PORTLAND PRESS

Research Article

Exploring the biosynthetic pathway of lignin in *Acorus tatarinowii* Schott using *de novo* leaf and rhizome transcriptome analysis

Shengxiang Zhang[1,2,*], Liqiang Zhao[1,2,*], Chunmiao Shan[1,2], Yuanyuan Shi[1,2], Kelong Ma[1,3] and
Jiawen Wu[1,2,4]

[1]Anhui University of Chinese Medicine and Anhui Academy of Chinese Medicine, Hefei 230038, China; [2]Key Laboratory of Xin'an Medicine, Ministry of Education, Anhui University of Chinese Medicine, Hefei 230038, China; [3]Clinical College of Integrated Traditional Chinese and Western Medicine, Anhui University of Chinese Medicine, Hefei 230012, China; [4]Synergetic Innovation Center of Anhui Authentic Chinese Medicine Quality Improvement, Hefei 230012, China

Correspondence: Kelong Ma (makelong210@126.com) or Jiawen Wu (wujiawen@ahtcm.edu.cn)

OPEN ACCESS

Acorus tatarinowii Schott is a well-known Chinese traditional herb. Lignin is the major biologically active ingredient and exerts a broad range of pharmacological effects: it is an antitumor, antioxidant and bacteriostatic agent, and protects the cardiovascular system. In the present study, the transcriptomes of the leaf and rhizome tissues of *A. tatarinowii* Schott were obtained using the BGISEQ-500 platform. A total of 141777 unigenes were successfully assembled, of which 76714 were annotated in public databases. Further analysis of the lignin biosynthesis pathway revealed a total of 107 unigenes encoding 8 key enzymes, which were involved in this pathway. Furthermore, the expression of the key genes involved in lignin synthesis in different tissues was identified by quantitative real-time PCR. Analysis of the differentially expressed genes (DEGs) showed that most of the up-regulated unigenes were enriched in rhizome tissues. In addition, 2426 unigenes were annotated to the transcriptome factor (TF) family. Moreover, 16 TFs regulating the same key enzyme (peroxidase) were involved in the lignin synthesis pathway. The alignment of peroxidase amino acid sequences and the analysis of the structural characteristics revealed that the key peroxidase enzyme had well-conserved sequences, spatial structures, and active sites. The present study is the first to provide comprehensive genetic information on *A. tatarinowii* Schott at the transcriptional level, and will facilitate our understanding of the lignin biosynthesis pathway.

Introduction

Acorus tatarinowii Schott (*A. tatarinowii*) is a well-known traditional Chinese herb that is abundant in eastern and southern Asia. *A. tatarinowii* is believed to have the potential to alleviate cognitive decline, epilepsy, and dementia [1]. The main biologically active ingredients are phenylpropanoids, lignins, sesquiterpenoids, and alkaloids [2,3]. The active ingredients of *A. tatarinowii* confer a broad range of pharmacological effects, including hypoglycemic, anti-Alzheimer's disease (AD), antiepileptic, neuroprotective, and memory-enhancing activities [4,5].

Lignin is the major bioactive ingredient in *A. tatarinowii*. Lignin is a composite phenolic polymer, composed mainly of hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units, which are formed through the oxidative polymerization of *p*-coumaryl, coniferyl, and sinapyl alcohols, respectively [6]. The main functions of lignin are to provide mechanical support for plants, water conductivity, and protection against plant

pathogens [7,8]. Reduced lignin contents often affect lodging resistance [9]. In addition, lignin has a waterproofing effect on the cell wall, forms a vascular system for the transport of water and solutes, and participates in the protection of plants from pathogens [10]. The initial three steps of lignin biosynthetic pathway, catalyzed by phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumaroyl CoA-ligase (4CL), are mandatory and provide the basis for all subsequent branches and resulting metabolites. PAL participates in the first step, catalyzing the nonoxidative deamination of phenylalanine to cinnamic acid and directing the carbon flow from the phenylalanine pathway [11]. Cinnamic acid forms *p*-cinnamic acid, which is catalyzed by C4H; *p*-cinnamic acid is then converted into *p*-coumaroyl-CoA and feruloyl-CoA by 4CL and caffeic acid 3-O-methyltransferase (COMT) [12]. Cinnamyl-CoA reductase (CCR) is responsible for catalyzing the most important metabolic reaction in the biosynthesis of lignin monomers [13]. Finally, monomeric lignin is catalytically synthesized by cinnamyl-alcohol dehydrogenase (CAD), peroxidase (POX), and ferulate-5-hydroxylase (F5H) [14,15]. Although researchers have studied *A. tatarinowii* for more than a century, the genomic information of *A. tatarinowii* remains incomplete.

RNA sequencing is a particularly effective technology for gene discovery, especially for species without a reference genome [16]. This technology has been used widely in the study of traditional medicinal plants, such as *Rehmannia glutinosa* [17], *Andrographis paniculata* [18], *Solanum elaeagnifolium* [19], and *Arisaema heterophyllum* Blume [20]. Comprehensive transcriptome information can be analyzed further to identify gene expression levels, gene sequences, biosynthetic pathway, and key enzymes, and to reveal candidate genes potentially involved in specific metabolic pathways. However, given that the biosynthesis of lignin and other secondary metabolites in *A. tatarinowii* is still largely unknown, further research on this species has been hindered. In the present study, we characterized the transcriptomes of the leaf tissue and rhizome tissue of *A. tatarinowii* and identified the unigenes encoding key enzymes in the lignin biosynthesis pathway. The transcriptome data for *A. tatarinowii* reported in the present study will provide an important resource for the study of lignin biosynthesis and other metabolic pathways. Furthermore, this database will provide important clues to explore the biological characteristics of other plants closely related to *A. tatarinowii*.

Results
RNA-Seq and assembly
In total, 124084 raw reads were obtained from the rhizome tissue and 76197 were obtained from the leaf tissue. The raw reads were generated using the BGISEQ-500 sequencing platform. The Q30 of each plant tissue was greater than 86.82%, and the GC content of each plant tissue was approximately 42.11%. In total, 141777 unigenes were assembled and clustered from high-quality reads, thorough quality control, and filtering. Among these unigenes, 51352 (36.22%) were longer than 1000 nt and 85212 (53.69%) were longer than 500 nt. The average length of the unigenes was 1108 nt and the N50 was 2062 nt. The distribution of the lengths of the unigenes is presented in Supplementary Figure S1. The quality of transcriptome sequencing was tested using a single copy orthologous database (BUSCO), and 96% of unigenes had exact matches in the BUSCO database (Supplementary Figure S2). The complete transcriptome data obtained in this sequencing have been deposited in the NCBI SRA database (accession number: SRA226668).

Functional annotation of unigenes
The 141777 identified unigenes were annotated using seven NCBI publicly available databases; the annotations were distributed as follows: Nr, 77368 unigenes (54.57%); Nt, 57501 unigenes (40.56%); Swissprot, 58557 unigenes (41.30%); KEGG, 62765 unigenes (44.27%); GO, 42681 unigenes (30.10%); KOG, 62928 unigenes (44.39%), and Pfam, 56187 unigenes (39.63%) (Table 1). A totaol of 76714 (68.22%) unigenes were annotated in at least one public functional database, and 23137 (16.32%) co-annotated unigenes were identified in five databases (Figure 1A). We obtained the species homology distribution of *A. tatarinowii* transcripts from the Nt annotations. The matched unigenes revealed that *A. tatarinowii* had the closest homology with *Nelumbo nucifera* (12.16%) in the Nr database, followed by *Elaeis guineensis* (12.14%), *Phoenix dactylifera* (8.35%), *Macleaya cordata* (8.28%), and *Ananas comosus* (3.72%) (Figure 1B).

GO functional annotations for *A. tatarinowii* unigenes were obtained from the Nr database: 42681 unigenes (30.10%) were categorized into main GO groups of molecular function, cellular components, and biological processes. In the biological processes group, most of the unigenes were related to 'cellular process' (13121 unigenes), 'biological regulation' (5009 unigenes), 'metabolic process' (3173 unigenes), and 'localization' (3038 unigenes). In the molecular function group, 21714 unigenes were related to 'binding,' and 20872 unigenes were most abundant in the

Table 1 Summary statistics of annotations for *Acorus tatarinowii* unigenes from seven publicly available databases

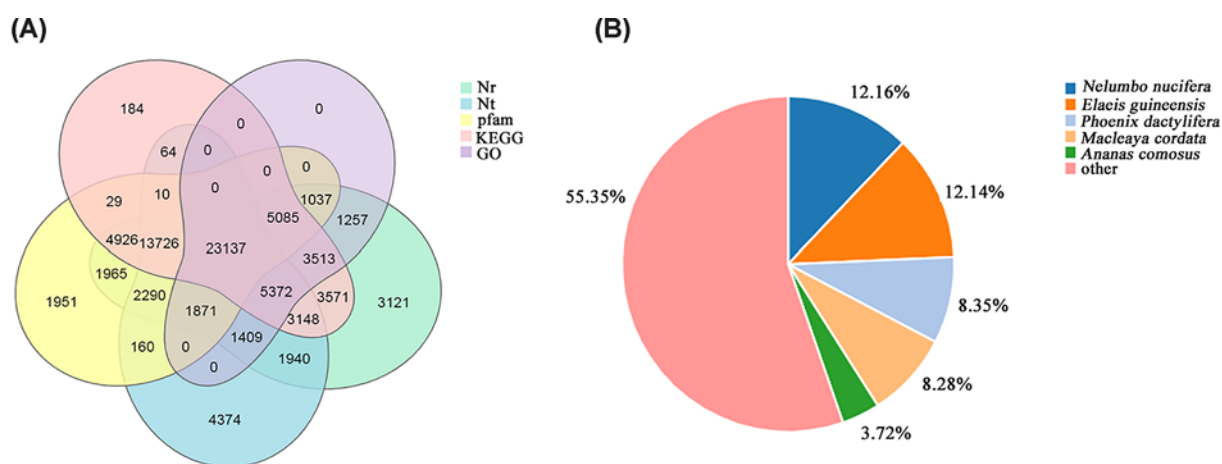| Database	Number annotated	Percentage (%)
Nr	77368	54.57
Nt	57501	40.56
Swissprot	58557	41.30
KEGG	62765	44.27
KOG	62928	44.39
Pfam	56187	39.63
GO	42681	30.10
Overall	84705	59.75

Figure 1. Unigene functional annotation

(**A**) Venn diagram of annotated unigenes in five different databases. (**B**) Species distribution of the annotations in the Nr database for *Acorus tatarinowii*.

'catalytic activity' class. In the cellular components group, most of the unigenes were involved in 'cell' (13439 unigenes), followed by 'membrane part' (11735 unigenes), and 'organelle part' (5289 unigenes) (Supplementary Figure S3).

In total, 62928 unigenes (44.39%) were classified into 25 unigene functional groups, with 11690 unigenes classified in the 'general function prediction only' group and 7981 unigenes classified in the 'signal transduction mechanisms' group.

Overview of expression and MAPMAN analysis

The expression value of all transcripts (fragments per kilobase of transcript per million mapped reads (FPKMs) > 1) in each tissue was counted. In total, 41327 and 57460 unigenes were expressed in the leaf tissue and rhizome tissue, respectively (Figure 2A). The overall gene expression level in rhizome tissue was higher than that in leaf tissue (Figure 2B).

The overall level of gene expression was analyzed in the *A. tatarinowii* by using MAPMAN software. The most common categories for the unigenes were protein (7.92%), followed by RNA (5.32%), signaling (3.03%), transport (2.46%), miscellaneous function (misc) (2.05%), and cell (1.76%) (Supplementary Figure S4). An overview of the metabolic and secondary metabolic pathways was prepared from the MAPMAN analysis based on the unigenes with FPKM > 1 (Supplementary Figure S5). In the 'metabolic pathway' subcategory, the highest number of unigenes was assigned to 'secondary metabolism.' The 'secondary metabolic pathways' subcategory contained 20 pathways, and most of the unigenes were in the 'phenylpropanoids' pathway.
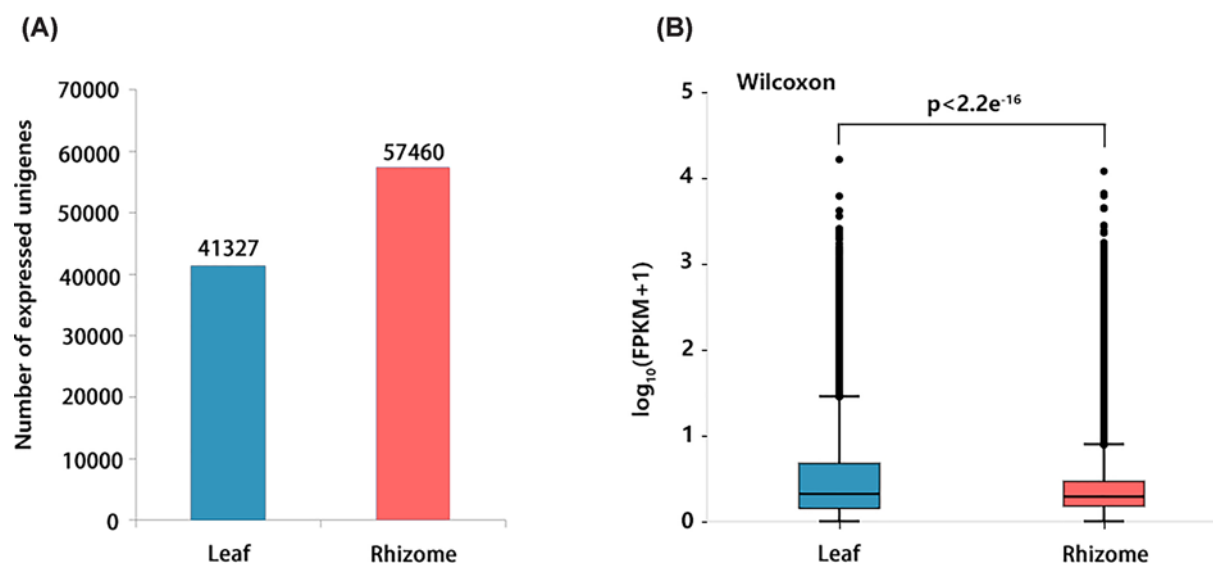
Figure 2. Overall expression profiles of *Acorus tatarinowii* leaf and rhizome tissues
(**A**) The distribution of the number of expressed unigenes in each tissue. (**B**) Boxplot of unigenes expressed in each tissue. The x-axis represents the two tissues and y-axis represents $\log_{10}$ (FPKM + 1) values.

Table 2 Number of unigenes encoding the key enzymes involved in the biosynthesis of lignin metabolism in *Acorus tatarinowii*

Abbreviation	EC number	Unigene number	Enzyme name
PAL	4.3.1.24	7	Phenylalanine ammonia-lyase
C4H	1.14.14.91	13	*trans*-cinnamate 4-monooxygenase
COMT	2.1.1.68	9	Caffeic acid 3-O-methyltransferase
4CL	6.2.1.12	13	4-coumarate-CoA ligase
CCR	1.2.1.44	15	Cinnamoyl-CoA reductase
CAD	1.1.1.195	12	Cinnamyl-alcohol dehydrogenase
POX	1.11.1.7	16	Peroxidase
F5H	1.14.-.-	22	Ferulate-5-hydroxylase

Characterization of functional genes involved in lignin biosynthesis via KEGG analysis

In total, 62765 unigenes (44.27%) were mapped to canonical pathways and assigned to 20 pathways based on the KEGG database. The most prevalent functional group in the KEGG pathways was 'global and overview maps' (13855 unigenes), followed by 'carbohydrate metabolism' (5653 unigenes), 'translation' (5232 unigenes), and 'folding, sorting and degradation' (4185 unigenes) (Supplementary Figure S6 and S7). The 'biosynthesis of other secondary metabolites' subcategory contained 16 biosynthetic pathways, including phenylpropanoid biosynthesis (ko00940), flavonoid biosynthesis (ko00941), isoquinoline alkaloid biosynthesis (ko00950), tropane, piperidine and pyridine alkaloid biosynthesis (ko00960), stilbenoid, diarylheptanoid and gingerol biosynthesis (ko00945), monobactam biosynthesis (ko00261), flavone and flavonol biosynthesis (ko00944), caffeine metabolism (ko00232), isoflavonoid biosynthesis (ko00943), betalain biosynthesis (ko00965), glucosinolate biosynthesis (ko00966), benzoxazinoid biosynthesis (ko00402), indole alkaloid biosynthesis (ko00901), anthocyanin biosynthesis (ko00942), biosynthesis of secondary metabolites-unclassified (ko00999), and carbapenem biosynthesis (ko00332) (Figure 3). Among these pathways, we annotated 1309 unigenes involved in phenylpropanoid biosynthesis and 219 unigenes involved in flavonoid biosynthesis. From further analysis of the lignin biosynthesis pathway, 107 unigenes encoding the key enzymes involved in lignin biosynthesis were identified, including PAL, C4H, COMT, 4CL, CCR, CAD, POX, F5H (Table 2). Most of the genes encoding CCR were more highly expressed in the leaf tissue, whereas the expression of most genes encoding PAL, C4H, POX, COMT, 4CL, and CAD was higher in rhizomes (Figure 4).
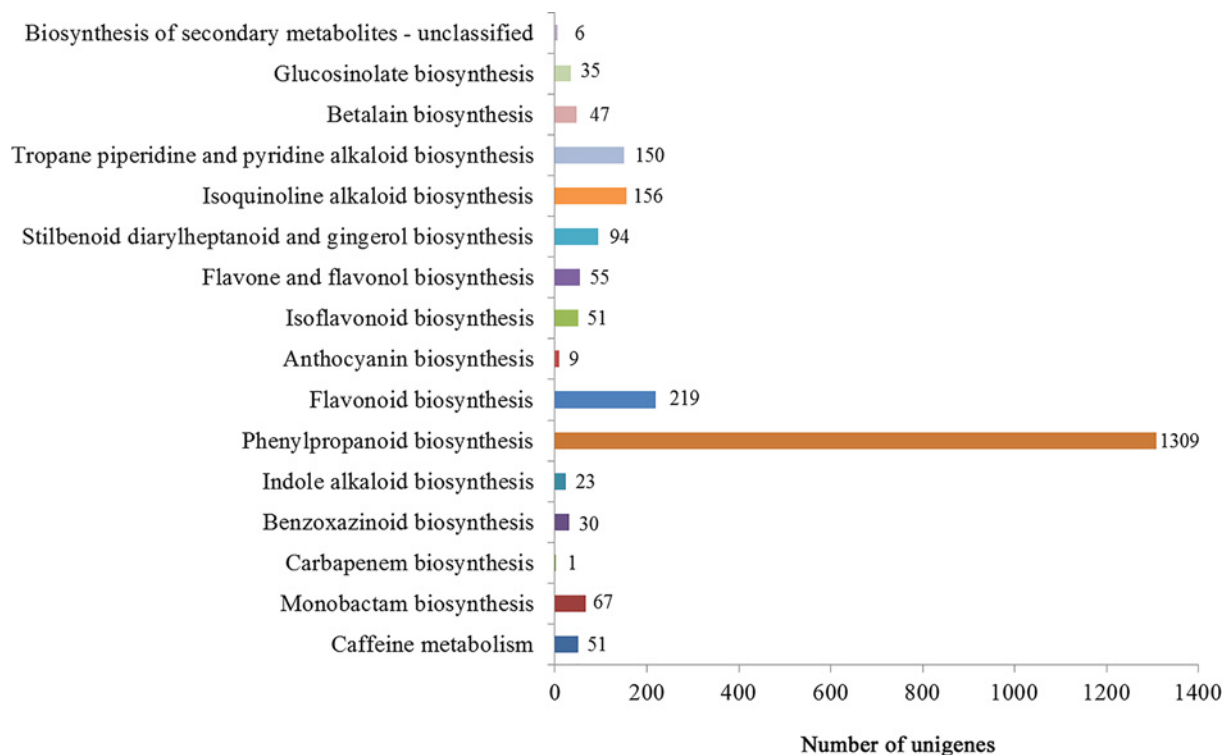
Figure 3. Pathway classifications for biosynthesis metabolism

Protein–protein interaction network construction

The protein–protein interaction network between the key enzymes was predicted using STRING database (https://string-db.org/) to identify functional protein association networks. From the analysis of 107 key enzyme genes using the STRING website, seven key enzymes involved in lignin biosynthesis were annotated as node proteins: PAL, C4H, 4CL, CCR, CAD, COMT, and F5H (Supplementary Table S1). The average node degree was 9.85 and the PPI enrichment P-value was $<1.0 \times 10^{-16}$. This result indicated a strong association among node proteins in the PPI networks. These key enzymes may interact with each other during lignin synthesis and play crucial roles in the regulation of lignin biosynthesis (Supplementary Figure S8).

Analysis of differentially expressed genes and specifically expressed genes in *Acorus tatarinowii*

In total, 68575 shared unigenes were identified in the two different tissue types of *A. tatarinowii* based on a Venn diagram analysis; 9997 and 51930 unigenes were only expressed in the leaf tissue and rhizome tissue, respectively (Figure 5A).

In the comparison of rhizome tissue with leaf tissue, 44879 unigenes were identified as significant differentially expressed genes (DEGs), which included 25602 up-regulated unigenes and 19277 down-regulated unigenes (Figure 5B). To determine the significantly enriched KEGG pathways, the 25602 up-regulated unigenes were mapped to the KEGG database. The four most common categories were 'phenylpropanoid biosynthesis' (288 genes), 'flavonoid biosynthesis' (36 genes), 'isoquinoline alkaloid biosynthesis' (23 genes), and 'stilbenoid, diarylheptanoid and gingerol biosynthesis' (21 genes). When these unigenes were mapped to the GO database, the most genes were enriched in 'integral component of membrane' (1952 genes), 'ATP binding' (973 genes), 'nucleus' (525 genes), and 'DNA binding' (344 genes) (Figure 5C, D).

The rhizome of *A. tatarinowii* is the main medicinal part of the plant. In the rhizome tissue, 51930 specifically expressed unigenes were identified; the significantly enriched KEGG pathways were 'phenylpropanoid biosynthesis' (468 unigenes), 'flavonoid biosynthesis' (80 unigenes), 'isoquinoline alkaloid biosynthesis' (61 unigenes), and 'tropane, piperidine and pyridine alkaloid biosynthesis' (55 unigenes). The significantly enriched GO terms were
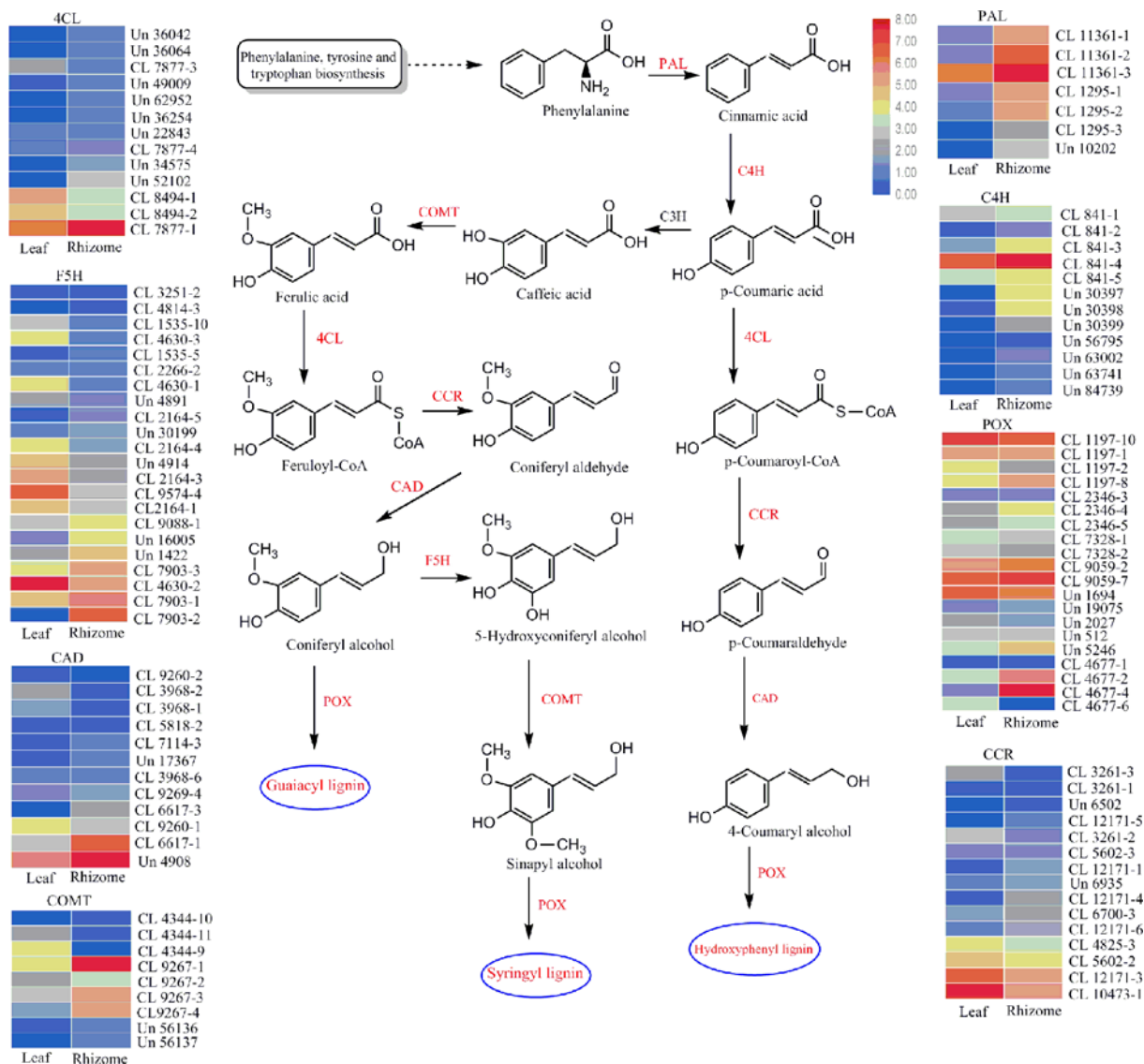
Figure 4. Proposed lignin biosynthetic pathways in *Acorus tatarinowii*

The relative expression of the unigenes encoding enzymes from each step, presented as heatmaps. Red and blue represent high and low expression levels, respectively. Non-dashed line arrows represent the identified enzymatic reactions, and dashed line arrows represent multiple enzymatic reactions through multiple steps.

'integral component of membrane' (3329 unigenes), 'ATP binding' (1958 unigenes), 'nucleus' (1226 unigenes), and 'cytoplasm' (904 unigenes) (Supplementary Figure S9 and S10).

Identification of transcriptome factors involved in lignin biosynthesis in *Acorus tatarinowii*

In the *A. tatarinowii* transcriptome database, 2426 unigenes were mapped to 57 transcriptome factor (TF) families. Mostly, unigenes in the MYB family were enriched (249 unigenes), followed by bHLH (179 unigenes), AP2-EREBP (170 unigenes), C2H2 (146 unigenes), C3H (135 unigenes), WRKY (130 unigenes), NAC (120 unigenes), and FAR1 (118 unigenes) (Figure 6A). From the KEGG pathway analysis of all TFs, 40 TFs related to the biosynthesis of other secondary metabolites were identified. Among these, 10 unigenes were enriched in the Trihelix family, and in the MYB (6 unigenes), zf-HD (6 unigenes), FAR1 (3 unigenes), C2H2 (3 unigenes), and GRAS (3 unigenes) families (Figure 6B). Sixteen unigenes involved in lignin biosynthesis were identified, including Trihelix (8 unigenes), MYB
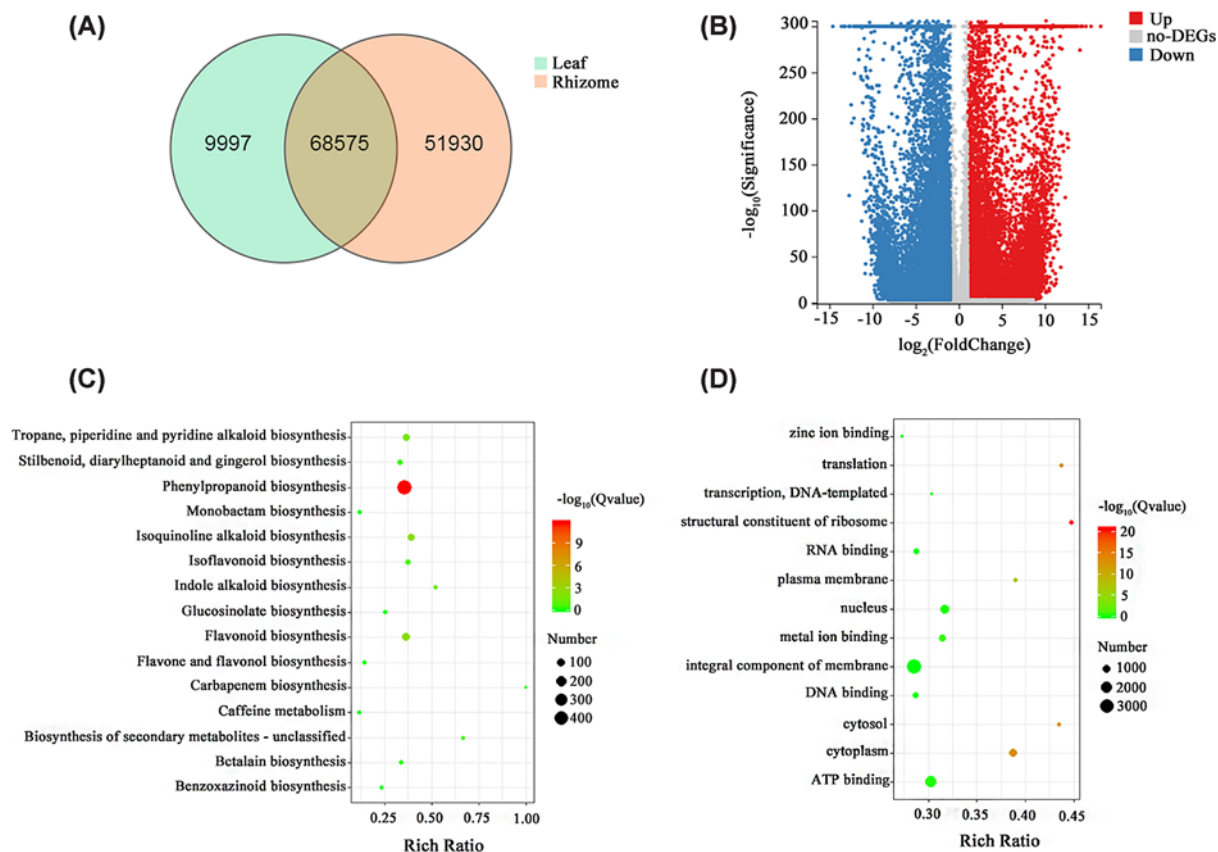
Figure 5. Unigenes expressed in *Acorus tatarinowii* tissues

(**A**) Venn diagram of unigenes expressed in leaf and rhizome tissues. (**B**) The number of DEGs in the *A. tatarinowii* leaf and rhizome tissues. A summary of unigene numbers that were up-regulated and down-regulated between the two samples is shown. The DEGs with higher expression levels in the rhizome tissue compared with the leaf tissue were defined as up-regulated, whereas those with lower expression levels in rhizome tissue were defined as down-regulated. (**C**) KEGG classification of the up-regulated unigenes in rhizome tissue compared with leaf tissue. (**D**) GO enrichment of the up-regulated unigenes in rhizome tissue compared with leaf tissue. Abbreviation: DEG, differentially expressed gene.

(4 unigenes), C2C2-Dof (1 unigene), OFP (1 unigene), zf-HD (1 unigene), and bHLH (1 unigene) (Figure 6C), and these TFs related to lignin biosynthesis regulated the same key enzyme (POX) (Supplementary Table S2).

Structural characteristics of peroxidase involved in lignin biosynthesis

The alignment of the four peroxidase amino acid sequences revealed that they shared a very high sequence identity (92.67%, Figure 7C). A 3D structural model of peroxidase (CL4977.Contig4, Range P29-N330, Coverage 0.85) was constructed based on the crystal structure of peroxidase from *Sorghum bicolor* (PDB ID: 5aog, 64.57% sequence identity) by using SWISS-MODEL (https://swissmodel.expasy.org/) and PyMOL software. The structure of peroxidase (CL4977.Contig4) is presented in Figure 7. The predicted tertiary structure indicated that the peroxidase contained two different domains, comprising 11 α-helixes. 'AAGLLRLHFHDC' (A66–C77) is the peroxidase active site signature (colored in pink, Figure 7B), and 'DLVALSGGHTI' (D196–I206) is the peroxidase proximal heme-ligand signature (colored in blue, Figure 7B). Moreover, the peroxidases contained four conserved disulfide bonds (C211–C238, C77–C82, C131–C326, and C125–C44, Figure 7A).

qRT-PCR validation of gene expression of key enzymes

The gene expression levels of POX (CL4977-4), F5H (CL7903-2), CAD (CL6617-1), COMT (CL9267-1), and PAL (CL11361-1 and CL11361-3) were evaluated in the rhizome and leaf tissues of *A. tatarinowii* by using qRT-PCR. The
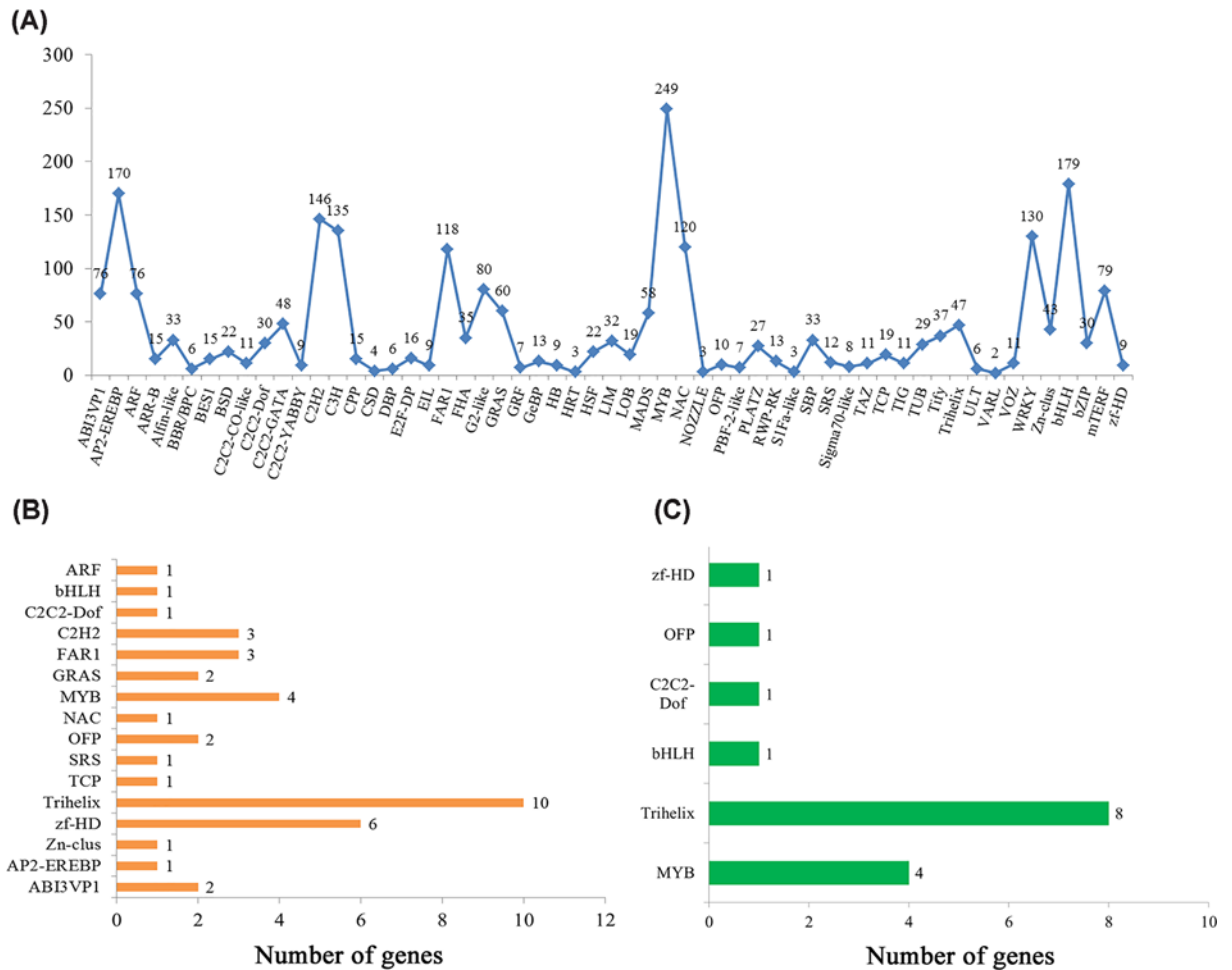
Figure 6. Analysis of transcription factors

(**A**) Classification of transcription factor families. (**B**) Classification of transcription factors involved in the biosynthesis of other secondary metabolites. (**C**) Classification of transcription factors involved in lignin biosynthesis.

relative expression of these genes was significantly higher in rhizome tissue than in leaf tissue, which was consistent with the transcriptome data (Figure 8).

Discussion

A. tatarinowii is a well-known traditional medicinal plant [21]: its sedative, digestive, analgesic, diuretic, and antifungal effects are recorded in the Chinese Pharmacopoeia [22]. Even though lignins are considered to be the major bioactive components of *A. tatarinowii*, there are comparatively few studies relating to lignins [23]. To identify the unigenes involved in the lignin biosynthesis pathway, we constructed a large number of gene sequences to help identify more key enzymes to assist research of lignin biosynthesis in *A. tatarinowii*. Based on the *A. tatarinowii* transcriptome database, 141777 unigenes were uncovered, with an average length of 1108 nt and an N50 length of 2062 nt; 51352 unigenes (36.22%) were longer than 1000 nt. Furthermore, the relative gene expression of POX, F5H, CAD, COMT, and PAL was examined using qRT-PCR, and found to be consistent with our transcriptional data. These results confirmed the high quality of the transcriptome assembly and the accuracy of our *A. tatarinowii* transcriptome database. The availability of *A. tatarinowii* transcriptome data will support the study of biological metabolism and functional genes of other homologous species.

When the DEGs were mapped to KEGG pathways, the expression level of 25602 unigenes in rhizome tissues was significantly up-regulated, these were mainly associated with the pathways of 'phenylpropanoid biosynthesis,' 'flavonoid biosynthesis,' and 'isoquinoline alkaloid biosynthesis.' A total of 51930 unigenes were identified as specifically expressed in rhizome tissue, and were mainly associated with the following KEGG pathways: 'phenylpropanoid
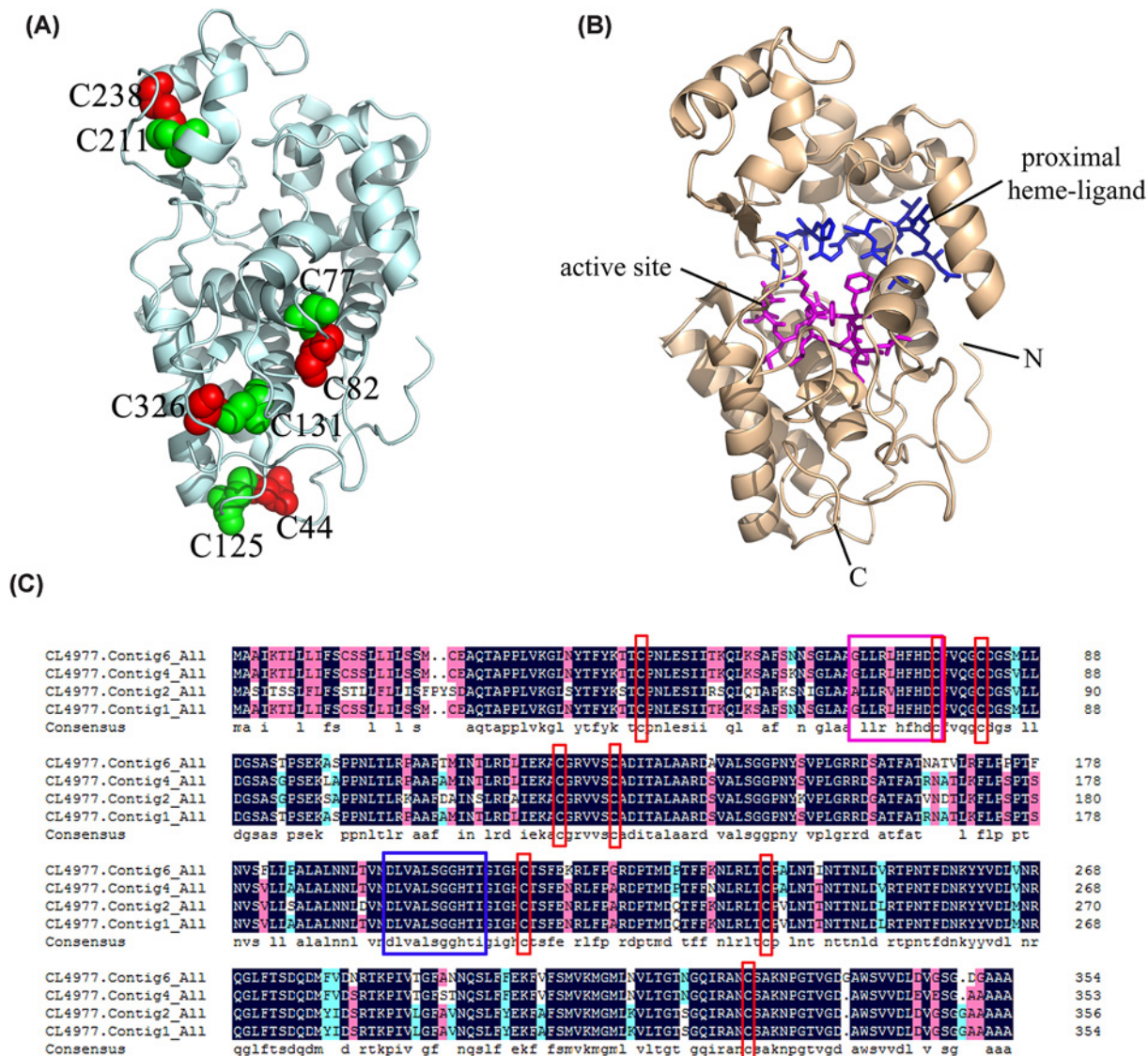
Figure 7. Spatial structure model of peroxidase
(**A**) Four disulfide bonds (C211–C238, C77–C82, C131–C326, and C125–C44) are shown as spheres and colored in red or green. (**B**) The active site and proximal heme-ligand of the peroxidases are shown in pink and blue, respectively). (**C**) Alignment of four amino acid sequences of peroxidases in *A. tatarinowii*. Black and pink show identical and similar amino acids, respectively. The red box highlights four conserved disulfide bonds, the blue box highlights the proximal heme-ligand signature, and the pink box highlights the active site.

biosynthesis,' 'flavonoid biosynthesis,' and 'isoquinoline alkaloid biosynthesis.' These data may confirm the medicinal value of *A. tatarinowii* from rhizome tissues at the molecular level. KEGG annotations revealed eight key enzymes related to lignin biosynthesis, namely PAL, C4H, COMT, 4CL, CCR, CAD, POX, and F5H. These key enzymes play important roles in the molecular mechanisms of lignin synthesis in *A. tatarinowii*. Compared with the leaf tissue, unigenes encoding the PAL, C4H, COMT, CAD, and POX enzymes were more highly expressed in rhizome tissue. COMT, CAD, and POX have been reported to play important roles in the process of lignin biosynthesis, and revealed that the expression levels of key enzymes genes may regulate the synthesis of lignin [24–26].

In the transcriptome data of *A. tatarinowii*, 2426 unigenes were successfully annotated to 57 transcription factor families. Mostly, unigenes in the MYB family were enriched. Studies have reported that MYB transcription factors are involved in lignin biosynthesis in many plant species, including *Arabidopsis*, *Loquat undergoes*, and *poplar*, and play a role in the formation of secondary cell walls [27–29]. Through KEGG pathway analysis of TFs, 16 TFs involved
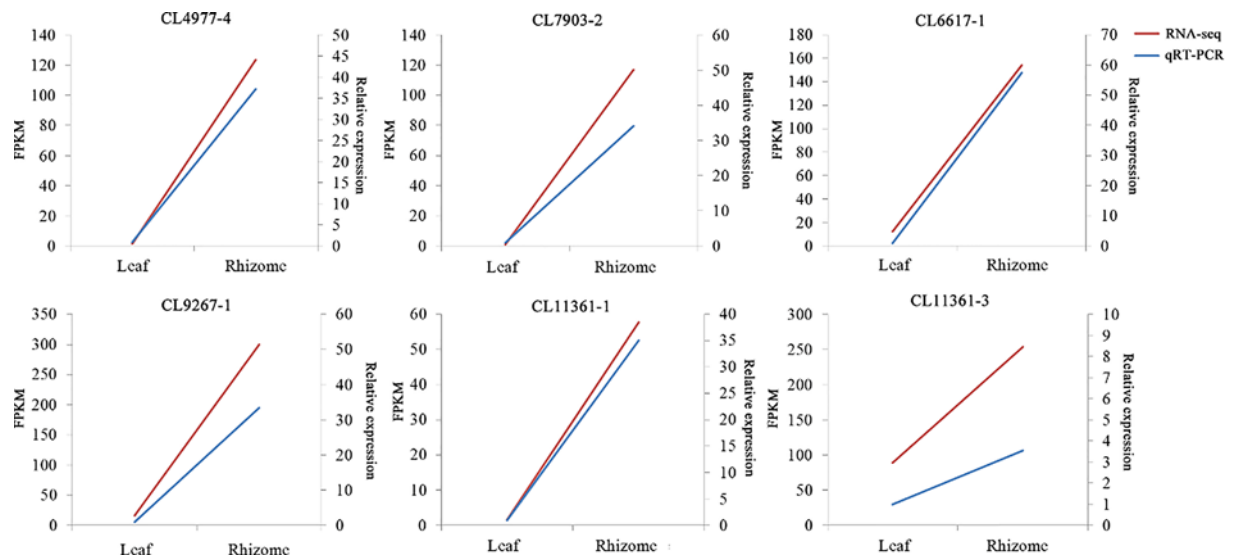
Figure 8. qRT-PCR and FPKM-value analysis of six genes encoding enzymes involved in lignin biosynthesis in *Acorus tatarinowii*

The expression of POX (CL4977-4), F5H (CL7903-2), CAD (CL6617-1), COMT (CL9267-1), and PALs (CL11361-1 and CL11361-3) was determined relative to the expression of the actin gene (Unigene12762); three technical replicates of each sample were analyzed.

in lignin biosynthesis in *A. tatarinowii* were identified, and these TFs regulated the same key enzyme (peroxidase). These results indicate that peroxidase may play a key role in the regulation of lignin biosynthesis in *A. tatarinowii*. Hence, the TFs regulating peroxidase are important candidates for future studies of the regulation of lignin biosynthesis in *A. tatarinowii*. However, RNA-Seq is not appropriate for the detection of regulatory genes with low expression levels [30,31]. For this reason, many TFs are hard to detect, which has obstructed research into their functions; however, they are still very important, because a small increase in the expression of TFs may have a dramatic effect.

The four sequences of peroxidase in *A. tatarinowii* had well-conserved amino acid sequences, spatial structures, active sites, and proximal heme-ligand signatures. Moreover, the peroxidase had eight conserved cysteine residues, which serve as the site of oxidation by peroxides. Other research has shown that the peroxidases active site (AAGLL-RLHFHDC) and proximal heme-ligand (DLVALSGGHTI) played a catalytic role by forming covalent bonds with Ca^{2+}, Na^+, and Fe^{2+}; of these, Ca^{2+} and Na^+ are the proximal and distal conserved binding sites. The incorporation of metal ions plays a regulatory role in the catalytic process of peroxidase [32,33].

Lignin is a promising material for pharmaceutical and biomedical applications owing to its biocompatibility, low cytotoxicity, antioxidant and antimicrobial properties. In medicine, it can be used as an auxiliary material to improve the oral bioavailability of the drug [34]. In addition, lignin can increase the stability of the antimicrobial properties of wound dressings [35,36]. Lignin-based hydrogels, which combine lignin with other polymers, have been used for gene delivery [37], drug delivery [38], and tissue engineering [39,40].

This is the first transcriptome sequencing study of two tissues of *A. tatarinowii* using RNA-Seq technology. The obtained data have provided comprehensive information on the gene sequences and expression level of genes in *A. tatarinowii*, and will support more detailed studies of the molecular mechanism of lignin biosynthesis.

Materials and methods
Plant material and RNA extraction

Three plants of *A. tatarinowii*, which were cultivated in the natural environment of the pharmaceutical garden of Anhui University of Chinese Medicine, were harvested in April 2019. Fresh plants were quickly cleaned with sterile water and wiped clean with filter paper. The tissues (leaf and rhizome) were separately collected into 50-mL centrifuge tubes, frozen quickly in liquid nitrogen for 6 h, and then stored at −80 °C for RNA extraction. An independent experiment was conducted for each *A. tatarinowii* tissue, with three biological replicates used for each treatment.

PORTLAND PRESS

cDNA library construction and sequencing (RNA-Seq)

Each plant tissue sample of RNA was extracted by using the RNA Plant Kit (Aidlab Biotech, Beijing, China) in accordance with the manufacturer's protocol. The mRNA from each sample was isolated from the total RNA by using beads with oligo(dT), and lysed into short fragments using fragmentation buffer. Random N6 primers were used to reverse-transcribe the short fragments of mRNA obtained; the first-strand cDNA was synthesized as a template, and two-strand cDNA was synthesized to form double-stranded cDNA by using dNTPs, RNase H, and DNA polymerase I. The obtained short cDNA fragment was ligated to the adapter, and the entire fragment was selected for PCR amplification. Finally, the quantification of cDNA library per tissue was detected, and cDNA library of plant tissues was constructed by BGISEQ-500 [41].

Transcriptome *de novo* assembly and unigene functional annotation

To obtain high-quality clean data, raw reads were filtered to remove low-quality reads with greater than 20% Q-value < 20 bases, ambiguous reads containing >5% unknown base (N), and for non-coding RNA by SOAPnuke and Trimmomatic software. Full-length transcripts were *de novo* assembled through a broad range of expression levels and sequencing depths by using Trinity software (v2.0.6), default parameters (K-mer = 25, group pairs distance = 400) [42]. These contigs were then further processed with Trinity to remove the redundant transcripts by using TGICL (v2.0.6) software, and finally, to generate unigenes.

Unigenes were mapped to several public databases using BLAST (version 2.2.23, e-value $\leq 1 \times 10^{-5}$), including Nt (ftp://ftp.ncbi.nlm.nih.gov/blast/db), Nr (ftp://ftp.ncbi.nlm.nih.gov/blast/db), COG (https://www.ncbi.nlm.nih.gov/COG/), KEGG (http://www.genome.jp/kegg), and Pfam (http://pfam.xfam.org). According to the Nr annotation, the GO (http://geneontology.org) annotation of unigenes was obtained using the Blast2GO program (https://www.blast2go.com) [43].

MAPMAN annotation and protein–protein interaction network construction

MAPMAN (http://mapman.gabipd.org/) is a tool for plant metabolism distribution research that performs large gene data enriched to diagrams of plant metabolic pathways [44]. All *A. tatarinowii* gene sequences were annotated by the MAPMAN database (http://www.plabipd.de/portal/mercator-sequence-annotation) to form independent species mapping. The expression level of unigenes in *A. tatarinowii* rhizome tissue was analyzed using MAPMAN software.

Based on *Arabidopsis thaliana* genomic data, the STRING online tool was used to analyze the genes encoding the key enzyme involved in the biosynthesis of lignin metabolism, and to predict the interactions between them. The protein–protein interactions of key enzymes were selected with score (median confidence) >0.4, and the node proteins were selected based on their association with other proteins. In a PPI network, an interactive network with a higher degree of association with other interactive networks plays a more important role [45].

Analysis of DEGs

RNA-seq has emerged as a powerful technology to measure gene expression and tissue specificity at the whole-transcriptome level. In the present study, the transcriptome data were evaluated to determine the differences in gene expression in the leaf and rhizome of *A. tatarinowii*. We calculated FPKM values to quantify the expression levels of each unigene. From the analysis of the DEGs, high-quality reads were assembled on the final transcriptome using Bowtie2 (version 2.2.5). Differential gene expression in two tissues was determined using the PossionDis method, which was performed as described: FC (fold change) ≥ 2.00 and FDR ≤ 0.001 [46]. Moreover, functional enrichment of the DEGs was annotated by the GO functional database by using GO-Term Finder software. KEGG pathway annotation was applied to significant DEGs.

Identification of TFs

In order to identify the transcription factor family represented in the *A. tatarinowii* transcriptome, Getorf (http://emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html) was applied to detect the open reading frames (ORFs) of the unigenes, and then to align them with the protein domains of transcription factors. In our transcriptome, unigene-encoded protein sequences were probed against and matched with all the transcription factor protein sequences in the Plant transcription factor database (PlantTFDB). Finally, the transcription factors were successfully predicted in the *A. tatarinowii* transcriptome [47].

Analysis of the structural characteristics of peroxidase

Four amino acid sequences of Pox were submitted to the SmartBlast in NCBI (http://www.ncbi.nlm.nih.gov/) to search for enzyme sequences of homologous species. The alignment of four ORFs of POX (peroxidase) was compared using DNAMAN software. SWISS-MODEL (https://swissmodel.expasy.org/) and PyMOL were applied to analyze the 3D structure model of the peroxidase protein [48].

qRT-PCR validation

The key enzyme genes involved in lignin synthesis were determined by quantitative real-time PCR in accordance with the instructions of Power SYBR® Green PCR Master Mix (Roche) and Quantstudio multiplex real-time PCR instrument (Life Technologies, U.S.A.). Based on transcriptome sequences, special PCR primers were designed for different genes using Primer Premier 6.0 and Beacon designer 7.8 software (Supplementary Table S3). We used 20 µl of reaction mixture, including 1 µL of diluted cDNA, 10 µl of Power SYBR® Green Master Forward Primer, 8 µL of SDW, 0.5 µL of forward primer (10 µM), and 0.5 µL of reverse primer (10 µM). The real-time PCR program used was 95 °C for 1 min, 40 cycles of 95 °C for 15 s, and 63 °C for 25 s. For each sample, three technical replicates of the qRT-PCR analyses were repeated, and the mean of the three replicates was used as the PCR result. The actin gene (Unigene12762) served as the internal control. The relative expression levels of the selected genes were calculated using the $2^{(-\Delta\Delta C_t)}$ method [49].

Data Availability

The RNA-seq datasets from two *A. tatarinowii* tissues were deposited in the NCBI Sequence Read Archive (SRA) database (accession: SRP213064).

Competing Interests

The authors declare that there are no competing interests associated with this paper.

CRediT Author Contribution

Shengxiang Zhang: Data curation, Writing—original draft. **Liqiang Zhao:** Data curation, Writing—original draft. **Chunmiao Shan:** Data curation. **Yuanyuan Shi:** Data curation. **Kelong Ma:** Funding acquisition, Data curation. **Jiawen Wu:** Funding acquisition, Validation.

Abbreviations

CAD, cinnamyl-alcohol dehydrogenase; CCR, cinnamyl-CoA reductase; COMT, caffeic acid 3-O-methyltransferase; C4H, cinnamate 4-hydroxylase; DEG, differentially expressed gene; FDR, False Discovery Rate; FPKM, fragments per kilobase of transcript per million mapped read; F5H, ferulate-5-hydroxylase; ORF, open reading frame; PAL, phenylalanine ammonia-lyase; POX, peroxidase; PPI, protein–protein interaction; qRT-PCR, real-time quantitative Polymerase Chain Reaction; SDW, sterile distilled deionized water; TF, transcriptome factor; 4CL, 4-coumaroyl CoA-ligase.

References

1　Shuang, L., Shu-Song, Y., Hong-Hua, W., Yan-Ting, L., Peng-Zhi, D., Yan, Z. et al. (2015) A novel sesquiterpene and three new phenolic compounds from the rhizomes of Acorus tatarinowii Schott. *Bioorg. Med. Chem. Lett.* **25**, 4214–4218, https://doi.org/10.1016/j.bmcl.2015.08.001

2　Zhang, F., Qi, P., Xue, R., Li, Z., Zhu, K., Wan, P. et al. (2015) Qualitative and quantitative analysis of the major constituents in Acorus tatarinowii Schott by HPLC/ESI-QTOF-MS/MS. *Biomed. Chromatogr.* **29**, 890–901, https://doi.org/10.1002/bmc.3371

3　Feng, Z., Peng, Q., Rui, X. et al. (2015) Qualitative and quantitative analysis of the major constituents in Schott by HPLC/ESI-QTOF-MS/MS. *Biomed. Chromatogr.* **29**, 890–901, https://doi.org/10.1002/bmc.3371

PORTLAND
PRESS

4   Liao, W.P., Chen, L.Y., Sun, W.W., Gao, M.M., Su, T. and Yang, S.Q. (2010) Study of antiepileptic effect of extracts from Acorus tatarinowii Schott. *Epilepsia* **46**, 21–24, https://doi.org/10.1111/j.0013-9580.2005.461007.x

5   Yuan, H., Mo, Y., Ping, L., Mu, L.H. and Hong, W. (2009) Effect of Acorus tatarinowii schott on ultrastructure and permeability of blood-brain barrier. *China J. Chin. Mater. Med.* **34**, 349–351

6   Boerjan, W., Ralph, J. and Baucher, M. (2003) Lignin biosynthesis. *Annu. Rev. Plant Biol.* **54**, 519–546, https://doi.org/10.1146/annurev.arplant.54.031902.134938

7   Vanholme, R., Demedts, B., Morreel, K., Ralph, J. and Boerjan, W. (2010) Lignin biosynthesis and structure. *Plant Physiol.* **153**, 895–905, https://doi.org/10.1104/pp.110.155119

8   Moerschbacher, B.M., Noll, U.M., Flott, B.E. and Reisener, H.J. (1988) Lignin biosynthetic enzymes in stem rust infected, resistant and susceptible near-isogenic wheat lines. *Physiol. Mol. Plant Pathol.* **33**, 33–46, https://doi.org/10.1016/0885-5765(88)90041-0

9   Sun, Q., Liu, X., Yang, J., Liu, W., Du, Q., Wang, H. et al. (2018) microRNA528 affects lodging resistance of maize by regulating lignin biosynthesis under nitrogen-luxury conditions. *Mol. Plant* **11**, 806–814, https://doi.org/10.1016/j.molp.2018.03.013

10 Cesarino, I., Araújo, P., Júnior, A.P.D. and Mazzafera, P. (2012) An overview of lignin metabolism and its effect on biomass recalcitrance. *Rev. Bras. Bot.* **35**, 303–311, https://doi.org/10.1590/S0100-84042012000400003

11 Morrison, T.A., Kessler, J.R., Hatfield, R.D. and Buxton, D.R. (2010) Activity of two lignin biosynthesis enzymes during development of a maize internode. *J. Sci. Food Agric.* **65**, 133–139, https://doi.org/10.1002/jsfa.2740650202

12 Ma, Q.H. and Xu, Y. (2008) Characterization of a caffeic acid 3-O-methyltransferase from wheat and its function in lignin biosynthesis. *Biochimie* **90**, 515–524, https://doi.org/10.1016/j.biochi.2007.09.016

13 Poke, F.S., Vaillancourt, R.E., Elliott, R.C. and Reid, J.B. (2003) Sequence variation in two lignin biosynthesis genes, cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase 2 (CAD2). *Mol. Breed.* **12**, 107–118, https://doi.org/10.1023/A:1026006828721

14 Humphreys, J.M., Hemm, M.R. and Chapple, C. (1999) New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10045–10050, https://doi.org/10.1073/pnas.96.18.10045

15 Mansell, R.L., Gross, G.G., Stöckigt, J., Franke, H. and Zenk, M.H. (1974) Purification and properties of cinnamyl alcohol dehydrogenase from higher plants involved in lignin biosynthesis. *Phytochemistry* **13**, 2427–2435, https://doi.org/10.1016/S0031-9422(00)86917-4

16 Daniel, S., Paniz-Mondolfi, A. and Monzon, F.A. (2018) Transcriptome sequencing (RNA-Seq). *Methods Mol. Biol.* **1754**, 15–27, https://doi.org/10.1007/978-1-4939-7717-8˙2

17 Chen, X.J. (2014) De novo characterization of the Rehmannia glutinosa leaf transcriptome and analysis of gene expression associated with replanting disease. *Mol. Breed.* **34**, 905–915, https://doi.org/10.1007/s11032-014-0084-5

18 Cherukupalli, N., Divate, M., Mittapelli, S.R., Khareedu, V.R. and Vudem, D.R. (2016) De novo assembly of leaf transcriptome in the medicinal plant Andrographis paniculata. *Front. Plant Sci.* **7**, 1203, https://doi.org/10.3389/fpls.2016.01203

19 Tsaballa, A., Nikolaidis, A., Trikka, F., Ignea, C., Kampranis, S.C., Makris, A.M. et al. (2015) Use of the de novo transcriptome analysis of silver-leaf nightshade (Solanum elaeagnifolium) to identify gene expression changes associated with wounding and terpene biosynthesis. *BMC Genomics* **16**, 504–519, https://doi.org/10.1186/s12864-015-1738-3

20 Wang, C., Zhu, J., Liu, M., Yang, Q., Wu, J. and Li, Z. (2018) De novo sequencing and transcriptome assembly of Arisaema heterophyllum Blume and identification of genes involved in isoflavonoid biosynthesis. *Sci. Rep.* **8**, 17643–17655, https://doi.org/10.1038/s41598-018-35664-1

21 Xiao-Xing, H.E. and Chang-Yu, D. (2008) The research progress on the role of Acorus Tatarinowii Schott for the central nerve system. *Chin. Arch. Tradit. Chin. Med.* **4**, 796–798

22 Tong, X.G., Zhou, L.L., Wang, Y.H., Xia, C., Wang, Y., Liang, M. et al. (2010) Acortatarins A and B, two novel antioxidative spiroalkaloids with a naturally unusual morpholine motif from Acorus tatarinowii. *Cheminform* **41**, 1844–1847, https://doi.org/10.1002/chin.201035205

23 Lu, Y., Xue, Y., Chen, S., Zhu, H., Zhang, J., Li, X.N. et al. (2016) Antioxidant lignans and neolignans from Acorus tatarinowii. *Sci. Rep.* **6**, 22909–22919, https://doi.org/10.1038/srep22909

24 Kawasaki, T., Koita, H., Nakatsubo, T., Hasegawa, K., Wakabayashi, K., Takahashi, H. et al. (2006) Cinnamoyl-CoA Reductase, a key enzyme in lignin biosynthesis, is an effector of small GTPase Rac in defense signaling in rice. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 230–235, https://doi.org/10.1073/pnas.0509875103

25 Hatfield, R.D. (2003) Structural and compositional modifications in lignin of transgenic alfalfa down-regulated in caffeic acid 3-O-methyltransferase and caffeoyl coenzyme A 3-O-methyltransferase. *Phytochemistry* **62**, 53–65, https://doi.org/10.1016/S0031-9422(02)00434-X

26 Janusz, G., Kucharzyk, K.H., Pawlik, A., Staszczak, M. and Paszczynski, A.J. (2013) Fungal laccase, manganese peroxidase and lignin peroxidase: Gene expression and regulation. *Enzyme Microb. Technol.* **52**, 1–12, https://doi.org/10.1016/j.enzmictec.2012.10.003

27 Zhou, J., Lee, C., Zhong, R. and Ye, Z.H. (2009) MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in Arabidopsis. *Plant Cell* **21**, 248–266, https://doi.org/10.1105/tpc.108.063321

28 Li, C., Wang, X., Lu, W., Liu, R., Tian, Q., Sun, Y. et al. (2014) A poplar R2R3-MYB transcription factor, PtrMYB152, is involved in regulation of lignin biosynthesis during secondary cell wall formation. *Plant Cell Tissue Organ Culture* **119**, 553–563, https://doi.org/10.1007/s11240-014-0555-8

29 Jing, Z., Hang, G., Chen, Z., Xian, L., Donald, G., Chen, K.S. et al. (2016) EjODO1, a MYB transcription factor, regulating lignin biosynthesis in developing Loquat (Eriobotrya japonica) fruit. *Front. Plant Sci.* **7**, 1360

30 Zhang, Z., Harland, R.M., Michal, J.J. et al. (2015) Whole transcriptome analysis with sequencing: methods, challenges and potential solutions. *Cell. Mol. Life Sci.* **72**, 3425–3439, https://doi.org/10.1007/s00018-015-1934-y

31 Malone, J.H. and Oliver, B. (2011) Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* **9**, 1–9, https://doi.org/10.1186/1741-7007-9-34

32 Edwards, S.L., Raag, R., Wariishi, H., Gold, M.H. and Poulos, T.L. (1993) Crystal structure of lignin peroxidase. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 750–754, https://doi.org/10.1073/pnas.90.2.750

33 Nnamchi, C.I., Parkin, G., Efimov, I., Basran, J., Kwon, H., Svistunenko, D.A. et al. (2016) Structural and spectroscopic characterisation of a heme peroxidase from sorghum. *J. Biol. Inorg. Chem.* **21**, 63–70, https://doi.org/10.1007/s00775-015-1313-z

34 Zhou, Y., Han, Y., Li, G., Yang, S., Xiong, F. and Chu, F. (2019) Preparation of targeted lignin–based hollow nanoparticles for the delivery of doxorubicin. *Nanomaterials* **9**, 188–201, https://doi.org/10.3390/nano9020188

35 Eun-Sil, L., Young-O, K., Yu-Mi, H., Daun, L., Yeon, H.J., Jaewoo, K. et al. (2018) Antimicrobial properties of lignin-decorated thin multi-walled carbon nanotubes in poly(vinyl alcohol) nanocomposites. *Eur. Polym. J.* **105**, 79–84, https://doi.org/10.1016/j.eurpolymj.2018.05.014

36 Aadil, K.R., Mussatto, S.I., Jha, H. et al. (2018) Synthesis and characterization of silver nanoparticles loaded poly(vinyl alcohol)-lignin electrospun nanofibers and their antimicrobial activity. *Int. J. Biol. Macromol.* **120**, 763–767, https://doi.org/10.1016/j.ijbiomac.2018.08.109

37 Liu, X., Hui, Y., Zhang, Z., Diao, B. and Li, J. (2015) Functionalization of lignin through ATRP grafting of poly(2-dimethylaminoethyl methacrylate) for gene delivery. *Colloids Surf. B.* **125**, 230–237, https://doi.org/10.1016/j.colsurfb.2014.11.018

38 Larrañeta, E., Domínguez-Robles, J., Coogan, M., Heaney, E., Stewart, S.A., Thakur, R.S. et al. (2019) Poly(methyl vinyl ether-co-maleic acid) hydrogels containing cyclodextrins and Tween 85 for potential application as hydrophobic drug delivery systems. *Macromol. Res.* **11**, 2660–2666, https://doi.org/10.1007/s13233-019-7074-8

39 Berglund, L., Forsberg, F., Jonoobi, M. and Oksman, K. (2018) Promoted hydrogel formation of lignin-containing arabinoxylan aerogel using cellulose nanofibers as a functional biomaterial. *RSC Adv.* **8**, 38219–38228, https://doi.org/10.1039/C8RA08166B

40 Quraishi, S., Martins, M., Barros, A.A., Gurikov, P., Raman, S.P., Smirnova, I. et al. (2015) Novel non-cytotoxic alginate–lignin hybrid aerogels as scaffolds for tissue engineering. *J. Supercrit. Fluids* **105**, 1–8, https://doi.org/10.1016/j.supflu.2014.12.026

41 Fang, C., Zhong, H., Lin, Y., Chen, B., Han, M., Ren, H. et al. (2017) Assessment of the cPAS-based BGISEQ-500 platform for metagenomic sequencing. *Gigascience* **7**, 1–8

42 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I. et al. (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**, 644–652, https://doi.org/10.1038/nbt.1883

43 Ana, C., Stefan, G.T., Juan Miguel, G.G., Javier, T., Manuel, T. and Montserrat, R. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676, https://doi.org/10.1093/bioinformatics/bti610

44 Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P. et al. (2010) Mapman: a user riven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939, https://doi.org/10.1111/j.1365-313X.2004.02016.x

45 Zhang, L., Huang, Y., Zhuo, W.L. et al. (2017) Identification and characterization of biomarkers and their functions for Lapatinib-resistant breast cancer. *Med. Oncol.* **34**, 89–97, https://doi.org/10.1007/s12032-017-0953-y

46 Audic, S. and Claverie, J.M. (1997) The significance of digital gene expression profiles. *Genome Res.* **7**, 986–995, https://doi.org/10.1101/gr.7.10.986

47 Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277, https://doi.org/10.1016/S0168-9525(00)02024-2

48 Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T. et al. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252, https://doi.org/10.1093/nar/gku340

49 Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**, 402–408, https://doi.org/10.1006/meth.2001.1262