

SCIENTIFIC REPORTS



OPEN

Comprehensive transcriptome profiling of *Salvia miltiorrhiza* for discovery of genes associated with the biosynthesis of tanshinones and phenolic acids

Wei Zhou^{1,3}, Qiang Huang², Xiao Wu², Zewen Zhou², Mingquan Ding³, Min Shi², Fenfen Huang², Shen Li³, Yao Wang² & Guoyin Kai^{1,2}

Tanshinones and phenolic acids are crucial bioactive compounds biosynthesized in *Salvia miltiorrhiza*. Methyl jasmonate (MeJA) is an effective elicitor to enhance the production of phenolic acids and tanshinones simultaneously, while yeast extract (YE) is used as a biotic elicitor that only induce tanshinones accumulation. However, little was known about the different molecular mechanism. To identify the downstream and regulatory genes involved in tanshinone and phenolic acid biosynthesis, we conducted comprehensive transcriptome profiling of *S. miltiorrhiza* hairy roots treated with either MeJA or YE. Total 55588 unigenes were assembled from about 1.72 billion clean reads, of which 42458 unigenes (76.4%) were successfully annotated. The expression patterns of 19 selected genes in the significantly upregulated unigenes were verified by quantitative real-time PCR. The candidate downstream genes and other cytochrome P450s involved in the late steps of tanshinone and phenolic acid biosynthesis pathways were screened from the RNA-seq dataset based on co-expression pattern analysis with specific biosynthetic genes. Additionally, 375 transcription factors were identified to exhibit a significant up-regulated expression pattern in response to induction. This study can provide us a valuable gene resource for elucidating the molecular mechanism of tanshinones and phenolic acids biosynthesis in hairy roots of *S. miltiorrhiza*.

Salvia miltiorrhiza Bunge (Dan shen in Chinese) is a traditional Chinese herb with significant medicinal and economic values. It has been used widely to treat many cardiovascular diseases such as menstrual disorder, blood circulation disturbance, inflammation and angina pectoris¹. The active ingredients of *S. miltiorrhiza* can be divided into two groups: one group is water-soluble phenolic acids, such as rosmarinic acids (RAs), salvianolic acids and lithospermic acid, functions as antibacterial, anti-oxidative and antiviral reagents²; the other group is diterpenoid tanshinone, including tanshinone I, tanshinone IIA, tanshinone IIB, dihydrotanshinone I and cryptotanshinone, exhibits various pharmacological activities including antioxidant, antitumor and anti-inflammatory properties. The active constituents obtained from cultivated *S. miltiorrhiza* is low and cannot meet the rapidly increasing market need^{3,4}. Genetic manipulation of active ingredients biosynthesis pathway in plants or hairy roots provide us a promising strategy⁵, which rely on the detail understanding of the biosynthesis pathway and regulation mechanism.

The unique chemical constituents of *S. miltiorrhiza*, diterpenoids tanshinones are derived predominantly from the plastidic methylerythritol phosphate (MEP) pathway or partly through the cytoplasmic mevalonate (MVA), with possible cross-talk between the common precursors of isopentenyl diphosphate (IPP) and dimethylallyl

¹College of pharmacy, Zhejiang Chinese Medical University, Hangzhou, Zhejiang, 310053, China. ²Laboratory of Plant Biotechnology, College of Life and Environment Sciences, Shanghai Normal University, Shanghai, 200234, China. ³The Key Laboratory for Quality Improvement of Agricultural Products of Zhejiang Province, School of Agriculture and Food Science, Zhejiang A&F University, Linan, Hangzhou, Zhejiang, 311300, China. Wei Zhou and Qiang Huang contributed equally to this work. Correspondence and requests for materials should be addressed to G.K. (email: guoyinkai@yahoo.com)

	Control	MeJA-1h	MeJA-6h	YE-1h	YE-12h	All Unigene
Raw Reads	37866226	42574704	42514138	49822250	54383514	—
Clean Reads	27364554	32801610	32161100	38973026	41483248	—
Number of contigs	46354	48181	44857	51660	46708	55588
Number of characters	40707368	40872184	39108985	45165688	44225189	71821817
Maximum contig length(bp)	11988	12449	12002	10940	12280	12449
Minimum contig length(bp)	201	201	201	201	201	300
Median contig length(bp)	589	558	580	588	646	1022
Average contig length(bp)	878.18	848.31	871.86	874.29	946.84	1292.04
Contig N50 length(bp)	1340	1304	1338	1348	1461	1772
GC% content after filtering (%)	49	49	49	49	49	49
Reads mapping to All-Unigene (%)	97.61	97.68	97.95	97.44	97.81	—

Table 1. Summary of *S. miltiorrhiza* transcriptome.

diphosphate (DMAPP)^{4–6}. The phenolic acids in *S. miltiorrhiza* are synthesized via the tyrosine-derived and phenylpropanoid pathways⁷.

Until now, some key enzyme genes in the upstream of biosynthesis pathways of bioactive compounds in *S. miltiorrhiza* have been cloned and characterized by conventional cloning methods with slow speed^{8–12}. To overcome the limitations of conventional methods, next generation (NG) sequencing methods such as Illumina, provide an easy way for quick analysis of one organism's transcriptome with numerous genes. Some studies have attempted to elucidate the entire biosynthesis processes of bioactive compounds in *S. miltiorrhiza* by analyzing its transcriptome and have cloned several transcription factors (TFs) associated with the biosynthesis and regulation of these bioactive compounds^{13–18}. It is reported that some cytochrome P450 proteins, decarboxylase, dehydrogenase and reductase probably participate in the catalytic reaction steps in the downstream biosynthesis of tanshinones through analyzing the structure traits of downstream compounds. But the exact information about the downstream enzymes catalyzing the several steps from ferruginol to tanshinones in tanshinone biosynthesis pathway and from rosmarinic acid to lithospermic acid in phenolic acids biosynthesis pathway is still unknown (Supplementary Fig. S1).

Methyl jasmonate (MeJA) and yeast extract (YE) are proposed to be effective elicitors to increase the accumulation of tanshinones and phenolic acids. Most of the genes involved in these pathways are also induced by MeJA or YE¹⁹. As we found that MeJA was an effective elicitor to enhance the production of phenolic acids and tanshinones simultaneously, while YE was an elicitor only for tanshinones accumulation. Little is known about the different molecular mechanism lay behind. To identify the downstream and regulatory genes involved in tanshinone and phenolic acid biosynthesis, we conducted comprehensive transcriptome profiling of *S. miltiorrhiza* using hair roots treated with MeJA or YE by RNA-seq strategy, respectively. Some important candidate downstream biosynthetic genes and regulatory genes were predicted. It would be an interesting work for functionally identifying the downstream biosynthetic genes in the biosynthesis of tanshinones and phenolic acids and elucidating the molecular mechanism of MeJA-mediated and YE-mediated biosynthesis of tanshinone and phenolic acid.

Results and Discussion

Transcriptome sequencing, de novo assembly and sequence clustering. The sequencing of five *S. miltiorrhiza* hairy root cDNA libraries generated 37,866,226 raw reads from the non-elicitor treated control, 42,574,704 from MeJA-1h (Sample was induced by MeJA for one hour), 42,514,138 from MeJA-6h, 49,822,250 from YE-1h and 54,383,514 from YE-12h. Approximately 25% of raw reads were removed post filtering of adapter sequences, low quality reads and short reads. Further, 27364554 valid reads in control, 32801610 in MeJA-1h, 32161100 in MeJA-6h, 38973026 in YE-1h and 41483248 in YE-12h were assembled de novo using trinity²⁰, which resulted in total 55588 unigenes. The statistical summary of data is outlined in Table 1. The average unigene length was 1292.04 bp, GC content was 49% and N50 was 1772 bp.

Unique sequence annotation. Unigenes were searched against non-redundant protein database (NR) in NCBI and the Swiss-Prot database and 42458 unigenes had at least one match to known protein sequences. After consolidation, 76.4% unigenes had been functionally annotated. *S. miltiorrhiza* belongs to *Lamiaceae* family, the whole genome of this family has not been assembled nicely so far²⁰. This may be one of the reasons for relatively low homology results. Although EST database of *S. miltiorrhiza* was developed earlier but only reported very few unique transcripts^{21,22}. The sequences with unknown homology may represent genes involved in metabolic processes which are unique to this plant and whose intermediates and enzymes have not been identified so far. Moreover, transcriptome studies of other plant species have also reported functional annotation of about half of unigenes. For example, only 49.88% of the unigene could be annotated in transcriptomes of *Cymbidium sinense*²³. Further, significant match could be assigned to only 55% of the unigenes in bamboo transcriptome²⁴ and only 52.89% of the unigenes from *Amaranthus tricolor* showed significant similarity to known genes²⁵.

Differential gene expression analysis and functional classification. In *S. miltiorrhiza* hairy root cultures, methyl jasmonate (MeJA) was an effective elicitor to enhance the production of phenolic acids and tanshinones, while yeast extract (YE) was only effective for tanshinones biosynthesis. In order to identify the different

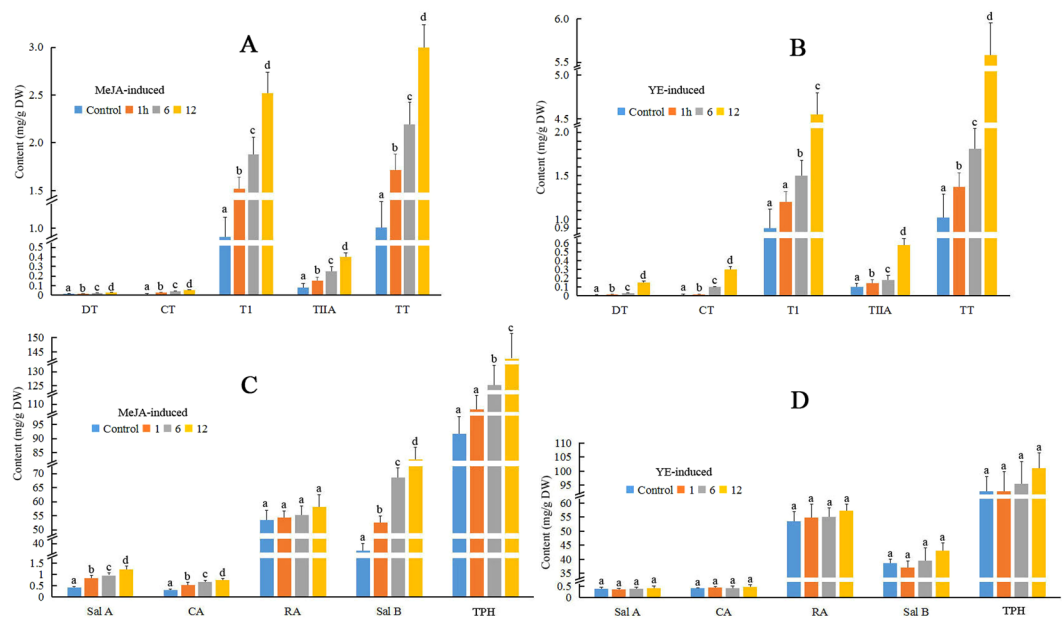


Figure 2. Content of tanshinones and phenolic acids assayed by HPLC. ((A,C) hairy root lines induced by MeJA; (B,D) hairy root lines induced by YE. Error bars represent standard error of the mean. The lower case letters above the column show the content of tanshinones and phenolic acids have a variance that are significantly different, based upon *T*-tests ($p < 0.05$); DT, dihydrotanshinone; CT, cryptotanshinone; T1, tanshinone I; TIIA, tanshinone IIA; TT, total of tanshinones; SalA, salvianolic acid A; CA, caffeic acid; RA, rosmarinic acid; SalB, salvianolic acid B; TPH, total of phenolic acids).

transcriptome implied that these genes responding to tanshinone and phenolic acid biosynthesis mainly exists in biosynthesis of secondary metabolites, phenylpropanoid, metabolic pathways, sesquiterpenoid and triterpenoid, plant hormone signal transduction and plant-pathogen interaction classes. These annotation processes for DGEs provide a valuable resource for identifying specific biosynthetic genes and related genes in signal transduction pathways that respond to MeJA and YE in *S. miltiorrhiza* separately, especially for providing useful information to examine the fine mechanism of plant secondary metabolites in *S. miltiorrhiza*.

Content of tanshinones and phenolic acids by HPLC. The total content of tanshinones (TT) and phenolic acids (TPH) in *S. miltiorrhiza* hairy roots were detected by HPLC. Tanshinones (including DT, CT, T1 and T2A) got a significant increase in hairy roots induced by MeJA and YE separately than the control. For YE-induced hairy roots, a maximum 4.47 folds increase of total tanshinones (TT) was observed, and a 1.94 folds increase were found in MeJA-induced hairy roots. MeJA, but not by YE, could induce the content of phenolic acids (TPH comprising of Sal A, Sal B and CA). For the hairy roots induced by MeJA, the total content of phenolic acid (TPH) got a maximum 0.56 fold increase compared with the control (Fig. 2).

MeJA and YE responsive genes related to tanshinones and phenolic acids biosynthesis. Previous studies found that MeJA and YE can induce the expression of genes involved in the biosynthesis of tanshinones and phenolic acids and lead to the accumulation of these compounds in *S. miltiorrhiza*^{3,19}. To elucidate the whole molecular mechanism of MeJA and YE-mediated tanshinones and phenolic acids biosynthesis, the expression of genes involved in tanshinones and phenolic acids biosynthesis were investigated. We observed 66 unigenes that encoded 26 enzymes involved in the biosynthetic pathways of tanshinones. In addition, we identified 37 unigenes that encoded 17 enzymes involved in the biosynthetic pathways of phenolic acid. These genes encode the enzymes involved in the MEP/MVA pathway and the downstream pathway for the biosynthesis of tanshinone, which include *DXS*, *DXR*, *CMK*, *MCS*, *HDS* and *IDS* in MEP pathway, *AACT*, *HMGR*, *HMGS*, *MK*, *PMK* and *MDC* in MVA pathway and *IPPI*, *GGPPS*, *GPPS*, *FPPS*, *KSL*, *CYP76AH1* and *CYP76AH3* in the downstream pathway for tanshinone biosynthesis (Supplementary Table S4). Almost all the genes in the biosynthesis pathway for tanshinone identified in our transcriptome dataset were confirmed by previous studies^{9–12}. Compared to MeJA, YE had a more significant effect on genes involved in biosynthesis of tanshinone, which got 50 unigenes (encoded 19 enzymes) having RPKM values >2 folds change more than the control while 33 unigenes (encoded 18 enzymes) got the RPKM values change 2-fold more than the control induced by MeJA (Supplementary Table S4 and Supplementary Fig. S2). The expression level of *KSL* response to MeJA is consistent with the result obtained by Ma *et al.*¹², but was not coincided with Luo *et al.*¹⁶, which may be caused by the different treated plant materials.

Unlike this, the expression patterns of genes involved in phenolic acid biosynthesis pathways were diverse. *PAL*, *C4H*, *4CL*, *TAT*, *HPPR*, *RAS* and *CYP98A14* were all induced by MeJA (with the maximum 13.6 folds RPKM values changes compared to the control). Nevertheless, YE didn't show any clear effect on these genes. It was also

consistent with change of active ingredients contents after treatment by these two elicitors, in which the content of tanshinones in hairy root could be induced by these two elicitors, but to phenolic acid, the contents could only be induced by MeJA, not by YE (Fig. 2).

Experimental validation of differential expressed genes by qRT-PCR. qRT-PCR was introduced to validate the differential gene expression obtained by RNA-seq. Twelve genes involved in tanshinone biosynthesis including *AACT1*, *CMK*, *DXR*, *DXS2*, *FPPS*, *GGPPS1*, *HMGR2*, *HMGs*, *IPPI*, *KSL1*, *CPS1* and *CYP76AH1*, were selected for qRT-PCR. The expression pattern of the these twelve genes as obtained by qRT-PCR was in line with those obtained by RNA-seq data. Statistical analysis a nice correlation between the two data was observed, with the correlation coefficient ranging from 0.932 to 0.979 (Fig. 3). Furthermore, several genes involved in phenolic acids biosynthesis including *PAL1*, *TAT1*, *4CL1*, *HPPR1*, *C4H1*, *RAS1* and *CYP98A14* were also selected for qRT-PCR detection. The result showed a nice coherence between the results of qRT-PCR and the dataset obtained by RNA-seq (Supplementary Fig. S3). It implied that the dataset obtained by RNA-seq is reliable for exploring the target downstream genes and regulatory genes involved in the biosynthesis of tanshinones and phenolic acids.

MeJA-responsive and YE-responsive P450 genes in *S. miltiorrhiza*. Cytochrome P450s being one of the biggest gene super families of plants, participate in various biochemical pathways especially in secondary metabolic pathways, including phenolic compounds, flavonoids, isoprenoids and alkaloids^{27, 28}. Co-expression analyses revealed that some P450s are coordinately expressed with key biosynthetic genes to constitute biological pathways. The combination of high-throughput transcriptome sequencing and co-regulation analysis allows for the formulation of highly precise hypotheses on the function of uncharacterized biosynthetic genes^{26, 27}. Previous reports elucidated that the expression of key genes such as *CPS1*, *KSL1* and *CYP76AH1* positively correlated with the tanshinone accumulation in hairy roots of *S. miltiorrhiza* under the treatment of MeJA and YE respectively^{3, 6, 15}. In this study, total of 45 candidate P450 unigenes in MeJA and 27 unigenes in YE-induced transcriptome were selected from the up-regulation P450 members based on co-expressed pattern analysis with the downstream biosynthetic genes involved in the biosynthesis of tanshinones, including *CPS1*, *KSL1* and *CYP76AH1*, among of which 19 P450 unigenes showed a co-expression pattern with *KSL1*, 13 P450 unigenes with *CPS1*, 16 P450 unigenes with *CYP76AH1* and only two P450 unigenes with both *KSL1* and *CPS1* (Supplementary Table S5 and Fig. 4A–D).

In the pathway of phenolic acid biosynthesis, total of 26 P450 unigenes exhibiting a co-expression pattern with genes in the metabolic pathway of phenolic acids (including *TAT1*, *HPPR1* and *RAS1*), were selected from up-regulation P450 members (RPKM values > 2 folds change between elicitor treatment and control), with 15 unigenes co-expressed with *TAT1*, 9 unigenes co-expressed with *HPPR1*, 7 unigenes co-expressed with *RAS1*, 4 unigenes co-expressed both with *HPPR1* and *RAS1* (Supplementary Table S6 and Fig. 4E–H). A cytochrome P450-dependent monooxygenase (*CYP98A14*), which can introduce 3-hydroxyl group in the late pathway to form RAs, was identified in *S. miltiorrhiza*²⁹. A P450, named *CYP76AH1* that catalyzes a unique four-electron oxidation cascade on miltiradiene to produce ferruginol, an important intermediate compound for tanshinone biosynthesis, was also recently characterized³⁰. The expression levels of these above two P450s were significantly induced by MeJA, and the studies indicated that P450s have key functions in tanshinone and phenolic acid biosynthesis^{29, 30}. Thus, the identification of these P450s contributes to the elucidation of the downstream biosynthetic pathways for tanshinone and phenolic acid biosynthesis.

CYP71 clan P450s was thought to contain the most P450 members and was known to be involved in plant secondary metabolism²⁷. In order to gain insight into the phylogenetic relationship between various P450s, a neighbour-joining phylogenetic tree was constructed with the deduced protein sequences of these selected P450s (45 P450 unigenes in MeJA and 27 P450 unigenes in YE-induced transcriptome) and P450s from other plant species. For P450 genes in tanshinones biosynthesis pathway, 29 unigenes in CYP71 clan (23 unigenes in CYP71 family, 5 unigenes in CYP76 family and 1 unigenes in CYP98 family), 9 unigenes in CYP72 clan and 7 unigenes in CYP85 clan were identified (Fig. 5A and Supplementary Fig. S4). To P450 genes in phenolic acids biosynthesis pathway, 17 unigenes in CYP71 clan (10 unigenes in CYP71 family, 5 unigenes in CYP76 family and 2 unigenes in CYP98 family), 6 unigenes in CYP72 clan and 4 unigenes in CYP85 clan were also identified (Fig. 5B and Supplementary Fig. S5). In rice, a diterpenoid biosynthetic gene cluster was identified on chromosome 2, in which two genes encoding CYP71 clan members of *CYP76M7* and *CYP71Z7* catalyze the early and later steps in phytocassane biosynthesis, respectively^{31, 32}. Thus, the *S. miltiorrhiza* CYP71 clan members can be of particular interests in further elucidation of biosynthetic pathway of tanshinones. Due to distinct catalytic activities of P450 monooxygenases, differentially expressed P450 genes of the CYP85 and CYP72 clans also hold the potential to be enzymes of the pathway.

Downstream of candidate genes involved in tanshinone and phenolic acid biosynthesis. Due to the hypothetical tanshinones biosynthetic pathway deduced by Yang *et al.*¹⁵ and Guo *et al.*¹⁷, biosynthesis of tanshinones involved in the downstream steps of decarboxylation, oxidation and reductions could be catalyzed by decarboxylase, dehydrogenase and reductase, respectively^{8, 15, 17, 30}. Unigene dataset was searched with *Zingiber zerumbet* short-chain dehydrogenase (*ZSD1*, AB480831) and *Artemisia annua* aldehyde Δ 11(13) reductase 2 (*DBR2*, EU704257) to identify candidate dehydrogenases and reductases^{33, 34}. 7 unigenes and 6 unigenes showed more than 42% protein sequence identity to *ZSD1* and *DBR2* respectively (Table 2). All the unigenes were expressed at significantly higher levels in elicitor treatment (MeJA or YE separately) after than before, with the maximum 6.96 folds change of RPKM values (FDR < 0.01), suggesting a high possibility of the dehydrogenase and reductase being involved in the biosynthesis of tanshinones.

For the biosynthesis of phenolic acids, it was confirmed that rosmarinic acid was the substrate of salvianolic acid B and the biosynthetic step was thought to be a catalytic reaction of dimerization²⁹. Ponzoni *et al.*³⁵ reported

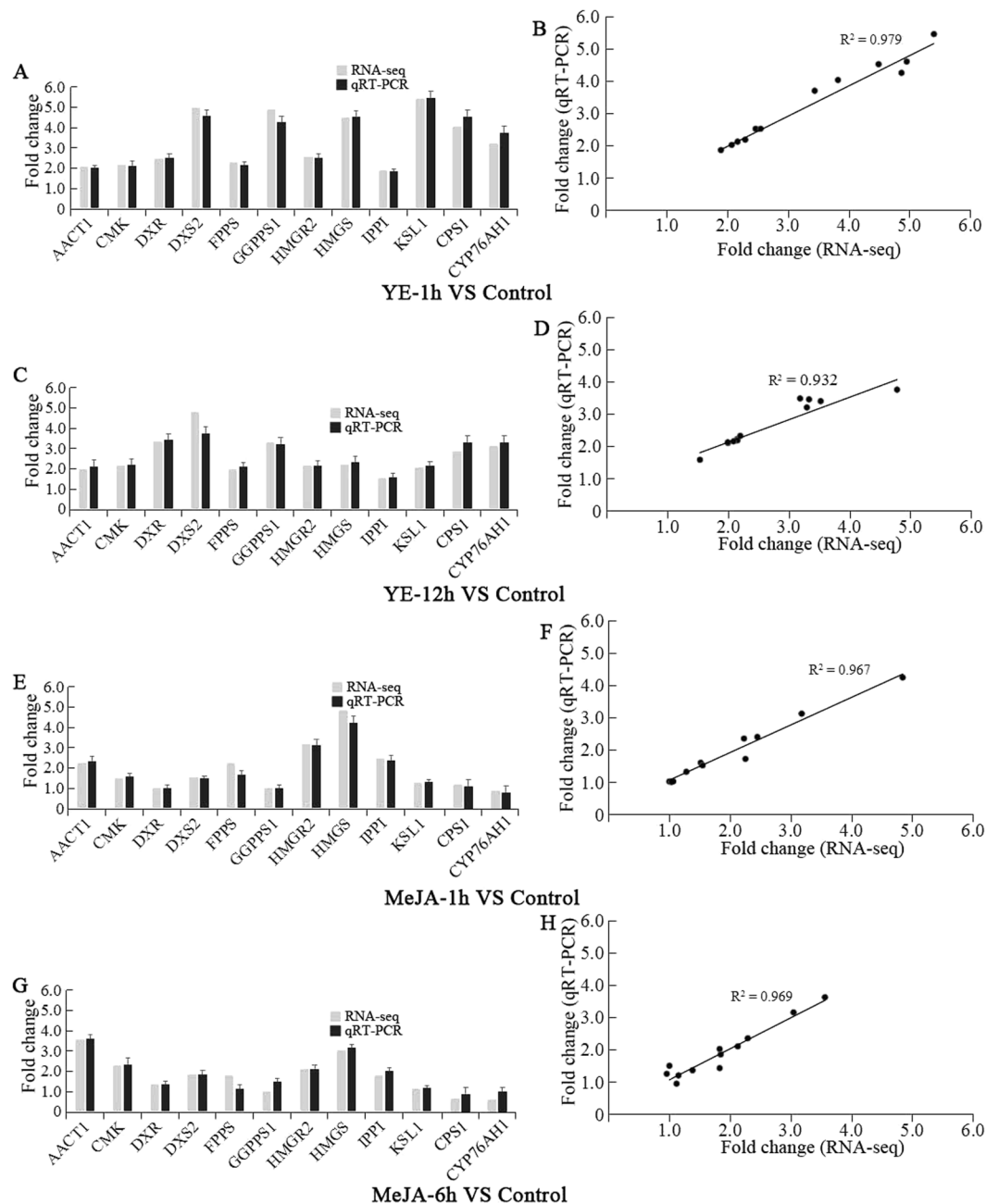


Figure 3. qRT-PCR validation of differentially expressed genes in RNA-seq dataset. (A,C,E and G) Expression of 12 genes was validated using qRT-PCR to compare with their expression in RNA-seq. Error bars indicate standard error of the mean. (B,D,F and H) Correlation of gene expression results obtained from qRT-PCR and RNA-seq).

that laccases derived from *Trametes pubescens*, could catalyze hydroxystilbene to form a dimerization reaction and produce three types of analogues (A, B and C), among of which A-type metabolite had a structure similar to salvianolic acid B. It implied that laccase might be a candidate enzyme to catalyze rosmarinci acid to form salvianolic acid B. Therefore, unigenes dataset was searched with *T. pubescens* laccase 1 A (Lap1A, AF414808) and laccase 2 (Lap2, AF414807) to identify candidate laccase involved in the biosynthesis of phenolic acids³⁶. Total of 4 unigenes being similar to *Lap1A* and 3 unigenes being similar to *Lap2*, were all identified (Table 2).

MeJA and YE-responsive TFs in *S. miltiorrhiza*. In plants, transcription factors (TFs) of different families regulate a series of life process, including development, evolution and the response to abiotic and biotic stresses. Many TFs have been found to regulate secondary metabolite biosynthesis and accumulation³⁷. One of the aims of the present work was to identify TFs which regulate the biosynthesis and accumulation of tanshinone and phenolic acid induced by YE and MeJA in *S. miltiorrhiza* hairy roots. The unigenes encoding TFs in our RNA-seq dataset were investigated. RPKM values of all the TFs in the elicitation treatment and the control were compared,

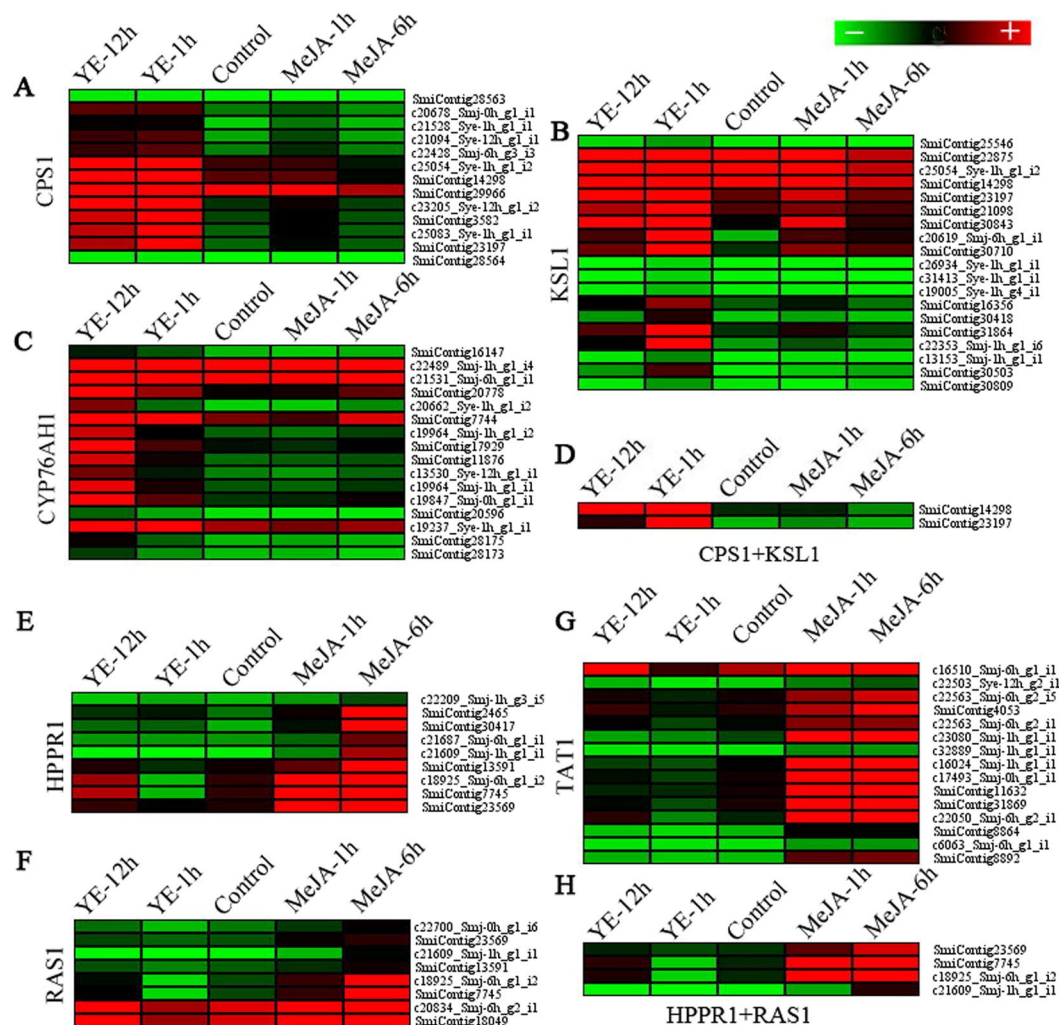


Figure 4. Co-expression analysis of P450 genes with downstream biosynthetic genes involved in the biosynthesis of tanshinones and phenolic acids. (A–H Co-expression with CPS1, KSL1, CYP76AH1, KSL1 + CPS1, HPPR1, RAS1, TAT1, HPPR1 + RAS1, separately).

which resulted in total 375 TFs with > 2-fold (FDR < 0.01) expression levels in YE or MeJA elicitation treatment than the control (Supplementary Table S7). Our dataset have more up-regulated TFs than Gao *et al.*¹⁴ and Yang *et al.*¹⁵, it might be due to different samples treated by different style. But the fact that the four major TFs showing a up-regulated expression pattern including WRKY, bHLH, AP2-EREBP and NAC is similar to the previous reports by Gao *et al.*¹⁴ and Yang *et al.*¹⁵, for which it implied their important effects in the biosynthesis of tanshinones and phenolic acids mediated by YE and MeJA (Supplementary Fig. S6).

Previous studies reported that TFs including WRKY, AP2-EREBP, MYB, GRAS and bHLH families showed more important roles in the regulation of plant secondary metabolism^{14,16,37}. In this study, we found that 59 WRKY, 28 AP2-EREBP, 38 MYB, 2 GRAS and 51 bHLH were up-regulated in YE- and/or MeJA-mediated transcriptomes, among which 23 WRKY, 6 AP2-EREBP, 4 MYB, 1 GRAS and 7 bHLH were only up-regulated by YE, not by MeJA; Reversely, to MeJA-mediated transcriptome, 8 WRKY, 13 AP2-EREBP, 15 MYB, 1 GRAS and 41 bHLH were up-regulated only by MeJA (Fig. 6 and Supplementary Table S7).

Among these up-regulated genes, AP2-EREBP, MYB and bHLH were more abundant in *S. miltiorrhiza* hairy roots treated with MeJA than those treated with YE, while WRKY family was more abundant in YE treated transcriptome which indicated that AP2-EREBP, MYB and bHLH families probably showed a more positive effect in participating the MeJA-mediated signal transduction and secondary metabolite biosynthesis, but WRKY families showed a reverse result in YE-mediated signal transduction and metabolism. The bHLH protein can physically interact with certain MYBs and WD-repeats and acts as an important regulator to modulate anthocyanin accumulation in plant species³⁸. Thus TFs responding to MeJA- and YE-mediated elicitation might interact with each other to regulate secondary metabolite biosynthesis and accumulation of tanshinone and phenolic acid. WRKY protein (CjWRKY1) is a transcriptional regulator of benzyloisoquinoline alkaloid biosynthesis in *Coptis japonica*³⁹. The AP2-EREBP members have been shown to regulate secondary metabolism pathways in several medicinal plants. In *Catharanthus roseus*, the AP2-EREBP proteins *ORCA2* and *ORCA3* positively regulate the expression

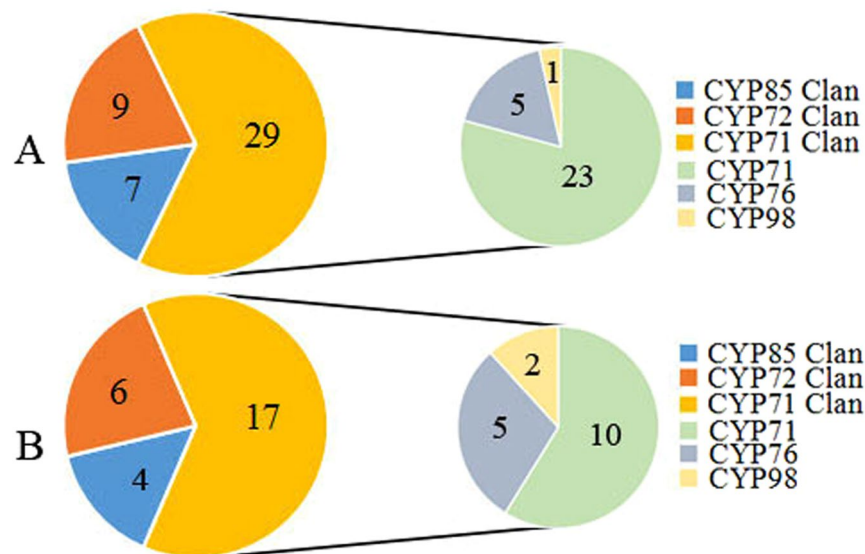


Figure 5. Candidate cytochrome P450 genes for tanshinones and phenolic acids biosynthesis. (A, Distribution of the 45 up-regulated CYPs in clans, as well as the family distribution of the 29 members from the CYP71 clan; (B) Distribution of the 27 up-regulated CYPs in clans, as well as the family distribution of the 17 members from the CYP71 clan).

Gene	Unigene ID	MeJA-1h	MeJA-6h	Control	YE-1h	YE-12h
Reductase	c23543_Smj-1h_g1_i2	321.25	184.58	91.40	75.48	112.25
	c4200_Smj-1h_g1_i1	51.46	28.91	25.91	19.84	26.72
	SmiContig16951	59.49	30.04	24.95	19.78	30.17
	SmiContig32803	397.26	184.30	129.36	143.80	161.71
	SmiContig6622	410.93	122.53	59.05	58.43	46.86
	SmiContig32802	232.74	45.01	47.77	171.43	78.49
Dehydrogenase	c16637_Smj-1h_g1_i2	10.86	3.96	4.90	14.30	4.78
	SmiContig20215	8.04	16.59	6.90	18.64	14.31
	SmiContig29584	11.32	1.89	3.16	9.49	2.58
	SmiContig29586	6.31	2.02	2.59	6.50	1.37
	SmiContig29955	8.66	3.29	3.29	10.15	3.57
	SmiContig26874	6.16	3.72	2.13	3.67	1.82
	SmiContig32840	106.37	86.31	33.41	29.99	47.66
Laccase 1 A	SmiContig4158	82.34	144.07	62.91	22.78	41.17
	c22112_Smj-6h_g1_i1	332.55	816.08	55.05	38.96	52.13
	SmiContig23578	213.25	378.59	42.47	37.30	43.50
	SmiContig4159	299.17	592.16	93.07	44.28	69.83
Laccase 2	SmiContig18382	206.52	266.23	112.76	138.61	181.98
	SmiContig6627	67.93	61.91	33.64	33.82	45.33
	SmiContig8962	19.50	1.19	9.25	11.16	9.32

Table 2. Expression of candidate downstream genes involved in the biosynthesis of tanshinones and phenolic acids.

of strictosidine synthase (*CrSTR*) and the biosynthesis of terpenoid indole alkaloid³³. In *Artemisia annua*, two JA-responsive AP2-ERE BP proteins, *AaERF1* and *AaERF2*, participate in artemisinin metabolism by binding to the promoter of *AaADS* and *AaCYP71AV1* genes, and finally activate their expression⁴⁰. In our dataset, 13 and 7 AP2-ERE BP members were induced by MeJA and YE respectively in which some members were even elevated to dozens of times compared to the control. It implied that AP2-ERE BP TFs might have a remarkable effect on secondary metabolite biosynthesis and accumulation in *S. miltiorrhiza* hairy roots.

Despite 375 TFs in *S. miltiorrhiza* hairy roots that be responsive to YE or MeJA elicitation were identified, it is possible that not all these TFs participate the biosynthesis of tanshinones and/or phenolic acids directly and the range of target TFs could be shortened by Weighted Gene Co-expression Network Analysis (WGCNA) strategy further^{26, 41, 42}. In fact, some similar research works on TFs have been studied in *Arabidopsis thaliana*,

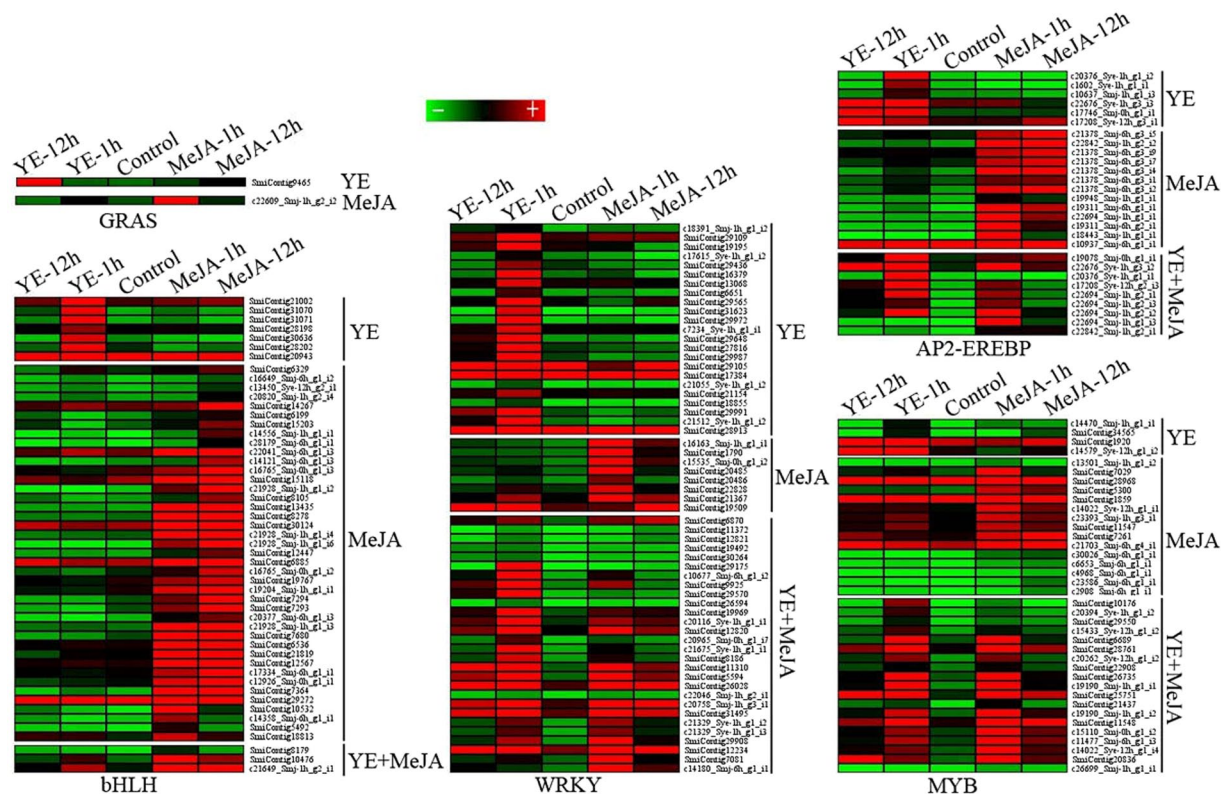


Figure 6. Expression of transcriptional factors (TFs) involved in tanshinones and phenolic acids biosynthesis. The heat graph shows TFs differential expression induced only by YE or MeJA, or by both MeJA and YE.

Catharanthus roseus, *Solanum tuberosum* and other plants^{14, 16, 33, 41, 42}. However, some TFs associated with regulation of the biosynthesis of tanshinones and phenolic acids, which were not up-regulated significantly by YE or MeJA, would be missed out in this study. Therefore it is important that identification of key biosynthetic genes and TFs using various strategies could help us to dissect the molecular mechanism underlying the biosynthesis of tanshinones and/or phenolic acids in the future. In a word, this work provides us a useful gene resource for discovery and functional identification of downstream biosynthetic genes and upstream TFs involved in the biosynthesis of tanshinones and phenolic acids.

Materials and Methods

Hairy root culture. Sterile *S. miltiorrhiza* plants were grown and maintained in our laboratory as reported previously^{9, 14}. To induce hairy roots, sterile leaf sections were used for infection with the disarmed *A. tumefaciens* strain C58C1 (pRi A4) as reported before by our groups¹⁰. The hairy root clones were routinely subcultured every 30 days on solid 1/2 MS medium. Rapidly growing hairy roots with normal growth status were used to establish hairy root lines, and the liquid 1/2 MS medium were changed every 20 days. Forty-day-old shake-flask cultured hairy roots were treated with elicitors, and the hairy roots were harvested from the culture medium at selected times after treatment by MeJA (0 h, 1 h and 6 h) or YE (1 h and 12 h) separately⁴⁻⁶. All the induced hairy root lines were harvested for RNA isolation.

Preparation of MeJA and YE elicitors. MeJA was dissolved in ethanol followed by filter-sterilization and added to the culture medium at a final concentration of 0.1 mM. A total of 25 g yeast extract from Bio Basic Inc was dissolved with 125 ml distilled water, and then yeast elicitors were prepared by two rounds of ethanol precipitation as reported before⁴³.

RNA isolation, quantification and qualification. Total RNAs from different hairy root lines were extracted using the *TRIzol* Reagent (Invitrogen) and treated with *DNase I* (Takara) according to manufacturer's protocols⁴⁴. Each sample was prepared by mixing three replicate samples. RNA were monitored using EtBr-stained 1% agarose gel. The RNA purity and concentration were determined using a Nanodrop spectrophotometer (Thermo).

cDNA library preparation and sequencing. Reverse transcription was performed with the SMART II™ cDNA Synthesis Kit (Clontech, USA) following the manufacturer's instructions. Double strand cDNAs were separated on 2% agarose gel, and those with size over 100 bp were recovered. The RNA-seq and construction of the libraries were performed by Novellbio Biotechnology Corporation (Shanghai, China) and the cDNA library

was sequenced using Illumina HiSeq™2000. All reads were deposited in the Short Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) public database under the accession GSE100970.

Data analysis and assembly. The sequence reads is processed to remove adapter sequences, low quality reads and very short length reads. Paired reads were quality filtered using NGS QC toolkit v 2.3⁴⁵. The cutoff for quality score is > 20 Q30 score with high quality bases > 70% of read length. High quality reads were used for de novo assembly using Trinity software with K-mer of 25. The assembly resulted in contigs and singletons which together form set of unigenes.

Unique sequence annotation and functional classification. Unigenes were used as query sequences to search against the non-redundant protein (NR) database at NCBI (<http://www.ncbi.nlm.nih.gov>) and the Swiss-Prot protein database (<http://www.ebi.ac.uk/uniprot>) with *E*-value cutoff of $1e^{-5}$. The annotations of the best hits were recorded. Gene Ontology (GO) (<http://www.geneontology.org/>) was further used to category the function of the unigenes by Blast2GO⁴⁶, and the unigenes were assigned to biological functions on the KEGG levels of ‘biological process’, ‘cellular component’ and ‘molecular function’. The unigenes were assigned by KEGG Automatic Annotation Server (KAAS) upon the KEGG pathways database (<http://www.genome.jp/kegg/>)⁴⁷.

Differential expression and co-expression network analysis. For differentially expressed genes assay, reads per kilobase per million reads (RPKM) values were used to normalize gene expression levels. Statistical comparison of RPKM values between elicitors-induced and non-induced sample was performed using software Bowtie⁴⁸. Statistical test analysis was done using FDR method⁴⁹. The corrected *P*-value (false discovery rate, FDR) of 0.05 was set as the threshold for significant differential expression. A gene expression correlation network between the candidate synthetic enzyme genes in the tanshinones and phenolic acids biosynthesis pathway and candidate P450s or transcriptional factors was constructed with the Weighted Gene Co-expression Network Analysis (WGCNA) method²⁶, co-expression correlation was defined based on the following criteria: $FDR \leq 0.001$ and $P\text{-value} < 0.01$. One color bar represents a range of coefficient values; the symbol ‘-’ denotes the declining expression pattern and the symbol ‘+’ denotes the increasing expression pattern. Hierarchical clustering was used to examine the correlation between the elicitors-induced and non-induced samples by the K-means clustering method. Heat map was produced by the software Heatmap 2.0⁵⁰.

Determination of tanshinones and phenolic acids by high-performance liquid chromatography (HPLC). Elicitors-induced and non-induced hairy root lines were harvested and dried. Dried hairy roots (200 mg) were pulverized and extracted with 16 mL of methanol/dichloromethane (3:1, v/v), sonicated for 1 h and then kept at room temperature for 24 h. Isolation of tanshinones and phenolic acids for HPLC analysis were performed as reported before¹³. Four tanshinone species comprising of dihydrotanshinone (DT), cryptotanshinone (CT), tanshinone I (T1) and tanshinone IIA (T2A), and four phenolic acids including Salvianolic acid A (Sal A), Salvianolic acid B (Sal B), Caffeic acid (CA) and Rosmarinic acid (RA) were detected and quantified by comparison with authentic standard curves and retention times. Total tanshinone comprising of DT, CT, T1 and T2A, and total phenolic acids comprising of Sal A, Sal B, CA and RA were designated as TT and TPH separately in this paper.

Validation of RNA-Seq data by quantitative real-time RT-PCR (qRT-PCR). First strand cDNA was synthesized from 100 ng total RNA using reverse transcriptase (Takara, Japan). For each reaction, 1/10th of the RT reaction mixes was used as the template for PCR. The sequences of primers used in this study are shown in Supplementary Table S8. Real-time PCR was performed using SYBR Pre-mix *Ex Taq*™ II kit (Promega, Beijing) and run on Step One Software v2.0 (ABI Step One, USA). Gene expression were calculated in relation to the reference gene *18srRNA* using $2^{-\Delta\Delta CT}$ method⁵¹. Each generated data point represented the average of three independent biological replicates.

Data availability statement. The authors declare that all the data in this manuscript is available.

Ethical Standard. The experiment conducted complies with the laws of China.

References

- Kai, G. Y. *et al.* Metabolic engineering tanshinone biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *Metabolic engineering* **13**, 319–327, doi:10.1016/j.ymben.2011.02.003 (2011).
- Hua, W. P. *et al.* De novo transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. *Genomics* **98**, 272–279, doi:10.1016/j.ygeno.2011.03.012 (2011).
- Kai, G. Y. *et al.* Molecular mechanism of elicitor-induced tanshinone accumulation in *Salvia miltiorrhiza* hairy root cultures. *Acta Physiol. Plant* **34**, 1421–1433 (2012).
- Shi, M. *et al.* Methyl jasmonate induction of tanshinone biosynthesis in *Salvia miltiorrhiza* hairy roots is mediated by JASMONATE ZIM-DOMAIN repressor proteins. *Sci. Rep.* **6**, 20919, doi:10.1038/srep20919 (2016).
- Shi, M. *et al.* Enhanced Diterpene Tanshinone Accumulation and Bioactivity of Transgenic *Salvia miltiorrhiza* Hairy Roots by Pathway Engineering. *J. Agric. Food Chem.* **64**, 2523–2530, doi:10.1021/acs.jafc.5b04697 (2016).
- Hao, X. L. *et al.* Effects of methyl jasmonate and salicylic acid on the tanshinone production and biosynthetic genes expression in transgenic *Salvia miltiorrhiza* hairy roots. *Biotechnol. App. Bioc.* **62**, 24–31, doi:10.1002/bab.1236 (2015).
- Laulle, O. *et al.* Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **100**, 6866–6871, doi:10.1073/pnas.1031755100 (2003).
- Gao, W. *et al.* A functional genomics approach to tanshinone biosynthesis provides stereochemical insights. *Org. Lett.* **11**, 5170–5173, doi:10.1021/ol902051v (2009).
- Liao, P. *et al.* Molecular cloning, characterization and expression analysis of a new gene encoding 3-hydroxy-3-methylglutaryl coenzyme A reductase from *Salvia miltiorrhiza*. *Acta Physiologiae Plantarum* **31**, 565–572, doi:10.1016/j.jplph.2010.06.008 (2009).

10. Song, J. & Wang, Z. Molecular cloning, expression and characterization of a phenylalanine ammonia-lyase gene (SmPAL1) from *Salvia miltiorrhiza*. *Mol. Biol. Rep.* **36**, 939–952, doi:10.1007/s11033-008-9266-8 (2009).
11. Dai, Z., Cui, G., Zhou, S. F., Zhang, X. & Huang, L. Cloning and characterization of a novel 3-hydroxy-3-methylglutaryl coenzyme A reductase gene from *Salvia miltiorrhiza* involved in diterpenoid tanshinone accumulation. *Journal of plant physiology* **168**, 148–157, doi:10.1016/j.jplph.2010.06.008 (2011).
12. Zhou, W. *et al.* Molecular cloning and characterization of two 1-deoxy-d-xylulose-5-phosphate synthase genes involved in tanshinone biosynthesis in *Salvia miltiorrhiza*. *Molecular breeding* **36**, 124, doi:10.1007/s11032-016-0550-3 (2016).
13. Chen, H., Wu, B., Nelson, D. R., Wu, K. & Liu, C. Computational Identification and Systematic Classification of Novel Cytochrome P450 Genes in *Salvia miltiorrhiza*. *PLoS One* **9**, e115149, doi:10.1371/journal.pone.0115149 (2014).
14. Gao, W. *et al.* Combining metabolomics and transcriptomics to characterize tanshinone biosynthesis in *Salvia miltiorrhiza*. *BMC Genomics* **15**, 73, doi:10.1186/1471-2164-15-73 (2014).
15. Yang, L. *et al.* Transcriptome analysis of medicinal plant *Salvia miltiorrhiza* and identification of genes related to tanshinone biosynthesis. *PLoS One* **8**, e80464, doi:10.1371/journal.pone.0080464 (2013).
16. Luo, H. *et al.* Transcriptional data mining of *Salvia miltiorrhiza* in response to methyl jasmonate to examine the mechanism of bioactive compound biosynthesis and regulation. *Physiol. Plant* **152**, 241–255, doi:10.1111/ppl.12193 (2014).
17. Guo, J. *et al.* Cytochrome P450 promiscuity leads to a bifurcating biosynthetic pathway for tanshinones. *New Phytol.* **210**, 525–534, doi:10.1111/nph.13790 (2016).
18. Scheler, U. *et al.* Elucidation of the biosynthesis of carnosic acid and its reconstitution in yeast. *Nat. Commun.* **7**, 12942, doi:10.1038/ncomms12942 (2016).
19. Xiao, Y. *et al.* Methyl jasmonate dramatically enhances the accumulation of salvianolic acids in *Salvia miltiorrhiza* hairy root cultures. *Physiologia plantarum* **137**, 1–9, doi:10.1111/j.1399-3054.2009.01257.x (2009).
20. Xu, H. *et al.* Analysis of the Genome Sequence of the Medicinal Plant *Salvia miltiorrhiza*. *Mol. Plant* **9**, 949–952, doi:10.1016/j.molp.2016.03.010 (2016).
21. Yan, Y., Wang, Z., Tian, W., Dong, Z. & Spencer, D. F. Generation and analysis of expressed sequence tags from the medicinal plant *Salvia miltiorrhiza*. *Sci China Life Sci* **53**, 273–285, doi:10.1007/s11427-010-0005-8 (2010).
22. Karaca, M., Ince, A. G., Aydin, A. & Ay, S. T. Cross-genera transferable e-microsatellite markers for 12 genera of the Lamiaceae family. *J. Sci. Food Agric.* **93**, 1869–1879, doi:10.1002/jsfa.5982 (2013).
23. Zhang, J. *et al.* Transcriptome analysis of *Cymbidium sinense* and its application to the identification of genes associated with floral development. *BMC Genomics* **14**, 279, doi:10.1186/1471-2164-14-279 (2013).
24. Zhang, X. M., Zhao, L., Larson-Rabin, Z., Li, D. Z. & Guo, Z. H. De novo sequencing and characterization of the floral transcriptome of *Dendrocalamus latiflorus* (Poaceae: Bambusoideae). *PLoS One* **7**, e42082, doi:10.1371/journal.pone.0042082 (2012).
25. Liu, S., Kuang, H. & Lai, Z. Transcriptome analysis by Illumina high-throughput paired-end sequencing reveals the complexity of differential gene expression during *in vitro* plantlet growth and flowering in *Amaranthus tricolor* L. *PLoS One* **9**, e100919, doi:10.1371/journal.pone.0100919 (2014).
26. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).
27. Guttikonda, S. K. *et al.* Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. *BMC plant biology* **10**, 243, doi:10.1186/1471-2229-10-243 (2010).
28. Geisler, K. *et al.* Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc. Natl. Acad. Sci. USA* **110**, E3360–E3367, doi:10.1073/pnas.1309157110 (2013).
29. Di, P. *et al.* ¹³C tracer reveals phenolic acids biosynthesis in hairy root cultures of *Salvia miltiorrhiza*. *ACS Chem. Biol.* **19**, 1537–1548, doi:10.1021/cb3006962 (2013).
30. Guo, J. *et al.* CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. *Proc. Natl. Acad. Sci. USA* **110**, 12108–12113, doi:10.1073/pnas.1218061110 (2013).
31. Swaminathan, S., Morrone, D., Wang, Q., Fulton, D. B. & Peters, R. J. CYP76M7 is an ent-cassadiene C11 α -hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *The Plant Cell* **21**, 3315–3325, doi:10.1105/tpc.108.063677 (2009).
32. Wu, Y., Hillwig, M. L., Wang, Q. & Peters, R. J. Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7. *FEBS letters* **585**, 3446–3451, doi:10.1016/j.febslet.2011.09.038 (2011).
33. Zhang, H. *et al.* The basic helix-loop-helix transcription factor *CrMYC2* controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in *Catharanthus roseus*. *Plant J.* **67**, 61–71, doi:10.1111/j.1365-3113X.2011.04575.x (2011).
34. Okamoto, S. *et al.* A shortchain dehydrogenase involved in terpene metabolism from *Zingiber Zerumbet*. *FEBS J.* **278**, 2892–2900, doi:10.1111/j.1742-4658.2011.08211.x (2011).
35. Ponzoni, C. *et al.* Laccase-catalyzed dimerization of hydroxystilbenes. *Advanced synthesis & catalysis* **349**, 1497–1506 (2007).
36. Pogni, R., Baratto, M. C., Sinicropi, A. & Basosi, R. Spectroscopic and computational characterization of laccases and their substrate radical intermediates. *Cell Mol. Life Sci.* **72**, 885–896, doi:10.1007/s00018-014-1825-7 (2015).
37. Yang, C. Q. *et al.* Transcriptional Regulation of Plant Secondary Metabolism. *Journal of integrative plant biology* **54**, 703–712, doi:10.1111/j.1744-7909.2012.01161.x (2012).
38. Ramsay, N. A. & Glover, B. J. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends in plant science* **10**, 63–70, doi:10.1016/j.tplants.2004.12.011 (2005).
39. Kato, N. *et al.* Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in *Coptis japonica*. *Plant Cell Physiol.* **48**, 8–18, doi:10.1093/pcp/pcl041 (2007).
40. Yu, Z. X. *et al.* The jasmonateresponsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. *Mol. Plant* **5**, 353–365, doi:10.1093/mp/ssr087 (2012).
41. Pillet, J., Yu, H. W., Chambers, A. H., Whitaker, V. M. & Folta, K. M. Identification of candidate flavonoid pathway genes using transcriptome correlation network analysis in ripe strawberry (*Fragaria × ananassa*) fruits. *J. Exp. Bot.* **66**, 4455–4467, doi:10.1093/jxb/erv205 (2015).
42. Saito, K., Hirai, M. Y. & Yonekura-Sakakibara, K. Decoding genes with coexpression networks and metabolomics- ‘majority report by precogs’. *Trends Plant Sci.* **13**, 36–43, doi:10.1016/j.tplants.2007.10.006 (2008).
43. Ge, X. & Wu, J. Tanshinone production and isoprenoid pathways in *Salvia miltiorrhiza* hairy roots induced by Ag⁺ and yeast elicitor. *Plant Science* **168**, 487–491, doi:10.1016/j.plantsci (2005).
44. Zhou, W. *et al.* Mapping of Ppd-B1, a Major Candidate Gene for Late Heading on Wild Emmer Chromosome Arm 2BS and Assessment of Its Interactions with Early Heading QTLs on 3AL. *PLoS ONE* **11**, e0147377, doi:10.1371/journal.pone.0147377 (2016).
45. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644–652, doi:10.1038/nbt.1883 (2011).
46. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676, doi:10.1093/bioinformatics/bti610 (2005).
47. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185, doi:10.1093/nar/gkm321 (2007).

48. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* 5, 621–628, doi:10.1038/nmeth.1226 (2008).
49. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. & Golani, I. Controlling the false discovery rate in behavior genetics research. *Behavioural brain research* 125, 279–284 (2001).
50. Toddenroth, D., Ganslandt, T., Castellanos, I., Prokosch, H. U. & Bürkle, T. Employing heat maps to mine associations in structured routine care data. *Artif. Intell. Med.* 60, 79–88, doi:10.1016/j.artmed.2013.12.003 (2014).
51. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408, doi:10.1006/meth.2001.1262 (2001).

Acknowledgements

This work was supported by National Natural Science Fund (81522049, 31270007, 31571735), New Century Talent Project (NECT-13-0902), Shanghai Science and Technology Committee Project (17JC1404300, 15430502700), Shuguang Program of Shanghai Education Development Foundation and Shanghai Municipal Education Commission (16SG38), the State Key Laboratory of Phytochemistry and Plant Resources in West China (P2015-KF02).

Author Contributions

G.K. conceived and designed the study. Q.H., W.Z., X.W., M.S., F.H. and S.L. performed research, W.Z., Q.H., X.W., Z.Z., M.S. and M.D. analyzed data, W.Z., Q.H., Y.W. and G.K. wrote the paper.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-10215-2

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017