UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Scienze Biologiche e Molecolari

XXVII Ciclo

Searching For New Genetic Pathways In Early Flower Development Of *Arabidopsis thaliana*

O. Mantegazza

PhD Thesis

Scientific Tutor: Prof. Martin Kater

Academic Year: 2013-2014

SSD: BIO/11; BIO/18

This thesis was performed at the Department of Biosciences; Università degli Studi di Milano

Contents

Abstract

During the production of flowers in *Arabidopsis thaliana* many key decisions are taken in a short lapse of time. The floral primordium has to be positioned correctly on the inflorescence meristem and it has to grow to the required dimension before flower organs are themselves positioned and differentiate.

All these tasks are strictly controlled at a molecular level and the genetic networks that underlies them have been intensively studied in the last 30 years. Nevertheless we are far from having a comprehensive knowledge on this process and the genetic mechanism controlling the arise, identity of the floral primordium and the timing of its developmental phases are widely unknown.

We have identified new genes potentially involved in early flower development with two approaches: (i) Analysis of the specific transcriptome of the earliest stages of flower development and (ii) Co-expression analysis using *APETALA1* and *LEAFY*, two genes that determine the identity of the floral meristem, which is the earliest stage of flower development. We have observed that multiple *REM* transcription factors are co-expressed with *APETALA1* and *LEAFY*.

Characterizing insertional mutants for genes potentially involved in early flower development and REM transcription factors, we have rarely observed a phenotype in the stages under study. This is consistent with the hypothesis that genes controlling early flower development are often functionally redundant. We are implementing various methods to overcome functional redundancy implementing analysis of gene families, multiple RNA interference and gene targeting strategies.

Part I

Introduction And Presentation Of My Work

Chapter 1

State Of The Art

1.1 Early flower development of *Arabidopsis thaliana*

Most of my thesis revolves around the genetic study of early flower development in *Arabidopsis thaliana*.

1.1.1 Morphology of early flower development.

In *Arabidopsis thaliana* flowers are generated at the apex of the inflorescence (inflorescence meristem - IM) in a spiral shaped series. Each flower is separated from the other by an angle of 137.5°(Guédon et al., 2013). The process of early flower development in *Arabidopsis* has been described in detail by Smyth, Bowman and Meyerowitz in 1990 using scanning electron microscopy (SEM) imaging (see figure 1.1). The IM is a meristematic dome-shaped structure of about 45 µm diameter located at the apex of the inflorescence. The first visible stage of floral development is the floral meristem (FM) which comprehends stage 1 and stage 2 of flower development. The FM is, again, a dome shaped meristematic structure, that buds from the flanks of the IM and grows in volume for approximately two days until it reaches about 35 µm diameter. Then the third stage of flower development (ST3) begins and organs differentiation starts. The transition from stage 1 to stage 2 of

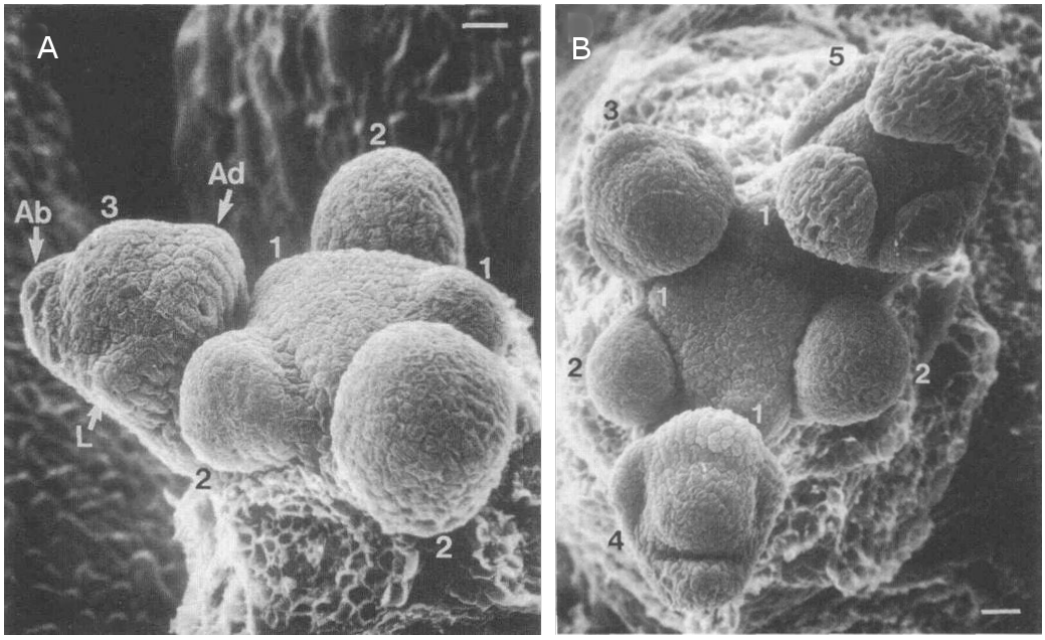


Figure 1.1: Lateral and top (respectively A and B) view of the inflorescence apex and younger buds after (older buds have been removed to make visible thee younger buds). The numbers indicate the developmental stage of the nearby flower. The position of the developing sepals in the stage 3 flower are indicated by: Ab=Abaxial, Ad=Adaxial L=Lateral (image from Smyth et al., 1990). The scale bar is 35 µm.

flower development (hereby both referred as floral meristems) is landmarked by the appearance of the stalk. In this work we will concentrate on the aforementioned structures.

1.1.2 Interest in early flower development

The mersitematic stages of flower development and the switch to organ differentiation are a prelude and bottleneck to the development of flower, fruit and seeds. As pointed out by J. L. Bowman in 2012, flowers, with their distinctive colors, shapes and perfumes, have attracted interests of scientists since the birth of genetics and even before (Bowman et al., 2012). On the other side, the agronomical interest in the control of fruit and seed development is self-evident.

In the last three decades, early flower development has been used as model system for many studies in molecular genetics. For example one of the first models to describe combinatorial functional and molecular interaction among homeotic transcription factors has been formulated on flowers and has been named ABC model of flower development. This model states that functional and physical interaction among three classes of proteins (A, B and C) specify the identity of the four flower organs (sepals, petals, stamens and carpels) (Coen and Meyerowitz, 1991; Melzer and Theißen, 2009). Moreover the meristematic stages (IM and FM) contain a stem cell pool that gets progressively lost from ST3 onward when the development of floral organs start. This stem cell pool has been used to describe the balance between stem cell maintenance and organ differentiation in plants. It has been shown that a molecular negative feedback loop determine the maintenance of the right amount of stem cells in the IM and FM (Fletcher et al., 1999; Brand et al., 2000). Years later, early flower development has been used to model how peaks in the concentration of the plant hormone auxin determine the position of lateral organs through mathematical modeling (Guédon et al., 2013). All these studies show that early flower development is an established model system for the scientific research in molecular genetics.

1.1.3 Genetics of early flower development

The genetic of early flower development in *Arabidopsis* has been subject of intense studies during the last 25 years (reviewed by (Ó'Maoiléidigh et al., 2014).

Determination of floral meristem identity

The appearance of the FM is marked by a sharp increase in the expression levels of the floral meristem identity (FMI) genes. Mutants lacking the function of the FMI genes produce meristem-like, leaf-like and inflorescence-like structures instead of flowers (Figure 1.2). The FMI genes are downstream to the floral integrator genes, such as *SUPPRESSOR OF OVEREXPRESSION OF COSTANS 1* (*SOC1*). The floral integrators collect the information from

the flowering time pathways which detect environmental and internal cues in order to establish the correct timing for the switch from vegetative to reproductive development.

The First FMI gene to be activated, in a group of cells morphologically indistinguishable from the IM, is *LEAFY* (*LFY*) (Weigel et al., 1992), which is activated by SOC1 itself (Lee et al., 2008). LFY is a transcription factor with two DNA binding domain and it is present as a single copy gene with no homologues in the *Arabidopsis* genome; LFY is necessary and sufficient to initiate flower development.

The activation of *LEAFY* is followed by the activation of the other flower meristem identity genes *LATE MERISTEM IDENTITY1 (LMI1) APETALA1 (AP1), CAULIFLOWER (CAL), AGAMOUS-LIKE 24 (AGL24)* and *SHORT VEGETATIVE PHASE (SVP)* (Bowman et al., 1993; Ferrándiz et al., 2000; Gregis et al., 2006; Saddic et al., 2006; Gregis et al., 2008).

Figure 1.2: Inflorescence of floral meristem identity loss of function mutants. A. wt Col-0 plant. B. Inflorescence of a *lfy-2* mutant, flower are converted in vegetative structures. C. Inflorescence of an *ap1 cal* double mutant; the lumpy cauliflower-like structures is composed by over-proliferating meristems that do not undergo differentiation. (Images A and C for courtesy of ABRC (https://abrc.osu.edu/).

The inter-regulatory pathways among the FMI form a tangled, highly redundant and complex network (Grandi et al., 2012). One of the most widely known phenotypes of the loss of floral meristem identiy is the one of the double mutant *ap1 cal* and of its phenocopy, the triple mutant *ap1 svp agl24*. In these mutant floral organs do not differentiate and the system is temporarily

blocked in a massive over-proliferation of meristematic tissues (see Figure 1.2) (**?**). *AP1, SVP, AGL24* and *CAL* are all MADS-box transcription factors (Parenicová et al., 2003).

The FMI transcription factors AP1, SVP and AGL24 repress the class B and C floral organ identity genes *PISTILLATA (PI), APETALA3 (AP3), AGAMOUS (AG)* preventing early differentiation of floral organs. Toward the end of stage 2 of flower development the expression of FMI genes *SVP, AGL24* and *CAL* decreases and the differentiation of floral organs begins.

Maintenance of stem cell pool

The complex molecular mechanisms allowing both the maintenance of a stem cell pool and the production of new tissues is regulated by a negative feedback loop. WUSCHEL (WUS), a homeodomain transcription factor required for the determination of stem cell identity, acts as a positive signal for stem cell proliferation (Laux et al., 1996; Mayer et al., 1998; Busch et al., 2010); while CLAVATA3 (CLV3) is a small secreted glycopeptide which acts together with its receptors CLAVATA1 (CLV1) and CLAVATA2 (CLV2) to negatively regulate stem cell proliferation (Clark et al., 1996; Fletcher et al., 1999; Guo and Clark, 2010). WUS and CLV3 regulate each other expression in a negative feedback loop that prevents both the collapse and overgrowth of stem cell niches in meristematic tissues (Brand et al., 2000). From floral stage 3, WUS activates the MADS-box floral homeotic class C gene *AG*, which, in turn, represses *WUS* stopping the indeterminate proliferation of the FM and promoting the development of the inner floral whorls (Lohmann et al., 2001).

1.1.4 Functional Information On Early Flower Development

As stated by the central dogma of molecular biology, the flow of genetic information within a living organism starts form DNA (Crick et al., 1970). Considering that we are trying to characterize the genetic/molecular pathways that underlie flower development in *Arabidopsis thaliana* and that the genetic information fed to these pathways comes from the genome, a logic question

is: how much do we know about the information encoded in the the genome of *Arabidopsis*? A good estimator of this is the fraction of the genes that have been functionally characterized with loss of function mutants which, for *Arabidopsis*, consists of about one tenth of the genes encoded in the genome (Lloyd and Meinke, 2012).

Many of the uncharacterized genes have been reported to be target of the floral meristem identity transcription factors LFY, AP1 and SVP with the experiment performed by Kaufmann et al. in 2010, Winter et al. in 2011 and Gregis et al. in 2013 and are thus very likely to be involved in early flower development. These genes might encode information on widely unknown processes such as the timing of the progression through developmental phases of the FM. We can estimate that, even if early flower development has been intensively studied and characterized, our knowledge of the genetic pathways that underlie it is limited.

1.1.5 The *ap1 cal AP1-GR* induction system

Many of the molecular characterization of floral meristem and early flower development have been carried out implementing the *ap1 cal AP1-GR* system. The inflorescence meristem and the early stages of flower developments of *Arabidopsis thaliana* have sizes that range from tens to hundreds of micrometers and therefore cannot be easily collected in the amount needed for their molecular characterization. The *ap1 cal AP1-GR* mutant plants are used to solve this problem. In the *ap1 cal* double mutant flower development is (temporarily) blocked in a massive over-proliferation of inflorescence-like meristems, leading to a cauliflower curd-like appearance (Figure 1.2 and 1.3A). The *ap1 cal AP1-GR* inducible system is based on the activation of the AP1-glucocorticoid receptor fusion protein (AP1-GR) by the synthetic steroid dexamethasone (DEX) in the *ap1 cal* double mutant. In the absence of DEX the GR domain excludes the AP1 transcription factor from the nucleus where it has to be to perform its physiological function. Induction of AP1 activity with DEX simultaneously turns all the inflorescence-like meristems of the *ap1 cal AP1-GR* plants into floral meristems and later to flowers (Figure 1.3B).

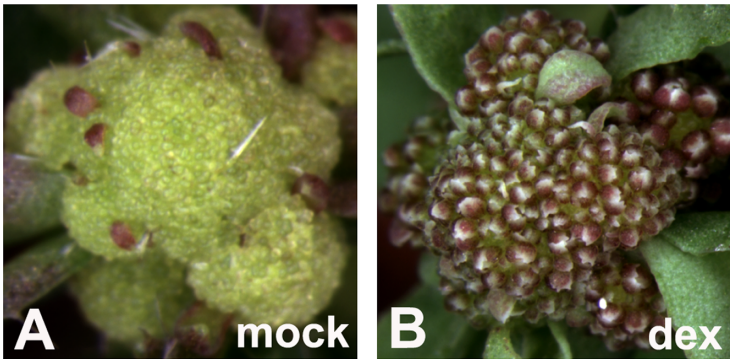


Figure 1.3: ***ap1 cal AP1-GR* system for synchronous flower induction.** A cauliflower curd like structure at the apex of an *ap1 cal* double mutant inflorescence, as shown in Figure 1.2 after treatment with (A.) a biologically inactive control solution (mock) and (B.) with dexamethasone. Treatment with dexhametasone causes the release of AP1 into the nucleus and triggers simultaneous development of multiple flowers (picture from Wellmer et al., 2006).

These flowers are numerous and synchronized and they can be easily collected and used to analyze the molecular pathways that underlie early flower development (Wellmer et al., 2006). Many molecular analysis of early flower development, such as the transcriptional profiling of these stages and the genome wide analysis of AP1 binding sites (Wellmer et al., 2006; Kaufmann et al., 2010; Pajoro et al., 2014), have been carried out using this system.

1.1.6 Attempts of comprehensive characterization of early flower developments

The reductionist approach used in molecular genetics, although consistent with the scientific method, have been partly criticized as insufficient to describe a complex systems, such as a living organism (Sauer et al., 2007). Studying molecular pathways underlying early flower development one gene/one pathway at a time is indeed a reductionist approach. A comprehensive approach instead addresses a biological system as a whole and relies on quantitative measurements of multiple components simultaneously,

these measurements can be supplied today by powerful technologies such as next generation sequencing.

Many attempts have been made to provide comprehensive quantitative data and to build mathematical models concerning the molecular pathways that underlie early flower development. Microarray data for loss of function mutants of the FMI gene *lfy* is publicly available at *AtgeneExpress* (Schmid et al., 2005). Numerous co-expressed groups of genes have been detected analyzing the transcriptome of the early developing flower using the *ap1 cal AP1-GR* system (Wellmer et al., 2006). Chromatin immunoprecipitation coupled to high throughput sequencing (ChIP-seq) (Johnson et al., 2007) has been used to detect the direct target and thus the gene regulatory network downstream of many of the transcription factors involved in early flower development, highlighting that SEP3 and AP1 are strongly involved in transcriptional control of hormonal pathways genes (Kaufmann et al., 2009, 2010) while SVP is controlling meristem development pathways (Gregis et al., 2013).

Finally, a combined study of gene expression, transcription factors binding and chromatin accessibility highlighted that gene sets controlled by homeotic genes involved in flower development are extensively but not completely overlapped (Pajoro et al., 2014), thus the molecular network that control flower devlopment is complex and redundant. Further application of comprehensive approaches will probably be needed in order to describe consistently early flower development.

1.2 The *REM* gene family

Performing a meta-analysis of the *Arabidopsis* expression data collected in the NASCarray repository (Craigon et al., 2004), we have observed that the expression levels of multiple *REPRODUCTIVE MERISTEM* (*REM*) trascription factors and of the FMI genes *LFY* and *AP1* are correlated. Thus, we have hypothesized that *REM* transcription factors are involved in early flower development. Moreover *REM* transcription factors are poorly functionally characterized; this makes them perfect candidates for our future

analysis. We are concentrating on *REM34* that is co-expressed with *LFY* and *AP1* and expressed during early flower development.

1.2.1 A general description of the *REM* gene family

The *REM* transcription factor family of *Arabidopsis thaliana* is composed by 45 genes (Romanel et al., 2009) all containing one or multiple copies of the B3 DNA binding domain (Swaminathan et al., 2008; Romanel et al., 2009). Moreover, *REM* genes are phylogenetically divergent and extensively duplicated and are sometimes located in clusters in the Arabidopsis genome (Swaminathan et al., 2008; Romanel et al., 2009). The largest cluster is located on chromosome 4 containing 9 *REM* genes (*REM34, REM35, REM36, REM37, REM38, REM39, REM41* and *REM42*) within 30 kilobases. that are closely phylogenetically related (Figure 1.4).

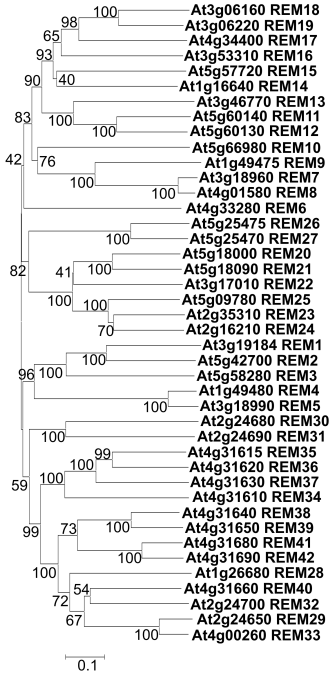


Figure 1.4: Phylogenetic tree of the whole *REM* gene family.

The B3 domain was first identified in maize (McCarty et al., 1991), is specific to plants (Swaminathan et al., 2008), and can bind DNA cooperatively in vitro (Suzuki et al., 1997). Five *Arabidopsis* transcription factor families contain the B3 DNA binding domain: the *LAV* (*LEC2* [*LEAFY COTYLEDON 2*]/*ABI3* [*ABSCISIC ACID INSENSITIVE 3*], *VAL* [*VP1/ABI3-LIKE*]), *RAV* (*RELATED to ABI3/VP1*), *ARF* (*AUXIN RESPONSE FACTOR*) and *REM* families. The DNA sequence recognized by the B3 domain is different in these five families (Swaminathan et al., 2008; Wang and Perry, 2013; Ulmasov et al., 1997;

Ezcurra et al., 2000; Kagaya et al., 1999; Romanel et al., 2009) and in the *REM* family it might bind DNA with no sequence specificity (King et al., 2013).

1.2.2 Function of *REM* transcription factors in vernalization and ovule development

Only two *REM* genes have been functionally characterized in *Arabidopsis*: *VERNALIZATION1/REM5 (VRN1)*, which is involved in the vernalization process (Levy et al., 2002; King et al., 2013) and *REM20/VDD*, which is essential for the development of the antipodal and synergid cells in the female gametophyte (Matias-Hernandez et al., 2010; Mendes et al., 2013). The expression pattern of the 45 *REM* genes suggests that they are involved in many other processes throughout the development of *Arabidopsis thaliana* but neither their function or mechanism of action have been reported yet.

1.3 New Technologies For Precise Genome Editing

The functional analysis of a genetic elements often relies on the observation of the phenotype caused mutation of the genetic element. For example, the function of the gene *LEAFY* has been first inferred from the phenotype of the *leafy* loss of function mutant in which flowers are converted in inflorescence-like and leaf-like structures (Weigel et al., 1992).

We are characterizing new genes involved in early flower development. New, precise and efficient ways of introducing modification in those genes can highly speed up the characterization process. New technologies of genomic modification that might introduce specific deletion in the genome are especially important for the functional characterization of the *REM* genes which are often concatenated on the genome, such as *REM34* with its closest homologues. This need can be addressed by the genome editing technologies (reviewed by Gaj et al., 2013). Since the genome editing technologies are relatively new, I

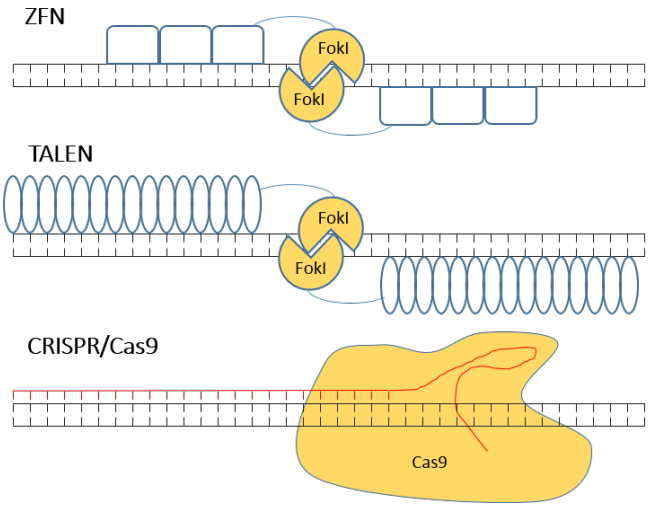


Figure 1.5: Schematic view oh the three main tools for Genome Editing ZFN, TALEN and CRISPR/Cas9, modified from AddGene. (https://www.addgene.org/)

will introduce them in detail hereby.

1.3.1 Genome editing tools

The rationale of Genome Editing is that repeatedly introducing a double strand break (DSB) at a specific locus will eventually cause its mutation. Nowaday we can introduce targeted DSB in vivo with engineered nucleases. The engineered nucleases are produced fusing a non-specific DNA cleavage domain, generally the FokI domain, with a customizable DNA binding domain (Kim et al., 1996). The DNA binding domains that have been used with greatest success are:

- **Zinc-Finger domains:** Zinc-Finger DNA binding domains are small and easy to transform in living organisms. One single Zinc-Finger domain individually binds 3 consecutive base pairs in the DNA and are generally used in arrays of 3 to 6 Zinc-Finger domains, which target 9 to 18 base pairs. Nevertheless engineering the Zinc Finger DNA binding domain can be challenging because the DNA specificity of one single Zinc-Finger domain is context-dependent (Ramirez et al.,

2008). The OPEN selection method (Maeder et al., 2009) has made this technology available to the broad scientific community providing guidelines to develop zinc-finger arrays with new DNA specificity testing a combinatorial array library of known Zinc-Finger domain. A Zinc-Finger array fused to the FokI DNA cleavage domain is called Zinc Finger Nuclease (ZFN) (Figure 1.5A).

- **Transcriptional Activator Like Effector (TALE):** TALE infectious bacterial transcription factors are a key component of the plant infection process by *Xanthomona* bacteria (Boch and Bonas, 2010). The DNA binding domain of TALE has specific features that makes it suitable for genome editing. These domains are composed of repetition of a highly conserved 33-34 amino acid motif; one single repetition/motif binds one single base pair (Boch et al., 2009) and the 12th and 13th amino acids residues (RVD) of each repetition determine the targeted nucleotide. Assembling the 4 different motives that target the four different base pairs in the correct sequence is sufficient to obtain a protein that binds the desired DNA sequence. The TALE DNA binding domain thus can be easily engineered to bind almost every DNA sequence.

  Nevertheless the repeated nature of the TALE DNA binding domains makes the cloning process difficult. Moreover a TALE domain is always considerably larger than a Zinc-Finger array targeting the same sequence. Fusion proteins of TALE and FokI DNA cleavage domain are called TALEN (Figure 1.5B).

Moreover, one of the latest tool for genome editing is the bacterial CRISPR/Cas9 DNA cleavage system (Brouns et al., 2008). This system has been engineered to specifically cleave DNA in mammals (Cong et al., 2013) and suitable to be used in virtually any organism. Since the specificity of this system is based on standard base paring between the target DNA and complementary CRISPR RNA (Brouns et al., 2008), CRIPSR/Cas9 can be easily designed to target whichever DNA sequence of interest (Figure 1.5B).

Genome editing is not a mature technology yet but has already been tested and implemented successfully in several plants such as *Arabidopsis*,

Tobacco (Townsend et al., 2009), rice (Miao et al., 2013) and wheat (Wang et al., 2014). Especially in Arabidopsis, CRISPR trasformed with floral dip method can introduce mutation in the germ line with a high but variable efficiency (Fauser et al., 2014; Feng et al., 2014).

The basal tools and concepts for genome editing can be further applied in order to introduce chromosomal deletions, inversions (Xiao et al., 2013), allele corrections (Ramirez et al., 2012) and even modifications in the epigenome (Mendenhall et al., 2013) of a living organism.

1.3.2 Gene targeting with engineered nucleases

As stated by Wikipedia: ''Gene targeting (also, replacement strategy based on homologous recombination) is a genetic technique that uses homologous recombination to change an endogenous gene''. Since the rationale of reverse genetics is studying the function of a genomic element through the effects of variations in its sequence, gene targeting is extremely useful tool in reverse genetics.

Genetic modification techniques in most plants relies mostly on random integration of the transformed DNA fragment in the genome and thus are not suitable for gene targeting. Nevertheless, gene targeting can be achieved in plants with an extremely low efficiency, ranging from 10^{-5} to 10^{-4}, introducing homology regions in the transformed DNA fragment (Offringa et al., 1990). The efficiency of gene targeting in plants can be raised to 10^{-2} inducing a double strand break in the targeted site with I-SceI meganuclease in order to stimulate repair through homologous recombination (Puchta et al., 1993, 1996). Even if complicated, it is feasible to engineer meganucleases for targeting a desired DNA sequence other then their natural target (Rosen et al., 2006) and they were used by Bayer CropScience to induce gene targeting in Cotton. Moreover, gene targeting was achieved in Tobacco with ZFN, introducing known mutation that confer herbicide resistance in the genes *SurA* and *SurB* with efficiency higher than 2% (Townsend et al., 2009).

Customizable nucleases such as ZFN, TALEN and the CRISPR/Cas9 system are likely to become the choice for gene targeting in plant and indeeed

CRISPR/Cas9 have been used to induce repairing of a partially duplicated reporter gene with homologous recombination with high efficiency (Fauser et al., 2014; Feng et al., 2014).

Gene targeting in plants have been reviewed by Puchta and Fauser in 2013.

Chapter 2

Aim Of The Work

Many genes that could be involved in the development of the FM and in early flower development may be still uncharacterized. This is preventing the formulation of functional model capable of describing how the plant uses the information contained in its genome to determine the molecular networks needed to build and develop correctly a new flower. We aimed at identifying new genes and genetic pathways involved in this process. Our work addresses the questions of how can we identify and characterize those genes.

2.1 Tasks

In order to identify new genes or molecular pathways involved early flower development, we have undertaken two approaches:

- **Definition of the precise transcriptome of the IM and of the first stages of flower development**: In order to find new genes involved in the development of the FM, we defined and analyzed the transcriptome of wild-type floral meristem and compared it with the transcriptome of the inflorescence meristem and of the differentiating flower (flowers at developmental stage 3) .

  We have dissected these developmental stages with micrometric precision at a laser microdissector, analyzed their transcriptome and used these data to define sets of genes that are differentially expressed among

these stages. The information contained in those genes is hopefully contributing to the developmental progression from IM to FM and later to the differentiating flower. This strategy has allowed us to select for candidate genes that we are characterizing with loss of function mutants right now. Moreover we have used the transcriptomic datasets in order to make hypothesis on cases of functional redundancy and regulatory pathways active in the stages under study.

- **Identification and characterization genes co-expressed with the two key FMI genes *LFY* and *AP1*:** We sought for genes co-expressed with *LFY* or *AP1* because two genes with a statistical correlation of mRNA expression levels may have similar function or be involved in the same biological process (Menges et al., 2007, 2008). Previously, in collaboration with Dr. Piero Morandini, we have observed that multiple *REM* transcription factors are co-expressed with the FMI genes *LFY* and *AP1*.

  The function of the *REM* genes is poorly understood and, often, their loss of function mutants have no peculiar phenotype (Franco-Zorrilla et al., 2002; Romanel et al., 2009, 2011). Interestingly, the expression pattern of *Arabidopsis REM* genes suggest a strong involvement in many developmental processes, especially in reproductive development.

  We have provided a detailed expression analysis and meta-analysis of existing data for the genes belonging to this family. We have identified the *REM34* cluster on chromosome 4 as a high profile candidate for involvement in early flower development and we have analyzed insertional mutants of genes belonging to this cluster. In this part of the work we started from statistical analysis of comprehensive transcriptomic datasets and we have progressively focused and reduced our approach selecting few genes (*REM34, REM35* and *REM36*) for further functional characterization.

Furthermore we are implementing genome editing and multiple RNA interference strategies and developing new ones in order to reveal the function of *REM* genes.

Chapter 3

Main Results

3.1 The Transcriptome of Early Flower Development

We used Laser microdissection coupled to RNA-seq to gain a snapshot of the transcriptome of the IM, FM and of the differentiating flower. This part of my work has been carried out in collaboration with Dr. David Horner and Dr. Matteo Chiara from the Beacon Bioinformatic Group at UNIMI. We have published this part of my work in Mantegazza et al. (2014a).

3.1.1 Setting of the microdissection of *Arabidopsis* inflorescences

In order to collect total RNA from the IM, FM and ST3 of wild type *Col-0* plants we dissected with micrometric precision the aforementioned stages at a Laser Microdissector (Figure 3.1). Setting up the condition for laser microdissection (LM) of *Arabidopsis* inflorescences has been a relevant part of the work of my first year of PhD.

In order to carry out LM, the tissue of interest must be fixed, embedded in paraplast, sectioned at the microtome and distended onto special slides before it can be microdissected (Day et al., 2005) To do so we adopted, with minor changes, a protocol from (Schmidt et al., 2012). The complete protocol

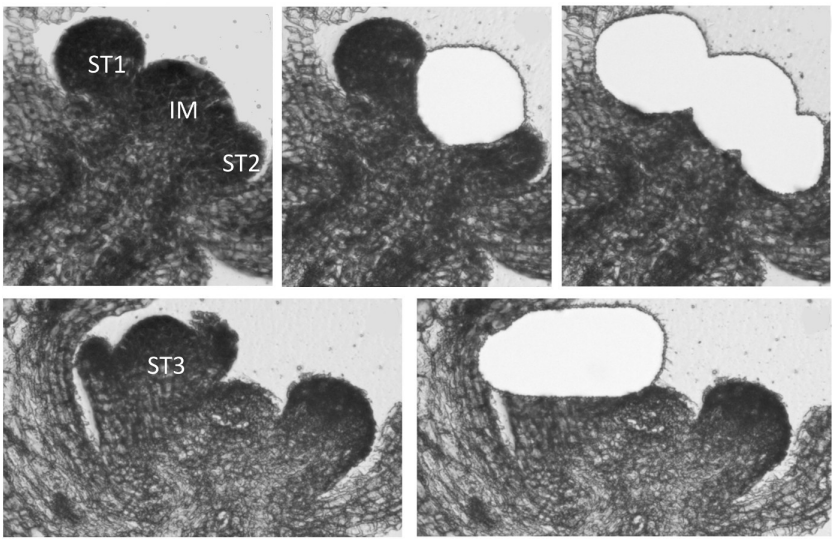


Figure 3.1: **Laser microdissection of *Arabidopsis* inflorescence** Inflorescence apex before and after microdissection of inflorescence meristem (IM), floral meristem (ST1 and ST2) and flower at developmental stage 3 (ST3)

is described in Mantegazza et al., (2014a). In our experience, the key steps essential for preservation of RNA integrity are: tissue fixation in EtOH:Acetic acid 9:1, embedding at 54 °C in special Paraplast with low melting temperature and to distending the tissue section on slides with methanol instead of water. This guaranteed a RNA Integrity Number (RIN, see Schroeder et al., 2006) above 6 for all the samples (see section 8.1). We found that pooled material from 15 inflorescences (Which is the limit for our pipeline of work) guaranteed over 10 ng of total RNA for each stage for each replicate. This is enough for subsequent RNA extraction, amplification, retrotranscription and sequencing with today's methods (Table 3.1).

3.1.2 Validation of our transcriptomic datasets

After turning reads into quantitative expression data with the Bowtie/Tophat/Cufflink pipeline (BTC, see Trapnell et al., 2010) , we have established a cutoff of 0.5 RPKM (the crossover of false negative and false discovery distribution, see section 8.2) for expressed genes. We thus detected 16,204 genes expressed in all the tissues, some at highly variable levels. 13,948 of

Sample	Dissected area (μm^2)	RNA conc. ($pg\,\mu L^{-1}$)	RIN	cDNA conc. ($ng\,\mu L^{-1}$)
IM - I	614,437	555	6.8	248
FM - I	830,466	384	6.6	242
ST3 - I	1,327,581	1167	7	264
IM - II	665,868	929	6.4	306
FM - II	914,178	1,639	7,1	320
ST3 - II	1,311,891	1966	6.9	300

Table 3.1: **Microdissected material** Dissected Area in μm^2, RNA concentration in $pg\,\mu L^{-1}$, RNA integrity number and cDNA concentration in $ng\,\mu L^{-1}$ for each stage in both replicates (I, II) used for sequencing.

these genes ($p - value < 10e - 20$) overlap with the transcriptome of the vegetative and transition shoot apical meristem (SAM), , characterized with a similar method by Torti et al. in 2012.

Then we have decided to test the consistency of our datasets and selected from scientific literature 13 marker genes that are known to:

1. be differentially expressed at least between two of the developmental stages under study (IM, FM and ST3),
2. have an enstablished biological function in one or more of these stages.
3. Have an expression pattern defined by publicly available in situ hybridization data in the stages under study.

These marker genes are:

***TERMINAL FLOWER 1* (*TFL1*)** TFL1 Determines the identity of the IM, where its mRNA is exclusively expressed (Bradley et al., 1997).

***AGAMOUS-LIKE 20* (*AGL20*)** AGL20 is a key factor that determines the IM identity and is also known as *SUPPRESSOR OF OVEREX-PRSSION OF COSTANS 1* (*SOC1*); Its mRNA is expressed exclusively in the IM (Samach et al., 2000).

***FLOWERING LOCUS* (*FD*)** FD promotes flowering acting in the SAM, its mRNA is expressed in the IM and not in later stages (Abe et al., 2005).

***AGAMOUS-LIKE 8* (*AGL8*)** AGL8 is also known as *FRUITFUL* (*FUL*) and is a regulator of IM development and of fruit development (Gu et al., 1998); its mRNA is expressed in the IM and not in the FM or ST3 (Mandel and Yanofsky, 1995).

***CLAVATA1* (*CLV1*)** CLV1 is a negative regulator of meristem size. Its mRNA is expressed in the IM and FM and in decreasing levels in the ST3 (Clark et al., 1996).

***SHORT VEGETATIVE PHASE* (*SVP*)** SVP is a key factor determining the identity of the FM (Gregis et al., 2008). It is expressed distinctively in the FM and not in ST3 (Hartmann et al., 2000).

***WUSCHEL* (*WUS*)** WUS is essential for the maintenance of the stem cell pool in IM and FM; its mRNA is expressed in the center of the IM, FM, and at lower levels in the ST3 where it get repressed by AGAMOUS (Mayer et al., 1998).

***LEAFY* (*LFY*)** LFY determines the identity of the FM and is one of the first FMI genes to be activated before the FM becomes morphologically distinguishable from the IM. *LFY* mRNA is expressed in the FM and in later stages of flower development (Weigel et al., 1992).

***PISTILLATA* (*PI*)** PI is a class B floral homeotic protein, it determines the identity of petals and stamens and it is expressed starting from ST3, when the identity of floral organs is determined (Goto and Meyerowitz, 1994).

***APETALA1* (*AP1*)** AP1 is both a FMI and a class A floral homeotic protein. It is expressed from FM on (Mandel et al., 1992).

***AGAMOUS* (*AG*)** AG is a class C floral homeotic transcription factor that determines the identity of stamens, carpels and later of the developing ovules. Moreover AG represses *WUS* blocking the meristem proliferation. *AG* is expressed in the ST3 and in later stages of flower development (Drews et al., 1991).

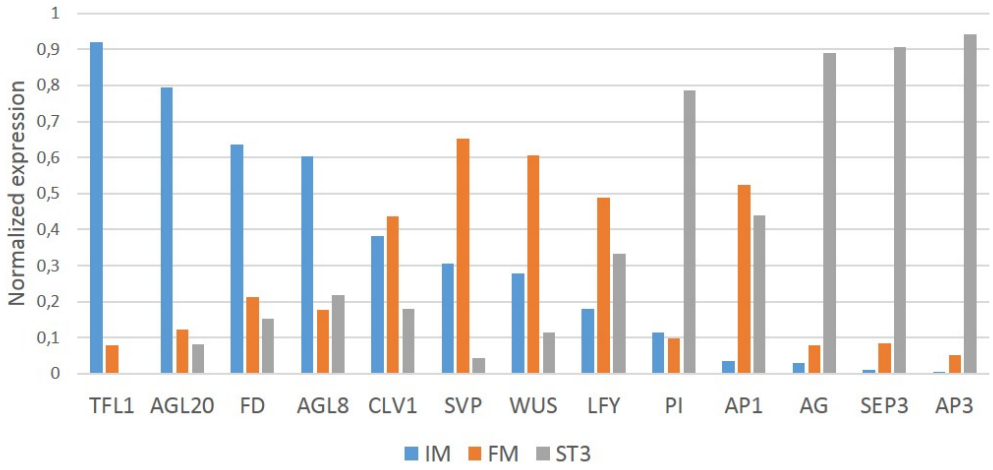


Figure 3.2: **Normalized expression of selected marker genes** The normalized expression levels of 13 selected marker genes confirms the high specificity of laser microdissection of floral tissues. The expression levels are shown in RPKM, normalized to a 0-1 scale relative for each gene.

***SEPALLATA3* (*SEP3*)** SEP3 is a class E floral homeotic protein necessary for the class A, B and C proteins to function. it is expressed from ST3 of flower development (Mandel and Yanofsky, 1998).

***APETALA3* (*AP3*)** AP3 is a class B floral homeotic protein that acts synergistically with PI. AP3, thus, determines the identity of petals and stamens and, accordingly with its function, is expressed in ST3 and later stages of flower development (Jack et al., 1992).

The 13 marker genes in our datasets behave as expected by their function and by the available expression data (Figure 3.2). This confirms that our dataset are reliable and rules out cross contamination of samples during the microdissection and amplification.

3.1.3 Characterization and confirmation of differentially expressed (DE) genes

We have identified differentially expressed genes using Cufflink (Trapnell et al., 2012). Setting a statistical cutoff of $FDR < 0.05$ we have detected:

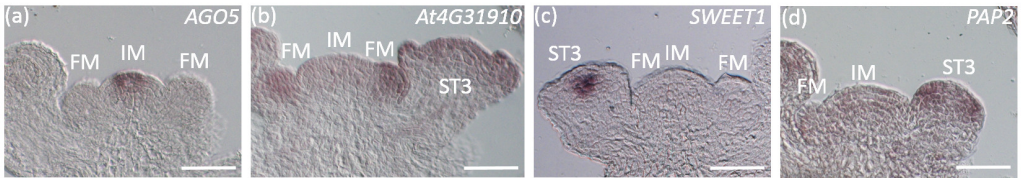


Figure 3.3: **Gene expression levels are confirmed independently by in situ hybridization** (a) Antisense probe targeting *AGO5*, which is detected in the inflorescence meristem (IM) and not in the floral meristem (FM), concordantly with what was expected from RNA-seq results (mean RPKM in IM=26,72 and in FM=3,69). (b) Antisense probe targeting *AT4G31910*, which is detected in the FM and flower at stage 3 (ST3) and not in the IM, concordantly with what was expected from RNA-seq results (mean RPKM in IM=1,05; in FM=19,36 and in ST3=11,01). (c) Antisense probe targeting *SWEET1*, which is detected exclusively in the ST3 concordantly with what was expected from RNA-seq results (mean RPKM in IM= ,88; in FM=2,55 and in ST3=19,86). (d) Antisense probe targeting *PAP2* which is detected more intensively in the ST3 compared to the IM and FM, concordantly with what was expected from RNA-seq results (mean RPKM in IM=7,20; in FM= 7,01 and in ST3=27,12). Scale bar is 50 micrometers.

- 46 genes differentially expressed between IM and FM,
- 171 genes differentially expressed between FM and ST3,
- 178 genes differentially expressed between IM and ST3.

The DE genes are listed in section 8.3.

In order to test for false positives, we selected four previously uncharacterized DE genes: AGO5 (AT2G27880), PAP2 (AT4G29080), SWEET1 (AT1G21460) and AT4G31910, and further characterize their expression patterns with in situ hybridization. We detected AGO5 mRNA in the IM, PAP2 and SWEET1 mRNA in ST3, and AT4G31910 mRNA in FM and ST3 concordantly with what we expected from RNA-sequencing data.

Then, we have searched for over-represented Gene Onthology (GO) terms in the DE genes sets in order to detect the main functional differences among IM FM and ST3. DE genes upregulated in IM or FM with respect to ST3 are enriched for GO terms of meristem development and maintenance and flower

development. On the other hand, the DE genes upregulated in ST3 versus IM or FM are enriched for GO related to specific organ development or response to endogenous stimulus. These data are consistent with the meristematic state of IM and FM, and the developmental and differentiation programs initiated during ST3.

We have searched for over-represented TF family members in the DE genes. Enrichment of member of a particular TF family is indicative of redundant or sinergystic action of its members in determining the differences between two developmental stages. We are particularly interested in redundancy cases, because they make null mutants ineffective in revealing the function of a gene.

At least 1 member of 14 TF families defined by Agris DB (Davuluri et al., 2003) is differentially expressed in the conditions tested. The *MADS-box* transcription factor family is overrepresented in almost all DE gene sets, consistent with its wide-ranging roles in reproductive meristem formation, development and differentiation (reviewed in Dornelas et al., 2011). Homeobox and *C2C2-YABBY* and *REM* families are overrepresented genes DE between FM and ST3, suggesting a role in the switch from meristematic state to differentiation of organs; while *SBP* and *ARF* TF families are characteristic of the IM.

3.1.4 Expression based clustering

In order to detect dynamic changes in the transcriptome, we clustered for expression levels the 1675 genes with the highest variation in the expression profile. We have defined six clusters of co-expressed genes. Clusters 4, 5 and 6 group genes that are expressed almost exclusively in one of the three developmental stages while clusters 1, 2 and 3 display similar but less emphatic expression patterns to the formers (Figure 3.4). GO terms enriched in the clusters comprehend all GO terms enriched in the DE genes plus others, such as transcriptional regulation and metabolic processes.

Moreover, in the clusters we detected strong enrichment of the *C2C2-YABBY* transcription factors family in clusters 1 and 6 (genes typical of ST3

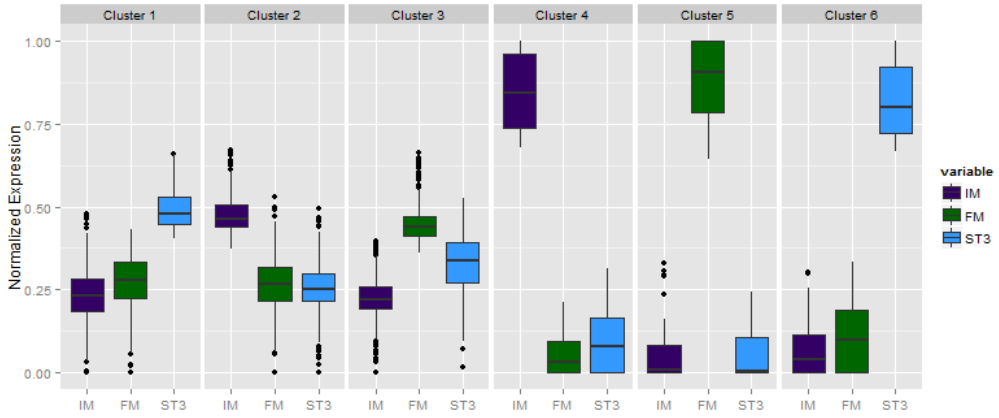


Figure 3.4: **Boxplot of normalized expression of genes in the six clusters** Clusters 1, 2 and 3 group genes preferentially expressed respectively in ST3, IM and FM, clusters 4, 5 and 6 group genes expressed in IM, FM and ST3 but with sharper changes in expression in respect to the first three clusters.

and thus of organ differentiation), of *SBP* family in cluster 2 (IM), of *ARF* and *Jumonji* in cluster 5 (FM) and of the *MADS-box* family members in cluster 6 and in clusters 4 and 5 to a lower level.

3.1.5 The *ap1 cal AP1-GR* system introduces more artifacts than laser microdissection of wild-type tissues

Our transcriptomic datasets of gene expression levels in IM, FM and ST3 should be similar to the one referring to the same stages produced with the *ap1 cal AP1-GR* system by Wellmer et al. in 2006 . We have decided to compare our datasets with the one produced with the *ap1 cal AP1-GR* system in order to detect transcriptomic differences between wild type and the induced *ap1 cal AP1-GR* plants.

We have quantified the correlation between the transcriptomes of IM, FM and ST3 in our experiments with the transcriptomes of the time points after AP1-GR induction analyzed by Wellmer et al. in order to identify precisely which time point after induction correspond to the FM and which

correspond to the ST3. The log fold changes of IM vs FM in our datasets correlated best with 0d vs 1d datasets from Wellem et al., ($\rho = 0.32$), and the log fold changes of FM vs ST3 comparison correlates best with 1d vs 2d dataset ($\rho = 0.26$) from Wellmer et al. Thus, as expected, the FM stage corresponds to the first day after AP1 induction (d1) and ST3 corresponds to the second day after AP1 induction in the datasets produced with *ap1 cal AP1-GR* system.

We detected a set of genes deregulated exclusively in the *ap1 cal AP1-GR* system which are enriched in GO terms including: response to stress, response to abiotic/endogenous stimulus, transcription regulator activity, while the genes with higher expression levels in our dataset show more pertinent functional enrichment, recovering terms including: post-embryonic development, flower development and multicellular organism development, and organellar components (see Chapter 6). The strong induction of AP1 activity in the *ap1 cal AP1-GR* system is therefore activating stress pathways that may perturb the floral development physiological transcriptional network.

3.2 Signs Of Redundancy During Early Flower Development

It is commonly accepted that functional redundancy is a key feature of plant molecular networks. Two transcription factors both homologous and expressed in the same cell are likely to be redundant (Briggs et al., 2006; Hauser et al., 2013). We analyzed our datasets in order to detect potential redundancies and to predict regulatory events.

3.2.1 Homology based redundancy

We compared phylogenetic analyses and expression profiles from our dataset for the TF families known to play pivotal roles in flower development such as *MADS* (Parenicová et al., 2003), *HOMEOBOX* (Kumar et al., 2007), *WRKY* (Rushton et al., 2010), *bHLH* (Carretero-Paulet et al., 2010), *NAC*

(Nakashima et al., 2012), *MYB* (Yanhui et al., 2006), and *WOX* (van der Graaff et al., 2009).

In collaboration with Dr. David Horner we detected potential redundancy cases by:

1. clustering by expression levels the TF family members that have a detectable change in expression profile among the three stages ,

2. comparing expression clusters and phylogenetic trees.

Using this method, we have detected known homology cases, such as AP1 and CAL (Bowman et al., 1993), as well as new potential cases, such as SAW1 and SAW2 which are known to redundantly regulate leaf margin growth (Kumar et al., 2007). Several additional candidates for functional redundancy in floral development include AGL6 and FUL, CAL and AGL87, IDD7 and IDD11 and two basic helix loop helix (bHLH) transcription factors At2G40200 and At3G56770.

3.2.2 Sub-Family wide redundancy

Always in order to detect functional redundancy cases, we applied the genome wide subfamily description defined by Friedrich Hauser and Julien Schroeder in 2013 to our datasets.

The gene subfamily definition that we have used is based on the genome-wide family definition performed by phytozome (http://www.phytozome.net/) but it is aimed at detecting potential redundancy rather than at describing phylogenetic relationships. Hauser et al. obtained this subfamily definition by ulterior clustering of the gene in the Phytozome families by sequence similarity (Hauser et al., 2013). The subfamilies should group together genes that are likely to be redundant if co-expressed.

We combined the genome-wide family definitions with our expression based clusters and retained as potential redundant two or more genes belonging to the same subfamily and co-expressed in either:

- clusters 2 and 4 (gene preferentially expressed in the IM),

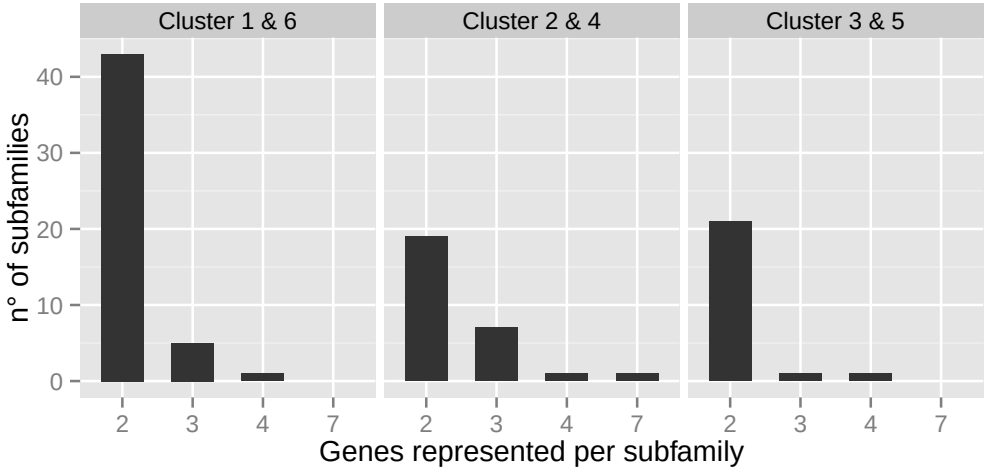


Figure 3.5: **Many subfamilies are represented in more than one potentially gene in co-expressed genes.** UP to 46 subfamilies are represented by 2 genes in our dataset of co-expressed genes, and we detected up to 7 genes from the same subfamily in clusters 2 and 4. Genes that are both co-expressed and in the same subfamilies are potentially redundant with each other.

- clusters 3 and 5 (genes preferentially expressed in the FM),
- clusters 1 and 6 (genes preferentially expressed in the ST3).

We detected 126 known or new potential redundant genes, 42 in the IM, 26 in the FM and 56 in the ST3. Out of 870 sub-families represented in the groups, 93 have at least two members in the same group and 15 have at least three member in the same group (see Figure 3.5).

Our approach correctly detects known redundancy cases, such as *AP1 CAL* and *AP1 SVP* in the FM and adds to them *AGL71*, which is a good testing candidate for future testing. Moreover we detect multiple new potential interesting cases such as

- TARGET OF MONOPTEROS 6 (TMO6) with its homolog AT5G65590 in the FM,
- CLV1 with its homologs AT4G08850 and AT1G16670 in the IM,

- the receptor protein kinase TMK1 (Chang et al., 1992) with its homologous AT5G49760 in ST3.

We hypothesized that if functional redundancy is such an important feature of plant development and if potential redundancy is correctly predicted by the subfamilies, we should find in co-expressed genes, more functional redundant genes couples than expected by chance.

We have defined as potentially redundant two genes that are both co-expressed and belonging to the same subfamily. Then we have sampled 1000 random gene sets of the same dimension of the co-expressed gene set in order to describe the distribution of genes belonging to the same subfamily under the null hypothesis. We tested the enrichment of members of the same subfamily in the co-expressed genes against this distribution. In FM and ST3 there are more potentially redundant genes than expected by chance with an FDR corrected attained significance level of respectively 2.7×10^{-04} and 8.7×10^{-17}.

3.2.3 Large mutant screening

We carried out a large screening of insertional mutant for a selection of genes that are differentially expressed in the stages under study and/or target of key transcription factors such as SVP, AP1, SEP3, LFY AP3 and PI. This screening is ongoing and by now we have screened 26 independent insertional lines for 22 different genes recovering 13 confirmed homozigous lines (see Table 3.2). We have screened these homozigous lines for phyllotaxis of the inflorescence, dimension and number of floral organs and dimension of the meristems without detecting any effect on flower development.

3.2.4 Inferring of regulatory events

We hypothesized a pipeline for inferring of regulatory network from expression data and putative promoter sequence of co-expressed genes.

First we defined as putative promoter the 1000 bp upstream of the transcription start site (TSS) of the genes contained in the 6 expression

TAIR ID	Feature ID	Mutant line	Confirmed	Expr. in
AT1G20910	AT1G20910	SALK_141443.13.85	NO	FM
AT1G21460	AT1G21460	SALK_029479	YES	ST3
AT1G21460	AT1G21460	0SAIL_883_C04	YES	ST3
AT1G25440	AT1G25440	GABI_458H02	NO	IM
AT1G26780	MYB117	SALK_025235	NO	IM
AT1G68825	RTFL15	SALK_013609.33.75	YES	ST3
AT1G68825	RTFL15	SALK_123407	YES	ST3
AT1G77080	MAF1	SALK_072871	NO	IM-ST3
AT1G77950	AGL67	GABI_340D03	YES	ST3
AT1G77950	AGL67	SALK_050367.42.65	NO	ST3
AT2G27880	AGO5	GABI_265A07	YES	IM
AT2G27880	AGO5	SALK_050544.37.75	NO	IM
AT2G37630	AS1	SALK_023987	NO	ST3
AT2G45650	AGL6	SALK_095121	NO	IM
AT3G03990	AT3G03990	WiscDsLoxHs137_07E	YES	FM-ST3
AT3G24420	AT3G24420	SALK_126829C	YES	ST3
AT3G50060	MYB77	WiscDsLox338D05	YES	FM-ST3
AT3G50630	KRP2	SALK_130744.55.00	YES	IM-ST3
AT4G11400	AT4G11400	SALK_016155.46.60	ETERO	FM
AT4G15620	AT4G15620	SALK_057616	YES	FM-ST3
AT4G29080	PAP2	SALK_070738.23.50	ETERO	ST3
AT4G31910	AT4G31910	SALK_123920.30.15.x	NO	FM
AT4G37260	MYB73	WiscDsLoxHs064_10A	YES	IM-FM
AT5G23000	MYB37	GABI_325E06	NO	IM
AT5G57340	AT5G57340	SALK_139344.52.30.x	NO	IM
AT5G60200	TMO6	SALK_201987	YES	FM

Table 3.2: List of insertional mutants analyzed in this work.

based clusters. Then we searched for recurrent motives in the putative promoters of the co-expressed genes and we retained only the motives similar ($p-value \leq 10E-6$) to the confirmed TFBS of *Arabidopsis thaliana* (Steffens et al., 2004). We detected 284 enriched motifs associated with 21 TF families.

We have linked the 144 putative TFs represented in the clusters to the 284 distinct motives hypothesizing that enrichment of particular TF families in an expression cluster and enriched of corresponding TFBS motives in the promoter of the genes belonging to a correlated or anti-correlated clusters might be an evidence for a regulatory event upon binding of the TF to the promoters.

Moreover, since motif occurrence alone is not predictive for functional TF binding (Moyroud et al., 2011) and transcription factor are known often to function in complexes (Smaczniak et al., 2012) and/or to act combinatorially (Molkentin and Olson, 1996; He et al., 2011; Feller et al., 2011), we have searched for combinations of TFBS motifs that exhibit significant patterns of co-occurrence in promoters of co-expressed genes and whose presence correlates (or anti-correlates) with the expression levels of the corresponding TF families within the clusters. In the promoter of the genes preferentially expressed in the IM (clusters 2 and 4), we have detected co-occurrence of:

Combinations MADS, AP2 and MYB binding site motives:

Clusters 2 and 4 contain *AGL20* (*SOC1, MADS*), *DREB2A* (*AP2*), *AT5G61590* (*AP2*), *MYB17* (*MYB*) and *TRFL10* (*MYB*) as well the bHLH transcription factor *bHLH071*. 54 putative targets for this regulatory module are significantly enriched in GO terms including: transcription factors, ABA signaling, reproductive structure development.

Combination of bZIP and AP2 binding site motives:

Clusters 2 and 4 contain both *FD* (*bZIP*) and *ERF12* (*AP2*) genes. The 204 genes potentially regulated by this module are enriched in GO terms including: thylakoid, organelle membrane and organelle parts, consistent with a potential role for this module in photoperiod dependent floral transition (Abe et al., 2005).

co-occurrence of AP2, MADS and E2F binding motives

SVP (*MADS*) and *CRF2* (*AP2*) have the expression pattern of a potential activator or and repressor of the clusters 3 and 5. 60 putative targets show a weak enrichment for the GO terms mitochondrion and amine biosynthesis, being little informative on the function of these genes.

Two putative regulatory modules were associated with ST3; the first contains a of binding motives combination of *MADS, SBP* and *bHLH* TF families, and a putative regulatory function can be assigned to *AGL101* (*MADS*), *SPL8* (*SBP*), and *AT1G05710* (*bHLH*).

The second predicted regulatory module associated with SP3 includes candidate binding sites for *HOMEOBOX* and *MADS* families represented in appropriate clusters by AG and AP3 (MADS) and the HOMEOBOX genes ATHB6 and KNAT4. 20 potential target genes show functional enrichment for the GO term endomembrane system although these result should be treated with care as the number of tested genes is low.

3.3 Analysis Of The *REM* Transcription Factor Family in Early Flower Development

We have got interested on *REM* transcription factors since we observed that *REM34* and *REM24* are co-expressed with *LFY* and *AP*. We have published most of the following results on REM genes in Mantegazza et al., (2014b) in collaboration with Prof. Lucia Colombo and Dr. Morandini from our same University and Prof. Marcio Alveis Ferreira group at the Federal University of Rio de Janeiro.

3.3.1 Expression Analysis

In order to describe the expression pattern of *REM* transcription factors, we both analyzed the expression profile of these genes ourself with qPCR, and collected publicly available microarray expression data.

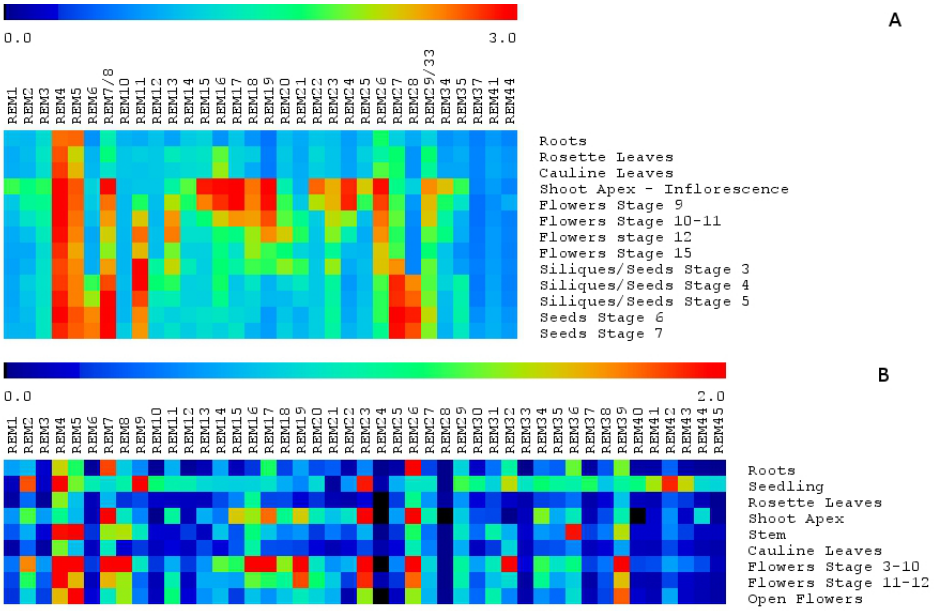


Figure 3.6: **Expression analysis of REM genes**(A) REM genes expression levels measured by microarray, data are shown in log10 transformed absolute values from AtGeneExpress developmental atlas. (B) REM genes expression levels measured by quantitative Real-time PCR shown in square root transformed $2^{-\Delta CT}$ values.

In collaboration with Dr Piero Morandini we have collected the data from the NASCarrays database (ftp://arabidopsis.info/pub/NASCArrays/Data/) and generated a heat-map of the expression patterns of REM genes. Eleven *REM* genes (*REM9, REM30, REM31, REM32, REM36, REM38, REM39, REM40, REM42, REM43* and *REM45*) have no corresponding probe on the ATH1 array; moreover the ATH1 array probe 256918_s_at does not distinguish between *REM7* and *REM8* and 257436_s_at does not distinguish between *REM29* and *REM33*, therefore we refer to measurements from these probes as *REM7/8* and *REM29/33*. *REM* genes are preferentially expressed during flower and seed development according to microarray data. Only *REM4* and *REM5/VRN1* are expressed in vegetative tissue; *REM10, REM37, REM41* and *REM44*) are almost undetected (Figure 3.6A).

We have refined the expression data with real-time qPCR, which is considered to be a gold standard for expression analysis (Wang et al., 2006).

The raw data that we analyzed were produced from our collaborator Dr. Camila Patreze from the Federal University of Rio de Janeiro. The qPCR expression analysis is concordant with microarray expression data and defines expression patterns of REM genes on a wider quantitative range. We have confirmed that three groups of REM genes ((i) REM4, REM5, REM7, REM8, (ii) REM15, REM16, REM17, REM18 REM19 and (iii) REM22, REM23) are highly expressed during early stages of flower development. Moreover, we have analyzed the peculiar expression pattern of *REM32*, which is distictively expressed during early flower development, of *REM39* which is strongly expressed during early flower development and of *REM42* which is distinctively expressed in seedlings (Figure 3.6B).

3.3.2 Co-expression

In order to infer the function of *REM* gene family we have quantified the correlation of the expression levels (Menges et al., 2007, 2008) of REMs and key flower transcription factors in collaboration with Dr. Piero Morandini. The rationale of this analysis is that given a statistical correlation in the expression level of two genes they are likely to be involved in the same pathway.

We analyzed the correlation among *REMs* and two groups of genes:

- the FMI genes *LFY*, *AP1* and *CAL* (Irish and Sussex, 1990; Weigel et al., 1992),

- the *MADS-box* floral homeotic genes, *APETALA3* (*AP3*), *PISTILLATA* (*PI*), *AGAMOUS* (*AG*) and again *AP1* (for review see Krizek and Fletcher, 2005).

As a positive control we used *STK*, which directly regulates the expression of *REM20/VDD* (Matias-Hernandez et al., 2010). Finally, as negative controls we used *TERMINAL FLOWER 1* (*TFL1*), whose transcripts accumulate only in the inflorescence meristem and are excluded from the flower (Ratcliffe et al., 1999), and *SHORT VEGETATIVE PHASE* (*SVP*) which is both a

FMI gene and a negative regulator of flowering active in vegetative tissues (Gregis et al., 2008).

The expression levels of *REM16, REM17, REM18, REM19, REM22, REM23, REM24* and *REM34* are correlated at $r > 0.7$ with the expression levels of *LFY* and to a lower level with *AP1* and *CAL*. Moreover, the expression levels of *REM1, REM4, REM7, REM8, REM26* and *REM29/33* are correlated to the first group of REMs and more weakly with the FMI genes. A third group, *REM11, REM13, REM20* and *REM21* is correlated to *STK*. We detected occasional correlation and anti-correlation of *REM*s and the floral homeotic genes *AP3*, *PI* and *AG* ($0.5 > r > -0.5$), no significant correlation with SVP ($0.2 > r > -0.2$) and only one gene (REM35) correlates with TFL1 although weakly ($r = 0.5627$).

3.3.3 Multiple evidences support a potential function of *REM34*, *REM35* and *REM36* in reproductive meristems

In order to restrict and refine the set of *REM* genes potentially involved in early flower development and to select candidates for functional characterization with insertional mutants, we have screened publicly available ChIPseq data and defined a set of *REM* genes that are target of key floral regulators. Screening the high confidence target datasets for: LFY (Winter et al., 2011), SVP (Gregis et al., 2013), AP1 (Kaufmann et al., 2010), SEP3 (Kaufmann et al., 2009), PI and AP3 (Wuest et al., 2012), AG (ÓMaoiléidigh et al., 2013), AGL15 (Zheng et al., 2009) and AP2 (Yant et al., 2010) we have observed that the ChIP-seq data are consistent with what expected from the co-expression analysis. *REM17* and *REM18* are both targets and coexpressed with LFY. On the other hand *REM34* is both target of and coexpressed with AP1.

AP3 and PI have multiple binding sites in the *REM34, REM35* and *REM36* cluster on chromosome 4. These three genes are linked on chromosome 4 within 10 kb and are part of a bigger linkage cluster containing nine *REM*s within 30 kb. Binding of AP3 and PI falls precisely in the short non coding

Figure 3.7: **AP1, AP3, PI and SVP binding sites in the *REM34, REM35, REM36, REM37* cluster on chromosome 4.**

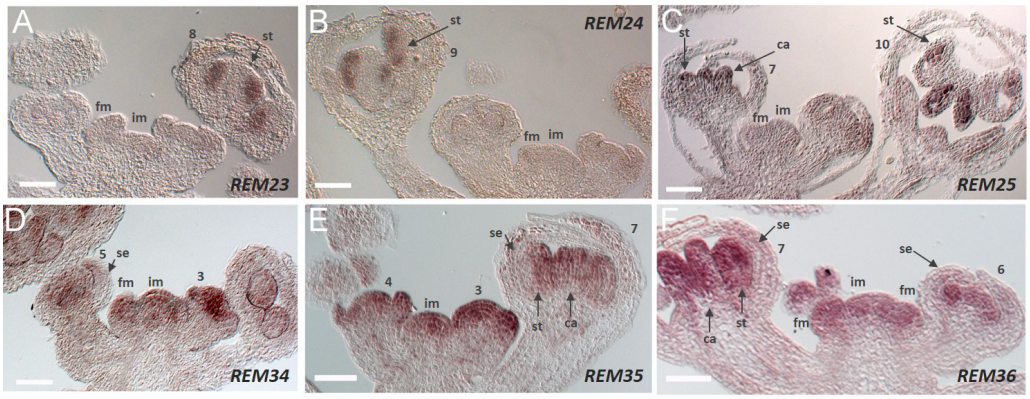


Figure 3.8: **Many subfamilies are represented in more than one potentially redundant gene in our dataset.** Number of subfamilies represented by more than on gene in the expression clusters.

region between *REM35* and *REM36*, which is 565 bp from the stop codon of *REM36* to the start codon of *REM35*, and in the small inter-genic region between *REM36* and *REM37*, which is 687 bp long from the stop codon of *REM37* to the start codon of *REM36* (Figure 3.7). *REM34, REM35* and *REM36* are upregulated in *pi-1* and *ap3-3* mutants (Wuest et al., 2012) while they do not change expression upon binding of AP1 (Kaufmann et al., 2010) and SVP (Gregis et al., 2013).

Then we characterized by RNA in situ hybridization the expression profiles of *REM23 REM24, REM25, REM34 REM35* and *REM36* during early stages of flower development with in situ hybridization experiments. We selected *REM23, REM24* and *REM25* because they are phylogenetically closely related (Romanel et al., 2009) and because they are co-expressed with the FMI genes. For the same reasons we have decided to characterize *REM34, REM35* and *REM36*, which are also direct targets of SVP, AP1, AP3 and PI

(see above).

REM23, REM24 and *REM25* are all expressed in stamens starting from stage 7/8 of flower development and are not expressed in earlier stages (Figure 3.8 A-B-C). *REM34, REM35* and *REM36*, instead, are expressed in the inflorescence meristem, floral meristem and, from later stages on, exclusively in the inner floral whorls and not in sepals (Figure 3.8 D-E-F).

3.3.4 Mutants

We could not detect peculiar phenotypes in the insertion mutant lines for *REM24* and *REM34* and of their closest homologues, *REM23* and *REM36*. In particular, we detected no variation in floral organ number, identity, structure, and inflorescence phyllotaxis. qPCR expression analysis revealed that only *rem23* is a confirmed complete knock-out, in the other mutants we detected mRNA at lower level than in the wild type. We produced:

- multiple mutant combinations of *rem24* and *rem34* with the mutants *ap1-10* and *lfy-2* (co-expressed FMI genes),
- the double mutant *rem23 rem24* since *REM23* and *REM24* are co-expressed and have highly similar amino acid sequence.

No peculiar phenotypes were detected in the *rem24 rem34* and *rem23 rem24* double mutants. No suitable mutant lines were found for *REM25*.

3.4 Ongoing work and tool development in our studies on the *REM34* gene cluster

We have taken the task of studying the complex *REM34* genomic region as a chance to develop new tools dedicated to plant reverse genetics.

The *REM34* genomic region is located on chromosome 4. This region contains 9 recently duplicated *REM* genes (*REM34, REM35, REM36, REM37, REM38, REM39, REM40, REM41* and *REM42*) and one unrelated gene (*UBP18*) within 30 kb (Swaminathan et al., 2008). It is unlikely to obtain a

multiple mutant for the *REM* genes contained in this cluster with current reverse genetic technologies. Moreover, since the B3 domain in REM genes is highly divergent (Romanel et al., 2009), it is unlikely to obtain a multiple knock-down with one single artificial small interfering RNA (siRNA) and expression of multiple siRNA is complicated and unlikely to deliver consistent results as the number of genes involved grows. We are testing little explored technologies and developing new ones in order to remove the function of the genes in the REM34 cluster from .

The experiments presented in this section are unpublished and ongoing and they will be reprise in detail in Part III.

3.4.1 Expression of multiple RNA interfering fragments within one single gene

RNA interference is a process by which RNA fragments can regulate and inhibit gene expression, generally targeting mRNA for degradation with base pair complementarity directed specificity (Fire et al., 1998). RNA interference pathways are conserved in eukaryotes (Saumet and Lecellier, 2006) and since they allow an extremely simple method to induce inheritable specific gene silencing, they have been widely used in functional genomic studies and biotechnologies.

RNAi technology of choiche

We have decided to implement a multiple RNA interference technology in which we express multiple dsRNA fragments targeting multiple mRNA within one single transcript (Bucher et al., 2006; Miki et al., 2005).

RNA interference in *Arabidopsis* can be triggered with double strand RNA (dsRNA), the easiest way to produce dsRNA in plants is probably to stably transform a gene that expresses a transcript with high self complementary which is capable of folding on itself producing an hairpin RNA (Wesley et al., 2001). The hairpin RNA is fragmented by DICER-LIKE (DCL) proteins in small dsRNA fragment activating the RNA interference cascade and its downstream effects.

Different methods are available to down-regulate multiple genes using an RNA interference approach in *Arabidopsis*. For example, multi-gene silencing can be obtained tranforming *Arabidopis* with one or multiple DNA fragments which express multiple hairpin RNA each targeting a single gene.

However, since hairpin RNA are cleaved in smaller fragments of 20-25 base pairs that are used for targeting, it is possible to assemble together multiple unrelated genomic fragments in one single hairpin RNA-producing transcript. These fragments, when cleaved by DCLs will target for degradation the genes from which they have been amplified (Bucher et al., 2006; Miki et al., 2005). It is thus possible to knock-down multiple unrelated genes with one single hairpin RNA. The advantages of this strategy are many:

- Only one construct has to be transformed into the plant,
- Only one promoter and one terminator are needed,
- The expression of only one hairpin RNA gene must be monitored,
- The down-regulation of multiple genes is likely to be more uniform when compared to other methods.

We are implementing this strategy in order to down-regulate *REM34, REM35* and *REM36*. Applying Golden Gate cloning (Engler et al., 2008) to the assembly of the construct we have simplified the cloning procedure making it easily implementable on virtually every set of genes.

Golden Gate cloning the *REM* fragments for RNAi

We have searched for three 200 base pairs long regions specific for the coding sequence of each the genes *REM34, REM35* or *REM36* and used BLAST to check the results against *Arabidopsis thaliana* genome for specificity. We have PCR amplified the three region adding the BsaI sites in the primers and performed one single Golden Gate reaction to directionally clone the *REM34, REM35* and *REM36* fragments altogether in a pENTR^{TM} vector previously modified to function as an Golden Gate acceptor, producing the pENTR-RNAiREM vector. Using with white/blue screening, we have

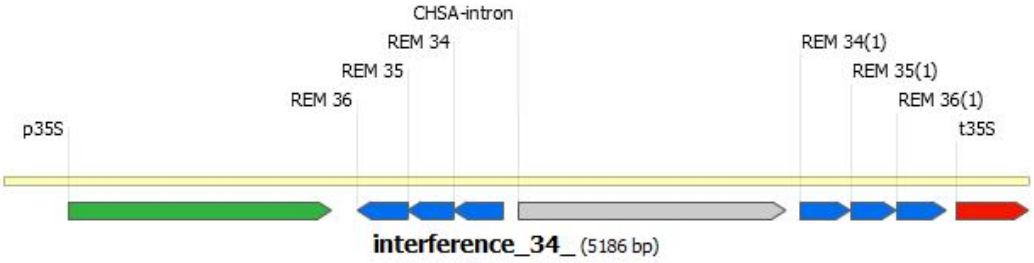


Figure 3.9: Schematic view of hairpin RNA producing gene to target RNA interfenrence on *REM34, REM35* and *REM36*.

estimated an efficiency of the Golden Gate reaction of 88.4%. We have checked 8 white colonies by PCR and they all produced the expected results.

Then we have used LR reaction to subclone the REM fragments into the pFGC5941 vector and used it to transform *Arabidopsis* in which the REM fragments will be inserted both in reverse and forward orientation separated by an intron (see Figure 3.9). Right now we are analyzing the transformant plants.

3.4.2 Plans for efficient deletion of *REM34* cluster

We are developing tools for gene targeting in *Arabidopsis thaliana* in order to delete *REM34, REM35* and *REM36* from the genome. Our system loosely resembles the famous strategy for gene targeting in mice developed by Thomas and Capecchi in 1987 .

Rationale and DNA parts needed

The technology that we are developing is based on *Agrobacterium* mediated transformation of a DNA fragment containing:

- **Custom nucleases:** we are using TALEN in order to introduce a double strand break (DSB) in the genomic region to be deleted. We are using Golden Gate assembled TALEN (Cermak et al., 2011).

- **Positive selection marker:**, We are using the *bar* expression cassette (see De Block et al., 1987) as positive selection marker flanked by two homology regions matching genomic regions upstream and downstream of the DSB site.

- **Negative selection marker**. We are using the *E. Coli* gene CodA, which is capable to convert 5-Fluorocytosine into its toxic metabolite 5-Fluorouracile in *Arabidopsis* (Perera et al., 1993) as a negative selection marker.

Transforming these parts into the plant altogether within one single construct will simplify downstream analysis. Once these parts are transformed into the plant, we expect this series of events to happen:

1. The nucleases will induce a DSB specifically in the *REM34* cluster,

2. The DSB will be repaired by homologous recombination (HR) or non homologous end joining (NHEJ) pathways:

    - If the DSB is repaired by NHEJ a mutation can be introduced, this mutation can heavily modify the site targeted by the nucleases making it . This step is irreversible if the target site is not recognized by the nucleases anymore.
    - if the DSB is repaired by the HR pathways two things can happen: (i) if the sister chromatid or the homologous chromosome are used as donor, the DSB will be repaired without introducing mutation and the nucleases can cut again in this locus. (ii) If the homology regions in the construct are used as donor for HR repair, the *bar* gene will substitute the *REM34* cluster deleting it from the genome. We expect this to happen with low efficiency but this reaction is irreversible, since it will delete from the genome the nuclease target site.

3. In the plants where *REM34* has been substituted by *bar* in the cell lineage that will produce gametes, the targeted deletion is inheritable.

4. When the substitution event is inherited in the T2, the positive selection marker *bar* have been de-concatenated form the negative selection marker *CodA* and is free to segregate. Plant surviving both selections are the deleted specimens.

Thus our strategy will allow easy detection of the rare event of specific deletion .

Molecular cloning of the parts

We have cloned the part needed from multiple sources:

- Homology regions

We have PCR amplified the homology regions from *Arabidopsis* genome flanking them with Golden Gates sites.

- *Bar* expression cassette

We have PCR amplified the *Bar* expression cassette from the plasmid set described in (Curtis and Grossniklaus, 2003) flanking it with Golden Gate sites.

- *CodA* expression cassette

We have cloned the *CodA* coding sequence from *E. Coli* and used golden gate to flank it with 35S promoter and tNOS terminator.

- TALEN

We have produced custom TALEN targeting *REM36* coding sequence using guidelines and material produced by Cermak et al. in 2011.

- pUBQ10

We are subcloning the TALENs under the pUBQ10 promoter, which was PCR amplified From *Arabidopsis* genome.

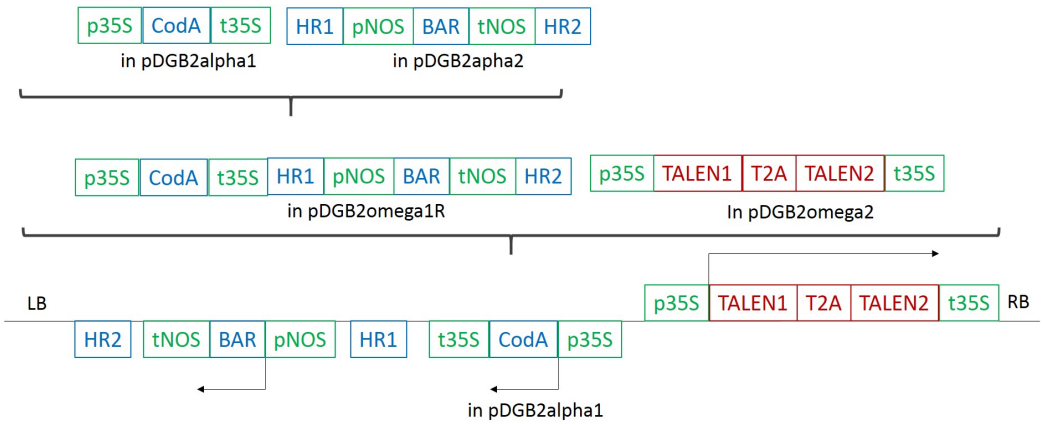


Figure 3.10: **Cloning Scheme** Series of Golden Gate reaction that we are using to produce the homologous recombination construct.

- T2A polypeptide

We are using the 2A peptide to translationally breakdown two TALEN coding sequences that are transcribed in the same mRNA as performed by Zhang et al. in 2010 for ZFN. We have synthesized the 2A sequence following the sequence guidlines drawn by Kim et al. in 2011.

We are using Golden Gate cloning and its iterative application Golden Braid (Sarrion-Perdigones et al., 2011) to assemble all these parts in one single construct, following the single scheme.

1. Assembling of the following modules in GoldenBraid plasmids with Golden Gate reactions
   - Homology Region 1 - *bar* - Homology region 2
   - p35S - *CodA* t35S
   - pUBQ10 - TALEN1 - T2A - TALEN2 - tNOS
2. Assembling with GoldenBraid reactions all the modules within one single self-sufficient construct that can be *Agrobacterium* transformed in plants with GoldenBraid reactions.

Right now we are producing the last module (pUBQ10 - TALEN1 - T2A - TALEN2 - tNOS). The whole cloning scheme is depicted in Figure 3.10

Chapter 4

Conclusions And Future Prospects

We have identified numerous genes potentially involved in early flower development. The loss of function mutant of those genes do not display a phenotype in the developmental stages under studies and we believe that this is caused by functional redundancy. We are implementing computational approaches to identify redundancy and gene targeting/multiple RNAi strategies to produce and analyze multiple mutant.

4.1 Transcriptomic Studies Of Native Organs

We believe that in a multicellular organism:

1. the transcripome varies continuously from cell to cell while the organism advance in its growth stages and respond to external clues (Lovatt et al., 2014),

2. transcriptome variations are one of the earliest manifestation of the processes that cause macroscopic changes in an organism (Sul et al., 2009),

3. a precise analysis of transcriptomic variations are needed to characterize the molecular regulation of macroscopic events.

Thus we have used laser microdissection (LM) in order to precisely dissect the early stages of flower development, a key step in plant life cycle. LM of early stages of flower development allowed us to study their transcriptome in wild type tissues grown in their physiological environment (Schmid et al., 2012).

Setting an efficient protocol for LM can be challenging but we have shown that with today's technologies LM can be used to reliably study the transcriptome of reporductive meristem from as little as 15 inflorescences and two replicates. After microdissecting the stages of interest and sequencing and assembling their transcriptomes, we tested the reliability of our results with multiple controls that returned encouraging outcomes:

1. Above all we were concerned with cross-contamination of samples, since (i) introduction of infinitesimal quantity of contaminants before the amplification reaction can have great repercussions on results and (ii) the LM protocols requires morphological identification of the developmental stages and thus is susceptible to human error. Controls on the expression levels of 13 marker genes selected from literature allowed us to confirm that the transcrptomes in our datasets indeed belong to the stages under study.

2. Next we have tested our DE genes datasets for false positives. In situ hybridization allowed us to test the expression of four DE genes (*AGO5, PAP2, AT4G31910* and *SWEET1*) in a LM independent manner with consistent results. The expression pattern of these four genes is highly comparable between the two methods.

3. We were highly concerned about the statistical power of our analysis, thus about genes which differential expression is decisive in determining the differences among the stages under study accidentally left out from the DE gene set (i.e. false negatives). We reasoned that false negatives may hamper the prediction of redundant gene pairs and the prediction of regulatory pathways.

In order to avoid this problem, we have clustered genes for expression levels bypassing the statistical analysis (Wertheim, 2012).The obtained gene clusters are enriched in the expected gene ontology terms, confirming the overall biological validity of this analysis. Anyway genes taken from the cluster set should be tested for differential expression by in situ hybridization before moving to single-gene-centered studies, since the risk of incurring in a false positive may be high in this case.

We have shown that LM-RNAseq is a reliable technology to study the transcriptome of early flower development in physiological condition.

Moreover, comparing our results with the correspondent datasets produced with the artificial *ap1 cal AP1-GR*, we have shown that LM introduces less perturbations in the system than the AP1-GR technology. We think that LM can be further applied in defining transcriptomes of mutant organs and of even more specific tissues as required for molecular network construction.

4.2 From transcriptomes to candidate genes

We have analyzed 13 confirmed homozygous insertional lines of genes potentially involved in early flower development without detecting any effect of the mutation on the stages under study. This can be caused by functionally redundancy. Functional redundancy often occurs among co-expressed homologues (Briggs et al., 2006; Hauser et al., 2013). Numerous cases of confirmed redundant or partially redundant genes in *Arabidopsis* are listed by Briggs et al. (2006) and Lloyd et al. (2012)

4.2.1 Hypothesis on functional redundancy

We reasoned that if we found genes with highly similar sequence co-expressed in our datasets, this might be a strong hint to functional redundancy during early flower development.

We have used the phantom DB (http://phantomdb.ucsd.edu/) as a source of gene sets predicted to be redundant by sequence similarity (namely gene subfamily) and we searched for co-expression patterns in our datasets. We

detected multiple genes that are co-expressed and potentially redundant that will be tested in the future with multiple loss of function mutants and amiRNA.

Another hypothesis for the lack of effect of single mutants on development can be functional degeneracy. Degeneracy happens when dissimilar components perform similar functions under certain conditions (Edelman and Gally, 2001). All these mechanism are thought to give robustness to biological systems (Whitacre and Bender, 2010) and are to be taken into account for future functional studies.

4.2.2 Inferring of gene regulatory circuits

We have built putative regulatory circuits using expression data, the sequence of putative promoters and known TFBS. We tested these predicted circuits with confirmed regulatory events, such as SVP, AP1, SEP3 and SOC1 which regulate *AG*, *SEP3*, *AP3* (Gregis et al., 2008, 2009, 2013; Immink et al., 2012; Kaufmann et al., 2009, 2010) *SEP2* (Immink et al., 2012), *SOC1* (Li et al., 2008), or for the TF AP2 which can directly repress *SEP3*, *SEP2* and *AG* probably in FM (Yant et al., 2010) and we found that the TFs and their known targets are recovered by our approach.

Additionally, we are able to tentatively predict combinations of TFs (at least at the level of families) that might contribute, in concert, to the regulation of groups of genes pertinent to diverse aspects of early floral development. In the future it would be interesting to test the regulatory interaction that we found and to further test how co-occourence of TFBS can be used to predict regulatory interaction. We are interested in refining this approach with new and more predictive methods for TFBS matching which take into account inter-positional sequence dependence and variable spacer lengths in searching for binding sites in the genome (Mathelier and Wasserman, 2013).

Alternatively regulatory circuits can be predicted with network analysis, which generally requires wider datasets than the one that we have produced (Zou and Conzen, 2005).

4.3 Future work on *REM* genes

Transcriptomic analysis and ChIP-seq data indicate that *REM* genes are implicated in many developmental processes but their function is still poorly understood.

REM are widely expressed throughout development and we conclude behaving as broad regulator of development with key function in reproduction. Our refined expression analysis can serve as a starting point for future functional characterization of this gene family.

The expression level of many REM genes and key floral regulators such AP1 LEAFY, AP3, PI is strongly correlated. Combining the co-expression data with publicly available ChIP-seq TF binding data we decided to further characterize REM34, REM35, REM36 since:

- they are target of AP1, LFY, AP3 and PI,
- they are expressed in the floral meristem and during earliest stages of flower development,
- *REM34* is co-expressed with the FMI genes,

We have defined and presented sets of REM genes likely to be involved in flower development and we have decided to focus on three of them for further functional studies.

REM34, REM35 and REM36 are closely related homologues clustered on chromosome 4 within less than 10 kbp. Single mutants of REM34 and REM35 apparently do not show any difference in flower development and thus we hypothesized that these genes may be highly redundant. Unfortunately they are in close linkage and we could not analyse multiple mutant combinations. In the future it would be interesting to knock-down all of these genes using a multiple RNA interference approach (Abbott et al., 2002) or to produce multiple null mutants using genome editing technologies (Miller et al., 2011; Cong et al., 2013).

Our analyses highlight the difficulties of studying this gene family due to redundancy and genomic positions, despite the huge amount of information

that are nowadays available in different databases as well as the enormous quantity of data arising from the high throughput studies, which all together clearly suggest that this family should be important for reproductive development.

4.4 Expected development of *REM34* cluster targeted deletion studies

Even if engineered nucleases are becoming the system of choice for gene targeting in plants, many questions are still open, such as:

- Which nuclease induces mutations with the highest efficiency and the highest specificity toward the chosen targeted site,
- in which cell the nuclease should be expressed to obtain the highest ratio of heritable targeted events,
- how the donor DNA sequence should be designed to cause a high rate of easily detectable gene targeting events.

With our approach we expect to find a viable mutant in the T2 of the plants transformed with the deletion inducing construct within a reasonable number of T2 plants. We will detect the deletion by PCR and then we plan to:

1. select plants in which the *REM34* cluster is deleted,
2. sequence the flanking region of the deletion in order to the structure of the deletion site in the genome,
3. self-cross the plant to obtain plant with homozigous mutation,
4. describe the phenotype and test for complementation with *REM34*

Possible drawbacks and eventualities of our strategy are:

- Partial integration of the construct in the genome can deconcatenate the positive and negative selection markers yielding false positives in the double selection step on the T2 plants,

- Higher efficiency of the NHEJ pathway can shift the results irreversible toward small mutation in the targeted site instead of inducing homologous recombination,
- The system is not suitable for inducing HR in the cell lineage that will produce gametes and yields only somatic mutation.

Moreover we are right now working on a parallel cloning scheme in which TALEN are substituted by CRISPR/Cas9. TALEN seems to be more specific than CRISPR, but TALEN highly repeated sequence causes complication in the cloning procedure, therefore we thing CRISPR can be more suitable for bulk application of targeted mutagenesis.

4.5 Final Conclusions

We have improved the knowledge on early flower development transcriptome, characterized a high number of single mutants and laid the basis for hopefully more successful multiple mutant screening which eventually will lead to a functional model of the transcriptional network that underlies early flower development.

Bibliography

James C Abbott, Abdellah Barakate, Gaelle Pinçon, Michel Legrand, Catherine Lapierre, Isabelle Mila, Wolfgang Schuch, and Claire Halpin. Simultaneous suppression of multiple genes by single transgenes. Down-regulation of three unrelated lignin biosynthetic genes in tobacco. *Plant Physiology*, 128(3):844--853, 2002.

Mitsutomo Abe, Yasushi Kobayashi, Sumiko Yamamoto, Yasufumi Daimon, Ayako Yamaguchi, Yoko Ikeda, Harutaka Ichinoki, Michitaka Notaguchi, Koji Goto, and Takashi Araki. FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science*, 309(5737): 1052--1056, 2005.

Jens Boch and Ulla Bonas. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology*, 48:419--436, 2010.

Jens Boch, Heidi Scholze, Sebastian Schornack, Angelika Landgraf, Simone Hahn, Sabine Kay, Thomas Lahaye, Anja Nickstadt, and Ulla Bonas. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959):1509--1512, 2009.

John L Bowman, John Alvarez, Detlef Weigel, Elliot M Meyerowitz, and David R Smyth. Control of flower development in *Arabidopsis thaliana* by APETALA 1 and interacting genes. *Development*, 119:721--721, 1993.

John L Bowman, David R Smyth, and Elliot M Meyerowitz. The ABC model

of flower development: then and now. *Development*, 139(22):4095--4098, 2012.

Desmond Bradley, Oliver Ratcliffe, Coral Vincent, Rosemary Carpenter, and Enrico Coen. Inflorescence commitment and architecture in Arabidopsis. *Science*, 275(5296):80--83, 1997.

Ulrike Brand, Jennifer C Fletcher, Martin Hobe, Elliot M Meyerowitz, and Rüdiger Simon. Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity. *Science*, 289(5479):617--619, 2000.

Georgette C Briggs, Karen S Osmont, Chikako Shindo, Richard Sibout, and Christian S Hardtke. Unequal genetic redundancies in *Arabidopsis*--a neglected phenomenon? *Trends in plant science*, 11(10):492--498, 2006.

Stan JJ Brouns, Matthijs M Jore, Magnus Lundgren, Edze R Westra, Rik JH Slijkhuis, Ambrosius PL Snijders, Mark J Dickman, Kira S Makarova, Eugene V Koonin, and John Van Der Oost. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321(5891):960--964, 2008.

Etienne Bucher, Dick Lohuis, Pieter MJA van Poppel, Christina Geerts-Dimitriadou, Rob Goldbach, and Marcel Prins. Multiple virus resistance at a high frequency using a single transgene construct. *Journal of General Virology*, 87(12):3697--3701, 2006.

Wolfgang Busch, Andrej Miotk, Federico D Ariel, Zhong Zhao, Joachim Forner, Gabor Daum, Takuya Suzaki, Christoph Schuster, Sebastian J Schultheiss, Andrea Leibfried, et al. Transcriptional control of a plant stem cell niche. *Developmental cell*, 18(5):841--853, 2010.

Lorenzo Carretero-Paulet, Anahit Galstyan, Irma Roig-Villanova, Jaime F Martínez-García, Jose R Bilbao-Castro, and David L Robertson. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant physiology*, 153(3):1398--1412, 2010.

Tomas Cermak, Erin L Doyle, Michelle Christian, Li Wang, Yong Zhang, Clarice Schmidt, Joshua A Baller, Nikunj V Somia, Adam J Bogdanove, and Daniel F Voytas. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research*, page gkr218, 2011.

Caren Chang, G Eric Schaller, Sara E Patterson, Shing F Kwok, Elliot M Meyerowitz, and Anthony B Bleecker. The TMK1 gene from Arabidopsis codes for a protein with structural and biochemical characteristics of a receptor protein kinase. *The Plant Cell Online*, 4(10):1263--1271, 1992.

Steven E Clark, Steven E Jacobsen, Joshua Z Levin, and Elliot M Meyerowitz. The CLAVATA and SHOOT MERISTEMLESS loci competitively regulate meristem activity in Arabidopsis. *Development*, 122(5):1567--1575, 1996.

Enrico S Coen and Elliot M Meyerowitz. The war of the whorls: genetic interactions controlling flower development. *nature*, 353(6339):31--37, 1991.

Le Cong, F Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D Hsu, Xuebing Wu, Wenyan Jiang, Luciano A Marraffini, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121):819--823, 2013.

David J Craigon, Nick James, John Okyere, Janet Higgins, Joan Jotham, and Sean May. NASCArrays: a repository for microarray data generated by NASCs transcriptomics service. *Nucleic acids research*, 32(suppl 1): D575--D577, 2004.

Francis Crick et al. Central dogma of molecular biology. *Nature*, 227(5258): 561--563, 1970.

Mark D Curtis and Ueli Grossniklaus. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant physiology*, 133(2):462--469, 2003.

Ramana V Davuluri, Hao Sun, Saranyan K Palaniswamy, Nicole Matthews, Carlos Molina, Mike Kurtz, and Erich Grotewold. AGRIS: Arabidopsis

gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. *BMC bioinformatics*, 4 (1):25, 2003.

Robert C Day, Ueli Grossniklaus, and Richard C Macknight. Be more specific! Laser-assisted microdissection of plant cells. *Trends in plant science*, 10 (8):397--406, 2005.

Marc De Block, J Botterman, M Vandewiele, J Dockx, C Thoen, V Gossele, N Rao Movva, C Thompson, Marc Van Montagu, and Jan Leemans. Engineering herbicide resistance in plants by expression of a detoxifying enzyme. *The EMBO journal*, 6(9):2513, 1987.

Gary N Drews, John L Bowman, and Elliot M Meyerowitz. Negative regulation of the Arabidopsis homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell*, 65(6):991--1002, 1991.

Gerald M Edelman and Joseph A Gally. Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences*, 98 (24):13763--13768, 2001.

Carola Engler, Romy Kandzia, and Sylvestre Marillonnet. A one pot, one step, precision cloning method with high throughput capability. *PloS one*, 3(11):e3647, 2008.

Inés Ezcurra, Paul Wycliffe, Lilian Nehlin, Mats Ellerström, and Lars Rask. Transactivation of the Brassica napus napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different cis-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. *The Plant Journal*, 24(1):57--66, 2000.

Friedrich Fauser, Simon Schiml, and Holger Puchta. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. *The Plant Journal*, 2014.

Antje Feller, Katja Machemer, Edward L Braun, and Erich Grotewold. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant Journal*, 66(1):94--116, 2011.

Zhengyan Feng, Yanfei Mao, Nanfei Xu, Botao Zhang, Pengliang Wei, Dong-Lei Yang, Zhen Wang, Zhengjing Zhang, Rui Zheng, Lan Yang, et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proceedings of the National Academy of Sciences*, 111(12):4632--4637, 2014.

Cristina Ferrándiz, Qing Gu, Robert Martienssen, and Martin F Yanofsky. Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development*, 127(4): 725--734, 2000.

Andrew Fire, SiQun Xu, Mary K Montgomery, Steven A Kostas, Samuel E Driver, and Craig C Mello. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *nature*, 391(6669):806--811, 1998.

Jennifer C Fletcher, Ulrike Brand, Mark P Running, Rüdiger Simon, and Elliot M Meyerowitz. Signaling of cell fate decisions by CLAVATA3 in Arabidopsis shoot meristems. *Science*, 283(5409):1911--1914, 1999.

José M Franco-Zorrilla, Pilar Cubas, José A Jarillo, Begoña Fernández-Calvın, Julio Salinas, and José M Martınez-Zapater. AtREM1, a member of a new family of B3 domain-containing genes, is preferentially expressed in reproductive meristems. *Plant physiology*, 128(2):418--427, 2002.

Thomas Gaj, Charles A Gersbach, and Carlos F Barbas III. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology*, 31(7):397--405, 2013.

Koji Goto and Elliot M Meyerowitz. Function and regulation of the Arabidopsis floral homeotic gene PISTILLATA. *Genes & Development*, 8(13): 1548--1560, 1994.

Valentina Grandi, Veronica Gregis, and Martin M Kater. Uncovering genetic and molecular interactions among floral meristem identity genes in Arabidopsis thaliana. *The Plant Journal*, 69(5):881--893, 2012.

Veronica Gregis, Alice Sessa, Lucia Colombo, and Martin M Kater. AGL24, SHORT VEGETATIVE PHASE, and APETALA1 redundantly control AGAMOUS during early stages of flower development in Arabidopsis. *The Plant Cell Online*, 18(6):1373--1382, 2006.

Veronica Gregis, Alice Sessa, Lucia Colombo, and Martin M Kater. AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis. *The Plant Journal*, 56(6):891--902, 2008.

Veronica Gregis, Alice Sessa, Carmen Dorca-Fornell, and Martin M Kater. The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes. *The Plant Journal*, 60 (4):626--637, 2009.

Veronica Gregis, Fernando Andrés, Alice Sessa, Rosalinda F Guerra, Sara Simonini, Julieta L Mateos, Stefano Torti, Federico Zambelli, Gian Marco Prazzoli, Katrine N Bjerkan, et al. Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis. *Genome biology*, 14(6):R56, 2013.

Qing Gu, Cristina Ferrándiz, Martin F Yanofsky, and Robert Martienssen. The FRUITFULL MADS-box gene mediates cell differentiation during Arabidopsis fruit development. *Development*, 125(8):1509--1517, 1998.

Yann Guédon, Yassin Refahi, Fabrice Besnard, Etienne Farcot, Christophe Godin, and Teva Vernoux. Pattern identification and characterization reveal permutations of organs as a key genetically controlled property of post-meristematic phyllotaxis. *Journal of theoretical biology*, 338:94--110, 2013.

Yongfeng Guo and Steven E Clark. Membrane distributions of two ligand-binding receptor complexes in the CLAVATA pathway. *Plant signaling & behavior*, 5(11):1442, 2010.

Ulrike Hartmann, Susanne Höhmann, Klaus Nettesheim, Ellen Wisman, Heinz Saedler, and Peter Huijser. Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. *The Plant Journal*, 21(4):351--360, 2000.

Felix Hauser, Wenxiao Chen, Ulrich Deinlein, Kenneth Chang, Stephan Ossowski, Joffrey Fitz, Gregory J Hannon, and Julian I Schroeder. A genomic-scale artificial microRNA library as a tool to investigate the functionally redundant gene space in Arabidopsis. *The Plant Cell Online*, 25(8):2848--2863, 2013.

Qiye He, Anaïs F Bardet, Brianne Patton, Jennifer Purvis, Jeff Johnston, Ariel Paulson, Madelaine Gogol, Alexander Stark, and Julia Zeitlinger. High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. *Nature genetics*, 43(5):414--420, 2011.

Richard GH Immink, David Posé, Silvia Ferrario, Felix Ott, Kerstin Kaufmann, Felipe Leal Valentim, Stefan De Folter, Froukje Van der Wal, Aalt DJ van Dijk, Markus Schmid, et al. Characterization of SOC1s central role in flowering by the identification of its upstream and downstream regulators. *Plant physiology*, 160(1):433--449, 2012.

Vivian F Irish and Ian M Sussex. Function of the apetala-1 gene during Arabidopsis floral development. *The Plant Cell Online*, 2(8):741--753, 1990.

Thomas Jack, Laura L Brockman, and Elliot M Meyerowitz. The homeotic gene *APETALA3* of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. *Cell*, 68(4):683--697, 1992.

David S Johnson, Ali Mortazavi, Richard M Myers, and Barbara Wold. Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 316 (5830):1497--1502, 2007.

Yasuaki Kagaya, Kunio Ohmiya, and Tsukaho Hattori. RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two

distinct DNA-binding domains uniquely found in higher plants. *Nucleic acids research*, 27(2):470--478, 1999.

Kerstin Kaufmann, Jose M Muino, Ruy Jauregui, Chiara A Airoldi, Cezary Smaczniak, Pawel Krajewski, and Gerco C Angenent. Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS biology*, 7(4): e1000090, 2009.

Kerstin Kaufmann, Frank Wellmer, Jose M. Muiño, Thilia Ferrier, Samuel E. Wuest, Vijaya Kumar, Antonio Serrano-Mislata, Francisco Madueño, Pawel Krajewski, Elliot M. Meyerowitz, Gerco C. Angenent, and José Luis Riechmann. Orchestration of Floral Initiation by APETALA1. *Science*, 328 (5974):85--89, 2010.

Jin Hee Kim, Sang-Rok Lee, Li-Hua Li, Hye-Jeong Park, Jeong-Hoh Park, Kwang Youl Lee, Myeong-Kyu Kim, Boo Ahn Shin, and Seok-Yong Choi. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PloS one*, 6(4):e18556, 2011.

Yang-Gyun Kim, Jooyeun Cha, and Srinivasan Chandrasegaran. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, 93(3):1156--1160, 1996.

Gordon J King, Aurélie H Chanson, Emily J McCallum, Masaru Ohme-Takagi, Karl Byriel, Justine M Hill, Jennifer L Martin, and Joshua S Mylne. The Arabidopsis B3 Domain Protein VERNALIZATION1 (VRN1) Is Involved in Processes Essential for Development, with Structural and Mutational Studies Revealing Its DNA-binding Surface. *Journal of Biological Chemistry*, 288(5):3198--3207, 2013.

Beth A Krizek and Jennifer C Fletcher. Molecular mechanisms of flower development: an armchair guide. *Nature Reviews Genetics*, 6(9):688--698, 2005.

Ravi Kumar, Kumuda Kushalappa, Dietmute Godt, Mark S Pidkowich, Sandro Pastorelli, Shelley R Hepworth, and George W Haughn. The

Arabidopsis BEL1-LIKE HOMEODOMAIN proteins SAW1 and SAW2 act redundantly to regulate KNOX expression spatially in leaf margins. *The Plant Cell Online*, 19(9):2719--2735, 2007.

Thomas Laux, KF Mayer, Jürgen Berger, and G Jurgens. The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. *Development*, 122(1):87--96, 1996.

Jungeun Lee, Mijin Oh, Hanna Park, and Ilha Lee. SOC1 translocated to the nucleus by interaction with AGL24 directly regulates LEAFY. *The Plant Journal*, 55(5):832--843, 2008.

Yaron Y Levy, Stéphane Mesnage, Joshua S Mylne, Anthony R Gendall, and Caroline Dean. Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. *Science*, 297(5579):243--246, 2002.

Dan Li, Chang Liu, Lisha Shen, Yang Wu, Hongyan Chen, Masumi Robertson, Chris A Helliwell, Toshiro Ito, Elliot Meyerowitz, and Hao Yu. A repressor complex governs the integration of flowering signals in *Arabidopsis*. *Developmental cell*, 15(1):110--120, 2008.

Johnny Lloyd and David Meinke. A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. *Plant physiology*, 158 (3):1115--1129, 2012.

Jan U Lohmann, Ray L Hong, Martin Hobe, Maximilian A Busch, François Parcy, Rüdiger Simon, and Detlef Weigel. A Molecular Link between Stem Cell Regulation and Floral Patterning in *Arabidopsis*. *Cell*, 105(6):793--803, 2001.

Ditte Lovatt, Brittani K Ruble, Jaehee Lee, Hannah Dueck, Tae Kyung Kim, Stephen Fisher, Chantal Francis, Jennifer M Spaethling, John A Wolf, M Sean Grady, et al. Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nature methods*, 2014.

Morgan L Maeder, Stacey Thibodeau-Beganny, Jeffry D Sander, Daniel F Voytas, and J Keith Joung. Oligomerized pool engineering (OPEN): an'open-

source'protocol for making customized zinc-finger arrays. *Nature protocols*, 4(10):1471--1501, 2009.

M Alejandra Mandel and Martin F Yanofsky. The Arabidopsis AGL8 MADS box gene is expressed in inflorescence meristems and is negatively regulated by APETALA1. *The Plant Cell Online*, 7(11):1763--1771, 1995.

M Alejandra Mandel, Cindy Gustafson-Brown, Beth Savidge, and Martin F Yanofsky. Molecular characterization of the Arabidopsis floral homeotic gene APETALA1. 1992.

MA Mandel and Martin F Yanofsky. The Arabidopsis AGL9 MADS box gene is expressed in young flower primordia. *Sexual Plant Reproduction*, 11(1):22--28, 1998.

Otho Mantegazza, Veronica Gregis, Matteo Chiara, Caterina Selva, Giulia Leo, David S Horner, and Martin M Kater. Gene coexpression patterns during early development of the native Arabidopsis reproductive meristem: novel candidate developmental regulators and patterns of functional redundancy. *The Plant Journal*, 79(5):861--877, 2014a.

Otho Mantegazza, Veronica Gregis, Marta Adelina Mendes, Piero Morandini, Márcio Alves-Ferreira, Camila M Patreze, Sarah M Nardeli, Martin M Kater, and Lucia Colombo. Analysis of the arabidopsis REM gene family predicts functions during flower development. *Annals of botany*, page mcu124, 2014b.

Anthony Mathelier and Wyeth W Wasserman. The next generation of transcription factor binding site prediction. *PLoS computational biology*, 9 (9):e1003214, 2013.

Luis Matias-Hernandez, Raffaella Battaglia, Francesca Galbiati, Marco Rubes, Christof Eichenberger, Ueli Grossniklaus, Martin M Kater, and Lucia Colombo. VERDANDI is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in Arabidopsis. *The Plant Cell Online*, 22(6):1702--1715, 2010.

Klaus FX Mayer, Heiko Schoof, Achim Haecker, Michael Lenhard, Gerd Jürgens, and Thomas Laux. Role of *WUSCHEL* in Regulating Stem Cell Fate in the *Arabidopsis* Shoot Meristem. *Cell*, 95(6):805--815, 1998.

Donald R McCarty, Tsukaho Hattori, Christian B Carson, Vimla Vasil, Mark Lazar, and Indra K Vasil. The *VIVIPAROUS-1* developmental gene of maize encodes a novel transcriptional activator. *Cell*, 66(5):895--905, 1991.

Rainer Melzer and Günter Theißen. Reconstitution of floral quartets in vitro involving class B and class E floral homeotic proteins. *Nucleic acids research*, 37(8):2723--2736, 2009.

Eric M Mendenhall, Kaylyn E Williamson, Deepak Reyon, James Y Zou, Oren Ram, J Keith Joung, and Bradley E Bernstein. Locus-specific editing of histone modifications at endogenous enhancers. *Nature biotechnology*, 2013.

Marta Adelina Mendes, Rosalinda Fiorella Guerra, Markus Christian Berns, Carlo Manzo, Simona Masiero, Laura Finzi, Martin M Kater, and Lucia Colombo. MADS domain transcription factors mediate short-range dna looping that is essential for target gene expression in Arabidopsis. *The Plant Cell Online*, 25(7):2560--2572, 2013.

Margit Menges, Giulio Pavesi, Piero Morandini, Laszlo Bögre, and James AH Murray. Genomic organization and evolutionary conservation of plant D-type cyclins. *Plant physiology*, 145(4):1558--1576, 2007.

Margit Menges, Róbert Dóczi, László Ökrész, Piero Morandini, Luca Mizzi, Mikhail Soloviev, James AH Murray, and László Bögre. Comprehensive gene expression atlas for the Arabidopsis MAP kinase signalling pathways. *New Phytologist*, 179(3):643--662, 2008.

Jin Miao, Dongshu Guo, Jinzhe Zhang, Qingpei Huang, Genji Qin, Xin Zhang, Jianmin Wan, Hongya Gu, and Li-Jia Qu. Targeted mutagenesis in rice using CRISPR-Cas system. *Cell research*, 23(10):1233, 2013.

Daisuke Miki, Rika Itoh, and Ko Shimamoto. RNA silencing of single and multiple members in a gene family of rice. *Plant Physiology*, 138(4):1903--1913, 2005.

Jeffrey C Miller, Siyuan Tan, Guijuan Qiao, Kyle A Barlow, Jianbin Wang, Danny F Xia, Xiangdong Meng, David E Paschon, Elo Leung, Sarah J Hinkley, et al. A TALE nuclease architecture for efficient genome editing. *Nature biotechnology*, 29(2):143--148, 2011.

Jeffery D Molkentin and Eric N Olson. Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. *Proceedings of the National Academy of Sciences*, 93(18):9366--9373, 1996.

Edwige Moyroud, Eugenio Gómez Minguet, Felix Ott, Levi Yant, David Posé, Marie Monniaux, Sandrine Blanchet, Olivier Bastien, Emmanuel Thévenon, Detlef Weigel, et al. Prediction of regulatory interactions from genome sequences using a biophysical model for the Arabidopsis LEAFY transcription factor. *The Plant Cell Online*, 23(4):1293--1306, 2011.

Kazuo Nakashima, Hironori Takasaki, Junya Mizoi, Kazuo Shinozaki, and Kazuko Yamaguchi-Shinozaki. NAC transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1819(2):97--103, 2012.

Carolyn Napoli, Christine Lemieux, and Richard Jorgensen. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *The Plant Cell Online*, 2(4): 279--289, 1990.

R Offringa, MJ De Groot, HJ Haagsman, MP Does, PJ Van Den Elzen, and PJ Hooykaas. Extrachromosomal homologous recombination and gene targeting in plant cells after Agrobacterium mediated transformation. *The EMBO journal*, 9(10):3077, 1990.

Diarmuid S ÓMaoiléidigh, Samuel E Wuest, Liina Rae, Andrea Raganelli, Patrick T Ryan, Kamila Kwaśniewska, Pradeep Das, Amanda J Lohan,

Brendan Loftus, Emmanuelle Graciet, et al. Control of reproductive floral organ identity specification in Arabidopsis by the C function regulator AGAMOUS. *The Plant Cell Online*, 25(7):2482--2503, 2013.

Diarmuid Seosamh Ó'Maoiléidigh, Emmanuelle Graciet, and Frank Wellmer. Gene networks controlling Arabidopsis thaliana flower development. *New Phytologist*, 201(1):16--30, 2014.

Alice Pajoro, Pedro Madrigal, Jose M Muiño, José Tomás Matus, Jian Jin, Martin A Mecchia, Juan M Debernardi, Javier F Palatnik, Salma Balazadeh, Muhammad Arif, et al. Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome biology*, 15(3):R41, 2014.

Lucie Parenicová, Stefan de Folter, Martin Kieffer, David S Horner, Cristina Favalli, Jacqueline Busscher, Holly E Cook, Richard M Ingram, Martin M Kater, Brendan Davies, et al. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis new openings to the MADS world. *The Plant Cell Online*, 15(7):1538--1551, 2003.

Ranian J Perera, Christian G Linard, and Ethan R Signer. Cytosine deaminase as a negative selective marker for Arabidopsis. *Plant molecular biology*, 23 (4):793--799, 1993.

Holger Puchta and Friedrich Fauser. Gene targeting in plants: 25 years later. *Int. J. Dev. Biol*, 57:629--637, 2013.

Holger Puchta, Bernard Dujon, and Barbara Hohn. Homologous recombination in plant cells is enhanced by in vivo induction of double strand breaks into DNA by a site-specific endonuclease. *Nucleic acids research*, 21(22): 5034--5040, 1993.

Holger Puchta, Bernard Dujon, and Barbara Hohn. Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proceedings of the National Academy of Sciences*, 93(10):5055--5060, 1996.

Cherie L Ramirez, Jonathan E Foley, David A Wright, Felix Müller-Lerch, Shamim H Rahman, Tatjana I Cornu, Ronnie J Winfrey, Jeffry D Sander, Fengli Fu, Jeffrey A Townsend, et al. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature methods*, 5(5):374--375, 2008.

Cherie L Ramirez, Michael T Certo, Claudio Mussolino, Mathew J Goodwin, Thomas J Cradick, Anton P McCaffrey, Toni Cathomen, Andrew M Scharenberg, and J Keith Joung. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic acids research*, 40(12):5560--5568, 2012.

Oliver J Ratcliffe, Desmond J Bradley, and Enrico S Coen. Separation of shoot and floral identity in Arabidopsis. *Development*, 126(6):1109--1120, 1999.

Elisson Romanel, Pradeep Das, Richard M Amasino, Jan Traas, Elliot Meyerowitz, and Marcio Alves-Ferreira. Reproductive Meristem22 is a unique marker for the early stages of stamen development. *International Journal of Developmental Biology*, 55(6):657, 2011.

Elisson AC Romanel, Carlos G Schrago, Rafael M Couñago, Claudia AM Russo, and Márcio Alves-Ferreira. Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. *PloS one*, 4(6):e5791, 2009.

Laura E Rosen, Holly A Morrison, Selma Masri, Michael J Brown, Brendan Springstubb, Django Sussman, Barry L Stoddard, and Lenny M Seligman. Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic acids research*, 34(17):4791--4800, 2006.

Paul J Rushton, Imre E Somssich, Patricia Ringler, and Qingxi J Shen. WRKY transcription factors. *Trends in plant science*, 15(5):247--258, 2010.

Louis A Saddic, Bärbel Huvermann, Staver Bezhani, Yanhui Su, Cara M Winter, Chang Seob Kwon, Richard P Collum, and Doris Wagner. The LEAFY target LMI1 is a meristem identity regulator and acts together

with LEAFY to regulate expression of CAULIFLOWER. *Development*, 133(9):1673--1682, 2006.

Alon Samach, Hitoshi Onouchi, Scott E Gold, Gary S Ditta, Zsuzsanna Schwarz-Sommer, Martin F Yanofsky, and George Coupland. Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. *Science*, 288(5471):1613--1616, 2000.

Alejandro Sarrion-Perdigones, Erica Elvira Falconi, Sara I Zandalinas, Paloma Juárez, Asun Fernández-del Carmen, Antonio Granell, and Diego Orzaez. GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PloS one*, 6(7):e21622, 2011.

Uwe Sauer, Matthias Heinemann, and Nicola Zamboni. Getting closer to the whole picture. *Science(Washington)*, 316(5824):550--551, 2007.

Anne Saumet and Charles-Henri Lecellier. Anti-viral RNA silencing: do we look like plants? *Retrovirology*, 3(1):3, 2006.

Marc W Schmid, Anja Schmidt, Ulrich C Klostermeier, Matthias Barann, Philip Rosenstiel, and Ueli Grossniklaus. A powerful method for transcriptional profiling of specific cell types in eukaryotes: laser-assisted microdissection and RNA sequencing. *PLoS One*, 7(1):e29685, 2012.

Markus Schmid, Timothy S Davison, Stefan R Henz, Utz J Pape, Monika Demar, Martin Vingron, Bernhard Schölkopf, Detlef Weigel, and Jan U Lohmann. A gene expression map of Arabidopsis thaliana development. *Nature genetics*, 37(5):501--506, 2005.

Anja Schmidt, Marc W Schmid, and Ueli Grossniklaus. Analysis of plant germline development by high-throughput RNA profiling: technical advances and new insights. *The Plant Journal*, 70(1):18--29, 2012.

Andreas Schroeder, Odilo Mueller, Susanne Stocker, Ruediger Salowsky, Michael Leiber, Marcus Gassmann, Samar Lightfoot, Wolfram Menzel, Martin Granzow, and Thomas Ragg. The RIN: an RNA integrity number

for assigning integrity values to RNA measurements. *BMC molecular biology*, 7(1):3, 2006.

Cezary Smaczniak, Richard GH Immink, Jose M Muiño, Robert Blanvillain, Marco Busscher, Jacqueline Busscher-Lange, QD Peter Dinh, Shujing Liu, Adrie H Westphal, Sjef Boeren, et al. Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proceedings of the National Academy of Sciences*, 109(5):1560--1565, 2012.

Nils Ole Steffens, Claudia Galuschka, Martin Schindler, Lorenz Bülow, and Reinhard Hehl. AthaMap: an online resource for in silico transcription factor binding sites in the Arabidopsis thaliana genome. *Nucleic acids research*, 32(suppl 1):D368--D372, 2004.

Jai-Yoon Sul, K Wu Chia-wen, Fanyi Zeng, Jeanine Jochems, Miler T Lee, Tae Kyung Kim, Tiina Peritz, Peter Buckley, David J Cappelleri, Margaret Maronski, et al. Transcriptome transfer produces a predictable cellular phenotype. *Proceedings of the National Academy of Sciences*, 106(18): 7624--7629, 2009.

Masaharu Suzuki, Chien Yuan Kao, and Donald R McCarty. The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. *The Plant Cell Online*, 9(5):799--807, 1997.

Kankshita Swaminathan, Kevin Peterson, and Thomas Jack. The plant B3 superfamily. *Trends in plant science*, 13(12):647--655, 2008.

Kirk R Thomas and Mario R Capecchi. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, 51(3):503--512, 1987.

Stefano Torti, Fabio Fornara, Coral Vincent, Fernando Andrés, Karl Nordström, Ulrike Göbel, Daniela Knoll, Heiko Schoof, and George Coupland. Analysis of the Arabidopsis shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. *The Plant Cell Online*, 24(2):444--462, 2012.

Jeffrey A Townsend, David A Wright, Ronnie J Winfrey, Fengli Fu, Morgan L Maeder, J Keith Joung, and Daniel F Voytas. High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature*, 459(7245): 442--445, 2009.

Cole Trapnell, Brian A Williams, Geo Pertea, Ali Mortazavi, Gordon Kwan, Marijke J van Baren, Steven L Salzberg, Barbara J Wold, and Lior Pachter. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology*, 28(5):511--515, 2010.

Cole Trapnell, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn, and Lior Pabchter. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7(3):562--578, 2012.

Tim Ulmasov, Jane Murfett, Gretchen Hagen, and Tom J Guilfoyle. Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *The Plant Cell Online*, 9(11): 1963--1971, 1997.

Eric van der Graaff, Thomas Laux, and Stefan A Rensing. The WUS homeobox-containing (WOX) protein family. *Genome Biol*, 10(12):248, 2009.

Fangfang Wang and Sharyn E Perry. Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. *Plant physiology*, 161(3):1251--1264, 2013.

Yanpeng Wang, Xi Cheng, Qiwei Shan, Yi Zhang, Jinxing Liu, Caixia Gao, and Jin-Long Qiu. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature biotechnology*, 32(9):947--951, 2014.

Yulei Wang, Catalin Barbacioru, Fiona Hyland, Wenming Xiao, Kathryn L Hunkapiller, Julie Blake, Frances Chan, Carolyn Gonzalez, Lu Zhang, and

Raymond R Samaha. Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays. *BMC genomics*, 7(1):59, 2006.

Detlef Weigel, John Alvarez, David R Smyth, Martin F Yanofsky, and Elliot M Meyerowitz. *LEAFY* controls floral meristem identity in Arabidopsis. *Cell*, 69(5):843--859, 1992.

Frank Wellmer, Márcio Alves-Ferreira, Annick Dubois, José Luis Riechmann, and Elliot M Meyerowitz. Genome-wide analysis of gene expression during early Arabidopsis flower development. *PLoS genetics*, 2(7):e117, 2006.

Bregje Wertheim. Beyond the Gene List: Exploring Transcriptomics Data in Search for Gene Function, Trait Mechanisms and Genetic Architecture. 2012.

S Varsha Wesley, Christopher A Helliwell, Neil A Smith, MingBo Wang, Dean T Rouse, Qing Liu, Paul S Gooding, Surinder P Singh, David Abbott, Peter A Stoutjesdijk, et al. Construct design for efficient, effective and high-throughput gene silencing in plants. *The Plant Journal*, 27(6):581--590, 2001.

James Whitacre and Axel Bender. Degeneracy: a design principle for achieving robustness and evolvability. *Journal of Theoretical Biology*, 263(1):143--153, 2010.

Cara M Winter, Ryan S Austin, Servane Blanvillain-Baufume, Maxwell A Reback, Marie Monniaux, Miin-Feng Wu, Yi Sang, Ayako Yamaguchi, Nobutoshi Yamaguchi, Jane E Parker, et al. LEAFY Target Genes Reveal Floral Regulatory Logic, *cis* Motifs, and a Link to Biotic Stimulus Response. *Developmental cell*, 20(4):430--443, 2011.

Samuel E Wuest, Diarmuid S OMaoileidigh, Liina Rae, Kamila Kwasniewska, Andrea Raganelli, Katarzyna Hanczaryk, Amanda J Lohan, Brendan Loftus, Emmanuelle Graciet, and Frank Wellmer. Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. *Proceedings of the National Academy of Sciences*, 109(33):13452--13457, 2012.

An Xiao, Zhanxiang Wang, Yingying Hu, Yingdan Wu, Zhou Luo, Zhipeng Yang, Yao Zu, Wenyuan Li, Peng Huang, Xiangjun Tong, et al. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic acids research*, page gkt464, 2013.

Ya Xue, Xuejun Liao, Lawrence Carin, and Balaji Krishnapuram. Multi-task learning for classification with Dirichlet process priors. *The Journal of Machine Learning Research*, 8:35--63, 2007.

Chen Yanhui, Yang Xiaoyuan, He Kun, Liu Meihua, Li Jigang, Gao Zhaofeng, Lin Zhiqiang, Zhang Yunfei, Wang Xiaoxiao, Qiu Xiaoming, et al. The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. *Plant molecular biology*, 60(1):107--124, 2006.

Levi Yant, Johannes Mathieu, Thanh Theresa Dinh, Felix Ott, Christa Lanz, Heike Wollmann, Xuemei Chen, and Markus Schmid. Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2. *The Plant Cell Online*, 22(7):2156--2170, 2010.

Feng Zhang, Morgan L Maeder, Erica Unger-Wallace, Justin P Hoshaw, Deepak Reyon, Michelle Christian, Xiaohong Li, Christopher J Pierick, Drena Dobbs, Thomas Peterson, et al. High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. *Proceedings of the National Academy of Sciences*, 107(26):12028--12033, 2010.

Yumei Zheng, Na Ren, Huai Wang, Arnold J Stromberg, and Sharyn E Perry. Global identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like15. *The Plant Cell Online*, 21(9):2563--2577, 2009.

Min Zou and Suzanne D Conzen. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, 21(1):71--79, 2005.

Chapter 5

Acknowledgments

All this work has been carried out under the supervision and with the collaboration of Dr. Veronica Gregis. The studies on REM genes have been carried out in collaboration with Prof. Lucia Colombo and Dr. Marta Miranda Mendes. A huge amount work on REM genes, yet unpublished, have been done by Francesco Gozzo, Francesca Caselli and Francesca Milani, The hybidization on REM genes, the hybridization on LM candidate genes, the large mutant screening and the cloning for the multiple RNAi and TALEN genome editing has been carried out by Caterina Selva and Giulia Leo. The GoldenBraid vector set have been provided by Alejandro Sarrion Perdigones and Diego Orzaez, the Golden Gate TALEN and TAL Effector Kit 2.0 was provided by Daniel Voytas through Addgene. I personally thank Dr. Simona Masiero for the support and constructive discussion during my work.

Part II

Papers published in peer reviewed journals

Chapter 6

Gene coexpression patterns during early development of the native Arabidopsis reproductive meristem: novel candidate developmental regulators and patterns of functional redundancy

The article ‘‘Gene coexpression patterns during early development of the native Arabidopsis reproductive meristem: novel candidate developmental regulators and patterns of functional redundancy’’ is available at the publisher's web site at the following link: `http://onlinelibrary.wiley.com/doi/10.1111/tpj.12585/abstract`

Chapter 7

Analysis of the arabidopsis REM gene family predicts functions during flower development

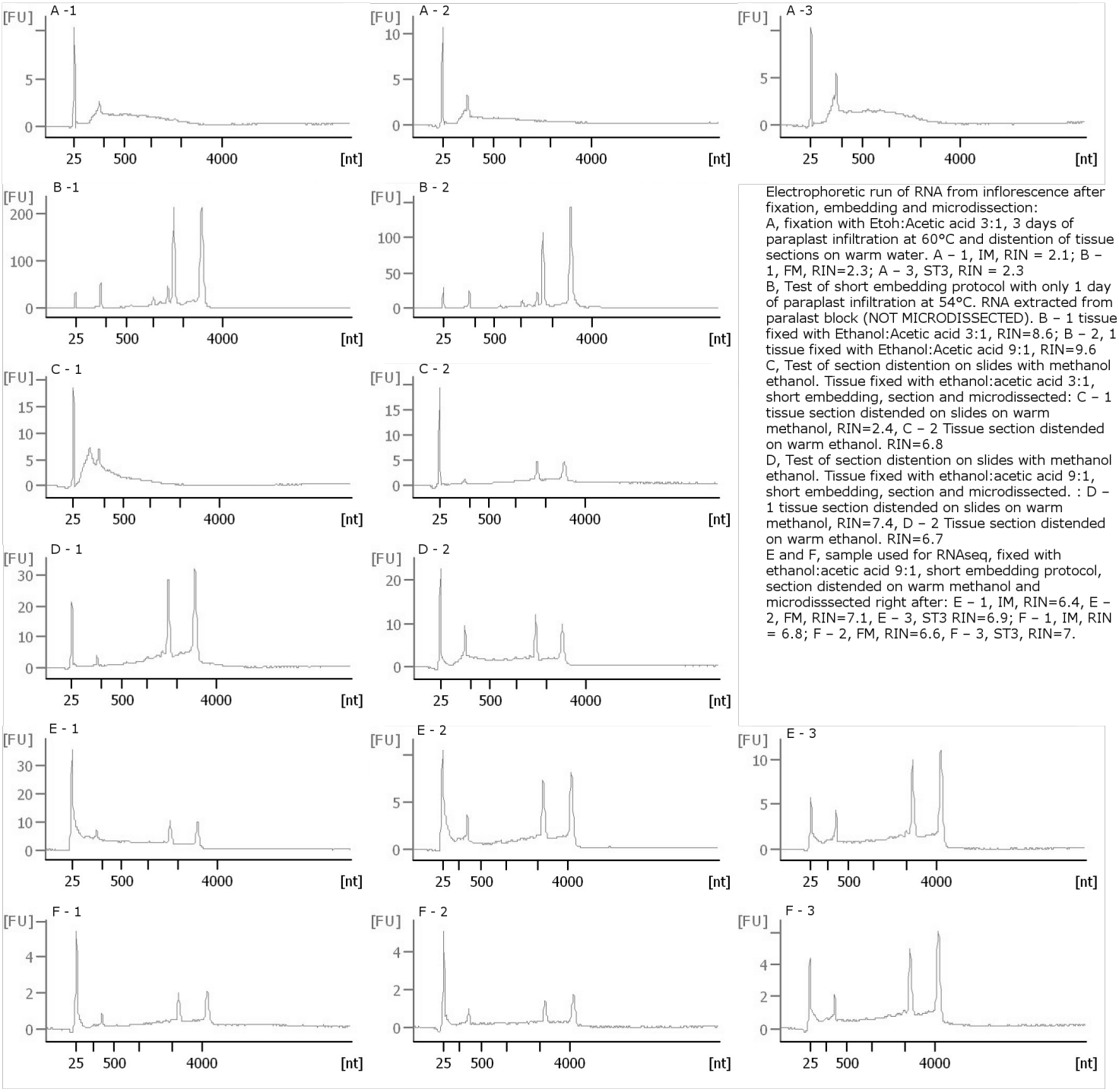
The article ‘‘Analysis of the arabidopsis REM gene family predicts functions during flower development’’ is available at the publisher’s web site at the following link: `http://aob.oxfordjournals.org/content/114/7/1507.abstract`

Part III

Supplementary Material

Chapter 8

Supplementary material to transcriptome of early flower development



Electrophoretic run of RNA from inflorescence after fixation, embedding and microdissection:
A, fixation with Etoh:Acetic acid 3:1, 3 days of paraplast infiltration at 60°C and distention of tissue sections on warm water. A – 1, IM, RIN = 2.1; B – 1, FM, RIN=2.3; A – 3, ST3, RIN = 2.3
B, Test of short embedding protocol with only 1 day of paraplast infiltration at 54°C. RNA extracted from paralast block (NOT MICRODISSECTED). B – 1 tissue fixed with Ethanol:Acetic acid 3:1, RIN=8.6; B – 2, 1 tissue fixed with Ethanol:Acetic acid 9:1, RIN=9.6
C, Test of section distention on slides with methanol ethanol. Tissue fixed with ethanol:acetic acid 3:1, short embedding, section and microdissected: C – 1 tissue section distended on slides on warm methanol, RIN=2.4, C – 2 Tissue section distended on warm ethanol. RIN=6.8
D, Test of section distention on slides with methanol ethanol. Tissue fixed with ethanol:acetic acid 9:1, short embedding, section and microdissected. : D – 1 tissue section distended on slides on warm methanol, RIN=7.4, D – 2 Tissue section distended on warm ethanol. RIN=6.7
E and F, sample used for RNAseq, fixed with ethanol:acetic acid 9:1, short embedding protocol, section distended on warm methanol and microdisssected right after: E – 1, IM, RIN=6.4, E – 2, FM, RIN=7.1, E – 3, ST3 RIN=6.9; F – 1, IM, RIN = 6.8; F – 2, FM, RIN=6.6, F – 3, ST3, RIN=7.

8.2 Intersection between the background and genic distribution

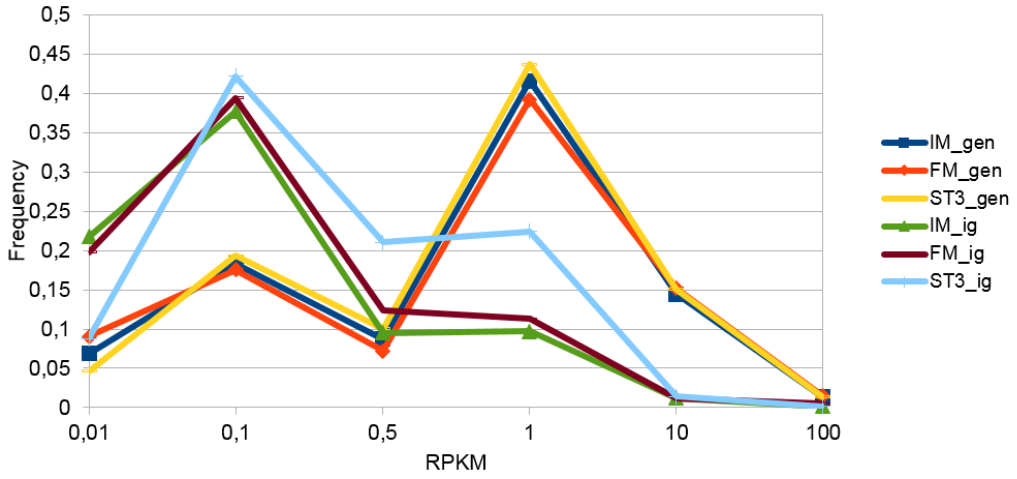


Figure 8.1: RPKM values were calculated for genes and intergenic background regions. The intergenic regions were matched to have the same length distribution as the genes and no mapping ESTs. For each tissue we binned the expression levels of all the genes and background regions (bins: RPKM¡=0.01, 0.1, 0.5, 1, 100) and computed the cumulative distribution. (RPKM¡=0.5) was arbitrarily chosen as the cutoff value for gene expression

8.3 Differentially Expressed Genes

Lists of genes differentially expressed in in all the pairwise comparisons among inflorescence meristems, floral meristems and flowers at developmental stage three, with False Discovery Rate smaller than 0.05.

8.3.1 IM vs. FM

TAIR-ID	AT1G69120	AT2G02100	AT1G62500	AT5G62230
AT1G62480	AT4G15620	AT1G71695	AT3G16400	AT4G15450
AT4G31910	AT4G15630	AT3G54400	AT5G15780	AT4G35900

AT1G21250	AT1G76110	AT5G54510	AT2G45660	AT5G28490
AT1G30950	AT2G41820	AT1G66970	AT1G43910	AT3G05650
AT5G60910	AT4G27460	AT5G49730	AT5G15960	AT5G03840
AT3G25020	AT5G49700	AT1G36060	AT2G27880	AT2G25510
AT2G31160	AT4G09760	AT3G55240	AT1G31580	
AT3G57920	AT4G39480	AT5G14920	AT5G42530	
AT4G24540	AT5G57340	AT5G23000	AT3G23290	

8.3.2 FM vs. ST3

TAIR-ID	AT2G45400	AT4G32295	AT4G37300	AT1G40104
AT5G03840	AT4G36260	AT2G13550	AT3G26740	AT1G09460
AT1G47990	AT2G15440	AT2G37420	AT3G02170	AT5G48180
AT2G22540	AT1G07480	AT4G21960	AT1G78830	AT3G51670
AT4G24540	AT1G31760	AT4G31620	AT5G07770	AT4G32980
AT2G45410	AT1G75520	AT2G24700	AT2G05380	AT2G15960
AT5G54050	AT5G52310	AT4G15910	AT2G07981	AT1G19570
AT3G25640	AT4G01460	AT5G48500	AT5G57130	AT2G42730
AT2G31160	AT3G01520	AT4G18700	AT5G43700	AT4G37390
AT5G28490	AT3G22550	AT4G11990	AT5G47500	AT1G70830
AT2G17950	AT1G75820	AT1G61570	AT2G33810	AT4G30250
AT3G59270	AT2G41070	AT3G01310	AT2G23170	AT1G77110
AT5G44630	AT5G03150	AT1G49240	AT2G08986	AT5G16000
AT2G14210	AT3G58770	AT1G33590	AT1G55110	AT5G47600
AT5G57785	AT1G69360	AT1G30490	AT4G29700	AT5G50915
AT1G49830	AT2G33860	AT3G22142	AT2G30860	AT1G52000
AT1G69850	AT2G21060	AT2G45190	AT1G25560	AT3G16240
AT2G33880	AT5G42900	AT3G53420	AT2G36270	AT5G49700
AT2G38530	AT3G58040	AT4G37430	AT2G02100	AT4G09030
AT4G38810	AT5G56530	AT1G12780	AT4G02290	AT3G54640
AT5G39850	AT4G23750	AT1G21830	AT2G45470	AT2G44670
AT5G18560	AT3G16770	AT5G24420	AT3G30775	AT1G12240
AT3G19200	AT5G15970	AT3G50630	AT5G14920	AT3G04290

AT5G24780	AT5G02380	AT1G52400	AT3G17580	AT4G36870
AT1G74100	AT5G05690	AT1G72260	AT2G34810	AT1G30650
AT3G26510	AT3G26520	AT3G09260	AT1G21460	AT5G20240
AT3G25560	AT4G00180	AT3G13400	AT3G28790	AT3G54340
AT3G63300	AT4G37470	AT3G28220	AT5G17760	AT3G14380
AT5G06870	AT5G40450	AT3G23130	AT4G24130	AT4G30270
AT1G10070	AT5G11060	AT3G03270	AT2G39010	AT3G28980
AT3G63210	AT1G52030	AT4G31500	AT1G52040	AT5G57720
AT2G13360	AT5G11070	AT1G54010	AT2G23760	
AT3G16470	AT4G33150	AT3G24450	AT1G24260	
AT4G21590	AT1G68780	AT1G02790	AT4G18960	
AT4G29080	AT1G69690	AT4G36740	AT2G31070	

8.3.3 IM vs. ST3

TAIR-ID	AT3G13400	AT2G36270	AT1G33590	AT5G47500
AT1G61566	AT5G24780	AT1G52030	AT5G41080	AT2G35370
AT1G78440	AT4G30270	AT1G69690	AT5G01750	AT2G21130
AT5G41050	AT2G05380	AT1G15330	AT1G71695	AT4G15630
AT3G54340	AT3G26520	AT5G11060	AT4G25100	AT1G20010
AT1G62480	AT1G78820	AT5G02380	AT4G09030	AT4G29080
AT1G24260	AT2G02100	AT4G32980	AT3G53420	AT3G30775
AT4G21870	AT5G20240	AT4G29700	AT1G23090	AT1G08630
AT2G41905	AT1G12240	AT4G37300	AT1G12900	AT4G13540
AT1G77110	AT3G16470	AT1G52400	AT3G26740	AT1G09570
AT3G28220	AT2G34810	AT4G00180	AT4G36250	AT5G10150
AT4G21590	AT1G69120	AT5G54770	AT1G26680	AT2G24150
AT1G68480	AT4G31910	AT3G26510	AT5G20250	AT4G19410
AT4G18960	AT1G21460	AT1G78830	AT1G52220	AT1G49240
AT4G12870	AT4G15620	AT5G15210	AT4G02290	AT3G11410
AT3G04290	AT1G09460	AT1G75500	AT2G33810	AT4G37390
AT1G52040	AT3G16240	AT1G54010	AT1G15380	AT1G19570
AT3G09260	AT1G52000	AT1G72260	AT4G26260	AT5G45670

AT3G61190	AT5G48500	AT1G62810	AT5G54510	AT3G21270
AT4G25050	AT4G21380	AT5G18560	AT2G33880	AT2G45660
AT1G67090	AT3G19290	AT5G57340	AT5G01310	AT3G23290
AT1G31310	AT2G33860	AT2G13550	AT1G76110	AT2G14210
AT1G21310	AT1G31760	AT5G42530	AT4G17800	AT3G25640
AT5G15780	AT4G38810	AT3G20340	AT3G10450	AT2G22540
AT4G30250	AT1G14440	AT4G35900	AT1G69850	AT4G29140
AT3G08030	AT5G60250	AT2G20550	AT2G27880	AT5G55450
AT2G15880	AT1G66970	AT1G72070	AT1G36060	AT1G47990
AT3G14990	AT4G01460	AT4G27460	AT2G42200	AT2G24850
AT1G12330	AT1G49050	AT5G49730	AT1G77660	AT5G15960
AT2G38540	AT1G46264	AT3G51810	AT3G47220	AT2G31160
AT5G65207	AT3G57920	AT5G52310	AT5G19110	AT2G25510
AT4G21960	AT5G20830	AT3G63200	AT5G39850	AT4G24540
AT5G10300	AT1G75170	AT1G43910	AT1G31580	AT5G28490
AT4G19200	AT5G15970	AT5G23000	AT3G05650	AT5G03840
AT1G21250	AT2G17840	AT3G22550	AT4G16770	
AT5G49120	AT3G26120	AT5G35770	AT2G45650	

Chapter 9

Experimental procedures for RNA interference and Genome editing technologies

In this section we describe the materials and method that we are using to generate the multiple RNA interference line and the lines with deletions in the *REM34* cluster.

9.1 multiple RNA interference lines

We have have amplified the lacZ fragment from the pUC19 with the subsequent primers:

The primer tails contain the CACC sequence and BsaI sites. We have cloned the amplicon in the pENTR-d-TOPO in order to make it a suitable Golden Gate acceptor, yielding the pENTR-GG-LacZ. Then we have aligned the coding sequences of *REM34, REM35* and *REM37* and we manually seleced three partially conserved region. We have blast searched these sequences in the Arabidopsis genome and confirmed that they are specific for the genes of interest.

Forward	CACCTGAGACCACGGTTGTGGGTCACAGCTTGTCTGTAAGCG
Reverse	AGAGACCGAATTCGCAGCTGGCACGACAGGTTTC

We then have PCR amplified one of the conserved regions from *REM34*, another from *REM35* and the third from *REM36* with the following primers:

REM34 fw	GGTCTCACACCTGAAGTTTCCAAAGGAAAGG
REM34 rev	GGTCTCTATCTCTCTCCAACCTCTTC
REM35 fw	GGTCTCAAGATTCCAAGTCCAAGGACAAG
REM35 rev	GGTCTCGTGTCAACAATAATCTGTTTC
REM36 fw	GGTCTCAGACATCATCAAGTCTAGAAGGGAAG
REM36 rev	GGTCTCGCCTTAATCATCCCACAAGCACAC

Each primers contains BsaI site designed so that in a single golden gate reaction all the fragment can be ligated directionally in the pENTR-GG-lacZ, yielding the pENTR-RNAiREM plasmid. We have sub-cloned the REM34-REM35-REM36 fragment in the pFGC5941 vector (available at ABRC under the accession CD3-447) with an LR Gateway reaction (Life technologies) producing the NOB218-RNAiREM binary vector and used this plasmid for *Agrobacterium* mediated transformation of *Arabidopsis thaliana*.

9.2 Assemply of the Deletion inducing vector

9.2.1 BASTA resistance and homology regions

We have PCR amplified the BASTA resistance cassette from the pMDC123 vector (Curtis and Grossniklaus, 2003) and the two homology regions (HR1 and HR2) from *Arabidopsis* genome with the following primers:

HR1 fw	GGTCTCAGGAGTCCCAAAAAGATCAAAAGAAAAATC
HR1 rev	GGTCTCAGGCTGTACCCATCATCTTTTTCTTACACC
Bar-1 fw	GGTCTCCCGAAATCAGCTTGCATGCCGGTCG
Bar-1 rev	GGTCTCATTTCCCCCCGCCACCAGCGGAC
Bar-2 fw	GGTCTCTGAAACGTACACGGTCGACTCGG
Bar-2 rev	GGTCTCAAGCCTATCATACATGAGAATTAAGGG
HR2-1 fw	GGTCTCATTCGTTGAGGTCTTTATGGATTTTTGG
HR2-1 rev	GGTCTCACGGCTCCGTCAAGATAAAGTTAC
HR2-2 fw	GGTCTCTGCCGTGAATGAGATAAAGGC
HR-2 rev	GGTCTCAAGCGGTGATGAGATAGAATCTGAG

We inserted a point mutation within the PCR primers in the BAR expression cassette and one in the HR2 region in order to delete internal BsaI

sites. Each primers contains BsaI site designed so that in a single golden gate reaction all the fragment can be ligated directionally in the pDGB2alphaR2.

9.2.2 CodA, p35S and t35S

We have PCR amplified the coding sequence of *CodA* from the genome of *E. Coli*, strain K-12 and p35S promoter and t35S terminator from the pMDC123 vector (Curtis and Grossniklaus, 2003) with the following primers:

p35S1 fw	GGTCTCGGGAGACTAGAGCCAAGCTGATCTCC
p35S1 rev	GGTCTCAAGTCCTGCCGCGTAGGCCTCTC
p35S2 fw	GGTCTCAGACTCATCAAGACGATCTACCC
p35S2 rev	GGTCTCTCATTTCGACTAGAATAGTAAATTG
CodA1 fw	GGTCTCGAATGTCGAATAACGCTTTACAAAC
CodA1 rev	GGTCTCGATTAGACGGTCGTATTTTTGCG
CodA2 fw	GGTCTCCTAATCGACGTTCACTGTGATG
CodA2 rev	GGTCTCAAAGCTCAACGTTTGTAATCGATGGC
HR2-2 fw	GGTCTCAGCTTCGGCCATGCTAGAGTCCG
HR-2 rev	GGTCTCGAGCGAGGTCACTGGATTTTGGTTTTAGG

We have inserted a point mutation in the p35S sequence and a silent point mutation in the coding sequence of CodA with the PCR primers in order to delete internal BsaI sites. Each primers contains BsaI site designed so that in one single golden gate reaction all the fragment can be ligated directionally in the pDGB2alpha1.

9.2.3 TALEN

We are right now working on the production of the vector containing the TALEN, the pUBI promoter t35S terminator. The TALEN were assembled according to the protocol published by Cermak et al. 2011 with the following RVD sequences:

- TALEN1 HD HD NI HD HD NI HD NI NG NG NG HD NG HD NG HD NG
- TALEN2 NH NH NG NG NG NH NG HD NG HD HD NH NH NG NI HD NH NI

we are planning of amplfying these sequences with the following primer:

pUBI fw	CGTCTCGGGAGTACCCGACGAGTCAGTA
pUBI rev	CGTCTCTCATTAGTGTTAATCAGAAAAACTCAG
TALEN1 fw	CGTCTCTAATGGCTTCCTCCCCTCCAAAG
TALEN1 rev	CGTCTCTAAAGTTTATCTCACCGTTATT
TALEN2 fw	CGTCTCTACCTATGGCTTCCTCCCCTCCAAAG
TALEN2 rev	GCGTCTCGAAGCTTAAAAGTTTATCTCACCG
t35S fw	CGTCTCAGCTTCGGCCATGCTAGAGTCCG
t35S rev	CGTCTCGAGCGAGGTCACTGGATTTTGGTTTTAGG

And moreover we have synthesized the T2A sequence flanked by BsmbI sites as the two complementary oligo CGTCTCACTTTGGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCTTGAGACG and CGTCTCAAGGTCCAGGATTCTCCTCGACGTCACCGCATGTTAGCAGACTTCCTCTGCCCTCTCCGCTTCCAAAGTGAGACG.

Both the primers and the T2A sequences contain BsmbI sites so that in one single one single golden gate reaction all the fragment can be ligated directionally in the pDGB2omega2

9.2.4 Golden Gate reaction

All Golden Gate reaction has been carried out in a volume of 20 µl containing:

- 40 ng of each PCR fragment
- 40 ng of undigested destination vector
- 10 U of restriction enzyme (Bsa or BsmbI)
- 1 X T4 ligase buffer
- 3 U of T4 ligase

and was run in an Eppendorf thermocycler with the program:

1. 37 °C for 5 minutes,

2. 15 °C for 10 minutes

3. Go to step 1 for 9 times

4. 50 °C for 5 minutes

5. 80 °C for 5 minutes

and used directly to transform E. Coli or stored at −20 °C

Chapter 10

Bibliography of supplementary materials

Mark D Curtis and Ueli Grossniklaus. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant physiology*, 133 (2):462--469, 2003.

Tomas Cermak, Erin L Doyle, Michelle Christian, Li Wang, Yong Zhang, Clarice Schmidt, Joshua A Baller, Nikunj V Somia, Adam J Bogdanove, and Daniel F Voytas. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research*, ePub e82, 2011.