



Genome-wide analysis of long non-coding RNAs responsive to multiple nutrient stresses in *Arabidopsis thaliana*

Jingjing Wang^{1,2,3} · Qi Chen⁴ · Wenyi Wu⁴ · Yujie Chen^{4,5} · Yincong Zhou⁴ · Guoji Guo^{1,2,3} · Ming Chen^{4,6,1}

Received: 31 May 2020 / Revised: 19 October 2020 / Accepted: 25 October 2020 / Published online: 31 October 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Nutrient stress is the most important environmental stress that limits plant growth and development. Although recent evidence highlights the vital functions of long non-coding RNAs (lncRNA) in response to single nutrient stress in some model plants, a comprehensive investigation of the effect of lncRNAs in response to nutrient stress has not been performed in *Arabidopsis thaliana*. Here, we presented the identification and characterization of lncRNAs under seven nutrient stress conditions. The expression pattern analysis revealed that aberrant expression of lncRNAs is a stress-specific manner under nutrient stress conditions and that lncRNAs are more sensitive to nutrient stress than protein-coding genes (PCGs). Moreover, competing endogenous RNA (ceRNA) network and lncRNA-mRNA co-expression network (CEN) were constructed to explore the potential function of these lncRNAs under nutrient stress conditions. We further combined different expressed lncRNAs with ceRNA network and CEN to select key lncRNAs in response to nutrient stress. Together, our study provides important information for further insights into the role of lncRNAs in response to stress in plants.

Keywords Long non-coding RNAs · Nutrient stress · Competing endogenous RNA · Co-expression network

Introduction

Generally, long non-coding RNAs (lncRNAs) are defined as transcripts that are longer than 200 nucleotides (nt) and are not translated into proteins. According to their genomic location, lncRNAs are further classified into three types: antisense lncRNA (antisense to coding genes), lincRNA (originated from intergenic regions away from coding genes), and

incRNA (located at introns of coding genes) (Derrien et al. 2012). Over recent years, numerous studies have shown that plant lncRNAs are developmentally regulated and responsive to abiotic stress in various biological processes (Chekanova 2015; Liu et al. 2015; Shafiq et al. 2016). They have emerged as key regulators at both transcriptional and posttranscriptional level by acting as competing endogenous RNA (ceRNAs), small RNA (smRNAs), precursors, and scaffolds (Chekanova

Jingjing Wang and Qi Chen contributed equally to this work.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10142-020-00758-5>.

✉ Guoji Guo
ggj@zju.edu.cn

✉ Ming Chen
mchen@zju.edu.cn

¹ Center for Stem Cell and Regenerative Medicine, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, People's Republic of China

² Zhejiang Provincial Key Lab for Tissue Engineering and Reenerative Medicine, Dr. Li Dak Sum & Yip Yio Chin Center for Stem Cell and Regenerative Medicine, Hangzhou, People's Republic of China

³ Zhejiang Laboratory for Systems & Precision Medicine, Zhejiang University Medical Center, Hangzhou, People's Republic of China

⁴ Department of Bioinformatics; the State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou, People's Republic of China

⁵ College of Life Science, Inner Mongolia University for the Nationalities, Tongliao, People's Republic of China

⁶ James D. Watson Institute of Genome Sciences, Zhejiang University, Hangzhou, People's Republic of China

2015; Liu et al. 2015; Wang et al. 2017). For example, Shumayla et al. (2017) have predicted 19 lncRNAs as precursors and 1047 as target mimics of known miRNAs in wheats (Shumayla et al. 2017).

Nutrient stress is one of the most severe abiotic stresses that seriously influence plant growth and development (Zeng et al. 2014). Vital macronutrients such as nitrogen (N), phosphorus (Pi), and sodium (Na) are essential for optimal plant growth and are required in relatively large amounts. Micronutrients usually serve as cofactors of metabolic enzymes and protein complexes in the electron transport chain, such as magnesium (Mg), iron (Fe), and cuprum (Cu). The acquisition, transport, and homeostasis of nutrients are fundamentally important for plant physiology, whereas nutrient stress largely affects the production and quality of crop plants. Recently, an increasing number of lncRNAs were reported to be responsive to nutrient stress in some model organisms. For example, 637 lncRNAs were identified under N deficiency in maize (Lv et al. 2016), and 98 antisense lncRNAs were identified in drought response in rice (Chung et al. 2016). In *Populus*, 14 N-responsive lncRNAs were identified as miRNA precursors (Chen et al. 2016), and 202 lincRNAs were differentially expressed under Pi starvation (Cruz de Carvalho et al. 2016). Additionally, Xu et al. (2016) identified 3170 altered expressed lincRNAs under Pi starvation (Xu et al. 2016). Moreover, 7361 and 7874 salt-responsive lncRNAs were identified in leaf and root tissues of *Medicago truncatula*, respectively. In *A. thaliana*, a recent study identified 254 ploy (A)+ and 58 ploy (A)-lncRNAs, which were aberrantly expressed under four stress conditions, including heat, cold, drought, and salt (Di et al. 2014). In particular, one lncRNA, called IPS1 (induced by phosphate starvation 1), was found to be up-regulated under Pi deficiency in *A. thaliana*, which acted as a target mimic for miR399 and blocked the interaction between miR399 and another miR399 target, phosphate2 (PHO2) (Franco-Zorrilla et al. 2007).

Currently, studies on stress-responsive lncRNAs were mostly carried out on single stress type, whereas there is no systematic screening of lncRNAs under multiple nutrient stresses and their convergent regulatory roles in stress-related pathway. To investigate the potential role of lncRNAs in response to multiple nutrient stresses, we performed genome-wide lncRNA identification under seven nutrient stress conditions in *A. thaliana*. Eight hundred seventy-four lncRNA transcripts with high-confidence were identified through four filtration steps, of which 224 were lincRNAs, 53 were incRNAs (lncRNA form intron), and 597 were antisense lncRNAs. The potential function of these lncRNAs under seven nutrient stress conditions was investigated based on ceRNA network and lncRNA-mRNA co-expression network (CEN). Finally, we retained 357 key stress-responsive lncRNA and constructed relevance network among the seven nutrient stresses. Two stress-responsive lncRNAs functionally

shared by seven nutrient stresses were extracted for deep analysis. In summary, our results demonstrate that lncRNAs play a significant role in response to multiple nutrient stresses in *A. thaliana* and that integrative analysis of ceRNA network and CEN provides new insights for further investigation on the molecular mechanisms of stress-responsive lncRNAs and related protein-coding genes (PCGs) in plants.

Results

Genome-wide identification and properties of lncRNAs in *A. thaliana*

We developed a pipeline to identify lncRNAs in *A. thaliana* under seven nutrient stress conditions (Table 1), which was initiated by RNA-seq data processing and followed by multiple filter stages (Fig. 1a). The details of the pipeline are showed in the section of “Materials and methods.” A total of 874 isoforms (721 loci) were obtained from the RNA-seq data (Supplementary Table S1). To verify the accuracy, we used COME software by Hu et al. (2017) to test our identified lncRNAs (Hu et al. 2017). The results show that 81.6% (713 out of 874) was marked as “non-coding,” and 97.5% (852 out of 874) was detected in poly A+ RNA libraries. In addition, 193 identified lncRNAs have been reported in CANTATA database (Szczesniak et al. 2019), and 41 have been annotated in GREENC database (Paytuy Gallart et al. 2016) (Supplementary Table S2). The comparison result between lncRNAs and long miRNAs suggested that long miRNA sequences have no effect on lncRNAs analysis. Although ath-MIR5647 gene completely overlaps with TCONS_00026952, ath-MIR5647 gene is a lot shorter than TCONS_00026952.

According to their proximity to coding genes in the genome, the predicted lncRNAs were divided into three types (Fig. 1b). Most lncRNAs (68.3%) were antisense to coding genes (antisense lncRNA). 25.6% originated from intergenic regions away from coding genes (lincRNA), and 6.1% were located at introns of coding genes (incRNA). We then compared the sequence length, exon number, expression level, and GC content between lncRNAs and PCGs in *A. thaliana* (Fig. 1c–f). Not surprisingly, lncRNAs were shorter than PCGs with fewer exons and lower expression levels (P value = $1.098e-14$, Student's t test), consistent with the previous study in rice and cassava (Li et al. 2017; Zhang et al. 2014). It is also worth mentioning that GC content volatility of lncRNAs was significantly higher than that of PCGs (Fig. 1f). GC content of lncRNAs was between 0.3 and 0.5, whereas GC content of PCG was focused largely on 0.4.

The expression pattern of stress-responsive genes was reported to be closely correlated with histone methylation (Haak et al. 2017). For instance, H3K4me3 is usually considered a stimulative epigenetic mark, while H3K27me3 is regarded as

Table 1 Summary of RNA-seq data used in the study

Project	Nutrient	Stress condition	Accession	Ref.
1	Sodium (Na)	Control High salt	ERR274309 ERR274308	Sani et al. (2013)
2	Cuprum (Cu)	Control Low Cu	SRR069565; SRR069566; SRR069567 SRR069568; SRR069569	Bernal et al. (2012)
3	Magnesium (Mg)	Control High Mg Low Mg	SRR1734689; SRR1734690 SRR1734693; SRR1734694 SRR1734697; SRR1734698	Niu et al. (2016)
4	Ferrum (Fe) & phosphate (Pi)	Control Low Fe Low Pi	SRR331227 SRR331228 SRR331229	Li et al. (2013)
5	Phosphate (Pi)	Control Low Pi	SRR420813; SRR420814; SRR420815 SRR420816; SRR420817; SRR420818	Lan et al. (2012)
6	Nitrogen (N)	Control High N	SRR671946; SRR671947 SRR671948; SRR671949	Vidal et al. (2013)

a repressive mark. Our histone methylation results showed that H3K4me3 mark of protein-coding genes was high around the transcription start site (TSS) but low around the transcription termination site (TTS) (Fig. 1g), suggesting the increased expression of stress-responsive genes was up-regulated by H3K4me3. We also observed different epigenetic modification signal distributions in our identified lncRNAs, compared with PCGs in *A. thaliana* (P value < 0.005, Wilcoxon rank sum test). The histone methylation of lncRNAs presented small ups and downs frequently, while there were no big overall changes. Comparatively, lncRNAs tend to have weaker H3K4me3 signals than PCGs around the transcription start site (TSS) but stronger signal around the transcription termination site (TTS), while the H3K27me3 signal of the lncRNAs was weaker than that of PCGs along the whole region (Fig. 1g and h). Altogether, the different signal distribution of H3K4me and H3K27me3 between lncRNAs and PCGs may give a hint to the potential effects of chromatin modification on lncRNA expression.

Expression pattern of lncRNAs responding to nutrient stress

Based on the lncRNAs identified in the *A. thaliana* genome mentioned above, we systematically analyzed their expression patterns under seven nutrient stress types, including salt_H, Cu_L, Mg_H, Mg_L, Fe_L, Pi_L, and N_H. We found that the expression levels and fold change (FC) values of lncRNAs changed with various stress conditions (Fig. 2a and Supplementary Fig. S1). Comparatively, Mg_H and Mg_L share similar expression pattern. More lncRNAs were up-regulated under salt_H and Cu_L stresses, while most lncRNAs were down-regulated under other five stresses (Fig. 2a). Subsequently, we defined lncRNAs that were differentially expressed under two stress conditions to have two

linked stress conditions. By that analogy, we observed that most lncRNAs (78.46%) had less than three linked stress conditions (Fig. 2b), suggesting that aberrant expression of lncRNAs was a stress-specific manner. To confirm these results, we calculated the stress specificity scores using Tau (τ) (Yanai et al. 2005), a widely used method for the tissue specificity scores. The stress specificity score τ values range from 0 for stress-independent genes to 1 for stress-specific genes. Our result showed that more than 80% associated lncRNAs had higher stress specificity scores ($\tau > 0.7$, P value < 2.2e-16, Wilcoxon rank sum test) (Fig. 2c). We then compared the altered expression pattern between lncRNAs and PCGs under seven stress conditions, respectively. The percentage of differential expressed lncRNAs and PCGs was calculated and shown in Fig. 2d. Comparing with PCGs, more lncRNAs tended to be differentially expressed under seven stress conditions (P value < 2.2e-16, Pearson chi-square test). All these observations demonstrate that lncRNAs have higher stress specificity than PCGs in response to various nutrient stresses.

Next, we analyzed the expression pattern between antisense lncRNAs and their sense genes. Compared with randomly selected gene pairs (P value < 0.05, Wilcoxon rank sum test), antisense lncRNAs could either present a positive (co-expression) relationship with their sense genes or show a negative (anti co-expression) relationship (Pearson correlation coefficient, $|r| > 0.5$, Supplementary Table S3). Studies dedicated to antisense lncRNAs have demonstrated their versatile mechanisms to elicit regulatory effects as positive and negative modulators of protein-coding genes (Villegas and Zaphiropoulos 2015). Accordingly, antisense lncRNAs in response to nutrient stress are likely to regulate their counterpart sense genes by bridging epigenetic effectors and regulatory complexes at specific genomic loci. Besides, antisense lncRNAs may be processed into nat-siRNAs in nutrient stress response based on our previous study that cis-nature antisense

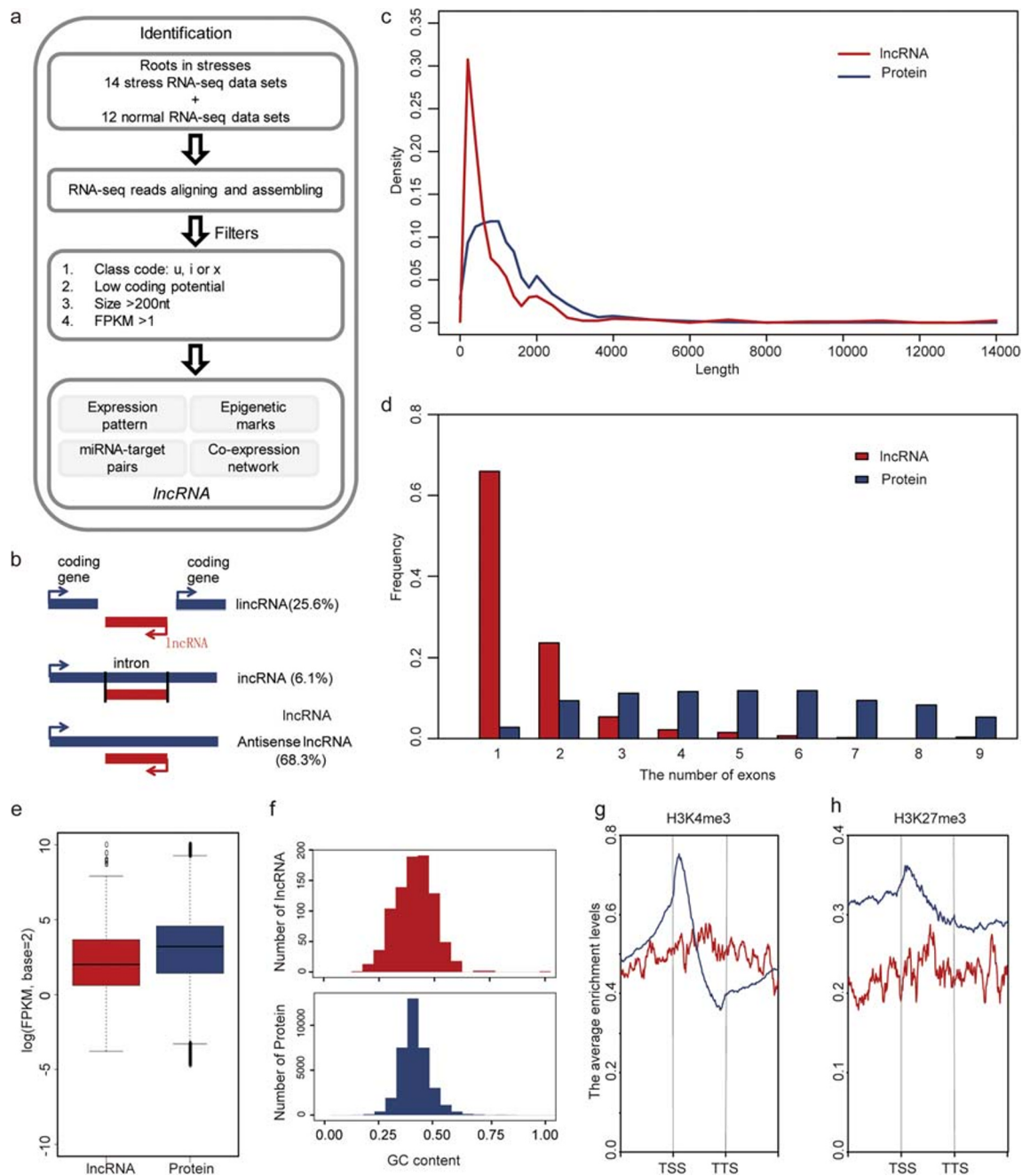


Fig. 1 Identification and characteristics of lncRNAs in *A. thaliana*. **(a)** The framework to identify and characterize lncRNAs. **(b)** Classification of lncRNAs according to their proximity to coding genes in the genome. **(c)** Transcript length comparison between protein-coding genes (PCGs) and lncRNAs. **(d)** Proportions of exons in PCGs and lncRNAs. **(e)** The expression level of the PCGs and lncRNAs. **(f)** GC content of PCGs and

lncRNAs. **(g–h)** The epigenetic markers level of PCGs and lncRNAs. The average enrichment levels of H3K4me3 **(g)** and H3K27me3 **(h)** in PCGs and lncRNAs, respectively. TSS means transcription start site and TTS means transcription termination site. In all cases, the red represents the lncRNA, and blue refers to the PCG

transcripts (cis-NATs) can produce nat-siRNAs (Yuan et al. 2015).

ceRNA network of nutrient-responsive lncRNAs

Previous studies have shown that lncRNAs could serve as competing endogenous RNAs (ceRNAs) to compete for

binding microRNAs (miRNAs). Those miRNA sponges can inhibit the activity of targeted miRNAs and relieve the repression to the originally paired mRNAs. The ceRNA phenomenon of lncRNAs has been widely found in plants (Yuan et al. 2016a). A recent study of rice lncRNAs has found a couple of lincRNAs competing with genes and playing important roles under Pi starvation (Xu et al. 2016). Based on the lncRNAs

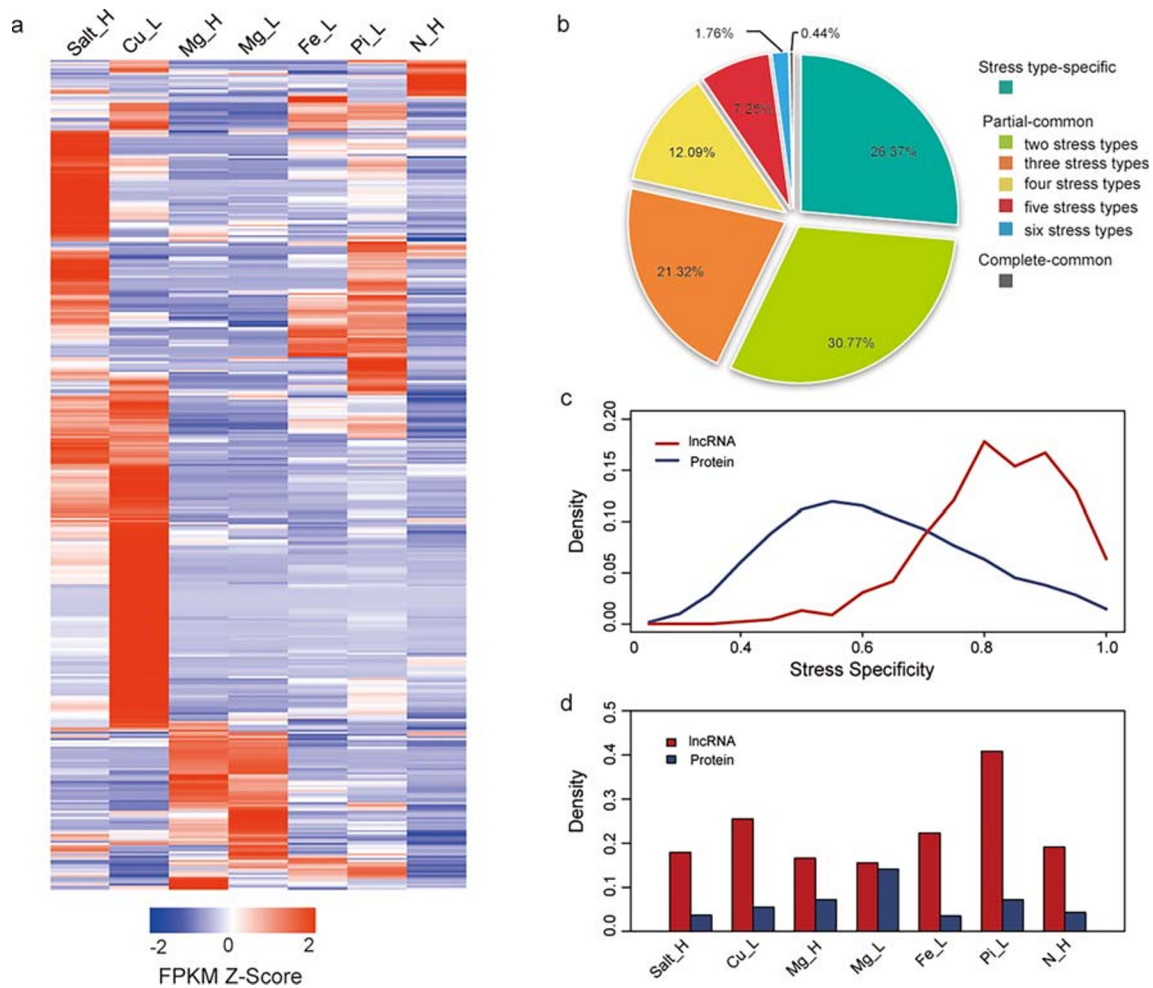


Fig. 2 Expression pattern of lncRNAs in response to different type of nutrient stress. **(a)** Heat map suggesting the differentially expressed lncRNAs expression pattern under seven nutrient stress conditions. Colors represent high (red), low (blue), or average (white) expression values based on Z-score normalized FPKM values. **(b)** Overlap of the

number of differentially expressed lncRNAs in different types of nutrient stress. **(c)** Stress-specific expression scores of PCGs and lncRNAs are calculated, and their densities are plotted. **(d)** The proportion of differentially expressed PCGs and lncRNAs in different nutrient stress conditions

under seven nutrient stress conditions in *A. thaliana*, we systematically assayed nutrient stress-associated competing endogenous RNA (ceRNA) network (Supplementary Fig. S2). We integrated multiple data sources to identify a total of 4157 miRNA-mediated ceRNA crosstalk and then constructed ceRNA network (see “Materials and methods”), which contained 4803 nodes (386 miRNAs, 59 lncRNAs, and 4353 mRNAs) and 6046 edges (miRNA-target pairs). The degree distribution of these nodes in ceRNA network showed a power law distribution with a slope of -1.292 and $R^2 = 0.697$. That is to say, most nodes had relatively few interacting partners, while only a small subset of nodes had a large number of interacting partners, suggesting our ceRNA network is scale-free. These findings are consistent with recent ceRNA studies in rice (Xu et al. 2016) and human (Ning et al. 2016; Xu et al. 2015; Zhou et al. 2016).

In order to explore the potential functional implication of lncRNAs in response to nutrient stress, we selected the sub-

ceRNA network that contains lncRNAs, related miRNAs, and related mRNAs (Fig. 3a, Supplementary Table S4). Studies dedicated to plant transcriptomes have revealed a large number of miRNAs and affirmed their critical involvement in numerous abiotic stresses during their growth cycle (Stauffer and Maizel 2014). Therefore, we first exacted 41 miRNAs in our sub-ceRNA network and compared them with differentially expressed miRNAs reported under different nutrient stress conditions in *A. thaliana* (Supplementary Table S5). We found that 82.93% of the selected miRNAs were responsive to high-salt conditions (Meng et al. 2015) and 21.95% were differentially expressed in response to carbon (C), sulfur (S), and N starvation (Liang et al. 2015). In addition, the involvement of ath-miR824-5p, ath-miR780.1, and ath-miR160c-3p has been observed to be up-regulated in Pi-deficient condition (Kumar et al. 2017), and 5 miRNAs (ath-miR854a-e) were induced by arsenic (As) stress (Srivastava et al. 2013). These stress-responsive miRNAs in our ceRNA network has well

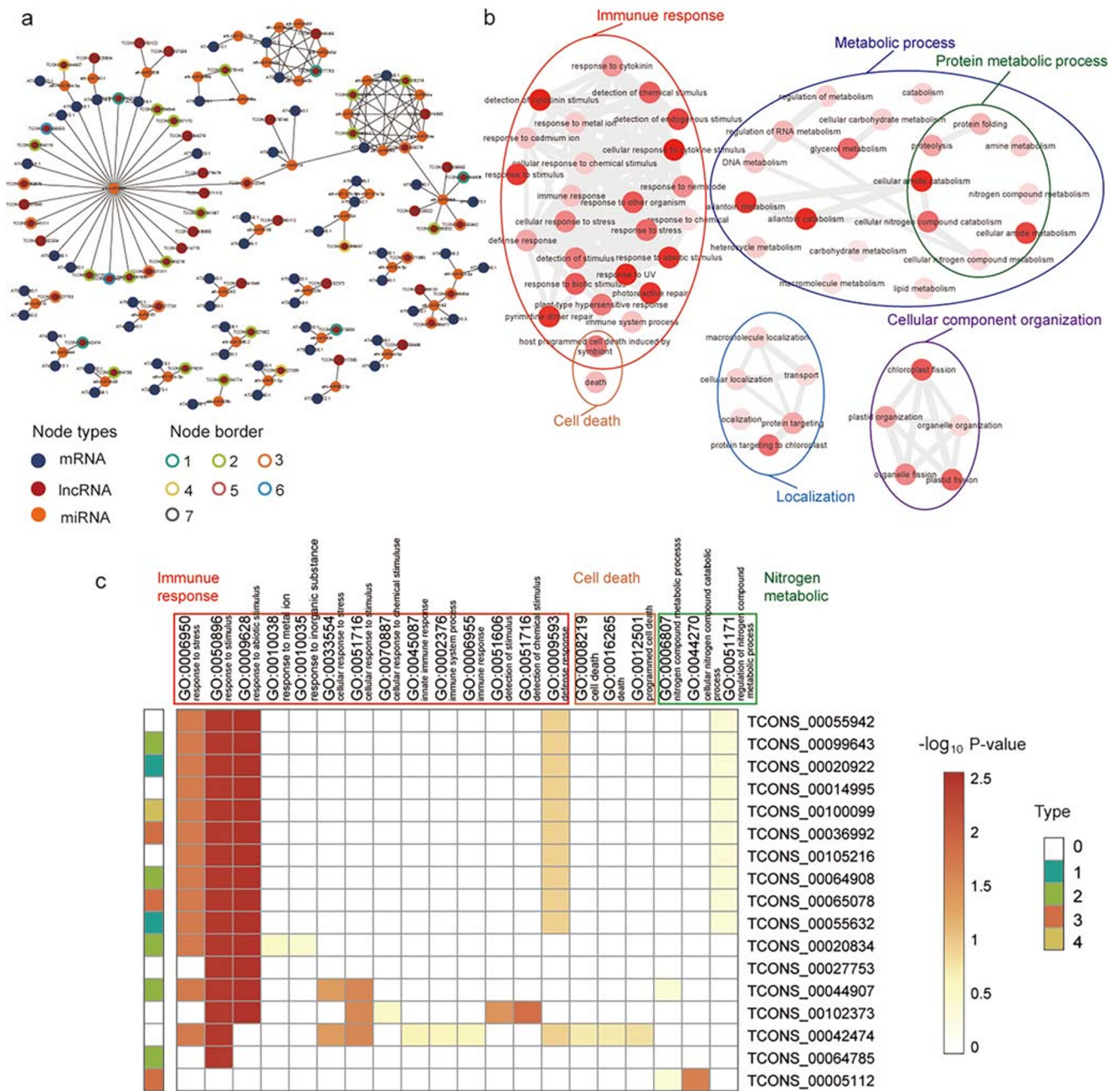


Fig. 3 The ceRNA network under nutrient stress conditions. **(a)** lncRNA-associated ceRNA network under nutrient stress conditions. mRNAs (blue), lncRNAs (red), and miRNAs (orange) are represented as circles. The color of circle border refers to the number of linked stress conditions. **(b)** Enrichment of nutrient stress-related GO terms of the coding genes in this ceRNA network. Among the enriched GO categories, immune

response, cell death, and nitrogen metabolic were related to stress response. **(c)** The function of lncRNAs that are involved in ceRNA network. Rows correspond to lncRNAs, and the columns represent GO terms. The type refers to the number of linked stress conditions. The heatmap color intensities indicate the level of enrichment score of each GO term

demonstrated the role of lncRNA regulation under multiple nutrient stresses. Accordingly, we speculate that some stress-responsive lncRNAs might participate in conservative stress-related signal pathways, thus being sensitive to various nutrient stresses.

We then conducted the functional analysis of the related mRNA genes in sub-ceRNA network. These genes were

found to be involved in immune response, cell death, localization, metabolic process, and cellular component organization (Fig. 3b). Based on the function of lncRNAs' nearest mRNAs and the second nearest mRNAs in the sub-ceRNA network, 17 lncRNAs were successfully annotated with stress-responsive related GO terms (Fig. 3c). Among these lncRNAs, 16 lncRNAs were annotated as “immune

response”-related GO terms, such as response to abiotic stimulus (GO:0009628), response to metal ion (GO:00100038), and detection of chemical stimulus (GO:0051606). Notably, TCONS_00042474 was enriched in cell death-related GO terms, and TCONS_00005112 was significant in nitrogen compound catabolic process. Altogether, our results suggest that lncRNAs could function as important regulatory nodes of different nutrient stress pathways to block the interactions between miRNAs and their mRNA targets.

lncRNA-mRNA co-expression network

To further identify functionally important lncRNAs, we constructed co-expression network (see “Materials and methods”). In the co-expression network, there are 68,622 connections between 468 lncRNAs and 10,668 coding genes. Similar to previous studies (Di et al. 2014), a small subset (20.8%) of lncRNA-mRNA co-expression pairs were within the same chromosome, and only 20 lncRNAs were located within 10 kb of their mRNA neighbors. Most lncRNAs were involved in *trans*-regulation, while only a small part of lncRNAs were involved in *cis*-regulation with their co-expression associated genes. Similar to the ceRNA network, network topology exhibited a hallmark of scale freeness (slope value = -1.059 and $R^2 = 0.731$). We defined top 1.2% (top 130) of nodes by degree as hubs and found that all these hubs are lncRNAs that co-expressed with almost 80.1% of nodes in the CEN, suggesting the centrality and essentiality of lncRNAs in CEN. The MCODE algorithm with default suggested parameters was applied to mine the subnetwork of CEN. In total, 29 modules were screened from the CEN, whereas there were two modules only containing lncRNAs. According to the previous studies, some lncRNAs might act synergistically to regulate genes via chromosome organization (Fu 2014). The co-expression pattern of lncRNAs in these two modules may be associated with the significant co-regulation of genes by a group of synergistic lncRNAs.

Next, we performed GO analysis for the coding genes in the whole CEN by BiNGO (Maere et al. 2005). There were three types of GO terms that are related to stress response: immune response, localization and transport, and death (Fig. 4a). Among the 29 modules of CEN, 18 modules were enriched in stress response GO terms (Fig. 4b). Based on the module function, 77 lncRNAs were successfully annotated with immune response, 61 were annotated as “localization and transport”-related GO terms, and 5 were identified as “death”-related GO terms (Fig. 4c). We found four stress-related GO terms enriched for lncRNAs in both ceRNA network and co-expression network. These four biological processes (detection of stimulus, respond to stimulus, respond to stress, and respond to metal ion) were closely related and intersect with each other, indicating versatile roles of lncRNAs in response to various nutrient stresses.

Key lncRNAs play critical roles in responding to nutrient stresses

We finally selected key lncRNAs as those were differentially expressed under at least one nutrient stress conditions and also involved in the ceRNA network or CEN (Supplementary Table S6). In total, there were 34 and 348 lncRNAs identified as key lncRNAs in ceRNA network and CEN, respectively. Among 34 key lncRNAs in the ceRNA network, 12 lncRNAs were annotated as being involved in stress response, suggesting that they may participate in the nutrient stress response by serving as ceRNAs to block the activity of miRNAs. On the other hand, 64 out of 348 key lncRNAs in the CEN were annotated to nutrient stress-responsive associated GO terms. Notably, we found that differentially expressed lncRNAs are more likely to be co-expressed with coding genes (P value $< 1.432e-12$, Fisher test), suggesting their significant roles in the gene regulatory network.

Currently, the characterization and the functional analysis of stress-responsive lncRNAs are very limited in plants, which are less well studied than that of miRNAs and mRNAs. To explore the stress-responsive function of our key lncRNAs, we compared them with lncRNAs identified by Yuan et al. (2016) and found 21 key lncRNAs involved in Pi starvation signaling pathways. Eight key lncRNAs work as targets of PHR1, a key transcription factor for transcriptional response to Pi starvation (Yuan et al. 2016b). Among 142 identified *A. thaliana* lncRNAs, 46 were found in the study of Liu et al. (2012) with 5 key lncRNAs regarded as the regulators involved in cold, drought, and high-salt response (Liu et al. 2012), while for antisense lncRNAs, 14% (89 out of 597) were involved in stress response as well as their sense genes, including 23 key antisense lncRNAs.

Based on the key lncRNAs in the seven stress conditions, we further constructed the nutrient stress-key lncRNAs network (Fig. 5a) and delineated the distribution features of key lncRNAs in different nutrient stress condition (Fig. 5b). There were the most stress-specific key lncRNAs in Pi deficiency condition, whereas only four stress-specific key lncRNAs were related to Mg abnormality. The majority of key lncRNAs were related with two or three stress conditions (Fig. 5 a and b), and those lncRNAs functionally shared by different types of nutrient stresses are likely to serve as “convergent points” of signal transduction, transcriptional regulation in nutrient stress-related pathways. Based on the result of key lncRNA network, we then calculated the “convergence degree” between each pair of the seven nutrient stresses (Fig. 5c). The Mg_L and Mg_H stresses showed the highest convergence with each other, whereas salt_H showed the lowest convergence compared with any of the other stresses (Fig. 5c). Among five lncRNAs linked to six stress conditions in key lncRNA network, four of them (XLOC_017401, XLOC_018003,

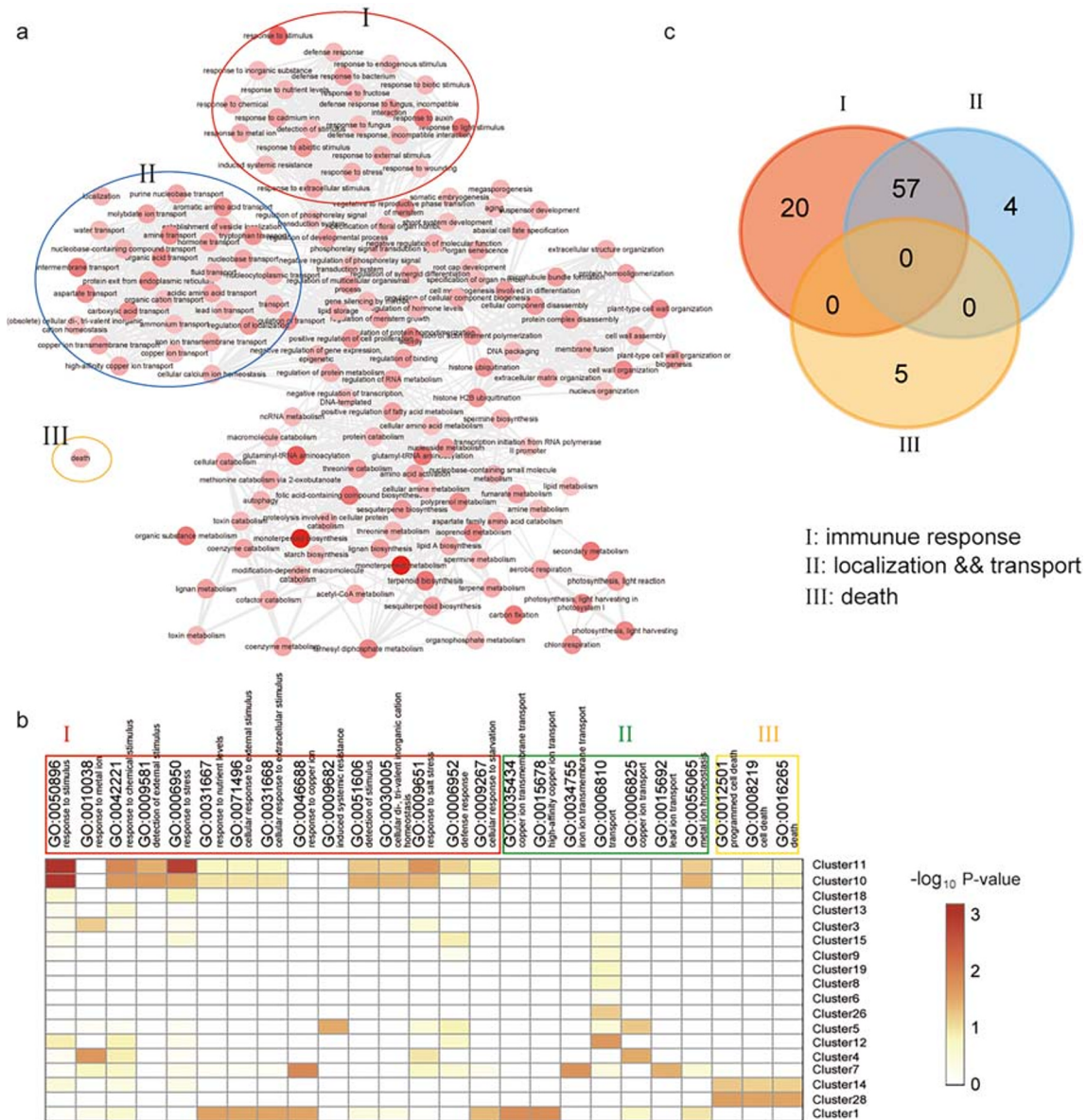


Fig. 4 The functional analysis of co-expression network under nutrient stress conditions. **(a)** Functional enrichment of the coding genes in the associated co-expression network. There are three types of GO categories that were related to stress response. **(b)** Functional GO term annotation of

co-expression modules. Rows correspond to one module, and the columns represent an associated GO term. The heatmap color intensities indicate the level of enrichment score of each GO term. **(c)** Venn diagrams showed the overlap of lncRNAs among the three GO categories

XLOC_018264, XLOC_030820) were involved in multiple nutrient stresses except for the salt_H.

Notably, two key lncRNAs (XLOC_014118, XLOC_001771) were related to all seven stress conditions. Expression analysis has revealed that XLOC_014118 was up-regulated in low Pi, low Mg, and high salt and down-regulated in other stress conditions, while XLOC_001771 was up-regulated in low Cu, low Mg, low Pi deficiency, and high-salt conditions and down-regulated in other stress conditions. XLOC_014118 is an antisense lncRNA with four isoforms (TCONS_00041596, TCONS_00042008, TCONS_00042400, TCONS_00042762) that were expressed

under different stress conditions via alternative splicing. For example, TCONS_00041596 was only up-regulated in low Pi deficiency condition; TCONS_00042008 was up-regulated both in low Pi deficiency and high-salt conditions, whereas TCONS_00042762 was up-regulated in multiple stress conditions, including low Pi deficiency, high salt, low Cu, low Fe, and low Mg deficiency conditions. XLOC_014118 is transcribed from antisense strand to the protein-coding gene (AT2G41510) (Supplementary Fig. S3) that encodes a protein whose sequence is similar to cytokinin oxidase/dehydrogenase, which catalyzes the degradation of cytokinins (CKs). Phytohormones were found critical in helping the plants to

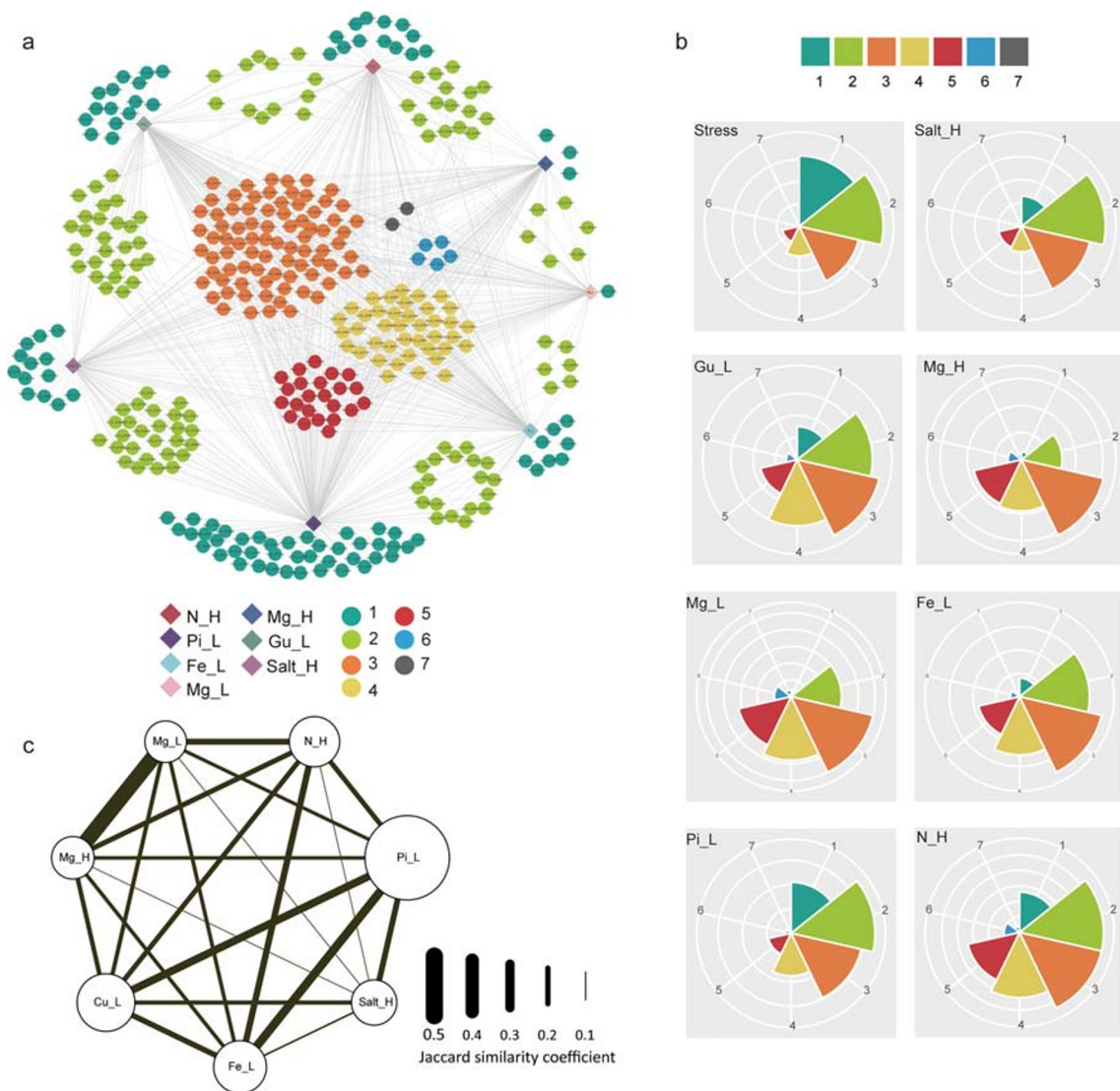


Fig. 5 The nutrient stress-key lncRNA network. **(a)** lncRNAs (circle) are differentially expressed under nutrient stress (rhombus). Different color rhombuses refer to different nutrient stress conditions, and different color circles refer to the number of linked stress conditions. **(b)** The distribution of key lncRNAs with respect to the number of linked stress conditions (1–

7) in each nutrient stress condition. The first figure (stress) refers to the total seven nutrient stress conditions. **(c)** “Convergence degree” between the seven nutrient stresses. Node degree refers to the number size of key lncRNAs under seven nutrient stress conditions; edge degree refers to convergence coefficient between two stress conditions

adapt to abiotic stress responses (Verma et al. 2016). Cytokinins could crosstalk with salicylic acid (SA) signaling cascade to regulate plant defenses, and the stabilized CK levels could exhibit enhanced resistance to abiotic stress (Jiang et al. 2013). In addition, there is a negative expression correlation between XLOC_014118 and AT2G41510 (Pearson correlation coefficient, $r = -0.396$). Therefore, we proposed that XLOC_014118 may down-regulate its anti-sense protein-coding gene and relieve the repression of degradation of cytokinins under multiple nutrient stress conditions. In addition, we found that splice sites of four lncRNA isoforms are in well-agreement with splice sites of identified lncRNAs in NONCODE (Fang et al. 2018), as well as with re-annotated new small peptide genes in Araport11 (Cheng et al.

2017). Therefore, the alternative splicing events of XLOC_014118 make sense. XLOC_001771 is an intergenic lncRNA between two protein-coding genes (AT1G33330, AT1G33340), whose expression patterns showed positive correlation (Pearson correlation coefficient, $r = 0.39$ between XLOC_001771 and AT1G33330, and $r = 0.34$ between XLOC_001771 and AT1G33340, respectively). Of note, AT1G33330 and AT1G33340 both are involved in stress response (Winter et al. 2007). Furthermore, one of the most well-characterized functions of plant lincRNAs is to epigenetically regulate gene expression by recruiting proteins for chromosome modification to specific loci (Yamada 2017). According to the prediction result of lncPro (Lu et al. 2013), XLOC_001771 is likely to interact with a chromatin structure-

remodeling complex protein (F4IHS2), which is involved in transcription regulation and nucleosome positioning.

Considering the merging trends of epigenetic regulation to nutrient stress response in plants, we further explored the DNA methylation pattern of XLOC_014118 and XLOC_001771 (Supplementary Fig. S4). We found that DNA methylation was enriched in the 5 kb upstream of TSS regions but not in ORF regions in leave (Ausin et al. 2012), endosperm (Hsieh et al. 2009), and flower (Qian et al. 2012b). Importantly, both lncRNAs were targeted by a histone acetyltransferase named IDM1 (for increased DNA methylation 1), a regulator of DNA demethylation in *A. thaliana* that can induce the demethylation and activate lncRNA transcription (Qian et al. 2012a).

Discussion

Numerous stress-responsive signaling pathways have incorporated miRNAs for many plants, while lncRNAs were less extensively studied. In this study, we presented a comprehensive investigation of lncRNAs in response to seven nutrient stresses in *A. thaliana* (salt_H, Cu_L, Mg_H, Mg_L, Fe_L, Pi_L, and N_H). Our integrative analysis of histone modification by ChIP-chip and RNA-seq data revealed several lncRNA properties regarding their epigenetic signatures, sequence features, and expression patterns. Key lncRNAs responsive to nutrient stresses were identified by conjoint analysis of ceRNA network and co-expression network. To our knowledge, this is the first attempt to characterize of lncRNAs under multiple nutrient stress conditions for *A. thaliana*, which sheds new light on the complex responsive signaling pathways involving lncRNAs to complement the miRNA participation in previous reports.

According to our lncRNA identification pipeline, 874 lncRNA transcripts were obtained (Supplementary Table S1), of which 455 were characterized and associated with seven types of nutrient stress response. It should be noted that they tend to be differentially expressed under no more than three nutrient stress conditions (Fig. 2b), suggesting that the aberrant expression of lncRNAs in response to nutrient stress is a stress-specific manner. Furthermore, these lncRNAs have high-stress specificity scores (Fig. 2c), and the ratio of differentially expressed lncRNAs is significantly higher than do PCGs in all seven stress conditions (Fig. 2d). This conservation is consistent with the previous study (Di et al. 2014) that lncRNAs are more sensitive to nutrient stress than PCGs.

To date, the function of lncRNAs in multiple nutrient stresses remains poorly understood. The construction of ceRNA network and CEN provides a glimpse into the potential functions of lncRNAs in nutrient stress response. On one hand, many miRNAs in the ceRNA network have been

reported as being dysregulated under different stress conditions, such as high salinity, nutrient starvation (C, S, N, Pi), and As stress. Those multifunctional miRNAs in the ceRNA network serve as good indicators of lncRNA regulation under different nutrient stresses. On the other hand, GO enrichment was carried out to speculate possible functions of lncRNAs involved in these two networks, respectively. Seventeen lncRNAs in ceRNA network and 86 lncRNAs in CEN were successfully annotated with stress-responsive related GO terms, respectively (Fig. 3c and Fig. 4b). In total, we found 67 key lncRNAs enriched in nutrient stress-responsive associated GO terms, such as “transport of copper ion and iron ion” and “response to salt stress.” Here, we propose that some lncRNAs may act as ceRNAs to block the miRNA functions, whereas some lncRNAs might be selectively regulated upon stress stimulus together with protein-coding genes to have synergetic responses.

Comparative analysis of reported work revealed that 21 key lncRNAs were responsive to Pi deprivation (Yuan et al. 2016b) and five lincRNAs were also involved in other stress types (Liu et al. 2012). Therefore, it is important to figure out the convergent downstream responses among different nutrient stress types and understand the involvement of lncRNAs in various biological processes, such as transcriptional, post-transcriptional, and epigenetic processes. However, determination of the precise mechanism of lncRNA function and elaborate experiments still remain challenging. Among seven nutrient stresses, Pi_L has more stress-specific expressed lncRNAs, while Mg_H and Mg_L have more lncRNAs shared between multiple stresses. In general, the majority of key lncRNAs were related with two or three stress conditions (Fig. 5b). The convergence degree has revealed highest convergence between Mg_L and Mg_H stresses (Fig. 5c), which is reasonable as two stress conditions refer to the abnormality of the same nutrient. The perception mode and several signaling pathways may be essential for both Mg deprivation and Mg excess, in which lncRNAs share largely overlapping transcriptional patterns. Comparatively, salt_H showed the lowest convergence compared with any of the other stresses (Fig. 5c).

In addition, two key lncRNAs (XLOC_014118, XLOC_001771) were related to all seven stress conditions. XLOC_014118 may function as the nature antisense transcript of mRNA that down-regulates cytokinin oxidase/dehydrogenase to maintain the stabilized CK levels. The role of CKs crosstalk with other phytohormones to function in drought and salinity stress responses has been demonstrated by several studies (Argueso et al. 2009; Jiang et al. 2013; Tran et al. 2007). The potential relationship between the key lncRNA and phytohormones has enhanced our understanding of lncRNAs' role in response to nutrient stresses. Based on the lncRNA-protein interaction, XLOC_001771 is possible to interact with chromatin structure-remodeling complex proteins involved in the transcription regulation and nucleosome

positioning. Our epigenetic analysis showed that the promoter regions of these two key lncRNAs all presented high DNA methylation levels in wide-type tissues (Supplementary Fig. S4). However, their detailed function mechanisms are unable to be disclosed in our study. Subsequent experimental characterization is needed for these two candidate lncRNAs.

To conclude, our genome-wide analysis of stress-responsive lncRNAs has provided a foundation for wet-bench experiments that are highly required to validate modeling results and to determine the precise roles of key lncRNAs related to multiple nutrient stresses.

Materials and methods

Data collection and processing

Raw data (RNA-seq datasets) were downloaded from NCBI SRA, including 14 nutrient stress samples and 12 untreated control (“normal”) samples (Table 1). The samples involve six nutrients, sodium (Na), cuprum (Cu), magnesium (Mg), ferrum (Fe), phosphate (Pi), and nitrogen (N), and seven stress conditions, high salt (salt_H), low Cu (Cu_L), high Mg (Mg_H), low Mg (Mg_L), low Fe (Fe_L), low Pi (Pi_L), and high N (N_H). For all the RNA-seq data, sequence reads were aligned to *A. thaliana* genome (TAIR10) using TopHat v2.1.1 with default parameters (Kim et al. 2013). The aligned files (in bam format) from TopHat were then fed to Cufflinks v2.1.1 to perform transcript assembly and abundance estimation (Trapnell et al. 2010). Finally, we merged all assemblies into final transcripts using Cuffcompare (Trapnell et al. 2010).

To evaluate differential lncRNA expression between stress and normal samples, we used sequence counts estimated by HTSeq (Anders et al. 2015) as input for DESeq (Anders and Huber 2010). Replicates were used independently for statistical analysis of gene expression. We filtered lncRNAs whose expression fold changes (FC) less than 2.

The ChIP-chip data for histone modification was GSE24710 (Roudier et al. 2011), which included H3K4me3 and H3K27me3 in the root. The normalized datasets of ChIP-chip were downloaded and converted into TAIR10 compatible coordinates using Perl script from <http://1001genomes.org/>. We applied average enrichment signal for two histone modifications in coding sequencing (ORF) and upstream regulatory (promoter) regions (500 bp upstream of the TSS).

lncRNAs identification pipeline

The assembled transcripts of 14 nutrient stress samples and 12 normal RNA-seq samples from Cuffcompare were used to identify lncRNAs through the following steps (Fig. 1a). Firstly, transcripts were retained as the potential lncRNAs with a class code of “u,” “i,” or “x,” standing for intergenic,

intronic, and antisense transcripts, respectively. Secondly, transcripts with protein-coding potential were discarded. The coding potential of each remaining transcript was estimated by two software packages: CPAT (Wang et al. 2013) and CNCI (Sun et al. 2013). We kept the transcripts that were marked as “non-coding” by two software packages. Third, the transcripts shorter than 200 nt were removed. At last, the remaining transcripts with FPKM values less than 1 were filtered out in all of the samples.

Furtherly, we compared the sequences of identified lncRNAs with known lncRNAs and miRNAs. NCBI *BLASTALL* (version 2.2.26) was run with options `-p blastn`. We assessed whether an identified lncRNA matched any annotated lncRNA in CANTATA database (Szczesniak et al. 2019) and GREENC database (Paytavi Gallart et al. 2016). *BLASTALL* was run against the miRBase (Kozomara et al. 2019) and PMRD (Zhang et al. 2010) to check whether the identified lncRNAs matched long miRNA hairpin sequences that are bigger than 200 nt. In all cases, an *E* value of 0.01 and a bits score of 50 were taken as thresholds.

Construction of competing endogenous RNA network

The ceRNA network was constructed based on the crosstalk of miRNAs and their targets. First, for each transcript (mRNA or lncRNA), all its miRNAs regulators were predicted using psRNA target (Dai and Zhao 2011) and TAPIR (Bonnet et al. 2010). Second, for a given RNA pair (A and B), the hypergeometric test is used to compute the significance of shared miRNAs for A and B.

$$P\text{-value} = 1 - F(x/U, M, N) = 1 - \sum_{i=0}^{x-1} \frac{\binom{M}{i} \binom{U-M}{N-i}}{\binom{U}{N}}$$

where *M* is the number of miRNAs regulating A, *N* is the number of miRNAs regulating B, *U* is the number of miRNAs that regulate A or B, and *x* presents the number of common miRNAs that regulate both A and B. RNA pairs with *P* value less than 0.05 were kept as the candidate ceRNA pairs.

Third, it has been reported that the expression of ceRNA pairs was positively correlated with each other (Salmena et al. 2011). Thus, Pearson correlation coefficient of each candidate ceRNA pairs was calculated based on gene expression data from all of the stress samples, and only positively correlated pairs with a *P* value less than 0.05 were set as the final ceRNA pairs. Finally, all the ceRNA pairs were combined together to construct the whole ceRNA network, and we then generated the lncRNA-related ceRNA network based on the whole ceRNA network.

In addition, the GO enrichment analysis was performed on all the known coding genes in the ceRNA network using Cytoscape plug-in BiNGO with the “GO Biological Process” category (Maere et al. 2005). REVIGO (Supek et al. 2011) was applied to remove redundant GO terms and visualize the GO enrichment results.

lncRNA-associated co-expression network and cluster construction

We built a stress-related co-expression network based on 14 stress samples. The Pearson correlation coefficient (Pcc) was calculated to define co-expression pairs ($Pcc > 0.9$ as the cut-off and FDR cutoff of $1e-4$). After assembling lncRNA-associated co-expression pairs, we generated lncRNA-associated CEN. Moreover, network clustering was performed on CEN with the MCODE package (Bader and Hogue 2003). Default suggested parameters were used, and then the GO enrichment analysis was conducted for the genes in each cluster using BiNGO plug-in (Maere et al. 2005). REVIGO software (Supek et al. 2011) was used to summarize GO terms together with their *P* value and visualize the enriched GO terms.

Statistical analysis and data visualization

Student's *t* test, Wilcoxon rank sum test, and Pearson chi-square test were used to compare the genetic and epigenetic features between lncRNAs and PCGs. *P* value less than 0.05 was considered statistically significant. All statistical analyses in this study were performed by using the R software (<https://www.r-project.org/>, release 3.3.0).

“Convergence degree” was deduced between each pair of the seven nutrient stresses based on the Jaccard similarity coefficient to measure the fractions of key lncRNAs that were shared between two stresses. The ratio value ranges from 1 (absolute convergence) to 0 (no convergence) and is defined as the size of the intersection divided by the size of the union of the sample sets: $J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$, where *X* and *Y* present the number of key lncRNAs in stress *X* and stress *Y*, respectively. Interactions between lncRNA-protein pairs were predicted by the software lncPro (Lu et al. 2013). The networks presented in this study were visualized using Cytoscape platform (Shannon et al. 2003). The network topology analysis was performed using NetworkAnalyzer (Assenov et al. 2008) plug-in for Cytoscape. In addition, The MeV (Saeed et al. 2003) software was used to view the gene expression profiling data.

Acknowledgements We are grateful to the other members of Ming Chen's laboratory for helpful discussions and helpful comments. The authors are also grateful for the critical comments of anonymous reviewers.

Authors' contributions J.W. and Q.C. designed the study. W.W., Y.C., and Y. Z. collected the data and performed the analysis. J.W. drafted the manuscript. G.G. and M. C. completed the final version of the manuscript. All co-authors read and approved the final manuscript.

Funding This work was supported by the National Natural Sciences Foundation of China (No. 31771477).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Consent for publication Not applicable.

Ethics approval and consent to participate Not applicable.

References

- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:Artn R106. <https://doi.org/10.1186/Gb-2010-11-10-R106>
- Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Argueso CT, Ferreira FJ, Kieber JJ (2009) Environmental perception avenues: the interaction of cytokinin and environmental response pathways. *Plant Cell Environ* 32:1147–1160
- Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. *Bioinformatics* 24:282–284. <https://doi.org/10.1093/bioinformatics/btm554>
- Ausin I et al (2012) INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in Arabidopsis. *Proc Natl Acad Sci USA* 109:8374–8381. <https://doi.org/10.1073/pnas.1206638109>
- Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4:2
- Bernal M, Casero D, Singh V, Wilson GT, Grande A, Yang H, Dodani SC, Pellegrini M, Huijser P, Connolly EL, Merchant SS, Kramer U (2012) Transcriptome sequencing identifies SPL7-regulated copper acquisition genes FRO4/FRO5 and the copper dependence of iron homeostasis in Arabidopsis. *Plant cell* 24:738–761. DOI 10.1105/tpc.111.090431
- Bonnet E, He Y, Billiau K, Van de Peer Y (2010) TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics* 26:1566–1568. <https://doi.org/10.1093/bioinformatics/btq233>
- Chekanova JA (2015) Long non-coding RNAs and their functions in plants. *Curr Opin Plant Biol* 27:207–216. <https://doi.org/10.1016/j.pbi.2015.08.003>
- Chen M, Wang C, Bao H, Chen H, Wang Y (2016) Genome-wide identification and characterization of novel lncRNAs in *Populus* under nitrogen deficiency. *Mol Genet Genom* : MGG 291:1663–1680. <https://doi.org/10.1007/s00438-016-1210-3>
- Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017) Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J* 89:789–804. <https://doi.org/10.1111/tj.13415>

- Chung PJ, Jung H, Jeong D-H, Ha S-H, Do Choi Y, Kim J-K (2016) Transcriptome profiling of drought responsive noncoding RNAs and their target genes in rice. *BMC Genomics* 17:563
- Cruz de Carvalho MH, Sun HX, Bowler C, Chua NH (2016) Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *New Phytol* 210:497–510. <https://doi.org/10.1111/nph.13787>
- Dai XB, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res* 39:W155–W159. <https://doi.org/10.1093/nar/gkr319>
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhatar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigo R (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22:1775–1789
- Di C et al (2014) Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J* 80:848–861. <https://doi.org/10.1111/tpj.12679>
- Fang S et al (2018) NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res* 46:D308–D314. <https://doi.org/10.1093/nar/gkx1107>
- Franco-Zorrilla JM et al (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* 39:1033–1037. <https://doi.org/10.1038/ng2079>
- Fu X-D (2014) Non-coding RNA: a new frontier in regulatory biology. *Natl Sci Rev* 1:190–204
- Haak DC, Fukao T, Grene R, Hua Z, Ivanov R, Perrella G, Li S (2017) Multilevel Regulation of Abiotic Stress Responses in Plants. *Front Plant Sci* 8:1564. <https://doi.org/10.3389/fpls.2017.01564>
- Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D (2009) Genome-Wide Demethylation of *Arabidopsis* Endosperm. *Science* 324:1451–1454. <https://doi.org/10.1126/science.1172417>
- Hu L, Xu ZY, Hu BQ, Lu ZJ (2017) COME: a robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. *Nucleic Acids Res* 45:Artn E2. <https://doi.org/10.1093/Nar/Gkw798>
- Jiang C-J, Shimono M, Sugano S, Kojima M, Liu X, Inoue H, Sakakibara H, Takatsuji H (2013) Cytokinins act synergistically with salicylic acid to activate defense gene expression in rice. *Mol Plant-Microbe Interact* 26:287–296
- Kim D, Perteza G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36. <https://doi.org/10.1186/gb-2013-14-4-r36>
- Kozomara A, Birgaoanu M, Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res* 47:D155–D162. <https://doi.org/10.1093/nar/gky1141>
- Kumar S, Verma S, Trivedi PK (2017) Involvement of small RNAs in phosphorus and sulfur sensing, signaling and stress: current update. *Front Plant Sci* 8:Artn 285. <https://doi.org/10.3389/fpls.2017.00285>
- Lan P, Li W, Schmidt W (2012) Complementary proteome and transcriptome profiling in phosphate-deficient *Arabidopsis* roots reveals multiple levels of gene regulation. *Mol Cell Proteomics* 11:1156–1166. DOI 10.1074/mcp.M112.020461
- Li S et al (2017) Genome-wide identification and functional prediction of cold and/or drought-responsive lncRNAs in cassava. *Sci Rep* 7:45981
- Li W, Lin WD, Ray P, Lan P, Schmidt W (2013) Genome-wide detection of condition-sensitive alternative splicing in *Arabidopsis* roots. *Plant Physiol* 162:1750–1763. DOI 10.1104/pp.113.217778
- Liang G, Ai Q, Yu DQ (2015) Uncovering miRNAs involved in crosstalk between nutrient deficiencies in *Arabidopsis*. *Sci Rep-Uk* 5:Artn 11813. <https://doi.org/10.1038/Srep11813>
- Liu J et al (2012) Genome-wide analysis uncovers regulation of long Intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 24:4333–4345. <https://doi.org/10.1105/tpc.112.102855>
- Liu X, Hao L, Li D, Zhu L, Hu S (2015) Long non-coding RNAs and their biological roles in plants. *Genomics Proteomics Bioinformatics* 13:137–147. <https://doi.org/10.1016/j.gpb.2015.02.003>
- Lu Q, Ren S, Lu M, Zhang Y, Zhu D, Zhang X, Li T (2013) Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics* 14:651
- Lv Y et al (2016) Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (*Zea mays* L.). *BMC Genomics* 17:350. <https://doi.org/10.1186/s12864-016-2650-1>
- Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21:3448–3449. <https://doi.org/10.1093/bioinformatics/bti551>
- Meng J, Liu D, Luan YS (2015) Inferring plant microRNA functional similarity using a weighted protein-protein interaction network. *BMC bioinformatics* 16:Artn 360. <https://doi.org/10.1186/S12859-015-0789-4>
- Ning S et al (2016) Construction of a lncRNA-mediated feed-forward loop network reveals global topological features and prognostic motifs in human cancers. *Oncotarget*. <https://doi.org/10.18632/oncotarget.10004>
- Niu Y, Ahammed GJ, Tang C, Guo L, Yu J (2016) Physiological and Transcriptome Responses to Combinations of Elevated CO₂ and Magnesium in *Arabidopsis thaliana*. *PLoS One* 11:e0149301. DOI 10.1371/journal.pone.0149301
- Paytvi Gallart A, Hermoso Pulido A, Anzar Martinez de Lagran I, Sanseverino W, Aiese Cigliano R (2016) GREENC: a Wiki-based database of plant lncRNAs. *Nucleic Acids Res* 44:D1161–D1166. <https://doi.org/10.1093/nar/gkv1215>
- Qian W et al (2012a) A histone acetyltransferase regulates active DNA demethylation in *Arabidopsis*. *Science* 336:1445–1448
- Qian WQ, Miki D, Zhang H, Liu Y, Zhang X, Tang K, Kan Y, la H, Li X, Li S, Zhu X, Shi X, Zhang K, Pontes O, Chen X, Liu R, Gong Z, Zhu JK (2012b) A Histone Acetyltransferase Regulates Active DNA Demethylation in *Arabidopsis*. *Science* 336:1445–1448. <https://doi.org/10.1126/science.1219416>
- Roudier F et al (2011) Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J* 30:1928–1938. <https://doi.org/10.1038/emboj.2011.103>
- Saeed AI et al (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–+
- Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell* 146:353–358. <https://doi.org/10.1016/j.cell.2011.07.014>
- Sani E, Herzyk P, Perrella G, Colot V, Amtmann A (2013) Hyperosmotic priming of *Arabidopsis* seedlings establishes a long-term somatic memory accompanied by specific changes of the epigenome. *Genome Biol* 14:R59. DOI 10.1186/gb-2013-14-6-r59
- Shafiq S, Li J, Sun Q (2016) Functions of plants long non-coding RNAs. *Biochim Biophys Acta* 1859:155–162. <https://doi.org/10.1016/j.bbagn.2015.06.009>
- Shannon P et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>
- Shumayla SS, Taneja M, Tyagi S, Singh K, Upadhyay SK (2017) Survey of high throughput RNA-Seq data reveals potential roles for lncRNAs during development and stress response in bread wheat. *Front Plant Sci* 8:1019. <https://doi.org/10.3389/fpls.2017.01019>

- Srivastava S, Srivastava AK, Suprasanna P, D'Souza SF (2013) Identification and profiling of arsenic stress-induced microRNAs in *Brassica juncea*. *J Exp Bot* 64:303–315. <https://doi.org/10.1093/jxb/ers333>
- Stauffer E, Maizel A (2014) Post-transcriptional regulation in root development. *Wiley Interdiscip Rev: RNA* 5:679–696
- Sun L et al (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res* 41:ARTN e166. <https://doi.org/10.1093/nar/gkt646>
- Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6: e21800. <https://doi.org/10.1371/journal.pone.0021800>
- Szczesniak MW, Bryzghalov O, Ciomborowska-Basheer J, Makalowska I (2019) CANTATAdb 2.0: expanding the collection of plant long noncoding RNAs. *Methods Mol Biol* 1933:415–429. https://doi.org/10.1007/978-1-4939-9045-0_26
- Tran L-SP, Urao T, Qin F, Maruyama K, Kakimoto T, Shinozaki K, Yamaguchi-Shinozaki K (2007) Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in *Arabidopsis*. *Proc Natl Acad Sci* 104:20623–20628
- Trapnell C et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515. <https://doi.org/10.1038/nbt.1621>
- Verma V, Ravindran P, Kumar PP (2016) Plant hormone-mediated regulation of stress responses. *BMC Plant Biol* 16:86. <https://doi.org/10.1186/s12870-016-0771-y>
- Vidal EA, Moyano TC, Krouk G, Katari MS, Tanurdzic M, McCombie WR, Coruzzi GM, Gutierrez RA (2013) Integrated RNA-seq and sRNA-seq analysis identifies novel nitrate-responsive genes in *Arabidopsis thaliana* roots. *BMC Genomics* 14:701. DOI 10.1186/1471-2164-14-701
- Villegas VE, Zaphiropoulos PG (2015) Neighboring gene regulation by antisense long non-coding RNAs. *Int J Mol Sci* 16:3251–3266
- Wang L, Park HJ, Dasari S, Wang SQ, Kocher JP, Li W (2013) CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res* 41:ARTN e74. <https://doi.org/10.1093/nar/gkt006>
- Wang J, Meng X, Dobrovolskaya OB, Orlov YL, Chen M (2017) Non-coding RNAs and their roles in stress response in plants genomics. *Proteomics Bioinformatics* 15:301–312. <https://doi.org/10.1016/j.gpb.2017.01.007>
- Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ (2007) An "electronic fluorescent pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One* 2:e718. <https://doi.org/10.1371/journal.pone.0000718>
- Xu J et al (2015) The mRNA related ceRNA-ceRNA landscape and significance across 20 major cancer types. *Nucleic Acids Res* 43: 8169–8182. <https://doi.org/10.1093/nar/gkv853>
- Xu XW, Zhou XH, Wang RR, Peng WL, An Y, Chen LL (2016) Functional analysis of long intergenic non-coding RNAs in phosphate-starved rice using competing endogenous. *RNA network Sci rep-Uk* 6:Artn 20715. <https://doi.org/10.1038/Srep20715>
- Yamada M (2017) Functions of long intergenic non-coding (linc) RNAs in plants. *J Plant Res* 130:67–73. <https://doi.org/10.1007/s10265-016-0894-0>
- Yanai I et al (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21:650–659. <https://doi.org/10.1093/bioinformatics/bti042>
- Yuan C, Wang J, Harrison AP, Meng X, Chen D, Chen M (2015) Genome-wide view of natural antisense transcripts in *Arabidopsis thaliana*. *DNA Res* 22:233–243. <https://doi.org/10.1093/dnares/dsv008>
- Yuan C, Meng X, Li X, Illing N, Ingle RA, Wang J, Chen M (2016a) PceRBase: a database of plant competing endogenous. *RNA Nucleic Acids Res* 45:D1009–D1014
- Yuan JP, Zhang Y, Dong JS, Sun YZ, Lim BL, Liu D, Lu ZJ (2016b) Systematic characterization of novel lincRNAs responding to phosphate starvation in *Arabidopsis thaliana*. *BMC Genomics* 17:Artn 655. <https://doi.org/10.1186/S12864-016-2929-2>
- Zeng HQ, Wang GP, Hu XY, Wang HZ, Du LQ ZYY (2014) Role of microRNAs in plant responses to nutrient stress. *Plant Soil* 374: 1005–1021. <https://doi.org/10.1007/s11104-013-1907-6>
- Zhang Z et al (2010) PMRD: plant microRNA database. *Nucleic Acids Res* 38:D806–D813. <https://doi.org/10.1093/nar/gkp818>
- Zhang Y-C et al (2014) Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol* 15:512
- Zhou M et al (2016) Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lincRNA biomarkers in human ovarian cancer. *Oncotarget* 7:12598–12611. <https://doi.org/10.18632/oncotarget.7181>

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