Molecular Plant • Volume 6 • Number 3 • Pages 704–715 • May 2013 REVIEW ARTICLE

Mechanisms of Small RNA Generation from *Cis*-NATs in Response to Environmental and Developmental Cues

Xiaoming Zhang[a], Yifan Lii[a], Zhigang Wu[b], Anton Polishko[c], Huiming Zhang[d], Viswanathan Chinnusamy[e], Stefano Lonardi[c], Jian-Kang Zhu[d,f,1], Renyi Liu[b,1] and **Hailing Jin[a,1]**

a Department of Plant Pathology and Microbiology, Center for Plant Cell Biology and Institute for Integrative Genome Biology, University of California, Riverside, CA 92521, USA
b Department of Botany and Plant Sciences and Institute for Integrative Genome Biology, University of California, Riverside, CA 92521, USA
c Computer Science and Engineering, Center for Plant Cell Biology and Institute for Integrative Genome Biology, University of California, Riverside, CA 92521, USA
d Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907, USA
e Division of Plant Physiology, Indian Agricultural Research Institute, New Delhi 110012, India
f Shanghai Center for Plant Stress Biology and Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

ABSTRACT A large proportion of eukaryotic genomes is transcribed from both positive and negative strands of DNA and thus may generate overlapping sense and antisense transcripts. Some of these so-called natural antisense transcripts (NATs) are possibly co-regulated. When the overlapping sense and antisense transcripts are expressed at the same time in the same cell in response to various developmental and environmental cues; they may form double-stranded RNAs, which could be recognized by the small RNA biogenesis machinery and processed into small interfering RNAs (siRNAs). *cis*-NAT-derived siRNAs (nat-siRNAs) are present in plants, animals, and fungi. In plants, the presence of nat-siRNAs is supported not only by Northern blot and genetic analyses, but also by the fact that there is an overall sixfold enrichment of siRNAs in the overlapping regions of *cis*-NATs and 19%–29% of the siRNA-generating *cis*-NATs in plants give rise to siRNAs only in their overlapping regions. Silencing mediated by nat-siRNAs is one of the mechanisms for regulating the expression of the *cis*-NATs. This review focuses on challenging issues related to the biogenesis mechanisms as well as regulation and detection of nat-siRNAs. The advantages and limitations of new technologies for detecting *cis*-NATs, including direct RNA sequencing and strand-specific RNA sequencing, are also discussed.

Key words: gene expression; gene regulation; gene silencing.

INTRODUCTION

Natural antisense transcripts (NATs) are endogenous transcripts that contain sequences complementary to each other and have the potential to be co-regulated either positively or negatively. Based on the site of their biogenesis, NATs are categorized into two subgroups: *cis*-NATs are transcribed from opposite DNA strands from the same genomic loci, whereas *trans*-NATs are transcribed from different genomic loci (Lapidot and Pilpel, 2006; Jin et al., 2008; Zhang et al., 2012). *cis*-NATs usually form long perfect matches with only one antisense transcript from the same generation site, while *trans*-NATs mostly form short imperfect matches with several antisense transcripts from different genomic loci. Genome-wide analysis indicates that 60%–80% of the human genome, more than 70% of the mouse genome, and around 30% of the plant genomes produce antisense transcripts, whose expression is possibly co-regulated with the sense transcripts (Yamada et al., 2003; Katayama et al., 2005; Werner et al., 2009). Silencing mediated by *cis*-NATs-derived small interfering RNAs is one of the mechanisms for regulating the expression of the NATs (Katiyar-Agarwal and Jin, 2010). This review addresses issues related to *cis*-NATs and nat-siRNAs, focusing on the biogenesis mechanisms

[1] To whom correspondence should be addressed. H.J. E-mail hailingj@ucr.edu, tel. +1-951-827-7995. R.L. E-mail renyi.liu@ucr.edu. J.-k.Z. E-mail jkzhu@purdue.edu.

doi:10.1093/mp/sst051, Advance Access publication 15 March 2013

Received 15 January 2013; accepted 28 February 2013

of nat-siRNAs and their potential regulatory functions in response to stress or developmental cues.

CIS-NATS ARE WIDESPREAD IN EUKARYOTES

Genome sequencing and annotation suggest that up to 22%–26% of human genes, 22%–29% of mouse genes, 15%–17% of fly genes, 0.5%–2.8% of worm genes, 9% of *Arabidopsis* genes, 11% of yeast genes, and 12% of *Plasmodium falciparum* genes are overlapping and can potentially generate *cis*-NATs (Gunasekera et al., 2004; Steigele and Nieselt, 2005; Lapidot and Pilpel, 2006; Zhang et al., 2006, 2012). Recently, NATs were identified from 69 plant species and a plant natural antisense transcripts database (PlantNATsDB) was developed (Chen et al., 2012). NATs can regulate gene expression at the transcriptional or post-transcriptional level through distinct mechanisms. The mechanisms include the following. (1) Transcriptional interference blocks the association of RNA polymerase complex to the antisense strand (Prescott and Proudfoot, 2002; Crampton et al., 2006; Hobson et al., 2012). (2) RNA masking blocks the sense transcript from proteins involved in mRNA processing (splicing, modification, transportation, and translation) and microRNA (miRNA) binding and degradation (Hastings et al., 1997; Kumar and Carmichael, 1997; Tufarelli et al., 2003; Prasanth et al., 2005; Ebralidze et al., 2008; Faghihi et al., 2010; Morrissy et al., 2011). (3) Double-stranded RNA (dsRNA) from the pairing of sense and antisense transcripts activates protein kinase R, which phosphorylates several substrates, including eukaryotic initiation factor 2, resulting in translational inhibition; it also influences the activity of certain transcription factors (Werner et al., 2009). (4) Small RNAs (sRNAs), processed from dsRNA produced by the pairing of sense–antisense transcripts, induce silencing of the NATs (Borsani et al., 2005). Thus, *cis*-NATs not only are widely present in eukaryotes, but also can be co-regulated through various mechanisms. Our discussion is focused on sRNA biogenesis from *cis*-NATs, an interesting regulatory pathway that remains to be elucidated.

OCCURRENCE OF NAT-SIRNAS IN PLANTS, ANIMALS, AND FUNGI

There are two major categories of sRNAs: miRNAs and siRNAs. miRNAs are processed from hairpin-structured precursors transcribed by RNA polymerase II. The biogenesis of miRNAs does not require any RNA-dependent RNA polymerases (RDR) or other RNA polymerases. siRNAs are processed from dsRNA precursors generated from antisense transcription by RNA polymerase II, and/or generated by the action of plant-specific RNA polymerase IV or V or RDRs (Pikaard et al., 2008; Wierzbicki et al., 2008; Mosher et al., 2009). These dsRNAs are processed into 20–30 nucleotide (nt) sRNAs by RNase III-type ribonuclease enzymes, Dicer, or Dicer-like proteins (DCL). Both miRNAs and siRNAs are loaded into Argonaute (AGO) proteins to guide silencing of their target transcripts (Jones-Rhoades et al., 2006; Mallory and Vaucheret, 2010; Molnar et al., 2011). nat-siRNA-mediated silencing is initiated by the formation of dsRNAs from the base-pairing of sense-antisense transcripts (Chapman and Carrington, 2007). Studies in different species have indicated that nat-siRNAs appear in multiple eukaryotic kingdoms. nat-siRNAs were first identified in plants in a study of a salt stress-regulated NATs pair in *Arabidopsis*–delta-pyrroline-5-carboxylate dehydrogenase gene (*P5CDH*) and Similar to Radicle Induced Cell Death One 5 gene (*SRO5*) (Borsani et al., 2005). The *SRO5* transcript is induced by salt stress and its 3′ region may form dsRNA with the constitutively expressed *P5CDH* transcript. The dsRNAs are then processed by DCLs into siRNAs, namely nat-siRNASRO5, which may be amplified by the plant-specific RNA polymerase IV (Pol IV) and RDR6. nat-siRNASRO5 was proposed to direct the cleavage of *P5CDH* transcripts, thereby reducing proline degradation and increasing salinity tolerance. A bacterial infection-induced nat-siRNA, nat-siRNAATGB2, was also identified in *Arabidopsis* (Katiyar-Agarwal et al., 2006). The *ATGB2* transcript from a GTP binding protein gene specifically induced by the infection of a bacterium *Pseudomonas syringae* DC3000 strain carrying an effector gene *avrRpt2* may base pair with a constitutively expressed antisense transcript encoding the Pentatrico Peptide Repeat-Like protein (PPRL). The dsRNA is processed into nat-siRNAATGB2 by DCL1. Biogenesis of nat-siRNAATGB2 is also dependent on RDR6, SGS3, and NRPD1 (the largest subunit of Pol IV). The nat-siRNA was proposed to direct the down-regulation of *PPRL*, a negative regulator of the *RPS2*-mediated resistance pathway, and thereby contributing to effector-triggered host immunity (Katiyar-Agarwal et al., 2006).

nat-siRNAs may also regulate plant development. ARIADNE14 (ARI14) is a putative ubiquitin E3 ligase and the overexpression of ARI14 in sperm impairs fertilization of plant. The pairing of sense transcript *KOKOPELLI* (*KPL*) and antisense transcript *ARI14* would give rise to nat-siRNAs in the *Arabidopsis* pollen to restrict the expression of *ARI14* (Ron et al., 2010). The absence of *KPL* increases the accumulation of *ARI14* transcript and thereby impairs plant fertilization. Mutations in the sRNA pathway proteins DCL1, HEN1, HYL1, RDR2, SGS3, and NRPD1 de-repress the expression of *ARI14*, further supporting the functions of nat-siRNAs in *ARI14* regulation (Ron et al., 2010). The nat-siRNAs generated from the overlapping region of the sense and antisense transcripts of the Shooting (SHO) gene locus in *petunia* × *hybrida* can be detected in all tissues except roots, allowing the *SHO* gene to direct cytokinin synthesis in the correct root locations (Zubko and Meyer, 2007). These two examples suggest that nat-siRNAs play important roles in gene regulation during plant development. Moreover, some nat-siRNAs can be specifically cloned from the reproductive organ of plants, further suggesting that nat-siRNA may

contribute to plant development and reproduction (Chen et al., 2010). Furthermore, genome-wide studies of nat-siRNAs in *Arabidopsis* and rice support the existence of nat-siRNAs as well as nat-siRNA-mediated regulation of NAT genes in response to environmental cues (Jin et al., 2008; Zhou et al., 2009; Zhang et al., 2012).

nat-siRNAs have also been found in other eukaryotic organisms including *Drosophila*, mouse, human, worm, and yeast. In *Drosophila*, *cis*-NAT transcripts are 10 times more likely to generate siRNAs than non-paired transcripts, and nat-siRNAs are one major siRNA class found in both somatic cells and gonadal tissues (Czech et al., 2008; Ghildiyal et al., 2008). Most of the nat-siRNAs in *Drosophila* are generated from the 3′–3′ overlapping regions of convergent *cis*-NATs. These 21-nt nat-siRNAs are processed by DICER2 and loaded into AGO2 with the help of R2D2 to direct silencing (Czech et al., 2008; Ghildiyal et al., 2008; Okamura et al., 2008, 2011). In mouse oocytes, bidirectional and antisense transcription are two major sources of nat-siRNAs. dsRNAs formed in these regions of the mouse genome are processed into 21-nt nat-siRNAs by Dicer and loaded into AGO2 (Watanabe et al., 2008). In addition to the siRNAs that accumulate from *cis*-NATs loci, siRNAs are also generated from *trans*-NAT loci from the pairing of sense transcripts from protein-coding genes and antisense transcripts from homologous pseudogenes (Tam et al., 2008). This study indicated that pseudogenes can also influence the regulation of homologous protein-coding genes through siRNAs. Similarly, siRNA enrichment in genomic regions with bidirectional transcription is also observed in human liver cells (Kawaji et al., 2008). Many nat-siRNAs were mapped to co-expressing sense and antisense transcripts within synaptic fractions of adult mouse hippocampus. A number of these genes are involved in Alzheimer's disease and/or synaptic signaling pathways (Smalheiser et al., 2011). nat-siRNAs were also found in the worm *Schistosoma japonicum*, although they only account for a minor proportion of the whole siRNA profile, and the authors only analyzed *trans*-nat-siRNAs in their report (Cai et al., 2011). Worm nat-siRNAs are predominately present in the cercaria life cycle stage where they regulate the expression of diverse genes that contribute to worm development. In the yeasts *Saccharomyces catellii* and *Candida albicans*, sense and antisense transcript pairs are one of the major sources of siRNAs (Drinnenberg et al., 2009). These dsRNAs can generate DCR1-dependent nat-siRNAs that are loaded into AGO1. A hyperthermophilic archaeon, *Sulfolobus solfataricus*, also has siRNAs enriched in overlapping transcript regions, and in some loci that produce antisense transcripts (Xu et al., 2012). Although the archaeal genomes are similar in size to those of bacteria, they encode many proteins with eukaryotic features. This study suggests that nat-siRNA-mediated regulation may be evolutionarily conserved and rather ancient. Taken together, available data support that nat-siRNAs are evolutionarily conserved, functional sRNAs that occur in many eukaryotes and archaea.

NAT-SIRNAS ARE GENERATED IN RESPONSE TO DEVELOPMENTAL AND ENVIRONMENTAL CUES

While there are many overlapping transcripts in plants, animals, and yeasts, and nat-siRNAs do occur in multiple eukaryotic lineages, nat-siRNAs do not appear constitutively produced in all cells. nat-siRNA-mediated regulation requires the co-expression of both sense and antisense transcripts at the same time and in the same cell. Many nat-siRNAs are likely specific to a tissue or developmental stage, or to an environmental condition. It would not be cost-effective for organisms if this regulatory strategy is 'on' all the time or in all cells. Therefore, it is not surprising that not many nat-siRNAs have been observed in most genome-wide analyses of whole plants under normal conditions. Many nat-siRNAs were detected only in specific tissues or under specific conditions in both plants and animals. In plants, nat-siRNASRO5 is induced by salt stress (Borsani et al., 2005), nat-siRNAATGB2 only accumulates in response to bacterial pathogen infection (Katiyar-Agarwal et al., 2006), nat-siRNA from the *KPL–ARI14* locus only occurs in sperm (Ron et al., 2010), and nat-siRNA from the *SHO* locus cannot accumulate in the root (Zubko and Meyer, 2007). Genome-wide analysis of nat-siRNAs in *Arabidopsis* and rice further suggested that the accumulation of many nat-siRNAs is condition-specific (Jin et al., 2008; Zhou et al., 2009; Chen et al., 2010; Lu et al., 2012; Zhang et al., 2012).

In animals, nat-siRNAs generated from antisense transcripts of *Slc34a* also display tissue specificity, as they are present only in mouse kidneys and testes (Carlile et al., 2009). The sense/antisense transcripts of this locus are only co-expressed within a very short period of time during embryonic development of zebra fish at 48 h post fertilization (hpf) (Carlile et al., 2008, 2009). Most strikingly, the orientation of the nat-siRNAs from this locus is also developmentally regulated. nat-siRNAs complementary to antisense transcripts were only expressed at 48 hpf, whereas nat-siRNAs complementary to sense transcripts were only detected at 72 hpf (Carlile et al., 2008, 2009; Werner et al., 2009). Similar strand bias is also observed for the Ppp4r1 siRNAs in mouse oocytes (Watanabe et al., 2008). Moreover, many plasticity-related genes that express both sense and antisense transcripts in the mammalian hippocampus, including some genes involved in Alzheimer's disease, schizophrenia, and synaptic signaling pathways, can generate nat-siRNAs in both orientations (Smalheiser et al., 2011). Several groups who analyzed the accumulation of NATs in various eukaryotic organisms also observed that nat-siRNAs only accumulate under specific circumstances or in specific cells (He et al., 2008; Jin et al., 2008; Faghihi and Wahlestedt, 2009). Although some of the nat-siRNAs occur at very low levels and are difficult to detect, they may still be functionally important, as they could be concentrated in *cis* at their generation site and thus may efficiently regulate the expression of transcripts from the same locus. Therefore,

many or most nat-siRNAs are likely produced in response to different environmental or developmental cues (Jin, 2008; Padmanabhan et al., 2009).

CHALLENGING ISSUES RELATED TO NAT-SIRNAS

Although nat-siRNAs are potentially an important regulatory mechanism in the regulation of NAT genes, studies on nat-siRNAs face several significant challenges. One challenge is their complex biogenesis mechanism that remains to be clearly defined. Recent deep sequencing studies have revealed various patterns of distribution of nat-siRNAs along the NATs (Lu et al., 2012; Zhang et al., 2012), which indicated that the biogenesis of nat-siRNAs is very complicated. Most nat-siRNAs in animals are generated within the overlapping regions of transcripts, with a few exceptions that nat-siRNAs are scattered throughout the entire transcript, such as *pgant35A* in Drosophila (Czech et al., 2008). Although some plant nat-siRNAs are also found only within the overlapping regions of NAT pairs (19% in *Arabidopsis* and 29% in rice), the majority of them are distributed beyond the overlapping regions of the NAT transcripts (Zhang et al., 2012). More than 50% of plant nat-siRNAs exhibit a strand bias (Zhang et al., 2012), possibly indicating a specific regulatory role of them on one transcript of a NAT pair. Moreover, some nat-siRNAs display a site-specific pattern in which the siRNAs are derived from one or several specific sites within the overlapping transcripts (Zhang et al., 2012). These different distribution patterns may be the result of distinct mechanisms of nat-siRNA biogenesis. It was proposed that all plant siRNAs are generated from dsRNAs by DCL proteins (Mallory and Vaucheret, 2006; Ruiz-Ferrer and Voinnet, 2009; Ding, 2010; Melnyk et al., 2011). The biogenesis of nat-siRNAs is not well defined, and it appears very complex and there are important unanswered questions. Based on available data, we speculate that there are at least five possible mechanisms to generate siRNAs from *cis*-NAT transcripts, depending on how the *cis*-NAT transcripts pair with each other and the participation of RNA-dependent RNA polymerases.

The most straightforward mechanism is the formation of dsRNAs by direct base-pairing of sense and antisense transcripts. As shown in Figure 1A, under specific conditions, such as in certain developmental stages or specific stress conditions, two transcripts may accumulate to significant levels in the same cell and pair up to form partial dsRNAs. sRNA machinery is then recruited to process the dsRNAs into nat-siRNAs. The nat-siRNAs generated from this mechanism would be restricted to the overlapping region of *cis*-NATs. Almost all nat-siRNAs from animals and a group of nat-siRNAs from plants are indeed only present in the overlapping regions of the *cis*-NATs. 19%–29% of the siRNA-generating *cis*-NATs in plants give rise to siRNAs only in their overlapping regions (Zhang et al., 2012). Importantly, the majority of siRNA-generating *cis*-NATs show a clear enrichment within the overlapping regions. The overlapping regions in plant *cis*-NATs generate over six times more siRNAs than the non-overlapping regions (Zhang et al., 2012).

In plants, the distribution of many nat-siRNAs extends beyond the overlapping regions. This may be due to the activities of RDRs, which use the single-stranded regions of NAT transcripts as a template to produce dsRNAs (Katiyar-Agarwal and Jin, 2010). The function of RDRs in RNA silencing was first reported when studying systemic silencing in plants (Voinnet et al., 1998; Dalmay et al., 2000). As shown in Figure 1B, primary nat-siRNAs processed by DCL proteins in the overlapping region by the first mechanism may serve as primers for RDRs, which synthesize dsRNAs using single-stranded regions of NAT transcripts as templates. The dsRNAs are processed by DCL proteins and give rise to secondary nat-siRNAs corresponding to either the overlapping region or the regions outside of it. The accumulation of a portion of nat-siRNAs in plants indeed is dependent on the function of RDRs, including RDR2 and RDR6 (Borsani et al., 2005; Katiyar-Agarwal et al., 2006; Zhang et al., 2012). The small RNAs dependent on RDR2 might extend to and target the promoter region of the sense and/or antisense transcripts and potentially regulate the accumulation of cognate transcripts at the transcriptional level.

Site-specific pattern is another distribution pattern of nat-siRNAs, which may be caused by the secondary or higher-order structures of a transcript in the NATs pair. Site-specific patterns of nat-siRNAs were observed in both *Arabidopsis* and rice (Zhang et al., 2012). As shown in Figure 1C and Figure 1D, both the sense and antisense strands are independently transcribed, and the pairing of the transcripts may induce the formation of a complex secondary or high-order structure (Figure 1C), or the independently transcribed RNAs may form complex secondary structures separately, which would limit the base-pairing regions between the sense and antisense transcripts (Figure 1D). RNAs with complex structures may be processed by DCL1 into site-specific nat-siRNAs. It is also possible that more extended dsRNAs are processed by DCL proteins into distributed siRNAs, but only one or a few are specifically selected and protected by AGOs or other RNA binding proteins and the rest of the distributed siRNAs are degraded, therefore resulting in the accumulation of siRNAs at only one or several positions. Just like the accumulation of miRNAs, although miRNA and miRNA* are generated at 1:1 ratio, often only miRNAs are loaded into AGOs and stabilized whereas miRNA*s are degraded.

We cannot rule out another possibility that the production of the siRNAs in the *cis*-NAT region does not require the pairing of sense and antisense transcripts. One transcript may be sufficient to form secondary structures and be processed by the RNA silencing machinery into siRNAs (Figure 1E). These siRNA products would not be genuine nat-siRNAs. These above scenarios did not consider the participation of PolIV, so, if one takes into consideration of PolIV, further complexities of nat-siRNA biogenesis are expected. Lastly, for distributed siRNAs that show a strong or even complete strand bias (Zhang et al., 2012), entirely new and hitherto unknown mechanisms may have to be proposed to account for their biogenesis.

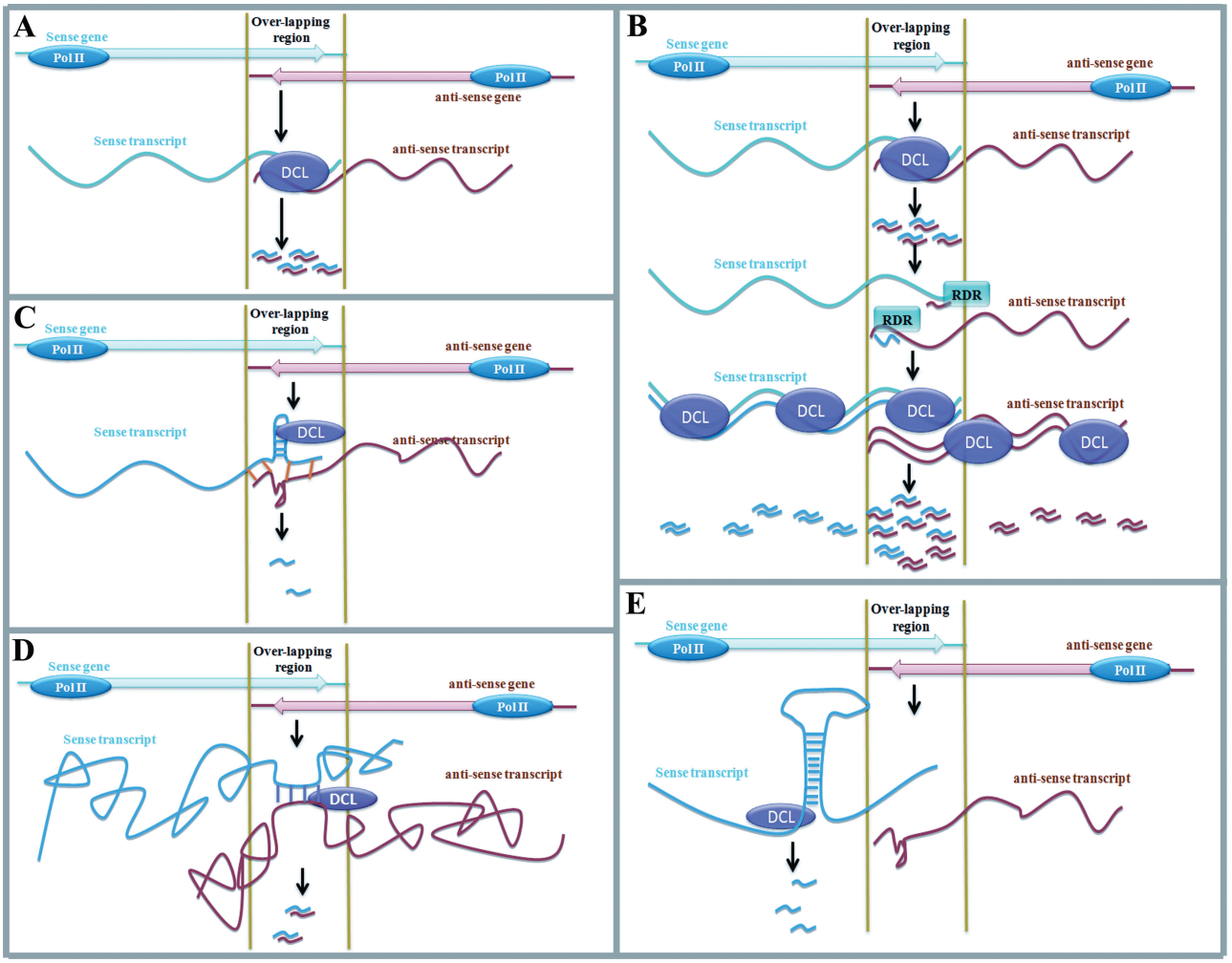

Figure 1. Models of the Possible Mechanisms for the Production of siRNAs from *cis*-NATs.
(A) Under specific conditions, sense and antisense genes are both transcribed by Pol II and accumulate in the same cell at the same time to form dsRNA, which is recognized by sRNA machinery. DCL processes the dsRNA into nat-siRNAs, which are solely derived from the overlapping region of the sense and antisense transcripts.
(B) Sense and antisense genes are transcribed by Pol II, and transcripts pair up to form dsRNA, which is processed by DCL to produce nat-siRNAs. These nat-siRNAs recruit RDR and/or Pol IV, which produce dsRNAs using the single-stranded transcripts as templates. The new dsRNAs are processed by DCL into secondary nat-siRNAs that match the overlapping region and regions extending beyond it.
(C) Sense and antisense genes are transcribed by Pol II, and transcripts pair up to form a complex secondary structure. DCL recognizes the secondary structure in only one of the transcripts and processes it into site-specific nat-siRNAs that are derived only from one transcript.
(D) Sense and antisense genes are transcribed by Pol II and the pairing of the transcripts forms a complex secondary structure that is recognized by DCL and processed into site-specific nat-siRNAs that are derived from both transcripts.
(E) Sense and antisense genes are transcribed by Pol II, but the secondary structure formed by one strand is sufficient to recruit DCL to form siRNAs. These siRNAs are solely derived from one transcript and do not require the pairing of the sense and antisense transcripts.

The second significant challenge facing nat-siRNA studies is the very low abundance of nat-siRNAs. *Cis*-regulation of NATs may not necessitate a high level of the nat-siRNAs, but the very low abundance makes nat-siRNA detection very difficult. Whereas most miRNAs can be detected on Northern blots using as low as a few micrograms of total RNAs, the detection of nat-siRNAs on Northern blots typically require 100 micrograms or more of enriched small RNAs per lane. For example, the first reported nat-siRNA, nat-siRNASRO5, has been very difficult to detect on Northern blots. Nevertheless, as shown in Figure 2, it could be independently detected using a high amount of enriched small RNAs from NaCl-treated *Arabidopsis* plants. In addition to the ~24-nt band, the oligonucleotide probe also detected an ~50-nt signal which was also NaCl-induced and is possibly an intermediate in the biogenesis of the 24-nt nat-siRNA (Figure 2). Some cell- or tissue-specific nat-siRNAs, such as pollen-specific nat-siRNAs generated from the *KPL–ARI14* locus, were too low to be even detected at all (Ron et al., 2010).

One might expect that deep sequencing may help detect specific nat-siRNAs. However, for siRNAs that are generated from long dsRNA precursors, such as repeat-associated siRNAs,

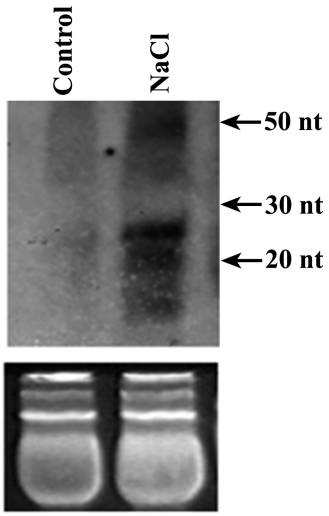

Figure 2. Northern Blot Detection of the Salt Stress-Induced *SRO5-P5CDH* nat-siRNA.
Arabidopsis seedlings were treated with half MS (control) or half MS supplemented with 150 mM NaCl (NaCl) for 6 h. Approximately 120 μg of small RNAs was loaded per lane. tRNA was used as loading control and similar results were obtained in two biological repeats.

nat-siRNAs (especially the distributed class), transgene-derived siRNAs, and virus-derived siRNAs, the positions of their generation are usually not the same, but are rather randomly distributed along the precursors. Even for abundant heterochromatic siRNAs, independent deep sequencing experiments do not usually detect exactly the same siRNAs due to the distributive pattern of the siRNAs. Therefore, when the abundance of an individual siRNA is very low, as in the case of nat-siRNAs, it would be even more difficult for deep sequencing to detect specific sequences every time, due to the many possible sequences that could be generated. One exception to the random siRNA generation positions across the dsRNA precursors is the biogenesis of tasiRNAs, which is initiated by miRNA-mediated cleavage that sets the phase of the processing and therefore ensures the specific siRNA generation positions.

Another problem is the potential bias introduced by sRNA library construction for deep sequencing. The cloning bias would lead to discrepancies between small RNA deep sequencing results and results obtained from Northern blot analysis and qPCR. Such discrepancies have been observed in both plant and animal systems (Reddy et al., 2009; Git et al., 2010; Kapranov et al., 2010; Zhang et al., 2011). Linsen et al. performed sRNA expression profiling using SREK-SOLiD and modban-Solexa deep sequencing methods and qPCR analysis on a sample composed of more than 400 synthetic human miRNAs at equal molarity (Linsen et al., 2009). They found up to four orders of magnitude of difference between the most and least abundant reads, indicating a strong bias generated by cloning different sRNAs for the deep sequencing. Even single nucleotide differences could influence the sRNA read frequencies. qPCR results from the same set of samples also differed substantially from the read frequencies (Linsen et al., 2009). Several steps within sRNA library preparation for deep sequencing could introduce bias, including: (1) ligation bias caused by the first 5′ or 3′ nucleotide of the RNA adaptors used in sRNA cloning; (2) possible secondary structure of the sRNA that may reduce or block the ligation; (3) sRNA modification that may reduce or block the ligation; and (4) the last PCR amplification step may enlarge the amount difference of sRNAs. The bias of RNA ligase was already confirmed and extensively studied (Hafner et al., 2011). Therefore, small RNA Northern blot analysis, when possible, remains more trustworthy compared to deep sequencing and other cloning-based methods.

NEW TECHNOLOGIES FOR DETECTING NATS

Even an accurate annotation of *cis*-NATs in the best-studied model plant, *Arabidopsis thaliana*, remains a challenge. Genome-wide cDNA sequencing, tiling arrays, direct RNA sequencing, and various other technologies have been used to identify *cis*-NATs in eukaryotes. Because antisense transcripts usually accumulate to relatively low levels compared to the sense transcripts, sometimes it is difficult to detect them (Wang et al., 2005; Werner et al., 2007). Moreover, the expression of antisense transcripts often appears to be developmental stage- and stress-condition-specific; therefore, whole-genome analysis under limited conditions cannot detect all *cis*-NATs (Wang et al., 2005; Sun et al., 2006; He et al., 2008; Mercer et al., 2008; Okada et al., 2008).

A recent study used Helicos direct RNA sequencing (DRS) to identify polyadenylated 3′ UTRs (Sherstnev et al., 2012). Using a surface coated with poly (dT), RNAs with 3′ polyA tail are captured and locked on the surface. A modified DNA-dependent DNA polymerase then adds the corresponding fluorescent nucleotide analog to the end of poly(dT) by base-pairing with the RNA it anchored. The sequencing of the RNA is finally obtained by fluorescence reading. Therefore, the DRS is achieved by synthesis with DNA-dependent DNA polymerase and anchored poly(dT), instead of indirectly sequencing cDNAs synthesized with poly(dT) and reverse transcriptase (Ozsolak et al., 2009). The authors asserted that the DRS method was 'sensitive with no bias' and that one-third of *Arabidopsis* genes in the TAIR database needed re-annotation based on their data. The study also claimed to have discovered many discrepancies in several published papers. The authors stated that the previous discoveries of non-templated base addition between cleavage sites and the poly (A) tail (Jin and Bian, 2004) and a large class of mRNAs that are cleaved in coding-exons (Meyers et al., 2004; Wu et al., 2011) were all artifacts from reverse transcriptase-dependent library construction or internal priming on A-rich sequences, respectively. They also stated that many previously identified protein-coding exosome targets (Chekanova et al., 2007) were artifacts from the combined use of reverse transcriptase and tiling arrays. This DRS study detected antisense transcription at only 3213 protein-coding genes, which is significantly fewer than what were found in the

previous studies that identified antisense transcripts at 7600 and 12 090 genes (Yamada et al., 2003; Stolc et al., 2005). The authors suggested that this discrepancy was also partially due to the artifacts of reverse transcriptase and tiling arrays. Moreover, their results indicated that the transcripts of At5g62520/At5g62530 and At4g35850/At4g35860 do not have sufficient overlap to produce the reported nat-siRNAs (Borsani et al., 2005; Katiyar-Agarwal et al., 2006).

To confirm the extent of overlap of the antisense transcripts, we used RT–PCR to amplify the transcript of At4g35850 using a forward primer before the last intron and several reverse primers downstream of the nat-siRNAATGB2 generation site. We detected RT–PCR products that extended 175 base pairs (bp) downstream of the nat-siRNAATGB2 site (Figure 3A and 3B), and the fragments were confirmed by sequencing, showing that the last intron was correctly spliced (Genbank No. JX893016). Similarly, a long transcript of the *SRO5* gene (At5g62520) was detected by 3′ RACE and confirmed by sequencing (Figure 3C and 3D) (GenBank No. JQ513374). According to the DRS study, the At5g62520/At5g62530 pair does not have enough overlap such that the *P5CDH* (At5g62530) 3′ probe used previously (Borsani et al., 2005) would detect only *SRO5* but not *P5CDH* itself. As shown in Figure 4, independent experiments found that, under salt stress conditions, *SRO5* gene induction (Figure 4A) is concomitant with *P5CDH* (At5g62530) gene repression (Figure 4B), as reported previously (Borsani et al., 2005). Importantly, both *SRO5* and *P5CDH* can be detected by a probe that corresponds to the last 189 bp of *P5CDH* 3′ UTR (Figure 4C). These results strongly support that the At5g62520/At5g62530 and At4g35850/At4g35860 NATs pairs do overlap as previously reported. In addition to detecting the *P5CDH* (#3 band) and *SRO5* (#4 band) bands, the short probe also detected several other signals that were either repressed (#1, #2, and #6 bands) or induced (#5 and #7 bands) (Figure 4C), which suggests a great complexity of transcripts from this region that has not been recognized thus far and is worthy of further investigation.

As DRS is a new technology that was performed only by Helicos and cannot be repeated by others, it is important that results in the Sherstnev et al. (2012) paper be validated by other means, especially when the paper claimed that more than 10 000 *Arabidopsis* genes need re-annotation and when many discrepancies were found. However, the authors only validated two transcripts by RT–PCR and 3′ RACE, and the RT–PCR primers they used did not cross introns and the reverse primers for 3′ RACE were incorrect. To further assess the quality and validity of this DRS data, we performed a comprehensive analysis of the DRS data set. We first examined the quality of the sequences. Using the sequence reads provided by the authors and the designated mapping software (Helisphere with the same parameters used in Sherstnev et al. (2012)), we were able to obtain only a fraction of the aligned reads claimed in the paper. To circumvent this problem, we instead started the analysis from the alignment files containing the details of 10 010 119 reads that were mapped to the TAIR10 genome, also provided by the authors. We took out the original RNA-Seq sequences and remapped the reads using Bowtie (Langmead et al., 2009), allowing up to three mismatches and no gaps. We were able to map 4 102 144 (41%) reads, among which only 1 366 779 (14%) reads had a perfect match and only 3 680 651 (37%) reads were mapped to a unique position in the genome. To ascertain that the large difference in these statistics compared to the ones in Sherstnev et al. (2012) were not solely caused by different mapping tools or parameters, we also mapped the original sequences using BWA (Li and Durbin, 2010) with two sets of parameters. With the default BWA parameters, namely (1) the fraction of missing alignments (given 2% uniform base error rate) was 0.04; (2) the maximum number or fraction of gap openings was 1; (3) long gaps were disabled; (4) indels within 5 bp towards the ends were disabled; (5) the maximum occurrences for extending a long deletion was 10; (6) the seed length was 32; and (7) the maximum number of differences in the seed was two. BWA mapped only 5 406 447 reads (54%) among which 5 173 080 (52%) were uniquely mapped. In the second set of parameters for BWA, we tried to emulate the parameters used by the authors for the Helisphere software: (1) the fraction of missing alignments was 0.07; (2) the maximum number or fraction of gap openings was 10; (3) long gaps were allowed; (4) indels towards the ends were allowed; (5) the maximum occurrence for extending a long deletion was 12; (6) the seed length was 32; and (7) the maximum number of differences in the seed was five. Even with these very relaxed parameters, we were able to map only 7 515 698 reads (75%) among which 6 964 455 reads (70%) were mapped uniquely.

We also analyzed the original alignments produced by the authors. Since the Helisphere software allows gaps when it searches for optimal mapping position(s), we counted a gap in the alignment as a mismatch when we evaluated the alignment quality. Among the 10 010 119 reads mapped by the Helisphere software, we found that 13.6%, 23.5%, 23.9%, 18.9%, 11.8%, and 8.4% of the reads were mapped with 0, 1, 2, 3, 4, and ≥5 mismatches or gaps, respectively. On average, there were 2.2 mismatches per alignment. These results taken together clearly indicate that the majority of the original sequence reads have a high rate of sequencing errors (in particular, indels), and thus they are of poor quality. The high error rate is an inherent problem of the Helicos DRS technology, with a typical raw base error of 4% (including 2%–3% deletions, 1%–2% insertions, and 0.1%–0.3% substitutions) (Ozsolak et al., 2009).

We next examined the raw read coverage at the TAIR10-annotated 3′ end and the proposed 3′ end for each of the 10 380 genes that were redefined by Sherstnev et al. (2012). Using the original alignments generated by the authors, we calculated the number of raw reads that cover each position on the positive and negative strands of the nuclear chromosomes, and we then extracted the raw coverage on the TAIR 10-annotated 3′ end and the authors-proposed 3′ end of each gene (Supplemental Table 1). We looked at the raw coverage on each proposed 3′ end and found that 50% of the

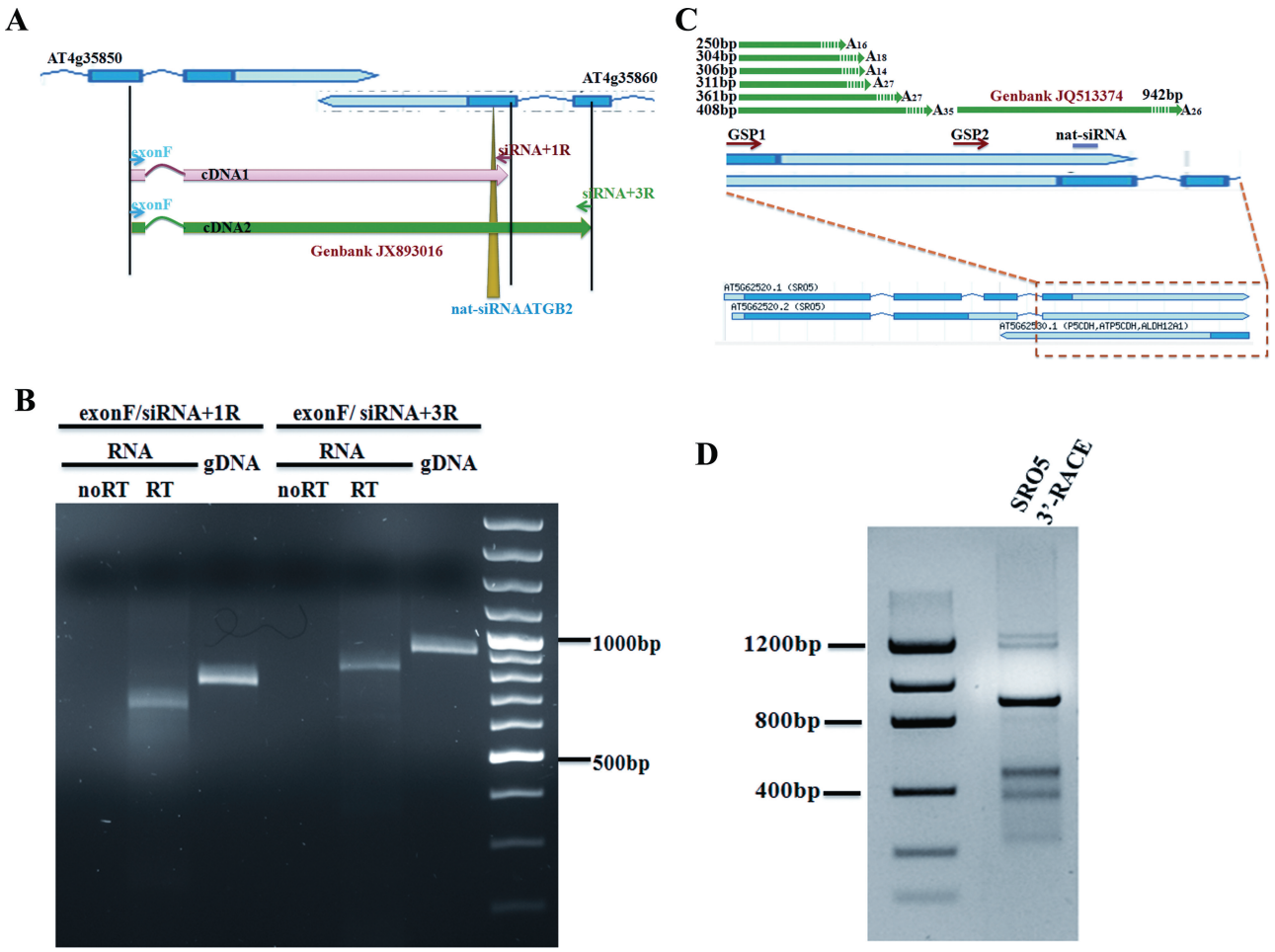

Figure 3. Schematic Diagrams and Gels Showing that Transcripts of At5g62520/At5g62530 and At4g35850/At4g35860 Overlap in the Region from which nat-siRNAs Are Produced.

(A) RT–PCR products cDNA1 (red bar) and cDNA2 (purple bar, Genbank No. JX893016) amplified from transcript At4g35850 using primer pairs exonF/siRNA+1R and exonF/siRNA+3R, respectively. Products include the nat-siRNAATGB2 generation site (green triangle). The products do not include the last intron in At4g35850, ruling out the possibility the products were from genomic DNA, and they include the last intron of At4g35860, ruling out the possibility that the products were from transcript At4g35860. Blue bars show gene models. Black vertical lines indicate the beginning and end of the RT–PCR products.

(B) Gel of PCR products amplified from (1) noRT (RNA only), (2) RT (cDNA from RT reaction), and (3) gDNA (genomic DNA) using primer pairs exonF/siRNA+1R and exonF/siRNA+3R. No bands in (1) show there is no DNA contamination in the RNA. Bands in (2) are shorter than that in (3), showing that cDNA1 and cDNA2 (from A) do not contain introns and are not from genomic DNA.

(C) Bottom blue bars are gene models for the transcript pair At5g62520/At5g62530 (*SRO5*/*P5CDH*). Inset shows 3′ RACE products (green) amplified using GSP1 and GSP2 primers (red arrows). A long transcript of the *SRO5* gene (GenBank No. JQ513374) contains the nat-siRNA generation site (green) and also overlaps with antisense gene *P5CDH*.

(D) Gel electrophoresis of *SRO5* 3′ RACE amplification products. The 942-bp amplicon was sequenced and identified as a poly-adenylated transcript of *SRO5* (GeneBank No. JQ513374.1).

proposed 3′ ends were covered by four or fewer reads and 928 (9%) of which were covered by no mapped reads. Considering that the authors used a smoothing algorithm for peak coverage detection, we also searched the maximum coverage within 2 bp of the proposed 3′ ends and found that 35% of the proposed ends had maximum coverage of four or fewer reads and 263 (2.5%) have zero maximum coverage within the 2-bp window (Supplemental Table 1). This indicates that many of their proposed 3′ ends were not well supported by the data. Compared to the raw coverage at the authors-proposed 3′ end, for 90% of the listed 10 380 genes, the raw coverage at the TAIR10-annotated 3′ end is greater. Specifically, 4.5%, 24.8%, 41.1%, and 69.4% of the genes have 50, 10, 5, and 2 times as many mapped reads at the TAIR10 annotated 3′ end, respectively. These numbers were 6.5%, 27.3%, 44%, and 70% if maximum coverage within the 2-bp window of the TAIR10 annotated and the Sherstnev et al. (2012) proposed ends were used. This suggests that, even if it could be established that the Sherstnev et al. (2012) proposed 3′ ends were real, they most likely resulted from alternative polyadenylation and represented a minor form of low frequency. Furthermore, we examined the minimum coverage between each pair of

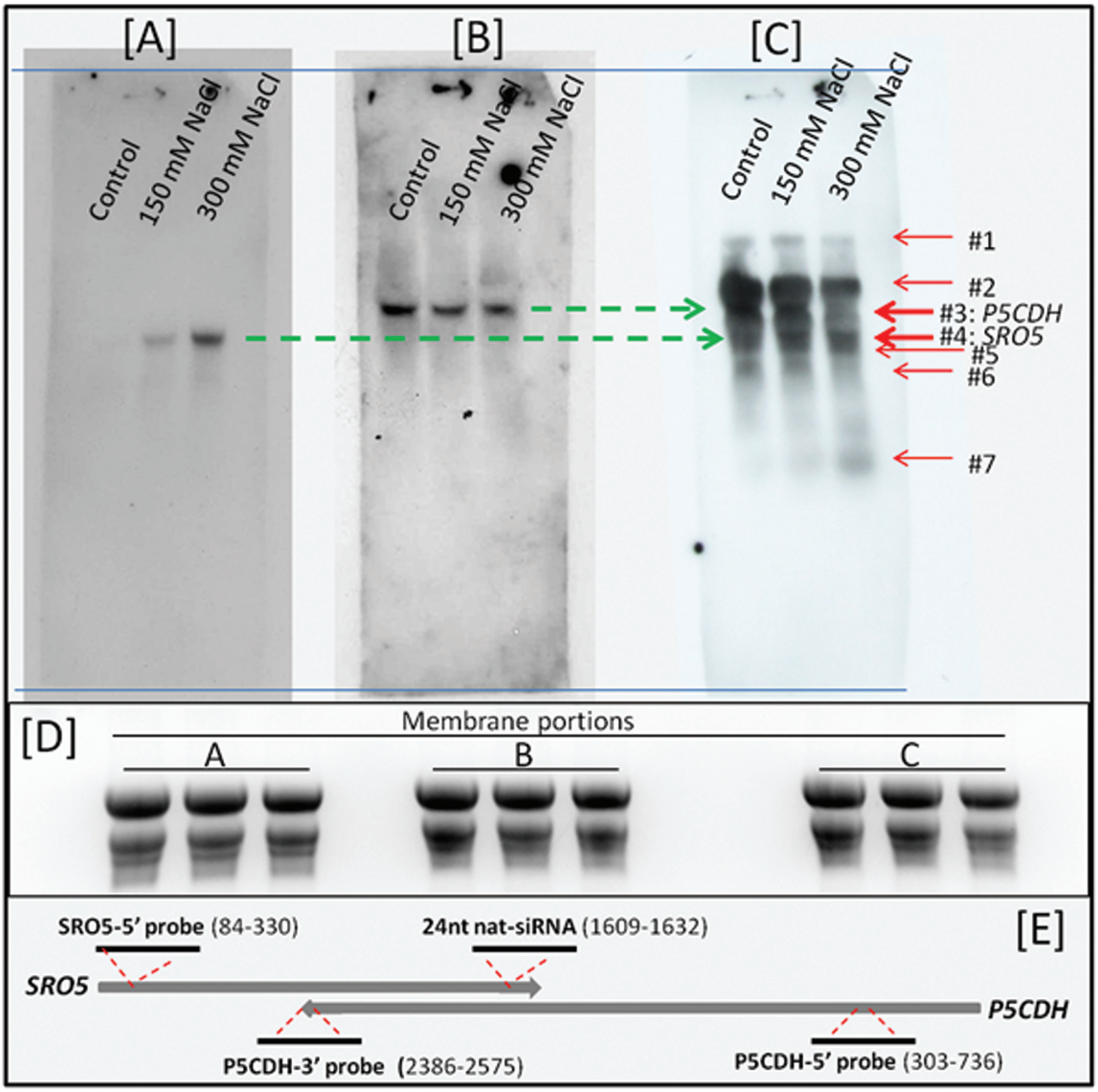

Figure 4. Northern Blot Analysis of *SRO5* and *P5CDH* Transcripts.
(A) *SRO5*, hybridized with the *SRO5-5'* probe as described (Supplementary Materials).
(B) *P5CDH*, hybridized with the *P5CDH-5'* probe.
(C) *SRO5* and *P5CDH*, hybridized with the *P5CDH-3'* probe. Seven identified bands were numbered from top to bottom as indicated. For best results, all membranes were new and hybridized with freshly prepared probes. The three membranes (A, B, and C) were three portions of a big membrane **(D)** to ensure that all of the samples were run simultaneously under the same conditions, so that the band sizes can be compared among the three membranes. Panel (D) shows EtBr staining of rRNA. The positions of the probes are indicated in **(E)**.

TAIR10-annotated and the Sherstnev et al. (2012) proposed 3' ends and found that, for 881 (8.5%) genes, the minimum coverage between two ends is zero. Because it is rare to have an intron from such a short distance from the 3' end, without continuous coverage, it is not possible to determine whether the expressed sequences at the proposed 3' end came from mRNAs of the corresponding gene, different neighboring genes, unannotated RNAs, or random noise.

We also compared the DRS study with a recent study using strand-specific RNA-Seq technology to define the 3' end formation of RNA transcripts in *Arabidopsis*. We started with the strand-specific RNA-Seq data that were generated from five libraries from wild-type *Arabidopsis* seedlings (Kurihara et al., 2012). We mapped the RNA-Seq reads to the TAIR10 reference genome using the Tophat software (Trapnell et al., 2009) and obtained the raw read depth on each position of the two strands of DNA. From the annotated 3' end of each gene, we extended the end if the next downstream position was covered by at least five (or two) raw reads. As shown in Supplemental Table 1, when a contiguous coverage of five raw reads was required, we were able to extend the 3' end of 3020 genes, among which 1118 genes were not

extended by the DRS study and 804 genes were extended to a longer length than the DRS study (Supplemental Table 2). When a contiguous coverage of two raw reads was required, these numbers were 6395, 2525, and 2238, respectively (Supplemental Table 3). These results indicate that the DRS method is not as sensitive as the authors claimed, and many 3′ ends of genes easily detected by other methods were missing from the DRS data set. Experimental validation is clearly necessary to validate the results from new technologies. In-depth functional analysis will help further understand the expression and regulation of NATs and NATs-derived siRNAs.

PERSPECTIVES

Most genome-wide studies have shown that large percentages of genomes have antisense transcription, and that the expression of antisense transcripts is tightly regulated (Chen et al., 2005; Jin et al., 2008; Werner et al., 2009). Antisense transcripts have biological functions, and most of them do not appear to be transcriptional noise. The base-pairing between sense and antisense transcripts forms dsRNAs, which can be processed into nat-siRNAs. However, the complex secondary structures of independently transcribed sense and antisense transcripts plus the involvement of RDRs and PolIV highly complicate the biogenesis mechanisms of nat-siRNAs. Nat-siRNA-mediated silencing is one of the regulatory mechanisms for controlling the expression of NATs, although the detection of nat-siRNAs is often very challenging due to their low abundances. More advanced sequencing technologies and improved small RNA detection techniques will highly benefit studies of *cis*-NATs and nat-siRNAs.

SUPPLEMENTARY DATA

Supplementary Data are available at *Molecular Plant Online.*

FUNDING

Research in Jin's Laboratory is supported by NSF Career Award MCB-0642843, NSF grant IOS-1257576, NIH grant R01GM093008, California Department of Food & Agriculture (CDFA) grant 23741, and California Citrus Board grant 5210–131.

ACKNOWLEDGMENTS

We apologize for not citing some publications owing to space limitations. No conflict of interest declared.

REFERENCES

Borsani, O., Zhu, J.H., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell. **123**, 1279–1291.

Cai, P., Hou, N., Piao, X., Liu, S., Liu, H., Yang, F., Wang, J., Jin, Q., Wang, H., and Chen, Q. (2011). Profiles of small non-coding RNAs in *Schistosoma japonicum* during development. PLoS Negl. Trop. Dis. **5**, e1256.

Carlile, M., Nalbant, P., Preston-Fayers, K., McHaffie, G.S., and Werner, A. (2008). Processing of naturally occurring sense/antisense transcripts of the vertebrate Slc34a gene into short RNAs. Physiol. Genomics. **34**, 95–100.

Carlile, M., Swan, D., Jackson, K., Preston-Fayers, K., Ballester, B., Flicek, P., and Werner, A. (2009). Strand selective generation of endo-siRNAs from the Na/phosphate transporter gene Slc34a1 in murine tissues. Nucleic Acids Res. **37**, 2274–2282.

Chapman, E.J., and Carrington, J.C. (2007). Specialization and evolution of endogenous small RNA pathways. Nat. Rev. Genet. **8**, 884–896.

Chekanova, J.A., Gregory, B.D., Reverdatto, S.V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., et al. (2007). Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome. Cell. **131**, 1340–1353.

Chen, D.J., Meng, Y.J., Ma, X.X., Mao, C.Z., Bai, Y.H., Cao, J.J., Gu, H.B., Wu, P., and Chen, M. (2010). Small RNAs in angiosperms: sequence characteristics, distribution and generation. Bioinformatics. **26**, 1391–1394.

Chen, D.J., Yuan, C.H., Zhang, J., Zhang, Z., Bai, L., Meng, Y.J., Chen, L.L., and Chen, M. (2012). PlantNATsDB: a comprehensive database of plant natural antisense transcripts. Nucleic Acids Res. **40**, D1187–D1193.

Chen, J., Sun, M., Hurst, L., Carmichael, G., and Rowley, J. (2005). Genome-wide analysis of coordinate expression and evolution of human *cis*-encoded sense-antisense transcripts. Trends Genet. **21**, 326–329.

Crampton, N., Bonass, W.A., Kirkham, J., Rivetti, C., and Thomson, N.H. (2006). Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. Nucleic Acids Res. **34**, 5416–5425.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., et al. (2008). An endogenous small interfering RNA pathway in Drosophila. Nature. **453**, U798–U797.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. (2000). An RNA-Dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell. **101**, 543–553.

Ding, S.W. (2010). RNA-based antiviral immunity. Nat. Rev. Immunol. **10**, 632–644.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. (2009). RNAi in budding yeast. Science. **326**, 544–550.

Ebralidze, A.K., Guibal, F.C., Steidl, U., Zhang, P., Lee, S., Bartholdy, B., Jorda, M.A., Petkova, V., Rosenbauer, F., Huang, G., et al. (2008). PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared *cis*-regulatory element. Genes Dev. **22**, 2085–2092.

Faghihi, M.A., and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nat. Rev. Mol. Cell Biol. **10**, 637–643.

Faghihi, M.A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M.P., Nalls, M.A., Cookson, M.R., St-Laurent, G., and Wahlestedt, C.

(2010). Evidence for natural antisense transcript-mediated inhibition of microRNA function. Genome Biology. **11**, R56.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C.J., Du, T.T., Lee, S., Xu, J., Kittler, E.L.W., Zapp, M.L., Weng, Z.P., et al. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science. **320**, 1077–1081.

Git, A., Dvinge, H., Salmon-Divon, M., Osborne, M., Kutter, C., Hadfield, J., Bertone, P., and Caldas, C. (2010). Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA. **16**, 991–1006.

Gunasekera, A.M., Patankar, S., Schug, J., Eisen, G., Kissinger, J., Roos, D., and Wirth, D.F. (2004). Widespread distribution of antisense transcripts in the *Plasmodium falciparum* genome. Mol. Biochem. Parasitol. **136**, 35–42.

Hafner, M., Renwick, N., Brown, M., Mihailović, A., Holoch, D., Lin, C., Pena, J.T., Nusbaum, J.D., Morozov, P., and Ludwig, J. (2011). RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. RNA. **17**, 1697–1712.

Hastings, M.L., Milcarek, C., Martincic, K., Peterson, M.L., and Munroe, S.H. (1997). Expression of the thyroid hormone receptor gene, erbAα, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. Nucleic Acids Res. **25**, 4296–4300.

He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008). The antisense transcriptomes of human cells. Science. **322**, 1855–1857.

Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA polymerase II collision interrupts convergent transcription. Mol. Cell. **48**, 365–374.

Jin, H. (2008). Endogenous small RNAs and antibacterial immunity in plants. FEBS Lett. **582**, 2679–2684.

Jin, H., Vacic, V., Girke, T., Lonardi, S., and Zhu, J.-K. (2008). Small RNAs and the regulation of *cis*-natural antisense transcripts in *Arabidopsis*. BMC Mol. Biol. **9**, 6.

Jin, Y.F., and Bian, T.F. (2004). Nontemplated nucleotide addition prior to polyadenylation: a comparison of *Arabidopsis* cDNA and genomic sequences. RNA. **10**, 1695–1697.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. Annu. Rev. Plant Biol. **57**, 19–53.

Kapranov, P., Ozsolak, F., Kim, S.W., Foissac, S., Lipson, D., Hart, C., Roels, S., Borel, C., Antonarakis, S.E., and Monaghan, A.P. (2010). New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism. Nature. **466**, 642–646.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al. (2005). Antisense transcription in the mammalian transcriptome. Science. **309**, 1564–1566.

Katiyar-Agarwal, S., and Jin, H.L. (2010). Role of small RNAs in host–microbe interactions. Annu. Rev. Phytopathol. **48**, 225–246.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Zhu, J.K., Staskawicz, B.J., and Jin, H.L. (2006). A pathogen-inducible endogenous siRNA in plant immunity. Proc. Natl Acad. Sci. U S A. **103**, 18002–18007.

Kawaji, H., Nakamura, M., Takahashi, Y., Sandelin, A., Katayama, S., Fukuda, S., Daub, C.O., Kai, C., Kawai, J., and Yasuda, J. (2008). Hidden layers of human small RNAs. BMC Genomics. **9**, 157.

Kumar, M., and Carmichael, G.G. (1997). Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. Proc. Natl Acad. Sci. U S A. **94**, 3542–3547.

Kurihara, Y., Schmitz, R.J., Nery, J.R., Schultz, M.D., Okubo-Kurihara, E., Morosawa, T., Tanaka, M., Toyoda, T., Seki, M., and Ecker, J.R. (2012). Surveillance of 3′ noncoding transcripts requires FIERY1 and XRN3 in *Arabidopsis*. G3: Genes| Genomes| Genetics. **2**, 487–498.

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

Lapidot, M., and Pilpel, Y. (2006). Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep. **7**, 1216–1222.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. **26**, 589–595.

Linsen, S.E., de Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R.K., Fritz, B., Wyman, S.K., de Bruijn, E., and Voest, E.E. (2009). Limitations and possibilities of small RNA digital gene expression profiling. Nat. Methods. **6**, 474–476.

Lu, T., Zhu, C., Lu, G., Guo, Y., Zhou, Y., Zhang, Z., Zhao, Y., Li, W., Lu, Y., and Tang, W. (2012). Strand-specific RNA-seq reveals widespread occurrence of novel *cis*-natural antisense transcripts in rice. BMC Genomics. **13**, 721.

Mallory, A., and Vaucheret, H. (2010). Form, function, and regulation of ARGONAUTE proteins. Plant Cell. **22**, 3879–3889.

Mallory, A.C., and Vaucheret, H. (2006). Functions of microRNAs and related small RNAs in plants. Nat. Genet. **38**, S31–S36.

Melnyk, C.W., Molnar, A., and Baulcombe, D.C. (2011). Intercellular and systemic movement of RNA silencing signals. EMBO J. **30**, 3553–3563.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. (2008). Specific expression of long noncoding RNAs in the mouse brain. Proc. Natl Acad. Sci. U S A. **105**, 716–721.

Meyers, B.C., Vu, T.H., Tej, S.S., Ghazal, H., Matvienko, M., Agrawal, V., Ning, J.C., and Haudenschild, C.D. (2004). Analysis of the transcriptional complexity of *Arabidopsis thaliana* by massively parallel signature sequencing. Nat. Biotechnol. **22**, 1006–1011.

Molnar, A., Melnyk, C., and Baulcombe, D.C. (2011). Silencing signals in plants: a long journey for small RNAs. Genome Biol. **12**, 215.

Morrissy, A.S., Griffith, M., and Marra, M.A. (2011). Extensive relationship between antisense transcription and alternative splicing in the human genome. Genome Res. **21**, 1203–1212.

Mosher, R.A., Melnyk, C.W., Kelly, K.A., Dunn, R.M., Studholme, D.J., and Baulcombe, D.C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*. Nature. **460**, U283–U151.

Okada, Y., Tashiro, C., Numata, K., Watanabe, K., Nakaoka, H., Yamamoto, N., Okubo, K., Ikeda, R., Saito, R., Kanai, A., et al. (2008). Comparative expression analysis uncovers novel features of endogenous antisense transcription. Hum. Mol. Genet. **17**, 1631–1640.

Okamura, K., Balla, S., Martin, R., Liu, N., and Lai, E.C. (2008). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. Nat. Struct. Mol. Biol. **15**, 581–590.

Okamura, K., Robine, N., Liu, Y., Liu, Q., and Lai, E.C. (2011). R2D2 organizes small regulatory RNA pathways in Drosophila. Mol. Cell. Biol. **31**, 884–896.

Ozsolak, F., Platt, A.R., Jones, D.R., Reifenberger, J.G., Sass, L.E., McInerney, P., Thompson, J.F., Bowers, J., Jarosz, M., and Milos, P.M. (2009). Direct RNA sequencing. Nature. **461**, 814–818.

Padmanabhan, C., Zhang, X., and Jin, H. (2009). Host small RNAs are big contributors to plant innate immunity. Curr. Opin. Plant Biol. **12**, 465–472.

Pikaard, C.S., Haag, J.R., Ream, T., and Wierzbicki, A.T. (2008). Roles of RNA polymerase IV in gene silencing. Trends Plant Sci. **13**, 390–397.

Prasanth, K.V., Prasanth, S.G., Xuan, Z.Y., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Egulating gene expression through RNA nuclear retention. Cell. **123**, 249–263.

Prescott, E.M., and Proudfoot, N.J. (2002). Transcriptional collision between convergent genes in budding yeast. Proc. Natl Acad. Sci. U S A. **99**, 8796–8801.

Reddy, A., Zheng, Y., Jagadeeswaran, G., Macmil, S., Graham, W., Roe, B., Desilva, U., Zhang, W., and Sunkar, R. (2009). Cloning, characterization and expression analysis of porcine microRNAs. BMC Genomics. **10**, 65.

Ron, M., Saez, M.A., Williams, L.E., Fletcher, J.C., and McCormick, S. (2010). Proper regulation of a sperm-specific *cis*-nat-siRNA is essential for double fertilization in *Arabidopsis*. Genes Dev. **24**, 1010–1021.

Ruiz-Ferrer, V., and Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. Annu. Rev. Plant Biol. **60,** 485–510.

Sherstnev, A., Duc, C., Cole, C., Zacharaki, V., Hornyik, C., Ozsolak, F., Milos, P.M., Barton, G.J., and Simpson, G.G. (2012). Direct sequencing of *Arabidopsis thaliana* RNA reveals patterns of cleavage and polyadenylation. Nat. Struct. Mol. Biol. **19**, 845–852.

Smalheiser, N.R., Lugli, G., Thimmapuram, J., Cook, E.H., and Larson, J. (2011). Endogenous siRNAs and noncoding RNA-derived small RNAs are expressed in adult mouse hippocampus and are up-regulated in olfactory discrimination training. RNA. **17**, 166–181.

Steigele, S., and Nieselt, K. (2005). Open reading frames provide a rich pool of potential natural antisense transcripts in fungal genomes. Nucleic Acids Res. **33**, 5034–5044.

Stolc, V., et al. (2005). Identification of transcribed sequences in *Arabidopsis thaliana* by using high-resolution genome tiling arrays. Proc. Natl Acad. Sci. U S A. **102**, 4453–4458.

Sun, M., Hurst, L.D., Carmichael, G.G., and Chen, J. (2006). Evidence for variation in abundance of antisense transcripts between multicellular animals but no relationship between antisense transcription and organismic complexity. Genome Res. **16**, 922–933.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature. **453**, U534–U538.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. **25**, 1105–1111.

Tufarelli, C., Stanley, J.A.S., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G., and Higgs, D.R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat. Genet. **34**, 157–165.

Voinnet, O., Vain, P., Angell, S., and Baulcombe, D.C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. Cell. **95**, 177–187.

Wang, X.J., Gaasterland, T., and Chua, N.H. (2005). Genome-wide prediction and identification of *cis*-natural antisense transcripts in *Arabidopsis thaliana*. Genome Biol. **6**, R30.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature. **453**, U539.

Werner, A., Carlile, M., and Swan, D. (2009). What do natural antisense transcripts regulate? RNA Biol. **6**, 43–48.

Werner, A., Schmutzler, G., Carlile, M., Miles, C.G., and Peters, H. (2007). Expression profiling of antisense transcripts on DNA arrays. Physiol. Genomics. **28**, 294–300.

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

Wu, X.H., Liu, M., Downie, B., Liang, C., Ji, G.L., Li, Q.Q., and Hunt, A.G. (2011). Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation. Proc. Natl Acad. Sci. U S A. **108**, 12533–12538.

Xu, N., Li, Y., Zhao, Y.-T., Guo, L., Fang, Y.-Y., Zhao, J.-H., Wang, X.-J., Huang, L., and Guo, H.-S. (2012). Identification and characterization of small RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. PLoS one. **7**, e353306.

Yamada, K., Lim, J., Dale, J.M., Chen, H.M., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., (2003). Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science. **302**, 842–846.

Zhang, W., Gao, S., Zhou, X., Chellappan, P., Chen, Z., Zhou, X., Zhang, X., Fromuth, N., Coutino, G., and Coffey, M. (2011). Bacteria-responsive microRNAs regulate plant innate immunity by modulating plant hormone networks. Plant Mol. Biol. **75**, 93–105.

Zhang, X., Xia, J., Lii, Y., Barrera-Figueroa, B., Zhou, X., Gao, S., Lu, L., Niu, D., Liang, W., Chen, Z., et al. (2012). Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. Genome Biol. **13**, R20.

Zhang, Y., Liu, X., Liu, Q., and Wei, L. (2006). Genome-wide *in silico* identification and analysis of *cis* natural antisense transcripts (*cis*-NATs) in ten species. Nucleic Acids Res. **34**, 3465–3475.

Zhou, X., Sunkar, R., Jin, H., Zhu, J.-K., and Zhang, W. (2009). Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in *Oryza sativa*. Genome Res. **19**, 70–78.

Zubko, E., and Meyer, P. (2007). A natural antisense transcript of the *Petunia hybrida* Sho gene suggests a role for an antisense mechanism in cytokinin regulation. Plant J. **52**, 1131–1139.