



Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses

Besma Ben Amor, Sonia Wirth, Francisco Merchan, et al.

Genome Res. published online November 7, 2008

Access the most recent version at doi:[10.1101/gr.080275.108](https://doi.org/10.1101/gr.080275.108)

P<P Published online November 7, 2008 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Comprehensive immune receptor profiling.
Discover the **DriverMap™ AIR Assay** difference.

LEARN
MORE



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © 2008, Cold Spring Harbor Laboratory Press

Title:

Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses

Besma Ben Amor^{1,*}, Sonia Wirth^{1,*,#}, Francisco Merchan¹, Philippe Laporte¹, Yves d'Aubenton-Carafa², Judith Hirsch^{3,##}, Alexis Maizel¹, Allison Mallory⁵, Antoine Lucas^{2,###}, Jean Marc Deragon⁴, Herve Vaucheret⁵, Claude Thermes² and Martin Crespi^{1,†}.

¹ Institut des Sciences du Végétal (ISV), CNRS, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette, France

² Centre de Génétique Moléculaire (CGM), CNRS, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette, France

³ BIOMEDAL S.L., Avda Américo Vespucio, 41092 Sevilla, Spain

⁴ Université de Perpignan Via Domitia, CNRS UMR5096 LGDP, 52 avenue Paul Alduy 66860 Perpignan Cedex France

⁵ INRA Laboratoire de Biologie Cellulaire, Route de St. Cyr, 78026 Versailles Cedex France

* Both authors have contributed equally to this work

† Corresponding author: Martin Crespi,

ISV, CNRS,

1 avenue de la terrasse,

91198 Gif-sur-Yvette cedex,

tel : ++ 33 1 69 82 37 03

fax : ++ 33 1 69 82 36 95

Present address: Laboratorio de Agrobiotecnología, Piso 2, Pabellón 2, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina.

Present address: INRA, Campus International de Baillarguet, UMR BGPI, TA A-54/K 34398, Montpellier, France.

Present address : Alten Ouest, 12i rue du Patis Tatelin, ZAC Saint-Sulpice, 35000 Rennes, France

Running title : long non-coding RNAs in *Arabidopsis*

Keywords: non-coding RNAs, small RNA precursors, microarray, *Arabidopsis* development, abiotic stresses

Abstract

Long non-protein coding RNAs (npcRNA) represent an emerging class of riboregulators, which act either directly in this long form or are processed to shorter miRNA and siRNA. Genome-wide bioinformatic analysis of full-length cDNA databases identified 76 *Arabidopsis* npcRNAs. Fourteen npcRNAs were antisense to protein-coding mRNAs, suggesting cis-regulatory roles. Numerous 24-nt siRNA matched to five different npcRNAs, suggesting that these npcRNAs are precursors of this type of siRNA. Expression analyses of the 76 npcRNAs identified a novel npcRNA that accumulates in a *dcl1* mutant but does not appear to produce trans-acting siRNA or miRNA. Additionally, another npcRNA was the precursor of miR869 and shown to be up-regulated in *dcl4* but not in *dcl1* mutants, indicative of a young miRNA gene. Abiotic stress altered the accumulation of 22 npcRNAs among the 76, a fraction significantly higher than that observed for the RNA binding protein-coding fraction of the transcriptome. Over-expression analyses in *Arabidopsis* identified two npcRNAs as regulators of root growth during salt stress and leaf morphology, respectively. Hence, together with small RNAs, long npcRNAs encompass a sensitive component of the transcriptome that have diverse roles during growth and differentiation.

Introduction

Non protein coding RNAs (npcRNAs) are a class of RNAs that do not encode proteins but instead their function lies on the RNA molecule. They are a heterogeneous group and have been divided into different classes according to their length and function. With respect to length, npcRNAs can range from 20 to 27 nucleotides for the families of microRNAs (miRNAs) and small interfering RNAs (siRNAs), 20–300 nucleotides for small RNAs commonly found as transcriptional and translational regulators, or up to and beyond 10,000 nucleotides for medium and large RNAs involved in other processes, including splicing, gene-inactivation and translation (Costa 2007). We use the term non-protein coding RNAs instead of non-coding RNAs as every sequence has the potential to be coding, and certain large npcRNAs might encode small oligopeptides, which could be translated under specific conditions as shown for a pentapeptide located inside rRNA, a canonical RNA in *E. coli* (Tenson et al. 1996). In recent years, numerous novel npcRNA candidates have been identified in a variety of organisms from *Escherichia coli* to *Homo sapiens* (Argaman et al. 2001; Storz et al. 2004; Washietl et al. 2005).

Several strategies have been employed to detect and discover novel npcRNAs, including both experimental and computational screenings (Huttenhofer et al. 2002). Genomic approaches, such as tiling arrays and systematic sequencing of full-length cDNA libraries, in model organisms have recently revealed that much larger portions of eukaryote transcriptomes represent non-protein-coding transcripts than previously believed (Okazaki et al. 2002; Numata et al. 2003; Rinn et al. 2003; Ota et al. 2004; Chekanova et al. 2007). Diverse npcRNAs, including a surprising number of antisense RNA transcripts, pseudogenes and truncated transcripts, have been described (Prasanth and Spector 2007). Certain npcRNAs, referred to as riboregulators, control the stability or translation of specific mRNAs and, in this way, regulate developmental events or stress responses in eukaryotic cells (Erdmann et al. 2001). As very few riboregulators involved in development were previously revealed by classical genetic approaches, it has been proposed that riboregulators may fine-tune mRNA levels in the cell and play a more critical role in the adaptation of developmental processes rather than in differentiation *per se*.

The most well-studied npcRNA species are single stranded 20- to 27-nucleotide small RNAs belonging to two classes, miRNAs and siRNAs, both known to have essential roles in the four eukaryote kingdoms (protists, fungi, plants and animals). In plants, miRNAs and siRNAs differ in their biogenesis, but both function by guiding target mRNA cleavage after integration into a ribonucleoprotein complex: the RISC (RNA Silencing Induced Complex) invariably containing a member of the AGO protein family (Vaucheret 2006). In contrast, most animal miRNAs appear to repress translation (Chapman and Carrington 2007). The miRNAs are single-stranded 21-nt RNA molecules deriving from partially-complementary RNA precursors, which are mainly transcribed by RNA polymerase II from intergenic regions, although few MIR genes are located in introns of protein-coding genes. It has been estimated that *MIRNA* genes could represent more than 1% of the expressed genome in worms and humans where it has been proposed that a single miRNA could regulate at least one hundred mRNA targets (Lim et al. 2005), underlining the relevance of this post-transcriptional regulatory mechanism. In *Arabidopsis*, 118 putative *MIRNA* loci have been identified (Jones-Rhoades et al. 2006), which represent 42 families, 20 families being conserved in other species such as rice (Bonnet et al. 2004; Sunkar et al. 2005) or poplar (Lu et al. 2005). DICER-LIKE 1 (DCL1) is the main enzyme responsible for mature miRNA production although DCL4 was shown to process at least two young MIRNA genes (Rajagopalan et al. 2006).

In plants, in addition to miRNAs, there exists a great diversity of siRNA: the heterochromatic siRNAs (hc-siRNAs), the trans-acting siRNAs (tasiRNAs) and the natural antisense siRNAs (nat-siRNAs). The common trait during the biogenesis of the first two siRNA species is the generation of a fully complementary double stranded RNA by the action of an RNA dependent RNA polymerase (RDR) (Vaucheret 2006). The highly diverse 24-nt hc-siRNAs are produced through the action of *DCL3* and *RDR2* and are linked to the formation of heterochromatin. The tasiRNAs derive from large npcRNAs that are targets of miRNAs. This miRNA-dependent processing generates shorter npcRNA molecules that are targeted by RDR6 to produce double-stranded RNAs (dsRNA) that are processed by DCL4 to generate 21-nt tasiRNAs that integrate into RISC complexes. In this way, novel tasiRNAs are generated from the action of non-homologous miRNAs, amplifying the diversity of small RNA regulated targets (Vaucheret

2006). The nat-siRNAs are generated from a pair of genes transcribed in antisense orientation generating natural double strand transcripts. They have been identified in *Arabidopsis* plants growing under high NaCl conditions where one overlapping gene is constitutively expressed and the other induced by salt stress. The resulting dsRNA molecules are cleaved by DCL2 and DCL1 enzymes to generate the nat-siRNA that cleave the mRNA of the constitutive gene and lead to salt tolerance (Borsani et al. 2005).

In contrast to small RNAs, much less is known about the large and diverse population of long npcRNAs. This heterogeneous class of transcripts generally does not contain any long ORF (e.g. no ORF coding more than 70 amino acids). Even though some of these genes may encode oligopeptides, it has been shown that their RNA moiety plays a critical role in their function. In order to achieve their function, they interact with proteins to regulate transcription, translation or mRNA stability (Yamashita et al. 1998; Campalans et al. 2004; Prasanth and Spector 2007; Filipowicz et al. 2008; Sasidharan and Gerstein 2008). Furthermore, several of these npcRNAs are precursors of miRNAs and tasiRNAs (Reinhart et al. 2002; Hirsch et al. 2006). Like some miRNAs, certain npcRNAs are induced in various developmental processes as well as during abiotic stress responses in plants and animals (Jones-Rhoades et al. 2006; Mendes Soares and Valcarcel 2006; Prasanth and Spector 2007; Sunkar et al. 2007). In *Caenorhabditis elegans*, 25 npcRNAs are regulated in seven developmental stages and two stimulated conditions (He et al. 2006). We have identified fifteen plant npcRNAs displaying diverse tissue-specific expression patterns and/or regulation by environmental stimuli (Hirsch et al. 2006).

In this work we performed a genome wide *in silico* screen of a full-length *Arabidopsis* cDNA library (Castelli et al. 2004) and identified 33 npcRNAs, including 13 that are antisense to protein-coding transcripts. Analyses of these new npcRNAs together with the 43 npcRNAs that we reported previously revealed that 34 potentially give rise to small RNAs and, consistently, we identified several novel mi/siRNA precursors from this group. A new npcRNA accumulated in *dcl1* mutants without forming known mi/siRNAs was identified, while abiotic stresses regulated the expression of 22 npcRNAs. Based on their expression profiles, stable secondary structures and/or sequence homologies, we selected 12 npcRNAs for functional studies in *Arabidopsis*

plants. Over-expression of two of these npcRNAs revealed roles in leaf differentiation and salt stress responses, respectively, suggesting that at least a subset of these newly identified npcRNAs has roles in developmental or stress adaptation programs.

RESULTS

Identification of new *Arabidopsis* npcRNAs

A number of npcRNAs detected in *Arabidopsis* databases could be truncated mRNAs, and, as such, further analyses are required to identify *bona fide* npcRNAs. Our previous (Hirsch et al. 2006) bioinformatic analyses, which filtered 172,495 ESTs and 24,985 mRNAs retrieved from NCBI databases, identified 43 npcRNA genes that shared certain functional elements, such as a high GC content, an atypical compositional skew (overrepresentation of T versus A) and the presence, in certain cases, of significantly stable RNA secondary structures. Using a similar approach, we analysed a collection of 18 000 full-length cDNAs (Castelli et al. 2004) and identified 22 additional candidate npcRNA genes, as well as another 11 npcRNAs that are antisense to annotated genes with an overlapping region of at least 50 bp (Supplemental Table 1 lists the 33 new npcRNAs and their names). Our 43 previously identified npcRNAs also were identified by our new analysis, although, in certain cases, we introduced small changes in their annotations, notably three were antisense to protein-coding mRNAs. The size of these 33 new npcRNAs ranges from 265- to 1879-nt with a mean of 1013-nt, and they harbour one or several introns in one third of the cases. We looked for potential ORF in the npcRNA, i.e. segments of RNA that, between a start and a stop codon, are multiples of three nucleotides. In 22 of them, the longest ORF (mean-100 nt) is preceded by 2 to 30 ATG codons rendering ribosome initiation at this ORF unlikely (Supplemental table 1). The majority of the selected npcRNAs displays a high GC content (mean GC%: 38%) and a significant abundance of T over A (skew: S already observed for the previous set of npcRNAs, calculated as in (Hirsch et al. 2006). Nineteen out of the 22 npcRNAs without an annotated antisense partner display a positive T/A skew further reinforcing their strand-specific transcription (Touchon et al. 2004). Highly significant Z-scores (Z-score >5, (Hirsch et al. 2006) suggest the presence of stable RNA secondary structures

for 5 of the new putative npcRNAs (Supplemental Fig. 1). BLASTN analysis revealed that 3 npcRNAs show significant nucleotide similarities independent of their encoded ORFs whereas 3 contain conserved sORFs (Supplemental table 1). When the sORF encoded in an npcRNA is much smaller than the similar protein, these genes may correspond to pseudogenes (indicated as pp in Supplemental table 1, 3 genes). Apart from antisense genes, we could not find any evidence of preferential silent codon nucleotide substitutions. Finally, npc511, npc513, npc520, npc530 and npc541 contain at least one 100 nucleotide box (conserved greater than 90%) in another region of the *A. thaliana* genome.

Nine npcRNAs correspond to miRNA, tasiRNA and 24-nt siRNA precursors

By mapping small RNA present in databases (Rajagopalan et al. 2006) to the complete collection of 76 npcRNAs (Supplemental Table 1 and (Hirsch et al. 2006), we revealed that 34 npcRNA loci give rise to at least one small RNA (Table 1 and Supplemental Fig. 2, npcRNA spanning small RNA loci). Some were already characterised small RNA precursors, such as the tasiRNA precursor *TAS3a* (npc41) or the *MIR162a* precursor (npc78) (Hirsch et al. 2006). The npc83 and npc521 corresponded to the miRNA precursors *MIR869a* and *MIR160c*, respectively. In addition, we identified new small RNA precursor candidates (npc34, npc351, npc375, npc520 and npc523). The majority of siRNAs deriving from these five npcRNAs are 24-nt long and map to both DNA strands of the npcRNA region, suggesting that these npcRNAs correspond to 24nt-small RNA precursors. Expression analysis in different organs (Fig. 1) revealed that npc351 is enriched in stems, npc34 in aerial parts whereas npc375 and npc523 show maximal levels in flowers. Specific oligonucleotides for real-time PCR studies could not be obtained for npc520. Whereas the npc375 and npc523 levels appeared unchanged in the *rdr2 rdr6* double mutant, the npc34 and npc351 were up-regulated in both seedlings and flowers of *rdr2 rdr6* (Fig. 1 right panel), suggesting that RDR2 or RDR6 produces a dsRNA from these npcRNAs, consistent with small RNA production from both DNA strands of these loci.

Transcriptomic analysis identifies a novel DCL4-processed young *MIRNA* gene and a novel npcRNA over-accumulating in *dcl1* mutants

In order to characterize further this set of npcRNAs, we constructed a “dedicated non-coding RNA biology” microarray (the “RIBOCHIP”) containing 274 different oligonucleotide probes (50-70-nt in length) corresponding to all 76 npcRNAs, the sense/antisense couples for those npcRNAs encoding antisense transcripts and a large collection of npcRNA-related RNA-binding proteins (see Supplemental table 2 for a complete list of genes, their identification and spotted oligonucleotides with their main characteristics). At first, the RIBOCHIP was hybridized with RNA from wild-type and *dcl1-9* inflorescences (Fig. 2). Among seven known miRNA precursors spotted on the RIBOCHIP, six (*MIR160c*, *MIR162a*, *MIR164b*, *MIR166a*, *MIR166b* and *MIR168a*) were up-regulated in *dcl1* inflorescences, consistent with DCL1-processed conserved miRNAs. In contrast, the non-conserved *MIR869a* precursor (npc83) did not over-accumulate in *dcl1*. This *MIRNA* precursor has an unusually stable secondary structure ((Hirsch et al. 2006), Z-score: 16.0) and other small RNAs are produced from the long stem-loop of this precursor (Fig. 3). RT-PCR analysis performed on other *dcl* mutants revealed that the *MIR869a* precursor over-accumulates in *dcl4* (Fig. 3C). These results strongly suggest that, like *MIR822* and *MIR839*, *MIR869a* is a young *MIRNA* gene and that its transcript is processed by DCL4 because it adopts a secondary structure closer to that of perfect dsRNA than that of a classic *MIRNA* precursor processed by DCL1 (Rajagopalan et al. 2006).

In addition to npcRNAs corresponding to conserved miRNAs, the npc531, which was not previously linked to RNA silencing mechanisms or known to encode a miRNA, also was upregulated in a *dcl1-9* mutant, a difference confirmed using quantitative RT-PCR (Fig. 4A, left panel). No known miRNAs map to this gene, suggesting that, alternatively, it could be the target of a miRNA. We identified a putative miR319 target site in npc531 (Fig. 4A right panel), but 5' RACE-PCR analyses did not reveal any specific cleavage of this transcript in this site (not shown).

Several npcRNAs are regulated by abiotic stresses

A deeper analysis of npcRNA accumulation was performed by hybridizing the RIBOCHIP array with RNA from wild-type leaves and roots, as well as roots of plants grown under stress conditions (phosphate starvation, salt stress, water stress) versus non stressed control plants. Of the 274 genes spotted on the RIBOCHIP, which include the 76 npcRNAs (Suppl. Table 1 and (Hirsch et al. 2006), 42 were differentially expressed in at least one of the conditions assayed. From these 42, 26 corresponded to npcRNAs and only 6 to putative or known RNA binding proteins (RBPs) (Table 2 and Figure 2). The remaining regulated genes corresponded to cis-antisense coding transcripts or positive controls. The observed bias towards npcRNAs (62% of the regulated genes were npcRNAs) points to a dynamic regulation of these functional RNAs over npcRNA-related protein-coding genes, which can be additionally controlled at translational and post-translational levels. Moreover, several of these genes were regulated in more than one condition suggesting pleiotropic roles (table 2).

Independent validations of the results observed in the microarray experiments were obtained for several npcRNAs using Northern analysis (Fig. 4B; that also confirmed the predicted size for npc43, npc72, npc60 and npc536) and quantitative RT-PCR (Figure 5). For the phosphate starvation assays, the up-regulation of npc43 and npc536 and the down-regulation of npc33 were confirmed (Figure 5a). We also observed a slight up- and down-regulation of npc60 and npc311, respectively, which also were detected in the microarray experiments (albeit with low statistical significance). Salt stress resulted in a dramatic 100-fold increase in npc60 accumulation. For the npc60, npc82 and npc536, the change in expression observed after 3h of NaCl treatment was maintained after 24h (Figure 5B), whereas for npc72 the induction was transient (Fig. 5B). Differential expression of *TAS3a*, npc43 and npc311 in roots and npc15 and npc156 in leaves was confirmed, as we previously reported using semi-quantitative RT-PCR (Hirsch et al. 2006). Additionally, the *MIR160c* precursor was specifically enriched in roots, as already shown (Wang et al. 2005a). Taken together, the combination of microarray and quantitative RT-PCR allowed us to identify 8 npcRNAs differentially expressed between roots and leaves and 15 npcRNAs regulated by diverse abiotic stresses.

Over-expression of two npcRNAs affects growth and differentiation in *Arabidopsis*

To explore the biological function of npcRNAs, we selected 12 candidates for further functional analyses. These candidates were chosen considering several criteria, which included conservation at the nucleotide level independent of encoded sORFs, the presence of statistically significant RNA secondary structures, tissue-specific expression patterns and stress-regulation (Table 3). As a first step to elucidate the function of these 12 npcRNAs, we over-expressed the complete cDNAs under the control of the strong constitutive cauliflower mosaic virus 35S promoter in *Arabidopsis*. Over-expression of 11 out of these 12 npcRNAs did not lead to obvious visible differences when compared to wild type plants in non-stressed growth conditions (a list of the over-expressed npcRNAs and the number of independent lines analyzed is indicated in Table 3). In contrast, T1 plants over-expressing the npc48 showed drastic developmental anomalies, including an increase in the rosette diameter, leaf serration and a delay in the flowering time compared with wild type plants (Fig. 6A). Quantitative RT-PCR analysis of npc48 accumulation revealed that all transgenic lines exhibiting this phenotype over-accumulated npc48 (Supplemental Figure 3). The phenotype of 35S::npc48 plants resembled that of 35S::MIR168 plants which have reduced levels of the miR168 AGO1 target and of AGO1-sensitive miRNAs such as miR166 (Vaucheret et al. 2006). 35S::npc48 plants also showed a decrease in miR166 accumulation, but accumulated AGO1 mRNA and miR168 at wild-type levels (Fig. 6B). In addition, 35S::npc48 plants, but not 35S::MIR168 plants, exhibited reduced accumulation of miR164 (Fig. 6B), suggesting that npc48 could regulate a subset of miRNAs. T-DNA insertions within npc48 are not available.

As the expression of several npcRNAs was affected by environmental stress, we explored the response to salt stress and phosphate starvation in lines over-expressing npc43, npc60, npc311 and npc536. 35S::npc536 transformants that actually overaccumulate npc536 (Supplemental Figure 4), displayed heightened root growth under salt stress conditions (Figure 7) compared to wild type plants. This increase is due to both primary root growth and secondary root length under salt stress (100 and 125 mM), whereas no differences were observed in the absence of salt treatment. No

phenotype could be observed in mutants that carry T-DNA insertions within npc536 (data not shown). The npc536 transcript is antisense to the At1g67930 mRNA (encoding a Golgi-transport complex related protein; Fig. 7C). The npc536 has a large dynamic range of expression across a wide range of tissue and hormonal, biotic or abiotic treatment, whereas At1g67930 is much less variable (AtGenExpress atlas, (Schmid et al. 2005). In response to certain abiotic stresses such as drought or cold treatment, the root expression of these two genes tends to be anti-correlated (pearson correlation coefficient of -0.636 and -0.983 respectively). However, accumulation of the At1g67930 transcript was not significantly modified in 35S::npc536 plants nor in npc536 mutants (Supplemental Figure 4). It remains possible that npc536 regulates the translation of At1g67930 mRNA. Altogether, over-expression of 2 of 12 npcRNAs affected *Arabidopsis* differentiation and growth responses to abiotic stresses.

DISCUSSION

Through this work we revealed a variety of *Arabidopsis* non-protein-coding transcripts, some of which likely regulate development and responses to abiotic stresses. Whole-genome mapping in the model plant *Arabidopsis*, based on the use of tiling arrays, revealed that >50% of observed transcription was intergenic and that numerous antisense RNA transcripts exist (Yamada et al. 2003). In *Drosophila*, related experiments based on tiling-arrays, performed at six developmental stages, detected RNA expression for 41% of the probes in intronic and intergenic regions (Stolc et al. 2004). In agreement, analysis of expressed sequence tag (EST) and full-length cDNA libraries identified many transcripts not previously assigned to genomic loci (Okazaki et al. 2002; Numata et al. 2003; Ota et al. 2004; Riano-Pachon et al. 2005; Hirsch et al. 2006). At least 13% and 26% of the unique full-length cDNAs in mice and humans, respectively, are thought to be poly (A) tail-containing mRNA-like npcRNAs. Since the functions of most of these npcRNAs are unknown, much work needs to be done to identify regulatory RNAs among these genes.

Using genome-wide mapping of full-length cDNAs, we describe 33 new npcRNAs that, when added to the 43 previously identified (Hirsch et al. 2006), make a set of 76

npcRNAs dispersed throughout the *Arabidopsis* genome. These npcRNAs belong to at least five classes:

- (i) Three npcRNAs correspond to MIRNA precursors such as the *MIR869a* precursor that overaccumulates in *dcl4* mutants, suggesting that *MIR869a* is a young MIRNA gene, similar to *MIR822* and *MIR839* (Rajagopalan et al. 2006).
- (ii) Five npcRNAs give rise to 24-nt siRNAs that map to both DNA strands, suggesting that these npcRNAs define 24-siRNA precursors. Among these five npcRNAs, two (*npc34* and *npc351*) over-accumulate in *rdr2/6* mutants, suggesting that these npcRNAs are likely RDR substrates or, alternatively, targets of RDR-dependent small RNAs such as the tasiRNAs.
- (iii) Five npcRNAs correspond to small RNA classes other than 21, 22 or 24 nt reminiscent of the class of 30-40-nt long-siRNAs (lsiRNAs) (Katiyar-Agarwal et al. 2007) recently shown to be involved in gene regulation in *Arabidopsis*.
- (iv) Fourteen npcRNAs corresponded to *cis* Natural Antisense Transcripts (*cis*-NATs), based on partial complementarity to other endogenous RNAs. The *cis*-NATs transcripts are derived from the same genomic loci as their sense counterparts, but from opposite strands. Genome-wide computational and experimental studies have shown that about 5% to 10% of gene transcripts in mammals and plants have *cis*-NATs (Wang et al. 2005b; Henz et al. 2007). cDNA sequence cluster analyses revealed that 7600 annotated genes in *Arabidopsis* (30%) had significant antisense expression (Yamada et al. 2003; Chekanova et al. 2007). Furthermore, characterization of 32 000 full-length rice cDNAs identified 600 antisense transcript pairs, half of which have no ORF in one member of the pair (Osato et al. 2003). Analysis of their regulation in comparison with their complementary transcripts indicates that there is a trend towards anticorrelated expression of *cis*-NAT pairs in *Arabidopsis*. However, currently available data do not produce a strong signature of small RNA mediated silencing for this process (Henz et al. 2007). The *Arabidopsis* transcriptome also contains a fairly large number of *trans*-NATs, around

1,320 putative *trans*-NAT pairs. Among those with available expression data, more than 85% were found in the same tissue as their sense partners (Wang et al. 2006). Several npcRNAs pairs of sense-antisense are co-regulated and potential NAT-based regulation systems in plants could be especially relevant for genes involved in developmental control or adaptive responses to changing environmental conditions. This assumption has recently been supported by the detection of NAT-specific small RNAs in plants that appear when antisense transcription is induced by salt stress (Borsani et al. 2005) or following pathogen attack (Katiyar-Agarwal et al. 2006). The npc536 forms a *cis*-NAT with At1g67930. Over-expression of npc536 allowed plants to grow under salt stress but did not modify At1g67930 mRNA accumulation. Furthermore, T-DNA mutants in this gene do not show mis-regulation of the antisense transcript. As this gene contains a sORF conserved in rice, it may act through this encoded peptide. Alternatively, it is possible that npc536 regulates translation of At1g67930 mRNA or that it acts as a *trans*-NAT with a yet to be identified gene that plays a role in salt stress.

- (v) The remaining npcRNAs did not correspond to known classes of small RNAs and were not antisense to known protein-coding RNAs. One possibility is that they act through the production of short peptides. Computational approaches based on an analysis of codon bias and cross species conservation suggest that approximately 5% of annotated genes in the genomes of yeasts, plants, flies, nematodes, mice and humans contain sORFs that could, in fact, be translated into exceptionally small peptides (Kastenmayer et al. 2006). Recently Kondo *et al.* (Kondo et al. 2007) reported that *polished rice (pri)*, which was identified previously as a gene for a non-coding RNA in *Drosophila*, is in fact transcribed into a polycistronic mRNA that contains evolutionarily conserved short ORFs that encode 11 or 32 amino acid-long peptides. The small PRI peptides seem to act non-cell autonomously to promote changes in epithelial-cell morphology. Hence, some Arabidopsis npcRNAs could encode sORFs that have regulatory roles. Alternatively, the RNA molecule may integrate

RNPs to determine the functional specificity of the complex (Prasanth and Spector 2007; Filipowicz et al. 2008) and orchestrate activities of other protein subunits during development (Brosius 2003; Wang et al. 2008). Other npcRNAs, such as *meiRNA* in yeast (*Saccharomyces cerevisiae*) and *enod40* in legume plants, are required for correct subcellular localization of RNP particles (Yamashita et al. 1998; Campalans et al. 2004). Finally, npcRNAs could also play roles in transcription-dependent mechanisms rather than being RNA-sequence dependent per se (e.g. genomic imprinting, intergenic transcripts in the *Drosophila Bithorax* complex (Prasanth and Spector 2007).

To explore the role of npcRNAs in *Arabidopsis* metabolism and physiology, we developed an array tool, the RIBOCHIP. This dedicated microarray contains 274 probes, representing not only the 76 identified npcRNAs, but also coding transcripts for many npcRNA-related proteins such as RNA-binding protein genes, RNA-metabolism genes, small RNA biogenesis pathways genes and other genes encoding proteins potentially linked to npcRNAs. Our RIBOCHIP analyses revealed a bias towards the regulation of npcRNAs over npcRNA-related protein-coding RNAs during stress conditions. One explanation for this bias is that npcRNAs, unlike protein coding RNAs, cannot be subjected to additional translational or post-translational controls that follow transcriptional regulations. Expression of many npcRNA candidates was regulated by growth conditions affecting root tissues, as seen for npcRNAs of the AtIPS1/At4 family (Franco-Zorrilla et al. 2002), and the *enod40* family in legumes (Campalans et al. 2004). We propose that npcRNAs are candidate regulators to adapt root growth and development to soil biotic and abiotic interactions.

Additional analysis was done by overexpression of 12 npcRNAs in *Arabidopsis*. The overexpression of each npcRNA was verified in order to discard any silencing effect due to transgene expression. Over-expression of npc48 led to leaf serration, a phenotype that has been observed in several *Arabidopsis* mutants such as *se* or *ago1*. SE is a general regulator of miRNA levels, affecting the processing of primary *MIRNA* precursors to miRNAs (Lobbes et al. 2006), while AGO1 acts as the major miRNA target slicing enzyme. Despite the resemblance of 35S::npc48 plants with hypomorphic *se* and

ago1 mutants, over-expression of *npc48* did not have a general effect on miRNAs or *AGO1* accumulation. However, *miR164* accumulation was significantly impaired. Interestingly, *mir164a* mutants display abnormal serrations of the leaf margins (Nikovics et al. 2006). Minor differences in global *miR164* accumulation were observed in *mir164a* mutants and localised mis-regulation of *miR164*-targets appeared critical for this phenotype. Nevertheless, the phenotype of *mir164a* mutants is milder than the strong leaf serrations induced by over-expression of *npc48*, suggesting that the reduction in *miR164* alone cannot account for the phenotype. Highly serrated leaves are a symptom of defects in cell proliferation along the margins of leaf primordia (March-Diaz et al. 2007) and the *npc48* might be linked to this process. Recently, it has been shown that the *npcRNA* At4 inhibits *in trans* the action of *miR399* through target mimicry (Franco-Zorrilla et al. 2007) without affecting *miR399* accumulation. No homology or complementarity was detected between *npc48* and any miRNA to suggest this. One possibility could be that *npc48* partially inhibit post-transcriptional regulation by miRNAs using another mechanism.

Inhibition of lateral root growth has been suggested to be an adaptive response to environmental stresses, notably drought and salt stress (Deak and Malamy 2005; Xiong et al. 2006). The hormone ABA has been implicated in this response although ABA independent pathways also exist. In addition, the cell layers surrounding the lateral root primordia conditioned the emergence of the lateral root, a process regulated by auxin signalling (Swarup et al. 2008). The overexpression of *npc536* could modify the expression of specific target genes in this cell layer or affect ABA responses in lateral roots in order to modulate the length of the lateral roots under a particular stress condition. This gene is also induced by phosphate starvation, another condition affecting root architecture and may play a role in the adaptation of root growth and development to the soil environment.

Although little is known about the biochemical activities of *npcRNAs*, an emerging hypothesis is that *npcRNAs* (long or small) incorporate into ribonucleoprotein particles (RNP) to determine their function and/or localisation. Additional analyses are necessary to reveal the molecular mechanisms that underlie the action of *npcRNAs* in eukaryotes. Our work offers new perspectives on the action of *npcRNAs*, which are a sensitive

component of the transcriptome, and reveals alternative riboregulatory mechanisms likely employed during plant growth and differentiation.

MATERIALS AND METHODS

Plant growth and RNA extraction

All experiments used the Columbia (Col0) ecotype of *Arabidopsis thaliana*. Seeds of the silencing-related mutants: *ago7-1*, *dcl1-9*, *rdr2/6*, *dcl4-2*, *dcl2-1* and *dcl3-1* have been described before (Gascioli et al. 2005; Adenot et al. 2006; Bouche et al. 2006; Vaucheret 2006). Plants were grown in long day conditions (16h light / 8h dark photoperiod) with 150 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of supplemental fluorescent light at 23°C.

For analysis of gene expression under salt stress, plants were grown in hydroponics in liquid 0.5X Murashige and Skoog (MS) salts (Sigma), 1% sucrose, during two weeks (until roots were 5-8 cm long). Then, the medium was replaced by fresh 0.5X MS (control) or 0.5X MS containing 150 mM NaCl. After 3h and 24h of incubation, roots of approximately 100-120 plants were separated from aerial part and pooled for RNA extraction. Phosphate (Pi) starvation treatments were performed according to Franco Zorrilla et al. (Franco-Zorrilla et al. 2002). Fourteen days after germination, roots of at least 100 plants were pooled and total RNA extracted. Water stress conditions were performed as described by Manavella et al. (Manavella et al. 2006). For analyzing stress responses, 45-50 plants of T1 segregating lines over-expressing an npcRNA were sterilized and germinated for two days and then transferred to a medium containing NaCl or not as indicated. After 14 days, primary and lateral root length for each plant is individually measured and photographed. All seedlings are transferred to the greenhouse and Basta selection and PCR-mediated genotyping determined homozygotes, heterozygotes and wild-type plants from the lot. Experiments were done per triplicate.

Inflorescences, stems, leaves of wild-type and mutant plants were collected from greenhouse-grown plants. Seedlings were grown *in vitro* for three weeks. Total RNA was extracted using Trizol (Invitrogen, Cergy-Pontoise, France) according to the manufacturer's instructions and further purified by passage through RNeasy columns (Qiagen, Courtaboeuf, France) and residual genomic DNA removed by on-column DNase I treatment (Qiagen, Courtaboeuf, France).

Bioinformatic analysis of npcRNAs

To identify candidate npcRNAs from the "full length" sequenced cDNA databank (Castelli et al. 2004), RNAs were first mapped with Sim4 (Florea et al. 1998) on the genome and selected by the length of the longest ORF (<210 nt). About 500 candidates were further tested for lack of significant Blast hits against NCBI nr protein databank. Mapping of other ESTs and cDNAs on these genomic regions did not bridge the gap with already annotated neighboring genes. Mapping of short RNAs (sRNAs) on the total set of 76 npcRNA was performed using a large dataset of massively sequenced *A. thaliana* sRNAs (Rajagopalan et al. 2006). Potential miRNAs regulating npc531 were searched in miRBASE database. Sequence comparisons were performed using BLASTN for nucleotide sequences and BLASTX for encoded sORFs sur TAIR, NCBI non-redundant and chromosome databases. Similarities between nucleotide (BLASTN) and sORF-encoded peptide (BLASTX) were searched up to an expected value e-05. Boxes of at least 21 nucleotides conserved between number of species N independent of the encoded ORFs are indicated as N(+). Conserved sORFs are indicated as N(XX) where N is the number of species and XX is the size of the sORF in *A. thaliana*. Those cases where the sORF show similarity with a larger ORF (more than 100 aa) in other species are indicated as putative pseudogenes (pp). Exceptionally stable RNA secondary structures were predicted as in (Hirsch et al. 2006). The sequences were scanned using sliding windows of increasing lengths. Briefly, for each window, the free-energy of folding of the most stable structure was computed using rnaFold from the Vienna Package (<http://www.tbi.univie.ac.at/~ivo/RNA/>) and compared to the distribution of free-energy of folding of random sequences similar in size and nucleotide composition. The Z score, defined as the difference between these free-energies divided by the standard deviation, was calculated and sequence regions with a Z score > 5 were displayed in Supplemental Fig. 1.

RIBOCHIP array design

Oligonucleotide primers were designed using OligoWiz 2.0 (Wernersson and Nielsen 2005). Each 45-60mer oligonucleotide probe was printed as triplicates on Quantifoil QMT epoxy slides in 16 grids in a 4 metacolumns x 4 metarows configuration with a 9

columns x 7 rows spot pattern for each sub-array by the Institute of Genome Research, Center of Biotechnology at Bielefeld University, Germany. Additionally, the array contains probes specifically designed as labeling and hybridization controls (positive) corresponding to 9 *A. thaliana* genes AT5G44200 (nuclear cap-binding protein), AT1G07920 (EF1 alpha), AT5G03240 (UBQ3), AT4G05320 (UBQ10) AT3G13920 (eIF-4A1), AT1G19910 (V-type H⁺ ATPase 16kD subunit) M21415 (beta-tubulin-4), AT1G49240 (actin 8 protein) and AT3G18780 (actin 2 protein). At least two independent biological replicates were performed for each condition analyzed in the RIBOCHIP.

Target preparation for RIBOCHIP array.

Two micrograms of total RNA from each biological sample (treatments and controls) were amplified using the Amino Allyl Message Amp II aRNA Amplification Kit (Ambion, Courtaboeuf, France), following the protocol provided by the manufacturer. Ten micrograms of amplified RNA (aRNA) were chemically labeled with Cy5 (Q15108, Amersham, Saclay, France), monofunctional reactive dye or Cy3 (Q13108, Amersham, Saclay, France), monofunctional reactive dye in coupling buffer (Ambion, Courtaboeuf, France). Labeled aRNA was purified from non-coupled dyes using the columns supplied by the Amino Allyl Message Amp II aRNA Amplification Kit.

Array hybridization

Before use, the printed slides were processed according to the manufacturer's recommendations (Institute of Genome Research, Center of Biotechnology, Bielefeld University) and then hybridized with 200 pmol each of Cy5- and Cy3-labeled targets (treatment and control) in 5X SSC (NaCl 0.75M, sodium citrate 75mM pH7), 25% formamide, 0.1% SDS at 42°C for 60h. For each slide 20µg of salmon sperm DNA and 4µg of yeast tRNA (Sigma) were added as blocking agents. After hybridization slides were washed once at 42°C in 1X SSC, 0.2% SDS for 4 min followed by a wash at room temperature in 1X SSC, 0.2% SDS for 4mins and twice of 4 min each in 0.1X SSC. A quick final wash at RT in 0.05X SSC for 1min was done to remove any salt residue and particles bind to the slides.

Slides were processed using an Axon 4000B microarray scanner (Axon Instruments) with 635 nm and 532 nm lasers and a scan resolution of 10 µm to generate TIFF images

on each channel. PMTs for lasers were automatically set by software for each slide (GenePix 6.0, Axon Instruments).

Microarray data processing and normalization

Quantification of images on 635nm and 532nm channels was performed with GenePix 6.0 software (Axon Instruments). Normalization was performed by the Loess and median by block method using MANGO software developed at the Gif-Orsay DNA micro-array platform (Centre de Génétique Moléculaire, CNRS, Gif-sur-Yvette campus, France). Replicate spots outliers were detected by Grubb's test. Statistical significance of regulated probes was analyzed by Newton's, Sapir and Churchill's and Chen's single slide methods.

Gene expression analysis

For real-time RT-PCR, cDNA was synthesized by reverse transcription performed on 1.5µg of total RNA using the SuperScript II Reverse Transcriptase (Invitrogen, Cergy Pontoise, France) and (T)16 A/G/C oligonucleotides. Real-time RT-PCR was performed on the Roche Light Cycler instrument using SYBR Green I dye (LightCycler FastStart DNA MasterPLUS SYBR Green I, Roche) using the primer pairs listed in Supplemental Table 3. Technical triplicates were done for each assay and at least two independent biological replicates were assayed.

5'RACE-PCR assays to detect npcRNA 531 cleavage products and Northern analysis of small RNAs were performed as described in (Hirsch et al. 2006).

Construction of *A. thaliana* npcRNA over expression lines

The coding sequences for each npcRNA were amplified by PCR from genomic *A. thaliana* Col0 DNA using the primers showed in Supplemental table 4. The full length DNAs were cloned into pENTR/D-TOPO (Invitrogen, Cergy Pontoise, France) and then in the pB7WG2 binary plasmid (Karimi et al. 2002), under the control of CaMV 35S promoter, using the gateway technology (Invitrogen, Cergy Pontoise, France) to be introduced into *Agrobacterium tumefaciens* (agl-o). These constructions were used to transform *Arabidopsis thaliana* ecotype 'Columbia' plants, by floral dip (Bechtold and Pelletier 1998). Transgenic plants were selected by spraying seedlings from T1

generation 7, 9, and 11 days after germination with a solution of 0.01% BASTA (200 g/l glufosinate ammonium). Basta-resistant plantlets (T1) were then grown in soil under long-day conditions (16 hours light). Expression levels of the transgene in the T1 and T2 generations were determined by qRT-PCR.

Acknowledgements

We would like to thank Raquel Chan, Argentina, for water stress RNA samples and Javier Paz Ares, Spain for helpful discussions. P.L. is the recipient of a fellowship from the “Ministère de l’Education Nationale et de la Recherche” (MENR), France whereas S.W. was a fellow from the “Fundacion Foro de Biotecnologia”, Argentina. F.M. was supported by the “Ministerio de Educacion y Ciencia”, Spain. This work was essentially supported by the European FP6 RIBOREG project N°: LSHG503 022 and partly by the French ANR-Genoplante RIBOROOT project.

Figure legends

Figure 1. *Expression patterns of four npcRNAs corresponding to new 24-nt siRNA precursors.* Left panels, the grey boxed region (above or below are Watson and Crick strands, respectively) indicate the npcRNA gene and transcript (black line) on the *Arabidopsis* genomic DNA coordinates. Small RNAs identified in public databases are indicated by small bands on both strands. The size of each siRNA is indicated in Supplemental Figure 2, npcRNAs spanning small RNA loci (following a color code). Central panels indicate expression of these npcRNAs in different tissues (Roots, Stems, Flowers, Cauline leaves and Rosette leaves) determined by qRT-PCR. Relative expression levels were normalized with ACT2 (At3g18780) and values for roots were arbitrarily fixed to 1 and standard deviations are shown. Right panels indicate expression in flowers or seedlings of Col0 and a *rdr2/6* or *ago7* mutant. Values for Col0 were fixed to 1.

Figure 2. *Regulation of genes in the RIBOCHIP.* A) Total number of genes that were either induced (red bars) or repressed (green bars) in each analyzed condition. Conditions include expression in roots versus leaves (Roots), roots of phosphate starved plants (-P), plants under water stress (-H₂O), roots of plants treated with 150mM NaCl (NaCl), and inflorescences of *dcl1-9* mutants (*dcl1*). B) Heat map showing expression level of each individual gene summarized in part a. *: npcRNAs identified by Hirsch et al. (Hirsch et al. 2006), **: npcRNAs identified in the present work. Genes without asterisks correspond to additional regulated genes from the RIBOCHIP (annotation is indicated: coding sense counterparts of npcRNAs, RNA-binding proteins, MIRNA precursors and control genes).

Figure 3. *The npc83 is the miR869a-precursor and is up-regulated in dcl4 mutants.* The npc83 transcript encodes several small RNAs, notably miR869, (A) that are produced from a highly stable RNA-secondary stem structure (Z-score: 16). The position of the miRNA and detected siRNAs on the stem-loop of this npcRNA is indicated in (B). Expression of the *MIR869a* precursor was analyzed in *dcl1-9*, *dcl2*, *dcl3* and *dcl4*

mutants and Col0 seedlings by real-time RT-PCR. Data was normalized with ACT2 (At3g18780) and values for Col0 were arbitrarily fixed to 1. Two biological replicates gave similar results and a representative example (standard deviations of technical replicates) is shown.

Figure 4. *Expression analysis of npcRNAs.* (A) *Accumulation of npc531 transcripts in dcl1 mutants.* Real time RT-PCR expression analysis of npc531 in wild type and *dcl1-9* inflorescences (a) was normalized with the ACT2 (At3g18780) gene and values for wild type were arbitrarily fixed to 1. On the right, predicted pairing between npc531 and miR319a. (B) *Regulation of npc43, npc60, npc72 and npc536 expression.* Northern analysis was performed on the indicated npcRNAs under different stress conditions (plants grown in phosphate starvation or under salt stress).

Figure 5. *Regulation of specific npcRNAs in different stress conditions.* Real time RT-PCR expression analysis of indicated npcRNAs in roots of plants grown in phosphate starvation (A) and after 3h and 24h in 150 mM NaCl (B). In every case data was normalized with ACT2 (At3g18780) and values for non-treated or wild type controls were arbitrary fixed to 1. For each cDNA synthesis, quantifications were made in triplicate and two biological replicates were analyzed. Values are means $\pm$ standard deviations.

Figure 6. *Phenotypic and molecular analysis of transgenic plants over-expressing npc48.* (A) Phenotype of *Arabidopsis* plants transformed with 35S::npc48 construct. A wild type (WT) plant (left) and two representative T1 35S::npc48 plants (OEnpc48) from independent lines displaying a characteristic serrated leaf phenotype at different growth stages (central photos) are shown. Rosette leaves from WT and OEnpc 48 plants (lower left panel). Image of whole plants (right-most panel) illustrates the delayed flowering of OEnpc48 plants. (B) Expression analysis of npc48 and AGO1 mRNA from leaves of two npc48 over-expressing lines and control WT using real-time RT-PCR (left panel). Northern blot analyses of miR164, miR166, miR168 and U6 RNAs (right panel). Quantification is indicated setting the Col0 value arbitrarily to 1 and normalizing to U6 values.

Figure 7. *Phenotypic and molecular analysis of transgenic plants overexpressing npcRNA 536.* Phenotype of *Arabidopsis* plants transformed with 35S::npc536 constructs (Oenpc536). (A) Control and a representative OEnpc536 transgenic plants grown under the indicated salt stress conditions (0mM, 100mM and 125 mM). (B) Quantification of primary root length and lateral root/primary root lengths in two independent transgenic OEnpc536 and control lines grown in 100 mM salt ($n < 45$ per experiment). Standard deviations are indicated. These differences are statistically significant (t-test 7.10^{-7}), whereas no differences could be detected under normal growth conditions. (C) Schematic of the genomic positions and transcripts deriving from the npc536 (At1g67920, white box) and At1g67930 (dark grey box) loci. The nucleotide sequence of the overlapping region (light grey) between the two transcripts is indicated.

Table legends.

Table 1. The *Arabidopsis* npcRNA collection and their matching small RNAs. Table shows npcRNA genes, annotation (name), antisense transcripts in this locus, gene position, length, chromosome, strand (Watson or Crick), number of exons, Z-score as well as the number of small RNA hits matching to either the sense or antisense npcRNA molecule (relative to the transcription direction). npcRNAs from npc2 to npc431 are derived from (Hirsch et al. 2006) whereas npcRNAs from npc502 to 541 are from this study.

Table 2. Analysis of gene regulation using the RIBOCHIP. Table shows genes spotted in the RIBOCHIP that were regulated (induced, positive numbers or repressed, negative numbers) in each analyzed condition: roots versus leaves (Roots), roots of phosphate starved plants (-P), plants under water stress (-H₂O), roots of plants treated with 150mM NaCl (NaCl), and inflorescences of *dcl1-9* mutants (*dcl1*). Numbers indicate fold-change induction (positive) or repression (negative) respect to several constitutive controls.

Table 3. Characteristics of selected candidate genes for functional analysis. Columns contain the name of candidate genes, sequence similarities (+, nucleotide similarities in N species independent of encoded sORFs, N(XX) number of species containing a conserved sORF of XX amino-acids in *Arabidopsis*, see Methods); the values of the Z scores (statistically significant stable RNA secondary structure were selected, Z scores >5); tissue specificity (+: preferentially expressed), regulation in response to stress condition (+: up regulated; P: phosphate starvation; H₂O: Water stress; NaCl: salt stress, *dcl1-9*: *dicer-like1* mutant) and their capacity to encode small RNAs (YES or NO).

Supplemental Figure 1. Proposed RNA secondary structures for the new npcRNAs, 511, 520, 521 (precursor of miR160c) and 524. Transcripts showing a stable structure with $Z > 5$ are depicted. For a better display, long secondary RNA structures are divided in parts to be stacked through common letters (A or B).

Supplemental Figure 2: npcRNAs spanning small RNA loci. Graphs showing the position of small RNAs in relation to gene transcript position for the 34 npcRNAs containing at least one small RNA. Colour codes are: blue (21-nt), red (24-nt), green (22-nt) and black (other size). The npcRNA gene is indicated in yellow and transcript in brown on the coordinate positions of the *Arabidopsis* genome. For reference, npc41 corresponds to TAS3, npc78 to the miR162a-precursor (Hirsch et al. 2006) and npc521 to miR160c.

Supplemental Figure 3. Real-time RT-PCR expression analysis of npc48 in various over-expressing transgenic lines either displaying early flowering and serrated leaves or showing no obvious developmental defects. Col0 values (first column) were arbitrarily fixed to 1.

Supplemental Fig. 4. Expression levels of npc536 and the overlapping At1g67930 gene in insertional T-DNA mutants of npc536 (TDNA21812-2 and 6) and in 3 independent overexpressing lines (L1, L2 and L5) as compared to Col 0 controls. Fold-induction was normalised to Col 0 controls.

Supplemental Table I. List of newly identified npcRNAs based on a full-length cDNA library (from 502 to 541, GSLT reference name according to (Castelli et al. 2004)). We include npcRNA name, number of ATGs before the initial ATG, size of the longest sORF and exon positions. Similarities between nucleotide (BLASTN) and sORF-encoded peptide (BLASTX) were searched up to an expected value e^{-05} . Boxes of at least 21 nucleotides conserved between the number of species N independent of the encoded ORFs are indicated as $N(+)$. Conserved sORFs are indicated as $N(XX)$ where N is the number of species and XX is the size of the sORF in *A. thaliana*. Those cases where the sORF show similarity with a larger ORF (more than 100 aa) in other species are indicated as putative pseudogenes (pp).

Supplemental Table 2. List of oligos used for the RIBOCHIP dedicated microarray, gene number (when relevant) with a short description of each gene, oligonucleotides properties and sequence.

We selected *Arabidopsis* genes coding for 129 RNA-binding proteins selected on the basis of their involvement in the production, maturation, stability and function of non-protein coding RNAs in plants and/or in other species. First, we searched for *Arabidopsis* genes coding for La and La-like proteins, involved in many aspects of non-protein coding RNA metabolism and present in a wide range of eukaryotes (Maraia 2001; Maraia and Intine 2001; Wolin and Cedervall 2002). From yeast to human, genetic and biochemical studies have shown that La protects non-protein coding RNAs from 3'-5' exonucleases and also fulfils an RNA chaperone activity involved in the assembly of several corresponding RNPs (Belisova et al. 2005; Fleurdepine et al. 2007). In *Arabidopsis*, two distantly related proteins with all structural characteristics of genuine La proteins have been characterized (At1g79880, At4g32720) and the La-motif is also found in a number of La-like proteins (At5g46250, At3g19090, At2g43970, At5g66100, At4g35890, At5g21160) (He et al. 2008). Second, we search for *Arabidopsis* genes coding for dsRBM (double-stranded RNA-binding motif) proteins. This motif is found in a diverse array of proteins that regulate npcRNA metabolism and function. Several of these proteins are known to be involved in small RNA pathways; the DICERs (At1g01040, At3g03300, At3g43920 and At5g20320), the DRBs (At1g09700, At2g28380, At3g26932, At3g62800 and At5g41070) and the RNA methylases (At4g20910 and At4g20920). Other dsRBM proteins are the DRB-like (At1g80650 and At4g00420), the RNases III (At3g20420 and At5g45150), the Helicases (At1g48650 and At5g04895) and the Phosphatases (At4g21670 and At5g01270). All *Arabidopsis* proteins containing one or several dsRBMs were selected. Third, we choose *Arabidopsis* SRP9 and SRP14 genes, as these two proteins are known to bind the 7SL npcRNA in all species and SINE npcRNAs in mammals (Chang et al. 1996) (At3g49100, At2g43640). Fourth, we selected *Arabidopsis* PAB protein genes as these proteins can bind to the poly(A) region of mRNAs and possibly to the equivalent region of long polIII npcRNAs (At2g36660, At1g22760, At3g16380, At4g34110, At1g71770, At1g34140, At2g23350 and At1g49760, (Kondrashov et al. 2005). Fourth, we selected several genes that had been linked to the action of tasi/miRNAs (such as AGO, HYL, HEN and related proteins; (Vaucheret 2006)) or RRM-containing proteins linked to the action of specific npcRNAs (homologues of ScMEI2 and MtRBP1 (Campalans et al. 2004; Jeffares et al. 2004)). Finally, we include several regions containing known miRNAs

(e.g. MIR164b, MIR166a, MIR166b, MIR168a), as well as several positive and negative controls.

Supplemental Table 3. List of oligos used in qRT-PCR studies.

Supplemental Table 4. List of oligos for the construction of npcRNA overexpression transgenic lines.

References

- Adenot, X., T. Elmayan, D. Lauressergues, S. Boutet, N. Bouche, V. Gascioli, and H. Vaucheret. 2006. DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7. *Curr Biol* 16: 927-932.
- Argaman, L., R. Hershberg, J. Vogel, G. Bejerano, E.G. Wagner, H. Margalit, and S. Altuvia. 2001. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* 11: 941-950.
- Bechtold, N. and G. Pelletier. 1998. In planta *Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Methods Mol Biol* 82: 259-266.
- Belisova, A., K. Semrad, O. Mayer, G. Kocian, E. Waigmann, R. Schroeder, and G. Steiner. 2005. RNA chaperone activity of protein components of human Ro RNPs. *RNA* 11: 1084-1094.
- Bonnet, E., J. Wuyts, P. Rouze, and Y. Van de Peer. 2004. Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *Proc Natl Acad Sci U S A* 101: 11511-11516.
- Borsani, O., J. Zhu, P.E. Verslues, R. Sunkar, and J.K. Zhu. 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* 123: 1279-1291.
- Bouche, N., D. Lauressergues, V. Gascioli, and H. Vaucheret. 2006. An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J* 25: 3347-3356.
- Brosius, J. 2003. Gene duplication and other evolutionary strategies: from the RNA world to the future. *J Struct Funct Genomics* 3: 1-17.
- Campalans, A., A. Kondorosi, and M. Crespi. 2004. Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell* 16: 1047-1059.
- Castelli, V., J.M. Aury, O. Jaillon, P. Wincker, C. Clepet, M. Menard, C. Cruaud, F. Quetier, C. Scarpelli, V. Schachter et al. 2004. Whole genome sequence comparisons and "full-length" cDNA sequences: a combined approach to evaluate and improve *Arabidopsis* genome annotation. *Genome Res* 14: 406-413.
- Chang, D.Y., K. Hsu, and R.J. Maraia. 1996. Monomeric scAlu and nascent dimeric Alu RNAs induced by adenovirus are assembled into SRP9/14-containing RNPs in HeLa cells. *Nucleic Acids Res* 24: 4165-4170.
- Chapman, E.J. and J.C. Carrington. 2007. Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet* 8: 884-896.
- Chekanova, J.A., B.D. Gregory, S.V. Reverdatto, H. Chen, R. Kumar, T. Hooker, J. Yazaki, P. Li, N. Skiba, Q. Peng et al. 2007. Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome. *Cell* 131: 1340-1353.
- Costa, F.F. 2007. Non-coding RNAs: lost in translation? *Gene* 386: 1-10.
- Deak, K.I. and J. Malamy. 2005. Osmotic regulation of root system architecture. *Plant J* 43: 17-28.
- Erdmann, V.A., M.Z. Barciszewska, M. Szymanski, A. Hochberg, N. de Groot, and J. Barciszewski. 2001. The non-coding RNAs as riboregulators. *Nucleic Acids Res* 29: 189-193.

- Filipowicz, W., S.N. Bhattacharyya, and N. Sonenberg. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 9: 102-114.
- Fleurdepine, S., J.M. Deragon, M. Devic, J. Guilleminot, and C. Bousquet-Antonelli. 2007. A bona fide La protein is required for embryogenesis in *Arabidopsis thaliana*. *Nucleic Acids Res* 35: 3306-3321.
- Florea, L., G. Hartzell, Z. Zhang, G.M. Rubin, and W. Miller. 1998. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res* 8: 967-974.
- Franco-Zorrilla, J.M., A.C. Martin, R. Solano, V. Rubio, A. Leyva, and J. Paz-Ares. 2002. Mutations at CRE1 impair cytokinin-induced repression of phosphate starvation responses in *Arabidopsis*. *Plant J* 32: 353-360.
- Franco-Zorrilla, J.M., A. Valli, M. Todesco, I. Mateos, M.I. Puga, I. Rubio-Somoza, A. Leyva, D. Weigel, J.A. Garcia, and J. Paz-Ares. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* 39: 1033-1037.
- Gascioli, V., A.C. Mallory, D.P. Bartel, and H. Vaucheret. 2005. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr Biol* 15: 1494-1500.
- He, H., L. Cai, G. Skogerbo, W. Deng, T. Liu, X. Zhu, Y. Wang, D. Jia, Z. Zhang, Y. Tao et al. 2006. Profiling *Caenorhabditis elegans* non-coding RNA expression with a combined microarray. *Nucleic Acids Res* 34: 2976-2983.
- He, N., N.S. Jahchan, E. Hong, Q. Li, M.A. Bayfield, R.J. Maraia, K. Luo, and Q. Zhou. 2008. A La-Related Protein Modulates 7SK snRNP Integrity to Suppress P-TEFb-Dependent Transcriptional Elongation and Tumorigenesis. *Mol Cell*.
- Henz, S.R., J.S. Cumbie, K.D. Kasschau, J.U. Lohmann, J.C. Carrington, D. Weigel, and M. Schmid. 2007. Distinct expression patterns of natural antisense transcripts in *Arabidopsis*. *Plant Physiol* 144: 1247-1255.
- Hirsch, J., V. Lefort, M. Vankersschaver, A. Boualem, A. Lucas, C. Thermes, Y. d'Aubenton-Carafa, and M. Crespi. 2006. Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts. *Plant Physiol* 140: 1192-1204.
- Huttenhofer, A., J. Brosius, and J.P. Bachellerie. 2002. RNomics: identification and function of small, non-messenger RNAs. *Curr Opin Chem Biol* 6: 835-843.
- Jeffares, D.C., M.J. Phillips, S. Moore, and B. Veit. 2004. A description of the Mei2-like protein family; structure, phylogenetic distribution and biological context. *Dev Genes Evol* 214: 149-158.
- Jones-Rhoades, M.W., D.P. Bartel, and B. Bartel. 2006. MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol* 57: 19-53.
- Karimi, M., D. Inze, and A. Depicker. 2002. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci* 7: 193-195.
- Kastenmayer, J.P., L. Ni, A. Chu, L.E. Kitchen, W.C. Au, H. Yang, C.D. Carter, D. Wheeler, R.W. Davis, J.D. Boeke et al. 2006. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res* 16: 365-373.
- Katiyar-Agarwal, S., S. Gao, A. Vivian-Smith, and H. Jin. 2007. A novel class of bacteria-induced small RNAs in *Arabidopsis*. *Genes Dev* 21: 3123-3134.
- Katiyar-Agarwal, S., R. Morgan, D. Dahlbeck, O. Borsani, A. Villegas, Jr., J.K. Zhu, B.J. Staskawicz, and H. Jin. 2006. A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci U S A* 103: 18002-18007.

- Kondo, T., Y. Hashimoto, K. Kato, S. Inagaki, S. Hayashi, and Y. Kageyama. 2007. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat Cell Biol* 9: 660-665.
- Kondrashov, A.V., M. Kieffmann, K. Ebnet, T. Khanam, R.S. Muddashetty, and J. Brosius. 2005. Inhibitory effect of naked neural BC1 RNA or BC200 RNA on eukaryotic in vitro translation systems is reversed by poly(A)-binding protein (PABP). *J Mol Biol* 353: 88-103.
- Lim, L.P., N.C. Lau, P. Garrett-Engele, A. Grimson, J.M. Schelter, J. Castle, D.P. Bartel, P.S. Linsley, and J.M. Johnson. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433: 769-773.
- Lobbes, D., G. Rallapalli, D.D. Schmidt, C. Martin, and J. Clarke. 2006. SERRATE: a new player on the plant microRNA scene. *EMBO Rep* 7: 1052-1058.
- Lu, S., Y.H. Sun, R. Shi, C. Clark, L. Li, and V.L. Chiang. 2005. Novel and mechanical stress-responsive MicroRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell* 17: 2186-2203.
- Manavella, P.A., A.L. Arce, C.A. Dezar, F. Bitton, J.P. Renou, M. Crespi, and R.L. Chan. 2006. Cross-talk between ethylene and drought signalling pathways is mediated by the sunflower Hahb-4 transcription factor. *Plant J* 48: 125-137.
- Maraia, R.J. 2001. La protein and the trafficking of nascent RNA polymerase iii transcripts. *J Cell Biol* 153: F13-18.
- Maraia, R.J. and R.V. Intine. 2001. Recognition of nascent RNA by the human La antigen: conserved and divergent features of structure and function. *Mol Cell Biol* 21: 367-379.
- March-Diaz, R., M. Garcia-Dominguez, F.J. Florencio, and J.C. Reyes. 2007. SEF, a new protein required for flowering repression in *Arabidopsis*, interacts with PIE1 and ARP6. *Plant Physiol* 143: 893-901.
- Mendes Soares, L.M. and J. Valcarcel. 2006. The expanding transcriptome: the genome as the 'Book of Sand'. *EMBO J* 25: 923-931.
- Nikovics, K., T. Blein, A. Peaucelle, T. Ishida, H. Morin, M. Aida, and P. Laufs. 2006. The balance between the MIR164A and CUC2 genes controls leaf margin serration in *Arabidopsis*. *Plant Cell* 18: 2929-2945.
- Numata, K., A. Kanai, R. Saito, S. Kondo, J. Adachi, L.G. Wilming, D.A. Hume, Y. Hayashizaki, and M. Tomita. 2003. Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res* 13: 1301-1306.
- Okazaki, Y. M. Furuno T. Kasukawa J. Adachi H. Bono S. Kondo I. Nikaido N. Osato R. Saito H. Suzuki et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420: 563-573.
- Osato, N., H. Yamada, K. Satoh, H. Ooka, M. Yamamoto, K. Suzuki, J. Kawai, P. Carninci, Y. Ohtomo, K. Murakami et al. 2003. Antisense transcripts with rice full-length cDNAs. *Genome Biol* 5: R5.
- Ota, T. Y. Suzuki T. Nishikawa T. Otsuki T. Sugiyama R. Irie A. Wakamatsu K. Hayashi H. Sato K. Nagai et al. 2004. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* 36: 40-45.
- Prasanth, K.V. and D.L. Spector. 2007. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes Dev* 21: 11-42.
- Rajagopalan, R., H. Vaucheret, J. Trejo, and D.P. Bartel. 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev* 20: 3407-3425.

- Reinhart, B.J., E.G. Weinstein, M.W. Rhoades, B. Bartel, and D.P. Bartel. 2002. MicroRNAs in plants. *Genes Dev* 16: 1616-1626.
- Riano-Pachon, D.M., I. Dreyer, and B. Mueller-Roeber. 2005. Orphan transcripts in *Arabidopsis thaliana*: identification of several hundred previously unrecognized genes. *Plant J* 43: 205-212.
- Rinn, J.L., G. Euskirchen, P. Bertone, R. Martone, N.M. Luscombe, S. Hartman, P.M. Harrison, F.K. Nelson, P. Miller, M. Gerstein et al. 2003. The transcriptional activity of human Chromosome 22. *Genes Dev* 17: 529-540.
- Sasidharan, R. and M. Gerstein. 2008. Genomics: protein fossils live on as RNA. *Nature* 453: 729-731.
- Schmid, M., T.S. Davison, S.R. Henz, U.J. Pape, M. Demar, M. Vingron, B. Scholkopf, D. Weigel, and J.U. Lohmann. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37: 501-506.
- Stolc, V., Z. Gauhar, C. Mason, G. Halasz, M.F. van Batenburg, S.A. Rifkin, S. Hua, T. Herreman, W. Tongprasit, P.E. Barbano et al. 2004. A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* 306: 655-660.
- Storz, G., J.A. Opdyke, and A. Zhang. 2004. Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol* 7: 140-144.
- Sunkar, R., V. Chinnusamy, J. Zhu, and J.K. Zhu. 2007. Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. *Trends Plant Sci* 12: 301-309.
- Sunkar, R., T. Girke, P.K. Jain, and J.K. Zhu. 2005. Cloning and characterization of microRNAs from rice. *Plant Cell* 17: 1397-1411.
- Swarup, K., E. Benkova, R. Swarup, I. Casimiro, B. Peret, Y. Yang, G. Parry, E. Nielsen, I. De Smet, S. Vanneste et al. 2008. The auxin influx carrier LAX3 promotes lateral root emergence. *Nat Cell Biol* 10: 946-954.
- Tenson, T., A. DeBlasio, and A. Mankin. 1996. A functional peptide encoded in the *Escherichia coli* 23S rRNA. *Proc Natl Acad Sci U S A* 93: 5641-5646.
- Touchon, M., A. Arneodo, Y. d'Aubenton-Carafa, and C. Thermes. 2004. Transcription-coupled and splicing-coupled strand asymmetries in eukaryotic genomes. *Nucleic Acids Res* 32: 4969-4978.
- Vaucheret, H. 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev* 20: 759-771.
- Vaucheret, H., A.C. Mallory, and D.P. Bartel. 2006. AGO1 homeostasis entails coexpression of MIR168 and AGO1 and preferential stabilization of miR168 by AGO1. *Mol Cell* 22: 129-136.
- Wang, H., N.H. Chua, and X.J. Wang. 2006. Prediction of trans-antisense transcripts in *Arabidopsis thaliana*. *Genome Biol* 7: R92.
- Wang, J.W., L.J. Wang, Y.B. Mao, W.J. Cai, H.W. Xue, and X.Y. Chen. 2005a. Control of root cap formation by MicroRNA-targeted auxin response factors in *Arabidopsis*. *Plant Cell* 17: 2204-2216.
- Wang, X., S. Arai, X. Song, D. Reichart, K. Du, G. Pascual, P. Tempst, M.G. Rosenfeld, C.K. Glass, and R. Kurokawa. 2008. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454: 126-130.
- Wang, X.J., T. Gaasterland, and N.H. Chua. 2005b. Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol* 6: R30.

- Washietl, S., I.L. Hofacker, M. Lukasser, A. Huttenhofer, and P.F. Stadler. 2005. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol* 23: 1383-1390.
- Wernersson, R. and H.B. Nielsen. 2005. OligoWiz 2.0--integrating sequence feature annotation into the design of microarray probes. *Nucleic Acids Res* 33: W611-615.
- Wolin, S.L. and T. Cedervall. 2002. The La protein. *Annu Rev Biochem* 71: 375-403.
- Xiong, L., R.G. Wang, G. Mao, and J.M. Koczan. 2006. Identification of drought tolerance determinants by genetic analysis of root response to drought stress and abscisic Acid. *Plant Physiol* 142: 1065-1074.
- Yamada, K., J. Lim, J.M. Dale, H. Chen, P. Shinn, C.J. Palm, A.M. Southwick, H.C. Wu, C. Kim, M. Nguyen et al. 2003. Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science* 302: 842-846.
- Yamashita, A., Y. Watanabe, N. Nukina, and M. Yamamoto. 1998. RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast. *Cell* 95: 115-123.

Figure 1

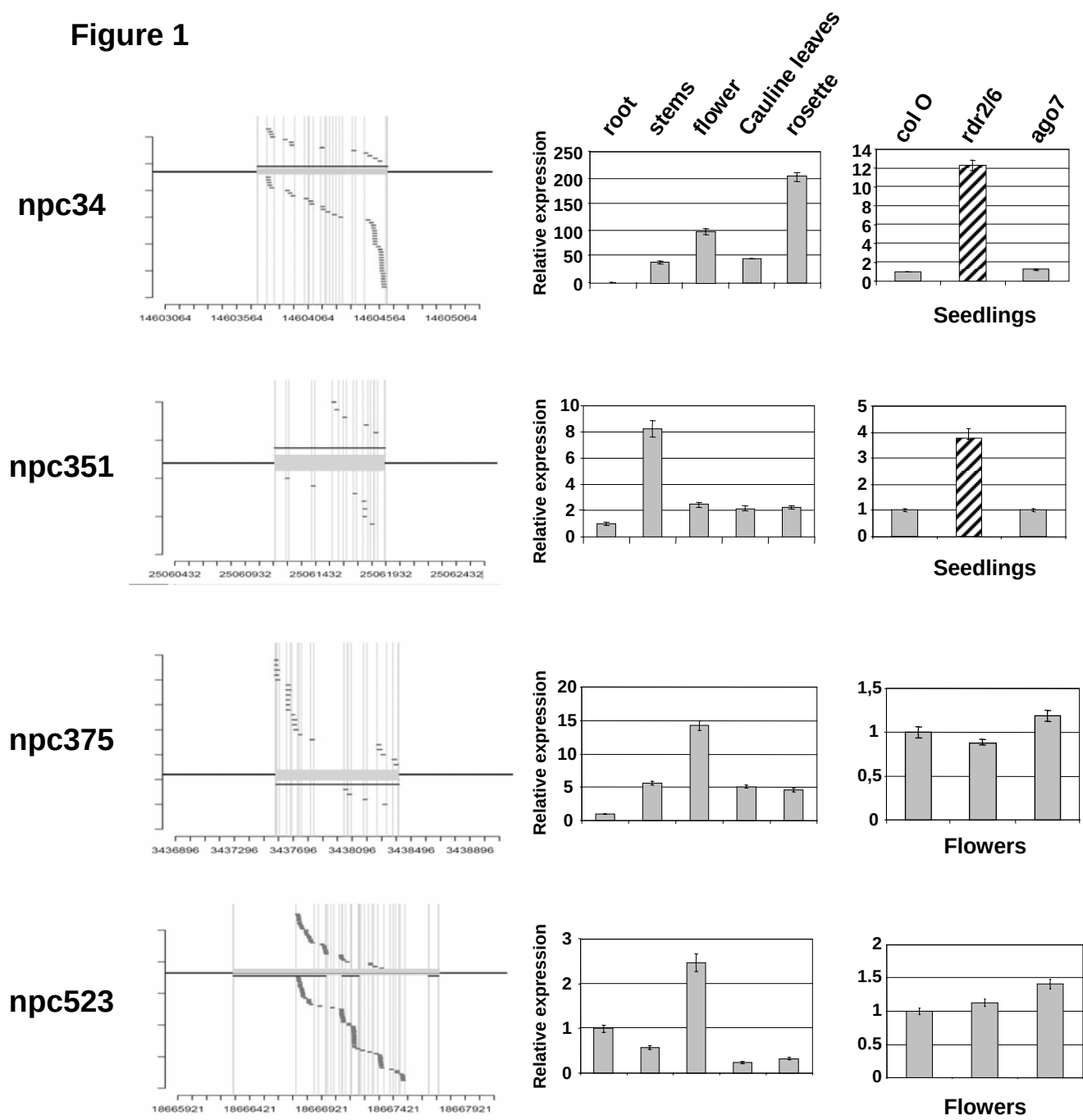


Figure 2

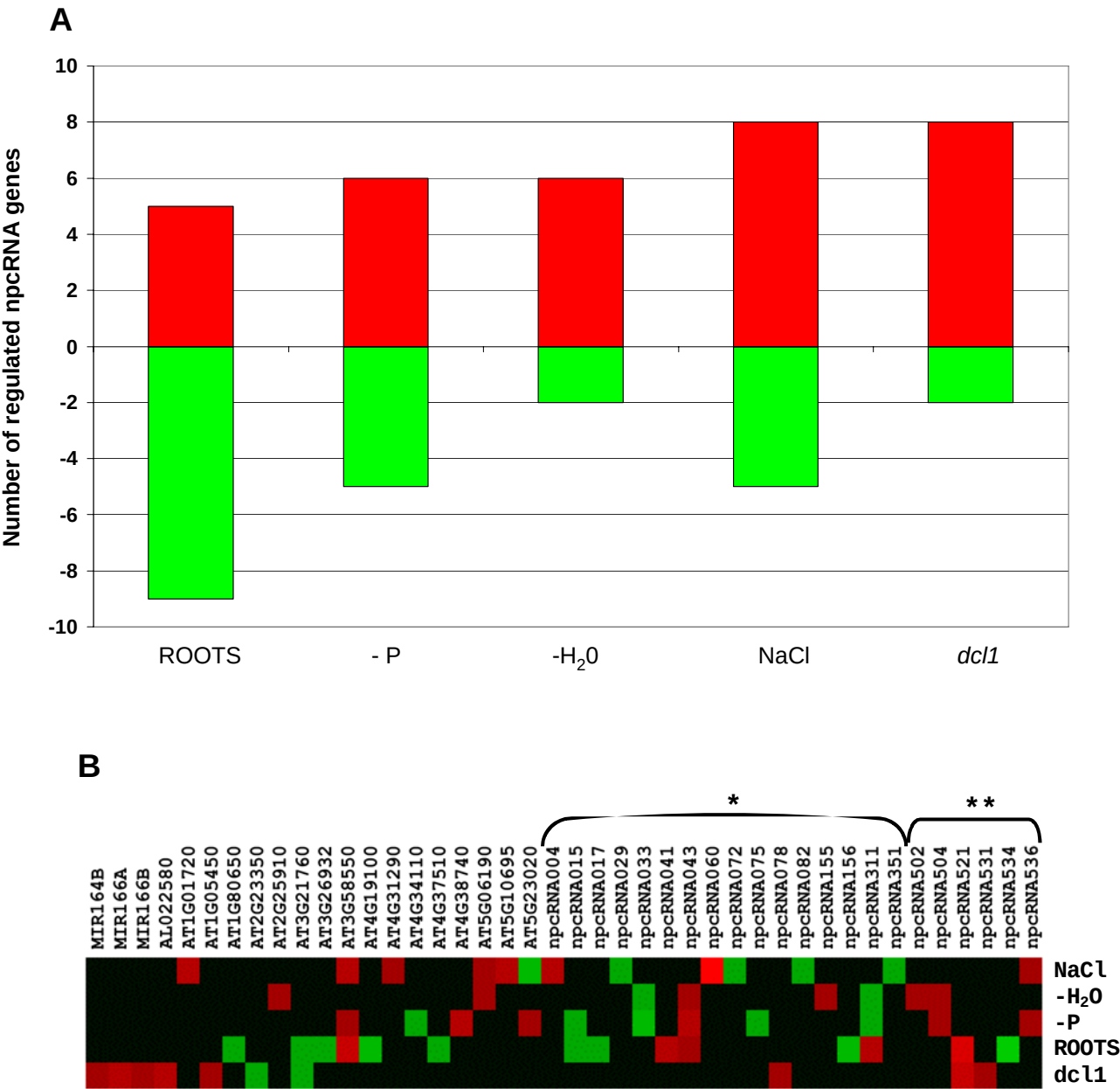
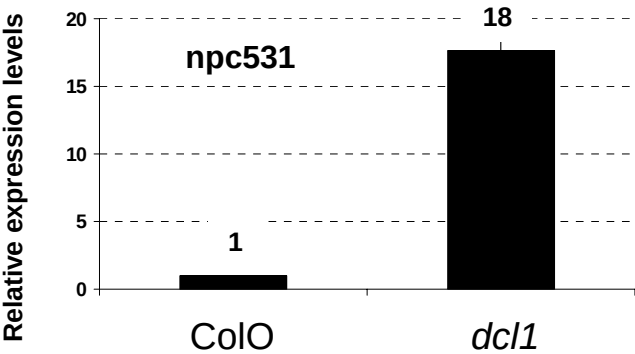


Figure 4

A



Target npc531 (1-900)
Fold.energy: -26.7 Kcal

26 47
GUUGAUCUUCUUUCAGUUUAU
|| |||||
UCCUCGAGGGAAGUCAGGUU
3' 5'
miR319a

B

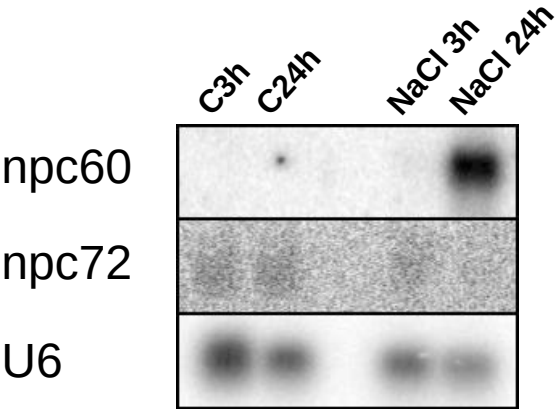
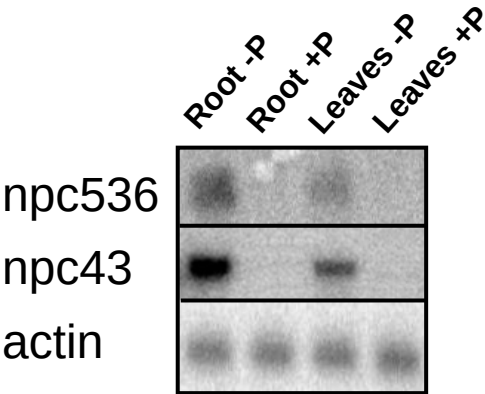


Figure 5

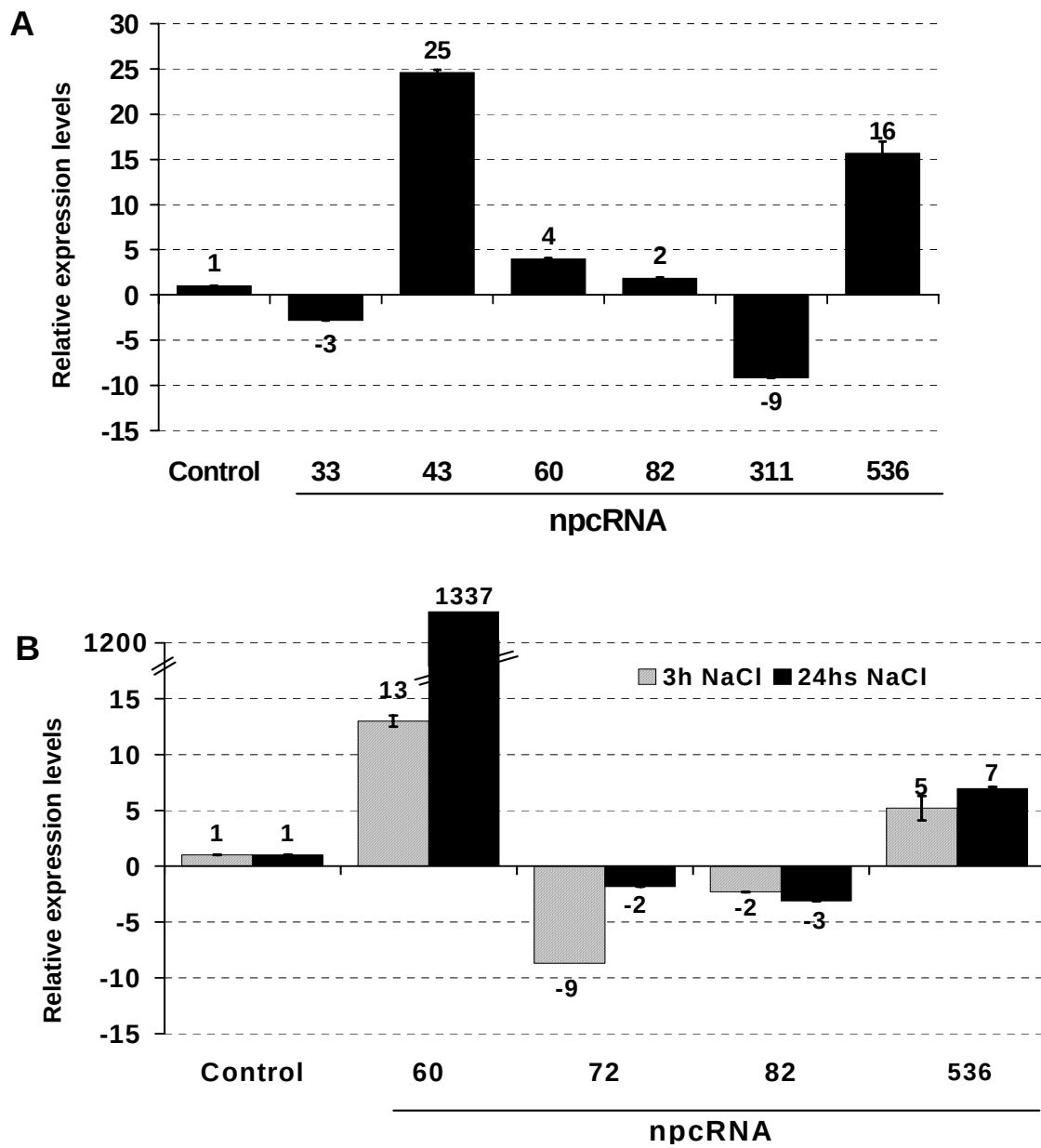


Figure 6

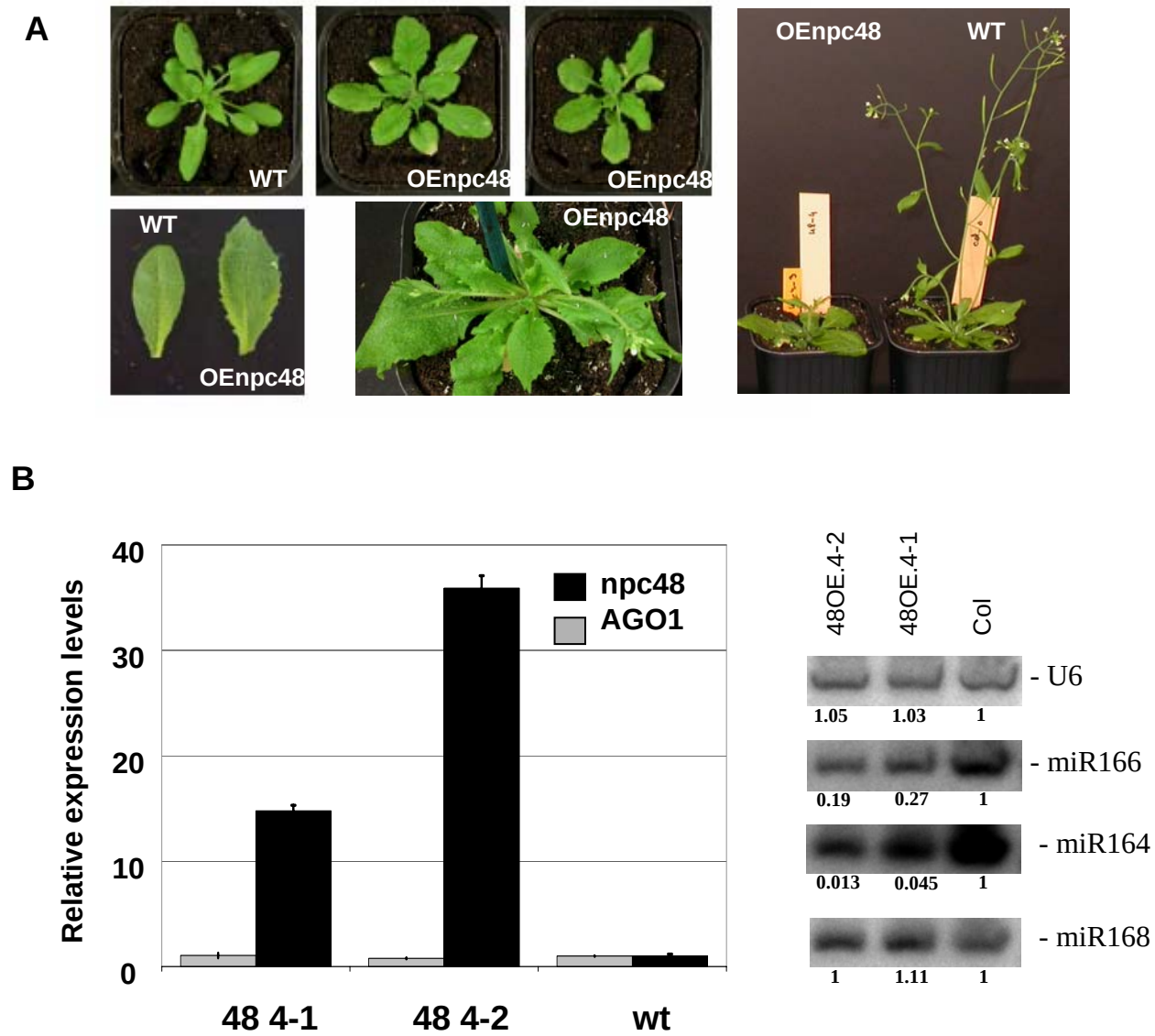


Figure 7

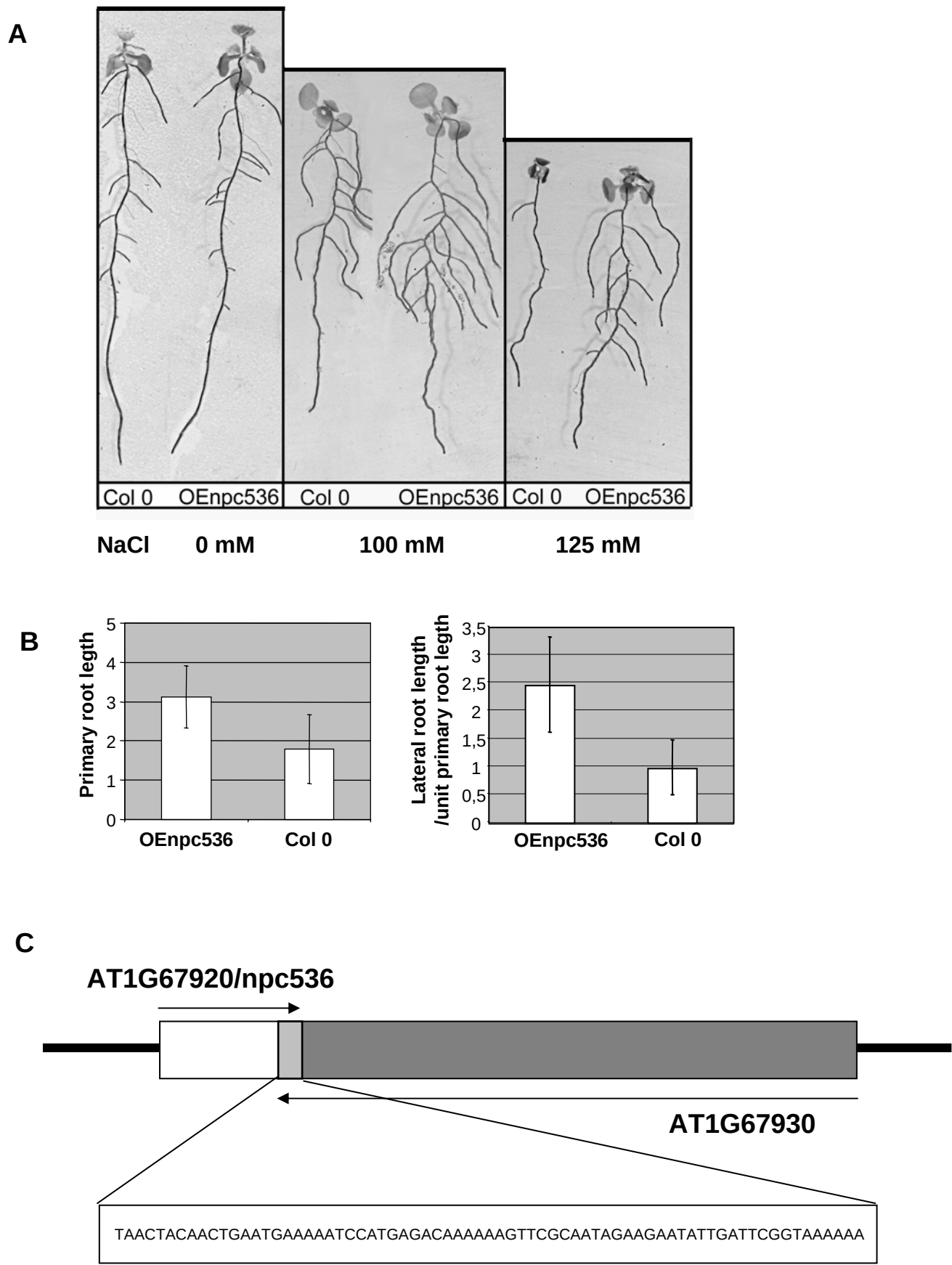


Table 1
The Arabidopsis npcRNA collection and their matching small RNAs

npcRNA	sense gene	antisense gene	gene begin	gene length	chr	strand	exon number	RNA structure	sense sRNA	antisense sRNA
2	AT1G03106		748102	933	1	c	1	5.7		
4	AT1G06135		1862196	546	1	w	1			
14			17298005	735	1	w	1			1
15			18141841	833	1	w	1			
17	AT1G52855		19686614	553	1	c	1			
21			26179316	3700	1	c	3			
26			2330796	698	2	w	1		1	
29			6651362	1253	2	c	1			
30			8129244	912	2	w	1			
33	AT2G22122		9410344	545	2	w	1			
34	AT2G34655		14603710	910	2	w	1		14	42
40	AT3G15395		5200646	876	3	w	2			
41	AT3G17185 (<i>ZAS3</i>)		5861497	947	3	w	1	6.4	120	116
43		npcRNA 504	6956996	692	3	w	1		1	1
48			11511953	983	3	w	1	5.6		
51	AT3G47965		17718903	992	3	c	2			
52	AT3G48115		17782155	1757	3	c	2			
58	AT4G04315	AT3G66658	2095257	1260	4	w	2			
60	AT4G12735		7493713	519	4	c	1			
62			9244083	589	4	w	1		1	
72			16617625	887	4	w	1			
75		AT4G38930	18151750	804	4	c	1	5.3	3	
78	AT5G08185 (<i>MIR162a</i>)		2634105	1333	5	c	4	5.2	14	1
79	AT5G10745		3396476	1589	5	w	3		1	
82	AT5G24105		8151927	944	5	w	2	9		
83			15908505	706	5	c	1	16.7	22	1
86	AT5G56795		22989135	603	5	c	2			
111			17255407	936	1	c	1		1	1
113			20126020	520	1	c	1		1	
131	AT2G31585		13453079	728	2	w	1			
149	AT3G24927		9103094	903	3	w	1			
150			9666799	788	3	w	1	5.6		
155			22001256	509	3	w	1		1	
156	AT4G04692		2375600	763	4	c	1			
157	AT4G08035		4866551	1790	4	w	1			
311			3544985	488	1	c	1	6.4		
325			11453507	415	1	w	2			
326	AT1G34418		12582100	600	1	w	2			
351			25061142	782	1	w	1	7.2	5	7
370		AT2G07783	3299858	522	2	c	1			1
375			3437581	832	2	c	1	5.1	4	22
415	AT3G09162		2807672	602	3	c	1		4	1
431			11155388	340	3	w	1		2	
502	AT1G16515		5647542	265	1	c	1			
504		npcRNA 43	6957092	591	3	c	1		1	1
507	AT3G21755	AT3G21760	7666815	1758	3	c	2			2
508			17460953	591	3	w	1		1	1
509	AT4G31985		15469901	665	4	w	2			
510	AT1G29357	AT1G29355	10274088	1617	1	c	2			
511	AT1G11592		3895179	1756	1	c	5	5.3	1	1
512			28940575	1500	1	c	5		7	5
513			11110207	1420	4	c	1		6	
514	AT3G27865 (snoRNA)		10329663	1360	3	w	1		8	
515		AT5G21020	7139131	542	5	c	1			
516		AT1G76500	28710088	1356	1	c	1			
518		AT1G58235	21583424	791	1	w	2			
519		AT4G39675	18413768	449	4	c	1			
520			16888083	1195	1	c	1	5	30	18
521	AT5G46845 (<i>MIR160c</i>)		19025187	1362	5	c	1	7.3	21	
522			8459600	705	4	w	1		11	
523	AT2G45245		18666309	1427	2	c	3		66	36
524			9832273	722	3	w	1	9.4	3	1
525	AT2G13665	AT2G13660	5699230	869	2	w	1			
526		AT1G12060	4076214	704	1	c	1		1	
527			11472056	740	1	w	1			
528		AT2G11240	4490324	1080	2	w	1			
529	AT1G67365	AT1G67370	25242073	826	1	w	1			
530	AT1G52347		19500175	1198	1	w	2			
531			17480129	900	1	w	1			
532	AT2G03972		1255245	1198	2	c	4			
534	AT1G67265		25179221	645	1	c	1			
535	AT3G05152		1443973	1121	3	w	2			
536	AT1G67920	AT1G67930	25477207	597	1	w	1		1	
539	AT3G59765		22087438	1879	3	c	3		1	4
540			20581088	850	1	c	1		1	
541			28298269	753	1	w	1			

Table 2. Regulated genes of the RIBOCHIP

Gene name	ROOTS	-P	-H2O	NaCl	<i>dcl1</i>	Description
164B					3	miRNA 164B
166A					5	miRNA 166A
166B					3	miRNA 166B
AL022580					5	T5K18 mirna168a
AT1G01720				5		no apical meristem (NAM) family protein putative TF
AT1G05450					2	protease inhibitor/seed storage/lipid transfer protein (LTP)
AT1G80650	-2					double-stranded RNA-binding domain (DsRBD)-containing protein
AT2G23350					-3	polyadenylate-binding protein, putative
AT2G25910			2			3-5 exonuclease domain-containing protein
AT3G21760	-3				-2	UDP-glucuronosyl/UDP-glucosyl transferase family protein
AT3G26932	-3					double-stranded RNA-binding domain (DsRBD)-containing protein
AT3G58550	8	3		4		protease inhibitor/seed storage/lipid transfer protein (LTP) family
AT4G19100	-6					expressed protein
AT4G31290				2		ChaC-like family protein
AT4G34110		-2				polyadenylate-binding protein 2 (PABP2)
AT4G37510	-3					ribonuclease III family protein
AT4G38740		4				peptidyl-prolyl cis-trans isomerase / cyclophilin / rotamase
AT5G06190			2	3		expressed protein
AT5G10695				4		Expressed protein
AT5G23020		2		-4		2-isopropylmalate synthase 2 (IMS2), identical to 2-isopropylmalate
npcRNA 004				4		npcRNA (AT1G06135)
npcRNA 015	-3	-2				npcRNA
npcRNA 017	-2					npcRNA AT1G52855
npcRNA 029				-2		npcRNA
npcRNA 033		-4	-2			npcRNA AT2G22122
npcRNA 041	4					TAS3
npcRNA 043	3	6	3			npcRNA
npcRNA 060				97		npcRNA AT4G12735
npcRNA 072				-2		npcRNA
npcRNA 075		-2				npcRNA
npcRNA 078					3	mir162a gene
npcRNA 082				-3		npcRNA AT5G24105
npcRNA 155			2			npcRNA
npcRNA 156	-5					npcRNA AT4G04692
npcRNA 311	4	-2	-2			npcRNA
npcRNA 351				-2		npcRNA
npcRNA 502			3			npcRNA AT1G16515
npcRNA 504		3	2			NpcRNA (AS 43)
npcRNA 521	14				8	mir160 gene
npcRNA 531					3	npcRNA
npcRNA 534	-7					npcRNA
npcRNA 536		3		2		npcRNA (AT1G67920)

Table 3

npRNA	Number of independent lines	Gene annotation	Sequence similarities	Z-score > 5	Tissue specificity	Small RNA	Regulation
34	19	AT2G34655			leaves	YES	
TAS3	14	AT3G17185	2 (+)	6,4	roots	YES	
43	21					YES	P, drought
48	236			5,6		NO	
60	9	AT4G12735				NO	NaCl
72	27					NO	NaCl
82	9	AT5G24105	9 (63)	9	flowers	YES	NaCl
MIR869a	14			16,7	stems	YES	
311	19		Bn (+)	6,4	roots	NO	P, NaCl
375	27		15 (+)	5,1	flowers	YES	
531	51					NO	Dcl1
536	8	AT1G67920	2 (67)			NO	P, NaCl