

RESEARCH ARTICLE

Open Access



# A transcriptomic profile of topping responsive non-coding RNAs in tobacco roots (*Nicotiana tabacum*)

Xi Chen<sup>1,2†</sup>, Shuo Sun<sup>1†</sup>, Fangjie Liu<sup>1,2</sup>, Enhui Shen<sup>1</sup>, Lu Liu<sup>1</sup>, Chuyu Ye<sup>1</sup>, Bingguang Xiao<sup>3</sup>, Michael P. Timko<sup>4</sup>, Qian-Hao Zhu<sup>5</sup>, Longjiang Fan<sup>1,2\*</sup> and Peijian Cao<sup>6\*</sup> 

## Abstract

**Background:** Non-coding RNAs (ncRNAs), including microRNAs (miRNAs), long ncRNAs (lncRNAs) and circular RNAs (circRNAs), accomplish remarkable variety of biological functions. However, the composition of ncRNAs and their interactions with coding RNAs in modulating and controlling of cellular process in plants is largely unknown. Using a diverse group of high-throughput sequencing strategies, the mRNA, miRNA, lncRNA and circRNA compositions of tobacco (*Nicotiana tabacum*) roots determined and their alteration and potential biological functions in response to topping treatment analyzed.

**Results:** A total of 688 miRNAs, 7423 non-redundant lncRNAs and 12,414 circRNAs were identified, among which, some selected differentially expressed RNAs were verified by quantitative real-time PCR. Using the differentially expressed RNAs, a co-expression network was established that included all four types of RNAs. The number of circRNAs identified were higher than that of miRNAs and lncRNAs, but only two circRNAs were present in the co-expression network. lncRNAs appear to be the most active ncRNAs based on their numbers presented in the co-expression network, but none of them seems to be an eTM (endogenous Target Mimicry) of miRNAs. Integrated with analyses of sequence interaction, several mRNA-circRNA-miRNA interaction networks with a potential role in the regulation of nicotine biosynthesis were uncovered, including a QS-circQS-miR6024 interaction network. In this network miR6024 was significantly down-regulated, while the expression levels of its two targets, circQS and its host gene QS, were sharply increased following the topping treatment.

**Conclusions:** These results illustrated the transcriptomic profiles of tobacco roots, the organ responsible for nicotine biosynthesis. mRNAs always play the most important roles, while ncRNAs are also expressed extensively for topping treatment response, especially circRNAs are the most activated in the ncRNA pool. These studies also provided insights on the coordinated regulation module of coding and non-coding RNAs in a single plant biological sample. The findings reported here indicate that ncRNAs appear to form interaction complex for the regulation of stress response forming regulation networks with transcripts involved in nicotine biosynthesis in tobacco.

**Keywords:** Co-expression network, Circular RNA, Long non-coding RNA, microRNA, *Nicotiana tabacum*, Topping, Transcriptomic profile

\* Correspondence: fanlj@zju.edu.cn; peijiancao@163.com; fanlj@zju.edu.cn

†Xi Chen and Shuo Sun contributed equally to this work.

<sup>1</sup>Institute of Crop Science, Zhejiang University, Hangzhou 310058, China

<sup>6</sup>Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China

Full list of author information is available at the end of the article



## Background

Non-coding RNAs (ncRNAs) consist of multiple classes of RNA molecules that are generated by genomic transcription but are not translated into proteins. They include microRNA (miRNA) with ~20–24 nucleotides, long ncRNA with a length > 200 bases, and circRNA that is a covalently closed RNA molecule without 5' cap and poly-A tail formed by back splicing of precursor RNA derived from exons (exonic circRNA), introns (intronic circRNA), or intergenic regions [1–4]. Many studies have looked into the crosstalk between different types of RNA molecules [1, 5, 6]. For example, lncRNA and circRNA may function as miRNA sponges to modulate the activities of both miRNA and its target genes [4, 7]. Noncoding endogenous target mimics (eTMs) for diverse miRNAs have been identified in *Arabidopsis thaliana*, *Oryza sativa* and *Glycine max*, and some of them have been shown to regulate plant development by repressing miRNA function [8–10]. Wang et al. (2017) identified six differentially expressed wheat circRNAs that respond to dehydration stress and play a role of putative miRNA sponges [11]. Increased evidences for broad functionality of ncRNAs and demonstration of their changing composition and interactions with mRNAs under different conditions in cells well illustrate the crosstalk between different types of RNAs and its importance in plant cellular metabolism. However, no attempt has been made to profile the four types of RNAs and to uncover their interactions in a single plant organ, such as tobacco roots producing nicotine.

The pyridine alkaloid nicotine is the signature compound of *Nicotiana* species, especially in tobacco, responsible for the addiction of smoking and functioning as one of the most effective plant defense metabolites in nature. In tobacco, nicotine is exclusively synthesized in young root tips and subsequently transported to leaves. Nicotine biosynthesis can be induced through topping for tobacco plants self-protection and significant increase of nicotine content can be observed in the topped tobacco plants [12]. The genes involved in nicotine biosynthesis and regulation, such as those encoding quinolinate synthase (QS), quinolinic acid phosphoribosyl transferase (QPT) and putrescine methyltransferase (PMT) have been intensively studied [12–15]. Transcription factors involved in the regulation of nicotine biosynthesis have also been reported, for example, *NIC1/NIC2* was shown to be a cluster of *AP2/ERFs* mediates nicotine formation [16]. A few studies have also investigated the composition and function of ncRNAs in tobacco, with the first topping-responsive miRNA identified in 2011 [17]. Shortly after, Tang et al. (2012) predicted 59 novel tobacco-specific miRNAs using small RNAs from roots of wounding- and topping-treated tobacco plants [18]. With the advent of genomic sequencing techniques

and high efficiency bioinformatic analysis, genome-wide scan for ncRNAs has become feasible. Gao et al. (2015) performed genome-wide identification of miRNAs after the releasing of the tobacco genome in 2014 and identified 165 miRNAs, including miRNA members from 55 conserved and 50 novel families, in leaves, stems and roots [19]. Recently, Li et al. (2018) identified 5206 tobacco lncRNAs based on RNA sequencing data, with 565 involved in nematode stress response [20]. ncRNAs have also been implicated in regulation of nicotine biosynthesis in tobacco. For instance, miR164 has been demonstrated to be a negative regulator of *NtNAC-RI*, a unique tobacco NAC (for NAM, no apical meristem) transcription factor responsive to topping and resulted in increase of lateral roots and nicotine contents [21]. In our previous study, a new miRNA, nta-miRX27, was shown to target *QPT2*, one of the key genes involved in nicotine biosynthesis; furthermore, targeting of *QPT2* by nta-miRX27 is attenuated by nta-eTMX27, a lncRNA functioning as an endogenous target mimicry (eTM) to compete against nta-miRX27 for *QPT2* [22]. These results led us to conclude that nicotine biosynthesis is regulated by both miRNAs and lncRNAs. Nevertheless, it is still unclear whether or not circRNAs are also one of the components of the RNA network related to nicotine biosynthesis in tobacco roots.

In this study, four types of RNA molecules (mRNA, miRNA, lncRNA, circRNA) in a single plant organ using high-throughput sequencing strategies were simultaneously characterized for the first time. On the base of identification of the four types of RNAs, their expression levels and abundance were quantified and compared based on a unified normalization approach. A co-expression network for the four types of RNAs was established and a few mRNA-circRNA-miRNA interaction networks were identified, including QS-circQS-miR6024, involved in the regulation of nicotine biosynthesis. The results presented in this study provide launch-pad for further functional characterization of both coding and non-coding RNAs in nicotine metabolism in tobacco.

## Results

### Sequencing and identification of four types of RNAs from a single organ: tobacco roots

*Nicotiana tabacum* cv. K326 was selected for transcriptomic (both coding and noncoding RNAs) profiling. RNA-seq, circRNA-Seq, small RNA-Seq and ssRNA-Seq were performed using roots from both topping-treated (i.e., removal of the floral head and upper young leaves) and untreated tobacco plants. In total, 102.66 Gb of mRNA, 148.86 Gb of circRNA, 25.1 Gb of small RNA and 71.46 Gb of lncRNA raw data were generated (Table 1). The number of the four types of RNA molecules identified and used in this study were shown in

**Table 1** Summary of raw sequencing data (Gb) from the topping treatment (TT) and control (CK) experiments

Sequencing methods	Control (CK)				Topping (TT)			
	Rep1	Rep2	Rep3	Total	Rep1	Rep2	Rep3	Total
RNA-Seq	17.74	16.79	16.02	50.55	17.71	17.37	17.03	52.11
circRNA-seq	20.30	22.44	25.52	68.26	24.61	29.24	26.75	80.6
miRNA-Seq	4.10	3.60	4.60	12.30	5.30	4.50	3.00	12.80
ssRNA-Seq	12.18	12.12	11.98	36.28	11.14	12.11	11.93	35.18

**Table 2.** More than 90% of the 85,570 gene models annotated by Siervo (2014) were supported by transcripts found in our RNA-seq data, with 77,849 and 78,467 genes expressed in the topping-treated and untreated roots, respectively. In this study, a total of 688 miRNAs were inferred, with 268 conserved miRNAs from 38 families and 420 novel miRNAs from 393 families (Additional file 1: Figure S1A). Collected with miRNAs deposited in the miRBase (<http://www.mirbase.org>, V22) and the ones identified by Li (2015), the total number of tobacco miRNAs analyzed in this study increased to 981 (Table 2). The majority predicted targets of the conserved miRNAs were transcription factors (TF). For example, miR156 and miR167 were predicted to target to squamosa promoter-binding-like (*SPL*) and auxin response factor (*ARF*), respectively, same as that reported in *Arabidopsis thaliana* [23, 24]. miR166 targeting HD-Zip transcription factors found in *Brassica napus* was also identified in tobacco roots [25]. For lncRNAs, in order to have an as complete as possible repository to be used in further analysis, the 6 poly(A) RNA-seq and the 6 ssRNA-seq datasets generated in this study and 82 poly(A) RNA-seq datasets downloaded from NCBI were used for lncRNA identification. As a result, 3635 (3688 transcripts) and 4121 non-redundant lncRNA loci (5352 transcripts) were identified from ssRNA-seq and poly(A) RNA-seq datasets, respectively (Additional file 2: Figure S2A). Only about 10% of the lncRNA loci were identified in both datasets (Additional file 2: Fig. S2B). A much higher proportion of single exon transcripts was found in the ssRNA-seq datasets (86.9%, 3206 transcripts) compared to the poly(A) RNA-seq datasets (4.0%, 214 transcripts) (Additional file 2: Figure S2C). These results confirmed our previous observation in oilseeds [26], suggesting that it is necessary to use both library creating and sequencing methods in lncRNA

identification in order to increase the number of lncRNAs identified. circRNA is an emerging star of ncRNA in plants. By integrating the prediction results from three bioinformatic tools (CIRI, find\_circ and PcircRNA-finder), 13,957 circRNA transcripts generated from 8244 host genes were identified. Most of which (11,782, 84.42%) were generated from exons of protein-coding genes, only 1085 (7.77%) were intergenic circRNAs and 7.81% were intronic circRNAs (Additional file 3: Figure S3A). Only around 5% circRNAs were detected by all three tools (Additional file 3: Fig. S3B) probably due to variable sensitivity of the prediction tools towards different types of circRNAs [27]. The general features of the identified tobacco circRNAs consist with those identified in other plants [28, 29], such as more than 80% of exonic circRNAs harbored 1–4 exons (Additional file 3: Figure S3C) and most genes produced only one isoform of circRNAs (Additional file 3: Fig. S3D). However, compared to other plants, tobacco has more circRNAs containing more than 7 exons.

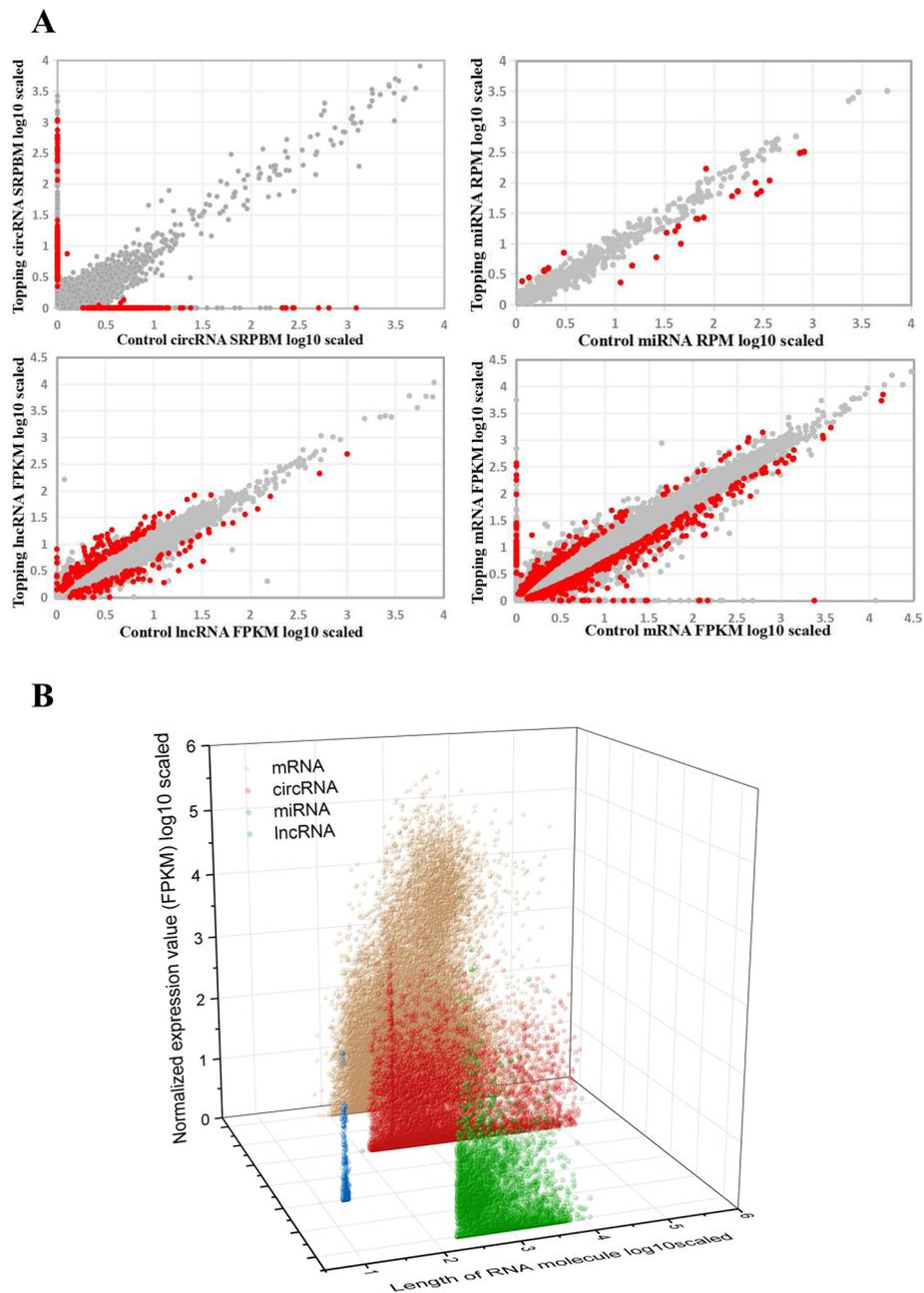
#### Expression profiling of coding RNAs and three types of ncRNAs in tobacco roots

Expression levels of mRNA and lncRNA were quantified using FPKM, while SRPBM and RPM were used to determine the expression levels of circRNA and miRNA, respectively. In the topping-treated sample, 1058 significantly differentially expressed genes (DEGs) ( $|\log_2(\text{fold-changes})| \geq 1$ ,  $\text{FDR} < 0.05$ ) were identified. As expected, most genes, such as *QPT2*, *QS* and *PMT2*, of the nicotine biosynthesis pathway were significantly up-regulated in response to topping. For ncRNAs, 315 (159 up-regulated and 156 down-regulated), 41 (15 up-regulated and 26 down-regulated) and 295 (178 up-regulated and 117 down-regulated) differentially expressed circRNAs, miRNAs and lncRNAs were identified, respectively (Fig. 1a). Nine mRNAs involved in nicotine biosynthesis and 10 mRNAs related to nicotine degradation, transportation, or topping responsive TFs were selected for qRT-PCR validation (Fig. 2a). All genes related to nicotine biosynthesis (*QPT2*, *QS*, *PMT2*) or transportation (*MATE1*, *MATE2*, *A622*) were confirmed to be significantly up-regulated, while topping responsive TFs like *NAC148* and *MYC2* were confirmed to be significantly down-regulated (Fig. 2a). Besides, the differential expression patterns of 6 of the 7 selected miRNAs, including 3 targeting nicotine biosynthesis genes, were also verified by qRT-

**Table 2** Number of coding and three types of non-coding RNAs identified in tobacco

Type	Loci	Transcripts	Identified by
mRNA	85,570	145,503	Siervo et al., 2014 (79,821 expressed in this study)
lncRNA	7423	8520	Only this study
miRNA*	981	981	This study, Li et al., 2015 and miRBase V22
circRNA	12,414	13,957	Only this study

\* 688 miRNAs were identified in this study



**Fig. 1** Expression profiling of mRNAs, circRNAs, miRNAs and lncRNAs in tobacco. **a** Normalized expression values of mRNAs (FPKM), circRNAs (SRPBM), miRNAs (RPM) and lncRNAs (FPKM) in control vs. topping-treated. Red and gray points represent significantly differentially expressed RNAs and not differentially expressed RNAs, respectively; **b** Normalized expression values (FPKM) of different length of mRNAs, circRNAs, miRNAs and lncRNAs in tobacco after topping treatment

PCR (Fig. 2b). qRT-PCR verification was also conducted for 12 lncRNAs and 17 circRNAs, including 15 circRNAs generated from nicotine biosynthesis genes. The expression changes in response to the topping-treatment were verified for 2 up-regulated and 10 down-regulated lncRNAs (Fig. 2d). For the circRNAs, the topping-induced expression patterns were verified for 9 out of the 17 selected ones, just as their topping-induced parental genes (Fig. 2c). Several

down-regulated circRNAs generated from *MATE1* and *QS* according to the circRNA-seq data failed to be detected.

In order to effectively compare the expression level of the four types of RNAs, a unified normalization standard is required. The expression level of circRNA and miRNA was converted into FPKM and plotted FPKM of the four types of RNAs based on their length (Fig. 1b). The mean expression levels of mRNAs, circRNAs, miRNAs and

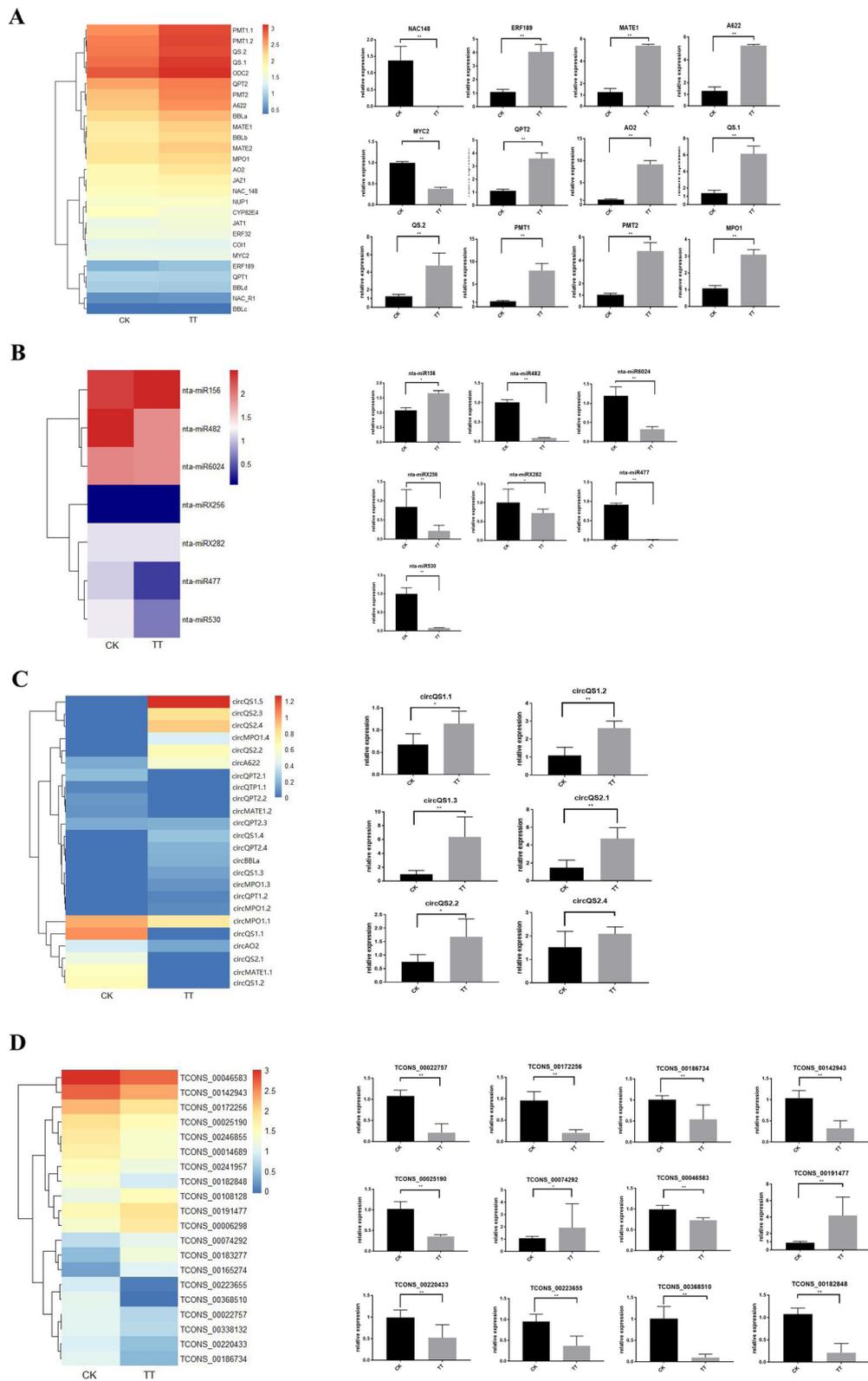


Fig. 2 (See legend on next page.)

(See figure on previous page.)

**Fig. 2** Normalized expression levels of 4 types of topping responsive RNA molecules and verified by qRT-PCR. **a** Normalized expression level of 27 nicotine biosynthesis and metabolism related mRNAs based on RNAseq data and partial RT-qPCR results. **b** Normalized expression level of 7 topping response miRNAs on RNAseq data and partial RT-qPCR results. **c** 24 circRNAs generated from nicotine biosynthesis and metabolism related genes on RNAseq data and partial RT-qPCR results. **d** Normalized expression level of 20 topping response lncRNAs based on RNAseq data and partial RT-qPCR results. Data are means  $\pm$  SD ( $n = 3$ ) \* $P < 0.05$ , \*\* $P < 0.01$  (One-way ANOVA followed by LDS test)

lncRNAs were 252.38, 7.51, 4.00, and 15.11, respectively, suggesting that ncRNAs were expressed at much lower levels than protein coding genes in tobacco roots. However, of all the RNAs detected in this study, both the most significantly down-regulated ( $\log_2|FC| = -18.8508$ ) and up-regulated ( $\log_2|FC| = 19.324$ ) transcripts in the topping-treated sample were lncRNAs. The transcript length of the expressed RNAs was also compared. The most highly expressed circRNAs have a length of 300 bp - 2000 bp even though the average circRNA length was around 5 kb, while the most highly expressed mRNAs and lncRNAs have a length of 2000 bp - 8000 bp and 200 bp - 1000 bp, respectively.

#### Co-expression of the four types of RNA molecules responsive to topping in tobacco roots

Most ncRNAs have not been annotated and an unknown function. To identify potential roles of the ncRNAs, a co-expression network for the four types of RNA molecules was constructed by calculating Pearson's correlation coefficient value (the absolute value of PCC  $> 0.99$ ,  $p$ -value  $< 0.05$  and FDR  $< 0.05$ ). From the network, we could see that coding and noncoding RNA molecules may affect plant defensive mechanism in various ways by interacting with each other.

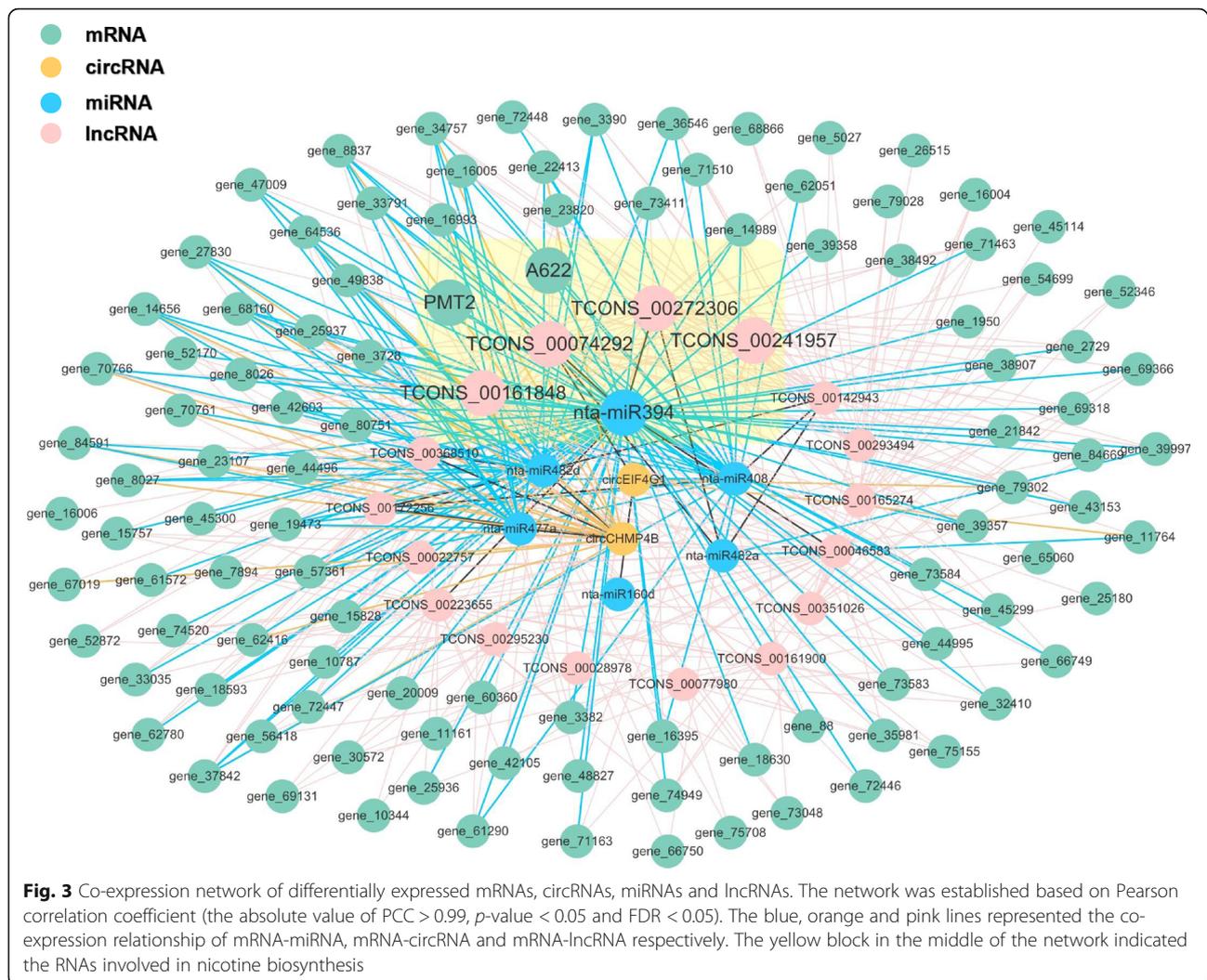
As shown in Fig. 3, after applying stringent filtering steps 6 conserved miRNAs were found in the co-expression network. Two miRNAs (miR482a and miR482d) of the 6 miRNAs identified belong to the miR482 family. miR482a/d had co-expression patterns with 27 mRNAs, including several targeting *NBS* genes as previously reported [30]. miR482a/d, miR394, miR408 and miR477a were co-expressed not only with mRNAs but also with lncRNAs. Both miR482d and miR477a also had a positive correlation with circCHMP4B. miR160d showed only negative correlation with circEIF4G1 generated from *EIF4G1*. circCHMP4B and circEIF4G1 were the only two circRNAs included in the network, since most co-expressed pairs between circRNA and other types of RNA molecules were not able to be connected to a third type of RNA. However, a larger number of lncRNAs (18) was presented in the network and showed a co-expression pattern with 106 mRNAs as well as the aforementioned 6 miRNAs and 2 circRNAs. With the network, it was not only possible to assess the function of ncRNAs by means of well-studied and co-expressed protein-coding mRNAs, but also to deduce the ncRNAs related to mRNAs of interest.

Closer examination of the genes of the nicotine biosynthesis pathway showed that *A622* and *PMT* were present in this co-expression network and that miR394 showed a negative correlation with these two genes. *PMT* also had a positive co-expression relationship with lncRNA TCONS\_00161848 and a negative co-expression relationship with lncRNAs TCONS\_00241957 and TCONS\_00272306. *A622* was also negatively co-expressed with TCONS\_00272306, although it had a positive co-expression relationship with another lncRNA (TCONS\_00074292). circCHMP4B and circEIF4G1 were present in the network, but their parental genes (*CHMP4B* and *EIF4G1*) were not related to nicotine metabolism, thus no putative circRNA-containing network related to nicotine biosynthesis was identified in this study. Nevertheless, it was possible to find 5 nicotine biosynthesis related co-expression networks containing miRNA, mRNA and lncRNA (i.e., miR394-A622-TCONS\_00272306, miR394-PMT2-TCONS\_00272306, miR394-A622-TCONS\_00074292, miR394-PMT2-TCONS\_00241957 and miR394-PMT2-TCONS\_00161848).

#### miRNA regulation network

To further provide convincing evidence for the crosstalk among miRNA, lncRNA, circRNAs and mRNA, the differential expression ( $|\log_2FC| \geq 1$ ) pattern of the RNAs involved and their physical interaction potential were checked.

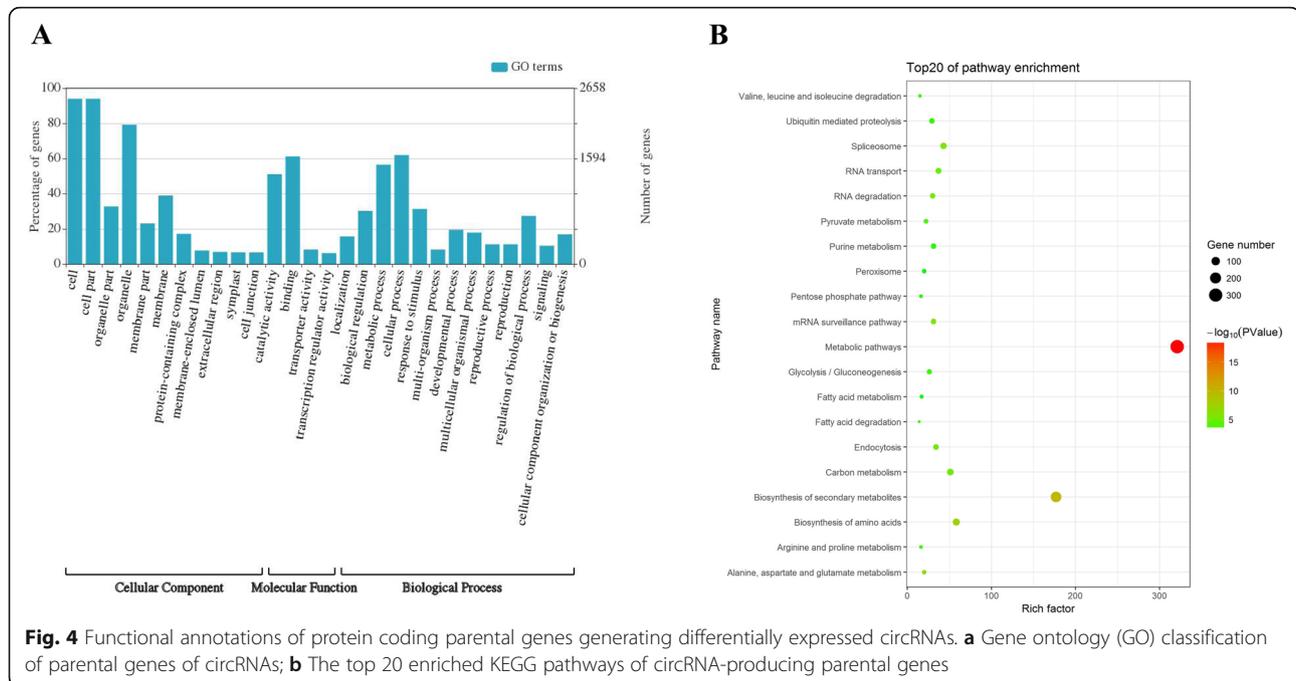
Based on target prediction, 678 mRNAs were predicted potentially regulated by 114 miRNAs, including the miRX27-*QPT2* pair studied in our previous work [22]. Many miRNA-targeted mRNAs were reported to be stress associated. The most significantly down-regulated miRNA identified in this study was miR530, which targets *KIP1* that was up-regulated after the topping treatment. Upon the topping treatment, miRX393 was significantly up-regulated while its target *CYP450* was down-regulated. Many circRNAs were predicted to having interactions with miRNAs. Linear regression analysis indicated that miRNAs had negative expression correlation with their predicted target genes ( $R^2 = 0.52$ , Additional file 4: Figure S4A) and target circRNAs ( $R^2 = 0.72$ , Additional file 4: Figure S4B). Interestingly, most miRNAs (88.2%) targeting circRNAs were up-regulated after the topping treatment. Majority of these miRNAs are conserved ones, for example, 40 of the 114 miRNAs involved in the interaction between miRNA and mRNA,



and 23 of the 36 miRNAs involved in the interaction between miRNA and circRNA were conserved. Some well-known miRNA families like miR156, miR482 and miR477 were found in both interactions, particularly miR156, more than 10 members were found to be involved in the two interactions. As for lncRNA-miRNA interaction, differentially-expressed miRNA and lncRNA that was predicted to be eTM of the miRNA were analyzed, and no statistical relationship could be concluded. The relationship between circRNAs and their parental genes was analyzed using the genes generating a single circRNA isoform. Forty-nine (49) circRNA-mRNA pairs with a highly positive co-expression relationship were found ( $R^2 = 0.34$ , Additional file 4: Figure S4C). The most significantly differential expressed circRNA was circLASPO, which was co-upregulated after the topping treatment with its parental gene LASPO (L-aspartate oxidase), a gene related to the production of aspartate.

### GO and KEGG analysis of parental genes generating differentially expressed circRNAs

The composition and potential function of circRNAs have not been previously reported in tobacco. Here GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses for the parental genes of differentially expressed circRNAs ( $|\log_2FC| \geq 1$ ) were carried out to understand the topping responsive pathways potentially regulated by circRNAs. Three GO categories were assigned to the 5147 parental genes of 7235 circRNAs (Fig. 4a). In the biological process category, “metabolic process”, “cellular process” and “regulation of biological process” were among those significant GO subcategories. Among them, GO: 0050896 (Response to stimulus) was significantly enriched for 673 mRNAs. According to KEGG analysis, the most highly represented pathways were those related to biosynthesis of secondary metabolites (Fig. 4b), implying that many parental genes of the



**Fig. 4** Functional annotations of protein coding parental genes generating differentially expressed circRNAs. **a** Gene ontology (GO) classification of parental genes of circRNAs; **b** The top 20 enriched KEGG pathways of circRNA-producing parental genes

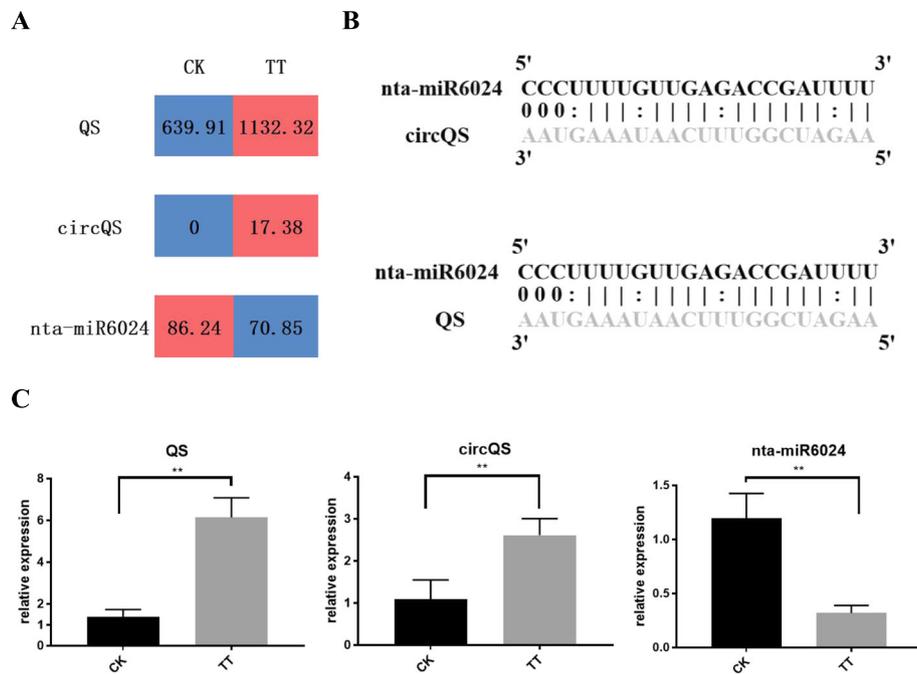
differentially expressed circRNAs in the topping-treated tobacco roots are involved in synthesis of secondary metabolites, including nicotine.

### The regulation network of the three types of RNA molecules involved in the nicotine biosynthesis pathway

Based on BLASTp search using known genes [31], 27 tobacco genes related to nicotine biosynthesis and metabolism were identified (Fig. 2a). From 10 of these genes, 24 exonic-circRNAs and one intronic circRNA were identified (Fig. 2c). Some genes generate multiple circRNAs, such as *QS.1* producing 5, and *QPT2*, *MPO1* and *QS.2* each producing 4. For the nicotine biosynthesis related miRNAs, in addition to those identified in our previous study [22], one conserved (miR6024) and 6 novel miRNAs were newly found to target 5 of the 27 genes related to nicotine biosynthesis (Fig. 2b). As the interaction principle between lncRNAs and other types of RNA molecules is still under debate and, thus, no prediction attempt was made, this study therefore focused on the construction of miRNA-circRNA-mRNA networks involved in regulation of nicotine biosynthesis.

Fourteen (14) out of the 27 genes related to nicotine biosynthesis were significantly up-regulated after the topping treatment (Fig. 2a). From 8 of those gene significantly up-regulated, 16 differentially expressed circRNAs ( $|\log_2FC| \geq 1$ , adjusted  $P$ -value  $\leq 0.05$ ) were identified based on analysis of circRNA-seq data. The differential expression pattern was verified by qRT-PCR for 9 of the 16 circRNAs (generated from 4 parental genes). Based

on the expression analysis and identification of the interaction sites, three networks involving mRNA, miRNA and circRNA were uncovered. *QS*, that contributes to both the nicotine and NAD pathways, was involved in the network *QS*-circ*QS*-miR6024 (Fig. 5). The target sites of miR6024 were identified in both *QS* and circ*QS*, which were significantly up-regulated while their repressor miR6024 was down-regulated (Fig. 5c). Due to *QS* and circ*QS* having identical target sites of miR6024, their repression by miR6024 would be competitive. Of the components of the network *AO2*-circ*AO2*-miRX282, *AO2* gene was significantly up-regulated after the topping treatment, while circ*AO2* and miRX282 were slightly down-regulated (Fig. 2b, Fig. 2c). In the third network *QPT2*-circ*QPT2*-miRX256, the miRNA did not target circ*QPT2* but *QPT2*. The expression level of *QPT2* was significantly up-regulated after the topping treatment while that of miRX256 was quite low and down-regulated (Fig. 2b). Multiple circRNAs were identified in *QPT2*. Two (circ*QPT2.1* and circ*QPT2.2*) were only identified in the control while the third circRNA (circ*QPT2.3*) was only found in the topping treatment. All three circRNAs had a low expression level. Of the three networks, *QS*-circ*QS*-miR6024 showed the highest expression levels of its components (Fig. 5c and Fig. 2), implying its importance in regulating nicotine biosynthesis. Conclusively, as shown in Fig. 6, coding genes were not the only determinant in nicotine biosynthesis in tobacco root, the interaction with non-coding RNAs, such as circRNAs and miRNAs, also extensively involved



**Fig. 5** The putative regulation network of non-coding RNAs involved in QS gene in nicotine pathway. **a** Normalized expression levels of QS, circQS and nta-miR6024 based on the RNA sequencing data; **b** Alignment of nta-miR6024 and its target RNAs. Base pairing between miRNA and its targets is shown, in which a vertical line means a Watson-Crick pair, two dots represent a G-U pair, and 0 means a mismatch; **c** qRT-PCR analyses of QS, circQS and nta-miR6024 in roots from the topping-treated and control plants

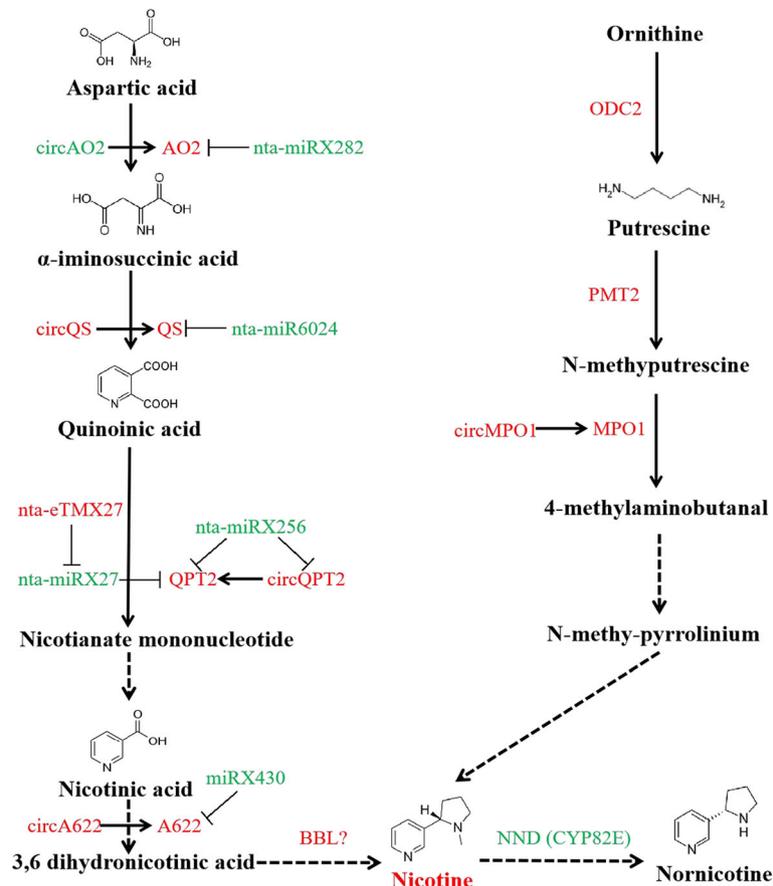
in it. To understand the regulation mechanism of non-coding RNAs involved in nicotine biosynthesis, further efforts are still needed.

## Discussion

Crosstalk and functional activities of ncRNAs and coding RNAs in plants are organized in complex modes. Interaction studies so far in plants mostly focused on two types of RNAs, such as miRNAs and their mRNA targets [1], circRNAs and their parental genes [32], and miRNAs and their eTM lncRNAs [2]. Regarding regulation of nicotine biosynthesis and other secondary metabolism in tobacco, a role of ncRNAs has rarely been demonstrated [33]. The present work investigated the dynamic changes of mRNAs, miRNAs, circRNAs and lncRNAs, and their interactions in response to topping treatment in tobacco, with a focus on coding and non-coding RNAs related to nicotine biosynthesis. Our results provide evidence for the involvement of ncRNAs and their interactions with coding genes in nicotine biosynthesis and the foundation for further functional characterization of the networks involving both coding and non-coding RNAs in tobacco.

The most abundant transcripts in tobacco were coding RNAs. Among the three types of ncRNAs, circRNA was the predominant one probably due to the high frequency of

alternative back splicing of linear genes in tobacco [34], and miRNA was the least abundant one. A large number of tobacco circRNAs were identified in this work, but *A. thaliana* or *O. sativa* have approximately 3 times more circRNAs than tobacco, even though tobacco has a much larger genome size and encodes many more genes (Additional file 5: Table S1). This difference could partially be the result of the many different bioinformatics tools having been used in circRNA identification in *A. thaliana* or *O. sativa* [29]. As the limitation of the circRNA enrichment technology and no single method dominated on all of the metrics such as precision and sensitivity to date, use of different bioinformatic tools would compensate the drawbacks of each individual tool to get comprehensive results. Some lowly expressed circRNAs failed to be constantly identified also lead to false negative results. A total of 688 miRNAs were identified in this study, including 268 conserved ones from 38 families and 420 novel miRNAs from 393 families. A significant portion of these miRNAs were newly identified in this study, which provided an important complementation for the previously reported ones [17–19, 22]. miRNAs are usually much more conserved than other ncRNAs. Nearly 39% of the miRNAs identified in this study were conserved ones (Additional file 1: Figure S1A). The abundance of lncRNAs identified in tobacco roots were also investigated and compared with the lncRNA number in *A.*



**Fig. 6** The potential interaction of RNA molecules involved in the nicotine biosynthesis pathway in tobacco root. The differential expressed circRNAs which generated from parental genes involved in nicotine biosynthesis pathway were presented. The differential expressed miRNAs which target to genes or circRNAs were also shown in the figure. The nicotine biosynthesis pathway diagram drew based on previous studies in tobacco. The up-regulated genes/non-coding RNAs after topping treatment were presented in red, while down-regulated ones were in green

*thaliana* and *O. sativa* [35, 36]. It showed a positive correlation between the abundance of lncRNAs and the complexity of plant genome.

This study, for the first time in plants, constructed a lncRNA-miRNA-circRNA-mRNA co-expression network based on differentially expressed RNAs with stringent filters (the absolute value of PCC > 0.99, *p*-value < 0.05 and FDR < 0.05). Interestingly, even though circRNAs were the predominant ncRNAs, they were the least ones with a co-expression relationship with mRNAs, instead lncRNA-mRNA was the most favorable co-expression pair (Fig. 3). The phenomenon can be explained that a large amount of circRNAs was only identified specifically in untreated or topping group. Most miRNAs (88.2%) targeting circRNAs were up-regulated after the topping treatment, and consequently the expression levels of all the targeted circRNAs were decreased. This hinted us that miRNAs could shift the balance between mRNA and circRNA towards mRNA to activate the cell self-protection. Nevertheless, most of the significantly differentially expressed circRNAs

exhibited an expression profile similar to their parental genes (Additional file 4: Figure S4C), suggesting a role of plant circRNAs in regulation of their parental genes [32]. Only two circRNAs represented in the co-expression network, largely due to the low expression level of circRNAs and their unique to one of the two samples used. This feature of circRNAs is quite different from that of other two types of ncRNAs, because 5 (miR394-A622-TCONS\_00272306, miR394-PMT2-TCONS\_00272306, miR394-A622-TCONS\_00074292, miR394-PMT2-TCONS\_00241957 and miR394-PMT2-TCONS\_00161848) co-expressed miRNA-mRNA-lncRNA modules were found. The conserved miR394 targeting F-box genes [37] was the only RNA molecule involved in all the 5 modules. Both A622 and PMT2 are not target genes of miR394, suggesting that co-expression of miR394 with A622 or PMT2 could be achieved through the lncRNAs of the modules. The majority nodes of the co-expression network were mRNAs, ncRNAs occupied only a small number of

nodes of the co-expression network, suggesting that ncRNAs-mediated regulation of cellular processes might be mainly achieved by interacting with mRNAs rather than with other ncRNA molecules.

Co-expression analysis uncovered the topping-responsive profile of ncRNAs and coding RNAs, interaction analysis further provided us the knowledge about the RNA crosstalk in tobacco roots. Among the miRNA-mRNA interaction pairs, 41 miRNAs were significantly differentially expressed in response to the topping treatment. Most (31) of these 41 were conserved miRNAs, including those that have been reported to be topping or wounding responsive, such as miR160, miR164, miR394, miR408, miR477 and miR1919 [18, 22]. For instance, miR164 and its target *NtNAC-R1* was down- and up-regulated in response to topping, respectively, consistent with the results reported previously [21]. The target genes of miR160, miR394, miR408, miR477 and miR482 generated large amount circRNAs, but only circRNAs from *CHMP4B* and *EIF4G1* were found to form the networks with three of these miRNAs (i.e., *CHMP4B*-miR477/482-circ*CHMP4B* and *EIF4G1*-miR160-circ*EIF4G1*).

Although *CHMP4B* and *EIF4G1* are not nicotine biosynthesis related genes, it was possible to find 3 networks involving mRNAs, circRNAs and miRNAs with a potential regulatory role in nicotine biosynthesis. In two of these networks, *QS*-circ*QS*-miR6024 and *AO2*-circ*AO2*-miRX282, miR6024 and miRX282 have target sites on both mRNA (*QS* and *AO2*, respectively) and circRNA (circ*QS* and circ*AO2*, respectively). Because each miRNA has the identical targets in the circRNA and its parental gene, binding of the miRNA to each of its two targets would be competitive. In view of the negative correlation between miRNA and its targets, it is expected that binding of the miRNA to circRNA would relax the miRNA-mediated cleavage of its target mRNA; therefore, the role of the circRNA is similar to that of eTM, i.e. function as the sponge of miRNAs but probably in a cleavage manner. This relationship among miRNA and its target mRNA and circRNA from the target mRNA suggests that the expression of *QS* and *AO2*, consequently nicotine biosynthesis, is under fine regulation mediated by miRNAs.

In summary, our comprehensive analyses of mRNAs and ncRNAs (miRNAs, lncRNAs, circRNAs) in tobacco roots and identification of coding and non-coding RNAs and their interaction networks in responsive to topping treatment enabled us to draw an overall picture of the transcriptome profile of all the four types of RNA molecules and offered us an insight into the crosstalk amongst these RNAs in tobacco.

## Methods

### Plant materials and treatment

All samples were collected from tobacco (*Nicotiana tabacum*) 'K326' root. Tobacco plants were grown at

25°C and 65% humidity in a growth chamber to reduce external environmental effects on the biosynthesis of nicotine alkaloids with a day-night cycle of 16 h light and 8 h darkness. At least three 50-d-old (days after seeding) plants were used for topping treatment (i.e., removal of the floral head and the upper young leaves of the tobacco plants), and plants of a similar size without topping treatment were used as control. After topping treatment, the plants were kept for another 48 h in the growth chamber before sample collection.

### Library preparation and Illumina sequencing

Total RNAs were isolated from root samples of topping-treated and untreated plants (each treatment with three biological replicates) using MiniBEST Plant RNA Extraction kit (TAKARA) following the manufacturer's procedure. Total RNA concentration and purity were assessed by Nanodrop ND 2000 (Thermo Fisher Scientific).

For mRNA sequencing library preparation, mRNA was enriched and purified with oligo (dT)-rich magnetic beads and then broken into short fragments. Six libraries (three libraries each for the topping-treated and untreated samples, same for miRNA-seq, circRNA-seq and lncRNA-seq) were generated and sequenced using an Illumina HiSeq™ 2500 platform (BIOMARKER, China). miRNA libraries were prepared by TruSeq Small RNA Sample Prep Kits (Illumina) and sequenced in a single-end (50 bp) manner on an Illumina HiSeq 2500 platform following the vendor's recommended protocol. lncRNA and circRNA libraries were created using the mRNA-Seq sample preparation kit (Illumina) after depletion of ribosomal RNAs by the Ribo-Zero Gold Kit (Illumina). Paired-end sequencing was then performed on an Illumina HiSeq 4000 and Illumina HiSeq X for lncRNA-seq and circRNA-seq, respectively.

### Computational identification of circRNAs

The adaptors and low-quality reads were removed from the sequencing raw data using Trimmomatic v 0.36 [38]. Filtered clean reads were mapped to the *N. tabacum* reference genome (Sierro et al., 2014) using Bowtie2 v 2.2.6 [39], TopHat v 2.1.1 [40] and STAR v 2.5.2 [41] to get different format of mapping results. All the unmapped reads were then used as input files to CIRI v 2.0.6 [42], find\_circ v 1.0 [43] and PcircRNA\_finder [44] for back-spliced junction and circRNAs identification.

### Prediction of miRNAs in tobacco

The clean reads were first compared with the annotated non-coding RNA sequences, including plant snoRNA (version 1.2; [http://bioinf.scri.sari.ac.uk/cgibin/plant\\_snorna/](http://bioinf.scri.sari.ac.uk/cgibin/plant_snorna/)), tRNA (<http://gtrnadb.ucsc.edu/>), rRNA (<http://rfam.xfam.org/>) and rasiRNA (<http://www.girinst.org/>) to remove those matching to these known non-coding RNAs.

The remaining clean reads were then aligned to the *N. tabacum* genome (Sierro et al., 2014) for identification of miRNAs. Mireap (<http://sourceforge.net/projects/mir-eap/>) was applied to predict secondary hairpin structures of miRNAs with the following parameters: the hairpin structure has a free energy lower than  $-18 \text{ kcal mol}^{-1}$ ; the space between miRNA and miRNA\* is less than 300 nt; miRNA and miRNA\* have at least 16 matched nucleotides and fewer than four unmatched nucleotides. Only small RNAs with at least two reads in a library were used for miRNA prediction. A miRNA was considered as conserved if its mature sequence had  $\leq 2$  mismatches with the miRNAs deposited in the miRBase (<http://www.mirbase.org>, release 21), otherwise the miRNA was considered as a novel miRNA [45].

### Computational identification of lncRNAs

All the raw reads from transcriptome sequencing were treated using Trimmomatic v 0.36 [38] with the default parameters for quality control. The clean data were then mapped to the *N. tabacum* genome using Tophat v 2.1.1 [40]. For each mapping result, Cufflinks v 2.1.1 [46] was used in transcript assembly. For strand-specific RNA-seq datasets, the parameter "--library-type fr-firststrand" was employed. All transcriptomes were then merged based on the annotation file of the reference genome to generate a final transcriptome using Cuffmerge. Cuffdiff was used to estimate the abundance of all transcripts based on the finally merged transcriptome. We then used the following six filters to shortlist the bona fide lncRNAs from the obtained final transcriptome assembly: (1) transcripts without strand information were removed; (2) all single-exon transcripts that are within a 500-bp flanking region of known transcripts and in the same direction as the known transcripts were discarded; (3) transcripts overlapped with mRNAs annotated in the reference genome were deleted; (4) transcripts with FPKM scores  $< 0.5$  (2 for single-exon transcripts) and shorter than 200 bp were discarded; (5) the coding potential value of each transcript was calculated using CPC and those with CPC scores  $> 0$  were discarded; (6) the remaining transcripts were searched against the Pfam database by HMMER to remove transcripts containing known protein domain. The remaining transcripts were regarded as expressed candidate lncRNAs.

### Expression analysis and mRNA-circRNA-miRNA-lncRNA interaction network construction

Quantifications of the expression level of genes, circRNAs and miRNAs were performed using FPKM (fragments per kilobase of transcripts per million mapped reads), SRPBM (number of circular reads/number of mapped reads (units in billion)/read length), and RPM (reads count of miRNAs/clean reads  $\times 10^6$ ), respectively. Differential expression of

genes, circRNAs and miRNAs was profiled with the Rpackage edgeR. *P*-values were adjusted using the Benjamini and Hochberg's approach. Differentially expressed mRNAs, circRNAs, miRNAs and lncRNAs were defined by different standard in terms of the adjusted *P*-value and the absolute  $\log_2$  (fold-changes) as shown in Additional file 8: Table S4, Additional file 9: Table S5, Additional file 10: Table S6, Additional file 11: Table S7. To compare the expression level of the four types of RNA molecules in a unified manner, they were normalized as  $\text{FPKM}_{\text{mRNA}}$  (number of mRNA reads/number of mapped reads (units in billion)/mRNA transcript length),  $\text{FPKM}_{\text{circRNA}}$  (number of circRNA reads/number of mapped reads (units in billion)/circRNA junction length),  $\text{FPKM}_{\text{miRNA}}$  (number of miRNA reads/number of mapped reads (units in billion)/miRNA transcript length),  $\text{FPKM}_{\text{lncRNA}}$  (number of lncRNA reads/number of mapped reads (units in billion)/lncRNA transcript length). Pearson's correlation coefficient was used to measure the strength of the linear dependence of two variables. Pearson's correlation value was calculated for the mRNA-circRNA-miRNA-lncRNA co-expression network. The networks with a Pairs' absolute coefficient value over 0.8 and a *P*-value below 0.05 were selected. The network results were visualized using Cytoscape v 3.6.0 to show the co-expression relationships [47]. Targets of miRNAs were predicted via the online server psRNATarget with the default settings [48].

### Annotation of nicotine biosynthesis and catabolism related genes

The clean reads of RNA-Seq data were mapped to the same *N. tabacum* reference genome mentioned above using hisat2 v 2.1.0 [49]. The annotation of nicotine-related genes in *N. tabacum* was carried out using the method described previously by Shen et al. (2015) with minor modifications, i.e. an *E*-value  $< 1e-5$  and sequence identify  $> 50\%$  were used in BLASTp search of homologous genes related to nicotine biosynthesis and catabolism [50].

### Expression validation of mRNAs and ncRNAs in tobacco

qRT-PCR was used to validate the expression levels of selected mRNAs, miRNAs, circRNAs and lncRNAs. Total RNAs isolated from roots of topping-treated or untreated plants were used in the first-strand cDNA synthesis using the HiScript II 1st Strand cDNA synthesis kit (Vazyme) according to the manufacturer's instruction. The Oligo (dT) and random hexamers supplied in the kit were used in generation of the 1st strand cDNA for quantification of mRNAs and lncRNAs, respectively. The 1st strand cDNAs generated using random hexamers were also used in analysis of circRNAs. To validate circRNAs identified in this study, polymerase chain reactions (PCRs) were performed using a set of divergent primers and a set of convergent

primers that were used as a control. For miRNA quantification, a stem loop primer was used in reverse transcription. qPCR was performed with the ChamQ SYBR qPCR Master Mix kit (Vazyme) on the LightCycler 96 platform (Roche). The PCR procedure for circRNAs, lncRNAs and target genes was as following: 95 °C for 30 s, 40 cycles at 95 °C for 10 s, 60 °C for 30 s and then 72 °C for 30 s. The PCR procedure for miRNAs was as following: 95 °C for 5 min, 40 cycles at 95 °C for 10 s, 60 °C for 30 s and then 72 °C for 20 s. In both programs, the melting curve and amplification curve were examined to evaluate the specificity of amplification. The relative expression levels were analyzed by the  $2^{-\Delta\Delta Ct}$  method. *U6* was used as the internal control for miRNAs. *EF-1 $\alpha$*  was used as the internal control for circRNAs, lncRNAs and target genes. All the qRT-PCR reactions were assayed in triplicates and the primers used in our study were listed in Additional file (Additional file 6: Table S2, Additional file 7: Table S3).

## Conclusions

In this study, 688 miRNAs, 7423 lncRNAs and 12,414 circRNAs were identified in tobacco roots. Although not all four types of RNAs, several networks involving three types of RNAs (e.g. mRNA-miRNA-circRNA) related to nicotine biosynthesis were uncovered. Our results illustrated the transcriptomic profiles of tobacco roots, an organ responsible for nicotine biosynthesis, suggesting that mRNAs always play the most important roles, while ncRNAs are also expressed extensively for topping treatment response, especially circRNAs are the most activated in the ncRNA pool. ncRNAs appear to form interaction complex or network for the regulation of stress response such as nicotine biosynthesis. Taken together, the study enhanced our understanding of the RNA landscape in tobacco and will facilitate functional characterization of mRNAs and ncRNAs in biotic/abiotic responses in tobacco.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-019-6236-6>.

**Additional file 1: Figure S1.** Genome-wide identification of miRNAs in tobacco.

**Additional file 2: Figure S2.** Genome-wide identification of lncRNAs in tobacco.

**Additional file 3: Figure S3.** Features of circRNAs identified in tobacco and compared with other plants.

**Additional file 4: Figure S4.** Co-expression patterns of coding RNAs and three types of non-coding RNAs.

**Additional file 5: Table S1.** Number of coding and three types of non-coding RNAs identified in other plants.

**Additional file 6: Table S2.** Primers used in qRT-PCR for miRNA.

**Additional file 7: Table S3.** Primers used in qRT-PCR for mRNA, lncRNA and circRNA.

**Additional file 8: Table S4.** Summary of expression changes of circRNAs between topping treatment (TT) and control (CK) group.

**Additional file 9: Table S5.** Summary of expression changes of miRNAs between topping treatment (TT) and control (CK) group.

**Additional file 10: Table S6.** Summary of expression changes of lncRNAs between topping treatment (TT) and control (CK) group.

**Additional file 11: Table S7.** Summary of expression changes of mRNAs between topping treatment (TT) and control (CK) group.

## Abbreviations

eTM: endogenous target mimicry; ncRNA: non-coding RNA; PMT: Putrescine methyltransferase; QPT: quinolinic acid phosphoribosyl transferase; QS: encoding quinolinate synthase; TF: Transcription factor

## Acknowledgements

We thank Hongbo Zhang for providing suggestions on experiments.

## Authors' contributions

LF and PC designed the research. XC and LL conducted experiments; SS, FL, ES and CY analyzed data; XC and SS wrote the paper, QHZ, MPT and LF edited the manuscript. All the authors read and approved the final version of the paper.

## Funding

This work was supported by the Fundamental Research Funds for the Central Universities of China (2017QNA6013) to XC and Jiangsu Collaborative Innovation Center for Modern Crop Production and 111 project of China (B17039) to LF. The authors are solely responsible for the experimental design, data interpretation, and conclusion drawn herein and all results and findings are in the public domain and freely distributed.

## Availability of data and materials

All the sequencing data generated in this study was submitted to NCBI with accession number PRJNA526387.

## Ethics approval and consent to participate

Tobacco seeds were provided by Dr. Bingguang Xiao from the Yunnan Academy of Tobacco Agricultural Sciences. The tobacco plants were grown in growth chambers at Zhejiang University. During the execution of these experiments, all plant materials collected were destroyed at the conclusion of the work. No field experiments were included in this study. All collection of plant material was done under compliance with any applicable institutional, national, or international guidelines, and as such no specific permissions and/or licenses were required in order to comply with the Convention on the Trade in Endangered Species of Wild Fauna and Flora since these plants fall outside of this jurisdiction.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Institute of Crop Science, Zhejiang University, Hangzhou 310058, China.

<sup>2</sup>Research Center for Air Pollution and Health, Zhejiang University, Hangzhou 310058, China. <sup>3</sup>Key Laboratory of Tobacco Biotechnological Breeding, Yunnan Academy of Tobacco Agricultural Sciences, Kunming 650021, China.

<sup>4</sup>Department of Biology, University of Virginia, Charlottesville, VA 22904, USA.

<sup>5</sup>CSIRO Agriculture and Food, GPO Box 1700, Canberra, ACT 2601, Australia.

<sup>6</sup>Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China.

Received: 11 March 2019 Accepted: 28 October 2019

Published online: 14 November 2019

## References

- Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009;136(2):215–33.

2. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell*. 2009;136(4):629–41.
3. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22(9):1775–89.
4. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013; 495(7441):384–8.
5. Kornienko AE, Guenzi PM, Barlow DP, Pauler FM. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol*. 2013;11:14.
6. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature*. 2014;505(7483):344–52.
7. Liang WC, Fu WM, Wong CW, Wang Y, Wang WM, Hu GX, et al. The lncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer. *Oncotarget*. 2015;6(26):22513–25.
8. Todesco M, Rubio-Somoza I, Paz-Ares J, Weigel D. A collection of target mimics for comprehensive analysis of MicroRNA function in *Arabidopsis thaliana*. *PLoS Genet*. 2010;6(7):10.
9. Ivashuta S, Banks IR, Wiggins BE, Zhang YJ, Ziegler TE, Roberts JK, et al. Regulation of gene expression in plants through miRNA inactivation. *PLoS One*. 2011;6(6):11.
10. Wu HJ, Wang ZM, Wang M, Wang XJ. Widespread long noncoding RNAs as endogenous target mimics for MicroRNAs in plants. *Plant Physiol*. 2013; 161(4):1875–84.
11. Wang YX, Yang M, Wei SM, Qin FJ, Zhao HJ, Suo B. Identification of Circular RNAs and Their Targets in Leaves of *Triticum aestivum* L. under Dehydration Stress. *Frontiers in Plant Science*. 2017;7:10.
12. Shoji T, Hashimoto T. Recruitment of a duplicated primary metabolism gene into the nicotine biosynthesis regulon in tobacco. *Plant J*. 2011;67(6):949–59.
13. Zhang HB, Bokowiec MT, Rushton PJ, Han SC, Timko MP. Tobacco transcription factors NtMYC2a and NtMYC2b form nuclear complexes with the NtJAZ1 repressor and regulate multiple Jasmonate-inducible steps in nicotine biosynthesis. *Mol Plant*. 2012;5(1):73–84.
14. Dewey RE, Xie JH. Molecular genetics of alkaloid biosynthesis in *Nicotiana tabacum*. *Phytochemistry*. 2013;94:10–27.
15. Shoji T, Hashimoto T. Stress-induced expression of NICOTINE2-locus genes and their homologs encoding ethylene response factor transcription factors in tobacco. *Phytochemistry*. 2015;113:41–9.
16. Sears MT, Zhang HB, Rushton PJ, Wu M, Han SC, Spano AJ, et al. NtERF32: a non-NIC2 locus AP2/ERF transcription factor required in jasmonate-inducible nicotine biosynthesis in tobacco. *Plant Mol Biol*. 2014;84(1–2):49–66.
17. Guo HX, Kan YC, Liu WQ. Differential expression of miRNAs in response to topping in flue-cured tobacco (*Nicotiana tabacum*) roots. *PLoS One*. 2011;6(12):15.
18. Tang S, Wang Y, Li ZF, Gui YJ, Xiao BG, Xie JH, et al. Identification of wounding and topping responsive small RNAs in tobacco (*Nicotiana tabacum*). *BMC Plant Biol*. 2012;12:15.
19. Gao J, Yin F, Liu M, Luo M, Qin C, Yang A, et al. Identification and characterisation of tobacco microRNA transcriptome using high-throughput sequencing. *Plant Biol*. 2015;17(3):591–8.
20. Li XH, Xing XX, Xu SX, Zhang MZ, Wang Y, Wu HY, et al. Genome-wide identification and functional prediction of tobacco lncRNAs responsive to root-knot nematode stress. *PLoS One*. 2018;13(11):15.
21. Fu YP, Guo HX, Cheng ZP, Wang R, Li GL, Huo G, et al. NtNAC-R1, a novel NAC transcription factor gene in tobacco roots, responds to mechanical damage of shoot meristem. *Plant Physiol Biochem*. 2013;69:74–81.
22. Li F, Wang W, Zhao N, Xiao B, Cao P, Wu X, et al. Regulation of nicotine biosynthesis by an endogenous target mimicry of MicroRNA in tobacco. *Plant Physiol*. 2015;169(2):1062–71.
23. Mallory AC, Bartel DP, Bartel B. MicroRNA-directed regulation of *Arabidopsis* AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell*. 2005;17(5):1360–75.
24. Wu G, Poethig RS. Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development*. 2006; 133(18):3539–47.
25. Shen E, Zou J, Hubertus Behrens F, Chen L, Ye C, Dai S, et al. Identification, evolution, and expression partitioning of miRNAs in allopolyploid *Brassica napus*. *J Exp Bot*. 2015;66(22):7241–53.
26. Shen E, Zhu X, Hua S, Chen H, Ye C, Zhou L, et al. Genome-wide identification of oil biosynthesis-related long non-coding RNAs in allopolyploid *Brassica napus*. *BMC Genomics*. 2018;19(1):745.
27. Chu QJ, Zhang XC, Zhu XT, Liu C, Mao LF, Ye CY, et al. PlantcircBase: A Database for plant circular RNAs. *Mol Plant*. 2017;10(8):1126–8.
28. Ye CY, Chen L, Liu C, Zhu QH, Fan LJ. Widespread noncoding circular RNAs in plants. *New Phytol*. 2015;208(1):88–95.
29. Chu Q, Bai P, Zhu X, Zhang X, Mao L, Zhu Q-H, et al. Characteristics of plant circular RNAs. *Brief Bioinform*. 2018.
30. Zhu QH, Fan LJ, Liu Y, Xu H, Llewellyn D, Wilson I. miR482 Regulation of NBS-LRR Defense Genes during Fungal Pathogen Infection in Cotton. *Plos One*. 2013;8(12):11.
31. Xu SQ, Brockmoller T, Navarro-Quezada A, Kuhl H, Gase K, Ling ZH, et al. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc Natl Acad Sci U S A*. 2017;114(23):6133–8.
32. Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell*. 2014;56(1):55–66.
33. Xie JH, Fan LJ. Nicotine biosynthesis is regulated by two more layers: small and long non-protein-coding RNAs. *Plant Signal Behav*. 2016;11(6):3.
34. Sierro N, Battey JND, Ouadi S, Bakaher N, Bovet L, Willig A, et al. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun*. 2014;5:9.
35. Di C, Yuan JP, Wu Y, Li JR, Lin HX, Hu L, et al. Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J*. 2014;80(5):848–61.
36. Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, et al. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol*. 2014;15(12):16.
37. Jones-Rhoades MW, Bartel DP. Computational identification of plant MicroRNAs and their targets, including a stress-induced miRNA. *Mol Cell*. 2004;14(6):787–99.
38. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
39. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9(4):357–U54.
40. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105–11.
41. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
42. Gao Y, Wang JF, Zhao FQ. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol*. 2015;16:16.
43. Memczak S, Jens M, Eleftheriotti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013;495(7441):333–8.
44. Chen L, Yu YY, Zhang XC, Liu C, Ye CY, Fan LJ. PcircRNA\_finder: a software for circRNA prediction in plants. *Bioinformatics*. 2016;32(22):3528–9.
45. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, et al. Criteria for annotation of plant MicroRNAs. *Plant Cell*. 2008;20(12):3186–90.
46. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat Protoc*. 2012;7(3):562–78.
47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
48. Dai XB, Zhao PX. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res*. 2011;39:W155–W9.
49. Kim D, Landmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–U121.
50. Xu SQ, Brockmoller T, Navarro-Quezada A, Kuhl H, Gase K, Ling ZH, et al. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc Natl Acad Sci U S A*. 2017;114(23):6133–8.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.