Global identification of *Arabidopsis* IncRNAs reveals the regulation of *MAF4* by a natural antisense RNA

Xinyue Zhao\(^1,2\), Jingrui Li\(^1,2\), Bi Lian\(^1,2\), Hanqing Gu\(^1,2\), Yan Li\(^1\) & Yijun Qi\(^1,2\)

Long non-coding RNAs (IncRNAs) have emerged as important regulators of gene expression and plant development. Here, we identified 6,510 IncRNAs in *Arabidopsis* under normal or stress conditions. We found that the expression of natural antisense transcripts (NATs) that are transcribed in the opposite direction of protein-coding genes often positively correlates with and is required for the expression of their cognate sense genes. We further characterized *MAS*, a NAT-IncRNA produced from the *MADS AFFECTING FLOWERING4* (*MAF4*) locus. *MAS* is induced by cold and indispensable for the activation of *MAF4* transcription and suppression of precocious flowering. *MAS* activates *MAF4* by interacting with WDR5a, one core component of the COMPASS-like complexes, and recruiting WDR5a to *MAF4* to enhance histone 3 lysine 4 trimethylation (H3K4me3). Our study greatly extends the repertoire of IncRNAs in *Arabidopsis* and reveals a role for NAT-IncRNAs in regulating gene expression in vernalization response and likely in other biological processes.
lncRNAs have emerged as important players in the regulation of gene transcription, splicing, and translation. Based on their relationship with protein-coding genes, lncRNAs can be classified as natural antisense transcripts (NATs), overlapping lncRNAs (OT-lncRNAs), long intergenic non-coding RNAs (lincRNAs), and intronic non-coding RNAs (incRNAs). NAT-lncRNAs are transcribed in the opposite direction of protein-coding genes, OT-lncRNAs partially or fully overlap protein-coding genes in the sense direction, whereas lincRNAs and incRNAs originate from intergenic and intronic regions, respectively.

NAT-lncRNAs are widespread in animals and plants. They and their cognate sense transcripts often exhibit concordant or discordant expression patterns. NAT-lncRNAs may positively or negatively regulate the expression of their sense transcripts using diverse transcriptional or post-transcriptional mechanisms. The transcriptional machineries of NAT-lncRNAs and their sense transcripts may compete for RNA Polymerase II (RNA Pol II) and regulatory transcription factors, or undergo collision resulting in transcriptional interference. Moreover, NAT-lncRNAs can serve as scaffolds to recruit DNA-modifying and histone-modifying enzymes, thereby facilitating DNA methylation, histone modifications, chromatin conformation changes, and eventually upregulation or downregulation of gene transcription.

In plants, thousands of lncRNAs have been identified and implicated in root development, seedling light response, flowering time control, reproduction, and stress response. They and their cognate sense transcripts often exhibit concordant or discordant expression patterns. NAT-lncRNAs may positively or negatively regulate the expression of their sense transcripts using diverse transcriptional or post-transcriptional mechanisms. The transcriptional machineries of NAT-lncRNAs and their sense transcripts may compete for RNA Polymerase II (RNA Pol II) and regulatory transcription factors, or undergo collision resulting in transcriptional interference. Moreover, NAT-lncRNAs can serve as scaffolds to recruit DNA-modifying and histone-modifying enzymes, thereby facilitating DNA methylation, histone modifications, chromatin conformation changes, and eventually upregulation or downregulation of gene transcription.

Post-transcriptionally, NAT-lncRNAs may affect mRNA decay by nucleases, mask miRNA binding sites, modulate protein translation or produce endogenous siRNAs to execute RNA interference (RNAi).

In plants, thousands of lncRNAs have been identified and implicated in root development, seedling light response, flowering time control, reproduction, and stress response. However, only a handful of plant lncRNAs have been experimentally characterized. CLOOAR is a set of alternatively spliced and polyadenylated transcripts transcribed from the FLOWERING LOCUS C (FLC) locus at an early stage of cold exposure and mediates the reduction of active histone mark H3 lysine 36 trimethylation (H3K36me3) and an increase of repressive histone mark H3K27me3. COLDAIR is induced at a later stage of cold exposure and cooperates with an FLC promoter-derived lncRNA COLDWRAP to establish high H3K27me3 and silence FLC. The lincRNA APOLIO is transcribed in response to auxin and regulates root development through mediating the formation of a chromatin loop encompassing the promoter of its neighboring gene PID and downregulating the transcription of PID.

The lncRNA HID1 induced by continuous red light also transcriptionally suppresses its target gene and promotes seedling photomorphogenesis. The elf18-induced lncRNA ELENA enhances PRI expression through interacting with MED19a and affecting its enrichment on the PRI promoter. Instead of being transcriptional regulators, ASCO-lncRNA was found to associate with the nuclear speckle RNA-binding protein (NSR) and mediate NSR-mediated alternative splicing events through mimicking and displacing pre-mRNA targets. Similarly, the lincRNA IPSI inhibits the activity of phosphate starvation-induced mR399 by mimicking and sequestering mR399 target mRNA. Two rice lincRNAs PMSIT and LDMAR were shown to regulate photoperiod-sensitive male sterility. Whereas PMSIT functions through generating phased small interfering RNAs (phasRNAs), the molecular basis of LDMAR function remains a mystery.

In this study, in order to explore the function of lncRNAs in gene regulation and the range of such regulation in plants, we employed high-depth strand-specific RNA sequencing (RNA-seq) to systematically identify lncRNAs in Arabidopsis thaliana. We annotated 6510 lncRNAs including 4050 NAT-lncRNAs and 2460 lincRNAs. We found that many NAT-lncRNAs and their cognate protein-coding sense transcripts are concordantly expressed in different tissues or under stress conditions and knocking down NAT-lncRNAs leads to decreased expression of sense transcripts. We further demonstrated that one NAT-lncRNA, MAS, positively regulates the transcription of its cognate sense gene MAF4 through interacting with and recruiting WDR5a, a core component of the COMPASS-like complexes, to MAF4, thereby regulating flowering time. Our study provides a resource for studying lncRNAs in Arabidopsis and reveals a mechanism for gene regulation by NAT-lncRNAs.

Results

Global identification of lncRNAs in Arabidopsis. To globally identify lncRNAs in Arabidopsis, we reconstructed an Arabidopsis transcriptome using high-depth strand-specific RNA sequencing (ssRNA-seq). We generated cDNA libraries for rRNA-depleted total, polyadenylated [poly(A)+] and non-polyadenylated [poly(A)−] RNAs in whole cell extract, nuclear and cytosolic fractions that were prepared from Arabidopsis grown under normal or stress conditions (Supplementary Data 1, RNA-seq datasets numbered 1–34). A total of 1.2 billion genome-matched reads were obtained. These reads, together with the reads obtained from 3 published RNA-seq datasets, were assembled to reconstruct the Arabidopsis transcriptome. This resulted in 106,421 unique transcripts from 64,987 genomic loci. Among these, 25,245 were previously annotated protein-coding transcripts (TAIR10), accounting for 93% of all annotated protein-coding transcripts. This indicates that the reconstructed transcriptome had reasonably high coverage and quality. After the removal of 39,082 transcripts corresponding to protein-coding transcripts, other known ncRNAs (e.g., miRNAs, tRNAs, and rRNAs), 29,463 transcripts with short length (< 150 nt) or low abundance (FPKM < 1), 25,270 transcripts with protein-coding potential (CPC score > 0), and 6096 transcripts partially or fully overlapping with protein-coding genes in the sense direction, we annotated 6510 lncRNAs ((Supplementary Fig. 1a and Supplementary Data 2). These lncRNAs include 4050 NAT-lncRNAs and 2460 lincRNAs (Fig. 1). NAT-lncRNAs were further classified into overlapping (2117), divergent (1296) and convergent (637) NAT-lncRNAs (Fig. 1).

Characteristics of Arabidopsis lncRNAs. We analyzed features of the identified lncRNAs including average size, exon number, isoform number, and expression level. Same analyses were also performed for protein-coding transcripts in parallel for comparison. We found that lncRNAs were much shorter than coding RNAs (mean length of 633 nt for lncRNAs versus 1408 nt for coding RNAs) (P-value < 0.0001, Mann–Whitney U-Test, one-tailed) (Supplementary Fig. 1b). The lncRNAs had fewer exons (mean = 3.7) than coding RNAs (mean = 5.9) (P-value < 0.0001, Mann–Whitney U-Test, one-tailed) (Supplementary Fig. 1c) and smaller number of isoforms (mean = 1.3) comparable to coding RNAs (mean = 1.4) (Supplementary Fig. 1d). The expression levels of lncRNAs and coding RNAs were estimated by fragments per kilobase of exonic sequence per million mapped reads (FPKM) using Cuffdiff. The expression levels of lncRNAs were lower than those of coding RNAs (P-value < 0.0001, Mann–Whitney U-Test, one-tailed) (Supplementary Fig. 1e).

We examined whether lncRNAs are polyadenylated, taking advantage of RNA-seq datasets for poly(A)+ and poly(A)− (Supplementary Data 1). By applying a strict criterion (P-value < 0.05 and fold-change ≥2), we found that 1352 lncRNAs were significantly enriched in the poly(A)+ fraction, whereas 198 lncRNAs were significantly enriched in the poly(A)− fraction (Supplementary Fig. 2a and
LncRNAs are developmentally and physiologically regulated.

To investigate whether the identified lncRNAs are developmentally and physiologically regulated, we estimated the expression levels of each lncRNA by calculating FPKM in different tissues (seedling, inflorescence, and siliques) or under different treatments (cold, ABA and drought) using the RNA-seq datasets, which include three biological replicates for each sample. The Pearson correlation coefficients close to 1 indicate high reproducibility of the RNA-seq experiments (Supplementary Fig. 3). We found that 627 lncRNAs had differential expression in different tissues (P-value < 0.05 and fold-change ≥2) (Fig. 2a and Supplementary Data 5). 510 and 509 lncRNAs showed inducible expression patterns at one time point upon ABA and drought treatment, respectively (Fig. 2a and Supplementary Data 6, 7). We also found that 196 lncRNAs including COOLAIR showed a significant increase or decrease in their expression levels after cold treatment (Fig. 2a and Supplementary Data 8). The expression patterns of several randomly selected lncRNAs were confirmed by quantitative RT-PCR (RT-qPCR) (Fig. 2b-d). These data show the dynamic changes of lncRNA expression in response to developmental and environmental cues and suggest their roles in development and stress responses.

NAT-lncRNAs regulate the expression of cognate sense genes.

To explore the function of lncRNAs in gene regulation, we first examined whether lncRNAs and their adjacent genes are concordantly or discordantly expressed. We calculated the Pearson correlation coefficients (p.c.c.) between the different types of lncRNAs and their adjacent protein-coding genes. The p.c.c. values between adjacent protein-coding gene pairs were calculated in parallel for comparison. We found that the p.c.c. values of overlapping NAT-lncRNA/sense gene pairs were significantly higher than the values between adjacent protein-coding pairs (Fig. 3a), suggesting that overlapping NAT-lncRNAs have a stronger tendency to have positively correlated expression patterns with their sense overlapping genes. The concordant expression patterns of 216 overlapping NAT-lncRNAs and their cognate sense genes (p.c.c. score > 0.6) are shown in Fig. 3b.

The finding of concordant expression of NAT-lncRNAs and their cognate genes led us to examine whether NAT-lncRNAs play a role in regulating the expression of their cognate genes. We knocked down 21 NAT-lncRNAs using artificial microRNAs (amiRNAs) (Supplementary Fig. 4). Interestingly, the reduction of 15 and 3 NAT-lncRNAs resulted in significantly decreased and increased expression of their cognate sense genes, respectively. The reduction of other 3 NAT-lncRNAs did not significantly change the expression of their cognate sense genes (Fig. 3c, Supplementary Figs. 5, 6). Alteration of sense gene expression in amiRNA knockdown lines was not due to targeting of sense genes by amiRNA*’s. Eight out of 21 amiRNA*’s do not base pair with sense mRNAs at all. The rest of the amiRNA’s have mismatches to corresponding sense mRNAs at critical positions (Supplementary Fig. 4). Furthermore, most of the amiRNA*’s do not have 5’ terminal uridine (Supplementary Fig. 4), making it less likely that they are loaded into the effector AGO1 to suppress gene expression. To further rule out the possibility that production of secondary siRNAs targeting sense genes leads to alteration of sense gene expression, we performed small RNA (sRNA) sequencing on 12 randomly chosen amiRNA knockdown lines. The results revealed that no secondary siRNAs were detected in these lines (Supplementary Fig. 7). Together, our data suggest that NAT-lncRNAs are involved in the regulation of cognate sense gene expression.

A natural antisense lncRNA regulates MAF4 gene expression.

The finding that NAT-lncRNAs regulates cognate sense gene expression prompted us to investigate the biological importance of such regulation. We focused on one NAT-lncRNA, NAT-lncRNA_2962. NAT-lncRNA_2962 is transcribed from the antisense strand of the cold-responsive MAF4 gene, a FLC family member that functions to prevent precocious vernalization response. We renamed it MAS for MAF4 antisense RNA (Fig. 4a). RACE analyses showed that the 5’ end of MAS is initiated at a site several nucleotides to the transcription termination site (TTS) of MAF4 and the 3’ end of MAS extends into the 1st intron of MAF4 and undergoes polyadenylation (Supplementary Fig. 8a).

MAF4 is induced during early periods of cold exposure and its expression peaks at 20th day of cold exposure. We validated the temporal expression pattern of MAF4 by RT-qPCR. Intriguingly, the expression pattern of MAS during cold treatment closely mimicked that of MAF4 (Fig. 4b). The concordant expression of MAF4 and MAS suggests that either MAF4 transcript promotes MAS expression or vice versa. We tested the first possibility by examining the expression of MAS in maf4-1 that contains a natural antisense lncRNA. We estimated the partitioning of each lncRNA between the nucleus and the cytoplasm by analyzing the RNA-seq datasets for cytosolic (SC_Total) and nuclear fractions (SN_Total) (Supplementary Data 1). We found that 239 lncRNAs had significantly higher levels in the nuclear fraction than that in the cytosolic fraction, whereas only 43 lncRNAs were more abundant in the cytosolic fraction (P-value < 0.05 and fold-change ≥2) (Supplementary Fig. 2c and Supplementary Data 4). RT-PCR analyses with fractionated nuclear and cytosolic extracts confirmed that all 10 randomly selected lncRNAs were predominantly localized in the nucleus (Supplementary Fig. 2d).

The results revealed that no secondary siRNAs were detected in these lines (Supplementary Fig. 7). Together, our data suggest that NAT-lncRNAs are involved in the regulation of cognate sense gene expression.
Fig. 2 LncRNAs are developmentally and physiologically regulated. a Heat maps showing the abundances of differentially expressed lncRNAs in different plant tissues and in plants treated with ABA, drought or cold. Rows are ordered based on a k-means clustering of lncRNAs. Color intensity represents the fractional density across the row of FPKM counts. b–d Detection of representative lncRNAs in the indicated samples by RT-qPCR. Error bars represent s.e. m (n = 3), asterisks indicate a significant difference (t-test, P-value < 0.05). Source data are provided as a Source Data file.
**Fig. 3** NAT-lncRNAs regulate the expression of cognate sense genes. **a** A boxplot showing the correlation of expression patterns between neighboring gene pairs. CCO, overlapped protein-coding gene pair; CCD, divergent protein-coding gene pair; CCC, convergent protein-coding gene pair; NCO, overlapping NAT-lncRNA and associated protein-coding gene pair; NCD, divergent NAT-lncRNA and closest neighboring gene pair; NCC, convergent NAT-lncRNA and closest neighboring gene pair; LC, lincRNA and closest neighboring gene pair. The central lines, bounds of box represent the median, 25% quartile and 75% quartile. The whiskers represent 1.5 × IQR of the lower or upper quartile. Asterisks indicate a significant difference between the indicated groups (Mann-Whitney U-test, P-value < 0.01). **b** Heat maps showing the expression patterns of 216 NCO pairs. Rows are ordered based on a k-means clustering of NAT-lncRNAs. Color intensity represents the fractional density across the row of FPKM counts. **c** Detection of NATs-lncRNAs and their cognate sense genes by RT-qPCR in Col-0 and indicated amiRNA knockdown lines. Error bars represent s.e.m (n = 3), asterisks indicate a significant difference (t-test, P-value < 0.05). Source data are provided as a Source Data file. Shown above the RT-qPCR results are genome browser views of RNA-seq signals at NAT-lncRNAs and cognate sense genes in Col-0, with normalized read counts per million along the y-axis. More examples are available in Supplementary Figs. 5, 6.
T-DNA insertion in the largest intron of MAF4 and has abolished MAF4 expression before and after cold treatment (Fig. 4c) and two amiRNA lines (amiR-MAF4-1 and amiR-MAF4-2) in which MAF4 transcript was knocked down (Fig. 4d). The basal expression and induction of MAS were not disturbed in both maf4-1 and amiR-MAF4-1/2 (Fig. 4c, d), indicating that MAF4 transcript does not affect MAS expression. We then tested whether MAS regulates MAF4 expression. We generated two amiRNA lines (amiR-MAS-1 and amiR-MAS-2) in which MAS transcript was knocked down (Fig. 4a, e). In both lines, the basal

---

**Fig. 4** A natural antisense lncRNA regulates MAF4 gene expression. A Genome browser view of MAF4 and MAS. Signals from RNA- and sRNA-seq are shown, with normalized read counts per million along the y-axis. T-DNA insertion and amiRNA target sites are indicated. TSS, transcription start site; TTS, transcription termination site. B Detection of MAF4 and MAS in Col-0 after 0–28 d of cold exposure by RT-qPCR. C Detection of MAF4 and MAS in Col-0 and maf4-1 before (0 d) and after 20 d of cold exposure. D Detection of MAF4 and MAS in Col-0 and two MAF4 amiRNA knockdown lines (amiR-MAF4-1 and amiR-MAF4-2) before (0 d) and after 20 d of cold exposure. E Detection of MAS and MAF4 in Col-0 and two MAS amiRNA knockdown lines (amiR-MAS-1 and amiR-MAS-2) before (0 d) and after 20 d of cold exposure. In b–e, error bars represent s.e.m (n = 3), asterisks indicate a significant difference between the indicated groups (t-test, P-value < 0.05). F Flowering-time phenotypes of Col-0, maf4-1, and amiR-MAS-1/2 lines grown in SD conditions after 20 d of cold exposure. Numbers of primary rosette leaves were counted when bolts were ~3–5 cm long. At least thirty plants of each genotype were used for statistical analysis. Asterisks indicate a significant difference (t-test, P-value < 0.05). Source data are provided as a Source Data file.

---

**ARTICLE** NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-07500-7

---

**NATURE COMMUNICATIONS** | (2018) 9:5056 | DOI: 10.1038/s41467-018-07500-7 | www.nature.com/naturecommunications
level of \(M4\) transcript was reduced and the induction of \(M4\) expression by cold was severely compromised as well (Fig. 4c). Similar to the reduction of sense gene expression in other amIR knockdown lines we generated (Supplementary Fig. 4), the reduction of \(M4\) expression was not due to amIRNA’s targeting \(M4\) mRNA (Supplementary Fig. 8b) or amIRNA-triggered production of secondary siRNAs (Fig. 4a). Thus, our data strongly support the notion that \(M4\) plays a positive role in \(M4\) expression. \(M5\), another FLC family member that functions to prevent precocious vernalization response, is near the \(M4\) gene. \(M5\) expression remained unaltered in amIR-\(M4\), suggesting that \(M4\) does not regulate the expression of \(M5\) (Supplementary Fig. 8c).

The role of \(M4\) in \(M4\) expression prompted us to examine whether \(M4\) also acts as a floral repressor. We examined the flowering phenotype of amIR-\(M4\) and the wild-type (Col-0) plants grown in short-day conditions after 20 days of cold exposure. We found that, like the \(M4\) mutant, amIR-\(M4\) flowered earlier than Col-0 (Fig. 4f). All together, these results suggest that \(M4\) transcript is necessary for the expression of \(M4\) and repression of flowering.

**MAS promotes \(M4\) expression at the transcriptional level.** We next investigated how \(M4\) regulates \(M4\). As \(M4\) is complementary to the \(M4\) transcript, it was possible that they form a double-stranded RNA to produce sRNAs. However, few sRNAs were detected at the overlapping region of \(M4\) and \(M4\) (Fig. 4a), excluding the possibility that \(M4\) regulates \(M4\) via a mechanism involving sRNAs.

To examine whether \(M4\) regulates \(M4\) expression through modulating the stability of \(M4\) transcripts, we measured the RNA decay rate of \(M4\) in Col-0 and amIR-\(M4\) lines treated with the transcriptional inhibitor actinomycin D (ActD). ActD effectively blocked the transcription of both \(M4\) and \(M4\) as well as that of \(GAPDH\). However, the decline rates of \(M4\) transcripts in Col-0 and amIR-\(M4\) were indistinguishable, suggesting that \(M4\) does not regulate the stability of \(M4\) transcript (Fig. 5a).

We next tested whether \(M4\) transcriptionally promotes \(M4\) expression. We performed nuclear run-on (NRO) assay to assess the transcriptional rate of \(M4\) in Col-0 and amIR-\(M4\) lines. As expected, \(M4\) expression prompted us to examine whether \(M4\) also acts as a floral repressor. We examined the flowering phenotype of amIR-\(M4\) and the wild-type (Col-0) plants grown in short-day conditions after 20 days of cold exposure. We found that, like the \(M4\) mutant, amIR-\(M4\) flowered earlier than Col-0 (Fig. 4f). All together, these results suggest that \(M4\) transcript is necessary for the expression of \(M4\) and repression of flowering.
overexpressed most likely at loci other than MAF4 (Fig. 5e). The expression level of MAF4 did not change in either transgenic line (Fig. 5e), suggesting that MAS functions to regulate MAF4 in cis, but not in trans. Taken together, our data indicate that cis-acting MAS activates MAF4 expression at the transcriptional level.

**Discussion**

The function and range of lncRNA-mediated regulation in plants have been increasingly appreciated. In this study, we identified a large number of lncRNAs and analyzed their expression profile in different tissues under normal or stress conditions using strand-specific RNA-seq. We sequenced poly(A)+ and poly(A)−, nuclear and cytoplasmic RNAs separately to increase the sensitivity of detecting lncRNAs with distinct features. Indeed, ~88% of the lncRNAs we identified have not been previously discovered by tilling arrays or conventional RNA-seq[11,20,25]. Thus, the lncRNAs we identified represent a valuable addition to the *Arabidopsis* lncRNA collection, and provide a rich resource for the community to further investigate the biology of lncRNAs in plants.

We focused on the function of a NAT-lncRNA MAS in the activation of MAF4 expression during cold treatment. We found that that MAS acts in cis to activate MAF4 expression at the transcriptional level (Figs. 4, 5). The transcriptional activating role of MAS is similar to that played by lncRNAs *HOTTIP*3, *NeST*5, *LAIR*3, and *EVX1as*6. However, the mechanisms adopted by these lncRNAs are varied. MAS binds WDR5a and then guides the COMPASS-like complexes to MAF4 to promote H3K4me3 (Fig. 6). Like MAS, *HOTTIP* also interacts with WDR5 and recruits the MLL complex to maintain H3K4me3 and activation of HOXA genes53. However, the cis-regulatory action of *HOTTIP* and *NeST* requires the chromosome looping that brings the *HOTTIP* or *NeST* locus into close spatial proximity to its target genes3. Ectopic overexpression of MAS cannot stimulate MAF4 expression, whereas ectopic overexpression of *LAIR* promotes the upregulation of LRK genes. *EVX1as* increases the transcription of *EVX1* through facilitating the binding of Mediator complex to *EVX1* region, leading to an active chromatin state.

Exemplified by MAS, many NAT-lncRNAs were found to be concordantly expressed with their sense genes (Fig. 3), suggesting co-upregulation of NAT-lncRNAs and their sense genes. Our results are consistent with previous findings that neighboring genes often have correlated expression irrespective of their orientations57. Also, previous study of an immediate-early gene (IEGs) revealed that the ripple effect plays an important role in transcriptional activation of IEGs and their neighboring genes58. However, the cases of NAT-lncRNAs and comparisons between the ripple effects triggered by lncRNAs and regular genes were not included in the previous analysis58. Our genome-wide analysis revealed that NAT-lncRNAs are significantly more likely to produce ripple effects and activate their sense overlapping genes than regular genes and other types of lncRNAs (Fig. 3a). On one hand, this could be because the average distances between the TSSs of NAT-lncRNAs and their paired genes are smaller. On the other hand, this may reflect the fact that NAT-lncRNAs play crucial roles in activating the expression of their paired genes. We found that some NAT-lncRNAs are indeed required for the expression of their sense overlapping genes, suggesting that this cis-regulatory mode could be common to many NAT-lncRNAs (Fig. 3c and Supplementary Figs. 3, 4). Whether these NAT-lncRNAs regulate their cognate sense genes through recruiting the COMPASS-like complexes or other mechanisms remains to be investigated.

Our finding that many lncRNAs are responsive to different stresses suggests that plant lncRNAs may play crucial biological roles. COOLAIR and COLDAIR have been found to mediate vernalization-induced repression of the floral repressor *FLC*.3,29 Here we demonstrate that the lncRNA MAS, by regulating the expression of an *FLC* family member, MAF4, fine-tunes the time of flowering. However, different from the repressive roles of COOLAIR and COLDAIR, MAS activates the expression of MAF4. Whereas COLDAIR associates with a subunit of the conserved repressive complex PRC232, MAS binds to the core component of the COMPASS-like complex that achieves transcriptional activation. Then why the floral repressor *FLC* and MAF4 are oppositely regulated upon cold exposure? It was suggested that MAF4 and MAF5 are transiently activated to prevent precocious flowering so that plants only flower after a long period.
Fig. 6 MAS mediates the recruitment of WDR5a to MAF4. a Upper panel, schematic representation of MAF4 locus, the positions of primers (R1 to R5) used for ChIP- and ChIRP-qPCR are indicated. Lower panel, detection of H3K4me3 levels in Col-0 and amiR-MAS-1/2 lines after 20 d of cold exposure by ChIP-qPCR. b Detection of MAS and GAPDH in the whole cell, cytoplasmic and nuclear fractions by RT-PCR. c Detection of MAS and GAPDH in histone 3 (H3) immunoprecipitates by RT-PCR. d ChIP-qPCR analyses of MAS association with MAF4 locus after 20 d of cold exposure. Left panel, ChIRP enrichment of MAS transcript, but not actin transcript in both odd and even probe pools. Right panel, qPCR detection of different regions of MAF4 locus in immunoprecipitated DNA. Probes targeting lacZ mRNA were used as negative controls. e Association between MAS and WDR5a detected by RIP with anti-FLAG Magnetic Beads in control and transgenic (FLAG-WDR5a and FLAG-WDR5aF250A) plants after 20 d of cold exposure. Purification of WDR5a and WDR5aF250A was validated by western blot (upper panel). The levels of MAS and MAF4 in the immunoprecipitates were determined by RT-qPCR (lower panel). f Detection of WDR5a at MAF4 locus by ChIP-qPCR with an antibody against WDR5 in Col-0 and amiR-MAS-1/2 lines after 20 d of cold exposure. g Detection of WDR5a and WDR5aF250A at MAF4 locus by ChIP-qPCR in Col-0 and the transgenic plants after 20 d of cold exposure. In (a), (d–g) error bars represent s.e.m (n = 3–4), asterisks indicate a significant difference between the indicated groups (t-test, P-value < 0.05). h A model for MAS-mediated activation of MAF4 gene expression during vernalization. Source data are provided as a Source Data file.
of cold when FLC is completely silenced. The dynamic and different expression profiles of FLC and MAF4 highlight the important role of IncRNAs in coordinating the vernalization response. However, the majority of IncRNAs, involved in flowering time control or other stress responses, are still awaiting functional characterization.

**Methods**

**Plant materials and growth conditions.** All plants used in this study are in the Col-0 background. Detailed information about mutants and generation of transgenic plants can be found in Supplementary Methods. Plants were grown on 1/2 MS medium with 30 g/L sucrose in long-day (LD, 16 h light, 22 °C / 8 h dark, 18 °C) or short-day conditions (SD, 8 h light, 22 °C / 16 h dark, 18 °C).

Stress treatments were performed as previously described with some modifications. For ABA treatment, 2-week-old seedlings were transferred to 1/2 MS liquid medium with 100 μM ABA. For dehydration treatment, 2-week-old seedlings were removed from the agar and desiccated in dishes. After being treated for different time periods (0, 2, 4, 6, 8 h), the plants were harvested for RNA isolation. For cold treatment, 2-week-old seedlings (grown under SD conditions) were transferred to 4 °C and cultured under SD conditions for different time periods. After treatment, the plants were harvested or transplanted into soil and grown under SD conditions for flowering time测定。

**RNA-seq dataset were assembled into transcripts in a reference annotation-based manner.** ArtSeq reads were subjected to a BlastX search against all plant protein sequences in the Swiss-Prot database with a cutoff e-value < 10^-4 and the transcripts with short hits (alignment length ≤30 nt, percent identity ≥ 35% and coverage of the alignment region in either query or subject sequence ≥35%) to known proteins were considered to have protein-coding potential. For antisense transcripts, open reading frames were checked. The CPC (Coding Potential Calculator) score ≥34, a value to assess protein-coding potential of a transcript based on six biologically meaningful sequence features, was calculated for each transcript. When the CPC score is positive, we considered the transcript to have protein-coding potential. Transcripts that passed the three filtering steps were annotated as IncRNAs.

**Co-expression analysis.** Pearson correlation coefficient was calculated between the expression levels of adjacent protein-coding genes and between the expression levels of IncRNAs and their closest protein-coding genes. LncRNA/protein-coding gene pairs with low abundance (FPKMmax < 1) were excluded from our analysis. LncRNA/protein-coding gene pairs with Pearson correlation coefficients greater than 0.6 were presented in the heat map.

**Quantitative RT-PCR.** Total RNA was isolated using TRIzol reagent (Invitrogen) and treated with DNase I (Invitrogen) for 30 min at 37 °C to eliminate contaminating genomic DNA. cDNAs were generated using 2 μg of total RNA with random or gene-specific primers and M-MLV (Invitrogen). Quantitative RT-PCR was performed using SYBR Premix Ex Taq (Takara) as described. Each sample was analyzed in triplicate. The level of GAPDH mRNA was detected in parallel and used for normalization. Primer sequences are provided in Supplementary Data 9.

**ChIP and ChIP-qPCR analyses.** ChIP was performed as described with some modifications. A detailed protocol is available in Supplementary Methods. qPCR was performed using SYBR Premix Ex Taq (Takara). Relative enrichment of H3K4me3, H3K27Ac, H3K27me3, H3K36me3 and WDR5a in each DNA region was normalized to input DNA. Primer sequences are provided in Supplementary Data 9.

**RNA IP.** RNA IP (RIP) was performed as described. Briefly, plants were harvested and crosslinked by using 1% formaldehyde for 20 min. RNA-protein complexes were immunoprecipitated by incubating with anti-FLAG M2 Magnetic Beads (M8823, Sigma-Aldrich) and rabbit polyclonal anti-H3 (ab1791, Abcam, Cambridge, UK) at 4 °C for 6 h. Then, the crosslinking was reversed and RNA was purified by TRizol.

**Nuclear run-on assay.** Nuclear run-on assay was performed as described except that nuclei were isolated from 10-day-old seedlings as described. A detailed protocol is available in Supplementary Methods.

**ChIRP and ChIRP-qPCR analyses.** ChIRP was performed as previously with some modifications. Antisense DNA probes which were separated into two groups (even and odd) were designed against the full-length transcript. Total RNA was used as input. The reaction was performed at 37 °C for 30 min with rotation. Then beads were washed twice with high-salt wash buffer (2 × SSC, 0.5% SDS, 1 mM DTT) and one time with low-salt wash buffer (0.1 × SSC, 0.5% SDS, 1 mM DTT, and 1 mM PMSF) for 5 min each at room temperature. DNA and RNA were purified and analyzed by qPCR. Probes and primer sequences are provided in Supplementary Data 9.

**RNA decay assay.** RNA decay assay was performed as described with some modifications. Two-week-old seedlings of Col-0, amir-MAS-1, and amir-MAS-2 were treated with cold for different time periods (0, 2, 4, 6, 8 h) and then harvested for RNA isolation. The level of GAPDH mRNA was detected in parallel and used for normalization. Primer sequences are provided in Supplementary Data 9.
were grown in 4 °C growth chamber for 20 d. After cold treatment, plants were transferred into 1/2 MS medium with 100 μg/mL actinomycin D (Sigma-Aldrich). Materials were harvested after 2, 4, 6, 8 h. Total RNA was extracted by TRizol reagent and used for RT-qPCR assays.

5′ and 3′ RACE. Poly(A)+ RNAs were isolated from 100 ug total RNAs using oligo (dT) Dynabeads (Thermo Fisher). The 5′ and 3′ RACE experiments were preformed according to the manuals of GeneRacer (Invitrogen). For 3′ RACE, poly(A)+ RNAs were reversely transcribed with GeneRacer oligo (dT) primers and then amplified with GeneRacer 3′ Primer/1st primer and MAS-3′ RACE-GSP2/3′. For 5′ RACE, poly(A)+ RNAs were reversely transcribed with MAS-5′ RACE-GSP1. After degradation of RNAs, the cDNA was tailcd by dCTP and the second strand cDNA was generated using the Abridged Anchor Primer (AAP). Final amplification was performed with the Abridged Universal Anchor Primer (AUAP) and MAS-5′ RACE-GSP2/3′. Primer sequences are provided in Supplementary Data 9.

Data availability
RNA-Seq and sRNA-seq datasets generated in this study can be found in the NCBI Gene Expression Omnibus under accession number GSE42695 and GSE120709. A reporting summary for this Article is available as a Supplementary Information file. The source data underlying Figs. 2-6 and Supplementary Figs. 2-3, 5-6 and 8-10 are provided as a Source Data file. All other data that support the findings of this study are available from the corresponding author upon request.

Received: 22 May 2018 Accepted: 6 November 2018
Published online: 29 November 2018

References
1. Ariel, F., Romero-Barrios, N., Jego, T., Benhamed, M. & Crespi, M. Battles and hijacks: noncoding transcription in plants. Trends Plant Sci. 20, 362–371 (2015).
2. Quinn, J. J. & Chang, H. Y. Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. 17, 47–62 (2016).
3. Mattick, J. S. & Rinn, J. L. Discovery and annotation of long noncoding RNAs. Nat. Struct. Mol. Biol. 22, 5–7 (2015).
4. Iyer, M. K. et al. The landscape of long noncoding RNAs in the human transcriptome. Nat. Genet. 47, 199–208 (2015).
5. Faghihi, M. A. & Wahlestedt, C. Regulatory roles of natural antisense transcripts. Nucleic Acids Res. 34, 3465–3476 (2006).
6. Katayama, S. et al. Antisense transcription in the mammalian transcriptome. Science 309, 1561–1566 (2005).
7. Cheng, I. et al. Transcriptional maps of 10 human chromosomes at 5′-nucleotide resolution. Science 308, 1149–1154 (2005).
8. Wang, X. J., Gaasterland, T. & Chua, N. H. Genome-wide prediction and identification of cis-natural antisense transcripts (cis-NATs) in ten species. Nucleic Acids Res. 34, 3465–3476 (2006).
9. Deng, P., Liu, S., Nie, X., Weining, S. & Wu, L. Conservation analysis of long non-coding RNAs in plants. Science 318, 453 (2007).
10. Katayama, S. et al. Antisense transcription in Arabidopsis. J. Genet. Genom. 33, 92–102 (2006).
11. Deng, P., Liu, S., Nie, X., Weining, S. & Wu, L. Conservation analysis of long non-coding RNAs in plants. Science 318, 453 (2007).
12. Liu, X. J., Gaasterland, T. & Chua, N. H. An estrogen receptor-based reporter system for studying the regulation of long noncoding RNAs in Arabidopsis and rice. Proc. Natl Acad. Sci. USA 111, 16160–16165 (2014).
13. Kim, D. H. & Sung, S. Vernalization-triggered intragenic chromatin loop formation by long non-coding RNAs. Dev. Cell 40, 302–312 e4 (2017).
14. Arai, E. et al. Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. Mol. Cell 55, 383–396 (2014).
15. Wang, Y. & Erdjument-Bromage, Y. Natural antisense transcripts mediate control of photonmorphogenesis by red light. Proc. Natl Acad. Sci. USA 111, 10359–10364 (2014).
16. Seo, J. S.et al. ELF18-INDUCED LONG-NONCODING RNA associates with mediator to enhance expression of innate immune response genes in Arabidopsis. Plant Cell 29, 1024–1038 (2017).
17. Bardou, M. et al. Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. Dev. Cell 30, 166–176 (2014).
18. Trappnell, C. et al. Differencial gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578 (2012).
19. Mi, S. et al. Sorting of small RNAs into Arabidopsis argonaute complexes is coordinated by the 5′-terminal nucleotide. Cell 151, 1544–1545 (2012).
20. Ietswaart, R., Wu, Z. & Dean, C. Flowering time control: another window to the connection between antisense RNA and chromatin. Trends Genet. 28, 445–453 (2012).
21. Chen, J., Sun, M., Hurst, L. D., Carmichael, G. G. & Rowley, J. D. Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. Trends Genet. 21, 326–329 (2005).
22. Wight, M. & Werner, A. The functions of natural antisense transcripts. Essays Biochem. 54, 91–101 (2013).
23. De Lucia, F. & Dean, C. Long non-coding RNAs and chromatin regulation. Curr. Opin. Plant Biol. 14, 168–173 (2011).
24. Magistri, M., Faghihi, M. A., St Laurent, G. 3rd & Wahlestedt, C. Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. Trends Genet. 28, 389–396 (2012).
25. Koziol, M. J. & Rinn, J. L. RNA traffic control of chromatin complexes. Curr. Opin. Genet. Dev. 20, 142–148 (2010).
26. Zhang, X. et al. Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. Genome Biol. 13, R20 (2012).
27. Ben Amor, B. et al. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. Genome Res. 19, 57–69 (2009).
28. Di, C. et al. Characterization of stress-responsive IncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. Plant J. 84, 848–861 (2014).
29. Li, L. et al. Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biol. 15, R40 (2014).
30. Dou, M. et al. Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biol. 15, R40 (2014).
50. Jiang, D., Kong, N. C., Gu, X., Li, Z. & He, Y. Arabidopsis COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. *PLoS Genet* 7, e1001330 (2011).

51. Jiang, D., Gu, X. & He, Y. Establishment of the winter-annual growth habit via FRIGIDA-mediated histone methylation at FLOWERING LOCUS C in Arabidopsis. *Plant Cell* 21, 1733–1746 (2009).

52. Gomez, J. A. et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. *Cell* 152, 743–754 (2013).

53. Wang, K. C. et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120–1215 (2011).

54. Yang, Y. W. et al. Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *Elife* 3, e02046 (2014).

55. Wang, Y. et al. Overexpressing IncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat. Commun.* 9, 3516 (2018).

56. Luo, S. et al. Divergent IncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell* 18, 637–652 (2016).

57. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* 26, 183–186 (2000).

58. Ebisuya, M., Yamamoto, T., Nakajima, M. & Nishida, E. Ripples from neighbouring transcription. *Nat. Cell Biol.* 10, 1106–1113 (2008).

59. Seki, M. et al. Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. *Func. Integr. Genom.* 2, 282–291 (2002).

60. Seki, M. et al. Monitoring the expression profiles of 7,000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31, 279–92 (2002).

61. Ono, Y. et al. Monitoring expression profiles of Arabidopsis gene expression during rehydration process after dehydration using ca 7,000 full-length cDNA microarray. *Plant J.* 34, 868–87 (2003).

62. Wang, W. et al. An importin beta protein negatively regulates MicroRNA activity in Arabidopsis. *Plant Cell* 23, 3565–3576 (2011).

63. Ye, R. et al. Cytoplasmic assembly and selective nuclear import of Arabidopsis Argonaute4/siRNA complexes. *Mol. Cell* 46, 859–870 (2012).

64. Levin, J. Z. et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* 7, 709–U67 (2010).

65. Lamesch, P. et al. The Arabidopsis Information Resource (TAIR): improved publication maps and institutional affiliations. *NATURFORSCHUNG* 28, 511–510 (2010).

66. Cabili, M. N. et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927 (2011).

67. Pauli, A. et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res* 22, 577–591 (2012).

68. Consortium, T. U. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 42, D191–D198 (2014).

69. Kong, L. et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35, W345–9 (2007).

70. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25 (2009).

71. Zhong, X. et al. DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat. Struct. Mol. Biol.* 19, 870–875 (2012).

72. Wierzbiicki, A. T., Haag, J. R. & Pikaard, C. S. Noncoding transcription by RNA polymerase Pol IV/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135, 635–648 (2008).

73. Roberts, T. C. et al. Quantification of nascent transcription by bromouridine immunocapture nuclear run-on RT-qPCR. *Nat. Protoc.* 10, 1198–1211 (2015).

74. Zhu, P. et al. Arabidopsis small nucleolar RNA monitors the efficient pre-rRNA processing during ribosome biogenesis. *Proc. Natl Acad. Sci. USA* 113, 11967–11972 (2016).

75. Narsai, R. et al. Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. *Plant Cell* 19, 3418–3436 (2007).

**Acknowledgements**

We are grateful to Dr. Yuehui He for providing us WDR5a mRNA lines. This work was supported by grants from National Natural Science Foundation of China to Y.Q. (Grant No. 31788103) and to J.L. (Grant No. 31400675) and from National Key R&D Program of China (Grant No. 2016YFA0500800) to Y.Q. Y.Q. is a visiting investigator of the CAS Center for Excellence in Molecular Plant Sciences.

**Author contributions**

X.Z., J.L., Y.L., and Y.Q. conceived this project, designed experiments, and analyzed data. X.Z., J.L., and H.G. performed experiments and B.L. did bioinformatic analyses. J.L., Y.L., and Y.Q. wrote the manuscript. All authors discussed the results and made comments on the manuscript.

**Additional information**

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07500-7.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

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

---

© The Author(s) 2019

---

[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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.