

Review

Molecular Functions of Long Non-Coding RNAs in Plants

Qian-Hao Zhu * and Ming-Bo Wang

CSIRO Plant Industry, GPO Box 1600, Canberra ACT 2601, Australia; E-Mail: ming-bo.wang@csiro.au

* Author to whom correspondence should be addressed; E-Mail: qianhao.zhu@csiro.au;
Tel.: +61-02-6246-4903; Fax: +61-02-6246-5000.

Received: 2 February 2012; in revised form: 28 February 2012 / Accepted: 29 February 2012 /
Published: 8 March 2012

Abstract: The past decade has seen dramatic changes in our understanding of the scale and complexity of eukaryotic transcriptome owing to the discovery of diverse types of short and long non-protein-coding RNAs (ncRNAs). While short ncRNA-mediated gene regulation has been extensively studied and the mechanisms well understood, the function of long ncRNAs remains largely unexplored, especially in plants. Nevertheless, functional insights generated in recent studies with mammalian systems have indicated that long ncRNAs are key regulators of a variety of biological processes. They have been shown to act as transcriptional regulators and competing endogenous RNAs (ceRNAs), to serve as molecular cargos for protein re-localization and as modular scaffolds to recruit the assembly of multiple protein complexes for chromatin modifications. Some of these functions have been found to be conserved in plants. Here, we review our current understanding of long ncRNA functions in plants and discuss the challenges in functional characterization of plant long ncRNAs.

Keywords: long non-coding RNA; RNA-seq; natural miRNA target mimic; chromatin modifier

1. Introduction

Whole-genome tiling array and RNA sequencing (RNA-seq) have revealed that the transcription landscape in eukaryotes is much more complex than had been expected, with a high proportion of novel transcripts generated from intergenic regions and promoters of annotated genes [1]. Meanwhile, natural antisense transcripts, which are RNA molecules transcribed from the opposite DNA strand and

overlapping in part or full with the sense transcripts [2], have been shown to be a pervasive feature of mammalian genomes [3,4]. Antisense transcripts were also found in ~30% of annotated genes in *Arabidopsis* [5]. Although ~90% of the human genome is transcribed [6], the ENCODE project demonstrated that only ~1.2% of the genome encodes proteins [7], suggesting that a large proportion of the eukaryotic genome produces RNA molecules that have no protein-coding capacity, namely non-coding RNAs (ncRNAs).

ncRNAs are arbitrarily grouped into short (<200 nt) and long ncRNAs (lncRNAs; >200 nt). The importance of short ncRNAs, including siRNAs, miRNAs and piRNAs, in transcriptional and posttranscriptional regulation of gene expression has been well recognized and the molecular mechanisms of short ncRNA-mediated regulation have been well understood [8,9]. In contrast, the regulatory roles of lncRNAs are only beginning to be recognized and the molecular basis of lncRNA-mediated gene regulation is still poorly understood. Studies on a small number of lncRNAs in animals have shown that they are involved in multiple levels of the gene regulation. These lncRNAs have been shown to mediate epigenetic changes through recruitment of the Polycomb repressive complex (PRC) [10–12], to act as decoy for splicing factors [13] and to compete for miRNA binding sites [14–18]. In comparison to animals, plants have fewer lncRNAs been identified [19,20] and functionally characterized [21–24]; however, the emerging picture is that the regulatory functions of plant lncRNAs are largely similar to animal lncRNAs. In this review we first provide a brief introduction of the methodologies used in lncRNA identification, and then summarize recent progresses in functional characterization of lncRNAs in plants. We also discuss the challenges in unveiling the functions of lncRNAs. Our focus in this review is on the lncRNA functions that are independent of siRNA-directed gene silencing pathways; functions of RNA polymerase V-dependent lncRNAs involved in RNA-directed DNA methylation and precursor transcripts of trans-acting siRNAs (tasiRNAs) will not be discussed. We refer readers to excellent recent reviews on these topics [25–27].

2. Discovery of lncRNAs

2.1. In Silico Identification

The rationale for *in silico* identification of lncRNAs is that they can be distinguished from protein-coding mRNAs based on the absence of discernible open reading frames (ORFs). The starting data for *in silico* identification can be sequences of cDNAs or Expressed Sequence Tags (ESTs) deposited in public databases or novel transcripts generated by full-length cDNA cloning, tiling arrays and RNA sequencing (see below). Usually cDNAs or EST sequences are first compared with genomic sequences to remove those overlapping with protein-coding genes; the remaining sequences are then subjected to ORF prediction. The threshold of ORF length is usually 70–100 amino acids, *i.e.*, RNAs with a predicted ORF of <70–100 amino acids would be treated as lncRNAs. Existing ORF prediction programs include GeneMark.hmm [28], GenScan [29], ESTScan2 [30], ANGLE [31] and ORF-Predictor [32]. More sophisticated bioinformatics tools for estimating the protein-coding potential of a RNA sequence include CRITICA [33], DIANA-EST [34], CSTminer [35], CONC [36], Coding Potential Calculator [37], integrated ncRNA finder [38] and RNAcode [39]. The *in silico* approach has been successfully applied to identifying lncRNAs in both plants [19] and animals [31,40,41].

2.2. De Novo Identification

2.2.1. Whole-Genome Tiling Array and RNA-seq Approaches

Full-length cDNA sequencing is the gold standard for determining exonic structure and coding or non-coding potential of a transcript; however, this approach is time-consuming and expensive. Furthermore, RNAs with low-level expression, a characteristic of most lncRNAs, would be hard to uncover using traditional cDNA cloning strategies. Tiling DNA microarray, designed for genome-wide high-resolution transcriptome analysis, provides an alternative for detection of lncRNAs and their expression. Using this approach, a large number of uniquely transcribed intergenic regions and stress-induced novel transcripts were found in rice [42] and *Arabidopsis* [43,44], respectively. These novel transcripts provided a rich source for lncRNA discovery. In addition, a single nucleotide resolution array designed for the *Arabidopsis FLC (FLOWERING LOCUS C)* locus and its 50-kb surrounding region uncovered a number of non-coding transcripts antisense to *FLC* [24]. However, tiling arrays rely on the existing knowledge of genome sequence. This technology allows for the identification of novel exons or transcriptional units but it does not provide information about their connections. Furthermore, it still suffers from a lack of high sensitivity in detecting rare transcripts due to high levels of background, cross-hybridization of related sequences and saturation of signals.

Some of the drawbacks with tiling arrays can be circumvented by RNA-seq, which has emerged as a new technology for tackling the complexity of eukaryotic transcriptomes in an unbiased manner [6,45,46]. RNA-seq is able to detect transcripts that are missing or incomplete in the reference genome and allows for accurate quantification of expression levels, making it an ideal approach for lncRNA discovery. With an ultra sequencing depth RNA-seq can be used to discover rare transcripts that are expressed in just a few cells within a tissue. For instance, by combining RNA-seq with targeted RNA capture, a ~4,607 fold coverage was achieved for the targeted human genomic regions. With this RNA-seq depth, rare alternative splicing variants of the lncRNA *HOTAIR* were identified, and lncRNAs expressed in only a small subpopulation of the cells sampled could be detected [47]. In another study, ~2,000 novel transcribed regions that do not link to any annotated gene models were identified by a comprehensive investigation of the *Drosophila melanogaster* transcriptome using tiling arrays in combination with RNA-seq. Approximately two thirds of these novel transcripts have an ORF less than 100 amino acids, including a multi-exon lncRNA in the well-studied Bithorax complex, which is expressed in embryos and adult males but not in females [48]. When combined with RNA immunoprecipitation, RNA-seq could facilitate the identification of lncRNAs associated with specific RNA-binding proteins and chromatin remodeling complexes [49,50].

2.2.2. Chromatin Signature-Based Approach

An actively transcribed region is usually defined by a K4-K36 domain, *i.e.*, an active promoter marked by H3K4me3 (trimethylation of lysine 4 of histone H3) in combination with a transcribed region marked by H3K36me3 (trimethylation of lysine 36 of histone H3). By searching K4-K36 domains in intergenic regions in the human and mouse genomes, a large number of lncRNAs, named as long intergenic ncRNAs (lincRNAs), were found in these two species. A significant number of these lincRNAs are conserved between human and mouse [49,51], suggesting that they are functional.

The finding that ~38% of these lincRNAs were physically associated with chromatin modifying complexes, such as Polycomb repressive complex 2 (PRC2) and/or CoREST, further suggests a regulatory role of these lincRNAs [49]. This approach has not yet been adopted in plants mainly due to lack of genome-wide H3K36me3 data although genome-wide H3K4me3 landscape has been established in *Arabidopsis* [52].

3. Molecular Functions of lincRNAs in Plants

3.1. LincRNA as Natural miRNA Target Mimic

Phosphate is an essential macronutrient for plant growth and development. Plants must not only absorb considerable amounts of phosphate from the soil but must also have a sophisticated regulatory mechanism to maintain phosphate homeostasis throughout the plant to meet the growth and metabolic requirements of each tissue. miRNAs have been shown to be an essential component of this complex regulatory system [53–57]. miR399, which is expressed in companion cells and phloem, is strongly induced by phosphate starvation [53]. Consequently the expression level of *PHO2*, a target of miR399 and encoding an E2 ubiquitin conjugase-related enzyme (UBC24), is repressed due to miR399-mediated mRNA cleavage [53–56]. Low *PHO2* activity leads to enhanced expression levels of two root-specific phosphate transporter genes, *Pht1;8* and *Pht1;9* [53,54], resulting in increased phosphate uptake. Besides miR399, *Induced by Phosphate Starvation1 (IPSI)*, a member of the *TPSI/Mt4* gene family that was first identified in tomato and *Medicago truncatula* [58,59] and then in other plant species including rice [60] and *Arabidopsis* [61,62], is also induced by phosphate starvation. *IPSI* does not encode a protein, and only a 23-nt long sequence motif is conserved among the members from different plant species [21,63,64]. This 23-nt motif is partially complementary to miR399 with a 3-nt central mismatch corresponding to positions 11–13 of miR399. As miRNA-mediated RNA cleavage usually occurs between nucleotides 10 and 11 relative to the 5' end of the miRNA, this central mismatch disrupts crucial base-pairing between miR399 and *IPSI* and hence inhibits miR399-mediated cleavage of *IPSI*. This observation leads to the hypothesis that *IPSI* functions as a non-cleavable target mimic of miR399 to sequester miR399 which in turn attenuates miR399-mediated repression of *PHO2* [21]. Indeed, transgenic plants overexpressing *IPSI* increased the transcript and protein levels of *PHO2*, whereas transgenic plants overexpressing a cleavable *IPSI* did not [21]. Thus, the increased expression of *IPSI* under phosphate starvation appears to counter-balance the effect of increased miR399 accumulation under the same condition, resulting in fine tuning of *PHO2* expression and phosphate uptake [63].

Such inhibition of miRNA activity by an endogenous noncleavable ncRNA target has been termed as target mimicry [21]. Recent discovery of competing endogenous RNAs (ceRNAs) in animal and human cells indicates that target mimicry may be a widespread phenomenon, where non-coding and coding RNAs with similar miRNA target sites could affect each other's activity.

The first example of ceRNA is the human pseudogene *PTENP1*, which is related to the tumor suppressor gene *PTEN* and produces a naturally occurring ncRNA. Both *PTEN* and *PTENP1* contain many conserved miRNA binding sites in their 3' untranslated regions (UTRs). *PTENP1* was found to regulate the expression of *PTEN* by acting as a decoy for miRNAs that bind to the common sites in the

3' UTRs of *PTENP1* and *PTEN* [14,16,18]. More recently, a muscle-specific lncRNA, *linc-MD1*, has been shown to regulate the expression of *MAML1* and *MEF2C* by sequestration of miR-133 and miR-135 that target the two genes. *MAML1* and *MEF2C* are two transcription factors that activate muscle-specific gene expression, controlling the timing of muscle differentiation. Consistently, downregulation or overexpression of *linc-MD1* resulted in a decreased or increased accumulation of myogenic marker genes in mouse myoblasts, which leads to retardation or acceleration of the muscle differentiation program, respectively [15]. Apart from these individual examples, transcripts of ~7,000 genes have been shown to potentially act as natural miRNA target mimics to regulate the establishment of oncogenic pathways in glioblastoma in human [17]. These results suggest that target mimicry or ceRNA network plays an important role in cell differentiation and tumorigenesis [17,65].

Besides its biological significance, target mimicry has provided an alternative approach for functional characterization of miRNAs. In plants, characterization of gene function has relied largely on the use of genetic knockout mutants caused by T-DNA or transposon insertion. However, because of the small size of *MIRNA* genes and the existence of multiple, highly conserved members in most plant miRNA families, it has been extremely laborious and time-consuming to obtain a corresponding null mutant plant line of a *MIRNA* gene [66]. Target mimicry has therefore been exploited as an alternative approach for functional characterization of miRNAs. The usefulness of this approach has been demonstrated by the closely resembled phenotypes observed in plants transformed with target mimicry constructs and in plants either overexpressing miRNA-resistant targets or harbouring a T-DNA insertion in *MIRNA* genes [21,67]. In animals and human, artificial miRNA sponge, a strategy similar to target mimicry in plants, has been widely used in characterization of miRNA functions [68,69]. In addition, artificial miRNA sponge has also been investigated for potential therapeutic applications in human diseases such as cancer and cardiac disorders associated with miRNA misregulation.

3.2. LncRNA Guiding Recruitment of Chromatin Modifiers

Studies in animals and plants have demonstrated that chromatin modifications are important for tissue-specific gene expression and for genome reprogramming during development [70,71]. Chromatin modifications at a certain locus are believed to be initiated by site-specific recruitment of chromatin modifying complexes. Several lncRNAs, such as *Air*, *HOTAIR*, *Xist* and *Kcnq1ot1*, have been shown to target repressive histone-modifying activities and direct epigenetic silencing through a molecular interaction with specific chromatin domains in animals and human [12,72–77]. In addition, hundreds of lncRNAs have been shown to co-purify with various components of chromatin modifying complexes in co-immunoprecipitation assays in human [49]. In plants, lncRNA-mediated chromatin modification has so far only been demonstrated in the *FLC* locus in *Arabidopsis* [23].

FLC acts as a floral repressor that confers a requirement for vernalization, a process by which certain plants acquire competence to flowering in spring by sensing prolonged exposure to winter cold [78,79]. Molecular studies have shown that both activation and repression chromatin remodelling complexes are involved in the regulation of *FLC* expression [80]. Vernalization induces a Plant HomeoDomain (PHD) finger containing protein, VERNALIZATION INSENSITIVE 3 (VIN3), and promotes association of VIN3 with PRC2 to stably repress the expression of *FLC* [81,82] through PRC2-mediated deposition of H3K27me3 marks at the *FLC* locus. The level of PRC2 occupancy at

FLC is correlated with the level of H3K27me3 and consequently the degree of repression of *FLC* [81,82]. Increased occupancy of PRC2 followed by increased level of H3K27me3 at the *FLC* chromatin is necessary for the stable maintenance of vernalization-induced *FLC* repression. PRC2 is a conserved repressive chromatin modifier [83]. In human, *HOTAIR*, an lncRNA generated from the *HOXC* locus, has been shown to mediate epigenetic changes at the *HOXD* locus *in trans* by recruiting PRC2 [12]. Further studies indicate that interaction between lncRNAs and chromatin modifying complex seems to be a general mechanism for epigenetic silencing in animals [84]. These findings encouraged plant scientists to investigate if lncRNAs are generated from the *FLC* locus and if they play a role in the repression of *FLC* expression.

Two classes of lncRNAs are identified from the *FLC* locus. The first class is *COOLAIR*, including long and short versions of lncRNAs that are transcribed in antisense orientation relative to *FLC* by a promoter located downstream of *FLC*. The expression levels of *COOLAIR* increase during vernalization, and induction of *COOLAIR* by vernalization coincides with a reduction of *FLC* but is earlier than the onset of other vernalization makers, such as *VIN3* [24]. This observation led to the suggestion that *COOLAIR* is involved in early, cold-dependent transcriptional silencing of *FLC* [24]. The nature of antisense orientation between *COOLAIR* and *FLC* and that the long version of *COOLAIR* transcripts extend beyond the transcriptional start site of *FLC* suggests a possible role of *COOLAIR* through transcriptional interference [24]. However, a more recent study, using multiple T-DNA insertion lines across the *FLC* and *COOLAIR*, showed that the transcription of *COOLAIR* is not required for the initial repression of *FLC*; instead the promoter and the first exon of the *FLC* gene are sufficient to initiate *FLC* repression during vernalization [85]. In addition, *COOLAIR* does not physically interact with PRC2 [23].

The second class of lncRNAs, *COLDAIR* that was uncovered by tiling RT-PCR, are transcribed from the first intron of *FLC* in the same direction as *FLC* [23]. Similar to *COOLAIR*, *COLDAIR* is also transiently induced by vernalization, but its peak expression time point is observed later than that of *COOLAIR*. The *COLDAIR* transcript interacts directly with CURLY LEAF (CLF), one of the components of PRC2, and can be co-purified with PRC2, indicating a direct role of *COLDAIR* in the recruitment of PRC2 to *FLC*. Recruitment and deposition of PRC2 at *FLC* increase the level of H3K27me3 at *FLC* chromatin after vernalization [23]. Knockdown of *COLDAIR* using RNAi compromises cold-mediated H3K27me3 enrichment and the vernalization response. In addition, the vernalization-induced repression of *FLC* is not maintained once plants return to warm conditions in the *COLDAIR* knockdown lines. These results together with the observation that the repression of *FLC* cannot be maintained in PRC2 component mutants suggest that *COLDAIR* is required for establishment and maintenance of the stable silencing state of *FLC* [23,86]. These results also suggest that lncRNA-mediated recruitment of PRC2 and gene repression is an evolutionally conserved mechanism in eukaryotes [23].

A growing body of evidence supports the notion that lncRNAs are key regulators of chromatin state through interacting and recruiting chromatin remodelling complexes to specific genomic loci. Several models, by which lncRNAs tether or guide chromatin modifying complexes to their specific destinations, have been proposed [12,84,87]. Meanwhile, genome-wide approaches for isolation of lncRNAs associated with chromatin or chromatin modifiers [50,88] and for identification of lncRNA occupancy [89] have been established. However, the nature and sites of lncRNA-chromatin interaction

are still largely unknown and more studies are required to uncover the exact mechanism(s) controlling the interaction between lncRNAs and chromatin modifying complexes.

3.3. *LncRNA as Molecular Cargo for Protein Re-Localization*

The early nodulin gene *Enod40*, first identified in soybean and *Medicago sativa* ssp. *varia* [90,91], is a plant gene that participates in the regulation of symbiotic interaction between leguminous plants and soil bacteria [91,92]. *Enod40* is rapidly induced by rhizobia in the root pericycle and in the dividing cortical cells of the nodule primordium during the symbiotic interaction [93]. Transgenic approach confirmed a role of *Enod40* in nodulation [94]. *Enod40* is highly conserved among legumes and is also present in various non-legume species, such as rice [95,96]. The *Enod40* transcript lacks long open reading frames, but encodes two short peptides (12 and 24 amino acid residues in soybean; and 13 and 27 amino acid residues in *M. truncatula*) [97,98]. Translation of these two short peptides is directly related to the biological activity of *Enod40* in *M. truncatula* [98]. In Soybean, these peptides were shown to bind specifically to sucrose synthase, suggesting a role of *Enod40* in the regulation of sucrose utilization in nodules [97]. However, two features of the *Enod40* transcript suggest that the general mechanism of action of *Enod40* may be achieved through its RNA molecule rather than the short peptides. Firstly, the *Enod40* RNA is highly structured and contains a highly stable RNA secondary structure. Analysis of *Enod40* transcripts from numerous leguminous species revealed five conserved domains [99] and at least two domains are absolutely conserved in all currently found *Enod40* homologues [95]. Secondly, one of the two short peptides is not always conserved and the highest conservation at the nucleotide level is observed in the region outside the conserved peptides [95]. In addition, the overall configuration of the secondary structure elements in the *Enod40* RNA is more conserved than the ORFs encoding short peptides [95].

The importance of the secondary structure of *Enod40* was demonstrated in *M. truncatula*. Plants transformed with an altered *Enod40*, in which the RNA structural elements were deleted while the proper translation of short peptides was retained, decrease its role in stimulation of cortical cell division and formation of nodules [98]. More importantly, *Enod40* has been shown to directly interact with MtRBP1 (*Medicago truncatula* RNA binding protein 1), a constitutively expressed RNA-binding protein identified by yeast three-hybrid screening, and play a role in re-localization of MtRBP1 from nuclear speckles into cytoplasmic granules during nodulation in *M. truncatula* [22]. This re-localization of MtRBP1 was only observed in *Enod40*-expressing plant cells and was not affected by impaired activity of peptide translation [22], suggesting that the *Enod40* RNA rather than the *Enod40*-encoded short peptides is important for the MtRBP1 re-localization. This study demonstrated that *Enod40*, like *Mei2p* in the fission yeast, is part of the nucleocytoplasmic trafficking machinery [100].

Recently, two small nodulin acidic RNA-binding proteins, MtSNARP1 (*Medicago truncatula* small nodulin acidic RNA-binding protein 1) and MtSNARP2, were also identified to interact with *Enod40* in *M. truncatula* [101]. However, the RNA-binding activity of MtSNARP2 does not seem to be sequence specific because MtSNARP2 is able to bind the entire *Enod40* RNA and synthetic RNA oligos as well. In addition, the exact binding sites in *Enod40* RNA have not yet been determined although the 5' and 3' regions of the *Enod40* transcripts are important for its interaction with MtRBP1 [22].

4. Challenges for Decoding the Functions of lncRNAs

RNA-seq technology provides a powerful tool for unbiased profiling of transcriptomes in complex organisms; however, it is still a long way from reaching the limit of the transcriptome as most lncRNAs are very lowly expressed [47]. Development of new methodologies, such as RNA CaptureSeq [47] and single-cell transcriptome profiling [102], has greatly increased the chance to identify rare transcriptional events taking place in specific genomic regions or cell types of interest. The ongoing development of the direct RNA sequencing technology, in which single RNA molecule is sequenced directly without prior conversion to cDNA, promises reduction of artifacts associated with the current RNA-seq technology during cDNA synthesis [103]. However, further technical innovation and revolution is required to routinely identify rare transcripts in a genome-wide scale in animals and particularly in plants.

Currently, distinguishing lncRNAs from protein-coding mRNAs depends solely on ORF prediction. However, several points need to be considered. First, the accuracy of ORF prediction relies on the completeness of the full-length status of the input sequences. Second, some transcripts such as *Enod40* lack long ORF but do contain short ORF(s) that has potential to encode short peptides [97,98]. For such transcripts, assigning a function to the RNA molecule or to the short ORF(s) is not a trivial task. While evolutionary conservation of amino acid sequences can be used to assess the functionality of short ORFs, and information on the intracellular localization of a RNA transcript also offers clues for its functionality, the ultimate solution should be to assess the activity of the RNA molecule or its encoded short peptide(s) using biochemical approaches [104]. Third, some protein coding RNAs could have additional structural functions that are unrelated to their translation product, as demonstrated by the identification of a class of coding transcripts with a role through their RNA molecules in maintaining a decondensed and biologically active interphase chromatin conformation in human and mice [105]. Thus, when it comes to the functional characterization of single transcripts, the presence of an ORF does not necessarily exclude the existence of additional regulatory functions at the RNA level; and *vice versa* [106].

Despite a number of lncRNAs have been demonstrated to function in guiding chromatin modifying complexes to specific genomic loci, the exact nature of the chromatin binding sites is yet to be determined. A new technique, Chromatin Isolation by RNA Purification (ChIRP)-seq, is suitable for genome-wide characterization of chromatin-binding sites [89]. Applying this technique to three lncRNAs has revealed that lncRNA binding sites resemble transcription factor binding sites in being focal, numerous and sequence-specific [89]. Future works are required to elucidate how a single stranded RNA molecule interacts with a specific double-stranded genomic region.

5. Conclusions

A large number of lncRNAs have been identified in animals. Sequence conservation and tissue-specific expression patterns strongly suggest that these lncRNAs are more than just transcriptional noise. With the increased cases of confirmed functional lncRNAs, it is becoming increasingly evident that lncRNAs play important roles in diverse cellular processes. However, for the majority of lncRNAs their biological functions remain unknown, and their precise mechanisms of action are yet to be determined. In plants, a number of lncRNAs have also been identified by *in silico*

or *de novo* approaches. While functional characterization of plant lncRNAs is still in its infancy, studies so far suggest that they function via similar mechanisms to animal lncRNAs. It can be anticipated that a diverse array of new molecular functions will emerge for plant lncRNAs with increased numbers of new plant lncRNAs being identified and characterized.

Acknowledgements

This study was supported by the CSIRO Transformational Biology Capability Platform.

References

1. Jacquier, A. The complex eukaryotic transcriptome: Unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.* **2009**, *10*, 833–844.
2. Faghihi, M.A.; Wahlestedt, C. Regulatory roles of natural antisense transcripts. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 637–643.
3. Katayama, S.; Tomaru, Y.; Kasukawa, T.; Waki, K.; Nakanishi, M.; Nakamura, M.; Nishida, H.; Yap, C.C.; Suzuki, M.; Kawai, J.; *et al.* Antisense transcription in the mammalian transcriptome. *Science* **2005**, *309*, 1564–1566.
4. He, Y.; Vogelstein, B.; Velculescu, V.E.; Papadopoulos, N.; Kinzler, K.W. The antisense transcriptomes of human cells. *Science* **2008**, *322*, 1855–1857.
5. Yamada, K.; Lim, J.; Dale, J.M.; Chen, H.; Shinn, P.; Palm, C.J.; Southwick, A.M.; Wu, H.C.; Kim, C.; Nguyen, M.; *et al.* Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **2003**, *302*, 842–846.
6. Wilhelm, B.T.; Marguerat, S.; Watt, S.; Schubert, F.; Wood, V.; Goodhead, I.; Penkett, C.J.; Rogers, J.; Bahler, J. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **2008**, *453*, 1239–1243.
7. The ENCODE Project Consortium; Birney, E.; Stamatoyannopoulos, J.A.; Dutta, A.; Guigo, R.; Gingeras, T.R.; Margulies, E.H.; Weng, Z.; Snyder, M.; Dermitzakis, E.T.; *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **2007**, *447*, 799–816.
8. Chitwood, D.H.; Timmermans, M.C. Small RNAs are on the move. *Nature* **2010**, *467*, 415–419.
9. Ghildiyal, M.; Zamore, P.D. Small silencing RNAs: An expanding universe. *Nat. Rev. Genet.* **2009**, *10*, 94–108.
10. Wang, K.C.; Chang, H.Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **2011**, *43*, 904–914.
11. Chaumeil, J.; Le Baccon, P.; Wutz, A.; Heard, E. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* **2006**, *20*, 2223–2237.
12. Rinn, J.L.; Kertesz, M.; Wang, J.K.; Squazzo, S.L.; Xu, X.; Brugmann, S.A.; Goodnough, L.H.; Helms, J.A.; Farnham, P.J.; Segal, E.; *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **2007**, *129*, 1311–1323.

13. Tripathi, V.; Ellis, J.D.; Shen, Z.; Song, D.Y.; Pan, Q.; Watt, A.T.; Freier, S.M.; Bennett, C.F.; Sharma, A.; Bubulya, P.A.; *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **2010**, *39*, 925–938.
14. Poliseno, L.; Salmena, L.; Zhang, J.; Carver, B.; Haveman, W.J.; Pandolfi, P.P. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **2010**, *465*, 1033–1038.
15. Cesana, M.; Cacchiarelli, D.; Legnini, I.; Santini, T.; Sthandier, O.; Chinappi, M.; Tramontano, A.; Bozzoni, I. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **2011**, *147*, 358–369.
16. Karreth, F.A.; Tay, Y.; Perna, D.; Ala, U.; Tan, S.M.; Rust, A.G.; DeNicola, G.; Webster, K.A.; Weiss, D.; Perez-Mancera, P.A.; *et al.* *In vivo* identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell* **2011**, *147*, 382–395.
17. Sumazin, P.; Yang, X.; Chiu, H.S.; Chung, W.J.; Iyer, A.; Llobet-Navas, D.; Rajbhandari, P.; Bansal, M.; Guarnieri, P.; Silva, J.; *et al.* An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* **2011**, *147*, 370–381.
18. Tay, Y.; Kats, L.; Salmena, L.; Weiss, D.; Tan, S.M.; Ala, U.; Karreth, F.; Poliseno, L.; Provero, P.; di Cunto, F.; *et al.* Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* **2011**, *147*, 344–357.
19. Ben Amor, B.; Wirth, S.; Merchan, F.; Laporte, P.; d'Aubenton-Carafa, Y.; Hirsch, J.; Maizel, A.; Mallory, A.; Lucas, A.; Deragon, J.M.; *et al.* Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Res.* **2009**, *19*, 57–69.
20. Wierzbicki, A.T.; Haag, J.R.; Pikaard, C.S. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **2008**, *135*, 635–648.
21. Franco-Zorrilla, J.M.; Valli, A.; Todesco, M.; Mateos, I.; Puga, M.I.; Rubio-Somoza, I.; Leyva, A.; Weigel, D.; Garcia, J.A.; Paz-Ares, J. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **2007**, *39*, 1033–1037.
22. Campalans, A.; Kondorosi, A.; Crespi, M. Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell* **2004**, *16*, 1047–1059.
23. Heo, J.B.; Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **2011**, *331*, 76–9.
24. Swiezewski, S.; Liu, F.; Magusin, A.; Dean, C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* **2009**, *462*, 799–802.
25. Dunoyer, P.; Brosnan, C.A.; Schott, G.; Wang, Y.; Jay, F.; Alioua, A.; Himber, C.; Voinnet, O. An endogenous, systemic RNAi pathway in plants. *EMBO J.* **2010**, *29*, 1699–1712.
26. Henderson, I.R.; Jacobsen, S.E. Epigenetic inheritance in plants. *Nature* **2007**, *447*, 418–424.
27. Allen, E.; Howell, M.D. miRNAs in the biogenesis of trans-acting siRNAs in higher plants. *Semin. Cell Dev. Biol.* **2010**, *21*, 798–804.
28. Lukashin, A.V.; Borodovsky, M. GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res.* **1998**, *26*, 1107–1115.

29. Burge, C.B.; Karlin, S. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* **1998**, *8*, 346–354.
30. Lottaz, C.; Iseli, C.; Jongeneel, C.V.; Bucher, P. Modeling sequencing errors by combining Hidden Markov models. *Bioinformatics* **2003**, *19*, ii103–ii112.
31. Shimizu, K.; Adachi, J.; Muraoka, Y. ANGLE: A sequencing errors resistant program for predicting protein coding regions in unfinished cDNA. *J. Bioinform. Comput. Biol.* **2006**, *4*, 649–664.
32. Jia, H.; Osak, M.; Bogu, G.K.; Stanton, L.W.; Johnson, R.; Lipovich, L. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA* **2010**, *16*, 1478–1487.
33. Badger, J.H.; Olsen, G.J. CRITICA: Coding region identification tool invoking comparative analysis. *Mol. Biol. Evol.* **1999**, *16*, 512–524.
34. Hatzigeorgiou, A.G.; Fiziev, P.; Reczko, M. DIANA-EST: A statistical analysis. *Bioinformatics* **2001**, *17*, 913–919.
35. Mignone, F.; Grillo, G.; Liuni, S.; Pesole, G. Computational identification of protein coding potential of conserved sequence tags through cross-species evolutionary analysis. *Nucleic Acids Res.* **2003**, *31*, 4639–4645.
36. Liu, J.; Gough, J.; Rost, B. Distinguishing protein-coding from non-coding RNAs through support vector machines. *PLoS Genet.* **2006**, *2*, e29.
37. Kong, L.; Zhang, Y.; Ye, Z.Q.; Liu, X.Q.; Zhao, S.Q.; Wei, L.; Gao, G. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **2007**, *35*, W345–W349.
38. Lu, Z.J.; Yip, K.Y.; Wang, G.; Shou, C.; Hillier, L.W.; Khurana, E.; Agarwal, A.; Auerbach, R.; Rozowsky, J.; Cheng, C.; *et al.* Prediction and characterization of noncoding RNAs in *C. elegans* by integrating conservation, secondary structure, and high-throughput sequencing and array data. *Genome Res.* **2011**, *21*, 276–285.
39. Washietl, S.; Findeiss, S.; Muller, S.A.; Kalkhof, S.; von Bergen, M.; Hofacker, I.L.; Stadler, P.F.; Goldman, N. RNAcode: Robust discrimination of coding and noncoding regions in comparative sequence data. *RNA* **2011**, *17*, 578–594.
40. Khachane, A.N.; Harrison, P.M. Mining mammalian transcript data for functional long non-coding RNAs. *PLoS One* **2010**, *5*, e10316.
41. Maeda, N.; Kasukawa, T.; Oyama, R.; Gough, J.; Frith, M.; Engstrom, P.G.; Lenhard, B.; Aturaliya, R.N.; Batalov, S.; Beisel, K.W.; *et al.* Transcript annotation in FANTOM3: Mouse gene catalog based on physical cDNAs. *PLoS Genet.* **2006**, *2*, e62.
42. Li, L.; Wang, X.; Stolc, V.; Li, X.; Zhang, D.; Su, N.; Tongprasit, W.; Li, S.; Cheng, Z.; Wang, J.; *et al.* Genome-wide transcription analyses in rice using tiling microarrays. *Nat. Genet.* **2006**, *38*, 124–129.
43. Matsui, A.; Ishida, J.; Morosawa, T.; Okamoto, M.; Kim, J.M.; Kurihara, Y.; Kawashima, M.; Tanaka, M.; To, T.K.; Nakaminami, K.; *et al.* *Arabidopsis* tiling array analysis to identify the stress-responsive genes. *Methods Mol. Biol.* **2010**, *639*, 141–155.
44. Rehrauer, H.; Aquino, C.; Gruissem, W.; Henz, S.R.; Hilson, P.; Laubinger, S.; Naouar, N.; Patrignani, A.; Rombauts, S.; Shu, H.; *et al.* AGRONOMICS1: A new resource for *Arabidopsis* transcriptome profiling. *Plant Physiol.* **2010**, *152*, 487–499.

45. Cloonan, N.; Forrest, A.R.; Kollé, G.; Gardiner, B.B.; Faulkner, G.J.; Brown, M.K.; Taylor, D.F.; Steptoe, A.L.; Wani, S.; Bethel, G.; *et al.* Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* **2008**, *5*, 613–619.
46. Mortazavi, A.; Williams, B.A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621–628.
47. Mercer, T.R.; Gerhardt, D.J.; Dinger, M.E.; Crawford, J.; Trapnell, C.; Jeddloh, J.A.; Mattick, J.S.; Rinn, J.L. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* **2011**, *30*, 99–104.
48. Graveley, B.R.; Brooks, A.N.; Carlson, J.W.; Duff, M.O.; Landolin, J.M.; Yang, L.; Artieri, C.G.; van Baren, M.J.; Boley, N.; Booth, B.W.; *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **2011**, *471*, 473–479.
49. Khalil, A.M.; Guttman, M.; Huarte, M.; Garber, M.; Raj, A.; Rivea Morales, D.; Thomas, K.; Presser, A.; Bernstein, B.E.; van Oudenaarden, A.; *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11667–11672.
50. Zhao, J.; Ohsumi, T.K.; Kung, J.T.; Ogawa, Y.; Grau, D.J.; Sarma, K.; Song, J.J.; Kingston, R.E.; Borowsky, M.; Lee, J.T. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **2010**, *40*, 939–953.
51. Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M.F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B.W.; Cassady, J.P.; *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **2009**, *458*, 223–227.
52. Zhang, X.; Bernatavichute, Y.V.; Cokus, S.; Pellegrini, M.; Jacobsen, S.E. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* **2009**, *10*, R62.
53. Aung, K.; Lin, S. I.; Wu, C.C.; Huang, Y.T.; Su, C.L.; Chiou, T.J. *pho2*, A phosphate overaccumulator, is caused by a nonsense mutation in a microRNA399 target gene. *Plant Physiol.* **2006**, *141*, 1000–1011.
54. Bari, R.; Datt Pant, B.; Stitt, M.; Scheible, W.R. PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. *Plant Physiol.* **2006**, *141*, 988–999.
55. Chiou, T.J.; Aung, K.; Lin, S.I.; Wu, C.C.; Chiang, S.F.; Su, C.L. Regulation of phosphate homeostasis by microRNA in *Arabidopsis*. *Plant Cell* **2006**, *18*, 412–421.
56. Fujii, H.; Chiou, T.J.; Lin, S.I.; Aung, K.; Zhu, J.K. A miRNA involved in phosphate-starvation response in *Arabidopsis*. *Curr. Biol.* **2005**, *15*, 2038–2043.
57. Pant, B.D.; Buhtz, A.; Kehr, J.; Scheible, W.R. MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. *Plant J.* **2008**, *53*, 731–738.
58. Burleigh, S.H.; Harrison, M.J. A novel gene whose expression in *Medicago truncatula* roots is suppressed in response to colonization by vesicular-arbuscular mycorrhizal (VAM) fungi and to phosphate nutrition. *Plant Mol. Biol.* **1997**, *34*, 199–208.
59. Liu, C.; Muchhal, U.S.; Raghothama, K.G. Differential expression of TPS11, a phosphate starvation-induced gene in tomato. *Plant Mol. Biol.* **1997**, *33*, 867–874.

60. Wasaki, J.; Yonetani, R.; Shinano, T.; Kai, M.; Osaki, M. Expression of the OsPI1 gene, cloned from rice roots using cDNA microarray, rapidly responds to phosphorus status. *New Phytol.* **2003**, *158*, 239–248.
61. Burleigh, S.H.; Harrison, M.J. The down-regulation of Mt4-like genes by phosphate fertilization occurs systemically and involves phosphate translocation to the shoots. *Plant Physiol.* **1999**, *119*, 241–248.
62. Martin, A.C.; del Pozo, J.C.; Iglesias, J.; Rubio, V.; Solano, R.; de La Pena, A.; Leyva, A.; Paz-Ares, J. Influence of cytokinins on the expression of phosphate starvation responsive genes in *Arabidopsis*. *Plant J.* **2000**, *24*, 559–567.
63. Rymarquis, L.A.; Kastenmayer, J.P.; Huttenhofer, A.G.; Green, P.J. Diamonds in the rough: mRNA-like non-coding RNAs. *Trends Plant Sci.* **2008**, *13*, 329–334.
64. Shin, H.; Shin, H.S.; Chen, R.; Harrison, M.J. Loss of At4 function impacts phosphate distribution between the roots and the shoots during phosphate starvation. *Plant J.* **2006**, *45*, 712–726.
65. Salmena, L.; Poliseno, L.; Tay, Y.; Kats, L.; Pandolfi, P.P. A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language? *Cell* **2011**, *146*, 353–358.
66. Eamens, A.L.; Wang, M.B. Alternate approaches to repress endogenous microRNA activity in *Arabidopsis thaliana*. *Plant Signal. Behav.* **2011**, *6*, 349–359.
67. Todesco, M.; Rubio-Somoza, I.; Paz-Ares, J.; Weigel, D. A collection of target mimics for comprehensive analysis of microRNA function in *Arabidopsis thaliana*. *PLoS Genet.* **2010**, *6*, e1001031.
68. Ebert, M.S.; Neilson, J.R.; Sharp, P.A. MicroRNA sponges: Competitive inhibitors of small RNAs in mammalian cells. *Nat. Methods* **2007**, *4*, 721–726.
69. Gentner, B.; Schira, G.; Giustacchini, A.; Amendola, M.; Brown, B.D.; Ponzoni, M.; Naldini, L. Stable knockdown of microRNA *in vivo* by lentiviral vectors. *Nat. Methods* **2009**, *6*, 63–66.
70. Ho, L.; Crabtree, G.R. Chromatin remodelling during development. *Nature* **2010**, *463*, 474–484.
71. Pfluger, J.; Wagner, D. Histone modifications and dynamic regulation of genome accessibility in plants. *Curr. Opin. Plant Biol.* **2007**, *10*, 645–652.
72. Nagano, T.; Mitchell, J.A.; Sanz, L.A.; Pauler, F.M.; Ferguson-Smith, A.C.; Feil, R.; Fraser, P. The air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **2008**, *322*, 1717–1720.
73. Pandey, R.R.; Mondal, T.; Mohammad, F.; Enroth, S.; Redrup, L.; Komorowski, J.; Nagano, T.; Mancini-Dinardo, D.; Kanduri, C. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **2008**, *32*, 232–246.
74. Zhao, J.; Sun, B.K.; Erwin, J.A.; Song, J.J.; Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **2008**, *322*, 750–756.
75. Mercer, T.R.; Dinger, M.E.; Mattick, J.S. Long non-coding RNAs: Insights into functions. *Nat. Rev. Genet.* **2009**, *10*, 155–159.
76. Ponting, C.P.; Oliver, P.L.; Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **2009**, *136*, 629–641.
77. Au, P.C.; Zhu, Q.H.; Dennis, E.S.; Wang, M.B. Long non-coding RNA-mediated mechanisms independent of the RNAi pathway in animals and plants. *RNA Biol.* **2011**, *8*, 404–414.

78. Sheldon, C.C.; Burn, J.E.; Perez, P.P.; Metzger, J.; Edwards, J.A.; Peacock, W.J.; Dennis, E.S. The FLF MADS box gene: A repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* **1999**, *11*, 445–458.
79. Michaels, S.D.; Amasino, R.M. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **1999**, *11*, 949–956.
80. Kim, D.H.; Sung, S. Environmentally coordinated epigenetic silencing of FLC by protein and long noncoding RNA components. *Curr. Opin. Plant Biol.* **2012**, *15*, 51–56.
81. de Lucia, F.; Crevillen, P.; Jones, A.M.; Greb, T.; Dean, C. A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16831–16836.
82. Wood, C.C.; Robertson, M.; Tanner, G.; Peacock, W.J.; Dennis, E.S.; Helliwell, C.A. The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 14631–14636.
83. Schwartz, Y.B.; Pirrotta, V. Polycomb silencing mechanisms and the management of genomic programmes. *Nat. Rev. Genet.* **2007**, *8*, 9–22.
84. Nagano, T.; Fraser, P. Emerging similarities in epigenetic gene silencing by long noncoding RNAs. *Mamm. Genome* **2009**, *20*, 557–562.
85. Helliwell, C.A.; Robertson, M.; Finnegan, E.J.; Buzas, D.M.; Dennis, E.S. Vernalization-repression of *Arabidopsis* FLC requires promoter sequences but not antisense transcripts. *PLoS One* **2011**, *6*, e21513.
86. Heo, J.B.; Sung, S. Encoding memory of winter by noncoding RNAs. *Epigenetics* **2011**, *6*, 544–547.
87. Koziol, M.J.; Rinn, J.L. RNA traffic control of chromatin complexes. *Curr. Opin. Genet. Dev.* **2010**, *20*, 142–148.
88. Mondal, T.; Rasmussen, M.; Pandey, G.K.; Isaksson, A.; Kanduri, C. Characterization of the RNA content of chromatin. *Genome Res.* **2010**, *20*, 899–907.
89. Chu, C.; Qu, K.; Zhong, F.L.; Artandi, S.E.; Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **2011**, *44*, 667–678.
90. Crespi, M.D.; Jurkevitch, E.; Poiret, M.; d'Aubenton-Carafa, Y.; Petrovics, G.; Kondorosi, E.; Kondorosi, A. Enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. *EMBO J.* **1994**, *13*, 5099–5112.
91. Yang, W.C.; Katinakis, P.; Hendriks, P.; Smolders, A.; de Vries, F.; Spee, J.; van Kammen, A.; Bisseling, T.; Franssen, H. Characterization of GmENOD40, a gene showing novel patterns of cell-specific expression during soybean nodule development. *Plant J.* **1993**, *3*, 573–585.
92. Kouchi, H.; Hata, S. Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Mol. Gen. Genet.* **1993**, *238*, 106–119.
93. Compaan, B.; Yang, W.-C.; Bisseling, T.; Franssen, H. Enod40 expression in the pericycle precedes cortical cell division in Rhizobium-legume interaction and the highly conserved internal region of the gene does not encode a peptide. *Plant Soil* **2001**, *230*, 1–8.
94. Charon, C.; Johansson, C.; Kondorosi, E.; Kondorosi, A.; Crespi, M. ENOD40 induces dedifferentiation and division of root cortical cells in legumes. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 8901–8906.

95. Gulyaev, A.P.; Roussis, A. Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. *Nucleic Acids Res.* **2007**, *35*, 3144–3152.
96. Kouchi, H.; Takane, K.; So, R.B.; Ladha, J.K.; Reddy, P.M. Rice ENOD40: Isolation and expression analysis in rice and transgenic soybean root nodules. *Plant J.* **1999**, *18*, 121–129.
97. Rohrig, H.; Schmidt, J.; Miklashevichs, E.; Schell, J.; John, M. Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1915–1920.
98. Sousa, C.; Johansson, C.; Charon, C.; Manyani, H.; Sautter, C.; Kondorosi, A.; Crespi, M. Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. *Mol. Cell Biol.* **2001**, *21*, 354–366.
99. Girard, G.; Roussis, A.; Gulyaev, A.P.; Pleij, C.W.; Spink, H.P. Structural motifs in the RNA encoded by the early nodulation gene enod40 of soybean. *Nucleic Acids Res.* **2003**, *31*, 5003–5015.
100. Yamashita, A.; Watanabe, Y.; Nukina, N.; Yamamoto, M. RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast. *Cell* **1998**, *95*, 115–123.
101. Laporte, P.; Satiat-Jeunemaitre, B.; Velasco, I.; Csorba, T.; van de Velde, W.; Campalans, A.; Burgyan, J.; Arevalo-Rodriguez, M.; Crespi, M. A novel RNA-binding peptide regulates the establishment of the *Medicago truncatula*-*Sinorhizobium meliloti* nitrogen-fixing symbiosis. *Plant J.* **2010**, *62*, 24–38.
102. Islam, S.; Kjallquist, U.; Moliner, A.; Zajac, P.; Fan, J.B.; Lonnerberg, P.; Linnarsson, S. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **2011**, *21*, 1160–1167.
103. Ozsolak, F.; Milos, P.M. RNA sequencing: Advances, challenges and opportunities. *Nat. Rev. Genet.* **2011**, *12*, 87–98.
104. Kageyama, Y.; Kondo, T.; Hashimoto, Y. Coding vs. non-coding: Translatability of short ORFs found in putative non-coding transcripts. *Biochimie* **2011**, *93*, 1981–1986.
105. Caudron-Herger, M.; Muller-Ott, K.; Mallm, J.P.; Marth, C.; Schmidt, U.; Fejes-Toth, K.; Rippe, K. Coding RNAs with a non-coding function: Maintenance of open chromatin structure. *Nucleus* **2011**, *2*, 410–424.
106. Solda, G.; Makunin, I.V.; Sezerman, O.U.; Corradin, A.; Corti, G.; Guffanti, A. An Ariadne's thread to the identification and annotation of noncoding RNAs in eukaryotes. *Brief. Bioinformatics* **2009**, *10*, 475–489.