universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

Characterization of microRNA-mediated cleavage events in *Arabidopsis* embryos

verfasst von / submitted by

Aleksandra Plotnikova

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 794 685 490
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Molecular Biology
Betreut von / Supervisor:	Michael D. Nodine, PhD

Во всем мне хочется дойти
До самой сути.
В работе, в поисках пути,
В сердечной смуте.

До сущности протекших дней,
До их причины,
До оснований, до корней,
До сердцевины...

Б. Пастернак

* * *

In everything I want to grasp
Its very core.
In work, in searching for the path,
In heart's uproar.

To see the essence of my days,
In every minute
To see its cause, its root, its base,
Its sacred meaning…

B. Pasternak
(translated by Natasha Gotskaya)

Acknowledgements

I would like to express my deepest gratitude to my supervisor Michael D. Nodine for the opportunity to work on this fascinating topic. His expertise and enthusiasm have guided me through the project. Furthermore, my sincere appreciation goes to Max Kellner, whose contribution was invaluable for completing this study. I am also very thankful to the rest of the lab, especially Falko, Magda, Michael, Olga, and Stefan, the world's most extraordinary colleagues and dear friends.

Further, I would like to thank my thesis advisory committee Louisa Cochella and Fred Berger, for their excellent advice and thoughtful comments during the running of this project. I wish to thank Tobias Müller and Pawel Pasierbek from BioOptics for sharing their expert knowledge and providing assistance with the microscopy. Many thanks to the Vienna RNA Community for their collaborative spirit and fruitful discussions. I want to pay my regards to Nicola Wiskocil, who helped me in any way she could during my first months in Vienna.

Special thanks to Andreas Finke for his unconditional support, encouragement, and patience. Finally, I wish to thank my parents, who continue to remain my most loyal supporters and who set me on the road to science a long time ago.

Characterization of microRNA-mediated cleavage events in *Arabidopsis* embryos

Contents

1. Abstract

MicroRNAs (miRNAs) are small regulatory RNAs that repress genes in plants and animals. Previously we found that miRNA-deficient *Arabidopsis* embryos exhibit widespread cellular differentiation and developmental timing defects (Nodine and Bartel, 2010). However, little was known about the contributions of individual miRNAs to pattern formation. To systematically characterize miRNA functions during embryogenesis, we profiled small RNAs throughout embryogenesis and identified dozens of miRNA families that are abundant during the morphogenesis phase of embryo development. The functions of miRNAs are defined by the genes they regulate. In plants, miRNAs have nearly perfect complementarity to their targets and typically guide their endonucleolytic cleavage. We have developed a high-throughput method, nanoPARE (parallel analysis of RNA ends), to profile miRNA cleavage products genome-wide from sub-nanogram amounts of total RNA. We applied nanoPARE to early embryos and identified dozens of miRNA:target interactions operating during embryogenesis. Our nanoPARE data, together with transcriptome analysis of wild-type and miRNA-deficient embryos, enabled the identification of several miRNA families that cleave and repress distinct sets of transcription factors. Because miRNA-mediated regulation of transcription factors may be especially crucial for embryonic pattern formation, we selected these miRNAs for further characterization. We used fluorescent protein-based methods to determine the activity patterns of these miRNAs at cellular resolution during embryogenesis. In addition, we generated transgenic plants expressing miRNA-resistant versions of these target transcription factors and examined the morphology of embryos to determine whether cell-specific repression of transcription factors is required for proper embryo morphogenesis. Altogether, our results demonstrate that miRNAs define the spatiotemporal localization of transcription factors, which is determinative for establishing the body plan at the beginning of plant life.

2. Zusammenfassung

MicroRNAs (miRNAs) sind eine Klasse kleiner regulatorischer RNAs, die Gene, sowohl in Pflanzen als auch in Tieren, unterdrücken. Bisher fanden wir heraus, dass miRNA-defiziente *Arabidopsis*-Embryonen eine Vielzahl an Zelldifferenierungs- und Entwicklungszeitdefekte aufweisen (Nodine and Bartel, 2010). Über die Beiträge spezifischer miRNAs zur Musterbildung ist jedoch wenig bekannt. Um die miRNA-Funktionen während der Embryogenese systematisch zu charakterisieren, haben wir kleine RNAs während der Embryogenese profiliert und dutzende von miRNA-Familien identifiziert, die während der Morphogenese-Phase der Embryonalentwicklung häufig vorkommen. Die Funktionen von miRNAs werden durch die Gene definiert, die sie regulieren. In Pflanzen haben miRNAs eine nahezu perfekte Komplementarität mit ihren Zielsequenzen und steuern typischerweise deren endonukleolytische Spaltung. In dieser Arbeit haben wir nanoPARE (parallele Analyse von RNA-Enden) entwickelt, eine Hochdurchsatzmethode, um miRNA-Spaltprodukte genomweit aus Sub-Nanogramm-Mengen der Gesamt-RNA zu profilieren. Wir haben nanoPARE auf RNA früher Embryonen angewendet und Dutzende von miRNA:Zielinteraktionen identifiziert, die während der Embryogenese ablaufen. Zusammen mit der Transkriptomanalyse von Wildtyp- und miRNA-defizienten Embryonen ermöglichten unsere nanoPARE-Daten die Identifizierung mehrerer miRNA-Familien, die unterschiedliche Familien von Transkriptionsfaktoren spalten und unterdrücken. Da die miRNA-vermittelte Regulation von Transkriptionsfaktoren für die Bildung embryonaler Muster besonders wichtig sein kann, haben wir diese miRNAs zur weiteren Charakterisierung ausgewählt. Wir verwendeten fluoreszierende proteinbasierte Methoden, um die Aktivitätsmuster dieser miRNAs bei zellulärer Auflösung während der Embryogenese zu bestimmen. Zusätzlich erzeugten wir transgene Pflanzen, die miRNA-resistente Versionen dieser Zieltranskriptionsfaktoren exprimierten, und untersuchten die Morphologie von Embryonen, um festzustellen, ob eine zellspezifische Repression von Transkriptionsfaktoren für eine ordnungsgemäße Embryonenmorphogenese erforderlich ist. Insgesamt zeigen unsere Ergebnisse, dass miRNAs die räumlich-zeitliche Lokalisierung von Transkriptionsfaktoren definieren, die für die Erstellung des Körperplans zu Beginn des Pflanzenlebens erforderlich ist.

3. Synopsis

This work consists of two interconnected studies and provides insights into the spatio-temporal functions of individual microRNAs (miRNAs) and transcription factors during embryo morphogenesis in *Arabidopsis thaliana*. The first study summarized in Chapter 5, "NanoPARE: parallel analysis of RNA 5′ ends from low-input RNA," describes a high-throughput method to profile miRNA-mediated mRNA cleavage events at a base-pair resolution. Transcriptome analysis, combined with the associated nanoPARE data from the same embryonic cDNA samples, enabled identifying several high-confidence miRNA:target interactions throughout *Arabidopsis* development. Some of these miRNA:target interactions were further investigated in Chapter 6, "MicroRNA Dynamics and Functions during *Arabidopsis* Embryogenesis." The second study focuses on the impact of specific miRNAs on the gene expression programs that guide developmental patterning and timing. Additional evidence obtained using genetic, microscopic, and histological approaches further validated these interactions and provided information about the dynamics, localization, and activity of miRNAs and their targets.

4. Introduction

4.1 Embryo development in *Arabidopsis thaliana*

In Angiosperms, the double fertilization of the female gametophyte leads to the formation of the embryo and the endosperm, an auxiliary tissue. Embryonic development is, therefore, a period of the plant life cycle from the emergence of the unicellular zygote to the formation of mature, nutrient-filled seed (Christensen *et al.* 1997; Faure *et al.* 2002; Owen and Makaroff 1995). Plant embryogenesis consists of morphogenesis and maturation phase (Braybrook and Harada 2008; Goldberg *et al.* 1994). The morphogenesis phase of embryo development is defined by the establishing of the basic body plan: radial and axial (root-shoot) patterning occur, and cell and tissue types are specified *de novo* (Braybrook and Harada 2008; Seefried *et al.* 2014). During the maturation phase, the embryo accumulates storage macromolecules, which serve as a reserve of nutrients for the dormant and germinating seedling (Baud *et al.* 2009; Pelletier *et al.* 2017; Santos-Mendoza *et al.* 2008). Although developing embryos' morphology and anatomy can differ substantially within the plant kingdom, there are highly robust species-specific principles of embryonic pattern formation. The basic body plan is built stereotypically, irrespective of different environmental conditions (Yoshida *et al.* 2014; Scheres *et al.* 1994; Abe *et al.* 2003).

In most flowering plants, including *Arabidopsis*, the zygote elongates and divides to form two asymmetric daughter cells (Mansfield and Briarty 1991; Zhang and Laux 2011). A small cytoplasmic-dense apical cell gives rise to the embryo proper the post-embryonic body, after seed germination. A large vacuolated basal cell forms an extraembryonic suspensor (Ueda and Laux 2012). The suspensor is a highly specialized and terminally differentiated tissue, which is crucial for embryo development. It connects the embryo with the surrounding tissues

(endosperm, seed coat) and participates in transporting growth factors, hormones, and nutrients to the embryo proper (Kawashima and Goldberg 2010; Yeung and Meinke 1993). The asymmetric nature of the first division of the zygote, which occurs 4-8 hours after fertilization (HAF), creates the apical-basal axis in the early embryo (van Dop *et al.* 2015, Mansfield and Briarty 1991; Kao and Nodine, 2019; Zhao *et al.* 2019). Later, 18 HAF, the apical cell divides vertically, producing two daughter cells half the parent cell size. The basal cell elongates and divides horizontally, keeping the same plane of divisions until the late globular stage. By 24 HAF, a change of division plane occurs: a 2-cell proembryo divides horizontally to form a 4-cell stage (quadrant) proembryo with upper and lower tiers to establish the apicobasal axis of the embryo proper. The quadrant then divides vertically, producing an 8-cell stage (octant) embryo proper (Fig 1). The transition from 8 to 16-cells (dermatogen) occurs at 36 HAF and requires tangential divisions of the outer embryonic cells towards the periphery. The radial axis is established at this stage, and the outermost protodermal cells (the precursors of the epidermal tissue) are specified (Seefried *et al.* 2014; van Dop *et al.* 2015; Mansfield and Briarty 1991; Yoshida *et al.* 2014; Scheres *et al.* 1994; Takada and Jürgens 2007; Nodine *et al.* 2007). By the globular stage, 48-60 HAF, the precursors of most plant tissues are formed: the inner cells of the embryo proper form the ground tissue and vasculature precursors (Palovaara *et al.* 2015; Radoeva and Weijers 2014; Yoshida *et al.* 2014). Suspensor is fully developed by the early globular stage; it consists of 7-9 cells connected by plasmodesmata. During the globular stage, the suspensor initiates programmed cell death and thus does not contribute to the next generation of the plant (Kawashima and Goldberg 2010; Yeung and Meinke 1993). When the embryo progresses from the early- to late-globular stage, the uppermost cell of suspensor (hypophysis) divides asymmetrically to produce basal and lens-shaped cells. The hypophysis will develop into columella initials, root cap initials, ground tissue initials and vascular initials of the root apical meristem (RAM) and the lens-shaped cell will give rise to quiescent center (QC), (Schlereth *et al.* 2010; Sabatini *et al.* 2003). By the end of the late globular stage, the frequency of protodermal divisions increases in the area of cotyledon primordia, and the embryo briefly assumes a triangular shape (transition stage). This stage is rather short-lived because cotyledon initials divide rapidly, and the embryo quickly progresses to the early heart stage (by approximately 80 HAF) (Mansfield and Briarty 1991; Jürgens and Mayer 1994). During the heart stage (66-84HAF), embryo cells continue to expand and differentiate. Due to the specification of the shoot apical meristem (SAM) and further elongation of the cotyledon primordia, the proembryo becomes heart-shaped. The endosperm starts to cellularize, and chlorophyll accumulation begins in the protoderm (Mansfield and Briarty 1991; Willmann *et al.* 2011).

Because photosynthesis seems to be required for the accumulation of macromolecules characteristic for seed maturation such as oleosins, embryo greening at the heart stage can be considered the onset of the maturation phase (Goffman *et al.* 2005; Liu *et al.* 2017; O'Neill *et*

al. 2019). In the following stages (torpedo, bent cotyledon), the cotyledon initials continue to elongate and fold to fill the space inside the seed. Storage oils and proteins (albumins, cruciferins, arabins) accumulate at the mature green stage which is followed by embryo dessication and quiescence (Baud *et al.* 2008; Goldberg *et al.* 1994; Pelletier *et al.* 2017; Santos-Mendoza *et al.* 2008). At the late stage of the maturation, the dehydrated seed enters a dormant state, which will last until germination (Leprince *et al.* 2016; Raz *et al.* 2001).

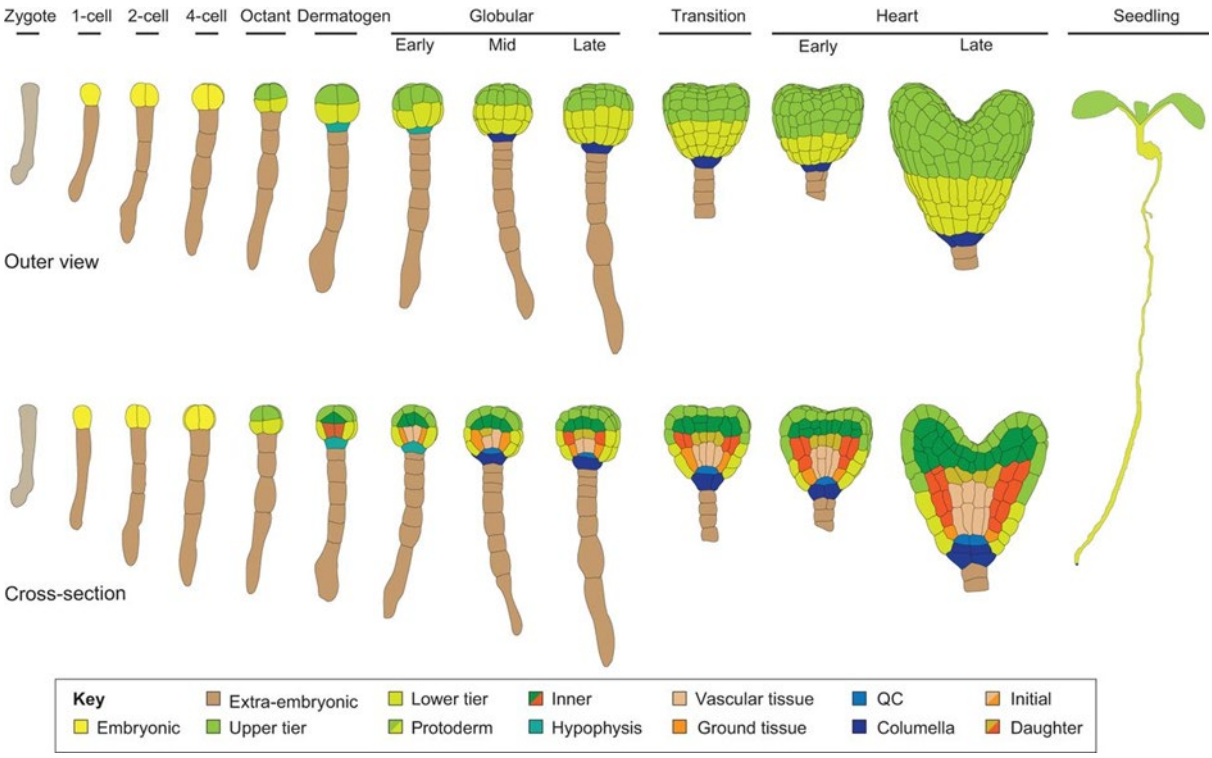


Figure 1. Embryo morphogenesis in *Arabidopsis*. Schematic overview of early *Arabidopsis* embryo progressing from the zygote to the heart stage embryo (the end of morphogenesis phase) through a series of stereotypical cell divisions. Surface view (upper row) and cross-sections (bottom row) of a developing. The colors represent cell lineages and tissues precursors, based on marker gene expression analysis, as indicated in the legend. The figure is modified from ten Hove *et al.* 2015 and Yoshida *et al.* 2014.

Despite their morphological simplicity (Figure 1), *Arabidopsis* embryos provide a blueprint of the basic plant body organization. The invariant timing and highly stereotypical nature of cell divisions make *Arabidopsis* embryo an ideal model for molecular characterization of pattern formation, cell differentiation, and tissue type specification (Mansfield and Briarty 1991; Jürgens and Mayer 1994; Laux *et al*, 2004). Embryo morphogenesis is under tight spatiotemporal regulation, and most genes previously identified as master regulators of embryonic patterning encode components of signaling cascades such as hormones and transcription factors (reviewed in ten Hove *et al.* 2015). A number of transcriptional regulators that orchestrate plant development are targeted by microRNAs (Jones-Rhoades *et al.* 2006; Nodine and Bartel 2010). Although substantial progress had been made elucidating the impact of miRNA:transcription factor interactions on post-embryonic development, our understanding of regulatory networks operating during plant embryogenesis remained limited.

4.2 MicroRNAs in *Arabidopsis thaliana*

MicroRNAs (miRNAs) are 20-22 nt long endogenous regulatory RNAs, which play integral roles in many vital processes in animals and plants by regulating the expression of genes, mostly post-transcriptionally. In plants, miRNAs influence development, growth, response to environmental changes, biotic and abiotic stresses (reviewed in Bartel 2009; Ambros 2011; Chen 2009; Sunkar *et al.* 2012). MiRNAs form effector complexes with Argonaute proteins (AGOs) and bind to complementary sites in target mRNAs, driving degradation or translational repression (Kawamata and Tomari 2010; Tang *et al.* 2003; Baumberger and Baulcombe 2005; Qi *et al.* 2003; Addo-Quaye *et al.* 2008; German *et al.* 2008).

Most plants' genomes typically encode a few hundred miRNA genes (MIR) mostly located in intergenic regions in the genome (Nozawa *et al.* 2012; Budak and Akpinar 2015). These genes are transcribed in the nucleus by DNA-dependent RNA Polymerase II (Pol II), which is recruited to MIR promoters by a multiprotein transcriptional activator complex Mediator (Fig 2) (Kim *et al.* 2011; Mathur *et al.* 2011; Buendía-Monreal and Gillmor 2016). Mediator is an evolutionarily conserved multi-subunit assembly that serves as a central interface, where signals from DNA-binding transcription factors (TFs) are communicated directly to Pol II machinery (Allen and Taatjes 2015). Hence, Mediator enables TF-dependent regulation of gene expression according to developmental or environmental context.

An average family of MIR genes consists of 4-5 members, which are under tight transcriptional regulation (Li and Mao 2007). Differential expression of individual MIR genes leads to spatiotemporal accumulation of particular miRNAs in response to developmental cues and changing conditions (Rogers and Chen 2013).

Pol II generates the primary transcripts (pri-miRNA), which are co-transcriptionally 5'-capped (7-methylguanosine), 3'-polyadenylated and often spliced (Hajheidari *et al.* 2012; Xie *et al.* 2005; Rogers and Chen 2013; Millar and Waterhouse 2005). The primary transcripts are, in turn, cleaved by the DICER-LIKE1 (DCL1) RNAse III-type nuclease to release the stem-loop-structured precursors (pre-miRNA), which are further processed into the mature double-stranded miRNA (Fig 2).

In animals, the processing of pri-miRNA into pre-miRNA and the following maturation of the miRNA from the pre-miRNA is mediated by the nuclear enzyme Drosha and cytoplasmic enzyme Dicer, respectively (Lee *et al.* 2003; Hutvagner *et al.* 2001; Kim 2005; Ketting *et al.* 2001). In plants, both steps are carried out by DCL1 (Reinhart *et al.* 2002; Kurihara and Watanabe 2004; Park *et al.* 2002). There are four members of the DCL protein family in *Arabidopsis*: DCL1 generates ~21-nt miRNAs, while other DCLs produce functionally distinct small RNAs of defined sizes (DCL4 - 21nt, DCL2 - 22nt and 24nt for DCL3) (Xie *et al.* 2004; Akbergenov *et al.* 2006).

DCL1 is a crucial factor of plant miRNA biogenesis *in vivo* (Park *et al.* 2002; Reinhart *et al.* 2002). miRNA-deficient *dcl1*-null embryos have widespread morphological defects and are

arrested early in development (Nodine and Bartel 2010; Schwarz *et al.* 1994). It was also shown that multiple miRNA targets are massively upregulated and ectopically expressed in *dcl1* embryos (Nodine and Bartel. 2010).

The processing of plant miRNAs is localized to subnuclear dicing bodies (D-bodies) and requires several RNA-binding proteins, including HYPONASTIC LEAVES1 (HYL1)/DOUBLE-STRANDED RNA BINDING (DRB1), SERRATE (SE), DAWDLE (DDL), TOUGH (TGH), which jointly enhance the activity and accuracy of DCL1 (Manavella *et al.* 2012; Fang and Spector 2007; Fujioka *et al.* 2007; Song *et al.* 2007). The DCL1-generated miRNA/miRNA* duplex consists of a "guide" (miRNA) strand, which is loaded into RNA-induced silencing complex (RISC), and a "passenger" (miRNA*) strand that undergoes degradation (Eamens *et al.* 2009). Mature miRNAs have 2-nt overhangs at 3' ends as well as 5'-phosphate and 3'-hydroxyl groups on both strands (Kurihara and Watanabe 2004; Reinhart *et al.* 2002). Most miRNAs in *Arabidopsis* are also methylated on the 2'-OH group of the 3'-terminal ribose of each strand (Park *et al.* 2002). HEN1 methyltransferase, which deposits the methyl groups, is crucial for the biogenesis and accumulation of mature miRNAs (Park *et al.* 2002; Yu *et al.* 2005). Unmethylated miRNAs in *hen1* mutants are uridylated and truncated at their 3'-ends, suggesting that 2'-O-methylation protects miRNAs from 3'-5' exonucleolytic cleavage and enhances its stability (Li *et al.* 2005; Zhao *et al.* 2012).In *Arabidopsis*, 3'-uridylation of miRNAs is carried out by HEN1 SUPRESSOR1 (HESO1), which adds non-templated Us to RNA oligonucleotides. HEN1-mediated 2'-O methylation inhibits this nucleotidyltransferase activity (Ren *et al.* 2012; Zhao *et al.* 2012).

Following methylation, the guide miRNA strand is incorporated into a ribonucleoprotein RNA-INDUCED SILENCING COMPLEX (RISC), while the passenger strand is removed (Figure 2). The strand selection and the RISC-loading process depend on miRNA strands' thermodynamic properties and is facilitated by HYL1 (Eamens *et al.* 2009). Argonaute proteins are the critical effector components of the RISC. *Arabidopsis* AGO family consists of ten paralogs with a similar domain architecture: the RNA unwinding N-terminal domain, RNA-binding PAZ and MID domains, and endonucleolytic PIWI domain (Hutvagner and Simard 2008; Zhang *et al.* 2014; Wang *et al.* 2009; Derrien *et al.* 2018;). The RNAseH-like PIWI domain is a slicer responsible for cleaving RNA targets. Unlike in animals, miRNAs are highly complementary to their targets in plants and mainly direct endonucleolytic slicing of the mRNA and translational repression (Baumberger and Baulcombe 2005; Brodersen *et al.* 2008; Li *et al.* 2013; Axtell and Meyers 2018).

Arabidopsis Argonauts bind small RNAs with characteristic 5'-terminal nucleotide biases: AGO1 and AGO10 prefer 5'-uracil, AGO5 to 5'-cytosines, and AGO2, AGO4, AGO6, AGO7, and AGO9 prefer 5'-adenines. Most plant miRNAs have 5'-uracil and are loaded into AGO1 (Mi *et al.* 2008; Takeda *et al.* 2008; Maunouny and Vaucheret 2010; Thieme *et al.* 2012).

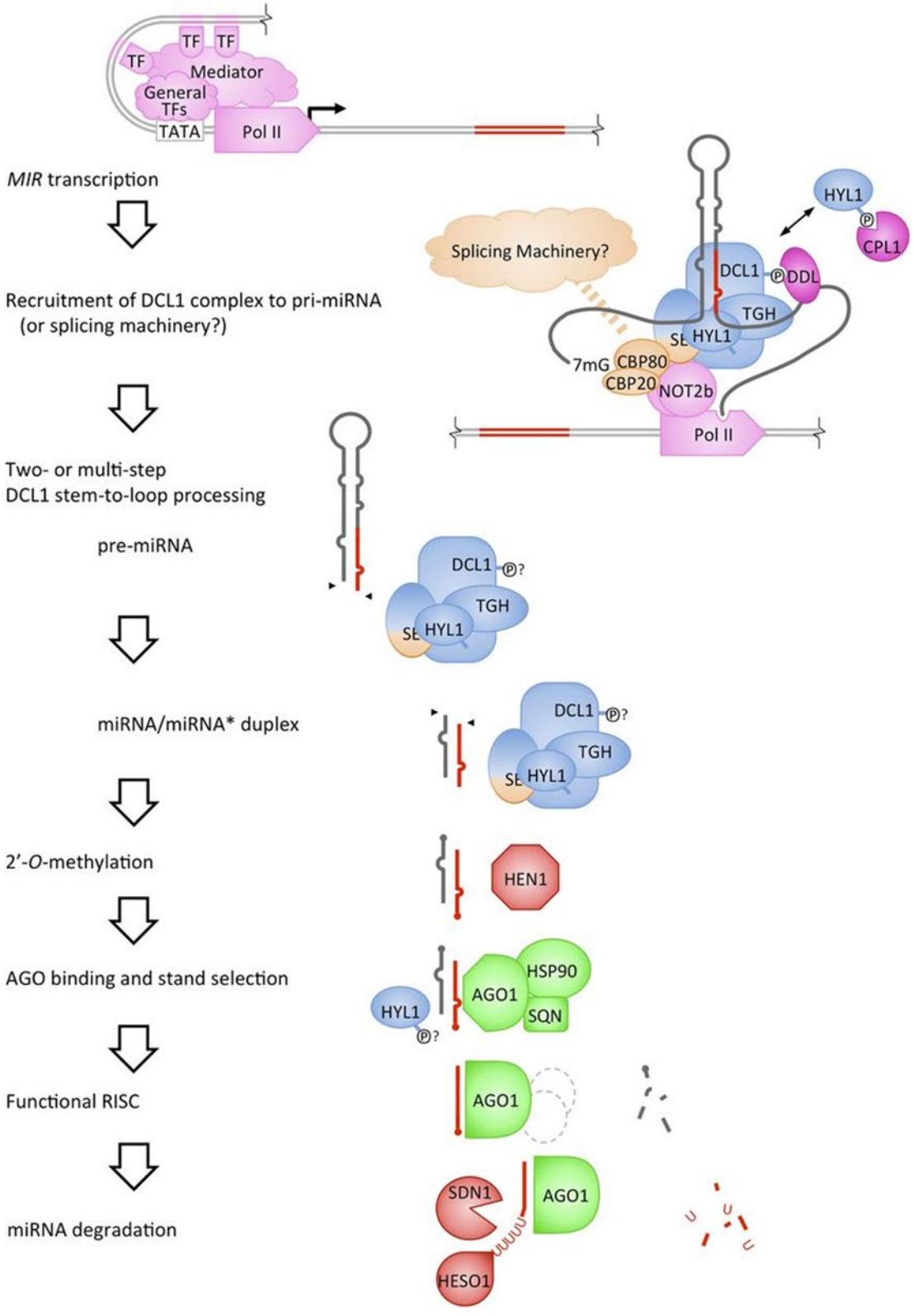


Figure 2. Biogenesis of plant microRNAs. Summary of miRNA processing and turnover, further details are described in the text. The color codes represent the known functions of proteins: transcription of MIR genes (pink), splicing (orange), processing by DCL1 (light blue), phospho-regulation (purple), RISC association (green), and miRNA stability and turnover (red). Dashed outlines depict unknown AGO1-binding proteins. The figure is taken with permission from Rogers and Chen 2013.

Until recently, it was not clear where miRNAs are loading into RISC. Based on the earlier reports, HASTY (HST), the homolog of the Exportin 5, transfers methylated miRNA/miRNA* from the nucleus to the cytosol before the RISC assembly (Park *et al.* 2005; Pontes *et al.* 2013; Yi *et al.* 2005). However, the latest evidence suggests that RISC is loaded in the nucleus and is then relocated to the cytoplasm by evolutionarily conserved protein CRM1/EXPORTIN1 (EXPO1) (Bologna *et al.* 2018). Recent studies suggest that HST functions as a scaffold for assembling the DCL1-Mediator complex on the genomic *MIRNA* loci, promoting pri-miRNA transcription and processing rather than nuclear export in *Arabidopsis (*Cambiagno et al. 2020).

In animals, exonuclease-mediated mRNA degradation begins with decapping and deadenylation. The degradation of the 3′-poly(A)-tail is carried out by two general deadenylation complexes: PAN2/3, which is recruited by poly(A)-binding protein (PABP) and CCR4-NOT, recruited by several mRNA-binding proteins, such as Pumilio. The assembly of Pan2/3 and CCR4-NOT complexes on mRNA is directed by AGO-bound miRNAs and adaptor proteins GW182/TNRC6 (Braun *et al.* 2011; Chekulaeva *et al.* 2011; Fabian *et al.* 2011). Deadenylation promotes the removal of a 5′- guanosine cap by decapping enzyme DCP2. The decapped mRNA undergoes 5′-3′ degradation by exoribonuclease XRN1 or 3′-5′ decay by the multisubunit exosome complex (Figure 3).

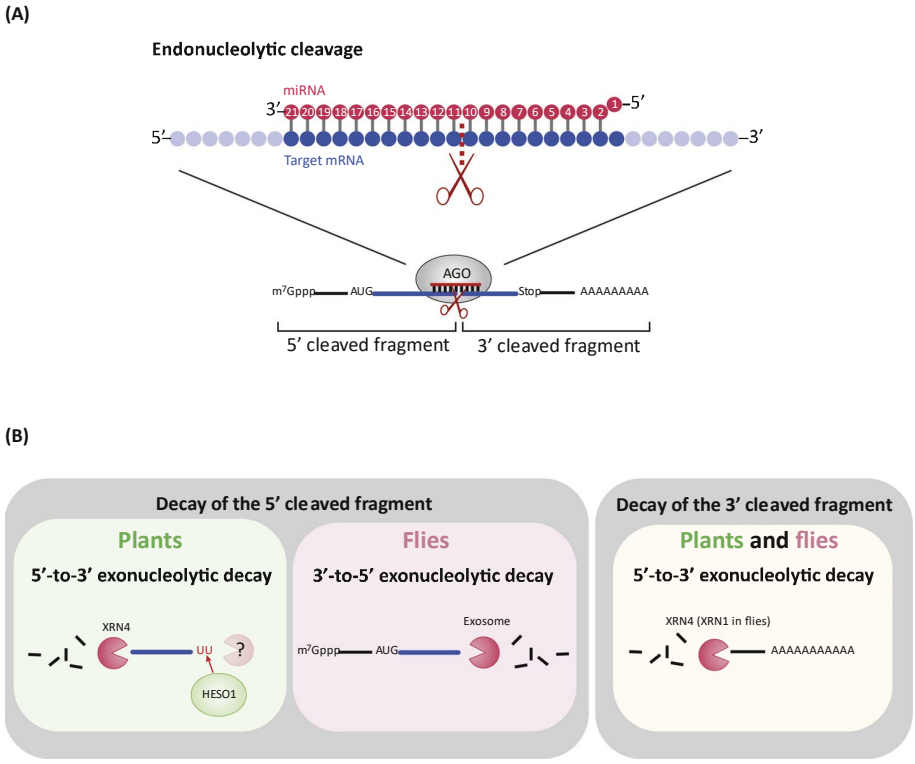


Figure 3. miRNA-Mediated mRNA Decay Pathways. (A) Plant miRNAs bind to the complementary target sites in mRNAs and trigger slicing at the position 10/11 nt. **(B)** In plants, the 5′-cleavage fragments undergo uridylation by HESO1 and 5′-3′ cleavage by XRN4. In flies, the 5′-product of siRNA-mediated cleavage is further degraded in the 3′-5′direction by the exosome. The 3′-cleaved products undergo 5′-3′ decay by XRN4 or XRN1 in plants or flies, respectively. Figure modified from Iwakawa and Tomari, 2015.

In contrast to yeast and animals, the knowledge about the mRNA turnover mechanism in plants is limited. However, because plants do not have a functional homolog of GW182 or miRNA-mediated mRNA deadenylation, this mechanism significantly differs from its animal counterpart resembling the mode of action of animal siRNAs. Plant miRNAs bind to mRNA with almost perfect complementarity and facilitate target RNA cleavage directly through AGO1 slicer activity. The resulting products of the initial endonucleolytic cleavage are further degraded by the 5'-to-3'endonuclease XRN4, a homolog of Drosophila XRN1 (Figure3). In contrast to animals, plant miRNAs repress fewer targets; however, many of these targets are critical regulatory proteins, such as transcription factors. The specific patterning events during embryo development is likely dependent on the steady-state cellular levels of miRNAs and their targets. Understanding the spatiotemporal dynamics of miRNA-mediated cleavage will also shed light on the complex regulatory networks underlying plant embryo development.

4.3 Goals of the thesis

- **Profile miRNA cleavage products genome-wide.** The RISC-mediated mRNA cleavage is a major mechanism of miRNA activity in plants. Because plant miRNAs generally bind their targets with nearly perfect complementarity, it is relatively easy to predict these targets. However, genome-wide validation of miRNA-cleavage products from developing embryos remains a challenging task. Early *Arabidopsis* embryos are small and deeply embedded in maternal tissues, making embryonic samples hard to collect and prone to heavy contamination from surrounding seed coat and endosperm (Schon and Nodine, 2017).

Standard biochemical approaches for individual miRNA:target validation, such as RT-PCR and Western Blotting or miRNA co-immunoprecipitation (co-IP) with RISC components, enable determining the levels of RNA transcripts or proteins, respectively; however, can not be applied on a global scale. Alternatively, 5'-RACE allows direct and high-throughput analysis of miRNA-cleavage products but requires large amounts of input material. We developed a powerful method, nanoPARE (parallel analysis of RNA ends), to profile miRNA cleavage products genome-wide from sub-nanogram amounts of total RNA. Using nanoPARE, we profiled miRNA:target interactions during eight embryonic developmental stages. Together with excellent transcriptomic resources available in the lab, nanoPARE data enabled identifying several miRNA families that cleave and repress dozens of diverse transcription factors.

- **Analyze the repressive activities of selected miRNAs at cellular resolution.**

During early development, miRNAs down-regulate their mRNA targets in specific domains to enable proper embryonic patterning. In order to examine the spatiotemporal repressive activities of miRNAs, I have used fluorescent protein-based reporters. These miRNA sensors consisted of miRNA-targeted nuclear-localized GFP under the control of the ubiquitously expressed promoter so that the GFP signal was silenced in the cells with active miRNAs.

Together with the miRNAs and transcription factor localization data, obtained from *in situ* hybridization experiments, this information provided an overview of cell-specific miRNA accumulation and activity dynamics throughout development.

- **Determine the embryonic functions of individual miRNA:target interactions.** Although miRNA-deficient *dcl1-5/+* mutant embryos exhibit pleiotropic developmental defects (Nodine and Bartel, 2010), little was known about the cell-specific contributions of individual miRNA:target interactions to embryo morphogenesis. I generated transgenic plants, which express miRNA-resistant transcription factors, including AUXIN RESPONSE FACTOR8 (ARF8), ARF17, SQUAMOSA PROMOTER BINDING PROTEIN-LIKE10 (SPL10), SPL11, PHABULOSA (PHB), TCP4 and examined their morphology. The results showed that miRNAs regulate embryo morphogenesis by affecting cell-specific accumulation and localization of transcription factors.

5. NanoPARE: parallel analysis of RNA 5′ ends from low-input RNA

type published

Genome Research 28, 1931 - 1942 (2018)

doi: 10.1101/gr.239202.118

authors Michael A. Schon* (MAS), Max J. Kellner *(MJK), Alexandra Plotnikova (AP), Falko Hofmann (FH) and Michael D. Nodine (MDN)

Contribution AP performed experiments including dissection of petals, sepals, anthers, stigmas, and ovules for the floral tissue atlas; preparation of nanoPARE libraries from floral tissues, *xrn4-5* mutant, Xrn1-treated and wild-type floral buds (used for Figures 2, 4, 5, S1, S3, S4, S6, S7, S9). Further author contributions are listed in section 5.7 on page 35.

5.1 Abstract

Diverse RNA 5′ ends are generated through both transcriptional and post-transcriptional processes. These important modes of gene regulation often varyacross cell types, and can contribute to the diversification of transcriptomes and thus cellular differentiation. Therefore, the identification of primary and processed 5′ ends of RNAs is important for their functional characterization. Methods have been developed to profile either RNA 5′ ends from primary transcripts or the products of RNA degradation genome-wide. However, these approaches either require high amounts of starting RNA or are performed in theabsence of paired gene-body mRNA-seq data. This limits current efforts in RNA 5′ end annotation to whole tissues and can prevent accurate RNA 5′ end classification due to biases in the datasets. To enable the accurate identification and precise classification of RNA 5′ ends from standard and low-input RNA, we developed a next-generation sequencing-based method called nanoPARE and associated software. By integrating RNA 5′ end information from nanoPARE with gene-body mRNA-seq data from the same RNA sample, our method enables the identification of transcription start sites at single-nucleotide resolution from single-cell levels of total RNA, as well as small RNA-mediated cleavage events from at least 10,000-fold less total RNA compared to conventional approaches. NanoPARE can therefore be used to accurately profile transcription start sites, noncapped RNA 5′ ends and small RNA targeting events from individual tissue types. As a proof-of-principle, we utilized nanoPARE to improve *Arabidopsis thaliana* RNA 5′ end annotations and quantify microRNA-mediated cleavage events across five different flower tissues.

5.2 Introduction

Diverse RNA 5′ ends are generated during and after transcription as the result of a variety of gene regulatory functions. Alternative transcription start sites (TSS) can generate RNA isoforms that differentially impact cellular activities. Alternative TSS have also been demonstrated to affect downstream translation and protein function through inclusion or exclusion of regulatory N-terminal peptides such as upstream open reading frames or protein localization sequences (Haberle *et al.* 2014; Cheng *et al.* 2018; Ushijima *et al.* 2017). Post-transcriptional maturation of non-coding RNAs such as those involved in splicing (snoRNA) or translation (rRNAs, tRNAs) also generates diverse RNA 5′ ends (Granneman *et al.* 2011; Wang *et al.* 1988; Henras *et al.* 2015; Filipowicz and Pogacić 2002). Moreover, small regulatory RNAs such as microRNAs (miRNAs) and small interfering RNAs (siRNAs) can mediate endonucleolytic cleavage of target RNAs and are important regulators of development, genome stability and defense (Bartel 2004; Borges and Martienssen 2015). Therefore, the identification of RNA 5′ ends derived from transcriptional and post-transcriptional processes is important for the functional characterization of RNA molecules. Next-generation sequencing (NGS) based methods have recently been used to identify RNA 5′ ends genome-wide. For example, TSS profiling using Cap Analysis of Gene Expression

(CAGE) led to the annotation of TSS from polyadenylated mRNA and long non-coding RNA (Andersson *et al.* 2014; Hon *et al.* 2017). TSS profiling has also provided fundamental insights into how RNA isoforms with different 5′ ends modulate gene function, as recently demonstrated in *Arabidopsis thaliana* (*Arabidopsis*) where the phytochrome photoreceptor regulates alternative promoter usage in a light-dependent manner that ultimately leads to gene products with distinct functions and subcellular localization (Ushijima *et al.* 2017). Methods referred to as PARE (Parallel Analysis of RNA Ends), or degradome sequencing enrich for 5′ monophosphorylated RNAs, which include small RNA (sRNA) mediated cleavage products. PARE methods, therefore, enable the genome-wide profiling of sRNA target sites and have been instrumental in characterizing the mechanistic basis of sRNA-mediated developmental and physiological processes (Addo-Quaye *et al.* 2008; German *et al.* 2008; Gregory *et al.* 2008).

Cell type-specific TSS and sRNA-mediated cleavage events contribute to the diversification of cellular transcriptomes and thus can impact cellular differentiation (Zhou *et al.* 2015; Knauer *et al.* 2013; Carlsbecker *et al.* 2010; Kidner and Martienssen 2004; Williams *et al.* 2005; Karlsson *et al.* 2017; Miyashima *et al.* 2013). However, due to technical limitations, RNA 5′ ends have traditionally been profiled on whole organisms or tissues composed of multiple cell types, and thus RNA 5′ ends that exist in specific cell types will be depleted in the corresponding final datasets. Recently, methods utilizing reverse transcriptase template-switching have been developed to profile TSS from the low amounts of RNA obtainable from specific cell types and individual cells (Arguel *et al.* 2017; Islam *et al.* 2011; Cole *et al.* 2018). However, it remains a challenge to identify and confidently assign bona fide TSS to their corresponding genes due to technical artifacts and variable performance between protocols (Cocquet *et al.* 2006; Tang *et al.* 2012; Adiconis *et al.* 2018). Moreover, PARE methods require high amounts of input RNA that limit their application to samples that can be collected in bulk.

5.3 Results

RNA 5′ end enrichment from low-input RNA samples

To profile RNA 5′ ends genome-wide from low amounts of total RNA, we developed an NGS based method called nanoPARE (Parallel Analysis of RNA 5′ Ends from low-input RNA; Fig. 1; see Methods). First, we followed the Smart-seq2 protocol through cDNA preamplification to produce full-length cDNAs with template-switching oligonucleotide (TSO) sequences at their 5′ ends (Picelli *et al.* 2013). Tn5 transposase was then used to fragment the cDNA and ligate adaptors for NGS library preparation (Ramsköld *et al.* 2012). To selectively amplify the 5′ ends of the cDNAs, we performed PCR on the tagmented products using primers complementary to TSO and inserted transposase adapter sequences. The resulting amplicons were then used for final PCR amplification with indexed Illumina-adapter primers for next-generation sequencing. Additionally, tagmented products corresponding to nonterminal, or

body, regions of transcripts were amplified according to the Smart-seq2 method (Picelli *et al.* 2013). By combining the 5′ end and body sequence information from the cDNA of a single sample, the 5′ ends of RNA can be precisely identified at a single-nucleotide resolution, as demonstrated below.

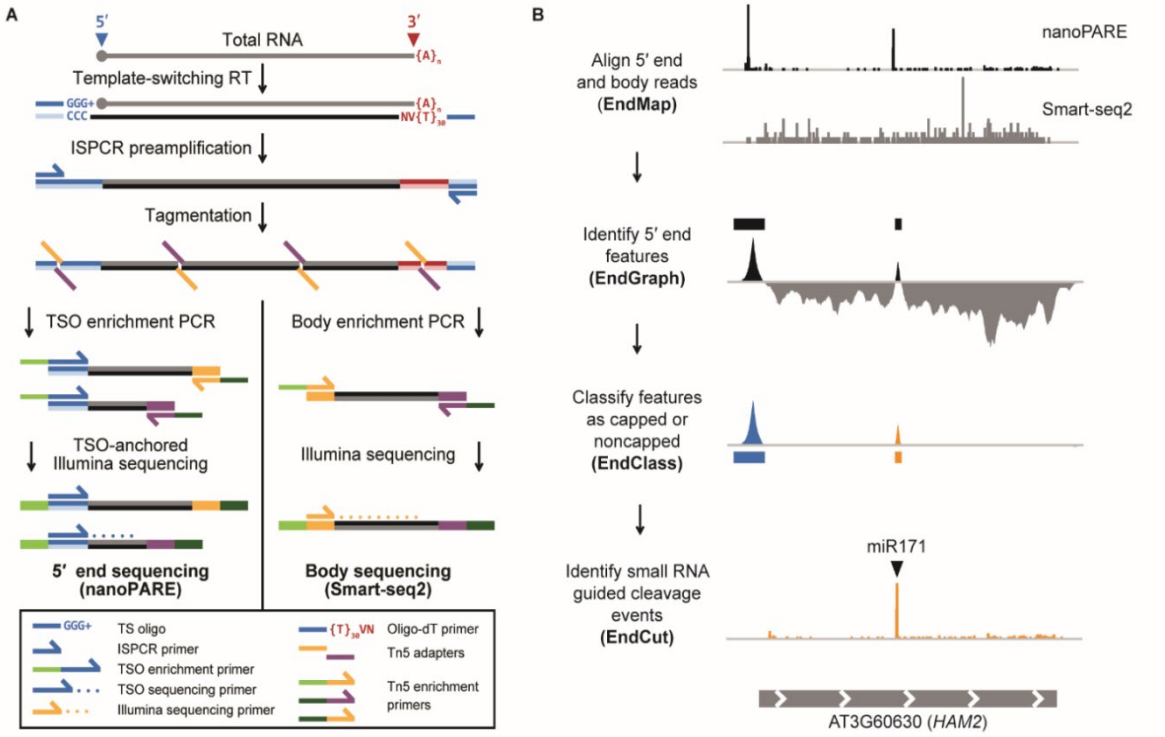


Figure 1. Workflow of nanoPARE and EndGraph. (A) Diagram of the nanoPARE protocol, which enables construction of a stranded 5′ end library (left) in parallel with a non-stranded transcript body library (Smart-seq2, (Picelli *et al.* 2013), from the same RNA sample. All oligonucleotides are labeled in the legend. **(B)** Workflow of the nanoPARE data analysis pipeline for identifying distinct capped and noncapped 5′ end features from a paired nanoPARE and Smart-seq2 sequencing library. Diagram represents the output of each step, using HAM2 as an example.

Identification of capped and noncapped 5′ end features

Template switching occurs at the 5′ ends of RNA templates with or without 7-methylguanosine (m7G) cap structures (Cloonan *et al.* 2008; Harbers *et al.* 2013). We reasoned that template switching could be used to identify RNA 5′ end features genome-wide regardless of their cap structure. However, spurious 5′ ends could be produced by a variety of technical artifacts, including random fragmentation of RNA *in vitro*, stalling of the reverse transcriptase enzyme, PCR amplification bias, or "strand invasion" by the template-switching oligo (Cocquet *et al.* 2006; Tang *et al.* 2012), which make it difficult to distinguish biological signal from noise. A major source of non-random bias is internal template switching at sites complementary to the TSO 3′ end, and signal at these sites can be removed *in silico* (Tang *et al.* 2012). For other sources of noise, we developed a "scaling factor" for comparing nanoPARE libraries to a counterpart Smart-seq2 library from the same cDNA (see Methods). The scaling factor estimates the expected ratio of 5′ end containing cDNA fragments to gene body fragments after tagmentation with the assumption that all RNA is full-length. After applying this scaling factor, genomic regions are identified that produce more terminal signal

than nonterminal signal, and these regions are isolated as 5′ end features using the software EndGraph (see Methods). We applied EndGraph to a paired collection of nanoPARE/Smart-seq2 libraries prepared from 5 ng of total RNA isolated from *Arabidopsis* floral buds in biological triplicate and reproducibly identified a total of 22,852 5′ end features from polyadenylated RNA in at least two of the three biological replicates (Supplemental Data S1; Supplemental Data S2). Reverse transcription produces untemplated cytosines at the template 5′ terminus that can base pair with a m7G cap to yield untemplated upstream guanosine (uuG) in the cDNA between the TSO and genome-matching sequence (Cumbie *et al.* 2015; de Rie *et al.* 2017). These uuGs can be used to filter 5′ end data produced by template-switching protocols and isolate 5′-capped transcription start sites (Cumbie *et al.* 2015). Indeed, uuG consistently appeared in roughly 15% of nanoPARE reads per library, occurring from 4 to 10 times more frequently than other nucleotides (Supplemental Fig. S1). We analyzed the total proportion of reads in all 5′ end features that contained uuG, and we could observe a striking bimodal distribution, with most 5′ end features (20,679; 90.5%) containing >10% uuG reads (Fig. 2A).

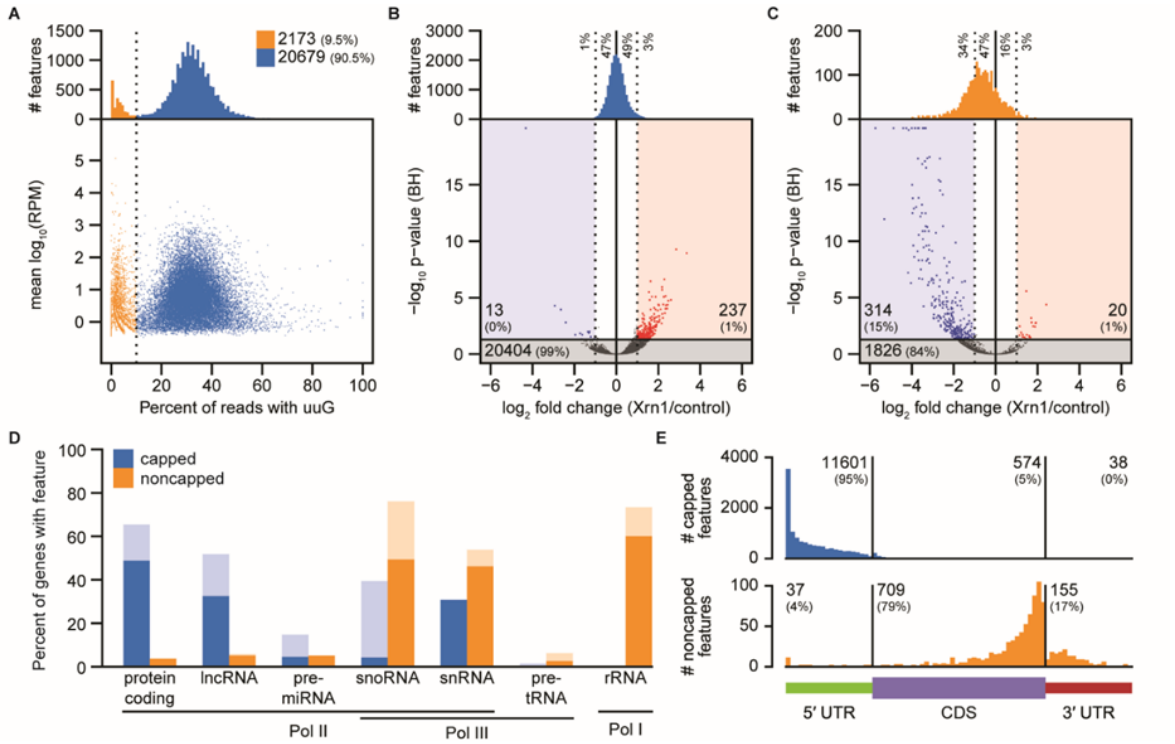


Figure 2. Identification of capped and noncapped 5′ end features with EndGraph.
(A) RNA 5′ end features identified from 5 ng of floral bud total RNA, distributed by the proportion of nanoPARE reads containing an upstream untemplated guanosine (uuG). The vertical line separates putative noncapped features (low-uuG, orange) from putative capped features (high-uuG, blue). **(B)** Volcano plot of the change in read abundance for putative capped features after digestion with Xrn1 exonuclease. Bar plots depict the distribution of all capped features by fold change versus control. Dotted lines delimit a two-fold change in feature abundance. Log2 fold change and Benjamini-Hochberg adjusted p-values (BH) were calculated by DEseq2. Horizontal line demarcates an adjusted p-value of 0.05. **(C)** Volcano plot as in **(B)** for putative noncapped features. **(D)** Capped and noncapped features overlapping TAIR10 genes classified by gene type. Lighter bars include features up to 500 nucleotides upstream of the annotation. **(E)** Positional distribution of capped (top) and noncapped (bottom) features that overlap protein-coding genes.

Genomic distributions of capped and noncapped 5′ features

Capped and noncapped features have distinct distributions in the genome. Capped features have an even distribution across the non-centromeric region of the nuclear genome, whereas noncapped features are highly concentrated on the mitochondrial and chloroplast genomes and more sparsely dispersed in the nuclear genome (Supplemental Fig. S2). The 5′ ends of chloroplast and mitochondrial RNA do not possess m7G caps (Monde *et al.* 2000; Grohmann *et al.* 1978; Legen *et al.* 2002). Accordingly, the 5′ features mapping to chloroplasts and mitochondria were classified as 94% and 95% noncapped, respectively. Moreover, capped and noncapped features localized to different gene types (Fig. 2D). Consistent with caps being associated only with RNA polymerase II (polII) transcripts, nanoPARE identified predominantly capped features for mRNA, long non-coding RNA, and primary microRNA (Supplemental Fig. S3). Several noncapped features mapping to pre-tRNA and rRNA loci, which are not transcribed by polII, could also be identified (Supplemental Fig. S4, S5). While these products normally do not possess a poly(A) tail, a subset will be transiently polyadenylated by the TRAMP complex prior to their degradation (Hopper *et al.* 2010), and nanoPARE may be detecting this subset. Even within a class of RNA, the behavior of capped and noncapped features is quite distinct. For example, a majority of snoRNAs in *Arabidopsis* are transcribed in tandem arrays of two or more species from a single polII precursor (Dieci *et al.* 2009). EndGraph identified both a noncapped feature at the mature 5′ terminus and a capped feature upstream of many annotated snoRNAs (Fig. 2D). Several primary snoRNAs contained additional noncapped features not predicted by TAIR10 annotations, and these are supported by genome-wide profiling of non-coding RNAs (Supplemental Fig. S3) (Wang *et al.* 2014).

Protein-coding mRNA also displays a distinct distribution of capped and noncapped 5′ end features. Among capped features overlapping nuclear protein-coding genes, 95% mapped to the annotated 5′ UTRs (Fig. 2E). Noncapped features showed an opposite trend, accumulating closer to the 3′ termini of protein-coding genes, particularly just upstream of the stop codon. This pattern closely resembles the reported pattern of co-translational decay detected in PARE data, which is mediated by the major cytoplasmic exonuclease EXORIBONUCLEASE4 (XRN4) (Yu *et al.* 2016; Hou *et al.* 2016). To test whether we detect the steady-state by-products of XRN4 digestion with nanoPARE, we generated nanoPARE and Smart-seq2 libraries for floral buds of *xrn4*-5 mutants and performed de novo 5′ feature identification. All features identified in wild-type floral buds (with or without Xrn1 digestion) and/or *xrn4-5* floral buds were combined to produce a unified set of 5′ end features (Supplemental Data S2). Because XRN4 is restricted to the cytoplasm (Kastenmayer and Green 2000), non-cytoplasmic 5′- monophosphorylated ends should not be increased in *xrn4-5* loss-of-function mutants. As predicted, non-cytoplasmic 5′ end features associated with mature or primary noncoding RNAs, including nuclear 3′ cleavage products of pri-snoRNA, pre-miRNA, as well as mitochondrial and chloroplast pri-tRNA (Supplemental

Data S3), were not increased in *xrn4-5* floral buds relative to wild-type. In contrast, these features were sensitive to Xrn1 digestion *in vitro* (Supplemental Fig. S6). Consistent with wild-type XRN4 digesting full-length decapped mRNA, noncapped features upstream and adjacent to stop codons were globally decreased in relative abundance in *xrn4-5* mutants, concomitant with a relative increase in reads contained by capped features (Supplemental Fig. S7). Together, these trends predict an increase in full-length transcripts with 5′-monophosphates in *xrn4* mutants. Indeed, the accumulation of full-length decapped mRNA has been reported for some transcripts in *xrn4* mutants (Gregory *et al.* 2008). We reanalyzed public PARE data from wild-type and *xrn4-5* inflorescences (German *et al.* 2008) and found a global average increase by more than three-fold in the proportion of 5′ monophosphorylated RNA ends mapping to all capped features defined by nanoPARE (Supplemental Fig. S7). Overall, the capped and noncapped 5′ features identified with nanoPARE support the existing model of XRN4 as a general RNA decay factor that acts downstream of decapping.

TSS characterization

To test the reproducibility of nanoPARE to detect 5′ end features from low-input RNA, we generated nanoPARE libraries from a dilution series of the original floral bud total RNA in triplicate: 1 ng, 100 pg, and 10 pg of total RNA input, which is typically less than or equal to the amount of total RNA found in a single cell (Brennecke *et al.* 2013; Ramsköld *et al.* 2012). We performed de novo feature identification using the Smart-seq2 libraries from 5 ng of RNA as a background model. To assess the sensitivity of the method at recovering genuine capped transcription start sites, we compared the capped features of the dilution series to a set of 9,326 transcription start sites identified by a cap-specific 5′ sequencing protocol (Paired-End Analysis of Transcription Start Sites; PEAT) applied to whole *Arabidopsis* roots (Morton *et al.* 2014). As a baseline, we tested whether the *Arabidopsis* reference annotations, TAIR10 and Araport11, contained a transcript model with a 5′ end within 50 bp of a given PEAT peak. TAIR10 detected 75% of PEAT peaks under this definition (Fig. 3A). The more recent Araport11 annotations performed much worse at accurately detecting transcription start sites. Only 2,239 (24%) of PEAT peaks fell within 50 bp of any transcript 5′ ends defined in Araport11, which is likely due to the systematic overextension of transcript UTRs during transcript model assembly. Despite the different tissue types and preparation methods, nanoPARE outperformed both reference annotations at detecting experimentally validated transcription start sites, down to 1 ng of total RNA (Fig. 3A). Furthermore, when comparing the precision of peaks, all nanoPARE dilutions, including those generated from 10 pg of total RNA, had a higher likelihood of agreeing with the PEAT data on the exact nucleotide position of the transcription start site peak (Fig. 3A, B). Finally, we examined the sensitivity of the method by comparing identified capped features with the transcript abundance as measured

by Smart-seq2 (Fig. 3C). Remarkably, nanoPARE reproducibly identified at least one capped feature overlapping 76.6% of all protein-coding transcripts detected at or above 0.1 transcripts per million (TPM) when libraries were generated from 5 ng of total RNA (18,295/23,900 genes). This value increased to 91.9% for transcripts detected at a threshold of 1 TPM (18,111/19,706 genes), and 99.0% for transcripts of at least 10 TPM (11,165/11,294 genes). In contrast, only 0.4% of transcripts not detected with Smart-seq2 (0 TPM, 20/5,419 genes) were assigned a capped feature. Overall, the 5 ng, 1 ng, and 100 pg samples performed similarly well, especially for robustly detected transcripts. Sensitivity reduced substantially between 100 pg and 10 pg of total RNA without affecting the precision of the capped features identified (Fig. 3A, B). Multiple distinct TSS could even be identified at all dilutions for certain highly expressed genes (Fig. 3D; Supplemental Data S2). Therefore, nanoPARE capped features represent genuine transcription start sites, and can be used for transcription start site annotation with as little as 10 pg of input RNA.

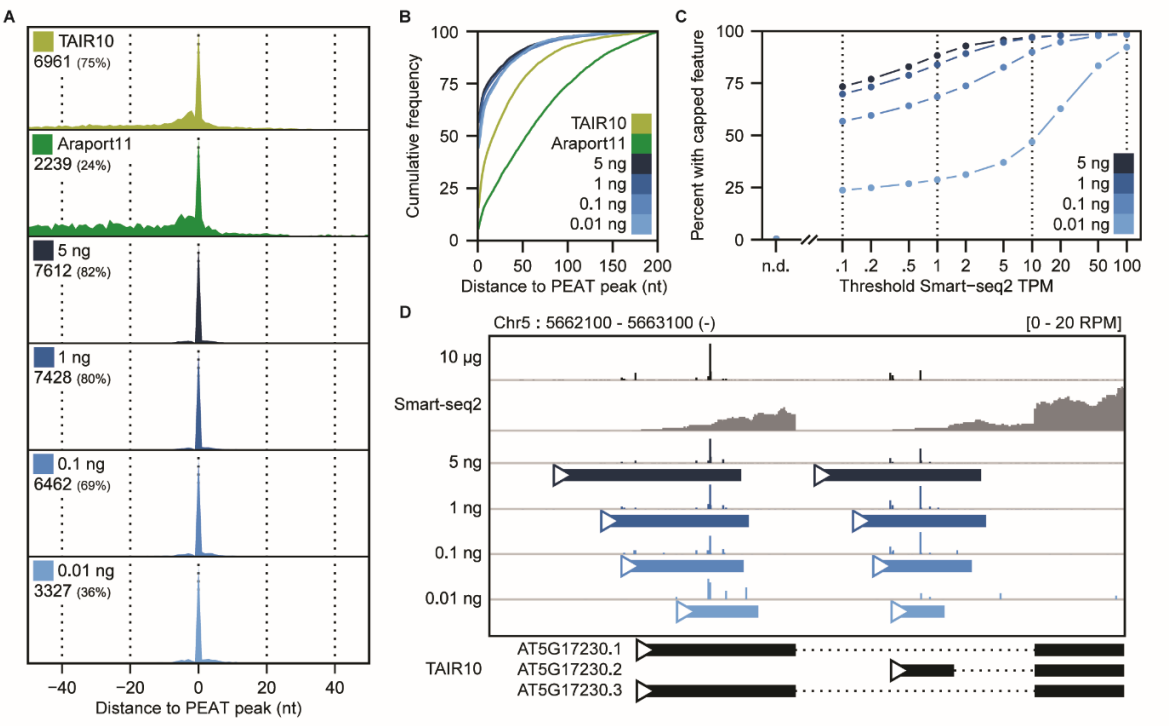


Figure 3. Sensitive low-input transcription start site detection with nanoPARE.
(A) Recall of capped peaks identified with PEAT (Morton *et al.* 2014) in two *Arabidopsis* reference annotations (TAIR10 and Araport11) and in nanoPARE features detected from a dilution series of total RNA input. Numbers indicate how many PEAT peaks have a 5′ end feature within 50 bp in the test dataset. **(B)** Cumulative frequency distribution of positional error for all 5′ features within 200 nt of a PEAT peak. **(C)** Sensitivity of nanoPARE in detecting capped 5′ features for nuclear protein-coding genes as a function of their abundance measured by Smart-seq2. Points indicate the percent of transcripts above the given threshold abundance (in transcripts per million, TPM) that contain a capped feature identified in at least 2 of 3 biological replicates. **(D)** Integrated Genomics Viewer (IGV) browser image of nanoPARE reads from the dilution series mapping to two transcription start sites of the PSY locus. Y-axis shows mean reads per million (RPM) across three biological replicates for each dilution. Solid colored bars mark capped features identified by EndGraph in each dilution.

Detecting sRNA-mediated cleavage sites

In plants, Argonaute-bound sRNAs recognize highly complementary 20-22 nt target sites and mediate target RNA cleavage precisely between the 10th and 11th nucleotides of the sRNA-target duplex (Jones-Rhoades and Bartel 2004; Llave *et al.* 2002; Kasschau *et al.* 2003). Because nanoPARE reads map to the 5′ ends of noncapped transcripts (Fig. 2), we reasoned that the first position of nanoPARE reads should also be enriched precisely at sRNA target cleavage sites and thus allow their identification from low-input RNA samples. To test whether nanoPARE reads from libraries generated with low-input RNA samples were enriched at sRNA target cleavage sites, we examined predicted cleavage sites for either miRNAs or trans-acting siRNAs (tasiRNAs) in libraries prepared from 5 ng of total RNA isolated from floral buds. The 5′ ends of nanoPARE reads were enriched at cleavage sites pairing to highly complementary miRNAs, including those from previously characterized miRNA cleavage sites in wild-type (Col-0) flowers (Fig. 4A, B; Supplemental Fig. S8, S9). As expected for sRNA-directed cleavage products that are 5′ monophosphorylated, the number of nanoPARE reads at cleavage sites was reduced when RNA was incubated with the Xrn1 exoribonuclease prior to library generation. Conversely, nanoPARE read enrichment at cleavage sites was increased in *xrn4-5* exoribonuclease mutants, which stabilize sRNA cleavage products (German *et al.* 2008; Souret *et al.* 2004) (Fig. 4A, B; Supplemental Fig. S8, S9). In addition to biological variation between tissues or genotypes, variability between nanoPARE/PARE libraries can also be largely due to technical differences in RNA quality and quantity, as well as library complexity and sequencing depth. Therefore, we developed software called EndCut that employs empirically determined null models from randomized versions of each sRNA to compute the likelihood that nanoPARE read 5′ ends are enriched at predicted target sites greater than expected by chance in each library (Fig 4C, D; Supplemental Fig. S10; see Methods). EndCut uses two metrics to calculate this likelihood: the level of sRNA-target complementarity (Allen score) and the number of read 5′ ends at predicted cleavage sites divided by the maximum number detected at a single site within 20 or 50 nt of flanking transcribed regions (fold-change). To assess the validity of EndCut, we examined the proportions of nanoPARE read 5′ ends within and adjacent to miRNA cleavage sites determined to be significant in at least one of the nanoPARE libraries prepared from Col-0 floral bud RNA without Xrn1 treatment (Col-0 (-Xrn1)). As expected for miRNA cleavage sites, the number of nanoPARE reads at these sites was significantly decreased 9.7-fold upon Xrn1 treatment (Col-0 (+Xrn1)) or significantly increased 1.8-fold in *xrn4-5* mutants (P-values =1.08×10-26 and 0.037, respectively; one-tailed K-S tests) (Fig. 4B). Moreover, we compared the number of significant cleavage sites identified by EndCut in Col-0 floral buds either treated or not treated with Xrn1 prior to library construction, as well as from *xrn4-5* and tasiRNA-deficient *dcl234* mutant floral buds (Gasciolli *et al.* 2005; Henderson *et al.* 2006; Howell *et al.* 2007; Yoshikawa *et al.* 2005).

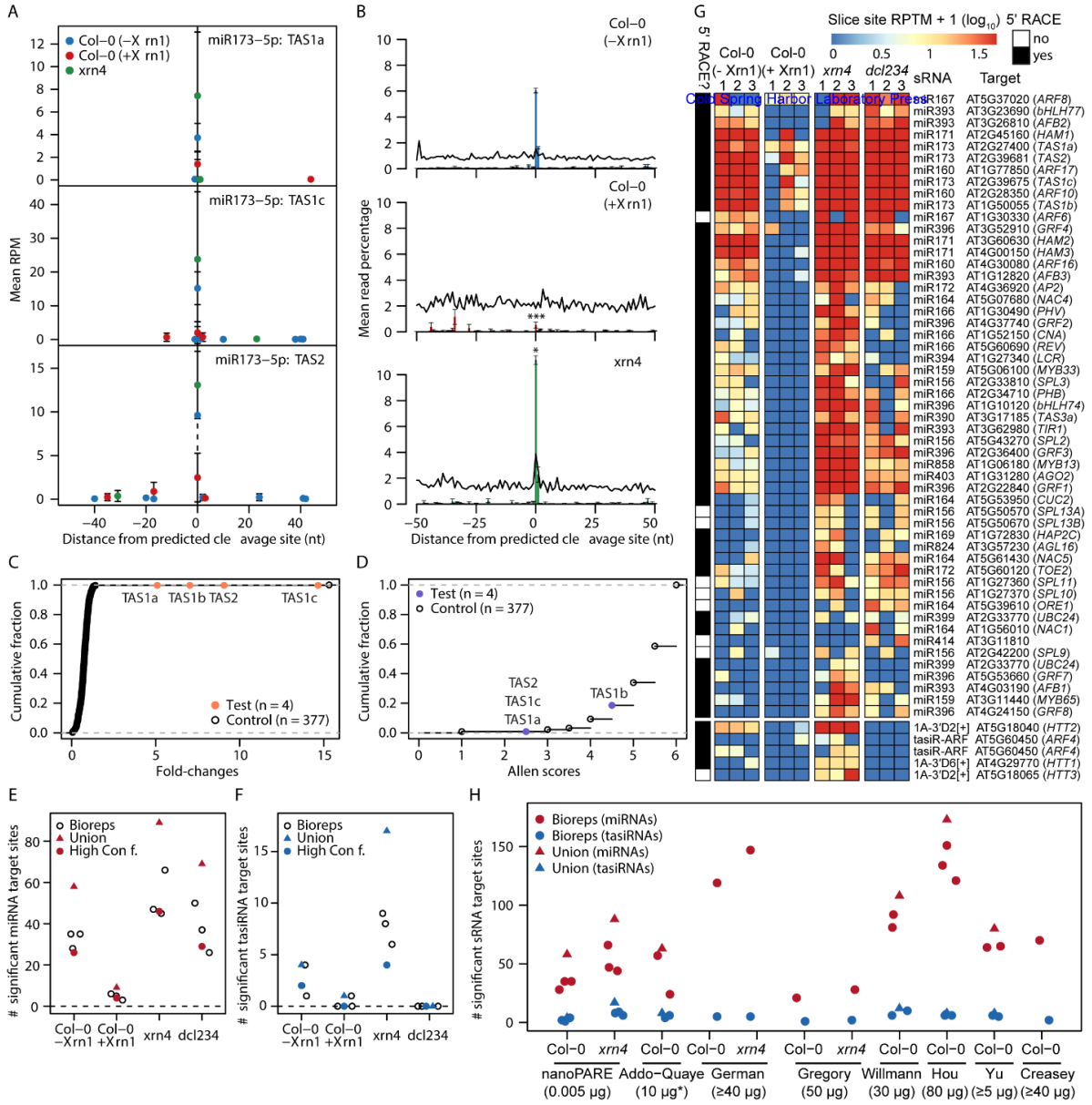


Figure 4. Detection of sRNA-mediated cleavage sites. (A) Scatter plot illustrating the number of nanoPARE read 5′ ends per million transcriptome-mapping reads (RPM) within 50 nt of predicted miR173-5p directed cleavage sites in TAS1a (top), TAS1c (middle) and TAS2 (bottom) transcripts. Mean RPM values of three biological replicates are shown for libraries prepared from 5 ng of total RNA from wild-type (Col-0) floral buds either not incubated with Xrn1 (Col-0 (-Xrn1)) or incubated with Xrn1 (Col-0 (+Xrn1)), or *xrn4-5* mutant floral buds (*xrn4*). Error bars represent standard errors of the means. **(B)** Number of nanoPARE read 5′ ends mapping within 50 nt of miRNA cleavage sites significantly detected by EndCut (Benjamini-Hochberg adjusted P-values < 0.05) in Col-0 (-Xrn1) libraries are shown as bar charts of the percentage of the total number of nanoPARE reads detected for each transcript in libraries prepared from Col-0 (-Xrn1) (top), Col-0 (+Xrn1) (middle) and *xrn4* (bottom) samples. Percentages of all predicted miRNA cleavage sites are shown as line graphs. * and *** indicate that the mean number of reads at predicted cleavage sites are significantly different in Col-0. (-Xrn1) libraries compared to either Col-0 (+Xrn1) or *xrn4* libraries (P-values < 0.05 and 0.001, respectively; one-tailed K-S tests). (C, D) Cumulative fractions of fold-changes **(C)** and Allen Scores **(D)** are shown for target sites predicted for either miR173-5p (test) or its randomized cohorts (control). **(E, F)** One-dimensional scatter plots illustrating the number of significant miRNA **(E)** or tasiRNA **(F)** target sites (Benjamini-Hochberg adjusted P-values < 0.05) detected in libraries prepared from Col-0 (–Xrn1), Col-0 (+Xrn1), *xrn4* or *dcl234* samples. Values for individual biological replicates (bioreps), all detected sites (union) and significant interactions observed in at least 2/3 bioreps (High conf.) are shown. **(G)** Heatmaps depicting the number of nanoPARE read 5′ ends per 10 million transcriptome-mapping reads (RPTM; log10) mapping to the high confidence miRNA (top) or tasiRNA (bottom) directed cleavage sites denoted in panels **E** and **F.** Small RNA families and corresponding targets are indicated beside each row, and targets previously verified by 5′ RACE are annotated.

Figure 4 (continued). (H) One-dimensional scatter plot showing the number of significant miRNA and tasiRNA target sites detected with EndGraph from nanoPARE libraries prepared from Col-0 or *xrn4* floral bud total RNA (nanoPARE), or published degradome/PARE libraries prepared from Col-0 or *xrn4* floral tissue total RNA. Published degradome/PARE libraries are indicated by the first author of the corresponding study: Addo-Quaye (Addo-Quaye *et al.* 2008), German (German *et al.* 2008), Gregory (Gregory *et al.* 2008), Willmann (Willmann *et al.* 2014), Hou (Hou *et al.* 2016), Yu (Yu *et al.* 2016) and Creasey (Creasey *et al.* 2014). The amounts of total input RNA (μg) used in each publication are indicated. The asterisk denotes that the Addo-Quaye samples were prepared from polyadenylated RNA instead of total RNA.

We identified 58 total miRNA target sites in Col-0 floral buds with a mean of 32.7 miRNA target sites among three biological replicates (Supplemental Data S4). The mean number of miRNA target sites identified upon Xrn1 treatment was significantly reduced 7-fold and significantly increased 1.6-fold in *xrn4-5* (P-values = 1.52×10-3 and 0.041, respectively; one-tailed t-tests) (Fig. 4E). We also identified 26 significant miRNA target cleavage sites in at least two biological replicates of Col-0, and these high-confidence sites had decreased and increased numbers of nanoPARE read 5′ ends in libraries from Xrn1-treated RNA and *xrn4-5* mutants, respectively (Fig. 4G). In addition to miRNAs, we also detected four significant tasiRNA target sites in Col-0 (-Xrn1) floral buds including two of which were detected in at least two biological replicates. Similar to what was observed for miRNA cleavage sites, tasiRNA cleavage sites significantly detected in Col-0 (-Xrn1) samples had reduced numbers of reads mapping to cleavage sites upon Xrn1 treatment (Fig. 4F, G; Supplemental Data S5). In contrast, both the number of significantly detected tasiRNA target sites and reads mapping to cleavage sites were increased in *xrn4-5* mutants (Fig. 4F, G). Importantly, whereas miRNA target sites significantly detected in Col-0 (-Xrn1) samples were generally unaffected in tasiRNA-deficient *dcl234* mutants, none of the corresponding tasiRNA target sites were significantly detected in *dcl234* mutants and zero reads corresponding to these target sites were observed (Fig. 4E-G). Out of the 58 high-confidence miRNA and tasiRNA target sites detected in either Col-0, *xrn4-5* or *dcl234,* 49 (84.5%) had been previously validated by a modified 5′ RACE technique (Supplemental Data S6), while 5 of the remaining 9 have genetic or expression data indicating that they are sRNA targets (Wu *et al.* 2006; Nodine and Bartel 2010; Wu *et al.* 2009; Wang *et al.* 2009). Based on the above biochemical and genetic tests, EndCut enables the accurate identification of sRNA-mediated cleavage events from nanoPARE data generated with as low as 5 ng of total RNA. PARE methods capture sRNA cleavage products through a series of adapter-RNA ligations, Type IIS restriction enzyme digestions and PCR amplifications (Addo-Quaye *et al.* 2008; German *et al.* 2008; Gregory *et al.* 2008). Because EndCut utilizes empirically determined null models based on randomized sRNAs, it can also be used to mitigate the effects of technical biases in conventional PARE datasets and help identify high-confidence target sites. As a proof-of-principle, EndCut was applied to 15 publicly available degradome/PARE datasets generated from at least 10 μg of total RNA from wild-type (Col-0) or *xrn4-5* mutant floral tissues. The number of significant cleavage sites detected with EndCut varied between publicly available datasets, but more sRNA-mediated cleavage events were detected from most PARE libraries compared to

nanoPARE libraries indicating that PARE libraries detect a greater diversity of cleavage sites when the starting amount of total RNA is not limiting (Fig. 4H; Supplemental Data S7, S8).

Tissue-specific miRNA-mediated cleavage sites

To test whether nanoPARE can detect small RNA-mediated cleavage sites that occur in specific tissue types, we applied the method to five different tissues dissected from whole flowers immediately after anthesis (Fig. 5A). The flower is comprised of four concentric whorls of tissues, which are specified by three transcription factor groups functioning in overlapping domains within the developing primordia. The coordinated action of these genes is described as the "ABC model" of flower development (Bowman *et al.* 1991b; Coen and Meyerowitz 1991). These transcription factors are known as group A, B, or C genes if they are expressed in the outer two whorls, middle two whorls, or inner two whorls of the developing flower, respectively. *Arabidopsis* possesses two A genes, two B genes, and a single C gene whose transcript spatial distributions are maintained through late flower development except for APETALA2 (AP2) mRNA, which has been observed in all four whorls of mature flowers (Jofuku *et al.* 1994; Mandel *et al.* 1992; Jack *et al.* 1992; Bowman *et al.* 1991a; Goto and Meyerowitz 1994). Upon comparing the relative abundance of 5′ capped transcript ends, we observed that nanoPARE faithfully recapitulated the expected spatial transcript patterns of all five homeotic genes (Fig. 5B). Therefore, these datasets can be used to quantify tissue-specific variation in RNA abundance. Tissue-specific variation of miRNA-target interactions on a genome-wide scale has not been reported in flowers, but individual studies indicate that miRNAs can suppress their targets in a tissue-specific manner (Wollmann *et al.* 2010; Wu *et al.* 2006; Liang *et al.* 2014). Upon performing nanoPARE on 10 ng of total RNA from either whole flowers or five individual floral tissues, we identified 41 miRNA target sites in at least two biological replicates of the same floral tissue (high-confidence sites; Fig. 5C and Supplemental Data S4). While the target cleavage sites directed by three miRNAs (miR160, miR171, and miR173) are robustly detected across all tissues examined, over half of the high-confidence interactions were enriched in specific tissues. To better estimate differences in miRNA-guided cleavage activity, we calculated the proportion of nanoPARE signal at the cleavage site relative to the total cleaved and capped (full-length) signal for each gene (Fig. 5D and Supplemental Fig. S11). For most miRNA target interactions, the proportion of cleaved transcripts varied strongly between tissues. For example, we found that AUXIN RESPONSE FACTOR 6 and 8 (ARF6/8) transcripts were preferentially cleaved in sepals, anthers and ovules, which is consistent with both MIR167 transcript levels as well as previous reports (Fig. 5E) (Rubio-Somoza and Weigel 2013; Wu *et al.* 2006). miR396 spatially restricts seven transcripts encoding GROWTH-REGULATING FACTOR proteins (GRF1/2/3/4/7/8/9) to the developing carpel (Liang *et al.* 2014). Although only one target (GRF1) was identified in wild-type flowers, six were identified in *xrn4-5* mutant flowers, indicating that the cleavage products from this gene family are efficiently cleared from wild-type cells (Fig. 4G,

Supplemental Fig. S11)

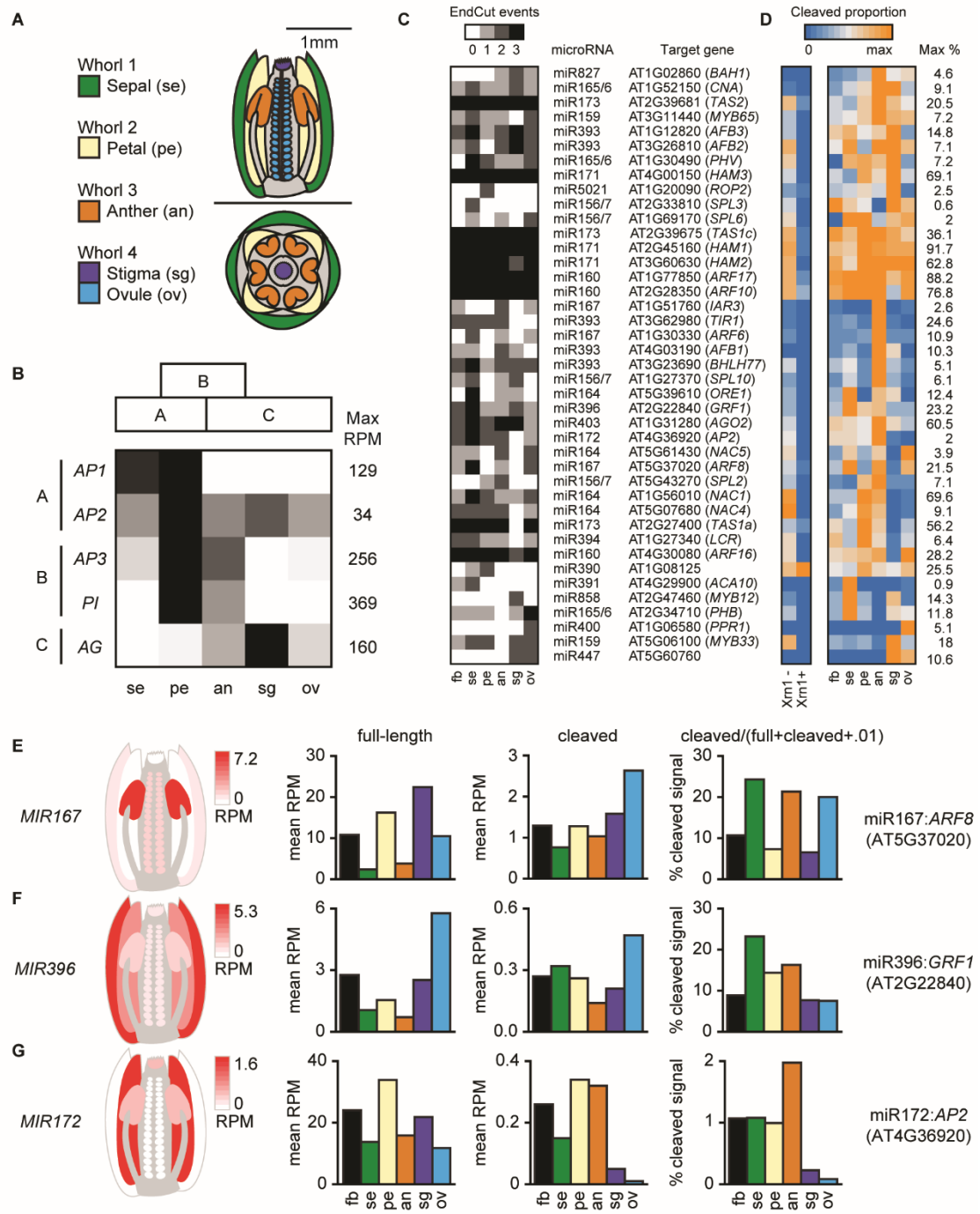


Figure 5. Tissue-specific miRNA-target interactions with nanoPARE. (A) Diagrams of a longitudinal section (top) and cross-section (bottom) of an *Arabidopsis* flower at the onset of anthesis. Tissue types isolated for nanoPARE libraries are color-coded as shown. **(B)** Relative expression of the five ABC model homeotic genes across the five tissue types in panel A. Each row is scaled from zero to the maximum observed reads per million (RPM) of a gene's capped feature. Expected spatial distributions based on the ABC model are shown as blocks above. **(C, D)** Heatmaps of 41 high confidence miRNA cleavage sites detected by nanoPARE in whole flowers (fb) and individual tissue types illustrating either the number of biological replicates in which the cleavage site was significantly detected (EndCut events) **(C)** or the proportion of cleaved signal to total full-length and cleaved signal **(D).** Each row is scaled to the maximum proportion observed for that interaction, which is indicated on the right. **(E-G)** (Left) Heatmaps of the summed primary transcript levels for three families of miRNA genes in flowers as measured by nanoPARE. Floral tissues match those labeled in panel **A**. (Right) Bar charts depicting the relative abundance of full-length RNA, truncated RNA with a 5′ end matching the miRNA cleavage site, and the proportion of cleaved RNA to the total cleaved and full-length signal, for the most strongly cleaved target of each of the three miRNA families to the left.

Despite their transient nature, miR396-directed GRF1 cleavage products accumulated to a higher proportion in non-carpel tissues than in stigmas or ovules (Fig. 5F), and full-length transcripts for the other targeted GRFs were restricted almost exclusively to ovules (Supplemental Fig. S11). Lastly, genetic data support a model whereby miR172 represses AP2 in whorl 3 to maintain stamen identity (Wollmann *et al.* 2010). Consistent with this model, we detected higher miR172 activity in anthers compared to other tissues (Fig. 5G). Because the tissue-enriched miRNA-guided cleavage events detected by nanoPARE are in good agreement with these experimentally supported examples, we conclude that nanoPARE can be used to detect sRNA-guided cleavage events in specific tissue types.

5.4 Discussion

NanoPARE can accurately profile RNA 5′ ends genome-wide from low amounts of total RNA. Because TSS partition the genome into transcribed and cis-regulatory regions, their accurate identification is critical for transcriptome assembly and prediction of regulatory binding sites. Moreover, TSS can vary among cell types and thus their identification from low-input RNA samples can increase our understanding of diverse RNA processing events and corresponding functions. NanoPARE's integrative approach of combining RNA 5′ end enrichment and full-length Smart-seq2 datasets from the same sample enables TSS identification at single-nucleotide resolution from single-cell to standard levels of total RNA. Accordingly, we have improved current *Arabidopsis* TSS annotations using this technique. NanoPARE's low-input RNA requirements and simplicity of the protocol should enable TSS annotation improvements in other eukaryotic species, as well as in rare tissues and individual cell-types.
The identification of sRNA directed cleavage targets is essential to understand the molecular basis of sRNA functions during cellular differentiation, physiology and defense. Conventional PARE/degradome methods have been key technologies for characterizing the molecular basis of sRNA-mediated regulation, but require high amounts of input RNA typically only obtainable from bulk samples. NanoPARE allows identification of sRNA-mediated cleavage products from at least 10,000-fold less input RNA compared to these conventional methods and thus can be applied to specific tissue types. As a case study, we utilized nanoPARE to quantify miRNA-mediated cleavage events across five different flower tissues. In addition to detecting the previously reported tissue-enriched activities of three miRNA families, nanoPARE also identified several novel tissue-enriched miRNA-guided cleavage events indicating that it can be used to profile differential sRNA activities across tissue-types.
Moreover, nanoPARE enables 5′ end RNA profiling from existing full-length Smart-seq2 libraries, which has become a commonly used single-cell sequencing method (Ziegenhain *et al.* 2017). Such re-sampling at the level of cDNA rather than tissue is unique and not possible with technologies such as CAGE, STRT-Seq, Tn5-Prime and PARE. We, therefore, envision future applications of nanoPARE on both existing and new datasets for dissecting cell-type-specific transcriptional and post-transcriptional RNA regulatory mechanisms.

5.5 Methods

Plant material and growth

Wild-type and mutant (*xrn4-5*, *dcl234*) seeds were in Col-0 accession backgrounds and were grown in climate-controlled growth chambers with 20-22° C temperature and 16h light/8h dark cycle. The *dcl234* mutants were composed of *dcl2-1*, *dcl3*-1 and *dcl4*-2 alleles (Henderson *et al.* 2006) and *xrn4-5* mutants were as previously described (Souret *et al.* 2004).

RNA extraction

Total RNA was extracted from stage 12 floral buds using TRIzol (Life Tech). Stage 13 flowers were collected in 1 ml of 500 μM DTSSP (3,3′-dithiobis-(sulfosuccinimidyl propionate)) (ThermoFisher Cat #21578) with 1× PBS, pH 7.4, vacuum-infiltrated for 5 minutes and incubated for 10 minutes. Individual floral tissues were dissected under a binocular microscope on a silanized slide and snap-frozen in liquid nitrogen. Tissues were then homogenized using Mixer Mill MM 400 (Retsch), and the resulting pellets were resuspended in 300 μl TRIzol (Life Tech). Total RNA was extracted using Direct-zol kit (ZymoResearch) according to the manufacturer's instructions. RNA integrity was assessed by a Fragment Analyzer (AATI) using the standard RNA sensitivity kit (DNF-471).

NanoPARE library preparation

A detailed protocol can be found in Supplemental Methods. In brief, cDNA library preparation from 5 ng or less total RNA was carried out according to the original Smart-seq2 method (Picelli *et al.* 2013). cDNA was tagmented using the Illumina Nextera DNA library preparation kit, purified using the Zymo 5× DNA Clean and Concentrator kit (ZymoResearch) and eluted with nuclease-free water. For final enrichment PCR, the purified reaction was split and amplified either with Tn5.1/TSO enrichment oligonucleotide or Tn5.2/TSO enrichment oligonucleotide primer sets (Supplemental Table S1). PCR reaction products with Tn5.1/TSO enrichment oligonucleotide and Tn5.2/TSO enrichment oligonucleotide primer sets were pooled and purified using AMPureXP DNA beads. For *in vitro* biochemical degradation of 5′ monophosphate-containing RNA, 100 ng of total RNA was treated with XRN1 exoribonuclease (NEB) for 60 minutes at 37 °C in a 20 μl reaction volume containing 1× NEB Buffer 3 and 1U of XRN1. The equivalent of 5 ng total RNA (1 μl of XRN1-treated reaction) was used for Smart-seq2 cDNA synthesis without additional purification.

Next-generation sequencing

To control for library quality, final nanoPARE libraries were checked on an Agilent DNA HS Bioanalyzer Chip. Libraries with size-ranges between 150 and 800 bp were diluted and sequenced to 10-15 million single-end 50 bp reads per sample using a custom sequencing primer (TSO_Seq) and a custom P5/P7 index primer mix on an Illumina HiSeq 2500 Illumina instrument (Supplemental Table S1; Supplemental Data S1).

Classification of RNA 5′ ends

The nanoPARE analysis pipeline was written to identify capped and noncapped 5' end features in the genome using paired nanoPARE (5P) reads Smart-seq2 (BODY) reads. The analysis is divided into four major steps:

(1) Mapping of 5P and BODY reads to the genome (EndMap)

(2) 5P end feature identification (EndGraph)

(3) Classification of capped and noncapped 5P features (EndClass)

(4) Transcript-level output of read noncapped reads (EndMask)

EndMap

FASTQ files were mapped to the *Arabidopsis thaliana* TAIR10 genome (Lamesch *et al.* 2012). EndMap first trims the appropriate adapter sequences using Cutadapt (Marcel 2011). To prevent reads with low sequence complexity from mapping to the genome, the I-complexity (Becher and Heiber 2012) of each FASTQ read was calculated and reads with a per-nucleotide I-complexity <0.15 were removed. The remaining reads were then aligned to the genome with STAR (Dobin *et al.* 2012). Mapping behavior differs slightly between BODY and 5P libraries. All reads were mapped using the STAR settings:

--alignIntronMax 10000; --alignMatesGapMax 11000;
--alignSJDBoverhangMin 1; --alignSJoverhangMin 10
--outFilterMismatchNmax 2; --outFilterMismatchNoverLmax .05;
--outFilterMultimapNmax 100; --outSAMprimaryFlag
AllBestScore; --outSAMtype BAM Unsorted

BODY reads were mapped with the additional settings:

--alignEndsType EndToEnd

5P reads were mapped with the additional settings:

--alignEndsType Local; --outFilterMatchNminOverLread 0.9

After alignment to the genome, a bias correction algorithm was applied to the aligned BAM file to adjust for sequence-specific biases in the BODY and 5P libraries. The bias correction method defined by Wang *et al.* (Wang *et al.* 2017) was used, with two modifications to make it suitable for RNA rather than DNA data: (1) only reads within exons of annotated genes were used to calculate the k-mer frequency matrix, and (2) the read depth for all positions with >1 read was set to 1 because RNA seq reads are not expected to have even coverage at all genomic locations. After bias correction, reads that mapped to more than one genomic location were assigned via a "rich-get-richer" algorithm similar to that employed by the software MuMRescue and MuMRescueLite (Faulkner *et al.* 2008; Hashimoto *et al.* 2009). First, the coverage depth of uniquely mapping reads is calculated for each position in the genome. Multimappers are then binned by their mapping multiplicity (i.e. a read that maps to 10 locations in the genome has a multiplicity of 10). Beginning with a multiplicity of 2, all reads in that bin are sorted from lowest possible genomic position to highest, and each read is

assigned in a multistep process: if ≥1 mapping position contains ≥1 existing read, the read is considered "unambiguous," and is assigned proportionally to its mapping locations using the formula $P_i = \frac{C_i}{\sum_{j=1}^{n} C_j}$, where P_i is the proportion of reads assigned to mapping location *i*, C_i is the total existing read coverage assigned to the genomic positions that comprise location *i,* and *n* is the number of mapping locations for the read. If the existing read coverage at all locations is 0, that read is not yet assigned. The process is repeated until no more unambiguous reads can be identified, then all remaining reads are assigned with equal weighting, or $P_i = 1/n$. This is repeated for multiplicities of 3–100. A bedGraph file of 5′ end counts is written for both strands of the genome. For 5P libraries, all nucleotides soft clipped from the 5′ end of reads are stored as upstream untemplated nucleotides (uuNs).

EndGraph

Discrete 5P features were identified genome-wide via subtractive kernel density estimations. bedGraph files output from EndMap corresponding to a sample's 5P and BODY libraries were evaluated together. First, strand invasion artifacts (Tang *et al.* 2012) were masked based on complementarity to the last 4 bases of the template-switching oligo, allowing up to one mismatch. Then, a scaling factor (S) was estimated to normalize the read depth of the 5P library against the BODY library using the formula:

$$S = \frac{2F*10^6}{\sum_{i=1}^{n}(TPM_i*L_i)} * \frac{R_B}{R_E},$$

where *n* is the total number of transcripts, TPM_i is the abundance of a transcript in transcripts per million, L_i is the length of a transcript in nucleotides, *F* is the mean fragment length of the BODY library, R_B is the total number of mapped BODY reads, and R_E is the total number of mapped 5P reads. Then, a Laplace kernel with a bandwidth of 15 nt was fit over the set of values (ER ∗ S) −BR, where ER is the set of 5P end read counts, and BR is the set of body read counts. Regions of continuous positive density were extracted and written as discrete features to a bed file.

EndClass

If a 5P experiment was designed with multiple replicates, EndClass merged all 5P features that could be reproducibly identified in ≥2 replicates. Then, the presence of a m7G cap was predicted for each replicable feature by calculating the proportion of reads containing upstream untemplated guanosine (uuG). A feature was considered capped if ≥10% of all reads from a sample type that map within the feature contained uuG; otherwise, the feature was considered noncapped.

EndMask

EndMask prepared a bedGraph file of 5P read positions relative to the start site of the dominant isoform of each gene in the reference annotation. Dominant isoforms were defined as the transcript isoform containing the most mapped reads. For nanoPARE libraries, this transcript-level bedGraph was generated with a cap-masked input in which 5P reads contained within replicable capped 5P features were discarded.

Detection of sRNA-mediated cleavage sites with EndCut

Sequences from miRNAs and tasiRNAs annotated in TAIR10 or miRBase21 (Lamesch *et al.* 2012; Kozomara and Griffiths-Jones 2013) were selected (i.e. anno.mir.tas.fa) and randomized one thousand times each by the Python script sRNA_shuffler.py to produce anno.mir.tas.i.fa files; where i is an integer between 0 and 999). For annotated miRNA, tasiRNA and the corresponding 1,000 randomized variants for each miRNA/tasiRNA, GSTAr.pl (https://github.com/MikeAxtell/GSTAr) was used to predict target sites in transcript models annotated as protein-coding genes, transposable element genes or other RNAs (i.e. TAIR10_pc_teg_other_transcripts.fasta). Target sites were determined based on the level of complementarity between sRNAs and transcripts computed using previously developed criteria based on the frequency and position of the miRNA-target duplex mismatches (i.e. Allen scores) (Allen *et al.* 2005). As described above, nanoPARE data was processed by EndMask to exclude capped regions of transcripts from further analyses. Publicly available PARE datasets were downloaded from the Sequence Read Archive (NCBI) (Supplementary Data S1), but alignments overlapping capped features were not excluded from downstream analyses. Predicted target sites and EndGraph output were used by EndCut_step1.sh to quantify the number of reads at predicted target sites and in adjacent 20 nt or 50 nt regions on the sense strand of the same transcript. Adjacent sites within one nucleotide of predicted cleavage sites were not considered in order to not penalize sites for sRNA isoforms with slightly offset target recognition sites.

The local enrichment of nanoPARE read 5′ ends at predicted cleavage sites relative to surrounding transcribed regions, or fold-changes, was calculated by dividing the numbers of nanoPARE read 5′ ends at predicted cleavage sites +1 by the maximum numbers of reads in adjacent transcript regions + 1. Allen scores were also assigned to each predicted cleavage site detected. For each randomized sRNA control set, EndCut_step2.R computed empirical cumulative distribution functions of fold-changes (ECDFFC) and Allen scores (ECDFAS). These were then used as null models to test whether the observed cleavage site fold-changes were not equal to or lesser than ECDFFC, as well as if the observed site Allen scores were not equal to or greater than ECDFAS. Final P-values were computed for each site by combining these two P-values using Fisher's combined probability test, and then adjusted for multiple testing using the Benjamini-Hochberg method. For our analyses, we defined significant cleavage sites that had adjusted P-values < 0.05, fold-changes > 1.0 and that were also represented by at least one read per ten million transcriptome-mapping reads.

Data Access

All sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/ under accession number GSE112869. All software code is publicly available at GitHub (https://github.com/Gregor-Mendel-Institute/NanoPARE) and is available as Supplemental Code S1.

5.6 Acknowledgements

We thank the VBCF NGS Unit for sequencing nanoPARE and Smart-seq2 libraries, the VBCF Plant Sciences Facility for plant growth chamber access, and Andreas Sommer, Michael Axtell and Nate Johnson for providing valuable NGS advice. This work was supported by funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant 637888 to M.D.N.) and the DK Graduate Program in RNA Biology (DK-RNA) sponsored by the Austria Science Fund (FWF, DK W 1207-B09).

5.7 Author Contributions

M.A.S., M.J.K. and M.D.N. conceived the ideas. M.J.K. and A.P. performed the experiments. M.A.S., F.H. and M.D.N. wrote software and M.A.S. and M.D.N. performed the analysis. M.A.S., M.J.K. and M.D.N. interpreted the results and wrote the manuscript.

5.8 Disclosure Declaration

M.A.S., M.J.K. and M.D.N. were employed by the Gregor Mendel Institute, a non-profit research organization that has filed a patent application on the nanoPARE technology described in this work.

5.9 References

Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. 2008. Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Current Biology* **18**: 758–762.

Adiconis X, Haber AL, Simmons SK, Levy Moonshine A, Ji Z, Busby MA, Shi X, Jacques J, Lancaster MA, Pan JQ, *et al.* 2018. Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nat Meth* **15**: 505–511.

Allen E, Xie Z, Gustafson AM, Carrington JC. 2005. microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. *Cell* **121**: 207–221.

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, *et al.* 2014. An atlas of active enhancers across human cell types and tissues. *Nature* **507**: 455–461.

Arguel M-J, LeBrigand K, Paquet A, Ruiz García S, Zaragosi L-E, Barbry P, Waldmann R. 2017. A cost effective 5 ' selective single cell transcriptome profiling approach with improved UMI design. *Nucleic Acids Research* **45**: e48–e48.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.

Becher V, Heiber P. A. 2012. A linearly computable measure of string complexity. *Theoretical Computer Science* **438**: 62–73.

Borges F, Martienssen RA. 2015. The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* **16**: 727–741.

Bowman JL, Drews GN, Meyerowitz EM. 1991a. Expression of the *Arabidopsis* floral homeotic gene AGAMOUS is restricted to specific cell types late in flower development. *The Plant Cell* **3**: 749–58.

Bowman JL, Smyth DR, Meyerowitz EM. 1991b. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**: 1–20.

Brennecke P, Anders S, Kim JK, Kołodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC, *et al.* 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Meth* **10**: 1093–1095.

Carlsbecker A, Lee J-Y, Roberts CJ, Dettmer J, Lehesranta S, Zhou J, Lindgren O, Moreno-Risueno MA, Vatén A, Thitamadee S, *et al.* 2010. Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature* **465**: 316–321.

Cheng Z, Otto GM, Powers EN, Keskin A, Mertins P, Carr SA, Jovanovic M, Brar GA. 2018. Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis. *Cell* **172**: 910–923.e16.

Cloonan N, Forrest ARR, Kolle G, Gardiner BBA, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, *et al.* 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Meth* **5**: 613–619.

Cocquet J, Chong A, Zhang G, Veitia RA. 2006. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **88**: 127–131.

Coen ES, Meyerowitz EM. 1991. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**: 31–37.

Cole C, Byrne A, Beaudin AE, Forsberg EC, Vollmers C. 2018. Tn5Prime, a Tn5 based 5' capture method for single cell RNA-seq. *Nucleic Acids Research* ***12****:* 1–12.

Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA. 2014. miRNAs trigger widespread epigenetically activated siRNAs from transposons in *Arabidopsis*. *Nature* **508**: 411–415.

Cumbie JS, Ivanchenko MG, Megraw M. 2015. NanoCAGE-XL and CapFilter: an approach to genome wide identification of high confidence transcription start sites. *BMC Genomics* 1–13.

de Rie D, Abugessaisa I, Alam T, Arner E, Arner P, Ashoor H, Åström G, Babina M, Bertin N, Burroughs AM, *et al.* 2017. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat Biotechnol* **35**: 872–878.

Dieci G, Preti M, Montanini B. 2009. Eukaryotic snoRNAs: A paradigm for gene expression flexibility. *Genomics* **94**: 83–88.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2012. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Faulkner GJ, Forrest ARR, Chalk AM, Schroder K, Hayashizaki Y, Carninci P, Hume DA, Grimmond SM. 2008. A rescue strategy for multimapping short sequence tags refines surveys of transcriptional activity by CAGE. *Genomics* **91**: 281–288.

Filipowicz W, Pogacić V. 2002. Biogenesis of small nucleolar ribonucleoproteins. *Curr Opin Cell Biol* **14**: 319–327.

Gasciolli V, Mallory AC, Bartel DP, Vaucheret H. 2005. Partially Redundant Functions of *Arabidopsis* DICER-like Enzymes and a Role for DCL4 in Producing trans-Acting siRNAs. *Current Biology* **15**:1494–1500.

German MA, Pillay M, Jeong D-H, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, *et al.* 2008. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat Biotechnol* **26**: 941–946.

Goto K, Meyerowitz EM. 1994. Function and regulation of the *Arabidopsis* floral homeotic gene PISTILLATA. *Genes & Development* **8**: 1548–1560.

Granneman S, Petfalski E, Tollervey D. 2011. A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *The EMBO Journal* **30**: 4006–4019.

Gregory BD, O'Malley RC, Lister R, Urich MA, Tonti-Filippini J, Chen H, Millar AH, Ecker JR. 2008. A Link between RNA Metabolism and Silencing Affecting *Arabidopsis* Development. *Developmental Cell* **14**: 854–866.

Grohmann K, Amairic F, Crews S, Attardi G. 1978. Failure to detect “cap” structures in mitochondrial DNA-coded poly(A)-containing RNA from HeLa cells. *Nucleic Acids Research* **5**: 637–651.

Haberle V, Li N, Hadzhiev Y, Plessy C, Previti C, Nepal C, Gehrig J, Dong X, Akalin A, Suzuki AM, *et al.* 2014. Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* **507**: 381–385.

Harbers M, Kato S, de Hoon M, Hayashizaki Y, Carninci P, Plessy C. 2013. Comparison of RNA- or LNA-hybrid oligonucleotides in template-switching reactions for high-speed sequencing library preparation. *BMC Genomics* **14**: 665–6.

Hashimoto T, de Hoon M.J.L, Grimmond SM, Daub CO, Hayashizaki Y, Faulkner GJ. 2009. Probabilistic resolution of multi-mapping reads in massively parallel sequencing data using MuMRescueLite. *Bioinformatics* **25**: 2613–2614.

Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE. 2006. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* **38**: 721–725.

Henras AK, Plisson-Chastang C, O'Donohue M-F, Chakraborty A, Gleizes P-E. 2015. An overview of pre-ribosomal RNA processing in eukaryotes. WIREs *RNA* **6**: 225–242.

Hon C-C, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJL, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J, *et al*. 2017. An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature* **543**: 199–204.

Hopper AK, Pai DA, Engelke DR. 2010. Cellular dynamics of tRNAs and their genes. *FEBS Letters* **584**: 310–317.

Hou CY, Lee WC, Chou HC, Chen AP, Chou SJ, Chen H-M. 2016. Global Analysis of Truncated RNA Ends Reveals New Insights into Ribosome Stalling in Plants. *The Plant Cell* **28**: 2398–2416.

Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, Kasschau KD, Carrington JC. 2007. Genome-Wide Analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 Pathway in *Arabidopsis* Reveals Dependency on miRNA- and tasiRNA-Directed Targeting. *The Plant Cell* **19**: 926–942.

Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, Linnarsson S. 2011. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research* **21**: 1160–1167.

Jack T, Brockman LL, Meyerowitz EM. 1992. The homeotic gene APETALA3 of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* **68**: 683–697.

Jofuku KD, Boer den BG, Van Montagu M, Okamuro JK. 1994. Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *The Plant Cell* **6**: 1211–1225.

Jones-Rhoades MW, Bartel DP. 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular Cell* **14**: 787–799

Karlsson K, Lönnerberg P, Linnarsson S. 2017. Alternative TSSs are co-regulated in single cells in the mouse brain. *Mol Syst Biol* **13**: 930–10.

Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, Krizan KA, Carrington JC. 2003. P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. *Developmental Cell* **4**: 205–217.

Kastenmayer JP, Green PJ. 2000. Novel features of the XRN-family in *Arabidopsis*: evidence that AtXRN4, one of several orthologs of nuclear Xrn2p/Rat1p, functions in the cytoplasm. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 13985–13990.

Kidner CA, Martienssen RA. 2004. Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. *Nature* **428**: 81–84.

Knauer S, Holt AL, Rubio-Somoza I, Tucker EJ, Hinze A, Pisch M, Javelle M, Timmermans MC, Tucker MR, Laux T. 2013. A protodermal miR394 signal defines a region of stem cell competence in the *Arabidopsis* shoot meristem. *Dev Cell* **24**: 125–132.

Kozomara A, Griffiths-Jones S. 2013. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* **42**: D68–D73.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, *et al.* 2012. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* **40**: D1202–10.

Legen J, Kemp S, Krause K, Profanter B, Herrmann RG, Maier RM. 2002. Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. *Plant J* **31**: 171–188.

Liang G, He H, Li Y, Wang F, Yu D. 2014. Molecular mechanism of microRNA396 mediating pistil development in *Arabidopsis*. *PLANT PHYSIOLOGY* **164**: 249–258.

Llave C, Xie Z, Kasschau KD, Carrington JC. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**: 2053–2056.

Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF. 1992. Molecular characterization of the *Arabidopsis* floral homeotic gene APETALA1. *Nature* **360**: 273–277.

Marcel M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJournal* **17**: 10–12.

Miyashima S, Honda M, Hashimoto K, Tatematsu K, Hashimoto T, Sato-Nara K, Okada K, Nakajima K. 2013. A Comprehensive Expression Analysis of the *Arabidopsis* MICRORNA165/6 Gene Family during Embryogenesis Reveals a Conserved Role in Meristem Specification and a Non-Cell-Autonomous Function. *Plant and Cell Physiology* **54**: 375–384.

Monde RA, Schuster G, Stern DB. 2000. Processing and degradation of chloroplast mRNA. *Biochimie* **82**: 573–582.

Morton T, Petricka J, Corcoran DL, Li S, Winter CM, Carda A, Benfey PN, Ohler U, Megraw M. 2014. Paired-end analysis of transcription start sites in *Arabidopsis* reveals plant-specific promoter signatures. *The Plant Cell* **26**: 2746–2760.

Nagarajan VK, Jones CI, Newbury SF, Green PJ. 2013. XRN 5′→3′ exoribonucleases: Structure, mechanisms and functions. *BBA - Gene Regulatory Mechanisms* **1829**: 590–603.

Nodine MD, Bartel DP. 2010. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes & Development* **24**: 2678–2692.

Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Meth* **10**: 1096–1098.

Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, *et al.* 2012. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* **30**: 777–782.

Rubio-Somoza I, Weigel D. 2013. Coordination of flower maturation by a regulatory circuit of three microRNAs. ed. L.-J. Qu. *PLoS Genetics* **9**: e1003374–10.

Souret FF, Kastenmayer JP, Green PJ. 2004. AtXRN4 Degrades mRNA in *Arabidopsis* and Its Substrates Include Selected miRNA Targets. *Molecular Cell* **15**: 173–183.

Tang DTP, Plessy C, Salimullah M, Suzuki AM, Calligaris R, Gustincich S, Carninci P. 2012. Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. *Nucleic Acids Research* **41**: e44–e44.

Ushijima T, Hanada K, Gotoh E, Yamori W, Kodama Y, Tanaka H, Kusano M, Fukushima A, Tokizawa M, Yamamoto YY, *et al.* 2017. Light Controls Protein Localization through Phytochrome-Mediated Alternative Promoter Selection. *Cell* **171**: 1316–1325.

Wang JR, Quach B, Furey TS. 2017. Correcting nucleotide-specific biases in high-throughput sequencing data. *BMC Bioinformatics* **18**: 357–10.

Wang JW, Czech B, Weigel D. 2009. miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* **138**: 738–749.

Wang MJ, Davis NW, Gegenheimer P. 1988. Novel mechanisms for maturation of chloroplast transfer RNA precursors. *The EMBO journal* **7**: 1567–1574.

Wang Y, Wang X, Deng W, Fan X, Liu T-T, He G, Chen R, Terzaghi W, Zhu D, Deng XW. 2014. Genomic Features and Regulatory Roles of Intermediate-Sized Non-Coding RNAs in *Arabidopsis*. *Molecular Plant* **7**: 514–527.

Williams L, Grigg SP, Xie M, Christensen S, Fletcher JC. 2005. Regulation of *Arabidopsis* shoot apical meristem and lateral organ formation by microRNA miR166g and its AtHD-ZIP target genes. *Development* **132**: 3657–3668.

Willmann MR, Berkowitz ND, Gregory BD. 2014. Improved genome-wide mapping of uncapped and cleaved transcripts in eukaryotes GMUCT 2.0. *Methods* **67**: 64–73.

Wollmann H, Mica E, Todesco M, Long JA, Weigel D. 2010. On reconciling the interactions between APETALA2, miR172 and AGAMOUS with the ABC model of flower development. *Development* **137**: 3633–3642.

Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS. 2009. The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* **138**: 750–759.

Wu MF, Tian Q, Reed JW. 2006. *Arabidopsis* microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development* **133**: 4211–4218.

Yoshikawa M, Peragine A, Park MY, Poethig RS. 2005. A pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*. *Genes & Development* **19**: 2164–2175.

Yu X, Willmann MR, Anderson SJ, Gregory BD. 2016. Genome-Wide Mapping of Uncapped and Cleaved Transcripts Reveals a Role for the Nuclear mRNA Cap-Binding Complex in Cotranslational RNA Decay in *Arabidopsis*. *The Plant Cell* **28**: 2385–2397.

Zhou Y, Honda M, Zhu H, Zhang Z, Guo X, Li T, Li Z, Peng X, Nakajima K, Duan L, *et al.* 2015. Spatiotemporal Sequestration of miR165/166 by *Arabidopsis* Argonaute10 Promotes Shoot Apical Meristem Maintenance. *Cell Reports* **10**: 1819–1827.

Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W. 2017. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular Cell* **65**: 631–643.

5.10 Supplemental Material

Table of Contents

Supplemental Data Files S1-S8 are provided online as Excel spreadsheets.

Data S1. Summary of high-throughput sequencing generated or reanalyzed for this study

Data S2. Floral bud 5′ end features identified by EndGraph

Data S3. EndGraph features associated with noncoding RNAs

Data S4. EndCut analysis of microRNA cleavage sites in nanoPARE data

Data S5. EndCut analysis of tasiRNA cleavage sites in nanoPARE data

Data S6. Published microRNA target interactions validated by 5′ RACE

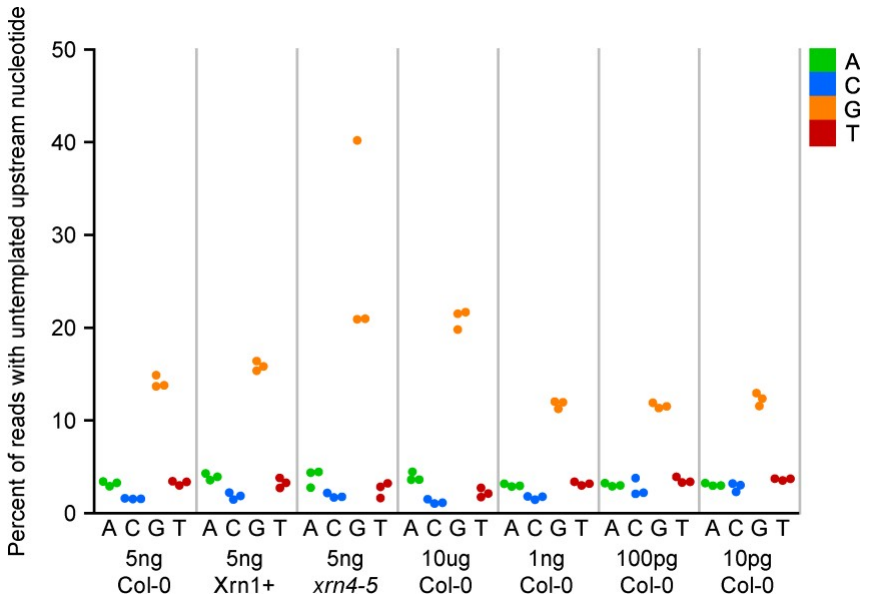
Data S7. EndCut analysis of microRNA cleavage sites in public degradome data

Data S8. EndCut analysis of tasiRNA cleavage sites in public degradome data

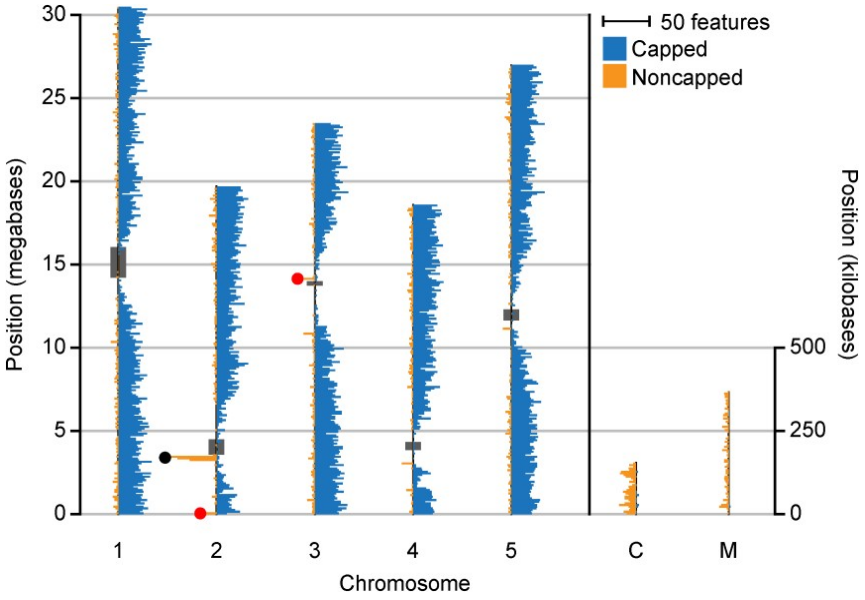
Supplemental Code S1. Scripts used to generate figures and datasets in this study.

Name	Sequence	step used	additional information
Anchored oligo-dT	AAGCAGTGGTATCAACGCAGAGTACT30VN	Reverse transcription	' N' is any base and 'V' is either 'A', 'C' or 'G'
TSO	AAGCAGTGGTATCAACGCAGAGTACrGrG+G	Reverse transcription	'+' is LNA base, 'r' is ribonucleotide, HPLC purified
TSO 5' Biotin	/5Biosg/AAGCAGTGGTATCAACGCAGAGTACrGrG+G	Reverse transcription	5'-biotinylated, '+' is LNA base, 'r' is ribonucleotide, HPLC purified, preferred to avoid concatemerization
ISPCR	AAGCAGTGGTATCAACGCAGAGT	PCR-preamplification	-
P5_TSO_N501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCCTAGCAAGCAGTGGTATCAACGCAGAGTACGGG	NanoPARE 5'end enrichment	Illumina Nextera N501 index, PAGE purified
P5_TSO_N502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATCTAGCAAGCAGTGGTATCAACGCAGAGTACGGG	NanoPARE 5'end enrichment	Illumina Nextera N502 index, PAGE purified
P5_TSO_N503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTCTAGCAAGCAGTGGTATCAACGCAGAGTACGGG	NanoPARE 5'end enrichment	Illumina Nextera N503 index, PAGE purified
P5_TSO_N504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGACTAGCAAGCAGTGGTATCAACGCAGAGTACGGG	NanoPARE 5'end enrichment	Illumina Nextera N504 index, PAGE purified
P5_TSO_N505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGCTAGCAAGCAGTGGTATCAACGCAGAGTACGGG	NanoPARE 5'end enrichment	Illumina Nextera N505 index, PAGE purified
P5_TSO_N506	AATGATACGGCGACCACCGAGATCTACACACTGCATACTAGCAAGCAGTGGTATCAACGCAGAGTACGGG	NanoPARE 5'end enrichment	Illumina Nextera N506 index, PAGE purified
P5_TSO_N507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTACTAGCAAGCAGTGGTATCAACGCAGAGTACGGG	NanoPARE 5'end enrichment	Illumina Nextera N507 index, PAGE purified
P5_TSO_N508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTCTAGCAAGCAGTGGTATCAACGCAGAGTACGGG	NanoPARE 5'end enrichment	Illumina Nextera N508 index, PAGE purified
P5_TSO_S510	AATGATACGGCGACCACCGAGATCTACACCGTCTAATCTAGCAAGCAGTGGTATCAACGCAGAGTACGGG	NanoPARE 5'end enrichment	Illumina Nextera S510 index, PAGE purified
P7_Tn5.1_N701	CAAGCAGAAGACGGCATACGAGATTAAGGCGATCGTCGGCAGCGTC	NanoPARE 5'end enrichment	Illumina Nextera N701 index, PAGE purified
P7_Tn5.2_N701	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGG	NanoPARE 5'end enrichment	Illumina Nextera N701 index, PAGE purified
TSO_seq_read1	CTAGCAAGCAGTGGTATCAACGCAGAGTACGGG	sequencing primer for i5 index	Illumina custom sequencing primer

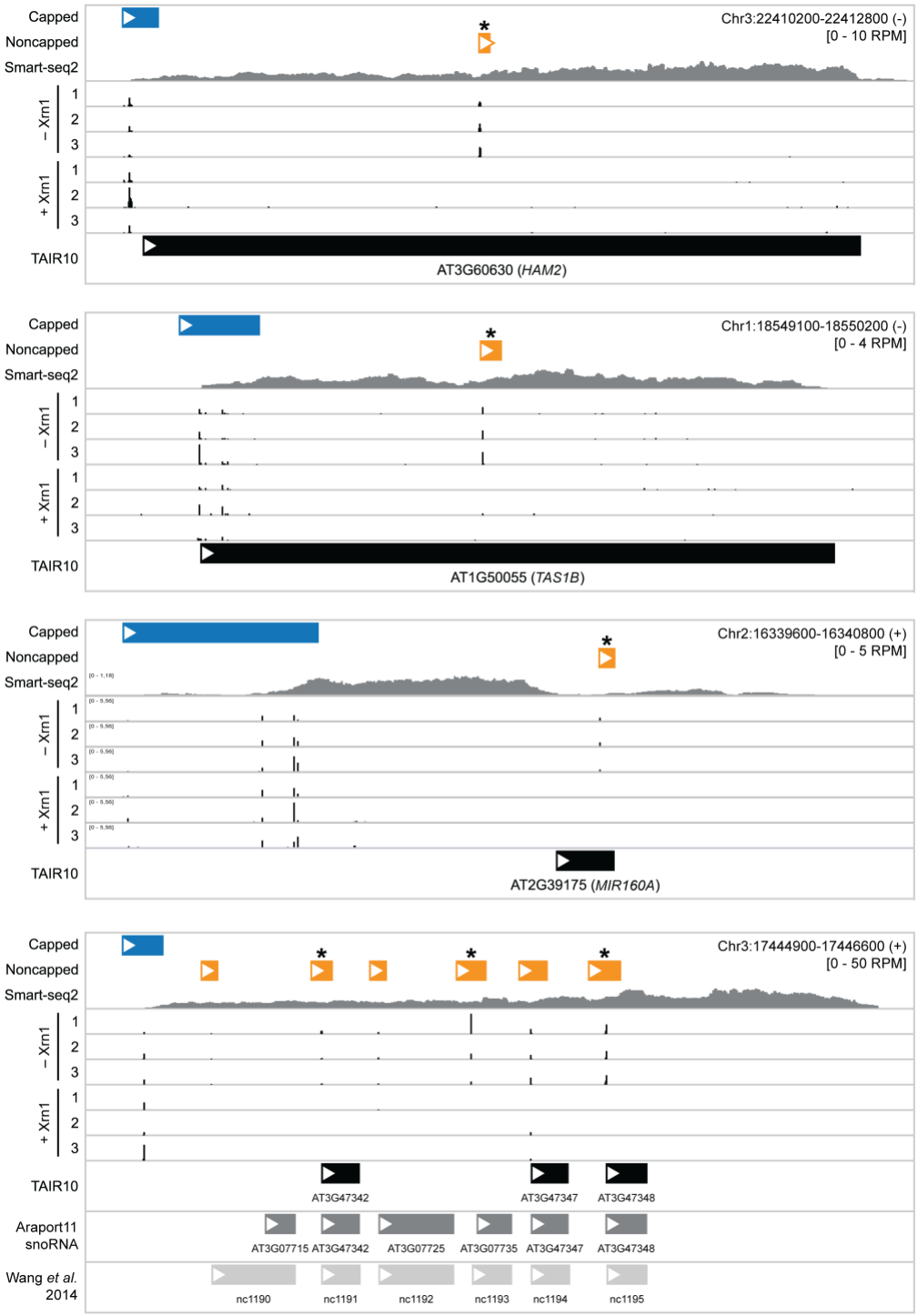
Table S1. Oligonucleotides used in this publication.



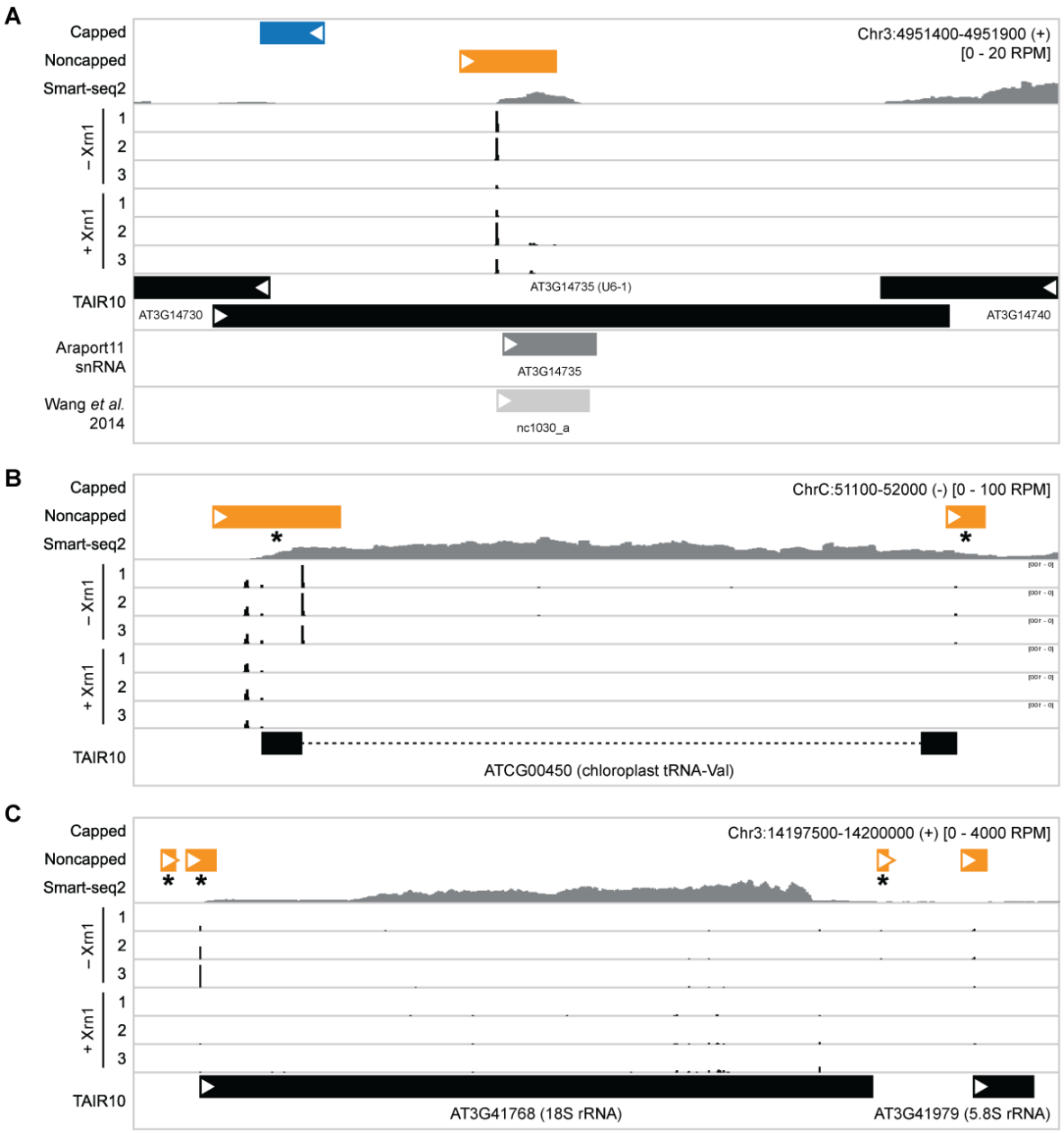
Supplemental Figure S1. Untemplated upstream nucleotide content. The proportion of genome-matching nanoPARE reads with untemplated upstream nucleotides (uuNs) in floral bud samples.



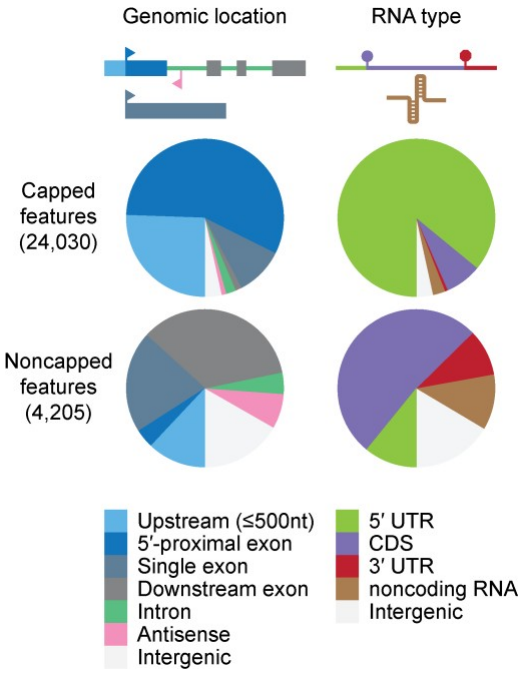
Supplemental Figure S2. Genomic distribution of capped and noncapped features. Barplots depicting the number of capped (blue) and noncapped (orange) features contained in bins across the genome. A bin size of 100 kilobases was used for nuclear chromosomes, and a bin size of 5 kilobases was used for mitochondrial and chloroplast genomes. Centromeres are marked by gray rectangles. Red circles are centered over the two 45S ribosomal RNA loci in the TAIR10 genome assembly on the short arms of chromosomes 2 and 3, and the black circle on chromosome 2 marks an insertion of the mitochondrial genome into the pericentromeric region.



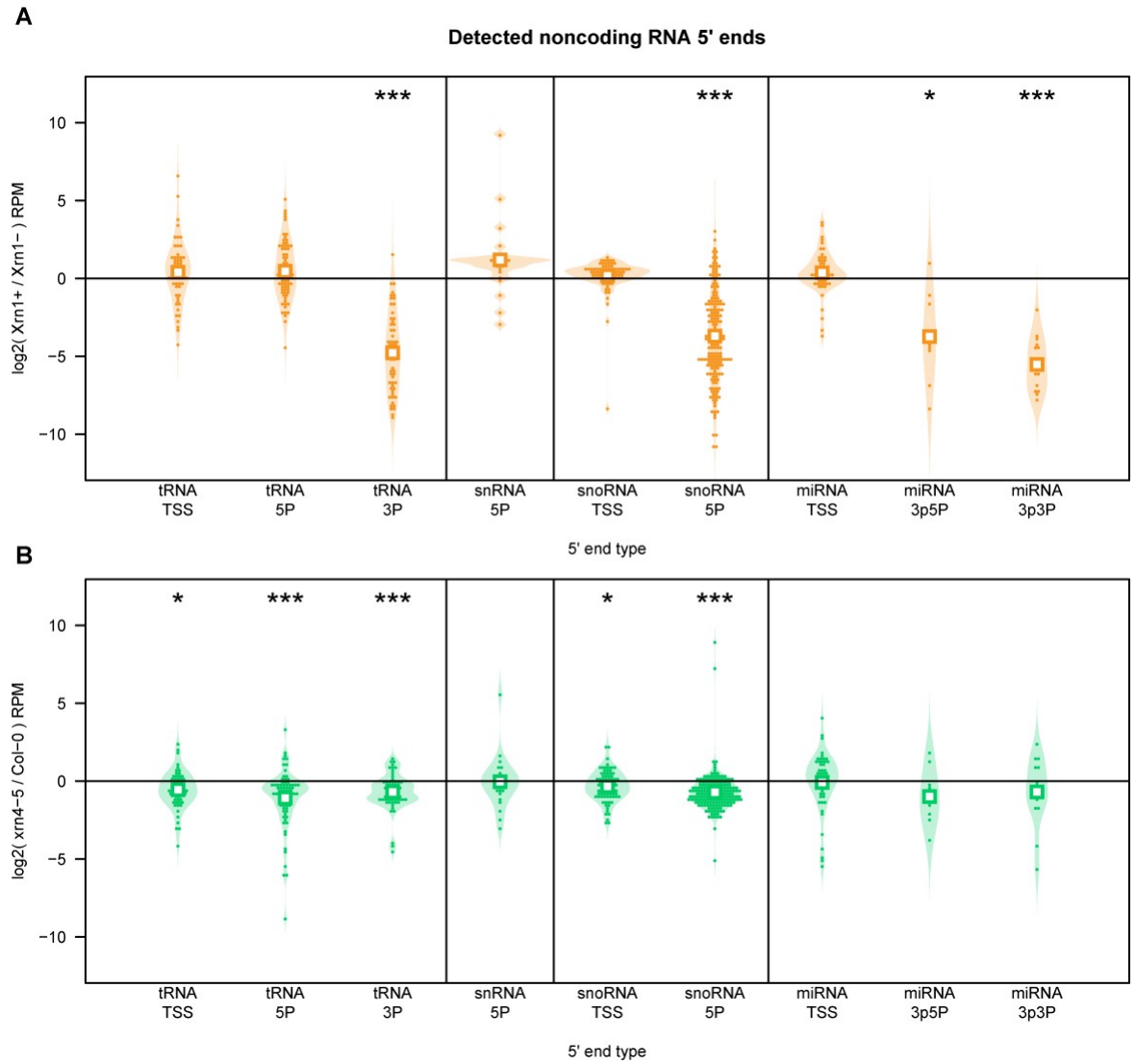
Supplemental Figure S3. IGV browser tracks with examples of RNA pol II transcripts. Blue boxes depict capped features, orange boxes depict noncapped features, and black boxes depict TAIR10 exons. Triangles show the feature orientation, and asterisk marks features that are significantly less abundant in Xrn1+ samples (*P*-values < 0.05, DEseq2). **(*A*)** Protein coding mRNA HAM2. **(*B*)** Long noncoding RNA TAS1B. **(*C*)** Primary microRNA MIR160A. **(*D*)** Unannotated primary snoRNA transcript containing six mature snoRNAs.



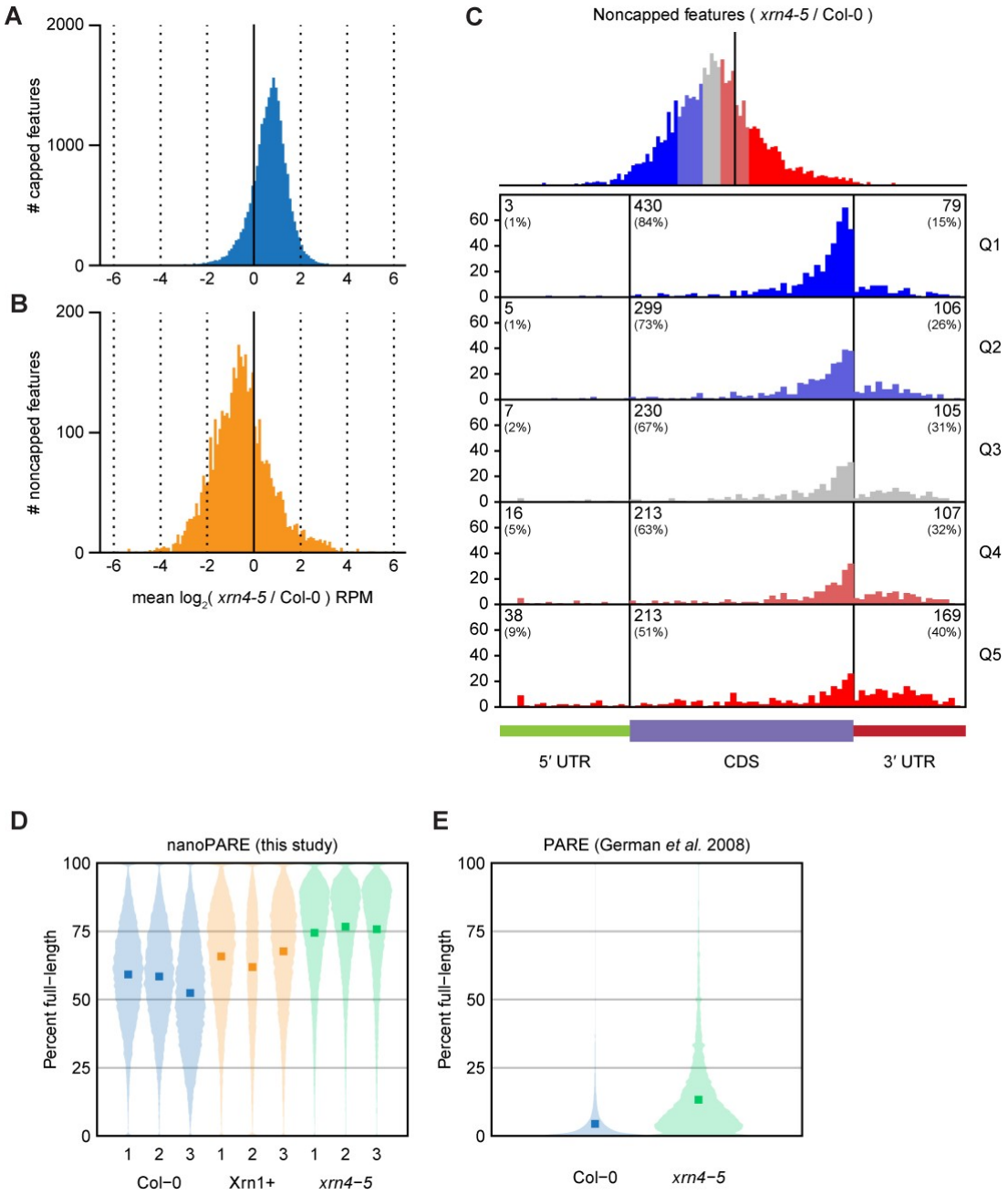
Supplemental Figure S4. IGV browser tracks with examples of non-pol II transcripts. Tracks are formatted as in Supplemental Figure 3. ***(A)*** Small nuclear RNA U6-1. ***(B)*** Chloroplast group-II intron-containing primary tRNA-Val. ***(C)*** 18S and 5.8S ribosomal RNA.



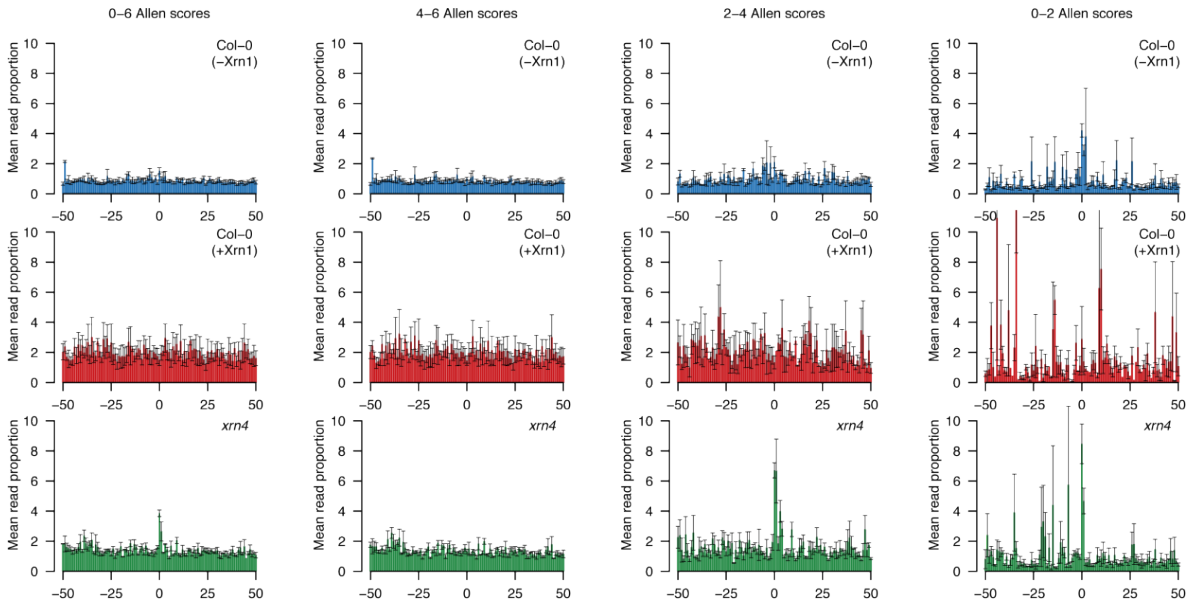
Supplemental Figure S5. Distribution of capped and noncapped features by gene type. Pie charts grouping all capped features (*top*) and noncapped features (*bottom*) by their location relative to TAIR10 gene annotations based on their genomic location (*left*) and the type of RNA (*right*).



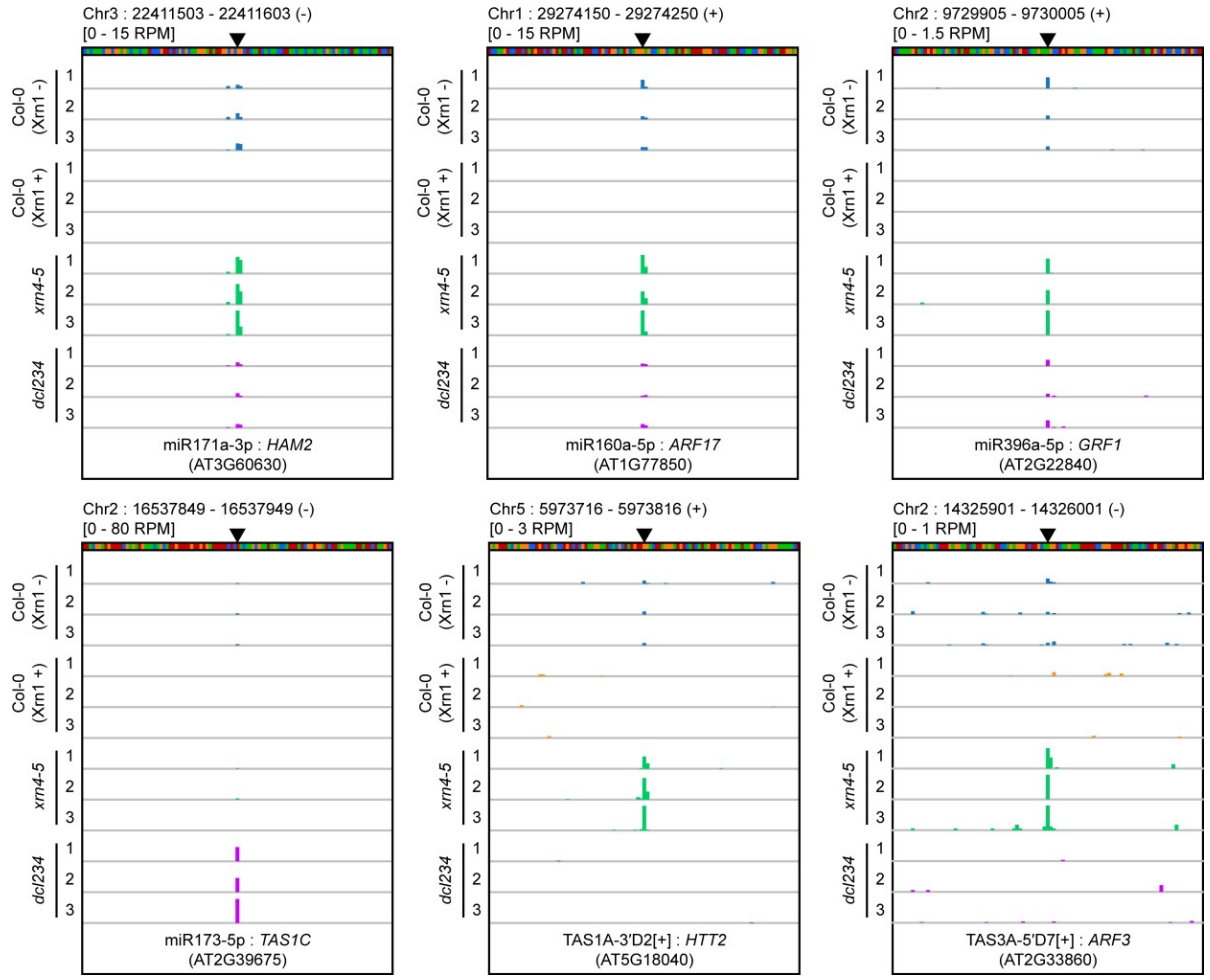
Supplemental Figure S6. Noncoding RNA 5′ ends detected by nanoPARE. Strip charts of all RNA 5′-ends identified by EndGraph from non-protein-coding transcripts showing fold change relative to untreated Col-0 nanoPARE reads. **(*A*)** Log_2 fold change after treatment with Xrn1 enzyme. **(*B*)** Log_2 fold change in *xrn4-5* mutant floral buds. * and *** indicate P-values < 0.05 and 0.001, respectively, based on two-tailed Student's t-test.



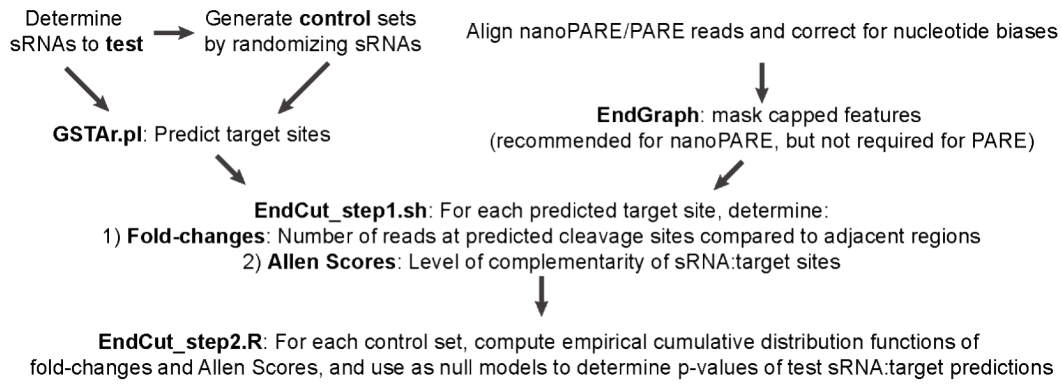
Supplemental Figure S7. Global changes to 5′ feature abundance in *xrn4-5* mutant floral buds. **(*A*)** Log_2 fold change of capped feature RPM. **(*B*)** Log2 fold change noncapped feature RPM. **(*C*)** mRNA metaplots as in Figure 2E for noncapped features grouped into quintiles based on change in abundance in *xrn4-5*. Violin plots showing the percent of nanoPARE reads contained within a gene's capped feature(s) for all genes where at least one capped feature could be identified. Squares mark the mean proportion of full-length reads across all genes. **(*E*)** Analysis of German *et al.* 2008 PARE data using the same criteria as in **D.**



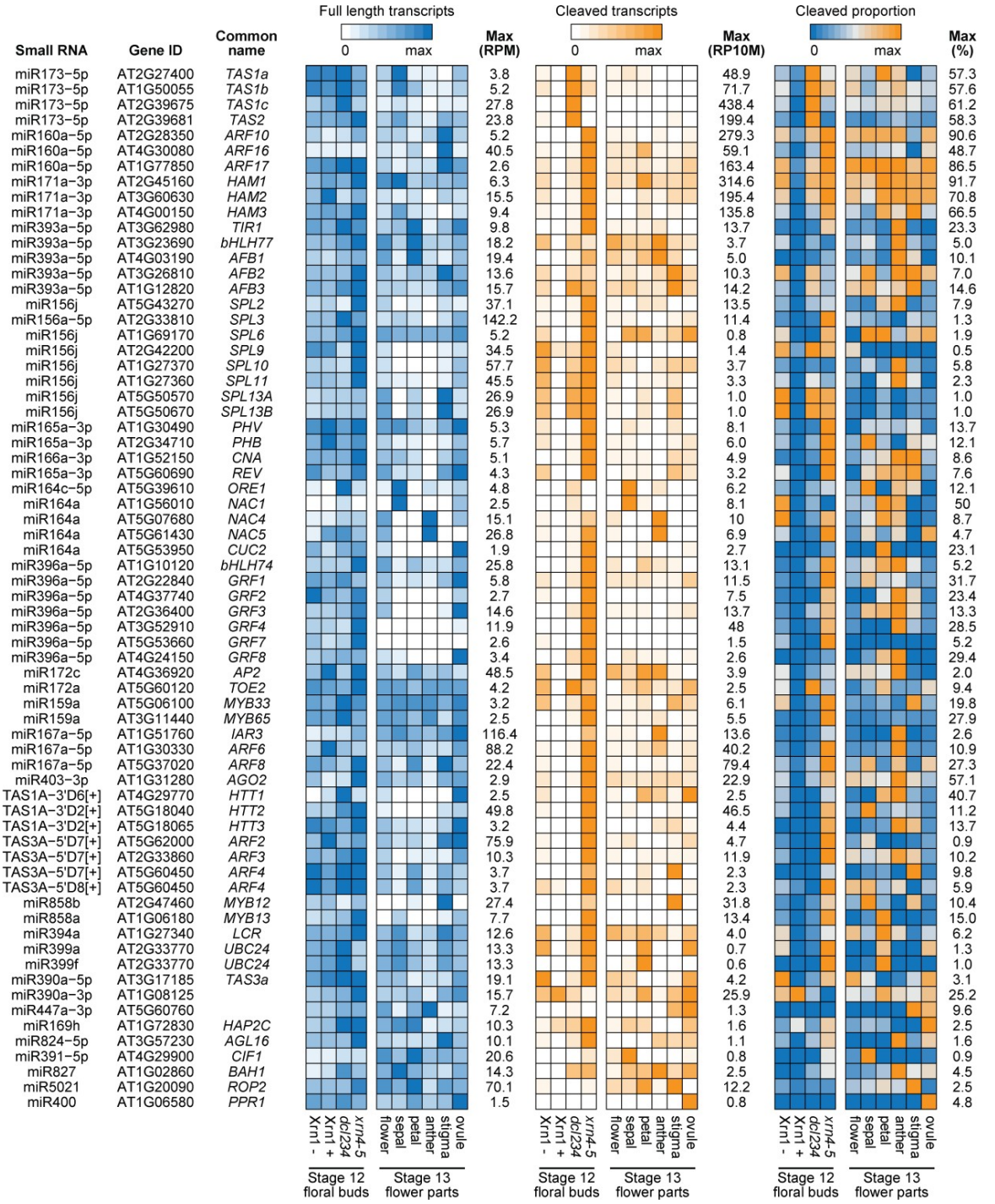
Supplemental Figure S8. Detection of predicted miRNA-mediated cleavage sites. The number of nanoPARE read 5′ ends mapping within 50 nt of all predicted miRNA cleavage sites are shown as a percentage of the total number of nanoPARE reads detected for each transcript. All predicted sites (0 – 6 Allen scores) and predicted sites with low (4 – 6 Allen scores), medium (2 – 4) or high (0 – 2 Allen scores) miRNA-target duplex complementarities are shown from *left* to *right*.



Supplemental Figure S9. IGV browser tracks with examples of small RNA cleavage sites. All tracks in each panel are normalized by the number of genome-matching reads per million (RPM) and were scaled by the values stated above the panels. Expected cleavage position of the listed small RNA species is marked by a black triangle. The nucleotide sequence is depicted by colored bars above each panel: green – A, blue – C, orange – G, red – T.



Supplemental Figure S10. Diagram of EndCut workflow used to detect sRNA-mediated cleavage sites. See Methods for details.



Small RNA	Gene ID	Common name	Max (RPM)	Max (RP10M)	Max (%)
miR173−5p	AT2G27400	*TAS1a*	3.8	48.9	57.3
miR173−5p	AT1G50055	*TAS1b*	5.2	71.7	57.6
miR173−5p	AT2G39675	*TAS1c*	27.8	438.4	61.2
miR173−5p	AT2G39681	*TAS2*	23.8	199.4	58.3
miR160a−5p	AT2G28350	*ARF10*	5.2	279.3	90.6
miR160a−5p	AT4G30080	*ARF16*	40.5	59.1	48.7
miR160a−5p	AT1G77850	*ARF17*	2.6	163.4	86.5
miR171a−3p	AT2G45160	*HAM1*	6.3	314.6	91.7
miR171a−3p	AT3G60630	*HAM2*	15.5	195.4	70.8
miR171a−3p	AT4G00150	*HAM3*	9.4	135.8	66.5
miR393a−5p	AT3G62980	*TIR1*	9.8	13.7	23.3
miR393a−5p	AT3G23690	*bHLH77*	18.2	3.7	5.0
miR393a−5p	AT4G03190	*AFB1*	19.4	5.0	10.1
miR393a−5p	AT3G26810	*AFB2*	13.6	10.3	7.0
miR393a−5p	AT1G12820	*AFB3*	15.7	14.2	14.6
miR156j	AT5G43270	*SPL2*	37.1	13.5	7.9
miR156a−5p	AT2G33810	*SPL3*	142.2	11.4	1.3
miR156j	AT1G69170	*SPL6*	5.2	0.8	1.9
miR156j	AT2G42200	*SPL9*	34.5	1.4	0.5
miR156j	AT1G27370	*SPL10*	57.7	3.7	5.8
miR156j	AT1G27360	*SPL11*	45.5	3.3	2.3
miR156j	AT5G50570	*SPL13A*	26.9	1.0	1.0
miR156j	AT5G50670	*SPL13B*	26.9	1.0	1.0
miR165a−3p	AT1G30490	*PHV*	5.3	8.1	13.7
miR165a−3p	AT2G34710	*PHB*	5.7	6.0	12.1
miR166a−3p	AT1G52150	*CNA*	5.1	4.9	8.6
miR165a−3p	AT5G60690	*REV*	4.3	3.2	7.6
miR164c−5p	AT5G39610	*ORE1*	4.8	6.2	12.1
miR164a	AT1G56010	*NAC1*	2.5	8.1	50
miR164a	AT5G07680	*NAC4*	15.1	10	8.7
miR164a	AT5G61430	*NAC5*	26.8	6.9	4.7
miR164a	AT5G53950	*CUC2*	1.9	2.7	23.1
miR396a−5p	AT1G10120	*bHLH74*	25.8	13.1	5.2
miR396a−5p	AT2G22840	*GRF1*	5.8	11.5	31.7
miR396a−5p	AT4G37740	*GRF2*	2.7	7.5	23.4
miR396a−5p	AT2G36400	*GRF3*	14.6	13.7	13.3
miR396a−5p	AT3G52910	*GRF4*	11.9	48	28.5
miR396a−5p	AT5G53660	*GRF7*	2.6	1.5	5.2
miR396a−5p	AT4G24150	*GRF8*	3.4	2.6	29.4
miR172c	AT4G36920	*AP2*	48.5	3.9	2.0
miR172a	AT5G60120	*TOE2*	4.2	2.5	9.4
miR159a	AT5G06100	*MYB33*	3.2	6.1	19.8
miR159a	AT3G11440	*MYB65*	2.5	5.5	27.9
miR167a−5p	AT1G51760	*IAR3*	116.4	13.6	2.6
miR167a−5p	AT1G30330	*ARF6*	88.2	40.2	10.9
miR167a−5p	AT5G37020	*ARF8*	22.4	79.4	27.3
miR403−3p	AT1G31280	*AGO2*	2.9	22.9	57.1
TAS1A−3'D6[+]	AT4G29770	*HTT1*	2.5	2.5	40.7
TAS1A−3'D2[+]	AT5G18040	*HTT2*	49.8	46.5	11.2
TAS1A−3'D2[+]	AT5G18065	*HTT3*	3.2	4.4	13.7
TAS3A−5'D7[+]	AT5G62000	*ARF2*	75.9	4.7	0.9
TAS3A−5'D7[+]	AT2G33860	*ARF3*	10.3	11.9	10.2
TAS3A−5'D7[+]	AT5G60450	*ARF4*	3.7	2.3	9.8
TAS3A−5'D8[+]	AT5G60450	*ARF4*	3.7	2.3	5.9
miR858b	AT2G47460	*MYB12*	27.4	31.8	10.4
miR858a	AT1G06180	*MYB13*	7.7	13.4	15.0
miR394a	AT1G27340	*LCR*	12.6	4.0	6.2
miR399a	AT2G33770	*UBC24*	13.3	0.7	1.3
miR399f	AT2G33770	*UBC24*	13.3	0.6	1.0
miR390a−5p	AT3G17185	*TAS3a*	19.1	4.2	3.1
miR390a−3p	AT1G08125		15.7	25.9	25.2
miR447a−3p	AT5G60760		7.2	1.3	9.6
miR169h	AT1G72830	*HAP2C*	10.3	1.6	2.5
miR824−5p	AT3G57230	*AGL16*	10.1	1.1	1.6
miR391−5p	AT4G29900	*CIF1*	20.6	0.8	0.9
miR827	AT1G02860	*BAH1*	14.3	2.5	4.5
miR5021	AT1G20090	*ROP2*	70.1	12.2	2.5
miR400	AT1G06580	*PPR1*	1.5	0.8	4.8

Supplemental Figure S11. Heatmaps of mean capped feature RPM. (*left*), cleaved transcript RP10M (*center*), and percent cleaved of capped and cleaved RPM (*right*) for all sRNA-target interactions identified in at least two bioreplicates of one or more sample types (i.e. high-confidence interactions). Each row is scaled to the maximum value measured for that row, displayed to the right of each heatmap.

5.11 Supplemental Methods

NanoPARE Protocol

The initial part of this protocol (Reverse Transcription through Tagmentation) is directly adapted from a Smart- seq2 protocol (Picelli et al. 2013). We have included it here for convenience and compatibility with the nanoPARE protocol.

Materials:

Oligonucleotides (see Table S1)

Reverse Transcription

- dNTP mix (10 mM each; ThermoFisher)
- Superscript II reverse transcriptase (200 U/µl; Invitrogen/Life Tech)
- Superscript II First-Strand Buffer (5×; Invitrogen/Life Tech)
- DTT (100 mM; Invitrogen/Life Tech)
- Murine RNAse inhibitor (40 U/µl; NEB)
- Betaine (5 M; Sigma)
- $MgCl_2$ (100 mM; Sigma)

PCR preamplification

- KAPA HiFi HotStart ReadyMix (2×; KAPPA Biosystems)
- EB solution (Qiagen) or Nuclease free water
- High-sensitivity DNA chip (Agilent Bioanalyzer)

Tagmentation reaction and final PCR

- Nextera DNA Sample Preparation kit (Illumina)

Reverse transcription

1. Prepare first-strand reaction mix as follows (7 µl per reaction):

Superscript II RT	0.5 µl
Murine RNase Inhibitor0.	25 µl
Superscript II first-strand buffer (5×)	2 µl
DTT (0.1 M)	0.25 µl
Betaine (5 M)	2 µl
$MgCl_2$ (100 mM)	0.9 µl
TSO (10 µM)	1 µl
Nuclease-free water	0.1 µl

2. Mix 1 µl of total RNA with 1 µl of anchored oligo-dT primer (10 µM) and 1 µl of dNTPs.
3. Denature at 72 ˚C for 3 minutes and place immediately on ice.
4. Add 7 µl of first-strand reaction to RNA/oligo-dT/dNTPs, pipette up and down 10 times.
5. In thermal cycler incubate at 42 ˚C for 90 minutes, then 10 cycles of (50˚C for 2 minutes, 42˚C for 2 minutes) and finally inactivate by incubating at 70 ˚C for 15minutes.

PCR preamplification

1. Make PCR master mix, add directly to inactivated RT reaction, and mix by pipetting up and down 10 times:

Nuclease-free water	14 µl
2× KAPA HiFi HotStart ReadyMix	25 µl
10 µM ISPCR Primer (1 uM final)	1 µl
RT reaction	10 µl

2. In thermal cycler incubate at 98˚C for 3 minutes, 11-20 cycles of (98˚C for 15 seconds, 67˚Cfor 20 seconds, 72˚C for 6 minutes) and a final extension at 72˚C for 5 minutes.

Note: *Adjust PCR preamplification cycle count for RNA input. Here are some guidelines for total RNA: 11 cycles for 5 ng, 13 cycles for 1 ng, 15-16 cycles for 100 pg and 16-18 cycles for 10 pg. We recommend performing negative controls by replacing the RT reaction products with water and using the maximum number of PCR cycles used for samples.*

3. Purify PCR product using a 1:1 ratio of AMPure XP beads:

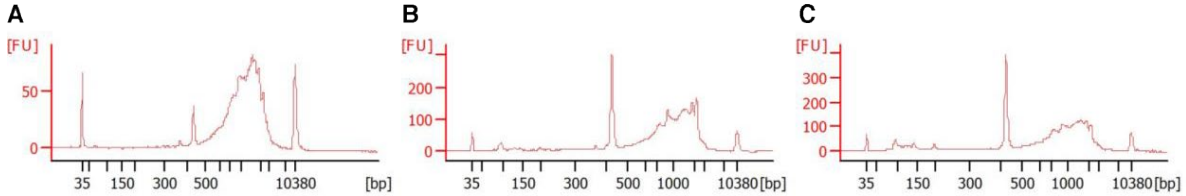
a. Keep the beads at room temperature for 30 minutes.
b. Vortex two times for 2 seconds and add 50 µl of beads to eachsample.
c. Pipette up and down 10 times and incubate at room temperature for 8minutes.
d. Place on the magnetic stand for five minutes.
e. Remove supernatant (do not disturb the beads). Keep on the magnetic stand and add 200 µl of fresh 80% ethanol, incubate for 30 seconds and remove supernatant.
f. Repeat the washing step and remove the supernatant completely.

Note: *Tn5 is sensitiveto ethanol.*

g. Incubate at room temperature for five minutes and remove from the magneticstand.
h. Add 16 µl of elution buffer (EB solution from Qiagen), pipette up and down 10 times, put on the magnetic stand for 2 minutes and collect 15 µl in new Eppendorf tube (1 µl is left to avoid taking up beads).

4. Run 1 µl on High-sensitivity Bioanalyzer DNA chip (Agilent) or Fragment Analyzer (Advanced Analytical) to determine median size and range. Measure concentration between 300 and 9000 bp (Supplemental Methods Fig. 1).

Note: *expected sizes vary by organism, but fragments < 300 bp should be negligible*



Supplemental Methods Figure 1. **Representative profiles of cDNA** generated from **(*A*)** 1 ng of total RNA input and 13 PCR cycles, **(*B*)** 100 pg of total RNA input and 15 PCR cycles and **(*C*)** 10 pg of total RNA input and 18 PCR cycles.

Note*: This is a good pausing point in the protocol. cDNA can be frozen and stored at 20˚C.*

Tagmentation using Nextera DNA Library preparation kit

- Use **5 ng** of cDNA for tagmentation:

cDNA	x µl (to a total of 5 ng)
Tn5 Tagmentation enzyme	2.5 µl
Tagmentation Buffer	12.5 µl
Nuclease free water	x µl (to a total of 25 µl)

- Incubate for 5 min at 55˚C and immediately cool to 4˚C or place on ice
- Clean up with DNA Clean & Concentrator-5 kit (Zymo Research):
  - Add 125 µl binding solution to 25 µl tagmentation reaction
  - Apply to column and spin for 30 seconds at >12000g
  - Wash 2× with 200 µl washing buffer
  - Elute in 21 µl of resuspension buffer

5’ end enrichment PCR (NanoPARE)

At this point, transcript body regions can be enriched from the above PCR amplification products using standard Nextera primers according to Picelli *et al.* (i.e. Smart-seq2; Picelli *et al.* 2013) or RNA 5′ ends can be enriched as described below.

- Split tagmentation reaction into two 10 µl aliquots and set enrichment PCR with either Tn5.1 or Tn5.2 as the reverse primer and P5 N5xx TSO enrichment primer as the forward primer

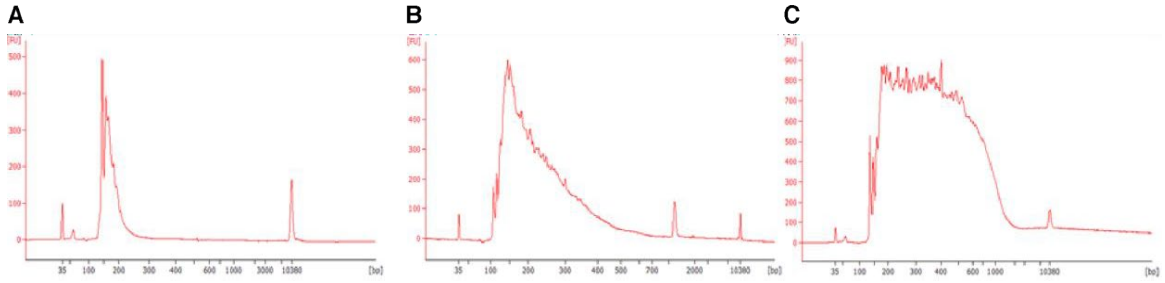
Nextera PCR Mastermix	7.5 µl
10µM P5 N5xx TSO enrichment Primer	2.5 µl (1 uM final)
10µM P7 Tn5.1/Tn5.2 enrichment Primer	2.5 µl (1 uM final)
Tagmentation reaction products	10 µl
Nuclease free water	2.5 µl

In thermal cycler perform enrichment PCR by incubating at 72 ˚C for 3 minutes to allow extension of single-stranded DNA adapters to form double-stranded DNA, 98 ˚C to denature the DNA, then 10 cycles of (98 ˚C for 10 seconds, 63 ˚C for 30 seconds, 72 ˚C for 3 minutes). Inactivate by incubating at 72 ˚C for 1 minute and hold at 12 ˚C.

- Pool both 25 µl reactions and clean up using AMPure XP beads in a 1:1 ratio as described above (50 µl beads).

Quality control

Run High-sensitivity Bioanalyzer DNA chip (Agilent) or Fragment Analyzer (Advanced Analytical) to check for mean fragment length. Informative fragments should have a length of 150-800 bp. Overtagmentation results in short fragments of about 135 bp long, which will have low mappability (Supplemental Methods Figure 2). In such cases, decrease the incubation time of the tagmentation reaction in the heating block. In case of undertagmentation check cDNA input and make sure it does not exceed 5 ng for 2.5 µl Tn5 enzyme mix.



Supplemental Methods Figure 2. Representative profiles of nanoPARE libraries where tagmentation reaction was performed either **(*A*)** excessively (i.e. overtagmented), **(*B*)** optimally or **(*C*)** insufficiently (i.e. undertagmented).

Next-generation sequencing

- Quantify library and pool in the equimolar ratio (about 4 nM each. Follow guidelines from Illumina Platform).
- For Illumina HiSeq2500 machines (this study): Sequence with custom sequencing primer listed in Supplemental Table S1 (TSO_seq_read1). Typically Single Read 50 base mode at a sequencing depth >10-15 million reads is sufficient. If possible, replace Index 1 (i7) Read primer with an equimolar mix of primers P7_Tn5.1_N701 and P7_Tn5.1_N701 to allow for dual indexing. Alternatively, run dual indexing with chemistry-only cycles for read 1 (i7) and use i5 sequences only for indexing.
- For Illumina NextSeq machines: Replace Index 2 (i5) Read primer with a custom primer (CCCGTACTCTGCGTTGATACCACTGCTTGCTAG), which is the reverse complement of custom sequencing primer listed in Supplemental Table S1 (TSO_seq_read1).

6. MicroRNA Dynamics and Functions During *Arabidopsis* Embryogenesis

type published
The Plant Cell Dec 2019, 31 (12) 2929-2946;

doi: 10.1105/tpc.19.00395

authors Alexandra Plotnikova (AP)*, Max J. Kellner (MJK)*, Michael A. Schon (MAS), Magdalena Mosiolek (MM), Michael D. Nodine (MDN)

Contribution Methodology, M.D.N., A.P., M.J.K. and M.M.; Investigation, A.P (including nanoPARE libraries preparation for *xrn4* mutants, XRN1-treated and wild-type floral buds; generation and microscopic analysis (CLSM) of miRNA activity sensors; generation, molecular characterization (qPCR), genetic (crosses) and phenotypic analysis (DIC) of miR-resistant and control (genomic) transgenic lines. Obtained data used for Fig4, 7, 8, S5), M.D.N., M.J.K. and M.M.; Writing – Review & Editing, M.D.N., M.A.S., M.J.K. and A.P.; Visualization, M.D.N., M.A.S. and A.P.

6.1 Abstract

MicroRNAs (miRNAs) are short non-coding RNAs that mediate the repression of target transcripts in plants and animals. Although miRNAs are required throughout plant development, relatively little is known regarding their embryonic functions. To systematically characterize embryonic miRNAs in *Arabidopsis thaliana*, we developed or applied high-throughput sequencing-based methods to profile hundreds of miRNAs and associated targets throughout embryogenesis. We discovered dozens of miRNAs that dynamically cleave and repress target transcripts, including 30 that encode transcription factors. Transcriptome analyses indicated that these miRNA:target interactions have profound effects on embryonic gene expression programs. Moreover, we demonstrated that the miRNA-mediated repression of six transcription factors is individually required for proper division patterns of various embryonic cell lineages. These data indicate that the miRNA-directed repression of multiple transcription factors is critically important for the establishment of the plant body plan, and they provide a foundation to further investigate how miRNAs contribute to these initial cellular differentiation events

6.2 Introduction

MicroRNAs (miRNAs) are a class of small regulatory RNAs (sRNAs) that post-transcriptionally repress gene expression and regulate cellular differentiation during plant and animal development (Bartel 2004; Plasterk 2006; Jones-Rhoades *et al.* 2006; Chen 2009). Plant miRNA precursors fold into characteristic RNA stem-loop structures that are recognized and processed into mature ~21-nt miRNAs by the RNase III domain-containing protein DICER-LIKE1 (DCL1) (Park *et al.* 2002; Reinhart *et al.* 2002). miRNAs are then loaded onto ARGONAUTE1 (AGO1) proteins and guide the complex to sequences in target RNAs that are almost perfectly complementary to the miRNA (Jones-Rhoades and Bartel, 2004; Allen *et al.* 2005). In general, miRNAs recognize single sites in target transcripts, and the high degree of miRNA:target duplex base-pairing results in target RNA cleavage, although translational repression has also been reported (Kasschau *et al.* 2003; Llave *et al.* 2002; Jones-Rhoades and Bartel 2004; Aukerman and Sakai 2003; Chen 2004; Gandikota *et al.* 2007). The miRNA-mediated cleavage and repression of transcripts including those encoding transcription factors are required throughout development (Jones-Rhoades *et al.* 2006; Chen, 2009; D'Ario *et al.* 2017). Although miRNAs have been implicated in an array of post-embryonic developmental processes, their functions during embryogenesis remain less well-characterized (Vashisht and Nodine, 2014). This is primarily due to early embryos being small and deeply embedded in maternal seed coat tissues, which makes it difficult to isolate them at high purity and characterize the corresponding RNA populations (Schon and Nodine, 2017).

Nevertheless, the precursors of the shoot and root meristems and three main radial tissue layers are precisely established during early embryogenesis, and miRNAs are required for

most of these early patterning events (Schwartz *et al.* 1994; Nodine and Bartel 2010; Willmann *et al.* 2011; Seefried *et al.* 2014). Moreover, miRNAs are required to prevent the precocious expression of genes involved in embryo maturation when storage macromolecules such as oil bodies accumulate (Willmann *et al.* 2011; Nodine and Bartel,2010). Embryonic miRNAs therefore help define cell-specific gene expression programs according to both spatial and temporal cues. For example, miR165/166 spatially restrict RNAs encoding homeobox-leucine zipper family transcription factors during embryogenesis (McConnell *et al.* 2001; Smith and Long 2010; Miyashima *et al.* 2013), and miR156/157-mediated repression of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) transcription factor genes is required for both the proper divisions of root meristem precursors and to prevent the precocious expression of maturation phase genes (Nodine and Bartel, 2010). *Arabidopsis thaliana mir160a* loss-of-function mutant embryos divide incorrectly, and the abnormal cotyledon phenotypes of seedlings expressing transgenes containing mutations in miR160, miR170/171 or miR319 target sites suggest that the corresponding miRNA activities are required for embryo morphogenesis (Mallory *et al.* 2005; Liu *et al.* 2010; Takanashi *et al.* 2018; Palatnik *et al.* 2003). The cell-type-specific miR394-mediated repression of transcripts encoding the LCR F-box protein is also required for patterning embryonic apical domains (Knauer *et al.* 2013).

Despite these individual examples of embryonic miRNA functions and miRNA profiling studies on late-stage plant embryos (Xu *et al.* 2018; Huang *et al.* 2013; Oh *et al.* 2008), a comprehensive understanding of embryonic miRNA populations and their individual contributions to embryogenesis is incomplete. *Arabidopsis* embryos are ideal model systems to investigate the roles of miRNAs during plant embryogenesis. Not only do the available genomic and genetic resources in *Arabidopsis* facilitate the functional characterization of miRNA, but *Arabidopsis* embryos undergo a series of highly stereotypical cell divisions to generate the basic body plan (Palovaara *et al.* 2016; Mansfield and Briarty, 1991). Therefore, abnormal cell division patterns in early *Arabidopsis* embryos can be screened for upon disrupting miRNA functions in order to test whether miRNAs are required for morphogenesis, and thus yield insights into the molecular basis of the corresponding patterning events. In the current study, we developed a low-input small RNA sequencing (sRNA-seq) method to generate profiles of hundreds of miRNAs and used the recently developed nanoPARE approach (Schon *et al.* 2018) to identify corresponding target transcripts throughout embryogenesis. We found that miRNAs dynamically cleave and repress at least 59 transcripts, including 30 encoding transcription factors belonging to eight different families. As a proof-of-principle of this dataset's utility, we selected individual miRNA/target interactions to investigate further and demonstrated that the miRNA-mediated repression of six RNAs encoding transcription factors are individually required for the proper cell division patterns of various post-embryonic tissue-type precursors. Therefore, this resource provides a foundation to further investigate how miRNAs help coordinate the

formation of the basic body plan by post-transcriptionally restricting their targets, including transcription factors, to specific stages and cell-types.

6.3 Results

Establishment of Low-input Small RNA Sequencing Method

To systematically characterize the dynamics and functions of individual embryonic miRNAs in *Arabidopsis*, it was first necessary to identify the miRNAs present in developing embryos. However, standard high-throughput sRNA-seq methods require relatively large amounts of total RNA, which are impractical to obtain from early embryos. The sequential ligation of adapters onto the hydroxyl and monophosphate groups at the respective 3′ and 5′ termini of sRNAs, followed by reverse transcription and PCR amplification during conventional sRNA-seq library preparation, requires ≥500 ng of total RNA, which is approximately 100 times more than can be obtained from early *Arabidopsis* embryos. More recent sRNA-seq methods can profile sRNAs from as little material as a single cell, but they do not enrich for sRNAs to the same extent as conventional methods (Faridani *et al.* 2016). Therefore, to enable the profiling of miRNAs present in developing *Arabidopsis* embryos, we developed a method employing the NEBNext Multiplex Small RNA Library Prep Set for Illumina kit (NEB) that is suitable for the low amounts of total RNA obtainable from early embryos (i.e. 1–5 nanograms (ng)). In brief, we included polyacrylamide gel-based size-selection methods to both enrich for sRNAs from total RNA before the first adapter ligation step, as well as to enrich for desired sRNA cDNAs after final PCR amplification (see Methods for details). We also reduced the amounts of 3′ adapters, reverse transcriptase primers, and 5′ adapters used in the NEBNext kit when starting with ≤500 ng of total RNA.

We compared sequencing data from libraries generated with 500, 50, 5, 1 or 0.5 ng of total RNA isolated from bent cotyledon stage Col-0 (hereafter referred to as wild-type) embryos to determine how well the method enriches for sRNAs, as well as the method's reproducibility and accuracy when starting with different amounts of total RNA. Approximately 21 nt miRNAs and 24 nt small interfering RNAs that typically begin with uridine- and adenosine-monophosphates, respectively, are characteristic features of plant sRNA populations (Borges and Martienssen, 2015). As expected for plant sRNAs, libraries generated from all input amounts of total RNA predominantly consisted of 21–24-base reads, with the first position of the 21- and 24-base reads enriched for thymine and adenine, respectively (Figure 1A and 1B; Supplemental Figure 1A to 1C). The distribution of sRNA-seq read sizes and 5′ nt biases indicated that the sRNA-seq protocol highly enriches for sRNAs from as little as 0.5 ng of total RNA. To determine the reproducibility of the method across various amounts of input RNA, we compared miRNA family levels between libraries constructed from 500 ng of total RNA with those generated from either 50, 5, 1 or 0.5 ng of total RNA. miRNA levels were highly correlated between biological replicate libraries generated from 500 ng of total RNA (Pearson's R >0.99) (Supplemental Figure 1D and 1E). Pearson's correlation coefficients were

>0.9 between 500 ng libraries and all libraries generated from ≥1 ng of total RNA (Figure 1C; Supplemental Figure 1 F to H).

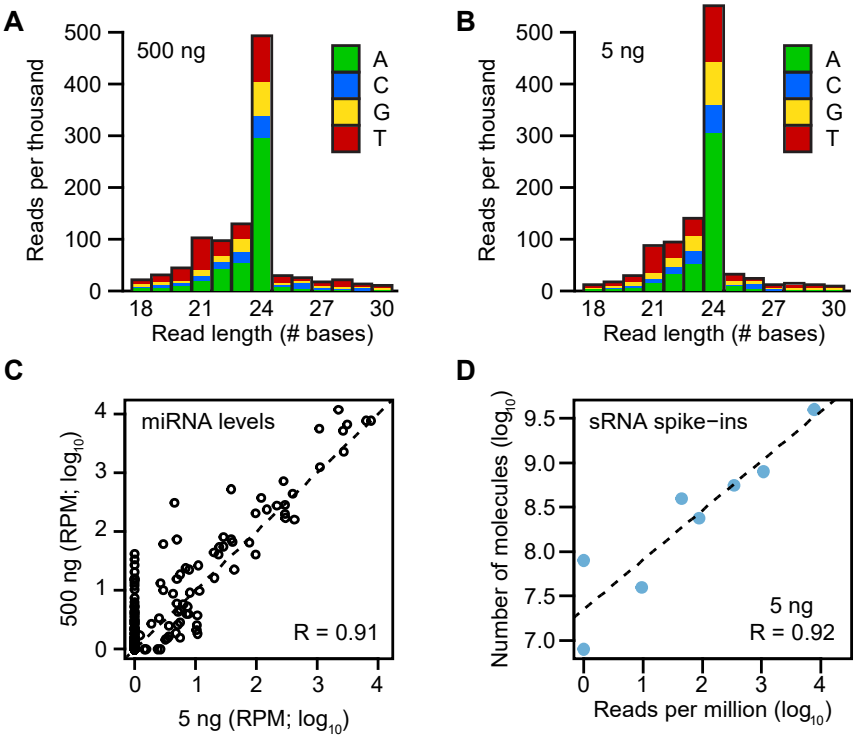


Figure 1. Establishment of Low-Input Small RNA Sequencing Method. (A and B) Stacked bar charts of normalized sRNA-seq read levels (reads per thousand genome-mapping reads) across different base lengths in libraries generated with either 500 ng **(A)** or 5 ng **(B)** of total RNA isolated from bent cotyledon stage embryos. Colors indicate the proportions of sRNA-seq reads that begin with various bases as indicated in the key. **(C)** Scatter plot of miRNA family levels in sRNA-seq libraries generated from 5 ng and 500 ng of total RNA. sRNA levels were normalized for reads per million genome-matching reads (RPM) and log10-transformed. Pearson's R value is indicated, as well as a dashed line with an intercept of 0 and slope of 1. **(D)** Scatter plot of relative sRNA spike-in levels (RPM; log10) compared to the absolute number of sRNA spike-in molecules (log10) added during RNA isolation for a sRNA-seq library generated from 5 ng of total RNA. Pearson's R value is shown, and the dashed line represents a linear model derived from the plotted data points.

We also assessed the accuracy of this low-input sRNA-seq method across the dilution series of input RNA by adding exogenous sRNA oligonucleotides (i.e. spike-ins) (Lutzmayer *et al.* 2017) during RNA isolation prior to library construction and examined spike-in levels in the resulting sRNA-seq datasets. If the method accurately quantified sRNA levels, we would expect a high correlation between the absolute number of spike-in molecules added and the number of sRNA-seq reads mapping to the spike-ins. Pearson's correlation coefficients between the absolute amounts of spike-ins added and the relative amounts of spike-ins sequenced were >0.9 for all libraries generated from ≥1 ng total RNA (Figure 1D; Supplemental Figure 1I to 1L). The progressive increase in the number of undetected miRNA families and sRNA spike-ins as total RNA amounts decreased indicated that the sensitivity of the method was reduced when starting with less than 50 ng of total RNA (Figure 1C to 1D; Supplemental Figure 1G and 1H, 1K and 1L). Regardless, the modified sRNA-seq library construction method allowed us to highly enrich for sRNAs and to reproducibly and accurately quantify miRNA levels when starting with 1–5 ng of total RNA, which are amounts obtainable from early *Arabidopsis*

embryos.

Embryonic miRNA Dynamics

We then used this low-input sRNA-seq method to generate libraries using total RNA isolated from embryos at eight developmental stages including three main phases of embryogenesis (Hofmann *et al.* 2019) (Figure 2A; Supplemental Data Set 1). Three pools of 50 embryos were isolated from each of the eight stages from different plants and on different days, and considered biological replicates (1,200 embryos in total). At least 80% of the total RNA isolated from each biological replicate was used to generate sRNA-seq libraries, and the remainder was used to generate full-length cDNAs to profile either transcriptomes (Hofmann *et al.* 2019) or miRNA-mediated cleavage products (see below).

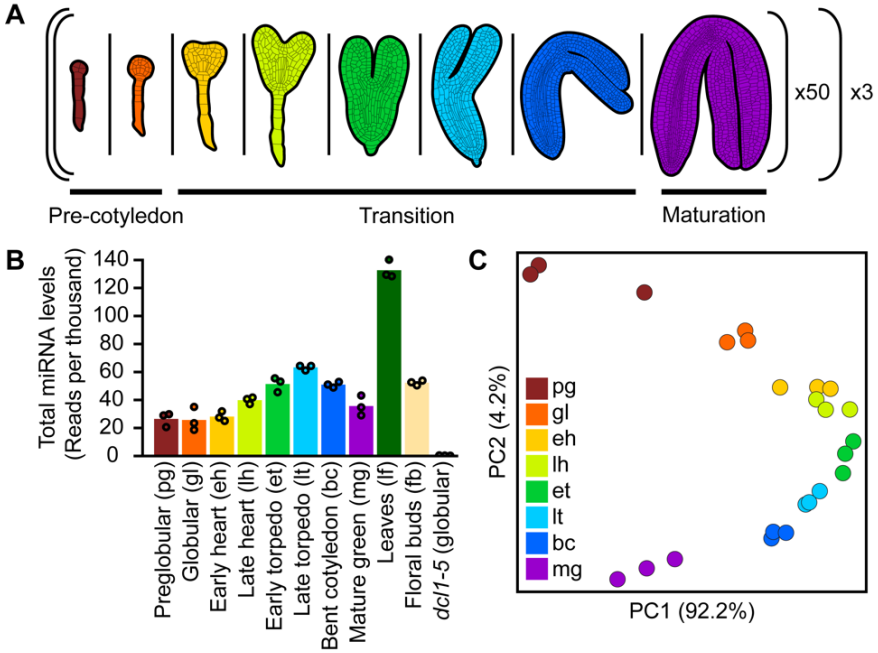


Figure 2. Application of the Low-Input sRNA Sequencing Method to *Arabidopsis* Embryos. (A) Schematic of sRNA profiling experiment across embryogenesis with the low-input sRNA-seq method. Fifty embryos from each of eight different embryonic stages spanning three indicated main phases of embryogenesis were pooled into individual biological replicates. This was repeated three times for each stage to generate three biological replicates for each of the eight developmental stages (i.e. 24 libraries from a total of 1,200 embryos). **(B)** Bar chart displaying the total amount of miRNAs detected across wild-type embryogenesis, leaves and floral buds, as well as in miRNA-deficient *dcl1-5* embryos isolated at the globular stage. miRNA levels were normalized by reads per thousand genome-mapping reads. Points indicate the mean levels of individual biological replicates. Individual stages, together with their abbreviations, are labelled. **(C)** Principal component analysis illustrating the relationships of the 24 sRNA-seq libraries generated from wild-type embryonic tissues based on miRNA levels. Embryonic stages are labelled according to the key.

Previous analysis of mRNA-seq libraries generated from an aliquot of the same total RNA demonstrated that the embryonic RNA samples were not significantly contaminated with non-embryonic RNAs (Hofmann *et al.* 2019), which had been a frequent problem in early embryonic *Arabidopsis* transcriptome datasets (Hofmann *et al.* 2019; Schon and Nodine, 2017). Total miRNA levels fluctuated in wild-type embryos according to their developmental stage but were almost completely lost in dicer-like1-5 (*dcl1-5*) null mutants (Figure 2B). Because DCL1 is required for miRNA biogenesis (Reinhart *et al.* 2002; Park *et al.* 2002), this

further supports the validity of the miRNAs identified in the sRNA-seq libraries. Principal component analysis of miRNA family levels in libraries generated from embryonic and post-embryonic tissues demonstrated that biological replicates clustered together (Figure 2C; Supplemental Figure 2).

Furthermore, the developmental stages of the embryonic samples were stratified along the second principal component and were clearly separated from the post-embryonic leaf and flower samples. By applying the low-input sRNA-seq method to developing embryos, we were able to generate high-quality profiles of embryonic miRNAs, which changed in composition across developmental stages. We detected 349 miRNAs belonging to 259 families in at least one embryonic stage (Supplemental Data Set 2). We then selected 59 miRNA families detected with an average of ≥10 reads per million genome-matching reads (RPM) in at least one embryonic stage to examine in greater detail. Three groups of miRNAs with similar dynamics across embryogenesis were observed (Figure 3A; Supplemental Figure 3). Twenty-two miRNA families accumulated during the late transition phase and persisted in mature green embryos. These included miR394, miR403 and miR170/171, as well as miR167 and miR390, which were both previously detected in late-stage embryos with whole-mount RNA *in situ* hybridizations (Ghosh Dastidar *et al.* 2016). Another set of 25 miRNA families, including miR156/157, miR161, miR164 and miR319, accumulated during the transition phase, but their levels were then reduced in mature embryos. Twelve miRNA families had relatively high levels during early embryogenesis and decreased thereafter. Based on further analysis of internally generated and publicly available sRNA-seq data from 26 tissue types (Xu *et al.* 2018), five miRNA families were highly enriched during the initial stages of embryogenesis, including miR156b-3p, miR831, miR845, miR866-3p, and miR3440b-3p (Figure 3B; Supplemental Figure 4). To examine whether miRNA levels vary between early embryonic cell types, we adapted a whole-mount sRNA *in situ* protocol (Ghosh Dastidar *et al.* 2016) to detect four selected miRNAs in sections of early embryos. Consistent with previous reports (Nodine and Bartel, 2010), miR156/157 was localized throughout wild-type embryos, and a similar pattern was also observed for miR159 (Figure 3C). miR165/166 confers repressive activities in the peripheral cell-types of embryos (Smith and Long, 2010; Miyashima *et al.* 2013; McConnell *et al.* 2001), and miR165/166 levels were accordingly higher in these outer cell types (Figure 3C). By contrast, miR160 levels were higher in the innermost vascular precursor cells (Figure 3C). sRNA *in situs* performed with probes antisense to the mouse-specific miR124 miRNA produced low signal compared to probes antisense to the four miRNAs in wild-type embryos or embryos with wild-type morphologies from *dcl1- 5/+* self-pollinated plants (i.e. wild-type or *dcl1-5/+* embryos) (Figure 3C). Moreover, probes antisense to the four miRNAs produced highly reduced signals when applied to miRNA-deficient *dcl1-5* embryos compared to wild-type or *dcl1-5/+* embryos (Figure 3C). These controls further support the specificity of the signal observed from the miRNA *in situ* hybridizations.

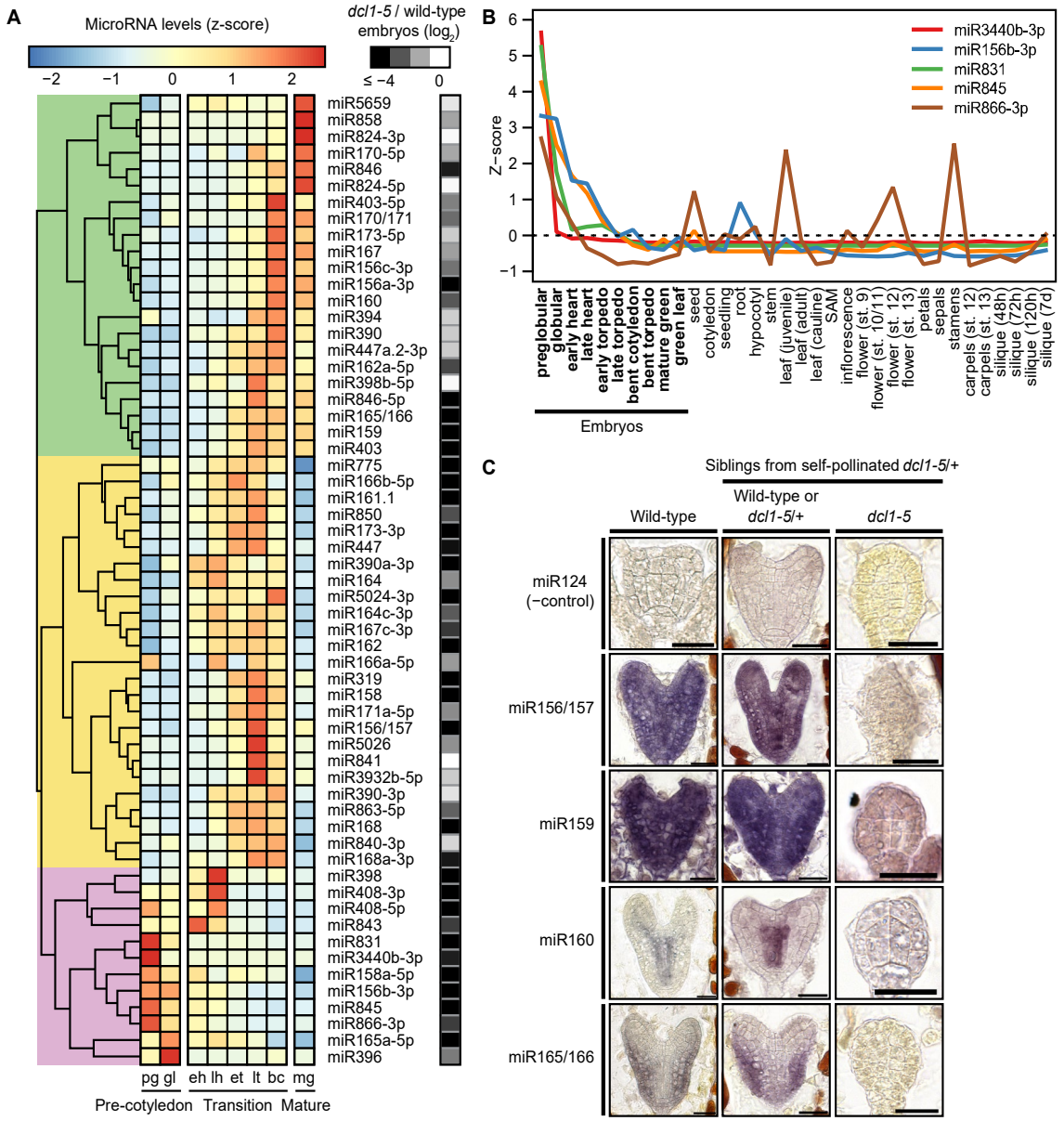


Figure 3. miRNA Dynamics During Embryogenesis.
(A) Heat map illustrating the relative levels of miRNA families across embryogenesis. miRNA families with ≥10 mean RPM in at least one embryonic stage are shown, and colors represent z-scores for each individual miRNA family according to the key. Log2-transformed levels of miRNAs in (*dcl1-5* +1)/(wild-type +1) are annotated. Three major phases of embryo development are labelled at the bottom, and individual columns are labelled according to the stage: pg, preglobular; gl, globular; eh, early heart; lh, late heart; et, early torpedo; lt, late torpedo; bc, bent cotyledon; mg, mature green. The dendrogram is highlighted in green, yellow or violet to indicate three clusters of miRNA families. **(B)** Line graph depicting relative levels (z-scores) of preglobular-enriched miRNA families across development. Five miRNA families were selected based on their enrichment in embryos compared to internally generated leaf and floral bud sRNA-seq libraries. sRNA-seq libraries generated internally are marked in bold, and published sRNA-seq data from 26 tissue types (Xu *et al.* 2018) are also shown. **(C)** Representative images of miRNA *in situ* hybridizations on sections of embryos. Probes antisense to four miRNAs detected in embryonic sRNA-seq libraries are shown, and the corresponding miRNA families are labelled. Probes antisense to the mouse miR124 were used as negative controls, as well as miRNA-deficient *dcl1-5* embryos. Scale bars are 20 µm.

Identification of Embryonic miRNA Targets

Based on our analyses, embryonic miRNA populations were distinct from those in post-embryonic tissues, and their levels frequently exhibited dynamic changes across developmental stages and sometimes cell types. These results suggest that miRNAs have distinct functions during different phases of embryogenesis. Because miRNA functions are largely defined by the targets they regulate, we next determined the targets of embryonic miRNAs. In plants, miRNAs typically bind to highly complementary binding sites within target RNAs and mediate their endonucleolytic cleavage (Kasschau *et al.* 2003; Llave *et al.* 2002; Jones-Rhoades and Bartel, 2004). miRNA-mediated cleavage of target RNAs produces cleavage products downstream of the slice site, which can be cloned and sequenced with high-throughput methods referred to as PARE (parallel analysis of RNA ends), GMUCT (genome-wide mapping of uncapped and cleaved transcripts), or degradome sequencing (Addo-Quaye *et al.* 2008; German *et al.* 2008; Gregory *et al.* 2008). Although these groundbreaking technologies have allowed miRNA target identification on a genome-wide scale, they require ≥10,000-fold more input RNA than what was obtainable from early *Arabidopsis* embryos. We previously developed a method called nanoPARE to enable the confident identification of miRNA-mediated target RNA cleavage products from low-input RNA (Schon *et al.* 2018). To identify embryonic miRNA targets, we generated nanoPARE libraries from the same eight stages of embryogenesis used for miRNA profiling in biological triplicates. In addition to these 24 libraries from wild-type embryos, we also generated nanoPARE libraries from three biological replicates of *dcl1-5* globular embryos as controls (Supplemental Data Set 1).

The nanoPARE datasets and target predictions for 164 miRNAs detected ≥1 RPM in ≥1 embryonic stage were used as input for EndCut software (Schon *et al.* 2018). We identified significant target transcript cleavage sites in ≥1 embryonic library (Benjamini-Hochberg adjusted P-values < 0.05) (Supplemental Data Set 3). These 115 target sites included 59 sites that were identified in ≥2 biological replicates from ≥1 developmental stage. We refer to these as high-confidence targets, and characterized these 59 sites corresponding to 22 miRNA families further. The first positions of nanoPARE reads mark RNA 5′ ends. The number of nanoPARE reads at the 59 high-confidence target sites detected in wild-type embryos was significantly reduced (40.5 fold) in miRNA-deficient *dcl1-5* globular embryos (P-value 0.0001; two-tailed K-S test) (Figure 4A). Moreover, no high-confidence targets were detected in *dcl1-5* embryos, and 58/59 of the high-confidence targets detected in developing wild-type embryos had decreased numbers of nanoPARE reads in *dcl1-5* embryos (Figure 4B). A lack of signal in *dcl1-5* embryos could be explained by either a loss of miRNA-mediated cleavage or technical differences in sample RNA quality. To differentiate between these two explanations, we measured nanoPARE signal mapping to the published transcription start sites (Schon *et al.* 2018) of all high-confidence targets detected in globular embryos.

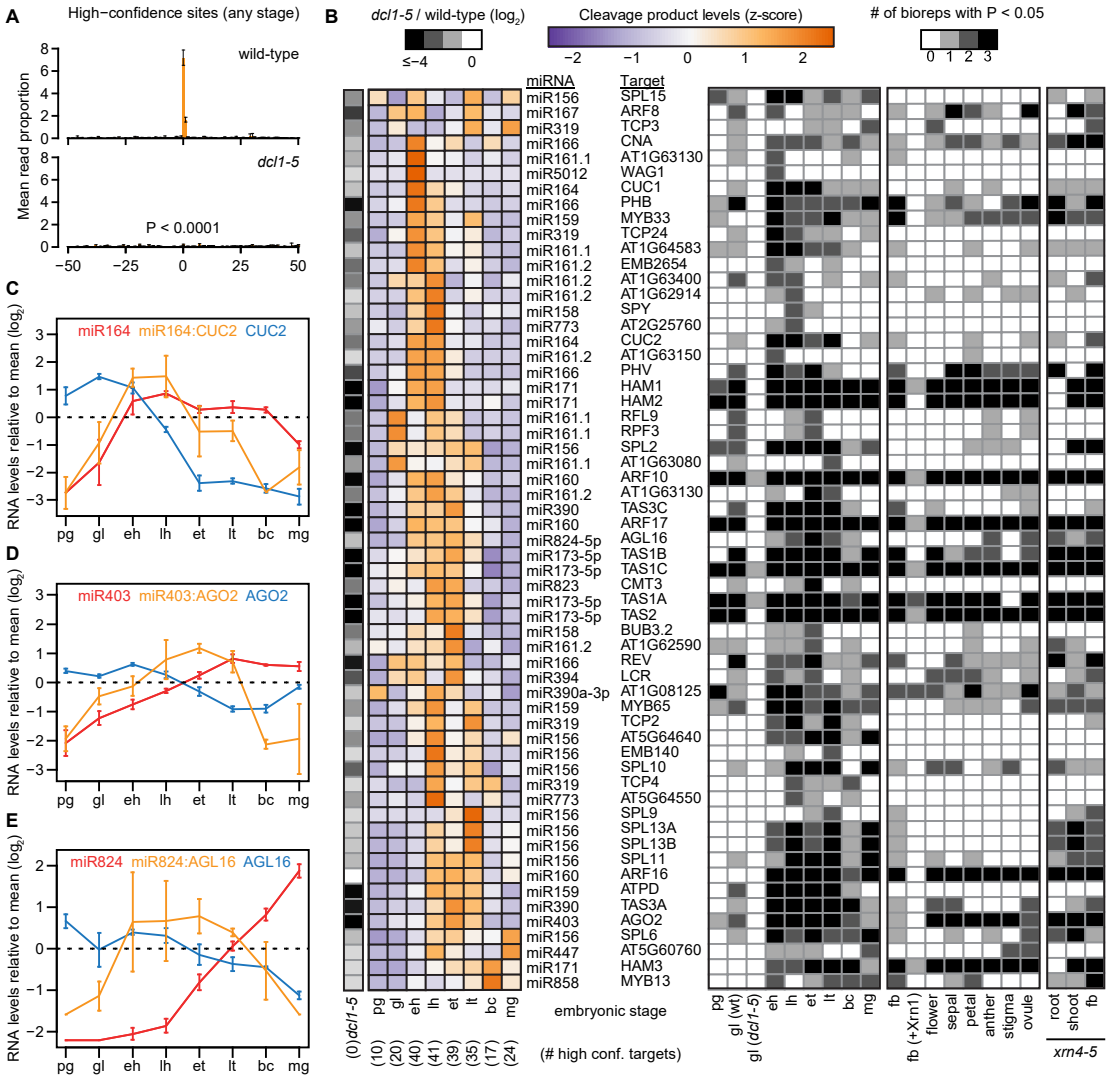


Figure 4. Identification of Embryonic miRNA Targets.
(A) The proportion of nanoPARE reads mapping within 50-nt of miRNA target sites significantly detected by EndCut (Benjamini-Hochberg adjusted P-values < 0.05) in ≥2 biological replicates from any embryonic stage (i.e. high- confidence miRNA cleavage sites; n = 59) are shown for wild-type (top) and *dcl1-5* (bottom) globular embryo libraries. The probability (P-value) that the mean number of reads at the predicted cleavage sites in *dcl1-5* is different from the wild-type mean due to chance is indicated (two-tailed K-S test). Error bars represent the standard errors of the means of three biological replicates. **(B)** Heat maps depicting the relative levels of miRNA-mediated cleavage products (left) and the number of biological replicate libraries in which cleavage products were significantly detected (right). High-confidence miRNA cleavage sites from any embryonic stage are shown together with miRNA families and target transcripts alongside the rows. Colors represent z-scores as indicated in the key. Log2-transformed levels of cleavage products in (*dcl1-5* + 1)/(wild- type + 1) are annotated, and embryonic stages are indicated below each column, as well as the number of high-confidence targets from that stage. (right) Shading densities in the heatmap on the right indicate the number of biological replicates for which the cleavage site was significantly detected by EndCut according to the key (Benjamini-Hochberg adjusted P-value < 0.05). Embryonic stages and post-embryonic tissues in wild-type and *xrn4-5* genotypes are indicated below each column. *dcl1-5*, *dcl1-5* globular; pg, preglobular; gl, globular; eh, early heart; lh, late heart; et, early torpedo; lt, late torpedo; bc, bent cotyledon; mg, mature green; fb, unopened floral buds; +Xrn1, RNA treated with Xrn1 exoribonuclease prior to library construction. **(C-E)** Line graphs illustrating the relative RNA levels of miRNAs (red), targets (blue) and miRNA-mediated cleavage products (orange) for miR164:CUC2 **(C)**, miR403:AGO2 **(D)** and miR824:AGL16 **(E).** Error bars represent the standard errors of the means of three biological replicates for each stage. Relative levels of miRNAs (RPM), cleavage products (reads per ten million genome-matching reads) and transcripts (TPM) for each stage were calculated by log2-transforming (stage levels + 1)/(mean levels across all stages + 1). Embryonic stage abbreviations below each graph are as indicated in panel **B**.

Full-length transcripts were more abundant in *dcl1-5* embryos for 17/20 of these high-confidence targets, demonstrating that the observed reduction of nanoPARE signal at miRNA-directed cleavage sites in *dcl1-5* embryos was not due to differences in RNA quality (Supplemental Figure 5). The loss of miRNA-mediated cleavage sites in miRNA-deficient *dcl1-5* embryos further supports the validity of the miRNA targets identified. miRNA-mediated cleavage products dynamically accumulated and were generally more abundantly detected during mid-embryogenesis (Figure 4B). To identify miRNA-mediated cleavage events that are enriched in embryos, we also analyzed nanoPARE libraries generated either previously from flowers and floral organs (Schon *et al.* 2018), or in this study from root or shoot tissues of exoribonuclease4-5 (*xrn4-5*) mutants, in which miRNA-directed cleavage products are stabilized (German *et al.* 2008; Souret *et al.* 2004).

We observed 11 high-confidence target transcripts enriched in developing embryos, including those encoding the EMB2654 (miR161.2) and SPY (miR158) tetratricopeptide repeat proteins involved in embryogenesis and gibberellic acid responses, respectively, an ATP synthase delta subunit (ATPD; miR159), a plant invertase/methylesterase inhibitor family protein (AT5G64640; miR156/157), and TCP4 and TCP24 (miR319) transcription factors. Interestingly, a simple linear relationship between miRNA abundance and cleavage products was sufficient to explain the dynamics of only a minority of the observed miRNA/target level dynamics during embryogenesis (Supplemental Figure 6). However, a few miRNA/target cleavage products accumulated during embryonic stages at which miRNA levels were increasing and full-length target transcripts were decreasing. For example, miR164:CUP-SHAPED COTYLEDON2 (CUC2) and miR824:AGAMOUS-LIKE16 (AGL16) cleavage products accumulated during mid-embryogenesis when miRNA and target levels were increasing and decreasing, respectively (Figure 4C to 4E). Similarly, miR403:ARGONAUTE2 (AGO2) products were present at relatively high levels during mid-embryogenesis when increasing miR403 levels were concomitant with decreasing AGO2 levels (Figure 4D).

Impact of miRNAs on the Embryonic Transcriptome

To assess how miRNAs influence embryonic transcript levels, we profiled transcriptomes from *dcl1-5* globular embryos in which miRNA levels and cleavage activities were highly reduced (Figure 2B; Figures 3A and 3C; Figures 4A and 4B). Principal component analysis of *dcl1-5* and wild-type embryonic transcriptomes (Hofmann *et al.* 2019) revealed that *dcl1-5* biological triplicates clustered together in a group that was separate from the wild-type transcriptomes (Figure 5A). This suggested that the loss of miRNAs resulted in large-scale changes in transcript populations. Indeed, 3,321 and 1,951 genes had at least two-fold significantly increased and decreased transcript levels in *dcl1-5* relative to wild-type globular embryos, respectively (DESeq2; Benjamini-Hochberg adjusted P-values < 0.01) (Figure 5B) (Supplemental Data Set 4).

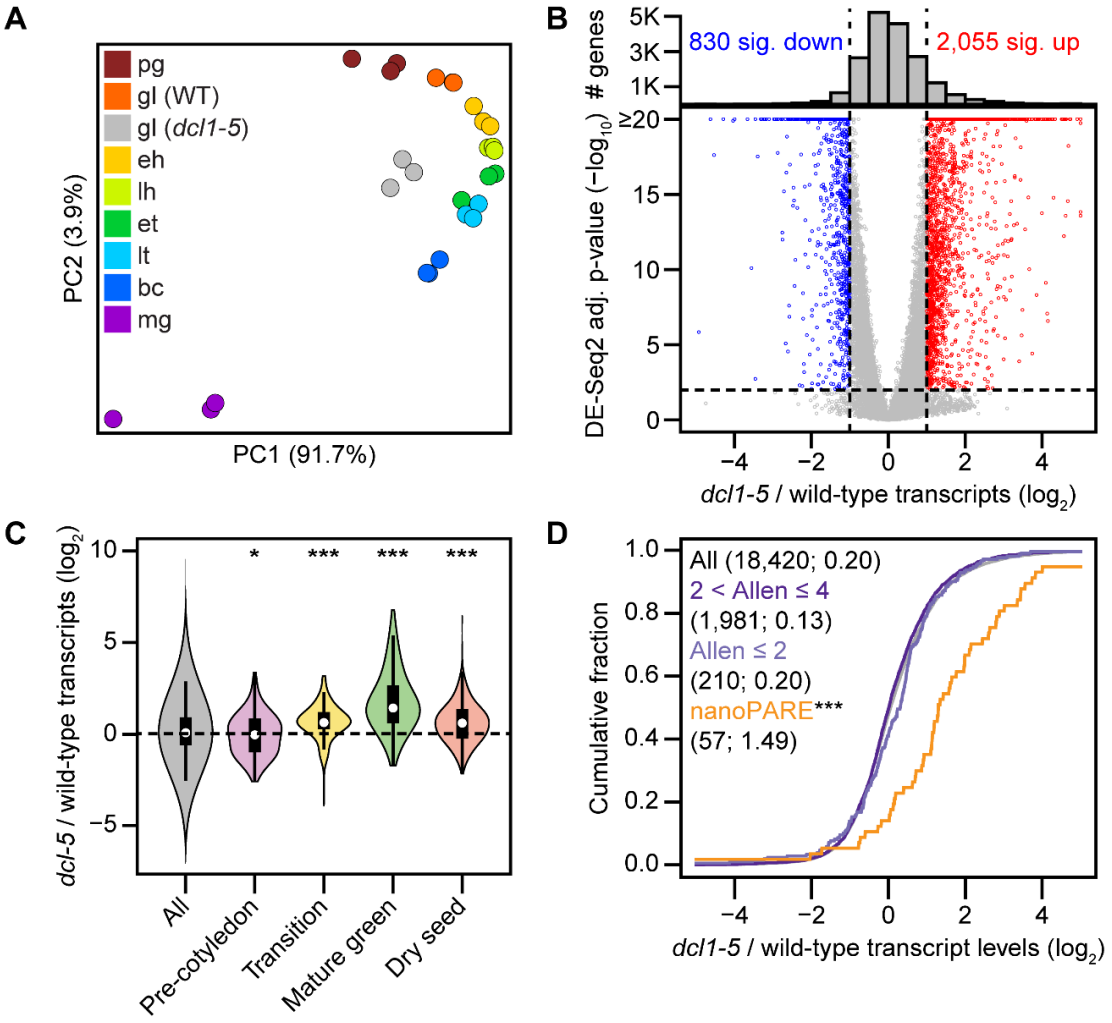


Figure 5. Impact of miRNAs on the Embryonic Transcriptome.
(A) Principal component analysis illustrating the relationships between mRNA-seq libraries generated from biological triplicates of either *dcl1-5* globular embryos or wild-type embryos from eight stages. As shown in the key, libraries are color-coded according to their stage: pg, preglobular; gl (WT), wild-type globular; gl (*dcl1-5*), *dcl1-5* globular; eh, early heart; lh, late heart; et, early torpedo; lt, late torpedo; bc, bent cotyledon; mg, mature green. **(B)** Volcano plot of log2-transformed transcript levels in (*dcl1-5* TPM + 1)/(wild-type TPM + 1) (x-axis) and log10- transformed Benjamini-Hochberg adjusted P-values based on DESeq2 (Love *et al.* 2014) in *dcl1-5* compared to wild- type (y-axis) (bottom). Red and blue indicate transcripts with P-values < 0.01 and ≥2-fold increased levels in *dcl1-5* or wild-type embryos, respectively. Histogram of the total numbers of transcripts across different *dcl1-5*/wild-type transcript fold-changes are shown (top). The number of significantly decreased (sig. dec.) and increased (sig. inc.) transcripts are indicated. **(C)** Violin plots of log2-transformed transcript levels in (*dcl1-5* + 1)/(wild-type + 1) for embryo phase-enriched marker transcripts as defined previously (Hofmann *et al.* 2019). Transcripts with ≥1 TPM in either *dcl1-5* or wild-type globular embryos (All; n = 18,420) and enriched in either pre-cotyledon (n = 107), transition (n = 127), mature green (n = 48) or dry seed (n = 183) phases. P-values < 0.01 and P-values < 10-6 are indicated by * and ***, respectively (two-tailed K-S test). **(D)** Cumulative distributions of (*dcl1-5* TPM + 1)/(wild-type TPM + 1) transcript levels (log2) for all transcripts with ≥1 TPM in either *dcl1-5* or wild-type globular embryos (All; black), transcripts confidently predicted computationally (2 < Allen scores ≤ 4, dark purple; Allen scores ≤ 2, light purple), and high-confidence targets detected by EndCut (nanoPARE, orange; *** indicates P-values = 3.96 × 10-11).

Considering that 18,420 genes had ≥1 transcripts per million (TPM) in either wild-type or *dcl1-5* globular embryos, this indicated that 28.6% of the transcriptome is significantly altered in *dcl1-5* embryos. Differences due to RNA contamination from non-embryonic seed tissues could be ruled out by applying the tissue-enrichment test (Schon and Nodine, 2017),

which revealed no significant RNA contamination in either the wild-type or *dcl1-5* samples (Supplemental Figure 7). These large-scale transcriptome changes may be related to the precocious activation of embryo maturation gene expression programs as previously reported (Nodine and Bartel, 2010; Willmann *et al.* 2011). To test whether transcriptomes from *dcl1-5* globular embryos resembled those from later stages of development, we examined the levels of transcripts that were specifically enriched during one of four main phases of embryogenesis and seed development (Hofmann *et al.* 2019) in *dcl1-5* compared to wild-type globular embryos. The levels of transition, mature green, and dry seed phase marker transcripts were significantly increased in *dcl1-5* globular embryos (P-values < 10-6, two-tailed K-S tests; Figure 5C). Therefore, miRNA-deficient *dcl1-5* embryos indeed prematurely activate late-stage gene expression programs. Because we detected relatively few miRNA targets with nanoPARE compared to the total number of differentially expressed genes (Supplemental Data Set 4; Figure 4B; Figure 5B), we reasoned that either many miRNA targets were not detected with nanoPARE or that the large-scale changes in *dcl1-5* transcriptomes were mostly a consequence of miRNA target de-repression. Plant miRNA targets can be predicted at various confidence levels depending on the frequency and position of the miRNA:target duplex mismatches (i.e. Allen scores) (Allen *et al.* 2005). Whereas the levels of miRNA targets detected with nanoPARE were significantly increased in *dcl1-5* relative to wild-type embryos compared to all expressed genes (2.8-fold; P-value = 3.96 × 10-11, two-tailed K-S test), the levels of targets confidently predicted computationally (i.e. Allen scores ≤2 or ≤4), including bona fide post-embryonic targets, were not substantially increased in *dcl1-5* (Figure 5D). These results are consistent with the nearly complete depletion of miRNAs in *dcl1-5* embryos resulting in the loss of cleavage and repression of dozens of targets and the consequential misregulated miRNA target activities having a large impact on embryonic gene expression programs. Thirty of the fifty-nine high-confidence miRNA targets detected with nanoPARE encoded transcription factors belonging to eight different families, including those containing ARF, GRAS, HD-ZIP, MADS-box, MYB, NAC, SBP and TCP domains (Figure 6A). Twenty-eight of these had transcripts >1 TPM in either wild-type or *dcl1-5* globular embryos, and remarkably, twenty-four (85.7%) were significantly up-regulated in *dcl1-5* compared to wild-type embryos (Figure 6B). RNA *in situ* hybridizations of three miR165/166 target RNAs encoding class III HD-ZIP transcription factors (i.e. PHABULOSA (PHB), CORONA (CNA) and PHAVOLUTA (PHV)) in wild-type embryos were congruous with previous reports (Prigge *et al.* 2005; McConnell *et al.* 2001; Smith and Long, 2010) and transcriptome analyses (Hofmann *et al.* 2019). Consistent with their up-regulation in *dcl1-5* embryos, PHB, CNA and PHV transcripts had increased signals throughout embryos, including ectopic localization in the basal and peripheral regions of *dcl1-5* embryos (Figure 6C).

Together with the observation that miR165/166 and its target transcripts had opposite localization patterns in heart-staged wild-type embryos (Figure 3C; Figure 6C), the ectopic

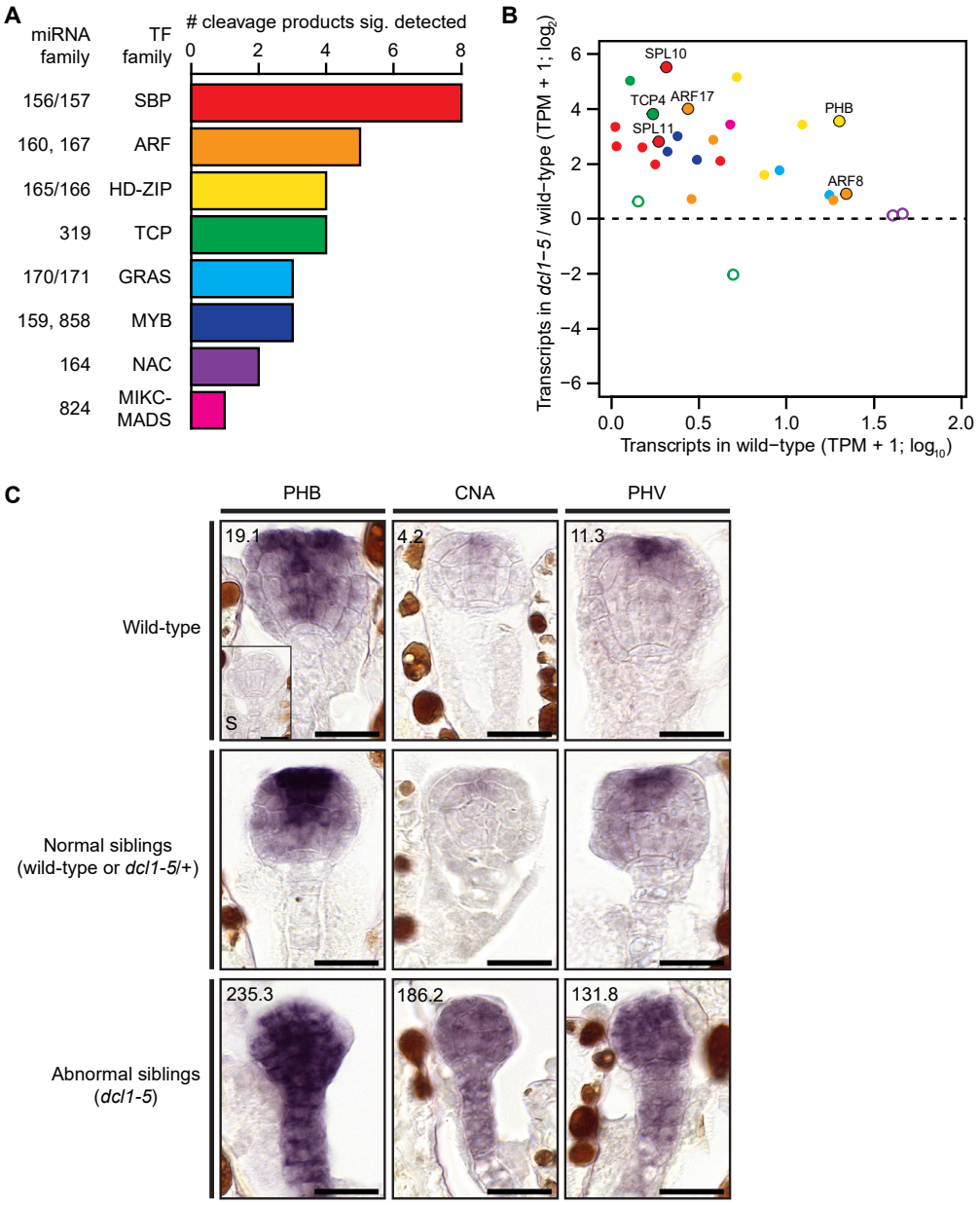


Figure 6. miRNA-Mediated Cleavage and Repression of Transcripts Encoding Transcription Factors. (A) Stacked bar chart indicating the number of transcription factor family members for which high-confidence cleavage products were detected during embryogenesis. Transcription factor families, as well as the miRNA families that mediate their cleavage, are labelled. **(B)** Scatter plots illustrating the levels of transcripts encoding transcription factors for which high-confidence cleavage products were detected according to their levels in wild-type (TPM + 1; log10) and relative fold-changes in log2- transformed (*dcl1-5* TPM + 1)/(wild-type TPM + 1). Transcripts with >1 TPM in either wild-type or *dcl1*-5 embryos are shown (n = 28). Significantly increased transcripts (n = 24; DESeq2 Benjamini-Hochberg adjusted P-values < 0.01; DESeq2) are indicated by points filled with colors representing various transcription factor families, as shown in **A**. Six targets selected for further analyses are labelled directly above each corresponding point and are also indicated with outlines. **(C)** Representative images of RNA *in situ* hybridizations with probes antisense to either PHB (left), CNA (middle) or PHV (right) transcripts on sections of embryos. RNA *in situs* were performed on embryos from either self-pollinated wild-type (top) or *dcl1-5*/+ plants. Embryos from self-pollinated *dcl1-5*/+ plants were further classified into either normal (middle, wild-type or *dcl1-5*/+) or abnormal (bottom, *dcl1-5*) siblings based on morphology. A sense control for PHB (S) is displayed in the inset of the top left panel. Numbers in the top left corners of wild-type and abnormal sibling images indicate transcript levels (TPMs) determined by mRNA-Seq. Scale bars represent 20 μm.

localization of class III HD-ZIP transcripts further supports the notion that miR165/166 helps define the proper localization patterns of their target transcription factors.

miRNA-Directed Repression Across Embryonic Cell-Types

Embryonic miRNAs direct the cleavage and repression of at least thirty transcripts encoding transcription factors (Figure 4B; Figures 6A and 6B). To examine miRNA-mediated gene repression at cellular resolution, we employed a green fluorescent protein (GFP)-based miRNA sensor system (Nodine and Bartel, 2010). Either a random 21 nt non-genome matching sequence (i.e. scrambled sensor) or 20-22 nt embryonic miRNA target sites detected by nanoPARE for miR156/157 (SPL10/11), miR160 (ARF17), miR165/166 (PHB), miR167 (ARF8) or miR319 (TCP4) were included in nuclear-localized GFP constructs under the control of a ubiquitous promoter. If the miRNA mediates repression, then the sensor transgene containing the corresponding target site was expected to produce less GFP signal. As expected, based on our sRNA-seq, nanoPARE, and *dcl1-5* mRNA-seq datasets, all five miRNA sensors had reduced GFP signal compared to scrambled sensors in at least one early embryonic stage (Figure 7). Sensors for miR156/157 and miR165/166 had strongly decreased levels throughout preglobular and globular embryos. At the heart stage, miR156/157 sensors were repressed throughout embryos, and miR165/166 sensors had increased levels in apical regions. miR156/157 and miR165/166 sensor activities were generally consistent with the results of RNA *in situ* analyses (Figure 3C; Figure 6C) and previous reports (Nodine and Bartel, 2010; Miyashima *et al.* 2013; Smith and Long, 2010). However, the miR165/166 sensors used in our study were repressed in more cell types, likely due to the use of different sensor constructs. The observed increases in miR165/166 target transcript levels throughout *dcl1-5* embryos further supports the idea that miR165/166 has broad repressive domains in early embryos (Figure 6C). The levels of miR160 sensors were also reduced throughout heart stage embryos (Figure 7). Because miR160 was detectable only in the provasculature by *in situ* hybridizations on sections of heart stage embryos (Figure 3C), the sensor approach appears to have better sensitivity than *in situ* hybridization. In fact, miR160 was detected throughout late-stage embryos when performing a more sensitive whole-mount *in situs* (Ghosh Dastidar *et al.* 2016). miR167 and miR319 sensors were weakly repressed in preglobular embryos and exhibited dynamic patterns thereafter (Figure 7). At the globular stage, miR167 sensor signals were reduced at the base of the suspensor and progressively decreased acropetally. By the heart stage, miR167 sensors had reduced signals throughout the suspensor and base of the embryo proper, as well as in the shoot meristem precursors. miR319 sensors were weakly repressed throughout globular-staged embryos, and exhibited stronger repression in the basal regions of heart-staged embryos. Altogether, the sensor dynamics support the nanoPARE and *dcl1-5* mRNA-seq datasets, and indicate that miRNAs can differentially mediate the cleavage and repression of targets, including those encoding transcription factors, across early embryonic cell types and developmental stages.

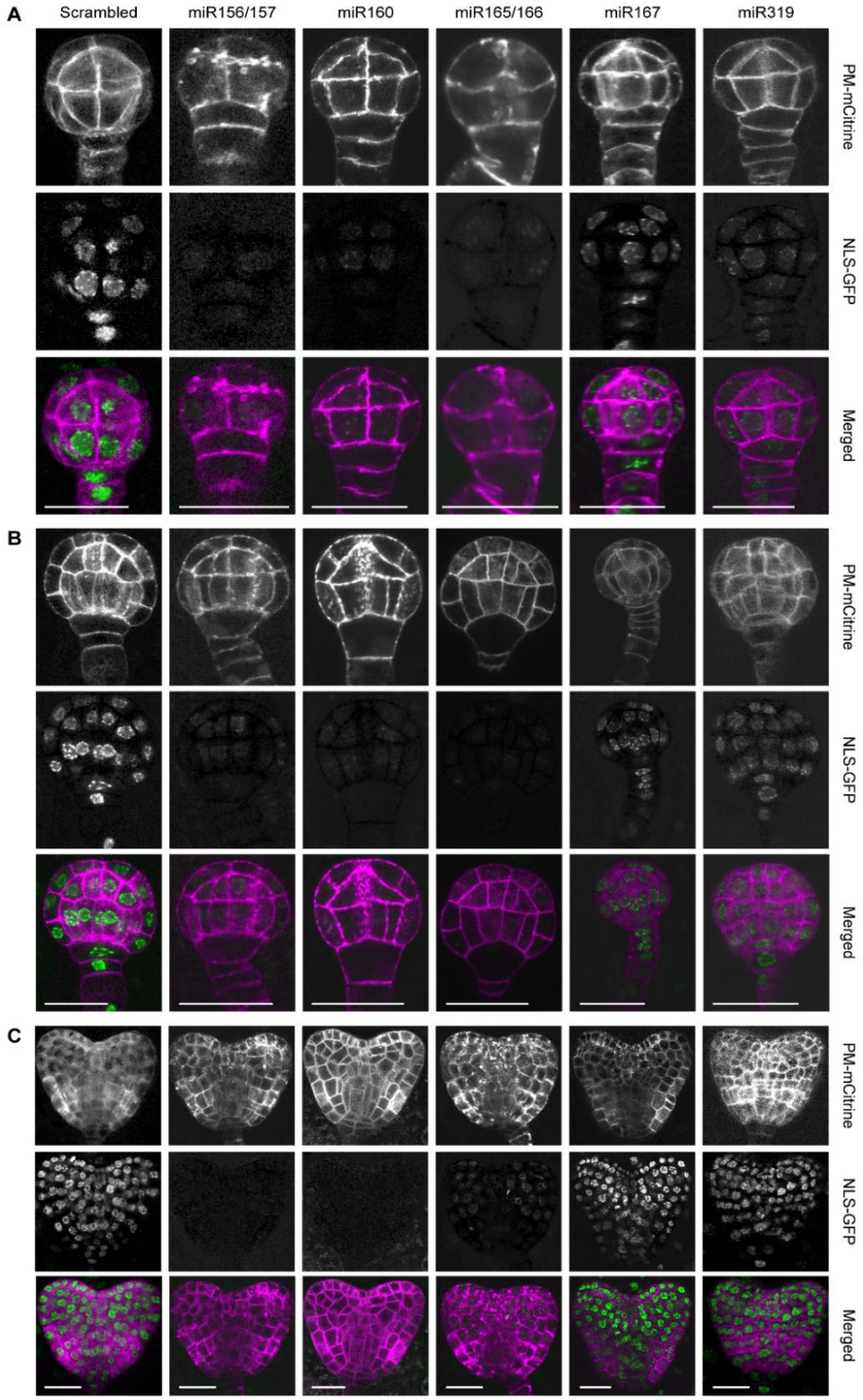


Figure 7. miRNA-Directed Repression Across Embryonic Cell Types.
Representative confocal microscopy images of preglobular **(A),** globular **(B)** and early heart **(C)** staged Col-0 embryos expressing a plasma membrane-localized mCitrine fluorescent protein under the control of the embryo-specific WOX2 promoter (pWOX2:mCitrine-SYP122) and ubiquitously expressing transcripts encoding nuclear-localized HTA6-GFP (pUBI3:HTA6-GFP) with 20-22 nt sequences that are either a non-genome matching random sequence (scrambled) or target sites detected in embryos with nanoPARE for SPL10/11 (miR156/157), ARF17 (miR160), PHB (miR165/166), ARF8 (miR167) or TCP4 (miR319), as labelled at the top. Images with signal corresponding to plasma-membrane localized pWOX2:mCitrine-SYP122 (PM-mCitrine) and pUBI3:HTA6-GFP (NLS-GFP) are indicated, as well as images from the merged channels color-coded in magenta and green, respectively. Scale bars represent 20 µm. Brightness and contrast were uniformly adjusted for each image using Zeiss ZEN software's Best Fit tool (see Methods).

miRNA-Mediated Repression of Transcription Factors is Required for Embryo Morphogenesis

The dynamic repressive activities directed by miRNAs help define the spatiotemporal domains of transcription factors and likely have a large impact on embryonic transcriptional regulatory networks, including those that help define the future plant body plan (Figures 4 to 7) (Nodine and Bartel, 2010; Seefried *et al.* 2014). To determine how miRNA-mediated cleavage of transcripts encoding transcription factors contributes to embryo morphogenesis, we selected six miRNA:target interactions to investigate further, including miR156/157:SQUAMOSA PROMOTER BINDING PROTEIN-LIKE10 (SPL10), miR156/157:SPL11, miR160:AUXIN RESPONSE FACTOR17 (ARF17), miR165/166:PHB, miR167:ARF8, and miR319:TCP4 (Figure 6B; Figure 8A). To generate target transgenes resistant to miRNA-mediated cleavage (i.e. resistant targets), we cloned these six target loci, including upstream and downstream intergenic sequences with endogenous cis-regulatory elements (4.9-7.8 kb depending on the locus), and introduced 2-7 synonymous mutations into the corresponding miRNA target sites using site-directed mutagenesis to abolish miRNA binding (Figure 8A). Because phenotypes resulting from miRNA-resistant transgenes should be interpreted with caution (Li and Millar, 2013), we performed the following experiments to control for transgene-induced artifacts. To control for potential effects unrelated to the disruption of the miRNA binding sites, we generated constructs for each resistant target whereby the corresponding miRNA binding sites were left intact (i.e. genomic targets) (Figure 8A). We transformed the resistant and genomic target constructs into wild-type plants and selected 6-17 independent transgenic lines for each construct. The post-embryonic phenotypes of resistant target lines were consistent with previous reports (Wu *et al.* 2009, 2006; Mallory *et al.* 2005) (Supplemental Figure 8A and 8B). To select representative lines for miRNA-resistant targets and their controls for further characterization, we performed quantitative RT-PCR (qRT-PCR) on target transcripts in floral bud total RNA from wild-type or their respective genomic or resistant target lines (113 total independent lines; Figure 8B). Significantly higher transcript levels were observed in all sets of resistant lines relative to wild-type and corresponding genomic target control lines (P-values < 0.001 and 0.05, respectively; t-tests) (Figure 8B). Based on the qRT-PCR analyses, we selected at least two representative lines for each construct to examine for embryo abnormalities (26 total independent lines plus wild-type; Figure 8B). The selection of multiple representative miRNA-resistant and control lines based on qRT-PCR together with miRNA-resistant lines phenocopying previously reported post-embryonic phenotypes upon miRNA knock-down strongly suggests that the phenotypes exhibited by the miRNA-resistant target lines used in this study are due to abolishing miRNA target sites rather than transgene-induced artifacts. Indeed, although we observed a modest increase in target transcript levels in control lines based on qRT-PCR, only lines with mutations in miRNA binding sites exhibited significantly

increased numbers of abnormal phenotypes (Figure 8C; Supplemental Figure 8C).

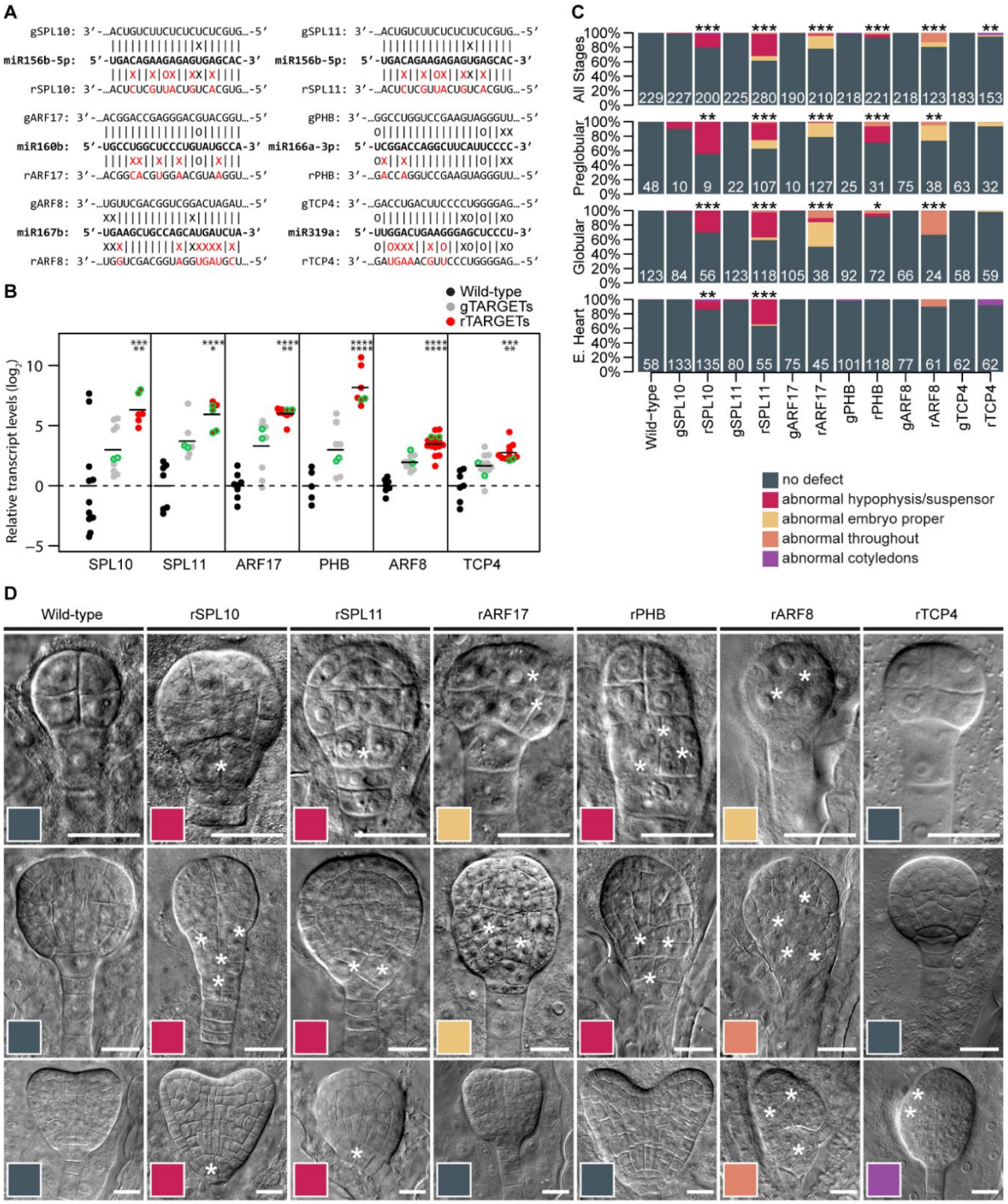


Figure 8. miRNA-Mediated Repression of Transcription Factors is Required for Embryo Morphogenesis. (A) Schematics of six miRNA target sites in transcripts encoding transcription factors selected for mutagenesis. The dominant miRNA isoforms in globular stage embryos for each family are shown. Base-pairing interactions with either wild-type target sites (genomic, gTARGET) or miRNA-resistant target sites (resistant, rTARGET) are indicated above and below, respectively. Mutations introduced by site-directed mutagenesis are labelled in red. Watson-Crick base-pairing (I), non-base-pairing (X) and G:U wobbles (O) for each pair are indicated. **(B)** miRNA target transcript levels in flowers from wild-type plants, or plants expressing either wild-type (gTARGET) or miRNA-resistant (rTARGET) versions of the target transcripts. qRT-PCR values were internally normalized with transcript levels from the eIF4A1 housekeeping gene and then divided by the levels observed in wild-type plants and log2-transformed.

Figure 8 (continued). Each dot represents the mean of two technical replicates from an independent transgenic line (first generation; T1 lines) and is color-coded according to the key. Horizontal bars represent the means between all lines. Asterisks indicate whether the transcript levels observed in rTARGET lines were significantly different compared to either wild-type (top) or gTARGET controls (bottom) (two-tailed Student's t-tests; ****, ***, ** and * represent P-values < 0.0001, P-values < 0.001, P-values < 0.01 and P-values < 0.05, respectively). Points corresponding to lines selected for phenotypic analyses are outlined in blue. **(C)** Stacked bar plot illustrating the proportions of phenotypes observed for embryos from self-pollinated wild-type plants, or T1 plants expressing either wild-type (gTARGET) or miRNA-resistant (rTARGET) versions of target transcripts in all stages examined, or in preglobular, globular or early heart (E. Heart) stages as indicated. Genotypes are labelled at the bottom, numbers at the base of each bar indicate the number of embryos examined, and phenotypes are color-coded according to the key. Asterisks indicate whether the number of abnormal embryos was significantly greater than for wild-type (one-tailed Fisher's exact test; ***, ** and * represent P-values < 0.0001, P-values < 0.001 and P-values < 0.01, respectively). **(D)** Representative Nomarski images of embryos from wild-type plants, or expressing miRNA-resistant targets at the preglobular (top), globular (middle) or early heart (bottom) stages. Genotypes are labelled above, and asterisks indicate abnormal cell divisions. Individual panels are color-coded according to the key in **C**. Scale bars represent 20 μm. Brightness and contrast were adjusted in Photoshop uniformly for each image to improve visibility (see Methods).

We phenotyped 2,682 preglobular to heart stage embryos from the 26 representative independent transgenic lines and wild-type. Distinct morphological defects were observed for each miRNA-resistant target (Figure 8C), and these were reproducible among independent transgenic lines (Supplemental Figure 8C). Consistent with our previous report, miR156/157-resistant SPL10 and SPL11 transgenic embryos exhibited abnormal divisions in their uppermost suspensor and hypophysis cells during early embryogenesis and failed to generate lens-shaped cells (i.e. precursors to the root quiescent center) (Figure 8D). Embryos with miR160-resistant ARF17 constructs had abnormal divisions in the embryo proper beginning at the preglobular stage when, for example, the protoderm layer failed to form on one side of the embryo (Figure 8D). Later during the globular stage, embryos with miR160-resistant ARF17 transgenes had abnormal divisions in the basal region of the subprotoderm. Preglobular and globular embryos with miR165/166-resistant PHB transgenes had abnormal divisions in the hypophysis (Figure 8D). Embryos from self-pollinated miR160-resistant ARF17 and miR165/166-resistant PHB lines appeared normal at the heart stage, but this may be due to lethality or developmental delay of embryos expressing the miRNA-resistant transgenes. Preglobular embryos with miR167-resistant ARF8 transgenes had defective divisions in the embryo proper, and then throughout globular and heart embryos (Figure 8D). Embryos with miR319-resistant TCP4 transgenes were morphologically normal during the preglobular and globular stages, but at the heart stage had a low, but significantly increased, number of embryos with defective cotyledon outgrowth (Figure 8D). Because most of the miRNA-resistant target lines were sterile and had abnormal flower morphologies (Supplemental Figure 8B), we crossed wild-type plants (as the female parent) with plants that were either wild-type or transgenic for miRNA-resistant constructs to determine whether the phenotypes observed from self-pollinated miRNA-resistant target lines were due to zygotic or maternal sporophytic effects of the transgenes. All crosses between wild-type maternal plants and resistant target lines yielded significantly more abnormal embryos than wild-type

crosses (P-values < 0.01; Fisher's exact test), and the progeny exhibited similar phenotypes to the self-pollinated resistant targets described above (Supplemental Figures 8D and 8E). Therefore, specific interactions between miRNAs and transcripts encoding transcription factors are morphologically required in a variety of embryonic cell types.

6.4 Discussion

We developed a sRNA-seq library preparation protocol that highly enriches for sRNAs, as well as reproducibly and accurately measures their levels from low amounts of total RNA (i.e. ≥1 ng). We expect this method to be useful for profiling sRNA populations from difficult-to-obtain samples, including plant and animal reproductive tissues. Here, we used this low-input sRNA-seq method to profile sRNA populations across *Arabidopsis* embryogenesis. The sRNA-seq and nanoPARE datasets reported in this study, as well as the transcriptome datasets produced from the same stages (Hofmann *et al.* 2019), provide a solid foundation for the characterization of non-coding and coding RNA populations in plant embryos. For example, miRNAs comprise only a fraction of the embryonic sRNA population, and these integrated sRNA-seq, nanoPARE and mRNA-seq datasets will also enable the systematic characterization of additional sRNAs present in early embryos, including small interfering RNAs involved in the establishment of epigenetic marks and associated transcriptional gene silencing events. In this study, we generated a catalog of miRNAs present during embryogenesis and applied our recently developed nanoPARE method to identify 59 high-confidence embryonic miRNA targets. We found high-confidence miRNA-directed cleavage products for only 22/115 detected embryonic miRNAs, suggesting that many miRNAs may not be directing target cleavage in the stages and conditions examined. Although this could be partially explained by the limited sensitivity of the nanoPARE method, our observation that targets detected by nanoPARE, but not those confidently predicted computationally, had globally increased transcript levels in *dcl1-5* embryos suggests that we have identified the majority of cleavage events. Moreover, we detected miRNA-directed cleavage products of all targets with published evidence supporting their existence during *Arabidopsis* embryogenesis (Knauer *et al.* 2013; Takanashi *et al.* 2018; Palatnik *et al.* 2003; Nodine and Bartel, 2010; Smith and Long, 2010; Miyashima *et al.* 2013; Mallory *et al.* 2005). We propose that many of the detected miRNAs function as fail-safes to prevent the aberrant accumulation of target transcripts or have already executed their functions during earlier stages of development. For instance, we were unable to detect targets for any of the five miRNA families that were abundant and enriched at the earliest stages of embryogenesis. The levels of these miRNAs decreased dramatically during early embryogenesis, and they may function directly after fertilization and prior to the earliest stage profiled with nanoPARE (i.e. preglobular or 8-cell/16-cell stage). As a proof-of-principle of this resource's utility, we focused on the miRNA-mediated regulation of transcription factors in the current study. We and others have demonstrated that miRNAs are required for pattern formation and proper

developmental timing of gene expression programs during *Arabidopsis* embryogenesis (Nodine and Bartel, 2010; Seefried *et al.* 2014; Willmann *et al.* 2011). Indeed, the more comprehensive *dcl1-5* embryo transcriptome dataset and analyses presented here further supports the concept that miRNAs have a large impact on the embryonic transcriptome, including the prevention of precocious expression of genes characteristic of the maturation phase of embryogenesis and related to oil body biogenesis, lipid storage, and other seed maturation processes (Hofmann *et al.* 2019). Because <5% of the transcripts whose levels significantly increased in *dcl1-5* embryos were directly cleaved and repressed by miRNAs, the vast majority of misregulated genes in *dcl1-5* embryos are likely downstream of miRNA targets. Interestingly, 30 of the 59 embryonic miRNA high-confidence targets identified encoded transcription factors. Their de-repression in miRNA-deficient *dcl1-5* embryos, along with associated misregulated downstream transcriptional cascades, may largely explain why thousands of transcripts are present at different levels in *dcl1-5* compared to wild-type embryos. Together with previous studies, our results indicate that multiple miRNAs are required for embryo morphogenesis and pattern formation. We previously demonstrated that miR156/157 prevents the accumulation of SPL transcription factors and the resulting expression of maturation phase genes during early embryogenesis (Nodine and Bartel, 2010). Although decreased SPL10/11 levels could suppress miRNA-deficient *dcl1-5* phenotypes, abolishing miR156/157:SPL10/11 interactions was not sufficient to fully phenocopy *dcl1-5* embryos. This suggested that additional miRNA:target interactions are required for embryonic pattern formation. Accordingly, the hypophysis and suspensor division defects observed in embryos expressing miR156/157-resistant SPL10/11 and miR165/166-resistant PHB, as well as the embryo proper defects of miR160-resistant ARF17 embryos, were both observed in *dcl1-5* embryos (Figure 8D) (Nodine and Bartel, 2010). Moreover, the pleiotropic defects exhibited by miR167-resistant ARF8 embryos and cotyledon initiation defects observed in miR319-resistant TCP4 further support the idea that multiple miRNAs are required for proper embryo morphogenesis. Interestingly, preglobular stage miR160-resistant ARF17, miR165/166-resistant PHB, and miR167-resistant ARF8 embryos often exhibit more severe defects than we observed in *dcl1-5* (Figure 8D) (Nodine and Bartel, 2010). Because homozygous *dcl1-5* embryos are lethal, they are derived from *dcl1-5*/+ parents. Therefore, it is possible that the DCL1 gene products inherited from diploid sporocytes are sufficient to produce miRNAs in gametophytes or early embryos, as previously proposed for other essential genes (Muralla *et al.* 2011). Additionally, redundant activities of other DCL genes may partially compensate for the loss of DCL1 in preglobular embryos. Consistent with both of these explanations, miR165/166 and miR167 levels were highly reduced, but not eliminated, in globular stage *dcl1-5* embryos (Supplemental Figure 3; Supplemental Data Set 2).

The developmental progression of miRNA-resistant target phenotypes generally corresponds well with the spatiotemporal dynamics of the corresponding miRNAs and their activities. miRNA-resistant transgenes generally caused phenotypes in the same cell-types in which the corresponding miRNAs were active (Figure 7; Figure 8D). One exception was the defective cotyledons observed in miR319-resistant TCP4 embryos. Although cotyledon initiation occurred at the heart stage when miR319 activities were increased (Figure 4B), and the cotyledon defects were in agreement with previously reported seedling phenotypes (Palatnik *et al.* 2003), miR319 was more active in basal regions of embryos (Figure 7).

Therefore, gene-regulatory processes downstream of miR319-mediated repression of TCP4 may be non-cell autonomously required for cotyledon formation. By contrast, miR160-resistant ARF17 and miR165/166-resistant PHB exhibited abnormal phenotypes in the cell types in which their highest levels or repressive activities were detected. For example, miR160-resistant ARF17 had defects in the subprotoderm of the embryo proper, which is congruent with higher miR160 levels in these cell types. Together with previously reported phenotypes of embryos with mir160a loss-of-function mutations (Liu *et al.* 2010), our results indicate that miR160-mediated repression of the ARF17 transcription factor is required for proper sub-protodermal cell division patterns.

The observation that miR165/166-mediated repression of target HD-ZIP transcripts occurs in basal embryonic regions indicates that miR165/166 helps define HD-ZIP transcription factor localization domains in early embryos (Miyashima *et al.* 2013; Smith and Long, 2010; McConnell *et al.* 2001) (Figure 4B; Figure 6; Figure 7). Accordingly, embryos expressing miR165/166-resistant PHB exhibited abnormal divisions typically in basal regions of the embryo (Figures 8C and 8D), indicating that miR165/166:PHB interactions are required in these cell types for proper morphogenesis. miR167 and its repression of ARF8 are required in the maternal sporophytic tissues for proper embryogenesis (Wu *et al.* 2006; Yao *et al.* 2019). Interestingly, plants containing miR167-resistant ARF8 transgenes had similar phenotypes when crossed as the pollen donors to wild-type maternal lines. This indicates that miR167-mediated repression of ARF8 is required in embryos for proper morphogenesis, which is similar to the five other miRNA:target interactions we characterized (Supplemental Figures 8D and 8E).

Altogether, these data support a model whereby the post-transcriptional regulation of transcription factor gene-regulatory networks by several miRNAs is critically important for the establishment of the plant body plan during early embryogenesis. The resources and phenotypes described in this study provide multiple entry points to further characterize how the miRNA-mediated repression of transcripts, including those encoding transcription factors, contributes to the initial cellular differentiation events that operate at the beginning of plant life.

6.5 Methods

Plant Material, Growth Conditions and RNA isolation

The *dcl1-5* (McElver *et al.* 2001) and *xrn4-5* (Souret *et al.* 2004) alleles in the *Arabidopsis thaliana* Col-0 accession background, together with Col-0, were grown in a climate-controlled growth chamber at 20˚C-22˚C under a 16h light/8h dark cycle. Plants were grown under incandescent lights at 130-150 μmol/m2/s. Embryos were dissected and total RNA was extracted at a similar time of day (13:00–17:00) as described previously (Hofmann *et al.* 2019). Except for the bent-cotyledon stage samples, all other total RNA samples pooled from 50 Col-0 embryos were used to generate mRNA-seq datasets (Hofmann *et al.* 2019) and the sRNA-seq and nanoPARE datasets reported in this study. Total RNA from 7-day-old *xrn4-5* seedling roots and shoots grown vertically on 0.5× MS plates were isolated as previously described (Schon *et al.* 2018). In this study, biological replicates of mRNA-seq, nanoPARE and mRNA-seq datasets were from pools of RNA collected from different embryos, leaves, flowers, roots, or shoots on different days.

Low-input sRNA-seq

18-30-nt RNAs were purified from ≥80% of the total RNA from each sample (from 50 pooled embryos) using denaturing polyacrylamide-urea gels as described previously (Grimson *et al.* 2008). Size-selected sRNAs were precipitated overnight at -20˚C with 2.5× volumes of ice-cold 100% ethanol and 1 μl of GlycoBlue (Thermo Fisher) and resuspended in 7.5 μl of nuclease-free water. This sample was used as input for the NEBNext Multiplex Small RNA Library Prep Set for Illumina kit (NEB #E7300) according to the manufacturer's recommendations with the following modifications. Adapters used for 3′ and 5′ ligations to sRNAs, and SR RT primers to generate sRNA cDNAs, were diluted to 25% of the amounts recommended for ≥500 ng of total RNA. Various numbers of PCR cycles were used to amplify cDNAs: 14, 16, 18 and 20 PCR cycles for early heart and later staged samples, and 18, 20, 22, and 24 PCR cycles for globular and earlier staged samples. Final amplicons were run on a 90% formamide/8% acrylamide gel at 5W for ~30-minutes, followed by 30W for ≥2 hours, and stained with SYBR Gold (1:10,000; Thermo Fisher). Fluorescence intensities of amplicons were examined across the PCR cycles, 137–149-bp products (corresponding to 18-30-nt sRNAs) with non-saturated signals were gel-purified, and after DNA precipitation, pellets were resuspended in 15 μl of Elution Buffer (Qiagen). To control for library quality, sRNA-seq libraries were examined on an Agilent DNA HS Bioanalyzer Chip, and those with the expected size range were sequenced on an Illumina HiSeq 2500 instrument in 50 base single-end mode (Supplemental Data Set 1). Cutadapt (Martin, 2011) was used to trim adapter sequences from sRNA-seq reads, and 18–30-base sequences that contained an adapter were retained. The trimmed sequences were aligned to the *Arabidopsis thaliana* TAIR10 genome (Lamesch *et al.* 2012) with STAR (Dobin *et al.* 2013), requiring no mismatches and allowing ≤100 multiple end-to-end alignments. The resulting SAM files were processed with the

readmapIO.py script to re-assign multimappers with a "rich-get-richer" algorithm as previously described (Schon *et al.* 2018). Output bedFiles were sorted, condensed, and normalized for total genome-matching reads. The BEDtools map function (Quinlan and Hall, 2010) was used to quantify the number of reads mapping to the same strand and overlapping ≥80% of mature miRNAs as annotated in TAIR10 and miRBase (Kozomara *et al.* 2019). Statistical analyses and associated figures were generated with the R statistical computing package (R Core Team, 2018).

nanoPARE and mRNA-seq

For transcriptome analyses, Smart-seq2 libraries (Picelli *et al.* 2013) were generated from *dcl1-5* embryos selected from self-fertilized *dcl1-5/+* plants based on their abnormal morphologies as previously described (Hofmann *et al.* 2019). These were sequenced on an Illumina HiSeq 2500 instrument in 50 base paired-end mode (Supplemental Data Set). Transcriptome analyses were performed as described (Hofmann *et al.* 2019), except that TAIR10 transposable element gene models were also included in the Kallisto-based pseudo-alignments (Bray *et al.* 2016). DESeq2 (Love *et al.* 2014) with default settings was used to compute P-values for the wild-type and *dcl1-5* transcriptome comparisons.

nanoPARE libraries presented in this study were generated as previously described (Schon *et al.* 2018) with the following exceptions. For all embryonic samples other than those from the bent-cotyledon stage, the same cDNA pools used in our previous transcriptome analysis (Hofmann *et al.* 2019) were also used as input for nanoPARE library preparation. The nanoPARE libraries were sequenced on an Illumina Hi-Seq 2500 instrument in 50 base single-end mode (Supplemental Data Set 1). Analysis of the nanoPARE data was performed as described (Schon *et al.* 2018), except that all capped features identified in the embryonic series were merged with those from published floral bud samples (Schon *et al.* 2018) and used to mask capped features from the transcript-level bedGraph files. Additionally, we used EndCut (Schon *et al.* 2018) to test for significant target sites for 164 miRNAs detected ≥1 RPM in at least one embryonic stage. nanoPARE libraries from all post-embryonic tissues were analyzed in an identical manner.

RNA *in situ* Hybridizations

miRNA *in situs* on embryo sections were performed based on a whole-mount *in situ* hybridization method (Ghosh Dastidar *et al.* 2016). Sample preparation leading up to probe hybridization was performed as described (Nodine *et al.* 2007), except that a LOGOS Microwave Hybrid Tissue Processor (Milestone Medical) was used for tissue embedding, and the samples were fixed with EDC solution (MN-(3- Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride in Methylimidazole-NaCl) after the proteinase K digestion step as follows. First, slides with adhered embryo sections were transferred to 1× PBS and washed 2×, and then incubated in a staining dish containing freshly prepared methylimidazole-NaCl for 10 minutes at room temperature (2×). The slides were then transferred to EDC solution and incubated for 2 hours at 60°C, and subsequently washed 2×

in 1× PBS for 5 minutes each prior to probe hybridization. Dual DIG-labelled LNA-modified oligos antisense to miR124, miR156a-f, miR159a, miR160a-c, or miR166a-f isoforms were used at a final concentration of 20nM (Supplemental Table 1), and the rest of the probe hybridization procedure, as well as subsequent washing, antibody, and colorimetric reactions were as described (Nodine *et al.* 2007). Slides were imaged on an automated Pannoramic SCAN 150 slide scanner (3DHISTECH) and collected with the associated Pannoramic Viewer software. Images of ≥50 embryos from >5 independent sets of experiments were recorded. mRNA *in situs* were performed as previously described (Nodine *et al.* 2007). Probes antisense to CNA, PHB, and PHV were generated from cDNAs by introducing T7 promoters via PCR as described previously (Hejátko *et al.* 2006) (Supplemental Table 1).

Generation of Transgenic Lines

Nuclear-localized GFP-based sensor constructs with miR156/157:SPL10/11 target sites (GTGCTCTCTCTCTTCTGTCA) in the 5′ UTR and under the control of the potato (*Solanum tuberosum*) UBI3 promoter were generated as previously described (Nodine and Bartel, 2010). A similar strategy was also employed to create constructs with the miR160:ARF17 (TGGCATGCAGGGAGCCAGGCA), miR165/166:PHB (TGGGATGAAGCCTGGTCCGG), miR167:ARF8 (TTAGATCAGGCTGGCAGCTTGT), and miR319:TCP4 (AGAGGGGTCCCCTTCAGTCCAG) target sites detected in embryos with nanoPARE. As a negative control, we also generated identical constructs except with a random 21-nt sequence (CCCCGTCTCGCGTCTCACGCA) that does not map to the *Arabidopsis* genome. Constructs were transformed into Col-0 plants harboring non-segregating transgenes for the mCitrine fluorescent protein fused to plasma membrane-localized SYP122 protein under the control of an embryo-specific WOX2 promoter (Breuninger *et al.* 2008) (pWOX2:mCitrine-SYP122) using the Agrobacterium floral dip method (Clough and Bent, 1998). At least two sensor lines were examined in the T1 and T2 generations.

Control genomic and miRNA-resistant SPL10 and SPL11 constructs were generated as previously described (Nodine and Bartel, 2010). For control genomic ARF17 (gARF17), PHB (gPHB), and TCP4 (gTCP4) transgenic constructs, target loci including upstream and downstream intergenic sequences were PCR amplified from Col-0 genomic DNA with primers containing overhangs for subsequent Gibson assembly. miR160-resistant ARF17 (rARF17), miR165/166-resistant PHB (rPHB), and miR319-resistant TCP4 (rTCP4) constructs were amplified as two separate fragments with overlaps to introduce specific mutations in the corresponding miRNA target sites. The backbones of the MultiSite- Gateway destination vectors pAlligatorG43 and pAlligatorR43 (Kawashima *et al.* 2014) were amplified for subsequent Gibson assembly, and genomic and resistant ARF17, PHB, and TCP4 plant transformation constructs were generated by Gibson Assembly (NEB) using the pAlligatorG43/R43 backbone and the target PCR fragments described above. For the control genomic ARF8 transgenic construct (gARF8), the ARF8 locus including upstream and

downstream intergenic sequences was PCR amplified from Col-0 genomic DNA and cloned into the pENTR/D-TOPO Gateway vector (Thermo Fisher). The miR167- resistant ARF8 construct (rARF8) was generated by PCR site-directed mutagenesis (NEB) of the gARF8 entry clone. Final plant transformation constructs were generated by Gateway LR reactions (Thermo Fisher) with pENTR-gARF8 or pENTR-rARF8, pDONR-L4R1-empty, and pDONR-R2L3-empty, and the Gateway destination vector pAlligatorR43 (RFP). All primers are listed in Supplemental Table 1. The constructs were transformed into Col-0 as described above, and transformants were selected based on GFP or RFP selection marker fluorescence from pAlligatorG43/R43 (Kawashima *et al.* 2014; Bensmihen *et al.* 2004).

qRT-PCR Analysis

Two clusters of floral buds were pooled from 8-week-old plants, snap-frozen in liquid nitrogen, homogenized using a Mixer Mill MM 400 (Retsch) for 30 seconds with max amplitude, and resuspended in 200 µl TRIzol (Life Technologies). Total RNA was extracted using a Direct-zol kit (Zymo Research) according to the manufacturer's instructions, and DNaseI treatment was performed on-column. Total RNA quality and quantity were determined with an Agilent Fragment Analyzer (AATI) using a standard RNA sensitivity kit (DNF-471). 200 ng of total RNA samples with RNA Quality Number (RQN) values >6.0 were used for cDNA synthesis together with the Oligo d(T)18 mRNA Primer (NEB) and SuperScript III Reverse Transcriptase (Thermo Fisher). The cDNA was diluted 10× with nuclease-free water, and 2 µl was used as a template for the qRT-PCR. qRT-PCR was performed on a LightCycler 96 Instrument (Roche) using gene-specific and control elF4A primers (Supplemental Table 1), and Fast SYBR™ Green Master Mix (Roche). Ct values were obtained using LightCycler 96 software, and relative quantification of transcripts (ΔΔCt values) was performed with an in-house R script. For each genotype, 6-17 individual first-generation transgenic (T1) lines were analyzed in technical duplicates.

Microscopy

Self-pollinated siliques from at least two representative and independently generated first-generation transgenic (T1) lines for each miRNA-resistant and control constructs were harvested, and ovules were fixed and cleared in a solution composed of 8 g chloral hydrate, 1 ml water, and 1 ml glycerol as described previously (Ohad *et al.* 1996). At least two representative T2 lines were also crossed as pollen donors to emasculated Col-0 flowers, and siliques were harvested 120 hours after pollination. Embryos were examined with Nomarski optics on a ZEISS Axio Observer Z1 with a sCMOS camera. Images were acquired using ZEISS ZEN (blue edition) imaging software and analyzed using ImageJ/Fiji processing software. To minimize potential bias, Nomarski images were examined by a person that did not acquire the images, and phenotypes were recorded prior to revealing sample identities. At least 38 embryos from 26 independent transgenic lines were examined for each construct (2,682 total embryos). For confocal microscopy, whole seeds containing preglobular, globular, or early

heart staged embryos were harvested, mounted in VectaShield antifade mounting medium (Vector Laboratories), and imaged directly on a Zeiss LSM 780 Axio Observer using a 488 nm excitation wavelength for both GFP and mCitrine; images were acquired using the same settings. The images were recorded and analyzed on Zeiss ZEN imaging software (black edition): spectral unmixing was performed to differentiate between emission spectra of GFP (~450-500nm) and mCitrine (~550-600nm), and contrast and brightness were uniformly adjusted using the Best Fit tool of Zeiss ZEN imaging software. All images were cropped and rotated in Photoshop (Adobe). To increase the resolution and uniformity of the image panels, Nomarski images were further processed in Photoshop by applying the following tools to the whole image: Image/Adjustments/Levels/Midtones Brighter adjustments, Auto Contrast adjustments, and the Unsharp Mask filter.

Accession Numbers, Data Acquisition, and Code Availability

All sequencing data generated in this study are available at the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE132066. Publicly available next-generation sequencing data were downloaded from the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) with the following accession numbers: Col-0 sRNA-seq (GSE79414 and GSE98553), Col-0 mRNA-seq (GSE121236) and nanoPARE (accession number GSE112869). Custom software used to align sRNA-seq data to annotated mature miRNAs is available on GitHub (https://github.com/Gregor-Mendel-Institute/Plotnikova.2019).

Supplemental Data

Supplemental Figure 1. Establishment of Low-input Small RNA Sequencing Method

Supplemental Figure 2. Principal Component Analysis of Embryonic and Post-Embryonic miRNA Populations

Supplemental Figure 3. Heat Map With miRNA Levels Normalized By Reads Per Million Genome-Matching Reads

Supplemental Figure 4. Embryo-Enriched miRNAs

Supplemental Figure 5. mRNA 5′ Ends of miRNA Targets in *dcl1-5* Mutant Embryos

Supplemental Figure 6. miRNA and miRNA-Mediated Cleavage Product Correlations

Supplemental Figure 7. Tissue-Enrichment Test of Wild-Type and *dcl1-5* Mutant Embryo Transcriptomes

Supplemental Figure 8. Control Experiments For Analysis of miRNA-Resistant Targets

Supplemental Table 1. Oligonucleotides Used in This Study

Supplemental Data Set 1. Summary of High-Throughput Datasets Generated In or Reanalyzed for This Study

Supplemental Data Set 2. Levels of miRNAs Detected During Embryogenesis

Supplemental Data Set 3. Predicted miRNA Cleavage Sites Detected in nanoPARE Datasets

Supplemental Data Set 4. Normalized Transcript Levels in Wild-type and miRNA-Deficient *dcl1-5* Globular Embryos

6.6 Acknowledgments

We thank the VBCF NGS Unit for sequencing sRNA-seq, nanoPARE and mRNA-seq libraries, the VBCF Plant Sciences Facility for plant growth chamber access, and the IMP- IMBA-GMI BioOptics Core Facility for microscopy access and support. We also thank Joop Vermeer, Yuree Lee and Niko Geldner for reagents and Ralf Jansen for technical assistance. This work was supported by funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant 637888 to M.D.N.) and the DK Graduate Program in RNA Biology (DK-RNA) sponsored by the Austrian Science Fund (FWF, DK W 1207-B09).

6.7 Author contributions

Conceptualization, M.D.N.; Methodology, M.D.N., A.P., M.J.K. and M.M.; Software, M.D.N. and M.A.S.; Formal Analysis, M.D.N. and M.A.S.; Investigation, A.P., M.D.N., M.J.K. and M.M.; Writing – Original Draft, M.D.N.; Writing – Review & Editing, M.D.N., M.A.S., M.J.K. and A.P.; Visualization, M.D.N., M.A.S. and A.P.; Supervision, M.D.N.; Funding Acquisition, M.D.N.

6.8 References

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P., and Axtell, M.J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr Biol.* **18**: 758–762.

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**: 207–221.

Aukerman, M.J. and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* **15**: 2730– 2741.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.

Bensmihen, S., To, A., Lambert, G., Kroj, T., Giraudat, J., and Parcy, F. (2004). Analysis of an activated ABI5 allele using a new selection method for transgenic *Arabidopsis* seeds. *FEBS Lett.* **561**: 127–131.

Borges, F. and Martienssen, R.A. (2015). The expanding world of small RNAs in plants. Nat. Rev. *Mol. Cell Biol.* **16**: 727–741.

Bray, N., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal RNA-Seq quantification with kallisto. *Nat. Biotechnol.* **34**: 525–527.

Breuninger, H., Rikirsch, E., Hermann, M., Ueda, M., and Laux, T. (2008). Differential expression of WOX genes mediates apical-basal axis formation in the *Arabidopsis* embryo. *Dev. Cell* **14**: 867–876.

Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science* **303**: 2022–2025.

Chen, X. (2009). Small RNAs and their roles in plant development. *Annu. Rev. Cell Dev. Biol.* **25**: 21–44.

Clough, S.J. and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* **16**: 735–743.

D'Ario, M., Griffiths-Jones, S., and Kim, M. (2017). Small RNAs: Big Impact on Plant Development. *Trends Plant Sci.* **22**: 1056–1068.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Faridani, O.R., Abdullayev, I., Hagemann-Jensen, M., Schell, J.P., Lanner, F., and Sandberg, R. (2016). Single-cell sequencing of the small-RNA transcriptome. *Nature Biotechnology* **34**: 1264–1266.

Gandikota, M., Birkenbihl, R.P., Höhmann, S., Cardon, G.H., Saedler, H., and Huijser, P. (2007). The miRNA156/157 recognition element in the 3′ UTR of the *Arabidopsis* SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. *Plant J.* **49**: 683–693.

German, M.A. *et al.* (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* **26**: 941–946.

Ghosh Dastidar, M., Mosiolek, M., Bleckmann, A., Dresselhaus, T., Nodine, M.D., and Maizel, A. (2016). Sensitive whole-mount in situ localization of small RNA s in plants. *Plant J.* **88**: 694–702.

Gregory, B.D., O'Malley, R.C., Lister, R., Urich, M.A., Tonti-Filippini, J., Chen, H., Millar, A.H., and Ecker, J.R. (2008). A link between RNA metabolism and silencing affecting *Arabidopsis* development. *Dev. Cell* **14**: 854–866.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**: 1193–1197.

Hejátko, J., Blilou, I., Brewer, P.B., Friml, J., Scheres, B., and Benková, E. (2006). In situ hybridization technique for mRNA detection in whole-mount *Arabidopsis* samples. *Nature Protocols* **1**: 1939–1946.

Hofmann, F., Schon, M.A., and Nodine, M.D. (2019). The embryonic transcriptome of *Arabidopsis thaliana*. *Plant Reprod.* **32**: 77–91.

Huang, D., Koh, C., Feurtado, J.A., Tsang, E.W.T., and Cutler, A.J. (2013). MicroRNAs and their putative targets in Brassica napus seed maturation. *BMC Genomics* **14**:140-148.

Jones-Rhoades, M.W. and Bartel, D.P. (2004). Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. *Molecular Cell* **14**: 787–799.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAs AND THEIR REGULATORY ROLES IN PLANTS. *Annual Review of Plant Biology* **57**: 19–53.

Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A., and Carrington, J.C. (2003). P1/HC-Pro, a Viral Suppressor of RNA Silencing, Interferes with *Arabidopsis* Development and miRNA Function. *Dev. Cell* **4**: 205–217.

Kawashima, T., Maruyama, D., Shagirov, M., Li, J., Hamamura, Y., Yelagandula, R., Toyama, Y., and Berger, F. (2014). Dynamic F-actin movement is essential for fertilization in *Arabidopsis thaliana*. *Elife* **3**.

Knauer, S., Holt, A.L., Rubio-Somoza, I., Tucker, E.J., Hinze, A., Pisch, M., Javelle, M., Timmermans, M.C., Tucker, M.R., and Laux, T. (2013). A protodermal miR394 signal defines a region of stem cell competence in the *Arabidopsis* shoot meristem. *Dev. Cell* **24**: 125–132.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**: D155–D162.

Lamesch, P. *et al.* (2012). The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**: D1202–10.

Li, J. and Millar, A.A. (2013). Expression of a microRNA-resistant target transgene misrepresents the functional significance of the endogenous microRNA: target gene relationship. *Mol. Plant* **6**: 577–580.

Liu, X., Huang, J., Wang, Y., Khanna, K., Xie, Z., Owen, H.A., and Zhao, D. (2010). The role of floral organs in carpels, an *Arabidopsis* loss-of-function mutation in MicroRNA160a, in organogenesis and the mechanism regulating its expression. *Plant J.* **62**: 416–428.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**: 2053–2056.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**: 550.

Lutzmayer, S., Enugutti, B., and Nodine, M.D. (2017). Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-Seq data. *Sci. Rep.* **7**: 5913.

Mallory, A.C., Bartel, D.P., and Bartel, B. (2005). MicroRNA-directed regulation of *Arabidopsis* AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell* **17**: 1360–1375.

Mansfield, S.G. and Briarty, L.G. (1991). Early embryogenesis in *Arabidopsis thaliana*. II. The developing embryo. *Can. J. Bot.* **69**: 461–476.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.

McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* **411**: 709–713.

McElver, J. *et al.* (2001). Insertional mutagenesis of genes required for seed development in *Arabidopsis thaliana*. *Genetics* **159**: 1751–1763.

Miyashima, S., Honda, M., Hashimoto, K., Tatematsu, K., Hashimoto, T., Sato-Nara, K., Okada, K., and Nakajima, K. (2013). A comprehensive expression analysis of the *Arabidopsis* MICRORNA165/6 gene family during embryogenesis reveals conserved role in meristem specification and a non-cell-autonomous function. *Plant Cell Physiol.* **54**: 375–384.

Muralla, R., Lloyd, J., and Meinke, D. (2011). Molecular foundations of reproductive lethality in *Arabidopsis thaliana*. *PLoS One* **6**: e28398.

Nodine, M.D. and Bartel, D.P. (2010). MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes Dev.* **24**: 2678– 879 2692.

Nodine, M.D., Yadegari, R., and Tax, F.E. (2007). RPK1 and TOAD2 Are Two Receptor-like Kinases Redundantly Required for *Arabidopsis* Embryonic Pattern Formation. *Developmental Cell 12:* 943–956.

Ohad, N., Margossian, L., Hsu, Y.C., Williams, C., Repetti, P., and Fischer, R.L. (1996). A mutation that allows endosperm development without fertilization. *Proceedings of the National Academy of Sciences* **93**: 5319–5324.

Oh, T.J., Wartell, R.M., Cairney, J., and Pullman, G.S. (2008). Evidence for stage-specific modulation of specific microRNAs (miRNAs) and miRNA processing components in zygotic embryo and female gametophyte of loblolly pine (Pinus taeda). *New Phytol.* **179**: 67–80.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. *Nature* **425**: 257–263.

Palovaara, J., de Zeeuw, T., and Weijers, D. (2016). Tissue and Organ Initiation in the Plant Embryo: A First Time for Everything. *Annu. Rev. Cell Dev. Biol.* **32**: 47–75.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12:** 1484–1495.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**: 1096–1098.

Plasterk, R.H.A. (2006). Micro RNAs in animal development. *Cell* **124**: 877–881.

Prigge, M.J., Otsuga, D., Alonso, J.M., Ecker, J.R., Drews, G.N., and Clark, S.E. (2005). Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *Plant Cell* **17**: 61–76.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. *Genes Dev*. **16**: 1616–1626.

Schon, M.A., Kellner, M.J., Plotnikova, A., Hofmann, F., and Nodine, M.D. (2018). NanoPARE: parallel analysis of RNA 5′ ends from low-input RNA. *Genome Research* **28**: 1931–1942.

Schon, M.A. and Nodine, M.D. (2017). Widespread Contamination of *Arabidopsis* Embryo and Endosperm Transcriptome Data Sets. *Plant Cell* **29**: 608–617.

Schwartz, B., Yeung, E., and Meinke, D. (1994). Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of *Arabidopsis*. *Development* **120**: 3235–3245.

Seefried, W.F., Willmann, M.R., Clausen, R.L., and Jenik, P.D. (2014). Global Regulation of Embryonic Patterning in *Arabidopsis* by MicroRNAs. *Plant Physiol.* **165:**670–687.

Smith, Z.R. and Long, J.A. (2010). Control of *Arabidopsis* apical-basal embryo polarity by antagonistic transcription factors. *Nature* **464**: 423–426.

Souret, F.F., Kastenmayer, J.P., and Green, P.J. (2004). AtXRN4 degrades mRNA in *Arabidopsis* and its substrates include selected miRNA targets. *Mol. Cell* **15**: 173– 183.

Takanashi, H., Sumiyoshi, H., Mogi, M., Hayashi, Y., Ohnishi, T., and Tsutsumi, N. (2018). miRNAs control HAM1 functions at the single-cell-layer level and are essential for normal embryogenesis in *Arabidopsis*. *Plant Molecular Biology* **96**: 627–640.

Vashisht, D. and Nodine, M.D. (2014). MicroRNA functions in plant embryos. *Biochem. Soc. Trans*. **42**: 352–357.

Willmann, M.R., Mehalick, A.J., Packer, R.L., and Jenik, P.D. (2011). MicroRNAs regulate the timing of embryo maturation in *Arabidopsis*. *Plant Physiol.* **155**: 1871– 1884.

Wu, G., Park, M.Y., Conway, S.R., Wang, J.-W., Weigel, D., and Scott Poethig, R. (2009). The Sequential Action of miR156 and miR172 Regulates Developmental Timing in *Arabidopsis*. *Cell* **138:** 750–759.

Wu, M.-F., Tian, Q., and Reed, J.W. (2006). *Arabidopsis* microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development* **133:** 4211–4218.

Xu, L., Hu, Y., Cao, Y., Li, J., Ma, L., Li, Y., and Qi, Y. (2018). An expression atlas of miRNAs in *Arabidopsis thaliana*. *Sci. China Life Sci.* **61**: 178–189.

Yao, X., Chen, J., Zhou, J., Yu, H., Ge, C., Zhang, M., Gao, X., Dai, X., Yang, Z.-N., and Zhao, Y. (2019). An Essential Role for miRNA167 in Maternal Control of Embryonic and Seed Development. *Plant Physiol.* **180**: 453–464.

6.9 Supplementary Information

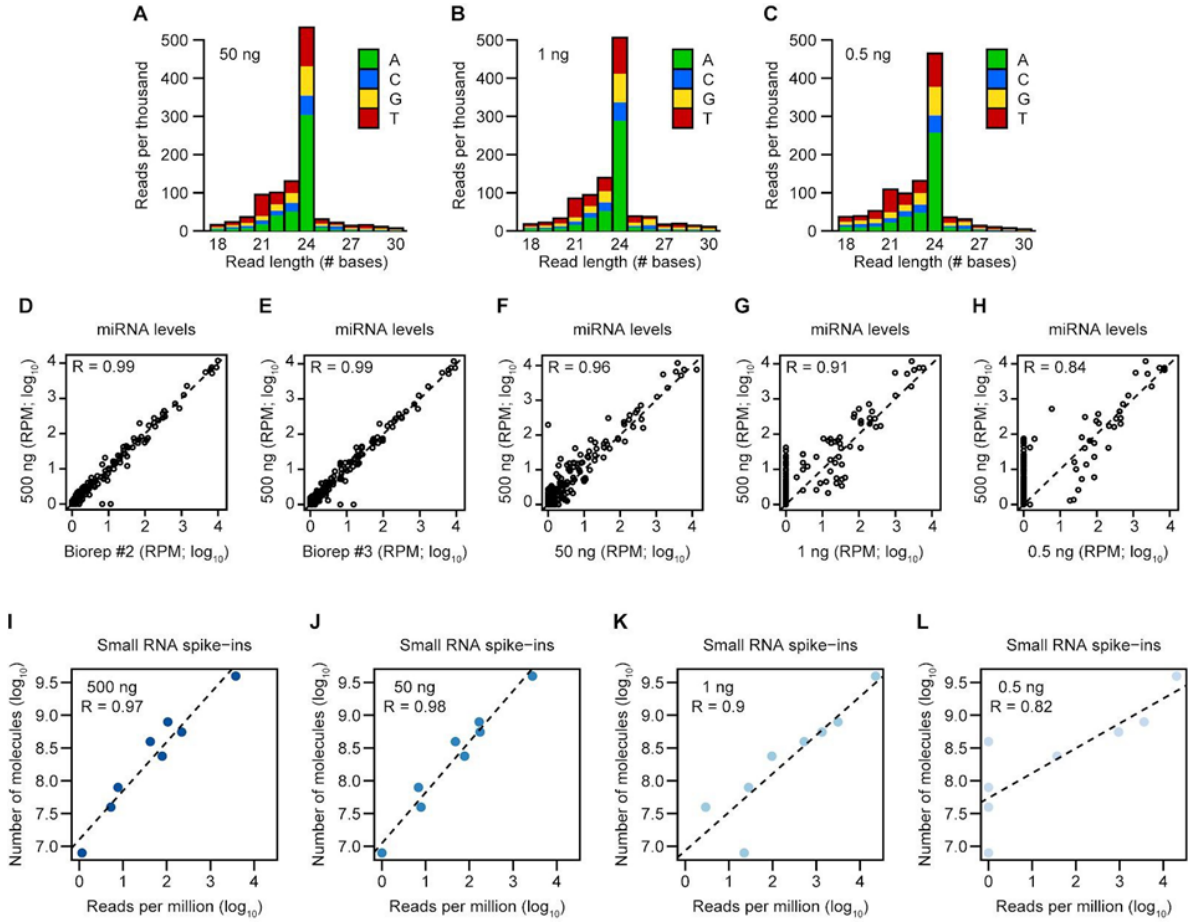


Figure S1. Establishment of Low-input Small RNA Sequencing Method, Related to Figure 1. **(A to C)** Stacked bar charts of normalized sRNA levels (reads per thousand genome-mapping reads) across different nucleotide (nt) lengths in libraries generated with either 50 ng **(A),** 1 ng **(B)** or 0.5 ng **(C)** of total RNA isolated from bent cotyledon stage embryos. Colors indicate proportions of sRNA-seq reads that begin with various bases as indicated in key. **(D to H)** Scatter plots of miRNA family levels in sRNA-seq libraries generated from either 500 ng (biological replicate #2) **(D),** 500 ng (biological replicate #3) **(E),** 50 ng **(F),** 1 ng **(G)** or 0.5 ng **(H)** compared to 500 ng of total RNA (biological replicate #1). sRNA levels were normalized for reads per million genome-mapping reads (RPM) and log10-transformed. Pearson's R values are indicated, as well as a dashed line with an intercept of 0 and slope of 1. **(I to L)** Scatter plots of relative sRNA spike-in levels (RPM; log10) compared to the absolute number of sRNA spike-in molecules (log10) added during RNA isolation for a sRNA-seq library generated from either 500 ng **(I),** 50 ng **(J),** 1 ng **(K)** or 0.5 ng **(L)** of total RNA. Pearson's R values are shown, and the dashed lines represent linear models derived from the plotted data points.

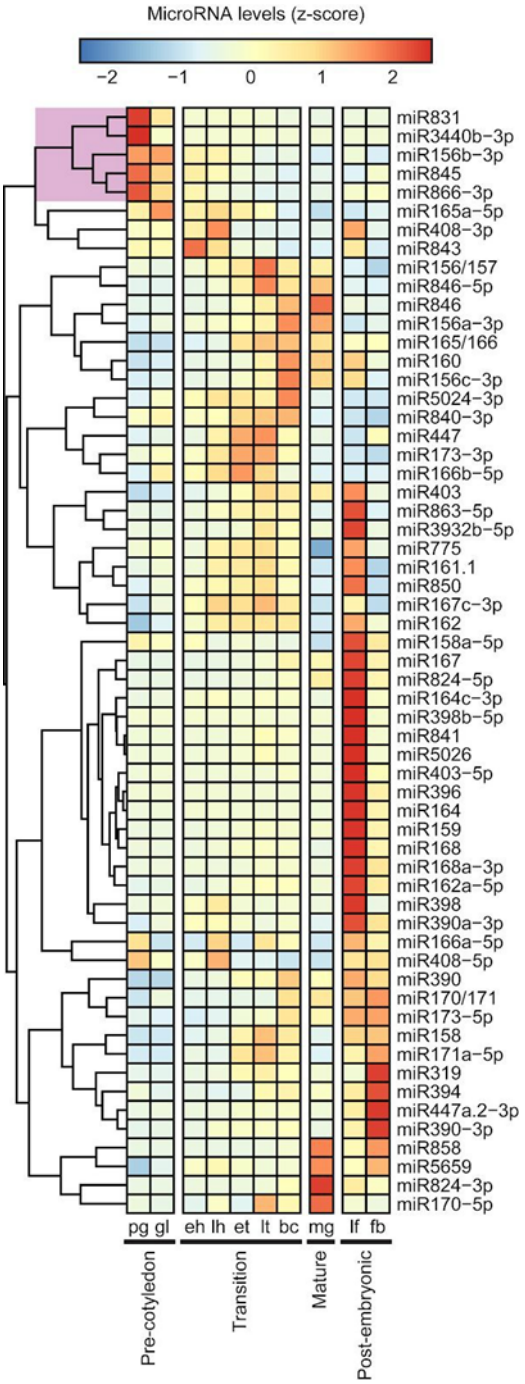


Figure S2. **Embryo-Enriched miRNAs,** Related to Figure 3.
Heat map illustrating the relative levels of miRNA families across embryogenesis, leaves and floral buds. miRNA families with ≥10 mean RPM in at least one embryonic stage are shown, and colors represent z-scores for each individual miRNA family according to the key. Three major phases of embryo development are indicated at the bottom and individual columns are labelled according to stage: pg, preglobular; gl, globular; eh, early heart; lh, late heart; et, early torpedo; lt, late torpedo; bc, bent cotyledon; mg, mature green; lf, leaves; fb, unopened floral buds. The dendrogram clade color-coded in violet indicates the five miRNA families enriched in early embryos.

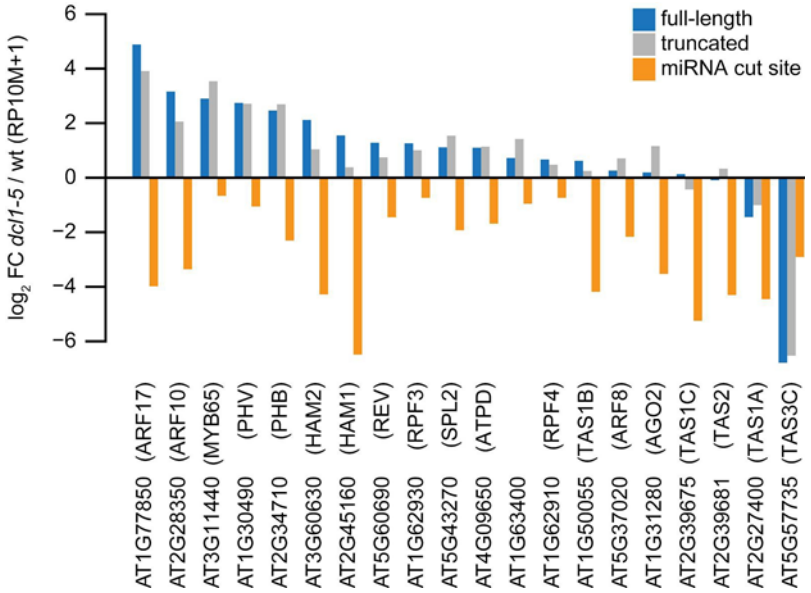


Figure S3. mRNA 5′ Ends of miRNA Targets in *dcl1-5* Mutant Embryos, Related to Figure 4. Comparison of the abundance of nanoPARE read 5′ ends mapping to different positions within all 20 high-confidence miRNA target transcripts identified in wild-type globular embryos. Each gene was subdivided into three regions: blue, annotated transcription start sites identified with nanoPARE (Schon *et al.* 2018); orange, positions 9 and 10 of the miRNA:target site; gray, all other exonic positions in the gene. Y-axis represents log2 fold change of the mean abundance of each gene feature in globular-stage *dcl1-5* mutant embryos compared to wild-type embryos of the same stage (reads per ten million genome-matching reads, RP10M).

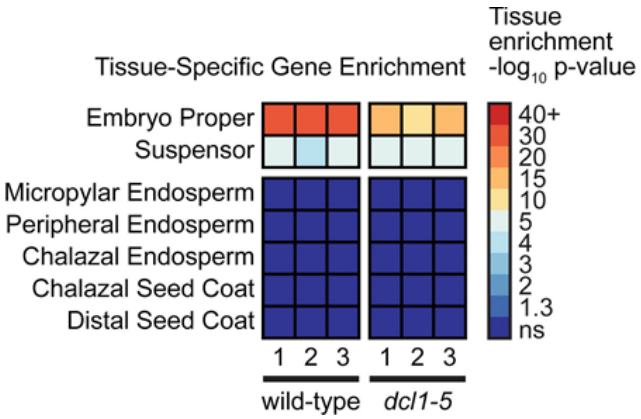


Figure S4. Tissue-Enrichment Test of Wild-Type and *dcl1-5* Mutant Embryo Transcriptomes, Related to Figure 5. Statistical enrichment of seven distinct seed tissue types inferred through the expression of tissue-enriched gene sets using the tissue-enrichment test (Schon and Nodine 2017) with default parameters. The three wild-type replicates are globular-stage mRNA-seq samples from GEO series GSE121236; *dcl1-5* replicates were generated for this study.

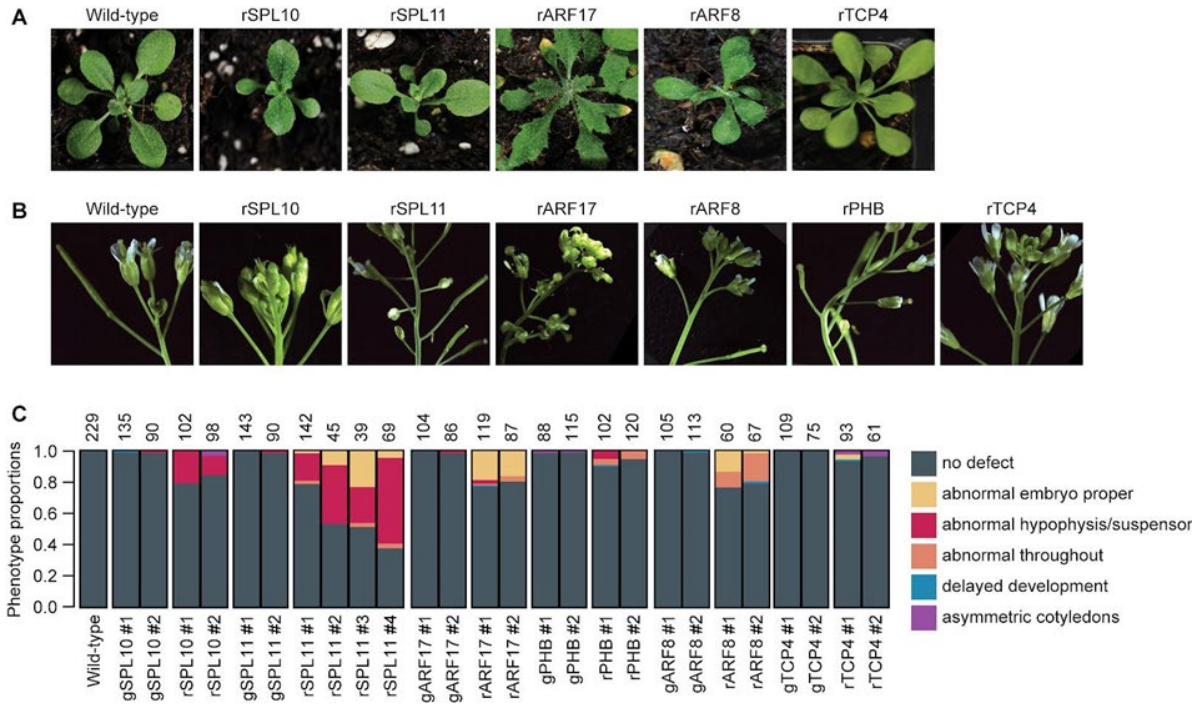


Figure S5. Post-Embryonic Phenotypes of Plants Expressing miRNA-Resistant Targets and Quantification Details, Related to Figure 7. **(A-B)** Representative images of vegetative **(A)** or flowering **(B)** plants **(A)** expressing miRNA-resistant targets. Genotypes are indicated above each panel. **(C)** Stacked bar plot illustrating the proportions of phenotypes observed for embryos derived from crosses between wild-type mothers and fathers that were either wild-type, or expressed transgenic copies of target transcripts containing wild-type (genomic; gTARGET) or abolished (resistant; rTARGET) miRNA binding sites 120 hours after pollination. Paternal genotypes used in the crosses, including transgenic line numbers, are labelled below. Numbers above each bar denote how many embryos were examined. Phenotypes are color-coded according to the legend.

Name	Sequence	General purpose	Specific purpose
miR124-AS-LNA	/5DIGN/ACTGATA+TC+AG+CTC+AGTAG GCAC/3DIG_N/	miRNA *in situs*	Negative control; antisense to animal-specific miR124 (+ indicates LNA position, /5DIGN/ and /5DIGN/ represents digoxigenin labels on the 5′ and 3′ ends)
miR156-AS-LNA	/5DIGN/GTGCT+CACT+CT+CTTCTG+TCA /3DIG_N/	miRNA *in situs*	Antisense to miR156a-f isoforms (+ indicates LNA position, /5DIGN/ and /5DIGN/ represents digoxigenin labels on the 5′ and 3′ ends)
miR159-AS-LNA	/5DIGN/TAGAG+CT+CC+CTT+CAATC+CA AA/3DIG_N/	miRNA *in situs*	Antisense to miR159a isoform (+ indicates LNA position, /5DIGN and /5DIGN/ represents digoxigenin labels on the 5′ and 3′ ends)
miR160-AS-LNA	/5DIGN/GGCATA+CAGG+GAG+CCAGG+ CAC/3DIG_N/	miRNA *in situs*	Antisense to miR160a-c isoforms (+ indicates LNA position, /5DIGN/ and /5DIGN/ represents digoxigenin labels on the 5′ and 3′ ends)
miR166-AS-LNA	/5DIGN/GGGGAA+TGAA+GC+CTGGTC+C GA/3DIG_N/	miRNA *in situs*	Antisense to miR166a-f isoforms (+ indicates LNA position, /5DIGN/ and /5DIGN/ represents digoxigenin labels on the 5′ and 3′ ends)
PHB F10	GTAGCGATGGTGCAGAGGATGT	mRNA *in situs*	Forward primer for amplifying PHB amplicon from cDNA
PHB R10	CGAACGACCAATTCACGAACAT	mRNA *in situs*	Reverse primer for amplifying PHB amplicon from cDNA
PHB F12	GTAGCGATGGTGCAGAGGATGTTACTG	mRNA *in situs*	Forward primer for incorporation of T7 site for PHB antisense probe generation
PHB R12T7	CCAAGCTTCTAATACGACTCACTATAGG GAGACGAACGACCAATTCACGAAC	mRNA *in situs*	Reverse primer for incorporation of T7 site for PHB antisense probe generation
PHB F11T7	CCAAGCTTCTAATACGACTCACTATAGG GAGAGTAGCGATGGTGCAGAGGAT	mRNA *in situs*	Forward primer for incorporation of T7 site for PHB sense probe generation
PHB R11	CGAACGACCAATTCACGAAC	mRNA *in situs*	Reverse primer for incorporation of T7 site for PHB sense probe generation
CNA_F1	TTCAAAGGCAACTGGAACCG	mRNA *in situs*	Forward primer for amplifying CNA amplicon from cDNA and for CNA antisense probe generation
CNA_R1	CCGCACAGGTTCCTACAGCA	mRNA *in situs*	Reverse primer for amplifying CNA amplicon from cDNA
CNA_AS_R1T7	CCAAGCTTCTAATACGACTCACTATAGG GAGAAGCAAGTGGAAGTATAACCT	mRNA *in situs*	Reverse primer for incorporation of T7 site for CNA antisense probe generation
PHV_F1	GGTCGCTGAAATCCTCAAAG	mRNA *in situs*	Forward primer for amplifying PHV amplicon from cDNA
PHV_R1	TTCGGATTTGTTTTTTGGTC	mRNA *in situs*	Reverse primer for amplifying PHV amplicon from cDNA
PHV_AS_F1	TCGTCCATCTTGGTTCCGTG	mRNA *in situs*	Forward primer for incorporation of T7 site for PHV antisense probe generation
PHV_AS_R1T7	CCAAGCTTCTAATACGACTCACTATAGG GAGAATTTTCATCAACGCCGCTAC	mRNA *in situs*	Reverse primer for incorporation of T7 site for PHV antisense probe generation
pAlligatorR/G43-F1	CTGCAGATCGTTCAAACATTTG	Cloning	Forward primer for generating the backbone of pAlligatorG43/R43 (use for g/mARF17, CNA, PHB, TCP4)
pAlligatorR/G43-R1	CTGCAGGTCGACCATAGTG	Cloning	Reverse primer for generating the backbone of pAlligatorG43/R43 (use for g/mARF17, CNA, PHB, TCP4)
pAlligatorR/G43-R2	ATAGCTTGGCGTAATCATGG	Cloning	Alternative reverse primer for generating the backbone of pAlligatorG43/R43 (use for g/mPHB)
g/mARF17-F1	ACACAACATATCCAGTCACTATGGTCGA CCTGCAGTGTCTTTTGTTTTTAGGTTTTT TTTTTAAC	Cloning	Forward primer for genomic ARF17 amplification and Gibson cloning into pAlligatorG43 /pAlligatorR43 destination vector
g/mARF17-R1	GAAACTTTATTGCCAAATGTTTGAACGAT CTGCAGTTTATTTAGTATTATTTGCTCTG TTTG	Cloning	Reverse primer for genomic ARF17 amplification and Gibson cloning into pAlligatorG43 /pAlligatorR43 destination vector
rARF17-F2	CTGGAATGCAAGGTGCACGGCAATATGA TTTTGGGTC	Cloning	Forward primer for the amplification of rARF17 Gibson piece 2 (with g/mARF17-R1)
rARF17-R2	ACCCAAAATCATATTGCCGTGCACCTTG CATTCCAG	Cloning	Reverse primer for the amplification of rARF17 Gibson piece 1 (with g/mARF17-F1)
gARF8-TOPO-F	CACCTCTCCAAGTGATACACTC	Cloning	Forward primer for genomic ARF8 amplification and cloning into pENTR/D-TOPO
gARF8-TOPO-R	TAAGTCTGATGTGTGTGCA	Cloning	Reverse primer for genomic ARF8 amplification and cloning into pENTR/D-TOPO
ARF8-SDM-F	CCGGTTGTTACGGAAAATACAAAAACAA C	Cloning	Forward site-directed mutagenesis primer to generate rARF8
ARF8-SDM-R	GGCCTGATTCCATTGGAATCATCG	Cloning	Reverse site-directed mutagenesis primer to generate rARF8

Name	Sequence	General purpose	Specific purpose
g/mPHB-F1	AAACAGCTATGACCATGATTACGCCAAG CTATTGGAGGGAAGAGGCTACAAAG	Cloning	Forward primer for genomic PHB amplification and Gibson cloning into pAlligatorG43 /pAlligatorR43 destination vector
g/mPHB-R1	ACTTTATTGCCAAATGTTTGAACGATCTG CAGTTGTCCGAGCATTGATTTTGTAC	Cloning	Reverse primer for genomic PHB amplification and Gibson cloning into pAlligatorG43 /pAlligatorR43 destination vector
rPHB-F2	AATAGAATCTGGTCCAGGCTACACCAGC AATGAAG	Cloning	Forward primer for the amplification of rPHB Gibson piece 2 (with g/mPHB-R1)
rPHB-R2	CATTGCTGGTGTAGCCTGGACCAGATTC TATTGGC	Cloning	Reverse primer for the amplification of rPHB Gibson piece 1 (with g/mPHB-F1)
g/mTCP4-F1	ACACAACATATCCAGTCACTATGGTCGA CCTGCAGCATTTTGATGAGGCGTATATA TATACATTTAATTAATATTG	Cloning	Forward primer for genomic TCP4 amplification and Gibson cloning into pAlligatorG43 /pAlligatorR43 destination vector
g/mTCP4-R1	GAAACTTTATTGCCAAATGTTTGAACGAT CTGCAGATATGATCTTTGTGTCATGACT	Cloning	Reverse primer for genomic TCP4 amplification and Gibson cloning into pAlligatorG43 /pAlligatorR43 destination vector
rTCP4-F2	GGTCCCTTGCAAAGTAGCTACAGTCCCA TGATCCGTG	Cloning	Forward primer for the amplification of rTCP4 Gibson piece 2 (with g/mTCP4-R1)
rTCP4-R2	ACGGATCATGGGACTGTAGCTACTTTGC AAGGGACC	Cloning	Reverse primer for the amplification of rTCP4 Gibson piece 1 (with g/mTCP4-F1)
eIF4A1 RTF	TGCAAGGCACTCTTTGATCTGATTT	qRT-PCR	Forward primer for detection of the housekeeping gene eIF4A1
eIF4A1 RTR	GAGATATGTTCGTAGCTGGGAGAGAGA G	qRT-PCR	Reverse primer for detection of the housekeeping gene eIF4A1
SPL10 RTF	TCAGGAGGCCTCCATGAATCTCA	qRT-PCR	Forward primer for SPL10 detection
SPL10 RTR	GGCCACGGGAGTGTGTTTGAT	qRT-PCR	Reverse primer for SPL10 detection
SPL11 RTF	CCAACCACATGTGCAGCCATTT	qRT-PCR	Forward primer for SPL11 detection
SPL11 RTR	GAACAGAGTAGAGAAAATGGCTGCA	qRT-PCR	Reverse primer for SPL11 detection
PHB RTF	GCTAGACAAGACCCTTGACGAACCT	qRT-PCR	Forward primer for PHB detection
PHB RTR	TCCCATGCTTGACGCACATACTC	qRT-PCR	Reverse primer for PHB detection
ARF8 RTF	CATGCAGATGTTGAGACGGATGAAG	qRT-PCR	Forward primer for ARF8 detection
ARF8 RTR	TTACTCGGTATCCCCAACTCAATCG	qRT-PCR	Reverse primer for ARF8 detection
ARF17 RTF	GTGCAGCAGCACCTGATCCAAG	qRT-PCR	Forward primer for ARF17 detection
ARF17 RTR	GGAGGATTTCCTCCAATGAATCCGG	qRT-PCR	Reverse primer for ARF17 detection
TCP4 RTF	CCAGTTCTTGGCCAAAGCCAAC	qRT-PCR	Forward primer for TCP4 detection
TCP4 RTR	ATGGTGGTGGTTGAGATCGTCG	qRT-PCR	Reverse primer for TCP4 detection

Supplementary Table1. List of oligonucleotides used in the study.

4. Discussion

The ground principles of plant body organization are established during early embryogenesis as a result of tightly coordinated spatiotemporal regulation of cell division and differentiation. The present work contributes to understanding how miRNA-mediated cleavage of transcription factors orchestrate Arabidopsis embryo morphogenesis. In this study, a combination of three powerful genome-wide approaches was used to characterize the miRNAs and their targets in developmental timecourse, spanning eight stages of embryogenesis:

- A low-input small RNA sequencing method described in Chapter 6 allows small RNA library preparation from 1-5 ng of total RNA, which is at least 100 times less than the input required by conventional strategies. This method includes gel-based size-selection steps to enrich for the small RNA populations. We detected hundreds of miRNAs present in the developing embryos. Moreover, multiple miRNA populations were differentially expressed at distinct developmental stages, suggesting that particular miRNAs are tissue- or cell-type-specific and operate at certain embryogenesis phases.
- Because miRNAs' functions are tightly coupled with the transcripts they regulate, we used low-input transcriptome analysis to identify these transcripts. We found that the reduction of mature miRNA levels in dcl1 mutants resulted in profound changes in almost 30% of the transcript populations. For example, dcl1-5 globular embryos untimely accumulate transcripts, characteristic for the late developmental stages, suggesting that miRNAs' cleavage activity prevents the precocious onset of maturation (Hofmann *et al.*, 2019) and strongly affects embryonic gene expression programs.
- The precise detection of these miRNA-induced cleavage events is crucial for molecular characterization of processes underlying cellular differentiation, tissue specification, physiology and development. The novel nanoPARE method described in Chapters 5 and 6 can reliably identify 5’-cleavage products on a genome-wide scale. Compared to other approaches for degradome analysis, such as PARE or CAGE, nanoPARE can be applied to high-range input material from standard amounts of total RNA to individual cell-types or even single-cells from hard-to-obtain samples. As a case study, we applied nanoPARE to the whole flowers and five floral tissue-types.

Angiosperm species share the main floral architecture principles, where sepals, petals, stamens and carpels are arranged in concentric whorls and have defined roles in reproduction (Bowman et al., 2012). The "ABC model" of flower development suggests that floral organs are specified by a unique combination of overlapping activities of homeotic 'A,' 'B' and 'C' genes in each whorl. The model also explains a wide variety of phenotypic variations in homeotic mutants, in which floral organs have altered identity or are spatially rearranged. The tissue specification within Arabidopsis flowers is regulated by two A (APETALA1 (AP1) and APETALA2 (AP2)), two B (APETALA3 (AP3) and PISTILLATA (PI)), and one C (AGAMOUS (AG)) homeotic genes expressed in the outer, middle or inner domains of the developing

flower, respectively (Bowman *et al.,* 2012). According to the model, AP1/AP2 define sepals (whorl 1), AP2 and AP3/PI specify petals (whorl 2), and AP3/PI and AG control stamen identity (whorl 3), and AG defines carpels (whorl 4). Interestingly, AP2 and AG are reciprocally antagonistic, and AP2 is post-transcriptionally regulated by miR172 (Bowman *et al.,* 2012, Aukerman and Sakai, 2003).

NanoPare data obtained from petals, sepals, anthers, stigmas, and ovules, manually dissected from the stage 13 flowers (after anthesis), is in good agreement with the spatial transcript distribution predicted by the "ABC model." We also found 41 high-confidence miRNA:target interactions, most of which significantly differed between tissue types. For instance, cleavage products of ARF6 and AFR8 were enriched in anthers, sepals and ovules, consistently with the expression pattern of MIR167. Despite the recent advances, little is known about the gene regulatory networks downstream of ABC genes or molecular processes behind the evolutionary diversity of flower morphology, missing or mosaic floral organs, or secondary flowers. The atlas of tissue-enriched 5'-cleavage products generated from flowers using scalable and straightforward nanoPARE method is a valuable resource for studying the molecular basis of floral development.

Because nanoPARE can utilize cDNAs from existing Smart-seq2 libraries, we applied it to the same embryonic samples used for transcriptome profiling, including dcl1-5 globular embryos. 59 high-confidence target sites corresponding to 22 miRNA families were identified across developmental timecourse. Remarkably, more than half of these high-confidence embryonic miRNA targets encoded transcription factors, including MADS-box, ARF, HD-ZIP and others. No high-confidence cleavage events, but full-length transcripts were detected in miRNA-deficient dcl1-5 embryos, further confirming the identified targets.

The above mentioned approaches (sRNA-seq, transcriptome analysis and nanoPARE) allowed us to profile embryonic miRNAs:transcription factor interactions genome-wide. However, for this analysis, we have sampled RNA obtained from the whole embryos. In order to visualize the miRNA-guided repression of transcription factors at cellular resolution, we examined fluorescent miRNA activity sensors (Nodine and Bartel, 2010).

We imaged the embryos expressing either non-genome matching "scrambled" sequence or miRNA target sites of miR156/157 (SPL10/11), miR160 (ARF17), miR165/166 (PHB), miR167 (ARF8) or miR319 (TCP4) identified by nanoPARE analysis. Compared to uniformly expressed scrambled control, the sensors had repressed GFP signal at least in one embryonic stage. Several sensors (miR156/157 and miR165/166) were repressed throughout the developmental time course, while the other exhibited dynamic repression patterns. For instance, the signal from the miR167 sensor was weak throughout the 16-cell embryos and strongly silenced in the shoot meristem precursors and suspensor at the heart stage. The observed repression patterns were mostly consistent with the miRNA localization data from whole-mount RNA in situ experiments. Therefore, these results confirm that miRNAs-mediated repression affects transcriptional regulatory programs across developmental stages

and individual cell types (Nodine and Bartel, 2010; Seefried *et al.* 2014).

In the course of embryogenesis, miRNAs restrict the ectopic or premature expression of their targets to facilitate proper pattern formation. To understand the impact of individual miRNAs and transcription factors on embryo morphogenesis, we generated and phenotyped six transgenic lines with mutated miRNA binding sites to determine the influence of individual miRNA:target interactions on embryo morphogenesis: miR156/157:SPL10, miR156/157:SPL11, miR160:ARF17, miR165/166:PHB, miR167:ARF8, and miR319:TCP4. Not only these transgenic lines phenocopied previously published post-embryonic defects, but they also had abnormal embryo morphology (Wu *et al.,* 2006; Mallory *et al.* 2005; Nodine and Bartel, 2010). For example, miR156/157-resistant SPL10 and SPL11 embryos had irregular division patterns in their uppermost suspensor and hypophysis. Their abnormal inflorescence and rosette morphology was consistent with previous reports (Wu *et al.* 2009; Kim *et al.* 2012; Yang *et al.* 2013; Guo *et al.* 2017).

Although follow-up research is required for the exhaustive characterization of developmental regulatory networks (e.g., identifying key molecules upstream of miRNAs, downstream of transcription factors, and their modes of action), the results presented above provide exciting insights into the molecular basis of embryonic patterning in Arabidopsis. We created a valuable toolbox for genome-wide analysis of miRNA:target interactions and identified several candidates directly implicated in transcriptional control of embryo morphogenesis.

8. References

Abe, M., Katsumata, H., Komeda, Y., Takahashi, T. (2003). Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in *Arabidopsis*. *Development* **130**(4): 635–643.

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P., Axtell, M.J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr. Biol.* **18**: 758–762.

Akbergenov, R., Si-Ammour, A., Blevins, T., Amin, I., Kutter, C., Vanderschuren, H., Zhang, P., Gruissem, W., Meins, F. Jr., Hohn, T., Pooggin, M.M. (2006). Molecular characterization of geminivirus-derived small RNAs in different plant species. *Nucleic Acids Res.* **34**: 462–471.

Ambros, V. (2011). MicroRNAs and developmental timing. *Curr. Opin. Genet. Dev.* **21**(4): 511-517.

Allen, B.L. and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription *Nat. Rev. Mol. Cell. Biol.* **16**:155-166.

Aukerman, M.J. and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* **15**: 2730– 2741.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281-297.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* **136**(2): 215-233.

Baud, S. and Lepiniec, L. (2009). Regulation of *de novo* fatty acid synthesis in maturing oilseeds of *Arabidopsis*. *Plant Physiol. Biochem.* **47**: 448–455.

Baud, S., Dubreucq, B., Miquel, M., Rochat, C. and Leipiniec, L. (2008). Storage reserve accumulation in *Arabidopsis*: metabolic and developmental control of seed filling. *Arabidopsis Book*. Rockville, *MD, American Society of Plant Biology* **6**, e0113.

Baumberger, N. and Baulcombe, D.C. (2005). *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci. USA*. **102**(33): 11928-11933.

Bowman, J.L., Smyth, D.R., Meyerowitz, E.M. (2012). The ABC model of flower development: then and now. *Development* **139**: 4095-4098.

Braybrook, S.A. and Harada, J.J. (2008). LECs go crazy in embryo development. *Trends Plant Sci.* **13**(12): 624–630.

Buendía-Monreal, M. and Gillmor, S.C. (2016). Mediator: A key regulator of plant development, *Dev. Biol.* **419**(1): 7-18.

Cambiagno, D.A., Giudicatti, A.J., Arce, A.L., Gagliardi, D., Li, L., Yuan,W., Derek S. Lundberg, Weigel, D. and Manavella, P.A. (2020). HASTY modulates miRNA biogenesis by linking pri-miRNA transcription and processing. *Molecular Plant* (in press).

Chen, X. (2009). Small RNAs and their roles in plant development. *Annu. Rev. Cell. Dev. Biol.* **25**: 21-44.

Christensen, C., King, E., Jordan, J.R., and Drews, G.N. (1997). Megagametogenesis in *Arabidopsis* wild-type and the Gf mutant. *Sex Plant Reprod.* **10**: 49–64.

Eamens, A.L., Smith, N.A., Curtin, S.J., Wang, M.-B., Waterhouse, P.M. (2009). The *Arabidopsis thaliana* double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA* **15**: 2219–2235.

Fang, Y., Spector, D.L. (2007). Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living *Arabidopsis* plants. *Curr. Biol.* **17:** 818–823.

Faure, J.E., Rotman, N., Fortune, P., Dumas, C. (2002). Fertilization in *Arabidopsis thaliana* wild-type: Developmental stages and time course. *Plant J.* **30**(4): 481-488.

Fujioka, Y., Utsumi, M., Ohba, Y., Watanabe, Y. (2007). Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in *Arabidopsis*. *Plant Cell Physiol.* **48**: 1243–1253.

German, M.A., Pillay, M., Jeong, D.-H., Hetawal, A., Luo, S., Janardhanan, P., Kannan, V., Rymarquis, L.A., Nobuta, K., German, R., *et al.* (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* **26**: 941–946.

Goffman, F.D., Alonso, A.P., Schwender, J., Shachar-Hill, Y. and Ohlrogge, J.B. (2005). Light enables a very high efficiency of carbon storage in developing embryos of rapeseed. *Plant Physiol.* **138**: 2269–2279.

Goldberg, R.B., de Paiva, G. and Yadegari, R. (1994). Plant embryogenesis: zygote to seed. *Science* **266**: 605–614.

Guo, C., Xu, Y., Shi, M., Lai, Y., Wu, X., Wang, H., Zhu, Z., Poethig, R.S., Wu, G. (2017). Repression of miR156 by miR159 Regulates the Timing of the Juvenile-to-Adult Transition in Arabidopsis. *The Plant Cell.* **29**(6):1293-1304.

Hofmann, F., Schon, M.A., and Nodine, M.D. (2019). The embryonic transcriptome of *Arabidopsis thaliana*. *Plant Reprod.* **32**: 77–91.

Hutvágner, G., McLachlan, J., Pasquinelli, A.E., Bálint E., Tuschl, T. and Zamore, P.D. (2001). A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA. *Science* **293** (5531): 834-838.

Jones-Rhoades, M.W. and Bartel, D.P. (2004). Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. *Molecular Cell* **14**: 787–799.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. *Annual Review of Plant Biology* **57**: 19–53.

Jürgens, G. and Mayer, U. (1994). *Arabidopsis*. In *Embryos: Color Atlas of Development* (J. Bard, ed). *London: Wolfe Publishing*, pp. 7–21.

Kao, P. and Nodine, M.D. (2019). Transcriptional Activation of *Arabidopsis* Zygotes Is Required for Initial Cell Divisions. *Sci. Rep.* **9**(1): 17159.

Kawamata, T., Tomari, Y. (2010). Making RISC. *Trends Biochem Sci.* **35**(7): 368-376.

Kawashima, T. and Goldberg, R.B. (2010). The suspensor: not just suspending the embryo. *Trends Plant Sci.* **15**(1): 23-30.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev.* **15**(20): 2654-9.

Kim, J.J., Lee, J.H., Kim, W., Jung, H.S., Huijser, P., Ahn J.H. (2012). The *microRNA156-SQUAMOSA PROMOTER BINDING PROTEIN-LIKE3* Module Regulates Ambient Temperature-Responsive Flowering via *FLOWERING LOCUS T* in Arabidopsis. *Plant Phys.* **159** (1):461-478.

Kim, V. (2005). MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6**: 376–385.

Kurihara, Y., Watanabe, Y. (2004). *Arabidopsis* microRNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci. USA.* **101**: 12753–12758.

Laux, T., Würschum, T. and Breuninger, H. (2004). Genetic regulation of embryonic pattern formation. *The Plant Cell* **16** Suppl: 190–202.

Law, J.A. and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Gen.* **11**: 204-220.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**(6956): 415-9.

Leprince, O., Pellizzaro, A., Berriri, S. and Buitink, J. (2016) Late seed maturation: drying without dying. *J. Exp. Bot.* **68**(4), 827–841.

Li, J., Yang, Z., Yu, B., Liu, J., Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in *Arabidopsis*. *Curr. Biol.* **15**: 1501–1507.

Liu, H., Wang, X., Ren, K., Li, K., Wei, M., Wang, W. and Sheng, X. (2017) Light deprivation-induced inhibition of chloroplast biogenesis does not arrest embryo morphogenesis but strongly reduces the accumulation of storage reserves during embryo maturation in *Arabidopsis*. *Front Plant Sci.* **8**, 1287.

Mallory, A.C., Bartel, D.P., and Bartel, B. (2005). MicroRNA-directed regulation of *Arabidopsis* AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *Plant Cell* **17**: 1360–1375.

Manavella, P.A., Koenig, D., Weigel, D. (2012). Plant secondary siRNA production determined by microRNA-duplex structure. *Proc. Natl. Acad. Sci. USA.* **109**: 2461–2466.

Mansfield, S.G. and Briarty, L.G. (1991). Early embryogenesis in *Arabidopsis thaliana*. II. The developing embryo. *Can. J. Bot.* **69**: 461–476.

Mathur, S., Vyas, S., Kapoor, S. and Tyagi, A.K. (2011). The Mediator Complex in Plants: Structure, Phylogeny, and Expression Profiling of Representative Genes in a Dicot (*Arabidopsis*) and a Monocot (Rice) during Reproduction and Abiotic Stress. *Plant Phys.* **157**(4): 1609-1627.

Nodine, M.D, Bartel, D.P. (2010). MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes Dev.* **24**: 2678–2692.

O'Neill, J.P., Colon, K.T. and Jenik, P.D. (2019). The onset of embryo maturation in *Arabidopsis* is determined by its developmental stage and does not depend on endosperm cellularization. *Plant J.* **99**: 286-301.

Owen, H.A, Makaroff, C.A. (1995). Ultrastructure of microsporogenesis and microgametogenesis in *Arabidopsis thaliana* (L.) Haynh. ecotype Wassilewskija (Brassicaceae). *Protoplasma* **185**: 7–21.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**: 1484–1495.

Pelletier, J.M., Kwong, R.W., Park, S., Le B.H., Baden R., Cagliari A., Hashimoto M., Munoz M.D., Fischer R.L., Goldberg R.B., Harada J.J. (2017) LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proc. Natl Acad. Sci. USA*. **114**(32): 6710–6719.

Plasterek, R.H. (2006). MicroRNAs in animal development. *Cell* **124**: 877-881.

Qi, Y., Denli, A.M. and Hannon, G. (2005). Biochemical specialization within *Arabidopsis* RNA silencing pathways. *Mol. Cell* **19**: 421-428.

Radoeva, T., and Weijers, D. (2014). A roadmap to embryo identity in plants. *Trends Plant Sci.* **19**: 709–716.7.

Raz, V., Bergervoet, J.H.W. and Koornneef, M. (2001) Sequential steps for developmental arrest in *Arabidopsis* seeds. *Development* **128**: 243–252.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. *Genes Dev*. **16**: 1616–1626.

Ren, G., Chen, X., Yu, B. (2012). Uridylation of miRNAs by *hen1* suppressor1 in *Arabidopsis*. *Curr. Biol.* **22**: 695–700.

Sabatini, S., Heidstra, R., Wildwater, M. and Scheres, B. (2003). SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes Dev*. **17**: 354-358.

Santos-Mendoza, M., Dubreucq, B., Baud, S., Parcy, F., Caboche, M., Lepiniec, L. (2008). Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *Plant J*. **54**(4): 608–620.

Scheres, B., Wolkenfelt, H., Willemsen, V., Terlouw, M., Lawson, E., Dean, C., Weisbeek, P. (1994). Embryonic origin of the *Arabidopsis* primary root and root meristem initials. *Development* **120**(9): 2475–2487.

Schlereth, A., Möller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E.H., Schmid, M., Jürgens, G. & Weijers, D. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature* **464**: 913-916.

Schwartz, B., Yeung, E., and Meinke, D. (1994). Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of *Arabidopsis*. *Development* **120**: 3235–3245.

Seefried, W.F., Willmann, M.R., Clausen, R.L., and Jenik, P.D. (2014). Global Regulation of Embryonic Patterning in *Arabidopsis* by MicroRNAs. *Plant Physiol.* **165:**670–687

Song, L., Han, M.-H., Lesicka, J., Fedoroff, N. (2007). *Arabidopsis* primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc. Natl. Acad. Sci. USA*. **104:** 5437–5442.

Sunkar, R., Li, Y.F., Jagadeeswaran, G.(2012). Functions of microRNAs in plant stress responses. *Trends Plant Sci.* **17**(4): 196-203.

Takada, S. and Jürgens, G. (2007). Transcriptional regulation of epidermal cell fate in the *Arabidopsis* embryo. *Development* **134**: 1141-1150.

Tang, G., Reinhart, B.J., Bartel, D.P. and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. *Genes Dev.* **17**: 49-63.

ten Hove, C.A., Kuan-Ju Lu, K.J., Weijers, D. (2015). Building a plant: cell fate specification in the early *Arabidopsis* embryo. *Development* **142**: 420-430.

Ueda, M. and Laux, T. (2012). The origin of the plant body axis. *Curr. Opin. Plant Biol.* **15**(6): 578-84.

van Dop, M., Liao, C.Y., Weijers D. (2015) Control of oriented cell division in the *Arabidopsis* embryo. *Curr. Opin. Plant Biol.* **23**: 25-30.

Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS. 2009. The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* **138**: 750–759.

Wu MF, Tian Q, Reed JW. 2006. *Arabidopsis* microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development* **133**: 4211–4218.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**: 104.

Yang, L., Xu, M., Koo, Y., He, J., Poethig, R.S. (2013). Sugar promotes vegetative phase change in *Arabidopsis thaliana* by repressing the expression of *MIR156* and *MIR156C*. *eLife* **2***:e00260*.

Yeung, E.C. and Meinke, D.W. (1993). Embryogenesis in angiosperms: development of the suspensor. *Plant Cell* **5**: 1371-1381.

Yoshida, S., Barbier de Reuille, P.B., Lane, B., Bassel, G.W., Prusinkiewicz, P., Smith, R.S. Weijers, D. (2014). Genetic control of plant development by overriding a geometric division rule. *Dev. Cell.* **29**(1): 75–87.

Zhang, Z. and Laux, T. (2011). The asymmetric division of the *Arabidopsis* zygote: from cell polarity to an embryo axis. *Sex Plant Reprod.* **24**: 161-169.

Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T.T., Meyers, B.C., Mo, B., Chen, X. (2012). The *Arabidopsis* nucleotidyl-transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr. Biol.* **22**: 689–694.

Zhao, P., Zhou, X., Shen, K., Liu, Z., Cheng, T., Liu, D., Cheng, Y., Peng, X., Sun, M. (2019). Two-Step Maternal-to-Zygotic Transition with Two-Phase Parental Genome Contributions. *Dev. Cell.* **49**(6): 882–893.