Gametogenesis-related small RNAs and Argonaute Proteins in *Arabidopsis thaliana*

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES
DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN
DER UNIVERSITÄT REGENSBURG

vorgelegt von

Marc Urban

aus

Hannover

im Februar 2016

Das Promotionsgesuch wurde eingereicht am: 12.02.2016

Die Arbeit wurde angeleitet von: PD Dr. S. Sprunck

Unterschrift:

TABLE OF CONTENTS

1 SUMMARY

The life cycle of flowering plants alternates between a diploid multicellular sporophyte generation and a highly reduced haploid gametophyte generation located within the reproductive tissues of the sporophyte. Emerging evidence indicates that the transition from sporophyte to a gametophytic life phase and the acquisition of reproductive fate is marked by extensive epigenetic reprogramming employing distinct members of the ARGONAUTE (AGO) family of proteins with roles in RNA-directed DNA methylation (RdDM) and heterochromatin formation. Nevertheless, almost nothing is known about RNA silencing machinery components and small noncoding RNAs associated with epigenetic reprogramming and the acquisition of gametic cell fate.

The aim of this work was the detailed investigation of two RdDM-associated Argonaute genes (*AGO8* and *AGO9*) regarding their expression pattern and function in the female reproductive lineage of *Arabidopsis thaliana*. Furthermore, a specialized Arabidopsis cell line with an egg cell-like transcriptome was generated to enable the identification of small RNAs and RNA silencing machinery components potentially associated with female reproductive fate.

RT-PCR based expression studies were performed for all ten Arabidopsis *AGO* genes, confirming that *AGO5*, *AGO8* and *AGO9* are predominantly expressed in floral tissues containing reproductive lineages, while other *AGO* genes are more or less ubiquitously expressed. Protein abundance was investigated by Western Blots using peptide antibodies available for six AGOs andp09 revealed that AGO5 and AGO9 are strongly represented in reproductive organs and in the egg cell-like callus. To investigate *AGO8* and *AGO9* promoter activities and the presence of AGO8 and AGO9 proteins in more detail during female gametophyte development, transgenic Arabidopsis reporterlines were generated and whole-mount immunolocalization experiments were performed. These studies showed that the *AGO8* promoter is only active in the egg cell. Nevertheless, *AGO8* expression only yielded aberrant spliced mRNA and a GFP-AGO8 fusion protein was not detectable in egg cells expressing *AGO8p:GFP-AGO8*, suggesting that *AGO8* represents a pseudogene.

AGO9 promoter-reporter lines showed activity in the nucellus cells, the megaspore mother cell (MMC) and the functional megaspore (FM) of young ovules. In the mature female gametophyte, *AGO9* promoter activity was detectable in the egg cell, the central cell and the chalazal region of the ovule. This is in remarkable contrast to results obtained by AGO9

whole-mount immunolocalization and expression of the GFP-AGO9 fusion protein: AGO9 was neither detectable in the MMC and FM nor in the egg cell nor central cell. The occurrence of aberrant splicing by intron retention in *AGO9* mRNA, emerging during ovules maturation, and the results obtained from ectopic expression of correctly spliced *AGO9* cDNA in the egg cell suggests that both, developmentally regulated differential splicing and decreased AGO9 protein stability is responsible for the lack of AGO9 protein in the egg cell. A few plants were obtained expressing GFP-AGO9 in the egg cell, which did, however, not correlate with developmental defects in the female gametophyte or in developing seeds. The impact of ectopic GFP-AGO9 on DNA methylation in the egg cell remains to be investigated.

To identify protein coding genes and small noncoding RNAs possibly involved in the acquisition of female reproductive fate or in egg cell specification a specialized transgenic Arabidopsis cell line with egg cell-like expression profile was established by ectopically expressing the transcription factor RKD2. Compared to the control cell line 5,511 genes were classified as differentially expressed by RNA-seq. Quantitative real-time PCR with eighteen selected genes confirmed mRNA-Seq expression data. A global look at genes involved in small RNA pathways revealed several differentially expressed genes in RKD2-induced cell line, including upregulated genes encoding the DNA methyltransferase *DOMAINS REARRANGED METHLYTRANSFERASE 1* (*DRM1*), the *DOUBLE-STRANDED RNA BINDING PROTEIN 3* (DRB3), the *HISTONE DEACETLYLASE 18* (*HDA18*), the putative chromatin remodeling protein *CHROMATIN REMODELING 34 (CHR34)* and a so far undescribed DNA-directed RNA polymerase V subunit 5A-like gene.

Small RNA profiling in the RKD2-induced cell line revealed noticeable differences compared to the control callus. The most abundant group of small RNAs in the control callus are miRNAs, whereas the biogenesis of siRNAs from transposable element (TE) transcripts is dominant in the egg cell-like callus. TE-derived siRNAs of the Ty3/Gypsy superfamily of long terminal repeat retrotransposons, containing members of the *ATHILA*, *ATLANTYS*, and *ATGP* family, form the largest fraction of siRNAs, supporting the idea that the plant reproductive lineage is protected by siRNAs against the activity of TEs. This is in line with the observation that aligned reads to respective TE transcripts are not present in the mRNA-Seq data of the egg cell-like cell line. Differential expression analysis of miRNAs revealed 96 miRNAs being at least up- or down regulated with a log2 fold change of 2, including only 9 miRNAs upregulated in the RKD2-induced callus. Furthermore, 32 novel miRNAs were discovered, with 20 miRNAs being derived from known miRNA precursors while 12

miRNAs are completely novel. Target gene predictions of differential expressed miRNAs revealed several inverse correlations between miRNA and mRNA target expression levels in both the egg cell-like callus and the control callus.

The discovery of so far undescribed transcripts in the egg cell-like cell line encoding small RNA pathway components involved in chromatin modification, together with the identification of differentially expressed known and novel small noncoding RNAs open up interesting possibilities for future investigations regarding their contribution to the acquisition of reproductive fate and egg cell specification.

2 ZUSAMMENFASSUNG

Der Lebenszyklus der Blütenpflanzen ist durch einen Generationswechsel geprägt. Der diploide und mehrzellige Sporophyt wechselt sich mit dem stark reduzierten, haploiden Gametophyten ab, welcher sich in den reproduktiven Geweben des Sporophyten befindet. Neuere Erkenntnisse weisen darauf hin, dass der Übergang vom Sporophyten zur gametophytischen Lebensphase und der Erwerb der Fähigkeit zur Reproduktion durch eine weitgehende epigenetische Neuprogrammierung des Genoms gekennzeichnet sind. Dabei sind bestimmte Mitglieder der ARGONAUTE (AGO) Proteinfamilie beteiligt. Diese spielen eine Rolle beim Prozess der RNA gesteuerten DNA-Methylierung (RdDM, RNA-directed DNA methylation). Trotz dieser Erkenntnisse ist wenig darüber bekannt, welche Komponenten der RNA- silencing Maschinerie und kleinen nicht-kodierenden RNAs an der epigenetischen Reprogrammierung in reproduktiven Geweben und am Erwerb eines gametischen Zellschicksals beteiligt sind.

Ziel dieser Arbeit war die detaillierte Untersuchung zweier RdDM-assoziierter AGO Proteine (AGO8 und AGO9) bezüglich ihres Expressionsmusters und ihrer Funktion in der weiblichen Keimbahn von *Arabidopsis thaliana*. Darüber hinaus wurde eine spezialisierte Arabidopsis Zelllinie mit einem eizellähnlichen Transkriptom erzeugt, um die Identifizierung von kleinen RNAs und RNA-silencing Komponenten bei der weiblichen Gametenbildung zu ermöglichen.

Für alle zehn Arabidopsis *AGO* Gene wurden RT-PCR basierte Expressionsstudien durchgeführt, diese bestätigten dass *AGO5*, *AGO8* und *AGO9* überwiegend in Blütengeweben, welche reproduktive Zellen enthalten, exprimiert sind. Im Gegensatz dazu kommen alle anderen *AGOs* mehr oder weniger ubiquitär vor. Die Präsenz von AGO-Proteinen wurde mittels Western Blot untersucht, wozu verfügbare Peptidantikörper gegen sechs AGO Proteine verwendet wurden. Diese Untersuchung ergab, dass AGO5 und AGO9 in reproduktiven Organen und in der eizellähnliche Zelllinie verstärkt vertreten sind. Um die Promoteraktivitäten und die Proteinpräsenz von AGO8 und AGO9 während der Entwicklung des weiblichen Gametophyten eingehender zu untersuchen, wurden transgene Arabidopsis Reporterlinien generiert und eine immunhistologische Untersuchung (whole mount-immunolocalization) durchgeführt. Diese Studien ergaben, dass der *AGO8* Promoter ausschließlich in der Eizelle aktiv ist. Nichtsdestotrotz ergab die Expression von AGO8 lediglich aberrant gespleißte mRNA. Auch in Eizellen die *AGO8p:GFP-AGO8* exprimierten,

konnte kein GFP-AGO8 Fusionsprotein nachgewiesen werden, was darauf hinweist, dass es sich bei *AGO8* um ein Pseudogen handelt.

AGO9 Promoter-Reporter-Linien zeigten Aktivität in den Nucelluszellen, in der Megasporenmutterzelle (megaspore mother cell, MMC) und der funktionellen Megaspore (FM) von jungen Samenanlagen. Im reifen weiblichen Gametophyten konnte *AGO9* Promoteraktivität in der Eizelle, der Zentralzelle und der chalazalen Region der Samenanlagen nachgewiesen werden. Dies ist ein bemerkenswerter Unterschied zu den Ergebnissen der Immunolokalization und der Expression des GFP-AGO9 Fusionsproteins. AGO9 war weder in der MMC oder FM noch in der Eizelle oder Zentralzelle nachweisbar. Das Auftreten aberranten Spleißens mit Intron Retention in der *AGO9* mRNA während der Reifung der Samenanlagen und die Ergebnisse der ektopischen Expression der korrekt gespleißten AGO9 mRNA in der Eizelle weisen darauf hin, dass sowohl entwicklungsspezifisch reguliertes alternatives Spleißen als auch verminderte Proteinstabilität für das Fehlen des AGO9-Proteins in der Eizelle verantwortlich sind. Nur wenige Pflanzen mit GFP-AGO9 Expression in der Eizelle konnten erzeugt werden, diese Expression korrelierte aber nicht mit Entwicklungsdefekten im weiblichen Gametophyten oder im sich entwickelnden Samen. Der Einfluss von ektopisch exprimiertem GFP-AGO9 auf die DNA-Methylierung in der Eizelle bleibt noch zu untersuchen.

Um protein-kodierende Gene und kleine nicht-kodierende RNAs mit möglicher Beteiligung an Prozessen der Zellspezifizierung im weiblichen Gametophyten identifizieren zu können, wurde eine spezialisierte transgene Arabidopsiszellline mit eizellähnlichem Expressionsprofil erzeugt, indem der Transkriptionsfaktor RKD2 ektopisch exprimiert wurde. Im Vergleich zur Kontrollzelllinie wurden 5.511 Gene als differentiell exprimiert identifiziert. Quantitative real-time PCR Analysen mit achtzehnzehn ausgewählten Genen bestätigte die mRNAseq-Expressionsdaten. Ein allgemeiner Blick auf Gene der RNA-silencing Maschinerie enthüllte einige differentiell exprimierte Gene in der RKD2-induzierten Zelllinie, so zum Beispiel hochregulierte Gene die für Proteine kodieren wie die DNA-methyltransferase *DOMAINS REARRANGED METHLYTRANSFERASE 1* (*DRM1*), das Doppelstrang-RNA-Bindeprotein the *DOUBLE-STRANDED RNA BINDING PROTEIN 3* (DRB3), die *HISTONE DEACETLYLASE 18* (*HDA18*), das mögliche Chromatinremodellierungsprotein *CHROMATIN REMODELING 34 (CHR34)* und ein bisher nicht beschriebenes Gen welches der Untereinheit 5A der RNA-Polymerase V ähnelt.

Die Untersuchung der kleinen RNAs in der RKD2-induzierten Zelllinie zeigt eindeutige Unterschiede im Vergleich zum Kontrollkallus. Die häufigste Gruppe kleiner RNAs im Kontrollkallus sind miRNAs, wohingegen im eizellähnlichen Kallus die Biogenese von siRNAs aus mRNAs transposabler Elemente (TEs) vorherrscht. TE-abgeleitete siRNAs der TY3/Gypsy Überfamilie (TEs mit long terminal repeats), welche Mitglieder der *ATHILA*, *ATLANTYS*, und *ATGP* Familie beinhalten, stellen dabei den größten Teil dar, was die Vorstellung unterstützt, das pflanzliche Keimzellen durch siRNAs gegen Transposonaktivität geschützt sind. Die Beobachtung, dass in den mRNA Sequenzdaten keine reads der entsprechenden TEs zu finden sind unterstützt ebenfalls diese Idee. Analyse der differentiellen Expression von miRNAs ergab 96 miRNAs die mindestens um einen log2 Faktor von zwei hoch- oder runterreguliert sind, wobei lediglich neun miRNAs im RKD2-induzierten Kallus hochreguliert sind. Weiterhin wurden 32 neuartige miRNAs entdeckt, wovon 20 aus bereits bekannten miRNA-Vorläufern entstehen, während 12 komplett neuartig sind. Die Voraussage möglicher Zielgene differentiell exprimierter miRNAs ergab einige inverse Korrelationen zwischen miRNA und mRNA Expressionslevels im eizellähnlichen und auch im Kontrollkallus.

Die Entdeckung von bisher nicht beschriebenen Transkripten in der eizellähnlichen Zelllinie, die für sRNA Maschinerie-Komponenten codieren und die an Chromatinmodifikationen beteiligt sind, zusammen mit der Identifizierung differentiell exprimierter bekannter und neuartiger kleiner nicht-kodierender RNAs eröffnet neue, interessante Möglichkeiten für zukünftige Untersuchungen. Diese könnten neue Erkenntnisse darüber hervorbringen, wie Zellen ein reproduktives Zellschicksal annehmen und wie die Spezifizierung der Eizelle vonstattengeht.

3 INTRODUCTION

3.1 Gametogenesis in *Arabidopsis thaliana*

In the life cycle of a plant, two generations are involved: the diploid sporophyte and the haploid gametophyte. In flowering plants, the sporophyte is the dominant generation and the gametophyte is reduced to a few cells: the three-celled male gametophyte (pollen), and 8 nucleated female gametophyte (embryo sac), which are characteristic of most flowering plants. In general, gametogenesis starts with the specification of a spore mother cell, which then meiotically divides into four spores.

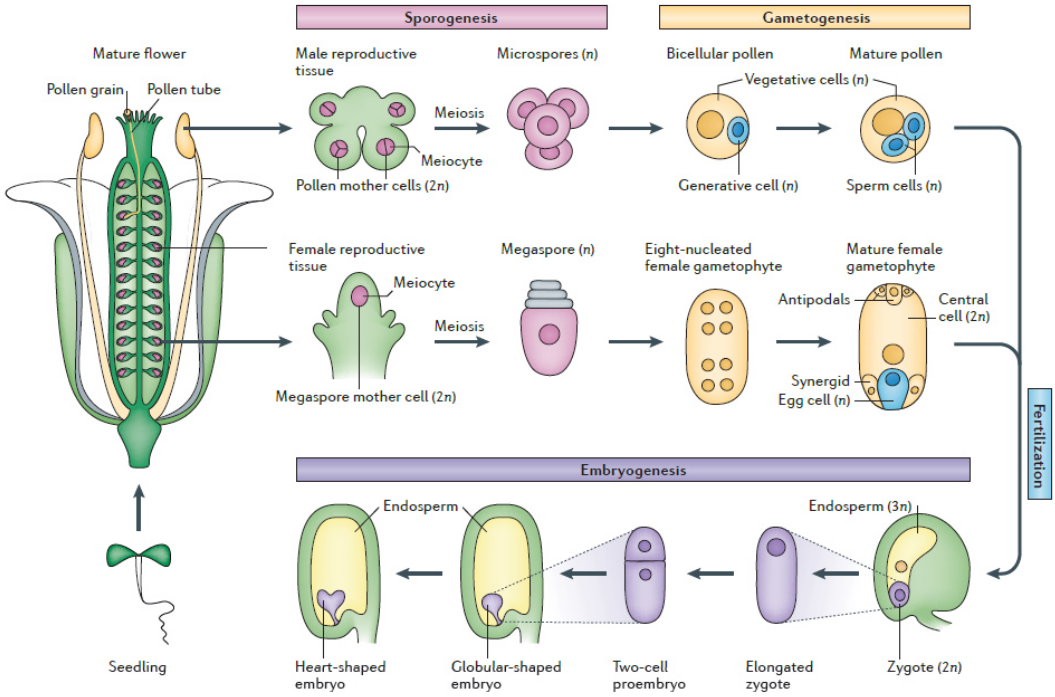


Figure 1: Sexual reproduction in *Arabidopsis thaliana*.
In the flower of *A.thaliana*, male and female reproductive tissues produce pollen mother cells and megaspore mother cells. They are generated in a position-dependent manner from somatic tissues.
During the process of sporogenesis, meiosis takes place, microspores are generated from pollen mother cells and megaspores from megaspore mother cells. In male gametogenesis, the mature pollen is formed. Female gametogenesis results in the formation of the mature female gametophyte, which consists of the egg cell, two synergids, three antipodal cells and a diploid central cell. The egg cell and the central cell are each fertilized by one sperm cell to produce the zygote and the endosperm, respectively. The zygote the elongation of the zygote precedes the first cell division, which initiates a proembryo that develops into the globular-shaped embryo and then the heart-shaped embryo to become the mature seed. Image taken from Kawashima and Berger, 2014.

In male gametogenesis this process is also called microsporogenesis, because the spore mother cell divides into four haploid microspores which first form a tetrad. After separation of the microspores each microspore develops into one pollen grain. The nucleus of a microspore undergoes nuclear migration and asymmetric division to generate the bicellular pollen grain (Berger and Twell 2011), which is composed of a generative and a vegetative cell. In Arabidopsis the generative cell divides once more before anthesis and the mature tri-cellular pollen is formed (Figure 1).

The generation of female gametes occurs within the female gametophyte (or embryosac) during ovule development. First, an archaesporial cell is specified from a cell from of the sporophytic nucellus tissue, which is called the megaspore mother cell (MMC). The MMC is localized directly under the nucellar epidermis and is relativly large compared to the other cells. It then undergoes meiosis to generate four haploid cells, from which three undergo programmed cell death. The remaining cell is called the functional megaspore (FM) and is the progenitor of the whole female gametophyte.

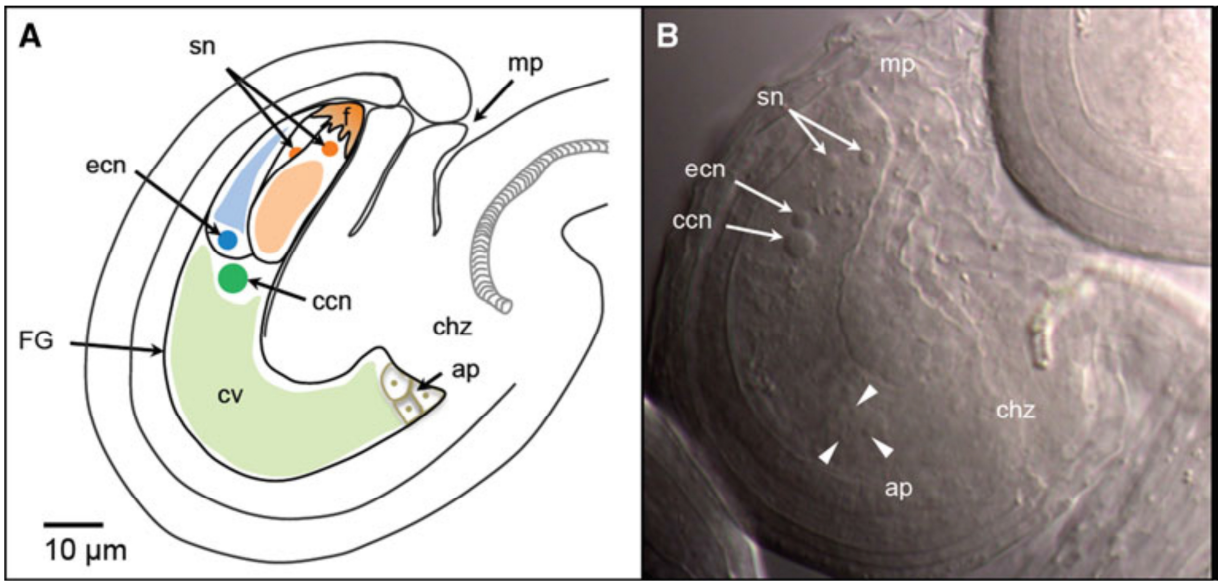


Figure 2: Schematic and microscopic view of a mature Arabidopsis ovule.
(A) Cartoon of an ovule, showing the position of the female gametophyte (FG) within the embedding tissues of the ovule. Within the curved morphology of the female gametophyte the egg cell nucleus (ecn) is always oriented towards the chalazal pole (chz), close to the nucleus of the much larger central cell (ccn). The micropylar region (mp) is the entry site of the pollen tube. The vacuole of the egg cell is oriented towards the micropylar pole. The large central cell vacuole (cv) occupies the chalazal part of the female gametophyte. Next to the egg cell are two synergid cells (in orange), which undergo cell death upon fertilization. The synergid nuclei (sn) are oriented towards the micropylar pole of the FG. On the chalazal pole lie the three antipodal cells (ap), which are a result of the mitotic divisions starting from the functional megaspore. **(B)** Differential interference contrast (DIC) microscopy image of an ovule showing the same cells as in (A). Picture taken from Sprunck *et al.* 2011.

After three rounds of nuclear divisions the female gametophyte has eight nuclei within a syncytium. The nuclei have to be repositioned and after cell specification and cellularization the female gametophyte consists of seven cells: three antipodal cells at the chalazal end, the homo-diploid central cell (resulting from fusion of two polar nuclei), and one egg cell and two synergid cells at the micropylar end (Figure 2). Recent studies showed that the antipodal cells persists through fertilization (Song et al. 2014).

During this whole process of gametogenesis, cell specification events are necessary, but little is known about the genes and molecular mechanisms involved. For example, the mechanisms of how somatic cells aquire a reproductive cell fate are not fully understood (reviewed in Yang et al. 2010). However, a few genes that influence cell fate during gametogenesis have been identified. The *SPOROCYTELESS* (*SPL*) gene of *Arabidopsis* has been reported to play a role in megasporogenesis, as *spl* mutants are able to initiate formation of MMCs but the meiosis of the MMCs does not occur, resulting in a complete lack of a germline. Later during the process of cellularization and cell specification, the nuclei within the embryo sac need positional information to aquire their cell fate. This positional information can be provided by polar expression of proteins. DEMETER (DME), a DNA glycosylase, which can also remove 5-methylcytosine, is localized in the micropylar region of the embryo sac before cellularization but restricted to the central cell after cellularization (Choi et al. 2002). Furthermore, an auxin gradient was suggested to provide positional information during female gametophyte development (Pagnussat et al. 2009), but a recent study provided evidence that only a shallow auxin gradient can be maintained in the female gametophyte (Lituiev et al. 2013). Mutagenesis of an egg cell-specific marker line identified three mutants which show misregulation of the egg cell reporter gene expression: *lachesis* (*lis*), *clotho* (*clo*) and *atropos* (*ato*). Loss of *lachesis* function results in expression of the egg cell marker in the synergid and the central cell (Groß-Hardt et al. 2007). Notably, *lachesis, clotho and atropos* are all genes encoding components of the splicing machinery. This is a hint that RNA splicing may play a role in cell specification during gametogenesis.

A transcriptome analysis of isolated egg cells revealed the enrichment of certain functional groups of proteins in the egg cell (Wuest et al. 2010). Especially genes involved in RNA interference are highly abundant in the egg cell, including genes encoding factors involved in biogenesis of microRNAs like *DICER-LIKE 1 (DCL1)* and *HYPONASTIC LEAVES (HYL1),* but also ARGONAUTE 1, a protein involved in gene silencing mediated by small RNAs (Wuest et al. 2010).

3.2 Argonaute proteins

A forward genetic screen for genes involved in developmental processes generated a mutant with extraordinary appearance (Böhmert el al. 1998). Seedlings had pointy, unexpanded cotyledons and a dark green hypocotyl. This phenotype reminded the researchers of a tiny squid and so this mutation was called Argonaute. Many homologues of Argonaute proteins were found in other species and it became clear that AGOs form a conserved family of proteins, which can be found in Archaea, Bacteria and almost all eukaryotes.

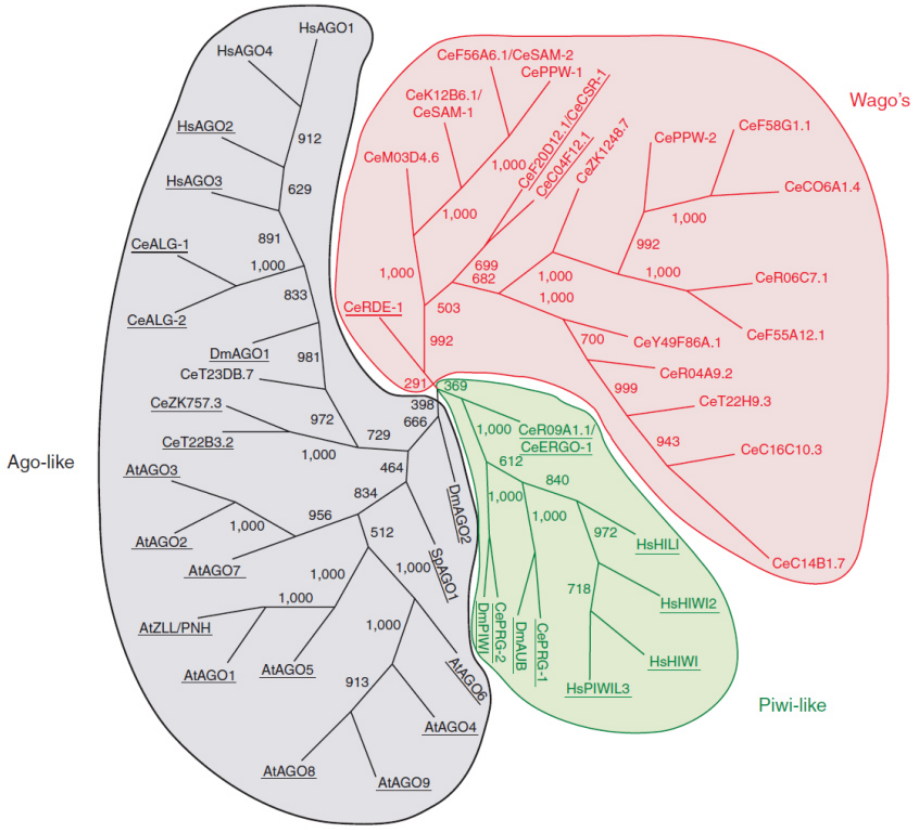


Figure 3: The three clades of the Argonautes.

The Ago-like clade, shown in black, is based on similarity to AtAGO1. The Piwi-like clade, shown in green, is based on similarity to *Drosophila melanogaster* Piwi and the Wago's, which is a worm-specific clade, is shown in red. Argonaute proteins that contain a complete catalytic motif are underlined. Image taken from Joshua-Tor and Hannon 2010.

AGOs can be grouped into three big subfamilies; the AGO group which is close to the Arabidopsis AGO1, the PIWI group containing the P-element Induced Wimpy testis protein from *Drosophila melanogaster,* and a third group which can be only found in *Caenorhabditis*

elegans, the WAGO or Worm AGO group (Hutvagner and Simard 2008, Joshua-Tor and Hannon 2010). Although almost every organism has at least one Argonaute protein but the total number of proteins varies strongly between species (Figure 3): the human genome contains eight Argonaute proteins, four from every subgroup. In Drosophila, five Argonaute proteins can be found (Höck and Meister, 2008), three AGOs and two PIWI-like proteins. *Saccharomyces cerevisiae*, a widely used model organism has no Argonaute protein, while its close relative *Schizosaccharomyces pombe* possesses one. The nematode *Caenorhabditis elegans* has 5 AGO, 4 PIWI and 18 WAGO proteins. The model plant *Arabidopsis thaliana* possesses ten different AGO proteins, but no members of PIWI-clade. Plant Argonautes will be discussed in detail in Chapter 3.2.6.

3.2.1 Structure of Argonaute proteins

The domain structure within the Argonaute protein family is highly conserved. In general, four domains characterize those proteins, the N-terminal, the PAZ domain (named after the PIWI, Argonaute and Zwille proteins containing it), the MID domain and the PIWI (P-element Induced Wimpy testis) domain (Figure 4A). Crystallization studies with prokaryotic AGOs provided insight into the overall protein structure (Tolia and Joshua-Tor 2007, Yuan et al. 2005). The tertiary structure of AGOs allows them to bind small RNAs and DNAs in a bilobal structure which is formed by the N-terminal and PAZ-domain on one side and the MID and PIWI domain on the other side. Figure 4 shows the domains and the tertiary structure of human AGO2 (Schürmann et al. 2013).

Each of the domains has its own specific function: the N-terminal domain takes not only part in forming the bilobal structure, but it also contains regulatory motifs, which can influence the activity of catalytic domains (Schürmann et al. 2013). The PAZ domain has been proven to bind small RNAs via the 3´ end, which was first shown for the Drosophila AGO2 protein (Lingel et al. 2003, Lingel et al. 2004). The MID domain is located between the PAZ and the PIWI domain. This domain binds small RNAs via the 5´ phosphate (Ma et al. 2005). Additionally, a highly conserved (m7G) cap-binding motif can be found in the MID domain, which is generally found in eukaryotic translational factor 4E (Kiriakidou et al. 2007). The PIWI domain has a strong similarity to RNAse H proteins, which are endonucleases. Some AGOs have an endonucleolytic activity and can cleave RNA targets that show perfect complementarity to the AGO bound small RNA. This activity is also referred to

as slicer activity. Three catalytic residues were found to be necessary for this endonucleolytic activity, the DDX motif consists of two aspartic acid residues (DD) and one histidine or another aspartic acid residue (X) (Liu et al. 2004, Rivas et al. 2005). Lately, it has been revealed that a fourth residue is important for slicer activity: crystallization studies of *Kluyveromyces polysporus* Argonaute together with its bound RNA showed that a conserved a glutamate residue, which protrudes into the catalytic pocket, is needed to stabilize the active catalytic center of the PIWI domain (Nakanishi et al. 2012b). Although the PIWI domain and its catalytic residues are needed for the slicing activity of Argonaute proteins, they are not the only requirement: also the N-terminal domain contributes to the slicing activity. Studies with chimeric human AGOs revealed the importance of two short sequence elements within the N-terminal domain (Hauptmann et al. 2013).

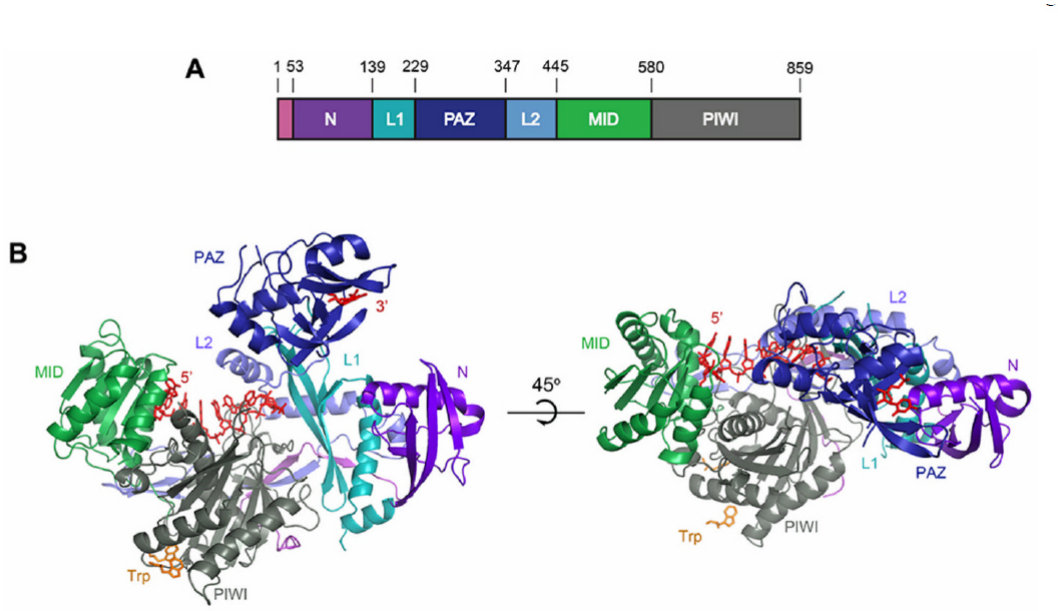


Figure 4: Structure of human AGO2.
(**A**) Domain structure of human AGO2 showing the N (purple), PIWI, Argonaute and Zwille =PAZ (navy), MID (green), P-element Induced Wimpy testis=PIWI (grey) domains and linkers L1 (teal) and L2 (**B**) Front and top views of AGO2 A generic guide RNA (red) can be traced for nucleotides 1-8 and 21. Tryptophan molecules (orange) bind to tandem hydrophobic pockets in the PIWI domain. Image taken from Schirle and McRae 2012.

3.2.2 Subcellular localization

Argonaute proteins localize to cytoplasm as well as to the nucleus. In the cytoplasm they are often concentrated in so-called processing bodies (P-bodies), in those defined regions they colocalize with proteins involved in mRNA decay, silencing and translational repression, like for example XRN1 (5'-3' exoribonuclease 1), an exonuclease, decapping enzymes and

deadenylases (Eulalio et al. 2007a). PIWI-clade AGOs are also present in a diffuse manner in the cytoplasm (Eulalio et al. 2007b). Mammalian AGO2 was reported to be concentrated in another cytoplasmic structure called stress-granule (Leung et al. 2006).

The presence of AGOs in the nucleus and their function in this cellular compartment has remained unclear for a long time and especially the presence of AGOs in the human nucleus has been a great matter of debate. A recent study revealed that human AGO2 colocalizes in the nucleus with other components of the RNAi machinery like for example DICER and GW182 (Gagnon et al. 2014). In plants, the Arabidopsis AGO4 is localized in the nucleus in specialized structures within the nucleolus; the so-called Cajal bodies (Li et al. 2006). Arabidopsis AGO1 has been reported to be at least partially associated the endoplasmic reticulum (Li et al. 2013).

3.2.3 Functions of Argonaute proteins

Argonaute proteins play a central role in regulatory pathways guided by small RNAs (sRNAs). Based on the sequence of the sRNA, the AGO/sRNA particle can bind to its complementary target and then regulate genes in different ways. AGOs can degrade mRNA via their slicer activity, but are also often incorporated into protein complexes, which are called RNA-induced silencing complexes or RISC (Hammond et al. 2001). Within those RISC complexes AGOs interact with other effectors of RNA silencing, for example poly(A)-binding proteins, decapping enzymes, deadenylases and 5´ to 3´ exonucleases (Braun et al. 2012). The interaction of AGOs with those effector protein within RISC complexes is mediated by the class of TNRC6 proteins (also known as GW proteins or GW182 family proteins), which interact with AGOs via conserved WG/GW motifs (Eulalio et al. 2008).

Different types of RISC complexes can be distinguished considering the type of small RNAs they bind and the protein components they consist of. For example the miRISC contains micro RNAs (miRNAs) and the siRISC incorporates small interfering RNAs (siRNAs). The role of miRISCs and siRISC is explained in the following sections.

3.2.4 The microRNA pathway

The biogenesis of microRNAs begins with a primary transcript produced by RNA-polymerase II. Those transcripts exhibit a characteristic structure, which contains one or more hairpins. The mature miRNA sequence is located within this hairpin. The primary transcript is processed by the microprocessor complex, which separates the hairpin from the transcript.

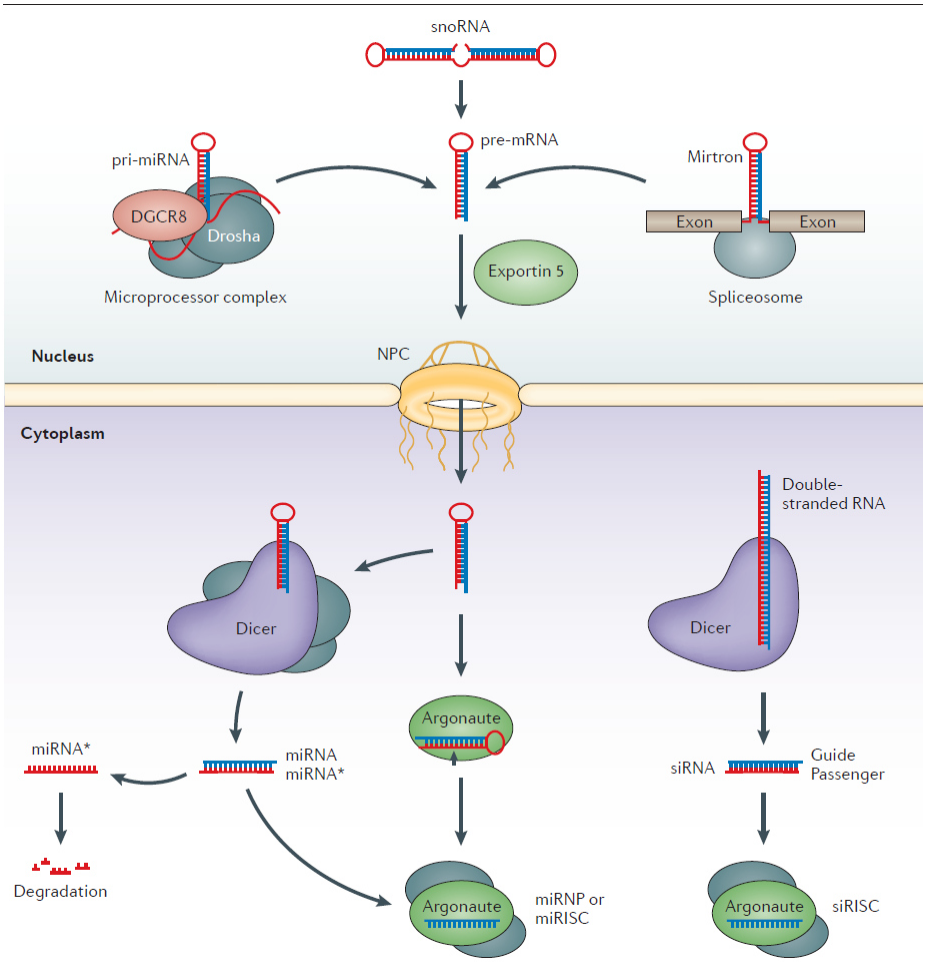


Figure 5: Biogenesis of miRNAs and siRNAs.
miRNAs and siRNA are generated by different mechanisms. Biogenesis of microRNAs (miRNAs) begins with a hairpin structure which processed in the nucleus by the microprocessor complex. The pre-mRNA is exported into the cytoplasm, processed by DICER and after degradation of the passenger strand the miRNA is incorporated into the RNA-induced silencing complex (RISC, or miRISC for miRNA-containing RISC or miRNP for microribonucleoprotein), Small interfering RNAs (siRNAs) are processed in the cytoplasm from double-stranded RNAs with the help of DICER. The guide strand is incorporated into RISC (or siRISC for siRNA-containing RISC). *DGCR8*, DiGeorge syndrome critical region gene 8. Image taken from Meister 2013.

Within the microprocessor complex the RNAse III like protein DROSHA is the catalytic component and it requires the cofactor DiGeorge syndrome critical region gene (DGCR8). After being transported into cytoplasm via Exportin 5, the pre-miRNA undergoes

further processing by Dicer, another RNAse III like protein Dicer removes the loop of the hairpin and leaves a double-stranded RNA of 21-22 nucleotides (nt) length. One strand of this dsRNA (miRNA*) is then degraded and the remaining strand (the mature miRNA) is incorporated into a miRNA-induced silencing complex (miRISC) and bound by an Argonaute. The miRISC can then be active as an RNP and degrade mRNAs, which have to be perfectly complementary to the bound miRNA. miRISCs can also mediate translational repression by preventing interaction of initiation factors with mRNAs (Humphreys et al. 2005) and cause deadenylation and subsequent decapping and degradation of the target mRNA (Fabian and Sonenberg 2012).

3.2.5 The siRNA pathway

Primary transcripts of small interfering RNAs (siRNAs) are generated from intergenic repetitive elements, pseudogenes, or endogenous siRNA clusters. They form long dsRNAs, which are transported into the cytoplasm and processed by DICER to generate a short double-stranded RNA of 20 to 30 nucleotides length (Kim et al. 2009). Only one strand of this dsDNA (the guide strand) is incorporated into an Argonaute within a siRISC effector complex (Figure 5). This siRISC targets and cleaves mRNAs with perfectly complementarity to the bound siRNA (Meister and Tuschl 2004). siRNAs from endogenous origins are also called endo-siRNAs. Exogenous siRNAs (exo-siRNAs) can be derived from viruses or transgenes (Meister 2013). *Drosophila melanogaster*, for example, uses a siRNA mechanism to defend itself against double stranded RNA viruses (Kingsolver et al. 2013). Small interfering RNAs take part in post-transcriptional and transcriptional gene silencing. In *Schizosaccharomcyes pombe* the only AGO protein forms a silencing complex at centromere regions, leading to histone methylation (Volpe et al. 2002).

3.2.6 Argonaute proteins of *Arabidopsis thaliana*

The model plant *Arabidopsis thaliana* possesses ten different AGO proteins. In the context of eukaryotic Argonautes, they all belong to the AGO subfamily. Phylogenetic analyzes group the Arabidopsis AGOs into three protein clades (Vaucheret 2008), the AGO1, AGO5, and AGO10 clade; the AGO2, AGO3, and AGO7 clade; and the AGO4, AGO6,

AGO8, and AGO9 clade (Mallory and Vaucheret, 2013) A phylogenetic tree comprising all Arabidopsis Argonautes is shown in Figure 6.

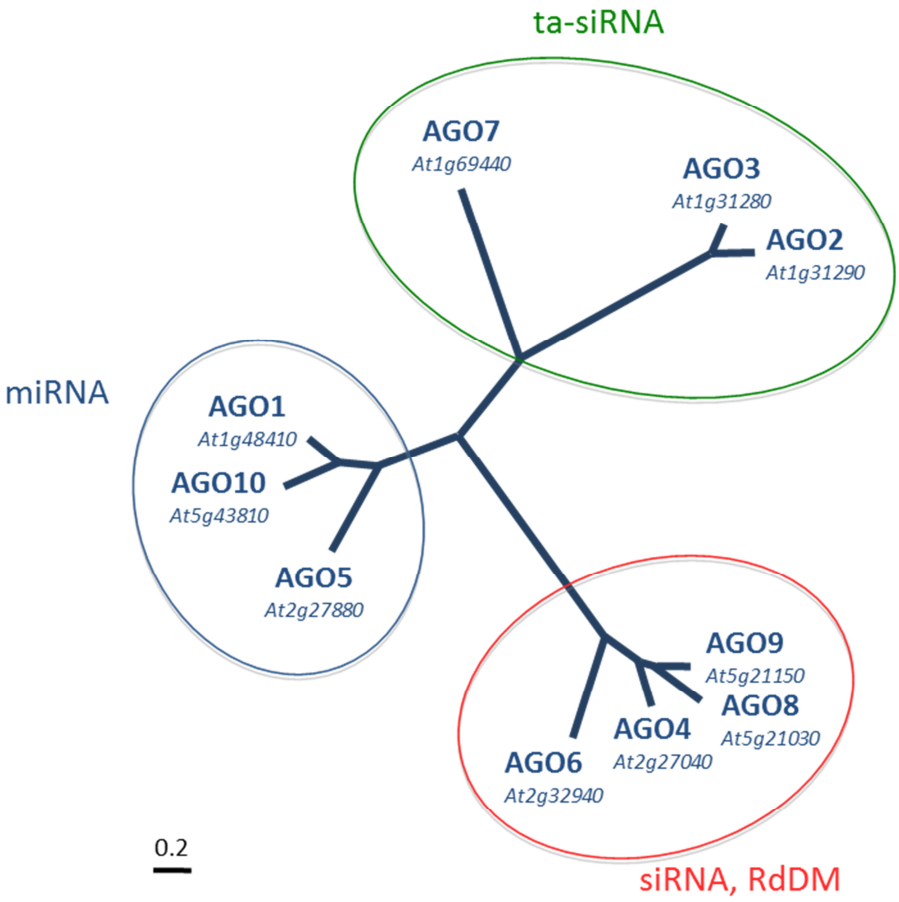


Figure 6: Phylogenetic tree illustrating the three clades of Arabidopsis Argonautes.
AGO1 and AGO10 act in the classic canonical miRNA pathway and the close relative AGO5 is involved in a novel miRNA pathway in the plant germline (Chapter 3.2.9). AGO7 binds miR390 is involved in the generation of trans-acting small interfering RNAs (ta-siRNAs), AGO2 is involved in antiviral and antibacterial defense and binds miRNAs beginning with an adenosine. AGO3 is part of a tandem repeat with AGO2 and it function is unknown. AGO4, AGO6 and AGO9 predominantly bind 24 nt siRNAs with 5' adenosine and are involved in the mechanism of RNA-directed DNA-methylation (RdDM). AGO8 is generally regarded as a pseudogene. Image by Stefanie Sprunck.

AGO1 is the main effector in the classic canonical miRNA pathway in plants, which follows similar principles as described above for animals. One difference of the miRNA pathway in plants is the fact that the microprocessor complex of *Drosophila* and mammals is not part of it (Denli et al. 2004). Instead, the Pol II transcribed miRNA hairpin is directly processed by a protein complex containing DCL1 (Figure 7A), which is one of four Arabidopsis homologues of the human DICER protein, methylated at the 2´O by HEN1, and

transported out of the nucleus with the assistance of HASTY, a homolog of Exportin 5 (Xie 2015). Finally, this results in 21-22 nucleotides long miRNAs which are bound by AGO1. In plants, microRNAs bound by AGO1 are characterized by a 5´ uracil (U) nucleotide and indeed, the identity of this nucleotide is important for sorting different small RNAs into their respective AGOs (Mi et al. 2008, Havecker et al. 2010). Like in other eukaryotes, AGO1 can, as part of a RISC complex, cleave mRNA targets (Baumberger and Baulcombe 2005) or mediate translational repression (Li et al. 2013). Beside these canonical pathways, AGO1 can also trigger the biogenesis of secondary siRNAs (Creasy et al. 2014). MiR/AGO1 complexes are involved in an initial cutting step which generates a template for an RNA-dependent RNA polymerase (RDR6), the resulting dsRNA is then processed into secondary siRNAs. This pathway is important for the defense against epigenetically reactivated transposons (Creasey et al. 2014).

AGO10 has been proposed to compete with AGO1 for miRNAs and thus keep up the expression of the HD-ZIP-III transcriptions factors, which is important for the maintenance of cell identity in the region of the shoot apical meristem (Zhang and Zhang 2012).

AGO5 shows a specific expression pattern in reproductive tissues of Arabidopsis (Mallory and Vaucheret 2010). The role of AGOs in plant reproduction will be discussed in detail in Chapter 3.2.9. Besides that, a recent work from Brosseau and Moffett (2015) showed that *AGO5* expression is induced by infection with potexvirus and is needed, together with *AGO2*, to restrict the viral infection.

AGO7 is the prominent member of the second clade, the so-called ta-siRNA-clade, which stands for trans-acting small interfering RNA. AGO7 is involved in the specialized ta-siRNA silencing pathway, in which the respective AGOs interact with their target via miRNA/mRNA interaction and exert an initial cleavage step on the transcript (Figure 7B). This processed transcript serves then as a template for an RNA-dependent RNA polymerase (RDR6), the resulting double-stranded RNA is cut by DCL4 resulting into a number of small dsRNAs of 21 to 22 nucleotides length. In the end, those RNAs can be incorporated into AGO1-RISC again to attack mRNA targets. In case of AGO7 only one miRNA is bound, which is miR390. The AGO7/miR390 complex then cleaves the mRNA of the *TAS3* gene, the resulting ta-siRNAs act via the miRNAs pathway to inhibit the auxin response factors ARF2, ARF3 and ARF4, which are involved in regulation of lateral root growth (Marin et al. 2010).

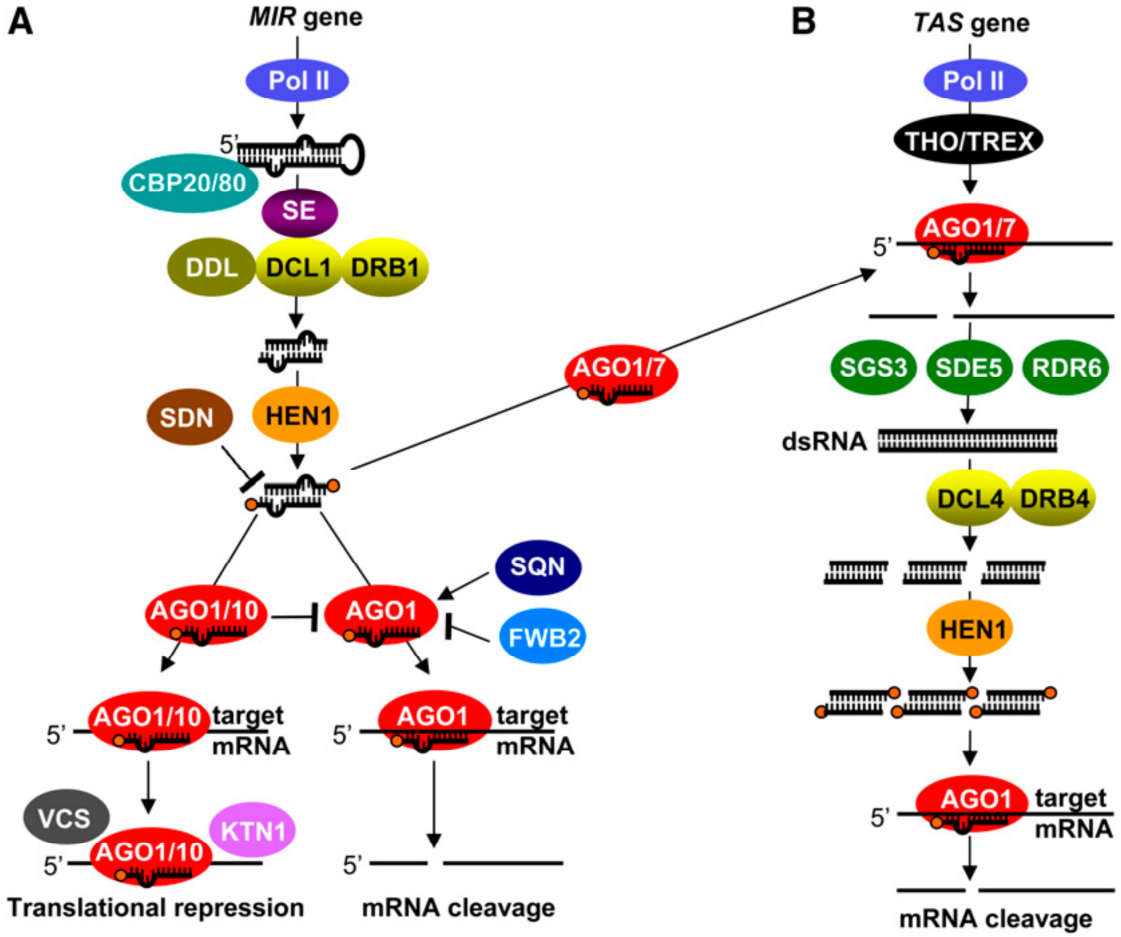


Figure 7: Trans-Silencing: the miRNA and ta-siRNA pathways.
(A) Micro RNA genes are transcribed by RNA polymerase II (PolII) and form a hairpin structure which is directly processed by a complex containing DCL1 and DRB1. The miRNA is then incorporated into an AGO1 or AGO10 containing RNA-induced silencing complex (RISC), which mediates translational repression or mRNA cleavage. **(B)** *TAS* genes are transcribed from PolII to form single-stranded RNAs, which are targeted by microRNAs, cleaved by AGO1 or AGO7 and copied into dsRNA by RDR6, SGS3, and SDE5, The dsRNA is processed into 21 to 22 nucleotide ta-siRNAs by the DRB4/DCL4 complex. The ta-siRNAs direct cleavage of complementary mRNAs. PolII, RNA-polymerase II; CBP20/80; cap binding protein20/80; SE, Serrate; DDL, DAWDLE; DCL1, dicer-like 1; DRB1; dsRNA-binding protein1; SDN, small RNA degrading nuclease; HEN1, HUA enhance 1; AGO1/10, Argonaute 1/10; SQN, squint; FWB2, F-BOX WITH WD-40 2; VCS, varicose, KTN1, katanin 1; THO/TREX, THO/Transcription export, SGS3, suppressor of gene silencing 3; SDE5, silencing defective 5; RDR6, RNA-dependent RNA polymerase 6; DCL4, dicer-like 4; DRB4, dsRNA-binding protein1. Image taken from Mallory and Vaucheret 2010.

Considering the fact that those small RNAs can trans-act on their target molecules and share the biogenesis of siRNAs, they are called ta-siRNAs.

AGO2 is reported to play a role in antiviral (Harvey et al. 2011) and antibacterial defense (Zhang et al. 2011). It predominantly binds miRNAs with a bias for a 5' adenosine (A) (Takeda et al. 2008, Mi et al. 2008). Additionally, it associates with virus-derived siRNAs (Harvey et al. 2011) and is involved in repair of DNA double strand breaks (Wei et al. 2012).

AGO3 is located closely to AGO2 on chromosome 2 and those two proteins are very similar to each other, which is a hint towards recent gene duplication (Vaucheret 2008).

The AGO4 clade AGOs are functionally associated with RNA-directed DNA methylation (RdDM), an epigenetic process in which 21-24 nucleotide siRNAs guide methylation of homologous DNA loci. For AGO4, AGO6 and AGO9 the predominant binding of siRNAs has been shown (Havecker et al. 2010), while AGO8 is regarded as a pseudogene (Takeda et al. 2008). The AGO4 clade AGOs will be discussed in detail in the next Chapter.

3.2.7 RNA-directed DNA-methylation and transposon silencing in plants

The AGO4 clade of Arabidopsis comprises AGO4, AGO6, AGO8, and AGO9. AGO4, AGO6 and AGO9 are characterized by the fact that they predominantly bind siRNAs of 24 nucleotides length (Mi et al. 2008, Borges and Martienssen 2015). Those siRNAs represent a central component of a plant specific siRNA pathway, which creates epigenetic modifications through siRNA guidance. This was first discovered in Arabidopsis plants which were infected with RNA viroids and it could be shown that viral RNA could direct DNA methylation towards specific genomics loci (Wassenegger 1994). This phenomenon was called RNA-directed DNA methylation (RdDM). Later it was discovered that components of the RNA silencing machinery are required for this mechanism (Chan 2004). A study by Zilbermann et al. (2003) identified AGO4 as a central component of this pathway. RdDM requires the plant specific RNA polymerases Pol IV and Pol V. Pol V generates the precursor transcripts, which are then converted into a double-stranded RNA (dsRNA) by RDR2 and processed by DCL3 (Haag and Pikaard 2011). The resulting small dsRNA is bound by an AGO4-clade Argonaute and the passenger strand is degraded (Ye et al. 2012). For AGO4 is has been shown that been shown that a nuclear localization sequence in the N-terminus of the protein is then accessible for nuclear import factors and AGO4 is imported back into the nucleus (Ye et al. 2012). The AGO4–RNP then interacts with Pol V transcripts and forms a complex with several RdDM–factors, like for example the largest Pol V subunit, the DNA and RNA-binding protein KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1), which is also known as SUPPRESSOR OF TY INSERTION 5-LIKE (SPT5L), and the DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2) which catalyzes the *de novo* methylation of DNA (Figure 8).

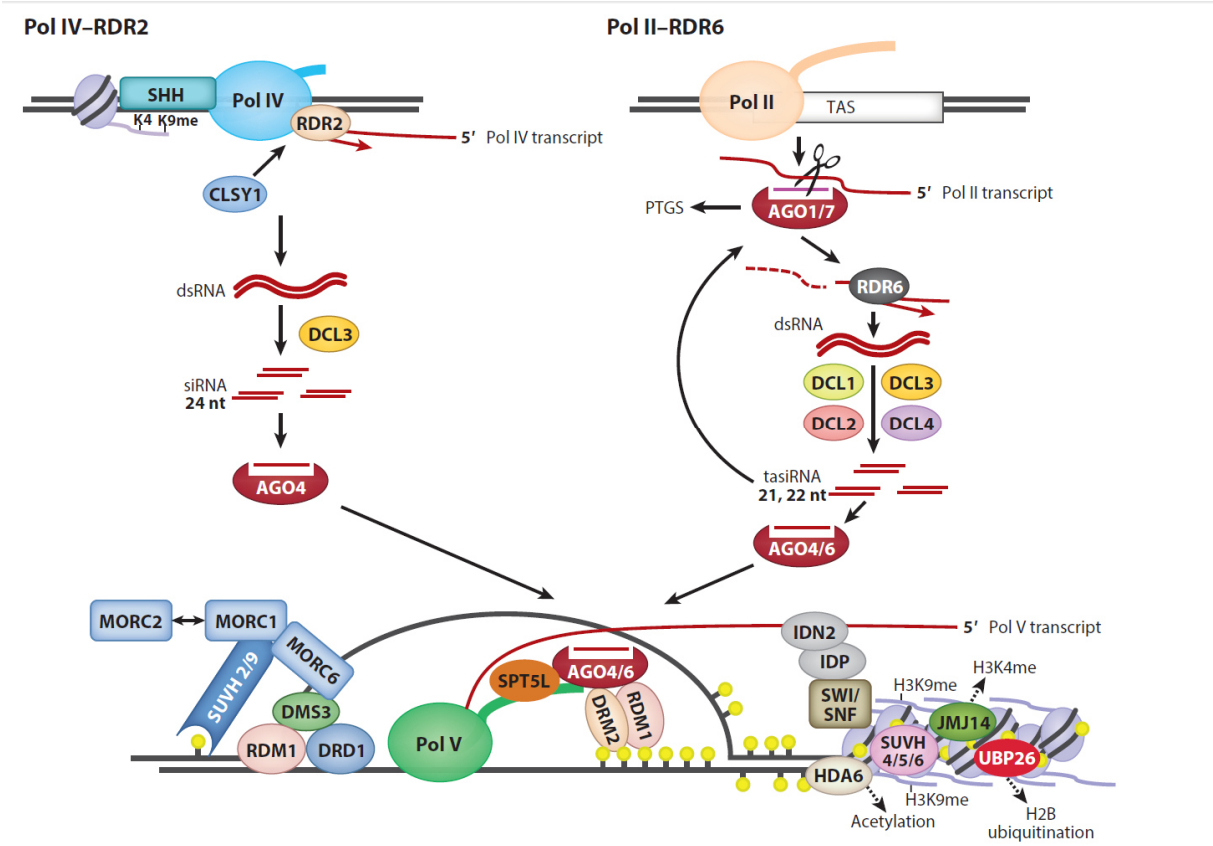


Figure 8: The RdDM pathway in *Arabidopsis thaliana*.
RNA-directed DNA methylation is a complex pathway of transcriptional silencing, preferentially directed towards transposons and other repetitive elements. The canonical RdDM pathway involves POL IV transcripts, which are converted into double-stranded RNA by the RNA-dependent RNA-polymerase RDR2. This dsRNA is processed into 24 nt long siRNA which cooperate with AGO protein and Pol V transcripts to direct DNA methylation. The Pol II/RDR6 pathway uses siRNAs generated by POLII, AGO1/7 and RDR6 to guide an AGO protein to the Pol V transcript. In both cases, DNA is methylated by DOMAINS REARRANGED METHYLTRANSFERASE 2. (DRM2). SHH, SAWADEE HOMEODOMAIN HOMOLOG; RDR2, RNA-DEPENDENT RNA-POLYMERASE 2; CLSY1, CLASSY1; DCL3, DICER-LIKE 3; AGO4, ARGONAUTE 4; MORC, MICRORCHIDIA; SUVH, SU(VAR)3-9 HOMOLOG; DMS3, DEFECTIVE IN MERISTEM SILENCING 3; RDM1, RNA-DIRECTED DNA METHYLATION 1; DRD1, DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1; SPT5L, SUPPRESSOR OF TY INSERTION 5-LIKE; IDN2, INVOLVED IN DE NOVO 2; IDP, IDN2 PARALOG, SWI/SNF, SWITCH/SUCROSE NON-FERMENTABLE; HDA6, HISTONE DEACETYLASE 6, JMJ14, JUMONJI 14, UBP26, UBIQUITIN-SPECIFIC PROTEASE 14, DCL, dicer-like. Image taken from Matzke et al. 2015.

This mechanism especially silences genes where POL V is active, such as transposons and other repetitive elements like telomeres (Verdel et al. 2004). The role of AGO6 is partially redundant with that of AGO4, it is reported to take part in transcriptional gene in root and shoot meristems (Havecker et al. 2010, Eun et al. 2011). A variant of this canonical RdDM pathway relies on RNA polymerase II (PolII) transcripts, which are subjected to an

initial cleavage step by an AGO, followed by generation on a dsRNA by RDR6, like in the TASI pathway (Figure 8).

Each of the of the four DCLs can generate 21 to 22 nt long ta-siRNAs, which are bound by one of the RdDM AGOs in order to initiate *de novo* DNA methylation via DRM1. In spite of this redundancy there seem to be differences in subnuclear localization and AGO6 seems to be even more connected with POL V than AGO4 (Figure 9). McCue et al. (2014) proposed that AGO6 plays a role in establishment of transposable element silencing. AGO9 is also a member of the AGO4 clade and binds 24 nt siRNAs, with a bias to a 5´ A.

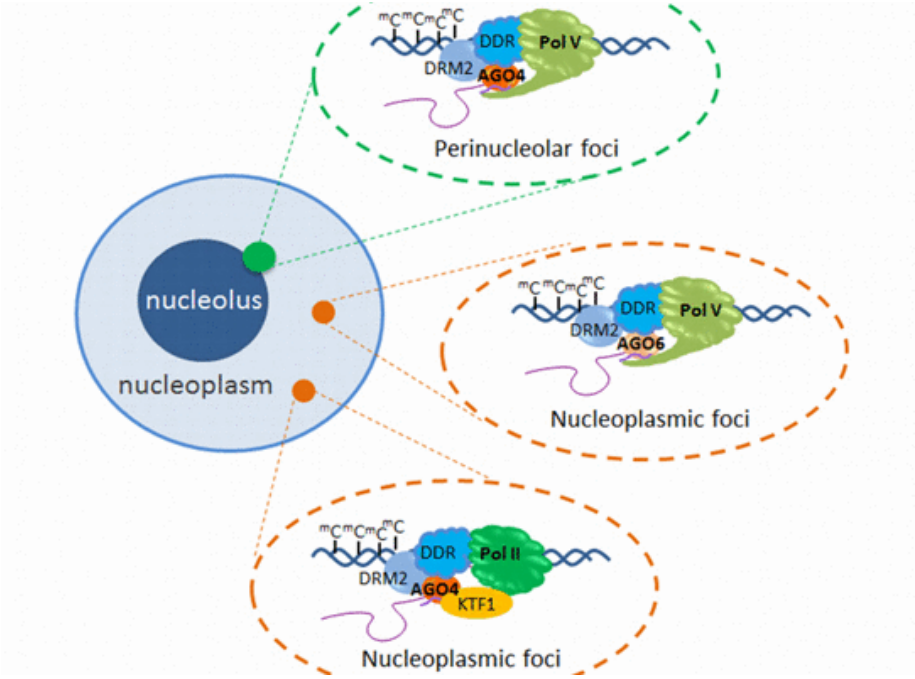


Figure 9: Distinct roles of AGO4 and AGO6 Proteins
AGO4 and Pol V are involved in RNA-directed DNA methylation (RdDM) in perinucleolar foci. RdDM activities in the nucleoplasm are mediated by AGO6 and Pol V, and by AGO4 and Pol II. DRM2, DOMAINS REARRANGED METHYLTRANSFERASE 2; DDR, a protein complex formed by DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1), RNA-DIRECTED DNA METHYLATION 1 (RDM1) and DEFECTIVE IN MERISTEM SILENCING 3 (DMS3). KTF1, KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1. Image taken from Duan et al. 2014.

Although its role in RdDM has not been clarified, AGO9 has been shown to silence repetitive elements (Olmedo-Monfil et al. 2010). Additionally, the enhanced sensitivity of the *ago9-1* mutant against reagents that cause DNA double strand breaks could indicate a role of AGO9 in DNA repair (Oliver et al. 2014). *AGO8* is found in close proximity to *AGO9* on chromosome 5 and is generally believed to be a pseudogene (Takeda et al. 2008). Silencing of

TEs is not only achieved by RdDM, which is a mechanism for *de novo* DNA-methylation occurring either after removal of DNA-methylation marks or after fresh insertion of TEs in the genome. Silencing of TEs is achieved by setting epigenetic marks on TE sequences and also other epigenetic mechanisms play a role in keeping TEs under control (see next Chapter).

3.2.8 Epigenetic reprogramming in plant sporo- and gametogenesis

The term "epigenetics" refers to heritable changes in gene expression, which are not caused by a change in DNA sequence. The biochemical modifications within an organism related to epigenetic changes are often called epigenetic marks. Silencing of transposons by modification of histones and DNA is only one part of the process of epigenetic silencing. Those marks can consist of methylation of DNA and modification of histones and have an influence on transcriptional activity and chromatin structure (Feng et al. 2010). DNA methylation occurs at specific DNA sites: CG, CHG, and CHH (where H = C, T, or A). In plants, DNA methylation in a GC-context is maintained by MET1 (DNA METHYLTRANSFERASE 1), with the help of the cofactor VIM (VARIATION OF METHYLATION). CHG-methylation is performed by CMT3 (CHROMOMETHYLASE 3), which interacts with the histone methyltransferase KYP (KRYPTONITE), forming a positive feedback loop. The maintenance of DNA methylation by CMT3 and MET1 can be distinguished from the *de novo* mechanism of RdDM (Chapter 3.2.6), which is active in all sequence contexts. In general, epigenetic marks have an important role in the formation of male and female gametes in plants as well. For example, the trimethylation of histone 3 at lysine 4 (H3K4) as an epigenetic mark is important for the maintenance of certain chromatin structures and is needed for gametophyte development (Berr et al. 2010). Until now, the state of epigenetic marks in plant meiocytes is unknown, but the increase of expression of transposable elements in male meiocytes indicates a change in histone modification and non-CG DNA-methylation, which normally keep TE expression under control (Kawashima and Berger 2014). In the central cell of the mature female gametophyte, MET1 is repressed, while DEMETER (which removes 5-methylcytosine) is active (Jullien et al. 2012). Thus, a general state of DNA-hypomethylation in the central cell is likely. The consequence of this hypomethylated state would be the transcriptional activation of transposable elements and other elements, like imprinted genes. In the egg cell, the maintenance DNA methyltransferases CMT3 and MET1 are barely detectable (Jullien et al. 2012). But on the

other hand, the *de novo* DNA methyltransferases DRM1 and DRM2 can be found in the egg cell, with DRM1 exclusively expressed in the egg cell (Jullien et al. 2012, Kawashima and Berger).

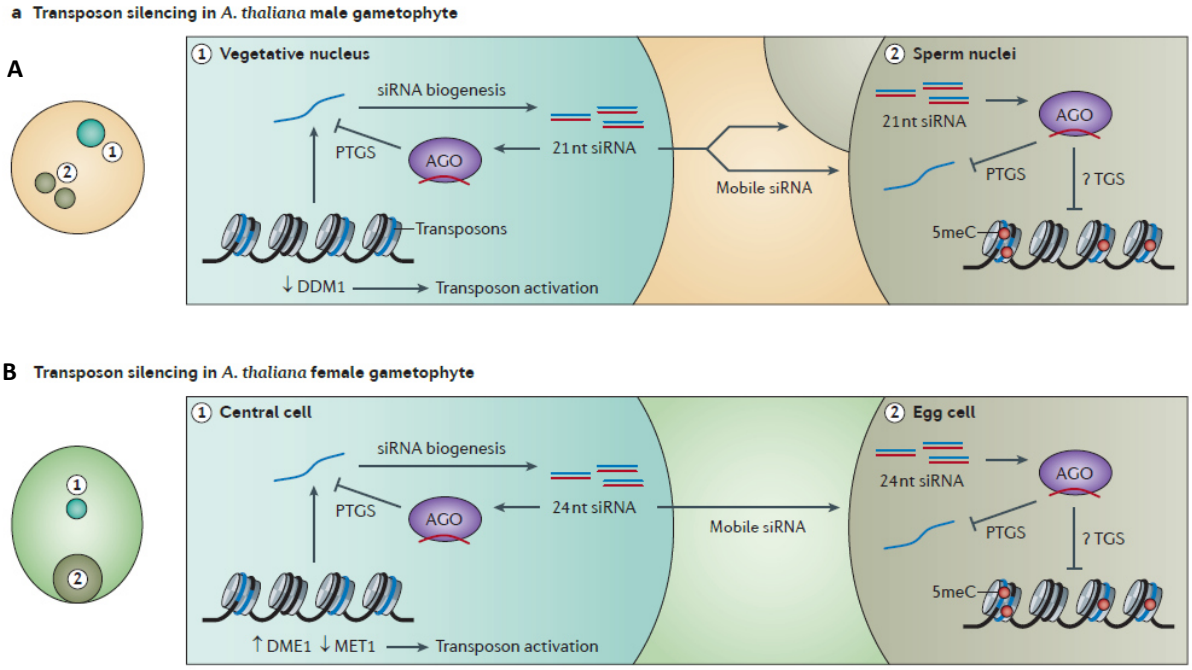


Figure 10: Transposon silencing in Arabidopsis male and female germline.

(A) In the *Arabidopsis* male gametophyte the chromatin remodeler DDM1 is repressed in the vegetative nucleus, allowing transposons to be expressed and mobile siRNAs to be generated. 24 nt siRNAs move to the sperm nuclei to trigger post-transcriptional gene silencing **(B)** In the female gametophyte, the central cell shows a general state of hypomethylation caused by repression of the DNA-methyltransferase MET1 and the demethylase DEMETER (DME1). This allows transposon activation and generation of mobile siRNAs. 21 nt siRNAs move to the egg cell nuclei to trigger post-transcriptional gene silencing. PTGS, post-transcriptional gene silencing; DDM1, DECREASED DNA METHYLATION 1; AGO, Argonaute; TGS, transcriptional gene silencing; 5meC, 5-methylcytosine, DME1, DEMETER; MET1, METHYTRANSFERASE1. Image taken from Castel and Martienssen 2012.

De novo DNA methylation via the RdDM pathway thus seems to play an important role in the egg cell. In the male gametophyte, de-repression of TEs in the vegetative nucleus of the pollen grain correlates with the repression of TEs in the sperm cell (Calarco et al. 2012). In plants, this involves movement of siRNAs from the vegetative nuclei to the sperm nuclei in the male gametophyte and movement of siRNAs from egg to central cell in the female gametophyte (Figure 10).

3.2.9 The role of AGO proteins in plant reproduction

Transcriptomic analyzes revealed that some *AGO* genes show a highly specific expression pattern in different cell types and developmental stages of plant development (Schmid et al. 2005, Mallory and Vaucheret 2013). For example, *AGO5* and *AGO9* transcripts are strongly present in reproductive tissues. In case of *AGO5*, the mutant *ago5-4* shows severe defects in the initiation of megasporogenesis (Tucker et al. 2012).

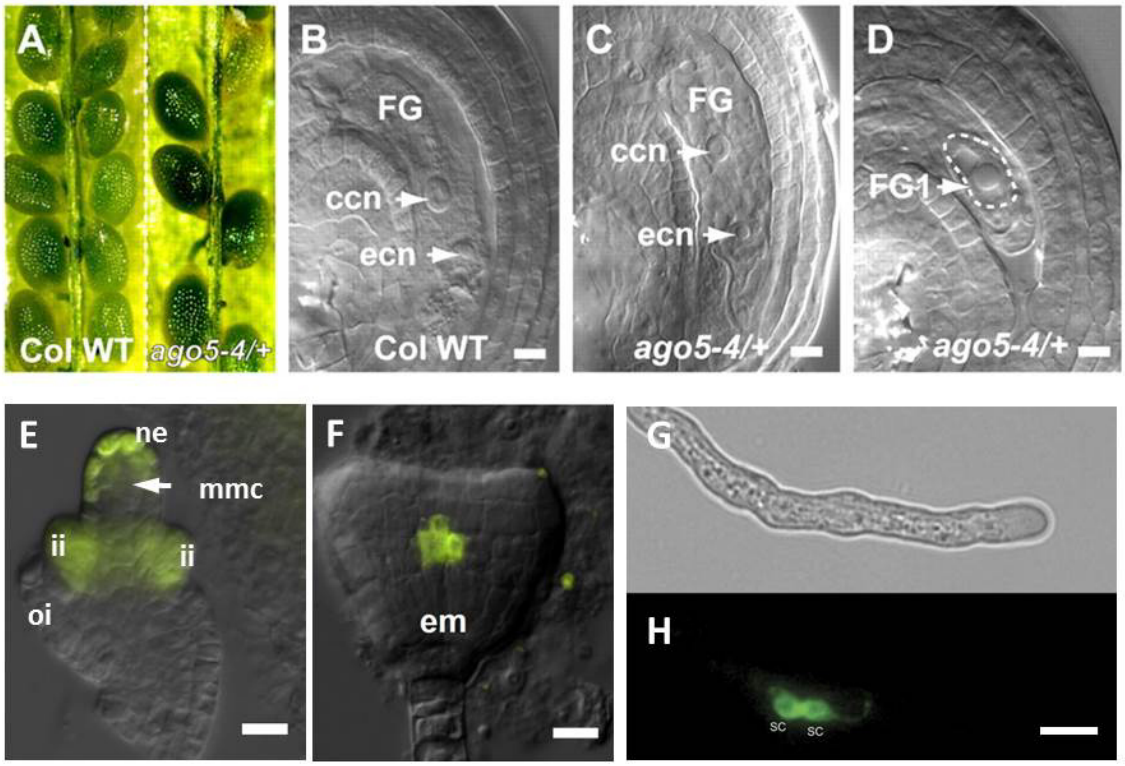


Figure 11: Role of AGO5 in ovule development and presence of AGO5 fusion proteins in developing ovules and in sperm cells.
(A) Comparison of seeds in mature WT (left) and *ago5-4*/+ (right) siliques. (**B**) WT anthesis ovule. (**C**) WT-like *ago5-4*/+ anthesis ovule. (**D**) Phenotypic *ago5-4*/+ anthesis ovule showing FG1 abortion (dashed line). (**E**) and **(F)** Expression of the YFPer reporter under control of the *AGO5* promoter. In immature ovules the *AGO5* promoter is active in the nucellar epidermis cells (ne) surrounding the functional megaspore (fm), and the developing inner integuments (ii). **(G)** and **(H)** Transgene expression of AGO5-eGFP in sperm cells (sc). GFP was C-terminally fused to the genomic *AGO5* sequence, under the control of the *AGO5* promoter. Images (A) to (F) taken from Tucker et al. 2012 and Borges et al. 2011 (G). Scale bars (A) to (F) are 10 µM. Scale bars (G) and (H) are 5 µM.

About 50 % of *ago5-4*/+ heterozygous plants show a developmental arrest at stage FG1 (Figure 11D), meaning that the functional megaspore mother cell does not progress to mitotic divisions. However, an YFP-AGO5 fusion protein expressed under control of the *AGO5* promoter is detected in the inner integuments and nucellus of the developing ovule but

neither in the functional megaspore nor in later stages of the developing female gametophyte (Tucker et al. 2012). In pollen, AGO5-GFP fusion protein is localized in the sperm cell cytoplasm (Figure 11G).

Arabidopsis ovules that carry a mutation in *AGO9* show a higher number of additional cells with germline identity (Olmedo-Monfil et al. 2010, Figure 12A). In these mutants, the expression level of otherwise silenced TEs is elevated, suggesting that AGO9 is involved in the silencing of TEs in the female germline.

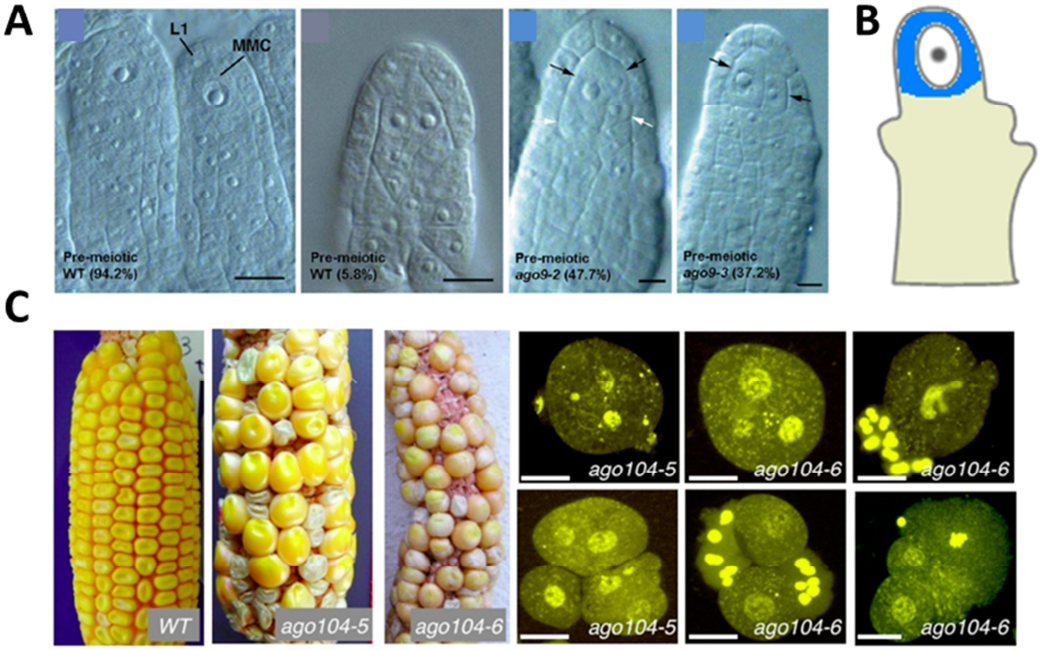


Figure 12: Function of AGO9 homologues in Arabidopsis and maize.
(A) In Arabidopsis *ago9* mutants, a high percentage of young ovules with additional megaspore mother cells (MMC) was observed. L1, L1 layer of the nucellus tissue. In wildtype plants, 94.2 % of the ovule had one MMC whereas 5.8% showed additional MMCs (left two pictures). In the mutants lines *ago9-2* and *ago9-3* 47.7% and 37.2% of ovules with additional MMCs were found, respectively (the two pictures on the right). Arrows indicate the position of additional MMCs. **(B)** In immunohistochemical staining experiments, AGO9 protein was detected in the nucellar tissue surrounding the MMC. Blue color indicates *AGO9* expression. **(C)** *Zea mays* loss of function mutant *ago104-5* and *ago 104-6* show kernel abortion (left images). In *ago104-5* and *ago 104-6* mutants defects in meiosis II, visualized using DAPI staining, lead to the production of abnormal tetrads, including triads and microspores with multiple nuclei (fluorescence images on the right). Images (A) and (B) (modifed) taken from Olmedo-Monfil et al. (2010). Images (C) and (D) taken from Singh et al. 2011. Scale bars (A) are 10 µM. Scale bars (B) are 50 µM.

Analysis of small RNAs bound to AGO9 showed a high number of sRNA derived from TEs and repetitive elements such as centromeres and telomeres (Duran-Figueroa and Vielle-Calzada, 2010). Recently, Rodríguez-Leal et al. (2015) linked *AGO9* activity to variations in cell differentiation in different Arabidopsis ecotypes. However, the maize

homologue of *AGO9*, which is called *AGO104*, seems to have an opposite function: egg cells of maize *ago104* mutant plants fail to undergo meiosis and generate diploid gametes (Singh et al. 2011), leading to kernel abortion. Those mutants showed an altered expression profile in ovules compared to the wild type, affecting genes as well as centromeric repeats and certain types of transposable elements.

3.3 Aims of this work

Small RNA pathways have been implicated to play an important role in plant reproduction. Argonaute proteins, which are the main effectors of such pathways, show a highly specific expression pattern, and transcripts of certain *AGOs* are highly abundant in the cells of the female gametophyte. In a previous work using isolated egg cells, central cells and synergid cells, microarray-based expression studies were performed (Soljic 2012).

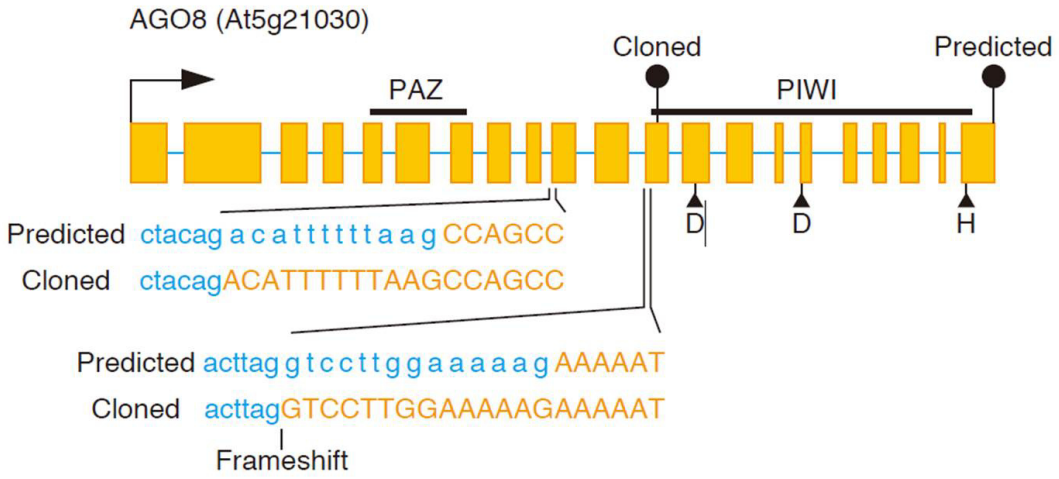


Figure 13: The *AGO8* locus and cloned *AGO8* transcript variant.
Cloning and sequencing of *AGO8* cDNA revealed the presence of alternative 3´ splice sites in exon 10 and exon 12. Orange boxes and letters represent exons. Blue lines and letters represent introns. The arrow indicates the translation start site. Black dots stand for a stop codon. Black bars: PAZ and PIWI domain (chapter 3.2.1). Triangles point at the position of amino acids needed for catalytic activity (DDH). Image taken from Takeda et al. 2008.

In these studies the expression of the putative *AGO8* pseudogene was found to be restricted to egg cells, while *AGO5* and *AGO9* are expressed in egg cells, central cells and sperm cells. In this work the expression of *AGO8* and *AGO9* genes and proteins should be investigated in more detail in transgenic plants, after generating suitable promoter-reporter constructs and constructs in which fluorescent AGO fusion proteins were expressed under control of their own promoters. Aspects such as expression pattern during female gametophyte development and the subcellular localization of AGO fusion proteins and native AGOs in the ovule and female gametophyte were one focus of this thesis. An approach of ectopically expressing *AGO8* and *AGO9* was chosen to address questions of *AGO8* and *AGO9*

function, since the generation of *ago8/*ago*9* double mutants is not possible due to their close physical proximity on chromosome 5 (Wartlsteiner 2010).

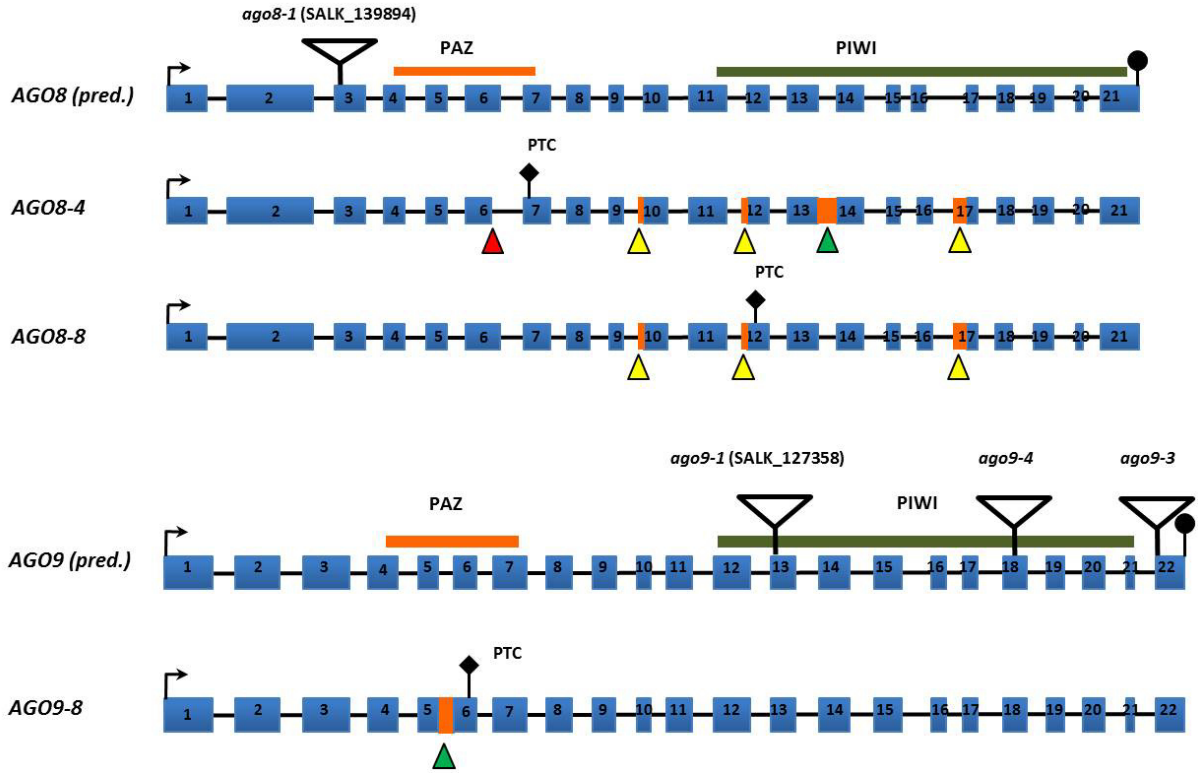


Figure 14: Alignment of *AGO8* and *AGO9* genomic DNA sequence with the cloned *AGO8* and *AGO9* transcript variants.
Genomic DNA and annotated cDNA sequence from *AGO8* and *AGO9* were retrieved from The Arabidopsis Information Resource (TAIR) and compared to previously discovered transcript variants. Blue bars represent exons as annotated by TAIR. Triangles indicate alternative splicing events. Red triangles stand for alternative 5'splice, yellow triangles represent alternative 3'splice sites and green triangles indicate intron retention. Orange bars indicate intronic sequences contained in the transcript. Image modified from Wartlsteiner 2010.

Moreover, a study from Takeda et al. (2008) revealed the presence of an aberrant transcript variant of *AGO8*. This variant differs from the predicted sequence of the *AGO8* coding sequence (CDS), showing alternative 3´ splice sites at the boundaries of exon 10 and 12 respectively. Those alterations in the coding sequence would lead to a frame shift and a premature termination signal (PTC) in exon 12 of the *AGO8* transcript (Figure 13).

Also in our group previous expression studies of *AGO8* by RT-PCR with cDNA, generated from mature pistils of emasculated Arabidopsis flowers, revealed amplicons that differ from the expected size (Wartlsteiner 2010). Amplification and sequencing of the full-length *AGO8* coding sequence exhibited the presence of two different aberrant spliced cDNAs (Figure 14). The two cloned cDNAs (*AGO8-4*, *AGO8-8*) both show alternative splicing events, including alternative donor and acceptor sites and a retained intron 13 (Wartlsteiner

2010). Additionally, a previously undescribed *AGO9* transcript variant was isolated from mature pistils of emasculated flowers (Figure 14). Sequencing of cloned transcripts revealed the presence of *AGO9* intron 6, which was retained (Wartlsteiner 2010). This retained intron leads to a frame shift, with the consequence of a premature termination codon (PTC) at base pair position 912 of the coding sequence.

To address the frequency of appearance of theses splicing variants during female gametophyte development, to screen for other/more *AGO8* and *AGO9* splicing variants, and to investigate the putative role of aberrantly spliced *AGO8* and *AGO9* transcript variants two approaches were used; an RT-PCR based approach was used to detect *AGO8* and *AGO9* transcript variants in different vegetative and reproductive tissues (Chapter 5.2.1).

In addition to this a so-called minigene approach was used. A minigene is a gene fragment from a gene of interest which contains an alternatively spliced region and its flanking genomic sequence (Cooper 2005). *AGO9* minigenes were cloned between green and red fluorescent proteins and the constructs were under the control of the endogenous AGO9 promoter. Alternative splicing of the minigenes would result in different fluorescent signals, which allows for the detection of different AGO9 transcript variants in the developing female gametophyte. This approach is explained in more detail in Chapter 4.4.10 (cloning) and 5.7.5 (principle).

Another goal of this work was to obtain more global information on transcripts and small RNAs that are active during gametogenesis in *Arabidopsis thaliana.* Since the generation of small RNA libraries for RNA sequencing (RNA-Seq) and any protein work with very limited starting material, such as isolated egg cells, is not possible, a specialized cell line with an egg cell-like transcriptome should be established. Köszegi et al. (2011) reported that plants overexpressing the RWP-RK family transcription factors RKD1 and RKD2 form an undifferentiated callus-like tissue running an egg cell-like transcription program. Therefore, an RKD2-induced cell line should be established during this thesis to be used for in depth analyzes of the transcriptome and the small RNA profile in comparison to standard callus tissue. The focus of these analyzes should be on the identification of small RNA pathway components upregulated in the female germline and the investigation of genes known to be involved in auxin and cytokinin signaling pathways. The generated small RNASeq data would be used to perform differential expression analysis of the RKD2-induced egg cell-like callus and the non-reproductive callus tissue, so that specific small RNAs playing a putative role in gamete formation could be identified

4 MATERIAL & METHODS

Chemicals and reagents used in this work had molecular grade or p.a. (pro analysis) grade purity. Molecular biology work was based on protocols from Sambrock et al. (1989).

4.1 Plant materials and growth conditions

Arabidopsis thaliana ecotype Columbia (Col-0) was grown on soil under a short photoperiod (9 h of light, 8,500 lux, 22°C, 60 to 70% humidity) for 3 to 4 weeks, followed by a long photoperiod (16 h of light, 8,500 lux, 22°C, 60 to 70% humidity). Seeds were stratified at 4°C for 48 hours before growth. The *ago9-1* T-DNA (SALK_127358) insertion line was ordered from the Nottingham Arabidopsis Stock Centre (NASC). For sterile growth and selection of transgenic seedlings on plates the seeds of transformed plants were surface-sterilized with 1% sodium hypochlorite (NaOCl) and 70% ethanol for 5 min respectively, washed 5 times with sterile water and then plated on half strength Murashige and Skoog (MS) medium (Duchefa) supplemented with 2% (w/v) sucrose and solidified with 1% phytoagar (Duchefa). Appropriate antibiotics were added when selection of transgenes was necessary.

4.2 Callus generation

In order to generate the RKD2-induced egg cell-like callus, the coding sequence (CDS) of RKD2 was amplified without a stop codon from pistil cDNA using the primer pair RKD2for/RKD2rev (chapter 9.2) and cloned into pENTR/D-TOPO® (Life Technologies). Subsequently, the CDS of RKD2 was fused to the N-terminus of GFP by cloning it into the GATEWAY® compatible destination vector pH7FWG2.0 (Karimi et al. 2002), generating the expression vector pURM18 (*35Sp:RKD2-GFP*). *Agrobacterium tumefaciens* strain C58C1 was transformed with pURM18 and used to transform Arabidopsis by the floral dip method according to Clough and Bent (1998). Seeds of transformed plants were surface-sterilized (Chapter 4.1) and grown on solid half-strength Murashige and Skoog (MS) medium (Duchefa) supplemented with 2% (w/v) sucrose and 30 µg/mL Hygromycin, which was prepared with 1% Phytoagar (Duchefa). RKD2-induced calli appeared at the hypocotyl of seedlings 20 to 30 days after sowing the seeds on selection medium. Calli were subcultured on new MS plates supplemented with 2% (w/v) sucrose and 30 µg/mL Hygromycin every 14

days. To induce calli from root segments (CIM callus), Arabidopsis Col-0 seeds were surface sterilized (chapter 4.1), plated on solid half strength MS-medium supplemented with 2% (w/v) sucrose and 1% Phytoagar (Duchefa) and grown for 10 to 14 day under sterile conditions. From those plantlets, 5-10 mm long explants from roots were cut out and placed on a callus induction medium (CIM) composed of full-strength Gamborg Medium B5 with vitamins and MES (Duchefa), 2% glucose, 0.5 mg/l 2,4-dichlorophenoxyacetic acid (2,4-D), and 0.05 mg/L kinetin with 1% phytoagar (Duchefa). Calli appeared after 7 to 10 days and were then subcultured every 14 days. For RNA and protein work, callus material was collected with a sterile scalpel blade and immediately frozen in liquid nitrogen.

4.3 Bacterial work

Escherichia coli was cultivated in liquid LB medium or on LB solidified with 1.5% Bacto Agar (w/v) and supplemented with the appropriate antibiotics at 37°C.

LB medium

1.0 %	Bacto Tryptone (w/v)
0.5 %	Yeast Extract (w/v)
1.0 %	NaCl (w/v)

Optional:

1.5 %	Bacto Agar (w/v)

YEP medium

1.0 %	Bacto Peptone (w/v)
1.0 %	Yeast Extract (w/v)
0.5 %	NaCl (w/v)

Optional:

1.5 %	Bacto Agar (w/v)

Strains of *Agrobacterium tumefaciens* were grown at 28°C in YEP medium or on plates with solidified YEP medium, supplemented with the appropriate antibiotics.

4.3.1 Generation and transformation of chemically competent *Escherichia coli* cells

In order to generate chemically competent *E. coli* cells, a protocol adapted from Nishimura et al. (1991) was used. 5 ml of LB medium (chapter 4.3) with the appropriate antibiotic was inoculated with a single bacterial colony and incubated overnight at 37°C with shaking at 180 rounds per minute (rpm). This overnight culture was diluted into 50 mL of fresh medium at a 1 to 100 ratio and grown at 37°C until an OD_{600} of 0.4 to 0.6 was reached. Cells were incubated on ice for ten minutes and afterwards harvested by centrifuging at 1,500 g and 4°C. The pellet was carefully resuspended in 0.5 mL LB, supplemented with 10 mM $MgS0_4$ and 0.2% glucose. The suspension was mixed with 2.5 mL of Nishmura medium (36% Glycerin,12%PEG, 12 mM $MgSO_4$ in LB) and divided into 100 or 200 µL aliquots, which were shock-frozen in liquid nitrogen.

In order to transform chemically competent *E. coli* cells, 5 µL of a conventional ligation reaction, 3 µL of a TOPO or LR clonase reaction, or 1 to 10 ng of plasmid DNA was added to thawed bacteria and kept on ice for half an hour. A heat shock of 30 sec at 42°C was applied and the cells were put on ice for one minute. Afterwards, 1 mL of LB was added and the cells were cultivated for one hour, while shaking at 37°C. The bacteria were harvested by centrifugation at 600 rpm, the pellet was resuspended in a small volume of medium and plated out on solid LB medium with the appropriate antibiotics.

4.3.2 Generation and transformation of chemically competent *Agrobacterium tumefaciens* cells

Single colonies of *A. tumefaciens* strains GV3101 (pMP90RK), GV3101 (pMP90, pSOUP) and C58C1 were grown overnight in 2 mL of YEP medium (chapter 4.3) at 28°C, with shaking at 200 rpm. The overnight cultures were used to inoculate 50 mL of YEP medium and the culture was incubated at 28°C until an OD600 of about 0.5 was reached. The cells were then collected by centrifugation for five minutes at 4,000 g and 4°C. The cell pellet was resuspended in 10 mL ice-cold 150 mM NaCl and centrifuged again, like described above. Cells were then resuspended on ice in one mL 20 mM $CaCl_2$ and 87% glycerol was added to a final concentration of 15%. Aliquots of 200 µL were frozen in liquid nitrogen and stored for further use at -80°C.

For transformation, one aliquot of cells was thawed on ice and 1 µg of plasmid DNA was added. After 30 min incubation on ice, cells were frozen in liquid nitrogen and thawed at 37°C. One mL of YEP or LB-Medium (Chapter 4.3) was added and cells were shaken at 28°C for two hours. The transformation was plated on solidified YEP medium (Chapter 4.3) with the appropriate antibiotics and incubated at 28°C for 2-3 days until colonies appeared.

4.4 DNA molecular biology work

4.4.1 Isolation of total DNA from leaves

Genomic DNA from leaves was extracted as previously described (Edwards et al. 1991). A small piece of leaf was frozen in liquid nitrogen and ground to a fine powder. 400 µL of buffer (200 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.5% SDS (w/v) and 25 mM EDTA) was added and the samples were centrifuged at full speed for five minutes. 350 µL of the supernatant was then mixed with an equal volume of isopropanol and after ten minutes incubation at room temperature, centrifuged again for ten minutes at full speed. The pellet was washed with 500 µL 70% ethanol and after drying resolved in 100 µL TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). In case of insoluble material, samples were again centrifuged for ten minutes at room temperatures and the clear supernatant was transferred to fresh tubes.

4.4.2 Isolation of high quality total DNA from plant material

For the amplification of promoter regions high quality genomic DNA was used. Genomic DNA was isolated as described in Dellaporta et al. (1983) with the following modifications: An extraction buffer consisting of 100 mM Tris-HCl, pH 7.5; 50 mM EDTA, pH 8.0; 1% sodium dodecyl sulfate (SDS) and 500 mM NaCl was used for the procedure. β-mercaptoethanol was added to a final concentration of 10 mM before use. 200 mg of plant material was frozen in liquid nitrogen and ground to a fine powder. 5 mL of prewarmed (65°C) extraction buffer was added and the samples were incubated for 10 min at 65°C in a water bath. After the addition of 1.67 mL of 5 M potassium acetate, samples were mixed and incubated for twenty minutes on ice, followed by centrifugation for 10 minutes at 25,000 x g at 4°C. The supernatant was mixed with 3.33 mL of Isopropanol and centrifuged again for 30 min at 20,000 g and 4°C. The pellet was resuspended in 1 mL of ice-cold TE buffer (Chapter

4.4.1) and 10 µL of RNAse A (10 mg/mL in TE buffer) were added, followed by incubation for 1 hour at 37°C. A phenol/chloroform extraction, followed by isopropanol precipitation was performed afterwards, following the protocol of Sambrock et al. (1989). The DNA pellet was dissolved in 100 µL of TE buffer.

4.4.3 Polymerase chain reaction (PCR)

PCR-based genotyping, RT-PCRs, and screening for positive transformants via colony-PCR were performed with Taq polymerase from MBI Fermentas (1 U/µL).

PCR with Taq DNA-Polymerase

2 µL	10 x Taq-Buffer $(NH4)_2SO4$
2 µL	25 mM $MgCl_2$
2 µL	dNTPs (2 mM)
1 µL	Primer Forward (10 µM)
1 µL	Primer Reverse (10 µM)
X µL	DNA template
0.5 µL	Taq DNA-Polymerase (1 U/µL)
Ad.20 µL	ddH_2O

Program: Taq™ DNA-Polymerase

95°C	3 min	
95°C	30 sec	30-40x
Tm +/- 10°C	30 sec	
72°C	60 sec/ kb	
72°C	10 min	
4°C	hold	

The extension time was calculated upon the lengths of the amplicon, given an amplification rate of the Taq-Polymerase of 1 kb/min. Amplication of cDNA and genomic

DNA-fragments for TOPO or conventional cloning was performed with Phusion™ DNA-Polymerase (Thermo Scientific), which is a DNA-polymerase with a proofreading activity and has a high amplification rate. The amount of DNA used as a template was dependent on the nature of the DNA. Of plasmid DNA 20 ng was used, while 100 to 200 ng were used from a preparation of genomic DNA.

PCR with Phusion™ DNA-Polymerase

10 µL	5 x HF-Buffer with 7.5 mM $MgCl_2$
5 µL	dNTPs
2.5 µL	Primer Forward (10 µM)
2.5 µL	Primer Reverse (10 µM)
X µL	DNA template
0.5 µL	Phusion™ DNA-Polymerase (1 U/µL)
Ad.50 µL	ddH_2O

Program: Phusion™ DNA-Polymerase

98°C	3 min	
98°C	15 sec	30-40x
Tm +/- 5°C	20 sec	
72°C	30 sec/ kb	
72°C	5 min	
4°C	hold	

4.4.4 Gel extraction of DNA fragments

Isolation of DNA fragments from Agarose gels was performed using the „Qiaquick Gel Extraction Kit“ from Qiagen® following the manufacturer’s instructions.

4.4.5 Isolation of plasmid DNA from *E.coli*

4 mL LB medium was inoculated with a single colony of *E. coli* and grown as described above (Chapter 4.3). Cells were collected by centrifugation for 1 min at maximum speed in an Eppendorf tabletop centrifuge. Plasmids were isolated using the High-Speed Plasmid Mini Kit from Avegene. The DNA was eluted in ddH_20, which was prewarmed to 50°C and concentration was measured with a NanoDrop spectrophotometer (Thermo Scientific). Midipreparation of plasmid DNA was carried out with a NucleoBond 100 Midiprep kit from Macherey & Nagel according to the provided protocol.

4.4.6 Isolation of plasmid DNA from *A.tumefaciens*

In order to isolate plasmid from *A.tumefaciens* an alkali lysis method as described in Sambrock et al. (1989) was used, omitting the step of Phenol/chloroform extraction.

4.4.7 Analytical and preparative restriction digest

Analytical restriction digests were performed in 20 µL reaction volumes using enzymes from MBI/Fermentas according to the manufacturer's instructions with 5 U of enzyme. 5 µL of isolated plasmid DNA (Chapter 4.4.5) was digested for one hour at 37°C and then separated on agarose gels.

Preparative reactions for the preparation of cloning vectors were performed with 2 µg of plasmid DNA in a 50 µL reaction and 20 units of each enzyme were used. Digestion was carried over night at 37°C and DNA fragments were isolated from a preparative agarose gel as described in Chapter 4.4.4.

4.4.8 DNA Ligation

Prior to ligation, vector DNA was dephosphorylated with FastAP Thermosensitive Alkaline Phosphatase (Life Technologies). A T4-DNA Ligase from MBI/Fermentas was used and the molecular vector/insert ratio was varied from 1:1 to 1:3 with 100 ng of vector. The reaction took place at 22°C for 2 hours and was continued at 4°C overnight.

The DNA ligase was inactivated for 5 minutes at 65°C and 5 µL of the ligation reaction was transformed into chemically competent *E.coli* cells as described (chapter 4.3.1).

4.4.9 Gateway®-based cloning

In order to subclone amplified PCR fragments into Gateway®-based vectors, the pENTR/D-TOPO cloning kit (Life Technologies) was used according to the manufacturers' manual to generate Entry vectors. The directional cloning of a PCR fragment into a Destination vector was achieved by adding a 5´ CACC overhang to each forward primer, which is complementary to a sequence in the pENTR/D-TOPO vectors cloning site. PCR-fragments with CACC overhang were generated using Phusion™ DNA-Polymerase (Thermo Scientific) as described in chapter 4.4.3.

The PCR fragment in the resulting Entry clone was sequenced. Subsequently, the Entry clone was used for LR reaction using the Gateway® LR Clonase® II enzyme mix from Life Technologies to recombine the PCR fragment into the Destination vector. LR reactions were performed according to the manufacturers' protocol, setting up half reactions with 50 ng Entry vector and 75 ng Destination vector. Entry- and Destination vectors used for Gateway® -based cloning and oligonucleotides used for cloning can be found in the appendix.

4.4.10 Cloning of PCR-fragments for the analysis of *AGO9* transcript variants

Overlapping DNA fragments derived from the *AGO9* coding sequence (Chapter 5.7.4) were generated by PCR using Taq-Polymerase (Chapter 4.4.3). The TOPO® TA Cloning®-Kit (Invitrogen) was used to clone generated fragments into the pCR®2.1 vector (Invitrogen), according to the manufacturer's manual. TOPO® reaction was transformed into One Shot® TOP10 chemically competent *E. coli* cells, supplied with the kit. Generated clones were analyzed by sequencing.

4.4.11 Cloning of *AGO9* minigene constructs

For the generation of minigene constructs genomic fragments of *AGO9* were amplified from genomic DNA using primers with added *NotI* restriction sites. The amplified fragments

were subcloned into the pCR®-Blunt II-TOPO® vector using Zero Blunt® TOPO® PCR Cloning Kit (Life Technologies) following the manufacturer's instructions.
Fragments were cloned into the pABB22 vector (kindly provided by Birgit Absmanner) which consists of the coding sequence of the mCherry fluorophore with pENTR/D-TOPO as backbone. Vector and insert were digested with *NotI* restriction enzyme (*NotI*-HF from New England Biolabs) and ligation was performed as described in chapter 4.4.8.

The destination vector for minigenes was pURM44 (Chapter 9.2) which is the modified binary expression vector pB7WGF2.0 (Karimi et al. 2002) in which the *35S* Promoter has been exchanged against the endogenous *AGO9* promoter. The *AGO9* promoter was amplified from genomic DNA and cloned into pB7WGF2.0 via *SacI* and *SpeI* restrictions sites. For the generation of the final minigene constructs a LR reaction was performed (chapter 4.5.9). Four minigene construct have been generated. Oligonucleotides used for the amplification of minigenes are listed in Chapter 9.2.

Summary of minigene constructs:

Entry vector	Exon/Intron-Structure	Lenght (bp)	Destination Vector	Expression clone
pURM48	E5-I6-E6	262	pURM44	pURM50
pURM49	E4-I5-E5-I6-E6-I7-E7	721	pURM44	pURM51
pURM54	E10-I11-E11	252	pURM44	pURM56
pURM55	E9-I10-E10-I11-E11-I12-E12	717	pURM44	pURM57

4.4.12 Sequencing

DNA sequencing was performed by 4base lab (Reutlingen), GATC Biotech (Konstanz), or LGC genomics (Berlin). Cloned plasmids were purified as described above and sent with sequencing primers in the recommended concentrations.

4.5 RNA work

4.5.1 Isolation of total RNA

Total RNAs including Low Molecular Weight (LMW) RNA from 200 to 300 mg of frozen RKD2-induced and CIM callus material was extracted using the RNeasy Plant Mini Kit (Qiagen), according to the manufacturer's instructions. Purity and integrity of the total RNA was assessed on the Agilent 2100 Bioanalyzer with the RNA 6000 Nano LabChip reagent set (Agilent).

4.5.2 mRNA isolation and cDNA synthesis

Messenger RNA was either isolated from 5-10 mg of plant tissue (leaf, shoot, root, RKD2-induced callus, CIM callus), or isolated from 5µg total RNA using Dynabeads® mRNA DIRECT™ Purification Kit (Life Technologies) with Dynabeads Oligo $(dT)_{25}$. First-strand cDNA synthesis was carried out using Oligo$(dT)_{18}$ primers (MBI Fermentas) and SuperScriptIII® Reverse Transcriptase (Life Technologies) following the manufacturers' protocols with the addition of 1 µl RiboLockTM Ribonuclease Inhibitor (MBI Fermentas).

4.5.3 Northern Blot hybridization

Small RNAs were separated on 12% Polyacrylamidgels containing 7.5M Urea. For preparation of the gels the SequaGel UreaGel system (National Diagnostics) was used according to the manufacturer's instructions. The gels were prerun at 400 V for 30 min until the gel was warm. RNA samples were mixed with equal volumes of 2x sample buffer (0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF in 95% deionized formamide) and incubated at 95°C for 5 minutes. Gels were run at 350 V until the bromophenol band reached 7 cm. After PAGE, the gels were soaked for 10 minutes in 200 mL TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) containing 0.5 µg/mL ethidium bromide).

RNA was transferred to neutral nylon membrane (Hybond-N, Amersham Biosciences) for 30 minutes at 20 V using a BioRad Trans-Blot® SD Semi-Dry transfer cell. Sterile bidistilled water was used as transfer buffer. The RNA was crosslinked to the membrane

using the EDC (1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide) method according to Pall and Hamilton (2008). A piece of Whatman 3MM paper (Fischer Scientific) was saturated with freshly prepared EDC crosslinking solution and the membrane was placed there with the RNA side face up. Incubation of the membranes took place in a hybridization oven at 50°C for one hour. After the crosslinking reaction the membranes were briefly rinsed with sterile bidestilled water.

EDC crosslinking reagent

184 mg	EDC
61.25 µL	12.5 M 1-methylimidazol
75 µL	1 M HCl
Ad. 6 mL	ddH_2O

Antisense oligonucleotides of the miRNAs of interest were 5`-end labeled using a T4 polynucleotidkinase (PNK) from MBI Fermentas. The reaction was set up as follows:

5′-End Labeling of RNA oligos with T4 PNK

1 µL	Oligonucleotide (20 µM)
2 µL	10xPNK reaction buffer A
14 µL	H_20
1 µL	PNK (10 u/µL)
2 µL	γ-^{32}P labelled ATP (10 µCi/µL)

The reaction was incubated for 30 to 45 min at 37°C. The labeling reaction was purified with Sephadex G25 Columns (GE Healthcare). The columns were vortexed, the bottom cap was removed and the columns were centrifuged at 700 g for one minute. The flow-through was discarded and the columns were placed into 1.5 mL microcentrifuge tubes containing 30 µL of 30 mM Ethylenediaminetetraacetic acid (EDTA). The labeling reaction was pipetted into the middle of the Sephadex resin and the columns were again centrifuged for 2 min at 700 g. The labelling efficiency was roughly estimated with a Geiger counter measuring the radioactivity of the column and the purified probe. Blots were prehybridized

for 30 min in hybridization solution (QuikHyb, Agilent Technologies) at 50 °C. For hybridization, the radiolabeled probe was added to the hybridization solution and incubation was continued overnight at 50°C. Membranes were washed twice with 5x sodium saline citrate (SSC), 1% sodium dodecyl sulfate (SDS) and once with 1xSSC, 1% SDS for 10 min at 50°C. Washing solutions were prepared from 20xSSC (3M sodium chloride, 0.3M sodium citrate, pH 7.0). After washing, membranes were exposed to a phosphor imager screen. Signals were detected with a cyclone plus phosphor imager (Perkin-Elmer). In order to remove hybridized probes and be able to reuse blot with different probes, blot were soaked in boiling 0.1% SDS and incubated for further 10 min in a 70°C shaking water bath. To control the removal of the probe membranes were put again on a phosphor imager screen and incubated over night. If all signals were removed from the blot, membranes were stored at -20°C between two sheets of Whatman 3MM paper (Fisher Scientific) and tin foil until the next use.

4.5.4 Quantitative real-time PCR Assay

Quantitative real-time PCR (qRT-PCR) was used to assess the transcription levels of different genes in RKD2-induced and CIM callus. Total RNA was extracted as described above (Chapter 4.5.1) and digested with RNase-free DNase I (Invitrogen). For 1 µg of total RNA 1 U of DNAse was used. The DNAse was inactivated by adding 1/10 volume of 25mM Ethylenediaminetetraacetic acid (EDTA) and heating the samples for 10 min at 65°C. Messenger RNA was isolated from 5µg total RNA using Dynabeads® mRNA DIRECT™ Purification Kit (Life Technologies) with Dynabeads Oligo $(dT)_{25}$ as described (chapter 4.6.2).

First-strand cDNA synthesis was carried out using Oligo$(dT)_{18}$ primers (MBI Fermentas) and SuperScriptIII® Reverse Transcriptase (Life Technologies) following the manufacturers protocols with the addition of 40 U RiboLock™ Ribonuclease Inhibitor (MBI Fermentas). One µL of cDNA template was used per PCR Reaction.

The PCR reaction was performed with 2x SYBR-Green Master Mix (Peqlab) on a Mastercycler ep realplex S (Eppendorf) with the following amplification parameters: 95 °C for 1 min, followed by 40 cycles of 95°C for 1 s, 63°C for 15 s and 72°C for 8 sec. *UBC* (*AT5G25760*) and *eIF4G* (*AT3G60240*) were used as reference genes (for primer sequences see chapter 9.2). A set of seventeen genes was used for validation of RNA-seq data.

Triplicates of each sample were analyzed. The change of expression between two samples was determined by the $2^{-\Delta\Delta CT}$ Method according to Livak and Schmittgen (2002). A Summary of genes used for qRT-PCR can be found in the appendix.

4.5.5 Library Construction and Illumina Next-Generation-Sequencing

The same batch of total RNA was used for mRNA-Seq and small RNA-Seq library preparation, respectively. Three biological replicates of each callus type (CIM and RKD2-induced callus material) were used for library preparation. Library preparation and deep sequencing was carried out at the Kompetenzzentrum für Fluoreszente Bioanalytik (KFB) at the University of Regensburg.

4.5.6 mRNA library preparation

Messenger RNA-Seq libraries were prepared from CIM and RKD2-induced callus material (three biological replicates each). Poly(A)+ mRNA was selectively enriched from 500 ng total RNA using Oligo$(dT)_{25}$ beads (NEBNext Poly(A) mRNA Magnetic Isolation (New England Biolabs). Messenger RNA libraries were prepared with the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs).

4.5.7 Small RNA library preparation

Small RNA library preparation was carried out according to the Illumina® TruSeq® Small RNA sample preparation guide with the following modifications: For each of the six samples (three biological replicates), 1 µg RNA was used for ligation of the RNA 3' and the RNA 5' adapters, followed by reverse transcription to create single stranded cDNA. The cDNA was then PCR amplified for 12 cycles using a universal and an index primer. Following magnetic bead clean up (Agencourt AMPure XP, Beckman Coulter), small RNA species were gel-purified using an automated nucleic acid fractionation/extraction system with a collection width of 30 bp (LabChip XT DNA 300 Kit, Caliper Life Sciences).

4.6 Bioinformatic analysis of RNA sequencing data

Analysis of RNA seq data was carried out by Julia Engelmann, Maurits Evers and Nicolas Strieder at the institute for functional genomics, department for statistical bioinformatics. Count tables required for differential expression analysis were produced using R/Bioconductor [GenomicAlignments] (Anders et al. 2015) for small RNAs and featureCounts (Liao et al. 2014) form RNAs. The DESeq2 software package (Love et al. 2014) was used for library size normalization and differential expression analysis.

4.6.1 Small RNA library analysis

The quality of the different RNA-Seq libraries was assessed by using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Afterwards, reads were size- and quality trimmed (smallRNA:TRAILING:20 LEADING:20 MINLEN:18, mRNA:TRAILING:3 LEADING:3 MINLEN:25) with Trimmomatic (Bolger et al. 2014). From the small RNA and mRNA libraries about 18%, respectively 3.6 %, of reads were dropped failing length or quality criteria. The resulting reads were aligned to the *Arabidopsis thaliana* reference genome TAIR10 (http://www.arabidopsis.org/). Small RNA reads were mapped with butter (Axtell 2014) distributing reads to repetitive elements by an iterative process relying on the distribution densities of confidentially aligned reads. Alignment of mRNA reads was performed with Tophat2.0 (Kim et al. 2013). The annotation was retrieved from EnsemblPlants (ftp.ensemblgenomes.org, Arabidopsis_thaliana. TAIR10.22.gtf, Kersey et al. 2013).

4.6.2 Prediction of small RNA targets

Potential target transcripts for miRNAs identified in our small RNA libraries were retrieved from a miRNA-target relation list from Dr. Jim Carrington (ftp://ftp.arabidopsis.org/home/tair/Genes/SmallRNAsCarrington/). Targets in this list were either predicted by 5'-RACE or a prediction algorithm proposed by Allen et al. (2005). In addition, psRNAtarget (Dai and Zhao 2011) was used to further potential targets.

4.7 Protein Methods

4.7.1 Total protein extracts from *Arabidopsis thaliana*

Fresh plant material was collected, frozen in liquid nitrogen and ground to a fine power in a precooled mortar. The material was transferred into ice-cold extraction buffer (20 mM Tris pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$, 0.1% Triton X-100, 1x complete Proteinase-Inhibitor Cocktail from Roche, 50 µM MG-132) and vortexed immediately. 200 µL of extraction buffer was used for 100 mg of plant material. After ten minutes of incubation on ice, samples were centrifuged for 10 minutes at 16.000 x g. The supernatant was transferred into a fresh 1.5 mL reaction tube. In case of the presence of cellular debris this step was repeated until the supernatant was clear.

4.7.2 Bradford protein assay

The protein concentration of extracts was determined by the Method of Bradford (1976), using the Bio-Rad protein assay kit according to the manufacturer's instructions. Extinction of standards and samples were measured with an Eppendorf BioPhotometer at 595 nm.

4.7.3 SDS-PAGE

Prior to SDS-Polyacrylamide gel electrophoresis (SDS-PAGE), extracts were mixed with one volume of 2x sample buffer (125 mM Tris, 4% SDS, 20% Glycerol, 10% ß-Mercaptoethanol) or the respective volume of 6x sample buffer and incubated for 10 min at 65°C. When cell pellets from *E. coli* were used, one mL of bacterial culture was pelleted, resuspended in ice cold PBS consisting of 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.2 (50 µL per OD 0.6) then mixed with sample buffer and treated as described above. Proteins were separated by SDS polyacrylamid gel electrophoresis (SDS-PAGE) according to Laemmli (1970). The Bio-Rad MiniProtean II system was used to cast and run gels. For the analysis and detection of Argonaute proteins in complex protein mixtures, 7.5 % resolving gels with a 4.5% stacking gel were used. Gels were run at 20mA per gel until the bromophenol blue front was running out from between the glass plates.

Gels were stained with Coomassie staining solution (0.25% Coomassie Brilliant Blue R-250 from Sigma-Aldrich, 10% acetic acid, 25% isopropanol) by incubating the gels overnight. The gels were destained with 10% acetic acid.

4.7.4 Western Blot

Proteins that were separated by SDS-PAGE were immobilized by transferring on a nitrocellulose membrane (Amersham Protran 0.45 NC). Transfer took place in a Mini-PROTEAN II cell from Bio-Rad at 400 mA for 60 minutes. After transfer the membrane was washed shortly in water and the transfer of the proteins was controlled by staining the membrane with Ponceau S solution (2% Ponceau S in 3% trichloroacetic acid). The protein bands of the transferred molecular weight standard were marked with a ball pen. The membrane was blocked with 5% milk powder in TBS-T (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.2% Tween-20) at room temperature for one hour. After washing of the membrane with TBS-T (at least 3 washes for 10 minutes), the primary antibody was added to the membrane and incubation took place at four degrees overnight. Primary antibodies against AGO proteins were purchased from Agrisera (Vännas, Sweden). All primary antibodies were polyclonal and specific peptides were used for immunization. Dilutions of the antibodies are summarized in the appendix. After washing the blot like above, the secondary antibody was added at a dilution of 1:5,000 and incubated with the blot for at least one hour. As secondary antibody a horseradish peroxidase (HRP)-coupled anti-rabbit IgG was used. After washing the blot with TTBS three times for 10 min, the detection reaction took place. A luminol-based HRP-substrate (HRP juice) from PJK was used for this reaction. Signals were visualized either by exposing the wrapped membrane on X-ray film or with an Multi image II imaging system (Alpha Innotec).

4.7.5 Peptide competition assay

Specificity of the anti-AGO9 antibody (Agrisera) was tested in a peptide competition assay. The AGO9 peptide (DSDEPNGSGLPPPC-CONH2) was ordered from Centic Biotec. The peptide was dissolved in 10 mM Tris/Cl buffer at pH 7.0, aliquoted and kept at -80°C. For peptide competition, 1 µg of anti-AGO9 antibody was combined with 5 µg blocking peptide in a volume of 1 mL TBS-T without milk powder. Samples were incubated on an

overhead shaker for two hours at room temperature. After this step, samples were filled up to 10 mL with blocking buffer, resulting in a 1:10,000 dilution of the antibody. Membranes were incubated with this mixture at 4°C overnight and the Western Blot procedure was continued as described above. As a control, a 1:10.000 dilution of the anti-AGO9 antibody in blocking buffer without peptide was used.

4.8 Cell biological work

4.8.1 Transient transformation of Tobacco BY2 cells by particle bombardment

Nicotiania tabacum (variant "Bright Yellow 2") cells were grown in single strength MS medium (Duchefa) for 3 to 4 days. The packed cell volume (PCV) of the cells was then determined. One PCV of transformation medium (MS medium containing 250 mM sorbitol and 250 mM mannitol) was added to the sedimented cells. The cell suspension was incubated for one hour at room temperature and then transferred onto MS plates containing 0.8% pyhtoagar.

Gold particles of 0.6 µM diameter were coated with plasmid DNA. To prepare the gold suspension for particle bombardment, 60 mg of gold was resuspended in 1 mL EtOH (abs) by vortexing and sonification. After at least two washing steps in EtOH (abs) the concentration of gold particles was adjusted with EtOH (abs) to 30 mg/mL and aliquoted to 50 µL in 1.5 mL reaction tubes. After adding 10 µL of plasmid DNA to 50 µL of gold suspension, the samples were mixed with 50 µL $CaCl_2$ and 20 µL 0.1 M spermidin, followed by intense vortexing. This mixture was centrifuged for ten seconds and the pellet was washed with 250 µL EtOH (abs). After another short centrifugation (10´´) the pellet was resuspended in 150 µL EtOH (abs). 7.5 µL of the gold-plasmid suspension was used for biolistic transformation.

Particle bombardment was performed using the PDS-1000/He Particle Delivery System from BioRad according to the manufacturer's instructions. Cell suspensions were bombarded by three shots at a pressure of 1.35 bar in a vacuum of -1.3 bar. After 3 days of incubation in the dark, cells were resuspended in 2 mL of MS medium and allowed to grow two more days in the dark. Finally, cells were analyzed by confocal laser scanning microscopy (CLSM).

4.8.2 Whole mount-immunolocalization

Immunolocalization of AGO9 protein was carried out on dissected ovules and embryonic stages as described by Sauer and Friml (2010) with slight modifications. Samples were fixed in 4% paraformaldehyd in 1x Microtubule Stabilization Buffer (MTSB) (50 mM PIPES, 5 mM MgSO4 x $7H_2O$, 5 mM EGTA, pH 6.9) under vacuum for one hour. If embryos were used, 0.1 % Triton-X100 (final concentration) was added to prevent floating. The samples were then washed three times with MTSB and twice with water and then mounted in a drop of water on adhesive microscope slides (Polysine Adhesion Microscope Slides, Thermo Scientific) were covered with a cover slip, immersed for 15 seconds in liquid nitrogen and then thawed. This was repeated four more times. The mounted plant material was dried overnight at room temperature. After drying, the area surrounding the samples was marked with a hydrophobic barrier pen (PAP pen, Sigma-Aldrich) and then rehydrated with MTSB for ten minutes. Cell walls were treated with 2% driselase in MTSB (Sigma-Aldrich) for 30 minutes at 37°C. After six ten minutes washes with MTSB, the tissues were permeabilized with 10 % DMSO and 3% IGEPAL CA-630 for one hour at room temperature. Slides were washed as above and then blocked with 3% BSA in MTSB for at least 2 hours at 37° C. Incubation with primary antibodies took place at 37° C and was continued overnight at 4° C.

Subsequently, slides were washed as described above and then incubated with CY3-conjugated anti-Rabbit secondary antibody diluted 1:250 in 3% BSA in MTSB for two hours at 37°C. After washing, specimens were covered with a drop of 50% glycerol in MTSB and cover slip was put on top. The edges of the coverslip were sealed with nail polish and the samples were examined by confocal microscopy as described below (Chapter 4.9.2).

4.9 Microscopy

4.9.1 Transmitted light and fluorescence microscopy

A Zeiss Discovery V8 stereomicroscope or a Zeiss Axioscope was used for transmitted light and fluorescence microscopy. For documentation, a Zeiss Axiocam MRc 5 was used.

4.9.2 Confocal microscopy

For confocal microcopy, a LSM 510 Meta confocal microscope from (Zeiss) was used. Green fluorescent protein and its fusions were excited with an argon laser at 488 nm and emission was detected using a 505 to 550 nm band pass filter. In immunolocalization studies, a Helium/Neon-Laser was utilized to excite the CY3 fluorescent dye at 543 nm, with detection between 585 and 615 nm.

5 RESULTS

5.1 Generation of an RKD2-induced cell line with egg cell-like identity

Approaches such as immunopurification or RNA-Seq require a considerable amount of cell material. Isolating egg cells from Arabidopsis ovules in sufficient amounts for such biochemical approaches is therefore an unattainable goal. In order to obtain sufficient egg cell material for these kinds of studies, a proliferating callus culture with egg cell-like identity was generated. Köszegi et al. (2011) reported that the ectopic expression of a plant specific RWP-RK domain-containing transcription factor termed RKD2, which is preferentially expressed in the egg cell of Arabidopsis, induces cell proliferation and a shift of gene expression towards an egg cell-like transcriptome.

For the generation of RKD2-induced callus, the coding sequence of the RKD2 gene was amplified from cDNA which was isolated from mature pistils of *Arabidopsis thaliana* ecotype Columbia (Col-0) plants. The coding sequence was subcloned into the pENTR/dTOPO vector Gateway technology was used to transfer the coding sequence into the plant expression vector pH7FWG2.0 (Karimi et al. 2002) in order to express RDK2-GFP fusion protein under the control of the *CaMV 35S* promoter (Chapter 4.4.9). The resulting expression vector pURM18 was transformed into *Arabidopsis* Col-0 plants (Chapter 4.2).

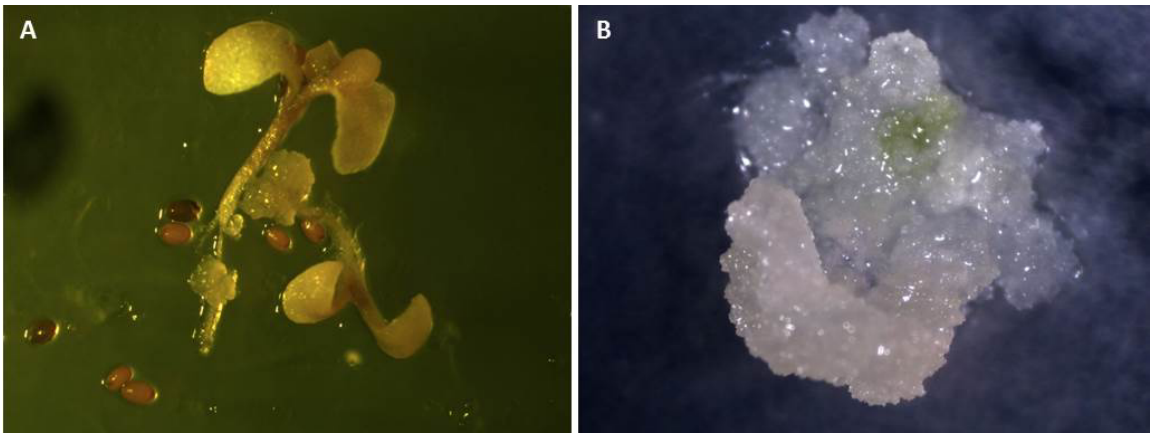


Figure 15: Generation of RKD2-induced egg cell-like callus.
(A) Transgenic seedling transformed with *35Sp:RKD2-GFP* and germinated on selection medium. The proliferating cell mass appears at the primary roots (arrows) or on the cotyledons. **(B)** RKD2-induced callus divided into a green differentiating part (arrow) and a colorless undifferentiated part.

For selection, seeds of transformed plants were laid out on half strength MS plates containing 2% sucrose and 30µg/mL Hygromycin. As a control callus an auxin-induced cell culture was used. This cell culture was generated from roots of 1 week old Arabidopsis Col-0 plants, which were grown under sterile conditions on MS-plates. Approximately one cm long segments were cut from the roots and placed on callus induction medium (CIM). This medium contains 0.5 mg/L 2,4-dichlorophenoxyacetic acid (2,4-D), and 0.05 mg/L Kinetin. By contrast, for the cultivation of RKD2 induced callus tissue, no addition of phytohormones was necessary.

In transgenic plants expressing *RKD2-GFP* under control of the *CaMV 35S* promoter, proliferating callus tissue appeared either at the hypocotyl, the primary root or the cotyledons, after seven to ten days of selection (Figure 15A). Callus tissue was transferred onto fresh medium every 14 days and subcultured into two parts when the diameter of the calli reached 5 to 7 mm.

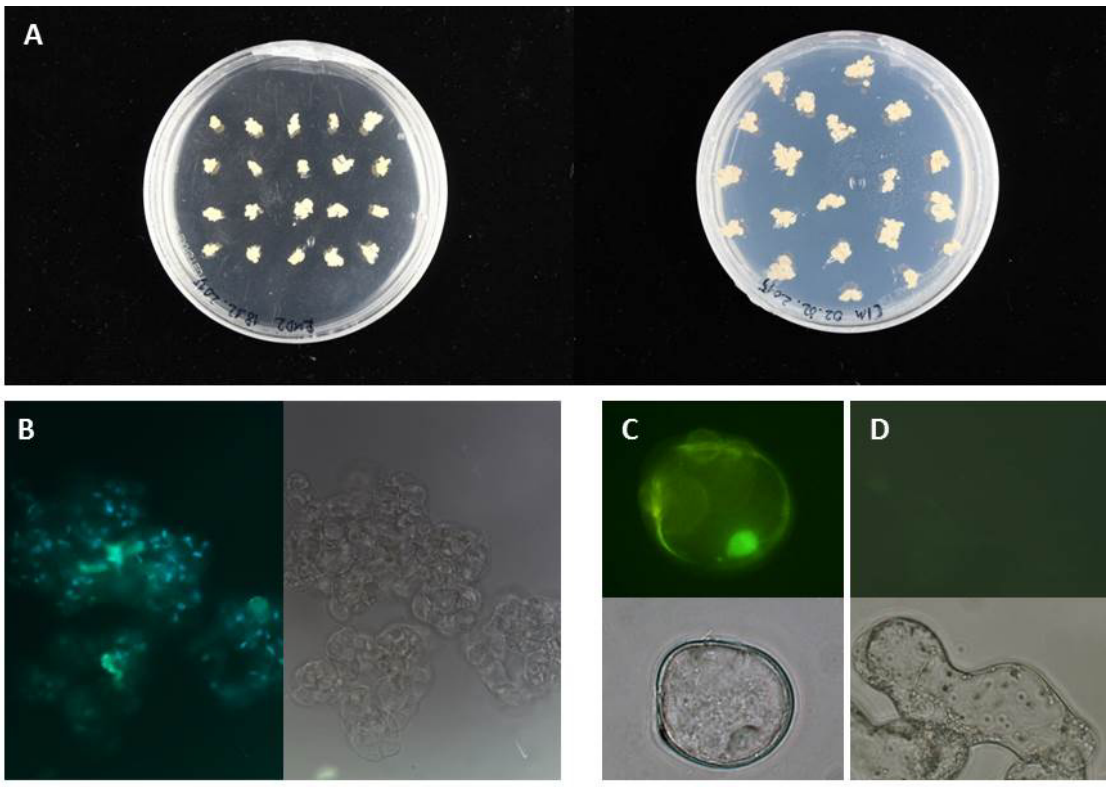


Figure 16: Cultivation of RKD2-induced callus.
(A) Plates with RKD2-induced callus, growing on hormone-free medium (left plate) and control callus (CIM) induced by auxin and cytokinin from seedling root segments. **(B)** GFP fluorescence of RKD2-induced callus expressing RKD2-GFP, examined by fluorescence microscopy. **(C)** Single cell derived from RKD2-induced callus showing RKD2-GFP fluorescence in the cell nucleus. **(D)** Single cell from CIM callus.

Proliferating RKD2-induced callus tissue appeared colorless and eventually formed some cell cluster of green differentiated cells (Figure 15B). According to Köszegi et al. 2011, only in the colorless tissue the RKD2 fusion protein is present, and this callus tissue could be cultivated for up to six months. For subcultivation, calli were divided into parts of approx. 2-3 mm diameter and transferred onto fresh plates (Figure 16A). Samples of RKD2 callus were checked for RKD2-GFP presence by fluorescence microscopy. RKD2-induced callus showed RKD2-GFP fluorescence in the cell nucleus (Figures 16B and C). As a control CIM callus cells were also checked for fluorescence and no signal could be detected (Figure 16D).

The first two generated RKD2-induced cell lines started to show signs of re-differentiation after five months and could be cultivated for only two more weeks. Re-differentiation manifested itself with the appearance of root, leaf-like or embryo-like structures (Figure 17A and B). Since those structures show no egg like identity, the cultivated cell lines were discarded.

However it was possible to generate two cell lines that kept dividing for more than six months (so far, up to 36 months), without showing any differentiation of root-like or embryo-like structures. Occasionally observed green cell clusters, indicating the upcoming of re-differentiating tissues, were carefully removed and the white egg cell-like part of the callus was subcultivated every 14 days.

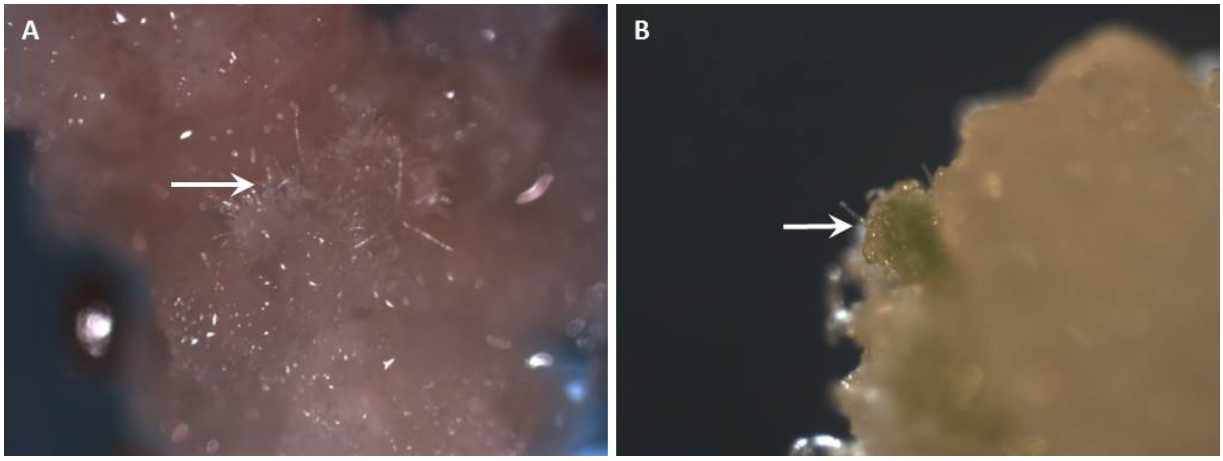


Figure 17: Re-differentiation of RKD2-induced callus.
(A) Root-like structure appearing on undifferentiated callus tissue (arrow).This callus has been cultivated for 2 and a half months. **(B)** Embryo-like green structure (arrow). This structure appeared after 3 months of cultivation.

Generation of the RKD2 induced cell line provided material for subsequent expression studies using RT-PCR (Chapter 5.2.1), for Western Blot analyses detecting AGO proteins (Chapter 5.2.2) and for the analysis of *AGO8* and *AGO9* transcript variants (Chapter 5.7). It was also used for the generation of mRNASeq and small RNA-Seq data (Chapters 5.8). In all those studies the CIM callus was used as a reference tissue.

5.2 Analysis of Argonaute expression patterns

To investigate the expression of all ten Arabidopsis AGOs in different plant tissues and to identify the members of this important gene family that are strongly present in the reproductive tissues and in the RKD2-induced egg cell-like cell line, RT-PCR studies and Western Blot experiments were performed. For *AGO8* and *AGO9,* both promoter studies and expression studies using gene-reporter fusions were performed during ovule and gametophyte development. Since no correct splice variant for *AGO8* was identified, whole mount immunolocalization experiments in developing ovules and embryos were focused on AGO9, using a specific AGO9 peptide antibody.

5.2.1 Differential expression of *Argonaute* genes in different plant tissues

The important role of sRNA pathways in plant reproduction has been implicated from array data analyzes using isolated female gametophyte cells (Soljic 2012). Additionally, *AGOs* show a very distinct expression pattern in vegetative and generative tissues (Mallory and Vaucheret 2011). Although some members of the AGO family have been reported to play a role in gamete formation (see Introduction), the exact role of AGOs and associated proteins in the female gametophyte is not yet clear.

A RT-PCR expression analysis of the ten *AGOs* in different vegetative and reproductive tissues of the Arabidopsis plant and included the RKD2-induced cell line and the CIM callus in these studies. Three members of the *AGO* family; *AGO5*, *AGO8* and *AGO9* show increased expression in reproductive tissues (Figure 18). *AGO9* is not detected in vegetative tissues, but strongly present in developing pistils in stages 10 to 14 and anthers from the same stages which the expression levels decreasing in stages 12 and 14. *AGO5* also shows strong presence in reproductive tissues, but can be detected at a low level in the shoot.

The *AGO8* gene, is found in almost all tissues at a very low level, but with expression peaks in developing ovules. On the transcript level, all ten *AGOS* are present in the RKD2-induced callus. The phytohormones-induced CIM callus shows a different expression pattern of *AGO* genes. The *AGO* members which were found predominantly in reproductive tissues are either weakly expressed (*AGO5* or *AGO8*) or not detectable (*AGO9*). All other Argonautes show high expression levels.

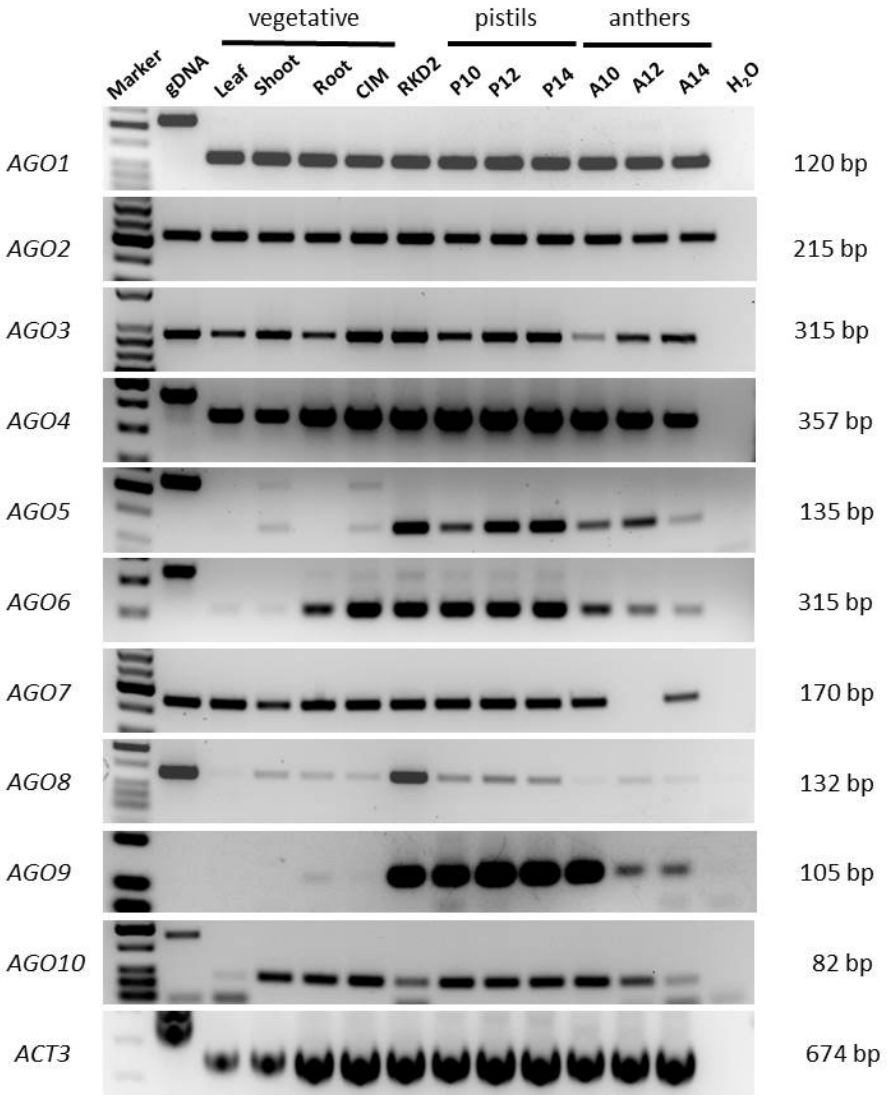


Figure 18: Expression of Arabidopsis *AGO* genes.
RT-PCR based expression study of the ten Arabidopsis *AGO* expression was tested in different vegetative and reproductive tissues, as well as in CIM and RKD2 callus tissues. *ACT3*, *ACTIN3*; P10, pistil stage 10, P12, pistil stage 12; P14, pistil stage 14; A10, anther stage 10; A12, anther stage 12; A14 anther stage 14. Anther and pistil stages according to Smyth et al. (1990). Primer pairs used for amplification can be found in the appendix. For *AGO8* 10 ng of plasmid DNA (pGRU1) was used as a positive control.

5.2.2 Differential expression of AGO proteins in different plant tissues

Available antibodies against Arabidopsis Argonaute proteins were obtained from Agrisera (Vännas, Sweden) and tested for their specificity in a Western Blot using different plant extracts (Figure 19). Antibodies against AGO1, 2, 4, 5, 6 and 9 were available. Protein extracts from leaves, root, mixed stage flowers and siliques as well as extracts from RKD2 and CIM callus cell were used to test the antibodies.

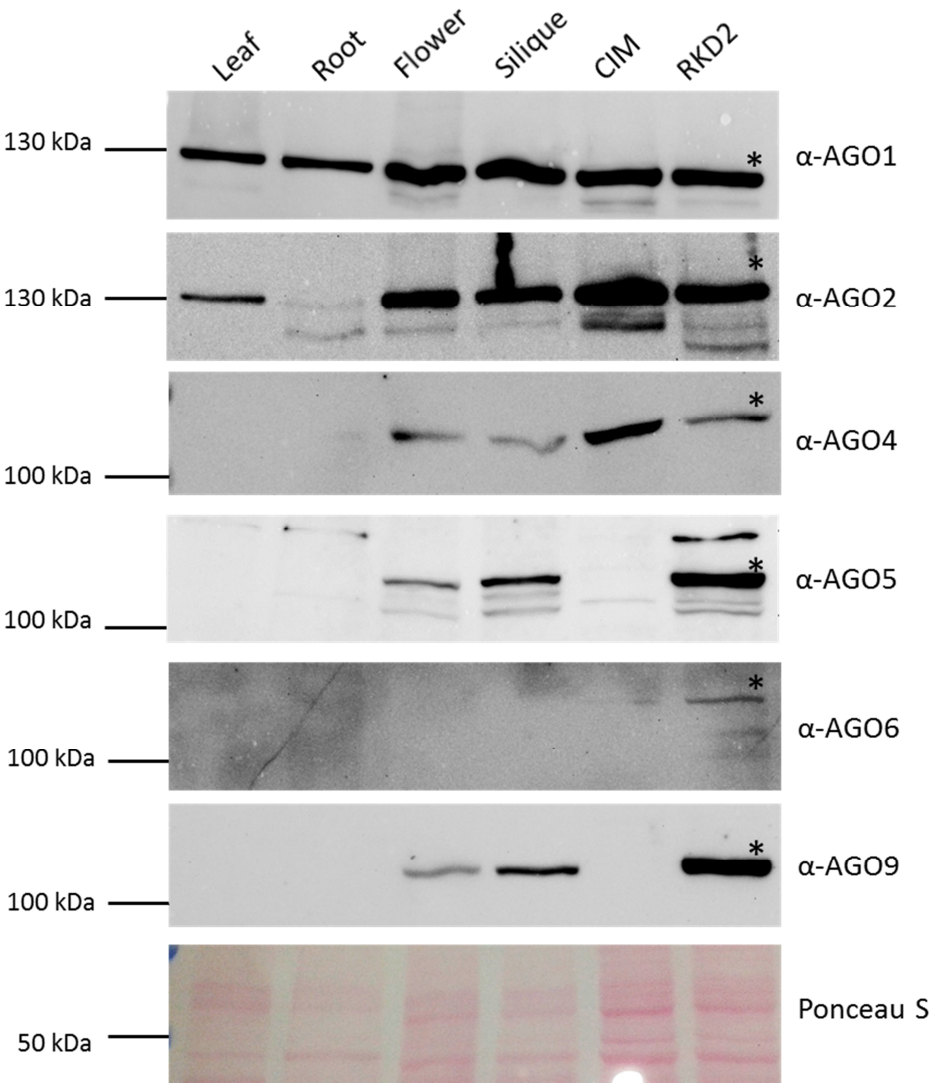


Figure 19: Western Blots showing the presence of AGO proteins in different Arabidopis tissues.
All available antiAGO antibodies were tested against different tissues as indicated. 30 µg of protein were loaded per lane. Specific bands recognized by the antibody are marked with an asterisk. Expected/apparent protein size: AGO1: 116/130 kDa, AGO2: 113/130 kDa, AGO4: 103/103 kDa, AGO5: 111/111 kDa, AGO6:116/116 kDA, AGO9: 100/115 kDa.

AGO1 was detected in every tested tissue, showing an apparent molecular weight of 130 kilo Dalton (kDa), which differs from the calculated molecular weight of 116.4 kDa. AGO2 is also present in all tissues but shows less expression in roots. The antiAGO2 antibody recognizes a 130 kDa band which is bigger than the predicted 113 kDa. AGO4 shows clear signals in flowers, siliques and the both callus tissues and a faint band in roots. The observed band has an apparent MW of 115 kDa, the calculated MW is 103 kDa. AGO6 is only detected in RKD2 and CIM callus. AGO5, as well as AGO9 is only present in reproductive organs and the egg cell-like, RDK2-induced callus.

Abundance of *AGO* transcripts corresponds in most of the examined tissues to the protein abundance. In case of *AGO2,* transcript can be found at a considerable amount in the roots, but the protein is hardly detectable. *AGO4* transcript is present in all tissues examined by RT-PCR, but the protein is only present in weak amounts in roots and is absent from leaves. This confirms the work of Havecker et al. 2010, where *AGO4* transcript is detected in mixed floral tissues. A difference between protein and transcript abundance is also observed for *AGO6*, which has high transcript levels in roots, reproductive organs as well as RKD2 and CIM callus, but in both callus tissue the protein could be detected at low levels.

Comparing RDK2 and CIM-induced callus, it is striking that in the RKD2 callus all *AGOs* could be detected at transcript level, also the corresponding proteins could be detected, in case there was an antibody available. In the CIM callus the presence of AGO1, 2, 4 and 6 could be confirmed on both transcript and protein level, but AGO5 and AGO9 were absent. AGO5 and AGO9 are strongly present as protein in reproductive tissues (Chapter 3.2.9). Testing of the AGO antibodies revealed the presence of unspecific bands in all Western Blot, except for the AGO9 antibody, which was used for whole-mount immunolocalization studies (Chapter 5.6).

5.3 Analysis of *AGO8* and *AGO9* promoter activities

The expression of *AGO8* and *AGO9* within the developing female gametophyte was studied using promoter-reporter constructs. The promoter regions (637 bp for *AGO8* and -2033 bp for *AGO9*) were amplified from genomic DNA and cloned into a Gateway-based plant expression vector from which a 3xGFP reporter protein with nuclear localization sequence (NLS) is expressed under the control of the cloned promoter.

Activity of the *AGO8* promoter

Expression analyses of *AGO8* in different tissues showed high amounts of transcript in pistils of stage 10, 12 and 14, as well as in anthers of those stages (Figure 18). Furthermore, transcriptome profiling using isolated gametophytic cells showed *AGO8* expression in the egg cell, but not the central cell or synergid cell (Soljic 2012).

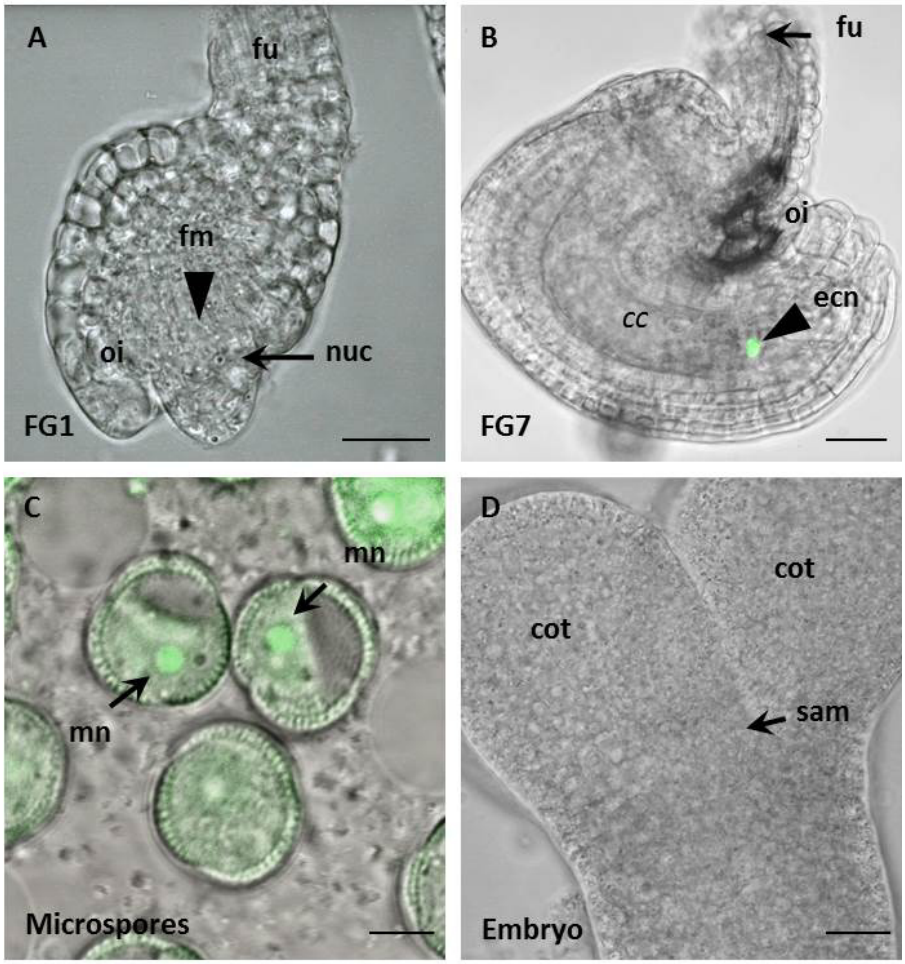


Figure 20: Activity of the *AGO8* promoter.
In the mature ovule, the *AGO8* promoter drives an egg cell-specific expression of the reporter protein NLS-(3x) GFP. Plants were analyzed by CSLM and single opical sections are shown **(A)** In FG1 ovules no activity of the *AGO8* is found **(B)** In mature ovules, at FG7 the promoter is active in the egg cell (C) In one of 15 lines a GFP-signal could be observed in the developing microspores. **(D)** No *AGO8* promoter activity is detected in torpedo stage embryos. cc, central cell; cot, cotyledons; ecn, egg cell nucleus; fm, functional megaspore; fu, funiculus; mn, microspore nucleus; nuc, nucellus, oi; outer integument scale bars: 20 µM (A, B, D) and 5 µM (C). Female gametophyte (FG) stages were defined according to Christensen et al. 1998.

The analysis of 15 independent transgenic lines (T1 generation) driving the NLS-3xGFP reporter under control of the *AGO8* promoter revealed that the reporter is detected in the mature ovule at female gametophyte stage 7 (FG7, according to Christensen et al. 1997) where a strong GFP-signal is observed only in the egg cell (Figure 20B). In all other stages of the developing ovule (FG0, FG1 – FG6) no GFP signal was detected (Figure 20). In contrast to this, in 1 out of 15 independent lines a signal was observed in the developing pollen, in the microspore (Figure 20C). No reporter activity was observed in any vegetative tissue investigated, including different embryo stages (Figure 20D).

Activity of the *AGO9* promoter

In contrast to the *AGO8* promoter, the *AGO9* promoter is not egg cell-specific, but active in other cell types and more developmental stages of the ovule. In the developing ovule at female gametophyte stage 0 (FG0), a signal from the NLS-(3x) GFP reporter is observed in the nucellus tissue surrounding the female gametophyte and in the funiculus (Figure 21 A and B).

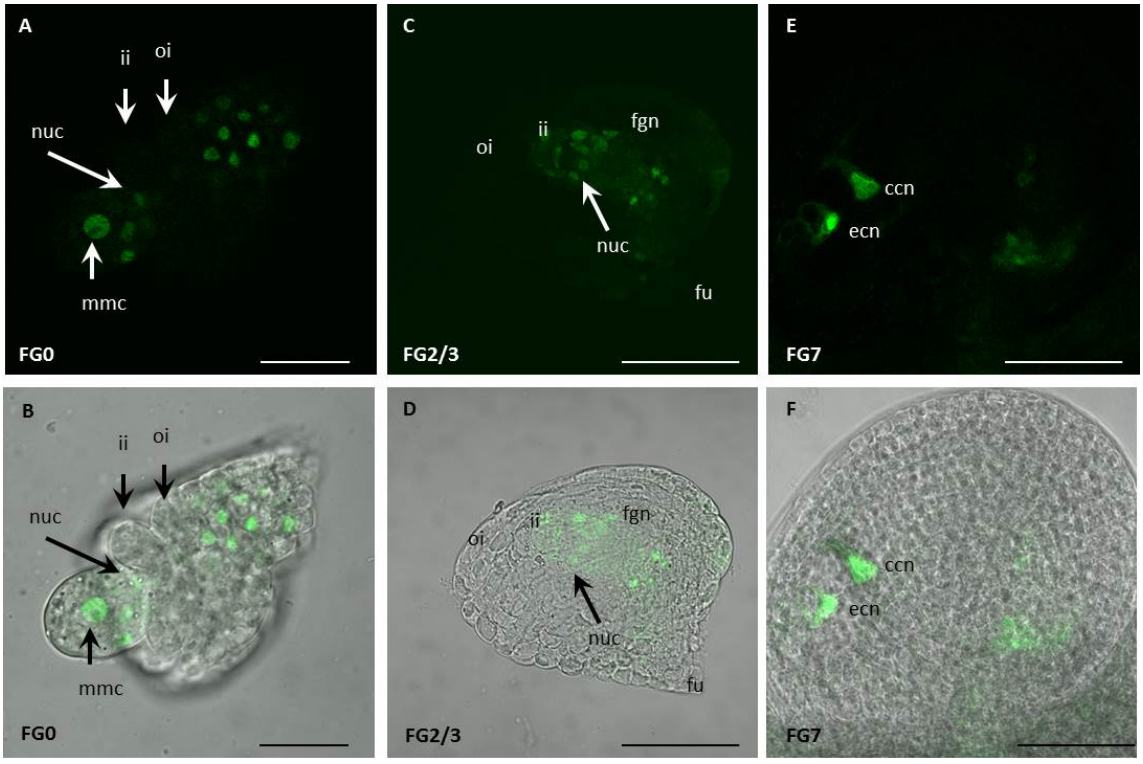


Figure 21: Activity of the *AGO9* promoter in the developing female gametophyte. (Legend on next page)

Figure 21: Activity of the *AGO9* promoter in the developing female gametophyte.
Ovules of plants expressing NLS-(3x) GFP reporter under the regime of the *AGO9* promoter were analyzed by CLSM. Single optical sections are shown. **(A)** and **(B)** AGO9 promoter activity is found in the megaspore mother cell and the surrounding nucellus cells of the ovule. **(C)** and **(D)** At female gametophyte stage 2/3 (FG2/3) the AGO9 promoter shows strong activity in the nucellar tissue of the female gametophyte. **(E)** and **(F)** The mature ovule (FG7) shows promoter activity in the nuclei of the egg cell and the central cell. mmc, megaspore mother cell; nuc, nucellus; ii, inner integument; oi, outer integument; fgn, female gametophyte nuclei; fu, funiculus; ccn, central cell nucleus; ecn, egg cell nucleus. Scale bars are 20 µM. FG stages were defined according to Christensen et al. 1998.

Additionally, *AGO9* promoter activity can be detected as well in the developing embryo (Figure 22). In the globular stage it can be found in a few cells of the protoderm (Figure 22 A and B). In the late heart stage embryo, the activity is concentrated in the region of the shoot apical meristem (Figure 22 C and D). Also in the fully developed seedling the AGO9 promoter activity is restricted to the shoot apical meristem (Figure 22 E and F).

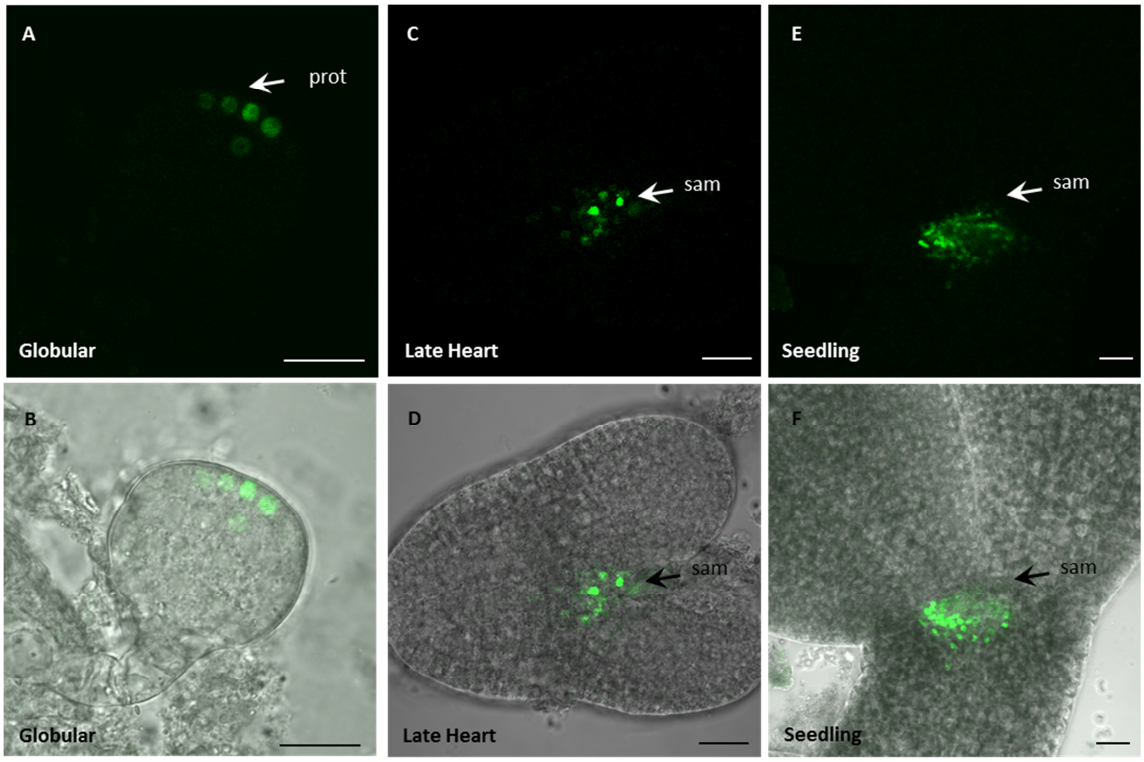


Figure 22: Activity of the *AGO9* promoter in the developing embryo.
Embryos of plants expressing NLS-(3x)GFP reporter under the regime of the *AGO9* promoter were analyzed by CLSM. Single optical sections are shown**. (A)** and **(B)** In the globular stage *AGO9* promoter activity is restricted to a few cells in the protoderm. **(C)** and **(D)** In the heart stage embryo, *AGO9* promoter activity is restricted to a few cell in the region of the shoot apical meristem (sam) **(E)** and **(F)** In the mature seedling *AGO9* promoter activity appears only in the shoot apical meristem scale bars are 20 µM prot, protoderm; sam, shoot apical meristem.

Table 1 summarizes the observed *AGO9* promoter activity in the developing ovule, classified according to the different developmental stages of the female gametophyte, and in the embryo. In total, twelve independent transgenic lines of the T1 generation have been analyzed. Nine lines showed a strong GFP signal, two lines showed a weak signal and only one line had no signal.

Table 1: *AGO 9* Promoter activity in different developmental stages of the ovule/embryo sac and the embryo. Twelve independent transgenic lines (T1) have been analyzed. Nuc = Nucellus, MMC = megaspore mother cell, EC = egg cell, CC = central cell, SAM = shoot apical meristem.

		GFP-Localization							
Line No.	**GFP-Signal**	**FG0**	**FG2/3**	**FG6/7**	**8/16 cell**	**Globular**	**Heart**	**Pollen**	**Seedling**
1	Strong	Nuc, MMC	Nuc	EC CC	Whole embryo	Protoderm	SAM	-	SAM
2	Weak	-	Nuc	EC, CC	-	-	-	-	
3	None	-	-	-	-	-	-	-	-
4	Strong	Nuc, MMC	Nuc	EC, CC	Whole embryo	Protoderm	SAM	-	SAM
5	Strong	Nuc, MMC	Nuc	EC, CC	Whole embryo	Protoderm	SAM	-	SAM
6	Strong	Nuc, MMC	Nuc	EC, CC	Whole embryo	Protoderm	SAM	-	SAM
7	Strong	Nuc, MMC	Nuc	EC, CC	Whole embryo	Protoderm	SAM	-	SAM
8	Weak	-	-	EC, CC	-	-	-	-	-
9	Strong	Nuc, MMC	Nuc	EC, CC	Whole embryo	Protoderm	SAM	-	SAM
10	Strong	Nuc, MMC	Nuc	EC, CC	Whole embryo	Protoderm	SAM	-	SAM
11	Strong	Nuc, MMC	Nuc	EC, CC	Whole embryo	Protoderm	SAM	-	SAM
12	Strong	Nuc, MMC	Nuc	EC, CC	Whole embryo	Protoderm	SAM	-	SAM

5.4 Expression and subcellular localization of AGO9 proteins

The subcellular localization of AGO8 and AGO9 was studied by transient expression of tobacco BY-2 cells and in roots of stable transformed Arabidopsis plants. Furthermore, the protein abundance and subcellular localization of AGO9 during ovule and embryo

development was investigated by analyzing transgenic plants expressing the GFP-AGO9 fusion protein under control of the *AGO9* promoter and by whole-mount immunolocalization.

5.4.1 Subcellular localization after transient expression

In previous studies, both *AGO8* aberrant spliced transcript variants were used in transient expression studies using tobacco BY-2 cells (Wartlsteiner 2010). The green fluorescent protein was fused N-terminally to the coding sequence of both variants and the expression was driven by the constitutive *CaMV 35S* promoter. In case of GFP-AGO8-4, in one cell a faint and diffuse GFP fluorescence could be detected in the nucleus and the cytoplasm, with signals accumulating at the periphery of the nucleolus and in a few speckles in the nucleoplasm. The fusion of GFP to the AGO8-8 variant was also used for transient expression in tobacco BY-2 cells, but no signal from GFP fusion protein could be observed (Wartlsteiner 2010). Transient expression of GFP-AGO8-4 was repeated and resulted in the same fluorescence pattern as described above.

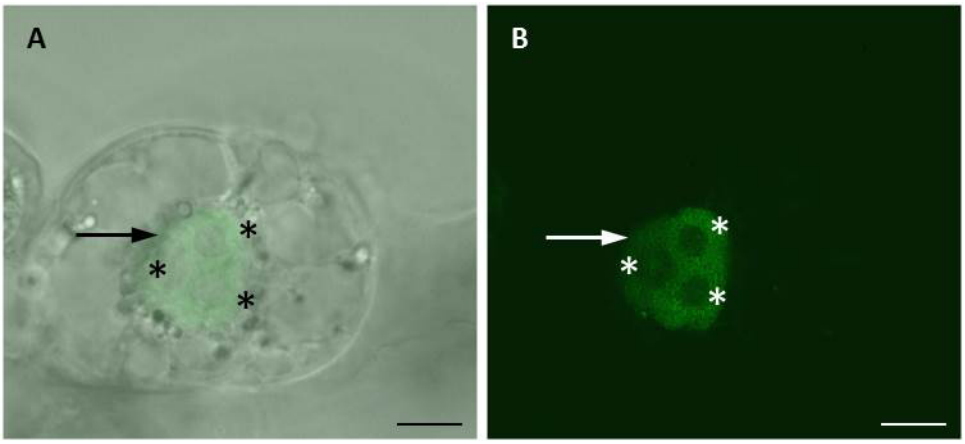


Figure 23: Transient expression of *GFP-AGO9* in tobacco BY2 cells.
A single tobacco BY2 cell expressing *GFP-AGO9* under the control of the constitutive *CaMV 35S* promote**r** is shown. **(A)** Bright field image. **(B)** Fluorescence image. GFP-AGO9 localizes within the entire nucleus (arrow) but excludes the nucleolus (asterisks). Scale bars are 10 µM.

The expression of *AGO9* variants gave different results: transient transformation with the intron-retained *AGO9-8* variant did not result in any fluorescence of the GFP reporter. In contrast to this, the correctly spliced *AGO9* CDS, fused to *GFP*, depicted a distinct subcellular localization pattern: GFP-AGO9 signals were always absent from the cytoplasm, instead fluorescence was exclusively found in the nucleus, with the nucleolus left out (Figure 23).

Generally, the observed fluorescence of GFP-AGO9 in BY-2 cells was weak and could be found in a few of the transformed cells. .

5.4.2 Subcellular localization in stably transformed plants

The same construct that was used for transient localization studies of GFP-AGO9 in tobacco BY-2 cells (*35Sp:GFP-AGO9*) was also used for stable transformation of Arabidopsis Col-0 plants. 12 independent transgenic lines were created. Progeny seeds of transgenic plants were surface sterilized and grown on half strength MS plates under sterile conditions.

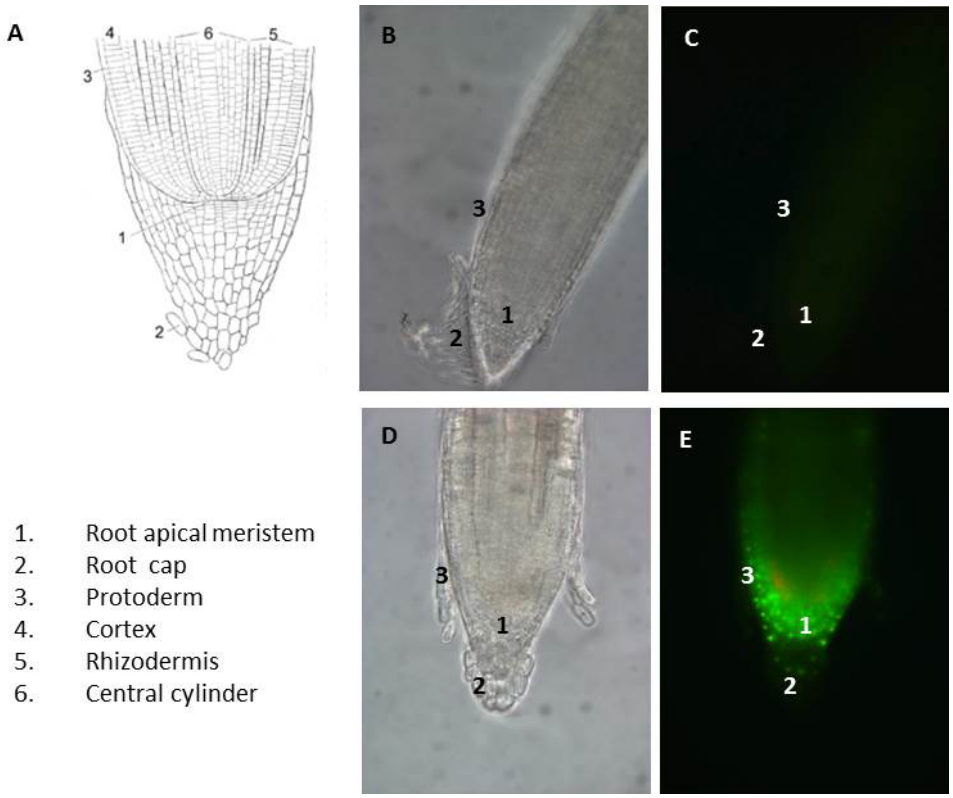


Figure 24: Ectopic expression of *GFP-AGO9* under control of the constitutive *CaMV 35S* promoter.
(A) Schematic representation of root tissues (Jennifer Szymanski on study.com). **(B)** and **(C)** Root of Col-0 control, analyzed by fluorescence microscopy. **(B)** Brightfield image, **(C)** fluorescence image. **(D)** and **(E)** Root of transgenic seedling expressing *35Sp:GFP-AGO9*, analyzed by fluorescence microscopy. **(D)** Brightfield image, **(E)** Fluorescence image of root shown in (D). The strongest signal of GFP-AGO9 is visible in the nuclei of the root apical meristem (1), the root cap (2) and the protoderm (3). The roots of *35Sp:GFP-AGO9*-expressing seedlings did not show any obvious developmental phenotypes.

14 days after germination, the roots of the seedlings were analyzed by CLSM for GFP-AGO9 fluorescence. Like in tobacco BY-2 cells (Figure 23), the GFP-AGO9 fusion protein was exclusively detected in the nucleus of expressing cells. Interestingly, the signal of the

GFP-AGO9 fusion protein was only detected in the nuclei of the root tip, including the lateral root cap, the root epidermis, the columnella and the root apical meristem (Figure 24). No expression was observed in the central cylinder of the root and in the vasculature.

This stands in contrast to previous observations that the *CaMV 35S* promoter is strongly active in the root vasculature (Benfey and Chua 1990, Jasik et al. 2011) and suggests that the stability of the GFP-AGO9 fusion protein in the cells of the root meristem and in the vasculature may be different.

5.4.3 Expression of *AGO8* transcript variants under control of their native promoter

Analysis of *AGO8* transcripts resulted in the discovery of two aberrantly spliced variants (Wartlsteiner 2010). No transcript encoding a full length AGO8 protein could be isolated. In order to investigate the potential presence of truncated proteins in the egg cell, Arabidopsis Col-0 plants were stably transformed with N-terminal GFP-fusions of the two AGO8 variants under the control of the endogenous AGO8 promoter.

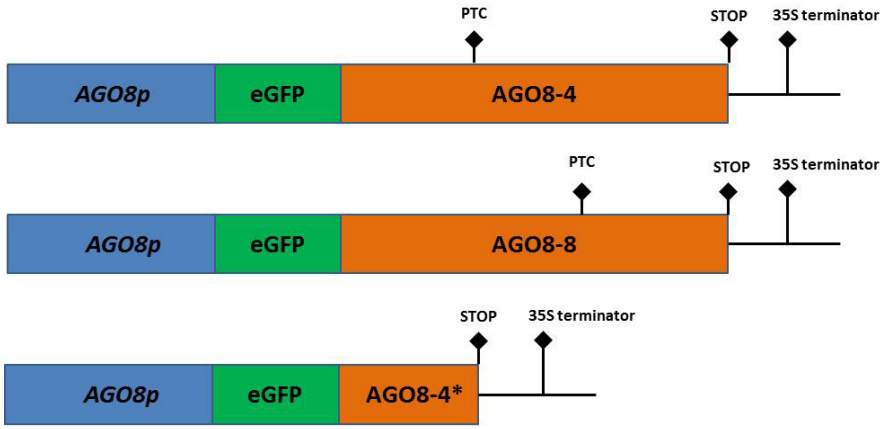


Figure 25: Expression of *AGO8* splicing variants under the control of the endogenous *AGO8* promoter
Schematic representation of constructs showing the positions of the *AGO8* promoter, the enhanced green fluorescent protein (eGFP) and the AGO8 cDNAs. The upper two constructs contain a premature stop codon (PTC) within the cDNA, whereas in the third construct, the cDNA ends with a true stop codon. All constructs contain a *35S* terminator.

The two aberrantly spliced *AGO8* transcript variants were cloned into the modified Gateway-based expression vector pB7WGF2.0 (Karimi et *al.*, 2002). As a result two transcripts encoding truncated N-terminal GFP-fusion proteins are expected to be produced in the egg cell, due to the premature termination codon (PTC) in the transcripts. Additionally, a

shortened version of *AGO8-4*, which was called *AGO8-4** was used. In this construct the open reading frame ends directly after the first PTC and is thus not prone to degradation by nonsense-mediated mRNA decay (NMD).

For each construct 12 independent transgenic lines were analyzed (Table 2). Considering the results of *AGO8* promoter-reporter studies (Chapter 5.3), the strongest signal of the truncated GFP-AGO8 protein would be expected in the egg cell at gametophyte stage FG7. Nevertheless, no signal was to be seen for any of the plant lines generated. This applies also for the transcript variant *AGO8-4**, in which the *AGO8* coding sequence ends directly after the first premature termination signal and cannot be degraded by NMD (Figure 25).

Table 2: Expression of *AGO8* splicing variants under the control of the endogenous *AGO8* promoter.

	Independent transgenic lines		
	GFP signal	weak signal	no GFP signal
AGO8p:GFP-AGO8-4	-	-	12/12
AGO8p:GFP-AGO8-8	-	-	12/12
*AGO8p:GFP-AGO8-4**	-	-	12/12

5.4.4 Expression of GFP-AGO9 under the control of the *AGO9* Promoter

Analyzing gene-reporter fusions under the control of endogenous promoter is a method to observe protein presence in different tissues and during development. GFP-AGO9, expressed under control of the *AGO9* promoter, was analyzed in the Col-0 ecotype background. In total, 12 independent transgenic lines were created (Table 3). One line only showed a weak and two lines presented a strong GFP-AGO9 signal. In the two strong expressing lines (*AGO9p:GFP-AGO9*-3 and -12) the fluorescence appeared in FG0 ovules in cell nuclei of the funiculus and of nucellar tissues but was missing in the megaspore mother cell (Figure 26A, B).

During further ovule development, in the mature female gametophyte at stage FG7, the GFP-AGO9 signal was only present in nuclei of nucellar cells and in nuclei of funiculus cells but is absent from all cells inside the female gametophyte, which clearly stands in contrast to the prevalence of the *AGO9* promoter activity, which also appears active in generative cells like in megaspore mother cell, the egg- and the central cell (Figure 21). After

fertilization GFP-AGO9 reappears in the zygote of the two lines *AGO9p:GFP-AGO9*-3 and -12 (Figure 27 A, B). In young embryos of the 4-cell stage GFP-AGO9 can be seen in the nuclei of every cell (Figure 27 C, D). In the globular embryo the signal is observed in four cells of the protoderm, but also in one cell below this tissue. In the late heart stage, the GFP-AGO9 signal is concentrated in the region of the shoot apical meristem. In the mature seedling GFP-AGO9 is detected in the shoot apical meristem. Table 3 summarizes the observed GFP-AGO9 signal in the developing ovule, classified according to the different developmental stages of the female gametophyte (Christensen et al. 1998), and in the embryo. Like shown in the transient assays (Chapter 5.4.1) and in plants expressing GFP-AGO9 under the control of the *35S* promoter the protein (Chapter 5.4.2) is confirmed to be exclusively expressed in the nucleus.

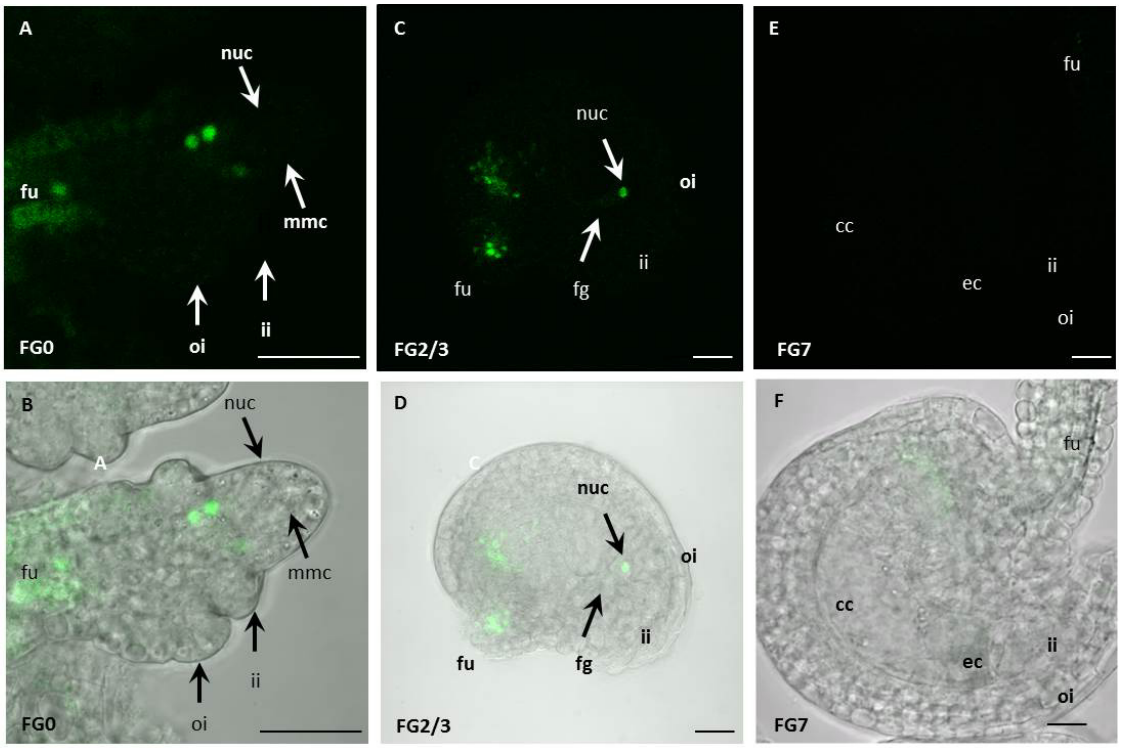


Figure 26: Localization of the GFP-AGO9 fusion protein under the control of the *AGO9* promoter during female gametogenesis. Samples were analyzed by CSLM and single optical sections are shown. **(A)** and **(B)** The fusion protein is found in the surrounding nucellus cells but not in the megaspore mother cell **(C)** and **(D).** At female gametophyte stage 2/3 (FG2/3) fusion protein is located in the funiculus and the nucellar tissue of the developing female gametophyte. **(E)** and **(F)** The mature ovule (FG7) shows no GFP-AGO9 signal. mmc, megaspore mother cell; nuc, nucellus; ii, inner integument; oi, outer integument; fu, funiculus; fg, female gametophyte, cc, central cell; ec, egg cell. Scale bars are 20 µM. FG stages were defined according to Christensen et al. 1998.

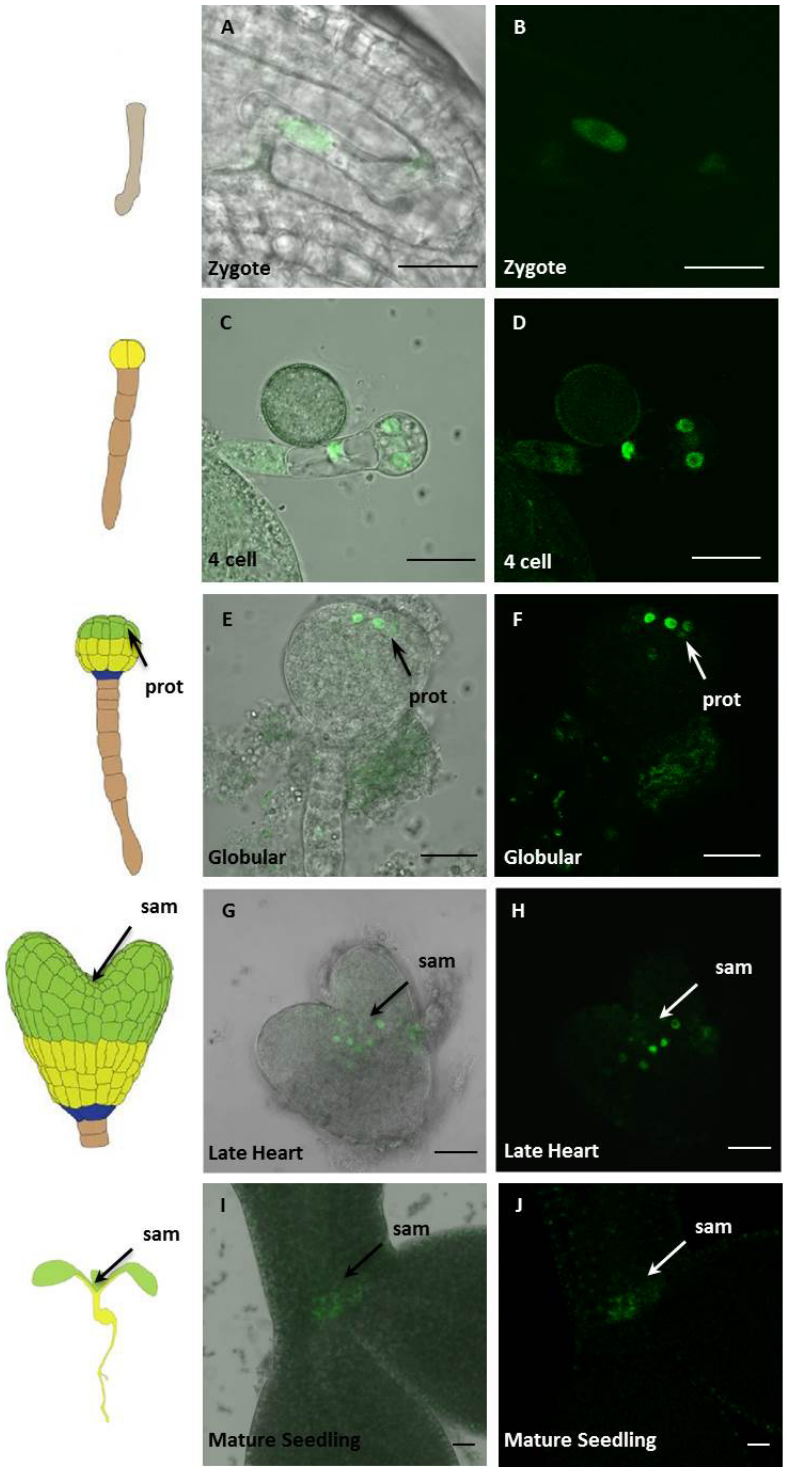


Figure 27: Expression of *AGO9p:GFP-AGO9* during embryogenesis.
Single optical sections (CSLM) are shown. **(A)** and **(B)** The fusion protein is the zygote nucleus. **(C)** and **(D)** GFP-AGO9 is detected in the nuclei of the four celled embryo. **(E)** and **(F)** At globular stage the signal is restricted to a few cells in the protoderm. **(G)** and **(H)** In late heart stage, the signal is concentrated in the region of the shoot apical meristem. prot, protoderm; sam, shoot apical meristem. Scale bars are 20 µM. Schematic representations of embryo stages adapted from ten Hove et al. 2015.

Table 3: GFP-AGO9 fluorescence in different developmental ovule and embryo stages of transgenic lines expressing *AGO9p:GFP-AGO9*. Twelve independent transgenic lines (T1) have been analyzed. Only three out of twelve independent transgenic lines showed fluorescence of GFP-AGO9. Nuc = Nucellus, Fu = Funiculus, EC = egg cell, CC = central cell, SAM = shoot apical meristem, Seedl. = Seedling (mature). Abbreviation in parentheses indicate weak expression.

		GFP-Localization							
Line No.	**GFP-Signal**	**FG0**	**FG2/3**	**FG6/7**	**Zyg.**	**4 cell**	**Globular**	**Heart**	**Seedl.**
1	None	-	-	-	-	-	-	-	-
2	None	-	-	-	-	-	-	-	-
3	Strong	Nuc, Fu	Fu, Nuc	(Fu, Nuc)	Yes	Yes	Protoderm	SAM	SAM
4	None	-	-	-	-	-	-	-	-
5	None	-	-	-	-	-	-	-	-
6	None	-	-	-	-	-	-	-	-
7	None	-	-	-	-	-	-	-	-
8	None	-	-	-	-	-	-	-	-
9	Weak	(Nuc)	-	-	-	-	-	-	-
10	None	-	-	-	-	-	-	-	-
11	None	-	-	-	-	-	-	-	-
12	Strong	Nuc	Fu, Nuc	(Fu, Nuc)	Yes	Yes	Protoderm	SAM	SAM

5.5 Ectopic overexpression of GFP-AGO9 in the egg cell

Homozygous T-DNA insertion lines for *ago8* and *ago9* did not show any obvious phenotype. Moreover, the generation of *ago8*/*ago9* double mutants was not possible due to their close physical proximity on chromosome 5. Therefore, the approach of ectopically expressing the functional *AGO9* transcript in the egg cell was chosen. The full-length cDNA of *AGO9* was expressed in plants under the regime of egg cell specific promoters, namely the *AGO8* promoter, which is active in the mature egg cell and the *EC1.1* promoter, which is also active in the egg cell but shows a much higher level of activity (Sprunck et al. 2012). As a control, transgenic lines expressing the TET9-GFP fusion protein (Sprunck et al. 2012) under the regime of the *EC1.1* promoter were generated.

5.5.1 Localization of *AGO8p:GFP-AGO9*

Transgenic plants with the *AGO8p:GFP-AGO9* construct only exhibited a GFP signal in a few cases. Only four of 21 independent transgenic plant lines showed a GFP signal, with two of those being very weak (Figure 28; lines *AGO8p:GFP-AGO9*-1, -3). In concordance with *AGO8* promoter-reporter studies (Figure 20), the GFP-AGO9 signal is exclusively found in the egg cell (Figure 29). However, the signal intensity of the GFP-AGO9 fusion protein is relatively weak. Furthermore, the fusion protein is localized to the nucleus, which was also shown for the GFP-AGO9 protein under the control of the native *AGO9* promoter (Figure 26). No correlation between ectopic GFP-AGO9 expression in the egg cell and a significant change in seed set was observed. Compared to the two weak expressers the lines *AGO8p:GFP-AGO9*-10 and *AGO8p:GFP-AGO9*-20 showed stronger GFP-AGO9 fluorescence but full seed set (Figure 28).

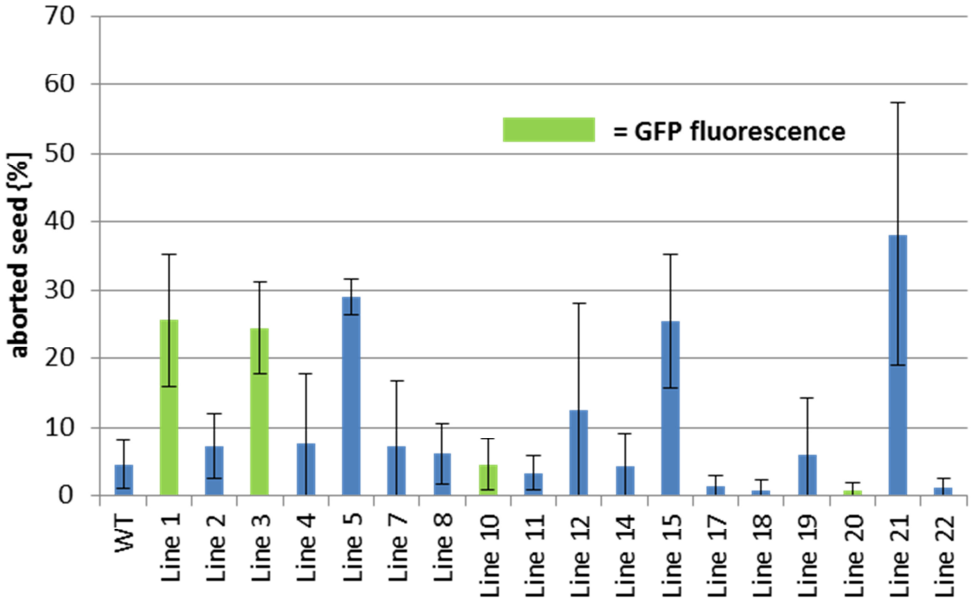


Figure 28: Ectopic expression of *GFP-AGO9* under control of egg cell specific *AGO8* promoter. Undeveloped seeds (in %) in siliques of the wild type and individual transgenic lines were evaluated. Plants of the first transgenic generation (T1) were examined. Seed abortion rates of transgenic plants with a GFP signal are shown in green. Plant lines -10 and -12 showed stronger GFP-AGO9 expression than lines -1 and -3. Seed abortion rates of plants without any GFP signal are shown in blue. In all lines the seeds from five siliques were counted. No correlation between GFP-AGO9 expression in the egg cell and the phenotype of undeveloped seeds was observed.

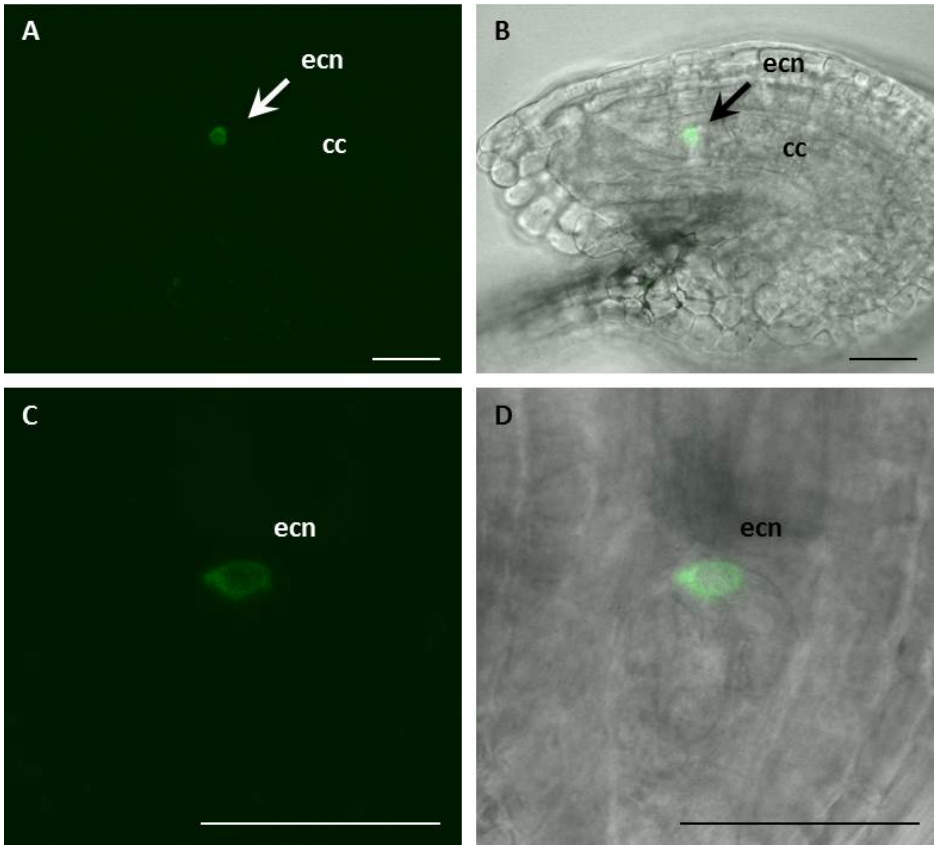


Figure 29: Ectopic expression of *GFP-AGO9* under control of the egg cell-specific *AGO8* promoter, analyzed by Confocal Laser Scanning Microscopy.
The transgenic line *AGO8p:GFP-AGO9*-10, which showed the strongest GFP-AGO9 signal, is shown. **(A)** and **(B)** show the position of the signal within the ovule, the signal is located in the egg cell. (A) Fluorescence image (B) Bright field image. **(C)** and **(D)** Detail of an ovule showing the subcellular localization of GFP-AGO9 in the nucleus of the egg cell. Scale bars are 20 µM.

5.5.2 Localization of *EC1.1p:GFP-AGO9*

The promoter of the *EC1.1* gene is specifically active in the egg cell at a high level (Sprunck et al. 2012). To ectopically express GFP-AGO9 in the egg cell the coding sequence of the fusion protein was cloned in an expression vector behind the *EC1.1* promoter to generate plants transgenic for *EC1.1p:GFP-AGO9*. In total, 12 independent transgenic lines were investigated for their GFP fluorescence in the egg cell and for the effect of ectopic GFP-AGO9 expression on seed set.

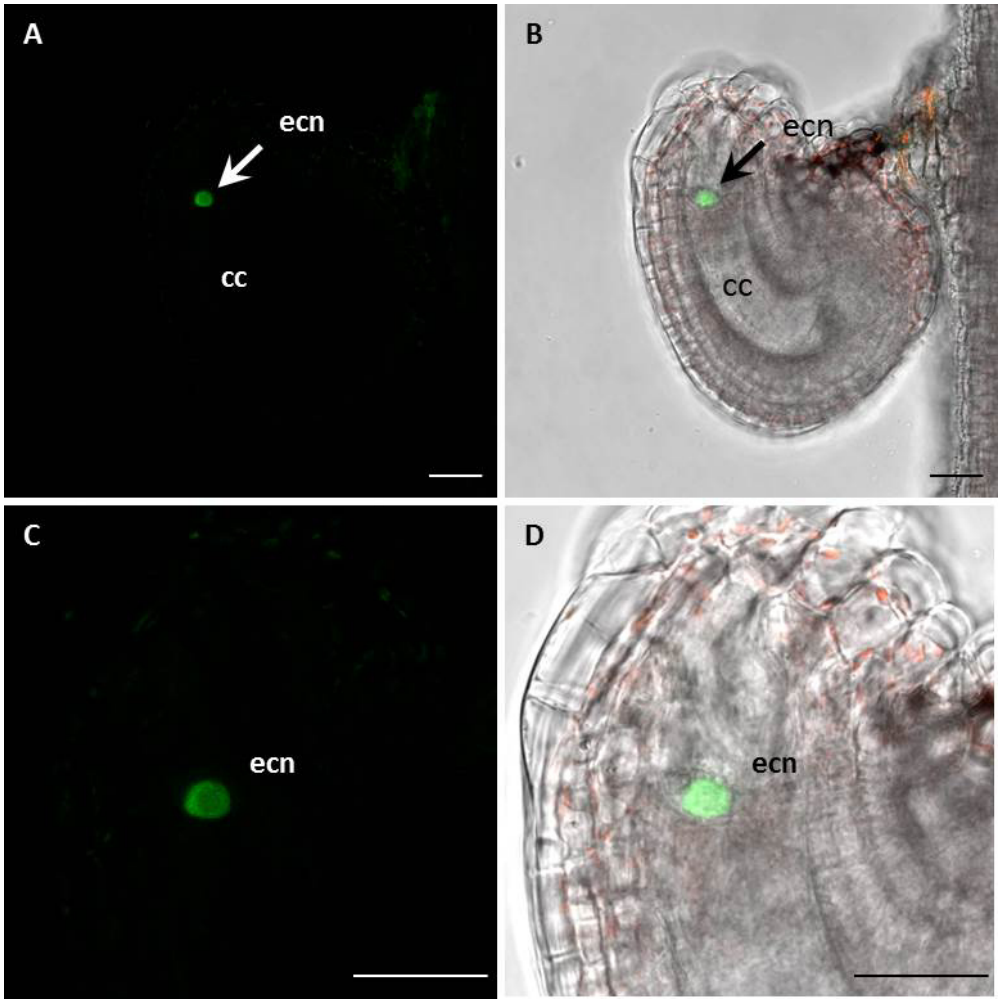


Figure 30: Ectopic expression of *GFP-AGO9* under control of the egg cell-specific *EC1.1* promoter, investigated by Confocal Laser Scanning Microscopy. Single optical sections are shown. **(A)** and **(B)** Ovule expressing the GFP-AGO9 fusion protein in the egg cell. **(C)** and **(D)** show a magnification of the same ovule. GFP-AGO9 locates to the egg cell nucleus (ecn). (A), (C) Fluorescence images (B), (D) Merged fluorescence and bright field images. Scale bars are 20 µM.

As a control, a construct expressing the fusion protein TET9-GFP under control of the EC1.1 promoter was transformed and resulting independent transgenic plant lines were investigated for TET9-GFP fluorescence. Three transgenic plant lines did not show any GFP-AGO9 fluorescence in the egg cell (lines *EC1.1p:GFP-AGO9*-1, 11, 12). Six plants (50%) bearing the *EC1.1p:GFP-AGO9* construct exhibited a relative strong GFP signal, which was only detected in the egg cell nucleus (*EC1.1p:GFP-AGO9*-2, -3, -4, -6, -8, -10). Three transgenic lines exhibited only a very weak signal in the egg cell (*EC1.1p:GFP-AGO9*-5, -7, -9, Figure 30).

Transgenic *EC1.1p:GFP-AGO9* lines without a signal showed no visible phenotype. The same is true for plants with weak or even strong signal. No correlation between signal strength and seed abortion could be observed.

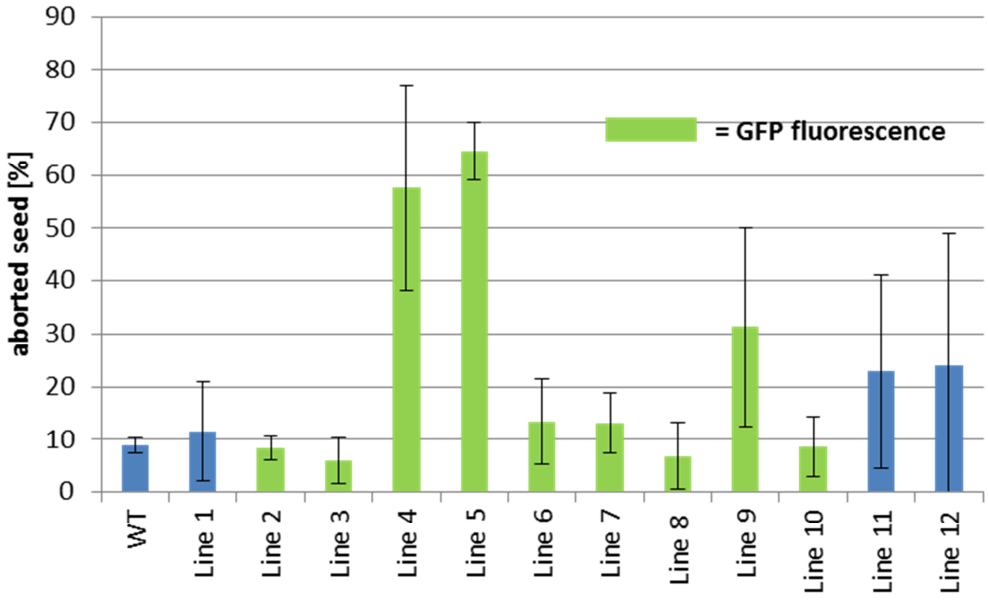


Figure 31: Ectopic expression of *GFP-AGO9* under control of egg cell specific *EC1.1* promoter. Plants of the first transgenic generation (T1) were examined. Undeveloped seeds (in %) in siliques of the wild type and individual transgenic lines were evaluated. Five siliques per plant were counted Seed abortion rates of transgenic plants with a GFP signal are shown in green. Seed abortion rates of the wild type and of transgenic plants showing no GFP fluorescence are shown in blue. No correlation between a stronger GFP-AGO9 expression in the egg cell (lines *EC1.1p:GFP-AGO9*-2, -3, -4, -6, -8, -10) and the phenotype of undeveloped seeds was observed.

Plant lines expressing GFP-AGO9 under the control of the *EC1.1* promoter show strong ectopic expression of AGO9 in the egg cell. Homozygous lines can be used to isolate egg cells. The isolated egg cells can be analyzed with respect to their methylation pattern. This will answer the question if ectopically expressed AGO9 is able to change the methylation pattern in egg cells. In summary, ectopic expression studies revealed that when the AGO9 coding sequence is expressed under the control of strong, egg cell-specific promoters, surprisingly few transgenic lines show GFP-AGO9 fluorescence (Table 4). By contrast, the expression of other reporter genes, such as NLS(3x) GFP or Tet9-GFP driven by the *EC1*, *AGO8* or *AGO9* promoters, revealed strong fluorescence signals from the reporter in almost all lines that have been generated (Table 4). These findings indicate that the stability of AGO9 protein is reduced in the egg cell.

Table 4: AGO9 protein stability is reduced in the egg cell. The table summarizes the results obtained from ectopic GFP-AGO9 expression in the egg cell using the egg cell specific *AGO8* and *EC1.1* promoters and the *AGO9* promoter, compared to the expression of controls such as Tet9-GFP in the egg cell (*EC1.1p:Tet9-GFP*) or expression of NLS(3x)GFP under control of the AGO8 or AGO9 promoters (*AGO8p:NLS(3x)GFP, AGO9p:NLS(3x)GFP).*

	Independent transgenic lines		
	GFP signal	weak signal	no GFP signal
AGO8p:NLS(3x)GFP	**15/15**	-	-
AGO9p:NLS(3x)GFP	9/12	2/12	1/12
AGO8p:GFP-AGO9 (CDS)	2/21	2/21	**17/21**
AGO9p:GFP-AGO9 (CDS)	2/12	1/12	**9/12**
EC1.1p:GFP-AGO9 (CDS)	6/12	3/12	**3/12**
EC1.1p:Tet9-GFP (CDS)	9/12	-	3/12

5.6 Whole mount-immunolocalization of AGO9

5.6.1 Control experiments

The polyclonal AGO9 antibody used for Western Blot (Chapter 5.2.2) was raised by using a synthetic peptide located at the N-terminus of AGO9 (Havecker et al., 2010). To be used for whole-mount immunolocalization in ovules the specificity of α-AGO9 was further evaluated using two approaches. First, a Western Blot analysis was performed using protein extracts from inflorescences from Arabidopsis wild type (Col-0) plants and the previously described T-DNA insertion line *ago9-1* (Havecker 2010) to investigate whether the α-AGO9 recognizes a truncated AGO9 protein or gives unspecific signals (Figure 32A).

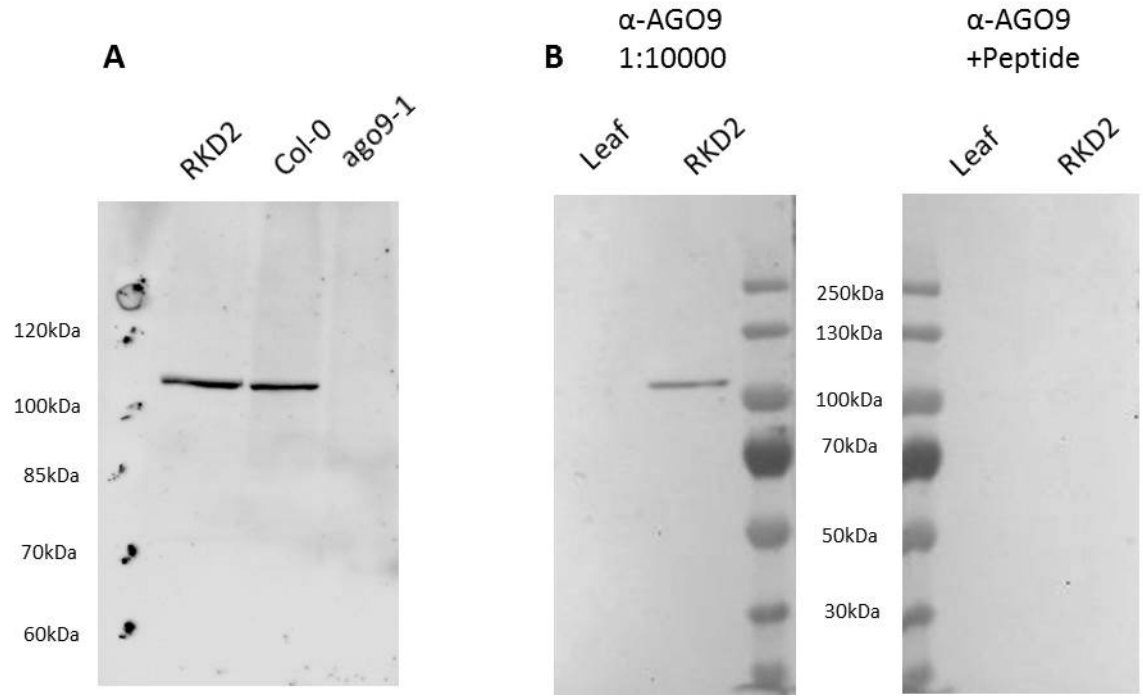


Figure 32: Evaluation of α-AGO9 antibody specificity. (A) The AGO9 peptide antibody was used in a Western Blot to detect AGO9 in protein extracts from RKD2 callus (positive control) and from Arabidopsis Col-0 mixed staged floral tissue. AGO9 was not detected in the homozygous *ago9-1* mutant **(B)** Peptide competition assay. The AGO9 peptide was used in a 5-fold excess over the antibody (by weight), which was diluted 1:10,000 to detect AGO9 in protein extracts from RKD2 callus. kDa, kilo Dalton.

Second, a peptide competition assay was conducted (Figure 32B), using the synthetic AGO9 peptide (DSDEPNGSGLPPPC-CONH2). For these assays protein extracts from the RKD2 callus and from Arabidopsis Col-0 inflorescences were used as positive controls.

In the homozygous T-DNA insertion line *ago9-1*, no AGO protein could be detected, like expected for a null mutant. In the wild type, one specific band of approximately 100 kDa, corresponding to the full length AGO9 protein, is detected (Figure 33A). In case a truncated protein of AGO9 would be present in the homozygous *ago9-1* T-DNA insertion line, a protein band of approximately 57.4 kDa would appear in the Western Blot of the *ago9-1* protein extract of mixed floral tissue, as the α-AGO9 antibody is directed against the N-terminus of AGO9 (Havecker et al. 2010). However, no truncated AGO9 protein is visible in the protein extract from *ago9-1* inflorescences (Figure 32A).

AGO9 antibody is preincubated with the AGO9 peptide (DSDEPNGSGLPPPC-CONH2).This experiment shows that the AGO9 peptide is specifically bound by the antibody and that the specific recognition sites of the antibody are blocked by bound AGO9 peptide. For this reason, the AGO9 protein on the Western Blot will not be recognized anymore by the antibody, and therefore no AGO9 immunosignal of approximately 100 kDa is visible (Figure 32B, α-AGO9 + peptide).

The α-AGO9 antibody was subsequently used to localize the AGO9 protein within the female gametophyte during different developmental stages before and after fertilization and in developing embryos. During the experimental procedure negative controls, which were only incubated with blocking solution instead of primary α-AGO9 antibody, followed by the incubation with the CY3-conjugated anti-rabbit secondary antibody, were carried out (Figure 33 E to H).

To verify that IgG antibodies are able to penetrate the mature female gametophyte, ovules expressing an EC1.1-GFP fusion protein under the control of the *EC1.1* promoter were treated with an anti-GFP primary antibody and a CY3-conjugated anti-rabbit secondary antibody. With this approach the EC1.1-GFP fusion protein could be detected in the egg cell of the mature ovule (Figure 33). Besides the red fluorescence of CY3, the green fluorescence of EC1-GFP could still be detected in the egg cell.

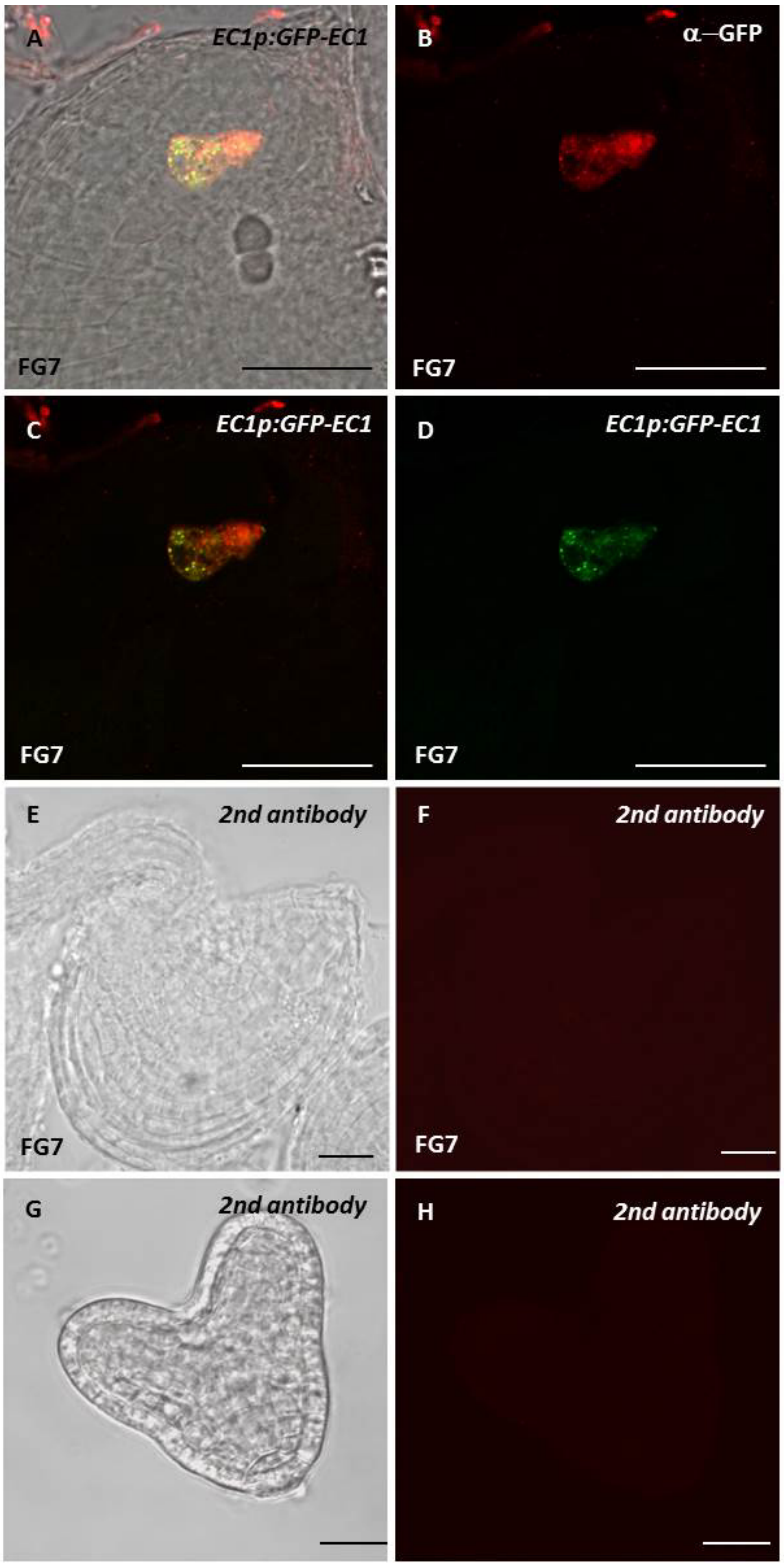


Figure 33: Whole mount-immunolocalization control experiments
(A) to **(D)** GFP-EC1 fusion protein under the control of the EC1.1 promoter is expressed in the egg cell and can be recognized by an antiGFP antibody. **(E)** and **(F)** Negative control. Ovules of the FG7 stage were incubated with blocking solution instead of primary antibody. The experimental procedure was continued as described in chapter 2.8.2. **(G)** and **(H)** Negative control. Heart stage embryos were treated as described above. Scale bars are 20 µM.

Colocalization of GFP and anti-GFP derived CY3 signal does occur in the whole egg cell. Nevertheless, at the micropylar pole of the egg cell the CY3 signal appears to be stronger than the GFP fluorescence. This could be the result of structural changes of the fusion protein caused by denaturing reagents during the experimental procedure, which influence the fluorescent properties of the GFP. This result showed that under the used conditions IgG antibodies are able to detect their epitopes within the female gametophyte.

5.6.2 Immunolocalization of AGO9

In order to investigate the protein localization of AGO9 in the developing ovules and to compare protein abundance with observed promoter-reporter activities, a whole-mount immunolocalization procedure was used as described in chapter 4.8.2. FG stages were defined according to Christensen et al. (1998). At female gametophyte stage 0 (FG0) AGO9 protein can be detected in nuclei of all cells of the ovule, except the MMC (Figure 35A and B). Strong signals are present in nuclei of the inner integuments (ii). Also in stage FG1 AGO9 is observed in the nucellus tissue (nu), the outer- and inner integuments and the funiculus, but not in the functional megaspore (fm) (Figure 35C and D). Later, at female gametophyte stage 2 (FG2) strong signals of AGO9 can be detected in the outer integuments (oi) and in the funiculus (fu) (Figure 35E and F). No AGO9 proteins can be detected in the two female gametes of the mature ovule, at FG7 (Figure 34K and L).

After fertilization, the AGO9 protein is first detected in the octant stage of the embryo. Also here, most of the protein is localized in the nucleus. During early embryogenesis, AGO9 is expressed in the whole embryo and can be found in the nuclei of 8 and 16 celled embryos (Figure 35A to D). In heart stage embryos a stronger expression of AGO9 can be observed in the nuclei of the protoderm and the shoot apical meristem (Figure 35E and F). In older embryonic stages, presence of AGO9 protein diminishes.

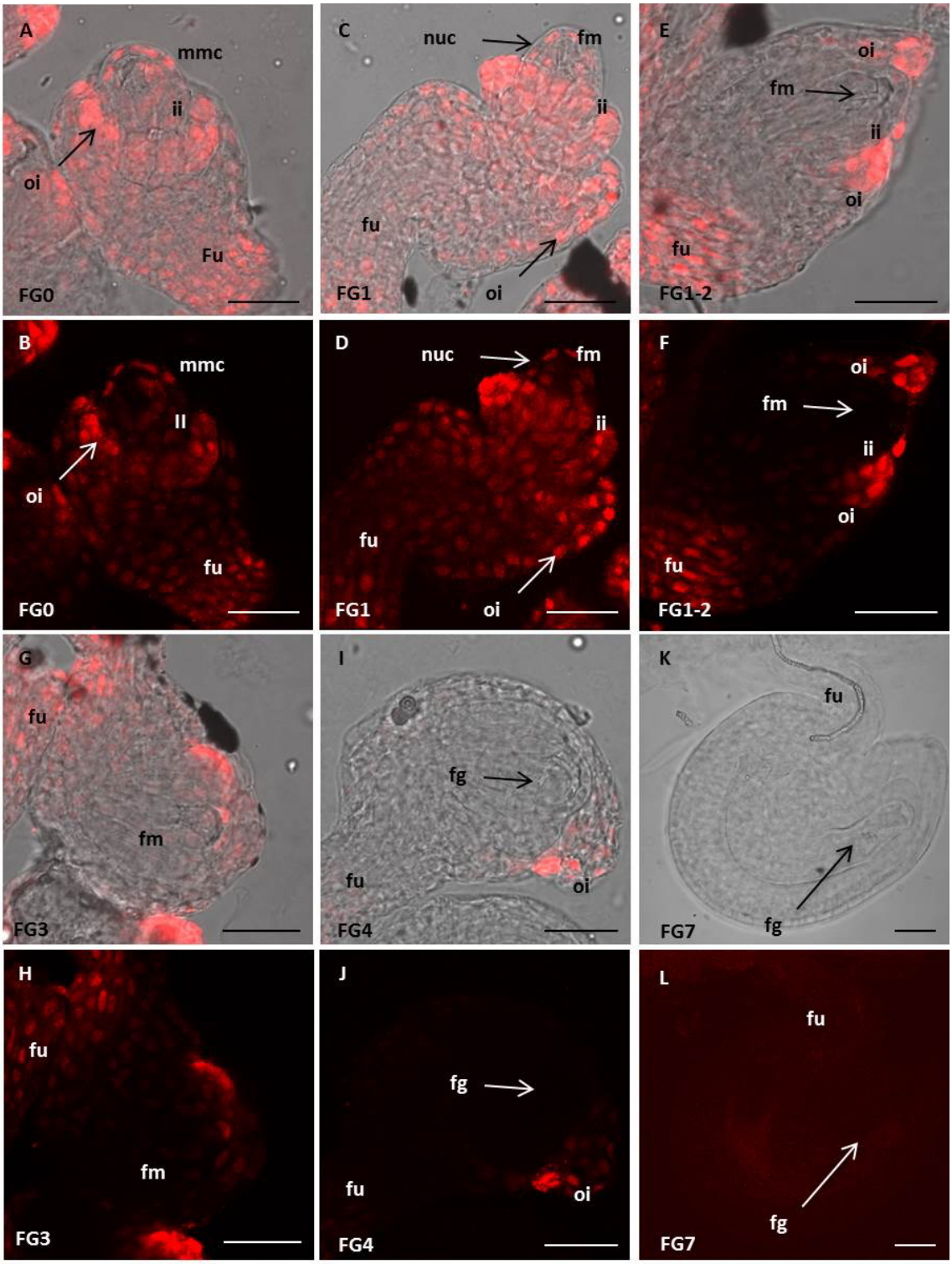


Figure 34: Whole mount-immunolocalization of AGO9 during female gametophyte development. **(A)** and **(B)** At female gametophyte stage 0 (FG0), AGO9 proteins can be detected in nuclei of all cells of the ovule, except the megasporemother cell (MMC). Strong signals are present in nuclei of the inner integuments (II). **(C)** and **(D)** At female gametophyte stage 1 (FG1) strong signals of AGO9 can be detected in in nuclei of all cells of the ovule, butnot in the functional megaspore (FM) **(E)** and **(F)** At FG2 of female gametophyte development AGO9 is detected in the inner (II) and outer integument (OI) and the funiculs (FU). **(G)** and **(H)** At FG3 AGO9 is also localized in the inner (II) and outer integument (OI). **(I)** and **(J)** At FG4 AGO9 is only detected in the outer integuments but not in the female gametophyte (FG). **(K)** and **(L)** In the mature ovule, at FG7 AGO9 cannot be detected. Scale bars are 20 µM.

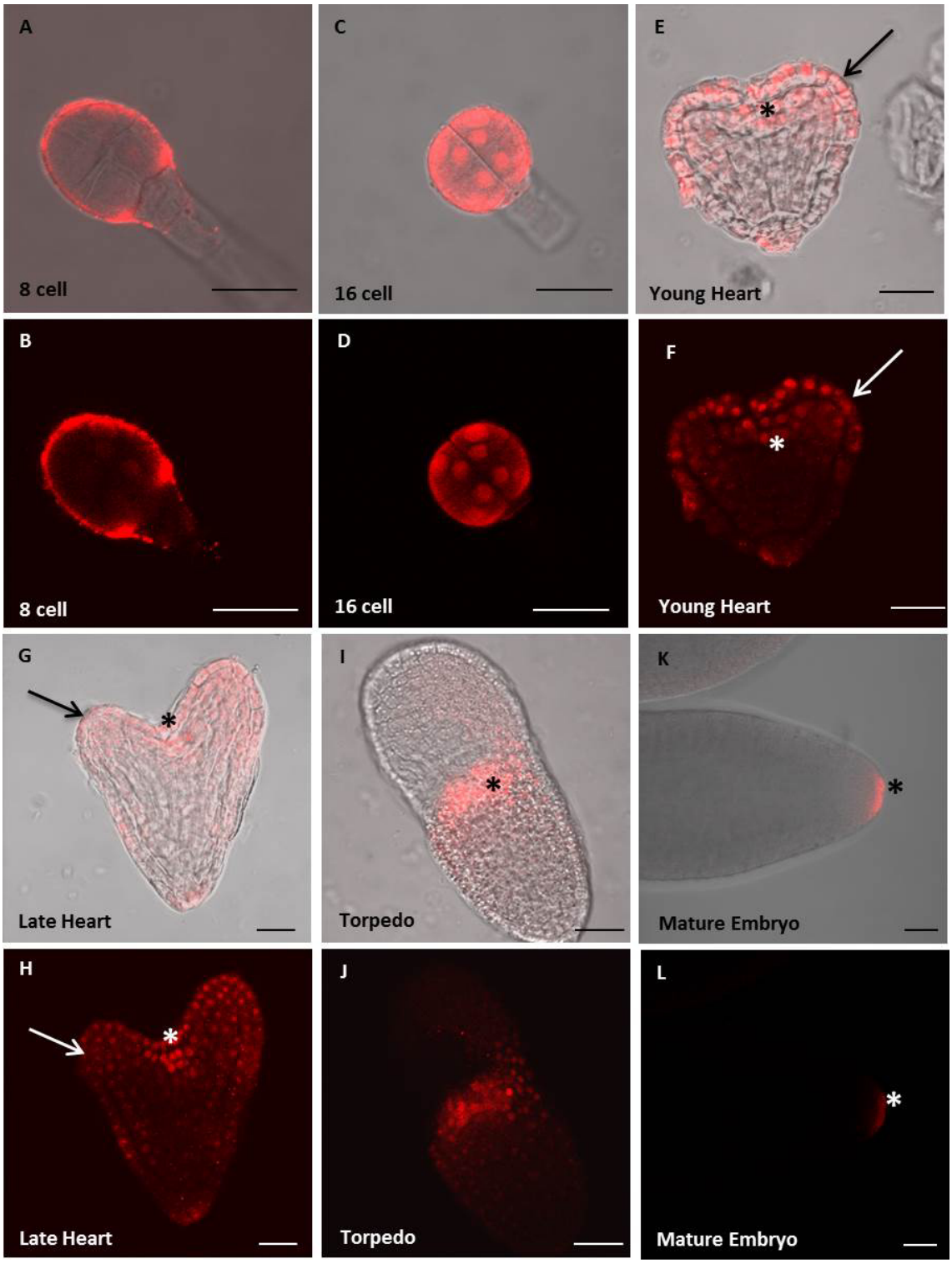


Figure 35: Whole mount-immunolocalization of AGO9 during embryo development.
(A) and **(B)** In the 8 cell embryo, AGO9 is detected in the nuclei of the whole embryo **(C)** and **(D)** In the 16 cell embryo, AGO9 also is detected in the nuclei of the whole embryo **(E)** and **(F)** Young heart embryo showing stronger signal in the protoderm (arrow) and the shoot apical meristem (asterisk). **(G)** and **(H)** Late heart embryo showing stronger signal in the protoderm (arrow) and the shoot apical meristem (asterisk). **(I)** and **(J)** Presence of AGO9 in torpedo stage embryo with strongest immunosignal in the shoot apical meristem (SAM) (asterisk). **(K)** and **(L)** Detection of AGO9 in the root apical meristem region (asterisk). Scale bars are 20 µM.

In the late heart stage, the signal in central region of the embryo gets weaker (Figure 35 G and H). In the torpedo stage, the strongest signal is observed in the region of the shoot apical meristem (Figure 35 I and J). In the mature embryo, a signal can be seen in the region of the root apical meristem (Figure 35 K and L). Also in embryogenesis, most signals, except for those in the root tip, are concentrated in the nuclei of the respective tissues

5.7 Transcript variants of *AGO8* and *AGO9*

Transcripts of *AGO8* and *AGO9* are subjected to alternative splicing. In case of *AGO9*, a transcript with retained intron 6 was found to be present in cDNA of emasculated pistils (Chapter 1.3) and alternative splicing may be the reason why AGO9 proteins are not detectable in the mature ovule (Chapter 5.6.2). Moreover, all *AGO8* transcripts described so far (Takeda et al. 2008, Wartlsteiner 2010) differ from the coding sequence, which is annotated in the TAIR database.

5.7.1 Expression pattern of *AGO8* transcript variants

Based on these results, the question arose if there is a correctly spliced transcript variant of *AGO8* present, for example during egg cell maturation, or in the RKD2-induced egg cell-like cell line. A PCR-based approach was conducted to address this question. Primers in the 3´ region of *AGO8* were created, which would result in different PCR products, according to the *AGO8* transcript variant. The forward primer is located within exon 11, the reverse primer in exon 16 of the *AGO8* mRNA (Figure 36A). Presence of the *AGO8-4* variant would result in an amplicon of 752 bp, the *AGO8-8* splicing variant yields a 672 bp amplicon and the correctly spliced transcript would create a PCR band of 601 bp in the agarose gel (Figure 36A). In transcript variant *AGO8-4* an alternative 5´ splice site results in generation of a premature stop codon (PTC) at position 1038 from the ATG, in *AGO8-8* two alternative 5´ splice sites create a PTC at position 1704 (Chapter 1.3).

PCR results showed that whenever *AGO8* transcripts were detected, two amplicons of 752 bp and 672 bp respectively appeared, corresponding to the two previously discovered splicing variants *AGO8-4* and *AGO8-8* (Wartlsteiner 2010). The *AGO8-8* fragment shows a stronger band than the *AGO8-4* fragment.

The 601 bp band, corresponding to the correctly spliced region of the *AGO8* transcript as it is annotated at TAIR (Accession AT5G21030), could not be detected in any of the analyzed tissues (Figure 36B). Importantly, no correct spliced amplicon of 601 bp could be amplified from cDNA derived from the RKD2-induced egg cell-like callus, suggesting that *AGO8* exon 12 and intron 13 do never yield functional AGO8 CDS and thus AGO8 protein.

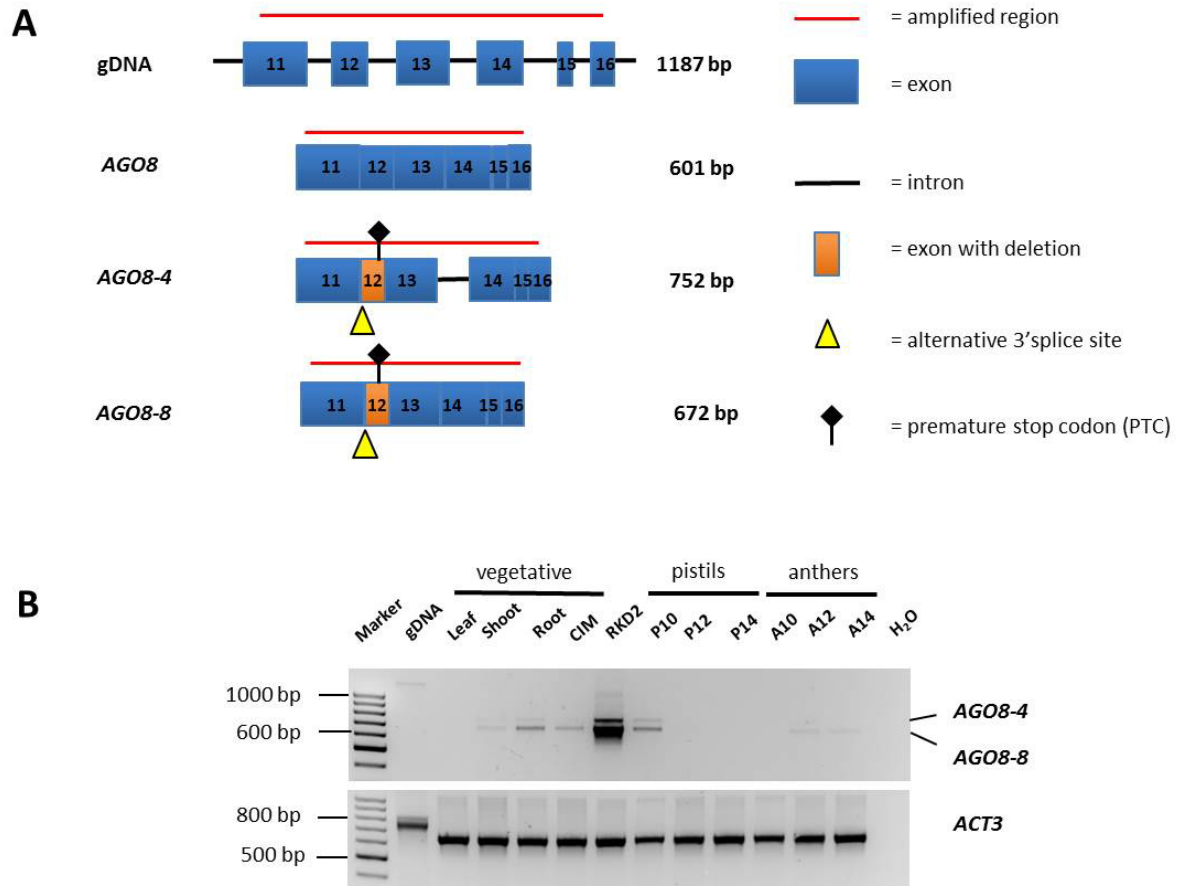


Figure 36: RT-PCR Analysis of AGO8 transcript variants.
(A) Schematic representation of PCR strategy. The primer combination AGO8seq5/AGO8GDrev (Wartlsteiner 2010) was used for ampflication of DNA fragments with Taq polymerase. The amplified region is marked with a red bar. Following bands and sizes could be expected: Genomic DNA (gDNA) *AGO8* (1187 bp), *AGO8 cds*(601 bp), *AGO8-4* (752 bp), *AGO8-8* (672 bp). A part of Exon 12 (marked in orange) is spliced out due to an alternative 5`prime splice site, resulting in a premature stop codon (PTC). **(B)** RT-PCR results. *AGO8* transcript variants were analyzed in different vegetative and reproductive tissues, as well as in CIM and RKD2 callus tissues. *ACT3*, *ACTIN3*; P10, pistil stage 10, P12, pistil stage 12; P14, pistil stage 14; A10, anther stage 10; A12, anther stage 12; A14 anther stage 14. Anther and pistil stages according to Smyth et al. (1990).

5.7.2 *AGO8* transcript variants in RKD2-induced callus

The presence of *AGO8* splicing variants in the RKD2-induced cell line was analyzed by amplifying the full coding sequence (2553 bp) of *AGO8* using the primer pair AGO 8 for/rev and cDNA generated from this cell line. For amplification, a proofreading DNA-polymerase was used (Chapter 4.4.3).

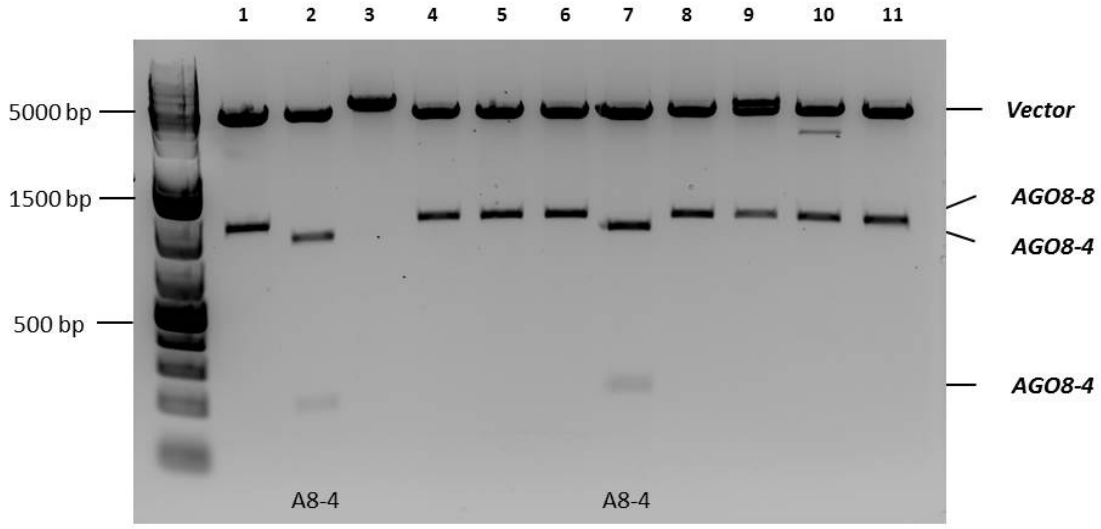


Figure 37: Detection of splicing variants in full-length *AGO8* coding sequences amplified from cDNA, generated from the RKD2-induced cell line. The cDNA was amplified with primer pair AGO8 for/rev and subsequently cloned into pENTR/dTOPO. Isolated plasmids from 11 different clones were subjected to restriction digest with EcoRI and EcoRV and separated by agarose gel electrophoresis. The clones in lane 2 and 7 show a restriction pattern ressembling *AGO8-4* consisting of a 4327 bp vector band and two 741 bp and 192 bp band from the *AGO8-4* transcript. In the other lanes a 4228 bp vector band and a 931 bp from the AGO8-8 transcript.

The amplified *AGO8* cDNA population was subcloned into the pENTR/dTOPO vector and isolated plasmids were subjected to restriction analysis using the restriction enzymes *EcoRI* and *EcoRV*.

The cloned fragments showed restriction patterns expected for the known variants *AGO8-4* and *AGO8*-8 (Introduction). However, sequencing of cloned fragments resembling *AGO8-4* revealed the presence of a novel *AGO8* transcript variant. This variant is similar to variant *AGO8-4*, but uses an alternative 3' splice junction (acceptor site) in exon 4, changing the 5' boundary of exon 4 in that 45 base pairs of its 5´ part is removed.

This generates the transcript variant *AGO8-30,* containing an open reading frame comprising a premature stop codon at cDNA position 1.641 relative to the start codon (Figure 38).

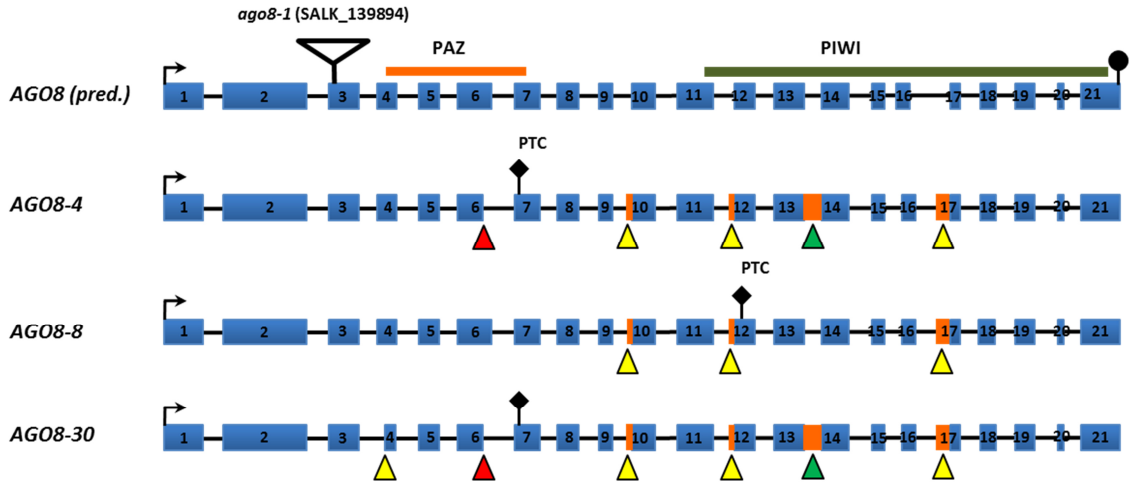


Figure 38: *AGO8* transcript variants identified in cDNA from the RKD2-induced cell line.
Schematic summary of *AGO8* transcript variants in the RKD2-induced cell line. Blue bars represent exons as annotated by TAIR. Green triangles indicate intron retention. Orange bars indicate intronic sequences contained in the transcript. Positions of the PIWI, Argonaute, Zwille (PAZ) and P-element Induced Wimpy testis (PIWI) domains are indicated by orange and olive bars.

5.7.3 Expression pattern of *AGO9* transcript variants

The full length transcript of *AGO9* was amplified using cDNA from emasculated pistils (Wartlsteiner 2010). Sequencing of cloned full length *AGO9* transcripts revealed the presence of a transcript variant of *AGO9* with a retained intron 6 (*AGO9-8*; Figure 14). This insertion leads to a frame shift in the coding sequence, with the consequence of a PTC at position 912 from the start codon.

Former expression analyzes of *AGO9* showed that the transcript is present in generative tissues such as mature and immature pistils and anthers, as well as siliques of 5, 9 and 13 mm length (Wartlsteiner 2010). However, the set of primer used for this experiment could not distinguish between the correctly spliced exon 5-exon 6 junction of *AGO9* and the splice variant with retained intron 6. In order to analyze the abundance of the two transcript variants in different Arabidopsis tissues, a primer pair flanking intron 6 was used (AGO9seq2/AGO91154rev, chapter 9.4), so that transcript variant *AGO9-8* would yield a 694 bp amplicon and the correctly spliced transcript a 620 bp amplicon after PCR.

The difference between the two amplification products of 65 bp corresponds to the length of the intron of 65 bp. Using the primer pair *AGO9*seq2/*AGO9*1154rev, *AGO9* transcript abundance in generative tissues and its absence from vegetative tissues was confirmed (Figure 39). Furthermore, *AGO9* transcript was found in the RKD2-induced callus but not in root-derived callus (CIM). Wherever *AGO9* was expressed, both transcript variants were detected.

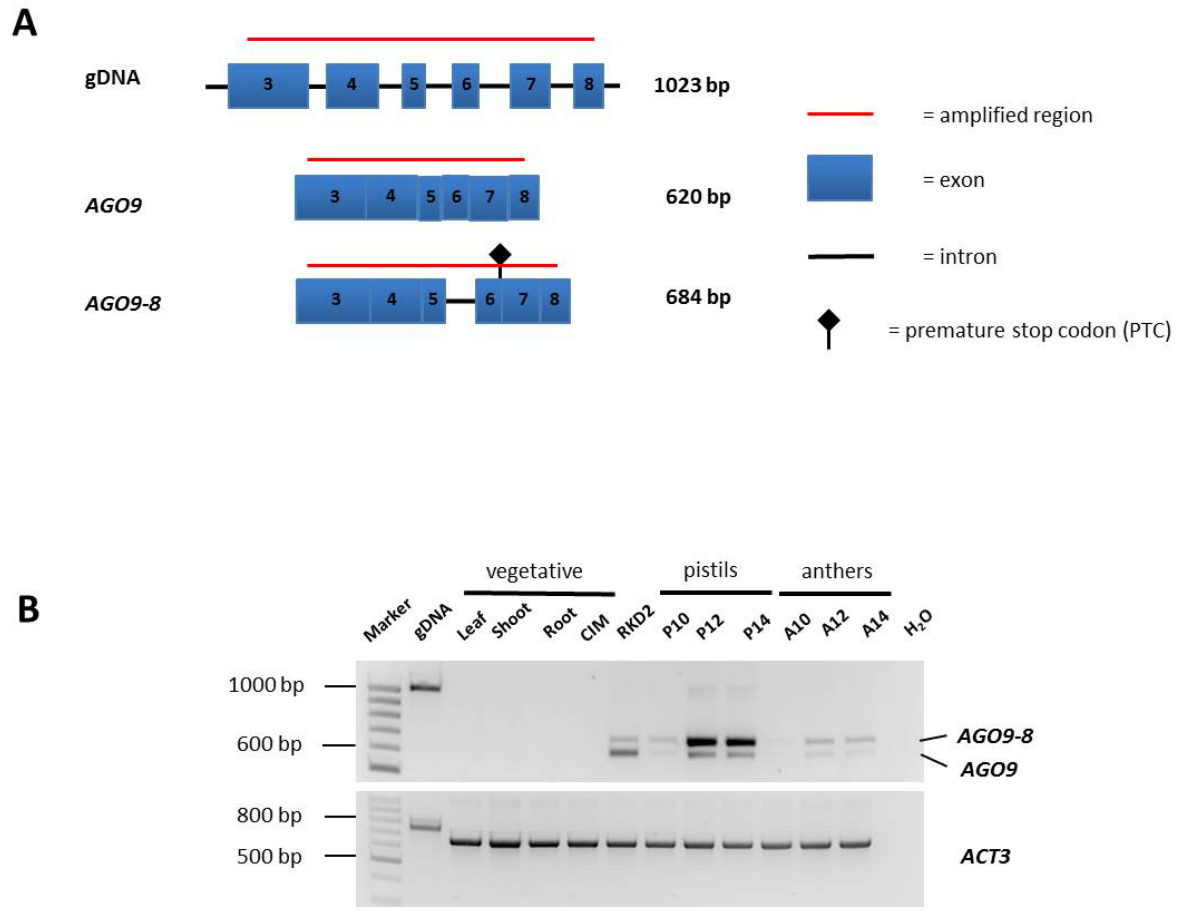


Figure 39: AGO9 variants in different tissues (A) Schematic representation of PCR strategy. The primer combination AGO9seq5/AGO9 1154rev (Wartlsteiner 2010) was used for ampflication of DNA fragments with Taq polymerase. The amplified region is marked with a red bar. Following bands and sizes could be expected: Genomic DNA (gDNA) *AGO9* (1023 bp), *AGO9 cds*(620 bp), *AGO9-8* (684 bp). **(B)** *AGO*9 transcript variants were analyzed in different vegetative and reproductive tissues, as well as in CIM and RKD2 callus tissues. *ACT3*, *ACTIN3*; P10, pistil stage 10, P12, pistil stage 12; P14, pistil stage 14; A10, anther stage 10; A12, anther stage 12; A14 anther stage 14. Anther and pistil stages according to Smyth et al. 1990.

The ratio between the two amplicons was variable; in pistils and anthers of flower stage 12 and 14 the *AGO 9-8* amplicon with retained intron 6 was dominant, whereas in the RKD2-induced callus the regularly spliced *AGO9* variant was more abundant (Figure 39B).

5.7.4 Systematic approach to identify other *AGO9* transcript variants

In order to address the question if there are more *AGO9* splicing variants, a systematic approach to identify putative new variants including rare ones was used. Since the cloning of *AGO9* full-length transcripts proved to be difficult, five overlapping PCR-fragments covering the whole *AGO9* coding sequence of 2.681 bp were amplified using different primer combinations (Figure 40A). Anthers and pistils of flower stages 10 and 12 were used to generate cDNA populations.

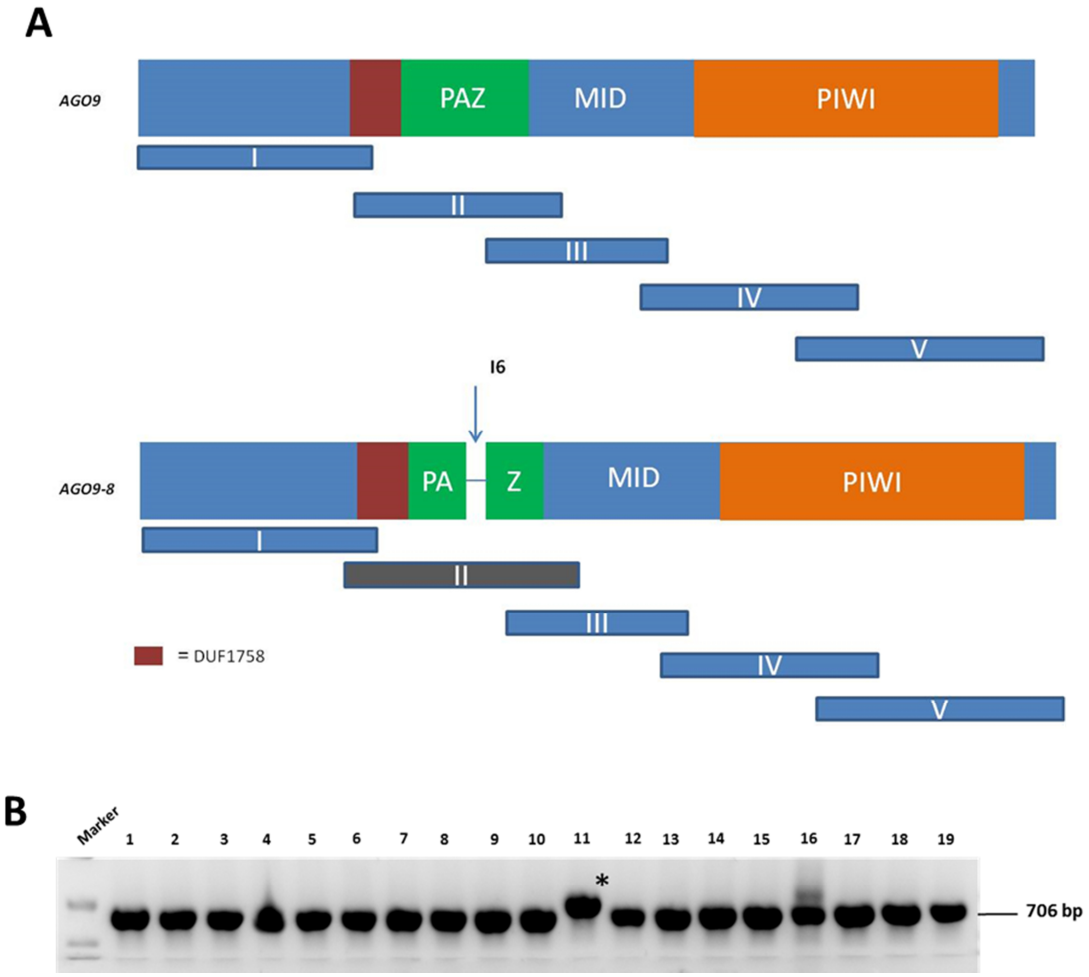


Figure 40: Overlapping *AGO9* cDNA fragments to identify putative novel *AGO9* transcript variants.
(A) Schematic representation of AGO9 gene structure and the length and position of PCR fragments. PCR fragments are represented by blue bars. The following primer combinations were used for amplification: fragment I: AGO9 for/AGO9 rev 700; fragment II: AGO9seq2/ AGO9 rev 1154, fragment III: AGO9seq3/ AGO9 rev 1595, fragment IV: AGO9seq3/ AGO9 rev 2158; fragment V: AGO9seq5/ AGO9 cds rev. Forward primer sequences were taken from Wartsteiner (2010), reverse primers were constructed during this work. Primer sequences and exact length of amplified fragments can be found in chapter 9.2 (table 18). **(B)** Colony PCR with bacteria colonies transformed with cloned *AGO9* IV fragments. The primer pair M13for/M13rev was used to detect cloned fragments in vector pCR2.1 via PCR. The fragment in lane 11 (asterisk) differs from the predicted size of 706 bp.

The size of the PCR fragments ranged from 543 to 739 base pairs. PCR fragments were cloned into the pCR2.1 vector and the inserts of at least ten individual positive clones from anthers and pistils were sequenced. The retained intron 6 of splicing variant *AGO9-8* is located in fragment II (Figure 39A). This splicing variant could be confirmed in 3 out of 19 sequenced clones, which were all derived from cDNA generated from flower stage 10 pistils (Table 7).

Fragment I was difficult to clone and only one single clone could be analyzed by sequencing, which was completely identical to this annotated sequence. In *AGO9* cDNA fragments III and V no sequence variations have been observed in 15 and 20 sequenced clones, respectively. Fragment IV, covering the exons 11 to 18, could be easily amplified from cDNA of anthers (5 clones) and pistils (11 clones).Although the amplicon of fragment IV showed the calculated size of 706 bp in the agarose gel, the following cloning and colony PCR revealed a few fragments (3 of 72 analyzed colonies) with a slightly larger size (Figure 40B). Sequencing of 16 cloned inserts including the three larger inserts revealed that the larger insert size is caused by an insertion in the amplified fragment IV which is identical to intron 13 of the genomic sequence. Therefore, intron retention between exon 13 and exon 14 of *AGO9* represents a second alternative splicing event in *AGO9* transcripts. The novel splice variant, thereafter named *AGO9-68*, is a rare one as it cannot be detected by RT-PCR, unlike variant *AGO9-8* which can easily detected in pistils, anthers and RKD2 induced callus. In spite of the high abundance of variant of *AGO9-8* variant only 3 of 19 sequenced clones of fragment II contained the retained intron 6, which might be caused by cloning issues. *AGO9* variants have also been discovered in reproductive tissues and screening of different tissues already confirmed the presence of variant *AGO9-8* in RKD2-induced tissues. Variant *AGO9-68* has also been found in the RKD2-induced callus. Altogether, the PCR-based approach to find new transcript variants resulted in the discovery of one novel and rare splice variant (Figure 41)

However it was not possible to clone all fragments that have been amplified from the different tissues, so it cannot be excluded that more *AGO9* splicing variants exist in other regions of the gene. Analysis of RNA seq data generated from different stages of developing ovules will be necessary to obtain an impression on the occurrence and frequency of all *AGO9* splicing variants.

Table 5: Identification of *AGO9* transcript variants using a systematic RT-PCR approach.
Summarized are the results of cloning and sequencing cDNA fragments I to V from anthers and pistils of flower stages 10 and 12 (flower stages according to Smyth et al. 1990). Two PCR fragments (II, IV) revealed alternative splicing. Primer sequences of AGO9 cds for/rev and seq1 to seq 5 from Wartlsteiner (2010). Numbers in parentheses marked with an asterisk indicate transcript variants with retained introns.

AGO9 fragment	Primer combination	Anthers	Pistils
I	AGO9cds for/AGO9 rev 700	1	-
II	AGO9seq2/AGO9 rev 1154	-	19 (3*)
III	AGO9seq3/AGO9 rev 1595	5	10
IV	AGO9seq4/AGO9 rev 2158	5	11 (2*)
V	AGO9seq5/AGO9cds rev	10	10

Nevertheless, it seems unlikely that the two splicing variants of *AGO9* described in this work yield truncated AGO9 proteins. Either a 33.6 kDa or a 73.1 kDa protein band would appear in the Western Blot using the α-AGO9 peptide antibody, representing the truncated translation products derived from the splice variants *AGO9-8* and *AGO-68*. The α-AGO9 peptide antibody is expected to detect both the truncated proteins as it is directed against the N-terminus of AGO9 (Havecker et al. 2010). However, truncated AGO9 proteins of 33.6 kDa or a 73.1 kDa could neither be detected in mixed inflorescences nor in the RKD2-induced cell line (Figure 32).

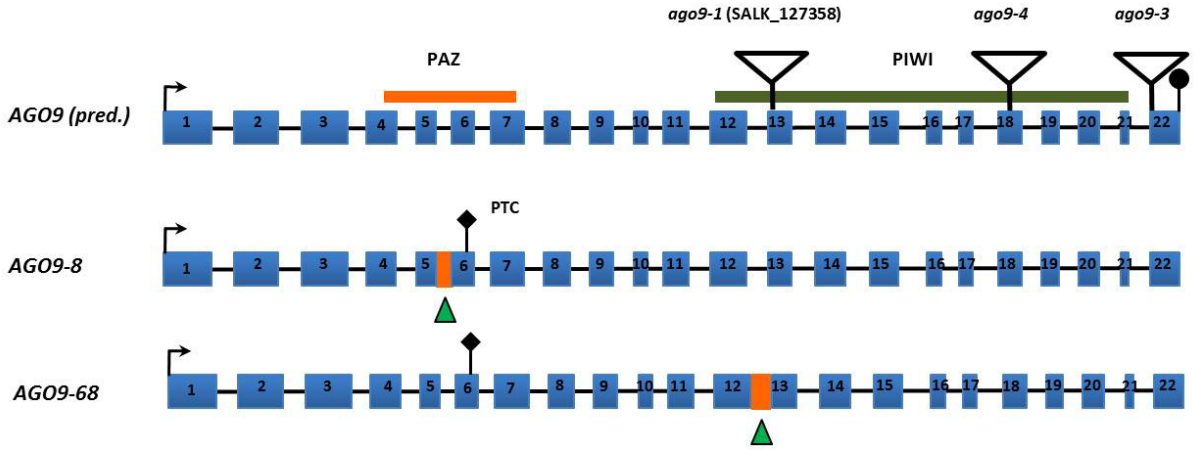


Figure 41: Summary of *AGO9* splice variants, detected in cDNA from Arabidopsis pistils and RKD2-induced callus
Schematic summary of *AGO9* transcript variants. Blue bars represent exons as annotated in TAIR. Green triangles indicate intron retention. Orange bars indicate intronic sequences contained in the transcript. Positions of the PIWI, Argonaute, Zwille (PAZ) and P-element Induced Wimpy testis (PIWI) domains are indicated by orange and olive bars.

5.7.5 *AGO9* minigene approach

The coding sequence of the *AGO9-8* splicing variant contains, due to intron 6 retention, a premature stop codon (PTC) at position 916 of the transcript (Figure 41). Expression of this variant would lead to a truncated protein of 304 amino acids with the predicted molecular weight of about 33.6 kDa and a calculated pI of 9.14. This protein would consist only of the N-terminal domain, the domain of unknown function (DUF) 1758 and a part of the PAZ domain (Figure 41). Since the PAZ domain is responsible for the slicer activity of Argonautes, the truncated protein derived from *AGO9-8* would not be able to degrade mRNAs. Nevertheless, in Drosophila, it has been shown that Argonaute proteins can exert biological functions independently of slicer activity (Darricarrère *et al.*, 2013). The question was, if transcript variants of *AGO9* play a role in defined stages of female gametogenesis and early embryonic development in Arabidopsis, like for example shortly after formation of the zygote (maternal-to-zygotic switch). Moreover, the question arose if alternative splicing events of *AGO9* depend on the developmental stage of the ovule/female gametophyte or the differentiation of the egg cell and central cell. To investigate alternative splicing of *AGO9* intron 6 during female gametophyte and gamete development, a minigen approach was chosen. The minigene approach has already been successfully used in human cells to perform a fluorescence-based quantitative analysis of alternative splicing of the human p53-inducible gene3 (*PIG3)* gene (Gurskaya 2012). Minigene reporter analyses were also part in a strategy to identify regulators of post-synaptic density gene 4 (*DLG4)* alternative splicing in mammals (Zheng 2013).

Constructs for this kind of experiments contain the exons flanking the differential spliced region of the gene of interest, fused to two different fluorescent proteins at the respective C- and N-termini. Exons must be in frame with the coding sequences of the two different fluorescent proteins (e.g., mCherry and GFP), so that translation of the correctly spliced transcript results in fluorescence of both mCherry and GFP. Intron retention would yield only translation of the N-terminal fused fluorescent protein, due to the presence of a premature termination codon (PTC) 5´ to the second fluorescent protein (Figure 42).

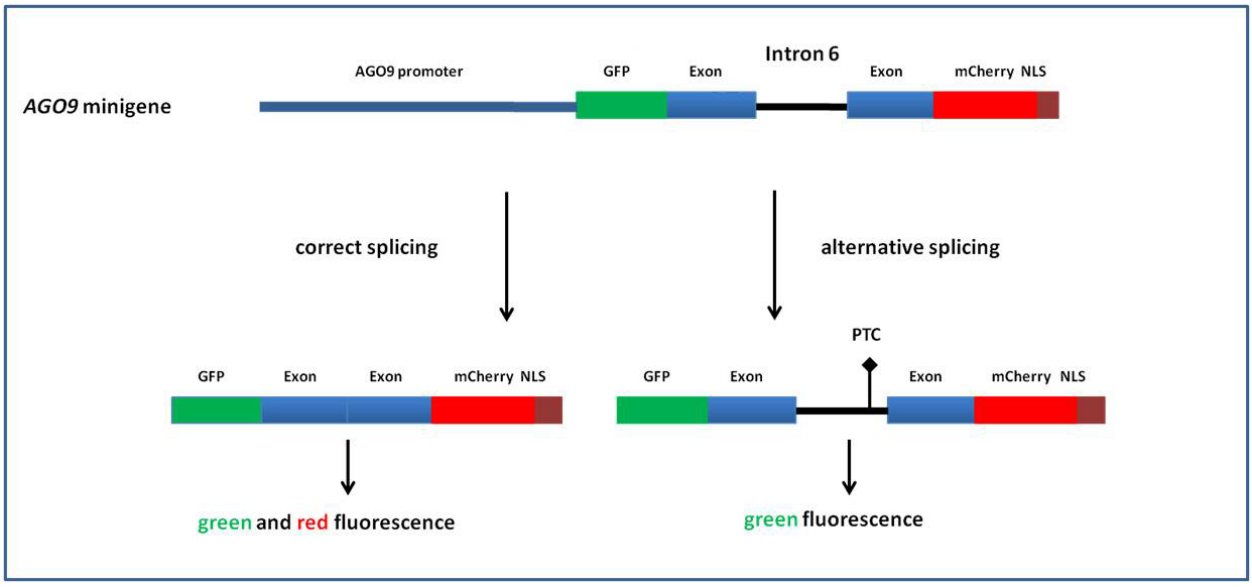


Figure 42: Monitoring the presence of splice variants using the minigene approach.
A genomic region including two exons that flank the alternatively spliced intron is translationally fused to the coding sequence of one fluorescent protein (in this case: *GFP*) at the 5´ end of the leading exon and a second fluorescent protein (in this case: *mCherry*) at the 3´ end of the following exon. In case of *AGO9*, alternative splicing of intron 6 results in two different mRNA species. The translational fusion of GFP at the 5´ end of exon 5 and nuclear localized *mCherry* at the 3´ end of exon 6 results in green and red fluorescence in case of correct splicing. A retained intron in the minigene construct leads to a premature termination codon (PTC) resulting in GFP derived green fluorescence only. GFP, green fluorescent protein; nls, nuclear localization signal; PTC, premature stop codon.

To investigate *AGO9* differential splicing two different *AGO9* intron 6 minigene-constructs and two *AGO9* minigene control constructs (covering *AGO9* intron 11) were cloned (Figure 43). The first *AGO9* intron 6 minigene construct included the genomic region of exon 6 to exon 7, while the second *AGO9* intron 6 minigene construct included the genomic region of exon 5 to exon 8. The genomic *AGO9* region that covers intron 11 was chosen as control because intron 11 has a similar size, compared with intron 6, and appears to be correctly spliced. This intron is located in the two overlapping fragments III and IV of AGO9 (Figure 40) and in 15 sequenced clones of fragment III and 16 sequenced clones of fragment IV no retained intron 11 was found. Also for the controls, one *AGO9* intron 11 minigene construct included the genomic region of exon 10 to exon 11, while the second control included the genomic region of exon 10 to exon 12 (Figure 43).

The longer minigene constructs were cloned in addition, as they would leave more space for the splicing machinery to interact with the minigene mRNA, because with about 720 bp they are approximately 470 bp longer than the single intron minigene constructs.

The exact length of the four cloned minigene fragments can be found in chapter 4.4.10. All minigenes were cloned under control of the *AGO9* promoter.

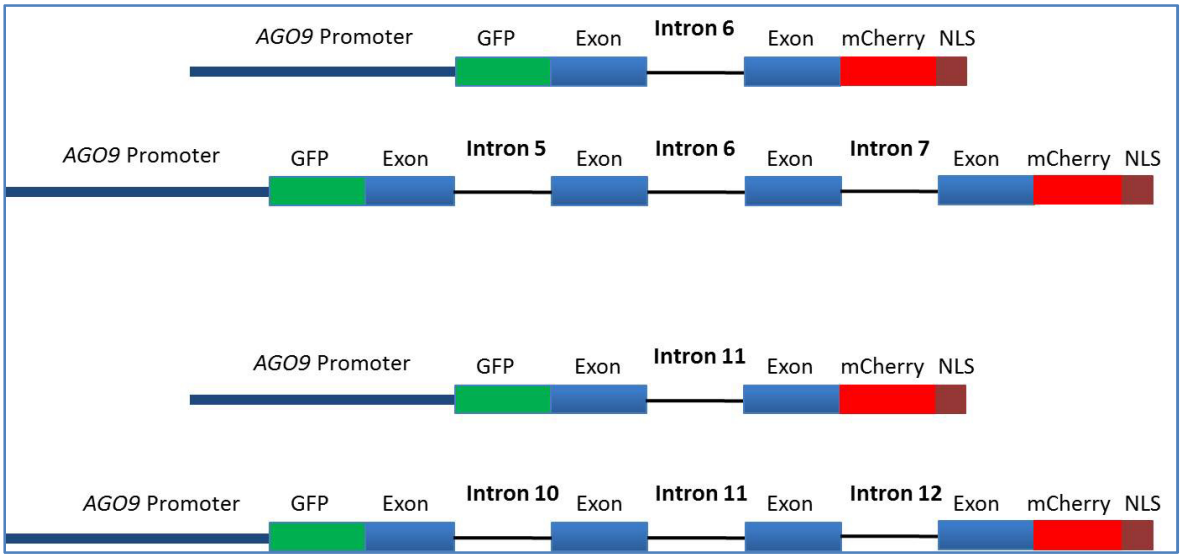


Figure 43: *AGO9* minigene constructs and control constructs to monitor differential splicing of *AGO9* during female gametophyte development.
Schematic representation of *AGO9* minigenes (1 and 2) and *AGO9* control minigenes (3 and 4). AGO9 minigenes contain intron 6 which is retained in part of AGO9 transcripts (chapter 5.7.1). AGO9 control minigenes control contain constitutively spliced intron 11 (chapter 5.7.4). NLS, nuclear localization signal; PTC, premature stop codon.

The minigene constructs were used for stable Arabidopsis transformation. In case of the *AGO9* intron 6 minigene constructs green fluorescence was expected in case of intron retention, whereas correct splicing of intron 6 and intron 11 would yield both red and green fluorescence in those cells expressing *AGO9*. For the *AGO9* minigene constructs 1 and 2 (Figure 43) 4 and 8 independent lines were generated respectively. 2 and 4 independent transgenic lines containing the control minigenes 3 and 4 were generated. Transgenic plants of the first transgenic generation were analyzed by confocal laser scanning microscopy.

However, in none of the analyzed transgenic lines a significant signal from GFP or mCherry could be detected. The reduced protein stability of GFP-AGO9 constructs expressed under the control of the AGO9 promoter has already been mentioned in chapter 5.5.3. This could be the reason that no fluorescent signal was found. More independent transgenic lines will have to be generated and analyzed.

5.8 Deep sequencing of mRNA and small RNAs in RKD2-induced cells with egg cell-like fate

5.8.1 Transcriptome profiling of RKD2-induced callus with egg cell-like fate

The transcription profile of the RKD2-induced callus was analyzed using RNA sequencing. A root-derived callus, which was induced by callus-induction medium (CIM), was used as a control (Chapter 5.1). Total RNA containing small RNAs was isolated from three biological replicates of each tissue.

5.8.2 RNA quality control and library generation

Preliminary quality control of the total RNA samples was performed by non-denaturing agarose gel electrophoresis. Two microliters of the RNA samples were loaded per lane. Ribosomal RNA bands of 28S, 18S and 5S were prominent on the gel and not degraded (Figure 44A). Concentration of the samples was measured with a nanodrop spectrophotometer. Before library preparation another quality control was carried out by the Kompetenzzentrum für Fluoreszente Bioanalytik (KFB) using the Agilent 2100 Bioanalyzer with the RNA 6000 Nano LabChip reagent set (Figure 44B and C). The Bioanalyzer electropherogram of the total RNA samples (Figure 44C) showed clear and distinct bands for ribosomal RNAs and small RNAs indicating a high integrity of the samples. The low noise between the sRNA peaks (Figure 44C) indicates a low degree of RNA degradation.

Poly(A)-enriched libraries were prepared at the KFB like as described (chapter 4.5.6). Quality of raw reads was assessed by FASTQC, with a mean quality score of 35.7 per sample. This value, which is also often called Phred quality score, stands for the probability of a base to be called incorrectly.

$$Quality\ score\ Q = -10 * log10\ (probability\ that\ the\ base\ is\ wrong)$$

A quality score of 35.7 corresponds to an error probability of 0.0269 %, indicating a high quality of the library.

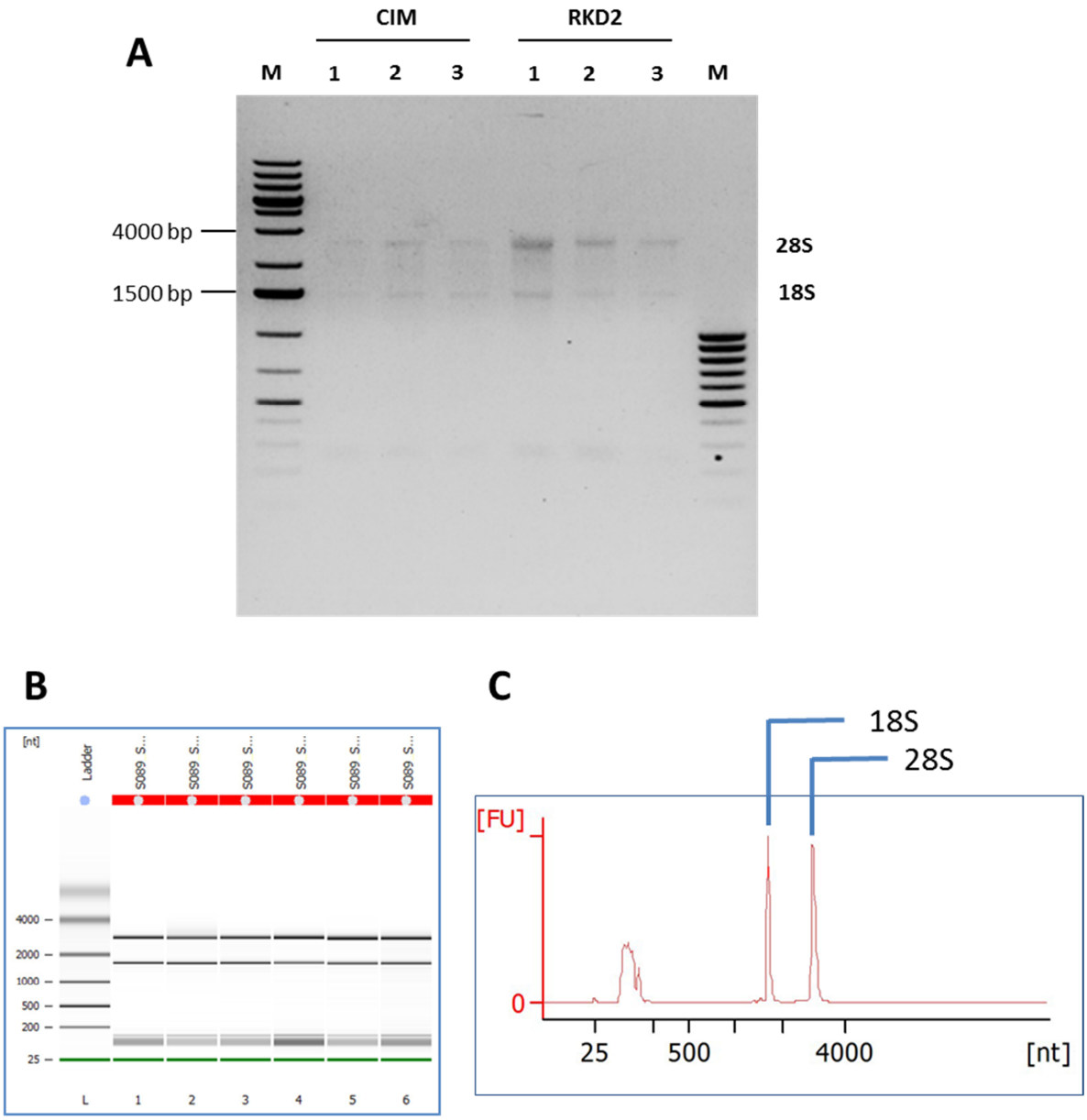


Figure 44: Quality control of total RNA samples
(A) Non-denaturing gel electrophoresis of RKD2 and CIM total RNA Positons of the ribosomal 28S, 18S and 5S rRNA are marked with an arrow. **(B)** and **(C)** Analysis of RNA quality with the Agilent Bioanalyzer **(B)** Electropherogramm showing all six total RNA samples used for library generation (1-3: Auxin-induced callus, 4-6: RKD2-induced callus. **(C)** Electropherogram from lane six. Relative amount of RNA is measured by fluorescence. Peaks at 3000 and 1800 bp represent the 28S and 18S ribosomal RNA respectively.

5.8.3 Data validation by qRT PCR

Fold changes of a selection of genes spanning the complete dynamic spectrum of expression were validated with qRT-PCR, including all ten Argonaute genes. The two reference genes *UBIQUITIN-CONJUGATING ENZYME 21 (UBC21, AT5G25760)* and *EUKARYOTIC TRANSLATION INITIATION FACTOR 4G* (*eIF4G, AT3G60240*) were used for normalization.

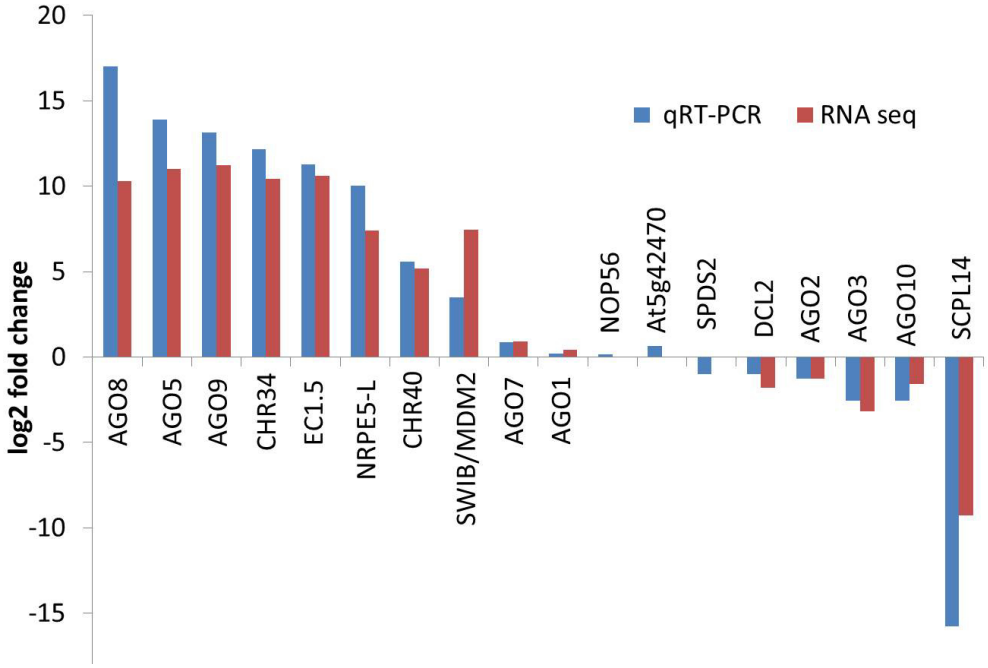


Figure 45: Validation of RNA-Seq expression data by qRT PCR.
Comparison of expression of selected genes determined by RNA-Seq (blue bars) and validated by qRT-PCR (red bars). X-Axis shows validated genes, Y-Axis shows log2 fold change of expression (RKD2-induced versus CIM callus). *AGO, ARGONAUTE; CHR34, CHROMATIN REMODELING 34; EC1.5, EGG CELL 1.5; NRPE5-L, RNA POLYMERASE II FIFTH LARGEST SUBUNIT, E CHR40, CHROMATIN REMODELING 40; SWIB/MDM2, SWIB/MDM2, PLUS-3 AND GYF DOMAIN-CONTAINING PROTEIN; NOP56, ARABIDOPSIS HOMOLOGUE OF NUCLEOLAR PROTEIN 56; AT5G42470, BRE (BRAIN/REPRODUCTIVE ORGAN EXPRESSED) DOMAIN-CONTAINING PROTEIN; SPDS2, SPERMIDINE SYNTHASE 2, DCL2, DICER-LIKE 2, SCPL14, SERINE CARBOXYPEPTIDASE 14.*

Genes marked as differentially expressed in RNA-Seq data were confirmed to be likewise expressed by statistical evaluation of qRT PCR dCT value of each sample. P-values were determined by a two-tailed paired Students t-test. P-values below 0.005 were considered as significant.

5.8.4 Data validation by analysis of the Köszegi dataset

The specific expression profile of the RKD2-induced cell culture was revealed by Köszegi et al. (2011) with the help of microarray-based transcriptome profiling. In order to validate the transcriptome data obtained by RNAseq, a set of 99 genes was chosen, which are reported in this study to be upregulated at least sevenfold compared to CIM callus and seedling (Figure 46). A similar expression pattern was observed in our mRNA seq data. Ninety of the analyzed genes showed also a fold change of seven or more compared to the CIM callus.

Seven genes were upregulated in the RKD2 callus, but with a fold change of less than seven. One example is the transcriptional activator *CUP-SHAPED COTYLEDON 2* (*CUC2, AT5G53950*) with a log2 fold change of 2.06, which equals a single fold change of 4.17. Another example is the *ARABIDOPSIS NAC DOMAIN CONTAINING PROTEIN 98* (*ANAC098*, *AT5G53950*) with a log2 fold change of 2.06. Nevertheless, differential expression analysis showed a corrected p-value of less than 0.0001, so those genes can be regarded as differentially expressed.

Only two of the 99 examined showed significant differences compared to the Köszegi dataset: The gene encoding the *CALLOSE SYNTHASE 1* (*CALS1, AT1G05570*) which was reported to be upregulated in the RKD2 callus is found in both callus types (log2 fold change 0.27) according to our RNAseq data. A gene encoding a putative protein kinase *(AT1G71530)* is only upregulated by a log2 fold change of 1.78 in the RKD2-induced callus, and shows a corrected p-value of more than 0.0001. In general, analysis of our expression data showed, with minor exceptions, a highly similar expression pattern compared to the dataset of Köszegi et al.

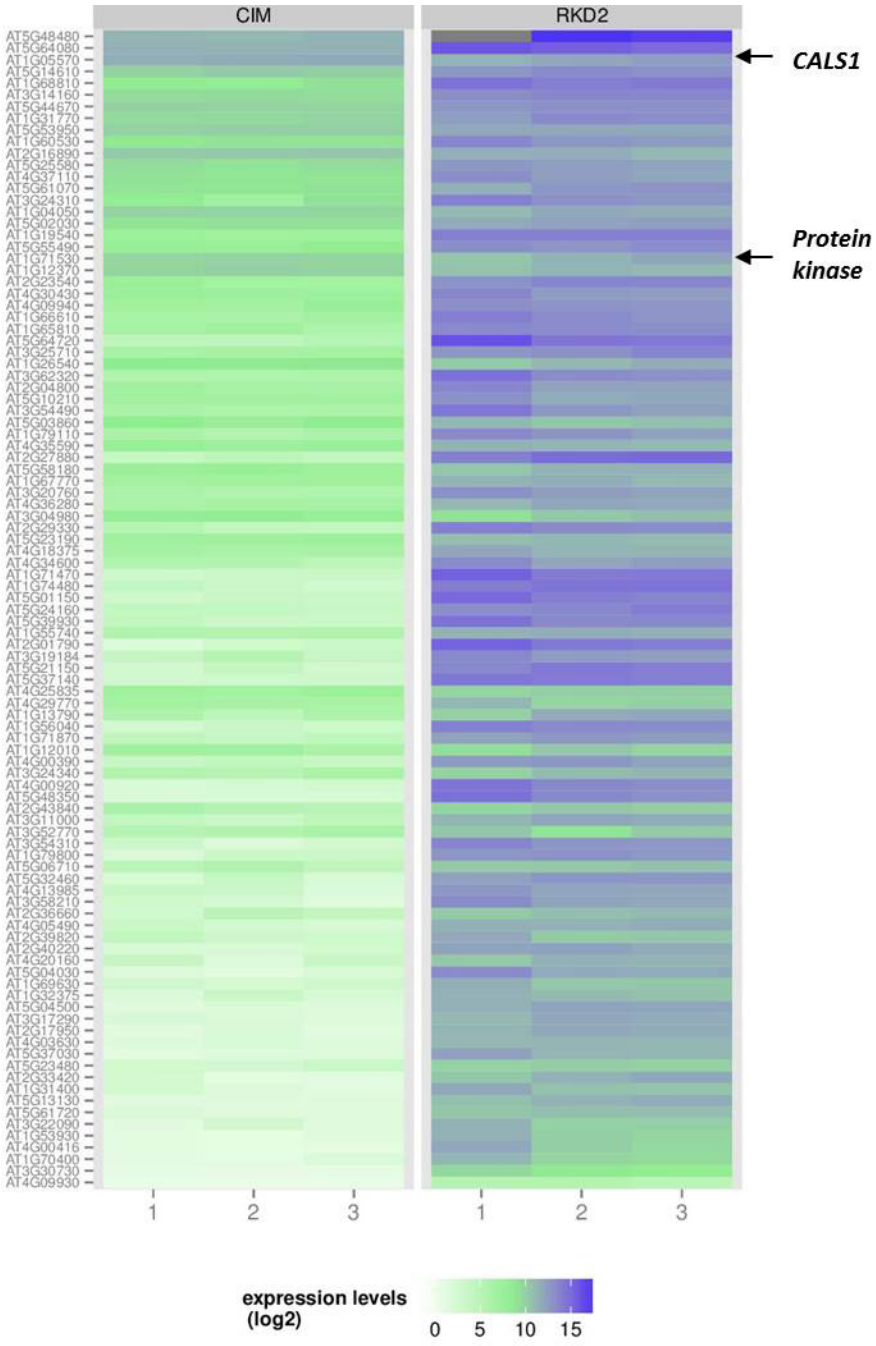


Figure 46: Microarray data from RKD2 callus.
A set of 99 genes which are reported to be upregulated at least sevenfold compared to CIM callus by Köszegi et al. (2010) was analyzed for expression in own RKD2-induced and CIM callus. Heat map shows count numbers according to mRNA-Seq data from the RKD2-induced cell line compared to auxin-induced callus (CIM) (three biological replicates each). Arrows mark the only two genes which are not significantly upregulated in our mRNAseq data. A list of all examined genes with fold changes is included in the digital supplement.

5.8.5 Differentially expressed mRNAs in the RKD2-induced cell lines, compared with the CIM callus

The transcript profiles derived from three biological replicates of the RKD2-induced cell line was compared with the mRNA-Seq data obtained from the phytohormone-induced CIM callus. Differential expression analysis of the mRNA-Seq data was conducted by Dr. Julia Engelmann and Nicolas Strieder institute for functional genomics, department for statistical bioinformatics, using the DESeq2 software package (Love et al. 2014). For heatmap visualization, regularized log (rlog) values from DESeq2 were used.

Comparing the RKD2-induced cell lines gene expression profile with that of the CIM callus, a lot of differentially expressed genes are found (Figure 47). In numbers, 5,511 genes are significantly differentially expressed (FDR adjusted p-value < 0.00001). 2,882 genes were upregulated and 2,629 genes were downregulated in the RKD2-induced cell line. A complete list of all differentially expressed genes can be found as digital file on the attached CD.

Genes found to be strongly upregulated in the RKD2-induced cell line are often also reported to be present in female gametophytic cells or even expressed exclusively in the egg cell (Table 8). The *EC1.5* gene, for example, belongs to the *EC1* gene family, which is egg cell-expressed and plays an important role during double fertilization (Sprunck et al. 2012). Microarray data from isolated cells of the female gametophyte showed presence of *EC1.5* transcript in the egg cell and synergids (Soljic, 2012). However, a look at the other genes from the *EC1* gene family shows that only EC1.1 is present in the transcriptome data, it is expressed at low levels in both CIM (22.73 reads) and RKD2-induced cells (34.93 reads). The other genes of this family (*EC1.2, EC1.3* and *EC1.3*) are neither present in the RKD2-induced nor in the CIM callus (expression values in the digital supplement). This stands in contrast to a previous study in which egg cell specific expression of all *EC1* genes is reported (Sprunck et al. 2012) and shows that the expression profile of the RKD2-induced is not absolutely egg cell-like. Also *AGO9* has been identified as differentially expressed between the RKD2-induced and CIM cell lines (Table 8). The high transcript abundance of *AGO9* in the egg cell was proved by microarray based expression studies on isolated egg cells (Soljic 2012) and by promoter-reporter studies (Figure 21). Another example is a gene encoding for a protein with domain of unknown function (*DUF674*, *AT5G37320*) which is upregulated in the RKD2 induced callus with log2 fold change of 11.37 (Table 8). Promoter-reporter studies identified this gene as exclusively expressed in the egg cell (T. Hackenberg, unpublished).

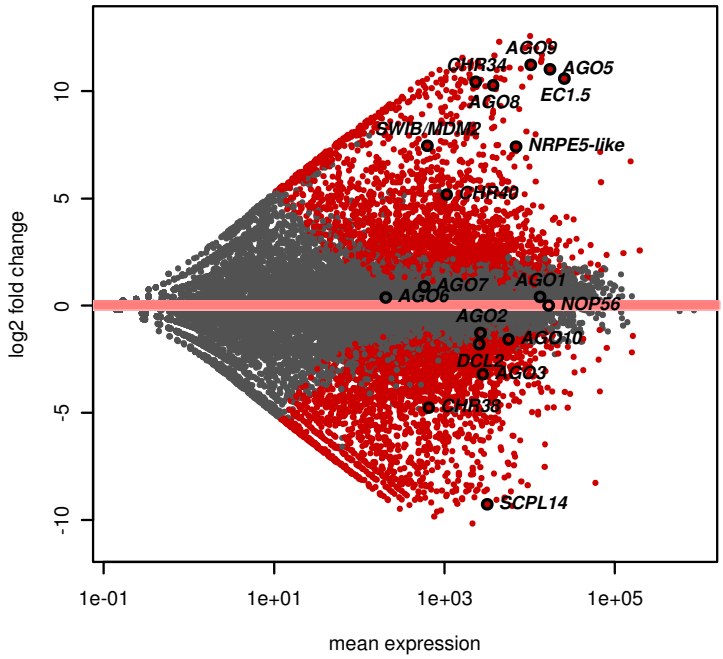


Figure 47: Gene expression differences between the RKD2 and the CIM callus.
Genes are shown as grey dots, genes significantly regulated between the two calli are highlighted in red. Fold changes of genes validated by RT-PCR are highlighted and labeled in black. Positive log2 fold changes represent genes induced in the RKD2 callus, negative fold changes indicate repressed genes compared to the CIM callus. The horizontal axis indicates the mean expression level (in counts) across both types of calli.

A general look at the most upregulated genes in the RKD2-induced cell line reveals the presence of a variety of gene classes. Genes encoding transcriptional regulators are found as well as genes encoding membrane proteins and proteins of unknown function (Table 6). Comparison of our differential expression analysis of mRNA-Seq data to previously published microarray data revealed that of the top 30 upregulated genes in RKD2-induced 17 are transcribed in cells of the embryo sac (Wuest et al. 2010, Soljic, 2012).

Eight of the top 30 upregulated genes cannot be found in microarray expression data and five genes are described as absent in the female gametophyte. One of those five genes is the *ISOPENTENYL ISOMERASE 7* (*IPT7*), a gene involved in cytokinin biosynthesis is upregulated in the RKD2-induced callus with a log2 fold change of 11.37 (Table 6), but according to the ATH1 gene chip data, *IPT7* is not expressed in the egg cell.

Moreover, some genes strongly upregulated in the RKD2-induced cell line are not represented on the ATH1 gene chip. This concerns, as mentioned above, 8 of the Top 30 upregulated genes in the RKD2 induced cell line.

Table 6: Upregulated genes in RKD2 callus and known expression within the female gametophyte.
Table showing the top 30 upregulated genes in RKD2 callus and their expression in the cell of female gametophyte (microarray data). Abbreviations: c, central cell es, embryo sac; ec, egg cell; s, synergid; N.A., not applicable

AGI identifier	Description	log2 (RKD2/CIM)	Expresssion in female gametophyte	Reference
AT5G48350	Polynucleotidyl transferase	12.58	ES	Wuest et al.2010, Soljic 2012
AT2G01790	TRAF-like family protein	12.34	EC, S	Soljic 2012
AT4G00320	F-box/RNI-like superfamily protein	12.31	ES	Soljic 2012
AT4G00920	COP1-interacting protein-related	12.12	EC	Wuest et al.2010
AT1G71470	Unknown protein	12.03	EC, S	Yu et al., 2005, Wuest et al.2010, Soljic 2012
AT1G60410	A paternally expressed imprinted gene	12.02	N.A.	
AT5G37140	P-loop containing nucleoside triphosphate hydrolases superfamily	12.02	EC, S	Soljic 2012
AT2G05420	TRAF-like family protein	11.71	N.A.	
AT5G04030	Unknown protein	11.67	EC	Soljic 2012
AT5G44585	Unknown protein	11.56	N.A.	
AT1G74480	RKD2	11.49	ES	Wuest et al.2010, Soljic 2012
AT5G04500	A member of the Glycosyltransferase Family 64	11.37	ES	Soljic 2012
AT5G01150	Protein of unknown function (DUF674)	11.37	EC, S	Wuest et al.2010, Soljic 2012
AT1G55950	DNA-binding storekeeper protein-related transcriptional regulator	11.36	Absent	
AT5G37030	P-loop containing nucleoside triphosphate hydrolases superfamily	11.28	(EC)	Wuest et al.2010, Soljic 2012
AT3G23630	ISOPENTENYLTRANSFERASE 7 (Cytokinin synthesis)	11.26	Absent	
AT5G21150	AGO9	11.24	ES	Wuest et al.2010, Soljic 2012
AT1G31390	TRAF-like family protein	11.15	N.A.	
AT2G17950	WUSCHEL	11.13	Absent	
AT3G54310	Unknown protein	11.11	EC, S	Soljic 2012
AT5G13130	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase family protein	11.09	EC, S	Soljic(2012
AT1G56040	HEAT/U-box domain-containing protein	11.06	EC, S	Soljic 2012
AT1G11510	DNA-binding storekeeper protein-related transcriptional regulator	11.06	N.A.	
AT4G01180	XH/XS domain-containing protein	11.06	C, EC, S	Soljic 2012
AT2G27880	AGO5	11.03	C, EC, S	Wuest et al. 2010, Soljic 2012
AT2G44745	WRKY DNA-BINDING PROTEIN 12	11.00	Absent	Soljic 2012
AT5G37320	Protein of unknown function (DUF674)	10.99	N.A.	
AT4G00416	METHYL-CPG-BINDING DOMAIN 3	10.99	N.A.	
AT5G37150	P-loop containing nucleoside triphosphate hydrolases superfamily protein	10.90	Absent	Wuest et al.2010, Soljic 2012
AT2G16310	Pseudogen, RER1-related protein	10.80	N.A.	

For example, two members of the tumor necrosis factor receptor associated factor (TRAF)-like gene family (*AT2G05420* and *AT1G31390*) are upregulated in the RKD2-induced cell line with log 2 fold changes of 11.71 and 11.15. Genes upregulated in the RKD2-induced cell which are not on the ATH1 gene chip is also *METHYL-CPG-BINDING DOMAIN 3* (*AT4G00416*), a paternally expressed gene (*AT1G60410*) and the already mentioned *DUF674*.

In general, differential expression analysis of mRNAs shows that many of the genes which are upregulated in the RKD2 induced can also be found in cells of the female gametophyte, especially the egg cell, but that there are some differences compared to the egg cell transcriptome.

5.8.6 Small RNA pathway components

The mRNAseq data could confirmed the results of the RT-PCR analysis, with *AGO5*, *AGO8* and *AGO9* being expressed predominantly in reproductive tissues and all other AGOs being present ubiquitously and both in the RKD2-induced cell line and in CIM callus (Chapter 5.21). However *AGO2*, *AGO3* and *AGO10* show significantly higher expression values in the CIM callus compared to the RKD2-induced callus, with the adjusted p-value being lower than 0.00001 (Table 7). A global look at genes involved in small RNA pathways reveals only a few genes differentially expressed in the RKD2-induced cell line (Figure 48). An example for a gene which is upregulated in RKD2 callus is *IDN2 PARALOG* (*IDP2*; *AT4G00380*), a double strand RNA binding protein, which is part of chromatin remodeling in RdDM (Zhang et al. 2012). *IDP2* is upregulated with a log2 fold change of 4.22 in the RKD2-induced callus. Strongly upregulated is also the histone deacetylase *HDA18*, which sets epigenetic marks on histones H3/H4 and affects gene expression at such modified loci (Table 21).

The DNA-glycosylase DEMETER is also upregulated in the CIM callus, with expression values being approximately seven times higher than in the RKD2 callus. Count values for the two maintenance DNA-methyltransferases CMT3 and MET1 are very high in both RKD2-induced cell line and CIM callus.

Table 7: Arabidopsis *AGO* expression.
Expression of the 10 Argonaute genes in RKD2-induced and CIM according to RNA-seq data compared to microarray transcriptome profiling of isolated cells of the female gametophyte (Soljic 2012). Absent calls in the microarray data are marked with an asterisk.

		RNA sequencing (normalized counts)			Microarray data expression value (Soljic 2012)			
AGI identifier	**Description**	**CIM**	**RKD2**	**log2 (RKD2/CIM)**	**Central cell**	**Egg cell**	**Synergid**	**Seedling**
AT1G48410	AGO1	11339	15066	0.41	1751	4317	1884	1012
AT1G31280	AGO2	3744.55	1541.99	-1.28	315	938	434	62
AT1G31290	AGO3	5070.49	551.77	-3.20	57	23*	103	51
AT2G27040	AGO4	4112.90	3031.76	-0.44	185	659	500	232
AT2G27880	AGO5	16.51	34524.41	11.03	2217	4418	1775	16*
AT2G32940	AGO6	177.34	230.78	0.38	126	367	289	45*
AT1G69440	AGO7	403.12	747.06	0.89	35	65	96	42
AT5G21030	AGO8	5.97	7421.41	10.28	17*	1027	191	25*
AT5G21150	AGO9	8.47	20485.57	11.24	1745	5901	2545	12*
AT1G43810	AGO10	8435.48	2841.14	-1.57	28*	32*	40*	12*

CMT3 has 1,768 reads in the CIM callus and 7,847 reads in the RKD2-induced callus. MET1 has 2,838 reads in the CIM callus and 2,722 reads in the RKD2-induced callus. This stands in contrast to the observation of Jullien et al. (2012), that reporter fusions of both genes, under the control of their own promoter, do not give rise to reporter-gene derived fluorescent signals in the egg cell. The cytosine methyltransferase *DRM2* (*DOMAINS REARRANGED METHYLTRANSFERASE 2*; *AT5G15380*) is expressed in both CIM and RKD2 callus at almost equal levels. *DRM3* (*DOMAINS REARRANGED METHYLTRANSFERASE 2*; *AT5G15380*) is highly expressed in both RKD2-induced cell line and CIM callus (1,246 and 2,111 reads). Notably, the third member of this gene family, *DRM1* (AT5G15380), is predominantly expressed in the RKD2-induced callus (log FC). According to egg cell transcriptome data (Soljic, 2012), *DRM1* transcripts are present in the egg cell and, to a lesser content, in the central cell and the synergid.

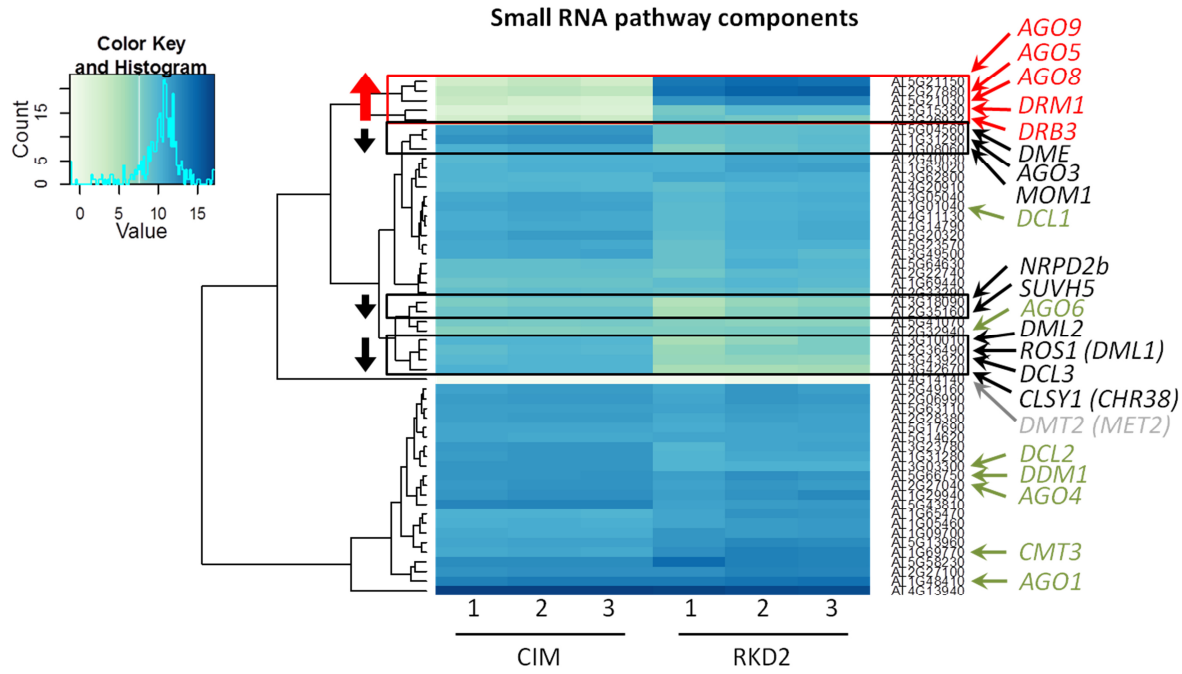


Figure 48: Heatmap of sRNA pathway components.

Several small RNA components appear to be differential expressed in own RKD2-induced callus, compared to CIM callus. Heat map shows count numbers according to mRNA-Seq data from the RKD2-induced cell line compared to auxin-induced callus (CIM) (three biological replicates each). Induced genes are labeled in red, repressed genes in black; genes with similar number of reads are in green. Genes without any reads are in grey. *AGO8*, *AGO9*, and *AGO5* are induced in RKD samples, as well as *DRM1* (*DOMAINS REARRANGED METHLYTRANSFERASE 1*) and *DRB3* (*DOUBLE-STRANDED RNA-BINDING PROTEIN 3*).

5.8.7 Pol IV and Pol V components

The RNA polymerases IV and V are plant specific and their function is restricted to their role in RdDM. Differential expression analysis revealed the presence of three Pol IV/V subunits which are upregulated in the RKD2-induced egg cell-like callus: *NRPE3B, NRPE4* and *NRPE5-L*. NRPE4 (AT4G15950) is a non-catalytic subunit of Pol IV and Pol V. In contrast to this, *NRPE3B* (*AT2G15400*) is only part of Pol V. The gene encoding *NRPE3B* is upregulated log2 3.24 times in the RKD2-induced cell line. NRPE5-L (*AT3G54490*), a putative Pol V subunit, is characterized by its homology to NRPE5. *NRPE5-L* shows a high expression level in the RKD2-induced callus, but is only detected in CIM callus tissue at a very low level, with a log2 fold change of 7.42 (which equals a single FC of 171.25).

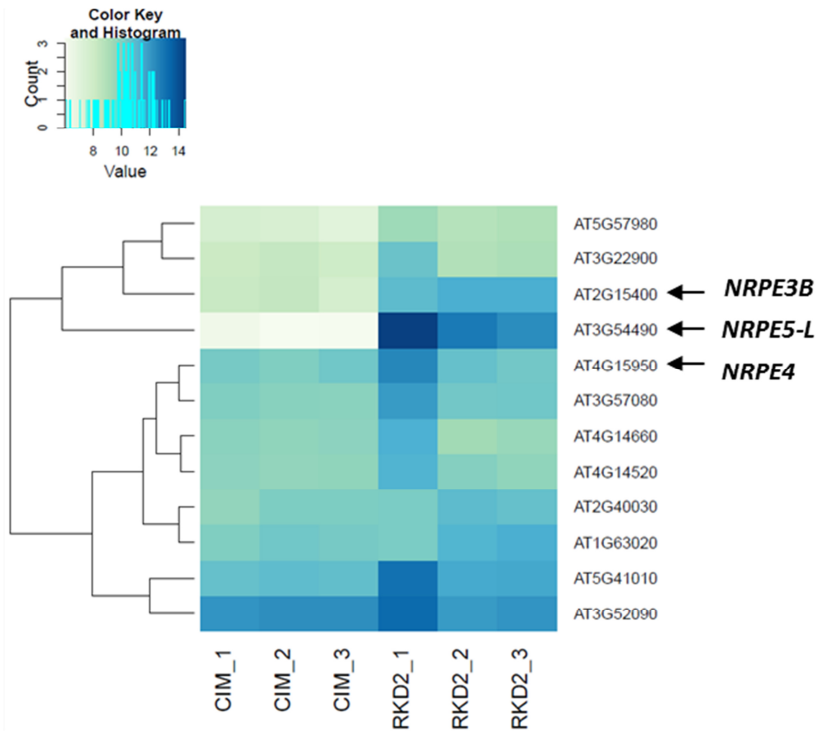


Figure 49: Differential expression of genes encoding subunits of the RNA-polymerase Pol IV and Pol V. Heat map shows count numbers according to mRNA-Seq data from the RKD2-induced cell line compared to auxin-induced callus (CIM) (three biological replicates each). The upregulated *NRPE3B*, *NRPE5-L* and *NRPE4* genes are marked by arrows.

5.8.8 Auxin and Cytokinin-related genes

In plant cell cultures, the in vitro culture medium must be supplemented with a constant ratio of cytokinin to auxin to induce callus formation from plant tissue (explants) and to maintain cell division. A shift towards more cytokinin or auxin will result in unwanted organogenesis. Notably, the RKD2-induced cell line is able to grow infinitely without any addition of phytohormones. This confirms the results of Köszegi et al. (2011), where cultures could be maintained in hormone-free medium. The observed phytohormone-independent growth of the RKD2-induced cell line led to the question if genes related to auxin and cytokinin biosynthesis, transport or signaling are differential expressed in the RKD2-induced callus, compared to the CIM callus that grows on callus induction medium supplemented with 0.5 mg/L auxin and 0.05 mg/L cytokinin. In order to answer this question the expression of auxin and cytokinin-related genes has been examined in RKD2-induced and CIM-induced cell lines.

The auxin signaling network in the plant includes the activity of genes involved in auxin biosynthesis and genes encoding auxin influx (*AUX/LAX*) and efflux carriers (*PIN*), auxin receptors (*TIR1/AFB*), a group of transcriptional repressors (*AUX/IAA* proteins) and a group of transcription factors that bind to auxin responsive elements, the so-called auxin responsive factors (*ARFs*). Auxin-responsive genes fall into three major families, the so-called *GH3*, small auxin up RNA (*SAUR*), and the *AUX/IAA* family. Some genes involved in the auxin network show differential expression in the two callus tissues (Figure 50). In the family of *ARFs* only two members are slightly upregulated in the RKD2-induced callus: *ARF11* (log2 fold change 3.83) and *ARF18* (log2 fold change 1.21). Among the IAA family members, only *IAA18* (log2 fold change 2.41) and *IAA26* are (log2 fold change 1.21) slightly upregulated in the RKD2-induced callus.

Expression of PIN-Proteins in RKD2 and CIM callus follows different expression patterns. In general, genes encoding PIN proteins are upregulated in the CIM callus. *PIN1, PIN3, PIN4* and *PIN7* are present in considerable amounts in the RKD2-induced callus but are down regulated compared to the CIM callus (log2 fold changes of -0.9, -2.17, -1.12 and -2.77 respectively. In contrast to this, *PIN2* is almost absent from RKD2 callus, with approx. 2,000 counts in CIM callus and only 8 in RKD2 callus. The auxin influx carrier *AUX1* is upregulated in the RKD2 callus (log2 fold change 1.75).

The *GH3* family of early auxin responsive can reduce the amount of free auxin in the cell by conjugating it with an amino acid, which is mostly isoleucin (Okrent and Wildermuth 2011). *GH3* genes shows distinct expression patterns in RKD2- induced and CIM callus. *GH3.1*, *GH3.3*, *GH3.5* and *GH3.6* are reported to be induced by exogenic auxin (Okrent and Wildermuth 2011) and are downregulated in the RDK2-induced callus with log2 fold changes of –6.65, -4.76, -1.96 and -1.92. This could be the result of the added auxin in the culture medium. *GH3.2* and *GH3.12* are strongly down regulated in the RKD2 callus (log2 fold change -5.08 and -5.23). GH3.2 can conjugate auxin to amino acid and GH3.12 can do the same with benzoates (Okrent et al. 2009). *GH3.11* and *GH3.17* are the only members of their family which are (slightly) upregulated in the RKD2 callus with a fold change of about 2. In contrast to other members of its family, GH3.11 can use jasmonat instead of auxin for conjugation with isoleucine. Genes involved in auxin biogenesis are present in both RKD2 and CIM callus (Figure 55). The enzyme tryptophan aminotransferase of Arabidopsis 1 (TAA1), which converts tryptophan into indole-3-pyruvic acid in an initial step of auxin biosynthesis (Mano and Nemoto 2012), is present in both cell types. This enzyme is

upregulated in the RDK2-induced callus with a log2 fold change of 2.45. Also the acylamidohydrolase AMI1, which catalyzes a final step of auxin bio synthesis, is upregulated in the RKD2 callus (log2 fold change 2.54).

A general look at auxin and its signaling network reveals slight differences between genes involved in auxin biosynthesis, were some crucial genes (*TAA1* and *AMI1*) are slightly but significantly upregulated, which might reflect the fact that the RKD2 callus grows on a medium without any phytohormones and synthesizes auxin in larger amounts than the CIM callus, in order to maintain the undifferentiated status of its cells. The situation concerning auxin transporters is different in the two compared tissues. PIN auxin efflux transporter proteins are generally higher expressed in the CIM callus, while the AUX1 influx carrier is more abundant in the RKD2-induced cell line.

Altogether clear differences concerning the auxin network in both compared tissues can be observed, which reflect the fact that auxin regulates cell fate and differentiation status in different ways, for example by synthesis rate (TAA1 and AMI1), by different intracellular concentrations and localization (PINS and AUX1) but also by modifying its biological activity (GH3 proteins).

Cytokinin is important for processes of plant growth and differentiation. It is required to promote cell division and is involved in root and shoot morphogenesis, in combination with auxin. In cell cultures equal levels of auxin and cytokinin promote formation of undifferentiated callus cells (Skoog and Miller, 1957). An excess of auxin promotes root formation while shoot formation requires more cytokinin. Genes involved in cytokinin biosynthesis, cytokinin degradation, cytokinin perception and cytokinin signaling were analyzed (Figure 49). The first step of cytokinin biosynthesis in higher plants is catalyzed by adenosine phosphate-isopentenyltransferases (IPTs), which catalyze *N*-prenylation of adenosine 5-phosphates (AMP, ADP, ATP) and produce the important cytokinin precursor isopentenyladenine (El-Showk et al. 2013). The RKD2-induced cell line and the CIM callus show a slightly different expression pattern of IPTs: *IPT1* is found at low levels in CIM callus and is upregulated in RKD2-induced cells. *IPT5* is upregulated in CIM callus, compared to the RKD2-induced cell line. Notably, *IPT7* is expressed very strongly in RKD2-induced cell line while it is not expressed in the CIM callus (Figure 50). Both cell cultures contain considerable amounts of *IPT2* and *IPT9* transcripts.

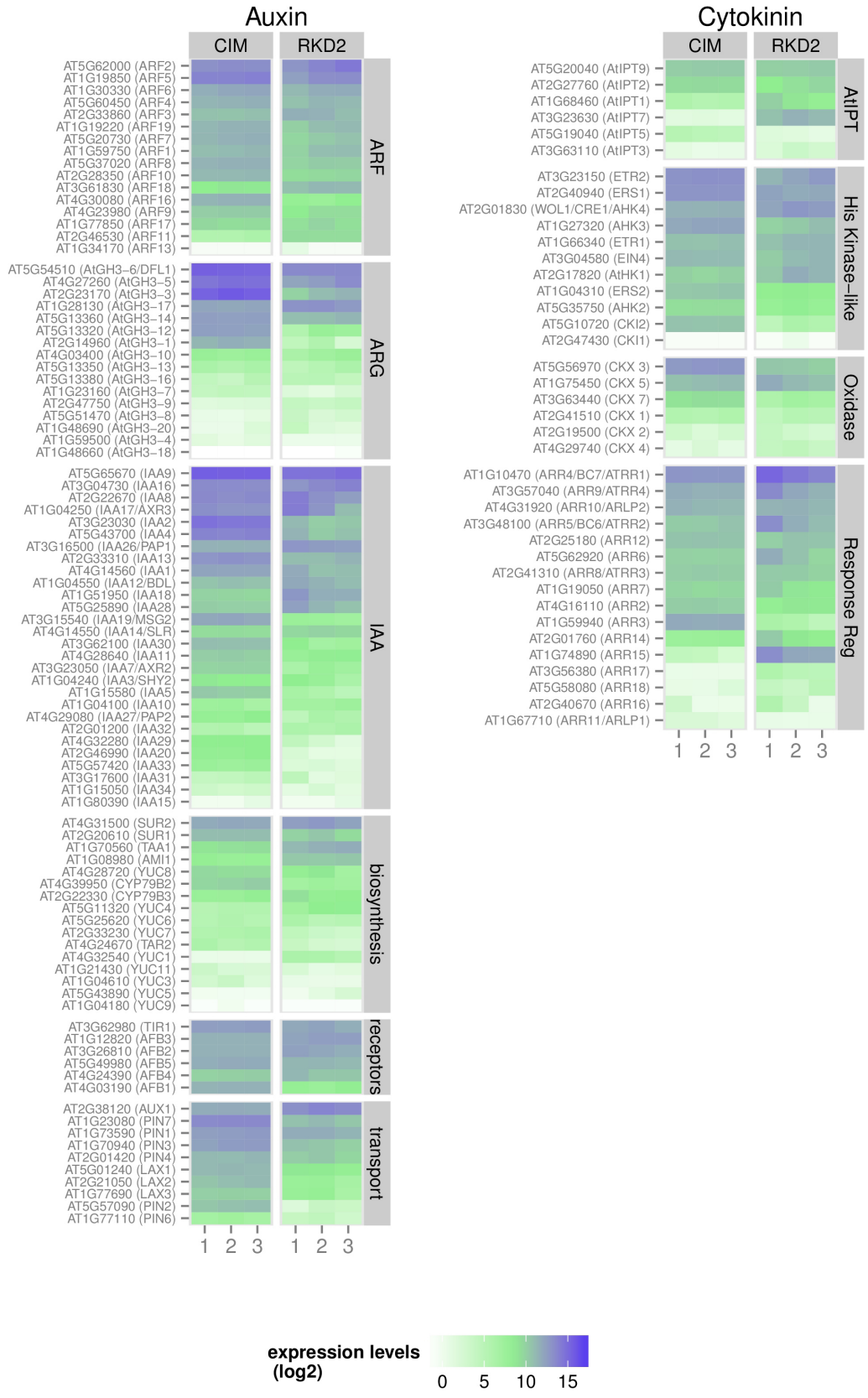


Figure 50: Differential expression of genes involved in auxin and cytokinin synthesis and signaling networks. Heat map shows count numbers according to mRNA-Seq data from the RKD2-induced cell line compared to auxin-induced callus (CIM) (three biological replicates each).

Those two genes belong to a subgroup of IPTs, which can add isopentenyl residues to certain types of tRNAs, which can, if modified in such a way, be further processed into Z-type cytokinins. In Arabidopsis, Cytokinin is perceived by membrane-located sensor histidine kinases, the *ARABIDOPSIS HISTIDINE KINASE2 (AHK2)*, *AHK3*, and *CYTOKININ RESPONSE1/AHK4*. In both RKD2 and CIM callus, those genes are expressed and are show only slight difference in expression (*AHK2* -1.38 log2 fold change, *AHK3* -1.72 log2 fold change and *AHK4* 1.03 log2 fold change.

Cytokinin signaling is mediated by a two-component system-based phosphorelay that transmits a signal from the cytokinin receptors, through histidine phosphotransfer proteins, to the downstream Arabidopsis Response Regulators (ARRs). ARRs comprise type-A and type-B *ARR* genes, based on domain structure and sequence (Imamura et al. 1999). Type-A *ARRs* (*ARR3* to *ARR9*, *ARR15* to *ARR17*) can be rapidly induced by cytokinin and act as negative regulators of cytokinin signaling. Three of the eleven type-B *ARRs* (*ARR1*, *ARR10*, *ARR12*) are known to act as positive regulators in the two-component cytokinin signaling pathway and induce the transcription of many genes (Sakai et al. 2001, Mason et al. 2005). In the RKD2-induced cell line two type-A *ARRs* appear to be differentially regulated (*ARR3* log2 fold change 6.32 down in RKD2 and *ARR15* 8.98 log2 fold change up in RKD2-induced callus) (Figure 55). *ARR3* is implicated to play a role in meristem maintenance and regeneration (Buechel et al. 2009). ARR15 is, together with ARR7 one of two auxin inducible ARRs (Müller and Sheen 2008) and plays a role in regulating cell fate in the shoot apical meristem (Zhao et al. 2010). ARR7 is not differentially regulated (-0.26 log2 fold change). Type-B *ARRs*, playing a pivotal role in the early response of plants to cytokinin, show no obvious differential expression comparing RKD2-induced and CIM callus.Degradation of cytokinin is catalyzed by cytokinin oxidase/dehydrogenase (CKX) enzymes. In general, the CIM callus shows higher expression of *CKX* genes; *CKX3* and *CKX7* are upregulated in this tissue with log2 fold changes of 2.76 and 2.67, respectively. *CKX5* is expressed in both tissues at a high level (log2 fold change 0.33 with 2600 reads in the RKD2 and 2100 in the CIM callus). The elevated expression levels of *CKX3* and *CKX7* might be the result of the cytokinin which contained in the medium.

Genes involved in the cytokinin network show different expression patterns in RKD2-induced and CIM callus: this concerns genes for cytokinin synthesis (*IPT7*), cytokinin degradation (*CKX3* and *CKX7*), as well as response regulators (*ARR3* and *ARR15*).

5.8.9 Small RNA profiling of RKD2-induced cells with egg cell-like fate

For each callus tissue (CIM callus and RKD2-induced cell line), three small RNA libraries were generated (three biological replicates). After size and quality trimming (chapter 4.5.7) 16 to 29 million clean reads per library could be obtained. From the CIM callus small RNA library, 17 to million reads could be mapped to the *Arabidopsis* genome, containing 11 to 12 million unique reads. The RKD2-induced cell line generated 8 to 9 million mapped small RNA reads, 1.5 to 2.3 million uniquely mapped reads included. Both tissues generated libraries with a sufficient number of reads. An important difference between CIM and RKD2 derived samples is the varying number of uniquely mapped reads. Ten percent of reads from the RKD2-induced callus could be mapped to a single genomic position, whereas in the CIM callus, 42% of reads are mapped uniquely. The lower percentage of uniquely mapped reads in the RKD2 callus may be caused by a higher number of small RNAs derived from repetitive elements, such as transposable elements, which could be mapped to multiple genomic positions.

5.8.10 Small RNA length distribution and annotation

Reads from the small RNA libraries were analyzed respective to their length and their annotated features, as found in TAIR10. Length distribution of small RNAs (sRNAs) showed the prevalence of 21 and 24 nucleotides long sRNAs, which correspond to the classes of siRNAs and miRNAs, respectively (Figure 51A). Comparing the callus types, some differences can be seen. In the size class of 21 nt sRNAs, the RKD2-induced callus only 10% of the total small RNA reads map to miRNAs, compared to the CIM callus in which about 28% of the 21 nt reads can be assigned to miRNAs. Instead more 21 nt sRNA reads map to transposable elements in the RKD2 callus, this class comprises 5 % of total reads whereas in the CIM callus only about 1% belong to that class.

Looking at the class of 24 nt reads, sRNAs derived from transposable elements are also found more often in the RKD2 callus (25% of total reads) than in the CIM callus (15% of total reads). siRNAs against protein coding genes (24 nt reads) are found more often in the CIM (10%) than in the RKD2 callus (4%, see Figure 51A).

In general, the RKD2-induced callus produces more small RNAs against transposable elements (45% versus 24%). Furthermore, the RKD2 callus has increased levels of 18, 19 and 24 nt long reads originating from tRNA loci (10% vs 1%).

Small RNA sequences derived from protein-coding genes are represented across the complete range of read lengths analyzed. These small RNAs could either represent siRNAs directed against proteins or be degradation products of protein coding mRNAs.

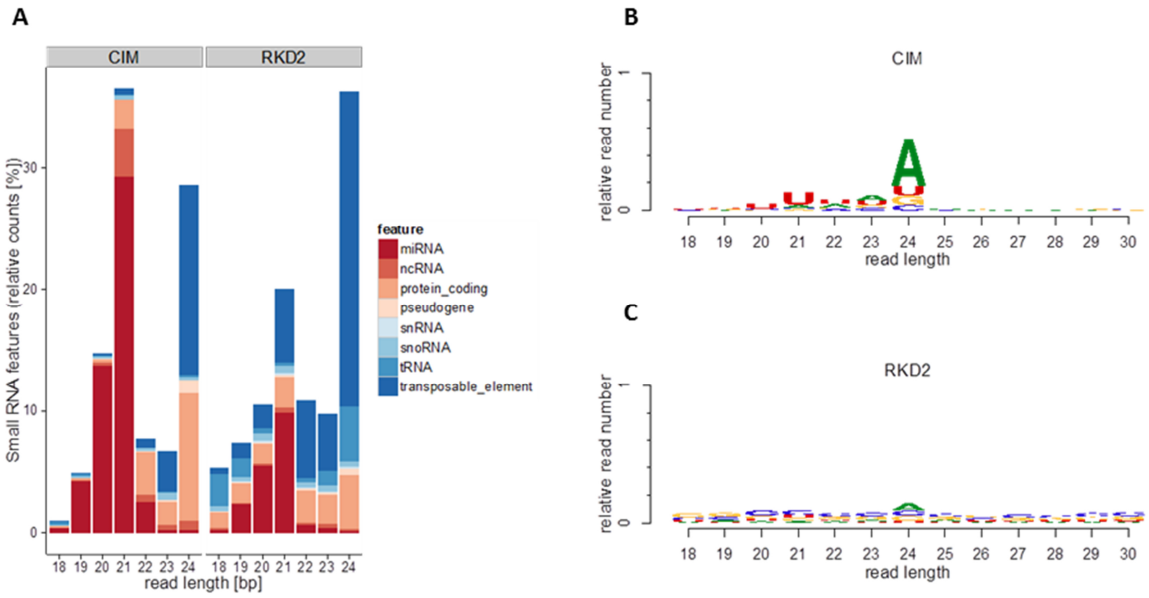


Figure 51: Small RNA feature and 5′ bias in CIM- and RKD2-induced cells.
(A) Relative counts of small RNA counts in CIM and RKD2 –induced callus, sorted by read length **(B)** Relative frequency of each nucleotide (A, U, G, or C) at the 5′-terminal position sorted by read length in the CIM callus sorted by read length **(C)** Relative frequency of each nucleotide (A, U, G, or C) at the 5′-terminal position sorted by read length in the CIM callus sorted by read length. Estimations of first base nucleotide ratios were performed using R/Bioconductor (Anders et al. 2014).

5.8.11 Analysis of 5′ terminal nucleotide frequency

Since it is known that sorting of small RNAs into Arabidopsis Argonaute complexes is directed by the 5' terminal nucleotide (Mi et al. 2008), we had a look at the first base of the sequenced small RNAs. In the CIM callus, 21 nt sRNAs preferentially begin with a U whereas 24 nt long sRNAs mostly start with an A. AGO1, the main effector in the miRNA pathway, mainly binds 21 nt miRNAs beginning with a U, so these findings indicate activity of the AGO1/miRNA pathway in the CIM callus. Small RNA reads of 24 nt length show a strong bias towards an 5' terminal A. Small interfering RNAs of this size are involved in RdDM, mostly begin with A and are bound by AGO4/6/9 (Havecker et al. 2010).

In the RKD2 callus, the situation is less clear; 24 nt RNAs shows a slight bias towards an 5' terminal A, but in the size class of 21 nt no clear tendency can be seen.

Recently, a method for the biochemical isolation of Argonaute protein complexes has been developed (Hauptmann et al. 2015). This method uses the high ability of a peptide from a member of the TNRC6/GW protein family to interact with Argonautes and is referred to as "Ago protein Affinity Purification by Peptides" (Ago-APP). Using this method all four human AGO complexes could be precipitated. Additionally, Ago-APP was performed on protein lysates of the RKD2-induced cell culture and AGO4, AGO6, and AGO9 could be efficiently pulled down. Further analyzes of small RNAs bound to Ago-APP precipitates revealed the enrichment of 24 nt siRNAs in those samples.

5.8.12 Identification of known miRNAs and analysis of differential expression

The expression profile of miRNA in the CIM and the RKD2 callus was analyzed by small RNA sequencing. Small RNA reads were mapped against the TAIR10 *Arabidopsis thaliana* reference genome with Butter (Axtell 2014), which is a script that depends upon two other freely available tools: samtools (Li et al. 2009) and bowtie (Langmead et al. 2009).

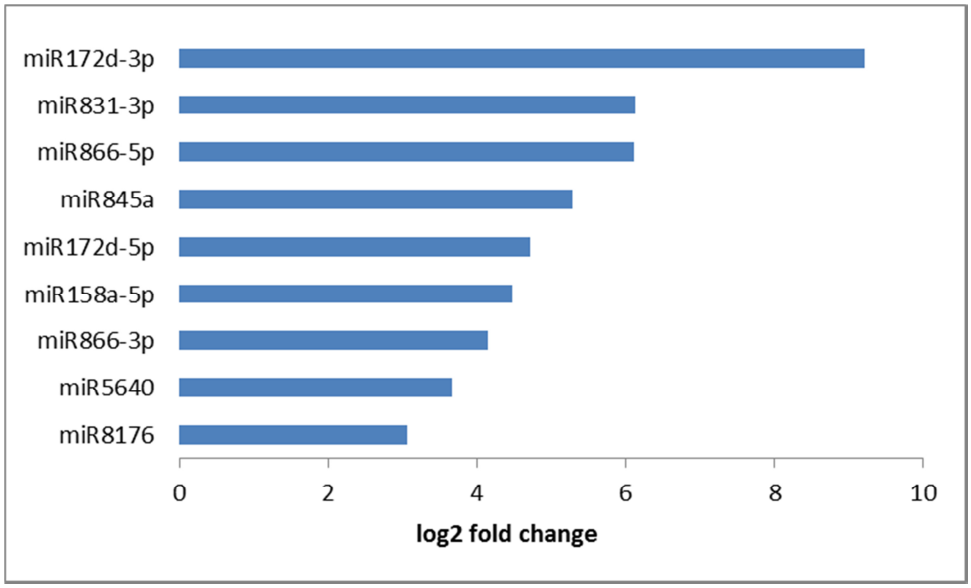


Figure 52: Differentially expressed known miRNAs in RKD2-induced callus, downregulated in CIM callus. X-axis represents the log2 fold change (RKD2-induced vs. CIM). 9 known miRNAs are upregulated with a log2 fold change of at least 2 in the RKD2 callus. All up- and downregulated known miRNAs are shown in Figure 60.

Differential expression analysis revealed clear differences between the miRNA profiles of the RKD2-induced cell line and the CIM callus. As mentioned above, the RKD2 induced cell line shows a reduced level of expressed miRNAs. In total, 96 known miRNAs are differentially regulated in the RKD2-induced callus with a log2 fold change of a least 2. Only 9 miRNAs are upregulated in the RKD2 callus (Figure 52) and 87 are upregulated in the CIM callus (Figure 53).

Micro RNA 172d is the most abundant one in the RKD2-induced callus (Figure 51) with a 9.221 log2 fold change for miR172d-3p and 4.727 log2 fold change for miR172-5p, compared to the CIM callus. Member of the miR172 family are known to regulate the mRNA abundance and/or translation of the plant-specific transcription factor genes belonging to the APETALA2 family (reviewed in Zhu and Hellwell 2011) and thus play a role in controlling flowering and juvenile to adult phase transition (Zhu et al. 2011).

Another miRNA upregulated in the RKD2-induced cell line is miR831-3p (Figure 51), which was previously reported to be present in pistils (Rajagopalan et al. 2006). In the CIM callus, there are miRNAs, which cannot be found in the small RNASeq data from the RKD2 callus. The most upregulated miRNA in the CIM callus is miR165b (Figure 53).

Micro RNAs of the miR166/165 family have been known to regulate meristem development and organ polarity by targeting members of the HD-ZIP III family of transcription factors: PHABULOSA (PHB), PHAVOLUTA (PHV), REVOLUTA (REV), CORONA (CNA), and ATHB8 (Zhu et al. 2007).

Highly upregulated in the CIM callus are also members of the miR156/157 family. For example miR157b-3p is 9.766 log2 fold downregulated in RKD2-induced callus (Table 10). Members of the miR156/157 family repress transcription factors of the SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) family and are also involved in matters of phase transition and regulation of flowering time (Huijser and Schmid 2011).

Another interesting miRNA in the CIM callus is mir5653 (9.744 log2 fold down in RKD2-induced callus), which is reported to be expressed in the meristematic zone of roots, with a specific expression pattern in different cells types (Breakfield et al. 2012).

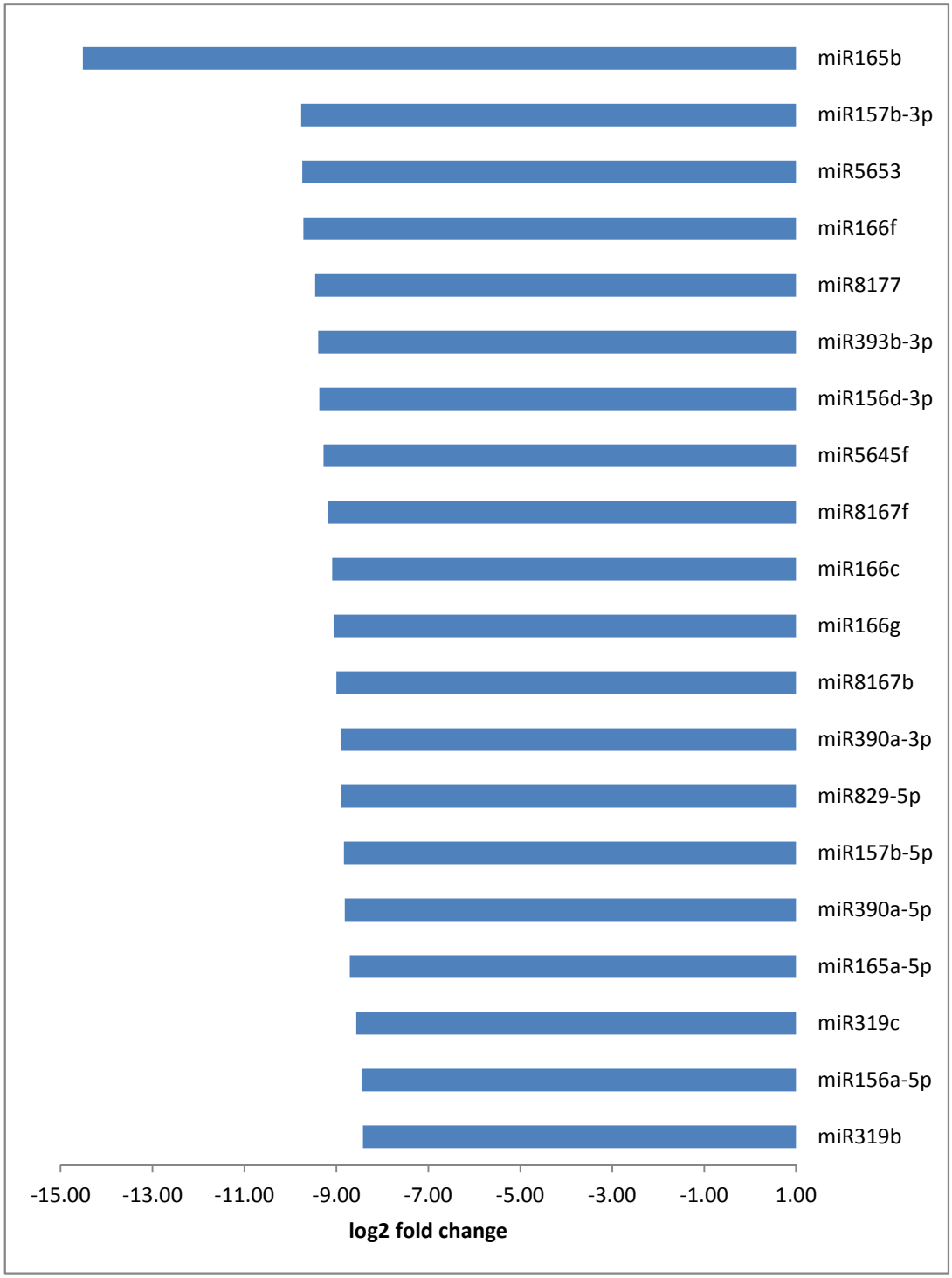


Figure 53: Differentially expressed miRNAs in CIM callus, down regulated in RKD2-induced callus (top 20 out of 87). X-axis represents the log2 fold change (RKD2-induced vs. CIM). All up- and downregulated known miRNAs are shown in Figure 60. 87 known miRNAs are upregulated with a log2 fold change of at least 2 in the RKD2 callus.

5.8.13 Identification of miRNA targets

As described in the material and methods section, potential targets of the miRNAs identified by small RNASeq were either obtained from the Carrington file (Allen et al. 2005) or predicted using the psTarget algorithm. As mentioned above, 96 miRNAs were differentially regulated in the RKD2-induced callus with a log2 fold change of a least 2, with 9 upregulated and 87 down regulated miRNAs. Table 8 shows the RKD2 upregulated miRs and their targets and the top 30 downregulated miRNAs. Some of the predicted targets have been identified in previous studies. While the majority of miRNAs targets protein coding genes, we found several miRNAs targeting transposable element genes and three miRNAs targeting TAS loci.

In Table 9, already known targets of miRNA 172 and the miRNA families 165/166 and 156/157 are shown. As described above, Ath-mir172d-3p has been predicted to target several members of the Apetala2-like transcription family members, so down regulation of AP2-family member transcripts in the RKD2-induced callus could be expected. However, our transcriptome data do not show a general repression of AP2-like transcripts. Only the transcripts of TOE3 and SMZ are significantly down regulated in the RKD2-induced callus (Table 9).

Table 8: Differential expressed known miRNAs and their predicted targets. Prediction of miRNAs was conducted using psRNAtarget (Dai and Zhao 2011). Experimentally validated targets are marked with an asterisk. The log2 fold change (RDK2 vs. CIM) is shown. In the column "target gene family" a family of target genes is indicated if more than member of that family is targeted. If only single genes are targeted, the target with the lowest expectancy value, computed by psRNAtarget is shown together with its AGI identifier.

miRNA	log2 FC	Target gene family
up in RKD2-induced callus		
ath-miR172d-3p	9.22	Apetala2-like transcription factors*
ath-miR831-3p	6.12	Methyl-CPG-binding domain *(AT4G00416)*
ath-miR866-5p	6.10	Nuclear factor Y, subunit A10 *(AT5G06510)*
ath-miR845a	5.28	Transposable element (Copia and Gypsy superfamilies)
ath-miR172d-5p	4.72	Transposable element Gypsy/ATHILA4C (*AT4G05582*)
ath-miR158a-5p	4.47	Chaperone DnaJ-domain (AT5G22080)
ath-miR866-3p	4.15	Cysteine-rich Receptor-like protein kinase 28 (AT4G21400)
ath-miR5640	3.67	Protein of unknown function DUF3741 (AT4G28760)
ath-miR8176	3.06	*AT2G22320*

miRNA	log2 FC	Target gene family
down in RKD2-induced callus		
ath-miR165b	-14,51	HD-ZIPIII transcription factors*
ath-miR157b-3p	-9.77	Squamosa-promoter binding protein-like*
ath-miR5653	-9.75	ABA insensitive 4 *(AT2G40220)*
ath-miR166f	-9.72	HD-ZIPIII transcription factors*
ath-miR8177	-9.46	indole-3-acetic acid 7 (IAA7) *(AT3G23050)*
ath-miR393b-3p	-9.39	Con A-like lectin protein kinase fam. prot. *(AT3G59740)*
ath-miR156d-3p	-9.37	Squamosa-promoter binding protein-like*
ath-miR5645f	-9.28	Transposable Element GYPSY (AT5G30247)
ath-miR8167f	-9.19	Cytochrome P450, family 705, subfamily A (3 members)
ath-miR166c	-9.09	HD-ZIPIII transcription factors*
ath-miR166g	-9.06	HD-ZIPIII transcription factors*
ath-miR8167b	-9.00	Cytochrome P450, family 705, subfamily A (3 members)
ath-miR390a-3p	-8.91	ATTAS3*
ath-miR829-5p	-8.90	Protein phosphatase 2C family protein (AT3G06270)
ath-miR157b-5p	-8.84	Squamosa-promoter binding protein-like*
ath-miR390a-5p	-8.82	ATTAS3*
ath-miR165a-5p	-8.71	HD-ZIPIII transcription factors*
ath-miR319c	-8.56	TCP domain protein 10 *(AT2G31070)*
ath-miR156a-5p	-8.46	Squamosa-promoter binding protein-like*
ath-miR319b	-8.42	TCP domain protein 10 *(AT2G31070)*
ath-miR393b-5p	-8.29	auxin signaling F-box 2 *(AT3G26810)*
ath-miR165a-3p	-8.26	HD-ZIPIII transcription factors*
ath-miR866-3p	-8.25	Other RNA *(AT4G13495)*
ath-miR8167b	-8.19	Cytochrome P450, family 705, subfamily A (3 members)
ath-miR319a	-8.19	TCP domain protein 10 *(AT2G31070)*
ath-miR157c-5p	-7.46	Squamosa-promoter binding protein-like*
ath-miR829-3p	-7.26	Transposable element gene GYPSY *(AT1G43060)*
ath-miR8167b	-7.18	Cytochrome P450, family 705, subfamily A, polypeptide
ath-miR822-5p	-7.15	Cysteine/Histidine-rich C1 domain family (9 members)
ath-miR824-5p	-7.03	Agamous-like 16 *(AT3G57230)*
ath-miR169f-5p	-6.97	nuclear factor Y, subunit A10 *(AT5G06510)*

In the CIM callus, members of the miR165/166 family are present in high amounts. It has been shown in several studies, that these miRNAs control HD-ZIP III transcripts through AGO1 and AGO10 in Arabidopsis (Zhong and Ye, 2007).

Table 9: Expression of known miRNA targets in RKD2-induced and CIM callus. Normalized count numbers of previously identified target gene families in RKD2-induced and CIM-callus and the corresponding log2 fold change (RKD2/CIM) is shown.

AGI identifier	miRNA	Target family	Descrip-tion	CIM (counts)	RKD2 (counts)	log2 (RKD2/CIM)
AT2G28550	172d-3p	Apetala 2-like transcription factors	TOE1	1997.07	1756.97	-0.18
AT5G60120			TOE2	3723.92	1472.04	-1.34
AT5G67180			TOE3	437.22	58.94	-2.89
AT4G36920			AP2	1793.54	2022.88	0.17
AT3G54990			SMZ	572.72	30.10	-4.25
AT2G39250			SNZ	N.A.	N.A.	N.A.
AT1G30490	165/166	HD-ZIPIII transcription factors	PHV	9098.39	4244.55	-1.1
AT1G52150			ATHB-15	11160.66	15620.30	0.49
AT5G60690			REV	3872.47	7907.69	1.03
AT2G34710			PHB	9242.58	11618.04	0.33
AT2G47070	156/157	Squamosa-promoter binding protein-like	SPL1	6302.15	3644.81	-0.79
AT1G54150			SPL2	1057.75	2283.11	1.11
AT2G33810			SPL3	52.41	28.67	-0.87
AT1G53160			SPL4	N.A.	N.A.	N.A.
AT3G15270			SPL5	15.99	30.05	-0.91
AT1G69170			SPL6	317.34	552.52	0.8
AT5G18830			SPL7	1806.69	1159.37	-0.64
AT1G02065			SPL8	0.85	12.57	3.88
AT2G42200			SPL9	115.18	7.66	-3.91
AT1G27370			SPL10	463.60	768.94	0.73
AT1G27360			SPL11	200.98	256.16	0.35
AT3G60030			SPL12	6761.64	3357.46	-1.01
AT5G50570			SPL13A	768.79	351.27	-1.13
AT5G50670			SPL13B	739.50	342.60	-1.11
AT1G20980			SPL14	4849.94	3405.74	-0.51
AT3G57920			SPL15	1.99	436.43	7.78

Transcripts of all four known family members are present in both cell types and although there are some differences in the expression levels the highest fold change of two compared genes do not exceed 2. A look at the miR156/157 targeted SPL family reveals the differential expression of just one gene namely SPL15, which has been discovered to play a role in controlling the juvenile-to-adult phase transition (Schwarz et al. 2008).

Table 10 shows the correlation between differential expression of known miRNA and their predicted targets. The strongest down regulated miRNA in the RKD2-induced cell line is mIR5653 (Table 12).

Table 10: Correlation between differential expression of known miRNAs and their predicted targets. Shown are combinations of miRNA and target with at least log2 fold 2 upregulated miRNA and at least log2 fold 2 downregulated mRNA or downregulated miRNA and upregulated mRNA. Counts were normalized to the size of the libraries (Chapter 4.6).

miRNA	log2 FC miRNA	target accession	(putative) target gene / gene family	CIM (counts)	RDK2 (counts)	log2 FC target
up in RKD2-induced egg cell-like cell line						
miR831-3p	6.12	AT1G63100	*GRAS transcription factor*	436.34	41.07	-3.40
miR866-5p	6.11	AT5G06510	*NF-YA10*	433.54	144.25	-2.96
miR172d-5p	4.73	AT3G23810	*SAHH2*	7231.88	891.54	-3.02
miR158a-5p	4.47	AT1G79680	*WAKL10*	115.36	7.84	-3.88
miR866-3p	4.15	AT1G43886	*copia-like retrotransposon family*	613.44	108.44	-2.50
down in RKD2-induced egg cell-like cell line						
miR5653	-9.74	AT2G40220	*ABA insensitive 4*	3.62	5356.26	10.53
miR390a-3p	-8.91	AT3G17185	*TASIR-ARF*	643.69	14068.87	4.45
miR829-5p	-8.9	AT1G73360	*EDT1*	127.66	1853.64	3.86
miR390a-5p	-8.82	AT3G17185	*TASIR-ARF*	643.69	14068.87	4.45
miR319c	-8.57	AT3G15030	*TCP4*	83.81	1721.07	4.36
miR319b	-8.42	AT3G15030	*TCP4*	83.81	1721.07	4.36
miR863-3p	-8.25	AT3G27620	*AOX1C*	57.00	1576.84	4.79
miR319a	-8.19	AT3G15030	*TCP4*	83.81	1721.07	4.36
miR169f-3p	-6.98	AT3G05690	*HAP2 transcription factor*	1722.26	8079.72	2.23
miR159b-3p	-6.97	AT3G15030	*TCP4*	83.81	1721.07	4.36
miR5635d	-6.69	AT1G05020	*ENTH/ANTH/VHS superfamily*	19.93	127.75	2.68
miR396a-3p	-6.52	AT2G45480	*Growth regulating factor 9*	40.53	2321.59	5.85
miR5635d	-6.06	AT1G05020	*ENTH/ANTH/VHS superfamily*	19.93	127.75	2.68
miR390b-5p	-6.01	AT3G17185	*TASIR-ARF*	643.69	14068.87	4.45
miR5635a	-5.65	AT1G05020	*ENTH/ANTH/VHS superfamily*	83.81	127.75	2.68
miR396a-5p	-5.11	AT2G45480	*Growth regulating factor 9*	40.53	127.75	5.84

miRNA	log2 FC miRNA	target accession	(putative) target gene / gene family	CIM (counts)	RDK2 (counts)	log2 FC target
miR398b-5p	-3.30	AT1G70440	*SRO3*	87.09	1943.43	4.48
miR159c	-3.23	AT2G32460	*MYB101*	13.49	228.17	4.08
miR396b-5p	-2.93	AT2G45480	*Growth regulating factor 9*	40.53	127.75	5.85
miR865-5p	-2.52	AT3G51895	*Sulfate Transporter 3*	6.25	54.69	3.13
miR781b	-2.49	AT5G23480	*SWIB/MDM2 domain*	7.10	1250.18	7.46
miR781a	-2.34	AT5G23480	*SWIB/MDM2 domain*	7.10	1250.18	7.46
miR846-3p	-2.12	AT1G52070	*Mannose-binding lectin superfamily*	2.58	21.58	3.05

Among the predicted targets of miR5653, the transcript of only one gene is significantly upregulated in the RKD2-induced cell line. It is a member of the APETALA2/ ETHYLENE-RESPONSIVE ELEMENT BINDING FACTOR (AP2/EREBP) family and is called ABA INSENSITIVE 4 (ABI4). It is involved in abscisic acid (ABA) signal transduction, it responds to nutrient status, plays a role in salt tolerance and seed germination (Gregorio et al. 2014) and takes part in retrograde signaling from chloroplast and mitochondria (Foyer et al. 2012).

5.8.14 Novel miRNAs and their targets

The prediction of novel miRNA was performed in collaboration with Dr. Julia Engelmann (Institute for Functional Genomics, University Regensburg), applying the miR-PREFeR software. The details of novel MIR genes including location and sequence are listed in the digital supplement. Small RNA sequence with no match in the miRNA databases were at first described as novel. At first, 32 previously undescribed different miRNAs were discovered. However, a manual rechecking of the miRNA sequences revealed that some of our new miRNAs corresponded to stem loop sequences of known miRNAs, but differ from the mature miRNAs derived from those stem loops. An example is miRNA_16, which is strongly upregulated in the RKD2-induced egg cell-like callus. It fits exactly to a part of the stem-loop of miR158b, but differs from its mature sequence (Figure 53). In total, 20 novel miRNAs could be linked to known precursors (Table 11).

```
mir158a-5p          --------------------CUUUGUCUACAAUUUUGGAAA--------------------
ath-MIR158a         ACACGUCAUCUCUGUGCUUCUUUGUCUACAAUUUUGGAAAAAGUGAUGACGCCAUUGCUC
athmiR158bmat       ------------------------------------------------------------
mir_16              --------------------CUUUGUCUACACUUUUGGAAA--------------------

mir158a-5p          ----------------------------------------
ath-MIR158a         UUUCCCAAAUGUAGACAAAGCAAUACCGUGAUGAUGUCGU
ath-miR158bmat      --CCCCAAAUGUAGACAAAGCA------------------
mir_16              UUCCCCAAAUGUAGACAAAGCAAUACCGUGAU--------
```

Figure 54: Novel miR_16 is processed from the known ath_MIR158b precursor.
The alignment shows sequences from novel miR_16, miR158a-5p and miR158b, including their precursors ath-MIR158a and ath-MIR158b. Novel miR_16 is processed from the 5p region of ath-MIR158b and therefore represents miR158b-5p. It differs in one nucleotide from miR158a-5p (yellow, C). Sequences were retrieved from miRBase (Kozamara and Griffith-Jones, 2014).

Target predictions for novel miRNAs revealed a variety of targets. Target predictions are based on the psTarget software (Dai and Zhao 2011). In addition to this, our transcriptome data was analyzed with respect to their differential expression in CIM and RKD2-induced callus. Only five miRNAs were upregulated in the RKD2-induced callus, of which novel miRNA 25-5p is most upregulated in the RKD2-induced callus. Interestingly, none of the predicted targets of this miRNA shows significant downregulation in RKD2 callus, instead of these two mRNAs seem to be upregulated, namely the GDSL-motif lipase 2 and another unknown protein (AT1G45165).

The novel miR_38 is predicted to target a gene of DNA-binding storekeeper protein-related transcriptional regulator family (AT4G00390), which shows high expression levels in the RKD2 callus but is absent in the CIM callus. Co-expression data from the atted.jp database (Obayashi et al. 2014) computes a number of genes which are upregulated in RKD2 callus to be co-expressed with this gene, including *AGO5*, *AGO9*, *Chr34* and the *POL V* subunit *NRPE5-L*. Also for the novel miRNAs, the mRNA expression data has been checked for correlation between the presence of a certain micro RNAs and the transcript level of its predicted target.

In total, ten novel miRNAs were found whose increase of expression was accompanied by a decrease of a predicted target mRNA. Two of those novel micro RNAs are upregulated in the RKD2-induced egg cell-like callus.

Table 11: Differentially expressed novel miRNAs from known or novel MIR precursors and their predicted targets, identified by psRNATarget. Shown are novel miRNAs, their precursor (if the precursors is a miRNA annotated in miRBase), the log2 fold change (RKD2 vs. CIM), and the putative target gene. The target with the lowest expectancy value, computed by psRNAtarget is shown together with its AGI identifier.

Novel miRs	Precursor	log2 FC	(putative) target gene / gene family
up in RKD2-induced egg cell-like cell line			
miR_25-5p	MIR777	6.87	*Transposable element gene (AT3G33181)*
miR_16-5p	MIR158b	6.01	*RNA-dependent RNA polymerase 5 (AT2G19930)*
miR_78-5p	MIR845a	4.12	*FtsH extracellular protease family (AT4G29130)*
miR_95	NOVEL	1.58	No putative targets found
miR_26-5p	MIR775	0.97	*SLAC1 homologue 1 (AT1G62280)*
down in RKD2-induced egg cell-like cell line			
miR_57-3p	NOVEL	- 8.53	*LSH10 (AT2G42610)*
miR_38-5p	NOVEL	- 8.16	*DNA-binding storekeeper protein-related transcriptional regulator (AT4G00390)*
miR_104-5p	MIR319b	-7.41	*CHR5 (AT2G13370)*
miR_79-5p	MIR319c	-7.23	*Calmodulin 2 (AT2G41110)*
miR_58-5p	NOVEL	-7.04	basic helix-loop-helix (bHLH) DNA-binding *(AT3G21330)*
miR_90-3p	MIR156e	-6.88	PATATIN-LIKE PROTEIN 9 *(AT3G63200)*
miR_64-5p	MIR827	-6.68	*HR-like lesion-inducing protein-related (AT3G23190)*
miR_109-5p	NOVEL	- 6.45	*Tubby like protein 3 (AT2G47900)*
miR_38-3p	NOVEL	- 6.42	*DHHC-type zinc finger family protein (AT4G24630)*
miR_109-3p	NOVEL	-6.39	*Arabidopsis Mediator Components 14 (AT3G04740)*
miR_73-3p	MIR850	-6.24	unknown protein (*AT5G13660*)
miR_57-5p	NOVEL	-5.78	*Pseudogene (putative DNA methyltransferase) (ATAT3G22555)*
miR_105-5p	MIR166f	-5.60	No putative targets found
miR_87	NOVEL	-5.35	*FAD-binding Berberine family protein (AT4G20800)*
miR_54-5p	NOVEL	-4.69	*YAO (Transducin/WD40 repeat-like) AT4G05410)*
miR_54-3p	NOVEL	-4.65	*CHUP1 (Hydroxyproline-rich glycoprotein family protein) (AT3G2569.)*
miR_43-3p	MIR778	-4.45	*ACT domain repeat 3 (AT1G76990)*
miR_8-3p	MIR781a	-3.48	*Protein of unknown function (DUF1350)* (AT3G43540.)
miR_9-3p	MIR400	-3.26	*CYP94D2 cytochrome P450 (AT3G56630)*
miR_88-5p	MIR8184	-3.24	*SPT5-Like (AT5G04290)*
miR_67-3p	MIR167b	-2.61	No putative targets found
miR_17-3p	MIR859	-2.40	*AT2G46200*
miR_65-3p	MIR1888b	-2.35	*DNAse I-like superfamily protein (AT3G48425)*
miR_59-3p	MIR5651	-1.91	No putative targets found
miR_19-5p	MIR842	-1.66	*VQ motif-containing protein (AT5G65170)*
miR_69-5p	MIR447a	-1.13	*Dof-type zinc finger DNA-binding family protein(AT1G28310)*

Table 12: Expression levels of selected putative target genes of differential expressed novel miRNAs. Shown are novel miRNAs and their putative target. miRNA/target pairs are shown if the up regulation of the miRNA and the down regulation of target corresponds to a log2 fold change of at least 2.

Novel miRs	Precursor	log2 FC	(putative) target gene	Locus (AGI)	log2 FC of target gene
up in RKD2-induced egg cell-like cell line					
miR_25-5p	MIR777	6.87	*Flavonol synthase 1*	AT5G08640	-3.73
miR_16-5p	MIR158b	6.01	*RNA-dependent RNA polymerase 5*	AT2G19930	-2.92
down in RKD2-induced egg cell-like cell line					
miR_57-3p	NOVEL	- 8.53	*Light sensitive hypocotyls 10*	AT2G42610	3.36
miR_38-5p	NOVEL	- 8.16	*DNA-binding storekeeper protein-related transcriptional regulator*	AT4G00390	9.11
miR_79-5p	MIR319a	-7.23	*Calmodulin 2*	AT2G41110	3.19
miR_58_5p	NOVEL	-7.04	*Basic helix-loop-helix (bHLH) DNA-binding*	AT3G21330	2.65
miR_90-3p	MIR156e	-6.88	*Patatin-like protein 9*	AT3G63200	3.36
miR_64-5p	MIR827	-6.68	*Glycosyl hydrolase 9C3*	AT4G11050	6.18
miR_73-3p	MIR850	-6.24	*Protein of unknown function (DUF1645)*	AT1G23710	3.08
miR_8-3p	MIR781a	-3.48	*Protein of unknown function (DUF626)*	AT3G44770	4.58
miR_9-3p	MIR400	-3.26	*CYP94D2 cytochrome P450*	AT3G56630	3.14

5.8.15 Validation of miRNA expression by Northern Blot analysis

In order to validate the differential expression of certain miRNAs in our callus tissues, selected miRNAs were analyzed by Northern Blot (Figure 55). The microRNAs mir390a and mir5653 are, according to the RNAseq data, only present in the CIM callus and this result can be confirmed by Northern Blot. Novel miRNA_16 (mir158b-5p) can be detected in the RKD2 callus both in RNASeq data (Table 13) and Northern Blot. MicroRNA 172d shows a strong presence in the RKD2 callus, but a faint signal can also be detected in the auxin-induced CIM callus. MiR 172d has 721.14 reads in the RKD2 callus but only 1.21 in CIM callus.

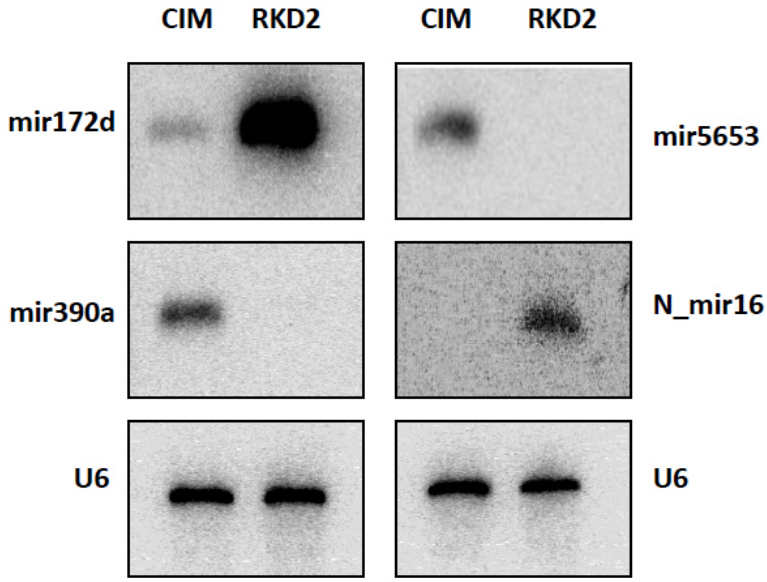


Figure 55: Validation of micro RNA by Northern Blot analysis.
The presence of miRNAs in RKD2 and CIM callus has been examined by Northern Blot as described in Chapter 4.5.3. *U6* small nuclear RNA (*snRNA*) was used as a loading control. 12.5 µg of total RNA was loaded per lane.

5.8.16 Transposable elements in RKD2-induced callus and CIM callus

In our mRNASeq data, reads for 315 different predicted transposable element (TE) transcripts were found. Of these, 65 were differentially regulated. 32 transcripts were upregulated in the RKD2-induced callus while 33 transcripts were upregulated in the CIM callus. Expression data of TE mRNAs was analyzed in RKD2-induced and CIM callus (Sense reads) and then classified into high and low expressed TEs. For high expression, a count value of at least 30 was assumed. Counts for antisense sRNA derived from TEs were then statistically analyzed for TE mRNAs with high and low expression level (for RKD2-induced and CIM callus). Using this statistic data, box plots were created in which the thick black bar in the middle of the box represent the median of TE antisense sRNA expression values (Figure 56). In general, TEs with a low expression level show high expression of antisense sRNA, this is observed in RKD2-induced and CIM callus. This indicates that least some transposable elements are downregulated by small RNA pathways in both tissues.

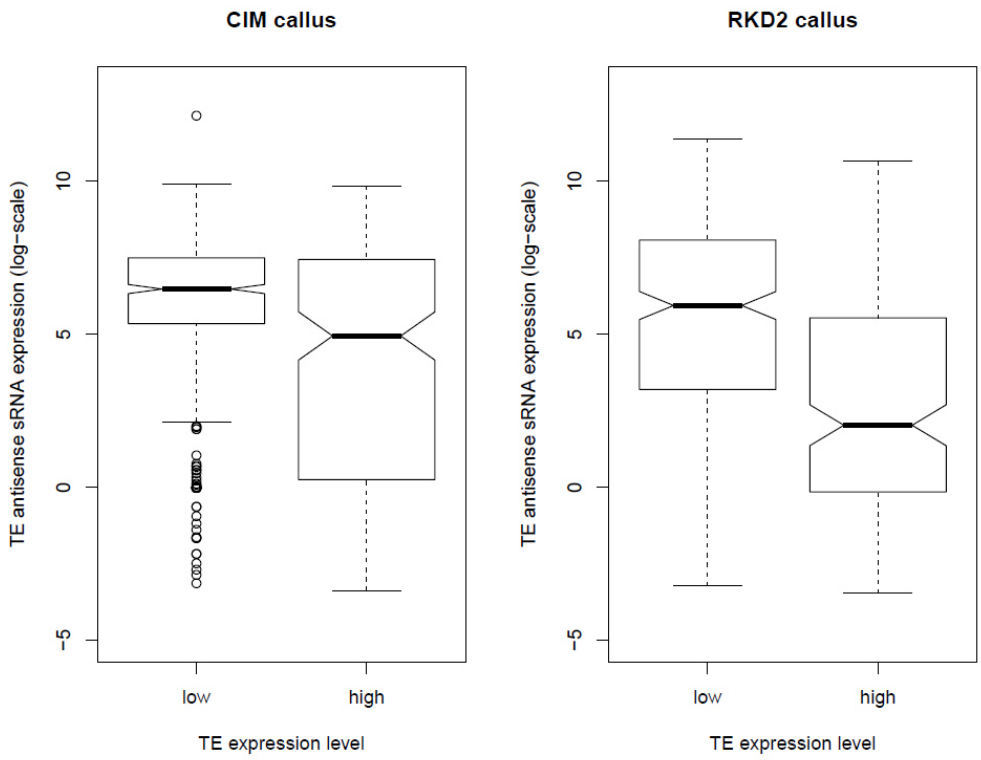


Figure 56: Notched boxplot of transposable element sRNA expression levels in relation to the TE expression levels in the RKD2-induced callus and the CIM callus**.** Boxplots are representing numbers of the TE antisense sRNA reads with the median depicted as thick black bar. Plots are shown for high and low expressed mRNAs in CIM (left) and RKD2 callus (right).

In the CIM callus a high expression of TE antisense sRNA is accompanied by high expression of TE mRNAs, meaning that in spite of the generation of sRNA from TEs the expression of some TEs is not effectively down regulated by small RNA pathways. In the RKD2-induced callus highly expressed TEs have relatively few antisense RNA reads from TEs, meaning that those TEs are not used for siRNA production.

Small RNA silencing seems to occur in RKD2 and CIM callus. The difference between the two tissues is that in the CIM callus production of sRNAs from TE is not enough to silence those elements. Four differentially regulated TE mRNAs, all down regulated in RKD2-induced callus, were investigated in detail for corresponding TE antisense small RNA reads. It turned out that all TEs are represented with a considerable number of sRNA reads (Table 13). This leads to the conclusion, that those TEs are least partly downregulated by small RNA pathways.

Table 13: Downregulated TEs in the RKD2-induced callus and corresponding TE antisense small RNAs. Shown are normalized counts of transposable elements with downregulated mRNAs in the RKD2-induced cell line and their respective sRNA counts.

		sRNA (counts)			mRNA (counts)		
AGI Identifier	**TE family**	**CIM**	**RKD2**	**log2 FC**	**CIM**	**RKD2**	**log2 FC**
AT5G19097	ATCOPIA89	363.16	914.67	1.33	1127.21	10.47	-6.75
AT4G27210	ATCOPIA8B	582.61	953.76	0.71	53.75	0.65	-6.38
AT1G45120	ATCOPIA1	2.67	60.02	4.49	527.35	175.21	-1.59
AT1G47950	copia-like retrotransposon	684.62	1312.02	0.94	66.59	4.45	-3.90

Comparing the TE-derived sRNAs, a clear difference concerning TE classes can be observed. In the CIM callus, sRNA from all superfamilies of transposable elements can be observed, whereas in the egg cell-like callus, one superfamily was dominant: The LTR/Gypsy superfamily comprises 61% of TE-derived small RNA reads in the RKD2-induced callus, containing members of the ATHILA, ATLANTYS and ATGP families (Figure 57).

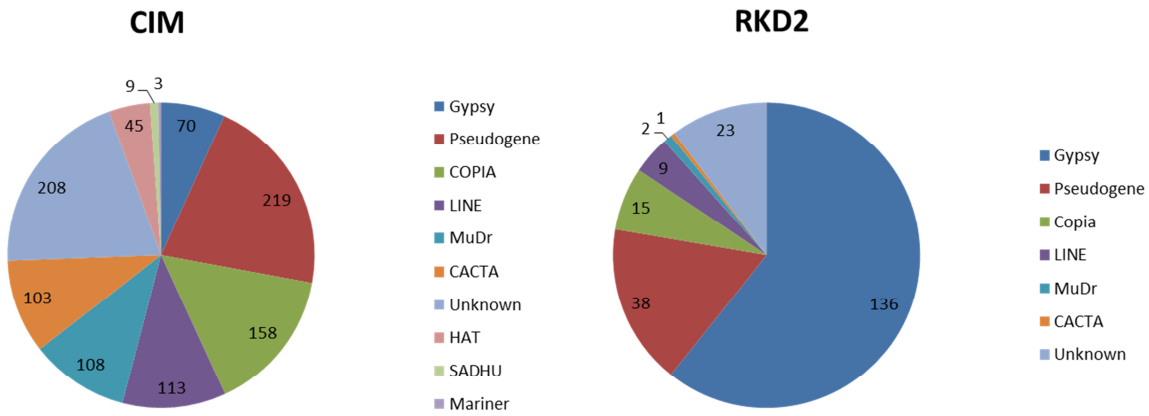


Figure 57: Absolute numbers of small RNAs derived from different classes of transposable elements in CIM and RKD2-induced callus. Small RNA reads in CIM and RKD2 callus were analyzed. Features of small RNAs was retrieved from EnsemblPlants (Chapter 2.6.1) In the CIM callus, small RNAs derived from transposable elements (TEs) are distributed evenly the different classes of TEs. In the RKD2-induced callus the Gypsy superfamily of retrotransposons is dominant.

In contrast to this, messenger RNAs from Gypsy-transposons are absent in the RKD2 callus, no reads can be found in the transcriptome data, suggesting that especially this class of transposable elements is down regulated by sRNA pathways in the RKD2-induced cell line.

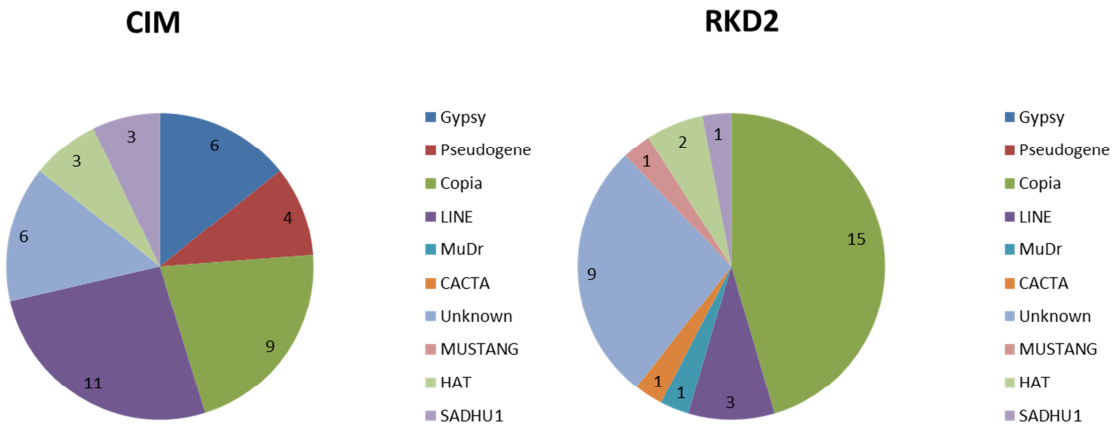


Figure 58: Absolute numbers of messenger RNAs derived from transposable elements in CIM and RKD2-induced callus. mRNASeq data from CIM and RKD2-induced callus was analyzed with respect to the abundance of mRNAs of transposable elements. In the CIM callus most TE transcripts belong to the LINE and COPIA superfamilies of retrotransposons. In the RKD2-induced callus the COPIA family alone is dominant.

Altogether, mRNAs of 33 TEs were discovered in the RKD2-induced cell line. Fifteen of those TEs belong to the COPIA superfamily of transposons (Figure 58). RKD2-induced and CIM cell cultures thus show clear differences concerning the composition of actively transcribed TE

6 DISCUSSION

Since the discovery of RNA interference (RNAi), many processes in the plant life cycle have been found to be influenced by this natural mechanism for sequence-specific gene silencing. In mammals, *de novo* DNA methylation directed by small noncoding RNAs has been shown to occur during both gametogenesis and early embryonic development and to be crucial to maintain genomic integrity through silencing transposable elements (Kawashima and Berger 2014). The strong expression of genes encoding components of small RNA pathways in plant reproductive organs implicate also an important role in plant reproduction, especially in gamete formation and the acquisition of gametic cell fate (Schmid et al. 2005).

Concerning small noncoding RNAs in the plant male germline, there have been studies that microRNAs (miRNAs) and short interfering RNAs (siRNAs) are present in both the vegetative cells and sperm cells of the pollen (reviewed in He et al. 2015). A clear difference is observed comparing small noncoding RNA species in the vegetative cell and the sperm cells. In the vegetative cell transposable element (TE)-derived 21 nt long siRNAs accumulate and are thought to move to the sperm cells to silence TE expression, thereby maintaining genomic integrity in the male gametes (Slotkin et al. 2009). However, it may also be possible that siRNAs, produced by microspores, are inherited upon differentiation into sperm cells, in which they target gene silencing through the RNA-directed DNA methylation (RdDM) pathway (Kawashima and Berger, 2014). Furthermore, after fertilization, siRNAs might be inherited from sperm cells to the zygote, in which they may have a role in reprogramming events during embryogenesis.

A similar situation as in sperm cells has been suggested for the female gametophyte, based on the expression pattern of genes encoding TEs and DNA methyltransferases in egg cell and central cell (Ibarra et al. 2012). In the female gametophyte of Arabidopsis certain TEs are repressed in the egg cell but depressed in the central cell (Olmedo-Monfil 2010). Moreover, in the developing ovule of Arabidopsis, AGO9 and de novo methyltransferases interact with TE-derived siRNAs, whose activity is required to silence TEs (Duran-Figueroa and Vielle-Calzada 2010, Olmedo-Monfil et al. 2010, Singh et al. 2011). Nevertheless, little is known about the presence and composition of small RNA species in the different cell types of the female gametophyte. Small RNA profiling of any cell within the female gametophyte has

not been performed so far, primarily due to the inaccessibility of the female gametophyte, which is deeply embedded within the female reproductive tissues of the sporophyte.

6.1 The role of AGO8 and AGO9 in female gametogenesis

Although Arabidopsis *AGO8* is regarded as a pseudogene (Takeda et al. 2008) *AGO8* transcripts appear to be abundant in the ovule. Reporter studies demonstrated that the *AGO8* promoter is specifically active in only one cell type of the mature ovule, which is the egg cell. The detailed sequence analysis of *AGO8* transcripts revealed that three different transcript variants exist. However, none of them is functional because none of them is free of alternative splicing events, which all lead to premature termination signals in the coding sequence. Homozygous *ago8* T-DNA insertion lines did not exhibit obvious phenotypes (Wartlsteiner 2010). Transgenic *AGO8* promoter-reporter lines showed egg cell specific promoter activity but expression of *AGO8* fusion proteins under the control of the constitutive CaMV35S yielded no GFP-signal. These facts confirm the common opinion that *AGO8* is a pseudogene that arose from recent gene duplication with the *AGO9* sequence as template and has not acquired a new function.

AGO9 was assigned a function in female gametogenesis in *Arabidopsis thaliana*. It was reported that the AGO9 protein is not expressed in the female gamete lineage but in subepidermal (somatic) nucellus cells of the developing ovule and that AGO9-dependent small RNA silencing restricts the specification of female gametophyte precursors in a non-cell-autonomous manner to the megaspore mother cell (MMC). This was concluded from the observation of an increased percentage of additional MMCs during gametogenesis in *ago9* mutant ovules (Olmedo-Monfil et al. 2010). However, the described phenotype could not be confirmed in this and also in a former work in our lab (Wartlsteiner 2010) using one of the described T-DNA insertion lines (*ago9-1*). Phenotypic differences in developing ovules of the same *ago9* T-DNA insertion line may be due to the higher temperatures that were used to grow Arabidopsis plants (24°C, according to Olmedo-Monfil et al. 2010). This temperature-dependent loss of cell fate restriction in developing ovules lacking AGO9 should be investigated in more detail in the future.

AGO9 is reported to preferentially interact with 24-nucleotide small RNAs derived from TEs and functions in down regulation of repetitive and transposable elements as in homozygous *ago9-1* mutants TEs and repetitive elements are no longer repressed the in

developing gynoecia (Olmedo-Monfil et al. 2010, Durán-Figueroa and Vielle-Calzada 2010). In the vegetative cell of the pollen, down regulation of the chromatin remodeler DDM1 leads to a hypomethylated state of the chromatin followed by epigenetic reactivation of transposable elements (Castel and Martienssen 2013). Despite their reputation as genomic parasites deleterious for the host through insertions into protein-coding genes, TEs can also influence neighboring genes and their regulatory sequences, for example by read through transcription, even at distances as long as 179 kb (Sasaki et al. 2008; Nakanishi et al. 2012a). For abiotic stress-induced maize genes in proximity to TEs it was suggested that stress-responsive gene expression is stimulated by local enhancer activities, provided by the TEs (Makarevitch et al. 2015), thereby contributing to regulatory variation in gene expression. In Arabidopsis pollen, epigenetically reactivated TEs are further processed into specific 21 nt long small RNAs (McCue et al. 2013) which were later termed transposon-derived epigenetically activated small interfering RNAs, shortly easiRNAs (Creasey et al. 2014). TE-derived sRNAs are also present in the RKD2 induced cell line and may also play a role in regulating neighboring genes in an egg cell-like context. The role *AGO9* in those gene regulatory processes in the egg cell remains to be investigated.

6.2 AGO9 protein expression and subcellular localization in ovules before and after fertilization

Former studies concerning AGO9 found this protein to be expressed in the L1 layer of the nucellus and not in the megaspore mother cell by using a polyclonal antibody directed against the N-terminal AGO9 peptide SSRNHAGNDTNDADRK (Olmedo-Monfil et al., 2010). In this work, the expression of AGO9 was investigated using a different peptide antibody, directed against the peptide DSDEPNGSGLPPPC-CONH2 (Havecker et al. 2010) at the N-terminus of AGO9. In these whole mount-immunolocalization experiments AGO9 was omnipresent during early stages FG0 and FG1 and not restricted to the L1 layer of the nucellus, but no protein could be detected in the MMC. During ovule development, AGO9 expression diminished, showing only residual signals in the tips of the integuments and the funiculus. There was no AGO9 detectable in cells of the mature female gametophyte, neither in the synergids nor in the central cell or the egg cell.

In the early stages of embryonic development, AGO9 can be detected by immunolocalization from the 8-cell embryo stage on. In the early heart stage embryo, AGO9

is found in the whole embryo with strongest signals in the protoderm and the region of the shoot apical meristem.

On subcellular level, Olmedo-Monfil et al. (2010) described the protein to be present in cytoplasmic foci of L1 layer nucellus cells. In a later study, a transient AGO9 expression in the nucleus was found (Rodriguez-Leal et al. 2015). However, in this study the presence of AGO9 in cytoplasmic foci was never observed while the nuclear localization of AGO9 could be proved by several approaches: transient GFP-AGO9 expression studies using tobacco BY-2 cells, analysis of stable transformed plant expressing the GFP-AGO9 fusion protein and whole-mount immunolocalization studies all showed clear signals in the nucleus. This is in concordance with the proposed role of AGO9 in small RNA-guided epigenetic modification of the genome that takes place in the nucleus, such as transposon silencing via the RdDM pathway.

AGO9 promoter activity was detected in the megaspore mother cell (MMC) and the functional megaspore (FM) of young ovules and in the egg cell, the central cell and the chalazal region of mature ovule. This stands in contrast to the above findings obtained by AGO9 whole mount-immunolocalization and expression of the GFP-AGO9 fusion protein, where the AGO9 protein is absent from the MMC, the FM, the egg cell and the central cell. Thus AGO9 protein stability is decreased in those cell types, which might play a role in regulating AGO9 activity at a post-translational level. The occurrence of *AGO9* transcript variants in developing ovules suggest a potential role of developmentally regulated alternative splicing in regulating AGO9 activity.

6.3 A cell line with egg cell-like expression profile as a tool to identify new small RNA pathway components in the female germline

Microarray-based gene expression profiling revealed that ectopic expression of the transcription factor RKD2 in transgenic Arabidopsis seedlings results in the formation of de-differentiated callus cells exhibiting an egg cell-like transcription profile (Köszegi, 2011). In this work, RKD2-induced cell lines were generated and used to investigate its transcriptome and small RNA profile by deep sequencing. Expression data of eighteen genes was validated by qRT-PCR. The comparison of our mRNA-Seq data with the microarray data from Köszegi et al. (2011) showed a similar expression pattern for the investigated 99 genes.

Besides *AGO8* and *AGO9*, where promoters were found to be active in the egg cell, egg cell specific promoter activity was confirmed for other genes upregulated in the RKD2-induced cell line, such as *EC1.5* (Sprunck, unpublished) and one *DUF674* gene (T.Hackenberg, unpublished). Furthermore, the RKD2-induced genes *ASPARTYL PROTEASE FAMILY PROTEIN* (*ASP*; *AT1G31450*) and *SUBTILASE 4.13* (*SBT4.13*; *AT5G59120*) are specifically expressed in the egg cell, proven by whole mount in situ hybridization experiments (Bleckmann and Dresselhaus, 2015).

Although the global expression profile of the RKD2-induced cell line mainly resembles that of an egg cell, there are some genes up regulated compared to the CIM callus that are not expressed in the egg cell according to array data. This comprises genes like the homeobox gene *WUSCHEL*, which is needed to keep meristem cells in an undifferentiated state (Laux et al. 1996) and the *ISOPENTENYL TRANSFERASE 7* (*IPT7*), a gene involved in cytokinin biosynthesis. These differences might reflect the fact that the cells of RKD2-induced cell line still possess the ability to perform cell division, which demands a complex regulatory network which involves for example phytohormones or homeobox proteins.

Other genes up regulated in the RKD2-induced callus, cannot be found in microarray data from isolated egg cell because of their absence on the ATH1 gene chip, for example two members of the tumor necrosis factor receptor associated factor (TRAF)-like gene family (*AT2G05420* and *AT1G31390)*. TRAF-proteins associate with membrane receptors and take part in signal transduction in humans (Xie 2014). In Arabidopsis, another member of this family, SEVEN IN ABSENTIA 2 (SINA2) is involved in drought-stress signaling (Bao et al. 2013). Another gene which has not been reported to be expressed in egg cell or other cells of the embryo sac is a methyl-CpG-binding domain protein (*AT4G00416*). This family of proteins can recognize methylated DNA and is conserved in all eukaryotes (Bogdanovic and Venstra 2009). In Arabidopsis, methyl-CpG-binding domain protein 7 (MBD7), takes part in active demethylation of some loci, as it is part of a protein complex which recruits the DNA-methylase DEMETER-LIKE 1 (DML1), which is also known as REPRESSOR OF SILENCING (ROS), to GpG-methylated DNA (Li et al. 2015). These genes might be part of an egg cell-specific regulatory network.

Auxin and Cytokinin-related genes show differential expression in RKD2-induced and CIM callus. The *PINOID* (*PIN*) family of auxin efflux facilitators directs polar auxin transport because of their asymmetric subcellular localizations (Adamowski and Friml 2015). Almost all *PIN* genes are present in both celltypes, but *PIN2* transcript is missing in the RKD2-

induced callus. PIN2 mainly acts in the lateral root tip and is involved in gravitropism (Krecek et al. 2009). *PIN2* presence in the CIM callus might reflect the fact that this cell culture is derived from root explants. The *ISOPENTENYLTRANSFERASE 7* (*IPT7*) gene, which is involved in cytokinin biosynthesis, is strongly upregulated in the RKD2 induced callus. Previous studies (Yanai et al. 2005) revealed that IPT7 can be activated by KNOTTED1-like homeobox transcription factors, like SHOOTMERISTEMLESS (STM), KNAT2 and KNOTTED1 from *Zea mays. STM* is downregulated in the RKD2 callus (-6.98 log2 fold change) but *KNAT2* is upregulated (5.34 log2 fold change). KNAT2 is reported to play a role in regulation of plant inflorescence architecture (Li et al. 2012). *IPT7* is reported to be absent from the egg cell, so its presence in the RKD2 callus might be an indicator for the ability of this cell culture to produce endogenous cytokinin in order to maintain cell division.

Altogether the RKD2-induced cell line provides a means for analyzing cells with egg cell-like character without having to use technically challenging and time-consuming methods like for example egg cell isolation by manual dissection of ovules. What is more, the RKD2-induced cell line established in this work did not cease growth after 6 months, like reported for cell lines generated by Köszegi et al. (2011), but so far it keeps its egg cell-like character and grows continuously, which give access to large amounts of cell material for biochemical approaches.

Small RNA-pathway components, especially Argonaute Proteins, play an important role in female gametophyte development. Interestingly, AGO proteins known to be expressed in nucellus cells surrounding the female gametophyte precursors (this study; Tucker et al. 2012) are highly present in the RKD2 callus. AGO5 and AGO9 proteins are specifically abundant in the RKD2-induced cell line but not in the control callus generated from root explants growing on auxin and cytokinin-supplemented medium, which contains considerable amounts of transcripts from every other *AGO*. This is in contrast to the Arabidopsis egg cell, where AGO5 and AGO9 proteins could not be detected (Tucker et al. 2012, this study), suggesting that the RKD2-induced cell line has not unequivocally obtained egg cell fate but may rather represent a mixture of cells with differentiation profiles similar to both egg cells and accessory cells surrounding the egg cell.

In general, all components of the miRNA, the tasi and the RdDM pathway are expressed in the RKD2 callus. Components of the miRNA pathway are expressed in both cell lines (RKD2-induced and control callus) and no significant differences in expression could be observed, with the exception of a gene encoding cap-binding protein *CBP20*, which is

upregulated in the RKD2 induced callus by a factor of 3 (log2 fold change of 1.61). *CBP20* is not only involved in miRNA processing, but also in mRNA processing and transport (Kierzkowski et al. 2009). Looking at specific components in detail, there are a few genes, which are overrepresented in the RKD2-induced callus. *INVOLVED IN DE NOVO 2* (*IDN2*), encoding a double-stranded RNA-binding protein involved in chromatin remodeling in RdDM (Zhang et al. 2012) is strongly upregulated in the RKD2-induced cell line. This indicates a high level of *de novo* methylation in this cell type. Transcripts encoding DNA methyltransferase DRM2, which is the key enzyme in RdDM, are present in both cell lines. This is also true for DRM3, a catalytically mutated paralog of DRM1, which plays a role in DNA methylation and influences Pol V transcript abundance (Zhong et al. 2015). Nevertheless, DRM3 is needed for Pol V activity in RdDM (Zhong et al. 2015) and regulates the methylation of other loci than the ones regulated by DRM2 (Costa-Nunes et al., 2014). The last member of the *DRM* family, *DRM1,* is exclusively expressed in the RKD2-induced cell line. A former study about DNA methylation dynamics in plant reproduction (Jullien et al. 2012) found the DRM1 protein to only be expressed in the mature egg cell (Figure 58). This may indicate that DRM1 plays a role in *de novo* methylation, but exclusively in the egg cell. The histone deacetlylase *HDA18*, which sets epigenetic marks on histones H3/H4 is also upregulated in the RKD2-induced callus. HDA18 is reported to influence cell fate in roots epidermis via histone acetylation at four kinase genes (Liu et al. 2013). This gene might play a role in creating epigenetic marks which are needed for cell specification.

The gene encoding a POL V NRPE5-like subunit is specifically expressed in the RKD2 callus, and its transcripts can be located in the egg cell, but its function is not yet resolved. The study of Ream et al. (2009) revealed the subunit composition of RNA-polymerases IV and V but only one homologue of *NRPE5* could be identified as part of Pol V, which leaves the function of NRPE5-L in RdDM open. A study from Johnston et al. (2007) using genetic subtraction and comparative profiling by microarray between the *Arabidopsis thaliana* wild type and a sporophytic mutant lacking an embryo sac revealed *NRPE5-L* to be embryo sac specific. It could replace its close homologue NRPE5 in the special context of the egg cell.

In the aforementioned study from Jullien et al. 2012, the DNA methyltransferases responsible for maintenance of symmetric CpG methylation (MET1, MET2a and MET2b) and maintenance of CHG methylation (CMT3) could not be detected as a protein in the egg cell while deep sequencing revealed transcripts to be present.

Transcripts of *MET1* and *CMT3* are highly abundant in the RKD2-induced cell line but it is not known if the corresponding proteins are translated in this cell line. The discrepancy of transcript and protein abundance could be the result of translational repression in the egg cell, or it points to a difference between the egg cell-like callus and the real egg cell.

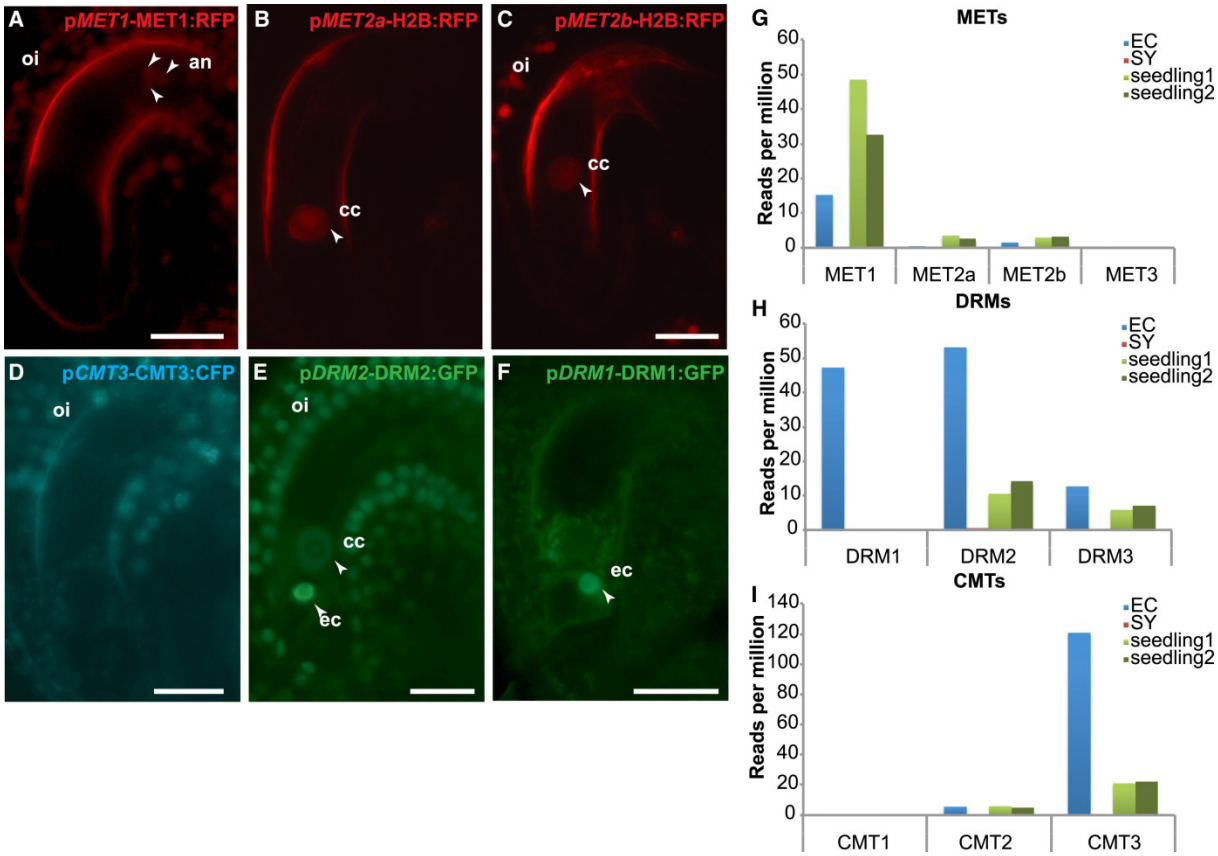


Figure 59. Expression of DNA methyltransferases in mature Arabidopsis ovules.
(A) pMET1-MET1:RFP is detected is the nuclei of the antipodals (an) and the outer integument (oi), but not in the gametes. **(B)** pMET2a-H2B:RFP expression in the central cell nucleus (cc) **(C)** pMET2b-H2B:RFP expression in the nucleus and in the outer layers of the outer integument (oi) **(D)** pCMT3-CMT3:CFP expression is detected in the nuclei of the outer integuments (oi) but not in the gametes. **(E)** pDRM2 is detected in the centrall cell (cc), the egg cell (ec) and in the nuclei of the outer integuments (oi) **(F)** pDRM2 is detected in the the egg cell (ec) **(G–I)** RNA deep-sequencing analysis of transcripts from isolated egg cells (EC), synergid cells (SY), and two seedling samples, used as a sporophytic tissue reference. Transcript abundance is indicated in reads per million for transcripts from genes encoding members of MET1 class (G), CMT3 class (H), and DRM class (I) of DNA methyltransferases. Image taken from Jullien et al. (2012).

6.4 Small RNA profiling reveals differential expression of known and novel miRNAs

The hormone-induced control callus and the RKD2-induced cell line exhibit different small RNA profiles. At first, the RKD2-induced callus has a lower content of small RNAs in general, which can be concluded from the smaller library size generated from this tissue. It is striking that in the RKD2-induced callus, transposable element-derived small RNAs are more prominent than in the hormone-induced control callus, which applies for all size classes of small RNA, from 18 to 24 nucleotides. This underlines the importance of de novo methylation by both canonical and non-canonical RdDM pathway in this cell line and most likely also in the egg cell.

Only a small number of miRNAs are upregulated in the RKD2-induced callus. This includes one completely novel miRNA, 5 novel miRNAs processed from known precursors and 9 known miRNAs annotated in miRBase. A well-studied miRNA highly upregulated in the RKD2-induced cell line is miR172d which is involved regulation of developmental phase transitions, increasing levels cause progression of the plant from juvenile into adult stage, from adult phase to flowering and flower development (Zhu and Helliwell 2011). This miRNA targets specifically members of the *APETALA 2 (AP2)* class of transcription factors, floral homeotic genes acting as flowering repressor. Transcript levels of *TOE1*, *TOE2* and *AP2* are high in the RKD2-induced callus, but from former studies it is known that miRNA172 acts mainly by translational repression (Chen et al. 2004). This indicates that the RKD2-induced cell line finds itself in an adult, flowering developmental stage. The up-regulated miR5640 was shown to target *PHOSPHOENOL PYRUVATE CARBOXYLASE 3* (*PPC3*), encoding an enzyme, which can be induced in roots by nitrate treatment (Vidal et al. 2013). Transcript levels of *PPC3* do not significantly differ in both cell cultures (control callus: 910, RKD2-induced: 705), thus the high expression level of mir5640 does not lead does not lead to degradation of *PPC3* mRNA. This could indicate a mechanism of translational inhibition, which differs from the post-transcriptional model in the aforementioned study.

The miR-PREFeR (miRNA PREdiction From small RNA-Seq data) software (Lei and Sun 2014) was used to predict miRNAs from the obtained small RNA sequencing data sets, but a number of sequences could not be assigned to known miRNA annotated in miRBase. Nevertheless, most of those miRNAs were not novel but identified as derived from precursors

of known miRNAs. Only five novel miRNAs were discovered that could not be linked to known precursors, one of them (miR_95) is upregulated in the RKD2 callus. Predicted targets of differentially expressed novel miRNAs from known or novel MIR precursors represent a variety of genes and may have multiple functions. Novel miR_25-5p (up in RKD2) for example is predicted to target a transposable element (AT3G33181), which belongs to the ATHILA family of retrotransposons. Small RNA reads from this locus are found in both tissues and those are not predicted to be miRNAs but may rather represent siRNAs. The mRNA of this TE is not present in our mRNA-Seq data in RKD2 induced or CIM callus, so is possible that the TE is silenced by non-canonical RdDM in both tissues.

Additionally, expression of miR_25 in the RKD2-induced cell line is accompanied by down regulation of *FLAVONOL SYNTHASE 1*, which is also a predicted target. Nevertheless, experimental evidence is needed to verify targets of novel miRNAs, like for example a RLM-RACE based approach to identify transcripts with miRNA cleavage sites.

Some of the predicted targets of down regulated miRNAs are upregulated in the RKD2-induced cell line. For example, the predicted target of novel miR_38-5p is a member of the DNA-binding storekeeper protein-related transcriptional regulators. This special member is expressed in seeds and in the expanded cotyledon stage (Schmidt et al. 2005). Another interesting target of a novel miRNA is the chromatin remodeling factor CHR5, which is targeted by novel miR_104-5p (processed from MIR3019b). Interestingly, the transcript level of its putative target *CHR* varies only slightly in the RKD2-induced cell line and the control cell line (Log2 FC 1.21). A translational inhibition is thus possible and has to be verified by protein biochemical methods (like for example Western Blot).

6.5 Regulation of egg cell-expressed genes via miRNAs

Transcriptome profiling using the RKD2-induced cell line revealed the presence of egg-cell specific expressed genes (Chapter 6.4). Genes involved in specification of gametes need to be repressed in somatic cells in order to prevent them from acquiring gametic fate. In this context it is striking that some transcription factors significantly upregulated in the RKD2-induced callus are predicted targets of miRNAs, which are highly expressed in the control callus. Members of the miR319 family, for example, regulate transcription factors of the *TCP* family (Palatnik et al. 2007). MiR319a, b and c are highly expressed in the control callus, whereas *TCP* transcription factors *TCP2*, 3, 4, 10 and 24 are only expressed at

moderate levels in this cell line. By contrast, *TCP4* is highly expressed in the RKD2-induced callus. It plays a role in leaf development, where it controls cell proliferation (Schommer et al. 2014). *TCP4* lack-of-function mutants are also impaired in embryonic development (Pagnussat et al. 2005). Plants with high TCP4 levels have a strong reduction in the expression of genes known to be active in G2-M phase of the cell cycle (Schommer et al. 2014). *TCP4* transcript can also be found in the egg cell (Soljic 2012). The role of TCP4 might be to prevent the egg from premature proliferation.

miRNA5653 is predicted to target the transcription factor *ABI4*, which is highly transcribed in the RKD2-induced cell line and in egg cells (Soljic 2012) but absent from the control callus. In correlation with this, miR5653 is strongly present in the control callus but absent in the RKD2-induced cell line, which has been confirmed by Northern Blot analyzes (Figure 59). ABI4 plays a role in seed development and response to abscisic acid (Södermann et al. 2000). Among its various functions, ABI4 represses transcription of the *ALTERNATIVE OXIDASE 1a (AOX1a)* (Giraud et al. 2011) and is involved in sugar signaling (Rook et al. 2006). The exact role of ABI4 in the egg cell remains to be investigated. The miRNA family 156/157 shows strong presence in the control callus, expressing all members of this family. Interestingly, almost no differential expression of predicted target genes, members of the SQUAMOSA promoter binding protein-like family (SPL), can be detected between the two compared tissues; Only *SPL9* is upregulated in the auxin-induced control callus and *SPL15* in the RKD2-induced callus.

A transcript for a SWIB/MDM2 domain containing protein (AT5G23480) is highly expressed in the RKD2-induced cell line. It is classified as egg cell specific by Köszegi et al. (2011) while microarray-based expression data from Soljic (2012) show its expression in all cells of the female gametophyte and in sperm cells. Differential expression of this gene has been verified by mRNA-Seq and qRT-PCR as well. SWIB/MDM2 proteins are conserved through all eukaryotic kingdoms and are involved in chromatin remodeling, repositioning of nucleosomes and transcriptional control (Bennett-Lovsey et al. 2002). This protein also possesses a Plus 3 domain, which may bind DNA (de Jong et al. 2007). Notably, the SWIB/MDM2 domain containing protein encoded by AT5G23480 is a predicted targeted of miRNA781a and miRNA781b, which are both down regulated in the RKD2-induced cell line.

AP2 transcription factors targeted by miR172 are present in the RKD2-induced callus at different levels. *TOE1* and *TOE2*, as well as *AP2* are highly expressed in this tissue despite the fact that miR172d is the most abundant miRNA. Interestingly, all three genes are

expressed in the auxin-induced control callus as well while miR172d is down regulated in this cell line. This may indicate that *TOE1*, *TOE2* and *AP2* are not targets of miR172d but others of the 147 *AP2* gene family members in Arabidopsis (Zheng et al. 2005). Nevertheless, miR172 family members have been reported to regulate gene expression by AGO-mediated cleavage as well as by translational repression (Zhu and Helliwell 2011). Presence of AP2 proteins in RKD2-induced cell lines and in egg cells has to be determined by immunhistochemical and biochemical methods. This would provide further inside into the putative role of miR172d and its target genes in specification of female gametes.

6.6 Transposon silencing and epigenetic modifications in the egg cell-like callus

In both the auxin-induced control callus and the RKD2-induced cell line are only few transcripts derived from TEs detected. This stands in contrast to the situation in the companion cells of male and female gametes of Arabidopsis. In the central cell a general state of hypomethylation leads to epigenetic reactivation of transposons. On top of that, the MEDEA (MEA) and FLOWERING LOCUS WAGENING (FWA) genes are activated by removal of epigenetic marks in their promoter region (Gehring et al. 2006, Kinoshita et al. 2004). In the RKD2-induced callus, transcripts of both loci are not detectable. Taken together, the egg cell-like callus does not seem to be in an epigenetically activated state like the central cell, even though DEMETER (DME) transcripts are present.

The presence of small RNAs derived from the Gypsy superfamily of TEs in the RKD2-induced callus on one hand and the absence of mRNA of those TEs indicates there is a functional mechanism of small RNA-mediated TE silencing in the egg cell, which is at least partially independent of mobile siRNAs generated in companion cells.

6.7 Outlook

Small RNAs play an important role in defining cell fate by regulation of mRNA expression. This is also true for the process of gametogenesis. Profiling of small RNAs and mRNAs in the RKD2-induced, egg cell-like callus may provide novel insights into the miRNA profile of the egg cell. Nine known and five novel miRNAs have been found to be upregulated in RKD2-induced cell line. The presence of these miRNAs in the egg cell needs

to be validated by whole mount in situ hybridization with Locked Nucleic Acid (LNA^{TM}) oligonucleotides to detect small RNAs in the ovule. Bioinformatic predictions of target genes have, together with comparative analysis of miRNA and predicted target mRNA levels, led to new insights into genes that may be specifically down regulated by these microRNAs. Further studies are needed in order to further validate miRNA/target mRNA interactions, using techniques like degradome analysis or parallel analysis of RNA ends (PARE). Another possibility would be the purification of miRNA:mRNA target by the immunoprecipitation of the RISC components (reviewed in Thomson et al. 2011). Transcriptome profiling revealed the presence of the maintenance DNA-methylases CMT3 and MET1 at transcript level in the RKD2-induced cell line. However, in previous studies the respective proteins are not detectable in the egg cell. Since the DNA-methylation level in CHG and CG context is rising again after fertilization in embryogenesis, the amount of transcripts of *CMT3* and *MET1* in the RKD2-induced callus might reflect the fact, that in the egg cell these transcripts are stored to be immediately translated after fertilization. In order to verify this assumption the RKD2-induced cell line would have to be examined for the presence of CMT3 and MET1 proteins. Further investigations about epigenetic modifications in the germline could be conducted by DNA methylome sequencing using RKD2-induced cell line and isolated egg cells.

AGO9 has been reported to play a role in transposon silencing in generative tissues. However in the mature female gamete, there is no AGO9 protein detectable. In early embryonic development, however, AGO9 protein is very abundant. This might indicate that *AGO9* transcript is stored in the egg cell to be translated directly after fertilization. Further studies should deal with the question of how *AGO9* translation is regulated. The *AGO9-8* transcript variant with the single retained intron 6 is present in the RKD2-induced callus along with its correctly spliced form. Previous studies revealed intron retention to be involved in mechanisms of gene regulation (reviewed in Wong et al. 2015). The storage of unspliced mRNA has been demonstrated in a recent study in the heterosporous fern, *Marsilea vestita* (Boothby et al. 2013). Those transcripts were found to be fully spliced only at specific developmental stage. In contrast to this, *AGO9* transcript variants appear both present at the same time, with increased intron retention during ovule development. Reduced AGO9 protein stability in the female gametes may add an additional level of regulation. A future challenge will be the detection of *AGO9* transcript variants on the cellular level in the female gametes and during early embryogenesis.

7 PUBLICATIONS

Parts of this work have been published in the following manuscript:

Hauptmann J, Schraivogel D, Bruckmann A, Manickavel S, Jakob L, Eichner N, Pfaff J, Urban M, Sprunck S, Hafner M, Tuschl T, Deutzmann R, Meister G (2015) Biochemical isolation of Argonaute protein complexes by Ago-APP. Proc Natl Acad Sci U S A **112:** 11841-11845

Manuscript in preparation:

Urban M, Strieder N, Bleckmann A, Šoljić L, Xia N, Hackenberg T, Evers M, Möhle C, Dresselhaus T, Engelmann JC and Sprunck S Identification of small RNA pathway genes and small noncoding RNAs in the Arabidopsis egg cell using a cell line with egg cell-like identity

8 BIBLIOGRAPHY

Adamowski M, Friml J (2015) PIN-dependent auxin transport: action, regulation, and evolution. Plant Cell **27:** 20-32

Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell **121:** 207-221

Anders S, Pyl PT, Huber W (2015) HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics **31:** 166-169

Aukerman MJ, 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

Axtell M (2014) Butter: High-precision genomic alignment of small RNA-seq data. BioRxiv: doi: http://dx.doi.org/10.1101/007427

Bao Y, Wang C, Jiang C, Pan J, Zhang G, Liu H, Zhang H (2014) The tumor necrosis factor receptor-associated factor (TRAF)-like family protein SEVEN IN ABSENTIA 2 (SINA2) promotes drought tolerance in an ABA-dependent manner in Arabidopsis. New Phytol **202:** 174-187

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

Benfey PN, Chua NH (1990) The Cauliflower Mosaic Virus 35S promoter: combinatorial regulation of transcription in plants. Science **250:** 959-966

Bennett-Lovsey R, Hart SE, Shirai H, Mizuguchi K (2002) The SWIB and the MDM2 domains are homologous and share a common fold. Bioinformatics **18:** 626-630

Berger F, Twell D (2011) Germline specification and function in plants. Annu Rev Plant Biol **62:** 461-484

Berr A, McCallum EJ, Menard R, Meyer D, Fuchs J, Dong A, Shen WH (2010) Arabidopsis SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development. Plant Cell **22:** 3232-3248

Bleckmann A, Dresselhaus T (2015) Fluorescent whole-mount RNA in situ hybridization (F-WISH) in plant germ cells and the fertilized ovule. Methods **Advance online publication:** doi: 10.1016/j.ymeth.2015.10.019

Bogdanovic O, Veenstra GJ (2009) DNA methylation and methyl-CpG binding proteins: developmental requirements and function. Chromosoma **118:** 549-565

Böhmert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J **17:** 170-180

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:** 2114-2120

Boothby TC, Zipper RS, van der Weele CM, Wolniak SM (2013) Removal of retained introns regulates translation in the rapidly developing gametophyte of *Marsilea vestita.* Dev Cell **24:** 517-529

Borges F, Pereira PA, Slotkin RK, Martienssen RA, Becker JD (2011) MicroRNA activity in the Arabidopsis male germline. J Exp Bot **62:** 1611-1620

Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. Nat Rev Mol Cell Biol

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry **72:** 248-254

Braun JE, Huntzinger E, Izaurralde E (2013) The role of GW182 proteins in miRNA-mediated gene silencing. Adv Exp Med Biol **768:** 147-163

Breakfield NW, Corcoran DL, Petricka JJ, Shen J, Sae-Seaw J, Rubio-Somoza I, Weigel D, Ohler U, Benfey PN (2012) High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis. Genome Res **22:** 163-176

Brosseau C, Moffett P (2015) Functional and Genetic Analysis Identify a Role for Arabidopsis ARGONAUTE5 in Antiviral RNA Silencing. Plant Cell **27:** 1742-1754

Buechel S, Leibfried A, To JP, Zhao Z, Andersen SU, Kieber JJ, Lohmann JU (2010) Role of A-type ARABIDOPSIS RESPONSE REGULATORS in meristem maintenance and regeneration. Eur J Cell Biol **89:** 279-284

Calarco JP, Borges F, Donoghue MT, Van Ex F, Jullien PE, Lopes T, Gardner R, Berger F, Feijo JA, Becker JD, Martienssen RA (2012) Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell **151:** 194-205

Castel SE, Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet **14:** 100-112

Chan SW, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE (2004) RNA silencing genes control de novo DNA methylation. Science **303:** 1336

Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet **6:** 351-360

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

Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. Cell **110:** 33-42

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16:** 735-743

Cooper TA (2005) Use of minigene systems to dissect alternative splicing elements. Methods **37:** 331-340

Costa-Nunes P, Kim JY, Hong E, Pontes O (2014) The cytological and molecular role of domains rearranged methyltransferase3 in RNA-dependent DNA methylation of *Arabidopsis thaliana*. BMC Res Notes **7:** 721

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

Christensen CA, Subramanian S, Drews GN (1998) Identification of gametophytic mutations affecting female gametophyte development in Arabidopsis. Dev Biol **202:** 136-151

Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res **39:** W155-159

Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ (2004) Processing of primary microRNAs by the microprocessor complex. Nature **432:** 231-235

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation:version II. Plant Mol Biol Rep **1**: 19-21.

de Jong RN, Truffault V, Diercks T, Ab E, Daniels MA, Kaptein R, Folkers GE (2008) Structure and DNA binding of the human Rtf1 Plus3 domain. Structure **16:** 149-159

Duan CG, Zhang H, Tang K, Zhu X, Qian W, Hou YJ, Wang B, Lang Z, Zhao Y, Wang X, Wang P, Zhou J, Liang G, Liu N, Wang C, Zhu JK (2014) Specific but interdependent functions for Arabidopsis AGO4 and AGO6 in RNA-directed DNA methylation. EMBO J **34:** 581-592

Duran-Figueroa N, Vielle-Calzada JP (2010) ARGONAUTE9-dependent silencing of transposable elements in pericentromeric regions of Arabidopsis. Plant Signal Behav **5:** 1476-1479

El-Showk S, Ruonala R, Helariutta Y (2013) Crossing paths: cytokinin signaling and crosstalk. Development **140:** 1373-1383

Eulalio A, Behm-Ansmant I, Izaurralde E (2007) P bodies: at the crossroads of post-transcriptional pathways. Nat Rev Mol Cell Biol **8:** 9-22

Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E (2007) P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. Mol Cell Biol **27:** 3970-3981

Eulalio A, Huntzinger E, Izaurralde E (2008) GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. Nat Struct Mol Biol **15:** 346-353

Eun C, Lorkovic ZJ, Naumann U, Long Q, Havecker ER, Simon SA, Meyers BC, Matzke AJ, Matzke M (2011) AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in Arabidopsis thaliana. PLoS One **6:** e25730 #

Fabian MR, Sonenberg N (2012) The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat Struct Mol Biol **19:** 586-593

Feng JX, Liu D, Pan Y, Gong W, Ma LG, Luo JC, Deng XW, Zhu YX (2005) An annotation update via cDNA sequence analysis and comprehensive profiling of developmental, hormonal or environmental responsiveness of the Arabidopsis AP2/EREBP transcription factor gene family. Plant Mol Biol **59:** 853-868

Feng S, Jacobsen SE, Reik W Epigenetic reprogramming in plant and animal development. Science **330:** 622-627

Foyer CH, Kerchev PI, Hancock RD (2012) The ABA-INSENSITIVE-4 (ABI4) transcription factor links redox, hormone and sugar signaling pathways. Plant Signal Behav **7:** 276-281

Gagnon KT, Li L, Chu Y, Janowski BA, Corey DR (2014) RNAi factors are present and active in human cell nuclei. Cell Rep **6:** 211-221

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell **124:** 495-506

Giraud E, Van Aken O, Ho LH, Whelan J (2009) The transcription factor ABI4 is a regulator of mitochondrial retrograde expression of ALTERNATIVE OXIDASE1a. Plant Physiol **150:** 1286-1296

Gross-Hardt R, Kagi C, Baumann N, Moore JM, Baskar R, Gagliano WB, Jurgens G, Grossniklaus U (2007) LACHESIS restricts gametic cell fate in the female gametophyte of Arabidopsis. PLoS Biol **5:** e47

Gurskaya NG, Staroverov DB, Zhang L, Fradkov AF, Markina NM, Pereverzev AP, Lukyanov KA (2012) Analysis of alternative splicing of cassette exons at single-cell level using two fluorescent proteins. Nucleic Acids Res **40:** e57

Haag JR, Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol **12:** 483-492

Hackenberg T (2016), Gamete surface proteins involved in gamete interaction and fusion, Unpublished PhD thesis, University of Regensburg

Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ (2001) ARGONAUTE2, a link between genetic and biochemical analyzes of RNAi. Science **293:** 1146-1150

Harvey JJ, Lewsey MG, Patel K, Westwood J, Heimstadt S, Carr JP, Baulcombe DC (2011) An antiviral defense role of AGO2 in plants. PLoS One **6:** e14639

Hauptmann J, Dueck A, Harlander S, Pfaff J, Merkl R, Meister G (2013) Turning catalytically inactive human Argonaute proteins into active slicer enzymes. Nat Struct Mol Biol **20:** 814-8174

Hauptmann J, Schraivogel D, Bruckmann A, Manickavel S, Jakob L, Eichner N, Pfaff J, Urban M, Sprunck S, Hafner M, Tuschl T, Deutzmann R, Meister G (2015) Biochemical isolation of Argonaute protein complexes by Ago-APP. Proc Natl Acad Sci U S A **112:** 11841-11845

Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC (2010) The Arabidopsis RNA-directed DNA methylation Argonautes functionally diverge based on their expression and interaction with target loci. Plant Cell **22:** 321-334

He H, Yang T, Wu W, Zheng B (2015) Small RNAs in pollen. Sci China Life Sci **58:** 246-252

Hejatko J, Blilou I, Brewer PB, Friml J, Scheres B, Benkova E (2006) In situ hybridization technique for mRNA detection in whole mount Arabidopsis samples. Nat Protoc **1:** 1939-1946

Höck J, Meister G (2008) The Argonaute protein family. Genome Biol **9:** 210

Huijser P, Schmid M (2011) The control of developmental phase transitions in plants. Development **138:** 4117-4129

Humphreys DT, Westman BJ, Martin DI, Preiss T (2005) microRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. Proc Natl Acad Sci U S A **102:** 16961-16966

Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol **9:** 22-32

Ibarra CA, Feng X, Schoft VK, Hsieh TF, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T, Rojas D, Fischer RL, Tamaru H, Zilberman D (2012) Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science **337:** 1360-1364

Ito H, Kakutani T (2014) Control of transposable elements in *Arabidopsis thaliana.* Chromosome Res **22:** 217-223

Jasik J, Schiebold S, Rolletschek H, Denolf P, Van Adenhove K, Altmann T, Borisjuk L (2011) Subtissue-specific evaluation of promoter efficiency by quantitative fluorometric assay in laser microdissected tissues of rapeseed. Plant Physiol **157:** 563-573

Jinek M, Doudna JA (2009) A three-dimensional view of the molecular machinery of RNA interference. Nature **457:** 405-412

Johnston AJ, Meier P, Gheyselinck J, Wuest SE, Federer M, Schlagenhauf E, Becker JD, Grossniklaus U (2007) Genetic subtraction profiling identifies genes essential for Arabidopsis reproduction and reveals interaction between the female gametophyte and the maternal sporophyte. Genome Biol **8:** R204

Joshua-Tor L, Hannon GJ (2010) Ancestral roles of small RNAs: an Ago-centric perspective. Cold Spring Harb Perspect Biol **3:** a003772

Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F (2012) DNA methylation dynamics during sexual reproduction in *Arabidopsis thaliana*. Curr Biol **22:** 1825-1830

Karimi M, Inze D, Depicker A (2002) GATEWAY vectors for Agrobacterium-mediated plant transformation. Trends Plant Sci **7:** 193-195

Kawashima T, Berger F Epigenetic reprogramming in plant sexual reproduction. Nat Rev Genet **15:** 613-624

Kersey PJ, Allen JE, Christensen M, Davis P, Falin LJ, Grabmueller C, Hughes DS, Humphrey J, Kerhornou A, Khobova J, Langridge N, McDowall MD, Maheswari U, Maslen G, Nuhn M, Ong CK, Paulini M, Pedro H, Toneva I, Tuli MA, Walts B, Williams G, Wilson D, Youens-Clark K, Monaco MK, Stein J, Wei X, Ware D, Bolser DM, Howe KL, Kulesha E, Lawson D, Staines DM (2014) Ensembl Genomes 2013: scaling up access to genome-wide data. Nucleic Acids Res **42:** D546-552

Kierzkowski D, Kmieciak M, Piontek P, Wojtaszek P, Szweykowska-Kulinska Z, Jarmolowski A (2009) The Arabidopsis CBP20 targets the cap-binding complex to the nucleus, and is stabilized by CBP80. Plant J **59:** 814-825

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14**(4):R36.

Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol **10:** 126-139

Kingsolver MB, Huang Z, Hardy RW (2013) Insect antiviral innate immunity: pathways, effectors, and connections. J Mol Biol **425:** 4921-4936

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T (2004) One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science **303:** 521-523

Kiriakidou M, Tan GS, Lamprinaki S, De Planell-Saguer M, Nelson PT, Mourelatos Z (2007) An mRNA m7G cap binding-like motif within human AGO2 represses translation. Cell **129:** 1141-1151

Komiya R, Ohyanagi H, Niihama M, Watanabe T, Nakano M, Kurata N, Nonomura K (2014) Rice germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. Plant J **78:** 385-397

Köszegi D, Johnston AJ, Rutten T, Czihal A, Altschmied L, Kumlehn J, Wust SE, Kirioukhova O, Gheyselinck J, Grossniklaus U, Baumlein H (2011) Members of the RKD transcription factor family induce an egg cell-like gene expression program. Plant J **67:** 280-291

Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res **42:** D68-73

Krecek P, Skupa P, Libus J, Naramoto S, Tejos R, Friml J, Zazimalova E (2009) The PIN-FORMED (PIN) protein family of auxin transporters. Genome Biol **10:** 249

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:** 680-685

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

Laux T, Mayer KF, Berger J, Jurgens G (1996) The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. Development **122:** 87-96

Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol **9:** e1003118

Lei J, Sun Y: miR-PREFeR: an accurate, fast and easy-to-use plant miRNA prediction tool using small RNA-Seq data (2014). Bioinformatics, **30**(19): 2837-2839

Leung AK, Calabrese JM, Sharp PA (2006) Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules. Proc Natl Acad Sci USA **103:** 18125-18130

Li CF, Pontes O, El-Shami M, Henderson IR, Bernatavichute YV, Chan SW, Lagrange T, Pikaard CS, Jacobsen SE (2006) An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana. Cell **126:** 93-106

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis, G, Durbin R and 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. Bioinformatics, **25**, 2078–2079.

Li Y, Pi L, Huang H, Xu L (2012) ATH1 and KNAT2 proteins act together in regulation of plant inflorescence architecture. J Exp Bot **63:** 1423-1433

Li S, Liu L, Zhuang X, Yu Y, Liu X, Cui X, Ji L, Pan Z, Cao X, Mo B, Zhang F, Raikhel N, Jiang L, Chen X (2013) microRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in Arabidopsis. Cell **153:** 562-574

Li Q, Wang X, Sun H, Zeng J, Cao Z, Li Y, Qian W (2015) Regulation of Active DNA Demethylation by a Methyl-CpG-Binding Domain Protein in Arabidopsis thaliana. PLoS Genet **11:** e1005210

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30:** 923-930

Liu C, Li LC, Chen WQ, Chen X, Xu ZH, Bai SN (2013) HDA18 affects cell fate in Arabidopsis root epidermis via histone acetylation at four kinase genes. Plant Cell **25:** 257-269

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ (2004) Argonaute2 is the catalytic engine of mammalian RNAi. Science **305:** 1437-1441

Lingel A, Simon B, Izaurralde E, Sattler M (2003) Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature **426:** 465-469

Lingel A, Simon B, Izaurralde E, Sattler M (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. Nat Struct Mol Biol **11:** 576-577

Lituiev DS, Krohn NG, Muller B, Jackson D, Hellriegel B, Dresselhaus T, Grossniklaus U (2013) Theoretical and experimental evidence indicates that there is no detectable auxin gradient in the angiosperm female gametophyte. Development **140:** 4544-4553

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods **25:** 402-408

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

Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ (2005) Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. Nature **434:** 666-670

Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM (2015) Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genet **11:** e1004915

Mallory AC, Elmayan T, Vaucheret H (2008) microRNA maturation and action--the expanding roles of ARGONAUTEs. Curr Opin Plant Biol **11:** 560-566

Mallory AC, Hinze A, Tucker MR, Bouche N, Gasciolli V, Elmayan T, Lauressergues D, Jauvion V, Vaucheret H, Laux T (2009) Redundant and specific roles of the ARGONAUTE proteins AGO1 and ZLL in development and small RNA-directed gene silencing. PLoS Genet **5:** e1000646

Mallory AC, Vaucheret H (2013) Form, function, and regulation of Argonaute proteins. Plant Cell **22:** 3879-3889

Mano Y, Nemoto K (2012) The pathway of auxin biosynthesis in plants. J Exp Bot **63:** 2853-2872

Marin E, Jouannet V, Herz A, Lokerse AS, Weijers D, Vaucheret H, Nussaume L, Crespi MD, Maizel A (2010) miR390, Arabidopsis TAS3 tasiRNAs, and their AUXIN RESPONSE FACTOR targets define an autoregulatory network quantitatively regulating lateral root growth. Plant Cell **22:** 1104-1117

Marin I, Llorens C (2000) Ty3/Gypsy retrotransposons: description of new Arabidopsis thaliana elements and evolutionary perspectives derived from comparative genomic data. Mol Biol Evol **17:** 1040-1049

Mason MG, Mathews DE, Argyros DA, Maxwell BB, Kieber JJ, Alonso JM, Ecker JR, Schaller GE (2005) Multiple type-B response regulators mediate cytokinin signal transduction in Arabidopsis. Plant Cell **17**: 3007-3018

Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet **15:** 394-408

Matzke MA, Kanno T, Matzke AJ (2015) RNA-Directed DNA Methylation: The evolution of a complex epigenetic pathway in flowering plants. Annu Rev Plant Biol **66:** 243-267

McCue AD, Nuthikattu S, Slotkin RK (2013) Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs. RNA Biol **10:** 1379-1395

McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK (2015) ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. EMBO J **34:** 20-35

Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ, Griffiths-Jones S, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi Y, Vaucheret H, Voinnet O, Watanabe Y, Weigel D, Zhu JK (2008) Criteria for annotation of plant microRNAs. Plant Cell **20:** 3186-3190

Meister G, Tuschl T (2004) Mechanisms of gene silencing by double-stranded RNA. Nature **431:** 343-349

Meister G (2013) Argonaute proteins: functional insights and emerging roles. Nat Rev Genet **14:** 447-459

Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, Chen S, Hannon GJ, Qi Y (2008) Sorting of small RNAs into Arabidopsis Argonaute complexes is directed by the 5' terminal nucleotide. Cell **133:** 116-127

Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, Gentleman R (2009) ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics **25:** 2607-2608

Müller B, Sheen J (2008) Cytokinin and auxin interaction in root stem-cell specification during early embryogenesis. Nature **453:** 1094-1097

Nakanishi A, Kobayashi N, Suzuki-Hirano A, Nishihara H, Sasaki T, Hirakawa M, Sumiyama K, Shimogori T, Okada N (2012) A SINE-derived element constitutes a unique modular enhancer for mammalian diencephalic FGF8. PLoS One **7:** e43785

Nakanishi K, Weinberg DE, Bartel DP, Patel DJ (2012) Structure of yeast Argonaute with guide RNA. Nature **486:** 368-374

Obayashi T, Okamura Y, Ito S, Tadaka S, Aoki Y, Shirota M, Kinoshita K (2014) ATTED-II in 2014: evaluation of gene coexpression in agriculturally important plants. Plant Cell Physiol **55:** e6

Okrent RA, Brooks MD, Wildermuth MC (2009) Arabidopsis GH3.12 (PBS3) conjugates amino acids to 4-substituted benzoates and is inhibited by salicylate. J Biol Chem **284:** 9742-9754

Okrent RA, Wildermuth MC (2011) Evolutionary history of the GH3 family of acyl adenylases in rosids. Plant Mol Biol **76:** 489-505

Oliver C, Santos JL, Pradillo M (2014) On the role of some Argonaute proteins in meiosis and DNA repair in Arabidopsis thaliana. Front Plant Sci **5:** 1774

Pagnussat GC, Yu HJ, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie LF, Ye D, Sundaresan V (2005) Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. Development **132:** 603-614

Pagnussat GC, Alandete-Saez M, Bowman JL, Sundaresan V (2009) Auxin-dependent patterning and gamete specification in the Arabidopsis female gametophyte. Science **324:** 1684-1689

Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, Rodriguez R, Warthmann N, Allen E, Dezulian T, Huson D, Carrington JC, Weigel D (2007) Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs miR159 and miR319. Dev Cell **13:** 115-125

Pall GS, Hamilton AJ (2008) Improved northern blot method for enhanced detection of small RNA. Nat Protoc **3:** 1077-1084

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev **20:** 3407-3425

Ramsak Z, Baebler S, Rotter A, Korbar M, Mozetic I, Usadel B, Gruden K (2014) GoMapMan: integration, consolidation and visualization of plant gene annotations within the MapMan ontology. Nucleic Acids Res **42:** D1167-1175

Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, Hagen G, Guilfoyle TJ, Pasa-Tolic L, Pikaard CS (2009) Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. Mol Cell **33:** 192-203

Rivas FV, Tolia NH, Song JJ, Aragon JP, Liu J, Hannon GJ, Joshua-Tor L (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol **12:** 340-349

Rodriguez-Leal D, Leon-Martinez G, Abad-Vivero U, Vielle-Calzada JP (2015) Natural variation in epigenetic pathways affects the specification of female gamete precursors in Arabidopsis. Plant Cell **27:** 1034-1045

Rook F, Hadingham SA, Li Y, Bevan MW (2006) Sugar and ABA response pathways and the control of gene expression. Plant Cell Environ **29:** 426-434

Sakai H, Honma T, Aoyama T, Sato S, Kato T, Tabata S, Oka A (2001) ARR1, a transcription factor for genes immediately responsive to cytokinins. Science **294:** 1519-1521

Sambrook, J, Fritsch, EF, and Maniatis, T (1989) Molecular cloning: a laboratory manual; Second edition vols. 1, 2 and 3. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, USA

Sasaki T, Nishihara H, Hirakawa M, Fujimura K, Tanaka M, Kokubo N, Kimura-Yoshida C, Matsuo I, Sumiyama K, Saitou N, Shimogori T, Okada N (2008)

Possible involvement of SINEs in mammalian-specific brain formation. Proc Natl Acad Sci USA **105:** 4220-4225

Schirle NT, MacRae IJ (2012) The crystal structure of human ARGONAUTE2. Science **336:** 1037-1040

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nat Genet **37:** 501-506

Schommer C, Debernardi JM, Bresso EG, Rodriguez RE, Palatnik JF (2014) Repression of cell proliferation by miR319-regulated TCP4. Mol Plant **7:** 1533-1544

Schwarz S, Grande AV, Bujdoso N, Saedler H, Huijser P (2008) The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis. Plant Mol Biol **67:** 183-195

Schürmann N, Trabuco LG, Bender C, Russell RB, Grimm D (2013) Molecular dissection of human Argonaute proteins by DNA shuffling. Nat Struct Mol Biol **20:** 818-826

Singh M, Goel S, Meeley RB, Dantec C, Parrinello H, Michaud C, Leblanc O, Grimanelli D (2011) Production of viable gametes without meiosis in maize deficient for an Argonaute protein. Plant Cell **23:** 443-458

Skoog F, Miller CO (1957) Chemical regulation of growth and organ formation in plant tissues cultured in vitro. Symp Soc Exp Biol **11:** 118-130

Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136:** 461-472

Smyth, DR, Bowman, JL, and Meyerowitz, EM (1990) Early flower development in *Arabidopsis*. Plant Cell **2,** 755-767.

Soderman EM, Brocard IM, Lynch TJ, Finkelstein RR (2000) Regulation and function of the Arabidopsis ABA-insensitive4 gene in seed and abscisic acid response signaling networks. Plant Physiol **124:** 1752-1765

Soljic S (2012) Microarray analysis of single isolated cells of the female gametophyte reveals potential regulators of female germline development in *Arabidopsis thaliana*, PhD thesis, University of Regensburg

Song X, Yuan L, Sundaresan V (2014) Antipodal cells persist through fertilization in the female gametophyte of Arabidopsis. Plant Reprod **27:** 197-203

Sparks E, Wachsman G, Benfey PN (2013) Spatiotemporal signalling in plant development. Nat Rev Genet **14:** 631-644

Sprunck S, Gross-Hardt R (2011) Nuclear behavior, cell polarity, and cell specification in the female gametophyte. Sex Plant Reprod **24:** 123-136

Sprunck S, Rademacher S, Vogler F, Gheyselinck J, Grossniklaus U, Dresselhaus T (2012) Egg cell-secreted EC1 triggers sperm cell activation during double fertilization. Science **338:** 1093-1097

Wong JJ, Au AY, Ritchie W, Rasko JE (2015) Intron retention in mRNA: no longer nonsense: known and putative roles of intron retention in normal and disease biology. Bioessays doi: 10.1002/bies.201500117

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

Takeda A, Iwasaki S, Watanabe T, Utsumi M, Watanabe Y (2008) The mechanism selecting the guide strand from small RNA duplexes is different among Argonaute proteins. Plant Cell Physiol **49:** 493-500

ten Hove CA, Lu KJ, Weijers D (2015) Building a plant: cell fate specification in the early Arabidopsis embryo. Development **142:** 420-430

Thomson DW, Bracken CP, Goodall GJ (2011) Experimental strategies for microRNA target identification. Nucleic Acids Res **39:** 6845-6853

Tucker MR, Okada T, Hu Y, Scholefield A, Taylor JM, Koltunow AM (2012) Somatic small RNA pathways promote the mitotic events of megagametogenesis during female reproductive development in Arabidopsis. Development **139:** 1399-1404

Vaucheret H (2008) Plant Argonautes. Trends Plant Sci **13:** 350-358

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science **303:** 672-676

Vidal EA, Moyano TC, Krouk G, Katari MS, Tanurdzic M, McCombie WR, Coruzzi GM, Gutierrez RA (2013) Integrated RNA-seq and sRNA-seq analysis identifies novel nitrate-responsive genes in Arabidopsis thaliana roots. BMC Genomics **14:** 701

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science **297:** 1833-1837

Xie P (2013) TRAF molecules in cell signaling and in human diseases. J Mol Signal **8:** 7

Xie M, Zhang S, Yu B (2015) microRNA biogenesis, degradation and activity in plants. Cell Mol Life Sci **72:** 87-99

Yang WC, Ye D, Xu J, Sundaresan V (1999) The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes Dev **13:** 2108-2117

Yang WC, Shi DQ, Chen YH (2010) Female gametophyte development in flowering plants. Annu Rev Plant Biol **61:** 89-108

Ye R, Wang W, Iki T, Liu C, Wu Y, Ishikawa M, Zhou X, Qi Y (2012) Cytoplasmic assembly and selective nuclear import of Arabidopsis ARGONAUTE4/siRNA complexes. Mol Cell **46:** 859-870

Wassenegger M, Heimes S, Riedel L, Sanger HL (1994) RNA-directed de novo methylation of genomic sequences in plants. Cell **76:** 567-576

Wartlsteiner U (2010) Analyze zur Charakterisierung der Rolle von AtAGO8 und AtAGO9 bei der Entwicklung des weiblichen Gametophyten von *Arabidopsis thaliana.* Diploma thesis University of Regensburg

Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenfuhrer J, von Mering C, Grossniklaus U (2010) Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. Curr Biol **20:** 506-512

Yanai O, Shani E, Dolezal K, Tarkowski P, Sablowski R, Sandberg G, Samach A, Ori N (2005) Arabidopsis KNOXI proteins activate cytokinin biosynthesis. Curr Biol **15:** 1566-1571

Zhao Z, Andersen SU, Ljung K, Dolezal K, Miotk A, Schultheiss SJ, Lohmann JU (2010) Hormonal control of the shoot stem-cell niche. Nature **465:** 1089-1092

Zhang CJ, Ning YQ, Zhang SW, Chen Q, Shao CR, Guo YW, Zhou JX, Li L, Chen S, He XJ (2012) IDN2 and its paralogs form a complex required for RNA-directed DNA methylation. PLoS Genet **8:** e1002693

Zhang X, Zhao H, Gao S, Wang WC, Katiyar-Agarwal S, Huang HD, Raikhel N, Jin H (2011) Arabidopsis ARGONAUTE 2 regulates innate immunity via miRNA393(*)-mediated silencing of a Golgi-localized SNARE gene, MEMB12. Mol Cell **42:** 356-366

Zhang Z, Zhang X (2012) Argonautes compete for miR165/166 to regulate shoot apical meristem development. Curr Opin Plant Biol **15:** 652-658

Zheng S, Damoiseaux R, Chen L, Black DL (2013) A broadly applicable high-throughput screening strategy identifies new regulators of DLG4 (PSD-95) alternative splicing. Genome Res **23:** 998-1007

Zhong R, Ye ZH (2007) Regulation of HD-ZIP III Genes by MicroRNA 165. Plant Signal Behav **2:** 351-353

Zhong X, Hale CJ, Nguyen M, Ausin I, Groth M, Hetzel J, Vashisht AA, Henderson IR, Wohlschlegel JA, Jacobsen SE (2015) DOMAINS REARRENGED METHYLTRANSFERASE3 controls DNA methylation and regulates RNA polymerase V transcript abundance in Arabidopsis. Proc Natl Acad Sci USA **112:** 911-916

Zhou GK, Kubo M, Zhong R, Demura T, Ye ZH (2007) Overexpression of miR165 affects apical meristem formation, organ polarity establishment and vascular development in Arabidopsis. Plant Cell Physiol **48:** 391-404

Zhu QH, Helliwell CA (2011) Regulation of flowering time and floral patterning by miR172. J Exp Bot **62:** 487-495

Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation

9 APPENDIX

9.1 Antibodies

Table 14: Summary of Antibodies used in this work.
WB = Western Blot, IL = (whole mount) immunolocalization, HRP = horse raddish peroxidase, Cy3 = Indocarbocyanine, fluorescent dye. All antibodies are polyclonal unless otherwise indicated.

Epitope	Host	Conjugate	Dilution	Manufacturer
AGO1	rabbit	None	1:5.000 WB	Agrisera (AS09527)
AGO2	rabbit	None	1:500 WB	Agrisera (AS132682)
AGO4	rabbit	None	1:4.000 WB	Agrisera (AS09617)
AGO5	rabbit	None	1:1.500 WB	Agrisera (AS10671)
AGO6	rabbit	None	1:1.500 WB	Agrisera (AS10672)
AGO9	rabbit	None	1:10.000 WB, 1:200 IL	Agrisera (AS10673)
rabbit IgG	goat	HRP	1:5.000 WB	Sigma Aldrich (M6150)
rabbit IgG	goat	Cy3	1:250 IL	Dianova (111-165-003)
mouse IgG	goat	Cy3	1:250 IL	Dianova (115-165-003)
GFP	mouse (monoclonal)	None	1:200 IL	Roche (11 814 460 001)

9.2 Oligonucleotides

Table 15: Cloning primers
Promoter regions were amplified from genomic DNA using pAGO8 and pAGO9 primers. Coding sequences from AGO8, AGO9 and RKD2 were amplified from cDNA. RKD2 primer sequences were taken from Köszegi et al. (2011). For minigene fragments, genomic DNA was used as template. Restriction endonuclease recognition sites are written in bold letters, overhangs for cloning into pENTR/D-TOPO (CACC) are underlined and written in italics.

			Fragment size (bp)	
Primer name	**Accession**	**Sequence (5' to 3 ')**	**gDNA**	**cDNA**
AGO8p for	AT5G21030	*CACC*AGCTTTTAGTATGAAAAAATGCAAGAGATTC	637	
AGO8p rev		TTTCTCGGCCTCCAGATGTTATGC		
AGO8p for (SacI)	AT5G21030	GGGG**GAGCTC**CTTTTAGTATGAAAAAATGCAAGAGATTC	637	
AGO8p rev (XbaI)		GGGG**TCTAGA**TTTCTCGGCCTCCAGATGTTATGC		
AGO9p for	AT5G21150	*CACC*CACCACTTGGGTCATGTAGCCACATCGATGC		2033
AGO9p rev		CTCTGGGATACAAATATCAGAAGGTC		

			Fragment size (bp)	
Primer name	**Accession**	**Sequence (5' to 3 ')**	**gDNA**	**cDNA**
AGO9p for (SacI)	AT5G21150	GGGG**GAGCTC**CACCACTTGGGTCATGTAGCCACATCGATGC		2033
AGO9p rev (XbaI)		GGGG**TCTAGA**TTTCTCGGCCTCCAGATGTTATGC		
AGO8 for	AT5G21030	*CACC*ATGGATACGACTCTACCGCCTCCTCA		2553
AGO8 rev		GCAAAAGAACATTGAACTAGC		
AGO9 for	AT5G21150	*CACC*ATGGATTCTGATGAACCGAATGGGAG		2691
AGO9 rev		TCAACAGAAGAACATGGAGGTTGAAACATTAT		
Mini A9 I6 for	AT5G21150	ATATA**GCGGCCGC**aGACACTTCAACTACGATGATAGTACAA	262	
Mini A9 I6 rev		ATATA**GCGGCCGC**AAAGTTGATCTTTGCAGCTGTGTTCACT		
Mini A9 I567 for	AT5G21150	ATATAGG**CGGCCGC**AGGTTGCCTCCTTGTTCGCCAGTC	721	
Mini A9 I567 rev		ATATA**GCGGCCG**CACTCAATGGGGAAGTAAGTGGGACGCTTCG		
Mini A9 I11 for	AT5G21150	ATATA**GCGGCCG**CACTGAAAGTGGGCAAGGGAGAAA	252	
Mini A9 I11 rev		ATATA**GCGGCCG**CAAATTCCTTTCTCCCGTCCACAT		
Mini A9 I101112 for	AT5G21150	ATATA**GCGGCCGC**AGGTCTGAAGGACAGCAATTACAATG	717	
Mini A9 I101112 rev		ATATA**GCGGCCGC**CATAAACATCAGAGTTTTTCCTTTCAGCGAGTA		
AtRKD2 forward	AT1G74480	*CACC*ATGGCTGATCACACAACCAAAGAACAGA		897
AtRKD2 reverse		CAAACCACTAGTAAATTCACTTGAGA		

Table 16: Primers for sequencing and analysis of transcript variants
List of primers used for PCR-based analysis of AGO8 and AGO9 transcript variants. Primers were also used for sequencing. Primer sequences taken from Wartlsteiner (2010).

			Fragment size (bp)	
Primer name	**Accession**	**Sequence (5' to 3 ')**	**gDNA**	**cDNA**
AGO8 seq5	AT5G21030	CTCAAGACAATGGCCACTCA	1187	601
AGO8 GD rev		GGGCTTTCGGTTTGGAAGAAC		
AGO9 for	AT5G21150	CACCATGGATTCTGATGAACCGAATGGGAG	987	700
AGO9rev700		CACCTCCACCAATAGGTACA		
AGO9 seq2		GCTATTGCAAGCGCTCTTCA	1023	620
AGO9 rev 1154		GTGTATCGCTGTAGAGACAC		
AGO9 seq3		GCGTTATTCAGGTGACTTCC	978	543
AGO9 rev 1595		CTCACAGTAGCTGGTGCATT		
AGO9 seq4		AACTTCTCTGCTCGCTGTGA	1405	706
AGO9 rev 2158		GAGACTCGCTTACACCATCC		
AGO9 seq5		CTCAAGACAATGGCCACTCA	1430	739
AGO9 rev		TCAACAGAAGAACATGGAGGTTGAAACATTAT		

Table 17: (q)RT-PCR primers
List of primers used for RT-PCR and qRT-PCR. Sequences of AGO8 primers from Köszegi (2011). Sequences for AGO2, AGO3, AGO4 and AGO6 from Soljic (2012). EC1.5 primers from Sprunck et al. (2012). qRT-PCR primers for AGO9 from Olmedo-Monfils et al. (2010).

			Fragment size (bp)	
Primer name	**Accession**	**Sequence (5' to 3 ')**	**gDNA**	**cDNA**
ACT3 for	AT3G53750	GATTTGGCATCACACTTTCTACAATG	740	657
ACT3 rev		GTTCCACCACTGAGCACAATG		
qAGO1 for	AT1G48410	GGCACTTCTCGACCTGCTCAT	221	120
qAGO1 rev		AACTGAGCGTGTGCATCTTG		
qAGO2 for	AT1G31280	ACTTTAAGCAGCCGCGGGGA	215	215
qAGO2 rev		CTGCGTTATTGGGGCGTTCCGA		
qAGO3 for	AT1G31290	GGCAATGTGCCTTCAGGTACGGT	315	315
qAGO3 rev		CGCTCCACGCGACTGCTTAGAATT		
qAGO4 for	AT2G27040	AAGCACCAGTGCCATTTCTG	445	357
qAGO4 for		AACCCCAACAGGCAAACAAG		
qAGO5 for	AT2G27880	GAACAAGCAGGCCGGCACAT	224	135
qAGO5 rev		GCTGGTGGCACAATTGACACAGA		
qAGO6 for	AT2G32940	CGTAGCACAACTGCAACTTC	217	341
qAGO6 for		GAGGAGGGTTTGGGTTTAGAC		
qAGO7 for	AT1G69440	CACGAGCAGGCCAACGCATT	170	170
qAGO7 rev		AGCCTTCCTCTGTACGCAGCAAG		
qAGO8 for	AT5G21030	CTTTGTGCTCATGCTGGAAAG	NA	132
qAGO8 rev		GCTCCGCTGGTAGACATAGG		
qAGO9 for	AT5g21150 (qRT-PCR)	TCCAGTCCACACGATAGCT	NA	150
qAGO9 rev		CCCACAAAAACCGACAGAAT		
AGO9 Exon 5-6 FP	AT5g21150 (RT-PCR)	GAATGGACTGGAACAAGGCTCGAC	NA	105
AGO9 Exon 6 RP		GTGGACTAAGTGAACACAGCTGCA		
qAGO10 for	AT5G43810	GTTATACCTATGCGCGGTGCACT	172	82
qAGO10 rev		CGTGCTCGAAATGCTGCAAG		
qAT5G42470 for	AT5G42470	ACGCATGGCTGAGCGACTGT	165	84
qAT5G42470 rev		TGTTGGAGCAGAGCTTCGTTACA		
qCHR34 for	AT2G21450	CAAGCTCGTAGCTGCGGATT	153	153
qCHR34 rev		GCATCACCGGAGTGATCAG		
qCHR40 for	AT3G24340	CGCATATCCGAACTGGTCTTCT	113	113
qCHR40 rev		TTCAGCTTCTCGTGCCGAAC		
qEC1.5 for	AT5G64720	GCGCCGGAAACTTGATGGACT	277	277
qEC1.5 rev		GGCGCCGGTGAAGGAGATAAT		
qeIF4G for	AT3G60240	CGGCGATGTTCTTGGGAGTG	124	124

			Fragment size (bp)	
Primer name	**Accession**	**Sequence (5' to 3 ')**	**gDNA**	**cDNA**
qeIF4G for		CCGGTTAGGTGCATGAGGTT		
qDCL2 for	AT3G03300	CGCCTCTCATCATCTCCATAAGC	352	141
qDCL2 rev		GCGCCTGCTAGAGATTCTATCAC		
qNOP56 for	AT1G56110	CACGCAAGAACGTGGATGTA	179	90
qNOP56 rev		TCTTCACCGAGGCATCAACT		
qNRPE5-L for	AT3G54490	TGAGGAAGCATGCACTTGAG	207	127
qNRPE5-L for		ATCCCCAACAGGCTCTTTAC		
qSCPL14 for	AT3G12330	ATGGAGGCCATGGATGATACTCG	352	152
qSCPL14 rev		GACCACTTATCCACCTCTTGAAC		
qSPDS2 for	AT1G70310	GAAGCCAGTGAGTCTAATCGATAC	188	103
qSPDS2 for		GCAAGCAGAAAGCAGCTGAGTG		
qSWIB/MDM2 for	AT5G23480	GACCAGCCAATATGGACTGCTTCT	300	107
qSWIB/MDM2 rev		CCTTCAGTGATAGAGTCCGTGTTC		
qUBC21 for	AT5G25760	TTGTGCCATTGAATTGAACCC	700	140
qUBC21 rev		GTCCTGCTTGGACGCTTCAGTCTG		

Table 18: Probes for small RNA blots

Primer name	**Sequence (5' to 3 ')**
U6 snRNA	TCATCCTTGCGCAGGGGCCA
miR172 cd	CTGCAGCATCATCAAGATTCT
miR_16	TTTCCAAAAGTGTAGACAAAG
miR5653	GCCAACTCAACTCAACTCAACCCA
miR390ab	GGCGCTATCCCTCCTGAGCTT

9.3 Plasmids

9.3.1 Summary of used vectors

Table 19: Entry vectors for TOPO cloning and subcloning vectors

Insert	**Vector backbone**	**Plasmid**
AGO8 promoter	pENTR™/ D-TOPO®	pURM1
AGO9 promoter	pCR®-Blunt II-TOPO®	pURM41
AGO9 promoter	pENTR™/ D-TOPO®	pURM42
AGO8-4 cDNA	pENTR™/ D-TOPO®	pGRU1
AGO8-8 cDNA	pENTR™/ D-TOPO®	pGRU2

Insert	Vector backbone	Plasmid
AGO8-4* cDNA	pENTR™/ D-TOPO®	pURM10
AGO9 cds (+ STOP)	pENTR™/ D-TOPO®	pURM24
RKD2 cds (- STOP)	pENTR™/ D-TOPO®	pURM16
AGO9 minigene i6	pCR®-Blunt II-TOPO®	pURM46
AGO9 minigene i567	pCR®-Blunt II-TOPO®	pURM47
AGO9 minigene i11	pCR®-Blunt II-TOPO®	pURM52
AGO9 minigene i101112	pCR®-Blunt II-TOPO®	pURM53
mCherry cds	pENTR™/ D-TOPO®	pABB22

Table 20: Destination vectors and binary vectors for expression in plants.

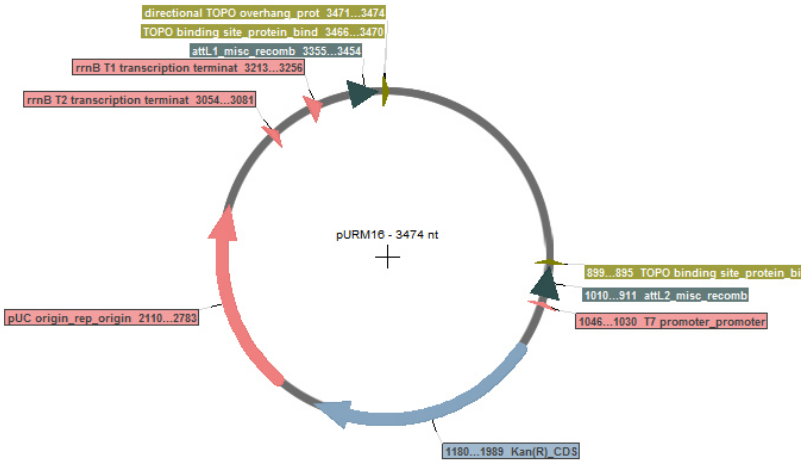
Insert	Vector backbone	Plasmid
AGO8 promoter	*pGW:NLS:3xEGFP::nost-pGII*	pURM2
AGO9 promoter	*pGW:NLS:3xEGFP::nost-pGII*	pURM43
AGO8 promoter	pB7WGF2.0	pURM3
AGO9 promoter	pB7WGF2.0	pURM44
AGO8-4 cDNA (pGRU1)	pB7WGF2.0	pGRU5
AGO8-8 cDNA (pGRU2)	pB7WGF2.0	pGRU6
AGO8-4 cDNA (pGRU1)	pURM3	pURM4
AGO8-8 cDNA (pGRU2)	pURM3	pURM5
AGO8-4* cDNA	pURM3	pURM12
AGO9 cds (+ STOP)	pB7WGF2.0	pURM28
AGO9 cds (+ STOP)	pURM3	pURM33
AGO9 cds (+ STOP)	pURM44	pURM45
RKD2 cds (- STOP)	pH7FWG2.0	pURM18
AGO9 minigene i6	pABB22	pURM48
AGO9 minigene i567	pABB22	pURM49
AGO9 minigene i11	pABB22	pURM54
AGO9 minigene i101112	pABB22	pURM55
AGO9 minigene i6:mCherry	pURM44	pURM50
AGO9 minigene i567:mCherry	pURM44	pURM51
AGO9 minigene i11:mCherry	pURM44	pURM56
AGO9 minigene i101112	pURM44	pURM57

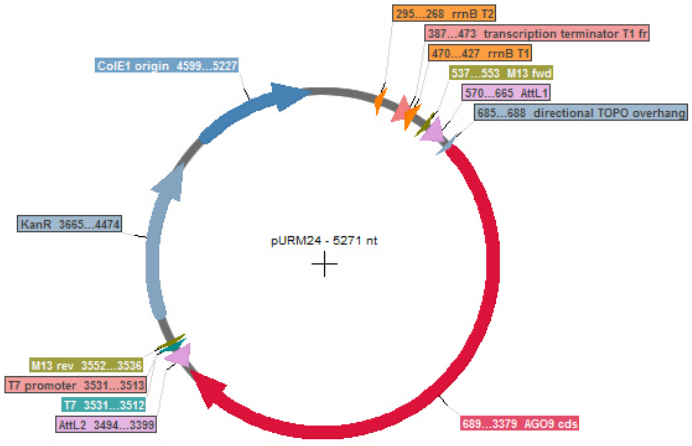
Insert	Vector backbone	Plasmid
EC1.1 promoter/AGO9 cds	pB7WGF2.0	pURM70
EC1.1 promoter/TET9 cds	pB7WGF2.0	pURM71

9.3.2 Plasmid sequences

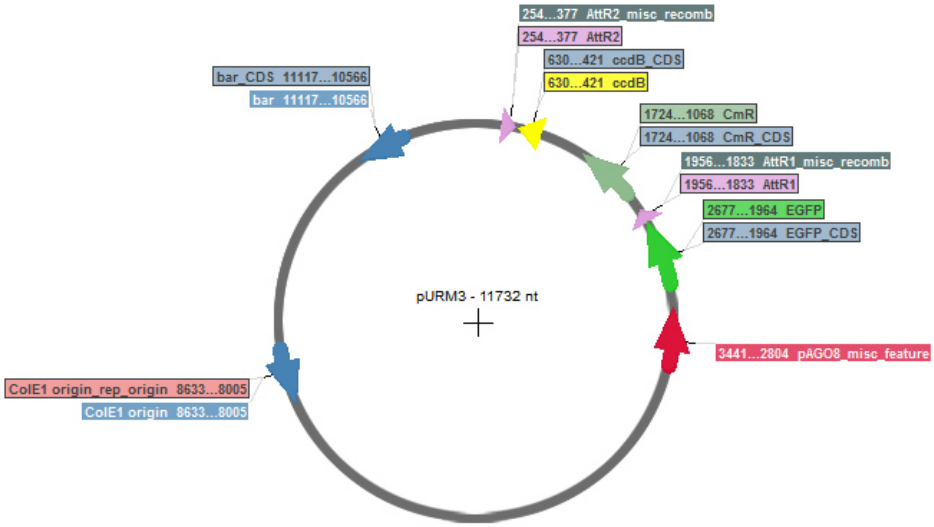
All Plasmids cloned and used during this thesis can be viewed as standard genebank files on the attached CD. Due to the number of cloned plasmids, only the most important vectors are shown below.

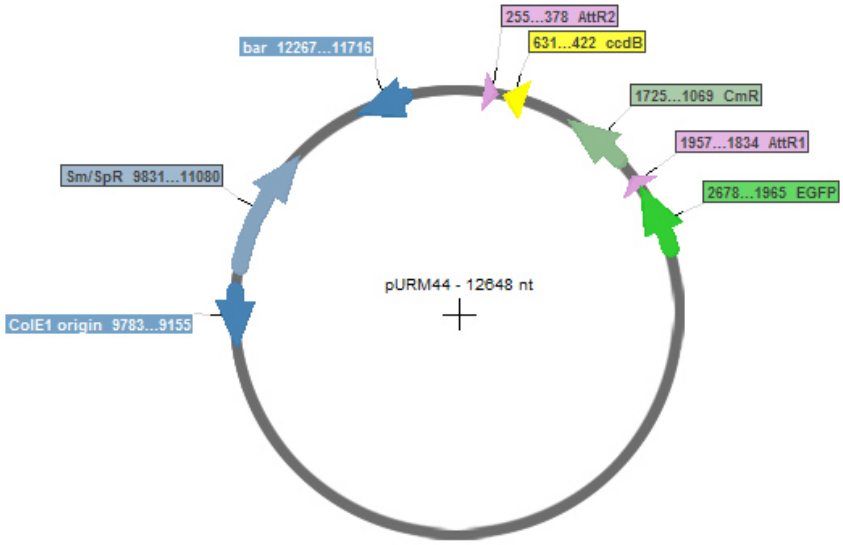
Entry Vectors for TOPO cloning



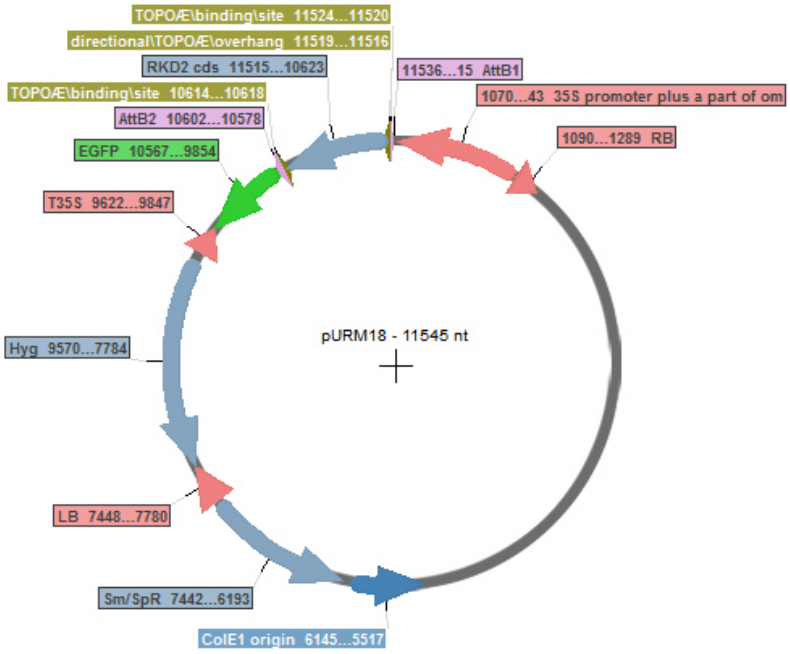


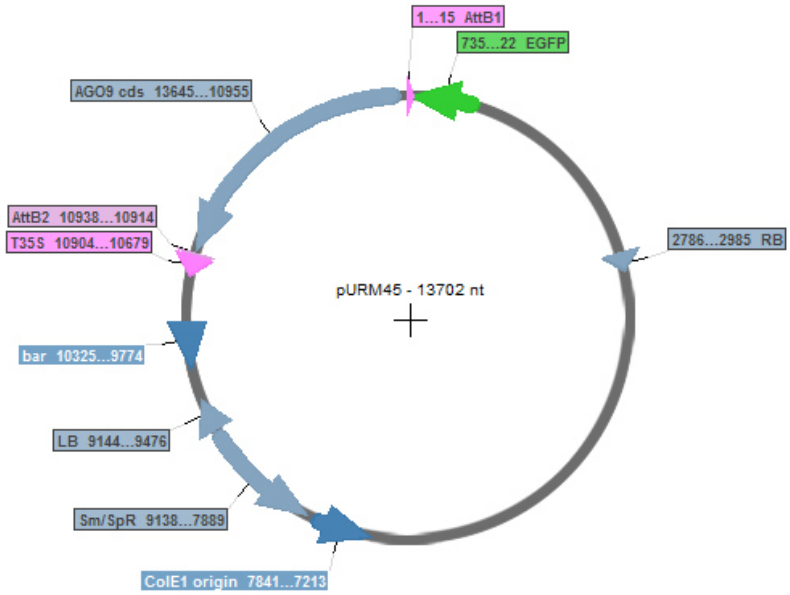
Destination vectors for gateway cloning





Binary vectors for expression in plants





The *pGW:NLS:3xEGFP::nost-pGII* expression vector was kindly provided by Mily Ron. Cloning of pURM70 and pURM71 was performed by Monika Kammerer.

9.4 RNA sequencing results

9.4.1 mRNA sequencing data

Log2 fold changes and mean expression values of all genes expressed in CIM and RKD2-induced callus are available in excel file format on the attached CD.

Table 21: Expression and log2 fold changes (RKD2/CIM) of genes known to be involved in sRNA pathways.
Shown are normalized counts in CIM and RKD2-induced cell lines and the log2 fold change (RKD2 vs. CIM). For gene description see Figures 7 and 8.

AGI identifier	Description	CIM (counts)	RKD2 (counts)	log2 FC (RKD2/CIM)	Function
AT5G44200	CBP20	3745.37	11432.81	1.61	mRNA and miRNA processing
AT2G13540	CBP80	2830.67	2220.89	-0.35	mRNA and miRNA processing
AT3G42670	CLSY (CHR38)	1261.45	46.55	-4.76	siRNA biogenesis
AT1G69770	CMT3	1768.14	7847.50	2.15	CHG methylation of DNA, Maintainance of methylation
AT1G01040	DCL1	2173.98	1079.48	-1,01	miRNA biogenesis
AT3G00330	DCL2	3965.18	1138.70	-1.80	siRNA biogenesis
AT3G43920	DCL3	1161.26	10.56	-3.40	siRNA biogenesis
AT5G20320	DCL4	2647.58	1269.86	-1.06	ta-siRNA biogenesis
AT5G66750	DDM1	3988.11	3799.23	-0.07	Chromatin remodeler, needed for DNA methylation
AT5G04560	DME	3837.51	543.43	-2.82	DNA glycosylase
AT2G36490	DML1 (ROS1)	815.49	170.25	-2.26	DNA demethylation
AT3G49250	DMS3 (IDN2)	1141.84	1386.42	-0.65	*de novo* methylation (RdDM)
AT3G26932	DRB3	34.45	358.10	5.11	dsRNA binding
AT5G15380	DRM1	2.14	792.88	8.53	Methyltransferase, de novo methylation (RdDM)
AT5G14620	DRM2	1829.51	2029.97	0.15	Methyltransferase, de novo methylation (RdDM)
AT3G17310	DRM3	2111.68	1246.94	-0.76	Methyltransferase, de novo methylation (RdDM)
AT1G65470	FAS1	1122.79	3475.17	1.63	Chromatin formation, DNA methylation
AT5G63110	HDA6	2512.84	2461.12	-0.03	Histone deacetylase
AT5G61070	HDA18	526.65	6939.87	3.72	Histone deacetylase

AGI identifier	Description	CIM (counts)	RKD2 (counts)	log2 FC (RKD2/CIM)	Function
AT1G09700	HYL1	1599.84	3405.64	1.09	Binds dsRNA, miRNA biogenesis
AT3G48670	IDN2	10128.37	2676.47	-1.92	RdDM
AT1G15910	IDP1	4193.50	1238.12	-1.76	RdDM
AT4G00380	IDP2	285.41	5318.89	4.22	RdDM
AT5G49150	MET1	2838.62	2722.98	-0.06	CpG methylation of DNA, Maintainance of methylation
AT1G08060	MOM1	1416.43	432.57	-1.90	Transcriptional silencer
AT4G36290	MORC1	1725.07	1643.47	-0.07	RdDM
AT4G36280	MORC2	94.43	4014.83	5.41	RdDM
AT1G19100	MORC6	1092.61	971.15	-0.17	RdDM
AT2G16485	NERD	3889.26	1504.72	-1.37	RdDM
AT3G18090	NRPD2b	278.61	88.77	-1.65	POL IV subunit
AT2G15400	NRPE3B	277.94	2625.98	3.24	POL V subunit
AT3G54490	NRPE5-L	80.07	13791.76	7.46	POL V subunit
AT1G31280	RDM1	1113.81	4644.45	2.06	RdDM, needed for Pol V transcription
AT4G11130	RDR2	1783.74	1136.74	-0.65	RdDM
AT3G49500	RDR6	2054.18	783.80	-1.39	ta-siRNA biogenesis
AT3G15390	SDE5	728.93	263.13	-1.47	ta-siRNA biogenesis
AT3G50100	SDN1	322.24	380.56	0.24	Degrades single-stranded small RNAs
AT2G27100	SE	4642.16	6749.58	0.54	miRNA biogenesis
AT5G23570	SGS3	1894.73	960.59	-0.98	ta-siRNA biogenesis
AT5G04290	SPT5L	3812.61	2909.51	-0.36	RdDM
AT3G33290	SUVH2	579.65	490.81	-0.24	histone methyltransferase
AT5G13960	SUVH4	2100.79	5429.93	1.37	histone methyltransferase
AT2G35160	SUVH5	497.06	146.76	-1.76	histone methyltransferase
AT2G27740	SUVH6	541.85	657.91	0.28	histone methyltransferase
AT4G13460	SUVH9	1719.19	1749.19	0.09	histone methyltransferase
AT3G49600	UBP26	2123.24	1307.00	-0.70	RdDM

Table 22: Log2 fold changes (RKD2/CIM) of genes validated by qRT-PCR compared to RNA-seq data.

AGI identifier	Description	qRT-PCR: log2 FC (RKD2/CIM)	RNA-seq: log2 FC (RKD2/CIM)
AT5G42470	-	0.63	0.00
AT1G48410	AGO1	0.18	0.41
AT1G31280	AGO2	-1.25	-1.28
AT1G31290	AGO3	-2.54	-3.20
AT2G27040	AGO4	0.55	-0.44
AT2G27880	AGO5	13.92	11.03
AT2G32940	AGO6	-0.40	0.38
AT1G69440	AGO7	0.87	0.89
AT5G21030	AGO8	17.00	10.28
AT5G21150	AGO9	13.15	11.24
AT5G43810	AGO10	-2.54	-1.57
AT2G21450	CHR34	12.18	10.44
AT3G24340	CHR40	5.58	5.18
AT3G03300	DCL2	-1.00	-1.80
AT5G64720	EC1.5	11.27	10.59
AT1G56110	NOP56	0.14	0.00
AT3G54490	NRPE5-L	10.02	7.42
AT3G12330	SCPL14	-15.76	-9.27
AT1G70310	SPDS2	-1.43	0.00
AT5G23480	SWIB/MDM2	3.51	7.46

9.4.2 Small RNA sequencing data

Log2 fold changes and mean expression values of small RNAs in CIM and RKD2-induced callus are available in excel file format on the attached CD. Target predictions of known and novel miRNAs are enclosed as separate excel files.

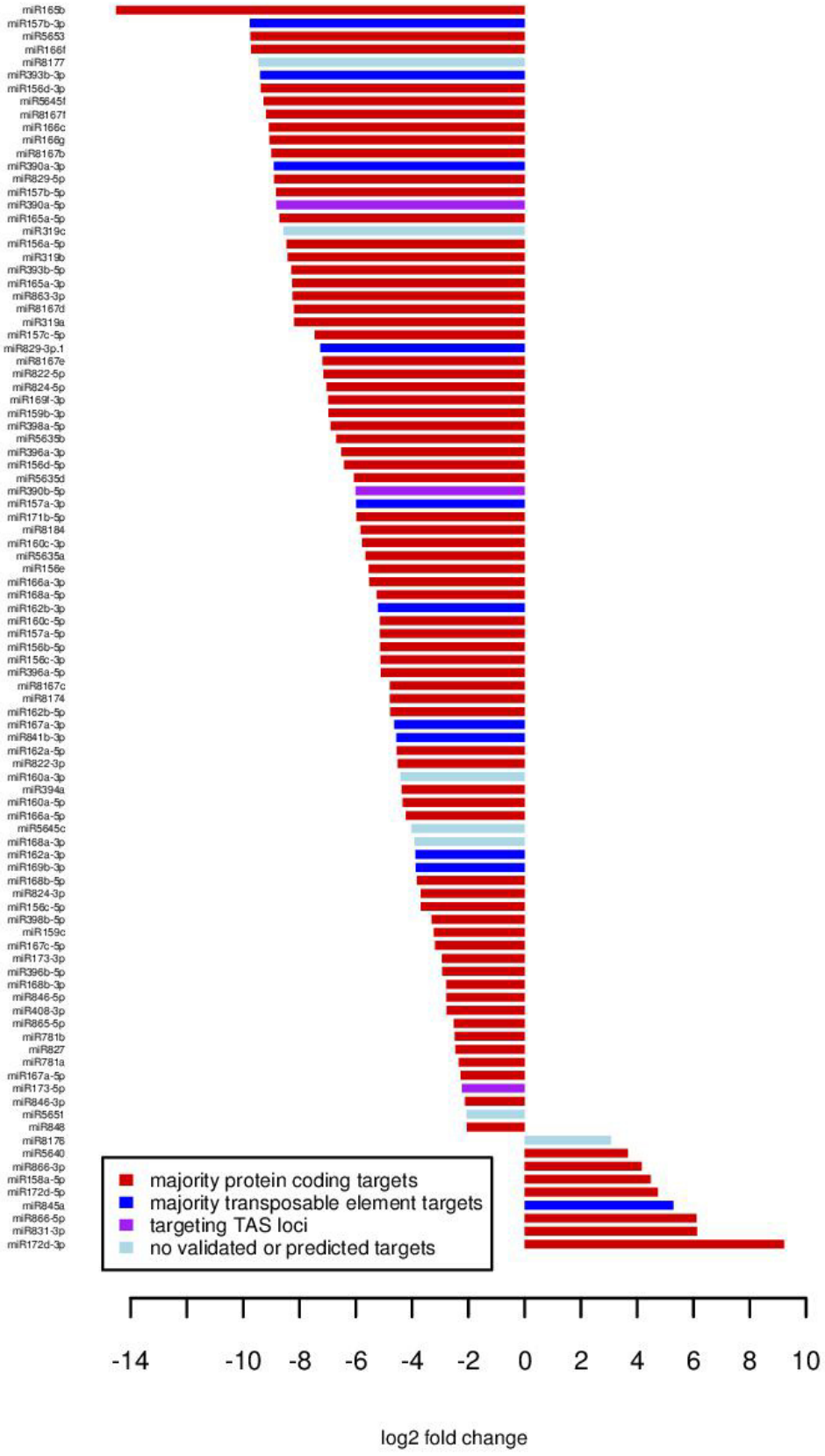


Figure 60: Annotated miRNAs with different expression levels in RKD2 and CIM callus. Significant differentially expressed miRNAs with a log2 fold change larger than two are shown. Bars are colored according to the majority of the targets falling either within protein coding genes, transposable element genes or TAS loci.

DANKSAGUNG

Ich bedanke mich bei Herrn Prof. Thomas Dresselhaus für die freundliche Aufnahme an seinem Lehrstuhl und die Möglichkeit diese Arbeit dort durchführen zu können.

Bei meiner Betreuerin PD Dr. Stefanie Sprunck bedanke ich mich dafür, dass ich an diesem interessanten Thema arbeiten durfte, für jede Menge fachliche Unterstützung und wertvollen Input und für das Korrekturlesen meiner Arbeit.

Dank geht auch an Prof. Dr. Klaus Förstemann für die Betreuung als externen Mentor und die Übernahme des Koreferats.

Bei PD Dr. Joachim Griesenbeck bedanke ich mich für die Bereitschaft, als Drittprüfer zu fungieren und bei Prof. Dr. Klaus D. Grasser für die Übernahme des Prüfungsvorsitzes.

Ich danke der DFG, die diese Arbeit in Rahmen des Sonderforschungsbereichs 960 finanziell gefördert hat.

Dr. Christoph Möhle und seinem Team am Kompetenzzentrum Fluoreszente Bioanalytik (KFB) hatten einen großen Anteil an dieser Arbeit durch die Erstellung der mRNA und sRNA-Bibliotheken, dafür vielen Dank.

Ein besonderer Dank geht an Dr. Julia Engelmann und Dr. Nicolas Strieder vom Institut für funktionelle Genomik, Abteilung Bioinformatik für die Hilfe bei der bioinformatischen Auswertung der RNAseq Ergebnisse.

Für die Unterstützung bei der Durchführung der Northern Blots bedanke ich mich bei Daniele Hassler, Judith Hauptmann und Brian Bitsch Sörensen.

Ich bedanke mich bei den vielen Menschen, die im Rahmen von Master- und Bachelorarbeiten, als Praktikant, TA oder wissenschaftliche Hilfskraft zu diesem Projekt beigetragen haben. Als da wären (in keiner besonderer Reihenfolge): Phillip Cyprys, Maura John, Joiselle Blanche Fernandes, Nina Wild und Monika Kammerer. Ein besonderer Dank geht an Ingrid Fuchs, die mir sehr geholfen hat.

Ich bedanke mich bei allen Mitarbeitern am Lehrstuhl für Zellbiologie und Pflanzenbiochemie für die freundliche und produktive Atmosphäre. Insbesondere sei meinen Kollegen im Sprunck-Lab gedankt : Svenja, Ulrike, Lucia, Phil, Frank, Maria, Karin und Tom. Danke Maria und Frank fürs Korrekturlesen meiner Arbeit.

Ein unschätzbarer Dank an Melanie, die mich immer unterstützt und an mich geglaubt hat.