MADS SPECIFICITY

Hilda van Mourik 2017

# MADS SPECIFICITY

**Unravelling the dual function of the MADS domain protein FRUITFULL**

Hilda van Mourik

# INVITATION

You are cordially invited to attend the public defence of my PhD thesis entitled:

**MADS SPECIFICITY**
**Unravelling the dual function of the MADS domain protein FRUITFULL**

on Friday, November 10th, 2017, at 13.30, in the Aula of Wageningen University
General Foulkesweg 1a,
Wageningen

**Hilda van Mourik**
hildavmourik@gmail.com

**Paranymphs:**
Alice Pajoro
alice.pajoro@gmail.com

and

Sam van Es
sam.vanes@wur.nl

# Propositions

1. Inter-familiar protein interactions influence *in vivo* DNA binding, adding another layer to TF binding specificity.
   (This thesis)
2. The current knowledge of *in vivo* TF DNA binding is insufficient to make full use of ChIP-seq data.
   (This thesis)
3. In many scientific fields most severe phenotypes are already found, therefore statistics will become increasingly important in future research.
4. Most definitions in science, like the definition of a gene, dependent on the context and the kind of research.
   (El-Hani, 2017, Mol Syst Biol. 2007; 3: 87)
5. Stress management courses should be mandatory, as the people that need them most, will not have time to participate in voluntary courses.
6. Serena Williams is the best tennis player in the world, not just the best female player.

Propositions belonging to the thesis, entitled:

**"MADS SPECIFICITY – Unravelling the dual function of the MADS domain protein FRUITFULL"**

Hilda van Mourik

Wageningen, 10 November 2017

# MADS SPECIFICITY

Unravelling the dual function of the MADS domain protein FRUITFULL

Hilda van Mourik

Thesis committee

Promotor

Prof. Dr Gerco C. Angenent

Personal chair at the Laboratory of Molecular Biology

Wageningen University & Research

Co-promotor

Prof. Dr Kerstin Kaufmann

Personal chair at the Institute for Biology

Humboldt University of Berlin, Germany

Other members

Prof. Dr Remko Offringa, Leiden University

Prof. Dr Dolf Weijers, Wageningen University & Research

Prof. Dr Martin M. Kater, The University of Milan, Italy

Dr Gabino F. Sanchez Perez, Genetwister Technologies B.V., Wageningen

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences (EPS)

# MADS SPECIFICITY

## Unravelling the dual function of the MADS domain protein FRUITFULL

Hilda van Mourik

**Thesis**

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 10 November 2017
at 1:30 p.m. in the Aula.

Hilda van Mourik

MADS SPECIFICITY - Unravelling the dual function of the MADS domain protein FRUITFULL, 202 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2017)

With references, with summary in English

ISBN 978-94-6343-672-4

DOI https://doi.org/10.18174/421563

# Contents

**Chapter 1**.....page 7

Towards understanding transcription factor DNA-binding specificity

**Chapter 2**.....page 31

Dual specificity and target gene selection by the MADS domain protein FRUITFULL

**Chapter 3**.....page 77

Characterization of *in vivo* DNA-binding events of plant transcription factors by ChIP-seq: experimental protocol and computational analysis

**Chapter 4**.....page 113

FRUITFULL controls *SAUR10* expression and regulates *Arabidopsis* growth and architecture

**Chapter 5**.....page 147

FRUITFULL directly regulates two flavonoid synthesis genes, indicating a possible link between flowering time and flavonoids

**Chapter 6**.....page 163

Concluding remarks and perspectives

**References.**.....page 174

**Summary**.....page 191

**Acknowledgements**.....page 194

**About the author**.....page 198

**EPS education statement**.....page 200

# Chapter 1

## General introduction

# Towards understanding transcription factor DNA-binding specificity

Encrypted in the DNA lays all information needed for the development of an organism. The transcription of this information in precise patterns of gene activity results in the development of different cell types, organs, and developmental structures. Moreover, transcriptional regulation enables an organism to respond to changing environmental conditions. Essential for the regulation of transcription are DNA-binding transcription factors (TFs). TFs bind the DNA in a sequence-specific fashion. Upon binding to a TF specific DNA motif, TFs typically activate or repress the transcription of nearby genes. Hence, TFs play a key role in transcriptional regulation.

In higher organisms, TFs function in complex networks to achieve precise patterns of gene expression (reviewed by Kaufmann *et al.* (2010a)). To correctly regulate genes, TFs need to recognize specific DNA sequences and bind to these sequences with high specificity. Hence, how TFs achieve DNA binding specificity *in vivo* is a fundamental question in molecular biology. Knowledge about TF-DNA binding mechanisms will enable us to better understand gene regulatory networks, to design TFs with new and improved properties, and ultimately would allow scientists to predict spatiotemporal expression patterns by studying the *cis*-regulatory elements of a gene. This review elaborates on the different high-throughput methods for determining TF-DNA interactions, focusses on the DNA binding specificities of TFs, and aims to bridge the current views on TF specificity.

## TF binding specificity *in vitro* versus *in vivo*

TFs bind different DNA sequences with different affinities. By comparing the sequence specific affinity of the TF to the affinity of the TF to all other possible DNA sequences, researchers can determine how well a TF can distinguish different sites, this is often referred to as the binding specificity of a TF. Traditional methods to determine TF-DNA interactions usually study one specific TF-DNA interaction, using these techniques to study multiple interactions is laborious. Hence, these techniques are not well suited to study TF binding specificity. However, in the last decades several techniques were developed that determine TF binding affinity to large numbers of DNA sequences, including methods to determine genome-wide TF binding. This chapter will discuss a selection of

these techniques that determine DNA binding affinities and genome-wide binding in a high-throughput fashion.

## Measuring *in vitro* DNA binding affinities

*In vitro* methods that determine binding affinity can be subdivided into three groups: techniques that measure absolute affinities directly, microarray-bases methods, and methods that require next-generation sequencing (NGS).

A technique that directly measures affinity is mechanical induced trapping of molecular interactions (MITOMI), this technique uses a microfluidic device aligned to a microarray (Maerkl and Quake, 2007; Fordyce *et al.*, 2010). Each cell of the microarray contains materials to allow the production of the studied TF from a synthetic gene. Besides the synthesized protein, each cell contains a specific fluorescently labelled DNA molecule in a particular concentration, which can be bound by the TF. Unbound DNA is washed out while mechanical trapping, using protein-specific antibodies attached to the surface of the cells, protects the existing protein-DNA interactions. The amount of protein-DNA interaction is determined using the fluorescent tag attached to the DNA. By determining the fraction of trapped DNA per cell and comparing the results of wells with different initial DNA concentrations, the relative affinity of each DNA sequence can be determined. Absolute affinities can be calculated by comparing the determined relative affinities with a reference DNA molecule with a known absolute affinity. The original MITOMI approach allowed the analysis of approximately 640 interactions (Maerkl and Quake, 2007), newer versions of this technique (MITOMI 2.0) have reported to analyse approximately 4,160 interactions per experiment (Fordyce *et al.*, 2010). The later MITOMI 2.0 method has for example been used to determine binding specificity of 28 *Saccharomyces cerevisiae* TFs from 10 different protein families to all possible 8-bp DNA sequences (Fordyce *et al.*, 2010).

Another technique that directly measures affinity is surface plasmon resonance (SPR) (Shumaker-Parry *et al.*, 2004; Campbell and Kim, 2007), while this technique is most often used to study protein-ligand interactions, it can also be used for protein-DNA interactions. This technique measures the angle of light reflected from a thin metal surface, the mass of molecules attached to the other side of the surface is directly influencing the angle of light reflection. This technique allows live tracking of the DNA-TF interaction (Boer *et al.*, 2014). By attaching DNA sequences to

the metal surfaces, this technique can measure changes in reflection angle over time, while increasing amounts of TF bind to the DNA until an equilibrium is reached. After measuring the binding constant, the dissociation constant of the interaction can be determined by washing away the TFs, again until an equilibrium is reached. By determining both the association and the dissociation constant this technique allows a direct calculation of absolute affinities where most techniques determine relative affinities. With the introduction of high throughput SPR (**HT-SPR**), this technique can be used to study more than 1000 different interactions simultaneously (Campbell and Kim, 2007).

A widely used microarray-based technique to study protein-DNA interactions is protein-binding microarray (**PBM**) (Mukherjee *et al.*, 2004). A PBM contains hundreds to thousands spots that each contains a specific dsDNA sequence. Purified TFs are added to the array and after sufficient time for the TF to bind the DNA, washing of the array removes non-specific protein-DNA interactions. The amount of bound protein to each DNA sequence is measured by adding and detecting a TF-specific fluorescent antibody. PBM probes are usually 8-10 bp long random sequences, this allows the microarray to contain almost all possible potential binding sites (reviewed in Stormo and Zhao (2010)). PBMs have for example been used to study the DNA binding preferences of nearly all members of the *Caenorhabditis elegans* bHLH family of TFs (Grove *et al.*, 2009). Grove *et al.* (2009) integrated information of bHLH dimerization, spatiotemporal expression, and DNA binding specificity to generate a gene-regulatory network. This network allowed them to study functional divergence between the different *C. elegans* bHLH proteins. Besides *C. elegance*, also the DNA binding specificities of 104 mouse TFs of 22 structural families have been determined using PBMs (Badis *et al.*, 2009). In a comprehensive study using PBMs, the DNA-binding specificities of 63 plant TFs representing 25 families were determined (Franco-Zorrilla *et al.*, 2014).

In addition to direct measurements and microarray based techniques, NGS is frequently used in techniques that study binding specificities. One example of such a technique is systematic evolution of ligands by exponential enrichment followed by sequencing (**SELEX-seq**), also called high-throughput SELEX (**HT-SELEX**) (Jolma *et al.*, 2010; Moyroud *et al.*, 2011; Sayou *et al.*, 2014). For this technique, purified TFs of interest are allowed to interact with a large pool of random or genomic DNA sequences. The DNA sequences contain primer binding sites on both ends to allow amplification by PCR. By washing away unbound DNA, DNA-protein complexes are retained. After

this washing step and PCR amplification, the bound DNA is re-used for subsequent rounds of SELEX. As higher affinity sites have a higher chance to be bound by protein, performing multiple SELEX rounds allows for the measurement of the enrichment of different DNA sequences over the course of multiple rounds of SELEX. The final library is analysed using NGS. The randomized DNA library used as starting point is very large and can comprise more than $10^{15}$ different sequences (Jolma *et al.*, 2010). Although usually not all these different variants are sequenced, the protein is able to select from an enormous population of sequences. This method was successfully used to study the binding specificities of at least one member of 14 different classes of human TFs (Jolma *et al.*, 2010). A similar approach was used to determine the preferred binding sites of the plant-specific TF LEAFY (LFY) (Moyroud *et al.*, 2011; Sayou *et al.*, 2014).

In summary, many *in vitro* techniques have been developed that provide high-throughput information about TF DNA binding specificity. These *in vitro* affinity measurements allowed for the identification of canonical binding sites for many TF families. However, while *in vitro* affinity measures are important for the determination of binding specificity, it is not easy to use these affinities to predict genome-wide binding of a TF. Several methods to determine genomic regions bound by a TF will be discussed in the next paragraph.

## Determining genome-wide TF binding *in vitro* and *in vivo*

Determination of genomic regions bound by a TF provides information about genes bound and possibly regulated by this TF. Studying the genome-wide binding patterns of many different TFs enable scientists to systematically generate large regulatory networks. Moreover, in combination with RNA-seq or other expression studies, genome-wide binding studies allow the identification of directly regulated genes. Most techniques used to study genome-wide TF-DNA binding make use of immunoprecipitation followed by microarray or by NGS.

An *in vitro* technique to study genome-wide TF-DNA interactions is DNA immunoprecipitation followed by tiling microarray detection (**DIP-chip)** (Liu *et al.*, 2005). With this technique, a purified TF is incubated with naked sheared genomic DNA. Immunoprecipitation, using an antibody against the TF of interest, separates bound and unbound DNA. Bound DNA sequences are sequentially purified, amplified, labeled fluorescently, and analysed by DNA microarray hybridization. A technique highly similar to DIP-seq is DNA affinity purification followed by high throughput

sequencing (**DAP-seq**) (Rajeev *et al.*, 2011; O'Malley *et al.*, 2016). Like DIP-chip, also this technique uses naked sheared genomic DNA as template, but applies NGS for the identification of bound DNA sequences. Since these techniques use the same DNA template as fully *in vivo* DNA binding assays (like Chromatin Immunoprecipitation, ChIP), the output can be directly compared with *in vivo* genome-wide DNA binding data. DAP-seq has recently been used to study the binding sites of 529 *Arabidopsis thaliana* TFs (O'Malley *et al.*, 2016). The authors compared DAP-seq results from sheared genomic DNA that contains methylation, and un-methylated sheared genomic DNA to determine the binding efficiency of the tested TFs to methylated DNA motifs.

*In vivo* genome-wide TF-DNA interactions are generally studied using chromatin immunoprecipitation followed by microarray hybridization (**ChIP-chip**) (Ren *et al.*, 2000) or high-throughput sequencing (**ChIP-seq**) (Johnson *et al.*, 2007; Kaufmann *et al.*, 2010b). With these techniques, *in vivo* DNA-protein complexes are cross-linked in intact tissues. Cross-linked chromatin is extracted and protein-DNA complexes containing the protein of interest are isolated using a TF specific antibody or antibodies against tagged proteins of interest. The precipitated DNA is analysed by hybridization to a DNA tiling microarray or by NGS. Disadvantage of these techniques is the low resolution, it is not possible to determine precisely where the TF binds as the obtained binding peaks are usually 100-300 bp wide. Motif enrichment algorithms, for example MEME (Bailey and Elkan, 1994), help to shed light on the motifs potentially bound by the protein within ChIP-seq/chip binding peaks. A more recent technique, called **ChIP-exo**, results in a better resolution (Rhee and Pugh, 2011). This technique extends the ChIP-seq protocol after immunoprecipitation by using an exonuclease to trim DNA sequences flanking the actual TF binding site. By shortening the DNA sequences, this technique increases the mapping resolution. These *in vivo* techniques allow analysis of TF-DNA interactions in a cell type-, tissue-, or stage-specific manner. The use of ChIP to determine *in vivo* genome-wide DNA binding of TFs in different developmental stages has for example been demonstrated by Pajoro *et al.* (2014). Dynamics of DNA binding of the MADS domain TFs APETALA1 (AP1) and SEPALLATA3 (SEP3) during flower development was analysed using ChIP-seq, providing a stage-specific gene regulatory network (Pajoro *et al.*, 2014).

These high-throughput techniques facilitated the identification of canonical binding sites of many TFs and provided genome-wide DNA binding data. However, the mechanisms by which TFs select

their targets *in vivo* is still not fully understood. To understand how DNA binding specificity is achieved by TFs, it is essential to firstly understand the interaction between the protein and 'naked' DNA.

## DNA recognition by TFs via base and shape readout

Most of the current knowledge about biophysical properties of TF-DNA interactions is derived from structural biology, co-crystallization, and nuclear magnetic resonance (NMR). These techniques revealed that interactions between a TF and the DNA depend on two protein-DNA recognition mechanisms: base and shape readout.

Most eukaryotic TFs have a modular protein structure. Having different functional domains is essential for TF function as the regulation of genes requires DNA-binding, transcriptional activation or repression, and formation of protein-protein interactions. The domain that mediates contact between the TF and the DNA is called the DNA binding domain (DBD). This DBD plays an important role in DNA binding preferences of a TF as DNA binding is dependent on the preference of amino acids in the DBD for particular nucleotides. NMR and co-crystallization studies have shown that this preference is influenced by specific contacts between the amino acid side chains of the DBD and the accessible edges of the DNA base pairs (Rohs *et al.*, 2010). The specific physical interactions between a TF and the DNA include electrostatic, hydrophobic, and hydrogen bond interactions (Seeman *et al.*, 1976; Luscombe *et al.*, 2001). Interactions between the DNA and TFs can also be bridged by water molecules, water bridges most likely decrease electrostatic repulsion between electron negative atoms (Jayaram and Jain, 2004). These direct interactions between the two macromolecules is referred to as 'base readout'. Besides base readout, TFs are also able to read the structure of the DNA this is called 'shape readout'. Shape readout depends on the sequence-dependent DNA structure and interactions between the protein and the sugar/ phosphate backbone of the DNA (Stella *et al.*, 2010; Chen *et al.*, 2013; Hancock *et al.*, 2013). The relative contributions of these two readouts is different for each protein family.

## Locating binding sites within a complex chromatin context

Although NMR and co-crystallography provided important knowledge about contacts between TFs and DNA, how a TF locates a functional binding site in a native chromatin context is less well

understood. In the nucleus the concentration of DNA is high, for most TFs millimolar amounts of potential binding sites are present (Stormo and Zhao, 2010). *In vitro* studies showed that most studied TFs bind to relatively short motifs, 8 – 12 bp (Jolma *et al.*, 2010; Franco-Zorrilla *et al.*, 2014; O'Malley *et al.*, 2016). By chance these motifs occur often in the genome, while ChIP-chip and ChIP-seq studies have shown that only a small proportion of these sequences are bound *in vivo* (Heyndrickx *et al.*, 2014; Yue *et al.*, 2014). In addition, to allow a quick response to stimuli, TFs need to locate and bind their functional binding sites quickly.

To explain how TFs scan and search for binding sites in the DNA, a rapid search model has been proposed (Slattery *et al.*, 2011). This rapid search model consists of two phases. In the first phase, TFs bind weakly and possibly non-specifically to the DNA. This weak binding is followed by migrations over the DNA, including rapid cycles of dissociations and re-binding the DNA or sliding over the DNA. In the second stage of this model, the TF identifies a high affinity binding site followed by specific binding to this side (Slattery *et al.*, 2011). Examples of TF that have been shown to fit this rapid search model are the rat Islet-1 (Isl1) protein (Ippel *et al.*, 1999) and the *Drosophila melanogaster* proteins HOXD9 and NK-2 (Vuzman and Levy, 2010). The Isl1 protein contains a domain that is largely unfolded domain at 37°C, upon binding to a high affinity binding site the Isl1 tail stabilizes. The stabilisation of this tail upon binding a high affinity site may prevent further sliding along the DNA (Ippel *et al.*, 1999). The proteins HOXD9 and NK-2 initially bind DNA with low affinity with their core helical homeodomains, after this weak binding the N-terminal tail of the protein scans the DNA for high affinity sites. This N-terminal tail locates high affinity binding sites via sequence-specific charge-to-charge interactions (Vuzman and Levy, 2010).

## Chromatin accessibility and TF binding

In order to fit the DNA inside the nucleus, condensation of the DNA is needed. Nucleosomes consist of 147 bp of DNA wrapped around eight histones and are the major means of chromatin condensation. Packing of the DNA by nucleosomes changes the accessibility of the DNA, and makes the DNA less accessible for TFs to bind their binding sites. The location and interaction strengths of nucleosomes can be influenced by histone modifications. More than hundred different histone modifications are known, making DNA accessibility a dynamic process (Tan et al., 2011). The

Table 1 | Comparison of high-throughput approaches for the determination of TF-DNA interactions.

| Method | Number of studied interactions | Probe type | Output | Advantage | Disadvantage | References |
|---|---|---|---|---|---|---|
| MITOMI 2.0 | $10^3$-$10^4$ | Oligo library | Absolute affinities | - Determines absolute affinities<br>- Can detect low-affinity and transient interactions<br>- No protein purification needed | - TF specific antibody or antibody against tagged protein needed<br>- Limited length of probe (probe length of max. 8-bp will allow the analysis of all possible sequences) | (Maerkl and Quake, 2007; Fordyce *et al.*, 2010) |
| HT-SPR | $10^3$ | Microarray | Absolute affinities | - Determines absolute affinities<br>- Can detect low-affinity and transient interactions (signal amplification needed)<br>- Real time measurement<br>- No antibodies needed | - Protein purification or *in vitro* production needed<br>- Cannot analyse proteins with a very low mass<br>- Sensitive to detecting non-specific binding | (Shumaker-Parry *et al.*, 2004; Campbell and Kim, 2007) |
| PBM | $10^5$-$10^6$ | Microarray | Binding models | - No antibodies needed<br>- Low costs<br>- Easy protocol | - TF specific antibody or antibody against tagged protein needed<br>- Protein purification or *in vitro* production needed<br>- Limited number of spots<br>- Often observed artefact: proteins prefer to bind the free-ends of the probes | (Mukherjee *et al.*, 2004; Badis *et al.*, 2009; Grove *et al.*, 2009; Franco-Zorrilla *et al.*, 2014) |

| | | | | | | |
|---|---|---|---|---|---|---|
| SELEX-seq | $10^{15}$ | Oligo library | Binding models | - No limitation to the length of DNA probes<br>- Number of analysed interactions only limited by the sequencing depth | - Recovers primarily strongly bound sequences<br>- TF specific antibody or antibody against tagged protein needed<br>- Protein purification or *in vitro* production needed<br>- Sensitive to PCR amplification bias | (Jolma *et al.*, 2010; Moyroud *et al.*, 2011) |
| DIP-chip | $10^{5}$ | Microarray | *In vitro* genome-wide binding | - Allows using naked genomic DNA<br>- Easy comparison to *in vivo* binding data | - TF specific antibody needed<br>- Protein purification or *in vitro* production needed<br>- Limited microarray positions<br>- Low-resolution | (Liu *et al.*, 2005) |
| DAP-seq | Genome-wide | Genomic fragments | *In vitro* genome-wide binding | - Allows to use naked genomic DNA<br>- Easy comparison to *in vivo* binding data<br>- Compared to DIP-chip higher resolution<br>- Genomic-coverage | - TF specific antibody or antibody against tagged protein needed<br>- Antibodies against tagged protein of interest needed<br>- Exact binding location not detected, low resolution | (Rajeev *et al.*, 2011; O'Malley *et al.*, 2016) |

| | | | | | | |
|---|---|---|---|---|---|---|
| ChIP-chip | $10^6$ | Microarray | *In vivo* genome-wide binding | - Provide genome-wide information<br>- Allows cell-type, tissue-, and stage-specific TF binding site determination<br>- Genomic coverage | - TF specific antibody needed<br>- Also captures indirect DNA binding<br>- Genome-wide arrays only for limited species available<br>- Low-resolution | (Ren *et al.*, 2000; Kaufmann *et al.*, 2010b) |
| ChIP-seq | Genome-wide | Genomic fragments | *In vivo* genome-wide binding | - Provide genome-wide information<br>- Allows cell-type, tissue-, and stage-specific TF binding site determination<br>- Genomic coverage<br>- Compared to ChIP-chip higher resolution | - TF specific antibody needed<br>- Exact binding location not detected | (Johnson *et al.*, 2007; Kaufmann *et al.*, 2010b) |
| ChIP-exo | Genome-wide | Genomic fragments | *In vivo* genome-wide binding | - Provides genome-wide information about promoter occupancy<br>- Allows cell-type, tissue-, and stage-specific TF binding site determination<br>- Genomic-coverage<br>- Very high resolution | - TF specific antibody needed<br>- Also captures indirect DNA binding<br>- Compared to ChIP-seq/-chip smaller amount of recovered DNA due to extra treatment, reducing the quality of the experiment | (Rhee and Pugh, 2011) |

impact of DNA accessibility and histone modifications on TF binding has been discussed in many studies. For example, a study on the *Saccharomyces cerevisiae* TF Leu3 has shown that the binding motif of Leu3 is similar for *in vivo* genomic binding compared to *in vitro* binding to naked genomic DNA. However, the exact binding locations of Leu3 differs in the two conditions. The authors showed that the differences in binding locations can partly be explained by differences in DNA accessibility due to presence of nucleosomes in the *in vivo* experiment (Liu *et al.*, 2006). Another example comes from the mammalian field, a genome-wide study using an inducible system to study DNA binding of the glucocorticoid receptor (John *et al.*, 2011). The data showed that chromatin accessibility has a strong effect on DNA binding of the glucocorticoid receptor, which led to the hypothesis that this influence of chromatin accessibility on DNA binding might explain the strong cell-type specific functions of this TF.

Although most TFs cannot bind compact DNA, TFs called pioneering factors can bind DNA that is inaccessible to other TFs (reviewed by Magnani *et al.* (2011)). Binding of pioneering factors to inaccessible DNA initiates the assembly of transcriptional complexes at promoters, which is a trigger for nucleosome repositioning, thereby increasing accessibility. Pioneering TFs are often not able to regulate gene expression themselves, but they play an important role in initiating the formation of functional TF-DNA complexes that control transcription. To summarize, although not all TF binding is influenced by chromatin accessibility, dynamics and cell type-specific DNA accessibility influence binding of many TFs. Hence, DNA accessibility can partly explain cell type-specific DNA binding events by TFs.

## DNA methylation and TF binding

Besides nucleosome occupancy, DNA methylation is known to influence the binding of TFs to DNA. DNA methylation is the covalent addition of a methyl group to nucleotides. In the traditional view of DNA methylation, the presence of methyl groups on the DNA results in silencing of genes (Holliday and Pugh, 1975; Riggs, 1975; Hsieh, 1994) and studies have shown that DNA methylation in TF binding sites can physically obstruct TF-DNA binding (Tate and Bird, 1993; O'Malley *et al.*, 2016). Recent studies on *in vivo* DNA methylation dynamics challenge this traditional view. In this traditional view, only proteins containing a methyl-CpG-binding domain can bind methylated DNA (Wade, 2001; Hendrich and Tweedie, 2003). However, in the last decades an increasing number of

studies have shown that TFs without a mCpG domain can bind methylated DNA *in vivo* (Prokhortchouk *et al.*, 2001; Lopes *et al.*, 2008). Interestingly, studies revealed that TFs can bind methylated and non-methylated sequences with different affinities (Rishi *et al.*, 2010; Hu *et al.*, 2013; O'Malley *et al.*, 2016). DNA methylation influences TF binding specificities by altering the DNA shape (Tippin *et al.*, 1997; Buck-Koehntop *et al.*, 2012; Lazarovici *et al.*, 2013), thereby changing the shape readout resulting in an altered TF binding affinity. Moreover, changes in DNA methylation are often linked to changes in DNA accessibility (Domcke *et al.*, 2015; Maurano *et al.*, 2015), thereby influencing TF-DNA binding. So, although most TFs do not bind methylated DNA and follow the traditional concept that methylation leads to interference of TF-DNA interactions, an increasing number of proteins are identified to interact with methylated DNA. As DNA methylation affects the DNA shape and accessibility, DNA methylation contributes to *in vivo* TF-DNA interaction specificity.

## How TFs achieve DNA binding specificity *in vivo*

Many TFs are members of large protein families, which possess highly conserved DBDs, resulting in the recognition of very similar DNA sequences. However, many *in vivo* studies have shown that family members with highly similar *in vitro* binding specificities bind different genomic targets. Part of the explanation can be found in different spatiotemporal expression patterns of family members. Differences in expression domains will allow them to encounter different DNA accessibilities and methylation status which can explain part of the *in vivo* genomic binding differences between family members. However, studies demonstrated that closely related proteins expressed in the same cells, therefore encountering the same DNA accessibility and DNA methylation patterns, bind different *in vivo* targets (Tao *et al.*, 2012; Pajoro *et al.*, 2014; Völkel *et al.*, 2015). Hence, although tissue-specific DNA accessibility and methylation influence *in vivo* DNA binding, these mechanisms are unlikely to account for the full *in vivo* specificity of TF family members. Therefore, how these individual family members achieve DNA-binding specificity *in vivo* remains an important question.

### Low-affinity binding sites

One possible explanation for the differences seen between *in vitro* and *in vivo* binding might arise from a prevalent focus on high affinity binding. Binding motifs are usually generated from high-affinity binding sites, while *in vivo,* TFs are also able to bind low-affinity sites (Tanay, 2006). Studies on *Drosophila* HOX proteins have shown that HOX proteins sharing highly similar high affinity

binding motifs bind different targets *in vivo* (Berger *et al.*, 2008; Noyes *et al.*, 2008). Recent work showed that the HOX protein Ubx binds specifically to clusters of low-affinity sites, changing these sites to high-affinity sites resulting in ectopic expression of Ubx targets (Crocker *et al.*, 2015). While high-affinity sites can be bound by all HOX proteins and potentially many more homeodomain proteins, low-affinity sites are specifically bound by unique HOX proteins. Thereby binding to these low-affinity sites provides specificity for the HOX family (Crocker *et al.*, 2015). However, this affinity-specificity trade-off accounts for some families of TFs, but it is unlikely to fit to all TF families.

## Specificity through protein-protein interactions

Although the DBDs of TF family members are often highly conserved, amino acids outside the DBD domains may differ and can indirectly influence DNA binding. This has for example been shown for example for mammalian zinc finger proteins (Lim *et al.*, 2016). Lim et al. (2016) fused the non-DNA binding domains of the human KRÜPPEL-LIKE FACTOR3 (KLF3) to an artificial zinc finger (AZF) protein. By comparing the *in vivo* DNA binding of both the AZF protein and the fusion protein, they demonstrated that adding the non-DNA binding domains of KLF3 can alter *in vivo* DNA binding of AZF. These results confirm earlier publications reporting that mutations in the N-terminal non-DNA binding domain of these zinc finger proteins influence genome-wide binding (Burdach *et al.*, 2013). Domains that do not interact with the DNA can alter DNA binding specificity in different ways. Differences in amino acids outside the DBD domain can result in different folding of the protein, which can influence the structure of the DBD domain and subsequently changes DNA binding preferences. It is also possible that differences in protein structure and sequence result in different abilities to bind family members, other proteins, or co-factors.

Many TF families bind DNA as obligatory homodimers and/or heterodimers. Most *in vitro* DNA binding studies focus on homodimeric TF-DNA interactions (e.g. Franco-Zorrilla *et al.* (2014); O'Malley *et al.* (2016)), while *in vivo* these proteins might preferentially form heterodimers with other family members. These heterodimeric interactions might have different DNA-binding preferences compared to homodimers. An example of a protein family that binds DNA as dimers is the basic leucine-zipper (bZIP) family. *In vitro* large-scale interactions studies with bZIP proteins from five metazoan and two unicellular eukaryotic species revealed that in the studied organisms only 5 - 30% of all possible heterodimeric combinations were formed, hence different bZIP

proteins have a highly specific interaction capability (Reinke *et al.*, 2013). As *in vitro* studies have shown that different combinations of bZIP heterodimers have different DNA-binding specificities (Hai and Curran, 1991), the protein dimerization specificity might be important for DNA-binding specificity of family members. Reinke et al. (2013) indeed showed that loss of protein interaction capability of bZIP proteins results in changes in *in vitro* DNA binding preferences. The interaction capability and *in vivo* availability of heterodimer partners could therefore be an important parameter for *in vivo* DNA binding specificity of different TF family members.

Besides interactions with family members, also interactions between TFs and other proteins or co-factors can influence DNA binding (Bemer *et al.*, 2017). An example is the interactions between the bZIP TFs PHYTOCHROME INTERACTING FACTOR 1 (PIF1) and ABSCISIC ACID INSENSITIVE 5 (ABI5) in *Arabidopsis* (Kim *et al.*, 2016). ChIP-chip studies on PIF1 showed that PIF1 binds *in vivo* to sites lacking its well-known binding motif, 5'-CACGTG-3', so-called G-box elements (Oh *et al.*, 2009). Recent work on the binding of PIF1 showed that the interaction between PIF1 and ABI5 targets PIF1 to specific genomic sites (Kim *et al.*, 2016). Interestingly, some TFs form interactions with multiple proteins in a combinatorial manner, different combinations of the proteins bind different genomic areas in a cell-type specific fashion. More and more examples of this phenomenon are discovered in various organisms. One example comes from the interactions and cooperative DNA binding of the *Arabidopsis* TFs AUXIN RESPONSE FACTOR 6 (ARF6), PIF4, and BRASSINAZOLE RESISTANT 1 (BZR1) (Oh *et al.*, 2014). Cooperative binding of these proteins to genomic targets allows different growth regulating pathways to act together as the different proteins integrate signals from auxin, brassinosteroid, light, and temperature. Besides interactions with PIF4 and BZR1, ARF6 is also capable of interacting with other bHLH family members, suggesting cell- or tissue-specific ARF6 complexes and thereby creating binding specificity (Oh *et al.*, 2014). Another example are the interactions of the human protein T-CELL ACUTE LYMPHOCYTIC LEUKEMIA PROTEIN 1 (TAL1). While TAL1 is essential for the specification, survival, and competence of haematopoietic stem cells, ectopic expression of this protein in lymphoid tissue results in oncogenesis (reviewed in Lécuyer and Hoang (2004)). ChIP-seq studies show that TAL1 genomic targets show limited overlap in normal erythroid cells versus leukemic T cells (Palii *et al.*, 2010). The observed differences in genomic-binding results from interactions of TAL1 with Protein ETS1 and RUNT-RELATED TRANSCRIPTION FACTOR 1/3 (RUNX1/3) in leukemic T-cells, these interactions target TAL1 to

specific genomic loci (Palii *et al.*, 2010). Besides ETS1 and RUNX1/3, TAL1 interacts with many more proteins, including HEB, E2A, LMO1/2, and GATA3, and it has been shown to bind multiple targets in a combinatorial way with these interactors (Wilson *et al.*, 2010; Sanda *et al.*, 2012). These results suggest that different TAL1 complexes have distinct DNA binding preferences.

## Effects of protein-protein interactions on TF functions

Interactions between proteins can influence DNA-binding specificity. Jolma *et al.* (2015) showed for 3,630 human TF-TF interactions that 315 of the tested interactions display cooperative DNA binding. Most TF pairs bind to both DNA-motifs with strong and specific orientation and/or spacing preferences. By binding to both motifs in concert, the number of potential genomic binding sites decreases, allowing more specific DNA binding. Moreover, many human TF pairs appear to bind DNA sites that differ considerably from their individual preferences (Jolma *et al.*, 2015). In addition, interactions can lead to recognition of novel motifs. This phenomenon is called latent specificity. An example of latent specificity through protein interaction comes from the anterior-posterior patterning *Drosophila* HOX proteins. The eight HOX homodimeric proteins bind to nearly identical sites *in vitro* (Noyes *et al.*, 2008). Despite overlapping DNA binding specificities, the HOX proteins have distinct functions *in vivo* (Rezsohazy *et al.*, 2015). Slattery *et al.* (2011) used SELEX-seq to analyse the binding specificity of the eight HOX proteins in complex with their shared cofactor Extradenticle (Exd). Using this method, the authors showed that the different HOX proteins acquire novel, unique DNA recognition properties when they bind DNA together with Exd (Slattery *et al.*, 2011).

Besides influencing the DNA binding specificity, interactions between proteins can also influence genomic TF binding in other ways (reviewed in Bemer *et al.*, 2017). Interactions can influence TF functionality by blocking or facilitating nuclear import (Balkunde *et al.*, 2011), by altering the stability of the TF (Rubio-Somoza *et al.*, 2014), or preventing DNA binding (Hao *et al.*, 2012). An example of an interaction that influences the TF localization is the nuclear trapping of TRANSPARENT TESTA GLABRA1 (TTG1) by GLABRA3 (GL3). TTG1 is a trichome-promoting factor in *Arabidopsis* leaves (Galway *et al.*, 1994; Walker *et al.*, 1999). By directly interacting with GL3, TTG1 is no longer able to move to other tissue layers ensuring correct trichome patterning during leaf organogenesis (Bouyer *et al.*, 2008; Pesch and Hulskamp, 2009; Balkunde *et al.*, 2011). Another example of an

effect of protein-protein interaction on TF functionality are the interactions between HLH and bHLH factors. The HLH TF subfamily lacks the basic domain required for DNA binding, and by forming heterodimers with bHLH proteins they inhibit DNA binding of the bHLH TFs (Hao *et al.*, 2012). Hence, protein interactions can influence DNA binding specificity and capability in many different ways.

Although many papers show that interactions with proteins and co-factors influence DNA binding specificity, most of the studies focus on a specific TF-target gene combination or are *in vitro*. More *in vivo* genome-wide data are needed to better understand the role of protein interactions on a genome-wide basis, and thereby getting a better understanding of TF specificity *in vivo*. Moreover, further studies into how closely related TF dimer combinations and other TF-complexes influence genomic binding will result in a better understanding of the mechanisms behind *in vivo* TF specificity.

## MADS domain TFs: one DNA binding consensus, many functions

The *Arabidopsis* genome encodes more than 1,500 transcription factors, many of these TFs belong to large families. The question how TFs of the same family that contain highly similar DBDs and hence bind comparable sequences *in vitro*, bind specific *in vivo* targets is therefore also highly relevant for plant TFs. As many studies into TF specificity are performed in animals and multicellular development originated independently in plants and animals, it is important to increase the knowledge on mechanisms underlying TF specificity in plants.

### Diverse functions of MADS domain proteins

In plants, a particularly large TF family is the MADS domain family. This protein family consists of over 100 members in *Arabidopsis,* many of which play diverse, critical roles in development. Well-known are the functions of MADS domain proteins in floral organ specification. Examples are the floral meristem identity factor AP1 (Alejandra Mandel *et al.*, 1992; Weigel *et al.*, 1992) and the petal and stamen homeotic proteins APETALA3 (AP3) and PISTILLATA (PI) (Goto and Meyerowitz, 1994; Jack *et al.*, 1994). Besides functions in floral organ development, MADS proteins function in many more developmental processes. For example in embryogenesis (e.g. AGAMOUS-LIKE15 (AGL15) (Heck *et al.*, 1995; Perry *et al.*, 1999), root development (e.g. SHATTERPROOF1 (SHP1) (Moreno-

Risueno *et al.*, 2010) and XAANTAL1 (XAL1) (Tapia-López *et al.*, 2008)) and flowering time (e.g. FLOWERING LOCUS C (FLC) (Michaels and Amasino, 1999), SHORT VEGETATIVE PHASE (SVP) (Hartmann *et al.*, 2000), and MADS AFFECTING FLOWERING 5 (MAF5) (Ratcliffe *et al.*, 2001); for a comprehensive review, see Smaczniak *et al.* (2012a)).

Besides the diverse functions of the individual MADS domain proteins, many MADS domain proteins have more than one function in seemingly unrelated processes. Examples of pleiotropic MADS domain proteins are SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) and FRUITFULL (FUL). SOC1 is best known for its role in the activation of flowering (Lee *et al.*, 2000; Seo *et al.*, 2009), and its regulation of the expression of the floral homeotic B-, C-, and E-class genes (Gregis *et al.*, 2009; Liu *et al.*, 2009). Additionally, downregulation of SOC1 and a second MADS domain protein FUL in the annual plant *Arabidopsis,* results in phenotypes common to perennial plants, hence SOC1 and FUL are important for maintenance of the annual life cycle of *Arabidopsis* (Melzer *et al.*, 2008). Like SOC1, also FUL has multiple functions during the life cycle of *Arabidopsis*. Firstly, during flower development FUL functions at two different stages. During the initiation of flowering, FUL regulates flowering time and meristem identity (Ferrandiz *et al.*, 2000b). Later during flower development, FUL controls fruit development by influencing valve, replum, and style morphology (Gu *et al.*, 1998; Ferrandiz *et al.*, 2000a). Besides the functions of FUL as annual life cycle regulator together with SOC1 (Melzer *et al.*, 2008) and during flower development, FUL also plays a role in determining cauline leaf morphology (Gu *et al.*, 1998), inflorescence architecture, and internode length (**Chapter 4**).

While these MADS proteins have many, diverse *in vivo* functions, all the family members appear to bind *in vitro* to a 10-bp motif called the CArG-box (consensus CC[AT]$_6$GG) (Pollock and Treisman, 1990; Schwarz-Sommer *et al.*, 1992; Huang *et al.*, 1993; Huang *et al.*, 1995; Tang and Perry, 2003). Genome-wide ChIP-seq studies have shown that MADS domain TFs bind hundreds to thousands of genomic loci. While part of the bound genomic regions overlap, also many regions are specific. In addition, these studies have shown that the majority of DNA sites bound do not contain a perfect CArG-box (Kaufmann *et al.*, 2009; Zheng *et al.*, 2009; Kaufmann *et al.*, 2010c; Deng *et al.*, 2011). To understand these results and to gain knowledge about the mechanisms determining TF-DNA specificity of MADS domain proteins, more research is needed on how MADS domain proteins achieve their DNA binding specificities *in vivo*.

## Protein interactions and binding specificity of MADS domain proteins

All MADS domain protein family members share a highly conserved, 56 amino acids long, MADS domain (Schwarz-Sommer *et al.*, 1990). This highly conserved domain makes contact with the DNA using its N-terminal and central parts. In the functionally best characterized class of plant MADS domain TFs, the type II MADS domain proteins, the MADS domain is followed by the I (intervening), K (keratin) domains, and the highly variable C-terminal domain. The I- and K-domains of the type II MADS proteins are essential for dimerization and higher-order complex formation, while the C-terminus is involved in protein complex formation and transcriptional regulation (reviewed by Kaufmann *et al.* (2005)). Since MADS domain proteins form complexes with other proteins, as discussed above, these protein complexes can influence *in vivo* binding of TFs.

MADS domain proteins bind DNA as obligate dimers and a large-scale interaction study has shown that these interactions are protein-specific (De Folter *et al.*, 2005), hence MADS dimerization is dependent both on the interaction specificity of a MADS protein and the *in vivo* availability of MADS binding partners. Most important for MADS dimerization specificity is the I-domain. Interestingly, domain-swap experiments have shown that this I domain, functioning in MADS dimerization, plays a role in DNA binding specificity as well (Riechmann *et al.*, 1996). This observation suggests that dimerization might play a role in *in vivo* DNA binding specificity of MADS proteins, however till now no *in vivo* studies have been conducted to elucidate the role of dimerization on MADS binding specificity.

Besides homo- and heterodimers, MADS domain proteins are also able to assemble into higher-order complexes (Egea-Cortines *et al.*, 1999; Honma and Goto, 2001), generating tetrameric protein complexes consisting of two dimers that bind to a target DNA sequence containing two CArG-boxes. By binding two CArG-boxes these tetrameric protein complexes generate a DNA loop between two binding sites. Although studies have shown that tetrameric MADS complexes are able to bind single CArG-boxes or a second low-affinity CArG-box, binding two CArG-boxes might provide stability to the complex and hence increase TF-DNA interaction stability (Melzer and Theißen, 2009; Smaczniak *et al.*, 2012b). The MADS domain proteins SEP1, SEP2, SEP3, and SEP4 play an important role in tetramer formation, but studies have shown that also other MADS domain proteins can mediate high-order complex formation (Immink *et al.*, 2009).

Besides interactions between family members, MADS proteins also interact with proteins from other families. These interaction partners include chromatin remodelling factors and co-repressors. Examples of MADS interactions with other protein families are the interactions with the co-repressors SEUSS (SEU) and LEUNIG (LUG) (Sridhar *et al.*, 2006). Interactions have been reported with chromatin regulatory proteins such as the Polycomb PRC1 analog TERMINAL FLOWER 2 (TFL2) and the SIN3 histone-deacetylase complex component SAP18 (Liu *et al.*, 2009). Interactions with those factors are proposed to influence the chromatin accessibility at bound loci resulting in transcriptional repression. Other interactions include the SWI2/SNF2-type ATP-dependent nucleosome remodellers BRAHMA (BRM) and SPLAYED (SYD), proteins that are involved in reversing Polycomb-mediated repression (Smaczniak *et al.*, 2012b; Wu *et al.*, 2012). A recent study on the relation between DNA accessibility and MADS domain protein DNA binding has shown that the MADS domain proteins AP1 and SEP3 are able to bind poorly accessible genomic regions. Moreover, this study shows that chromatin flanking MADS binding sites becomes more accessible upon MADS domain TF binding (Pajoro *et al.*, 2014). The interaction between chromatin remodellers and other non-MADS proteins might allow MADS domain proteins to bind poorly accessible DNA or may enable changes in DNA accessibility following MADS domain binding.

Despite these hypotheses about mechanisms of MADS domain protein action, currently no high-throughput study has been conducted to determine the influence of protein-protein interactions on DNA binding specificity of MADS domain proteins. Hence, studying the relation between MADS protein interactions and DNA binding specificity will provide insights into *in vivo* DNA binding specificity of MADS domain proteins and will increase our knowledge about *in vivo* DNA binding specificity of TFs.

## Scope of this study

The aim of the research described in this thesis is to study the influence of protein interactions on TF DNA binding specificity. The project focusses on the pleiotropic MADS domain TF FRUITFULL (FUL). As mentioned previously, FUL has distinct functions in the plant life cycle and has a dual role during flower development. The distinct functions of FUL suggest that this protein regulates different sets of target genes in different tissues, this makes FUL an excellent candidate to study mechanisms underlying target gene selection of MADS domain transcription factors.

In **Chapter 2**, we investigate the role of protein interactions on DNA binding specificity of MADS domain proteins. We show that gene regulation and genomic DNA binding by the MADS domain protein FUL is tissue-specific and that the observed differences relate to the tissue-specific functions. Tissue-specific immunoprecipitation experiments show that the interactions of FUL with MADS domain proteins and other proteins are partly tissue-specific. By studying the *in vitro* DNA binding specificities of the different tissue-specific FUL heterodimers and correlating these results with the *in vivo* genomic binding profiles, we were able to show that dimerization between MADS domain proteins influences DNA binding specificity and these differences correlate with the observed differences in *in vivo* genomic binding.

**Chapter 3** provides a step-by-step guide for ChIP-seq experiments and computational analysis designed for wet-lab biologists to perform the experiments and analyse their ChIP-seq data.

In **Chapter 4**, we study in more detail a direct FUL target gene, *SMALL AUXIN UPREGULATED RNA 10 (SAUR10)*. We show that *SAUR10* expression is regulated by FUL in multiple tissues, among others cauline leaves, stems, and branches. The expression of *SAUR10* at the abaxial side of branches is influenced by a combination of hormones, light conditions, and FUL. Additionally, we discuss several other FUL target genes involved in hormone pathways and light conditions. These data demonstrate the connection between development and environment in growth-regulated programs.

**Chapter 5** focusses on the putative direct targets of FUL in inflorescence meristem (IM) tissue. Among the putative direct targets two genes involved in flavonoid synthesis were identified, *FLAVONOID SYNTHESE 1 (FLS1)* and *UDP-GLUCOSYL TRANSFERASE 78D3 (UGT78D3)*. We report the first link between MADS domain protein FUL and flavonoids synthesis in *Arabidopsis*. Moreover, our results indicate a possible link between flavonoids and flowering time.

To conclude, in **chapter 6** discusses the results of this thesis and expand on possible further implications of the work performed in this thesis.

# Chapter 2

## Dual specificity and target gene selection by the MADS domain protein FRUITFULL

Hilda van Mourik
Jose M Muiño
Cezary D Smaczniak
Marian Bemer
Dijun Chen
Gerco C Angenent
Kerstin Kaufmann

Manuscript in preparation

## Abstract

How transcription factors of a single family confer different functional specificities *in vivo*, is an important question in molecular biology. Even more intriguingly, a single transcription factor can regulate context- or tissue-specific target genes to achieve distinct functions. Here we show, using a variety of genome-wide techniques, that gene regulation and DNA binding site selection by the MADS domain protein FRUITFULL (FUL) is tissue-specific. FUL has a dual role in regulating floral transition and fruit development. The tissue-specifically bound and regulated genes fit the known dual functions of FUL. Protein complex isolation in the two studied tissues shows that the interactions of FUL with other MADS domain proteins are tissue-specific. Using SELEX-seq we studied the DNA binding affinities of the different FUL-MADS dimers. Although all tested dimers prefer to bind to DNA sequences corresponding to the canonical binding motif of MADS domain proteins, the CArG-box, the dimers show different preferences for A and T nucleotides within the CArG-box and for A- and T-stretches surrounding the CArG-box. Mapping the SELEX-seq affinities to the genome allowed comparison with *in vivo* ChIP-seq data that were generated from the two different tissues. This analysis revealed a correlation between dimer-specific DNA-binding preferences and tissue-specific genomic binding site selection of FUL. Hence, we show that the choice of MADS dimerization partner influences DNA binding affinity, leading to differences in tissue-specific DNA binding.

# Introduction

Most studied plant transcription factors (TFs) bind to relatively short motifs, 8 – 12 bp (Franco-Zorrilla *et al.*, 2014; O'Malley *et al.*, 2016). Statistically, these motifs occur often in the genome, while ChIP-chip and ChIP-seq studies have shown that only a small portion of these motifs are bound *in vivo* (De Folter and Angenent, 2006; Heyndrickx *et al.*, 2014). Moreover, many TF families contain numerous members, all with different roles and functions. This raises the question how TFs of a single family achieve functional specificity.

The large MADS domain protein family, consisting of over 100 members in *Arabidopsis* (Parenicová *et al.*, 2003), is known for their broad array of functions during the entire life-cycle of the plant (reviewed by Smaczniak *et al.*, 2013)). MADS domain proteins bind regulatory elements in DNA through their highly conserved 56 amino acid N-terminal MADS domain. *In vitro* studies have shown that MADS domain proteins bind to a 10-bp DNA consensus sequence, $CC[AT]_6GG$, called the CArG-box (reviewed in De Folter and Angenent (2006)). Despite their overlapping DNA binding specificities, MADS domain proteins have distinct functions *in vivo* (reviewed by Smaczniak *et al.* (2013)). Although part of the specificity is achieved by cell type specific expression of different family members, ChIP-seq data show that MADS domain family members bind partly non-overlapping sets of target genes in the same tissue (reviewed by Yan *et al.* (2016)). Moreover, some MADS box genes acquired more than one function in different organs or developmental stages, suggesting that they regulate different sets of target genes in the different tissues.

MADS domain proteins can exclusively bind DNA as dimers, therefore interactions between MADS domain proteins are essential for their function (Schwarz-Sommer *et al.*, 1992; Huang *et al.*, 1996). De Folter *et al.* (2005) developed an interaction map of the *Arabidopsis* MIKC-type MADS TFs showing that different family members have specific sets of intra-family interaction partners. Clustering MADS proteins with similar interaction patterns grouped proteins with overlapping functions. The importance of protein-protein interactions is highlighted in several studies. For example for animal HOX proteins, showing that changes in protein-protein interactions can alter phenotypes (Löhr and Pick, 2005; Brayer and Lynch, 2011). In addition, in Snapdragon, variation in protein-protein interaction patterns between paralogous class C homeotic proteins allowed their evolutionary sub-functionalization in specification of male and female reproductive organs (Airoldi *et al.*, 2010). Although some studies addressed the influence of changes in protein-protein

interactions on TF DNA binding specificity (Slattery *et al.*, 2011; Kong *et al.*, 2012; Sayou *et al.*, 2014), the *in vivo* effect of dimer composition of TF on DNA-binding specificity is far from understood.

To study how and to what extent DNA binding specificity of MADS domain proteins is determined by dimer composition, this chapter focusses on one specific MADS domain protein: FRUITFULL (FUL). Throughout development of *Arabidopsis*, *FUL* is expressed in a broad array of tissues (See **Chapter 4** of this thesis), suggesting FUL to function at different stages during plant development. Well studied are the roles of FUL during flower development. In the inflorescence meristem (IM), FUL functions as a positive regulator of flowering and is important for meristem identity (Mandel and Yanofsky, 1995; Hempel *et al.*, 1997; Ferrandiz *et al.*, 2000b; Balanzà *et al.*, 2014). We will refer to this function as the "IM function" of FUL. At a later stage, FUL is expressed in the pistil where it is essential for correct patterning of the fruit (Gu *et al.*, 1998; Ferrandiz *et al.*, 2000a; Liljegren *et al.*, 2000; Roeder *et al.*, 2003; Liljegren *et al.*, 2004; Ripoll *et al.*, 2011), from now on referred to as the "pistil function" of FUL. The ability of FUL to regulate distinct processes, suggests that FUL might regulate different sets of target genes by binding to distinct regulatory DNA sequences in the two tissues. Moreover, FUL has been shown to interact with several MADS domain proteins (De Folter *et al.*, 2005; Smaczniak *et al.*, 2012b), making FUL an excellent candidate to study the role of MADS homo- and heterodimerization on TF specificity.

In this chapter we used high-throughput techniques to show that tissue-specific complex formation by MADS domain TFs influences DNA binding specificity. Using RNA- and ChIP-seq we demonstrate that FUL regulates partly different sets of target genes in IM and pistil, showing that DNA binding specificity of FUL is partly tissue-specific. Tissue-specific FUL-MADS dimers bind with different affinities to diverse CArG-box-like sequences. These *in vitro* determined dimer-specific binding specificities correlate with *in vivo* tissue specific DNA binding events determined by ChIP-seq. In summary, by combining multiple techniques we are able to obtain a better understanding of *in vivo* DNA binding site selection by MADS domain proteins and their heteromeric complexes.

# Results

## Tissue specific gene regulation by FUL

The MADS domain protein FUL functions in a tissue-dependent manner, raising the question whether FUL may regulate different sets of target genes in different tissues. To answer this question we studied the tissue-specific gene regulation of FUL using an RNA-seq approach, comparing gene expression in wild type versus *ful-1* mutant tissues. To determine tissue-specific gene regulation in the IM and the pistil we collected IM tissue including young flower buds, stages 1 to 9, and pistils from stage 12 to 16 flowers (stages according to Smyth *et al.* (1990)), respectively. Identification of tissue-specific differentially expressed genes (DEGs) resulted in a list of genes directly and indirectly regulated by FUL (**Supplementary file S1**). For the IM we identified a total of 45 significant DEGs, while many more DEGs were identified for the pistil, 521 DEGs (p-value ≤ 0.05 & |FC| ≥ 1.8) (**Figure 1A**). In both stages, the number of upregulated and downregulated genes is approximately the same (**Figure 1A**). Although our approach identifies both direct and indirect targets these results suggest that FUL may act both as a repressor and as activator of transcription. When comparing DEGs in IM and pistil tissue we found that 40% of the IM DEGs are also identified in the pistil, 27 IM DEGs are unique. For pistil tissue we identified 503 unique DEGs (**Figure 1B**). These results show that FUL is able to regulate different sets of genes in the IM versus the pistil.

A more detailed look at the differences in DEGs identified in IM versus pistil tissue was performed using gene ontology (GO)-term analysis on clusters of DEGs grouped based on differential expression. A heat map of 684 DEGs with an adjusted p-value ≤ 0.05 could be subdivided in seven clusters based on fold change (**Figure 1C**). Complete GO-term analysis can be found in **Supplementary file S2**. Genes having a comparable fold change in both tissues are found in clusters A, E, and F. Together, the small clusters A and F contain 12 genes that are upregulated in the mutant compared to wild type, including a gene related to cytokinin signalling *CYTOKININ RESPONSE FACTOR 4 (CRF4)* (Rashotte *et al.*, 2006). The phytohormone cytokinin is linked to meristem activity (Bartrina *et al.*, 2011), flowering (D'Aloia *et al.*, 2011), and pistil development (reviewed in Sehra and Franks (2015)). Down-regulated DEGs are found in cluster E, and the GO terms of this cluster are "response to auxin", "flavonoid biosynthesis" and "auxin transport". Like cytokinin, also auxin can be linked to many processes in plant development, including inflorescence development (Heisler *et al.*, 2005) and pistil development (Marsch-Martínez and de Folter, 2016;

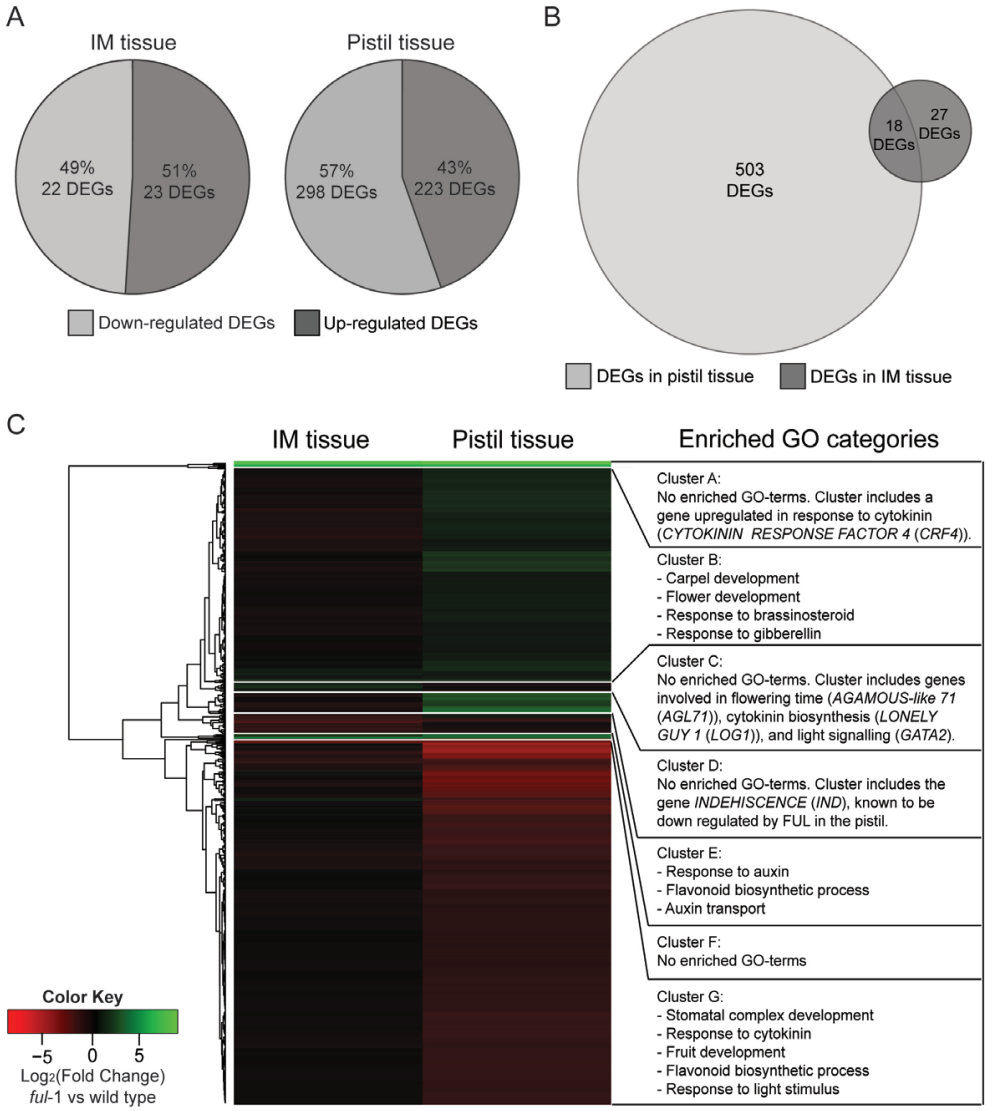



**Figure 1 | Tissue-specific gene regulation by FUL.** (**A**) Pie chart showing the number of genes differentially up- and down-regulated in *ful*-1 mutants compared to wild type for the IM and pistil tissues (|FC| ≥ 1.8 & adjusted p-value ≤ 0.05). (**B**) Venn-diagram showing overlap in DEGs of IM tissue and pistil tissue (|FC| ≥ 1.8 & adjusted p-value ≤ 0.05). (**C**) Transcriptional profile of 684 DEGs with an adjusted p-value ≤ 0.05 for genes expressed in *ful*-1 mutant compared to wild type in at least one of the tissues. Expression change is represented in $\log_2$ scale. A subset of GO terms enriched in the different clusters is shown and examples of representative genes are indicated (for full GO term analysis see **Supplementary file S2**).

Van Gelderen et al., 2016). The role of flavonoids in flower development is less well understood, but it has been proposed that flavonoids may negatively regulate auxin thereby ensuring the correct distribution of auxin (Brown *et al.*, 2001). Examples of auxin- and flavonoid-related genes in this cluster are *AUXIN-RESPONSE PROTEIN 3 (IAA3)*, *SMALL AUXIN UP RNA 79 (SAUR79)*, and *FLAVONOID SYNTHASE 1* (*FLS1).* Among the tissue-specific clusters, the small cluster C is specific for the IM. This cluster contains 14 genes upregulated in mutant IM tissues, while unchanged in the pistil. Among those, genes with functions in flowering time (*AGAMOUS-like 71 (AGL71))* (Dorca-Fornell *et al.*, 2011) and meristem determinacy (*LONELY GUY 1 (LOG1))* (Han *et al.*, 2014; Han and Jiao, 2015) can be found. Clusters D and G contain DEGs specific for the pistil function of FUL. These clusters contain genes like *INDEHISCENCE (IND)* and show overrepresentation of genes with the GO-term "fruit development", which have a direct link to the known function of FUL in pistil development (Liljegren *et al.*, 2004). In summary, FUL regulates different sets of targets genes in different tissues and the differences in target genes regulated reflect the known tissue-specific functions of FUL.

## Genome-wide mapping of FUL binding sites in IM and pistil

To test whether tissue-specific gene regulation by FUL is reflected by differences in *in vivo* DNA binding, ChIP-seq was used to identify genome-wide binding sites of FUL in the IM and pistil. ChIP-seq was performed using an anti-GFP antibody. For the IM function of FUL, IM tissue was collected from non-induced *ap1 cal* pAP1::AP1-GR plants expressing pFUL::FUL-GFP. Expression of pFUL::FUL-GFP in *ap1 cal* pAP1::AP1-GR is shown in **Supplementary figure S1**. To determine genomic binding sites of FUL in pistils, pistils of stage 12 – 16 flowers were dissected from pFUL::FUL-GFP *ful-1* plants (Urbanus *et al.*, 2009). Two biological replicates were generated for each ChIP-seq experiment. Pearson correlation between individual ChIP-seq experiments showed a Pearson correlation of 0.73 and 0.94 for the IM ChIP-seq and the pistil ChIP-seq, respectively (**Supplemental figure S2**). Following the methodology of other studies (Kaufmann *et al.*, 2009; Deng *et al.*, 2011; Gregis *et al.*, 2013), good correlation between biological replicates allowed for pooling of the mapped reads to obtain the final list of direct targets. In the IM we identified a total of 2538 significant binding sites assigned to 4055 genes (with peaks located 3 kb upstream of the transcription start site (TSS) to 1 kb downstream of the gene) (FDR ≤ 0.05). In the pistil, 2459 binding sites corresponding to 3598 genes were identified (FDR ≤ 0.05) (**Supplementary file S3** & **Supplementary table S1**). As seen for other MADS domain TFs (Kaufmann *et al.*, 2010c; Gregis *et al.*, 2013; Pajoro *et al.*, 2014), binding

sites were mostly found in promoter regions 5' of the TTS (**Supplementary figure S3**). A MEME-ChIP search for enriched sequence motifs in the significant peaks revealed a perfect consensus MADS domain TF-binding site (CArG-box) (Machanick and Bailey, 2011)(**Supplementary table S2**). The ChIP-seq data recovered genes previously associated with FUL function in IM and pistil. For example, the MADS box gene SUPPRESSOR OF *OVEREXPRESSION CONSTANTS 1 (SOC1)* has been shown to be a direct FUL target using ChIP-qPCR (Balanzà *et al.*, 2014), also in our hands *SOC1* is identified as a direct target of FUL. Other known potentially direct target genes of FUL identified in our ChIP-seq experiments include *SHATTERPROOF1* (*SHP1), SHP2*, and *IND* (Liljegren *et al.*, 2004). Two known genomic targets not identified in our ChIP-seq data are *MIR172C* and *LEAFY* (*LFY). LFY* is not significantly bound in our experiments, in contrast to Balanzà *et al.*, 2014 where FUL was shown to directly bind the *LFY* promoter. Balanzà *et al.* (2014) used a 35S::FUL-GFP line, causing high and ectopic expression of *FUL* in this line, which may lead to the identification of false-positive binding sites not occurring under native expression levels, possibly explaining the observed difference in results. In the pistil, direct binding of FUL to the promoter of *MIR172C* is shown by Ripoll *et al.* (2015) in inflorescences including open flowers and stage 13 -14 fruits, while our data does not show binding to the promoter of *MIR172C* most likely due to the difference in tissue used. However, we do find direct binding of FUL to the promoter of *APETALA2* (*AP2*) in the pistil, suggesting that FUL may regulate *AP2* expression both indirectly via *MIR172C* (Ripoll *et al.*, 2015) and directly via promoter binding. ChIP-seq peak patterns in *SHP1*, *SHP2*, *IND*, *SOC1* and *AP2* loci can be found in **Supplementary figure S4**. These results show that our ChIP-seq results are comparable to previously performed ChIP-seq experiments on MADS domain proteins and consistent with the current knowledge of FUL genomic binding, highlighting the good quality of our datasets.

## Tissue-specific putative FUL targets

To determine whether FUL binds DNA in a tissue-specific manner, the ChIP-seq results of IM and pistil tissue were compared. GO term analysis of the significantly bound genes in the IM and pistil ChIP-seq showed differences in enriched GO annotations. A selection of enriched GO terms is shown in **Figure 2A**, the complete list of GO-enriched GO-term can be found in **Supplementary file S4**. As expected, putative target genes in the pistil are enriched for GO terms related to fruit development and cell growth, while these terms are much weaker or not enriched among the IM putative targets. Although GO terms related to flowering time were enriched with similar

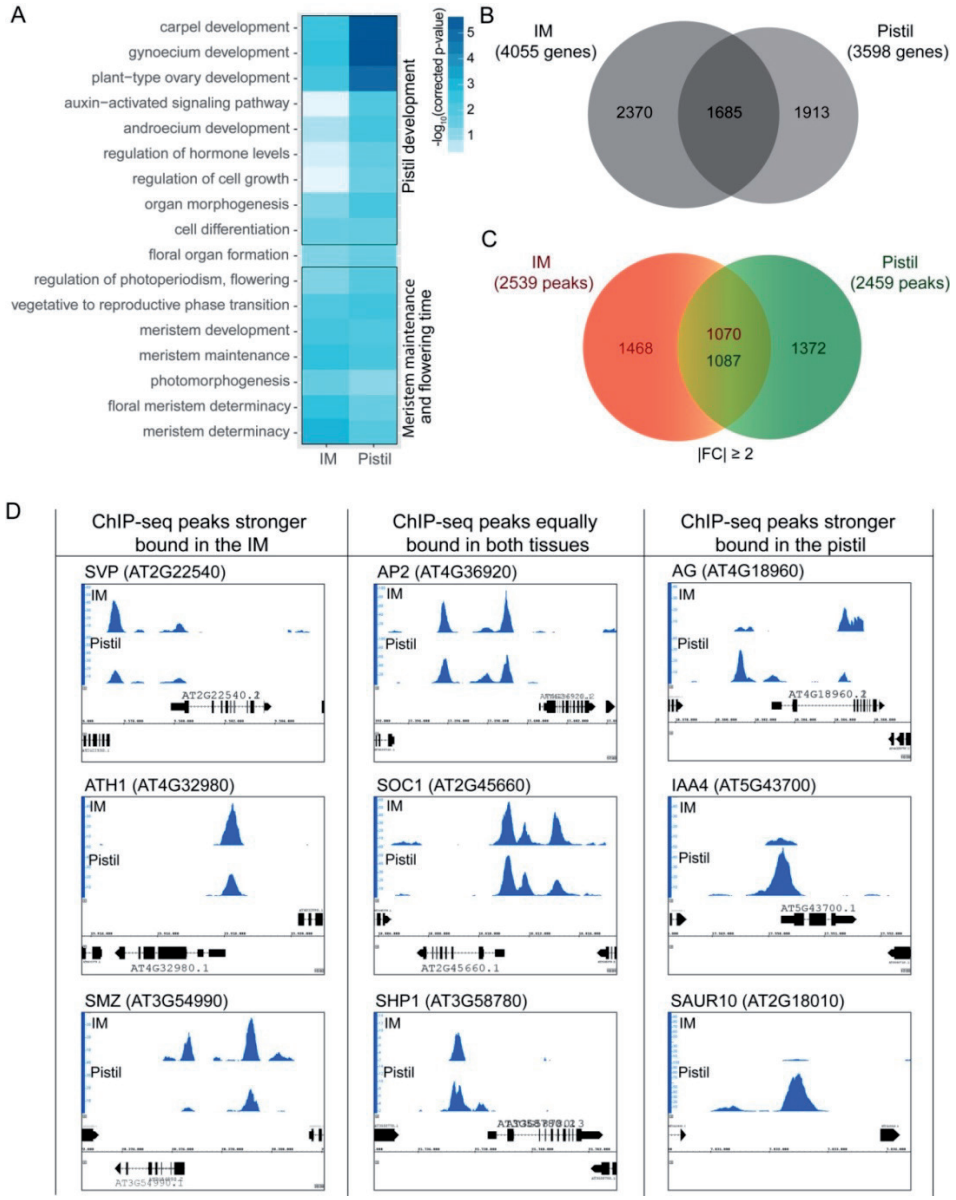
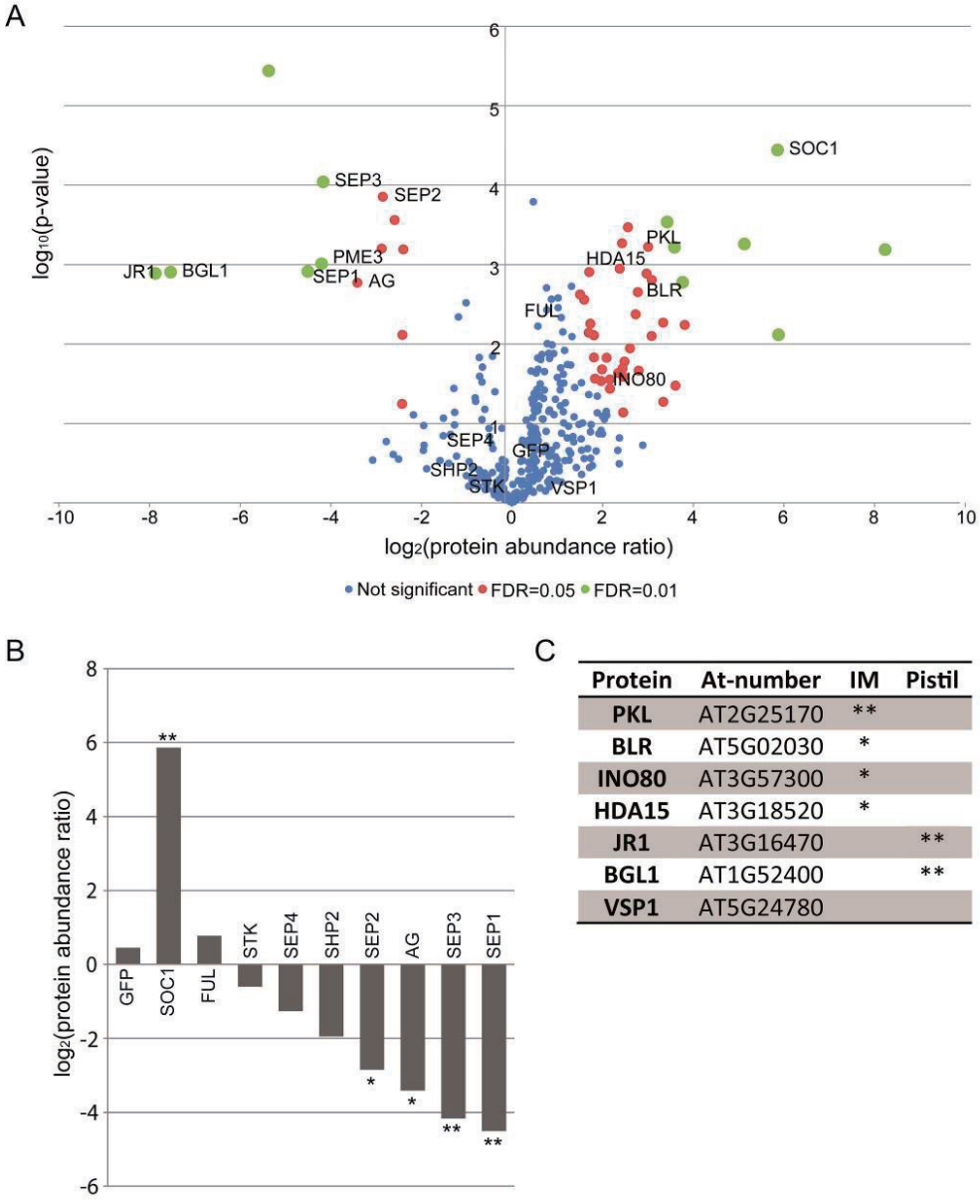



**Figure 2 | Tissue-specific DNA binding patterns of FUL.** (**A**) GO-term analysis of all putative target genes bound by FUL in the IM versus all putative target genes bound in the pistil (FDR ≤ 0.05). A selection of GO-terms is shown. (**B**) Venn diagram showing the overlap between significantly bound targets of FUL in the IM versus the pistil. (**C**) Quantitative analysis of peak heights between ChIP-seq binding peaks of FUL in the IM versus the pistil. Venn diagram shows the number of significant tissue specific binding sites that have at least a two-fold higher score compared to the other tissue. (**D**) Examples of peak profiles of FUL ChIP-seq peaks, peaks chosen based on quantitative analysis of peak height (FDR ≤ 0.05, |FC| ≥ 2).




C

| Protein | At-number | IM | Pistil |
|---|---|---|---|
| **PKL** | AT2G25170 | ** | |
| **BLR** | AT5G02030 | * | |
| **INO80** | AT3G57300 | * | |
| **HDA15** | AT3G18520 | * | |
| **JR1** | AT3G16470 | | ** |
| **BGL1** | AT1G52400 | | ** |
| **VSP1** | AT5G24780 | | |

**Figure 3 |** ***In planta*** **FUL protein interactions.** (**A**) Volcano plot of the IP results of FUL in IM versus pistil tissue. Protein abundance ratios between IM and pistil samples are given. (**B**) $Log_2$ protein abundance of FUL-MADS domain protein interactors in IM versus pistil. (**C**) List of proteins other than MADS domain proteins identified in either of the two IP experiments (complete list of proteins is given in **Supplementary file S5**). Asterisks indicate significant results (* p-value ≤ 0.05; ** p-value ≤ 0.01).

significance in both tissues, a stronger enrichment for GO terms linked to meristem maintenance and determinacy was observed in the IM. A closer look at the overlap of the putative targets in the IM and pistil shows that 2370 genes are specific for the IM and 1913 specific for the pistil, and 1685 genes are bound in both tissues. To get a better idea of quantitative differences in peak height between the two tissues, we performed a quantitative analysis of the ChIP-seq peak heights of peaks that are significant in at least one of the datasets. In total 6085 peaks were identified to be significantly enriched in at least one of the two tissues. Of these 6085 genes, 1468 peaks were identified to have a 2-fold higher peak score in IM tissue compared to pistil tissue. For pistil tissue, 1372 peaks were identified to be 2-fold higher compared to IM tissue (**Figure 2C**).

Examples of tissue-specifically bound putative targets are shown in **figure 2D**. Among the genes stronger bound by FUL in the IM we identified *SHORT VEGETATIVE PHASE (SVP) (Hartmann et al., 2000; Balanzà et al., 2014)*, *HOMEOBOX GENE 1 (ATH1)* (Gómez-Mena and Sablowski, 2008; Rutjens *et al.*, 2009) and *SCHLAFMUTZE (SMZ) (Mathieu et al., 2009)*, known for their roles in floral transition. Genes stronger bound in the pistil are the MADS-box gene *AGAMOUS (AG)* and *INDOLE-3-ACETIC ACID INDUCIBLE 4 (IAA4)*. *AG* is well known for its role in pistil development (Yanofsky *et al.*, 1990). IAA4 is an auxin inducible protein and plays a role in auxin signalling to control pistil development and fruit initiation (Marsch-Martínez and de Folter, 2016; Van Gelderen *et al.*, 2016). Another auxin related gene stronger enriched in the pistil ChIP-seq is *SAUR10*, more information about this gene can be found in **Chapter 4** of this thesis. These results in combination with the RNA-seq data show that FUL is indeed able to bind and regulate genes in a tissue-specific manner and that the regulation is linked to the tissue-specific functions of FUL.

## Stage-specific protein interactions of FUL

To study the role of MADS domain protein complex formation in tissue-specific DNA binding specificity, we determined the tissue-specific protein-protein interactions of FUL *in planta*. IMs of gAP1-GR *ap1 cal* gFUL-GFP plants were collected to identify meristem-specific protein complexes of FUL, and stage 12 - 16 pistils of gFUL-GFP *ful-1* plants were collected for the pistil complexes. Protein complexes were isolated by immunoprecipitation using anti-GFP antibodies and protein identification was performed using LC-MS/MS, followed by label-free protein quantification (Smaczniak *et al.*, 2012c). This approach allowed identification of tissue-specific protein interaction partners of FUL. Multiple MADS domain proteins have been shown to interact with FUL based on yeast-two-hybrid studies, BiFC experiments and IP studies (De Folter *et al.*, 2005; Smaczniak *et al.*,

2012b; Balanzà *et al.*, 2014) (**Supplementary table S3**). Our tissue-specific IP data showed that FUL is indeed able to interact with several MADS domain proteins *in planta* (**Figure 3A**; **Supplementary file S5**). In total, eight MADS domain proteins were identified in the IM and pistil datasets: AG, SEPALLATA 1 (SEP1), SEP2, SEP3, SEP4, SHP2, SOC1, and SEEDSTICK (STK). So, FUL is able to interact *in vivo* with multiple different MADS domain proteins.

When comparing MADS domain protein abundances between the IM and the pistil complex isolation data, SOC1 was identified as predominant FUL partner in the IM, while in the pistil we found higher levels of AG, SEP1, SEP2, and SEP3 in the FUL complexes (**Figure 3B**). The difference in protein complex abundance follows the expression levels in IM and pistil tissue for most interaction partners (**Supplementary figure S5**). However, protein abundance in the FUL complexes between IM and pistil is not significantly different for SHP2 and STK, while *SHP2* and *STK* expression is significantly higher in the pistil compared to the IM. This difference is likely due to the limited overlap in expression domain of *FUL* and *SHP2* and *STK* within the pistil, these proteins may interact with FUL only in very few cells reflecting the low abundances of these proteins in our IP data. In conclusion, *in planta* interactions of FUL with other MADS domain proteins confirm several previously reported FUL interactions, and show clear tissue-specific differences between the IM and the pistil.

## Non-MADS interaction partners in the ChIP-seq and the IP data

Besides MADS domain proteins, also members of other protein families are identified by our LC-MS/MS based complex isolation experiments (**Figure 3C**). In the IM, we found FUL to interact with histone modifying and chromatin remodelling factors PICKLE (PKL), INO80, HISTONE DEACETYLASE 15 (HDA15), and the homeodomain TF BELLRINGER (BLR). Several of these factors were previously identified as MADS protein interactors (Smaczniak *et al.*, 2012b). Many of the proteins that interact with FUL are directly or indirectly linked to flowering time: *pkl*, *ino80,* and *blr* loss-of-function mutants are late flowering (Kanrar et al., 2008; Andrés et al., 2015; Zhang et al., 2015; Fu et al., 2016) and several HDA15 homologs have been linked to flowering (reviewed in Liu et al. (2014)). In the pistil, the IP experiments showed enrichment for the hormone related factors JASMONATE RESPONSIVE 1 (JR1) and BETA-GLUCOSIDASE HOMOLOG 1 (BGL1). VEGETATIVE STORAGE PROTEIN 1 (VSP1) is identified in both tissues. VSP1 has been shown to interact with the MADS domain protein AG and is thought to be involved in nutrient storage, plant defence, and flower development (Gamboa *et al.*, 2001; Berger *et al.*, 2002; Guerineau *et al.*, 2003).

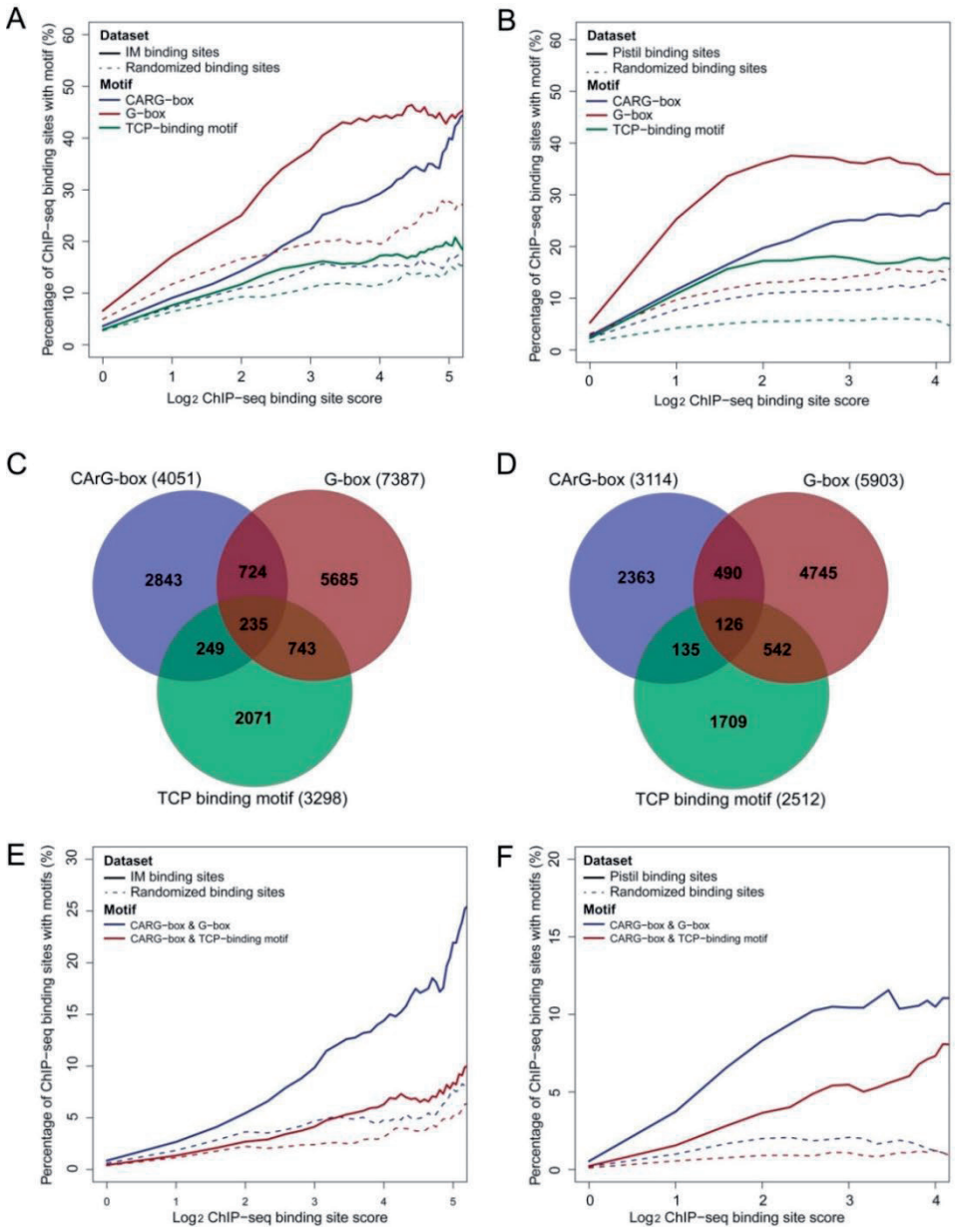



**Figure 4 | Motif enrichment in ChIP-seq datasets.** Enrichment of CArG boxes (CC[AT]$_6$GG), G-boxes (CACGTG), and TCP binding consensus sites (GGNCCCAC & GTGGNCCC) in ChIP-seq data of the (A) IM and (B) pistil. Venn diagram showing the number of motifs found in complete ChIP-seq binding peak list (FDR ≤ 0.05) for the (C) IM and the (D) pistil ChIP-seq data. Enrichment of binding sites containing both a CArG-box and a G-box or TCP binding motif for the (E) IM and (F) pistil ChIP-seq data. Dashed lines show enrichment in a control set containing binding sites with the same nucleotide content in a randomized order.

MEME-ChIP motif search performed on both ChIP-seq experiments, using 100 bp around the peak summit (FDR ≤ 0.05), showed enrichment of a motif other than a CArG-box, possibly pin-pointing to another FUL interaction partner. In both studied tissues we identified a GA-rich motif enriched (**Supplementary table S2**). Previous MADS domain ChIP-seq data also reported this GA-rich sequence (Pajoro *et al.*, 2014), and candidate proteins that bind this sequence are BASIC PENTACYSTEINE (BPC) transcriptional regulators that control multiple aspects of plant development (O'Malley *et al.*, 2016). It has been shown that BPC proteins interact with MADS domain proteins to regulate their targets (Simonini *et al.*, 2012). In order to search for known TF binding motifs in the ChIP-seq data, a TOMTOM motif comparison search was performed (Gupta *et al.*, 2007) using the complete sequence underneath ChIP-seq binding peaks (FDR ≤ 0.05). As expected, for both tissues CArG-box-like motifs were discovered, however also other motifs (**Supplementary table S3**). TOMTOM identified a G-box motif and a TCP-binding motif in the two tissues. Both motifs have been shown to be enriched in other MADS protein ChIP-seq datasets as well (Kaufmann *et al.*, 2009; Deng *et al.*, 2011; Tao *et al.*, 2012). To determine the enrichment of these motifs in the ChIP-seq datasets we visualized the motif enrichment compared to a randomized dataset (**Figure 4A and B**). These enrichment studies show that G-boxes are highly enriched in both tissues, while TCP binding motif enrichment is specific for the pistil binding sites. We found that approximately 25% of binding sites in either tissue containing a CArG-box also contain a G-box and/or a TCP binding motif (**Figure 4C and D**). Again we found for the IM a clear enrichment of the CArG-box and G-box motif combination in one binding region, while the CArG-box and TCP binding site combination does not show a clear enrichment (**Figure 4E**). For the pistil binding region our data shows a clear enrichment of CArG-boxes in combination with either G-boxes or TCP binding sites. These data suggest that FUL might cooperatively bind promoter regions with bHLH and TCP proteins (**Figure 4F**).

## DNA binding specificities of heteromeric FUL complexes

The DNA binding capacity of FUL dimers was tested using Electrophoretic Mobility Shift Assays (EMSAs) using a *SEP3* promoter probe containing two CArG box sequences previously published by (Smaczniak *et al.*, 2012b). ChIP-seq data showed that FUL strongly binds the *SEP3* locus at the position of the probe in both the IM and the pistil, confirming the *SEP3* probe to be suitable for testing the DNA binding of FUL dimers (**Supplemental figure S6**). EMSA results confirmed that

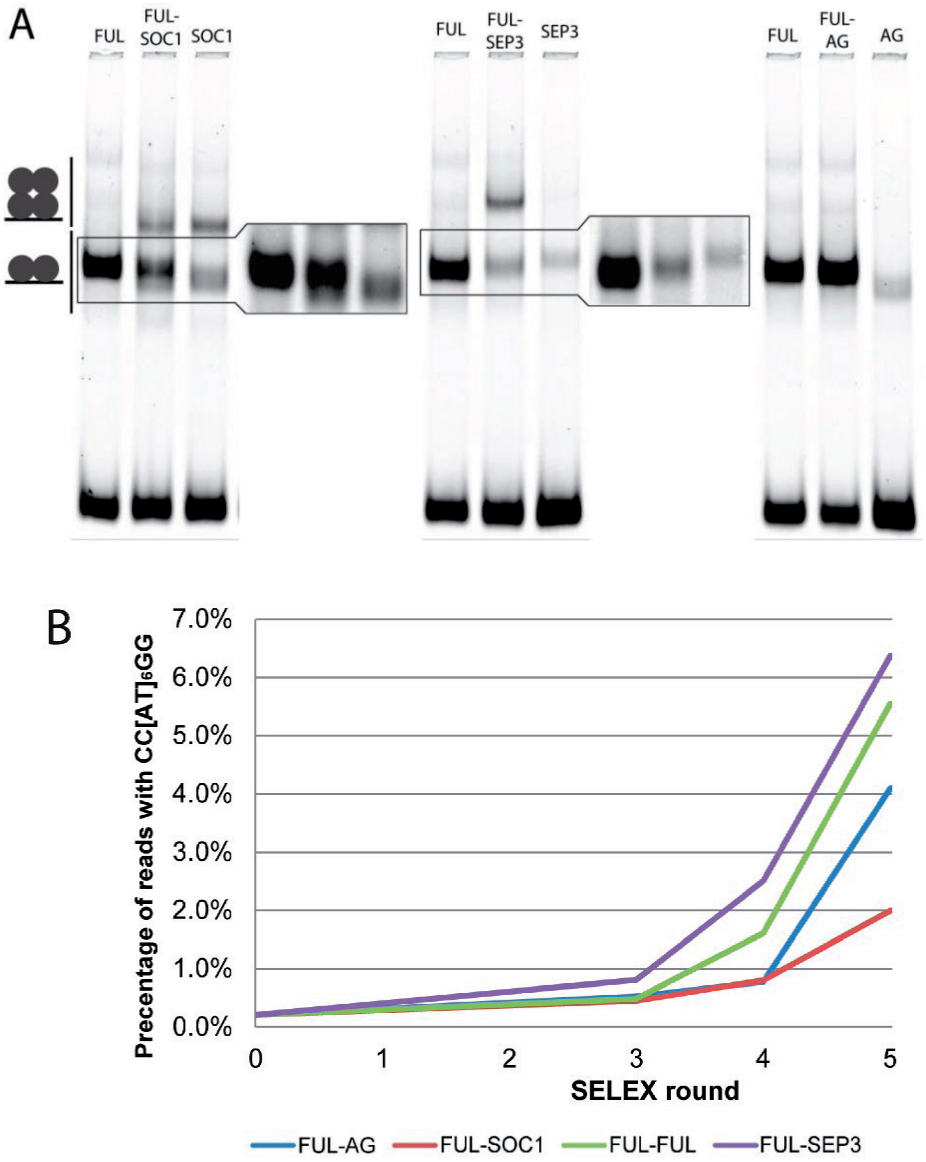



**Figure 5 | FUL dimer DNA binding and SELEX-seq CArG-box enrichment.** (**A**) Analysis of binding of FUL homo- and heterodimers to the *SEP3* promoter probe and model representation of formed protein-DNA sequences. Separate gels are shown for FUL/SOC1, FUL/SEP3, and FUL/AG. Next to the gels an enlargement of the gel. (**B**) Enrichment of reads with a canonical CArG-box consensus sequence in sequential SELEX-seq rounds.

FUL-SOC1, FUL-SEP3, FUL-AG, and FUL-FUL are able to bind DNA as a dimer, and at least in case of FUL-SOC1 and FUL-SEP3, may also form a tetrameric complex (**Figure 5A**).

To determine whether tissue-specific DNA binding of FUL reflects differences in MADS dimer specific DNA binding affinities, we used SELEX-seq. For the SELEX-seq approach we incubated *in vitro* translated MADS domain proteins with a dsDNA library containing a region of 40 random nucleotides (Jolma *et al.*, 2010). After incubation, FUL-DNA complexes were isolated using immobilized FUL-specific antibody. The resulting dsDNA was amplified and used for a subsequent round of SELEX. Five rounds of SELEX enrichment were performed in total. Complex-specific DNA binding specificity was measured by SELEX-seq for the IM specific complex FUL-SOC1 and the pistil specific complexes FUL-SEP3 and FUL-AG. As a control the non-tissue specific FUL-FUL complex was used to ensure that the resulting DNA-binding specificities are indeed derived from the FUL heteromeric complexes. Several rounds of incubation of *in vitro* translated MADS domain proteins with the dsDNA libraries resulted in enrichment of CArG-boxes for all protein combinations (**Figure 5B**), confirming the quality of our data (**Supplementary file S6**).

SELEX-seq relative affinities of the tested FUL-MADS dimer are visualized in a heat map (**Figure 6**), which shows clear clusters of k-mers specific to the different dimers. Clusters A to D are dimer-specific clusters, clusters E and F are found for all libraries. It is possible that cluster E and F partly consist of FUL homodimeric affinities that will be present in all libraries. Clusters A to D show clear differences in affinity between the FUL-FUL complexes and the other libraries, ensuring these affinities to arrive from heteromeric complexes. To visualize k-mer sequences in the different clusters, we extracted the 40 nucleotide full sequences and performed motif enrichment analysis using MEME-ChIP (Bailey *et al.*, 2009) (**Figure 6**). Although the motifs seem similar, there are important differences between them. The FUL-homodimer motif (cluster A) is unique in its strong preference for an A at position 7 of the CArG-box, where other dimers are more flexible in the use of an A or T nucleotide at this position. Like most other dimers, FUL-FUL prefers an A at position 3 of the CArG-box, but FUL homodimer seems to have the strongest preference for this nucleotide. As for the homodimer, many clusters have a preference for a T-stretch before and/or an A-stretch after the CArG-box, suggesting the binding motif of FUL-dimers to exceed the 10-bp CArG-box. Cluster B is linked to the FUL-SOC1 dimer, and the corresponding FUL-SOC1 motif is less strict in the need for a G at position 9 in the CArG-box, suggesting that CArG-boxes with a single G and a longer AT-stretch can also be bound by this complex. In contrast to other dimers, this complex also does not have the preferred T-stretch before the CArG-box as seen for clusters A, D, E, and F.

Cluster C, specific for FUL-SEP3 is unique in that it does not require the A and T stretches flanking the CArG-box which is seen for most other clusters. Moreover, like FUL-SOC1 this motif shows that the G at position 9 of the CArG-box can be replaced by an A, resulting in a longer AT-stretch and a single G in the CArG-box. Cluster D, specific for FUL-AG, is much less conserved in its preference for A or T nucleotides in the CArG-box. While the other dimers have a high preference for an A at position 3, this preference is not seen for FUL-AG. Also, the strong preference for a T at position 8 identified for most dimers is not found for FUL-AG. We found that FUL-dimer high affinity binding preferences are complex specific (**Supplementary figure S7**), suggesting that for the FUL-dimers, high affinity sites confer specificity. These data show that although all tested dimers preferably bind CArG-boxes, differences in DNA binding affinities are found for the FUL MADS dimers.

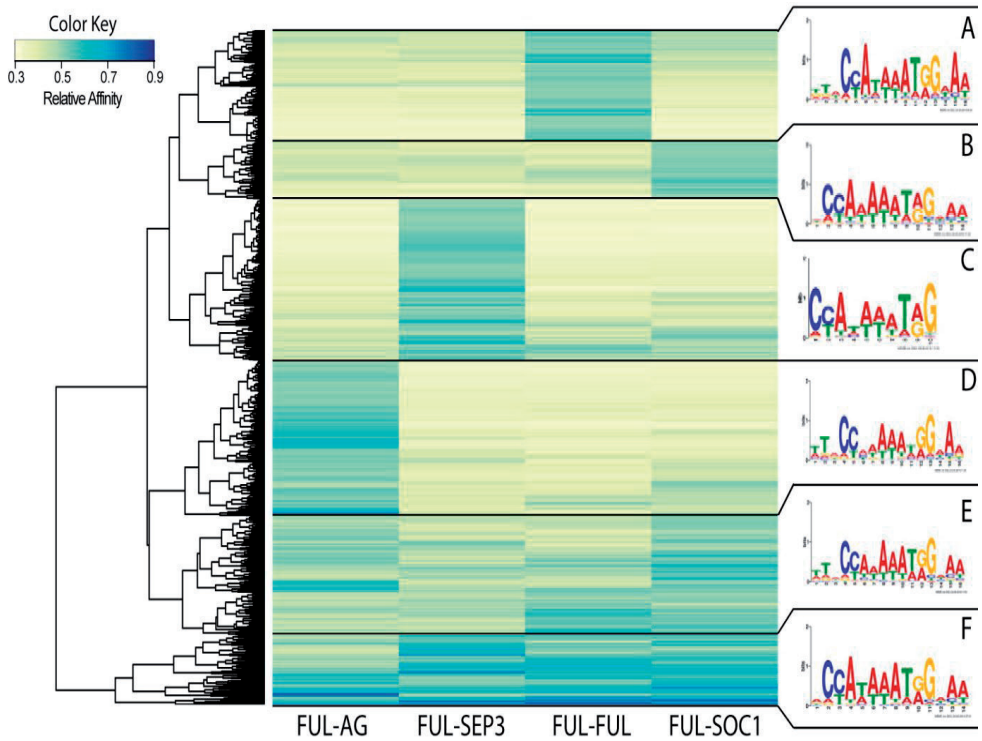



**Figure 6 | FUL-MADS dimer DNA binding specificities.** Heat map of 10mer normalized affinities enriched in round 5 of the SELEX with an affinity of at least 0.5 for one of the dimers plus corresponding sequence motifs build from 40N reads containing these 10mers.

## Complex-specific DNA binding preferences contribute to tissue specificity of *in vivo* DNA binding events

Different DNA binding preferences between FUL complexes could possibly explain the observed differences in *in vivo* DNA binding determined by ChIP-seq. To allow comparison between *in vivo* ChIP-seq data with *in vitro* SELEX-seq data, SELEX k-mers were mapped to the *Arabidopsis* genome generating a genome-wide SELEX-seq dataset. To verify mapping of the SELEX-seq to the genome we performed EMSA experiments on the promoter of *SCHLAFMUTZE (SMZ)*. In the *SMZ* promoter we find two ChIP-seq peaks of different height. Mapped SELEX-seq affinities fit with the observed differences in ChIP-seq peak heights, we find the peak closest to the gene to have a lower ChIP-seq peak score and a corresponding lower SELEX-seq affinity compared to the second binding site in the same promoter for all mapped SELEX-seq libraries (Figure 7). In agreement with these data, EMSA experiments confirmed that indeed the first binding site (probe 1A for *SMZ*) is weaker bound *in vitro* compared to the second binding site (probe 1B for *SMZ*), confirming the quality of our SELEX-seq results.

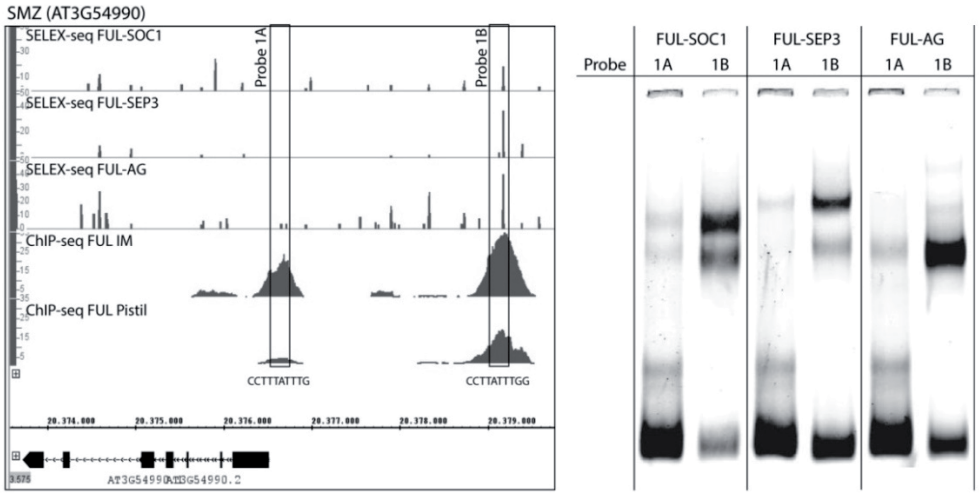



**Figure 7 | Confirmation of *in vitro* SELEX-seq data.** ChIP-seq peak profiles and SELEX-seq mapped relative affinity data of the *SMZ (AT3G54990)* promoter, boxes indicate the probe locations. (Right panel) EMSA results of FUL-dimers binding the two *SMZ* promoter probes.

More examples of SELEX-seq mapped relative affinity data can be found in Figure 8 where the promoter regions of *SVP*, *AG*, and *IAA4* are shown. Like for *SMZ*, also for these genes we see a good alignment of the SELEX-seq 'peaks' with *in vivo* ChIP-seq binding sites. For *SVP* we found a

higher ChIP-seq peak score in the IM, compared to the pistil. In agreement with the ChIP-seq data we find the highest SELEX-seq affinity for the IM-specific complex FUL-SOC1. Similarly, for *AG* and *IAA4* a higher ChIP-seq peak score is found for the pistil and indeed the pistil specific complex FUL-AG contains the highest affinity for this binding site. These data are in line with the high correlation found between SELEX-seq affinity ratios and ChIP-seq score ratios.

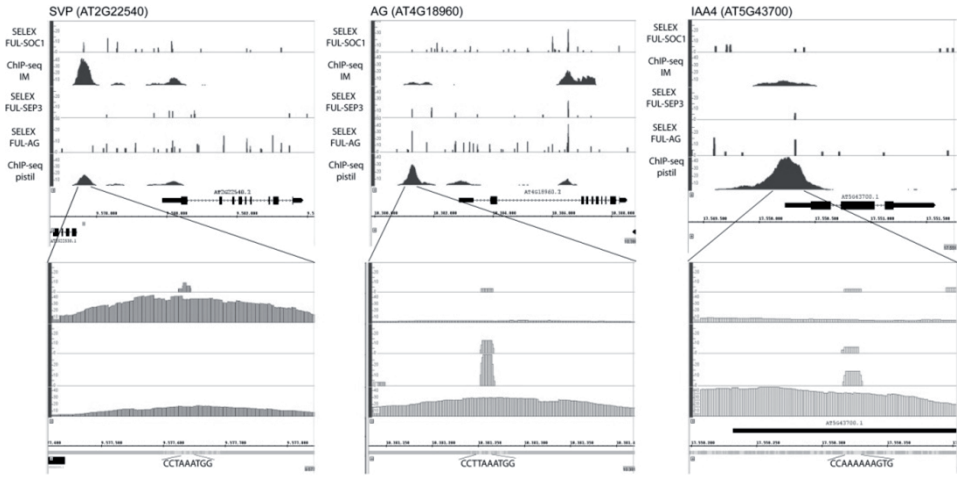



**Figure 8 | Examples of differentially bound ChIP-seq targets predicted by differences in SELEX-seq relative affinities.** Promoter regions of *SVP*, *AG*, and *IAA4* with ChIP-seq peaks and FUL-dimer binding sites predicted by SELEX-seq. At the bottom a zoom of the center of the ChIP-seq binding site, including SELEX-seq prediction and corresponding CArG-box sequence.

Although overall a good fit was found between ChIP-seq binding and SELEX-seq affinity peaks we also found discrepancies between the two datasets. An example is the *SMZ* locus, where the DNA binding affinities of the stage-specific complex FUL-SOC1 does not fit with the IM stage-specific binding patterns observed in our ChIP-seq data. Besides the SELEX-seq binding peaks found within ChIP-seq enriched regions, we also see that many SELEX-seq 'peaks' do not correspond with a ChIP-seq binding event, most likely due to additional *in vivo* mechanisms not accounted for in our *in vitro* approach, e.g. control of binding site accessibility by chromatin structure. Furthermore, not all ChIP-seq peaks contain a CArG-box or only a CArG-box-like sequence with low affinity, suggesting that cooperative DNA binding of a MADS domain protein with each other in higher-order complexes, or with other co-factors can change binding affinities. These results show that although in general SELEX-seq can predict tissue-specific DNA binding, individual predicted binding sites alone need to be interpreted with caution.

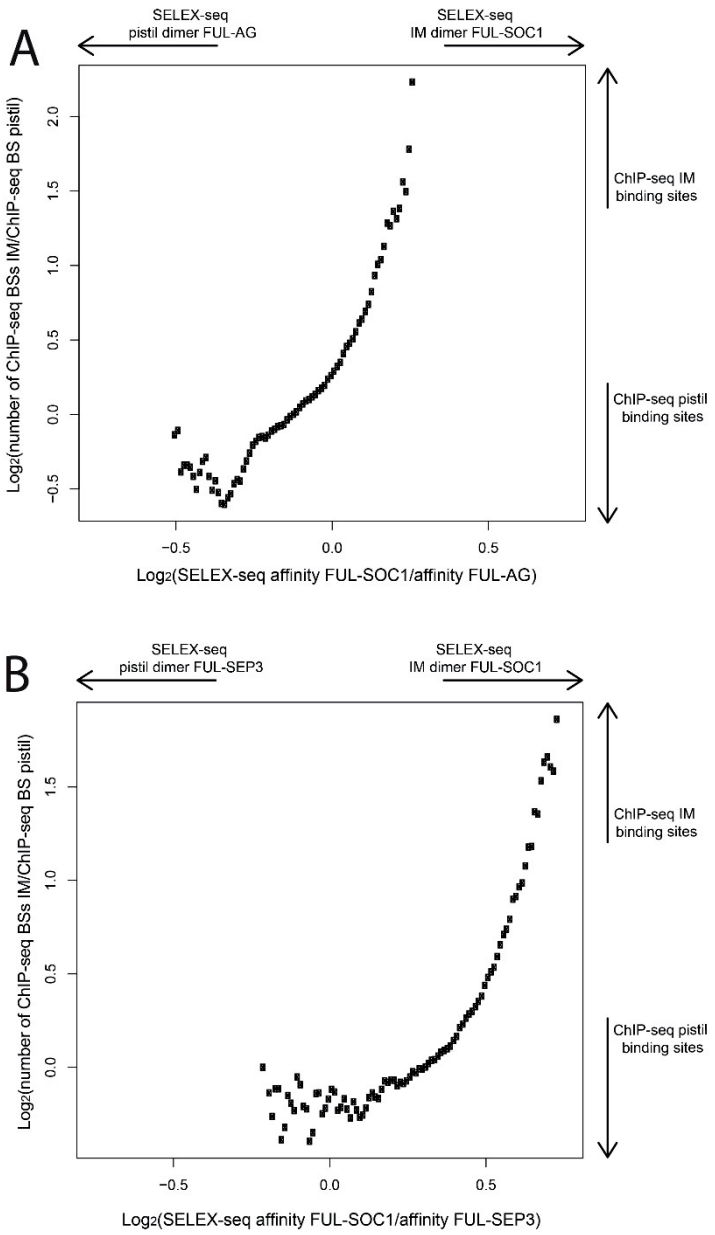



**Figure 9 | Correlation between SELEX-seq affinities and ChIP-seq DNA binding.** Ratio of number of ChIP-seq peaks in comparison with the difference in SELEX-seq affinities between the IM specific dimer (FUL-SOC1) and the pistil specific dimers (**A**) FUL-AG and (**B**) FUL-SEP3.

Because we observed in general a good fit of the *in vitro* SELEX-seq affinities with *in vivo* ChIP-seq binding sites we tested whether the tissue-specific FUL-dimer affinities globally correlate with the stage-specific ChIP-seq data. For this we compared the ChIP-seq peak height ratios in the IM versus the pistil with the SELEX-seq affinity ratios between different FUL-dimers. This comparison resulted in a high correlation between tissue-specific dimer DNA-binding affinities and tissue-specific ChIP-seq binding. For affinities of both FUL-AG and FUL-SEP3, specific for the pistil, compared to the IM specific FUL-SOC1 dimer, we found a highly significant correlation with the stage-specific ChIP-seq results (**Figure 9**, p-value $\leq 2 \cdot 10^{-16}$). These data show that DNA-binding specificity can predict DNA binding *in vivo*, and strongly point towards a role of MADS protein interactions in tissue-specific gene regulation by the MADS domain protein FUL.

# Discussion

How TFs obtain different context-specific functions *in vivo* is an important question. In this chapter we show that gene regulation and genomic DNA binding by the MADS domain protein FUL differs for different tissues and that these observed differences relate to the tissue-specific functions of FUL. Besides DNA binding, also the interaction partners of FUL are tissue-specific. Our combined results suggest that combinatorial gene regulation with non-MADS TFs on the one hand, and MADS protein complex-specific DNA binding preferences on the other hand, contribute to tissue specificity of FUL functions. We also identified tissue-specific differences in the recruitment of other regulatory proteins to FUL complexes, which also affect the target gene repertoire.

## Tissue-specific regulation of gene expression and DNA binding

FUL is expressed at two stages during reproductive development, this dual expression pattern results in two distinct functions of FUL: a role in regulating floral transition and a role in fruit development. In this chapter we demonstrate using RNA-seq that FUL regulates different sets of target genes in the two tissues. A core set of DEGs is regulated at both stages, but most genes are unique for the different tissues. The differences in genes regulated reflect the known functions of FUL at both stages. When comparing the DEGs in both stages, it is clear that the total number of DEGs in both tissues is very different, 45 for the IM and 521 for the pistil. The low number for IM is most likely due to the redundancy between *FUL* and the MADS-box gene *SOC1* in the IM (Torti and Fornara, 2012; Balanzà *et al.*, 2014), which limits the number of genes that show altered expression in the *ful* mutant IM. The high number of DEGs in the pistil may partly be the result from the identity change of valve cells into valve margin cells in *ful-1* mutant pistils as the used approach identifies both direct and indirect targets of FUL. To identify putative direct targets of FUL we used ChIP-seq. Similar to what was found in the RNA-seq, a core set of genes is bound at both stages, but we also identified a large number of tissue-specific putative targets. Other data on MADS domain protein DNA binding confirm that a core set of genomic targets is conserved among the studied MADS domain proteins, but also supports the idea that each MADS domain protein binds a protein specific set of targets (reviewed in Yan *et al.* (2016)).

## FUL protein interactions are tissue-specific

Interactions between MADS domain proteins are essential for their function. Our tissue-specific IP data showed that FUL is able to interact with several MADS domain proteins *in planta:* AG, SEP1, SEP2, SEP3, SEP4, SHP2, SOC1, and STK. Yeast-two-hybrid data confirmed the interactions of FUL with AG, SEP1, SEP3, SEP4, and SOC1, suggesting that FUL is able to directly interact with these proteins. No interactions were found between FUL and SHP2 or STK in yeast-two-hybrid, therefore FUL-SHP2 and FUL-STK most likely do not form dimers but might be found in higher-order complexes, possibly via SEP3 as all three proteins have been shown to interact with SEP3 (Immink *et al.*, 2009). No yeast-two-hybrid studies have been performed testing the interaction between SEP2 and FUL. SVP has been shown to interact with FUL in BiFC experiments in tobacco leaves (Balanzà *et al.*, 2014), but our complex isolation experiments did not identify interactions between SVP and FUL, most likely due to limited overlap in expression domain (Gregis *et al.*, 2009; Urbanus *et al.*, 2009). Similarly, previously detected interaction with the MADS domain proteins AGL6, AGL14, AGL21, and AGL24 in Y2H were not detected in this study possibly due to limited overlap in gene expression domain with *FUL* in the tested tissues (reviewed in Alvarez-Buylla *et al.* (2010)). *In vivo* IP experiments in complete inflorescences showed that FUL interacts with AP1, but also SOC1 was identified as interaction partner (Smaczniak *et al.*, 2012b). In line with the results of Smaczniak *et al.* (2012b), we indeed identified SOC1 as interaction partner in the IM. Due to the use of the *ap1* mutant plants, AP1 could not be detected as partner. However, *AP1* expression in the IM is limited, *FUL* and *AP1* expression overlaps transiently in stage 1 to 2 floral buds (Urbanus *et al.*, 2009), hence we do not expect FUL-AP1 complexes to be of high importance for the function of FUL in the IM. Comparison of the interactions in both tissues showed that FUL interacts tissue-specifically with MADS domain proteins. Expression data of these MADS interactors suggested that the observed differences in MADS complexes are most likely caused by differences in expression levels of the different interactors, although it should be noted that RNA levels are not always a direct measure for the resulting amount of produced protein.

Besides MADS domain protein interactions, FUL also interacts with other proteins. The IP data showed interactions with several histone modifying and chromatin remodelling proteins, partly confirming previous results (Smaczniak *et al.*, 2012b). Our results showed that interaction between FUL and these proteins is tissue-specific, hence it could be hypothesized that differences in non-MADS FUL interactions may in part influence tissue-specific DNA binding. To get a more complete view on other factors that could influence DNA-binding or gene regulation by FUL we studied

motif enrichment in the tissue-specific DNA-binding sites of FUL. As expected many CArG-box-like motifs were identified, however also motifs similar to the bZIP binding motif, the G-box, and the TCP binding motifs were identified. The G-box is enriched in ChIP-seq data of IM and pistil tissue, while the TCP binding motifs are specifically enriched in the pistil ChIP-seq dataset. We also showed a significant enrichment of ChIP-seq binding sites containing both a CArG-box and a G-box or TCP binding site. Although we did observe clear motif enrichments for G-boxes and TCP binding motifs, the IP data did not show interactions between FUL and G-box binding proteins or TCPs. It is however possible that the interactions between FUL and these proteins is not stable enough to be maintained during complex isolation, or that the interaction between FUL and these proteins is stabilised upon interaction with DNA. Although no direct interactions between FUL and G-box binding proteins or TCPs have been shown, these results suggest regulatory cooperativity between MADS domain proteins and the G-box binding proteins, and TCP proteins.

## MADS protein complexes have different DNA binding preferences

To test whether the observed differences in MADS dimerization partners could account for the observed differences in DNA binding, *in vitro* binding preferences were determined for the different FUL complexes. Previous studies on MADS domain binding preferences revealed the consensus DNA binding motif of MADS domain proteins, the CArG-box (Huang *et al.*, 1993; Shiraishi *et al.*, 1993; Huang *et al.*, 1995; Huang *et al.*, 1996; Tang and Perry, 2003). In line with these experiments, all tested FUL complexes bind to CArG-boxes. Despite that all complexes bind CArG-boxes, the SELEX-seq results show that the tested FUL complexes have different preferences for A- or T-nucleotides at specific positions within the CArG-box. For example, where FUL-AG, FUL-SOC1, and FUL-SEP3 all prefer a T after the double C, FUL-AG does not have a strong preference for a T or an A at this position.

Additional differences were found in the length of the binding sites. Most of the tested dimers prefer a binding site longer than the 10 bp, these longer motifs have a T-stretch before and an A-stretch after the CArG-box. The idea of longer CArG-boxes is in agreement with motif search for our ChIP-seq data and ChIP-seq data published for other MADS domain proteins (Deng *et al.*, 2011; Immink *et al.*, 2012; Pajoro *et al.*, 2014). Interestingly, not all tested motifs have the same preferences for T- and A-bases flanking the CArG-box. Where both FUL-FUL and FUL-AG prefer a T-stretch before and an A-stretch after the CArG-box, FUL-SOC1 motifs show only an A-stretch after the CArG-box. FUL-SEP3 does not prefer a longer CArG-box. Hence, our data show that

although all MADS domain dimers tested bind with high affinity to CArG-boxes, different dimers have different preferences both for nucleotide positions within the CArG-box, but also for motif length.

It has been hypothesized that these flanking nucleotides alter DNA shape and thereby facilitating the TF-DNA interaction (Jolma *et al.*, 2013; Muiño *et al.*, 2013). To visualize DNA binding specificities of the FUL complexes we used Position Weight Matrixes (PWM). PWMs are commonly used for the visualization of DNA binding motifs. However, they lack information about preferred DNA shape parameters and interdependencies between nucleotides within the motif. Hence, additional studies on the influence of DNA shape and interdependencies may further reveal differences in binding preferences of FUL complexes.

## Differences in DNA-binding affinities predict tissue specific DNA binding

Mapping the SELEX-seq to the genome showed that affinities of tissue-specific FUL-dimers are correlated with tissue-specific binding sites. This suggests that tissue-specificity of FUL functions is (at least partly) achieved by complex-specific differences in DNA binding site selection. However, despite the strong correlation found between DNA binding specificity and *in vivo* tissue specific DNA binding, the data also highlight the complexity of binding site selection by showing that not all high affinity SELEX-seq peaks correspond with ChIP-seq peaks. On the other hand, we also identified strongly enriched ChIP-seq peaks that are not linked to a SELEX-seq peak. These discrepancies between the *in vitro* SELEX-seq and *in vivo* ChIP-seq data are most likely due to *in planta* mechanisms not accounted for in our SELEX-seq experiments.

One possible explanation lays in the use of 'naked' DNA in SELEX-seq experiments. Nuclear DNA is packed into a higher-order chromatin structure which influences binding of TFs by blocking or enhancing DNA accessibility. Several studies showed the importance of nucleosome organization for TF binding in eukaryotes (Kaplan *et al.*, 2009; Nie *et al.*, 2014). While most TFs factors cannot evict nucleosomes, some TFs are able to influence nucleosomes, these TFs are called 'pioneer factors' (Zaret and Carroll, 2011). A study that combines analysis of MADS binding dynamics and chromatin accessibility shows that MADS domain TFs AP1 and SEP3 can select their binding sites independent of chromatin accessibility (Pajoro *et al.*, 2014). This suggests that MADS proteins can by itself or via other pioneer partners, alter DNA accessibility upon DNA binding. Also for FUL we found interaction partners that could potentially function to alter DNA accessibility, PKL, INO80,

and HDA15. Our results showed that interaction with these nucleosome remodelling factors is tissue-specific, suggesting that the ability of MADS domain proteins to influence chromatin structure might also be tissue specific. Moreover, besides nucleosomes also other types of high-affinity DNA binding proteins may interfere with DNA binding of MADS domain proteins, including other MADS domain proteins that may have a higher affinity for the specific site, resulting in binding site competition. The *in vitro* SELEX-seq data lack information about DNA accessibility and competing factors, and this can in part explain the discrepancies between the *in vitro* SELEX-seq and *in vivo* ChIP-seq results.

A reason for the apparent lack of high affinity FUL binding sites in genomic regions bound by FUL, could be related to the ability of MADS domain proteins to form quaternary complexes that are able to loop the DNA and recruit another dimer to low affinity binding sites in close proximity (Egea-Cortines *et al.*, 1999; Melzer *et al.*, 2009). Also, FUL homodimers, FUL-SOC1 and FUL-SEP3 are able to form quaternary complexes on the DNA. The SELEX-seq probes have a length of 40 bp, while a longer fragment is needed for DNA looping by MADS domain proteins (Melzer *et al.*, 2009), as a result the effect of looping and consequently cooperativity on DNA binding affinity is not accounted for in the SELEX-seq results. As looping may influence the recruitment to specific DNA sites, this could influence the correlations between *in vivo* binding and *in vitro* DNA binding sites. When looking more specifically at the *SMZ* promoter we find two CARG-boxes in close proximity. We find the SELEX-seq relative affinities of the two CArG-boxes to match dimer affinities in EMSA results, however the resulting affinities are not in line with the ChIP-seq peak heights, possibly due to differences in affinity between dimers and quarternary complexes.

Finally, we find a subset of ChIP-seq peaks that do not align with canonical binding sites of the different MADS protein complexes and therefore also do not show a SELEX-seq peak. This phenomenon is not unique for FUL, it has also been observed for other MADS domain proteins, e.g. SEP3 and SVP (Kaufmann *et al.*, 2009; Tao *et al.*, 2012), and for other TFs (O'Malley *et al.*, 2016). These binding sites are most-likely results of recruitment by other factors or reflect binding to multiple low affinity binding sites. Our immunoprecipitation results show that FUL is able to interact with other DNA binding proteins, which may facilitate indirect binding of FUL with the DNA.

In summary, our data pinpoints the important role of protein interactions on FUL DNA-binding specificity and thereby its function. However, the observed differences between *in vitro* SELEX-seq and *in vivo* ChIP-seq data also highlight the complexity of DNA binding site selection *in vivo*.

Future studies into other possible mechanisms determining DNA binding site recognition will result in a more complete picture of the underlying molecular mechanisms. These studies will also allow researchers to better predict regulatory specificity based on the *cis*-regulatory information.

# Material and methods

## Plant material

All plants were grown on rock-wool at 21°C under long day conditions (16 h light, 8 h dark).

## Tissue collection

For RNA-seq approximately 0.1 gram tissue of *ful-1* and *Landsberg erecta* wild type plants was collected and immediately frozen in liquid nitrogen. For the IM function of FUL, IM to stage 9 flowers were collected (floral stages according to Smyth *et al.* (1990)). For the function of FUL in the pistil stage 12 – 16 pistils were collected. Three biologically independent samples were collected per tissue. For ChIP-seq approximately 0.5 gram tissue was collected per sample, two independent biological samples were used per experiment. For the function of FUL in the IM uninduced inflorescence-like meristems of 35S:AP1:AP1-GR pFUL:FUL-GFP *ap1-1 cal-1* plants were collected. For the function of FUL in pistils stage 12 – 16 pistils of pFUL:FUL-GFP *ful-1* plants were collected.

## RNA-seq experiment

Total RNA was isolated from tissue using InviTrap Spin Plant RNA Mini Kit (Stratec). RNA was treated with DNase to remove remaining DNA using the TURBO DNA-*free*™ Kit (Ambion). The quality of the RNA was verified using the Agilent 2100 Bioanalyzer and a RNA 6000 pico kit (Agilent). RNA concentration were determined using a Nanodrop ND-100 spectrophotometer (Thermo Fisher Scientific). The TruSeqTM RNA Sample Preparation Kit (Illumina) was used for the cDNA library construction essentially following manufacturer's instructions. For synthesis of the first strand cDNA SuperScript III Reverse Transcriptase (Invitrogen) instead of Superscript II in similar amounts. The quality and size of the cDNA libraries was measured using an Agilent 2100 Bioanalyzer and a DNA 1000 kit (Agilent). DNA concentration were measured with a Qubit® 2.0 Fluorometer using the Qubit™ dsDNA HS assay kit (Invitrogen). Libraries of 12 samples were pooled and sequenced single-end with a read length of 50 bp in two sequence lanes, of the Illumina HiSeq2000.

## RNA-seq data analysis

RNA-seq analysis was adapted from Verk et al., 2013 (Van Verk *et al.*, 2013). Base calling was performed using CASAVA version 1.8. Sequenced reads reported by illumine CASAVA v1.8 pipeline as low quality reads were removed before further analysis. Quality of the reads was checked using FastQC (Andrews, 2010). The Illumina single-end reads were mapped to the *Arabidopsis* transcriptome TAIR10 (Huala *et al.*, 2001) using Bowtie version 0.12.7 (Langmead *et al.*, 2009). Parameters used are: -m 1, keep reads with maximum one reportable alignment and using the --best setting, guaranteeing that bowtie reports singleton alignments that are 'best' in terms of stratum and quality values at the mismatch positions. An overview of the mapping efficiency of the sequences data is reported in **Supplementary table S4**. DESeq was used to identify differentially expressed genes using default parameters (Anders and Huber, 2010). Genes are reported as differentially expressed when the adjusted P-value ≤ 0.05 and applying an absolute fold-change (FC) cutoff of 1.8. Reproducibility data (Pearson correlations and PCA plot) can be found in **Supplementary table S5 and S6** and **Supplementary figure S8**.

## ChIP-seq experiments

ChIP-seq experiments were performed following a previously published protocol (Van Mourik *et al.*, 2015). using an anti-FUL antibody. ChIP-seq experiment input samples were used as negative controls. For each tissue two biological replicates and one input sample were sequencing on the Illumina HiSeq with a read length of 50 bp. For number of reads per sample see **Supplementary table S7**.

## ChIP-seq data analysis

Base calling was performed using CASAVA version 1.8. Sequenced reads reported by Illumina CASAVA v1.8 pipeline as low quality reads were removed before further analysis. Quality of the reads was checked using FastQC (Andrews, 2010). Reads were mapped to the *Arabidopsis* genome (TAIR10)(Huala *et al.*, 2001) using soap2.20 (Li *et al.*, 2008) allowing 2 mismatches. ChIP-seq analysis was performed following a previously published protocol (Van Mourik *et al.*, 2015) using the CSAR software (Muiño *et al.*, 2011). Pearson correlation between individual ChIP-seq experiments showed a Pearson correlation of 0.73 and 0.94 for the IM ChIP-seq and the pistil ChIP-seq, respectively (**Supplemental figure S2**). Following the methodology of other studies (Kaufmann *et al.*, 2009; Deng *et al.*, 2011; Gregis *et al.*, 2013), the good correlation between biological replicates allowed for

pooling of mapped reads to obtain the final list of direct targets. CSAR background settings were set on 25 for both ChIP-seq data-sets. The Integrated Genome browser was used for visualization of ChIP-seq binding peaks (Nicol *et al.*, 2009).

## Quantitative comparison of ChIP-seq experiments

The differential binding analysis was performed using the DiffBind package (Stark and Brown, 2013). Specially, peaks identified from IM and pistil samples were first merged to get a common set of peaks. Then, the number of reads overlapping each peak were counted for each sample to get a matrix of counts based on FDR < 0.05. Lastly, differential analysis was performed using the dba.analyze function with default parameter "method=DBA_EDGER" (FC ≥ 2).

## Confocal scanning laser microscopy (CSLM)

FUL-GFP tagged protein localization was observed using CSLM on Leica SPE DM5500 upright microscope using an ACS APO 40x/1.15 oil lens and the LAS AF 1.8.2 software. GFP was excited with the 488-nm line of the Argon ion laser. Optical slices were median filtered and merged together into a three-dimensional projection using the LAS AF 1.8.2 software package.

## Protein immunoprecipitation and LC-MS/MS data analysis

Protein immunoprecipitation and LC-MS/MS data analysis was essentially performed as described in (Smaczniak *et al.*, 2012c).

## GO-term analysis

GO-term analysis was performed using the Cytoscape plugin BINGO (Maere *et al.*, 2005). Gene ontology for *Arabidopsis thaliana* and annotation file go-basic version 1.2 were downloaded from the Gene Ontology Consortium (Ashburner *et al.*, 2000; Consortium, 2015).

## EMSA

EMSA DNA probes were obtained from the promoter region of the SMZ gene and cloned into p-GEM-T vector (Promega). SEP3 promoter probes were obtained from Smaczniak *et al.* (2012b). Oligonucleotides were labelled with DY-682. Labelling was performed by PCR using vector-specific

DY-682-labelled primers followed by agarose gel extraction. Primers can be found in **Supplementary File S7**.

FUL and SOC1 coding sequences (CDS) were amplified from cDNA and cloned in pSPUTK expression vectors, primers used for cloning van be found in **Supplementary File S7**. pSPUTK::SEP2 and pSPUTK::AG vectors were obtained from Smaczniak *et al.* (2012b). Proteins were synthesized using TNT SP6 Quick Coupled Transcription/ Translation System (Promega) according to manufactures instructions. EMSA's were performed as described by Smaczniak *et al.* (2012b). Gel-shifts were visualized using a LiCor Odyssey imaging system at 700 nm.

## SELEX-seq

SELEX-seq dsDNA libraries were generated from ssDNA sequences by a single-cycle PCR amplification round with complementary primers essentially as described by Jolma *et al.* (2010). dsDNA sequences contain a 40 bp random sequence flanked by barcodes needed for later characterization after multiplex sequencing. In addition, the dsDNA sequences contained all features needed for direct sequencing on an HiSeq 2000 sequencer (Illumina).

First, purified FUL antibodies resuspended in 1X PDS were coupled to magnetic beads according to manufacturer's instructions (MyOne, Invitrogen). As for EMSA, proteins were synthesized using TNT SP6 Quick Coupled Transcription/ Translation System (Promega) according to manufactures instructions in a total volume of 20 µl. Binding reaction mix was prepared essentially as for EMSA experiments with a total volume of 120 µl (Smaczniak *et al.*, 2012b), the mix contained 20 µl *in vitro* synthesized protein and 50-100 ng dsDNA. The binding mix was incubated on ice for 1 hour. Followed by, immunoprecipitation using 0.5 mg FUL-antibody coupled to magnetic beads (MyOne, Invitrogen) in a thermomixer at 4ºC at 700 rpm. After immunoprecipitation, magnetic beads were washed 5 times with 150 µl binding buffer without salmon-sperm DNA. Bound DNA was elucidated in 50 µl 1X TE in a 90ºC thermomixer at full speed. Next, magnetic beads were immobilized and supernatant transferred to a new tube. To allow a second round of SELEX, DNA fragments were amplified with 8 to 16 cycles of PCR with SELEX round specific primers (Jolma *et al.*, 2010)(**Supplementary File S7**). The amplification efficiency was checked on agarose gel by comparison to a sample of known concentration. The total amplicon was used for a subsequent round of SELEX. Round for sequencing were cut out from gel after PCR amplification using

MinElute Gel Extraction Kit (Qiagen). Multiple SELEX samples were multiplexed in quimolar amounts, sequencing was performed in a HiSeq 2000 sequencer (Illumina).

## SELEX-seq data analysis

Base calling was performed using CASAVA version 1.8. Sequence reads that did not pass the filter quality of CASAVA 1.8 or mapped with no mismatches to the phix174 genome were eliminated. The remaining sequences in fasta format were extracted and grouped according to library specific barcodes, allowing no mismatches. Barcodes were removed leading to 40 bp sequence libraries used in the data analysis. The 40 bp sequences that were present in libraries in an unexpected high number (>1000) were eliminated, as well as 40 bp reads containing the sequence "TCGTATGCCG" which is part of the Illumina adapter sequence used for sequencing.

Data analysis was essentially performed as described before (Slattery *et al.*, 2011). We calculated which k-mer length should be used to obtain a good gain in information. For this, we computed the information gain (the Kullback–Leibler divergence of Round 5 relative to Round 0) for each k-mer length. For most of libraries the length 10 bp gave a good information gain, therefore, we based the further analysis on the 10-mer sequences.

Frequencies of 10-mer sequences in each round except Round 0 was calculated directly from the data using the function oligonucleotideFrequency from the Bioconductor R package: Biostrings. The 10-mer sequences that were present in libraries in an unexpected high number (>1,000) were eliminated at this step.

Sequences in Round 0 represent a set of randomly synthetized oligonucleotides and their complexity did not allow for the direct calculation of 10-mer frequencies. Therefore, the sequence frequency in Round 0 was estimated by the sixth-order Monte Carlo model, as proposed before (Slattery *et al.*, 2011). We chose the sixth-order Monte Carlo model, because when the model was trained using 75% of the sequencing data, it resulted in the highest prediction value as measured by the Pearson correlation coefficient between the predicted and observed frequencies in the other 25% of the sequencing data. Relative affinity for each possible 10-mer was calculated as the ratio between the frequencies of 10-mers in Round 5 to Round 0, and normalize to 1 by dividing for the highest affinity-predicted 10-mer.

### Combining ChIP-seq and SELEX-seq data to predict FUL-dimer binding sites

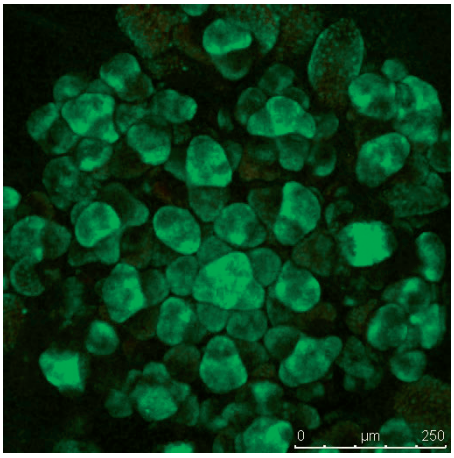
To *in silico* predict genomic regions bound by a given MADS domain dimer based on our SELEX-seq experiments, we obtained the affinity value for each k-mer of length 13 bp. For each studied MADS domain complex, we estimated the average affinity value to a particular 13-mer sequence and its reverse complementary sequence. We chose 13-mers instead of 10-mers for the technical reason: the mapping software 'soapv2' (Li *et al.*, 2009) is only able to map sequences with a minimum length of 13 bp. We created fasta files for each library containing 13-mer sequences in a number equal to their estimated relative affinity multiplied by 100 and rounded up to the closest integer (e.g. a sequence with the relative affinity of 0.98 was present in 98 copies in the fasta file). Next, we mapped these fasta files to the TAIR10 genome with 'soapv2' (Li *et al.*, 2009) allowing no mismatches. Later, the function mappedReads2Nhits from the peak caller CSAR (Muiño *et al.*, 2011) was used to identify enriched regions and their associated score, extending the 13 bp reads and using no control since the relative affinities were already corrected by the Round 0 enrichment. This resulted in a genomic SELEX-seq score proportional to the affinity of the sum of 13-mers located on the particular genomic region.

To relate ChIP-seq scores with the SELEX-seq scores obtained by CSAR, we analysed the FUL IM and pistil data using CSAR. Next, we link the ChIP-seq and SELEX-seq peak to the SEP3 ChIP-seq when their distance was shorter than 500 bp. Only the significant ChIP-seq peaks in at least one conditions were retained. Quantile normalization was used to normalize the SELEX-seq peaks scores, and independently to normalize the ChIP-seq peak scores aligned to the FUL ChIP-seq peaks.

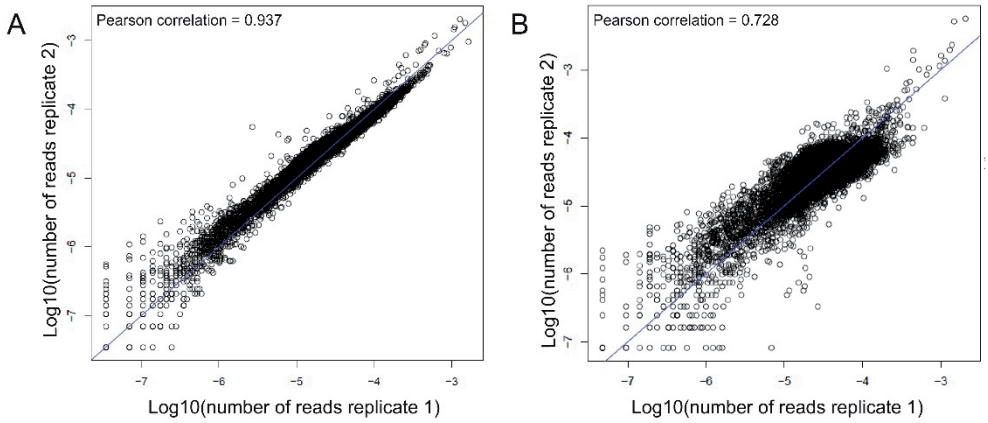
## Acknowledgements

The authors would like to thank Elio Schijlen and Bas te Lintel Hekkert for generating the HiSeq 2000 data. Marco Busscher for assistance in the laboratory. We thank Alice Pajoro for comments on the manuscript and advise on ChIP-seq experiments, and Arttu Jolma and Jussi Taipale for providing barcoded ssDNA. This work was supported by an NWO-VIDI grant to Kerstin Kaufman.

# Supplementary material



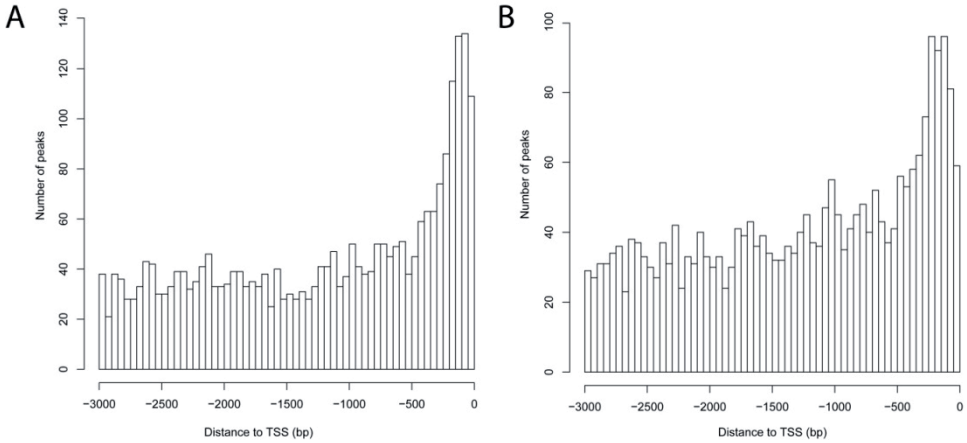

**Supplementary figure S1** | Confocal image of pFUL::FUL-GFP in IM tissue of un-induced *ap1 cal* 35S::AP1-GR lines.




**Supplementary figure S2** | Scatter plot based on the raw sequence data showing the average number of overlapping extended reads at each nucleotide position for the two biological replicates for the (**A**) IM and (**B**) pistil ChIP-seq experiments. The averages based on non-overlapping windows of 5,000 bp.

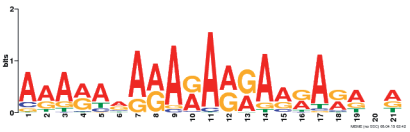
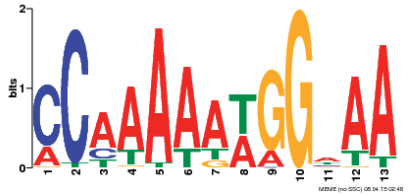
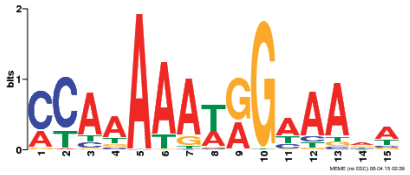
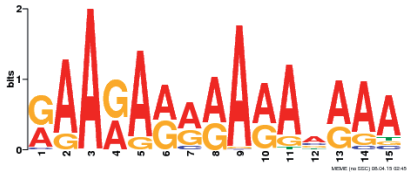
**Supplementary table S1** | ChIP-seq data: number of significant genes and peaks

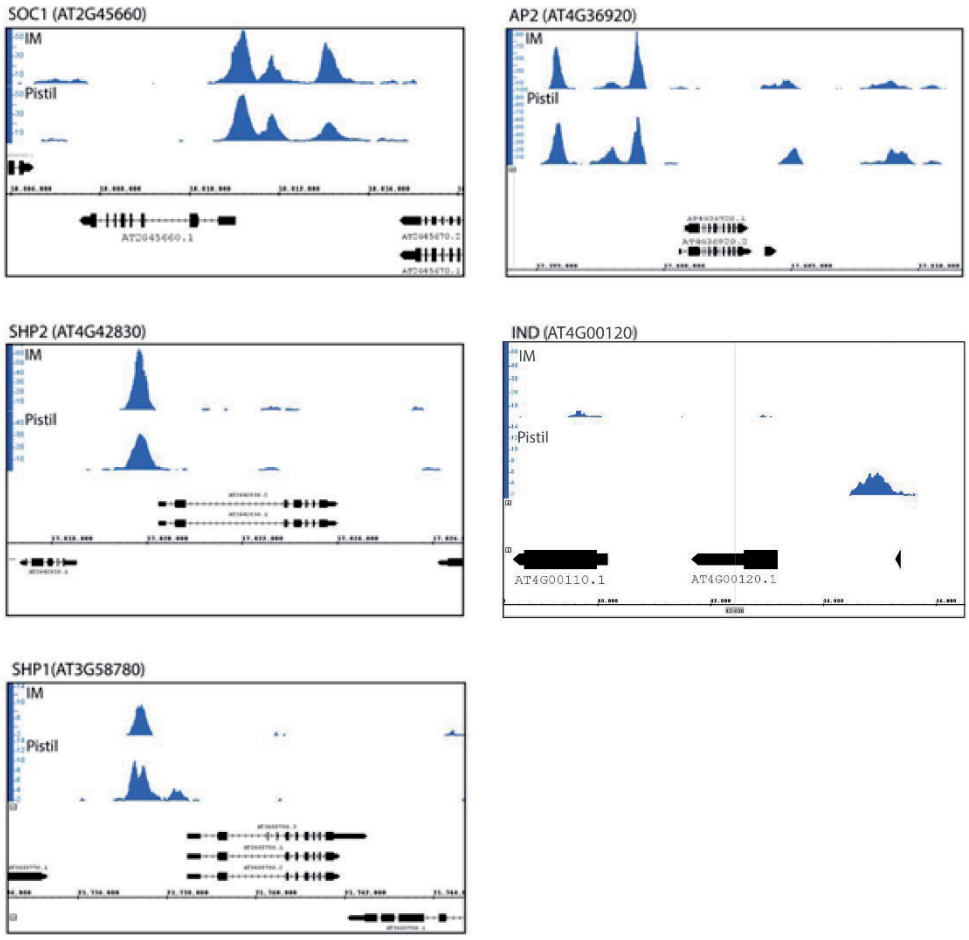
| | | FDR thresholds | | | |
|---|---|---|---|---|---|
| Tissue | | 0.05 | 0.01 | 0.005 | 0.001 |
| IM | Peak height threshold | 7.99 | 11.62 | 12.58 | 14.84 |
| | Total number of peaks | 2538 | 1194 | 1031 | 765 |
| | Total number of putative targets | 4055 | 1940 | 1692 | 1252 |
| Pistil | Peak height threshold | 5.19 | 6.40 | 7.00 | 7.65 |
| | Total number of peaks | 2459 | 1740 | 1534 | 1314 |
| | Total number of putative targets | 3598 | 2549 | 2265 | 1950 |




**Supplementary figure S3** | Number of ChIP-seq peaks relative to the transcriptional start site (TSS) of genes (zero position) for the (**A**) IM and the (**B**) pistil tissue (FDR 0.05).

**Supplementary table S2** | List of motifs identified by MEME-ChIP for the IM and pistil peaks (FDR 0.05) within 100bp around center of the peaks. Motifs with at least 50 occurrences are shown.

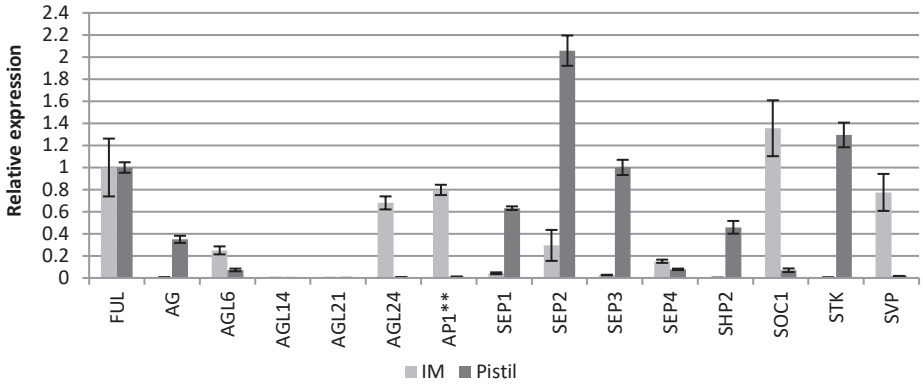
| Motif | E-value | Occurrences | Identified TFs (TOMTOM, only plants) |
|---|---|---|---|
| IM | | | |
|  | 1.2e-081 | 148 | MA0554.1 (SOC1)<br>MA0559.1 (PI)<br>MA0555.1 (SVP)<br>MA0045.1 (HMG-I/Y)<br>MA0120.1 (id1)<br>MA0558.1 (FLC)<br>MA0548.1 (AGL15)<br>MA0556.1 (AP3) |
|  | 3.5e-026 | 76 | MA0555.1 (SVP)<br>MA0548.1 (AGL15)<br>MA0556.1 (AP3)<br>MA0558.1 (FLC)<br>MA0559.1 (PI)<br>MA0563.1 (SEP3)<br>MA0082.1 (squamosa)<br>MA0584.1 (SEP1)<br>MA0001.2 (SEP4)<br>MA0005.2 (AG) |
| Pistil | | | |
|  | 1.5e-136 | 263 | MA0559.1 (PI)<br>MA0555.1 (SVP)<br>MA0548.1 (AGL15)<br>MA0556.1 (AP3)<br>MA0563.1 (SEP3)<br>MA0558.1 (FLC)<br>MA0082.1 (squamosa)<br>MA0554.1 (SOC1)<br>MA0584.1 (SEP1)<br>MA0578.1 (AtSPL8) |
|  | 7.8e-025 | 108 | MA0554.1 (SOC1)<br>MA0559.1 (PI)<br>MA0045.1 (HMG-I/Y)<br>MA0120.1 (id1)<br>MA0555.1 (SVP)<br>MA0558.1 (FLC)<br>MA0548.1 (AGL15)<br>MA0556.1 (AP3) |




**Supplementary figure S4** | ChIP-seq profiles of identified known FUL targets. The graphs represent for each locus the ChIP-seq peak profile for both the IM and the pistil. Chromosomal position (TAIR10) and gene models are shown at the bottom of each panel.

**Supplementary table S3** | Summary of interaction data and expression data of FUL interaction partners. * = no or limited overlap in expression domain with FUL (reviewed in Alvarez-Buylla *et al.* (2010) and Roeder and Yanofsky (2006))

| | *Reported interactions* | | | *Expression* | |
|---|---|---|---|---|---|
| Interaction partner | de Folter *et al.*, 2005 | Balanza *et al.*, 2014 | Smaczniak *et al.*, 2012 | IM | Pistils stage 12 - 16 |
| AP1 | | | + | - | - |
| SEP1 | + | | + | - | + |
| SEP2 | ND | | | + | + |
| SEP3 | + | | + | - | + |
| SEP4 | + | | | + | + |
| SOC1 | + | + | + | + | - |
| AGL6 | + | | | +* | +* |
| AGL14 | + | | | - | - |
| AGL21 | + | | | - | - |
| AGL24 | + | | | +* | - |
| AG | + | | | - | + |
| FUL | | | + | + | + |
| SVP | | + | | +* | - |
| STK | | | | - | + |
| SHP2 | | | | - | +* |

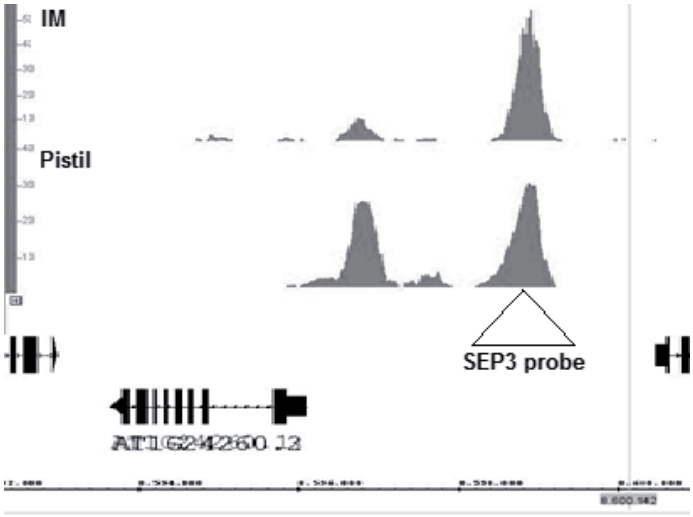



**Supplementary figure S5** | Gene expression of FUL interaction partners in IM and pistil stage 12-16 tissues. Based on microarray data from Pajoro *et al.* (2014) before induction and RNA-seq from this study (pistils stage 12-16). IM tissue is *ap1 cal* mutant x pAP1-AP1-GR non-induced. This figure does not provide information about the expression domains of these genes in the given tissues. ** AP1 is expressed as AP1-GR, the protein is however not moving to the nucleus as the tissue is not induced by dexamethasone. Upon AP1 induction *FUL* is downregulated.

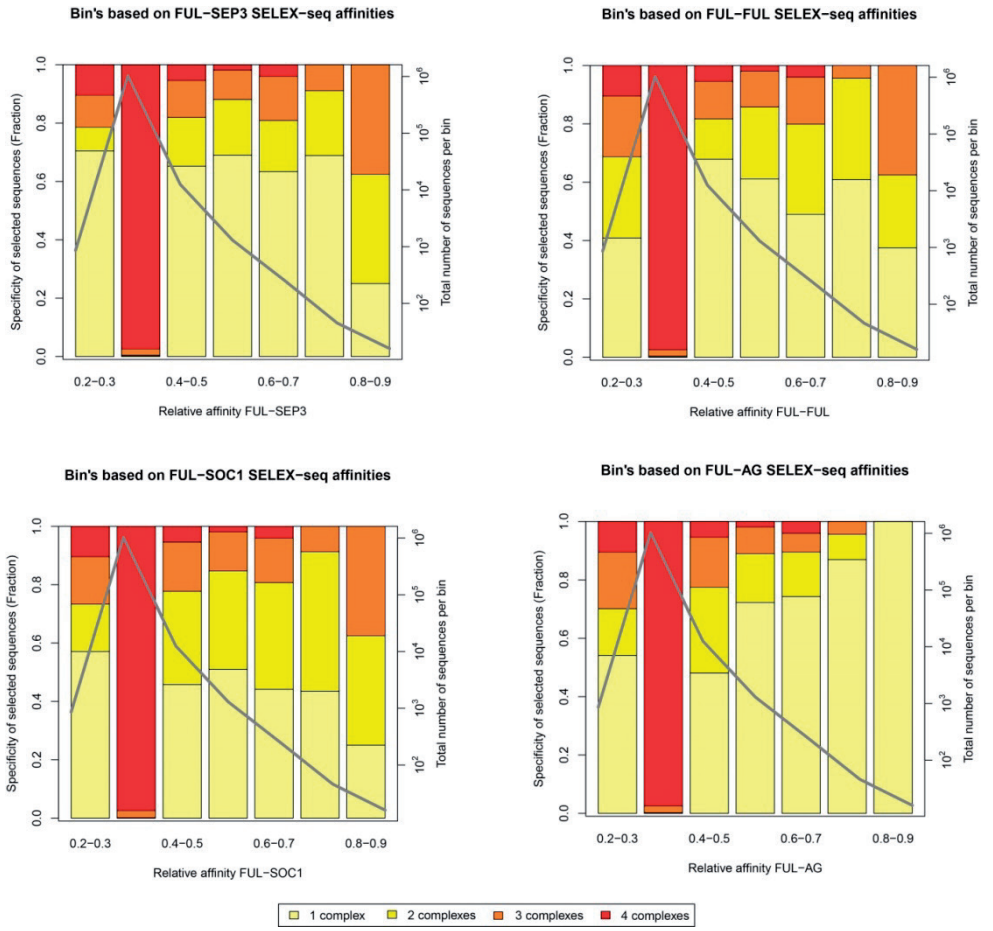
**Supplementary table S3:** TOMTOM output full sequence under peaks FDR0.05.

| Consensus | TOMTOM matches | List |
|---|---|---|
| IM | | |
| CACGYG | 16 | MA0561.1 (PIF4), MA0566.1 (MYC2), MA0569.1 (MYC4), MA0549.1 (BES1), MA0568.1 (MYC3), MA0560.1 (PIF3), MA0562.1 (PIF5), MA0550.1 (BZR1), MA0552.1 (PIL5), MA0580.1 (DYT1) |
| AAAWGGAA | 12 | MA0555.1 (SVP), MA0548.1 (AGL15), MA0559.1 (PI), MA0556.1 (AP3), MA0558.1 (FLC), MA0045.1 (HMG-I/Y), MA0082.1 (squamosa), MA0563.1 (SEP3), MA0120.1 (id1), MA0584.1 (SEP1) |
| CATAAAWG | 9 | MA0563.1 (SEP3), MA0548.1 (AGL15), MA0555.1 (SVP), MA0082.1 (squamosa), MA0556.1 (AP3), MA0584.1 (SEP1), MA0559.1 (PI), MA0008.1 (HAT5), MA0558.1 (FLC) |
| ANAGARA | 6 | MA0554.1 (SOC1), MA0559.1 (PI), MA0120.1 (id1), MA0558.1 (FLC), MA0548.1 (AGL15), MA0053.1 (MNB1A) |
| GCCRC | 6 | MA0570.1 (ABF1), MA0561.1 (PIF4), MA0560.1 (PIF3), MA0097.1 (bZIP911), MA0567.1 (ERF1), MA0551.1 (HY5) |
| TATATR | 2 | MA0584.1 (SEP1), MA0001.2 (SEP4) |
| GGYCCA | 2 | MA0587.1 (TCP16), MA0110.2 (ATHB5) |
| CTTCBTC | 1 | MA0127.1 (PEND) |
| CATAAAWG | 9 | MA0563.1 (SEP3), MA0548.1 (AGL15), MA0555.1 (SVP), MA0082.1 (squamosa), MA0556.1 (AP3), MA0584.1 (SEP1), MA0559.1 (PI), MA0008.1 (HAT5), MA0558.1 (FLC) |
| AGYTTCC | 0 | |
| AACCCTAR | 0 | |
| Pistil | | |
| CACRYG | 14 | MA0562.1 (PIF5), MA0561.1 (PIF4), MA0566.1 (MYC2), MA0568.1 (MYC3), MA0560.1 (PIF3), MA0569.1 (MYC4), MA0128.1 (EmBP-1), MA0549.1 (BES1), MA0580.1 (DYT1), MA0550.1 (BZR1) |
| ATAAAWGG | 12 | MA0563.1 (SEP3), MA0082.1 (squamosa), MA0548.1 (AGL15), MA0555.1 (SVP), MA0001.2 (SEP4), MA0556.1 (AP3), MA0584.1 (SEP1), MA0110.2 (ATHB5), MA0008.1 (HAT5), MA0558.1 (FLC) |
| ATTTRTGG | 8 | MA0584.1 (SEP1), MA0563.1 (SEP3), MA0082.1 (squamosa), MA0548.1 (AGL15), MA0001.2 (SEP4), MA0556.1 (AP3), MA0558.1 (FLC), MA0555.1 (SVP) |
| TTTKGB | 7 | MA0559.1 (PI), MA0563.1 (SEP3), MA0548.1 (AGL15), MA0556.1 (AP3), |

| | | |
|---|---|---|
| | | MA0558.1 (FLC), MA0555.1 (SVP), MA0082.1 (squamosa) |
| TRAGGAAA | 6 | MA0559.1 (PI), MA0120.1 (id1), MA0548.1 (AGL15), MA0555.1 (SVP), MA0556.1 (AP3), MA0082.1 (squamosa) |
| ATATRKA | 3 | MA0584.1 (SEP1), MA0082.1 (squamosa), MA0001.2 (SEP4) |
| AGAARM | 2 | MA0559.1 (PI), MA0554.1 (SOC1) |
| GGHCCA | 1 | MA0587.1 (TCP16) |
| AGAGASD | 0 | |
| AGTASTA | 0 | |
| CCCSACA | 0 | |




**Supplementary figure S6** | ChIP-seq profiles of the *SEP3* locus, triangle points towards the *SEP3* EMSA probe used in **figure 5A**. The graphs represent the ChIP-seq peak profiles for IM and pistil. Chromosomal position and gene models are shown at the bottom of the panel (TAIR10).




**Supplementary figure S7** | Correlation between DNA sequence affinity and specificity. The proportion of 10mer sequences bound by various FUL protein complexes versus the relative affinity of these 10mers. The number of 10mers in each bin is plotted in grey. Sequences specific for a single FUL complex (light yellow) are not more prevalent in high affinity bins than in low-affinity bins.

**Supplementary table S4** | Bowtie mapping results RNA-seq data.

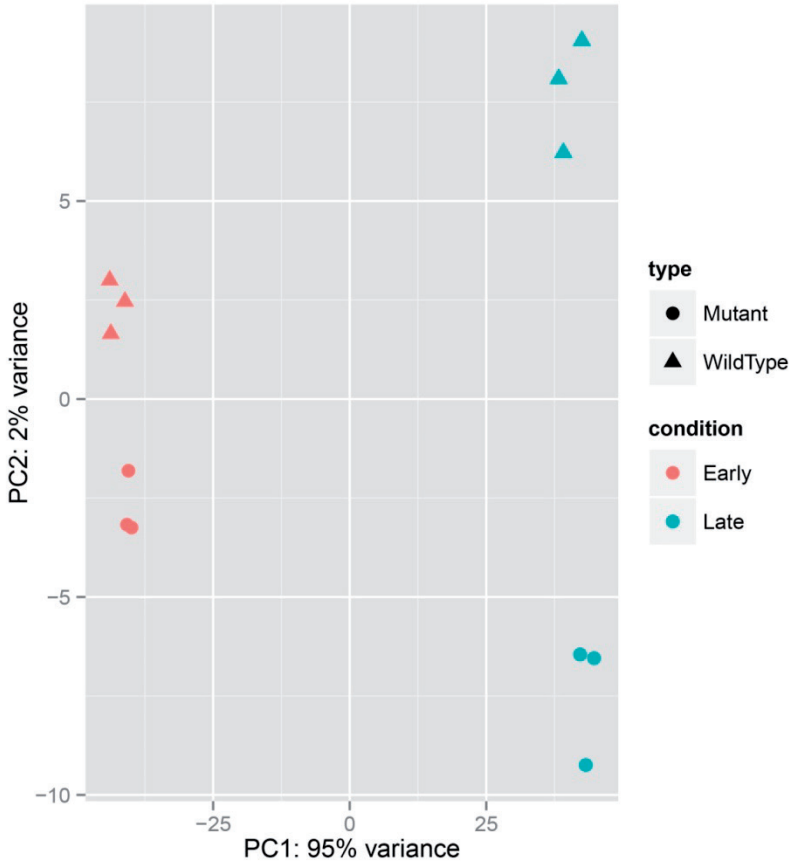
| Sample name | Sample | | Total number of reads | % of uniquely mapped reads | % of non-aligned reads | % of reads with >1 alignment |
|---|---|---|---|---|---|---|
| A | IM | Ler WT | 38,227,466 | 86.9 | 6.8 | 6.3 |
| B | | | 21,829,279 | 85.9 | 7.7 | 6.4 |
| C | | | 24,789,845 | 86.7 | 7.0 | 6.4 |
| D | | *ful*-1 | 33,037,994 | 86.6 | 7.0 | 6.4 |
| E | | | 36,173,776 | 85.5 | 8.3 | 6.2 |
| F | | | 26,521,861 | 84.3 | 9.7 | 6.1 |
| G | Pistil | Ler WT | 27,858,491 | 82.9 | 10.7 | 6.5 |
| H | | | 30,146,897 | 87.1 | 6.3 | 6.5 |
| I | | | 32,266,440 | 86.5 | 7.4 | 6.1 |
| J | | *ful*-1 | 36,256,815 | 86.9 | 7.2 | 5.9 |
| K | | | 25,940,758 | 85.4 | 8.7 | 5.9 |
| L | | | 32,780,774 | 87.6 | 6.6 | 5.8 |

**Supplementary table S5** | Pearson correlation of number of hits per gene for the IM plus stage 1 – 10 flower tissue.

| | | | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|
| | | | IM | | | | | |
| | | | Ler WT | | | *ful*-1 | | |
| A | IM | Ler WT | 1 | | | | | |
| B | | | 0.929638 | 1 | | | | |
| C | | | 0.94604 | 0.977148 | 1 | | | |
| D | | *ful*-1 | 0.914935 | 0.96066 | 0.985818 | 1 | | |
| E | | | 0.94517 | 0.920074 | 0.973093 | 0.974046 | 1 | |
| F | | | 0.986337 | 0.932519 | 0.949434 | 0.925024 | 0.949474 | 1 |

**Supplementary table S6** | Pearson correlation of number of hits per gene for the pistil tissue.

| | | | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|
| | | | Pistil | | | | | |
| | | | Ler WT | | | *ful*-1 | | |
| G | Pistil | Ler WT | 1 | | | | | |
| H | | | 0.930478 | 1 | | | | |
| I | | | 0.927856 | 0.956015 | 1 | | | |
| J | | *ful*-1 | 0.850265 | 0.904893 | 0.926968 | 1 | | |
| K | | | 0.884109 | 0.935182 | 0.921909 | 0.979297 | 1 | |
| L | | | 0.895157 | 0.88878 | 0.897817 | 0.948626 | 0.971783 | 1 |




**Supplementary figure S8** | PCA plot of RNA-seq (data analysed by DE-seq).

**Supplementary table S7** | SOAPv2 mapping results of the ChIP-seq data

| Sample | Total number of reads | # reads mapped | % of uniquely mapped reads | % filtered out |
|---|---|---|---|---|
| IM_rep1 | 39,831,046 | 30,277,332 | 76.0% | 24.0 % |
| IM_rep2 | 38,269,308 | 27,143,142 | 70.9% | 29.1% |
| Siliques_rep1 | 29,144,539 | 21,759,771 | 74.7% | 25.3% |
| Siliques_rep2 | 39,037,290 | 12,999,859 | 33.3% | 66.7% |
| Siliques_input | 124,995,421 | 84,161,263 | 67.3% | 32.7% |

# Description of additional files

**Supplementary file S1** | Differentially expressed genes (DEGs) as identified in RNA-seq comparing wild type and *ful*-1 mutant tissue.

**Supplementary file S2** | Complete GO-term analysis on DEGs identified by RNA-seq.

**Supplementary file S3** | ChIP-seq peak calling for FUL in IM and pistil tissue.

**Supplementary file S4** | Complete GO-term enrichment of significantly bound genes in IM and pistil ChIP-seq experiments.

**Supplementary file S5** | Complete IP LC-MS/MS analysis for FUL comparing IM and pistil tissue.

**Supplementary file S6** | SELEX-seq result for MADS domain dimers FUL-FUL, FUL-AG, FUL-SOC1, and FUL-SPE3.

**Supplementary file S7** | List of primers used.

# Chapter 3

## Characterization of *in vivo* DNA-binding events of plant transcription factors by ChIP-seq: experimental protocol and computational analysis

Hilda van Mourik
Jose M Muiño
Alice Pajoro
Gerco C Angenent
Kerstin Kaufmann

Methods Mol Biol. 2015;1284:93-121

## Abstract

Chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) is a powerful technique for the genome-wide identification of *in vivo* binding sites of DNA-binding proteins. The technique had been used to study many DNA-binding proteins in a broad variety of species. The basis of the ChIP-seq technique is the ability to covalently cross-link DNA and proteins that are located in very close proximity. This allows the use of an antibody against the (tagged) protein of interest to specifically enrich DNA fragments bound by this protein. ChIP-seq can be performed using antibodies against the native protein or against tagged proteins. Using a specific antibody against a tag to immunoprecipitate tagged proteins eliminates the need for a specific antibody against the native protein and allows more experimental flexibility. In this chapter we present a complete workflow for experimental procedure and bioinformatic analysis that allows wet-lab biologists to perform and analyse ChIP-seq experiments.

# Introduction

Chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) is a powerful technique for the genome-wide identification of *in vivo* binding sites of DNA-binding proteins. The method can be used for most DNA-binding proteins, including transcription factors, transcriptional co-activators, chromatin regulators, and (modified) histones (Furey, 2012). The information obtained from ChIP-seq experiments extends our knowledge about transcriptional regulation, chromatin structure and dynamics, and other processes that are fundamental to gene regulation.

The ChIP-seq technique is based on the ability to covalently cross-link DNA and proteins that are located in close proximity using a chemical agent. In ChIP experiments, cross-linked DNA–protein complexes containing the protein of interest are immunoprecipitated using a specific antibody. High quality of the antibody, with strong antigen-binding and low levels of cross-reaction, is an important prerequisite for specific ChIP enrichment and therefore for a successful outcome of ChIP-seq experiments.

An antibody against the native protein is usually preferred. However, a specific antibody of sufficient quality is not always available nor can it be generated for all proteins. To overcome this problem, ChIP can be performed using proteins fused to a so called tag. In this approach, the protein of interest is transgenically expressed fused to a tag (e.g. Green Fluorescent Protein (GFP), myc, flag), allowing the use of a specific antibody against the tag. Besides bypassing the need for an antibody against the native protein, tagged proteins can also increase experimental flexibility. An example is the expression of the tagged protein from a tissue specific promoter within wild type plants, instead of the endogenous promoter; this will enable the identification of tissue-specific DNA binding sites.

This chapter provides a hands-on protocol for ChIP-seq experiments using tagged proteins. The protocol is based on a previously published protocol (Kaufmann *et al.*, 2010b) with some modifications (Kallesen and Rosen, 2001). We choose to describe a protocol for GFP-tagged proteins since commercial high-quality antibodies are available against GFP. Moreover, the use of a fluorescent tag allows the visualization of protein expression and localization *in vivo* by confocal microscopy. The protocol can, with only minor modification, be used for proteins fused to other

tags. We present a complete workflow for experimental procedure and bioinformatic analysis to allow wet-lab biologists to perform and analyse their ChIP-seq experiments.

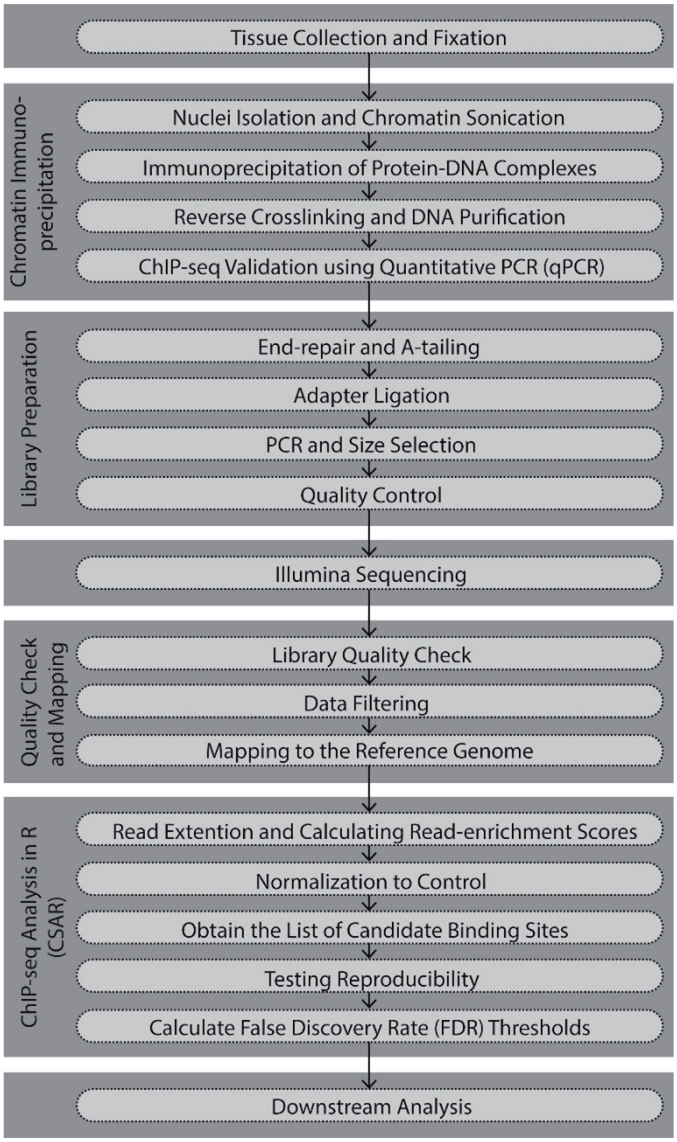



**Figure 1 |** Overview of the different steps of the complete ChIP-seq procedure, from experimental protocol to downstream bioinformatics analysis.

# Materials

Before starting the protocol, cool down all the buffers on ice.

## Tissue collection and fixation

1. Liquid nitrogen
2. MC buffer (fresh): 10mM sodium phosphate buffer pH 7, 50mM NaCl, 0.1M sucrose
3. Formaldehyde (37 %), preferably less than a year old
4. 1.25M glycine (store at 4 °C)
5. Desiccator linked to vacuum pump
6. 50-ml centrifuge tubes (Greiner)

## Nuclei isolation and chromatin sonication

1. Mortar and pestle
2. Liquid nitrogen
3. Nylon mesh (55 µm pore size)
4. Nitrile gloves
5. Glass funnel, recommended size 75 mm funnel top
6. M1 buffer (fresh): 10mM sodium phosphate buffer pH 7, 0.1M NaCl, 1M 2-methyl 2.4-pentanediol, 10mM β-mercaptoethanol, 1×complete protease inhibitor cocktail (Roche)
7. M2 buffer (fresh): 10mM sodium phosphate buffer pH 7, 0.1M NaCl, 1M 2-methyl 2.4-pentanediol, 10mM β-mercaptoethanol, 10mM $MgCl_2$, 0.5% Triton X-100, 1 × complete protease inhibitor cocktail (Roche)
8. M3 buffer (fresh): 10mM sodium phosphate buffer pH 7, 0.1M NaCl, 10mM β-mercaptoethanol, 1×complete protease inhibitor cocktail (Roche)
9. Low-adhesion microcentrifuge tubes, 2 ml (e.g. Protein LoBind Tubes, Eppendorf)
10. Probe sonicator, e.g. MSE Soniprep 150 (see Note 1)
11. Sonic buffer (store at −20°C): 10mM sodium phosphate pH 7, 0.1M NaCl, 0.5% Sarkosyl, 10mM EDTA, 1 × complete protease inhibitor cocktail (Roche) (add freshly)
12. IP buffer (store at −20°C): 50mM HEPES pH 7.5, 150mM NaCl, 5mM $MgCl_2$, 10µM $ZnSO_4$, 1% (vol/vol) Triton X, 0.05% (wt/vol) SDS
13. Low-binding tips (VWR)

## Immunoprecipitation of protein–DNA complexes

1. Nitrile gloves
2. Rotation Incubator for tubes (e.g. Stuart® rotator with micro tube holder (Stuart))
3. µMACS GFP Isolation Kit (Miltenyi Biotec): µMACS anti-GFP microbeads (see Note 2)
4. µ-Columns (Miltenyi Biotec)
5. µMACS Separator (Miltenyi Biotec)
6. IP buffer (store at –20°C): 50mM Hepes pH 7.5, 150mM NaCl, 5mM $MgCl_2$, 10µM $ZnSO_4$, 1% (vol/vol) Triton X-100, 0.05% (wt/vol) SDS
7. High-salt buffer: 500mM NaCl, 1% (vol/vol) Triton X-100, 0.1% (wt/vol) SDS, 2mM EDTA, 20mM Tris–HCl pH 8
8. LiCl buffer: 10mM Tris–HCl pH 8, 1mM EDTA, 1% (vol/vol) NP-40, % (vol/vol) sodium deoxycholate, 0.25M LiCl
9. 1 × TE buffer: 10mM Tris–HCl pH 8, 1mM EDTA
10. Elution buffer: 1% (wt/vol) SDS, 50mM Tris–HCl pH 8, 10mM EDTA, 50mM DTT (add freshly)
11. Low-adhesion microcentrifuge tubes, 1.5 and 2 ml (e.g. Protein LoBind Tubes (Eppendorf))
12. Low-binding tips (VWR)

## Reverse cross-linking and DNA purification

1. Nitrile gloves
2. 1 × TE buffer: 10mM Tris–HCl pH 8, 1mM EDTA
3. Proteinase K (20mg/ml) (Roche)
4. 100% ethanol
5. 3M Sodium acetate, pH 5.4
6. Glycogen (20µg/µl)
7. Ultrapure water
8. Qiaquick PCR purification kit (Qiagen)
9. Agarose MP (Invitrogen)
10. 1 × TE buffer: 1mM EDTA pH 8, 10mM Tris–HCl pH 8
11. Ethidium bromide

12. Smart Ladder (Eurogentec)
13. Low-adhesion microcentrifuge tubes, 1.5 ml (e.g. Protein LoBind Tubes (Eppendorf))
14. Low-binding tips (VWR)

## ChIP-seq validation using quantitative PCR (qPCR)

1. Nitrile gloves
2. Forward and reverse qPCR primers
3. iQ SYBR Green Supermix (Bio-Rad)
4. Realtime PCR machine (e.g. Bio-Rad iQ5)
5. 96 × 0.2 ml Plate (BIOplastics)
6. Opti-Seal Optimal Disposal Adhesive (BIOplastics)
7. Low-binding tips (VWR)
8. Low-adhesion microcentrifuge tubes, 1.5 ml (e.g. Protein LoBind Tubes (Eppendorf))

## DNA library preparation for Illumina sequencing: end-repair and A-tailing

1. Nitrile gloves
2. Low-adhesion microcentrifuge tubes, 1.5 and 0.5 ml (e.g. Protein LoBind Tubes (Eppendorf))
3. End-It DNA end repair kit (Epicentre)
4. Klenow enzyme (NEB)
5. Klenow buffer, e.g. NEB buffer 2 (New England BioLabs)
6. QIAquick PCR Purification Kit (Qiagen)
7. 1mM dATP
8. Klenow fragment (3′ to 5′ exo-minus) (NEB)
9. MinElute PCR purification kit (Qiagen)
10. Low-binding tips (VWR)

## Adapter ligation

1. Nitrile gloves
2. Adaptors from Illumina ChIP-seq kit (Illumina)
3. 10 × T4 DNA ligase buffer (see **Note 3**)

4. T4 DNA ligase
5. QIAquick PCR purification Kit (Qiagen)
6. Low-adhesion microcentrifuge tubes, 1.5ml (e.g. Protein LoBind Tubes (Eppendorf))
7. Low-binding tips (VWR)

## PCR and size selection

1. Nitrile gloves
2. Illumina adaptors and PCR primers (Illumina, cat. no. IP-102-1001 or PE-400-1001) (see Note 4)
3. Phusion high-fidelity DNA polymerase enzyme (ThermoScientific)
4. 4.5 × Phusion buffer
5. 2.5mM dNTP
6. QIAquick MinElute PCR Purification Kit (Qiagen)
7. Glycerol
8. QIAgen MinElute Gel purification Kit (Qiagen)
9. Loading Dye from the QIAgen MinElute Gel purification Kit (Qiagen)
10. Agrose gel running and visualization (UV) devices
11. Low-adhesion microcentrifuge tubes, 1.5 ml (e.g. Protein LoBind Tubes (Eppendorf))
12. Low-binding tips (VWR)

## Quality control

1. Nitrile gloves
2. Forward and reverse qPCR primers for negative (non-bound) and positive (bound) genomic regions
3. iQ SYBR Green Supermix (Bio-Rad)
4. Real-time PCR machine (e.g. Bio-Rad iQ5)
5. 96 × 0.2ml Plate (BIOplastics)
6. Opti-Seal Optimal Disposal Adhesive (BIOplastics)
7. Low-adhesion microcentrifuge tubes, 1.5 ml (e.g. Protein LoBind Tubes (Eppendorf))
8. Low-binding tips (VWR)

## Illumina sequencing

1. Nitrile gloves
2. Qubit® dsDNA HS Assay Kit (Invitrogen)
3. Qubit® 2.0 fluorometer (Invitrogen)
4. Agilent BioAnalyzer system (Agilent)
5. Agilent BioAnalyzer DNA 1000 Kit (Agilent)
6. Elution buffer from the QIAgen MinElute Gel purification Kit (Qiagen)
7. Illumina sequencer (e.g. HiSeq (Illumina))
8. Illumina sequencer reagent kit (Illumina)
9. Low-binding tips (VWR)

## Equipment and software for bioinformatic analysis

For the computational part of this protocol, a computer workstation either running a Unix-based operating system or with access to a Unix-based server is needed. This protocol provides commands runnable in the Unix shell. Part of the protocol uses the R statistical computing environment. Commands to be run in the UNIX shell are prefixed with "`$`", commands meant to run from a R script are prefixed with "`>`".

Required software are: R version 3.0.2 or higher installed (http://www.r-project.org/), JAVA runtime environment (https://www.java.com/), Perl (http://www.perl.org/), SOAPv2 (Li *et al.*, 2009) (http://soap.genomics.org.cn/), FASTQC (http://www.Bioinformatics.babraham.ac.uk/projects/fastqc/), and the R packages: CSAR (Muiño *et al.*, 2011) (http://www.Bioconductor.org/packages/release/bioc/html/CSAR.html) and Biostrings (Pages *et al.*, 2017) (http://www.bioconductor.org/packages/release/bioc/html/Biostrings.html).

To install the R packages CSAR and Biostrings, start an R session and do the following:

```
$ R
> source("http://bioconductor.org/biocLite.R")
> biocLite("CSAR")
> biocLite("Biostrings")
```

Before installing SOAPv2, create a directory to store executable programs (if it does not yet exist):

```
$ cd <yourhomedirectory>
$ mkdir software
```

Download SOAPv2 from http://soap.genomics.org.cn/soapaligner.html to the "software" directory. Unpack and install SOAPv2:

```
$ cd <yourhomedirectory>/software
$ tar zxvf SOAPaligner.tar.gz
```

This directory contains two executable files: *2bwt-builder* and *soap* .

For filtering CASAVA version 1.8 FASTQ-files, a Perl script called "covert_export_to_fasta.pl" is used. To create this Perl script, do the following:

Create a directory to store scripts (if it does not yet exist):

```
$ cd <yourhomedirectory>
$ mkdir code
```

Generate the Perl file script:

```
$ cd <yourhomedirectory>/code
$ cat > convert_export_to_fasta.pl
#!/usr/bin/perl –w
use strict;
my $infile = $ARGV[0];
my $seq="";
open (IN, $infile);
while (<IN>) {
my @fields = split (/:/);
if(@fields>7){
if($fields[7] eq "N"){$seq=<IN>;
if(!($seq=~"N")){print ">\n$seq";}};
}
}
```

type Crtl-D to save and exit.

Besides hardware and software, also the assembled genome sequence and gene annotation files are needed. For most plant species, these files can be download from several databases, for example Phytozome (Goodstein *et al.*, 2012)(http://phytozome.jgi.doe.gov/pz/portal.html) or Ensembl plants (http://plants.ensembl.org). As an example, instructions on how to download the *Arabidopsis* reference genome from Ensembl plants are shown below:

Create a directory to store the genome-files (if it does not yet exist):

```
$ cd <yourhomedirectory>
$ mkdir genomes
```

Both the reference genome and the genomic features files can be downloaded via the Web Browser or using the shell window:

– Via the Web Browser:

- Go to the website of Ensembl Plant (http://plants.ensembl.org/)
- Go to "Downloads" at the top of the page
- Select "Download data via FTP" under "Download databases & software"
- Search for the organism of interest (e.g. *Arabidopsis thaliana* )

For the reference genome:

- Click on "FASTA (DNA)" of the species of interest
- Download the file ending with ".dna.genome.fa.gz", this contains the unmasked genome sequence
- Move the downloaded file to the "<yourhomedirectory>/genomes" directory

For the gene annotation:

- Click on the "GFF3" of the species of interest
- Download the file contains the gene annotation of the whole genome (i.e., "*Arabidopsis*_thaliana.TAIR10.25.gff3.gz")
- Move the downloaded file to the "<yourhomedirectory>/genomes" directory

– Via the shell window (see **Note 5**):

```
$ cd <yourhomedirectory>/genomes
$ wget -nd ftp://ftp.ensemblgenomes.org/pub/plants/release-
25/fasta/Arabidopsis_thaliana/dna/Arabidopsis_thaliana.TAIR10.25.
dna.genome.fa.gz
$ wget -nd ftp://ftp.ensemblgenomes.org/pub/release-
25/plants/gff3/Arabidopsis_thaliana/Arabidosis_thaliana.TAIR10.25
.gff3.gz
```

Both files are provided as compressed documents, and therefore, unzipping is needed:

```
$ cd <yourhomedirectory>/genomes
$ gunzip Arabidopsis_thaliana.TAIR10.25.dna.genome.fa.gz
$ gunzip Arabidopsis_thaliana.TAIR10.25.gff3.gz
```

The GFF3 file contains all sequence features known by the Ensembl group, for example protein coding genes, ncRNA, exons, and UTRs. It is possible to use the complete GFF3 files. However, for most ChIP-seq experiments we are primarily interested in the location of the binding peaks in relation to genes. Therefore, the GFF3 file can be adjusted to contain only genomic features for genes. For this, the software environment R is used. First, define your working directory and open the .gff3 file:

```
$ R
> setwd("<yourhomedirectory>/genomes")
> gff<-read.table("Arabidopsis_thaliana.TAIR10.25.gff3", header =
FALSE, sep="\t", quote="")
```

Next, select only those rows containing the word "gene" in the third column and save the file (see Notes 6 and 7):

```
> gff<-gff[gff$V3=="gene",]
> write.table(gff,file="TAIR10_GFF3_onlygenes.gff")
> q()
```

# Methods

## ChIP-seq protocol

The ChIP protocol starts with tissue collection and covalently linking protein-DNA interactions. Cross-linking is followed by nuclei isolation and shearing of the chromatin. Both nuclei extraction and shearing generates to insoluble material in the sample. Insoluble material is a major source of 'background' in the immunoprecipitate and may block the column when purifying protein-DNA complexes. Therefore, to eliminate this insoluble material several centrifugation steps are integrated in the protocol. Next, specific protein-DNA complexes are isolated by immunoprecipitation using a specific antibody against the tagged protein of interest. After immunoprecipitation, the DNA is reverse cross-linked and purified. As the quality of the generated library reflects the level of enrichment of specific genomic regions bound by the protein of interest, the approximate amount of DNA as well as the quality of the purified ChIP DNA is tested by qPCR. When the quality of the DNA is sufficient, the sample is prepared for next generation sequencing. Our protocol describes library preparation for an Illumina sequencing platform. For an overview of the different steps in the ChIP procedure, see **Figure 1**.

When a ChIP experiment is performed using a tagged protein of interest, it is crucial that the protein remains functional when fused to the tag. The functionality of the tagged-protein construct is influenced by the used tag, the sequence between the tag and the protein (linker), the position of the tag (N- or C- terminal of the protein), and the choice of promoter sequence. The functionality of the tagged protein can be tested by introducing it in the mutant background: a functional tagged protein is expected to rescue the mutant phenotype (De Folter *et al.*, 2007). Also, make sure the tag is not cleaved off from the protein. This can be tested using western blot: a band corresponding to the size of free tag should not be detectable.

To reduce signal-to-noise ratio, it is essential to obtain high amounts of specifically enriched DNA while avoiding non-specifically precipitated DNA. Several things influence the signal-to-noise ratio, the most import are: the choice of tissue, the expression level of the protein, the choice of negative control, and, as mentioned in the introduction, the specificity of the antibody (Landt *et al.*, 2012). It also matters how stable the protein binds to the DNA and whether the protein-DNA binding is

direct or indirect. In the latter case, additional cross-linking agents could be beneficial, since formaldehyde mainly crosslinks direct protein-DNA and protein-protein interactions (Nowak *et al.*, 2005; Zeng *et al.*, 2006).

Many proteins are active in more than one type of tissue and/or developmental stage. The tissue of interest should be chosen to match the research question. It is preferable to use homogenous tissue rather than a mix of different tissues or stages. As this protocol is aimed for fusion proteins, it is recommended to express the fusion protein in its mutant line. Related to this, the choice of promoter is also important. For most applications it is best to use the native promoter, as it will drive the fusion protein expression at its 'natural' level, location, and stages. However, in some cases it can be beneficial to use a different type of promoter. Examples of other promoters are constitutive promoters (e.g. CaMV 35S or UBQ10 promoter), inducible promoters, or cell-type-specific promoters. In case a non-native promoter is used, it is crucial to confirm the binding sites detected in the ChIP by independent methods, e.g. ChIP qPCR using the natively expressed proteins, other DNA-binding essays and/or *in vivo* reporter gene studies.

Another consideration before starting a ChIP-seq experiment is the choice of control sample. Control experiments are critical for adjustment of bias of non-uniform DNA fragmentation during sonication caused by differences in chromatin accessibility (Auerbach *et al.*, 2009). Preferably, a control should also detect non-homogeneous 'background' precipitation of genomic DNA in the immunoprecipitation step. In the past, different types of controls have been used: input DNA, which is isolated from the sonicated chromatin prior to immunoprecipitation; IgG control, with a ChIP performed using a non-specific IgG fraction; and mock IP, with a ChIP done on a tissue that does not express the tagged protein of interest (Kaufmann *et al.*, 2010b; Landt *et al.*, 2012). For tagged proteins, the best control is the same background line expressing only the tag (without the protein of interest) from the same promoter as used for the tagged protein. If using an inducible version of the protein of interest, material from un-induced plants can be used as control. Negative control ChIP experiments often result in a (too) low DNA yield; this can be overcome by pooling several samples. It is important to sequence the sample and the control at a comparable depth to prevent bias caused by peak calling.

As mentioned in the introduction, a good quality of the specific antibody will increase the enrichment and statistical power of the ChIP. There are many ChIP-grade antibodies available for a wide range of tags, however the quality of the antibodies can differ between companies and even between different lots.

Before generating libraries for next-generation sequencing, the quality of a ChIP experiment is determined by qPCR. ΔCt values between sample and control ChIP sample for known targets and non-targets determine the specific enrichment of a ChIP experiment. The primers used for the qPCR should be close to 100% in their amplification efficiency (determined by a template dilution series) and should amplify between 90 and 150 bp of the target genomic region. Therefore, careful primer design is needed. Positive control primers can be derived based on knowledge of known binding sites of the protein of interest. If such information is not available, other evidence can be used to determine candidate-binding sites, e.g. based on presence of binding motifs in promoters of genes that are known to be regulated by the protein of interest. We recommend using at least two positive (target) and two negative (non-target) control primer pairs.

Finally, experimental replicates give information about the quality of a ChIP experiment. It is recommended to generate at least two replicates per experiment (Landt *et al.*, 2012).

## Tissue collection and fixation

1. Obtain homozygous plant lines expressing the tagged protein of interest.
2. Collect the plant tissue of interest on ice into a 50-ml tube containing 25 ml of MC buffer. For one ChIP-seq experiment generally 0.5-0.8 gram of material is needed (see **Note 8**). To prevent withering of the tissue, tissue collection should not take longer than 40 min to maximal 1 h (see **Note 9**).
3. Add 676 µl of formaldehyde (see **Note 10**) to the 25 ml of MC buffer to reach a 1% concentration. Quickly fix the tissue on ice under vacuum. Apply vacuum for 15 min, release vacuum and mix the sample briefly, re-apply vacuum for another 14 minutes (total fixation time is 30 minutes) (see **Note 11**).
4. Stop the fixation by adding 2.5 ml of 1.25 M glycine, mix well and apply vacuum for another 2 min.
5. Wash tissue three times with 25 ml of MC buffer (each wash).

6. 'Dry' the tissue using paper towels (see Note 12). Transfer the dried tissue to a new 50-ml tube and freeze in liquid nitrogen (see Note 13).

## Nuclei isolation and chromatin sonication

Wear nitrile gloves for all steps. After the step 1 grinding, make sure to work in a fume hood as buffers M1, M2 and M3 contain β-mercaptoethanol. From step 8 onwards, use low-adhesion tubes and tips at all steps and work in a laminar flow cabinet.

1. Grind the cross-linked tissue in liquid nitrogen in a mortal until the tissue is completely homogeneous (see Note 14).
2. Resuspend the homogeneous tissue in 20 ml of ice-cold M1 buffer.
3. Filter the resulting slurry through a 55 µm cloth mesh in a glass funnel, collect the flow-through in a 50-ml tube on ice. Wash the mesh with an additional 5 ml of M1 buffer to collect all tissue.
4. Centrifuge the filtrate at 1,000 x g for 20 min at 4°C.
5. Remove the supernatant and keep the pellet: the pellet contains the nuclei. Resuspend the nuclear pellet in 5 ml of ice cold M2 buffer, and centrifuge at 1,000 x g for 10 minutes at 4°C.
6. Repeat step 6 four times (5 washing steps with M2 buffer in total).
7. Resuspend the pellet in 5 ml of ice cold M3 buffer and centrifuge at 1,000 x g for 10 min at 4°C.
8. Add 42 µl of 25x protease inhibitor cocktail to 1 ml sonic buffer. Resuspend nuclear pellet in 1 ml of ice cold sonic buffer and transfer to a 2 ml low adhesion microcentrifuge tube (see Note 15).
9. Next, the DNA is sheared using a probe sonicator. Sonicate three times for 15 sec, with 45 sec cooling between repetitions. Invert the tube between sonication steps. Make sure the tube is placed on ice the entire time. After sonication leave the tube on ice for 3-4 min before continuing to the next step (see Note 16).
10. To remove insoluble materials, the sonicated chromatin is centrifuged for 15 min at 4°C at top speed in a microcentrifuge. After centrifuging, transfer the supernatant to a low-adhesive 2 ml tube.
11. Repeat step 10 once (see Note 17).

12. Keep 100 µl of sonicated chromatin aside on ice, this will function as input control (see Note 18).
13. Add 1 ml of IP buffer and 10 µl of 25 x protease inhibitor cocktail.

## Immunoprecipitation of protein-DNA complexes

In this protocol, we describe the use of anti-GFP microbeads, however, also microbeads for other tags are available and can be used in a similar manner. Steps 2 – 9 are performed at room temperature.

1. Add 50 µl of anti-GFP microbeads to the lysate and incubate for 1 hour gently rotating on a tube rotating device (10 rpm) at 4°C to 8°C.
2. Place the µ-Column in the µMACS separator and equilibrate the column with 200 µl of IP-buffer.
3. Apply the lysate onto the column and let it run through by gravity flow.
4. Wash the immobilized beads two times with 400 µl of IP buffer. Drain the column after each wash step by gravity flow.
5. Wash the immobilized beads two times with 200 µl of high salt buffer.
6. Wash the immobilized beads two times with 200 µl of LiCl buffer.
7. Wash the immobilized beads two times with 200 µl of TE.
8. Apply 20 µl of hot elution buffer (95°C) to the column and incubate for 5 min.
9. Elute three times with 50 µl of hot elution buffer (95°C) each into a new 1.5 ml low-binding microcentrifuge tube.

## Reverse crosslinking and DNA purification

1. Add 150 µl of TE to the combined eluate (from step 9 of section "Immunoprecipitation of protein-DNA complexes") and 200 µl of TE to the input DNA (from step 12 of section "Nuclei isolation and chromatin sonication"). Add 11.25 µl of proteinase K (20 mg/ml) to each sample, mix and incubate overnight at 37°C.
2. Next morning, add a second aliquot of 11.25 µl proteinase K to each sample, mix and incubate at 65°C for 6 - 10 hours.
3. Precipitate the DNA by adding 2.5 vol of 100% ethanol, 1/10 vol of 3 M sodium acetate pH 5.4, and 1 µl of glycogen and incubate overnight at -20°C.

4. Centrifuge for 30 min at 4°C at top speed in a microcentrifuge.
5. Remove the supernatant and air dry the pellet for approximately 15 min, do not over dry the pellet this will reduce resuspension efficiency.
6. Resuspend the pellet in 100 µl ultrapure water (see Note 19).
7. Purify the DNA using the QIAquick PCR purification kit according to the manufacturer's instructions and elute with 34 µl elution buffer into 1.5 ml low-adhesive microcentrifuge tubes (see Note 20).

## ChIP quality validation using quantitative PCR (qPCR)

To test the quality of the ChIP, qPCR is used to confirm enrichment. We recommend to use at least two positive (bound genomic regions) and two negative (non-bound regions) primer pairs. Enrichment can be roughly determined using the ΔΔCT determined by comparing CT thresholds of negative and positive control genes, normalized using the sample primers on the negative control DNA sample. Fold enrichments of 8-16 times the levels of the negative controls is normally an indication of a good ChIP sample. The CT values of the negative control primer pairs give an indication about the DNA amount: the CT values should be less than 34. When these CT values are higher than 34, the amount of DNA may be too low to continue with library preparation.

1. Take a 1 µl aliquot of the ChIP sample as well as the control sample and dilute it 1:5 with milliQ water. Input DNA controls need to be diluted 1:1000 to 1:5000 with miliQ water.
2. Prepare the primer master mixes by combining forward and reverse primer at a final concentration of 1 µM per primer.
3. Combine 5 µl of Primer master mix, 2.5 µl of milliQ water, 5 µl of diluted DNA sample and 12.5 µl of iQ SYBR Green Supermix.
4. Set-up and run the real-time PCR machine. The PCR program is (Bio-Rad iQ5): 3 min incubation at 95°C, followed by 40 cycles of [15 sec at 95°C, 1 min at 60°C] (see Note 21).

## DNA library preparation for Illumina sequencing: end-repair and A-tailing

1. Combine and mix the following components in a 1.5 ml low-adhesion microfuge tube on ice:

| | |
|---|---|
| ChIP DNA | 34 µl |
| 10x End-repair buffer | 5 µl |

| | |
|---|---|
| 2.5 mM dNTP Mix | 5 µl |
| 10 mM ATP | 5 µl |
| End-repair enzyme mix | 1 µl |
| | --------- + |
| Total reaction volume | 50 µl |

2. Incubate for 45 minutes at room temperature (18-20°C).
3. Purify the end-repaired DNA using the QIAquick PCR Purification Kit and protocol, elute with 34 µl of elution buffer.
4. For A-tailing, combine and mix the following components in a 0.5 ml low-binding microcentrifuge tube:

| | |
|---|---|
| Purified DNA from end-repair reaction | 34 µl |
| NEB buffer 2 | 5 µl |
| 1 mM dATP | 10 µl |
| Klenow fragment (3' to 5' exo minus) | 1 µl |
| | --------- + |
| Total reaction volume | 50 µl |

5. Incubate the A-tailing mix for 30 min at 37°C.
6. Purify the A-tailed DNA on a QIAquick MinElute column using the minElute PCR Purification Kit and protocol, elute with 17 µl of elution buffer.

## Adapter ligation

1. Combine and mix the following components in a 1.5 ml low-adhesion microfuge tube (see Note 22):

| | |
|---|---|
| End-repaired and A-tailed DNA | 16.5 µl |
| 10x T4 DNA ligase buffer | 2.0 µl |
| Illumina adaptor oligo mix (diluted 1:10 – 1:50) | 1.0 µl |
| T4 DNA ligase | 0.5 µl |
| | --------- + |
| Total reaction volume | 20.0 µl |

Incubate for 20-22 hours at 16°C for efficient adapter ligation.

2. Purify the DNA using the QIAquick PCR purification kit and protocol, elute with 30 µl of elution buffer.

## PCR and size selection

1. Prepare the following master mix:

Per reaction mix (always prepare 1 reaction mix more than needed):

| | |
|---|---|
| 5 x Phusion buffer | 10.0 µl |
| 2.5 mM dNTP | 4.0 µl |
| Phusion enzyme | 0.8 µl |
| PCR primer 1.1 | 0.5 µl |
| PCR primer 2.1 | 0.5 µl |
| $dH_2O$ | 4.2 µl |
| | --------- + |
| Total reaction volume | 20.0 µl |

2. Combine and mix in a PCR tube:

| | |
|---|---|
| DNA from Step 2 of section "Adapter Ligation " | 30 µl |
| Phusion master mix from Step 1 of section "PCR and Size Selection" | 20 µl |
| | -------- + |
| Total reaction volume | 50 µl |

3. Amplify using the following PCR protocol:

30 sec at 98°C

[10 sec at 98°C, 30 sec at 65°C, 30 sec at 72°C]

Cycles total 14 – 20 (depending of your starting amount) (see **Note 23**)

5 min at 72°C

Hold at 4°C

4. Purify the PCR reaction using the QIAquick MinElute PCR Purification Kit and protocol. Elute with 11 µl elution buffer.
5. For loading on a gel, add 2 µl of 10x gel loading dye and 1 µl of glycerol to the DNA.
6. For size selection, load DNA on a clean 2% (wt/vol) agarose gel in 1x TE and 0.5 µg/mL ethidium bromide with a Smart ladder 1 kb-100 bp as marker, run at 135 V for 30 min.

7. Cut a large band of 200 or 250 bp to 500 bp in size from the gel using a clean scalpel (see **Figure 2**). Take pictures before and after excising the DNA to allow calculation of the median product size. If multiple libraries are being prepared, be cautious of cross contamination and run only one sample per gel.
8. Purify the DNA from the gel using QIAgen MinElute Gel purification Kit. Dissolve the gel at room temperature and after purification elute the DNA with 18 µl of elution buffer.

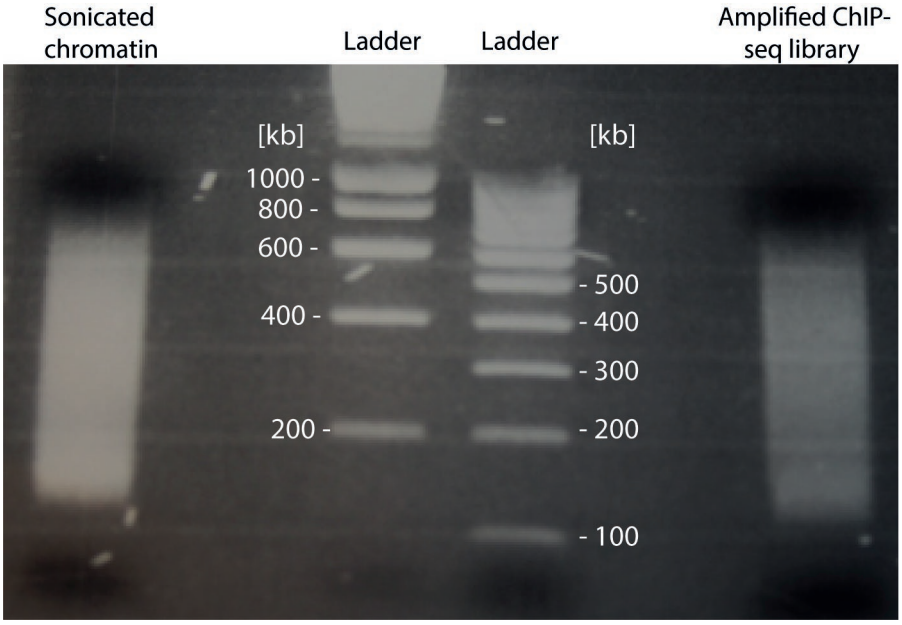



**Figure 2 | Example of sonicated chromatin (left) and amplified ChIP-seq library (right) on a 2% agarose gel.** The size of the DNA ladders are indicated on the side of the gel. Prior to sequencing the amplified ChIP-seq libraries, a gel band representing 200 to 500 bp needs to be cut out and purified.

## Quality control

To test that the library preparation was successful, test the enriched DNA using qPCR. For the criteria to determine library quality see "ChIP Quality Validation using Quantitative PCR (qPCR)"

1. Take a 0.25 µl aliquot of the library and dilute 20 fold.
2. Prepare a primer master mix by combining forward and reverse primer at a final concentration of 1 µM per primer.

3. Combine 5 µl of primer master mix, 2.5 µl of milliQ water, 5 µl of diluted DNA sample and 12.5 µl of iQ SYBR Green Supermix.
4. Set-up and run the real-time PCR machine. The PCR program is (Bio-Rad iQ5): 3 min incubation at 95°C, followed by 40 cycles of [15 sec at 95°C, 1 min at 60°C].

### Illumina sequencing

1. Measure the DNA concentration and size by Invitrogen Qubit measurement (Qubit® dsDNA HS Assay Kit) and Agilent BioAnalyzer DNA 1000 chip. Use the average size of DNA as indicated by the Bioanalyzer and the concentration of the Qubit to determine the concentration of the sample.
2. Dilute the sample to 10 nM using elution buffer of the QIAgen MinElute Gel purification Kit (see Note 24).
3. Prepare Illumina flow-cell following the Illumina protocol (see Note 25).

## Bioinformatic analysis of ChIP-seq data

Sequencing of ChIP-seq libraries leads to millions of short read sequences representing mainly genomic regions originally bound by the protein of interest. Therefore, genuine DNA binding sites should show an enrichment in sequence reads when comparing ChIP-seq versus the control libraries. This enrichment should be detected by a proper statistical methodology. In this section, we present a protocol for computational analysis of ChIP-seq data.

Sequence reads obtained by an Illumina sequencer are normally provided in a FASTQ format. This format contains not only the read sequence, but also gives information about the sequencing quality of the read. We advise to use the FastQC software (Babraham Bioinformatics) to check the sequence quality of our libraries. This software enables easy identification of common problems that can arise from either sequencing or library preparation.

If the libraries are of sufficient quality (see section "Library Quality Check"), reads can be mapped to the reference genome using one of the many available mapping tools (e.g. SOAP (Li *et al.*, 2009), Bowtie (Langmead *et al.*, 2009), and BWA (Li and Durbin, 2009)). Typically, the mapping procedure

is done allowing up to two mismatches, but the user may allow a higher number of mismatches when working with samples of species distant from the reference genome used (e.g. a different ecotype than Col-0 for *Arabidopsis thaliana*). Reads that map to more than one location in the genome are usually discarded, since it is not possible to determine the origin of these reads.

After mapping of the ChIP and control libraries, DNA-bound regions can be identified by detection of genomic regions with higher amount of reads in the ChIP libraries compared to control libraries. There are many publicly available packages for the analysis of ChIP-seq datasets, for example MACS (Zhang *et al.*, 2008), PeakSeq (Rozowsky *et al.*, 2009), CisGenome (Ji *et al.*, 2008), and CSAR (Muiño *et al.*, 2011). In this protocol, we choose to use CSAR for the statistical detection of protein binding sites in the DNA. CSAR is a package written in the popular R language and has proven to be an efficient and accurate tool for the analysis of plant ChIP-seq data . CSAR has been used for multiple ChIP-seq studies in *Arabidopsis* (Kaufmann *et al.*, 2009; Schiessl *et al.*, 2014), but also in other non-plant species (Outchkourov *et al.*, 2013).

In this section, we provide instructions on the analysis of ChIP-seq datasets. All code is provided as Unix shell to allow line-by-line processing of the dataset. In the protocol we work with three example datasets, two samples and a control, to which we will refer as "sequences-IP1.fq.gz", "sequences-IP2.fq.gz", and "sequences-control.fq.gz". We assume that these ChIP-seq datasets are generated from an *Arabidopsis thaliana* background. However, the protocol can easily be adapted for other plant species. If you are unfamiliar with working in a shell, please have a look at http://www.ks.uiuc.edu/Training/Tutorials/Reference/unixprimer.html for some basic commands that will help to better understand this protocol.

## Library quality check

It is recommended to start the ChIP-seq data analysis by performing a library quality check. FASTQC is a JAVA program able to visually report the quality of Illumina sequencing libraries. It has an intuitive graphical interface that allows a user-friendly experience. Important information provided by the software is:

- Per base sequence quality:

When the program indicates a problem for per base sequence quality, the user should consider to increase the maximum number of mismatches allowed in the mapping process.

- Adapter content:

  The program will report proportion of adapter sequences found in one's library.

- Sequence duplication level:

  A high duplication level may result from a problem during the PCR amplification usually associated with low amount of starting material.

For details on how to install and use the program, please, refer to its manual http://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/.

## Data filtering

Make a directory to place the raw sequence files (if not already exist):

```
$ cd <yourhomedirectory>
$ mkdir raw-data
$ cd raw-data
```

Enter your raw sequence files into this directory.

Often the files obtained from the sequencer are compressed (e.g. .gz or .zip files). In the case of a compressed file, unpack the file before further analysis:

```
$ gunzip -c sequences-IP1.fq.gz > sequences-IP1.fq
$ gunzip -c sequences-IP2.fq.gz > sequences-IP2.fq
$ gunzip -c sequences-control.fq.gz > sequences-control.fq
```

or

```
$ unzip -n sequences-IP1.fq.zip > sequences-IP1.fq
$ unzip -n sequences-IP2.fq.zip > sequences-IP2.fq
$ unzip -n sequences-control.fq.zip > sequences-control.fq
```

When using Illumina sequencing platforms, reads are provided in CASAVA FASTQ format. Since the version 1.8 of the CASAVA FASTQ format, the file generated contains reads that passed Illumina quality filter and reads that did not pass the filter. Only the reads that did pass the filter should be used as input for SOAPv2.

A CASAVA 1.8 FASTQ file containing four sequences may look like this (see Note 26):

```
@HWI-ST995:196:D1FU6ACXX:5:1101:1424:1984 1:N:0:
TTTTCCTGAAGGGATCCTTGAATATTGCCTGAGGGTATAGAATGACTTCAC
+
=1=DDDDD>3D?FFFDEFHEHDEH@HHHD?FBA@GGFFF>GFFFAFFGFBG
@HWI-ST995:196:D1FU6ACXX:5:1101:1508:1972 1:N:
ATAGCAACGCGGAACAGTCGACCGTATCAGGAGATAGCATGTCTCAGTTGG
+
G1:BDDDDHHHHCEHBHDGEDGHE@DHIGGGHHGIGEGG9@CHEHHHH>CE
@HWI-ST995:196:D1FU6ACXX:5:1101:1556:1981 1:N:0:
GTCAATAGAATCCTCGATCTTCTTCTTGTCTGCAGCCGGGAGCTTCTCACC
+
11=?DBEFFHHHHJGIIIJI@HDEFHCDHDIJJGIJIJJB@FHGGIJJIII
@HWI-ST995:196:D1FU6ACXX:5:1101:1943:1919 1:Y:0:
NTCANAACNNNNNNNNAGTCCACTCTCAGGTTGGAAGCGAANCATGTCTGC
+
#07?#2==########22@==>>>?@@???=?<??>?9==;#0<=???=??
```

The first line, starting with "@", is the identifier line. This line gives information about the sequence platform used, the flow cell, the location of the cluster on the flow-cell, etc. Since CASAVA version 1.8, this line also reports whether the read is of sufficient quality ("N") or it should be filtered out ("Y"). For read mapping, only use the reads indicated by "N".

For filtering reads, use the Perl script 'convert_export_to_fasta.pl' (see section "Equipment and software for Bioinformatic analysis"):

```
$ cd <yourhomedirectory>/raw-data
$ perl <yourhomedirectory>/code/convert_export_to_fasta.pl
sequences-IP1.fq > sequences-IP1_filtered.fa
$ perl <yourhomedirectory>/code/convert_export_to_fasta.pl
sequences-IP2.fq > sequences-IP2_filtered.fa
$ perl <yourhomedirectory>/code/convert_export_to_fasta.pl
sequences-control.fq > sequences-control_filtered.fa
```

## Mapping to the reference genome

There are many programs developed for fast and accurate alignment of reads to a reference genome. In our protocol, SOAP2v2 is used. SOAPv2 is chosen because this program trims non-mappable reads at the 3′ end until the sequence can be mapped to the genome or it is too short to be mapped. 3′ end trimming is beneficial as reads usually have more errors towards the 3′ end of the sequence. Typically, only reads that map to one unique location in the genome with maximum of two mismatches are considered.

Like other short read aligners, SOAPv2 uses index files of the reference genome for fast alignment of reads to the genome. Therefore, one should reformat the sequence genome in the following way:

```
$ cd <yourhomedirectory>/genomes
$ <yourhomedirectory>/software/SOAPaligner/2bwt-builder
Arabidopsis_thaliana.TAIR10.22.dna.genome.fa
```

Next, the reads need to be mapped to the reference genome using the SOAPv2 and the generated indexes (see Note 27):

```
$ cd <yourhomedirectory>
$ mkdir mapped
$ cd mapped
$ <yourhomedirectory>/software/soap2.20/soap -r 0 -a
<yourhomedirectory>/raw-data/sequences-IP1_filtered.fa -D
<yourhomedirectory>/genomes/Arabidopsis_thaliana.TAIR10.22.dna.ge
nome.fa.index -o sequences-IP1_filtered.soap2
$ <yourhomedirectory>/software/soap2.20/soap -r 0 -a
<yourhomedirectory>/raw-data/sequences-IP2_filtered.fa -D
<yourhomedirectory>/genomes/Arabidopsis_thaliana.TAIR10.22.dna.ge
nome.fa.index -o sequences-IP2_filtered.soap2
$ <yourhomedirectory>/software/soap2.20/soap -r 0 -a
<yourhomedirectory>/raw-data/sequences-control_filtered.fa -D
<yourhomedirectory>/genomes/Arabidopsis_thaliana.TAIR10.22.dna.ge
nome.fa.index -o sequences-control_filtered.soap2
```

When read mapping is finished, SOAPv2 provides an alignment report. This report states the percentage of mapped reads and the elapsed time. The provided percentage of mapped reads is a measure for the statistical power of the sequencing data. Usually a good ChIP-seq library produces at least 30% uniquely mapped reads (Kaufmann *et al.*, 2010b).

## Using CSAR for ChIP-seq analysis

After read mapping, CSAR can be used for the detection of genome-wide DNA-binding regions. First, create a directory for the ChIP-analysis and load the mapped libraries (.soap2 files):

```
$ cd <yourhomedirectory>
$ mkdir CSAR
> R
> setwd("<yourhomedirectory>/CSAR")
> library(CSAR)
```

```
> IP1<- loadMappedReads("<yourhomedirectory>/mapped/sequences-
IP1_filtered.soap2", format="SOAP", header=FALSE)
> IP2<- loadMappedReads("<yourhomedirectory>/mapped/sequences-
IP2_filtered.soap2", format="SOAP", header=FALSE)
> control<- loadMappedReads
("<yourhomedirectory>/mapped/sequences-control_filtered.soap2",
format="SOAP", header=FALSE)
```

The first step of peak detection is directional, according to the DNA strand to which the read is mapped, to count the number of reads mapping to each position in the genome. This is done with the function *mappedReads2Nhits*. For this function, some parameters need to be specified: specify with *chr* the chromosome names exactly as present in the input file; *chrL* specifies the chromosome length in base pairs; and *w* specifies the average read-length of the library. This average is determined from running the library on electrophoresis gel (see Note 28).

```
> IP1 <-mappedReads2Nhits(IP1, file="sequences-IP1_w300", chr =
c("Chr1", "Chr2", "Chr3", "Chr4", "Chr5"), chrL= c(30427671,
19698289, 23459830, 18585056, 26975502), w=300)
> IP2 <-mappedReads2Nhits(IP2, file="sequences-IP2_w300", chr =
c("Chr1", "Chr2", "Chr3", "Chr4", "Chr5"), chrL= c(30427671,
19698289, 23459830, 18585056, 26975502), w=300)
> control <-mappedReads2Nhits(control, file="sequences-
control_w300", chr = c("Chr1", "Chr2", "Chr3", "Chr4", "Chr5"),
chrL= c(30427671, 19698289, 23459830, 18585056,26975502), w=300)
```

The generated non-normalized nucleotide read count scores can be visualized in a genome browser (e.g. IGB (http://bioviz.org/igb/)). Visualization is a good way to get a first impression of the ChIP experiments and an idea about their quality. To visualize the read count values per nucleotide in a genome browser, we need to generate a file in Wiggle (WIG) format. Generating a WIG file representing the whole genome will result in files with a too big size. For this reason, we use the parameter *t* to indicate that the WIG file should only contain regions with at least `t` reads mapped. In the next example, we use `t=10`:

```
> IP1$digits<-0
> IP2$digits<-0
> control$digits<-0
> score2wig(IP1, "sequences-IP1_w300.wig", t=10,
description="Read count-IP1")
> score2wig(IP2, "sequences-IP2_w300.wig", t=10,
description="Read count-IP2")
```

```
> score2wig(control, "sequences-control_w300.wig", t=10,
description="Read count-Control")
```

Next, calculate the normalized read enrichment score for each nucleotide position:

```
> test1<-ChIPseqScore(control = control, sample = IP1, file =
"control.v.IP1@TAIR10", backg= -1, norm= -1)
> test2<-ChIPseqScore(control = control, sample = IP2, file =
"control.v.IP2@TAIR10", backg= -1, norm= -1)
```

Generate a WIG file, to be able to visualize the results of the analysis in a genome browser. For a detailed description about visualization in a genome browser, see . The parameter *t* indicates that only regions with a score greater than *t* will be represented by the WIG file:

```
> score2wig(test1, "control.v.IP1.wig", t=2,
description="IP1.v.control")
> score2wig(test2, "control.v.IP2.wig", t=2,
description="IP2.v.control")
```

Obtain the list of candidate binding sites and save it into a .csv file (see **Note 29**):

```
> win1<-sigWin(test1, t=1, g = 100)
> write.csv(as.data.frame(win1),file="IP1.v.control-
BindingSites.csv")
> win2<-sigWin(test2, t=1, g = 100)
> write.csv(as.data.frame(win2),file="IP2.v.control-
BindingSites.csv")
```

Next, to obtain the list of genes with binding site(s) in their proximity, the function *distance2Genes* can be used. This will generate a list of candidate target genes (see **Note 30**):

```
> genes<-
read.table("<yourhomedirectory>/genomes/TAIR10_GFF3_onlygenes.gff")
> distance1<-distance2Genes(win1, genes)
> genes1 <- genesWithPeaks(distance1)
> write.csv(genes1,file="IP1.v.control-targets.csv")
> distance 2<-distance2Genes(win2,Genes)
> genes2 <- genesWithPeaks(distance2)
> write.csv(genes2,file="IP2.v.control-targets.csv")
```

The generated .csv file contains a tab-delimited table that lists genes with candidate enriched regions located near them. The file can be opened with Excel.

## Testing reproducibility

As stated earlier, to ensure reproducibility, it is recommended to have at least two biological replicates per sample. Reproducibility between two replicates can be measured using Pearson correlation coefficients (PCC). Unrelated samples usually have a Pearson correlation between 0.3-0.4, highly reproducible experiments have a correlation of more than 0.9 (Bardet *et al.*, 2012).

Open R and generate a pdf file to store the scatterplots:

```
$ R
> pdf("scatterplots_ChIPrepoducibility.pdf")
```

Combine the datasets of the biological replicates:

```
> bs<-merge(genes1, genes2, by="name")
```

Generate scatterplots and determine Pearson correlation:

```
> cor=cor(log10(bs$u1000.x), log10(bs$u1000.y), method =
"pearson")
> plot(log10(bs$u1000.x), log10(bs$u1000.y), xlab="sequences_IP1",
ylab="sequences_IP2", cex = 0.4, main = paste("Reproducibility
ChIPseq\nR=", cor))
> abline(a = 0, b = 1, col = "blue")
```

Save and close the pdf file:

```
> dev.off()
```

The produced scatterplots and Pearson correlation coefficients can be found in the generated pdf file: "scatterplots_ChIPrepoducibility.pdf".

## Calculate FDR thresholds

To calculate FDR thresholds, the distribution of the normalized scores under the null hypothesis using permutations should be generated (see **Note 31**). In this example, 20 sets of permutations will be produced (indicated by the variable *nn*).

```
> nn=20
> for(i in 1:nn){permutatedWinScores(nn=i, control, IP1,
fileOutput="IP1.v.input_b-1",chr = c("Chr1", "Chr2", "Chr3",
"Chr4", "Chr5"), chr = c("Chr1", "Chr2", "Chr3", "Chr4", "Chr5"),
chrL= c(30427671, 19698289, 23459830, 18585056,26975502), w=300,
backg=-1, norm=-1)}
> for(i in 1:nn){permutatedWinScores(nn=i, control, IP2,
fileOutput="IP2.v.input_b-1",chr = c("Chr1", "Chr2", "Chr3",
"Chr4", "Chr5"), chr = c("Chr1", "Chr2", "Chr3", "Chr4", "Chr5"),
chrL= c(30427671, 19698289, 23459830, 18585056,26975502), w=300,
backg=-1, norm=-1)}
```

Make sure to have at least 50,000 scores are obtained after permutation. More permutations can be generated by increasing the value of *nn*.

Next, obtain the FDR thresholds (see Note 32):

```
> nulldist1 <- getPermutatedWinScores(file = "IP1.v.input_b-1", nn
= 1:20)
> getThreshold(winscores = values(win1)$score, permutatedScores =
nulldist1, FDR = 0.05)
> nulldist2<- getPermutatedWinScores(file = "IP2.v.input_b-1", nn
= 1:20)
> getThreshold(winscores = values(win2)$score, permutatedScores =
nulldist2, FDR = 0.05)
```

This can be done for all desired FDR cut-offs by changing the value of *FDR* to the FDR threshold desired.

## Downstream analysis: genomic peak sequence extraction

CSAR provides lists of genes in close proximity of DNA-binding sites including significance, representing potential "direct" target genes, and generates a DNA-binding map in a WIG format for visualization in a genome browser. After the initial bioinformatic analysis, downstream analysis is needed to provide additional information and allow a better interpretation of the data. A commonly performed downstream analysis is *in vivo* motif discovery. As a first step in motif discovery, a FASTA-file containing DNA-sequences associated with ChIP-seq peaks needs to be generated. The script for extraction of these sequences is given below. In this example, a sequence of 300 bp, 150 bp left and right of the peak summit, is extracted. Please see Muino *et al.* 2011 for an extensive explanation about *in vivo* motif discovery and other downstream analyses, e.g. peak visualization and GO classification.

```
$ R
> setwd("<yourhomedirectory>/CSAR")
> library(Biostrings)
> seq<-
readDNAStringSet("<yourhomedirectory>/genomes/Arabidopsis_thalian
a.TAIR10.22.dna.genome.fa")
> test1<-read.csv(file="IP1.v.control-BindingSites.csv")
> test2<-read.csv(file="IP2.v.control-BindingSites.csv")
```

```
Take only those peaks having a peak score higher than the FDR
threshold value wanted (in this case at FDR 0.05 the determined
FDR threshold value is 2.00 for both ChIP experiments)
> test1<-test1[test1$score>2.00,]
> test2<-test2[test2$score>2.00,]
```

Extract *w* bp sequence region around the peak position and save it in a FASTA format (see **Note 33**):

```
> w=300
> test1<-test1[(test1$posPeak > 150),]
> test2<-test2[(test2$posPeak > 150),]
> finalseq1<-DNAStringSet()
for (i in unique(test1$seqnames)){
print(i)
tempwin<-test1[test1$seqnames==i,]
seq1<-DNAStringSet(seq[[i]],tempwin$posPeak-w,tempwin$posPeak+w)
names(seq1)<-
paste(tempwin$seqnames,tempwin$posPeak,tempwin$score,sep="_")
finalseq1<-c(finalseq1,seq1)}
>
writeXStringSet(finalseq1,file="control.v.IP1_300bp.fasta",format
="fasta")
> finalseq2<-DNAStringSet()
for (i in unique(test2$seqnames)){
print(i)
tempwin<-test2[test2$seqnames==i,]
seq2<-DNAStringSet(seq[[i]],tempwin$posPeak-w,tempwin$posPeak+w)
names(seq1)<-
paste(tempwin$seqnames,tempwin$posPeak,tempwin$score,sep="_")
finalseq1<-c(finalseq2,seq2)}
>
writeXStringSet(finalseq2,file="control.v.IP2_300bp.fasta",format
="fasta")
```

## Notes

1. Although a probe sonicator (e.g. MSE Soniprep 150) had proven to work well, also newer types of sonicators (e.g. BIORUPTOR, Diagenode, Liege, Belgium) have been used successfully in ChIP-seq experiments.
2. The µMACS istope Isolation Kit is available for many different tags, including HA, c-myc, GST and His. Kits using magnetic beads can also be obtained from other vendors, though the protocol may need to be adapted in this case.
3. It is recommended to use 10x T4 ligase buffer instead of the 2x buffer supplied by the Illumina ChIP-seq kit. Buffer needs to be vortexed before use to resuspend DTT.
4. Also other ChIP-seq preparation kits have been used successfully, e.g. NEXTflex™ ChIP-seq Kit (BIOO Scientific). The library preparation procedure may need to be adjusted when using another kit.
5. To obtain the link needed to download the files using the shell, go to the Ensembl website. Downloading the files using the shell is potentially faster.
6. When interested in genomic features other than genes (e.g. miRNAs, ncRNAs, etc.), the script can be adjusted for extracting more than one genomic feature by: `> gff<-gff[gff$V3=="gene"| gff$V3=="miRNA",]`
7. For different species the exact "name" for genomic features in the gff3 file can be different. Therefore, check the different "names" using: "`> unique gff$V3`" and adjust the script accordantly.
8. The amount of tissue can differ based on the expression level and expression domain of the protein.
9. The stability of the plant material is dependent on the tissue used. For example, water content and sample size can influence tissue stability.
10. Formaldehyde is most often used as cross-linking agent for ChIP experiments. In some cases, such as fixing proteins bound to DNA in a complex, additional fixatives can be beneficial.
11. Make sure the tissue is submerged in the buffer to allow proper fixation of the tissue.

12. Drying of the tissue is needed to prevent ice-crystals from forming. Ice-crystals can make sample grinding more difficult and residual buffer will 'dilute' the M1 buffer used in the first nuclei isolation step.
13. After this step, the tissue can be stored for several months at -80ºC.
14. Ground tissue can be stored at -80ºC for up to two days.
15. The resuspended nuclear pellet is viscous, so a cut or wide-bore pipet tip is recommended.
16. The sonication step depends strongly on the sonicator used. It is recommended to test the sonication efficiency before ChIP experiments are conducted. Sheared chromatin can be analysed after de-crosslinking and DNA purification using gel electrophoresis; the majority of the DNA should have a size between 200 and 800 bp.
17. If a pellet is visible after two rounds of centrifugation, add another centrifugation step to make sure all insoluble material is removed.
18. The input sample is not used for immunoprecipitation. Continue processing the input-DNA at section "Reverse Crosslinking and DNA Purification". It is important to handle the input DNA strictly separate from the ChIP sample to avoid contamination.
19. To check shearing efficiency, after de-cross linking, run 10 µl of input DNA on gel (see Figure 2).
20. Purified DNA can be stored at -20ºC for up to 4 months.
21. When the sample has a low signal-to-noise ratio, corresponding to a low enrichment in the ChIP compared to the control, this could be caused by: wrong choice of a positive control (genomic region is not strongly bound by the protein of interest); primer efficiently of less than 100%; too much insoluble material in the sample (see Note 17); antibody of low quality; low protein abundance level in collected tissue; too long or too short fixation; loss of proteins during nuclei preparation; and/or low protein stability and/or sub-optimal sonication conditions.
22. Use adapters diluted 10 to 50 times to adjust for the smaller amount of DNA. Excess adapters can interfere with sequencing. Optimal concentrations of adapters may need to be determined empirically.
23. Optimal number of PCR cycles depends on the amount of DNA in your sample. The optimal number of cycles can vary between 14 and 20 cycles. The minimum number of

cycles should be used to avoid bias due to PCR over-amplification, while obtaining sufficient material for sequencing.

24. The 10 nM library can be stored at -20ºC for several months.
25. For *Arabidopsis thaliana,* we prefer to run 50-bp single-end sequencing, and it is recommended to have at least 1-10x coverage (Pepke *et al.*, 2009). However, the optimal read-length and preferred sequencing depth differs between species. Longer sequences or paired-end will improve the mapping efficiency of the sample.
26. The format can be checked by printing to the screen the first lines of your sequence file using "`> less sequence-IP1.fq`".
27. Different options can be selected for SOAPv2. The options used here are: `-r` defines how SOAPv2 reports multi-hits (`0`= none, `1`= random, `2`=all); `-a` specifies the input file containing the reads; `-D` specifies the SOAPv2 index files; and `-o` specifies the output file. To specify the number of threads used by SOAPv2, use the parameter `-p`: using more threads makes the mapping faster.
28. For the function `mappedReads2Nhits` the parameters `chr` and `chrL` should be adjusted depending on the species used. `w` should be adjusted depending on the average fragment size of the library.
29. *sigWin* allows specification of a score cut-off (`t`): only read-enrichment regions with score higher than this cut-off are reported as candidate binding site regions.
30. The `distance2Genes` function parameters can be adjusted to the preferred distances to genes by changing: "`d1= negative integer`", minimum relative position with respect to the start of the gene to be considered; and "`d2= positive integer`", maximum relative position with respect to the end of the gene to be considered.
31. When making permutated datasets make sure to use the same parameters values as used for the analysis of the sample and control.
32. Dependent on the preferred FDR value, the threshold can be determined using the function `getThreshold` by changing "`FDR = 0.05`" to another value.
33. When using another length of extracted sequence, change the values of the variable `w` to the desired lengths.

# Chapter 4

## FRUITFULL controls *SAUR10* expression and regulates *Arabidopsis* growth and architecture

Marian Bemer
Hilda van Mourik
Jose M Muiño
Cristina Ferrándiz
Kerstin Kaufman
Gerco C Angenent

Modified from JEXBOT/2017/196188

## Abstract

MADS domain transcription factors are well-known for their roles in plant development, and regulate sets of downstream genes that have been uncovered by high-throughput analyses. A considerable number of these targets are predicted to function in hormone responses or responses to environmental stimuli, suggesting that there is a close link between developmental and environmental regulators of plant growth and development. Here, we show that the *Arabidopsis* MADS domain factor FRUITFULL (FUL) executes several functions in addition to its noted role in fruit development. Among the direct targets of FUL, we identified *SMALL AUXIN UPREGULATED RNA 10* (*SAUR10*), a growth regulator that is highly induced by a combination of auxin and brassinosteroids, and in response to reduced R:FR light. Interestingly, we discovered that *SAUR10* is repressed by FUL in stems and inflorescence branches. *SAUR10* is specifically expressed at the abaxial side of these branches, and this localized activity is influenced by hormones, light conditions and by FUL, which has an effect on branch angle. Furthermore, we identified a number of other genes involved in hormone pathways and light signalling as direct targets of FUL in the stem, demonstrating a connection between the developmentally and environmentally regulated growth programs.

# Introduction

Plant growth and development are regulated by an interplay between internal and external factors. The timely expression of different sets of transcription factors regulates the default program of plant growth and development, but this program is highly influenced by external factors that allow the plant to adapt its growth according to the environmental conditions. As a result, plants with the same genotype show distinct phenotypic differences when for example grown at different temperatures or under different light conditions. This response to environmental conditions is mainly regulated via hormonal pathways, and involves auxin, gibberellic acid (GA), cytokinin and brassinosteroids (BR). In particular auxin, which induces cell-elongation, has been shown to be essential for growth responses to environmental conditions such as phototropism and gravitropism (Fankhauser and Christie, 2015; Paponov *et al.*, 2008). Recently, the light-regulated growth of *Arabidopsis* hypocotyls has been thoroughly investigated, and revealed to depend on physical interactions between transcription factors involved in auxin, GA, BR and light responses (Bai *et al.*, 2012; Oh *et al.*, 2014; Ross and Quittenden, 2016). Downstream growth-regulating genes can be induced or repressed by either the hormone-mediated environmental response pathway or by the internal developmental pathway, thus integrating these two pathways in the growth response.

A group of growth regulators that has been shown to be highly responsive to auxin and other hormonal stimuli is the SAUR family of Small Auxin-Upregulated RNAs. *SAUR* transcripts were first discovered in soybean and found to be rapidly upregulated after addition of auxin. Additional research in soybean and other species has revealed that SAUR activity is highly dynamic, as both transcript and protein half-lives were reported to be extremely short (Knauss *et al.*, 2003; McClure and Guilfoyle, 1987). In *Arabidopsis*, the *SAUR* gene family contains 79 genes (Ren and Gray, 2015), of which approximately two-third have been found to respond to auxin in certain tissues (Bargmann *et al.*, 2013; Chapman *et al.*, 2012; Paponov *et al.*, 2008). In addition, several *SAUR* genes have been found to be influenced by other hormones like ABA, ethylene, GA and BR (Kodaira *et al.*, 2011; Li *et al.*, 2015; Oh *et al.*, 2014; Stamm and Kumar, 2013; Walcher and Nemhauser, 2012). The function of several *Arabidopsis SAUR* genes has been investigated using overexpression studies, unveiling their general capacity to promote cell-elongation in growth-related processes (Chae *et al.*, 2012; Ren and Gray, 2015; Spartz *et al.*, 2012; Stamm and Kumar, 2013; Sun *et al.*, 2016). For a

long time, it was unknown how induced *SAUR* gene expression could result in increased cell elongation, but a study by Spartz *et al.* (2014) recently unveiled that SAURs act according to the earlier postulated acid-growth theory (Rayle and Cleland, 1992). They interact with protein phosphatases of the PP2C-D family to inhibit their function, thereby preventing dephosphorylation of plasma membrane $H^+$-ATPases, resulting in activation of these membrane pumps. Activation of the $H^+$-ATPases leads to membrane acidification, which enables cell elongation. Different *Arabidopsis* SAURs were tested and they were all able to interact with PP2C-Ds (Spartz *et al.*, 2014; Sun *et al.*, 2016).

In addition to being responsive to hormones, *SAUR* genes have also been reported as targets of several transcription factors involved in plant development, like the MADS domain transcription factors SEPALLATA3 (SEP3) and APETALA1 (AP1), and the TCP (TEOSINTE BRANCHED1/CYCLOIDEA/ PROLIFERATING CELL FACTOR1) family protein TCP200 (Kaufmann *et al.*, 2009; Kaufmann *et al.*, 2010c; Danisman *et al.*, 2012), suggesting that growth-regulators of the SAUR family can act as integrators of the developmental and environmental growth pathways. However, the interaction between both pathways during plant growth is poorly understood.

We performed a ChIP-seq experiment with the *Arabidopsis* MADS domain factor FRUITFULL (FUL) and identified two closely related *SAUR* genes, *SAUR10* and *SAUR16*, as strongly bound target genes. FUL is a major player in the network that regulates *Arabidopsis* fruit development, and determines both fruit patterning and growth (Ferrándiz *et al.*, 2000b; Gu *et al.*, 1998). In addition, *ful* mutants were also reported to flower later than wild type (Ferrándiz *et al.*, 2000a), and to exhibit an altered cauline leaf shape (Gu *et al.*, 1998). Here, we demonstrate that FUL plays additional and novel roles in plant growth, and is able to directly regulate genes involved in hormone- and light-induced cell elongation, such as the DELLA genes *RGL2* and *GAI*, *PHYTOCHROME INTERACTING FACTOR 3-LIKE 1* (*PIL1*), the *CYTOKININ OXIDASES CKX5* and *CKX6*, and *SAUR10*. The architecture phenotype of *ful* mutants, which exhibit more vertical branch growth, can be explained by the de-repression of *SAUR10*, which is specifically expressed at the abaxial side of the branch. *SAUR10* is repressed by FUL in the stem, but can be highly induced by a combination of auxin and brassinosteroids, and is upregulated by simulated shade. Both the activity of FUL and the light conditions influence the specific expression of *SAUR10* in branches, and thereby affect the

*Arabidopsis* branch angle phenotype. *SAUR10* is thus responsive to both developmental and environmental cues, and integrates both in the growth response.

# Results

## FUL binds many genes involved in hormone pathways, among which *SAUR10* and *SAUR16*

The MADS domain transcription factor FUL is well-known for its role in pistil and silique patterning (Gu *et al.*, 1998; Ferrandiz *et al.*, 2000a), and is expressed in a broad range of tissues (**Supplementary figure S1**). To identify direct targets of FUL, we initially focused on pistil and silique tissues (hereafter called silique) and performed a ChIP-seq experiment with siliques from floral stages 12-16 (Smyth *et al.*, 1990). This resulted in a list of 2538 significantly enriched binding sites (FDR<0.05, Supplementary **table S1**), corresponding to 1544 putative target genes (genes with a peak within 3kb upstream of the transcriptional start site till 1kb downstream of the stop codon). This list showed a ~58% overlap with loci identified for the putative FUL interaction partner SEPALLATA3 (SEP3) (De Folter *et al.*, 2005; Kaufmann *et al.*, 2009). The FUL ChIP list contains *SHATTERPROOF2* (*SHP2*), a previously identified target of FUL in the silique (Ferrandiz *et al.*, 2000a), and *INDEHISCENT* (*IND*) another well-described target of FUL (Liljegren *et al.*, 2004).

To unveil the processes in which FUL is predominantly acting, a gene ontology enrichment analysis was performed (**Figure 1A**). Interestingly, the gene category with the highest enrichment in the FUL target set was 'response to hormones', directly followed by 'response to abiotic processes', indicating that FUL may play an important role in the cross-talk between developmentally and environmentally-regulated processes. A closer inspection of the target list revealed in particular many auxin-response genes, in addition to GA-, cytokinin-, ABA- and ethylene-pathway genes (**Supplementary table S1**). As highly enriched genes, we identified two closely related *Small Auxin Up-Regulated (SAUR)* genes, *SAUR10* and *SAUR16*, which had FUL binding sites in their promoters, approximately 1380 bp and 1970 bp upstream of the start codon, respectively (**Figure 1B**). At the FUL binding positions, the promoter of *SAUR10* contains a canonical CArG-box, reported to be commonly bound by MADS domain proteins (Kaufmann *et al.*, 2009), while no clear CArG-box was identified in the *SAUR16* promoter (**Supplementary figure S2**). A second, smaller peak was identified in the *SAUR10* promoter about 1200 bp upstream of the large peak, suggesting that *SAUR10* could be regulated by a tetrameric FUL-containing complex that binds to two CArG boxes, as has been shown for MADS complexes involved in floral organ formation (Theissen and Saedler,

2001; Smaczniak *et al.*, 2012b). *SAUR10* and *SAUR16* belong to a clade of eight highly homologous *SAUR* genes, comprising of *SAUR8*, *SAUR9*, *SAUR10*, *SAUR12*, *SAUR16*, *SAUR50*, *SAUR51* and *SAUR54* (Kodaira *et al.*, 2011). The genes in this clade have not been functionally characterized yet, but both SAUR9 and SAUR50 have been reported to strongly inhibit PP2C-D activity (Spartz *et al.*, 2014; Atamian *et al.*, 2016), suggesting that the proteins of the SAUR10-clade inhibit PP2C-Ds to induce cell expansion via modification of $H^+$-ATPases, as has been reported for SAUR19 (Spartz *et al.*, 2014).

To confirm that FUL is able to bind to the upstream regions of *SAUR10* and *SAUR16*, we performed an electrophoretic mobility shift assay (EMSA) using the sequence below the peaks as probes (**Figure 1C**, see **Supplementary table S2** for the probe sequences). A shift was clearly detected for the *SAUR10* fragment, confirming that FUL can physically bind to this fragment. However, only a faint band indicating a shift was observed for the *SAUR16* fragment, suggesting that FUL is not able to efficiently bind this fragment as a homodimer, in line with the lack of a canonical CArG-box in this fragment. Possibly, FUL needs to interact with other transcription factors to strongly bind the *SAUR16* upstream region. To investigate whether the CArG-box in the *SAUR10* fragment is essential for FUL binding, we also generated a probe in which the CArG box was disturbed by the mutation of AT to CG in the mid region of the motif (*mSAUR10*). We did not observe a shift for this fragment (**Figure 1C**), confirming the importance of the CArG-box for the binding of FUL. Because FUL is able to bind strongly to sequences in the *SAUR10* promoter *in vivo* and *in vitro*, we further focused on *SAUR10* to unravel its function as direct target of FUL.

## SAUR10 induces growth

To get a first indication about the function of SAUR10, we generated an overexpression construct under control of the 35S promoter (35S:SAUR10) and transformed it to Col-0 *Arabidopsis*. The transgenic lines showed pleiotropic growth-related phenotypes, comprising of longer organs and tissues such as sepals, filaments, etiolated hypocotyls, cauline leaves, pistils and siliques, and a wavy stem (**Figure 1D** and **Supplementary figure S3**). These data indicate that SAUR10 can promote cell elongation similar to other SAURs (Chae *et al.*, 2012; Spartz *et al.*, 2012; Stamm and Kumar, 2013; Ren and Gray, 2015). As reported before for SAUR36 (Hou *et al.*, 2013), the leaves of the overexpression lines were also senescing earlier than the wild type leaves (**Supplementary figure**

S3). We also tested different T-DNA insertion lines for SAUR10. However, the insertion in FLAG_590D09 could not be confirmed, while the insertion in SM_3_1724 was found to be located at the 3' end of the coding sequence and didn't affect the expression of SAUR10. The phenotypes of these lines were similar to the wild type.

## FUL is a pleiotropic regulator of plant growth and architecture

To investigate the role of FUL in the regulation of *SAUR10* in the silique, we performed a quantitative RT-PCR experiment (qPCR) to compare the expression of *SAUR10* in wild type and *ful-7* (SALK_033647) mutant siliques from flower stages 11-14 (Smyth *et al.*, 1990). However, we didn't find an expression difference, suggesting that *SAUR10* is not regulated by FUL in the silique (Figure 2A).

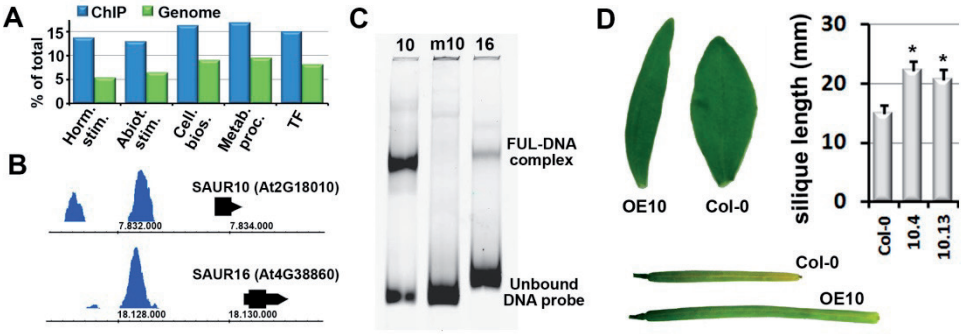



**Figure 1 | FUL binds many genes involved in hormone pathways, among which *SAUR10* and *SAUR16*.** (**A**) Graph showing gene ontology categories in which the FUL targets are significantly over-represented. The y-axis indicates the percentage of genes belonging to this ontology category. The generic GO term finder was used for the analysis (www.go.princeton.edu). The bars represent from left to right: i) response to hormones; ii) response to abiotic stimulus; iii) regulation of cellular biosynthetic process; iv) regulation of primary metabolic process; and v) regulation of transcription, DNA-templated. (**B**) FUL binding peaks in the upstream regions of *SAUR10* (top) and *SAUR16* (bottom). (**C**) Binding of the FUL homodimer to different DNA probes in an EMSA assay. Lane 1, *SAUR10* promoter fragment; Lane 2, *SAUR10* promoter fragment with mutated CArG-box; Lane 3, *SAUR16* promoter fragment (**D**) The overexpression phenotypes of the 35S:SAUR10 lines include longer siliques (picture stage 17B siliques) and differently shaped cauline leaves. See **Supplementary figure S4** for additional phenotypes. Significant differences from the wild type (t-test, $p<0.05$) are indicated with an asterisk.

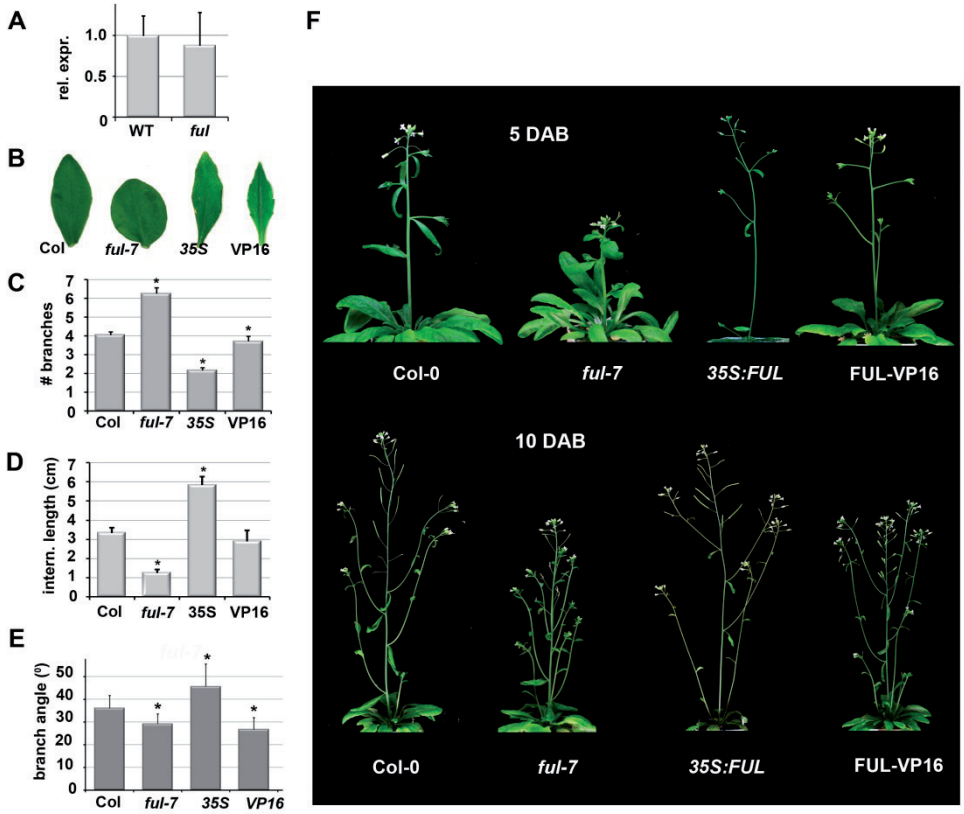



**Figure 2 | FUL is a pleiotropic regulator of plant growth and architecture.** (**A**) Expression of *SAUR10* in wild type and *ful-7* siliques from stages 12-16. (**B**) Cauline leaf phenotypes of Col-0, *ful-7*, *35S:FUL* and FUL-VP16 leaves (from left to right). (**C**) The number of branches formed along the main inflorescence. (**D**) Average internode length between the branches in plants at 10 days after bolting (DAB). (**E**) Average branch angle of all side branches along the primary stem of plants around 12 days after bolting (DAB). (**F**) Architecture phenotypes of Col-0, *ful-7*, *35S:FUL* and FUL-VP16 plants at 5 DAB (upper panel) and 10 DAB (lower panel). In A, the error bars represent the SE based on two biological replicas. In **C-E**, the error bars represent the SE based on at least 20 measurements. Significant differences from the control (t-test, $p<0.05$) are indicated with an asterisk.

Therefore, we hypothesized that there could be other tissues in which FUL is regulating *SAUR10*, possibly explaining some of the growth-related phenotypes observed in *ful* mutants (**Figure 2B-F**). In addition to the well-described phenotypes in the silique and inflorescence meristem (Gu *et al.*, 1998; Ferrandiz *et al.*, 2000a), *ful* mutants also exhibit an altered cauline leaf shape (**Figure 2B**, (Gu *et al.*, 1998), and we noticed distinct differences in stem development and architecture as well, including enhanced branching, decreased inflorescence branch angles, and shorter stem and internode lengths in *ful-7* mutants (**Figure 2C-F** and **Supplementary figure S4**). Shortly after bolting,

wild type stems elongate considerably, reaching a length of on average 12.8 cm (+/-5.4cm) at 5 days after bolting (DAB), while *ful-7* stems have hardly elongated at that stage and are only on average 5.3 cm (+/- 2.8cm, $p<0.001$) long. The difference in stem length is still visible at 10 DAB, but disappears when the plants grow older (**Figure 2F** and **Supplementary figure S4**). *35S:FUL* overexpression lines exhibit a phenotype opposite to *ful-7*, with enhanced stem elongation and increased internode size in combination with reduced branch numbers and increased branch angles (**Figure 2B-F**).

We compared the phenotypes of *ful-7* mutants and wild type plants also to a *pFUL:FUL-VP16* line (FUL-VP16). This line consists of a translational fusion of FUL with the strong transcriptional activation domain of the herpes virus protein VP16, driven by the *FUL* promoter. Genes that are in wild type plants repressed by *FUL* are expected to become activated in this line. *ful* mutant phenotypes that are caused by target gene de-repression should thus be similar in the FUL-VP16 line, albeit probably to a lesser extent, because FUL-VP16 has been generated in the Col-0 background and still contains an endogenous FUL copy that can repress the targets. The FUL-VP16 plants showed aberrant phenotypes that were probably a mix of enhanced target gene activation (for those targets that are in wild type tissues activated by FUL) and activation of targets that are normally repressed by FUL. For example, FUL-VP16 siliques exhibited 'shoulders' and a short style similar to *35S:FUL* siliques, while their overall phenotype resembled more *ful-7* siliques (**Supplementary figure S5A**). To verify that FUL-VP16 is not co-suppressing the endogenous *FUL*, we tested *FUL* transcript abundance in stem and branches and found a 1.5-2 fold higher expression of *FUL* (**Figure S5B**), indicative of the presence of an additional *FUL* copy (FUL-VP16) and not of co-suppression. Several FUL-VP16 phenotypes, including stem and cauline leaf growth, were more similar to *35S:FUL* than to *ful-7*, suggesting that these traits are largely regulated by activation of target genes in the wild type (**Figure 2B** and **2F**).

## FUL represses *SAUR10* in the stem

To investigate whether FUL can regulate *SAUR10* in other tissues than the silique, we generated reporter lines for *SAUR10* (*pSAUR10:GUS*) and crossed these into the *ful-7*, *35S:FUL* and *FUL-VP16* backgrounds. The *GUS* expression patterns of the *pSAUR10:GUS* lines was rather specific, with staining predominantly in the veins and petioles of rosette leaves and cauline leaves. In the context

of the flower, expression appeared only in stage 12 flowers in the vasculature of the style, and in stage 13 flowers in the apical parts of the stamen filaments and petals. Apart from the expression in the style, no expression was observed in the pistil or silique (**Figure 3A**).

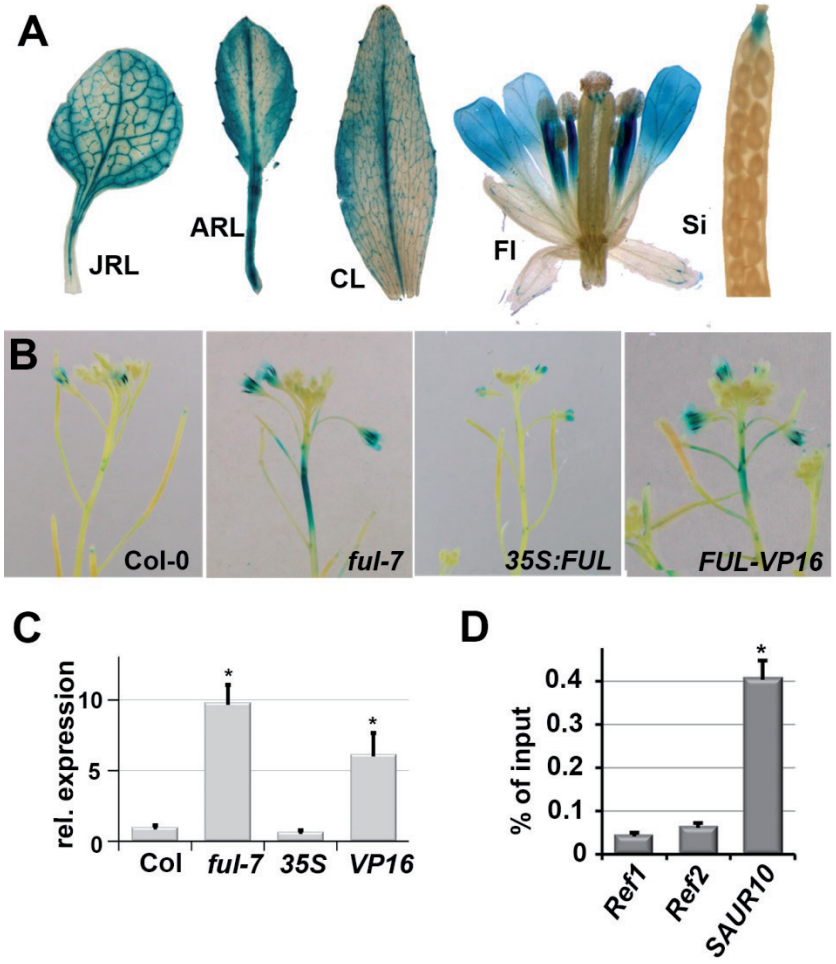



**Figure 3 | *SAUR10* is regulated by FUL in stems.** (**A**) GUS staining pattern in different organs and tissues of a pSAUR10:GUS line. JRL, juvenile rosette leaf; ARL, adult rosette leaf; CL, fully expanded cauline leaf; Fl, flower stage 13; Si, silique stage 17. (**B**) GUS staining pattern of pSAUR10:GUS inflorescence stems from different FUL backgrounds. (**C**) Relative expression of SAUR10 in the upper stem segment of Col-0, *ful-7*, 35S:FUL and FUL-VP16 plants 8-10 DAB. (**D**) Graph from a ChIP-qPCR experiment showing the enrichment of the *SAUR10* fragment relative to two reference sequences in a ChIP sample from stem tissue. The enrichment of the fragments was calculated as a percentage of the input sample. Error bars represent the SE of three biological replicas in the case of expression analyses, and two replicas for the ChIP-PCR. Significant differences from the control (t-test, $p<0.05$) are indicated with an asterisk.

*SAUR10* expression in *ful-7* and FUL-VP16 stems was clearly different from expression in wild type plants. In wild type inflorescences, *pSAUR10:GUS* expression was not, or only very weakly observed in the stem. However, in both *ful-7* and FUL-VP16 inflorescences, GUS expression was present in the upper stem region, indicating that *SAUR10* is de-repressed in *ful-7* and activated by FUL-VP16 in inflorescence stems (**Figure 3B**). To investigate whether *SAUR10* transcript levels were in line with these results, RNA was extracted from a 1 cm stem segment directly below the inflorescence meristem, and a qPCR was performed. *SAUR10* expression was found to be approximately 10-fold higher in *ful-7* and 6-fold higher in FUL-VP16, while the levels in *35S:FUL* were non-significantly decreased, probably because *SAUR10* expression is already very low in wild type stems (**Figure 3C**). These data are consistent with direct repression of *SAUR10* by FUL, and are in line with our observation that *pFUL:GUS* is active in the inflorescence stem (**Supplementary figure S1**).

Since our ChIP-seq experiment had been performed with silique tissue, we were not certain if FUL could directly bind to *SAUR10* in the stem. To test this, we performed ChIP-qPCR experiments using stem tissue of primary and secondary inflorescences. We found a distinct enrichment for *SAUR10* in both replicates (**Figure 3D**), showing that FUL can directly bind to the *SAUR10* locus in stems as well. In conclusion, we show here that FUL can directly repress *SAUR10* expression in the stem.

## Auxin and BR induce *SAUR10* expression synergistically

Auxin application experiments have identified *SAUR10* as one of the *Arabidopsis SAUR* genes clearly upregulated in response to auxin treatment in seedlings (Goda *et al.*, 2004; Paponov *et al.*, 2008; Chapman *et al.*, 2012; Bargmann *et al.*, 2013), and *SAUR10* has also been reported to be responsive to brassinosteroids (Yu *et al.*, 2011). To investigate the interaction between FUL-controlled repression and hormone-induced upregulation, we treated wild type, *ful-7*, *35S:FUL* and FUL-VP16 seedlings with auxin (5 µM IAA) and/or brassinosteroids (5 µM brassinolide (BL)) for 4 hrs. Wild type seedlings treated with BL showed a 4-fold induction of *SAUR10* compared to mock-treated seedlings, while the seedlings treated with auxin showed a 10-fold increase, confirming the previously published hormone responses (**Figure 4A**). Interestingly, a combination of IAA and BL resulted in a synergistic effect on the induction of transcription and led to an impressive 65-fold higher expression of *SAUR10* in seedlings. In all treatments, the expression of *FUL* did not change,

while the *GUS* transcript levels in treated *pSAUR10:GUS* plants showed a response similar to *SAUR10*, indicating that the hormone induction is regulated by the promoter rather than through post-transcriptional mechanisms. To test whether FUL could influence the hormone-induced increase in *SAUR10* expression, we performed the IAA-BL treatments also in *ful-7*, *35S:FUL* and FUL-VP16 seedlings. This revealed no significant differences (**Figure 4B**), suggesting that FUL does not influence the hormonal upregulation of *SAUR10* in seedlings.

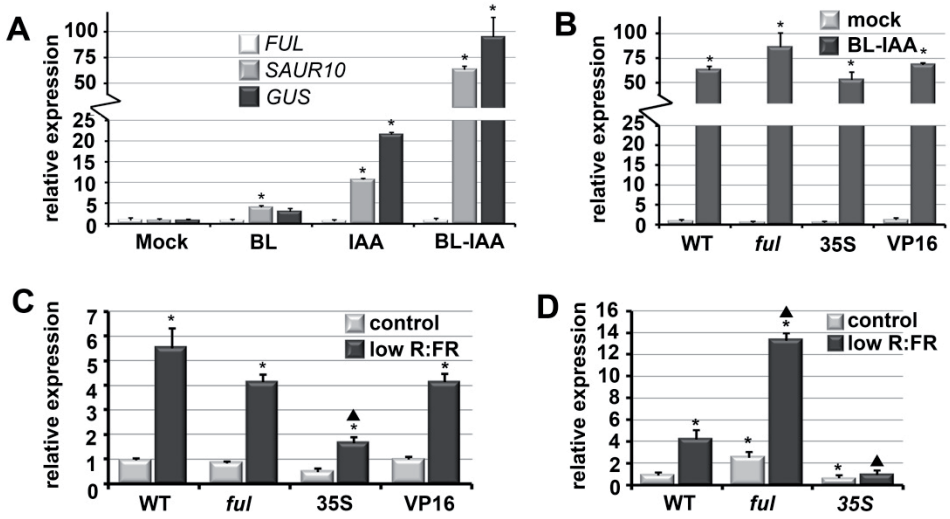



**Figure 4 | *SAUR10* is induced by a combination of BR and auxin, and by reduced R:FR ratios, which can be repressed by FUL in the stem.** (**A**) Expression of *SAUR10*, *GUS* (in *pSAUR10:GUS* lines) and *FUL* in 10 day-old seedlings after a 4 hour treatment with auxin (IAA), brassinosteroids (BL) or a combination of both (BL-IAA). (**B**) Expression of *SAUR10* in 10 day-old Col-0, *ful-7*, *35S:FUL* and FUL-VP16 seedlings after a 4 hour treatment with a combination of auxin and BR (BL-IAA). (**C**) Expression of *SAUR10* in Col-0, *ful-7*, *35S:FUL* and FUL-VP16 12 day-old seedlings grown for 4h under reduced R/FR light conditions or under control conditions. (**D**) Expression of *SAUR10* in Col-0, *ful-7*, *35S:FUL* and FUL-VP16 stems 7 DAB, grown for 4h under reduced R/FR light conditions or under control conditions. The error bars represent the SE of three biological replicas. In (**A**) and (**B**), significant differences from the mock control (t-test, $p<0.05$) are indicated with an asterisk. In (**C**) and (**D**), significant differences from the WT control situation (t-test, $p<0.05$) are indicated with an asterisk, while significant differences from the WT low R:FR situation are indicated with a triangle.

## The response of *SAUR10* to shade is influenced by FUL

*SAUR10* can be highly induced by a combination of auxin and brassinosteroids, two hormones that have together been associated with shade responses (Pierik *et al.*, 2009; Keuskamp *et al.*, 2011). When exposed to shade, plants sense a decreased R:FR ratio, as well as depletion of blue light, and

respond by phenotypic changes such as stem, internode and petiole elongation, and hyponastic leaf movement, together referred to as the shade-avoidance syndrome (SAS). To investigate whether *SAUR10* is responsive to simulated shade, we transferred 12 day old seedlings to low R:FR light conditions and compared the expression of the genes with seedlings from the control condition. *SAUR10* showed a marked increase of expression after 4 hours of low R:FR, indicating that it can indeed positively respond to shade. This response was not different in *ful* or FUL-VP16 seedlings, but was significantly reduced in *35S:FUL* seedlings, suggesting that the ectopic/over-expression of FUL represses the shade-induced upregulation of *SAUR10* (**Figure 4C**).

To investigate whether the effect of FUL was more distinct in the tissue where it is actually repressing *SAUR10*, we transferred wild type, *ful* and *35S:FUL* plants to low R:FR conditions and harvested stem segments after 4 hrs. The expression in wild type stems was four times up-regulated compared to control stems, while the up-regulation in *ful* mutant stems increased to 14 times (**Figure 4D**). The up-regulation in *ful* stems is higher than can be explained by an additive effect of de-repression and shade-induced upregulation, suggesting that loss-of-FUL allows a greater response to the light conditions. No significant up-regulation compared to control conditions could be detected at all in *35S:FUL* stems (**Figure 4D**), pointing to a much stronger effect of FUL repression in the regulation of *SAUR10* expression in the stem than in seedlings. In conclusion, these experiments show that *SAUR10* is a distinct responder to both hormone and light stimuli, and that the light response can be attenuated by FUL in the stem.

## De-repressed *SAUR10* expression correlates with longer cells in the stem

We inspected the stem and architecture phenotype of the *ful* mutants further to identify phenotypic features that could be attributed to *SAUR10* de-repression. Given the longer-organ phenotype of the *SAUR10* overexpression lines, we expected the inflorescence stem of *ful-7* plants, in which *SAUR10* is higher expressed, to be longer than the wild type, and that of *35S:FUL* to be shorter. However, we found an opposite effect in young inflorescences, which were shorter in *ful-7* plants, with a significantly smaller distance between side branches (**Figure 2D**) and silique internodes (**Supplementary figure S4**). To determine whether the cell sizes in the stem were in line with the internode sizes, we measured cell length in wild type, *ful-7* and *35S:SAUR10* stems between internodes one and two. Interestingly, this revealed longer cells in the *ful-7* and

*35S:SAUR10* stems compared to the wild type (**Figure 5A-B**), showing that *ful-7* stems have longer cells despite having shorter internodes. Thus, the upregulation of *SAUR10* in the *ful-7* stem probably does result in longer cells, but the shorter stem phenotype is caused by reduced cell division as a result of de-regulation of other target genes.

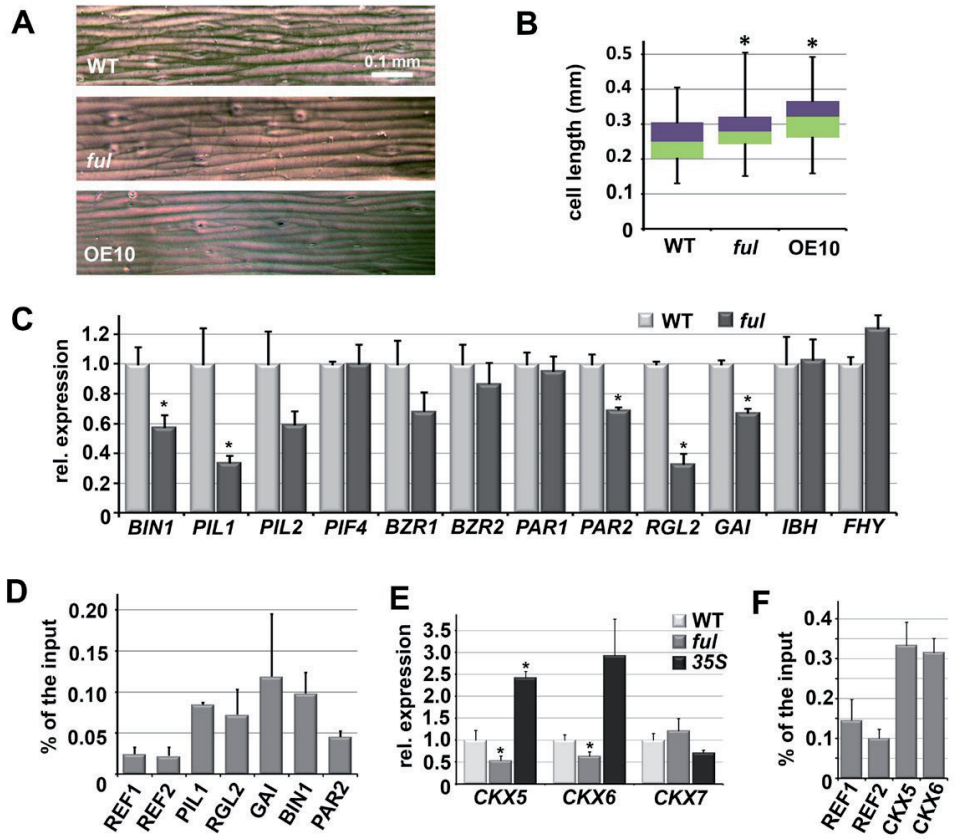



**Figure 5 | The *ful* mutant stem phenotype is due to a combination of de-regulated genes.** (**A**) Prints of stem segments between internode 1 and 2 from WT, *ful-7* and *35S:SAUR10* stems. The scale bar represents 0.1 mm. (**B**) Boxplot showing the cell lengths of the stem segments such as in panel (**A**). The distribution of the lengths is depicted as follows: purple, upper quartile; green, lower quartile; upper error bar, maximum; lower error bar, minimum. Cell lengths were measured with ImageJ and based on at least 40 cells from three different stem segments. (**C**) Relative expression of a number of genes from the ChIP-seq target list in wild type and *ful-7* stems. (**D**) Graph from a ChIP-qPCR experiment showing the enrichment of the *PIL1, RGL2, RGA2* and *BIN1* fragments relative to two reference sequences in a ChIP sample from stem tissue. The enrichment of the fragments was calculated as a percentage from the input sample. (**E**) Relative expression of *CKX5, CKX6* and *CKX7* in the upper stem segment of wild type, *ful-7* and *35S:FUL* inflorescences.

**Figure 5 | continued.** The significance of the differential expression in the *ful-7* samples compared to the wild type was calculated based on five biological replicates. (**F**) Graph from a ChIP-qPCR experiment showing the enrichment of the *CKX5 and CKX6* fragments relative to two reference sequences in a ChIP sample from stem tissue. The enrichment of the fragments was calculated as a percentage of the input sample. Error bars represent the SE of three biological replicas in the case of expression analyses. Significant differences from the control (t-test, $p<0.05$) are indicated with an asterisk. Two replicas were performed for the ChIP-PCR.

## The *ful* mutant stem phenotype is caused by a combination of de-regulated genes

FUL represses *SAUR10* in the stem, but the stem phenotype of the *ful-7* line, which shows retarded cell division, indicates that FUL additionally activates other targets to regulate stem growth and architecture. We therefore examined the FUL ChIP target list in more detail for genes that have been associated with growth responses. In addition to genes involved in cytokinin and auxin signalling, such as the cytokinin degradases *CKX5*, *CKX6* and *CKX7*, and genes encoding the AUX/IAA proteins IAA8 and IAA16, we also found a remarkable number of genes that are implicated in the light-sensitive growth of hypocotyls, encoding transcription factors involved in the BZR-PIF-ARF-DELLA pathway (Bai *et al.*, 2012; Oh *et al.*, 2014). These include the PIF genes *PIL1*, *PIL2* and *PIL4*, the DELLA genes *RGL2* and *GAI*, the BR-pathway genes *BZR1*, *BZR2* and *BRASSINOSTEROID INSENSITIVE 1* (*BIN1)* and also the photoreceptor Phytochrome A (FHY), and its targets *PHYTOCHROME RAPIDLY REGULATED 1* (*PAR1)* and *PAR2*, which are negative regulators of the shade response and reduce the expression of several *SAURs* (Roig-Villanova *et al.*, 2007).

To determine if any of these genes was regulated by FUL in the stem, we performed qRT-PCR analysis to compare the transcription levels of *ful-7* mutant stems with the wild type. We detected significantly lower transcript levels for *BIN1*, *PIL1*, *RGL2*, *GAI* and *PAR2*, indicating that FUL activates these genes in wild type stems (**Figure 5C**). We selected these targets to test whether they were also bound by FUL in the stem, and found enrichment for all five tested genes in two independent ChIP experiments (**Figure 5D**). However, the decreased expression of these genes will rather have an effect on cell elongation than on cell division (Nam and Li, 2002; Salter *et al.*, 2003; Roig-Villanova *et al.*, 2007; Li *et al.*, 2012), and can thus not entirely explain the *ful* stem phenotype. *CKX5*, *CKX6* and *CKX7,* on the other hand, are cytokinin oxidase/dehydrogenase (CKX) genes, which can catalyze the degradation of cytokinin, and have been reported to determine the activity of the shoot meristem. We found that the transcript levels of *CKX5* and *CKX6* are significantly

decreased in *ful-7* stems, while being upregulated in *35S:FUL* stems (Figure 5E). This would lead to reduced cytokinin breakdown in *ful* mutants and a higher meristem activity (Werner *et al.*, 2003; Bartrina *et al.*, 2011), which could explain the increased branching and shorter internodes as observed. We also tested whether *CKX5* and *CKX6* were bound by FUL in the stem, and detected a clear enrichment for the *CKX5* and *CKX6* loci compared to two reference loci (Figure 6D). FUL thus appears to regulate a complex network of genes that are likely to have opposite functions in stem growth. The outcome of this regulation probably depends on other factors that interfere with this network, such as hormone concentration and light quality.

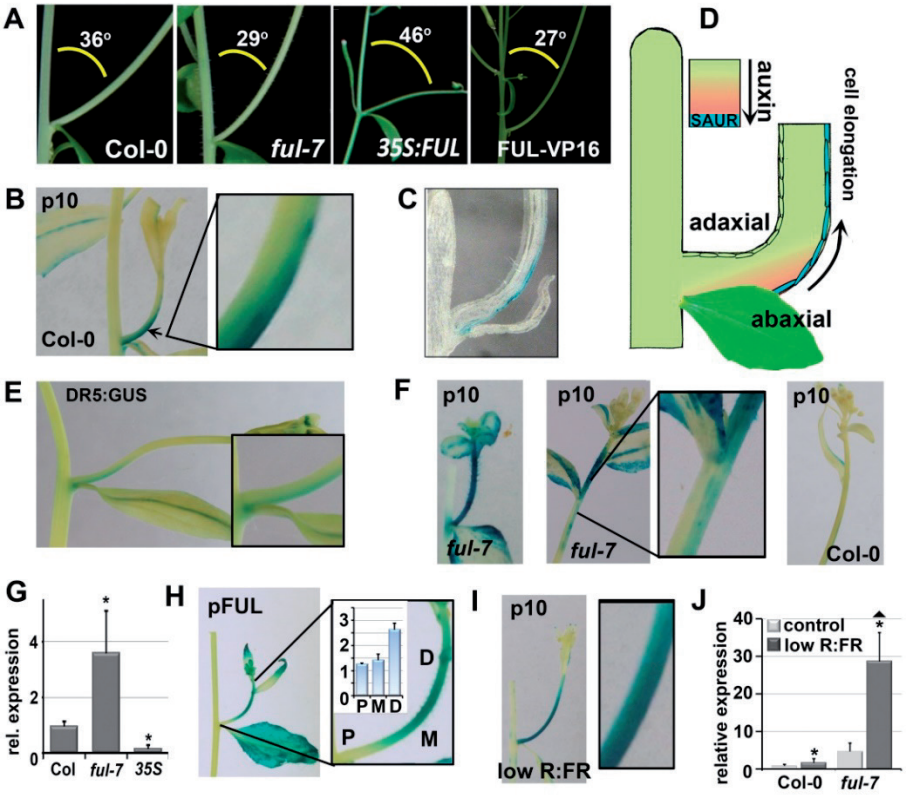



**Figure 6 | The branch angle phenotype of the *ful* mutant correlates with specific abaxial expression of *SAUR10* in branches.** (**A**) Pictures showing the relation between branch angle and plant architecture in the different FUL backgrounds. The smaller the angle, the more vertical the branches grow. The average branch angle in the different backgrounds is indicated. (**B**) Localization of *pSAUR10:GUS* at the abaxial side of a young branch. The arrow points to the region that was enlarged. (**C**) Cross section of a *pSAUR10:GUS* branch. (**D**) Model showing the correlation between auxin-induced accumulation of growth factors (SAURs) at the abaxial side and branch bending. (**E**) Localization of DR5:GUS at the abaxial side of the branch.

**Figure 6 | continued.** (**F**) Expression and localization of *pSAUR10:GUS* in the branches of Col-0 and *ful-7*. Left panel: *pSAUR10:GUS* can be observed on both sides of emerging *ful-7* branches. The middle and right panels show older, elongating branches, where no signal is visible any more at the proximal part near the primary stem, but a difference in the distal part can be observed. *pSAUR10:GUS* is clearly visible in the distal region of elongating *ful-7* branches, and is specifically abaxially expressed directly below that region (middle panel), whereas *pSAUR10:GUS* is not visible in the distal region of elongating (older) Col-0 branches (right panel). ; Right panel: (**G**) Relative transcript levels of *SAUR10* in young branches from different FUL backgrounds. (**H**) *FUL* is highly expressed throughout branches, with the highest expression in the distal region. P = proximal region, M = middle region D = distal region. (**I**) Localization of *pSAUR10:GUS* on both sides of branches grown under reduced R:FR light conditions. (**J**) Relative transcript levels of *SAUR10* in Col-0 and *ful-7* branches under control conditions and reduced R:FR light conditions. p10, pictures from *pSAUR10:GUS* stained tissue; pFUL, pictures from pFUL:GUS stained tissue. The error bars depict the SE based on three biological replicas. Significant differences (t-test, $p<0.05$) are indicated with an asterisk.

## De-repressed *SAUR10* expression correlates with decreased branch angles

We also noticed that the branch angles in *ful-7* and FUL-VP16 plants are significantly smaller than in the wild type, while being larger in *35S:FUL* (**Figure 2E** and **6A**). Branch angle is highly dynamic and depends on gravitropic and phototropic signals, both of which depend on auxin gradients (Roychoudhry *et al.*, 2013; Liscum *et al.*, 2014). In addition, brassinosteroids have also been found to influence auxin-mediated phototropic responses (Liscum *et al.*, 2014). We therefore investigated whether the altered branch angles in *ful* mutants could be correlated to *SAUR10* expression. *pSAUR10:GUS* activity was specifically observed at the abaxial side of young Col-0 branches, and sectioning revealed that the expression was located in the abaxial epidermal layer (**Figure 6B-C**), pointing to a role for *SAUR10* in branch bending. Since directed auxin-induced hypocotyl/stem elongation occurs in accordance with the SAUR-mediated acid-growth theory and growth is predominantly regulated by the epidermis (Fendrych *et al.*, 2016; Procko *et al.*, 2016), we reasoned that *SAUR* expression in the abaxial epidermal cell layer is expected to enhance cell elongation on this side, resulting in more vertical growth of the branch (**Figure 6D**). To determine if auxin concentrations are higher at the abaxial side of the branch, we examined the DR5:GUS auxin reporter line (Ulmasov *et al.*, 1997). This revealed a weak GUS signal in DR5:GUS branches specifically at the abaxial side after prolonged staining (**Figure 6E**), indicating that auxin levels are indeed higher at the abaxial side of the branch, probably inducing *SAUR10* expression.

Similar to the de-repression of *SAUR10* in the primary inflorescences of *ful-7* and *FUL-VP16* plants, *SAUR10* was also clearly upregulated in the stem of the branch inflorescences. We examined the

*pSAUR10:GUS* pattern in detail for the *ful-7* line, and found that in young, emerging branches, the GUS signal was higher than in the wild type and could also be observed at the adaxial side (**Figure 6F**). This upregulated *pSAUR10:GUS* signal remained distally visible when the branch elongated, and was accompanied by a region with distinct abaxial GUS expression just below the distal region. In contrast, *pSAUR10:GUS* was rarely observed in wild type branches at this stage (**Figure 6F**). In accordance with FUL repressing *SAUR10* in branches, *pSAUR10:GUS* signal was mostly absent in *35S:FUL* branches (**Supplementary figure S6A**). These *pSAUR10:GUS* data were confirmed by qRT-PCR analysis of *SAUR10* transcript levels in branches 4-6 days after bolting (DAB), which revealed a 3-4 fold higher expression in *ful-7*, and a 5-fold reduction in *35S:FUL* (**Figure 6G**). The higher abaxial expression in the region below the *ful-7* (**Figure 6F**) and FUL-VP16 (**Figure S6A**) inflorescences can well explain the more vertical branching in these lines. The expression of *FUL* itself also corresponds with the observed *SAUR10* expression, as *FUL* is highly expressed throughout branches, and exhibits the highest levels just below the inflorescences (distal part, **Figure 6H**). This is in line with the observation that *SAUR10* de-repression is most prominent in this region.

## Abaxial *SAUR10* expression is affected by the light conditions

To investigate whether the expression of *pSAUR10:GUS* in branches was sensitive to the light conditions, we reduced the R:FR ratio and determined the GUS pattern in the inflorescences after 24h. We observed a relocation of the GUS signal to both sides of the branch (**Figure 6I**), consistent with a positive response of *SAUR10* to the shaded (low R:FR) conditions at both the adaxial and abaxial sides and thus an increase of the Ad/Ab ratio. In line with this result, plants that were for a longer period exposed to reduced R:FR conditions exhibited substantially increased branch angles (**Supplementary figure S7A-B**), suggesting a more homogenous growth factor distribution. If *SAUR10* is indeed regulating branch bending, branch angle should also be disturbed in *35S:SAUR10* plants, where *SAUR10* is ectopically expressed at the adaxial side. The expected increased Ad/Ab ratio would then result in more horizontal branch growth. Indeed, we observed larger branch angles in the *35S:SAUR10* line, although branch growth was highly variable and irregular (**Supplementary figure S7A-B**). This irregular growth was even stronger under reduced R:FR conditions (**Supplementary figure S7B**), suggesting that the *SAUR10* overexpression phenotype is enhanced by the SAS response.

We also tested to what extent FUL could influence the effect of simulated shade on *SAUR10* expression levels and branch angle phenotype. In branches of the *ful* mutant, de-repression of *SAUR10* combined with shade-induced expression resulted in an almost 30-fold upregulation of *SAUR10* in young branches (**Figure 6J**). This is again more than can be explained by an additive effect alone, suggesting that the absence of FUL allows an enhanced response to the light conditions, causing an increase in expression on both the abaxial and adaxial sides. In line with this, simulated shade resulted in a more horizontal branch growth in all backgrounds (**Supplementary figure S6B**).

In conclusion, we found *SAUR10* to be abaxially expressed in branches, presumably as a result of auxin accumulation at the shaded side of the branch. The de-repression of *SAUR10* in the *ful* mutant results in increased abaxial expression in the distal part of the branch, which can cause the increased bending of the branches in *ful* mutants. In wild type branches, FUL represses *SAUR10*, thereby possibly preventing over-bending of the branch and attenuating responses to the light conditions.

# Discussion

We show here that the MADS domain transcription factor FUL is a pleiotropic regulator of plant development, which plays important roles in plant growth and architecture in addition to its well-known functions in fruit development and flowering time. In particular the deviating branch angles in the *ful* mutants are very interesting, since little is known about this trait, which is in particular important for crop yield. Loci in other species that could be linked to branch angle have been associated with the auxin pathway, such as the LA1 locus in rice (Li *et al.*, 2007) and a *GRETCHEN HAGEN 3* (*GH3*) gene in *Brassica napus* (Liu *et al.*, 2009), and our analysis indicates that also members of the *SAUR* family can play an important role in branch bending, especially in response to environmental conditions like high plant density. We demonstrate that *SAUR10* is specifically expressed at the abaxial side of the branch, thereby affecting branch angle. Enhanced and prolonged abaxial expression of *SAUR10* in the *ful* mutant can explain its more vertical branching phenotype. The activity of *SAUR10* in stems and branches appears to be regulated by an interplay between hormone-induced upregulation and FUL-controlled repression. In addition, our data reveal that FUL directly regulates a number of other genes involved in hormone and light signalling, of which the de-regulation contributes to the *ful* mutant phenotype. To what extent these genes are contributing to the *ful* mutant phenotype needs to be further investigated. However, the picture emerges that FUL can regulate plant growth and architecture in concert with the environment by balancing the expression of hormone and light responsive factors. It will be interesting to study how the expression of FUL changes during plant development, and if for example older plants are less responsive to environmental signals through increased FUL expression.

## FUL interacts with the IAA/BR pathway to repress *SAUR10*

Despite the broad expression pattern of FUL, de-repression of *SAUR10* in the *ful* mutant only occurs in a limited number of tissues, indicating that *SAUR10* activation requires additional tissue-specific factors, such as high auxin and brassinosteroid levels. In addition to repressing *SAUR10* under control conditions, we also found that FUL can buffer the hormone- or shade-induced expression of *SAUR10* in stems and branches, suggesting that FUL can attenuate the activity of the auxin and/or brassinosteroid response transcription factors. Since *SAUR10* has been identified as a

direct target of both ARF6 and BZR1 by ChIP-seq analyses (Oh *et al.*, 2014), it is possible that FUL can interact with either or both of these factors, thereby repressing transcription. FUL binds to a CArG-box in the *SAUR10* promoter, which is located 260 bp upstream of a canonical ARF binding motif and only 100 bp downstream of an AuxRE-related element identified by Walcher and Nemhauser (2012). Binding of FUL to the CArG-box may disturb the interaction between ARF6 and BZR1 (previously reported by Oh *et al.* (2014)). Our results indicate that FUL is not required to determine the *SAUR10* expression domain, but rather to fine-tune or buffer the response to hormonal stimuli.

## SAURs integrate environmental, hormonal and developmental signals in the growth response

Different studies in *Arabidopsis*, soybean and maize have identified *SAUR* genes as hormone-responsive growth regulators. However, *SAUR*s have also been found as direct targets of several transcription factors functioning in plant development, suggesting that they function downstream of both developmental and hormonal regulators to direct plant growth. We demonstrate here that *SAUR10* is regulated by hormonal stimuli, light signals, as well as by the developmental regulator FUL, and can thereby integrate a plethora of signals in the growth response. Several recent reports have strengthened the idea that *SAUR*s can in general respond to a variety of upstream factors, integrating these in the regulation of cell elongation through interaction with PP2C-Ds (Spartz *et al.*, 2014; Challa *et al.*, 2016; Procko *et al.*, 2016; Sun *et al.*, 2016). This has been most thoroughly investigated in seedlings, where *SAUR* genes have been grouped according to their response to light conditions in hypocotyls and cotyledons (Sun *et al.*, 2016), indicating that the growth response is controlled by a cluster of similarly regulated *SAUR*s, rather than by single genes. In that respect, *SAUR10* may not be the only *SAUR* with differential expression between the abaxial and adaxial side of the *Arabidopsis* branches. Interestingly, a *SAUR50*-like gene (*SAUR50* belongs to the *SAUR10*-clade) has recently been identified to be responsible for heliotropism in sunflower (Atamian *et al.*, 2016). The gene is higher expressed on the east side of the stem during the day, enabling the shoot apex to move gradually from east to west along with the sun. In addition, several other *SAUR*s have been reported to be responsive to shade (Roig-Villanova *et al.*, 2007; Spartz *et al.*, 2012; Procko *et al.*, 2016), indicating that the dynamic response to shade may to a

large extent be executed by SAUR proteins. Differential expression of *SAUR* genes may thus in general allow directional growth in a variety of species.

# Materials and Methods

## Plant materials and growth conditions

Most plants used in this study were in the Col-0 background, including the overexpression and reporter lines, and the *ful-7* (SALK_033647) mutant. For the ChIP analysis, FUL-GFP lines were used from a mixed Ler (*ful-1*)/Col-0 background (Urbanus *et al.*, 2009). The JIC SM T-DNA insertion line SM_3_1724 was received from NASC (Tissier *et al.*, 1999), the FLAG T-DNA line FLAG_590D09 (Samson *et al.*, 2002) was received from the IJPB in Versailles. Plants were grown on rockwool blocks watered with HYPONeX® solution (1.5 g/l), in a long-day climate chamber (16/8) at 22 °C. The climate chamber was equipped with LED lights, resulting in the following control conditions: 87.6 µmol $m^{-2}$ $s^{-1}$ photosynthetically active radiation (PAR); R:FR ratio = 30.1). Reduced R:FR conditions were achieved by supplemental FR (730 nm) irradiation, resulting in a PAR of 83,5 µmol $m^{-2}$ $s^{-1}$, and a R:FR ratio of 1.15.

## ChIP-seq analysis

For ChIP-seq analysis, pistils/siliques in stages 12-16 were harvested from gFUL-GFP lines (Urbanus *et al.*, 2009), and the ChIP-seq and subsequent data analysis were performed according to Kaufmann *et al.* (2010b). Input samples were used as controls. The data analyses were largely performed as described in Van Mourik *et al.* (2015). Sequences from each ChIP library were mapped to the unmasked *A. thaliana* genome (TAIR9) using SOAPv2 (Li *et al.*, 2009). A maximum of two mismatches and no gaps were allowed. Only uniquely mapped reads were retained. Sequence reads mapping to the plastid and mitochondrial genomes were eliminated. The R package CSAR was used for peak calling (Muiño *et al.*, 2011).

## EMSA

*SAUR10* and *SAUR16* CDSs were amplified from wild type Col-0 cDNA and cloned into pSPUTK (**see Supplementary table S2** for all primer sequences). The pSPUTK promoter allowed *in vitro* protein synthesis using the TnT® SP6 High-Yield Wheat Germ Protein Expression System (Promega) according to manufactures instructions. For *SAUR10*, the probe fragment consisted of a region of 100 bp with the canonical CArG-box in the centre. For *SAUR16*, the fragment consisted of a region

of 128 bp below the peak summit (see **Supplementary figure S2** and **table S2** for the primer sequences). Promoter fragments were amplified from genomic DNA; the complete FUL coding sequence was amplified from cDNA. The mutated *SAUR10* fragment was generated by overlapping PCR using primers that replaced the canonical CCAAATATGG CArG-box by CCAACGATGG. EMSAs were performed essentially as described by Smaczniak *et al.* (2012b) with minor modifications. Oligonucleotides were fluorescently labelled using DY-682. Labelling was performed by PCR using vector-specific DY-682-labelled primers followed by agarose gel extraction. Gel-shifts were visualized using a LiCor Odyssey imaging system at 700 nm.

## Generation of transgenic lines

The Gateway technology (Invitrogen) was used for generation of the overexpression and reporter constructs, using the entry vector pDONR221 and the destination vectors pK2GW7 (*35S:SAUR10* overexpression) and pBGWFS7 (*pSAUR10:GUS* and *pFUL:GUS*) (Karimi *et al.*, 2002) (See **Supplementary table S2** for all primers). The *FUL-VP16* line was generated by cloning a genomic fragment of the FUL locus (including 3.9 kB upstream region), which was fused in frame with the coding sequence of the strong activation domain of VP16, and followed by the CaMV 35S terminator, into the pBIN19 vector. Constructs were checked by sequencing, transformed into Agrobacterium strains LBA4404 or EHA105 and transformed to Col-0 plants using floral dip (Clough and Bent, 1998).

## Expression analysis

RNA was extracted using the InviTrap® Spin Plant RNA Mini kit (Stratec Molecular), or with a CTAB/LiCl protocol. The RNA concentrations were adjusted to 200 ng/µl, and a DNase treatment was performed using Ambion Turbo DNase (AM1907). For qRT-PCR analysis, the RNA was reverse transcribed using the iScript cDNA synthesis kit (BioRad), and the qRT-PCR reaction was performed with iQ SybrGreen supermix from BioRad. The quantitative RT-PCR analyses were performed on the BioRad iCycler. The UBC21 and/or TIP41 genes were used as reference genes (Czechowski *et al.*, 2005).

## Hormone treatments and shade experiments

For the hormone treatments, 10 day old seedlings were removed from plates (2.2 g/l Murashige and Skoog medium (MS), 10% sucrose, 0.8% agar) and incubated in liquid 2.2 g/l MS medium with or without hormones on the shaker for 4 hours (the seedlings were floating with their roots submerged in the medium and their leaves contacting the liquid medium). The following hormone concentrations were used: 5 µM IAA (according to Bargmann *et al.*, 2013), 5 µM Brassinolide, 5 µM IAA + 5 µM Brassinolide. After incubation, seedlings were frozen in liquid nitrogen and stored at −80°C prior to RNA isolation. To investigate the response to simulated shade, seedlings were grown on plates (2.2 g/l MS, 10% sucrose, 0.9% agar) under control conditions for 12 days, and then transferred to reduced R:FR for 4 hrs (see above) or kept for another 4 hrs under control conditions. To determine the branching phenotypes under simulated shade conditions, plants were placed under reduced R:FR conditions upon bolting and grown for another 2-3 weeks.

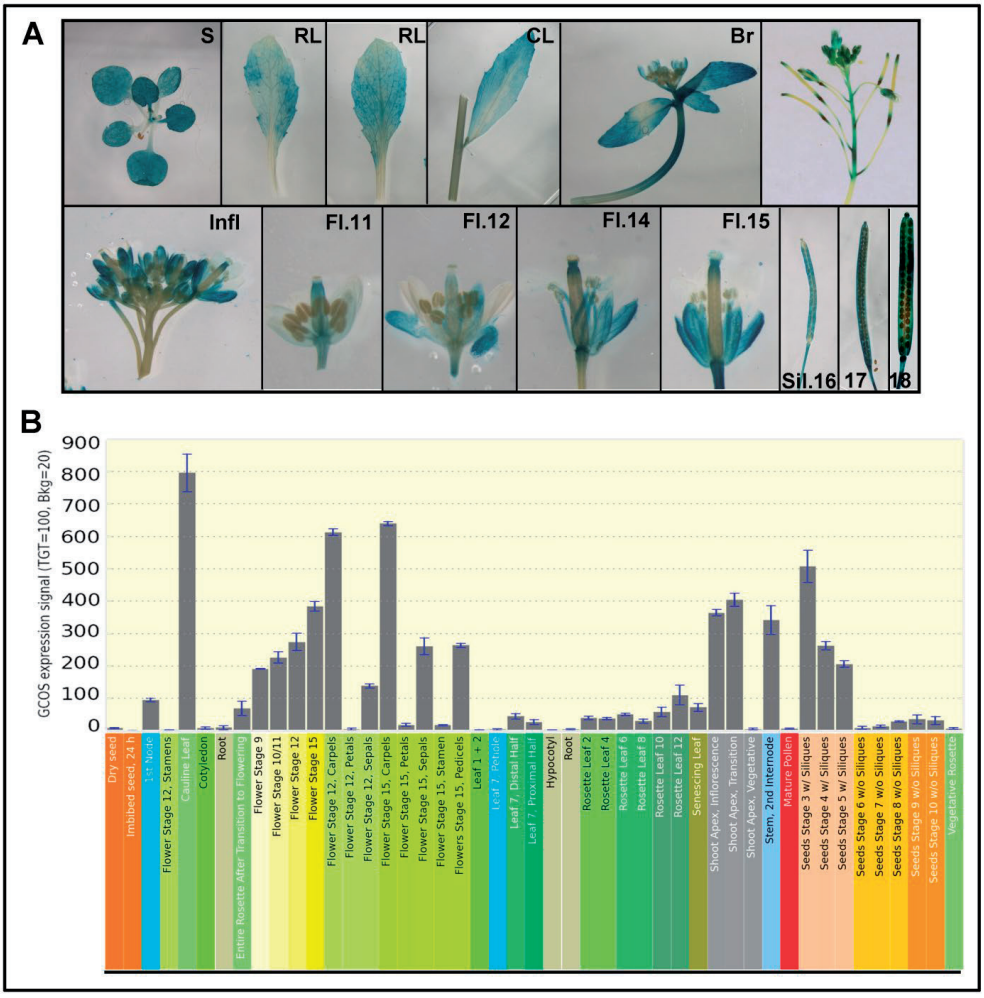
## Histochemical analyses

GUS activity was analysed by staining various tissues overnight at 37 °C in staining solution (10 mM EDTA, 0.2% Triton X-100, 1 mM $Fe^{2+}CN$, 1 mM $Fe^{3+}CN$ and 1 mg $mL^{-1}$ 5-bromo-4-chloro-3-indolyl-β-glucuronic acid in 50 mM phosphate buffer, pH 7.2). DR5:GUS tissue was stained for two days, refreshing the staining buffer after day one. The tissue was destained for several days in 70% EtOH, and mounted in 30% glycerol for microscopical analysis or photographed while being submerged in 70% EtOH. For DIC microscopy of epidermal cells from the stem, the two components of President coltene(R) dental paste (REF4667), catalyst and base, were mixed together on a glass slide. Before this mixture became solid, stem segments were softly pressed into the paste on the slide. After solidification, the stem segments were removed leaving a print of the outer cell layer. This print was subsequently covered with transparent nail polish, which was removed after hardening. The nail-polish layer that contained the cell shapes was then mounted in 100% glycerol and observed under the microscope using DIC optics.

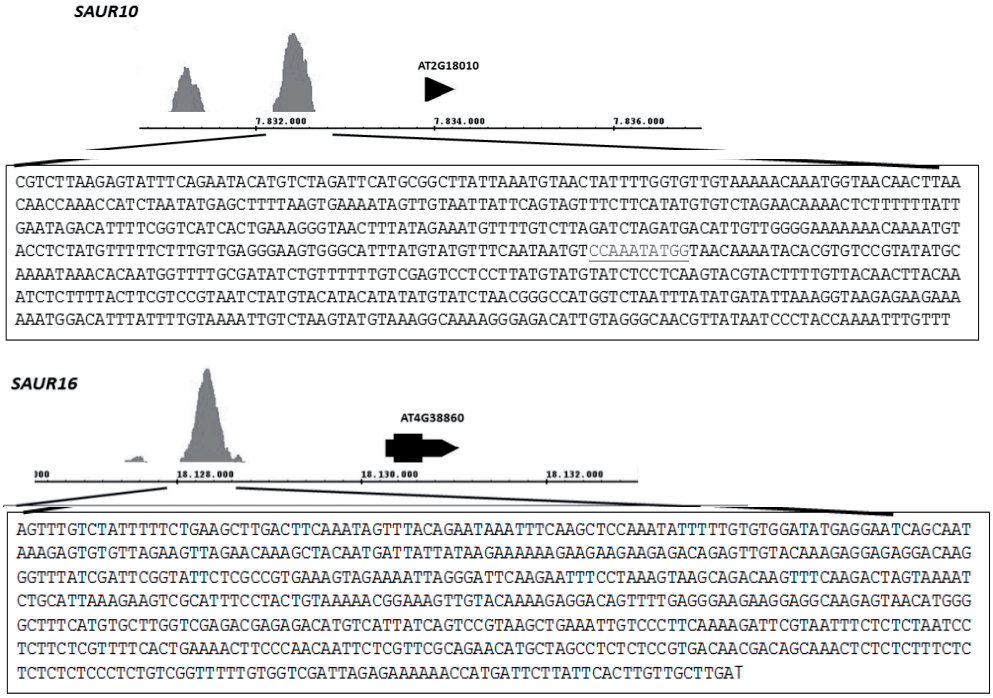
## Acknowledgements

We thank Arjo Meijering for assistance with the light measurements, Niek Stortenbeker for contributions to the manuscript, and Ueli Grossniklaus (University of Zürich) for financial and technical support. M.B. was supported by the Dutch Organization for Scientific research (NWO) in the framework of the ERA-NET on Plant Genomics (ERA-PG) program project CISCODE and by an NWO Veni-grant. In part, this work was performed in Ueli Grossniklaus' laboratory at the University of Zürich with support through an EMBO LT Fellowship to M.B. and a grant from the Swiss National Science Foundation to Ueli Grossniklaus. H.v.M was supported by an NWO Vidi-grant, granted to K.K.

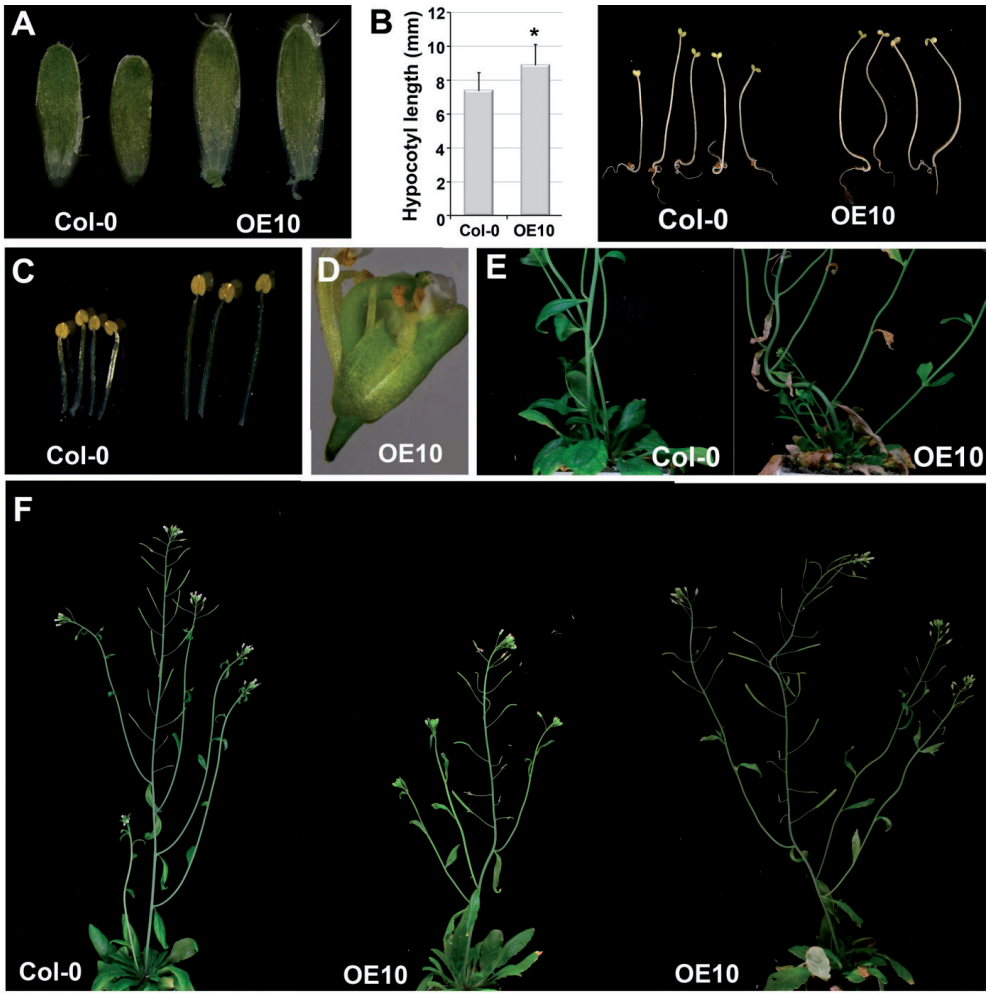
# Supplementary data



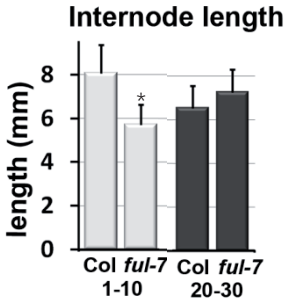

**Supplementary figure S1 | FUL is widely expressed in *Arabidopsis*.** (**A**) GUS staining of tissues from a pFUL:GUS line after overnight staining. S = 14 day old seedling/rosette, RL = rosette leaf, CL = cauline leaf, Br = side branch, Infl = inflorescence, FL.11 = flower stage 11, Fl.12 = flower stage 12, Fl.14 = flower stage 14, Fl.15 = flower stage 15, Sil. 16 = silique stage 16, 17 = silique stage 17, 18 = silique stage 18 (stages according to Smyth *et al.* (1990)). (**B**) Expression profile for FUL based on large-scale transcriptome datasets, visualized with the eFP browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) (Winter *et al.*, 2007).



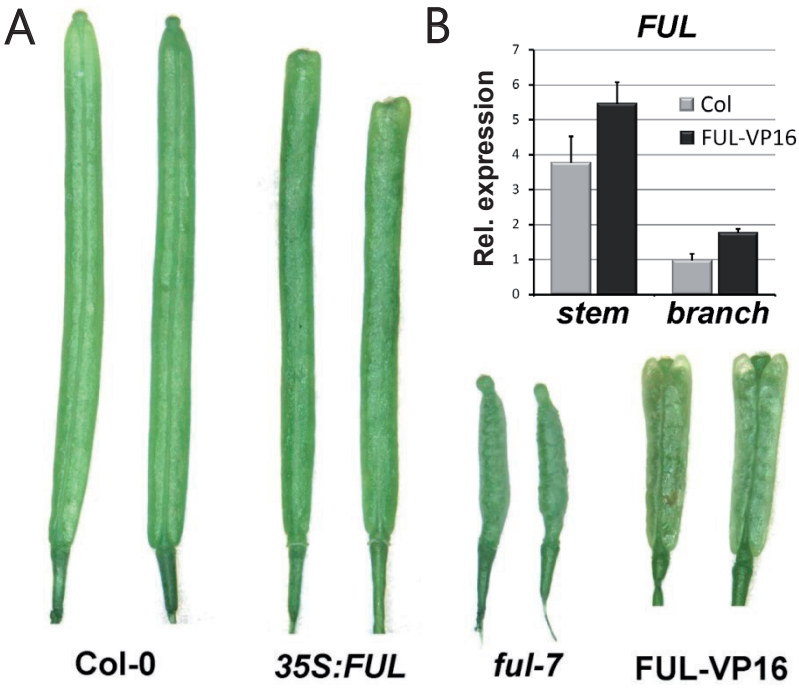

**Supplementary figure S2 | Binding sites of FUL in the *SAUR10* and *SAUR16* upstream regions.** The depicted sequences are part of the upstream regions of *SAUR10* (upper panel) and *SAUR16* (lower panel), spanning the peak region. The canonical CArG-boxes in the SAUR10 promoter is shown in gray and underlined.



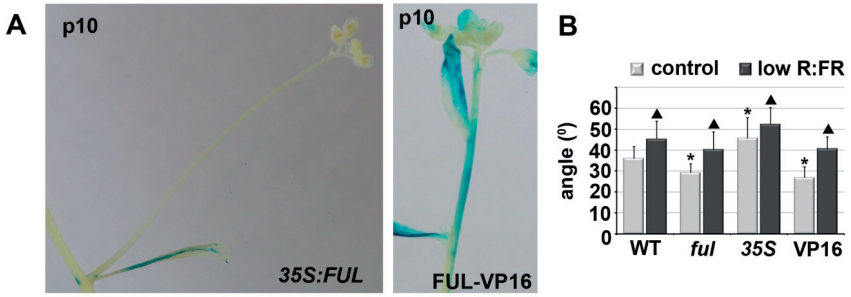

**Supplementary figure S3 | Phenotypes of the *SAUR10* overexpression lines.** (**A**) longer sepals of stage 13 flowers, (**B**) longer etiolated hypocotyls at 7 days after stratification, (**C**) longer filaments of stage 13 flowers, (**D**) very long pistils that often remain unpollinated (maximum silique length is only reached in pollinated pistils), (**E**) early senescing rosette and cauline leaves, and (**F**) Whole-plant phenotypes of Col-0 and OE10 around 8-10 days after bolting. The overexpression lines display wavy main stems and side branches, show reduced fertility due to the long pistil size, and have an irregular phyllotaxy. Significant differences (t-test, $p<0.05$) are indicated with an asterisk.




**Supplementary figure S4 | Distance between the silique internodes.** *ful-7* plants have a significantly shorter internode distance between the first 10 siliques that appear. Internode distances later in development (e.g. 20-30) are similar to the wild type. Significant differences (t-test, $p<0.05$) are indicated with an asterisk.




**Supplementary figure S5 | Characterization of the FUL-VP16 plants.** (**A**) Siliques stage 17 (Smyth et al., 1990). The apical phenotype of FUL-VP16 siliques resembles 35S:FUL siliques, with distinct 'shoulders' and a short style, but the valve tissues and the overall appearance are more similar to *ful-7* siliques (**B**) FUL is 1.5-2 times higher expressed in the FUL-VP16 plants. Stem, 0.5 cm of the stem just below the inflorescence; branch, 0.5 cm of the proximal part of the branch. The expression is depicted relative to the lowest expression (Col-0 in branch). A significant difference from the corresponding Col-0 sample ($p < 0.05$) is indicated with an asterisk.




**Supplementary figure S6 | FUL represses *SAUR10* in branches, which can be correlated to branch angle.** (**A**) Left panel: in *35S:FUL* branches, *pSAUR10:GUS* signal is only weakly visible or completely absent. Right panel: the pattern in FUL-VP16 branches resembles the pattern in *ful-7* branches with de-repressed *pSAUR10:GUS* just below the inflorescence. (**B**) Branch angles are larger in *35S:FUL*, and smaller in *ful-7* and FUL-VP16 lines. Simulated shade resulted in a more horizontal branch growth in all backgrounds. Significant differences from the wild type (t-test, $p<0.05$) are indicated with an asterisk; significant differences from control light conditions in the same background (t-test, $p<0.05$) are indicated with a triangle.

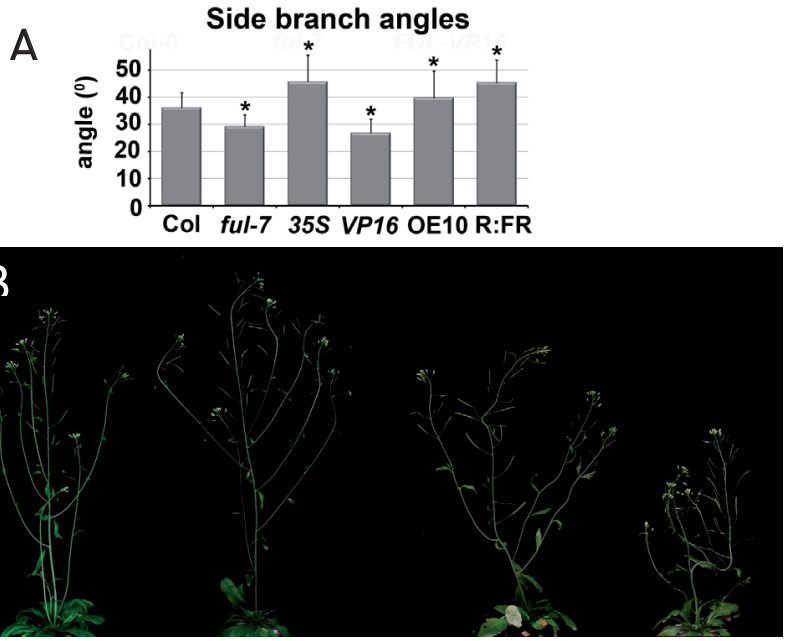



**Figure S7 | The architecture of Col-0 and *35S:SAUR10* plants changes under reduced R:FR conditions.** (**A**) Average branch angle measured in 35S:SAUR10 and under reduced R:FR conditions, compared with the angles in the different FUL backgrounds. (**B**) Architecture phenotypes of Col-0 and 35S:SAUR10 under control conditions and reduced R:FR conditions. Each bar represents the average of at least 20 measured branches. The plants were transferred to reduced R:FR light upon bolting.

**Supplementary table S1** | Identified loci with significant enrichment in the FUL ChIP-Seq

**Supplementary table S2** | Primer list

# Chapter 5

## FRUITFULL directly regulates two flavonoid synthesis genes, indicating a possible link between flowering time and flavonoids

Hilda van Mourik
Marco Busscher
Kerstin Kaufman
Gerco C Angenent

## Abstract

Flowering time is regulated by complex genetic networks that integrate environmental and developmental cues. Within this network many MADS domain proteins play important roles. One of these MADS domain proteins if FRUITFULL (FUL). Several studies have reported the role of FUL in meristem maintenance, flowering time, and shoot maturation. However, little is known about the down-stream effectors of this transcription factor. This chapter uses ChIP-seq and RNA-seq data to determine putative direct targets of FUL in the inflorescence meristem. Among the putative direct targets two genes involved in flavonoid biosynthesis were identified, *FLAVONOID SYNTHESE 1 (FLS1)* and *UDP-GLUCOSYL TRANSFERASE 78D3 (UGT78D3)*. Interestingly, similar to the *ful-7* mutant, the *fls1* mutant is late flowering. Moreover, expression data reveal an increased gene expression for both *FLS1* and *UGT78D3* in developing meristems and *FLS1* expression to be influenced by light conditions. Hence, this chapter is the first report linking the MADS domain protein FUL and flavonoids biosynthesis in *Arabidopsis*. Moreover, our initial results indicate a possible link between flavonoids and flowering time.

# Introduction

After a period of vegetative growth, plants switch to reproductive development to generate flowers. The correct timing of this transition is essential for the reproductive success of the plant. To ensure flowering at favourable conditions, flowering time is regulated by complex genetic networks that integrate both environmental and developmental signals. Within these complex regulatory networks MADS box transcription factors (TF) play important roles (reviewed in Smaczniak *et al.* (2012a)).

One of the many MADS domain proteins that regulate flowering time is FRUITFULL (FUL). Although FUL is best known for its role in pistil development, FUL also functions in flowering time regulation, meristem identity specification, and shoot maturation (Hempel *et al.*, 1997; Gu *et al.*, 1998; Ferrandiz *et al.*, 2000b; Ferrandiz *et al.*, 2000a; Balanzà *et al.*, 2014). Recently we discovered that *FUL* also plays a role in the architecture of inflorescences, more specifically it determines the side branch angle in the inflorescence (Bemer *et al.*, 2017 ; Chapter 4). *ful* mutants show a minor late flowering phenotype under continuous light and long day conditions (Ferrandiz *et al.*, 2000b; Balanzà *et al.*, 2014), in short day this phenotype is stronger (Balanzà *et al.*, 2014). A much stronger flowering time phenotype was seen for the double mutant between *FUL* and *SUPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*) (Melzer *et al.*, 2008; Balanzà *et al.*, 2014), demonstrating the redundancy between these two MADS-box genes. Besides the partial redundancy with *SOC1*, *FUL* also has a redundant function with *APETALA1 (AP1)* and *CAULIFLOWER (CAL)* in flower meristem identity specification. The *ap1 cal ful* triple mutant shows a dramatic non-flowering phenotype where leafy shoots are generated in place of flowers caused by a lack of *LEAFY (LFY)* upregulation and ectopic expression of the flowering time repressor *TERMINAL FLOWER1 (TFL1)* (Ferrandiz *et al.*, 2000b).

Studies have shown that *FUL* is one of the earliest genes to respond to changes in photo-inductive signals as a target of the flowering time promotor complex FLOWERING LOCUS T (FT) - FLOWERING LOCUS D (FD) (Hempel *et al.*, 1997; Schmid *et al.*, 2003; Teper-Bamnolker and Samach, 2005). Besides being involved in the photoperiod pathway of flowering, FUL is also influenced by the age pathway leading to flowering time regulation via SQUAMOSA PROMOTER BINDING LIKE (SPL) proteins (Shikata *et al.*, 2009; Wang *et al.*, 2009; Yamaguchi *et al.*, 2009).

Moreover, it has been suggested that FUL is able to affect flowering time via the ambient temperature pathway as a gene that responds to miR156/SPL3 and FT (Kim *et al.*, 2012).

Little is known about the downstream effectors of FUL. A small step in this direction was made in a recent paper, showing that FUL directly regulates the expression of the MADS box TF *SOC1* and the flowering regulator *LFY* (Balanzà *et al.*, 2014). Balanzà *et al.* (2014) proposed a model where, during vegetative growth, expression of *SOC1* is repressed by the MADS heterodimer FLOWERING LOCUS C (FLC) - SHORT VEGETATIVE PHASE (SVP). Upon *FUL* accumulation the resulting FUL protein competes with FLC for dimerization with SVP. The newly formed FUL-SVP dimers can also bind the *SOC1* promoter, thereby competing with FLC-SVP dimers. By lowering the concentration of repressive FLC-SVP complexes on the *SOC1* promoter and possibly also by direct activation of *SOC1* expression, FUL triggers *SOC1* accumulation. In its turn, SOC1 also forms dimers with FUL, this FUL-SOC1 dimer binds and activates the *LFY* promoter thereby triggering *LFY* expression and hence flower initiation.

To further expand our knowledge about the downstream effectors of FUL, this chapter aims to characterize putative direct target genes of FUL in the inflorescence meristem (IM). Interestingly, two genes out of the identified 9 putative direct targets function in subsequent steps in the flavonoids biosynthesis pathway. Flavonoids are important plant metabolites that are involved in many signalling pathways. Examples of functions of flavonoids are as pigments of flowers and fruits, responses to ambient temperature, UV filtering, regulation of auxin transport, and plant protection against pathogen and herbivores (Brown *et al.*, 2001; Pollastri and Tattini, 2011; Kuhn *et al.*, 2017). Although flavonoid concentrations are increased by light and reports have suggested a possible link between flavonoids and photomorphogenesis (Nguyen *et al.*, 2016), to our knowledge there are no reports on a possible link between flavonoids and flowering.

This chapter shows that the flowering time regulator FUL directly binds and regulates the expression of two flavonoid biosynthesis genes. Mutant phenotyping shows delayed flowering for these flavonoid synthesis genes. Hence, this chapter shows initial results on a possible link between flowering time and flavonoids.

# Results and discussion

Chapter 2 of this thesis shows that gene regulation and genomic DNA binding by the MADS domain protein FUL is tissue specific. To study the mechanisms that determine tissue-specific FUL functions, several techniques were used including chromatin immunoprecipitation followed by sequencing (ChIP-seq) and RNA sequencing (RNA-seq). Combining ChIP-seq to identify genome-wide binding and RNA-seq to determine genes regulated by FUL allows to identify putative direct targets of FUL. This chapter takes a detailed look at the potential direct targets of FUL in the IM tissue.

## Identification of putative direct targets

To identify putative direct targets in the IM, ChIP-seq was performed on IM tissue of *ap1 cal* pAP1::AP1-GR (no DEX treatment) plants expressing pFUL::FUL-GFP using an anti-GFP antibody. ChIP-seq experiments resulted in 2538 significant FUL-bound genomic regions in the IM (FDR ≤ 0.05). We defined bound genes as genes containing a significant ChIP-seq peaks located up to 3 kb upstream of the transcriptional start site (TSS) to 1 kb downstream of the gene. A total of 4055 significant genes were identified (FDR ≤ 0.05)(Figure 1). As expected from TFs, binding sites are mostly found in the promoter regions 5' of the TSS (Figure 1A). A MEME-chip search identified a MADS domain protein binding site, a CArG-box (CC[AT]$_6$GG) to be significantly enriched in peak centers of FUL ChIP-seq binding peaks (Figure 1B). These result highlight the good quality of the data.

**Table 1 | ChIP-seq data, numbers of significant genes and peaks**

| | | FDR Thresholds | | | |
|---|---|---|---|---|---|
| Tissue | | 0.05 | 0.01 | 0.005 | 0.001 |
| IM | Peak height threshold | 7.99 | 11.62 | 12.58 | 14.84 |
| | Total number of peaks | 2538 | 1194 | 1031 | 765 |
| | Total number of putative targets | 4055 | 1940 | 1692 | 1252 |

As mentioned in the introduction, little is known about the downstream targets of FUL in the IM. Till now two genes have been reported as direct targets of FUL in the IM: *SOC1* and *LFY* (Balanzà *et al.*, 2014). Similar to the data of Balanzà *et al.* (2014), we identified a FUL binding peak close to

the *SOC1* locus (**Figure 1C**). In our experiments no significant binding to the *LFY* locus was detected. Balanzà *et al.* reported *LFY* to be bound by the FUL-SVP heterodimer, but since SVP is hardly expressed in the IM, the absence of binding might be caused by a lack of heterodimerization partner (Gregis *et al.*, 2009). Alternatively, the differences between the results might arise from the use of 35S::FUL-GFP plants by Balanzà *et al.*, instead of the endogenous *FUL* promoter. The use of this construct results in high and partially ectopic expression of FUL which can lead to false-positive binding.

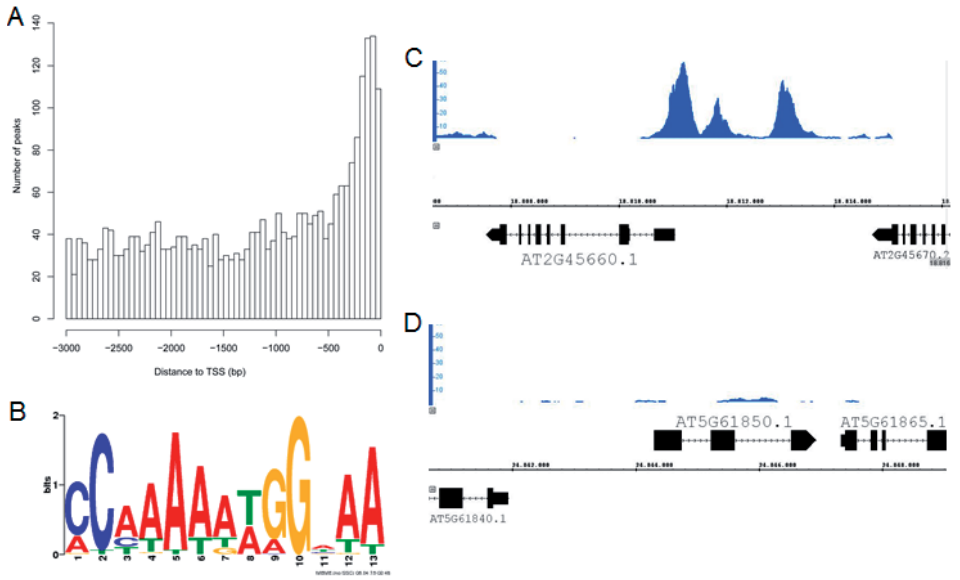



**Figure 1 | ChIP-seq results on genome-wide binding of FUL**. (**A**) FUL ChIP-seq peak positions relative to the TSS. (**B**) MADS binding motif, CArG-box, significantly enriched in ChIP-seq data based on MEME-chip (E-value = $3.5e^{-26}$). And ChIP-seq binding profiles of (**C**) *SOC1 (AT2G45660)* and (**D**) *LFY (AT5G61850).* The graphs represent for each locus the ChIP-seq peak profile. Chromosomal position (TAIR10) and gene models are shown at the bottom of each panel.

To identify genes regulated by FUL, RNA-seq was performed on IM tissue (stage 1-9 flower buds according to Smyth *et al.* (1990)), comparing wild type versus *ful-1* mutant. We identified a total of 45 significant differentially expressed genes (DEGs) (p-value ≤ 0.05 & |FC| ≥ 1.8). The number of DEGs is small and the list does not include the known FUL target, *SOC1.* The low number of DEG's and lack of differential expression of *SOC1* is most likely caused by redundancy of FUL with other MADS domain proteins in the IM: SOC1, AP1 and CAL (Ferrandiz *et al.*, 2000b; Melzer *et al.*, 2008;

Torti and Fornara, 2012; Balanzà *et al.*, 2014). As a consequence of these redundancies, we expect that the identified 45 DEGs are specifically regulated by FUL and independently of AP1, SOC1 and CAL.

To determine genes directly regulated by FUL we combined the ChIP- and RNA-seq results, identifying genes that are both bound and regulated by FUL. Eight genes were overlapping between the ChIP- and RNA-seq data sets (**Figure 2, Table 2**). ChIP-seq peaks patterns of these genes are shown in **Figure 3**. Most of the identified putative direct targets do not have a known function in meristem development or flowering time regulation. In the next paragraphs we aim to give more insight into the putative direct targets and reflect on possible links between the targets and the functions of FUL.

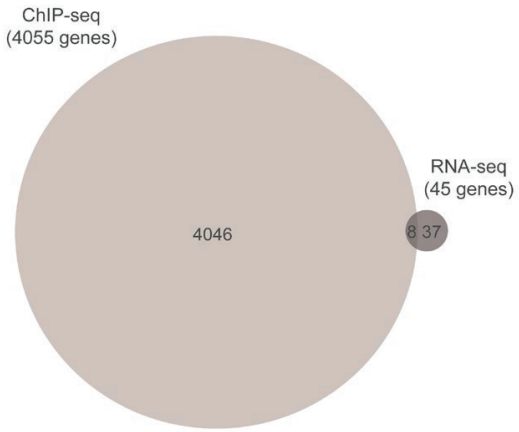



**Figure 2 | Putative direct targets of FUL in IM tissue.** Venn diagram showing the overlap between ChIP-seq bound genes and differentially expressed genes (DEGs) identified by RNA-seq comparing wild type and *ful-1* mutant.

**Table 2 | Putative direct targets of FUL in the IM**

| AT-number | Name | ChIP-seq score* | RNA-seq | |
|---|---|---|---|---|
| | | | $log_2$(FC)** | Adjusted P-value |
| At1g01070 | *UMAMIT28* | 8,51 | 2,02 | 0,023 |
| At1g45201 | *TLL1* | 12,98 | 1,62 | 0,004 |
| At2g34655 | *npcRNA* | 9,32 | 4,74 | 3,55E-22 |
| At2g41475 | *ATS3* | 9,41 | 1,00 | 0,005 |
| At4g27950 | *CRF4* | 27,82 | 5,62 | 7,48E-08 |
| At5g08640 | *FLS1* | 12,7 | -0,97 | 0,001 |
| At5g17030 | *UGT78D3* | 20,63 | -1,16 | 0,002 |
| At2g45050 | *GATA2* | 26,1 | 1,81 | 0,044 |

* score represents maximum peak height 3 kb upstream to 1 kb downstream of the gene

** $log_2$ Fold Change reflects *ful-1* mutant versus the wild type

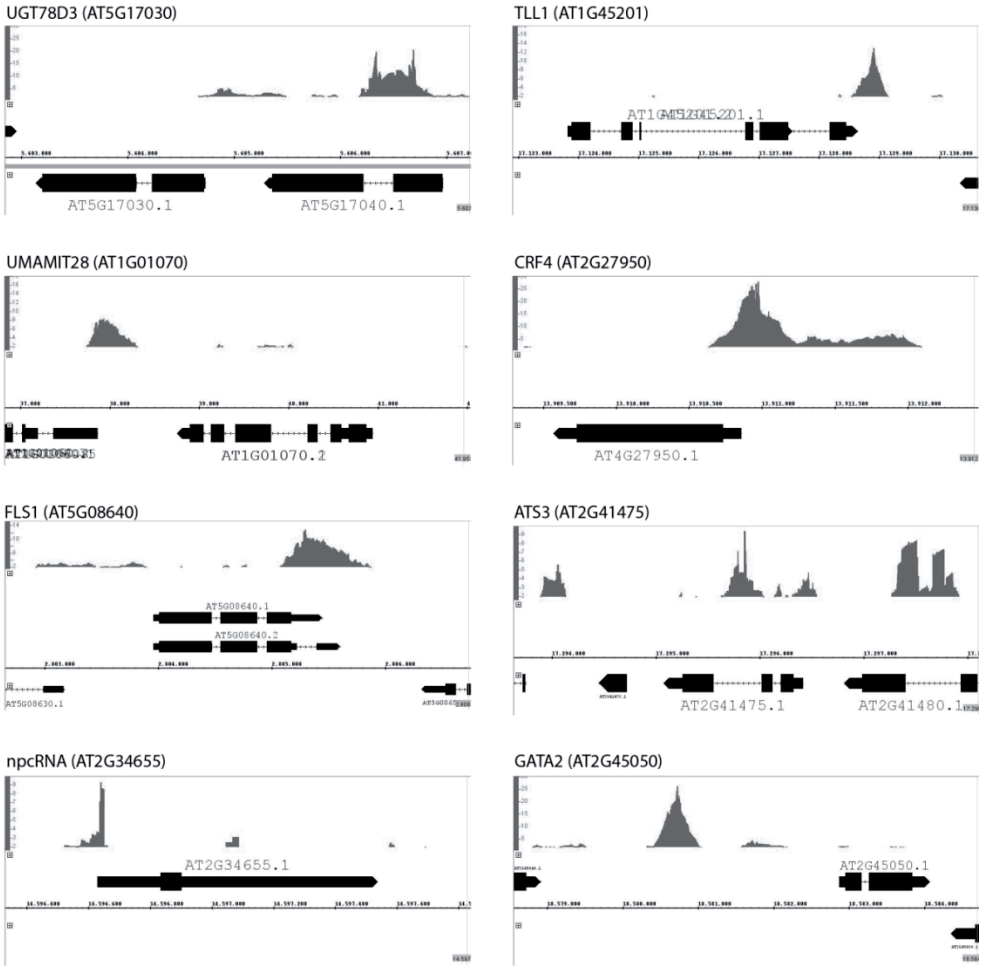



**Figure 3 | ChIP-seq binding profiles of FUL at the genomic locus of the eight putative direct targets.** The graphs represent for each locus the ChIP-seq peak profile. Chromosomal position (TAIR10) and gene models are shown at the bottom of each panel.

## Phenotypic screening of T-DNA insertion lines

The eight putative direct targets represent genes regulated by FUL independent from the redundant MADS domain proteins SOC1, AP1 and CAL. To determine the FUL-specific roles in the IM we screened the *ful-7* mutant (SALK_033647) for flowering time and cauline leaf number. As the phenotype of *ful* mutants has been reported to be stronger under short day (SD) conditions (Balanzà *et al.,* 2014), we screened the *ful-7* mutant in SD. The *ful-7* mutant showed a clear

phenotypical difference compared to wild type plants (**Figure 4**). *Ful-7* generated many more cauline leaves (p-value 7.01 x $10^{-36}$) and had a longer stem (not quantified) (**Figure 4A & B**). The high number of cauline leaves in the *ful-7* mutant, suggest a role in floral meristem initiation. Besides a difference in number of cauline leaves, we also found a significant late flowering phenotype (p-value 4.9 x $10^{-15}$)(**Figure 4C**). These phenotypes are in agreement with the results of Balanzà *et al.* (2014) for *ful-2* and *ful-1* mutant plants.

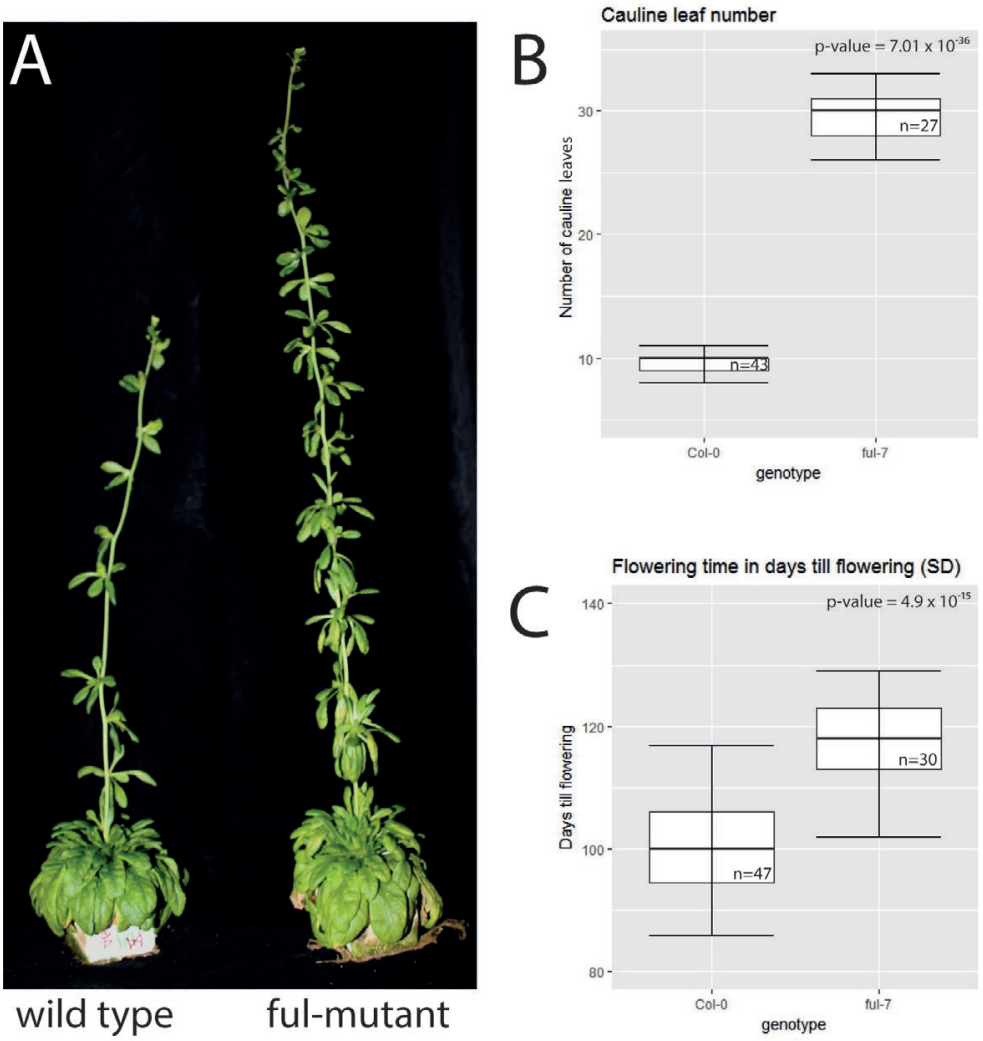



**Figure 4 | *ful-7* mutant phenotype in short day (SD) growth conditions.** (**A**) *ful-7* mutant phenotype compared to a wild type plant. (**B**) Box-plot showing the number of cauline leaves until flowering for both wild type and *ful-7* mutants. (**C**) Box-plot showing the number of days till flowering in SD for both wild type and *ful-7* mutants.

To determine whether the *ful* mutant phenotypes in SD could explained by the differential expression of the eight putative direct targets we selected T-DNA insertion lines for *UGT78D3* (SALK_114099), *UMAMIT28* (SALK_099741), *CRF4* (FLAG_423H06), and *FLS1* (SALK_106244) (**Figure 6A-D**). All insertions were confirmed by PCR. T-DNA insertion lines of *CRF4*, *FLS1* and *UMAMIT28* are exon insertions (**Figure 6B-D**). As SALK_114099 is a promoter insert (**Figure 6A**), downregulation of *UGT78D3* expression was confirmed by qPCR (**Figure 5**). For the other putative direct targets no insertion lines were available, or insertions could not be confirmed.

Analysis of the number of cauline leaves showed no significant difference compared to wild type for any of the analysed lines (**Figure 6F-I**). However, the flowering time results showed significantly different flowering time for *umamit28* (p-value 0.0006) and *fls1* (p-value 0.0003), the other lines did not show a significant result (**Figure 6J-M**). The late flowering phenotypes of *umamit28* and *fls1* suggest that both genes, similar to FUL, are positive regulators of flowering. In our RNA-seq experiment *UMAMIT28* is ~4 times higher expressed in the mutant versus the wild type, hence FUL negatively regulates *UMAMIT28*. The negative regulation of *UMAMIT28* by FUL does not fit the observed flowering time phenotype of *umamit28*, suggesting that the flowering time phenotype of FUL cannot be explained via *UMAMIT28*. *FLS1* has shown to have a 0.511 fold change in our RNA-seq experiments, *ful-1* mutant versus the wild type, hence FUL positively regulated *FLS1*. This fits the observed flowering time phenotype, where both FUL and FLS1 positively regulate flowering. Hence, *FLS1* is a candidate gene to be downstream of FUL to regulate flowering.

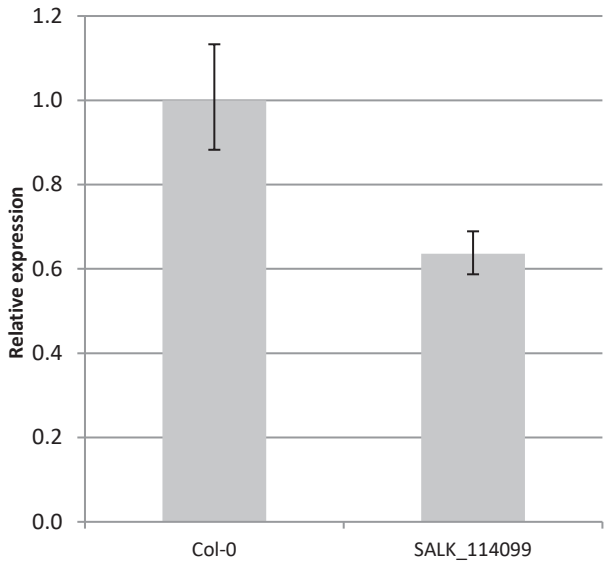



**Figure 5 | RT-qPCR expression of *UGT78D3* in SALK_114099.** Error bars represent standard error, n=3.

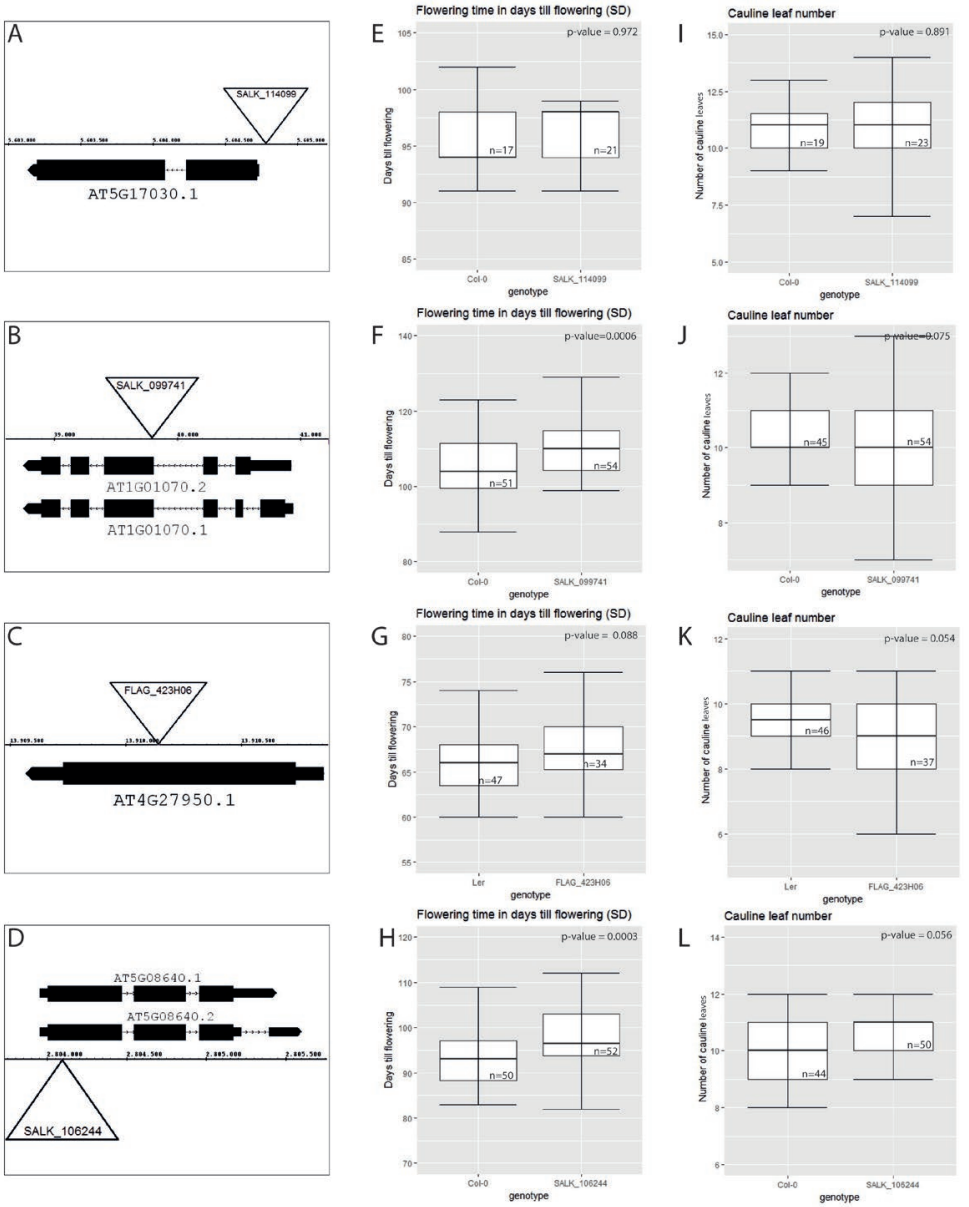



**Figure 6 | T-DNA insert location and phenotypes.** (**A – D**) Schematic representation of *UGT78D3*, *UMAMIT28*, *CRF4*, and *FLS1* loci, respectively. T-DNA insertion of corresponding lines are indicates as triangles. (**E – H**) Box-plot showing the number of days till flowering in SD for wild type and T-DNA insertion lines, outliers were removed. (**I – L**) Box-plot showing the number of cauline leaves until flowering for wild type and T-DNA insertion lines, outliers were removed.

## Relation between flowering time and flavonoid synthesis

*FLS1* encodes a flavonoid synthase that has a central role in flavonoid synthesis by catalysing the formation of flavonols from dihydroflavonols (Owens *et al.*, 2008). The *Arabidopsis* genome encodes six FLS genes, *FLS1* is functionally the most important (Owens *et al.*, 2008). Mutating or overexpressing the *FLS1* gene has been shown to influence flavonoid levels in plants (Nguyen *et al.*, 2016; Kuhn *et al.*, 2017). Interestingly, expression studies in various plant tissues show that *FLS1* is expressed most strongly in inflorescences, floral buds, and flowers (Owens *et al.*, 2008; Nguyen *et al.*, 2016). Also, publically available microarray data (AtGeneExpress & eFP browser) shows high expression of *FLS1* in inflorescences and show increasing expression in developing meristems (Redman *et al.*, 2004; Winter *et al.*, 2007)(**Figure 7**). This expression increase in developing meristems seems to follow the expression of FUL in these meristems (**Figure 7**). Genevestigator shows that the expression of *FLS1* is influenced by light intensity, light quality, among several other external influences (Hruz *et al.*, 2008; Owens *et al.*, 2008). Hence, the spatiotemporal expression pattern of *FLS1* is in line with the expectation that *FLS1* is a direct targets of FUL to regulate flowering time, possibly mediating light responses.

Among the putative direct targets of FUL *FLS1* is not the only gene involved in flavonoid biosynthesis. Also *UGT78D3* functions in flavonoid biosynthesis, it encodes an UDP-dependent glycosyltransferase. UGT78D3 uses one of the products of FLS1, quercetin, as a sugar acceptor to produce quercetin-3-O-arabinside (Yonekura-Sakakibara *et al.*, 2008). The *UGT78D3* T-DNA insertion line did not show a mutant phenotype, possibly due to the redundancy of *UGT78D3* with its close relatives *UGT78D2* and *UGT78D1* (Yonekura-Sakakibara *et al.*, 2008) or due to the subtle down-regulation in the used T-DNA insertion line (**Figure 5**). Similar to *FLS1* and *FUL*, publicly available expression data show an increase in expression of *UGT78D3* in the shoot apex at floral transition (**Figure 7**).

Although this is the first report showing a link between flavonoids and FUL in *Arabidopsis*, a link between flavonoids and FUL has been reported in Tomato. Tomato studies report that FUL binds directly to a selection of flavonoid-related genes and show that the tomato *ful* mutants have a reduced flavonoid concentration (Jaakola *et al.*, 2010; Bemer *et al.*, 2012; Fujisawa *et al.*, 2014), suggesting that the regulation of tomato flavonoid genes by FUL causes an increase in flavonoid

concentration. These data suggest that the regulation of flavonoid synthesis by FUL might be a conserved regulatory pathway in plants.

In summary, this report suggests that FUL directly regulates the expression of two flavonoid synthesis genes. The T-DNA insertion mutant *fls1,* which contains lower flavonoid concentrations (Nguyen *et al.*, 2016; Kuhn *et al.*, 2017), shows a mild late flowering phenotype. Besides the late flowering phenotype, expression studies show an increase in expression of both *FLS1* and *UGT78D3* in developing meristems, an expression pattern that seems to follow the expression of FUL. Hence, here we report the first study linking the MADS domain protein FUL and flavonoid biosynthesis in *Arabidopsis*. Moreover, initial results show a possible link novel link between flavonoids and flowering time, a link that is worth studying in more detail.

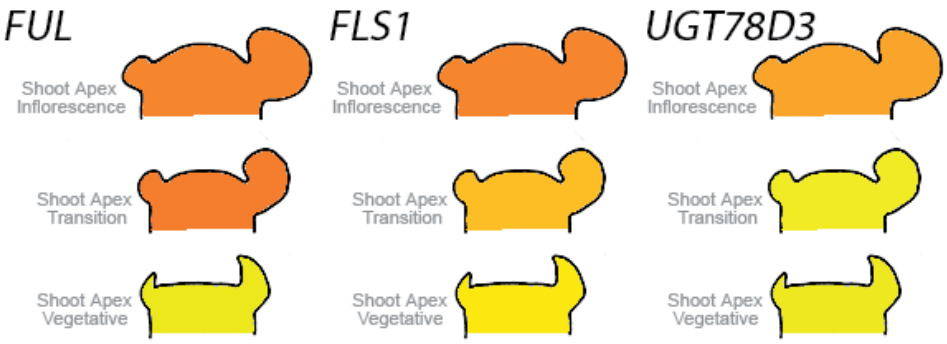



**Figure 7 | Expression profile of *FUL*, *FLS1*, and *UGT78D3* in developing meristems of** *Arabidopsis* based on publicly available microarray data (Winter *et al.*, 2007).

# Materials and methods

## Plant lines and growth conditions

Plants for ChIP-seq and RNA-seq were grown at 21°C under long day conditions (16 h light, 8 h dark). Phenotyping of wild type, *ful-7* (SALK_033647) and T-DNA insertion lines was performed on plants grown at 21°C under short day conditions (8 h light, 16 h dark). All plants were grown on rock-wool and watered with HYPONeX® solution (1.5 g/l).

## Genotyping and qRT-PCR

Before phenotyping all T-DNA insertion lines were genotyped to confirm the insert. Primers and T-DNA insertion lines can be found in Table 3. For qRT-PCR IM tissue including stages 1 – 9 flower buds (stages according to Smyth *et al.* (1990)) was collected from Col-0 and *ugt78D3* plants. For RNA extraction the InviTrap® Spin Plant RNA Mini kit (Stratec Molecular) was used. DNase treatment was performed using Ambion Turbo DNase (AM1907). For qRT-PCR analysis, the RNA was reverse transcribed using the iScript cDNA synthesis kit (BioRad). qRT-PCR reactions were performed with iQ SybrGreen supermix from BioRad and qRT-PCR was performed on the BioRad iCycler, primers used can be found in Table 4.

## ChIP-seq and RNA-seq

See chapter 2 of this thesis.

Table 3 | List of T-DNA insertion lines and primers used

| T-DNA insertion line | Genotyping Primers | | |
|---|---|---|---|
| | Fw | Rv | Lb |
| FLAG_423H06 | TCGTGCAACTCAAAGAAATTG | GATTCTACGGTTCAAAAGCCC | CGTGTGCCAGGTGCCCACGGAATAGT |
| SALK_099741 | AATGGTCGATCATTTCGTCAG | AAGGCTCAAGAGAGCACATTG | ATTTTGCCGATTTCGGAAC |
| SALK_106244C1 | TCTCGCTCTCAGCTCTTTTTG | AGCTCGAAGAATTTTCTTCCG | CGTGTGCCAGGTGCCCACGGAATAGT |
| SALK_114099 | TCGTGCAACTCAAAGAAATTG | GATTCTACGGTTCAAAAGCCC | CGTGTGCCAGGTGCCCACGGAATAGT |

Table 4 | List qRT=PCR primers used

| Target | Gene | qRT-PCR primers | |
|---|---|---|---|
| | | Fw | Rv |
| AT5G17030 | UGT78D3 | ACACCGCCTCCTGTAGAGCTT | GTACACCTGCCGACACACTCTC |

# Chapter 6

## Concluding remarks and perspectives

Regulation of gene expression is crucial for the development of different cell types, organs, and developmental structures. Essential for gene regulation are DNA binding transcription factors (TFs). To correctly regulate genes, these TFs need to bind DNA with high specificity. Hence, how TFs achieve DNA binding specificity is a fundamental question in molecular biology.

## High-throughput methods enable genome-wide binding studies

In the last decades, technological advances and decreasing sequencing costs allowed the development of several high-throughput methods to study protein-DNA interactions. Traditional *in vitro* methods to determine TF-DNA interactions usually study one or a few interactions, while new high-throughput methods can determine TF specificity by measuring the DNA binding affinity against a large collection of all possible binding sites. Several high-throughput techniques to study TF-DNA interactions are discussed in Chapter 1 of this thesis. These new technologies and methods resulted in a fast growing number of studies on DNA binding specificities of TFs. Examples are the determination of 830 binding profiles of different human and mouse TFs using systematic evolution of ligands followed by high-throughput sequencing (SELEX-seq) (Jolma *et al.*, 2015), the characterization of DNA-binding specificities of 63 plant TFs using protein binding microarrays (PBMs) (Franco-Zorrilla *et al.*, 2014), and hundreds of plant TFs using DNA affinity purification sequencing (DAP-seq) (O'Malley *et al.*, 2016). These *in vitro* affinity studies greatly advanced our understanding of DNA binding specificity.

In this thesis the *in vitro* DNA binding specificity of the MADS domain protein FRUITFULL (FUL) was studied using the high-throughput technique SELEX-seq. Other high-throughput techniques used during my PhD are: RNA-seq to study the gene expression, immunoprecipitation followed by LC-MS/MS to study *in vivo* protein-protein interactions, and chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) to study *in vivo* genome-wide DNA binding events. ChIP-seq enables researchers to determine the genome-wide binding patterns of a DNA binding protein and, by combining several ChIP-seq studies, generate large gene regulatory networks.

In the last decade, genome-wide ChIP studies were performed for many different TFs from many different organisms. Interestingly, although always thought that TFs would bind to a few hundred genes, ChIP-seq studies report thousands of binding sites in the genome. Yet, based on known *in vitro* generated DNA binding motifs, which usually are between 6 to 12 bp long many more

binding events are expected (Jolma *et al.*, 2010; Franco-Zorrilla *et al.*, 2014; O'Malley *et al.*, 2016). In line with these findings, our study on the binding of FUL shows that also this protein binds to thousands of sites, despite the presence of many more CArG-box motifs in the *Arabidopsis* genome (De Folter and Angenent, 2006). Moreover, we showed that FUL binds sites that lack the canonical binding site of MADS domain proteins, the CArG-box (**Chapter 2**). These findings show that *in vivo* DNA binding is more complex than binding affinities determined by *in vitro* methods and suggest additional *in vivo* mechanisms that determine binding.

The complexity of *in vivo* binding is moreover highlighted by the discovery of tissue-specific binding events. FUL has many distinct functions in different tissues and organs during the life cycle of *Arabidopsis* (see **Chapter 4**), the FUL binding events in two of these tissues/organs were studied in more detail in **Chapter 2** of this thesis. In the inflorescence meristem (IM), FUL functions as a positive regulator of flowering and is important for meristem identity (Mandel and Yanofsky, 1995; Hempel *et al.*, 1997; Ferrandiz *et al.*, 2000b; Balanzà *et al.*, 2014). At a later stage during flower development, FUL is expressed in the pistil where it is essential for correct patterning of the fruit (Gu *et al.*, 1998; Ferrandiz *et al.*, 2000a; Liljegren *et al.*, 2000; Roeder *et al.*, 2003; Liljegren *et al.*, 2004; Ripoll *et al.*, 2011; Ripoll *et al.*, 2015). Our tissue specific ChIP-seq study showed that only 47% of the sites bound in the IM were also significantly bound in the pistil, while the remaining 53% was uniquely bound in the IM. Also in the pistil we find 58% of the binding sites to be unique (**Chapter 2**). These results demonstrate that *in vivo* conditions and mechanisms influence genomic DNA binding.

## The cell nucleus is a crowded place

The cell nucleus is a complex organelle that, next to millions of DNA base pairs, contains numerous other molecules. Within this crowded place, DNA binding by TFs is not solitary dependent on the shape- and base-readout of the DNA and TF. *In vivo* processes influence the binding of TF to the DNA. Important parameters that influence DNA binding are chromatin accessibility, competition with nucleosomes, DNA methylation, and the availability of collaborating proteins. A review on *in vivo* TF specificity can be found in **Chapter 1**. The work in this thesis focusses on the role of protein-interactions on TF binding specificity.

Interactions between proteins can directly influence DNA binding of a TF, for example by preventing the TF to enter the nucleus. Furthermore, protein interactions can indirectly influence the binding by altering the binding specificity of a TF, for instance by influencing the structure of the DNA binding domain (DBD) (reviewed in **Chapter 1**). Recently, several studies were published aiming to unravel the role of protein interactions on DNA binding specificity. An interesting *in vitro* study on DNA binding of interacting human and mouse TFs (Jolma *et al.*, 2015) shows that many mouse and human TFs bind DNA cooperatively. Cooperative DNA binding by these TFs can influence DNA binding in different ways. Protein interactions can result in additive specificity when the specificity of the complex is a combination of the binding specificities of both proteins, alternatively, interactions can lead to altered binding where the complex binds to a completely new motif, a phenomenon called latent specificity (for review see Bemer *et al.* (2017)). Hence, protein complex formation plays an important role DNA binding specificity of TFs.

Another interesting form of protein-protein interactions is homo- and heteromerization. Most *in vitro* DNA binding studies focus on homodimeric TF-DNA interactions (e.g. Franco-Zorrilla *et al.*, 2014; O'Malley *et al.*, 2016), while many TF families form heterodimers *in vivo*. For basic leucine-zipper (bZIP) proteins, a large class of multifunctional TFs, it has been shown that different heterodimeric combinations bind slightly different DNA binding motifs (Hai and Curran, 1991; Schutze *et al.*, 2008). A recent study showed that mutations altering the interaction capability of the different bZIP proteins, changes the *in vitro* DNA binding capability (Reinke *et al.*, 2013). Hence, differences in heterodimerization potential within a family can possibly influence *in vivo* DNA binding. To test this hypothesis, the work presented in this thesis uses the MADS domain TF FUL as model to determine the influence of heterodimerization on genome-wide TF binding.

## Dimerization influences DNA binding specificities

MADS domain TFs bind DNA obligatorily as homo- and heterodimers. A comprehensive plant protein-protein interaction map of more than 100 MADS domain proteins revealed that the interactions between MADS domain proteins are highly specific (De Folter *et al.*, 2005). This study shows that FUL is able to form heterodimers with nine other MADS domain proteins. Since the interactions between MADS domain proteins are highly specific, we questioned whether the tissue-

specific binding patterns of FUL might arise from differences in available interaction partners in the two studied tissues.

In Chapter 2 we used immunoprecipitation followed by LC-MS/MS to determine *in vivo* protein interactions of FUL in IM and pistil tissues. These experiments showed that the interactions of FUL with MADS domain proteins are indeed tissue-specific, with SUPPRESSOR OF OVEREXPRESSION OF CONTSTANS 1 (SOC1) as main interactor in the IM, and SEPALLATA proteins and AGAMOUS (AG) as main partners in the pistil. Whether FUL forms homodimers in these tissues could not be determined by our MS-experiments. Using SELEX-seq we were able to confirm that, similarly to the earlier mentioned bZIP proteins, different MADS heterodimers have different *in vitro* DNA binding specificities. To determine whether the tissue-specific MADS heterodimeric specificities reflect the *in vivo* DNA binding of FUL, we mapped the SELEX-seq affinities onto the genome. Comparison of these 'genomic SELEX-seq' data with tissue-specific FUL ChIP-seq data revealed a strong correlation between tissue-specific dimer affinities and the tissue-specific genomic binding sites of FUL. Hence, our results suggest that dimerization of MADS domain proteins plays an important role in *in vivo* DNA binding specificities. Thereby, this chapter brings us a step closer to understanding how TFs can find their specific binding sites *in vivo*.

For further research it would be interesting to understand how heterodimerization can alter DNA binding specificity. For this, protein-DNA crystallography, to study how dimerization alters the interactions between the TFs and DNA, would be highly useful. Additionally, a more detailed analysis of the SELEX-seq affinities could be valuable, for example, to study whether different complexes prefer different DNA shapes or whether dependencies between nucleotides exist within the heterodimer specific motifs. Moreover, the role of MADS quaternary complexes in *in vivo* DNA binding specificity is an interesting study subject.

## FUL regulates tissue specific sets of genes

As mentioned, FUL is expressed at two stages during flower development, this dual expression pattern results in two distinct functions of FUL. In this thesis we demonstrate using ChIP-seq that FUL binds different sets of target genes in the two tissues (Chapter 2). However, a core set of genes is bound at both stages. These results confirm the idea that a core set of genomic targets is conserved among MADS domain proteins, but each MADS domain protein also binds a protein

and tissue specific set of targets. In chapter 4 and 5 of this thesis we study some of the putative direct targets of FUL in more detail.

Chapter 5 focusses on the putative direct targets of FUL in IM tissue. Among the putative direct targets two genes involved in flavonoid synthesis were identified, *FLAVONOID SYNTHESE 1* (*FLS1*) and *UDP-GLUCOSYL TRANSFERASE 78D3* (*UGT78D3*). Interestingly, similar to the *ful-7* mutant, the *fls1* mutant is late flowering. Moreover, expression data exposed an increased gene expression for both *FLS1* and *UGT78D3* in developing meristems and showed *FLS1* expression to be influenced by light conditions. Hence, these results show a possible link between MADS domain protein FUL and flavonoids synthesis in *Arabidopsis*. Moreover, our results indicate a possible link between flavonoids and flowering time. Interestingly, one of the genes, *FLS1*, is also significantly bound in pistil tissue, RNA-seq data however do not show differential expression of *FLS1* in pistil tissue.

Another example of a gene differentially bound by FUL in the two studied tissues is *SMALL AUXIN UPREGULATED RNA 10 (SAUR10).* The promoter of this gene shows two significant binding peaks in the pistil, while in the IM no significant enrichment was found. In Chapter 4 we elaborate on the function of SAUR10 and we showed that FUL plays additional and novel roles in plant growth. SAUR10 is a growth regulator that is down-regulated by FUL. The expression of SAUR10 is highly induced by a combination of auxin and brassinosteroids and by R:FR light. Our study shows that SAUR10 is expressed at the abaxial side of branches and all together we propose a model , in which SAUR10 regulates branch angle by integrating both environmental cues via auxin, brassinosteroids and R:FR light, and developmental cues via FUL. Hence, we report a new function of the MADS domain protein FUL in the regulation of branch angle via regulation of SAUR10.

Interestingly, we observed that *SAUR10* is repressed by FUL in several tissues, including cauline leaves, inflorescence stems, and branches. However, additional quantitative RT-PCR experiments with wild type and *ful-7* mutant did not show *SAUR10* expression difference in the pistil, suggesting that *SAUR10* is not regulated by FUL in the pistil. So, despite the binding of FUL to the promotor of *SAUR10* in the pistil, this binding does not result in gene regulation. Also, *FLS1* is bound by FUL in pistil-tissue but not differentially expressed. This raises the question how TF occupancies and gene regulation relate.

## The relation between DNA occupancy and gene regulation

ChIP-seq binding studies usually result in thousands of binding peaks that correspond to high numbers of putative target genes. However, many studies have shown that only a small portion of the bound genes is regulated. For the MADS domain protein FUL we found maximum 15% of bound genes to be differentially expressed (Chapter 2). However, for our RNA-seq experiment we have used mutant lines, while a better estimation of regulated genes might come from inducible lines in a corresponding mutant background. Consistent with the limited overlap between the genomic binding sites of FUL and the number of genes regulated by FUL, also genome-wide studies on other MADS domain TFs report low percentages of overlap (Kaufmann *et al.*, 2009; Tao *et al.*, 2012). For example, only 44% of the genes bound by the MADS domain proteins APETALLA1 (AP1) are differentially expressed in the mutant (Kaufmann *et al.*, 2010c). The limited overlap between binding and expression has also been shown for other species (Vokes *et al.*, 2008; Gorski *et al.*, 2011; Whitfield *et al.*, 2012; Hughes and de Boer, 2013; Bianco *et al.*, 2014). Hence, the observed limited number of binding sites that result in differential expression is widespread.

Several studies focused on the relevance of DNA binding by TFs. One aspect of ChIP-seq datasets is peak height, which is used by researchers to try to find a relation between ChIP-seq peak height and gene regulation, but different studies show conflicting results (Fisher *et al.*, 2012; Lickwar *et al.*, 2012). Moreover, ChIP experiments are usually based on millions of cells, therefore peak scores represent an average of all the cells. Hence, an average binding peak could be the result of a high affinity binding site in only part of the cells, while a high binding peak may arise from medium binding in all cells. For our FUL ChIP-seq study, increasing the peak height cut-off did not lead to a higher percentage of regulated genes (results not shown). In the future, single-cell sequencing may provide more information on the relation between peak height and gene regulation.

Besides the interpretation of ChIP-seq binding, also the mode of action of a TF can influence the overlap between ChIP-seq binding and gene regulation. Not all TFs regulate the expression of a gene directly, some TFs influence gene regulation indirectly for example by opening the chromatin or by recruiting additional factors needed for transcription. In addition, models for enhancer activity state that the regulation of a gene is determined by the cooperative occupancy and activity of all TFs at the enhancer (Spitz and Furlong, 2012). Hence, the outcome of a single TF binding to the DNA is dependent on the context of other proteins bound to the promoter, and the binding of

a single TF to DNA might not be sufficient to alter gene regulation. Alternatively, the limited overlap between ChIP-seq and expression studies might be caused by redundancies between proteins. This is most likely the case for the early function of FUL. In *ful* mutant IM tissues, two redundant floral activators SOC1 and AGL24 are still active, which partially masks the regulatory role of FUL in the IM (Melzer *et al.*, 2008; Torti and Fornara, 2012). Redundancy also explains the very short list of differentially expressed genes in the IM of wild type and *ful* mutant plants.

Binding dynamics has also been proposed to influence the functionality of a binding site. A competitive ChIP study with the yeast Rap1 TF measured binding kinetics, a measure for TF residence time on the DNA (Lickwar *et al.*, 2012). By combining the binding kinetics with transcriptional activity data Lickwar et al. (2012) showed that stable, long-term binding is linked to high transcriptional activity, while short-term binding with a rapid-turnover, referred to as 'treatmilling', correlates with low transcriptional regulation. Another model for transcriptional regulation, the "Hit- and Run" model, proposes that short TF binding events can trigger the assembly of a stable transcriptional complex, thereby a TF can influence transcription even after this protein is no longer bound to the DNA (Schaffner, 1988). After "hitting" a DNA binding site, the TF can bind to many more sites in the genome to catalyse a broad transcriptional response. Recent studies in yeast (Knight *et al.*, 2014) and *Arabidopsis* (Para *et al.*, 2014) support this model. Although both theories concern short-term binding, each model refers to a different type of short-term binding event. Lickwar et al. (2012) talked about 'treatmilling' TFs that shortly bind the same locus repeatedly, while "Hit-and-Run" TFs bind to many consecutive binding loci. In a canonical ChIP study 'treatmilling' TF will be detected, but "hit-and-run" TF are less likely to be captured in the snapshot of the ChIP experiment.

The above described mechanisms provide explanations for the low number of 'functional' binding sites found in ChIP-seq experiments. It is generally accepted that a reasonable fraction of detected ChIP-seq binding sites do not influence the expression of genes. Therefore, we may question whether ChIP alone is a good method to elucidate the function of a TF. In contrast, gene expression studies do provide directly and in-directly regulated genes, hence expression studies give information about the effects of the TF in the selected tissue. Yet, despite the limitation of ChIP studies in the discrimination of functional from non-functional binding, ChIP studies can be used together with expression studies to discriminate direct from indirect regulation. Moreover,

ChIP is an important tool for further studies on *in vivo* DNA binding specificities. The inability to determine functional binding sites using ChIP studies reflects our limited knowledge on *in vivo* gene regulation. Hence, future studies are needed to better understand how TFs regulate their target genes.

## Role of protein interactions on gene regulation

As mentioned earlier, in the last couple of years the role of protein interactions in DNA binding and gene regulation has become a focus of many studies (reviewed by Spitz and Furlong (2012)). Moreover, recent studies have shown that cross-family protein interactions are more common than initially assumed and are supposed to be a novel layer of gene regulation (Yu *et al.*, 2006; Dreze *et al.*, 2011; Rhee *et al.*, 2014; Jolma *et al.*, 2015; Yazaki *et al.*, 2016; Bemer *et al.*, 2017). In line with this, also for the MADS domain protein FUL we identified cross-family protein interactions. Additionally, among the FUL binding sites we find enrichment of non-CArG-box like motifs (G-box motifs and TCP binding sites) (Chapter 2). The role of these cross-family protein interactions in cooperative gene regulation is far from understood and will be an important research topic in the coming years.

In summary, although the work described in this thesis has increased our knowledge on genome-wide DNA binding specificities of TFs, more work is needed to understand how *in vivo* DNA binding relates to gene regulation. The growing numbers of genome-wide DNA binding studies and our increasing knowledge of protein interactions will be the basis for further studies.

References
Summary
Acknowledgements
About the author
EPS educational statement

# References

**Airoldi, C.A., Bergonzi, S., and Davies, B.** (2010). Single amino acid change alters the ability to specify male or female organ identity. Proceedings of the National Academy of Sciences **107,** 18898-18902.

**Alejandra Mandel, M., Gustafson-Brown, C., Savidge, B., and Yanofsky, M.F.** (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene APETALA1 **360,** 273-277.

**Alvarez-Buylla, E.R., Benítez, M., Corvera-Poiré, A., Chaos Cador, Á., de Folter, S., Gamboa de Buen, A., Garay-Arroyo, A., García-Ponce, B., Jaimes-Miranda, F., Pérez-Ruiz, R.V., Piñeyro-Nelson, A., and Sánchez-Corrales, Y.E.** (2010). Flower Development. The *Arabidopsis* Book / American Society of Plant Biologists **8,** e0127.

**Anders, S., and Huber, W.** (2010). Differential expression analysis for sequence count data. Genome biology **11,** R106-R106.

**Andrés, F., Romera-Branchat, M., Martínez-Gallegos, R., Patel, V., Schneeberger, K., Jang, S., Altmüller, J., Nürnberg, P., and Coupland, G.** (2015). Floral Induction in *Arabidopsis* by FLOWERING LOCUS T Requires Direct Repression of BLADE-ON-PETIOLE Genes by the Homeodomain Protein PENNYWISE. Plant Physiology **169,** 2187-2199.

**Andrews, S.** (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

**Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., and Sherlock, G.** (2000). Gene Ontology: tool for the unification of biology **25,** 25-29.

**Atamian, H.S., Creux, N.M., Brown, E.A., Garner, A.G., Blackman, B.K., and Harmer, S.L.** (2016). Circadian regulation of sunflower heliotropism, floral orientation, and pollinator visits. Science **353,** 587.

**Auerbach, R.K., Euskirchen, G., Rozowsky, J., Lamarre-Vincent, N., Moqtaderi, Z., Lefrançois, P., Struhl, K., Gerstein, M., and Snyder, M.** (2009). Mapping accessible chromatin regions using Sono-Seq. Proceedings of the National Academy of Sciences **106,** 14926-14931.

**Babraham Bioinformatics.** FASTQC: A quality control tool for high througput sequencing data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

**Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C.F., Coburn, D., Newburger, D.E., Morris, Q., Hughes, T.R., and Bulyk, M.L.** (2009). Diversity and complexity in DNA recognition by transcription factors. Science **324,** 1720-1723.

**Bai, M.-Y., Shang, J.-X., Oh, E., Fan, M., Bai, Y., Zentella, R., Sun, T.-p., and Wang, Z.-Y.** (2012). Brassinosteroid, gibberellin, and phytochrome impinge on a common transcription module in *Arabidopsis*. Nature cell biology **14,** 810-817.

**Bailey, T.L., and Elkan, C.** (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings. International Conference on Intelligent Systems for Molecular Biology **2,** 28-36.

**Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S.** (2009). MEME Suite: tools for motif discovery and searching. Nucleic Acids Research **37,** W202-W208.

**Balanzà, V., Martínez-Fernández, I., and Ferrándiz, C.** (2014). Sequential action of FRUITFULL as a modulator of the activity of the floral regulators SVP and SOC1. Journal of Experimental Botany **65,** 1193-1203.

**Balkunde, R., Bouyer, D., and Hülskamp, M.** (2011). Nuclear trapping by GL3 controls intercellular transport and redistribution of TTG1 protein in *Arabidopsis*. Development **138,** 5039-5048.

**Bardet, A.F., He, Q., Zeitlinger, J., and Stark, A.** (2012). A computational pipeline for comparative ChIP-seq analyses **7,** 45-61.

**Bargmann, B.O.R., Vanneste, S., Krouk, G., Nawy, T., Efroni, I., Shani, E., Choe, G., Friml, J., Bergmann, D.C., Estelle, M., and Birnbaum, K.D.** (2013). A map of cell type-specific auxin responses. Molecular Systems Biology **9**.

**Bartrina, I., Otto, E., Strnad, M., Werner, T., and Schmülling, T.** (2011). Cytokinin Regulates the Activity of Reproductive Meristems, Flower Organ Size, Ovule Formation, and Thus Seed Yield in *Arabidopsis* thaliana. The Plant cell **23,** 69-80.

**Bemer, M., van Dijk, A.D.J., Immink, R.G.H., and Angenent, G.C.** (2017). Cross-Family Transcription Factor Interactions: An Additional Layer of Gene Regulation. Trends in Plant Science **22,** 66-80.

**Bemer, M., Karlova, R., Ballester, A.R., Tikunov, Y.M., Bovy, A.G., Wolters-Arts, M., Rossetto, P.d.B., Angenent, G.C., and de Maagd, R.A.** (2012). The Tomato FRUITFULL Homologs TDR4/FUL1 and MBP7/FUL2 Regulate Ethylene-Independent Aspects of Fruit Ripening. The Plant cell **24,** 4437-4451.

**Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., Khalid, F., Zhang, W., Newburger, D., Jaeger, S., Morris, Q.D., Bulyk, M.L., and Hughes, T.R.** (2008). Variation in homeodomain DNA-binding revealed by high-resolution analysis of sequence preferences. Cell **133,** 1266-1276.

**Berger, S., Mitchell-Olds, T., and Stotz, H.U.** (2002). Local and differential control of vegetative storage protein expression in response to herbivore damage in *Arabidopsis* thaliana. Physiologia Plantarum **114,** 85-91.

**Bianco, S., Brunelle, M., Jangal, M., Magnani, L., and Gevry, N.** (2014). LRH-1 governs vital transcriptional programs in endocrine-sensitive and -resistant breast cancer cells. Cancer research **74,** 2015-2025.

**Boer, D.R., Freire-Rios, A., van den Berg, Willy A.M., Saaki, T., Manfield, Iain W., Kepinski, S., López-Vidrieo, I., Franco-Zorrilla, Jose M., de Vries, Sacco C., Solano, R., Weijers, D., and Coll, M.** (2014). Structural Basis for DNA Binding Specificity by the Auxin-Dependent ARF Transcription Factors. Cell **156,** 577-589.

**Bouyer, D., Geier, F., Kragler, F., Schnittger, A., Pesch, M., Wester, K., Balkunde, R., Timmer, J., Fleck, C., and Hulskamp, M.** (2008). Two-dimensional patterning by a trapping/depletion mechanism: the role of TTG1 and GL3 in *Arabidopsis* trichome formation. PLoS biology **6,** e141.

**Brayer, K.J., and Lynch, V.J.** (2011). Evolution of a derived protein–protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation. Proceedings of the National Academy of Sciences.

**Brown, D.E., Rashotte, A.M., Murphy, A.S., Normanly, J., Tague, B.W., Peer, W.A., Taiz, L., and Muday, G.K.** (2001). Flavonoids Act as Negative Regulators of Auxin Transport in Vivo in *Arabidopsis*. Plant Physiology **126,** 524-535.

**Buck-Koehntop, B.A., Stanfield, R.L., Ekiert, D.C., Martinez-Yamout, M.A., Dyson, H.J., Wilson, I.A., and Wright, P.E.** (2012). Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proceedings of the National Academy of Sciences **109,** 15229-15234.

**Burdach, J., Funnell, A.P.W., Mak, K.S., Artuz, C.M., Wienert, B., Lim, W.F., Tan, L.Y., Pearson, R.C.M., and Crossley, M.** (2013). Regions outside the DNA-binding domain are critical for proper in vivo specificity of an archetypal zinc finger transcription factor. Nucleic Acids Research.

**Campbell, C.T., and Kim, G.** (2007). SPR microscopy and its applications to high-throughput analyses of biomolecular binding events and their kinetics. Biomaterials **28,** 2380-2392.

**Chae, K., Isaacs, C.G., Reeves, P.H., Maloney, G.S., Muday, G.K., Nagpal, P., and Reed, J.W.** (2012). *Arabidopsis* SMALL AUXIN UP RNA63 promotes hypocotyl and stamen filament elongation. The Plant Journal **71,** 684-697.

**Challa, K.R., Aggarwal, P., and Nath, U.** (2016). Activation of YUCCA5 by the Transcription Factor TCP4 Integrates Developmental and Environmental Signals to Promote Hypocotyl Elongation in *Arabidopsis*. The Plant cell.

**Chapman, E.J., Greenham, K., Castillejo, C., Sartor, R., Bialy, A., Sun, T.-p., and Estelle, M.** (2012). Hypocotyl Transcriptome Reveals Auxin Regulation of Growth-Promoting Genes through GA-Dependent and -Independent Pathways. PLoS ONE **7,** e36210.

**Chen, Y., Zhang, X., Dantas Machado, A.C., Ding, Y., Chen, Z., Qin, P.Z., Rohs, R., and Chen, L.** (2013). Structure of p53 binding to the BAX response element reveals DNA unwinding and compression to accommodate base-pair insertion. Nucleic Acids Res **41,** 8368-8376.

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

**Consortium, T.G.O.** (2015). Gene Ontology Consortium: going forward. Nucleic Acids Research **43,** D1049-D1056.

**Crocker, J., Abe, N., Rinaldi, L., McGregor, Alistair P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., Mann, Richard S., and Stern, David L.** (2015). Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. Cell **160,** 191-203.

**Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.-R.** (2005). Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in *Arabidopsis*. Plant Physiology **139,** 5-17.

**D'Aloia, M., Bonhomme, D., Bouché, F., Tamseddak, K., Ormenese, S., Torti, S., Coupland, G., and Périlleux, C.** (2011). Cytokinin promotes flowering of *Arabidopsis* via transcriptional activation of the FT paralogue TSF. The Plant Journal **65,** 972-979.

**Danisman, S., van der Wal, F., Dhondt, S., Waites, R., de Folter, S., Bimbo, A., van Dijk, A.D., Muino, J.M., Cutri, L., Dornelas, M.C., Angenent, G.C., and Immink, R.G.H.** (2012). *Arabidopsis* Class I and Class II TCP Transcription Factors Regulate Jasmonic Acid Metabolism and Leaf Development Antagonistically. Plant Physiology **159,** 1511-1523.

**De Folter, S., and Angenent, G.C.** (2006). Trans meets cis in MADS science. Trends Plant Sci **11**.

**De Folter, S., Urbanus, S., van Zuijlen, L., Kaufmann, K., and Angenent, G.** (2007). Tagging of MADS domain proteins for chromatin immunoprecipitation. BMC Plant Biology **7,** 47.

**De Folter, S., Immink, R.G., Kieffer, M., Parenicova, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M., Davies, B., and Angenent, G.C.** (2005). Comprehensive interaction map of the *Arabidopsis* MADS Box transcription factors. The Plant cell **17,** 1424-1433.

**Deng, W., Ying, H., Helliwell, C.A., Taylor, J.M., Peacock, W.J., and Dennis, E.S.** (2011). FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of *Arabidopsis*. Proceedings of the National Academy of Sciences **108,** 6680-6685.

**Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., and Schübeler, D.** (2015). Competition between DNA methylation and transcription factors determines binding of NRF1 **528,** 575-579.

**Dorca-Fornell, C., Gregis, V., Grandi, V., Coupland, G., Colombo, L., and Kater, M.M.** (2011). The *Arabidopsis* SOC1-like genes AGL42, AGL71 and AGL72 promote flowering in the shoot apical and axillary meristems. The Plant Journal **67,** 1006-1017.

**Dreze, M., Carvunis, A.-R., Charloteaux, B., Galli, M., Pevzner, S.J., Tasan, M., Ahn, Y.-Y., Balumuri, P., Barabási, A.-L., Bautista, V., Braun, P., Byrdsong, D., Chen, H., Chesnut, J.D., Cusick, M.E., Dangl, J.L., de los Reyes, C., Dricot, A., Duarte, M., Ecker, J.R., Fan, C., Gai, L., Gebreab, F., Ghoshal, G., Gilles, P., Gutierrez, B.J., Hao, T., Hill, D.E., Kim, C.J., Kim, R.C., Lurin, C., MacWilliams, A., Matrubutham, U., Milenkovic, T., Mirchandani, J., Monachello, D., Moore, J., Mukhtar, M.S., Olivares, E., Patnaik, S., Poulin, M.M., Przulj, N., Quan, R., Rabello, S., Ramaswamy, G., Reichert, P., Rietman, E.A., Rolland, T., Romero, V., Roth, F.P., Santhanam, B., Schmitz, R.J., Shinn, P., Spooner, W., Stein, J., Swamilingiah, G.M., Tam, S., Vandenhaute, J., Vidal, M., Waaijers, S., Ware, D., Weiner, E.M., Wu, S., and Yazaki, J.** (2011). Evidence for Network Evolution in an *Arabidopsis* Interactome Map. Science **333,** 601-607.

**Egea-Cortines, M., Saedler, H., and Sommer, H.** (1999). Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in Antirrhinum majus. Embo j **18,** 5370-5379.

**Fendrych, M., Leung, J., and Friml, J.** (2016). TIR1/AFB-Aux/IAA auxin perception mediates rapid cell wall acidification and growth of *Arabidopsis* hypocotyls. eLife **5,** e19048.

**Ferrandiz, C., Liljegren, S.J., and Yanofsky, M.F.** (2000a). Negative regulation of the SHATTERPROOF genes by FRUITFULL during *Arabidopsis* fruit development. Science **289**.

**Ferrandiz, C., Gu, Q., Martienssen, R., and Yanofsky, M.F.** (2000b). Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. Development **127,** 725-734.

**Fisher, W.W., Li, J.J., Hammonds, A.S., Brown, J.B., Pfeiffer, B.D., Weiszmann, R., MacArthur, S., Thomas, S., Stamatoyannopoulos, J.A., Eisen, M.B., Bickel, P.J., Biggin, M.D., and Celniker, S.E.** (2012). DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. Proceedings of the National Academy of Sciences of the United States of America **109,** 21330-21335.

**Fordyce, P.M., Gerber, D., Tran, D., Zheng, J., Li, H., DeRisi, J.L., and Quake, S.R.** (2010). De Novo Identification and Biophysical Characterization of Transcription Factor Binding Sites with Microfluidic Affinity Analysis. Nature biotechnology **28,** 970-975.

**Franco-Zorrilla, J.M., Lopez-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R.** (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. Proceedings of the National Academy of Sciences of the United States of America **111,** 2367-2372.

**Fu, X., Li, C., Liang, Q., Zhou, Y., He, H., and Fan, L.-M.** (2016). CHD3 chromatin-remodeling factor PICKLE regulates floral transition partially via modulating LEAFY expression at the chromatin level in *Arabidopsis*. Science China Life Sciences **59,** 516-528.

**Fujisawa, M., Shima, Y., Nakagawa, H., Kitagawa, M., Kimbara, J., Nakano, T., Kasumi, T., and Ito, Y.** (2014). Transcriptional Regulation of Fruit Ripening by Tomato FRUITFULL Homologs and Associated MADS Box Proteins. The Plant cell **26,** 89-101.

**Furey, T.S.** (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions **13,** 840-852.

**Galway, M.E., Masucci, J.D., Lloyd, A.M., Walbot, V., Davis, R.W., and Schiefelbein, J.W.** (1994). The TTG gene is required to specify epidermal cell fate and cell patterning in the *Arabidopsis* root. Developmental biology **166,** 740-754.

**Gamboa, A., Paéz-Valencia, J., Acevedo, G.F., Vázquez-Moreno, L., and Alvarez-Buylla, R.E.** (2001). Floral Transcription Factor AGAMOUS Interacts in Vitro with a Leucine-Rich Repeat and an Acid Phosphatase Protein Complex. Biochemical and Biophysical Research Communications **288,** 1018-1026.

**Goda, H., Sawa, S., Asami, T., Fujioka, S., Shimada, Y., and Yoshida, S.** (2004). Comprehensive Comparison of Auxin-Regulated and Brassinosteroid-Regulated Genes in *Arabidopsis*. Plant Physiology **134,** 1555-1573.

**Gómez-Mena, C., and Sablowski, R.** (2008). *ARABIDOPSIS* THALIANA HOMEOBOX GENE1 Establishes the Basal Boundaries of Shoot Organs and Controls Stem Growth. The Plant cell **20,** 2059-2072.

**Gorski, J.J., Savage, K.I., Mulligan, J.M., McDade, S.S., Blayney, J.K., Ge, Z., and Harkin, D.P.** (2011). Profiling of the BRCA1 transcriptome through microarray and ChIP-chip analysis. Nucleic Acids Res **39,** 9536-9548.

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

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

**Gregis, V., Andres, F., Sessa, A., Guerra, R.F., Simonini, S., Mateos, J.L., Torti, S., Zambelli, F., Prazzoli, G.M., Bjerkan, K.N., Grini, P.E., Pavesi, G., Colombo, L., Coupland, G., and Kater, M.M.** (2013). Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in *Arabidopsis*. Genome biology **14,** R56.

**Grove, C.A., De Masi, F., Barrasa, M.I., Newburger, D.E., Alkema, M.J., Bulyk, M.L., and Walhout, A.J.M.** (2009). A Multiparameter Network Reveals Extensive Divergence between C. elegans bHLH Transcription Factors. Cell **138,** 314-327.

**Gu, Q., Ferrándiz, C., Yanofsky, M.F., and Martienssen, R.** (1998). The FRUITFULL MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. Development **125,** 1509-1517.

**Guerineau, F., Benjdia, M., and Zhou, D.X.** (2003). A jasmonate-responsive element within the A. thaliana vsp1 promoter. Journal of Experimental Botany **54,** 1153-1162.

**Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S.** (2007). Quantifying similarity between motifs. Genome biology **8,** R24-R24.

**Hai, T., and Curran, T.** (1991). Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proceedings of the National Academy of Sciences **88,** 3720-3724.

**Han, Y., and Jiao, Y.** (2015). APETALA1 establishes determinate floral meristem through regulating cytokinins homeostasis in *Arabidopsis*. Plant Signaling & Behavior **10,** e989039.

**Han, Y., Zhang, C., Yang, H., and Jiao, Y.** (2014). Cytokinin pathway mediates APETALA1 function in the establishment of determinate floral meristems in *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America **111,** 6840-6845.

**Hancock, S.P., Ghane, T., Cascio, D., Rohs, R., Di Felice, R., and Johnson, R.C.** (2013). Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. Nucleic Acids Res **41,** 6750-6760.

**Hao, Y., Oh, E., Choi, G., Liang, Z., and Wang, Z.-Y.** (2012). Interactions between HLH and bHLH Factors Modulate Light-Regulated Plant Development. Molecular Plant **5,** 688-697.

**Hartmann, U., Höhmann, S., Nettesheim, K., Wisman, E., Saedler, H., and Huijser, P.** (2000). Molecular cloning of SVP: a negative regulator of the floral transition in *Arabidopsis*. The Plant Journal **21,** 351-360.

**Heck, G.R., Perry, S.E., Nichols, K.W., and Fernandez, D.E.** (1995). AGL15, a MADS domain protein expressed in developing embryos. The Plant cell **7,** 1271-1282.

**Heisler, M.G., Ohno, C., Das, P., Sieber, P., Reddy, G.V., Long, J.A., and Meyerowitz, E.M.** (2005). Patterns of Auxin Transport and Gene Expression during Primordium Development Revealed by Live Imaging of the *Arabidopsis* Inflorescence Meristem. Current Biology **15,** 1899-1911.

**Hempel, F.D., Weigel, D., Mandel, M.A., Ditta, G., Zambryski, P.C., Feldman, L.J., and Yanofsky, M.F.** (1997). Floral determination and expression of floral regulatory genes in *Arabidopsis*. Development **124,** 3845-3853.

**Hendrich, B., and Tweedie, S.** (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends in genetics : TIG **19,** 269-277.

**Heyndrickx, K.S., de Velde, J.V., Wang, C., Weigel, D., and Vandepoele, K.** (2014). A Functional and Evolutionary Perspective on Transcription Factor Binding in *Arabidopsis* thaliana. The Plant cell **26,** 3894-3910.

**Holliday, R., and Pugh, J.E.** (1975). DNA modification mechanisms and gene activity during development. Science **187,** 226-232.

**Honma, T., and Goto, K.** (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs **409,** 525-529.

**Hou, K., Wu, W., and Gan, S.-S.** (2013). SAUR36, a SMALL AUXIN UP RNA Gene, Is Involved in the Promotion of Leaf Senescence in *Arabidopsis*. Plant Physiology **161,** 1002-1009.

**Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P.** (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Advances in bioinformatics **2008,** 420747.

**Hsieh, C.L.** (1994). Dependence of transcriptional repression on CpG methylation density. Molecular and cellular biology **14,** 5487-5494.

**Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C., Xia, S., Liu, S., Lyu, H., Ming, G.-L., Wade, H., Song, H., Qian, J., and Zhu, H.** (2013). DNA methylation presents distinct binding sites for human transcription factors. eLife **2,** e00726.

**Huala, E., Dickerman, A.W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W., Mueller, L.A., Bhattacharyya, D., Bhaya, D., Sobral, B.W., Beavis, W., Meinke, D.W., Town, C.D., Somerville, C., and Rhee, S.Y.** (2001). The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Research **29,** 102-105.

**Huang, H., Mizukami, Y., Hu, Y., and Ma, H.C.** (1993). Isolation and characterization of the binding sequences for the product of the *Arabidopsis* floral homeotic gene AGAMOUS. Nucleic acids research **21,** 4769-4776.

**Huang, H., Tudor, M., Weiss, C.A., Hu, Y., and Ma, H.** (1995). The *Arabidopsis* MADS-box gene AGL3 is widely expressed and encodes a sequence-specific DNA-binding protein. Plant molecular biology **28,** 549-567.

**Huang, H., Tudor, M., Su, T., Zhang, Y., Hu, Y., and Ma, H.C.** (1996). DNA binding properties of two *Arabidopsis* MADS domain proteins: binding consensus and dimer formation. The Plant cell **8,** 81-94.

**Hughes, T.R., and de Boer, C.G.** (2013). Mapping yeast transcriptional networks. Genetics **195,** 9-36.

**Immink, R.G., Tonaco, I.A., de Folter, S., Shchennikova, A., van Dijk, A.D., Busscher-Lange, J., Borst, J.W., and Angenent, G.C.** (2009). SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. Genome biology **10,** R24.

**Immink, R.G., Posé, D., Ferrario, S., Ott, F., Kaufmann, K., Valentim, F.L., de Folter, S., van der Wal, F., van Dijk, A.D., Schmid, M., and Angenent, G.C.** (2012). Characterization of SOC1's central role in flowering by the identification of its upstream and downstream regulators. Plant Physiol **160**.

**Ippel, H., Larsson, G., Behravan, G., Zdunek, J., Lundqvist, M., Schleucher, J., Lycksell, P.-O., and Wijmenga, S.** (1999). The solution structure of the homeodomain of the rat insulin-gene enhancer protein Isl-1. Comparison with other homeodomains1. Journal of Molecular Biology **288,** 689-703.

**Jaakola, L., Poole, M., Jones, M.O., Kämäräinen-Karppinen, T., Koskimäki, J.J., Hohtola, A., Häggman, H., Fraser, P.D., Manning, K., King, G.J., Thomson, H., and Seymour, G.B.** (2010). A SQUAMOSA MADS Box Gene Involved in the Regulation of Anthocyanin Accumulation in Bilberry Fruits. Plant Physiology **153,** 1619-1629.

**Jack, T., Fox, G.L., and Meyerowitz, E.M.** (1994). *Arabidopsis* homeotic gene APETALA3 ectopic expression: Transcriptional and posttranscriptional regulation determine floral organ identity. Cell **76,** 703-716.

**Jayaram, B., and Jain, T.** (2004). The role of water in protein-DNA recognition. Annual review of biophysics and biomolecular structure **33,** 343-361.

**Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M., and Wong, W.H.** (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data **26,** 1293-1300.

**John, S., Sabo, P.J., Thurman, R.E., Sung, M.-H., Biddie, S.C., Johnson, T.A., Hager, G.L., and Stamatoyannopoulos, J.A.** (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns **43,** 264-268.

**Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B.** (2007). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. Science **316,** 1497-1502.

**Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J.** (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature **527,** 384-388.

**Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpää, M.J., Bonke, M., Palin, K., Talukder, S., Hughes, T.R., Luscombe, N.M., Ukkonen, E., and Taipale, J.** (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome research **20,** 861-873.

**Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., Palin, K., Vaquerizas, J.M., Vincentelli, R., Luscombe, N.M., Hughes, T.R., Lemaire, P., Ukkonen, E., Kivioja, T., and Taipale, J.** (2013). DNA-binding specificities of human transcription factors. Cell **152,** 327-339.

**Kallesen, M., and Rosen, J.M.** (2001). ChIP Assay Protocol. https://www.bcm.edu/rosenlab/index.cfm?pmid=12979.

**Kanrar, S., Bhattacharya, M., Arthur, B., Courtier, J., and Smith, H.M.S.** (2008). Regulatory networks that function to specify flower meristems require the function of homeobox genes PENNYWISE and POUND-FOOLISH in *Arabidopsis*. The Plant Journal **54,** 924-937.

**Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E.** (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature **458,** 362-366.

**Kaufmann, K., Melzer, R., and Theissen, G.** (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene **347,** 183-198.

**Kaufmann, K., Pajoro, A., and Angenent, G.C.** (2010a). Regulation of transcription in plants: mechanisms controlling developmental switches **11,** 830-842.

**Kaufmann, K., Muino, J.M., Jauregui, R., Airoldi, C.A., Smaczniak, C., and Krajewski, P.** (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the *Arabidopsis* flower. PLoS Biol. **7**.

**Kaufmann, K., Muino, J.M., Osteras, M., Farinelli, L., Krajewski, P., and Angenent, G.C.** (2010b). Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP) **5,** 457-472.

**Kaufmann, K., Wellmer, F., Muiño, J.M., Ferrier, T., Wuest, S.E., Kumar, V., Serrano-Mislata, A., Madueño, F., Krajewski, P., Meyerowitz, E.M., Angenent, G.C., and Riechmann, J.L.** (2010c). Orchestration of Floral Initiation by APETALA1. Science **328,** 85-89.

**Keuskamp, D.H., Sasidharan, R., Vos, I., Peeters, A.J.M., Voesenek, L.A.C.J., and Pierik, R.** (2011). Blue-light-mediated shade avoidance requires combined auxin and brassinosteroid action in *Arabidopsis* seedlings. The Plant Journal **67,** 208-217.

**Kim, J., Kang, H., Park, J., Kim, W., Yoo, J., Lee, N., Kim, J., Yoon, T.-y., and Choi, G.** (2016). PIF1-Interacting Transcription Factors and Their Binding Sequence Elements Determine the in Vivo Targeting Sites of PIF1. The Plant cell **28,** 1388-1405.

**Kim, J.J., Lee, J.H., Kim, W., Jung, H.S., Huijser, P., and Ahn, J.H.** (2012). The microRNA156-SQUAMOSA PROMOTER BINDING PROTEIN-LIKE3 module regulates ambient temperature-responsive flowering via FLOWERING LOCUS T in *Arabidopsis*. Plant Physiol **159,** 461-478.

**Knight, B., Kubik, S., Ghosh, B., Bruzzone, M.J., Geertz, M., Martin, V., Dénervaud, N., Jacquet, P., Ozkan, B., Rougemont, J., Maerkl, S.J., Naef, F., and Shore, D.** (2014). Two distinct promoter architectures centered on dynamic nucleosomes control ribosomal protein gene transcription. Genes & Development **28,** 1695-1709.

**Kodaira, K.-S., Qin, F., Tran, L.-S.P., Maruyama, K., Kidokoro, S., Fujita, Y., Shinozaki, K., and Yamaguchi-Shinozaki, K.** (2011). *Arabidopsis* Cys2/His2 Zinc-Finger Proteins AZF1 and AZF2 Negatively Regulate Abscisic Acid-Repressive and Auxin-Inducible Genes under Abiotic Stress Conditions. Plant Physiology **157,** 742-756.

**Kong, Q., Pattanaik, S., and Feller, A.** (2012). Regulatory switch enforced by basic helix-loop-helix and ACT-domain mediated dimerizations of the maize transcription factor R. Proceedings of the National Academy of Sciences.

**Kuhn, B.M., Nodzyński, T., Errafi, S., Bucher, R., Gupta, S., Aryal, B., Dobrev, P., Bigler, L., Geisler, M., Zažímalová, E., Friml, J., and Ringli, C.** (2017). Flavonol-induced changes in PIN2 polarity and auxin transport in the *Arabidopsis* thaliana rol1-2 mutant require phosphatase activity. Scientific Reports **7**.

**Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K.I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A.J., Hoffman, M.M., Iyer, V.R., Jung, Y.L., Karmakar, S., Kellis, M., Kharchenko, P.V., Li, Q., Liu, T., Liu, X.S., Ma, L., Milosavljevic, A., Myers, R.M., Park, P.J., Pazin, M.J., Perry, M.D., Raha, D., Reddy, T.E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J.A., Tolstorukov, M.Y., White, K.P., Xi, S., Farnham, P.J., Lieb, J.D., Wold, B.J., and Snyder, M.** (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome research **22,** 1813-1831.

**Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L.** (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology **10,** 1-10.

**Lazarovici, A., Zhou, T., Shafer, A., Dantas Machado, A.C., Riley, T.R., Sandstrom, R., Sabo, P.J., Lu, Y., Rohs, R., Stamatoyannopoulos, J.A., and Bussemaker, H.J.** (2013). Probing DNA shape and methylation state on a genomic scale with DNase I. Proceedings of the National Academy of Sciences **110,** 6376-6381.

**Lécuyer, E., and Hoang, T.** (2004). SCL: From the origin of hematopoiesis to stem cells and leukemia. Experimental Hematology **32,** 11-24.

**Lee, H., Suh, S.-S., Park, E., Cho, E., Ahn, J.H., Kim, S.-G., Lee, J.S., Kwon, Y.M., and Lee, I.** (2000). The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. Genes & Development **14,** 2366-2376.

**Li, H., and Durbin, R.** (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25,** 1754-1760.

**Li, P., Wang, Y., Qian, Q., Fu, Z., Wang, M., Zeng, D., Li, B., Wang, X., and Li, J.** (2007). LAZY1 controls rice shoot gravitropism through regulating polar auxin transport. Cell Res **17,** 402-410.

**Li, Q.-F., Wang, C., Jiang, L., Li, S., Sun, S.S.M., and He, J.-X.** (2012). An Interaction Between BZR1 and DELLAs Mediates Direct Signaling Crosstalk Between Brassinosteroids and Gibberellins in *Arabidopsis*. Sci. Signal. **5,** ra72.

**Li, R., Li, Y., Kristiansen, K., and Wang, J.** (2008). SOAP: short oligonucleotide alignment program. Bioinformatics **24,** 713-714.

**Li, R., Yu, C., Li, Y., Lam, T., Yiu, S., Kristiansen, K., and Wang, J.** (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics **25,** 1966 - 1967.

**Lickwar, C.R., Mueller, F., Hanlon, S.E., McNally, J.G., and Lieb, J.D.** (2012). Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function **484,** 251-255.

**Liljegren, S.J., Ditta, G.S., Eshed, Y., Savidge, B., Bowman, J.L., and Yanofsky, M.F.** (2000). SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis* **404,** 766-770.

**Liljegren, S.J., Roeder, A.H.K., Kempin, S.A., Gremski, K., Ostergaard, L., Guimil, S., Reyes, D.K., and Yanofsky, M.F.** (2004). Control of fruit patterning in *Arabidopsis* by INDEHISCENT. Cell **116**.

**Lim, W.F., Burdach, J., Funnell, A.P.W., Pearson, R.C.M., Quinlan, K.G.R., and Crossley, M.** (2016). Directing an artificial zinc finger protein to new targets by fusion to a non-DNA-binding domain. Nucleic Acids Research **44,** 3118-3130.

**Liscum, E., Askinosie, S.K., Leuchtman, D.L., Morrow, J., Willenburg, K.T., and Coats, D.R.** (2014). Phototropism: Growing towards an Understanding of Plant Movement. The Plant cell **26,** 38-55.

**Liu, C., Xi, W., Shen, L., Tan, C., and Yu, H.** (2009). Regulation of Floral Patterning by Flowering Time Genes. Developmental Cell **16,** 711-722.

**Liu, X., Noll, D.M., Lieb, J.D., and Clarke, N.D.** (2005). DIP-chip: Rapid and accurate determination of DNA-binding specificity. Genome research **15,** 421-427.

**Liu, X., Lee, C.K., Granek, J.A., Clarke, N.D., and Lieb, J.D.** (2006). Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. Genome research **16,** 1517-1528.

**Liu, X., Yang, S., Zhao, M., Luo, M., Yu, C.-W., Chen, C.-Y., Tai, R., and Wu, K.** (2014). Transcriptional Repression by Histone Deacetylases in Plants. Molecular Plant **7,** 764-772.

**Löhr, U., and Pick, L.** (2005). Cofactor-interaction motifs and the cooption of a homeotic Hox protein into the segmentation pathway of Drosophila melanogaster. Current biology : CB **15,** 643-649.

**Lopes, E.C., Valls, E., Figueroa, M.E., Mazur, A., Meng, F.G., Chiosis, G., Laird, P.W., Schreiber-Agus, N., Greally, J.M., Prokhortchouk, E., and Melnick, A.** (2008). Kaiso contributes to DNA methylation-dependent silencing of tumor suppressor genes in colon cancer cell lines. Cancer research **68,** 7258-7263.

**Luscombe, N.M., Laskowski, R.A., and Thornton, J.M.** (2001). Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. Nucleic Acids Research **29,** 2860-2874.

**Machanick, P., and Bailey, T.L.** (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics **27,** 1696-1697.

**Maere, S., Heymans, K., and Kuiper, M.** (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. Bioinformatics **21,** 3448-3449.

**Maerkl, S.J., and Quake, S.R.** (2007). A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. Science **315,** 233-237.

**Magnani, L., Eeckhoute, J., and Lupien, M.** (2011). Pioneer factors: directing transcriptional regulators within the chromatin environment. Trends in Genetics **27,** 465-474.

**Mandel, M.A., and Yanofsky, M.F.** (1995). The *Arabidopsis* AGL8 MADS box gene is expressed in inflorescence meristems and is negatively regulated by APETALA1. The Plant cell **7,** 1763-1771.

**Marsch-Martínez, N., and de Folter, S.** (2016). Hormonal control of the development of the gynoecium. Current opinion in plant biology **29,** 104-114.

**Mathieu, J., Yant, L.J., Mürdter, F., Küttner, F., and Schmid, M.** (2009). Repression of Flowering by the miR172 Target SMZ. PLoS biology **7,** e1000148.

**Maurano, Matthew T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, John A.** (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Reports **12,** 1184-1195.

**Melzer, R., and Theißen, G.** (2009). Reconstitution of 'floral quartets' in vitro involving class B and class E floral homeotic proteins. Nucleic Acids Research **37,** 2723-2736.

**Melzer, R., Verelst, W., and Theissen, G.** (2009). The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in floral quartet-like complexes in vitro. Nucleic Acids Res **37**.

**Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A., and Beeckman, T.** (2008). Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis* thaliana **40,** 1489-1492.

**Michaels, S.D., and Amasino, R.M.** (1999). FLOWERING LOCUS C Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering. The Plant cell **11,** 949-956.

**Moreno-Risueno, M.A., Van Norman, J.M., Moreno, A., Zhang, J., Ahnert, S.E., and Benfey, P.N.** (2010). Oscillating Gene Expression Determines Competence for Periodic *Arabidopsis* Root Branching. Science **329,** 1306-1311.

**Moyroud, E., Minguet, E.G., Ott, F., Yant, L., Pose, D., Monniaux, M., Blanchet, S., Bastien, O., Thevenon, E., Weigel, D., Schmid, M., and Parcy, F.** (2011). Prediction of regulatory interactions from genome sequences using a biophysical model for the *Arabidopsis* LEAFY transcription factor. The Plant cell **23,** 1293-1306.

**Muiño, J., Kaufmann, K., van Ham, R., Angenent, G., and Krajewski, P.** (2011). ChIP-seq Analysis in R (CSAR): An R package for the statistical detection of protein-bound genomic regions. Plant Methods **7,** 11.

**Muiño, J.M., Smaczniak, C., Angenent, G.C., Kaufmann, K., and Dijk, A.D.J.** (2013). Structural determinants of DNA recognition by plant MADS-domain transcription factors. Nucleic Acids Res. **42**.

**Mukherjee, S., Berger, M.F., Jona, G., Wang, X.S., Muzzey, D., Snyder, M., Young, R.A., and Bulyk, M.L.** (2004). Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. Nature genetics **36,** 1331-1339.

**Nam, K.H., and Li, J.** (2002). BRI1/BAK1, a Receptor Kinase Pair Mediating Brassinosteroid Signaling. Cell **110,** 203-212.

**Nguyen, N.H., Kim, J.H., Kwon, J., Jeong, C.Y., Lee, W., Lee, D., Hong, S.W., and Lee, H.** (2016). Characterization of *Arabidopsis* thaliana FLAVONOL SYNTHASE 1 (FLS1) -overexpression plants in response to abiotic stress. Plant physiology and biochemistry : PPB **103,** 133-142.

**Nicol, J.W., Helt, G.A., Blanchard, S.G., Raja, A., and Loraine, A.E.** (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics **25,** 2730-2731.

**Nie, Y., Cheng, X., Chen, J., and Sun, X.** (2014). Nucleosome organization in the vicinity of transcription factor binding sites in the human genome. BMC Genomics **15,** 1-14.

**Nowak, D.E., Tian, B., and Brasier, A.R.** (2005). Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. BioTechniques **39,** 715-725.

**Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., and Wolfe, S.A.** (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell **133,** 1277-1289.

**O'Malley, R.C., Huang, S.S., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R.** (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell **165,** 1280-1292.

**Oh, E., Zhu, J.-Y., Bai, M.-Y., Arenhart, R.A., Sun, Y., and Wang, Z.-Y.** (2014). Cell elongation is regulated through a central circuit of interacting transcription factors in the *Arabidopsis* hypocotyl. eLife **3,** e03031.

**Oh, E., Kang, H., Yamaguchi, S., Park, J., Lee, D., Kamiya, Y., and Choi, G.** (2009). Genome-Wide Analysis of Genes Targeted by PHYTOCHROME INTERACTING FACTOR 3-LIKE5 during Seed Germination in *Arabidopsis*. The Plant cell **21,** 403-419.

**Outchkourov, Nikolay S., Muiño, Jose M., Kaufmann, K., van Ijcken, Wilfred F.J., Koerkamp, Marian J.G., van Leenen, D., de Graaf, P., Holstege, Frank C.P., Grosveld, Frank G., and Timmers, H.T.M.** (2013). Balancing of Histone H3K4 Methylation States by the Kdm5c/SMCX Histone Demethylase Modulates Promoter and Enhancer Function. Cell Reports **3,** 1071-1079.

**Owens, D.K., Alerding, A.B., Crosby, K.C., Bandara, A.B., Westwood, J.H., and Winkel, B.S.J.** (2008). Functional Analysis of a Predicted Flavonol Synthase Gene Family in *Arabidopsis*. Plant Physiology **147,** 1046-1061.

**Pages, H., Aboyoun, P., Gentleman, R., and DebRoy, S.** (2017). Biostrings: String objects representing biological sequences, and matching algorithms.

**Pajoro, A., Madrigal, P., Muino, J.M., Matus, J.T., Jin, J., Mecchia, M.A., Debernardi, J.M., Palatnik, J.F., Balazadeh, S., Arif, M., O'Maoileidigh, D.S., Wellmer, F., Krajewski, P., Riechmann, J.L., Angenent, G.C., and Kaufmann, K.** (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome biology **15,** R41.

**Palii, C.G., Perez-Iratxeta, C., Yao, Z., Cao, Y., Dai, F., Davison, J., Atkins, H., Allan, D., Dilworth, F.J., Gentleman, R., Tapscott, S.J., and Brand, M.** (2010). Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. The EMBO Journal **30,** 494.

**Paponov, I.A., Paponov, M., Teale, W., Menges, M., Chakrabortee, S., Murray, J.A.H., and Palme, K.** (2008). Comprehensive Transcriptome Analysis of Auxin Responses in *Arabidopsis*. Molecular Plant **1,** 321-337.

**Para, A., Li, Y., Marshall-Colón, A., Varala, K., Francoeur, N.J., Moran, T.M., Edwards, M.B., Hackley, C., Bargmann, B.O.R., Birnbaum, K.D., McCombie, W.R., Krouk, G., and Coruzzi, G.M.** (2014). Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America **111,** 10371-10376.

**Parenicová, L., de Folter, S., Kieffer, M., Horner, D.S., Favalli, C., Busscher, J., Cook, H.E., Ingram, R.M., Kater, M.M., Davies, B., Angenent, G.C., and Colombo, L.** (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. The Plant cell **15**.

**Pepke, S., Wold, B., and Mortazavi, A.** (2009). Computation for ChIP-seq and RNA-seq studies **6,** S22-S32.

**Perry, S.E., Lehti, M.D., and Fernandez, D.E.** (1999). The MADS-Domain Protein AGAMOUS-Like 15 Accumulates in Embryonic Tissues with Diverse Origins. Plant Physiology **120,** 121-130.

**Pesch, M., and Hulskamp, M.** (2009). One, two, three...models for trichome patterning in *Arabidopsis*? Current opinion in plant biology **12,** 587-592.

**Pierik, R., Djakovic-Petrovic, T., Keuskamp, D.H., de Wit, M., and Voesenek, L.A.C.J.** (2009). Auxin and Ethylene Regulate Elongation Responses to Neighbor Proximity Signals Independent of Gibberellin and DELLA Proteins in *Arabidopsis*. Plant Physiology **149,** 1701-1712.

**Pollastri, S., and Tattini, M.** (2011). Flavonols: old compounds for old roles. Annals of Botany **108,** 1225-1233.

**Pollock, R., and Treisman, R.** (1990). A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res **18,** 6197-6204.

**Procko, C., Burko, Y., Jaillais, Y., Ljung, K., Long, J.A., and Chory, J.** (2016). The epidermis coordinates auxin-induced stem growth in response to shade. Genes & Development **30,** 1529-1541.

**Prokhortchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E.** (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes & Development **15,** 1613-1618.

**Rajeev, L., Luning, E.G., Dehal, P.S., Price, M.N., Arkin, A.P., and Mukhopadhyay, A.** (2011). Systematic mapping of two component response regulators to gene targets in a model sulfate reducing bacterium. Genome biology **12,** R99.

**Rashotte, A.M., Mason, M.G., Hutchison, C.E., Ferreira, F.J., Schaller, G.E., and Kieber, J.J.** (2006). A subset of *Arabidopsis* AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. Proceedings of the National Academy of Sciences **103,** 11081-11085.

**Ratcliffe, O.J., Nadzan, G.C., Reuber, T.L., and Riechmann, J.L.** (2001). Regulation of Flowering in *Arabidopsis* by an FLCHomologue. Plant Physiology **126,** 122-132.

**Redman, J.C., Haas, B.J., Tanimoto, G., and Town, C.D.** (2004). Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. The Plant Journal **38,** 545-561.

**Reinke, A.W., Baek, J., Ashenberg, O., and Keating, A.E.** (2013). Networks of bZIP protein-protein interactions diversified over a billion years of evolution. Science **340,** 730-734.

**Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L., Wilson, C.J., Bell, S.P., and Young, R.A.** (2000). Genome-Wide Location and Function of DNA Binding Proteins. Science **290,** 2306-2309.

**Ren, H., and Gray, William M.** (2015). SAUR Proteins as Effectors of Hormonal and Environmental Signals in Plant Growth. Molecular Plant **8,** 1153-1164.

**Rezsohazy, R., Saurin, A.J., Maurel-Zaffran, C., and Graba, Y.** (2015). Cellular and molecular insights into Hox protein action. Development **142,** 1212-1227.

**Rhee, D.Y., Cho, D.-Y., Zhai, B., Slattery, M., Ma, L., Mintseris, J., Wong, C.Y., White, K.P., Celniker, S.E., Przytycka, T.M., Gygi, S.P., Obar, R.A., and Artavanis-Tsakonas, S.** (2014). Transcription Factor Networks in Drosophila melanogaster. Cell reports **8,** 2031-2043.

**Rhee, H.S., and Pugh, B.F.** (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell **147,** 1408-1419.

**Riechmann, J.L., Wang, M., and Meyerowitz, E.M.** (1996). DNA-Binding Properties of *Arabidopsis* MADS Domain Homeotic Proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. Nucleic Acids Research **24,** 3134-3141.

**Riggs, A.D.** (1975). X inactivation, differentiation, and DNA methylation. Cytogenet Genome Res **14,** 9-25.

**Ripoll, J., Bailey, L.J., Mai, Q.A., Wu, S.L., Hon, C.T., Chapman, E.J., Ditta, G.S., Estelle, M., and Yanofsky, M.F.** (2015). microRNA regulation of fruit growth. Nature Plants **1,** 15036.

**Ripoll, J.J., Roeder, A.H.K., Ditta, G.S., and Yanofsky, M.F.** (2011). A novel role for the floral homeotic gene APETALA2 during *Arabidopsis* fruit development. Development **138,** 5167-5176.

**Rishi, V., Bhattacharya, P., Chatterjee, R., Rozenberg, J., Zhao, J., Glass, K., Fitzgerald, P., and Vinson, C.** (2010). CpG methylation of half-CRE sequences creates C/EBPα binding sites that activate some tissue-specific genes. Proceedings of the National Academy of Sciences **107,** 20311-20316.

**Roeder, A.H.K., and Yanofsky, M.F.** (2006). Fruit Development in *Arabidopsis*. The *Arabidopsis* Book / American Society of Plant Biologists **4,** e0075.

**Roeder, A.H.K., Ferrándiz, C., and Yanofsky, M.F.** (2003). The Role of the REPLUMLESS Homeodomain Protein in Patterning the *Arabidopsis* Fruit. Current Biology **13,** 1630-1635.

**Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S.** (2010). Origins of specificity in protein-DNA recognition. Annual Review of Biochemistry **79,** 233-269.

**Roig-Villanova, I., Bou-Torrent, J., Galstyan, A., Carretero-Paulet, L., Portolés, S., Rodríguez-Concepción, M., and Martínez-García, J.F.** (2007). Interaction of shade avoidance and auxin responses: a role for two novel atypical bHLH proteins. The EMBO Journal **26,** 4756-4767.

**Roychoudhry, S., Del Bianco, M., Kieffer, M., and Kepinski, S.** (2013). Auxin Controls Gravitropic Setpoint Angle in Higher Plant Lateral Branches. Current Biology **23,** 1497-1504.

**Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., and Gerstein, M.B.** (2009). PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls **27,** 66-75.

**Rubio-Somoza, I., Zhou, C.M., Confraria, A., Martinho, C., von Born, P., Baena-Gonzalez, E., Wang, J.W., and Weigel, D.** (2014). Temporal control of leaf complexity by miRNA-regulated licensing of protein complexes. Current biology : CB **24,** 2714-2719.

**Rutjens, B., Bao, D., Van Eck-Stouten, E., Brand, M., Smeekens, S., and Proveniers, M.** (2009). Shoot apical meristem function in *Arabidopsis* requires the combined activities of three BEL1-like homeodomain proteins. The Plant Journal **58,** 641-654.

**Salter, M.G., Franklin, K.A., and Whitelam, G.C.** (2003). Gating of the rapid shade-avoidance response by the circadian clock in plants. Nature **426,** 680-683.

**Samson, F., Brunaud, V., Balzergue, S., Dubreucq, B., Lepiniec, L., Pelletier, G., Caboche, M., and Lecharny, A.** (2002). FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis* thaliana T-DNA transformants. Nucleic Acids Research **30,** 94-97.

**Sanda, T., Lawton, Lee N., Barrasa, M.I., Fan, Zi P., Kohlhammer, H., Gutierrez, A., Ma, W., Tatarek, J., Ahn, Y., Kelliher, Michelle A., Jamieson, Catriona H.M., Staudt, Louis M., Young, Richard A., and Look, A.T.** (2012). Core Transcriptional Regulatory Circuit Controlled by the TAL1 Complex in Human T Cell Acute Lymphoblastic Leukemia. Cancer Cell **22,** 209-221.

**Sayou, C., Monniaux, M., Nanao, M.H., Moyroud, E., Brockington, S.F., Thévenon, E., Chahtane, H., Warthmann, N., Melkonian, M., Zhang, Y., Wong, G.K., Weigel, D., Parcy, F., and Dumas, R.** (2014). A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. Science (New York, N.Y.) **343,** 645-648.

**Schaffner, W.** (1988). A hit-and-run mechanism for transcriptional activation? **336,** 427-428.

**Schiessl, K., Muiño, J.M., and Sablowski, R.** (2014). *Arabidopsis* JAGGED links floral organ patterning to tissue growth by repressing Kip-related cell cycle inhibitors. Proceedings of the National Academy of Sciences **111,** 2830-2835.

**Schmid, M., Uhlenhaut, N.H., Godard, F., Demar, M., Bressan, R., Weigel, D., and Lohmann, J.U.** (2003). Dissection of floral induction pathways using global expression analysis. Development **130,** 6001-6012.

**Schutze, K., Harter, K., and Chaban, C.** (2008). Post-translational regulation of plant bZIP factors. Trends Plant Sci **13,** 247-255.

**Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H.** (1990). Genetic Control of Flower Development by Homeotic Genes in Antirrhinum majus. Science **250,** 931-936.

**Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P.J., Hansen, R., Tetens, F., Lonnig, W.E., Saedler, H., and Sommer, H.** (1992). Characterization of the Antirrhinum floral homeotic MADS-box gene deficiens: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. Embo j **11,** 251-263.

**Seeman, N.C., Rosenberg, J.M., and Rich, A.** (1976). Sequence-specific recognition of double helical nucleic acids by proteins. Proceedings of the National Academy of Sciences **73,** 804-808.

**Sehra, B., and Franks, R.G.** (2015). Auxin and cytokinin act during gynoecial patterning and the development of ovules from the meristematic medial domain. WIREs Dev Biol **4,** 555-571.

**Seo, E., Lee, H., Jeon, J., Park, H., Kim, J., Noh, Y.-S., and Lee, I.** (2009). Crosstalk between Cold Response and Flowering in *Arabidopsis* Is Mediated through the Flowering-Time Gene SOC1 and Its Upstream Negative Regulator FLC. The Plant cell **21,** 3185-3197.

**Shikata, M., Koyama, T., Mitsuda, N., and Ohme-Takagi, M.** (2009). *Arabidopsis* SBP-box genes SPL10, SPL11 and SPL2 control morphological change in association with shoot maturation in the reproductive phase. Plant & cell physiology **50,** 2133-2145.

**Shiraishi, H., Okada, K., and Shimura, Y.** (1993). Nucleotide sequences recognized by the AGAMOUS MADS domain of *Arabidopsis* thaliana in vitro. The Plant journal : for cell and molecular biology **4,** 385-398.

**Shumaker-Parry, J.S., Aebersold, R., and Campbell, C.T.** (2004). Parallel, Quantitative Measurement of Protein Binding to a 120-Element Double-Stranded DNA Array in Real Time Using Surface Plasmon Resonance Microscopy. Analytical Chemistry **76,** 2071-2082.

**Simonini, S., Roig-Villanova, I., Gregis, V., Colombo, B., Colombo, L., and Kater, M.M.** (2012). BASIC PENTACYSTEINE Proteins Mediate MADS Domain Complex Binding to the DNA for Tissue-Specific Expression of Target Genes in *Arabidopsis*. The Plant cell **24,** 4163-4172.

**Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, Harmen J., and Mann, Richard S.** (2011). Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins. Cell **147,** 1270-1282.

**Smaczniak, C., Immink, R.G.H., Angenent, G.C., and Kaufmann, K.** (2012a). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development **139,** 3081-3098.

**Smaczniak, C., Immink, R.G., Muino, J.M., Blanvillain, R., Busscher, M., and Busscher-Lange, J.** (2012b). Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. Proc Natl Acad Sci U S A. **109**.

**Smaczniak, C., Li, N., Boeren, S., America, T., van Dongen, W., Goerdayal, S.S., de Vries, S., Angenent, G.C., and Kaufmann, K.** (2012c). Proteomics-based identification of low-abundance signaling and regulatory protein complexes in native plant tissues **7,** 2144-2158.

**Smyth, D.R., Bowman, J.L., and Meyerowitz, E.M.** (1990). Early flower development in *Arabidopsis*. The Plant cell **2,** 755-767.

**Spartz, A.K., Ren, H., Park, M.Y., Grandt, K.N., Lee, S.H., Murphy, A.S., Sussman, M.R., Overvoorde, P.J., and Gray, W.M.** (2014). SAUR Inhibition of PP2C-D Phosphatases Activates Plasma Membrane H+-ATPases to Promote Cell Expansion in *Arabidopsis*. The Plant cell **26,** 2129-2142.

**Spartz, A.K., Lee, S.H., Wenger, J.P., Gonzalez, N., Itoh, H., Inzé, D., Peer, W.A., Murphy, A.S., Overvoorde, P.J., and Gray, W.M.** (2012). The SAUR19 subfamily of SMALL AUXIN UP RNA genes promote cell expansion. The Plant journal : for cell and molecular biology **70,** 978-990.

**Spitz, F., and Furlong, E.E.M.** (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet **13,** 613-626.

**Sridhar, V.V., Surendrarao, A., and Liu, Z.** (2006). APETALA1 and SEPALLATA3 interact with SEUSS to mediate transcription repression during flower development. Development **133,** 3159-3166.

**Stamm, P., and Kumar, P.** (2013). Auxin and gibberellin responsive *Arabidopsis* SMALL AUXIN UP RNA36 regulates hypocotyl elongation in the light. Plant Cell Rep **32,** 759-769.

**Stark, R., and Brown, G.** (2013). DiffBind: differential binding analysis of ChIP-Seq peak data.

**Stella, S., Cascio, D., and Johnson, R.C.** (2010). The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. Genes & Development **24,** 814-826.

**Stormo, G.D., and Zhao, Y.** (2010). Determining the specificity of protein–DNA interactions **11,** 751-760.

**Sun, N., Wang, J., Gao, Z., Dong, J., He, H., Terzaghi, W., Wei, N., Deng, X.W., and Chen, H.** (2016). *Arabidopsis* SAURs are critical for differential light regulation of the development of various organs. Proceedings of the National Academy of Sciences **113,** 6071-6076.

**Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B., and Zhao, Y.** (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell **146,** 1016-1028.

**Tanay, A.** (2006). Extensive low-affinity transcriptional interactions in the yeast genome. Genome research **16,** 962-972.

**Tang, W., and Perry, S.E.** (2003). Binding Site Selection for the Plant MADS Domain Protein AGL15: AN IN VITRO AND IN VIVO STUDY. Journal of Biological Chemistry **278,** 28154-28159.

**Tao, Z., Shen, L., Liu, C., Liu, L., Yan, Y., and Yu, H.** (2012). Genome-wide identification of SOC1 and SVP targets during the floral transition in *Arabidopsis*. The Plant journal : for cell and molecular biology **70,** 549-561.

**Tapia-López, R., García-Ponce, B., Dubrovsky, J.G., Garay-Arroyo, A., Pérez-Ruíz, R.V., Kim, S.-H., Acevedo, F., Pelaz, S., and Alvarez-Buylla, E.R.** (2008). An AGAMOUS-Related MADS-Box Gene, XAL1 (AGL12), Regulates Root Meristem Cell Proliferation and Flowering Transition in *Arabidopsis*. Plant Physiology **146,** 1182-1192.

**Tate, P.H., and Bird, A.P.** (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Current opinion in genetics & development **3,** 226-231.

**Teper-Bamnolker, P., and Samach, A.** (2005). The Flowering Integrator FT Regulates SEPALLATA3 and FRUITFULL Accumulation in *Arabidopsis* Leaves. The Plant cell **17,** 2661-2675.

**Theissen, G., and Saedler, H.** (2001). Plant biology: Floral quartets. Nature **409,** 469-471.

**Tippin, D.B., Ramakrishnan, B., and Sundaralingam, M.** (1997). Methylation of the Z-DNA decamer d(GC)5 potentiates the formation of A-DNA: crystal structure of d(Gm5CGm5CGCGCGC)1. Journal of Molecular Biology **270,** 247-258.

**Tissier, A.F., Marillonnet, S., Klimyuk, V., Patel, K., Torres, M.A., Murphy, G., and Jones, J.D.G.** (1999). Multiple Independent Defective Suppressor-mutator Transposon Insertions in *Arabidopsis*: A Tool for Functional Genomics. The Plant cell **11,** 1841-1852.

**Torti, S., and Fornara, F.** (2012). AGL24 acts in concert with SOC1 and FUL during *Arabidopsis* floral transition. Plant Signaling and Behavior **7**.

**Ulmasov, T., Murfett, J., Hagen, G., and Guilfoyle, T.J.** (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. The Plant cell **9,** 1963-1971.

**Urbanus, S., de Folter, S., Shchennikova, A., Kaufmann, K., Immink, R., and Angenent, G.** (2009). In planta localisation patterns of MADS domain proteins during floral development in *Arabidopsis* thaliana. BMC Plant Biol **9**.

**Van Gelderen, K., Van Rongen, M., Liu, A., Otten, A., and Offringa, R.** (2016). An INDEHISCENT-Controlled Auxin Response Specifies the Separation Layer in Early *Arabidopsis* Fruit. Mol Plant **9,** 857-869.

**Van Mourik, H., Muiño, J.M., Pajoro, A., Angenent, G.C., and Kaufmann, K.** (2015). Characterization of in vivo DNA-binding events of plant transcription factors by chIP-seq: Experimental protocol and computational analysis. In Methods in Molecular Biology, pp. 93-121.

**Van Verk, M.C., Hickman, R., Pieterse, C.M.J., and Van Wees, S.C.M.** (2013). RNA-Seq: revelation of the messengers. Trends in Plant Science **18,** 175-179.

**Vokes, S.A., Ji, H., Wong, W.H., and McMahon, A.P.** (2008). A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. Genes Dev **22,** 2651-2663.

**Völkel, S., Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G.** (2015). Zinc Finger Independent Genome-Wide Binding of Sp2 Potentiates Recruitment of Histone-Fold Protein Nf-y Distinguishing It from Sp1 and Sp3. PLoS Genetics **11,** e1005102.

**Vuzman, D., and Levy, Y.** (2010). DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. Proceedings of the National Academy of Sciences **107,** 21004-21009.

**Wade, P.A.** (2001). Methyl CpG-binding proteins and transcriptional repression. BioEssays : news and reviews in molecular, cellular and developmental biology **23,** 1131-1137.

**Walcher, C.L., and Nemhauser, J.L.** (2012). Bipartite Promoter Element Required for Auxin Response. Plant Physiology **158,** 273-282.

**Walker, A.R., Davison, P.A., Bolognesi-Winfield, A.C., James, C.M., Srinivasan, N., Blundell, T.L., Esch, J.J., Marks, M.D., and Gray, J.C.** (1999). The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. The Plant cell **11,** 1337-1350.

**Wang, J.W., Czech, B., and Weigel, D.** (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis* thaliana. Cell **138,** 738-749.

**Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F., and Meyerowitz, E.M.** (1992). LEAFY controls floral meristem identity in *Arabidopsis*. Cell **69,** 843-859.

**Werner, T., Motyka, V., Laucou, V., Smets, R., Van Onckelen, H., and Schmülling, T.** (2003). Cytokinin-Deficient Transgenic *Arabidopsis* Plants Show Multiple Developmental Alterations Indicating Opposite Functions of Cytokinins in the Regulation of Shoot and Root Meristem Activity. The Plant cell **15,** 2532-2550.

**Whitfield, T.W., Wang, J., Collins, P.J., Partridge, E.C., Aldred, S.F., Trinklein, N.D., Myers, R.M., and Weng, Z.** (2012). Functional analysis of transcription factor binding sites in human promoters. Genome biology **13,** R50.

**Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schütte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E., Pimanda, J.E., de Bruijn, M.F.T.R., and Göttgens, B.** (2010). Combinatorial Transcriptional Control In Blood Stem/Progenitor Cells: Genome-wide Analysis of Ten Major Transcriptional Regulators. Cell Stem Cell **7,** 532-544.

**Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J.** (2007). An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. PLOS ONE **2,** e718.

**Wu, M.-F., Sang, Y., Bezhani, S., Yamaguchi, N., Han, S.-K., Li, Z., Su, Y., Slewinski, T.L., and Wagner, D.** (2012). SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors. Proceedings of the National Academy of Sciences **109,** 3576-3581.

**Yamaguchi, A., Wu, M.-F., Yang, L., Wu, G., Poethig, R.S., and Wagner, D.** (2009). The microRNA regulated SBP-box transcription factor SPL3 is a direct upstream activator of LEAFY, FRUITFULL, and APETALA1. Developmental cell **17,** 268-278.

**Yan, W., Chen, D., and Kaufmann, K.** (2016). Molecular mechanisms of floral organ specification by MADS domain proteins. Current opinion in plant biology **29,** 154-162.

**Yanofsky, M.F., Ma, H., Bowman, J.L., Drews, G.N., Feldmann, K.A., and Meyerowitz, E.M.** (1990). The protein encoded by the *Arabidopsis* homeotic gene agamous resembles transcription factors **346,** 35-39.

**Yazaki, J., Galli, M., Kim, A.Y., Nito, K., Aleman, F., Chang, K.N., Carvunis, A.-R., Quan, R., Nguyen, H., Song, L., Alvarez, J.M., Huang, S.-s.C., Chen, H., Ramachandran, N., Altmann, S., Gutiérrez, R.A., Hill, D.E., Schroeder, J.I., Chory,**

**J., LaBaer, J., Vidal, M., Braun, P., and Ecker, J.R.** (2016). Mapping transcription factor interactome networks using HaloTag protein arrays. Proceedings of the National Academy of Sciences **113,** E4238-E4247.

**Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., Watanabe-Takahashi, A., Inoue, E., and Saito, K.** (2008). Comprehensive Flavonol Profiling and Transcriptome Coexpression Analysis Leading to Decoding Gene–Metabolite Correlations in *Arabidopsis*. The Plant cell **20,** 2160-2176.

**Yu, X., Lin, J., Masuda, T., Esumi, N., Zack, D.J., and Qian, J.** (2006). Genome-wide prediction and characterization of interactions between transcription factors in Saccharomyces cerevisiae. Nucleic Acids Res **34,** 917-927.

**Yu, X., Li, L., Zola, J., Aluru, M., Ye, H., Foudree, A., Guo, H., Anderson, S., Aluru, S., Liu, P., Rodermel, S., and Yin, Y.** (2011). A brassinosteroid transcriptional network revealed by genome-wide identification of BESI target genes in *Arabidopsis* thaliana. The Plant Journal **65,** 634-646.

**Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., Shen, Y., Pervouchine, D.D., Djebali, S., Thurman, R.E., Kaul, R., Rynes, E., Kirilusha, A., Marinov, G.K., Williams, B.A., Trout, D., Amrhein, H., Fisher-Aylor, K., Antoshechkin, I., DeSalvo, G., See, L.H., Fastuca, M., Drenkow, J., Zaleski, C., Dobin, A., Prieto, P., Lagarde, J., Bussotti, G., Tanzer, A., Denas, O., Li, K., Bender, M.A., Zhang, M., Byron, R., Groudine, M.T., McCleary, D., Pham, L., Ye, Z., Kuan, S., Edsall, L., Wu, Y.C., Rasmussen, M.D., Bansal, M.S., Kellis, M., Keller, C.A., Morrissey, C.S., Mishra, T., Jain, D., Dogan, N., Harris, R.S., Cayting, P., Kawli, T., Boyle, A.P., Euskirchen, G., Kundaje, A., Lin, S., Lin, Y., Jansen, C., Malladi, V.S., Cline, M.S., Erickson, D.T., Kirkup, V.M., Learned, K., Sloan, C.A., Rosenbloom, K.R., Lacerda de Sousa, B., Beal, K., Pignatelli, M., Flicek, P., Lian, J., Kahveci, T., Lee, D., Kent, W.J., Ramalho Santos, M., Herrero, J., Notredame, C., Johnson, A., Vong, S., Lee, K., Bates, D., Neri, F., Diegel, M., Canfield, T., Sabo, P.J., Wilken, M.S., Reh, T.A., Giste, E., Shafer, A., Kutyavin, T., Haugen, E., Dunn, D., Reynolds, A.P., Neph, S., Humbert, R., Hansen, R.S., De Bruijn, M., Selleri, L., Rudensky, A., Josefowicz, S., Samstein, R., Eichler, E.E., Orkin, S.H., Levasseur, D., Papayannopoulou, T., Chang, K.H., Skoultchi, A., Gosh, S., Disteche, C., Treuting, P., Wang, Y., Weiss, M.J., Blobel, G.A., Cao, X., Zhong, S., Wang, T., Good, P.J., Lowdon, R.F., Adams, L.B., Zhou, X.Q., Pazin, M.J., Feingold, E.A., Wold, B., Taylor, J., Mortazavi, A., Weissman, S.M., Stamatoyannopoulos, J.A., Snyder, M.P., Guigo, R., Gingeras, T.R., Gilbert, D.M., Hardison, R.C., Beer, M.A., and Ren, B.** (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature **515,** 355-364.

**Zaret, K.S., and Carroll, J.S.** (2011). Pioneer transcription factors: establishing competence for gene expression. Genes & Development **25,** 2227-2241.

**Zeng, P.Y., Vakoc, C.R., Chen, Z.C., Blobel, G.A., and Berger, S.L.** (2006). In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. BioTechniques **41,** 694, 696, 698.

**Zhang, C., Cao, L., Rong, L., An, Z., Zhou, W., Ma, J., Shen, W.-H., Zhu, Y., and Dong, A.** (2015). The chromatin-remodeling factor AtINO80 plays crucial roles in genome stability maintenance and in plant development. The Plant Journal **82,** 655-668.

**Zhang, Y., Liu, T., Meyer, C., Eeckhoute, J., Johnson, D., Bernstein, B., Nusbaum, C., Myers, R., Brown, M., Li, W., and Liu, X.S.** (2008). Model-based Analysis of ChIP-Seq (MACS). Genome biology **9,** R137.

**Zheng, Y., Ren, N., Wang, H., Stromberg, A.J., and Perry, S.E.** (2009). Global Identification of Targets of the *Arabidopsis* MADS Domain Protein AGAMOUS-Like15. The Plant cell **21,** 2563-2577.

# Summary

Encrypted in the DNA lays most information needed for the development of an organism. The transcription of this information into precise patterns of gene activity results in the development of different cell types, organs, and developmental structures. Moreover, transcriptional regulation enables an organism to respond to changing environmental conditions. Essential for the regulation of transcription are DNA-binding transcription factors (TFs). TFs bind the DNA in a sequence-specific fashion. Upon binding of a TF to its DNA binding site, TFs typically activate or repress the transcription of nearby genes. To better understand transcriptional regulation it is essential to study DNA binding specificity of TFs.

In the last decades, technological advances allowed the development of high-throughput methods to study protein-DNA interactions. Traditional *in vitro* methods study one or a few interactions, while new high-throughput methods can determine TF specificity by measuring relative DNA-binding affinities against a large collection or even all possible binding sites. Several high-throughput techniques to study TF-DNA interactions are discussed in Chapter 1 of this thesis. These new technologies and methods resulted in a fast growing number of studies on DNA binding specificities of TFs, expanding the knowledge about TF specificity. A review on the current knowledge of TF DNA binding specificity is described in Chapter 1.

One aspect that influences DNA binding of TFs are differences in ability to form protein-protein interactions. The aim of this thesis was to study the role of protein-protein interactions in determining DNA binding specificity of a developmental regulatory MADS domain TF in *Arabidopsis thaliana*. While the members of the MADS-box protein family have many, diverse *in vivo* functions, all members bind *in vitro* to a 10-bp motif called the CArG-box. Moreover, studies demonstrated that closely related MADS proteins are expressed in the same cells, therefore encountering the same DNA accessibility and DNA methylation patterns, but bind different *in vivo* targets. Interestingly, MADS domain proteins bind DNA obligatorily as homo- and heterodimers and the interactions between MADS domain proteins are highly protein specific. Hence, MADS domain proteins are a perfect model system to study the influence of intra-family protein interactions on DNA binding specificity.

To study the influence of protein-protein interactions on DNA-binding specificity this work focusses on one specific MADS domain protein, FRUITFULL (FUL). FUL is expressed at two stages during flower development and, in both stages FUL has highly diverse functions. In **Chapter 2** we demonstrate using RNA-seq that FUL regulates different sets of target genes in the two stages. Moreover, using ChIP-seq we show that FUL genomic DNA binding is partly tissue-specific. These tissue-specifically bound and regulated genes are in line with the known dual functions of FUL during development. Interestingly, using protein complex immunoprecipitation for the two studied tissues/stages we show that the interactions of FUL with other MADS domain proteins are also tissue-specific. To determine whether the tissue-specific *in vivo* binding pattern are due to differences in DNA binding specificity of the FUL-MADS dimers, we studied the DNA binding specificities of the different protein complexes using SELEX-seq. The SELEX-seq results show that although all tested dimers preferably bind the canonical binding motif of MADS domain proteins, different dimers have different preferences for nucleotides within and surrounding the canonical binding site. Hence, different MADS domain dimers have different *in vitro* DNA binding specificities. By mapping the SELEX-seq affinities to the genome we were able to compare these results with *in vivo* tissue-specific ChIP-seq data. This analysis revealed a strong correlation between tissue-specific dimer affinities and tissue-specific genomic binding sites of FUL. Hence, we show that the choice of MADS dimerization partner influences DNA binding specificity, highlighting the role of intra-family protein interactions in defining DNA binding specificity.

To allow other researchers to determine genome-wide DNA binding of TFs **Chapter 3** provides a step-by-step guide for ChIP-seq experiments and computational analysis. The protocol is designed for wet-lab biologists to perform ChIP-seq experiments and analyse their own ChIP-seq data.

Using the genome-wide DNA binding patterns determined by ChIP-seq, **Chapter 4** and **Chapter 5** take a more detailed look at some of the genes directly bound by FUL. In **Chapter 4**, we demonstrate a connection between developmentally and environmentally regulated growth programs. We studied a gene directly bound by FUL in pistil tissue, *SMALL AUXIN UPREGULATED RNA 10 (SAUR10)*. *SAUR10* expression is regulated by FUL in multiple tissues, among others cauline leaves, stems, and branches. The results show that the expression of *SAUR10* at the abaxial side of branches is influenced by a combination of environmental and developmental regulated growth

programs: hormones, light conditions, and FUL binding. This spatial regulation possibly affects the angle between the side branches and the main inflorescence stem. Additionally, we discuss several other FUL target genes involved in hormone pathways and light conditions.

Chapter 5 focusses on the putative direct targets of FUL in IM tissue. Among the putative direct targets two genes involved in flavonoid synthesis were identified, *FLAVONOID SYNTHESE 1* (*FLS1*) and *UDP-GLUCOSYL TRANSFERASE 78D3* (*UGT78D3*). Interestingly, similar to the *ful-7* mutant, the *fls1* mutant is late flowering. Moreover, expression data exposed an increased gene expression for both *FLS1* and *UGT78D3* in developing meristems and showed *FLS1* expression to be influenced by light conditions. We report the first link between the MADS domain protein FUL and flavonoid synthesis in *Arabidopsis*. Moreover, our results indicate a possible link between flavonoids and flowering time.

In Chapter 6 I discuss the findings of this thesis and make suggestions for further research. Taken together, the work in this thesis shows that intra-family protein interactions can influence DNA-binding specificity of a protein. Thereby these protein-protein interactions can influence genome-wide binding patterns and, as a result, the function of a protein. Moreover, by studying several putative direct targets of FUL in more detail, we demonstrated a connection between development and environment in growth-regulated programs. Interestingly, the FUL target *SAUR10* is repressed by FUL in several tissues, including cauline leaves, inflorescence stems, and branches. However, no influence of FUL on *SAUR10* expression could be detected in the pistil. So, despite the binding of FUL to the promotor of *SAUR10* in the pistil, this binding does not result in gene regulation. This finding reflects the complex relation between TF occupancy and gene regulation, further research is needed to better understand this relation. Moreover, besides MADS domain protein interactions, we found FUL to interact with several proteins of other families. The role of these cross-family protein interactions in cooperative gene regulation is not fully understood and will be an important research topic in the coming years.

# Acknowledgements

Although I am the only one getting the title of Docter, I could not have finished this PhD alone. I am thankful that I got the opportunity to meet many incredible people during my time as a PhD candidate. I want to thank all those people for being part of my PhD journey! This part of the thesis I want to use to write to people I would like to especially thank.

My promoter and daily supervisor Gerco. Your door is always open and I greatly enjoyed our Friday morning discussions. I am thankful that you made time in your busy schedule to take over the daily supervision when Kerstin moved to Germany. Thank you for all your help, advise, and the scientific discussions. And thank you for showing me that being a great group leader is more than being a great scientist.

Kerstin, my supervisor and co-promoter, you gave me the opportunity to start my PhD and helped me along the way. Thank you for all the advises and help you gave me! I was happy to be able to also be part of your journey. Not only did you set up three different labs and started your own research group in Germany, you also became mother of a great little boy, Lucas. I want to thank you for all your help and encouragements.

All the people from our "Flower Power" group. My paranymph, Alice, you are one of the strongest persons I know. I learned so much from you, not only in the lab but also as a person. Thank you for all your encouraging words and all the talks and all the laughs we sheared together! I am so happy to see that you and Luigi are starting a family, the world needs more people like you guys! Cezary, although I still cannot not pronounce your name correctly, I loved working together with you! When I started at PDS we were an inseparable duo in the lab, it was great having you as my lab table buddy. Thank you for learning me so many things in the lab and for all the fun we had. I wish you and Aline much joy and love! Soon you will also welcome your first child, I know you will be great parents! Sweet Suzanne, I miss our walks around the pond behind the Lumen building. Thank you for all the walks we walked, all the talks we had, all the tea we shared. Thank you for sharing this PhD journey with me, especially during the last stretch of our PhDs. And of course, Marco. Whether I needed help in the lab, had a question or came for a friendly chat, you always had time for me. Thank you.

The PDS group: I could not have joined a better group. Besides creating a great scientific environment, I also enjoyed all the fun we had. Thanks to all the technicians in the lab that allowed the lab to function so smoothly and for making sure that all materials are always in stock: Marco, Michiel, Tjitske, Mieke, Froukje, Jan, and Jaqueline. My desk-partner and paranymph, Sam. It was great having you as my desk neighbor. Although I usually talk a lot, the two of us didn't need many words to make each other laugh. Thank you for joining me in transforming the patio into a butterfly garden. Leonie, we started our PhDs in the same week. During a difficult period in our lives, both for different reasons, we often sat at the little bridge over the pond to talk. Thank you for your listening ear when I needed it so much. And my Indian friend, Suraj, you are the initiator of many activities, thank you all the drinks, dinners, and game nights! And thank you for always making everyone smile with your thoughtfulness. Marian, my fellow FRUITFULL-girl. You are a great scientist and an amazing supervisor. It was great to see how you supervised your students with so much patience. Thank you for all your help and advice. Anneke, you are always happy and singing along with the radio in the lab. Good luck with your Veni and much love to your expanding family! And thank you to all the other PhD students that brighten up the open office workspace and the lab: Jenny, Violeta, Manjunath, Mengfan, Baojian, Rufang, Vera, and Han. Thanks to all the senior scientist for their input and the scientific discussions: Kim, Ruud, Richard, Steven, and Rumyana. Tjitske and Martijn for organizing the unforgettable group outings. Mieke for arranging the Sinterklaasborrel. Hana, whatever I asked you, you had it finished in record speed. Michiel and Marco for being the filming crew and producers of so many great PhD movies. And to all other staff, postdocs, PhDs, guests, and students from the PDS group for creating a wonderful and fun environment. And for making the countless BBQs, drinks, and outings so enjoyable.

Everyone from PRI-biosciences, for the inspiring talks and input during the weekly PRI-meeting. In particular everyone from the bioinformatics department for their help, input, and problem solving. And to Elio and Bas for their help with sequencing of my samples.

I also want to thank the people in the 'Plant Developmental Biology' group in Germany, who were my far-a-way colleagues. In particular Johanna and Dijun. Johanna, thank you for all the work you did on phenotyping of the several mutant lines. Dijun, thank you for your help with the comparative ChIP-seq analysis.

Thanks to Jose for the help with the ChIP-seq and SELEX-seq bioinformatics. Thank you for your advice on the ChIP-seq data analysis and all the time you put in analyzing the SELEX-seq data. I greatly enjoyed my stay in Germany where you helped me with the analysis of my data.

Thanks to all the co-authors of the different chapters in this thesis and of my other research papers. Thank you all for your help, input, and ideas.

My friends who I met in Wageningen. The Goldwasser group, Mathijs, Sander, Tobie, Reinier, Tom, Geert, Jens, Sjoerd, Frank, Patrick, and Christa for all the BBQs and game nights! My forever neighbor, Anne-Ruth, for the countless cups of tea we shared and for always having my back. Thanks to all the people from frisbee club WAF, for all the nice trainings, games, and tournaments. Especially my frisbee-friends Djoeke en Henrieke with whom I shared most of my frisbee adventures.

Thanks to my longtime friends. Thérèse, at high school we dreamed of very different lives. And although we chose very different professions, it seems that our goals in life are becoming more and more the same. I am happy we are living closer together now and look forward to many more cycling days with you! And my longtime friend, Renny, we met at kindergarten and during our lives you always seem to follow me: from Woerden to Utrecht, from Utrecht to Wageningen, I am waiting for you to join me in Leiden ;). Thank you for all the great walks!

And of course my wonderful family. Dear mam and pap, thank you for always believing in me. I had the best childhood I could ever imagine and I am happy that I can still come home to you for a place to feel completely safe and happy. To my big sister, Krista, who has an incredible dedication to everything she loves and works for. Thank you for arranging so many wonderful family outings. And to my brother, Willem, who always works hard and never complains. Thank you for your incredible sense of humor that always makes me laugh! And to Jochen, who joined your family more than 10 years ago. Thank you for being the tour guide in many of our family outings!

For the family of Frank, who welcomed me into their family. Your family is incredibly close, it is great to be able to be part of such a warm family. Thank you.

And last, but definitely not least, Frank. We met in the first month of my PhD and became inseparable right from the start. You are an amazing soul and the most courageous person I know.

I am thankful that you want to share your life with me. Thank you for all you support and patience, especially at the end of the PhD were I needed it most. Thank you for making my life so wonderful!

09 – 08 – 2017,

Alphen aan den Rijn

# About the author

Hilda van Mourik was born on the 8th of May 1987 in Woerden, The Netherlands. She completed the higher general secondary education (HAVO) degree in 2004 at "Kalsbeek College" in Woerden. Directly after she started the bachelor Applied Sciences at Hogeschool Utrecht, studying Life Sciences with a specialization in Molecular Biology. During this study she completed an internship at the laboratory of Phytopathology under supervision of Prof. Dr Bart Thomma. She worked on the development of a plant model for the human pathogen *Staphylococcus aureus*. During her bachelor Hilda worked part-time as a microbiology technician for Aseptix Technologies and as a student assistant at Hogeschool Utrecht. Moreover, at the end of her bachelor studies Hilda started following courses at Wageningen University.

After receiving her bachelor degree in 2008, Hilda decided to continue her studies in the academic world and enrolled into the master program Biology at Wageningen University. During this master program, Hilda specialized in Molecular Plant Biology and Mathematical Modelling. She finished two internships. For her first internship she joined the laboratory of Biochemistry in the group of Prof. Dr Dolf Weijers, under supervision of Dr Bert de Rybel. She studied the role of the transcription factor TARGET OF MONOPTEROS 7 (TMO7) during *Arabidopsis thaliana* embryonal development. For her second internship she worked at the research group Biometris at Wageningen University under supervision of Prof. Dr Jaap Molenaar and Dr Maarten de Gee. Her internship project was performed in collaboration with the laboratory of Genetics together with Prof. Dr Arjan de Visser. During this internship Hilda developed a mathematical model for the competition of toxin-producing and toxin-sensitive *Saccharomyces cerevisiae* colonies.

Directly after finishing her master degree, Hilda started a PhD program at Plant Developmental Systems in the group of Prof. Dr Gerco Angenent under supervision of Prof. Dr Kerstin Kaufmann. The work she performed during her time as a PhD student resulted in the publication of this thesis. She is currently working as a researcher and lecturer bio-informatics at Hogeschool Leiden.

# Publications

**Van Mourik, H.**, Muiño, J. M., Pajoro, A., Angenent, G. C., & Kaufmann, K. (2015). Characterization of *In Vivo* DNA-Binding events of plant transcription factors by ChIP-seq: experimental protocol and computational analysis. Plant Functional Genomics: Methods and Protocols, 93-121.

**Van Mourik, H.**, Muiño, J. M., Smaczniak, C. D., Bemer, M., Chen, D., Angenent, G. C. & Kaufmann, K. (In preparation). Dual specificity and target gene selection by the MADS domain protein FRUITFULL.

**Van Mourik, H.**, Van Dijk, A. D. J., Stortenbeker, N., Angenent, G. C. &, Bemer, M. (Submitted). Divergent Regulation of *Arabidopsis* SAUR genes: a focus on the SAUR10-clade.

Bemer, M., **Van Mourik, H.**, Muiño J. M., Ferrándiz, C., Kaufmann, K & Angenent, G. C. (2017). FRUITFULL controls *SAUR10* expression and regulates *Arabidopsis* growth and architecture. J Exp Bot 2017 erx184.

Ruelens, P., Zhang, Z., **Van Mourik, H.**, Maere, S., Kaufmann, K., & Geuten, K. (2017). The origin of floral organ identity quartets. The Plant Cell, 29(2), 229-242.

Lu, K. J., De Rybel, B., **Van Mourik, H.**, & Weijers, D. (2017). Regulation of intercellular TARGET OF MONOPTEROS 7 protein transport in the *Arabidopsis* root. bioRxiv, 122960.

De Gee, M., **Van Mourik, H.**, De Visser, A., & Molenaar, J. (2016l). Modeling competition between yeast strains. In Y. Takeuchi, & M. Apri (Eds.), AIP Conference Proceedings (Vol. 1723, No. 1, p. 020001). AIP Publishing.

Aerts, N., De Bruijn, S., **Van Mourik, H.**, Angenent, G. C. & Van Dijk, A. D. J. (Submitted). Comparative Analysis of Binding Patterns of MADS domain Proteins in *Arabidopsis thaliana*.

De Rybel, B., van den Berg, W., Lokerse, A. S., Liao, C. Y., **Van Mourik, H.**, Möller, B., Peris, C. L. & Weijers, D. (2011). A versatile set of ligation-independent cloning vectors for functional studies in plants. Plant physiology, 156(3), 1292-1299.

# Education Statement of the Graduate School

# Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

**Issued to:** Hilda van Mourik
**Date:** 10 November 2017
**Group:** Bioscience-PDS and Laboratory of Molecular Biology
**University:** Wageningen University & Research

| 1) Start-up phase | *date* |
|---|---|
| ► **First presentation of your project** | |
| *Title:* Target gene selection by the MADS-domain protein FRUITFULL: the role of proteins complexes, DNA accessibility and higher order chromatin organization. | 06 Mar 2012 |
| ► **Writing or rewriting a project proposal** | |
| *Title:* Target gene selection by the MADS-domain protein FRUITFULL: the role of proteins complexes, DNA accessibility and higher order chromatin organization. | Dec 2011 |
| ► **Writing a review or book chapter** | |
| "Characterization of In Vivo DNA-Binding Events of Plant Transcription Factors by ChIP-seq: Experimental Protocol and Computational Analysis" Springer New York,Methods in Molecular Biology, in Book: Plant Functional Genomics, p. 93-121, 2015 | Jan 2015 |
| ► **MSc courses** | |
| ► **Laboratory use of isotopes** | |
| *Subtotal Start-up Phase* | *6.5 credits** |

| 2) Scientific Exposure | *date* |
|---|---|
| ► **EPS PhD student days** | |
| EPS PhD student day, Amsterdam, NL | 30 Nov 2012 |
| EPS PhD student day, Leiden, NL | 29 Nov 2013 |
| EPS PhD Student Days 'Get2Gether', Soest, NL | 29-30 Jan 2015 |
| ► **EPS theme symposia** | |
| EPS theme 1 'Developmental Biology of Plants', Wageningen, NL | 19 Jan, 2012 |
| EPS theme 1 'Developmental Biology of Plants', Leiden, NL | 17 Jan, 2013 |
| EPS theme 1 'Developmental Biology of Plants', Leiden, NL | 24 Jan, 2014 |
| EPS theme 1 'Developmental Biology of Plants', Leiden, NL | 08 Jan, 2015 |
| ► **National meetings (e.g. Lunteren) and other National Platforms** | |
| Annual meeting 'Experimental Plant Sciences', Lunteren, NL | 02-03 Apr, 2012 |
| Annual meeting 'Experimental Plant Sciences', Lunteren, NL | 22-23 Apr, 2013 |
| Annual meeting 'Experimental Plant Sciences', Lunteren, NL | 14-15 Apr, 2014 |
| Annual meeting 'Experimental Plant Sciences', Lunteren, NL | 13-14 Apr, 2015 |
| National PhD day, Delft, NL | 29 Oct, 2016 |
| ► **Seminars (series), workshops and symposia** | |
| Frank Pugh, talk about ChIP-exo (Utrecht) | 14 Jun 2012 |
| Doris Wagner, 'Auxin mediated organogenesis in Arabidopsis' (Dolf's lab) | 10 Jul 2012 |
| Cristel Carles, Chromatin dynamics and cell fate in plants: From Genetics to Epigenetics (Radix) | 16 Jan 2013 |
| Andrew Sugden, Editor of Science (Radix) | 08 Feb 2013 |
| Flying Seminar Detlef Weigel (MPI Developmental Biology, Tübingen, Germany) | 27 Feb 2013 |
| Start symposium Plant Developmental Biology (Wageningen) | 14 Oct 2013 |
| Robert Paul Zimmermann, "Detection of convoluted aptamers from high-throughput SELEX data", MPI Berlin | 11 Sep 2013 |
| Maria-Helena de Souza Goldman, "A tissue-specific cell cycle regulator and its potential mechanism of action" (Wageningen) | 22 May, 2014 |
| Javier Palatnik, "Control of cell proliferation and differentiation by microRNAs." (Wageningen) | 03 Jun 2014 |
| Prof. Eric Schranz, "Influence of Genome and Gene Duplications to trait Evolution in Flowering Plants" | 04 No 2014 |
| Prof. Dr. Ortum Mittelstein Scheid, "Genetics and epigenetics: a complex relationship." Wageningen | 19 Nov 2014 |
| Prof. Dr. George Coupland, "Seasonal flowering in annual and perennial plants" EPS flying Seminar | 19 Jan 2015 |
| Prof. Dr. Marcelo C. Dornelas, "Using the non-model genus Passiflora to study the evolution of novelty in plant reproductive development" (Wageningen) | 27 Jan 2015 |
| Prof. Yves van de Peer, "The evulutionary significance of gene and genome duplications" | 03 Feb 2015 |
| Prof. Martin Kater, "Mining for Floral Meristem Regulatory Pathways in Arabidopsis and Rice" | 11 Mar 2015 |
| Prof. Malgorzata Gaj, "Hormone-related functions of LEC2 in Somatic Embryogenesis Induction in Arabidopsis" | 16 Apr 2015 |
| Dr. Francois Parcy, "An integrated structural biology approach to flower development" | 15 Oct 2015 |
| Dr. Jingyu Zhang, "The response to cold stress in rice: signaling, transcriptional and metabolic regulation" | 01 Dec 2015 |
| Cristina Ferrandiz (IBMCP, Valencia, Spain) "The secrets of a long life: genetic control of life span in monocarpic plants" | 21 Jan 2016 |
| ► **Seminar plus** | |
| ► **International symposia and congresses** | |
| PhD Workshop Golm "Molecular Genetics of Plant Development and Stress Response", Germany | 17 Apr 2013 |
| PhD Workshop Golm "Molecular Plant Sciences", Germany | 11 Dec 2014 |
| Flower Development Workshop, France | 08-12 Jun 2013 |
| Flower Development Workshop, Spain | 07-11 Jun 2015 |

***CONTINUED ON NEXT PAGE***

| | | |
|---|---|---|
| ► | **Presentations** | |
| | *Talk:* International PhD School in Plant Development, Sienna, Italy | 26 Sep 2012 |
| | *Poster:* Annual meeting 'Experimental Plant Sciences', Lunteren, NL | 22 Apr 2013 |
| | *Flash talk:* "Start symposium Plant Developmental Biology", Wageningen, NL | 24 May 2013 |
| | *Flash talk & poster:* Flower Development Workshop, France | 09 Jun 2013 |
| | *Talk:* Annual meeting 'Experimental Plant Science', Lunteren, NL | 14 Apr 2015 |
| | *Talk:* Thematic Meeting Bioscience, Wageningen, NL | 19 Mar 2013 |
| | *Talk:* Thematic Meeting Bioscience, Wageningen, NL | 18 Mar 2014 |
| | *Talk:* Thematic Meeting Bioscience, Wageningen, NL | 17 Mar 2015 |
| | *Talk:* Thematic Meeting Bioscience, Wageningen, NL | 13 Oct 2015 |
| | *Talk:* PhD Workshop 'Molecular Genetics of Plant Development and Stress Response', Golm, Germany | 11 Dec 2013 |
| | Talk: PhD Workshop 'Molecular Plant Sciences', Golm, Germany | 17 Apr 2014 |
| | *Flash Talk & Poster:* Flower Development Workshop, Spain | 07-11 Jun 2015 |
| ► | **IAB interview** | |
| ► | **Excursions** | |
| | Rijk Zwaan Excursie, De Lier, NL | 27 Sep 2013 |
| | *Subtotal Scientific Exposure* | *22.7 credits** |

| **3) In-Depth Studies** | | *date* |
|---|---|---|
| ► | **EPS courses or other PhD courses** | |
| | International PhD School in Plant Development, Sienna, Italy | 25-28 Sep 2012 |
| | EMBL advanced course: "Towards Next-Generation Sequencing Data Integration", Heidelberg, Germany | 23-26 Sep 2013 |
| | Postgraduate Course 'Transcription Fators and Transcriptional Regulation', Wageningen, NL | 17-19 Sep 2013 |
| ► | **Journal club** | |
| | Member of literature discussion group at Bioscience-PDS | 2012-2016 |
| ► | **Individual research training** | |
| | Analysing ChIP-seq with Jose Muino, MPI Berlin, Germany | 10-13 Sep 2013 |
| | *Subtotal In-Depth Studies* | *7.6 credits** |

| **4) Personal development** | | *date* |
|---|---|---|
| ► | **Skill training courses** | |
| | WGS Course: Project & time management | Nov-Dec, 2012 |
| | WGS Course: Scientific writing | Oct-Dec, 2013 |
| | Workshop Tools for networkers | 02 Jun 2013 |
| | WGS PhD Workshop Carousel (Pitch workshop, Be a vital PhD, Scientific Publishing) | 02 Jun 2014 |
| | WGS PhD Workshop Carousel (Speed Reading, writing propositions) | 17 Apr 2015 |
| | WGS PhD Workshop Carousel (Networking, online personal branding, transferable skills) | 08 Apr 2016 |
| | WGS Course: Career Orientation | Mar 2016 |
| ► | **Organisation of PhD students day, course or conference** | |
| ► | **Membership of Board, Committee or PhD council** | |
| | *Subtotal Personal Development* | *5.8 credits** |

| **TOTAL NUMBER OF CREDIT POINTS*** | **42,6** |
|---|---|

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

** A credit represents a normative study load of 28 hours of study.*

This research was supported by a NWO VIDI network grant

Financial support from Wageningen University for printing this thesis is gratefully acknowledged

Thesis layout and cover design by the author with help of proefschriftmaken.nl.

Printed by proefschriftmaken.nl