

Genome-wide identification of novel long non-coding RNAs in *Populus tomentosa* tension wood, opposite wood and normal wood xylem by RNA-seq

Jinhui Chen · Mingyang Quan · Deqiang Zhang

Received: 4 August 2014 / Accepted: 7 September 2014 / Published online: 18 September 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Increasing evidence shows that long non-coding RNAs (lncRNAs) function as important regulatory factors during plant development, but few reports have examined lncRNAs in trees. Here, we report our genome-scale identification and characterization of lncRNAs differentially expressed in the xylem of tension wood, opposite wood and normal wood in *Populus tomentosa*, by high-throughput RNA sequencing. We identified 1,377 putative lncRNAs by computational analysis, and expression and structure analyses showed that the lncRNAs had lower expression levels and shorter lengths than protein-coding transcripts in *Populus*. Of the 776 differentially expressed ($\log_2FC \geq 1$ or ≤ -1 , $FDR \leq 0.01$) lncRNAs, 389 could potentially target 1,151 genes via *trans*-regulatory effects. Functional annotation of these target genes demonstrated that they are involved in fundamental processes, and in specific mechanisms such as response to

stimuli. We also identified 16 target genes involved in wood formation, including cellulose and lignin biosynthesis, suggesting a potential role for lncRNAs in wood formation. In addition, three lncRNAs harbor precursors of four miRNAs, and 25 were potentially targeted by 44 miRNAs where a negative expression relationship between them was detected by qRT-PCR. Thus, a network of interactions among the lncRNAs, miRNAs and mRNAs was constructed, indicating widespread regulatory interactions between non-coding RNAs and mRNAs. Lastly, qRT-PCR validation confirmed the differential expression of these lncRNAs, and revealed that they have tissue-specific expression in *P. tomentosa*. This study presents the first global identification of lncRNAs and their potential functions in wood formation, providing a starting point for detailed dissection of the functions of lncRNAs in *Populus*.

Keywords Long non-coding RNAs · *Populus* · Tension wood · Wood formation

Electronic supplementary material The online version of this article (doi:10.1007/s00425-014-2168-1) contains supplementary material, which is available to authorized users.

J. Chen · M. Quan · D. Zhang (✉)
National Engineering Laboratory for Tree Breeding,
College of Biological Sciences and Technology,
Beijing Forestry University, No. 35, Qinghua East Road,
Beijing 100083, People's Republic of China
e-mail: DeqiangZhang@bjfu.edu.cn

J. Chen · M. Quan · D. Zhang
Key Laboratory of Genetics and Breeding in Forest Trees
and Ornamental Plants, College of Biological Sciences
and Technology, Beijing Forestry University, No. 35,
Qinghua East Road, Beijing 100083, People's Republic of China

Abbreviations

lncRNAs	Long non-coding RNAs
4CL	4-Coumarate-CoA ligase
CCOMT	Caffeoyl-CoA 3-O-methyltransferase
CESA	Cellulose synthase
CSLD4	Cellulose synthase-like protein D4
CW	Compression wood
GA2ox	Gibberellin 2-oxidase
GO	Gene ontology
NW	Normal wood
OW	Opposite wood
ppo	Polyphenol oxidase
TF	Transcription factor
TW	Tension wood

Introduction

Long non-coding RNAs (lncRNAs) are non-coding RNA transcripts of more than 200 nucleotides in length (Kapranov et al. 2007). Although most lncRNAs are transcribed by RNA polymerase II, some are transcribed by RNA polymerase III (Gibb et al. 2011). Most lncRNAs exist in the nucleus with only a few lncRNAs in the cytoplasm, unlike mRNAs, which mostly exist in the cytoplasm (Kapranov et al. 2007). In general, lncRNAs occur in very low abundance, and in specific patterns in tissue types and subcellular compartments (Wang et al. 2011b). lncRNAs are classified into i, o, u and x types, according to their genomic locations relative to their neighboring genes (Mercer et al. 2009; Ponting et al. 2009). For example, the class i intron lncRNAs, are entirely within a gene and the majority of the intron lncRNAs have similar expression patterns to their host genes (Sigova et al. 2013). Following the advent of the high-throughput sequencing, especially RNA sequencing (RNA-seq), emerging work has revealed that lncRNAs regulate gene expression in many biological processes, including in transcriptional, posttranscriptional and epigenetic mechanisms, in either *cis* or *trans* (Mercer et al. 2009; Yoon et al. 2013).

Since the first report of lncRNAs in humans by Lukiw et al. (1992), recent work has made significant advances in the identification and functional analysis of lncRNAs. Recently, many studies have reported that lncRNAs are associated with human diseases such as cancer, Alzheimer's disease, and heart diseases (Taft et al. 2010). The systematic identification of lncRNAs in zebrafish embryos revealed that lncRNAs are relatively short, with low, tissue-specific expression, characteristics shared with their mammalian counterparts (Pauli et al. 2012). In plants, lncRNAs have been identified in *Arabidopsis thaliana*, *Zea mays*, *Triticum aestivum* and *Oryza sativa*, indicating that lncRNA-mediated gene expression occurs widely in plants (Lu et al. 2012; Boerner and McGinnis 2012). Many studies have indicated that plant lncRNAs may be involved in the regulation of flowering, male sterility, metabolism, and biotic and abiotic stress responses (Zhang et al. 2013a). For instance, the lncRNA of *LDMAR* regulates photoperiod-sensitive male sterility in rice, and its low expression in photoperiod-sensitive male sterile rice Nongken 58S leads to male sterility under long-day conditions (Ding et al. 2012).

In addition, some studies reported that a subset of lncRNAs can be processed to small RNAs, indicating a potential regulatory function of lncRNAs (Jalali et al. 2012). The large intergenic lncRNA *H19* gene includes a miRNA, miR-675, which is involved in the developmental regulation of genes (Cai and Cullen 2007). lncRNAs could also potentially interact with other classes of non-coding

RNAs, including miRNAs (Jalali et al. 2013). Like mRNAs, lncRNAs could also be targeted and regulated by miRNAs (Zhang et al. 2013b). For example, analysis of lncRNAs by reverse transcription-polymerase chain reaction array revealed that miR-21 could suppress the lncRNA of *growth arrest-specific 5 (GAS5)* (Zhang et al. 2013b). lncRNAs have been described as a part of the “dark matter” of the genome and are annotated only rarely; particularly, identification and characterization of lncRNAs in trees remain to be addressed.

As perennial plants and complex multicellular organisms, trees experience many different biological conditions and undergo complex developmental processes during their lives. For example, wood formation (xylogenesis) is precisely regulated and distinguishes trees from herbaceous plants. Cellulose, lignin, and other biopolymers make up of the majority of woody tissues, providing a vital raw material for pulp and paper, biofuels, and solid wood products (Mellerowicz and Sundberg 2008). In addition to wood formation during normal growth, trees develop a special kind of wood, tension wood (TW) in branches that are bent or displaced by environmental conditions. In angiosperms, TW forms on the upper side of a bent stem or branch, and has decreased lignin and increased cellulose contents (Timell 1969; Joseleau et al. 2004). In the study and refinement of wood development, TW has been reported as an ideal model to identify the functions of genes involved in wood formation (Kwon 2007). To discover the molecular mechanisms that condition the different lignin and cellulose contents in opposite wood (OW), TW and normal wood (NW), TW models have been used to identify key genes, especially carbohydrate-active genes (Du et al. 2004; Nugroho et al. 2012). A microarray and metabolite analysis of TW-forming tissues in *Populus tremula* suggested that an increased C flux to cellulose was mainly caused by the up-regulation of sucrose synthase transcripts (Andersson-Gunnerås et al. 2006). Also, in pine species, *4-coumarate-CoA ligase (4CL)* expression is stimulated during the formation of compression wood (CW) (Zhang et al. 2013b). Additionally, transgenic suppression of *4CL* in *Pinus radiata* decreased lignin contents (Wagner et al. 2009). The genes involved in the synthesis of plant hormones such as auxin, ethylene and gibberellin have important roles in the formation of TW (Du et al. 2004; Nugroho et al. 2012). For example, gibberellin is required for the formation of TW and for stem gravitropism in seedlings of *Acacia mangium* (Nugroho et al. 2012). Recently, regulation by ncRNAs such as miRNAs has been implicated in the control of normal and TW formation in trees (Lu et al. 2005, 2013). Transgenic studies of miRNAs discovered that overexpression of *Ptr-miR397a* could reduce lignin content, revealing that the *Ptr-miR397a* is a

negative regulator of laccase genes affecting lignin content in *Populus trichocarpa* (Torr. & Gray) (Lu et al. 2013). However, whether lncRNAs have functions in wood formation, and how they work in wood formation, remains unknown.

To explore whether lncRNAs have a role in wood formation, we used RNA-seq to identify lncRNAs differentially expressed in the xylem tissues of NW, TW and OW in *Populus tomentosa* (Carr.). We also found lncRNAs that potentially target genes for *trans*-regulatory effects and lncRNAs confirmed by quantitative RT-PCR showed tissue-specific expression. Gene ontology (GO) analysis showed that these potential targets include important genes involved in response to stimulus, and in cellulose, lignin and gibberellin biosynthesis. In addition, we identified a network of interactions between ncRNAs, and mRNAs. Thus, our RNA-seq and computational analysis identified novel lncRNAs that have potentially important roles in the regulation of wood formation; moreover, the identification of the lncRNA–miRNA regulatory network places these ncRNAs in a functional context for gene regulation.

Materials and methods

Plant materials

Three 30-year-old individuals of *P. tomentosa* Carr. with well-developed branches in the national nursery of Guan Xian County (Shandong Province, China; N36°28'28",

E115°26'17") were used as biological replicates in our study. The largest branch from each tree was further selected for study in natural condition without any treatment. The mature xylem tissues of normal, opposite and tension wood (NW, OW and TW) were isolated from the same tree with a sharp chisel after removing the bark of the sampling area and the tissues were immediately frozen in liquid nitrogen, then stored at –80 °C for RNA extraction. The TW and OW were taken from the upper and lower sides of the same branch.

Total RNA isolation, RNA library construction, and Solexa sequencing

Total RNA from TW, OW and NW were isolated by a modified CTAB method (Chang et al. 1993) that used isopropanol instead of lithium chloride for RNA precipitation. DNase was applied to eliminate the DNA from the total RNA. The total RNA quality was monitored by NanoDrop ND-1000 (A260/A280: TW 2.11, OW 2.17 and NW 2.09) and Agilent Bioanalyzer 2100 (28S/18S: TW 1.6, OW 1.7 and NW 1.5; RIN: TW 8.3, OW 8.6 and NW 8.2) and was of sufficient quality (2100 RIN ≥7.0 and 28S/18S ≥0.7) for RNA-seq. Extracted RNA was used for RNA library construction. Qualified total RNA was purified with the RNeasy micro kit (Cat#74004, Qiagen) which only RNAs longer than 200 bp were purified, and then processed according to the TruSeq RNA Sample Preparation Guide to build a cDNA library. The sequencing of the cDNA library was carried out by the Shanghai Biotechnology Corporation

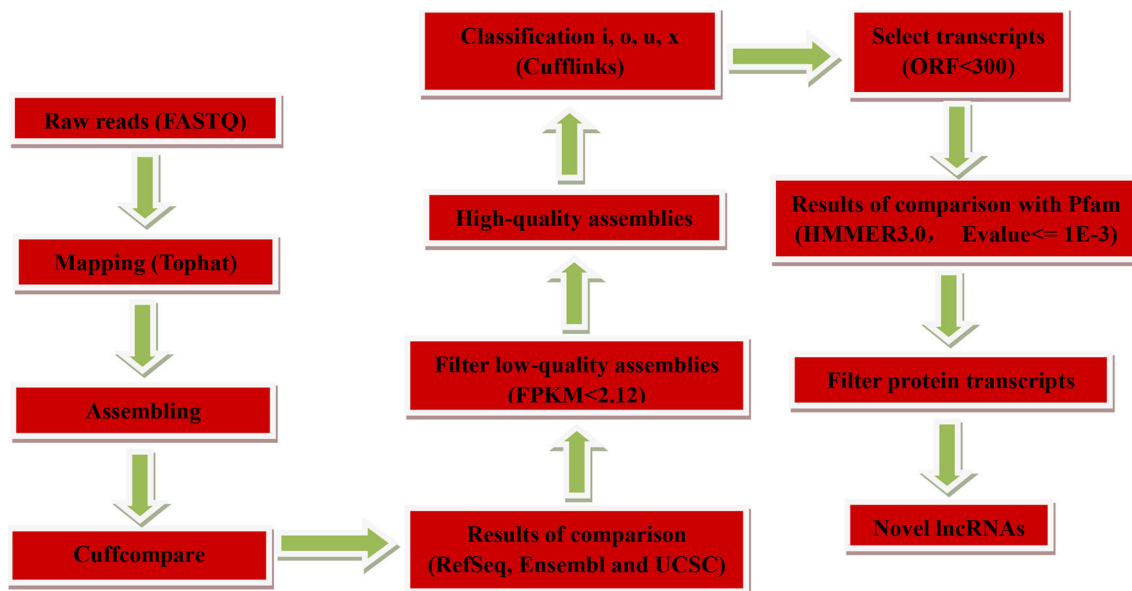


Fig. 1 Pipeline for predicting novel lncRNAs. Raw reads are first mapped onto the reference *Populus trichocarpa* genome. The initial assemblies are categorized by cuffcompare, compared with the combined gene annotations. The low-quality transcripts are then

filtered according to the optimum FPKM (2.12). The transcripts that encode protein were deleted from the remaining high-quality assemblies

(Shanghai, China) using Illumina HiSeq 2000, following the cBot User Guide and HiSeq 2000 User Guide with paired-end program.

Prediction of lncRNAs

Pre-processing and assembly of raw sequence data were performed using the CLC Genomics Workbench (version 4.9). Pre-processing included removal of the adapter sequences, low-quality (Q20 <20) sequences, and sequences shorter than 20 nucleotides. Clean reads were then aligned with the *Populus trichocarpa* genome (<http://www.phytozome.net/poplar>) (Tuskan et al. 2006) using Tophat (Trapnell et al. 2012) with two base mismatches allowed. The transcripts were assembled using cufflinks (Trapnell et al. 2012). The cuffcompare program (Trapnell et al. 2012) was used to merge the RefSeq-, Ensembl- and UCSC *Populus*-known genes into one set of gene annotations for comparison with the assembled transcripts. All the sequencing data were uploaded to SRA database with accession number SRP040531. The low-quality transcripts were then filtered according to the optimum FPKM (fragments per kilobase of exon per million fragments mapped) (2.12). The prediction of lncRNAs was

performed according to Sun et al. (2012) based on RNA-Seq data, as described in Fig. 1. The lncRNA classification was carried out according to their genomic location, referring to the neighboring genes, and lncRNAs were classified to 'i', 'u', 'o' and 'x' categories as defined by cufflinks (Roberts et al. 2011). Specifically, the 'i' category contained lncRNAs that fall entirely within an intron of a known gene. The 'u' category contained intergenic lncRNAs (lincRNAs). The 'o' category contained the lncRNAs that have generic exonic overlap with a known transcript, and the 'x' category contained lncRNAs that also have exonic overlap with known transcripts, but on the opposite strand.

The cufflink program (version 2.0.2) was used to calculate the expression of transcripts (Trapnell et al. 2010). The fold change (FC) lncRNA expression level equals the FPKM of TW divided FPKM of NW, and so on. Here, the FPKM(OW)/FPKM(NW) was defined as OW vs. NW, FPKM(TW)/FPKM(NW) as TW vs. NW and FPKM(OW)/FPKM(TW) as OW vs. TW. The differentially expressed lncRNAs were selected with a \log_2FC (OW vs. NW, TW vs. NW, or OW vs. TW) ≥ 1 or ≤ -1 , and with statistical significance ($P < 0.05$ and false discovery rate (FDR) < 0.01).

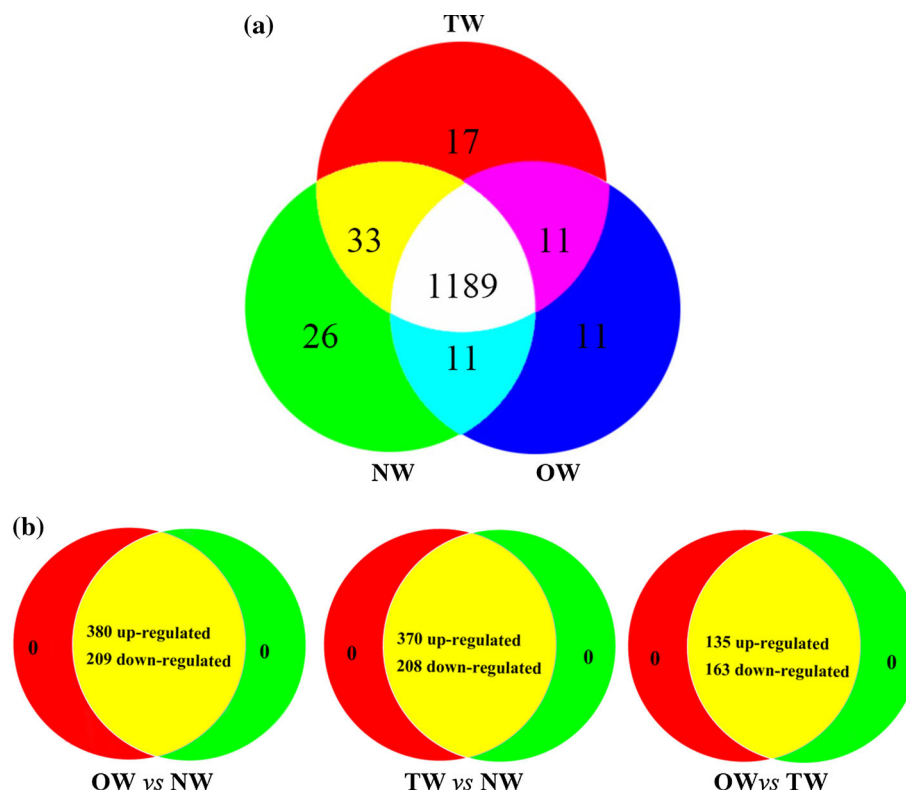


Fig. 2 Venn diagram of lncRNAs in TW, NW and OW. **a** Of the total 1,377 lncRNAs, 17, 11 and 26 were uniquely expressed in TW, OW and NW, respectively. Also, 1,189 lncRNAs were expressed in all three wood tissues. There are 11, 11 and 33 lncRNAs expressed in TW

and OW, OW and NW, and TW and NW, respectively. **b** The summary of differentially expressed lncRNAs in OW vs. NW and OW vs. TW

Genomic characterization of lncRNAs

The transcription factors (TFs) binding sites of differentially expressed lncRNAs were identified by BLAST against the Plant Transcription Factor Database v3.0 (<http://plantfdb.cbi.pku.edu.cn/>) (Jin et al. 2013) using the 2,000-bp (base pair) sequence upstream of the lncRNAs with *E* value <1e−20. To explore whether some lncRNAs may act as precursors of miRNAs, the published miRNAs in miRBase (Release 20, <http://www.mirbase.org/>) were aligned to the sequences of the novel lncRNAs, to find known miRNA precursors in the lncRNAs. We used the prediction software Mireap (<https://sourceforge.net/projects/mireap/>) to predict novel miRNAs by exploring the secondary structures, DL1 cleavage sites, and minimum free energies of the unannotated sRNA tags that could be

mapped to the *P. trichocarpa* reference genome and selected by basic criteria according to Meyers et al. (2008), with low free energy characteristic of stable hairpin structures (<−25 kcal/mol). The secondary structure was predicted with the Vienna RNA package RNAfold web (<http://rna.tbi.univie.ac.at/>). The target genes of miRNAs were predicted by psRNATarget (<http://plantgrn.noble.org/psRNATarget/>) (Dai and Zhao 2011) with expectation ≤3.

Prediction of the potential target genes of lncRNAs

Identification of the target genes of differentially expressed lncRNAs via *trans*-regulatory effects used sequence complementarity, as previously described (Han et al. 2012). The first step uses BLAST to select a target sequence that was complementary to the lncRNA, setting *E* value = 1e−5 and

Table 1 Putative transcription factors binding sites in the upstream 2,000-bp sequence of significantly differentially expressed lncRNAs

lncRNA ID	Protein ID	Protein length	<i>E</i> value	Score	Percent identity
TCONS_00000615	MIKC	187	6.00E−14	32	96.88
TCONS_00003815	bZIP	269	2.00E−130	189	97.4
TCONS_00005875	C3H	220	3.00E−101	177	65.87
TCONS_00006471	bHLH	629	4.00E−33	64	100
TCONS_00008831	EIL	659	0	504	100
TCONS_00008832	EIL	659	0	529	100
TCONS_00013357	WRKY	268	1.00E−65	104	100
TCONS_00014982	NAC	171	3.00E−48	79	100
TCONS_00018254	NAC	1,524	4.00E−78	144	75.29
TCONS_00019603	NAC	1,524	5.00E−69	132	75
TCONS_00027324	FAR1	714	7.00E−75	138	71.43
TCONS_00030474	HB-other	340	6.00E−35	52	81.97
TCONS_00030475	HB-other	340	6.00E−35	52	81.97
TCONS_00032949	ZF-HD	184	6.00E−101	151	87.79
TCONS_00032950	ZF-HD	281	1.00E−59	118	79.86
TCONS_00033008	C3H	454	8.00E−10	92	20.77
TCONS_00033288	MYB	301	3.00E−31	60	100
TCONS_00035639	C3H	1,951	3.00E−93	150	56.52
TCONS_00038421	bHLH	680	9.00E−50	85	100
TCONS_00042101	ARF	548	6.00E−49	100	59.28
TCONS_00043309	FAR1	883	4.00E−17	40	95.12
TCONS_00045090	bHLH	264	2.00E−66	105	99.05
TCONS_00047929	MYB_related	297	5.00E−08	42	46.03
TCONS_00049680	NAC	1,524	6.00E−22	60	50.45
TCONS_00054035	MYB	197	4.00E−29	55	100
TCONS_00060049	HSF	393	6.00E−177	247	100
TCONS_00062368	NAC	213	6.00E−108	112	99.11
TCONS_00064035	NAC	213	5.00E−121	112	99.11
TCONS_00065712	NAC	1,524	1.00E−39	106	50.31
TCONS_00066456	M-type	216	1.00E−17	92	32.57
TCONS_00066783	WRKY	237	5.00E−10	32	80
TCONS_00066845	GRAS	1,044	4.00E−24	82	44.83

The transcription factor-binding sites of differentially expressed lncRNAs were predicted by hidden Markov model using the upstream 2,000-bp (base pair) sequence of lncRNAs together with the Plant Transcription Factor Database v3.0 (<http://plantfdb.cbi.pku.edu.cn/>)

Table 2 The significant GO categories of significantly expressed lncRNAs target genes

			GO terms	FDR	Num.		
GO:0008150 biological_process	Up-regulated	OW vs. NW	GO:0032501 multicellular organismal process	1.50e−10	20		
			GO:0023052 signaling	2.60e−27	65		
			GO:0050789 regulation of biological process	5.90e−12	84		
			GO:0002376 immune system process	2.10e−36	44		
			GO:0000003 reproduction	8.50e−12	20		
			GO:0065007 biological regulation	3.60e−12	87		
			GO:0016265 death	8.10e−99	121		
			GO:0051704 multi-organism process	2.60e−11	20		
			GO:0009987 cellular process	7.80e−37	266		
			GO:0022414 reproductive process	8.50e−12	20		
			GO:0023046 signaling process	1.70e−17	45		
			GO:0050896 response to stimulus	3.70e−12	45		
			OW vs. TW	GO:0032501 multicellular organismal process	1.40e−10	14	
				GO:0023052 signaling	1.90e−11	26	
				GO:0050789 regulation of biological process	2.50e−07	37	
	GO:0002376 immune system process	5.60e−07		10			
	GO:0000003 reproduction	1.90e−11		14			
	GO:0065007 biological regulation	5.60e−07		37			
	GO:0016265 death	7.30e−08		17			
	GO:0051704 multi-organism process	4.10e−11		14			
	GO:0009987 cellular process	0.0018		68			
	GO:0022414 reproductive process	1.90e−11		14			
	GO:0023046 signaling process	0.00068		12			
	TW vs. NW	GO:0032501 multicellular organismal process		1.00e−09	19		
		GO:0023052 signaling		0.0002	27		
		GO:0002376 immune system process		0.022	8		
		GO:0000003 reproduction		7.00e−11	19		
		GO:0065007 biological regulation	0.07	50			
		GO:0016265 death	9.70e−60	86			
		GO:0051704 multi-organism process	2.00e−10	19			
		GO:0009987 cellular process	2.40e−21	223			
		GO:0022414 reproductive process	7.00e−11	19			
	Down-regulated	OW vs. NW	GO:0008152 metabolic process	0.028	88		
OW vs. TW			GO:0032501 multicellular organismal process	5.50e−08	12		
			GO:0023052 signaling	0.014	12		
		GO:0050789 regulation of biological process	0.03	25			
		GO:0000003 reproduction	1.10e−08	12			
		GO:0065007 biological regulation	0.045	25			
		GO:0051704 multi-organism process	2.00e−08	12			
		GO:0022414 reproductive process	1.10e−08	12			
		GO:0008152 metabolic process	0.012	76			
		TW vs. NW	None	None	None		
			Up-regulated	OW vs. NW	GO:0060089 molecular transducer activity	1.70e−23	44
					GO:0030528 transcription regulator activity	6.30e−05	39
		GO:0005488 binding			9.60e−47	327	
OW vs. TW		GO:0060089 molecular transducer activity		0.00065	10		
		GO:0030528 transcription regulator activity		2.70e−06	24		
	GO:0005488 binding	5.50e−05		89			
TW vs. NW	GO:0030528 transcription regulator activity	7.40e−05		38			
	GO:0005488 binding	3.00e−38		301			
	Down-regulated	OW vs. NW		None	None	None	
OW vs. TW		None	None	None			
TW vs. NW		GO:0005488 binding	4.60e−05	122			
	GO:0009055 electron carrier activity	0.0032	15				
GO:0003674 molecular_function	Up-regulated	OW vs. NW	GO:0060089 molecular transducer activity	1.70e−23	44		
			GO:0030528 transcription regulator activity	6.30e−05	39		
			GO:0005488 binding	9.60e−47	327		
			OW vs. TW	GO:0060089 molecular transducer activity	0.00065	10	
				GO:0030528 transcription regulator activity	2.70e−06	24	
				GO:0005488 binding	5.50e−05	89	
	TW vs. NW	GO:0030528 transcription regulator activity	7.40e−05	38			
		GO:0005488 binding	3.00e−38	301			
		Down-regulated	OW vs. NW	None	None	None	
OW vs. TW	None		None	None			
TW vs. NW	GO:0005488 binding		4.60e−05	122			
			GO:0009055 electron carrier activity	0.0032	15		

The gene ontology terms were analyzed with agriGO (<http://bioinfo.cau.edu.cn/agriGO/index.php>), and significant GO categories were checked using a FDR-adjusted value of ≤ 0.05 as the cutoff

identity = 95 %. Then RNAplex was used to select targets with E value = -30 (Tafer and Hofacker 2008). Target gene annotations were carried out by PopGenie (<http://www.popgenie.org/>) (Sjödin et al. 2009) and gene ontology terms were analyzed with agriGO (<http://bioinfo.cau.edu.cn/agriGO/index.php>) (Du et al. 2010). The enriched GO categories were checked using an FDR-adjusted value of ≤ 0.05 as the cutoff to find significant GO categories. The interactions among mRNAs, lncRNAs and miRNAs were modeled with Cytoscape 3.2 (Saito et al. 2012).

5'-RACE

RNA ligase-mediated 5'-RACE (RLM-RACE) was performed with the First Choice RLM-RACE Kit (Ambion) according to the manufacturer's instructions with slight modifications. PCR was performed with 5' adaptor primers and 3' gene-specific primers (Supplementary material S1) using cDNA as the template. Four putative lncRNAs for four categories were chosen for verification, including i: TCONS_00025385 and host gene POPTR_0016s00250; o: TCONS_00016255 and host gene POPTR_0012s02210; x: TCONS_00058236 and refer gene POPTR_0008s02100; u: TCONS_00012272. The RACE products were gel-purified, cloned, and sequenced.

qRT-PCR

qRT-PCR was performed as described, using the TaKaRa ExTaq R PCR Kit and SYBR green dye (TaKaRa, Dalian, China) on a DNA Engine Opticon 2 machine (MJ

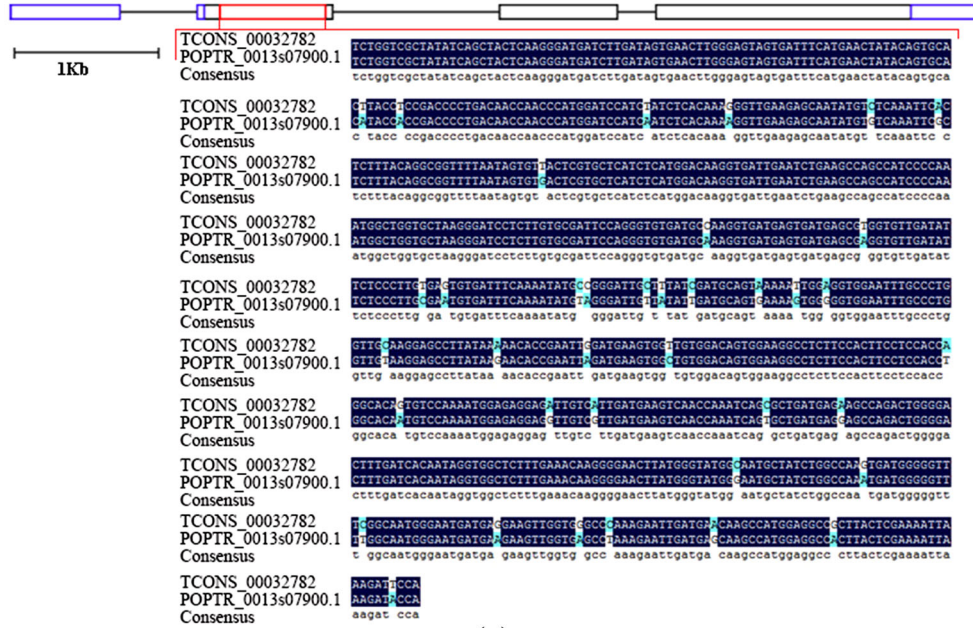
Research, Waltham, MA, USA) (Zhang et al. 2011). All reactions were carried out in triplicate technical and triplicate biological repetitions. Fourteen differentially expressed lncRNAs from the four classes (Supplementary material S1) were chosen for qRT-PCR verification. Among them, the expression of three lncRNAs (R4, R5 and R8) was detected in five tissues including leaf, phloem, cambium, root, and developing xylem along with TW, NW and OW. To discover the relationship between expression of the lncRNA and expression of its host gene, expression levels of lncRNA and its host gene (Supplementary material S1) were measured by qRT-PCR. Similarly, to find the relationship between lncRNA expression and expression of its potential target gene, expression levels of four pairs (Supplementary material S1) were also measured by qRT-PCR. Among the four genes, *lnTqP1* encodes a protein similar to cellulose synthase-like D4 (PtCSLD4), *lnTqP3* encodes a protein similar to gibberellin 2-oxidase (GA2ox) and *lnTqP4* encodes a protein similar to a putative glucosyltransferase. At the same time, the expression levels of miRNAs and their potential targets (genes and lncRNAs) were also measured, as well as their host lncRNAs (Supplementary material S1). Small RNAs (<200 nt) were isolated using the mirVana miRNA Isolation Kit (Ambion, Foster City, CA, USA) following the manufacturer's instructions. qRT-PCR of miRNAs precursors was carried out as previously described (Schmittgen et al. 2008). Primers are shown in Supplementary material S1. The efficiency of the primers was calculated by performing real-time PCR on several dilutions of first-strand cDNAs. Efficiencies of the different primer sets were similar. The

Table 3 The important target genes related to wood formation

LncRNA ID	Class	Trans targets	Trans E value	Trans target description ^a
TCONS_00032782	o	POPTR_0013s07900.1	-88.2	Similar to CELLULASE SYNTHASE 3
TCONS_00056386	o	POPTR_0017s06100.1	-81.1	Similar to putative glucosyltransferase
TCONS_00056386	o	POPTR_0001s39980.1	-80.8	UDP-glucuronosyl/UDP-glucosyl transferase family protein
TCONS_00056386	o	POPTR_0007s00890.1	-75.9	Similar to putative glucosyltransferase
TCONS_00060049	o	POPTR_0010s11500.1	-72	Similar to caffeoyl-CoA 3-O-methyltransferase
TCONS_00056386	o	POPTR_0007s00870.1	-71.5	UDP-glucuronosyl/UDP-glucosyl transferase family protein
TCONS_00022311	o	POPTR_0010s15950.1	-71.3	Similar to gibberellin 2-oxidase
TCONS_00045819	o	POPTR_0010s14680.1	-61.3	Similar to glucosyltransferase like protein
TCONS_00029839	u	POPTR_0015s12170.1	-61	Predicted protein
TCONS_00026974	o	POPTR_0009s04390.1	-59.9	UDP-glucuronosyl/UDP-glucosyl transferase family protein
TCONS_00027462	o	POPTR_0007s12400.1	-59.8	UDP-glucuronosyl/UDP-glucosyl transferase family protein
TCONS_00036943	o	POPTR_0006s20340.1	-57.6	Similar to sugar carrier protein A
TCONS_00053930	o	POPTR_0001s07400.1	-57.1	similar to 4-coumarate-CoA ligase (EC 6.2.1.12)
TCONS_00032782	o	POPTR_0003s09580.1	-56.4	Similar to cellulose synthase-like D1 protein (<i>PtCSLD1</i>)
TCONS_00002143	o	POPTR_0007s12390.1	-54.5	Similar to UTP-glucose glucosyltransferase
TCONS_00026694	o	POPTR_0002s09860.1	-52.8	Similar to probable UDP-glucuronosyltransferase (EC 2.4.1.-)

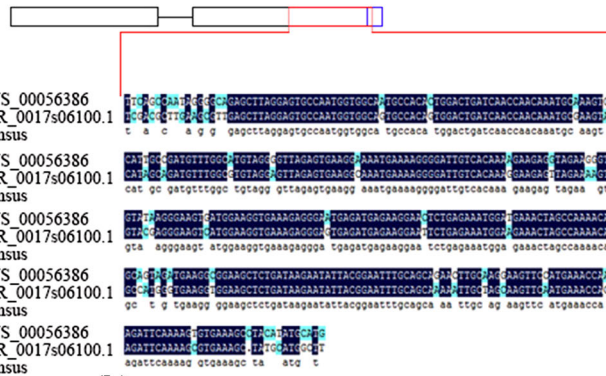
^a Target gene annotations were carried out by PopGenie (<http://www.popgenie.org/>)

The structure of POPTR_0013s07900.1



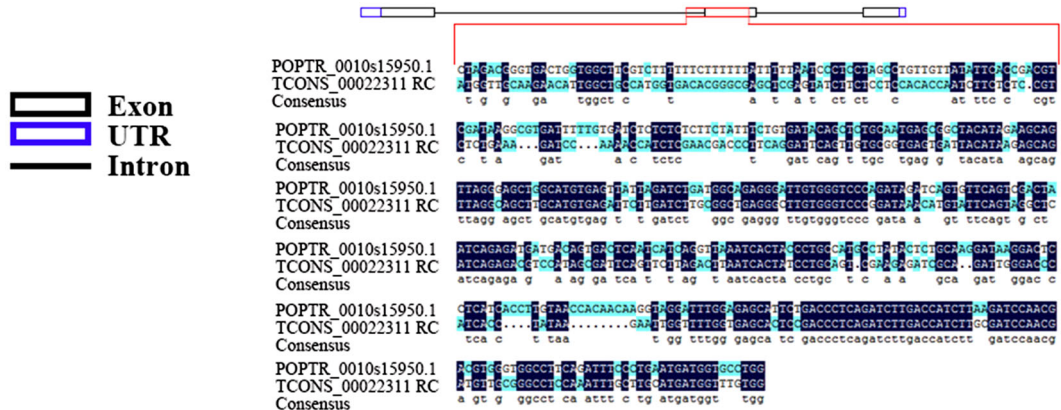
(a)

The structure of POPTR_0017s06100.1



(b)

The structure of POPTR_0010s15950.1



(c)

Fig. 3 The sites where lncRNAs match their potential target genes. **a** The match sites of *TCONS_00032782* and *POPTR_0013s07900.1* (*Pt-CESA3.1*). **b** The match sites of *TCONS_00056386* and *POPTR_0017s06100.1* (*Pt-ZOG1.12*). **c** The match sites of *TCONS_00022311* and *POPTR_0010s15950.1* (*Pt-GA2.3*). The results show the lncRNAs and their potential target genes match very well and probably interact to cause *trans*-regulatory effects

specificity of each primer set was checked by sequencing PCR products (Zhang et al. 2010). The reactions were carried out in a 20- μ l volume containing 2 μ l of diluted cDNA, 200 nM of each primer, and 19 PCR Master Mix with the following conditions: 95 °C for 30 s, and 45 cycles of 95 °C for 5 s, 58 °C for 15 s, and 72 °C for 20 s. Then, a thermal denaturing cycle of 95 °C for 15 s and 60 °C for 1 min was applied to determine the dissociation curves, which were used to verify the specificity of PCR amplifications. Relative expression levels of candidate genes, miRNAs and lncRNAs were calculated by the $2^{-\Delta C_t}$ method. The results obtained for the different tissues were analyzed and standardized to the mRNA levels of poplar *ACTIN1-like* (accession number: EF145577), which shows stable expression.

Results

Characterization of novel lncRNAs in *Populus*

In total, our sequencing generated 128,972,228, 117,672,362 and 140,978,316 raw reads in OW, NW and TW, respectively. After trimming, 108,187,750, 101,399,718 and 119,716,602 reads remained in OW, NW and TW, respectively (Supplementary material S2). Our computational analysis identified 1,377 putative lncRNAs from wood xylem tissues that were isolated from three different parts of *P. tomentosa* (Supplementary material S3). Structural analysis showed that these *Populus* lncRNAs range from 201 to 4,524 nt, with a mean of 694 nt. Here, the lower range is normal because the RNeasy micro kit does not recover RNAs with size lower than 200 nt that agreed with the definition of lncRNAs longer than 200 nt. The average length of lncRNAs was 2.2-fold shorter than the average length of protein-coding transcripts, which is 1,560.5 nt. For expression levels, the average counts for lncRNAs were less than the counts for coding transcripts

by about 0.13-fold for OW, 0.14-fold for NW, and 0.18-fold for TW. As determined by their genomic locations with respect to nearby genes, the 1,377 lncRNAs included members of all four classes of lncRNAs, with most in the o class, which overlap exonic regions (15 members of the i class, 1,030 in o, 202 in u and 130 in x). In order to validate the novel lncRNAs, the sequences of four lncRNAs from four categories were cloned by 5'-RACE in our study (Fig. S1; Supplementary material S4). The cloning of RACE-PCR products showed that the entire contig can be amplified, indicating that the contigs were not assembly artifacts. As shown in Fig. S1a, the cloned sequence of *POPTR_0016s00250*, which includes the class i lncRNA *TCONS_00025385*, shows that the lncRNA was transcribed from the first intron. Also, the class o lncRNA *TCONS_00016255* overlaps with the last exon and 3'-UTR of *POPTR_0012s02210* (Fig. S1b). Moreover, the reverse-complement sequence of the class x lncRNA *TCONS_00058236* covers three parts of *POPTR_0008s02100* (Fig. S1c). Lastly, the cloning of a class u lncRNA showed that it could be cloned and no gene was located near it (Supplementary material S4).

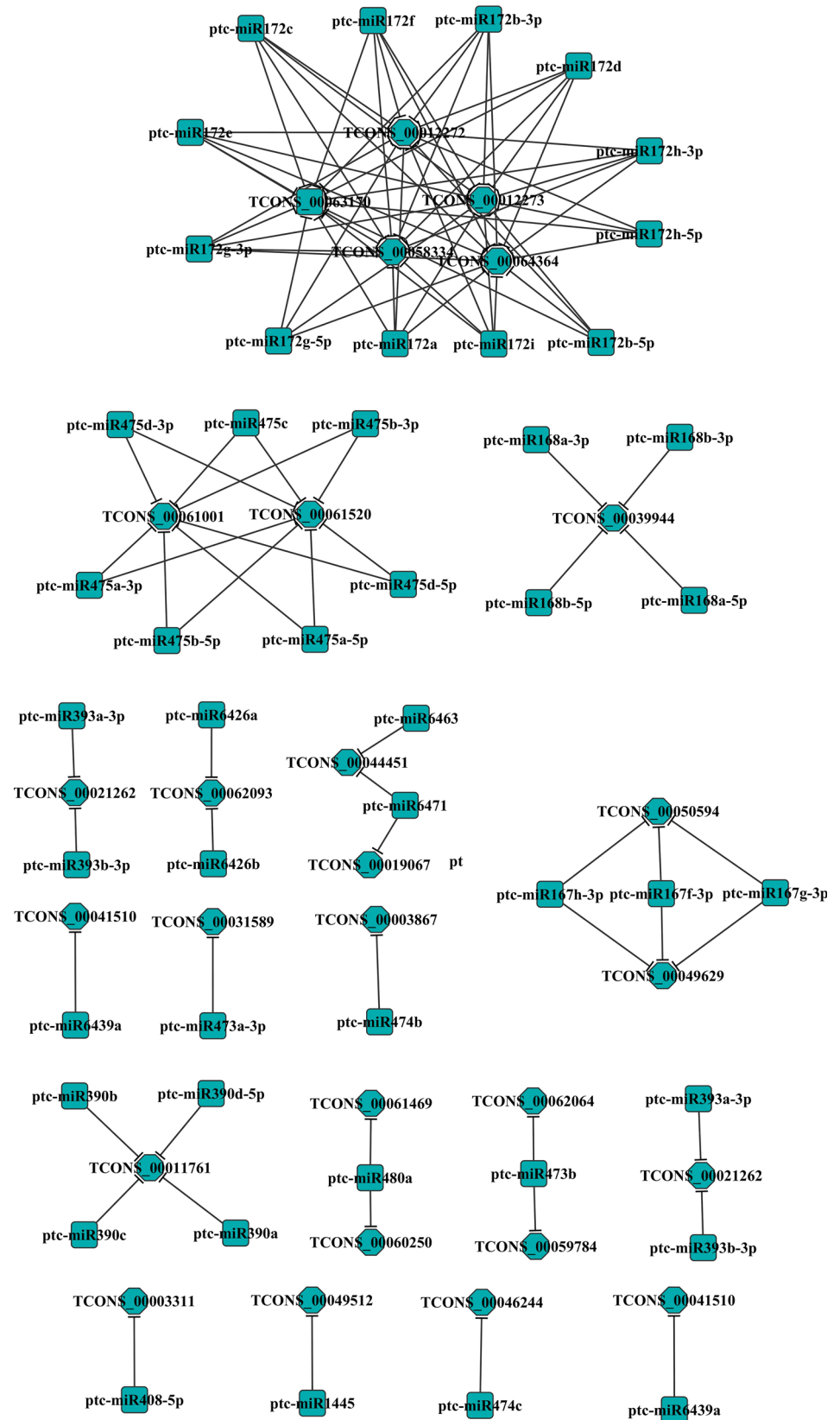
Differentially expressed lncRNAs

As shown in Fig. 2a, among the 1,377 putative lncRNAs, 54 lncRNAs were only identified in one library, not in the others, with 17 uniquely expressed in TW, 11 in OW and 26 in NW. Also, 766 lncRNAs were differentially expressed in TW, OW and NW with $\log_2FC \geq 1$ or ≤ -1 , and $FDR \leq 0.01$. Of these, 32 TFs from 16 families identified by TFs discovery analysis could bind to the 2-kb upstream sequences of 32 lncRNAs (Table 1). In comparison to NW, 589 lncRNAs were differentially expressed in OW, including 380 up-regulated and 209 down-regulated (Fig. 2b; Supplementary material S5). Similarly, 578 lncRNAs were differentially expressed in TW compared to NW, with 370 up-regulated and 208 down-regulated (Fig. 2b; Supplementary material S5). Lastly, 298 lncRNAs were differentially expressed in OW compared to TW, with 135 up-regulated and 163 down-regulated (Fig. 2b; Supplementary material S5). The expression patterns of the common differentially expressed lncRNAs in three wood tissues (109) could be categorized into five major types, A to E (Supplementary material S6). The expression of type

Table 4 lncRNAs transcripts corresponding to miRNA precursors

lncRNA ID	miRNA ID	Mature sequence of miRNA
TCONS_00061773	Ptc-miR1448	CUUUCCAACGCCUCCCAUAC
TCONS_00061773	Ptc-miR482a.2	UCUUGCCUACUCCUCCCAUU
TCONS_00023843	Ptu-172	AUUCCCUAUACUCACAUC
TCONS_00048079	Ptu-37	CAUAGUAGGAUUCUGGGGCU

Fig. 4 The networks of miRNAs and lncRNAs. 25 lncRNAs were predicted as potential target genes of 44 miRNAs, in psRNATarget with expectation ≤ 3



A and B lncRNAs such as *TCONS_00026694* and *TCONS_00028236* was NW > TW > OW and NW > OW > TW, and these lncRNAs were similarly suppressed by tension and compression stresses (Supplementary material S6). The lncRNAs showing lowest expression in NW belonged to type C (TW > OW > NW) and D (OW > TW > NW), and included *TCONS_00044376* and *TCONS_00055243* (Supplementary material S6). The expression of type E lncRNAs, such as *TCONS_00028236*, was TW > NW > OW (Supplementary material S6).

Trans-acting target genes of differently expressed lncRNAs

Computational prediction of potential *trans*-regulated target genes of the differentially expressed lncRNAs identified 1,151 potential target genes for 389 lncRNAs (Supplementary material S7). To functionally annotate and predict their biological roles, the predicted targets of the differentially expressed lncRNAs were sorted into GO-term categories for molecular function, biological processes, and cellular component using agriGO based on the Blast2GO program (Du et al. 2010) (Supplementary material S8). The significant GO categories were selected with an FDR-adjusted value of ≤ 0.05 as the cutoff (Table 2). For biological process, nine common significant terms were identified among potential target genes of up-regulated lncRNAs in OW vs. NW, OW vs. TW and OW vs. TW (Table 2). The “response to stimulus” category was identified in OW vs. NW (Table 2), suggesting that target genes predicted from the up-regulated lncRNAs among OW vs. NW were involved in response to stimulus. Within the GO category, “molecular function”, the “transcription regulator activity” and “binding” terms were identified in OW vs. NW, OW vs. TW and OW vs. TW (Table 2). The “molecular transducer activity” was identified among OW vs. NW and OW vs. TW (Table 2). Regarding the “biological process” GO category, of the target genes of down-regulated lncRNAs, eight terms were enriched in OW vs. TW and one in OW vs. NW, including “metabolic process” (Table 2). For “molecular function”, there were two terms, “binding” and “electron carrier activity” identified in TW vs. NW (Table 2).

After the functional exploration of target genes in general, we next selected genes related to cellulose synthase (UDP-forming) activity, lignin biosynthesis, sugar carrier, UDP-glucuronosyl and UTP-glucose glucosyltransferase, as potentially important target genes in wood formation (Table 3). The lncRNA *TCONS_00032782* was predicted to regulate POPTR_0013s07900.1 (*Pt-CSLD4*), which encodes a protein similar to cellulase synthase 3. As shown in Fig. 3a, the match sites were in the first exon of *Pt-CSLD4* and occurred in good alignment. Genes encoding

glucosyltransferases were predicted as targets of *TCONS_00056386* (Table 3; Fig. 3b). Besides the genes related to cellulose, lignin and glucosyltransferase, POPTR_0010s15950.1 (*Pt-GA2.3*) encodes GA2ox and was predicted as a target of *TCONS_00022311* (Fig. 3c). Another target gene, predicted to be targeted by *TCONS_00053930*, encodes a protein similar to 4CL, the central enzyme of the plant-specific phenylpropanoid pathway, which has an important role in lignin biosynthesis (Table 3).

lncRNA transcripts as targets of miRNAs and as miRNA precursors

To examine the regulatory functions of the lncRNAs, we next examined their sequences to determine if they could be targets of known miRNAs. The psRNATarget program, with expectation ≤ 3 , predicted 25 lncRNAs as targets of 44 miRNAs from 15 families (Supplementary material S9). Based on this, we also modeled the network of miRNAs regulating lncRNAs, showing that one lncRNA could potentially be targeted by 1–12 miRNAs and one miRNA could target 1–5 lncRNAs (Fig. 4). As shown in Fig. 5, there are perfect seed-match sequences between miRNAs and lncRNAs, indicating that miRNAs may target the lncRNAs. The five lncRNAs that were predicted to be targeted by miRNAs also may interact with 11 target genes via *trans*-regulatory effects, as indicated by the computational analyses described in the previous section. Also, 204 transcripts were predicted to be target genes of the 44

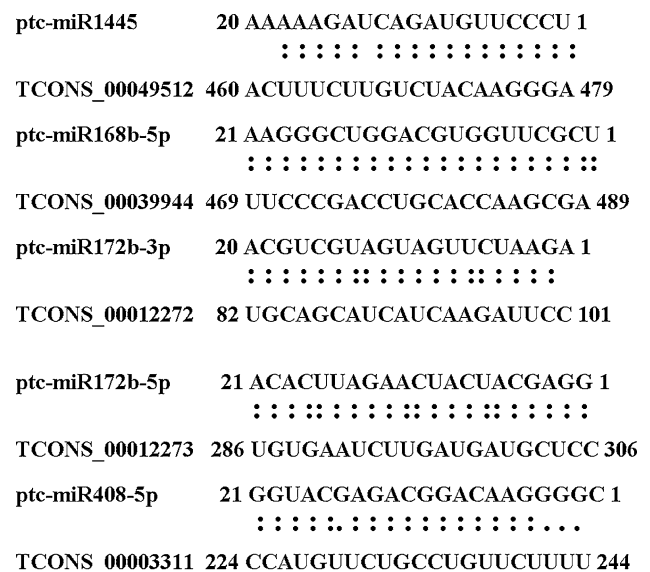
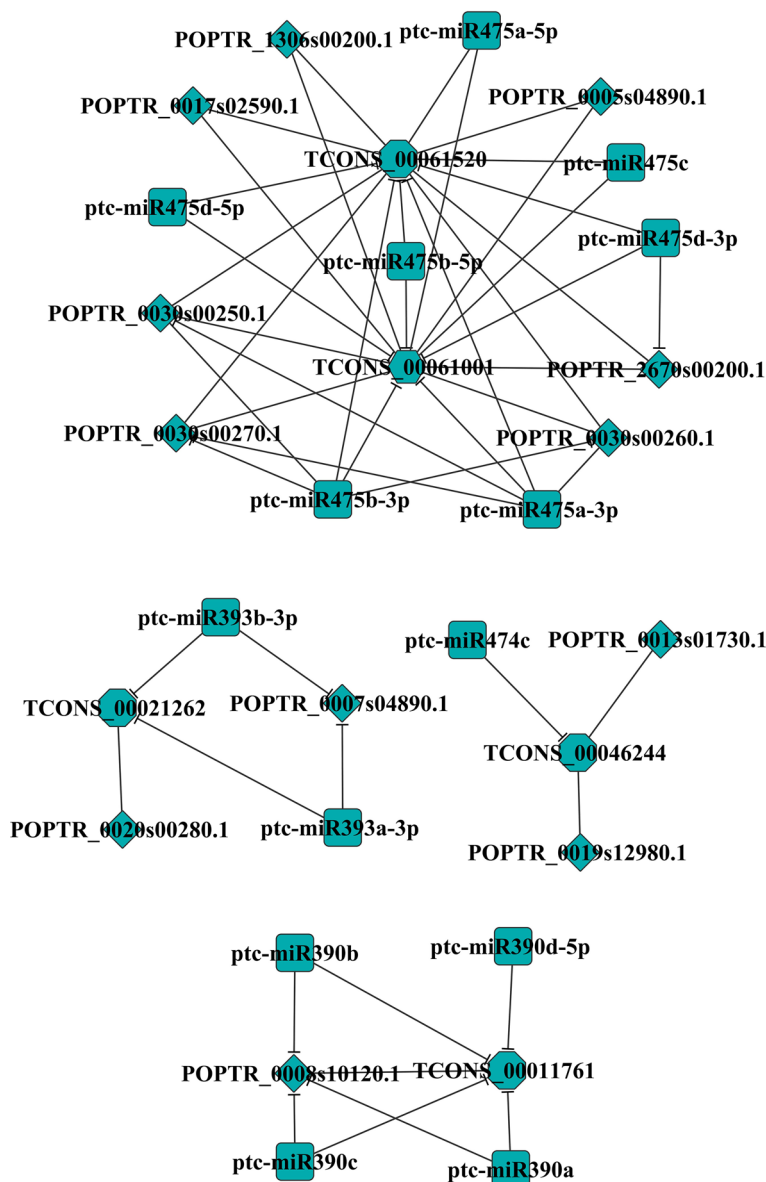


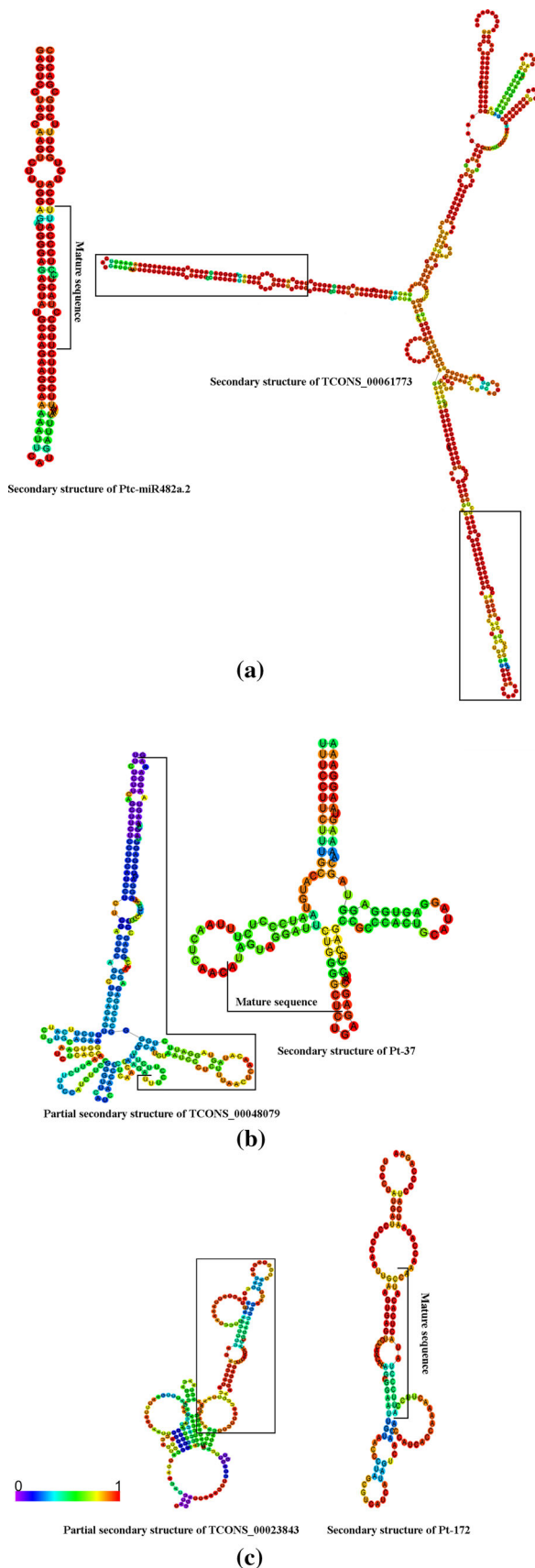
Fig. 5 The binding alignments of miRNAs targeting predicted lncRNAs in *Populus*. The figure shows predicted binding alignment of Ptc-miR1445, Ptc-miR168b-5p, Ptc-miR172b-3p, Ptc-miR172b-5p and Ptc-miR408-5p with predicted *Populus* lncRNAs

Fig. 6 The interaction networks of miRNAs, lncRNAs and mRNA. Only the mRNAs involved in interactions with miRNAs and lncRNAs are shown



miRNAs that probably regulate lncRNAs. The interaction network of miRNAs, lncRNAs and mRNAs, in which miRNAs regulated lncRNAs and mRNAs, and lncRNAs targeted mRNAs, was modeled using Cytoscape 3.2 (Fig. 6). Since we focused on the network of mRNAs, lncRNAs and miRNAs, only the mRNAs involved in interactions with miRNAs and lncRNAs are shown in Fig. 6. The interaction network involved in 44 miRNAs, 25 lncRNAs and 210 mRNAs (5 were commonly regulated by both lncRNAs and miRNAs), showing mRNAs could be regulated by both lncRNAs and miRNAs (Fig. 6). Thus, the network of lncRNAs, miRNAs and mRNAs shows that lncRNAs could potentially interact with miRNAs and may be a component of regulatory interactions with miRNAs.

We also checked the sequences for lncRNAs that correspond to miRNA precursors, and found that only three of the 1377 lncRNAs (0.2 %), i.e., *TCONS_00061773*, *TCONS_00023843* and *TCONS_00048079*, harbor complete precursors, for four miRNAs (Ptc-miR1448, Ptc-miR482a.2, Ptu-172 and Ptu-37) (Table 4; Fig. 7). Among the four miRNAs, two have been reported in miRBase and another two were novel miRNAs with a characteristic hairpin structure. Prediction of the secondary structure for the four transcripts using the Vienna RNA package RNAfold program showed that these four miRNA precursors had stable hairpin structures (Fig. 7). The precursors of two known miRNAs, Ptc-miR1448 and Ptc-miR482a.2 were located in one lncRNA, *TCONS_00061773*. None of the target genes of these three lncRNAs was predicted, but



◀ **Fig. 7** The predicted secondary structures of lncRNAs transcripts and miRNAs sequences. **a** *TCONS_00061773* consists of Ptc-miR1448 and Ptc-miR482a.2 precursors. **b** *TCONS_00023843* contains Pt-172 precursors. **c** *TCONS_00048079* contains Pt-37 precursors. The secondary structure was created by Vienna RNA package RNAfold web (<http://rna.tbi.univie.ac.at/>)

psRNATarget predicted 39 target genes for the four miRNAs (Supplementary material S10).

Verification of RNA-seq data and examination of tissue-specific expression of lncRNAs by qRT-PCR

The expression patterns identified by RNA-seq were further validated by qRT-PCR with sequence-specific primers. Figures 8 and 9 illustrate the relative transcript levels of lncRNAs, miRNAs and mRNAs normalized to the expression level of *Pt-ACTIN*. For this verification, we focused primarily on the transcripts of significantly expressed lncRNAs (R1-R14). As shown in Fig. 8a, similar expression patterns were detected by RNA-seq and qRT-PCR; for instance, R5 (*TCONS_00056386*) was detected as having its highest expression in OW, then TW and lowest in NW (Supplementary material S5) by RNA-seq and a similar expression pattern was also detected by qRT-PCR. The expression of five lncRNAs representing five expression patterns (R11:A, R12:B, R13:C, R3:D and R14:E) was measured by qRT-PCR and agreed with the results of RNA-seq. As the RNA-seq revealed that the lncRNAs have specific expression in the three xylem tissues, the lncRNAs also have tissue-specific expression patterns in five different tissues examined by qRT-PCR (Fig. 8b). As indicated in Fig. 8b, the lncRNAs displayed distinct tissue-expression patterns with abundance in root, stem and leaf in poplar. The highest expression of R5 lncRNA was observed in flowers and the lowest in NW (Fig. 8b). The expression of R5 lncRNA in developing xylem is higher than in mature xylem (NW, TW and OW) or in phloem and cambium, suggesting that lncRNAs are differentially expressed in vascular tissues. Of the 14 lncRNAs, seven were located in one known gene (detail in Supplementary material S11); thus, the correlation of expression levels between the host gene and lncRNAs was further analyzed. As shown in Fig. 8c, the lncRNAs and host genes have almost the same expression in TW, NW and OW. For example, both R9 and G9 expressed highest in TW (TW > OW > NW). However, the expression relationship between lncRNAs and their host lncRNAs in one tissue was not the same; for example, R8 expressed higher than G8, while R9 lower than G9. To explore the expression relationship between lncRNAs and their potential target genes, four potential target genes were measured by qRT-PCR. The expression of lncRNAs and their potential target

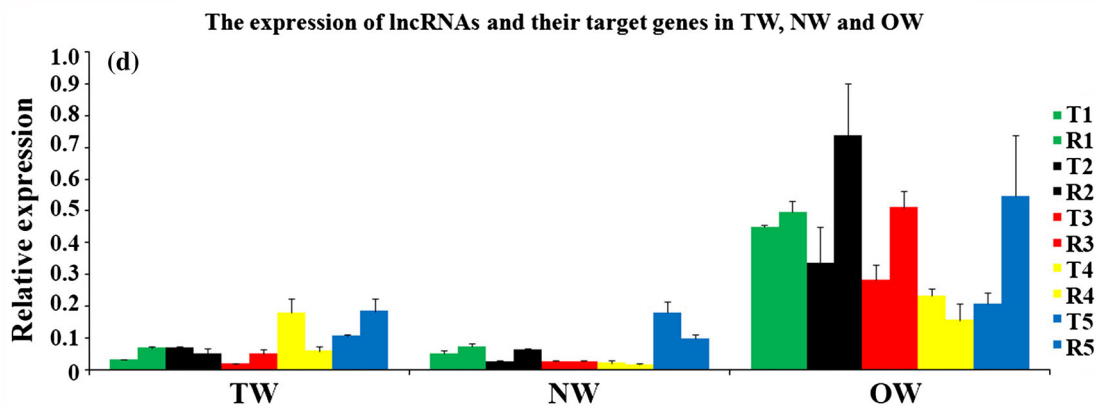
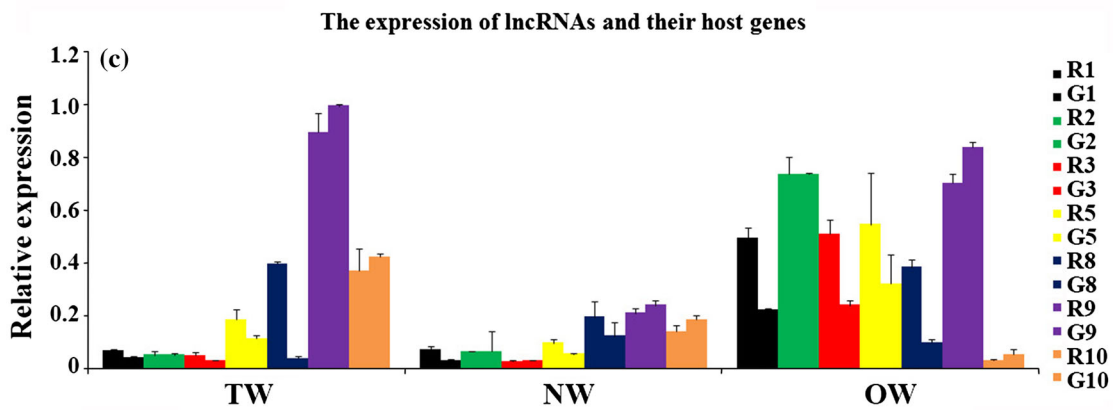
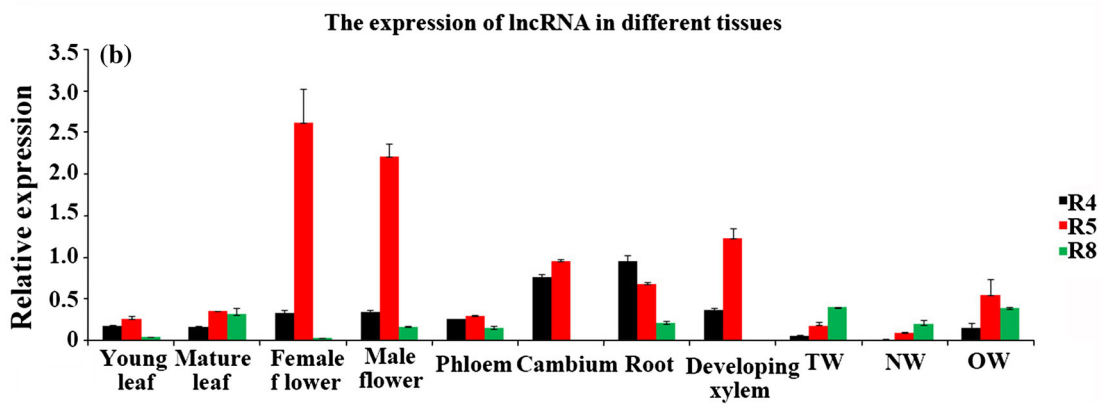
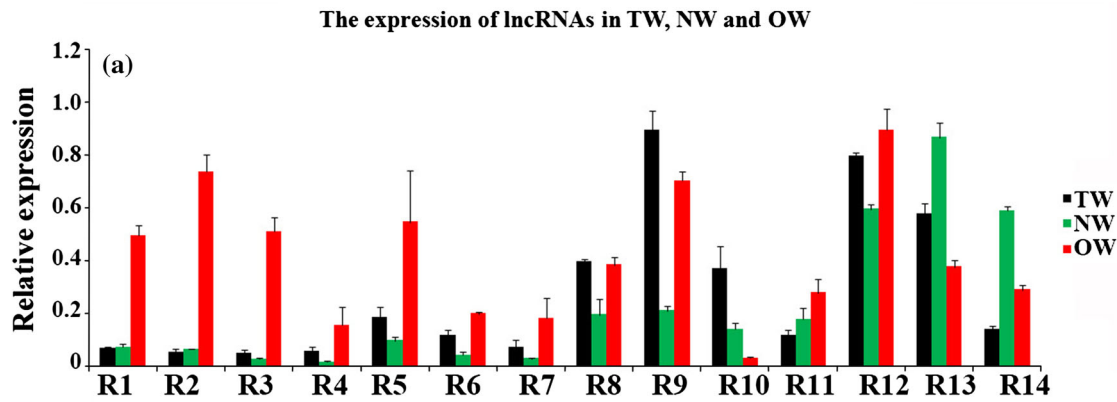


Fig. 8 The expression levels as measured by qRT-PCR. **a** The expression level of lncRNAs in TW, NW and OW. **b** The expression level of lncRNAs in different tissues. **c** The expression level of lncRNAs and host genes in TW, NW and OW. **d** The expression level of lncRNAs and their target genes in TW, NW and OW. The relative expression levels were standardized to the levels of *Pt-ACTIN* mRNA

genes identified by qRT-PCR showed the two pairs R1–T1 and R4–T4 had the same expression pattern in three xylem tissues. The lncRNAs and their potential target genes of the remaining two pairs were all expressed highest in OW, while their expression in NW and TW was opposite. Expression levels measured by qRT-PCR thus suggested that lncRNAs and their target genes might have similar tissue-specific expression (Fig. 8d).

In order to examine the relationship between expression of miRNAs and their potential targets, including genes and lncRNAs, we used qRT-PCR to measure the expression of miRNAs and their potential targets (Fig. 9). Figure 9a shows that a negative relationship between miRNAs and their potential target lncRNAs were detected; for example, miR2 (*Ptc-miR168b-5p*) is expressed in OW (OW > TW > NW), while its potential target lncRNA *lmR2* (*TCONS_00039944*) showed its lowest expression in OW (NW > TW > OW). Also, the miRNA miR3 (*Ptc-miR1448*) showed the highest expression in TW (TW > NW > OW) and its potential target gene *miT3* showed the lowest expression in TW (OW > NW > TW), indicating a negative expression relationship between them (Fig. 9b). Two miRNAs (*miR7:Ptu-172* and *miR8:Ptu-37*) that we predicted in our study were expressed only in TW, and their potential genes (*miT5* and *miT6*) showed their lowest expression in TW (Fig. 9b). Since *Ptc-miR1448* (*miR3*) and *Ptc-miR482a.2* (*miR4*) were located in *TCONS_00061773* (*lmR3*), we also measured their expression (Fig. 9c), finding the *Ptc-miR1448* is much more abundant than *Ptc-miR482a.2*. However, no significant correlation between miRNAs and their host lncRNAs was found. In summary, the qRT-PCR analysis verified the data from the RNA-seq analysis, and revealed that the lncRNAs had tissue-specific patterns of expression in *Populus*.

Discussion

The *P. tomentosa* lncRNAs are shorter and expressed at lower levels than protein-coding transcripts

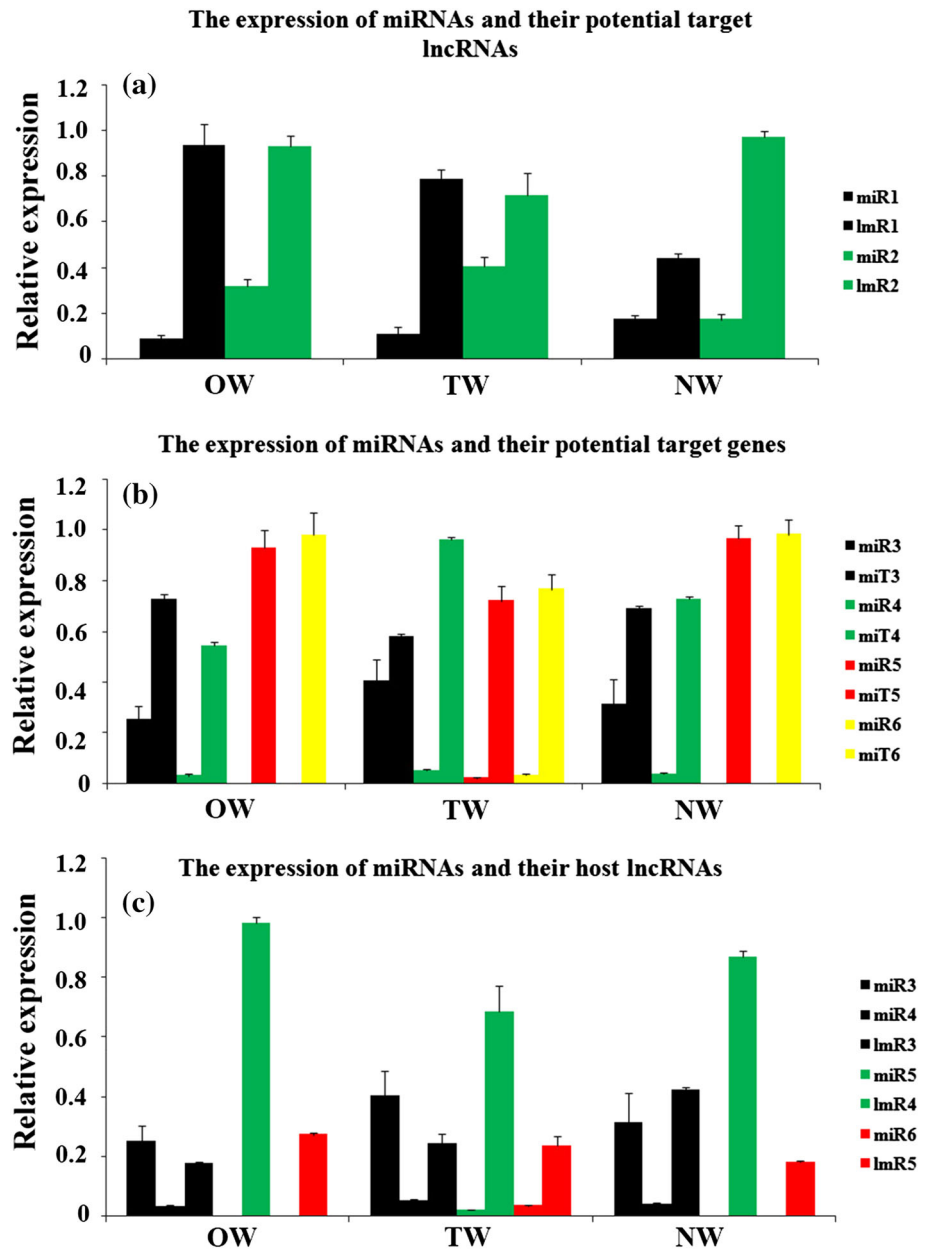
Our study identified 1,377 putative lncRNAs in three xylem tissues by RNA-seq, and used computational analysis to classify these into four categories to explore their probable functions. Since the whole genome sequence of *P. tomentosa* has not been generated and hybrids between *P.*

trichocarpa and *P. tomentosa* are fertile, the genome of *P. trichocarpa* was used as a reference in our study. Intron sequences are rapidly broken down by nuclease, therefore the number of intronic lncRNAs detected by RNA-seq was low, less than class o. However, the transcriptional regulation of lncRNAs is unclear and explaining this phenomenon will require further analysis. Previous studies in mammals indicated that lncRNAs are shorter and expressed at significantly lower levels than protein-coding transcripts (Pauli et al. 2012). In our study, the lncRNAs were 2.2-fold shorter, on average, than coding transcripts, and the expression level of these lncRNAs was lower than the coding transcripts, in agreement with previous studies (Pauli et al. 2012). LncRNAs also have specific expression patterns in tissue types and subcellular compartments (Wang et al. 2011b) and are only expressed in some conditions and biological processes (Zhang et al. 2013a). In our study, 54 lncRNAs displayed unique expression patterns in different parts of the xylem tissue, suggesting that some lncRNAs are involved in the formation of special wood types (Fig. 2). Among the significant differently expressed lncRNAs, we found five expression patterns in OW, TW and NW, suggesting that lncRNAs are possibly associated with the formation of specialized wood types. According to our definitions, lncRNAs of types A and C can be considered as responding to mechanical stress, but such responses generally do not discriminate between tension and compression stresses, revealing potentially important roles for type A and C lncRNAs in counteracting overall mechanical stimuli. As indicated earlier, lncRNAs of the types B, D, and E, which showed differences between TW and OW, are possibly associated with more specialized regulation that may lead to preferential development of either TW or OW. High-throughput genomic technologies such as lncRNA microarray and RNA-Seq often generate a set of lncRNAs of interest, but little is known about the transcriptional regulation of lncRNA genes. Here, we identified TFs that may bind to the upstream of putative *Populus* lncRNAs and found 32 TFs including members of the NAC and MYB families, providing clues on which TFs could be common transcriptional regulators of lncRNA genes. Together, our data represent the first characterization of lncRNAs in woody plants and provide many clues for the further study of regulation of wood formation in *Populus* by lncRNAs.

LncRNAs and miRNAs are involved in an interaction network

Our study discovered 25 lncRNAs that may be targeted by 44 miRNAs and our qRT-PCR studies indicated a probable negative relationship between expression of lncRNAs and their targets, in agreement with earlier studies that showed

Fig. 9 The expression levels of miRNAs, lncRNAs and potential target genes as measured by qRT-PCR. **a** The expression of miRNAs and their potential target lncRNAs. **b** The expression of miRNAs and their potential target genes. **c** The expression of miRNAs and their host lncRNAs. The relative expression levels were standardized to the levels of poplar *ACTIN* mRNA



that miRNAs can target lncRNAs (Zhang et al. 2013b) (Fig. 5). In our study, *ptc-miR1445* was predicted to target *TCONS_00049512*, which has been predicted to function in cleaving the transcripts of development-related or stress-responsive genes in *Populus* (Lu et al. 2008). Besides *ptc-miR1445*, miRNAs that were specifically expressed in xylem and reportedly participate in formation of tension wood (Lu et al. 2005, 2008; Puzey et al. 2012) were predicted to regulate some of the lncRNAs identified in our study. For example, *ptr-miR172*, which was only suppressed in compression tissue (Lu et al. 2005), may target *TCONS_00012272* and *TCONS_00012273*, which were detected in our study and showed a type C expression

pattern (TW > OW > NW). In addition, our study constructed a comprehensive network of RNA-mediated interactions, putting together miRNA–mRNA, miRNA–lncRNA and lncRNA–mRNA interactions, using a computational approach (Fig. 6). Our results agreed with former studies (Jalali et al. 2013), and suggested that there are widespread regulatory interactions between non-coding RNAs and mRNAs in *Populus*.

Our study identified three lncRNAs that contain miRNA precursors, indicating that some lncRNAs are processed into miRNAs in *Populus* (Fig. 7), in agreement with earlier studies (Cai and Cullen 2007). *TCONS_00061773* harbors precursors for two known miRNAs, *Ptc-miR1448* and *Ptc-*

miR482a.2 (Fig. 7). Previous studies supported the idea that Ptc-miR1448 may be involved in the resistance of plants to biotic and abiotic stresses (Lu et al. 2008) and Ptc-miR482a.2 has been reported as a tree-specific miRNA in tension wood (Lu et al. 2005, 2008). The expression measured by qRT-PCR revealed that Ptc-miR1448 is much more abundant than Ptc-miR482a.2, although both miRNAs occur in one cluster in the genome (Lu et al. 2005, 2008). These findings supported the idea that *TCONS_00061773* may be associated with abiotic stress tolerance in *Populus*. Although we identified only two novel miRNAs by examination of our lncRNAs, lncRNAs might be considered as a resource for identifying novel miRNAs. Whether lncRNAs function by themselves or as precursors of miRNAs requires more evidence. All these findings supported a functional network of lncRNAs and miRNAs, but this network requires further study, and identifying well-established miRNAs that bind lncRNAs may help to infer the function of lncRNAs.

LncRNAs may function via *trans*-regulatory effects on genes involved in many processes

Studies have implicated lncRNAs in transcriptional regulation via *trans* effects in many different ways, such as by activation of TFs, which is the predicted mechanism of action of the lncRNA *Eyf2* (Bond et al. 2009). The prediction of *trans*-regulatory target genes of lncRNAs provides a useful way to identify which process lncRNAs participated in and to infer their potential functions. Our functional prediction according to GO categories (Table 2) showed that the target genes of lncRNAs are involved in many biological processes such as “multicellular organismal process”, and “signaling and biological regulation”. ncRNAs can target transcriptional activators or repressors, different components of the transcription reaction, including RNA polymerase II and even the DNA duplex to regulate transcription (Goodrich and Kugel 2006). Here, significant GOs for “transcription regulator activity” and “binding” were identified among the potential target genes of lncRNAs, which agreed with previous studies (Yoon et al. 2013). The “response to stimulus” term was only detected in the target genes of lncRNAs up-regulated in OW vs. NW, indicating a role for lncRNAs on the lower side of a branch, which receives more stimulus from gravity. Since this process is important to regulate plant adaptation to the environment, and the lncRNAs potentially regulated these genes, it could be proposed that lncRNAs might play important roles in plant environmental adaptation. Lastly, all the functional predictions implicated the potential target genes of lncRNAs in many biological processes, including in development and regulation of biotic and/or abiotic stresses, as well as in TW formation induced by gravity.

LncRNAs may regulate genes involved in wood formation

Twelve lncRNAs were identified that potentially regulate 16 important genes involved in wood formation (Table 3). Cellulose produced by the cellulose synthase (CesAs) complex is well known as a basic component of plant cell walls, but the roles of a number of CesAs and cellulose synthase-like proteins in cellulose biosynthesis require further study. CSLD family members were proposed to have glucosyltransferase activity in synthesizing a polysaccharide that has a specialized structural role in the cell walls of tip-growing cells in *Arabidopsis* (Bernal et al. 2008). *Arabidopsis CSLD1* and *CSLD4* are important for normal growth of pollen tubes, likely through cellulose deposition (Wang et al. 2011a). Ptc-miRX41 is predicted to target a *CSLD4* in *P. trichocarpa* (Puzey et al. 2012), and this study opened the door to understanding the role of ncRNAs in regulation of genes involved in cellulose biosynthesis in trees. In our study, *TCONS_00032782*, which was predicted to target *Pt-CSLD4*, revealed that lncRNA might function in wood formation through regulation of *CSLD* in *Populus*.

A prominent product of the phenylpropanoid pathway is lignin, the second most abundant naturally occurring polymer, after cellulose; lignin confers the rigidity and mechanical strength needed for plant growth. 4CL is the central enzyme of the plant-specific phenylpropanoid pathway and has an important role in lignin biosynthesis (Hu et al. 2010). *TCONS_00053930* was predicted to target 4CL (Table 3), indicating that there might be a regulatory relationship between lncRNA and 4CL. Another important enzyme involved in the phenylpropanoid pathway is caffeoyl-CoA 3-*O*-methyltransferase (CCOMT). The methylation steps in the biosynthesis of lignin monomers catalyzed by CCOMT (Maury et al. 1999) are essential in lignin biosynthesis in poplar (Zhong et al. 2000). Strong down-regulation of *CCOMT* led to reduced lignin levels and influenced lignin composition of guaiacyl and syringyl subunits (Guo et al. 2001). Although *CCOMTs* play important roles in lignin content, little is known about the regulators of *CCOMT*. *TCONS_00060049* was predicted to target *CCOMT*, suggesting that it may have a potential role in lignin formation by regulation of *CCOMT* (Table 3). Besides 4CL and *CCOMT*, other important genes, such as the genes encoding UDP-glucuronosyl/UDP-glucosyl transferase family members, were predicted as target genes of some lncRNAs identified in our study (Table 3). More recently, the discovery that Ptr-miR397a functions as a negative regulator of laccase genes, and in turn affects lignin content in transgenic *P. trichocarpa* (Lu et al. 2013), revealed that miRNAs targeting ncRNAs function as an important factor in regulation of genes involved in lignin

biosynthesis. All these results indicated that lncRNAs might take part in regulation of lignin biosynthesis via *4CL* and *CCOMT*, uncovering an area of regulation of lignin biosynthesis genes by ncRNAs.

Apart from genes involved in cellulose and lignin biosynthesis, genes related to plant hormones also play important roles in controlling cambial activity and the differentiation of cambial derivatives in woody plants, and also affect the formation of TW induced by gravity (Du et al. 2004; Israelsson et al. 2005; Dayan et al. 2012). Gibberellin can promote the formation of TW on the upper sides of branches (Yoshida et al. 1999). Recently, Nugroho et al. (2012) reported that gibberellin is required for the formation of tension wood and stem growth in *Acacia mangium* seedlings. All these results showed that wood formation is likely to be strongly affected by gibberellin. GA2ox is a key enzyme in the gibberellin biosynthetic pathway. In our study, *TCONS_00022311* was predicted to target a *GA2ox*, suggesting that lncRNAs may take part in regulating wood formation through gibberellin biosynthesis.

Conclusions

Using the TW in poplar as a model, we have identified differentially expressed lncRNAs in NW, TW and OW by RNA-seq, and predicted their potential roles in wood formation. This study identified and characterized 1,377 putative lncRNAs in *Populus*. Of these, 589, 578, 298 were significantly differentially expressed between OW vs. NW, TW vs. NW, and OW vs. TW, respectively. We found that the predicted target genes of lncRNAs are involved in many biological processes, such as cellulose, lignin and gibberellin biosynthesis, suggesting lncRNAs' potential roles in wood formation. In addition, an interaction network between lncRNAs, miRNAs and mRNAs was constructed, showing that lncRNAs may function through miRNA pathways and indicating interactions between lncRNAs and miRNAs. This study provides new insights into the role of lncRNAs in the molecular mechanisms underlying reaction wood formation. This lncRNA resource can be mined to explore the functions of lncRNAs and for further genetic studies in *Populus*.

Acknowledgments This work was supported by the State Key Basic Research Program of China (No. 2012CB114506), and the Projects of the National Natural Science Foundation of China (No. 31170622, 30872042), and Shandong Province Agriculture Improved Variety Project (No. 2012213).

References

Andersson-Gunnerås S, Mellerowicz EJ, Love J, Segerman B, Ohmiya Y, Coutinho PM, Nilsson P, Henrissat B, Moritz T,

Sundberg B (2006) Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *Plant J* 45:144–165

Bernal AJ, Yoo C-M, Mutwil M, Jensen JK, Hou G, Blaukopf C, Sørensen I, Blancaflor EB, Scheller HV, Willats WG (2008) Functional analysis of the cellulose synthase-like genes *CSLD1*, *CSLD2*, and *CSLD4* in tip-growing *Arabidopsis* cells. *Plant Physiol* 148:1238–1253

Boerner S, McGinnis KM (2012) Computational identification and functional predictions of long noncoding RNA in *Zea mays*. *PLoS ONE* 7:e43047

Bond AM, VanGompel MJ, Sametsky EA, Clark MF, Savage JC, Disterhoft JF, Kohtz JD (2009) Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat Neurosci* 12:1020–1027

Cai X, Cullen BR (2007) The imprinted *H19* noncoding RNA is a primary microRNA precursor. *RNA* 13:313–316

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep* 11:113–116

Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res* 39:W155–W159

Dayan J, Voronin N, Gong F, Sun T-p, Hedden P, Fromm H, Aloni R (2012) Leaf-induced gibberellin signaling is essential for internode elongation, cambial activity, and fiber differentiation in tobacco stems. *Plant Cell* 24:66–79

Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, Yao J, Xu C, Li X, Xiao J, Zhang Q (2012) A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *PNAS* 109:2654–2659

Du S, Uno H, Yamamoto F (2004) Roles of auxin and gibberellin in gravity-induced tension wood formation in *Aesculus turbinata* seedlings. *IAWA J* 25:337–348

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 38:W64–W70

Gibb EA, Brown CJ, Lam WL (2011) The functional role of long non-coding RNA in human carcinomas. *Mol Cancer* 10:1–17

Goodrich JA, Kugel JF (2006) Non-coding-RNA regulators of RNA polymerase II transcription. *Nat Rev Mol Cell Biol* 7:612–616

Guo D, Chen F, Inoue K, Blount JW, Dixon RA (2001) Downregulation of *caffeic acid 3-O-methyltransferase* and *caffeoyl CoA 3-O-methyltransferase* in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and S lignin. *Plant Cell* 13:73–88

Han L, Zhang K, Shi Z, Zhang J, Zhu J, Zhu S, Zhang A, Jia Z, Wang G, Yu S (2012) LncRNA profile of glioblastoma reveals the potential role of lncRNAs in contributing to glioblastoma pathogenesis. *Int J Oncol* 40:2004–2012

Hu Y, Gai Y, Yin L, Wang X, Feng C, Feng L, Li D, Jiang X-N, Wang D-C (2010) Crystal structures of a *Populus tomentosa* 4-coumarate: CoA ligase shed light on its enzymatic mechanisms. *Plant Cell* 22:3093–3104

Israelsson M, Sundberg B, Moritz T (2005) Tissue-specific localization of gibberellins and expression of gibberellin-biosynthetic and signaling genes in wood-forming tissues in aspen. *Plant J* 44:494–504

Jalali S, Jayaraj GG, Scaria V (2012) Integrative transcriptome analysis suggest processing of a subset of long non-coding RNAs to small RNAs. *Biol Direct* 7:25

Jalali S, Bhartiya D, Lalwani MK, Sivasubbu S, Scaria V (2013) Systematic transcriptome wide analysis of lncRNA–miRNA interactions. *PLoS ONE* 8:e53823

Jin J, Zhang H, Kong L, Gao G, Luo J (2013) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res* 42:D1182–D1187

- Joseleau J-P, Imai T, Kuroda K, Ruel K (2004) Detection in situ and characterization of lignin in the G-layer of tension wood fibres of *Populus deltoides*. *Planta* 219:338–345
- Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hacker Muller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316:1484–1488
- Kwon M (2007) Review: Tension wood as a model system to explore the carbon partitioning between lignin and cellulose biosynthesis in woody plants. *J App Biol Chem* 50:83–87
- Lu S, Sun Y-H, Shi R, Clark C, Li L, Chiang VL (2005) Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell* 17:2186–2203
- Lu S, Sun YH, Chiang VL (2008) Stress-responsive microRNAs in *Populus*. *Plant J* 55:131–151
- Lu T, Zhu C, Lu G, Guo Y, Zhou Y, Zhang Z, Zhao Y, Li W, Lu Y, Tang W (2012) Strand-specific RNA-seq reveals widespread occurrence of novel *cis*-natural antisense transcripts in rice. *BMC Genom* 13:721
- Lu S, Li Q, Wei H, Chang M-J, Tunlaya-Anukit S, Kim H, Liu J, Song J, Sun Y-H, Yuan L (2013) Ptr-miR397a is a negative regulator of *laccase* genes affecting lignin content in *Populus trichocarpa*. *Proc Natl Acad Sci USA* 110:10848–10853
- Lukiw WJ, Handley P, Wong L, Crapper McLachlan DR (1992) BC200 RNA in normal human neocortex, non-Alzheimer dementia (NAD), and senile dementia of the Alzheimer type (AD). *Neurochem Res* 17:591–597
- Maury S, Geoffroy P, Legrand M (1999) Tobacco *O*-methyltransferases involved in phenylpropanoid metabolism. The different caffeoyl-coenzyme a/5-hydroxyferuloyl-coenzyme a 3/5-*o*-methyltransferase and caffeic acid/5-hydroxyferulic acid 3/5-*o*-methyltransferase classes have distinct substrate specificities and expression patterns. *Plant Physiol* 121:215–224
- Mellerowicz EJ, Sundberg B (2008) Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties. *Curr Opin Plant Biol* 11:293–300
- Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10:155–159
- Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ (2008) Criteria for annotation of plant microRNAs. *Plant Cell* 20:3186–3190
- Nugroho WD, Yamagishi Y, Nakaba S, Fukuhara S, Begum S, Marsoem SN, Ko J-H, Jin H-O, Funada R (2012) Gibberellin is required for the formation of tension wood and stem gravitropism in *Acacia mangium* seedlings. *Ann Bot Lond* 110:887–895
- Pauli A, Valen E, Lin MF, Garber M, Vastenhout NL, Levin JZ, Fan L, Sandelin A, Rinn JL, Regev A (2012) Systematic identification of long noncoding RNAs expressed during Zebrafish embryogenesis. *Genome Res* 22:577–591
- Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. *Cell* 136:629–641
- Puzey JR, Karger A, Axtell M, Kramer EM (2012) Deep annotation of *Populus trichocarpa* microRNAs from diverse tissue sets. *PLoS ONE* 7:e33034
- Roberts A, Pimentel H, Trapnell C, Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27:2325–2329
- Saito R, Smoot ME, Ono K, Ruschinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T (2012) A travel guide to Cytoscape plugins. *Nat Methods* 9:1069–1076
- Schmittgen TD, Lee EJ, Jiang J, Sarkar A, Yang L, Elton TS, Chen C (2008) Real-time PCR quantification of precursor and mature microRNA. *Methods* 44:31–38
- Sigova AA, Mullen AC, Molinie B, Gupta S, Orlando DA, Guenther MG, Almada AE, Lin C, Sharp PA, Giallourakis CC (2013) Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *PNAS* 110:2876–2881
- Sjödin A, Street NR, Sandberg G, Gustafsson P, Jansson S (2009) The *Populus* genome integrative explorer (PopGenIE): a new resource for exploring the *Populus* genome. *New Phytol* 182:1013–1025
- Sun L, Zhang Z, Bailey TL, Perkins AC, Tallack MR, Xu Z, Liu H (2012) Prediction of novel long non-coding RNAs based on RNA-Seq data of mouse *Klf1* knockout study. *BMC Bioinform* 13:331
- Tafer H, Hofacker IL (2008) RNAplex: a fast tool for RNA–RNA interaction search. *Bioinformatics* 24:2657–2663
- Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS (2010) Non-coding RNAs: regulators of disease. *J Pathol* 220:126–139
- Timell T (1969) The chemical composition of tension wood. *Svensk Papp Tidn* 72:173–181
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc* 7:562–578
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604
- Wagner A, Donaldson L, Kim H, Phillips L, Flint H, Steward D, Torr K, Koch G, Schmitt U, Ralph J (2009) Suppression of 4-*coumarate-CoA ligase* in the coniferous gymnosperm *Pinus radiata*. *Plant Physiol* 149:370–383
- Wang W, Wang L, Chen C, Xiong G, Tan X-Y, Yang K-Z, Wang Z-C, Zhou Y, Ye D, Chen L-Q (2011a) *Arabidopsis CSLD1* and *CSLD4* are required for cellulose deposition and normal growth of pollen tubes. *J Exp Bot* 62:5161–5177
- Wang XT, Song XY, Glass CK, Rosenfeld MG (2011b) The long arm of long noncoding RNAs: roles as sensors regulating gene transcriptional programs. *CSH Perspect Biol* 3:a003756
- Yoon JH, Abdelmohsen K, Gorospe M (2013) Post-transcriptional gene regulation by long noncoding RNA. *J Mol Biol* 425:3723–3730
- Yoshida M, Yamamoto H, Okuyama T, Nakamura T (1999) Negative gravitropism and growth stress in GA3-treated branches of *Prunus spachiana* Kitamura f. *spachiana* cv. *Plenarosea*. *J Wood Sci* 45:368–372
- Zhang D, Du Q, Xu B, Zhang Z, Li B (2010) The actin multigene family in *Populus*: organization, expression and phylogenetic analysis. *Mol Genet Genomics* 284:105–119
- Zhang D, Xu B, Yang X, Zhang Z, Li B (2011) The sucrose synthase gene family in *Populus*: structure, expression, and evolution. *Tree Genet Genomes* 7:443–456
- Zhang J, Mujahid H, Hou Y, Nallamilli BR, Peng Z (2013a) Plant long ncRNAs: a new frontier for gene regulatory control. *Am J Plant Sci* 4:1038–1045
- Zhang Z, Zhu Z, Watabe K, Zhang X, Bai C, Xu M, Wu F, Mo Y (2013b) Negative regulation of lncRNA *GAS5* by *miR-21*. *Cell Death Differ* 20:1558–1568
- Zhong R, Morrison WH, Himmelsbach DS, Poole FL, Ye Z-H (2000) Essential role of caffeoyl coenzyme A *O*-methyltransferase in lignin biosynthesis in woody poplar plants. *Plant Physiol* 124:563–578