Briefings in Functional Genomics Advance Access published June 9, 2014

doi:10.1093/bfgp/elu017

RNA regulatory networks in animals and plants: a long noncoding RNA perspective

Youhuang Bai, Xiaozhuan Dai, Andrew P Harrison and Ming Chen

Abstract

A recent highlight of genomics research has been the discovery of many families of transcripts which have function but do not code for proteins. An important group is long noncoding RNAs (lncRNAs), which are typically longer than 200 nt, and whose members originate from thousands of loci across genomes. We review progress in understanding the biogenesis and regulatory mechanisms of lncRNAs. We describe diverse computational and high throughput technologies for identifying and studying lncRNAs. We discuss the current knowledge of functional elements embedded in lncRNAs as well as insights into the lncRNA-based regulatory network in animals. We also describe genome-wide studies of large amount of lncRNAs in plants, as well as knowledge of selected plant lncRNAs with a focus on biotic/abiotic stress-responsive lncRNAs.

Keywords: *lncRNA; miRNA; NAT; RNA regulatory network*

INTRODUCTION

The majority of DNA in the human genome is transcribed but only a small proportion (up to 2%) is covered by sequences known to code for proteins. The bulk of the human genome is not 'junk' DNA but instead exerts a pivotal effect on the regulation of genes [1, 2]. Pervasive transcription appears to be prevalent in different genomes ranging from animals to plants. The great mass of transcripts from unexpected regions, such as introns and intergenic parts of the genome, is not known to function as templates for protein synthesis and is termed noncoding RNAs (ncRNAs). Many different types of ncRNAs have been identified, including microRNAs (miRNAs), small interfering RNAs (siRNAs), small nucleolar RNAs, transfer RNAs, ribosomal RNAs and long noncoding RNAs (lncRNAs).

The wide range of different types of lncRNAs is changing our previous knowledge about gene concepts. According to their proximity to protein-coding genes in the genome, lncRNAs can be mainly grouped into the following three categories: (i) intergenic lncRNA (lincRNA) without any overlap with other loci, (ii) long intronic overlap that overlaps with intronic region of any other loci, and (iii) antisense exonic lncRNA that overlaps with exon(s) of other loci on the opposite strand [3].

In addition to Pol I–III, there are two structurally and functionally distinct plant-specific RNA polymerases (Pol IV and V). In *Arabidopsis thaliana* two

Corresponding author. Ming Chen, Department of Bioinformatics, State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou, 310058, People's Republic of China. Tel./Fax: +86-(0)571-88206612; E-mail: mchen@zju.edu.cn

Youhuang Bai is a PhD student in Ming Chen's laboratory in Zhejiang University. His research focuses on integrated approaches for identification of long noncoding RNA in plants and *in silico* analysis of noncoding RNA-mediated regulatory networks.

Xiaozhuan Dai is a Masters student in Ming Chen's laboratory in Zhejiang University. Her research focuses on secondary structure and evolution of long noncoding RNA in animals and plants.

Andrew Harrison is a Senior Lecturer in the Department of Mathematical Sciences, University of Essex. His research concentrates on identifying biases in scientific datasets that are available in the public domain.

Ming Chen is a Professor in the Department of Bioinformatics, College of Life Sciences, Zhejiang University. His current research focuses on construction of small RNA-mediated regulatory networks and establishing useful web servers and platforms to help biologists browse and analyze massive biological datasets.

lncRNAs—*AtR8* and *AtR18*, transcribed by RNA polymerase III (Pol III), were identified by efficient *in vitro* transcription in tobacco nuclear extracts [4]. Other studies have revealed that a subset of lncRNAs are the product of Pol IV and/or Pol V [5, 6]. Pol V-dependent long intergenic ncRNAs can be detectable by reverse transcription-PCR (RT-PCR) [7].

Many lncRNAs have been found to be natural antisense transcripts (NATs), a group of endogenous RNA molecules containing sequences that are complementary to other transcripts [8, 9]. NATs can be grouped into two categories, *cis*-NATs and *trans*-NATs. *Cis*-NAT pairs are transcribed at the same genomic locus from opposite DNA strands, whereas *trans*-NAT pairs are transcribed from different loci and are partially complementary. Although underlying mechanisms are largely unknown, the fact that the fraction of antisense transcripts is quite large, and do not encode for proteins, suggests that NATs play important regulatory roles in many aspects of gene regulation, including genomic imprinting, transcriptional interference, RNA masking, RNA editing, RNA interference (RNAi) and translational regulation [9–11].

Several genome-wide studies have identified lncRNAs expressed in many organisms (including human, mouse, zebrafish and *Drosophila melanogaster*) [12–16]. Here, we review the striking progress in understanding the biological impact of lncRNAs. We describe the computational and high throughput technologies for identifying and studying lncRNAs, as well as the current knowledge of the functional elements embedded in lncRNAs, and the lncRNA-based regulatory networks. Furthermore, we review and summarize the knowledge about the large amount of lncRNAs in plants, describe selected plant lncRNAs functional studies with a particular emphasis on biotic/abiotic stress-responsive lncRNAs.

GENOME-WIDE STUDIES OF lncRNAS

Recent advances in technology such as tiling arrays and deep sequencing have led to the discovery of many classes of ncRNAs. Re-annotation of microarrays by bioinformatics approaches has also led to the identification of novel lncRNAs in human and mouse [17]. Analysis of RNA-seq and ChIP-seq data provides further information about unexpected transcripts without protein-coding potential [13, 18].

Evidence for functional roles of lncRNAs

There has been a surge in studies establishing the diverse impacts of lncRNAs [7–10]. Although many lncRNAs have been identified by different computational and experimental methods, there are extensive debates about their function in the cell [19, 20]. Roughly half of the lncRNA candidates that were recently shown to be required for maintenance of pluripotency contain regions with high translation efficiency, comparable to protein-coding genes [21, 22]. But none of lncRNA candidates was shown to have clear translation efficiency or codon bias according to proposed standards by Guttman *et al.* [23].

The evolutionary rate of protein-coding genes shows a universal negative correlation with expression: highly expressed genes are, on average, more conserved during evolution than genes with lower expression levels [24]. It has also been found that the universal dependency between evolution and expression holds true for lincRNA genes and is comparable in magnitude to the anticorrelation detected for protein-coding genes. *Sphinx* is a newly evolved lncRNA involved in the regulation of male courtship behavior of *Drosophila melanogaster* [25]. It has been reported that the *sphinx* gene was formed by the insertion of a retroposed sequence of the ATP synthase F-chain gene from chromosome 2 into the 102F region of chromosome 4, recruiting sequences upstream to form a new exon and intron [26]. The rate of evolution for this young gene *sphinx* is significantly above neutral expectations, suggesting rapid adaptive evolution.

It is worthy to note that a specific kind of DNA fragments, termed as ultra-conserved elements (UCEs), are located in longer than 200 bp noncoding regions. These UCEs are absolutely conserved between orthologous regions of the human, rat and mouse genomes [27]. Despite often being noncoding DNA, the UCEs have been found to be transcriptionally active and are proposed to be involved in diverse biological processes.

Sequence composition of lncRNAs

Several methods have been used to distinguish between protein-coding and noncoding sequences. There is low GC content in lncRNAs compared with protein-coding transcripts [18]. By analyzing

204 lncRNAs in the functional lncRNA database, researchers have revealed significant similarities between the lncRNAs and the 3′-untranslated regions (3′-UTRs) of mRNAs both in structural features and in sequence composition [28].

Some of the earliest bioinformatics analysis employed BLAST to search ortholog of all three frame translated peptides deduced from novel transcripts [29, 30]. Coding Potential Calculator is a support vector machine (SVM)-based method, using information on open reading frame (ORF) and ortholog relationships to known protein-coding transcripts [31]. This computational identification mainly depends on the process of BLAST which searches the most likely coding region against comprehensive database of known proteins, such as Swiss-Prot and nonredundant database. It is therefore usually very time-consuming process for the prediction of hundreds of thousands of sequences in a genome-wide manner. To overcome such disadvantages, additional tools attempt to evaluate the coding potential of genomic areas by analyzing only the primary sequence. iSeeRNA employed 10 features related to conservation, ORF and nucleotide sequence-based information to separate intergenic lncRNAs from protein-coding transcripts [32]. Coding-Potential Assessment Tool (CPAT) uses a logistic regression model, built with features extracted from ORF and hexamer usage bias [33]. Both iSeeRNA and CPAT allow users to submit sequences and receive the prediction results almost instantly. In addition, Coding–NonCoding Index (CNCI) software was developed to effectively distinguish between protein-coding and noncoding sequences independent of known annotations using codon usage frequency information by profiling adjoining nucleotide triplets (ANT), as well as information on ORF [34]. Firstly, an ANT matrix was built by analyzing the usage frequency of ANT in coding sequences and ncRNA sequences. Then, sequence-score of each transcript in six reading frames was calculated based on the ANT score matrix. Finally, the sequence-score, combined with other information on the sequences, was utilized to build an SVM model to distinguish protein-coding sequences from the noncoding sequences. In vertebrates, prediction of coding potential heavily relies on evolutionary signatures produced from multiple sequence alignment, as in 'phylogenetic codon substitution frequencies' (PhyloCSF) and RNAcode softwares [35, 36].

Functional elements embedded in lncRNAs

More and more researchers have recognized the pivotal regulatory role of lncRNA. However much less is known about the regulatory elements in lncRNAs compared with protein-coding genes. Until now, three types of element have been described.

miRNA binding site

A muscle-specific lncRNA *linc-MD1* interacts with two miRNAs (*miR-133* and *miR-135*) to regulate the expression of *MAML1* and *MEF2C* (Figure 1A), which encode two transcription factors that activate muscle-specific gene expression [37]. *LincRNA-p21* is

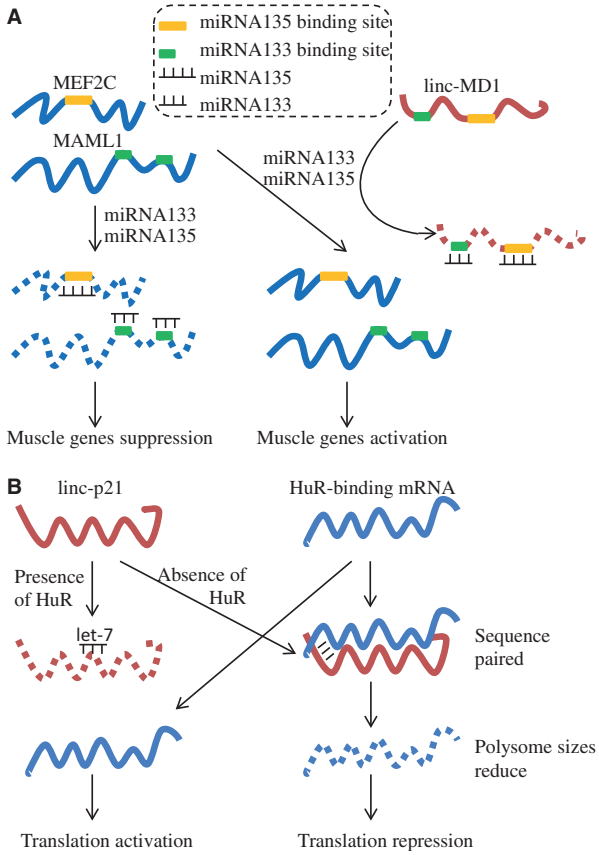

Figure 1: lncRNA-mediated regulatory network. (A) *Linc-MD1* sponges *miR-133* and *miR-135* to regulate the expression of *MAML1* and *MEF2C* transcription factors that activate muscle-specific gene expression. (B) The lncRNA *lincRNA-p21* coordinates expression of HuR-associated mRNA in HuR-dependant manner. In the presence of HuR, *lincRNA-p21* is unstable through the recruitment of *let-7*/Ago2, whereas in the absence of HuR, *lincRNA-p21* is stable and accumulates. Thus, the association of *lincRNA-p21* with other HuR-associated *CTNNB1* and *JUNB* mRNAs leads to their translation.

targeted by miRNA *let-7*, and associated with RNA-binding protein (RBP) HuR, leads to the instability [38]. Then Rck promotes the association of *lincRNA-p21* with *CTNNB1* and *JUNB* mRNAs, repressing their HuR-dependent translation activation through a mechanism that includes reduced polysome sizes.

Transposable element

Antisense *Uchl1* (FANTOM2 clone number: 64305 96G22) is a 5′ head-to-head antisense lncRNA of the *Uchl1* gene, which encodes a neuron-restricted protein associated with Parkinson's disease and Alzheimer's disease. Antisense *Uchl1*-mediated up-regulation of UCHL1 expression at a post-transcriptional level requires the repetitive sequence SINEB2 instead of the adjacent Alu element [39]. Antisense *Uchl1* localization can be regulated by the mTOR pathway, and its cytoplasmic level correlates with the expression of UCHL1 protein.

Highly structured motif

The 3′-ends of lncRNAs *MALAT1* and the *MEN β* are protected from 3′–5′ exonucleases by highly conserved U- and A-rich motif which are predicted to form an triple helical structures [40, 41]. This triple helix structure strongly promotes both RNA stability and translation, despite the absence of a poly(A) tail.

lncRNA–protein and lncRNA–chromatin interactions

High-throughput identification of RNA–DNA and RNA–protein interactions has facilitated the progress for exploring the mechanisms governing ncRNA–chromatin interactions. Ribonucleoprotein (RNP) immunoprecipitation followed by high-throughput sequencing (RIP-seq) has recently been developed to discover RNA transcripts that interact with a specific protein or protein complex. Co-immunoprecipitation of lncRNAs involves immunoprecipitation of a protein from cross-linked cell lysate followed by reverse-cross-linking, isolation, and deep sequencing of RNAs, leading to the identification of all lncRNAs that are associated with a specific protein complex [42]. By using a modified RIP-seq method without cross-linking, two analyses have identified that vast numbers of lncRNAs are capable of directly associating with several chromatin-modifying complexes including PRC2, CoREST and SMCX [43, 44].

Simon D *et al.* had developed CHART (capture hybridization analysis of RNA targets), a hybridization-based purification strategy that can be used to map the genomic binding sites for endogenous RNAs, which is analogous to chromatin immuno precipitation (ChIP) for proteins [45]. CHART is a new technique to purify lncRNAs together with their targets (proteins and DNA fragments), in order to determine the genome-wide localization of a specific lncRNA in chromatin as well as the protein content by Western blot analysis. CHART was successfully applied to lncRNAs of different lengths from human and fruit fly. At the same time, Chu *et al.* had developed chromatin isolation by RNA purification to allow high-throughput discovery of DNA–RNA–protein interactions. In this method, specific lncRNAs bound with protein(s) and DNA sequences are retrieved by tiling oligonucleotides, and followed by high throughput sequencing [46].

Localization

A method termed as 'combined knockdown and localization analysis' has been recently developed [47]. lncRNAs can be targeted by designed endoribonuclease prepared siRNAs (esiRNAs), which have been proven to be particularly suitable for RNAi screening as they efficiently deplete the target transcript without causing prominent off-target effects owing to their inherent complex pools of siRNAs. Simultaneously, specific riboprobes are generated to determine lncRNAs localization in cells. Hundreds of mouse lncRNAs were experimentally studied and about 80% probes have detected diverse localization patterns at varying expression levels for different lncRNAs, suggesting so diverse roles of lncRNAs in different biological processes [47].

Databases

Databases providing information on lncRNAs can be divided into four types, according to their information content: sequence, expression, regulation and association. These databases (i.e. Rfam, fRNAdb, NONCODE) contain various kinds of ncRNAs, including a large amount of lncRNAs [48, 49]. In addition to these broad RNA warehouses, many lncRNA specific databases are emerging for different sources since the last two years. The lncRNAdb is a database providing comprehensive annotations of 194 eukaryotic lncRNAs, and their biological functions have been experimentally verified [50].

Recently, the Functional LncRNA Database has been published, containing 204 well-studied lncRNAs and their splicing variants manually culled from the literatures [28]. A public repository Noncoding RNA Expression Database provides gene expression information for thousands of long ncRNAs in human and mouse [51]. PLncDB contains comprehensive information related to *Arabidopsis* lncRNAs, such as genomic information, expression profiles, siRNA information and associated epigenetic markers [52].

The lncRNA regulation mediated by the miRNAs and transcription factors is being studied. The miRcode database offers a map of putative miRNA target sites in the long noncoding transcriptome, as well as protein-coding genes by using the TargetScan program [53]. In the 2013 database issue of *Nucleic Acids Research*, an integrated database ChIPbase was published to provide transcription factor binding maps and information on transcriptional regulations of coding genes and ncRNAs derived from hundreds of ChIP-Seq data sets [54]. In the same issue, another paper descripted the LncRNADisease database, containing more than 480 lncRNA-disease related entries and 475 lncRNA interaction entries, covering 208 lncRNAs and 166 diseases derived from ~500 publications [55]. The interaction between few well-studied lncRNAs (such as *XIST*, *H19*) and other molecules is available from the NPInter database [56].

lncRNA-MEDIATED REGULATORY NETWORK

The emerging picture of transcriptional and post-transcriptional regulation is that an extremely rich landscape of diverse RNAs is transcribed by a large fraction of the genome in a spatiotemporally dependent manner. The regulation of lncRNA may be similar to that of protein-coding transcripts, which is highly regulated by small ncRNAs, such as miRNA. At the same time, there are NAT annotation databases in both human and plant genomes [57, 58]. By combining the information of miRNA-target relationship and of NAT regulation between mRNA and lncRNA transcripts, a comprehensive RNA regulatory network can be obtained, providing convincing evidence for the existence of a layer of lncRNA-based regulation of gene expression.

lncRNAs play pivotal roles in transcriptional and post-transcriptional regulation and the influence among RBPs, lncRNAs and miRNAs is beginning to emerge [38]. RBP HuR is good example of a node in the lncRNA-based regulatory network. HuR associates with many mRNAs, which can influence cell proliferation, survival, carcinogenesis and stress-/immune-responses [59, 60]. In the presence of HuR, *lincRNA-p21* is unstable through the recruitment of *let*-7/Ago2 (Figure 1B). HuR promotes the translation of targets *CTNNB1* and *JUNB* mRNAs by favoring their association with polysomes. In the absence of HuR, *lincRNA-p21* is stable and accumulates, and Rck promotes the association of *lincRNA-p21* with *CTNNB1* and *JUNB* mRNAs, repressing their translation through a mechanism that includes reduced polysome sizes (Figure 1B). Therefore HuR-dependent translation requires rapid degradation of *lincRNA-p21* in order to prevent the recruitment of translation repressors onto target mRNAs. Similar regulation mechanisms may affect other mRNAs whose translation is increased by HuR [61].

Both HuR and *lincMD1* are involved in muscle differentiation and are under the repressive control of *miR-133*. Legnini *et al.* identified HuR as another component of the *lincMD1*-regulated circuitry, which binds *linc-MD1* and protects it from Drosha cleavage at the expense of *miR-133b* biogenesis [62]. Thus, these findings help to point out an established positive feed forward control among miRNA, lncRNA and products of mRNA.

GENOME-WIDE IDENTIFICATION OF lncRNA IN PLANTS

It has been reported that thousands of distinct noncoding regions result in transcription units in *Arabidopsis* [63, 64]. There are more than 6000 intergenic lncRNAs and 37 238 sense–antisense transcript pairs in *Arabidopsis* discovered from analyzing tiling array, other expression arrays and RNA-seq [65, 66]. These results showed that about 70% of *Arabidopsis* protein-coding genomic loci retain antisense transcripts with no coding potential. Expression profiling of antisense transcripts indicates that organ-specific or condition-specific lncRNAs are under specific regulation, suggesting that they are not generated by spurious transcriptional noise [66].

A computational method was developed to identify endogenous miRNA target mimic sites within

long intergenic noncoding gene for 20 conserved miRNAs in *Arabidopsis* and rice [67]. Using strand-specific RNA sequencing, Song *et al.* applied a method based on model comparison (NASTI-seq) to identify *cis*-NAT pairs, which lead to an increase in the number of known *cis*-NAT pairs in *Arabidopsis* by more than 60% [68].

Intermediate-size ncRNAs (im-ncRNAs) are defined as RNA transcripts with a length ranging from 50 to 300 nt. They are particularly difficult to study experimentally due to their stable secondary structures, lack of a poly (A), and low efficiency of reverse transcription. In *A. thaliana*, analysis of 521 novel im-ncRNAs has shown that these im-ncRNAs are independently regulated components and mostly evolutionary divergent. Similarly, by 454 deep sequencing, 754 novel im-ncRNAs were identified in *Oryza sativa* [69]. The chromosome location of im-ncRNAs is without strand bias and chromosome bias.

Transcriptome analysis of early developing maize seed has identified more than 1000 lncRNA candidates. About 45% of the lncRNAs showed above 2-fold expression change between embryo and endosperm, indicating that a part of these candidate lncRNAs are tissue-preferentially expressed [70]. Global patterns of allelic gene expression in developing maize endosperms from reciprocal crosses between inbreds B73 and Mo17 revealed that 38 lncRNAs expressed in the endosperm as well as more than 179 protein-coding genes are imprinted [71]. Very recently, Li *et al.* integrated RNA-seq datasets from 30 different experiments to identify 1704 high-confidence lncRNAs and revealed that maize lncRNAs are less affected by *cis*- than by *trans*-genetic factors by expression quantitative locus mapping [72].

Transcriptome reconstruction by RNA sequencing in foxtail millet has identified more than 500 lncRNAs [73] and demonstrated sequence features similar to those observed for lncRNA in animals. An increasing amount of papers seek to identify the novel intergenic transcribed region and antisense transcription by using RNA sequencing technology in plants [74, 75].

EXAMPLES OF lncRNA FUNCTIONS IN PLANTS

In comparison to animals, the studies of lncRNA in plants are just beginning to emerge. Several lncRNAs have been identified by the genome-wide screens of cDNA libraries, tiling arrays and RNA-seq data in plants (for more details see next section below) [65, 76–78].

PHO1 plays an important role in phosphate homeostasis of plant [79]. Although *Os-PHO1;2* is likely the functional ortholog of *At-PHO1*, an interesting feature distinguishing all three rice *PHO1* genes from their *Arabidopsis* homologs is the presence of *cis*-NATs associated with all of them [80]. NAT of *Os-PHO1;2* increased its expression in phosphate-deficient condition and leads to an increase of PHO1;2 protein concentration without changes of the levels of expression, sequence, or nuclear export of *PHO1;2* mRNA [81]. Polysome profiles revealed that both the sense *PHO1;2* and the antisense NAT shifted toward the translationally active polysomes, leading to new insights into how lncRNA enhances protein

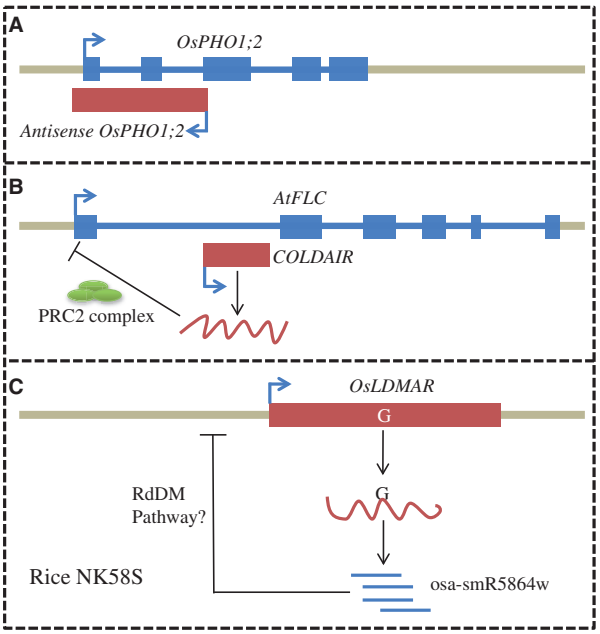

Figure 2: Schematic representation of mechanism of different kinds of lncRNAs in plants. (A) Antisense of *PHO1;2* regulates the expression of the sense of *PHO1;2* transcript in phosphate-deficient condition. (B) The lncRNA *COLDAIR* is transcribed from the intronic region of *FLC* locus, and functions to recruit PRC2 to induce histone modifications at the *FLC* locus, reducing the level of *FLC* expression. (C) *LDMAR* regulates the transcription of itself by producing an lncRNA. *LDMAR* is normally expressed in NK58N, whereas the C-to-G mutation altered the secondary structure of *LDMAR* and leads to DNA methylation of the promoter region through the RdDM pathway in NK58S.

expression independently of mRNA level (Figure 2A).

Another lncRNA, called *IPS1,* is induced upon phosphate starvation and accumulates into the roots and shoots in plants. *IPS1* has complementarity to *miR-399*, but contains a mismatched loop that renders it uncleavable when *miR-399* binds. lncRNA *IPS1* is not cleaved but instead sequesters *miR-399*, leading to increased expression of *miR-399* targets including *PHO2* transcript [82]. This phenomenon in plants is called 'miRNA mimic', which is also found in mammalian genomes. These transcripts are named as 'competing endogenous RNA' [83, 84].

A third example, the lncRNA *COLDAIR* (COLD ASSISTED INTRONIC NONCODING RNA), is related to vernalization. Vernalization is an environmentally induced epigenetic switch, during which winter cold triggers epigenetic silencing of floral repressors and thus provides competence to flower in spring. In *Arabidopsis*, winter cold triggers enrichment of histone modification H3K27me3 at chromatin of the floral repressor, *FLOWERING LOCUS C* (FLC), and results in the epigenetically stable repression of *FLC* [85]. This epigenetic change is mediated by an evolutionarily conserved repressive complex, polycomb repressive complex 2 (PRC2). In *A. thaliana*, the long intronic ncRNA *COLDAIR* was identified to confer vernalization-mediated epigenetic repression by physically associating with a component of PRC2 and targets PRC2 to FLC [86] (Figure 2B).

It is well known that photoperiod- and temperature-sensitive genic male sterile lines (PGMS and TGMS) have made a great contribution to two-line hybrid breeding in rice since 1990s. Recently, Ding *et al.* have identified that PGMS in rice is regulated by a lncRNA of 1236 bases, termed as long-day-specific male-fertility-associated RNA (LDMAR), which is required for normal pollen development of plants under long-day conditions [87]. Meanwhile, Zhou *et al.* have independently revealed that a P/TMS12-1 locus encodes a unique ncRNA, which generates a 21 nt small RNA osa-smR5864w (Figure 2C), conferring PGMS in the japonica rice line Nongken 58S (NK58S) and TGMS in the indica rice line Peiai 64S [88]. Although these two publications have independently identified two different ncRNA molecules, both researches reported that a C-to-G nucleotide transversion took place at position 22258571 of chromosome 12 in the primitive male sterility trait. Though recent papers provide a starting point for figuring out how lncRNA is responsible for the male fertility regulation, their precise mechanisms are still unclear. So in this regard, it is most likely that the primary transcript *LDMAR* may be processed first into the 136 nt intermediate precursor and then into osa-smR5864w [89]. It is suggested that the RNA-directed DNA methylation (RdDM) is involved in regulating photoperiod-sensitive male sterility in rice. A spontaneous mutation causing a single nucleotide polymorphism (SNP) between the wild-type and mutant alters the secondary structure of *LDMAR*. Moreover, many human SNPs located on lncRNAs are found to be associated with prostate cancer risk [90]. The SNP density in regions of lncRNA was similar to that in protein-coding regions, but they were less polymorphic than surrounding regions. It was also reported that SNP in lncRNA *Igf2as* associates with increased muscle damage (strength loss and soreness) after eccentric exercise, similar to SNPs found in protein-coding gene *IGF2* [91].

It is suggested that some lncRNA sequences are precursors to short ncRNA in human and other genomes. By analyzing the full length cDNA library of *A. thaliana*, 76 novel lncRNAs were identified, with nine non-protein coding RNAs (npcRNAs) representing precursors of distinct type of small RNA, such as miRNA, tasiRNA and 24 nt siRNA [77, 78]. RT-PCR analysis demonstrated that *npc83*, the *miR869a* precursor, overaccumulates in *dcl4* mutant, suggesting that *miR869a* is a young miRNA gene [78].

lncRNA *CsM10* was isolated in *Cucumis sativus* and showed differential expression patterns in different tissues, developmental stages and photoperiods. In maize, the putative lncRNA *Zm401* is expressed specifically in pollen. Genetic studies show a function for *Zm401* in regulating the expression of critical genes necessary for pollen development including *MZm3-3* (up-regulated), *ZmMADS2* and *ZmC5* (down-regulated). The lncRNA *Enod40* directed the re-localization of *MtRBP1* (*Medicago truncatula* RBP 1) from the nucleus to cytoplasmic granules during specific stages of legume (*M. truncatula*) root nodule organogenesis.

It must also be noted that viroids are a class of sub-viral plant-pathogenic lncRNAs, composed of a single-stranded, circular molecule with a size in the range of 246–400 nt. Because viroids need to subvert the pathways regulating the lncRNA compartmentalization after host entry, understanding the

mechanism of the viroids compartmentalization may provide insights into the traffic of both foreign and endogenous lncRNA into different organelles in cell.

lncRNAs are also present in telomerase which is an RNP reverse transcriptase. Telomerase contains two essential lncRNA subunits *TER1* and *TER2* that are essential for telomere repeat synthesis in *A. thaliana*. *TER1* is required for telomere maintenance and provides the major template for telomerase [92], whereas *TER2* is assembled in distinct RNP complex to modulate telomerase in response to DNA damage [93].

RESPONES TO BIOTIC/ABIOTIC STRESSES

Increasing evidence points to the fact that lncRNAs play important roles in the regulation of gene expression within biotic/abiotic stress responses. As mentioned above, *COLDAIR* expression is induced by exposure to a cold exposure environment. Also, the Pol III-transcribed lncRNA *AtR8* is conserved among *Arabidopsis* and Brassicaceae and expression of *AtR8* responds negatively to hypoxic stress [4]. Another example *TER2* is a DNA damage-induced lncRNA that works in concert with the *TER1* to promote genome integrity in *Arabidopsis* [93].

In *Arabidopsis*, more than a thousand intergenic lncRNAs have been found to be significantly altered after drought, cold, high-salt and/or abscisic acid treatments. Treatment by elf18 (EF-Tu), which triggers pathogen-associated molecular pattern responses, could also increase the expression level of one of the representative stress responsive lincRNAs [65]. It was reported that abiotic stress altered the accumulation of 22 npcRNAs among the 76 npcRNAs identified through genome-wide analysis of full-length cDNAs [78].

In rice, mining of strand-specific RNA-seq data identified thousands of antisense transcripts, including 84, 74 and 128 *cis*-NAT pairs related with drought, salt and cold stresses, respectively [94]. A reference annotation-based transcript assembly was generated using RNA-seq coupled with a comprehensive time-course experiment, identifying 438 unannotated loci that were differentially expressed under Pi starvation. Several new loci encode small proteins with no homology to known proteins and are enriched in the nonpolysomal fraction, suggesting that these phosphate-responsive transcripts are likely to be ncRNA [95].

Twenty *Fusarium oxysporum*-responsive lncRNAs, from 159 novel intergenic transcriptional regions, were identified using a strand-specific RNA sequencing approach in *Arabidopsis* [96]. Co-induction of multiple neighboring protein-coding genes and lncRNAs suggested that some fungal-responsive promoter elements may contribute to regulation of mRNA and *F. oxysporum*-responsive lncRNA during the transcriptomic changes in response to *F. oxysporum* infection [96].

Drought-regulated expression patterns were shown in two distinct NATs of *Si003758m* (a homolog of *Arabidopsis SAG21*, involving oxidative stress tolerance) and *Si038715m* (a hydroxyproline-rich glycoprotein family protein, involving the process of defense response to pathogen attack) [97]. Another study identified 125 putative stress (powdery mildew infection and heat stress) responsive lncRNAs in wheat [98].

CONCLUSIONS

Even though sustained efforts have been made to characterize lncRNAs by Sanger sequencing, histone mark ChIP-seq, or (non)strand-specific RNA-seq, the current annotation of lncRNAs is likely to be far from complete, especially in plants. Emerging functions for lncRNAs are their contribution to various transcriptional and post-transcriptional gene regulation, epigenetic regulation, and response to different environmental conditions, such as biotic/abiotic stresses. In the future, more discoveries will probably go far beyond the identification of new mechanisms of lncRNAs that regulate gene expression during developmental and differentiation processes, and will likely continue to expand knowledge about the biological significance of lncRNA in a more comprehensive manner.

The complex organization of lncRNA loci and protein-coding genes on chromosomes leads to the sophisticated transcriptional regulation of both lncRNA and mRNA in RNA regulatory networks with some specific modules enriched for different biological function. At the same time, one main regulatory role of lncRNA is that they can be miRNA sponge, because they contain different miRNA binding sites. Combined with the known

relationship between protein-coding transcripts, the lncRNA-mediated RNA regulatory network covers the most important biological processes. Therefore, it is important to predict the functional relevance of lncRNAs from the comprehensive genome-scale network of lncRNA-mediated interactions.

Key Points

- lncRNAs play widespread roles in transcriptional and post-transcriptional gene regulation, epigenetic regulation, and response to different environmental stress in diverse species.
- Next generation sequencing, combined with advanced computational methods, provides powerful capabilities for identifying lncRNAs and inferring their functions in a genome-wide manner.
- Constructing and mining the lncRNA-mediated regulatory networks in both animals and plants facilitates a more comprehensive understanding of lncRNAs functions.

Acknowledgements

The authors thank Dr Christian Klukas, Guy J. Baudoux and Zeeshan Gillani for critical reading of the manuscript. They also thank the anonymous reviewer for many valuable suggestions to improve this manuscript.

FUNDING

This work was supported by the National Natural Sciences Foundation of China (No. 31371328, 30971743), National Science and Technology Project of China (No. 2008AA10Z125, 2009DFA32030), the Fundamental Research Funds for the Central Universities, and the Program for Innovative Research Team in University.

References

1. Carninci P, Kasukawa T, Katayama S, *et al*. The transcriptional landscape of the mammalian genome. *Science* 2005; **309**:1559–63.
2. Katayama S, Tomaru Y, Kasukawa T, *et al*. Antisense transcription in the mammalian transcriptome. *Science* 2005;**309**: 1564–6.
3. Huang Y, Liu N, Wang JP, *et al*. Regulatory long noncoding RNA and its functions. *J Physiol Biochem* 2012;**68**: 611–8.
4. Wu J, Okada T, Fukushima T, *et al*. A novel hypoxic stress-responsive long non-coding RNA transcribed by RNA polymerase III in *Arabidopsis*. *RNA Biol* 2012;**9**:302–13.
5. Rowley MJ, Bohmdorfer G, Wierzbicki AT. Analysis of long non-coding RNAs produced by a specialized RNA polymerase in *Arabidopsis thaliana*. *Methods* 2013;**63**:160–9.
6. Zheng Q, Rowley MJ, Bohmdorfer G, *et al*. RNA polymerase V targets transcriptional silencing components to promoters of protein-coding genes. *Plant J* 2013;**73**:179–89.
7. Wierzbicki AT, Haag JR, Pikaard CS. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 2008;**135**:635–48.
8. Lapidot M, Pilpel Y. Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep* 2006;**7**:1216–22.
9. Faghihi MA, Wahlestedt C. Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol* 2009;**10**:637–43.
10. Borsani O, Zhu J, Verslues PE, *et al*. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* 2005;**123**:1279–91.
11. Li Q, Su Z, Xu X, *et al*. AS1DHRS4, a head-to-head natural antisense transcript, silences the DHRS4 gene cluster in *cis* and *trans*. *Proc Natl Acad Sci USA* 2012;**109**:14110–5.
12. Guttman M, Amit I, Garber M, *et al*. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 2009;**458**:223–7.
13. Young RS, Marques AC, Tibbit C, *et al*. Identification and properties of 1,119 candidate lincRNA loci in the *Drosophila melanogaster* genome. *Genome Biol Evol* 2012;**4**: 427–42.
14. Cabili MN, Trapnell C, Goff L, *et al*. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 2011;**25**: 1915–27.
15. Guttman M, Garber M, Levin JZ, *et al*. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol* 2010;**28**:503–10.
16. Derrien T, Johnson R, Bussotti G, *et al*. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012;**22**:1775–89.
17. Liao Q, Liu C, Yuan X, *et al*. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res* 2011;**39**: 3864–78.
18. Ulitsky I, Shkumatava A, Jan CH, *et al*. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 2011;**147**:1537–50.
19. Pang KC, Frith MC, Mattick JS. Rapid evolution of non-coding RNAs: lack of conservation does not mean lack of function. *Trends Genet* 2006;**22**:1–5.
20. Ponting CP, Hardison RC. What fraction of the human genome is functional? *Genome Res* 2011;**21**:1769–76.
21. Guttman M, Donaghey J, Carey BW, *et al*. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 2011;**477**:295–300.
22. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 2011;**147**: 789–802.
23. Guttman M, Russell P, Ingolia NT, *et al*. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 2013;**154**:240–51.
24. Managadze D, Rogozin IB, Chernikova D, *et al*. Negative correlation between expression level and evolutionary rate of long intergenic noncoding RNAs. *Genome Biol Evol* 2011; **3**:1390–404.
25. Dai H, Chen Y, Chen S, *et al*. The evolution of courtship behaviors through the origination of a new gene in *Drosophila*. *Proc Natl Acad Sci USA* 2008;**105**:7478–83.

26. Wang W, Brunet FG, Nevo E, *et al*. Origin of sphinx, a young chimeric RNA gene in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 2002;**99**:4448–53.
27. Bejerano G, Pheasant M, Makunin I, *et al*. Ultraconserved elements in the human genome. *Science* 2004;**304**:1321–5.
28. Niazi F, Valadkhan S. Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3′ UTRs. *RNA* 2012;**18**: 825–43.
29. Boerner S, McGinnis KM. Computational identification and functional predictions of long noncoding RNA in *Zea mays*. *PLoS One* 2012;**7**:e43047.
30. Jia H, Osak M, Bogu GK, *et al*. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA* 2010;**16**:1478–87.
31. Kong L, Zhang Y, Ye ZQ, *et al*. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* 2007;**35**: W345–9.
32. Sun K, Chen X, Jiang P, *et al*. iSeeRNA: identification of long intergenic non-coding RNA transcripts from transcriptome sequencing data. *BMC Genomics* 2013;**14**(Suppl. 2):S7.
33. Wang L, Park HJ, Dasari S, *et al*. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* 2013;**41**:e74.
34. Sun L, Luo H, Bu D, *et al*. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res* 2013;**41**:e166.
35. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* 2011;**27**:i275–82.
36. Washietl S, Findeiss S, Müller S, *et al*. RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA (New York, NY)* 2011;**17**:578–94.
37. Cesana M, Cacchiarelli D, Legnini I, *et al*. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 2011;**147**:358–69.
38. Yoon JH, Abdelmohsen K, Srikantan S, *et al*. LincRNA-p21 suppresses target mRNA translation. *Mol Cell* 2012;**47**: 648–55.
39. Carrieri C, Cimatti L, Biagioli M, *et al*. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 2012;**491**:454–7.
40. Wilusz JE, JnBaptiste CK, Lu LY, *et al*. A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev* 2012;**26**:2392–407.
41. Brown JA, Valenstein ML, Yario TA, *et al*. Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc Natl Acad Sci USA* 2012;**109**:19202–7.
42. Moran VA, Niland CN, Khalil AM. Co-Immunoprecipitation of long noncoding RNAs. *Methods Mol Biol* 2012;**925**:219–28.
43. Khalil AM, Guttman M, Huarte M, *et al*. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* 2009;**106**:11667–72.
44. Zhao J, Ohsumi TK, Kung JT, *et al*. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* 2010;**40**:939–53.
45. Simon MD, Wang CI, Kharchenko PV, *et al*. The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci USA* 2011;**108**:20497–502.
46. Chu C, Qu K, Zhong FL, *et al*. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell* 2011;**44**:667–78.
47. Chakraborty D, Kappei D, Theis M, *et al*. Combined RNAi and localization for functionally dissecting long non-coding RNAs. *Nat Methods* 2012;**9**:360–2.
48. Kin T, Yamada K, Terai G, *et al*. fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Res* 2007;**35**: D145–8.
49. Bu D, Yu K, Sun S, *et al*. NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res* 2012;**40**:D210–5.
50. Amaral PP, Clark MB, Gascoigne DK, *et al*. lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res* 2011;**39**:D146–51.
51. Dinger ME, Pang KC, Mercer TR, *et al*. NRED: a database of long noncoding RNA expression. *Nucleic Acids Res* 2009; **37**:D122–6.
52. Jin J, Liu J, Wang H, *et al*. PLncDB: plant long non-coding RNA database. *Bioinformatics* 2013;**29**:1068–71.
53. Jeggari A, Marks DS, Larsson E. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics* 2012;**28**:2062–3.
54. Yang JH, Li JH, Jiang S, *et al*. ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res* 2013;**41**:D177–87.
55. Chen G, Wang Z, Wang D, *et al*. LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res* 2013;**41**:D983–6.
56. Wu T, Wang J, Liu C, *et al*. NPInter: the noncoding RNAs and protein related biomacromolecules interaction database. *Nucleic Acids Res* 2006;**34**:D150–2.
57. Chen D, Yuan C, Zhang J, *et al*. PlantNATsDB: a comprehensive database of plant natural antisense transcripts. *Nucleic Acids Res* 2012;**40**:D1187–93.
58. Yin Y, Zhao Y, Wang J, *et al*. antiCODE: a natural sense-antisense transcripts database. *BMC Bioinformatics* 2007;**8**: 319.
59. Kim MY, Hur J, Jeong S. Emerging roles of RNA and RNA-binding protein network in cancer cells. *BMB Rep* 2009;**42**:125–30.
60. Srikantan S, Gorospe M. HuR function in disease. *Front Biosci (Elite Ed)* 2012;**17**:189–205.
61. Srikantan S, Tominaga K, Gorospe M. Functional interplay between RNA-binding protein HuR and microRNAs. *Curr Protein Pept Sci* 2012;**13**:372–9.
62. Legnini I, Morlando M, Mangiavacchi A, *et al*. A feedforward regulatory loop between HuR and the long noncoding RNA linc-MD1 controls early phases of myogenesis. *Mol Cell* 2014;**53**:506–14.
63. Loraine A, McCormick S, Estrada A, *et al*. RNA-seq of *Arabidopsis* pollen uncovers novel transcription and alternative splicing. *Plant Physiol* 2013;**162**:1092–109.
64. Moghe G, Lehti-Shiu M, Seddon A, *et al*. Characteristics and significance of intergenic polyadenylated RNA transcription in *Arabidopsis*. *Plant Physiol* 2013;**161**:210–24.

65. Liu J, Jung C, Xu J, *et al*. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 2012;**24**:4333–45.
66. Wang H, Chung P, Liu J, *et al*. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res* 2014;**24**: 444–53.
67. Wu H-J, Wang Z-M, Wang M, *et al*. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol* 2013;**161**:1875–84.
68. Li S, Liberman L, Mukherjee N, *et al*. Integrated detection of natural antisense transcripts using strand-specific RNA sequencing data. *Genome Res* 2013;**23**:1730–9.
69. Liu TT, Zhu D, Chen W, *et al*. A global identification and analysis of small nucleolar RNAs and possible intermediate-sized non-coding RNAs in *Oryza sativa*. *Mol Plant* 2013;**6**: 830–46.
70. Lu X, Chen D, Shu D, *et al*. The differential transcription network between embryo and endosperm in the early developing maize seed. *Plant Physiol* 2013;**162**:440–55.
71. Zhang M, Zhao H, Xie S, *et al*. Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci USA* 2011; **108**:20042–7.
72. Li L, Eichten S, Shimizu R, *et al*. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol* 2014;**15**:R40.
73. Qi X, Xie S, Liu Y, *et al*. Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. *Plant Mol Biol* 2013;**83**:459–73.
74. Wang L, Zhao S, Gu C, *et al*. Deep RNA-Seq uncovers the peach transcriptome landscape. *Plant Mol Biol* 2013;**83**:365–77.
75. Yu X, Yang J, Li X, *et al*. Global analysis of cis-natural antisense transcripts and their heat-responsive nat-siRNAs in *Brassica rapa*. *BMC Plant Biol* 2013;**13**:208.
76. Song D, Yang Y, Yu B, *et al*. Computational prediction of novel non-coding RNAs in *Arabidopsis thaliana*. *BMC Bioinformatics* 2009;**10**(Suppl. 1):S36.
77. Hirsch J, Lefort V, Vankersschaver M, *et al*. Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts. *Plant Physiol* 2006;**140**:1192–204.
78. Ben Amor B, Wirth S, Merchan F, *et al*. Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Res* 2009;**19**: 57–69.
79. Hamburger D, Rezzonico E, MacDonald-Comber Petetot J, *et al*. Identification and characterization of the *Arabidopsis* PHO1 gene involved in phosphate loading to the xylem. *Plant Cell* 2002;**14**:889–902.
80. Secco D, Baumann A, Poirier Y. Characterization of the rice PHO1 gene family reveals a key role for OsPHO1;2 in phosphate homeostasis and the evolution of a distinct clade in dicotyledons. *Plant Physiol* 2010;**152**:1693–704.
81. Jabnoune M, Secco D, Lecampion C, *et al*. A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. *Plant Cell* 2013;**25**:4166–82.
82. Franco-Zorrilla JM, Valli A, Todesco M, *et al*. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* 2007;**39**:1033–7.
83. Salmena L, Poliseno L, Tay Y, *et al*. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011; **146**:353–8.
84. Tay Y, Kats L, Salmena L, *et al*. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* 2011;**147**:344–57.
85. Swiezewski S, Liu F, Magusin A, *et al*. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* 2009;**462**:799–802.
86. Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 2011;**331**: 76–9.
87. Ding J, Lu Q, Ouyang Y, *et al*. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc Natl Acad Sci USA* 2012;**109**:2654–9.
88. Zhou H, Liu Q, Li J, *et al*. Photoperiod- and thermo-sen sitive genic male sterility in rice are caused by a point mutation in a novel noncoding RNA that produces a small RNA. *Cell Res* 2012;**22**:649–60.
89. Zhu D, Deng XW. A non-coding RNA locus mediates environment-conditioned male sterility in rice. *Cell Res* 2012;**22**:791–2.
90. Jin G, Sun J, Isaacs SD, *et al*. Human polymorphisms at long non-coding RNAs (lncRNAs) and association with prostate cancer risk. *Carcinogenesis* 2011;**32**:1655–9.
91. Devaney JM, Hoffman EP, Gordish-Dressman H, *et al*. IGF-II gene region polymorphisms related to exertional muscle damage. *J Appl Physiol (1985)* 2007;**102**: 1815–23.
92. Cifuentes-Rojas C, Kannan K, Tseng L, *et al*. Two RNA subunits and POT1a are components of *Arabidopsis telomerase*. *Proc Natl Acad Sci USA* 2011;**108**:73–8.
93. Cifuentes-Rojas C, Nelson AD, Boltz KA, *et al*. An alternative telomerase RNA in *Arabidopsis* modulates enzyme activity in response to DNA damage. *Genes Dev* 2012;**26**:2512–23.
94. Lu T, Zhu C, Lu G, *et al*. Strand-specific RNA-seq reveals widespread occurrence of novel cis-natural antisense transcripts in rice. *BMC Genomics* 2012;**13**:721.
95. Secco D, Jabnoune M, Walker H, *et al*. Spatio-temporal transcript profiling of rice roots and shoots in response to phosphate starvation and recovery. *Plant Cell* 2013;**25**: 4285–304.
96. Zhu QH, Stephen S, Taylor J, *et al*. Long noncoding RNAs responsive to *Fusarium oxysporum* infection in *Arabidopsis thaliana*. *New Phytol* 2013;**201**:574–84.
97. Qi X, Xie S, Liu Y, *et al*. Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. *Plant Mol Biol* 2013; **83**:459–73.
98. Xin M, Wang Y, Yao Y, *et al*. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol* 2011;**11**:61.