analysis toolbox. *Nucleic Acids Research* 40:W22–W28. DOI: 10.1093/nar/gks554.
716 Xiao Y, Zhang L, Gao S, Saechao S, Di P, Chen J, Chen W. 2011. The c4h, tat, hppr and hppd
717   Genes Prompted Engineering of Rosmarinic Acid Biosynthetic Pathway in *Salvia*
718   *miltiorrhiza* Hairy Root Cultures. *PLoS ONE* 6:e29713. DOI:
719   10.1371/journal.pone.0029713.
720 Xu Z, Peters RJ, Weirather J, Luo H, Liao B, Zhang X, Zhu Y, Ji A, Zhang B, Hu S, Au KF,
721   Song J, Chen S. 2015. Full-length transcriptome sequences and splice variants obtained by a
722   combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza*
723   and tanshinone biosynthesis. *The Plant Journal* 82:951–961. DOI: 10.1111/tpj.12865.
724 Xu H, Song J, Luo H, Zhang Y, Li Q, Zhu Y, Xu J, Li Y, Song C, Wang B, Sun W, Shen G,
725   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,
726   Liu J, Wang Y, Hayward A, Nelson D, Ning Z, Peters RJ, Qi X, Chen S. 2016. Analysis of
727   the Genome Sequence of the Medicinal Plant *Salvia miltiorrhiza*. *Molecular Plant* 9:949–
728   952. DOI: 10.1016/j.molp.2016.03.010.
729 Yamaguchi S. 2008. Gibberellin Metabolism and its Regulation. *Annual Review of Plant Biology*
730   59:225–251. DOI: 10.1146/annurev.arplant.59.032607.092804.
731 Yang Y, Hou S, Cui G, Chen S, Wei J, Huang L. 2010. Characterization of reference genes for
732   quantitative real-time PCR analysis in various tissues of *Salvia miltiorrhiza*. *Molecular*
733   *Biology Reports* 37:507–513. DOI: 10.1007/s11033-009-9703-3.
734 Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: A
735   tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*
736   13:134. DOI: 10.1186/1471-2105-13-134.
737 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
738   of microRNAs and long non-coding RNAs involved in fatty acid biosynthesis in tree peony
739   seeds. *Gene* 666:72–82. DOI: 10.1016/j.gene.2018.05.011.
740 Yu T, Tzeng DTW, Li R, Chen J, Zhong S, Fu D, Zhu B, Luo Y, Zhu H. 2019. Genome-wide
741   identification of long non-coding RNA targets of the tomato MADS box transcription factor
742   RIN and function analysis. *Annals of Botany* 123:469–482. DOI: 10.1093/aob/mcy178.
743 Zhang B, Gunawardane L, Niazi F, Jahanbani F, Chen X, Valadkhan S. 2014. A Novel RNA
744   Motif Mediates the Strict Nuclear Localization of a Long Noncoding RNA. *Molecular and*
745   *Cellular Biology* 34:2318–2329. DOI: 10.1128/MCB.01673-13.
746 Zhang X, Luo H, Xu Z, Zhu Y, Ji A, Song J, Chen S. 2015. Genome-wide characterisation and
747   analysis of bHLH transcription factors related to tanshinone biosynthesis in *Salvia*
748   *miltiorrhiza*. *Scientific Reports* 5:11244. DOI: 10.1038/srep11244.
749 Zhang L, Wan X, Xu Y, Niyitanga S, Qi J, Zhang L. 2020. De novo assembly of transcriptome
750   and genome-wide identification reveal GA3 stress-responsive WRKY transcription factors
751   involved in fiber formation in jute (*Corchorus capsularis*). *BMC Plant Biology* 20:403.
752   DOI: 10.1186/s12870-020-02617-8.

753   Zhou D, Du Q, Chen J, Wang Q, Zhang D. 2017. Identification and allelic dissection uncover
754   roles of lncRNAs in secondary growth of Populus tomentosa. *DNA Research* 24:473–486.
755   DOI: 10.1093/dnares/dsx018.
756   Zhou J, Liu R, Shuai M, Yan Z-Y, Chen X. 2021. Comparative transcriptome analyses of
757   different *Salvia miltiorrhiza* varieties during the accumulation of tanshinones. *PeerJ*
758   9:e12300. DOI: 10.7717/peerj.12300.
759

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.

(A)

(B)
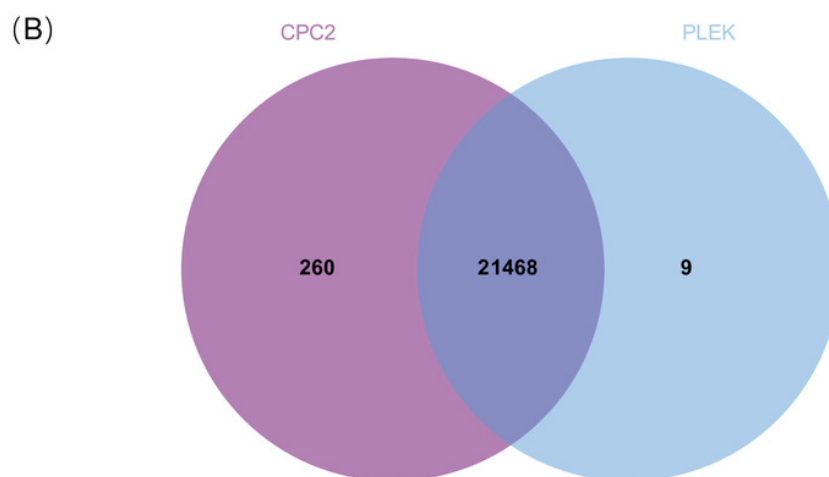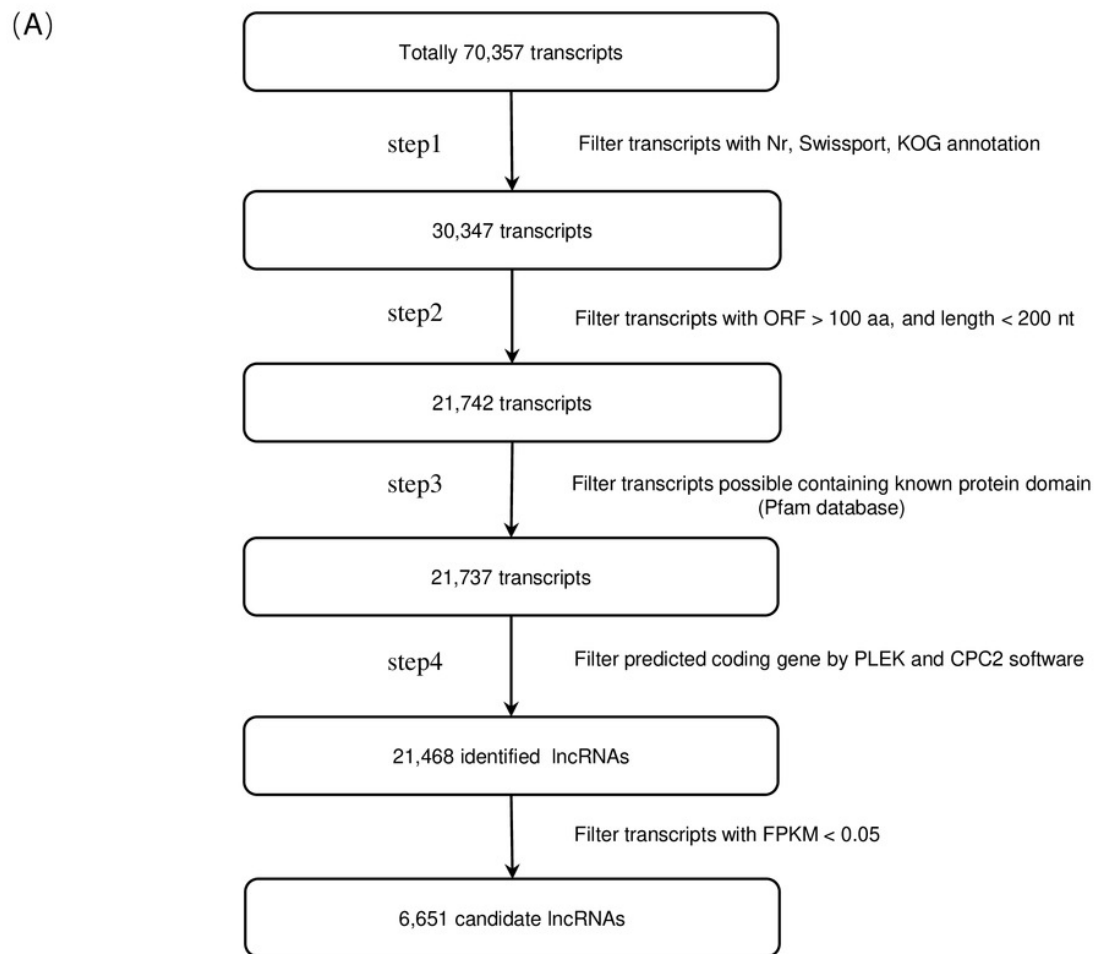
Figure 2

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*.
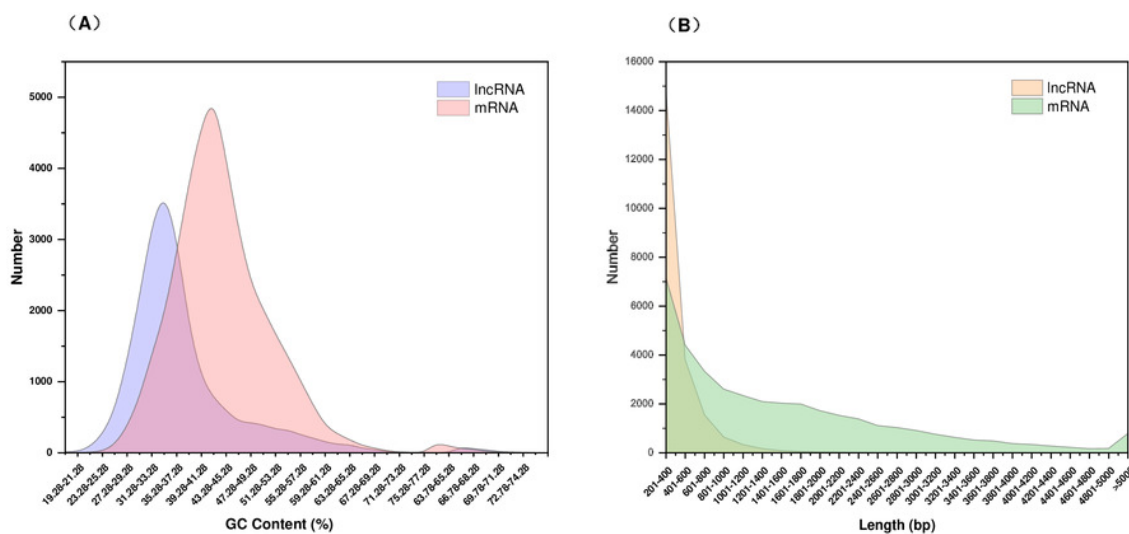
Figure 3

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.
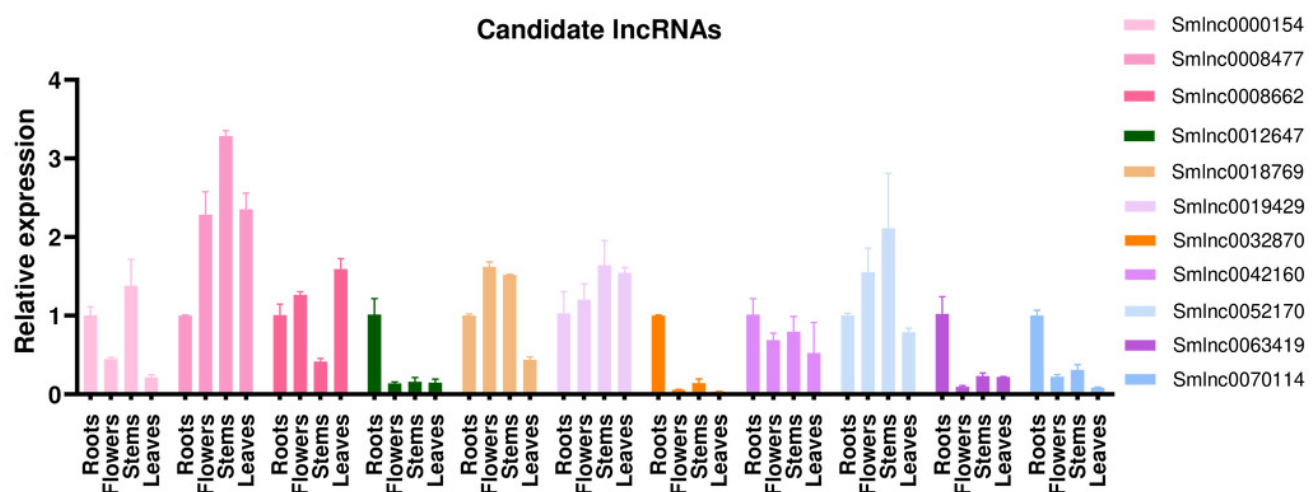
Figure 4

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 5 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.
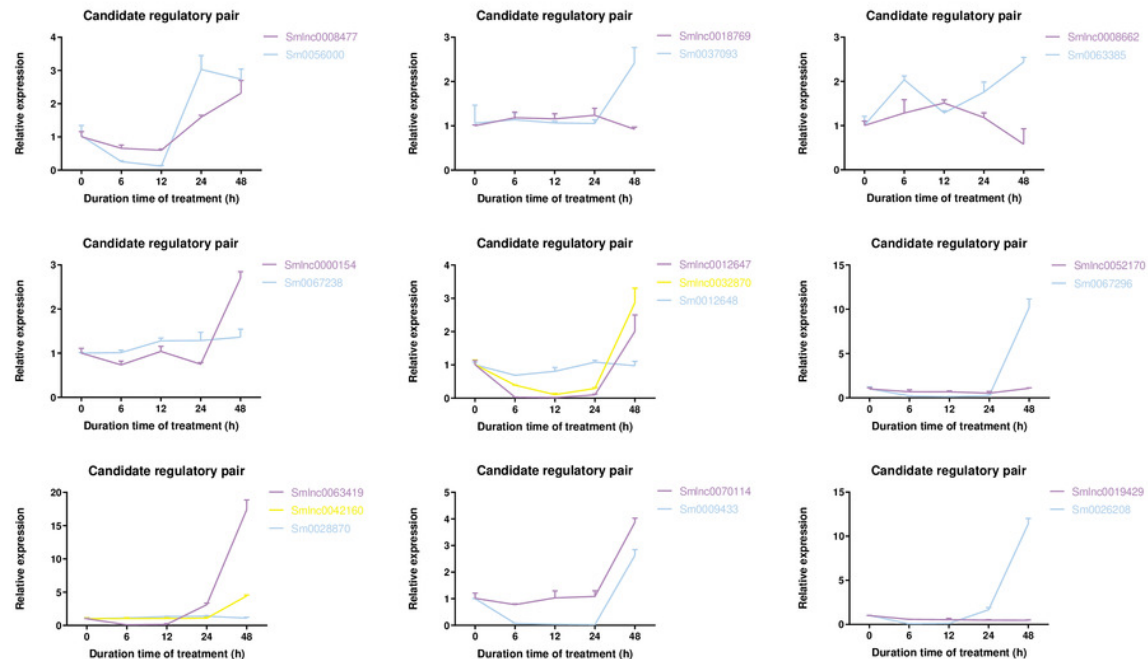
Figure 5

LncRNA-mRNA-TF module, lncRNA-TF module, lncRNA-mRNA module and hub genes in
S. miltiorrhiza.

(A) (B) (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. The Rank is closer
to 1, the higher the rank.