CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL
IRAPUATO

Regulación transcripcional de miRNAs por el módulo CDK8 del Mediador en *Arabidopsis thaliana*

Tesis que presenta

IBQ Joel Rodríguez-Medina

Para obtener el grado de

Maestría en Ciencias en Biología Integrativa

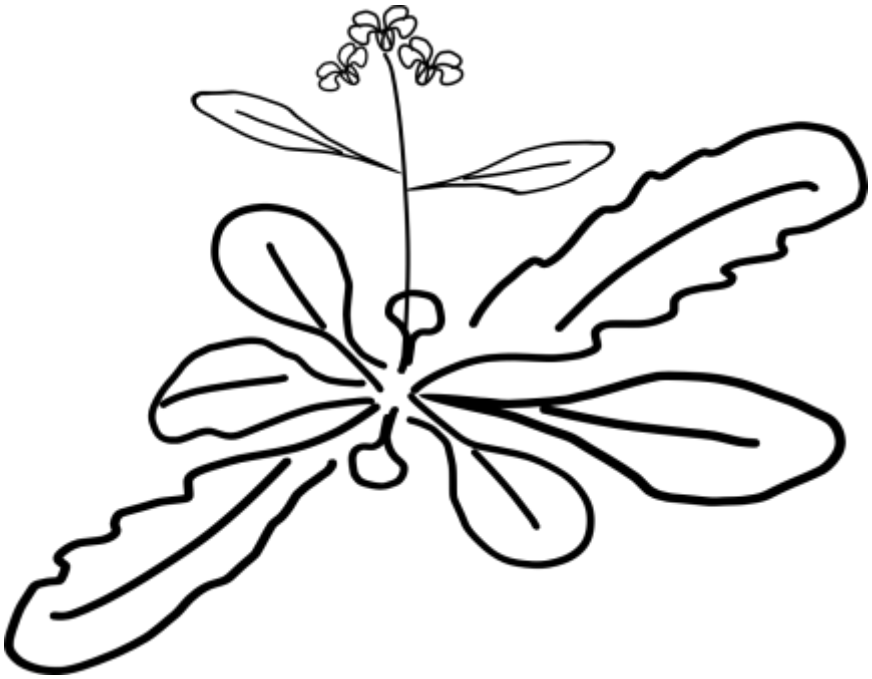
Directores de tesis:

Dr. Charles Stewart Gillmor III

Dr. Cei Leander Gastón Abreu Goodger

Irapuato, Guanajuato Agosto, 2015

This study was done in the RNA Computational Genomics and the Plant Development and Morphogenesis laboratories at LANGEBIO, Cinvestav in Irapuato.



Dedication

To my father Joel,

my mother Raquel,

and my sister Sarón.

For their constant love and support.

Acknowledgments

I want to thank God, my father Joel, my mother Raquel and my sister Sarón, for everything; for their love and support.

To my advisors, Dr. Stewart Gillmor and Cei Abreu: for their patience and continuous guidance and motivation.

To my committee, Drs. Alfredo Cruz and Paco Barona, for their valuable comments and discussions.

To Alma, Marce, Roberto and Cesare: for all the constant discussions and help.

To Alex, Beto, Carol, Fibo, Felipe and Jhon: for our journey together in this adventure.

To both, the Embryo and RNA labs: for their friendship and partnership.

To Lily, Sara, Emmy, Yuli, Claudia, Christian David and Diana; to my family and friends, whose names may not be here but their friendship is not forgotten.

To all of you, for when our paths split, may our memories hold us together.

Contents

Figures

Tables

List of Abbreviations.

- **RISC.** RNA Induced Silencing Complex
- **RNA pol II**. RNA polymerase II
- **sRNA**. small RNA
- **miRNA**. microRNA.
- **PTGS**. Post-transcriptional gene silencing.
- **PIC**. pre-initiation complex.
- **TF**. Transcription factor.
- **nt**. nucleotide

Resumen

El Mediador es un complejo multiprotéico conservado evolutivamente que actúa como regulador transcripcional en eucariotas (Malik and Roeder 2010). El Mediador Principal está formado por módulos (Cabeza, Mitad y Cola) además de un cuarto módulo desacoplable llamado el módulo CDK8 (Ciclina Dependiende de Cinasa 8, por sus siglas en inglés). El módulo CDK8 está compuesto por cuatro subunidades codificadas por los genes MED12, MED13, HEN3 y CYCC. El módulo está involucrado principalmente en la regulación negativa de la transcripción (Björklund and Gustafsson 2005). La regulación negativa se da por la inhibición de la interacción entre el Mediador Principal y la RNA pol II (Tsai et al. 2013). En *Arabidopsis thaliana* el módulo CDK8 participa en varios programas claves para el desarrollo incluyendo embriogénesis (Gillmor et al. 2010), desarrollo vegetativo (Conaway and Conaway 2011; Kidd et al. 2011), y floración (Imura et al. 2012).

Resultados previos de nuestro laboratorio demuestran que el módulo CDK8 regula los niveles de miR156. Para investigar el papel general de las subunidades del módulo en la transcripción de RNAs pequeños (sRNAs) realizamos sequenciación masiva de sRNAs de plantas con genotipo silvestre y mutantes sencillas para *med12*, *med13*, *hen3* y una doble mutante *med12;hen3* a 18 días.

Entre los sRNAs se incluyen miRNAs que participan en respuestas a estrés y otros que son parte del control de desarrollo. La familia miR166 se encontró reprimida en las mutantes; estos miRNAs participan en la regulación de la identidad abaxial/adaxial en hojas al restringir espacialment a sus blancos (los factores de transcripción del tipo HD-ZIP III). Por el contrario, la familia miR164 se encontraba sobre-expresada en las mutantes. Se sabe que miR164A controla la morfología de la hoja regulando el gen *CUC2* para determinar las dentaciones en el margen de la hoja durante el desarrollo vegetativo.

Nuestro análisis a nivel genómico de la transcripción de miRNAs respalda el papel del módulo CDK8 del Mediador en la transcripción de miRNAs que son importantes para el desarrollo al controlar su expresión espacio-temporal. Además, sugiere una participación del módulo en respuesta a estreses abióticos.

Como una aproximación para validar nuestros resultados bioinformáticos generamos líneas marcadoras para investigar el patrón de expresión de MIR164A y cómo es reprimido por el gen *CCT*. Para este propósito utilizamos marcadores transcripcionales para explorar las diferencias en la expresión entre plantas silvestres y mutantes utilizando herramientas genéticas disponibles y además creando un marcador fluorescente para investigar dicho patrón *in vivo*. Resultados preliminares sugieren que *CCT* reprime la expresión de miR164A de manera espacial y temporal, sin embargo se requieren de análisis más detallados para confirmar esta observación.

Abstract

Mediator is an evolutionarily conserved multiprotein complex that is a key transcriptional regulator in eukaryotes (Malik and Roeder 2010). The complex consists of three modules of Core Mediator (Head, Middle and Tail); plus a fourth, detachable module called the Cyclin Dependent Kinase 8 (CDK8). The CDK8 module is composed by four subunits coded by the genes MED12, MED13, Cyclin C and HEN3 and is mainly involved in negative regulation of transcription (Björklund and Gustafsson 2005). The CDK8 module regulates transcription initiation in a negative way by inhibition the interaction between Core Mediator and the RNA pol II (Tsai et al. 2013). In *Arabidopsis thaliana* the CDK8 module participates in several key developmental programs, including embryogenesis (Gillmor et al. 2010), vegetative development (Conaway and Conaway 2011; Kidd et al. 2011), and flowering (Imura et al. 2012).

Previous results from our laboratory demonstrate that the CDK8 module regulates miR156 levels (Gillmor et al. 2014). To investigate the general role of CDK8 module subunits on small RNA (sRNA) transcription, we performed high-throughput sequencing of sRNAs from Arabidospsi from wild-type, *med12*, *med13*, *cdk8*, and a *med12;cdk8* double mutant genotypes at 18 days old.

Deregulated sRNAs include miRNAs that participate in stress response and others that are part of developmental control. The miR166 family was repressed in the mutants; these miRNAs participate in the regulation of abaxial/adaxial identity of leaves by spatially restricting

their targets (transcription factors of the type HD-ZIP III). Conversely the miR164 family was upregulated in the mutants; miR164A is known to control leaf morphology by targeting the *CUC2* gene to control margin serrations during vegetative development.

Our genome level analysis of microRNA transcription supports a role for the CDK8 module of Mediator in the transcription of miRNAs important for development by controlling their spatial and temporal expression. It also suggests an involvement of the CDK8 module in the response to abiotic stresses.

As an approach to validate our bioinformatics results we generated marker lines to investigate the expression pattern of MIR164A, and how it is repressed by the *CCT* gene. For this purpose we used transcriptional markers to explore differences in the expression between wild type and *cct* mutants using available genetic tools and further creating a fluorescent marker to look at the pattern *in vivo*. Preliminar results suggest that *CCT* represses expression of miR164A in a spatial and temporal manner; however more detailed analyses are needed to confirm this observation.

Chapter I Introduction and Background

I.I Introduction

I.I.1 Importance of Genetic Control

Multicellular organisms form by the process of development. Although every cell in an organism carries the same genetic information in its DNA, differences in gene expression determine the identity of each cell. Genetic control of cell fate specification at specific times and places is crucial for development (Benfey and Weigel 2001; Levine and Davidson 2005; Riechmann 2002).

Regulation of gene expression can occur at different stages. At the level of DNA, epigenetic modifications can alter the expression of a gene without affecting its sequence. This is achieved through changes in the methylation states of the DNA or by compacting the chromatin, silencing genes by making them inaccessible to RNA polymerases. At the RNA level, controlling the transcription from DNA to RNA and, at the protein level, translation is regulated by determining how and when a protein is made (Alberts B, Johnson A, Lewis J, Raff M, Roberts K 2002).

I.I.2 Transcriptional Regulation

The Central Dogma of Molecular Biology describes the direction of genetic information flow. Figure 01 exemplifies this: information stored in the form of DNA can undergo two processes, replication or transcription. Replication copies the DNA to ensure its propagation. During transcription, DNA is converted into RNA, which serves as an intermediary for the production of proteins during translation. In some cases, RNA can be retrotranscribed to DNA, however proteins cannot be converted back to their nucleic acid counterparts (Harvey Lodish, Arnold Berk, Paul Matsudaira, Chris A. Kaiser 2003).

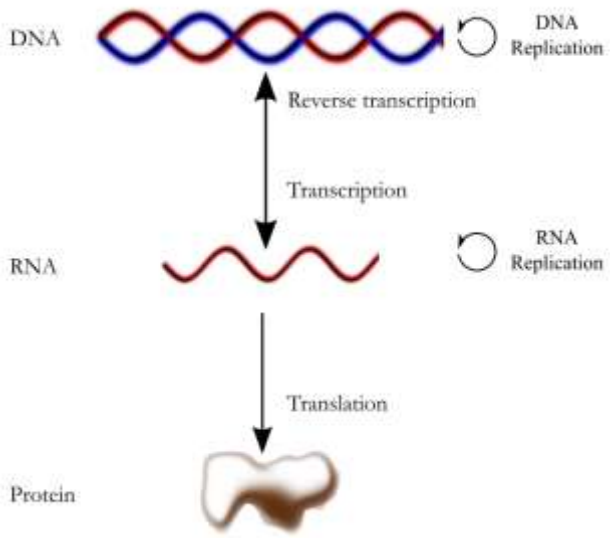


Figure 1. Conceptual representation on the Central Dogma of Molecular Biology.

I.I.3 Coding and non-coding genes

Genes can be classified into two major groups: coding and non-coding. This denotes whether a gene will be translated into a protein (i.e. protein coding) or not.

In eukaryotes, transcription is a complex process as it can be carried out by different polymerases. RNA pol I and III mainly transcribe genes encoding ribosomal and transfer RNAs, respectively. All protein coding genes are transcribed by RNA pol II. In plants, two specific RNA polymerases, pol IV and V, participate in gene silencing and genome protection mediated by RNA (Haag and Pikaard 2011). Initiation of pol II mediated transcription, represented conceptually in figure 02, involves a series of steps in which several proteins and enzymes are recruited to the regulatory regions (promoters) of genes in order to transcribe the information encoded in the DNA to RNA.

I.I.4 Transcription initiation

Transcription factors (TF) are proteins with the ability to recognize and bind promoters upstream of specific genes, to induce transcription. They can function either as activators or repressors. When functioning as an activator, a TF recruits the pre-initiation complex (PIC),

formed by a set of General Transcription Factors (GTFs) which helps the RNA polymerase (RNA pol) settle at the promoter of a gene and initiate transcription. A major component in the initiation of eukaryotic transcription is the Mediator complex; it promotes transcription by acting as a bridge between gene specific transcription factors and RNA pol II (Harvey Lodish, Arnold Berk, Paul Matsudaira, Chris A. Kaiser 2003; Latchman 2005).

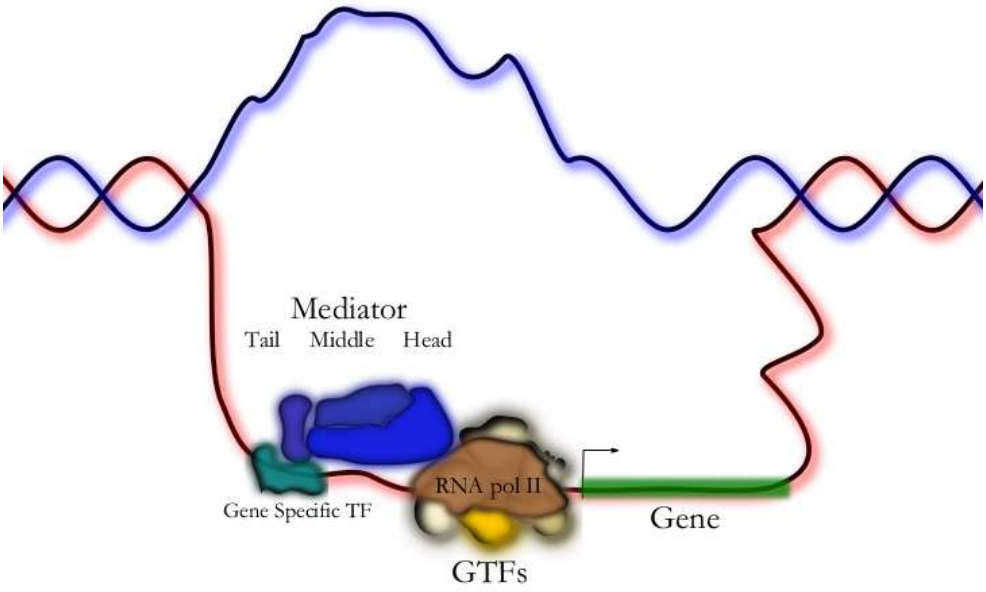


Figure 2. Conceptual representation of transcription initiation. A TF binds to the regulatory region of the gene. The GTF's recruit the RNA polymerase II to the starting site of transcription. Mediator acts as a bridge between the gene specific TF and the RNA pol II to initiate transcription.

I.I.5 The Mediator Complex is required for activated transcription

Mediator is a key regulatory element in RNA pol II mediated transcription. First discovered in yeast for its ability to stimulate transcription *in vitro*, it was subsequently identified in a wide range of eukaryotes, including humans and plants. This discovery led to the Nobel prize of chemistry in 2006 as an advance in our understanding of eukaryotic transcription (Conaway and Conaway 2013; Flanagan et al. 1991; Kidd et al. 2011; Tóth-Petróczy et al. 2008).

Core Mediator is composed of about 30 proteins. The actual number of subunits varies between organisms. The core complex is divided in three segments or modules: Head, Middle and Tail. A fourth segment, called the CDK8 module, mainly acts as a negative regulator of the

activity of the Core Mediator; it has a dynamic interaction with the core Mediator as it can be found either attached or free from the complex (Elmlund et al. 2006; Larivière, Seizl, and Cramer 2012). Evidence in Drosophila and human cell lines has shown that in certain circumstances the CDK8 module can act as a transcriptional activator (Carrera et al. 2008; Donner et al. 2007; Rau, Fischer, and Neumann 2006).

Given its large size, analysis of the Mediator as a whole is a difficult task; its function has been studied through analysis of the loss of function of individual components. Studies show that individual subunits are required for correct development in model organisms such as mice, *C. elegans*, *Drosophila* and zebrafish (Carrera et al. 2008; Malik and Roeder 2010). Mediator complex also plays a crucial role in plants: mutations in different subunits affect developmental processes as well as hormone and stress responses in Arabidopsis (Gillmor et al. 2010, 2014; Hentges 2011; Kidd et al. 2011).

The negative regulatory submodule of Mediator is called the CDK8 module or Kinase module (figure 03). It is formed by four different proteins. Their Arabidopsis names are *CENTER CITY (CCT)*, *GRAND CENTRAL (GCT)*, *HUA ENHANCER 3 (HEN3)*, and *CYCLCIN C (CYCC)* (corresponding to MED12, MED13, CDK8 and CYCC, according to the unified nomenclature for Mediator). When attached, the Kinase module blocks the interaction of Mediator with RNA pol II (Elmlund et al. 2006).

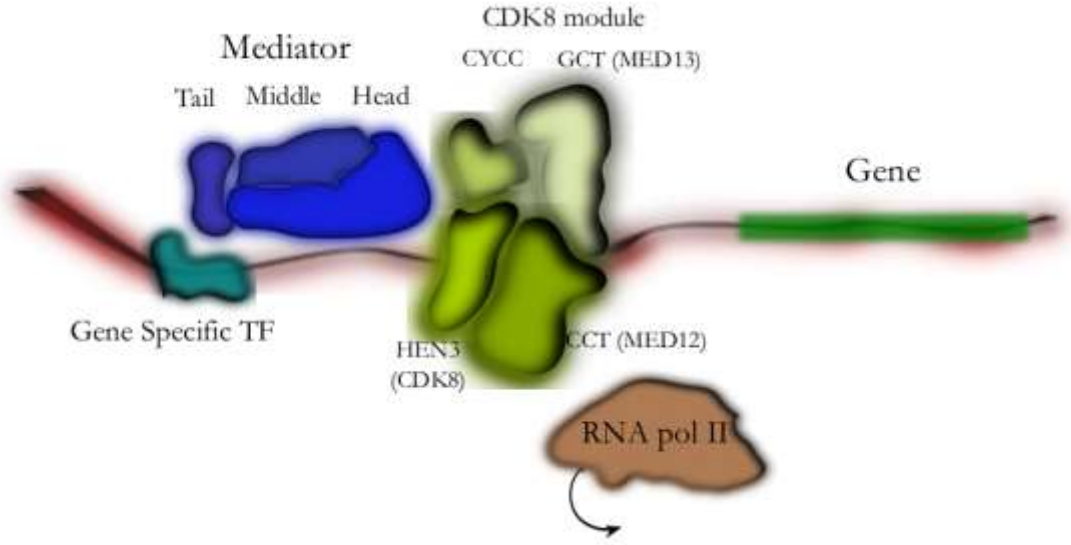


Figure 3. The CDK8 module blocks the interaction of the RNA pol II with the Mediator complex thus inhibiting transcription. It is composed by four subunits.

I.I.6 Discovery of regulatory small RNAs

In 1993, the groups of Victor Ambros and Gary Ruvkun described *lin-4* in *C. elegans,* a non-coding gene that is crucial for the temporal control of development. They reported that this gene negatively controls expression of the TF *lin-14*; although the transcript levels remained the same, its protein levels decreased over time. Thus, regulation occurs at the posttranscriptional level. Correct expression of lin-4 promotes the transition between L1 and L2 larval developmental stages. The product of the gene is a small RNA (conserved and known in other species as miR125) of 22 nucleotides (nt) that is complementary to several regions of the 3' untranslated region (UTR) of *lin-14*. This was the first report of a microRNA (miRNA) (R. C. Lee, Feinbaum, and Ambros 1993; Wightman, Ha, and Ruvkun 1993). In 1999 Hamilton and Baulcombe described small RNA molecules (of about 25 nt) that participate in the posttranscriptional silencing by two different cues in plants (Hamilton and Baulcombe 1999). These small RNAs were complementary to the sequence of either transgenes or virus mRNAs. In 2002, the Bartel group confirmed the presence of miRNAs producing loci in *Arabidopsis* (Reinhart et al. 2002).

Later, the elucidation of the RNAi mechanism by Andrew Fire and Craig Mello (Fire et al. 1998), proposed a new mode of genetic control: small transcripts, between 20-30nt long, that could post-transcriptionally regulate gene expression. This discovery won them the Nobel prize in Physiology or Medicine in 2006, for the advance in our understanding of how genes are regulated. The activity of these small RNAs (sRNAs) can range from effects on chromatin organization to transcriptional or post-transcriptional regulation of gene activity (van Wolfswinkel and Ketting 2010).

I.I.7 Biogenesis of microRNAs

Transcription of MIR genes, as opposed to that of other non-coding genes, is mediated by the RNA Pol II. This process is essentially the same as protein coding genes, in which the transcripts are capped and polyadenylated.

In Arabidopsis, miRNA biogenesis (represented in figure 04), starts with transcription of a *MIR* gene into a primary miRNA transcript (pri-miRNA). This transcript contains a region of

self-complementarity that forms a hairpin structure. The secondary structure of the pri-miRNA is recognized by a group of proteins that further process the transcript. SERRATE (SER) and HYPONASTIC LEAVES1 (HLY1) form a complex with the endonuclease DICER-LIKE1 (DCL1); the latter enzyme is in charge of cleaving the hairpin into a double stranded RNA molecule called a precursor miRNA (pre-miRNA). Until this point, pre-processing of the miRNA occurs inside the nucleus of the cell. To stabilize the pre-miRNA and protect it from degradation, it is methylated at both 3' ends by the HUA ENHANCER1 (HEN1) methyltransferase before being transported to the cytoplasm by the HASTY (HST) exportin (reviewed in Chen, 2009).

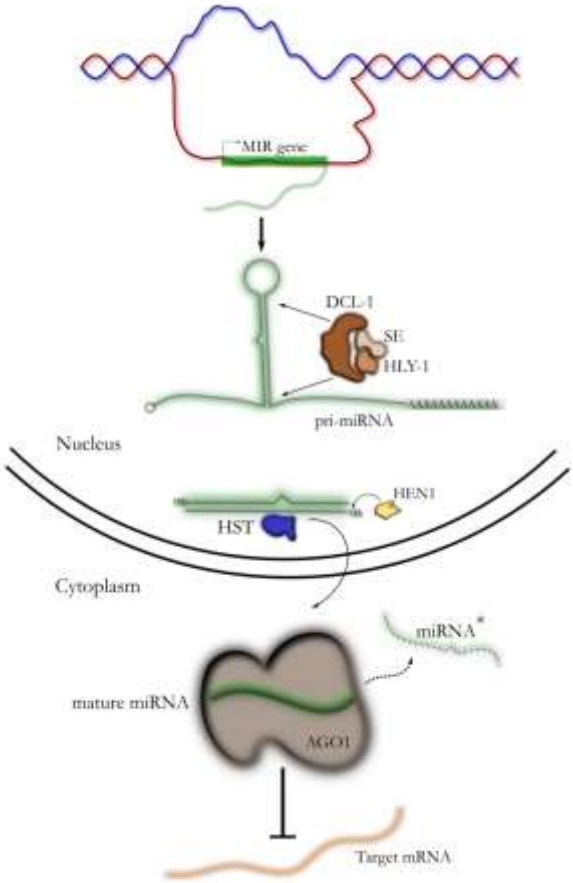


Figure 4. Biogenesis of miRNAs in plants.

The double stranded RNA molecule is usually 21 base pairs (bp) long; one of the strands is the mature miRNA while the other, known as the miRNA star strand (miRNA*), is eventually degraded. Once in the cytoplasm the duplex is separated and the mature miRNA is loaded into the ARGONAUTE1 (AGO1) protein forming the RNA-induced silencing complex (RISC). The function of this AGO1-microRNA complex (RISC) is to control gene expression by interfering with translation or by degrading their target mRNA in a process called post-transcriptional gene silencing (PTGS) (X. Chen 2005; K. Rogers and Chen 2012; Kestrel Rogers and Chen 2013).

Genetic studies point for a role of different subunits of the Mediator complex in the biogenesis of small RNAs (sRNAs) in *Arabidopsis*. Mutations in the subunits MED17, MED18, and MED20 resulted not only in less accumulation of microRNAs but also in developmental phenotypes such as organ abnormalities and a delay in flowering time. Reports also show that RNA pol II binds to regulatory regions of miRNAs and that this occupancy is significantly reduced when these Mediator subunits are absent (Kidd et al. 2011; Y. J. Kim et al. 2011). This proves that Mediator does interact with the pol II to promote transcription of MIR genes. MED12 and MED13 have also been shown to influence the level of miR156 and miR172 causing delays in developmental transitions (Gillmor et al., 2014). This work further explores other miRNAs that could be under regulation by subunits of the Arabidopsis CDK8 module.

I.I.8 Function of microRNAs

miRNAs act in *trans*, binding their targets in a sequence complementary fashion and inducing the translational repression or mRNA cleavage by the RISC. The miRNA sequence is partially complementary to a region of its targets. In animals, a small stretch of 6-8 nt, termed the *seed* region, is complementary to the mRNA. In plants, the sequence of the miRNA is almost a perfect match with its target (Mallory and Bouché 2008). miRNAs can be assigned to families, according to the similarity of their mature sequences, which often regulate the same targets.

Defects in the miRNA biogenesis pathway reveal their importance during development. One of the functions of miRNAs is to modulate the expression levels of target messengers and maintaining homeostasis in the cell. Some miRNAs can act as molecular switches to establish cell fate. In plants they mainly participate in developmental processes which are temporally

regulated and help to define spatial patterning. Many miRNA targets are transcription factors (Ebert and Sharp 2012; J. H. Kim et al. 2009; Mallory and Bouché 2008; Mallory and Vaucheret 2006). This study explores the role of subunits of CDK8 module in the production of miRNAs and how they relate to the phenotype observed in mutants of this module.

From the hundreds of miRNAs in Arabidopsis, only a few have a validated role during embryogenesis or post-embryonic development. Alterations in the function of distinct miRNAs or their targets have specific developmental implications (Mallory and Vaucheret 2006). The miR156 family controls the transition from vegetative to reproductive stage. This miRNA family is highly expressed during vegetative phase, when the plant produces leaves with juvenile traits. During growth, the levels of the miR156 mature sequence decrease, thus expression of its targets increase. miR172 acts downstream of this regulatory cascade to promote flowering (Wu et al. 2009).

Organs in plants are also produced post-embryonically; thus, maintenance of stem cell niches is crucial for growth and development. In Arabidopsis, the shoot apical meristem (SAM) is maintained by the expression of several TFs, including the *CUP SHAPED COTYLEDON* (*CUC)* genes. These are members of the NAC family which is targeted by miR164. *CUC* genes act redundantly during organ formation as repressors of growth, defining organ boundaries. In turn, the miR164 family represses *CUC* genes, limiting expansion of these boundaries. Absence of *CUC* expression, either by mutations or overexpression of MIR164 genes, results in failure to develop a SAM. Less severe phenotypes show an enlargement of the boundary domain as well as fused organs (Laufs et al. 2004; Sharma and Fletcher 2002).

Other miRNAs have a role in response to a variety of stresses such as temperature or low nutrient availability. Even miRNAs with a known role in development, such as miR156 and miR172, respond to temperature changes (Jones-Rhoades, Bartel, and Bartel 2006; H. Lee et al. 2010). This is an example of how different pathways are interconnected to confer robustness to biological processes.

I.I.9 Development of multicellular organisms

The life cycle of a complex organism starts with a process called embryogenesis. Here, a single totipotent cell called the zygote will produce a fully functional organism. Two important

programs are intertwined during embryo development: cell division and cell fate specification. It is not sufficient to produce more cells through the division of existing ones; they have to acquire their final identity to perform correctly. The synchronous action of different cell types allows the formation of tissues and organs for the appropriate functioning of a complex organism (Capron et al. 2009; Gilbert 2010; Scott 2000).

In animals, the final morphology of the organism is established during embryo development. Limbs are already specified and, in general, no development of new tissues or organs occurs, only maturation of the existing ones (Capron et al. 2009; Gilbert 2010). On the other hand, during plant embryogenesis only the basic body of the plant is specified. The balance between division and fate specification of cells patterns the embryo. Here, the apical and basal axes are defined, and they provide a supply of stem cells to form the major tissues and organs in the plant. All plant organs (roots, flowers, fruits, and leaves) are developed post-embryonically from root and shoot stem regions (Capron et al. 2009; Goldberg, de Paiva, and Yadegari 1994; ten Hove, Lu, and Weijers 2015; Lack and Evans 2001; Mayer et al. 1991; S. Park and Harada 2008).

Unlike animals, plant cells do not migrate, so their final identity within the embryo is specified by temporal and spatial cues. Each specialized cell type has unique functions for the organism. Cells can be defined by the specific set of genes expressed that allow them to carry out their functions. This differential expression of genes at specific cells occurs in all multicellular organisms, and involves coordination of specific transcriptional programs to control different aspects of cell biology such as division, fate and communication (Harvey Lodish, Arnold Berk, Paul Matsudaira, Chris A. Kaiser 2003; ten Hove, Lu, and Weijers 2015).

In plants and other multicellular organisms, genetic control of cell differentiation and organ development is important for embryo pattern formation and morphogenesis (Goldberg, de Paiva, and Yadegari 1994). This is achieved by integrating internal and external signals that control different aspects of gene expression, like the rate of transcription or the rate of translation from RNA to protein (Singh 1998). This highlights the importance of spatial and temporal regulation of specific transcription factors driving developmental processes (Bolouri and Davidson 2003).

I.I.10 Case Study: Leaf Development

"From first to last, the plant is nothing but leaf"
Wolfgang von Goethe (The Metamorphosis of Plants, 1679)

A fundamental question proposed by developmental biology is that of morphogenesis: how are different cells produced and organized to form precise functioning structures of organs and tissues? (Gilbert 2010).

Plants continuously produce organs as they develop (Geuten and Coenen 2013). Leaves are an example of this. They carry out important metabolic processes such as photosynthesis, and play important roles in sensing environmental cues such as photoreception and respiration. As time progresses, leaves change in their shape and size; after the vegetative to reproductive transition, plants produce, instead of leaves, flowers. They constitute the most important organs for several species of the diverse plant lineages.

The analysis of leaf development is not new; as early as the 18th century, Wolfgang von Goethe and other botanists began studying the morphological changes during plant growth. Through their observations, they hypothesized that flowers were modified leaves and they were not wrong. Genetic studies subsequently provided evidence that floral organs are indeed transformed leaves (Tsukaya 2002; Bowman, Smyth, and Meyerowitz 1991; von Goethe 2006 [originally, 1790]).

The first step in the making of a new leaf is the formation of organ primordia at the flanks of the shoot meristem; these leaf primordia will give rise to all the cell types that form part of leaves. The interplay of gene networks and plant hormones, such as auxin and cytokinins, drive the establishment of new cell types to form leaves. An increase in the concentration of the phytohormone auxin correlates with the initiation of organ primordia (Tsukaya 2002).

Spatiotemporal gene expression drives the specification of morphological traits and physiological functions of the leaf. For example, the specification of adaxial (upper) and abaxial (lower) polarity is key not only in development of the flat shape of the leaf but also in physiological processes; the adaxial layer plays important roles in photosynthesis, while the abaxial part specializes in gas exchange. KANADI and the HD-ZIP III genes specify these upper and lower sides and are mutually exclusive. Other genes such as miR166 and auxin

response factors also form part of the network driving the specification of leaf polarity (Barkoulas et al. 2007; Tsukaya 2002). PHABULOSA (PHB), PHAVOLUTA (PHV) and REVOLUTA (REV), members of the HD-ZIP III family, along with the *KANADI* genes, act antagonistically to establish adaxial and abaxial polarity in leaves and roots. The miR165 and miR166 families target these HD-ZIP III transcription factors, and their expression patterns overlap with that of the *KANADIs,* suggesting that they both limit the HD ZIP III domain. The control of HD-ZIP III factors through miR165/166 is evolutionarily conserved among plants, and is important for the establishment of bilateral symmetry during embryo development (X. Chen 2005; Izhaki and Bowman 2007). Dominant gain-of-function alleles of HD-ZIP III genes cause changes in the recognition site for miR165/6, thus impeding cleavage of the HD-ZIP mRNAs. This causes up-regulation of the HD-ZIP targets and also increases their expression domain and cells, which normally have abaxial identity, become adaxialized.

Although leaf initiation is the same for all leaf primordia, the traits of the fully developed leaf change during the course of vegetative development. These changes in leaf traits during vegetative growth are referred to as heteroblasty (figure 05). The first leaves made by the plant are referred to as juvenile leaves, while leaves produced later during vegetative development are called adult leaves. Juvenile leaves are small and round, lack abaxial trichomes (leaf hairs), and have smooth margins. Conversely, adult leaves are larger and elongated, have trichomes on their abaxial side, and their margins are serrated (fig. 05) (Poethig 2010; Tsukaya 2002).

Figure 5. Changes in leaf morphology at 17 days. Arabidopsis produces leaves sequentially during vegetative growth. Here, I highlight the shape of the leaf: changes in width, length and the margin are clearly visible.

Activity of miR164 was first reported by the Bartel groups (Mallory, Dugas, et al. 2004) using plants that either overexpressed miR164, or expressed *CUC* genes resistant to regulation by miR164 (the *CUC* mRNAs are are targeted by miR164). They observed that at the embryonic and vegetative stages, organ position, number and formation of boundaries were affected. Other studies demonstrated the importance of this family of miRNAs in the regulation of petal

number (Baker et al. 2005), leaf margin serrations (Nikovics et al. 2006), and control of age-induced cell death (J. H. Kim et al. 2009). Three loci, MIR164A, B, and C, produce the mature miR164 sequence. The A and B miRNAs differ from the C mature sequence by only one nucleotide (X. Chen 2009).

The miR164 family regulates several members of the NAC domain transcription factors (TFs). This large family of plant-specific TFs participates in key developmental processes from embryogenesis to vegetative and floral development (Jensen et al. 2010; Olsen et al. 2005). The NAC domain is conserved in the plant kingdom; the term is coined from names of genes from different plants containing the same domain: the petunia **N**AM, and Arabidopsis **A**TAF and **C**UC genes (Aida et al. 1997). In petunia, mutants in the *NAM* (*NO APICAL MERISTEM*) gene lack a Shoot Apical Meristem (SAM), hence the name, and seedlings fail to grow; plants that manage to develop often display fused cotyledons and aberrant flowers (Olsen et al. 2005).

Among the first NAC factors discovered were the *CUC* genes (*CUP-SHAPED COTYLEDON*). Mutants in these genes also produced fused organs and affect SAM formation. In *Arabidopsis*, the *CUC* family is composed of three members: *CUC1*, *CUC2* and *CUC3*. The functions of *CUC1* and *CUC2* partially overlap and show additive effects as demonstrated by the phenotypes of their single and double mutants (Aida et al. 1997; Hasson et al. 2011; Olsen et al. 2005). *CUC3* is not regulated by miR164 because it lacks the miRNA recognition site (Laufs et al. 2004).

The involvement of miR164 in cell death through aging shows the extent of the role of this miRNA: a pathway involving ORE1 (another NAC gene), miR164 and EIN2, an ethylene-responsive gene, regulates leaf senescence in Arabidopsis in an aging-dependent manner (J. H. Kim et al. 2009).

I.I.11 Leaf margin development

The Laufs group demonstrated that the balance between miR164A and CUC2 specifically regulates leaf serrations in Arabidopsis (Nikovics et al. 2006). When regulation by miR164A is absent, leaves display a higher degree of serrations at the margin than the wild type. Figure 06 shows the heteroblasty of different genotypes at 17 days, focusing on the margin of leaves. The wild type row shows the normal progression of leaves, from small and round to enlarged and

serrated along the margin. The *cct* mutant, marked by a delay in development, has fewer leaves and they are smaller, with smooth margins. The last two rows show the involvement of miR164A and *CUC2* in the control of serration formation. Both genotypes represent a lack of miR164 regulation. The first one is a null mutant allele of miR164A and the second is a *CUC2* gene in which the miRNA regulatory site was modified to avoid recognition and cleavage. Both genotypes are characterized by an earlier appearance of serrations and higher frequency than the wild type. The miR164a-4 and CUC2g-m4 lines were kindly donated to us by Dr. Patrick Laufs (Nikovics et al. 2006).

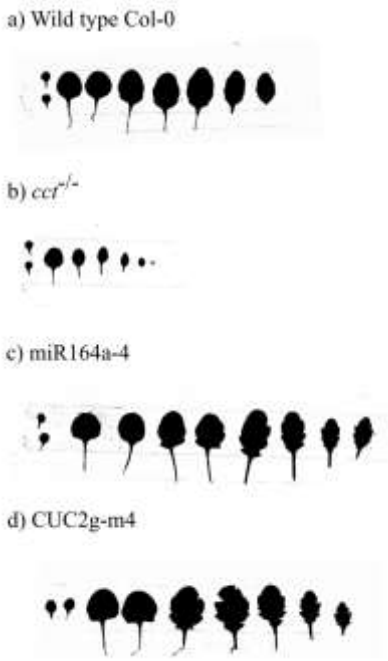


Figure 6. Heteroblasty of different genotypes at 17 days. The margin of the leaf is the important part of this figure. The two small leaves at the beginning of each row are the cotyledons (embryonic leaves).

Using computational approaches, the Tsiantis group showed that serrations at the margin are the outcome of a growth-repression mechanism. Auxin and miR164 promote the growth of the tips while spatially confining *CUC2* to no-growth zones. Zones with *CUC2* expression are spaced along the margin of the leaf, between points of auxin/miR164 expression. These mutually exclusive areas result in the formation of serrations and indentations (Bilsborough et al. 2011).

I.I.12 Measuring Gene Expression

Molecular analysis of the expression pattern of a gene can help to discover its function in a specific biological process. This analysis can be done at the level of RNA or protein. Different methods can be used to characterize static or dynamic gene expression at the level of a single gene or all the genes within a genome. Methods like qRT-PCR quantitatively measure the expression level of individual genes. Other techniques such as *in situ* hybridization can spatially localize expression of a gene in the organism. The use of reporter genes such as GUS or fluorescent proteins can detect the dynamic spatial pattern of a gene. Their advantage is that they can be used to monitor expression at the level of RNA or protein, by using transcriptional or translational reporters, respectively (Meneely 2009).

While genetic analyses generally focus on the function of a single gene at a time, the field of genomics offers a way to characterize the expression patterns of all the genes transcribed in a specific stage, organ treatment or cell type at once. Microarrays were the first genomic technique that offered a way to characterize the transcriptome of an organism by measuring expression of many genes at once (Lashkari et al. 1997; Schena et al. 1995). Further advances in high throughput methods allowed the large scale sequencing of nucleic acids in a fast and inexpensive way, based on massive sequencing of individual transcripts of multiple samples in parallel. Application of these technologies at the level of RNA allowed for finer quantitative measurements of gene expression (compared with microarrays). RNA Sequencing (RNA-seq) provides several advantages over microarrays, such as the ability to quantify poorly expressed transcripts, and to detect new genes and different transcript isoforms. Reproducibility among technical and biological replicates is higher than in microarrays (Lister et al. 2008; Nagalakshmi et al. 2008). Contrary to microarrays, the relative number of molecules present in the sample and their nucleotide composition is precisely determined.

Work on model and non-model organisms using RNA-seq of small transcripts (sRNA-seq) proves the usefulness of this technology for elucidating the role of miRNAs and other sRNAs in biological processes. Examples include studies of miRNAs in renal cell carcinoma (Zhou et al. 2010), viral responses on bovine cells (Glazov et al. 2009), human B-cell development (Jima et al. 2010) as well as the discovery of new loci producing miRNAs in mammalian cells (Castellano and Stebbing 2013). In Arabidopsis, root miRNAs responding to different nitrate

conditions were discovered (Vidal et al. 2013), and in *Brassica napus* miRNAs involved in seed development were found. (Huang et al. 2013).

I.I.13 Use of bioinformatics approaches to measure gene expression

Zhan and Lukens integrated RNA-seq and microarrays to identify microRNAs important for transitions during Arabidopsis development. Using data sets from specific tissues and from plants defective in genes involved in miRNA biogenesis, they identified targets of miRNAs whose expression changed during different developmental stages. Furthermore, they were able to discover new miRNA loci, and showed evidence for miRNAs with a tissue specific expression pattern. They concluded that miRNAs in Arabidopsis control developmental transitions, and spatially restrict their targets at specific time points during development (Zhan and Lukens 2010).

As an example of the importance of bioinformatic tools is the case of annotation of sRNA data. Most currently available tools cannot discriminate the source of a short sequence when it is aligned to a location where two genes overlap (data not shown). This could mean that when a read is assigned a location overlapping the primary transcript of a miRNA and the mature miRNA, will either discard it or assign it to both. This situation is undesirable because information of the counts will be altered (reduced or duplicated), biasing the differential expression results. Figure 07 conceptually exemplifies one of the issues I tackled during this study (Chapter I).

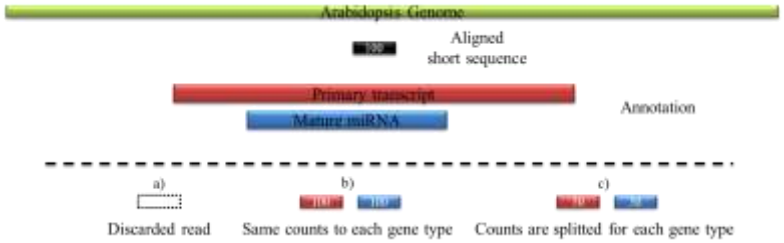


Figure 7. After aligning the short sequences to the genome, overlapping annotations represents a problem. Possible solutions for this are a) discarding the read; b) assigning the same number of counts to each type or c) splitting the counts between different types. Either solution may alter the counts and influence the differential expression results.

I.II Background

In 2010, Gillmor and collaborators described for the first time the Arabidopsis homologues of MED12 and MED13. The name of the genes *CENTER CITY* (*CCT* for MED12) and *GRAND CENTRAL* (*GCT* for MED13) illustrate the enlarged apical meristem phenotype in mutant embryos. The expression patterns of *GCT* and *CCT* change through embryo development: they are expressed in all cells during early stages but are then restricted to the vasculature, and the shoot and root apical meristems (SAM and RAM) in the mature embryo. The role of these genes in Arabidopsis is to control developmental timing events such as phase change or cell fate specification. During embryo development, mutants displayed morphological abnormalities as a result of the separation of cell fate specification from cell division. The transition from globular to heart stage is also delayed. Misshapen cotyledons and an expanded SAM are characteristic of the mutant mature embryos.

Post embryonically, *gct* and *cct* mutant plants are smaller than their wild type counterparts; also, their leaves display juvenile traits such as late appearance of abaxial trichomes and low level of serrations at the margin, in addition to late flowering. Alteration in the timing of onset of leaf traits was shown to be due to misregulation of the microRNA miR156 in *gct* and *cct* mutants (Gillmor et al., 2014). miR156 had previously been demonstrated to be a master regulator of developmental timing during vegetative development (Wu and Poethig, 2007; Wu et al., 2009). The late flowering phenotype of *gct* and *cct* mutants was demonstrated to be mostly due to overexpression of the flowering regulator *FLC* (Gillmor et al., 2014).

Additionally, two groups in Japan independently isolated different alleles for CDK8 subunits. *MAB2/MED13* (*MACCHI-BOU 2*) also showed aberrations in the pattern of cell division during embryogenesis leading to morphological abnormalities and late flowering. *CRP/MED12* (*CRYPTIC PRECOCIOUS*), a dominant negative allele of MED12, displayed an early flowering phenotype (Imura et al. 2012; Ito et al. 2011). This supports the role of the CDK8 module in the temporal expression of specific genes that act at different developmental stages; also, since the both mutants have very similar phenotypes they must act in related biological pathways.

The *HEN3* (*HUA ENHANCER 3*) gene codes for the mammalian CDK8 homologue in Arabidopsis. The Chen group first described this gene based on its developmental defects during floral determinacy (Wang and Chen 2004); opposed to the delayed flowering phenotypes of *cct* and *gct*, *hen3* mutants cause alterations in the identities of stamen and carpel cells. The function of CYCLIN C (CYCC) subunit is not well documented during Arabidopsis development. As other cyclins, it partners with a CDK. CYCC binds specifically to the CDK8 protein, however, CYCC does not oscillate with the cell cycle (Banyai et al. 2014). Biochemical studies show that, in mammals, the CYCC-CDK8 pair represses transcription by phosphorylation of CYCH, thus blocking the ability of TFIIH to initiate transcription. Another mechanism of control is through the phosphorylation of the CTD (C Terminal Domain) of the RNA pol II, which also represses transcription. However, to date, there are no studies showing the developmental relevance of the *CYCC* gene. Dr. Claudia Silva-Ortega, from our group, tried to obtain CYCC mutants however, the gene is positioned in tandem with another CYCC copy (AT5G48630 and AT5G48630), making it difficult to obtain a double mutant.

Morphological defects show the importance of the CDK8 module in Arabidopsis embryogenesis and growth (Gillmor et al. 2010, 2014; Wang and Chen 2004). At the molecular level, this module regulates the expression of genes important for developmental processes like phase transitions in embryos and adult plants. Our group showed that loss of function mutations in components of the CDK8 cause a delay in embryo patterning, as well as a delay in the acquisition of adult leaf traits, and in flowering time; this is explained by an increase of the miR156 levels (Gillmor et al. 2010, 2014). In this study I seek to address the role of distinct subunits of the CDK8 module in the transcription of other miRNAs that.

For this study, Dr. Claudia Silva-Ortega performed the RNA extraction and library preparation experiments from 18 days old *Arabidopsis thaliana*. At this stage wild type plants are in the transition to the reproductive phase; we selected this stage to know which processes are misregulated during this transition in the mutants we had available. We aimed to investigate the role of GCT, CCT and HEN3 subunits in the production of miRNAs important for development in *Arabidopsis thaliana*.

Hypothesis

"Subunits of the CDK8 module controls miRNAs involved in development and morphogenesis."

Objectives

General objective.

Identify microRNAs whose abundance is controlled (either directly or indirectly) by the CDK8 module subunits MED12(CCT), MED13(GCT), and CDK8(HEN3) in *Arabidopsis thaliana*.

Specific objectives.

- Characterize the small RNA transcriptomic landscape of mutants in subunits of the CDK8 module.
- Discover miRNAs that are differentially expressed in mutants of the subunits CCT, GCT and HEN3 of the Arabidopsis CDK8 submodule.
- Validate the role of a CDK8 subunit in the regulation of a miRNA.

Strategies.

- Use a combination of custom and available bioinformatics tools to characterize small RNA-Seq transcriptomes of wild type and CDK8 module mutants.
    - Design and develop a bioinformatic pipeline for the analysis of sRNA-Seq data.
    - Search for miRNAs misregulated by mutations in subunits of the CDK8 module.Identify differentially expressed miRNAs in the sRNA-seq data.
- Characterize the spatiotemporal expression pattern of a miRNA in abscense of regulation of a CDK8 subunit.
    - Use of marker lines for the miRNA in wild type and mutant genotypes to investigate changes in the expression pattern.

Chapter II Materials and Methods

In this section I describe the methods and tools used in both the bioinformatics analysis and the experimental validation.

II.I Extraction and sequencing of sRNAs

Sequencing of small RNAs was done on the Illumina Genome Analyzer IIx platform. Illumina along with SOLiD are good technologies for sequencing of sRNAs (McCormick, Willmann, and Meyers 2011).

For the RNA extraction and preparation of libraries we used 18 days old *Arabidopsis thaliana* plants of the Columbia ecotype. Five different genotypes were used: wild type, *cct*, *gct*, *hen3* and the double mutant *cct;hen3*. Two biological replicates were used for this study, pooling together ~10 genetically identical plants for the RNA extraction. This generated a total of 10 independent samples. To increase read coverage, all samples were sequenced in three different lanes; obtaining a total of 30 different sequencing files. sRNA libraries were prepared starting with a standard trizol protocol for RNA extraction. RNA was purified with the Qiagen RNeasy Mini Kit using the RNeasy spin column; large transcripts (mRNA and rRNA) stay in the column but the sRNA fraction is eluted. The sRNA fraction is then purified and concentrated with RNeasy MinElute Cleanup Kit.

II.II Computational Methods

I wrote custom scripts in bash and R using packages from the Bioconductor project. Quality of the 30 different raw sequencing files was assessed before concatenating files belonging to the same biological samples.

Table 1 represents a summary of the programs and software used for this study along with each program's version and reference.

Table 1. Software and versions used in the transcriptomic analysis.

Program.	Version.	Reference.
FastQC	0.11.2	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Kraken	13-274	(Davis et al. 2013)
Bowtie	1.0.0	(Langmead et al. 2009)
R	3.1.0	(R Development Core Team 2008)
Bioconductor	3.0	(Gentleman et al. 2004)
GenomicRanges	1.16.4	(Lawrence et al. 2013)
rtracklayer	1.24.2	(Lawrence, Gentleman, and Carey 2009)
Biostrings	2.32.1	(Pages H, Aboyoun P n.d.)
edgeR	3.6.8	(Robinson, McCarthy, and Smyth 2009)

1. Quality control

A bash script uses FastQC to analyze the quality of the sequencing files. The parameters used to analyze quality of the 30 libraries were:

- -t 4. To process 4 files simultaneously.
- -f fastq = Format of the sequencing files is fastq.

The output file is an HTML file with the quality report for each sequencing file. Taking advantage of the FastQC results, I wrote another script in bash to extract the total number of reads in each file. The output is used by a script in R to plot the size of individual libraries.

2. Pre-processing of short reads

I used the Kraken toolkit (Davis et al. 2013) for preprocessing the sequencing files. This toolkit contains three programs: Minion, Reaper and Tally. Minion, a program to automatically detect adapter sequences, was not used since we already knew the sequence of the adapter used during sequencing.

3. Reaper

I used Reaper directly to remove the adapter and eliminate low quality segments. Reaper aligns

the sequence of an adapter to a read; if there is a match within certain parameters (explained below) such segment is trimmed or discarded, depending on where the match occurs.

The parameters I used for Reaper are shown in Table 02. Numbers separated by slashes (as in l/e/g) in certain parameters represent:

- **l** - Minimum length of a match of an adapter sequence with a read.
- **e** - Maximum number of substitutions needed for a perfect match.
- **g** - Maximum number of gaps between adapter and read.

Table 2. List and explanation of the parameters used for trimming adapters.

Parameter	Explanation
-geom no-bc	No extra bar code in the sequences.
-3pa {Sequence}	Indicates the sequence of the 3' adapter.
-tabu {Sequence}	5' adapter sequence. If found, the whole read is discarded.
-mr-tabu 14/2/1	The match requirements for the 5' alignment. 14 is the minimum length of a match of the adapter with the read. 2, the maximum number of edits for a perfect match. 1 is the maximum number of gaps.
-3p-global 6/1/0	Indicates the criteria for a match anywhere between the adapter and the read.
-3p-prefix 11/2/1	Criteria for a match between the start of adapter and the end of a read.
-3p-head-to-tail 1	At which length a perfect match between adapter start and end of a read should be removed.
-dust-suffix 20	Trims low complexity sequences formed by poly-N (A,C,T or G)
-format-clean %C%t%L%n	Output format: Sequence of the trimmed read (%C) followed by a spacer (%t), then the length of the read (%L). Each read in a new line (%n).
-nnn-check 1/1	Removes any ambiguous base (N)
-qqq-check 35/10	Trims a sequence according to quality. Any consecutive segment of 10 bases with a median phred quality score less than 35 is trimmed.

4. Tally

Tally uses Reaper's output to minimize memory usage by eliminating duplicated reads but conserving their number of occurrences. In Table 03, I describe the parameters used to run Tally in a bash script.

Table 3. Parameters used in Tally.

Parameter	Explanation
-format >seq%I_w%L_x%C%n%R%n	Changes the output format: **>** is the FASTA identifier. Prefix "seq" is followed by an identifier (%I); w is the length of the read (%L); x is the number of ocurrences. The sequence of the read is in a new line (%n%R).
-l 12	Remove reads that are less than 12 bases.
-record-format %R%t%F	To identify the input format. %R indicates that the sequence of the read comes first. Everything else after is discarded.

5. Mapping short reads to the reference genome

I used Bowtie to map the processed reads to the Arabidopsis genome. The script first creates an index of the Arabidopsis genome that allows fast alignments of the small reads.

First, I wrote a script in bash to create the index using the *bowtie-build* command. Since we don't have paired-end sequences, the –noref attribute keeps the builder from creating extra files. The genome sequence is in FASTA format with all the chromosomes or contigs in the same file. I used the TAIR10 version of the Arabidopsis genome.

Bowtie reports the alignment in a tab delimited file. The following is an example of bowtie output.

seq29_w30_x6798 - Chr2 12975988 15 8:C>T

- It contains the name of the read (along with its length and number of occurrences from the Tally output).
- The strand to which the read mapped.
- Genomic region where the match occurred (Chromosome)
- Position where the match begins.
- The next field indicates how many valid matches occurred besides the one reported.
- If a mismatch occurs, bowtie indicates the position and the base change with a Position:Base>Change format. If no mismatch occurred, this field is left blank.

The script for bowtie processes one file at a time. Table 04 describes the attributes used for each parameter.

Table 4. Parameters and attributes used in the script to align short reads to the Arabidopsis genome.

Parameter	Explanation
--time	Prints the time taken to process each file.
-v 3	Maximum number of mismatches allowed for alignment.
--best	If there is more than one optimum alignment, bowtie will report the best in terms of fewer mismatches.
-k 100	Reports a maximum number of optimum matches if the read was aligned to multiple places in the genome.
--strata	If there are more "best" alignments, they are sorted by their number of mismatches. The one with fewer mismatches is reported.
--chunkmbs 512	Memory allocation for processing.
-f	Indicates files are in FASTA format.
--suppress 5,6	Supresses 5th and 6th columns on the output. See below.

A script in R uses the output from bowtie to create a special table with two main contents: the genomic coordinates of the reads and metadata such as counts, mismatches, and annotation. The script uses Bioconductor, GenomicRanges, rtracklayer, and Biostrings packages, and reads the bowtie files and creates a table with the genomic features from the reads.

Table 5. Description of the fields after mapping the reads.

Genomic information	
Seqname	Chromosome location
Range	Start and end positions of the sequence.
Strand	Direction of the sequence (+/-).
Metadata	
Read	Unique sequence identifier
OtherPositions	Number of alternate locations
Mismatch	Number of mismatches
Length	Size of the aligned read.
Counts	Total number of identical reads.
MatchRatio	To know how good the match is. Ratio between 0 and 1. Cutoff at 0.8 (80% of identity). $MatchRatio = \frac{(Length - No.Mismatches)}{Length}$
CountsPerPosition	For reads that map to more than one position, divide the number of counts equally between each position.

6. Annotation of the mapped reads

After assigning the reads a location in the genome, I wrote a script in R to assign a functional annotation to each mapped read. I used the *Arabidopsis thaliana* annotation from TAIR10. For the miRNA annotation I used the miRBase v21 for Arabidopsis. This miRNA annotation contains sequences for hairpin precursor and mature miRNAs. I manually removed miRNA annotation from the TAIR database and added the one from miRBase instead.

Table 6. Description of the added fields after annotation.

Metadata	Description
Class	Refers to the biotype annotated (exon, intron, mature miRNA, etc.)
Name	AGI identifier or miRNA name.
Description	Short description of the gene function.

The script uses Bioconductor and GenomicRanges packages. Here we implemented the annotation algorithm we devised (see Chapter III) comparing the position of the annotation against that of a mapped read to assign functional annotation. When a read matches its smallest annotation possible, the metadata of the table with genomic features is updated to add functional information (as described in Table 06).

After annotation, this script also plots the total number of classes annotated as total annotated classes and as a distribution according to read length.

7. Quantification of the annotated reads

This script creates a matrix of gene counts (counts table), with genes as rows and genotypes as columns. It also contains information (metadata) on the gene function. It also filters the reads by class (only mature miRNAs, for example) and by read length (for this study we used reads between 16 and 26nt).

8. Differential Expression Analysis

This R script uses edgeR to detect differentially expressed genes. Since we are interested in miRNAs, the counts table is filtered for mature miRNAs.

Genes are filtered according to the counts per million (cpm) across libraries. Filtering this way does not bias the analysis because there is no *a priori* selection of the conditions, only genes that match this condition in any 2 libraries are selected.

For normalization, edgeR provides various algorithms. I used the TMM method to calculate normalization factors and scale counts to account for differences in size of individual libraries.

To compare the expression levels between genotypes, the script fits a generalized linear model to the data. The contrasts made are:

- *gct* VS wt
- *cct* VS wt
- *hen3* VS wt
- *hen3*/*cct* VS wt

Using edgeR to estimate dispersions serves to get an insight of the variability of the counts among samples. For this, the common, tagwise, and trended dispersions were estimated using the commands *estimateGLMCommonDisp* and *estimateGLMTagwiseDisp*, and *estimateGLMTrendedDisp* with default parameters.

Differential expression for each gene is assessed with the command *glmLRT*, with default settings. It fits the read counts of each gene from the contrasts selected to a NB distribution. To test for truly differentially expressed miRNAs the *decideTestsDGE* command classifies genes according to their significance (FDR < 0.05).

To find miRNAs that were possibly regulated by the whole module I filtered the resulting table for miRNAs that were significant (FDR < 0.05) across all conditions and whose logFC was higher than 1.5. I also filtered for significant miRNAs in individual conditions.

Using the PMRD database, I added information about the targets of Arabidopsis miRNAs (Zhang et al. 2009). With this data I performed GO term enrichment analysis on the targets of upregulated genes. This analysis was made using targets from the set of miRNAs upregulated across all conditions as well as in individual contrasts. For this I used the binGO plugin (Maere, Heymans, and Kuiper 2005) in Cytoscape (Shannon et al. 2003).

II.III Experimental Methods

1. Plant materials

Genotypes of the plants used in this study are summarized in Table 07. All genotypes are of the Columbia ecotype.

Table 7. Genotypes used in this study.

Genotype	Description	Reference	Source
wt	Wild type (Col-0).		Lab stock
cct-1	MED12 homologue.	(Gillmor et al. 2010)	
gct-2	MED13 homologue.		
hen3-564	CDK8 homologue.	(Wang and Chen 2004)	
cct-2;hen3-564	Double mutant		
miR164a-4	miR164a null mutant.	(Nikovics et al. 2006)	Dr. Patrick Laufs
CUC2g-m4	*CUC2* resistant to miR164A regulation.		
pMIR164A::GUS	GUS expression driven by the MIR164A promoter.		

2. Plant growth

Seeds were sown in peat moss soil and put at 4°C for 48 hours for vernalization. After this they were placed on laboratory growth racks under 16h/day fluorescent light for germination. When seeds germinated, flats were moved to greenhouse conditions.

3. Heteroblasty analysis

To see the differences in leaf morphology of different genotypes wt, *cct*, *miR164a-4* and *CUC2g-m4* plants were grown and their leaves dissected after different time points. Sequential leaves were placed on white paper sheets using double sided adhesive tape. Sheets were scanned and processed with GIMP (http://www.gimp.org) and Inkscape (https://inkscape.org) to highlight the silhouettes.

4. Crosses

We crossed the pMIR164A:GUS line with *cct* and took it to the F3 generation by selecting independent lines that were positive for GUS staining and also segregated the *cct* phenotype. We obtained plants that were homozygous for GUS and heterozygous for *cct* (*pMIR164A:GUS*$^{+/+}$;*cct*$^{+/-}$).

5. GUS Staining

GUS solution was prepared following the protocol described in Table 08.

Table 8. Protocol for the GUS staining solution.

Reagent	Stock concentration	Final concentration	Volume for 50 mL
PBS (pH 7)	100 mM	50 mM	25 mL
EDTA (pH 8)	50 mM	10 mM	1 mL
Potassium ferricyanide	50 mM	5 mM	5 mL
Potassium ferrocyanide	50 mM	5 mM	5 mL
Triton X-100	100 %	0.1 %	50 µL
X-Gluc		1 mg/mL	Dissolve 50mg in 1 mL DMSO or DMF
Take to 50 mL with dH2O. Pass through a syringe filter and store in the dark at -20°C.			

6. Construction of a pMIR164A:eGFP marker line

To amplify the promoter of the MIR164A gene I modified the primers reported by the Laufs group (Nikovics et al. 2006); these oligos were used to make the MIR164A:GUS line also used in this work. These primers amplify a 2.1kb region upstream of the start site of miR164A. Because this region complements 92% of the miR164a mutant phenotype (Nikovics et al. 2006).

The construct was made using pGEM-T-Easy® as an entry vector; after checking the quality of the promoter, restriction enzymes were used to cut from the entry vector and paste it into the final vector. The final vector is a modified pCAMBIA 3301® carrying the sequence for eGFP. The eGFP has an endoplasmic reticulum (ER) signaling tag for subcellular localization.

The primers I used are in table 09. The original primer sequence has also a sequence used for recombining into a vector. Since I used a "cut/paste" approach, I screened for restriction sites

that were present in both of our vectors (pGEM and pCAMBIA) but absent in the pMIR164A sequence. The sites that matched these requirements were PstI and BglII.

Table 9. Primers used to amplify the upstream region of MIR164A.

	Sequence (5'-3')	Restriction site.
Forward	ATcccgggAGATGCTCATCACGTATGCCAA	PstI
Reverse	GTTagatctGGAGATTCTCACCCGCATTT	Bgl II

The primers sequences on table 09 can be divided into three parts: upper-case letters on the right side belong to the actual sequence of the promoter; lower-case bases at the middle correspond to the restriction sites; upper-case letters at the left are added for cutting efficiency.

I used Top10 *E.coli* cells to amplify plasmid and *Agrobacterium tumefaciens* to transform Arabidopsis plants. I directly transformed the plasmids carrying our pMIR164A:eGFP construct into wild type and *cct* plants. The transformation technique was done by floral dipping. I independently transformed different plants (five wild type and three *cct* homozygous plants). Then, I collected seeds from each plant.

Chapter III - Development of a bioinformatics pipeline for the analysis for sRNA-Seq data.

In this chapter I present the description of our bioinformatics approach to characterize sRNA transcriptomes.

Wild type plants at 18 day old wild type plants were in the transition from vegetative to reproductive phase. However, mutants for subunits in the CDK8 module are delayed in this transition. Although plants are clearly in different developmental stages, this first approach was to explore transcriptional differences in equally aged plants, at the time where the transition normally occurs. After analyzing these results, we planned an experimental approach using different developmental time points to validate our computational findings.

Briefly, we used the Illumina Genome Analyzer IIx platform to sequence libraries of small-RNAs. Each library (10 samples, consisting of two biological replicates of 5 genotypes) was sequenced three times using different lanes of the flow cell, giving a total of 30 sequencing files. The output files contain information of each individual read analyzed, along with the nucleotide sequence itself and its quality per base. More details of the experimental design are described in the Materials and Methods section.

III. 1 Data Analysis Description

Samples were prepared by pooling ~10 genetically identical plants at the same age and then extracting RNA. Each pool of plants represents a biological replicate. Two biological replicates were used per genotype. Furthermore, each sample was sequenced independently three times (technical replicates) in different lanes of the flow cell to increase read number in the libraries. Counts among technical replicates from the same sample were highly consistent (Appendices 31).

A total of 30 sequencing files were produced. I concatenated the same biological replicates sequenced in different lanes (technical replicates) into a single file, resulting in a total of 10 files, two samples per genotype. This is represented in figure. 08. I assessed the quality of each independent library prior to combining files from technical replicates.

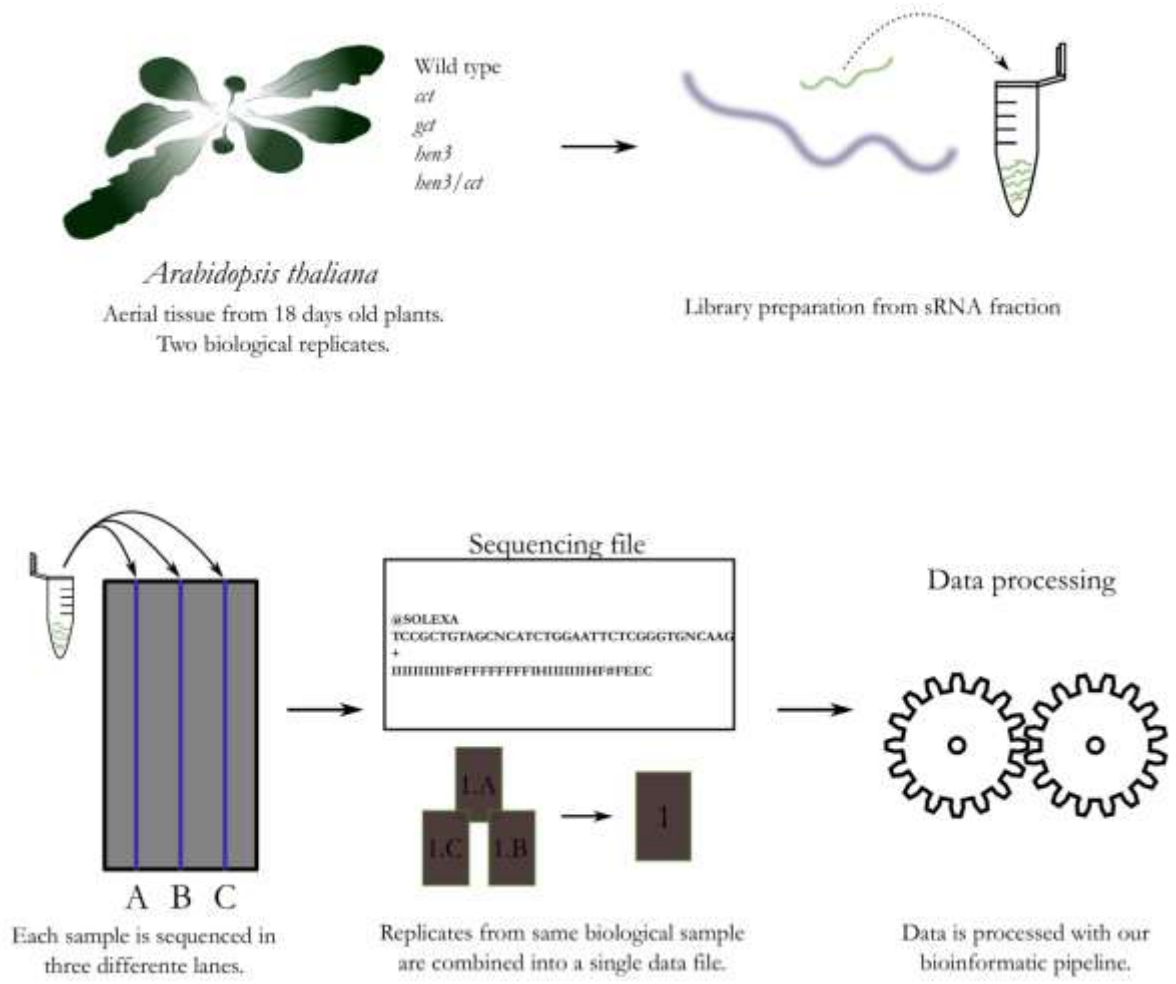


Figure 8. Conceptual representation of sample preparation for sequencing and data pre-processing in this study.

To analyze the libraries produced from sequencing we designed a semiautomatic pipeline which uses a combination of open source software and custom scripts for the analysis of multiple samples at once. The output of each section serves as the input for the next. The pipeline runs on UNIX based systems (Mac OSX, Linux, etc.) and the custom scripts are written in either bash or in R using packages from Bioconductor.

R is a programming language widely used for statistical analysis. It provides a broad range of tools and open source software, such as Bioconductor, for the exploration, processing and

visualization of genomic data sets (Gentleman et al. 2004). It is extensively used in the processing of biological data such as that from RNA-sequencing.

The pipeline starts with quality assessment of the 30 independent sequencing files. Next, I concatenated the files from technical replicates belonging to the same biological replicates and processed the data to trim the sequences. Oligonucleotide adapters that are added during library preparation, and which are needed for the sequencing process, have to be removed. Also, sequences with low quality or complexity can be trimmed or removed.

The next step involves aligning the processed sequences to a reference. For this I used the latest version of the Arabidopsis genome, TAIR10 (http://www.arabidopsis.org/portals/genAnnotation/gene_structural_annotation/annotation_data.jsp). Once the sequences were aligned along the genome, I used a custom R script to annotate the reads. This helps us to know which genes producing small RNAs were expressed in the transcriptomes. The final step of the pipeline involves differential expression analysis in order to discover small RNAs whose regulation was affected by mutations in subunits of the CDK8 module of Mediator in Arabidopsis.

Next, I present a description of each module, the programs used and the format of the results.

III.2 Quality Control.

For this purpose I used the FastQC program. The aim of this software is to provide a report on the quality of the sequences to ensure that there are no biases or problems in the data.

FastQC produces the quality reports as HTML files. It evaluates different aspects of the data, from quality per base and GC content to overrepresented sequences. Furthermore, each quality assessment is graphically represented and evaluated in three categories (pass, attention, fail) representing if data looks good, slightly abnormal or totally unusual. Yet, FastQC recommends taking into account the context of the libraries before taking decisions on the pass/attention/fail results of the quality assessment.

There are eleven different analysis modules in the report. One of the most important ones is the "Per base sequence quality" report. It graphically summarizes the quality ranges of each position for all reads in a library.

III.3 Sequence pre-processing.

For the purpose of processing the short reads, I used the Kraken toolkit (Davis et al. 2013). This software is designed for analyzing short sequences produced by NGS.

Minion is more of a utility; its purpose is to infer the presence of a 3' adapter contaminant in the sequences, if such a sequence is not previously known.

Reaper is designed to quickly trim and filter short reads. I used this tool to search and remove adapter contamination at the 3' side of the sequence; also, to remove sequences of low complexity, such as those with a string of repeated nucleotides and finally to discard or trim sequences based on the quality of the bases. After this process, the length of many reads changed, producing sequences that may have 0 bases (where the whole read was discarded) to 36 bases (where no trim was necessary) based on the parameters used.

Tally, the third tool in the set, is used to minimize redundancy and file size by compressing data. After removing and trimming the sequences, the program eliminates duplicated reads but retains the number of occurrences. The program can output the results in different formats, including fasta. To facilitate downstream analyses the name of the sequence was changed to a unique identifier. The following represents an example of the output format:

>seq1_w14_x123074
GGGATGGGTCGGCCT

The first part is the FASTA identifier (**>**) followed by (**seq##**) a unique identifier for each read. Next, (**_w##)** identifies the length of the sequence while (**x##)** represents the number of occurrences, or counts, for each specific sequence.

III. 4 Sequence alignment.

With the output of the previous module, we have a clean set of sequences but we don't know where they originated from in the Arabidopsis genome. To solve this situation there are two possibilities. The first is to compare the sequences against the complete genome and search for a match. The second is to compare the sequence against a database of genomic regions that are known to be genes or other genomic elements in the plant.

I chose to align the sequences in the whole genome because this approach opens the possibility of finding new sRNA-producing loci. For this purpose I used an open source software called Bowtie (Langmead et al. 2009). This program is designed to align short reads against large genomes in a fast and efficient manner. The Burrows-Wheeler transform that Bowtie uses creates an index for the reference genome. Indexing the genome of the reference creates a compressed version of it; thus, it provides a guide for a faster localization of the reads to a place in the genome.

Basically, this algorithm transforms text into a matrix by making permutations on the elements of the string. It sorts the matrix in alphabetical order, thus clustering similar characters together, improving compressibility. It then compares the short sequences against the indexed genome reporting the alignments with the fewest mismatches in a fast way.

III.5 Sequence annotation.

In order to maximize the information available I combined annotation data from TAIR (http://arabidopsis.org/) and miRBase (Kozomara and Griffiths-Jones 2014). miRBase is a resource for the nomenclature, annotation and sequence information of microRNAs from various organisms, including *Arabidopsis thaliana*. The reason for combining both databases was that the annotation of miRNAs in TAIR lacks information on the mature miRNA sequences and genomic locations. Thus, I manually removed all miRNA information from TAIR and added the miRBase information, creating a more complete annotation reference.

Once I identified the genomic positions producing the small reads, I wanted to add functional annotation for the short reads in our data. For this purpose I used a custom annotation algorithm written in the R programming language.

As stated before, an actual bioinformatic problem is to assign annotation of reads to overlapped genes. To solve this, I designed a sieve-like annotation algorithm that first sorts genes according to different criteria such as a) class (henceforth referred to as *biotypes*), for example mRNA, exon, transposon, etc., b) average length of each biotype and c) potential source of small RNAs.

Briefly, our algorithm compares the position of a read with that of an annotated gene. If they match, the read is annotated. Until here, it behaves like other annotation programs. What our

algorithm then does differently is to separate annotated genes according to their biotype and compare read and sequence iteratively to assign the properties to find the smallest gene that matches a read.

The first step is to sort, in decreasing order, the biotypes according to their average gene length (in nucleotides). To identify genes that could possibly be transcribed from the antisense strand we inverted the biotypes to the contrary transcription orientation; this creates an "antisense" class for each biotype. Since the algorithm iterates several times over, this helps discriminate a read matching to overlapped genes from different biotypes.

To annotate a read, the algorithm compares the genomic position of the read with that of an annotated gene. If both positions have a match, the read acquires the properties of the annotated gene. This process is iterated for each of the biotypes selected, going from the biotypes with the largest genes to the smallest ones. If an already annotated read matches a smaller biotype, it will acquire this newest annotation. In this fashion we ensure that the reads always take the smallest annotation, eliminating the problem of splitting or discarding reads. This is especially helpful in the case of MIR genes where reads can map to any part of the primary transcript, however only a small section is most relevant: the mature miRNA.

The annotation procedure is conceptually represented in figure 09. Although the results presented in this thesis are solely based on differential expression of miRNAs, the reason for designing a general algorithm was to find all classes of small RNAs that were regulated by the CDK8 module. For example, studies show that subunits of both Mediator and CDK8 are important for gene silencing through epigenetic mechanisms in animals and plants (Chaturvedi et al. 2012; Kidd et al. 2011).

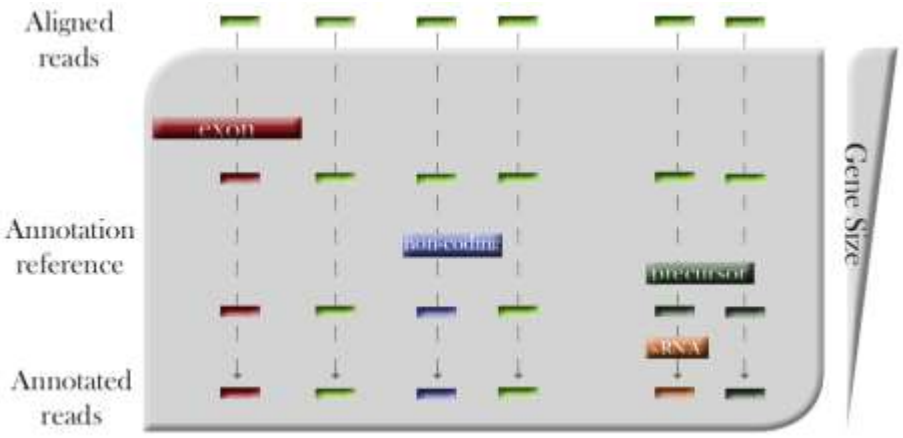


Figure 9. Graphical representation of our annotation algorithm. Green bars represent aligned reads but without annotation. Each biotype is represented by a bar of different color and a name; they are ordered by decreasing order of average size. For each iteration, if a sequence matches a biotype its annotation (color) changes. Thus if the read overlaps two genes, it will receive the annotation of the smallest one.

The biotypes considered are exons, introns, transposable elements, ncRNA, tRNA, pri-miRNA, rRNA, snRNA, snoRNA and the mature miRNA in both sense and antisense.

III.6 Quantification.

To quantify the number of reads for each gene I took advantage of the unique identifier from the Tally pre-processing step: information on the number of occurrences (counts) for each sequence is saved in the name of the read.

Counts are stored in a matrix with the columns being the genotypes and the rows all the genes for which a read had a match in the annotation. This matrix also stores metadata such as biotype, locus identifier in the Arabidopsis Gene Identifier (AGI) format and gene symbol.

III.7 Differential Expression.

Our main interest is to test if mutations in subunits of the CDK8 module of Arabidopsis Mediator have a significant impact on expression of particular small RNAs.

To answer this question I used the *edgeR* package from Bioconductor (Robinson, McCarthy, and Smyth 2009). This package was designed for testing for differential expression when expression values are discrete. For sequencing technology, the abundance of a transcript is

quantified as a number of reads, unlike microarrays in which abundance is typically measured in terms of luminescence intensity.

There are several programs designed for the analysis of differential gene expression, and diverse studies have compared their performance using simulated and real datasets. Results indicate consistency among these different methods (Nookaew et al. 2012; Rapaport et al. 2013). However, in the end the recommendations of different studies agree that the best method depends on the type of analysis that will be performed, and that parameters must be chosen carefully (Kvam, Liu, and Si 2012; Soneson and Delorenzi 2013).

In the next chapter I present the results obtained from each part of the pipeline.

Chapter IV - Results and characterization of the sRNA transcriptomes of CDK8 mutants.

In this chapter I present the results obtained by implementing the pipeline described in the previous chapter on our sRNA-Seq data.

IV.1 Quality Control.

Figure 10 shows the "Per base sequence quality" of one of the wt libraries. In this boxplot, the X-axis represents each position in the sequence and the Y-axis represents the quality in Phred format. The inter-quartile range (25% - 75%) of quality at each position is represented with the yellow box with the upper and lower values represented by the whiskers. The blue and red lines represent the mean and median quality values respectively.

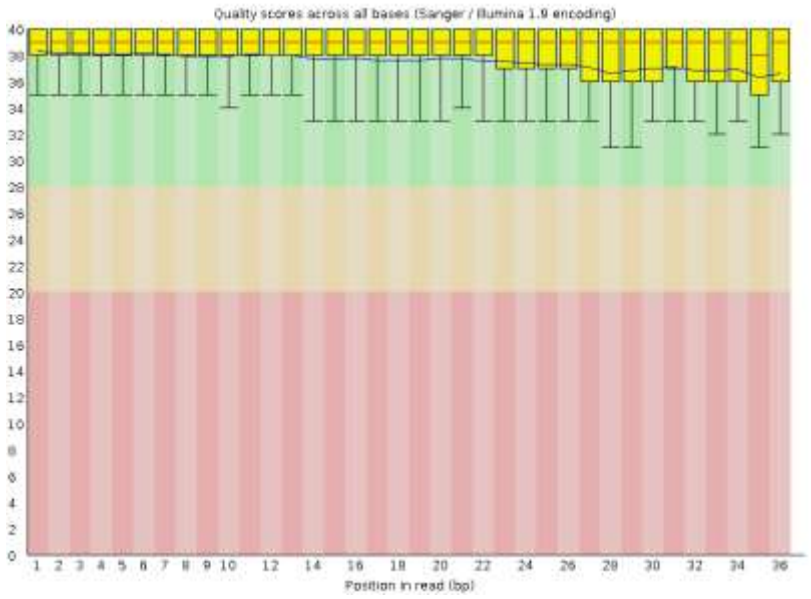


Figure 10. Quality scores of a wt library.

Phred scores represent the probability that a nucleotide in the sequence is incorrect. The actual score is the logarithm of the probability of a base being called erroneously. For example, if the Phred score of a nucleotide is 20, there is a 1% chance (0.01 probability) that such nucleotide is wrong.

Scores of over 30 are good, since they mean that there is only a probability of 1 in 1000 that those nucleotides are incorrect (Illumina 2011). The rest of the quality reports are included in the appendices (Appendices 1-30).

IV.2 Library Sizes.

After assessing the quality of the sequences, and confirming that they were of sufficient quality, I concatenated the fastq files coming from the same library as stated in the Data Analysis Description.

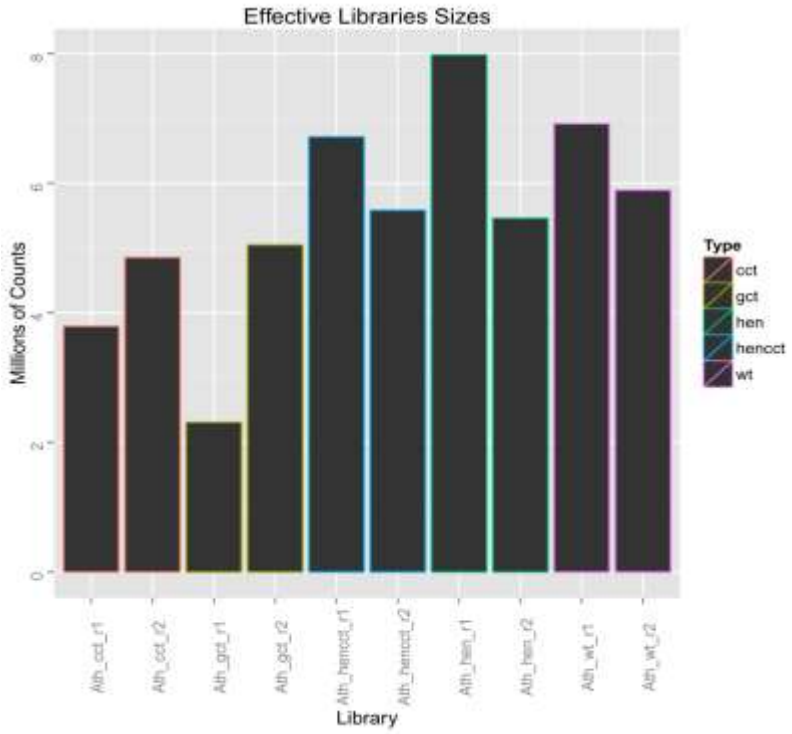


Figure 11. Sizes of sRNA libraries from biological replicates. Each sample is the result of concatenating technical replicates.

Figure 11 represents the effective size of each library used in this study after joining files from technical replicates. The plots of individual library sizes are shown in the appendix (Appendices 31-32).

IV.3 Sequence length distribution after trimming.

The original length of each sequence is 36nt. After removing the adapters and other contaminants using Reaper, the sizes of the sequences can change. Figure 12 represents the distribution of read lengths after trimming. Note that many reads are of length zero, which means that the whole sequence was trimmed. Also, the number of reads may not match with the effective size of the libraries due to discarded sequences of low complexity before trimming.

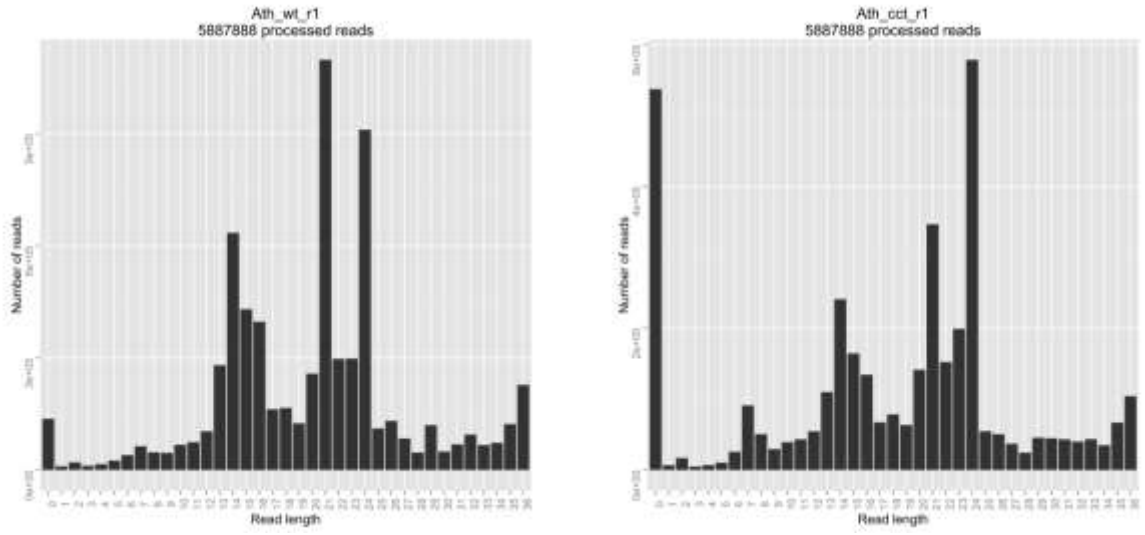


Figure 12. **Representative histogram of trimmed sequences of wt and cct libraries. The histograms are in different scale but they show the frequency of each sequence length.**

Three peaks dominate most of the distribution in all libraries: a 14nt peak, which I didn't explore further, because it was beyond the scope of this work, and two major peaks at 21 and 24 nt. They correspond to canonical sizes of miRNAs and siRNAs, respectively. As represented in fig. 12, in mutants for subunits of the CDK8 module, the 24 nt peak was the highest.

The rest of the histograms are in the appendix (Appendices 33-42). For the annotation and further analyses I used only reads between 16 and 26 nt, considering a range of sizes that would allow us to discover differentially expressed miRNAs. This is because the smaller the

read, the more difficult it is to have an unambiguous match with the genome annotation, while bigger reads are too far from canonical sizes of small RNAs.

IV.4 Sequence Annotation

To annotate the reads I used biotypes that were potential sources of small RNAs. Reads that had a match in the genome but did not match any of the biotypes were classified into an "intergenic" category. I also used the antisense version of every biotype in order to detect potential sRNAs being transcribed from opposite strands.

Figure 13 shows the total number of reads annotated for each biotype in all libraries. It represents the total number of reads (12-36 nt) classified into each possible class.

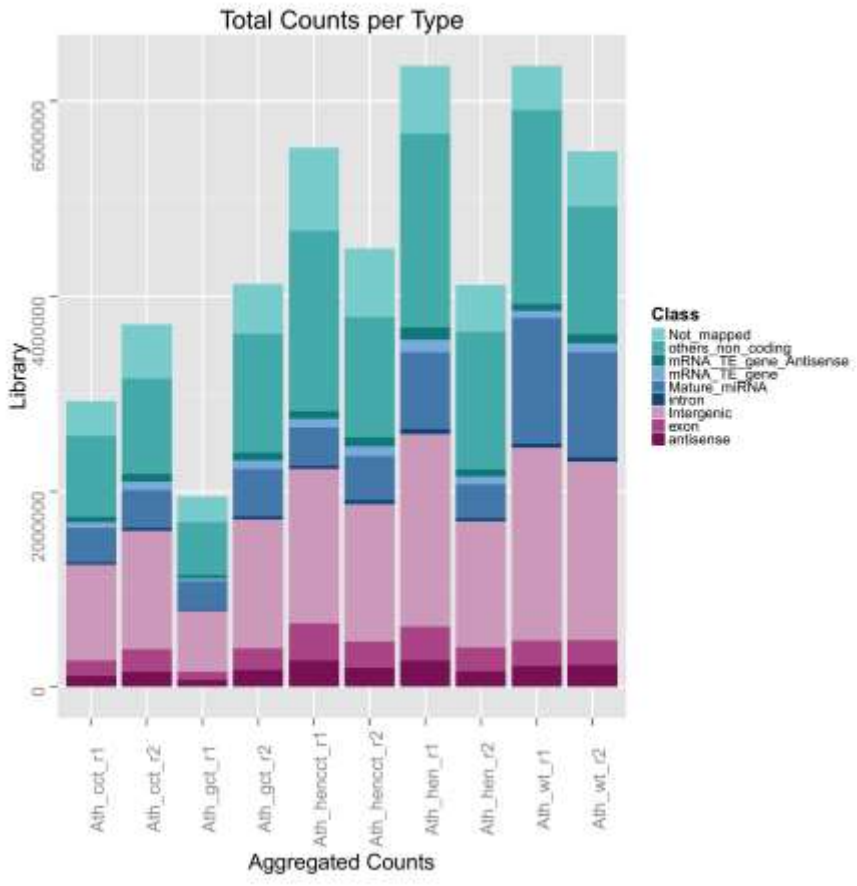


Figure 13. **Total annotated classes for all libraries.**

Figure 14 shows the distribution of annotated reads, according to the read length. Notably, the 24nt sequences have a large proportion of reads mapping to unannotated regions of the Arabidopsis genome. Most of the 21nt reads, the canonical miRNA length, map to regions annotated as miRNAs. This behavior is consistent across all conditions. Figure 14b presents the same information as relative abundance of reads.

The distribution of annotation by length for all the libraries is in the Appendices (43-52).

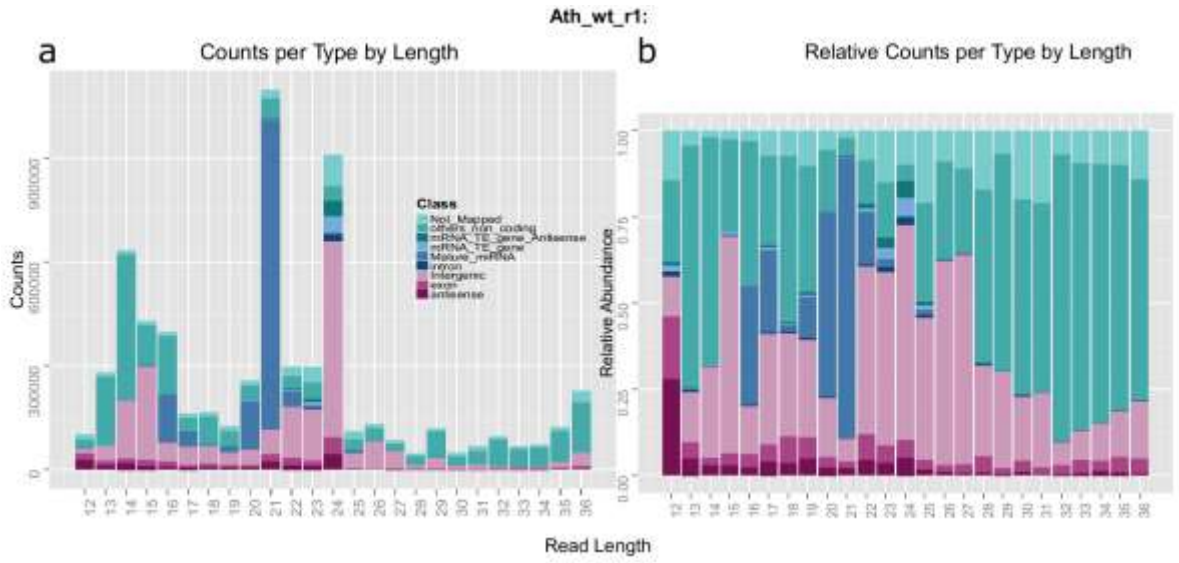


Figure 14. Distribution of annotated reads according to the sequence length. a) Raw counts of a wt library, b) the same library showing the relative abundance of each biotype.

IV.5 miRNA annotation

Figure 15 summarizes the ten most expressed families of miRNAs after library normalization. This figure shows the consistency between different libraries (in this case wt and *hen3*), with miR166 and miR158 being the two highest expressed miRNAs. The top ten most abundant miRNAs for all libraries are in the Appendices (53-57).

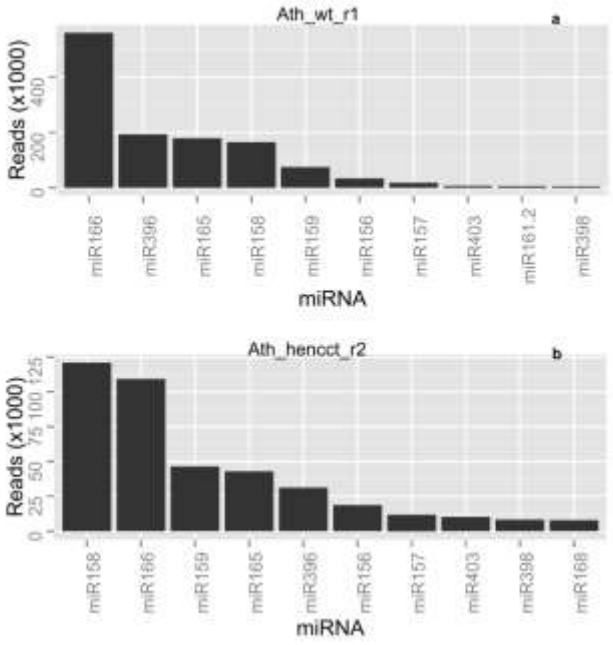


Figure 15. **The ten miRNAs with the highest counts in a representative wild type and mutant library.**

IV.6 Differential Expression

The basis for differential gene expression (DGE) analysis is to compare data under the null hypothesis that any given condition (a mutant genotype, for example) does not have an effect on the expression of the majority of genes. Thus, the data is adjusted to fit a model that accounts for biological and technical variation, and then normalization factors are calculated. These factors are mostly used to correct for differences in the sequencing depth of the libraries.

In any given RNA-seq experiment, the abundance of read counts coming from the same gene in different samples will vary; this is due to the sampling process (selecting RNA molecules from a larger pool), to measurement errors due to the technology used, and biological noise (how gene expression actually varies between cells and organisms). Sequencing experiments sample only fragments of the entire transcriptome, leading to random sampling errors. This creates a problem when aiming to calculate the abundance of each transcript in the population.

Modeling the data with a negative binomial (NB) distribution estimates variation between conditions beyond random sampling, including technical and biological variation. Adjusting

RNA-seq data to an NB distribution also allows gene-specific variation to be estimated. The Biological Coefficient of Variation (BCV) takes into account all sources of variation, from technical to biological among samples; the difference between NB and Poisson distributions accounts for the BCV (McCarthy, Chen, and Smyth 2012; Rapaport et al. 2013).

DGE assesses the difference between the observed counts between two conditions for all genes. If the change is greater than the expected variation would allow, then it is predicted to be differentially expressed (DE). The magnitude of DE is reported in terms of a logFC. It represents the log2 of the ratio between the expression levels in two conditions. This value can be either positive or negative, indicating the direction of change.

For the differential expression analysis I focused solely on the genes annotated as mature miRNAs in the sRNA-seq files. However, results from our group demonstrate that GCT and CCT are involved in regulation of miR156 and miR172 to control developmental phase transitions, while the expression of some other miRNAs does not change in the mutants (Gillmor et al. 2014). Hence we can consider that changes in miRNA abundance can be due to mutations in *GCT* and *CCT*.

Multidimensional scaling (MDS) measures the global similarity or dissimilarity among different data sets at different levels (dimensions). The outcome is then represented in a graphical manner as points in a space (commonly in two dimensions) with the distances between them reflecting their differences. Figure 16 represents an MDS plot showing the relationships between samples. Libraries coming from wild-type plants separate from the mutant genotypes along the first dimension (X-axis). Along the second dimension (Y-axis) the libraries from *hen3* and *gct* genotypes are grouped together, while the *cct* and the double mutant *cct/hen3* form their own clusters.

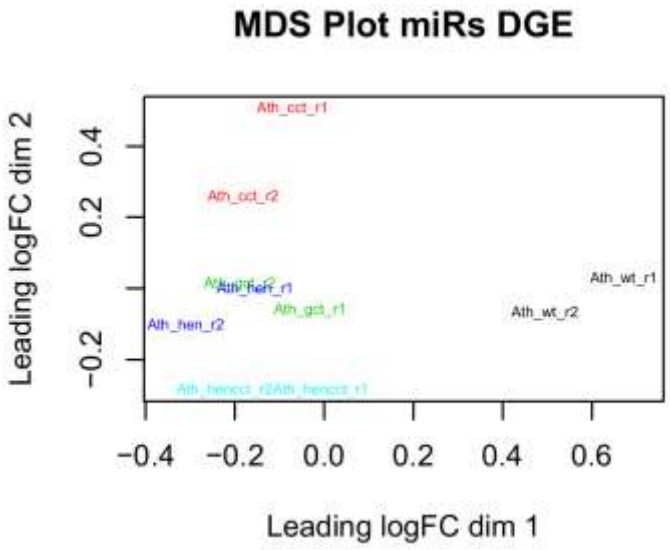


Figure 16. MDS plot for miRNA data. Multiple libraries are transformed and represented in two dimensions. Libraries coming from similar genotypes group together forming clusters within the plot.

In Figure 17, the BCV of each gene is plotted against its expression level. Common variability (red line) although at a high BCV value (almost 0.25), is constant. However, as the trend (blue line) shows, the higher the expression of a gene the higher its biological variation.

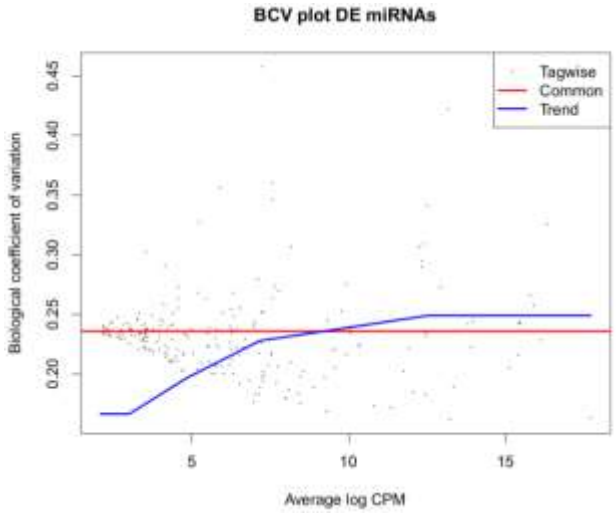


Figure 17. Graph of the relationship between biological variance and level of expression. Each dot represents a miRNA.

I evaluated the effects of the different genotypes on the expression of miRNAs comparing each mutant against the wild type reference. Thus, I made four different assessments: *cct*, *gct*, *hen3* and the double mutant *cct/hen3* compared to the miRNA expression in wt. Figure 18 shows a MA plot representing the relationship between the logFC and the mean expression level for each miRNA in the DE analysis of *cct* against wt. miRNAs that participate in developmental processes such as phase transitions or leaf development (Mallory, Reinhart, et al. 2004; Nikovics et al. 2006; Wu et al. 2009) are highlighted and color coded. Red dots indicate differentially expressed miRNAs. Each name corresponds to a single point representing one miRNA; note how miRNAs from the same family cluster together in the graph.

The sign of the logFC value represents the direction of change. In this case, a positive value indicates that counts in the mutant conditions were higher than in the reference wild type. On the other hand, a negative logFC indicates that the expression in the wild type was higher than the mutant.

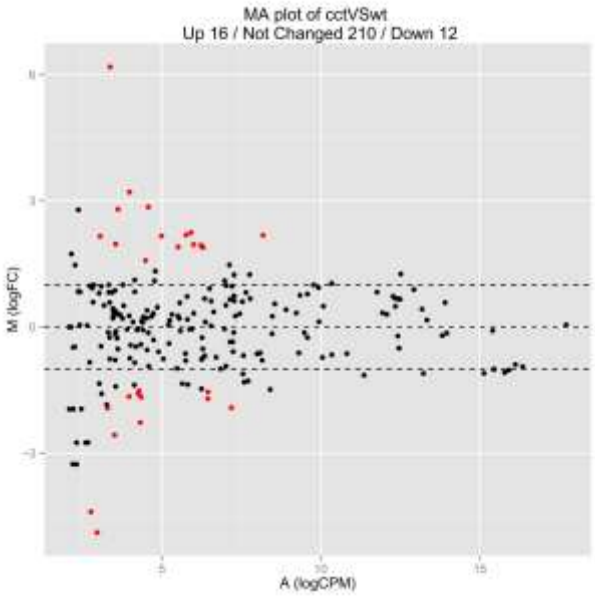


Figure 18. MA plot of cct VS wt. Some miRNAs involved in developmental processes are highlighted. The X-axis represents the average level of expression (log2 of Counts Per Million sequenced reads) ranging from low to high (left to right). The Y-axis represents the ratio of change between of gene expression (log2 of fold change).

The MA plot (figure 18) highlights the relationship between average expression and ratio of expression between conditions for all 238 miRNAs analyzed. Each point represents a miRNA; the X axis (A value) is the average expression level for each miRNA while the Y axis (M value) represents the change of expression between the mutant compared to the wild type. Points in red represent miRNAs whose expression change is statistically significant. MA plots for the rest of the contrasts are in Appendices (58-61).

IV.7 Common misregulated miRNAs in all conditions

Figure 19 represents miRNAs whose change (up or down regulation) is either specific for each contrast or common across conditions. Areas of the circles that don't overlap with others display the number of miRNAs whose change is specific for each condition (for example there are 13 miRNAs uniquely misregulated in *cct*). 32 miRNAs are commonly misregulated across all four contrasts.

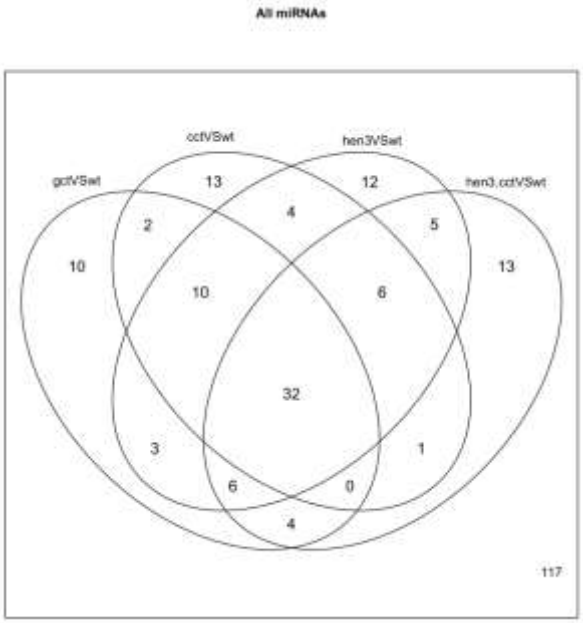


Figure 19. Venn Diagram of miRNAs changing across different conditions. From all the 238 miRNAs analyzed, 32 are commonly misexpressed in all conditions.

After assessing the DE results for each condition (Appendices 62), I generated a heatmap (Figure 20) in order to visually represent the changes in gene expression common across all

conditions (Fig19, overlap of all circles). The X-axis represents the evaluated conditions, while the Y-axis includes all miRNAs that were predicted as differentially expressed in one or more conditions. Colors tending toward the blue spectrum represent overexpression of the miRNA in the mutant relative to the wild type condition; the red spectrum indicates repression in the mutants. Heatmap in Appendices 62 shows the change in expression prior to filtering by FDR.

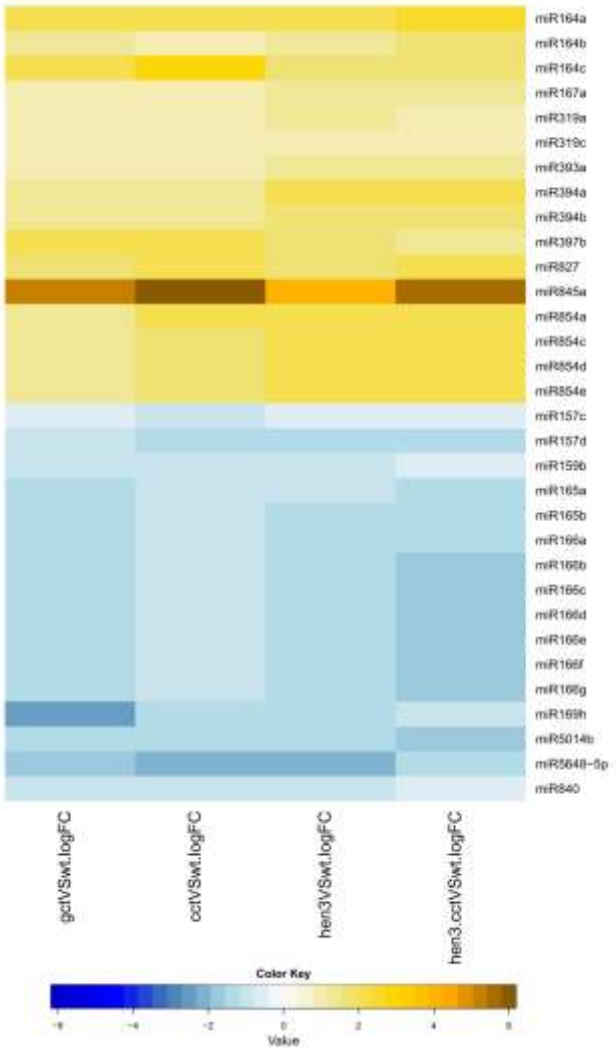


Figure 20. Heatmap of significantly DE miRNAs across all conditions. miRNAs (Y-axis) showed only one of two states across conditions (X-axis): they were either up- or down-regulated.

Since each of the studied genotypes represents a part of the same inhibitor complex, I was interested to find genes that followed the same mode of regulation across all conditions. These

miRNAs would likely be regulated by the full CDK8 module. To do this, I selected miRNAs that significantly changed in all conditions (FDR ≤ 0.05) even if the magnitude of change was small. In total, 32 miRNAs passed these filters. From the differentially expressed set, 16 miRNAs were more abundant and 16 were less abundant across all conditions (Figure 20).

Two whole families of miRNAs showed either higher or lower abundance, respectively. The miR165/6 family was repressed by mutations of CDK8 subunits. Relative expression was reduced to a quarter compared to the wild type (logFC = -2). Conversely, the miR164 family turned out to be consistently more abundant in the mutants compared to wild type.

The miRNA with the largest logFC value was miR845a, but its average expression is very low in normalized libraries. Furthermore, very little is known about its biological function. One study reports that it is conserved and expressed in *Brassica napus* seeds (Table 09) (Huang et al. 2013). The raw counts for this particular miRNA were also low. Assessment of the mature miRNA sequence to look for putative targets using psRNATarget (Dai and Zhao 2011), found mostly transposable elements matching to the miRNA mature sequence, indicating inhibition by cleavage (Appendices 64).

Chapter V - Biological interpretation of bioinformatics results.

Previous results from our lab demonstrated that *CCT* and *GCT* affect expression of miR156 as well as miR172 (Gillmor et al. 2014); the miR156 family is overexpressed in mutants of the CDK8 module, while miR172 is repressed. In my differential expression analysis, however, change of the miR156 family was not significant. Despite of this, our data showed miR169 repressed in all conditions (fig. 19); this miRNA is reported to have a role during transition to flowering; as miR156, this miR169 is also responsive to temperature changes in Arabidopsis (H. Lee et al. 2010).

Of the 32 misregulated miRNAs found in this study, 16 of them were overexpressed in the mutant plants. Surprisingly, most of them are reported to be involved in stress and abiotic responses (table 10). It is possible that these miRNAs also have a function in Arabidopsis development. As an example of genes participating in both developmental and stress responses, Taylor-Teeples and collaborators (2014), reported that genes participating in xylem cell specification are also activated in response to iron, salt and sulphur stresses.

Thus, deregulated miRNAs in our analysis that are reported to change in stress responses could also have a yet-to-discover function in developmental processes. This also points a role for subunits of the CDK8 module as an integrator of external signals for the plant to respond.

V.1 miR165/166 family

The downregulation of miR166 can be explained through a more complex layer of genetic control involving one (or multiple) intermediate(s) controlling the miRNA through CDK8 (fig. 21b). The alternative would be for the CDK8 to act in this case as an activator instead of its canonical repressive function (Nemet et al. 2013).

A study from Ochando and collaborators (Ochando et al. 2006) reports that a gain of function allele of CORONA (a target of miR165/166) shows a late flowering phenotype similar to that of *cct* and *gct* mutants. Interestingly, the allele they studied is insensitive to miR166 regulation; plants carrying this allele showed a late flowering phenotype in addition to an increased number of rosette leaves, indicating an expanded vegetative phase similar to the phenotype of *cct* and *gct* mutants.

A recent study showed that loss of miR166 regulation increases expression of seed maturation genes (Tang et al. 2012). Furthermore our group found a role for CCT and GCT in the repression of the seed maturation program in Arabidopsis. From microarray analyses found that seed-specific genes were overexpressed in the *cct* and *gct* mutants (Gillmor et al. 2014). Upregulation of these genes involved in seed maturation program could be explained by the downregulation of the miR166 family that we found in this study.

V.2 miR164 family

The miR164 family is known to participate in different developmental processes. It is comprised of 3 members, namely miR164A, B and C. Our analysis indicates consistent upregulation of the whole family; the mean logFC value indicates a twofold increase in expression all conditions (mean logFC ~2, FDR < 0.05).

In 2005, the Meyerowitz group reported that miR164C controls petal number independently of miR164A or miR164B (Baker et al. 2005). The Bartel group studied a null allele of miR164B, but did not find a phenotype in aerial tissues of seedlings, suggesting that miR164A and miR164B might be partially redundant (Mallory, Dugas, et al. 2004). Guo and collaborators (Guo et al. 2005) reported that mutations in either miR164A or B resulted in the development of more lateral roots in *Arabidopsis* supporting the idea of these two miRNAs are partially redundant

The Meyerowitz lab also studied the effects of abolishing expression of the entire miR164 family. They concluded that, although the functions of the three miRNAs partially overlap, they also present a degree of functional specialization. The *miR164abc* triple mutant plants displayed minor differences with the wild type in vegetative states; however, they showed more aberrations when plants transitioned to flowering stages. Decreasing expression of miR164 affected leaf margin at the vegetative stage and floral organs and their phyllotaxis at the reproductive stage (Sieber et al. 2007).

Interestingly, Laufs group reported that a null allele of miR164A affected vegetative leaf traits by showing an increased level of serrations at the margin (Nikovics et al. 2006). Higher serrations at leaf margins were phenocopied by expressing a resistant version of one of

miR164 targets, *CUC2*, but not *CUC1*. This suggests that the miR164A-*CUC2* module is in charge of controlling margin development in Arabidopsis leaves.

V.3 Exploring upregulated miRNAs in CDK8 mutants

Since the CDK8 complex acts as a repressive module that blocks the interaction of Mediator with RNA pol II (Elmlund et al. 2006), I focused on the microRNAs that had a positive logFC across conditions, reasoning that if the CDK8 directly represses transcription of the miRNA (Fig. 21a) we would see an overexpression, as in the case of miR164. Conversely, if regulation occurs through an intermediate, like a transcription factor, that in turn represses miRNA expression (Fig. 21b) the result would be a repression of miRNA expression, which is a likely scenario for the case of miR166.

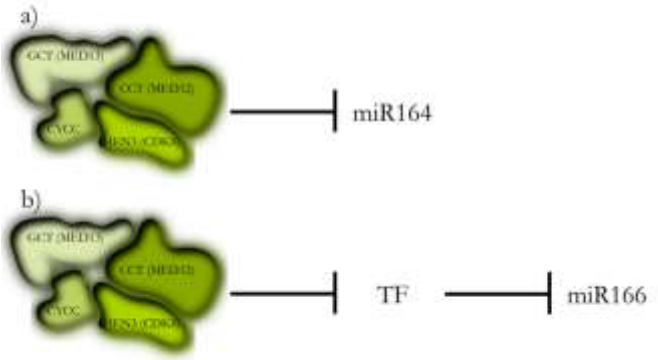


Figure 21. Plausible scenarios to explain the CDK8 module regulation of its different targets. In a), the mode of regulation is direct while in b), regulation of the targets occurs through a cascade of negative regulation.

I summarize the 16 miRNAs that met our criteria in Table 10. Notably, many of them are reported to participate in abiotic stress responses and just a few participate in developmental processes.

Table 10. Functions of miRNAs that were significantly upregulated in all mutants.

miRNA	Function	Short Description	Reference
miR164a	Controls leaf serrations.	Leaf/Root.	(Nikovics et al. 2006)
miR164b	Partially redundant with MIR164a.	Root/leaf.	(Mallory, Dugas, et al. 2004)
miR164c	Controls petal number.	Petals.	(Baker et al. 2005)
miR167a	Salinity and drought.	Abiotic stress.	(H. Lee et al. 2010)
miR319a miR319c	Distorted leaf development. Late flowering. Male sterility.	Regulates TCP and leaf morphogenessis.	(Schommer et al. 2012)
miR393a	Salinity, cold and drought.	Abiotic stress.	(H. Lee et al. 2010)
miR394a miR394b	Salt and cold stresses. Bacterial immunity.	Abiotic stress.	(H. Lee et al. 2010; Sunkar 2012)
miR397b	Cold stress.	Abiotic stress.	(H. Lee et al. 2010)
miR827	H_2O_2-responsive miRNA. Phosphate response.	Abiotic stress.	(Sunkar 2012)
miR845a	Seed development in *B. napus*.	Seeds.	(Huang et al. 2013)
miR854a miR854c miR854d miR854e	Hypoxia. TE-derived miRNA.	Abiotic stress.	(Arteaga-Vázquez, Caballero-Pérez, and Vielle-Calzada 2006; Piriyapongsa and Jordan 2008; Sunkar 2012)

V.4 Gene Ontology enrichment analysis of miRNA targets

In order to investigate predominant biological processes misregulated in the mutants, I did an enrichment of GO terms using the BinGO plugin from Cytoscape (Maere, Heymans, and Kuiper 2005; Shannon et al. 2003), taking the predicted and confirmed targets of all

upregulated miRNAs common to all mutants. Zhang and collaborators (Zhang et al. 2009) created a public repository for plant microRNA information, integrating available information into a single database. Additionally, they provide a curated Arabidopsis miRNA target list that I used in this study. The results are in Figure 23. The overexpression of these miRNAs in mutants would suggest that their targets are normally not repressed in wild type conditions. The biological processes described in figure 23 should thus be commonly affected in the CDK8 module mutants. Predictions include developmental processes like shoot formation, root and shoot meristem development, morphogenesis and metabolic and signaling processes.

Pollen development and maturation is one of the clusters in the network of GO terms that is affected by these mutations. This is consistent with reports from our group showing that *gct* and *cct* mutants are sterile (Gillmor et al. 2014). Formation of organ boundaries is another process predicted to be affected. Stamens are occasionally fused in *gct* and *cct* mutants. Interestingly, this term is connected to other nodes related to organ morphogenesis. Often, mutations in subunits for the CDK8 module have malformed cotyledons; in rare cases they are fused together displaying a cup-shape cotyledon phenotype (figure 22). Note how the wild type (figure 22a) cotyledons are well defined whereas the cotyledons from *gct* (figure 22b) failed to separate and form a cup-shape.

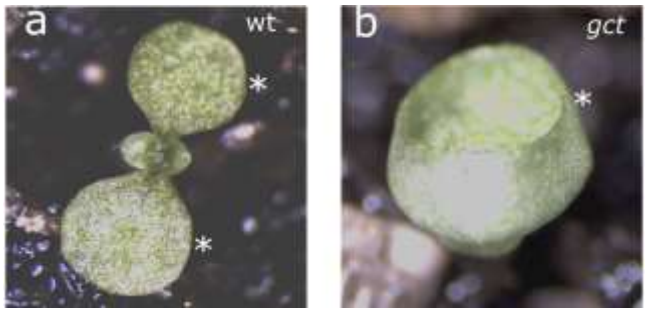


Figure 22. Examples of wild type and gct plants at 5 days after germination. Asterisks mark the cotyledons. The wild type (a) has normal cotyledons whereas mutant plants (b) display fused cotyledons (cup-shaped cotyledon phenotype).

The inositol triphosphate metabolic pathway regulates calcium concentration in the cytoplasm. In turn, calcium is required for the regulation of various cell developmental processes. Calcium concentration is also important for polarized growth (Hepler 2005). Interestingly, *gct* and *cct*

were discovered through a screening of genes affecting polarity in Arabidopsis embryos (Gillmor et al. 2010).

V.5 Analysis of miRNAs in individual mutants

To understand the specific differences between individual CDK8 subunits and wild type, I selected miRNAs that were differentially expressed in each condition. In contrast to the previous analysis, this approach is aimed at discovering subunit-specific miRNAs.

In *cct,* I found 18 significantly upregulated miRNAs whose change in expression was greater than 1.5 fold-change relative to the wild type. Biological processes suggested to be affected in *cct* mutants are shared with those of the general CDK8 subunits (figure 24). Following the same trend, individual analysis of the other mutants showed similar affected biological processes. Among those that were specific was lipid A biosynthetic process, only present in *gct* and *hen3/cct* double mutants. In the single mutant *hen3* the auxin mediated signaling pathway was also affected. The GO enrichment networks for the rest of the genotypes are included as Appendices (65-69).

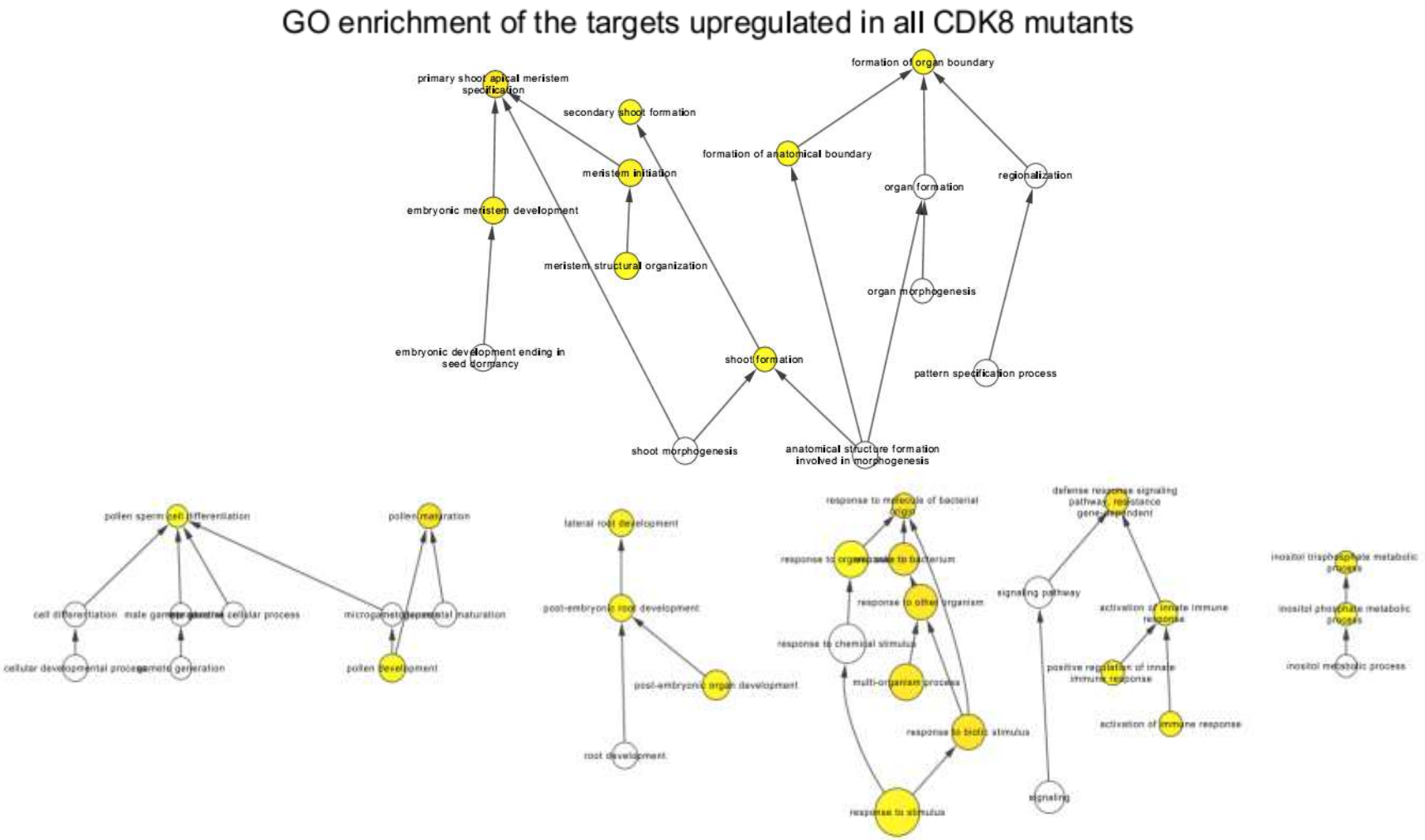


Figure 23. Biological processes related to targets of overexpressed miRNA in CDK8 mutants. The full network is shown in Appendices (69) and represents a hierarchical tree of GO terms. This image shows the upmost part of the tree, which represent more specific GO terms. The size of the node is proportional of the number of genes in that category. White nodes were not significantly enriched as opposed to colored nodes.

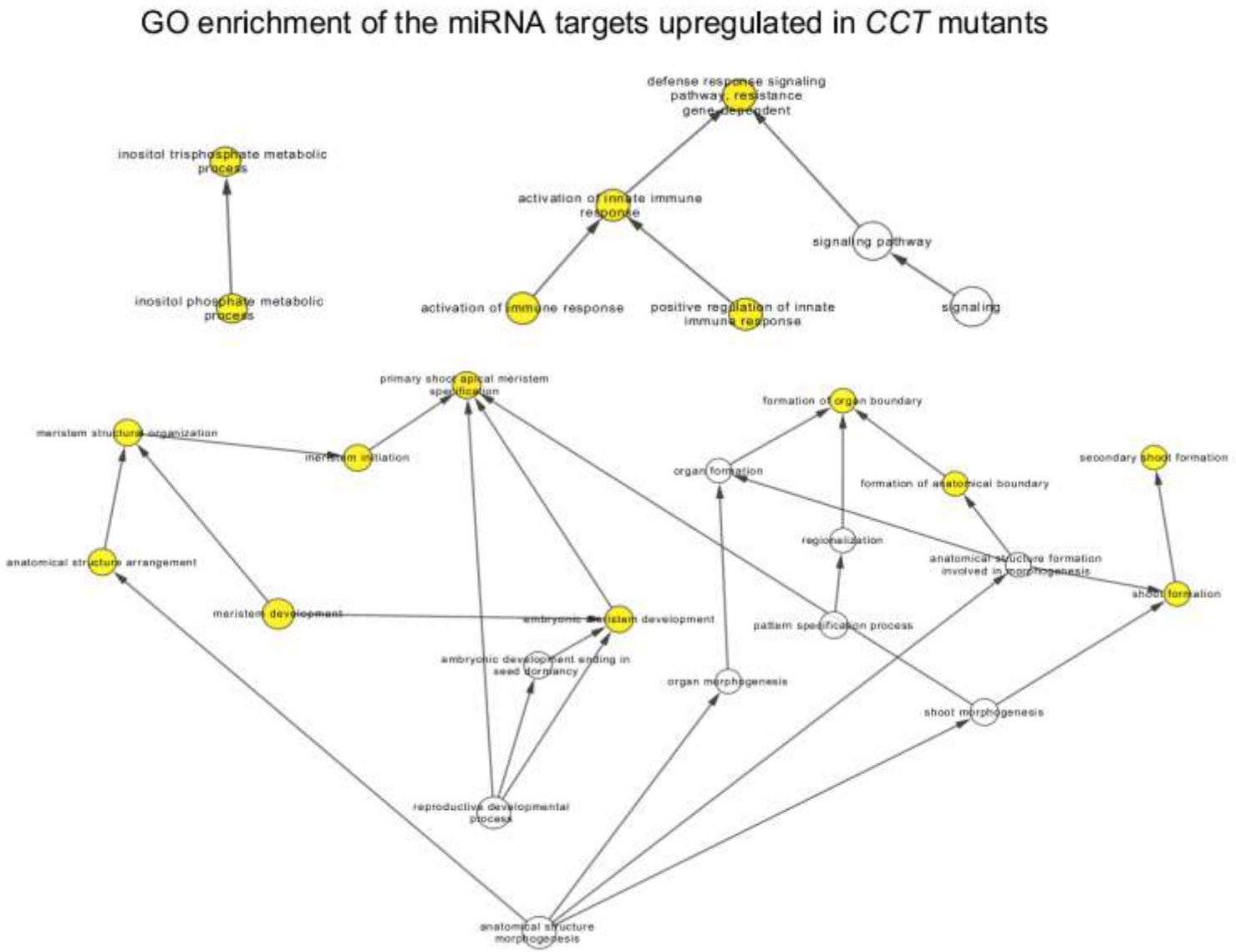


Figure 24. Biological processes related to overexpressed miRNA targets in cct mutants. Three different clusters can be distinguished. The size of the node is proportional of the number of genes in that category. Yellow nodes are significantly enriched as opposed to white nodes.

GO terms were enriched in developmental processes that recapitulate the mutant phenotypes of the plants, such as pollen maturation and formation of organ boundaries reflecting the sterility and the *CUC* phenotype of *cct* and *gct* plants. Interestingly, enriched biological processes such as organ boundary formation and pollen maturation were consistent in the 16 miRNAs set and the specific sets for individual genotypes.

My results, in conjunction with reports from the literature, demonstrate the importance of the CDK8 module for the correct regulation of miRNAs involved in developmental processes. These miRNAs have to be tightly regulated in space and time for the correct progression of developmental programs. Future work could benefit from the study of the role of miRNAs in development and subsequently the role of the CDK8 module in regulating miRNAs that might also control developmental transitions.

Based on the results I obtained, we opted for a more detailed analysis on the spatial and temporal regulation of miR164 by the *CCT (MED12)* subunit, and its impact in Arabidopsis development

V.6 Experimental validation on CCT regulating miR164A

To test if the CDK8 module acts upstream of miR164A and at what extent it plays a role in leaf morphology we explored the role of *CCT* in the spatial and temporal repression of a miRNA involved in leaf development. We used transcriptional marker lines for miR164A in wild type and *cct* mutant genotypes in Arabidopsis leaves at different time points. This approach would permit us to characterize the spatiotemporal expression pattern of miR164A in absence of *CCT* regulation.

For this, I decided to focus the interaction between the *CCT (MED12)* subunit and MIR164A, a miRNA I found to be regulated by the CDK8 module, and which is known to affect the development of leaf serrations (Nikovics et al. 2006).

Since the CDK8 module mutants had previously been shown to affect timing of leaf serrations (Gillmor et al., 2014), the study of MIR164 regulation by *CCT* allowed me to integrate the previously identified spatial role of miR164 in creating serrations on the margins of leaves, with a newly defined temporal role for *CCT* in controlling the timing of onset of leaf serrations by potentially regulating miR164. The study of *CCT* regulation of miR164 was also a good fit due

to the interest of our lab in temporal regulation of leaf development, and because many genetic stocks for the study of miR164 were already present in the Gillmor laboratory. I selected to focus on *CCT* (instead of a different CDK8 module gene) due to the higher penetrance of the *cct* mutation relative to other phenotypes of the CDK8 module.

V.7 Expression pattern of MIR164A using GUS marker

I generated a $cct^{+/-}$ line expressing GUS under the MIR164A promoter (pMIR164A:GUS) and evaluated the GUS expression patterns of wild type and *cct* homozygous plants at different time points (explained thoroughly in the Materials and Methods chapter). GUS staining reveals a blue color in the region where the gene is expressed. In the wild type, MIR164A expression begins at the primordia of the developing leaf, observed along the margin of the new leaf. As the leaf develops, expression due to the miRNA promoter is restricted to the vasculature and the tips of the serrations in leaves (Nikovics et al. 2006). If *CCT* negatively regulates miR164A we would expect ectopic expression in the *cct* mutant background. I show the preliminary results in figure 25.

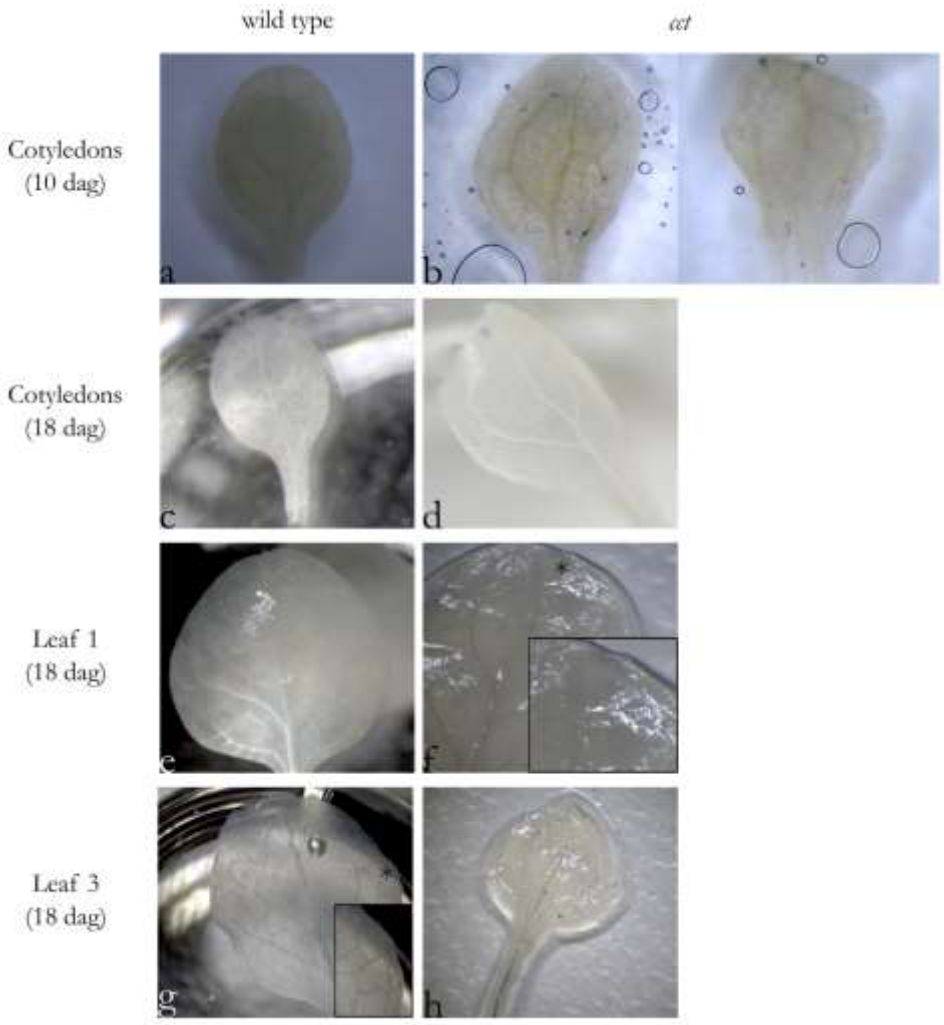


Figure 25. *MIR164A* expression in individual leaves. First column is the wild type background; second column is *cct* genotype. Each row represents a different leaf at the indicated time point. In general, *cct* leaves have more GUS signal than their wild type counterparts. Insets highlight GUS expression.

Figure 25 (a-d) displays cotyledons from wild type and *cct* plants. In addition to the abnormal morphology of *cct* leaves, they display small areas of blue GUS signal whereas wild type organs show no GUS signal. In the first and third leaves (fig. 25 e-h), the wild type leaf again lacks GUS signal, whereas the *cct* leaves have blue staining (although less than in cotyledons).

Our preliminary results of this analysis suggest ectopic expression of miR164A in *cct* seedlings relative to their wild type counterparts, although this need to be further explored.

V.8 Design of a pMIR164A:eGFP transcriptional marker

To enhance our analysis on the expression pattern of miR164A, I designed and constructed transcriptional markers lines expressing eGFP (see Materials and Methods).

Due to time constraints further analyses on the lines are still needed to ensure that not only transformation did not affect plants but that the markers are being expressed correctly. Thus, screening leaves from plants using confocal microscopy and genotyping/phenotyping *cct* homozygous plants to make certain that the transgene is fixed and the mutation is segregating.

Figure 26 shows preliminary results from a screen test of plants carrying the pMIR164A:eGFP transgene. It has a endoplasmic reticulum (ER) localization signal.

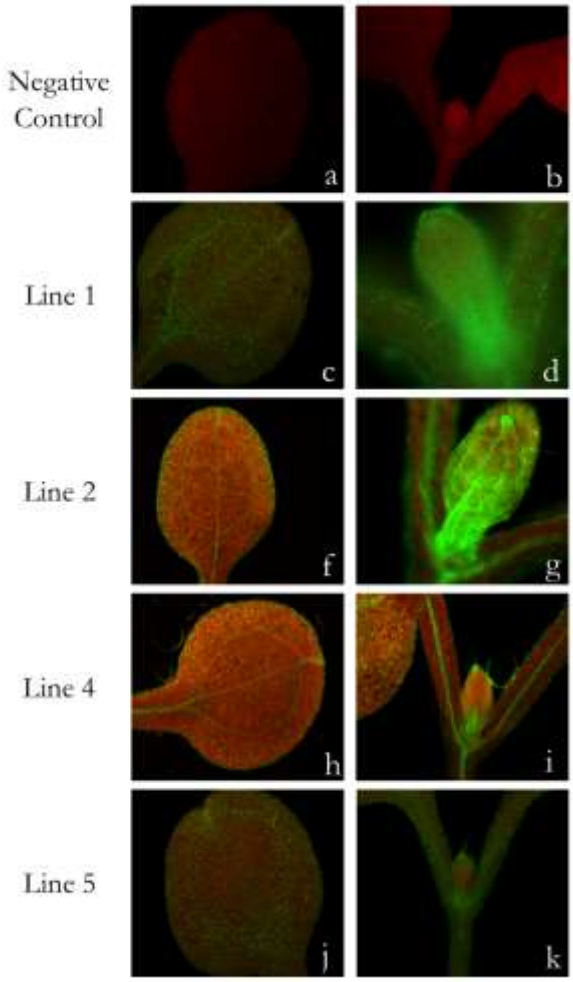


Figure 26. Expression pattern of eGFP under the promoter of MIR164A. These images are the representative

pattern of different plants from the same lines. The signal on line 3 only in the tips of the cotyledons (data not shown).

Figure 26 demonstrates the transformation process was successful. The expression pattern from our lines is in accord with that reported in wild type GUS lines (Koyama et al. 2010).

One possibility is that our pMIR164A:GUS cross with *cct* plants caused a misregulation on the expression of the transgene. However, results from the GUS expression pattern in *cct* are still preliminary and have to be further analyzed to be able to confirm that CCT indeed represses the expression of MIR164A. The eGFP results offer the possibility to investigate the pattern *in vivo* in both wild type and mutants to enhance our analysis.

Conclusions

This study focused on the general transcriptomic landscape of miRNAs in wild-type and mutants in subunits of the CDK8 module of Arabidopsis Mediator. I characterized the small RNA-seq transcriptomes through a bioinformatics pipeline that uses a combination of open software and custom programs.

Overall, I found that the expression of 32 of miRNAs is altered (induced or repressed) by mutations in the subunits of the CDK8 module used in this study. Whether the alterations are due to direct or indirect regulation by the CDK8 module is beyond the scope of this work. Ongoing work in the lab aims to uncover how the CDK8 module regulates not only miR156 in developmental transitions but also if this module also controls targets independently of this miRNA, such as directly affecting the SPL transcription factors that are misregulated in *gct* and *cct* mutants (Gillmor et al., 2014).

The miR319 and miR164 families were among the upregulated genes known to be involved in developmental processes. Misregulation of miR319 and miR164 families has an impact on leaf development in Arabidopsis. As an example, overexpression of miR164A generates leaves with smooth margin, consistent with the leaf phenotype of CDK8 mutant plants (Gillmor et al., 2014). Our results show increased expression of both of these miRNAs.

Based on our bioinformatic results we asked if the CDK8 module is indeed repressing miR164A and if this could account for the lack of leaf serrations in the mutants. We addressed this question using genetics and molecular approaches.

As a first approach, results may point to a role that CCT represses miR164A and that, at some extent, this could explain the leaf margin phenotype on the mutants, yet, more analyses on the GUS expression pattern are needed to validate our hypothesis.

Preliminary results from the eGFP lines demonstrate that the transformation was indeed successful and comparing with the literature, the expression pattern recapitulates that of the GUS expression. Further work is needed to select homozygous plants carrying the eGFP

transgene and look for plants segregating *ct* mutants or cross wild type plants with CDK8 mutants to observe differences in expression.

Discussion

The approach I presented here aimed to discover microRNAs regulated by the CDK8 module of the Arabidopsis Mediator complex. Through a genome wide analysis we used NGS technology to generate data of small RNAs expressed in plants with mutant subunits of the CDK8 module, using the wild type genotype as a reference. For this purpose the aerial tissue of 18 day old plants was used; at this stage plants undergo the transition from vegetative to reproductive phase.

The use of custom scripts and open source programs was useful not only to allow us to make use of our own datasets for annotation (instead of relying on the default settings of available programs), but also to be able to choose from a different range of programs designed for this purpose. This pipeline proved useful not only for this study but it was also adapted to analyze sRNA-seq data from a non-model organism in collaboration with the group of Alfredo Herrera-Estrella.

It is possible that the differences in developmental stages between wt and mutant plants may have an impact on our results. However, our goal of this first exploratory study was to assess the differences in miRNA expression role in mutants for the CDK8 module at a stage approximating vegetative to reproductive transition.

Our annotation algorithm allowed us to identify different genes being expressed in our libraries. While other reports studying small RNAs also focused on mature miRNAs (F. Chen et al. 2010; Huang et al. 2013; Zhou et al. 2010) it could prove useful to explore other sRNAs potentially regulated by the CDK8 module of Mediator, since miRNAs are only a small part of the complex pathways of sRNAs and their regulatory biological processes (Sunkar 2012).

Results from the distribution of annotated reads for different sequence lengths indicates that 24nt sequences are mapping to unknown regions of the genome (fig. 14). This could suggest an active process of gene silencing through canonical 24nt siRNAs. Thus, a future study could assess the importance of the CDK8 module on gene silencing through the RdDM pathway.

For this work, I focused mainly on the deregulated miRNAs that were common across all conditions. I found 32 misregulated miRNAs. Due to the fact that CDK8 is a transcriptional

repressor, I then focused on miRNAs that were likely to be directly regulated by the CDK8 module (for example as in Fig 21a), i.e. those that had a positive logFC.

Two different approaches can be used to discover new miRNAs; the first is through the use of bioinformatics tools to predict potential miRNA regions for example (Adai et al. 2005; Arteaga-Vázquez, Caballero-Pérez, and Vielle-Calzada 2006; W. Park et al. 2002). This approach permits identification of novel loci actively expressing microRNAs, but it doesn't evaluate their biological function. A second approach involves screening mutants with affected developmental processes and mapping the mutation to a MIR gene; the case of miR319 (Schommer et al. 2012; Sunkar 2012) and miR172 (Aukerman and Sakai 2003) are examples of miRNAs found this way.

Conversely, in Arabidopsis, many miRNAs reported as misregulated during stress responses were found using high throughput technologies measuring changes in expression of small RNAs in different abiotic conditions (H. Lee et al. 2010; Piriyapongsa and Jordan 2008; Sunkar 2012). Interestingly some of the miRNAs known to have a function in developmental processes often are reported as also being misregulated during stress responses (H. Lee et al. 2010).

Taken together with previous results from our laboratory showing that CCT regulates leaf traits such as leaf shape, abaxial trichome production, and leaf hairs by regulating the miRNA miR156, this experimental approximation suggests a role for CCT in controlling production of leaf serrations by repression of MIR164A. The result that CCT regulates miR164 in addition to miR156 suggests that the CDK8 module of Mediator acts as a master regulator of heteroblasty by regulating multiple miRNAs.

The role of the CDK8 module as a repressor suggests that miRNAs whose levels decrease in CDK8 module mutants may be indirect targets of the CDK8 module. The case of miR165/166 and its target CORONA is an interesting candidate to study indirect regulation of a miRNA and also of its target; as the dominant negative CNA allele presents a late flowering phenotype similar to *cct* and *gct* mutants. Exploring regulation of these miRNAs could shed light on the downstream targets of the CDK8 module and how it acts at different levels in regulatory cascades. Thus, an interesting future direction would be to investigate how the

miR166-CORONA regulation is affecting the miR156 cascade to control flowering time and in turn how are these modules regulated by the CDK8 subcomplex.

A more thorough analysis of the data could shed light on the general role of subunits of the CDK8 module in the regulation of important small RNAs such as those involved in gene silencing.

Including other biotypes such as transposable elements and other sRNAs in the DGE analysis of the transcriptomes could broaden our understanding on the involvement of the CDK8 module in other regulatory pathways such as epigenetic silencing. For example, analyzing differential expression of sense and anti-sense transposable elements could indicate a role for subunits of the module in the silencing of these genes. The RdDM pathway works with two different polymerases (RNA pol IV and V) in this process. Although Mediator and thus, the CDK8 module, is known to interact with RNA pol II, this would expand our view on the specificity or plasticity of the CDK8 module interactions.

For the in silico analysis of related GO terms I reasoned that the mRNA targets of upregulated miRNAs would be repressed in the mutants. I decided to focus on the upregulated miRNAs due to the canonical function of the CDK8 module as a repressor of transcription. This is certainly a coarse approach, due to some of them possibly not being expressed at all for reasons like tissue specificity or due to the age of the plants. However, as a first strategy, it offers a general overview of the biological processes affected in CDK8 subunits mutants and how they correlate with the displayed phenotypes.

To investigate the role of all these miRNAs in development would require a thorough analysis; first, validating the role of the CDK8 module in controlling such miRNAs and then exploring their involvement in development using target mimics and/or overexpressing transgenic lines. These experiments were out of the scope of my thesis.

A twofold strategy to validate the role of the CDK8 module on the regulation of miRNAs would be to generate marker lines to follow the expression pattern of subunits of the module and the miRNA. If the CDK8 module represses the miRNA, their expected patterns would be complementary. The second approach is to investigate the developmental function of individual miRNAs generating mimicry and overexpressing lines.

Even though our bioinformatics results show an increase of miR164A expression, our results show slight ectopic expression of the GUS signal. This can be due to low expression of the miRNA. Counts in wild type plants for the miRNA family are low (fig. 26); however these changes are enough to be detected by our differential expression analysis as significant. This could account for the small differences in the GUS analysis. Furthermore, since the sequence of the three members of the MIR164 family is highly similar our mapping step could produce artifacts. Due to the lack of replicates and a low GUS signal more experiments are needed to test our hypothesis.

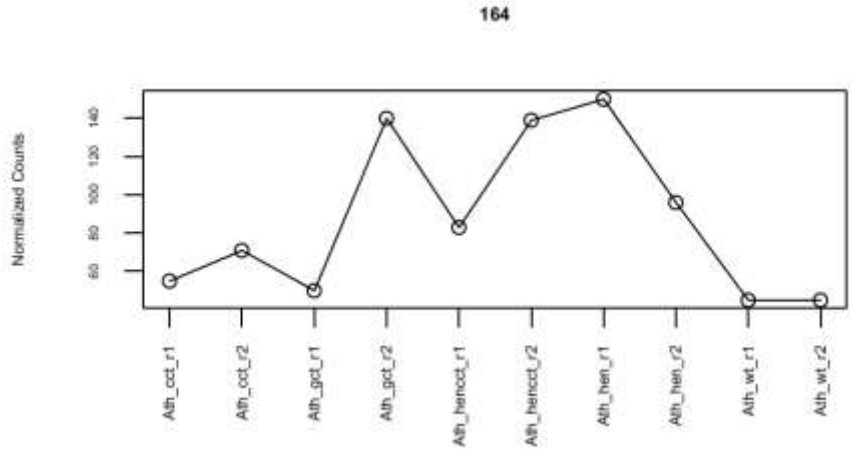


Figure 27. Normalized counts (TPM) of miR164A family.

More detailed analyses on the spatiotemporal expression of miR164 using our transcriptional reporters are necessary to validate our hypothesis, since the results from our first exploratory approximations are still inconclusive.

The fact that the miR164 family also affects other leaf traits such as senescence could point to a more general role of the CDK8 module in regulating not only developmental growth but also age induced cell death. Thus, more functional analysis of the relationship between CCT and miR164A during other processes, such as embryogenesis or flowering, can help us gain better insight on the extent of the CDK8 module controlling other aspects of Arabidopsis development.

Our results from the fluorescent offer the possibility to look at the expression pattern *in vivo* and further analyses can be made crossing plants carrying the eGFP transgene with other CDK8 mutants and observe how their expression patterns change.

Perspectives

In this study I investigated the role of subunits of the CDK8 module in regulating miRNAs important for Arabidopsis development. Here I propose six future directions to address the functions of subunits of the CDK8 module during Arabidopsis growth and development.

1. To make full use of our annotation approach, we can analyze expression differences of sense and anti-sense transcription within transposable elements. This approach could be used to infer if the CDK8 module is also involved in the silencing of mobile elements.

2. We can predict new miRNA or sRNA-producing loci in Arabidopsis from our sRNA data using existing bioinformatics tools.

3. Our results demonstrate that the miR164 family is affected by the mutations in the different CDK8 subunits. Mutations in CDK8 subunits produces leaf margin and blade length phenotypes. However, the length of the blade is independent of miR164, suggesting different regulatory pathways. An interesting next step would be to investigate the extent of different subunits in regulating leaf elongation, as well as the adaxial/abaxial cell fates, vascular patterning on leaves and the number of abaxial trichomes. Since these are traits that change during development, it would expand our view of the CDK8 module as an upstream regulator of leaf development.

4. Is the CDK8 module regulating the miR164 family in an organ specific manner? Investigating the regulation of MIR164 by CDK8 and its involvement in other organs such as flowers or roots can shed light on how extensive is the regulation by the CDK8 in other developmental stages. The lines expressing eGFP under the miR164A promoter will be helpful to answer this question.

5. Plant hormones play an important role in growth. Understanding the interplay of the CDK8 module and factors such as auxin could help us gain insight into the role of the module on integrating endogenous signals to direct developmental processes.

6. Many of the missexpressed miRNAs are reported to have a role in stress responses. Investigating the role of the CDK8 module in the response to abiotic stresses such as temperature or nutrient availability would enhance our view on the CDK8 module as an integrator of external signals for plant defense.

References

Adai, Alex et al. 2005. "Computational Prediction of miRNAs in Arabidopsis Thaliana." : 78–91.

Aida, M et al. 1997. "Genes Involved in Organ Separation in Arabidopsis: An Analysis of the Cup-Shaped Cotyledon Mutant." *The Plant cell* 9(June): 841–57. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=156962&tool=pmcentrez&rendertype=abstract.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2002. *Molecular Biology of the Cell.* 4th Editio. New York: Garland Science.

Arteaga-Vázquez, Mario, Juan Caballero-Pérez, and Jean-Philippe Vielle-Calzada. 2006. "A Family of microRNAs Present in Plants and Animals." *The Plant cell* 18(12): 3355–69.

Aukerman, Milo J, and Hajime Sakai. 2003. "Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2-like Target Genes." *The Plant cell* 15(11): 2730–41.

Baker, Catherine C., Patrick Sieber, Frank Wellmer, and Elliot M. Meyerowitz. 2005. "The Early Extra petals1 Mutant Uncovers a Role for microRNA miR164c in Regulating Petal Number in Arabidopsis." *Current Biology* 15: 303–15.

Banyai, G., M. D. Lopez, Z. Szilagyi, and C. M. Gustafsson. 2014. "Mediator Can Regulate Mitotic Entry and Direct Periodic Transcription in Fission Yeast." *Molecular and Cellular Biology* 34: 4008–18. http://mcb.asm.org/cgi/doi/10.1128/MCB.00819-14.

Barkoulas, Michalis, Carla Galinha, Stephen P. Grigg, and Miltos Tsiantis. 2007. "From Genes to Shape: Regulatory Interactions in Leaf Development." *Current Opinion in Plant Biology* 10: 660–66.

Benfey, P N, and D Weigel. 2001. "Transcriptional Networks Controlling Plant Development." *Plant physiology* 125: 109–11.

Bilsborough, Gemma D et al. 2011. "Model for the Regulation of Arabidopsis Thaliana Leaf Margin Development." *Proceedings of the National Academy of Sciences of the United States of America* 108(8): 3424–29. http://www.pnas.org/content/108/8/3424.short (January 21, 2014).

Björklund, Stefan, and Claes M Gustafsson. 2005. "The Yeast Mediator Complex and Its Regulation." *Trends in biochemical sciences* 30(5): 240–44. http://www.ncbi.nlm.nih.gov/pubmed/15896741 (May 29, 2014).

Bolouri, Hamid, and Eric H Davidson. 2003. "Transcriptional Regulatory Cascades in Development: Initial Rates, Not Steady State, Determine Network Kinetics." *Proceedings of the National Academy of Sciences of the United States of America* 100(16): 9371–76.

Bowman, J L, D R Smyth, and E M Meyerowitz. 1991. "Genetic Interactions among Floral Homeotic Genes of Arabidopsis." *Development (Cambridge, England)* 112: 1–20.

Capron, Arnaud, Steven Chatfield, Nicholas Provart, and Thomas Berleth. 2009. "Embryogenesis: Pattern Formation from a Single Cell." *The Arabidopsis book / American Society of Plant Biologists* 7: e0126. http://marcin_filipecki.users.sggw.pl/PDFY/embryogenesis_tab.pdf.

Carrera, Inés et al. 2008. "Pygopus Activates Wingless Target Gene Transcription through the Mediator Complex Subunits Med12 and Med13." *Proceedings of the National Academy of Sciences of the United States of America* 105(18): 6644–49.

Castellano, Leandro, and Justin Stebbing. 2013. "Deep Sequencing of Small RNAs Identifies Canonical and Non-Canonical miRNA and Endogenous siRNAs in Mammalian Somatic Tissues." *Nucleic Acids Research* 41(5): 3339–51.

Chaturvedi, Chandra-Prakash et al. 2012. "Maintenance of Gene Silencing by the Coordinate Action of the H3K9 Methyltransferase G9a/KMT1C and the H3K4 Demethylase Jarid1a/KDM5A." *Proceedings of the National Academy of Sciences of the United States of America* 109(46): 18845–50. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3503177&tool=pmcentrez&rendertype=abstract (January 22, 2014).

Chen, Fangfang et al. 2010. "Expression Analysis of miRNAs and Highly-Expressed Small RNAs in Two Rice Subspecies and Their Reciprocal Hybrids." *Journal of integrative plant biology* 52(11): 971–80. http://www.ncbi.nlm.nih.gov/pubmed/20977655 (January 23, 2014).

Chen, Xuemei. 2005. "microRNA Biogenesis and Function in Plants." *FEBS Letters* 579(26): 5923–31.

———. 2009. "Small RNAs and Their Roles in Plant Development." *Annual review of cell and developmental biology* 25: 21–44. http://www.ncbi.nlm.nih.gov/pubmed/19575669 (October 21, 2013).

Conaway, Ronald C, and Joan Weliky Conaway. 2011. "Origins and Activity of the Mediator Complex." *Seminars in cell & developmental biology* 22(7): 729–34. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3207015&tool=pmcentrez&rendertype=abstract (September 11, 2013).

———. 2013. "The Mediator Complex and Transcription Elongation." *Biochimica et biophysica acta* 1829(1): 69–75.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3693936&tool=pmcentrez&rendertype=abstract (July 2, 2014).

Dai, Xinbin, and Patrick Xuechun Zhao. 2011. "PsRNATarget: A Plant Small RNA Target Analysis Server." *Nucleic Acids Research* 39(SUPPL. 2): 1–5.

Davis, Matthew P a et al. 2013. "Kraken: A Set of Tools for Quality Control and Analysis of High-Throughput Sequence Data." *Methods (San Diego, Calif.)* 63(1): 41–49. http://dx.doi.org/10.1016/j.ymeth.2013.06.027 (February 2, 2014).

Donner, Aaron Joseph, Stephanie Szostek, Jennifer Michelle Hoover, and Joaquin Maximiliano Espinosa. 2007. "CDK8 Is a Stimulus-Specific Positive Coregulator of p53 Target Genes." *Molecular Cell* 27(1): 121–33.

Ebert, Margaret S., and Phillip a. Sharp. 2012. "Roles for MicroRNAs in Conferring Robustness to Biological Processes." *Cell* 149(3): 505–24. http://dx.doi.org/10.1016/j.cell.2012.04.005.

Elmlund, Hans et al. 2006. "The Cyclin-Dependent Kinase 8 Module Sterically Blocks Mediator Interactions with RNA Polymerase II." *Proceedings of the National Academy of Sciences of the United States of America* 103(43): 15788–93. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1635081&tool=pmcentrez&rendertype=abstract.

Fire, A et al. 1998. "Potent and Specific Genetic Interference by Double-Stranded RNA in Caenorhabditis Elegans." *Nature* 391(6669): 806–11.

Flanagan, P M et al. 1991. "A Mediator Required for Activation of RNA Polymerase II Transcription in Vitro." *Nature*.

Gentleman, Robert C et al. 2004. "Bioconductor: Open Software Development for Computational Biology and Bioinformatics." *Genome biology* 5(10): R80. http://www.ncbi.nlm.nih.gov/pubmed/15461798.

Geuten, Koen, and Heleen Coenen. 2013. "Heterochronic Genes in Plant Evolution and Development." *Frontiers in plant science* 4(September): 381. http://www.scopus.com/inward/record.url?eid=2-s2.0-84900840917&partnerID=tZOtx3y1.

Gilbert, Scott F. 2010. *Developmental Biology*. 9th Ed. ed. Mass Sunderland. Sinauer Associates.

Gillmor, C Stewart et al. 2010. "The MED12-MED13 Module of Mediator Regulates the Timing of Embryo Patterning in Arabidopsis." *Development (Cambridge, England)* 137(1): 113–22. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2796935&tool=pmcentrez&rendertype=abstract (August 7, 2013).

———. 2014. "The Arabidopsis Mediator CDK8 Module Genes CCT (MED12) and GCT (MED13) Are Global Regulators of Developmental Phase Transitions." *Development (Cambridge, England)* 141(November): 4580–89. http://www.ncbi.nlm.nih.gov/pubmed/25377553.

Glazov, Evgeny a et al. 2009. "Repertoire of Bovine miRNA and miRNA-like Small Regulatory RNAs Expressed upon Viral Infection." *PloS one* 4(7): e6349. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2713767&tool=pmcentrez&rendertype=abstract (March 3, 2014).

Von Goethe, Johann Wolfgang. 2006. 6 KronoScope *The Metamorphosis of the Plants*.

Goldberg, R B, G de Paiva, and R Yadegari. 1994. "Plant Embryogenesis: Zygote to Seed." *Science (New York, N.Y.)* 266(ii): 605–14.

Guo, Hui-shan, Qi Xie, Ji-feng Fei, and Nam-hai Chua. 2005. "MicroRNA Directs mRNA Cleavage of the Transcription Factor NAC1 to Downregulate Auxin Signals for Arabidopsis Lateral Root Development." 17(May): 1376–86.

Haag, Jeremy R, and Craig S Pikaard. 2011. "Multisubunit RNA Polymerases IV and V: Purveyors of Non-Coding RNA for Plant Gene Silencing." *Nature reviews. Molecular cell biology* 12(8): 483–92. http://www.ncbi.nlm.nih.gov/pubmed/21779025 (January 27, 2014).

Hamilton, Andrew J, and David C Baulcombe. 1999. "A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants." *Science (New York, N.Y.)* 286(5441): 950–52.

Harvey Lodish, Arnold Berk, Paul Matsudaira, Chris A. Kaiser, Monty Krieger. 2003. *Molecular Cell Biology*. 5th ed. Freeman, W. H. & Company.

Hasson, Alice et al. 2011. "Evolution and Diverse Roles of the CUP-SHAPED COTYLEDON Genes in Arabidopsis Leaf Development." *The Plant cell* 23(1): 54–68. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3051246&tool=pmcentrez&rendertype=abstract (January 21, 2014).

Hentges, Kathryn E. 2011. "Mediator Complex Proteins Are Required for Diverse Developmental Processes." *Seminars in cell & developmental biology* 22(7): 769–75. http://www.ncbi.nlm.nih.gov/pubmed/21854862 (February 1, 2014).

Hepler, Peter K. 2005. "Calcium: A Central Regulator of Plant Growth and Development." *The Plant cell* 17(8): 2142–55.

Ten Hove, C. a., K.-J. Lu, and D. Weijers. 2015. "Building a Plant: Cell Fate Specification in the Early Arabidopsis Embryo." *Development* 142: 420–30. http://dev.biologists.org/cgi/doi/10.1242/dev.111500.

Huang, Daiqing et al. 2013. "MicroRNAs and Their Putative Targets in Brassica Napus Seed Maturation." *BMC genomics* 14(1): 140. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3602245&tool=pmcentrez&rendertype=abstract (February 26, 2014).

Illumina. 2011. "Quality Scores for Next-Generation Sequencing." http://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf.

Imura, Yuri et al. 2012. "CRYPTIC PRECOCIOUS/MED12 Is a Novel Flowering Regulator with Multiple Target Steps in Arabidopsis." *Plant & cell physiology* 53(2): 287–303. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3278046&tool=pmcentrez&rendertype=abstract (February 17, 2014).

Ito, Jun, Takako Sono, Masao Tasaka, and Masahiko Furutani. 2011. "MACCHI-BOU 2 Is Required for Early Embryo Patterning and Cotyledon Organogenesis in Arabidopsis." *Plant & cell physiology* 52(3): 539–52. http://www.ncbi.nlm.nih.gov/pubmed/21257604 (February 17, 2014).

Izhaki, Anat, and John L Bowman. 2007. "KANADI and Class III HD-Zip Gene Families Regulate Embryo Patterning and Modulate Auxin Flow during Embryogenesis in Arabidopsis." *The Plant cell* 19(2): 495–508.

Jensen, Michael K et al. 2010. "The Arabidopsis Thaliana NAC Transcription Factor Family: Structure-Function Relationships and Determinants of ANAC019 Stress Signalling." *The Biochemical journal* 426(2): 183–96.

Jima, Dereje D. et al. 2010. "Deep Sequencing of the Small RNA Transcriptome of Normal and Malignant Human B Cells Identifies Hundreds of Novel microRNAs." *Blood* 116.

Jones-Rhoades, Matthew W, David P Bartel, and Bonnie Bartel. 2006. "MicroRNAS and Their Regulatory Roles in Plants." *Annual review of plant biology* 57: 19–53. http://www.ncbi.nlm.nih.gov/pubmed/16669754 (November 13, 2013).

Kidd, Brendan N et al. 2011. "Diverse Roles of the Mediator Complex in Plants." *Seminars in cell & developmental biology* 22(7): 741–48. http://www.ncbi.nlm.nih.gov/pubmed/21803167 (August 14, 2013).

Kim, Jin Hee et al. 2009. "Trifurcate Feed-Forward Regulation of Age-Dependent Cell Death Involving miR164 in Arabidopsis." *Science (New York, N.Y.)* 323(February): 1053–57.

Kim, Yun Ju et al. 2011. "The Role of Mediator in Small and Long Noncoding RNA Production in Arabidopsis Thaliana." *The EMBO journal* 30(5): 814–22. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3049218&tool=pmcentrez&rendertype=abstract (October 26, 2013).

Koyama, Tomotsugu et al. 2010. "TCP Transcription Factors Regulate the Activities of ASYMMETRIC LEAVES1 and miR164, as Well as the Auxin Response, during

Differentiation of Leaves in Arabidopsis." *The Plant cell* 22(11): 3574–88. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3015130&tool=pmcentrez&rendertype=abstract (January 27, 2014).

Kozomara, Ana, and Sam Griffiths-Jones. 2014. "MiRBase: Annotating High Confidence microRNAs Using Deep Sequencing Data." *Nucleic Acids Research* 42(November 2013): 68–73.

Kvam, V. M., P. Liu, and Y. Si. 2012. "A Comparison of Statistical Methods for Detecting Differentially Expressed Genes from RNA-Seq Data." *American Journal of Botany* 99(2): 248–56.

Lack, Andrew, and David Evans. 2001. *Plant Biology*. Garland Science.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L Salzberg. 2009. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome biology* 10(3): R25.

Larivière, Laurent, Martin Seizl, and Patrick Cramer. 2012. "A Structural Perspective on Mediator Function." *Current opinion in cell biology* 24(3): 305–13. http://www.ncbi.nlm.nih.gov/pubmed/22341791 (August 14, 2013).

Lashkari, D a et al. 1997. "Yeast Microarrays for Genome Wide Parallel Genetic and Gene Expression Analysis." *Proceedings of the National Academy of Sciences of the United States of America* 94(24): 13057–62.

Latchman, David S. 2005. *Gene Regulation - A Eukaryotic Perspective*. 5th ed. Taylor & Francis Group.

Laufs, Patrick, Alexis Peaucelle, Halima Morin, and Jan Traas. 2004. "MicroRNA Regulation of the CUC Genes Is Required for Boundary Size Control in Arabidopsis Meristems." *Development (Cambridge, England)* 131(17): 4311–22.

Lawrence, Michael et al. 2013. "Software for Computing and Annotating Genomic Ranges." *PLoS Computational Biology* 9(8): 1–10.

Lawrence, Michael, Robert Gentleman, and Vincent Carey. 2009. "Rtracklayer: An R Package for Interfacing with Genome Browsers." *Bioinformatics* 25(14): 1841–42.

Lee, Hanna et al. 2010. "Genetic Framework for Flowering-Time Regulation by Ambient Temperature-Responsive miRNAs in Arabidopsis." *Nucleic acids research* 38(9): 3081–93. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2875011&tool=pmcentrez&rendertype=abstract (February 17, 2014).

Lee, Rosalind C, R. L. Feinbaum, and V. Ambros. 1993. "The C. Elegans Heterochronic Gene Lin-4 Encodes Small RNAs with Antisense Complementarity to Lin-14." *Cell* 75(5): 843–54.

Levine, Michael, and Eric H Davidson. 2005. "Gene Regulatory Networks for Development." *Pnas* 102(14): 4936–42. http://www.pnas.org/content/102/14/4936.long.

Lister, Ryan et al. 2008. "Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis." *Cell* 133(3): 523–36.

Maere, Steven, Karel Heymans, and Martin Kuiper. 2005. "BiNGO: A Cytoscape Plugin to Assess Overrepresentation of Gene Ontology Categories in Biological Networks." *Bioinformatics* 21(16): 3448–49.

Malik, Sohail, and Robert G Roeder. 2010. "The Metazoan Mediator Co-Activator Complex as an Integrative Hub for Transcriptional Regulation." *Nature reviews. Genetics* 11(11): 761–72. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3217725&tool=pmcentrez&rendertype=abstract (October 20, 2013).

Mallory, Allison C, Brenda J Reinhart, et al. 2004. "MicroRNA Control of PHABULOSA in Leaf Development: Importance of Pairing to the microRNA 5' Region." *The EMBO journal* 23(16): 3356–64. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=514513&tool=pmcentrez&rendertype=abstract (February 2, 2014).

Mallory, Allison C, and Nicolas Bouché. 2008. "MicroRNA-Directed Regulation: To Cleave or Not to Cleave." *Trends in plant science* 13(7): 359–67. http://www.ncbi.nlm.nih.gov/pubmed/18501664 (October 26, 2013).

Mallory, Allison C, and Hervé Vaucheret. 2006. "Functions of microRNAs and Related Small RNAs in Plants." *Nature genetics* 38 Suppl(June): S31–36. http://www.ncbi.nlm.nih.gov/pubmed/16736022 (November 17, 2013).

Mallory, Allison C., Diana V. Dugas, David P. Bartel, and Bonnie Bartel. 2004. "MicroRNA Regulation of NAC-Domain Targets Is Required for Proper Formation and Separation of Adjacent Embryonic, Vegetative, and Floral Organs." *Current Biology* 14: 1035–46.

Mayer, Ulrike et al. 1991. "Mutations Affecting Body Organization in the Arabidopsis Embryo." *Nature* 353: 402–7.

McCarthy, Davis J., Yunshun Chen, and Gordon K. Smyth. 2012. "Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation." *Nucleic Acids Research* 40(10): 4288–97.

McCormick, Kevin P, Matthew R Willmann, and Blake C Meyers. 2011. "Experimental Design, Preprocessing, Normalization and Differential Expression Analysis of Small RNA Sequencing Experiments." *Silence* 2(1): 2. http://www.silencejournal.com/content/2/1/2.

Meneely, Philip. 2009. *Advanced Genetic Analysis: Genes, Genomes, and Networks in Eukaryotes*. OUP Oxford.

Nagalakshmi, Ugrappa et al. 2008. "The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing." *Science (New York, N.Y.)* 320(5881): 1344–49.

Nemet, Josipa, Branka Jelicic, Ivica Rubelj, and Mary Sopta. 2013. "The Two Faces of Cdk8, a Positive/negative Regulator of Transcription." *Biochimie*: 6–11. http://www.ncbi.nlm.nih.gov/pubmed/24139904 (October 23, 2013).

Nikovics, Krisztina et al. 2006. "The Balance between the MIR164A and CUC2 Genes Controls Leaf Margin Serration in Arabidopsis." *The Plant cell* 18(11): 2929–45. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1693934&tool=pmcentrez&rendertype=abstract (November 17, 2013).

Nookaew, Intawat et al. 2012. "A Comprehensive Comparison of RNA-Seq-Based Transcriptome Analysis from Reads to Differential Gene Expression and Cross-Comparison with Microarrays: A Case Study in Saccharomyces Cerevisiae." *Nucleic acids research* 40(20): 10084–97. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3488244&tool=pmcentrez&rendertype=abstract (January 20, 2014).

Ochando, Isabel et al. 2006. "Mutations in the microRNA Complementarity Site of the INCURVATA4 Gene Perturb Meristem Function and Adaxialize Lateral Organs in Arabidopsis." *Plant physiology* 141(June): 607–19.

Olsen, Addie Nina, Heidi a Ernst, Leila Lo Leggio, and Karen Skriver. 2005. "NAC Transcription Factors: Structurally Distinct, Functionally Diverse." *Trends in plant science* 10(2): 79–87. http://www.ncbi.nlm.nih.gov/pubmed/15708345 (January 20, 2014).

Pages H, Aboyoun P, Gentleman R and DebRoy S. "Biostrings: String Objects Representing Biological Sequences, and Matching Algorithms."

Park, Soomin, and John J Harada. 2008. "Arabidopsis Embryogenesis." *Methods in molecular biology (Clifton, N.J.)* 427: 3–16.

Park, Wonkeun et al. 2002. "CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in Arabidopsis Thaliana." *Current Biology* 12(17): 1484–95.

Piriyapongsa, Jittima, and I King Jordan. 2008. "Dual Coding of siRNAs and miRNAs by Plant Transposable Elements." : 814–21.

Poethig, R Scott. 2010. "The Past, Present, and Future of Vegetative Phase Change." *Plant physiology* 154(2): 541–44. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2949024&tool=pmcentrez&rendertype=abstract (January 27, 2014).

R Development Core Team. 2008. "R: A Language and Environment for Statistical Computing." http://www.r-project.org.

Rapaport, Franck et al. 2013. "Comprehensive Evaluation of Differential Gene Expression Analysis Methods for RNA-Seq Data." *Genome biology* 14(9): R95. http://www.ncbi.nlm.nih.gov/pubmed/24020486 (January 22, 2014).

Rau, Marlene J., Sabine Fischer, and Carl J. Neumann. 2006. "Zebrafish Trap230/Med12 Is Required as a Coactivator for Sox9-Dependent Neural Crest, Cartilage and Ear Development." *Developmental Biology* 296(1): 83–93.

Reinhart, Brenda J. et al. 2002. "MicroRNAs in Plants." *Genes and Development* 16(13): 1616–26.

Riechmann, José Luis. 2002. "Transcriptional Regulation: A Genomic Overview." *The Arabidopsis Book* 16(1): 1.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2009. "edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26(1): 139–40.

Rogers, K., and X. Chen. 2012. "MicroRNA Biogenesis and Turnover in Plants." *Cold Spring Harbor Symposia on Quantitative Biology* 77: 183–94.

Rogers, Kestrel, and Xuemei Chen. 2013. "Biogenesis, Turnover, and Mode of Action of Plant microRNAs." *The Plant cell* 25(7): 2383–99. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3753372&tool=pmcentrez&rendertype=abstract (January 21, 2014).

Schena, M, D Shalon, R W Davis, and P O Brown. 1995. "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray." *Science (New York, N.Y.)* 270(5235): 467–70.

Schommer, Carla, Edgardo G Bresso, Silvana V Spinelli, and Javier F Palatnik. 2012. "Role of MicroRNA miR319 in Plant Development." *MicroRNAs in Plant Development and Stress Responses* 15: 29–48. http://www.springerlink.com/index/10.1007/978-3-642-27384-1.

Scott, Matthew P. 2000. "Development: The Natural History of Genes." *Cell* 100(7): 27–40.

Shannon, Paul et al. 2003. "Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks." (Karp 2001): 2498–2504.

Sharma, Vijay K, and Jennifer C Fletcher. 2002. "Maintenance of Shoot and Floral Meristem Cell Proliferation and Fate." *Plant physiology* 129(1): 31–39.

Sieber, Patrick et al. 2007. "Redundancy and Specialization among Plant microRNAs: Role of the MIR164 Family in Developmental Robustness." *Development (Cambridge, England)* 134(6): 1051–60. http://www.ncbi.nlm.nih.gov/pubmed/17287247 (January 29, 2014).

Singh, Karam B. 1998. "Update on Gene Regulation Transcriptional Regulation in Plants : The Importance of Combinatorial Control." : 1111–20.

Soneson, Charlotte, and Mauro Delorenzi. 2013. "A Comparison of Methods for Differential Expression Analysis of RNA-Seq Data." *BMC bioinformatics* 14(1): 91. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3608160&tool=pmcentrez&rendertype=abstract (January 20, 2014).

Sunkar, Ramanjulu. 2012. *MicroRNAs in Plant Development and Stress Responses*.

Tang, Xurong et al. 2012. "MicroRNA-Mediated Repression of the Seed Maturation Program during Vegetative Development in Arabidopsis." *PLoS Genetics* 8(11): 20–22.

Tóth-Petróczy, Agnes et al. 2008. "Malleable Machines in Transcription Regulation: The Mediator Complex." *PLoS computational biology* 4(12): e1000243. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2588115&tool=pmcentrez&rendertype=abstract (October 20, 2013).

Tsai, Kuang-Lei et al. 2013. "A Conserved Mediator-CDK8 Kinase Module Association Regulates Mediator-RNA Polymerase II Interaction." *Nature structural & molecular biology* 20(5): 611–19. http://www.ncbi.nlm.nih.gov/pubmed/23563140 (October 20, 2013).

Tsukaya, Hirokazu. 2002. "Leaf Development." *Arabidopsis Book* 1(13): e0072. http://dx.doi.org/10.1199/tab.0072.

Vidal, Elena a et al. 2013. "Integrated RNA-Seq and sRNA-Seq Analysis Identifies Novel Nitrate-Responsive Genes in Arabidopsis Thaliana Roots." *BMC genomics* 14: 701. http://www.ncbi.nlm.nih.gov/pubmed/24119003 (January 24, 2014).

Wang, Wenming, and Xuemei Chen. 2004. "HUA ENHANCER3 Reveals a Role for a Cyclin-Dependent Protein Kinase in the Specification of Floral Organ Identity in Arabidopsis." *Development (Cambridge, England)* 131(13): 3147–56. http://www.ncbi.nlm.nih.gov/pubmed/15175247 (October 25, 2013).

Wightman, B., I. Ha, and G. Ruvkun. 1993. "Posttranscriptional Regulation of the Heterochronic Gene Lin-14 by Lin-4 Mediates Temporal Pattern Formation in C. Elegans." *Cell* 75(5): 855–62.

Van Wolfswinkel, Josien C, and René F Ketting. 2010. "The Role of Small Non-Coding RNAs in Genome Stability and Chromatin Organization." *Journal of cell science* 123: 1825–39.

Wu, Gang et al. 2009. "The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis." *Cell* 138(4): 750–59.

Zhan, Shuhua, and Lewis Lukens. 2010. "Identification of Novel miRNAs and miRNA Dependent Developmental Shifts of Gene Expression in Arabidopsis Thaliana." *PLoS ONE* 5(4).

Zhang, Zhenhai et al. 2009. "PMRD: Plant microRNA Database." *Nucleic Acids Research* 38(SUPPL.1): 806–13.

Zhou, Liang et al. 2010. "Integrated Profiling of microRNAs and mRNAs: microRNAs Located on Xq27.3 Associate with Clear Cell Renal Cell Carcinoma." *PloS one* 5(12): e15224. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3013074&tool=pmcentrez&rendertype=abstract (January 31, 2014).

Appendices