Distinct impact of CURLY LEAF and SWINGER on the Arabidopsis Histone H3 Lysine 27 trimethylation pattern is linked to the underlying genetic code

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Theodoros Zografou
aus Athen

Köln, Mai 2013

Die vorliegende Arbeit wurde am Max-Planck-Institut für Züchtungsforschung in Köln in der Abteilung für Entwicklungsbiologie der Pflanzen (Direktor Prof. Dr. G. Coupland) angefertigt.

Prüfungsvorsitzender:

Prof. Dr. Martin Hülskamp

Berichterstatter:

Prof. Dr. George Coupland

Prof. Dr. Ute Höcker

Tag der Disputation: 27. Juni 2013

Contents

Abstract iv

Zusammenfassung vi

1 Introduction 1

1.1 Mechanisms and concepts of Polycomb Group mediated transcriptional control . . . 1

1.2 The Polycomb Repressive Complexes and their composition in Arabidopsis . . . 3

1.3 Plant Polycomb Repressive Complex 2 proteins in an wider evolutionary context . . . 5

1.4 Polycomb Repressive Complex 2 recruitment to target site . . . 6

1.5 Function of the Polycomb Repressive Complex 2 and its central output, trimethylation at Histone H3 Lysine 27 . . . 8

1.6 Genome wide distribution of trimethylated H3K27 in Arabidopsis . . . 11

1.7 The Arabidopsis histone-methyltransferase proteins CURLY LEAF and SWINGER . . . 11

1.8 EMBRYONIC FLOWER 2 -, VERNALIZATION RESPONSE 2 - and FERTILIZATION INDEPENDENT SEED 2 - Polycomb Repressive Complexes 2 and their role in Arabidopsis development . . . 13

1.9 Transcriptional repression of *FLOWERING LOCUS C* . . . 14

2 Aims of the study 17

3 Materials and Methods 18

3.1 Phylogenetic analysis . . . 18

3.1.1 Extraction of sequences . . . 18

3.1.2 Calculation of evolutionary distances and protein alignment . . . 18

3.2 Biomolecular analysis . . . 18

3.2.1 Plasmid constructions and plant transformation . . . 18

3.2.2 Transgenic Arabidopsis lines . . . 19

3.2.3 Plant growing conditions . . . 20

3.2.4 Protein abundance determination . . . 20

3.2.5 mRNA abundance determination . . . 21

3.2.6 Chromatin immunoprecipitation 21
3.2.7 Barcoded library preparation and sequencing 23
3.2.8 Tukey test 24
3.2.9 Buffers, Solutions, Medium 25
3.3 Computational analysis 26
3.3.1 Tools and data used 26
3.3.2 Mapping short reads to Col genome 26
3.3.3 Identification of H3K27me3-enriched regions and genes 27
3.3.4 Clustering analysis of gene expression data 27
3.3.5 Chromosomal distribution calculation of genes 27
3.3.6 Identification and confirmation of conserved sequence motifs . . 27

4 Results 28
4.1 Phylogenetic analysis of CLF and SWN 28
4.1.1 The diversification of CLF and SWN is conserved in the plant kingdom 28
4.1.2 CLF, SWN and MEA are highly similar but differ in the middle part 32
4.2 Chromatinimmunoprecipitation of PRC2 proteins 35
4.2.1 CLF-HA abundance is low and probably mainly regulated post transcriptionally 35
4.2.2 Binding of CLF-HA to AG and FLC is weak but is statistically significant compared to loci were no enrichment is expected . . . 37
4.2.3 ChIP-seq of GFP-CLF 42
4.2.4 Proteasome inhibitor degradation enhances slightly the abundance of CLF but ChIP efficiency of GFP:CLF does not increase . 46
4.3 H3K27me3 profiles of *clf* and *swn* mutants 50
4.3.1 High overlap between H3K27me3 targeted genes in *clf*, *swn* and wt 50
4.3.2 *clf*, *swn* and wt have different H3K27me3 distribution patterns . 52
4.4 Analysis of CLF dependent genes 58
4.4.1 A small fraction of H3K27me3 target genes is significantly H3K27me3 depleted in *clf*. 58
4.4.2 CLF and SWN function in combination or independently from each other as activators and repressors of CLF dependent genes. 59
4.4.3 CLF dependent genes are distributed in clusters across the chromosomes 62
4.4.4 CLF dependent genes are enriched for Telo boxes. 65

5 Discussion **70**
5.1 Phylogenetic relationships between CLF SWN and MEA 70
5.2 Attempts to ChIP on PRC2 proteins 73
5.2.1 CLF is a low binding to chromatin is at the detection limit of ChIP-PCR 73
5.2.2 CLF abundance is regulated at the post-translational level 74
5.2.3 ChIP-PCRs should be normalized against an internal control . . . 76
5.3 H3K27me3 profiles of *clf* and *swn* mutants 77
5.3.1 Three types of H3K27me3 patterns 77
5.4 Analysis of CLF dependent genes 80
5.4.1 CLF and SWN regulate gene expression together, independently, alone or antagonistically 80
5.4.2 CLF dependent genes are distributed in clusters across the chromosomes 81
5.4.3 Telo boxes are associated with CLF dependent genes 82
5.4.4 A new model of CLF and SWN function in Arabidopsis 84

Abbreviations **107**

6 Appendix **108**

Abstract

Polycomb Group Proteins (PcG) are a conserved class of transcriptional repressors in animals and plants, controlling the expression of hundreds of genes. Specifically, they repress gene expression by histone H3 trimethylation at lysine 27 (H3K27me3) via the Polycomb Repressive Complex 2 (PRC2). In *Arabidopsis thaliana* CURLY LEAF (CLF), a PRC2 methyltransferase, is partially redundant with the closely related protein SWINGER (SWN). The phenotype of *clf* null mutant plants is less severe than the phenotype of *clf/swn* double mutants. Plants that are mutated in *swn* exhibit no obvious phenotype. The large degree of sequence conservation of both CLF and SWN throughout the plant kingdom, suggests that they have distinct and conserved functions in plant development.

I set out to identify the individual target genes of CLF and SWN and to uncover their functions. To this end, I used chromatin immunoprecipitation followed by deep sequencing to create genome-wide H3K27me3 profiles of wild type, *clf* and *swn* seedlings. The three genotypes share many H3K27me3 target regions but they exhibit partially distinct methylation profiles. In *clf* null mutants, H3K27me3 is significantly reduced at 643 genes, indicating that at these genes CLF is the major catalytic component of PRC2. A part of the CLF dependent genes is significantly physically clustered along chromosomes and surprisingly, 60% of CLF dependent genes contain a conserved AAACCCTA or telo box which has been previously reported to enhance transcription. Unexpectedly, reduction of H3K27me3 in *clf* mutant plants does not correlate frequently with an increase of gene expression. The *swn* mutation can either have the same effect, oppose or contribute to the effect of *clf* on the expression of these genes. Moreover, in *swn* null mutants, H3K27me3 seems to be slightly increased in 294 genes compared to wild-type. These results suggest that the interplay between CLF and SWN might be important to establish wild-type-like H3K27me3 and expression levels. Finally, in *swn*, expression of some CLF dependent genes changes dramatically, although H3K27me3 is unaffected. This indicates that SWN might have an H3K27me3 independent function.

In short, these results demonstrate that CLF and SWN have partially distinct effects on the *Arabidopsis thaliana* H3K27me3 pattern and that their function is linked to the underlying genetic code. SWN might control expression in a H3K27me3 independent pathway and the interplay between CLF and SWN is presumably important to maintain establish wild-type-like H3K27me3 and expression levels.

Zusammenfassung

Polycomb Group Proteins (PcG) sind eine konservierte Klasse von Transkriptionsrepressoren in Tieren und Pflanzen, die die Expression von hunderten von Genen steuern. Insbesondere unterdrücken sie die Genexpression durch Histon H3 an Lysin-Trimethylierung 27 (H3K27me3) über den *Polycomb Repressive Complex 2* (PRC2). In *Arabidopsis thaliana* ist CURLY LEAF (CLF), eine PRC2-Methyltransferase, teilweise redundant zu dem eng verwandten Protein SWINGER (SWN). Der Phänotyp der *clf*-Nullmutante ist weniger deutlich ausgeprägt als der Phänotyp der *clf/swn*-Doppelmutanten. Pflanzen in denen *swn* mutiert ist zeigen keinen offensichtlichen Phänotyp. Das große Maß an Sequenzkonservierung sowohl bei CLF als auch SWN im ganzen Pflanzenreich legt nahe, dass sie klare und konservierte Funktionen in der Entwicklung von Pflanzen haben.

Das Ziel dieser Arbeit ist es die einzelnen Zielgene von CLF und SWN zu identifizieren und ihre Funktionen aufzudecken. Zu diesem Zweck habe ich chromatin immunopräzipitation gefolgt von *deep sequencing* verwendet um genomweite H3K27me3 Profile des Wildtyps zu erstellen, sowohl für *clf*- als auch *swn*-Setzlinge. Die drei Genotypen haben viele H3K27me3 ähnliche Zielregionen, weisen jedoch teilweise unterschiedliche Methylierungsmuster auf. In *clf*-Nullmutanten ist H3K27me3 deutlich an 643 Genen reduziert, was darauf hinweist, dass bei diesen Genen CLF die zentrale katalytische Komponente von PRC2 ist. Ein Teil der CLF-abhängigen Gene ist signifikant räumlich geclustert entlang der Chromosomen und überraschenderweise enthalten 60 % der CLF-abhängigen Gene eine konservierte AAACCCTA oder *telo box*, die, wie an anderer Stelle berichtet wurde, die Transkription steigern. Unerwarteter Weise korreliert die Reduktion von H3K27me3 in *clf* mutierten Pflanzen nicht mit einer Erhöhung der Genexpression. Die *swn* Mutation kann den selben Effekt haben, der Expression dieser Gene entgegenwirken oder zu ihr beitragen. Darüber hinaus scheint H3K27me3 in *swn*-Nullmutanten im Vergleich zum Wildtyp an 294 Genen leicht erhöht zu sein. Diese Ergebnisse deuten darauf hin, dass das Zusammenspiel zwischen CLF und SWN wichtig sein könnte, um das dem Wildtyp entsprechendes H3K27me3 und Expressionsniveau zu erhalten. Zuletzt kann festgestellt werden, dass sich in *swn* die Expression einiger von CLF abhängiger Gene stark verändert, obwohl H3K27me3 nicht verändert ist. Dies deutet darauf hin, dass SWN möglicherweise eine von H3K27me3 unabhängige Funktion hat.

Zusammenfassend zeigen diese Ergebnisse, dass CLF und SWN teilweise unter-

schiedliche Auswirkungen auf die H3K27me3-Muster von *Arabidopsis thaliana* haben und das ihre Funktion verknüpft ist mit dem zugrunde liegenden genetischen Code. SWN könnte die Expression in einem H3K27me3 unabhängigen Pathway kontrollieren. Das Wechselspiel zwischen CLF und SWN ist vermutlich wichtig, um das dem Wildtyp entsprechende H3K27me3- und Expressionsniveau zu erhalten.

1 Introduction

1.1 Mechanisms and concepts of Polycomb Group mediated transcriptional control

Polycomb Group (PcG) proteins are evolutionary conserved master regulators that navigate a plethora of cellular processes during development. They act mainly to maintain cell differentiation status by repressing genomic programs that are needed only at specific and temporally restricted moments (Hennig & Derkacheva, 2009; Margueron & Reinberg, 2011). PcGs are used as model systems to study epigenetic mechanisms and transcriptional regulation operating in most eucaryotic organisms including unicellular algae, vascular plants, insects and mammals. PcG proteins were first identified in *Drosophila melanogaster* (Drosophila) as nuclear proteins which target homeotic genes and regulate transcription by remodeling chromatin. Importantly, they are distinct from transcription factors which regulate transcription by binding to specific DNA sequences and activating or blocking the recruitment of RNA Polymerase II (RNAPII). Mutations in PcG proteins can result in uncoordinated pattern development and early lethality in Drosophila, mice and *Arabidopsis thaliana* (Arabidopsis). Moreover, mutations in PcG complexes have been associated with aggressive cancers (Ben-Porath *et al.*, 2008). Genome wide studies illustrated that PcG proteins are involved in the control of 5 - 15 % of genes in multicellular eukaryotes, including master regulators of developmental processes (Margueron & Reinberg, 2011; Schuettengruber & Cavalli, 2009). Thus, PcG proteins are conserved transcriptional regulators which maintain transcriptional states.

The two main Drosophila components of PcG are the Polycomb Repressive Complexes 1 (PRC1) and 2 (PRC2). In Drosophila each PRC1 and PRC2 component is represented by one or two proteins. Interestingly, PcG components in mammals and plants have evolved into small protein families (Farrona *et al.*, 2008). Moreover, they probably have undergone sub- and neofunctionalization. PcG complexes act in diverse compositions to regulate distinct developmental processes. They act through different mechanisms to control different steps during transcription at different positions of the gene body (Czermin *et al.*, 2002; How Kit *et al.*, 2010; Margueron *et al.*, 2008; Schmitges *et al.*, 2011).

The hallmark of the PcG machinery is the trimethylation of histone 3 at lysine 27 (H3K27me3) which is deposited by the PRC2 complex (Cao *et al.*, 2002; Czermin *et al.*,

2002). The core PRC1 function is the recognition of H3K27me3 via the chromodomain protein Polycomb (PC) (Cao & Zhang, 2004). PRC1 proteins catalyze the monoubiquitylation of histone H2A at lysine 119 (H2AK119ub). The canonical PcG model suggests a mechanism in which PRC1 is recruited by the deposition of H3K27me3 through PRC2 to catalyze H2AK119ub and confer gene repression (Margueron & Reinberg, 2011). PcG mediated gene repression is antagonized by Trithorax Group (TrxG) proteins which deposit H3K4me3, H3K36me2 and H3K36me3 marks. These marks are associated with active transcription (Pien & Grossniklaus, 2007).

Recent reports have illustrated that PRC2 and PRC1 are also important components of H3K27me3 and H2AK119ub independent mechanisms (Eskeland *et al.*, 2010; Francis *et al.*, 2004; Xu *et al.*, 2012). The two complexes can act independently from each other and PRC1 can also recruit PRC2 (Simon & Kingston, 2009; Tavares *et al.*, 2012). Moreover, PRC2 subunits confer H3K27me3 independent transcriptional repression through chromatin compaction but also PcG independent transcriptional activation (Margueron *et al.*, 2008; Xu *et al.*, 2012). Finally, there is differential modulation of the catalytic activity of PRC2 in animals which depends on the composition of the complex (Schmitges *et al.*, 2011). Thus, in the current model, PcG complexes and their subunits are implicated in a variety of mechanisms to control transcription (Simon & Kingston, 2013).

Nevertheless, in most cases, both PRC2 and PRC1 act together to achieve gene repression. While the role of PRC2 is to initiate, coordinate and define the PcG mode of action, PRC1 seems be important for the stabilization of the chromatin compaction.

Between animals and plants, PcG proteins are highly conserved and form similar complexes (Köhler *et al.*, 2003b; Wood *et al.*, 2006). However, in Arabidopsis, the function and biochemical properties of PcG proteins are less clear, partially because there is lack of established biochemical approaches and specific antibodies. Nothing is known about the genome wide occupancy of PRC2 proteins but binding to single genes that are marked with H3K27me3 has been demonstrated together with a reduction of H3K27me3 in PRC2 mutants (see below Bouyer *et al.* (2011); Farrona *et al.* (2011); Schubert *et al.* (2006); Zhang *et al.* (2007b)). The allocation of the PRC1 complex protein LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) highly correlates with domains that are marked by H3K27me3 and gene repression is less stable when LHP1 function is abolished (Mylne *et al.*, 2006; Turck *et al.*, 2007). Although it is assumed that the basic mechanisms between plants and animals are similar it remains unclear how plant PcG proteins actually function.

It is an open question whether histone modifications trigger differences between chromatin states, or if the differences in modifications are mainly consequences of dynamic processes, such as transcription and nucleosome remodeling. Therefore, the mechanism through which PcG complexes confer transcriptional repression are not

fully understood. The precise functional outputs of PcG, H3K27me3 and H2AK119ub remain unclear (Henikoff & Shilatifard, 2011; Simon & Kingston, 2013).

Moreover, it is not yet fully understood how PcG is recruited to target genes. In Drosophila, Polycomb Response Elements (PRE) are the major recruiters of the PcG machinery. In mammals and plants multiple components, including non coding RNAs (ncRNA), proteins and DNA elements operate probably partially in cooperative fashion to correctly target PcG complexes (see section 1.4, Margueron & Reinberg (2011)).

In the following, the composition and evolutionary context of PRC2 proteins will be described. In addition the recruitment and function of PRC2 proteins as their key catalytic output, the H3K27me3 mark will be highlighted. Major focus will be invested in the possible functions of different PRC2 complexes and the Arabidopsis histone-methyltransferase (HMT) proteins CURLY LEAF (CLF) and SWINGER (SWN). Finally, PcG mediated repression in plants will be highlighted on the example of *FLOWERING LOCUS C* (*FLC*).

1.2 The Polycomb Repressive Complexes and their composition in Arabidopsis

The evolutionary conserved PRC2 core complex in Drosophila is composed of four components (Fig. 1.1): 1. Enhancer of Zeste (E(z), Ezh2 in mammals), a histone methyltransferase with a SET domain 2. Suppressor of Zeste (Su(z)12, Suz12 in mammals), a DNA/Protein binding C2H2 Zn-finger protein 3. Extra Sex Combs (ESC, Embryonic Ectorderm Development (Eed) in mammals), a protein with a WD40 beta-propeller 4. Nucleosome remodeling factor 55 (N55, Retinoblastoma binding protein (RbpA) -46 / -48 in mammals), a WD40 domain protein (Adrian *et al.*, 2009; Farrona *et al.*, 2008; Schwartz & Pirrotta, 2007).

In multicellular plants the PRC2 proteins are conserved in small families, which associate in different compositions to target different loci (Pien & Grossniklaus, 2007; Schwartz & Pirrotta, 2007). The homologues of E(z) are CURLY LEAF (CLF), SWINGER (SWN) and MEDEA (MEA) while EMBRYONIC FLOWER 2 (EMF2) , REDUCED VERNALIZATION RESPONSE 2 (VRN2) and FERTILIZATION-INDEPENDENT SEED 2 (FIS2) are related to Su(z)12 (Goodrich *et al.*, 1997; Chanvivattana *et al.*, 2004; Grossniklaus *et al.*, 1998; Yoshida *et al.*, 2001; Gendall *et al.*, 2001; Luo *et al.*, 1999). The WD-40 protein ESC is encoded by FERTILIZATION INDEPENDENT ENDOSPERM (FIE) (Ohad *et al.*, 1999). Finally, N55 has 5 homologues named MULTICOPY SUPPRESSOR OF IRA 1 to 5 (MSI1-5) (Ach *et al.*, 1997; Hennig *et al.*, 2003; Kenzior & Folk, 1998).

The Drosophila PRC1 complex comprises Polycomb (PC), Polyhomeotic (PH), Posterior Sex Combs (PSC), RING and Sex Combs Extra (SCE). PC is the founding member of PcG proteins and mediates binding to H3K27me3 via its chromodomain, while the

catalytic activity to modify lysine 119 at histone H2A is conferred by the RING E3 Ubiquitinligases SCE and RING. PH is a Zinc-finger protein while SCM is characterized by a sterile alpha motif (SAM) domain (Müller & Kassis, 2006; Pien & Grossniklaus, 2007; Vandamme *et al.*, 2011). The PRC2 complex in Arabidopsis was discovered in 1997, however, until recently it was unclear if PRC1 also exists in plants because there seemed to be no real homologues (Farrona *et al.*, 2008). Novel discoveries proposed a PRC1 like complex composed out of AtRING1a, AtRING1b, LHP1, AtBMI1a, AtBMI1b and EMF1 (Bratzel *et al.*, 2010; Chen *et al.*, 2010).

Two further complexes, PHO repressive complex (Pho-RC) and Polycomb repressive deubiquitinase (PR-DUB) were characterized in Drosophila (Klymenko *et al.*, 2006; Scheuermann *et al.*, 2010). These complexes have not been identified in Arabidopsis.

In short, the Drosophila PRC2 complex is composed of four core components while in mammals and plants the PRC2 complex components have evolved into small families. The homologues of the four core components are optionally interchangeable.

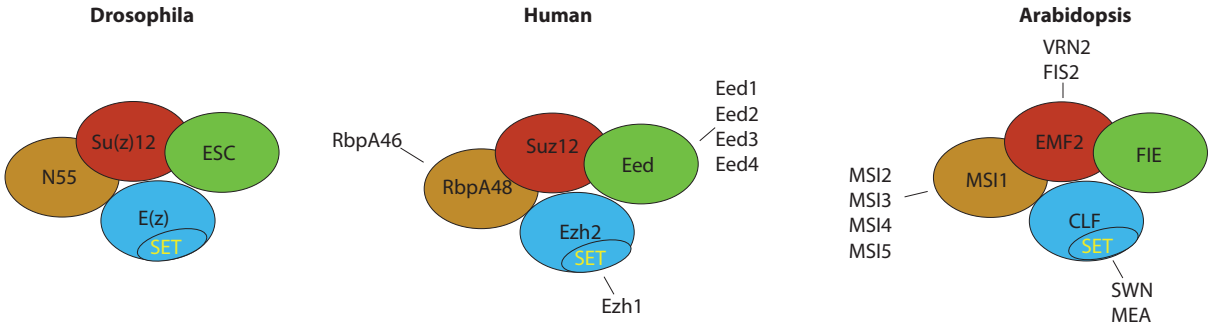


Figure 1.1: **PRC2 complex core components in Drosophila, human and Arabidopsis.** Schematic representations.
Drosophila: The Drosophila PRC2 complex is composed of four core components: Enhancer of Zeste (E(z)); Suppressor of Zeste (Su(z)12) ; Extra Sex Combs (ESC); Nucleosome remodeling factor 55 (N55).
Human: The human PRC2 complex components have evolved into small families. Homologues of the four core components are optionally interchangeable: Enhancer of Zeste (Ezh)1/2; Supressor of Zeste 12 (Suz12); Embryonic Ectoderm Development (Eed) 1/2/3/4; Retinoblastoma binding protein (RbpA) -46 / -48.
Arabidopsis: The Arabidopsis PRC2 complex components have evolved into small families. Homologues of the four core components are optionally interchangeable: CURLY LEAF (CLF), SWINGER (SWN), MEDEA (MEA); EMBRYONIC FLOWER 2 (EMF2) , REDUCED VERNALIZATION RESPONSE 2 (VRN2) and FERTILIZATION-INDEPENDENT SEED 2 (FIS2); MULTICOPY SUPPRESSOR OF IRA (MSI) 1/2/4/5 ; FERTILIZATION INDEPENDENT ENDOSPERM (FIE).
E(z) homologues in blue; Su(z)12 homologues in red; ESC homologues in green; N55 homologues in brown. Lines indicate protein variants from a single gene or alternative subunts derived from multiple gene copies.

1.3 Plant Polycomb Repressive Complex 2 proteins in an wider evolutionary context

PRC2 proteins and their functions are conserved among most extant eukaryotes. It is assumed that they existed in the last common unicellular ancestor before the emergence of plants and animals. They could have been involved in defense responses against intragenomic parasites such as transposable elements (Shaver *et al.*, 2010). However, PRC2 proteins are not essential in eukariotic development as they were probably lost in the unicellular fungi *Saccharomyces cerevisiae* and *Saccharomyces pombe* (Butenko & Ohad, 2011). A detailed phylogenetic analysis found that PRC2 proteins are abundant in three of the six major eukaryotic groups: Opisthokonta (animals and fungi), Archeplastida (red algae, green algae and plants) and Chromalveolata (Alveolate and Stramenopiles) (Shaver *et al.*, 2010). Hence, PRC2 proteins are widely spread in eukariotes but not crucial for their developmental program.

The MSI1 homolog N55 is most conserved among the eukariotic PRC2 proteins, probably because it is also a core feature of the NURF and CAF-1 chromatin assembly machineries (Shaver *et al.*, 2010). In Arabidopsis, MSI like proteins can be divided into three diverse families: MSI1-likes, MSI2/3-likes and MSI4/5-likes (Gleason & Kramer, 2012). N55 likely has a function that is generally important for chromatin modification.

The VEFS and C2H2 domains of Su(z)12 are highly conserved. However, the protein has probably undergone a complicated evolution since it is absent in *Caenorhabditis elegans* and several algae (Shaver *et al.*, 2010). Most plant species contain either only EMF2 or proteins that diversified from EMF2. In the Brassicaceae family to which Arabidopsis belongs, FIS2 probably diversified from VRN2 which arose from EMF2 (Luo *et al.*, 2009). Su(z)12 might be less important for PRC2 function.

ESC homologues are probably monophyletic and nearly every protein from animals to plants contains seven WD40 repeat domains. ESC-likes are highly divergent in nematodes but highly conserved in the plant lineage. Strikingly, mutation of the *Physcomitrella patens* ESC homologue *PpFIE* can be partially complemented by *FIE* from Arabidopsis (Mosquna *et al.*, 2009). This depicts the importance of a conserved ESC structure and function in the PRC2 machinery.

Proteins that are homologues to E(z) were identified in all species were the H3K27-me3 mark has been identified. Most animals contain a single E(z) homologue, however the clade underwent considerable expansion in vascular plants (Shaver *et al.*, 2010). Strikingly, the annotated domains between rice, Arabidopsis and Drosophila are highly conserved (How Kit *et al.*, 2010). The lower plants *Physcomitrella patens* and *Selaginella moellendorffii* have a single CLF homologue. Thus, CLF and SWN must have diversified before the divergence of monocots and dicots. MEA probably arose from SWN as a result of a whole genome duplication and became imprinted during its neo-

functionalization as a seed development regulator. Prior to this study, homologues of MEA have only been identified in Brassicaceae. In contrast, SWN evolved under purifying selection and probably retained its ancestral function (Spillane *et al.*, 2007). It has been proposed that CLF is the ancestral gene since E(z) from *Selaginella moellendorffii* is included in the CLF clade (Luo *et al.*, 2009). Accordingly, the E(z) is probably essential for H3K27me3 deposition and vascular plants CLF homologues have undergone diversification .

In conclusion, PRC2 proteins and especially their key domains are highly conserved among evolutionary distant organisms. However, PRC2 proteins were lost several times during the evolution of eucaryotes. Especially in plants, PRC2 core components have undergone duplications to control distinct developmental processes and transitions.

1.4 Polycomb Repressive Complex 2 recruitment to target site

In Drosophila, PRC2 recruitment is mainly mediated by PREs. These are large nucleosome-depleted sequence stretches that are located far from their target genes and form looping structures. PREs bind PRC2 recruiting proteins like the central PRC2 recruiter Pleiohomeotic (PHO). In contrast, only a few PREs have been identified in mammals and plants (Berger *et al.*, 2011; Lodha *et al.*, 2013; Margueron & Reinberg, 2011; Schuettengruber & Cavalli, 2009).

In mammals, PRC2 target sequences are highly enriched in GC repeat rich regions (CpG islands), but no specific recruiting element could be detected in these (Ku *et al.*, 2008). It is suggested, that mainly unmethylated CpG islands that are devoid of transcription factor binding sites can attract PRC2 (Simon & Kingston, 2013). PRE-kr and D11.12 were identified to be involved in PRC2 recruitment. The transcription factors Ying Yang 1 (YY1) and RING1 and YY1-binding -protein (RYBP) were shown to be involved in the mediation of this binding suggesting a common mechanism between Drosophila and mammals (Sing *et al.*, 2009; Woo *et al.*, 2010). However, genome wide studies showed that there is no clear overlap between these proteins and PcG target genes (Xi *et al.*, 2007). Genetic studies have identified the involvement of several DNA-binding proteins that mediate PcG recruitment, however, genome wide occupancy studies indicate only a partial overlap between these *trans*-acting factors and PcG targets. Additionally, mutants of these PcG recruiting proteins have a milder phenotype than PcG mutants (Li *et al.*, 2010). Consequently, it is likely that mammalian PRC2s are only partially recruited by PREs and associated PRC2 recruiting proteins.

Recent studies have raised the role of long non coding RNAs (lncRNA) in PcG recruitment. In mammals, Xist and RepA lncRNAs recruit and spread PRC2 during X-Chromosome inactivation in *cis*. Moreover, the *trans*-acting lncRNA HOTAIR recruits

PRC2 to repress the homeobox gene HOX-D (Plath *et al.*, 2003; Rinn *et al.*, 2007). It has also been shown that short RNAs (ncRNA; 50-200nt) derived from the proximal promoter of the target genes play a role for the association of PRC2 with its targets (Kanhere *et al.*, 2010).

None of these *cis*- and *trans*-acting factors are essential for global PcG function. It was therefore suggested that a combination of PREs, PTMs, regulatory proteins, ncRNAs, specific Histone variants and CpG islands might be responsible for the recruitment of PcG proteins (Beisel & Paro, 2011; Margueron & Reinberg, 2011).

In the current model, the PRC2 holoenzyme is recruited to the target genes in a quartered mechanism of non-consecutive steps (Margueron & Reinberg, 2011). In one layer, the *trans*-acting factors AE binding protein 2 (Aebp2) and Jumonji, AT rich interactive domain 2 (Jarid 2) interact directly with DNA. A second stage is defined by histone to protein interaction via RbpA48 (Kim *et al.*, 2009; Li *et al.*, 2010; Song *et al.*, 2008). A third layer involves the interaction of the ESC homologue Eed with H3K27me3 and interaction of the Polycomb-likes (Pcl) with an unknown histone mark (Margueron *et al.*, 2009). Finally, lncRNAs mediate the binding of PRC2 to target loci, for example via their tertiary structure (Tsai *et al.*, 2010). As a result, the specificity and the binding intensity can be modulated by a combination of histones, histone modifications and abundance of certain proteins as well as ncRNAs.

PRC1 containing the H3K27me3 binding CBX protein (Polycomb in Drosophila) was depicted to co-localize extensively with H3K27me3 and loss of H3K27me3 dislodges CBX-PRC1 distribution. Allocation of PRC1 that lacks CBX is not affected by loss of H3K27me3. Hence, two types of PRC1 recruitment have been suggested, H3K27me3 dependent and H3K27me3 independent (Tavares *et al.*, 2012). Furthermore, it was demonstrated that the PRC1 family can be recruited directly to promoter regions by physical interactions with transcription factors operating at promoters (Yu *et al.*, 2012). lncRNAs have been implicated to mediate PRC1 complex recruitment (Yap *et al.*, 2010). In short, recruitment of PRC2 and PRC1 appears not to be not always connected as assumed in earlier studies.

However, the primary signal that recruits PRC1 independently of H3K27me3 is not known. Recruitment of the PRC1 complex can also be mediated by lncRNAs .

In Arabidopsis the situation remains unclear. Homologues of Aebp2, Jarid 2 and RbAp-46/-48 have not been reported to date. However, PHD-finger proteins like VIN3, VRN5 and VEL1 are functionally related to the Pcls (De Lucia *et al.*, 2008; Sung & Amasino, 2004). In addition, the importance of trans acting factors, PRE-like elements and ncRNAs in PRC2 recruitment is emerging (Heo & Sung, 2011b; Liu *et al.*, 2011; Lodha *et al.*, 2013).

It was demonstrated that AGAMOUS (AG) directly represses *WUSCHEL* (*WUS*) expression by binding to the *WUS* locus and recruiting directly or indirectly the PRC2.

This leads to methylation of H3K27 at *WUS* (Liu *et al.*, 2011). Furthermore, repression of *BREVIPEDICELUS* (*BP*) and *KNAT2* is mediated through ASYMMETRIC LEAVES1 (AS1) and AS2 dependent PcG recruitment (Lodha *et al.*, 2013). Accordingly, non PcG proteins play a role in the specific recruitment of the PcG machinery.

In the same report, it is illustrated that H3K27me3 is reduced at *BP* when the promoter region of a *BP* transgene is mutagenized. The authors suggest that this region might reconstitute a plant PRE. In addition, the RLE *cis*-acting element found in the *LEAFY COTYLEDON2 LEC2* promoter can trigger H3K27me3 deposition when inserted into a novel genomic context (Berger *et al.*, 2011). However, it remains unclear, if the proposed PRE-like elements can directly recruit PRC2.

Repression of the floral integrator *FLC* involves intronic RNA mediated recruitment of the PRC2s (see chapter 1.9), (Swiezewski *et al.*, 2009; Heo & Sung, 2011b). The PRC2 target genes *FLOWERING LOCUS T* (*FT*) and *AG* also contain extensive regulatory regions within introns, similar to *FLC*. These regions could comprise loci coding for hidden regulatory RNAs. Thus, it is likely that ncRNAs are involved in the recruitment of plant PRC2.

Until recently, it was unclear if plants contain a PRC1-like complex, very little is known about its recruitment. However, H3K27me3 dependent and independent pathways were suggested (personal communication, Miriam Calonje, University of Heidelberg).

In short, recruitment of PRC2 in mammals and plants is based on a multifactorial system, in which DNA elements, proteins and RNAs recruit PRC2 to target cite.

1.5 Function of the Polycomb Repressive Complex 2 and its central output, trimethylation at Histone H3 Lysine 27

A central function of PRC2 is the deposition of mono-, di- and tri-methyl at H3K27 (Cao *et al.*, 2002). A possible mechanism was recently proposed based on the molecular architecture of the mammalian PRC2 (Fig. 1.2 A; Ciferri *et al.* (2012)). To perform catalysis, mammalian PRC2 binds to Histone tails via RbpA48. The Ezh2 methyltransferase domain is structurally coupled with the WD40 domain of Eed and the VEFS domain of Suz12 (Margueron *et al.*, 2009; Schmitges *et al.*, 2011; Ciferri *et al.*, 2012). The PRC2 catalytic activity is mainly regulated by modifications of Histone tails, nucleosome density and regulatory proteins (Simon & Kingston, 2013).

Catalysis is probably performed as following: PRC2 is located between two nucleosomes and RbpA48 binds via its WD40 domain to H3 tails. At compact loci, H3K27me3 stimulates the activity of PRC2 to trimethylate proximate nucleosomes at H3K27. In this positive feedback loop, H3K27me3 is recognized by the WD40 beta propeller of Eed. This recognition is transferred via the SANT domains of Ezh2 to its SET do-

main to increase its methyltransferase activity and boosts chromatin compaction (Margueron *et al.*, 2009; Ciferri *et al.*, 2012). Furthermore, Suz12 recognizes densely packed loci and enhances PRC2 catalytic activity dramatically (Yuan *et al.*, 2012). At actively transcribed loci that are marked by H3K4me3, H3K36me2 and H3K36me3, binding of RbpA48 is inhibited. In addition, Suz12 recognizes these positive marks via its VEFS domain and inhibits allosterically the catalytic efficiency of the Ezh2 SET domain (Schmitges *et al.*, 2011; Ciferri *et al.*, 2012). Recent reports illustrated that PRC2 methyltransferase activity is also impeded by H3K27 acetylation (H3K27ac) and H3 Serine 28 phosphorylation (H3S28P) (Simon & Kingston, 2013). It is hypothesized that a molecular switch phosphorylates H3S28 to convert silenced PcG targets from H3K27me3 (OFF) to H3K27ac (ON) (O'Meara & Simon, 2012). To this end, H3K27me3 deposition by PRC2 is promoted upon recognition of H3K27me3 or nucleosome rich loci and mitigated in the presence of H3K4me3, H3K36me2, H3K36me3, H3K27ac and H3S28P.

The Pcl proteins and Jarid 2 comprise important cofactors that modulate PRC2 association to chromatin in different genomic contexts. They mediate recruitment to target sites and fine tune the PRC2 catalytic properties (Simon & Kingston, 2013).

Apart from its role in depositing H3K27me3, PRC2 has been implicated in more regulatory activities. Mechanisms of PRC2 action change depending on the composition of the complex. The mammalian Ezh1-PRC2 and Ezh2-PRC2 regulate transcription through different pathways. Ezh1-PRC2 directly compacts chromatin in a RbpA48 dependent manner which is largely independent of the SET domain and Eed proteins. Ezh2-PRC2 represses genes via deposition of H3K27me3. For this, the methyltransferase activity of the SET domain as well as Eed are essential (Fig. 1.2 B; Margueron *et al.* (2008)). In a Polycomb independent manner, phosphorylated Ezh2 can activate transcription (Fig. 1.2 C; Xu *et al.* (2012)). In short, PRC2 components regulate gene expression through a variety of mechanisms.

H3K27me3 is appointed to be the major determinant of CBX containing PRC1 recruitment (Tavares *et al.*, 2012). PRC1 can silence chromatin through ubiquitilation of H2AK119, chromatin compaction and interaction with the general transcription factor machinery. However, the true function of H3K27me3 remains unclear. As discussed above, it is unknown whether PRC2 induces gene silencing or follows gene repression, which it then reinforces (Henikoff & Shilatifard, 2011). It is hypothesized that H3K27me3, possibly together with PRC2, can block RNAPII recruitment to target gene promoters. Moreover, H3K27me3 might affect nucleosome dynamics, interact with chromatin remodelers (other than PRC1) and antagonize activating marks such as H3K27ac (Simon & Kingston, 2013). Like PRC2, the H3K27me3 mark regulates gene expression through different mechanisms.

Briefly, three mechanisms of PRC2 mediated gene regulation are known in mammals. Firstly, in the canonical model, Ezh2-PRC2 trimethylates H3K27 and the catalytic

activity is modulated by Eed and Suz12. Repressive transcription histone marks and compact chromatin promote H3K27me3 deposition while in loci of active transcription PRC2 catalytic activity is inhibited. Secondly, the non-canonical Ezh1-PRC1 represses transcription by compacting chromatin in a H3k27me3 independent manner. Thirdly, the non-canonical Ezh2 can activate transcription in an PcG independent manner. Finally, H3K27me3 can also control transcription through multiple pathways.

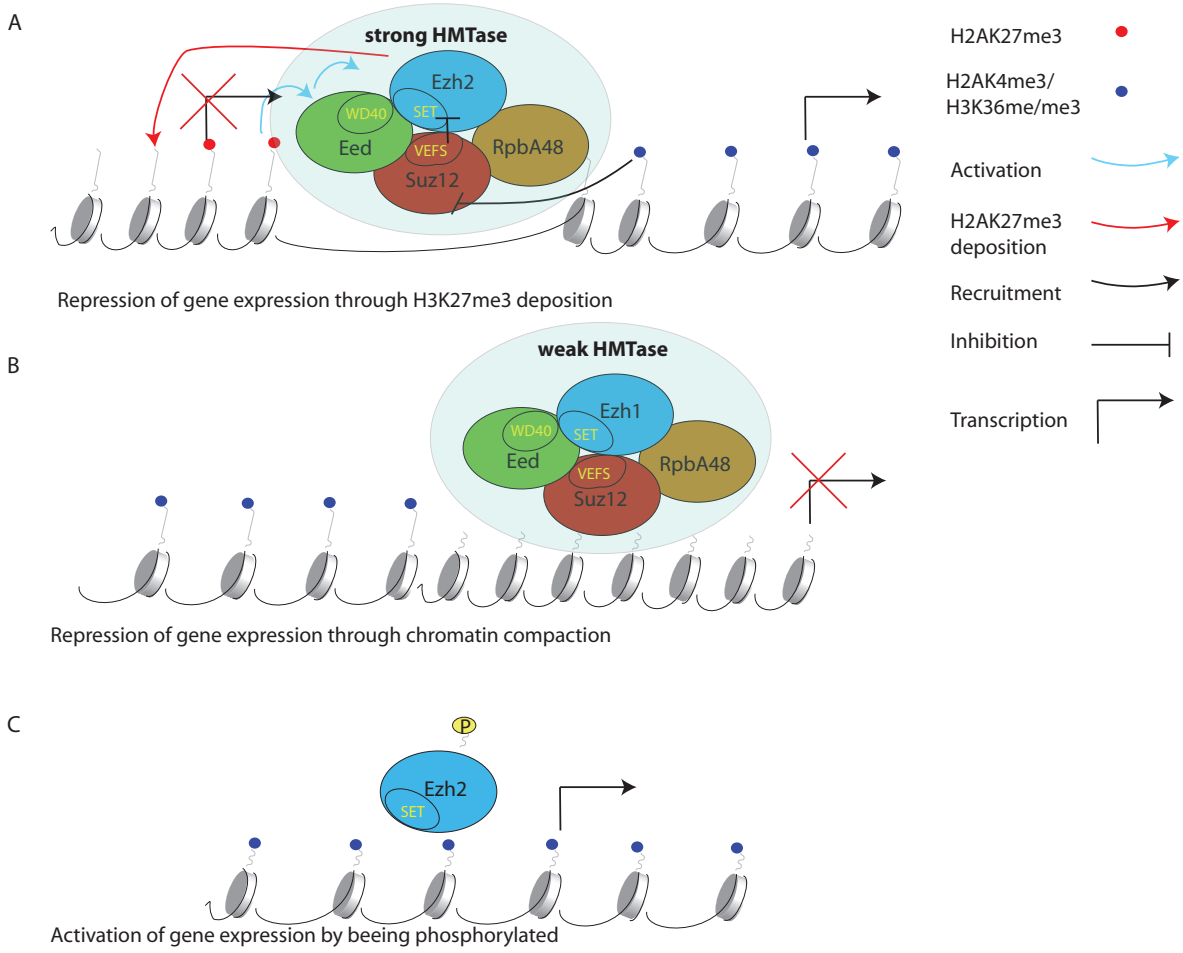


Figure 1.2: **Models of mammalian PRC2 transcription regulation mechanisms**. Schematic representations.
A: Canonical Ezh2-PRC2. Rbp48 can recognize an unmodified histone tail preferentially at dinucleosomes. Eed can recognize H3K27me3 via its WD40 beta-propeller. Upon recognition a positive feedback loop is initiated through enhancement of the Ezh2 SET domain methyltransferase activity via the Ezh2 SANT domains. Suz12 is modified by H3K4me3 and H3K36me2/me3 so that it allosterically inhibits the catalytic activity of Ezh2 SET domain (Ciferri *et al.*, 2012).
B: Non-canonical Ezh1-PRC2. Ezh1 can directly compact and repress open chromatin via Rbp48 in a partially SET domain independent manner (Margueron *et al.*, 2008).
C: Non-canonical Ezh2. Posphorylated Ezh2 can directly activate transcription in a PRC2/H3K27me3 independent manner (Xu *et al.*, 2012).
Ezh2 homologues in blue; Suz12 homologues in red; Eed homologues in green; RbpAb48homologues in brown.

1.6 Genome wide distribution of trimethylated H3K27 in Arabidopsis

Approximately 5% of the Drosophila and human genome are covered by H3K27me3 (Ku *et al.*, 2008; Margueron & Reinberg, 2011). The H3K27me3 mark covers broad domains including the entire transcription unit and regulatory regions (Schwartz *et al.*, 2006; Boyer *et al.*, 2006). In Arabidopsis 4.400 to 8.000 of the ca. 23.000 genes of the genes are marked by H3K27me3. On average these genes are expressed at significantly lower levels than genes without H3K27m23. In contrast to animals, H3K27me3 covers mainly transcribed regions in the euchromatic chromosome arms and it is not associated with low nucleosome density regions. Nevertheless, about 70 % of the H3K27me3 target regions are found in the promoters (200 bp upstream of the transcription start sites) (Zhang *et al.*, 2007b; Turck *et al.*, 2007; Lafos *et al.*, 2011). The amount of identified H3K27me3 target genes varies, however the overlap of targets between studies is very high in spite of differences in analysis methods and plant materials. H3K27me3 target genes are enriched in transcription factors, and many of them are key factors in plant development. These results illustrate that the Arabidopsis PcG machinery covers a broad range of target genes and that the H3K27me3 pattern is distinct between plants and animals.

H3K27me3 target genes are expressed only in a few specific tissues (Zhang *et al.*, 2007b). However, comparison of H3K27me3 profiles in the endosperm displayed that only 240 loci are differentially marked by H3K27me3 compared to whole seedlings. Thus, PcGs are only partially acting in a tissue specific manner. Moreover, comparison of leaf and meristem H3K27me3 profiles demonstrated changes in H3K27me3 distribution upon differentiation. These results imply that regulation of H3K27me3 is highly dynamic between different tissues during plant development (Lafos *et al.*, 2011). In a recent report where H3K27me3 levels were profiled in calli, it was demonstrated that H3K27me3 is essential for the leaf to callus transition and that H3K27me3 is critical for the repression of leaf regulatory genes (He *et al.*, 2012). Consequently, the PcG machinery is important for developmental transitions.

Taken together, H3K27me3 in Arabidopsis targets mainly genic regions and is involved in the regulation of a variety of genes. In addition, H3K27me3 changes in a dynamic manner at a few loci depending on the tissue and the developmental stage.

1.7 The Arabidopsis histone-methyltransferase proteins CURLY LEAF and SWINGER

In Arabidopsis, the full function of CLF is masked by partial redundancy to SWN (Chanvivattana *et al.*, 2004). The null *clf* mutant plants have a mild phenotype that is

mainly caused by misexpression of *AG*, *FLC*, *FT* and *SEP-3*. At these loci, H3K27me3 is reduced in *clf* (Goodrich *et al.*, 1997; Lopez-Vernaza *et al.*, 2012). *Swn* null mutants have a wild-type (wt) -like phenotype. However, the *clf/swn* double mutant is severely impaired and develops to a callus like structure and H3K27me3 is probably globally substantially mitigated (Goodrich *et al.*, 1997; Chanvivattana *et al.*, 2004). Thus, CLF or SWN activity are essential for Arabidopsis development, probably because of their ability to trimethylate H3K27. Callus like phenotypes are generally observed in plants which completely lack the PcG members of one protein family (Bouyer *et al.*, 2011; Bratzel *et al.*, 2011; Chen *et al.*, 2010).

CLF and SWN are reported to share 5 conserved domains with the eukariotic E(z) proteins: EZD1 (E(z)-domain 1), C5 (or EZD2 domain), SANT (Swi3, Ada2, N-Cor, and TFIIIB), CXC (cystein-rich / Pre-SET motif) and SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) (Luo *et al.*, 2009). The SET domain is involved in H3K27 trimethylation. The CXC domain is involved in the control of E(z) catalytic activity in vitro and can interact with single-stranded DNA. In addition the Arabidopsis specific protein BLISTER CLF binds to the CLF CXC domain (Schatlowski *et al.*, 2010; Krajewski *et al.*, 2005). The SANT domain has been implicated in protein-protein interactions and histone binding while the function of EZD1 is unknown. In Arabidopsis the C5 domain is involved in the interaction between E(z) homologues and the VEFS domain of Su(z) proteins (Chanvivattana *et al.*, 2004).

CLF and SWN are expressed in most tissues, especially in regions containing proliferating cells. It is thought that they have a general role in vegetative and reproductive development. Nevertheless, their protein levels are (equally) increased upon vernalization, the promotion of flowering after exposure to long periods of cold (Chanvivattana *et al.*, 2004; Wood *et al.*, 2006; Spillane *et al.*, 2007). Moreover, CLF and SWN bind to EMF2 and VRN2. Strikingly, in a Y2H and in co immunoprecipitation assays, FIS2 binding to SWN and MEA is considerably stronger than to CLF. This suggests that CLF might not be part of the FIS-PRC2 complex (Chanvivattana *et al.*, 2004; Katz *et al.*, 2004; Wang *et al.*, 2006).

Size exclusion chromatography suggested that CLF, SWN, VRN2, FIE and VERNALIZATION INDEPENDENT 3 (VIN3) are part of a 650 kDa high molecular mass complex that is heavily increased upon vernalization. Notably, in vernalized plants SWN occurs in a complex with VRN2 and FIE were CLF cannot be detected (Wood *et al.*, 2006). This implies that SWN might have a CLF independent role during vernalization.

Expression studies in *clf* and *clf/swn* implied that CLF and SWN have distinct roles on gene expression. Some genes were similarly changed in *clf* and *clf/swn*. However other genes changed in an additive, opposing or cooperative way (Farrona *et al.*, 2011). CLF and SWN have therefore probably distinct function in Arabidopsis development.

Chromatin immunoprecipitation assays revealed that CLF is binding to *AG*, *WUSCHEL* (*WUS*), *FLC*, *FT*, *BREVIPEDICELLUS* (*BP*) and *KNOTTED 1-LIKE HOMEOBOX* (*KNOX*). H3K27me3 at these genes is reduced in *clf* compared to wt (Liu *et al.*, 2011; Lodha *et al.*, 2013; Schubert *et al.*, 2006). These observations depict that CLF directly binds to specific loci and mediates the trimethylation of H3K27. There are no reports demonstrating SWN binding to chromatin.

Overexpression of *SWN* cannot complement mutations in *clf* suggesting that the unique phenotype of *clf* is not mediated by differences in expression of the two proteins (Chanvivattana *et al.*, 2004). In the current model CLF and SWN can act interchangeably in a complex to deposit the H3K27me3 mark with their SET domain. It is assumed that at some genes either CLF or SWN activity is sufficient for adequate H3K27me3 deposition. These genes can be significantly misregulated only in *clf/swn*. The third Arabidopsis E(z) protein, MEA, is mainly expressed in the female gametophyte, the endosperm and in the central cell. It probably forms a complex with MSI1, FIE and FIS2 specifically regulate seed development. Although *MEA* is expressed in *clf/swn*, H3K27me3 is undetectable in western blot (Luo *et al.*, 1999; Köhler *et al.*, 2003b; Wang *et al.*, 2006).

1.8 EMBRYONIC FLOWER 2 -, VERNALIZATION RESPONSE 2 - and FERTILIZATION INDEPENDENT SEED 2 - Polycomb Repressive Complexes 2 and their role in Arabidopsis development

In Arabidopsis, PRC2 proteins form at least three distinct complexes named EMF2-, VRN2- and FIS-PRC2 which are important at different developmental transitions. Although they have a clearly distinct function, genetic analysis implies that they share some target genes (Makarevich *et al.*, 2006; Pien & Grossniklaus, 2007).

EMF2 forms a complex with FIE, MSI1 and CLF/SWN and controls the floral transition by targeting *FT* and *FLC* (Jiang *et al.*, 2008; Kim *et al.*, 2010; Schönrock *et al.*, 2006; Yoshida *et al.*, 2001). The *emf2* mutant displays pleitropic phenotypes indicating a broad role in the regulation of vegetative plant development (Yoshida *et al.*, 2001). Moreover, H3K27me3 is only partially reduced in *emf2*, probably because of the redundancy of EMF2 with VRN2 (Kim *et al.*, 2012). *EMF2* is expressed throughout the life cycle of Arabidopsis and is especially active in proliferating tissues (Winter *et al.*, 2007). Accordingly, EMF2 is important for floral and vegetative development and has overlapping functions with VRN2.

fie mutants, develop to an unorganized callus like structure in which H3K27me3 distribution is globally lost. However, *fie* seedlings germinate and start developing nor-

mally, suggesting that PRC2 is not important for the embryo development, but crucial for the embryo to seedling transition. These mutants also display extremely early ectopic flowering, highlighting the importance of FIE in the control of flowering time (Bouyer *et al.*, 2011). The *MSI1* insertion allele results in embryonic lethality when maternally inherited while in heterozygous mutants seeds develop in a high percentage independent of fertilization (Köhler *et al.*, 2003b). Subsequently, FIE and MSI1 are essential for Arabidopsis development, likely because of their unique role in the PRC2s.

The VRN2-complex is involved in the control of flowering time by vernalization and VRN2 can be co-immunoprecipitated with FIE, CLF, and SWN and bind the floral integrator *FLC* (Wood *et al.*, 2006; De Lucia *et al.*, 2008). However, VRN2 is not essential to deposit H3K27me3 on *FLC* but on the MADS AFFECTING FLOWERING (MAF) gene family (Sheldon *et al.*, 2009). *VRN2* is expressed throughout development and protein levels but not mRNA abundance is altered by long periods of cold. Interestingly, unlike EMF2-PRC2, VRN2-PRC2 was shown to trimethylate H3K27 in a H3K4 independent manner (Schmitges *et al.*, 2011). These observations display that the VRN2-PRC2 is a central regulator of vernalization and this function might be associated with its distinct catalytic properties.

The FIS-PRC2 complex mainly controls seed development and is constituted by MEA, FIS2, FIE and MSI1. Null mutants of MEA and FIS2 components initiate aspects of reproductive and maternal development in the absence of fertilization (Ohad *et al.*, 1999; Köhler *et al.*, 2003a). One of the central FIS-PRC2 functions is the maternal imprinting of the seed development master regulator *PHERES1*. Interestingly, FIS2-PRC2 regulates itself by depositing H3K27me3 on the *MEA* paternal allele. However, *PHERES1* and *MEA* are also targeted by other PRC2 complexes. A recent report shows that the FIS-PRC2 complex is required for the repression of a small group of genes orchestrating the endosperm cellularization (Weinhofer *et al.*, 2010). Hence, FIS-PRC2 controls embryonic development by imprinting central seed development regulators.

To this end, EMF2-PRC2 has a general role in Arabidopsis development during vegetative growth, floral transition and flower development. VRN2-PRC2 has a central role in vernalization response and thus flower development. Finally, FIS-PRC2 is mainly involved in seed development. It remains widely unclear if these complexes differ from a mechanistic point of view. However, taken the functional variation of the animal PRC2 machinery and the many possible PRC2 combinations into account it is likely that plant PRC2s have gained special functions and regulatory features.

1.9 Transcriptional repression of *FLOWERING LOCUS C*

The regulation of *FLC* expression in the course of vernalization is one of the best studied epigenetic mechanisms in plants Amasino (2009). In the current model, a multi-

component system involving ncRNAs, the VRN-PRC2-complex and three Plant Homeodomain (PHD) proteins controls the transcription of *FLC* in five steps (Heo & Sung, 2011b).

1. Before cold, the VRN2-PRC2 complex is constitutively loosely associated over the *FLC* chromatin and *FLC* is actively transcribed (De Lucia *et al.*, 2008).

2. In the early phases of cold, *COOLAIR* antisense ncRNAs increases and declines in the third stage simultaneously with *FLC* transcript. *COOLAIR* ncRNAs are transcribed from the full *FLC* and repress the transcriptional activity of *FLC* (Swiezewski *et al.*, 2009).

3. Shortly before *FLC* transcript levels begin to decrease, *COLDAIR* RNA is rapidly increased and declines simultaneously with *COOLAIR* and *FLC* mRNA. *COLDAIR* is an unspliced non-coding RNA transcribed from within the first *FLC* intron involved in the recruitment VRN-PRC2 to *FLC* (Heo & Sung, 2011b).

4. In the late cold period, the expression of the PHD-finger protein *VIN3* peaks when *FLC*, *COOLAIR* and *COLDAIR* levels surpass their peak phase (De Lucia *et al.*, 2008).

Induction of cold correlates with the increase of chromatin marks that are generally associated with high gene repression (H3K27me3) and the decrease of PTMs that coincide with high gene expression levels (H3K4me3) (Zhao *et al.*, 2005; Finnegan & Dennis, 2007). VIN3 interacts biochemically with VIL1/VRN5 PHD-finger protein and forms the PHD-PRC2 complex.

5.After the prolonged cold period, VIN3 is no longer present and VIL1/VRN5 associates more widely with PRC2 throughout the *FLC* locus. This correlates with a significant increase in H3K27me3 and LHP1 levels (Mylne *et al.*, 2006; Sung *et al.*, 2006; De Lucia *et al.*, 2008).

Knockdown of one of the components of this system inhibits H3K27me3 accumulation after vernalization and results in unstable repression of *FLC* transcription (Mylne *et al.*, 2006; Sung *et al.*, 2006; De Lucia *et al.*, 2008).

Briefly, constant H3K27me3 mediated *FLC* repression during prolonged periods of cold is controlled by a five step mechanism. The mechanism involves controlled subsequent occurrence of ncRNAs, VRN2-PRC2, and PHD-PRC2.

Two recent studies suggested a bistable mechanism to control H3K27me3 mediated silencing of an individual *FLC* locus using mathematical modeling (Angel *et al.*, 2011; Satake & Iwasa, 2012). In this model it is assumed that *FLC* can be in an actively methylated (A), a neutral unmethylated (U) or in a repressed methylated (M) state. The chromatin states antagonize each other and promote their own deposition through an autocatalytic positive feedback loop (Margueron *et al.*, 2009; Schmitges *et al.*, 2011). Transition from A to M state and vice versa can occur only via U. These steps possibly involve histone demethylases like RELATIVE OF EARLY FLOWERING (REF6) (Lu *et al.*, 2011). Before vernalization, *FLC* loci in individual cells are mostly in an active

state (A). From the onset of vernalization *VERNALIZATION INSENSITIVE 3* (*VIN3*) expression raises continuously causing an increase in H3K27me3. In case a certain H3K27me3 threshold is reached, *FLC* switches from an active state to a repressed state (M). This model was supported by *FLC* - reporter gene studies. At the beginning of cold, most cells expressed *FLC* while on the course of vernalization the number of cells expressing *FLC* decreases.

In conclusion, memory of *FLC* expression is based on a quantitive bistable system. Length of cold is quantitively reflected in VIN3 abundance and controls a PcG based bistable system.

It is not known if the basic mechanism of PRC2 mediated regulation of *FLC* is universal or if it comprises a unique mechanism. Studies on *FT* indicate several striking differences. Detailed analysis of the regulation of PRC2 target genes other than *FLC* will help understand that basic mechanisms of the PcG machinery in plants.

2 Aims of the study

This study aims to unravel differences between the catalytic PRC2 components CLF and SWN in Arabidopsis. Several functions have been assessed for the different PRC2 components. The mammalian Ezh1- and Ezh2-PRC2 confer gene repression by different mechanisms and in Arabidopsis EMF2- and VRN2-PRC2 complexes have distinct substrates (Margueron *et al.*, 2008; Schmitges *et al.*, 2011). Moreover, in plants and mammals specific PRC2 complexes have evolved that are active only in distinct tissues or developmental processes and target specific genes (How Kit *et al.*, 2010; Margueron *et al.*, 2009; Spillane *et al.*, 2007). Finally, the diversification between CLF and SWN exists in most vascular plants and global expression analysis of *clf* and *clf/swn* mutant Arabidopsis seedlings indicated that the two proteins have distinct functions (Farrona *et al.*, 2011; Gleason & Kramer, 2012). To test this hypothesis four approaches were used: phylogenetic analysis of CLF and SWN (1), creation of genome wide CLF allocation maps (2), creation of genome wide H3K27me3 allocation maps mediated by CLF, SWN or both and bionformatic analysis of genes in which H3K27me3 deposition depends on CLF or SWN (3) . The results were summarized in four parts.

In the first part of the study, the conservation between CLF and SWN in vascular plants and in the Arabidopsis genus was illustrated. Conserved amino acid differences in the catalytic domain and the middle domain of the proteins were highlighted.

In the second part of the study, the identification of CLF target genes in Arabidopsis was assessed by Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq). Tests on the effect of the 26S-Proteasome on CLF stability and statistical confirmation of CLF binding to chromatin were additionally performed.

In the third part of the study, the ChIP-Seq against H3K27me3 in *clf*, *swn* and wt Arabidopsis seedlings is highlighted. Genes where H3K27me3 deposition depends on CLF or SWN and the specific H3K27me3 pattern mediated by each of the two proteins were depicted.

In the fourth part of the study, the genes at which H3K27me3 deposition depends on CLF and SWN are assessed. CLF dependent genes are examined with respect to their expression in Arabidopsis PcG mutants and their chromosomal distribution. Further analysis are performed to identify conserved nucleotide sequences within the CLF target genes.

3 Materials and Methods

3.1 Phylogenetic analysis

3.1.1 Extraction of sequences

All sequences were obtained from Phytozome 9.0 (Table 3.1). *Arabidopsis* sequences were used as a template and blasted against the sequences of respective organisms.

3.1.2 Calculation of evolutionary distances and protein alignment

The evolutionary history was inferred using the neighbor-joining method or the maximum parsimony method (Fitch, 1971; Saitou & Nei, 1987). The bootstrap consensus tree inferred from 100.000 replicates represented the evolutionary history of the proteins analyzed (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the number of differences method and are in the units of the number of base differences per sequence (Nei & Kumar, n.d.). Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011). Protein sequences were aligned with CLUSTAL OMEGA using default parameters and visualized with Jalview (Clamp *et al.*, 2004; Sievers *et al.*, 2011). Evolutionary conserved sequence stretches were extracted and analyzed with Simple Modular Architecture Research Tool (SMART) to determine conserved domains (Letunic *et al.*, 2004).

3.2 Biomolecular analysis

3.2.1 Plasmid constructions and plant transformation

PCR products for genes were amplified from Arabidopsis cDNA (genes) of Columbia (Col) ecotype. For amplification of sequences specific primers with GATEWAY tails were used. The forward primers contained the attB1 extension (5′- GGGGACAAGTTTGTACAAAAAAGCAGGCTTA - 3′); reverse primers contained the attB2 tail (5′ - GGGGACCACTTTGTACAAGAAAGCTGGGTA - 3′).

Long PCR fragments were amplified with the TaKaRa Ex Taq polymerase (Takara Bio Inc.), while for shorter ones the Expand High Fidelity PCR system (Roche) was used. PCR reactions were performed as shown in table 3.2. To generate the entry clones, PCR products were introduced into the GATEWAY pDONR201 vector (Invitrogen) through

Table 3.1: **Gene names of CLF-, SWN- and MEA- homologues**

Organism	CLF homologgue	SWN homologue	MEA homologue
Brachypodium distachyon	*Bradi1g64460*	*Bradi1g48340*	-
Brassica rapa	*Bra032169*	*Bra036300*	*Bra033334*
Capsella rubella	*Carubv10024520m*	*Carubv10000337m*	*Carubv10012183*
Chlamydomonas reinhardtii	*g18041.t1*	-	-
Drosophila melanogaster	E(Z)	-	-
Glycine max	*Glyma11g05760*	*Glyma10g01580*	*Glyma19g40430*
Hordeum vulgare	.	*BAJ84922*	-
Malus domestica	*MDP0000854536*	*MDP0000298253*	-
Medicago truncatulla	*Medtr7g109560*	*Medtr1g086980*	*Medtr5g021590*
Oryza sativa	*Os06g16390*	*Os03g19480*	-
Osteococcus lucimarinus	*19017*	-	-
Petunia hybrida	*PhCLF1*	*PhCLF3*	-
Physcomitrella patens	*Pp1s100-146V6*	-	-
Populus trichocarpa	*Potri007G045800*	*Potri002G195700*	*Potri014G120100*
Prunus persica	*Ppa001213m*	*Ppa001254m*	-
Ricinus communis	*30076m004652*	-	-
Selaginella moellendorffii	*111860*	-	-
Sorghum bicolor	*Sb10g004560*	*Sb01g037450*	-
Solanum lycopersicum	*Solyc03g044380*	*Solyc01g079390*	-
Theobroma cacao	*Thecc1EG000655t3*	*Thecc1EG004817t1*	-
Vitis vinifera	*GSVIVT0102297001*	*GSVIVT 01028124001*	-
Zea mays	*GRMZM2G157820*	*GRMZM2g043484*	-

BP reactions. Sequencing verified the successful introduction of the correct sequence into the vector. The expression clones were created by performing LR reactions with the GW:HA-pAM-Kan destination vector (Franziska Turck, MPIPZ, Cologne). The sequences for each construct are listed in table 3.4. Promoter sequences were amplified from gDNA extracted from Col plants with primers that contained the AscI and AbsI restriction sites. After restriction of amplified *CLF* promoter as well as GW:HA-pAM-Kan vectors that contained *CLF* with AscI and AbsI, restriction products (2,7 kb and 8,0 kb respectively) were gel purified and ligated with each other. The The same procedure was performed with the*SWN* (1,3 kb) promoter and GW:HA-pAM-Kan (8,0 kb) vectors that contained *SWN*. The produced *35S:PcG:HA* and *pPcG:PcG:HA*, vectors were introduced into *Agrobacterium tumefaciens* strain GV3101 (pRSK90) (Koncz & Schell, 1986). Plasmids were transformed into Arabidopsis plants by the floral dip method described by (Clough & Bent, 1998).

3.2.2 Transgenic Arabidopsis lines

Plants of the F1-generation carrying the *35S:PcG:HA* or *pPcG:PcG:HA* transgenes were selected via their Kanamycin resistance. The next generation was tested for single locus insertion of the transgene based on a 3:1 segregation on 1/2 strength Murashige and Skoog (GM) medium. Only T2 lines that showed a segregation ratio between 2:1 and 4:1 for the transgene were transferred to soil and analyzed for homozygocity in the T3 generation. 5 homozygous lines per transgene identified for *35S:EMF2:HA*,

Table 3.2: **PCR conditions for Gateway cloning**

	step	temperature [°C]	duration (s)
	1	94	120
	2	94	15
35x	3	58	30
	4	72	30
	5	72	300
	6	4	∞

35S:VRN2:HA and *35S:CLF:HA*. *35S:SWN:HA*. All the following experiments were performed with homozygous T3 and T4 plants.

3.2.3 Plant growing conditions

Seeds of Arabidopsis were sown on soil or sterilized in 70 % EtOH and 100% EtOH for 5 minutes, respectively before being sown on solid media (0.5x MS / 1% Sucrose). Seedlings were grown in controlled conditions of 80 μmol/m-2sec-1 white light and 22°C (Intellus® environmental controller, Percival) or in the greenhouse for 10 to 14 days. All plants were grown at long day (LD) conditions of 16 hours light and 8 hours dark. Seedlings were either harvested in liquid nitrogen or treated with either 100μM 26S proteasome specific inhibitor MG-132 (MG-132 assays) or 1% formaldehyde (in ChIP assays). For MG-132 assays, treatment was done in 1xMS liquid media supplemented with 3% sucrose. Seedlings were vacuum infiltrated for 10 minutes and transferred back to LD conditions for 10 or 24 hours. After treatment, plants were either analyzed with the Zeiss LSM510Meta microscope or frozen in liquid Nitrogen and kept at -80 C. For ChIP assays, treatment was done in 1xPBS media. Seedlings were vacuum infiltrated for 5 minutes, steered, vacuum infiltrated for 5 minutes, frozen in liquid Nitrogen and kept at -80 C.

3.2.4 Protein abundance determination

Protein expression analysis was performed as described by Searle and colleagues (Searle *et al.*, 2006). A nuclear pellet from roughly 200 μl extract was precipitated with Aceton (80%) and resuspended in 20-40 μl Laemmli buffer. The mixture was incubated at 95 °C for 10-30 min. Proteins were separated on 10-15% Bis-Tris NuPAGE gels, transferred to nitrocellulose membranes and probed with monoclonal a-HA (Roche; 1:1000 dilution) followed by a horseradish peroxidase-conjugated secondary antibody (1:7500 dilution). Immunoreactive proteins were visualized by a 1:1 mix of Pico and Femto chemiluminescence substrate system (Pierce). In some cases, the membranes were subsequently or directly probed with the native antibodies a-VRN2 (Eurogenetec), a-

Table 3.3: **a-EMF2, a-VRN2, a-CLF and a-SWN antibodies**

Antibody	Source	Amino acid sequence of epitope	Aminoacid interval
a-EMF2-1	Rat Rabbit	CEVGSTEKSPYSSFSY CEVGSTEKSPYSSFSY	
a-EMF2-2	Rat Rabbit	CVRSEKSRIPPGKHYE CVRSEKSRIPPGKHYE	
a-VRN2-1	Rat Rabbit	unknown unknown	
a-VRN2-2	Rat Rabbit	CSKPRKRRQRGGR CSKPRKRRQRGGR	
a-CLF-1	Rabbit Rabbit		325 - 530 " "
a-SWN-1	Rabbit Rabbit		313 - 488 " "

EMF2 (Eurogenetec) or a-CLF (Eurogenetec) (see table 3.3). A-actin (Abcam) served as a loading control.

3.2.5 mRNA abundance determination

Seedlings were grown on soil for 14 days and frozen in liquid Nitrogen. RNA was extracted with the RNeasy Mini Kit (Qiagen) and 10 μg was DNaseI-treated using the DNA-freeTM kit (Ambion). cDNA synthesis was performed using dT18 primer and the Superscript II reverse transcriptase enzyme (Roche). cDNA was diluted to 150 μl with RNase free water (Invitrogen) and 4 μl were used for quantitative real-time-PCR (qRT-PCR) using a Roche LifeCycler 480 II apparatus and SYBR Green (Biorad) detection. PCR conditions used are shown in table 3.5. A dilution series of GW:HA-pAM-Kan plasmids carrying the tested gene was used as standard for each primer pair and allowed calculation of molar ratios. For the quantification of gene expression the primers shown in table 3.4 were used. Primers against *Protein Phosphatase 2* (*PP2A*) were used as controls to account for variation between samples.

3.2.6 Chromatin immunoprecipitation

ChIP experiments were performed as described by De Lucia *et al.* (2008). Plants were grown as described in chapter 3.2.3. Briefly, 5 to 7 g of soil grown and formaldehyde treated seedlings were ground in liquid nitrogen for 10 minutes. Nuclei were extracted by suspending powder in ChIP extraction buffer 1 and passing the suspension through Miracloth filters. Nuclei were centrifuged (4000xg, 20 minutes) and enriched in two consecutive steps of treatment with extraction buffers 2 and 3, respectively, followed by centrifugation (14.000xg, 10 minutes and 14.000xg, 60 minutes). Pelleted nuclei were lysed in nuclear lysis buffer and sonicated in iced water for 10 minutes at HIGH setting to a DNA fragment size of 300 to 1000 bp using a Bioruptor apparatus. De-

Table 3.4: **Primers used for Gateway cloning, mRNA abundance determination and ChIP-PCR**

Locus	Primer	Nucleotide sequence of primer (5′ - 3′)
Gateway cloning		
attEMF2-CDS	B009-att1EMF2 B010-att2EMF2	(attB1)-CCAGGCATTCCTCTTGTTA (attB2)-AATTTGGAGCTGTTCGAGAAAG
attVRN2-CDS	B011-att1VRN2 B012-att2VRN2	(attB1)-ATGTGTAGGCAGAATTGTC (attB2)-CTTGTCTCTGCTGTTATT
attCLF-CDS	B028-att1CLF B029-att2CLF	(attB1)-ATGGCGTCAGAAGCTTCGCCTT (attB2)-AGCAAGCTTCTTGGGTCTACCA
attSWN-CDS	Z012-attB1-SWN Z013-attB2-SWN	(attB1)-ATGGTGACGGACGATAGCAACTC (attB2)-ATGAGATTGGTGCTTTCTGGCTCTAC
CLF promoter	pCLFlong-fw pCLFlong-rw	CGTAGGGCGCGCCGTATATATATAATCTCCACG CCTCGCCTCGAGGTGTCAAGAAACCAGATCGGA
SWN promoter	pSWN pSWN	CGTAGGGCGCGCCAACCATCAGATATACAAATA CCTCGCCTCGAGGTGATGACTCCTCGAGCTTTC
mRNA abundance		
PP2A	PP2A-fw PP2A-rw	CGGGTTCGACTTGACGTGGAGC ACTCCATCAACCCCCAATCCACCA
CLF-CDS	CLF-cds-fw CLF-cds-rw	AGAGGCAAAGTGCGGCGCTT GGACACTGGCGGCTTCGACA
LHP1	LHP1-cds-fw LHP1-cds-rw	ATGGACCCCCACCCACGAGG CCGTCTCCGCCTCACCAGCT
ChIP-PCR		
At1g67090	At1g67090-fw At1g67090-rw	GGC AAG TAA AAT GAG CAA GC AAC AAG CTT TGG AGT GAT CG
AT4G24420	AT4G24420-fw AT4G24420-rw	AGGACCCAACGAGTGATGAA ACTTGTTGTCAAGGCCAACC
FLC	FLC-B-fw FLC-B-rw	GAAAGAAATAAAGCGAGAAAAGGAA GGCTTTGTGCCCTAATTTGAT
AT5G20710-*BGAL7*	BGAL7-fw BGAL7-rw	GTCCCAACGAAGAAGTGCAT GATGTAGTAAGTTTGATGGTGTGG
AT3G22770-*F-BOX*	F-BOX-fw F-BOX-rw	AGGCGAGTGGTTCTTTGTTTT CTTTGCTACTTTTGCATTGCC
AT5G45960-*GDSL*	GDSL-fw GDSL-rw	TTGCATCACTCTCGTTTACCC GCGTCACAGAGAACAGAAAGC
AT3G01345-*EP-HYD*	EP-HYD-fw EP-HYD-rw	TTGTTTCTCGGTTCTGTGTTTG TCACTCAACAACATCGAGCAC
AT1G30040-*Ga2ox2*	Ga2ox2-fw Ga2ox2-rw	TCATTGGATCTCTCGGATTTG GAATCGCAATCGATAACCAGA
AT1G30040-*Ga2ox2*	Ga2ox2-fw Ga2ox2-rw	CGCGCTATCACCTCTTTCTT TTGAAAAGGCCACTGGAAAC

bris was removed by centrifugation (14.000g, 10 minutes) and fragmented chromatin extract was divided into 4 equal portions that were immunoprecipitated with 5μl of a-HA (Sigma-Aldrich, from rabbit), 1μl of a-H3K27me3 (Milipore, from rabbit) or a-H3 (Abcam, from rabbit). DNA of 6,7% of chromatin extract was extracted and used as input. No antibody was used as a negative control and for some IPs, a-H3 antibody was used for normalization. Immunoprecipitated samples were washed two times with low-, high-salt, LiCl and TE wash buffers before eluting DNA by adding 65 °C elution buffer and incubating at 65 °C for 30 minutes. Samples were reverse cross linked by adding NaCl and incubating for 12 hours at 65 °C and subsequently incubating with 0.5 M EDTA, 20 ul Tris-HCl 1M (pH 6.5), and proteinase K (10 mg/ml) at 45 °C for 60 minutes. DNA was recovered by phenol chloroform extraction and over-night precipitation at minus 20 °C with 1/10th volume of 3M NaAc (pH 5.2), 2 volumes of ethanol and 1μl of glycogen (Chomczynski & Sacchi, 2006). For the qPCR-analysis, a Roche LifeCycler 480 II apparatus and SYBR Green (Biorad) detection were used (Table 3.5). The signal obtained with fragments on the At1g67090 or *ACTIN* were used as a negative controls for PRC2 recruitment (Adrian *et al.*, 2010). For some PCRs, signals obtained with primers on At4g24420 were used for normalization. PCRs on the input were performed to quantify IP efficiency. The primers which were used are listed in table 3.4. PCR conditions are shown in table 3.5.

Table 3.5: **qPCR conditions**

	step	temperature [°C]	duration
	1	95	5′
45x	2	95	10
	3	60	10
	4	72	10

3.2.7 Barcoded library preparation and sequencing

ChIPed DNA could not be measured with the Nanodrop apparatus (Peqlab) but quantification with qPCR indicated that ChIP DNA amounts correspond approximately to a 1/100 dilution of chromatin input samples.

Library preparation with adaptors containing distinct sequence stretches (barcodes) was performed with three different protocols:

1.ChIP-Seq library. Construction protocol for 10ng ChIP-DNA using the Illumina TrueSeq adapters. The protocol from http://ethanomics.wordpress.com/protocols/ was used with a few changes.

In short, end repair of ChIP-DNA or diluted input (1/10, 1/100 and 1/1000) was performed (blunt ends) and poly-A tails were added. A-tailed fragments were purified

using AMPure XP beads and ligated to the diluted (1/250, 1/500, 1/750 and 1/1000) Illumina True-Seq (original concentration 0.25 M) adaptors. Fragments were amplified by performing 18 PCR cycles using the KaPa HiFi polymerase (Peqlab) or Phusion High Fidelity polymerase (NEB) and purified two times with Agencourt AMPure XP beads according to the manufacturer's instructions.

2. Bioo Scientific NEXTflex ChIPSeq Kit for 10ng ChIP-DNA. Experiment was conducted with ChIP-DNA or input (1/10, 1/100 and 1/1000) as described in the manufacturers instructions for "Gel-Free Size Selection Clean-Up" procedure (Option 3). End repair was followed by gel free size selection clean-up using AMPure XP beads as described in the protocol and 3'adenylation. Supplied adapter mixes were used in different concentrations (1/250, 1/500, 1/750 and 1/1000) and samples were cleaned-up using AMPure XP beads as described in the protocol prior to PCR amplification (18 cycles) using the provided master and primer mixes. Libraries were purified two times after PCR amplification using Agencourt AMPure XP beads as described in the protocol..

3. Nugen Ovation Ultralow DR Multiplex Kit for 1ng ChIP-DNA. Experiment was conducted with ChIP-DNA or input (undiluted and 1/100) as described in the manufacturers instructions. End repair was followed by ligation of supplied adapter mixes that were used in different concentrations (undiluted, 1/10 and 1/100). Fragments were cleaned-up using Agencourt RNAClean XP Beads as described in the instruction book, amplified by performing 18 PCR cycles and purified with Agencourt RNAClean XP Beads as described in the instructors manual.

Library amplification efficiency was calculated by comparing abundance of samples before (undiluted samples) and after library preparation (1/500 diluted samples) by qPCR. Analysis by qPCRs was performed using FLC-B primers and SYBR green (Bio-rad) detection on a Roche LifeCycler 480 II apparatus (Tables 3.4, 3.5). Fragment size distribution was determined with the Bioanalyzer apparatus

Barcoded libraries were gel sized (250 to 500 bp) using a 2 % low melt agarose gel and purified with the MinElute gel purification kit (Illumina). Agarose was melted without heating and subsequently, the manufacturers instructions were followed.

Concentration of individual libraries was determined with a Qbit apparatus and a pool of equally concentrated libraries was sequenced with an Illumina HiSeq2000 apparatus (performed by the Genomecentre, MPIPZ, Cologne).

3.2.8 Tukey test

Fragment concentration of ChIP-PCRs was calculated relative to the input. A numeric value (1-16) was assigned to every tested condition (Fig. 4.6 B). In the conducted Tukey test using R, every possible pair of mean values from the ChIP-PCRs is compared based

on the students t-test . Distance between two mean values and corresponding 5 % and 95% confidence intervals were plotted.

3.2.9 Buffers, Solutions, Medium

ChIP Extraction Buffer 1 0.4M sucrose, 10mM Tris-HCl pH 8, 10mM MgCl, 5mM BME, plant proteinase inhibitor cocktail (PI) (cOmplete; Roche), 1 mM EDTA

ChIP Extraction Buffer 2 0.25M sucrose, 10mM Tris-HCl pH 8, 10mM MgCl2, 1% Triton X-100, 5mM BME, plant PI (cOmplete; Roche), 1 mM EDTA

ChIP Extraction Buffer 3 1.7M sucrose, 10mM Tris-HCl pH 8, 0.15% Triton X-100, 2mM MgCl2, 5mM BME, plant PI (cOmplete; Roche), 1 mM EDTA

ChIP Nuclei Lysis Buffer 50mM Tris-HCl ph 8, 10mM EDTA, 1% SDS, plant PI (cOmplete; Roche),

ChIP Dilution Buffer 1.1% Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl pH 8, 167mM NaCl, plant PI (cOmplete; Roche)

ChIP Elution Buffer 1% SDS, 0.1M NaHCO3

ChIP Low Salt Wash Buffer 150mM NaCl 0.1% SDS 1% Triton X-100 2mM EDTA 20mM Tris-HCl pH 8

ChIP High Salt Wash Buffer 500mM NaCl 0.1% SDS 1% Triton X-100 2mM EDTA 20mM Tris-HCl pH 8

ChIP LiCl Wash Buffer 0.25M LiCl 1% NP-40 1% sodium deoxycholate 1mM EDTA 10mM Tris-HCl pH 8

TE Buffer 10mM Tris-HCl pH 8 1mM EDTA

1/2 strength Murashige Skoog medium

4,4 g x l-1 Murashige Skoog basal salt mixture, 1 ml x l-1 Murashige Skoog vitamines, 3 ml x l-1 2-(N-morpholino)ethanesulfonic acid (MES) buffer, 10 g x l-1 sucrose, pH 5,7 with KOH, 0,9 % Agar for solid media

Lysogeny Broth (LB) medium 10 g tryptone, 5 g yeast extract, 5 g NaCl, ad. to 1 l with dH2O, pH 7,2 with NaOH, 8 g Agar for solid media

5 x Laemmli buffer Dissolve 2g SDS in 6,25 ml Tris-HCl (1 M; pH 6,8), 9 ml glycerol, 12 mg bromphenol blue, 5 ml beta-mercaptoethanol, plant PI (cOmplete; Roche)

1 x Transfer buffer 3,03 g Tris, 14,4 g glycine, 200 ml methanol, fill up to 1 l with dH2O

10 x TBS buffer 500 mM Tris, 1,5 M NaCl, fill up to 1 l with dH2O, pH 7,4 with HCl, for TBS-T add 0,1 % Tween 20

3.3 Computational analysis

3.3.1 Tools and data used

The following list of tools was used in this project for data analysis.

1. Perl (http://www.perl.org), a programming language.

2. R (http://cran.r-project.org) (version 2.10.1), a programming language and environment for statistical computing.

3. Bionconductor (version 2.6), a free, open source software project to provide tools for the analysis of genomic data. It is based primarily on the R language (Gentleman *et al.*, 2004).

4. GBrowse, a local instance of GBrowse for visualization of genomic data (Stein, 2013).

5. BWA 0.5.9, a short reads mapper for efficient mapping of probes or reads to a reference genome (Li & Durbin, 2009).

6. SAMtools (version $0.1.6_X86_64 - linux$), provides various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per position format (Li *et al.*, 2009).

7.ChIPR (Göbel et al. 2010), a R package with Ringo incorporated for ChIP-chip data analysis.

8. SICER (version v1.1), a clustering approach for identification of enriched domains from ChIP-Seq data for histone modifications (Zang *et al.*, 2009).

9. Picard (http://picard.sourceforge.net) (version 1.16), Java-based command-line utilities that manipulate SAM files.

10. IGV, a local instance for visualisation of genomic data (Thorvaldsdóttir *et al.*, 2013).

11. MEME 4.8., a tool for the identification of conserved sequence motifs (Bailey *et al.*, 2009).

12. The Arabidopsis Information Resource (TAIR), the database for downloading genome sequences and annotation of Arabidopsis thaliana (Lamesch *et al.*, 2012).

13. Genesis, Java suite for large-scale gene expression analysis (Sturn *et al.*, 2002).

3.3.2 Mapping short reads to Col genome

Single end 100bp reads of barcoded libraries were produced to detect the H3K27me3 targets of *clf*, *swn* and wt seedlings. The raw data from Illumina was in FASTQ format. The reads were mapped to Col genome with BWA Li & Durbin (2009). A maximal edit distance of n=3 including a maximal gap of one were allowed during mapping. The low quality bases at the end of short reads were trimmed with BWA while mapping by setting parameter q=15. The mapped reads were sorted according to their genome

coordinates using SAMtools (Li *et al.*, 2009). In order to completely exclude the effect of potential PCR artefacts, redundant reads mapped to the same position in the genome were cleaned with Picard by keeping just one copy.

3.3.3 Identification of H3K27me3-enriched regions and genes

The ChIP-Seq peak detection tool SICER was used for identification of H3K27me3 marked regions in hybrids (Zang *et al.*, 2009). The alignment result file containing the mapping information of non-redundant reads was used as input for SICER. The adjacent predicted peaks with gaps smaller than 200bp were merged into broader regions, which were further mapped to TAIR9 gene annotation with ChIPR (Gobel *et al.*, 2010). Genes were considered as H3K27me3 targets, if either at least 20% of the gene or at least 500 bp of it were covered by H3K27me3. Intergenic H3K27me3 positive regions were not used for further analysis.

3.3.4 Clustering analysis of gene expression data

Expression data was loaded into Genesis and analyzed using distance Pearson correlation (Sturn *et al.*, 2002). Hierarchical clusters of log 2 transformed values were visually divided into major branches (major expression patterns in the expression set). Analysis of the dataset with the elbow shoulder method using R was performed to confirm correct number of major branches. K-means (n=number of major branches) clustering was applied to group genes together according to their expression pattern. The calculation was repeated 10 times to determine if the genes stably distribute into the clusters.

3.3.5 Chromosomal distribution calculation of genes

The genome was divided into 0,2 Mio bp windows and the number of CLF dependent genes per window was calculated. The same number of H3K27me3 targets was sampled from all H3K27me3 targets randomly for 1000 times to determine the background. The number of H3K27me3 targets fallen into each window was calculated and plotted as the boxplot in Figure. All analysis was performed using R.

3.3.6 Identification and confirmation of conserved sequence motifs

The ChIP-Seq peak detection tools SICER and diffReps 1.5.4 were used to identify H3K27me3 enriched regions (Zang *et al.* (2009), http://code.google.com/p/diffreps/). Division of fragments according to size was performed with PERL. Fragments were analyzed with the MEME 4.8.1 sequence identification tool using the zoops (zero or one occurrence) model. Search for conserved motifs in the fragments was conducted with PERL.

4 Results

To identify structural differences between the PRC2 proteins CLF and SWN and detect plant PREs we used four approaches:

1. We performed phylogenetic analysis of CLF and SWN to assess the evolutive conservation and identify conserved domains and sequence stretches.

2. We developed native antibodies against CLF, SWN and the Zn-finger proteins EMF2 and VRN2 with the aim to perform ChIP-Seq and identify the binding sites and target genes of each of these proteins

3. We created HA-tagged lines of CLF, SWN, EMF2, VRN2 that were expressed under the control of the *CaMV-35S* promoter to perform ChIP-seq and confirm the specificity of the native antibodies. In a parallel approach we created HA-tagged lines under the control of the native promoter.

4. We performed ChIP-Seq experiments with anti-H3K27me3 antibodies in *clf*, *swn* and wt to identify regions that potentially lack H3K27me3 in the mutants compared to the wt. We speculated that regions which are differentially trimethylated at H3K27 in *clf* or *swn* depend on CLF or SWN respectively.

4.1 Phylogenetic analysis of CLF and SWN

4.1.1 The diversification of CLF and SWN is conserved in the plant kingdom

Unique roles have been assigned to the Arabidopsis E(z) proteins CLF and MEA, however, the role of the third family member, SWN, is still unclear (Spillane *et al.*, 2007; Chanvivattana *et al.*, 2004). To gain further insights into the role of SWN and understand the difference to CLF I analyzed the evolutionary relationship of this family. The Arabidopsis sequence of each of these proteins was blasted against the sequences of species across the plant kingdom using the Phytozome 9.0 platform (http://www.phytozome.net/).

CLF and SWN homologues display major differences in their middle parts. In all tested plant species apart from the evolutionary distant Brassicaceae and Fabaceae family members, blasts with MEA resulted in proteins that had a higher homology score to SWN. Some plant species contained many E(z) family proteins. In most of these cases, the homologous proteins grouped with SWN when phylogenetic tests were performed (data not shown). These homologues were removed from the dataset to simplify the phylogenetic tree. In addition, proteins that contained sequence errors

were removed. The sequences of the remaining proteins were aligned with CLUSTAL W and phylogenetic relations were determined by the Neighbor Joining method (Felsenstein, 1985; Sievers *et al.*, 2011).

E(z) proteins diverge into 6 major branches (Fig. 4.1). CLF and SWN are divided into a monocotyledon and a dicotyledon clade. It was proposed that MEA like proteins exist only in Brassicaceaes (Spillane *et al.*, 2007). I identified a novel group of MEA-like proteins within the Fabaceae. These proteins differ from the Brassicaceae MEA-like proteins and phylogenetic analysis with the maximum persimony method, proposed that they are part of the SWN dicot clade (data not shown, Fitch (1971)). The middle part of the Fabaceae MEA-like proteins is similar to the middle part of the MEA-like proteins from Brassicaceae but all other parts of the proteins resemble the SWN sequence. This diversification could be recapitulated when SET domains of CLF SWN and MEA were analyzed.

The E(z) proteins from *Drosophila melanogaster*, *Chlamydomonas reinhardtii* and *Osteococcus lucimarinus* that were used as out-groups are surprisingly in the same brach as the MEA-like proteins, that evolved after CLF and SWN (Spillane *et al.*, 2007). The moss *Physcomitrella patens* and the primitive vascular plant *Selaginella moellendorffii* form a distinct clade in which tested species contain one E(z) family member that is evolutionary closer to the monocot SWN branch than to the CLF lineage. Surprisingly, the alignment illustrates that the E(z) homologues of *Physcomitrella patens* and *Selaginella moellendorffii* are CLF homologues.

SWN and MEA diverge from a common ancestor and MEA but not SWN is driven by positive darwinian selection (Spillane *et al.*, 2007). Moreover, the phenotype of Arabidopsis *swn* mutants resembles wt. This could mean that SWN might have lost its unique role in Arabidopsis. However, comparison of the *SWN* sequences of 80 *Arabidopsis* accessions on the 1001 genomes platform illustrates one SNP in the exons of *SWN*. Compared to the genetically highly diverse gene *FRIGIDA*, *SWN* depicts a remote amount of low-frequency polymorphisms in exons (13 low-frequency polymorphisms in 80 accessions) between *Arabidopsis* accessions (Le Corre *et al.*, 2002). This indicates that *SWN* is conserved as a functional gene and suggests that the SWN function is similar within *Arabidopsis* (Fig. 4.2, $http://gbrowse.weigelworld.org/cgi-bin/gb2/gbrowse/ath_reseq_mpicao2010/$).

From a parsimonious point of view, these results suggest that CLF is the ancestral gene. Strikingly, SWN, exists together with CLF in all analyzed euphylophyta including the Arabidopsis genus. Thus, I propose that the SWN function is conserved.

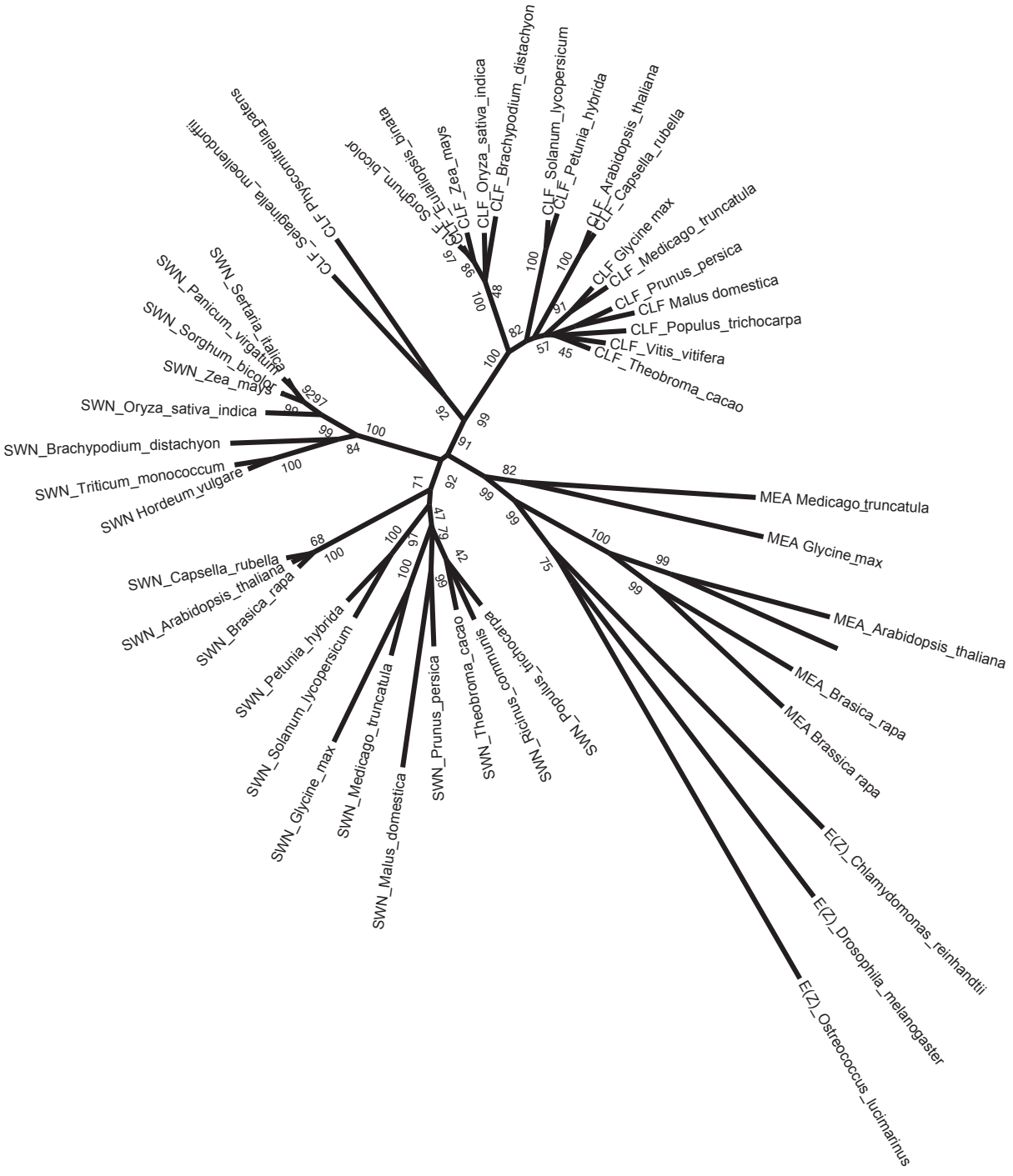


Figure 4.1: **CLF and SWN are evolutionary conserved in higher plants and diverge including MEA into 6 branches.**
Neighbor Joining Analysis of CLF, SWN and MEA proteins. CLF and SWN homologs exist in all tested spermatophyte and both form distinct branches in monocots and dicots. Brassicaceae and Fabaceae contain in addition a MEA homolog that is evolutionary close to the E(z) proteins of *Drosophila* and *Chlamydomonas reinhardtii*. Mosses contain E(z)-like protein that is similar to CLF. Arabidopsis sequences of the respective proteins were blasted against the sequences of the indicated organisms on the Phytozome 9.0 platform. *Drosophila* E(z) was used as an outlier. Alignment was performed with CLUSTALW and tree was calculated with MEGA 5.0. Bootstrap values of 10.000 replicates are indicated on the nodes.

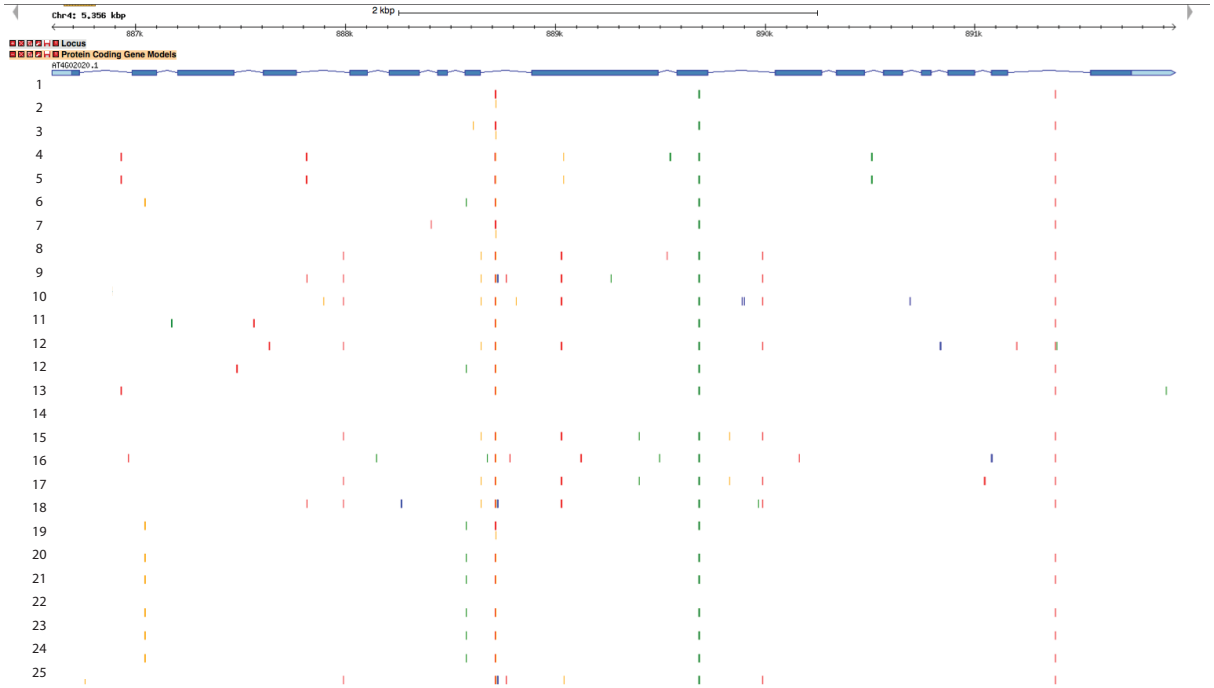


Figure 4.2: **SWN is conserved among Arabidopsis accessions.**
Genome browser view of *SWN* in 25 representative Arabidopsis accessions. *SWN* contains one conserved exotic SNP and 9 low-frequency polymorphisms in *Arabidopsis*. Col-0 has SNPs in the 10th exon as well as in the 7th and the 15th intron.
Data was obtained from the 1001 genomes database ($http://gbrowse.weigelworld.org/cgi-bin/gb2/gbrowse/ath_reseq_m picao2010/$). 25 of 80 analyzed accessions are presented.

4.1.2 CLF, SWN and MEA are highly similar but differ in the middle part

It has been reported that the Arabidopsis E(z) proteins contain one of each of a C5, SANT, CXC and SET domain. The C5 domain is important for protein protein interaction while CXC and SANT domains influence the catalytic activity of the SET domain(Chanvivattana *et al.*, 2004; Ciferri *et al.*, 2012; Schmitges *et al.*, 2011). I observed that CLF and SWN homologues are similar at the NH2 (N)- and COOH (C)- terminal fragment but different in their middle part (Section 4.1.1). To assess the amino acid conservation of the plant E(z) family proteins I aligned CLF-like, SWN-like and MEA-like from Brassicaceae using CLUSTAL OMEGA (Sievers *et al.*, 2011). Conserved domains were identified with Jalview 2.8 (Clamp *et al.*, 2004). Protein sequences with substantial variation were removed and conserved regions were analyzed with SMART (Letunic *et al.*, 2004).

CLF-like proteins share 8 conserved amino acid stretches that correspond to the following domains: H15, Blast SANT, C5, CXC and SET (Fig. 4.3 A). Two unknown conserved sequence blocks were identified between H15 and Blast SANT as well as between C5 and CXC domains. Since they have not been described before, I named these domains MYSTERY (MYR) and CLF RIDDLE (CRI) respectively. BLAST of MYR and CRI did not reveal conservation of these domains in other proteins. In contrary to CLF-like, SWN-like lack conservation in the middle part of the protein which corresponds to the CRI domain in CLF and contain 7 conserved sequence regions (fig. 4.3 B). The conserved domains are positioned and correspond to those found in CLF. MEA proteins share 8 conserved amino acid stretches that are positioned as in CLF (Fig. 4.3). However, the domain that corresponds to CRI in CLF is different and was annotated as MEA RIDDLE (MRI).

The amino acid conservation between CLF-likes, SWN-likes and MEA-likes was determined. CLF-likes, SWN-likes and MEA-likes share 7 conserved core sequence stretches (Fig. 4.3 D). The middle part of the proteins displays substantial amino acid variation. CLF- and SWN-likes are more similar to each other than to MEA-likes.

In conclusion, CLF-, SWN- and MEA-like proteins share 7 conserved domains that display a few amino acid changes and display striking variation in the middle part. In contrast to SWN-likes, CLF- and MEA-likes have in the middle part unique domains.

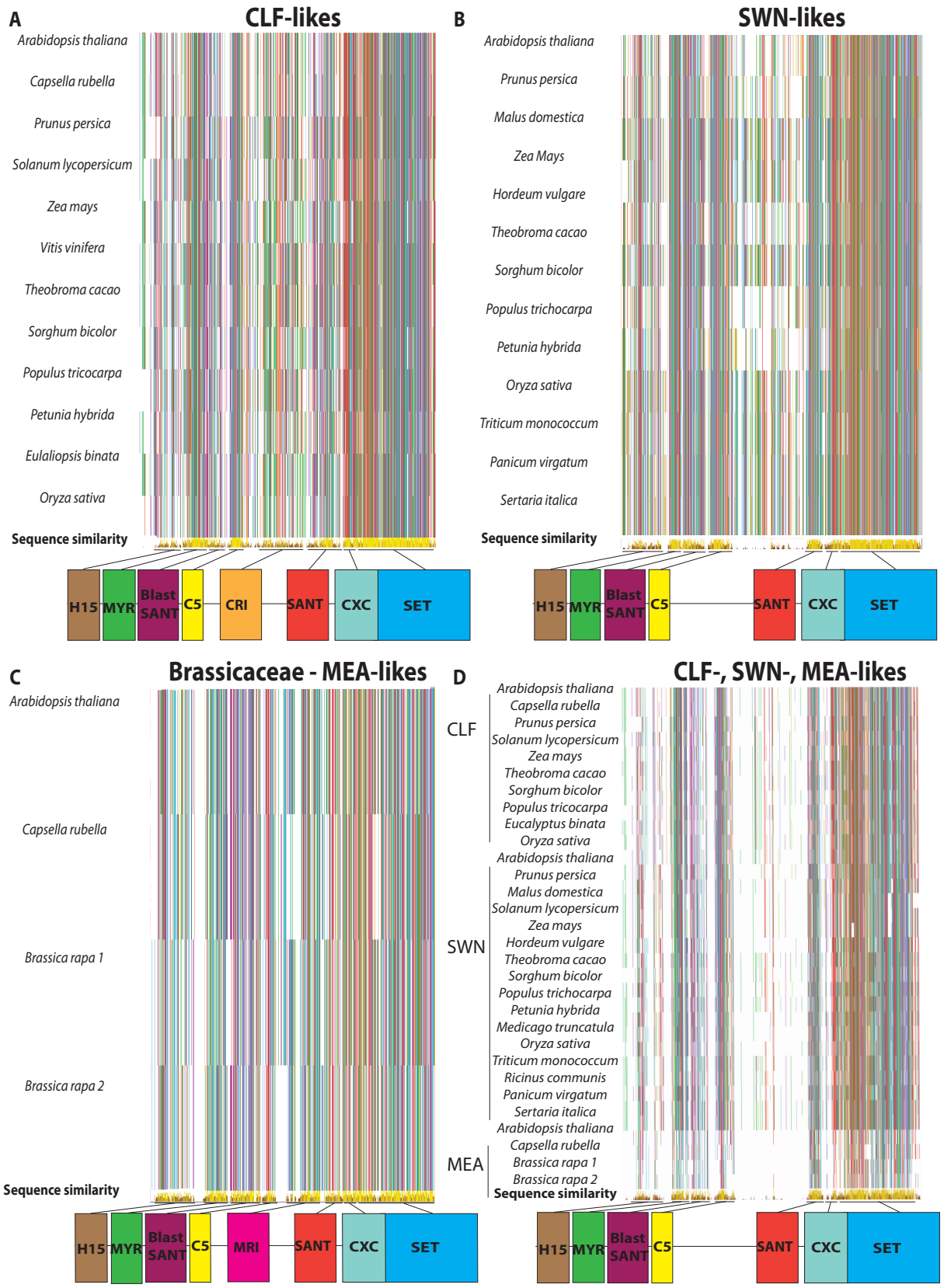


Figure 4.3: **CLF, SWN and MEA share 7 conserved domains and differ in the middle part.**

Figure 4.3: **A**: Protein alignment and schematic representation of CLF-like proteins. CLF-like proteins contain highly conserved amino acid stretches. They correspond from the NH2 (N) - to the COOH (C) -terminus to H15, Blast SANT, C5, CXC and SET domains. A conserved region that I named MYSTERY (MYR) could be identified at the beginning of the protein that is conserved also in SWN (B, D) and MEA (C, D). The middle part of the protein depicts a conserved CLF specific stretch that I annotated as CLF RIDDLE (CRI).
B: Protein alignment and schematic representation of SWN-like proteins. SWN-like proteins contain highly conserved amino acid stretches. They correspond from the N- to the C-terminus to H15, MYR, Blast SANT, C5, CXC and SET domains. The middle part of the protein is not conserved.
C: Protein alignment and schematic representation of MEA-like proteins. MEA-like proteins contain highly conserved amino acid stretches. They correspond from the N- to the C-terminus to H15, MYR, Blast SANT, C5, CXC and SET domains. The middle part of the protein depicts a conserved MEA specific stretch that I annotated as MEA RIDDLE (MRI).
D: Protein alignment and schematic representation of CLF-, SWN- and MEA-like proteins. CLF-, SWN- and MEA-like proteins share 7 conserved domains (H15, MYR, Blast SANT, C5, CXC and SET) and vary substantially in the middle part. MEA proteins are considerable different compared to CLF and SWN.
Arabidopsis sequences of the respective proteins were blasted against the sequences of the indicated organisms on the Phytozome 9.0 platform. Alignment was performed with CLUSTAL OMEGA and homologies were visualised with Jalview 2.8. Clustal color code is used. Relative sequence similarity is indicated in the bottom line in yellow conserved domains were represented as cartoons.

4.2 Chromatinimmunoprecipitation of PRC2 proteins

4.2.1 CLF-HA abundance is low and probably mainly regulated post transcriptionally

To analyze the binding specificity of different PRC2 subunits to chromatin via ChIP I created transgenic lines that expressed HA tagged *CLF*, *SWN*, *EMF2*, *VRN2* that were expressed under the control of the *CaMV-35S* promoter. The same vector was used as in *35S:LHP1:HA* transgenic plants, which were previously used to identify genome wide binding sites of LHP1.HA (Engelhorn *et al.*, 2012; Turck *et al.*, 2007). In addition I planed to create HA-tagged lines of the four PRC2 components under the control of the native promoter. Finally, we generated native antibodies against CLF, SWN, EMF2 and VRN2.

To determine which of the 5 available homozygous *35S:CLF:HA* lines contained the highest CLF-HA abundance and was therefore most eligible for ChIP-Seq, Western blot analysis was performed. CLF-HA (~100kDA) was detected in lines #4 and #11 but the signal was much weaker than that of the positive controls LHP1-HA (normally detected at ~80kDA) and the unrelated protein 3xHA-CDF2 in *35S:LHP1:HA* and *SUC2:3xHA:CDF* (normally detected at ~72kDA) respectively (Fig. 4.4).

Abundance of the PCR2 proteins VRN2-HA and EMF2-HA in 5 independent homozygous lines was similarly low or lower than that of CLF-HA compared to positive controls (Master thesis, Theodoros Zografou). In the same work, no signals at the expected size were detected with antibodies that were developed against peptide epitopes of VRN2 , EMF2 or CLF. I assumed that the antibody sera should be purified to increase the titer and be able to detect the respective antibodies. Therefore a-SWN antibody was not tested. a-VRN was purified, but VRN2 protein could not be detected. a-CLF, a-SWN and a-EMF were not purified. I did not test protein abundance levels of SWN-HA because the homozygous *35S:SWN:HA* lines were not available. Segregation of these lines was not continued because I speculated that there might be a general problem detecting PRC2 proteins in the created transgenic lines. For the same reason, I did not continue cloning CLF, SWN, EMF2 and VRN2 lines that were expressed under the native promoter. Because it was difficult to detect VRN2-HA and EMF2-HA I decided to focus my work on *35S:CLF:HA* #11 in which I could detect the highest CLF-HA amount.

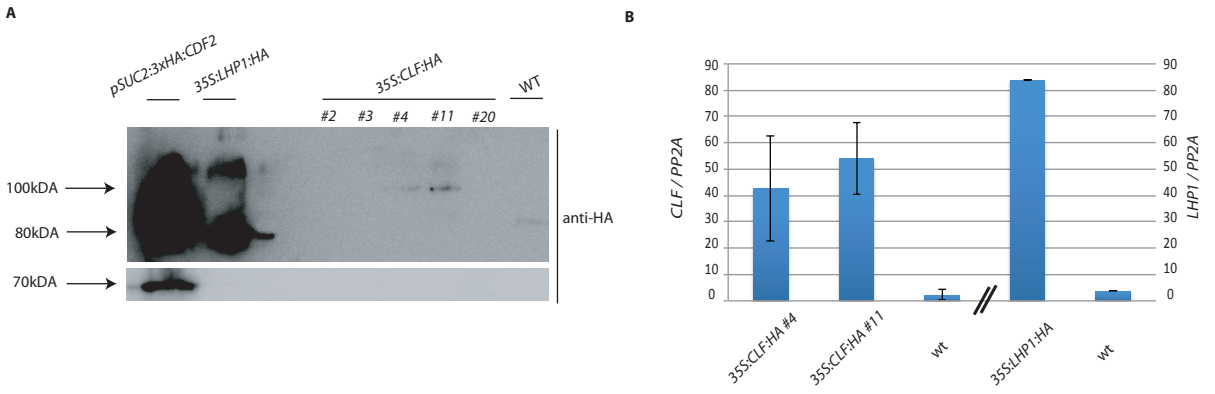


Figure 4.4: **CLF:HA abundance in *35S:CLF:HA* does not correlate with *CLF* expression levels and is low compared to positive controls.**
A: Western blot analysis. CLF-HA was detected in *35S:CLF:HA* lines #4 and #11. LHP1-HA and 3xHA-CDF2 display stronger signal intensity in the respective positive control lines *35S:LHP1:HA* and *SUC2:3xHA:CDF*. No signal could be observed in the negative control (wt). Analysis was performed on crosslinked nuclear extracts from 12-day-old seedlings. Equal protein amounts were loaded and run on a 12% Polyacrylamide gel under non-native conditions. Membrane was probed with a-HA from (Sigma) and developed for 2 sec (lower membrane) or 60 sec (upper membrane) . A representative membrane is illustrated.
B: q-RT PCR analysis. Lines that overexpress HA-tagged versions of *CLF* display higher abundance of *CLF* mRNA than in wt and similar levels to *LHP1* in *35S:LHP1:HA*. Quantification was performed my measuring absolute levels of vectors carrying *CLF* and *LHP1* respectively. Experiments were performed with 14-day-old seedlings grown on soil. The abundance of *CLF* was measured in three biological replicates. The data on LHP1 control is based on one experiment. Error bars represent standard error of the mean.

Based on our results and a previous report, I hypothesized, that plants with and high ectopic PRC2-HA levels might be embryonic lethal (Jeong *et al.*, 2011). In this case, mRNA levels of PRC2 components in the transgenic lines that overexpress these genes would be low because only plants with low expression would have survived the selection. In contrast, if the failure to detect high levels of CLF-HA protein was independent of transgene selection, the ratio of *CLF* mRNA to CLF-HA protein in the transgenic lines should correspond to the proportion of *LHP1* mRNA to LHP1-HA protein in *35S:LHP1:HA* since the same vector had beed used for the generation of both lines. qRT-PCR experiments revealed slightly lower levels of *CLF* mRNA in *35S:CLF:HA* in line #4 and #11 to the quantity of *LHP1* mRNA in *35S:LHP1:HA*. This indicates that the *CLF:HA* transgene can be overexpressed and is not silenced (Fig. 4.4 B). However, CLF-HA abundance was substantially lower than LHP1-HA abundance (Fig. 4.4 A). These results suggest, that CLF-HA quantity is probably regulated at the post-transcriptional or translational level. Similar observations have been reported in previous studies (Wood *et al.*, 2006).

4.2.2 Binding of CLF-HA to AG and FLC is weak but is statistically significant compared to loci were no enrichment is expected

To test if CLF-HA is binding to chromatin I conducted ChIP-PCR analysis with a-HA on chromatin prepared from *35S:CLF:HA* lines. Only a weak CLF-HA signal could be measured at *AG* and *FLC* compared to H3K27me3 (Fig. 4.5). However, binding to *AG* and *FLC* was higher than to At1g67090. No CLF-HA is expected to bind At1g67090 because it is not targeted by H3K27me3 (Adrian *et al.*, 2010). In replicated experiments, the signal of At1g67090 was often similarly high or slightly lower than at loci were binding is expected. Furthermore, the enrichment of LHP1-HA in *35S:LHP1:HA* and of H3K27me3 was in all tested conditions substantially higher that the enrichment of CLF-HA (Master thesis, Theodoros Zografou).

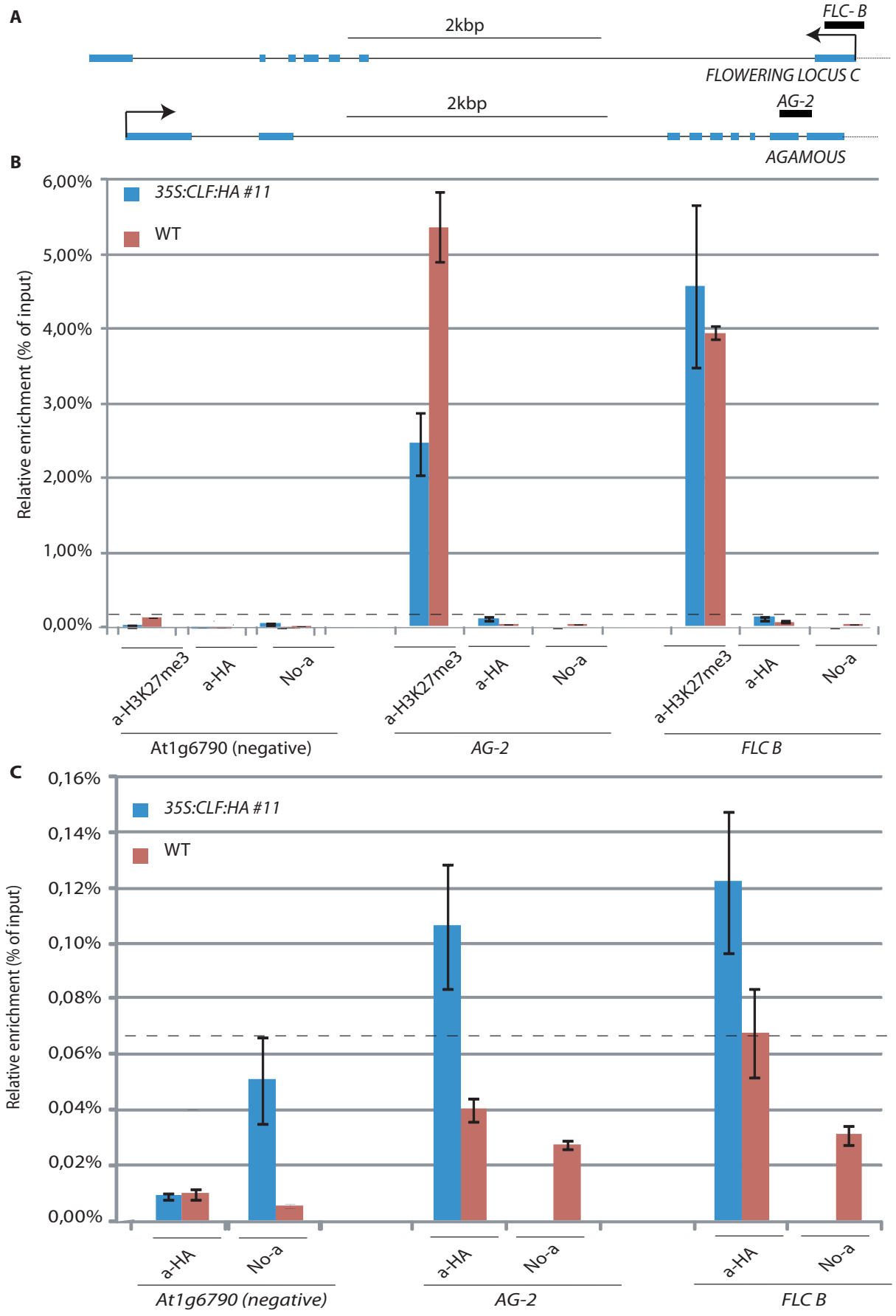


Figure 4.5: **CLF-HA is enriched on *AG*)and *FLC*.**

Figure 4.5:
A: Schematic representation of *AG* and *FLC*. Exons are illustrated as blue boxes, introns as black lines and positions of amplicons used for ChIP analysis as black labeled boxes. Dotted line represents region upstream of the TSS.
B: ChIP-PCR analysis of CLF-HA compared to the H3K27me3 signal at *AG*, *FLC* and At1g67090 loci. In *35S:CLF:HA* (#11, in blue), abundance of CLF-HA on *AG* and *FLC* corresponds to the background levels of the of a-H3K27me3 IPs. Wt (in red), At1g67090 and sample without antibody (No-a) are used as a negative controls.
C: ChIP-PCR analysis of CLF-HA compared to the H3K27me3 signal at *AG*, *FLC* and At1g67090 loci. The same experiment as in B is presented but without the abundance of trimethylated H3K27. CLF-HA in *35S:CLF:HA* displays higher abundance than amplicons of the IPs with a-H3K27me3 at At1g67090 or IPs with No antibody (No-a) and a-HA in wt (in red). Wt, At1g67090 and sample without antibody are used as a negative controls.
Polyclonal a-HA (Roche) and a-H3K27me3 (Millipore) from rabbit were used for IPs on crosslinked chromatin extract from 14-day-old seedlings grown on soil. Enrichment is presented relative to chromatin input (% of input). Error bars represent standard error of the mean. Dashed line represents background, determined as maximum of qRT-PCR signal of negative controls (At1g67090, No antibody). One representative experiment is depicted.

Two possibilities were proposed to explain the weak CLF-HA binding to chromatin that was observed: 1. CLF-HA binding to chromatin is real but close to detection limit. 2. CLF-HA binding is not true. GFP-CLF binds to chromatin in *35S:GFP:CLF ; clf/clf* (in the following *35S:GFP:CLF*) (Schubert *et al.*, 2006). To rule out the possibility that the CLF-HA is interfering with the function of the native CLF allele subsequent experiments were conducted with the *35S:GFP:CLF* line in which GFP-CLF is complementing the lack of native CLF.

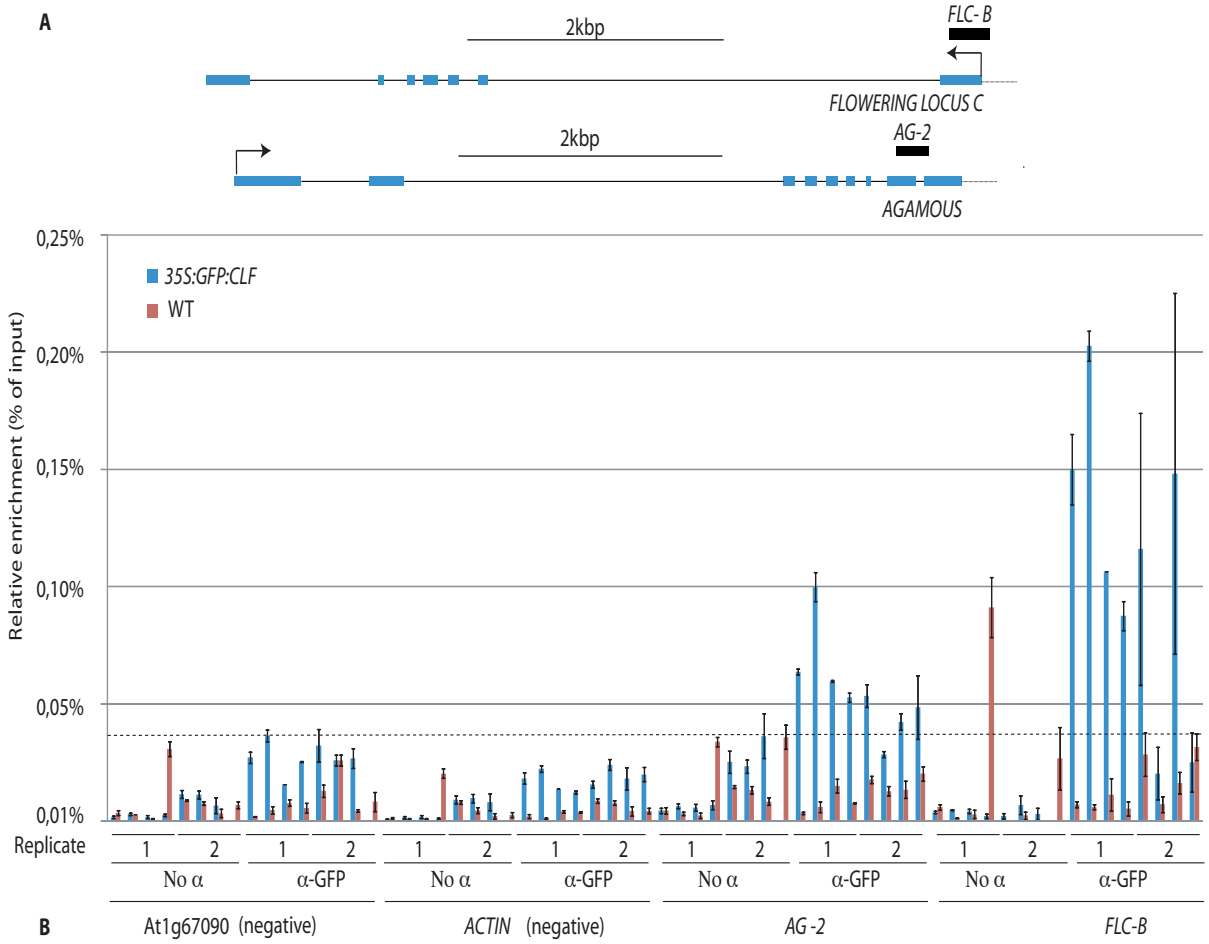
Two ChIPs, each with two biological replicates of each *35S:GFP:CLF* and wt were performed (8 ChIPs in total). Each of the eight chromatin extracts was used for two IPs with a-GFP and two IPs with No antibody (32 IPs in total). Binding of GFP:CLF in the H3K27me3 targets *FLC* and *AG* (positive controls) as well as the H3K27me3 free loci *Actin* and At1g67090 (negative controls) in each IP was tested by qPCR in three technical triplicates.

Although the results from the qPCRs were highly variable, I detected in IPs with a-GFP on *35S:GFP:CLF* samples a higher signal in *FLC* and *AG* compared to *Actin* and At1g67090 (Fig. 4.6 A).

To statistically confirm the observations of this experiment we conducted a Tukey test using R in which every condition is compared to all others (conducted with Benedict Drosse and Jonas Klasen, MPIPZ, Cologne). We defined 16 different conditions for each of the two ChIP experiments (Fig 4.6 B). The Tukey test was used to determine the distances and 95% confidence intervals between the enrichment obtained in two conditions (Fig. 4.6 C).

Condition 1 (*35S:GFP:CLF;* a-GFP; *FLC*) was statistically enriched compared to conditions 2-16 demonstrating that GFP-CLF is enriched at *FLC* and that this enrichment

is higher than that observed at *AG*). Condition 2 (*35S:GFP:CLF;* a-GFP; *AG*) is statistically enriched compared to conditions 3-16, indicating that GFP-CLF is binding to *AG*. Comparison of conditions 3 (*35S:GFP:CLF;* a-GFP; *ACTIN*) and 9 (*35S:GFP:CLF;* a-GFP; At1g67090) to conditions 4-16 and 5-8;10-16 respectively is higher than the distance in comparisons between conditions 5-8;10-16 between each other. These results suggest that ChIP-PCRs in regions that are not expected to be enriched for the protein of interest are the most stringent controls



Genotype	Antibody	Amplicon	Condition number	Control type
35S:GFP:CLF	anti-GFP	FLC-B	1	positive
35S:GFP:CLF	anti-GFP	AG-2	2	positive
35S:GFP:CLF	anti-GFP	At1g67090	3	negative
35S:GFP:CLF	No a	At1g67090	4	negative
wt	No a	At1g67090	5	negative
wt	a-GFP	At1g67090	6	negative
35S:GFP:CLF	No a	Actin	7	negative
wt	No a	Actin	8	negative
35S:GFP:CLF	a-GFP	Actin	9	negative
wt	a-GFP	Actin	10	negative
35S:GFP:CLF	No a	AG-2	11	negative
wt	No a	AG-2	12	negative
wt	a-GFP	AG-2	13	negative
35S:GFP:CLF	No a	FLC-B	14	negative
wt	No a	FLC-B	15	negative
wt	a-GFP	FLC-B	16	negative

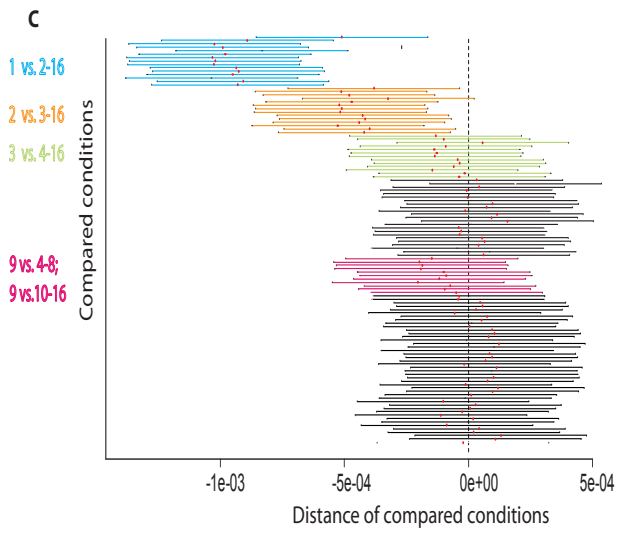


Figure 4.6: **Enrichment of GFP:CLF on *FLC* and *AG* is statistically significant compared to loci where no enrichment is expected which are the most stringent negative controls.**

Figure 4.6:
A: ChIP-PCR analysis of CLF-HA compared to the H3K27me3 signal at *AG*, *FLC*, *ACTIN* and At1g67090 loci and schematic representation of *AG* and *FLC*. *35S:GFP:CLF* samples (in blue) with a-GFP displays in most PCRs higher enrichment at *AG* and *FLC* compared to the negative controls At1g67090, *ACTIN* and IPs performed with wt samples (in red) or in the absence of antibody (no-antibody). Exons are illustrated as blue boxes, introns as black lines and positions of amplicons used for ChIP analysis as black labeled boxes. Dotted line represents region upstream of the TSS.
Polyclonal a-GFP (Abcam) from rabbit was used for IPs on crosslinked chromatin extract from 14-day-old seedlings grown on soil. Enrichment is presented relative to chromatin input (% input). The data are based on one experiment with two biological replicates per genotype. Error bars represent standard error of the mean. Dashed line represents background, calculated as maximum of qPCR signals of negative controls.
B: Table of tested conditions. Each of the two ChIP experiments was divided in 16 conditions. Positive controls in *35S:GFP:CLF* with a-GFP on *FLC* and *AG* are highlighted in blue and orange respectively. Positive controls in *35S:GFP:CLF* with a-GFP on *FLC* and *AG* are marked in green and purple respectively.
C: Plot of mean distances of differences between the enrichment in two conditions depicted in A. *FLC* and *AG* in *35S:GFP:CLF* with a-GFP antibody (a-GFP) (conditions 1 in blue and 2 in orange respectively) differ in average more from ChIP-PCRs at *FLC, AG, ACTIN* and At1g67090 in *35S:GFP:CLF* without antibody (No-a) or wt samples probed with a-GFP or without antibody (conditions 5-8;10-16 in gray). ChIP-PCRs at the H3K27me3 free *ACTIN* and At1g67090 in *35S:GFP:CLF* with a-GFP antibodies (conditions 3 in green and 9 in purple) differ in average less from ChIP-PCRs at *FLC* and *AG* in *35S:GFP:CLF* with a-GFP (blue and orange) than other ChIP-PCRs (gray). Controls for low enriched loci are the most stringent negative controls in ChIP-PCR experiments.
Red dots indicate the average value of a ChIP-PCR and lines represent the 5 and 95% confidence interval. The distance between the dotted line and average of a ChIP-PCR corresponds to the difference between two compared ChIP-PCRs. Statistical analysis was performed with R using the tukey test.

4.2.3 ChIP-seq of GFP-CLF

1. Library preparation

Binding of GFP-CLF to chromatin could be statistically confirmed, therefore I wanted to assess the genome wide binding sites of GFP-CLF. Because the ChIPed DNA amount was very low, the library preparation protocol must be extremely optimized. The short reads (100bp) that are produced by the HiSeq2000 (Illumina) platform are sufficient to sequence and correctly map ChIP-DNA fragments. In addition, up to 120 Gbp can be sequenced in one run with the HiSeq2000. Because of the high amount of reads per run (120 Gbp) many samples can be sequenced at once. Compared to SOLiD (ABI; 3Gbp/per run) and 454 (Roche 0,6 Gbp/per run) it is also the cheaper system. Therefore, I choose to sequence with the HiSeq2000 apparatus. In previous ChIP-Seq experiments against H3K27me3 it was depicted that from ~20 Mbp gained reads, approximately 1 Mbp uniquely mapped to the Arabidopsis genome and were sufficient to determine genome wide H3K27me3 allocation (PhD thesis, Dr. Xue Dong, MPIPZ,

Cologne). To ensure a sufficient sequencing depth I decided to sequence 20 Mbp per library. Because in the HiSeq2000, up to 120 Gbp can be sequenced per lane, I decided to barcode the libraries. However, Illumina did not provide a barcoded library preparation kit for ChIP samples that is optimized for sequencing on a HiSeq2000 platform. In addition, the common ChIP library preparation protocols are optimized for 10ng ChIP-DNA. However, qPCR analysis of ChIP samples from Arabidopsis chromatin depicted that precipitated DNA quantity was ~0,01ng or lower. To find optimal conditions for library preparation that are compatible with the HiSeq2000 and barcoding I tested three library preparation protocols: 1. ChIP-Seq library construction protocol for 10ng ChIP-DNA using the Illumina TrueSeq adapters (http://ethanomics.wordpress.com/protocols/) . 2. BiooScientific NEXTflex ChIP-Seq Kit for 10ng ChIP-DNA. 3. Nugen Ovation Ultralow DR Multiplex Kit for 1ng ChIP-DNA. To adjust to the ChIP DNA quantity of ~0,01ng, protocols 1 and 2 were tested as described in the instructions but also with dilutions of the adapters (1/250, 1/500, 1/750 and 1/1000) to avoid formation of primer dimers. Tests were performed with 1/10, 1/100, 1/1000 dilutions of the input. Protocol 3 was tested with undiluted, 1/10 and 1/100 diluted adaptors using a 1/100 dilution of the input. Amplification rate before and after library preparation was tested by qPCR. Fragment size distribution and relative concentration was determined with a Bioanalyzer (Agilent) apparatus. Because 18 PCR cycles were used to amplify the ChIP-DNA libraries, a maximum of 262.144 fold amplification was expected. However, in experiments conducted with approach 1 and 2, no amplification could be determined. 20 - 1000 fold amplification was measured in libraries prepared with the Nugen Ovation Ultralow DR Multiplex Kit. The highest amplification was detected using 1/10 diluted adaptors. After amplification, fragments were distributed in sizes between 200-1500bp. The Nugen Ovation Ultralow DR Multiplex Kit was used for further experiments, since it was the only protocol with which libraries could be amplified.

2.ChIP-Seq of GFP-CLF

Four biological replicates of *35S:GFP:CLF* using a-GFP antibody were processed in parallel ChIP experiments. Three IPs of each ChIP were pooled to increase the starting DNA quantity for library construction (Fig. 4.7). Simultaneously processed ChIPs with the same antibody on wt seedlings were used as a control to determine unspecific binding of the antibody. GFP-CLF enrichment was confirmed by ChIP-PCR and two samples that depicted the highest GFP-CLF levels and one wt sample were used for library construction. To identify further regions of unspecific immunoprecipitation, additional libraries were created with a 1 / 100 dilution of the input that corresponded approximately to 1ng of DNA. To control library construction efficiency, the increase of expected fragments after library construction was determined by qPCR. All samples displayed at least 10fold enrichment before sequencing. The four barcoded probes

were sequenced on a single lane of an Illumina HighSeq2000 flow cell (performed by the Genomecenter, MPIPZ, Cologne).

12.979.309, 7.526.430, 9.602.170 and 7.005.033 high quality reads were identified in the two *35S:GFP:CLF* replicates, wt and *35S:GFP:CLF* input respectively (Table 4.1; complete bioinformatic analysis performed by Dr. Xue Dong). The received reads were mapped to the TAIR10 Arabidopsis reference genome with BWA allowing a maximum of 3 mismatches, including 1 gap (Li & Durbin, 2009). SAMtools and Picard were used to pick out reads that mapped specifically and non ambiguous to the Col-0 genome (Li *et al.* (2009) and http://picard.sourceforge.net/index.shtml). 6.502.657 and 4.045.771 reads were mapped to TAIR10 for *35S:GFP:CLF* sample 1 and 2 respectively. For wt 3.286.172 and for input samples 2.172.910 reads could be mapped to unique positions respectively. After removing redundant reads that mapped to the same position in the Col-0 genome 5.731.886, 3.269.130, 2.787.545 and 1.039.437 specific reads were kept for the four samples. Regions that were GFP-CLF enriched were identified with the SICER software and signal from the wt (a-GFP) and input samples were used as background control (Zhang *et al.*, 2007b). Any region that was GFP-CLF enriched and overlapped with a gene was used for further analysis and determined as CLF-enriched. Enriched regions that mapped to intergenic regions were not analyzed further. identified CLF-enriched regions were mapped to TAIR10 gene annotation and data was uploaded to the Integrated Genome Viewer (IGV) browser (http://www.broadinstitute.org/igv/).

To determine statistically differentially enriched regions in *35S:GFP:CLF* compared to wt, a two step approach was applied using SICER: 1. CLF-enriched regions with a minimum size of 200 bp and gaps smaller than 400 bp were combined into islands. 2. differentially enriched islands were determined. 857 genes were identified in replicate 1 and 890 in replicate 2. 166 common target genes could be identified in the two replicates, indicating a high amount of false positives (Fig. 4.8 A). Notably, a recent study suggested that the overlap between to ChIP-Seq replicates should at least 75% to account the detected peaks as trustworthy (Furey, 2012). No GFP-CLF binding could be detected at the known GFP-CLF binding loci *AG*, *FLC*, *FT* and *SEP3*. When a set of the 166 common target genes was analyzed using the genome visualization tool IGV browser, marginal difference could be determined between replicate 1, 2 and wt (exemplified on AT2G21770, Fig. 4.8 B, Thorvaldsdóttir *et al.* (2013)). Moreover, I could observe high similarities in the GFP-CLF signal in replicate 1, 2 and wt (exemplified on chromosome 2, Fig. 4.8 C). In conclusion, these results imply that no true enrichment of GFP-CLF was detected in the performed ChIP-Seq experiment.

.

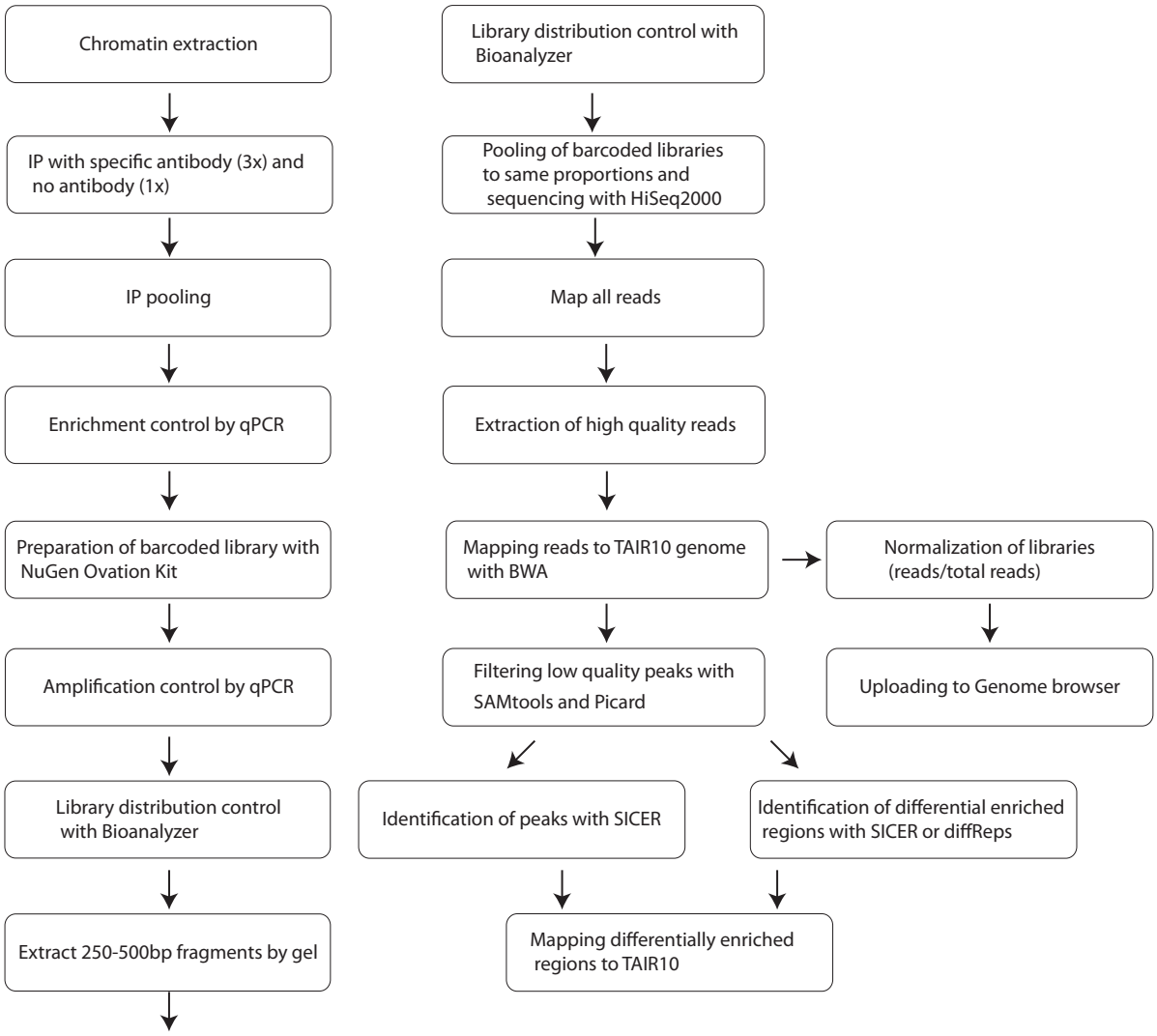


Figure 4.7: **Work flow of ChIP-seq experiment**

Table 4.1: **Number of reads of two *35S:GFP:CLF* replicates after sequencing and analysis.**

	***35S:GFP:CLF* 1**	***35SGFP:CLF* 2**	**wt**	***35S:GFP:CLF* input**
Total reads	12.979.309	7.526.430	9.602.170	7.005.033
After filtering reads mapping to multiple position	6.502.657	4.045.771	3.286.172	2.172.910
Unique reads	5.731.886	3.269.130	2.787.545	1.039.437

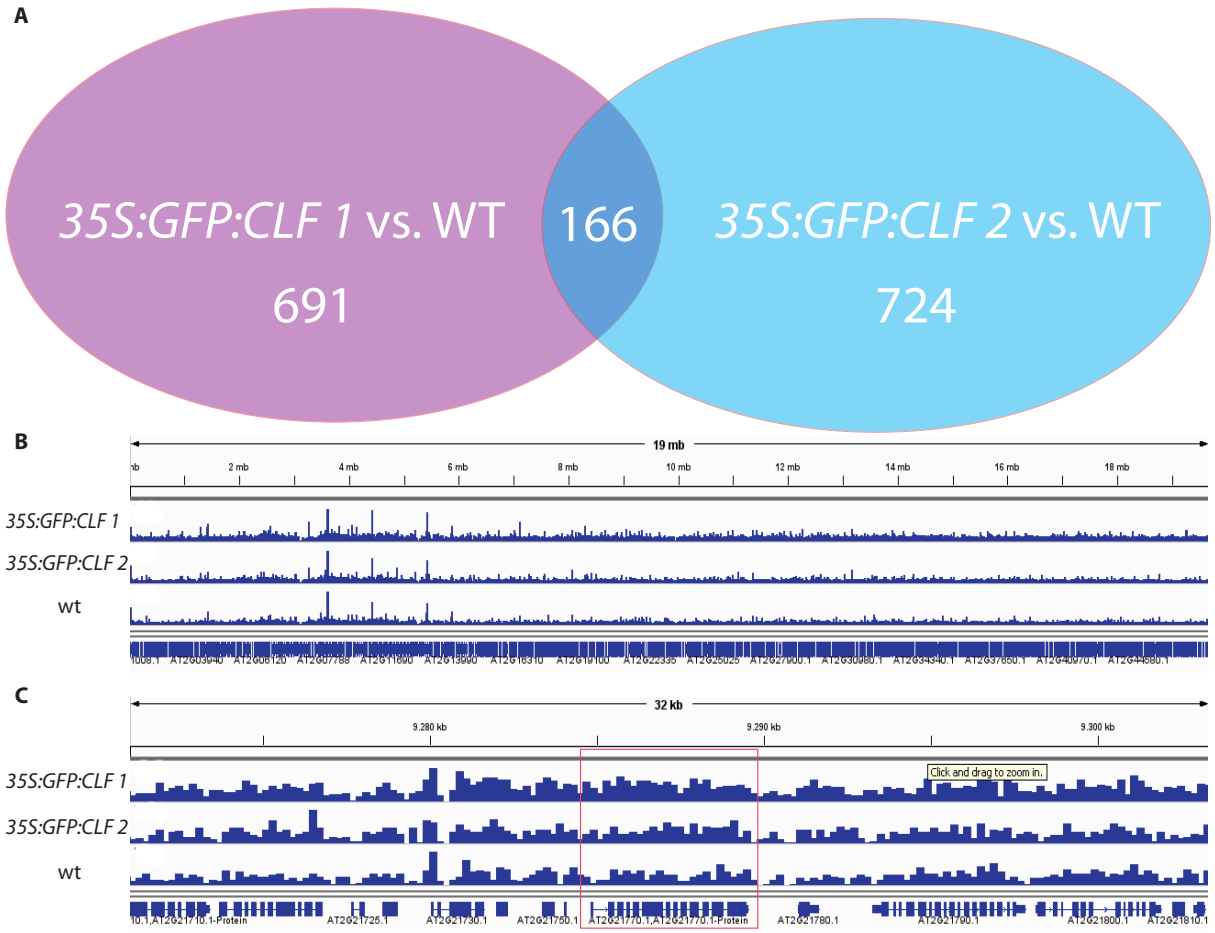


Figure 4.8: **High similarity between positive and negative controls in ChIP-Seq of *35:GFP:CLF***
A:Venn-Diagram of ChIP-Seq analysis. 166 commonly GFP-CLF enriched genes could be identified in *35:GFP:CLF* replicate 1 (purple) and 2 (blue). .
Analysis was performed with SICER 1.1. CLF-enriched regions with a minimum size of 200 bp and gaps smaller than 400 bp were combined into islands and 2fold differentially enriched islands (FDR:0,001) were determined. Polyclonal a-GFP (Abcam) from rabbit was used for IPs on crosslinked chromatin extract from 14-day-old seedlings grown on soil. Similarly treated wt samples were used as controls for background enrichment.
B: Integrated Genome Viewer image of *35S:GFP:CLF* replicate 1, 2 and wt (from top to bottom) on chromosome 2. Signal distribution on Chromosome 2 is highly similar between *35S:GFP:CLF* and wt.
C: Integrated Genome Viewer image of *35S:GFP:CLF* replicate 1, 2 and wt (from top to bottom) on AT2G21770 (red square). AT2G21770 is not differentially enriched between *35S:GFP:CLF* and wt as determined by SICER (panel A).

4.2.4 Proteasome inhibitor degradation enhances slightly the abundance of CLF but ChIP efficiency of GFP:CLF does not increase

Although statistical dissection could confirm that GFP-CLF is bound to chromatin, the low abundance of DNA fragments after ChIP was probably not sufficient to subsequently create DNA-libraries for sequencing with a next generation sequencer (Fig.

4.6). Recent publications revealed that CLF is negatively regulated by the E3-ligase UPWARD CURLY LEAF 1 (UCL1) at the post translational level (Jeong *et al.*, 2011). Furthermore, proteins that are associated with the protein degradation machinery were identified in a screen for enhancers of *clf* (personal communication, Justin Goodrich, Edinburgh). Finally, the results in section 4.2.1 suggested that CLF might be regulated at the post-translational level. Therefore, the data in section 4.2.1 indicated that CLF might be regulated by the 26S-Proteasome. Consequently treatment of plants with the proteasome inhibitor MG-132 could increase CLF protein abundance in such a way that ChIP efficiency would be higher and ChIP-DNA-library creation facilitated.

35S:GFP:CLF plants were infiltrated and subsequently incubated for 24h with MG-132. The intensity of nuclear signal for GFP-CLF in *35S:GFP:CLF* was significantly stronger than in mock treated plants (Fig. 4.9 A 1, 5 and B). Unexpectedly, GFP-CLF was also more abundant in the cytoplasm of MG-132 treated plants. To confirm these observations, I tested if MG-132 treatment increased CLF-HA abundance to similar levels as has been reported for the positive control 3xHA-CONSTANS (3xHA-CO) (personal communication Samon Simon, MPIPZ, Cologne). Compared to the positive control only a weak increase of CLF-HA could be detected (Fig. 4.9 C;)

To assess if inhibition of the 26S-Proteasome can increase GFP-CLF binding to chromatin, a ChIP-PCR assay with MG-132 and mock treated seedlings (24h) was conducted (Fig. 4.9 D). Accordingly with previous experiments, ChIP-PCRs on *35S:GFP:CLF* samples with a-GFP antibodies that were incubated with the proteasome inhibitor displayed strong variation in signal intensity and no significant increase in binding to target regions was detected (Fig. 4.9 D; data not shown).

These experiments lead to the conclusion that inhibition of 26S-Proteasome leads to increased GFP-CLF in the nucleus. However, ChIP efficiency is not increased in such a way that the assay can be used to immunoprecipitate a superior amount of DNA to create DNA-libraries.

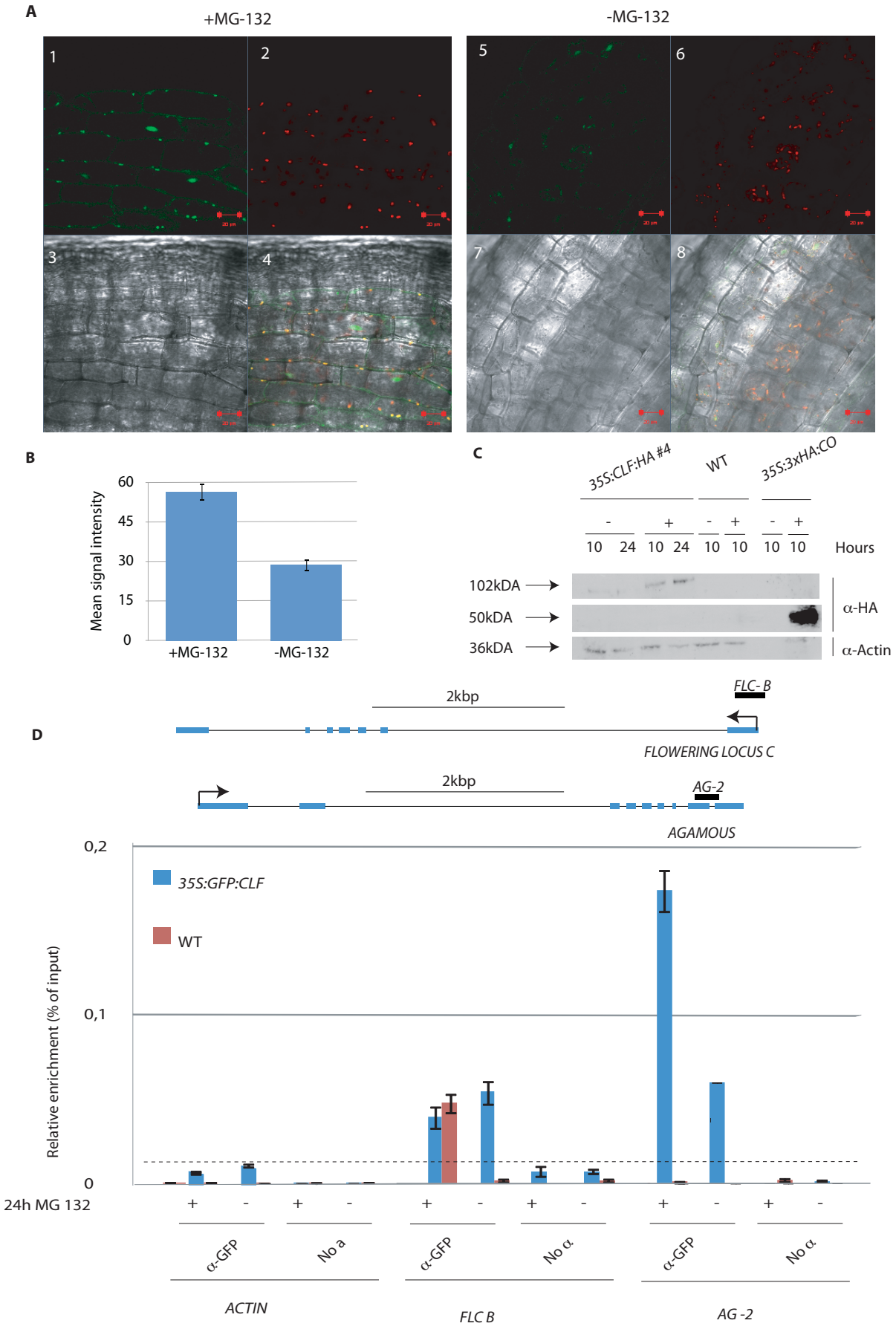


Figure 4.9: **MG-132 treatment increases abundance of CLF**

Figure 4.9:
A: Confocal microscope analysis of hypocotyls. GFP-CLF displays stronger signal intensity after treatment with MG-132 than mock treated *35S:GFP:CLF* plants. Green filter (1, 5), red filter (2, 6), bright field (3, 7) and merge (4,8) are depicted. 20 μm scale bar is illustrated at the bottom of each panel.
14-day-old *35S:GFP:CLF* seedlings were incubated for 24h with MG-132 (+) or DMSO as mock treatment (-). Representative image is presented.
B: Mean signal intensity of GFP-CLF. Signal intensity of nuclei illustrated in image B1 (+MG132) and B5 (-MG132) was determined.
Quantification was performed with ImageJ.
C: Western blot analysis. CLF-HA abundance in *35S:CLF:HA* increases after 10 and 24 hours of MG-132 treatment. The positive control 3xHA-CONSTANS displays increase in abundance after 10 hours or MG-132 treatment. No signal was detected in wt.
Analysis was performed with total extract of 14-day-old seedlings incubated for 10 or 24 hours with MG-132 (+) or DMSO (-). Membranes were probed with a-HA (Sigma), stripped and reprobed with a-Actin. Actin was used as a loading control. A representative membrane is presented.
D: ChIP-PCR analysis and schematic representation of *AG* and *FLC*. GFP-CLF is slightly more enriched at *AG* after 24h of MG-132 (+) incubation compared to DMSO as mock treatment (-). No increase could be observed on *FLC*. Exons are illustrated as blue boxes, introns as black lines and positions of amplicons used for ChIP analysis as black labeled boxes.
Polyclonal a-GFP antibodies (Abcam) and a-H3K27me3 (Millipore) from rabbit were used for IPs on crosslinked chromatin extract from 14-day-old seedlings grown on soil. Enrichment is presented relative to chromatin input (% of input). Error bars represent standard error of the mean. Dashed line represents background, determined as maximum of qRT-PCR signal of negative controls (*ACTIN* and absence of antibody (No-a). One representative experiment is depicted.

4.3 H3K27me3 profiles of *clf* and *swn* mutants

4.3.1 High overlap between H3K27me3 targeted genes in *clf*, *swn* and wt

The results in section 4.1.2 demonstrated that the CLF and SWN proteins have conserved differences in their catalytic SET domain. The two proteins are also strikingly different in their middle part, where CLF relatives depict a conserved domain that is not shared by SWN-likes. This indicated that they could have different catalytic properties and potentially different target genes, binding partners or other unknown functions. Moreover, studies on the tomato CLF and SWN homologues have indicated, these proteins have distinct target genes (How Kit *et al.*, 2010). Nevertheless, CLF and SWN have seven common domains, partially redundant function and overlapping target genes (Chanvivattana *et al.*, 2004). All so far tested known PcG target genes have lost H3K27me3 in the *clf/swn* double mutant seedlings while H3K27me3 was only reduced in *clf* (Farrona *et al.*, 2008). Moreover, ChIP experiments on CLF illustrated that it is difficult to determine the direct CLF target genes (Section 4.2.4). I hypothesized that if CLF and SWN are required for the trimethylation of H3K27 at a certain region, this mark would be absent in the majority of targets in the respective mutants. Furthermore, I considered the fact that ChIPs performed with H3K27me3 antibody depict at least 10-fold higher amount of precipitated chromatin than ChIPs performed with a-GFP in *35S:GFP:CLF* as a technical advantage for library preparation and sequencing (Section 4.2.2). Therefore, I decided to determine the differences in the roles of CLF and SWN in H3K27me3 deposition via ChIP-seq and identify individual target genes, by comparing H327me3 distribution maps in *clf*, *swn* and wt. I took into account that a minority of H3K27me3 target regions might also depict differential H3K27me3 due to indirect effects. Nevertheless, apart from possible indirect effects, the H3K27me3 pattern that results from CLF or SWN should be visible in the *swn* and *clf* mutant respectively.

The same procedure as described in section 4.2.3 was performed using the a-H3K27me3 instead of a-GFP antibodies (Fig. 4.7). After peak detection and identification of enriched regions with SICER we considered genes that were covered at a proportion of 20% or at least 500 bp as positive for H3K27me3. Three biological replicates of each, *clf*, *swn*, wt and the corresponding inputs were sequenced and analyzed. Inputs from each sample were also sequenced for later background calculation (Table 4.3). After identifying the common H3K27me3 target genes between the three biological replicates of each genotype among the 6.611 genes that were enriched for H3K27me3 in wt 5.513 and 6.256 genes where still positive for this mark in *clf* and *swn*, respectively (Fig. 4.10). These results illustrate that the absence of CLF or SWN has a minor impact on the set of H3K27me3 targeted genes, suggesting that CLF and SWN can partially compensate for each others function in H3K27me3 deposition. The few changes in

Table 4.2: **Definitions of terms.**

Feature name	Description	Section determined
Peak	Region that is enriched for H3K27me3	4.3.1
Island	Connected H3K27me3 enriched regions with intervening gaps smaller than 400bp	4.3.1
Type	Pattern of H3K27me3 abundance within an island	4.3.1
Block	H3K27me3 covered region larger than 10 kb	4.4.3
Enriched window	0.2Mio bp window enriched for CLF target genes	4.4.3
Cluster	Sum of CLF target genes that have a common expression pattern in wt *clf*, *swn* and *clf/swn*	4.4.2
Category	Group of clusters based on the expression pattern in wt *clf*, *swn* and *clf/swn*	4.4.2
CLF dependent gene	Gene that has reduced H3K27me3 in *clf* compared to wt	4.4.1
SWN affected gene	Gene that has enriched H3K27me3 in *swn* compared to wt	4.4.1

H3K27me3 positive genes reflect the mild phenotypic differences in *clf* compared to wt. High overlap between *swn* and wt H3K27me3 target genes mirrors the absence of differences between the two genotypes. 5.331 were shared between *clf* and wt while 6.068 between *swn* and wt. 5.088 genes are commonly marked by H3K27me3 in *clf* and*swn* but not in wt.

From the total genes that were trimethylated at H3K27 in the three genotypes, 5063 (73 %) were found in *clf*, *swn* and wt. Among these, I could identify the major flowering time regulators *AG*, *FLC*, *FT* and *SEP3*. It has been reported that these genes are the key genes mediating the *clf* phenotype and partially loose H3K27me3 in the *clf* mutant (Lopez-Vernaza *et al.*, 2012). Importantly, it is not clear whether the H3K27me3 reduction in *clf* is global or if it follows a particular pattern. It is assumed that because H3K27me3 is not completely absent in *clf*, SWN can partially deposit H3K27me3 at *AG*, *FLC*, *FT* and *SEP3*. Because these genes are marked by H3K27me3 but differentially expressed in *clf*, the SWN mediated H3K27me3 pattern might be different than the H3K27me3 pattern resulting from CLF and SWN.

157 and 163 unique target genes were identified in *clf* and *swn* respectively. Analysis in the genome visualization tool Gbrowse suggested that these genes are false positives (Stein, 2013).

Taken together, mainly the same genes are targeted by H3K27me3 in wt and upon loss of CLF or SWN. The change of genes that are targeted by H3K27me3 in *clf* and *swn* compared to wt correlates with the phenotypic changes in the mutants compared to wt.

Table 4.3: **Number of reads of three replicates of clf, swn and wt after sequencing and analysis.**

	Replicate	*clf*	*swn*	wt	*clf* input	*swn* input	wt input
Total reads	1	5.073.076	13.620.052	5.752.201	6.232.953	6.029.780	755.711
	2	8.147.923	18.226.173	15.751.513	9.701.620	13.873.387	5.412.440
	3	18.397.718	16.186.004	16.459.698	7.088.752	8.452.724	2.535
After filtering reads	1	3.913.176	3.033.544	4.818.883	3.911.846	3.556.269	288.940
mapping to multiple	2	6.499.050	6.949.973	13.164.325	6.311.815	9.581.492	3.330.670
position	3	13.680.070	14.527.816	13.945.572	4.659.913	5.909.930	1.536
Unique reads	1	3.569.656	2.871.292	4.423.631	3.620.300	3.556.269	230.901
	2	2.880.297	3.325.801	10.207.892	5.201.891	8.253.646	2.343.016
	3	11.351.279	11.769.8711	10.999.594	4.229.224	4.816.543	1.091

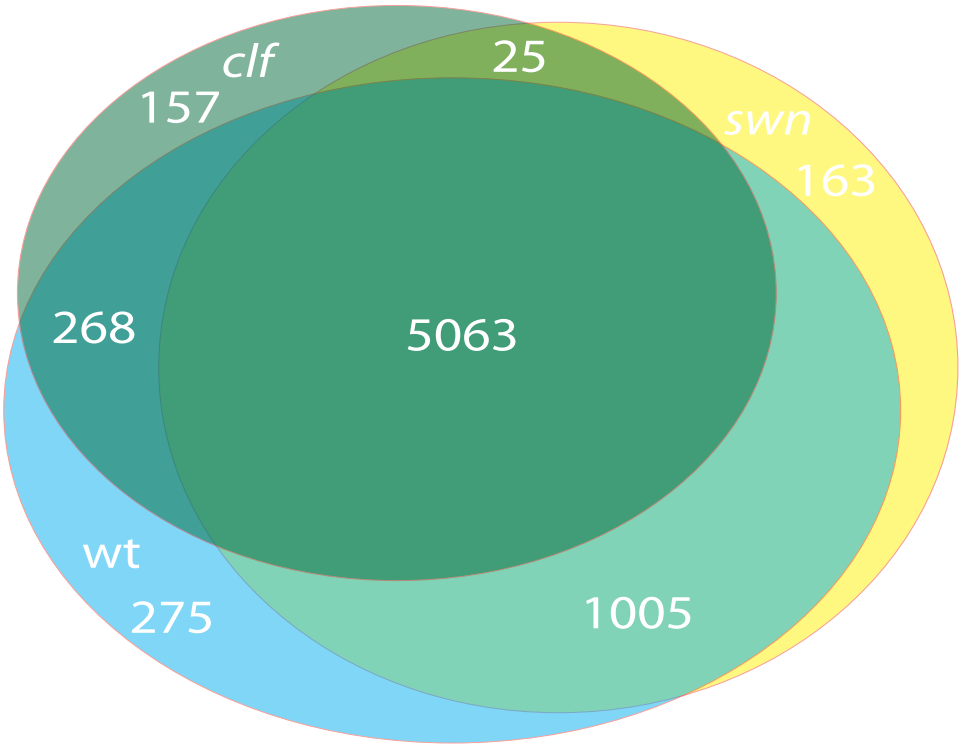


Figure 4.10: **High overlap between H3K27me3 target genes in *clf* and *swn* and wt plants.**
Venn-Diagram of ChIP-seq analysis. Less genes are positive for H3K27me3 in *clf* (5513, green) compared to *swn* (6256, yellow) and wt (6611, blue). 20 percent of a gene or at least 500 bp have to be covered by H3K27me3 to be determined as positive. Analysis was performed with SICER 1.1. Consensus gene lists of three independent experiments are presented.

4.3.2 *clf*, *swn* and wt have different H3K27me3 distribution patterns

Since the three genotypes shared many target genes, I wanted to know if CLF and SWN are important for specific patterns of H3K27me3 distribution. I speculated that

H3K27me3 abundance might be changed at certain loci in *clf* or *swn*, indicating that these loci rely on CLF or SWN for wt-like H3K27me3 coverage. To assess this question, the data was normalized by dividing the unique reads at each position through the total number of unique reads of each library and then uploaded to a custom Gbrowser for empirical analysis (performed by Julia Engelhorn, CEA, Grenoble). Because I observed that the H3K27me3 target genes of *clf*, *swn* and wt are highly overlapping I wanted to avoid loosing resolution of the H3K27me3 patterns by merging the three biological replicates and treated samples separately. The amount of unique reads was highly variable between the biological replicates and genotypes (Table 4.3). However, in the third biological replicate *clf* and *swn* and wt had an equal amount of total unique reads and were therefore comparable after normalization. Thus the third biological replicate was used for further analysis in Gbrowse. Importantly, patterns identified in replicate 3 were usually corroborated in the other replicates (data not shown).

H3K27me3 positive regions were divided into islands (Table 4.2). Each island consisted of a continuous H3K27me3 enriched region that had maximum gaps of 400 bp. After empirical analysis of roughly 2000 H3K27me3 marked regions, the islands were divided into three major types, depending of the H3K27me3 ratio between *clf*, *swn* and wt (Fig. 4.11):

Type 1 (CLF related): H3K27me3 abundance in *swn* is similar or slightly reduced compared to wt and substantially reduced in *clf* (Fig. 4.11 A). In most islands, one or more peaks of H3K27me3 are preserved in *clf*. However, these peaks are diminished in *clf* compared to *swn* and wt. CLF is probably important to fully cover these islands with H3K27me3.

Type 2 (SWN related): H3K27me3 abundance in *swn* is higher than wt and *clf* (Fig. 4.11 B). In these regions, the interplay between CLF and SWN is probably important to cover these islands properly with H3K27me3.

Type 3 (CLF or SWN related): H3K27me3 abundance is similar in *clf*, *swn* and wt (Fig. 4.11 C). At these regions either CLF or SWN are sufficient to deposit H3K27me3 properly.

Among the regions that were positive for H3K27me3, type 1 regions were most abundant, including *AG*, *FLC*, *FT* and *SEP3*. Several additional types of islands were observed, yet they were infrequent. Thus, I mainly focused on the three major island types.

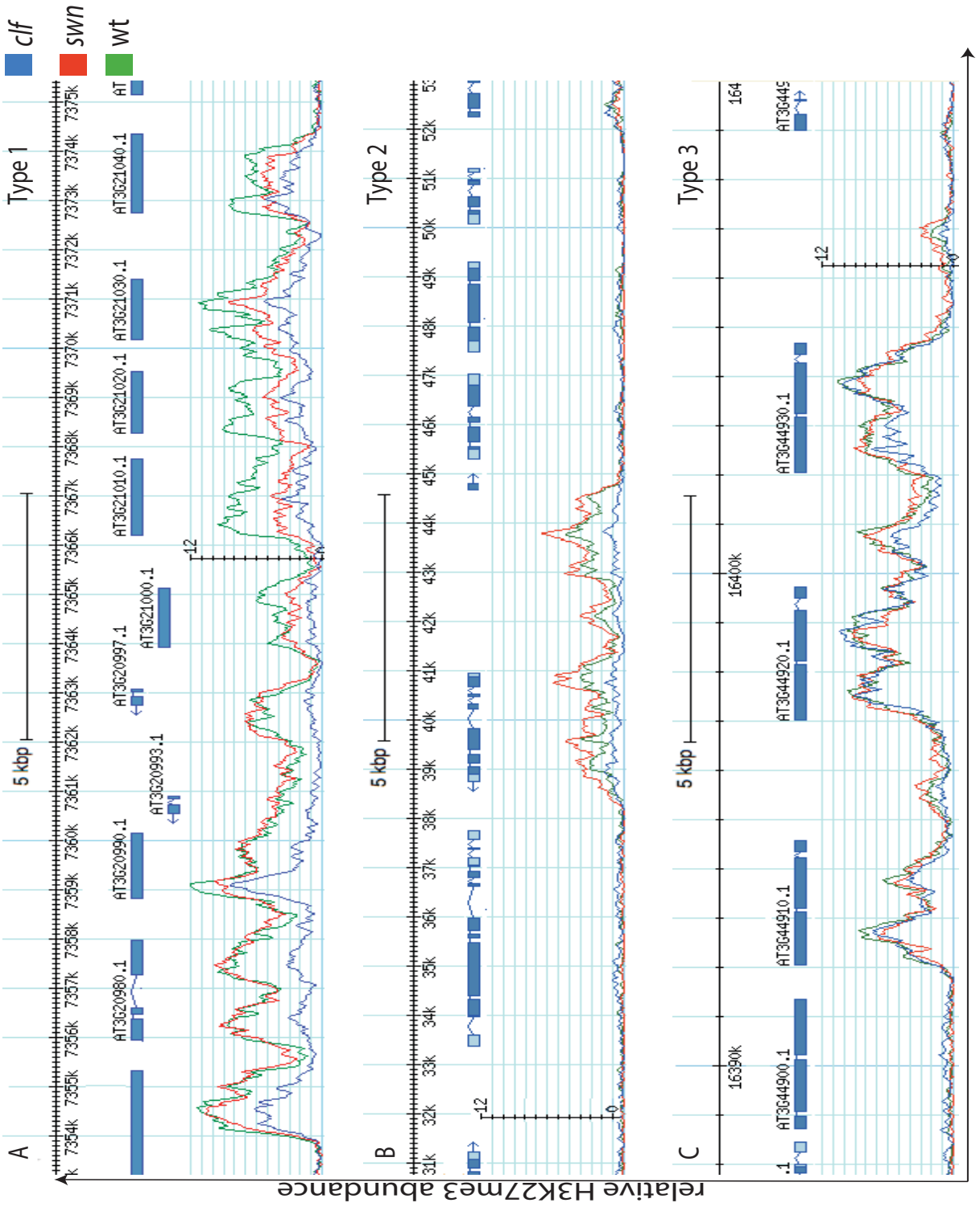


Figure 4.11: **Three types of H3K27me3 distribution patterns between *clf*, *swn* and wt.**

Figure 4.11: Genome browser view of three exemplary regions.
A: Type 1 pattern (CLF related). The H3K27me3 abundance of *clf* is strongly reduced in the non peak regions and slightly reduced at peak regions; *swn* displays a similar signal as wt.
B: Type 2 pattern (SWN related). *clf* H3K27me3 allocation resembles the type 1 pattern while H3K27me3 in *swn* is slightly more abundant than in wt.
C: Type 3 (CLF or SWN related). *clf*, *swn* and wt have similar H3K27me3 profiles.
Three exemplary regions of replicate 3 are presented. H3K27me3 levels are calculated as reads per million reads (total reads). *clf* is marked in blue, *swn* in red and wt in green.

To confirm the ChIP-seq results and the observed pattern differences, ChIP-PCR analysis of genes that were representative for each observed type was performed. ChIP experiments accumulate a high amount of extract -specific bias and results can vary depending on the normalization method. This makes minor differences between samples difficult to assess. In ChIP-PCR, results can vary depending on the normalization method. Because in ChIP-seq experiments, approximately same quantities are sequenced and individual loci are normalized against the total amount of reads, the bias at single loci is low compared to ChIP-PCR if the amount of unique reads is similar (personal communication, Julia Engelhorn, CEA, Grenoble; discussed in Section 5.2.3). Three normalization approaches were used in ChIP-PCRs to overcome the limitations of this approach.

1. Normalization against the input of each genotype.

2. Normalization against a parallel ChIP performed against unmodified H3 for each genotype.

3. Normalization against an internal locus of each genotype.

Based on ChIP-seq results, At4g24420 had similar H3K27me3 abundance between *clf*, *swn* and wt. In ChIP-PCRs normalized against against the input or H3 this absence of change in H3K27me3 was confirmed. Therefore, subsequent ChIP-PCRs were normalized against At4g24420 (Fig. 4.12 A, B).

Three biological replicates were averaged and differences between *clf*, *swn* and wt were statistically assessed using the students T test. I could confirm enrichment at *FLC* and the absence of H3K27m3 at At1g67090 (Adrian *et al.*, 2010; Angel *et al.*, 2011) (Fig. 4.12 C, D). Strikingly, the H3K27me3 mark at *FLC* was strongly varying between between replicates compared to the other tested loci. As in ChIP-seq, H3K27me3 at *Ga2ox2* is significantly higher in *swn* than in *clf* or wt (Fig. 4.12 A, C). Higher H3K27me3 abundance in *swn* compared to *clf* can also be confirmed at *Ga2ox7* and *GDSL*. These results underline, that H3K27me3 is partially increased in *swn*. Increased H3K27me3 signal in *swn* compared to wt at *BGAL7* and *EP-HYD* cannot be confirmed as statistically significant although the tendency is observed as in ChIP-seq. Although rarely occurring, reduction of H3K27me in *clf* and *swn* compared to wt can be confirmed by ChIP-PCR at *F-BOX*. In summary, in most experiments H3K27me3 enrichment and pattern changes between genotypes at the tested loci can be reproduced by

ChIP-PCR.

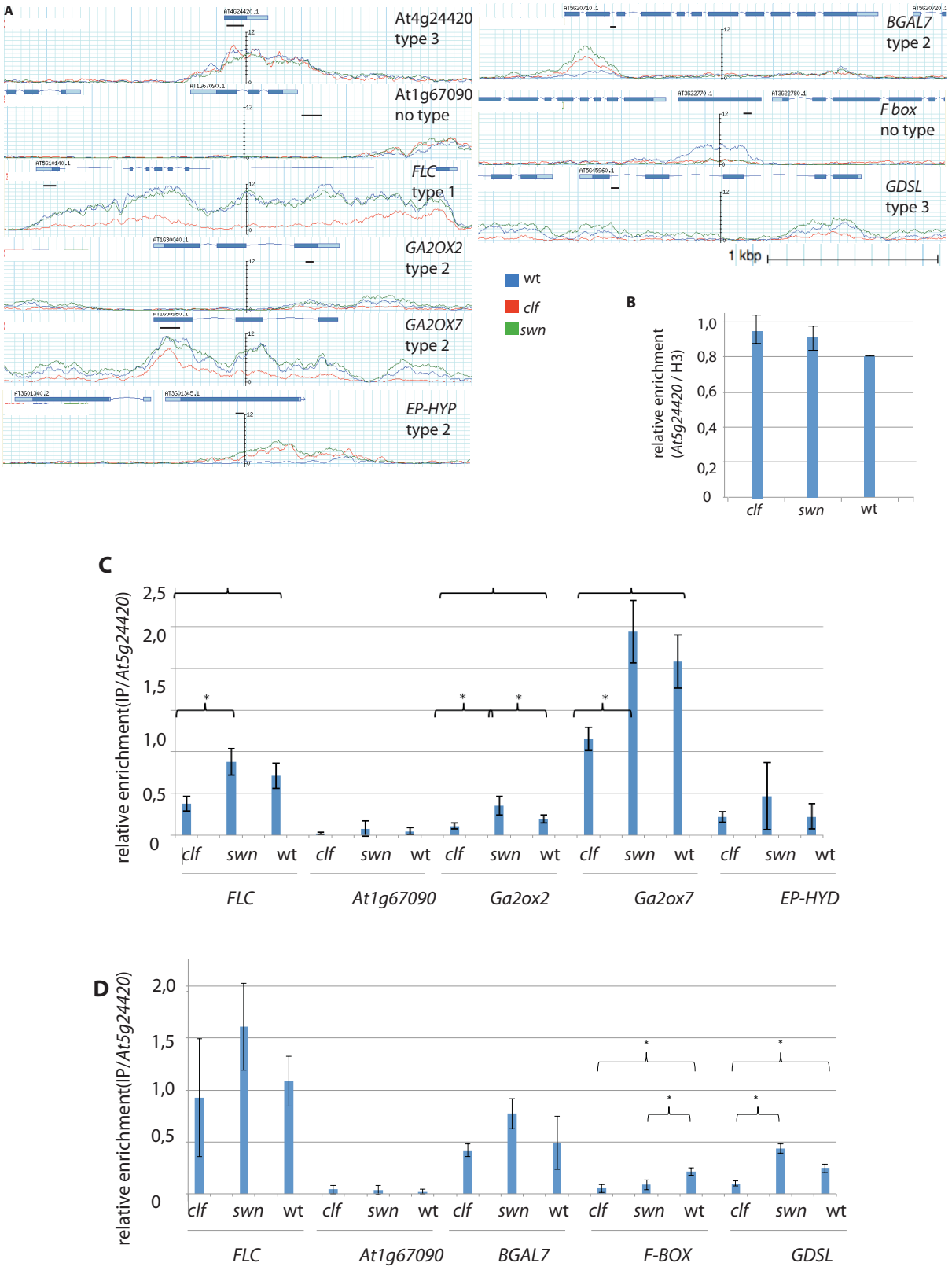


Figure 4.12: **Type of H3K27me3 distribution patterns can be confirmed by ChIP-PCR.**

Figure 4.12:
A: Gbrowser view of H3K27me3 enrichment in *clf* (red), *swn* (green) and wt (blue). At4g24420 is similarly enriched between *clf*, *swn*. At1g67090 is not enriched in *clf*, *swn* and wt. *FLC*, *Ga2ox2*, *Ga2ox7*, *EP-HYD*, *BGAL7*, *F-BOX* and *GDSL* are differentially enriched between *clf*, *swn* and wt.
B: ChIP-PCR of H3K27me3 enrichment in *clf*, *swn* and wt at At4g24420. At4g24420 is enriched at similar levels between *clf*, *swn* and wt and can be used for normalization. One representative experiment is illustrated. Relative abundance against H3 is plotted.
C: ChIP-PCR of H3K27me3 enrichment in *clf*, *swn* and wt. *Ga2ox2*, *Ga2ox7* and are significantly differentially enriched between *clf*, *swn* and wt and reflect relative abundance observed in ChIP-seq (A). H3K27me3 signal at *EP-HYD* is weak and variable. Average of 3 biological replicates is illustrated and error bars represent standard error of the mean. Signal intensity relative to At4g24420 is plotted. PCRs at *FLC* and At1g67090 were used as positive and negative controls respectively.
D: ChIP-PCR of H3K27me3 enrichment in *clf*, *swn* and wt. *BGAL7*, *F-BOX* and *GDSL* are differentially enriched between *clf*, *swn* and wt and reflect relative abundance observed in ChIP-seq (A). Average of 3 biological replicates is illustrated and error bars represent standard error of the mean. Signal intensity relative to At4g24420 is plotted. PCRs at *FLC* and *At1g67090* were used as positive and negative controls respectively.
Polyclonal a-H3 (Abcam) and a-H3K27me3 (Millipore) from rabbit were used for IPs on crosslinked chromatin extract from 14-day-old seedlings grown on soil. Error bars represent standard error of the mean. No enrichment was obtained with No - antibody control.

4.4 Analysis of CLF dependent genes

4.4.1 A small fraction of H3K27me3 target genes is significantly H3K27me3 depleted in *clf*.

clf, *swn* and wt share many target genes but in several regions H3K27me3 is reduced in *clf* compared to *swn* and wt (Section 4.3.2). Furthermore, several regions had elevated H3K27me3 in the *swn* mutant. To determine if these observations are statistically significant (FDR: 0,001) we used SICER to determine which genes are at least 2 fold decreased or increased between *clf*, *swn* and wt (Section 4.2.3; analysis performed by Xue Dong).

When *clf* and *swn* where compared to wt, 6575 genes were unchanged and 643 were decreased in *clf* versus wt (Fig. 4.13 A). These genes were assigned as CLF dependent genes (Table 6.1). 5 genes were increased in *clf* versus wt. Because the FDR was set at 0,001, comparisons were less than 20 genes were determined as differentially trimethylated at H3K27 were assigned as false positives. They were not used for further analysis.

Unexpectedly, although 163 genes were identified as uniquely marked by H3K27me3 in *swn* (Fig. 4.10), only a few changes in H3K27me3 were statistically significant when *swn* was compared to wt. 5 genes were decreased in *swn* compared to wt while 4 genes were more increased in *swn* than in wt. However, 825 genes had varying H3K27me3 abundance levels in *clf* compared to *swn* (Fig. 4.13 B). 294 of these genes were not included in the genes that were reduced in *clf* versus wt indicating that they have more H3K27me3 abundance in *swn* than wt. This could corroborate the observed islands where H3K27me3 was increased in *swn* and decreased in *clf* although these had not been difficult to confirm by ChIP-PCR (Section 4.3.2). I analyzed the relevant genes in Gbrowse and observed that they were located within type 2 regions. These genes were assigned as SWN affected (Table 6.2). 112 genes were uniquely reduced in *clf* versus wt. At these loci, probably CLF but not SWN can function to establish H3K27me3.

To test if genes differentially affected are functionally connected I conducted GO term analysis. CLF dependent genes were found to be enriched for "Cyclase activity" (p=0.000009078). No functional connection could be detected when SWN affected genes or the 112 genes that were uniquely reduced in *clf* (genes unaffected in *swn but reduced in clf*) were analyzed. When CLF dependent genes were analyzed for genes that are important for flowering time, two genes that are involved in the GA degradation pathway (*Ga2ox2*, *Ga2ox7*) and two genes that are involved the aging pathway were detected (*mir156b*, *SNZ*). *FLC* and *AG* were included in the list of genes that are CLF dependent, but unexpectedly *FT* and *SEP3* not. This suggested that the parameters that were chosen in the analysis of differential genes are very stringent. Genes where patterns were mildly affected were probably discarded. Adding to this, genes

that were increased in H3K27me3 in *swn* were identified only in the comparison of *clf* versus *swn* but not when *swn* was compared to wt. In summary, I demonstrate that most genes loose H3K27me3 in a CLF dependent manner and that in some regions genes gain H3K27me3 in a SWN dependent manner. Because the H3K27me3 changes in *swn* were marginal, these were not detected in the comparison to wt. Further analyses were focused on CLF dependent genes.

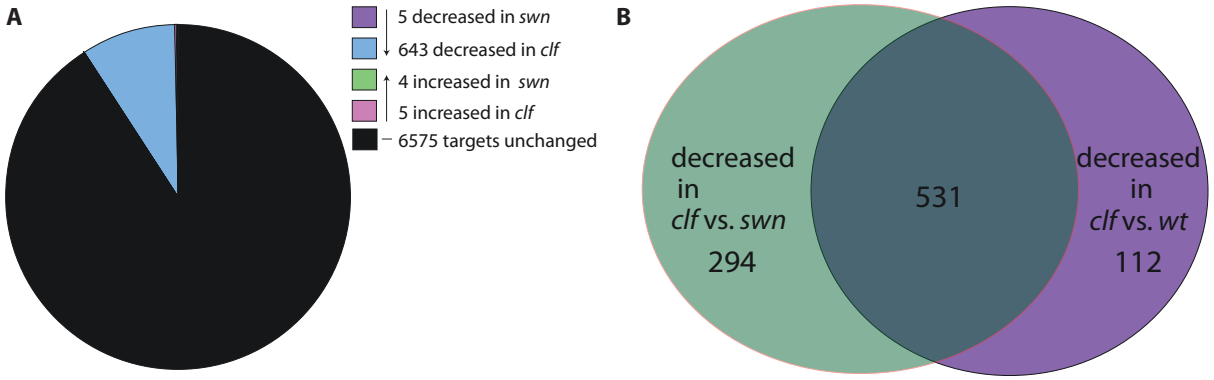


Figure 4.13: **Strong H3K27me3 decrease in *clf* and subtle increase in *swn* .**
A: Pie diagram of differentially trimethylated genes compared to wt. 10 % of the H3K27me3 target genes (black) are decreased in *clf* (blue) compared to wt. A few genes are increased in *clf* (pink) or differentially trimethylated at H3K27 in *swn* compared to wt (purple and green correspond to decreased and increased genes respectively).
B: Venn diagram of differentially trimethylated genes compared to *clf*. 294 genes appear as differentially methylated at H3K27me3 in *swn* when compared to *clf* (green) than when compared to wt. 112 genes are uniquely decreased in *clf* (purple) 531 genes *clf* and *swn* display changes. Only genes with an FDR of 0,001 and above 2fold change of H3K27me3 abundance were determined as differentially marked. Analysis was performed with SICER 1.1 using the SICER-df.sh script.

4.4.2 CLF and SWN function in combination or independently from each other as activators and repressors of CLF dependent genes.

In mammals, the function of the E(Z) homologue Ezh2 correlates with expression of Ezh2 target genes either positively or negatively in a cell dependent manner. This proves that Ezh2 can function as an activator and as a repressor of transcription (Xu *et al.*, 2012). In Arabidopsis, the effect of PcG-mediated regulation varies for target genes and for *FT*, H3K27me3 is not the major expression determinant (Farrona *et al.*, 2011). Reduction of H3K27me3 at *FT* in *clf* is linked to increased expression of this gene. Unexpectedly loss H3K27me3 in *clf/swn* results in decreased *FT* expression (Farrona *et al.*, 2011). However, the authors suggest that this is a indirect effect. In their model, they propose that the lack of chromatin-mediated repression leads to elevated expression only in the presence of additional positive regulators (Farrona *et al.*, 2011). Furthermore, H3K27me3 directly represses only specific transcription factor families,

but indirectly activates others through H3K27me3-mediated silencing of microRNA genes (Lafos *et al.*, 2011).

I wanted to further understand the role of CLF and H3K27me3 in the regulation of gene expression. Therefore I analyzed the expression of CLF dependent genes (reduced in H3K27me3 in *clf*) in wt, *clf*, *swn* and *clf/swn* plants using a previously published dataset (Farrona *et al.*, 2011). From the 643 CLF dependent genes, 360 genes were represented in the ATH1 microarray. The 360 genes that were reduced in H3K27me3 in *clf* were classified into 8 clusters depending on their expression in the PcG mutants. Classification was done based on hierarchical clustering that was performed using the Genesis 1.7.6 software. Elbow point analysis using the GMD package in R confirmed that the used dataset can be grouped into 8-10 clusters (Elbow point analysis performed by Vimal Rawat; Zhao *et al.* (2011)). Depending on which mutants expression is mainly altered, I additionally divided the 8 clusters into 4 major categories (A-D):

Category A (clusters 1-4; mainly changed in *clf/swn*): Genes in which H3K27me3 is decreased in *clf* and expression is substantially increased (A1; clusters 1 and 2) or decreased (A2; clusters 3 and 4) in *clf/swn*. At these loci, both, CLF and SWN can partially take over each other's function and only one of the two proteins is necessary to maintain wt-like expression. In cluster 2, genes are upregulated in *clf* but downregulated in *swn*. This implies that CLF and SWN might have antagonistic functions. Similar observations were also made in clusters 1 and 4. Expression of category A genes depends on CLF or SWN.

Category B: (cluster 5; reduced in *swn*): Genes in which H3K27me3 is decreased in *clf* and expression is substantially reduced in *swn*. Analysis in Gbrowse illustrated that at approximately 50% of these genes, H3K27me3 is not affected. This indicated that expression is probably mainly regulated by SWN but without affecting the enrichment of H3K27me3. Therefore, the data might suggest that SWN has a role in transcription independent of H3K27me3. Gbrowse analysis depicted that at 50% of the category B genes H3K27me3 is increased. This suggested that these genes are directly positively regulated by SWN and that the observed increase in H3K27me3 is associated with reduced gene expression. However, It remains unclear if and at which loci SWN is a direct or an indirect regulator. Notably, in *clf/swn* expression is increased compared to *swn*. Thus, the *clf* allele counteracts the effects of the *swn* allele. This implies that CLF and SWN might have antagonistic functions. Expression of category B these genes depends mainly on SWN.

Category C (cluster 6; reduced in *clf*): Genes in which H3K27me3 and expression are decreased in *clf*. Expression of these genes is mainly regulated by CLF. Analogously to cluster B genes, in *clf/swn* expression is increased compared to *swn*. Thus, the *swn* allele counteracts the effects of the *clf* allele. This implies that CLF and SWN might

have antagonistic functions. Expression of category C genes depends mainly on CLF.

Category D (cluster 7-8; changed in *clf*, *swn* and *clf/swn*): Genes in which H3K27me3 is decreased in *clf* and expression is substantially increased (D1; cluster 7) or decreased (D2; cluster 8) in *clf*, *swn* and *clf/swn*. At these loci, the absence of CLF and SWN is non-additive. Neither CLF nor SWN are sufficient to maintain wt-like gene expression. Expression of category D genes depends on CLF and SWN.

Since H3K27me3 is a repressive mark, I expected to identify clusters were expression is increased only in *clf* or in *clf* and *clf/swn*. Strikingly, only *AG* expression is substantially increased in *clf*. However, genes where expression is considerably increased mainly in *clf* might not be represented on the microarray. A cluster were expression is mainly increased in *clf* might appear in case the expression of CLF dependent genes in PcG mutants is analyzed.

FLC and *AG* are in cluster D1 implying that their expression is therefore dependent on both CLF and SWN. Strikingly, the H3K27me3 pattern at these genes does not change in *swn* compared to wt (Fig. 4.12; data not shown). *FT* and *SEP* are not included in the list of genes were H3K27me3 is at least 2 fold reduced and therefore not included in this analysis. However, their expression is altered in *clf* and *clf/swn* but not in *swn* indicating that several genes that are not included in this analysis depend only on CLF. Importantly, the observed changes might be caused by indirect effects through the changes of expression at other loci. Genes in the 9 identified clusters were not connected by specific GO terms and were not expressed in a tissue specific manner (data not shown).

In short, expression of CLF dependent genes can depend on 1: CLF or SWN; 2:SWN; 3:CLF; 4:CLF and SWN. I propose that SWN might function in a H3K27me3 independent pathway and that slight increase in H3K27me3 observed in *swn* might have a negative effect on expression. Furthermore, I depict that gene expression of CLF dependent genes changes in an H3K27me3 dependent or independent manner and cannot be predicted.

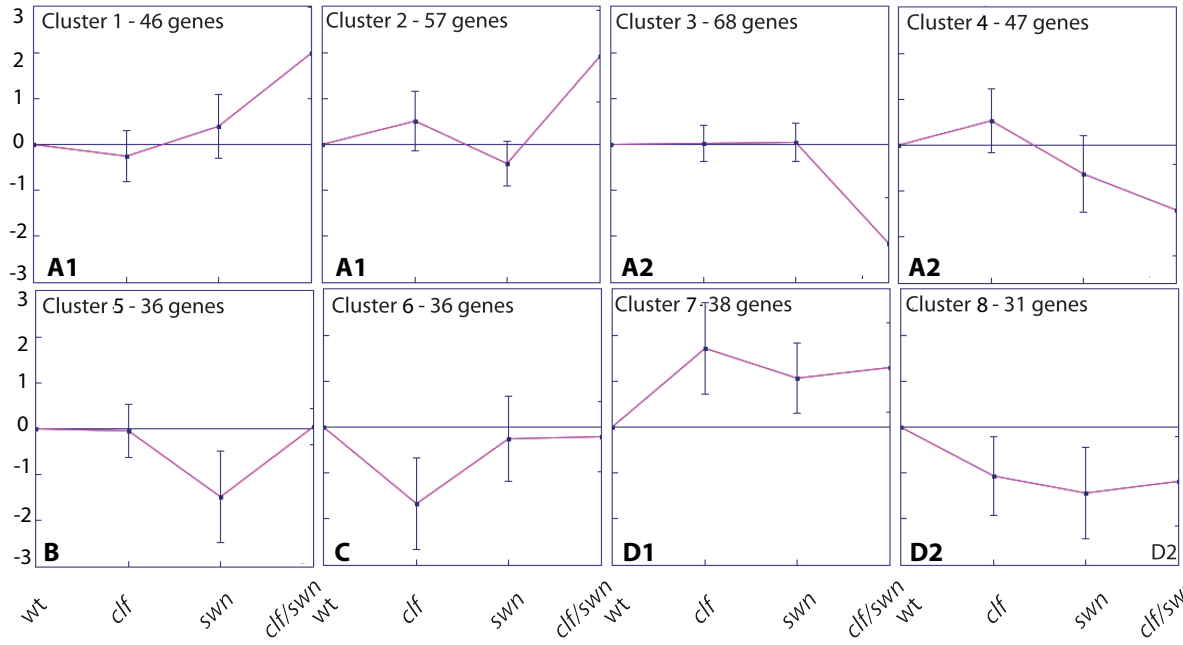


Figure 4.14: **CLF and SWN activate and repress CLF dependent genes in combination or independently from each other**.
K-means clustering (k = 8) of expression of CLF dependent genes in PcG mutant seedlings. Depending on which mutants expression is mainly altered, 4 categories of stable clusters were identified.
Category A: Expression is slightly altered in *clf* and *swn* but considerably increased (A1; cluster 1 and 2) or decreased (A2; cluster 2 and 3) in *clf/swn*.
Category B: Expression is substantially reduced in *swn* (cluster 5).
Category C: Expression is substantially decreased in *clf* (cluster 6).
Category D: Expression is increased (D1; cluster 7) or decreased (D2; cluster 8) in *clf*, *swn* and *clf/swn*.
Analysis was performed with 360 of 643 genes that have reduced H3K27me3 abundance in *clf* compared to wt that are represented on the microarray of PcG mutant seedlings (Farrona *et al.*, 2011). Clustering was performed on log2-transformed fold- change values using the software Genesis 1.7.6. The Pearson correlation coefficient was used as a distance measurement to cluster according to patterns rather than absolute values. Expression values for each gene were calculated as fold change compared with the wild-type (wt). Each graph represents one cluster. Squares represent the average expression, and bars indicate the normalized SD.

4.4.3 CLF dependent genes are distributed in clusters across the chromosomes

In line with previous reports, I found that the H3K27me3 allocation in wt resembles the distribution of genes and is highly enriched in the euchromatic arms (Zhang *et al.*, 2007b). However, the ratio of H3K27me3 abundance between *clf*, *swn* and wt was different depending on the chromosomal region (Fig. 4.11). Thus, I hypothesized that CLF dependent genes might be distributed in a specific pattern along the chromosomes.

To determine how CLF dependent regions are distributed along the chromosomes

we determined the chromosomal distribution of the 643 CLF dependent genes. The genome was divided into 0,2 Mio bp windows and the number of CLF dependent genes per window was calculated. Furthermore the distribution of randomly chosen H3K27me3 target genes per window was calculated to determine the background. 643 H3K27me3 target genes were randomly chosen 1000 times and the average of genes that was included in a given window was plotted (analysis performed by Xue Dong). Strikingly, 18 windows contained significantly more CLF dependent genes than H3K27me3 target genes (Fig. 4.15 A). They were assigned as enriched windows. Enriched windows tend to be located close to the centromeres and the middle of the chromosomes rather than on the ends of arms. The amount of enriched windows per chromosome correlates with the length of the chromosome. The analysis was repeated with windows of different sizes (0,1; 0,5 ; 1 Mio bp) and similar results were obtained (data not shown).

To assess how CLF dependent genes are distributed within the enriched windows, I compared enriched windows with non enriched windows in Gbrowse. In contrary to non enriched windows, enriched windows contain one to three H3K27me3 covered stretches that are between 10 to 30 kb (Fig. 4.15 B and C; Table 4.4). Notably, most H3K27me3 covered regions are smaller than 5 kb (Fig. 4.16 A). H3K27me3 covered regions with sizes above 10 kb were entitled as blocks (Table 4.2). Several H3K27me3 blocks contain large (>2 kb) intergenic regions suggesting that they might have a function. Moreover, we extracted the CLF dependent genes in each enriched window and block and performed GO term analysis to determine if they functionally connected. The genes under five of the eighteen peaks had common GO term enrichment and detailed analysis of genes below blocks indicate that they often belong to one protein family (Table 4.4). These genes might be structurally connected in case they result from tandem duplications. These results suggest that CLF controls individual genes but also gene families.

AG, *FLC*, *FT* and *SEP3* are not located under statistical significant peaks or functionally blocks .

Briefly, I demonstrate that CLF dependent genes are not distributed randomly along the chromosomes and are partially organized in functionally or structurally connected blocks.

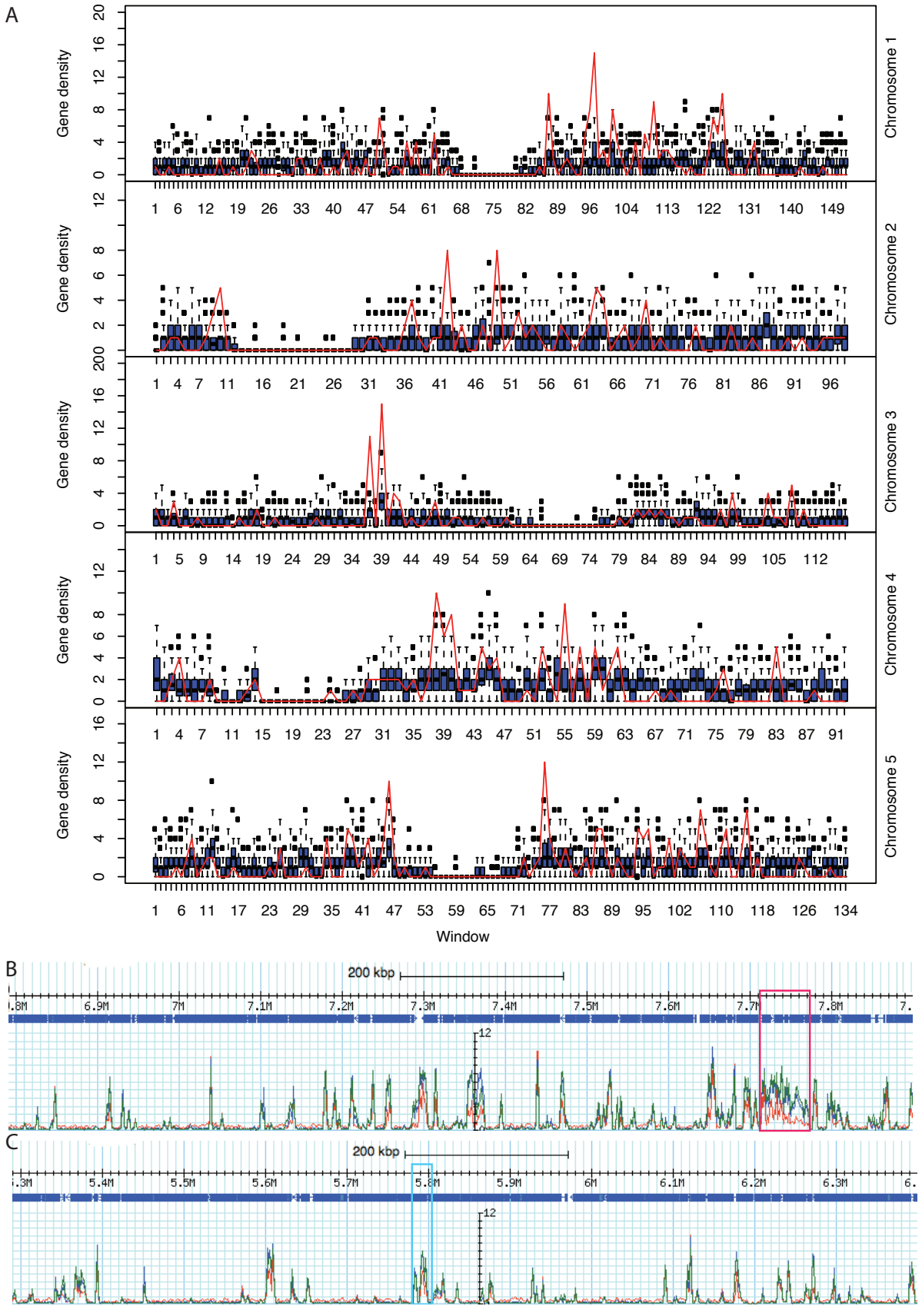


Figure 4.15: **CLF dependent genes are distributed in clusters across the chromosomes**

Figure 4.15:
A: Gene density map of CLF dependent genes within 0,2 Mio bp windows. CLF dependent genes (red line) are significantly enriched in 18 windows. The distribution of 1000 times randomly chosen 644 H3K27me3 target genes is displayed in the box plots (blue) as a background. Regions where the red line was higher than the background were determined as enriched.
For each chromosome, the gene density (y-axis) is plotted against sequentially ordered windows on the (x- axis). The whiskers represent the 2nd and 98th percentile. Analysis was performed in R.
B: Genome browser view of 0,1 Mio bp of window 37 on chromosome 3. CLF dependent genes form ~30kb continuous H3K27me3 covered blocks (red square).
C: Genome browser view of 0,1 Mio bp of window 3-19 on chromosome 3. CLF dependent genes form <5kb H3K27me3 covered regions (blue square).

Table 4.4: **Summary and features of CLF dependent genes in enriched windows.**

Chromosome	Window	Start	Stop	Nr of blocks	Comments
1	50	AT1G28300	AT1G28430	1	Large H3K27me3 covered intergenic region
1	87	AT1G46408	AT1G47410	1	Large H3K27me3 covered intergenic region
1	96	AT1G51240	AT1G51330	1	-
1	97	AT1G51810	AT1G52100	2	Mannose binding proteins
1	110	AT1G59530	AT1G59730	2	-
1	125	AT1G66475	AT1G66870	3	Large H3K27me3 covered intergenic region at AT1G66475; highly covered intergenic region(inclutes SWEET and miRNA157A/B); -
2	10	AT2G05400	AT2G05435	-	-
2	42	AT2G19010	AT2G19070	1	Lipid metabolic process and RALF like genes
2	49	AT2G22620	AT2G22980	1	Carbopeptidases
3	37	AT3G20975	AT3G21040	2	Transposons and unknown proteins
3	39	AT3G21730	AT3G22060	1	-
4	38	AT4G12520	AT4G12950	2	Large H3K27me3 covered intergenic region; 50percent lipid metabolism ; 50 percent IPR016140
4	40	AT4G13480	AT4G13575	1	Large H3K27me3 covered intergenic region
4	55	AT4G19970	AT4G20240	1	Cytochrome and Terpenoid cyclases
5	46	AT5G25955	AT5G26114	1	Plant self incopatibility S1 proteins
5	76	AT5G37800	AT5G38096	2	Large H3K27me3 covered intergenic region; Aquaporines, Alkohol dehydrogenases
5	106	AT5G51845	AT5G52160	1	-
5	115	AT5G56368	AT5G56560	1	Strongly H3K27me3 covered region

4.4.4 CLF dependent genes are enriched for Telo boxes.

CLF dependent genes are partially enriched in specific regions along the chromosome (Fig. 4.15). Because of the non random distribution of these regions, I hypothesized that they might share common DNA sequences. In addition, functionally conserved DNA sequences that are associated with the recruitment of PcG proteins have

been identified in Drosophila, mammals and Arabidopsis (Lodha *et al.*, 2013; Simon & Kingston, 2013). To identify common *cis*-elements within CLF dependent genes, we conduced MEME analysis on the regions that have less H3K27me3 in *clf* compared to wt (performed by Vimal Rawat). However, analysis with MEME can be performed efficiently only with regions that are smaller than 5 kb. Differentially enriched regions were determined with SICER (Section 4.4.1, Zang *et al.* (2009)). SICER combines H3K27me3 positive regions to islands before determination of differentially enriched regions (Zang *et al.*, 2009). 1266 differentially enriched regions are distributed in fragments between 500 - 27.000 bp (Fig. 4.16 A). Since 207 of 1266 fragments are larger than 5 kp, they cannot be reliably analyzed with MEME. Furthermore, we could not find a conserved sequence motif in the 1059 fragments that were smaller than 5 kb (data not shown).

The diffReps 1.54 software uses the sliding window approach with which H3K27me3 positive regions are not fused to large windows (http://code.google.com/p/diffreps/). Using a window size of 600 bp and a step size of 300 bp, 3502 differential regions with a range size between 500bp - 13.000 bp were identified that were at least 2 fold decreased in H3K27me3 (FDR=0,001) in *clf* compared to wt (Fig. 4.16 B) (performed by Xue Dong). 3284 fragments were smaller than 4 kb and could therefore be analyzed with MEME. To increase MEME analysis speed and efficiency the 3502 fragments were divided by size (<0,8kb; 0,8-1kb; 1-1,5kb; 1,5-2kb; 2-3kb; 3-4kb;>5kb) and scanned for conserved *cis*-elements (performed by Vimal Rawat). Results from the 1,5-2kb window and depicted in the following. Similar observations were made in all the analyzed windows.

As in previous scans for conserved *cis*-elements that were performed in our lab on H3K27me3 enriched regions, a strong overrepresentation in AT rich regions was determined (Fig. 4.16 A, personal communication, Ulrike Göbel, MPIPZ). Interestingly, we identified in 25% of the 1,5-2 kb fragments (58 out 239) the highly conserved telo box (AAACCCTA) motif (eValue=3,4*e-9). To test if this motif occurs randomly, each base pair of the AAACCCTA element was randomly exchanged and the occurrence of the mutated telo box in the 3502 fragments was determined. AAACCCTA occurred 1006 (29%) times while mutated versions occurred maximally 639 times (Fig. 4.16 B). Additionally, we determined the occurrence of random 8bp k-mers within the 3500 fragments. The test was repeated 1000times to determine the frequency of occurrence. The probability that a certain 8 bp long fragment occurs 1000 times in the available sequences is marginal (Fig. 4.16 C). The 3502 fragments correspond to 1236 genes and 769 (62%) contain telo boxes within the genic region (Fig. 4.16 D).

GO term analysis of the telo box containing CLF dependent genes, revealed overrepresentation of genes with "transcription factor activity" and InterPro analysis revealed significant enrichment of "Cyclin-F-box" (p=2.972e-22) genes. *AG* contains two telo

boxes in the first intron and two consecutive in the proximate promoter while the *SEP3* proximal promoter and intron comprise one telo box each. *FLC* and *FT* do not contain telo boxes.

In summary, we found that CLF dependent genes are significantly enriched for the telo box motif AAACCCTA and that these genes are associated with "transcription factor activity". Mutated variants of telo boxes were less frequent within CLF dependent genes. Moreover, it is highly unlikely, that the telo box occurrence in 62% of the CLF dependent genes is a random event. Lastly, we ascertained that CLF dependent genes are heavily enriched for A and T rich streches.

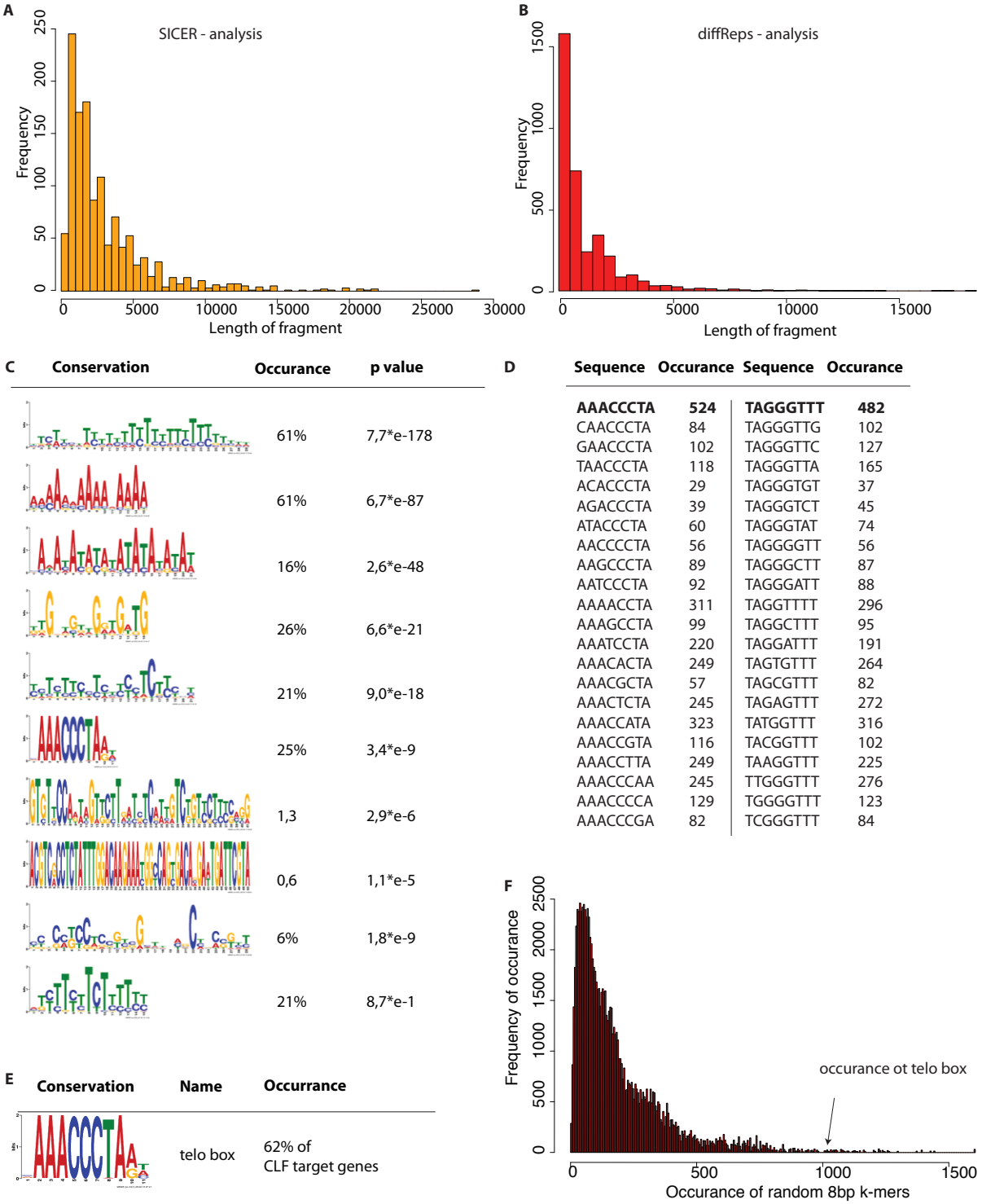


Figure 4.16: **Telo boxes are enriched in CLF dependent regions.**
A: Frequency and size distribution of regions with reduced H3K27me3 in *clf* compared to wt analyzed with SICER (Zang *et al.*, 2009). Fragments distribute in a range between 500 -27.000 bp. 1059 fragments are smaller than 5000 bp.
Differentially enriched fragments (FDR=0,001; 2 fold difference) were determined in a two step approach: 1. combination of H3K27me3 positive regions with a minimum size of 200 bp and gaps smaller than 400 bp into islands. 2. determination of differentially enriched islands .

Figure 4.16:
B: Frequency and size distribution of regions with reduced H3K27me3 in *clf* compared to wt analyzed with diffReps 1.5.4 (http://code.google.com/p/diffreps/). Fragments distribute in a range between 500 -13.000 bp. 3284 fragments are smaller than 4000 bp.
Differentially enriched fragments (FDR=0,001; 2 fold difference) were detected using the sliding window approach (window size=600 bp; step size 300 bp)
C: MEME analysis of regions with reduced H3K27me3 in *clf* compared to wt. Regions with reduced H3K27me3 abundance in *clf* are enriched in poly A streches and a conserved 8bp long cis-element termed Telo box. Conservation is illustrated as a WEBLOGO.
MEME 4.8.1 identification was in performed in the 239 regions with a size between 1500 bp and 2000bp using the zoops (zero or one occurance) model.
D: Summary of Telo-box occurrence. 769 genes out of 1236 that had reduced H3K27me3 abundance in *clf* (diffReps analysis) contained a conserved 8bp long cis-element termed Telo box. Conservation is illustrated as a WEBLOGO.
Identification was performed with MEME and occurrence in sequences confirmed reciprocally with PERL.
E: Occurrence of mutated telo boxes in 3500 fragments. At least one telo box (AAACCCTA and TAGGGTTT) occurs in 1006 fragments. Fragments with one base pair exchange compared to telo boxes occur maximally 639 times.
Experiments were conducted with PERL.
F: Frequency of 8bp kmers in 3500 fragments. A random 8bp kmer occurs with a very low frequency in 1000 of 3500 fragments. Test was repeated 1000 times to calculate frequency of occurrence.
Experiments were conducted with R.

5 Discussion

5.1 Phylogenetic relationships between CLF SWN and MEA

Phylogenetic analysis demonstrated that the diversification between CLF and SWN occurred within the lineage of vascular plants. E(z) family proteins in analyzed green plants form 6 major groups. SWN and CLF homologues exist in all analyzed plant families after mosses and *Selaginella moellendorffii* which both contain an E(z) protein that is evolutionary closer to CLF. In addition to previous reports which detected MEA only in Brassicaceae, I found MEA homologues in Fabaceae (Fig. 4.1 Spillane *et al.* (2007)). Moreover, *SWN* homologues were found to be conserved in the genomes of 80 Arabidopsis accessions (Fig. 4.2).

My results suggest that both CLF and SWN have a function in green plants and that the diversification between the two proteins is important in Arabidopsis. Nonetheless, it is not known to what extent the functions of CLF and SWN have evolved. If the functional difference is mainly at the mechanistic layer (e.g H3K27me3 catalysis) or due to tissue subfunctionalization remains unclear. Generally, at least three PRC2 complexes have been suggested to coexist in Arabidopsis to control different developmental processes (Butenko & Ohad, 2011). In tomato, *Solanum lycopersicum CLF* (*SlCLF*) cannot compensate the lack of *SlSWN* and the two genes are expressed differentially between different tissues (How Kit *et al.*, 2010). *MEA* and *Mez1* in Arabidopsis and maize, respectively have undergone subfunctionalization (Gleason & Kramer, 2012). Arabidopsis *MEA* is expressed mainly in the gametophyte and the developing embryo while *CLF* and *SWN* are mainly expressed in the sporophyte. Furthermore, *MEA* cannot complement the *clf* mutation when it is ectopically overexpressed (Köhler *et al.*, 2003b; Jeong *et al.*, 2011; Chanvivattana *et al.*, 2004). Thus, it is likely that the diversification of CLF and SWN might have evolved to coordinate specific sporophytic transitions and organogenesis. Conditional knockout and complementation studies of *CLF* and *SWN* in basal vascular plant species could unravel the specific role of these two genes in sporophytic transitions. I determined the role of CLF and SWN by creating H3K27me3 patterns of young Arabidopsis *clf* and *swn* null mutant seedlings. Analogous studies in non-Brassicaceae species could elucidate functions of CLF and SWN in H3K27me3 deposition that might have been lost in the evolution of Brassicaceae. In addition, roles of CLF and SWN in specific tissues could be assessed by creating tissue specific H3K27me3 profiles of Arabidopsis plants that carry *clf* or *swn* mutations.

CLF, SWN and MEA share 7 conserved domains and differ mainly in the middle part (Fig. 4.3). The C terminal SET and CXC domains that have been demonstrated to be important for the PRC2 HMTase activity are most conserved between the analyzed proteins (Fig. 4.3, Baumbusch *et al.* (2001); Cao *et al.* (2002); Czermin *et al.* (2002)). Alignment of the SET domains of the two proteins could reconstruct the phylogenetic relationships that were determined with full length proteins. This implies conserved differences in the catalytic function of CLF and SWN. Recent reports in mammals and Drosophila suggest that the catalytic activity of the Ezh2 SET domain is negatively regulated through interaction of the VEFS domain of Suz12 proteins upon recognition of H3K4me3, H3K36me2 or H3K36me3 (Ciferri *et al.*, 2012; Schmitges *et al.*, 2011). In contrast, Eed recognizes H3K27me3 and probably enhances through its WD40 domain the HMTase activity of Ezh2 SET domain (Ciferri *et al.*, 2012; Margueron *et al.*, 2009). The conserved sequence differences in the CLF and SWN SET domains could modulate the interaction between Eed and Suz12. SWN and CLF could therefore have different catalytic activities depending on the epigenetic landscape. The detected differences in H3K27me3 landscape between *clf* and *swn* could depend on the observed differences in the SET domain. Crystal structure determination of CLF- and SWN-PRC2 complexes could unravel if the two complexes have similar structures and therefore functions like the mammalian PRC2 complexes. Putative CLF and SWN specific inter- and intramolecular interaction sites could be determined by studies were crosslinking is combined with electron microscopy (Ciferri *et al.*, 2012). The importance of these interaction sites in catalyzing the trimethylation of H3K27 could be confirmed by performing histone methyl transferase (HMTase) assays with CLF- and SWN-PRC2 complexes that carry mutations at these sites. Isothermal titration calorimetry (ITC) assays with versions of CLF and SWN that carry mutations at the putative interaction sites could uncover the importance of these sites for the complex stability.

The mammalian E(z) homologues Ezh1 and Ezh2 form a part of mechanistically distinct PRC2 complexes Ezh1-PRC2 is responsible for gene repression through chromatin compaction and Ezh2-PRC2 represses gene expression through H3K27me3 deposition (Margueron *et al.*, 2008). Besides, the Su(z) homologues EMF2 and VRN2 function differently depending on the nucleosomal context. EMF2-PRC2 mono-, di- and trimethylates unmodified nucleosomes but VRN2-PRC2 di- and trimethylates unmodified but also H3K4me3 marked nucleosomes (Schmitges *et al.*, 2011). This illustrates that changes in the amino acid sequences of PRC2 core components can result in substantial functional differences. Differences in CLF-PRC2 and SWN-PRC2 substrate specificity and HMTase activity could be assessed by performing HMTase assays with recombinant PRC2 complexes and nucleosomal templates that carry different histone modifications. In addition, performing these experiments with versions of CLF and SWN where the conserved domains are exchanged between the two proteins could

unravel the functional differences between CLF and SWN.

The strongest difference between CLF, SWN and MEA amino acid sequences were detected in the middle part. The middle part was conserved within CLF and MEA homologues but not within SWN-likes. However, monocot SWN proteins are conserved in the middle part. Moreover, the Arabidopsis SWN is the only Brassicaceae representative in this alignment. It is therefore possible that the middle part is conserved within the Brassicaceae. These results illustrate that the middle part of E(z) proteins underwent substantial changes during evolution but might be conserved within plant families. Therefore the middle part of E(z) proteins might define the functional differences between plant E(z) proteins. In a possible scenario, this middle part could be important for the interaction with a functionally but not structurally conserved factor. Specific interacting proteins could be identified in a Y2H assay with the middle parts of CLF, SWN and MEA . Alternatively, plants that carry E(z) proteins with deletions in the middle part could be created. In immunoprecipitation experiments followed mass spectrometric analysis, components that bind to the middle part should be absent in mass spectrograms of mutated proteins.

When the sequences of the middle part of CLF (CRI-domain) and MEA (MRI-domain) were analyzed with SMART, no homology to other domains could be detected (Letunic *et al.*, 2004). This suggests that the function of this part of the protein is highly specific for plant E(z) proteins. However, it was depicted that some amino acids within the middle part of mammalian Ezh2 are important for intramolecular binding, indicating that amino acids in the middle part of E(z) proteins are diverse but important (Ciferri *et al.*, 2012). It is likely that the middle part of plant E(z) proteins is responsible for an inter- or intramolecular interaction that defines at least partially the specific function of CLF, SWN and MEA. The importance of this protein part could be assessed by testing if a CLF protein with a SWN middle part can complement the *clf* mutant phenotype. In a reciprocal approach, introgression of the CLF middle part into the analogous position in SWN should restore the phenotype of wt plants in a *clf* mutant background. Whether the middle domain is important for H3K27me3 deposition could be tested by creating H3K27me3 profiles of plants that contain versions of CLF which have the middle part of SWN instead of the CLF middle part.

Importantly, I also observed conserved differences in the SET domains of CLF and SWN. To assess the effects of the middle part and the SET domain, these sequence stretches of CLF and SWN, respectively could be interchanged and transformed into *clf/swn* mutants. If the middle part and the SET domain of CLF are sufficient to confer full CLF function, a SWN version with a CLF middle part and a CLF SET domain should complement the *clf* mutation. Analogously, CLF with a SWN middle part and SET domain should not restore the wt phenotype in *clf* mutant plants. To assess the evolutionary importance of the middle part of SWN, middle parts of SWN homologues

from basal vascular plants could be exchanged with the native Arabidopsis SWN middle part. In case the function of the middle part is conserved, these fusion proteins should restore the phenotype of *clf* mutant plants in the *clf/swn* mutant background.

Taken together, I propose that the differences in the SET domains of CLF and SWN cause differences in the enzymatic properties of these two proteins. These differences might influence the catalytic activity or substrate specificity of CLF and SWN. The differences in the middle parts of CLF and SWN might additionally modulate the functions of the proteins and be important for the interaction with specific proteins.

5.2 Attempts to ChIP on PRC2 proteins

5.2.1 CLF is a low binding to chromatin is at the detection limit of ChIP-PCR

ChIP-PCR experiments and statistic analysis demonstrated that CLF binds to *AG* and *FLC*. However, the enrichment of target DNA was at least 10-fold lower than observed for H3K27me3 (Fig.4.5). Moreover, I could not detect GFP-CLF binding to chromatin by ChIP-seq. Previous studies had detected genome wide binding of an overexpressed version of LHP1-HA to chromatin (Zhang *et al.*, 2007a; Turck *et al.*, 2007). However, western blot demonstrated that CLF-HA abundance is low compared to LHP1-HA (Fig. 4.4).

GFP-CLF signal in ChIP is constitutively higher at loci where no binding is expected compared to ChIP-signal obtained with wt plants. This suggests that CLF could randomly and loosely bind to chromatin in a cell autonomous manner. It is proposed that upon vernalization, H3K27me3 accumulates at *FLC* in a cell dependent manner (Angel *et al.*, 2011). This implies that CLF might bind strongly to specific loci only in some cells. Moreover, VRN2 binding to *FLC* is increased upon vernalization and VRN2 accumulates in shoot tips of vernalized plants (De Lucia *et al.*, 2008; Wood *et al.*, 2006). Hence, CLF, like VRN2, might actively bind chromatin only upon a specific signal and this binding might be cell specific. I speculate that CLF binds generally loosely to chromatin but accumulates at nucleation regions only under certain conditions in a specific cell population to trimethylate H3K27. CLF abundance is heavily increased upon vernalization (Wood *et al.*, 2006). Furthermore, *CLF* mRNA abundance is different between young and mature leaves (Finnegan & Dennis, 2007). The authors suggest that these differences in *CLF* result from differences in cell proliferation rate. To test if CLF binding to chromatin is increased at nucleation sites in proliferating cells after perception of an inductive signal could be assessed by performing ChIP-PCR experiments in cold treated plants in tissues with high and low division rates at putative nucleation sites.

CLF and SWN share the same binding partners (Chanvivattana *et al.*, 2004). More-

over, ectopic CLF expression in the endosperm developing seeds induces embryo abortion like in the *mea* mutant, probably because CLF antagonizes MEA in the binding of partner proteins (Jeong *et al.*, 2011). Therefore, it might be difficult to perform ChIP against CLF, because it competes with SWN for binding partners and possibly for DNA binding positions. ChIP signal of CLF in a *swn* mutant could be therefore higher than in wt. In this case, comparing ChIP signals of CLF abundance in wt and in the *swn* mutant plants could confirm such a hypothesis.

In Western blots and ChIPs that were conducted, CLF abundance levels between different experiments were highly variable. Furthermore, CLF could not be detected in nuclear extracts in case they were frozen and thawed. This implied that CLF is a highly unstable protein. In addition, although CLF is localized in the nucleus I did not observe a significant increase in CLF after nuclear extraction compared to total extracts (Fig. 4.9 Schubert *et al.* (2006)). The protein might be degraded during nuclear extraction but less degraded in total extracts. Application of a variety of proteinase inhibitor cocktails and testing different nuclear extraction protocols could decrease CLF degradation.

I detected ~800 GFP-CLF enriched genes per biological replicate in the conducted ChIP-seq experiment but only 166 were overlapping. The 166 detected peaks could not be confirmed in the IGV browser and known CLF target genes were not within the 166 identified target genes. In addition, visual analysis of the identified target genes depicted that there was no enrichment between target genes and non-target genes. These results suggested that the performed ChIP-seq experiment did not work technically. However, I could depict that CLF binds to chromatin and I was able to create ChIP-DNA libraries to get millions of high quality unique reads upon sequencing from ChIPs against H3K27me3. I speculate that the enrichment of CLF at target genes might be not clear in ChIP-seq because of ChIP-DNA amounts are low and relative abundance of ChIP-DNA fragments are changed during library preparation. Reliable genome wide CLF distribution maps could be generated by creating ChIP-DNA libraries using ChIP samples from plants or tissues where CLF enrichment at chromatin is increased at known target genes (as discussed above).

In short, it is unclear why I detect remote amounts of CLF bound to chromatin. I expect that the signal in ChIPs against CLF will be increased if experiments are conducted in highly proliferating tissues in vernalized plants. Using these samples, enrichment of CLF at target genes might also be detectable in ChIP-seq experiments.

5.2.2 CLF abundance is regulated at the post-translational level

The fact that plants that express *CLF* under the control of the *MEA* promoter have *mea*-like seed abortion phenotype indicates that ectopic CLF expression is toxic (Jeong *et al.*, 2011). Thus, *35S:CLF:HA* plants with high CLF abundance might not be viable.

However, the observation that CLF, SWN and FIE proteins but not their respective mRNAs are increased after vernalization indicates that CLF abundance is regulated at the level of translation or protein turn-over (Wood *et al.*, 2006). This hypothesis is supported by the observation that MEA transcript levels decrease gradually after pollination while MEA protein disappears quickly (Dumbliauskas *et al.*, 2011). Moreover, CLF-HA protein could be detected but the abundance was low compared to LHP1 and did not correlate with the mRNA amount to same proportions as LHP1-HA. Hence, I speculate that I selected for *35S:CLF:HA* plants with low CLF abundance. An inductive system in which CLF fused to the GR domain of the glucocorticoid receptor could be created to overcome embryonic lethality. In the absence of a steroid ligand, the GR domain retains a nuclear factor in the cytoplasm but nuclear localization is restored in the presence of the synthetic glucocorticoid Dexamethasone (Dex, Aoyama & Chua (1997)). Higher CLF abundance would be expected in transgenic *35S:CLF:GR* plants treated with Dex compared to mock treated plants.

It was recently shown that UCL1 forms an Skp, Cullin, F-box (SCF) complex, binds CLF, and reduces CLF protein levels, suggesting that CLF is regulated by the 26S-Proteasome pathway (Jeong *et al.*, 2011). I tested if CLF is regulated by the 26S-Proteasome, assuming that the CLF amount should be increased when the 26S-Proteasome is inhibited. Unexpectedly, I detected only a mild increase in CLF abundance compared to the positive control, suggesting that the observed increase in CLF abundance is only indirectly mediated by the 26S-Proteasome (Fig. 4.9). It is proposed that the effect of the CLF associated E3 Ubiquitin ligase CUL4 on cell cycle control is more dominant and counteracts CUL4 involvement in FIS PRC2 regulation (Dumbliauskas *et al.*, 2011). Thus, inhibition of the 26S-Proteasome might diminish direct CLF degradation by the 26S-Proteasome but simultaneously increase the abundance of compounds that degrade CLF. Transient inhibition of putative CLF regulators e.g. by creating plants in which the putative CLF regulators are transported into the nucleus only upon Dex treatment combined with 26S-Proteasome treatment of these plants could increase CLF abundance. Alternatively, CLF abundance could be increased by mutating the CLF polyubiquitilation and UCL1 binding sites.

The FIS-PRC2 complex protein MSI1 and its close homolog MSI4 interact with the CUL4-DDB1 E3-Ubiquitin ligase proteins and regulate flowering time together with PRC2 complexes (Dumbliauskas *et al.*, 2011; Pazhouhandeh *et al.*, 2011). Additionally, CUL4-DDB1 is required for the FIS-PRC2 dependent paternal imprinting. It is speculated that CUL4-DDB1 directly ubiquitilates FIS-PRC2 for degradation (Dumbliauskas *et al.*, 2011). However, activity reduction of CUL4 leads to a decrease in H3K27me3, and the mammalian CUL4 and DDB1 interact with Eza2 and Suz12 and are necessary for full H3K27m3 and H2Ak119 activity (Hu *et al.*, 2012). These results suggest that CUL4 - PRC2 interaction is probably not primarily mediating CLF ubiquitination and subse-

quent 26-Proteasome dependent degradation. Therefore, CLF abundance is probably regulated additionally by a 26S-Proteasome independent pathway.

An alternative mechanism to control CLF abundance could be mediated by the autophagy pathway. However, tests with autophagy inhibitors performed in our lab do not support this hypothesis (data not shown).

In summary, CLF and other PRC2 complex proteins in animals and plants are probably regulated at the post-transcriptional level. Post-transcriptional regulation is likely mediated only indirectly through the 26S-Proteasome pathway. Approaches to increase CLF protein abundance and therefore presumably ChIP-seq efficiency by protein degradation were not successful but the use of inducible expression remains to be tested.

5.2.3 ChIP-PCRs should be normalized against an internal control

I suggest that in ChIP-seq experiments the bias at single loci is low compared to ChIP-PCR. However, this is only the case if the amount of unique reads is similar and individual loci are normalized against the unique amount of reads (personal communication, Julia Engelhorn, CEA, Grenoble). Nevertheless, differences in signal abundance should be additionally assessed by statistical approaches (e.g. SICER, Zang *et al.* (2009)) and ChIP-PCRs.

Normalization approaches in ChIP-PCR experiments vary between different and sometimes even within the same reports:

1. normalization against the input of each IP (De Lucia *et al.*, 2008; Lopez-Vernaza *et al.*, 2012).

2. normalization against H3 (Lodato *et al.*, 2013).

3. normalization against an internal control (Heo & Sung, 2011a; De Lucia *et al.*, 2008).

Generally, variation between ChIP-PCR experiments is probably due to the very low amount of immunoprecipitated DNA and a difference in ChIP efficiency between chromatin preparations. The choice of the normalization method and the robustness between biological replicates is therefore crucial. Naturally, each normalization approach has drawbacks.

If samples are normalized against the input, bias can be introduced because the input is processed independently from the immunoprecipitated sample. Furthermore, ChIP efficiency varies between chromatin extracts.

It is assumed that within an extract of seedlings, the position of H3 is on average not changed. Therefore, H3 can be used to quantify the amount of histone marks or proteins bound to chromatin. However, variation can be introduced if there is variation in antibody efficiency. To determine H3 IP efficiency, IPs with serially diluted H3 antibodies could be performed. In addition, variation in H3K27me3 abundance might

change nucleosome allocation (Simon & Kingston, 2013). Therefore, histone position might be different between *clf*, *swn* and wt.

In ChIP-PCRs that are normalized relative to the abundance of a locus which remains stable between extracts, bias compared to normalization relative to the input or H3 is minor. However, it may be impossible to determine a locus that does not change between extracts prepared from different genotypes or conditions.

In my experiments, I concluded that normalization could be carried out relative to the enrichment of a single gene, At4g24420, since H3K27me3 abundance in ChIP-seq was highly similar (type 3 island) at this locus between *clf*, *swn* and wt. I assume that the bias at a single locus in ChIP-seq experiments with a similar amount of unique reads is lower than in ChIP-PCR. Normalization is performed relative to the total amount of unique reads using the ChIP-seq data normalization software SICER (Zang *et al.*, 2009).

However, H3K27me3 abundance at At4g24420 varied in *clf*, *swn* and wt when ChIP-PCRs were normalized relative to the input or H3 (Fig. 4.12). The fact that no genotype dependent variation of H3K27me3 could be observed between two biological replicates at At4g24420 suggested that the observed changes result from technical variation between the ChIP-PCRs.

In summary, I suggest that relative enrichment in ChIP-PCRs in IPs against histone marks should be normalized relative to the enrichment of a single gene that does not change between tested conditions if this is possible. The consistency of the internal control and in general ChIP results should be determined by a combined reciprocal approach: ChIP-seq with statistical analysis and repetitions of ChIP-PCRs using different normalization methods.

5.3 H3K27me3 profiles of *clf* and *swn* mutants

5.3.1 Three types of H3K27me3 patterns

CLF and SWN have overlapping functions and H3K27me3 is only partially reduced in *clf* at *AG*, *FT*, *FLC* and *SEP3* (Chanvivattana *et al.*, 2004; Farrona *et al.*, 2011; Lopez-Vernaza *et al.*, 2012). In the *clf/swn* mutant though, H3K27me3 is virtually absent at *FT* (Farrona *et al.*, 2011). I determined that most PcG target genes remain marked by H3K27me3 in *clf* or *swn* mutants, which underscores the hypotheses that CLF and SWN can complement for each others function in H3K27me3 deposition (Fig. 4.10, Chanvivattana *et al.* (2004)). To confirm this assumption full H3K27me3 profiles of *clf/swn* mutants should be examined. MEA is expected to target only a few genes since it is mainly involved in the control of seed development Köhler *et al.* (2003b). Apart from MEA target genes, most PcG target loci should be depleted of H3K27me3 marks.

Although the PcG target genes remain marked by H3K27me in *clf* and *swn*, I found that the H3K27me3 distribution patterns change (Fig. 4.11 and 4.12). I determined 3 types of islands:

Type 1. Regions that where substantially reduced in *clf* but not in *swn*

Type 2. Regions that where mainly increased in *swn*

Type 3. Regions that where unchanged

In type 1 regions at least one peak remained fully H3K27me3 positive in most islands, suggesting that SWN partially can take over the CLF function at specific sequences or specific positions within islands (Fig. 4.11). It is unclear what determines the localization of this peak. Additionally, it is unclear whether this peak exists due to global H3K27me3 reduction, or if I observe a mixed signal from loci where H3K27me3 is completely abolished and tissue where H3K27me3 is maintained. In a possible scenario, CLF has specific target genes that are determined by co-factors. SWN might have a more general role, for example by keeping a low amount of H3K27me3 at loci to ease redeposition of the mark after e.g. nucleosome turnover. To assess the role of these peaks, the underlying DNA sequences could be extracted and analyzed with MEME. In case these peaks have common motifs, it could be tested if they comprise plant PREs. In striking difference to animal PREs, that consist of many kilo bases long sequence stretches far from the H3K27me3 site, plant PREs would resemble ~200bp long sequences within the H3K27me3 marked region (Simon & Kingston, 2013). However, PRC2 could be recruited to these peaks although they are not connected by a specific motif.

H3K27Ac, H3K36me3 and H3K4me3 are marks that antagonize H3K27me3 and are associated with highly expressed genes (Pien & Grossniklaus, 2007; Schmitges *et al.*, 2011; Simon & Kingston, 2013). It is tempting to speculate that the abundance of these marks is partially increased upon H3K27me3 reduction in *clf*. In a recent report mathematical modeling suggested that *FLC* expression is regulated in a bistable manner through histone marks (Angel *et al.*, 2011). In this model, *FLC* is either covered by activating or repressing histone marks. In case a certain threshold of activating or repressing histone mark is reached, expression switches to "ON" or "OFF" respectively. However, it remains unclear at what levels H3K27me3 has to be reduced and activating marks have to be increased to affect gene expression. Answering this question will be particularly difficult, since the control of gene expression is multifactorial and often very gene specific. Mathematical modelling approaches coupled with gene expression studies and chromatin IPs could shed light on the precise transcriptional regulation of individual genes. Increase of H3K27me3 antagonizing marks in *swn* would indicate a possible involvement of SWN in the regulation of these marks.

H3K27me3 plays a major role in the PRC1 complex recruitment (Gao *et al.*, 2012; Simon & Kingston, 2013; Tavares *et al.*, 2012). In Arabidopsis, LHP1 is associated

with H3K27me3 and at the *FT* locus, LHP1 is reduced in *clf* and *clf/swn* analogously to H3K27me3 (Farrona *et al.*, 2011; Lodha *et al.*, 2013; Turck *et al.*, 2007). However, H3K27me3 probably controls gene expression also by antagonizing activating epigenetic marks, reducing accessibility of trans activating factors or RNA polymerase II, changing nucleosome dynamics (Simon & Kingston, 2013). It remains unclear how the LHP1 abundance changes upon H3K27me3 reduction in a genome wide level and how changes in expression in Arabidopsis *clf* or *swn* are associated with loss of LHP1.

In mammals, the chromatin remodeling proteins Brahma (BRM) and Brahma-related gene 1 (BRG1) are highly similar. Because of the mild phenotype of *brm* mutants, it has been generally thought that BRM plays a similar but mostly auxiliary role to BRG1 in regulation of tissue specific gene expression (de la Serna *et al.*, 2006). However, a recent study showed that these proteins have an antagonistic function during differentiation (Flowers *et al.*, 2009). Analogously, CLF and SWN could have overlapping antagonistic function (Fig. 4.3.2, A1, B and C and Fig. 4.14 A). This could be induced by higher accessibility of CLF to partner proteins. Embryonic lethality of *pCLF:MEA* also suggests that CLF and MEA compete for binding partners (Jeong *et al.*, 2011).

To further assess the antagonistic role between CLF and SWN, comparative ChIPs of CLF or SWN in *swn* or *clf* mutant background could be performed. CLF or SWN binding to type 2 loci should be higher in the absence of SWN or CLF, respectively.

Unexpectedly, no regions where H3K27me3 was reduced could be detected in *swn*. However, type 2 regions displayed increased H3K27me3 abundance in the absence of SWN. This observation could indicate that CLF has a higher HMTase activity than SWN. Since the experiments were performed in complete young seedlings, H3K27me3 might also be increased only in a certain cell type. Such a hypothesis could be supported by the fact that H3K27me3 enrichment changes dynamically during development in a tissue dependent manner (Lafos *et al.*, 2011; Weinhofer *et al.*, 2010). Therefore, SWN related effects might be identified by performing ChIP-PCRs against H3K27me3 in different tissues and developmental stages with *swn* mutant plants. Indeed, CLF dependent genes are differentially expressed in 11 different tissues (data not shown; ATAX microarray).

Taken together, CLF and SWN regulate H3K27me3 abundance differentially depending on the locus. They act together or independently and they might antagonize each other at some loci. It is possible, that CLF and SWN have partially distinct functions in different tissues.

5.4 Analysis of CLF dependent genes

5.4.1 CLF and SWN regulate gene expression together, independently, alone or antagonistically

I showed that in Arabidopsis expression of H3K27me3 target genes is not predictable in PcG mutants and that CLF dependent genes form 4 major groups of expression clusters (Fig. 4.14): Genes that depend on 1: CLF or SWN (cluster A); 2: SWN (cluster B); 3: CLF (cluster C); 4: CLF and SWN (cluster D).

Expression of cluster A genes probably depend on the coordinated action of CLF and SWN. However, H3K27me3 does not change in *swn* but in *clf/swn*. In addition, expression of cluster B is substantially reduced in *swn*. Nonetheless, only ~50% of the cluster B CLF dependent genes are also SWN target genes (increased H3K27me3 in *swn*). It has been reported that the mammalian E(z) homologue can function in a PRC2 /H3K27me3 independent manner to control transcription (Xu *et al.*, 2012). Thus, it is tempting to speculate that SWN might have an PcG/H3K27me3 independent role in controlling gene expression. However, it remains unclear to what extent changes in H3K27me3 levels affect expression the respective gene. In mammals, hypertrimethylation of H3K27 can lead to cancer (McCabe *et al.*, 2012). In Arabidopsis, the gain-of-function *clf-59*-allele induces higher H3K27me3 at *FLC* (Doyle & Amasino, 2009). Assuming that most genes can be trimethylated at H3K27 by CLF, global H3K27me3 and expression profiles of the *clf-59* mutant could be performed to assess the effects of hypertrimethylation at H3K27 in Arabidopsis.

The full effects of SWN (and CLF) on H3K27me3 abundance might be limited to a specific context and can therefore not be detected in the applied experimental setup (discussed section 5.2.1). H3K27me3 independent downregulation of genes in *swn* might also be explained by systemic/indirect effects caused by deregulation of other H3K27me3 target genes (discussed by Lafos *et al.* (2011)).

In clusters 1, 2 and 8, expression changes to the opposite direction in *clf* and *swn*. Expression is reduced in cluster B and C in *swn* and *clf* ,respectively, but expression is restored in *clf/swn*. Moreover, H3K27me3 is increased at some loci upon SWN depletion (SWN affected genes). Lastly, a recent report illustrated that expression of a group of genes can be reduced in *clf* but increased in *clf/swn* plants (Farrona *et al.*, 2011). These observations suggest that CLF and SWN could have opposing roles at these loci (discussed in section 5.3.1).

H3K27me3 is associated with transcriptional repression. Thus, I expected to detect genes that have increased H3K27me3 in *clf*. Unexpectedly, I did not detect a cluster where genes are up-regulated only in *clf* or in *clf* and *clf/swn*. This could imply, that there are no genes at which PcG mediated repression relies uniquely on CLF. This could partially explain the mild phenotype of *clf*. However, the expression data was

gathered under different experimental conditions than the ChIP-Seq data. The two experiments should be repeated under same conditions to be comparable.

5.4.2 CLF dependent genes are distributed in clusters across the chromosomes

I observed that CLF dependent genes are distributed non randomly along the chromosomes and form large continuous H3K27me3 covered islands which I named blocks (Fig. 4.15 A and Fig. 4.16 A). These blocks mark sometimes genes that belong to one gene family and have probably undergone gene duplication (Table 4.4). H3K27me3 distribution of all PcG dependent genes resembles the distribution of genes (Zhang *et al.*, 2007b).

It is unclear which factors determine whether a gene or a region is targeted by CLF, SWN or both. One possible explanation could be that these regions are differentially accessible by CLF or SWN. The tomato CLF and SWN homologues are located at distinct locations within the nucleus (How Kit *et al.*, 2010). Confocal microscopy studies could give further insights on the localization of CLF and SWN in Arabidopsis nuclei and help explain the non random distribution of CLF dependent genes on the chromosomes.

CLF dependent genes form often H3K27me3 blocks and these cover in some cases gene families. It was recently suggested that H3K27me3 constrains expression divergence after gene duplication (Berke *et al.*, 2012). I speculate, that gene families in H3K27me3 blocks have undergone gene duplication and H3K27me3 confers equal expression of these genes. Alignment of these DNA sequences with the respective sequences from other plant species could unravel if these regions arose from gene duplication. Testing the expression pattern of these genes at different time points and in different tissues could depict if they are simultaneously expressed. Simultaneous expression would suggest that these genes are regulated by a similar mechanism.

H3K27me3 blocks of CLF dependent genes contain partially large intergenic regions. The mammalian lncRNAs Xist and HOTAIR are expressed in intergenic regions and are responsible for the recruitment of the PRC2 complex (Zhao *et al.*, 2008; Tsai *et al.*, 2010). In Arabidopsis *FLC*, the lincRNAs *COLDAIR* and *COOLAIR* are expressed from intronic and exonic regions to regulate *FLC* expression, partially via PRC2 (Swiezewski *et al.*, 2009; Heo & Sung, 2011b). Large intronic regions within H3K27me3 covered blocks could comprise regulatory units that control expression of the adjacent genes. A possibility is that they encode for lncRNAs that recruit PRC2. The role of the intergenic regions in the H3K27me3 blocks, could be analyzed by determining if these regions are conserved in other spiecies by performing phylogenetic shadowing. In a subsequent step, the expression of genes that are flanking the conserved regions could be measured in plants that carry transgenes with mutated and wt versions of these

highly conserved parts and the flanking genes. Finally, it was predicted that Arabidopsis contains ~2700 lncRNAs (Liu *et al.*, 2012). It could be tested if predicted lncRNAs are located at H3K27me3 covered blocks.

In summary, I propose that CLF marks gene families that arose from gene duplication with H3K27me3 to confer equal expression patters of these genes. I also hypothesize that PcGs control apart from genes regulatory elements like lncRNAs. It remains unclear, which factors are responsible for the non-random localization of CLF target genes.

5.4.3 Telo boxes are associated with CLF dependent genes

I found that 62 % (~1000) of CLF dependent genes are substantially enriched for the telo box element (AAACCCTA) and that these genes are associated with "transcription factor activity". Mutated versions of telo boxes were considerably less abundant in these genes and the probability that an 8 bp long sequence occurs in 1000 genes is ~0,04%. In addition, CLF dependent genes were highly enriched for A and T rich stretches (Fig. 4.16).

It is not clear yet, if telo boxes are enriched only in CLF dependent genes or in all H3K27me3 target genes. Previous search for conserved DNA elements in Arabidopsis H3K27me3 marked sequences has not revealed conserved sequence motifs but A and T rich stretches (personal communication, Ulrike Göbel, MPIPZ, Cologne). Thus, Iexpect that telo boxes are not significantly enriched in H3K27me3 target genes. A and T are more abundant in Arabidopsis than G and C. This explained that A and T enrichment in H3K27me3 target genes is not significantly enriched compared to the Arabidopsis genome.

Telo boxes resemble the Arabidopsis telomeric repeat sequence ((AAACCCT)n) and are located mostly in the 5'prime UTR and at the telomeres (Gaspin *et al.*, 2010; RICHARDS & AUSUBEL, 1988). They have been reported as enriched in promoters of genes that are overexpressed in cycling cells and promoters of genes that are important for DNA replication (Michael *et al.*, 2008; Trémousaygue *et al.*, 2003; Wang *et al.*, 2011). The telo box is also part of a transcription module that is important for the expression of midnight expressed genes (Michael *et al.*, 2008). Furthermore, telo boxes are important for the coordinative expression of genes that are expressed in the mitochondria and the cytosolic ribosomal proteins (Wang *et al.*, 2011). Although telo boxes cannot activate transcription by RNAPII, a GUS assay with mutated and wt versions of telo box elements demonstrated that they enhance transcription (Trémousaygue *et al.*, 2003).

The link between CLF, CLF dependent genes and telo boxes is not clear yet since telo boxes occur in promoters of expressed genes while CLF is rather associated with gene repression. Whether telo boxes counteract or act together with CLF to repress

genes remains to be answered. The distribution of telo box containing genes in expression clusters of CLF dependent genes in PcG mutants could answer this question. Increase of expression of telo box containing genes in PcG mutants would suggest that telo boxes possibly counteract CLF mediated repression. Otherwise, if telo boxes are mostly abundant in genes in which expression in reduced in PcG mutants it would be probable that telo boxes act together with CLF.

A recent publication identified a plant PRE element in the *BP* promoter (Lodha *et al.*, 2013). Upon mutation of the putative PRE, less H3K27me3 could be recruited to the transgene. Interestingly, the BP promoter contains a telo box repeat. Because telo boxes are enriched in CLF dependent genes Ihypothesize that although telo boxes are associated with increased expression, they might also have a role in recruiting CLF. In case the telo box in the BP promoter is involved in CLF recruitment, mutation of the telo box should reduce CLF and H3K27me3 abundance. Expression of BP should be increased.

In an alternative scenario, CLF dependent genes are genes that are mainly repressed but need to be strongly expressed only in certain tissues or during certain developmental processes for short periods (e.g. at midnight, during replication or during transition to flowering). These genes could recruit factors that antagonize H3K27me3 via telo boxes or other *cis*-elements. A transgene with a mutated telo box should therefore have more H3K27me3 and possibly SWN-PRC2 abundance compared to a transgene with a non mutated telo box. Expression of such a mutated transgene should be decreased. .

To further complicate the picture, a screen for modulators of LHP1 conducted in our laboratory, identified mutations in TELOMERE REPEAT BINDING FACTOR 1 (TRB1) and TRB3 to enhance the phenotype of *lhp1* plants (unpublished, Ben Hartwig, MPIPZ, Cologne). *TRB1* and *TRB3* are expressed globally, the proteins bind to telomeric DNA via their Telobox domain and are located at the nucleolus where ribosome biogenesis takes place (Dvořáčková *et al.*, 2010). Interestingly, telo boxes are together with other *cis*-elements (site II motifs) enriched in the majority of genes encoding ribosomal proteins (Trémousaygue *et al.*, 2003; Wang *et al.*, 2011). TRB proteins could mediate binding to telo boxes via their telobox domain and act together with CLF to specifically regulate ribosome homeostasis. To test this hypothesis, expression of ribosomal proteins could be assessed in *clf* and *trb* mutant plants.

In summary, telo boxes are enriched in CLF dependent genes, however their association with the Polycomb machinery is not clear yet. Because telo boxes are associated with increased expression, I hypothesize that they antagonize CLF mediated gene repression. In addition I propose that CLF and TRBs work partially together and probably have a specific role in the regulation of ribosomal proteins.

5.4.4 A new model of CLF and SWN function in Arabidopsis

Phylogenetic comparisons revealed that CLF and SWN are both present in all analyzed plant species after the basal vascular plant *Selaginella moellendorffii*. They both display substantial differences in their middle part but have also conserved amino acid changes in the catalytic SET domain. ChIP-Seq and gene expression studies indicated that CLF and SWN cooperatively, antagonistically or independently influence H3K27me3 patterns and gene expression in a locus specific manner. Moreover, I detected an overrepresentation of the telo box *cis*-element in CLF dependent genes.

PRC2 complexes in plants control different genes and developmental processes by depositing the H3K27me3 mark depending on their subunit composition (Butenko & Ohad, 2011). In *Drosophila* and mammals, PRC2 complex components form smaller families than in Arabidopsis but both mammalian and Arabidopsis PRC2 proteins have undergone subfunctionalization (Simon & Kingston, 2013; Butenko & Ohad, 2011). The catalytic activity of the mammalian E(z) homologue Ezh2 is controlled through H3K27me3 in an auto activating feedback loop (Margueron *et al.*, 2009). H3K4me3, H3K36me2 and H3K36me3 inhibit Ezh2 activity via Suz12 and Esc (Schmitges *et al.*, 2011). Pcls and Jarid2 also fine-tune PRC2 activity and targeting (Simon & Kingston, 2013). Notably, Ezh2 can function in a PRC2 / H3K27me3 independent manner to activate transcription (Xu *et al.*, 2012). Ezh1 and Ezh2 form highly similar PRC2 complexes in mammals, but they confer repression through different mechanisms. The Ezh1-PRC2 complex compacts chromatin independently of H3K27me3. Its SET domain confers weak HMTase activity and is not necessary to repress gene expression. In contrast, Ezh2-PRC2 catalyses H3K27me3 and in the absence of Ezh2 genes are upregulated (Margueron *et al.*, 2008). In Arabidopsis, a SWN-PRC2 complex interacts with VRN2 to deposit H3K27me3 at *FLC* during vernalization while CLF is important for the H3K27me3 deposition at a broad range of genes like *AG*, *BP*, *FLC*, *FT* (De Lucia *et al.*, 2008; Lopez-Vernaza *et al.*, 2012). Moreover, H3K27me3 deposition is also supported by co-factors like the PHD finger proteins VIN3, VRN5 and VEL1 (Angel *et al.*, 2011; De Lucia *et al.*, 2008).

Based on our results and on the published literature, I suggest that CLF and SWN are the main catalytic components of PRC2 and that they have highly overlapping functions in H3K27me3 deposition. However, CLF and SWN have partially distinct effects on the Arabidopsis H3K27me3 pattern which might result from the differences in the middle part of the proteins and their SET domain. I also propose that the interplay between the proteins is important for wt-like H3K27me3 deposition and expression levels. Moreover, the function of CLF and SWN is presumably linked to the underlying genetic code and SWN lacks the ability to deposit H3K27me3 at some loci (Fig 5.1). I suggest three likely scenarios that could explain SWN function: Firstly, SWN could

have a different catalytic activity than CLF. Secondly, SWN could have a tissue specific function. Thirdly, SWN could have a PcG independent function. In all scenarios, the proposed SWN function might depend on an unknown binding partner.

In the first scenario, the increased H3K27me3 abundance at type 2 islands in the absence of SWN, could suggest that SWN might have a different enzymatic activity from CLF. In such a scenario, the catalytic activity of SWN would not be sufficient to mark genes with enhancing transcriptional elements such as telo boxes as it was observed in CLF dependent genes. It is possible that SWN also cannot methylate substrates carrying active chromatin marks. Alternatively, the putatively weaker HMTase SWN might antagonize strong HMTase CLF for binding to PCR2 components and methylation of target genes. In such a scenario, SWN would be needed to maintain proper H3K27me3 homeostasis and thus confer more flexibility to the PcG pathway. This hypothesis is supported by the observation that expression of CLF dependent genes in *clf* and *swn* mutants can be opposing. However, these transcriptional changes might also be explained by indirect effects. In such a scenario, SWN affected genes could encode for repressors of CLF dependent genes. In the absence of SWN, abundance of these repressors would be increased and mitigate expression of some CLF dependent genes.

In the second scenario, the fact that no genes could be detected that have reduced H3K27me3 in *swn* could suggest that SWN is not essential for the PRC2 mediated regulation of a specific developmental process. However, SWN could be detected in a CLF free complex during vernalization (De Lucia *et al.*, 2008; Wood *et al.*, 2006). Thus SWN might have a role in catalyzing H3K27me3 deposition at conditions or tissues that have not been assessed in this approach. A *swn* specific phenotype might therefore be visible under different growing conditions.

In the third scenario, the fact that expression of several genes is modulated by SWN although H3K27me3 levels do not directly change could indicate that SWN has a PRC2 independent function in gene regulation. Like the mammalian Ezh1, SWN could mediate chromatin compaction independently of H3K27me3 (Margueron *et al.*, 2008). Alternatively, SWN could function like the mammalian Ezh2 as a transcriptional activaror (Xu *et al.*, 2012). Interaction via the middle part to transcription factors could mediate transcriptional changes. In such a case, the SWN transcriptional activator complex could also antagonize CLF-PRC2.

New tools for fast ethyl methanesulfonate-induced screens have emerged through next generation sequencing (Hartwig *et al.*, 2012). A screen for SWN dependent phenotypes in the *swn* mutant, could shed more light on the role of SWN and reveal novel links between the Polycomb machinery and plant development. Y2H assays using the middle part of SWN could also unravel new interaction partners of SWN. Depending on the potential SWN interacting proteins, *swn* null mutant plants could be grown under a variety of environmental conditions to identify phenotypic changes that do

not occur in wt plants. However, Y2H screens should also be conducted with the CLF middle part since it is more conserved than the SWN middle part. Genome wide patterns of chromatin marks that are associated with enhanced transcription could be determined in *swn* and *clf* to assess if these marks are affected by CLF and SWN. In addition, analysis of H3K27me3 patterns and CLF or SWN abundance in wt plants compared to plants containing mutated versions of telo boxes could uncover if telo boxes have an inhibitory effect on the CLF and SWN HMTase function (Lodha *et al.*, 2013). Development of specific antibodies against CLF and SWN would help dissect the biochemical functions of these proteins (Hyun *et al.*, 2013). In case the proteins could be stably expressed in cell cultures, their catalytic activity could be assessed in HMT assays (Schmitges *et al.*, 2011). Finally, the direct effect of CLF-PRC2 and SWN-PRC2 on chromatin packaging could be assessed by electron microscopy (Margueron *et al.*, 2008).

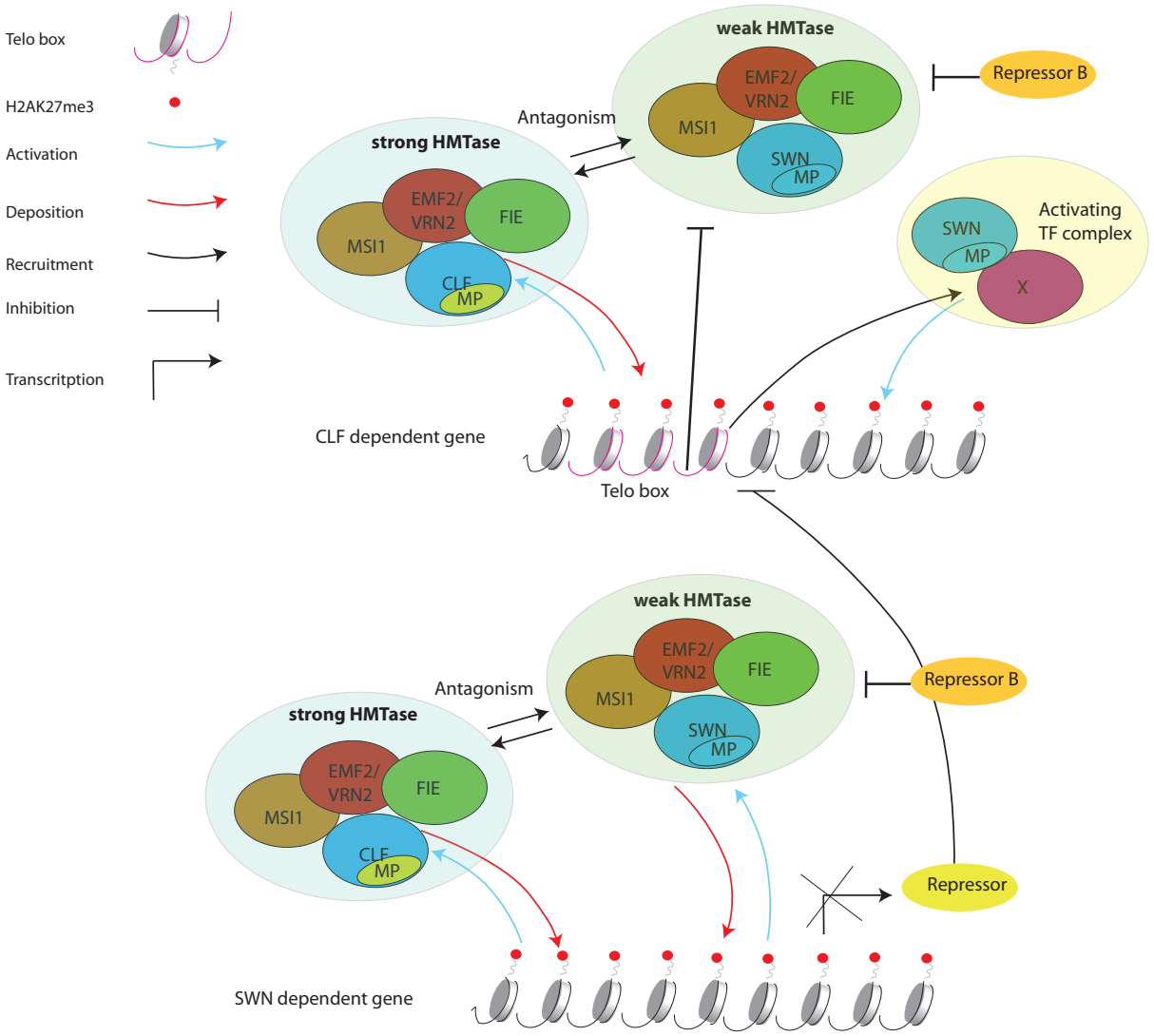


Figure 5.1: **CLF-PRC2 and SWN-PRC2 which have distinct enzymatic activities work together, independently and antagonistically to regulate transcription in a PRC2 dependent and maybe also a PRC2 independent pathway** .
A: CLF dependent genes. A PRC2 complex containing CLF deposits H3K27me3 to confer gene repression on CLF dependent genes. Telo boxes and other enhancing transcriptional elements inhibit H3K27me3 deposition and at these loci the enzymatic activity of SWN-PRC2 is not sufficient to deposit H3K27me3. SWN-PRC2 enzymatic activity might also be inhibited by an unknown factor (Repressor B). At some loci SWN interact via its MP with TFs to work as a transcriptional activator in a PRC2 independent and possibly antagonistic manner.
B: SWN affected genes and CLF or SWN dependent genes. The SWN-PRC2 complex slightly inhibits the CLF-PRC2 complex at highly accessible SWN affected genes to establish a H3K27me3 equilibrium at SWN affected genes. SWN affected genes might encode for repressors of CLF dependent genes. At less accessible SWN or CLF dependent genes, the two PRC2 complexes act similarly.
CLF in blue, EMF2 / VRN2 in red, FIE in green, MSI1 in brown, Middle part (MP) in yellow, Transcription factor (TF) in light yellow, Repressor A in dark yellow, Repressor B in orange, SWN inter protein (X) in magenta.

Bibliography

ACH, R A, TARANTO, P, & GRUISSEM, W. 1997. A Conserved Family of WD-40 Proteins Binds to the Retinoblastoma Protein in Both Plants and Animals. *Plant Cell*, **9**(9), 1595–1606. [3]

ADRIAN, J, FARRONA, S, REIMER, J J, ALBANI, M C, COUPLAND, G, & TURCK, F. 2010. cis-Regulatory Elements and Chromatin State Coordinately Control Temporal and Spatial Expression of FLOWERING LOCUS T in Arabidopsis. *Plant Cell*, **22**(5), 1425–1440. [23, 37, 55]

ADRIAN, JESSIKA, TORTI, STEFANO, & TURCK, FRANZISKA. 2009. From Decision to Commitment: The Molecular Memory of Flowering. *Mol. Plant*, **2**(4), 628–642. [3]

AMASINO, RICHARD. 2009. Floral induction and monocarpic versus polycarpic life histories. *Genome Biol.*, **10**(7), 228. [14]

ANGEL, ANDREW, SONG, JIE, DEAN, CAROLINE, & HOWARD, MARTIN. 2011. A Polycomb-based switch underlying quantitative epigenetic memory. *Nature*, **476**(7358), 105–108. [15, 55, 73, 78, 84]

AOYAMA, T, & CHUA, N H. 1997. A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J.*, **11**(3), 605–612. [75]

BAILEY, TIMOTHY L, BODEN, MIKAEL, BUSKE, FABIAN A, FRITH, MARTIN, GRANT, CHARLES E, CLEMENTI, LUCA, REN, JINGYUAN, LI, WILFRED W, & NOBLE, WILLIAM S. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**(Web Server issue), W202–8. [26]

BAUMBUSCH, L O, THORSTENSEN, T, KRAUSS, V, FISCHER, A, NAUMANN, K, ASSALKHOU, R, SCHULZ, I, REUTER, G, & AALEN, R B. 2001. The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res.*, **29**(21), 4319–4333. [71]

BEISEL, CHRISTIAN, & PARO, RENATO. 2011. Silencing chromatin: comparing modes and mechanisms. *Nat. Rev. Genet.*, **12**(2), 123–135. [7]

BEN-PORATH, ITTAI, THOMSON, MATTHEW W, CAREY, VINCENT J, GE, RUPING, BELL, GEORGE W, REGEV, AVIV, & WEINBERG, ROBERT A. 2008. An embryonic

stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.*, **40**(5), 499–507. [1]

BERGER, NATHALIE, DUBREUCQ, BERTRAND, ROUDIER, FRANÇOIS, DUBOS, CHRISTIAN, & LEPINIEC, LOÏC. 2011. Transcriptional Regulation of Arabidopsis LEAFY COTYLEDON2 Involves RLE, a cis-Element That Regulates Trimethylation of Histone H3 at Lysine-27. *Plant Cell*, **23**(11), 4065–4078. [6, 8]

BERKE, LIDIJA, SANCHEZ-PEREZ, GABINO F, & SNEL, BEREND. 2012. Contribution of the epigenetic mark H3K27me3 to functional divergence after whole genome duplication in Arabidopsis. *Genome Biol.*, **13**(10), R94. [81]

BOUYER, DANIEL, ROUDIER, FRANÇOIS, HEESE, MAREN, ANDERSEN, ELLEN D, GEY, DELPHINE, NOWACK, MORITZ K, GOODRICH, JUSTIN, RENOU, JEAN-PIERRE, GRINI, PAUL E, COLOT, VINCENT, & SCHNITTGER, ARP. 2011. Polycomb Repressive Complex 2 Controls the Embryo-to-Seedling Phase Transition. *PLoS Genet.*, **7**(3), e1002014. [2, 12, 14]

BOYER, LAURIE A, PLATH, KATHRIN, ZEITLINGER, JULIA, BRAMBRINK, TOBIAS, MEDEIROS, LEA A, LEE, TONG IHN, LEVINE, STUART S, WERNIG, MARIUS, TAJONAR, ADRIANA, RAY, MRIDULA K, BELL, GEORGE W, OTTE, ARIE P, VIDAL, MIGUEL, GIFFORD, DAVID K, YOUNG, RICHARD A, & JAENISCH, RUDOLF. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, **441**(7091), 349–353. [11]

BRATZEL, FABIAN, LÓPEZ-TORREJÓN, GEMA, KOCH, MARCUS, DEL POZO, JUAN C, & CALONJE, MYRIAM. 2010. Keeping Cell Identity in Arabidopsis Requires PRC1 RING-Finger Homologs that Catalyze H2A Monoubiquitination. *Curr. Biol.*, **20**(20), 1853–1859. [4]

BRATZEL, FABIAN, YANG, CHAO, ANGELOVA, ALEXANDRA, LÓPEZ-TORREJÓN, GEMA, KOCH, MARCUS, DEL POZO, JUAN CARLOS, & CALONJE, MYRIAM. 2011. Regulation of the New Arabidopsis Imprinted Gene AtBMI1C Requires the Interplay of Different Epigenetic Mechanisms. *Mol. Plant*, Sept. [12]

BUTENKO, YANA, & OHAD, NIR. 2011. Polycomb-group mediated epigenetic mechanisms through plant evolution. *Biochim. Biophys. Acta*, **1809**(8), 395–406. [5, 70, 84]

CAO, RU, & ZHANG, YI. 2004. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr. Opin. Genet. Dev.*, **14**(2), 155–164. [2]

CAO, RU, WANG, LIANGJUN, WANG, HENGBIN, XIA, LI, ERDJUMENT-BROMAGE, HEDIYE, TEMPST, PAUL, JONES, RICHARD S, & ZHANG, YI. 2002. Role of histone H3

lysine 27 methylation in Polycomb-group silencing. *Science*, **298**(5595), 1039–1043. [1, 8, 71]

CHANVIVATTANA, YINDEE, BISHOPP, ANTHONY, SCHUBERT, DANIEL, STOCK, CHRISTINE, MOON, YONG-HWAN, SUNG, Z RENEE, & GOODRICH, JUSTIN. 2004. Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. *Development*, **131**(21), 5263–5276. [3, 11, 12, 13, 28, 32, 50, 70, 73, 77]

CHEN, DONGHONG, MOLITOR, ANNE, LIU, CHUNLIN, & SHEN, WEN-HUI. 2010. The Arabidopsis PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth. *Cell Res.*, **20**(12), 1332–1344. [4, 12]

CHOMCZYNSKI, PIOTR, & SACCHI, NICOLETTA. 2006. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat. Protoc.*, **1**(2), 581–585. [23]

CIFERRI, CLAUDIO, LANDER, GABRIEL C, MAIOLICA, ALESSIO, HERZOG, FRANZ, AEBERSOLD, RUEDI, & NOGALES, EVA. 2012. Molecular architecture of human polycomb repressive complex 2. *Elife*, **1**, e00005. [8, 9, 10, 32, 71, 72]

CLAMP, MICHELE, CUFF, JAMES, SEARLE, STEPHEN M, & BARTON, GEOFFREY J. 2004. The Jalview Java alignment editor. *Bioinformatics*, **20**(3), 426–427. [18, 32]

CLOUGH, STEVEN J, & BENT, ANDREW F. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.*, **16**(6), 735–743. [19]

CZERMIN, BIRGIT, MELFI, RAFFAELLA, MCCABE, DONNA, SEITZ, VOLKER, IMHOF, AXEL, & PIRROTTA, VINCENZO. 2002. Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. *Cell*, **111**(2), 185–196. [1, 71]

DE LA SERNA, IVANA L, OHKAWA, YASUYUKI, & IMBALZANO, ANTHONY N. 2006. Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nat. Rev. Genet.*, **7**(6), 461–473. [79]

DE LUCIA, FILOMENA, CREVILLEN, PEDRO, JONES, ALEXANDRA M E, GREB, THOMAS, & DEAN, CAROLINE. 2008. A PHD-Polycomb Repressive Complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proc. Natl. Acad. Sci. USA*, **105**(44), 16831–16836. [7, 14, 15, 21, 73, 76, 84, 85]

DOYLE, MARK R, & AMASINO, RICHARD M. 2009. A single amino acid change in the enhancer of zeste ortholog CURLY LEAF results in vernalization-independent, rapid flowering in Arabidopsis. *Plant Physiol.*, **151**(3), 1688–1697. [80]

DUMBLIAUSKAS, EVA, LECHNER, ESTHER, JACIUBEK, MIŁOSŁAWA, BERR, ALEXANDRE, PAZHOUHANDEH, MAGHSOUD, ALIOUA, MALEK, COGNAT, VALERIE, BRUKHIN, VLADIMIR, KONCZ, CSABA, GROSSNIKLAUS, UELI, MOLINIER, JEAN, & GENSCHIK, PASCAL. 2011. The Arabidopsis CUL4-DDB1 complex interacts with MSI1 and is required to maintain MEDEA parental imprinting. *EMBO J.*, **30**(4), 731–743. [75]

DVOŘÁČKOVÁ, MARTINA, ROSSIGNOL, PASCALE, SHAW, PETER J, KOROLEVA, OLGA A, DOONAN, JOHN H, & FAJKUS, JIRÍ. 2010. AtTRB1, a telomeric DNA-binding protein from Arabidopsis, is concentrated in the nucleolus and shows highly dynamic association with chromatin. *Plant J.*, **61**(4), 637–649. [83]

ENGELHORN, JULIA, REIMER, JULIA J, LEUZ, IRIS, GOBEL, ULRIKE, HUETTEL, BRUNO, FARRONA, SARA, & TURCK, FRANZISKA. 2012. DEVELOPMENT-RELATED PcG TARGET IN THE APEX 4 controls leaf margin architecture in Arabidopsis thaliana. *Development*, **139**(14), 2566–2575. [35]

ESKELAND, RAGNHILD, LEEB, MARTIN, GRIMES, GRAEME R, KRESS, CLÉMENCE, BOYLE, SHELAGH, SPROUL, DUNCAN, GILBERT, NICK, FAN, YUHONG, SKOULTCHI, ARTHUR I, WUTZ, ANTON, & BICKMORE, WENDY A. 2010. Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. *Mol. Cell*, **38**(3), 452–464. [2]

FARRONA, S, COUPLAND, G, & TURCK, F. 2008. The impact of chromatin regulation on the floral transition. *Semin. Cell Dev. Biol.*, **19**(6), 560–573. [1, 3, 4, 50]

FARRONA, SARA, THORPE, FRAZER L, ENGELHORN, JULIA, ADRIAN, JESSIKA, DONG, XUE, SARID-KREBS, LIRON, GOODRICH, JUSTIN, & TURCK, FRANZISKA. 2011. Tissue-specific expression of FLOWERING LOCUS T in Arabidopsis is maintained independently of polycomb group protein repression. *Plant Cell*, **23**(9), 3204–3214. [2, 12, 17, 59, 60, 62, 77, 79, 80]

FELSENSTEIN, J. 1985. Confidence Limits on Phylogenies With a Molecular Clock. *Syst. Biol.*, **34**(2), 152–161. [18, 29]

FINNEGAN, E JEAN, & DENNIS, ELIZABETH S. 2007. Vernalization-Induced Trimethylation of Histone H3 Lysine 27 at FLC Is Not Maintained in Mitotically Quiescent Cells. *Curr. Biol.*, **17**(22), 1978–1983. [15, 73]

FITCH, W M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Biol.* [18, 29]

FLOWERS, STEPHEN, NAGL, NORMAN G, BECK, GEORGE R, & MORAN, ELIZABETH. 2009. Antagonistic roles for BRM and BRG1 SWI/SNF complexes in differentiation. *J. Biol. Chem.*, **284**(15), 10067–10075. [79]

FRANCIS, NICOLE J, KINGSTON, ROBERT E, & WOODCOCK, CHRISTOPHER L. 2004. Chromatin compaction by a polycomb group protein complex. *Science*, **306**(5701), 1574–1577. [2]

FUREY, TERRENCE S. 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.*, **13**(12), 840–852. [44]

GAO, ZHONGHUA, ZHANG, JIN, BONASIO, ROBERTO, STRINO, FRANCESCO, SAWAI, AYANA, PARISI, FABIO, KLUGER, YUVAL, & REINBERG, DANNY. 2012. PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol. Cell*, **45**(3), 344–356. [78]

GASPIN, CHRISTINE, RAMI, JEAN-FRANÇOIS, & LESCURE, BERNARD. 2010. Distribution of short interstitial telomere motifs in two plant genomes: putative origin and function. *BMC Plant Biol.*, **10**, 283. [82]

GENDALL, ANTHONY R, LEVY, YARON Y, WILSON, ALLISON, & DEAN, CAROLINE. 2001. The VERNALIZATION 2 Gene Mediates the Epigenetic Regulation of Vernalization in Arabidopsis. *Cell*, **107**(4), 525–535. [3]

GENTLEMAN, ROBERT C, CAREY, VINCENT J, BATES, DOUGLAS M, BOLSTAD, BEN, DETTLING, MARCEL, DUDOIT, SANDRINE, ELLIS, BYRON, GAUTIER, LAURENT, GE, YONGCHAO, GENTRY, JEFF, HORNIK, KURT, HOTHORN, TORSTEN, HUBER, WOLFGANG, IACUS, STEFANO, IRIZARRY, RAFAEL, LEISCH, FRIEDRICH, LI, CHENG, MAECHLER, MARTIN, ROSSINI, ANTHONY J, SAWITZKI, GUNTHER, SMITH, COLIN, SMYTH, GORDON, TIERNEY, LUKE, YANG, JEAN Y H, & ZHANG, JIANHUA. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**(10), R80. [26]

GLEASON, EMILY J, & KRAMER, ELENA M. 2012. Characterization of Aquilegia Polycomb Repressive Complex 2 homologs reveals absence of imprinting. *Gene*, **507**(1), 54–60. [5, 17, 70]

GOBEL, ULRIKE, REIMER, JULIA, & TURCK, FRANZISKA. 2010. Genome-wide mapping of protein-DNA interaction by chromatin immunoprecipitation and DNA microarray hybridization (ChIP-chip). Part B: ChIP-chip data analysis. *Methods Mol. Biol.*, **631**, 161–184. [27]

GOODRICH, JUSTIN, PUANGSOMLEE, PREEYA, MARTIN, MARTA, LONG, DEBORAH, MEYEROWITZ, ELLIOT M, & COUPLAND, GEORGE. 1997. A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. *Nature*, **386**(6620), 44–51. [3, 12]

GROSSNIKLAUS, UELI, VIELLE-CALZADA, JEAN-PHILIPPE, HOEPPNER, MARILU A, & GAGLIANO, WENDY B. 1998. Maternal Control of Embryogenesis by MEDEA, a Polycomb Group Gene in Arabidopsis. *Science*, **280**(5362), 446–450. [3]

HARTWIG, BENJAMIN, JAMES, GEO VELIKKAKAM, KONRAD, KATHRYN, SCHNEEBERGER, KORBINIAN, & TURCK, FRANZISKA. 2012. Fast isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks. *Plant Physiol.*, **160**(2), 591–600. [85]

HE, CHONGSHENG, CHEN, XIAOFAN, HUANG, HAI, & XU, LIN. 2012. Reprogramming of H3K27me3 is critical for acquisition of pluripotency from cultured Arabidopsis tissues. *PLoS Genet.*, **8**(8), e1002911. [11]

HENIKOFF, STEVEN, & SHILATIFARD, ALI. 2011. Histone modification: cause or cog? *Trends Genet.*, **27**(10), 389–396. [3, 9]

HENNIG, LARS, & DERKACHEVA, MARIA. 2009. Diversity of Polycomb group complexes in plants: same rules, different players? *Trends Genet.*, **25**(9), 414–423. [1]

HENNIG, LARS, TARANTO, PATTI, WALSER, MARCEL, SCHÖNROCK, NICOLE, & GRUISSEM, WILHELM. 2003. Arabidopsis MSI1 is required for epigenetic maintenance of reproductive development. *Development*, **130**(12), 2555–2565. [3]

HEO, JAE BOK, & SUNG, SIBUM. 2011a. Encoding memory of winter by noncoding RNAs. *Epigenetics*, **6**(5), 544–547. [76]

HEO, JAE BOK, & SUNG, SIBUM. 2011b. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*, **331**(6013), 76–79. [7, 8, 15, 81]

HOW KIT, A, BOUREAU, L, STAMMITTI-BERT, L, ROLIN, D, TEYSSIER, E, & GALLUSCI, P. 2010. Functional analysis of SlEZ1 a tomato enhancer of zeste (E(z)) gene demonstrates a role in flower development. *Plant Mol. Biol.*, **74**(3), 201–213. [1, 5, 17, 50, 70, 81]

HU, HUILI, YANG, YANG, JI, QINGHONG, ZHAO, WEI, JIANG, BAICHUN, LIU, RUIQIONG, YUAN, JUPENG, LIU, QIAO, LI, XI, ZOU, YONGXIN, SHAO, CHANGSHUN, SHANG, YONGFENG, WANG, YAN, & GONG, YAOQIN. 2012. CRL4B catalyzes H2AK119 monoubiquitination and coordinates with PRC2 to promote tumorigenesis. *Cancer Cell*, **22**(6), 781–795. [75]

HYUN, YOUBONG, YUN, HYEIN, PARK, KYUNGHYUK, OHR, HYONHWA, LEE, OKCHAN, KIM, DONG-HWAN, SUNG, SIBUM, & CHOI, YEONHEE. 2013. The catalytic subunit of Arabidopsis DNA polymerase alpha ensures stable maintenance of histone modification. *Development*, **140**(1), 156–166. [86]

JEONG, CHEOL WOONG, ROH, HYUNGMIN, DANG, TUONG VI, CHOI, YANG DO, FISCHER, ROBERT L, LEE, JONG SEOB, & CHOI, YEONHEE. 2011. An E3 ligase complex regulates SET-domain polycomb group protein activity in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. USA*, **108**(19), 8036–8041. [36, 47, 70, 74, 75, 79]

JIANG, DANHUA, WANG, YUQI, WANG, YIZHONG, & HE, YUEHUI. 2008. Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb repressive complex 2 components. *PLOS ONE*, **3**(10), e3404. [13]

KANHERE, ADITI, VIIRI, KEIJO, ARAÚJO, CARLA C, RASAIYAAH, JANE, BOUWMAN, RUSSELL D, WHYTE, WARREN A, PEREIRA, C FILIPE, BROOKES, EMILY, WALKER, KIMBERLY, BELL, GEORGE W, POMBO, ANA, FISHER, AMANDA G, YOUNG, RICHARD A, & JENNER, RICHARD G. 2010. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell*, **38**(5), 675–688. [7]

KATZ, AVIVA, OLIVA, MORAN, MOSQUNA, ASSAF, HAKIM, OFIR, & OHAD, NIR. 2004. FIE and CURLY LEAF polycomb proteins interact in the regulation of homeobox gene expression during sporophyte development - Katz - 2004 - The Plant Journal - Wiley Online Library. *Plant J.*, **37**(5), 707–719. [12]

KENZIOR, ALEXANDER L, & FOLK, WILLIAM R. 1998. AtMSI4 and RbAp48 WD-40 repeat proteins bind metal ions. *FEBS Lett.*, **440**(3), 425–429. [3]

KIM, HANA, KANG, KEUNSOO, & KIM, JOOMYEONG. 2009. AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res.*, **37**(9), 2940–2950. [7]

KIM, S Y, ZHU, T, & SUNG, Z R. 2010. Epigenetic Regulation of Gene Programs by EMF1 and EMF2 in Arabidopsis. *Plant Physiol.*, **152**(2), 516–528. [13]

KIM, SANG YEOL, LEE, JUNGEUN, ESHED-WILLIAMS, LEOR, ZILBERMAN, DANIEL, & SUNG, Z RENEE. 2012. EMF1 and PRC2 cooperate to repress key regulators of Arabidopsis development. *PLoS Genet.*, **8**(3), e1002512. [13]

KLYMENKO, TETYANA, PAPP, BERNADETT, FISCHLE, WOLFGANG, KÖCHER, THOMAS, SCHELDER, MALGORZATA, FRITSCH, CORNELIA, WILD, BRIGITTE, WILM, MATTHIAS, & MÜLLER, JÜRG. 2006. A Polycomb group protein complex

with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev.*, **20**(9), 1110–1122. [4]

KÖHLER, CLAUDIA, HENNIG, LARS, BOUVERET, ROMARIC, GHEYSELINCK, JACQUELINE, GROSSNIKLAUS, UELI, & GRUISSEM, WILHELM. 2003a. Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. *EMBO J.*, **22**(18), 4804–4814. [14]

KÖHLER, CLAUDIA, HENNIG, LARS, SPILLANE, CHARLES, PIEN, STÉPHANE, GRUISSEM, WILHELM, & GROSSNIKLAUS, UELI. 2003b. The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev.*, **17**(12), 1540–1553. [2, 13, 14, 70, 77]

KONCZ, CSABA, & SCHELL, JEFF. 1986. The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Mol. Genet.*, **204**(3), 383–396. [19]

KRAJEWSKI, WLADYSLAW A, NAKAMURA, TATSUYA, MAZO, ALEXANDER, & CANAANI, ELI. 2005. A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. *Mol. Cell. Biol.*, **25**(5), 1891–1899. [12]

KU, MANCHING, KOCHE, RICHARD P, RHEINBAY, ESTHER, MENDENHALL, ERIC M, ENDOH, MITSUHIRO, MIKKELSEN, TARJEI S, PRESSER, AVIVA, NUSBAUM, CHAD, XIE, XIAOHUI, CHI, ANDREW S, ADLI, MAZHAR, KASIF, SIMON, PTASZEK, LEON M, COWAN, CHAD A, LANDER, ERIC S, KOSEKI, HARUHIKO, & BERNSTEIN, BRADLEY E. 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.*, **4**(10), e1000242. [6, 11]

LAFOS, MARCEL, KROLL, PHILLIP, HOHENSTATT, MAREIKE L, THORPE, FRAZER L, CLARENZ, OLIVER, & SCHUBERT, DANIEL. 2011. Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. *PLoS Genet.*, **7**(4), e1002040. [11, 60, 79, 80]

LAMESCH, PHILIPPE, BERARDINI, TANYA Z, LI, DONGHUI, SWARBRECK, DAVID, WILKS, CHRISTOPHER, SASIDHARAN, RAJKUMAR, MULLER, ROBERT, DREHER, KATE, ALEXANDER, DEBBIE L, GARCIA-HERNANDEZ, MARGARITA, KARTHIKEYAN, ATHIKKATTUVALASU S, LEE, CYNTHIA H, NELSON, WILLIAM D, PLOETZ, LARRY, SINGH, SHANKER, WENSEL, APRIL, & HUALA, EVA. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.*, **40**(Jan.), D1202–10. [26]

LE CORRE, VALÉRIE, ROUX, FABRICE, & REBOUD, XAVIER. 2002. DNA polymorphism at the FRIGIDA gene in Arabidopsis thaliana: extensive nonsynonymous variation is consistent with local selection for flowering time. *Mol. Biol. Evol.*, **19**(8), 1261–1271. [29]

LETUNIC, I, COPLEY, R R, & SCHMIDT, S. 2004. SMART 4.0: towards genomic data integration. *Nucleic Acids Res.* [18, 32, 72]

LI, GANG, MARGUERON, RAPHAËL, KU, MANCHING, CHAMBON, PIERRE, BERNSTEIN, BRADLEY E, & REINBERG, DANNY. 2010. Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.*, **24**(4), 368–380. [6, 7]

LI, HENG, & DURBIN, RICHARD. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**(14), 1754–1760. [26, 44]

LI, HENG, HANDSAKER, BOB, WYSOKER, ALEC, FENNELL, TIM, RUAN, JUE, HOMER, NILS, MARTH, GABOR, ABECASIS, GONCALO, DURBIN, RICHARD, & 1000 GENOME PROJECT DATA PROCESSING SUBGROUP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**(16), 2078–2079. [26, 27, 44]

LIU, JUN, JUNG, CHOONKYUN, XU, JUN, WANG, HUAN, DENG, SHULIN, BERNAD, LUCIA, ARENAS-HUERTERO, CATALINA, & CHUA, NAM-HAI. 2012. Genome-Wide Analysis Uncovers Regulation of Long Intergenic Noncoding RNAs in Arabidopsis. *Plant Cell*, Nov. [82]

LIU, X, KIM, Y J, MULLER, R, YUMUL, R E, LIU, C, PAN, Y, CAO, X, GOODRICH, J, & CHEN, X. 2011. AGAMOUS Terminates Floral Stem Cell Maintenance in Arabidopsis by Directly Repressing WUSCHEL through Recruitment of Polycomb Group Proteins. *Plant Cell*, **23**(10), 3654–3670. [7, 8, 13]

LODATO, MICHAEL A, NG, CHRISTOPHER W, WAMSTAD, JOSEPH A, CHENG, ALBERT W, THAI, KEVIN K, FRAENKEL, ERNEST, JAENISCH, RUDOLF, & BOYER, LAURIE A. 2013. SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet.*, **9**(2), e1003288. [76]

LODHA, MUKESH, MARCO, CRISTINA F, & TIMMERMANS, MARJA C P. 2013. The ASYMMETRIC LEAVES complex maintains repression of KNOX homeobox genes via direct recruitment of Polycomb-repressive complex 2. *Genes Dev.*, **27**(6), 596–601. [6, 7, 8, 13, 66, 79, 83, 86]

LOPEZ-VERNAZA, MANUEL, YANG, SUXIN, MÜLLER, RALF, THORPE, FRAZER, DE LEAU, ERICA, & GOODRICH, JUSTIN. 2012. Antagonistic roles of SEPALLATA3, FT and FLC genes as targets of the polycomb group gene CURLY LEAF. *PLoS ONE*, **7**(2), e30715. [12, 51, 76, 77, 84]

LU, FALONG, CUI, XIA, ZHANG, SHUAIBIN, JENUWEIN, THOMAS, & CAO, XIAOFENG. 2011. Arabidopsis REF6 is a histone H3 lysine 27 demethylase. *Nat. Genet.*, **43**(7), 715–719. [15]

LUO, MING, BILODEAU, PIERRE, KOLTUNOW, ANNA, DENNIS, ELIZABETH S, PEACOCK, W JAMES, & CHAUDHURY, ABDUL M. 1999. Genes controlling fertilization-independent seed development in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. USA*, **96**(1), 296–301. [3, 13]

LUO, MING, PLATTEN, DAMIEN, CHAUDHURY, ABED, PEACOCK, W J, & DENNIS, ELIZABETH S. 2009. Expression, imprinting, and evolution of rice homologs of the polycomb group genes. *Mol. Plant*, **2**(4), 711–723. [5, 6, 12]

MAKAREVICH, GRIGORY, LEROY, OLIVIER, AKINCI, UMUT, SCHUBERT, DANIEL, CLARENZ, OLIVER, GOODRICH, JUSTIN, GROSSNIKLAUS, UELI, & KÖHLER, CLAUDIA. 2006. Different Polycomb group complexes regulate common target genes in Arabidopsis. *EMBO Rep.*, **7**(9), 947–952. [13]

MARGUERON, RAPHAËL, & REINBERG, DANNY. 2011. The Polycomb complex PRC2 and its mark in life. *Nature*, **469**(7330), 343–349. [1, 2, 3, 6, 7, 11]

MARGUERON, RAPHAËL, LI, GUOHONG, SARMA, KAVITHA, BLAIS, ALEXANDRE, ZAVADIL, JIRI, WOODCOCK, CHRISTOPHER L, DYNLACHT, BRIAN D, & REINBERG, DANNY. 2008. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell*, **32**(4), 503–518. [1, 2, 9, 10, 17, 71, 84, 85, 86]

MARGUERON, RAPHAËL, JUSTIN, NEIL, OHNO, KATSUHITO, SHARPE, MIRIAM L, SON, JINSOOK, DRURY, WILLIAM J, VOIGT, PHILIPP, MARTIN, STEPHEN R, TAYLOR, WILLIAM R, DE MARCO, VALERIA, PIRROTTA, VINCENZO, REINBERG, DANNY, & GAMBLIN, STEVEN J. 2009. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*, **461**(7265), 762–767. [7, 8, 9, 15, 17, 71, 84]

MCCABE, MICHAEL T, GRAVES, ALAN P, GANJI, GOPINATH, DIAZ, ELSIE, HALSEY, WENDY S, JIANG, YONG, SMITHEMAN, KIMBERLY N, OTT, HEIDI M, PAPPALARDI, MELISSA B, ALLEN, KIMBERLY E, CHEN, STEPHANIE B, DELLA PIETRA, ANTHONY, DUL, EDWARD, HUGHES, ASHLEY M, GILBERT, SETH A, THRALL, SARA H, TUMMINO, PETER J, KRUGER, RYAN G, BRANDT, MARTIN, SCHWARTZ, BENJAMIN, & CREASY, CARETHA L. 2012. Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc. Natl. Acad. Sci. USA*, **109**(8), 2989–2994. [80]

MICHAEL, TODD P, MOCKLER, TODD C, BRETON, GHISLAIN, MCENTEE, CONNOR, BYER, AMANDA, TROUT, JONATHAN D, HAZEN, SAMUEL P, SHEN, RONGKUN,

PRIEST, HENRY D, SULLIVAN, CHRISTOPHER M, GIVAN, SCOTT A, YANOVSKY, MARCELO, HONG, FANGXIN, KAY, STEVE A, & CHORY, JOANNE. 2008. Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet.*, **4**(2), e14. [82]

MOSQUNA, ASSAF, KATZ, AVIVA, DECKER, EVA L, RENSING, STEFAN A, RESKI, RALF, & OHAD, NIR. 2009. Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. *Development*, **136**(14), 2433–2444. [5]

MÜLLER, JÜRG, & KASSIS, JUDITH A. 2006. Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Curr. Opin. Genet. Dev.*, **16**(5), 476–484. [4]

MYLNE, JOSHUA S, BARRETT, LYNNE, TESSADORI, FEDERICO, MESNAGE, STEPHANE, JOHNSON, LIANNA, BERNATAVICHUTE, YANA V, JACOBSEN, STEVEN E, FRANSZ, PAUL, & DEAN, CAROLINE. 2006. LHP1, the Arabidopsis homologue of HETEROCHROMATIN PROTEIN1, is required for epigenetic silencing of FLC. *Proc. Natl. Acad. Sci. USA*, **103**(13), 5012–5017. [2, 15]

NEI, M, & KUMAR, S. Mol. Evol. Phylogenet., year = 2000. [18]

OHAD, NIR, YADEGARI, RAMIN, MARGOSSIAN, LINDA, HANNON, MIKE, MICHAELI, DAPHNA, HARADA, JOHN J, GOLDBERG, ROBERT B, & FISCHER, ROBERT L. 1999. Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. *Plant Cell*, **11**(3), 407–416. [3, 14]

O'MEARA, M MAGGIE, & SIMON, JEFFREY A. 2012. Inner workings and regulatory inputs that control Polycomb repressive complex 2. *Chromosoma*, **121**(3), 221–234. [9]

PAZHOUHANDEH, MAGHSOUD, MOLINIER, JEAN, BERR, ALEXANDRE, & GENSCHIK, PASCAL. 2011. MSI4/FVE interacts with CUL4-DDB1 and a PRC2-like complex to control epigenetic regulation of flowering time in Arabidopsis. *Proc. Natl. Acad. Sci. USA*, **108**(8), 3430–3435. [75]

PIEN, STÉPHANE, & GROSSNIKLAUS, UELI. 2007. Polycomb group and trithorax group proteins in Arabidopsis. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, **1769**(5-6), 375–382. [2, 3, 4, 13, 78]

PLATH, KATHRIN, FANG, JIA, MLYNARCZYK-EVANS, SUSANNA K, CAO, RU, WORRINGER, KATHLEEN A, WANG, HENGBIN, DE LA CRUZ, CECILE C, OTTE, ARIE P, PANNING, BARBARA, & ZHANG, YI. 2003. Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science*, **300**(5616), 131–135. [7]

RICHARDS, E J, & AUSUBEL, F M. 1988. Isolation of a Higher Eukaryotic Telomere From Arabidopsis Thaliana. *Cell*, **53**(1), 127–136. [82]

RINN, JOHN L, KERTESZ, MICHAEL, WANG, JORDON K, SQUAZZO, SHARON L, XU, XIAO, BRUGMANN, SAMANTHA A, GOODNOUGH, L HENRY, HELMS, JILL A, FARNHAM, PEGGY J, SEGAL, ERAN, & CHANG, HOWARD Y. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**(7), 1311–1323. [7]

SAITOU, N, & NEI, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**(4), 406–425. [18]

SATAKE, AKIKO, & IWASA, YOH. 2012. A stochastic model of chromatin modification: cell population coding of winter memory in plants. *J. Theor. Biol.*, **302**(June), 6–17. [15]

SCHATLOWSKI, NICOLE, STAHL, YVONNE, HOHENSTATT, MAREIKE L, GOODRICH, JUSTIN, & SCHUBERT, DANIEL. 2010. The CURLY LEAF interacting protein BLISTER controls expression of polycomb-group target genes and cellular differentiation of Arabidopsis thaliana. *Plant Cell*, **22**(7), 2291–2305. [12]

SCHEUERMANN, JOHANNA C, DE AYALA ALONSO, ANDRES GAYTAN, OKTABA, KATARZYNA, LY-HARTIG, NGA, MCGINTY, ROBERT K, FRATERMAN, SVEN, WILM, MATTHIAS, MUIR, TOM W, & MÜLLER, JÜRG. 2010. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature*, **465**(7295), 243–247. [4]

SCHMITGES, FRANK W, PRUSTY, ARCHANA B, FATY, MAHAMADOU, STÜTZER, ALEXANDRA, LINGARAJU, GONDICHATNAHALLI M, AIWAZIAN, JONATHAN, SACK, RAGNA, HESS, DANIEL, LI, LING, ZHOU, SHAOLIAN, BUNKER, RICHARD D, WIRTH, URS, BOUWMEESTER, TEWIS, BAUER, ANDREAS, LY-HARTIG, NGA, ZHAO, KEHAO, CHAN, HOMAN, GU, JUSTIN, GUT, HEINZ, FISCHLE, WOLFGANG, MÜLLER, JÜRG, & THOMÄ, NICOLAS H. 2011. Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell*, **42**(3), 330–341. [1, 2, 8, 9, 14, 15, 17, 32, 71, 78, 84, 86]

SCHÖNROCK, NICOLE, BOUVERET, ROMARIC, LEROY, OLIVIER, BORGHI, LORENZO, KÖHLER, CLAUDIA, GRUISSEM, WILHELM, & HENNIG, LARS. 2006. Polycomb-group proteins repress the floral activator AGL19 in the FLC-independent vernalization pathway. *Genes Dev.*, **20**(12), 1667–1678. [13]

SCHUBERT, DANIEL, PRIMAVESI, LUCIA, BISHOPP, ANTHONY, ROBERTS, GETHIN, DOONAN, JOHN, JENUWEIN, THOMAS, & GOODRICH, JUSTIN. 2006. Silencing by

plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *EMBO J.*, **25**(19), 4638–4649. [2, 13, 39, 74]

SCHUETTENGRUBER, BERND, & CAVALLI, GIACOMO. 2009. Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development*, **136**(21), 3531–3542. [1, 6]

SCHWARTZ, YURI B, & PIRROTTA, VINCENZO. 2007. Polycomb silencing mechanisms and the management of genomic programmes. *Nat. Rev. Genet.*, **8**(1), 9–22. [3]

SCHWARTZ, YURI B, KAHN, TATYANA G, NIX, DAVID A, LI, XIAO-YONG, BOURGON, RICHARD, BIGGIN, MARK, & PIRROTTA, VINCENZO. 2006. Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nat. Genet.*, **38**(6), 700–705. [11]

SEARLE, IAIN, HE, YUEHUI, TURCK, FRANZISKA, VINCENT, CORAL, FORNARA, FABIO, KR BER, SANDRA, AMASINO, RICHARD A, & COUPLAND, GEORGE. 2006. The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis. *Genes Dev.*, **20**(7), 898–912. [20]

SHAVER, SCOTT, CASAS-MOLLANO, J ARMANDO, CERNY, RONALD L, & CERUTTI, HERIBERTO. 2010. Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. *Epigenetics*, **5**(4), 301–312. [5]

SHELDON, CANDICE C, FINNEGAN, E JEAN, PEACOCK, W JAMES, & DENNIS, ELIZABETH S. 2009. Mechanisms of gene repression by vernalization in Arabidopsis. *Plant J.*, **59**(3), 488–498. [14]

SIEVERS, FABIAN, WILM, ANDREAS, DINEEN, DAVID, GIBSON, TOBY J, KARPLUS, KEVIN, LI, WEIZHONG, LOPEZ, RODRIGO, MCWILLIAM, HAMISH, REMMERT, MICHAEL, SÖDING, JOHANNES, THOMPSON, JULIE D, & HIGGINS, DESMOND G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539. [18, 29, 32]

SIMON, JEFFREY A, & KINGSTON, ROBERT E. 2009. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.*, **10**(10), 697–708. [2]

SIMON, JEFFREY A, & KINGSTON, ROBERT E. 2013. Occupying chromatin: polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol. Cell*, **49**(5), 808–824. [2, 3, 6, 8, 9, 66, 77, 78, 79, 84]

SING, ANGELA, PANNELL, DYLAN, KARAISKAKIS, ANGELO, STURGEON, KENDRA, DJABALI, MALEK, ELLIS, JAMES, LIPSHITZ, HOWARD D, & CORDES, SABINE P.

2009. A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell*, **138**(5), 885–897. [6]

SONG, JI JOON, GARLICK, JOSEPH D, & KINGSTON, ROBERT E. 2008. Structural basis of histone H4 recognition by p55. *Genes Dev.*, **22**(10), 1313–1318. [7]

SPILLANE, CHARLES, SCHMID, KARL J, LAOUEILLÉ-DUPRAT, SYLVIA, PIEN, STÉPHANE, ESCOBAR-RESTREPO, JUAN-MIGUEL, BAROUX, CÉLIA, GAGLIARDINI, VALERIA, PAGE, DAMIAN R, WOLFE, KENNETH H, & GROSSNIKLAUS, UELI. 2007. Positive darwinian selection at the imprinted MEDEA locus in plants. *Nature*, **448**(7151), 349–352. [6, 12, 17, 28, 29, 70]

STEIN, LINCOLN D. 2013. Using GBrowse 2.0 to visualize and share next-generation sequence data. *Brief. Bioinformatics*, **14**(2), 162–171. [26, 51]

STURN, ALEXANDER, QUACKENBUSH, JOHN, & TRAJANOSKI, ZLATKO. 2002. Genesis: cluster analysis of microarray data. *Bioinformatics*, **18**(1), 207–208. [26, 27]

SUNG, SIBUM, & AMASINO, RICHARD M. 2004. Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3. *Nature*, **427**(6970), 159–164. [7]

SUNG, SIBUM, HE, YUEHUI, ESHOO, TIFANI W, TAMADA, YOSUKE, JOHNSON, LIANNA, NAKAHIGASHI, KENJI, GOTO, KOJI, JACOBSEN, STEVE E, & AMASINO, RICHARD M. 2006. Epigenetic maintenance of the vernalized state in Arabidopsis thaliana requires LIKE HETEROCHROMATIN PROTEIN 1. *Nat. Genet.*, **38**(6), 706–710. [15]

SWIEZEWSKI, SZYMON, LIU, FUQUAN, MAGUSIN, ANDREAS, & DEAN, CAROLINE. 2009. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature*, **462**(7274), 799–802. [8, 15, 81]

TAMURA, KOICHIRO, PETERSON, DANIEL, PETERSON, NICHOLAS, STECHER, GLEN, NEI, MASATOSHI, & KUMAR, SUDHIR. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, **28**(10), 2731–2739. [18]

TAVARES, LIGIA, DIMITROVA, EMILIA, OXLEY, DAVID, WEBSTER, JUDITH, POOT, RAYMOND, DEMMERS, JEROEN, BEZSTAROSTI, KAREL, TAYLOR, STEPHEN, URA, HIROKI, KOIDE, HIROSHI, WUTZ, ANTON, VIDAL, MIGUEL, ELDERKIN, SARAH, & BROCKDORFF, NEIL. 2012. RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell*, **148**(4), 664–678. [2, 7, 9, 78]

THORVALDSDÓTTIR, HELGA, ROBINSON, JAMES T, & MESIROV, JILL P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics*, **14**(2), 178–192. [26, 44]

TRÉMOUSAYGUE, DOMINIQUE, GARNIER, LIONEL, BARDET, CLAUDE, DABOS, PATRICK, HERVÉ, CHRISTINE, & LESCURE, BERNARD. 2003. Internal telomeric repeats and 'TCP domain' protein-binding sites co-operate to regulate gene expression in Arabidopsis thaliana cycling cells. *Plant J.*, **33**(6), 957–966. [82, 83]

TSAI, M C, MANOR, O, WAN, Y, MOSAMMAPARAST, N, WANG, J K, LAN, F, SHI, Y, SEGAL, E, & CHANG, H Y. 2010. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science*, **329**(5992), 689–693. [7, 81]

TURCK, FRANZISKA, ROUDIER, FRANÇOIS, FARRONA, SARA, MARTIN-MAGNIETTE, MARIE-LAURE, GUILLAUME, ELODIE, BUISINE, NICOLAS, GAGNOT, SÉVERINE, MARTIENSSEN, ROBERT A, COUPLAND, GEORGE, & COLOT, VINCENT. 2007. Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet.*, **3**(6), e86–866. [2, 11, 35, 73, 79]

VANDAMME, JULIEN, VÖLKEL, PAMELA, ROSNOBLET, CLAIRE, LE FAOU, PERRINE, & ANGRAND, PIERRE-OLIVIER. 2011. Interaction Proteomics Analysis of Polycomb Proteins Defines Distinct PRC1 Complexes in Mammalian Cells. *Mol. Cell. Proteomics*, **10**(4). [4]

WANG, DONGFANG, TYSON, MARK D, JACKSON, SHAWN S, & YADEGARI, RAMIN. 2006. Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in Arabidopsis. *Proc. Natl. Acad. Sci. USA*, **103**(35), 13244–13249. [12, 13]

WANG, JIE, WANG, YU, WANG, ZHUO, LIU, LEI, ZHU, XIN-GUANG, & MA, XIAOTU. 2011. Synchronization of cytoplasmic and transferred mitochondrial ribosomal protein gene expression in land plants is linked to Telo-box motif enrichment. *BMC Evol. Biol.*, **11**, 161. [82, 83]

WEINHOFER, ISABELLE, HEHENBERGER, ELISABETH, ROSZAK, PAWEL, HENNIG, LARS, & KÖHLER, CLAUDIA. 2010. H3K27me3 Profiling of the Endosperm Implies Exclusion of Polycomb Group Protein Targeting by DNA Methylation. *PLoS Genet.*, **6**(10), e1001152. [14, 79]

WINTER, DEBBIE, VINEGAR, BEN, NAHAL, HARDEEP, AMMAR, RON, WILSON, GREG V, & PROVART, NICHOLAS J. 2007. An Electronic Fluorescent Pictograph Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLoS ONE*, **2**(8), e718. [13]

WOO, CAROLINE J, KHARCHENKO, PETER V, DAHERON, LAURENCE, PARK, PETER J, & KINGSTON, ROBERT E. 2010. A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell*, **140**(1), 99–110. [6]

WOOD, CRAIG C, ROBERTSON, MASUMI, TANNER, GREG, PEACOCK, W JAMES, DENNIS, ELIZABETH S, & HELLIWELL, CHRIS A. 2006. The Arabidopsis thaliana vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc. Natl. Acad. Sci. USA*, **103**(39), 14631–14636. [2, 12, 14, 36, 73, 75, 85]

XI, HUALIN, YU, YONG, FU, YUTAO, FOLEY, JONATHAN, HALEES, ANASON, & WENG, ZHIPING. 2007. Analysis of overrepresented motifs in human core promoters reveals dual regulatory roles of YY1. *Genome Res.*, **17**(6), 798–806. [6]

XU, KEXIN, WU, ZHENHUA JEREMY, GRONER, ANNA C, HE, HOUSHENG HANSEN, CAI, CHANGMENG, LIS, ROSINA T, WU, XIAOQIU, STACK, EDWARD C, LODA, MASSIMO, LIU, TAO, XU, HAN, CATO, LAURA, THORNTON, JAMES E, GREGORY, RICHARD I, MORRISSEY, COLM, VESSELLA, ROBERT L, MONTIRONI, RODOLFO, MAGI-GALLUZZI, CRISTINA, KANTOFF, PHILIP W, BALK, STEVEN P, LIU, X SHIRLEY, & BROWN, MYLES. 2012. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science*, **338**(6113), 1465–1469. [2, 9, 10, 59, 80, 84, 85]

YAP, KYOKO L, LI, SIDE, MUÑOZ-CABELLO, ANA M, RAGUZ, SELINA, ZENG, LEI, MUJTABA, SHIRAZ, GIL, JESÚS, WALSH, MARTIN J, & ZHOU, MING-MING. 2010. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell*, **38**(5), 662–674. [7]

YOSHIDA, NOBUMASA, YANAI, YUKIHIRO, CHEN, LINGJING, KATO, YOSHIHIRO, HIRATSUKA, JUNZO, MIWA, TATSUSHI, SUNG, Z RENEE, & TAKAHASHI, SHIGERU. 2001. EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates shoot development and flowering in Arabidopsis. *Plant Cell*, **13**(11), 2471–2481. [3, 13]

YU, MING, MAZOR, TALI, HUANG, HUI, HUANG, HSUAN-TING, KATHREIN, KATIE L, WOO, ANDREW J, CHOUINARD, CANDACE R, LABADORF, ADAM, AKIE, THOMAS E, MORAN, TYLER B, XIE, HUAFENG, ZACHAREK, SIMA, TANIUCHI, ICHIRO, ROEDER, ROBERT G, KIM, CARLA F, ZON, LEONARD I, FRAENKEL, ERNEST, & CANTOR, ALAN B. 2012. Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. *Mol. Cell*, **45**(3), 330–343. [7]

YUAN, WEN, WU, TONG, FU, HANG, DAI, CHAO, WU, HUI, LIU, NAN, LI, XIANG, XU, MO, ZHANG, ZHUQIANG, NIU, TIANHUI, HAN, ZHIFU, CHAI, JIJIE, ZHOU, XIANGHONG JASMINE, GAO, SHAORONG, & ZHU, BING. 2012. Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science*, **337**(6097), 971–975. [9]

ZANG, CHONGZHI, SCHONES, DUSTIN E, ZENG, CHEN, CUI, KAIRONG, ZHAO, KEJI, & PENG, WEIQUN. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*, **25**(15), 1952–1958. [26, 27, 66, 68, 76, 77]

ZHANG, XIAOYU, GERMANN, SOPHIE, BLUS, BARTLOMIEJ J, KHORASANIZADEH, SEPIDEH, GAUDIN, VALERIE, & JACOBSEN, STEVEN E. 2007a. The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat. Struct. Mol. Biol.*, **14**(9), 869–871. [73]

ZHANG, XIAOYU, CLARENZ, OLIVER, COKUS, SHAWN, BERNATAVICHUTE, YANA V, PELLEGRINI, MATTEO, GOODRICH, JUSTIN, & JACOBSEN, STEVEN E. 2007b. Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *PLoS Biol.*, **5**(5), e129. [2, 11, 44, 62, 81]

ZHAO, JING, SUN, BRYAN K, ERWIN, JENNIFER A, SONG, JI JOON, & LEE, JEANNIE T. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, **322**(5902), 750–756. [81]

ZHAO, XIAOBEI, VALEN, EIVIND, PARKER, BRIAN J, & SANDELIN, ALBIN. 2011. Systematic Clustering of Transcription Start Site Landscapes. *PLoS ONE*, **6**(8). [60]

ZHAO, ZHONG, YU, YU, MEYER, DENISE, WU, CHENGJUN, & SHEN, WEN-HUI. 2005. Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. *Nat. Cell Biol.*, **7**(12), 1256–1260. [15]

Abbreviations

Table 5.1: **Abbreviations**

Abbreviation	**Full name**
Aebp2	AE binding protein 2
AG	AGAMOUS
AS1 and AS2	ASYMMETRIC LEAVES 1 and 2
AtBMI1a and AtBMI1b	Arabidopsis B lymphoma Mo-MLV insertion region 1 homolog
BP	BREVIPEDICELUS
CAF-1	Chromatin assembly factor 1
Cbx	Chromobox gene
ChIP	Chromatin immunoprecipitation
CLF	CURLY LEAF
co-IP	co-immunoprecipitation
CpG island	GC repeat rich region
E(z)	Enhancer of Zeste (Drosophila)
Eed	Embryonic ectorderm Development
EMF1 and EMF2	EMBRYONIC FLOWER 1 and 2
ESC	Extra Sex Combs
EZD1	(E(z)-domain 1)
Ezh1 and Ezh2	Enhancer of zeste (mammals)
FIE	FERTILIZATION INDEPENDENT ENDOSPERM
FIS2	FERTILIZATION-INDEPENDENT SEED 2
FLC	FLOWERING LOCUS C
FT	FLOWERING LOCUS T
GFP	Green fluorescent protein
H2A	Histone 2A
H2AK119ub	Monoubiquitylation of histone H2A at lysine 119
H3K27me3	Trimethylation of histone 3 at lysine 27
H3K36me2	Dimethylation of histone 3 at lysine 36
H3K36me3	Trimethylation of histone 3 at lysine 36
H3K4me3	Trimethylation of histone 3 at lysine 4
HMT	Histone-methyltransferase

Continued on next page

Table 5.1 – continued from previous page

Abbreviation	Full name
Jarid2	Jumonji AT rich interactive domain 2
KNAT2	KNOTTED LIIKE ARABIDOPSIS 2
KNOX	KNOTTED LIKE HOMEOBOX
LEC2	LEAFY COTYLEDON2
LHP1	LIKE HETEROCHROMATIN PROTEIN 1
lncRNA	long non coding RNA
MAF	MADS AFFECTING FLOWERING
MEA	MEDEA
MSI	MULTICOPY SUPPRESSOR OF IRA
N55	Nucleosome remodeling factor 55
ncRNA	non coding RNAv
Nurf	Nucleosome remodeling factor
PC	Polycomb
PcG	Polycomb Group
Pcl	Polycomb like
PH	Polyhomeotic
PHD	Plant Homeodomain
PHO	PRC2 recruiter Pleiohomeotic
Pho-RC	PHO repressive complex
PR-DUB	Polycomb repressive deubiquitinase (PR-DUB)
PRC	Polycomb Repressive Complex
PRE	Polycomb Response Elements
PSC	Posterior Sex Combs
PTM	Post translational modification
RbpA	Retinoblastoma binding protein
REF6	RELATIVE OF EARLY FLOWERING
RNAPII	RNA Polymerase II
RYBP	RING1 and YY1-binding -protein
SAM	sterile alpha motif
SCE	Sex Combs Extra
SET	Su(var)3-9, E(z) and Trx
Su(z)12	Suppressor of Zeste (Drosophila)
Suz12	Suppressor of zheste (mammals)
SWN	SWINGER
TrxG	Trithorax Group
VEFS	VRN2, EMF2, FIS2 and Su(z)12

Continued on next page

Table 5.1 – continued from previous page

Abbreviation	Full name
VIN3	VERNALIZATION INDEPENDENT 3
VRN2	REDUCED VERNALIZATION RESPONSE 2
WUS	WUSCHEL
Y2H	Yeast two hybrid
YY1	Ying yang 1

6 Appendix

Table 6.1: **CLF dependent genes**

Number	Atg name	Short description
1	AT1G12667	-
2	AT1G18230	-
3	AT1G23724	-
4	AT1G26762	-
5	AT1G28304	-
6	AT1G28307	-
7	AT1G28447	-
8	AT1G28630	-
9	AT1G31245	-
10	AT1G31983	-
11	AT1G35320	-
12	AT1G44382	-
13	AT1G47389	-
14	AT1G47395	-
15	AT1G47400	-
16	AT1G47410	-
17	AT1G47590	-
18	AT1G47595	-
19	AT1G49110	-
20	AT1G49310	-
21	AT1G51000	-
22	AT1G51010	-
23	AT1G51020	-
24	AT1G51030	-
25	AT1G51035	-
26	AT1G51913	-
27	AT1G51920	-
28	AT1G53970	-
29	AT1G57835	-
30	AT1G58007	-

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
31	AT1G58010	-
32	AT1G59535	-
33	AT1G59722	-
34	AT1G65481	-
35	AT1G65483	-
36	AT1G66475	-
37	AT1G66790	-
38	AT2G01990	-
39	AT2G02330	-
40	AT2G04495	-
41	AT2G04795	-
42	AT2G04800	-
43	AT2G14700	-
44	AT2G17064	-
45	AT2G18690	-
46	AT2G20070	-
47	AT2G20080	-
48	AT2G24460	-
49	AT2G24748	-
50	AT2G24945	-
51	AT3G09280	-
52	AT3G09285	-
53	AT3G21985	-
54	AT3G22022	-
55	AT3G22053	-
56	AT3G22057	-
57	AT3G43900	-
58	AT3G46360	-
59	AT3G52700	-
60	AT3G58347	-
61	AT4G01516	-
62	AT4G03566	-
63	AT4G09850	-
64	AT4G09860	-
65	AT4G10596	-
66	AT4G11211	-

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
67	AT4G12930	-
68	AT4G12940	-
69	AT4G13572	-
70	AT4G13575	-
71	AT4G14723	-
72	AT4G15990	-
73	AT4G16008	-
74	AT4G18850	-
75	AT4G18860	-
76	AT4G19970	-
77	AT4G19980	-
78	AT4G20190	-
79	AT4G22160	-
80	AT4G23090	-
81	AT4G23370	-
82	AT4G30970	-
83	AT5G13825	-
84	AT5G19710	-
85	AT5G19729	-
86	AT5G22970	-
87	AT5G25422	-
88	AT5G26100	-
89	AT5G26114	-
90	AT5G37730	-
91	AT5G37732	-
92	AT5G38080	-
93	AT5G38096	-
94	AT5G38393	-
95	AT5G43196	-
96	AT5G46300	-
97	AT5G49260	-
98	AT5G50190	-
99	AT5G52130	-
100	AT5G52145	-
101	AT5G54145	-
102	AT5G54206	-

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
103	AT5G56240	-
104	AT5G56368	-
105	AT5G57340	-
106	AT1G52800	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
107	AT4G27420	ABC-2 type transporter family protein
108	AT5G25320	ACT-like superfamily protein
109	AT4G12440	adenine phosphoribosyl transferase 4
110	AT2G28700	AGAMOUS-like 46
111	AT1G28450	AGAMOUS-like 58
112	AT2G45650	AGAMOUS-like 6
113	AT3G30260	AGAMOUS-like 79
114	AT1G46408	AGAMOUS-like 97
115	AT3G25760	allene oxide cyclase 1
116	AT3G25770	allene oxide cyclase 2
117	AT3G25780	allene oxide cyclase 3
118	AT4G20990	alpha carbonic anhydrase 4
119	AT4G21000	alpha carbonic anhydrase 6
120	AT1G51300	alpha/beta-Hydrolases superfamily protein
121	AT2G22960	alpha/beta-Hydrolases superfamily protein
122	AT4G15960	alpha/beta-Hydrolases superfamily protein
123	AT5G22960	alpha/beta-Hydrolases superfamily protein
124	AT5G24200	alpha/beta-Hydrolases superfamily protein
125	AT5G46600	Aluminium activated malate transporter family protein
126	AT5G46610	Aluminium activated malate transporter family protein
127	AT2G47240	AMP-dependent synthetase and ligase family protein
128	AT1G28300	AP2/B3-like transcriptional factor family protein
129	AT4G08870	Arginase/deacetylase superfamily protein
130	AT2G16960	ARM repeat superfamily protein
131	AT5G43120	ARM-repeat/Tetratricopeptide repeat (TPR)-like protein
132	AT5G51590	AT hook motif DNA-binding family protein
133	AT5G15100	Auxin efflux carrier family protein
134	AT4G12550	Auxin-Induced in Root cultures 1

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
135	AT1G33890	Avirulence induced gene (AIG1) family protein
136	AT3G10960	AZA-guanine resistant1
137	AT4G15370	baruol synthase 1
138	AT1G12540	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
139	AT1G25330	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
140	AT2G22750	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
141	AT2G31215	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
142	AT5G43650	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
143	AT5G65320	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
144	AT1G59530	basic leucine-zipper 4
145	AT5G28510	beta glucosidase 24
146	AT2G44470	beta glucosidase 29
147	AT5G49250	Beta-galactosidase related protein
148	AT1G61810	beta-glucosidase 45
149	AT4G12360	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
150	AT4G12520	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
151	AT4G12530	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
152	AT4G12545	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
153	AT4G12825	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
154	AT5G07230	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
155	AT5G52160	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
156	AT5G56480	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
157	AT1G13400	C2H2 and C2HC zinc fingers superfamily protein
158	AT5G22990	C2H2-like zinc finger protein
159	AT2G26940	C2H2-type zinc finger family protein
160	AT4G01420	calcineurin B-like protein 5
161	AT1G54530	Calcium-binding EF hand family protein
162	AT1G08860	Calcium-dependent phospholipid-binding Copine family protein
163	AT1G66855	Carbohydrate-binding X8 domain superfamily protein
164	AT1G66870	Carbohydrate-binding X8 domain superfamily protein
165	AT2G30933	Carbohydrate-binding X8 domain superfamily protein
166	AT4G15290	Cellulose synthase family protein
167	AT2G32530	cellulose synthase-like B3
168	AT4G15320	cellulose synthase-like B6
169	AT1G09260	Chaperone DnaJ-domain superfamily protein
170	AT2G21890	cinnamyl alcohol dehydrogenase homolog 3
171	AT1G66860	Class I glutamine amidotransferase-like superfamily protein
172	AT5G49270	COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family
173	AT3G56240	copper chaperone
174	AT5G52670	Copper transport protein family
175	AT5G52680	Copper transport protein family
176	AT5G52690	Copper transport protein family
177	AT5G52760	Copper transport protein family
178	AT5G25970	Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein
179	AT1G51915	cryptdin protein-related
180	AT1G59800	Cullin family protein
181	AT5G26330	Cupredoxin superfamily protein
182	AT4G03565	Cystatin/monellin superfamily protein
183	AT5G05040	Cystatin/monellin superfamily protein
184	AT5G05060	Cystatin/monellin superfamily protein
185	AT3G44780	Cysteine proteinases superfamily protein
186	AT5G05050	Cysteine proteinases superfamily protein
187	AT1G66440	Cysteine/Histidine-rich C1 domain family protein
188	AT5G44770	Cysteine/Histidine-rich C1 domain family protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
189	AT5G54030	Cysteine/Histidine-rich C1 domain family protein
190	AT5G54215	Cysteine-rich protein
191	AT1G65670	cytochrome P450, family 702, subfamily A, polypeptide 1
192	AT4G15300	cytochrome P450, family 702, subfamily A, polypeptide 2
193	AT4G15310	cytochrome P450, family 702, subfamily A, polypeptide 3
194	AT1G28430	cytochrome P450, family 705, subfamily A, polypeptide 24
195	AT5G48000	cytochrome P450, family 708, subfamily A, polypeptide 2
196	AT4G13290	cytochrome P450, family 71, subfamily A, polypeptide 19
197	AT4G13310	cytochrome P450, family 71, subfamily A, polypeptide 20
198	AT4G20240	cytochrome P450, family 71, subfamily A, polypeptide 27
199	AT4G20235	cytochrome P450, family 71, subfamily A, polypeptide 28
200	AT5G23190	cytochrome P450, family 86, subfamily B, polypeptide 1
201	AT1G12740	cytochrome P450, family 87, subfamily A, polypeptide 2
202	AT5G56490	D-arabinono-1,4-lactone oxidase family protein
203	AT1G31772	Defensin-like (DEFL) family protein
204	AT2G22805	Defensin-like (DEFL) family protein
205	AT2G36255	Defensin-like (DEFL) family protein
206	AT5G46873	Defensin-like (DEFL) family protein
207	AT5G46877	Defensin-like (DEFL) family protein
208	AT5G51845	Defensin-like (DEFL) family protein
209	AT5G56369	Defensin-like (DEFL) family protein
210	AT5G05070	DHHC-type zinc finger family protein
211	AT3G21730	Dihydroneopterin aldolase
212	AT1G58400	Disease resistance protein (CC-NBS-LRR class) family
213	AT5G38350	Disease resistance protein (NBS-LRR class) family

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
214	AT5G46260	disease resistance protein (TIR-NBS-LRR class) family
215	AT4G11180	Disease resistance-responsive (dirigent-like protein) family protein
216	AT3G12710	DNA glycosylase superfamily protein
217	AT1G28310	Dof-type zinc finger DNA-binding family protein
218	AT2G28510	Dof-type zinc finger DNA-binding family protein
219	AT4G21030	Dof-type zinc finger domain-containing protein
220	AT4G21040	Dof-type zinc finger domain-containing protein
221	AT3G21970	Domain of unknown function (DUF26)
222	AT3G21980	Domain of unknown function (DUF26)
223	AT3G21990	Domain of unknown function (DUF26)
224	AT3G22000	Domain of unknown function (DUF26)
225	AT3G22050	Domain of unknown function (DUF26)
226	AT1G32030	Domain of unknown function (DUF313)
227	AT4G09450	Duplicated homeodomain-like superfamily protein
228	AT4G18870	E2F/DP family winged-helix DNA-binding domain
229	AT1G54720	early-responsive to dehydration protein-related / ERD protein-related
230	AT3G45245	ECA1 gametogenesis related family protein
231	AT3G01323	ECA1-like gametogenesis related family protein
232	AT3G01324	ECA1-like gametogenesis related family protein
233	AT4G13235	embryo sac development arrest 21
234	AT3G29070	emp24/gp25L/p24 family/GOLD family protein
235	AT4G05110	equilibrative nucleoside transporter 6
236	AT1G61630	equilibrative nucleoside transporter 7
237	AT3G02970	EXORDIUM like 6
238	AT3G49307	Expressed protein
239	AT5G44400	FAD-binding Berberine family protein
240	AT5G56470	FAD-dependent oxidoreductase family protein
241	AT2G29430	Family of unknown function (DUF572)
242	AT4G09440	Family of unknown function (DUF577)
243	AT4G12950	Fasciclin-like arabinogalactan family protein
244	AT3G44540	fatty acid reductase 4
245	AT5G22420	fatty acid reductase 7
246	AT1G55660	FBD, F-box and Leucine Rich Repeat domains containing protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
247	AT3G49030	FBD, F-box and Leucine Rich Repeat domains containing protein
248	AT3G56780	FBD, F-box and Leucine Rich Repeat domains containing protein
249	AT5G56560	FBD, F-box and Leucine Rich Repeat domains containing protein
250	AT1G13200	F-box and associated interaction domains-containing protein
251	AT1G47390	F-box and associated interaction domains-containing protein
252	AT1G51290	F-box and associated interaction domains-containing protein
253	AT1G51320	F-box and associated interaction domains-containing protein
254	AT1G54550	F-box and associated interaction domains-containing protein
255	AT2G16220	F-box and associated interaction domains-containing protein
256	AT3G22700	F-box and associated interaction domains-containing protein
257	AT5G51000	F-box and associated interaction domains-containing protein
258	AT1G46912	F-box associated ubiquitination effector family protein
259	AT1G59680	F-box associated ubiquitination effector family protein
260	AT1G59690	F-box associated ubiquitination effector family protein
261	AT3G22770	F-box associated ubiquitination effector family protein
262	AT3G52510	F-box associated ubiquitination effector family protein
263	AT5G48550	F-box associated ubiquitination effector family protein
264	AT1G30920	F-box family protein
265	AT1G32020	F-box family protein
266	AT1G46840	F-box family protein
267	AT1G59675	F-box family protein
268	AT1G60410	F-box family protein
269	AT3G22710	F-box family protein
270	AT3G55900	F-box family protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
271	AT4G12820	F-box family protein with a domain of unknown function (DUF295)
272	AT4G22165	F-box family protein with a domain of unknown function (DUF295)
273	AT1G58310	F-box/RNI-like superfamily protein
274	AT1G60400	F-box/RNI-like superfamily protein
275	AT5G41630	F-box/RNI-like superfamily protein
276	AT3G52680	F-box/RNI-like/FBD-like domains-containing protein
277	AT5G56370	F-box/RNI-like/FBD-like domains-containing protein
278	AT1G65860	flavin-monooxygenase glucosinolate S-oxygenase 1
279	AT3G20980	Gag-Pol-related retrotransposon family protein
280	AT3G21000	Gag-Pol-related retrotransposon family protein
281	AT1G60450	galactinol synthase 7
282	AT5G15140	Galactose mutarotase-like superfamily protein
283	AT3G27150	Galactose oxidase/kelch repeat superfamily protein
284	AT4G12890	Gamma interferon responsive lysosomal thiol (GILT) reductase family protein
285	AT1G54000	GDSL-like Lipase/Acylhydrolase superfamily protein
286	AT1G54010	GDSL-like Lipase/Acylhydrolase superfamily protein
287	AT1G54020	GDSL-like Lipase/Acylhydrolase superfamily protein
288	AT2G19010	GDSL-like Lipase/Acylhydrolase superfamily protein
289	AT2G19050	GDSL-like Lipase/Acylhydrolase superfamily protein
290	AT3G27950	GDSL-like Lipase/Acylhydrolase superfamily protein
291	AT4G18970	GDSL-like Lipase/Acylhydrolase superfamily protein
292	AT5G03610	GDSL-like Lipase/Acylhydrolase superfamily protein
293	AT5G45950	GDSL-like Lipase/Acylhydrolase superfamily protein
294	AT5G45960	GDSL-like Lipase/Acylhydrolase superfamily protein
295	AT1G53940	GDSL-motif lipase 2
296	AT1G53990	GDSL-motif lipase 3
297	AT4G14630	germin-like protein 9
298	AT1G50960	gibberellin 2-oxidase 7
299	AT2G29120	glutamate receptor 2.7
300	AT4G23100	glutamate-cysteine ligase
301	AT1G02940	glutathione S-transferase (class phi) 5
302	AT2G29490	glutathione S-transferase TAU 1
303	AT1G59670	glutathione S-transferase TAU 15

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
304	AT1G59700	glutathione S-transferase TAU 16
305	AT2G29480	glutathione S-transferase tau 2
306	AT2G29440	glutathione S-transferase tau 6
307	AT5G17650	glycine/proline-rich protein
308	AT5G25425	glycine-rich protein
309	AT5G42635	glycine-rich protein
310	AT4G11050	glycosyl hydrolase 9C3
311	AT3G47050	Glycosyl hydrolase family protein
312	AT1G31140	GORDITA
313	AT4G22110	GroES-like zinc-binding dehydrogenase family protein
314	AT4G39770	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
315	AT5G54210	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
316	AT5G56250	hapless 8
317	AT3G22830	heat shock transcription factor A6B
318	AT3G02960	Heavy metal transport/detoxification superfamily protein
319	AT1G58300	heme oxygenase 4
320	AT2G22800	Homeobox-leucine zipper protein family
321	AT5G42630	Homeodomain-like superfamily protein
322	AT5G62110	Homeodomain-like superfamily protein
323	AT3G23100	homolog of human DNA ligase iv-binding protein XRCC4
324	AT3G23190	HR-like lesion-inducing protein-related
325	AT2G42820	HVA22-like protein F
326	AT5G26070	hydroxyproline-rich glycoprotein family protein
327	AT2G36260	Iron-sulphur cluster biosynthesis family protein
328	AT1G33790	jacalin lectin family protein
329	AT1G48500	jasmonate-zim-domain protein 4
330	AT1G30135	jasmonate-zim-domain protein 8
331	AT4G18960	K-box region and MADS-box transcription factor family protein
332	AT5G10140	K-box region and MADS-box transcription factor family protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
333	AT1G54540	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
334	AT5G53730	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
335	AT4G13560	Late embryogenesis abundant protein (LEA) family protein
336	AT5G22590	Leucine Rich Repeat protein family
337	AT1G49100	Leucine-rich repeat protein kinase family protein
338	AT1G51810	Leucine-rich repeat protein kinase family protein
339	AT1G51820	Leucine-rich repeat protein kinase family protein
340	AT1G51830	Leucine-rich repeat protein kinase family protein
341	AT1G51910	Leucine-rich repeat protein kinase family protein
342	AT2G28990	Leucine-rich repeat protein kinase family protein
343	AT2G29000	Leucine-rich repeat protein kinase family protein
344	AT3G46350	Leucine-rich repeat protein kinase family protein
345	AT5G49770	Leucine-rich repeat protein kinase family protein
346	AT5G49780	Leucine-rich repeat protein kinase family protein
347	AT3G24240	Leucine-rich repeat receptor-like protein kinase family protein
348	AT1G53420	Leucine-rich repeat transmembrane protein kinase
349	AT5G24180	Lipase class 3-related protein
350	AT5G24190	Lipase class 3-related protein
351	AT4G10595	low-molecular-weight cysteine-rich 2
352	AT4G09984	low-molecular-weight cysteine-rich 34
353	AT3G20997	low-molecular-weight cysteine-rich 55
354	AT3G20993	low-molecular-weight cysteine-rich 56
355	AT5G54225	low-molecular-weight cysteine-rich 83
356	AT1G51260	lysophosphatidyl acyltransferase 3
357	AT1G52190	Major facilitator superfamily protein
358	AT1G69860	Major facilitator superfamily protein
359	AT2G16970	Major facilitator superfamily protein
360	AT2G16980	Major facilitator superfamily protein
361	AT2G16990	Major facilitator superfamily protein
362	AT3G05155	Major facilitator superfamily protein
363	AT3G25280	Major facilitator superfamily protein
364	AT1G52050	Mannose-binding lectin superfamily protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
365	AT1G52060	Mannose-binding lectin superfamily protein
366	AT1G52070	Mannose-binding lectin superfamily protein
367	AT1G52100	Mannose-binding lectin superfamily protein
368	AT5G42600	marneral synthase
369	AT1G66780	MATE efflux family protein
370	AT1G10747	Maternally expressed gene (MEG) family protein
371	AT2G16225	Maternally expressed gene (MEG) family protein
372	AT3G02980	MEIOTIC CONTROL OF CROSSOVERS1
373	AT2G24940	membrane-associated progesterone binding protein 2
374	AT4G30972	MIR156B; miRNA
375	AT1G66783	MIR157A; miRNA
376	AT1G66795	MIR157B; miRNA
377	AT3G55734	MIR393B; miRNA
378	AT1G69792	MIR395D; miRNA
379	AT1G69795	MIR395E; miRNA
380	AT4G05105	MIR397A; miRNA
381	AT4G23387	MIR845a; miRNA
382	AT4G23375	MIR845b; miRNA
383	AT4G13493	MIR850a; miRNA
384	AT4G13494	MIR863a; miRNA
385	AT1G56380	Mitochondrial transcription termination factor family protein
386	AT1G24020	MLP-like protein 423
387	AT2G31180	myb domain protein 14
388	AT5G40430	myb domain protein 22
389	AT4G13480	myb domain protein 79
390	AT5G39700	myb domain protein 89
391	AT3G47600	myb domain protein 94
392	AT5G62470	myb domain protein 96
393	AT1G54560	Myosin family protein with Dil domain
394	AT1G66800	NAD(P)-binding Rossmann-fold superfamily protein
395	AT1G12110	nitrate transporter 1.1
396	AT5G37810	NOD26-like intrinsic protein 4;1
397	AT5G37820	NOD26-like intrinsic protein 4;2
398	AT1G01070	nodulin MtN21 /EamA-like transporter family protein
399	AT1G60050	Nodulin MtN21 /EamA-like transporter family protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
400	AT4G01440	nodulin MtN21 /EamA-like transporter family protein
401	AT4G01450	nodulin MtN21 /EamA-like transporter family protein
402	AT4G08300	nodulin MtN21 /EamA-like transporter family protein
403	AT4G30420	nodulin MtN21 /EamA-like transporter family protein
404	AT1G66770	Nodulin MtN3 family protein
405	AT2G47810	nuclear factor Y, subunit B5
406	AT4G15970	Nucleotide-diphospho-sugar transferase family protein
407	AT3G45760	Nucleotidyltransferase family protein
408	AT2G38995	O-acyltransferase (WSD1-like) family protein
409	AT1G21110	O-methyltransferase family protein
410	AT1G21120	O-methyltransferase family protein
411	AT4G22540	OSBP(oxysterol binding protein)-related protein 2A
412	AT1G32172	other RNA
413	AT1G56242	other RNA
414	AT2G22955	other RNA
415	AT3G20978	other RNA
416	AT4G13495	other RNA
417	AT4G26255	other RNA
418	AT4G30975	other RNA
419	AT5G24205	other RNA
420	AT5G24206	other RNA
421	AT5G38005	other RNA
422	AT2G23830	PapD-like superfamily protein
423	AT4G37050	PATATIN-like protein 4
424	AT5G15110	Pectate lyase family protein
425	AT5G19730	Pectin lyase-like superfamily protein
426	AT5G27530	Pectin lyase-like superfamily protein
427	AT5G06720	peroxidase 2
428	AT2G18140	Peroxidase superfamily protein
429	AT3G21770	Peroxidase superfamily protein
430	AT5G06730	Peroxidase superfamily protein
431	AT5G52120	phloem protein 2-A14
432	AT1G56240	phloem protein 2-B13
433	AT2G32960	Phosphotyrosine protein phosphatases superfamily protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
434	AT3G51690	PIF1 helicase
435	AT3G51700	PIF1 helicase
436	AT5G50040	Plant invertase/pectin methylesterase inhibitor superfamily protein
437	AT1G65985	Plant protein of unknown function (DUF247)
438	AT1G62530	Plant protein of unknown function (DUF863)
439	AT2G24470	Plant protein of unknown function (DUF869)
440	AT1G26799	Plant self-incompatibility protein S1 family
441	AT1G28305	Plant self-incompatibility protein S1 family
442	AT1G28306	Plant self-incompatibility protein S1 family
443	AT1G51240	Plant self-incompatibility protein S1 family
444	AT1G51250	Plant self-incompatibility protein S1 family
445	AT5G04347	Plant self-incompatibility protein S1 family
446	AT5G04350	Plant self-incompatibility protein S1 family
447	AT5G26050	Plant self-incompatibility protein S1 family
448	AT5G26060	Plant self-incompatibility protein S1 family
449	AT5G26090	Plant self-incompatibility protein S1 family
450	AT1G66100	Plant thionin
451	AT3G16340	pleiotropic drug resistance 1
452	AT1G33870	P-loop containing nucleoside triphosphate hydrolases superfamily protein
453	AT5G41310	P-loop nucleoside triphosphate hydrolases superfamily protein with CH (Calponin Homology) domain
454	AT5G22430	Pollen Ole e 1 allergen and extensin family protein
455	AT1G24000	Polyketide cyclase/dehydrase and lipid transport superfamily protein
456	AT1G24010	Polyketide cyclase/dehydrase and lipid transport superfamily protein
457	AT4G01455	pre-tRNA
458	AT1G23720	Proline-rich extensin-like family protein
459	AT1G31250	proline-rich family protein
460	AT5G26080	proline-rich family protein
461	AT1G51870	protein kinase family protein
462	AT1G51040	Protein kinase superfamily protein
463	AT4G08800	Protein kinase superfamily protein
464	AT5G57670	Protein kinase superfamily protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
465	AT1G51840	protein kinase-related
466	AT5G39430	Protein of unknown function (DUF1336)
467	AT4G17860	Protein of Unknown Function (DUF239)
468	AT4G23080	Protein of Unknown Function (DUF239)
469	AT4G23380	Protein of Unknown Function (DUF239)
470	AT4G23390	Protein of Unknown Function (DUF239)
471	AT5G25415	Protein of Unknown Function (DUF239)
472	AT5G25960	Protein of Unknown Function (DUF239)
473	AT5G46810	Protein of Unknown Function (DUF239)
474	AT5G46820	Protein of Unknown Function (DUF239)
475	AT5G50150	Protein of Unknown Function (DUF239)
476	AT1G65740	Protein of unknown function (DUF295)
477	AT2G04810	Protein of unknown function (DUF295)
478	AT5G55890	Protein of unknown function (DUF295)
479	AT4G07940	Protein of unknown function (DUF3245)
480	AT4G09890	Protein of unknown function (DUF3511)
481	AT1G33840	Protein of unknown function (DUF567)
482	AT1G63410	Protein of unknown function (DUF567)
483	AT5G42610	Protein of unknown function (DUF607)
484	AT3G44770	Protein of unknown function (DUF626)
485	AT5G41640	Protein of unknown function (DUF626)
486	AT5G54480	Protein of unknown function (DUF630 and DUF632)
487	AT5G42690	Protein of unknown function, DUF547
488	AT4G13965	Protein with RNI-like/FBD-like domains
489	AT5G49240	pseudo-response regulator 4
490	AT5G43110	pumilio 14
491	AT1G18220	purine permease 9
492	AT1G56360	purple acid phosphatase 6
493	AT5G06300	Putative lysine decarboxylase family protein
494	AT1G65352	Putative membrane lipoprotein
495	AT5G46871	Putative membrane lipoprotein
496	AT5G46874	Putative membrane lipoprotein
497	AT5G36150	putative pentacyclic triterpene synthase 3
498	AT2G21880	RAB GTPase homolog 7A
499	AT5G39620	RAB GTPase homolog G1
500	AT1G75250	RAD-like 6

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
501	AT2G19020	ralf-like 10
502	AT2G19030	ralf-like 11
503	AT2G19040	RALF-like 12
504	AT2G19045	RALF-like 13
505	AT2G32835	RALF-like 16
506	AT1G65790	receptor kinase 1
507	AT3G23110	receptor like protein 37
508	AT3G22030	Receptor protein kinase-related
509	AT3G22010	Receptor-like protein kinase-related family protein
510	AT3G22020	Receptor-like protein kinase-related family protein
511	AT3G22060	Receptor-like protein kinase-related family protein
512	AT1G51900	Regulator of Vps4 activity in the MVB pathway protein
513	AT2G27070	response regulator 13
514	AT5G62120	response regulator 23
515	AT3G58350	RESTRICTED TEV MOVEMENT 3
516	AT4G20220	Reverse transcriptase (RNA-dependent DNA polymerase)
517	AT2G22620	Rhamnogalacturonate lyase family protein
518	AT5G37800	RHD SIX-LIKE 1
519	AT5G19720	Ribosomal protein L25/Gln-tRNA synthetase, anti-codon-binding domain
520	AT1G33850	Ribosomal protein S19 family protein
521	AT5G43640	Ribosomal protein S19 family protein
522	AT3G45555	RING/U-box protein
523	AT3G45480	RING/U-box protein with C6HC-type zinc finger
524	AT1G18780	RING/U-box superfamily protein
525	AT1G51930	RING/U-box superfamily protein
526	AT1G77830	RING/U-box superfamily protein
527	AT2G25360	RING/U-box superfamily protein
528	AT2G34000	RING/U-box superfamily protein
529	AT2G46493	RING/U-box superfamily protein
530	AT2G46494	RING/U-box superfamily protein
531	AT2G46495	RING/U-box superfamily protein
532	AT4G13490	RING/U-box superfamily protein
533	AT5G07225	RING/U-box superfamily protein
534	AT5G52140	RING/U-box superfamily protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
535	AT5G52150	RING/U-box superfamily protein
536	AT5G38960	RmlC-like cupins superfamily protein
537	AT1G61700	RNA polymerases N / 8 kDa subunit
538	AT3G52690	RNI-like superfamily protein
539	AT3G59230	RNI-like superfamily protein
540	AT3G59240	RNI-like superfamily protein
541	AT1G66470	ROOT HAIR DEFECTIVE6
542	AT4G04900	ROP-interactive CRIB motif-containing protein 10
543	AT2G29125	ROTUNDIFOLIA like 2
544	AT1G18790	RWP-RK domain-containing protein
545	AT2G32160	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
546	AT5G37990	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
547	AT5G38020	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
548	AT5G38780	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
549	AT4G34770	SAUR-like auxin-responsive protein family
550	AT4G34780	SAUR-like auxin-responsive protein family
551	AT4G34790	SAUR-like auxin-responsive protein family
552	AT4G34800	SAUR-like auxin-responsive protein family
553	AT4G15735	SCR-like 10
554	AT4G15733	SCR-like 11
555	AT4G22115	SCR-like 14
556	AT4G22105	SCR-like 26
557	AT1G60986	SCR-like 4
558	AT1G60985	SCR-like 6
559	AT1G60983	SCR-like 8
560	AT4G34580	Sec14p-like phosphatidylinositol transfer family protein
561	AT2G22920	serine carboxypeptidase-like 12
562	AT2G22980	serine carboxypeptidase-like 13
563	AT2G24000	serine carboxypeptidase-like 22
564	AT5G22980	serine carboxypeptidase-like 47
565	AT1G51330	Serine protease inhibitor (SERPIN) family protein

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
566	AT2G39200	Seven transmembrane MLO family protein
567	AT2G19060	SGNH hydrolase-type esterase superfamily protein
568	AT2G19070	spermidine hydroxycinnamoyl transferase
569	AT5G22580	Stress responsive A/B Barrel Domain
570	AT1G51270	structural molecules;transmembrane receptors;structural molecules
571	AT1G66220	Subtilase family protein
572	AT1G66210	Subtilisin-like serine endopeptidase family protein
573	AT5G37180	sucrose synthase 5
574	AT1G22150	sulfate transporter 1;3
575	AT4G20200	Terpenoid cyclases/Protein prenyltransferases superfamily protein
576	AT4G20210	Terpenoid cyclases/Protein prenyltransferases superfamily protein
577	AT4G20230	Terpenoid cyclases/Protein prenyltransferases superfamily protein
578	AT4G13280	terpenoid synthase 12
579	AT4G13300	terpenoid synthase 13
580	AT1G52191	Thioesterase superfamily protein
581	AT1G52990	thioredoxin family protein
582	AT1G59730	thioredoxin H-type 7
583	AT1G47370	Toll-Interleukin-Resistance (TIR) domain family protein
584	AT1G57830	Toll-Interleukin-Resistance (TIR) domain family protein
585	AT1G69650	TRAF-like family protein
586	AT1G69660	TRAF-like family protein
587	AT2G05410	TRAF-like family protein
588	AT2G05420	TRAF-like family protein
589	AT3G58340	TRAF-like family protein
590	AT3G58360	TRAF-like family protein
591	AT3G58370	TRAF-like family protein
592	AT5G26320	TRAF-like family protein
593	AT2G32820	Transcription elongation factor (TFIIS) family protein
594	AT4G19000	Transcription elongation factor (TFIIS) family protein
595	AT1G30340	transposable element gene

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
596	AT1G47405	transposable element gene
597	AT1G54955	transposable element gene
598	AT1G65750	transposable element gene
599	AT2G05435	transposable element gene
600	AT2G14430	transposable element gene
601	AT2G28980	transposable element gene
602	AT2G40680	transposable element gene
603	AT3G20975	transposable element gene
604	AT3G20990	transposable element gene
605	AT3G21010	transposable element gene
606	AT3G21020	transposable element gene
607	AT3G21030	transposable element gene
608	AT3G21040	transposable element gene
609	AT4G01515	transposable element gene
610	AT4G21363	transposable element gene
611	AT5G25955	transposable element gene
612	AT5G36293	transposable element gene
613	AT5G38230	transposable element gene
614	AT5G43105	transposable element gene
615	AT5G46875	transposable element gene
616	AT1G53020	UBC26 (UBIQUITIN-CONJUGATING ENZYME 26); ubiquitin-protein ligase
617	AT1G53023	Ubiquitin-conjugating enzyme family protein
618	AT1G53950	Ubiquitin-like superfamily protein
619	AT1G53980	Ubiquitin-like superfamily protein
620	AT2G05400	Ubiquitin-specific protease family C19-related protein
621	AT2G05430	Ubiquitin-specific protease family C19-related protein
622	AT3G55710	UDP-Glycosyltransferase superfamily protein
623	AT5G38010	UDP-Glycosyltransferase superfamily protein
624	AT2G27370	Uncharacterised protein family (UPF0497)
625	AT4G15610	Uncharacterised protein family (UPF0497)
626	AT5G44550	Uncharacterised protein family (UPF0497)
627	AT5G39630	Vesicle transport v-SNARE family protein
628	AT3G48260	with no lysine (K) kinase 3
629	AT5G52830	WRKY DNA-binding protein 27
630	AT5G22570	WRKY DNA-binding protein 38

Continued on next page

Table 6.1 – continued from previous page

Number	Atg name	Short description
631	AT5G46310	WRKY family transcription factor
632	AT5G25420	Xanthine/uracil/vitamin C permease
633	AT4G13080	xyloglucan endotransglucosylase/hydrolase 1
634	AT4G18990	xyloglucan endotransglucosylase/hydrolase 29
635	AT1G65730	YELLOW STRIPE like 7
636	AT3G55890	Yippee family putative zinc-binding protein
637	AT1G57820	Zinc finger (C3HC4-type RING finger) family protein
638	AT5G43200	Zinc finger, C3HC4 type (RING finger) family protein
639	AT5G37940	Zinc-binding dehydrogenase family protein
640	AT5G37980	Zinc-binding dehydrogenase family protein
641	AT5G38000	Zinc-binding dehydrogenase family protein
642	AT3G10950	Zinc-binding ribosomal protein family protein
643	AT1G67270	Zinc-finger domain of monoamine-oxidase A repressor R1 protein

Table 6.2: **SWN affected genes**

Number	Atg name	Short description
1	AT1G01073	-
2	AT1G03820	-
3	AT1G12665	-
4	AT1G13605	-
5	AT1G21395	-
6	AT1G33785	-
7	AT1G33820	-
8	AT1G33860	-
9	AT1G33925	-
10	AT1G35183	-
11	AT1G46336	-
12	AT1G47317	-
13	AT1G47705	-
14	AT1G53625	-
15	AT1G53633	-
16	AT1G56480	-
17	AT1G60783	-
18	AT1G61830	-

Continued on next page

Table 6.2 – continued from previous page

Number	Atg name	Short description
19	AT1G64370	-
20	AT1G65510	-
21	AT1G66010	-
22	AT1G66235	-
23	AT1G75480	-
24	AT1G78470	-
25	AT1G78476	-
26	AT2G02440	-
27	AT2G04480	-
28	AT2G14206	-
29	AT2G14692	-
30	AT2G15029	-
31	AT2G15840	-
32	AT2G16955	-
33	AT2G19000	-
34	AT2G19146	-
35	AT2G23834	-
36	AT2G31700	-
37	AT2G40745	-
38	AT3G13950	-
39	AT3G30110	-
40	AT3G30281	-
41	AT3G43410	-
42	AT3G43420	-
43	AT3G44783	-
44	AT3G44980	-
45	AT3G47510	-
46	AT3G49305	-
47	AT3G62336	-
48	AT3G62350	-
49	AT4G09987	-
50	AT4G10201	-
51	AT4G13285	-
52	AT4G27415	-
53	AT5G05025	-
54	AT5G09976	-

Continued on next page

Table 6.2 – continued from previous page

Number	Atg name	Short description
55	AT5G24557	-
56	AT5G25870	-
57	AT5G38090	-
58	AT5G41420	-
59	AT5G41660	-
60	AT5G42146	-
61	AT5G42591	-
62	AT5G44555	-
63	AT5G44585	-
64	AT5G54790	-
65	AT3G46490	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
66	AT1G60040	AGAMOUS-like 49
67	AT1G28460	AGAMOUS-like 59
68	AT5G26120	alpha-L-arabinofuranosidase 2
69	AT1G52440	alpha/beta-Hydrolases superfamily protein
70	AT1G56630	alpha/beta-Hydrolases superfamily protein
71	AT5G41900	alpha/beta-Hydrolases superfamily protein
72	AT5G23810	amino acid permease 7
73	AT5G54620	Ankyrin repeat family protein
74	AT1G28130	Auxin-responsive GH3 family protein
75	AT1G33880	Avirulence induced gene (AIG1) family protein
76	AT3G29970	B12D protein
77	AT3G28857	basic helix-loop-helix (bHLH) DNA-binding family protein
78	AT1G68810	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
79	AT2G20100	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
80	AT2G44480	beta glucosidase 17
81	AT5G24550	beta glucosidase 32
82	AT1G61820	beta glucosidase 46
83	AT1G32280	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
84	AT3G22142	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein

Continued on next page

Table 6.2 – continued from previous page

Number	Atg name	Short description
85	AT4G22485	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
86	AT5G38195	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
87	AT1G11190	bifunctional nuclease i
88	AT5G46370	Ca2+ activated outward rectifying K+ channel 2
89	AT5G46360	Ca2+ activated outward rectifying K+ channel 3
90	AT1G21550	Calcium-binding EF-hand family protein
91	AT5G41380	CCT motif family protein
92	AT1G68795	CLAVATA3/ESR-RELATED 12
93	AT3G10116	COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family
94	AT4G12432	conserved peptide upstream open reading frame 26
95	AT5G52770	Copper transport protein family
96	AT4G03570	Cystatin/monellin superfamily protein
97	AT2G03955	Cysteine-rich protein
98	AT1G66540	Cytochrome P450 superfamily protein
99	AT3G44970	Cytochrome P450 superfamily protein
100	AT3G30290	cytochrome P450, family 702, subfamily A, polypeptide 8
101	AT5G42590	cytochrome P450, family 71, subfamily A, polypeptide 16
102	AT1G34047	Defensin-like (DEFL) family protein
103	AT2G21725	Defensin-like (DEFL) family protein
104	AT1G06080	delta 9 desaturase 1
105	AT4G34060	demeter-like protein 3
106	AT1G66090	Disease resistance protein (TIR-NBS class)
107	AT4G08450	Disease resistance protein (TIR-NBS-LRR class) family
108	AT4G11340	Disease resistance protein (TIR-NBS-LRR class) family
109	AT1G59725	DNAJ heat shock family protein
110	AT3G21920	Domain of unknown function (DUF26)
111	AT3G22040	Domain of unknown function (DUF26)
112	AT3G01326	ECA1 gametogenesis related family protein
113	AT3G45243	ECA1 gametogenesis related family protein
114	AT3G01322	ECA1-like gametogenesis related family protein
115	AT3G01327	ECA1-like gametogenesis related family protein

Continued on next page

Table 6.2 – continued from previous page

Number	Atg name	Short description
116	AT3G01329	ECA1-like gametogenesis related family protein
117	AT5G09980	elicitor peptide 4 precursor
118	AT5G09978	elicitor peptide 7 precursor
119	AT4G05130	equilibrative nucleoside transporter 4
120	AT5G10770	Eukaryotic aspartyl protease family protein
121	AT5G20260	Exostosin family protein
122	AT1G66490	F-box and associated interaction domains-containing protein
123	AT3G20690	F-box and associated interaction domains-containing protein
124	AT4G10190	F-box and associated interaction domains-containing protein
125	AT3G44090	F-box family protein
126	AT5G25860	F-box/RNI-like superfamily protein
127	AT5G38390	F-box/RNI-like superfamily protein
128	AT5G25850	F-box/RNI-like/FBD-like domains-containing protein
129	AT5G53840	F-box/RNI-like/FBD-like domains-containing protein
130	AT1G66000	Family of unknown function (DUF577)
131	AT1G06090	Fatty acid desaturase family protein
132	AT1G06100	Fatty acid desaturase family protein
133	AT5G22500	fatty acid reductase 1
134	AT3G44550	fatty acid reductase 5
135	AT5G53592	FBD-like domain family protein
136	AT1G01590	ferric reduction oxidase 1
137	AT5G50160	ferric reduction oxidase 8
138	AT5G07800	Flavin-binding monooxygenase family protein
139	AT1G62570	flavin-monooxygenase glucosinolate S-oxygenase 4
140	AT1G33811	GDSL-like Lipase/Acylhydrolase superfamily protein
141	AT5G03590	GDSL-like Lipase/Acylhydrolase superfamily protein
142	AT1G30040	gibberellin 2-oxidase
143	AT5G17330	glutamate decarboxylase
144	AT5G61660	glycine-rich protein
145	AT1G66270	Glycosyl hydrolase superfamily protein
146	AT2G14690	Glycosyl hydrolase superfamily protein
147	AT3G60140	Glycosyl hydrolase superfamily protein
148	AT5G17500	Glycosyl hydrolase superfamily protein

Continued on next page

Table 6.2 – continued from previous page

Number	Atg name	Short description
149	AT5G37960	GroES-like family protein
150	AT2G22190	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
151	AT4G12430	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein
152	AT5G40290	HD domain-containing metal-dependent phosphohydrolase family protein
153	AT2G18550	homeobox protein 21
154	AT1G14440	homeobox protein 31
155	AT3G50460	homolog of RPW8 2
156	AT1G54400	HSP20-like chaperones superfamily protein
157	AT3G30280	HXXXD-type acyl-transferase family protein
158	AT1G61080	Hydroxyproline-rich glycoprotein family protein
159	AT1G49120	Integrase-type DNA-binding superfamily protein
160	AT5G03570	iron regulated 2
161	AT4G04080	ISCU-like 3
162	AT4G08360	KOW domain-containing protein
163	AT5G28030	L-cysteine desulfhydrase 1
164	AT5G48100	Laccase/Diphenol oxidase family protein
165	AT4G13230	Late embryogenesis abundant protein (LEA) family protein
166	AT5G53820	Late embryogenesis abundant protein (LEA) family protein
167	AT3G58190	lateral organ boundaries-domain 29
168	AT2G14440	Leucine-rich repeat protein kinase family protein
169	AT3G27940	LOB domain-containing protein 26
170	AT5G35900	LOB domain-containing protein 35
171	AT1G49430	long-chain acyl-CoA synthetase 2
172	AT1G57550	Low temperature and salt responsive protein family
173	AT1G52110	Mannose-binding lectin superfamily protein
174	AT1G11670	MATE efflux family protein
175	AT1G66760	MATE efflux family protein
176	AT1G10745	Maternally expressed gene (MEG) family protein
177	AT5G39635	MIR169C; miRNA
178	AT5G41663	MIR319/MIR319B; miRNA
179	AT2G39885	MIR393A; miRNA

Continued on next page

Table 6.2 – continued from previous page

Number	Atg name	Short description
180	AT1G78478	MIR833a; miRNA
181	AT5G39693	MIR869a; miRNA
182	AT1G35310	MLP-like protein 168
183	AT5G14750	myb domain protein 66
184	AT1G52040	myrosinase-binding protein 1
185	AT3G04070	NAC domain containing protein 47
186	AT4G17980	NAC domain containing protein 71
187	AT1G78390	nine-cis-epoxycarotenoid dioxygenase 9
188	AT2G37460	nodulin MtN21 /EamA-like transporter family protein
189	AT3G30340	nodulin MtN21 /EamA-like transporter family protein
190	AT4G08290	nodulin MtN21 /EamA-like transporter family protein
191	AT5G53190	Nodulin MtN3 family protein
192	AT1G03810	Nucleic acid-binding, OB-fold-like protein
193	AT2G16230	O-Glycosyl hydrolases family 17 protein
194	AT3G44765	other RNA
195	AT1G33830	P-loop containing nucleoside triphosphate hydrolases superfamily protein
196	AT1G33930	P-loop containing nucleoside triphosphate hydrolases superfamily protein
197	AT4G11310	Papain family cysteine protease
198	AT4G11320	Papain family cysteine protease
199	AT2G19150	Pectin lyase-like superfamily protein
200	AT3G17060	Pectin lyase-like superfamily protein
201	AT5G18990	Pectin lyase-like superfamily protein
202	AT5G44840	Pectin lyase-like superfamily protein
203	AT3G49110	peroxidase CA
204	AT2G18150	Peroxidase superfamily protein
205	AT1G33920	phloem protein 2-A4
206	AT1G80110	phloem protein 2-B11
207	AT5G24560	phloem protein 2-B12
208	AT2G02340	phloem protein 2-B8
209	AT1G58320	PLAC8 family protein
210	AT3G55680	Plant invertase/pectin methylesterase inhibitor superfamily protein
211	AT2G18660	plant natriuretic peptide A
212	AT5G22560	Plant protein of unknown function (DUF247)

Continued on next page

Table 6.2 – continued from previous page

Number	Atg name	Short description
213	AT2G24870	Plant self-incompatibility protein S1 family
214	AT2G24880	Plant self-incompatibility protein S1 family
215	AT3G55672	Plant self-incompatibility protein S1 family
216	AT3G55677	Plant self-incompatibility protein S1 family
217	AT5G63087	Plant thionin family protein
218	AT2G01660	plasmodesmata-located protein 6
219	AT1G66480	plastid movement impaired 2
220	AT5G05020	Pollen Ole e 1 allergen and extensin family protein
221	AT1G78480	Prenyltransferase family protein
222	AT3G49300	proline-rich family protein
223	AT1G22720	Protein kinase superfamily protein
224	AT1G69790	Protein kinase superfamily protein
225	AT5G25840	Protein of unknown function (DUF1677)
226	AT3G50350	Protein of unknown function (DUF1685)
227	AT1G10750	Protein of Unknown Function (DUF239)
228	AT4G10220	Protein of Unknown Function (DUF239)
229	AT4G17505	Protein of Unknown Function (DUF239)
230	AT5G05030	Protein of Unknown Function (DUF239)
231	AT5G11660	Protein of Unknown Function (DUF239)
232	AT5G25410	Protein of Unknown Function (DUF239)
233	AT5G06610	Protein of unknown function (DUF620)
234	AT2G21727	Protein of unknown function (DUF784)
235	AT4G13968	Protein of unknown function (PD694200)
236	AT2G23067	Putative membrane lipoprotein
237	AT5G17340	Putative membrane lipoprotein
238	AT1G23147	RALF-like 3
239	AT4G13075	RALF-like 30
240	AT5G26594	response regulator 24
241	AT2G23640	Reticulan like protein B13
242	AT4G13240	RHO guanyl-nucleotide exchange factor 9
243	AT4G10600	RING/FYVE/PHD zinc finger superfamily protein
244	AT1G35330	RING/U-box superfamily protein
245	AT1G72220	RING/U-box superfamily protein
246	AT4G10160	RING/U-box superfamily protein
247	AT5G37270	RING/U-box superfamily protein
248	AT4G10290	RmlC-like cupins superfamily protein

Continued on next page

Table 6.2 – continued from previous page

Number	Atg name	Short description
249	AT5G38940	RmlC-like cupins superfamily protein
250	AT5G03580	RNA-binding (RRM/RBD/RNP motifs) family protein
251	AT4G15417	RNAse II-like 1
252	AT5G04370	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
253	AT5G37970	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
254	AT4G10457	SCR-like 1
255	AT4G13345	Serinc-domain containing serine and sphingolipid biosynthesis protein
256	AT5G22860	Serine carboxypeptidase S28 family protein
257	AT3G02110	serine carboxypeptidase-like 25
258	AT1G60060	Serine/threonine-protein kinase WNK (With No Lysine)-related
259	AT5G03600	SGNH hydrolase-type esterase superfamily protein
260	AT5G42170	SGNH hydrolase-type esterase superfamily protein
261	AT5G39440	SNF1-related protein kinase 1.3
262	AT1G68800	TCP domain protein 12
263	AT1G66020	Terpenoid cyclases/Protein prenyltransferases superfamily protein
264	AT3G29110	Terpenoid cyclases/Protein prenyltransferases superfamily protein
265	AT1G45145	thioredoxin H-type 5
266	AT5G37478	TPX2 (targeting protein for Xklp2) protein family
267	AT1G35170	TRAM, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein
268	AT1G58330	transcription factor-related
269	AT1G32040	transposable element gene
270	AT1G33817	transposable element gene
271	AT1G33835	transposable element gene
272	AT1G47320	transposable element gene
273	AT1G53630	transposable element gene
274	AT1G62695	transposable element gene
275	AT2G14420	transposable element gene
276	AT2G27375	transposable element gene
277	AT2G38990	transposable element gene

Continued on next page

Table 6.2 – continued from previous page

Number	Atg name	Short description
278	AT3G22136	transposable element gene
279	AT3G26605	transposable element gene
280	AT3G28855	transposable element gene
281	AT4G37570	transposable element gene
282	AT5G24915	transposable element gene
283	AT5G28240	transposable element gene
284	AT2G29330	tropinone reductase
285	AT1G65990	type 2 peroxiredoxin-related / thiol specific antioxidant / mal allergen family protein
286	AT1G53930	Ubiquitin-like superfamily protein
287	AT5G40630	Ubiquitin-like superfamily protein
288	AT5G37950	UDP-Glycosyltransferase superfamily protein
289	AT1G12600	UDP-N-acetylglucosamine (UAA) transporter family
290	AT2G28760	UDP-XYL synthase 6
291	AT3G15300	VQ motif-containing protein
292	AT5G46350	WRKY DNA-binding protein 8
293	AT3G62340	WRKY family transcription factor
294	AT1G31260	zinc transporter 10 precursor

List of Figures

1.1 **PRC2 complex core components in Drosophila, human and Arabidopsis.** 4
1.2 **Models of mammalian PRC2 transcription regulation mechanisms.** . . 10

4.1 **CLF and SWN are evolutionary conserved in higher plants and diverge including MEA into 6 branches.** 30
4.2 **SWN is conserved among Arabidopsis accessions.** 31
4.3 **CLF, SWN and MEA share 7 conserved domains and differ in the middle part.** 33
4.4 **CLF:HA abundance in *35S:CLF:HA* does not correlate with *CLF* expression levels and is low compared to positive controls.** 36
4.5 **CLF-HA is enriched on *AG*)and *FLC*.** 38
4.6 **Enrichment of GFP:CLF on *FLC* and *AG* is statistically significant compared to loci where no enrichment is expected which are the most stringent negative controls.** 41
4.7 **Work flow of ChIP-seq experiment** 45
4.8 **High similarity between positive and negative controls in ChIP-Seq of *35:GFP:CLF*** 46
4.9 **MG-132 treatment increases abundance of CLF** 48
4.10 **High overlap between H3K27me3 target genes in *clf* and *swn* and wt plants.** 52
4.11 **Three types of H3K27me3 distribution patterns between *clf*, *swn* and wt.** 54
4.12 **Type of H3K27me3 distribution patterns can be confirmed by ChIP-PCR.** 56
4.13 **Strong H3K27me3 decrease in *clf* and subtle increase in *swn* .** 59
4.14 **CLF and SWN activate and repress CLF dependent genes in combination or independently from each other.** 62
4.15 **CLF dependent genes are distributed in clusters across the chromosomes** 64
4.16 **Telo boxes are enriched in CLF dependent regions.** 68

5.1 **CLF-PRC2 and SWN-PRC2 which have distinct enzymatic activities work together, independently and antagonistically to regulate transcription in a PRC2 dependent and maybe also a PRC2 independent pathway** 87

List of Tables

3.1 **Gene names of CLF-, SWN- and MEA- homologues** 19

3.2 **PCR conditions for Gateway**20table.caption.7 . . . . . 3.3

a-EMF2, a-VRN2, a-CLF and a-SWN antibodies 21table.caption.8

3.4 **Primers used for Gateway cloning, mRNA abundance determination and ChIP-PCR** 22

3.5 **qPCR conditions** 23

4.1 **Number of reads of two *35S:GFP:CLF* replicates after sequencing and analysis.** 45

4.2 **Definitions of terms.** 51

4.3 **Number of reads of three replicates of clf, swn and wt after sequencing and analysis.** 52

4.4 **Summary and features of CLF dependent genes in enriched windows.** 65

5.1 **Abbreviations** 105

6.1 **CLF dependent genes** 108

6.2 **SWN affected genes** 128

Danksagung

Ich möchte mich bedanken bei:

Franziska, für das Projekt, die Betreuung, die Freiheit, die Geduld, die enorme Unterstützung, die Kreativität, die Härte, die Inspiration, für die Art und Weise wie Sie mir Wissenschaft vorgelebt hat. Ich habe es wirklich sehr genossen mit zu arbeiten und von dir zu lernen.

Prof. George Coupland for the opportunity to work in this wonderful department, for having created such a positive and stimulating environment and for taking time when I needed support and help.

Prof. Ute Höcker für die Übernahme des Koreferats und dafür Enthusiasmus und Spass an der Wissenschaft und am wissenschaftlichen Denken in mir gefördert zu haben.

Prof. Martin Hülskamp für die Übernahme des Prüfungsvorsitzes.

Sara, para la instrucción, el soporte, el motivación para mejorarse, la creatividad, la risa, la electricidad, la paciencia y para tomar tiempo cada vez que tenia preguntas y necesito algo. Servs de modelo cientfico para mi.

Xue, Vimal and **Julia** for a lot of support in the bioinformatic part of this thesis. I don't know what I would have done without your help and support!!!

Samson, Julieta und **Maida** für viele Ideen, Unterstützung beim ChIP-Seq, Korrekturlesen, Lachen. Ihr ein sehr wichtiger Teil von tollen Zeit die ich hier hatte!

Petra für die Hilfe bei meinem versuch so wenig wie möglich zu klonieren :)

Benny und **Jonas** für die Hilfe mit R und dem Tukey-Test.

Elizabeth und **Leon** für eine coole Zeit und dafür dass die mich gezwungen haben strukturierter zu arbeiten.

The **Groups Turck** and **Coupland** for the very nice time and the scientific input discussions.

Martin und **Jan** für die Hilfe mit Latex/Korrekturlesen und da sein wenn ich Hilfe brauche.

David und **Simon** für die Herberge im Büro, Freundschaft, da sein, Spass und lebenserhaltende Maßnahmen.

Victor für Freundschaft und dafür, dass du mich immer wieder erdest.

Ineke für innere Ruhe, Wärme und Herberge.

Megatron für super geile Ultrakräfte die letzten 10 Monate much zu disziplinieren und kaum etwas anderes zu machen ausser an meinen Ergebnissen und später an meiner Thesis zu arbeiten.

Nicolas Jaar, Alt-J und Hiphop für Musik die mir das Gefühl gegeben hat cool zu sein während ich schreibe.

Die Cafs Brussels und Halmackenreuther für schöne abwechslungsreiche Atmosphäre.

Nico und ganze Familie **Matentzoglu** für Freundschaft, geistige Unterstützung und Familie für mich sein.

Meine Familie für die wahnsinnige und bedingungslose Unterstützung und Liebe.

Die vorliegende Arbeit wurde in der Zeit von Januar 2010 bis Mai 2013 in der Arbeitsgruppe Transcription Control and Chromatin Structure im Institut für Entwicklungsbiologie der Pflanzen, am Max Planck Institut für Pflanzenzüchtungsforschung Anleitung von Herrn Prof. Dr. George Coupland angefertigt.

Ich versichere hiermit, dass ich meine Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

The work in hand, was conducted from January 2010 to April 2013 in the Department for Plant Developmental Biology, at the Max Planck Institute for Plant Breeding Research under the supervision of Prof. Dr. George Coupland.

I hereby declare that this thesis presents my own work, that I used no other than the denoted resources/tools and marked citations accordingly.

Köln, den 21. Mai 2013
Cologne, Mai 21st 2013

_____
(Theodoros Zografou)