

RESEARCH ARTICLE

Open Access



# Genome-wide identification and analysis of long noncoding RNAs (lncRNAs) during seed development in peanut (*Arachis hypogaea* L.)

Xingli Ma<sup>1</sup>, Xingguo Zhang<sup>1</sup>, Sy Mamadou Traore<sup>2</sup>, Zeyu Xin<sup>1</sup>, Longlong Ning<sup>1</sup>, Ke Li<sup>1</sup>, Kunkun Zhao<sup>1</sup>, Zhongfeng Li<sup>1</sup>, Guohao He<sup>2</sup> and Dongmei Yin<sup>1\*</sup>

## Abstract

**Background:** Long noncoding RNAs (lncRNAs) have several known functions involving various biological regulatory processes in plant. However, the possible roles of lncRNAs during peanut seed development have not been fully explored.

**Results:** In this study, two peanut recombinant inbred lines (RIL<sub>8</sub>) that differ in seed size were used to investigate comprehensive lncRNA profiles derived from the seed development at 15 and 35 days after flowering (DAF). We identified a total of 9388 known and 4037 novel lncRNAs, from which 1437 were differentially expressed lncRNAs. Interestingly, the expression patterns of a number of lncRNAs can be very different between two closely related inbred lines and these lncRNAs were expressed predominantly in only one RIL at 35 DAF. Some differentially expressed lncRNAs were found related to putative cis-acting target genes and predicted to be involved in transcription, transport, cell division, and plant hormone biosynthesis. The expression patterns of several representative lncRNAs and 12 protein-coding genes were validated by qPCR. Same expression pattern was observed between most lncRNAs and their target genes. 11 lncRNAs, XR\_001593099.1, MSTRG.18462.1, MSTRG.34915.1, MSTRG.41848.1, MSTRG.22884.1, MSTRG.12404.1, MSTRG.26719.1, MSTRG.35761.1, MSTRG.20033.1, MSTRG.13500.1, and MSTRG.9304.1 and their cis-acting target genes may play key roles in peanut seed development.

**Conclusions:** These results provided new information on lncRNA-mediated regulatory roles in peanut seed development, contributing to the comprehensive understanding of the molecular mechanisms involved in peanut seed development.

**Keywords:** Peanut, lncRNA, Seed development, RNA-Seq, Transcriptional regulation

## Background

In recent years, the use of the next generation sequencing approaches revealed that transcription in eukaryotes is complex. A large number of eukaryotic genomes are universally transcribed producing coding and non-coding RNAs (ncRNAs). Non-coding RNAs (ncRNAs)

are a class of RNAs that cannot code for proteins; however, they play important regulatory roles in numerous biological processes [1]. ncRNAs can be classified based on their lengths into small RNAs (< 200 nt) and long non-coding RNAs (lncRNAs, longer than 200 nt). Small RNAs are further classified as microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), small interfering RNAs (siRNAs), and small nuclear RNA (snRNAs) [2]. lncRNAs are categorized as long intergenic noncoding

\* Correspondence: [yindm@126.com](mailto:yindm@126.com)

<sup>1</sup>College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China  
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

RNAs (lincRNAs), natural antisense transcripts (NATs) and intronic RNAs (incRNAs) according to the genomic location and context [3, 4].

lncRNAs exhibit tissue and cell-specific expression patterns and tend to show poor conservation across different species [5, 6]. Previous studies found that lncRNAs involve in several activities, including gathering, transporting proteins, regulating promoter activities through proximal *cis*-acting or *trans*-acting sequences, and epigenetic modification, silencing or repression [7, 8]. lncRNAs also involve in growth and development, disease occurrence in mammals [9, 10].

Several biological functions of lncRNAs have been characterized in various plants. For example, *COOLAIR* and *COLDAIR* regulate the expression of *FLOWERING LOCUS C (FLC)* which determines *Arabidopsis* flowering time [11, 12]. In hybrid rice, photoperiod-sensitive male sterility is correlated with a lncRNA, known as long-day-specific male-fertility-associated RNA (LDMAR) [13]. In cotton seedlings, two lncRNAs, GhlncNAT-ANX2 and GhlncNAT-RLP7, provided an enhanced resistance against fungal pathogens *Verticillium dahliae* and *Botrytis cinerea* [14]. In *Medicago truncatula*, the lncRNA *Enod40* interacts with soil rhizobia for nodule formation by inducing a re-localization of a nuclear RBP [15]. Despite this progress, there is still a paucity of studies addressing the role of the lncRNA in plants, more importantly peanut. To date, genome-wide analyses of lncRNAs have been carried out in several plants, but the numbers and characteristics of lncRNAs involved in seed development have not yet been explored in peanut.

Peanut (*Arachis hypogaea* L.), an allotetraploid species ( $2n = 4x = 40$ ; AABB), is an important crop grown worldwide for both oil and protein production. Peanut seed size is a main agronomic trait for breeders, therefore understanding the molecular mechanisms underlying the development of the peanut seed is a high priority for peanut researchers and breeders. Recent research studies have focused on the identification of some important genes contributing highly to peanut seed development [16]. Moreover, the function of several miRNAs and their target genes involved in peanut seed development has been characterized [17]. However, there is no report provided any information on the involvement of lncRNAs and their target genes in peanut seed development mechanisms. In this study, we used two peanut RILs to gain a better understanding of the function of lncRNA in peanut seed development. The expression profiles of lncRNAs at 15 and 35 days after flowering (DAF) were characterized in two sister lines from a RIL<sub>8</sub>, where line 8106 has medium-sized pods and line 8107 has super-large pods with different seed sizes. We identified and validated lncRNAs in peanut; then identified the differentially expressed lncRNAs (DELs) by

comparing and analyzing the expression profile of the lncRNAs in different seed developmental stages between the two peanut RILs; and therefore, understanding the possible roles of DELs in peanut seed development.

## Results

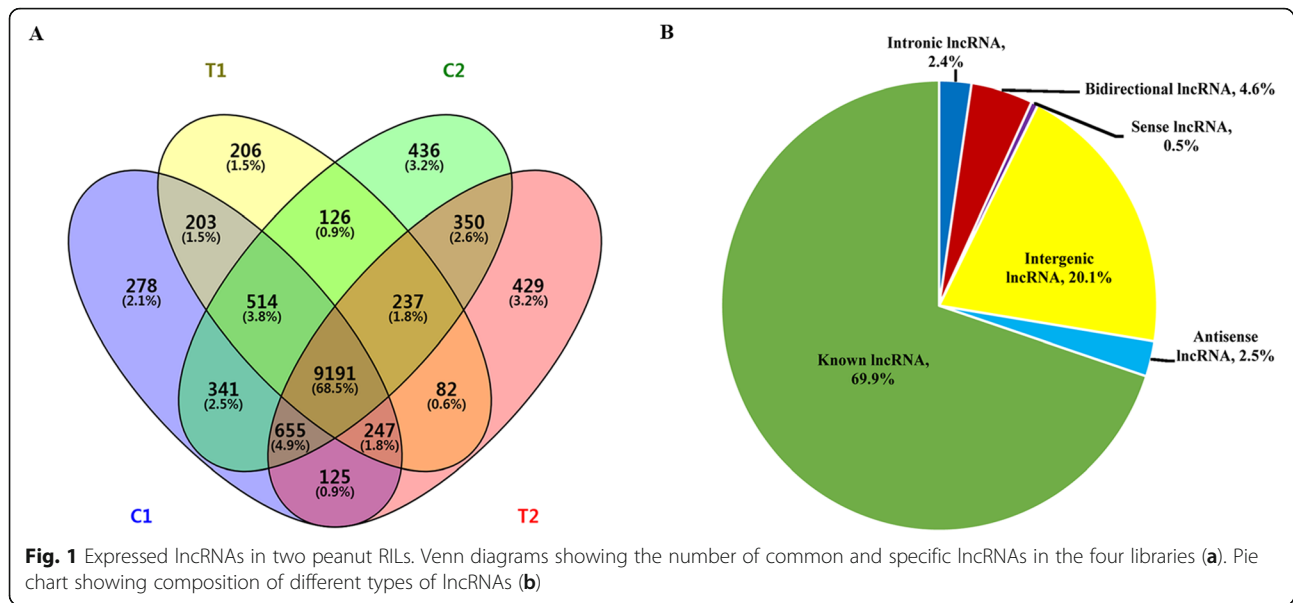
### Sequencing of peanut lncRNAs

In order to investigate the dynamic variation of lncRNAs during peanut seed development, the whole-transcriptome strand-specific RNA sequencing for two peanut RILs at two seed developmental stages (15 DAF and 35 DAF) was performed with three biological replicates. In total, more than 100 million raw reads were generated by high-throughput sequencing. Fast QC with a phred-like algorithm provided a mean quality score (Q30%) greater than 95%, indicating the RNA-seq data was highly reliable. From the raw reads, more than 97% reads were clean reads (Additional file 1: Table S1).

### Identification and characterization of lncRNAs in peanut

The clean reads were mapped to the two diploid *Arachis* reference genomes (*A. duranensis* and *A. ipaensis*) using the TopHat. Then, the transcripts were assembled and annotated using the StringTie. Mostly, 73,180, 72,725, 74,702, 74,324, 74,027, 73,964, 69,645, 69,155, 70,184, 70,516, 69,935, and 72,215 unique mRNAs from the twelve cDNA libraries were identified, including C1(RIL8106-15DAF), C2(RIL8107-15DAF), T1(RIL8106-35DAF), and T2 (RIL8107-35DAF), each with three replicated samples, respectively (Additional file 2: Table S2). The remaining transcripts were filtered according to their length and coding potential; transcripts shorter than 200 bp were discarded, and transcripts with potential CPC score  $> -1$  and CNCI score  $> 0$  were removed [18, 19]. The remaining transcripts were considered as lncRNAs, from which, 9528, 9429, 10,177, 10,274, 10,089, 10,131, 9005, 8867, 9044, 9388, 9398, and 9693 unique lncRNAs were identified from the twelve samples, respectively (Additional file 2: Table S2). In total, 13,425 unique lncRNAs were identified in this study (Additional file 3: Table S3). The numbers of lncRNAs found in the two RILs at 15 DAF (11,554 in C1 and 11,850 in C2) were considerably higher than those detected at 35 DAF (10,806 in T1 and 11,316 in T2), and 9191 lncRNAs were common to all four samples (Fig. 1a).

All lncRNAs were mapped to the 20 chromosomes of the peanut genome. Our result indicated that the lncRNAs were evenly distributed across these chromosomes without a preference of location (Fig. 2a and b). According to the genomic locations of these lncRNAs, 2693 intergenic (20.1%), 616 bidirectional (4.6%), 340 antisense (2.5%), 320 intronic (2.4%), 68 sense lncRNAs (0.5%) were identified. In addition, 9388 known lncRNAs (69.9%) were also detected in this study (Fig. 1b). The



estimation of the length of these lncRNAs revealed that most lncRNA possessed more than 1000 bp (Fig. 2c). Furthermore, the assessment of the expression level of each transcript using fragments per kilobase of exon model per million mapped reads (FPKM) showed that the overall expression level and number of lncRNAs were lower than those of mRNAs (Fig. 2d and e). The result was consistent with previous study in the upland cotton [20].

#### Analysis of differentially expressed lncRNAs in peanut

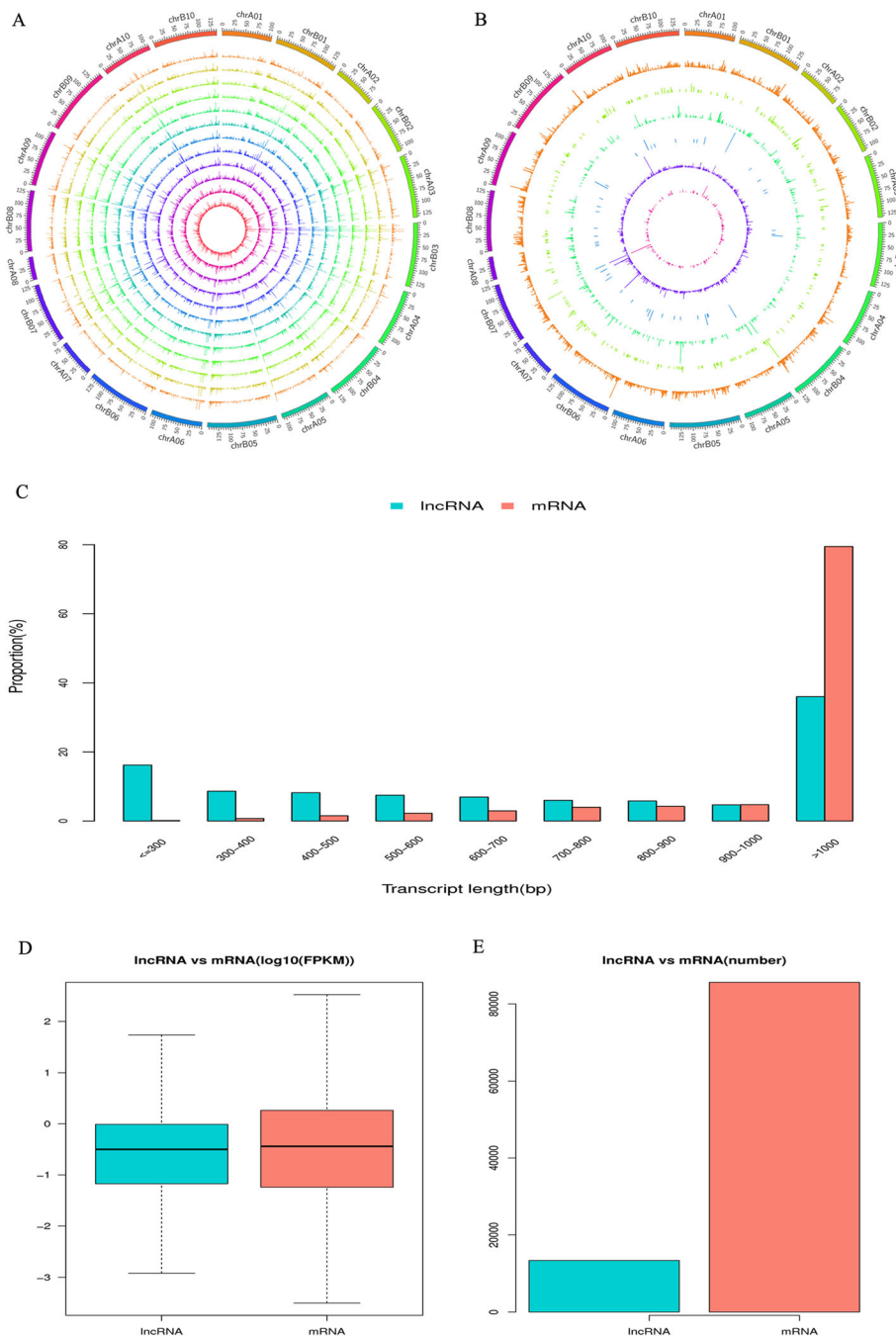
To analyze the difference of expression of lncRNAs at 35 DAF and 15 DAF between the two peanut RILs, we compared the normalized expression (FPKM) of lncRNAs amongst all libraries. The following criteria were used to identify differentially expressed lncRNAs in the comparison of different groups: (1)  $\log_2$  (fold change)  $> 1$  or  $\log_2$  (fold change)  $< -1$ , and (2) statistical significance ( $p$  value  $< 0.05$ ). We finally identified 2178 differentially expressed lncRNAs from four different comparisons. Among these, 594, 1019, 691, and 746 differentially expressed lncRNAs were found in comparisons of C2 vs. C1, T2 vs. T1, T1 vs. C1, and T2 vs. C2, respectively, and 24 differentially expressed lncRNAs were common in the four comparisons (Fig. 3 and Additional file 4: Table S4–1, S4–2, S4–3, and S4–4). Moreover, the differentially expressed lncRNAs in four comparisons between two RILs showed that the number of up- and down-regulated lncRNAs was close, such as 294 up and 397 down-regulated lncRNAs in T1 vs. C1, 378 up and 368 down in T2 vs. C2, and 492 up and 527 down in T2 vs. T1 (Fig. 4a). However, significant difference of 392 up- and 202 down-regulated lncRNAs was identified in C2 vs. C1 where the large-seeded line

(RIL8107) compared with the medium-seeded line (RIL8106) at the 15 DAF stage (Fig. 4a).

In addition, we clustered the 67 differentially expressed lncRNAs on the basis of their expression patterns in the two developmental stages between the two peanut RILs. Our result showed a difference of expression between the lncRNAs at 35 DAF stage in the two RILs. For example, XR\_001590629.1 and MSTRG.37421.1 were down-regulated, while XR\_001593099.1 was up-regulated in the two RILs (Fig. 4b). Interestingly, many lncRNAs were expressed predominantly in only one of the RILs at the 35 DAF stage. Fifteen (15) lncRNAs expression, including XR\_001593511.1, MSTRG.18462.1, MSTRG.31185.2, MSTRG.34915.1, MSTRG.41848.1, were up-regulated only in medium-size seed line of RIL8106, while only nine (9) lncRNAs including XR\_001614585.1, XR\_001593524.1, MSTRG.22884.1, XR\_001621358.1, XR\_001621694.1, MSTRG.12404.1, MSTRG.13652.1, MSTRG.26719.1, and MSTRG.35761.1 were up-regulated only in large-size seed line of RIL 8107 (Fig. 4b). These results suggested that these differentially expressed lncRNAs might play an important role in regulating peanut seed development.

#### Validation of differentially expressed lncRNAs

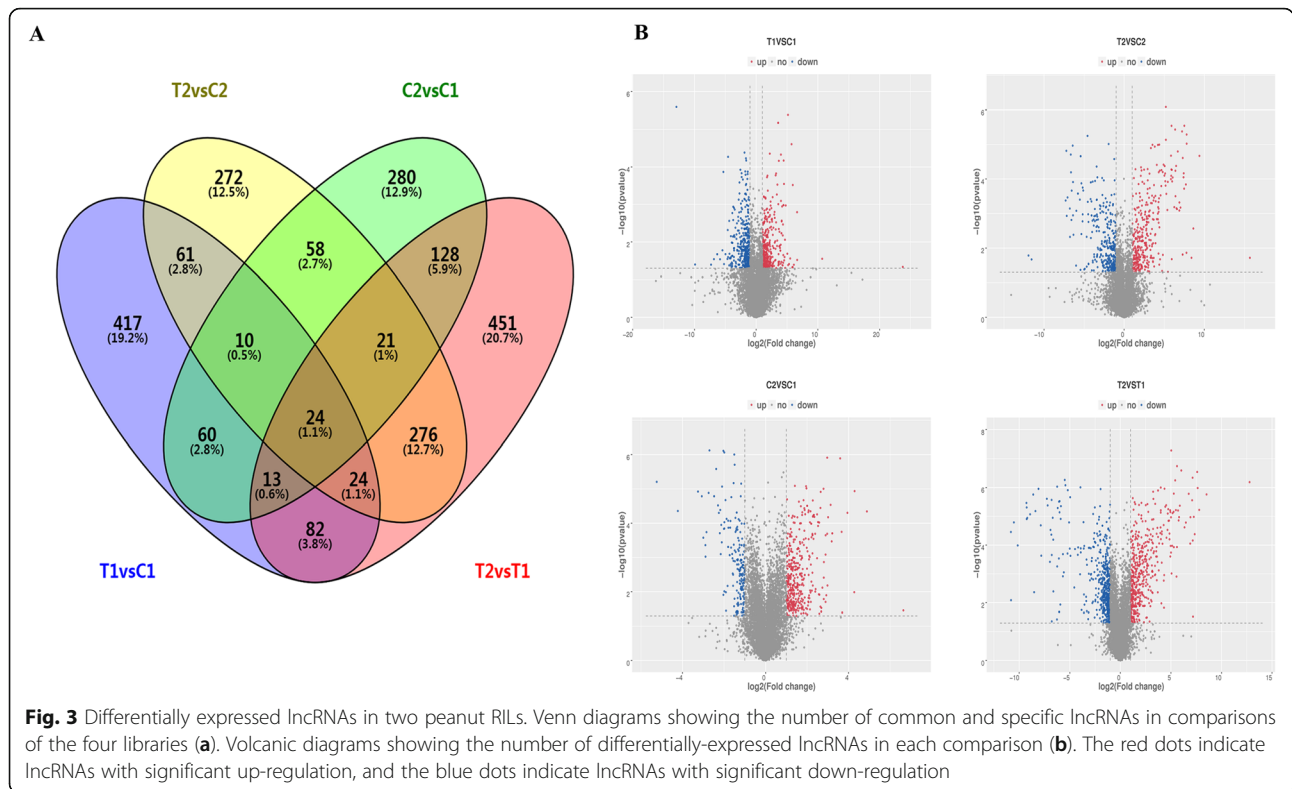
To confirm the data from the RNA-seq, we randomly selected 12 lncRNAs to verify their expression patterns by qPCR (Fig. 5). The qPCR results of these lncRNAs were consistent with those by high-throughput sequencing. For instance, the expression level of lncRNA XR\_001593099.1 was confirmed as up-regulated in both peanut RILs, while lncRNA XR\_001590629.1 was down-regulated in both RILs. However, lncRNAs XR\_001593511.1, MSTRG.18462.1, MSTRG.34915.1, and



**Fig. 2** The expression level of lncRNAs along the twenty peanut chromosomes (a); comprising twelve concentric rings, and each corresponding to a different sample including C1(RIL8106-15DAF), C2(RIL8107-15DAF), T1(RIL8106-35DAF), and T2 (RIL8107-35DAF) from outer to inner, respectively. Distribution of different types of lncRNAs (b). The known, intronic, bidirectional, sense, intergenic, and antisense lncRNAs are represented by different concentric rings according to the loci of lncRNAs in the genome. Length distribution of lncRNAs (c). Expression level (d) and number (e) of lncRNAs and mRNAs in samples

MSTRG.41848.1 showed up-regulated expression only in RIL8106 by two methods. Similarly, lncRNAs XR\_001614585.1, MSTRG.22884.1, MSTRG.12404.1, MSTRG.26719.1, and MSTRG.35761.1 displayed up-regulated only in RIL8107 by both methods. Moreover,

the result from the qPCR method also verified that lncRNA MSTRG.20033.1 was down-regulated predominantly in RIL8107 as showed by RNA-seq. These results indicate that RNA-seq analysis is highly reliable prediction of lncRNA expression patterns.



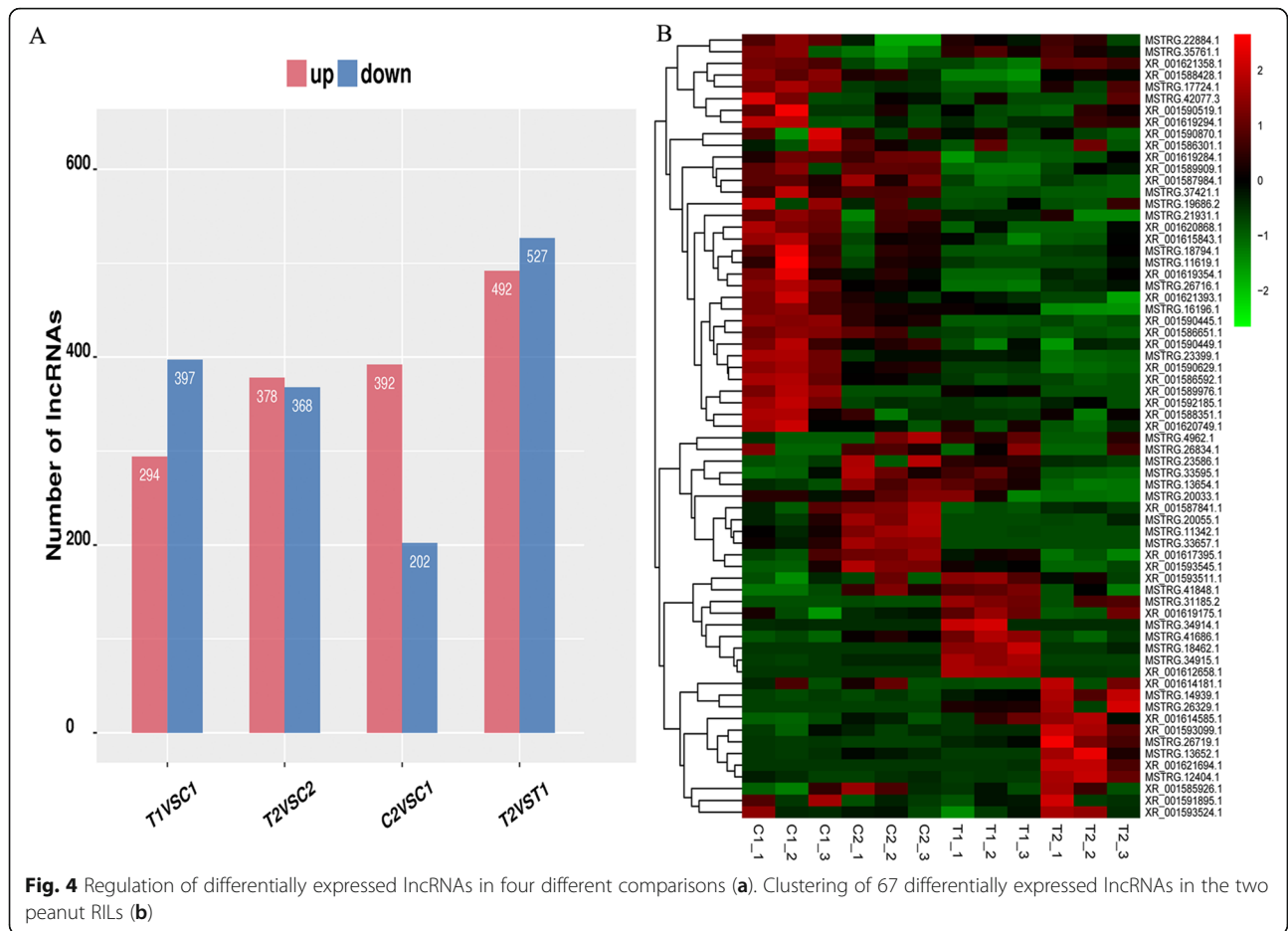
### Function of differentially expressed lncRNAs during peanut seed development

lncRNAs are proved to be proximal to their target genes preferentially [21–23]. In order to explore the potential functions of lncRNAs, we defined co-expressed protein-coding genes that located within 100 kb from each corresponding regulated lncRNAs as predictable targets. Differentially expressed lncRNAs were identified between two peanut RILs at the 15 DAF and 35 DAF stages, respectively, in order to dissect the lncRNAs at the seed development stage (Additional file 5: Table S5). Gene ontology (GO) analysis was performed to categorize these protein-coding genes. There were 25 classes of biological processes and these protein-coding genes were mainly enriched in “regulation of transcription”, “translation”, and “transport”. Moreover, some important growth and development-related genes were identified as lncRNAs targets, including “cell division”, “carbohydrate metabolic process”, “cell cycle”, and so on (Fig. 6a). These findings suggested that these differentially expressed lncRNAs might be involved in seed development of peanut by regulating expression of related-protein-coding genes.

To further understand the function of differentially expressed lncRNAs in seed development, putative interactive networks were constructed to disclose the relationship between lncRNAs and protein-coding genes (Fig. 6b). Among these, five protein-coding genes

involved in material transportation, including protein transportation (XM\_016309191.1), iron transportation (XM\_016101863.1), sugar transportation (XM\_016101862.1), auxin transportation (XM\_016102040.1), and translation process (XM\_016102046.1), were found to be regulated by sixteen lncRNAs in the seed development (Fig. 6B-1 and 6B-2). Three transcription factors, B3 domain-containing transcription factor (XM\_016091512.1), squamosa promoter-binding-like protein (XM\_016091385.1), and zinc-finger homeodomain protein (XM\_016335443.1), were embraced in the network of Fig. 6B-3, which involved in activating growth-related genes in the downstream during seed development. In addition, several important biological processes, such as GO:0009734, auxin mediated signaling pathway; GO:0009740, gibberellic acid mediated signaling pathway; GO:0051301, cell division, also were found in this study (Additional file 5: Table S5). These results implied that the differentially expressed lncRNAs might regulate genes involved in various biological processes, including transcription, transport, hormone signal transduction, and translation controlling seed development in peanut.

Moreover, based on the KEGG analysis, these protein-coding genes were significantly enriched in 20 pathways in the two peanut RILs, respectively (Fig. 7a and b). Among these pathways, there are three most over-represented categories, including ribosome, purine metabolism, and oxidative phosphorylation, were

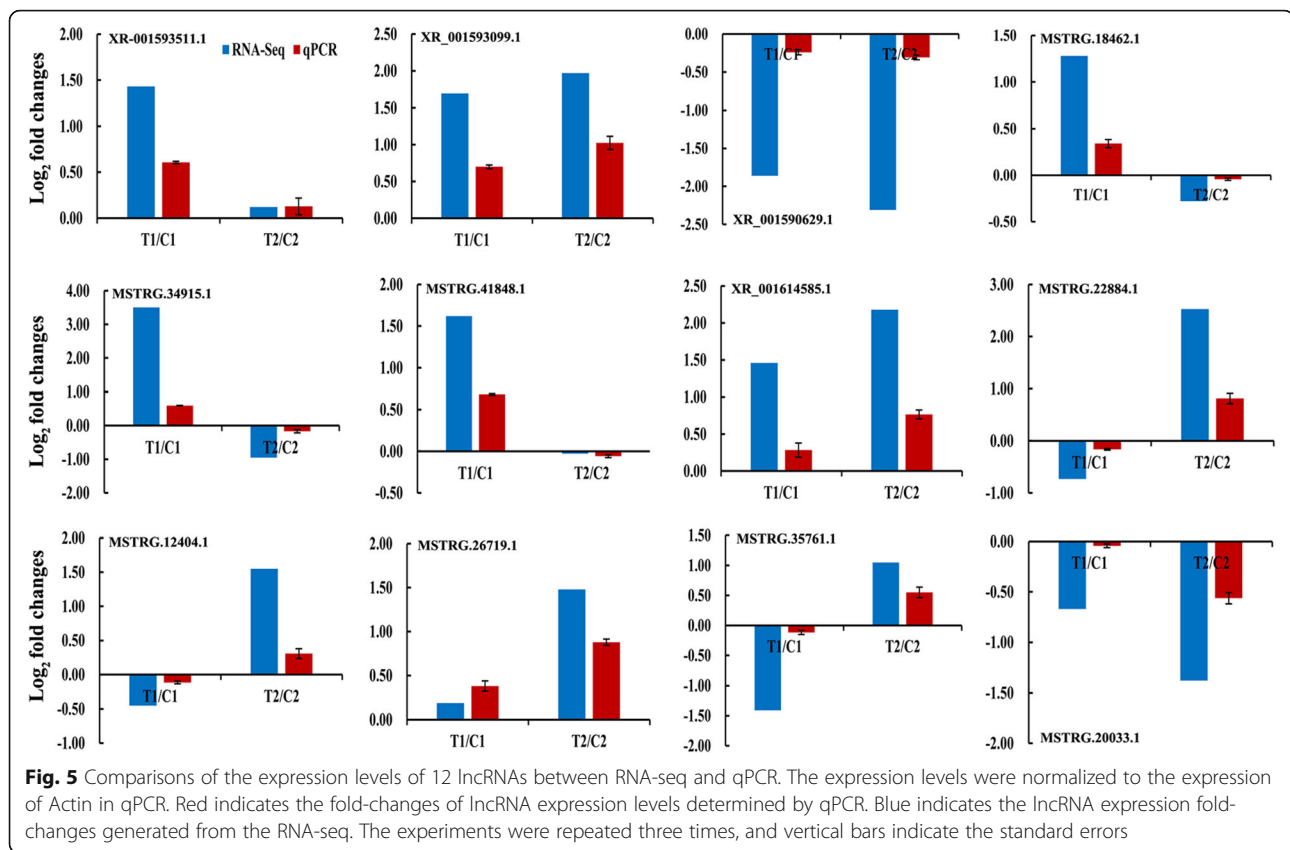


significantly enriched in the medium size seed line RIL 8106. However, the three most frequent pathways are glycolysis/gluconeogenesis, flavonoid biosynthesis, and galactose metabolism in the large size seed line RIL 8107. Interestingly, some of lncRNAs target genes were found to be involved in the zeatin synthesis only in RIL8107 (Fig. 7b). In this pathway, MSTRG.13500.1, MSTRG.13501.1, and MSTRG.13501.2 were involved in zeatin biosynthesis, and the accumulation level of the CYP735A protein (Additional file 5: Table S5). These findings suggested that lncRNAs have a significant effect on the regulation of peanut seed development by effecting endogenous hormone accumulation level.

**Validation of lncRNA target genes by qPCR**

To confirm the relationship of lncRNAs and their related protein-coding genes, the expression patterns of 12 putative cis-acting proteins as targets for 12 differentially expressed lncRNAs were verified by qPCR. Most lncRNAs and their putative cis-acting targets were co-expressed and up- or down-regulated at the 35 DAF stage (Fig. 5 and Fig. 8). The expression of lncRNA XR\_001593099.1 and the target for

embryonic protein DC-8-like (XM\_016114848.1) were up-regulated in both peanut RILs. However, the expression of lncRNA MSTRG.18462.1 and the target for MADS-box transcription factor 23-like (XM\_016087708.1), MSTRG.34915.1 and the target for protein transport protein sec31-like (XM\_016309191.1), MSTRG.41848.1 and the target for B3 domain-containing transcription factor VRN1-like (XM\_016324297.1) were up-regulated at the 35 DAF only in the RIL 8106. Meanwhile, the expressions of MSTRG.22884.1 and E3 ubiquitin-protein ligase UPL4 (XM\_016327810.1), MSTRG.12404.1 and amino acid permease 6-like (XM\_016116309.1), MSTRG.26719.1 and zinc-finger homeodomain protein 8-like (XM\_016335443.1), MSTRG.35761.1 and EPIDERMAL PATTERNING FACTOR-like protein 2 (XM\_016310265.1) were up-regulated at the 35 DAF only in the RIL 8107. Moreover, lncRNA MSTRG.20033.1 and squamosa promoter-binding-like protein 14 (XM\_016091385.1) was down-regulated in the RIL 8107. These results indicate that lncRNAs are involved in regulating peanut seed development by modulating the expression of their cis-acting target genes.



## Discussion

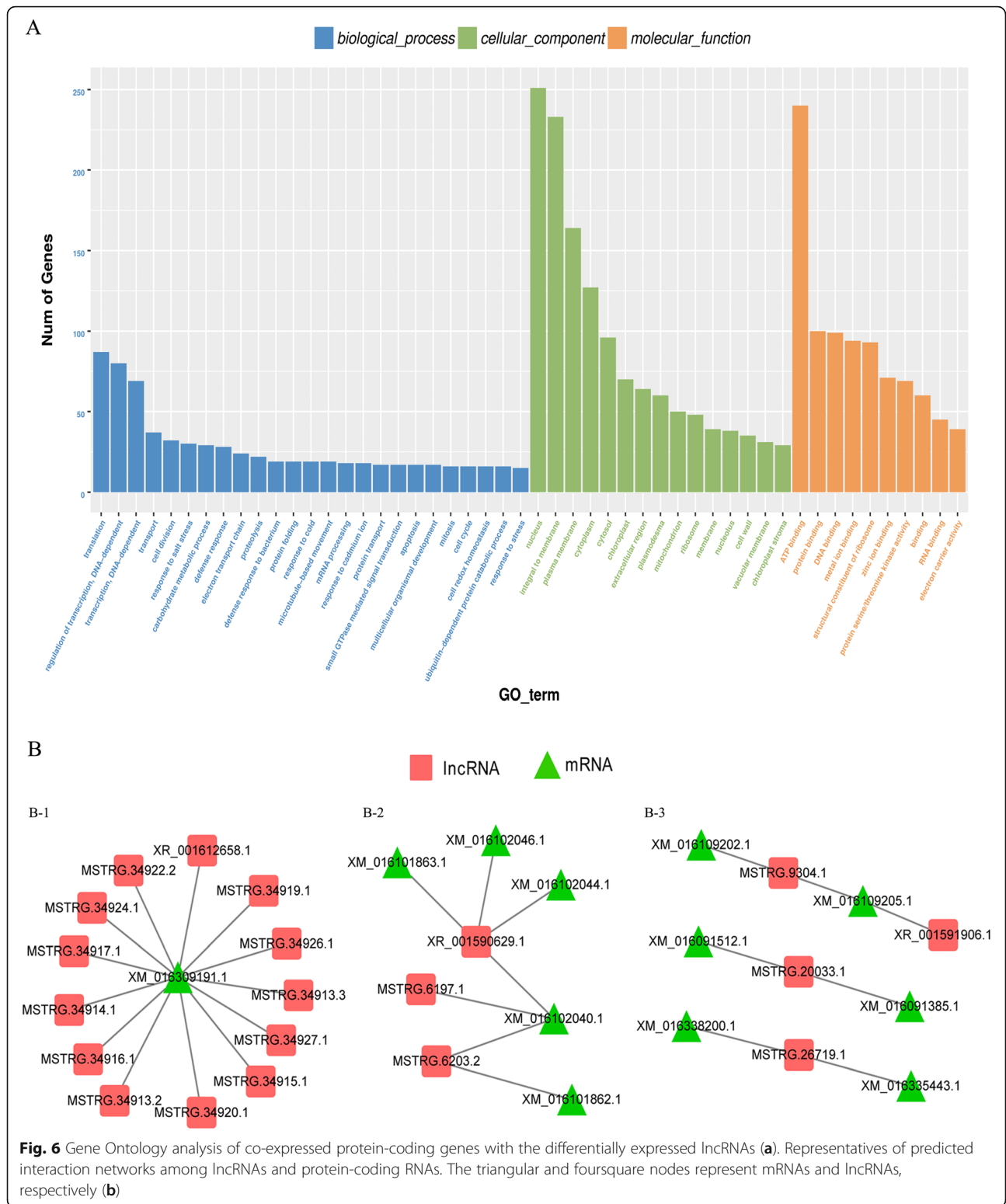
### Identification of lncRNAs and verification of their functions during peanut seed development

The regulatory roles of lncRNAs are increasingly being unraveled in plants. lncRNAs have been found to be involved in plant growth, development, reproduction and abiotic stress responses in plants such as maize, rice, and Arabidopsis [24–27]. In peanut, previous studies on non-coding RNA have predominantly focused on miRNA identification and functional analysis [17, 28, 29], however, the role of lncRNAs, especially in association with peanut seed development has not been reported yet. Peanut seed development is a complex biological process regulated by coordinated gene expression. According to previous description about the whole stages of seed development in peanut [17], the first sign of pod development is seen at 15 DAF, and the pods enlarge to the maximum size at about 35 DAF, which is named the stereotyped fruit. Hence, we selected seeds at these two stages from two peanut RILs to perform a genome-wide analysis of lncRNAs using a high-throughput sequencing technology. We identified a total of 13,425 unique lncRNAs in different development stages between two peanut RILs. Of these lncRNAs, 4037 novel lncRNAs, including MSTRG.37421.1, MSTRG.18462.1, MSTRG.31185.2, MSTRG.34915.1, MSTRG.41848.1, MSTRG.22884.1,

MSTRG.12404.1, MSTRG.13652.1, MSTRG.13500.1, and MSTRG.13501.1, etc., were first reported in peanut seeds, indicating their roles in seed development. Moreover, we identified 1437 differentially expressed lncRNAs at the 35 DAF stage compared to 15 DAF between two RILs. To better understand the function of differentially expressed lncRNAs in peanut seed development, we analyzed differentially expressed protein-coding genes that were co-expressed with these lncRNAs (Additional file 5: Table S5). Some lncRNAs co-expressed with some protein-coding genes appeared at 35 DAF stage between the two RILs regardless of seed size. For example, lncRNA XR\_001593099.1 was up-regulated in both peanut RILs and the expression of the target gene of lncRNA XR\_001593099.1 was also increased. The target gene, embryonic protein DC-8-like (*DC8*), has been shown to be strongly expressed during embryogenesis and in cell walls of endosperm tissues in plant [30, 31], indicating that *DC8* plays a role in peanut seed development. Our work provides an important resource of peanut lncRNAs that can be useful to the research community.

### Specific lncRNAs involved in seed development in the medium size seed line RIL 8106

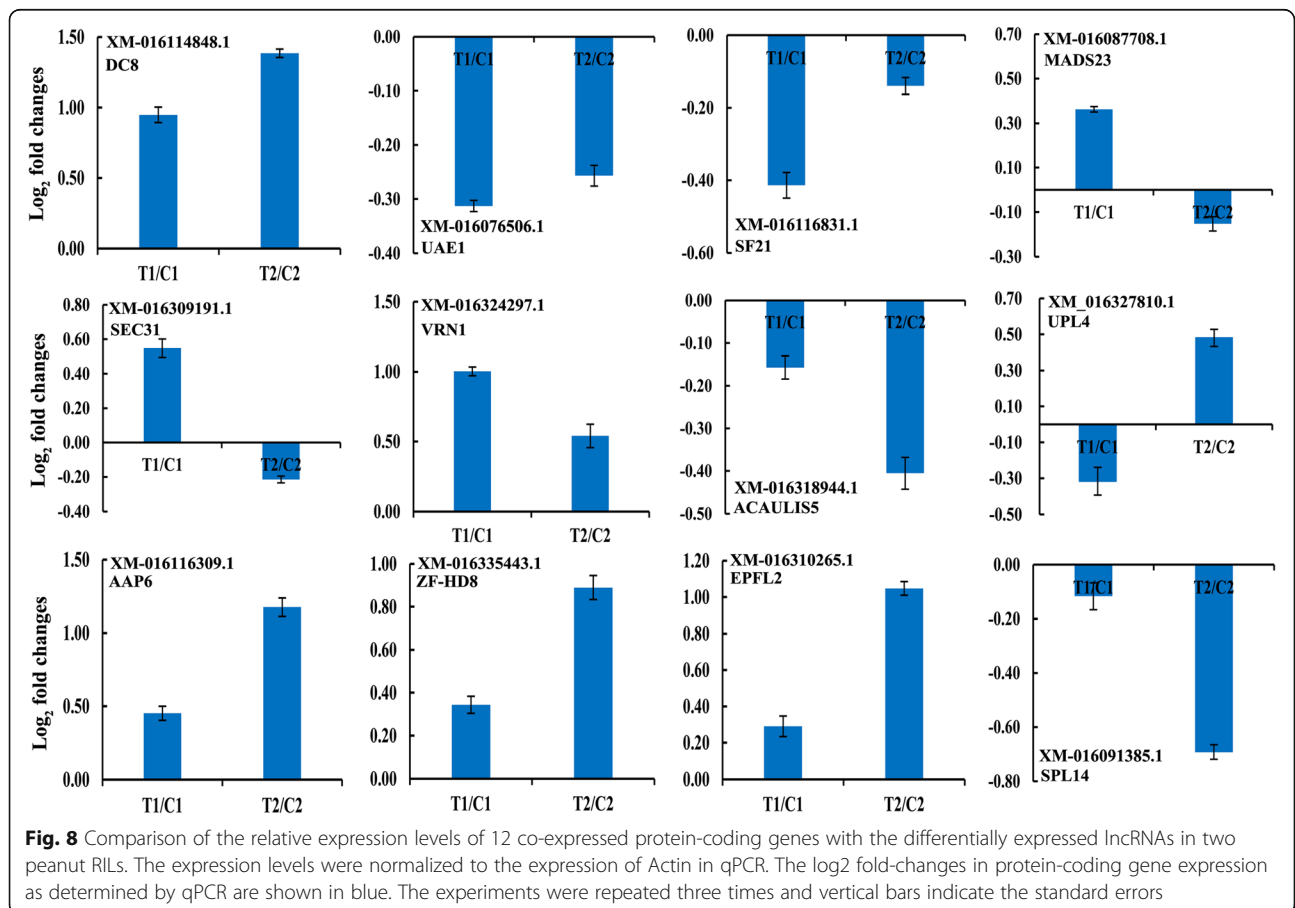
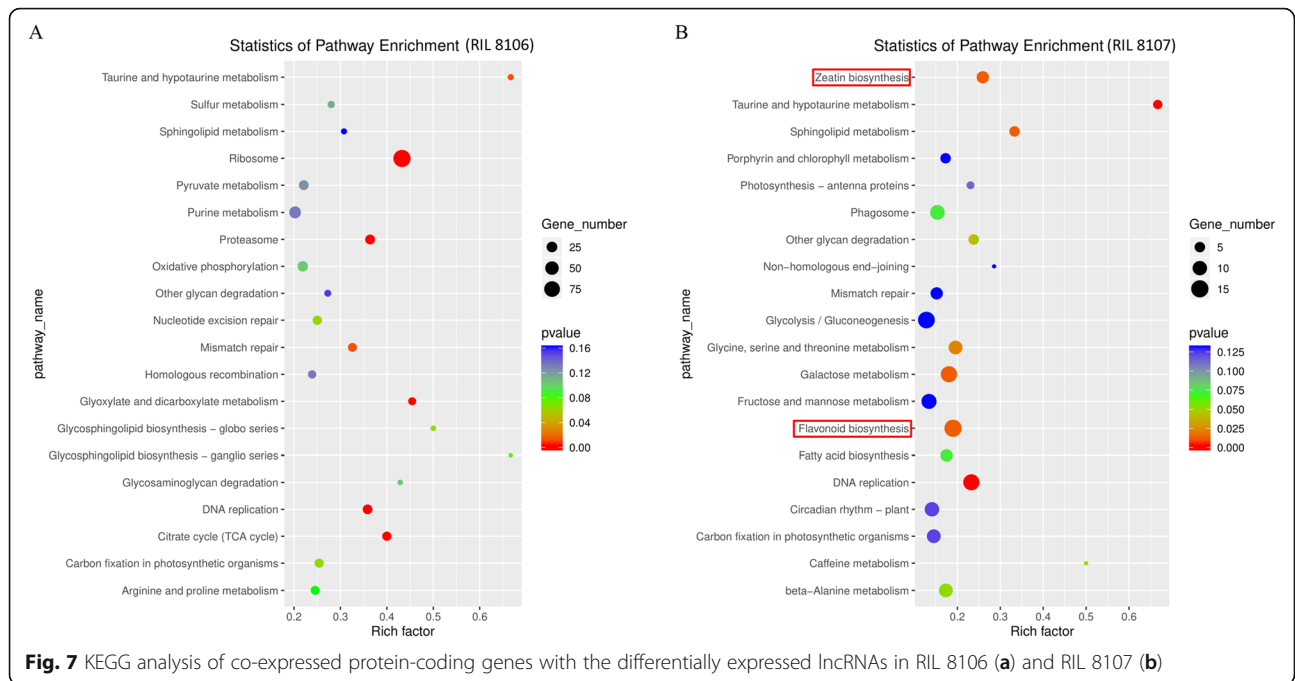
In this study, many lncRNAs and their putative cis-acting target genes were co-expressed predominantly only in the RIL8106 at the 35 DAF stage. For example,



three lncRNAs, MSTRG.18462.1, MSTRG.34915.1, and MSTRG.41848.1, were up-regulated only in the RIL 8106 at 35 DAF (Fig. 5). The target gene of lncRNA MSTRG.18462.1 encodes the MADS-box transcription

factor 23-like (*MADS*). Previous study showed that *MADS* family members are key elements of the genetic networks that control flowering and fruit development [32]. TAGL1 is a transcription factor belonging to the





family of the MADS-box and it has been shown to be involved in many biological processes of the fruit ripening in tomato [33]. In this work, the level of expression of the *MADS* gene was increased as the lncRNA MSTRG.18462.1 performed, which may regulate the expression of *MADS* for the process of seed development. Another up-regulated lncRNA MSTRG.34915.1 was expressed upstream of the coding region of the transport protein sec31-like (*SEC31*) at the 35 DAF stage. In Arabidopsis, *SEC* gene is proved to be essential for gametophyte development, and is required for secretory trafficking in developing pollen [34]. These results suggested that lncRNA MSTRG.34915.1 may regulate the expression of *SEC31*, which could then contribute to seed development in peanut. Similar to MSTRG.34915.1, the lncRNA MSTRG.41848.1 was located approximately 10 kb upstream of the coding region of B3 domain-containing transcription factor VRN1-like (*VRN1*) (Additional file 5: Table S5). Both of them were up-regulated at 35 DAF in RIL8106. Previous studies have demonstrated that B3 domain-containing proteins are involved in seed development, hormone response, and flowering time [35, 36]. Therefore, the lncRNA MSTRG.41848.1 could possibly be involved in regulating seed development in peanut by modulating the expression of *VRN1*. The high expression of these lncRNAs and their putative cis-acting target genes was an assumed mechanism for a specific seed development in RIL 8106.

#### Specific lncRNAs involved in seed development in the large size seed line RIL 8107

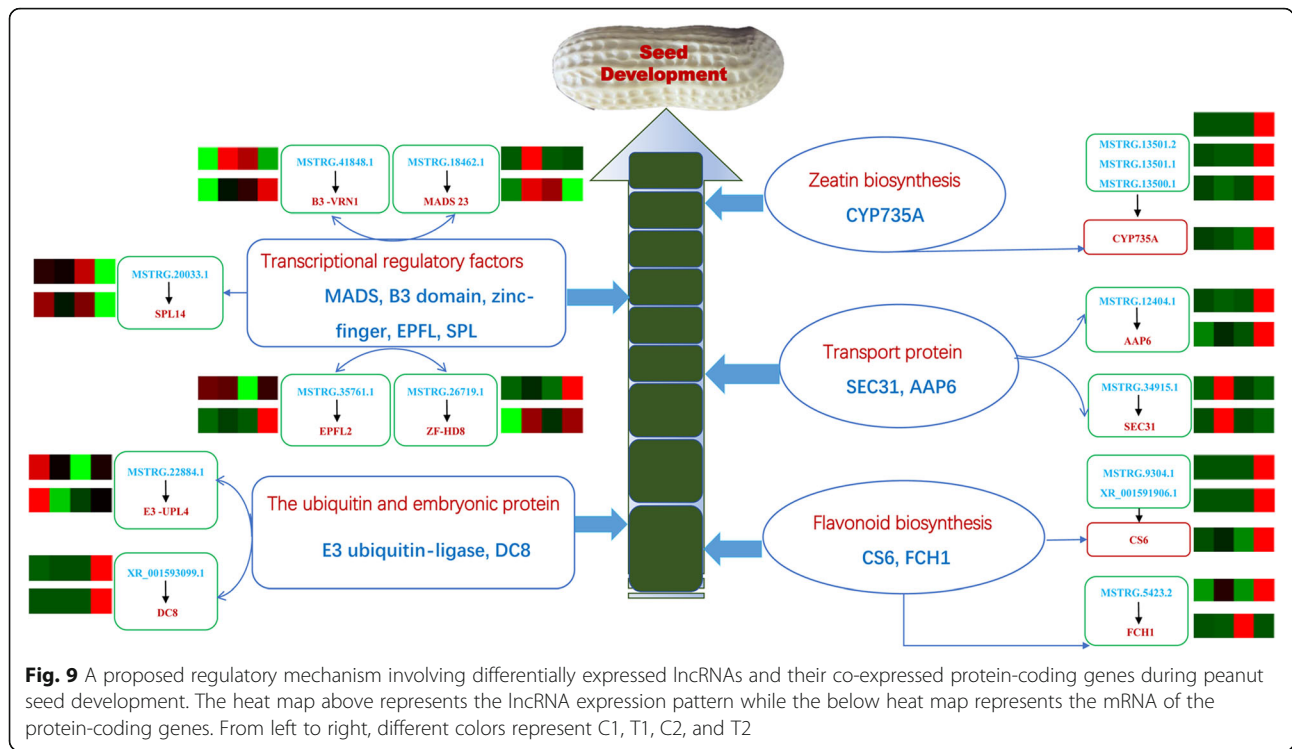
A number of lncRNAs and their putative protein-coding genes was up-regulated only in RIL8107, including lncRNAs MSTRG.22884.1, MSTRG.12404.1, MSTRG.26719.1, and MSTRG.35761.1 (Fig. 5 and Fig. 8). Transcription of lncRNA MSTRG.22884.1 was significantly increased in RIL 8107 as well as its target E3 ubiquitin-protein ligase (*UPL4*). E3 family members play an important role in regulating gametogenesis and cell cycle processes during seed development in Arabidopsis [37–39]. The increased expression of the *UPL4* gene could be related with its role in peanut seed development. Another up-regulated lncRNA MSTRG.12404.1 targets amino acid permease 6-like (*AAP6*) gene, which encodes plant-specific amino acid transmembrane transporter and involved in the amino acid uptake [40]. In the present study, the expression of *AAP6* gene was up-regulated at 35 DAF, indicating that *AAP6* may also be involved in peanut seed development. The target gene of lncRNA MSTRG.26719.1 is the zinc-finger homeodomain protein 8-like (*ZF-HD8*). Previous studies have shown that *ZF-HD* genes encode a group of transcriptional regulators in floral and leaf development [41, 42]. We identified the lncRNA MSTRG.26719.1 located upstream of the coding sequence of *ZF-HD8* (Additional file 5: Table S5). The results suggested

that MSTRG.26719.1 may be involved in regulating seed development by regulating the expression of *ZF-HD8*. Similarly, the target gene of another up-regulated lncRNA MSTRG.35761.1 is EPIDERMAL PATTERNING FACTOR-like protein 2 (*EPFL2*), which encode plant-specific transcription factors involved in cell differentiation. In rice, *RAE2* (an *EPFL* gene) can promote the proliferation of vasculature cells for awn elongation [43]. A similar phenomenon has been reported in Arabidopsis that *AtEPFL4* and *AtEPFL6* coordinate development of inflorescence architecture [44]. In our present work, we found that the level of *EPFL2*-specific mRNA was increased at 35 DAF stage (Fig. 5 and Fig. 8). The increased expression of *EPFL2* may also play a role in peanut seed development by regulating cell differentiation.

In contrast, the lncRNA MSTRG.20033.1 as well as its predictable target gene, squamosa promoter-binding-like protein 14 (*SPL*), were significantly down-regulated and predominantly occurred only in the RIL8107 (Fig. 5 and Fig. 8). The *SPL* plays important roles in plant developmental phase transition, flower and fruit development, gibberellins signaling etc. [45–47]. In our present study, the *SPL* gene was predicted to be regulated by the lncRNA MSTRG.20033.1, and also participated in peanut seed development simply prolonging the developmental phase change. Our findings suggest that specific regulation of these lncRNAs and their putative protein-coding genes only in RIL 8107 might explain the difference of seed size between RIL 8107 and RIL 8106.

#### lncRNA-dependent regulatory networks might involve in seed development of peanut

Plant endogenous hormones play vital roles in various developmental processes. Auxin regulates a vast array of growth and developmental processes in plants [48]. Cytokinins are considered to be the main hormone participating in many biological processes of the plant, such as apical dominance [49], root proliferation [50], and reproductive development [51]. Previous study has shown that *CYP735A* gene plays an important role in trans-zeatin (an endogenous cytokinin) biosynthesis in *Jatropha curcas* [52]. The *Jccyp735a* mutants using the CRISPR-Cas9 system has been found to have a significant decrease of the concentrations of trans-zeatin and trans-zeatin-riboside displaying severe retarded growth. In this study, we found that several lncRNAs, including MSTRG.13500.1, MSTRG.13501.1, and MSTRG.13501.2, participate in zeatin biosynthesis by regulating the expression of *CYP735A* only in RIL8107 (Fig. 7b, Fig. 9, and Additional file 5: Table S5). These findings suggest that these lncRNAs may play key roles in peanut seed development by affecting endogenous cytokinin biosynthesis. In addition, flavonoids play an important role in plant development. In Arabidopsis, the accumulation of flavonoids including flavonols, anthocyanins and proanthocyanidins in specific tissue leads



to the regulation of biological processes including light-attenuation, oxidative stress protection and more importantly seed coat development [53]. In our present work, we found that lncRNAs MSTRG.9304.1 and XR\_001591906.1 participate in flavonoid biosynthesis by regulating the expression of chalcone synthase 6 like (CS6) only in RIL8107 (Fig. 7b, Fig. 9, and Additional file 5: Table S5), indicating their potential roles in peanut seed development.

The functions of their putative protein-coding genes in peanut seed development were assessed in addition to the aforementioned lncRNAs (Fig. 9). For example, lncRNAs MSTRG.18462.1 and target *MADS23*; MSTRG.41848.1 and target *B3-VRNI*; MSTRG.22884.1 and target *E3-UPL4*; MSTRG.26719.1 and target *ZF-HD8*; MSTRG.35761.1 and target *EPFL2*; MSTRG.20033.1 and target *SPL14* were also identified to be involved in seed development. The identification of these growth and development-related lncRNAs, and the characterization of their regulatory networks can enhance the understanding of the molecular mechanisms involved in peanut and plant seed development in general.

**Conclusions**

Peanut seed development is a complex process that involves a series of physiological, morphological, and transcriptional changes. This study represents the first report characterizing the expression landscape of lncRNAs involved in peanut seed development process. A total of 9388 known and 4037 novel lncRNAs were discovered. Among them, 1437 lncRNAs were differentially expressed during

seed development between the two RILs. Functional analysis of the lncRNAs and their close related protein-coding genes revealed at least 11 regulatory modules of these lncRNA-mRNAs playing important roles in peanut seed development. Our results provide new insights into lncRNAs and their possible functions in peanut, as well as their expression pattern in the context of seed development; therefore, may provide new approaches for the genetic improvement of yield-related traits in peanut.

**Methods**

**Plant materials**

We obtained two peanut cultivars, Huayou 7 and Huayou 4, and manually crossed them to generate two eighth-generation recombinant inbred lines (RILs), RIL8106 and RIL8107. These two RILs are both erect Virginia-types with 8–10 branches, but their seed sizes displayed a main difference. RIL 8106 has smaller seeds compared with RIL8107, which was supported by a corresponding 100-seed weight of 100 g of RIL8106 and a corresponding 100-seed weight of 182 g of RIL8107. Fresh seeds were harvested from the two RILs at 15 DAF and 35 DAF, respectively, and these plants were designated as C1, C2, T1 and T2. Four seed samples were randomly chosen and subjected to liquid nitrogen before storing them at -80 °C. For each treatment, three replicates of samples were collected.

### Construction of cDNA libraries and deep sequencing

Total RNA was extracted from each peanut seed sample using Trizol reagent (Invitrogen, CA, USA) following the manufacturer's procedure. According to the protocol provided in the Epicentre Ribo-Zero Gold Kit (Illumina, San Diego, USA), ribosomal RNA was depleted from approximately 10 µg of total RNA representing a specific adipose type. Following purification, the poly(A)- or poly(A)+ RNA fractions were cleaved into small fragments and reverse-transcribed to construct the final cDNA library based on the protocol described in the RNA-Seq sample preparation kit (Illumina, San Diego, USA). We then performed the paired-end sequencing (150 bp) on an Illumina Hiseq4000 sequencer at the LC Biotech (Hangzhou, China) following the vendor's recommended protocol.

### Reads mapping and transcriptome assembly

Firstly, the low quality reads (including adaptor contamination, low quality bases, and undetermined bases) were removed using Cutadapt [54]. Then sequence quality was verified using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). We then used TopHat [55] to obtain clean paired-end reads by mapping them to reference genomes of two diploid peanut species, *A. duranensis* and *A. ipaensis*, acquired from the peanut-base database (<https://www.peanutbase.org/>). To construct transcriptome, the mapped reads were assembled using StringTie [56]. After the final transcriptome was generated, StringTie and Ballgown [57] were used to estimate the expression levels of all transcripts.

### LncRNA identification

According to the characteristics of lncRNA, we adopted the following steps to identify lncRNAs from the transcripts of transcriptome assemblies [58]. First, all the transcripts that overlapped with reference genome were defined as "known lncRNA". Second, among the remaining transcripts, those with length longer than 200 bp were selected for the protein-coding-score test to calculating the Coding Potential Calculator (CPC) [18] and Coding-Non-Coding Index (CNCI) [19]. Finally, the transcripts with CPC score < -1 and CNCI score < 0 were defined as novel lncRNAs.

### Analysis of differentially expressed patterns

StringTie [56] was used to perform expression level of all transcripts, including mRNAs and putative lncRNAs by calculating FPKM [59]. Differentially expression analysis was performed using R package – Ballgown [57] with  $|\log_2(\text{fold change})| > 1$  and  $p$  value < 0.05.

### Target gene functional analysis of lncRNAs

To explore the function of lncRNAs in peanut seed development, we predicted the cis-target genes of lncRNAs. The lncRNAs may play a cis role acting on neighboring target genes [60]. In this study, coding genes in 100,000 bp up- and downstream from the lncRNA, were selected by python script [61]. Moreover, we showed functional analysis of the target genes for lncRNAs by using the BLAST2GO [62] and significance was expressed as a  $p$  value < 0.05.

### Quantitative real-time PCR

We determined the transcript levels of selected lncRNAs and protein-coding genes via RT-qPCR (Additional file 6: Table S6) according to the manufacturer's instructions for the CFX96 Real-Time System (Bio-Rad, Hercules, CA, USA) and the TB Green Premix Ex Taq II (TaKaRa, Dalian, China). Total RNA were extracted from seeds at different developmental stages using RNAsimple Total RNA Kit (TIANGEN biotech, Beijing, China), and reverse-transcribed using the PrimeScript RT reagent Kit (TaKaRa, Dalian, China). The reverse transcription reactions were performed according to the supplier's protocol. The peanut gene Actin 7 was used as the internal control for RT-qPCR. The reactions were conducted at following conditions: 94 °C for 30 s, followed by 40 cycles of 94 °C for 5 s and 60 °C for 30 s. All reactions were performed with three replicates. The relative gene expression values were calculated by the  $2^{-\Delta\Delta CT}$  method [63].

### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12870-020-02405-4>.

**Additional file 1: Table S1.** Mean sample quality.

**Additional file 2: Table S2.** Statistical data of the RNA-Seq reads for four samples.

**Additional file 3: Table S3.** All lncRNAs in peanut.

**Additional file 4: Table S4.** Differentially expressed known and novel lncRNAs four comparisons. (S4-1) Differentially expressed lncRNAs in RIL8106 at 35 DAF (T1 vs C1). (S4-2) Differentially expressed lncRNAs in RIL8107 at 35 DAF (T2 vs C2). (S4-3) Differentially expressed lncRNAs in C2 vs C1. (S4-4) Differentially expressed lncRNAs in T2 vs T1.

**Additional file 5: Table S5.** Target genes prediction of differentially expressed lncRNAs. (S5-1) Target genes prediction of differentially expressed lncRNAs in T1 vs C1. (S5-2) Target genes prediction of differentially expressed lncRNAs in T2 vs C2.

**Additional file 6: Table S6.** All primers in this study.

### Abbreviations

lncRNAs: Long noncoding RNAs; RIL: Recombinant inbred lines; DAF: Days after flowering; DELs: Differentially expressed lncRNAs; FPKM: Fragments per kilobase of exon model per million fragments mapped reads; GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; CPC: Coding potential calculator; CNCI: Coding-non-coding index

### Acknowledgements

We are grateful to the anonymous reviewers for their helpful suggestions about the manuscript.

### Authors' contributions

DY designed and conceived the research. XM and GH wrote the manuscript. XM, XZ, and ST analyzed the data. LN, KL, KZ and ZL performed the qPCR experiments. ZX, GH, ST, and DY edited the manuscript. The authors read and approved the manuscript.

### Funding

This work was supported by grants from the Key Program of National Natural Science Foundation of China (NSFC)-Henan United Fund (No. U1704232), Key Scientific and Technological Project of Henan Province (No. 161100111000; S2012-05-G03), Innovation Scientists and Technicians Troop Construction Projects of Henan Province (No. 2018JR0001). These funding bodies have had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

### Availability of data and materials

All data pertaining to the present study has been included in the Figures/ supplementary files of the manuscript. The two peanut recombinant inbred lines analyzed in this article (RIL 8106 and RIL 8107), are saved at Henan Agricultural University (Zhengzhou, China), and are available from the corresponding author on reasonable request. The RNA-seq datasets are available in the Sequence Read Archive (SRA) database (Accession ID: PRJNA627334; <https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA627334>).

### Ethics approval and consent to participate

Collection of plant material in this study complies with institutional, national, or international guidelines.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China. <sup>2</sup>College of Agriculture, Environment and Nutrition Sciences, Tuskegee University, Tuskegee 36088, AL, USA.

Received: 8 October 2019 Accepted: 27 April 2020

Published online: 06 May 2020

### References

- Bai Y, Dai X, Harrison AP, Chen M. RNA regulatory networks in animals and plants: a long noncoding RNA perspective. *Brief Funct Genomics*. 2015;14(2):91–101.
- Zhou ZY, Li AM, Adeola AC, Liu YH, Irwin DM, Xie HB, et al. Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs. *Genome Biol Evol*. 2014;6(6):1387–92.
- Dogini DB, Pascoal VD, Avansini SH, Vieira AS, Pereira TC. Lopes-Cendes I. the new world of RNAs. *Genet Mol Biol*. 2014;37:285–93.
- Wang H, Niu QW, Wu HW, Liu J, Ye J, Yu N, et al. Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. *Plant J*. 2015;84:404–16.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25(18):1915–27.
- Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. *RNA Biol*. 2013;10(6):924–33.
- Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol*. 2013;11:59.
- Liu J, Wang H, Chua N-H. Long noncoding RNA transcriptome of plants. *Plant Biotechnol J*. 2015;13(3):319–28.
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007;129(7):1311–23.
- Valencia P, Dias AP, Reed R. Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proc Natl Acad Sci U S A*. 2008;105(9):3386–91.
- Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*. 2011;331(6013):76–9.
- Swiezewski S, Liu FQ, Magusin A, Dean C. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature*. 2009;462(7274):799–802.
- Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, Yao J, et al. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc Natl Acad Sci U S A*. 2012;109(7):2654–9.
- Zhang L, Wang M, Li N, Wang H, Qiu P, Pei L, et al. Long noncoding RNAs involve in resistance to Verticillium dahliae, a fungal disease in cotton. *Plant Biotechnol J*. 2018;16(6):1172–85.
- Campalans A, Kondorosi A, Crespi M. Eno40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell*. 2004;16:1047–59.
- Zhang J, Liang S, Duan J, Wang J, Chen S, Cheng Z, et al. De novo assembly and characterization of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). *BMC Genomics*. 2012;13:90.
- Ma X, Zhang X, Zhao K, Li F, Li K, Ning L, et al. Small RNA and degradome deep sequencing reveals the roles of microRNAs in seed expansion in peanut (*Arachis hypogaea* L.). *Front Plant Sci*. 2018;9:349.
- Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35(suppl 2):W345–9.
- Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013;41(17):e116.
- Deng F, Zhang X, Wang W, Yuan R, Shen F. Identification of *Gossypium hirsutum* long non-coding RNAs (lncRNAs) under salt stress. *BMC Plant Biol*. 2018;18:23.
- Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*. 2008;451(7175):202–6.
- Hanano S, Goto K. Arabidopsis TERMINAL FLOWER1 is involved in the regulation of flowering time and inflorescence development through transcriptional repression. *Plant Cell*. 2011;23(9):3172–84.
- Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res*. 2012;22(3):577–91.
- Zhang YC, Chen YQ. Long noncoding RNAs: new regulators in plant development. *Biochem Biophys Res Commun*. 2013;436(2):111–4.
- Li L, Eichten SR, Shimizu R, Petsch K, Yeh C-T, Wu W, et al. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol*. 2014;15(2):R40.
- Zhang Y, Liao J, Li Z, Yu Y, Zhang J, Li Q, 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;15:512.
- Amor BB, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, et al. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Res*. 2009;19(1):57–69.
- Chi X, Yang Q, Chen X, Wang J, Pan L, Chen M, et al. Identification and characterization of microRNAs from peanut (*Arachis hypogaea* L.) by high-throughput sequencing. *PLoS One*. 2011;6(11):e27530.
- Gao C, Wang P, Zhao S, Zhao C, Xia H, Hou L, et al. Small RNA profiling and degradome analysis reveal regulation of microRNA in peanut embryogenesis and early pod development. *BMC Genomics*. 2017;18:220.
- Hatzopoulos P, Fong F, Sung ZR. Abscisic acid regulation of DC8, a carrot embryonic gene. *Plant Physiol*. 1990;94:690–5.
- Cheng J, Seeley KA, Goupil P, Sung ZR. Expression of DC8 is associated with, but not dependent on embryogenesis. *Plant Mol Biol*. 1996;31(1):127–41.
- Hu P, Li G, Zhao X, Zhao F, Li L, Zhou H. Transcriptome profiling by RNA-Seq reveals differentially expressed genes related to fruit development and ripening characteristics in strawberries (*Fragaria xananassa*). *PeerJ*. 2018;6:e4976.
- Garceau DC, Batson MK, Pan IL. Variations on a theme in fruit development: the PLE lineage of MADS-box genes in tomato (TAGL1) and other species. *Planta*. 2017;246(2):313–21.
- El-Kasbi F, Pacher T, Strompen G, Stierhof Y-D, Müller LM, Koncz C, et al. Arabidopsis SNARE protein SEC22 is essential for gametophyte

- development and maintenance of Golgi-stack integrity. *Plant J.* 2011;66(2): 268–79.
35. Swaminathan K, Peterson K, Jack T. The plant B3 superfamily. *Trends Plant Sci.* 2008;13:647–55.
  36. King GJ, Chanson AH, McCallum EJ, Ohme-Takagi M, Byriel K, Hill JM, et al. The Arabidopsis B3 domain protein VERNALIZATION1 (VRN1) is involved in processes essential for development, with structural and mutational studies revealing its DNA-binding surface. *J Biol Chem.* 2013;288(5):3198–207.
  37. Xia T, Li N, Dumenil J, Li J, Kamenski A, Bevan MW, et al. The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in Arabidopsis. *Plant Cell.* 2013;25:3347–59.
  38. Li N, Li Y. Signaling pathways of seed size control in plants. *Curr Opin Plant Biol.* 2016;33:23–32.
  39. Downes BP, Stupar RM, Gingerich DJ, Vierstra RD. The HECT ubiquitin-protein ligase (UPL) family in Arabidopsis: UPL3 has a specific role in trichome development. *Plant J.* 2003;35(6):729–42.
  40. Song P, Chen X, Wu B, Gao L, Zhi H, Cui X. Identification for soybean host factors interacting with P3N-PIPO protein of soybean mosaic virus. *Acta Physiol Plant.* 2016;38:131.
  41. Tan QK-G, Irish VF. The Arabidopsis zinc finger-homeodomain genes encode proteins with unique biochemical properties that are coordinately expressed during floral development. *Plant Physiol.* 2006;140:1095–108.
  42. Khatun K, Nath UK, Robin AFK, Park J-I, Lee D-J, Kim M-B, et al. Genome-wide analysis and expression profiling of zinc finger homeodomain (ZHD) family genes reveal likely roles in organ development and stress responses in tomato. *BMC Genomics.* 2017;18:695.
  43. Bessho-Uehara K, Wang DR, Furuta T, Minami A, Nagai K, Gamuyao R. Loss of function at RAE2, a previously unidentified EPFL, is required for awnlessness in cultivated Asian rice. *Proc Natl Acad Sci U S A.* 2016;113(32): 8969–74.
  44. Uchida N, Lee JS, Horst RJ, Lai H-H, Kajita R, Kakimoto T, et al. Regulation of inflorescence architecture by intertissue layer ligand-receptor communication between endodermis and phloem. *Proc Natl Acad Sci U S A.* 2012;109(16):6337–42.
  45. Usami T, Horiguchi G, Yano S, Tsukaya H. The more and smaller cells mutants of Arabidopsis thaliana identify novel roles for SQUAMOSA PROMOTER BINDING PROTEIN-LIKE genes in the control of heteroblasty. *Development.* 2009;136:955–64.
  46. Yu S, Galvão VC, Zhang Y, Horrer D, Zhang T, Hao Y, et al. Gibberellin regulates the Arabidopsis floral transition through miR156-targeted SQUAMOSA PROMOTER BINDING-LIKE transcription factors. *Plant Cell.* 2012; 24:3320–32.
  47. Wang Z, Wang Y, KohalmiLisa SE, Amyot L, Hannoufa A. SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 2 controls floral organ development and plant fertility by activating ASYMMETRIC LEAVES 2 in Arabidopsis thaliana. *Plant Mol Biol.* 2016;92(6):661–74.
  48. Salehin M, Bagchi R, Estelle M. SCFTIR1/AFB-based auxin perception: mechanism and role in plant growth and development. *Plant Cell.* 2015;27: 9–19.
  49. Shimizu-Sato S, Tanaka M, Mori H. Auxin–cytokinin interactions in the control of shoot branching. *Plant Mol Biol.* 2009;69(4):429–35.
  50. Kudo T, Kiba T, Sakakibara H. Metabolism and long-distance translocation of cytokinins. *J Integr Plant Biol.* 2010;52(1):53–60.
  51. Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, et al. Cytokinin oxidase regulates rice grain production. *Science.* 2005;309(5735): 741–5.
  52. Cai L, Zhang L, Fu Q, Xu Z. Identification and expression analysis of cytokinin metabolic genes IPTs, CYP735A and CKXs in the biofuel plant *Jatropha curcas*. *Peer J.* 2018;6:e4812.
  53. Kleindt CK, Stracke R, Mehrtens F, Weisshaar B. Expression analysis of flavonoid biosynthesis genes during Arabidopsis thaliana silique and seed development with a primary focus on the proanthocyanidin biosynthetic pathway. *BMC Res Notes.* 2010;3:255.
  54. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal.* 2011;17(1):10.
  55. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25(9):1105–11.
  56. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290–5.
  57. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol.* 2015;33(3):243–6.
  58. Kapusta A, Feschotte C. Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends Genet.* 2014;30: 439–52.
  59. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010;28(5):511–5.
  60. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet.* 2009;10(3):155–9.
  61. Liao Q, Liu CN, Yuan XY, Kang SL, Miao RY, Xiao H, et al. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res.* 2011;39(9):3864–78.
  62. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21(18):3674–6.
  63. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods.* 2001;25:402–8.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

