# Propositions

1. Proteomics approach is a powerful technology to identify protein complex partners.
(this thesis)

2. The real challenge is not in identifying plant chromatin interactions but in unravelling their function.
(this thesis)

3. Refusing a gene-editing technology for ethical reasons ignores the ethical consequences of not using it.
(Gaudelli, N.M. et al. Nature Oct. 2017)

4. Looks matter, even in science.
(Gheorghiu, A. I et al. PNAS April. 2017)

5. The drive to remove fear from society is causing it.

6. The mere presence of a tradition does not justify its practice.

7. Rather than looking for water on distant planets, we should find a solution on this planet for the 844 million people who don't have access to clean drinking water.

Propositions belonging to the thesis, entitled:

**MADS FLORAL INTEGRATORS**

**Insights into molecular mechanisms of MADS domain proteins in the floral transition**

Suraj Babasaheb Jamge

Wageningen, 9 February 2017

Suraj B. Jamge

# MADS FLORAL INTEGRATORS

## Insights into molecular mechanisms of MADS domain proteins in the floral transition

# MADS FLORAL INTEGRATORS

Insights into molecular mechanisms of MADS domain proteins in the floral transition

**Suraj B. Jamge**

**Thesis committee**

**Promotor**
Prof. Dr Gerco C. Angenent
Personal chair at the Laboratory of Molecular Biology
Wageningen University & Research

**Co-promotor**
Prof. Dr Richard G.H. Immink
Special Professor Physiology of Flower Bulbs
Wageningen University & Research

**Other members**
Prof. Dr Mark G.M. Aarts, Wageningen University & Research
Prof. Dr Christa S. Testerink, Wageningen University & Research
Prof. Dr Claire Périlleux, University of Liège, Belgium
Dr Paul F. Fransz, University of Amsterdam

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences.

# MADS FLORAL INTEGRATORS

Insights into molecular mechanisms of MADS domain proteins in the floral transition

Suraj B. Jamge

**Thesis**

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of 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 9 February 2018
at 4 p.m. in the Aula.

Suraj B. Jamge

MADS FLORAL INTEGRATORS Insights into molecular mechanisms of MADS domain proteins in the floral transition,

206 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, with summaries in Dutch and English

ISBN 978-94-6343-735-6
DOI https://doi.org/10.18174/430710

**प्रथमेश,**

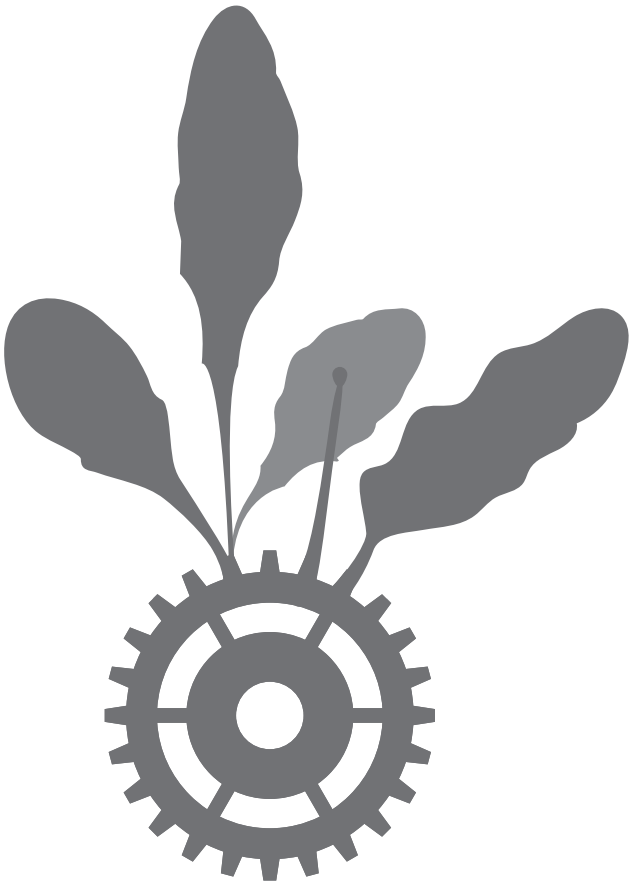
(१९८८ - २०१६)

मम **हृदये** तव नित्य **स्मरण**

# Contents

# CHAPTER 1

## Introduction: Molecular regulation of flowering time in *Arabidopsis thaliana*



How plants sense and respond to their changing environment is an intriguing question. Although substantial progress has been made in the past decade, partly due to technical advances, our understanding of this communication is still far from complete. At any given time, plants interact with several biotic (insects and pathogens) and abiotic factors (light, temperature, salinity, drought etc.), and these interactions might have a dramatic effect on plants fitness. Since plants cannot escape their environment, they must find ways to adapt and deal with any adversities. Developmental decisions and functioning of cellular processes are tightly controlled internally, but also take into account the environment. The transition to flowering is one of these developmental decisions and an important trait in the life cycle of flowering plants, as it is directly linked to seed production, and survival of the species. Agronomically, it is also of utmost importance, because it determines yield and harvest time. Studies in the dicot model species *Arabidopsis thaliana* have resulted in a remarkable source of information, guiding the identification of genes, governing the time of flowering, and unravelling their genetic and molecular interactions. Flowering, which is the transition from the vegetative to the reproductive phase of development, is the result of sensing various endogenous and environmental signals, being integrated to result in a simple flowering stimulus. During the vegetative phase of Arabidopsis, the shoot apical meristem (SAM) produces leaf primordia on its flanks. Upon transition to flowering, the apical meristem undergoes an abrupt transition into an inflorescence meristem, after which flower primordia (floral meristems) are produced instead of leaf primordia. Unlike the vegetative SAM and inflorescence meristem that continue to produce primordia indefinitely, floral meristems are determinate structures that produce a defined number of organ primordia. These organ primordia reside in four concentric whorls, with four sepals in the outermost first whorl, followed by four petals, six stamens (male reproductive organs), and two fused carpels (female structures) in the 2nd, 3rd, and 4th whorls, respectively. Later on in development, the fertilized ovary will give rise to the fruit (silique). An impressive number of studies in the past decades has culminated in the insight that flowering is determined by a complex network of genes and gene products under influence of environmental and endogenous signals. As such, the photoperiod, autonomous, vernalization, ambient temperature, hormonal (e.g. gibberellic acid (GA) signalling), and aging pathways can be distinguished as major flowering time controlling signalling pathways (Figure 1). More than 300 genes involved in flowering time control have been identified, mainly by classical genetic studies and mutant analyses (Putterill et al., 2004; Fornara et al., 2010; Andres and Coupland, 2012; Bouche et al., 2016). We are beginning to understand how these genes are regulated and thereby integrate the various internal and external cues to control the onset of flowering. In this chapter I introduce how the different layers of gene regulation can influence and profoundly affect the time of flowering in the model plant *Arabidopsis thaliana*. An overview will be given from the different ways of transcriptional control and the mode of action of some of the regulatory proteins involved.

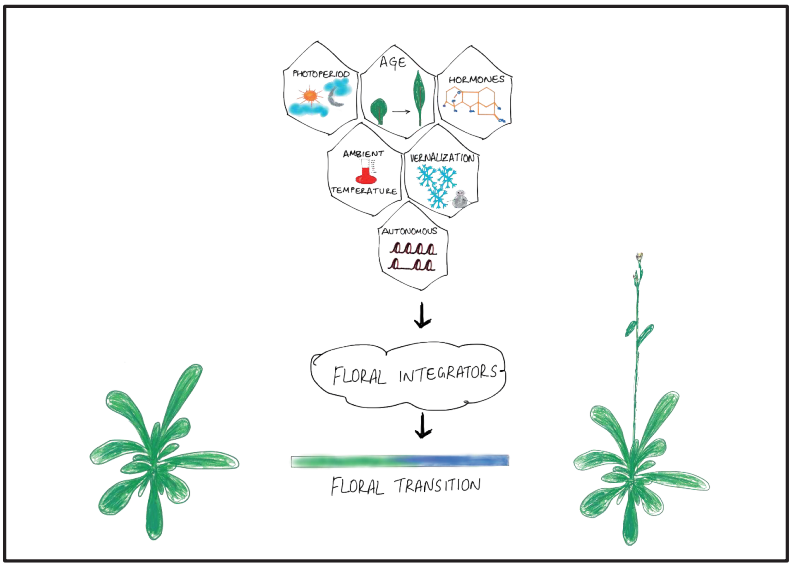



Figure 1. **Major flowering pathways in *Arabidopsis***: Photoperiod, Autonomous, Vernalization, Photoperiod, Aging, Ambient temperature and Hormonal Pathways.

**Transcriptional control of flowering time**

Eukaryotic organisms have a much larger genome than prokaryotes and their genomes are organized into multiple chromosomes with a higher sequence complexity and packed in chromatin (Lynch and Conery, 2003). Despite the fact that every diploid cell of an organism carries the same DNA sequence, yet the tissues arising from these cells drastically differ in appearance and function. The reason for such variation simply resides in the manner of how certain genes are turned “on” and “off” during development, i.e. differential gene regulation. By regulating their genes, organisms respond to internal and external signals to orchestrate development (Scott, 2000; Scheres, 2007). The first level of gene regulation is the transcriptional control. DNA-binding proteins, such as Transcription Factors (TFs), are involved in the process of transcribing DNA into mRNA. One of the basic features of TFs is that they contain DNA-binding domains that recognise and bind to specific sequences within the genome. Certain TFs bind to DNA sequences on the promoter close to the Transcription start site (TSS) and help to form the transcription initiation complex, whereas TFs can also bind to distant regulatory sequences that are thousands of bases upstream or downstream from the gene to be transcribed (Ohler and Wassarman, 2010; Lenhard et al., 2012; Spitz and Furlong, 2012). Binding of TFs to these regulatory sequences can either increase or decrease expression of the associated gene and thereby these TFs are thus acting as activators or repressors. TFs can operate alone or interact with other molecules to achieve their activation or repression function. Therefore, complex formation is another key feature for TF function. In short, the ability of TFs to recognize and bind to specific DNA sequences (as well as with other

components) with high affinity aid in specification and fine tuning of gene expression across tissue types and development stages (Babu, 2010; Kaufmann et al., 2010; Spitz and Furlong, 2012).

Plant TFs are classified into different families based on their structural features and on their DNA-binding domain (Riechmann et al., 2000), for example helix-loop-helix proteins and MADS domain-containing proteins, respectively. Plant TF families are involved in variety of developmental processes. Some have specialized and defined functions at certain developmental stages, whereas others are active across all stages of development. A substantial number of TFs have been identified to be co-ordinately regulating the floral transition process. These TFs are grouped into different pathways, and have been schematically illustrated in the Flowering Interactive Database (FLOR-ID) (http://www.phytosystems.ulg.ac.be/florid/)(Bouche et al., 2016). A majority of the flowering time TFs belongs to MCM1/AGAMOUS/DEFICIENS/SRF (serum response factor; MADS) domain transcription factor family, although also members of other families have been identified. The focus of this thesis is on selected MADS TFs and their role in flowering time control, but also important key regulatory flowering time regulators of other families will be briefly discussed. In Arabidopsis, *CONSTANS (CO)* and *FLOWERING LOCUS C (FLC)* perceive different environmental flowering controlling signals and regulate the vegetative to reproductive phase change antagonistically(Putterill et al., 1995; Samach et al., 2000). The *CO* gene encodes a putative zinc finger transcription factor which acts as a floral activator and mediates the photoperiod pathway, whereas the *FLC* gene is encoding a MADS domain protein that acts as a floral repressor and mediates the autonomous and vernalization pathways. In turn, the CO and FLC transcription factors regulate the expression of important downstream genes, *SUPPRESSOR OF OVEREXPRESSION OF CO 1 (SOC1) and FLOWERING LOCUS T (FT)(Simpson and Dean, 2002; Parcy, 2005)*. Notably, these genes regarded as floral integrators, can receive and integrate signals from multiple pathways and thereby are key in determining the exact flowering time(Simpson and Dean, 2002). Ultimately, this results in reprogramming of the shoot apex from vegetative identity to reproductive identity due to the activation of the floral meristem identity genes *APETALA1 (AP1)* and *LEAFY (LFY)*. Subsequently, floral organ identity genes such as SEPALLATA3 (SEP3), *AGAMOUS (AG)* and APETALA3 (*AP3)* get expressed(Fornara et al., 2010; Kaufmann et al., 2010; Andres and Coupland, 2012).

SOC1 is a MADS-domain transcription factor initially identified as suppressors of CO overexpression and later found to be regulated as well by gibberellin (GA) signalling (Lee et al., 2000; Onouchi et al., 2000; Samach et al., 2000; Hepworth et al., 2002; Moon et al., 2003). Detailed analyses revealed its central integrative role as a positive regulator of flowering (Lee and Lee, 2010; Immink et al., 2012; Torti et al., 2012). *SOC1* is mainly expressed in developing leaves and the SAM to promote flowering in *Arabidopsis (Samach et al., 2000; Lee and Lee, 2010)*. Notably, during the vegetative phase, *SOC1* expression is suppressed by a repressor

complex that consist of FLC and SHORT VEGETATIVE PHASE (SVP), whereas towards floral transition, autonomous, GA, ambient temperature and vernalization pathways down-regulate *FLC* and *SVP*, and thereby de-repress *SOC1(Lee et al., 2007; Li et al., 2008)*. In addition, the *SOC1* encoding protein acts with another MADS-domain protein, AGL24, to promote flowering by regulating each other's expression at the shoot apex during floral transition in a GA-dependent manner (Yu et al., 2002; Michaels et al., 2003; Lee and Lee, 2010). Furthermore, SOC1 and AGL24 are responsible for the regulation of *LFY* in response to the photoperiod (Liu et al., 2008). Overall, transcriptional regulation of the above discussed MADS-domain TFs have a direct impact on the flowering time process. Below I will describe the various ways of transcriptional control and mode of actions of the factors involved in the different flowering pathways.

**Co-transcriptional control / Role of alternative splicing**

Alternative splicing, is a process that results in different mature mRNAs from a single precursor mRNA. Approximately 60% of the intron containing genes in *Arabidopsis* undergo alternative splicing, resulting among others in increased diversity and functionality of the cell's proteome. It is known that *Arabidopsis* Columbia (Col-0) accession flowers earlier when exposed to warm ambient temperatures, whereas exposure to low ambient temperatures delays flowering. Genetic evidence links the flowering phenotype of *Arabidopsis* with the presence of an alternative spliced variant of the MADS-domain gene *FLOWERING LOCUS M (FLM),* upon exposure to varying ambient temperature (Balasubramanian et al., 2006). A combined genetic and biochemical study demonstrated that production of the FLM-δ splice variant is favoured in warm, whereas FLM-β is more abundant at cold ambient temperature(Lee et al., 2013; Pose et al., 2013). The authors showed that the proteins putatively encoded by both splice variants interact with the floral repressor *SVP,* and that the composition of this complex defines the flowering response. The FLMβ-SVP complex acts as a potent repressor of flowering by direct binding at the regulatory sequences of floral integrator genes, such as *SOC1.* In contrast, interaction of FLM-δ with SVP renders SVP inactive and out-competes FLMβ binding to SVP(Lee et al., 2013; Pose et al., 2013). Additionally, the formation of *FLM* isoforms that get targeted for nonsense mediated RNA decay (NMD) at higher ambient temperatures seems to play a role in the ambient temperature flowering time response (Sureshkumar et al., 2016) and the closely related *MADS AFFECTING FLOWERING2 (MAF2)* gene appears to be regulated in a similar way (Airoldi et al., 2015). Together, these examples clearly highlight that intricate mechanisms, such as alternative splicing, are directly linked to flowering time control.

**Non-coding RNA based regulation**

Since the complete genome sequences of multicellular model-organisms became available, a striking similarity across them was that only a few percent of the genomes comprised protein coding sequences and the remainder was intronic, intergenic and non-coding sequences. Advances in next generation sequencing methodologies and analysis has led to detection of

pervasive transcription and identification of small RNAs and long noncoding RNAs (lncRNAs) at an unprecedented scale. microRNAs (miRNAss), 20–24 nucleotide small noncoding RNAs, have been very well described for their role in the plant growth and development(Jones-Rhoades et al., 2006; Zhang et al., 2006; Zhang et al., 2007; Banerjee et al., 2016; Martinez and Kohler, 2017). miRNAs function by either targeting mRNAs for cleavage or translation inhibition (Bartel, 2004). In *Arabidopsis*, miR156 and miR172 have been identified as important players in regulation of flowering time under the age-dependent pathway (Aukerman and Sakai, 2003; Wu and Poethig, 2006; Jung et al., 2007; Wang et al., 2009). Overexpression of miR156 results in delayed flowering response and overexpression of miR172 accelerates flowering in *Arabidopsis*. A decrease in expression of miR156, whereas a concomitant increase in miRNA172 occurs just before the switch from vegetative to reproductive stage. miRNA156 have been shown to target SQUAMOSA PROMOTER-BINDING PROTEIN-LIKEs (SBPs/SPLs) genes, whereas miR172 target members of the AP2 and AP2-like genes (Aukerman and Sakai, 2003; Wu and Poethig, 2006; Jung et al., 2007; Wang et al., 2009; Yamaguchi et al., 2009) and the sequential action of these two miRNAs regulate the developmental phase transition in *Arabidopsis* (Fornara and Coupland, 2009; Wu et al., 2009; Spanudakis and Jackson, 2014; Teotia and Tang, 2015). Besides smallRNAs, lncRNAs also appear to be crucial in different developmental process. Based on recent reports, ~70% of Arabidopsis protein-coding loci encode potential natural antisense transcript (NATs) (Wang et al., 2014). This finding is line with studies done in humans (~61-72%) and mouse (70%). Although thousands of long non-coding transcripts have been identified in *Arabidopsis*, the role and function of only a handful have be recently explored. Among them, *COOLAIR*, is one of the well characterized example of antisense transcript in *Arabidopsis*. *COOLAIR,* an antisense transcript of the *FLC* locus, is initiated and regulated independently of the *FLC* sense transcript (Swiezewski et al., 2009; Liu et al., 2010b). In a series of late and early flowering Arabidopsis mutants within the autonomous and vernalization pathways, Swiesewski *et al*. closely examined the levels of *FLC* sense and non-coding antisense transcripts. *COOLAIR* transcription was found to be positively correlated with the *FLC* sense transcription. However, upon 2 weeks of cold treatment, a 10-fold increase in the antisense transcript was observed, whereas the sense *FLC* transcript was downregulated. Seven days after returning the cold treated plants to warm conditions, the levels of the antisense RNA were similar to plants that had no cold exposure, indicating the 10-fold increase to be transient and mainly cold dependent. Further studies on *COOLAIR* antisense transcripts revealed that it is both alternatively spliced as well as alternatively polyadenylated. Closer examination of targeted 3' processing of the *COOLAIR* antisense transcripts identified a role for a set of RNA binding proteins (FCA, FY, FPA) within the autonomous pathways to affect the alternative polyadenylation of the antisense RNAs at the *FLC* locus. It is also worthwhile to mention that the mutant backgrounds of these RNA binding proteins do not affect alternative polyadenylation of the *FLC* sense transcripts. Increased proximal polyadenylation of *COOLAIR* is now linked with low *FLC* sense expression, whereas distal polyadenylation is associated with high *FLC* transcription. However, the exact mechanism of how *COOLAIR* processing is linked to

changes in transcription is a matter of ongoing research. Nevertheless, *COOLAIR* mediated repression of *FLC* appears to be an early event in the cold-induced epigenetic silencing of *FLC (Swiezewski et al., 2009)* to control the time of flowering. Overall, these findings offer an exciting insight into how non-coding antisense transcripts can regulate its sense partners in a developmental stage- and condition-dependent manner, and the role of the RNA processing machinery in this process.

**Post-translation modifications in flowering time control:**

Post-translation modifications (PTMs) play a pivotal role in regulating protein stability and function. After translation, numerous proteins are subjected to covalent modifications. Adding covalent modifications or functional groups to amino-acid residues of proteins expands the functional repertoire of proteins. To date more than 400 types of PTMs have been assumed to exist in nature (Pagel et al., 2015) (Doll and Burlingame, 2015). The most frequently reported PTMs comprises phosphorylation, ubiquitination, methylation, sumolyation, and acetylation. Besides single modifications, proteins are often undergoing a series of PTMs and this is a prerequisite for their maturation or activation. Thus, protein modifications affect and at large define variety of developmental functions. In context of flowering time, biochemical evidence of PTMs has been unearthed for several key floral regulators.

Sufficient amount of day length can trigger plants to flower. *CO* is regulated by the circadian clock and its expression peaks at the end of the day (Suarez-Lopez et al., 2001). The peak in CO protein abundance is limited to long day conditions, when plants receive daylight for more than 12 h. At the post translational level, CO protein is stabilized when plants are exposed to light, whereas during night CO is subjected to degradation via the ubiquitin proteasome system (UPS) (Valverde et al., 2004b). In LDs, the stabilized CO ensures transcriptional activation of the floral integrator *FT (An et al., 2004)*. Recent studies have identified CO being subjected to different PTMs as described below (Sarid-Krebs et al., 2015).

Phosphorylation, i.e. covalent addition of a phosphate group to Ser/Thr/Tyr residues by protein kinases, is a very dynamic and reversible PTM. To date, phosphorylation is one of the most studied modifications in eukaryotes (Manning et al., 2002; Zulawski et al., 2013). The rate of turnover of the CO protein is influenced by phosphorylation (Sarid-Krebs et al., 2015). Throughout the diurnal cycle, both the phosphorylated and unphosphorylated forms of CO were found in varying ratio. The relative abundance of the phosphorylated form of CO was higher during the daytime, whereas the non-phosphorylated variant was more abundant at night. This difference is a result of rapid degradation of the phosphorylated form of CO at night through the 26S proteasome pathway mediated by CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1). COP1 is well known for its role in protein ubiquitination, a hallmark for degradation of proteins by the proteosome system. It is an ATP-dependent process, carried out by three different classes of ubiquitin enzymes namely activating enzyme -E1, conjugating enzyme E2, and ligating enzyme - E3. COP1 is a E3 Ubiquitin ligase. In conditions where the activity of COP1

was compromised, the phosphorylated form of CO was more abundant than the non-phosphorylated form. In conditions with high COP1 activity, the relative abundance of the two forms appeared to be reversed. Based on these observations CO phosphorylation appears to contribute to the photoperiod flowering response and this is mainly by enhancing the rate of CO turnover mediated by the activity of COP1 (Valverde et al., 2004a) (Sarid-Krebs et al., 2015; Xu et al., 2016).

In *Arabidopsis*, temperature-dependent ubiquitination was recently described in flowering time control. Certain *Arabidopsis* accessions require a prolonged cold exposure (i.e. vernalisation) before they can switch to flowering. On contrary any intermittent or short exposure to cold treatments rather delays the transition to flowering. In addition to photoperiod regulation of CO as described earlier, CO protein was also shown to undergo temperature-dependent ubiquitination. In a study, when plants were exposed to cooler temperatures, no discernible changes at the *CO* mRNA level occurred, however, the CO protein abundance dramatically varied. These changes in CO abundance was the result of CO degradation, but rather independent of *COP1* activity. Instead, another E3 ubiquitin ligase HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENE1 (HOS1), a cold signalling attenuator, directly interacts with CO and triggers CO degradation upon cold stress and regulates flowering in *Arabidopsis* (Jung et al., 2012). Ubiquitination has been traditionally viewed as a PTMs resulting in death sentence for proteins. However, recent studies have shown it to have non-proteolytic functions, such as protein trafficking, subcellular localization, signalling, DNA repair and transcriptional regulation (Schnell and Hicke, 2003; Sun and Chen, 2004; Wang and Deng, 2011). The implications of ubiquitination are on a variety of development processes.

Sumoylation, another type of post-translational regulatory process, has been shown to function in flowering time control as well. Similar to ubiquitin, SUMO (Small Ubiquitin-related Modifier) a small polypeptide, covalently attaches to targeted lysine residues. Like ubiquitination, sumolyation is also carried out by the action of three different classes of enzymes E1, E2 and E3, sharing similar non-proteolytic functions. Sumolyation is a relatively unexplored PTMs in plants, until recently. For instance, AtSIZ1, a SUMO E3 ligase, was shown to sumolyate FLOWERING LOCUS D (FLD) (Jin et al., 2008). FLD is a well-known repressor of *FLC* in the *Arabidopsis* FRIGIDA background (Jin et al., 2008, Kwak, 2016 #1275). This Sumolyation results in inhibition of FLD activity, indicating that AtSIZ1 act as a floral repressor. In another study, mutant alleles of a SUMO nuclear protease *EARLY IN SHORT DAYS (ESD4)* were shown to result in extremely early flowering (Reeves et al., 2002) (Murtas et al., 2003). Loss of ESD4 function results in reduced levels of free SUMO and an increase in SUMO conjugates, postulating a role for ESD4 in regulating the abundance of SUMO conjugates (Murtas et al., 2003). Similar to *eds4* mutants, the *nuclear pore anchor (nua)* mutant shows an early flowering response, elevated SUMO conjugates, and reduced free SUMO (Xu et al., 2007) (Jin et al., 2008). Overall, these studies demonstrate how post translational systems are involved in modulation of the flowering stimulus.

**Chromatin based regulation of flowering time**

In eukaryotes, including plants, the genetic information is organized in chromatin. Throughout the course of development, different patterns of chromatin states are established, disrupted or maintained, for a variety of *in vivo* processes. Chromatin, a dynamic fibre made of nucleosomal units, in which ~147bp of DNA is tightly wrapped around structural proteins, namely histones, represents the first level of chromatin organization (Luger et al., 1997; Luger et al., 2000; Davey et al., 2002). At the centre of each nucleosome core particle sits a histone octamer, a protein complex made up of two copies of each of the histones H2A, H2B, H3 and H4, respectively. The binding of histone H1 to linker DNA sequences between nucleosomes further assembles the chromatin into higher-order structures. A variety of distinct but linked processes lead to changes in the structural organization of chromatin. These includes the exchange of canonical histones by histone variants (Kamakaka and Biggins, 2005), post translation modifications of histones (Kouzarides, 2007), and ATP-dependent chromatin remodelling (i.e. sliding or eviction of nucleosomes across the DNA) (Clapier et al., 2017). Here we discuss key determinants of chromatin organization and their role in governing the flowering time response in *Arabidopsis*.

The canonical histones sit at the core of the chromatin. Similar to PTMs discussed for TF proteins, histone proteins are also subjected to PTMs and this is tightly linked with orchestrating gene transcription (Kouzarides, 2007). Specific enzymes, such as histone acetyltransferases (HATs), histone deacetylases (HDAs), histone methyltransferases (HMTs), and histone kinases (HKs) carry out these modifications. As a result, this can activate or repress transcription, or alternatively, keep genes in a so-called poised state. At the chromatin level, one or combinatorial modifications at histones are proposed to directly alter chromatin structure, or to serve as a platform to recruit additional factors (Zhang, 2008). In fact, plants possess machinery's that enable them to deposit, remove or maintain histone modifications throughout development. Over the last two decades, a vast amount of epigenomic studies have identified the role of histone modification in molecular processes such as, DNA replication, control of alternative splicing, transcriptional regulation, and biological processes such as, timing of flowering, pollen tube formation and floral organ development (Pfluger and Wagner, 2007; Farrona et al., 2008; Liu et al., 2010a; Luco et al., 2011; Kawashima and Berger, 2014).

In the context of flowering time, one of the best studied example of effects of histone modifications on gene regulation, is the floral repressor *FLC*. *Arabidopsis* winter-annuals require vernalization to accelerate flowering. *FLC* is a key component of the vernalization pathway in *Arabidopsis* and its expression is initially high in Arabidopsis winter annuals. The elevated expression of *FLC* prevent *Arabidopsis* floral transition before winter, an evolutionary adaptation increasing the chance of reproductive success (Michaels and Amasino, 1999) (Sheldon et al., 1999). Upon exposure to winter cold, *FLC* expression is epigenetically

repressed and this is stably maintained even after plants are returned to warmer conditions (Berry and Dean, 2015). Based on solid evidence, now it's widely accepted that *FLC* expression is modulated by post translational modification of histones before and after vernalization process (Shafiq et al., 2014; Yang et al., 2014). Detailed studies of histone marks on the chromatin of the *FLC* locus revealed correlation of histone acetylation, histone H2B ubiquitination (H2Bub1), and tri-methylation of histone H3 at lysine 4 and lysine 36 (H3K4me3/H3K36me3, respectively) with the active state of *FLC* before vernalization (Zhao et al., 2005; Pien et al., 2008). All these particular histone marks are generally associated with actively transcribing genes in both plants and yeast (Li et al., 2007). Upon vernalization, the active histone marks on the *FLC* chromatin appeared to be replaced by tri-methylation at lysine 27 of histone H3 (H3K27m3), an evolutionarily conserved hallmark of silenced genes (Bastow et al., 2004; Wood et al., 2006; De Lucia et al., 2008). Mutant studies done in *Arabidopsis* resulted in the identification of proteins (and protein complexes) that are responsible for depositing or replacing active or repressive histone marks at the *FLC* locus (He et al., 2004; Kim et al., 2005; Zhao et al., 2005; Pien et al., 2008; Tamada et al., 2009). For instance, interaction between POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) proteins and PLANT HOMEODOMAIN PROTEIN (PHD) proteins are found to be important during vernalization. At the *FLC* locus, physical interaction of a PHD-PRC2 complex is essential for maintenance of repressed state of *FLC* expression after vernalization. Together, these studies highlight the significance of histone based modification in control of *FLC* expression to regulate flowering time in *Arabidopsis*.

Besides the canonical histones, all eukaryotic organisms possess histone variants (Talbert and Henikoff, 2010b) . Expressed at a very low level, these specialized variants differ in the amino acid sequence compared to the core counterparts. With an exception to H4, variants for all the core histone proteins have been identified and a large number of histone variants belong to the H2A Family, that includes H2A.Z, H2A.X, H2A.W (Talbert and Henikoff, 2010a; Bonisch and Hake, 2012; Millar, 2013; Jiang and Berger, 2016). Incorporation of histone variants lead to changes in the chromatin landscape, affect recruitment of distinct binding partners and as such, influence gene transcription. Genome-wide chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) studies of H2A.Z variants in yeast has revealed occupancy of Z variants at the promoters of inactive genes (Guillemette et al., 2005). In contrast with the observation in yeast, in mice and humans, H2A.Z variants were found to be localized at the 5' end of many actively transcribing genes. Specific proteins (and protein-complexes) that deposit the histone variants have been unearthed in different species. For example, the evolutionary conserved H2A.Z variant is deposited by chromatin remodelling protein complexes consisting of SWR1, SRCAP, and Tip60, in yeasts, human, and fruit fly, respectively. In *Arabidopsis,* four out of the 13 H2A's are H2A.Z variants. Plant H2A.Z variants are more closely related to H2A.Z proteins of other organisms than to its own H2A Histones. Putative homologs of the SWR1/SCARP complexes, namely ACTIN-RELATED PROTEIN 6 (ARP6) and the PHOTOPERIOD-INDEPENDENT EARLY FLOWERING 1 (PIE1) have been characterized. Loss-of-function mutants of *arp6* and *pie1* display molecular and developmental phenotypes,

including early flowering, reduced leaf size and mis-regulation of floral repressors, *FLC* and *MAF* genes (Deal et al., 2007). ARP6 and PIE1 act together as part of the SWR1-like complex, in mediating deposition of H2A.Z at multiple gene loci, including *FLC*, in *Arabidopsis*. H2A.Z occupancy at the 3'end of the *FLC* gene is observed to be inversely correlated with transcript levels of *FLC (Deal et al., 2007)*.

Besides histone modifying enzymes, a novel class of enzymes has been identified that modulates rapid rearrangements in chromatin structure, i.e. chromatin remodelling ATPases. These enzymes are highly conserved in eukaryotes and are divided into four subfamilies based on the conserved ATPase domains (INO80SWR1, CHD, ISWI and SNF/SNF) (Han et al., 2015). Recent studies place a spotlight on chromatin remodelling proteins for their involvement in different protein complexes with specific implications on either gene transcription, chromatin assembly, or maintenance of higher-order chromatin structure. For instance, high resolution LC-MS/MS on immuno-precipitated CHROMATIN REMODELLING 11 and 17 (CHR11 and 17) protein complexes, co-purified MADS-domain transcription factors primarily involved in vegetative phase transition and specification of floral organ identity (Smaczniak et al., 2012). Both CHR11 and CHR17, belonging to ISWI family, directly repress expression of the flowering time gene *FT* and the floral organ identity gene *SEP3*. Single mutants of *chr11* and *chr17* flowered earlier than wild type *(Li et al., 2012)*. Analysis of the double mutant *chr11chr17* showed abnormal floral organs, confirming their function in specification and maintenance of floral organ identity via MADS-domain protein complexes (Smaczniak et al., 2012). Similarly, loss-of-function mutants of chromatin remodelers from the SWI/SNF family affect flowering time and floral morphology in *Arabidopsis* (Han et al., 2015). For example, Loss of *BRAHMA (BRM)* activity results in elevated expression of *FT* and has an early flowering phenotype under both long and short day conditions compared to wild type, suggesting *BRM* to be a floral repressor (Farrona et al., 2004; Farrona et al., 2011). The repression of *FT* by BRM is directed through *SVP*, which in turn is a direct repressor of *FT*. BRM does so by binding directly to the *SVP* locus (Li et al., 2015). In *brm* mutants, increased levels of H3K27me3 at the *SVP* locus results in decreased *SVP* expression (Li et al., 2015). In summary, these studies demonstrate the importance of chromatin remodelling factors in control of phase transitions, such as the switch to flowering.

Large-scale chromatin structures and spatial arrangements of chromatin within the nucleus are inherently linked with physiological processes. Juxtaposition of gene and regulatory sequences enable distant genomic elements to come in close physical proximity (also referred as "looping"). A diverse array of looping structures have been identified e.g. promoter-terminator contacts, promoter-enhancers contacts, boundary element interactions, and so on. Similar to studies done in animals and humans, with methods like chromosome conformation capture (3C), analysis of chromatin loops at individual loci have been performed in plants. The first observation of plant chromatin loops using 3C was at *booster1 (b1)* locus in maize (Louwers et al., 2009). Since then, many chromatin loops have been identified in the plant model species *Arabidopsis thaliana (Liu and Weigel, 2015)*.

Promoter-enhancer contacts have been reported for transcriptional regulation of the flowering time gene *FT (Cao et al., 2014)*. Sequences upstream of the *FT* transcription start site contribute to this transcriptional response. In particular, BLOCK C, an enhancer sequence ~5.3kb upstream of *FT*, physically interacts with the *FT* transcription start site. BLOCK C contains a CCAAT motif, preferably bound by Nuclear factor-Y (NF-Y) transcription factors. CONSTANS (CO), key transcriptional activator of *FT*, is shown to bind to some NF-Y proteins (Hou et al., 2014). This example clearly illustrates how chromatin looping is a result of interaction of enhancer-regulatory factor complexes with proximal-promoter sequences. On contrary, a chromatin loop at the *FT* homolog *TERMINAL FLOWER 1 (TFL1)* locus occurs in a very different fashion. A regulatory sequences downstream (3'region) of *TFL1,* interacts with the transcription start site, and is relevant for *TFL1* transcription (Liu et al., 2013). This 3' region can be bound by a complex of MADS-domain transcription factors (SOC1, SVP, AGL24 or SEP4). Upon binding of the MADS-complexes to the 3' region, the chromatin loop dissociates, and in turn, results in reduced *TFL1* expression (Liu et al., 2013). Together, these studies highlight how several factors directly or indirectly modulate chromatin architecture to precisely control the time of flowering in *Arabidopsis*.

**Scope of this thesis:**

The main aim of this thesis is understanding the molecular regulation of flowering time in *Arabidopsis thaliana*. More specifically, we focus on key regulatory genes of flowering that integrate several internal and external flowering signals and examine in detail how they are regulated at the transcriptional and post-transcriptional level. Many of the key regulatory genes encode transcription factors (TFs), which are often functioning in larger protein complexes and are part of complex gene regulatory networks. Two important regulators are the MADS-domain TFs SHORT VEGETATIVE PHASE (SVP) and SUPPRESSOR OF OVEREXPRESSION OF CO 1 (SOC1) and we studied the protein-protein interactions, chromosomal interactions and TF-DNA interactions, all connections that are part of the gene regulatory networks involved in flowering control. In **Chapter 2**, we aimed at identifying novel upstream regulators of *SOC1*. To this extent, a matrix-based yeast-one-hybrid approach is employed to screen the different promoter elements of *SOC1*. This led to identification of a few dozen of TFs that supposedly bind to *SOC1* promoter and may regulate its expression. With a focus on flowering time control, we narrowed-down the putative SOC1 regulators by performing co-expression analysis. The putative candidates that co-express with *SOC1* during the switch from vegetative to reproductive stages, were examined for expression Quantitative Trait Loci (eQTL). Overall, this chapter combines and integrates *in planta*, *in vitro* and *in silico* approaches to identify and characterize putative upstream regulators of *SOC1*. In addition to investigating *SOC1* at the transcriptional level, in this thesis we also devised strategies to study how SOC1 is regulated at the protein level. **Chapter 3** describes a robust methodological approach for isolation and identification of putative interactors in complex with TFs of interest. This chapter is the basis for all the proteomics data that we generated and later being discussed in the thesis. In **Chapter 4**, we unearth the protein complex

composition of the floral integrator SOC1. Our study identifies important interactions partners of SOC1 before and after the switch to flowering. In this chapter we follow up on one particular interaction partner, i.e. SVP, which was identified in both flowering and non-flowering developmental stages in complex with SOC1. In context of flowering, SOC1 acts as an activator and SVP as a repressor. Therefore, we question the function and significance of this activator-repressor complex. To investigate this further, we analysed previously published genome-wide ChIP seq studies for both TFs and closely examine specific and overlapping targets of each. Our data from *in vitro* Protein-DNA bindings assay (EMSA) on certain targets hint a preference for SOC1-SVP heterodimer over homodimer. Using protoplast transient assays, we measured the transcriptional response of the targets in the presence of a homo or heterodimer. Overall, our results prime the idea that combinatorial activity of these two TFs can influence the transcriptional behaviour of their target genes, enabling a certain degree of plasticity, which is essential to govern the gene regulatory functions in the floral transition process. Besides studying SOC1, here we also unravel the protein-protein interactions of AGAMOUS LIKE 24 (AGL24), another MADS-domain TF involved in floral promotion. The outcome of the both SOC1 and AGL24 proteomics dataset is furthered compared to obtain a clear insight into their independent and redundant function during the floral transition and early stages of flower development. In **Chapter 5**, we unravel the *in vivo* protein complex composition for the floral repressor SVP in two different developmental stages. We analysed and put forward the stage-specific interactome of SVP. Our *in vivo* interaction data confirms the previously identified and reported *in vitro* SVP interaction partners. In addition, we report several novel proteins, such as chromatin remodelers, co-transcriptional regulators, histone proteins, non-MADS TFs as complex partners with SVP. We further specifically investigate the role of a novel and previously uncharacterized SVP interactor, namely *POUSHP*, and underline its role in flowering time control. In **Chapter 6** & **7**, we focus on the role of chromatin organization in regulation of gene expression and describe in detail a method (3C) to study in vivo chromatin interactions. The 3C method is gaining traction among the plant research communities. In this chapter, we review most of the recent 3C studies in plants. As a case study we performed 3C analysis on *SOC1* locus and discuss the outcome (**Chapter 7**). Overall we provides detailed guidelines for the use of the 3C method and discuss intricacies of this resourceful yet challenging method. In **Chapter 8**, we discuss all the findings outlined in this thesis and provide concluding remarks on future research avenues to take this research further ahead.

## REFERENCES


**Airoldi, C.A., McKay, M., and Davies, B.** (2015). MAF2 Is Regulated by Temperature-Dependent Splicing and Represses Flowering at Low Temperatures in Parallel with FLM. PloS one **10,** e0126516.

**An, H.L., Roussot, C., Suarez-Lopez, P., Corbesler, L., Vincent, C., Pineiro, M., Hepworth, S., Mouradov, A., Justin, S., Turnbull, C., and Coupland, G.** (2004). CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. Development **131,** 3615-3626.

**Andres, F., and Coupland, G.** (2012). The genetic basis of flowering responses to seasonal cues. Nature reviews. Genetics **13,** 627-639.

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

**Babu, M.M.** (2010). Structure, evolution and dynamics of transcriptional regulatory networks. Biochemical Society transactions **38,** 1155-1178.

**Balasubramanian, S., Sureshkumar, S., Lempe, J., and Weigel, D.** (2006). Potent induction of Arabidopsis thaliana flowering by elevated growth temperature. PLoS genetics **2,** e106.

**Banerjee, A., Roychoudhury, A., and Krishnamoorthi, S.** (2016). Emerging techniques to decipher microRNAs (miRNAs) and their regulatory role in conferring abiotic stress tolerance of plants. Plant Biotechnol Rep **10,** 185-205.

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

**Bastow, R., Mylne, J.S., Lister, C., Lippman, Z., Martienssen, R.A., and Dean, C.** (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. Nature **427,** 164-167.

**Berry, S., and Dean, C.** (2015). Environmental perception and epigenetic memory: mechanistic insight through FLC. The Plant journal : for cell and molecular biology **83,** 133-148.

**Bonisch, C., and Hake, S.B.** (2012). Histone H2A variants in nucleosomes and chromatin: more or less stable? Nucleic acids research **40,** 10719-10741.

**Bouche, F., Lobet, G., Tocquin, P., and Perilleux, C.** (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. Nucleic acids research **44,** D1167-1171.

**Cao, S., Kumimoto, R.W., Gnesutta, N., Calogero, A.M., Mantovani, R., and Holt, B.F., 3rd.** (2014). A distal CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the FLOWERING LOCUS T promoter and regulates the timing of flowering in Arabidopsis. The Plant cell **26,** 1009-1017.

**Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L.** (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nature reviews. Molecular cell biology.

**Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J.** (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 angstrom resolution. Journal of molecular biology **319,** 1097-1113.

**De Lucia, F., Crevillen, P., Jones, A.M., Greb, T., and Dean, C.** (2008). A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. Proceedings of the National Academy of Sciences of the United States of America **105,** 16831-16836.

**Deal, R.B., Topp, C.N., McKinney, E.C., and Meagher, R.B.** (2007). Repression of flowering in Arabidopsis requires activation of FLOWERING LOCUS C expression by the histone variant H2A.Z. The Plant cell **19,** 74-83.

**Doll, S., and Burlingame, A.L.** (2015). Mass spectrometry-based detection and assignment of protein posttranslational modifications. ACS chemical biology **10,** 63-71.

**Farrona, S., Coupland, G., and Turck, F.** (2008). The impact of chromatin regulation on the floral transition. Seminars in cell & developmental biology **19,** 560-573.

**Farrona, S., Hurtado, L., Bowman, J.L., and Reyes, J.C.** (2004). The Arabidopsis thaliana SNF2 homolog AtBRM controls shoot development and flowering. Development **131,** 4965-4975.

**Farrona, S., Hurtado, L., March-Diaz, R., Schmitz, R.J., Florencio, F.J., Turck, F., Amasino, R.M., and Reyes, J.C.** (2011). Brahma is required for proper expression of the floral repressor FLC in Arabidopsis. PloS one **6,** e17997.

**Fornara, F., and Coupland, G.** (2009). Plant Phase Transitions Make a SPLash. Cell **138,** 625-627.

**Fornara, F., de Montaigu, A., and Coupland, G.** (2010). SnapShot: Control of flowering in Arabidopsis. Cell **141,** 550, 550 e551-552.

**Guillemette, B., Bataille, A.R., Gevry, N., Adam, M., Blanchette, M., Robert, F., and Gaudreau, L.** (2005). Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS biology **3,** e384.

**Han, S.K., Wu, M.F., Cui, S., and Wagner, D.** (2015). Roles and activities of chromatin remodeling ATPases in plants. The Plant journal : for cell and molecular biology **83,** 62-77.

**He, Y., Doyle, M.R., and Amasino, R.M.** (2004). PAF1-complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit in Arabidopsis. Genes & development **18,** 2774-2784.

**Hepworth, S.R., Valverde, F., Ravenscroft, D., Mouradov, A., and Coupland, G.** (2002). Antagonistic regulation of flowering-time gene SOC1 by CONSTANS and FLC via separate promoter motifs. The EMBO journal **21,** 4327-4337.

**Hou, X., Zhou, J., Liu, C., Liu, L., Shen, L., and Yu, H.** (2014). Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of Arabidopsis. Nature communications **5,** 4601.

**Immink, R.G., Pose, 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 physiology **160,** 433-449.

**Jiang, D., and Berger, F.** (2016). Histone variants in plant transcriptional regulation. Biochimica et biophysica acta.

**Jin, J.B., Jin, Y.H., Lee, J., Miura, K., Yoo, C.Y., Kim, W.Y., Van Oosten, M., Hyun, Y., Somers, D.E., Lee, I., Yun, D.J., Bressan, R.A., and Hasegawa, P.M.** (2008). The SUMO E3 ligase, AtSIZ1, regulates flowering by controlling a salicylic acid-mediated floral promotion pathway and through affects on FLC chromatin structure. The Plant journal : for cell and molecular biology **53,** 530-540.

**Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B.** (2006). MicroRNAs and their regulatory roles in plants. Annual review of plant biology **57,** 19-53.

**Jung, J.H., Seo, P.J., and Park, C.M.** (2012). The E3 ubiquitin ligase HOS1 regulates Arabidopsis flowering by mediating CONSTANS degradation under cold stress. The Journal of biological chemistry **287,** 43277-43287.

**Jung, J.H., Seo, Y.H., Seo, P.J., Reyes, J.L., Yun, J., Chua, N.H., and Park, C.M.** (2007). The GIGANTEA-regulated microRNA172 mediates photoperiodic flowering independent of CONSTANS in Arabidopsis. The Plant cell **19,** 2736-2748.

**Kamakaka, R.T., and Biggins, S.** (2005). Histone variants: deviants? Genes & development **19,** 295-310.

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

**Kawashima, T., and Berger, F.** (2014). Epigenetic reprogramming in plant sexual reproduction. Nature Reviews Genetics **15,** 613-624.

**Kim, S.Y., He, Y., Jacob, Y., Noh, Y.S., Michaels, S., and Amasino, R.** (2005). Establishment of the vernalization-responsive, winter-annual habit in Arabidopsis requires a putative histone H3 methyl transferase. The Plant cell **17,** 3301-3310.

**Kouzarides, T.** (2007). Chromatin modifications and their function. Cell **128,** 693-705.

**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.

**Lee, J., and Lee, I.** (2010). Regulation and function of SOC1, a flowering pathway integrator. Journal of experimental botany **61,** 2247-2254.

**Lee, J.H., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.S., and Ahn, J.H.** (2007). Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes & development **21,** 397-402.

**Lee, J.H., Ryu, H.S., Chung, K.S., Pose, D., Kim, S., Schmid, M., and Ahn, J.H.** (2013). Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. Science **342,** 628-632.

**Lenhard, B., Sandelin, A., and Carninci, P.** (2012). REGULATORY ELEMENTS Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature Reviews Genetics **13,** 233-245.

**Li, B., Carey, M., and Workman, J.L.** (2007). The role of chromatin during transcription. Cell **128,** 707-719.

**Li, C., Chen, C., Gao, L., Yang, S., Nguyen, V., Shi, X., Siminovitch, K., Kohalmi, S.E., Huang, S., Wu, K., Chen, X., and Cui, Y.** (2015). The Arabidopsis SWI2/SNF2 chromatin Remodeler BRAHMA regulates polycomb function during vegetative development and directly activates the flowering repressor gene SVP. PLoS genetics **11,** e1004944.

**Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E., and Yu, H.** (2008). A repressor complex governs the integration of flowering signals in Arabidopsis. Developmental cell **15,** 110-120.

**Li, G., Zhang, J., Li, J., Yang, Z., Huang, H., and Xu, L.** (2012). Imitation Switch chromatin remodeling factors and their interacting RINGLET proteins act together in controlling the plant vegetative phase in Arabidopsis. The Plant journal : for cell and molecular biology **72,** 261-270.

**Liu, C., and Weigel, D.** (2015). Chromatin in 3D: progress and prospects for plants. Genome biology **16,** 170.

**Liu, C., Chen, H., Er, H.L., Soo, H.M., Kumar, P.P., Han, J.H., Liou, Y.C., and Yu, H.** (2008). Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis. Development **135,** 1481-1491.

**Liu, C., Teo, Z.W., Bi, Y., Song, S., Xi, W., Yang, X., Yin, Z., and Yu, H.** (2013). A conserved genetic pathway determines inflorescence architecture in Arabidopsis and rice. Developmental cell **24,** 612-622.

**Liu, C.Y., Lu, F.L., Cui, X., and Cao, X.F.** (2010a). Histone Methylation in Higher Plants. Annual Review of Plant Biology, Vol 61 **61,** 395-420.

**Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C.** (2010b). Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. Science **327,** 94-97.

**Louwers, M., Bader, R., Haring, M., van Driel, R., de Laat, W., and Stam, M.** (2009). Tissue- and expression level-specific chromatin looping at maize b1 epialleles. The Plant cell **21,** 832-842.

**Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T.** (2011). Epigenetics in Alternative Pre-mRNA Splicing. Cell **144,** 16-26.

**Luger, K., Mader, A., Sargent, D.F., and Richmond, T.J.** (2000). The atomic structure of the nucleosome core particle. Journal of biomolecular structure & dynamics**,** 185-188.

**Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J.** (1997). Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. Nature **389,** 251-260.

**Lynch, M., and Conery, J.S.** (2003). The origins of genome complexity. Science **302,** 1401-1404.

**Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S.** (2002). The protein kinase complement of the human genome. Science **298,** 1912-1934.

**Martinez, G., and Kohler, C.** (2017). Role of small RNAs in epigenetic reprogramming during plant sexual reproduction. Current opinion in plant biology **36,** 22-28.

**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.

**Michaels, S.D., Ditta, G., Gustafson-Brown, C., Pelaz, S., Yanofsky, M., and Amasino, R.M.** (2003). AGL24 acts as a promoter of flowering in Arabidopsis and is positively regulated by vernalization. The Plant journal : for cell and molecular biology **33,** 867-874.

**Millar, C.B.** (2013). Organizing the genome with H2A histone variants. Biochemical Journal **449,** 567-579.

**Moon, J., Suh, S.S., Lee, H., Choi, K.R., Hong, C.B., Paek, N.C., Kim, S.G., and Lee, I.** (2003). The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. The Plant journal : for cell and molecular biology **35,** 613-623.

**Murtas, G., Reeves, P.H., Fu, Y.F., Bancroft, I., Dean, C., and Coupland, G.** (2003). A nuclear protease required for flowering-time regulation in Arabidopsis reduces the abundance of SMALL UBIQUITIN-RELATED MODIFIER conjugates. The Plant cell **15,** 2308-2319.

**Ohler, U., and Wassarman, D.A.** (2010). Promoting developmental transcription. Development **137,** 15-26.

**Onouchi, H., Igeno, M.I., Perilleux, C., Graves, K., and Coupland, G.** (2000). Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among Arabidopsis flowering-time genes. The Plant cell **12,** 885-900.

**Pagel, O., Loroch, S., Sickmann, A., and Zahedi, R.P.** (2015). Current strategies and findings in clinically relevant post-translational modification-specific proteomics. Expert review of proteomics **12,** 235-253.

**Parcy, F.** (2005). Flowering: a time for integration. The International journal of developmental biology **49,** 585-593.

**Pfluger, J., and Wagner, D.** (2007). Histone modifications and dynamic regulation of genome accessibility in plants. Current opinion in plant biology **10,** 645-652.

**Pien, S., Fleury, D., Mylne, J.S., Crevillen, P., Inze, D., Avramova, Z., Dean, C., and Grossniklaus, U.** (2008). ARABIDOPSIS TRITHORAX1 dynamically regulates FLOWERING LOCUS C activation via histone 3 lysine 4 trimethylation. The Plant cell **20,** 580-588.

**Pose, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G.C., Immink, R.G., and Schmid, M.** (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. Nature **503,** 414-417.

**Putterill, J., Laurie, R., and Macknight, R.** (2004). It's time to flower: the genetic control of flowering time. BioEssays : news and reviews in molecular, cellular and developmental biology **26,** 363-373.

**Putterill, J., Robson, F., Lee, K., Simon, R., and Coupland, G.** (1995). The Constans Gene of Arabidopsis Promotes Flowering and Encodes a Protein Showing Similarities to Zinc-Finger Transcription Factors. Cell **80,** 847-857.

**Reeves, P.H., Murtas, G., Dash, S., and Coupland, G.** (2002). early in short days 4, a mutation in Arabidopsis that causes early flowering and reduces the mRNA abundance of the floral repressor FLC. Development **129,** 5349-5361.

**Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C.Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K., and Yu, C.L.** (2000). Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. Science **290,** 2105-2110.

**Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., and Coupland, G.** (2000). Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. Science **288,** 1613-1616.

**Sarid-Krebs, L., Panigrahi, K.C., Fornara, F., Takahashi, Y., Hayama, R., Jang, S., Tilmes, V., Valverde, F., and Coupland, G.** (2015). Phosphorylation of CONSTANS and its COP1-dependent degradation during photoperiodic flowering of Arabidopsis. The Plant journal : for cell and molecular biology **84,** 451-463.

**Scheres, B.** (2007). Stem-cell niches: nursery rhymes across kingdoms. Nat Rev Mol Cell Bio **8,** 345-354.

**Schnell, J.D., and Hicke, L.** (2003). Non-traditional functions of ubiquitin and ubiquitin-binding proteins. Journal of Biological Chemistry **278,** 35857-35860.

**Scott, M.P.** (2000). Development: The natural history of genes. Cell **100,** 27-40.

**Shafiq, S., Berr, A., and Shen, W.H.** (2014). Combinatorial functions of diverse histone methylations in Arabidopsis thaliana flowering time regulation. The New phytologist **201,** 312-322.

**Sheldon, C.C., Burn, J.E., Perez, P.P., Metzger, J., Edwards, J.A., Peacock, W.J., and Dennis, E.S.** (1999). The FLF MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. The Plant cell **11,** 445-458.

**Simpson, G.G., and Dean, C.** (2002). Arabidopsis, the Rosetta stone of flowering time? Science **296,** 285-289.

**Smaczniak, C., Immink, R.G., Muino, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Parcy, F., Xu, L., Carles, C.C., Angenent, G.C., and Kaufmann, K.** (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proceedings of the National Academy of Sciences of the United States of America **109,** 1560-1565.

**Spanudakis, E., and Jackson, S.** (2014). The role of microRNAs in the control of flowering time. Journal of experimental botany **65,** 365-380.

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

**Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G.** (2001). CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature **410,** 1116-1120.

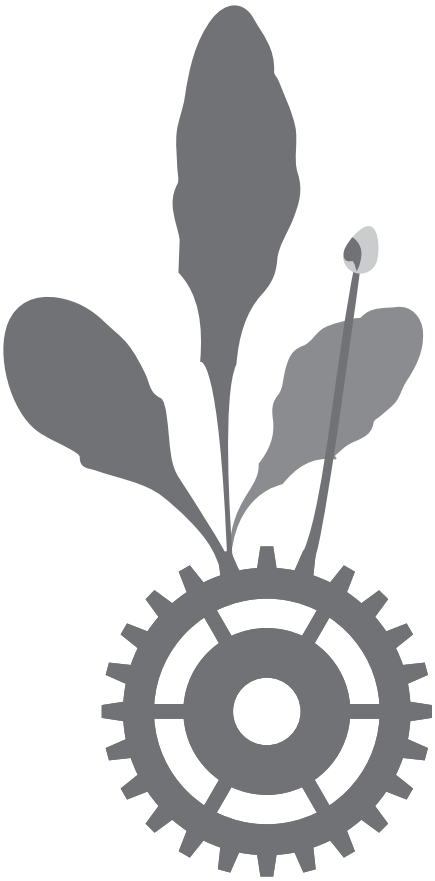
**Sun, L.J., and Chen, Z.J.** (2004). The novel functions of ubiquitination in signaling. Current opinion in cell biology **16,** 119-126.

**Sureshkumar, S., Dent, C., Seleznev, A., Tasset, C., and Balasubramanian, S.** (2016). Nonsense-mediated mRNA decay modulates FLM-dependent thermosensory flowering response in Arabidopsis. Nature plants **2,** 16055.

**Swiezewski, S., Liu, F., Magusin, A., and Dean, C.** (2009). Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. Nature **462,** 799-802.
**Talbert, P.B., and Henikoff, S.** (2010a). Histone variants - ancient wrap artists of the epigenome. Nat Rev Mol Cell Bio **11,** 264-275.
**Talbert, P.B., and Henikoff, S.** (2010b). Histone variants--ancient wrap artists of the epigenome. Nature reviews. Molecular cell biology **11,** 264-275.
**Tamada, Y., Yun, J.Y., Woo, S.C., and Amasino, R.M.** (2009). ARABIDOPSIS TRITHORAX-RELATED7 is required for methylation of lysine 4 of histone H3 and for transcriptional activation of FLOWERING LOCUS C. The Plant cell **21,** 3257-3269.
**Teotia, S., and Tang, G.** (2015). To bloom or not to bloom: role of microRNAs in plant flowering. Molecular plant **8,** 359-377.
**Torti, S., Fornara, F., Vincent, C., Andres, F., Nordstrom, K., Gobel, U., Knoll, D., Schoof, H., and Coupland, G.** (2012). Analysis of the Arabidopsis shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. The Plant cell **24,** 444-462.
**Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., and Coupland, G.** (2004a). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science **303,** 1003-1006.
**Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., and Coupland, G.** (2004b). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science **303,** 1003-1006.
**Wang, F., and Deng, X.W.** (2011). Plant ubiquitin-proteasome pathway and its role in gibberellin signaling. Cell Res **21,** 1286-1294.
**Wang, H., Chung, P.J., Liu, J., Jang, I.C., Kean, M.J., Xu, J., and Chua, N.H.** (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in Arabidopsis. Genome research **24,** 444-453.
**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.
**Wood, C.C., Robertson, M., Tanner, G., Peacock, W.J., Dennis, E.S., and Helliwell, C.A.** (2006). The Arabidopsis thaliana vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. Proceedings of the National Academy of Sciences of the United States of America **103,** 14631-14636.
**Wu, G., and Poethig, R.S.** (2006). Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3. Development **133,** 3539-3547.
**Wu, G., Park, M.Y., Conway, S.R., Wang, J.W., Weigel, D., and Poethig, R.S.** (2009). The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis. Cell **138,** 750-759.
**Xu, D., Zhu, D., and Deng, X.W.** (2016). The role of COP1 in repression of photoperiodic flowering. F1000Research **5**.
**Xu, X.M., Rose, A., Muthuswamy, S., Jeong, S.Y., Venkatakrishnan, S., Zhao, Q., and Meier, I.** (2007). NUCLEAR PORE ANCHOR, the Arabidopsis homolog of Tpr/Mlp1/Mlp2/megator, is involved in mRNA export and SUMO homeostasis and affects diverse aspects of plant development. The Plant cell **19,** 1537-1548.
**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.
**Yang, H., Howard, M., and Dean, C.** (2014). Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at Arabidopsis FLC. Current biology : CB **24,** 1793-1797.
**Yu, H., Xu, Y., Tan, E.L., and Kumar, P.P.** (2002). AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals. Proceedings of the National Academy of Sciences of the United States of America **99,** 16336-16341.
**Zhang, B., Pan, X., Cobb, G.P., and Anderson, T.A.** (2006). Plant microRNA: a small regulatory molecule with big impact. Developmental biology **289,** 3-16.
**Zhang, B.H., Wang, Q.L., and Pan, X.P.** (2007). MicroRNAs and their regulatory roles in animals and plants. Journal of cellular physiology **210,** 279-289.
**Zhang, X.** (2008). The epigenetic landscape of plants. Science **320,** 489-492.
**Zhao, Z., Yu, Y., Meyer, D., Wu, C., and Shen, W.H.** (2005). Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. Nature cell biology **7,** 1256-1260.
**Zulawski, M., Braginets, R., and Schulze, W.X.** (2013). PhosPhAt goes kinases--searchable protein kinase target information in the plant phosphorylation site database PhosPhAt. Nucleic acids research **41,** D1176-1184.

# CHAPTER 2

## Spotlight on potential upstream regulators of *SOC1* in flowering time control



Suraj B. Jamge
Francesca Bellinazzo
Aalt D.J. van Dijk
Gerco C. Angenent
Richard G.H. Immink

**ABSTRACT**

*SOC1* encodes a MADS-domain transcription factor that integrates endogenous and environmental signals controlling flowering time in *Arabidopsis thaliana* and acts as a hub in the gene regulatory network initiating flowering. Likewise, it is involved in flower development, stomatal opening, cold tolerance, plant growth and longevity and other developmental functions. Hence, spatiotemporal control of SOC1's expression is key to ensure its multifaceted role. In the present study, we identified and investigated potential upstream regulators of *SOC1.* Using *in silico* and yeast-based approaches, we infer key potential regulators of *SOC1* in flowering time control.

***INTRODUCTION***

Plants are fixed in one place and cannot escape their environment. In order to optimize their chance of survival, they sense environmental cues and adapt their development to these circumstances. To flower is a major developmental decision in the life cycle of flowering plants and this switch from vegetative to reproductive development is controlled by an intricate network of genes and regulatory pathways. Major flowering pathways include photoperiod, vernalization, gibberellic acid (GA), ambient temperature, aging and the autonomous pathway. Signals from these distinct pathways come together at the level of the so-called floral integrator genes, which ultimately initiate flowering (Simpson and Dean, 2002; Boss et al., 2004; Putterill et al., 2004; Parcy, 2005; Baurle and Dean, 2006; Fornara et al., 2010; Bouche et al., 2016).

In *Arabidopsis*, *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) is an important floral integrator gene, acting downstream of *FLOWERING LOCUS T (FT) (Borner et al., 2000; Lee et al., 2000; Onouchi et al., 2000; Samach et al., 2000; Wigge et al., 2005; Yoo et al., 2005; Lee and Lee, 2010)*. *SOC1* is widely expressed during plant development. During vegetative growth, *SOC1's* expression gradually increases in leaves and the shoot apical meristem (SAM), until it reaches a particular threshold in the SAM which marks the onset of flowering (Immink et al., 2012). *SOC1* encodes a MADS-domain protein and promotes the transition to flowering in response to GA, Long Day (LD) photoperiod and the autonomous pathway (Lee and Lee, 2010).

During the floral transition process, the transcriptional regulation of *SOC1* is known to be shaped by key players from different flowering pathways (Lee and Lee, 2010). In the photoperiod pathway, under inductive LD conditions, CONSTANS upregulates *SOC1* expression through the FT protein (Hepworth et al., 2002; Wigge et al., 2005; Yoo et al., 2005). Under non-inductive Short Day (SD) conditions *SOC1* induction is delayed, but ultimately, its expression is positively regulated by GA signalling (Moon et al., 2003). In the age-dependent pathway, the action of miRNA156-targeted *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) genes result in *SOC1* upregulation (Wang et al., 2009; Xu et al., 2016). Eleven out of the seventeen *SPLs* genes in *Arabidopsis* are targeted by miR156 (Wang and Wang, 2015), of which *SPL2, SPL9, SPL11, SPL13* and *SPL15* directly promote *SOC1* transcription *(Xu et al., 2016)*. In addition to these pathways, signals from the autonomous and vernalization pathway negatively regulate *SOC1* via the repressive action of the MADS-domain transcription factors (TFs) SHORT VEGETATIVE PHASE (SVP) and FLOWERING LOCUS C (FLC) (Michaels and Amasino, 1999; Hartmann et al., 2000; Hepworth et al., 2002; Searle et al., 2006; Lee et al., 2007; Li et al., 2008; Seo et al., 2009; Deng et al., 2011). Both FLC and SVP can bind to *SOC1* regulatory sequences and repress its expression (Helliwell et al., 2006; Li et al., 2008; Tao et al., 2012; Gregis et al., 2013; Mateos et al., 2015). Furthermore, SOC1 can repress its own expression by binding to its 5' untranslated region (UTR) (Immink et al., 2012). Analysis of the 5'UTR and ~1kb upstream of the *SOC1* transcription start site (TSS) resulted in the identification of seven

conserved putative CArG motifs, which represent MADS-domain TF binding motifs (Immink et al., 2012). Since MADS-domain proteins can bind DNA as dimers, Immink et al further studied the binding of Type II MADS proteins to these regulatory sequences and identified binding by several MADS dimers containing proteins with different functions. For instance, protein dimers involved in floral transition, such as SVP-AGAMOUS-LIKE15 (AGL15) and FRUITFUL (FUL)-SOC1, and floral organ specification, such as AGAMOUS (AG)-SEPALATA3 (SEP3). Additionally, dimers consisting of proteins involved in floral transition and floral organ development (e.g. AGL24-AP1 and SOC1-SEP3) were identified, suggesting a role for a large number of MADS domain proteins in *SOC1* regulation during both the floral transition and subsequent flower development (Immink et al., 2012). In addition to transcriptional control, *SOC1* is also regulated at the post-transcriptional level through the action of the RNA binding protein EARLY FLOWERING 9 (ELF9). ELF9 reduces *SOC1* transcripts through nonsense-mediated mRNA decay (NMD) (Song et al., 2009). At the protein level, PIN1-type parvulin (Pin1At), a PPIase enzyme involved in cis/trans-isomerization of phosphorylated Ser/Thr-Pro motifs, can influence SOC1 activity (Wang et al., 2010).

Despite being a central player in the flowering network, SOC1 function is not limited to flowering time control. Recent reports suggest SOC1 has a role in floral patterning, floral meristem identity, cold tolerance, stomatal opening, plant growth and longevity, greening, dark-induced chlorophyll degradation and leaf senescence (Melzer et al., 2008; Liu et al., 2009; Seo et al., 2009; Richter et al., 2013; Kimura et al., 2015; Davin et al., 2016; Chen et al., 2017). This multifaceted role of SOC1 and its prominence in flowering time regulation demands tight control of its expression and function. To gain new insights into how *SOC1* is regulated, therefore in this study we aimed at identifying novel direct upstream regulators of *SOC1*. A yeast one-hybrid (Y1H) approach was implemented that identified a plethora of potential *SOC1* regulators. A meta-analysis approach was used to obtain additional evidence of these putative upstream regulators of *SOC1* in flowering time control.

## *RESULTS*

### Putative cis-regulatory elements present at the *SOC1* promoter

Gene promoters are not highly conserved, except for short stretches of sequence bound by TFs, known as cis-elements. In order to identify TFs that can directly control *SOC1* expression, we first examined the promoter sequence of *SOC1* for the presence of known cis-regulatory elements. The publically available Plant Transcription Factor DataBase (PlantTFDB 4.0) (Jin et al., 2015; Jin et al., 2017), via the PlantRegMap webtool, was used for the detection of transcription factor binding sites (TFBSs). The PlantTFDB is composed of high quality, manually curated and non-redundant TF motifs that were experimentally derived from >150 species. Using a FIMO algorithm (Grant et al., 2011), the PlantRegMap tool reports TFBSs for a given input DNA sequence and also predicts potential TF that binds to these sites. In this analysis we used a ~1 kb sequence upstream of *SOC1's* transcription start site (TSS). Based on empirical

evidence, the majority of proteins involved in transcription control are known to bind in this region in *Arabidopsis* (Yu et al., 2016). Furthermore, we also included the 5'UTR region of *SOC1*, which was shown to be important in auto-regulatory feedback (Immink et al., 2012). PlantRegMap analysis identified 330 binding sites for 131 TFs at a default threshold p-value of ≤1e-4. The outcome represented a diverse range of cis-regulatory motifs (Supplemental File S1). Potential binding sites for TFs belonging to 27 different families such as MADS, bZIP, MADS, C2H2, bHLH , Dof, MYB and MYB-related (Figure 1) were reported.

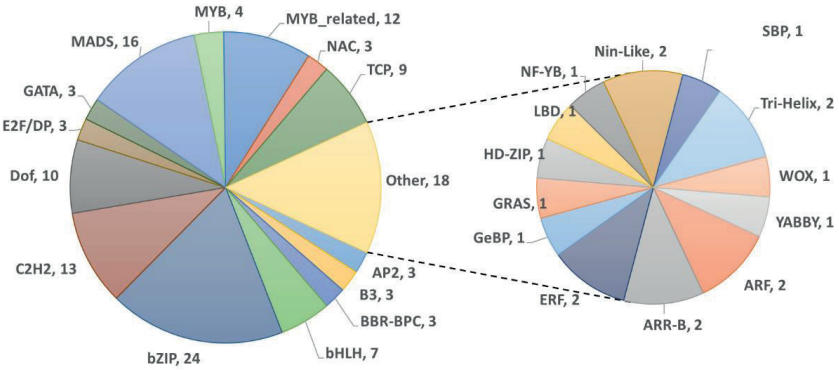



Figure 1. **Overview of TF families with putative TFBSs at *SOC1's* 1kb promoter and 5'UTR sequence.** Pie chart highlighting the predicted 131 TFs associated with at least one TFBS at the SOC1 locus using PlantRegMap tool, organized according to their representative TF families. In total 330 TFBSs were identified spanning both strands. For the complete list predicted TFs and their associated binding sites please refer to Supplemental file S1.

### *SOC1* is targeted by multiple TF families

The *in silico* analysis of publicly available database for putative TFBSs in *SOC1* regulatory sequences resulted in an large number of potential binding motifs for a variety of TFs families (Figure 1). However, our knowledge of TF binding sites is rather limited and known consensus binding sites are short and degenerate in nature (O'Malley et al., 2016). Consequently, the predictive power based solely on the presence of short sequence motifs is limited. Furthermore, most plant TFs belong to large families, therefore identifying a consensus cis-element for a particular family of TFs is only a first step towards the identification of binding by a specific TF. Taking this knowledge into account, we decided to apply the yeast-one-hybrid (Y1H) technique in order to identify potential *SOC1*-binding TFs. A matrix based Y1H approach was used to screen a TF library comprising of over 1200 Arabidopsis TFs (Castrillo et al., 2011). This library was screened with three previously isolated *SOC1* sequences spanning the 5'-UTR and ~1kb upstream of the *SOC1* TSS ((Immink et al., 2012); Figure 2a). The screen was performed in duplicate and identified binding of >90 TFs to the *SOC1* promoter (Figure 2b, Supplemental Table 1). Among others, we identified CONSTANS (CO), a previously reported direct transcriptional regulator of *SOC1*, revealing the relevance and quality of our study.

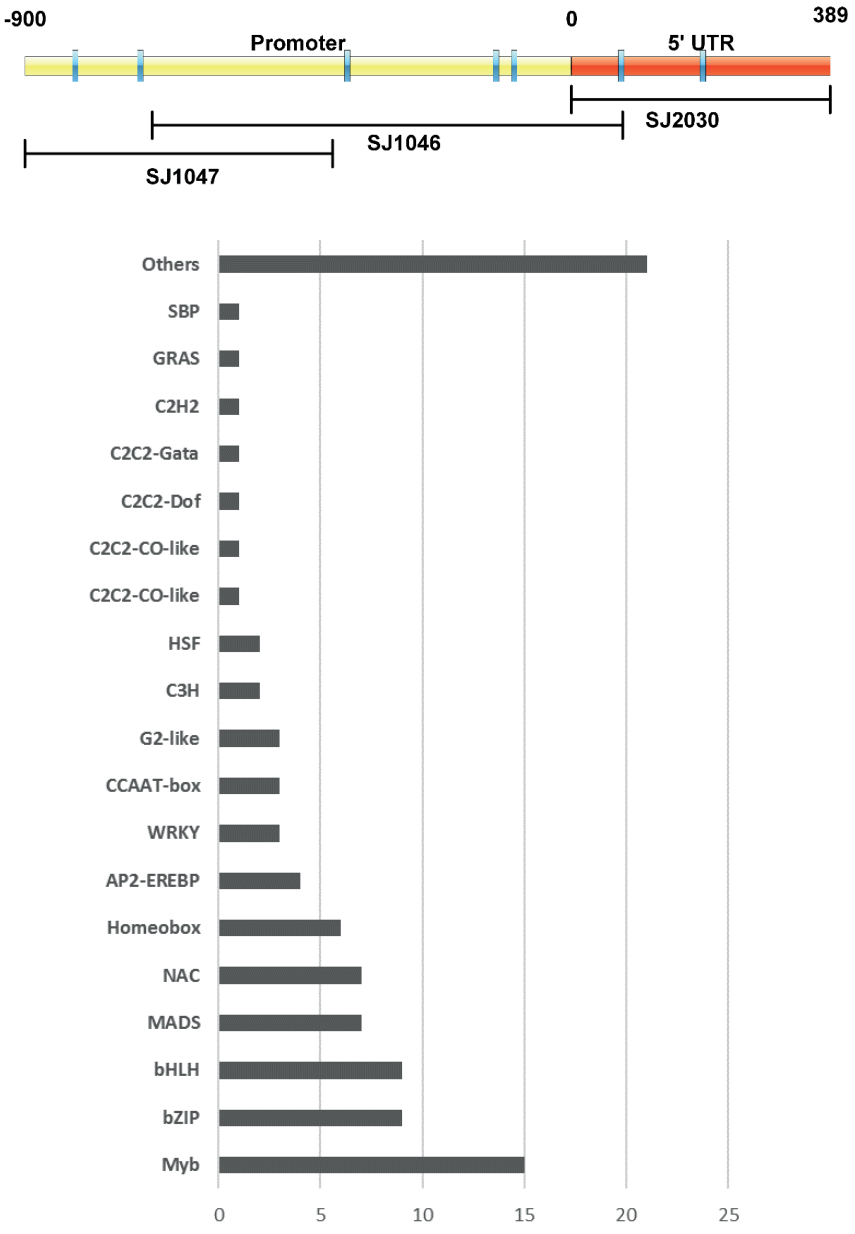



**Figure 2.** *SOC1* genomic region and identified putative regulators**.** A) Schematic representation of the *Arabidopsis SOC1* promoter and 5'UTR region. The upstream promoter region is shown in yellow and the 5'UTR region is indicated in orange. The positions of seven previously described CArG motifs is indicated in blue. B) Graph represents the type and number of transcription factors identified in a Y1H assay using the *SOC1* promoter and UTR regions as indicated in (A).

### *SOC1* is co-expressed with some of its putative transcriptional regulators

Our Y1H assay identified a large and diverse set of TFs that bind upstream of *SOC1* and may potentially regulate *SOC1* expression. However, binding of TFs to regulatory elements does not always lead to changes in gene expression and besides this, the Y1H approach fails to capture the biological context in which the TF binding occurs and will probably yield false positive interactors as well. Since *SOC1* has a major role during the floral transition process, we asked whether any of the above identified TFs are co-expressed with *SOC1* in the SAM and thus may be relevant as a positive *SOC1* regulator during the flowering process. For our co-expression analysis we utilized a recently published RNA-seq time series expression study in *Arabidopsis*, which comprises gene expression profiles from germination up-to the moment of floral transition (Klepikova et al., 2015). We specifically assessed co-expression in the SAM at different time points (before and after transition), where *SOC1* expression is known to be of significance for floral transition (Immink et al., 2012). Upon performing the co-expression analysis for all the identified TFs together with SOC1, the co-expression analysis yielded a Pearson correlation value between -1 (perfect negative correlation) and 1 (perfect positive correlation). TFs that activate *SOC1* expression are expected to be positively correlated, whereas TF expression that lead to *SOC1* repression should appear negatively correlated. A correlation cut-off of + 0.6 or - 0.6 was used to define strong positive or negative co-expression, respectively. Approximately 8% of all the identified putative upstream regulators identified by Y1H are positively correlated with *SOC1* expression (Table 1). *CO* and *SPL15,* for example are positively correlated (r>0.60), which is in line with the positive effects of these TFs on flowering time control (Yoo et al., 2005; Hyun et al., 2016; Xu et al., 2016). We also observed a strong negative correlation for 13% of the identified TF genes. These negative correlated genes included *NAC6*, *WRKY6* and *bZIP*42, among others (Table 1).

| Candidate AGI | Gene Name | TF Family Name | Binding Region See Fig.2A | Pearson Correlation (r) |
|---|---|---|---|---|
| AT2G45660 | SOC1 | MADS | NA | **1.000** |
| AT3G57920 | AtSPL15 | SBP | UTR | **0.905** |
| AT5G62610 | bHLH79 | bHLH | UTR | **0.733** |
| AT4G08150 | KNAT1 | Homeobox | UTR | **0.728** |
| AT1G72210 | AtbHLH96 | bHLH | 1047 | **0.654** |
| AT1G06850 | AtbZIP52 | bZIP | 1046, 1047 | **0.618** |
| AT2G23290 | AtMYB70 | MYB | 1047 | **0.610** |
| AT1G26780 | AtMYB117 | MYB | UTR/ 1047 | **0.608** |
| AT5G15840 | CO | C2C2-Co-like | UTR | **0.605** |
| AT5G39610 | ORE1 | NAC | UTR | **-0.610** |
| AT1G19000 | - | - | 1047 | **-0.622** |
| AT3G30530 | AtbZIP42 | bZIP | UTR | **-0.700** |
| AT4G36020 | AtCSP1 | - | UTR | **-0.702** |
| AT5G01200 | - | Homeodomain-like | 1046, 1047 | **-0.710** |
| AT1G62300 | AtWRKY6 | WRKY | 1046, 1047 | **-0.725** |
| AT4G27950 | CRF2 | AP2-EREBP | 1047 | **-0.814** |
| AT5G13180 | NAC83, VNI2 | NAC | UTR | **-0.854** |
| AT5G07690 | AtMYB29 | MYB | 1046, 1047 | **-0.857** |
| AT1G13600 | AtbZIP58 | bZIP | UTR | **-0.879** |
| AT5G59780 | AtMYB59 | MYB | 1046, 1047 | **-0.907** |
| AT3G17100 | AtbHLH147 | bHLH | 1047 | **-0.936** |
| AT4G36710 | SCL15 | GRAS | UTR | **-0.936** |

Table 1. **Co-expression analysis between *SOC1* and putative upstream regulators identified by Y1H.** Candidate transcription factor genes with a Spearman correlation coefficient of > 0.6 or <- 0.6 are represented.

### Correlating eQTLs observed for *SOC1* and its transcriptional regulators

Transition to flowering is a complex developmental trait and quantitative in nature. Many Quantitative trait locus (QTL) mapping studies have been performed in relation to the flowering time trait (El-Din El-Assal et al., 2001; Balasubramanian et al., 2006; Filiault et al., 2008; Caicedo et al., 2009; Schwartz et al., 2009; Rosloski et al., 2010; Salome et al., 2011), associating genomic loci with phenotypic variations. A rapid increase in availability of genome-wide gene expression data has enabled the incorporation of transcriptomic variation in QTL studies, thereby identifying expression QTLs (eQTLs). eQTLs are genomic regions that directly influence the expression of one or more genes. In the framework of this study, eQTL analysis was used to further strengthen the experimental data obtained from the Y1H and co-expression studies. eQTL profiles for *SOC1* were examined across different eQTL studies using the araQTL workbench (Figure 3a; (Nijveen et al., 2017)). Multiple significant co-occurring trans-eQTLs were observed within chromosome one, four and five (Figure 3a), hinting at variation in these genomic regions being of importance for differential *SOC1* expression.

Consequently, we further examined if any of the putative regulators of *SOC1,* initially identified through our Y1H screenings, are located within one of the significant *SOC1* eQTL regions. KNAT4, AtMYB29 and bHLH12 (Table 2) were located in significant eQTL genomic regions, providing additional support for their proposed role in *SOC1* regulation.

As a second approach, we performed a closer examination of the eQTL peak at chromosome 4 from the study by West et al 2007. We investigated which other genes share an eQTL peak at this genomic region (Figure 3b). At a LOD score cut-off of 3, 282 genes were found with an eQTL associated with this genomic region on chromosome 4. Among them, we identified flowering related genes such as *FLC* (LOD 56.74), *FRIGIDA (FRI)* (LOD 25.30), *AGL42* (LOD 16.35), and *FUL (AGL8)* (LOD 4.50). *AGL42*, is a *SOC1*-like gene that acts redundantly with two other *SOC1*-like genes (*AGL71*, *AGL72*) and are of importance during the floral transition process in axillary meristems (AMs) (Dorca-Fornell et al., 2011). Their activity overlaps with *SOC1* function in AMs. Similarly, *FUL* is also described to be partially redundant with *SOC1* in promotion of flowering (Melzer et al., 2008). These few examples demonstrate that other genes partially acting at the same position in the flowering time gene regulatory network as *SOC1* and possibly regulated by the same locus underlying the eQTL on chromosome 4.

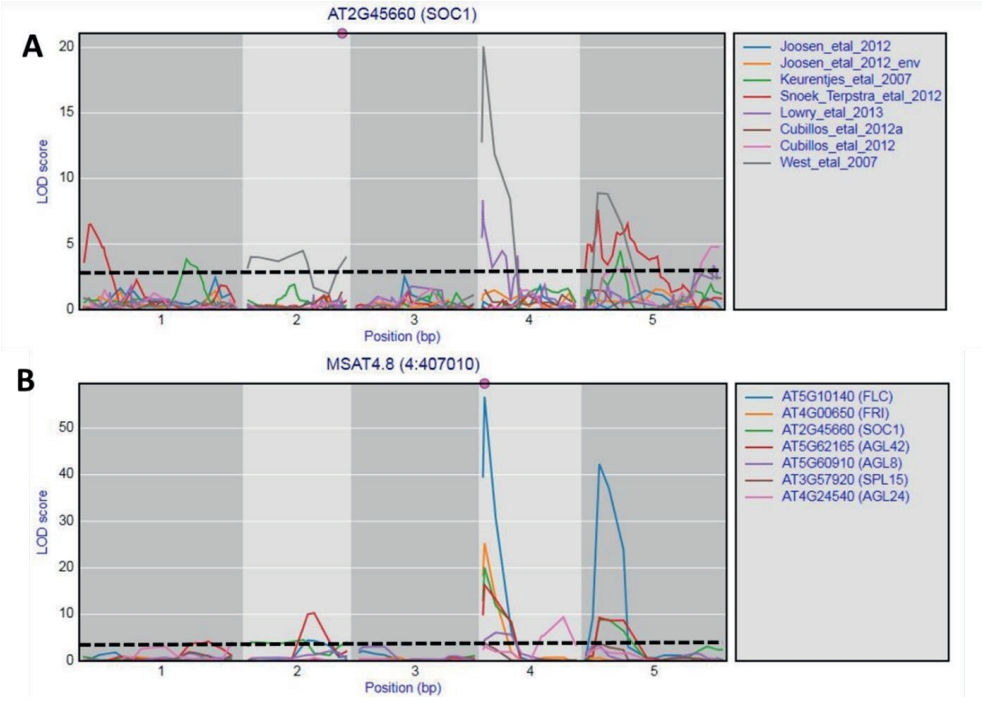



Figure 3. **Visualization of eQTL profiles** A) Snapshot of *SOC1's* eQTL profile from allstudies present in the AraQTL workbench. In total three prominent peaks are observed on chromosome 1, 4 and 5. B) Closer examination of the most prominent *SOC1* peak from west et al 2007 at chromosome 4. The eQTL profile of seven flowering related genes peaking at this genomic region have been highlighted. The similarity of the profiles hint at shared regulation by factors positioned at the beginning of chromosome 4.

| TF | Binding region | Correlation | eQTL | Flowering time |
|---|---|---|---|---|
| **KNAT4** | UTR | -0.02 | yes<br>Snoek Terpstra et al 2012 | NA |
| **AtMYB29** | Promoter | -0.86 | Yes<br>West et al 2007 | NA |
| **bHLH12** | UTR | 0.01 | Yes<br>Lowry et al 2013 | NA |

Table 2. **Overview of potential upstream regulators of *SOC1* with eQTL support.** TFs with a correlation coefficient of at least + 0.6 or - 0.6 are referred as positive or negative correlation. Literature search for any reported flowering time phenotype for the TF mutants was performed. NA refers to no flowering time data available in present literature.

***DISCUSSION***

*SOC1* is a major floral integrator, incorporating signals from various flowering-time pathways into flowering initiation. Besides floral transition, *SOC1* is able to fulfil many other developmental functions, including its role during flower development, stomatal opening, and dark-induced senescence. TFs with such multifaceted functions demand a tight spatial-temporal control of gene expression. In this study, we provide novel insights into transcriptional regulation of *SOC1* during the floral transition process. We used a combination of *in silico* and wet lab approaches to identify and further characterize potential upstream regulators of *SOC1*.

Direct recognition of DNA elements by regulatory proteins such as TFs is a well-established concept. For example, MADS-domain TFs bind to a conserved sequence motif (CC[A/T]$_6$GG) called a CArG-box. Seven CArG motifs have been identified in the *SOC1* promoter sequence (Immink et al., 2012). Several MADS-domain TF dimers were identified as potential interactors of these elements and binding at the third CArG site appeared to be biological relevant in limiting *SOC1* expression in vegetative (Hepworth et al., 2002) and floral tissues (Immink et al., 2012). Here we systematically investigated the upstream sequence of *SOC1* and demonstrated the presence of a large number of other putative TFBSs (Figure 1), implying regulation of *SOC1* by a plethora of different TFs. However, the reliability of such *in silico* analysis is fairly limited. Since the algorithms that assess motifs and predict binding sites largely depend on available knowledge of TF binding sites. In recent years there have been major improvements in methods that capture *in vivo* TF occupancy and our knowledge on the binding specificity of different TFs is growing at a rapid pace. Nonetheless, with the available methodologies, identifying and assessing TF-DNA interactions that are transient and dynamic (time-specific) still remain a challenge. In spite of these limitations, *in silico* prediction of binding sites is informative, relatively easy to perform and represents a good starting point for exploring transcriptional regulation of target genes.

A wide variety of TF motifs upstream of *SOC1* was found and the performed Y1H assay, in which binding of >1200 TFs at *SOC1's* upstream regions was assessed, resulted in the

identification of nearly 12% of all tested TFs as *SOC1* interactors (Figure 2, Supplemental Table 1). Among the identified transcriptional regulators, CONSTANS (CO) was present, which has been shown to regulate *SOC1* indirectly via FT (Yoo et al., 2005). However, our results suggest a direct binding of CO to the *SOC1* UTR region. Besides CO, we also showed direct binding of SPL15 at the *SOC1* UTR region and is co-expressed with *SOC1* in the SAM during floral transition. *SPL15* has been well characterized for its promotion of juvenile-to-adult phase transitioning in *Arabidopsis* (Schwarz et al., 2008). Interestingly, very recently Hyun et al. reported *SPL15*'s role in regulating flowering time under non-inductive SD photoperiod conditions (Hyun et al., 2016). In their study, *SOC1* expression appeared to be unaltered in the *spl15* mutant under LDs or SDs and *SOC1* expression in the shoot apex of the mutant seems to be insufficient to induce flowering under non-inductive conditions. *SPL15* was also shown to be important for its action on the downstream targets of *SOC1*, namely *FUL* and *miR172b* and cooperation between SOC1 and SPL15 in the regulation of downstream targets seems to be key for flowering time control. Here we identified direct binding of the *SOC1* 5' UTR region by SPL15, suggesting the presence of a feedback mechanism that has not been notified in any of the previous studies.

Among others we also identified MYB TFs associated with the *SOC1* promoter. MYB TFs are known to play a role in plant growth and development. *AtMYB29* was identified as a candidate in our eQTL analysis, an upstream binding factor and negative co-expressor of *SOC1.* AtMYB29 is a major regulator of aliphatic glucosinolate biosynthesis acting redundantly with AtMYB28 (Li et al., 2013) . As of now, no flowering phenotype has been reported for *myb29* and hence, the function of *MYB29* in flowering time control still remains to be investigated.

Y1H is a robust method for studying TF binding to genomic regions, but due to the nature of the assay we cannot rule out the identification of false positives. In order to minimize this, our Y1H screen was performed in duplicate and only reproducible interactions were considered for further analysis. Technologies like EMSA and chromatin immunoprecipitation (ChIP) are alternatives that can be used to validate the interactions reported in this study. Performing these assays on all the candidates would be challenging considering the number of TFs involved. Therefore, we made use of *in silico* approaches, such as co-expression analysis and eQTL profiling, to reduce the number of candidates. These methods have their own technical limitations. For instance, eQTL studies are influenced by multiple parameters, e.g. the design of the experiments, sample size, data quality, the data analysis approach and most importantly the heterogeneity of the tissues samples. Given that flowering time is such a complex trait, the identified eQTLs may not indicate any direct biological relationship.

The presence of binding sequences and evidence of direct binding by a TF are signs of potential transcriptional regulation. Nevertheless, binding of a TF to its targets may not always induce changes in gene expression and hence will not result in an eQTL. Other factors such as the chromatin structure, the interaction partners of TF and their post-transcriptional regulation can also influence the behaviour of TF binding and its resulting transcriptional output.

Moreover, it is important to realize that the Y1H system fails to capture the biological conditions in which TF binding takes place. Therefore further follow-up studies are essential to confirm the findings presented in this study. For instance, *in planta* DNA-binding analysis with ChIP and dexamethasone-inducible TF expression in transgenic lines will enable confirmation and analysis of the function of these potential transcriptional regulators of *SOC1* during the floral transition process.

To conclude this study reports a large number of previously undescribed potential *SOC1* regulators. Our analysis utilized publically available RNA-seq data, eQTLs studies and knowledge of flowering time phenotypes previously reported.

***MATERIALS AND METHODS***

Analysis of SOC1 regulatory sequences (TF/TFBS prediction): A ~1.3kb *SOC1* regulatory sequence (900 bp upstream of TSS and 409 bp 5'UTR downstream of TSS) was used for transcription factor binding site perdition analysis. Binding sites were predicted using the web portal of the PlantTFDB 4.0. (Jin et al., 2017).

Yeast-One-Hybrid Assay: As described previously by Immink et al 2012, we PCR amplified the two *SOC1* upstream promoter fragments and the 5' UTR region. All the fragments were cloned into a gateway compatible reporter plasmid pABAi vector CZN1018 (Danisman et al., 2012). This resulted in bait constructs SJ1046 (SOC1 Promoter -663 to +82), SJ1047 (SOC1 promoter -864 to-397) and SJ2030 (SOC1-5'-UTR +1 to +409). The constructs were transformed into yeast strain PJ69-4alpha. Auto-activation tests were performed to determine background expression of the reporter bait constructs. We observed yeast growth in a aurobasidin concentration of up to 50ng/ml. Therefore, for the final binding analysis, screening was done at 75, 100, 125 and 150 ng/ml aurobasidin. We used the REGIA transcription factor collection as previously reported (Castrillo et al., 2011), which are in the PJ68-4A yeast strain. The mating was performed as described previously (Danisman et al., 2012). After 5 days of incubation on selection media at 20 °C, plates were scored for growth. All putative positives identified at 125 ng/ml were rescreened at 125, 150 and 175 ng/ml. All the above indicated constructs were tested by restriction analysis and sanger sequencing for the inserted fragment.

Co-expression analysis : RNA-seq time series data as published by were used (Klepikova et al., 2015), which consist of DEseq-normalized counts. Shoot apical meristem (SAM) samples i.e. Time points M1 to M10 (each consisting of two replicates) were used. Pearson correlation of the various TFs with SOC1 was calculated using the R-function cor (R Developmental Core Team, 2010) http://www.R-project.org.

eQTL analysis: AraQTL workbench was used to investigate the eQTL profiles of *SOC1* and its correlating peaks. A detailed workflow of AraQTL has been described previously (Nijveen et al., 2017). In order to identify TF linked to the significant SOC1 eQTL peaks, using markers as described for each eQTL, we scanned a 300kb reference window around the eQTL peak.

**Supplemental Material**

Supplemental Table 1. **List of all potential upstream regulators of *SOC1***. The first column represents the Locus identifiers of the transcription factors (TFs) that were able to bind at SOC1 regulatory sequence in Y1H assay and the second column gives the names of these TFs. The third column indicates to which *SOC1* upstream region binding was found in the Y1H assay: UTR = 5'UTR (+1 to +409), 1046 = -663 until +82 promotor region, 1047 = -864 until -397 promotor region; Fig.2A. The fourth column reports the level of co-expression between *SOC1* and the identified binding TFs, and the last column gives the annotation of the TFs.

| TF Locus Id | Gene alias | Binding Region at *SOC1* (See Fig.2A) | | | Pearson Correlation (r) | Annotation |
|---|---|---|---|---|---|---|
| | | UTR | 1046 | 1047 | | |
| **AT2G45660** | SOC1 | NA | NA | NA | 1 | AGL20_ATSOC1_SOC1__AGAMOUS-like 20 |
| **AT3G57920** | SPL15 | UTR | | | 0.91 | SPL15__squamosa promoter binding protein-like 15 |
| **At5g62610** | BHLH79 | UTR | | | 0.73 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein |
| **AT4G08150** | KNAT1 | UTR | | | 0.73 | BP_BP1_KNAT1__KNOTTED-like from Arabidopsis thaliana |
| **At1g72210** | BHLH96 | | | 1047 | 0.65 | bHLH096_bHLH96__basic helix-loop-helix (bHLH) DNA-binding superfamily protein |
| **AT1G06850** | BZIP52 | | 1046 | 1047 | 0.62 | AtbZIP52_bZIP52__basic leucine-zipper 52 |
| **AT2G23290** | MYB70 | | | 1047 | 0.61 | AtMYB70_MYB70__myb domain protein 70 |
| **AT1G26780** | LOF1 | UTR | | 1047 | 0.61 | AtMYB117_LOF1_MYB117__myb domain protein 117 |
| **AT5G15840** | CO | UTR | | | 0.61 | BBX1_CO_FG__B-box type zinc finger protein with CCT domain |
| **AT2G41710** | | UTR | | | 0.6 | Integrase-type DNA-binding superfamily protein |
| **AT1G68640** | BZIP46 | | | 1047 | 0.57 | PAN_TGA8__bZIP transcription factor family protein |
| **AT3G09370** | | UTR | 1046 | | 0.56 | AtMYB3R3_MYB3R-3__myb domain protein 3r-3 |
| **AT3G14020** | NFYA6 | | 1046 | 1047 | 0.54 | NF-YA6__nuclear factor Y, subunit A6 |
| **AT2G45650** | AGL6 | UTR | | | 0.5 | AGL6_RSB1__AGAMOUS-like 6 |
| **AT3G44350** | NAC061 | | 1046 | 1047 | 0.49 | NAC061_anac061__NAC domain containing protein 61 |
| **AT3G04450** | | | | 1047 | 0.49 | Homeodomain-like superfamily protein |

| TF Locus Id | Gene alias | Binding Region at *SOC1* (See Fig.2A) | | | Pearson Correlation (r) | Annotation |
|---|---|---|---|---|---|---|
| | | UTR | 1046 | 1047 | | |
| **At5g26990** | DI19-6 | | | 1047 | 0.48 | Drought-responsive family protein |
| **AT3G58120** | BZIP61 | | | | 0.48 | ATBZIP61_BZIP61__Basic-leucine zipper (bZIP) transcription factor family protein |
| **AT3G11440** | MYB65 | | | 1047 | 0.47 | ATMYB65_MYB65__myb domain protein 65 |
| **AT2G20180** | PIF1 | UTR | | | 0.46 | PIF1_PIL5__phytochrome interacting factor 3-like 5 |
| **AT3G02310** | SEP2 | | 1046 | | 0.42 | AGL4_SEP2__K-box region and MADS-box transcription factor family protein |
| **AT5G15800** | SEP1 | | 1046 | 1047 | 0.42 | AGL2_SEP1__K-box region and MADS-box transcription factor family protein |
| **AT2G31220** | BHLH10 | UTR | | | 0.38 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein |
| **AT3G24650** | ABI3 | | | 1047 | 0.34 | ABI3_SIS10__AP2/B3-like transcriptional factor family protein |
| **AT4G32040** | KNAT5 | UTR | | | 0.34 | KNAT5__KNOTTED1-like homeobox gene 5 |
| **AT1G18330** | RVE7 | | | 1047 | 0.34 | EPR1_RVE7__Homeodomain-like superfamily protein |
| **AT5G05410** | DREB2A | | | 1047 | 0.25 | DREB2_DREB2A__DRE-binding protein 2A |
| **AT3G09600** | RVE8 | | 1046 | 1047 | 0.23 | LCL5_RVE8__Homeodomain-like superfamily protein |
| **AT2G31210** | BHLH91 | UTR | | | 0.22 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein |
| **AT3G13445** | TBP1 | | | 1047 | 0.21 | TBP1_TFIID-1__TATA binding protein 1 |
| **AT1G79580** | SMB | | | 1047 | 0.18 | ANAC033_SMB_URP7__NAC (No Apical Meristem) domain transcriptional regulator superfamily protein |
| **AT4G26640** | WRKY20 | | 1046 | 1047 | 0.12 | AtWRKY20_WRKY20__WRKY family transcription factor family protein |
| **AT1G72830** | NFYA3 | | 1046 | | 0.11 | ATHAP2C_HAP2C_NF-YA3__nuclear factor Y, subunit A3 |
| **AT1G73100** | SUVH3 | UTR | 1046 | 1047 | 0.11 | SDG19_SUVH3__SU(VAR)3-9 homolog 3 |
| **AT5G67060** | HEC1 | UTR | | 1047 | 0.06 | HEC1__basic helix-loop-helix (bHLH) DNA-binding superfamily protein |
| **AT5G04940** | SUVH1 | | | 1047 | 0.06 | SUVH1__SU(VAR)3-9 homolog 1 |
| **AT2G20400** | | UTR | | | 0.03 | myb-like HTH transcriptional regulator family protein |
| **AT4G00480** | BHLH12 | UTR | | | 0.01 | ATMYC1_myc1__basic helix-loop-helix (bHLH) DNA-binding superfamily protein |

| TF Locus Id | Gene alias | Binding Region at *SOC1* (See Fig.2A) | | | Pearson Correlation (r) | Annotation |
|---|---|---|---|---|---|---|
| | | UTR | 1046 | 1047 | | |
| **AT5G38800** | BZIP43 | UTR | | 1047 | 0 | AtbZIP43_bZIP43__basic leucine-zipper 43 |
| **AT4G01350** | | | 1046 | 1047 | -0.01 | Cysteine/Histidine-rich C1 domain family protein |
| **AT4G26930** | MYB97 | UTR | | | -0.01 | AtMYB97_MYB97__myb domain protein 97 |
| **AT5G11060** | KNAT4 | UTR | | | -0.02 | KNAT4__KNOTTED1-like homeobox gene 4 |
| **AT2G26150** | HSFA2 | UTR | | | -0.03 | ATHSFA2_HSFA2__heat shock transcription factor A2 |
| **AT3G05480** | | | | | -0.06 | ATRAD9_RAD9__cell cycle checkpoint control protein family |
| **AT4G35550** | WOX13 | UTR | | | -0.07 | ATWOX13_HB-4_WOX13__WUSCHEL related homeobox 13 |
| **AT2G33290** | SUVH2 | | 1046 | 1047 | -0.07 | ATSUVH2_SDG3_SUVH2__SU(VAR)3-9 homolog 2 |
| **AT1G35515** | HOS10 | | 1046 | | -0.08 | HOS10_MYB8__high response to osmotic stress 10 |
| **AT1G14200** | | | | 1047 | -0.08 | RING/U-box superfamily protein |
| **AT2G16720** | MYB7 | UTR | | | -0.09 | ATMYB7_ATY49_MYB7__myb domain protein 7 |
| **AT4G37180** | | | | 1047 | -0.1 | Homeodomain-like superfamily protein |
| **AT3G61150** | HDG1 | UTR | | | -0.14 | HD-GL2-1_HDG1__homeodomain GLABROUS 1 |
| **AT3G23250** | MYB15 | UTR | | | -0.17 | ATMYB15_ATY19_MYB15__myb domain protein 15 |
| **AT5G37260** | RVE2 | | | 1047 | -0.18 | CIR1_RVE2__Homeodomain-like superfamily protein |
| **AT1G08320** | BZIP21 | UTR | | | -0.22 | TGA9_bZIP21__bZIP transcription factor family protein |
| **AT5G25830** | GATA12 | UTR | | | -0.23 | GATA12__GATA transcription factor 12 |
| **AT5G50480** | NFYC6 | | 1046 | 1047 | -0.24 | NF-YC6__nuclear factor Y, subunit C6 |
| **AT1G18710** | MYB47 | | | 1047 | -0.24 | AtMYB47_MYB47__myb domain protein 47 |
| **AT5G65790** | MYB68 | | | 1047 | -0.24 | ATMYB68_MYB68__myb domain protein 68 |
| **AT1G62975** | BHLH125 | UTR | | | -0.25 | basic helix-loop-helix (bHLH) DNA-binding superfamily protein |
| **AT1G80730** | ZFP1 | | | 1047 | -0.29 | ATZFP1_ZFP1__zinc-finger protein 1 |
| **AT3G04420** | NAC048 | | 1046 | 1047 | -0.32 | NAC048_anac048__NAC domain containing protein 48 |

| TF Locus Id | Gene alias | Binding Region at *SOC1* (See Fig.2A) | | | Pearson Correlation (r) | Annotation |
|---|---|---|---|---|---|---|
| | | UTR | 1046 | 1047 | | |
| **AT5G62020** | HSFB2A | UTR | | | -0.32 | AT-HSFB2A_HSFB2A__heat shock transcription factor B2A |
| **AT4G04450** | WRKY42 | | 1046 | | -0.35 | AtWRKY42_WRKY42__WRKY family transcription factor |
| **AT1G71692** | AGL12 | | 1046 | 1047 | -0.35 | AGL12_XAL1__AGAMOUS-like 12 |
| **AT1G42990** | BZIP60 | UTR | | | -0.4 | ATBZIP60_BZIP60__basic region/leucine zipper motif 60 |
| **AT3G10590** | | | 1046 | 1047 | -0.43 | Duplicated homeodomain-like superfamily protein |
| **AT3G16500** | IAA26 | | 1046 | 1047 | -0.44 | IAA26_PAP1__phytochrome-associated protein 1 |
| **AT3G61120** | AGL13 | | | 1047 | -0.46 | AGL13__AGAMOUS-like 13 |
| **AT1G28050** | COL15 | | | 1047 | -0.47 | BBX13__B-box type zinc finger protein with CCT domain |
| **AT3G11260** | WOX5 | | | 1047 | -0.47 | WOX5_WOX5B__WUSCHEL related homeobox 5 |
| **AT1G77450** | NAC032 | UTR | | | -0.48 | NAC032_anac032__NAC domain containing protein 32 |
| **AT2G04038** | BZIP48 | UTR | 1046 | 1047 | -0.54 | AtbZIP48_bZIP48__basic leucine-zipper 48 |
| **AT4G24060** | DOF4.6 | | | 1047 | -0.56 | Dof-type zinc finger DNA-binding family protein |
| **AT1G17950** | MYB52 | UTR | | | -0.57 | ATMYB52_BW52_MYB52__myb domain protein 52 |
| **At4g10480** | | | | 1047 | -0.58 | Nascent polypeptide-associated complex (NAC), alpha subunit family protein |
| **AT2G43000** | JUB1 | UTR | | | -0.6 | ANAC042_JUB1_NAC042__NAC domain containing protein 42 |
| **AT2G22670** | IAA8 | | | 1047 | -0.6 | IAA8__indoleacetic acid-induced protein 8 |
| **AT5G39610** | NAC6 | UTR | | | -0.61 | ANAC092_ATNAC2_ATNAC6_NAC2_NAC6_ORE1__NAC domain containing protein 6 |
| **At1g19000** | | | | 1047 | -0.62 | Homeodomain-like superfamily protein |
| **AT3G30530** | BZIP42 | UTR | | | -0.7 | ATBZIP42_bZIP42__basic leucine-zipper 42 |
| **AT4G36020** | CSP1 | UTR | | | -0.7 | AtCSP1_CSDP1_CSP1__cold shock domain protein 1 |
| **AT5G01200** | | | 1046 | 1047 | -0.71 | Duplicated homeodomain-like superfamily protein |
| **AT1G62300** | WRKY6 | | 1046 | 1047 | -0.72 | ATWRKY6_WRKY6__WRKY family transcription factor |

| TF Locus Id | Gene alias | Binding Region at *SOC1* (See Fig.2A) | | | Pearson Correlation (r) | Annotation |
|---|---|---|---|---|---|---|
| | | UTR | 1046 | 1047 | | |
| **AT3G23050** | IAA7 | | | | -0.74 | AXR2_IAA7__indole-3-acetic acid 7 |
| **At4g27950** | CRF4 | | | 1047 | -0.81 | CRF4__cytokinin response factor 4 |
| **AT5G13180** | NAC083 | UTR | | | -0.85 | ANAC083_NAC083_VNI2__NAC domain containing protein 83 |
| **AT5G07690** | MYB29 | | 1046 | 1047 | -0.86 | ATMYB29_MYB29_PMG2__myb domain protein 29 |
| **AT1G13600** | BZIP58 | UTR | | | -0.88 | AtbZIP58_bZIP58__basic leucine-zipper 58 |
| **AT5G59780** | MYB59 | | 1046 | 1047 | -0.91 | ATMYB59_ATMYB59-1_ATMYB59-2_ATMYB59-3_MYB59__myb domain protein 59 |
| **At3g17100** | BHLH147 | | | 1047 | -0.94 | AIF3__sequence-specific DNA binding transcription factors |
| **AT4G36710** | SCL15 | UTR | | | -0.94 | AtHAM4_HAM4__GRAS family transcription factor |



**Description of additional files :** (Will be provided upon request)

**Supplemental file S1| Complete** List of TFBSs and predicted TFs binding at the SOC1 locus resulting from the PlantRegMap analysis.

## REFERENCES


**Balasubramanian, S., Sureshkumar, S., Agrawal, M., Michael, T.P., Wessinger, C., Maloof, J.N., Clark, R., Warthmann, N., Chory, J., and Weigel, D.** (2006). The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of Arabidopsis thaliana. Nature genetics **38,** 711-715.

**Baurle, I., and Dean, C.** (2006). The timing of developmental transitions in plants. Cell **125,** 655-664.

**Borner, R., Kampmann, G., Chandler, J., Gleissner, R., Wisman, E., Apel, K., and Melzer, S.** (2000). A MADS domain gene involved in the transition to flowering in Arabidopsis. The Plant journal : for cell and molecular biology **24,** 591-599.

**Boss, P.K., Bastow, R.M., Mylne, J.S., and Dean, C.** (2004). Multiple pathways in the decision to flower: enabling, promoting, and resetting. The Plant cell **16 Suppl,** S18-31.

**Bouche, F., Lobet, G., Tocquin, P., and Perilleux, C.** (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. Nucleic acids research **44,** D1167-1171.

**Caicedo, A.L., Richards, C., Ehrenreich, I.M., and Purugganan, M.D.** (2009). Complex rearrangements lead to novel chimeric gene fusion polymorphisms at the Arabidopsis thaliana MAF2-5 flowering time gene cluster. Molecular biology and evolution **26,** 699-711.

**Castrillo, G., Turck, F., Leveugle, M., Lecharny, A., Carbonero, P., Coupland, G., Paz-Ares, J., and Onate-Sanchez, L.** (2011). Speeding cis-trans regulation discovery by phylogenomic analyses coupled with screenings of an arrayed library of Arabidopsis transcription factors. PloS one **6,** e21524.

**Chen, J., Zhu, X., Ren, J., Qiu, K., Li, Z., Xie, Z., Gao, J., Zhou, X., and Kuai, B.** (2017). Suppressor of Overexpression of CO 1 Negatively Regulates Dark-Induced Leaf Degreening and Senescence by Directly Repressing Pheophytinase and Other Senescence-Associated Genes in Arabidopsis. Plant physiology **173,** 1881-1891.

**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.** (2012). Arabidopsis class I and class II TCP transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. Plant physiology **159,** 1511-1523.

**Davin, N., Edger, P.P., Hefer, C.A., Mizrachi, E., Schuetz, M., Smets, E., Myburg, A.A., Douglas, C.J., Schranz, M.E., and Lens, F.** (2016). Functional network analysis of genes differentially expressed during xylogenesis in soc1ful woody Arabidopsis plants. The Plant journal : for cell and molecular biology **86,** 376-390.

**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 of the United States of America **108,** 6680-6685.

**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 : for cell and molecular biology **67,** 1006-1017.

**El-Din El-Assal, S., Alonso-Blanco, C., Peeters, A.J., Raz, V., and Koornneef, M.** (2001). A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2. Nature genetics **29,** 435-440.

**Filiault, D.L., Wessinger, C.A., Dinneny, J.R., Lutes, J., Borevitz, J.O., Weigel, D., Chory, J., and Maloof, J.N.** (2008). Amino acid polymorphisms in Arabidopsis phytochrome B cause differential responses to light. Proceedings of the National Academy of Sciences of the United States of America **105,** 3157-3162.

**Fornara, F., de Montaigu, A., and Coupland, G.** (2010). SnapShot: Control of flowering in Arabidopsis. Cell **141,** 550, 550 e551-552.

**Grant, C.E., Bailey, T.L., and Noble, W.S.** (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics **27,** 1017-1018.

**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.

**Hartmann, U., Hohmann, 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 : for cell and molecular biology **21,** 351-360.

**Helliwell, C.A., Wood, C.C., Robertson, M., James Peacock, W., and Dennis, E.S.** (2006). The Arabidopsis FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. The Plant journal : for cell and molecular biology **46,** 183-192.

**Hepworth, S.R., Valverde, F., Ravenscroft, D., Mouradov, A., and Coupland, G.** (2002). Antagonistic regulation of flowering-time gene SOC1 by CONSTANS and FLC via separate promoter motifs. The EMBO journal **21,** 4327-4337.

**Hyun, Y., Richter, R., Vincent, C., Martinez-Gallegos, R., Porri, A., and Coupland, G.** (2016). Multi-layered Regulation of SPL15 and Cooperation with SOC1 Integrate Endogenous Flowering Pathways at the Arabidopsis Shoot Meristem. Developmental cell **37,** 254-266.

**Immink, R.G., Pose, 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 physiology **160,** 433-449.

**Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., and Gao, G.** (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic acids research **45,** D1040-D1045.

**Jin, J., He, K., Tang, X., Li, Z., Lv, L., Zhao, Y., Luo, J., and Gao, G.** (2015). An Arabidopsis Transcriptional Regulatory Map Reveals Distinct Functional and Evolutionary Features of Novel Transcription Factors. Molecular biology and evolution **32,** 1767-1773.

**Kimura, Y., Aoki, S., Ando, E., Kitatsuji, A., Watanabe, A., Ohnishi, M., Takahashi, K., Inoue, S., Nakamichi, N., Tamada, Y., and Kinoshita, T.** (2015). A flowering integrator, SOC1, affects stomatal opening in Arabidopsis thaliana. Plant & cell physiology **56,** 640-649.

**Klepikova, A.V., Logacheva, M.D., Dmitriev, S.E., and Penin, A.A.** (2015). RNA-seq analysis of an apical meristem time series reveals a critical point in Arabidopsis thaliana flower initiation. BMC genomics **16,** 466.

**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.

**Lee, J., and Lee, I.** (2010). Regulation and function of SOC1, a flowering pathway integrator. Journal of experimental botany **61,** 2247-2254.

**Lee, J.H., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.S., and Ahn, J.H.** (2007). Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes & development **21,** 397-402.

**Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E., and Yu, H.** (2008). A repressor complex governs the integration of flowering signals in Arabidopsis. Developmental cell **15,** 110-120.

**Li, Y.M., Sawada, Y., Hirai, A., Sato, M., Kuwahara, A., Yan, X.F., and Hirai, M.Y.** (2013). Novel Insights Into the Function of Arabidopsis R2R3-MYB Transcription Factors Regulating Aliphatic Glucosinolate Biosynthesis. Plant and Cell Physiology **54,** 1335-1344.

**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.

**Mateos, J.L., Madrigal, P., Tsuda, K., Rawat, V., Richter, R., Romera-Branchat, M., Fornara, F., Schneeberger, K., Krajewski, P., and Coupland, G.** (2015). Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in Arabidopsis. Genome biology **16,** 31.

**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. Nature genetics **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.

**Moon, J., Suh, S.S., Lee, H., Choi, K.R., Hong, C.B., Paek, N.C., Kim, S.G., and Lee, I.** (2003). The SOC1 MADS-box gene integrates vernalization and gibberellin signals for flowering in Arabidopsis. The Plant journal : for cell and molecular biology **35,** 613-623.

**Nijveen, H., Ligterink, W., Keurentjes, J.J., Loudet, O., Long, J., Sterken, M.G., Prins, P., Hilhorst, H.W., de Ridder, D., Kammenga, J.E., and Snoek, B.L.** (2017). AraQTL - workbench and archive for systems genetics in Arabidopsis thaliana. The Plant journal : for cell and molecular biology **89,** 1225-1235.

**O'Malley, R.C., Huang, S.S.C., 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 (vol 165, pg 1280, 2016). Cell **166,** 1598-1598.

**Onouchi, H., Igeno, M.I., Perilleux, C., Graves, K., and Coupland, G.** (2000). Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among Arabidopsis flowering-time genes. The Plant cell **12,** 885-900.

**Parcy, F.** (2005). Flowering: a time for integration. The International journal of developmental biology **49,** 585-593.

**Putterill, J., Laurie, R., and Macknight, R.** (2004). It's time to flower: the genetic control of flowering time. BioEssays : news and reviews in molecular, cellular and developmental biology **26,** 363-373.

**R Developmental Core Team.** (2010). R: A Language and Environment for Statistical Computing (R Foundation for statistical computing.).

**Richter, R., Bastakis, E., and Schwechheimer, C.** (2013). Cross-repressive interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the control of greening, cold tolerance, and flowering time in Arabidopsis. Plant physiology **162,** 1992-2004.

**Rosloski, S.M., Jali, S.S., Balasubramanian, S., Weigel, D., and Grbic, V.** (2010). Natural diversity in flowering responses of Arabidopsis thaliana caused by variation in a tandem gene array. Genetics **186,** 263-276.

**Salome, P.A., Bomblies, K., Laitinen, R.A., Yant, L., Mott, R., and Weigel, D.** (2011). Genetic architecture of flowering-time variation in Arabidopsis thaliana. Genetics **188,** 421-433.

**Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., and Coupland, G.** (2000). Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. Science **288,** 1613-1616.

**Schwartz, C., Balasubramanian, S., Warthmann, N., Michael, T.P., Lempe, J., Sureshkumar, S., Kobayashi, Y., Maloof, J.N., Borevitz, J.O., Chory, J., and Weigel, D.** (2009). Cis-regulatory changes at FLOWERING LOCUS T mediate natural variation in flowering responses of Arabidopsis thaliana. Genetics **183,** 723-732, 721SI-727SI.

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

**Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasino, R.A., and Coupland, G.** (2006). The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. Genes & development **20,** 898-912.

**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.

**Simpson, G.G., and Dean, C.** (2002). Arabidopsis, the Rosetta stone of flowering time? Science **296,** 285-289.

**Song, H.R., Song, J.D., Cho, J.N., Amasino, R.M., Noh, B., and Noh, Y.S.** (2009). The RNA binding protein ELF9 directly reduces SUPPRESSOR OF OVEREXPRESSION OF CO1 transcript levels in arabidopsis, possibly via nonsense-mediated mRNA decay. The Plant cell **21,** 1195-1211.

**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.

**Wang, H., and Wang, H.** (2015). The miR156/SPL Module, a Regulatory Hub and Versatile Toolbox, Gears up Crops for Enhanced Agronomic Traits. Molecular plant **8,** 677-688.

**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.

**Wang, Y., Liu, C., Yang, D., Yu, H., and Liou, Y.C.** (2010). Pin1At encoding a peptidyl-prolyl cis/trans isomerase regulates flowering time in Arabidopsis. Molecular cell **37,** 112-122.

**Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U., and Weigel, D.** (2005). Integration of spatial and temporal information during floral induction in Arabidopsis. Science **309,** 1056-1059.

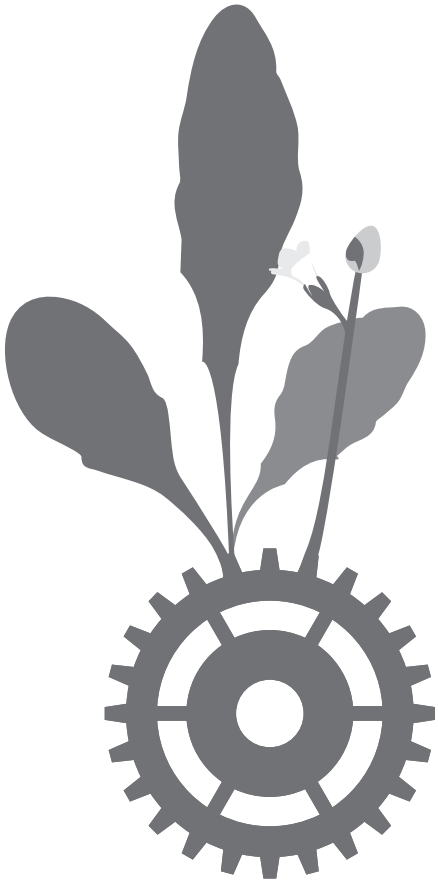
**Xu, M., Hu, T., Zhao, J., Park, M.Y., Earley, K.W., Wu, G., Yang, L., and Poethig, R.S.** (2016). Developmental Functions of miR156-Regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) Genes in Arabidopsis thaliana. PLoS genetics **12,** e1006263.

**Yoo, S.K., Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S., and Ahn, J.H.** (2005). CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in Arabidopsis. Plant physiology **139,** 770-778.

**Yu, C.P., Lin, J.J., and Li, W.H.** (2016). Positional distribution of transcription factor binding sites in Arabidopsis thaliana. Scientific reports **6,** 25164.

# CHAPTER 3

## Identification of in planta protein-protein interactions using IP-MS



Suraj B. Jamge
Gerco C. Angenent
Marian Bemer

Plant Chromatin Dynamic Methods and Protocol
(MIMB, volume 1675)

***ABSTRACT***

Gene regulation by transcription factors involves complex protein interaction networks, which include chromatin remodelling and modifying proteins as an integral part. Decoding these protein interactions is crucial for our understanding of chromatin-mediated gene regulation. Here, we describe a method for the immunoprecipitation of *in planta* nuclear protein complexes followed by mass spectrometry (IP-MS) to identify interactions between transcription factors and chromatin remodelers/modifiers in plants. In addition to a step-by-step bench protocol for immunoprecipitation and subsequent mass spectrometry, we provide guidelines and pointers on necessary controls and data analysis approaches.

Keywords: Arabidopsis, protein-protein interactions, immunoprecipitation, Mass spectrometry, label free quantification, chromatin remodelers.

## *1. INTRODUCTION*

Proteins interact with each other and form multimeric complexes that execute unique functions. Protein-protein interactions occur at different places in the cell and play crucial roles in a variety of processes, such as cell-to-cell signalling by effector proteins, chromatin organization by histones proteins, and gene regulation by transcription factor proteins. Gene transcription is usually regulated by large protein complexes that can include transcription factors, transcriptional co-factors and chromatin remodelers and modifiers (Smaczniak et al., 2012b). While more and more functions of individual genes and their protein products have been elucidated, the interactions between the different factors in protein complexes and their role in gene regulatory networks are still far from understood. Although gene regulation was initially considered to be independently regulated by chromatin remodelers and transcription factors, recent data have revealed many physical interactions between these two types of proteins, suggesting that they act together in large regulatory complexes (Smaczniak et al., 2012b; Liang et al., 2015; Del Olmo et al., 2016). These regulatory protein modules are highly dynamic and different combinations of proteins are involved in the regulation of specific target genes in particular tissues or under specific conditions. To unravel the composition of the *in planta* complexes that regulate gene activity, the entire complex can be immunoprecipitated (IP) using a specific antibody against one of the proteins, or using an antibody against a tagged protein.

In the classical approach, this IP is followed by western blotting (co-IP). Although relatively easy to perform, a drawback of this approach lies in the need for specific antibodies against each of the potential complex partners, which requires prior knowledge of the putative interaction partners, and thereby prevents the identification of novel interactors. This problem has been overcome by technical advances in liquid chromatography coupled with mass-spectrometry (LC-MS), which have enabled the detection of low-abundance proteins, as well as high-throughput identification of hundreds of proteins from a single sample in a relatively short time. The additional availability of user-friendly data analyses tools and the choice of label free quantification (LFQ) makes LC-MS an attractive option for protein interaction research. In this chapter, we describe a label-free method that enables the user to study the composition of *in planta* nuclear protein complexes and to identify the interactions between transcription factors and chromatin remodelers.

In this method (see **Fig. 1**), protein complexes are isolated from native plant tissue by immuno-precipitation. Subsequently, the identification of all proteins requires a proteolytic step with trypsin, followed by purification of the sample. Digested peptides are then eluted and further injected into a mass spectrometer, where the molecules are ionized, accelerated and separated based on their mass-to-charge ratio, enabling the deduction of peptide identity. In short, the peptide identification is done by liquid chromatography tandem-mass spectrometry (LC-MS/MS) followed by searches of the deduced peptides against a protein database,

resulting in a list of proteins present in the IP sample. This is a straightforward procedure for Arabidopsis, because an exhaustive peptide database is available, but may be more challenging for other species. Here, we provide a detailed step-by-step protocol for performing IPs and sample preparation for LC-MS/MS analysis. In addition, guidelines on data analysis tools and label-free quantification as well as recommendations on necessary controls have been summarized.

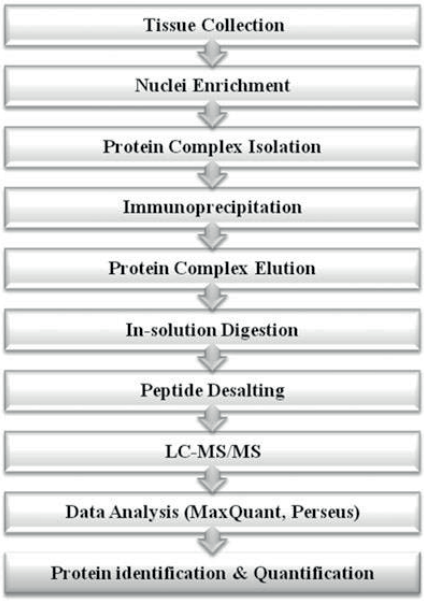



Fig. 1. **Schematic workflow of the IP-MS/MS protocol.**

## *2. MATERIALS*

The immunoprecipitation of *in planta* nuclear protein complexes can be achieved using a specific antibody against the protein of interest, or by using a specific antibody against a protein tag that has been added to the protein of interest. The advantage of the latter approach is that a specific antibody of high quality can be selected, but it requires the generation of stable transgenic lines that express the tagged protein of interest (*see* **Note 1**). Because the use of fluorophore-tagged proteins works best in our hands, this protocol describes the immunoprecipitation based on the use of magnetic anti-GREEN FLUORESCENT PROTEIN (GFP) microbeads. However, other combinations of antibodies and beads can also be used, but this will have to be optimized by the user (*see* **Note 2**). All the buffers and solutions described in this protocol are prepared with autoclaved milliQ water (before the IP) or HPLC water (after the IP). The recommended waste disposal regulations for each reagents should be followed.

**2.1 Tissue collection & preparation**

1. Mortar and pestle
2. Liquid nitrogen
3. Nitrile gloves
4. 50 ml centrifuge tubes

**2.2 Protein Complex Isolation**

1. Liquid nitrogen
2. Nitrile gloves
3. Nylon mesh (pore size 55 µm)
4. Glass funnel
5. 50 ml centrifuge tubes
6. 100 mM Sodium phosphate buffer pH 7.0: prepare 100 mM $Na_2HPO_4$ and 100 mM $NaH_2PO_4$. While measuring the pH, add 100 mM $NaH_2PO_4$ to the 100 mM $Na_2HPO_4$ solution until a pH of 7.0 is reached.
7. 20 % Triton X-100: prepare fresh in advance, allowing the Triton to mix with the milliQ on a roller for several hours.
8. Complete Protease Inhibitor Cocktail tablets (e.g. from Roche).
9. M1 buffer (prepare freshly from stock solutions): 0.1 M NaCl, 10 mM sodium phosphate buffer pH 7.0, 1 M 2-methyl 2,4-pentanediol. Per 50 ml M1 buffer, add 35.4 µl 2-mercaptoethanol and 1 tablet of protease inhibitor cocktail (*see* **Note 3**). Bring up to a final volume of 50 ml using milliQ water.
10. M2 buffer (prepare freshly from stock solutions): 0.1 M NaCl, 10 mM sodium phosphate pH 7.0, 10 mM $MgCl_2$, 1 M 2-methyl 2,4-pentanediol, 0.5 % Triton X-100 (from a fresh 20 % stock solution). For 50 ml M2 buffer, add 35.4 µl and 1 tablet protease inhibitor cocktail (*see* **Note 3**). Bring up to a final volume of 50 ml using milliQ water.
11. M3 buffer (prepare freshly from stock solutions): 0.1 M NaCl, 10 mM sodium phosphate pH 7.0. For 10 ml buffer, add 7.1 µl 2-mercaptoethanol and ¼ tablet of protease inhibitor cocktail (*see* **Note 3**). Bring up to a final volume of 50 ml using milliQ water.
12. Lysis buffer (*see* **Note 4)**: 1 % triton X100, 50 mM Tris-HCl pH 8.0.
13. Benzonase (25 u/µl)
14. Sonicator
15. 2 ml low protein-binding tubes
16. 25X protease inhibitor mix: dissolve 1 tablet of Complete Protease Inhibitor Cocktail in 2 ml of HPLC water in a 2 ml tube. Store 100 µl aliquots in 1.5 ml tubes at -20 °C.

**2.3 Protein Immunoprecipitation, Elution and Tryptic Digestion**

1. Nitrile gloves.

2. Because our proteins are tagged with GFP, we use anti-GFP magnetic microbeads for the IP. However, various other options are available.
   µMACS™ GFP Isolation Kit (Miltenyi Biotech) containing (*see* **Note 4**):
   a. anti-GFP microbeads.
   b. Lysis buffer: 1 % triton X100, 50 mM Tris-HCl pH 8.0.
   c. Wash buffer 2: 20 mM Tris-HCl pH 7.5.
3. Magnetic stand for columns/tubes (e.g. µMACS Separator with MultiStand, Miltenyi Biotec).
4. Microbeads-binding matrix (e.g. µ-Columns, Miltenyi Biotec).
5. 8 M urea in HPLC water (fresh).
6. Rotating device.
7. 50 mM ammonium bicarbonate in HPLC water (fresh).
8. Low-binding microcentrifuge tubes (2 ml and 1.5 ml)
9. Low-binding tips
10. 45 mM dithiothreitol in 50 mM ammonium bicarbonate (fresh).
11. 100 mM iodoacetamide in 50 mM ammonium bicarbonate (fresh).
12. Trypsin Gold, Mass Spectrometry Grade. Prepare 0.1 µg/µl trypsin in HPLC water. Aliquots can be prepared in advance and stored at -80 °C.

**2.4 Peptide Preparation**

For the desalting of large numbers of samples simultaneously, we use the MultiScreen Vacuum Manifold system (Merck Millipore) in combination with a 96-wells Oasis HLB µElution plate. However, single samples can also be desalted using special desalting columns or tips, such as C18 ZipTips (Millipore).

1. Nitrile gloves.
2. 100 % Acetonitrile HPLC grade
3. 100 % Formic acid HPLC grade
4. 10 % Formic acid (in HPLC water)
5. 50 % acetonitrile / 5 % formic acid (v/v)
6. HPLC water
7. 96-well plate with sorbent for purification (e.g. Oasis HLB µElution plate, Waters)
8. Device for vacuum filtration of 96-well plates (e.g. Multiscreen HTS vacuum manifold apparatus, Millipore)
9. Vacuum pump
10. Vacuum centrifuge concentrator (e.g. SpeedVac, Thermo Fischer Scientifics)

**2.5 LC-MS equipment and Software for data analysis**

1. Ion Trap-Orbitrap Mass Spectrometer such as LTQ-Orbitrap XL (Thermo Fischer Scientifics)
2. MaxQuant, http://www.coxdocs.org/doku.php?id=:maxquant:start
3. Perseus, http://www.coxdocs.org/doku.php?id=perseus:start

## *3. METHODS*

At all times, work in a clean and sterile environment and use nitrile gloves during the handling of the samples. Based on the capacity of certain instruments and the handling and waiting time between steps, we recommend not to process more than six samples at once during the IP. For handling more than six samples, please make use of one of the pause points during which the samples can be stored for a few days or up to several weeks as stated. The trypsin digestion and desalting procedure can be performed on a large number of samples simultaneously if the 96-wells plate is used for purification.



### 3.1 Starting Material and Preparations

1. Harvest the tissue of interest, determine the weight and freeze immediately in liquid nitrogen. The tissue can be stored at -80 °C until use. The time of harvest, growth stage, amount of starting material, number of replicates and controls are key factors to consider before performing the protocol (*see* **Notes 1** and **Notes 5-8**).
2. Using a chilled mortar and pestle, grind the harvested plant material using liquid nitrogen. Make sure the tissue is ground into a fine powder.
3. Transfer the powdered tissue with liquid nitrogen to a 50 ml centrifuge tube and allow the liquid nitrogen to vaporize. Once vaporized, the ground tissue can be used immediately or stored at -80 °C for up to 2-3 days. We recommend to grind the material a day before performing the protein extraction steps.

### 3.2 Protein Isolation and Sonication

In this step, the nuclei are purified to enrich for protein complexes that play a role in the nucleus, such as those containing transcription factors and chromatin modifiers (*see* **Note 6**). After isolation of the nuclei, the cells are disrupted by sonication in lysis buffer to release the proteins.

1. Freshly prepare all the buffers (M1, M2, M3) on ice under a sterile fume hood.
2. Prepare the filter apparatus by placing an open 50 ml tube on ice, topped with a glass funnel, which contains a clean piece of cloth mesh (55 µm).
3. Take the grinded tissue sample from the -80 °C freezer (see step 3 of Subheading 3.1) in a Dewar flask containing liquid nitrogen.
4. Take out the sample from the liquid nitrogen with a long forceps. Add approximately 20 ml M1 buffer per gram of starting material (weighed prior to grinding, see **Note 5**) and resuspend gently by shaking until it becomes homogenous.
5. Once homogenous, carefully pour the sample through the filter apparatus, in the glass funnel on ice. Allow the sample to flow through by gravity.
6. Pipet 5 ml of additional M1 buffer to wash the mesh of the filter apparatus and allow residual sample to flow through.
7. Centrifuge the filtrate at 1000 x g for 20 min at 4 °C.

8. Discard the supernatant in the appropriate waste container and place the sample back on ice.
9. To wash the cell pellet, add 5 ml of M2 buffer and resuspend the pellet gently by shaking. Do not vortex.
10. Centrifuge the resuspended pellet at 1000 x g for 10 min at 4 °C and discard the supernatant in the appropriate waste container and place the sample back of ice.
11. Repeat the washing step with 5 ml of M2 buffer four times.
12. After completing five washing steps with M2 buffer, resuspend the semi-pure nuclei gently shaking in 5 ml of M3 buffer.
13. Centrifuge the semi-pure nuclei at 1000 x g for 10 min at 4 °C.
14. Discard the supernatant in the appropriate waste container and place the tube back on ice.
15. Resuspend the crude nuclear pellet in 1 ml of lysis buffer with 45 µl of 25X protease inhibitor mix and 5 µl benzonase (to eliminate the DNA and RNA) and transfer the sample to a clean 2 ml tube.
16. Sonicate the nuclei for 10 seconds on ice (*see* **Note 9**). After sonication, invert the tube several times and place on ice for 45 seconds. Repeat the sonication step twice.
17. After 3 sonication cycles, place the sample on ice for 5 minutes.
18. Place the sample into a pre-cooled microcentrifuge and centrifuge at maximum speed for 10 minutes at 4 °C.
19. Transfer the supernatant (soluble protein extract) to a new microcentrifuge tube and centrifuge again at maximum speed for 10 minutes at 4 °C. Repeat this step until no visible pellet is present anymore (*see* **Note 10**).
20. Upon final centrifugation, transfer the supernatant to a 2 ml low-protein binding tube and place on ice.
21. Optional: Take a small aliquot (50 µl) of the protein extract to check using Western blotting (*see* **Note 11**).

**3.3 Protein Immunoprecipitation and Elution**

In the IP step, the native protein complexes that contain the protein of interest will be pulled down. We describe here the pull down of a GFP-tagged protein using anti-GFP magnetic beads (*see* **Note 2**). For all the following steps in this protocol, work in a clean sterile environment (e.g. flow hood). Any contamination of the sample may affect the detection of immunoprecipitated proteins during the LC- MS/MS analysis. Therefore, while handling samples, the use of protective nitrile gloves is highly advised.

1. Resuspend the anti-GFP microbeads by vortexing and add 50 µl of anti-GFP microbeads to the soluble protein extract.
2. Incubate the sample on a rotating device for 1 hour at 4°C.
3. Place the µColumn onto the µMACS™ separator in the flow hood.
4. First calibrate the µColumn with 200 µl lysis buffer.

5. Load the immunoprecipitate (sample from step 2) onto the calibrated µColumn and allow it to flow through by gravity (*see* **Note 12**).
6. Wash the immobilized beads six times (*see* **Note 13**) with 200 µl µMACS lysis buffer.
7. To remove detergents, wash the immobilized beads twice with 200 µl of µMACS wash buffer 2.
8. Freshly prepare 1 ml 8M Urea, preferably with HPLC water.
9. Add 20 µl (dead volume of the column) of 8M Urea to the µColumn, which contains the washed immobilized beads.
10. Incubate the µColumn for 5 min at room temperature.
11. If a small droplet remains on the end of the µColumn, remove it using a pipette tip.
12. To collect the eluate, place a clean, labelled, 1.5 ml low-protein binding tube under the µColumn.
13. Add 50 µl of 8M Urea on to the µColumn to elute the protein immunoprecipitate. Be patient and collect all the droplets during the elution.
14. The protein immunoprecipitate can be stored at -20 °C (for several weeks) until use.



**3.5 Protein Digestion with Trypsin**

For analysis of the proteins with LC-MS/MS, the proteins have to be digested into smaller peptides using the serine protease trypsin (*see* **Note 14**).

1. Freshly prepare all the solution in 50 mM ammonium bicarbonate and continue working in a clean environment while wearing protective nitrile gloves.
2. Dilute the eluated proteins (i.e. the 50 µl protein immunoprecipitate from step 6 of Subheading 3.4) four times by adding 150 µl of 50 mM ammonium bicarbonate to adjust the final concentration of Urea to 2M.
3. In order to reduce the number of protein disulphide bonds, add 10 µl 45 mM DTT to the sample.
4. Incubate at 37 °C for 30 min in a thermomixer at 550 rpm.
5. Allow the sample to cool down by leaving it at room temperature for 5 min.
6. Add 10 µl of 100 mM iodocetamide and place the sample in the dark at room temperature for 30 min. This will alkylate any free thiol groups.
7. Add 15 µl of 0.1 ug/µl trypsin (aliquots of trypsin can be prepared in advance and stored at -80 °C).
8. To digest the proteins into peptides, incubate the sample at 37 °C for 16 – 24 h.
9. The digested mixture can be stored at -80 °C for several days.

**3.6 Peptide Clean Up**

Prior to LC-MS/MS analysis, the peptides must be cleaned to remove salts and urea from the mixture. Freshly prepare all the solutions and work under the flow hood.

1. Add 26 µl 100% formic acid to the digested sample (final concentration of formic acid 10% v/v), vortex gently and briefly spin down.

2. Assemble the multiscreen HTS vacuum manifold apparatus with the vacuum pump and place the 96-well Oasis HLB µelution plate in it on top of a wash plate (96 well low-binding plate (*see* **Note 15**)). Mark the wells that are going to be used.
3. Turn on the vacuum pump and set the pressure to 2.5 inHg.
4. Equilibrate the HLB µelution plate by pipetting 150 µl of 100 % Acetonitrile in the marked wells and collect the flow-through into the wash plate. Increase the pressure to 10-12 inHg in order to allow all the solution to pass through. Reduce the pressure back to 2.5 inHg and stop the vacuum.
5. Repeat the wash with 150 µl of 100% acetonitrile under vacuum.
6. Discard the flow-through from the wash plate and place it back.
7. Start the vacuum pump and set the pressure to 2.5 inHg.
8. Wash the marked wells of the HLB µelution plate twice with 150 µl 10 % formic acid. Vacuum at a high pressure (i.e. 10-12 inHg) in order to allow all the solution to pass through. Then reduce the pressure back to 2.5 inHg and then stop the vacuum. Discard the flow-through.
9. Switch on the vaccum pump and set the pressure to 2.5 inHg.
10. Load the sample into the marked well of the washed HLB µelution plate and allow the sample to flow through without increasing the pressure. Leaving the vacuum at low pressure provides maximum binding capacity to the column. Once all the sample mixture has run through, stop the vacuum and discard the waste.
11. Turn on the vacuum pump and set the pressure at 2.5 inHg.
12. Wash twice with 150 µl 10 % formic acid. Increase the pressure to 10-12 inHg in order to allow all the solution to pass through. Then reduce the pressure back to 2.5 inHg and stop the vacuum to discard the waste.
13. Replace the wash plate from the apparatus with a collection plate (96 well low-binding plate).
14. Start the vacuum pump at a pressure of 2.5 inHg.
15. To elute the sample peptides, add 50 µl of 50 % Acetonitrile / 5 % formic acid (v/v). At the end of the elution the pressure can be increased to 15 inHg. Stop the vacuum after 1-2 min (when the eluate reached the collection plate). Repeat the elution step once. Transfer the eluate to a new 1.5 ml low-protein binding tube (*see* **Note 16**).
16. Place the peptide eluate in a SpeedVac to evaporate the elution solution and recover a dried peptide pellet.
17. The dried peptides are ready for LC-MS/MS analysis.

**3.7 Peptide Sequencing**

In reverse phase chromatography, the peptides are separated and then identified by tandem mass spectrometry (MS/MS) on an LTQ-Orbitrap XL (Wendrich et al., 2017). Owing to the complexity of the peptide sample, a nano-LC setup with a 250 mm (or more) column of C18 particles should be used. For the separation, the dried peptides are dissolved in 50 µl 0.1 % formic acid in water and 18 µl is injected onto a pre-concentration column. The injection is

performed at a constant pressure of 270 bar to obtain a constant flow of around 7 µl /min. The peptides are then eluted onto an analytical column with an acetonitril gradient at a flow rate of 0.5 µl/min with a Proxeon EASY nanoLC. The MS/MS measurements are recorded with a gradient increase of 0.5 % (v/v) acetonitrile per minute or less. FTMS spectrum between m/z 380 and 1,400 is acquired by the orbitrap at a high resolution (60,000) with a target value of 1,000,000 or a maximum ion injection time of 500 ms. Data-dependent MS/MS spectra are obtained by the LTQ for the highest four multiple-charged peaks at a threshold target value of 5,000 with an exclusion list of 500 m/z values and a 60-s exclusion duration (*see* **Note 17**).



### 3.8 Data Analysis

The LC-MS/MS raw data generated from the LTQ-orbitap XL can be analysed with the MaxQuant software package, a tool for protein identification and quantification that is freely available (Cox and Mann, 2008). A detailed description of the MaxQuant suite that summarizes the algorithms it uses, including those for peak detection, scoring peptides, and protein identification has been described by Cox and Mann (Cox and Mann, 2008). For step-by-step processing of LC-MS/MS datasets with MaxQuant, please refer to Smackzniak et al. (Smaczniak et al., 2012a) and Cox and Mann (see **Note 18**) (Cox and Mann, 2008). After the completion of the Max Quant analysis the output dataset can now be further processed for statistical analysis with the Perseus package (*see* **Note 19**) (Tyanova et al., 2016). Below we described key steps for filtering and statistical analysis.

#### 3.8.1 Protein groups filtering and Statistical analysis with Perseus

1. Using Generic matrix upload icon to load the proteinGroup.txt file of your MaxQuant output data in the Peruses software (version 1.5.3.2).
2. Selectively place the output data from the text file into the right categories as described below and click OK.
   a) Expression (Main): Add 'LFQ' and 'iBAQ' intensities of all samples (both test & control samples).
   b) Numerical Columns: Add 'Unique peptides', 'peptides', 'number of proteins', 'iBAQ'
   c) Text Columns: Add 'Protein IDs', 'Fasta headers'
   d) Categorical Columns : Add 'Potential contaminants', 'Reverse' and 'Only identified by sites'
3. Using the 'Filter rows' tab, perform the following filtering :
   a) Filter the rows based on categorical columns: filter for 'Only identified by sites' and then for 'Reverse'. Optionally, one can also filter out the 'potential contaminants' from the dataset. However, it is good to realize that your fluorophore e.g. GFP and trypsin are also characterized as contaminants. These proteins should not be filtered out. Therefore, it is best to filter out contaminants using the 'select rows manually' tab and remove contaminants excluding GFP and trypsin.

b) Filter the rows based on numerical/main columns: filter for 'peptides' with at least two identified peptides per protein. Similarly filter for 'unique peptides' with at least one unique peptide identified per protein.

4. Using the Basic tab, perform logarithmic transformation (log2) for the intensity (expression) data.
5. With the Imputation tab, replace missing values by a constant that is marginally lower than the lowest (Log) value measured.
6. Assemble the data set into Groups, e.g. test and control, and use an appropriate statistical test to calculate the relative protein abundance. Performing a *T*-test (e.g. FDR =0.01 and $S_0$ = 1) allows to test the significant differences between the means.
7. In addition to data filtering, data imputation and data normalization, Perseus offers several tools for quantification and visualization of the proteomics data, and the generation of volcano plots and scatter plots can be performed within the software package.
8. The analysed data from Perseus can be exported into excel format for further analysis.

***4. NOTES***

1. We prefer to use fluorophore tagged lines and antibodies against the tag for the immunoprecipitation experiments. In our case, this approach gave better results than antibodies against the native protein. This requires the generation of transgenic lines that express the tagged protein of interest, preferably in the corresponding mutant background (e.g. *SEP3:GFP* in the *sep3* mutant background). If expressing the transgene in the mutant background, complementation should be observed. If this is not the case transferring the tag from the C-terminus to the N-terminus or vice versa may help.
2. When using an antibody against a native protein, test whether the antibody is specific for your protein of interest prior to the IP-MS. Subsequently, the antibody should be coupled to magnetic or agarose beads according to the manufacturer's description.
3. Dissolving a complete protease inhibitor tablet directly in the buffer will take some time. To facilitate the dissolution of the protease inhibitor tablet, the tablet can be disintegrated in a tube with 1 ml milliQ using tapping and up- and down pipetting and the complete 1 ml can be added to 50 ml buffer.
4. The buffers and wash solutions present in the µMACS™ GFP Isolation Kit may not be sufficient to process all samples. The user may have to prepare extra solutions as per the kit's instruction manual. The lysis buffer from the kit contains 1 % Triton X-100. Depending on your experimental requirements, useful variations in lysis buffer can be performed. For example, lowering the detergent concentrations or switching to other detergents such as NP-40 or Digitonin may help to increase the signal to noise ratio for

proteins that have higher or lower affinity for the antibody/beads than the average protein.

5. The amount and type of plant tissue used depends on the user's research question. This protocol has been successfully tested with 2 g of starting material and can be scaled down to as low as 0.5 g of tissue. This protocol is suitable for different Arabidopsis tissues and has been used for seedlings, rosette leaves, inflorescences, stems, cauline leaves and siliques without modifications.
6. Moderate to high concentrations of the bait protein in the nucleus are essential. For proteins which are of low or variable abundance in the cell, the nuclear enrichment steps (steps 1 to 14 of Subheading 3.2) can be skipped and IPs can be performed on crude protein extract.
7. We advise a minimum of 3-4 independent biological replicates for reliable reproducibility and estimation of protein abundance ratios between test and control samples. Replicates aid in normalization of the datasets and an improved comparison of protein abundance across multiple samples. The statistical power is increased with multiple biological replicates.
8. A reliable IP control is necessary for accurate comparison in quantitative proteomics. One possible control would be Arabidopsis wild-type plants for comparison against the transgenic GFP (or other fluorophore )-tagged line. This control can correct for non-specific binding of GFP with other proteins. However, the best and most reliable control would be an Arabidopsis line expressing nuclear localized GFP under the same promoter as the bait protein.
9. Dependent on the type of sonicator used, or when using an ultrasonic bath, pulse times may need to be adjusted for proper cell disruption. However, the samples must be kept on ice during the entire procedure.
10. This step is critical as any debris/pellet may obstruct the column during IP steps.
11. To check if your protein of interest and/or protein tag can be detected in the sample, a western blot can be performed using the same antibody as used for the immunoprecipitation. At several points in the protocol, such as before the IP, and after the washing and elution steps, a small aliquot can be saved for western analysis. This analysis can be performed when the samples are stored after elution of the IP sample (step 14 of Subheading 3.3).
12. If pipetting of the sample or the washing buffers on the column results in the formation of an air bubble, this can block the flow-through of the solutions (visible as a small circle in the column). This is probably the case if the flow through in one column is much slower than in the others. If an air bubble has formed, it can be removed by carefully pipetting up and down without touching the beads.
13. Depending on the type of nuclear protein, type of TF, strength of interactions etc. the number (lowering/increasing) of washing steps can be optimized. For example, washing can be reduced if you expect only weak interactions. For optimization, it can

be useful to perform western blot analysis with the flow- through collected in different washing steps.

14. Depending on your experimental purpose, for example when the protein of interest is very lysine-rich, digestion with an alternative proteases (e.g. Arg-C, Thermolysin, Pepsin) can be an option. Choice of an alternative protease demands optimization by the user.
15. Correct assembly of the two plates, i.e. the 96-well Oasis HLB µelution plate and the 96-well wash plate on to the vacuum manifold apparatus and of the vacuum manifold apparatus to the vacuum pump is crucial. Gaps between the plates and the vacuum manifold will result in an incomplete vacuum, prohibiting efficient flow-through of the solutions. To fit the 96-well Oasis HLB µelution plate on top of the 96-well wash plate in to the vacuum manifold, we built an additional support. This in-house built support raised the platform of the 96-well wash plate, bringing it closer to the Oasis HLB µelution plate and closing any gaps. Before starting the clean-up of your sample using the HTS apparatus, 96-well wash plate and 96-well Oasis HLB µelution plate, it is recommended to first test whether the assembly is correctly set up using some test solutions. If only a few samples need to be processed, clean-up can also be performed using desalting columns/tips such as C18 ZipTips (Millipore).
16. Droplets of 50 % Acetonitrile / 5 % formic acid (v/v) can stick to the sides of the wells and reduce the elution volume. Try to pipet the elution solution into the centre of the tube to avoid such droplets and if necessary, remove droplets from the walls with a pipette and then pipet them into the centre without making contcat with the column. Typically, 50-80 µl of sample eluate is recovered.
17. Identification of post-translationally modified peptides, such as acetylated or phosphorylated peptides, is not standardly performed, but can be useful (although largely extending the time for the MS/MS run), as a considerable proportion of the proteome is post-translationally modified.
18. Additional insights and online training/tutorials on the use of these software packages have been made available by the MaxQuant developers at the following link: https://www.youtube.com/channel/UCKYzYTm1cnmc0CFAMhxDO8w
19. Statistical analysis can only be performed if three or more biological replicates have been performed. If certain peptides are not detected at all in the control sample (e.g. extracted from a line expressing a free GFP, ideally under control of the promoter of the gene of interest ), they may not be recognized as statistically significant, although they can still represent important interactors. Thus, non-significant data may also be relevant. Therefore, authors often present a table which summarizes the number of peptides identified and their abundance.

***ACKNOWLEDGEMENTS***

We thank Sjef Boeren and Twan America for helpful suggestions regarding the data analysis and Tom Brabbs and Richard Immink for critical reading of the manuscript. S.J. is supported by the European Commission Seventh Framework-People-2012-ITN Project EpiTRAITS, GA-316965 (Epigenetic regulation of economically important plant traits) and M.B. by an NWO-Veni grant.

***REFERENCES***

**Cox, J., and Mann, M.** (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology **26,** 1367-1372.

**Del Olmo, I., Lopez, J.A., Vazquez, J., Raynaud, C., Pineiro, M., and Jarillo, J.A.** (2016). Arabidopsis DNA polymerase recruits components of Polycomb repressor complex to mediate epigenetic gene silencing. Nucleic acids research **44,** 5597-5614.

**Liang, S.C., Hartwig, B., Perera, P., Mora-Garcia, S., de Leau, E., Thornton, H., de Lima Alves, F., Rappsilber, J., Yang, S., James, G.V., Schneeberger, K., Finnegan, E.J., Turck, F., and Goodrich, J.** (2015). Kicking against the PRCs - A Domesticated Transposase Antagonises Silencing Mediated by Polycomb Group Proteins and Is an Accessory Component of Polycomb Repressive Complex 2. PLoS genetics **11,** e1005660.

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

**Smaczniak, C., Immink, R.G., Muino, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Parcy, F., Xu, L., Carles, C.C., Angenent, G.C., and Kaufmann, K.** (2012b). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proceedings of the National Academy of Sciences of the United States of America **109,** 1560-1565.

**Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J.** (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature methods **13,** 731-740.

**Wendrich, J.R., Boeren, S., Moller, B.K., Weijers, D., and De Rybel, B.** (2017). In Vivo Identification of Plant Protein Complexes Using IP-MS/MS. Methods in molecular biology **1497,** 147-158.

# CHAPTER 4

## Unravelling the protein-protein interactions of flowering time regulators in *Arabidopsis thaliana*

Suraj B. Jamge
Twan America
Bhagyshree B. Jamge
Jan Cordewener
Gerco C. Angenent
Richard G.H. Immink

***ABSTRACT***

*Proteins* are active in many developmental processes in the cell and often act in larger protein complexes. Here, we unravelled the protein-protein interactions of two flowering–time-related proteins, SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) and AGAMOUS-LIKE 24 (AGL24). We observed differences in protein-complex compositions of the SOC1 and AGL24 MADS-domain transcription factors before and after the transition to flowering. In addition to MADS-domain proteins, many different potential transcriptional regulators were found as their interaction partners, suggesting the presence of multiple different higher-order protein complexes containing SOC1 and/or AGL24. In this study, a SOC1-SVP complex was found to be enriched during vegetative as well as reproductive development. Since SOC1 and SVP are known to antagonistically regulate the flowering time process, we followed up on the role and molecular function of this "flowering activator-repressor" complex. To this end, we tested the DNA binding capacity of this complex and examined its potential to either activate or repress expression of selected target genes.

***INTRODUCTION***

Protein-protein interactions (PPIs) are the driving force for the precise control of biological processes in multicellular organisms. These interactions can generally modulate protein function and this notion holds true for transcription factor (TF) proteins, DNA-binding proteins and other important regulators of gene expression. The transcriptional activity of TF proteins depends, to a large extent, on the proteins they interact with. Here we examined the PPI capacity of two Arabidopsis MADS-domain TFs, SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1) and AGAMOUS-LIKE 24 (AGL24), which are involved in the regulation of flowering time and flower development.

Transition to flowering is a major developmental event in angiosperms, but the significance of PPIs during this transition phase has not been studied in much detail. During the vegetative stage of development, FLOWERING LOCUS C (FLC) and SHORT VEGETATIVE PHASE (SVP), both members of the MADS-domain TF family, have been shown to physically interact and to repress flowering. FLC-SVP binds as a complex to the promoter region of the floral integrator genes *SOC1* and *FLOWERING LOCUS T* (*FT*), which results in their repression (Li et al., 2008). FLOWERING LOCUS M (FLM), another MADS-domain TF and member of the *FLC* clade, also interacts with SVP and regulates flowering time in response to signals from the ambient temperature pathway (Pose et al., 2013; Capovilla et al., 2015). Another example of PPIs in flowering involves the 'florigen' signal encoded by *FT*, which is synthesized in leaves and transported to the shoot apex where it interacts with the bZIP TF protein FLOWERING LOCUS D (FD) in order to initiate flowering. In *Arabidopsis*, the FT-FD interaction is likely bridged by 14-3-3 proteins, as shown in rice, and the resulting 'florigen-complex' activates the expression of floral-meristem identity genes such as *LEAFY* (*LFY*) and *APETALA1* (*AP1*) (Abe et al., 2005; Wigge et al., 2005; Taoka et al., 2011; Niwa et al., 2013; Taoka et al., 2013; Ho and Weigel, 2014). Together these examples clearly highlight the significance of PPIs in flowering time control. Knowledge of the interactions of TF proteins in space and time is therefore vital to our understanding of developmental processes in higher plants.
In Arabidopsis, 107 genes encode for MADS-domain TFs, which contribute to a variety of plant developmental events. Most importantly, MADS TFs are prominently involved in flowering time control and floral organ identity specification. Biochemical and genetic interaction studies reveal that MADS TFs function as dimers and as multimeric complexes. According to the Quartet model (Theissen and Saedler, 2001), combination of these TF complexes can specify the different floral organ identities. Recent advances also suggest cofactors and accessory proteins associating with MADS TFs provide their specificity necessary for unique target gene regulation (Smaczniak et al., 2012a; Smaczniak et al., 2012b).

Over the years, direct targets of MADS-domain TFs have been identified, providing insights into their molecular function. However, little is known of the native protein complex composition of MADS TF and their involvement and role in target gene regulation. In this report we studied protein complex composition of the SOC1 and AGL24 MADS TFs.

SOC1 plays a central role in the flowering time control network (Lee and Lee, 2010; Immink et al., 2012). Under favourable environmental and plant developmental conditions, the level of *SOC1* in the shoot apical meristem (SAM) reaches its threshold and triggers the floral transition by activating expression of the floral meristem identity genes *LFY* and *AP1* (Immink et al., 2012). Besides SOC1's role in floral induction, it also acts redundantly with AGL24 and SVP to repress the expression of B, C and E class floral homeotic genes in inflorescence meristems (IMs) and early floral meristems (FM) (Liu et al., 2009). As a consequence, cells stay in a meristematic state for longer and do not directly differentiate into a particular floral organ identity. Once the FM is formed, *SOC1* expression is temporarily diminished in stage 1 and 2 flowers by the action of AP1 (Lee and Lee, 2010; Immink et al., 2012). *SOC1* expression reappears in later stages of flower development and is found in floral organs such as stamens and carpels (Lee and Lee, 2010). However, the level of *SOC1* in these tissues is much lower in comparisons to *SOC1* in the SAM.

Similarly to *SOC1*, *AGL24* has a role in floral induction *(Yu et al., 2002; Michaels et al., 2003)*. *AGL24* acts as a positive regulator of *SOC1 (Michaels et al., 2003)* and promotes flowering in a dosage-dependent manner. It has been shown that AGL24 interacts directly with SOC1, which results in the translocation of the dimer into the nucleus to regulate *LFY* expression (Lee et al., 2008). In addition, AGL24 promotes inflorescence identity (Yu et al., 2004), while it is directly repressed by LFY and AP1 during flower development. *AGL24* has been described to play a redundant role with other flowering time genes during both the floral transition and flower development. For instance, in conjunction with *SOC1* and *FRUITFUL (FUL)* during floral induction (Torti and Fornara, 2012) and with *SOC1* and *SVP* at the early stages of flower development (Liu et al., 2009).

Since both SOC1 and AGL24 have different but overlapping developmental functions and are important players in the floral transition process, we aimed to investigate the protein interaction networks of these two TFs before and after the switch to flowering. We used an affinity purification based mass spectrometry approach to isolate and study PPIs in different development tissues in which the two proteins possess known developmental functions. Besides confirming previously reported interactions, this study reports many novel and stage-specific PPIs. Our study also suggests that AGL24 has a more limited role in the floral induction process than previously proposed. Furthermore, we explore the molecular role and function of a flowering activator-repressor complex composed of SOC1 and SVP.

## *RESULTS*

### Gradual increase in SOC1 protein abundance upon transition to flowering

The model plant *Arabidopsis thaliana* is a facultative long day plant and inductive long day signals greatly expedite the floral transition. The expression of the floral integrator *SOC1* is crucial during the vegetative to reproductive switch. In three weeks old *Arabidopsis* plants grown under non-inductive short day (SD) conditions, SOC1 is absent at the shoot apical meristem (SAM). When these plants are exposed to inductive long day (LD) conditions, there is a gradual increase in SOC1 at the SAM (Immink et al 2012). After seven days of LD exposure the developmental transition to the reproductive stage is morphologically visible (Immink et al 2012). Using GFP-tagged SOC1, Immink et al. (2012) showed that SOC1 protein levels correlate with published mRNA expression patterns, but this method is qualitative. In order to obtain quantitative data on protein abundance levels, a time series experiment was performed. To allow for comparison, the design of this time series was identical to the study by Immink et al. (2012) (Figure 1) . Enriched dissected meristems (EDMs) were collected at each time point from a control line (*pSOC1::GFP*; *GFP* driven by the *SOC1* promoter) and the previously described *gSOC1::GFP* line (a *SOC1-GFP* fusion protein driven by the *SOC1* promoter). Immunoprecipitation of SOC1 protein complexes were performed for all four time points (Figure 1a), followed by mass spectrometry (MS) and quantification analysis with Progenesis QI. We observed a gradual increase in SOC1 protein abundance in the SAM during the transition to flowering (Figure 1b). Despite this increase of SOC1, we did not observe the microscopic switch to flowering as described by Immink et al 2012 at.

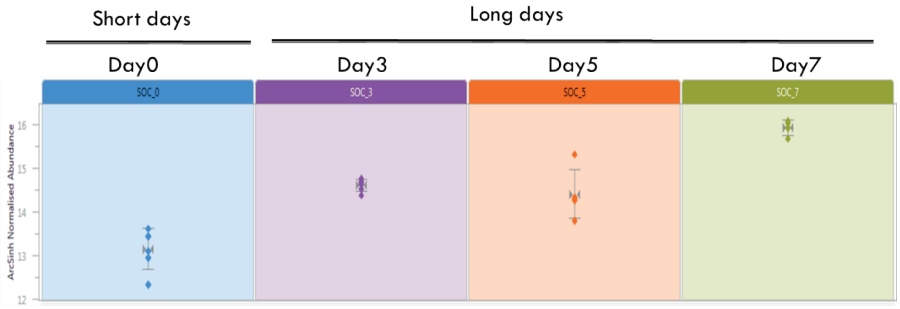
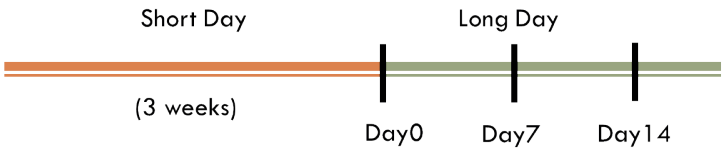



Figure 1. **SOC1 protein abundance during transition to flowering.** Normalized protein abundances (on a log scale) of SOC1 for each biological replicate are plotted (per time point in different colours), average and standard deviations are indicated by grey line and arrow heads. Protein abundance is the sum of peptide-peak intensities determined by Progenesis QI software. Top panel indicates the schematic of experimental setup showing the light conditions and the time points when plant material was collected.

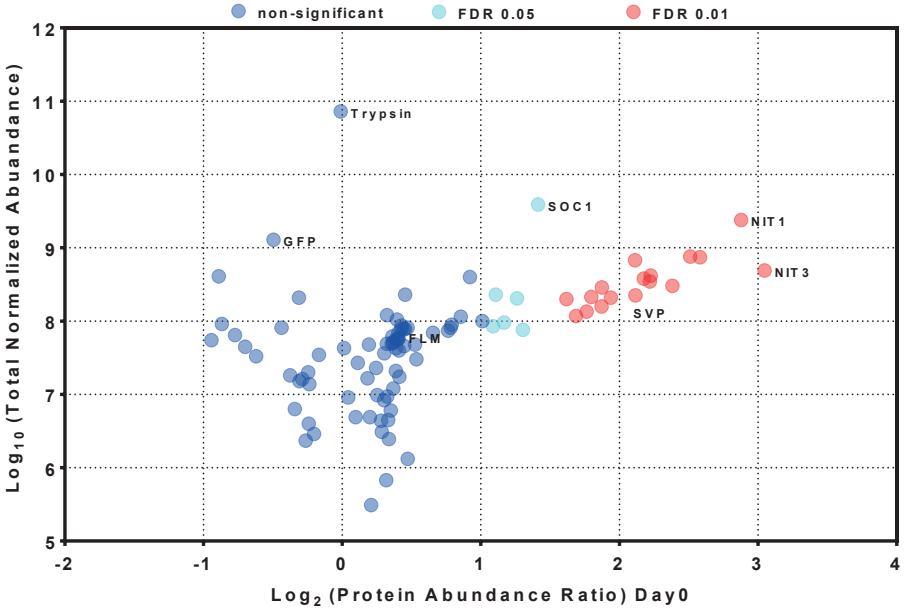
**Dynamics in SOC1 Interactome upon transition to flowering**

To study the differences in the SOC1 interactome during the switch to flowering, we examined SOC1 protein interactions in a different experimental setup spanning three time points (Figure 2a). These time points were, Day0 when *SOC1* expression is not detectable in the SAM, Day7 when SOC1 is present at the SAM and the molecular switch to flowering should have taken place according to previous published work (Immink et al 2012), and Day14 when the first signs of the developing bolt were visible to the eye under our growth conditions. EDMs were collected for this three different time point from a control line (*pSOC1::GFP*) and *gSOC1::GFP* line. Protein complexes were isolated by affinity purification, followed by MS and label-free quantification. For each time point, we identified SOC1 interaction partners that had previously been identified by yeast-based methods (de Folter et al., 2005), revealing the technical soundness of the experiments. In addition, we also immuno-precipitated many novel and previously undescribed SOC1-interacting proteins (Supplemental File S1). At the vegetative stage (Day0) the floral repressors SVP and FLM were found to be in complex with SOC1 (Figure 2b). Upon transition to flowering (Day7), in addition to SVP and FLM, we detected another floral repressor protein, AGAMOUS-LIKE 16 (AGL16), the floral promoter AGL24 and the floral meristem identity protein AP1 (Figure 2c). At Day14, the MADS-domain TF FRUITFULL (FUL) was present amongst the significant interactors in addition to SVP, FLM, and AGL24 (Figure 2d). Apart from the various MADS transcription factors identified (Figure 3), two enzymes, namely NITRILASE 1 (NIT1) and NIT3, were also significantly enriched during the transition to flowering (Figure 2). These have been previously shown to have a role in cell proliferation (Doskocilova et al., 2013), root morphogenesis (Lehmann et al., 2017) and plant-microbe interaction (Howden and Preston, 2009).

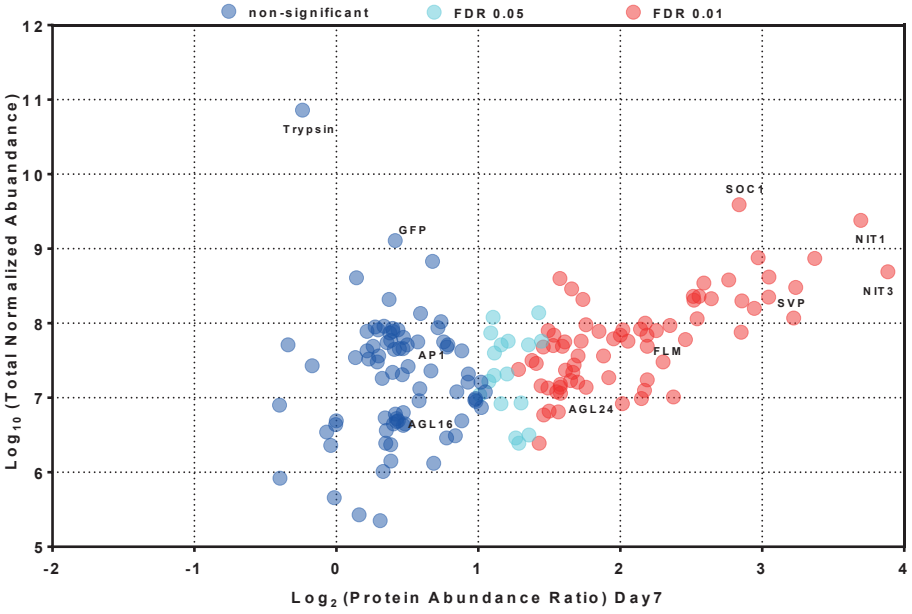
**A)**



B)


non-significant
FDR 0.05
FDR 0.01
Trypsin
SOC1
NIT1
GFP
NIT3
SVP
FLM
Log10 (Total Normalized Abundance)
Log2 (Protein Abundance Ratio) Day0

C)


non-significant
FDR 0.05
FDR 0.01
Trypsin
SOC1
GFP
NIT1
NIT3
SVP
AP1
FLM
AGL24
AGL16
Log10 (Total Normalized Abundance)
Log2 (Protein Abundance Ratio) Day7

**D)**

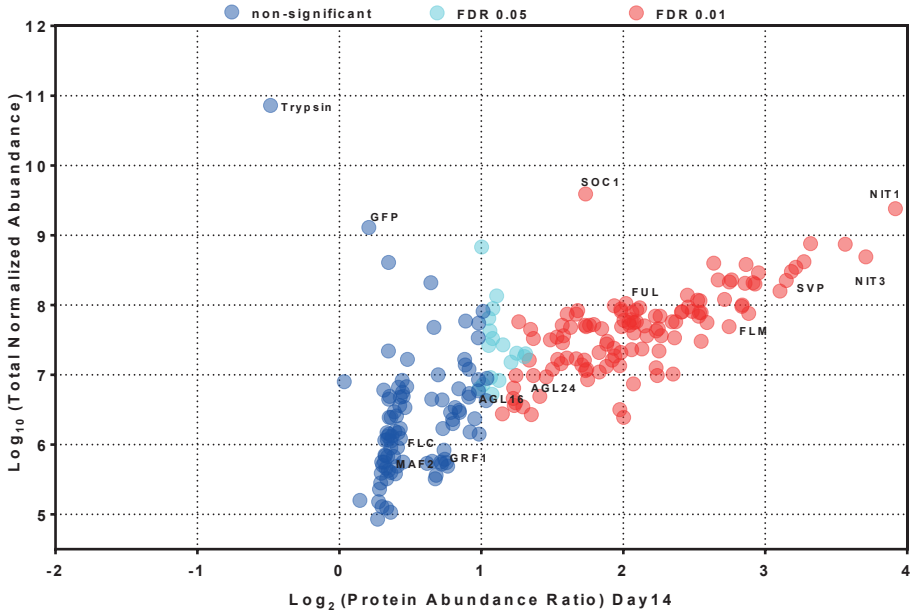



Figure 2. ***In planta* SOC1 protein interaction profiles during floral transition.** A) Schematic of experimental setup, showing the light conditions and the time points when plant material was collected. B) Day0; C) Day7; D) Day14. Graphs represent the normalized protein abundance ratio between *pSOC1:SOC1-GFP* and the *pSOC1::GFP* control plotted against the iBAQ intensities for a particular protein. Significant protein abundance differences between sample and control at FDR 0.01 (s0 =2) or FDR 0.05 (s0 =2) are shown in red and cyan circles, respectively. Blue circles denote non-significant hits. Note that IPs also capture abundant ribosomal proteins, which are generally categorized as common contaminants in affinity purification based mass spectrometry. For a complete list of SOC1 interaction partners identified in the Day0, Day7 and Day14 please refer the Supplemental File S2.

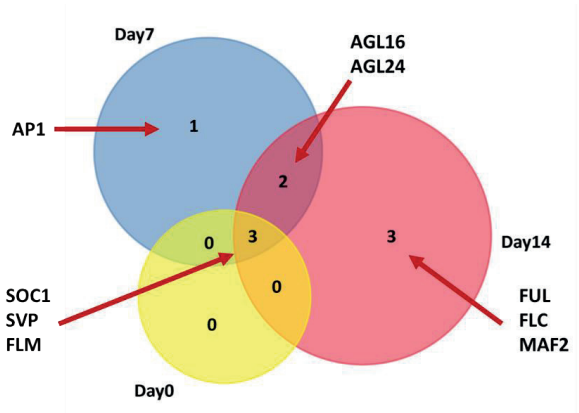



Figure 3. **Overview of SOC1 interactions with MADS-domain TFs.** Venn diagram highlighting all identified SOC1 interactions with MADS TFs in Day0, Day7, Day14 samples.

**SOC1 interacts with floral homeotic proteins during flower development**

In light of SOC1's function during flower development, we also studied SOC1 protein complexes in reproductive tissues (Figure 4). In line with our previous observations, affinity of SOC1 for MADS TFs was clearly noticeable. Among the MADS TFs, we identified the ABC(E) class proteins AP1, AGAMOUS (AG), SEPALLATA1 (SEP1), SEP2, SEP3 and SEP4, flowering promoters such as AGL6, AGL24 and FUL, and floral repressors, such as SVP and FLM (Figure 4). Interactions with other types of proteins were found. For instance with FLOR1 (FLR1), a leucine rich repeat (LRR) protein described as a interaction partner of AG and having a role in flower development (Gamboa et al., 2001; Acevedo et al., 2004). AG itself was also found in complex with SOC1, making it tempting to speculate that SOC1, AG and FLR1 form a multimeric complex. Vegetative storage protein 1 (VSP1) and VSP2, classified as acid phosphateses with diverse functions in nutrient storage and mobilization, plant development and defence, were also identified as part of the SOC1 interactome. Furthermore, NIT1 and NIT3, initially identified during the transition to flowering (Figure 2), were significantly present in the SOC1 interactome in inflorescence tissue (Figure 4).

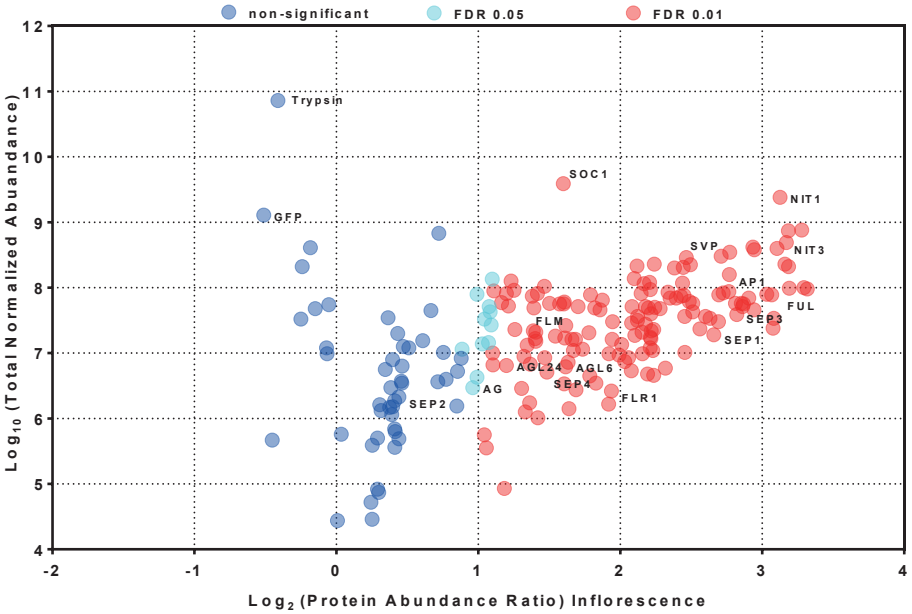
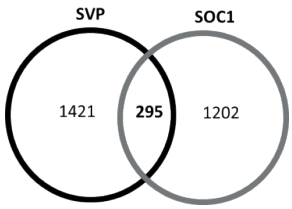



Figure 4. ***In planta*** **SOC1 protein interaction profile in Arabidopsis inflorescences.** Graph represents the normalized protein abundance ratio between the samples *pSOC1:SOC1-GFP* and the *pSOC1::GFP* control, plotted against the iBAQ intensities for a particular protein. Significant protein abundance differences between sample and control at FDR 0.01 (s0 =2) or FDR 0.05 (s0 =2) are shown in red and cyan circles, respectively. Blue circles denotes non-significant hits. For a complete list of SOC1 interaction partners identified in the Inflorescence please refer the Supplemental File S1.

**Common targets of SOC1 and SVP**

SOC1 acts as a flowering activator, whereas SVP functions as a flowering repressor. Nevertheless, amongst all the MADS TFs that we have identified as potential SOC1-interactors, the enrichment of SVP in complex with SOC1 appeared to be highly significant (FDR 0.01, S0 =2). Since both proteins are TFs and can directly bind and regulate gene expression, we asked the question, what would be the transcriptional output of a SOC1-SVP complex. To gain more insight into the potential function of this flowering activator-repressor complex, we looked into the reported genome-wide targets of each TF. Occupancy of genomic loci by TF proteins is often investigated using chromatin immuno-precipitation (ChIP) and a number of such studies have been reported for SOC1 and SVP (Immink et al., 2012; Tao et al., 2012; Gregis et al., 2013; Mateos et al., 2015). In order to compare and explore the common targets and binding sites, we reanalysed the publically available ChIP-seq datasets of SOC1 (Immink et al 2012) and SVP (Mateos et al 2015) (Figure 5). Upon re-analysis, approximately 295 peaks overlapped for SOC1 and SVP and 48.8% of the overlapping peaks are located in promoter regions (Figure 5a and Supplementary Figure 1). The overlapping peaks represent 278 genomic loci (Supplemental File S2). Among the common targets were other TF- and flowering-related genes, including AP2-TFs and the *SOC1* and *SVP* loci themselves, suggesting auto-regulation of these genes (Figure 6, Supplemental File S2). A gene ontology (GO) slim enrichment analysis of the common targets identified several GO categories that were significantly overrepresented, when compared with the *Arabidopsis* genome as background. For molecular functions, GO terms such as DNA binding and transcription factor activity were significantly overrepresented at an FDR $p<0.001$. This result suggests that SOC1 and SVP are master transcriptional regulators acting on other transcription factors. Overrepresentation of the biological processes terms reproductive process, developmental process and responses to stimulus were found (Figure 5b).

**A)**



**B)**

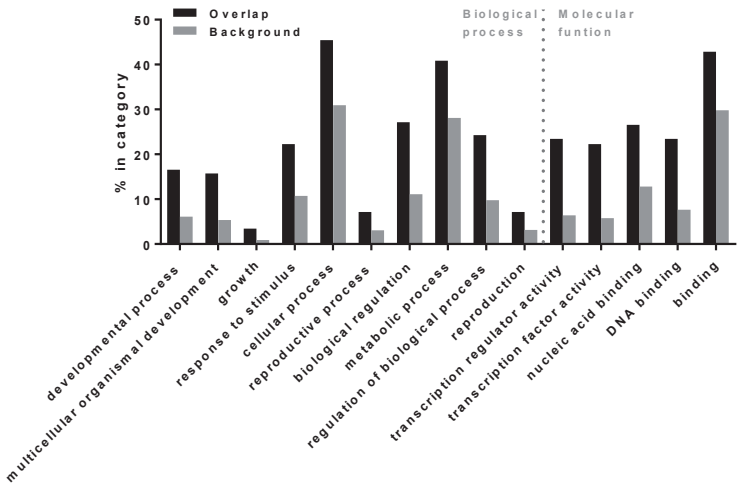



Figure 5. **Distribution and Classification of SOC1 and SVP bound loci.** A) Comparison of the overlapping peaks upon reanalysis of SOC1 and SVP ChIP-seq data. B) GO enrichment analysis of the common targets of SOC1 and SVP.

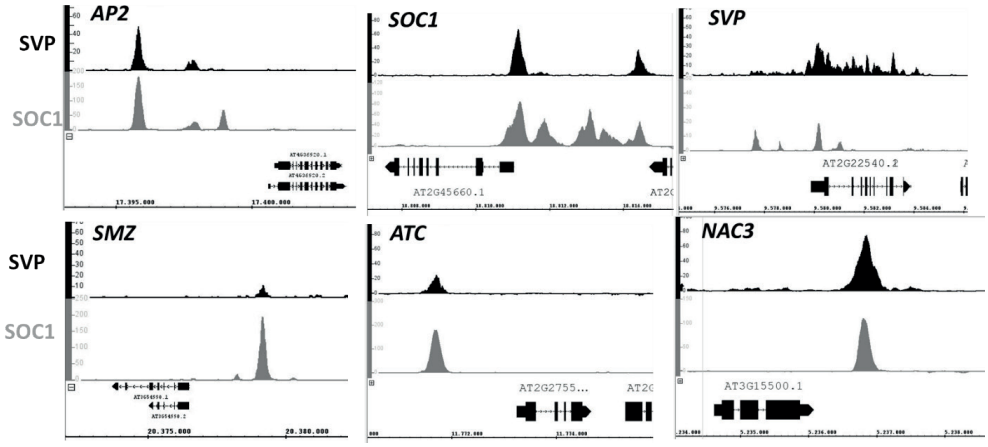



Figure 6. **Binding profiles of selected target genes of SOC1 and SVP**. Snapshot of the Integrative genome browser (IGB) highlighting the enrichment profiles of SVP (in Black) and SOC1 (in grey) at six loci along with the TAIR9 annotated gene models. For a complete list of ChIP-seq overlapping and specific targets of SOC1 and SVP please refer the Supplemental File S2.

**SOC1-SVP complex bind at the common targets *in vitro***

Based on the ChIP-seq analysis, individual binding sites of SOC1 and SVP overlapped at 278 genomic loci. In order to understand whether they bind to their targets independently of each other or as a heteromeric complex, we performed electrophoretic mobility shift assays (EMSAs). Our EMSA results showed a strong binding of both SOC1 and SVP to the CArG motifs in probe fragments representing the centre of the SOC1 and SVP ChIP-seq peaks in the *SMZ*, *ATC*, and *NAC3* promoters. All three loci were enriched for SOC1 and SVP binding according to the ChIP-seq studies (Figure 7). This outcome demonstrates that both proteins can bind individually as a dimer and tetramer to their targets in an *in vitro* assay (Figure 7). However, differences in DNA mobility were observed when the two proteins were incubated together with the probes (Figure 7). This suggests a preference for heterodimeric complexes over homodimers when both TFs are present (Figure 7).

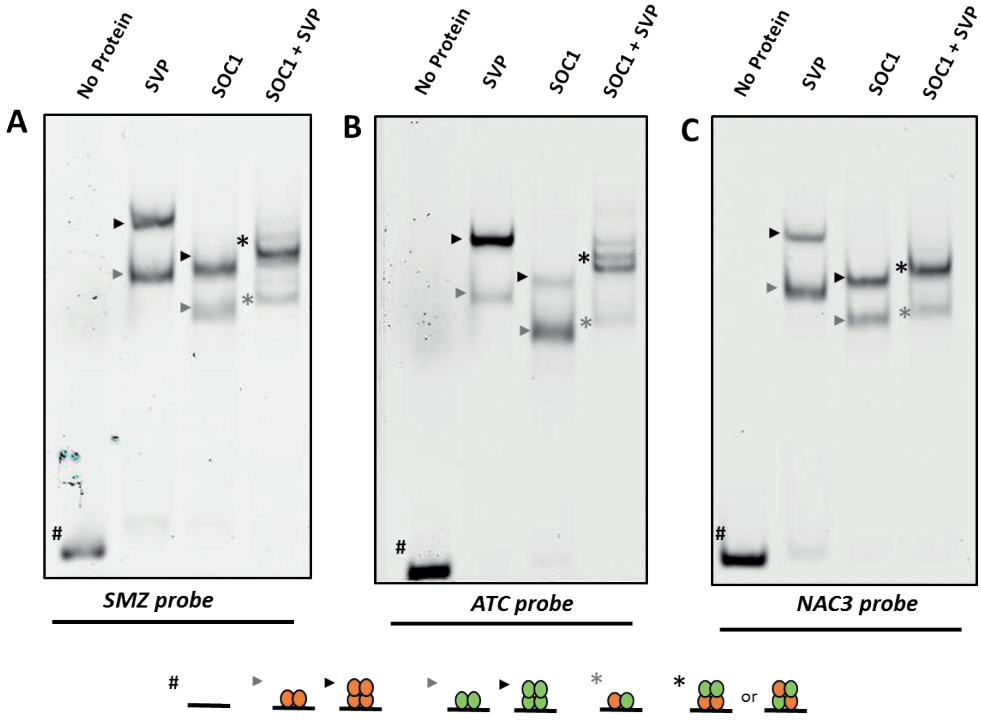



Figure 7**. Direct binding of SOC1-SVP complex at common targets.** Electrophoretic mobility shift assay (EMSA) to test the binding of SOC1, SVP and SOC1-SVP to promoter regions of *SMZ* (A), *ATC* (B) and *NAC3* (C). Lane 1: DNA probe (P) only, Lane 2: P plus SVP protein, Lane 3: P plus SOC1 protein and Lane 4 P plus SOC1-SVP proteins. Different order complex are represented by black arrowheads and asterisk for homo or heterotetramers, respectively, and with grey arrowheads and asterisk for homo or heterodimers respectively. '#' denotes free DNA probe, Orange and green eclipse represent SVP and SOC1 protein respectively bound to the probe.

### Mode of action of SOC1-SVP complex

Preference for binding of the selected promoter fragments by SOC1-SVP heterocomplexes can be observed in the *in vitro* EMSA assays (Figure 7). To gain an insight into the molecular mode of action of this activator-repressors complex, we investigated the transcriptional response of several targets in the presences of SOC1, SVP and the SOC1-SVP heterodimer. We employed a transactivation system, where SOC1, SVP and SOC1-SVP were transiently expressed in *Arabidopsis* leaf mesophyll protoplasts and the transcriptional response of selected target genes was measured by qPCR. Our preliminary results showed no difference in the endogenous expression of *SOC1*, *SVP, PIF4 and AGL16* (Figure 8). A difference in expression of *SMZ* and *AGL19* was evident when both TFs were present, an indication that the SOC1-SVP complex might act as transcriptional activator for these two targets. In the case of *FCA* and *LEJ2*, increase in expression was observed in the presence of SOC1, but SOC1-SVP resulted in no significant difference when compared with SOC1 alone. Together, these preliminary results highlight that there is combinatorial activity of SOC1 and SVP in regard to target gene expression.



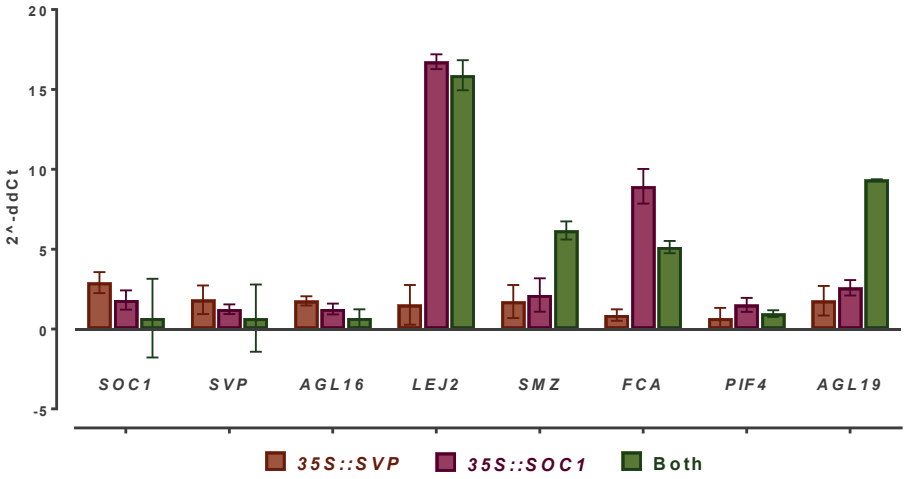
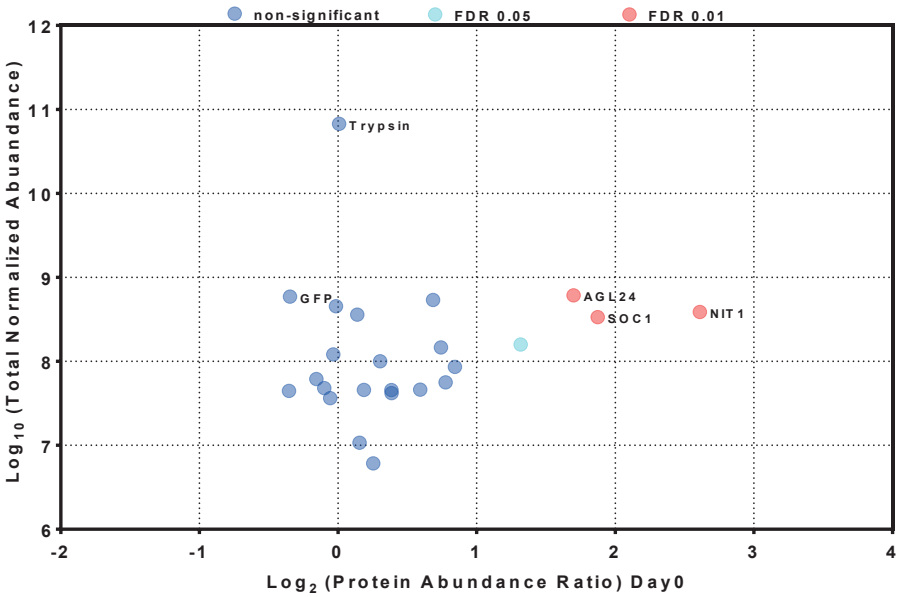



Figure 8. **Endogenous expression of SOC1/SVP target genes in a transient assay.** Leaf protoplast were transfected with *35S::SVP* (orange), *35S::SOC1* (pink) and both constructs (green) and the endogenous expression of a select set of potential target genes was measured by qPCR. The graph shows the average between two biological replicates. Bars indicate standard deviation between the replicates.

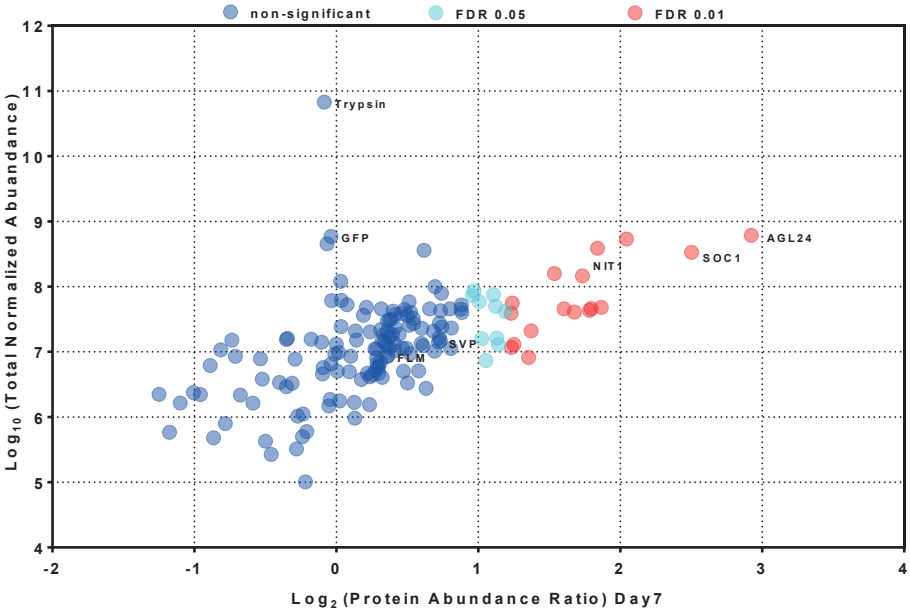
**AGL24 and SOC1 have common interaction partners during transition to flowering**

*AGL24* is often reported together with *SOC1* for its involvement in floral induction. Therefore, we decided to unravel the protein-protein interactions of AGL24 using the same method as for SOC1 (Figure 2a). At Day0 the only MADS-domain TF identified in complex with AGL24 was SOC1 (Figure 9a). Contrary to this, in our SOC1 immuno-precipitation (IP) at Day0, we did not identify AGL24 (Figure 2b), instead the floral repressors SVP and FLM were found. An explanation for this result could be the overall lower abundance of AGL24 at the Day0 time point. Upon induction of flowering at Day7, AGL24 interactions with SOC1 and floral repressors, such as FLM and SVP, were identified (Figure 9b). With the exception of AP1, all other identified AGL24-interacting MADS TFs were similar to what was observed for SOC1 at Day7 (Figure 2c). The results for Day14 were again similar to that of SOC1 (Figure 2d), with the MADS TFs SOC1, AGL24, FUL, SVP, FLM, and AGL16 showing a significant interaction with AGL24 (Figure 9c).

**A)**



**B)**




**C)**

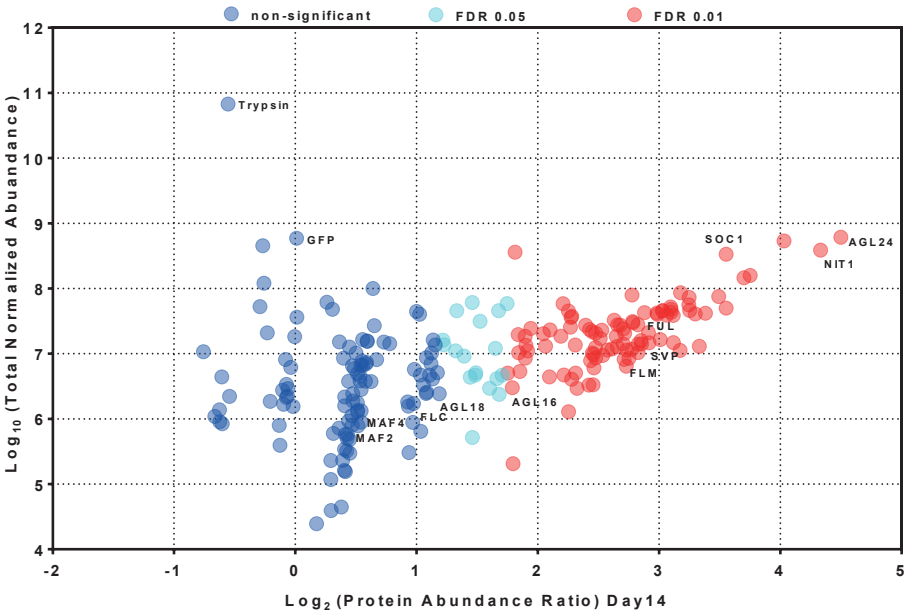



Figure 9. ***In planta*** **AGL24 protein interaction profile during floral transition.** A) Day0; B) Day7; C) Day14. All graphs represent the normalized protein abundance ratio between *pAGL24:AGL24-GFP* and the *pAGL24::GFP*

control, plotted against the iBAQ intensities for a particular protein. Significant protein abundance difference between sample and control at FDR 0.01 (s0 =2) or FDR 0.05 (s0 =2) are shown in red and cyan circles, respectively. Blue circles denote non-significant hits. For a complete list of AGL24 interaction partners identified in the Day0, Day7 and Day14 please refer the Supplemental File S3.

### The AGL24 inflorescence interactome

Considering the function of AGL24 during floral organ development, we examined the *in planta* protein-protein interactions of AGL24 in inflorescence tissue. Since genetic redundancy was shown between *SOC1* and *AGL24* in their role during reproductive development. Except for AP1, none of the other SOC1-interacting ABC class MADS proteins were found in complexes with AGL24 (Figure 10). The most significant interactions were with FUL and AP1. SOC1 and SVP were also present among the non-significant hits (Figure 10, Supplemental File S3).

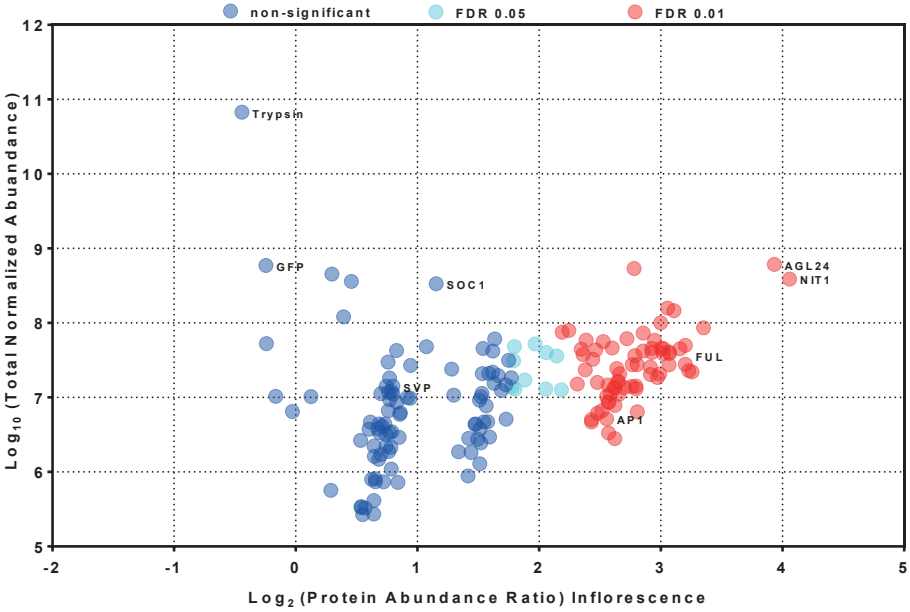



Figure 10. ***In planta*** **AGL24 protein interaction profile during floral development.** Graph represents the normalized protein abundance ratio between *pAGL24:AGL24-GFP* and the *pAGL24::GFP* control, plotted against the iBAQ intensities for a particular protein. Significant protein abundance difference between sample and control at FDR 0.01 (s0 =2) or FDR 0.05 (s0 =2) are shown in red and cyan circles, respectively. Blue circles denote non-significant hits. For a complete list of AGL24 interaction partners identified in the Inflorescence please refer the Supplemental File S3.

### *SOC1* acts partially independent of *AGL24*

The *Arabidopsis soc1* and *agl24* mutants show a delayed flowering response and both are referred to as flowering promoting genes (Valverde et al., 2004). A previous study described partial independence between *SOC1* and *AGL24 in* regulation of flowering based on mutant analyses and overexpression in a L*er* background (Michaels et al., 2003). However, another

study showed that SOC1 and AGL24 can heterodimerize and the authors proposed that AGL24 is key for translocation of biologically active SOC1 from cytosol into the nucleus, where the heterodimer activates *LFY* expression to commence the floral transition process (Lee et al., 2008). This observation and hypothesis suggest full dependency on AGL24 of SOC1's role in flowering time control in the Col-0 background. Nevertheless, *soc1* mutants exhibit a very late flowering time phenotype, whereas only a mild delay was observed for *agl24* mutants when compared to wild type. Considering these observations and that there is no clear consensus on the importance of AGL24 in relation to SOC1's function in the flowering time process, we decided to re-examine their genetic interaction. We investigated to what extent SOC1's function in flowering time regulation is dependent on the presence of AGL24. To answer this, we generated a transgenic line overexpressing SOC1 and showing precocious flowering, and used this line to cross into *agl24-2* mutant background, resulting in *agl24-2,35S::SOC1*. The flowering time phenotype was scored and compared between the various lines (Figure 11). Our data shows that *SOC1* overexpression results in early flowering and that *agl24* mutants flowered later than wild type. These observations are in line with the previously described phenotypes (Michaels et al., 2003). Interestingly, a transgenic line overexpressing *SOC1* in an *agl24* mutant background, flowered at the same time as wild type plants and significantly earlier than the *agl24* mutant. This reconfirms the notion that *SOC1* can still induce flowering in the absence of *AGL24* and therefore, *SOC1* is at least partially independent of *AGL24* during the floral transition process.



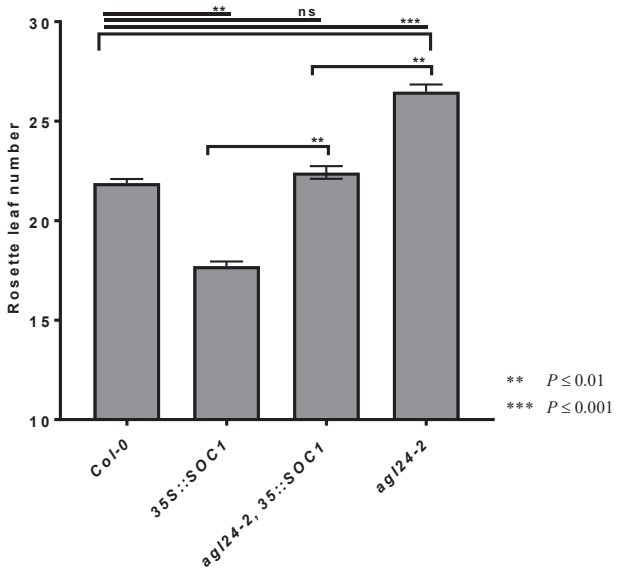



Figure 11. **Effect of overexpression (OE) of *SOC1* on *AGL24*.** Flowering time of wild type Col-0, OE:SOC1, *agl24-2,* and *agl24-2* and OE:SOC1 under long day conditions and scored by total number of rosette leaves (RLN). Asterisks indicate significant results in comparisons to Col-0 or *agl24-2, 35::SOC1* (** p-value ≤ 0.01; *** p-value ≤ 0.001). P-values were calculated using student T-test in graphpad prism.

***DISCUSSION***

The composition of transcriptional complexes recruited at regulatory sequences can influence the transcriptional machinery and final gene regulation response. Here, we uncover the developmental dynamics of protein complex composition for two well-known MADS-domain proteins. Our protein complex isolation results suggest spatio-temporal differences in protein complex formation upon transition to flowering.

SOC1, regarded as a floral integrator in *Arabidopsis*, is primarily involved in the promotion of flowering (Lee and Lee, 2010). We isolated protein complexes of SOC1 during the floral transition process and found MADS-domain proteins to be the major transcriptional regulators in complex with SOC1. Upon exposure to inductive photoperiod signals, changes in SOC1 protein complex stoichiometry is evident. The presence of AP1 and FUL in complex with SOC1 clearly reflects the role of SOC1 in the transition to flowering and floral meristem and organ development. Before the transition (the Day0 time point) there was a significant interaction between SOC1 and SVP. This is unexpected since SVP has an opposite function to SOC1 in flowering time regulation. A possible explanation could be that by forming a complex with SOC1, SVP sequesters SOC1 and thereby SOC1 is no longer able to bind to its target genes under short day conditions. A similar role has been described for SVP's interaction with FLM isoforms (Pose et al., 2013; Capovilla et al., 2015). As SVP is involved in different flowering pathways, it would be worthwhile to check if this interaction recurs during the vegetative phase of plants growing under long day conditions or different ambient-temperature. The double mutant *soc1 svp* phenotype varies depending on photoperiod conditions. Under LD conditions, *soc1 svp* is similar to wild-type, whereas under SD conditions the double mutant appears to phenocopy *svp* mutants, i.e. an early flowering phenotype (Torti et al., 2012).

Besides the SOC1 interactome, we also reported on the interaction partners of AGL24, another MADS-domain protein with a role in floral induction (Yu et al., 2002; Michaels et al., 2003). SOC1 appeared to be the only MADS-domain protein in complex with AGL24 in before floral induction. At the Day7 and Day14 time points, most MADS-domain protein AGL24 interactions overlap with those observed for the SOC1 IPs, suggesting both are part of a similar higher-order complex upon transition to flowering.

Genetic and expression studies have shown that *SOC1* and *AGL24* positively regulate each other and, in addition, can independently promote flowering (Liu et al., 2008). The expression patterns of both genes overlap during vegetative growth and upon transition to flowering in the SAM (Torti and Fornara, 2012), whereas differences in their expression profiles are observed during flower development. *AGL24* is expressed early on in floral primordia, whereas *SOC1* only appears from floral stage two (Smyth et al 1990). Despite their expression in these tissues, single or double mutants of *soc1* and *agl24* do not exhibit any floral defects. Only in triple mutants, such as in *soc1 agl24 svp*, do their role in floral organ development becomes apparent (Gregis et al., 2006, 2008; Liu et al., 2009; Lee and Lee, 2010). Thus, in addition to

the complexes formed during floral transition, we also unravelled the protein complex composition of SOC1 and AGL24 during flower development.

Our SOC1 IP on closed inflorescence buds uncovered an array of MADS-domain proteins, including proteins that are part of the floral-quartet model, such as the SEP subfamily proteins, along with AG and AP1. Identification of these floral homeotic proteins suggest that in addition to SOC1's early function in repression of the B, C and E class genes, (Liu et al., 2009; Lee and Lee, 2010), SOC1 also plays a role at later floral organ development stages together with the floral ABCE MADS-domain proteins. Upon examination of protein complexes in closed inflorescence buds for AGL24, we observed remarkable differences in the protein-protein interaction composition of AGL24 in comparisons to SOC1 for this tissue type. AGL24 has a significant interaction with FUL and AP1, whereas for SOC1 many other floral homeotic proteins were also identified. This observation strengthens the previously proposed idea that SOC1 acts as a major hub and through its specific interactions it participates in a number of processes during flower development (de Folter et al., 2005). According to the ABC model of flower development, formation of each floral whorl is defined by a specific MADS tetramer (Theissen and Saedler, 2001). Here we show occurrence of *in planta* higher-order complexes, where SOC1 is able to associate with the proteins from the floral quartet model (e.g. AG, SEPs). The molecular function of SOC1 in such a complex still remains unclear. One likely explanation could be that SOC1's activation domain could be essential for transcriptional activation, especially when the interaction partners such as AG do not possess one. An *Arabidopsis* flower undergoes a series of developmental stages before it sets seeds (Smyth et al., 1990; Alvarez-Buylla et al., 2010,). The interactome data presented here is derived from a mix of early and late developmental tissue (stage one to twelve). To better understand the flower development process, isolation of protein complexes from individual developmental stages may shed light on the role, extent and involvement of each MADS-domain proteins throughout this process. To achieve this, the synchronised system in the *ap1 cal* double mutant, in which floral development can be induced by AP1:GR, would be of great help (Pajoro et al., 2014).

Genome-wide studies indicate that a single TF is able to directly target hundreds of genes within a genome. Our proteomic analysis suggests that TFs likely function in complexes and depend on specific interactions for their activity. Therefore, it is becoming more clearer that the stoichiometry of TFs in a complex is key in defining target gene regulation (Smaczniak et al., 2017). To test this hypothesis, the combinatorial effects of two TFs (SOC1 and SVP) with opposite functions in flowering time control were investigated. SOC1 acts as an activator of flowering, whereas SVP is referred as the floral repressors. Analysis of previously published ChIP-seq data showed that they both have common shared targets. With the *in vitro* protein-DNA binding assay study, we confirmed the preference for binding by the SOC1-SVP heterodimer at these target loci. Our preliminary results using a transient expression assay in *Arabidopsis* leaf protoplasts indicate that the combinatorial activity of SOC1-SVP can influence the transcriptional behaviour of certain target genes. Further examination of several other

common targets will be necessary to obtain a clear understanding on the function of a SOC1-SVP complex. Since MADS TFs act in higher-order complexes, it would be worthwhile to study the effect of different TF complexes on target gene regulation. Currently this type of *in planta* quantitative analysis still remains a technical challenge. In addition, prior knowledge of the protein complex composition is also necessary. With the recent advancements in methodologies and their further optimizations, knowledge about *in vivo* protein interactions from individual cell types (Long et al., 2017) and their effect on target gene regulation is starting to emerge.

***MATERIAL AND METHODS***

Plants material and growth conditions: Most plants used in this study were in the Col-0 background, including the GFP tagged lines and the overexpression lines. All seeds were sown on rock wool blocks in trays and were placed in a cold room (4°C) for two nights for stratification. Following cold stratification, trays were moved to a walk-in-growth chamber with LED light based illumination. Plants were grown under a relative humidity (RH) of 70% at 22 °C and with 94.35 µmol m−2 s−1 photosynthetically active radiation (PAR). Plants were regularly watered with HYPONEX® plant nutrient solution (1gm/L). We obtained the seeds for the *gSOC1:GFP* line from Richard Immink and for *gAGL24:*GFP line from Cezary Smaczniak. In addition, the *agl24-2* line was kindly provided by Martin Kater. For long day conditions the plants were exposed to 16h light / 8h dark and for short day conditions 8h light/ 16h dark. Flowering time experiments were performed under long day conditions. For the floral transition immuno-precipitation experiments, plants were initially grown under short day conditions for three weeks before being moved to long day conditions for another two weeks. The day on which plants were transferred from short day to long day is designated as day0 for all time course experiments. Enriched dissected meristems (EDM) were harvested at Day0, Day7 and Day14. For inflorescence immuno-precipitation experiments, plants were grown under long day condition and inflorescence were collected. All sampling was performed on four biological replicates and was done at the end of the afternoon (zeitgeber 14h).

Generation of Transgenic lines: In order to increase the sensitivity of our method and to minimise the identification of false positives we generated transgenic lines. *GFP* was expressed under control of the native plant promoters *(SOC1, AGL24*) and these lines were used as background control for the immuno-precipitation (IP) experiments. These lines were generated by amplify their promoter sequences from gDNA of WT plants. To PCR amplify promoter sequence of SOC1 the primer pair PDS2051 GGTGTTTGCTCCTCTAGTTCTGA and PDS7338 ATCTTCTTCTTTAGTTAATTTCCCTTG was used and for amplification of AGL24 promoter sequence PDS2047 TCGTTCCTTATAGCGGTGGAT and PDS7339 TTTACCAGATCTCTCCTTCACTACTG was used. These promoter fragments were cloned into Gateway entry vector pCR8/GW/TOPO (Invitrogen), followed by a subcloning via Gateway LR reaction into destination vector CZN1493 (GW-NLS-GFP/pGREEN). The generated GFP constructs were transformed in to Arabidopsis thaliana ecotype Col-0 plants using standard floral dip method (Clough and Bent, 1998).

LC-MS/MS based complex isolation : High-resolution LC-MS/MS and quantitative data analysis with MaxQuant was performed as described previously in Chapter 3 (Jamge et al., 2018), on “crude extracts” for four biological replicates for *gSOC1:GFP*, *gAGL24:GFP* and their respective control lines. In addition to the LC-MS/MS analysis as described in Chapter 3, LC-MS/MS have also been analysed with Progenesis software (see Figure1). In Progenesis LC-MS/MS analysis peak detection and quantitation is performed separately from the peptide identification with database search engine. Peptides Identifications are then linked to quantified peaks. Peak

intensity data are normalised between individual injection data by a correction factor , which brings the median value of all data points (excluding major outliers) per injection to the same value for all injections. Protein quantification is then calculated by summing the normalised peak intensities of all peptides matched per protein. (For details see: http://www.nonlinear.com/progenesis/qi-for-proteomics/v2.0/faq/how-normalisation-works.aspx )

EMSA : The EMSA probe sequences were derived from the promoter regions of the *SMZ*, *ATC*, and *NAC3* promoter regions, taking into account the SOC1 and SVP ChIP-seq binding sites within these regions. The probes for EMSA contain a canonical CArG-box in the centre. The probe fragments were amplified from wild type Col-0 genomic DNA and cloned into the pGMET vector system (See table below for all primers and probe sequences). The coding sequences (CDS) of SVP and SOC1 were amplified from wild type Col-0 cDNA and cloned into the pSPUTK expression vector. Proteins were synthesized using the TnT® SP6 High-Yield Wheat Germ Protein Expression System (Promega) according to manufactures instructions. EMSAs were performed as described previously by Smaczniak et al 2012 (Smaczniak et al., 2012b) with minor modifications. Such as, the probes were fluorescently labelled using DY-682 and the labelling was performed by PCR using vector-specific DY-682-labelled primers followed by agarose gel extraction. EMSA gels were visualized using a LiCor Odyssey imaging system at 700 nm.

ChIP-seq analysis: Raw ChIP-seq data for SOC1 and SVP was downloaded from NCBI and had the GEO accession numbers GSE45846 and GSE54881, respectively. Reads were mapped to TAIR10 with Bowtie 2 (Langmead and Salzberg, 2012). PCR duplicates were removed using SAMtools (Li et al., 2009). Replicates were merged and peaks were called using MACS v2 (Zhang et al., 2008) with input as control and a minimum FDR cut off of 0.01. Peak regions were further analysed in R using the Bioconductor packages ChIP seeker (Yu et al., 2015) and ClusterProfiler (Yu et al., 2012) to identify unique and overlapping peaks between the two transcription factosr. Plant GO slim term analysis was performed using the AgriGO v2.0 platform (Tian et al., 2017).

Construction of the fluorescent plasmids: Complete ORFs of MADS box genes *SOC1* (At2g45660) and *SVP* (AT2G22540) were amplified with gene-specific primers to remove the stop codon and for C-terminal in-frame fusion with the coding region of CFP and YFP (from pECFP and pEYFP; CLONTECH catalog nos. 6075-1 and 6004-1, respectively). Subsequently, the SOC1-CFP and SVP-YFPYFP" products were cloned in the expression vector pGD120 {Immink, 2002 #2286}. The resulting MADS-box CFY/YFP fusion vectors were maxi prepped using the ZymoPURETM Plasmid maxiprep kit as per the manufacture protocol and used for protoplast transfection assays. All the above indicated constructs were tested by restriction analysis and sanger sequencing for the inserted fragment.

Protoplast assay: The "Tape-Arabidopsis sandwich" method was followed to isolate protoplast from three week old long day Arabidopsis leafs (Wu et al., 2009). Protoplast isolation and

transfection assay were performed as described previously (Rios et al., 2017). Transfected protoplast were transferred to a six well petri dish with 1ml WI buffer as described previously (Niu and Sheen, 2012). The petri dish were incubated at 25°C in dark and 16 hour after incubation confocal imaging was performed.

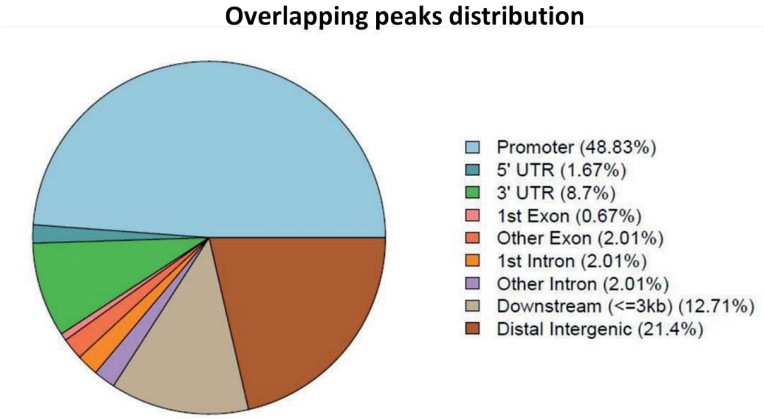
RNA isolation from protoplast and real time PCR analysis: RNA extraction was performed using Trizol Method as per the manufactures protocol, total RNA was extracted from transfected protoplast. Total RNA was subjected to DNase I digestion using the TURBO DNA-free™ kit (Ambion) and RNA integrity was checked on 1% (w/v) agarose gel after DNase treatment. First-strand cDNA was synthesized from 400ng of total RNA using iScript™ RT supermix kit (BIO-RAD), following the manufactures protocol. 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.



Primers used in this study :

| Name | Primer Sequence 5'-> 3'' | Description |
|---|---|---|
| **PDS9876** | CCCAGCCGTTCAATTTATCCAG | SMZ probe |
| **PDS9877** | ACCTATTGGAACTAGAATGAACAGAGA | SMZ probe |
| **PDS9882** | CATGGGCTGCTTGTTTGACT | LEJ2 probe |
| **PDS9883** | ACGAACAAAAAGCGTAGATGTGT | LEJ2 probe |
| **PDS9945** | TGGGTCGCCAACATTAACAT | ATC probe |
| **PDS9946** | GGATGTATTGCAAAGAATATATCCC | ATC probe |
| **PZN1** | ACGAAGACTGGAAATTGGTTGGG | LEJ2 qPCR |
| **PZN2** | TTTCCAGGTGCTGTCAACCTCAG | LEJ2 qPCR |
| **PZN765** | CGCAATTGTTCATTGGGTTA | SOC1 qPCR |
| **PZN766** | TCTCAGTACTGCTAAACCTGTTTTT | SOC1 qPCR |
| **PZN767** | GAGCTCCGACACTTCCCTTA | SVP qPCR |
| **PZN768** | GACATTGTCTCTTGTTACTACCGAGT | SVP qPCR |
| **PDS8442** | GGTTTGTGTTTTGGGGCCTTG | UBQ10 qPCR HKG |
| **PDS8443** | CGAAGCGATGATAAAGAAGAAGTTCG | UBQ10 qPCR HKG |
| **PDS8480** | CCCATCACAGAACGATCTCGAT | PIF4 qPCR |
| **PDS8481** | AGGAGCCACCTGATGAGGAACT | PIF4 qPCR |
| **PDS8492** | TGCATCAATGCCTTCTCCAAGCAA | AGL19 qPCR |
| **PDS8493** | TCAGCAAGCGAGAGACGAAACATC | AGL19 qPCR |
| **PDS8484** | TGTTCGAACGAGAGCAACAG | FCA qPCR |
| **PDS8485** | AACGGCTGTAATTGGGTCTG | FCA qPCR |
| **PDS8414** | AGCAAGTTTATTTGGGCGGGTTTG | SMZ qPCR |
| **PDS8415** | TGATAGCAGCTCGGTCGTAAGC | SMZ qPCR |
| **PDS2231** | ACATGAAAAGGTTTCAGAGGTCGAG | AGL16 qPCR |
| **PDS2232** | AGATGGACATGTTCGTTCGAGGTAT | AGL16qPCR |

**Supplementary Material**




**Supplementary Figure 1.** Distributions of the overlapping peaks found for SOC1 and SVP. Pie charts represent peak distributions among genomic constituents.

**Description of additional files :** (Will be provided upon request)

**Supplemental File S1 |** Complete list of identified protein interaction partners for SOC1 (Time points : Day0, Day7, Day14 and inflorescences).

**Supplemental File S2 |** List of all the identified overlapping and specific targets of SOC1 and SVP upon re-analysis of the ChIP-seq.

**Supplemental File S3 |** Complete list of identified protein interaction partners for AGL24 (Time points : Day0, Day7, Day14 and inflorescences).

***REFERENCES***

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309, 1052-1056.

Acevedo, F.G., Gamboa, A., Paéz-Valencia, J., Jiménez-García, L.F., Izaguirre-Sierra, M., and Alvarez-Buylla, E.R. (2004). FLOR1, a putative interaction partner of the floral homeotic protein AGAMOUS, is a plant-specific intracellular LRR. Plant Science 167, 225-231.

Alvarez-Buylla, E.R., Benitez, M., Corvera-Poire, A., Chaos Cador, A., de Folter, S., Gamboa de Buen, A., Garay-Arroyo, A., Garcia-Ponce, B., Jaimes-Miranda, F., Perez-Ruiz, R.V., Pineyro-Nelson, A., and Sanchez-Corrales, Y.E. (2010). Flower development. The Arabidopsis book 8, e0127.

Capovilla, G., Schmid, M., and Pose, D. (2015). Control of flowering by ambient temperature. Journal of experimental botany 66, 59-69.

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

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.

Doskocilova, A., Kohoutova, L., Volc, J., Kourova, H., Benada, O., Chumova, J., Plihal, O., Petrovska, B., Halada, P., Bogre, L., and Binarova, P. (2013). NITRILASE1 regulates the exit from proliferation, genome stability and plant development. The New phytologist 198, 685-698.

Gamboa, A., Paez-Valencia, J., Acevedo, G.F., Vazquez-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.

Gregis, V., Sessa, A., Colombo, L., and Kater, M.M. (2006). AGL24, SHORT VEGETATIVE PHASE, and APETALA1 redundantly control AGAMOUS during early stages of flower development in Arabidopsis. The Plant cell 18, 1373-1382.

Gregis, V., Sessa, A., Colombo, L., and Kater, M.M. (2008). AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis. The Plant journal : for cell and molecular biology 56, 891-902.

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.

Ho, W.W., and Weigel, D. (2014). Structural features determining flower-promoting activity of Arabidopsis FLOWERING LOCUS T. The Plant cell 26, 552-564.

Howden, A.J., and Preston, G.M. (2009). Nitrilase enzymes and their role in plant-microbe interactions. Microbial biotechnology 2, 441-451.

Immink, R.G., Pose, 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 physiology 160, 433-449.

Jamge, S., Angenent, G.C., and Bemer, M. (2018). Identification of In Planta Protein–Protein Interactions Using IP-MS. In Plant Chromatin Dynamics: Methods and Protocols, M. Bemer and C. Baroux, eds (New York, NY: Springer New York), pp. 315-329.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods 9, 357-359.

Lee, J., and Lee, I. (2010). Regulation and function of SOC1, a flowering pathway integrator. Journal of experimental botany 61, 2247-2254.

Lee, J., Oh, M., Park, H., and Lee, I. (2008). SOC1 translocated to the nucleus by interaction with AGL24 directly regulates leafy. The Plant journal : for cell and molecular biology 55, 832-843.

Lehmann, T., Janowitz, T., Sanchez-Parra, B., Alonso, M.P., Trompetter, I., Piotrowski, M., and Pollmann, S. (2017). Arabidopsis NITRILASE 1 Contributes to the Regulation of Root Growth and Development through Modulation of Auxin Biosynthesis in Seedlings. Frontiers in plant science 8, 36.

Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E., and Yu, H. (2008). A repressor complex governs the integration of flowering signals in Arabidopsis. Developmental cell 15, 110-120.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

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, C., Chen, H., Er, H.L., Soo, H.M., Kumar, P.P., Han, J.H., Liou, Y.C., and Yu, H. (2008). Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis. Development 135, 1481-1491.

Long, Y., Stahl, Y., Weidtkamp-Peters, S., Postma, M., Zhou, W., Goedhart, J., Sanchez-Perez, M.I., Gadella, T.W.J., Simon, R., Scheres, B., and Blilou, I. (2017). In vivo FRET-FLIM reveals cell-type-specific protein interactions in Arabidopsis roots. Nature 548, 97-102.

Mateos, J.L., Madrigal, P., Tsuda, K., Rawat, V., Richter, R., Romera-Branchat, M., Fornara, F., Schneeberger, K., Krajewski, P., and Coupland, G. (2015). Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in Arabidopsis. Genome biology 16, 31.

Michaels, S.D., Ditta, G., Gustafson-Brown, C., Pelaz, S., Yanofsky, M., and Amasino, R.M. (2003). AGL24 acts as a promoter of flowering in Arabidopsis and is positively regulated by vernalization. The Plant journal : for cell and molecular biology 33, 867-874.

Niu, Y., and Sheen, J. (2012). Transient expression assays for quantifying signaling output. Methods in molecular biology 876, 195-206.

Niwa, M., Daimon, Y., Kurotani, K., Higo, A., Pruneda-Paz, J.L., Breton, G., Mitsuda, N., Kay, S.A., Ohme-Takagi, M., Endo, M., and Araki, T. (2013). BRANCHED1 interacts with FLOWERING LOCUS T to repress the floral transition of the axillary meristems in Arabidopsis. The Plant cell 25, 1228-1242.

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.

Pose, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G.C., Immink, R.G., and Schmid, M. (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. Nature 503, 414-417.

Rios, A.F., Radoeva, T., De Rybel, B., Weijers, D., and Borst, J.W. (2017). FRET-FLIM for Visualizing and Quantifying Protein Interactions in Live Plant Cells. Methods in molecular biology 1497, 135-146.

Smaczniak, C., Immink, R.G., 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., Muino, J.M., Chen, D.J., Angenent, G.C., and Kaufmann, K. (2017). Differences in DNA Binding Specificity of Floral Homeotic Protein Complexes Predict Organ-Specific Target Genes. The Plant cell 29, 1822-1835.

Smaczniak, C., Immink, R.G., Muino, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Parcy, F., Xu, L., Carles, C.C., Angenent, G.C., and Kaufmann, K. (2012b). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proceedings of the National Academy of Sciences of the United States of America 109, 1560-1565.

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

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.

Taoka, K., Ohki, I., Tsuji, H., Kojima, C., and Shimamoto, K. (2013). Structure and function of florigen and the receptor complex. Trends in plant science 18, 287-294.

Taoka, K., Ohki, I., Tsuji, H., Furuita, K., Hayashi, K., Yanase, T., Yamaguchi, M., Nakashima, C., Purwestri, Y.A., Tamaki, S., Ogaki, Y., Shimada, C., Nakagawa, A., Kojima, C., and Shimamoto, K. (2011). 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. Nature 476, 332-335.

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

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic acids research.

Torti, S., and Fornara, F. (2012). AGL24 acts in concert with SOC1 and FUL during Arabidopsis floral transition. Plant signaling & behavior 7, 1251-1254.

Torti, S., Fornara, F., Vincent, C., Andres, F., Nordstrom, K., Gobel, U., Knoll, D., Schoof, H., and Coupland, G. (2012). Analysis of the Arabidopsis shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. The Plant cell 24, 444-462.

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., and Coupland, G. (2004). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Science 303, 1003-1006.

Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U., and Weigel, D. (2005). Integration of spatial and temporal information during floral induction in Arabidopsis. Science 309, 1056-1059.

Wu, F.H., Shen, S.C., Lee, L.Y., Lee, S.H., Chan, M.T., and Lin, C.S. (2009). Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. Plant methods 5, 16.

Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382-2383.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology 16, 284-287.

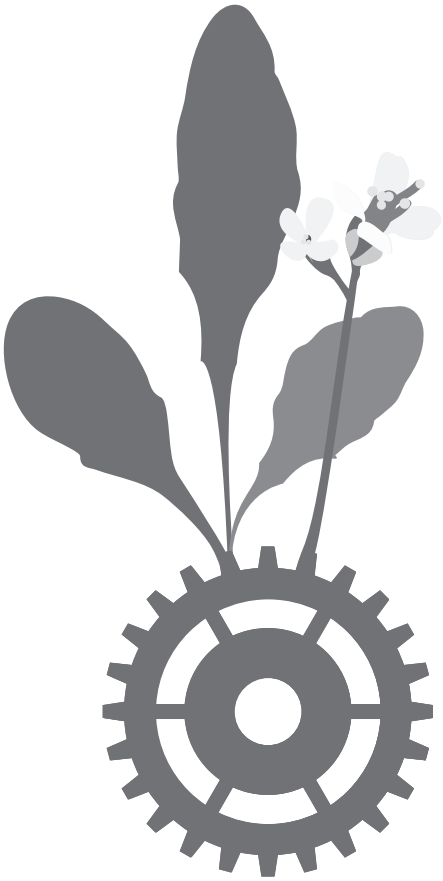
Yu, H., Xu, Y., Tan, E.L., and Kumar, P.P. (2002). AGAMOUS-LIKE 24, a dosage-dependent mediator of the flowering signals. Proceedings of the National Academy of Sciences of the United States of America 99, 16336-16341.

Yu, H., Ito, T., Wellmer, F., and Meyerowitz, E.M. (2004). Repression of AGAMOUS-LIKE 24 is a crucial step in promoting flower development. Nature genetics 36, 157-161.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology 9, R137.

# CHAPTER 5

## POUSHP a novel interactor of SHORT VEGETATIVE PHASE and involved in flowering time control



Suraj B. Jamge
Francesca Bellinazzo
Gerco C. Angenent
Richard G.H. Immink

***ABSTRACT***

Flowering is an important developmental switch in the life cycle of angiosperms and members of the MADS-domain transcription factor (TF) family act as central players in the flowering gene regulatory network (GRN). MADS-domain proteins are known to bind DNA as dimers and are able to form higher-order complexes. We aimed to unravel the composition of MADS-domain complexes regulating the floral transition in *Arabidopsis thaliana* and focussed our study on SHORT VEGETATIVE PHASE (SVP). SVP acts as a floral inhibitor during the vegetative stage of development and represses precocious expression of ABCE-class MADS-domain TFs during early flower development. To understand how SVP can fulfil these different functions, native SVP protein complexes have been isolated from vegetative and reproductive tissues. Particular SVP interactors were identified in both tissue types, but also differences in complex components were found. We identified a large number of previously reported SVP interactors, demonstrating the efficiency of our approach, but also various potential novel interaction partners. Here we report the functional analysis of a gene coding for one of the identified new SVP-interactors. Plants mutated for this gene were late flowering and therefore, we named it *POUSHP (PSH)*. We provide evidence that PSH is mainly nuclear localised, interacting with SVP and positively regulating flowering time only in the presence of the floral inhibitor SVP.

***INTRODUCTION***

Transition from vegetative to reproductive growth is a defining moment in the life of all flowering plants. Environmental and endogenous cues are critical in the timely execution of this developmental phase switch i.e. transition to flowering. The model plant *Arabidopsis thaliana* has been extensively studied in this aspect, resulting in identification of multiple genetic pathways controlling flowering time (Fornara et al., 2010; Bouche et al., 2016). Some pathways mediate flowering by monitoring internal cues, including hormone levels and the developmental age of the plant, whereas other pathways involve environmental signals, such as ambient temperature, exposure to winter cold, durations of day-length, and changes in light quality. Ultimately, the signals from these individual pathways converge at the level of the floral integrator genes *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1* (*SOC1*) and *FLOWERING LOCUS T* (*FT*) (Fornara et al., 2010). It is the spatio-temporal regulation of these floral integrators that defines the moment of flowering and hence the onset of reproductive growth.

During vegetative growth, the shoot apical meristem (SAM) develops leaf primordia on its flanks. Upon transition to reproductive development, the SAM developmental fate changes into inflorescence meristem (IM) identity and this IM develops multiple determinate floral buds in *Arabidopsis*. Vegetative growth in *Arabidopsis* is maintained by the function of key flowering repressor genes. Among these, *SHORT VEGETATIVE PHASE* (*SVP*), a gene encoding a MADS-domain transcription factor (TF) is known for its role in the ambient temperature, gibberellin and autonomous flowering pathways (Hartmann et al., 2000). *SVP* is broadly expressed during the vegetative phase, but its activity is absent in the IM and later on reappearing during early stages of flower development (stages 1 and 2) (Hartmann et al., 2000).

Molecular analysis and functional studies revealed that *SVP* is a central player in the flowering regulatory network (Lee et al., 2007; Gregis et al., 2008; Li et al., 2008). During the vegetative phase, the SVP protein acts in complexes with other floral repressors, such as FLOWERING LOCUS C (FLC) (Li et al., 2008) and FLOWERING LOCUS M (FLM) (Pose et al., 2013), and these complexes bind to promoter regions of the floral integrator genes *SOC1* and *FT* to suppress flowering (Li et al., 2008). Apart from its repressive function during vegetative growth, SVP is also of importance during flower development. This is evident from genetic studies and ectopic expression phenotypes of *SVP* in *Arabidopsis (Gregis et al., 2009)*. In addition to a late-flowering phenotype, overexpression of *SVP* resulted in altered floral morphology leading to formation of secondary flowers, flowers with shoot-like structures, and flowers terminating with carpelloid leaves and stamens (Masiero et al., 2004; Liu et al., 2007; Severing et al., 2012). Mutant analyses revealed that SVP acts in the floral meristem redundantly with the MADS domain proteins AGAMOUS-LIKE 24 (AGL24) and SOC1 to maintain the floral meristem state by suppressing the E-class gene SEPALLATA3 (SEP3) (Gregis et al., 2006, 2008; Liu et al., 2009).

Application of chromatin-immunoprecipitation followed by sequencing (ChIP-seq) has led to the identification of genome-wide binding sites of SVP. These ChIP-seq studies identified genes involved in flowering time and meristem development as putative direct targets of SVP (Tao et al., 2012; Gregis et al., 2013; Mateos et al., 2015). SVP also auto-regulates its own gene expression during the vegetative and reproductive developmental phases. *SVP* is phylogenetically related to another MADS-box gene *AGL24*. Interestingly, *AGL24* has an opposite effect on flowering time as *SVP*. However, during flower development, *AGL24* and *SVP* are co-expressed at early stages ( Stage 1 and 2) and function redundantly (Gregis et al., 2006, 2008; Gregis et al., 2009). Protein interaction studies revealed that SVP is able to interact with other MADS-domain proteins, including FLC, FLM, SOC1, AGL15, AGL16, and AGL21 (Pelaz et al., 2001; de Folter et al., 2005; Gregis et al., 2006; Li et al., 2008; Lee et al., 2013; Balanza et al., 2014; Hu et al., 2014; Hwan Lee et al., 2014). MADS-domain proteins bind to DNA as dimers and quaternary complexes and recent studies showed that these proteins recruit also other transcription factors and cofactors (Smaczniak et al., 2012b). For SVP, only a handful of potential interactions with non-MADS TFs are known. For instance, a J-domain protein AtJ3 was identified as a direct interactor of SVP and thereby regulates the downstream targets of SVP, such as *SOC1* and *FT* (Shen et al., 2011). In another study, a chromatin regulator, TERMINAL FLOWER 2 / LIKE HETEROCHROMATIN PROTEIN 1 (TFL2/LHP1), supposedly orchestrates *SEPALLATA3* (*SEP3)* expression during floral patterning in a complex with SVP (Liu et al., 2009). Despite these few examples, we are far from understanding in which complexes SVP is active and how this single protein can act as both a floral repressor during the vegetative stage of development and a regulator of the ABC-class MADS-box genes during flower development.

Here, we aimed at unravelling the SVP interactome i.e. the protein complex composition, during the vegetative and reproductive stages of development. To this extent, we implemented a robust proteomics approach that allowed us to isolate native SVP protein complexes by affinity purification. This effort resulted in the confirmation of various previously identified SVP interactors and the identification of potential novel interaction partners of SVP. The function of one of these SVP interacting proteins, POUSHP (PSH), was investigated in detail, showing its role as a novel regulator of flowering time.

***RESULTS***

**Analysis of SVP complex formation during the vegetative stage of development**

To provide insight into the protein complex formation of the floral repressor SVP before the switch to flowering, protein complex isolations were performed on three weeks old *Arabidopsis* plants grown under non-inductive short day (SD) conditions. We used a previously reported transgenic line expressing SVP from its native promoter and C-terminally tagged with the GREEN FLUORESCENT PROTEIN (GFP) (Gregis et al., 2013). SVP protein complexes were isolated from rosette tissue using GFP antibodies and further characterized by liquid chromatography (LC-MS/MS) (Smaczniak et al., 2012a). Several hundred proteins (>600) were

identified in these immuno-precipitated protein complex mixtures. Subsequently, filtering and post-processing using stringent cut-offs, followed by label free quantification analysis was performed to identify interactions of high confidence (Figure 1a) (Smaczniak et al., 2012a; Jamge et al., 2018). Our results confirmed previously reported interactions of SVP, e.g. MADS-domain proteins FLM and SOC1 (Figure 1b), which were previously identified using yeast two-hybrid technology (de Folter et al., 2005). Unexpectedly, the most well studied and characterized interaction partner of SVP, the floral repressor FLC (Li et al., 2008), was not present in our stringently selected list of interactors. Upon closer examination, we found one peptide of FLC in the unfiltered IP dataset. We reasoned that the low coverage might be due to the fact that *FLC* is hardly expressed in the Col0 genetic background due to the *frigida (fri)* mutation. To test this hypothesis, the GFP tagged genomic SVP construct was crossed into the SF2 FRI+ background (Michaels and Amasino, 1999), followed by native complex isolations on rosette tissue of three week old plants grown under SD conditions. As expected, FLC was now in the list of significantly enriched proteins (Supplemental Figure 1).

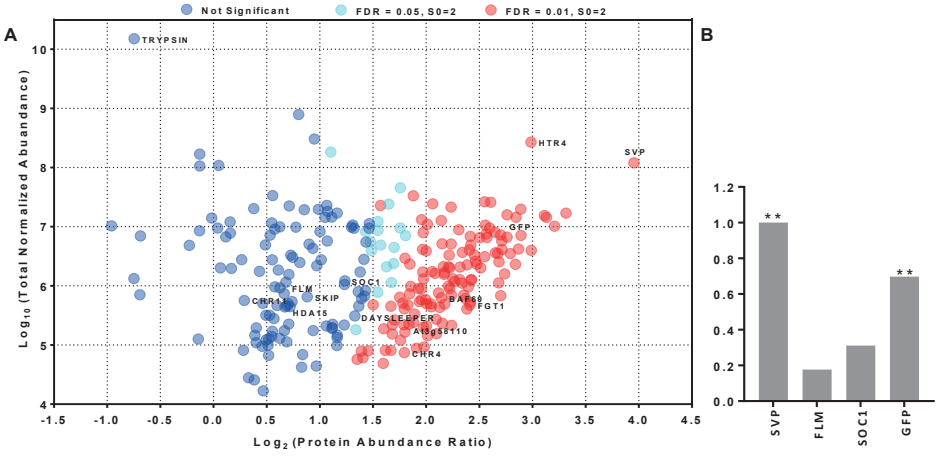



Figure 1. ***In planta* SVP protein interaction profile from vegetative stage.** A) Graph represents the (normalized) protein abundance ratio between pSVP:SVP-GFP and the wild type control, plotted against the iBAQ intensities for each protein in rosette tissues. Significant protein abundance difference between sample and control at FDR 0.01 (s0 =2) or FDR 0.05 (s0 =2) are shown in red and cyan circles, respectively. Blue circles denote non-significant hits. B) Average protein abundance between IP and control, scaled to the ratio of bait protein, for MADS-domain proteins and GFP. Asterisks indicate significant results (* p-value ≤ 0.05; ** p-value ≤ 0.01). For a complete list of SVP interaction partners identified in the vegetative stage please refer the Supplemental File S1.

### Analysis of SVP protein complexes in the reproductive phase

SVP is expressed and functional in both the vegetative and reproductive phase of *Arabidopsis* development and therefore, we also isolated SVP protein complexes from inflorescences, using a similar approach as for the vegetative rosette tissue (Figure 2a). As expected, based on the results of previously performed yeast two-hybrid studies and genetic analyses, other MADS-domain proteins were identified as SVP interactors in the inflorescence, such as SOC1 and AP3 (Figure2b).

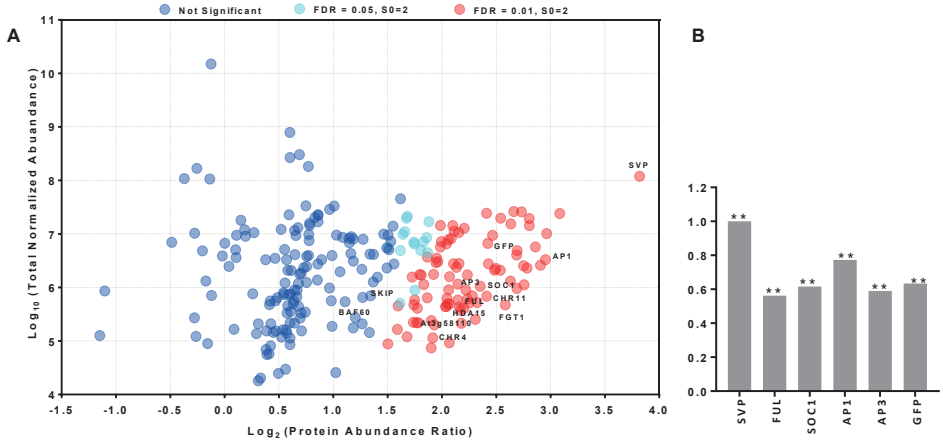



Figure 2. ***In planta* SVP protein interaction profile during the reproductive phase.** A) Graph represents the (normalized) protein abundance ratio between pSVP:SVP-GFP and the wild type control, plotted against the iBAQ intensities for each protein in reproductive tissue. Significant protein abundance difference between sample and control at FDR 0.01 (s0 =2) or FDR 0.05 (s0 =2) are shown in red and cyan circles, respectively. Blue circles denote non-significant hits. B) Average protein abundance between IP and control samples, scaled to the ratio of bait protein, for MADS-domain proteins and GFP. Asterisks indicate significant results (* p-value ≤ 0.05; ** p-value ≤ 0.01). For a complete list of SVP interaction partners identified in the reproductive phase please refer the Supplemental File S1.

### Analysis and comparison of tissue specific SVP interactomes

As a next step, the SVP interactomes from the two analysed developmental stages, the adult vegetative stage, represented by rosette tissue and the reproductive stage of development, represented by inflorescence tissue, were compared. We hypothesized that the different SVP functions in these two developmental stages are reflected in different protein complexes. Indeed, specific protein-protein interactions were confined to a tissue type, whereas others were identified in both developmental stages (Table 1 and Figure 3). For instance, SVP interactions with MADS-domain TFs were more abundant in the reproductive stage than during the vegetative non-flowering stage (Figure 3). One of the best examples is the identification of SVP in complex with FLM in non-flowering plants (Figure 1b), which is in accordance with the proposed role of this complex in flowering repression (Pose et al., 2013).

During the reproductive stage, specific interactions with floral homeotic proteins were clearly evident (Figure 2b and 3). For example, SEP3, an E class protein, known to act redundantly with other SEP proteins in floral patterning, was identified in inflorescences only (Ditta et al., 2004; Goto et al., 2001; Pelaz et al., 2000; Theissen and Saedler, 2001). Besides these tissue-specific enriched interactions, we observed a significant number of proteins overlapping between the two tissues types.

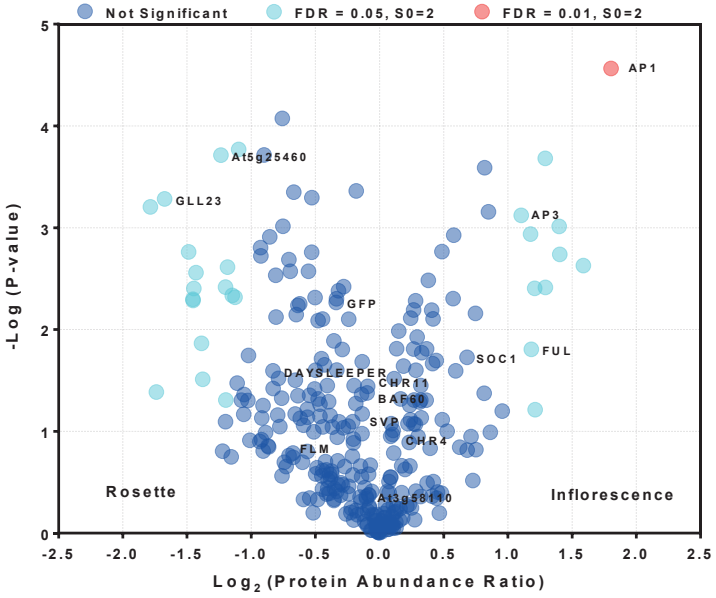



Figure 3. ***In planta* SVP protein interactions.** Volcano plot of the IP results of SVP in reproductive (inflorescence) versus vegetative (rosette) tissues. $Log_2$ protein abundance ratios of reproductive versus vegetative samples are shown. The circles indicate significant protein abundance differences between the two tissues, types with FDR set to 0.01 or 0.05 and s0 parameter set to 2. Complete IP LC-MS/MS analysis for SVP comparing vegetative and reproductive tissue please refer the Supplemental File S2.

### Novel potential SVP interactors

Since we aimed to identify and characterize novel and previously unknown SVP interactors, we focused our study on potential interactors that appeared to be of high confidence and that were previously unreported for SVP. Among these, we identified several proteins that are either directly or indirectly related to transcriptional control and therefore, obvious candidates as interactor of the SVP TF (Table 1, Supplemental File S1). This list includes various SWI/SNF chromatin-remodelling factors, such as CHROMATIN-REMODELLING PROTEIN 4 (CHR4), CHR11, and BAF60 (http://www.chromdb.org, (Verbsky and Richards, 2001; Jerzmanowski, 2007)). We identified FORGETTER1 (FGT1), known to interact with chromatin remodeller proteins and to affect nucleosome dynamics during transcription (Brzezinka et al., 2016). SKIP, which is a transcriptional co-regulator and an important component of the

spliceosome machinery (Wang et al., 2012), was also identified. Furthermore, a histone deacetylase enzyme, HDA15, mainly involved in chromatin compaction and thereby transcriptional repression (Liu et al., 2014), was among the SVP interactors. Besides these, many uncharacterized proteins were found to be enriched in SVP complexes (Table 1 and Supplemental File S1 )

| Protein Name | All peptides | Unique peptides | Sequence coverage [%] | Vegetative Phase | Reproductive Phase |
|---|---|---|---|---|---|
| **MADS Transcription Factors** | | | | | |
| **SVP** | 13 | 13 | 47.2 | ** | ** |
| **FLM** | 2 | 2 | 13.8 | * | |
| **AP1** | 8 | 8 | 21.5 | | ** |
| **SOC1** | 4 | 4 | 21 | ns | ** |
| **AP3** | 2 | 2 | 9.1 | | ** |
| **FUL** | 3 | 3 | 22.2 | | ** |
| **Chromatin remodellers** | | | | | |
| **BAF60** | 5 | 5 | 9.6 | ** | ns |
| **CHR11** | 9 | 9 | 12.9 | ns | ** |
| **CHR4** | 3 | 3 | 1.9 | ** | ** |
| **Other known Factors** | | | | | |
| **DAYSLEEPER** | 2 | 2 | 4.2 | ns | |
| **FGT1** | 9 | 9 | 9 | ** | ** |
| **HDA15** | 2 | 2 | 3.6 | ns | ** |
| **KNAT4** | 2 | 1 | 6.1 | ns | |
| **SKIP** | 4 | 4 | 8 | ns | ns |
| **Uncharacterized Factors** | | | | | |
| **At1g10580** | 3 | 3 | 5.8 | ns | ns |
| **At1g48610** | 4 | 4 | 30.3 | ** | ** |
| **At3g58110** | 3 | 3 | 4.5 | ** | ** |
| **At4g21520** | 3 | 3 | 10.8 | ns | ns |

Table 1. **Subset of identified interaction partners of SVP.** List of known and novel interaction partners of SVP along with the number of identified (all and unique) peptides and the % sequence coverage. Statistical significance was calculated by permutation based FDR estimation (Hubner et al., 2010) with FDR set to 0.01 and the s0 parameter set at 2. Asterisks indicate significant results (* p-value ≤ 0.05; ** p-value ≤ 0.01), ns = not significant. For a complete list of SVP interaction partners identified in this study please refer the Supplemental File S1.

### Flowering time analysis of SVP interactors

*SVP* is prominently involved in floral repression and loss of function *svp* mutants display an early flowering phenotype. Based on this knowledge, we decided to score flowering time for mutants in a selected set of genes encoding putative novel SVP interaction partners, identified in rosette tissue before the switch to flowering. Flowering time was scored under long day conditions and for four out of the six characterized mutant lines, a significant flowering time phenotype was found (Supplemental Figure 2). One candidate in particular caught our

interest, because it is so far a uncharacterized gene and the mutant line was late flowering, in contrast to the early flowering phenotype observed for the *svp* mutant (Figure 4). To confirm and validate our findings, a second mutant allele was identified, which also had a significant late flowering phenotype under long day conditions (Figure 4). Based on these observations, the gene was named *POUSHP (PSH)* (Sanskrit: पौष्प, transliteration: pauSpa, meaning – the one who allows it to flower or blossom).

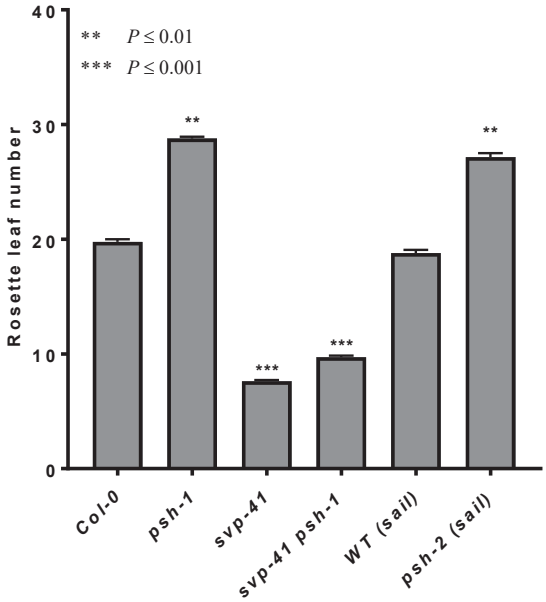



Figure 4. ***POUSHP mutants are*** **late flowering under long day conditions in *Arabidopsis*.** Flowering time screen of wild type (WT) and mutant lines in long day conditions. The single mutants *psh-1, and svp-41* and the double *svp-41 psh-1* mutant are in the Col-0 background. The *psh-2* allele was identified from the SAIL collection and was therefore compared to the WT SAIL line. Please refer to supplemental figure 3 for the schematic details of *psh* mutants. The experiment was performed on three biological replicates and the bar indicates the standard error. Asterisks indicate significant results in comparisons to their wild type (Col-0) (** p-value ≤ 0.01; *** p-value ≤ 0.001). P-values were calculated using student T-test in graphpad prism.

**POUSHP is expressed in broad range of tissue throughout *Arabidopsis* development**

To obtain insights into the temporal and tissue specific expression of *PSH*, qRT-PCR was performed. *PSH* expression was found to be present in all tissues that were analysed in wild type *Arabidopsis* Col-0 (Figure 5). Two weeks old *Arabidopsis* seedlings showed very low expression in comparison to other tissues tested. Expression levels of *PSH* were relatively higher in flowers (closed and open) and developing siliques compared to seedlings. Overall, *PSH* appears to be present across many developmental stages during the life cycle of *Arabidopsis.*

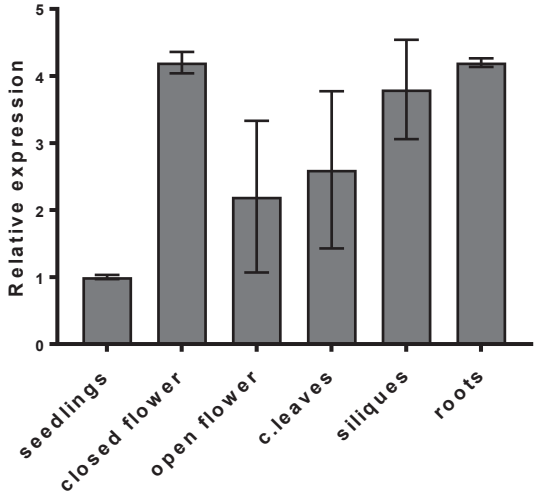



Figure 5. **qRT-PCR analysis of *POUSHP* expression in Col-0 wild type plant tissues.** The relative expression of *PSH* in different plant tissue is shown in comparison to expression of *PSH* in two week old seedlings. The experiment was performed in triplicates and the bars indicate the standard error. Expression was normalized against *UBQ10* expression.

### *SVP* is epistatic to *POUSHP*

As previously described (Hartmann et al., 2000; Masiero et al., 2004) and shown in this study, *svp-41* mutants exhibit a clear early flowering phenotype. On the contrary, *psh* mutants displayed a significant late flowering phenotype under long day conditions. Our *in planta* proteomics approach revealed PSH in complex with SVP during both the vegetative and reproductive stage of development. To better understand the genetic interaction between *PSH* and *SVP*, we constructed a double mutant of these genes. *psh-1* was crossed with *svp-41* and a stable double mutant line s*vp-41 psh-1* was generated. Subsequently, flowering time analyses were performed on this double mutant line, together with the single mutants and wild type Col-0. The *svp-41 psh-1* double mutant flowered much earlier than the *psh-1* single mutant and was not significantly different in flowering time response than the *svp-41* mutant (Figure 4). Hence *SVP* is epistatic to *PSH*, suggesting that the PSH protein fulfils its function in flowering time control in an SVP dependent manner.

### *POUSHP* protein sequence and evolutionary analysis

The full-length PSH (At3G58110) protein sequence was recovered from ARAPORT11 (Cheng et al., 2017) and blastp was performed across different databases. PSH appears to be a plant specific protein and conserved in Angiosperms. No protein blast hits were found in gymnosperms, mosses and ferns. Phylogenetic analysis identified one paralog for PSH in *Arabidopsis thaliana*, which is At2G42370 (Figure 6). Protein sequence comparison of PSH with this paralog (designated POUSHP-like, PSL) showed 47.7% identity (72.3% similarity).

Orthologs of PSH were found in both monocot and dicot species. *In silico* protein sequence motif analysis of PSH predicts the presence of a nuclear localization signal (NLS), an indication that PSH is localized in the nucleus. We also identified three coiled coil regions, D/E rich repeats and an SMC_N domain in PSH (Supplemental Figure 4). D/E rich repeats are often associated with providing binding specificity for target proteins and act as an activation domain for transcription factor proteins (Chou and Wang, 2015).

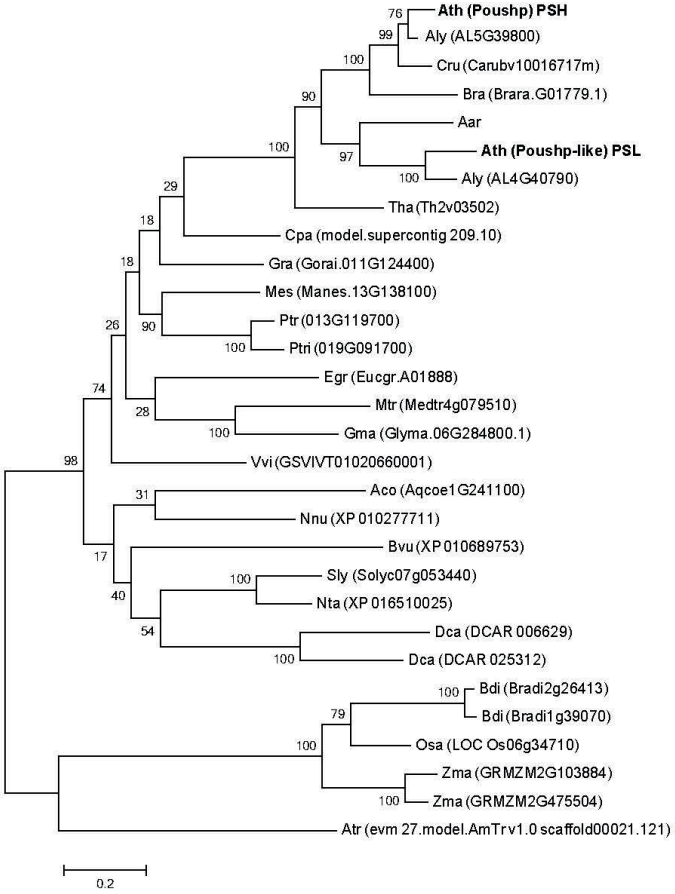



Figure 6. **Phylogenetic tree of *At3g58110* (*POUSHP)* and it orthologs in selected monocot and dicot species**. Protein alignment was made by MUSCLE using the maximum-likelihood method with a 1000 replicate bootstrap analysis. The phylogenetic tree was constructed in MEGA5.

**POUSHP is nuclear localized**

An overexpression construct containing YFP (yellow fluorescent protein) tagged to the N terminus of PSH (35S::YFP-PSH) was created. Transient expression of this fusion protein in *Arabidopsis* mesophyll leaf protoplasts was analysed using confocal microscopy to study *in planta* localization of PSH. PSH localization was observed in the nucleus as well as the cytoplasm (Figure 7). The nuclear signal was more pronounced, although we observed mainly larger aggregates in the nucleus.

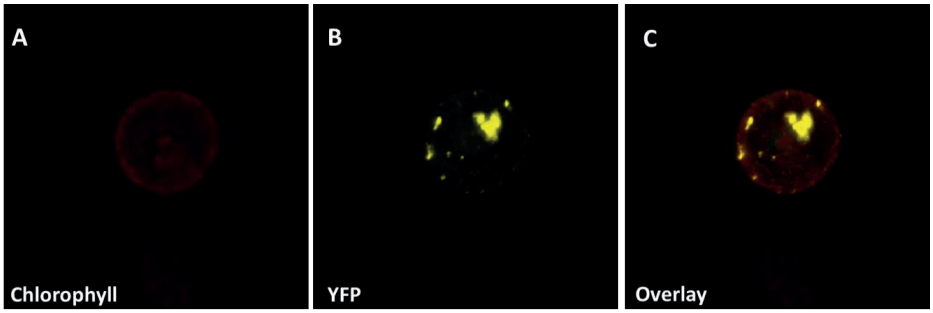



Figure 7. **POUSHP protein localization.** Leaf mesophyll protoplast transfected with a *35S:YFP-PSH* construct. Confocal images were taken eight hours after transfection.

***DISCUSSION***

Phase transition from vegetative to reproductive development in *Arabidopsis* involves the combined action of multiple flowering pathways. Interplay of different flowering stimuli leads to modulation of transcription thereby precisely controlling this developmental event. To ensure proper transcription, transcription factors are important components, often recruiting co-factors and other transcriptional regulators to fulfil their specific functions. Previous studies that isolated transcription-associated protein complexes in *Arabidopsis* have been mainly restricted to components of the general transcriptional machinery. Only recently has progress been made in the identification of protein complex components for specific transcription factors and chromatin-remodeller proteins (Kaufmann et al., 2011; Smaczniak et al., 2012b; Debernardi et al., 2014).

In the present study, we showed that the well-known SVP MADS-domain transcription factor works in a combinatorial fashion not only with other MADS-domain proteins, but in complex with a plethora of other transcriptional regulators and cofactors. A comprehensive *in vitro* yeast-2-hybrid screen by de Folter et al 2005, had shed light on the protein-protein interaction (PPI) network of MADS-domain proteins. This and other studies revealed that SVP has the potential to interact with the flowering time regulators FLC, FLM and SOC1 and the floral organ identity proteins AP1 and AP3 (de Folter et al., 2005). However, due the methods *in vitro* nature and technological constraints, these studies may not reflect indirect and transient PPIs

and do not take into account the co-expression of the interactors *in planta*. Furthermore, these screens had a matrix-based set-up, which is restricted to only a selected set of potential interactors.

Here, our *in planta* approach allowed us to unravel the protein complex composition of SVP in two distinct developmental tissues, where SVP is known to be expressed and to have specialized functions. By incorporating the prior knowledge of tissue-specific gene expression and combining it with the interaction data obtained in this study, we were able to distinguish enrichment of SVP interactions for each tissue type. For instance, in vegetative tissue, it was clearly evident that the floral repressor complex SVP-FLC is of importance, due to its previously described role in maintenance of vegetative growth. In the inflorescence, AP1 is a major interaction partner of SVP and both are expressed in the floral meristem (FM) and early stages of flower formation. It is known that SVP is active during these early stages to prevent precocious expression of the floral homeotic genes. Possibly, SVP is performing this function in concert with AP1. Similarly, we also found enrichment for SVP-SOC1, in line with the role of this complex in ensuring proper development of FMs (Gregis et al., 2006, 2008; Gregis et al., 2009; Lee and Lee, 2010).

Besides tissues-specific enrichments of interactions, it appears that some components of the SVP interactome are SVP partners in both developmental stages. The chromatin associated factors, such as CHR11, CHR14, BAF60, FGT1 and histone modifying proteins (namely HDA15), have been identified as an interactor in both developmental tissues. ATP-dependent chromatin remodelling proteins are highly conserved proteins in eukaryotes and have versatile functions with implications in transcription, replication and DNA repair (Han et al., 2015). Their ability to alter chromatin structure enables recruitment of various interacting protein complexes at unique target genes for their tight regulation. Observation of these proteins in both developmental tissues suggests a shared mechanisms through which SVP recruits these chromatin remodelers to target specific loci.

This study also led to the identification and characterization of a previously uncharacterized gene and protein, namely *POUSHP (PSH)*. Plants lacking *PSH* show a significant change in flowering response in *Arabidopsis* and surprisingly, this was opposite to the *svp* mutant (Figure 4). Flowering time analysis of the *svp-41 psh-1* double mutant, revealed that it pheno-copied the *svp-41* response, suggesting *SVP* to be genetically epistatic to *PSH*. Both genes appear to act in the same pathway. Overall, our genetic and *in vivo* protein interaction data suggests that *PSH* is a positive regulator of *SVP* with respect to flowering time control under long day conditions. Although we only quantified the flowering time response, it's worthwhile to mention that no visible pleiotropic effects were displayed in *PSH* mutants. However, expression analysis of *PSH* did suggest it to be ubiquitously present throughout development (Figure 5) and therefore *PSH* could have other developmental roles. Therefore, it would be of interest to examine and quantify other phenotypes of *PSH mutants*. In addition, it of interest

to investigate the involvement of PSH in other flowering time pathways such as the ambient temperature pathway.

BLAST searches across different EST and genomic libraries indicate that the *PSH* gene is unique to angiosperms. Phylogenetic analysis also led to the identification of a *PSH* paralog in *Arabidopsis thaliana*, which we named *POUSPH-like (PSL)* (Figure 6). Although the functional overlap of these two genes still remains to be determined, a distinct late flowering phenotype for *psh* single mutant suggests that these two genes most likely are not functionally redundant. *In silico* examination of PSH protein sequence revealed the presence of a nuclear localization signal (NLS), three coiled coil regions, D/E rich repeats and a SMC_N domain (Supplemental Figure 4). The SMC_N domain is a distinct feature found at the N terminus of structural maintenance of chromosome (SMC) proteins. In eukaryotes, the SMC proteins represent a large family of ATPases and are essential for higher order complex formation and involved in chromatin dynamics (Hirano, 2006; Uhlmann, 2016). The presence of D/E-rich repeats in PSH also gives some insights into the possible molecular function of PSH. D/E amino acids are acidic in nature. Proteins with D/E rich repeats are known to exhibit a strong negative charge distribution and have been implicated to play roles in DNA mimicry, mRNA processing and regulation of transcription (Chou and Wang, 2015). To what extent PSH is involved in any of these process remains to be determined.

Upon fusion of a fluorescent protein to the N-terminal of PSH and its transient expression in protoplasts, localization of PSH was observed in the nucleus and cytoplasm (Figure 7). The *in vivo* proteomics data presented here cannot distinguish whether the observed interaction is a direct bimolecular interaction or an indirect one. Experimental evidence from approaches such as Yeast-2-hybrid, FRET-FLIM or reciprocal IPs will allow us to address this question. Indirect interaction between SVP and PSH would necessitate the involvement of other proteins to form a multimeric SVP complex. The presence of a SMC_N domain in the PSH protein also supports the hypothesis that it is part of a higher order-complex. Overall, based on *in vivo* interaction and genetic analysis, *PSH* appears to be of importance in the flowering time process for long day photoperiod and dependent on the presence of SVP. We speculate that PSH acts as an inhibitor of SVP when it is incorporated in an SVP complex. Follow up studies on how PSH interacts with SVP in different environmental conditions will be necessary to better understand their positions in the flowering regulatory network.

To conclude, in this study we were able to identify both previously known and novel SVP interactors. Some of the novel SVP interactors correspond to previously identified uncharacterized genes. Besides deciphering the tissue-specific SVP interactome, we have characterized the role of PSH, a novel SVP interacting protein, involved in the control of flowering time.

***MATERIAL AND METHODS***

Plant Material and Growth conditions: All seeds were sown on rock wool blocks in trays and were placed in a cold room (4°C) for two nights of stratification. Following stratification, trays were moved to walk-in-growth chamber with LED light based illumination. Plants were grown in a relative humidity (RH) of 70%, at 22°C and with 94.35µmol m−2 s−1 of photosynthetically active radiation (PAR). Plants were regularly watered with HYPONEX® plant nutrient solution (1gm/L). For long day conditions, the plants were exposed to 16h light / 8h dark and for short day conditions 8h light/ 16h dark. Seeds for knock-out lines SALK_010002 (N510002) and SAIL_136_F10 (N806634) were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC).

Tissue collection, Isolation of RNA and real time PCR analysis: For the time series experiment, tissue collection was performed every two days starting from nine days after germination until all plants bolted. Five to seven plants were harvested per replicate at the end of the afternoon (zeitgeber 14h) and sampling was performed in triplicate. Using the Invitek RNA isolation kit, as per the manufactures protocol, total RNA was extracted from harvested plant material. Total RNA was subjected to DNase I digestion using the TURBO DNA-free™ kit (Ambion) and RNA integrity was checked on 1% (w/v) agarose gel after DNase treatment. First-strand cDNA was synthesized from 400ng of total RNA using iScript™ RT supermix kit (BIO-RAD), following the manufactures protocol. For *PSH* expression, a primer pair was designed within the 3'UTR region (PDS8917 Forward: 5-'GCTGCTGCTATGGAAGAAACTGC-3') and ( PDS8918 reverse: 5'-ACGGATTTGCAAACCTTGAGAGC-3'). 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.

Primers used in this Study:

| Name | Sequences 5' to 3' | Details |
|---|---|---|
| **PDS8917-Fwd** | GCTGCTGCTATGGAAGAAACTGC | At3G58110 qRTPCR |
| **PDS8918-Rev** | ACGGATTTGCAAACCTTGAGAGC | At3G58110 qRTPCR |
| **PDS8791-Fwd** | GTGCTGCAGATTCAAGAAAGG | At3G58110 (SALK) genotyping |
| **PDS8792-Rev** | GCTAGAACTCGTTGCATCCTG | At3G58110 (SALK) genotyping |
| **PDS9403-Fwd** | CAGAACTTGCAGGATCTGGAG | At3G58110 (SAIL) genotyping |
| **PDS9404-Rev** | TTCCTCGACATCTTCAATTGG | At3G58110 (SAIL) genotyping |
| **PDS2958** | GCTTCCTATTATATCTTCCCAAATTACC | LB primer for SAIL line |
| **PDS8798** | ATTTTGCCGATTTCGGAAC | LBb1.3 primer for SALK line |

LC-MS/MS based complex isolation: A previously reported transgenic line expressing SVP from its native promoter and C-terminally tagged with the GREEN FLUORESCENT PROTEIN (GFP) was used for tissues collection (Gregis et al., 2013).Immunoprecipitation was performed using a GFP antibody coupled to magnetic beads on three week old rosette material (SD conditions)

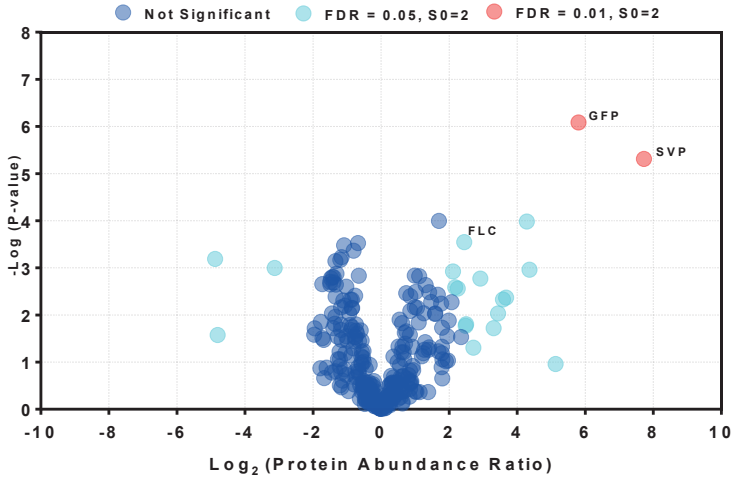
or inflorescence tissue of gSVP:GFP plants. The experiments were performed in triplicate as described previously (Smaczniak et al., 2012a; Jamge et al., 2018). WT Col-0 was used as a control. For the three week old rosette material (SD conditions), where the gSVP:GFP line was crossed into the FRIGIDA+ (FRI+) background, FRI+ was used as a control.

Plasmid Constructions: CDS sequence of PSH (At3g58110.1) was ordered from gene script in pDEST32 gateway destination vector. Using the Gateway BP reaction, PSH CDS was initially cloned into the entry vector pDONR201. Using the gateway LR reaction, the obtained entry vector was sub cloned into the compatible pARC428, from which expression is driven by the constitutive CaMV35S promoter and that contains the coding regions of the fluorophore YFP(Tonaco et al., 2006). The resulting overexpression PSH-YFP vector was maxi prepped using the ZymoPURE$^{TM}$ Plasmid maxiprep kit as per the manufacture protocol and used for protoplast transfection assays. All the above indicated constructs were tested by restriction analysis and sanger sequencing for the inserted fragment.

Protoplast assay: The “Tape-Arabidopsis sandwich” method was followed to isolate protoplast from three week old long day *Arabidopsis* leafs (Wu et al., 2009). Protoplast isolation and transfection assay were performed as described previously (Rios et al., 2017). Transfected protoplast were transferred to a six well petri dish with 1ml of WI buffer as described previously (Niu and Sheen, 2012). The petri dish were incubated at 25°C in dark and 16 hour after incubation confocal imaging was performed.

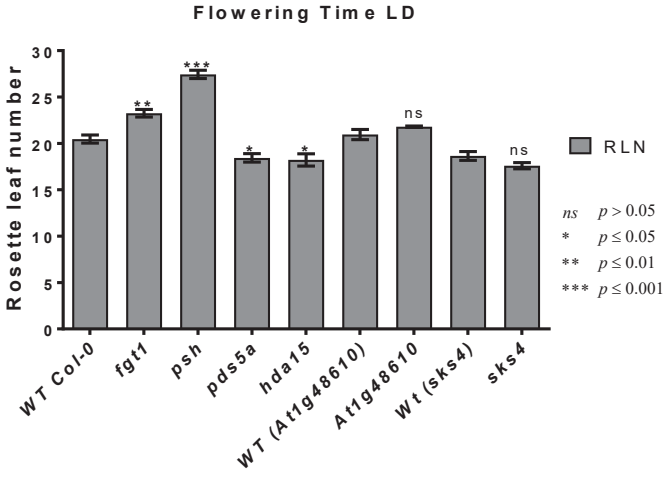
Microscopy: YFP tagged protein localization was observed through confocal laser scanning microscopy (CLSM) on Leica SPE DM5500 upright microscope using an ACS APO 40x/1.15 oil lens and with LAS AF 1.8.2 imaging software. YFP was excited with the 488-nm line of an Argon ion laser. At a bandwidth of 510–540 nm the YFP, whereas the red auto-fluorescence of chloroplast was detected at a bandwidth of 650-800nm. Upon image acquisition, optical slices were median filtered and we generated three-dimensional projection using the LAS AF 1.8.2 software package.

## Supplementary material

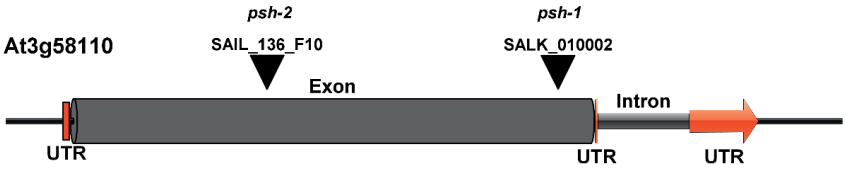



Supplemental Figure 1. ***In planta* SVP protein interaction profile in SF2 FRI+ background during vegetative phase.** Volcano plot of the IP results of SVP in vegetative (rosette) tissues. Protein abundance ratio between the samples pSVP:SVP-GFP/FRI+ and the wild type FRI+ control, plotted against the - Log(P-value). Significant protein abundance difference between sample and control, with FDR set to 0.01 or 0.05 and the s0 parameter set to 2, are shown. For a complete list of SVP interaction partners identified in the vegetative tissue of pSVP:SVP-GFP/FRI+ line please refer the Supplemental File S3.





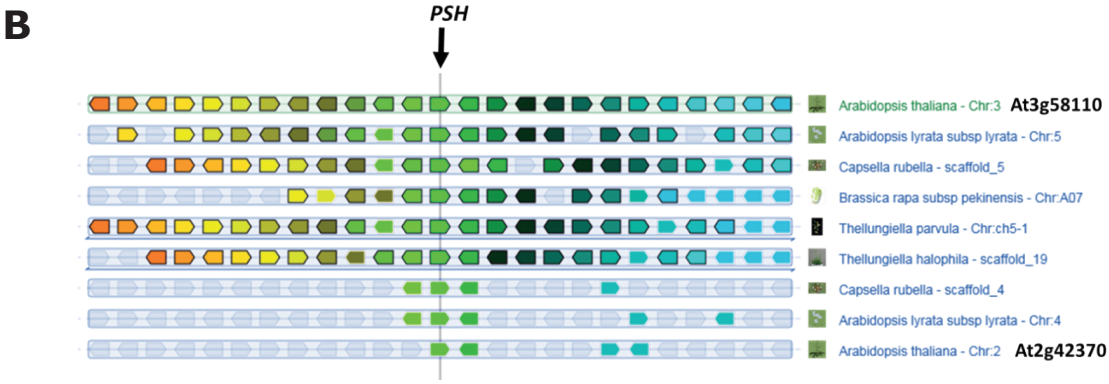

Supplemental Figure 2. **Flowering time of plant lines mutant for genes encoding putative interactors of SVP during the vegetative phase.** Flowering time screen of WT (Col-0) and homozygous single mutants of significant SVP interactors scored under long day conditions and measured as total number of rosette leaves (RLN). The experiment was performed in three biological replicates and the bar indicates the standard error. P-values were calculated using student T-test in graphpad prism.




Supplemental Figure 3. **Gene Structure of POUSHP and location of T-DNA insertions.** A schematic diagram showing the positions of T-DNA insertions in *psh* alleles. Exons, introns and the UTR sequences are labelled. Black triangles indicated the T-DNA insertion sites.

**A**

```
1   MASSPPSDPTDRDAETLSQNPSLIEKPSVVEQGSLSVENVAEKALNLESTQDEETQNLQD  60
1   MASSPPSDPTDRDAETLSQNPSLIEKPSVVEQGSLSVENVAEKALNLESTQDEETQNLQD  60
    ************************************************************

61  LEESDGRDQQLEASLEESRNEEDDMDTTQAVSSSYYRRGGGPKRKKGNQKKRKQLEKSKE  120
61  LEESDGRDQQLEASLEESRNEEDDMDTTQAVSSSYYRRGGGPKRKKGNQKKRKQLEKSKE  120
    ************************************************************

121 KLEVLLKTLKPIAFAPCKTLDFARHEKLLKTLGLWDFVHLDFDQNIREDLVANLVAYYNS  180
121 KLEVLLKTLKPIAFAPCKTLDFARHEKLLKTLGLWDFVHLDFDQNIREDLVANLVAYYNS  180
    ************************************************************

181 ERRCSYVNGARINVSRPDLARALKLPMKKDFVVTEEERELLENDESVRFIDEIVSTCVLL  240
181 ERRCSYVNGARINVSRPDLARALKLPMKKDFVVTEEERELLENDESVRFIDEIVSTCVLL  240
    ************************************************************

241 QRDDMWIMPVEIVEWTRDIKQKHLEKLDWPKLLWFMVEKELKAEPPLGDCFFASHLQLLI  300
241 QRDDMWIMPVEIVEWTRDIKQKHLEKLDWPKLLWFMVEKELKAEPPLGDCFFASHLQLLI  300
    ************************************************************

301 KSQKEDLLKEKCKADDEEDDDDDDDDVKEVDFLVKSPKEDCLEVKEEDVGAADSRKDDGA  360
301 KSQKEDLLKEKCKADDEEDDDDDDD----------------------------DDDGA  330
    ***************************                         ****

361 VDLKEDKYVEEHMLELNLGQETVSEMVSGEERGPVEGQPMDVEENKKEEDERWAWNGDSH  420
331 VDLKEDKYVEEHMLELNLGQETVSEMVSGEERGPVEGQPMDVEENKKEEDERWAWNGDSH  390
    ************************************************************

421 AGSHFLRRCNHSSAREGDEDNHIEGSMEMGEDEPIEDVEEEETEEDTEKHEGGFPFFPNG  480
391 AGSHFLRRCNHSSAREGDEDNHIEGSMEMGEDEPIEDVEEEETEEDTEKHEGGFPFFPNG  450
    ************************************************************

481 DSLQGVGQGNLMLGDASPLGYNSGLQIHGNSIGGDFLASRGEMHMAMGSGSSSLFGNGNN  540
451 DSLQGVGQGNLMLGDASPLGYNSGLQIHGNSIGGDFLASRGEMHMAMGSGSSSLFGNGNN  510
    ************************************************************

541 KREIEHENGITYHSHNPINKRLRTEEPSWDEKPPPVDMCLDQMAYWAEKARLSFAEKDRE  600
511 KREIEHENGITYHSHNPINKRLRTEEPSWDEKPPPVDMCLDQMAYWAEKARLSFAEKDRE  570
    ************************************************************

601 REQSVINQQYLMNELQSKTAMIQELERTKFEEQQRKDIMIYKLESELRMMTSVVEGYRKA  660
571 REQSVINQQYLMNELQSKTAMIQELERTKFEEQQRKDIMIYKLESELRMMTSVVEGYRKA  630
    ************************************************************

661 LKITQKASREHRKRCPLRDDKQVYMDVKGSGGLVLSTTEIEKLRLKQEEEDRMQRVLAKR  720
631 LKITQKASREHRKRCPLRDDKQVYMDVKGSGGLVLSTTEIEKLRLKQEEEDRMQRVLAKR  690
    ************************************************************

721 QIDDFEHNWLNKFEEHMEAVELLNERLIENEDEVKILRETLSESKNIETSEVAAAMEETA  780
691 QIDDFEHNWLNKFEEHMEAVELLNERLIENEDEVKILRETLSESKNIETSEVAAAMEETA  750
    ************************************************************

781 TEET  784
751 TEET  754
    ****
```




Supplemental Figure 4. **Protein sequence and evolutionary analysis of *POUSPH (PSH)* At3g58110**

A) Sequence similarity of the two splice variants of POUSPH (At3g58110.1 (784 aa) and At3g58110.2 (754aa) ) is displayed in grey colour. The yellow colour highlights the coiled coil regions. The pink colour denotes the acidic D/E rich regions of the POUSHP protein. B) Synteny analysis of POUSHP (PSH). Coloured arrows represent the genes, orthologues have the same colour. *POUSPH-like (PSL),* At2g42370 is indicated at the bottom, showing limited synteny with *PSH*.



## **Description of additional files** (Will be provided upon request)

**Supplementary file S1|**Complete IP LC-MS/MS analysis for SVP in vegetative and reproductive tissue

**Supplementary file S2|**Complete IP LC-MS/MS analysis for SVP comparing vegetative and reproductive tissue

**Supplementary file S3 |** Complete IP LC-MS/MS analysis for FLC in vegetative tissue

***ACKNOWLEDGEMENTS***

We like to acknowledged Tatyana Radoeva, Jan Willem Borst and Prof. Jan Sheen for their valuable suggestions during the set-up of the protoplast method in our laboratory. We thank Suzanne de Bruijn for her assistance in the phylogenetic analysis, Sjef Boeren for his technological support for the LC-MS/MS analysis and we sincerely thank Tom Brabbs for critical reading of the manuscript.

***REFERENCES***

**Balanza, V., Martinez-Fernandez, I., and Ferrandiz, 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.

**Bouche, F., Lobet, G., Tocquin, P., and Perilleux, C.** (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. Nucleic acids research **44,** D1167-1171.

**Brzezinka, K., Altmann, S., Czesnick, H., Nicolas, P., Gorka, M., Benke, E., Kabelitz, T., Jahne, F., Graf, A., Kappel, C., and Baurle, I.** (2016). Arabidopsis FORGETTER1 mediates stress-induced chromatin memory through nucleosome remodeling. eLife **5**.

**Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D.** (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. The Plant journal : for cell and molecular biology **89,** 789-804.

**Chou, C.C., and Wang, A.H.** (2015). Structural D/E-rich repeats play multiple roles especially in gene regulation through DNA/RNA mimicry. Molecular bioSystems **11,** 2144-2151.

**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.

**Debernardi, J.M., Mecchia, M.A., Vercruyssen, L., Smaczniak, C., Kaufmann, K., Inze, D., Rodriguez, R.E., and Palatnik, J.F.** (2014). Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity. The Plant journal : for cell and molecular biology **79,** 413-426.

**Fornara, F., de Montaigu, A., and Coupland, G.** (2010). SnapShot: Control of flowering in Arabidopsis. Cell **141,** 550, 550 e551-552.

**Gregis, V., Sessa, A., Colombo, L., and Kater, M.M.** (2006). AGL24, SHORT VEGETATIVE PHASE, and APETALA1 redundantly control AGAMOUS during early stages of flower development in Arabidopsis. The Plant cell **18,** 1373-1382.

**Gregis, V., Sessa, A., Colombo, L., and Kater, M.M.** (2008). AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis. The Plant journal : for cell and molecular biology **56,** 891-902.

**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 : for cell and molecular biology **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.

**Han, S.K., Wu, M.F., Cui, S.J., and Wagner, D.** (2015). Roles and activities of chromatin remodeling ATPases in plants. Plant Journal **83,** 62-77.

**Hartmann, U., Hohmann, 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 : for cell and molecular biology **21,** 351-360.

**Hirano, T.** (2006). At the heart of the chromosome: SMC proteins in action. Nature reviews. Molecular cell biology **7,** 311-322.

**Hu, J.Y., Zhou, Y., He, F., Dong, X., Liu, L.Y., Coupland, G., Turck, F., and de Meaux, J.** (2014). miR824-Regulated AGAMOUS-LIKE16 Contributes to Flowering Time Repression in Arabidopsis. The Plant cell **26,** 2024-2037.

**Hwan Lee, J., Sook Chung, K., Kim, S.K., and Ahn, J.H.** (2014). Post-translational regulation of SHORT VEGETATIVE PHASE as a major mechanism for thermoregulation of flowering. Plant signaling & behavior **9,** e28193.

**Jamge, S., Angenent, G.C., and Bemer, M.** (2018). Identification of In Planta Protein–Protein Interactions Using IP-MS. In Plant Chromatin Dynamics: Methods and Protocols, M. Bemer and C. Baroux, eds (New York, NY: Springer New York), pp. 315-329.

**Jerzmanowski, A.** (2007). SWI/SNF chromatin remodeling and linker histones in plants. Bba-Gene Struct Expr **1769,** 330-345.

**Kaufmann, K., Smaczniak, C., de Vries, S., Angenent, G.C., and Karlova, R.** (2011). Proteomics insights into plant signaling and development. Proteomics **11,** 744-755.

**Lee, J., and Lee, I.** (2010). Regulation and function of SOC1, a flowering pathway integrator. Journal of experimental botany **61,** 2247-2254.

**Lee, J.H., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.S., and Ahn, J.H.** (2007). Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes & development **21,** 397-402.

**Lee, J.H., Ryu, H.S., Chung, K.S., Pose, D., Kim, S., Schmid, M., and Ahn, J.H.** (2013). Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. Science **342,** 628-632.

**Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E., and Yu, H.** (2008). A repressor complex governs the integration of flowering signals in Arabidopsis. Developmental cell **15,** 110-120.

**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, C., Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T., and Yu, H.** (2007). Specification of Arabidopsis floral meristem identity by repression of flowering time genes. Development **134,** 1901-1910.

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

**Masiero, S., Li, M.A., Will, I., Hartmann, U., Saedler, H., Huijser, P., Schwarz-Sommer, Z., and Sommer, H.** (2004). INCOMPOSITA: a MADS-box gene controlling prophyll development and floral meristem identity in Antirrhinum. Development **131,** 5981-5990.

**Mateos, J.L., Madrigal, P., Tsuda, K., Rawat, V., Richter, R., Romera-Branchat, M., Fornara, F., Schneeberger, K., Krajewski, P., and Coupland, G.** (2015). Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in Arabidopsis. Genome biology **16,** 31.

**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.

**Niu, Y., and Sheen, J.** (2012). Transient expression assays for quantifying signaling output. Methods in molecular biology **876,** 195-206.

**Pelaz, S., Gustafson-Brown, C., Kohalmi, S.E., Crosby, W.L., and Yanofsky, M.F.** (2001). APETALA1 and SEPALLATA3 interact to promote flower development. The Plant journal : for cell and molecular biology **26,** 385-394.

**Pose, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G.C., Immink, R.G., and Schmid, M.** (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. Nature **503,** 414-417.

**Rios, A.F., Radoeva, T., De Rybel, B., Weijers, D., and Borst, J.W.** (2017). FRET-FLIM for Visualizing and Quantifying Protein Interactions in Live Plant Cells. Methods in molecular biology **1497,** 135-146.

**Severing, E.I., van Dijk, A.D., Morabito, G., Busscher-Lange, J., Immink, R.G., and van Ham, R.C.** (2012). Predicting the impact of alternative splicing on plant MADS domain protein function. PloS one **7,** e30524.

**Shen, L., Kang, Y.G., Liu, L., and Yu, H.** (2011). The J-domain protein J3 mediates the integration of flowering signals in Arabidopsis. The Plant cell **23,** 499-514.

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

**Smaczniak, C., Immink, R.G., Muino, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Parcy, F., Xu, L., Carles, C.C., Angenent, G.C., and Kaufmann, K.** (2012b). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proceedings of the National Academy of Sciences of the United States of America **109,** 1560-1565.

**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.

**Tonaco, I.A., Borst, J.W., de Vries, S.C., Angenent, G.C., and Immink, R.G.** (2006). In vivo imaging of MADS-box transcription factor interactions. Journal of experimental botany **57,** 33-42.

**Uhlmann, F.** (2016). SMC complexes: from DNA to chromosomes. Nature reviews. Molecular cell biology **17,** 399-412.

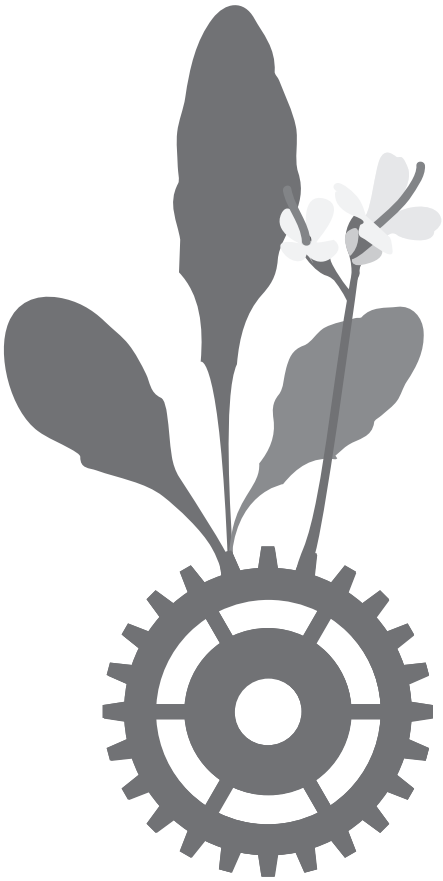
**Verbsky, M.L., and Richards, E.J.** (2001). Chromatin remodeling in plants. Current opinion in plant biology **4,** 494-500.

**Wang, X.X., Wu, F.M., Xie, Q.G., Wang, H.M., Wang, Y., Yue, Y.L., Gahura, O., Ma, S.S., Liu, L., Cao, Y., Jiao, Y.L., Puta, F., McClung, C.R., Xu, X.D., and Ma, L.G.** (2012). SKIP Is a Component of the Spliceosome Linking Alternative Splicing and the Circadian Clock in Arabidopsis. The Plant cell **24,** 3278-3295.

**Wu, F.H., Shen, S.C., Lee, L.Y., Lee, S.H., Chan, M.T., and Lin, C.S.** (2009). Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. Plant methods **5,** 16.

# CHAPTER 6

## 3C in Maize and Arabidopsis



Blaise Weber[1]
Suraj B. Jamge[1]
Maike E. Stam

Plant Chromatin Dynamic Methods and Protocol (MIMB, volume 1675)
[1]: These authors contributed equally to this work

***ABSTRACT***

Chromosome Conformation Capture (3C) allows studying the relative frequency of interaction of one chromosomal fragment with another. The technique is especially suited for unravelling the 3D organization of specific loci when focusing on aspects such as enhancer-promoter interactions or other topological conformations of the genome. 3C has been extensively used in animal systems, among others providing insight into gene regulation by distant *cis*-regulatory elements. In recent years, the 3C technique has been applied in plant research. However, the complexity of plant tissues prevents direct application of existing protocols from animals. Here we describe an adapted protocol suitable for plant tissues, especially *Arabidopsis thaliana* and *Zea mays*.

Keywords (5-10)

Chromosome Conformation Capture, 3C, protocol, plant, *Arabidopsis thaliana*, *Zea mays*

## 1. *INTRODUCTION*

Since the first microscopic observation of nuclei, it has become clear that chromatin is not randomly organized [1]. A specific 3D architecture of the genome is established in each and every cell's nucleus to ensure proper regulation of gene expression [2, 3]. Such architecture includes large-scale chromatin domains as well as specific enhancer-promoter interactions that together shape the cell's transcriptome and its fate [4, 5]. Chromosome Conformation Capture (3C) allows to study chromatin organization and helps to understand how the spatiotemporal organization of chromatin influences gene expression [6]. 3C technology was first implemented in yeast [7] and rapidly adapted to other organisms including mice, human and Drosophila [8–10]. The 3C technique provides the attractive opportunity to study chromosomal interactions at a resolution that was previously difficult to achieve with cytological methods [6]. The method relies on the proximity ligation concept (Figure 1). In this method, chromatin from fixed nuclei is first subjected to digestion with a methylation insensitive restriction enzyme, followed by ligation in conditions favoring intramolecular ligations. Hereby, regions that are in close contact in 3D have a higher chance of being ligated together. Subsequently, interaction frequencies of specific ligation events are quantified, providing insight into the 3D organization of chromatin at the genomic locus of interest. Frequencies of interaction are measured using quantitative PCR (qPCR) in combination with primers specific for each interaction one desires to examine. The use of primers recognizing specific fragments makes 3C an hypothesis-driven approach. Since the first publication in 2002, variants of the 3C method have been developed (e.g. 4C, 5C and Hi-C) that allow a more systematic analysis of chromosomal interactions [11–14]. These methods identify many more interactions at the same time (one to all, many to many, all to all) by including deep sequencing techniques. Hence they are associated with higher costs, and given the complexity and amount of data generated, also time consuming data analysis pipelines. Therefore, 3C remains a method of choice when special focus is given to one specific locus. It allows faster results and often offers better resolution.

The specific nature of plant cells hampers the direct application of published 3C protocols from other species. Therefore, 3C on plant tissues requires plant specific steps. In literature, 3C protocols have been described for Arabidopsis and Maize [15, 16]. In this article, we provide a step-by-step bench protocol, starting from the design and setup of a 3C experiment, up to the analysis and interpretation of the 3C data. In addition, we provide critical notes on different aspects that need to be adapted when applying this method to other plants or tissues of interest.

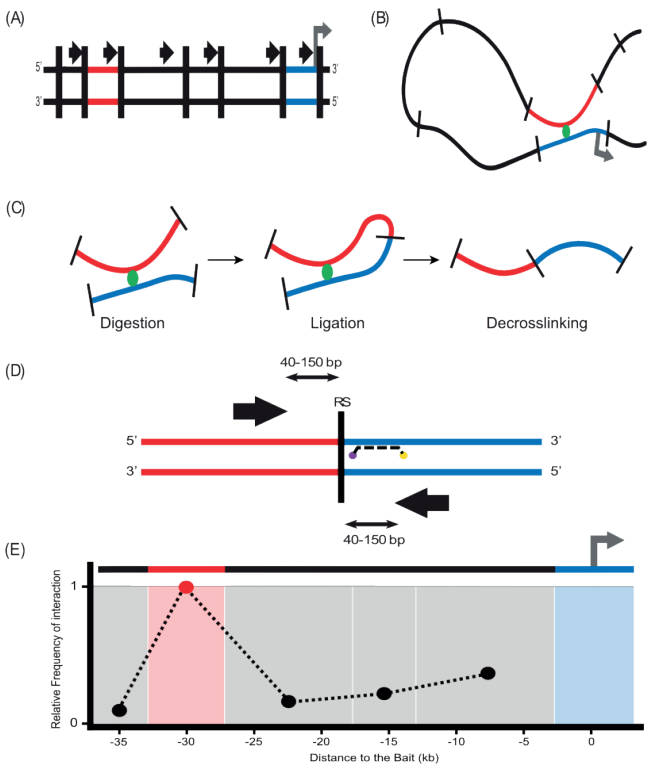



**Fig. 1** Graphical overview of the 3C technique**.** (A) Schematic representation of a potential locus of interest. Primers (black arrows) are designed on one and the same strand for all fragments monitored. The bait is indicated in blue. (B) Schematic representation of an hypothetical chromosome conformation at the potential locus of interest. (C) Fixed chromatin is digested, ligated in a large volume and then de-crosslinked. The products of these processes are shown. (D) Schematic representation of a 3C template. Black vertical bar indicates the ligated restriction site (RS). Primers (black arrows) anneal on each side of the RS and now form a primer pair. A TaqMan probe can be used for more specific quantification by qPCR (*see* **Note 15**) and should be designed on the bait fragment, on the strand complementary to the strand on which the bait primer is designed. (E) Graphical representation of a 3C plot. The relative frequency of interaction of each fragment with the bait is plotted on the y-axis (black and red circles). The distances to the middle of the bait fragment are reported on the x-axis. Vertical boxes are used to indicate the different fragments monitored.

## *2. MATERIALS*

Prepare all solutions using autoclaved milliQ water and analytical grade reagents. Sterilization by autoclaving is performed at 121 °C for 10 min unless indicated otherwise. Diligently follow all waste disposal regulations when disposing of waste materials.

**2.1 Tissue Fixation and nuclei isolation**

1. 1 X Phosphate Buffered Saline (PBS): add 800 mL of milliQ water into a graduated cylinder together with a magnetic stirrer. Weigh 8 g of NaCl, 0.2 g of KCl, 1.44 g of $Na_2HPO_4$, 0.24 g of $KH_2PO_4$ and add it to the measuring column. Adjust the pH to 7.4 with HCl and add milliQ water up to 1 L. Autoclave the solution and store at room temperature (RT).
2. 4% Paraformaldehyde (PFA) in PBS: Prepare this solution in the fume hood. Pour 100 mL of autoclaved 1 X PBS in a 250 mL glass bottle. Weigh 4 g of powdered paraformaldehyde and transfer it to the PBS-filled bottle. Adjust the pH to 9 with KOH, close the bottle and transfer it to a 65 °C water bath. Shake the bottle from time to time until the PFA is completely dissolved. Transfer it back to the fume hood and allow the bottle to cool down before adjusting the pH back to 7-7.5 with HCl. Prepare aliquots of 10 mL to be stored at -80 °C.
3. 2 M Glycine: Add 80 mL of milliQ water to a glass graduated cylinder or beaker, together with a magnetic stirrer. Weigh 15 g of Glycine and add it into the graduated cylinder or beaker. Dissolution can be enhanced by raising the temperature. After complete homogenization, adjust the volume to 100 mL with milliQ water and transfer the solution to a 100 mL glass bottle. Autoclave and store at 4 °C.
4. 20 % Triton X-100: pipette twice 1 mL of Triton X-100 with a cut pipette tip and add to a 15 mL tube containing 8 mL of autoclaved milliQ water. Allow all Triton to get out of the tip by pipetting up and down. Avoid foaming as much as possible. Shield the tube from light with opaque tape or aluminium foil. Place the tube in a rotating wheel at RT overnight to allow complete homogenization. 20 % Triton X-100 can be stored shielded from light at RT for a month.
5. 100 mM Phenylmethylsulfonyl fluoride (PMSF): weigh 174 mg of PMSF and dissolve it in 10 mL of isopropanol. Prepare aliquots of 100 µL and store at -20 °C.
6. Complete Protease inhibitor (Roche): dissolve 1 complete protease inhibitor tablet in 2 mL of autoclaved milliQ water by vortexing vigorously. Dissolved tablets can be stored for 1-2 weeks at 4 °C, or up to 6 weeks at -20 °C.
7. Nuclei Extraction Buffer (100 mL, prepare fresh): 2 mL 1M Hepes pH 8, 25 mL 1M Sucrose, 0.1 mL 1M MgCl2, 0.5 mL 1M KCl, 46 mL 87% Glycerol, 1.25 mL 20% Triton X-100, 26 mL autoclaved milliQ, 100 µl 100 mM PMSF, 100 µl Beta-Mercaptoethanol.
8. 50 mL tubes.
9. Miracloth.
10. Sefar Nitek nylon filter 50 µm pore size.
11. Morter and pestle.
12. Liquid Nitrogen.
13. Cooling centrifuge for 50 mL tubes.
14. Cooling microcentrifuge.

**2.2 Digestion and ligation of 3C and control samples**

1. BAC-clone or other large plasmid that contains the region of interest (see step 3.2.1)
2. Suitable restriction enzyme (see Table 1 and section 3.1) with 10x restriction buffer.
3. Phenol:Chloroform:Isoamyl alcohol (25:24:1 v/v): under the fume hood, pipette 25 mL of Phenol equilibrated with 10 mM Tris-HCl, 1mM EDTA, pH 8 and transfer it to a 50 mL glass bottle. Add 24 mL of Chloroform and 1 mL of Isoamyl alcohol. Close the lid and mix. Allow the phases to separate, protect the bottle from light and store at 4 °C.
4. Chloroform:Isoamyl alcohol (24:1 v/v): under the fume hood, pipette 24 mL of Chloroform and transfer it to a 50 mL glass bottle. Add 1 mL of Isoamyl alcohol. Close the lid and mix. Protect the bottle from light and store at 4 °C.
5. 2 M Sodium Acetate (NaOAc) pH 5.6: Add 80 mL of milliQ water to a graduated cylinder. Weigh 16.4 g of Anhydrous Sodium Acetate and dissolve it in the 80 mL of milliQ water. Adjust pH to 5.6 with HCl and bring volume up to 100 mL with milliQ water. Autoclave and store at RT.
6. Glycogen 20 mg/mL
7. 96% and 70% Ethanol
8. 10 mM Tris pH 7.5
9. 1 M Tris-HCl pH 7.8: add 800 mL of milliQ water into a graduated cylinder, together with a magnetic stirrer. Weight 157.6 g of Tris and gradually add it to the column while stirring. Set the pH to 7.8 with HCl and adjust the volume to 1 L with milliQ water. Autoclave solution and store at RT.
10. 1 M DL-Dithiothreitol (DTT): dissolve 1.53 g of DTT powder into 10 mL of autoclaved milliQ water. Prepare aliquots of 100 µL and store at -20 °C.
11. 10x Ligase buffer for 3C samples (1 mL, prepare freshly): 300 µL 1M Tris-HCl pH7.8, 100 µL1M MgCl2, 100 µL 1M DTT, 6 mg ATP (final concentration 10 mM), 500 µL autoclaved milliQ water.
12. Highly concentrated T4 DNA ligase
13. 20 % Sodium dodecyl sulfate (SDS): open SDS container under the fume hood and weigh 10 g of SDS. Still under the fume hood, transfer SDS to a 50 mL glass bottle filled with 35 mL of autoclaved milliQ water. Ensure complete homogenization and bring volume to 50 mL with autoclaved milliQ water. Store bottle at RT.Agarose
14. Gel running device and UV transilluminator.
15. Water bath
16. Heat block
17. Cooling microcentrifuge
18. Cooling centrifuge for 50 mL tubes.
19. 50 mL tubes

**2.3 De-crosslinking, DNA purification and qPCR analysis**

1. 10 mg/mL Proteinase K: weight 100 mg of proteinase K and dissolve in 10 mL of autoclaved milliQ water. Prepare aliquots of 100 µL and store at -20 °C.

2. 10 mg/mL RNase A: for a final volume of 1 mL, add 900 µL of 10mM NaOAc pH5.6to a 1.5 mL eppendorf. Add 10 mg of RNase A. Mix until complete dissolved and place the tube into boiling water for 15 min. Allow solution to cool down at RT and add 100 µL of Tris-HCl pH 7.4. Store at -20°C.
3. Phenol:Chloroform:Isoamyl alcohol (25:24:1 v/v), see step 4 of Subheading 2.2.
4. Chloroform:Isoamyl alcohol (24:1 v/v), see step 5 of Subheading 2.2
5. Agarose.
6. Gel running device and UV transilluminator.
7. 10 µM primers for the ligation products to be tested (see step 3 of Subheading 3.1).
8. 10 mM dNTPs.
9. Quantitative PCR machine.
10. Reagents for real-time PCR analysis (e.g. SYBR Green Mastermix)

## *3. METHODS*

### 3.1 Study design

1. **Define your region of interest:** 3C can be performed at any specific locus or gene of interest for which the DNA sequence is known. In general, one should select a restriction enzyme (RE) that will generate a restriction pattern compatible with intramolecular ligation (fragments smaller than 300 bp are more difficult to ligate) and allows the verification of hypothesized interactions (regions of interest should be located in fragments >300 bp). Define a “bait” or “viewpoint” (see Figure 1) by selecting the fragment for which you will quantify the frequency of interactions with other fragments.
2. **Selection of the Restriction Enzyme:** REs used in the 3C procedure need to efficiently digest crosslinked chromatin, which is challenging. Different REs or combinations of REs that allow digestion of fixed chromatin have been reported in published 3C protocols and studies. Using such REs (see Table 1) is a safe choice. However, if the restriction patterns generated by these enzymes are not compatible with one’s hypothesis at the locus of interest, newly selected enzymes should fulfill a number of requirements. First, the selected RE should be methylation insensitive, since methylation sensitive enzymes may result in partial digestion and thus introduce a bias. Second, the selected RE should ideally display optimal efficiency for digesting fixed chromatin at 37 °C, and preferably maintain its activity over a long period of time (e.g. overnight). In case of a short survival time, aliquots of restriction enzymes can be added sequentially during the digestion time. Thirdly, high ligation efficiency is crucial for a successful 3C experiment. Thus, favor enzymes generating sticky ends (the larger the overhang, the better), and ensure that the re-ligation efficiency is high without the need of adding any macromolecular crowding agent such as polyethylene glycol (PEG).

Once the digestion is complete, to stop digestion the RE needs to be inactivated by elevated temperatures. In addition, for some enzymes detergents need to be added for inactivation (see Table 1). Note that the addition of SDS is associated with negative effects on ligation efficiency.

| Enzyme | Heat Inactivation[1] | Reference |
|---|---|---|
| *Hin*dIII | Yes | [17, 18] |
| *Eco*RI | Yes | [7, 18] |
| *Bgl*II | No | [18–20] |
| *Bam*HI | No | [18, 20, 21] |
| *Dpn*II | Yes | [22, 23] |
| *Mfe*I | No | [24] |
| *Nla*III | Yes | [25] |
| *Xho*I | Yes | [18] |

**Table 1**: List of REs regularly used in 3C experiments. [1]This column indicates if an enzyme can be heat inactivated without the addition of SDS.

3. **Type of tissue and number of nuclei:** chromosome conformations are mostly cell type-dependent [8, 19, 26]. Therefore, ideally, for 3C one uses fresh (*see* **Note 1**), homogeneous and synchronized cell populations. Plant tissues have an heterogeneous cell type composition, and their cells are not synchronized. The heterogeneity of plant tissues does not preclude the use of 3C technology, however, one should keep in mind that the results obtained will reflect an average of the chromosomal interactions occurring in different and unsynchronized cell types. Combining the 3C procedure with methods that allow the isolation of specific cell types (e.g. FACS or INTACT) could, although technically very challenging, allow studying cell type-specific interactions [27–29].

   The type of plant tissues used for 3C analysis is dependent on the research question. Theoretically, most tissues are compatible with 3C analysis, however, tissues that are highly lignified or have a high starch content pose difficulties at grinding and/or downstream process. In addition, the amount of starting material needs to be sufficient to allow isolating 1 to 10 million nuclei (to assess the number of nuclei *see* **Note 2**). Tissues that are difficult to harvest in bulk, such as meristems, are therefore challenging.

4. **Primer Design:** Specific primers need to be designed for all DNA fragments one wishes to study. 3C primers need to be orientated uni-directionally, meaning that all primers are designed complementary to the same strand. This way, only ligation events between two different fragments will result in amplicon formation (see Fig. 1). Primers are usually designed 40-150 bp away from the restriction sites. Furthermore, they are preferably 18-27 bp long with a GC content of about 50 %, a Tm between 57-63 °C and no more than 2 °C difference in Tm between them. Primers need to be very specific, which should be determined using BLAST (High Throughput Genome Sequence database). Select primers only if one perfect match is found; homology to sequences

elsewhere in the genome should be less than 75 % of the primer length and exclude the 3' end of the primer. Primers should be tested experimentally on control template (Random Ligation Library; see section 3.2.1) and genomic DNA to ensure that only the correct amplicon size is amplified when using the Random Ligation Library.

### 3.2 Controls in the 3C procedure

The correct interpretation of 3C results demands the use of a number of controls. For instance, controls are used to ensure an optimal digestion efficiency (a non-digested 3C sample and digested gDNA), correct for primer efficiency (Random Ligation Library), and also account for technical and biological variation between 3C libraries (endogenous locus).

#### 3.2.1 Random Ligation Library (RL-Library)

To determine the relative frequency of interaction of one fragment with another, one needs to ensure that no quantification bias arises due to primer pairs with different amplification efficiencies. Therefore, to normalize for the primer efficiency, a random ligation library (RL-Library) is prepared that consists of all possible ligation products that need to be analyzed by qPCR (see 3.7). The RL-Library can be used as a template to first test primer efficiency and specificity. An RL-Library can be generated in different ways. One option is by digesting a bacterial artificial chromosome (BAC) or other large plasmid that contains the locus of interest, followed by re-ligation. The digestion of the BAC is performed with the same RE selected to generate the 3C library but in a smaller volume compared to the ligation step for the 3C library. This allows all fragments to randomly ligate with one another. Alternative to using a BAC-derived RL-Library in the qPCR experiments, one can PCR amplify all potential ligation products from a 3C library, and mix those in equimolar amounts. Finally, note that a different RL-Library has to be prepared for every locus one examines (locus of interest and the endogenous locus).

1. Prepare a 1.5 mL eppendorf tube with 10 µg of template (e.g. BAC) in a final volume of 100 µL with a two-fold excess of restriction enzyme (e.g. 1µl of 20,000 units/mL *Hin*dIII) and its recommended restriction buffer at a final concentration of 1x.
2. Incubate for 5 hours at 37 °C.
3. Check the digestion efficiency by running 10 µL of the digestion mixture on a 0.8 % agarose gel. Satellite bands should be visible.
4. Add 210 µL of sterile milliQ to bring the volume up to 300 µL.
5. Add 1 Volume of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) to the digestion mixture. Mix thoroughly.
6. Centrifuge at full speed for 10 min at RT.
7. Transfer the top aqueous phase to a new 1.5 mL eppendorf tube and add 1 volume of Chloroform:Isoamyl Alcohol (24:1). Mix thoroughly.
8. Repeat the centrifugation step at full speed for 10 min at RT.

9. Transfer the top (aqueous) phase to a new 1.5 mL eppendorf tube and precipitate the DNA with 1/10 volume of 2 M NaOAc (pH 5.6), 2 volumes of 96 % Ethanol and 10 µL of Glycogen (20 mg/ml).
10. Store the tube at -80 °C for 2 hours to overnight.
11. Centrifuge at full speed for 30 min at 4 °C.
12. Gently pour off the supernatant and wash the pellet with 1 mL of 70 % ethanol.
13. Centrifuge at full speed for 10 min at 4 °C.
14. Gently pour off the supernatant and use a pipette tip to remove the remaining droplets.
15. Let the pellet air dry for 2-5 min and resuspend the DNA in 22 µL of 10 mM Tris pH 7.5.
16. Transfer 20 µL of the digested DNA to a new eppendorf tube. Add 5 µL of fresh 10x ligation buffer, sterile milliQ water and 20 Units of T4 DNA ligase up to a final volume of 50 µL (*see* **Note 3**). Store the remaining 2 µL of digested BAC DNA at -20 °C.
17. Incubate the ligation reaction overnight at 16 °C.
18. Bring the volume to 300 µL by addition of 250 µL of sterile milliQ water.
19. Add 1 Volume of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) to the ligation mixture. Mix thoroughly.
20. Centrifuge at full speed for 10 min at RT.
21. Transfer the top phase to a new 1.5 mL eppendorf tube and add 1 volume of Chloroform:Isoamyl Alcohol (24:1). Mix thoroughly.
22. Centrifuge at full speed for 10 min at RT.
23. Transfer the top phase to a new 1.5 mL eppendorf tube and precipitate the DNA with 1/10 volume of 2 M NaOAc (pH 5.6), 2 volumes of 96 % Ethanol and 10 µL of Glycogen (20 mg/ml).
24. Store the tube at -80 °C for 2 hours to overnight.
25. Centrifuge at full speed for 30 min at 4 °C.
26. Gently pour off the supernatant and wash the pellet with 1 mL of 70 % ethanol.
27. Centrifuge at full speed for 10 min at 4 °C.
28. Gently pour off the supernatant and use a pipette tip to remove the remaining droplets.
29. Let the pellet air dry for 2-5 min and resuspend the DNA in 22 µL of 10 mM Tris pH 7.5.
30. Prepare a 0.8 % Agarose gel and mix both the 2 µL of the RL-Library, and the 2 µL of digested 3C sample (3.6.1 step 16) with 7 µL of sterile milliQ and 1 µL of 10x loading buffer.
31. Store the RL-Library at -20 °C.
32. Run the digested and ligated samples on the prepared 0.8 % agarose gel to check the ligation efficiency.
33. Prepare 1/10, 1/100, 1/1000 and 1/10,000 dilutions of the RL-Library in sterile milliQ, and add non-digested gDNA to each dilution (final concentration of 50 ng gDNA/µL). Addition of non-digested gDNA mimics the PCR conditions with the 3C library as a template. Measure the Ct values for the different primer pairs of interest using qPCR.

For each 20 µL PCR reaction prepare a qPCR mix according to your own setup. Use 1 µL of template (concentration depends on the dilution) and 1 µL of each primer (10 µM). When analyzing the 3C library for the first time with qPCR, use the different dilution series of your RL-Library complemented with gDNA. Subsequently, for normalization use the dilution showing Ct values in the range of the Ct values obtained for the 3C library. In later qPCR analyses, one can use only the relevant dilution(s) of the RL-Library.

### 3.2.2 Endogenous Control

Another crucial 3C control, an endogenous control locus, accounts for technical and biological variation between samples. Technical variation hereby refers to differences in quantity and quality of the sample, biological variation refers to differences in interaction frequencies at the locus of interest between different tissues. To this end, an endogenous locus is chosen that shows similar RNA expression levels across the different tissues examined by 3C. A similar RNA expression level indicates a similar chromatin conformation. Typically, genic loci referred to as housekeeping genes are known to be similarly expressed and can be assumed to show similar chromatin conformation in different tissues. The frequency of interactions at such loci can therefore be used to normalize the data between biological samples.

1. Identify a proper endogenous control locus (e.g. *SAM* (GRMZM2G154397) in *Z. mays* or *TIP41* (AT4G34270.1) in *A. thaliana*). Do this by checking if RNA transcript levels are similar in the tissues of interest.
2. Design primers complementary to multiple restriction fragments at the endogenous control locus and its flanking sequences.
3. Using qPCR and your 3C library as a template, measure the frequencies of interaction between the selected bait fragment and the other fragments of the endogenous locus. Usually the bait fragment consists of the fragment that contains the TSS of the selected endogenous locus. Select a primer pair displaying a frequency of interaction comparable to the mean of your frequency of interactions at the studied locus and take this primer pair along at subsequent 3C experiments for normalization.

### 3.2.3 Positive Digestion Control

The positive digestion control is used as a reference to determine the efficiency of digestion. The positive digestion control consists of fully digested gDNA. In this sample, the pattern of digestion (size range of the smear and satellite bands) should be clearly visible.

1. Prepare a 1.5 mL eppendorf tube with 10 µg of gDNA template in a final volume of 100 µL with a two-fold excess of restriction enzyme and its recommended buffer at a final concentration of 1x.
2. Incubate overnight at 37 °C.

3. Check the digestion efficiency by running 10 µL of the digestion mixture on a 0.8 % agarose gel. Satellite bands should be visible. If the pattern of digestion is not clearly visible, extend the incubation time.
4. Add 200 µL of sterile milliQ to bring the volume up to 300 µL.
5. Add 1 Volume of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) to the digestion mixture. Mix thoroughly.
6. Centrifuge at full speed for 10 min at RT.
7. Transfer the top aqueous phase to a new 1.5 mL eppendorf tube and add 1 volume of Chloroform:Isoamyl Alcohol (24:1). Mix thoroughly.
8. Repeat the centrifugation step at full speed for 10 min at RT.
9. Transfer the top (aqueous) phase to a new 1.5 mL eppendorf tube and precipitate the DNA with 1/10 volume of 2 M NaOAc (pH 5.6), 2 volumes of 96 % Ethanol and 10 µL of Glycogen (20 mg/ml).
10. Store the tube at -80 °C for 2 hours to overnight.
11. Centrifuge at full speed for 30 min at 4 °C.
12. Gently pour off the supernatant and wash the pellet with 1 mL of 70 % ethanol.
13. Centrifuge at full speed for 10 min at 4 °C.
14. Gently pour off the supernatant and use a pipette tip to remove the remaining droplets.
15. Let the pellet air dry for 2-5 min and resuspend the DNA in 50 µL of 10 mM Tris pH 7.5.
16. Store the tube at -20 °C.
17. Load 5 µL of the positive digestion control on a 0.8% agarose gel when checking the digestion efficiency of the 3C library (see 3.6.16).

**3.3 Plant tissue fixation and nuclei isolation**

The following procedure describes the handling of one biological sample. Multiple samples can be handled at the same time. In our hands, working with more than four samples at once is cumbersome and might result in suboptimal 3C library quality.

1. Prepare Nuclei Extraction Buffer (NEB) and place on ice. Also, pre-cool the centrifuge for 50 ml tubes (swing-out) and eppendorf centrifuge to 4 °C.
2. Fill a 50 mL centrifuge tube (preferably PPCO tubes – Poly Propylene COpolymer) and fill it with 10 mL of 1x PBS. Place the tube on ice.
3. Harvest tissue of interest (1 to 3 g; *see* **Note 2**) and place it in a large petri-dish on ice. If necessary, cut tissue into pieces with a sharp scalpel to improve penetration of the fixative (*see* N**ote 4**). Place the tissue on top of a 12x12 cm piece of Miracloth (*see* **Note 5**). Enclose the tissue into the Miracloth by folding and stapling the corners, generating a “tea bag” (see Figure 2). Completely submerge the bag into the PBS solution.

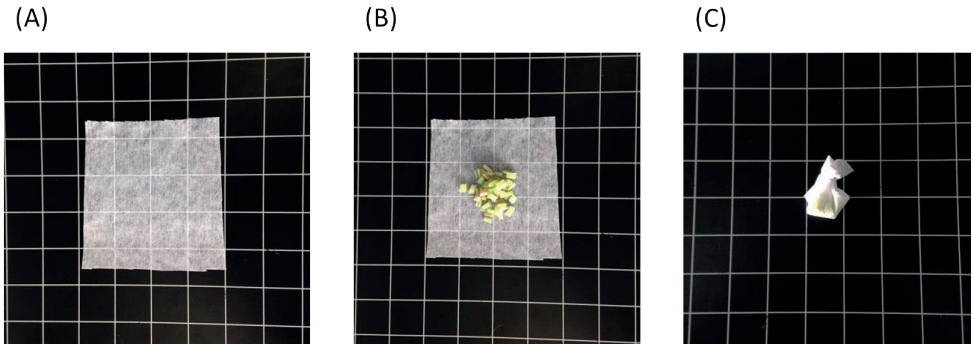



**Fig. 2** Preparation of a "tea bag" from Miracloth. One black square is 4 by 4 cm. (A) Cut a 12 by 12 cm piece of Miracloth. (B) Place the plant tissue at the center of the piece of Miracloth. (C) Fold by joining all corners and staple them together.

4. Place the tube under the fume hood. Add 10 mL of 4 % PFA.
5. Vacuum infiltrate the tissue at RT (to determine the time of fixation, *see* **Note 4**). During fixation turn vacuum OFF and ON again 3 times to ensure good penetration of the fixative.
6. Stop fixation by adding 1.25 mL of ice-cold 2 M Glycine (final concentration is 0.125 M) and vacuum infiltrate for 5 min.
7. Place the tube back on ice under the fume hood and discard PFA solution. Add autoclaved milliQ water to the tube, close the lid and shake vigorously to wash the tissue. Repeat this step twice. Discard PFA solution and milliQ water used for the washes according to your waste disposal regulations for fixative.
8. Place the tea bag in between two stacks of paper towels and press to dry the tissue. Repeat this process with new paper towels until the tissue is sufficiently dry (*see* **Note 6**).
9. Open the tea bag and place the tissue into a pre-chilled mortar containing liquid nitrogen. Grind the tissue into a fine powder. Avoid thawing the ground material.
10. Add sufficient NEB to submerge all the ground material (usually 10 to 20 mL). The NEB may freeze upon addition into the frozen mortar. Wait for it to thaw and mix from time to time. Avoid the suspension to warm up higher than 4 °C.
11. Place a new 50 mL tube on ice with a funnel on top.
12. Prepare a 12 by 12 cm piece of Sefar Nitek nylon filter (50 µm pore size) and of Miracloth. First place the Sefar Nitek filter in the funnel and then cover it with the piece of Miracloth, resulting in a two-layer filter.
13. Pipette the tissue suspension (from step 10) onto the two-layer filter and allow it to flow through by gravity. Rinse the mortar with an additional 5 to 10 mL of NEB and pipette it on top of the filter. Do not compress the filter! Upon squeezing you also obtain undesirable debris. Let gravity do its work. The filtrate contains your nuclei.

14. Centrifuge filtered nuclei at 1900 g for 15 min at 4 °C.
15. Promptly place the tube back on ice and gently pour off the supernatant. Resuspend the pellet in 1 mL of NEB and transfer the nuclei suspension into a 1.5 mL eppendorf tube. At this step the number and quality of extracted nuclei can be determined (*see* **Note 2**).
16. Centrifuge at 1900 g for 5 min at 4 °C.
17. Promptly place the eppendorf tube back on ice and gently pipette off the supernatant. Resuspend the pellet in 1 mL of NEB and repeat the centrifugation step (1900 g for 5 min at 4 °C).
18. Promptly place the tube back on ice.

**3.4 Digestion**

1. Gently pipette off the supernatant and resuspend the nuclei into 400 µL of 1.2x restriction buffer (refer to the manufacturer's instructions for the optimal restriction buffer).
2. Centrifuge at 1900 g for 5 min at 4 °C.
3. Promptly place the tube back on ice and gently pipette off the supernatant. Resuspend the nuclei into 500 µL of 1.2x restriction buffer.
4. Add 7.5 µL of 20 % SDS (final concentration 0.3 %) to permeabilize the nuclei and inactivate endogenous nucleases. Incubate at 65 °C for 40 min in a shaker at 900 rpm (*see* **Note 7**).
5. Place the tube at 37 °C for 20 min, still shaking at 900 rpm.
6. Add 50 µL 20 % Triton X-100 (final concentration is 2 %). Incubate at 37 °C for 1 hour while shaking at 900 rpm. The Triton X-100 will sequester the SDS, preventing a negative impact on the digestion efficiency. The susceptibility to SDS varies from one restriction enzyme to another. In case of poor digestion efficiency, the final concentration of SDS might have to be adjusted (*see* **Note 8**).
7. For a non-digested control sample: transfer 28 µL of the nuclei suspension to a new 1.5 mL eppendorf tube containing 272 µL of milliQ water. Store the tube at -20 °C until all samples will be decrosslinked (see step 7 of Subheading 3.5.).
8. Add 400 Units of Restriction Enzyme and incubate overnight at 37 °C while shaking at 900 rpm (*see* **Note 9**).

**3.5 Intra-molecular Ligation and de-crosslinking**

The ligation of fragments crosslinked together needs to be favored. Therefore, the volume of ligation needs to be sufficiently large to favor intramolecular ligations. At the same time, a too large volume will result in low DNA recovery. Thus, genome size-specific adaptations are required. For small-genome organisms such as *A. thaliana* the volume of ligation needs to be decreased compared to the volume used for large-genome organisms such as *Z. mays.* The following part of the protocol describes volumes based on the *Z. mays* genome size. To

determine in which volume intramolecular ligation should be performed for other organisms see **Note 10** and Table 2.

| Organism | *S. cerevisiae* | *A. thaliana* | *Z. mays* | *M. musculus* | *H. sapiens* |
|---|---|---|---|---|---|
| Genome size (Mbp) | ~12,5 | ~135 | ~2100 | ~2800 | ~3300 |
| Volume ligation reaction (mL) | 0,8 | 2 | 7 | 7,5 | 7,5 |
| Reference | [7] | [16] | [15, 19] | [25] | [18] |

**Table 2:** Reported volumes of ligation reactions in different 3C protocols and organisms with their respective haploid genome size.

1. Prepare 10x ligation Buffer and store at RT (*see* **Note 11**).
2. Inactivate the restriction enzyme either by heat inactivation (refer to the manufacturer's instructions, shake at 900 rpm) or by addition of 40 µL of 20 % SDS (final concentration is 1.6 %) followed by incubation for 25 min at 65 °C, 900 rpm.
3. Transfer the digested sample to a 50 mL tube and add 7 mL of 1x ligation Buffer (700 mL of 10x Ligation buffer plus 6.3 mL of sterile milliQ water).
4. Sequester the SDS by addition of 375 µL of 20 % Triton X-100 (final concentration is 1 %) and incubation for 1 hour at 37 °C.
5. For a digested control sample, pipette 300 µL of the digested sample into a 1.5 mL eppendorf tube and store the tube at -20 °C until the de-crosslinking step (see step 7).
6. Add 100 units of highly concentrated T4 DNA ligase to the ligation mix and incubate for 5 hours at 16 °C, followed by 45 min at RT (*see* **Note 3**).
7. Add 30 µL of 10 mg/mL proteinase K to the ligation mix, and 5 µL to the non-digested and digested samples previously stored at -20 °C (see step 7 of Subheading 3.4 and step 5 of Subheading 3.5).
8. Incubate all tubes overnight in a 65 °C water bath.

### 3.6 DNA purification

1. Place the Phenol:Chloroform:Isoamyl Alcohol and Chloroform:Isoamyl Alcohol solutions under the fume hood at RT at least 2 hours before starting (*see* **Note 12**).
2. Add 30 µL of 10 mg/mL RNase A to the ligation sample, and 5 µL to the non-digested and digested samples.
3. Incubate all tubes at 37 °C for 30 to 45 min.
4. Place all tubes under the fume hood and add 10 mL of Phenol:Chloroform:Isoamyl Alcohol to the ligated sample and 300 µL to the non-digested and digested samples. Close all lids tightly and shake vigorously.
5. Spin all tubes at 4500 g for 10 min at RT.
6. Transfer the aqueous phase to a new 50 or 1,5 mL tube.
7. Precipitate the DNA by first doubling the volume with milliQ, followed by addition of 1/10 volume of 2M NaOac (pH 5.6) and 2 volumes of 96 % Ethanol. To promote high

precipitation efficiency the addition of Glycogen is strongly recommended. Typically, 20-40 µL (20 mg/ml) is added to the ligated sample and 5 µL to the non-digested and digested samples.

8. Incubate all tubes at -80 °C for at least 2 hours.
9. Cool the centrifuge for 50 ml tubes (swing-out) and eppendorf centrifuge to 4 °C.
10. Centrifuge all tubes at 4500 g for 60 min at 4 °C.
11. Pour off the supernatant and wash the pellet of the ligated sample with 10 mL and the non-digested and digested samples with 1 mL of 70 % Ethanol.
12. Centrifuge all tubes at 4500 g for 20 min at 4 °C.
13. Promptly place the tubes back on ice and gently pour off the supernatant. With a pipette tip, remove the residual droplet of ethanol and let the pellet air-dry for 2-5 min. Alternatively, when dealing with multiple samples one can use a pump to remove the supernatant and then dry the pellet. Using a pump is certainly faster, but be cautious as loose pellets might get lost.
14. Resuspend the ligated sample in 150 µL, and the non-digested and digested samples in 15 µL Tris-HCl pH 7.5.
15. Incubate all tubes at 4 °C overnight to optimally resuspend the DNA pellet.
16. To assay the digestion and ligation efficiency, mix 2 µL of the generated 3C library with 1.5 µL of 10x loading buffer and 11.5 µL of milliQ water. Load the sample, together with the entire 15 µL of the non-digested and digested samples, and a positive control for digestion consisting of digested gDNA (see step 3 of Subheading 3.2), all with 10x loading buffer, on a 0.8 % Agarose gel. For the expected results, see Fig. 3 and **Note 13**.
17. The 3C library concentration can be estimated using a Qubit fluorometer. Alternatively, the concentration can be estimated on gel by comparison to a dilution series of a gDNA sample of known concentration. For more accuracy, load several dilutions of the 3C library on gel (1:1, 1:2, 1:4, 1:8). Note that nanodrop measurement is not reliable for quantification of complex DNA samples such as 3C libraries.

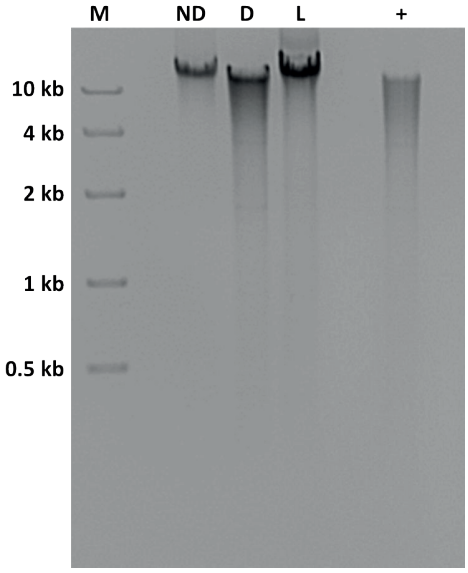



**Fig. 3** Non-digested (ND), digested (D) and Ligated 3C library (L) samples on a 0.8% agarose gel. The ND and L samples show one high molecular weight band, indicating intact DNA (ND sample) and efficiently ligated DNA (L sample). The presence of low molecular weight products in the ND and L samples would indicate DNA degradation and inefficient ligation, respectively. The digestion (D) sample has to show a similar banding pattern, including signs of satellite bands, as the positive digestion control (+), which consists of digested genomic DNA. M indicates the size marker lane.



### 3.7 Quantification of interaction frequencies and Data analysis

In 3C experiments the Relative Frequency of Interaction (RFI) is measured between a fixed fragment, the bait or viewpoint, and another fragment at the locus of interest. To estimate the RFI of a given fragment (e.g. the red fragment in Fig. 1) with the bait fragment (blue fragment in Fig. 1), one needs to compare it to the RFI of other fragments (black fragments in Fig. 1) with the bait. Below, we provide step-by-step qPCR and RFI quantification methods. The qPCR protocol is based on using SYBR green. The use of TaqMan probes is an alternative to SYBR green technology. For this, please see **Note 14**.

1. Prepare the qPCR mix, excluding primers, according to your own setup. The final volume for each reaction should be 20 µL.
2. Use 50 ng of 3C library as a template per reaction. Prepare similar reactions for the locus of interest and the endogenous control locus.
3. For the RL-Library controls, use the previously prepared dilution(s) complemented with gDNA (see step 33 of Subheading 3.2.1). Prepare similar reactions for the RL-Library dilution of the locus of interest and the endogenous control locus.
4. Add 2 µL of primer (1 µL bait primer + 1 µL fragment primer from 10 nM stock) to each well.

5. For each primer pair, perform a triplicate qPCR reaction on the 3C library and RL-Libraries.
6. For each primer pair, the RFI is calculated as follows:
   $RFI = (2^{-(Ct_{3Ci} - Ct_{RLi})})/(2^{-(Ct_{3Ce} - Ct_{RLe})})$

   $Ct_{3Ci}$ = mean value from the 3C triplicate for a given primer pair *i.*

   $Ct_{RLi}$ = mean value from the RL-Library triplicate for primer pair *i.*

   $Ct_{3Ce}$ = mean value from the 3C triplicate for the primer pair of the endogenous locus.

   $Ct_{RLe}$ = mean value from the RL-library triplicate for the endogenous locus.

7. Normalize all RFIs to the highest RFI value and plot the normalized RFIs as a function of their distance to the bait (Fig. 1E).

## *4. NOTES*

1. Use always fresh plant material if possible. Fresh material results in the most optimal digestion of fixed chromatin. In case material collection and the 3C protocol cannot be performed simultaneously, we advise to store fixed, dried material rather than fixed, ground material. This appeared more effective in our hands.
2. The amount of tissue to be processed per sample ranges between 1 to 3 grams of fresh plant material. This range should not be exceeded as too much tissue will affect the efficiency of fixation. At the same time, the amount of tissue used in each experiment should yield a sufficient number of nuclei (1 to 10 x $10^6$). To estimate the number of nuclei isolated, take a 2 µL aliquot after the first resuspension of nuclei in 1 mL of NEB (step 15 of Subheading 3.3). The nuclei have the tendency to sink to the bottom of the tube. Therefore, invert the tube with the resuspended nuclei gently three times before pipetting to avoid underestimating the actual yield. Add 2 µL of DAPI stain to the sample (final DAPI concentration of 2 µg/mL; Dilution Factor of 2) and count the nuclei on a hemocytometer using fluorescence microscopy (10-20x magnification, use DAPI filter). The total number of nuclei is estimated with the following formula: Total number of nuclei = (total number of counted nuclei x Dilution Factor x $10^4$)/(number of chambers counted). In case of a low number of nuclei per gram of tissue, one could consider generating multiple tissue samples and subsequently pool the nuclei together when resuspending the nuclear pellets (step 15 of Subheading 3.3).
3. Crowding reagents such as PEG cannot be used to increase the efficiency of the 3C ligation as it compromises the intramolecular nature of the ligation. Note that the addition of PEG 4000 (10 % final concentration) can be used to increase the ligation efficiency of the RL-library.
4. Some tissue types can be used directly for fixation, other types need to be cut in smaller pieces. For instance, relatively permeable tissue like Arabidopsis rosettes can

be used directly, while maize inner stem and husk tissue needs to be cut in ~1 $cm^2$ pieces before fixation. For each tissue type the optimal fixation time needs to be determined. Under-fixation will negatively affect the ability to detect chromosomal interactions. Over-fixation will negatively affect digestion and de-crosslinking efficiency, but also increase the background level of interactions. A good indication that tissue gets fixed is when it gets a translucent appearance. To define the optimal fixation time, one should perform a time series, fixing tissue samples for different time periods, and proceed with nuclei isolation and digestion (from step 6 of Subheading 3.3 to step 8 of Subheading 3.4), followed by de-crosslinking and DNA precipitation (from step 7 of Subheading 3.5 to step 14 of Subheading 3.6). Run the DNA samples on a 0.8 % agarose gel in parallel with an unfixed, de-crosslinked sample (no fixative added, no vacuum infiltration, positive control for DNA isolation) and a fixed, non-decrosslinked sample (negative control). At the optimal time of fixation, samples display efficient digestion and a relatively high DNA recovery after de-crosslinking and DNA isolation. Alternatively, one could test the effect of different formaldehyde concentrations while using a fixed incubation time.

5. For efficient fixation, wrap the tissue or pieces of tissue in a 12 by 12 cm piece of Miracloth, and close the Miracloth with a staple, generating a “tea bag” (see Fig. 2). The tea bag ensures complete submersion of the tissue during fixation, prevents spilling of tissue, and allows easier handling of the sample during subsequent washes.
6. To efficiently dry the tissue, place the Miracloth “tea bag” containing the tissue between a stack of paper towels and use a bottle or something similar as a roller to remove water. Repeat this procedure till the paper towels do not get wet anymore. Then the tissue is considered dry. Remaining water negatively impacts the grinding efficiency and hence the yield of nuclei.
7. The incubation of the chromatin for 40 min at 65 °C prior to digestion is crucial for inactivation of endogenous nucleases that would otherwise become active during the digestion step at 37 °C.
8. Digestion efficiency is sensitive to the chemicals present in a solution, including the SDS used to permeabilize the nuclei (step 4 of Subheading 3.4). In case of poor digestion efficiency, we recommend testing the effect of different SDS concentrations (0.05 to 0.3 % final concentration) on the digestion efficiency of the restriction enzyme.
9. Efficient digestion is key to a successful 3C experiment. For enzymes with a low performance over a long incubation period, we recommend to add fractions of the total number of enzyme units (400) at different time points during the digestion procedure. This helps to maintain high digestion efficiency.
10. Good intramolecular ligation conditions are important to ensure that only fragments crosslinked with one another are being ligated. To favor intramolecular ligation events, the ligation reaction is carried out in a large volume. This volume needs to be adapted to the genome size, as a low DNA concentration hampers an efficient precipitation of the ligation products, while a too high DNA concentration also allows intermolecular

ligation events. In Table 2 we report volumes of intramolecular ligation used in different published 3C protocols for different organisms. Note that the appropriate ligation volume (step 3 of Subheading 3.5) is influenced by the concentration of SDS present before the ligation step. If one wishes to lower the ligation volume, the volume of digestion should be adapted such that the SDS concentration will not hamper ligation efficiency. For instance, when working with *Arabidopsis thaliana*, in our hands a ligation volume of 2 mL (*see* **Table 2**) requires a digestion volume of 250 µL (step 3 of Subheading 3.4) followed by addition of SDS to a final concentration of 0.2 % (step 4 of Subheading 3.4).

11. We highly recommend preparing a fresh 10x ligation buffer each time to ensure efficient ligation. Addition of extra ATP (0.6 mg/mL) after 2 to 3 hours of incubation of the ligation reaction can further improve the ligation efficiency.
12. In case of low DNA recovery after the ligation step: make sure that thePhenol:Chloroform:Isoamyl-Alcohol and Chloroform:Isoamyl-Alcohol solutions are well equilibrated at RT before adding them to the samples. Cold Phenol:Chloroform:Isoamyl-Alcohol and Chloroform:Isoamyl-Alcohol are more prone to phase inversion. Increasing the amount of Glycogen added and incubation of the precipitation reaction overnight at -80 °C can help to increase precipitation efficiency.
13. Digestion efficiency can be evaluated on an agarose gel as indicated at step 16 of Subheading 3.6 (see Fig. 3). A more accurate evaluation of digestion efficiency can be achieved by designing a few primer pairs spanning restriction sites, followed by qPCR. The digestion efficiency of the 3C sample should be evaluated by comparison to a non-digested and digested gDNA template. Ideally the digestion efficiency is above 80 %. To check for variation in the amount of each template, use a primer pair amplifying a region not cut by the restriction enzymes used.
14. Plant genomes can harbor a very high density of repetitive elements (see e.g. *Zea mays*) [30]. Specific amplification of the desired amplicons can therefore be challenging. When analyzing the qPCR results, always perform melting curve analyses for all primer pairs to check amplicon specificity. If obtaining specific primers appears to be very difficult, the design and use of a TaqMan probe (Fig. 1) can help to increase signal specificity. With a TaqMan probe one does not rely on the use of a non-sequence-specific fluorescent dye such as SYBR. A TaqMan probe should be designed for the bait fragment, on the opposite strand of the bait primer (see Fig. 1). In this way, the quencher from the probe can only be removed when a new strand is synthesized using the primer annealed at the ligated fragment. TaqMan probes are usually designed as an approximately 30 bp oligo with a Tm 7-10 °C above the Tm of the primers [25].

***5. ACKNOWLEDGEMENT***

We gratefully acknowledge the support of the European Commission Seventh Framework-People-2012-ITN Project EpiTRAITS, GA-316965 (Epigenetic regulation of economically important plant traits) for Blaise Weber and Suraj Jamge. We thank Iris Hövel for her advice on 3C experiments on plant material.

**6. REFERENCES**

1. Heitz E (1928) Das heterochromatin der moose. Jahrb Wiss Bot 69:762–818.

2. Manuelidis L (1990) A view of interphase chromosomes. Science 250:1533–1540. doi: 10.1126/science.2274784

3. Gorkin DU, Leung D, Ren B (2014) The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell 14:762–775. doi: 10.1016/j.stem.2014.05.017

4. Sakabe NJ, Savic D, Nobrega MA (2012) Transcriptional enhancers in development and disease. Genome Biol 13:238. doi: 10.1186/gb-2012-13-1-238

5. Kolovos P, Knoch TA, Grosveld FG, Cook PR, Papantonis A (2012) Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin 5:1. doi: 10.1186/1756-8935-5-1

6. Wit E de, Laat W de (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev 26:11–24. doi: 10.1101/gad.179804.111

7. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295:1306–1311. doi: 10.1126/science.1067799

8. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10:1453–1465.

9. Vernimmen D, Gobbi MD, Sloane-Stanley JA, Wood WG, Higgs DR (2007) Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J 26:2041–2051. doi: 10.1038/sj.emboj.7601654

10. Comet I, Schuettengruber B, Sexton T, Cavalli G (2011) A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. Proc Natl Acad Sci 108:2294–2299. doi: 10.1073/pnas.1002059108

11. Zhao Z, Tavoosidana G, Sjölinder M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, Singh Sandhu K, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38:1341–1347. doi: 10.1038/ng1891

12. Grob S, Schmid MW, Luedtke NW, Wicker T, Grossniklaus U (2013) Characterization of chromosomal architecture in Arabidopsis by chromosome conformation capture. Genome Biol 14:R129. doi: 10.1186/gb-2013-14-11-r129

13. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485:381–385. doi: 10.1038/nature11049

14. Lieberman-Aiden E, Berkum NL van, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science 326:289–293. doi: 10.1126/science.1181369

15. Louwers M, Splinter E, van Driel R, de Laat W, Stam M (2009) Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C). Nat Protoc 4:1216–1229. doi: 10.1038/nprot.2009.113

16. Hövel I, Louwers M, Stam M (2012) 3C Technologies in plants. Methods 58:204–211. doi: 10.1016/j.ymeth.2012.06.010

17. Palstra R-J, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W (2003) The β-globin nuclear compartment in development and erythroid differentiation. Nat Genet 35:190–194. doi: 10.1038/ng1244

18. Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, Palstra R-J, Wendt KS, Grosveld F, van Ijcken W, Soler E (2013) Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. Nat Protoc 8:509–524. doi: 10.1038/nprot.2013.018

19. Louwers M, Bader R, Haring M, van Driel R, de Laat W, Stam M (2009) Tissue- and Expression Level-Specific Chromatin Looping at Maize b1 Epialleles. PLANT CELL ONLINE 21:832–842. doi: 10.1105/tpc.108.064329

20. Crevillén P, Sonmez C, Wu Z, Dean C (2013) A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. EMBO J 32:140–148. doi: 10.1038/emboj.2012.324

21. Kang H, Wiedmer A, Yuan Y, Robertson E, Lieberman PM (2011) Coordination of KSHV latent and lytic gene control by CTCF-cohesin mediated chromosome conformation. PLoS Pathog 7:e1002140. doi: 10.1371/journal.ppat.1002140

22. Kyrchanova O, Toshchakov S, Parshikov A, Georgiev P (2007) Study of the functional interaction between Mcp insulators from the Drosophila bithorax complex: effects of insulator pairing on enhancer-promoter communication. Mol Cell Biol 27:3035–3043. doi: 10.1128/MCB.02203-06

23. Cao S, Kumimoto RW, Gnesutta N, Calogero AM, Mantovani R, Holt BF (2014) A Distal CCAAT/NUCLEAR FACTOR Y Complex Promotes Chromatin Looping at the FLOWERING LOCUS T Promoter and Regulates the Timing of Flowering in Arabidopsis. Plant Cell Online tpc.113.120352. doi: 10.1105/tpc.113.120352

24. Jégu T, Domenichini S, Blein T, Ariel F, Christ A, Kim S-K, Crespi M, Boutet-Mercey S, Mouille G, Bourge M, Hirt H, Bergounioux C, Raynaud C, Benhamed M (2015) A SWI/SNF

Chromatin Remodelling Protein Controls Cytokinin Production through the Regulation of Chromatin Architecture. PLOS ONE 10:e0138276. doi: 10.1371/journal.pone.0138276

25. Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forné T (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc 2:1722–1733. doi: 10.1038/nprot.2007.243

26. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, Herman B, Happe S, Higgs A, LeProust E, Follows GA, Fraser P, Luscombe NM, Osborne CS (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 47:598–606. doi: 10.1038/ng.3286

27. Deal RB, Henikoff S (2011) The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat Protoc 6:56–68. doi: 10.1038/nprot.2010.175

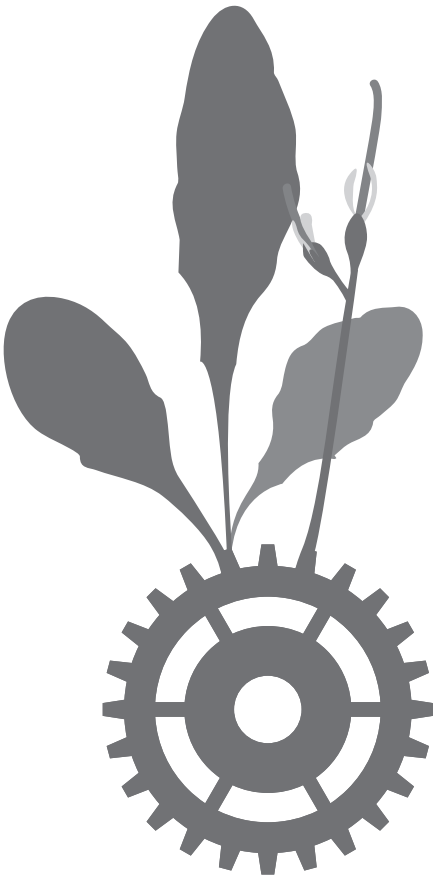
28. Wang D, Deal RB (2015) Epigenome profiling of specific plant cell types using a streamlined INTACT protocol and ChIP-seq. Methods Mol Biol Clifton NJ 1284:3–25. doi: 10.1007/978-1-4939-2444-8_1

29. Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EEM (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 44:148–156. doi: 10.1038/ng.1064

30. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115. doi: 10.1126/science.1178534

# CHAPTER 7

## A cautionary note on the use of chromosome conformation capture in plants



Suraj B. Jamge
Maike E. Stam
Gerco C. Angenent
Richard G.H. Immink

Plant Methods 2017, Volume 13 , Number 101

***ABSTRACT***

**Background:** The chromosome conformation capture (3C) technique is a method to study chromatin interactions at specific genomic loci. Initially established for yeast the 3C technique has been adapted to plants in recent years in order to study chromatin interactions and their role in transcriptional gene regulation. As the plant scientific community continues to implement this technology, a discussion on critical controls, validations steps and interpretation of 3C data is essential to fully benefit from 3C in plants.

**Results:** Here we assess the reliability and robustness of the 3C technique for the detection of chromatin interactions in *Arabidopsis*. As a case study, we applied this methodology to the genomic locus of a floral integrator gene *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1)*, and demonstrate the need of several controls and standard validation steps to allow a meaningful interpretation of 3C data. The intricacies of this promising but challenging technique are discussed in depth.

**Conclusions:** The 3C technique offers an interesting opportunity to study chromatin interactions at a resolution infeasible by microscopy. However, for interpretation of 3C interaction data and identification of true interactions, 3C technology demands a stringent experimental setup and extreme caution.

Keywords: 3C, *Arabidopsis*, plants, chromatin interactions

***BACKGROUND***

Perception and response to internal and external stimuli is the fundamental nature of cellular life. The transcriptional regulatory system plays an integral role in fulfilling the needs of the cell and organism by ensuring proper gene activity. In comparison to bacteria, transcriptional control of an eukaryotic cell is far more complex, involving several layers of regulation inside the nucleus. It takes more than just the action and sufficient quantity of activator or repressor proteins to modulate gene expression.

Many modules, such as transcription factors (TFs), RNA polymerase, chromatin remodellers and associated proteins, and regulatory DNA sequences, are determinants of eukaryotic transcription (Urnov and Wolffe, 2001; Coulon et al., 2013). All together these factors create an open chromatin structure, which is essential to initiate eukaryotic gene transcription. General TFs recognize and bind to discrete DNA sequences (also referred to as *cis*-elements) located in the core promoter region close to the transcription start site (TSS). For instance, the TATA box is one such evolutionarily conserved core promoter *cis*-acting element found upstream of most eukaryotic genes (Burley and Roeder, 1996; Patikoglou et al., 1999; Weber et al., 2016). Upon association of the general TFs to *cis*-elements, they interact with other proteins and form complexes to recruit RNA polymerase II, thereby initiating transcription. Examples of these other proteins include specific TFs, which can bind to *cis*-elements more distantly located from the core promoter elements (Krivega and Dean, 2012; Marsman and Horsfield, 2012). When these distant *cis*-elements are involved in the specific activation of gene expression they are called transcriptional enhancers. These enhancers can be found upstream, downstream, or within introns of coding regions and are reported to be located as far as several hundred kilo bases (kb) from the TSS (Louwers et al., 2009b; Marand et al., 2016). These distant enhancers can come into close proximity of their target sequences by protein-mediated chromatin interaction. In this respect, transcriptional gene regulation relies to a great extent on proteins that bind to DNA, not only close to the genes that they regulate, but also at distal DNA sites that can interact with the transcription initiation site by looping the intervening DNA. Thus DNA looping is speculated to be crucial to allow multiple proteins to regulate the core transcriptional machinery, resulting in a correct and controlled transcriptional response (Ptashne, 1986; Schleif, 1988; Matthews, 1992).

Over the last two decades much attention has been paid to the role of chromatin and its conformation in the regulation of gene expression. Various processes, including the differential deposition of histone variants, histone modifications such as methylation and acetylation, DNA methylation, and the activity of other non-histone architectural proteins are known to regulate the structure of chromatin (Rosa and Shaw, 2013; Jiang and Berger, 2016; Rodriguez-Granados et al., 2016). Empirical evidence add to the notion that the dynamics of higher-order chromatin conformation plays a crucial role not just in transcription, but also in other nuclear processes inherent to DNA (DNA replication, DNA repair, chromosome

transmission etc.). Therefore understanding the conformation of the chromatin within the cell nucleus has become a fundamental topic in biology.

Over the years, different imaging methods have been deployed to study chromosome conformation (Daban, 2011; Rapkin et al., 2012). However, detailed and local analysis of chromatin contacts with these methods has been complicated due to technical constraints. For instance, scanning electron microscopy (SEM) provides high resolution, but this technology is laborious, and most importantly, not suitable to study specific loci. Light microscopy has a limited resolution (200nm), and therefore is inadequate to define local chromosome conformation. Direct *in vitro* evidence of DNA looping has been shown using very-high resolution three-dimensional atomic force microscopy (AFM) (Puranik et al., 2014). Nevertheless, this method is labour intensive and an *in vitro* based approach. Artificial TFs fused with fluorescent proteins such as GFP do allow to spatially visualize and temporally track repetitive genome sequences *in vivo*, but the method still needs optimization in order to visualize unique individual loci and to detect chromatin interactions (Fujimoto et al., 2016). Fluorescence *in situ* hybridisation (FISH) is another alternative. However, this method involves stringent preparation treatments that can influence the chromatin organization itself and it was originally only suitable for the visualization of repetitive sequences (Fransz et al., 2002). Though, recent improvements and coupling of FISH with rolling-circle amplification of gene-specific circularizable oligonucleotides makes it possible to visualize the dynamics of individual loci (Feng et al., 2014a). Overall, microscopy studies have been crucial in defining chromosome territories and nuclear architecture at a single-cell level and new developments will probably allow to image individual chromatin contacts *in vivo* in the near future. Additionally, a new molecular approach has become available in recent years to study spatial organization of chromosomes at a high resolution, and this molecular tool is called **C**hromosome **C**onformation **C**apture (3C) (Dekker et al., 2002).

In 3C, chromatin in the intact nucleus is cross-linked by formaldehyde, followed by digestion with a restriction enzyme (RE) and intramolecular ligation (Dekker et al., 2002). The 3D conformation of the region or locus of interest is then studied by detecting ligation events occurring between non-neighbouring restriction sites. Possible interactions occurring between different chromosomal locations within the nucleus can be quantified as fused sequences by quantitative PCR (Hagege et al., 2007). The 3C method is cell population based and results in information about the relative frequency of interactions. The 3C method was initially developed for yeast by Dekker and co-workers and has been widely adapted to different model organisms shortly after. For plants, this method was also successfully applied to study chromatin conformation (Louwers et al., 2009a) and, since then it has become a powerful method to study gene looping in plants, as summarized below.

The first report exploring 3C in plants comes from a study in maize (*Zea mays*) that describes the role of a distant enhancer sequence at the *b1* locus. At the *b1* Locus, a hepta-repeat around 100 kb upstream of the transcription start site (TSS) appeared to interact with the TSS

region in a tissue and epiallele-specific manner (Louwers et al., 2009b). Since then a number of studies have highlighted the occurrence of chromosomal interactions in Thale Cress, *Arabidopsis thaliana*.

Crevillén *et al.* reported the presence and condition-dependent disruption of a chromatin loop at the *FLOWERING LOCUS C* (*FLC*) locus upon vernalization (Crevillen et al., 2013). *FLC,* a potent floral repressor and a polycomb target gene, is under tight control of winter cold. Vernalization is a classical epigenetic process in which prolonged cold exposure quantitatively affects the time of flowering. A robust gene loop, due to an interaction between the 5' and 3' flanking sequences of the *FLC* locus, has been reported and this interaction is independent of the level of *FLC* transcript in different genetic backgrounds and genomic contexts. However, upon vernalization, within the first two weeks of cold exposure, the loop is disrupted and it has been proposed that this disruption is an early event in the transition of the *FLC* locus to an epigenetically silent stage. Subsequently, other DNA contacts in the *FLC* locus increase in frequency under control of the COLDAIR and COLDWRAP long non-coding RNAs (LncRNAs), giving rise to polycomb-dependent and stable repression of *FLC* expression (Kim and Sung, 2017).

Another study by Liu *et al.* in 2013 reported the occurrence of a conformational change in chromosome looping at the *TERMINAL FLOWER1* (*TFL1*) locus that appears to be regulating *TFL1* transcription (Liu et al., 2013). In this case, disruption of the gene loop between the TSS and 3' distal region of the *TFL1* locus results in *TFL1* suppression.

Likewise, two independent studies have identified gene loops at another flowering related gene, *FLOWERING LOCUS T* (*FT*), that are associated with the photoperiod-dependent flowering response (Cao et al., 2014; Liu et al., 2014). *FT,* a floral integrator, can unite signals from multiple pathways to induce flowering. The first study reports the occurrence of multiple loops between a distal enhancer element (that contains CCAAT boxes) and core *cis* regulatory sites located in the promoter of *FT (pFT*) (Cao et al., 2014). Nuclear Factor-Y (NF-Y) is known to bind *CCAAT*-boxes and these CCAAT-bound NF-Y complexes are hypothesised to come into close proximity with core *pFT* sites, enabling improved recruitment and stabilized binding of CONSTANS (CO), together initiating photoperiod-dependent flowering in *Arabidopsis* (Cao et al., 2014). A second study showed the folding of the *FT* locus into a three-dimensional structure, favouring interactions between two regulatory regions (named as *Block A* and *Block C,* ~5.6kb apart) with another region called *Block ID,* an intermittent promoter region between *Block C* and *A* (Liu et al., 2014). An introduced change in the promoter length of *FT*, i.e. an increase in the distance between *Block C* and *Block ID,* by a T-DNA insertion, abolished the C-ID interaction and resulted in reduced chromatin interactions of Block C with Block A. It is noteworthy to mention that the chromatin interactions identified in these two studies do not overlap.

Together, the 3C studies discussed above provide intriguing insights into the possible roles of chromatin interactions to regulate gene expression in plants, similar to the studies reported in yeast and other model organisms (Levine et al., 2014; Pombo and Dillon, 2015).

3C provides an interesting opportunity to study *in vivo* chromatin interactions at a high-resolution and thus has become a standard method for studying chromatin contacts at single gene loci (Louwers et al., 2009b; Louwers et al., 2009a). However, like every other method, 3C has its own shortcomings. The challenges and technical issues of this method can at times outweigh its advantages. Therefore a good experimental setup, rigorous controls, and unbiased data analysis are crucial for meaningful interpretation of 3C data. This is clearly evident from studies performed in other model organism (mammalians, yeast etc.), where several papers have highlighted the importance of necessary 3C controls and appropriate data analysis (Dekker, 2006; Hagege et al., 2007; Miele and Dekker, 2009; Gavrilov et al., 2013; O'Sullivan et al., 2013). However, cautionary notes are largely missing in the plant science community. In this study we assessed the reliability and robustness of the qPCR-based 3C method in *Arabidopsis*. Based on this investigation, we provide detailed guidelines on necessary controls and how interaction data should be interpreted in a 3C experiment. Intricacies of this promising but challenging technique are further discussed.

## *RESULTS AND DISCUSSION*

### Chromosome Conformation Capture (3C) in *Arabidopsis*

To assess the reliability and robustness of the 3C technique for the detection of chromatin interactions in *Arabidopsis*, we used this methodology to investigate the chromatin conformation at the locus of the floral integrator gene *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1)* (Lee and Lee, 2010). Initially, the *SOC1* locus, including the ~3.8 kb promoter, the gene body, and ~1 kb downstream region, was divided into fragments using the four-cutter RE *Fsp*BI, as schematically shown in Fig. 1. Twelve distinct fragments of varying lengths (smallest fragment IX of 276bp and longest fragment VII of 1475bp), spanning the entire *SOC1* locus, were tested for chromatin contacts. Fragment VII, which contains the transcriptional start site (TSS), was used as the bait (also referred to as 3C anchor) to generate a chromatin interaction profile (Fig. 1). Throughout this study, proper controls were used as described previously (Louwers et al., 2009a) (also see methods) to ensure that only valid chromatin contacts are detected and quantified. As seen in Fig. 1, multiple contacts between the anchor and other regions of the *SOC1* locus were identified. Overall the 3C interaction profiles observed were consistent and reproducible across independent biological samples.

In a 3C experiment the fragment(s) that show(s) the highest interaction frequency with the bait fragment is (are) considered as chromatin contact(s). For the TSS region in the *SOC1* locus the highest interaction frequencies were observed with fragments X, XI and XII, all downstream of the VII-bait (Fig. 1). Furthermore, a potential contact with a promoter region, Fragment IV, was identified.

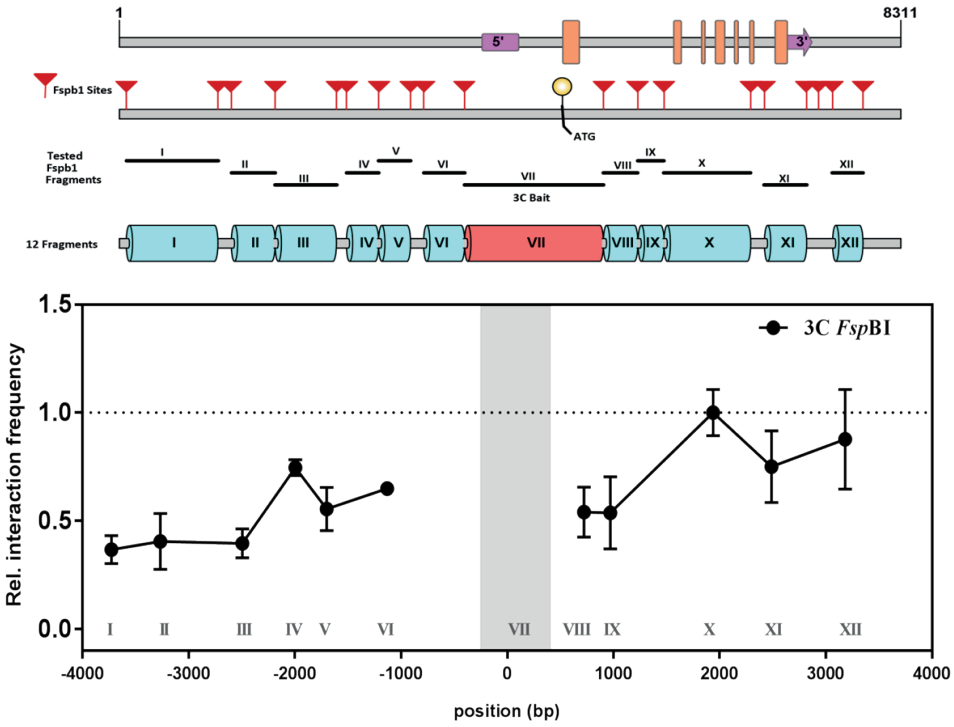



Fig. 1 **Chromosome Conformation Capture at the *SOC1* locus of *Arabidopsis* using the *Fsp*BI restriction enzyme**. In the top panel a schematic diagram of the *SOC1* locus is shown indicating the positions of all *Fsp*BI restriction sites (red arrowheads), and the fragments tested for interaction with a region spanning the Transcriptional Start Site (TSS; Fragment VII). The pink blocks indicate the 5' and 3'UTR and the orange blocks highlight the position of exons along the *SOC1* locus. The graph shows the observed relative interaction frequency of regions with the anchor VII, performed on three week-old wild type (Col-0) seedlings. Relative interaction frequencies are plotted on the y-axis. Distances in base pairs (bp) relative to the TSS of *SOC1* are plotted along the x-axis. Mean (±SD) derived from three independent biological samples is indicated. High interaction frequencies with the bait fragment indicate potential chromatin interactions.



## Cross-Validation of Chromatin contacts

One way to validate the putative chromatin contacts identified from a 3C experiment, is by performing a reciprocal 3C (r3C) experiment. In an r3C experiment, the fragment showing the highest interaction frequency with the 3C bait in the initial assay is used as a new 3C bait to generate an interaction profile. Thus we performed r3C experiments using one of the potential interacting fragments downstream of the *SOC1* TSS as bait (Fragment X; Fig. 1) aiming to verify the identified contacts. The chromatin interaction profile generated using fragment X as bait is shown in Fig. 2. As expected, we identified a high interaction frequency for the combination X-VII, validating the initial identified 3C contact point (Fig. 1). Moreover, another region (XI)

showed an equally high interaction frequency when X was used as bait. However, since XI is in immediate proximity of the 3C bait, this high interaction frequency might be the result of random collisions of neighbouring fragments, a phenomenon often observed in 3C experiments (Dekker, 2006).

The key difference between the 3C and r3C experiment is the fragment that is selected as bait and consequently, the combinations of qPCR primers used to detect the relative interaction frequency. For the *SOC1* locus we identified in this way e.g. an interaction between bait VII and fragment X (Fig. 1) and therefore fragment X was used as bait in the r3C experiment (Fig. 2). In any PCR-based 3C experiment, the primer of the bait is kept constant and is combined with a unique primer annealing specifically to one of the fragments that is tested for interaction (Louwers et al., 2009b). It is good to realize that in a 3C and r3C experiment the combination of primers to test the interactions between one specific combination of fragments (VII and X in the example of *SOC1*), is identical. Performing an r3C experiment is certainly of value, since a comprehensive profile of chromatin interaction of the locus will be obtained from yet another viewpoint. However, it is good to realize that the outcome of 3C experiments might be biased due to different characteristics of the used REs or technical constraints of PCR (Frohler and Dieterich, 2009). Therefore, it is desirable to perform another independent type of validation, besides the r3C experiment.

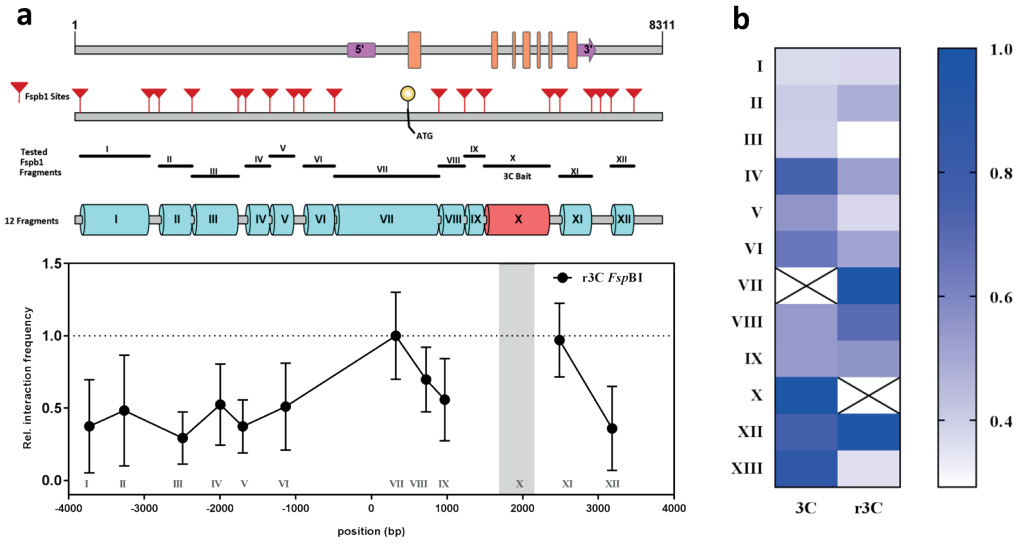



Fig. 2 **Reciprocal Chromosome Conformation Capture (r3C) at the *SOC1* locus of *Arabidopsis* using the *Fsp*BI restriction enzyme.** A) Schematic diagram of *SOC1* locus showing the position of all *Fsp*BI restriction sites, and the fragments tested for interaction. The pink blocks indicate the 5' and 3'UTR and the orange blocks highlight the position of exons along the *SOC1* locus. The graph shows the outcome of the r3C analysis of *SOC1* with *Fsp*BI using fragment X as bait. The experiments were performed on the same 3C libraries used in the original 3C experiment presented in Figure 1. Mean (±SD) derived from three biological replicates is plotted. B) heat map summarizing and comparing the 3C and r3C interaction profile obtained with *Fsp*BI. The cross indicates the

fragment used as bait. Note that the highest reciprocal contact frequency was identified between fragment VII and X. The intensity of blue colouring is an indication of the relatively interaction frequency (scale from 0-1.0).

**Validation of 3C interaction profiles with another restriction enzyme**

One of the best options to confirm and validate the outcome of a 3C experiment is to repeat the 3C experiment with yet another RE. A similar 3C interaction profile obtained from two independent REs strengthens the reliability of identified chromatin contacts. Further, it allows a more precise identification of the specific chromosomal regions that interact. With this in mind, we re-examined the chromatin interaction profile for the *SOC1* locus using a different four-cutter enzyme (*Nla*III). Now 16 distinct fragments of varying length spanning the entire *SOC1* locus were tested for chromatin contacts (Fig. 3). The interaction profile of this validation 3C assay is shown in Fig 3a. Fragment I spanning the TSS, which to a large part overlaps with Fragment VII used as bait upon the *Fsp*BI digestion (Fig. 1), was used as bait. The highest interaction frequency was observed for the combination 'I-N'. In addition to that, bait I also interacted with fragments L and C at a relatively high frequency. By comparing the 3C profiles obtained with the two REs (Fig. 3b), we identified at least one contact to overlap in both 3C experiments with the region spanning the TSS, and this is represented by the regions X and L. However, region N was found as novel interacting region for bait I, but this region was not represented in the *Fsp*BI run due to multiple closely located *Fsp*BI restriction sites. Consequently, this part of the locus became too fragmented for reliable qPCR primer design and amplification and was not monitored in the *Fsp*BI-based experiment. Besides this lack of coverage of some regions due to the selected RE, tested fragments do not completely overlap and this can result in differences. In the case of *SOC1 for example*, it is possible that the interaction between bait VII and region XII detected in the *Fsp*BI experiment is due to a contact between a sequence in the 5'-end of bait VII, and therefore not identified in the *Nla*III experiment fragment I (Fig. 3b).

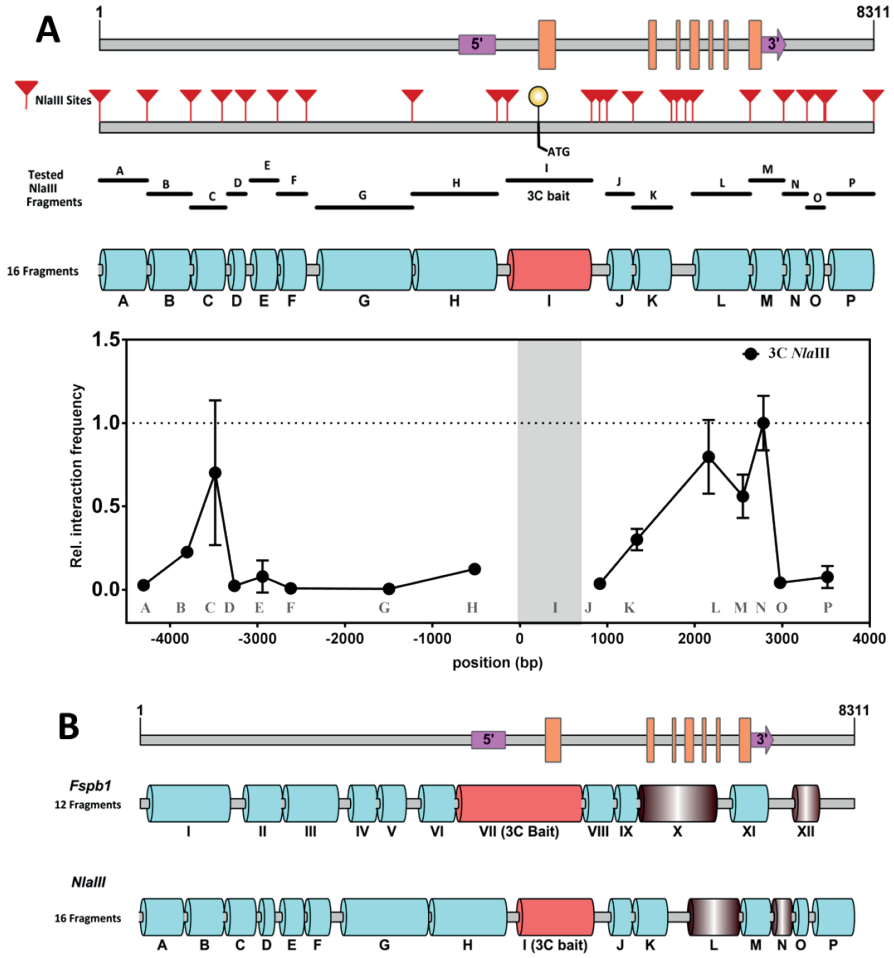



Fig. 3 **Chromosome Conformation Capture at the *SOC1* locus of *Arabidopsis* using the *Nla*III restriction enzyme.** A) Schematic diagram of *SOC1* locus showing the positions of all *Nla*III restriction sites (red arrowheads), and the fragments tested for interaction with a region spanning the Transcriptional Start Site (TSS; Fragment I). The pink blocks indicate the 5' and 3'UTR and the orange blocks highlight the position of exons along the *SOC1* locus. The graph represents the results of the 3C analysis on three-week old wild-type (Col-0) seedlings using fragment I as 3C bait. Mean (±SD) for two biological replicates plotted. B) A schematic representation of the top two interacting fragments (in brown) identified at the *SOC1* locus with *Fsp*BI and *Nla*III.

Although not all potential interactions were validated, the results obtained with the second RE supports the initially identified interaction between a fragment around the *SOC1* TSS and a fragment towards the end of the coding region of the gene (X for *Fsp*BI and L for *Nla*III, respectively). The next logical step was to perform an r3C experiment using *Nla*III. For this purpose fragment N, located in the 3'-region of *SOC1* (Fig. 3 and 4a) was used as a bait, as this region showed the highest interaction frequency with region I (Fig. 3a). Surprisingly, the observed relative interaction frequency for the combination N-I in the r3C experiments was

extremely low, suggesting no interaction (Fig. 4). Instead we identified two other potential contact points, both located in the *SOC1* promoter, i.e. fragments A and D. Notably, we obtained a similar deviating result in all biological replica's that were tested. Although we cannot exclude the possibility that this result is due to the way interaction frequencies are calculated and interpreted, the outcome (Fig. 4) shows the importance of applying multiple validation and confirmation experiments, such as r3C and the use of a second RE.

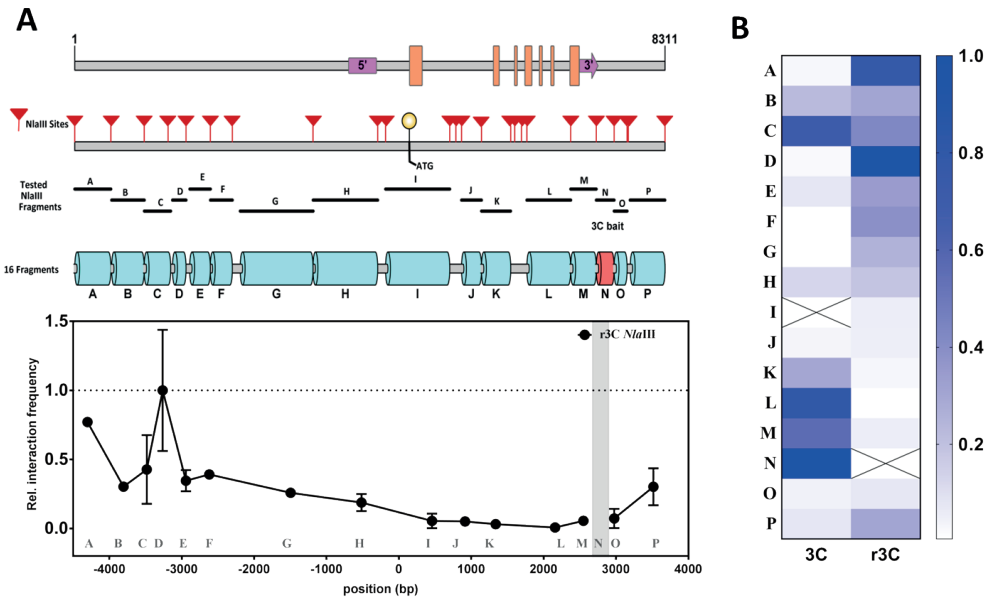



Fig. 4 **Reciprocal Chromosome Conformation Capture (r3C) at the *SOC1* locus of *Arabidopsis* using the *Nla*III restriction enzyme.** A) Schematic diagram of the *SOC1* locus showing the position of all *Nla*III restriction sites (red arrowheads), and the fragments tested for interaction. The pink blocks indicate the 5' and 3'UTR and the orange blocks highlight the position of exons along the *SOC1* locus. The graph represents the results of the r3C analysis on three-week old wild type (Col-0) seedlings with fragment N as bait. Mean (±SD) for two biological replicates is plotted. B) A heat map summarizing and comparing the 3C and r3C interaction profile performed with *Nla*III. The cell with a cross indicates the bait. The intensity of blue colouring is an indication of the relatively interaction frequency (scale from 0-1.0).

The observed contradicting results prompted us to investigate potential reasons of miss-interpretation of 3C results due to the lack of sufficient controls or technical constrains of the 3C technology. A user of 3C defines e.g. the bait region and the size of the region of interest to study, and hence the number of potentially interacting fragments to be monitored. These choices affect the 3C outcome, since the measured interactions are relative to one another with the highest interaction frequency set as one. In literature, we commonly come across 3C studies focusing on promoter regions only, the entire gene locus, or a specific distal enhancer region to identify e.g. promoter-enhancer contacts. When we re-analyzed our data starting from the hypothesis that there is an interaction between the TSS and an upstream *SOC1* promoter region, and therefore monitored this part of the locus only, we observed the highest

interaction between bait VII and promoter fragment IV with the *Fsp*BI restriction profile (Fig. 5). Although this chromosome interaction was also detected in our initial experiment (Fig. 1), our attention was directly drawn towards the region in the 3'-end of the *SOC1* gene, for which the highest relative interaction frequencies were found using bait VII. However, more worrying is the non-overlapping pattern at the *SOC1* promoter observed for the *Nla*III restriction profile, in which bait I interacts with promoter fragment C (Figs. 3A, 5). Surprisingly, this interaction pattern is not overlapping at all with the *Fsp*BI relative interaction pattern. Once more, it is possible that this deviation is caused by the lack of full overlap between the bait fragments. Sequences in the 3'-end of fragment I might be essential for the interaction with fragment C. Nevertheless, these results show how much the outcome depends on the choice of the RE to be used and which chromosomal region is taken and by that, the inclusion or exclusion of particular high interacting regions.

Overall, upon performing 3C at the *SOC1* locus independently with two different REs, we found various similarities, but also differences in the generated 3C profiles. One clear trend observed was the interaction of the TSS-spanning bait region with a region close to the 3' UTR of the *SOC1* locus for both restriction profiles. A striking discrepancy were the interaction patterns identified for the TSS bait fragment with the promoter region. Hence, our data show the potential subjectivity of 3C results and prompt for cautiousness when interpreting 3C interaction patterns.

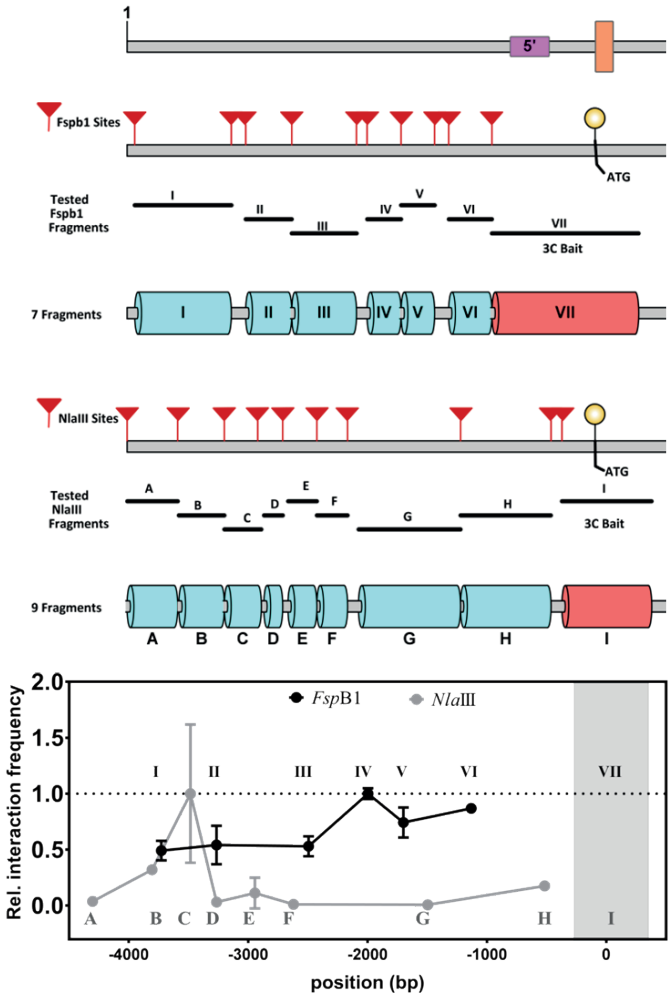



Fig. 5 **Chromosome Confirmation Capture at the *SOC1* promoter only.** In the top panel a schematic diagram of *SOC1* promoter is given, showing the position of all *Fsp*BI and *Nla*III restriction sites, along with the fragments monitored for interaction. The graph represents the results of the 3C analysis based on the *FspB*I and *Nla*III data sets and using fragment VII and I as bait, respectively.

## Classification of 3C interactions

All the above discussed results reveal the subjectivity of 3C data interpretation. Furthermore, in a 3C experiment, for regions in close proximity to the bait fragment under study, usually a high occurrence of ligations due to random (non-specific) collisions is observed (Dekker, 2006; Dekker et al., 2013). Thus, mere identification of ligation events does not necessarily mean the occurrence of an existing *in vivo* interaction. In order to be able to detect interactions above the background of random interactions, it's important to carefully consider the proximity between the bait and the supposed interacting fragments. The *Arabidopsis* genome

in nature is very small and compact (Arabidopsis Genome, 2000), hence studying regulatory interactions at individual gene loci is complex, due to the small distances between neighbouring restriction fragments thereby resulting in high potential for random interactions (Dekker et al., 2013). In conclusion, the combination of compact genome of *Arabidopsis* and the flexible nature of chromatin fibres makes interpretation of 3C interaction on individual loci challenging and demands high caution.

Reviewing published literature we found only a limited number of 3C studies that have been performed in plants. Among these studies we observed considerable shortcomings within the 3C experimental set up (e.g. the PCR method, lack of endogenous normalization and random ligation libraries as controls, use of only one enzyme and no r3C), and differences in the way the 3C interaction data have been interpreted and represented. Furthermore, comparing 3C results across experiments and laboratories is complicated due to differences in the experimental set up, such as PCR method, normalization method, plant growth conditions, and the way the interaction frequency is determined. For example, in literature, one can find chromatin interactions determined using 3C by either semi-quantitative PCR or by qPCR-based approaches. Only the latter provide numeric peak interactions and is therefore a much more trusted and widely adopted method in recent years. A recent publication manually curated more than 3000 interactions from 5000 publications across 17 species into a database called 3C database (3CDB) (Yun et al., 2016). This 3CDB classified the strongest 3C interactions into four distinct classes based on their reliability. Class I and II cover the semi-quantitative PCR data, which nowadays is not an accepted detection method, whereas Class III and IV refer to numeric peak interactions. According to the 3CDB classification, interactions belonging to class IV are considered to be the most reliable, due to the fact that they are validated with an r3C experiment (Yun et al., 2016). The results we described in this study are all numeric peak interactions and fall into class III and IV. However, based on our interaction profiles from independent 3C experiments that used different REs, we see a need to further extend this set of classifications. We suggest introducing a class V for interactions that have been identified and r3C-validated using two different REs. Taking into account that even in this set up the experimental procedure is the same and provides relative and hence, subjective data, 3C experiments should be confirmed by an alternative and independent method to get full proof for a potential chromosome interaction.

**Challenges of the 3C method**

Ligation based methods, such a 3C, heavily rely on a sound experimental design. Many technical biases may be introduced if the design and set up of the experiment is not optimal. For example, optimization of crosslinking conditions is necessary, as over-fixed chromatin often renders digestion with REs inefficient. Similarly, biases may arises from the choice of RE and its digestion efficiency. One of the prerequisites for REs in 3C is their ability to digest crossed-linked chromatin efficiently, but at the same time providing the desired resolution at the locus of interest. By far not all REs behave optimally in 3C, e.g. because buffering

conditions during digestion are sub-optimal (e.g. due to presence of detergents). Hence, optimization of several parameters is essential in order to overcome technical biases in a PCR-based 3C experiment. Most of these technical aspects and their optimization procedures have been very-well addressed in literature (Splinter et al., 2004; Dekker, 2006; Hagege et al., 2007; Louwers et al., 2009a; Hovel et al., 2012; Naumova et al., 2012). Besides these technical issues, biological variation may arise from differences in the growing conditions, the time of tissue collection in relation to the developmental age of the plant, and due to sampling itself. These aspects are very critical, especially when the goal is to study dynamics in chromatin interactions in time or upon a change in condition, as was e.g. done for the *FLC* locus [25]. To exemplify this, we performed a 3C assay on a developmentally different tissue where *SOC1* is known to be actively expressed (Immink et al., 2012). For this experiment we sampled the same type of plant material exactly one week later, but upon transfer of the plants to flowering-inducing long day (LD) conditions and after growth at these conditions for seven days (7dai). At this time point, we observed a distinct chromatin interaction profile at *SOC1* locus, when using the *Nla*III RE (Supplemental Fig. 1), suggesting dynamics in relation to *SOC1's* transcriptional state. However, when 3C was repeated on this sample using *Fsp*BI (Supplemental Fig. 2), we obtained a pattern resembling the pattern after three weeks growth under SD conditions (Fig. 4b), suggesting lack of dynamics. This example reveals that extra caution is required when studying dynamic interactions and that it is of utmost importance to keep biological variation at minimum.

**Beyond the traditional 3C**

Since the development of the 3C method, many variants of this technology have been rolled out (reviewed in (de Wit and de Laat, 2012; Denker and de Laat, 2016), Table 1). These variants enabled the user among others to study chromatin interaction at a genome-wide scale. 4C combines the traditional 3C assay with microarrays (3C-on-chip) (Simonis et al., 2006), and thereby a user can examine one-to-all contacts throughout the genome, instead of exploring one-to-one locus specific contacts as is done in classical 3C experiments. Advancements and development of cost-effective sequencing techniques gave birth to a wide range of sequencing variants of the 3C method (Table 1), improving the overall resolution of the interaction profile. Instead of using one viewpoint, some 3C variants, such as multiplex 3C-seq (many–to-all) and Hi-C (all–to-all), provide the opportunity to explore genome-wide interactions from multiple viewpoints simultaneously. More and more computational tools and packages are now publically available making it easier to process and analyse the vast amount of genome-wide interaction data (Paulsen et al., 2014; Ay and Noble, 2015; Lajoie et al., 2015; Sato and Suyama, 2015; Schmid et al., 2015; Yardimci and Noble, 2017). Thus in comparisons to traditional 3C, nowadays some 3C variants might be more attractive, robust, and cost effective to perform. Therefore, we recommend users to compare the ins-and-outs of all these methodologies taking into account their research question, before deciding on the appropriate 3C method of choice (reviewed in (Denker and de Laat, 2016)). However note that the full potential of all these variants still remains to be exploited in plants. A few studies did

make use of 4C and Hi-C approaches to gain insight into the three-dimensional chromatin configuration of Arabidopsis genomes (Moissiard et al., 2012; Grob et al., 2013; Feng et al., 2014b; Grob et al., 2014; Wang et al., 2015). In comparison to the majority of other plant genomes, the Arabidopsis genome is densely packed with a gene density of one gene per 4.5kb. However, most of the chromosome conformation capture technologies are best suited to study mid-range and long-range chromatin interaction and therefore, less suitable for Arabidopsis. Nevertheless, one Hi-C study did report contact maps of up to 2kb resolution (Wang et al., 2015). But, when it comes to studying short-range interactions, all the above discussed variants are limited in resolution in comparison to the (q)PCR-based 3C methods. Further improvements in sequencing depths, choice of REs (e.g. micrococcal nuclease, four cutters) and overcoming computational barriers may drastically improve the resolution of these technologies in the near future, enabling the generation of unbiased high-resolution chromatin interaction maps.

| Ligation based Chromatin capture method | Application | References |
|---|---|---|
| **qPCR-based 3C** | One-to-one | (Hagege et al., 2007; Louwers et al., 2009a) |
| **3C-seq, 4C** | One-to-all | (Soler et al., 2010; van de Werken et al., 2012) |
| **5C** | many-to-many | (Dostie and Dekker, 2007) |
| **ChIA-PET** | many-to-many | (Fullwood et al., 2009) |
| **Multiplex 3C-seq** | many-to-all | (Ribeiro de Almeida et al., 2011; Stadhouders et al., 2012) |
| **HiCap,CHi-C** | many-to-all | (Dryden et al., 2014; Jager et al., 2015; Sahlen et al., 2015) |
| **Capture-C** | many-to-all | (Hughes et al., 2014) |
| **T2C** | many-to-all | (Kolovos et al., 2014) |
| **Hi-C, Dnase Hi-C, Micro-C, Micro-CXL,** | all-to-all | (Lieberman-Aiden et al., 2009; Hsieh et al., 2015; Jager et al., 2015; Hsieh et al., 2016; Ramani et al., 2016) |
| **TCC** | all-to-all | (Kalhor et al., 2012) |

**Table 1:** Overview of existing and recently developed 3C based methods

***CONCLUSIONS***

3C is a powerful tool when it comes to studying chromatin interactions at a gene specific locus. However, identification of valid interactions via PCR-based 3C demands multiple controls and validation steps. Only when the results are consistent across the proper control and validation experiments, an interaction can be considered of high confidence. Subsequently, it is of interest to unveil for every high confidence interaction whether it exists because of the regulation of gene expression, a particular nuclear or chromatin organisation, or spatial restrictions in the nucleus or the flexible nature of chromatin. Hence, identified interactions do not reveal the underlying mechanism behind its co-localization, neither do they distinguishing if it's a functional or non-functional interaction. To shed more light on the functionality of an observed interaction, genetic studies are essential. For instance, making use of T-DNA insertion lines, or targeted disruption of the DNA regions involved in the observed interactions with CRISPR-Cas9 genome editing (Woo et al., 2015), can aid further functional characterization of identified *in vivo* interactions.

Since the establishment of the 3C technique, hundreds of potential interactions have been reported supporting the potential role of chromatin interactions in transcriptional control. As the plant scientific field is gaining momentum in deciphering this new layer of transcriptional regulation of intricate gene regulatory networks, the 3C technique will play a prominent role in expanding our knowledge on this new fundamental topic of plant biology. Nevertheless, utmost care should be taken in assigning meaningful 3C interactions, as described here.

***METHODS***

*Plant material, growth conditions and tissue collection.*

Col-0 wild type plants were grown on rock-wool for three weeks at 20$^{o}$C under short day (SD) conditions (8 h light, 16 h dark). Two grams of seedling material (above ground tissues) per biological sample were collected during the afternoon hour of the day. In addition, material was sampled seven days later and after transfer and growth of the plants for a week at long day photoperiod conditions (16 h light, 8 h dark).

*Chromatin conformation Capture (3C)*

3C was performed on the *SOC1* locus using the previously described protocol with some adaptations for *Arabidopsis* (Louwers et al., 2009a). Two grams of *Arabidopsis* above-ground seedling material was crosslinked with 2% paraformaldehyde PBS buffer under vacuum for 30 minutes (mins) on ice. The cross-linking reaction was stopped by addition of ice-cold 2M glycine (final concentration = 0.125M) under vacuum for 5 mins on ice. The crosslinked tissue was ground and nuclei were isolated and purified using nuclei extraction (NE) buffer. Before digestion, the purified nuclei pellets were re-suspended in 1.2x restriction buffer and treated with 0.2% SDS at 65$^{o}$C for 20 mins. Later, SDS was sequestered by incubating with 2%Triton X-100 for 30 mins. 3C analysis was performed on the *SOC1* locus using two different REs (namely *Fsp*BI and *Nla*III) independently. 400 U of RE was used for overnight digestion at 37$^{o}$C. Digestion was stopped by incubation at 65$^{o}$C for 20mins. Ligation was performed using 100 U of T4 DNA ligase, initially at 16$^{o}$C for 5 h, followed by room temperature for 45 mins. Reverse cross-linking was done overnight with a treatment of proteinase K at 65$^{o}$C. After reversal of the crosslinks, phenol/chloroform extraction and ethanol precipitation was performed for recovery of the DNA.

*3C Primers, Controls and Quantification*.

All the primers used in this study are listed in Table 2 and 3. For a detailed discussion on controls we highly recommend these published studies (Splinter et al., 2004; Dekker, 2006). The relative interaction frequencies of one fragment to another were calculated based on quantitative PCR (qPCR) data using SYBR Green I master mix. For the analysis of the specific ligation events, two controls were used. First, in order to correct for the primer amplification efficiencies, for each primer pair the qPCR dataset was normalized with an random ligation (RL) control sample. The RL control sample was obtained by digestion of a BAC clone containing the *SOC1* locus and followed by re-ligation in small volumes to obtain all the possible random ligation events. Secondly, in order to control for the quality and quantity of each 3C sample, the 3C data needed to be further normalized to 3C values measured for an endogenous control locus (usually a reference gene) unrelated to the *SOC1* locus. The chromatin state of such a reference gene is assumed to be stable across samples. Therefore, for each 3C sample, 3C values were also obtained for the reference gene *TIP41-like*. The reference gene primer amplification efficiencies were also corrected with a RL control

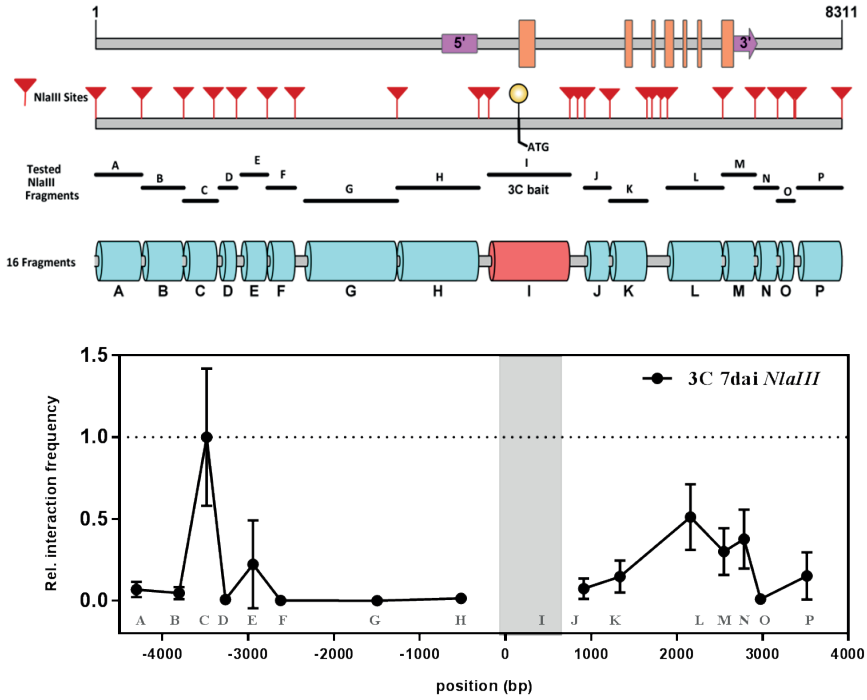
obtained by digestion and re-ligation of a BAC clone containing the *TIP41-like* locus. The 3C data of *SOC1* was normalized to the 3C values measured for the *TIP41-like* locus to obtain relative interaction frequencies. For more details on step-by-step data analysis of the qPCR-based 3C method see (Weber et al., 2018). All figures shown in this study are the mean of two or three independent biological samples.

| **Table 2** | **3C primers for *Fsp*BI restriction profile** | | |
|---|---|---|---|
| **Name** | **Gene locus and Restriction enzyme** | **Primer on** | **Seq 5' to 3'** |
| **PDS6848** | SOC1_*Fsp*BI | Fragment I | AGATTCTCAAACATCAGTCGGA |
| **PDS6849** | SOC1_*Fsp*BI | Fragment II | ACAAAAGGAGTAGGTTTCTGGA |
| **PDS6850** | SOC1_*Fsp*BI | Fragment III | TGAGCTTATGACTGGTAAACTC |
| **PDS6851** | SOC1_*Fsp*BI | Fragment IV | GTTTTGGATTTGTCTCAACCAG |
| **PDS7489** | SOC1_*Fsp*BI | Fragment V | TGGTCCTCCTCCCGATATAGA |
| **PDS6852** | SOC1_*Fsp*BI | Fragment VI | ACGAGAGAGTGTTTGTGTCC |
| **PDS6847** | SOC1_*Fsp*BI | Fragment VII (Bait) | GACGTTTGCTTTGAGAGGTG |
| **PDS6853** | SOC1_*Fsp*BI | Fragment VIII | GCTTCATTTCATGCTCATTCC |
| **PDS6854** | SOC1_*Fsp*BI | Fragment IX | ACTTCTTTCTCTCGAACCTACT |
| **PDS6855** | SOC1_*Fsp*BI | Fragment X | AGTAAGTAAGCCTCTTGTGCT |
| **PDS6856** | SOC1_*Fsp*BI | Fragment XI | AGCTGCTTCTCTCTTGTTGT |
| **PDS6857** | SOC1_*Fsp*BI | Fragment XII | AAGGGCCTACTTTGCGATAA |
| **PDS7307** | TIP41_Like *Fsp*BI | Bait-TSS | GTTTCGATCTCCCAGTCATG |
| **PDS7308** | TIP41_Like *Fsp*BI | -500bp | AACTAAACCAAAGCAAATACGA |

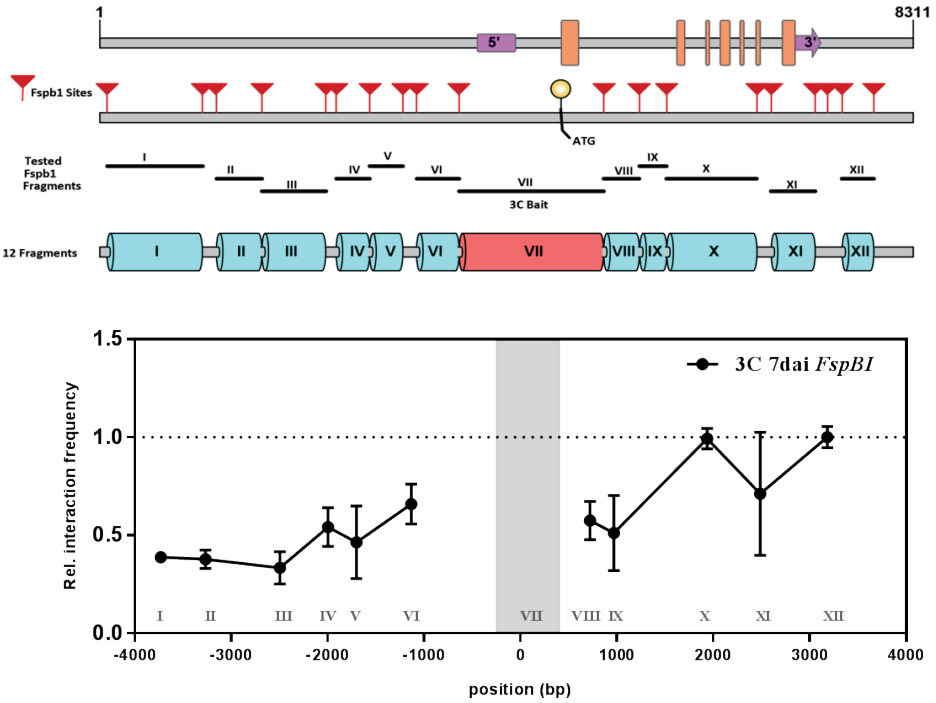
| **Table 3** | **3C primers for *Nla*III restriction profile** | | |
|---|---|---|---|
| **Name** | **Gene locus and Restriction enzyme** | **Primer on** | **Seq 5' to 3'** |
| **PDS7922** | SOC1_*Nla*III | Fragment A | ACCGTTGGATGAAAGAGCAT |
| **PDS7923** | SOC1_*Nla*III | Fragment B | CGCGTCTACAGAAAGTTAACCA |
| **PDS7924** | SOC1_*Nla*III | Fragment C | TGACCTTACCCACATAGAAACAC |
| **PDS7925** | SOC1_*Nla*III | Fragment D | GCCAAACCAACATCACAAAA |
| **PDS7926** | SOC1_*Nla*III | Fragment E | GAAAACAAAAGGAGCGAAAAA |
| **PDS7927** | SOC1_*Nla*III | Fragment F | TTTTTCCCACCCTTATTTCTC |
| **PDS7928** | SOC1_*Nla*III | Fragment G | CATTGCCCCATTGTCTCTGT |
| **PDS7920** | SOC1_*Nla*III | Fragment H | ATCCTCGAAAGCTTCCTCCT |
| **PDS7929** | SOC1_*Nla*III | Fragment I | AATCATCTGTCTCTCTCTTTCTCAA |
| **PDS7930** | SOC1_*Nla*III | Fragment J | TGAAAATGCCAGCTTTTGAT |
| **PDS7931** | SOC1_*Nla*III | Fragment K | GAGCGGTAATGAATATAACCACAA |
| **PDS7932** | SOC1_*Nla*III | Fragment L | TTGGTTATCTTCAATCATCAACCT |
| **PDS7933** | SOC1_*Nla*III | Fragment M | TGATTCTGAACTGCTTGTGTTATG |
| **PDS7934** | SOC1_*Nla*III | Fragment N | ATCCATTGGCCAAAAATCAA |
| **PDS7935** | SOC1_*Nla*III | Fragment O | GAGGCTTTTAGCCCATCAAA |

| | | | |
|---|---|---|---|
| **PDS7936** | SOC1_*Nla*III | Fragment P | CGACGTCGCACGATTTATTA |
| **PDS7939** | TIP41_Like *Nla*III | Bait-TSS | CCGGCCTAGTTTCATTTTAGTT |
| **PDS7940** | TIP41_Like *Nla*III | -1000 | CGAGCACAAATACAAAACCG |

***Supplementary Material***




**Supplemental Figure 1** 3C analysis at the *SOC1* locus seven days after the transfer of short day grown plants to long day flowering inducing conditions (7dai) on *Arabidopsis* rosette tissue using the *Nla*III restriction enzyme.




**Supplemental Figure 2** 3C analysis at the *SOC1* locus seven days after the transfer of short day grown plants to long day flowering inducing conditions (7dai) on *Arabidopsis* rosette tissue using the *Fsp*BI restriction enzyme.


### *Acknowledgements*

We kindly thank Iris Hovel and Blaise Weber for their advice on the methodology. We thank the PDS group members at Wageningen University & Research and the EpiTRAITS community, for useful ideas and fruitful discussions. We gratefully acknowledge the support from the European Commission Seventh Framework-People-2012-ITN Project EpiTRAITS, GA-316965 (Epigenetic regulation of economically important plant traits) to S.B.J, M.E.S and G.C.A.

## *REFERENCE*

**Arabidopsis Genome, I.** (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408,** 796-815.

**Ay, F., and Noble, W.S.** (2015). Analysis methods for studying the 3D architecture of the genome. Genome biology **16,** 183.

**Burley, S.K., and Roeder, R.G.** (1996). Biochemistry and structural biology of transcription factor IID (TFIID). Annual review of biochemistry **65,** 769-799.

**Cao, S., Kumimoto, R.W., Gnesutta, N., Calogero, A.M., Mantovani, R., and Holt, B.F., 3rd.** (2014). A distal CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the FLOWERING LOCUS T promoter and regulates the timing of flowering in Arabidopsis. The Plant cell **26,** 1009-1017.

**Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R.** (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nature reviews. Genetics **14,** 572-584.

**Crevillen, P., Sonmez, C., Wu, Z., and Dean, C.** (2013). A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. The EMBO journal **32,** 140-148.

**Daban, J.R.** (2011). Electron microscopy and atomic force microscopy studies of chromatin and metaphase chromosome structure. Micron **42,** 733-750.

**de Wit, E., and de Laat, W.** (2012). A decade of 3C technologies: insights into nuclear organization. Genes & development **26,** 11-24.

**Dekker, J.** (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. Nature methods **3,** 17-21.

**Dekker, J., Marti-Renom, M.A., and Mirny, L.A.** (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nature reviews. Genetics **14,** 390-403.

**Dekker, J., Rippe, K., Dekker, M., and Kleckner, N.** (2002). Capturing chromosome conformation. Science **295,** 1306-1311.

**Denker, A., and de Laat, W.** (2016). The second decade of 3C technologies: detailed insights into nuclear organization. Genes & development **30,** 1357-1382.

**Dostie, J., and Dekker, J.** (2007). Mapping networks of physical interactions between genomic elements using 5C technology. Nature protocols **2,** 988-1002.

**Dryden, N.H., Broome, L.R., Dudbridge, F., Johnson, N., Orr, N., Schoenfelder, S., Nagano, T., Andrews, S., Wingett, S., Kozarewa, I., Assiotis, I., Fenwick, K., Maguire, S.L., Campbell, J., Natrajan, R., Lambros, M., Perrakis, E., Ashworth, A., Fraser, P., and Fletcher, O.** (2014). Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. Genome research **24,** 1854-1868.

**Feng, C.M., Qiu, Y.J., Van Buskirk, E.K., Yang, E.J., and Chen, M.** (2014a). Light-regulated gene repositioning in Arabidopsis. Nature communications **5**.

**Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E.** (2014b). Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. Molecular cell **55,** 694-707.

**Fransz, P., de Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I.** (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. Proceedings of the National Academy of Sciences of the United States of America **99,** 14584-14589.

**Frohler, S., and Dieterich, C.** (2009). 3PD: Rapid design of optimal primers for chromosome conformation capture assays. BMC genomics **10,** 635.

**Fujimoto, S., Sugano, S.S., Kuwata, K., Osakabe, K., and Matsunaga, S.** (2016). Visualization of specific repetitive genomic sequences with fluorescent TALEs in Arabidopsis thaliana. Journal of experimental botany **67,** 6101-6110.

**Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., Chew, E.G., Huang, P.Y., Welboren, W.J., Han, Y., Ooi, H.S., Ariyaratne, P.N., Vega, V.B., Luo, Y., Tan, P.Y., Choy, P.Y., Wansa, K.D., Zhao, B., Lim, K.S., Leow, S.C., Yow, J.S., Joseph, R., Li, H., Desai, K.V., Thomsen, J.S., Lee, Y.K., Karuturi, R.K., Herve, T., Bourque, G.,**

**Stunnenberg, H.G., Ruan, X., Cacheux-Rataboul, V., Sung, W.K., Liu, E.T., Wei, C.L., Cheung, E., and Ruan, Y.** (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature **462,** 58-64.

**Gavrilov, A.A., Golov, A.K., and Razin, S.V.** (2013). Actual ligation frequencies in the chromosome conformation capture procedure. PloS one **8,** e60403.

**Grob, S., Schmid, M.W., and Grossniklaus, U.** (2014). Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Molecular cell **55,** 678-693.

**Grob, S., Schmid, M.W., Luedtke, N.W., Wicker, T., and Grossniklaus, U.** (2013). Characterization of chromosomal architecture in Arabidopsis by chromosome conformation capture. Genome biology **14,** R129.

**Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forne, T.** (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nature protocols **2,** 1722-1733.

**Hovel, I., Louwers, M., and Stam, M.** (2012). 3C Technologies in plants. Methods **58,** 204-211.

**Hsieh, T.H.S., Fudenberg, G., Goloborodko, A., and Rando, O.J.** (2016). Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. Nature methods **13,** 1009-+.

**Hsieh, T.H.S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J.** (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell **162,** 108-119.

**Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R.** (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nature genetics **46,** 205-212.

**Immink, R.G., Pose, 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 physiology **160,** 433-449.

**Jager, R., Migliorini, G., Henrion, M., Kandaswamy, R., Speedy, H.E., Heindl, A., Whiffin, N., Carnicer, M.J., Broome, L., Dryden, N., Nagano, T., Schoenfelder, S., Enge, M., Yuan, Y., Taipale, J., Fraser, P., Fletcher, O., and Houlston, R.S.** (2015). Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. Nature communications **6,** 6178.

**Jiang, D., and Berger, F.** (2016). Histone variants in plant transcriptional regulation. Biochimica et biophysica acta.

**Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L.** (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nature biotechnology **30,** 90-U139.

**Kim, D.H., and Sung, S.** (2017). Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. Developmental cell **40,** 302-312 e304.

**Kolovos, P., van de Werken, H.J., Kepper, N., Zuin, J., Brouwer, R.W., Kockx, C.E., Wendt, K.S., van, I.W.F., Grosveld, F., and Knoch, T.A.** (2014). Targeted Chromatin Capture (T2C): a novel high resolution high throughput method to detect genomic interactions and regulatory elements. Epigenetics & chromatin **7,** 10.

**Krivega, I., and Dean, A.** (2012). Enhancer and promoter interactions-long distance calls. Current opinion in genetics & development **22,** 79-85.

**Lajoie, B.R., Dekker, J., and Kaplan, N.** (2015). The Hitchhiker's guide to Hi-C analysis: practical guidelines. Methods **72,** 65-75.

**Lee, J., and Lee, I.** (2010). Regulation and function of SOC1, a flowering pathway integrator. Journal of experimental botany **61,** 2247-2254.

**Levine, M., Cattoglio, C., and Tjian, R.** (2014). Looping back to leap forward: transcription enters a new era. Cell **157,** 13-25.

**Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., and Dekker, J.** (2009).

Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science **326,** 289-293.

**Liu, C., Teo, Z.W., Bi, Y., Song, S., Xi, W., Yang, X., Yin, Z., and Yu, H.** (2013). A conserved genetic pathway determines inflorescence architecture in Arabidopsis and rice. Developmental cell **24,** 612-622.

**Liu, L., Adrian, J., Pankin, A., Hu, J., Dong, X., von Korff, M., and Turck, F.** (2014). Induced and natural variation of promoter length modulates the photoperiodic response of FLOWERING LOCUS T. Nature communications **5,** 4558.

**Louwers, M., Splinter, E., van Driel, R., de Laat, W., and Stam, M.** (2009a). Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C). Nature protocols **4,** 1216-1229.

**Louwers, M., Bader, R., Haring, M., van Driel, R., de Laat, W., and Stam, M.** (2009b). Tissue- and expression level-specific chromatin looping at maize b1 epialleles. The Plant cell **21,** 832-842.

**Marand, A.P., Zhang, T., Zhu, B., and Jiang, J.** (2016). Towards genome-wide prediction and characterization of enhancers in plants. Biochimica et biophysica acta.

**Marsman, J., and Horsfield, J.A.** (2012). Long distance relationships: enhancer-promoter communication and dynamic gene transcription. Biochimica et biophysica acta **1819,** 1217-1227.

**Matthews, K.S.** (1992). DNA looping. Microbiological reviews **56,** 123-136.

**Miele, A., and Dekker, J.** (2009). Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C). Methods in molecular biology **464,** 105-121.

**Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., Hale, C.J., Feng, W., Michaels, S.D., Frand, A.R., Pellegrini, M., Dekker, J., Kim, J.K., and Jacobsen, S.E.** (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science **336,** 1448-1451.

**Naumova, N., Smith, E.M., Zhan, Y., and Dekker, J.** (2012). Analysis of long-range chromatin interactions using Chromosome Conformation Capture. Methods **58,** 192-203.

**O'Sullivan, J.M., Hendy, M.D., Pichugina, T., Wake, G.C., and Langowski, J.** (2013). The statistical-mechanics of chromosome conformation capture. Nucleus **4,** 390-398.

**Patikoglou, G.A., Kim, J.L., Sun, L., Yang, S.H., Kodadek, T., and Burley, S.K.** (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes & development **13,** 3217-3230.

**Paulsen, J., Sandve, G.K., Gundersen, S., Lien, T.G., Trengereid, K., and Hovig, E.** (2014). HiBrowse: multi-purpose statistical analysis of genome-wide chromatin 3D organization. Bioinformatics **30,** 1620-1622.

**Pombo, A., and Dillon, N.** (2015). Three-dimensional genome architecture: players and mechanisms. Nature reviews. Molecular cell biology **16,** 245-257.

**Ptashne, M.** (1986). Gene regulation by proteins acting nearby and at a distance. Nature **322,** 697-701.

**Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., Marcellin, R., Melzer, R., Brown, E., Hart, D., Theissen, G., Silva, C.S., Parcy, F., Dumas, R., Nanao, M., and Zubieta, C.** (2014). Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in Arabidopsis. The Plant cell **26,** 3603-3615.

**Ramani, V., Cusanovich, D.A., Hause, R.J., Ma, W.X., Qiu, R.L., Deng, X.X., Blau, C.A., Disteche, C.M., Noble, W.S., Shendure, J., and Duan, Z.J.** (2016). Mapping 3D genome architecture through in situ DNase Hi-C. Nature protocols **11,** 59-76.

**Rapkin, L.M., Anchel, D.R., Li, R., and Bazett-Jones, D.P.** (2012). A view of the chromatin landscape. Micron **43,** 150-158.

**Ribeiro de Almeida, C., Stadhouders, R., de Bruijn, M.J., Bergen, I.M., Thongjuea, S., Lenhard, B., van Ijcken, W., Grosveld, F., Galjart, N., Soler, E., and Hendriks, R.W.** (2011). The DNA-binding protein CTCF limits proximal Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus. Immunity **35,** 501-513.

**Rodriguez-Granados, N.Y., Ramirez-Prado, J.S., Veluchamy, A., Latrasse, D., Raynaud, C., Crespi, M., Ariel, F., and Benhamed, M.** (2016). Put your 3D glasses on: plant chromatin is on show. Journal of experimental botany **67,** 3205-3221.

**Rosa, S., and Shaw, P.** (2013). Insights into chromatin structure and dynamics in plants. Biology **2,** 1378-1410.

**Sahlen, P., Abdullayev, I., Ramskold, D., Matskova, L., Rilakovic, N., Lotstedt, B., Albert, T.J., Lundeberg, J., and Sandberg, R.** (2015). Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. Genome biology **16,** 156.

**Sato, T., and Suyama, M.** (2015). ChromContact: A web tool for analyzing spatial contact of chromosomes from Hi-C data. BMC genomics **16,** 1060.

**Schleif, R.** (1988). DNA looping. Science **240,** 127-128.

**Schmid, M.W., Grob, S., and Grossniklaus, U.** (2015). HiCdat: a fast and easy-to-use Hi-C data analysis tool. BMC bioinformatics **16,** 277.

**Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W.** (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nature genetics **38,** 1348-1354.

**Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J.C., Thongjuea, S., Stadhouders, R., Palstra, R.J., Stevens, M., Kockx, C., van Ijcken, W., Hou, J., Steinhoff, C., Rijkers, E., Lenhard, B., and Grosveld, F.** (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. Genes & development **24,** 277-289.

**Splinter, E., Grosveld, F., and de Laat, W.** (2004). 3C technology: analyzing the spatial organization of genomic loci in vivo. Methods in enzymology **375,** 493-507.

**Stadhouders, R., Thongjuea, S., Andrieu-Soler, C., Palstra, R.J., Bryne, J.C., van den Heuvel, A., Stevens, M., de Boer, E., Kockx, C., van der Sloot, A., van den Hout, M., van IJcken, W., Eick, D., Lenhard, B., Grosveld, F., and Soler, E.** (2012). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. Embo Journal **31,** 986-999.

**Urnov, F.D., and Wolffe, A.P.** (2001). Chromatin remodeling and transcriptional activation: the cast (in order of appearance). Oncogene **20,** 2991-3006.

**van de Werken, H.J., Landan, G., Holwerda, S.J., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Oz, Y., Bouwman, B.A., Verstegen, M.J., de Wit, E., Tanay, A., and de Laat, W.** (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nature methods **9,** 969-972.

**Wang, C., Liu, C., Roqueiro, D., Grimm, D., Schwab, R., Becker, C., Lanz, C., and Weigel, D.** (2015). Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. Genome research **25,** 246-256.

**Weber, B., Jamge, S., and Stam, M.** (2018). 3C in Maize and Arabidopsis. Methods in molecular biology **1675,** 247-270.

**Weber, B., Zicola, J., Oka, R., and Stam, M.** (2016). Plant Enhancers: A Call for Discovery. Trends in plant science.

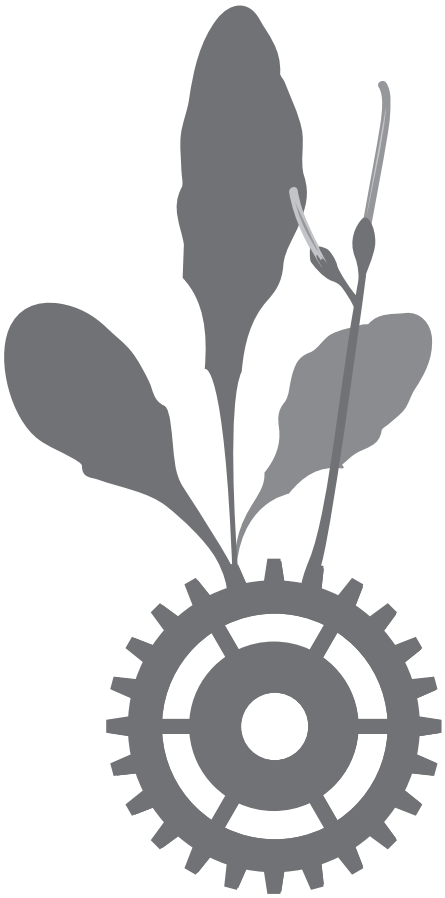
**Woo, J.W., Kim, J., Kwon, S.I., Corvalan, C., Cho, S.W., Kim, H., Kim, S.G., Kim, S.T., Choe, S., and Kim, J.S.** (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nature biotechnology **33,** 1162-1164.

**Yardimci, G.G., and Noble, W.S.** (2017). Software tools for visualizing Hi-C data. Genome biology **18,** 26.

**Yun, X., Xia, L., Tang, B., Zhang, H., Li, F., and Zhang, Z.** (2016). 3CDB: a manually curated database of chromosome conformation capture data. Database : the journal of biological databases and curation **2016**.

# CHAPTER 8

## Concluding remarks and perspectives



The precise timing of flowering is paramount for reproductive success in plants. On account of this, plants have evolved elaborate strategies to ensure they blossom at the right time. Over the years, extensive genetic research on the model plant *Arabidopsis thaliana* has resulted in a thorough understanding of how environmental and internal flowering signals, together with a complex network of gene interactions, fine-tune the robust developmental switch from vegetative to reproductive development. In this thesis, I dive deeper into the molecular mechanisms of how protein-protein and protein-DNA interactions help define the floral transition process, with a particular focus on flowering-related proteins from the MADS-domain transcription factor (TF) family. This TF family comprises of around one hundred proteins in *Arabidopsis*, and are essential for a variety of developmental process throughout the plant life cycle (Smaczniak et al., 2012a). At present, the biological role of many plant MADS-domain proteins have been elucidated by genetic studies and the majority of them were characterized with more than one biological function. How this ability of MADS-domain proteins to fulfil different developmental functions in time or place is laid down in their protein sequence, structure and function, is an open research avenue. The studies described in this thesis brings us a step closer towards a full understanding of MADS-domain protein's functional diversity.

***Order within chaos: Establishment of higher-order complexes during floral transition***

*In vivo* protein-DNA interaction studies with flowering time related MADS-domain proteins, such as SUPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1), SHORT VEGETATIVE PHASE (SVP), FLOWERING LOCUS C (FLC) and FLOWERING LOCUS M (FLM), have uncovered specific binding of these master regulatory proteins to thousands of genomic regions. However, only a small portion of these regions for a specific MADS-domain protein give rise to a transcriptional response of the neighbouring genes, which therefore got designated as their potential target genes. Among these, some targets are shared by multiple MADS-domain TFs (e.g. the potential targets of SVP and SOC1, Chapter 4). Furthermore, in many of these cases, overlap in the binding sites is observed (Chapter 4). A likely explanation for this is that MADS-domain TFs bind in the form of heterodimers or as multimeric protein complexes consisting of different MADS-domain proteins. This joint binding could increase the selectivity for protein-DNA interaction and therefore, a wide variety of unique DNA binding complexes can be formed with just a limited number of MADS TF proteins. This thesis brings forward the idea that MADS-domain TFs with different biological functions work together in complexes. Moreover, our data strengthens the notion that the MADS-domain TFs controlling flowering time act in higher-order complexes; as was previously demonstrated for floral organ identity specifying MADS proteins (Smaczniak et al., 2012b). This would suggest, that this is a more common mode-of-action for these transcriptional regulators. Besides other MADS-domain proteins, many other transcriptional regulators, co-regulators, chromatin remodellers and accessory proteins have been identified and can be part of such a higher-order complex. However, a question remaining is on the size and their exact stoichiometry of the individual and functional complexes. In this respect, we hypothesize a modular mode-of-action for the

studied MADS domain TF proteins, in which a number of different and unique protein complexes are assembled at specific developmental time points and in different tissues. Each complex possess a selective DNA binding capacity and, as such, perform a unique biological function. Our proteomics approach provides a glimpse into the intricate network of *in planta* MADS-domain protein interactions. Proteins with similar function, such as SOC1 and AGL24, should have similar interaction partners, but upon decoding the protein complex compositions of the two flowering promoters SOC1 and AGAMOUS-LIKE 24 (AGL24) (Chapter 4), we observed striking differences. During flower development, SOC1 interacts with a broad range of floral homeotic MADS-domain TFs, including SEPs, AG, and AP1. In contrast, AGL24 interactions were confined to a select number of MADS-domain TFs such as AP1 and FUL. Overall our findings indicate that SOC1 is a major hub protein in floral transition and flower development. A future avenue of research that would be of benefit to the field is the application of systems biology approaches (e.g. graph theory (Barabasi and Oltvai, 2004)) on interactome datasets accompanied by dynamic visualization of the protein interaction networks (PINs). Furthermore, comparison of the PINs from different developmental time points or tissues will allow us to categorise and identify universal versus specific protein complexes, of which the latter are supposed to be unique to a particular developmental state or biological function. At a higher level, comparison of such PINs between organisms could shed light on their conservation vs. diversification, providing insight in evolution of PINs and its possible role in species diversification (Sharan et al., 2005; Sharan et al., 2007; Dittrich et al., 2008; Safari-Alighiarloo et al., 2014).

***Opposite attracts: Function of an 'flowering activator-repressor' complex***

The experiments performed for this thesis resulted in the identification of a SOC1-SVP interaction *in planta,* suggesting the presence of these flowering time regulators in a joint protein complex (Chapter 4 and 5). These two TF proteins antagonistically regulate the flowering time process, with SOC1 acting as a activator and SVP as repressor (Hartmann et al., 2000; Lee et al., 2000; Samach et al., 2000). The finding that proteins with opposite biological roles can function in one complex was initially difficult to reconcile. As the molecular function of such an intriguing 'flowering activator-repressor' complex has never been described before, we aimed at unravelling its function by investigating its DNA binding capacity and transcriptional response of such a heterodimer in comparison to SOC1 and SVP homodimer complexes. Genome-wide examination of SOC1 and SVP binding sites do indicate certain level of interdependency, because approximately 25% overlap was found in their lists of potential targets. Upon overexpression of these two TFs in protoplasts, our preliminary results do indicated a different transcriptional response depending on whether SOC1 and SVP were present as homodimer or as heterodimer. Our observations suggest that the combinatorial activity of such complexes most likely provides the degree of plasticity necessary to govern their gene regulatory functions in the flowering time control network. Similar combinatorial activities of TF complexes in target gene regulation have been studied before (Mateos et al., 2015). When proteins such as SOC1-SVP with opposite function interact,

we believe that such interaction could be necessary for the regulation of specific target genes independent of their specific function in flowering time process. In order to confirm this belief, transcriptional regulation of more target genes will have to be tested in the presence of SOC1-SVP, but also in *svp* and *soc1* mutant backgrounds. Over the years, conventional ChIP-seq studies combined with RNA-seq expression data has resulted in identification of direct targets for individual TFs. However, inference of targets shared by multiple TFs from such studies may not be very accurate. Therefore, performing assays such as Re-ChIP combined with RNA pol II occupancy may prove valuable for our understanding of the role and function of multiple TFs in a complex. For example, Re-ChIP (also known as double ChIP, sequential ChIP, serial ChIP) enables the possibility to capture if two TFs (or chromatin marks) can bind at the particular loci in certain developmental stage simultaneously, or whether they bind individually at the same loci but at a different developmental stage or time. For example, co-localization of the BR-activated transcription factor, BZR1 and phytochrome-interacting factor 4 (PIF4) has been shown at some of the common targets with Re-ChIP (Oh et al., 2012). Although Re-ChIP technology has been around for quite some time, the application of this methodology for genome-wide localization (i.e. Re-ChIP-seq) is technically challenging (e.g. lower yield of targeted DNA, choice of antibodies, cross-linking time, etc.). As of now Re-ChIP-seq methodology has been mainly applied to study co-localization of chromatin marks in *Arabidopsis* (Luo et al., 2013; Sequeira-Mendes et al., 2014). Confronting some of the technical challenges, Zhang and co-workers applied this methodology to study genome-wide associations between ethylene signalling proteins and histone acetylation marks (Zhang et al., 2017). Similarly, application of Re-ChIP to study SOC1-SVP binding and associated with chromatin marks will allow for the identification of their common targets. Besides this, transcription regulation is known to be determined not just by TFs, but also by associating co-factors (e.g. chromatin modellers, co-activators or repressors). Our proteomics data have identified such proteins in complex with SVP and SOC1. As of now, there is no direct evidence of these co-factors associating with the SOC1-SVP heterodimer nor their impact on this heterodimers targets. It would be of interest to identify the exact stoichiometry of each SOC1-SVP complex variants and their influence on gene expression. Our knowledge of the molecular mechanism of the SOC1-SVP complex has been expanded in this study, but is still far from complete necessitating a more detailed analysis to identify targets and the transcriptional response of these targets to complex binding.

***Multi-layered regulation of flowering time***

A glance into a cell's nucleus under a microscope tells us that the genomic content is packed and folded in a non-random fashion. Ever since the discovery of enhancers, the regulation of gene expression by distantly acting genomic elements has been a common school of thought. Currently, most of our understanding of chromatin organization within a cell's nucleus comes from non-plant based studies (Gibcus and Dekker, 2013; Pombo and Dillon, 2015; Sexton and Cavalli, 2015). With the advent of 3C and its variants, the contacts between chromatin regions and DNA loops have now been directly linked to nuclear processes; providing insights into

mechanisms that were simply not possible to study a decade ago. For example, promoter-enhancer contacts have been emphasized in the regulation of gene expression and linked to development and diseases (Matharu and Ahituv, 2015; Weber et al., 2016). 3C-based studies have provided us probabilistic (and not deterministic) association of DNA structures, thereby drawing a picture of the DNA topology within a cell population. Majority of long-range chromatin interaction have been identified in larger genomes such as mammalians and maize (Bonev and Cavalli, 2016; Denker and de Laat, 2016). Reports on DNA looping and their functional relevance for smaller and dense genomes, such as *Arabidopsis* have been very limited (Liu, et al., 2015). Although the 3C method to detect local chromatin contacts (or DNA loops) is more than a decade old (Dekker et al., 2002), its application in plants is still very limited. One possible reason is the need for extensive optimizations and complex experimental design before the start of the experiment (Dekker, 2006; Louwers et al., 2009). In this thesis, we have provided detailed guidelines for the application of this method in plants and have discussed the caveats of this important yet challenging technique (Chapter 6 and 7). In *Arabidopsis*, locus-specific short distance DNA loops have been mostly studied and identified with a qPCR-based 3C method (reviewed in chapter 7, (Liu and Weigel, 2015)). A recent study has suggested a mechanism on how long noncoding RNAs (lncRNAs) can modulated gene expression through short distance looping in *Arabidopsis (Kim and Sung, 2017)*. They showed that the silencing of the floral repressor *FLC* upon vernalization is meditated by the formation of intergenic loops at the *FLC* locus, which is coordinated by the action of two lncRNAs, COLDWRAP and COLDAIR (Kim and Sung, 2017). According to this study, during the initial days of a cold period, COLDAIR recruits the repressive PRC2 polycomb complex at the intron of the *FLC* locus and forms an intergenic loop. As a consequence, nearby histone 3s are trimethylated at lysine 27 (H3K27me3). Later on the interaction of COLDWRAP enables the spread of the repressive H3K27me3 mark across the *FLC* promoter (Kim and Sung, 2017). This intriguing multi-layered regulation that, combines seasonal cues, protein-RNA interaction, and epigenetic repression of target genes, demonstrates the complexity of flowering time regulation. So far thousands of lncRNAs have been identified in *Arabidopsis* and it remains to be seen if a similar mode of regulation exists at other flowering related loci. Moving beyond 3C, in *Arabidopsis* 4C and Hi-C studies (variants of the 3C method) are now shaping our understanding of plant chromatin architecture (Grob et al., 2013; Feng et al., 2014; Grob et al., 2014; Wang et al., 2015). Currently there is no clear consensus whether the topological associated domains (TADs), i.e. chromosomes regions that show higher interaction frequencies, as described for metazoans are also prominent in *Arabidopsis* genomes (Grob and Grossniklaus, 2017). The resolution of the majority of results from 3C and its variants has been at the kilo bases or above. A new variant of 3C performed in yeast, called micro-C (Hsieh et al., 2015), relies on MNase-based digestion and takes the resolution of chromatin interactions down to the base pair level (150bp to 1kb). Application of this method in *Arabidopsis* will add a valuable tool to the current tool-box for identifying short distance loops similar to as described previously for the *FLC* locus. Combining our knowledge of chromatin-

interaction, protein-protein and protein-DNA interactions will aid in our understating of the multi-layered regulation of the flowering time process.

***Going beyond the flowering time function***

MADS-domain proteins were initially identified in yeast and mammals, and before being identified and characterized as a large and extended family of regulatory proteins in plants. Plant MADS-domain proteins have received major attention for their role in flowering time and the flower development. This is, because of the importance of these developmental processes and the distinct mutant phenotypes of the MADS-box genes involved in flowering time and floral organ identity. But there is more to MADS-domain protein functions than what first catches the eye. In recent years, a number of studies have started to explore the role of these transcriptional regulators in a variety of other developmental processes. For instance, a combination of physiological, transcriptomic and metabolomics studies resulted in the elucidation of SVP's role as a major hub in the drought-stress response (Bechtold et al., 2016). SOC1 has been described to have a role in cold tolerance, stomatal opening, plant growth and longevity, greening, dark-induced chlorophyll degradation and leaf senescence, besides its well-known function in flowering time control (Melzer et al., 2008; Liu et al., 2009; Seo et al., 2009; Richter et al., 2013; Kimura et al., 2015; Davin et al., 2016; Chen et al., 2017). The MADS-domain interactome of SOC1, AGL24 and SVP, discussed in Chapter 4 & 5 ,identified numerous significant interactions with completely different classes of proteins. To be more specific, the Nitrilase proteins NIT1 and NIT3 were significantly enriched in the SOC1 immuno-precipitations across various developmental stages (chapter 4). NIT proteins have been characterized for their role in cell proliferation (Doskocilova et al., 2013), root morphogenesis (Lehmann et al., 2017) and plant defence response (Howden and Preston, 2009). The identification of NIT proteins in complex with SOC1 and AGL24 (Chapter 4), suggests potential roles of SOC1 and AGL24 beyond their well-known flowering time regulation function. Apart from protein-protein interactions, we also studied which transcriptions factors might regulate *SOC1*'s expression (Chapter 2). We reported a large number of potential upstream regulators, for instance JUNGBRUNNEN1 (JUB1), a NAC domain TF that is involved in leaf senescence, GA/BR metabolism and negative regulation of plant defence (Shahnejat-Bushehri et al., 2012; Wu et al., 2012; Shahnejat-Bushehri et al., 2016a; Shahnejat-Bushehri et al., 2016b). It is very well possible that a number of these potential novel upstream regulators of *SOC1* and identified SOC1 interaction partners are essential for the various different development functions of SOC1. It would therefore be of interest to follow up on these potential regulators and mediators of *SOC1* functioning in the future.

***New Kid on the block: SVP as a floral pathway integrator***

The first appearance of the term floral pathway integrator in the literature was back in 2002, to describe the role of three genes, namely *SOC1*, *FLOWERING LOCUS T (FT)* and *LEAFY (LFY) (Simpson and Dean, 2002)*. According to this definition, floral integrators are genes who's expression or function is influenced or regulated by more than one flowering input signal and

transfer these integrated signals to a final flowering regulating response (Simpson and Dean, 2002). Substantial progress has since been made in identifying regulatory modules within the different flowering pathways resulting in an intricate flowering time control gene regulatory network (GRN) (Fornara et al., 2010; Bouche et al., 2016). Biochemical and genetic studies have identified *SVP* as one of the critical nodes in this flowering GRN, integrating signals from multiple pathways (Hartmann et al., 2000; Lee et al., 2007; Gregis et al., 2008; Li et al., 2008; Lee et al., 2013; Pose et al., 2013). Specifically, SVP is regulated by fluctuations in ambient-temperature and upon exposure to cold, and is involved in the autonomous, photoperiodic and GA signalling pathways (Lee et al., 2007; Fujiwara et al., 2008; Li et al., 2008; Andres et al., 2014). Allelic variation in *SVP* observed in naturally occurring populations suggests that SVP is able to fine-tune flowering time according to local environmental conditions (Mendez-Vigo et al., 2013). Our SVP interactome dataset (Chapter5) identified SVP interactions with different types of proteins, from transcription factors and co-factors to chromatin remodelling proteins, suggesting the involvement of SVP in a large number of different processes. In light of this knowledge, here we propose *SVP* as a new floral pathway integrator, joining the club of *SOC1*, *FT* and *LFY*. The major difference between SVP and the latter is that SVP acts as repressor in the flowering time control, whereas the others are flowering promoters.

**Tip of the Iceberg: the flowering protein network**

We have immuno-precipitated a large number of proteins and our data confirms several previously reported interactions as well as many novel interactions (Chapter 4 and 5). However, it is noteworthy to mention that the protein-protein interactions reported in this thesis may represent only a subset of the complete SOC1, AGL24 and SVP interactomes, due the stringency of the performed assay. On the other hand, the identified interactome is most likely contaminated by various false positives. For instance, the likelihood of co-purifying abundant proteins (e.g ribosomal) is relatively high in affinity based purification assays, but, the numerous purification steps within the protocol might have resulted in washing away weak or low-abundant interactions. The lack of standard controls and reference sets makes it difficult to decipher true interactions from false positives. Identified protein interactions therefore require validation by another independent assay. Validation assays such as yeast-two-hybrid are relatively straight forward to perform, assuming the interaction is direct. Since IP based LC-MS/MS identifies both direct and indirect interactions between proteins in a complex, reciprocal IPs or FRET-FILM are better alternatives to confirm the identified interactions *in vivo (Braun et al., 2013)*. Assays such as yeast-three-hybrid might also be feasible, but these demands prior knowledge of the bridging partners (co-factor).

The latest *Arabidopsis* genome annotation describes 27,655 protein coding loci with 48,359 transcripts and 39% of protein coding loci coding for two or more splice variants (Cheng et al., 2017). In addition, proteins undergo reversible or irreversible post transcriptionally modifications (PTMs), such as phosphorylation, which can define protein function and its interaction with other molecules. Both PTMs and alternative splicing are important biological

processes with direct relevance for flowering time control (Chapter1). For example, SVP has been shown to interact with both isoforms of FLM (FLM beta and FLM delta), and interaction with a specific isoform of FLM can either delay or induce flowering in the ambient temperature pathway (Pose et al., 2013). Our IPs with SVP identified the FLM protein, but our proteomics approach was not suitable to distinguish the specific isoform nor was aimed at protein modification detection. Moving forward, methodologies such as targeted proteomics (Schubert et al., 2017) or employing PTM enrichment strategies (Su et al., 2015) in the mass spectrometry work flow will be essential to identifying and quantifying these isoforms or modifications to fully understand the flowering regulatory network.

***Variation is all around us: Limitations of heterogeneous population based methods***

Since the advent of next-generation sequencing, the omics technologies such as genomics, transcriptomic, metabolomics and proteomics have become routine practices in biological research. Currently, we are facing a big challenge in improving the temporal and spatial resolution of the approaches. Most techniques require the use of large amounts of starting material (tissues), which usually represents a very heterogeneous population of cells. As a result, the final outcome often yields information that may not be relevant for a particular developmental stage or tissue. Gene and protein expression, when observed at a single-cell level, may appear to be stochastic in nature (Kaern et al., 2005; Raj and van Oudenaarden, 2008). For example, in the case of *FLC* regulation, upon vernalisation a gradually repression of *FLC* is observed at tissue level as well as at a whole plant level. However, at the single cell level, repression of *FLC* is binary with the gene either in an "ON" state or "OFF" state (Angel et al., 2015); this is referred to as digital repression. It is the sum of each cell's binary response that results in the gradual repression at the tissue level. Drawing observations from a heterogeneous population may therefore mask the real difference that occurs at the single cell level. Similarly, in a single cell based Hi-C study performed in vertebrates, cell-to-cell variability in chromatin interactions has been reported (Nagano et al., 2013). Such tissue-specific and cell-specific variation may also occur for protein complex formation. Therefore, the next decade of single cell omics studies in plants will be key to distinguish the variation that is random from variation that is biologically relevant.

***On the move: Subcellular localizations of MADS-domain proteins***

Moving from the single cell level to the subcellular level, here proteins play multiple roles in subcellular processes and can localize to different cellular compartments for their function (e.g cytoplasm, nucleus, ER membrane etc). Recent evidence suggests that subcellular localization of a protein or as a complex can affect its function. The subcellular function of 14-3-3 proteins have been well described in plants and includes: nuclear-cytoplasmic shuttling, binding and sequestering of TFs in cytoplasm, activation of mitochondrial enzymes, and acting as co-activators or transcriptional activators. Intercellular transport of certain MADS-domain proteins have been described during flower development and a few subcellular studies indicate localization of these TFs in different subcellular organelles (Urbanus et al., 2009).

However, very little is known about the functional dynamics of MADS-TFs in the different subcellular environments. Besides TFs, the proteomics data presented in this thesis also identified different classes of proteins including, mediatory complex subunits, chaperones proteins, protein kinases, enzymes, such as nitrilases, and chromatin modellers such as histone deacetylases. Some of these proteins are reported to have specific functions within a single subcellular environment. For example, HDA15, a class II histone deacetylase and potential interaction partner of SVP (Chapter 5), has been described to have a role in nuclear-cytoplasmic shuttling in response to light exposure (Alinsug et al., 2012). Histone deacetylase have been described to stimulate a closed chromatin configuration thereby acting as transcriptional repressor. HDA15 is associated with deacetylation of the nuclear core complex proteins and it is also speculated that it has a similar function to other class II HDAs in deacetylating non-histone proteins in the cytosol (Alinsug et al., 2012). Therefore, the investigation of interactions of MADS-domain TFs with other cellular molecules and their function in other subcellular compartments is an important research avenues that still remains to be explored.

To conclude, this thesis advances our knowledge of the protein-protein and protein-DNA interactions of MADS-domain TFs. Additional work is necessary in order to fully understand the stoichiometry of the protein complexes and how these complexes define target gene regulation *in vivo* to control flowering time. With rapid improvements in existing methodologies and development of novel techniques, we will be able to elucidate the dynamics of protein, RNA and DNA interactions at a single cell level. The data from such studies in *Arabidopsis* will serve as a basis for transferring this knowledge to crop species to steer breeding programs for the improvement of agronomical traits, such as flowering time.

***REFERENCE***

**Alinsug, M.V., Chen, F.F., Luo, M., Tai, R., Jiang, L.W., and Wu, K.Q.** (2012). Subcellular Localization of Class II HDAs in Arabidopsis thaliana: Nucleocytoplasmic Shuttling of HDA15 Is Driven by Light. PloS one **7**.

**Andres, F., Porri, A., Torti, S., Mateos, J., Romera-Branchat, M., Garcia-Martinez, J.L., Fornara, F., Gregis, V., Kater, M.M., and Coupland, G.** (2014). SHORT VEGETATIVE PHASE reduces gibberellin biosynthesis at the Arabidopsis shoot apex to regulate the floral transition. Proceedings of the National Academy of Sciences of the United States of America **111,** E2760-2769.

**Angel, A., Song, J., Yang, H., Questa, J.I., Dean, C., and Howard, M.** (2015). Vernalizing cold is registered digitally at FLC. Proceedings of the National Academy of Sciences of the United States of America **112,** 4146-4151.

**Barabasi, A.L., and Oltvai, Z.N.** (2004). Network biology: understanding the cell's functional organization. Nature reviews. Genetics **5,** 101-113.

**Bechtold, U., Penfold, C.A., Jenkins, D.J., Legaie, R., Moore, J.D., Lawson, T., Matthews, J.S.A., Vialet-Chabrand, S.R.M., Baxter, L., Subramaniam, S., Hickman, R., Florance, H., Sambles, C., Salmon, D.L., Feil, R., Bowden, L., Hill, C., Baker, N.R., Lunn, J.E., Finkenstadt, B., Mead, A., Buchanan-Wollaston, V., Beynon, J., Rand, D.A., Wild, D.L., Denby, K.J., Ott, S., Smirnoff, N., and Mullineaux, P.M.** (2016). Time-Series Transcriptomics Reveals That AGAMOUS-LIKE22 Affects Primary Metabolism and Developmental Processes in Drought-Stressed Arabidopsis. The Plant cell **28,** 345-366.

**Bonev, B., and Cavalli, G.** (2016). Organization and function of the 3D genome. Nature reviews. Genetics **17,** 772.

**Bouche, F., Lobet, G., Tocquin, P., and Perilleux, C.** (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. Nucleic acids research **44,** D1167-1171.

**Braun, P., Aubourg, S., Van Leene, J., De Jaeger, G., and Lurin, C.** (2013). Plant Protein Interactomes. Annual Review of Plant Biology, Vol 64 **64,** 161-187.

**Chen, J., Zhu, X., Ren, J., Qiu, K., Li, Z., Xie, Z., Gao, J., Zhou, X., and Kuai, B.** (2017). Suppressor of Overexpression of CO 1 Negatively Regulates Dark-Induced Leaf Degreening and Senescence by Directly Repressing Pheophytinase and Other Senescence-Associated Genes in Arabidopsis. Plant physiology **173,** 1881-1891.

**Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D.** (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. The Plant journal : for cell and molecular biology **89,** 789-804.

**Davin, N., Edger, P.P., Hefer, C.A., Mizrachi, E., Schuetz, M., Smets, E., Myburg, A.A., Douglas, C.J., Schranz, M.E., and Lens, F.** (2016). Functional network analysis of genes differentially expressed during xylogenesis in soc1ful woody Arabidopsis plants. The Plant journal : for cell and molecular biology **86,** 376-390.

**Dekker, J.** (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. Nature methods **3,** 17-21.

**Dekker, J., Rippe, K., Dekker, M., and Kleckner, N.** (2002). Capturing chromosome conformation. Science **295,** 1306-1311.

**Denker, A., and de Laat, W.** (2016). The second decade of 3C technologies: detailed insights into nuclear organization. Genes & development **30,** 1357-1382.

**Dittrich, M.T., Klau, G.W., Rosenwald, A., Dandekar, T., and Muller, T.** (2008). Identifying functional modules in protein-protein interaction networks: an integrated exact approach. Bioinformatics **24,** i223-231.

**Doskocilova, A., Kohoutova, L., Volc, J., Kourova, H., Benada, O., Chumova, J., Plihal, O., Petrovska, B., Halada, P., Bogre, L., and Binarova, P.** (2013). NITRILASE1 regulates the exit from proliferation, genome stability and plant development. The New phytologist **198,** 685-698.

**Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E.** (2014). Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. Molecular cell **55,** 694-707.

**Fornara, F., de Montaigu, A., and Coupland, G.** (2010). SnapShot: Control of flowering in Arabidopsis. Cell **141,** 550, 550 e551-552.

**Fujiwara, S., Oda, A., Yoshida, R., Niinuma, K., Miyata, K., Tomozoe, Y., Tajima, T., Nakagawa, M., Hayashi, K., Coupland, G., and Mizoguchi, T.** (2008). Circadian clock proteins LHY and CCA1 regulate SVP protein accumulation to control flowering in Arabidopsis. The Plant cell **20,** 2960-2971.

**Gibcus, J.H., and Dekker, J.** (2013). The hierarchy of the 3D genome. Molecular cell **49,** 773-782.

**Gregis, V., Sessa, A., Colombo, L., and Kater, M.M.** (2008). AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis. The Plant journal : for cell and molecular biology **56,** 891-902.

**Grob, S., and Grossniklaus, U.** (2017). Chromosome conformation capture-based studies reveal novel features of plant nuclear architecture. Current opinion in plant biology **36,** 149-157.

**Grob, S., Schmid, M.W., and Grossniklaus, U.** (2014). Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Molecular cell **55,** 678-693.

**Grob, S., Schmid, M.W., Luedtke, N.W., Wicker, T., and Grossniklaus, U.** (2013). Characterization of chromosomal architecture in Arabidopsis by chromosome conformation capture. Genome biology **14,** R129.

**Hartmann, U., Hohmann, 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 : for cell and molecular biology **21,** 351-360.

**Howden, A.J.M., and Preston, G.M.** (2009). Nitrilase enzymes and their role in plant-microbe interactions. Microbial biotechnology **2,** 441-451.

**Hsieh, T.H.S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O.J.** (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell **162,** 108-119.

**Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J.** (2005). Stochasticity in gene expression: From theories to phenotypes. Nature Reviews Genetics **6,** 451-464.

**Kim, D.H., and Sung, S.** (2017). Vernalization-Triggered Intragenic Chromatin Loop Formation by Long Noncoding RNAs. Developmental cell **40,** 302-312 e304.

**Kimura, Y., Aoki, S., Ando, E., Kitatsuji, A., Watanabe, A., Ohnishi, M., Takahashi, K., Inoue, S., Nakamichi, N., Tamada, Y., and Kinoshita, T.** (2015). A flowering integrator, SOC1, affects stomatal opening in Arabidopsis thaliana. Plant & cell physiology **56,** 640-649.

**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.

**Lee, J.H., Yoo, S.J., Park, S.H., Hwang, I., Lee, J.S., and Ahn, J.H.** (2007). Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. Genes & development **21,** 397-402.

**Lee, J.H., Ryu, H.S., Chung, K.S., Pose, D., Kim, S., Schmid, M., and Ahn, J.H.** (2013). Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. Science **342,** 628-632.

**Lehmann, T., Janowitz, T., Sanchez-Parra, B., Alonso, M.P., Trompetter, I., Piotrowski, M., and Pollmann, S.** (2017). Arabidopsis NITRILASE 1 Contributes to the Regulation of Root Growth and Development through Modulation of Auxin Biosynthesis in Seedlings. Frontiers in plant science **8,** 36.

**Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E., and Yu, H.** (2008). A repressor complex governs the integration of flowering signals in Arabidopsis. Developmental cell **15,** 110-120.

**Liu, C., and Weigel, D.** (2015). Chromatin in 3D: progress and prospects for plants. Genome biology **16,** 170.

**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.

**Louwers, M., Splinter, E., van Driel, R., de Laat, W., and Stam, M.** (2009). Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C). Nature protocols **4,** 1216-1229.

**Luo, C., Sidote, D.J., Zhang, Y., Kerstetter, R.A., Michael, T.P., and Lam, E.** (2013). Integrative analysis of chromatin states in Arabidopsis identified potential regulatory mechanisms for natural antisense transcript production. The Plant journal : for cell and molecular biology **73,** 77-90.

**Mateos, J.L., Madrigal, P., Tsuda, K., Rawat, V., Richter, R., Romera-Branchat, M., Fornara, F., Schneeberger, K., Krajewski, P., and Coupland, G.** (2015). Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in Arabidopsis. Genome biology **16,** 31.

**Matharu, N., and Ahituv, N.** (2015). Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. PLoS genetics **11,** e1005640.

**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. Nature genetics **40,** 1489-1492.

**Mendez-Vigo, B., Martinez-Zapater, J.M., and Alonso-Blanco, C.** (2013). The flowering repressor SVP underlies a novel Arabidopsis thaliana QTL interacting with the genetic background. PLoS genetics **9,** e1003289.

**Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P.** (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature **502,** 59-+.

**Oh, E., Zhu, J.Y., and Wang, Z.Y.** (2012). Interaction between BZR1 and PIF4 integrates brassinosteroid and environmental responses. Nature cell biology **14,** 802-809.

**Pombo, A., and Dillon, N.** (2015). Three-dimensional genome architecture: players and mechanisms. Nature reviews. Molecular cell biology **16,** 245-257.

**Pose, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G.C., Immink, R.G., and Schmid, M.** (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. Nature **503,** 414-417.

**Raj, A., and van Oudenaarden, A.** (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell **135,** 216-226.

**Richter, R., Bastakis, E., and Schwechheimer, C.** (2013). Cross-repressive interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the control of greening, cold tolerance, and flowering time in Arabidopsis. Plant physiology **162,** 1992-2004.

**Safari-Alighiarloo, N., Taghizadeh, M., Rezaei-Tavirani, M., Goliaei, B., and Peyvandi, A.A.** (2014). Protein-protein interaction networks (PPI) and complex diseases. Gastroenterol Hepatol Bed Bench **7,** 17-31.

**Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., and Coupland, G.** (2000). Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. Science **288,** 1613-1616.

**Schubert, O.T., Rost, H.L., Collins, B.C., Rosenberger, G., and Aebersold, R.** (2017). Quantitative proteomics: challenges and opportunities in basic and applied research. Nature protocols **12,** 1289-1294.

**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.

**Sequeira-Mendes, J., Araguez, I., Peiro, R., Mendez-Giraldez, R., Zhang, X.Y., Jacobsen, S.E., Bastolla, U., and Gutierrez, C.** (2014). The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. The Plant cell **26,** 2351-2366.

**Sexton, T., and Cavalli, G.** (2015). The role of chromosome domains in shaping the functional genome. Cell **160,** 1049-1059.

**Shahnejat-Bushehri, S., Mueller-Roeber, B., and Balazadeh, S.** (2012). Arabidopsis NAC transcription factor JUNGBRUNNEN1 affects thermomemory-associated genes and enhances heat stress tolerance in primed and unprimed conditions. Plant signaling & behavior **7,** 1518-1521.

**Shahnejat-Bushehri, S., Nobmann, B., Devi Allu, A., and Balazadeh, S.** (2016a). JUB1 suppresses Pseudomonas syringae-induced defense responses through accumulation of DELLA proteins. Plant signaling & behavior **11,** e1181245.

**Shahnejat-Bushehri, S., Tarkowska, D., Sakuraba, Y., and Balazadeh, S.** (2016b). Arabidopsis NAC transcription factor JUB1 regulates GA/BR metabolism and signalling. Nature plants **2,** 16013.

**Sharan, R., Ulitsky, I., and Shamir, R.** (2007). Network-based prediction of protein function. Mol Syst Biol **3,** 88.

**Sharan, R., Suthram, S., Kelley, R.M., Kuhn, T., McCuine, S., Uetz, P., Sittler, T., Karp, R.M., and Ideker, T.** (2005). Conserved patterns of protein interaction in multiple species. Proceedings of the National Academy of Sciences of the United States of America **102,** 1974-1979.

**Simpson, G.G., and Dean, C.** (2002). Arabidopsis, the Rosetta stone of flowering time? Science **296,** 285-289.

**Smaczniak, C., Immink, R.G., 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., Busscher-Lange, J., Dinh, Q.D., Liu, S., Westphal, A.H., Boeren, S., Parcy, F., Xu, L., Carles, C.C., Angenent, G.C., and Kaufmann, K.** (2012b). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proceedings of the National Academy of Sciences of the United States of America **109,** 1560-1565.

**Su, N., Zhang, C.P., Zhang, Y., Wang, Z.Q., Fan, F.X., Zhao, M.Z., Wu, F.L., Gao, Y., Li, Y.C., Chen, L.S., Tian, M.M., Zhang, T., Wen, B., Na, S.S., Xiong, Z., Wu, S.F., Liu, S.Q., Yang, P.Y., Zhen, B., Zhu, Y.P., He, F.C., and Xu, P.** (2015). Special Enrichment Strategies Greatly Increase the Efficiency of Missing Proteins Identification from Regular Proteome Samples. Journal of proteome research **14,** 3680-3692.

**Urbanus, S.L., de Folter, S., Shchennikova, A.V., Kaufmann, K., Immink, R.G., and Angenent, G.C.** (2009). In planta localisation patterns of MADS domain proteins during floral development in Arabidopsis thaliana. BMC plant biology **9,** 5.

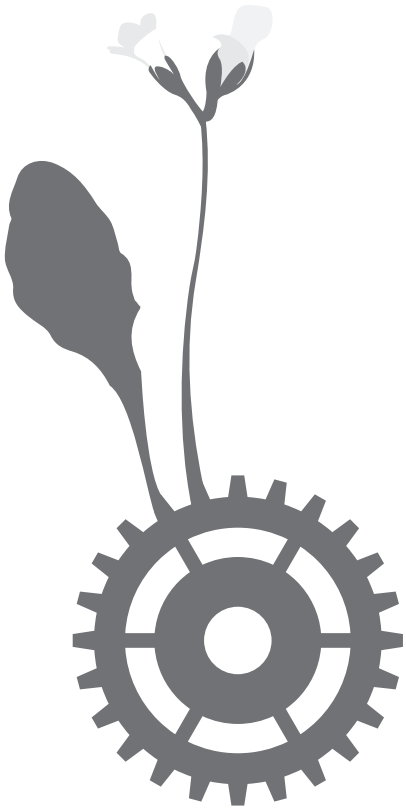
**Wang, C., Liu, C., Roqueiro, D., Grimm, D., Schwab, R., Becker, C., Lanz, C., and Weigel, D.** (2015). Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. Genome research **25,** 246-256.

**Weber, B., Zicola, J., Oka, R., and Stam, M.** (2016). Plant Enhancers: A Call for Discovery. Trends in plant science.

**Wu, A., Allu, A.D., Garapati, P., Siddiqui, H., Dortay, H., Zanor, M.I., Asensi-Fabado, M.A., Munne-Bosch, S., Antonio, C., Tohge, T., Fernie, A.R., Kaufmann, K., Xue, G.P., Mueller-Roeber, B., and Balazadeh, S.** (2012). JUNGBRUNNEN1, a reactive oxygen species-responsive NAC transcription factor, regulates longevity in Arabidopsis. The Plant cell **24,** 482-506.

**Zhang, F., Wang, L., Qi, B., Zhao, B., Ko, E.E., Riggan, N.D., Chin, K., and Qiao, H.** (2017). EIN2 mediates direct regulation of histone acetylation in the ethylene response. Proceedings of the National Academy of Sciences of the United States of America **114,** 10274-10279.

# SUMMARY



## Summary

In nature, the importance of flowers can be seen everywhere. From every fruit to every grain of rice, it all comes about as a result of flowers. Flowers do not just feed our planet, but they also fuel our imaginations, which is nicely illustrated in the famous floral paintings of Vincent van Gogh from the 1880s. With their vivid colours and complex structures, flowering plants add to the rich source of biodiversity. For plants, the decision to make flowers is in their own interest, as it guarantees having offspring and ultimately ensures the survival of the species. With this in mind, the transition to flowering is one of the most important steps in the life cycle of a flowering plant, a developmental switch that is influenced by multiple environmental and endogenous cues. The last two decades of research on the model plant *Arabidopsis thaliana* has deepened our understanding of the distinct flowering pathways and key regulators that control the floral transition process. Ever since this time, genetic studies have led to the identification of hundreds of genes and gene products that are now linked to the floral transition process, revealing the existence of an intricate network controlling flowering. This network is composed of numerous different types of proteins, protein-protein interactions, protein-DNA interactions, and chromatin states, which combined control gene activity and ultimately developmental programs. The current state-of-the-art and knowledge about the underlying mechanisms, which has led to the major research questions addressed in this thesis, are introduced in **Chapter 1**. A large number of the identified genes and regulators in the flowering network encode MADS-domain transcription factors (TF). I mainly focused on three MADS-domain TFs, namely SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), AGAMOUS-LIKE 24 (AGL24) and SHORT VEGETATIVE PHASE (SVP) and aimed at gaining insight into the multi-layered regulation of flowering time in *Arabidopsis thaliana*.

*SOC1* is a so-called floral pathway integrator and integrates endogenous and environmental signals that control the floral transition process. Besides its major importance in flowering time regulation, this TF is also involved in other developmental programs such as flower organ development, stomatal opening, plant growth and plant longevity. In order to shed light on the tight transcriptional regulation of *SOC1* related to its multifaceted role in plant development, we identified potential upstream regulators of this gene through employing a yeast-one-hybrid approach to screen more than 1,200 *Arabidopsis* TFs (**Chapter 2**). As a result, we identified a diverse array of TFs that bind to promoter regions of *SOC1*. Among them, previously characterized regulators such as CONSTANS (CO) and SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 15 (SPL15), were present. Publicly available RNA-seq data was used to perform co-expression studies between *SOC1* and its potential upstream regulators, followed by eQTL analyses. These additional *in silico* analyses resulted in the identification of a large number of confident and potential new upstream regulators of *SOC1*.

The majority of TFs interact with other proteins and perform their specialized functions in unique multimeric complexes. Yeast-based protein interaction studies indicated that MADS-

domain TFs have in general multiple, but very specific interactions. Recent advancements in protein detection methodologies and optimizations of protein immunoprecipitation (IP) techniques enabled us to study MADS-domain protein complexes from native plant tissues in a robust fashion. An optimized protocol to perform native protein complex isolation is described in **Chapter 3**, and this served as a basis for all the proteomics work described in **Chapters 4** and **5** of this thesis.

Protein complex isolation for SOC1 and AGL24 in shoot apical meristem enriched material before and after the switch to flowering resulted in the identification of various other MADS TFs as major interaction partners (**Chapter 4**). A significant overlap of AGL24 and SOC1 interactions with MADS TFs was observed in the IP data, 14 days after switching two week old plants from short-day to flowering-inducing long-day conditions, highlighting their redundant function during the early stages of flower development. Nevertheless, in inflorescence tissue, different proteins were identified as potential interaction partner of SOC1 and AGL24. Beyond the MADS TFs, other classes of proteins were identified in the SOC1 and AGL24 complexes, such as Nitrilases. Since, SOC1's interaction with the floral repressor SVP was significantly enriched in all developmental tissues and both proteins have opposite function in flowering time control, we selected this flowering activator-repressor complex to investigate its molecular function. Re-analysis of genome wide ChIP-seq studies for both TFs, showed that these TFs have common targets, where they most likely work in a complex. To confirm direct binding of these targets by a SOC1-SVP complex, *in vitro* EMSA assays were employed. In addition, transactivation assays were done in protoplasts to determine transcriptional response of the potential target genes. Our results suggest that the two TFs act in a combinatorial fashion and can influence the transcriptional behaviour of their target genes. This kind of combinatorial activity might provide a certain degree of flexibility or plasticity, which is essential to govern the gene regulatory functions in the floral transition process.

In **Chapter 5,** we switched gears and moved from flowering promoters to a flowering repressor protein. We examined the protein complex compositions of SVP. SVP functions as floral inhibitor during the vegetative stage of development, whereas during reproductive stages i.e. early flower development, it represses precocious expression of ABCE-class MADS-domain TFs. Protein complexes were isolated form these two different developmental stages to understand how SVP can fulfil these two different functions. Our SVP IP showed enrichment of specific protein interactions in both developmental tissues, but also differences in protein complex components were observed. Besides confirmation of previously reported SVP interactions, our approach also identified several potential novel SVP interaction partners, including a protein that we named POUSHP (PSH). PSH appears to be a plant specific protein, which is conserved in angiosperms. Transient expression of PSH in protoplast showed that it is mainly nuclear localised and *in silico* analyses suggest that it acts as a transcriptional activator, because of the presence of D/E-rich repeats. Detailed genetic characterization of *PSH* revealed that it positively regulates flowering time in the presence of the floral repressor SVP.

In **Chapter 6**, we described a protocol for application of Chromosome Conformation Capture (3C) in plants, which is a technique especially suited to study physical interactions of distant DNA regions, thereby providing information on the 3D chromatin organization of specific loci. Although 3C has been extensively used in other model organisms to study chromatin interactions (e.g. promoter-enhancer contacts) and their role in transcriptional gene regulation, usage of this technique in plants is very recent. The protocol we described is specifically suitable for plant tissues, especially for *Arabidopsis thaliana* and *Zea mays*. In **Chapter 7**, in addition to reviewing some of the important 3C literature in plants, we implemented 3C to study chromatin interactions at the *SOC1* locus. Unfortunately, this resulted not in reproducible results and on the basis of our experience, we describe that for meaningful interpretation of 3C data several controls and standard validation steps are required. We discuss in-depth the technicalities and intricacies associated with this promising but challenging technique.

In **Chapter 8** I discussed all the major findings of this thesis and provide guidelines for future research. Overall, this thesis provided insight into *in planta* protein-protein interactions (PPIs) of flowering time-controlling MADS-domain TFs. The multitude of biologically relevant PPIs identified for SOC1 and SVP and their involvement in various flowering pathways, strengthen the notion that both these proteins act as major hubs in the flowering time network and can be regarded as floral integrators. This thesis reports several novel interaction partners of MADS-domain TFs, which represent potential new players in the flowering time controlling gene regulatory network, such as POUSHP. Nevertheless, the exact stoichiometry of the identified protein complexes and their impact on gene regulation needs further study, which is an interesting and important research topic to follow-up in the near future.

## Samenvatting

Overal in de natuur is zichtbaar hoe belangrijk bloemen zijn. Elke vrucht en elke rijstkorrel is het resultaat van een bloem. Bloemen voeden niet alleen onze planeet, maar ze voeden ook onze verbeelding, zoals goed te zien is in de beroemde bloemenschilderijen van Vincent van Gogh van rond de jaren 1880-90. Met hun levendige kleuren en complexe vormen zijn bloeiende planten een rijke bron van biodiversiteit. Voor planten is het besluit om bloemen te maken in hun eigen belang, omdat het nakomelingen garandeert en daarmee uiteindelijk het voortbestaan van de soort. Met dit in gedachte kan de overgang tot bloei als de meest belangrijke stap in de levenscyclus van de plant worden beschouwd, een ontwikkelingsschakelaar die beïnvloed wordt door verschillende omgevings- en interne factoren. In de afgelopen 20 jaar heeft onderzoek aan de modelplant *Arabidopsis thaliana* ons begrip van verschillende routes naar bloei, en van de sleutelregulatoren die de overgang naar bloei sturen, verdiept. In die tijd hebben genetische studies geleid tot de identificatie van honderden genen en genproducten die nu met de overgang naar bloei verbonden zijn, en daarmee het bestaan van een ingewikkeld bloeiregulatienetwerk blootgelegd. Dit netwerk is samengesteld uit vele verschillende eiwitten, eiwit-eiwitinteracties, eiwit-DNA interacties, en chromatinetoestanden, welke samen genactiviteit en uiteindelijk ontwikkelingsprogramma's sturen. De huidige stand van zaken in dit onderzoeksveld en de kennis van de onderliggende mechanismen die hebben geleid tot de belangrijkste onderzoeksvragen die worden behandeld in dit proefschrift,, worden in **hoofdstuk 1** uitgelegd. Een groot aantal van de regulators in het bloeiregulatienetwerk behoort tot de familie van MADS-domein transcriptiefactoren (TFs). We hebben ons vooral geconcentreerd op drie MADS-domein transcriptiefactoren, SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1), AGAMOUS-LIKE 24 (AGL24) en SHORT VEGETATIVE PHASE (SVP), en richtten ons op het verkrijgen van inzicht in de veel-gelaagde regulatie van bloeitijd in *Arabidopsis thaliana*.

*SOC1* is een zogenaamde bloeiroute-integrator welke endogene en omgevingssignalen die de overgang naar bloei sturen, combineert. Naast zijn belangrijke rol in bloeitijdregulatie is deze transcriptiefactor ook betrokken bij andere ontwikkelingsprocessen zoals bloemorgaanontwikkeling, opening van huidmondjes, en groei en levensduur van planten. Om inzicht te krijgen in de regulatie van *SOC* zelf in relatie tot zijn veelzijdige rol in plantontwikkeling hebben we regulators van *SOC* transcriptie gezocht door middel van een yeast-one-hybrid strategie en het testen van 1,200 *Arabidopsis* transcriptiefactoren (**Hoofdstuk 2**). Dit resulteerde in de identificatie van een grote verscheidenheid aan transcriptiefactoren die aan de promoter van *SOC1* binden. Daaronder bevonden zich eerder beschreven regulators zoals CONSTANS (CO) en SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 15 (SPL15). Publiekelijk beschikbare RNA-seq datasets werden gebruikt voor het bestuderen van co-expressie van *SOC1* en zijn mogelijke transcriptieregulators, gevolgd door eQTL-experimenten. Deze toegevoegde *in silico* analyses leidden tot de identificatie van mogelijke nieuwe transcriptieregulators van *SOC1* met hoge betrouwbaarheid.

Eiwit complex isolatie voor SOC1 en AGL24 in scheutmeristeem-verrijkt materiaal voor en na de schakeling naar bloei resulteerde in de identificatie van verschillende andere MADS transcriptiefactoren als interactie-partners (**hoofdstuk 4**). een significante overlap van AGL24 en SOC1 interacties met MADS transcriptiefactoren werd geobserveerd, veertien dagen nadat twee weken oude planten van korte dag, naar bloei-inducerende, lange-dag condities overgezet waren, wat benadrukt dat deze eiwitten een redundante functie hebben tijdens de vroege stadia van bloemontwikkeling. Niettemin werden er verschillen gevonden in potentiele interactiepartners van SOC1 en AGL24 in materiaal van bloeiwijzes. Naast de MADS transcriptiefactoren werden ook andere klassen van eiwitten geïdentificeerd in de SOC1 en AGL24 complexen, zoals nitrilases. Omdat SOC1's interactie met de bloeirepressor SVP significant verrijkt was in weefsel van alle verschillende ontwikkelingsstadia, en beide eiwitten een tegenovergestelde functie hebben in bloeitijdregulatie, selecteerden we dit bloei 'activatie-repressie' complex voor verdere karakterisatie van de moleculaire functie. Heranalyse van genoomwijde ChIP-seq studies voor beide transcriptiefactoren laat zien dat deze transcriptiefactoren gemeenschappelijke targets hebben, waar ze hoogstwaarschijnlijk in één complex aan binden. Om deze directe binding van een SOC1-SVP complex aan targets te bevestigen werden *in vitro* EMSA assays gebruikt. Daarnaast werden transactivatie assays gedaan in protoplasten om de transcriptionele reactie van de potentiele targetgenen te bepalen. Onze resultaten suggereren dat de twee transcriptiefactoren in een combinatorische manier werken on transcriptie van de target genen te beïnvloeden. Dit soort combinatorische activiteit kan in een zekere mate van flexibiliteit of plasticiteit voorzien, wat essentieel is voor genregulatie in de schakeling naar bloei.

In **hoofdstuk 5**, veranderen we van richting en gaan we van bloei-promotoren naar een bloei-repressor eiwit. We bestudeerden de composities van eiwit-complexen die SVP bevatten. SVP fungeert als een bloei-repressor tijdens de vegetatieve ontwikkeling, terwijl tijdens de reproductieve stadia, bijvoorbeeld vroege bloem-ontwikkeling, SVP voorkomt dat de ABCE-klasse MADS-domein transcriptiefactoren te vroeg tot expressie komen. Eiwitcomplexen werden geïsoleerd uit weefsel in deze twee verschillende ontwikkelingsstadia om te begrijpen hoe SVP deze twee verschillende functies kan vervullen. Onze SVP immunoprecipitaties lieten verrijking zien van specifieke eiwit-interacties in beide ontwikkelingsstadia, maar ook verschillen in eiwit-complex componenten werden gevonden. Naast het bevestigen van eerder gevonden SVP interacties, heeft onze aanpak ook geleid tot de identificatie van verschillende nieuwe potentiele interactiepartners van SVP, waaronder een eiwit dat we POUSHP (PSH) genoemd hebben. PSH lijkt een plant-specifiek eiwit te zijn, wat geconserveerd is in angiosperms. Transiënte expressie van PSH in protoplasten toonde aan dat dit eiwit voornamelijk gelokaliseerd is in de celkern, en *in silico* analyses suggereren dat dit eiwit werkt als een transcriptie-activator, vanwege de aanwezigheid van D\E-rijke repeats. Gedetailleerde genetische karakterisatie van PSH laat zien dat dit eiwit bloeitijd positief beïnvloed als de bloeirepressor SVP aanwezig is.

In **Hoofdstuk 6** beschrijven we een protocol voor de toepassing van Chromosome Conformation Capture (3C) in planten. Deze techniek wordt gebruikt om de fysieke interacties van verafgelegen stukken DNA te bestuderen, en geeft hiermee informatie over de 3D organisatie van specifieke loci. Hoewel 3C al veel gebruikt is in andere modelorganismen om interacties van chromatine (bv promoter-enhancer contacten) en hun rol in de regulatie van gentranscriptie te bestuderen, wordt deze techniek pas recentelijk toegepast in planten. Het protocol dat wij beschrijven is specifiek geschikt voor weefsels van planten, voornamelijk *Arabidopsis thaliana* en *Zea mays*.

In **Hoofdstuk 7** bespreken we de belangrijkste literatuur over 3C in planten, en passen we 3C toe op het *SOC1* locus. Helaas waren onze resultaten niet reproduceerbaar en we geven aan, op basis van onze ervaring, welke controles en standaard validatiestappen nodig zijn om een zinvolle interpretatie van 3C data te kunnen geven. We gaan diep in op de technische aspecten en moeilijkheden van deze veelbelovende maar uitdagende techniek.

In **Hoofdstuk 8** bespreek ik de belangrijkste bevindingen van dit proefschrift en geef ik richtlijnen voor toekomstig onderzoek. In zijn geheel geeft dit proefschrift inzicht in *in planta* eiwit-eiwit interacties van bloeitijd-regulerende MADS-domein transcriptiefactoren. De hoeveelheid biologisch relevante eiwit-eiwit interacties die gevonden zijn voor SOC1 en SVP en hun betrokkenheid bij verschillende 'bloei-routes', dragen bij aan het idee dat deze eiwitten de hoofdschakelaars zijn in het bloeitijd-netwerk en dat ze beschouwd mogen worden als 'bloei-integratoren'. Dit proefschrift rapporteert een aantal nieuwe interactie-partners van MADS-domein transcriptiefactoren, zoals POUSHP, die mogelijk nieuwe spelers in het bloeitijd-regulerende genregulatienetwerk zijn. Desalniettemin moet de precieze stoichiometrie van de geïdentificeerde eiwitcomplexen en hun invloed op genregulatie beter bestudeerd worden, wat een interessant en belangrijk onderwerp is voor toekomstig onderzoek.

# ACKNOWLEDGEMENTS

## A personal note of thanks …

As I begin to write this final part of my dissertation, I find myself looking back in time and re-witnessing some of the most exciting years of my life. The journey of the last four years (and a bit) has been a tremendous learning experience. Besides developing into scientific and research skills, this journey has shaped me deeply on a very personal level. Therefore, in this last piece of writing, I take the opportunity to reflect and thank all the wonderful people whom I met on this PhD journey.

Foremost I owe a big gratitude to all my mentors. Without their support and encouragement I may have never witnessed this moment. There is only one thought that comes to my mind and I like to recite the first two lines of guru stotram (in Sanskrit) to express my gratitude:

गुरुर्ब्रह्मा गुरुर्विष्णुर्गुरुर्देवो महेश्वरः ।
गुरुरेव परं ब्रह्म तस्मै श्रीगुरवे नमः ॥१॥

Here, I especially like to thank my two gurus with whom I embarked on this journey. I want to thank my thesis promoter, Professor. G.C. Angenent (or simply **Gerco** - as he prefers), for opening the very first door and accepting me as your student. The opportunity to work in your group and at the same time being part the EpiTRAITS ITN network has been very unique. Gerco, I cannot thank you enough for all the scientific discussion, your valuable feedback on different projects and your input while drafting this dissertation. In addition to acquiring scientific knowledge, I have learned a great deal of leadership and interpersonal skills from you. Cluster PDS is a perfect embodiment of all the special things you do. Heel erg bedankt!

I sincerely like to thank my daily supervisor and my co-promoter Professor. R.G.H Immink. Dear **Richard**, I have been very lucky to have you as my daily mentor. Just like any other journey, this PhD journey was not all peaches and cream. There have been many roadblocks and every time you helped me face them with strength. Without your motivation and trust I would have never seen the light at the end of the tunnel. Under your guidance and supervision, I tend to believe that I have left the safety of my nest (comfort zone) and learned to fly on my own. Although there's still a lot of turbulence in the flight, I believe it will only get better. I truly admire your passion and vision in the scientific pursuits. I just hope I acquired one or two during my tenure. It was a real honour to work with you. Bedankt dat je altijd voor me klaar stond !

Next in line, I like to thank the two beautiful ladies, who happen to also be my paranymphs and are supporting me to get across the finish line. **Froukje**, I don't know where to begin to describe our bond. From day one, you have been my search engine. No matter what doubts or questions I had or how busy you were, you always are ready to help with a magic wand (smile). I learned a great deal of skills from you (including the IPs). Even during the final phase

of writing, you kept our spirits high with all tasty treats and banana breads. **Suzanne**, in spite the fact that for the first two years you relocated to Germany, we still keep seeing each other at so many occasions. Our passion to travel and organizing fun events further led to strengthen our friendship. You are the exception to the Dutchness, I always wished most of my other Dutch friends could be as spontaneous as you. I'm so happy to have you both as my paranymphs and feel very content that you both will be at my side through the Final Hour.

Now, I extend my thanks to the beautiful minds at the cluster PDS. Ruud, Kim, Steven and Martijn thank you for all the discussion and feedback over the years. It was always a pleasure to discuss and obtain your critical insights on my project. Besides this I'm glad that we have also connected outside the work environment. In **Ruud**, I found a great and very reliable squash buddy, which I certainly will miss. We discussed science not just in formal settings, but also off court after intense meltdown sessions. Thank you for all the encouragement and advice during this journey. **Kim**, you are a very inspiring and charismatic scientist (a role model to look up to). You are also the only person I know in Netherlands who truly understands what Indian cuisine is all about. It has always been a pleasure to interact with you. **Steven**, you introduced me to the amazing wonders of life underwater, a hobby that has developed into a passion. You set a perfect example of how to balance science and art together. **Martijn**, besides your endless and useful tips on cloning the mission impossible *UNK2* gene, I really admire your playful persona and the confidence you resonate. You are one of your kind and just like *UNK2* maybe unclonable and irreplaceable in the cluster PDS. I will also miss the snacking Fridays and officially now the top spot is up for grabs.

I would like to thank each and every past and present members of the **PDS cluster**. I will try my best to not miss out on anyone. I begin with a very important person, **Michiel**, a.k.a the man with the KEY. He is also the one who is always cheerful and will remain forever 28. Even on a bad day you lighten up the mood. The lab radio is only bearable when you start your own gig. Without you, the small lab would be boring and dull. Sorry for bothering you with so many ordering requests, sometimes I even called you on the weekends and you came to my rescue. Thanks for your unwavering support and fun memories in the lab. Also thanks to **Jacqueline** for moving into our small lab and making it even more happening. Now, I like to thank the two angels who bound the of PDS cluster together- **Mieke & Tjitske**. I will certainly miss the sinterklaas borrel and the amazing lab outings. Thank you for organizing all these special events. And special thanks to **Martijn** for pushing the wet activities at those occasions. **Rumyana**, thank you for all the science related discussion during the cluster meetings and non-work related stories at coffee corner. Now, I thank the odd man out at the Friday snacks day- **Jan**. Thanks for all the statistics related help and discussions with planning the large scale flowering time experiments. Most importantly, I like to thank the "real master" of the PDS lab and the one who makes sure everything runs smooth both online and offline, Thank you **Marco**!

Now I want to thank the open work space (a.k.a "**My WUR Buddies**"). First, my lab bench mate and someone who is kinder than kind- **Alice**. I have to admit being proactive, is a virtue I learnt from you. Thanks for all your help both in and out of the lab. I could rely on your advice on almost anything. **Marian**, yet another remarkable (and inspiring) person who was around for troubleshooting and critical discussions when things were in a frenzy. **Anneke** (one desk to my right), thanks for your scientific advice and by the way I learned a lot just observing you (super organized and focused even in open work space). **Cezary**, I'm grateful for your hands-on training on analysis of proteomics data. In addition, you introduced me to completely new sports which has become a routine. Your competitive spirit is something I admire and follow both on and off court. I look forward to our one-on-one encounters in future. **Hilda**, one stop for all the know-hows from EMSAs to Dutch traditions. Thanks to you my very first EMSA gels were publication ready (still have to publish– minor detail). I thoroughly enjoyed the gourmets, sinterklass games you introduced to the internationals. Talking about Dutch traditions, Dutch etiquettes are best learnt from Uncle **Sam**. Hallo Sir, I was so happy when you joined the group (six months later), being the only male PhD had its own minuses. For a change, I felt less left out during the gossips and dinner evenings. Since we were in the same year, we totally understood each other. You are a great companion at work and outside. Thank you for being around and I'm very glad your band will perform at the party. And best of luck with the final phase. **Leonie** (one desk to the left from my old spot), it was a pleasure to have your company during the Friday beer hour (me with my tea cup), also the discussion within our flowering time group were very useful. **Francesca** (my student), you joined our lab at a very crucial time of my PhD journey. It was nice to have you when a tremendous amount of work was ongoing and it was equally fun to work with you. Thank you for your hard work that went into genotyping and counting all those leaves. I wish you the best for the next steps of your career.

As I look around the open work space and I see all new faces since the time I started. It's time to thank these new blood who are carrying on the spirit of PDS cluster - **Manjunath, Vera, Rufang, Baojian, Lena, Mengfan, Han, Charlotte, Ellen, Wilma, Xiaobing, Christine, Elwin and et al.** It's amazing to have you guys and work alongside you. It was nice to see your active participation in and outside work activities. I think the Lunch club is a real success (Vera ;) ). It will be equally hard to say goodbye to you all. I wish each one of you the very best for the upcoming challenges ahead of you.

Special thanks to **Team Europe-** Patricia, Alba, Nicolay, Tom and Gian. I'm happy that we met and I look forward to seeing you soon. Our fun memories (and encounters) are best to leave outside this text. Honourable mention and gazillions thanks and hugs to **Team Brazil**, Diego, Jose, Greice, Dennis, Fabia, Eveline, Sylvia and et al. You all visited the PDS cluster for a short duration, yet we had great connection during this time. I was happy that I could visit your home in brazil and even present some of my PhD work at the University of Sao Paolo ( Jose's lab). Thank you for all the amazing time and remember it's a small world and our paths will cross again!

Also a big thanks to my **Proteomics Team-** Twan , Jan Co. and Sjef. Thanks for all your help with the MS/MS runs and lengthy discussions on the data analysis. Most of our discovery proteomics were a big success. For me, the questions we could address with targeted proteomics were even more exciting and it had a lot of potential. But, as we have seen, these methodologies are challenging and demanding. I hope for the best in the continuation of these projects. **Aalt-Jan**, thank for helping out with the RNA-seq and co-expression analysis and our discussions. **Biosciences** members I thank you all for the fruitful discussion over the years and for the yearly outings and X-mas dinners (which I managed to eventually attend).

**HANA,** the heart and soul of the entire Bioscience. You deserves a big applause for all the help you do for the new and old members of the group. A big thank you and warm hug for always helping us out with almost anything related with administration. I'm equally grateful to **Maria** and **Marie-Jose** for their assistance for all official help.

Now I move outside the Bioscience and reach out to thank my **EpiTRAITS family** which made this journey a really unique one. **Maike,** thank you for establishing this network. Without this ITN, I cannot imagine meeting and getting to know so many incredible scientists across this continent. We fellows are very lucky to be part of this consortium and equally happy for all the collaborations that worked out (including the publications and dissemination activities). I like to specially thank the big shots, Gerco, Maike, Manuel, Daniel, Franziska, Paul F, Valerie, Pawel K, Philippe and Charlie for all the discussions. Thanks to the advisory board members for guiding us during our PCDP meetings and a huge applause for one and only **Helen.**

**EpiTRAITS Fellows**, we all embarked on this journey together, and without a doubt (and you will agree) our ITN is the best thing that could happen to any of us. I personally enjoyed the company of each one of you. A few words about you from the three plus years we had together. **Blaise,** (Gulia and Matildi a.ka chandani), while working on chromatin looping, might sound cliché our lives looped into each other. **Stefi**, the most adorable friend and you are full of joy. **Dorota**, the girl with so much creativity & style. **Pawel**, the man from another galaxy, there's never a dull moment when you are in the room. Mitra, our little adventure with Hrishi is something we will never forget. **Till**, our storyteller, you enlighten us with all the science and non-science facts. **Jean-louis**, the young and bright mind, with a big appetite for almost everything. It was fun working with you in Wageningen. **Johan**, our critical expert and the one who keeps track we are in time everywhere. **Rurika**, very shy and quiet, yet mighty with her super powers (bioinfo). **Javi**, our navigator, finding the best places no matter in which part of the world we were. **Dimitris**, our entertainer, after all the long day of discussions and meetings. **Tom**, I better leave you with one word – 'bollocks' ☺. I'm glad I joined this journey with you guys and some of you even ended up at my home in India. I'm confident that each one of you will climb great heights (and some of you already started) in the next step of your careers. All the best EpiTRAITS fellows !!

I like to thank the **EPS PhD Council and WPC** members (and those that helped us) it was really great to work besides you. We managed to introduce a lot of new activities and initiatives for the PhD students, not just at Wageningen University, but across the Netherlands. I was fortunate to be part of some of them right from the drafting to the finalizing phase. Special thanks to Tanya, Kris, Hanna, Setareh, Masha, Mannos, Amalia, Sara, Francesca, Magda, Lot, Lotte, Tom and Jordi. Douwe, Ingrid, Claudius and Ton thank you for encouraging and supporting the PhD councils. A big cheers for Ria, working behind the scenes with all the emailing.

Now it's time to thank the friends and people who I met outside work. These are the people who made my stay in Wageningen even more special after a long day of lab work. Let start with **Marianna & Marc** (M&Ms), these are the two wonderful souls that happened upon me during my stay here. No matter the country you are in or the comforts you have around you, staying away from home and loved ones is always difficult. But you both being my neighbours (and landlords - minor detail) made it less difficult. I will see what my parents have to say about the proposal of you both adopting me. No matter where I move from here, CW18 will always be my home away from home and I'm happy to call you my Dutch parents for a change and my time here was Gezellig!!

**Team Hoevestein-** Balaji, Neeraj, Tanya, Abhiroop, Aarti, Sindhuja, Liana, Jana, Nitin, Surabhi, Raka, Somya and et al (jazz club). Special thanks to Balaji and Neeraj, you have been an amazing support and we literally can be in each other's shoes. By the way, Hoevestein 243 is without doubt the finest Veg kitchen in town and the time spent in here is something I will certainly miss. Thank you all ! Now I thank, **Team India.** Trupti, my go to desi friend to call and ask for khaana khanne aajoa mein?. You, Pranesh and Pranjali always welcomed me at your home and were my little Indian family in Wageningen. Priyanka, since you moved to the north, I no longer run into a familiar face while biking on Wageningen streets. Bandan, Gurnoor, Sandeep, Sindhu and Sneha, yet another gang of bandits. It was a joy with all those midnight birthday parties. Wishwas ( and Bhagyashree) thanks to all of your invites for the events outside Wageningen. One of them even led to us to meet our Prime Minister. Thank You!

Another very important group of friends belong to **Team Wageningen -** Jochem, Viet, Edgar(big brother), Madelon, Sacha, Anne, Julius, Nadia, Jarst, Kaleb, Aline, Rita, Casper, Rony, Delano, Daniel, Diego, Carlos, friends at WOWV and et al. From day trips, game nights, lekker avondeten, bongerd action, condition training, to the climbing weekends in Germany, each one of you made my time inhere much more exciting. Also a shout-out is also well deserved for the friends I made and a new passion I found for meeples, tiles and dices over the last year. Sam, Irene, Erik, Mark and Matt thank you for inviting me over for all the evenings at your home (and @Thuis). It was real fun to let go of the thesis pressure (for a while) and enter the realms of betrayal, revenge and victory for a change. Thanks to you all, when someone says to me nothing much happens in this little town, I say otherwise. I just wished we had known each other way earlier so I could also make some trips to Essen and

come back with some amazing stuff. Irene, I eagerly look forward to resuming the GOT club in 2019.

**Team Germany -** Siyang, Cas, Sebas, Lize, Daniel, Eve, Iuliia and Ganesh. We stayed in touch even after the master's program. It was a real pleasure to witness and be part of those important and special days in your lives. Our journey together only gets better. There is definitely a lot of good karma going around, that's why I have you on my side, even on the day of my defence. I look forward to more of our reunions and meetups. All the love and best of luck with the adventures ahead.

Special thanks to **Ajit** mama for suggesting the Sanskrit name *POUSHP*, for the gene I have described in Chapter 5 of this thesis. I'm also very grateful to **Bhagyashree** (Soni), for all the bioinformatics related help and assistance with data analysis. Now, I can proudly say we have a bioinformatician in our family. You just embarked on your own PhD journey and I truly believe it's going to be an outstanding one. Lots of love and success on the way.

I like to thank Ruud, Anneke and Suzanne for translation of the summary in Dutch and Richard for the final revision. A big thanks to Tom B, for his valuable feedback while proofreading this thesis. Here, I also like to acknowledge the erudite members of my **thesis committee** and thank them for the opportunity to defend my research work. Most importantly, I want to extend my thanks to **Wageningen University & Research**, an incredible organization that provides a pleasant and progressive working environment for students and researchers alike.

Finally, now it's time to acknowledged the people for whom I literally fall short of words - **my Family**. स्वतःच्या कुटुंबापासून इतके दिवस लांब राहणं हाच एक मोठा संघर्ष आहे. माझ्या कल्पनेपेक्षाही जास्त अडचणींचा सामना माझ्या या कुटुंबाने केला. या सगळ्या काळात तुमच्या आधारानं, माझ्यावरच्या विश्वासानं मी आत्त्ता पर्यंतची वाटचाल केली.

**मम्मी,** मी जे काही करतो, भविष्यात जे काही मला करायचं आहे त्यासाठी तू माझं प्रेरणास्थान आहेस. **पपा,** तुम्ही मला शिकवलंत, नुसती स्वप्न बघायची नाहीत तर, वास्तवात आणायची. या स्वप्नपूर्तीच्या दिशेनं आत्त्ता कुठे माझी वाटचाल सुरु झाली आहे. मला माहित आहे कि माझी डिग्री, शिक्षण या छोट्या छोट्या मिळकती आहेत पण, म्हनतात नं, " पिक्चर तो अभी बाकी है ....."

मी तिथे असायला हवं होतं पण, मी नव्हतो तिथे. **आई,** मी इथे तुला खूप मिस केलं. **भैय्या**, या जामगे परिवाराची काळजी घेणारा, प्रेमळ मोठा भाऊ आहेस. आणि माझ्या सगळ्या बहिणी - **हिराताई, मीराताई, नंदाताई, सीमा, सोनी, गीता "बाई", सायली , सिमरन** तुम्ही सगळ्यांनी हे कुंटुंब पूर्ण, केलंत. **सौरभ, आकाश, रोहित** तुम्ही आपल्या परिवारचे "आँखों का तारा" आहेत.

माझ्या सर्व काकी-काका, मावशी-मामा,भावजी, आजी-आजोबा, माझे सर्व (आप्तेष्ट) आप्त, इष्ट, माझे गावाकडचे मित्र या सगळ्यांसाठी माझ्या मानत कृतज्ञता आहे, माझं स्वप्न साकारण्या साठी मला बळ

दिलंत म्हणून. आता माझ्या गावाकडे कौडगावकडे वळतो. सध्या माझी स्थिती हि "एकला चलो रे" अशी आहे. मी आशा करतो की, गावातले विदयार्थी स्वतःच्या चिकित्सक, संशोधक होतकरू वृत्त्तीचा वापर आपल्या देशाच्या भल्यासाठी करतील.

पक्षी कितीही लांब गेला तरी त्याला घरी जायची ओढ असतेच. शेवटी एवढंच म्हणेन,

**"ने मजसी ने परत मातृभूमीला,**

**सागरा, प्राण तळमळला, सागरा."**

Thank you / धन्यवाद,
Suraj

**About the Author**

Suraj Jamge was born on August 9th 1987, in Gangakhed, Parbhani, India. In 2003, he completed his 10th Grade from Sanjeewan Vidyalaya, Panchgani. Later on, he continued his higher secondary education in the Science faculty at S.P. College, Pune. In 2005, he moved to Aurangabad to study a four years bachelor's program specializing in Agricultural Biotechnology. After graduating with a Bachelor's Degree of Science from Marathwada Agricultural University, he moved to Germany to pursue a Master's program in Agrobiotechnology. He performed his master thesis in the group of Prof. K. H. Kogel and Dr. Patrick Schäfer, at the Institute of Phytopathology and Applied Zoology, Giessen. During this thesis, his research focused on the identification of proteins that are participating in the initiation and execution of the endoplasmic reticulum stress-induced cell death process.

Shortly thereafter, he acquired an Erasmus fellowship and visited Gregor Mendel Institute of Molecular Plant Biology, Vienna to perform a six months internship in the group of Dr. Ortrun Mittelsten Scheid. Here, his research aimed at understanding the epigenetic control of plant retrotransposons under heat stress. In 2013, he obtained his Master's Degree of Science from Justus-Liebig-Universität, Giessen and immediately thereafter, he moved to the Netherlands to continue his scientific interest in plant sciences in form of a PhD. He joined as a PhD student in the group of Prof. Gerco C. Angenent at Wageningen University & Research. His PhD project was part of the EU Marie-Curie-ITN network named EpiTRAITS (Epigenetic regulation of economically important plant traits). In his project, he studied molecular mechanisms of MADS-domains proteins in control of the floral transition process and the results obtained during the course of his PhD are presented in this thesis.

## Publications

**Jamge, S.**, Angenent, G. C., & Bemer, M. (2018). Identification of In Planta Protein-Protein Interactions Using IP-MS. Methods Mol Biol, 1675, 315-329.

Weber, B., **Jamge, S.**, & Stam, M. (2018). 3C in Maize and Arabidopsis. Methods Mol Biol, 1675, 247-270.

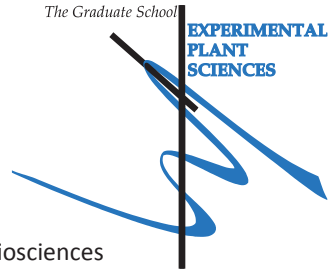
**Jamge, S.**, Stam, M., Angenent, G. C., & Immink, R. G. H. (2017). A cautionary note on the use of chromosome conformation capture in plants. Plant Methods, 13, 101.

Bey, T., **Jamge, S.**, Klemme, S., Komar, D. N., Le Gall, S., Mikulski, P., et al. (2016). Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 - 15, 2016 - Strasbourg, France. Epigenetics, 11(8), 625-634.

Cavrak, V. V., Lettner, N., **Jamge, S.**, Kosarewicz, A., Bayer, L. M., & Mittelsten Scheid, O. (2014). How a retrotransposon exploits the plant's heat stress response for its activation. PLoS Genet, 10(1), e1004115.

Deshpande B.D., Deshpande Kshama G., & **Jamge S.** (2010). Studies on enzymatic degradation of sugarcane tops. BIOINFOLET, 7(1), 0973-1431.

# Education Statement of the Graduate School Experimental Plant Sciences




**Issued to:** Suraj B. Jamge

**Date:** 9 February 2018

**Group:** Laboratory of Molecular Biology & Business Unit Biosciences

**University:** Wageningen University & Research

| 1) Start-up phase | *date* |
|---|---|
| ► **First presentation of your project** | |
| *Title:* Role of chromosome conformation in regulation of flowering genes | 28 Sep 2013 |
| ► **Writing or rewriting a project proposal** | |
| *Title:* The role of MADS-box TFs in chromosomal interactions during flower development | Aug 2013 |
| ► **Writing a review or book chapter** | |
| Identification of In Planta Protein-Protein Interactions Using IP-MS, Marian Bemer and Celia Baroux (eds.), Plant Chromatin Dynamics: Methods and Protocols, Methods in Molecular Biology, vol. 1675, DOI 10.1007/978-1-4939-7318-7_18, Springer Science+Business Media LLC 2018 | Dec 2017 |
| 3C in Maize and Arabidopsis, Marian Bemer and Celia Baroux (eds.), Plant Chromatin Dynamics: Methods and Protocols, Methods in Molecular Biology, vol. 1675, DOI 10.1007/978-1-4939-7318-7_15, Springer Science+Business Media LLC 2018 | Dec 2017 |
| ► **MSc courses** | |
| ► **Laboratory use of isotopes** | |
| *Subtotal Start-up Phase* | *6.0 credits** |

| 2) Scientific Exposure | *date* |
|---|---|
| ► **EPS PhD student days** | |
| EPS PhD Student day, Leiden, NL | 29 Nov 2013 |
| EPS PhD Student Days Get2Gether 2015, Soest, NL | 29-30 Jan 2015 |
| EPS PhD Student Days Get2Gether 2016, Soest, NL | 28-29 Jan 2016 |
| ► **EPS theme symposia** | |
| EPS Theme 1 Symposium 'Developmental Biology of Plants', Wageningen, NL | 24 Jan 2014 |
| EPS Theme 1 Symposium 'Developmental Biology of Plants', Leiden, NL | 08 Jan 2015 |
| EPS Theme 1 Symposium 'Developmental Biology of Plants', Wageningen, NL | 21 Jan 2016 |
| ► **National meetings (e.g. Lunteren days) and other national platforms** | |
| 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 |
| Annual meeting 'Experimental Plant Sciences', Lunteren, NL | 11-12 Apr 2016 |
| ► **Seminars (series), workshops and symposia** | |
| *Workshops* | |
| Workshop 3D regulation, Amsterdam, NL | 18, 20, 22 & 23 Mar 2013 |
| EU workshop Plant Chromatin, Madrid, Spain | 29-30 Aug 2013 |

| | | |
|---|---|---|
| | *Seminars* | |
| | Caroline Dean, Green Life Science Seminar, Chromatin and antisense transcript dynamics regulating the switch to flowering | 14 Apr 2014 |
| | Eric Schranz, Influences of Genome and Gene duplications | 04 Nov 2014 |
| | Ortrun Mittlestein-Schied, Genetics and Epigenetics: a complex relationship | 19 Nov 2014 |
| | Geroge Coupland, EPS flying seminar, Seasonal flowering in annual and perennial plants | 19 Jan 2015 |
| | Martin Kater, Mining for floral Meristem Regulatory Pathways in Arabidopsis and Rice | 11 Mar 2015 |
| | Marcelo C. Dornelas, Using the non-model genus Passiflora to study the evolution of novelty in plant reproductive development | 27 Jan 2015 |
| | Francois Parcy, An integrated structural biology approach to flower development | 15 Oct 2015 |
| ▶ | **Seminar plus** | |
| ▶ | **International symposia and congresses** | |
| | EpiTRAITS Start-up Meeting, Amsterdam, NL | 21 Mar 2013 |
| | EpiTRAITS Annual Meeting 2014, Wageningen, NL | 25-26 Mar 2014 |
| | EpiTRAITS Annual Meeting 2015, Colonge, Germany | 30-31 Mar 2015 |
| | Workshop on molecular mechanisms controlling flower development, Aiguablava, Spain | 07-11 Jun 2015 |
| | EpiTRAITS Annual Meeting 2016, Strasbourg, France | 14-15 Jan 2016 |
| | EpiTRAITS Final Meeting 2017, Amsterdam, NL | 27 Sep 2016 |
| ▶ | **Presentations** | |
| | *Poster:* A role for chromosome looping in flowering time control? | 25 Mar 2014 |
| | *Talk:* Finding the missing links in flowering time control by floral integrators, Wageningen, NL | 25 Mar 2014 |
| | *Talk:* Mid Term Review EpiTRAITS progress report | 08 Jul 2014 |
| | *Talk:* MADS domain proteins in floral transition, Cologne, Germany | 31 Mar 2015 |
| | *Flash Talk & Poster:* Unravelling the protein-complex composition of the floral repressors SVP, before and after the switch to flowering, Aiguablava, Spain. | 07 Jun 2015 |
| | *Flash Talk & Poster:* Decoding the protein-complex configuration of key flowering time regulators, before and after the switch to flowering, Strasbourg, France | 14 Jan 2016 |
| | *Talk;* A proteomics based approach to study flowering time control in Arabidopsis, Lunteren, NL | 12 Apr 2016 |
| | *Talk:* MADS TFs in Flowering Time Control, Amsterdam, NL | 27 Sep 2016 |
| ▶ | **IAB interview** | |
| ▶ | **Excursions** | |
| | *Subtotal Scientific Exposure* | *20.1 credits** |

Continued on next page

| 3) In-Depth Studies | *date* |
|---|---|
| ► **EPS courses or other PhD courses** | |
| Image analysis & spatial modelling, INRA, Paris, France | 08-11 Jul 2013 |
| Mathematical Modelling for Biologists, CPIB, University of Nottingham, UK | 09-12 Sep 2013 |
| Bioinformatics I, University of Amsterdam, NL | 31 Mar-01 Apr 2014 |
| Next Generation sequencing, Key Gene, Wageningen, NL | 31 Mar-03 Apr 2014 |
| ► **Journal club** | |
| Literature discussion at the PRI-PDS cluster | 2013-2017 |
| ► **Individual research training** | |
| Chromosome conformation capture (3C) technique, University of Amsterdam, NL | May-Jul 2013 |
| Practical Maize Breeding, Somlice HR, Poland | May 2015 |
| Immunolabeling-FISH, University of Amsterdam, NL | Nov 2015 |
| *Subtotal In-Depth Studies* | *9.9 credits** |

| 4) Personal development | *date* |
|---|---|
| ► **Skill training courses** | |
| Plant Breeding, IP & Marketing, Fijnaart, NL | 19 Mar 2013 |
| Project management, Amsterdam, NL | 25-26 Mar 2013 |
| Entrepreneurship and business skills, Nottingham, UK | 13 Sep 2013 |
| Ethics in practice, Wageningen, NL | 24 Mar 2014 |
| Course Presentation and Communication Skills, Amsterdam, NL | 20-21 Mar 2014 |
| Standardisation of biological data, Amsterdam, NL | 16-17 Sep 2014 |
| Entrepreneurship: business plan, Amsterdam, NL | 18-19 Sep 2014 |
| Dissemination Skills, Amsterdam, NL | 22-23 Sep 2014 |
| Personal branding via social media, Amsterdam, NL | 24 Sep 2014 |
| Entrepreneurship III : Business Game. Düsseldorf, Germany | 23 Mar 2015 |
| Scientific writing I: Writing Papers and Theses in the Life sciences, Düsseldorf, Germany | 24-25 Mar 2015 |
| Scientific writing II: Writing Grant Applications, Düsseldorf, Germany | 26-27 Mar 2015 |
| Workshop: Setting up a business, Köln, Germany | 01 Apr 2015 |
| ► **Organisation of PhD students day, course or conference** | |
| PhD movie screening 1 & 2, Wageningen, NL | 08 Jan 2016 |
| EpiTRAITS Open Conference, Strasbourg, France | 14-15 Jan 2016 |
| Get2Gether, PhD Student Retreat, Soest, NL | 28-29 Jan 2016 |
| ► **Membership of Board, Committee or PhD council** | |
| Member EPS PhD Council | May 2014-Apr 2015 |
| President of the EPS PhD Council | May 2015-Apr 2016 |
| Member Wageningen PhD Council (WPC) | May 2015-Apr 2016 |
| PhD Representative EpiTRAITS ITN | Mar 2015-Feb 2016 |
| *Subtotal Personal Development* | *12.0 credits** |

| **TOTAL NUMBER OF CREDIT POINTS*** | **48.0** |
|---|---|

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 described in this thesis was financially supported by a European Commission Seventh Framework-People-2012-ITN Project EpiTRAITS, GA-316965 (Epigenetic regulation of economically important plant traits).

Financial support from Wageningen University & Research for printing this thesis is gratefully acknowledged.

On the Cover: Environmental and internal stimuli are integrated at the level of floral integrators thereby allowing the plant to decide when to flower (Front). In response to the stimuli, action of several molecular machineries together aid the plant in making this developmental switch (back).

Cover design by the author and Fenna Schaap
Printed by proefschriftmaken.nl