

---

Electronic Thesis and Dissertation Repository

---

10-1-2019 2:30 PM

# Genome-wide occupancy of Polycomb group proteins and chromatin remodeler SPLAYED and their interplay in Arabidopsis

Jie Shu, *The University of Western Ontario*

Supervisor: Cui, Yuhai, *The University of Western Ontario*

Co-Supervisor: Kohalmi, Susanne, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biology

© Jie Shu 2019

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Plant Biology Commons](#)

---

## Recommended Citation

Shu, Jie, "Genome-wide occupancy of Polycomb group proteins and chromatin remodeler SPLAYED and their interplay in Arabidopsis" (2019). *Electronic Thesis and Dissertation Repository*. 6564.  
<https://ir.lib.uwo.ca/etd/6564>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

Genome-wide occupancy of Polycomb group proteins and chromatin remodeler  
SPLAYED and their interplay in *Arabidopsis*

by

Jie Shu

Graduate Program in Biology

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies  
Western University  
London, Ontario, Canada

© Jie Shu 2019

## Abstract

The Polycomb group (PcG) proteins form two protein complexes, Polycomb repressive complex 1 (PRC1) and PRC2, which are key epigenetic regulators in eukaryotes. PRC2 represses gene expression by catalyzing trimethylation of histone H3 lysine 27 (H3K27me3). In *Arabidopsis thaliana*, CURLY LEAF (CLF) and SWINGER (SWN) are two major H3K27 methyltransferases, playing essential roles in plant growth and development. Despite their importance, genome-wide occupancy profiles of CLF and SWN have not yet been determined and compared. In this thesis, I generated transgenic lines expressing GFP-tagged CLF/SWN and determined the genome-wide distributions of CLF and SWN in *Arabidopsis* seedlings. I also compared the global H3K27me3 levels in wild-type (WT) and PcG mutants (*clf*, *swn*, and *clf swn*). The data show that CLF and SWN co-targeted a large number of genes, except that SWN had a few hundred more targets. The GAGA-like and Telo-box-like motifs were found enriched in CLF- and SWN-occupied regions. The global H3K27me3 levels in *clf*, but not in *swn*, were markedly reduced compared to WT, and H3K27me3 was undetectable in *clf swn*. Thus, this work provides a useful resource for the plant epigenetics community for dissecting the functions of PRC2 in plant growth and development.

SPLAYED (SYD) is an SWI/SNF-type chromatin remodeler that plays critical roles in the regulation of gene expression. SYD is the closest homolog of BRAHMA (BRM), functioning redundantly and/or differentially with BRM in *Arabidopsis*. Recently, the genome-wide occupancy of BRM has been profiled. However, the occupancy of SYD was still missing. Therefore, I generated a transgenic line expressing GFP-tagged SYD which was used for profiling the genome-wide occupancy of SYD at the seedling and reproductive stages. SYD and BRM co-localized at over three thousand genes, suggesting that SYD and BRM function redundantly at these genes. When analyzing the global distribution of H3K27me3 in *syd*, it was discovered that the loss of SYD activity resulted in changes in H3K27me3 levels at over several hundred genes compared to WT. To summarize, this work demonstrates the genome-wide occupancy of SYD and emphasizes a global functional interplay between SYD and PcG repression.

## **Keywords**

Polycomb group proteins, CURLY LEAF, SWINGER, H3K27me3, floral transition, SPLAYED, BRAHMA, ChIP-seq, RNA-seq, *Arabidopsis thaliana*

## Summary for Lay Audience

A seed can develop into a whole plant that has many different tissues. However, every cell has the same genome containing the same genes. In general, it is the expression levels of genes that contribute to the production of different cell types, thereby different tissue types. There are many proteins playing important roles by changing the expression levels of different genes. In this thesis, I focus on examining two types of proteins by using a model plant organism *Arabidopsis thaliana*. These two types of proteins play very important roles in plants, as plants cannot grow normally when the functions of genes (encoding these two types of proteins) are disrupted.

The first type is the Polycomb group (PcG) proteins. They usually form two complexes: Polycomb repressive complex 1 (PRC1) and PRC2. PRC2 lowers gene expression levels by adding three methyl groups to histone H3 at lysine 27 residue (referred to H3K27me3). PRC2 complex has two key subunits called CURLY LEAF (CLF) and SWINGER (SWN), which can change expression levels of different genes by producing H3K27me3 in plant cells. Therefore, it is important to find genes whose expression levels are controlled by CLF and SWN. Thus, in this thesis, I conducted experiments to find CLF- and SWN-controlled genes. Further, I also found that CLF and SWN function similarly in controlling expression levels of genes in plants. The second type is SPLOYED (SYD) which can change chromatin structures. SYD plays important roles in plant growth and development, as plants grow abnormally when SYD function is removed. Finally, I found many genes that are controlled by SYD. Previous literature proposed that there is interplay between SYD and PRC2. Therefore, I also examined the interplay between SYD and PRC2. My data suggest that SYD and PRC2 play opposite and/or cooperative roles in controlling gene expression levels in plants.

## Co-Authorship Statement

My supervisor, Dr. Yuhai Cui, provided insight and strategic direction for the projects and edited my thesis. My co-supervisor, Dr. Susanne Kohalmi, provided insightful comments as well as edited my thesis.

I performed all of the experiments described in the following thesis except for the following:

Next-generation sequencing in section 2.9 was performed at The Hospital for Sick Children in Toronto. ChIP-seq data analysis was conducted by myself, our current lab member Dr. Chen Chen, and Dr. Chenlong Li from Sun Yat-sen University, China.

RNA-seq in section 2.10.2 was performed at Novogene in China. The data were analyzed by myself, and our lab members Dr. Chen Chen, and Raj K. Thapa.

Sections 3.1 and 4.1 contain material that has been published in:

Shu, J., Chen, C., Thapa, R.K., Bian, S., Nguyen, V., Yu, K., Yuan, Z.C., Liu, J., Kohalmi, S.E., Li, C. and Cui, Y. (2019) Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in *Arabidopsis* seedlings. *Plant Direct*, 3, e00100.

## **Acknowledgments**

The completion of my Ph.D. project would have not been possible without the support of many people.

First of all, I would like to express my sincere gratitude to my supervisor, Dr. Yuhai Cui, for his unconditional support and encouragement throughout the years. It has been my great pleasure to join his research group and participate in exciting projects. I deeply appreciate all of the support and help from my co-supervisor, Dr. Susanne Kohalmi, whose enthusiasm for plant science has always been inspiring me. My sincere thanks also go to my advisory committee members, Dr. Mark Gijzen and Dr. Danielle Way, for their insightful comments and encouragement during my graduate study.

I would like to send my thanks to all my fellow lab mates: Dr. Chenlong Li, Dr. Behnaz Saatian, Dr. Chen Chen, Dr. Jun Liu, Dr. Wenqun Fu, Dr. Shaomin Bian, Dr. Zhengfeng Jiang, Jakir Hossan, and Raj K. Thapa. I genuinely appreciate their collaboration, insightful discussion, and encouragement. I am thankful to Ms. Vi Nguyen for providing lab operational support and editing my thesis. Many thanks are also conveyed to all of my friends at AAFC-London, without their friendship and encouragement, I would not fully enjoy my stay here.

I specifically would like to thank Shanwu Lyu for all his endless support and encouragement. Thanks for making my life wonderful.

Last but not least, I thank my grandparents, parents, and sister for supporting me throughout my whole life. I am profoundly grateful for everything that they have done for me.

## Table of Contents

Abstract .....	ii
Keywords .....	iii
Summary for Lay Audience .....	iv
Co-Authorship Statement.....	v
Acknowledgments.....	vi
Table of Contents .....	vii
List of Figures .....	xi
List of Appendices .....	xiii
List of Abbreviations .....	xiv
1 INTRODUCTION .....	1
1.1 Histone modification-mediated gene expression .....	1
1.2 Polycomb group proteins .....	2
1.3 Recruitment of PcG components to targets .....	9
1.4 PcG proteins play essential roles in <i>Arabidopsis</i> developmental transitions.....	10
1.4.1 The floral transition.....	10
1.4.2 CLF plays critical roles in floral transition .....	11
1.5 SPLAYED, an SWI/SNF-type chromatin remodeler .....	13
1.6 PcG and SWI/SNF function antagonistically or synergistically .....	14
1.7 Thesis objectives .....	15



2	MATERIAL AND METHODS .....	16
2.1	Plant material and growth conditions .....	16
2.2	Crossing of <i>Arabidopsis</i> plants .....	16
2.3	Generation of plasmid constructs for plant transformations.....	17
2.4	Western blot.....	17
2.5	PCR-based genotyping.....	18
2.6	Flowering time measurement.....	19
2.7	Chromatin immunoprecipitation (ChIP) .....	19
2.8	ChIP DNA library preparation.....	19
2.9	ChIP-seq and data analysis .....	20
2.10	RNA protocols .....	21
2.10.1	RNA extraction .....	21
2.10.2	Gene expression analysis .....	21
2.10.3	RNA-sequencing analysis.....	22
2.11	Statistical analysis.....	22
2.12	Primer design .....	23
2.13	Accession numbers .....	23
3	RESULTS .....	24
3.1	Genome-wide occupancy of H3K27 methyltransferases CLF and SWN in <i>Arabidopsis</i> seedlings .....	24

3.1.1	<i>GFP</i> -tagged <i>CLF</i> and <i>SWN</i> transgenes function <i>in vivo</i> .....	24
3.1.2	Genome-wide occupancy of <i>CLF</i> and <i>SWN</i> .....	27
3.1.3	Identification of two DNA motifs from <i>CLF</i> and <i>SWN</i> peaks .....	36
3.1.4	Functions of <i>CLF</i> and <i>SWN</i> targets in plant growth and development....	36
3.1.5	Genome-wide profiling of H3K27me3 in <i>clf-29</i> , <i>swn-4</i> , and <i>clf-29 swn-4</i> .....	43
3.1.6	Transcriptome profiling in <i>clf-29</i> , <i>swn-4</i> , and <i>clf-29 swn-4</i> .....	61
3.1.7	Implication of <i>CLF</i> in floral transition.....	61
3.2	Genome-wide characterization of <i>Arabidopsis</i> SWI/SNF chromatin remodeler <i>SYD</i> .....	64
3.2.1	<i>GFP</i> -tagged <i>SYD</i> transgene functions <i>in vivo</i> .....	71
3.2.2	Genomic regions occupied by <i>SYD</i> at two developmental stages .....	74
3.2.3	Patterns of <i>SYD</i> occupancy .....	77
3.2.4	<i>SYD</i> may mainly activate its targets.....	77
3.2.5	Interplay between <i>SYD</i> and <i>BRM</i> .....	77
3.3	Interplay between SWI/SNF and PcG .....	82
3.3.1	Loss of <i>SYD</i> activity leads to changes of H3K27me3 at hundreds of genes .....	82
3.3.2	Increased occupancy of <i>CLF</i> / <i>SWN</i> at target loci in <i>brm-1</i> .....	92
4	DISCUSSION .....	95
4.1	Genome-wide characterization of <i>CLF</i> and <i>SWN</i> .....	95

4.1.1	CLF and SWN are the only active H3K27 methyltransferases at the seedling stage .....	95
4.1.2	CLF and SWN play both redundant and differential roles in depositing H3K27me3 .....	96
4.1.3	Implications of CLF and SWN in gene expression .....	97
4.1.4	CLF regulates the floral transition .....	98
4.2	Global characterization of SYD .....	99
4.2.1	A truncated <i>SYD</i> fully rescues <i>syd-5</i> phenotype .....	99
4.2.2	SYD may mainly function as a transcription activator .....	100
4.2.3	SYD and BRM share similar occupancy patterns .....	100
4.3	Antagonistic and synergistic interactions between SWI/SNF and PcG .....	101
4.4	Concluding remarks and future direction .....	102
	References .....	104
	Appendices .....	124
	Curriculum Vitae .....	139

## List of Figures

Figure 1 Conserved subunits of PcG complexes in <i>Drosophila</i> and <i>Arabidopsis</i> .....	4
Figure 2 Domain architecture of CLF and SWN proteins. ....	7
Figure 3 Transgenic <i>Arabidopsis</i> lines expressing GFP-tagged CLF or SWN.....	25
Figure 4 Pearson correlation plots for two CLF and SWN ChIP-seq replicates. ....	28
Figure 5 Genome-wide occupancy of CLF and SWN.....	30
Figure 6 ChIP-seq signals at representative loci co-occupied by CLF and SWN. ....	32
Figure 7 ChIP-seq signals at representative loci uniquely occupied by SWN. ....	34
Figure 8 Motifs enriched in CLF and SWN peaks. ....	37
Figure 9 Functional categorization of CLF and SWN targets. ....	39
Figure 10 Representative genes co-targeted by CLF and SWN. ....	41
Figure 11 Distribution of H3K27me3 in <i>Arabidopsis</i> genome.....	44
Figure 12 Pattern of CLF and SWN peaks. ....	46
Figure 13 Overlaps between H3K27me3-marked and CLF/SWN-occupied genes. ....	49
Figure 14 BRM occupancy overlaps with H3K27me3/CLF/SWN occupancies.....	51
Figure 15 Genome-wide profiling of H3K27me3 in different genetic backgrounds.....	53
Figure 16 Histone H3 levels in different genetic backgrounds.....	55
Figure 17 Overlaps of genes occupied by CLF/SWN and those showing changes in H3K27me3. ....	57
Figure 18 Three patterns of H3K27me3 reduction in <i>clf-29</i> .....	59

Figure 19 Transcriptome profiling of PcG mutants. ....	62
Figure 20 ChIP-seq signals at representative genes involved in the control of flowering .....	65
Figure 21 CLF directly targets the flowering activator <i>AGL17</i> . ....	67
Figure 22 Partial contribution of <i>AGL17</i> to the early flowering phenotype of <i>clf-29</i> . ....	69
Figure 23 Truncated <i>SYD</i> fully completes <i>syd-5</i> .....	72
Figure 24 Genome-wide occupancy of SYD at two developmental stages.....	75
Figure 25 Pattern of SYD peaks. ....	78
Figure 26 Transcriptome profiling of <i>syd-5</i> . ....	80
Figure 27 SYD and BRM target common genes. ....	83
Figure 28 Genome-wide profiling of H3K27me3 in <i>syd-5</i> .....	85
Figure 29 Representative genes showing changes in H3K27me3 levels in <i>syd-5</i> . ....	88
Figure 30 Antagonistic and synergistic functions between SYD and CLF at representative genes. ....	90
Figure 31 Altered occupancies of CLF and SWN in <i>brm-1</i> . ....	93

## **List of Appendices**

Appendix A: Primers used in this thesis .....	124
Appendix B: Total read numbers of all ChIP-seq samples.....	132
Appendix C: Total read numbers of all RNA-seq samples.....	134
Appendix D: List of genes regulated by BRM and PcG.....	136

## List of Abbreviations

The International System of Units (SI) were used unless otherwise stated.

%	percent
-/-	homozygous
::	fused to (in the context of constructs)
+/-	heterozygous
aa	amino acid
AAFC	Agriculture and Agri-Food Canada
ABI3	ABSCISIC ACID INSENSITIVE 3
ABRC	<i>Arabidopsis</i> Biological Resource Center
AG	AGAMOUS
AGL	AGAMOUS-LIKE
AP1	APETALA 1
APOLO	AUXINREGULATED PROMOTER LOOP RNA
AS1/2	ASYMMETRIC LEAVES 1/2
<i>AT</i>	<i>Arabidopsis thaliana</i>
ATP	adenosine triphosphate
ATX1	ARABIDOPSIS TRITHORAX 1
BAC	bacterial artificial chromosome
BAH	bromo adjacent homology

BP	BREVIPEDICELLUS
BRM	BRAHMA
CGI	CpG island
ChIP-qPCR	chromatin immunoprecipitation combined with quantitative PCR
ChIP-seq	chromatin immunoprecipitation combined with next-generation sequencing
Chr	chromosome
CLF	CURLY LEAF
CO	CONSTANS
Col-0	Columbia
COLD AIR	COLD ASSISTED INTRONIC NONCODING RNA
CTAB	cetyltrimethylammonium bromide
CXC	cysteine-rich region
DNA	deoxyribonucleic acid
E(z)	Enhancer of zeste
EBS	EARLY BOLTING IN SHORT DAYS
EDTA	ethylenediaminetetraacetic acid
EMF1/2	EMBRYONIC FLOWER 1/2
Esc	Extra sex comb
EZD1/2	E(z) domain 1/2



FC	fold change
FCA	FLOWERING CONTROL LOCUS A
FDR	false discovery rate
FIE	FERTILIZATION INDEPENDENT ENDOSPERM
FIS2	FERTILIZATION INDEPENDENT SEED 2
FLC	FLOWERING LOCUS C
FLD	FLOWERING LOCUS D
FRI	FRIGIDA
FT	FLOWER LOCUS T
FUS3	FUSCA 3
FVE	FLOWERING LOCUS VE
GA	gibberellic acid
GAF	GAGA factor
GFP	green fluorescent protein
GO	Gene Ontology
H2AK119ub	monoubiquitination of histone H2A lysine 119
H3K27me	monomethylation of histone H3 lysine 27
H3K27me2	dimethylation of histone H3 lysine 27
H3K27me3	trimethylation of histone H3 lysine 27

IGB	Integrated Genome Browser
IP-MS	immunoprecipitation together with mass spectrometry
KNAT2	KNOTTED-LIKE FROM ARABIDOPSIS THALIANA 2
KNU	KNUCKLES
LB	left border primer of the T-DNA insertion
LEC2	LEAFY COTYLEDON 2
LFY	LEAFY
LHP1	LIKE HETEROCHROMATIN PROTEIN 1
lncRNA	long non-coding RNA
LP	primer specific to the 5' flanking sequence of T-DNA insertion
MEA	MEDEA
MINU1/2	MINUSULE 1/2
MLL	MIXED-LINEAGE LEUKEMIA
MP	MONOPTEROS
MS	Murashige and Skoog
MSI1-5	MULTICOPY SUPPRESSOR OF IRA 1-5
MYB2	MYB DOMAIN PROTEIN 2
NCBI	National Center for Biotechnology Information
NDG	Nearest Downstream Gene

NF-YC	nuclear factor-Y C
PcG	Polycomb group
PCR	polymerase chain reaction
Ph	Polyhomeotic
PID	PINOID
PKL	PICKLE
PRC1/2	Polycomb repressive complex 1/2
PRE	Polycomb response element
Psc	Posterior sex combs
PTM	post-translational modification
PVDF	polyvinylidene fluoride
REF6	RELATIVE OF EARLY FLOWERING 6
RNA-seq	ribonucleic acid sequencing
RP	primer specific to the 3' flanking sequence of the T-DNA insertion
SAM	shoot apical meristem
SANT	SWI3, ADA2, N-CoR, and TFIIIB
SAUR	SMALL AUXIN UPREGULATED RNA
Sce	Sex combs extra
SDS	sodium dodecyl sulfate

SET	Su(var)3-9, E(z), Trithorax
SHL	SHORT LIFE
SNC1	SUPPRESSOR OF NPR1, CONSTITUTIVE 1
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
STM	SHOOT MERISTEMLESS
Su(z)12	Suppressor of zeste 12
SVP	SHORT VEGETATIVE PHASE
SWI/SNF	SWItch/Sucrose Non-Fermentable
SWN	SWINGER
SYD	SPLAYED
SYD $\Delta$ C	GFP-tagged SYD fusion generated excluded its C-terminal domain
TAIR	The <i>Arabidopsis</i> Information Resource
TBST	tris-buffered saline with Tween 20
TCP	TEOSINTE BRANCHED1, CYCLOIDEA, and PCF
T-DNA	transfer DNA
TF	transcription factor
TMO3	TARGET OF MONOPTEROS 3
TRB	telomere-repeat factor
TrxG	Trithorax group

TSS	transcription start site
TTS	transcription termination site
ULT1	ULTRAPETALA 1
UTR	untranslated region
VEFS	VRN2, EMF2, FIS2, and Su(z)12
VRN1/2	VERNALIZATION 1/2
WT	wild-type
WUS	WUSCHEL

# 1 INTRODUCTION

## 1.1 Histone modification-mediated gene expression

In eukaryotes, chromatin is mainly composed of repeating units called nucleosome core which consists of 146 base pairs (bp) of DNA wrapped around a central histone octamer (two copies of each core histone H2A, H2B, H3, and H4; Luger *et al.*, 1997). N-terminal regions of the core histones, termed histone tails, undergo various post-translational modifications (PTMs) to regulate chromatin structure and dynamics (Huang *et al.*, 2014; Kim and Sung, 2014; Stillman, 2018). The chromatin has two distinct states, one is open and active called euchromatin, which is accessible to transcription factors (TFs) and co-factors. The other state is closed and silent called heterochromatin, which is inaccessible to TFs and co-factors (Gilbert *et al.*, 2004). Histone modifications occur mainly on the histone tails and play critical roles in determining the two states of chromatin during plant growth and development. Numerous PTMs, including methylation, acetylation, ubiquitination, phosphorylation, and sumoylation, have been identified (Kouzarides, 2007; Bannister and Kouzarides, 2011; Lawrence *et al.*, 2016; Villanueva *et al.*, 2017). Even though genetic instructions are carried by DNA, histone modifications are another layer of genetic information in eukaryotes. Generally, these histone modifications alter DNA-histone interactions and thereby, control gene expression. Depending on their effect on gene transcription, most PTMs are described as active or repressive marks (Venkatesh and Workman, 2015). For example, histone acetylation is usually linked to transcriptional activation. Similarly, methylation of histone H3 at lysine 4 and lysine 36 is associated with transcriptional activation, whereas methylation of H3 at lysine 9 and lysine 27 is implicated in transcriptional repression (Verdin and Ott, 2015; Becker *et al.*, 2016; Howe *et al.*, 2016; Xiao *et al.*, 2016; Xu *et al.*, 2019).

The trimethylation of histone H3 lysine 27 (H3K27me3) results from a series of mono-methylations of H3K27 and is a stable mark that is linked to transcriptional repression (Barski *et al.*, 2007; Zee *et al.*, 2010; Li *et al.*, 2018b). In mammalian embryonic stem cells, methylation of H3K27 is very abundant, displaying with 5%-10% H3K27me of all histone H3, 50%-70% H3K27me2, and 5%-10% H3K27me3 (Peters *et al.*, 2003; Laugesen *et al.*,

2019). In *Drosophila melanogaster* (*Drosophila*), the H3K27me<sub>3</sub>-marked domains usually cover more than 10-kb-long regions of the genome (Schwartz *et al.*, 2006). There are two different patterns of H3K27me<sub>3</sub>-marked regions in mammals: some are very large regions ( $\geq 100$  kb) such as those containing the *Hox* loci, while others are much smaller regions spanning only a few kilobases (Zhao *et al.*, 2007; Hawkins *et al.*, 2010; Margueron and Reinberg, 2011).

In *Arabidopsis thaliana* (*Arabidopsis*), the H3K27me<sub>3</sub>-marked regions were found to be remarkably shorter than those in *Drosophila* and mammals. At least 25% of all protein-coding regions were reported to be covered by H3K27me<sub>3</sub>, and the majority of H3K27me<sub>3</sub> was found in the promoters and gene bodies (Turck *et al.*, 2007; Zhang *et al.*, 2007; Lafos *et al.*, 2011; Wiles and Selker, 2017). Furthermore, when comparing the genome-wide patterns of H3K27me<sub>3</sub> in undifferentiated shoot apical meristems (SAMs) and differentiated leaves, it was found that plant H3K27me<sub>3</sub> is highly dynamic during differentiation and is associated with tissue-specific repression of genes (Lafos *et al.*, 2011). Additionally, it was reported that the genome-wide reprogramming of H3K27me<sub>3</sub> is essential in the regulation of key genes that lead to cell fate transition (He *et al.*, 2012). In conclusion, H3K27me<sub>3</sub> covers up a large percentage of the *Arabidopsis* genome and plays critical roles in the regulation of plant growth and development.

## 1.2 Polycomb group proteins

In recent years, much attention has been given to the Polycomb group (PcG) proteins, among the many factors that are involved in the regulation of chromatin dynamics. The PcG proteins play important roles in epigenetic regulation through repressing target genes (Dellino *et al.*, 2004). Initially described in *Drosophila* as regulators of gene expression (Lewis, 1978), PcG proteins form two functionally distinct complexes known as Polycomb repressive complex 1 (PRC1) and PRC2 (Margueron and Reinberg, 2011). PRC2 catalyzes H3K27me<sub>3</sub> (Cao *et al.*, 2002; Hansen *et al.*, 2008; Lafos *et al.*, 2011); whereas PRC1 catalyzes the monoubiquitination of histone H2A at lysine 119 (H2AK119ub) (Cao and Zhang, 2004).

The subunits of PRC2 are highly conserved in multicellular organisms, but the number of genes encoding each subunit varies among species (Mozgová and Hennig, 2015). The *Drosophila* PRC2 has four main components: Enhancer of zeste (E(z)), Suppressor of zeste 12 (Su(z)12), Extra sex comb (Esc), and p55 (Schwartz and Pirrotta, 2007; Margueron and Reinberg, 2011). In humans, two copies of *E(z)* - *EZH1* and *EZH2* exist (Ciferri *et al.*, 2012). In *Arabidopsis*, there are three homologous genes for the *E(z)* subunit: *CURLY LEAF* (*CLF*), *SWINGER* (*SWN*), and *MEDEA* (*MEA*); three for *Su(z)12*: *EMBRYONIC FLOWER 2* (*EMF2*), *VERNALIZATION 2* (*VRN2*), and *FERTILIZATION INDEPENDENT SEED 2* (*FIS2*); one gene for *Esc*: *FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE*); and five genes for *p55*, *MULTICOPY SUPPRESSOR OF IRA 1-5* (*MSI1-5*) (Figure 1; Whitcomb *et al.*, 2007).

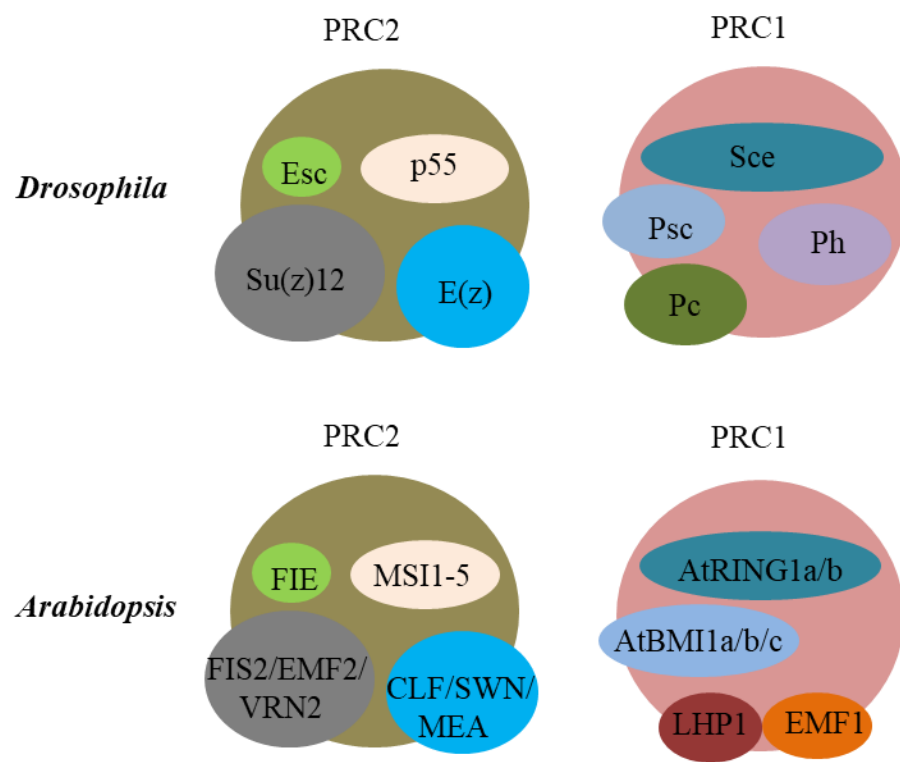
There are four core PRC1 subunits in *Drosophila*, which include Sex combs extra (Sce), Posterior sex combs (Psc), Pc, and Polyhomeotic (Ph) (Whitcomb *et al.*, 2007; Schuettengruber *et al.*, 2017). In *Arabidopsis*, two homologs of Sce (AtRING1a/1b) and three homologs of Psc (AtBMI1a/1b/1c) have been identified (Kim and Sung, 2014; Wang and Shen, 2018). Based on sequence similarity, no obvious homologs of Pc or Ph could be found. However, several proteins which show similar functions to Pc have been characterized. For example, LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) physically interacts with PRC1 subunits such as AtRING1a and largely co-localizes with H3K27me3-marked regions genome-wide, displaying PRC1-like functions in *Arabidopsis* (Blackledge *et al.*, 2015; Veluchamy *et al.*, 2016). The plant bromo-adjacent homology (BAH) domain-containing proteins SHORT LIFE (SHL) and EARLY BOLTING IN SHORT DAYS (EBS) recognize H3K27me3 and form a complex with EMBRYONIC FLOWER 1 (EMF1) to display PRC1-like functions (Li *et al.*, 2018b; Qian *et al.*, 2018; Yang *et al.*, 2018).

In *Arabidopsis*, two genes encoding H3K27 methyltransferases, *CLF* and *SWN*, are highly redundant. The expression patterns of *CLF* and *SWN* are highly similar, with both showing expression during vegetative and reproductive processes (Goodrich *et al.*, 1997; Chanvivattana *et al.*, 2004). *CLF* has long been known to be necessary for the control of



**Figure 1** Conserved subunits of PcG complexes in *Drosophila* and *Arabidopsis*.

Core components of PRC2 and PRC1 in *Drosophila* and *Arabidopsis* are shown. Same colors represent the functional equivalents of PcG components in *Drosophila* and *Arabidopsis*. LHP1 and EMF1 are PRC1-like complex(es) in *Arabidopsis*. Figure redrawn based on Kim and Sung (2014).

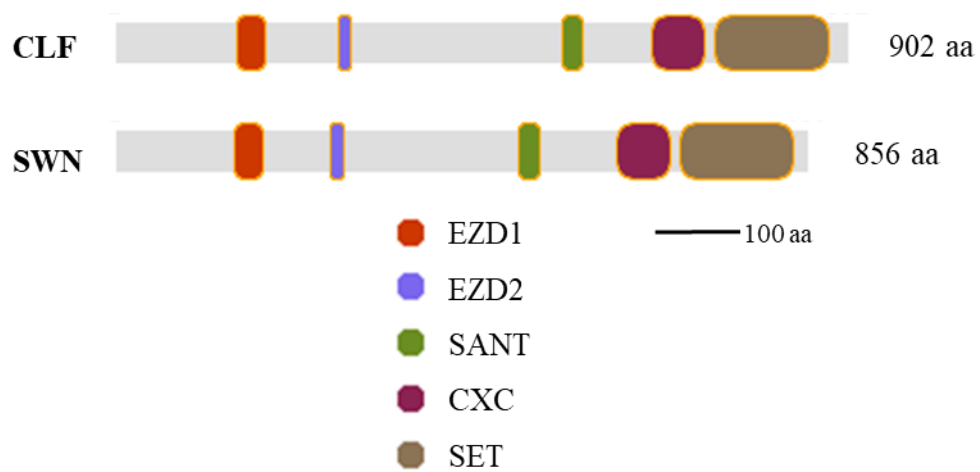


leaf and flower morphology, likely through repressing *AGAMOUS* (*AG*) and *SHOOT MERISTEMLESS* (*STM*) (Goodrich *et al.*, 1997; Schubert *et al.*, 2006). CLF is also necessary for *WUSCHEL* (*WUS*) repression, contributing to the termination of floral stem cells (Liu *et al.*, 2011), and is required for maintaining root meristem activity by antagonizing the chromatin remodeler PICKLE (*PKL*) (Aichinger *et al.*, 2011). *clf* mutants exhibit severe phenotypes including dwarf stature, curly leaf, and early flowering (Schatlowski *et al.*, 2010; Farrona *et al.*, 2011b; Lopez-Vernaza *et al.*, 2012). On the other hand, *swn* mutants only show weak changes during vegetative phase transition (Xu *et al.*, 2016a, b). In addition, *clf swn* double mutants lose the capacity to differentiate and form massive somatic embryo-like structures (Chanvivattana *et al.*, 2004; Farrona *et al.*, 2011a; Lu *et al.*, 2011).

CLF and SWN share five conserved domains with the *Drosophila* E(z), including EZD1 (E(z) domain 1), EZD2 (E(z) domain 2), SANT (SWI3, ADA2, N-CoR, and TFIIIB), CXC (cysteine-rich region), and SET (Su(var)3-9, E(z), Trithorax) (Figure 2; Springer *et al.*, 2003; Ng *et al.*, 2007). The function of EZD1 is unknown, while EZD2 was shown to mediate the interaction of plant E(z) homologs with the VEFS (VRN2, EMF2, FIS2, and Su(z)12) domain of Su(z)12 homologs in *Arabidopsis* (Chanvivattana *et al.*, 2004). The SANT domain is involved in regulating chromatin structure and histone binding (Yu *et al.*, 2003; Tsai and Fondell, 2004). In *Drosophila*, mutations in the CXC domain of E(z) lead to loss-of-function phenotypes and genome-wide reduction of H3K27me3 (Ketel *et al.*, 2005). It was speculated that BLISTER (BLI) interacts with the CXC domain of *Arabidopsis* CLF, thereby modulating CLF enzymatic activity (Schatlowski *et al.*, 2010). In addition, the CXC domain of CLF may be involved in the recruitment and stabilization of the PRC2 complex (Krajewski *et al.*, 2005). The SET domain possesses methyl catalytic activity, which is highly conserved in higher eukaryotes (Dillon *et al.*, 2005; Herz *et al.*, 2013; Jiao and Liu, 2015).

**Figure 2** Domain architecture of CLF and SWN proteins.

EZD: E(z) domain; SANT: SWI3, ADA2, N-CoR, and TFIIB; CXC: cysteine-rich region; SET: Su(var)3-9, E(z), Trithorax. Scale bar: 100 aa. Figure redrawn based on Springer *et al.* (2003).



### 1.3 Recruitment of PcG components to targets

Although PcG components are highly conserved in higher eukaryotes, the mechanisms underlying PcG recruitment are remarkably different. In *Drosophila*, the Polycomb response elements (PREs) work as major docking sites for PcG recruitment through DNA-binding factors (Simon *et al.*, 1993; Müller and Kassis, 2006; van Kruijsbergen *et al.*, 2015). In mammals, it was first found that the genome-wide occupancy of PcG components has a large overlap with CpG islands (CGIs) (Ku *et al.*, 2008). It was also shown that the CGIs are sufficient for PcG recruitment and stable repression (Lynch *et al.*, 2012; Riising *et al.*, 2014). This gave rise to the idea that CGIs are PRE-like regulatory elements in mammals. In addition, it has been shown that RNAs also play essential roles in mediating PcG recruitment. For example, it was proposed that the non-coding RNA (ncRNA) *XIST-RepA* drives PRC2 recruitment to the X chromosome during X inactivation (Plath *et al.*, 2003; Kohlmaier *et al.*, 2004). However, *RepA* is dispensable for PRC2 occupancy (da Rocha *et al.*, 2014). So far, it is still uncertain whether ncRNAs directly recruit PRC2 to targets (Laugesen *et al.*, 2019).

In *Arabidopsis*, several PREs have been identified that are required for the recruitment of PRC2 to specific genes. For instance, a 50-bp-long *cis*-acting sequence, located at the promoter region of the seed regulator *LEAFY COTYLEDON 2* (*LEC2*), was shown to be required for repressing transcription and initiating H3K27me3 deposition at this locus (Berger *et al.*, 2011). The ASYMMETRIC LEAVES 1 and 2 (AS1-AS2) complex physically interacts with PRC2 and recruits PRC2 to the homeobox genes *BREVIPEDICELLUS* (*BP*) and *KNOTTED-LIKE FROM ARABIDOPSIS THALIANA 2* (*KNAT2*) in differentiating leaves (Lodha *et al.*, 2013). The PRE-like sequence in the promoter region of *KNUCKLES* (*KNU*) can be recognized by PRC2 and AG, and AG displaces PRC2 from *KNU* (Sun *et al.*, 2014). In addition to these single locus-based findings, a few consensus motifs, such as the GAGA-like and *Telo*-box-like motifs, were reported to be involved in the recruitment of PcG proteins to target loci in *Arabidopsis* (Deng *et al.*, 2013; Xiao *et al.*, 2017; Zhou *et al.*, 2018). The PRE-based GA-repeats that are recognized by GAGA factors (GAFs) were reported earlier in *Drosophila* (Farkas *et al.*, 1994; Katsani *et al.*, 1999; Hodgson *et al.*, 2001). The *Telo*-box motif is a short

sequence identical to telomere repeat units (AAACCCTA)<sub>n</sub>, which was originally observed in the 5' flanking regions of genes encoding ribosomal proteins and the translation elongation factor EF1 $\alpha$  (Gaspin *et al.*, 2010). Furthermore, there are two examples showing that H3K27me<sub>3</sub> deposition depends on ncRNAs in *Arabidopsis*. First, it was reported that the *COLD ASSISTED INTRONIC NONCODING RNA (COLDAIR)*, a long intronic noncoding RNA, interacts with PRC2 and is required for the vernalization-mediated repression of *FLOWERING LOCUS C (FLC)* (Heo and Sung, 2011). Second, the PcG protein LHP1 targets *AUXIN REGULATED PROMOTER LOOP RNA (APOLO)* to regulate the gene *PINOID (PID)* (Ariel *et al.*, 2014).

#### **1.4 PcG proteins play essential roles in *Arabidopsis* developmental transitions**

*Arabidopsis* undergoes major important phase transitions during its life cycle, from gametophyte to sporophyte, embryo to vegetative, and vegetative to reproductive stages. Each phase transition is controlled by distinct endogenous and environmental cues (Poethig, 2003; Bäurle and Dean, 2006; Huijser and Schmid, 2011). In particular, PcG proteins play important roles in these phase transitions (Xiao and Wagner, 2015). In the following section, I will focus on the floral transition from vegetative to reproductive phase, which is the most relevant phase change to this thesis.

##### **1.4.1 The floral transition**

The floral transition, which results in flowering, is precisely controlled for maximum reproductive success and optimal seed sets. Correct timing for flowering is important as it can affect plant fitness. The transition to flowering is controlled by a complex network that integrates signals from endogenous and environmental cues (Araki, 2001; Simpson and Dean, 2002; Bäurle and Dean, 2006; Huijser and Schmid, 2011). To date, five genetic pathways have been identified to regulate flowering in *Arabidopsis* (Fornara *et al.*, 2010; Srikanth and Schmid, 2011).

The first pathway is known as vernalization, which accelerates the flowering process by a long period of cold treatment. In previous studies, genetic analysis has revealed that the dominant alleles of *FRIGIDA (FRI)* confer a vernalization requirement in *Arabidopsis*

(Napp-Zinn, 1987; Clarke and Dean, 1994). Further studies proved that *FLC*, a floral repressor, also plays important roles in vernalization, and *FRI* up-regulates the expression of *FLC* (Michaels and Amasino, 1999; Geraldo *et al.*, 2009). *FLC* was shown to repress the expression of *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)* and *FLOWER LOCUS T (FT)* by directly binding to CArG boxes on their respective chromatin (Helliwell *et al.*, 2006). *VERNALIZATION 1 (VRN1)* and *VRN2* were found to be similar to the *Drosophila* PcG protein Su(z)12, and are required to maintain the repressed state of *FLC* epigenetically (Gendall *et al.*, 2001; Levy *et al.*, 2002). The second pathway is the photoperiod pathway, which refers to the regulation of flowering in response to day length and quality of light perceived (Simpson and Dean, 2002; Srikanth and Schmid, 2011). *CONSTANS (CO)* works as the key regulator, integrating internal and external signals into the photoperiod pathway (Shim *et al.*, 2017). The third pathway is the gibberellin pathway, which refers to the requirement of one hormone, gibberellic acid (GA), for flowering. In *Arabidopsis*, GA promotes floral transition through the activation of floral integrators such as *SOC1*, *LEAFY (LFY)*, and *FT* (Mutasa-Göttgens and Hedden, 2009; Davière and Achard, 2013). The fourth pathway is the autonomous pathway, which comprises endogenous factors promoting flowering independent of photoperiod and GA pathways. In general, the endogenous factors accelerate flowering indirectly by repressing the expression of *FLC* (Simpson, 2004). Lastly, the fifth pathway, namely the aging pathway, refers to the endogenous pathway which contributes to the control of flowering time (Wang *et al.*, 2009). Taken together, these pathways function as a cross-talking network to promote precise plant flowering (Boss *et al.*, 2004; Srikanth and Schmid, 2011).

#### **1.4.2 CLF plays critical roles in floral transition**

A major player in floral transition in *Arabidopsis* is *FLC*, a MADS-box TF that represses the vegetative to reproductive phase transition (Michaels and Amasino, 1999; Berry and Dean, 2015; Whittaker and Dean, 2017). The expression of *FLC* is under the control of various activators and repressors. For example, *FLC* is constitutively repressed by the activators in the autonomous pathway such as *FLOWERING LOCUS VE (FVE)*, *FLOWERING CONTROL LOCUS A (FCA)*, and *FLOWERING LOCUS D (FLD)* (Macknight *et al.*, 1997; He *et al.*, 2003; Ausin *et al.*, 2004; Kim *et al.*, 2004).



The vernalization pathway represses *FLC* expression to accelerate flowering in response to prolonged cold treatment (Sheldon *et al.*, 2000). *FLC* prevents flowering partially by repressing the key floral integrator *FT* (Searle *et al.*, 2006). Initially described as a component of the photoperiod pathway, *FT* is activated by *CO* through day-length perception (Kardailsky *et al.*, 1999; Suarez-Lopez *et al.*, 2001). *FT* is expressed in the leaf vascular tissue, where the protein is further translocated from the phloem to SAMs to induce flowering (Takada and Goto, 2003; Corbesier *et al.*, 2007). In summary, *FT* functions as a floral integrator that integrates signals from the vernalization and photoperiod pathways to accelerate flowering in *Arabidopsis*.

Studies revealed the roles of CLF and other regulators at the *FLC* and *FT* loci to control floral transition. For example, it was reported that CLF directly mediates the deposition of H3K27me3 on the *FLC* locus during vegetative development (Jiang *et al.*, 2008). On the other hand, CLF also strongly represses the expression of *FT* to delay flowering during vegetative development (Farrona *et al.*, 2011b; Lopez-Vernaza *et al.*, 2012; Müller-Xing *et al.*, 2014). A more recent study showed that, at the *FT* locus, the binding of nuclear factor-Y C (NF-YC) antagonizes the CLF-dependent deposition of H3K27me3, leading to the de-repression of *FT* and floral transition under long-day conditions (Liu *et al.*, 2018). In addition, it was shown that the floral activator *AGAMOUS-LIKE 19* (*AGL19*) is repressed by CLF and the de-repression of *AGL19* in *clf* is partially responsible for the early flowering phenotype of *clf* mutants (Schönrock *et al.*, 2006).

*AGAMOUS-LIKE 17* (*AGL17*), a MADS-box gene, was reported to be preferentially expressed in roots (Rounsley *et al.*, 1995). A later study reported that *AGL17* transcripts can be detected in various plant tissues with the highest expression in roots (Han *et al.*, 2008). Although *AGL17* is positively regulated by the photoperiod pathway regulator *CO*, it does not appear to be affected by *FT*. Moreover, loss-of-function of *AGL17* delays flowering, while over-expression of *AGL17* promotes flowering (Han *et al.*, 2008). It was concluded that *AGL17* acts in an *FT*-independent photoperiod pathway to promote flowering in *Arabidopsis* (Han *et al.*, 2008).

## 1.5 SPLAYED, an SWI/SNF-type chromatin remodeler

Chromatin remodeling plays crucial roles in the regulation of gene expression through dynamic modifications of chromatin architecture. There are two kinds of key chromatin remodeling players. One involves enzymes that specifically add or remove covalent modifications on DNA and histone residues (Li *et al.*, 2007; Jones, 2012), such as CLF and SWN described in section 1.2. The other player, such as SWItch/Sucrose Non-Fermentable (SWI/SNF) complex which is the focus of this section, utilizes the energy derived from adenosine triphosphate (ATP) hydrolysis to alter DNA-histone interactions (Clapier and Cairns, 2009; Hargreaves and Crabtree, 2011; Han *et al.*, 2015).

The SWI/SNF ATP-dependent chromatin remodeling complex, initially identified in *Saccharomyces cerevisiae*, has been extensively studied (Neugeborn and Carlson, 1984; Peterson and Herskowitz, 1992; Saha *et al.*, 2006; Wang *et al.*, 2007; Clapier and Cairns, 2009; Kadoch and Crabtree, 2015; Mashtalir *et al.*, 2018; Barisic *et al.*, 2019; Pan *et al.*, 2019). The SWI/SNF complex is evolutionally conserved from yeast to mammals and plants (Martens and Winston, 2003; Sif, 2004; Flaus *et al.*, 2006; Euskirchen *et al.*, 2012). A number of SWI/SNF homologs have been identified in *Arabidopsis* (Verbsky and Richards, 2001; Knizewski *et al.*, 2008), which include four SWI/SNF ATP-driven chromatin remodelers: SPLAYED (SYD), BRAHMA (BRM), MINUSCULE 1 (MINU1), and MINU2 (Wagner and Meyerowitz, 2002; Farrona *et al.*, 2004; Mlynárová *et al.*, 2007; Tang *et al.*, 2008; Sang *et al.*, 2012).

The *Arabidopsis* SWI/SNF ATPase SYD plays pivotal roles in the regulation of several aspects of development, including regulation of floral homeotic gene expression and stem cell pool maintenance (Wagner and Meyerowitz, 2002; Kwon *et al.*, 2005, 2006). Moreover, within biotic stress signaling networks, SYD acts in specific stress pathways, e.g., negatively regulating SUPPRESSOR OF NPR1, CONSTITUTIVE 1 (SNC1)-mediated immunity (Walley *et al.*, 2008; Johnson *et al.*, 2015). The *Arabidopsis* SYD and BRM share several conserved domains, of which the ATPase domain is the most conserved with nearly 50% identity (Wagner and Meyerowitz, 2002; Farrona *et al.*, 2004). Unlike BRM that harbors a C-terminal bromodomain, SYD contains a large unique C-terminal

extension with two repetitive regions (Farrona *et al.*, 2004; Su *et al.*, 2006). However, even an SYD truncated version lacking the C-terminal extension is able to rescue the *syd* null mutant (Su *et al.*, 2006). SYD and BRM have different and redundant functions in the regulation of plant development, as their null mutants exhibit similar but distinct defects (Bezhani *et al.*, 2007). For example, both *syd* and *brm* mutants are dwarfed, slow-growing, and have curling leaves (Wagner and Meyerowitz, 2002; Farrona *et al.*, 2004; Hurtado *et al.*, 2006). BRM promotes vegetative development by antagonizing PcG at key loci such as *SHORT VEGETATIVE PHASE* (*SVP*), a flowering repressor (Li *et al.*, 2015). However, it is not clear whether SYD functions similarly to BRM in antagonizing PcG repression. Recently, the genome-wide occupancy of BRM has been profiled (Li *et al.*, 2016; Archacki *et al.*, 2017), which aids in deciphering BRM's important roles in plant developmental processes. Although a few direct targets of SYD have already been identified by chromatin immunoprecipitation combined with quantitative PCR (ChIP-qPCR) during the past decade (Kwon *et al.*, 2005; Walley *et al.*, 2008; Wu *et al.*, 2015), the global occupancy of SYD still needs to be established.

## 1.6 PcG and SWI/SNF function antagonistically or synergistically

SWI/SNF-type chromatin remodelers play important roles in plant development through either activating or repressing transcription (Kwon *et al.*, 2006; Reyes, 2014; Li *et al.*, 2015; Yang *et al.*, 2015; Archacki *et al.*, 2017). Increasing data have suggested that *Arabidopsis* SWI/SNF family can act antagonistically or synergistically with PcG (Reyes, 2014; Paz Sanchez *et al.*, 2015; Pu and Sung, 2015). For example, in leaves, BRM represses seed maturation genes seemingly through cooperating with PcG (Tang *et al.*, 2008). It has been shown that SYD and BRM function opposite with CLF at *AG* and *AP3* in developing flower primordia (Wu *et al.*, 2012). In *brm*, *FLC* expression is up-regulated and accompanied by the reduction of H3K27me3 (Farrona *et al.*, 2011a). Furthermore, prior work from our laboratory has shown that BRM can both antagonize and cooperate with PcG at several loci during plant development (Li *et al.*, 2015). Although these studies emphasize the interplay between PcG and SWI/SNF at some loci, the genome-wide interplay between these two complexes remains to be uncovered.

## 1.7 Thesis objectives

During the past decades, numerous works, based largely on mutant examination and single-gene analysis, have clearly demonstrated the importance of PcG and SWI/SNF in plant growth and development. Even though a number of H3K27me3 genome-wide profiles have been recently reported (Luo *et al.*, 2013; Li *et al.*, 2015, 2016; Cui *et al.*, 2016; Wang *et al.*, 2016; Yang *et al.*, 2016; Carter *et al.*, 2018), studies on genome-wide occupancy of CLF, SWN, and SYD proteins are still rather limited. In addition, the interplay between PcG and SWI/SNF is poorly understood. To address these questions, the following objectives were set for this thesis:

1. To profile the genome-wide occupancy of CLF and SWN.
2. To perform a genome-wide comparison of H3K27me3 profiles between wild-type and *clf*, *swn*, and *clf swn* mutants.
3. To examine the effects of *CLF* and *SWN* on global transcription.
4. To identify SYD direct targets genome-wide and compare them to the targets of BRM.
5. To investigate the interplay between PcG and SWI/SNF.

## 2 MATERIAL AND METHODS

### 2.1 Plant material and growth conditions

*Arabidopsis* WT and mutants, *clf-29* (SALK\_021003), *swn-4* (SALK\_109121), *agl17-3* (SALK\_024428), and *syd-5* (SALK\_023209) used in this thesis, were all in Columbia (Col-0) background, and have been previously described (Bouveret *et al.*, 2006; Wang *et al.*, 2006; Han *et al.*, 2008; Walley *et al.*, 2008). Transfer DNA (T-DNA) insertion mutants were obtained from the *Arabidopsis* Biological Resource Center (ABRC).

All seeds were sterilized in 20% bleach with 0.1% sodium dodecyl sulfate (SDS) for 15 min with gentle shaking at room temperature. Sterilized seeds were washed five times with sterilized double-distilled water (ddH<sub>2</sub>O) and stratified for 3 days at 4°C in darkness before sowing in soil (ProMix-BX, Premier Horticulture, Quebec) or on agar plates containing 4.3 g/L Murashige and Skoog (MS) nutrient mix (Sigma-Aldrich), 1.5% sucrose (pH 5.75), and 0.8% agar. Plants were grown in either a growth room or cabinet under long-day conditions (16 h light/8 h dark) at 22°C.

Homozygous T-DNA insertion mutants were identified by polymerase chain reaction (PCR)-based genotyping as described in section 2.5. Two-week-old plants, grown on media under long-day conditions, were used for all analysis unless specified otherwise.

### 2.2 Crossing of *Arabidopsis* plants

To cross *Arabidopsis* plants from different genetic traits, 5- to 6-week-old healthy plants were chosen as pollen recipients. First, flower buds with newly emerging and barely visible petals were selected, fine forceps were used to open the flower buds to remove their sepals, petals, and stamens. Great care was taken not to injure the pistils. Then, healthy pollen-donor flowers were cut from the flowering stalk and the petals and sepals were removed by using fine forceps. The anthers brimming with pollen were used to pollinate the stigma of the prepared pistils. Next, the pollinated pistils were marked with a small piece of tape describing the genetic backgrounds of the pollen-donors and recipients.

### 2.3 Generation of plasmid constructs for plant transformations

Three bacterial artificial chromosome (BAC) clones F26B6, T10M13, and T3B23, obtained from ABRC, harbor the *CLF*, *SWN*, and *SYD* loci, respectively. The BAC plasmids were isolated using QIAGEN Large-Construct Kit and served as templates to amplify the *CLF*, *SWN*, and *SYD* genomic sequences including the promoter/regulatory regions. The PCR products were purified using the Geneaid GenepHlow Gel/PCR Kit and subcloned into the binary vector *pMDC107* (Curtis and Grossniklaus, 2003), between the *PmeI* and *AscI* sites, to obtain the following constructs: *pMDC107-gCLF*, *pMDC107-gSWN*, and *pMDC107-gSYDΔC*. After subsequent sequencing to confirm the in-frame coding sequences for *CLF*, *SWN*, and *SYD*, the constructs were then introduced, by electroporation (den Dulk-Ras and Hooykaas, 1995), into *Agrobacterium tumefaciens* strain *GV3101*, which were then used to transform into the mutant plants using the floral dip method (Zhang *et al.*, 2006). Homozygous transgenic lines were selected from T3 generations, in which the functional GFP-tagged proteins were detected.

### 2.4 Western blot

Two-week-old seedlings (~2 g) were collected, and nuclei were isolated according to the ChIP protocol (Gendrel *et al.*, 2005) except for the tissue fixation step. The nuclear protein was released by dissolving the nuclei mixture in 300 µl of lysis buffer (50 mM Tris-HCl, 10 mM ethylenediaminetetraacetic acid (EDTA), 1% SDS, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), and 1× cComplete™ Protease Inhibitor Cocktail (Roche)) and then sonicated. The protein solution was centrifuged at 16,000g for 10 min at 4°C to remove debris. To separate the proteins, the supernatant was loaded onto an 8% SDS polyacrylamide gel and size separated for 2 h at 120 Volts. The proteins were then transferred to a polyvinylidene fluoride (PVDF) membrane which was subsequently saturated in 5% non-fat milk/Tris-buffered saline with 0.1% Tween 20 (TBST) solution for 1 h with gentle shaking at room temperature. Incubation of the membrane in the primary antibody solution (anti-GFP: Abcam, ab290, 1:20,000 dilution; anti-H4: Millipore, 07-108, 1:20,000 dilution; and anti-Actin: Abcam, ab1801, 1:20,000 dilution) was carried out overnight at 4°C with gentle shaking. After washing the membrane five times with TBST

buffer to remove the excess primary antibody, the membrane was incubated in the secondary antibody (A0545, Sigma) solution for 1 h. The excess secondary antibody was removed by washing five times with TBST buffer. The protein bands were visualized by using the Western Blotting Reagents (RPN2106, GE Healthcare) and chemiluminescence was captured on X-ray films (Curix Ultra UV-G plus, Agfa) or using MicroChemi (DNR Bio-Imaging System).

## 2.5 PCR-based genotyping

PCR-based genotyping was employed to identify homozygous and heterozygous plants for all T-DNA insertion lines. Two-week-old leaves (~10 mg) were homogenized into fine powders and dissolved in 300  $\mu$ l cetyltrimethylammonium bromide (CTAB) extraction buffer (0.1 M Tris-HCl pH 8, 1.4 M NaCl, 0.02 M EDTA pH 8, 2% CTAB). The mixture was thoroughly vortexed and 270  $\mu$ l of chloroform were added. The tube was vortexed for 30 s and centrifuged at 11,000g for 8 min. The top phase (~220  $\mu$ l) was transferred to a new tube and 250  $\mu$ l of isopropanol was added. The tube was gently inverted five times and centrifuged at 16,000g for 10 min. The aqueous phase was discarded and 200  $\mu$ l of 70% ethanol was added and centrifuged at 10,000g for 5 min. Ethanol was removed and the open-tube was kept under vacuum for 10 min at room temperature to completely dry the pellet, which was finally resuspended in 20  $\mu$ l of ddH<sub>2</sub>O.

For a 20  $\mu$ l of PCR reaction, the following components were added into 8-strip tubes: 10  $\mu$ l of 2x Taq FroggaMix (FBTAQM, FroggaBio), 0.5  $\mu$ l of 10  $\mu$ M each of forward and reverse primers, 2  $\mu$ l DNA template, and 7  $\mu$ l ddH<sub>2</sub>O. The PCR reaction was carried out using a thermocycler, set for 3 min at 94°C, followed by 30 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 1 min, followed by a final extension phase of 72°C for 3 min.

To identify the T-DNA lines, three types of primers were used: RP, LP, and LB, which are specific to the 3' and 5' genomic sequences and the left border of the T-DNA insertion, respectively. For WT plants, a single band should be produced in the RP+LP reaction, but not in the RP+LB reaction. For hemizygous T-DNA insertion plants, a band should be observed in both RP+LP and RP+LB reactions. For homozygous T-DNA insertion plants, a band should be visible in the RP+LB reaction, but not in the RP+LP reaction. Primers for

genotyping T-DNA mutants were designed using the website: <http://signal.salk.edu/tdnaprimers.2.html>, and are listed in Appendix A.

## 2.6 Flowering time measurement

WT, mutant, and T-DNA transgenic plants were grown side by side in soil. The number of rosette leaves was counted when the inflorescence stem grew to 1 cm in length. For each genotype, at least 10 plants were analyzed for each of the three biological replicates.

## 2.7 Chromatin immunoprecipitation (ChIP)

ChIP experiments were carried out as described (Gendrel *et al.*, 2005; Li *et al.*, 2016; Chen *et al.*, 2017) with minor modifications. Typically, 5 g of plant material (two-week-old seedlings or four-week-old *clf swn* double mutants) grown on media were harvested and cross-linked with 1% formaldehyde for 20 min under vacuum and ground in liquid nitrogen into fine powders. The chromatin was isolated and sheared into 200-800 bp fragments by sonicating five times for 15 s each at 15% power (Fisherbrand Model 120). The sonicated chromatin was incubated with anti-GFP (ab290, Abcam), anti-H3K27me3 (07-449, Millipore), or anti-H3 (ab1791, Abcam) antibody overnight at 4°C with gentle rotating. The precipitated DNA was recovered using the MinElute PCR Purification Kit (Cat.# 28004, Qiagen) according to the manufacturer's instruction. The ChIP DNA was used for Illumina single-end (1 × 50 bp) sequencing or qPCR. Three biological replicates of ChIP-qPCR were performed and the results were presented as percentage of input-DNA according to the Champion ChIP-qPCR user manual (SABioscience). The primers used for qPCR are listed in Appendix A. DNA quantity and quality were checked using a Qubit fluorometer (ThermoFisher Scientific).

## 2.8 ChIP DNA library preparation

At least 2 ng of each ChIP DNA were used to construct libraries for ChIP-seq, and two biological replicates were carried out for each ChIP DNA. End repair, adapter ligation, PCR amplification, and cleanup of PCR reaction were performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (E7645S, NEB) according to the manufacturer's



protocol. Each DNA library was then checked using Agilent High Sensitivity DNA Kit (5067-4626, Agilent Technologies) following the manufacturer's instruction.

## 2.9 ChIP-seq and data analysis

HiSeq 2500 platform was used for high-throughput sequencing of the libraries. Sequence data were analyzed as previously described (Li *et al.*, 2015, 2016; Chen *et al.*, 2017). Briefly, after sequencing, the raw sequence reads were cleaned by removing the bases with low-quality score and cutting sequencing adapter followed by filtering out short reads. The cleaned sequence reads were then mapped to the *Arabidopsis* genome (TAIR10) (Lamesch *et al.*, 2012) by Bowtie mapper (Langmead *et al.*, 2009) with default mismatch parameters, and only reads that can be mapped uniquely to the genome were retained for further analysis. The total number of ChIP-seq reads for all experiments, in this thesis, is listed in Appendix B. To identify reads enriched regions (peaks), MACS 1.4 (Zhang *et al.*, 2008) was employed to perform peak calling with default settings. High-confidence target regions were defined as the strict overlap of the MACS peaks from two biological replicates. The data were imported to the Integrated Genome Browser (IGB) (Nicol *et al.*, 2009) for visualization.

Peak analysis was performed using PeakAnalyzer (Salmon-Divon *et al.*, 2010). PeakAnnotator was employed to identify functional elements proximal to peak loci through the Nearest Downstream Gene (NDG) subroutine. The target genes of each peak were defined as those genes closest to a given peak localized around the gene body (from 1 kb upstream of transcription start sites (TSSs) to transcription termination sites (TTSs)). The gene annotation file was downloaded from the EnsemblPlants homepage (<https://plants.ensembl.org/index.html>). SICER program (window size = 600, gap size = 200, false discovery rate (FDR) = 1%) (Zang *et al.*, 2009) was used to quantitatively compare H3K27me3 levels of WT and mutants. Peaks with at least a two-fold change were kept for further analysis.

To identify DNA motifs enriched at CLF- and SWN-occupied peaks, 300 bp surrounding each peak summit (150 bp upstream and downstream) were extracted and searched for

enriched DNA motifs using the DREME/MEME software suite (<http://meme-suite.org/tools/meme-chip>) (Machanick and Bailey, 2011) with default settings.

The distribution of peaks was performed by ChIPseek, a web-based analysis tool (<http://chipseek.cgu.edu.tw>) (Chen *et al.*, 2014), with default settings. Briefly, the information of chromosome, start and end sites for each peak was fed to ChIPseek. Based on the annotation file from the UCSC TAIR10 assembly, the location of each peak was further placed into one of these categories: promoter-TSS (1 kb upstream of TSS), intergenic, exon, intron, 5' untranslated region (UTR), 3' UTR, and TTS. Heatmaps of ChIP-seq were generated using the computeMatrix and plotHeatmap utilities in deepTools2 (Ramirez *et al.*, 2016).

Functional annotation was performed using the online “AgriGo” Gene Ontology (GO) analysis toolkit (<http://bioinfo.cau.edu.cn/agriGO/>) with default settings.

## **2.10 RNA protocols**

### **2.10.1 RNA extraction**

Total RNA was isolated from ~50 mg of two-week-old WT and mutant material using RNeasy Plant Mini Kit (Qiagen) according to supplier's instruction. Total RNA was treated with DNase TURBO (Ambion) to remove the genomic DNA.

### **2.10.2 Gene expression analysis**

200 ng of RNA were reverse transcribed into cDNA by using High Capacity cDNA Reverse Transcript Kit (ThermoFisher Scientific). Real-time quantitative PCR (qPCR) was performed on the CFX96 Real-time OCR Instrument (Bio-Rad) using the SsoFast EvaGreen Supermix (Bio-Rad). Data of three biological replicates were analyzed by using the Bio-Rad CFX Manager 3.1 software. The expression levels were normalized to that of *GAPDH* (Czechowski *et al.*, 2005).

### 2.10.3 RNA-sequencing analysis

A total amount of 1.5 µg RNA per sample was used as input material for RNA-sequencing (RNA-seq) preparation. Sequencing libraries were generated using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample. Libraries of three independent biological replicates were sequenced on the Illumina HiSeq 4000 platform and 150 bp paired-end reads were generated. The number of RNA-seq reads for all of the experiments in this thesis is listed in Appendix C. Raw sequence reads were cleaned by removing bases with low-quality scores and by cutting sequencing adapters followed by removing short reads. Cleaned reads were mapped to the TAIR10 *Arabidopsis* genome using TopHat v2.0.4 with default settings, with the exception of a minimum and maximum intron lengths of 20 bp and 4000 bp, respectively. To identify differentially expressed genes, transcript assembly and calculations were carried out by the Cufflinks package (Trapnell *et al.*, 2012). Briefly, the alignment files, produced from running TopHat, were provided to Cufflinks to generate a transcriptome assembly for each sample. These assemblies were then merged together using the Cuffmerge utility to provide a uniform basis for calculating transcript levels. The reads and merged assemblies were fed to Cuffdiff, where transcript levels were calculated and the statistical significance of changes were determined. In this thesis, genes with at least a 1.5-fold (for *clf-29*, *swn-4*, *clf-29 swn-4*) or 2-fold (for *syd-5*) change in transcript levels (FDR = 5%,  $P < 0.05$ ) were considered to be differentially expressed.

### 2.11 Statistical analysis

Simple statistics such as means and standard deviations were calculated using Excel 2016 (Microsoft Corp., Redmond, Washington). The student *t*-test (two-tailed) or one-way ANOVA was employed for determining the significance of the difference between two or more than two independent datasets, respectively. A *P* value of 0.05 or less was used as a statistically significant difference.

## **2.12 Primer design**

Unless otherwise specified, all primers used in this thesis were designed using National Center for Biotechnology Information (NCBI) Primer-BLAST at <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>, to generate a PCR product of size 80-200 bp having primer melting temperatures of 58-61°C. The primers listed in Appendix A are shown in 5' to 3' direction.

## **2.13 Accession numbers**

ChIP-seq and RNA-seq data have been deposited in the NCBI GEO database under accession numbers GSE108960 and GSE126889.

### 3 RESULTS

#### 3.1 Genome-wide occupancy of H3K27 methyltransferases CLF and SWN in *Arabidopsis* seedlings

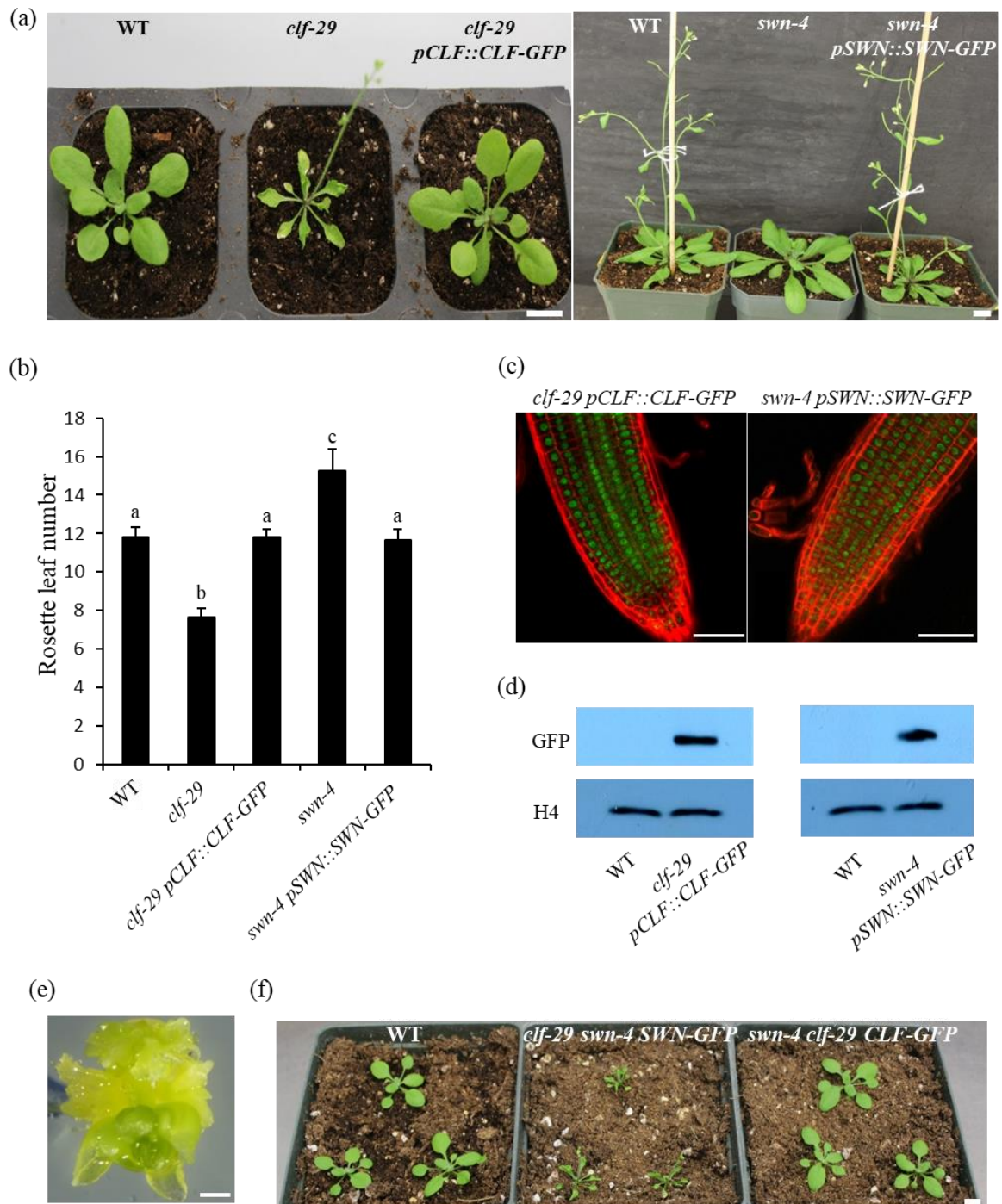
Epigenetic regulation of gene expression through histone modifications is fundamental for plant growth and development. In particular, the repressive histone modification H3K27me3 covers a large fraction of the *Arabidopsis* genome dynamically and plays critical roles in regulating plant development (Zhang *et al.*, 2007; Derkacheva and Hennig, 2014; Gan *et al.*, 2015; Li *et al.*, 2015; Wiles and Selker, 2017). However, the occupancy profiles of the two major H3K27 methyltransferases, CLF and SWN, have not yet been determined and compared genome-wide. In this section, the goal is to identify and compare genome-wide occupancies of CLF and SWN, as well as examine how these two methyltransferases deposit H3K27me3 in *Arabidopsis* seedlings.

##### 3.1.1 GFP-tagged CLF and SWN transgenes function *in vivo*

In preparation for profiling the genome-wide occupancy of CLF and SWN, GFP-tagged CLF and SWN under the control of their own promoters (named *pCLF::CLF-GFP* and *pSWN::SWN-GFP*) were transformed into *clf-29* and *swn-4* to generate *clf-29 pCLF::CLF-GFP* and *swn-4 pSWN::SWN-GFP* transgenic lines, respectively. As described in section 1.2, the *clf-29* mutants display severe defects, including dwarf stature, curly leaf, and early flowering (Schatlowski *et al.*, 2010; Farrona *et al.*, 2011b; Lopez-Vernaza *et al.*, 2012). The *swn-4* mutants flower later compared to WT (Xu *et al.*, 2016a and b). As shown in Figure 3a and b, *clf-29 pCLF::CLF-GFP* and *swn-4 pSWN::SWN-GFP* transgenic plants flowered at the same time as WT, indicating that the transgenes rescue the phenotypes of the *clf-29* and *swn-4* mutants, respectively. In addition, GFP signals in roots were observed by confocal microscopy (Figure 3c), and fusion proteins using an anti-GFP antibody were detected by Western blot (Figure 3d), demonstrating that CLF and SWN fusion proteins are expressed *in vivo*. To further confirm these two *pCLF::CLF-GFP* and *pSWN::SWN-GFP* transgenes are fully functional, *clf-29 pCLF::CLF-GFP* and *swn-4 pSWN::SWN-GFP* transgenic plants were individually crossed with *clf-29<sup>+/-</sup> swn-4<sup>-/-</sup>* plants (section 2.2).

**Figure 3** Transgenic *Arabidopsis* lines expressing GFP-tagged CLF or SWN.

(a) Plants showing complementation of the *clf-29* and *swn-4* flowering phenotypes by the *CLF*- and *SWN*-GFP fusion genes driven by their native promoters, *pCLF::CLF-GFP* and *pSWN::SWN-GFP*, respectively. Left: 4-week-old plants are shown, WT and *clf-29 pCLF::CLF-GFP* plants are not bolting, while the *clf-29* plant is already bolting. Right: 6-week-old plants are shown, WT and *swn-4 pSWN::SWN-GFP* plants are already bolting, while the *swn-4* plant is just starting to bolt. Scale bar: 1 cm. (b) Rosette leaf number at bolting of plants in different genetic backgrounds (WT, *clf-29*, *clf-29 pCLF::CLF-GFP*, *swn-4*, and *swn-4 pSWN::SWN-GFP*). Flowering time was recorded as the number of rosette leaves at bolting, thus late flowering plants have more rosette leaves. Error bars show standard deviation determined for at least 30 plants for each genetic background. Lowercase letters show significant differences among genetic backgrounds, as determined by Post-hoc Tukey's HSD test. (c) GFP signals detected by confocal microscopy in 4-day-old *clf-29 pCLF::CLF-GFP* and *swn-4 pSWN::SWN-GFP* roots. Propidium iodide was used to stain cell walls (red). Scale bar: 50  $\mu$ m. (d) Western blot analysis of nuclear extracts from two-week-old *clf-29 pCLF::CLF-GFP* and *swn-4 pSWN::SWN-GFP* seedlings. Antibodies used: anti-GFP and anti-histone H4. WT was used as a negative control and total histone H4 protein was used as a loading control. (e) 30-day-old *clf-29 swn-4* showing somatic embryo-like phenotypes. Scale bar: 1 mm. (f) The *CLF-GFP* and *SWN-GFP* transgenes rescue the *clf-29 swn-4* somatic embryo-like phenotypes. *clf-29 swn-4 CLF-GFP* and *clf-29 swn-4 SWN-GFP* plants show no somatic embryo-like phenotypes. Image of 3-week-old plants are taken. Scale bar: 1 cm.



The *clf-29 swn-4* double mutant forms massive somatic embryo-like structures which are lethal (Figure 3e). As expected, no embryo-like calli were observed in the F3 progeny of *clf-29 swn-4 pSWN::SWN-GFP* and *clf-29 swn-4 pCLF::CLF-GFP* (Figure 3f). Taken together, these data confirm that *GFP*-tagged *CLF* and *SWN* transgenes are functional *in vivo*.

### 3.1.2 Genome-wide occupancy of CLF and SWN

To profile CLF and SWN targets genome-wide, ChIP DNA was extracted from two-week-old *clf-29 pCLF::CLF-GFP* and *swn-4 pSWN::SWN-GFP* seedlings followed by sequencing (see section 2.7-2.9). Two independent biological replicates of ChIP DNA were sequenced. Strong correlations between these two replicates were observed (Pearson coefficient is 0.89 for CLF, and 0.92 for SWN; Figure 4). A total of 1,041 and 1,298 genomic regions (peaks) were occupied by CLF and SWN, respectively.

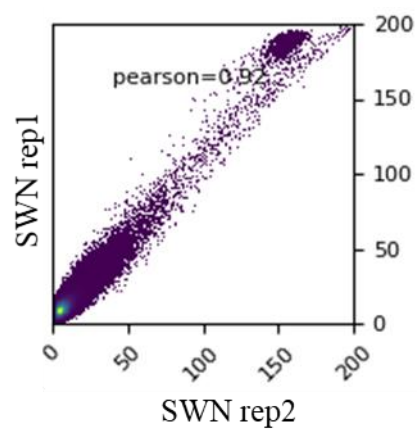
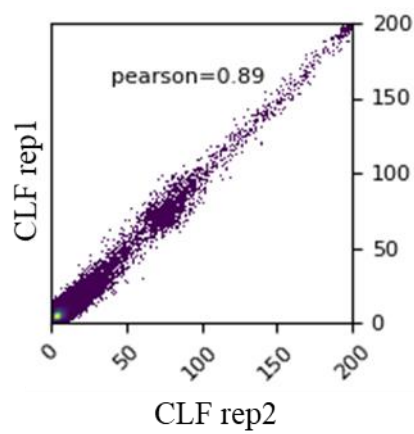
By applying a web-based analysis tool (ChIPseeker; see section 2.9), it was found that more than half of the CLF-occupied peaks are located in exons (39%) and promoter regions (24%), and about 16% in introns (Figure 5a). Similarly, 43% of the SWN-occupied peaks were found in exons, followed by 24% in promoter regions and 16% in introns (Figure 5a). Furthermore, it was shown that both CLF and SWN signals are highly enriched at TSSs and gradually decline towards the TTSs (Figure 5b).

The peaks of CLF and SWN were then assigned to corresponding genes and found that CLF and SWN target 1,391 and 1,877 genes, respectively. By comparing the genome-wide targets between CLF and SWN, it was shown that almost all of the CLF-occupied genes are also occupied by SWN (1,385 out of the 1,391 CLF targets). Only 6 genes are uniquely occupied by CLF, while 492 genes are occupied solely by SWN (Figure 5c). Then ChIP-qPCR was performed on several representative targets to validate the ChIP-seq results. It was shown that CLF and SWN co-target these genes by ChIP-qPCR, in agreement with the ChIP-seq results (Figures 6). Figure 7a shows representative genes that were targeted by SWN, and no CLF occupancy was detected in these genes. It was also confirmed by ChIP-qPCR (Figures 7b). These results show the quality and reliability of the ChIP-seq data.



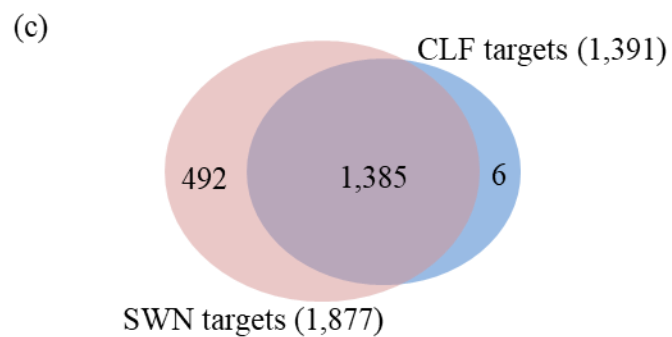
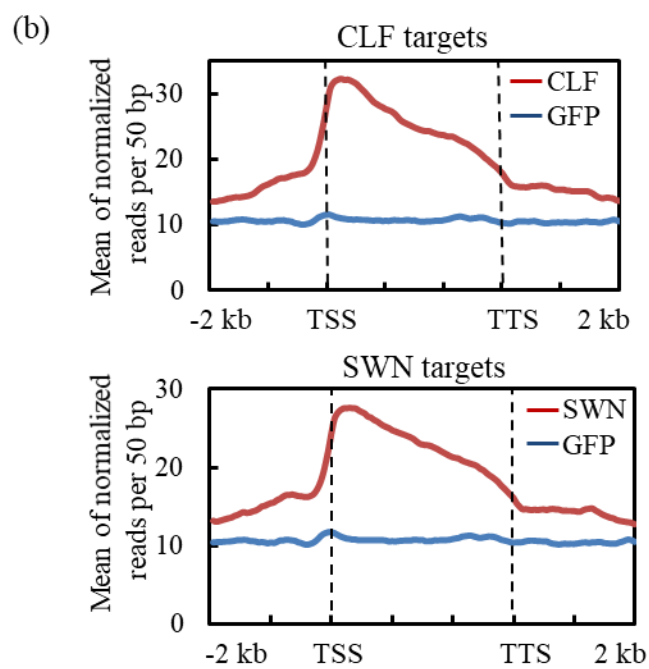
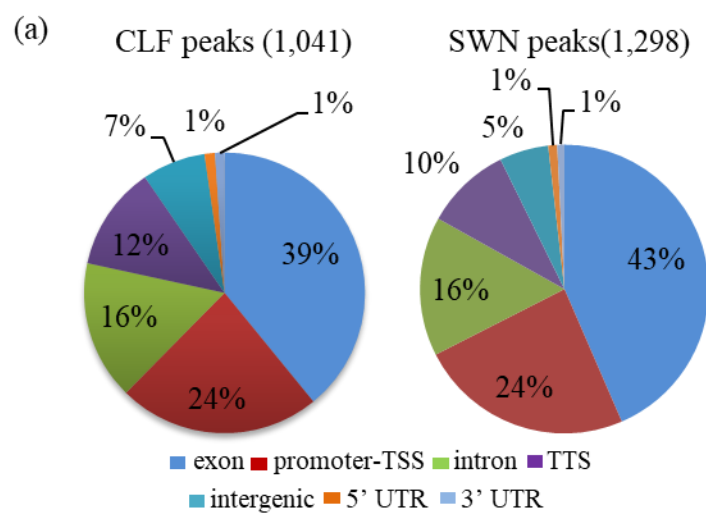
**Figure 4** Pearson correlation plots for two CLF and SWN ChIP-seq replicates.

The whole genome is divided into non-overlapping bins (50 bp), and normalized reads of each replicate (normalization is performed by randomly selecting the same number of reads from each replicate) are filled in each bin. The numbers of normalized reads in each bin are plotted. *X* and *Y* axis show the number of reads in each bin for each biological replicate. The correlation between CLF (*clf-29 pCLF::CLF-GFP*) and SWN (*swn-4 pSWN::SWN-GFP*) ChIP-seq replicates is very high (median  $r = 0.89$  and  $0.92$ , respectively).



**Figure 5** Genome-wide occupancy of CLF and SWN.

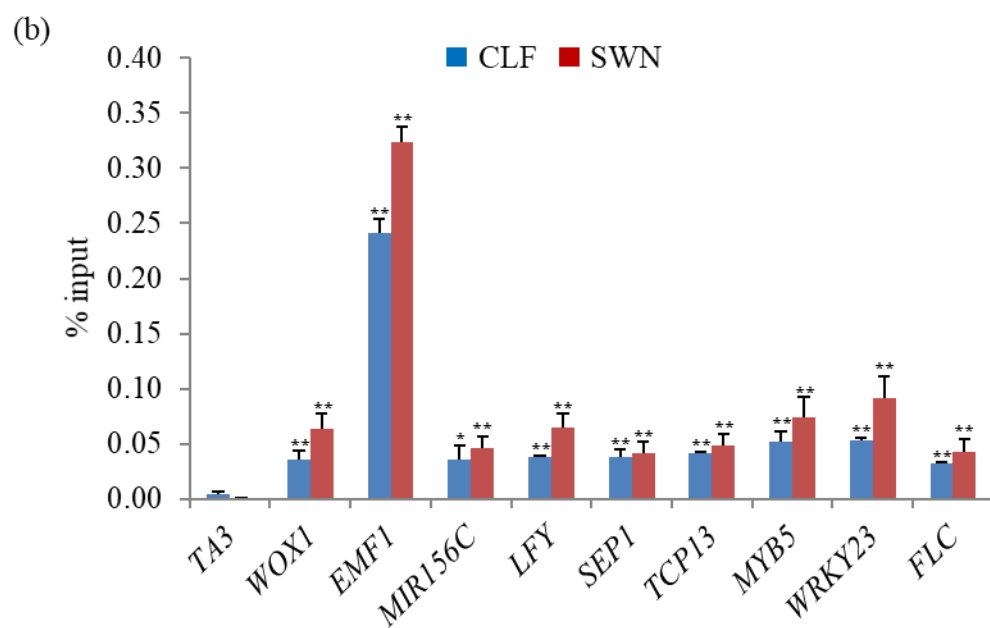
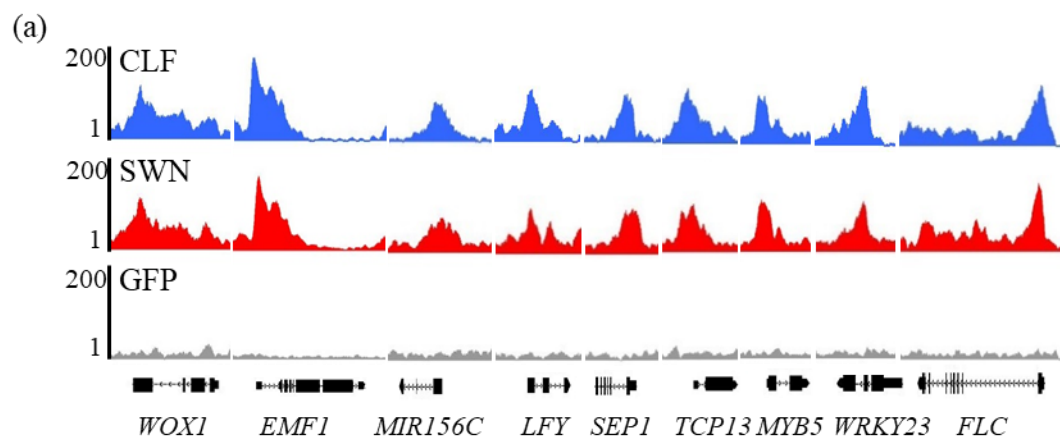
(a) Pie charts showing the distribution of CLF and SWN peaks at annotated genic and intergenic regions in the genome. (b) Mean density of CLF and SWN occupancy at all targets. GFP (p35S::GFP) was used as a negative control. The gene bodies (TSS to TTS) are scaled to 3 kb, 5' ends (-2 kb to TSS) and 3' ends (TTS to downstream 2 kb) are not scaled. (c) Venn diagram showing overlap between genes targeted by CLF and SWN.



**Figure 6** ChIP-seq signals at representative loci co-occupied by CLF and SWN.

(a) Genome browser views of CLF and SWN co-occupancy at representative genes. GFP (p35S::GFP) was used as a negative control. Gene structures are shown in black underneath the panel. (b) ChIP-qPCR validation. ChIP signals are shown as percentage of input. *TA3*, a transposable element gene, was used as a negative control locus. Error bars show standard deviations among three biological replicates. Student's *t*-test, \* $P < 0.05$ , \*\* $P < 0.01$ .

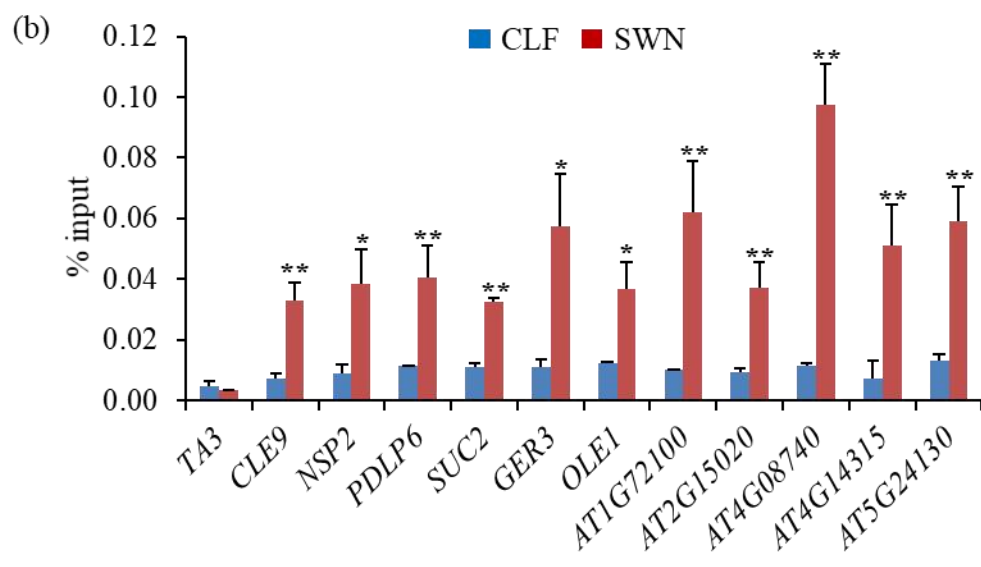
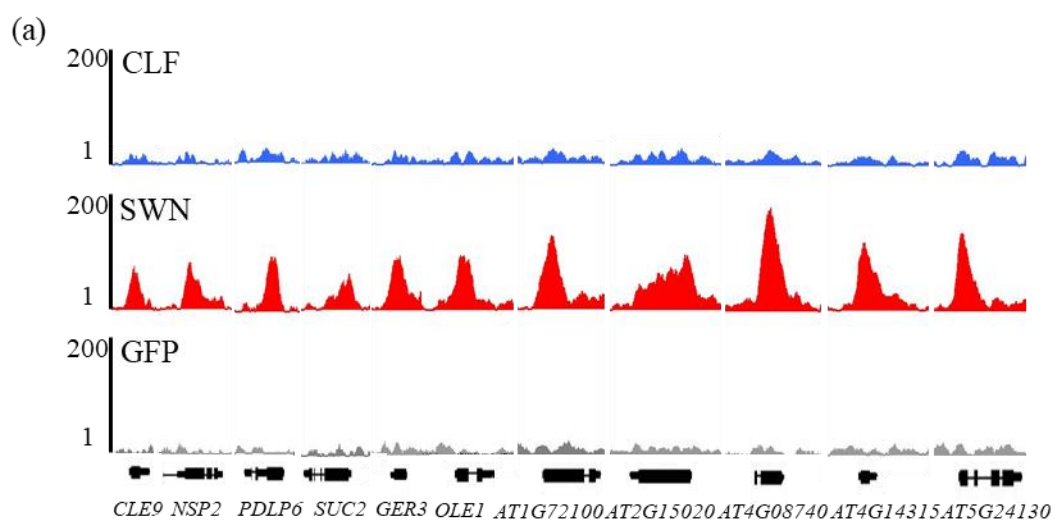
*WOX1*: *WUSCHEL RELATED HOMEODOMAIN 1*; *EMF1*: *EMBRYONIC FLOWER 1*; *MIR156C*: *MICRORNA 156C*; *LFY*: *LEAFY*; *SEP1*: *SEPALLATA 1*; *TCP13*: *TEOSINTE BRANCHED 1, CYCLOIDEA, and PCF 13*; *MYB5*: *MYB DOMAIN PROTEIN 5*; *WRKY23*: *WRKY DNA-BINDING PROTEIN 23*; *FLC*: *FLOWERING LOCUS C*.



**Figure 7** ChIP-seq signals at representative loci uniquely occupied by SWN.

(a) Genome browser views of SWN unique occupancy at representative genes. GFP (p35S::GFP) was used as a negative control. Gene structures are shown in black underneath the panel. (b) ChIP-qPCR validation. ChIP signals are shown as percentage of input. *TA3*, a transposable element gene, was used as a negative control locus. Error bars show standard deviations among three biological replicates. Student's *t*-test, \* $P < 0.05$ , \*\* $P < 0.01$ .

*CLE9*: *CLAVATA3/ESR-RELATED 9*; *NSP2*: *NITRILE SPECIFIER PROTEIN 2*; *PDLP6*: *PLASMODESMATA-LOCATED PROTEIN 6*; *SUC2*: *SUCROSE-PROTON SYMPORTER 2*; *GER3*: *GERMIN 3*; *OLE1*: *OLEOSIN 1*; AT1G72100, AT2G15020, AT4G08740, AT4G14315, and AT5G24130 are uncharacterized genes.





### 3.1.3 Identification of two DNA motifs from CLF and SWN peaks

To examine whether any particular DNA motifs are enriched, in the identified CLF and SWN peaks, stringent motif searches were performed (see section 2.9), which identified two DNA motifs for both CLF and SWN, a *GAGA*-like motif (*GA* repeats) and a *Telo*-box-like motif (*AAACCCTA*) (Figure 8). The *GAGA*-like motif was the most enriched motif as it was present in 24.6% of CLF peaks and 25.6% of SWN peaks with a high significance ( $E\text{-value} \leq 8.5 \times 10^{-113}$ ). Although not as highly enriched in CLF and SWN peaks (9.2% of CLF peaks; 6.5% of SWN peaks), the *Telo*-box-like motif was discovered with a high significance ( $E\text{-value} \leq 7.1 \times 10^{-67}$ ). The *Telo*-box-like motif has been recently shown to be recognized by the telomere-repeat factors (TRBs) which then recruit CLF/SWN through direct interaction with CLF/SWN (Zhou *et al.*, 2018). It is noteworthy to mention that the two motifs were also previously identified in the genomic targeting regions of other PcG proteins, such as FIE and LHP1 (Deng *et al.*, 2013; Molitor *et al.*, 2016; Zhou *et al.*, 2016; Xiao *et al.*, 2017).

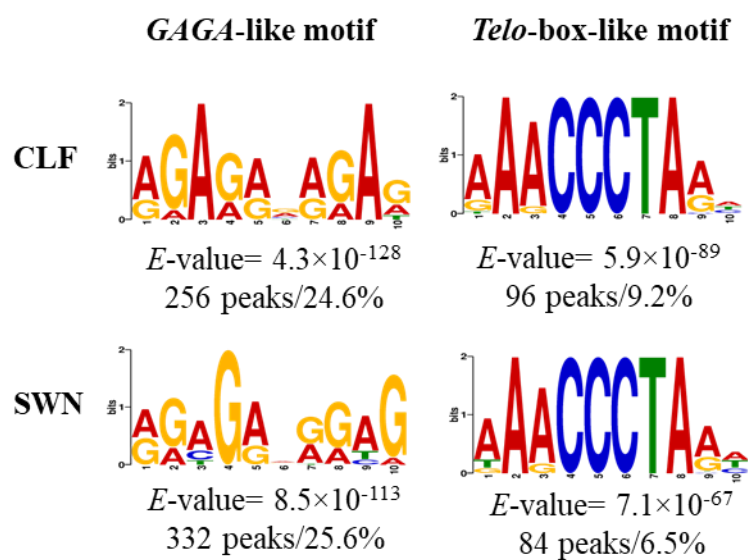
### 3.1.4 Functions of CLF and SWN targets in plant growth and development

To reveal the potential functions of CLF and SWN targets, Gene Ontology (GO) analysis was performed (see section 2.9) and it was found that CLF and SWN preferentially target genes that are involved in developmental pathways as well as abiotic and biotic stress responses (Figure 9a). Particularly, genes encoding members of several TF families, that are involved in developmental processes and stress responses, are over-represented in the CLF and SWN co-targets, such as the Homeobox, WRKY, and MADS-box families (Figure 9b). In contrast, SWN unique targets are highly involved in lipid localization/storage, cell wall modification, and post-embryonic development (Figure 9c).

Figure 10 shows ChIP-seq signals, at representative loci, that play important roles in plant growth and developmental processes. For example, major seed regulatory genes including *LEC2*, *FUSCA 3* (*FUS3*), and *ABSCISIC ACID INSENSITIVE 3* (*ABI3*) (To *et al.*, 2006; Tang *et al.*, 2008, 2012a and b) are co-occupied by CLF and SWN (Figure 10a), suggesting that CLF and SWN may be involved in suppression of the seed maturation program at

**Figure 8** Motifs enriched in CLF and SWN peaks.

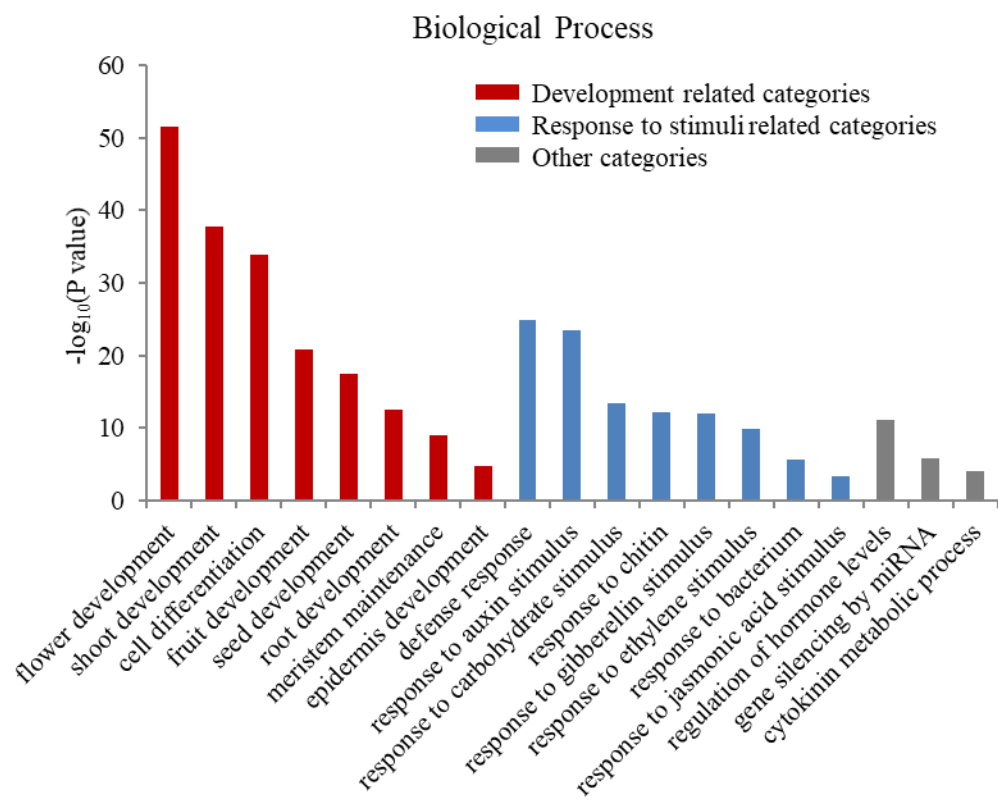
Number and percentage of peaks containing the motifs are shown. *E*-value indicates the estimated statistical significance of a motif. Motifs are generated from <http://meme-suite.org/tools/meme-chip>.



**Figure 9** Functional categorization of CLF and SWN targets.

(a) Gene Ontology (GO) analysis of CLF and SWN co-target genes. GO analysis was performed using the online “AgriGo” toolkit (<http://bioinfo.cau.edu.cn/agriGO/>) with default settings.  $P$  value determines the significance of the term.  $P < 0.05$  indicates the term is significant. (b) Table showing transcription factor (TF) gene families, co-targeted by CLF and SWN. TF gene families can be found at the *Arabidopsis* genome resources website (<http://arabidopsis.med.ohio-state.edu/AtTFDB/>). TCP: TEOSINTE BRANCHED 1, CYCLOIDEA, and PCF; MADS: MCM1, AGAMOUS, DEFICIENS, and SRF; bHLH: basic/helix-loop-helix; AP2-EREBP: APETALA 2 and ethylene-responsive element binding protein; NAC: NAM, ATAF, and CUC. (c) GO analysis of SWN unique targets. GO analysis was performed using the online “AgriGo” toolkit (<http://bioinfo.cau.edu.cn/agriGO/>) with default settings.  $P$  value determines the significance of the term.  $P < 0.05$  indicates the term is significant.

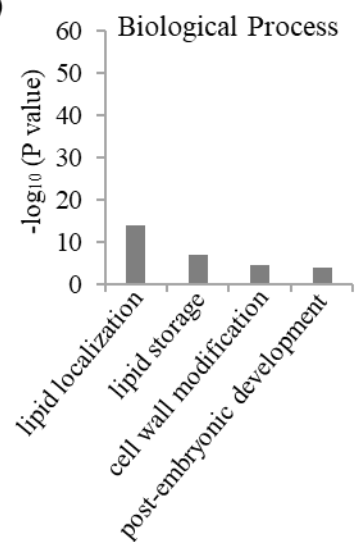
(a)



(b)

TF Families	Number of Genes in the Genome	Number/Percent of Genes Co-targeted by CLF/SWN
Total TFs	1,700	307/18.1%
Homeobox	91	31/34.1%
WRKY	72	20/27.8%
TCP	26	7/26.9%
MADS	109	28/25.7%
bHLH	161	36/22.4%
AP2-EREBP	138	28/20.3%
NAC	96	19/19.8%

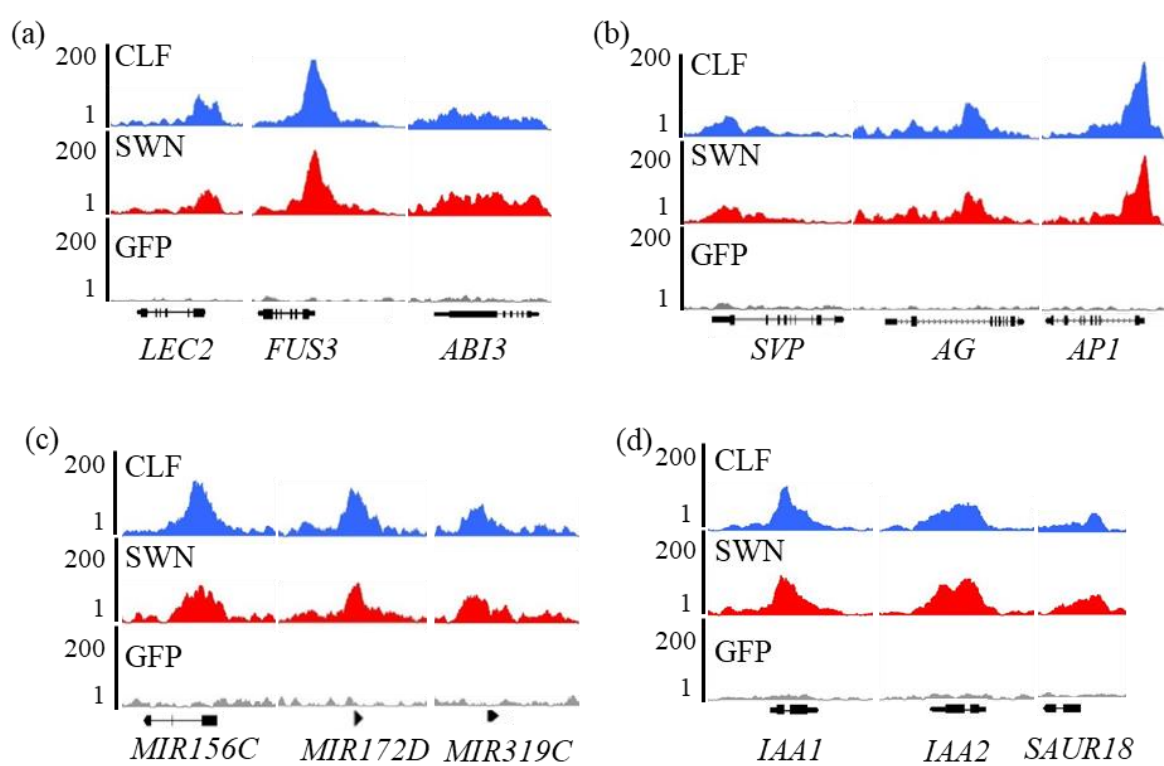
(c)



**Figure 10** Representative genes co-targeted by CLF and SWN.

(a-d) Genome browser views of CLF and SWN co-occupancy at representative genes. CLF occupancy shown in blue, SWN occupancy in red, and GFP (p35S::GFP), used as a negative control, in grey. Gene structures are shown in black underneath each panel.

*LEC2: LEAFY COTYLEDON 2; FUS3: FUSCA 3; ABI3: ABSCISIC ACID INSENSITIVE 3; SVP: SHORT VEGETATIVE PHASE; AG: AGAMOUS; API: APETALA 1; MIR156C: MICRORNA 156C; MIR172D: MICRORNA 172D; MIR319C: MICRORNA 319C; IAA1: INDOLE-3-ACETIC ACID INDUCIBLE 1; IAA2: INDOLE-3-ACETIC ACID INDUCIBLE 2; SAUR18: SMALL AUXIN UPREGULATED RNA 18.*



the seedling stage by directly targeting key maturation genes. Moreover, CLF and SWN also target several major floral transition genes including *SVP* and flower organ identity genes like *AG* and *APETALA 1 (API)* (Figure 10b; Hartmann *et al.*, 2000; Michaels *et al.*, 2003; Chiang *et al.*, 2009; Smaczniak *et al.*, 2012; Li *et al.*, 2015), indicating that CLF and SWN play an important role in silencing the flower development program during vegetative growth. It is well documented that the antagonistic activities of microRNAs *MIR156* and *MIR172* coordinate the juvenile-to-adult leaf transition. *MIR319* represses the onset of senescence by targeting *TEOSINTE BRANCHED 1*, *CYCLOIDEA*, and *PCF (TCP)* TFs (Chen, 2004; Schommer *et al.*, 2008; Wu *et al.*, 2009; Rubio-Somoza and Weigel, 2011). The gene loci encoding these microRNAs are co-targets of CLF and SWN (Figure 10c). Plant hormones such as auxin mediate plant growth and developmental processes (Bari and Jones, 2009). It was noticed that CLF and SWN target genes that are involved in hormonal signaling pathways, such as genes encoding key players in auxin signaling pathways, *Aux/IAA* transcriptional repressors and *SMALL AUXIN UPREGULATED RNA (SAUR)* family (Figure 10d; Woodward and Bartel, 2005).

### 3.1.5 Genome-wide profiling of H3K27me3 in *clf-29*, *swn-4*, and *clf-29 swn-4*

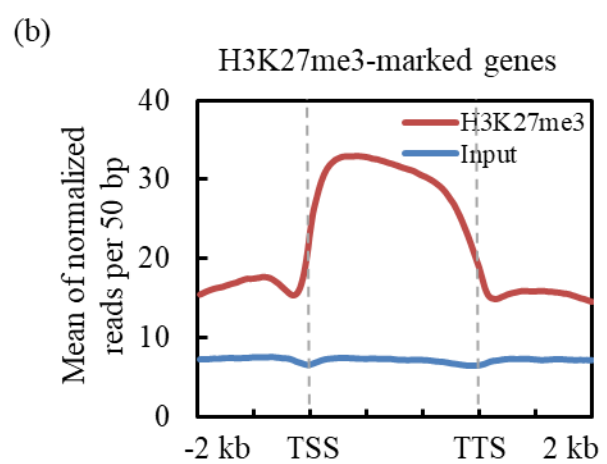
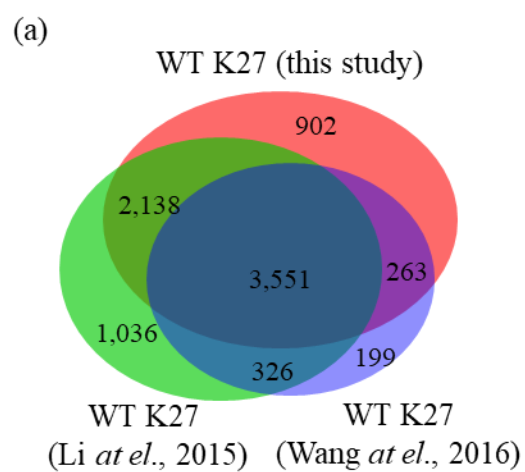
To further examine the functional redundancy between CLF and SWN in *Arabidopsis* seedlings, the genome-wide distributions of H3K27me3 in WT, *clf-29*, *swn-4*, and *clf-29 swn-4* were examined. In WT, 6,854 genes marked by H3K27me3 were identified, which largely overlap the published data (Figure 11a; Li *et al.*, 2015; Wang *et al.*, 2016). By analyzing the distribution pattern of H3K27me3 peaks over annotated genomic features, it was found that H3K27me3 has a relatively even distribution from TSSs to TSSs genome-wide (Figure 11b). Next, the size distribution of H3K27me3 in WT and CLF/SWN peaks was examined. Peaks of *RELATIVE OF EARLY FLOWERING 6 (REF6)*, a DNA binding factor that shows typical narrow peaks (Li *et al.*, 2016), was used as a control. It was shown that H3K27me3 and CLF/SWN peaks are generally broader compared to REF6 peaks (Figure 12).

By comparing the genome-wide occupancy between H3K27me3 and CLF/SWN, it was found that the large majority of CLF- and SWN-occupied genes are marked by H3K27me3



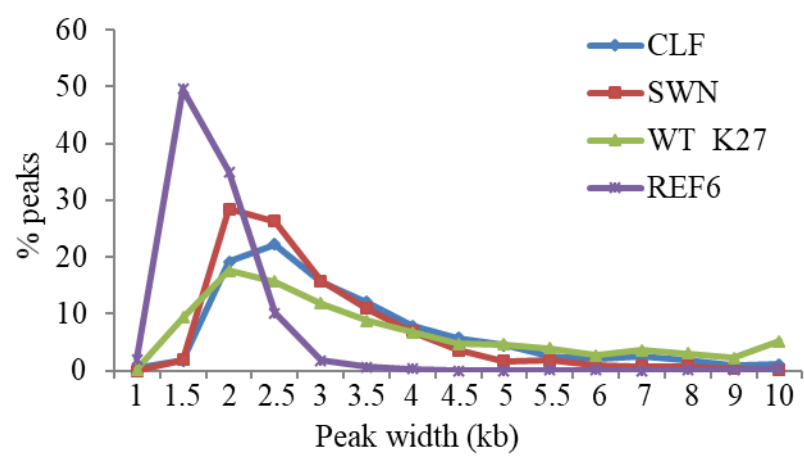
**Figure 11** Distribution of H3K27me3 in *Arabidopsis* genome.

(a) Comparison of H3K27me3 ChIP-seq (WT K27) in this study with two recently published ChIP-seq data (Li *et al.*, 2015; Wang *et al.*, 2016). (b) Mean density of H3K27me3 occupancy at all target genes in WT. The mean density of input is used as a negative control. The gene bodies (TSS to TTS) are scaled to 3 kb, 5' ends (-2 kb to TSS) and 3' ends (TTS to downstream 2 kb) are not scaled.



**Figure 12** Pattern of CLF and SWN peaks.

Width ranges for CLF, SWN, and WT H3K27me3 (WT K27) peaks. *X* axis shows the width of peaks within the ranges (e.g. 1 kb: width < 1 kb; 1.5 kb:  $1 \text{ kb} \leq \text{width} < 1.5 \text{ kb}$ ). *Y* axis represents percentage of peaks in each range. REF6 peaks are used as representative “narrow peaks” (Li *et al.*, 2016).



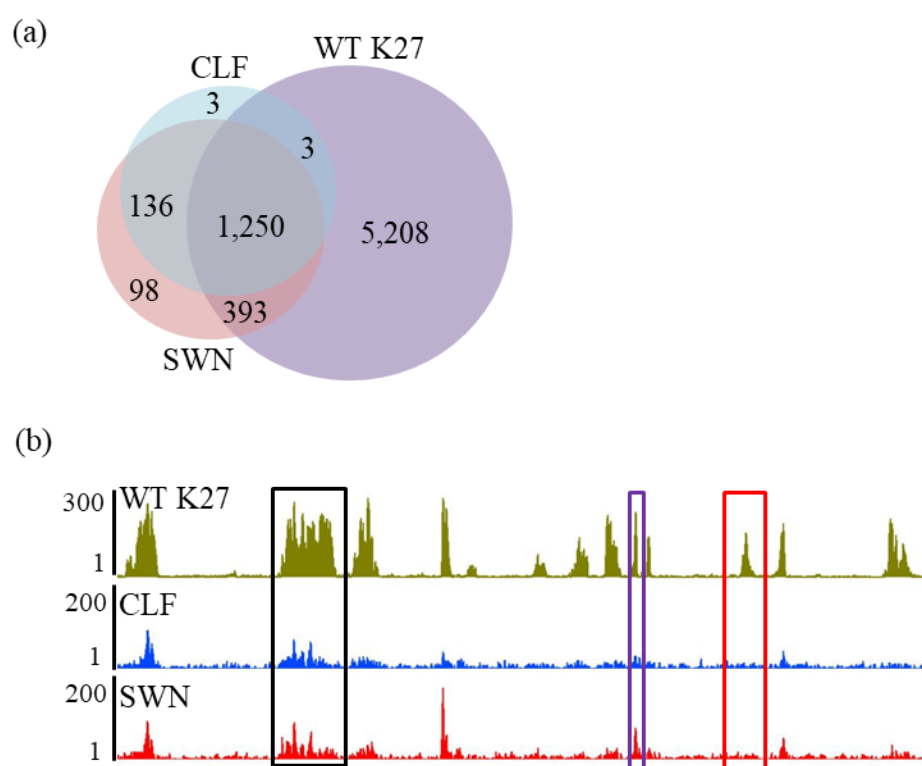
(1,253 out of 1,391 for CLF; 1,643 out of 1,877 for SWN; Figure 13a), which is consistent with the notion that both enzymes directly target genes to deposit H3K27me3. Surprisingly, on the other hand, the majority (5,208 out of 6,854) of the H3K27me3-marked genes are not occupied by either CLF or SWN in WT (Figure 13). Intriguingly, there was a small group of genes (237) that are CLF/SWN targets but showed no or very low levels of H3K27me3 (Figure 13). This was reminiscent of previous findings at some genomic loci such as the floral repressor locus *SVP* (Li *et al.*, 2015). In that work, it was shown that the chromatin remodeler BRM can promote gene expression by limiting the deposition of H3K27me3 by suppressing the recruitment and/or activity of CLF/SWN. To find out whether the lack of H3K27me3 deposition is due to the presence of BRM, BRM occupancy (Li *et al.*, 2016) at CLF and SWN co-targets was checked. It was observed that nearly half of these genes (63 out of 136) are occupied by BRM (Figure 14; Appendix D).

The global level of H3K27me3 was almost identical in *swn-4* and markedly lower in *clf-29* compared to the H3K27me3 level in WT (Figure 15). In *clf-29 swn-4* double mutant, nearly no H3K27me3 was detected (Figure 15). To eliminate the possibility that the loss of H3K27me3 was due to a change in nucleosome compaction, the levels of H3 were checked in the *clf-29 swn-4* double mutant. It was found that the H3 levels in *clf-29 swn-4* were very similar to those in WT, with a reduction at some loci (Figure 16), suggesting that the loss of H3K27me3 in *clf-29 swn-4* is not because of changes in H3 levels. Therefore, the lack of H3K27me3 in *clf-29 swn-4* indicates that CLF and SWN are probably the only active enzymes catalyzing H3K27me3 at the seedling stage.

When comparing H3K27me3 levels between *clf-29* and WT, I identified 552 and 91 genes whose H3K27me3 levels in *clf-29* were reduced and increased, respectively (Figure 17a). However, only 55 out of 643 genes were CLF direct targets (Figure 17a). In *swn-4*, changes in H3K27me3 levels were only detected in 68 genes, of which 32 were increased and 36 decreased (Figure 17b). After close examination of regions with reduced H3K27me3 levels in *clf-29* compared to WT, three types of patterns were identified (Figure 18a). In type I, loss of *CLF* had only a weak or no effect on the H3K27me3 peak summit but caused a drastic reduction of H3K27me3 in the flanking regions. The signal of H3K27me3 at

**Figure 13** Overlaps between H3K27me3-marked and CLF/SWN-occupied genes.

(a) Venn diagram showing overlaps among genes marked by H3K27me3 in WT (WT K27) and those occupied by CLF or SWN. (b) Genome browser views of one genomic region showing overlaps among WT K27, CLF, and SWN. The black box indicates the region is occupied by H3K27me3, CLF, and SWN; the purple box indicates the region is occupied by H3K27me3 and SWN; the red box indicates the region is only occupied by H3K27me3.

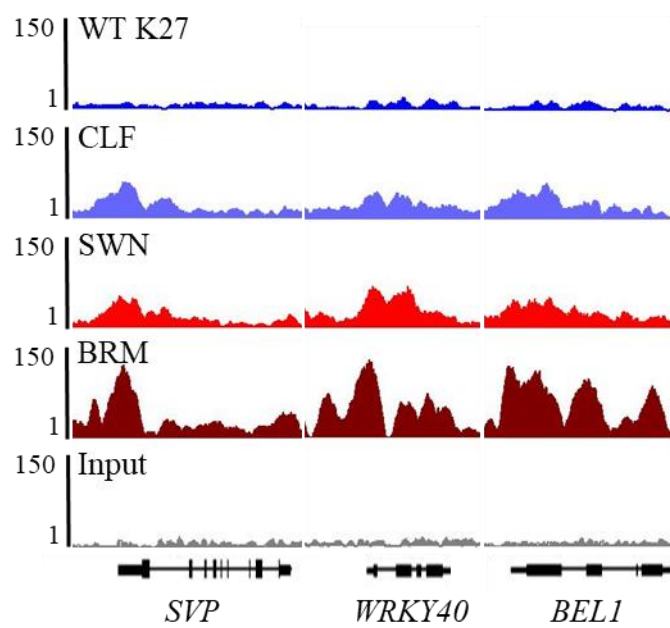


**Figure 14** BRM occupancy overlaps with H3K27me3/CLF/SWN occupancies.

Genome browser views of representative genes showing BRM occupancy (Li *et al.*, 2016) and the occupancies of H3K27me3 in WT (WT K27), CLF, and SWN. BRM input signals at these genes are also shown as a negative control. Gene structures are shown in black underneath the panel.

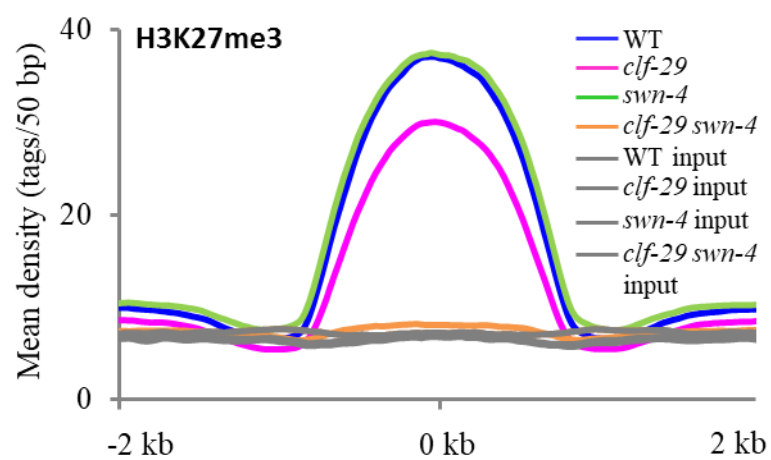
*SVP: SHORT VEGETATIVE PHASE; WRKY40: WRKY DNA-BINDING PROTEIN 40; BEL1: BELL 1.*





**Figure 15** Genome-wide profiling of H3K27me3 in different genetic backgrounds.

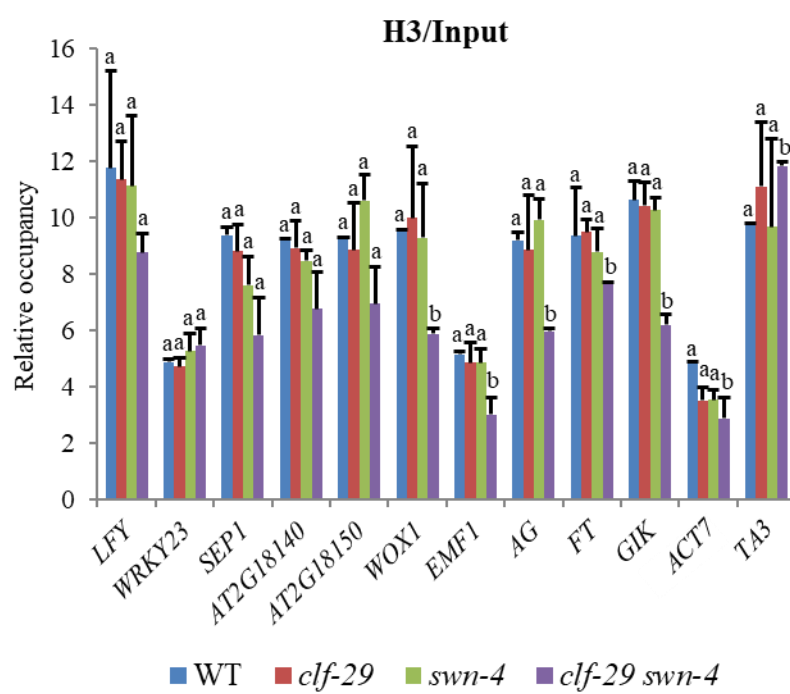
Mean density of H3K27me3 levels in WT, *clf-29*, *swn-4*, and *clf-29 swn-4*. Inputs are negative controls for ChIP-seq, which are shown in grey. The average signal within 2 kb genomic regions, flanking the center of the H3K27me3 peaks in WT, is shown.



**Figure 16** Histone H3 levels in different genetic backgrounds.

Mean  $\pm$  SD; three biological replicates are included; lowercase letters show significant differences of histone H3 levels among different genetic backgrounds, one-way ANOVA.

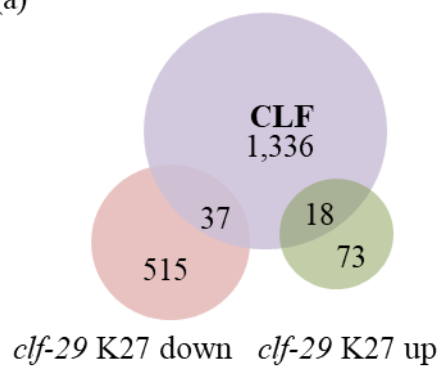
*LFY: LEAFY; WRKY23: WRKY DNA-BINDING PROTEIN 23; SEP1: SEPALLATA 1; WOX1: WUSCHEL RELATED HOMEBOX 1; AG: AGAMOUS; FT: FLOWER LOCUS T; GIK: GIANT KILLER; ACT7: ACTIN 7; TA3: a transposable element gene; AT2G18140 and AT2G18510 are uncharacterized genes.*



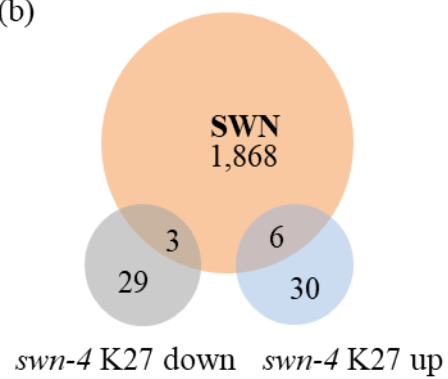
**Figure 17** Overlaps of genes occupied by CLF/SWN and those showing changes in H3K27me3.

(a) Venn diagram showing overlaps between genes occupied by CLF and genes with decreased (*clf-29* K27 down) and increased H3K27me3 levels (*clf-29* K27 up) in *clf-29* compared to WT. (b) Venn diagram showing overlaps between genes occupied by SWN and genes with decreased (*swn-4* K27 down) and increased H3K27me3 levels (*swn-4* K27 up) in *swn-4* compared to WT.

(a)



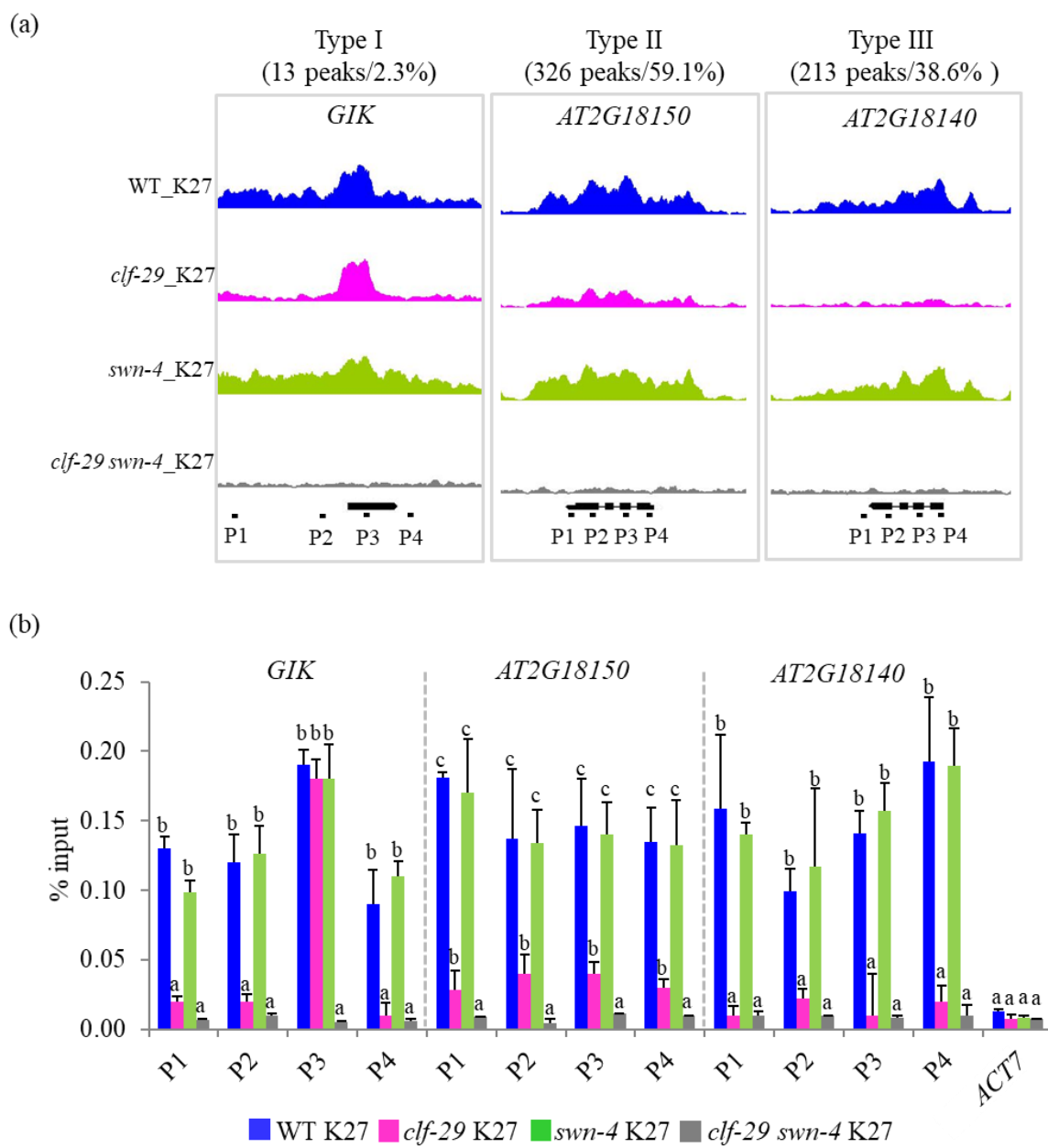
(b)



**Figure 18** Three patterns of H3K27me3 reduction in *clf-29*.

(a) Genome browser views at representative genomic loci showing three distinct patterns of H3K27me3 reduction in *clf-29* (*clf-29* K27; pink) compared to WT (WT K27; blue). H3K27me3 levels in *swn-4* (*swn-4* K27; green) and *clf-29 swn-4* (*clf-29 swn-4* K27; grey) are shown. Gene structures are shown in black underneath each panel. P1, P2, P3, and P4 are four pairs of primers used for ChIP-qPCR validation. (b) ChIP-qPCR validation. ChIP signals are shown as percentage of input. *ACT7* was used as a negative control locus. H3K27me3 (K27) levels in WT, *clf-29*, *swn-4*, and *clf-29 swn-4* are shown in blue, pink, green, and grey, respectively. Error bars show standard deviations among three biological replicates. Statistically significant differences are shown by different lowercase letters, one-way ANOVA.





the summits, as shown in the ChIP-seq, completely disappeared in *clf-29 swn-4*. In type II, a partial reduction of H3K27me3 across the entire H3K27me3-marked region was observed in *clf-29*, while the signal was completely abolished in *clf-29 swn-4*. In type III, a complete loss of H3K27me3 was observed in *clf-29* compared to WT. These three patterns were also validated by ChIP-qPCR (Figure 18b).

### 3.1.6 Transcriptome profiling in *clf-29*, *swn-4*, and *clf-29 swn-4*

To further understand the action of CLF and SWN, RNA-seq was performed to profile transcriptome changes in *clf-29*, *swn-4*, and *clf-29 swn-4* mutants compared to WT, respectively. In *clf-29*, it was found that 591 genes were transcriptionally up-regulated compared to WT, whereas the expression levels of 426 genes were down-regulated (Figure 19). In *swn-4*, a total of 374 and 266 genes showed increased and decreased expression levels, respectively, compared to WT (Figure 19). Overall, fewer genes showing differential expression were detected in *swn-4* (640) compared to *clf-29* (1,017), which may explain why *swn-4* mutants display less severe morphological defects compared to *clf-29* mutants. In *clf-29 swn-4*, more than 10,000 genes (6,058 up-regulated; 4,601 down-regulated) showed differentially expressional levels compared to WT (Figure 19a), in agreement with the *clf-29 swn-4* embryo-like phenotype. Next, I also examined whether there is a correlation between the loss of H3K27me3 and the expression levels of transcripts in the *clf-29 swn-4* double mutants. It was observed that many genes are up-regulated in *clf-29 swn-4* compared to WT, particularly those that were heavily marked by H3K27me3 in WT plants (Figure 19b). However, many genes were also down-regulated in *clf-29 swn-4* compared to WT, regardless of the loss of H3K27me3 in *clf-29 swn-4* (Figure 19b).

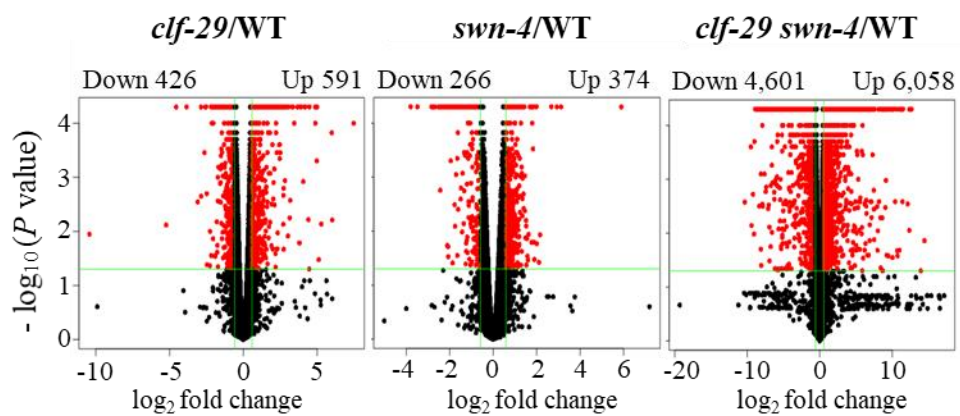
### 3.1.7 Implication of CLF in floral transition

It is well known that PcG proteins play important roles in developmental transitions (Pu and Sung, 2015; Xiao and Wagner, 2015), but the underlying mechanisms are still not completely understood. By integrating CLF/SWN and H3K27me3 ChIP-seq and RNA-seq data, it was attempted to show how CLF/SWN may act to regulate floral transition. The data provide some clues for the early flowering phenotype of *clf* mutants. As shown in

**Figure 19** Transcriptome profiling of PcG mutants.

(a) Volcano plots displaying up-regulated and down-regulated genes in PcG mutants (*clf-29*, *swn-4*, and *clf-29 swn-4*) compared to WT. *X* axis represents  $\log_2$  value of fragments per kilobase per million (FPKM) mapped reads in each mutant vs WT. *Y* axis shows the  $-\log_{10}$  of *P* value. Red dots represent genes that are differentially expressed in PcG mutants compared to WT ( $P < 0.05$ , fold change  $> 1.5$ ). Black dots represent genes that are not differentially expressed in PcG mutants compared to WT ( $P \geq 0.05$  or fold change  $\leq 1.5$ ). The horizontal green lines indicate  $P = 0.05$  (upper:  $< 0.05$ ; bottom:  $P > 0.05$ ). The vertical green lines indicate fold change = 1.5 (fold change  $< 1.5$  between two lines). (b) Heatmaps illustrating ChIP-seq density in WT and PcG mutants (*clf-29*, *swn-4*, and *clf-29 swn-4*), ranked by H3K27me3 intensity within  $\pm 2$  kb of peak summits in WT (green), and RNA-seq intensity for the H3K27me3-marked genes in WT with the same order. Each horizontal line represents an H3K27me3 peak. Columns show the genomic region surrounding each peak summit. Signal intensities are indicated by the shade of green. The expression intensity is measured by  $\log_{10}$  (FC), FC = fold change (mutant vs WT FPKM).

(a)



(b)

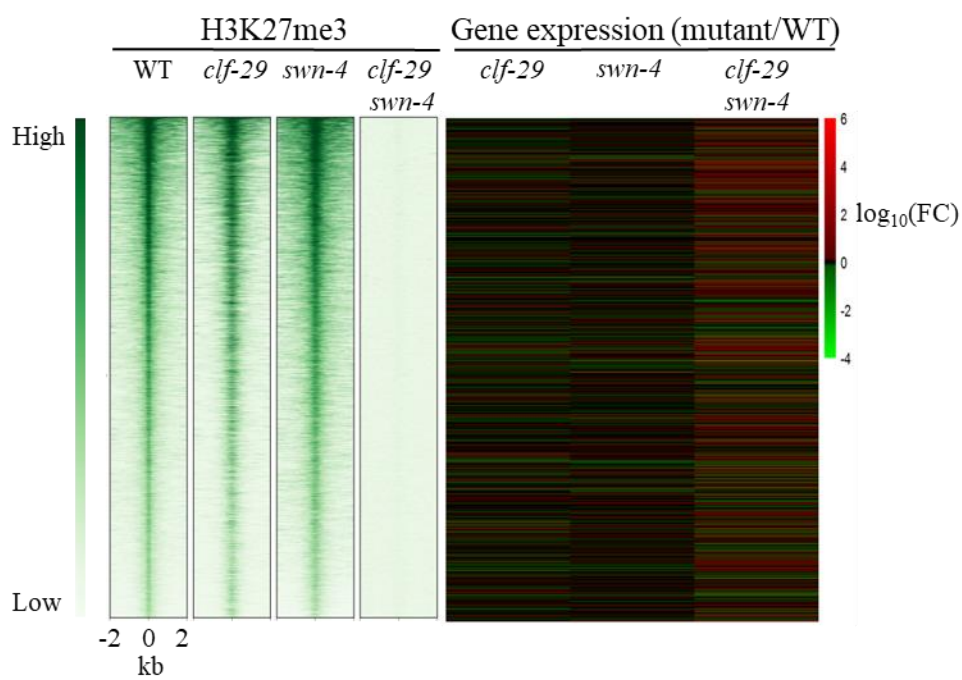


Figure 20, the expression levels of several genes that encode positive regulators of floral transition such as *AGL17*, *AGL19*, *AGL24*, *SOC1*, and *FT*, were increased in *clf-29*, but not in *swn-4* compared to WT. Two flowering repressor-encoding genes *FLC* and *SVP* showed differential expression (up- and down-regulated, respectively) compared to WT. Based on a literature search, previous publications reported that all these genes, except for *AGL17*, have roles in PcG repression (Schönrock *et al.*, 2006; Lopez-Vernaza *et al.*, 2012; Torti and Fornara, 2012; Hou *et al.*, 2014; Müller-Xing *et al.*, 2014; Li *et al.*, 2015). Thus, it is interesting to examine whether *AGL17* is also involved in PcG repression.

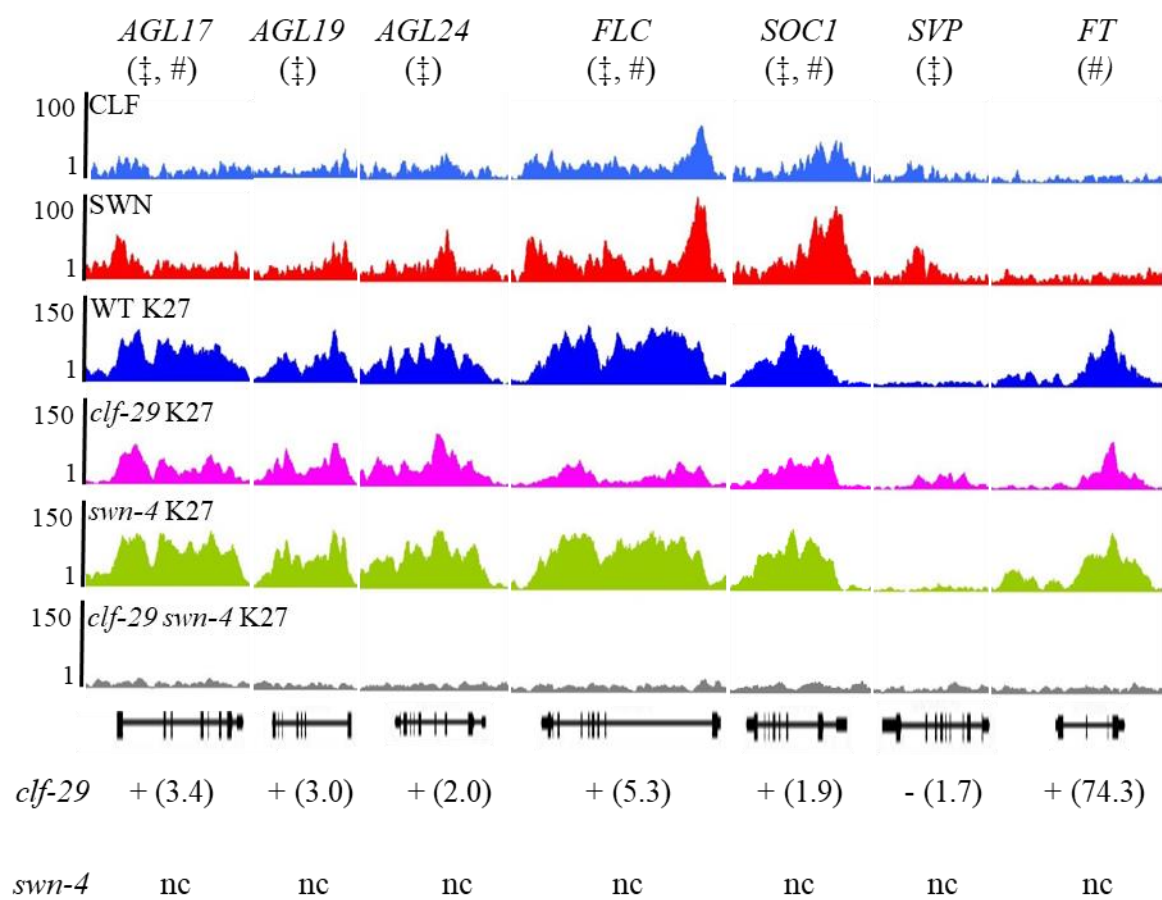
CLF was shown to target *AGL17* by ChIP-seq (Figure 21a). To confirm the occupancy of CLF at *AGL17*, ChIP-qPCR was performed to examine the CLF enrichment at *AGL17*. The results show that CLF is significantly enriched in the first exon of *AGL17* (Figure 21b). In addition, the level of H3K27me3 was decreased at *AGL17* in *clf-29* compared to WT (Figure 21c). Consistently, the H3K27me3 level was significantly decreased in *clf-29* compared to WT as determined by ChIP-qPCR (Figure 21d). It was previously shown that *AGL17* acts in an *FT*-independent photoperiod pathway to promote flowering (Han *et al.*, 2008). To examine whether *AGL17* up-regulation contributes to the early flowering phenotype of the *clf-29* mutant, a T-DNA mutant line *agl17-3* was crossed with *clf-29*. From the F3 progeny, the *clf-29 agl17-3* double mutants were identified by PCR-based genotyping (see section 2.5). Then WT, *clf-29*, *agl17-3*, and *clf-29 agl17-3* were grown in soil side by side to monitor their flowering time. Interestingly, *clf-29 agl17-3* flowered earlier than WT, but later than *clf-29* (Figure 22a and b). Furthermore, the expression of *AGL17* was significantly up-regulated in *clf-29* compared to WT (Figure 22c). These data suggest that *AGL17* up-regulation is partially responsible for the early flowering phenotype of the *clf-29* mutant.

### **3.2 Genome-wide characterization of *Arabidopsis* SWI/SNF chromatin remodeler SYD**

The functions of SWI/SNF chromatin remodelers have been extensively examined in yeast and animals (Hargreaves and Crabtree, 2011; Euskirchen *et al.*, 2012; Kadoch and Crabtree, 2015). Recent studies have improved our understanding of the roles of SWI/SNF

**Figure 20** ChIP-seq signals at representative genes involved in the control of flowering.

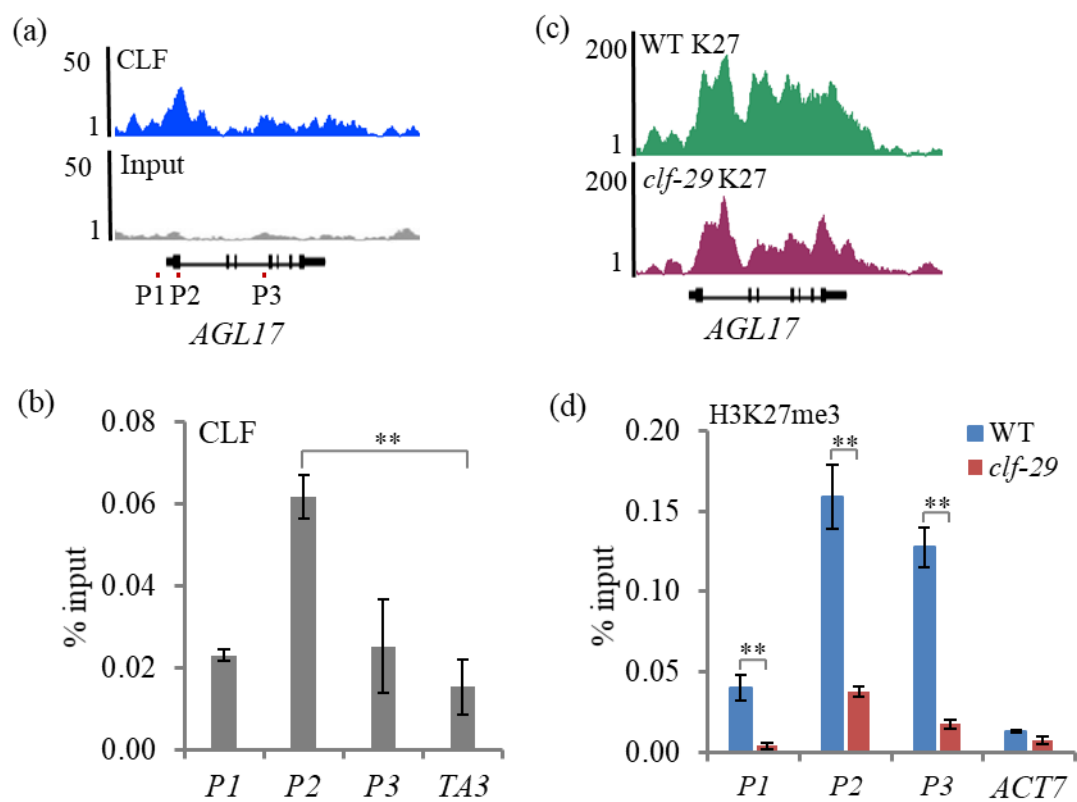
Top panel: ChIP-seq genome browser views of CLF and SWN occupancy as well as H3K27me3 (K27) levels in WT, *clf-29*, *swn-4*, and *clf-29 swn-4* at major floral transition genes. Gene structures are shown in black underneath each panel. “‡” represents genes occupied by CLF, “#” represents decrease of H3K27me3 levels in *clf-29* compared to WT. Bottom panel: fold change of expression levels at these representative genes in *clf-29* and *swn-4* compared to WT, respectively. “+”, “-”, and “nc” indicate up-regulation, down-regulation, and no-change, respectively. Numbers in brackets show fold change compared to WT.



**Figure 21** CLF directly targets the flowering activator *AGL17*.

(a) Genome browser view of CLF occupancy at *AGL17* locus. Input is used as a negative control. P1, P2, and P3 are the locations of three pairs of primers used in (b) and (d). Gene structure is shown in black underneath the panel. (b) ChIP-qPCR validation to confirm CLF targets *AGL17* by using ChIP DNA extracted from two-week-old *clf-29 pCLF::CLF-GFP* seedlings. *TA3*, a transposable element gene, was used as a negative control locus. Error bars show standard deviations among three biological replicates. Student's *t*-test,  $**P < 0.01$ . (c) Genome browser view of H3K27me3 in WT (WT K27) and *clf-29* (*clf-29* K27) at *AGL17* locus. (d) ChIP-qPCR validation to confirm the reduction of H3K27me3 in *clf-29* (*clf-29* K27) compared to WT (WT K27). *ACT7* was used as a negative control. Error bars show standard deviations among three biological replicates. Student's *t*-test,  $**P < 0.01$ .





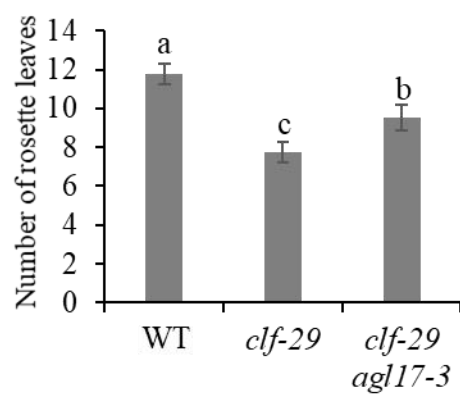
**Figure 22** Partial contribution of *AGL17* to the early flowering phenotype of *clf-29*.

(a) Image of 5-week-old WT, *clf-29*, *agl17-3*, and *clf-29 agl17-3* plants. Scale bar: 1 cm.  
 (b) Rosette leaf number at the time of bolting of plants in different genetic backgrounds (WT, *clf-29*, and *clf-29 agl17-3*). Flowering time was recorded as the number of rosette leaves at bolting, thus late flowering plants have more rosette leaves. Error bars show standard deviation from at least 20 plants for each genetic background. Lowercase letters indicate significant differences between genetic backgrounds, as determined by Post-hoc Tukey's HSD test. (c) Expression of *AGL17* in two-week-old WT, *clf-29*, *agl17-3*, and *clf-29 agl17-3* seedlings. The expression levels were normalized to that of *GAPDH*. Error bars show standard deviations among three biological replicates. Student's *t*-test,  $**P < 0.01$ .

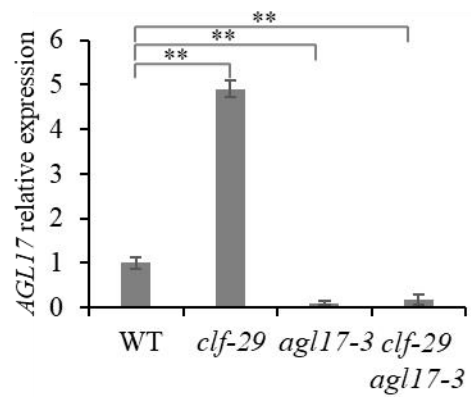
(a)



(b)



(c)



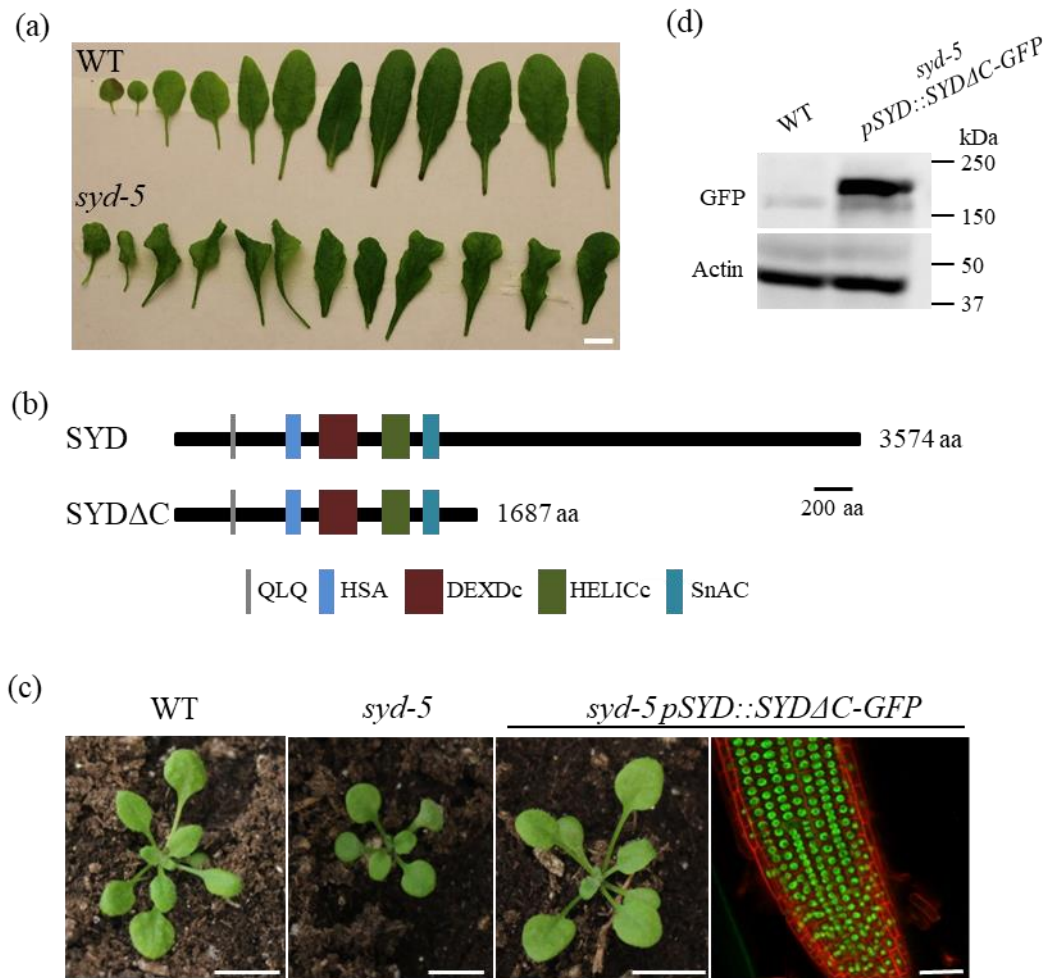
proteins in plants, including their suppression of PcG function (Li *et al.*, 2015; Wu *et al.*, 2015; Yang *et al.*, 2015). My focus in this section is to understand the biological roles and functions of *Arabidopsis* SWI/SNF chromatin remodeler SYD by employing global approaches, which is in preparation for investigating the global interactions between SYD and BRM.

### 3.2.1 GFP-tagged SYD transgene functions *in vivo*

It was reported that *syd* mutants exhibit a variety of developmental defects, such as short stature, female-sterility, and loss of maintenance of SAMs (Wagner and Meyerowitz, 2002; Kwon *et al.*, 2005). In addition, after close observation of *syd-5*, it was found that the rosette leaves are smaller and curl downward compared to WT (Figure 23a). To profile the genome-wide occupancy of SYD, transgenic *Arabidopsis* lines expressing GFP-tagged SYD under the control of its own promoter in *syd-5* genetic background (*syd-5 pSYD::SYDΔC-GFP*) were generated. Of note, the N-terminal ATPase AT-hook-containing region of SYD has been shown to be sufficient for its biological activity (Su *et al.*, 2006). Thus, the GFP-tagged SYD fusion generated in this thesis excluded its C-terminal domain (referred to SYDΔC-GFP; Figure 23b). To check whether the transgene functions, the morphological phenotypes of the *syd-5 pSYD::SYDΔC-GFP* transgenic plants were examined. No morphological defects were observed in transgenic plants compared to WT (Figure 23c), suggesting that the *pSYD::SYDΔC-GFP* transgene can fully rescue the *syd-5* phenotype. Then, GFP signals were observed in the nucleus by using 4-day-old *syd-5 pSYD::SYDΔC-GFP* roots under a confocal microscope (Figure 23c), showing that SYD is localized in the nucleus. Further, a Western blot experiment was conducted to examine the size of the SYDΔC-GFP protein by using an anti-GFP antibody. Compared to WT, a band (~200 kDa) was observed in *syd-5 pSYD::SYDΔC-GFP*, showing that the transgene is expressed and translated as expected *in vivo* (Figure 23d). Taken together, these data suggest that the truncated version of SYD can fully complement the *syd-5* mutant.

**Figure 23** Truncated *SYD* fully completes *syd-5*.

(a) Representative leaves of 38-day-old WT and *syd-5* plants. Cotyledon and rosette leaves are shown. Scale bar: 1 cm. (b) Conserved domains of the SYD protein. SYDAC represents the C-terminal truncated version of the SYD protein. QLQ: Gln, Leu, Gln motif; HSA: helicase/SANT-associated domain; DEXDc: DEAD-like helicases superfamily; HELICc: helicase superfamily C-terminal domain; SnAC: Snf2-ATP coupling. Domains were identified using the NCBI Conserved Domain Database. (c) Image of 3-week-old WT, *syd-5*, and *syd-5 pSYD::SYDAC-GFP* plants. GFP signals detected by confocal microscopy in 4-day-old *syd-5 pSYD::SYDAC-GFP* roots. Propidium iodide was used to stain cell walls (red). Scale bar: 50  $\mu$ m. (d) Western blot analysis of nuclear extracts from two-week-old *syd-5 pSYD::SYDAC-GFP* seedlings. Antibodies used: anti-GFP and anti-Actin. Nuclear extract from WT seedlings used as a negative control, total Actin protein used as a loading control.



### 3.2.2 Genomic regions occupied by SYD at two developmental stages

To profile the global occupancy of SYD, ChIP-seq was performed by using ChIP DNA extracted from two-week-old *syd-5 pSYD::SYDAC-GFP* seedlings. Two independent biological replicates of ChIP DNA were sequenced. A strong correlation between the two replicates was observed (the Pearson coefficient is 0.88; Figure 24a). In total, there were 4,865 SYD peaks detected by ChIP-seq, corresponding to 6,714 genes in the *Arabidopsis* genome (Figure 24b). Inspection of SYD occupancy at previously published targets by ChIP-qPCR confirmed the results generated by ChIP-seq in this study (Walley *et al.*, 2008; Wu *et al.*, 2015). For example, the *MYB DOMAIN PROTEIN 2* (*MYB2*), *LFY*, and *TARGET OF MONOPTEROS 3* (*TMO3*) loci were shown to be SYD direct targets (Figure 24c). In addition, it was reported that SYD plays important roles in carpel and ovule development (Wagner and Meyerowitz, 2002). To further investigate the dynamics and tissue specificity of SYD occupancy in the genome, another ChIP-seq experiment was conducted to identify the genome-wide occupancy of SYD in inflorescences. As a result, 4,264 SYD peaks (corresponding to 5,975 genes) were identified in inflorescences (Figure 24b), of which, 750 SYD target genes were found to be specific to seedlings, and only 11 were specific to inflorescences (Figure 24d). For example, SYD was found to be uniquely enriched at *ZINC FINGER OF ARABIDOPSIS THALIANA 11* (*ZAT11*) at the seedling stage (Figure 24e). *ZAT11* is highly expressed in seedlings, especially in primary roots, where it plays a positive role in primary root growth (Liu *et al.*, 2014). Further, SYD targets *SEPALLATA 2* (*SEP2*) at the inflorescence stage but not the seedling stage (Figure 24e). *SEP2*, a MADS-box TF that is strongly expressed in inflorescences, plays vital roles in the identity determination of petals, stamens, and ovules (Favaro *et al.*, 2003), indicating that SYD may regulate ovule development by directly targeting and regulating *SEP2*.

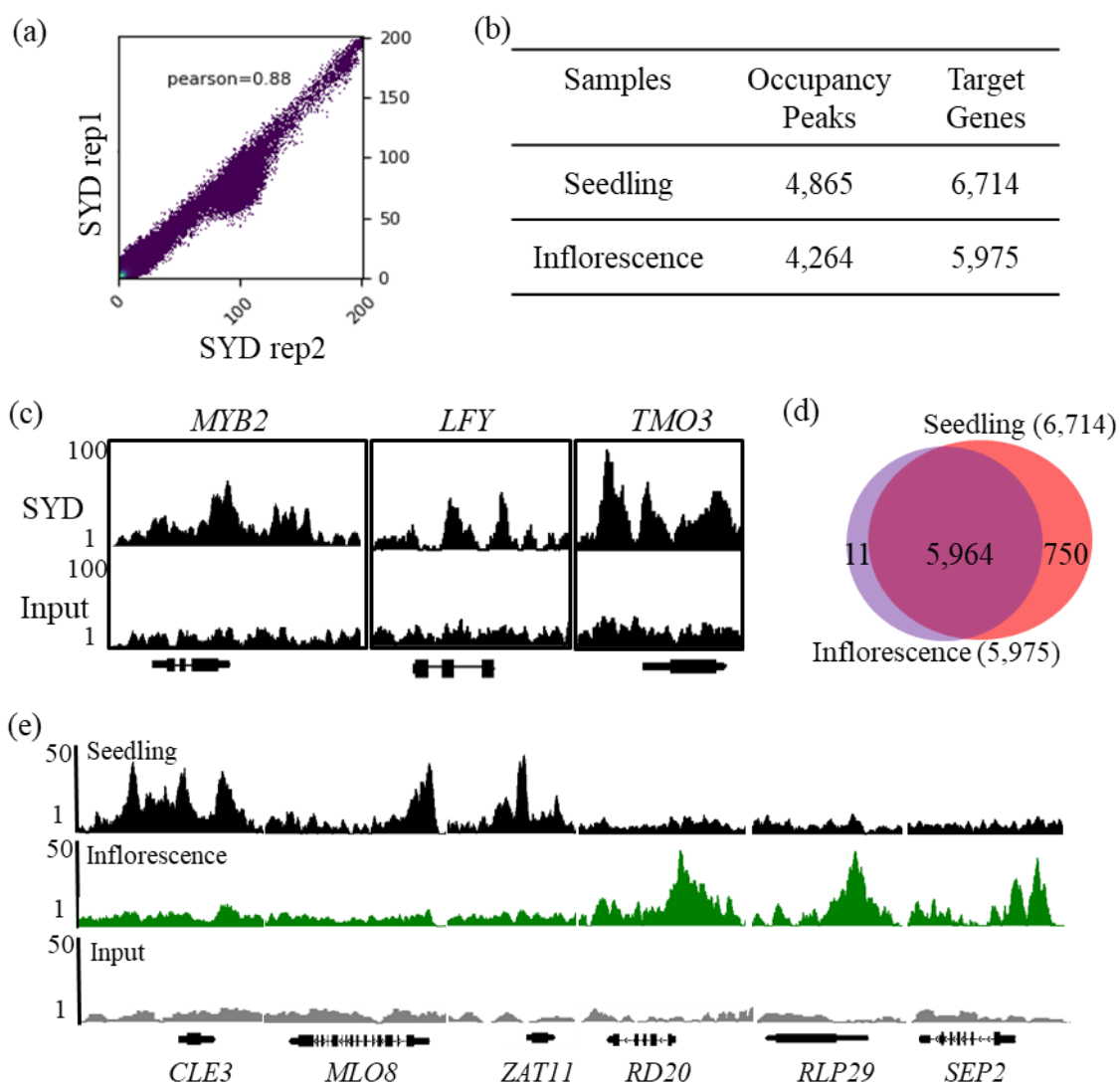
Due to a strong similarity of SYD occupancy between the two developmental stages, I only focused, in the following section, on the examination of SYD functions at the seedling stage.

**Figure 24** Genome-wide occupancy of SYD at two developmental stages.

(a) Pearson correlation plot for two SYD (*syd-5 pSYD::SYDAC-GFP*) ChIP-seq replicates at the seedling stage. The whole genome is divided into non-overlapping bins (50 bp), and normalized reads of each replicate (normalization is performed by randomly selecting the same number of reads from each replicate) are filled in each bin. The numbers of normalized reads in each bin are plotted. *X* and *Y* axis show the number of reads in each bin for each biological replicate. (b) Table showing the number of SYD peaks and corresponding genes at two developmental stages. (c) Genome browser views of SYD occupancy at representative genes at the seedling stage. Input signals are used as negative controls. Gene structures are shown in black underneath the panel. (d) Venn diagram showing overlap between seedling and inflorescence SYD targets. (e) Genome browser views of SYD unique targets in seedlings (black) and inflorescences (green), respectively. Input signals (grey) are used as negative controls. Gene structures are shown in black underneath the panel.

*CLE3: CLAVATA 3/ESR-RELATED 3; MLO8: MILDEW RESISTANCE LOCUS O 8; ZAT11: ZINC FINGER OF ARABIDOPSIS THALIANA 11; RD20: RESPONSIVE TO DESICCATION 20; RLP29: RECEPTOR LIKE PROTEIN 29; SEP2: SEPALLATA 2.*





### 3.2.3 Patterns of SYD occupancy

To examine the distribution of SYD peaks genome-wide, ChIPseeker, a web-based analysis tool, was employed (see section 2.9). It was found that ~50% of identified SYD peaks are located in the promoter regions and nearly a quarter in intergenic regions (Figure 25a). Further, two major types of SYD peaks were observed, namely single, narrow peaks and clustered broad peaks (see section 2.9; Figure 25b). In terms of the size distribution of SYD peaks, it was found that ~60% peaks are within 1-2 kb and ~40% are larger than 2 kb in length (Figure 25c). Furthermore, the signals of SYD occupancy were found to be highest near TSSs, with weaker signals near TTSs (Figure 25d), sharing a similar occupancy pattern with BRM (Li *et al.*, 2016; Archacki *et al.*, 2017).

### 3.2.4 SYD may mainly activate its targets

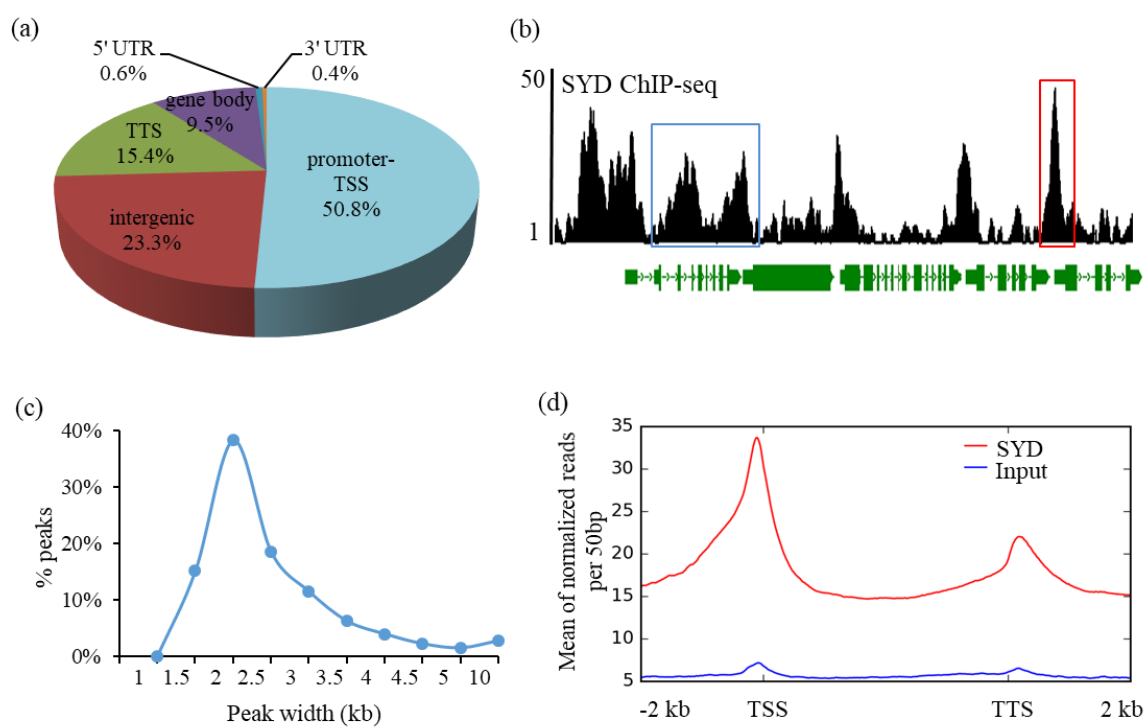
To further examine the role of SYD in the regulation of gene expression, RNA-seq was performed to profile transcriptome changes in *syd-5* compared to WT (see section 2.10). It was found that while 601 genes were transcriptionally up-regulated compared to WT, the expression of 1,535 genes was down-regulated (Figure 26a). Next, the average levels of gene expression in WT and *syd-5* were compared. It was found that there is a decrease in *syd-5* compared to WT (Figure 26b). To better understand the relationship between SYD occupancy and gene expression, the levels of SYD occupancy and transcriptional levels were examined. It was shown that higher expression level is associated with higher occupancy of SYD (Figure 26c). In addition, approximately 11% (755 out of 6,714) of genes occupied by SYD exhibited differentially expressed levels in *syd-5* compared to WT. The majority of SYD direct targets were down-regulated (522 out of 755) in *syd-5* compared to WT (Figure 26d). Together, these findings suggest that SYD plays a role as a transcription activator.

### 3.2.5 Interplay between SYD and BRM

Previous genetic work suggested that *Arabidopsis* SYD and BRM exhibit redundant as well as differential roles in regulating plant growth and development (Bezhani *et al.*, 2007; Wu *et al.*, 2015). Interestingly, one recent publication also revealed that BRM can

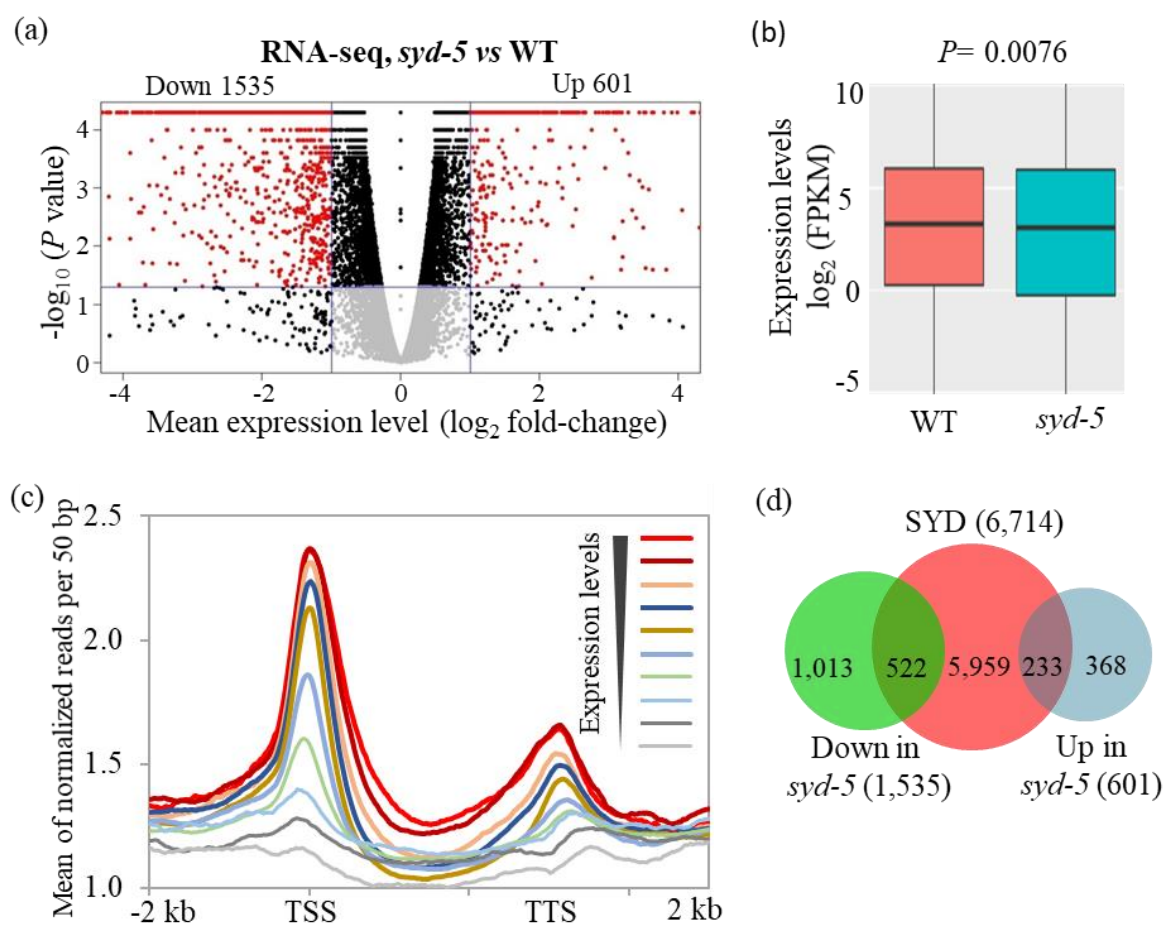
**Figure 25** Pattern of SYD peaks.

(a) Pie chart showing the distribution of SYD peaks at annotated genic and intergenic regions in the genome. (b) ChIP-seq genome browser view of SYD occupancy in a region on chromosome 4. Blue and red boxes highlight clustered broad and single narrow peaks, respectively. Gene structures are shown in green underneath the panel. (c) The size distribution of SYD peaks. *X* axis shows the width of peaks within the ranges (e.g. 1 kb: width < 1 kb; 1.5 kb: 1 kb ≤ width < 1.5 kb; 10 kb: 5 kb ≤ width < 10 kb). *Y* axis represents percentage of peaks in each range. (d) Mean density of SYD occupancy at all target genes. The mean density of input is used as a negative control. The gene bodies (TSS to TTS) are scaled to 3 kb, 5' ends (-2 kb to TSS) and 3' ends (TTS to downstream 2 kb) are not scaled.



**Figure 26** Transcriptome profiling of *syd-5*.

(a) Volcano plot showing up-regulated and down-regulated genes in *syd-5* compared to WT, respectively. *X* axis represents  $\log_2$  value of fold-change (fragments per kilobase per million (FPKM) mapped reads in *syd-5* vs WT). *Y* axis shows the  $-\log_{10}$  of *P* value. Red dots represent genes that are differentially expressed in *syd-5* compared to WT ( $P < 0.05$ , fold change  $> 2$ ). Black dots represent genes that are not differentially expressed in *syd-5* compared to WT ( $P \geq 0.05$  or fold change  $\leq 2$ ). The horizontal grey line indicates  $P = 0.05$  (upper:  $P < 0.05$ ; bottom:  $P > 0.05$ ). The vertical grey lines indicate fold change = 2 (fold change  $< 2$  between two lines). (b) Box plot representing the average expression level (FPKM) of genes in WT and *syd-5*. (c) Enrichment of SYD occupancy is associated with transcriptional levels. WT transcriptional levels (FPKM) are grouped from high (1<sup>st</sup> group; red) to low (10<sup>th</sup> group; grey). Mean signals of SYD occupancy along the gene bodies (scaled to 3 kb from TSS to TTS) and 2 kb region flanking the TSS or TTS are plotted. (d) Venn diagram showing overlap between genes targeted by SYD and genes showing differential expression in *syd-5* compared to WT.



physically interact with SYD (Li *et al.*, 2016). Taking advantage of the published BRM genome-wide occupancy (Li *et al.*, 2016), the overlap between SYD and BRM genome-wide targets was examined. First, the signals of SYD occupancy were plotted with BRM ChIP-seq signals (Li *et al.*, 2016). As shown, there were co-occupancy signals between SYD and BRM (Figure 27a). About half of SYD targets (3,387 out of 6,714) were occupied by BRM (Figure 27b), suggesting that SYD and BRM probably function redundantly on their common targets. On the other hand, the finding that loss-of-function of *SYD* and *BRM* display different morphological defects implies that SYD and BRM have different roles at their unique targets. To further examine the relationship between SYD and BRM, the expression of genes that are co-occupied by SYD and BRM in WT, *syd-5*, and *brm-1* was checked. The majority of SYD and BRM co-targeted genes (2,805 out of 3,387) did not show any differential expression in either *syd-5* or *brm-1* compared to WT (Figure 27c), implying that SYD and BRM may function redundantly on these genes. However, 454 co-targeted genes were differentially expressed in *syd-5* compared to WT (Figure 27c). Furthermore, 78 co-targeted genes differentially expressed in *brm-1* compared to WT (Figure 27c). To summarize, the data suggest that SYD and BRM function not only redundantly but also differentially on specific genes.

### 3.3 Interplay between SWI/SNF and PcG

#### 3.3.1 Loss of *SYD* activity leads to changes of H3K27me3 at hundreds of genes

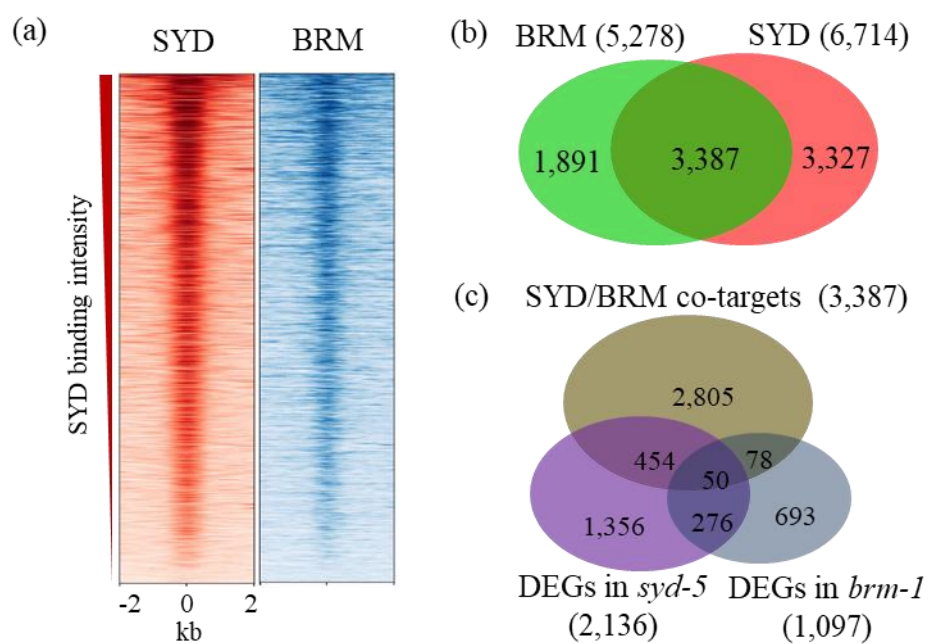
H3K27me3 is deposited by PcG catalytic enzymes CLF and SWN in *Arabidopsis* seedlings (Zhou *et al.*, 2017). The SWI/SNF complex has been proposed to function antagonistically with PcG repression (Reyes, 2014; Han *et al.*, 2015). To investigate the interplay between SWI/SNF subunit SYD and PcG, I first examined whether SYD affects the genome-wide distribution of H3K27me3. Thus, ChIP-seq, with an anti-H3K27me3 antibody, using DNA from two-week-old WT and *syd-5* seedlings was performed.

The correlations between two biological ChIP-seq replicates of each ChIP-seq were strong (Figure 28a). Then, the genome-wide profiles of SYD and H3K27me3 were compared by plotting average signals of occupancy (see section 2.9). However, there was no H3K27me3 enrichment at TSSs or TTSs where SYD occupies genome-wide (Figure 28b). It was

**Figure 27** SYD and BRM target common genes.

(a) Heatmaps representing co-occupancy of SYD (red) and BRM (blue) in the genome. Heatmaps are sorted according to the intensity of SYD occupancy. Each horizontal line represents an SYD peak. Columns show genomic region surrounding each peak summit. (b) Venn diagram showing overlap between SYD targets in this study and BRM targets from published data (Li *et al.*, 2016). (c) Venn diagram showing overlaps between SYD and BRM co-targets and differentially expressed genes (DEGs) in *syd-5* (this study) and *brm-1* (Li *et al.*, 2016) compared to WT, respectively.



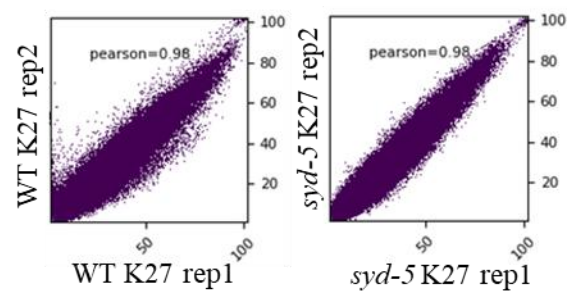


**Figure 28** Genome-wide profiling of H3K27me3 in *syd-5*.

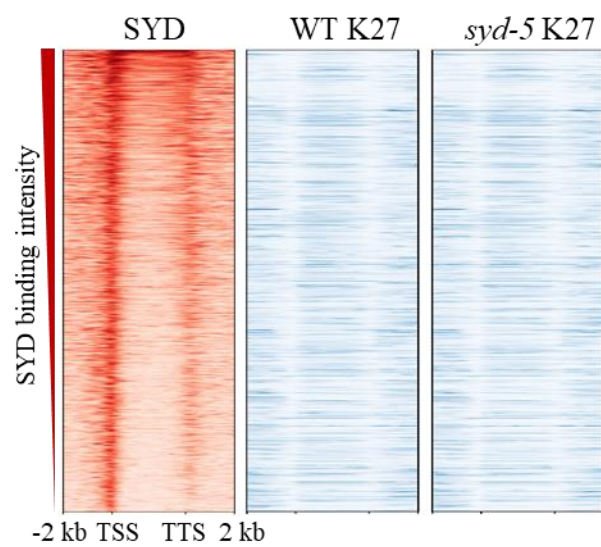
(a) Pearson correlation plots for two H3K27m3 ChIP-seq replicates in WT and *syd-5*. The whole genome is divided into non-overlapping bins (50 bp), and normalized reads of each replicate (normalization is performed by randomly selecting the same number of reads from each replicate) are filled in each bin. The numbers of normalized reads in each bin are plotted. *X* and *Y* axis show the number of reads in each bin for each biological replicate.

(b) Heatmaps representing distribution of SYD occupancy (red) and H3K27me3 (blue) in WT and *syd-5*. Each horizontal line represents an SYD-targeted gene. Mean signals along the gene bodies (TSS to TTS; scaled to 3 kb) and 2 kb region flanking the TSS or TTS are plotted.

(a)



(b)



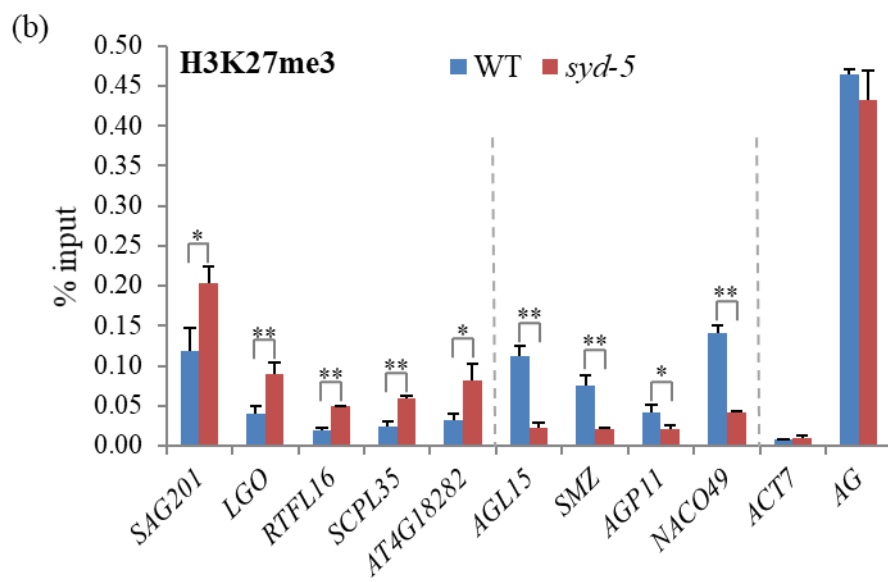
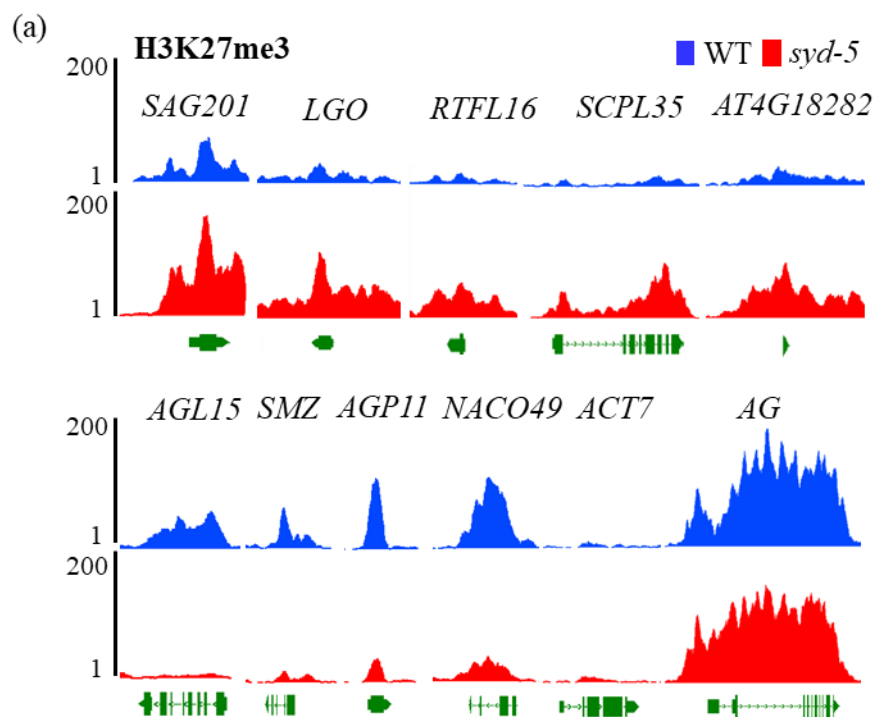
reasoned that changes may occur only at a small number of genes and not clearly be seen genome-wide. Therefore, the H3K27me3 signals in *syd-5* and WT at individual genes throughout the genome were checked. Indeed, I identified 291 genes whose H3K27me3 levels changed more than two-fold in *syd-5* compared to WT, with 122 and 169 genes showing an increase and a reduction, respectively. Further, a few representative loci that show changes in H3K27me3 levels in *syd-5* compared to WT by ChIP-seq are shown (Figure 29a). In addition, *ACTIN 7 (ACT7)* and *AG* which are CLF/SWN non-target and target, respectively, were used as controls. ChIP-qPCR was performed to validate these results. In agreement with the ChIP-seq results, changes of H3K27me3 levels in *syd-5* compared to WT were also observed by ChIP-qPCR (Figure 29b), suggesting that H3K27me3 levels at these 291 genes are changed in *syd-5* compared to WT.

In addition to this finding, nearly one-third (100 out of 291) of these H3K27me3-altered genes are SYD direct targets (Figure 30a and b), suggesting that SYD can directly affect local levels of H3K27me3 at specific loci. For example, it was found that SYD and CLF both target *BELLI (BEL1)* that encodes a homeodomain protein required for ovule identity, and loss of *SYD* activity leads to the gain of H3K27me3 at this locus (Figure 30c), suggesting that SYD functions antagonistically with PcG at this locus. Consistent with this notion, previous work from our laboratory revealed that BRM antagonizes PcG repression through physically preventing PRC2 occupancy at *BEL1* locus (Li *et al.*, 2015). In contrast, *MIR156* locus was an example showing the synergistic interaction of SYD and PcG in regulating the level of H3K27me3. *MIR156* has been shown to be a master regulator of the juvenile-to-adult phase transition in plants (Wu *et al.*, 2009). It was shown that SYD and CLF co-target *MIR156A/C*, and loss-of-function of *SYD* results in a decrease in H3K27me3 level at these loci (Figure 30c), implying that SYD functions synergistically with PcG to regulate the juvenile-to-adult transition in plants. To summarize, loss-of-function of *SYD* results in changes in H3K27me3 levels over several hundred genes, suggesting that SYD may antagonize and/or collaborate with PcG at specific loci across the *Arabidopsis* genome.

**Figure 29** Representative genes showing changes in H3K27me3 levels in *syd-5*.

(a) Genome browser views showing changes in H3K27me3 levels at representative genes in *syd-5* (red) compared to WT (blue). Five genes (upper panel) show an increase and four (bottom panel) show a decrease in H3K27me3 levels. *ACT7* and *AG* are used as control loci that exhibit no change in H3K27me3 levels in *syd-5* compared to WT. Gene structures are shown in green underneath each panel. (b) ChIP-qPCR validation. ChIP signals are shown as percentage of input. Error bars show standard deviations among three biological replicates. Student's *t*-test, \**P* < 0.05, \*\**P* < 0.01.

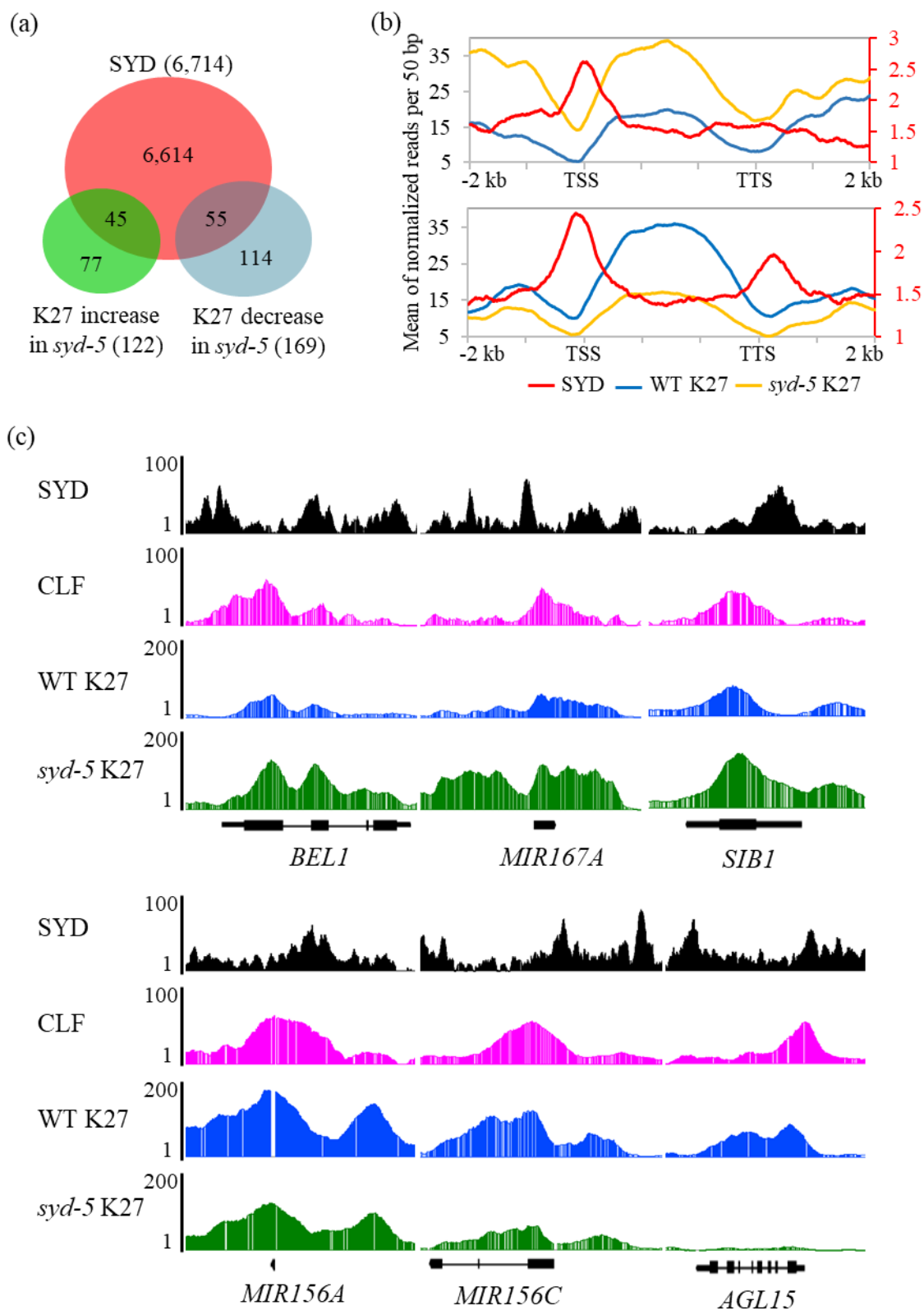
*SAG201*: SENESCENCE-ASSOCIATED GENE 201; *LGO*: LOSS OF GIANT CELLS FROM ORGANS; *RTFL16*: ROTUNDIFOLIA LIKE 16; *SCPL35*: SERINE CARBOXYPEPTIDASE-LIKE 35; *AGL15*: AGAMOUS-LIKE 15; *SMZ*: SCHLAFMUTZE; *AGP11*: ARABINOGALACTAN PROTEIN 11; *NAC049*: NAC DOMAIN CONTAINING PROTEIN 49; *ACT7*: ACTIN 7; *AG*: AGAMOUS; AT4G18282 is an uncharacterized gene.



**Figure 30** Antagonistic and synergistic functions between SYD and CLF at representative genes.

(a) Venn diagram showing overlaps between genes targeted by SYD and genes with altered changes in H3K27me3 levels in *syd-5* compared to WT. (b) Average enrichment of SYD occupancy and H3K27me3 levels in WT (WT K27) and *syd-5* (*syd-5* K27). Plot on the top shows 45 genes occupied by SYD and an increase in H3K27me3 levels in *syd-5* compared to WT. Plot for the remaining 55 genes (occupied by SYD and a decrease in H3K27me3 levels in *syd-5* compared to WT) is shown on the bottom. (c) Genome browser views showing occupancies of SYD (black) and CLF (pink), as well as changes in H3K27me3 levels at representative genes in *syd-5* (*syd-5* K27; green) compared to WT (WT K27; blue). Gene structures are shown in black underneath each panel.

*BEL1*: *BELL1*; *MIR167A*: *MICRORNA 167A*; *SIB1*: *SIGMA FACTOR BINDING PROTEIN 1*; *MIR156A/C*: *MICRORNA 156A/C*; *AGL15*: *AGAMOUS-LIKE 15*.





### 3.3.2 Increased occupancy of CLF/SWN at target loci in *brm-1*

Previous work from our laboratory showed that BRM prevents the occupancy of CLF/SWN at several genes examined by ChIP-qPCR (Li *et al.*, 2015). However, how BRM interplays with CLF/SWN at the genome-wide scale is not fully understood. Thus, ChIP-seq was performed to determine the global changes of CLF and SWN occupancy in *brm-1* background. First, *GFP*-tagged *CLF* and *SWN* transgenic lines (see section 3.1.1) were crossed with *brm-1*<sup>+/-</sup> plants, respectively (see section 2.2). Then *brm-1 clf-29 pCLF::CLF-GFP* and *brm-1 swn-4 pSWN::SWN-GFP* plants were identified by PCR-based genotyping (see section 2.5) from the F3 progeny. Second, ChIP experiments were performed by using two-week-old *brm-1 clf-29 pCLF::CLF-GFP* and *brm-1 swn-4 pSWN::SWN-GFP* seedlings and the ChIP DNA was sequenced (see section 2.7-2.9).

Compared to the occupancy of CLF described in section 3.1.2, loss-of-function of *BRM* activity led to an increase and a decrease of CLF occupancy for 100 and 9 genes, respectively (Figure 31a). Then the occupancy of SWN in *swn-4 pSWN::SWN-GFP* and *brm-1 swn-4 pSWN::SWN-GFP* was compared. It was observed that, in the absence of *BRM*, the occupancy of SWN is increased at 85 genes and decreased at 25 genes (Figure 31a). For example, the occupancies of CLF and SWN are dramatically increased at *SVP* locus and decreased at *WRKY DNA-BINDING PROTEIN 23 (WRKY23)* locus when *BRM* is absent (Figure 31b), further confirming the previous results determined by ChIP-qPCR (Li *et al.*, 2015). Thus, loss-of-function of *BRM* contributes to increased occupancies of CLF and SWN at a small number of genes (100 for CLF; 85 for SWN) genome-wide, suggesting that BRM antagonizes PcG repression by preventing PcG's occupancy on these genes.

**Figure 31** Altered occupancies of CLF and SWN in *brm-1*.

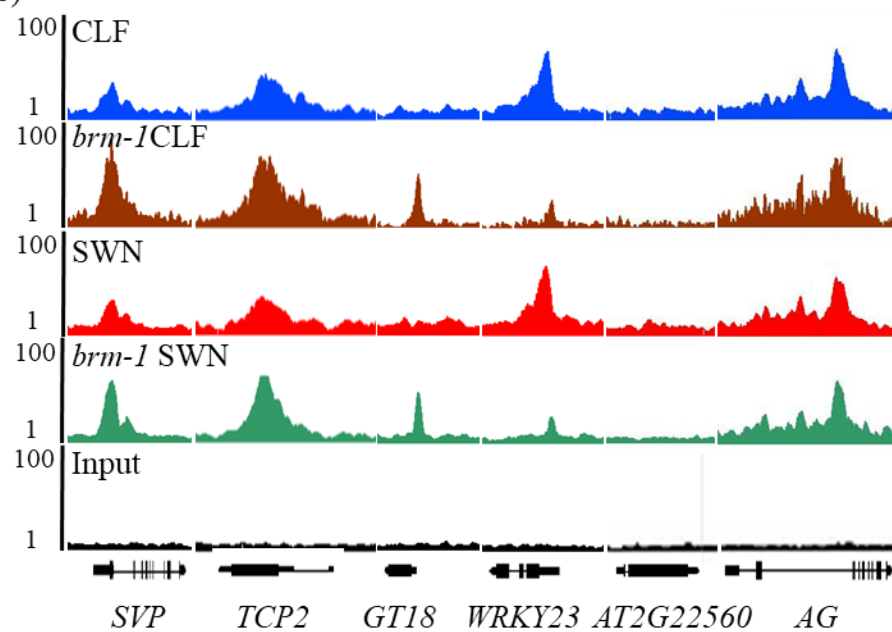
(a) Table summarizing the number of genes showing altered enrichment of CLF and SWN during lost *BRM* activity at the genome-wide scale. (b) Genome browser views showing occupancies of CLF (*clf-29 pCLF::CLF-GFP*), *brm-1* CLF (*brm-1 clf-29 pCLF::CLF-GFP*), SWN (*swn-4 pSWN::SWN-GFP*), and *brm-1* SWN (*brm-1 swn-4 pSWN::SWN-GFP*) at representative genes. Input signals are used as negative controls. *AT2G22560* and *AG* are used as control loci exhibiting no occupancy signals and strong signals of CLF and SWN, respectively. Gene structures are shown in black underneath the panel.

*SVP*: *SHORT VEGETATIVE PHASE*; *TCP2*: *TEOSINTE BRANCHED 2*; *GT18*: *GLYCOSYLTRANSFERASE 18*; *WRKY23*: *WRKY DNA-BINDING PROTEIN 23*; *AG*: *AGAMOUS*; *AT2G22560* is an uncharacterized gene.

(a)

Protein	Enrichment increase in <i>brm-1</i>	Enrichment decrease in <i>brm-1</i>
CLF	100	9
SWN	85	25

(b)



## 4 DISCUSSION

### 4.1 Genome-wide characterization of CLF and SWN

In this study, the *CLF*- and *SWN-GFP* fusion transgenic *Arabidopsis* lines, driven by their corresponding native promoters, were generated (described in section 3.1.1). These lines can be used to study PcG function in many developmental processes and stress response pathways in *Arabidopsis*. In addition, the CLF and SWN ChIP-seq data (section 3.1.2), together with the H3K27me3 ChIP-seq (section 3.1.5) and transcriptome data (section 3.1.6), generated in this study, are valuable resources for the plant epigenetics research community to further decipher PcG activities and their regulation during plant growth and development. In the following sections, I will highlight and discuss several interesting findings that provide new biological insights into the mechanisms underlying PcG activities, particularly the functional interplay between CLF and SWN.

#### 4.1.1 CLF and SWN are the only active H3K27 methyltransferases at the seedling stage

Previously published data have indicated that CLF and SWN are probably the only two H3K27 methyltransferases capable of catalyzing H3K27me3 at the seedling stage (Lafos *et al.*, 2011; Zhou *et al.*, 2017). In these reports, no H3K27me3 was detected in *clf swn* by Western blot. In this thesis, H3K27me3 was completely lost in the *clf-29 swn-4* double mutants as determined by ChIP-seq (Figure 15; section 3.1.5), suggesting that CLF and SWN are the only active H3K27 methyltransferases at the seedling stage. Since CLF and SWN are required to catalyze H3K27me3 during the seedling stage, it is therefore expected that H3K27me3-marked genes are occupied by CLF and/or SWN. However, my ChIP-seq data presented in this thesis show that the majority (5,208 out of 6,854) of H3K27me3-marked genes are surprisingly not occupied by either CLF or SWN (Figure 13). It was also reported that many genes are marked by H3K27me3 without PRC2 occupancy by comparing the genome-wide occupancies of H3K27me3 and PRC2 in *Drosophila* (Schwartz *et al.*, 2006). This apparent discrepancy/mystery can be explained possibly by technical and/or biological reasons. Technically, it is very hard to detect enrichment signals by ChIP-seq at loci where the abundance of CLF/SWN is very low. A biological

explanation for the low CLF and SWN occupancy over the genes that are occupied by H3K27me3 could be that CLF and SWN are recruited to chromatin transiently, and might leave the occupied regions as soon as the H3K27me3 is deposited. Therefore, in support of this hypothesis, a fluorescence recovery after photobleaching (FRAP) study in *Drosophila*, found that PcG occupancy is dynamic at analyzed loci in the genome (Ficz *et al.*, 2005). It is also necessary to bear in mind that ChIP-seq profiling is a snapshot by capturing the occupancy at the time of tissue fixation, whereas the H3K27me3 observed represents a culmination of PcG activity during the plant development to the time of tissue fixation. Thus, the occupancies of CLF and SWN cannot be recorded for a single-time point.

#### **4.1.2 CLF and SWN play both redundant and differential roles in depositing H3K27me3**

In this thesis, another interesting finding is the differential roles of CLF and SWN in depositing H3K27me3 at the same locus. It was observed that the genome-wide levels of H3K27me3 are lower in *clf-29*, and the levels of H3K27me3 are similar in *swn-4* to those in WT (Figure 15; section 3.1.5). Furthermore, three patterns of H3K27me3 reduction were identified in *clf-29* compared to WT (Figure 18; section 3.1.5). In type I, it was discovered that some genomic regions where the loss of *CLF* leads to a marked reduction of H3K27me3 in the flanking regions of peaks but not in the peak summits, suggesting that CLF might help to spread H3K27me3 from a “core” area to flanking regions. This pattern of H3K27me3 reduction when *CLF* is lost has also been proposed earlier (Li and Cui, 2016; Wang *et al.*, 2016; Yuan *et al.*, 2016; Yang *et al.*, 2017). The H3K27me3 signal at the summit in *clf-29* mutants might be catalyzed by SWN. Indeed, the H3K27me3 signal at the summit completely disappeared in the *clf-29 swn-4* double mutant. Thus, SWN might play redundant roles in catalyzing H3K27me3 only at the peak summit, and not in the flanking regions. In type II, a partial reduction of H3K27me3 was found across the entire H3K27me3-marked regions in *clf-29* compared to WT, while the H3K27me3 is completely abolished in *clf-29 swn-4*, suggesting that both CLF and SWN are required to catalyze H3K27me3 at these regions. In type III, H3K27me3 is completely lost in *clf-29* and *clf-29 swn-4*, while the level of H3K27me3 is the same in *swn-4* as those in WT, indicating that only CLF, but not SWN, is required to catalyze H3K27me3 at these regions.

Taken together, these three patterns of H3K27me3 reduction suggest that CLF and SWN play either partially redundant or non-redundant roles in depositing H3K27me3 at specific regions genome-wide.

These observations present questions that warrant future investigation, such as the locus-specific recruitment of CLF and SWN, and their functional interplay in depositing H3K27me3 at specific loci. For locus-specific recruitment of PcG proteins, this study (Figure 8; section 3.1.3) and other recent studies have identified *GAGA*-like and *Telo*-box-like motifs that are enriched in PcG-occupied regions (Deng *et al.*, 2013; Molitor *et al.*, 2016; Xiao *et al.*, 2017; Zhou *et al.*, 2018), suggesting that PcG proteins may be recruited by GAFs and TRBs, respectively, to specific regions where *GAGA*-like and *Telo*-box-like motifs are present. Further studies are required to decipher the roles of these GAFs and TRBs in mediating genomic targeting of PcG proteins. One possible way to examine the role of these DNA motifs in recruiting PcG proteins is to use the CRISPR/Cas9 system to disrupt the motif *in vivo* and then check the enrichment levels of PcG proteins (Li *et al.*, 2018a). It is also crucial to identify the DNA sequence-specific players that mediate the locus-specific recruitment of PcG proteins.

#### **4.1.3 Implications of CLF and SWN in gene expression**

Transcriptome analysis revealed that the expression of many more genes is changed in the *clf-29 swn-4* double mutant than in either the *clf-29* or *swn-4* single mutants (Figure 19a; section 3.1.6). This observation is consistent with the proposed genetic redundancy of *CLF* and *SWN* and the recognized functions of PcG in repressing gene expression (Jiang *et al.*, 2008; Xiao and Wagner, 2015; Wang *et al.*, 2016; Mozgová *et al.*, 2017). Although many genes are clearly up-regulated in the *clf-29 swn-4* double mutant, the expression of many other genes remained unchanged (Figure 19b). Thus, there is no overall correlation between the loss of H3K27me3 and the up-regulation of gene expression. A similar observation was reported in a previous transcriptome analysis for *brm* and *ref6* mutants that many BRM and REF6 targets do not show changes in transcript levels compared to WT (Li *et al.*, 2015). A possible explanation might be that many other factors such as TFs are also involved with these genes, and the overall transcript levels may not be affected by

only one of them. In addition, there are genes showing down-regulation in the *clf-29 swm-4* mutants (Figure 19b), which are not likely directly affected by the loss of H3K27me<sub>3</sub>, but rather indirectly *via* some intermediate effectors.

#### 4.1.4 CLF regulates the floral transition

In this thesis, it was found that CLF delays flowering by directly targeting and repressing *AGL17*, *AGL19*, *AGL24*, and *SOC1* (Figure 20-22; section 3.1.7). It was reported that the flower activator *AGL19* is repressed by CLF and the de-repression of *AGL19* in *clf* is partially responsible for the early flowering phenotype of *clf* mutants (Schönrock *et al.*, 2006). Therefore, the data are consistent with the previous report. In addition, the floral regulators *AGL24* and *SOC1* directly upregulate each other, thereby providing a positive-loop during the floral transition (Liu *et al.*, 2008). This work showed that CLF suppresses *AGL24* and *SOC1* by directly targeting these two genes, adding the H3K27me<sub>3</sub>-mediated regulation into the positive-loop of *AGL24* and *SOC1*. Moreover, *AGL17* was shown to accelerate flowering in the photoperiod pathway (Han *et al.*, 2008). This work first determined that CLF directly targets *AGL17*, thereby down-regulating its expression to delay the floral transition. Altogether, this work suggests a critical role for CLF in the H3K27me<sub>3</sub>-mediated floral transition.

Further, it was found that the two flowering repressor-encoding genes *FLC* and *SVP* show up- and down-regulation, respectively, in *clf-29* compared to WT (Figure 20; section 3.1.7). The increase in *FLC* expression was consistent with the reduction of H3K27me<sub>3</sub>, and the decrease in *SVP* could be (at least partially) due to the slight increase of H3K27me<sub>3</sub> level at this locus in *clf-29* (Figure 20). As *FLC* and *SVP* have been shown to work together by forming a repressor complex (Li *et al.* 2008; Mateos *et al.*, 2015), it is reasonable to speculate that the accumulation of the *FLC-SVP* complex is lower in *clf-29* compared to WT. Therefore, the *clf-29* mutant flowers earlier compared to WT regardless of *FLC* up-regulation. This is likely the case, as previous reports have shown that *SVP* represses flowering in a dosage-sensitive manner (Hartmann *et al.*, 2000; Li *et al.*, 2015). Notably, others and I have consistently observed an up-regulation of *FT* in the *clf-29* mutants (Figure 20; Jiang *et al.*, 2008; Farrona *et al.*, 2011b; Liu *et al.*, 2018). However, unlike the

previous studies, I did not detect CLF and SWN occupancies at this locus. This apparent inconsistency could be due to the use of different promoters in these independent studies, i.e., the *CLF-GFP* fusion gene used in this work was driven by the native *CLF* promoter, rather than the constitutive promoters used in the other studies.

## 4.2 Global characterization of SYD

ChIP-seq and RNA-seq were performed to reveal the functions of SYD in *Arabidopsis*. It was found that SYD directly targets nearly 25% of protein-coding genes in the genome, suggesting critical roles for SYD in plants. By taking advantage of published BRM genome-wide targets (Li *et al.*, 2016), I found that SYD and BRM co-target over three thousand genes, indicating that SYD and BRM may function redundantly at these co-targets. Further, the genome-wide occupancy of SYD (section 3.2.2), together with the transcriptome profiling (section 3.2.4), suggests that SYD mainly acts to activate its targets genome-wide.

### 4.2.1 A truncated SYD fully rescues *syd-5* phenotype

In section 3.2.1, it was shown that a truncated SYD, lacking the large C-terminal domain, can fully rescue the *syd-5* null mutant (Figure 23), which is consistent with the results of a previous report (Su *et al.*, 2006). Hence, the truncated SYD may be sufficient for biological activity. However, the ratio of the full-length and truncated SYD protein is dynamic in development (Su *et al.*, 2006), indicating a potential developmentally regulated process involved in generating truncated SYD. Su *et al.* (2006) proposed that an endopeptidase may cleave the full-length SYD to yield the truncated version, as it was shown that alternative splicing does not contribute to this process. Endoproteolysis provides a molecular mechanism for generating biologically active proteins from inactive polypeptide precursors in eukaryotes (Seidah and Chrétien, 1997; Kageyama, 2002). For example, the maize histone deacetylase ZmHDA1 is processed from an enzymatically inactive protein to the active truncated form by proteolytic removal of its C-terminal domain (Kageyama, 2002). Human MIXED-LINEAGE LEUKEMIA (MLL), a large protein homologous to *Drosophila* Trithorax group (TrxG), is cleaved by an endopeptidase Taspase1 (threonine aspartase 1) into N- and C-terminal fragments (Hsieh *et al.*, 2003). To investigate whether



the full-length SYD is active and the differences exist between the two forms of SYD, I successfully generated a transgenic *Arabidopsis* line expressing full-length SYD driven by its native regulatory region (a ~18 kb genomic fragment). Despite several attempts, I was unable to profile the genome-wide occupancy of full-length SYD by ChIP-seq, likely due to its large size (approximately 420 kDa). Therefore, I then decided to determine the genome-wide targets of SYD by using the *syd-5 pSYD::SYDΔC-GFP* transgenic plants instead.

#### **4.2.2 SYD may mainly function as a transcription activator**

By examining the transcriptome profiles in *syd-5* and WT, it was found that 601 and 1,535 genes are up- and down-regulated, respectively, in *syd-5* compared to WT (Figure 26a; section 3.2.4). Furthermore, it was shown that the signals of SYD occupancy are correlated with transcript levels, with higher signals corresponding to higher transcript levels (Figure 26c). In addition, more SYD directly targeted genes are down-regulated (522) than those up-regulated (233) in *syd-5* compared to WT (Figure 26d). Therefore, these data suggest that SYD targets genes to mainly activate their expression genome-wide.

#### **4.2.3 SYD and BRM share similar occupancy patterns**

In section 3.2.2, 6,714 genes were identified to be targeted by SYD at the seedling stage in *Arabidopsis* (Figure 24). Comparing the global occupancies of SYD and BRM (Li *et al.*, 2016), it was found that SYD and BRM share 3,387 co-targets (Figure 27b; section 3.2.5), indicating SYD and BRM may function redundantly on these common targets and differentially on their unique targets. It also gives clues for the functional interplay between these two homologs in plants. Further, the majority of the common genes (2,805 out of 3,387) were not differentially expressed in *syd-5* or *brm-1* compared to WT (Figure 27c), consistent with the genetic redundancy between *SYD* and *BRM* in *Arabidopsis* reported previously (Bezhani *et al.*, 2007; Wu *et al.*, 2015).

In section 3.2.3, it was shown that SYD occupancy peaks have two general patterns: single narrow or clustered broad peaks (Figure 25b). Analysis of the genome-wide signals of SYD occupancy showed that SYD mainly enriches around TSSs (Figure 25d). These

patterns show similarities with the published genome-wide patterns for BRM (Li *et al.*, 2016; Archacki *et al.*, 2017). Li *et al.* (2016) also reported that BRM occupancy peaks have narrow and broad patterns. However, the biological relevance of narrow and broad peaks is not clear. In addition, the occupancy of BRM was also found to be enriched near TSSs (Li *et al.*, 2016; Archacki *et al.*, 2017). These similarities between SYD and BRM are consistent with the expected conserved functions between these two homologs in plants.

### 4.3 Antagonistic and synergistic interactions between SWI/SNF and PcG

Early studies in *Drosophila* proposed the antagonistic functions of TrxG and PcG proteins that are generally considered to be required for activating and repressing their target genes, respectively (Brock and Fisher, 2005; Kadoch and Crabtree, 2015). To date, several putative TrxG proteins have been proposed in *Arabidopsis*, including SWI/SNF ATPases SYD and BRM (Wu *et al.*, 2012; Li *et al.*, 2015, 2016; Xu *et al.*, 2016), CHD3-type chromatin remodeling factors PKL and PICKLE RELATED 2 (PKR2) (Aichinger *et al.*, 2009, 2011), SAND domain protein ULTRAPETALA 1 (ULT1) (Carles and Fletcher, 2009), histone H3K4 methyltransferase ARABIDOPSIS TRITHORAX 1 (ATX1) (Alvarez-Venegas *et al.*, 2003), and H3K27 demethylase REF6 (Lu *et al.*, 2011; Li *et al.*, 2016). However, the functional interplay between SWI/SNF ATPases (SYD and BRM) and PcG proteins (CLF and SWN) are not fully understood.

In section 3.3.1, it was shown that SYD and PcG can act either antagonistically or synergistically. When comparing the global distribution of H3K27me3 in WT and *syd* mutants, although there was no obvious overall genome-wide increase in H3K27me3 levels in *syd-5*, I did, however, detect a few hundred genes showing altered levels of H3K27me3 in the genome. Noticeably, almost half of these genes displayed an increased level of H3K27me3, indicating that SYD is able to not only antagonize PcG function but also cooperate with PcG at some genes. For example, as shown in Figure 30, the levels of H3K27m3 were either increased or decreased at SYD and CLF co-targeted genes in *syd-5* compared to WT, suggesting that SYD may either antagonize or enhance PcG-mediated H3K27me3 deposition, thereby specifically regulating gene expression.

Previous work from our laboratory found that 258 and 18 genes show an increase and a reduction, respectively, in the levels of H3K27me3 in *brm-1* compared to WT (Li *et al.*, 2015), suggesting that BRM mainly acts to antagonize PcG activity during vegetative development. In section 3.3.2, it was found that the occupancies of CLF and SWN are changed at a small number of genes (100 increase and 9 decrease for CLF occupancy; 85 increase and 25 decrease for SWN occupancy) when *BRM* function is lost at the genome-wide scale (Figure 31), suggesting that BRM mainly prevents PcG's occupancy at these specific genes. These findings are consistent with the published observation in other organisms (Ho *et al.*, 2011; Kadoch *et al.*, 2016; Schuettengruber *et al.*, 2017). For instance, a study in human embryonic stem cells found that the chromatin remodeling ATPase BRG1, a BRM homolog, regulates the activity of PcG in both antagonistic and synergistic modes across the genome (Ho *et al.*, 2011).

#### 4.4 Concluding remarks and future direction

In summary, the work presented in this thesis provides useful resources by determining the occupancies of the PcG proteins, CLF and SWN, as well as the SWI/SNF chromatin remodeler SYD during vegetative development in *Arabidopsis*. These genome-wide data can be useful for the plant biology community to further deciphering PcG and SWI/SNF activities.

In this thesis, I have shown that CLF and SWN target many common genes, but it is necessary to keep in mind that the occupancy signals were captured from whole seedlings (i.e., mixed cells/tissues). Therefore, it is not clear whether CLF and SWN target the same chromatin regions. It would be worthwhile, in the future, to explore whether CLF and SWN target same genes in the same region of chromatin by undertaking a sequential ChIP (also called ChIP-re-ChIP) experiment (Beischlag *et al.*, 2018), where one protein is precipitated followed by a second round of precipitation of another protein. Thus, ChIP-re-ChIP can be performed to show that CLF and SWN occupy the same region of chromatin. Furthermore, it will be interesting to identify the unknown factors that contribute to different morphological defects of *clf-29* and *swn-4*. One proposed approach is to determine CLF- and SWN-interacting proteins by conducting immunoprecipitation

followed by mass spectrometry (IP-MS) experiments. In addition, it was shown, based on the changes in H3K27me3 levels in *syd-5* relative to WT, that SYD may antagonize and/or cooperate with PcG at specific loci genome-wide. It will be of importance to investigate whether SYD impairs and/or enhances PcG occupancy at these loci by examining PcG occupancy in a *syd-5* genetic background.

## References

- Aichinger, E., Villar, C.B., Farrona, S., Reyes, J.C., Hennig, L. and Kohler, C.** (2009) CHD3 proteins and Polycomb group proteins antagonistically determine cell identity in *Arabidopsis*. *PLOS Genetics*, **5**, e1000605.
- Aichinger, E., Villar, C.B., Di Mambro, R., Sabatini, S. and Kohler, C.** (2011) The CHD3 chromatin remodeler PICKLE and Polycomb group proteins antagonistically regulate meristem activity in the *Arabidopsis* root. *The Plant Cell*, **23**, 1047-1060.
- Alvarez-Venegas, R., Pien, S., Sadler, M., Witmer, X., Grossniklaus, U. and Avramova, Z.** (2003) ATX-1, an *Arabidopsis* homolog of Trithorax, activates flower homeotic genes. *Current Biology*, **13**, 627-637.
- Araki, T.** (2001) Transition from vegetative to reproductive phase. *Current Opinion in Plant Biology*, **4**, 63-68.
- Archacki, R., Yatusovich, R., Buszewicz, D., Krzyczmonik, K., Patryn, J., Iwanicka-Nowicka, R., Biecek, P., Wilczynski, B., Koblowska, M., Jerzmanowski, A. and Swiezewski, S.** (2017) *Arabidopsis* SWI/SNF chromatin remodeling complex binds both promoters and terminators to regulate gene expression. *Nucleic Acids Research*, **45**, 3116-3129.
- Ariel, F., Jegu, T., Latrasse, D., Romero-Barrios, N., Christ, A., Benhamed, M. and Crespi, M.** (2014) Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Molecular Cell*, **55**, 383-396.
- Ausin, I., Alonso-Blanco, C., Jarillo, J.A., Ruiz-Garcia, L. and Martinez-Zapater, J.M.** (2004) Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nature Genetics*, **36**, 162-166.
- Bannister, A.J. and Kouzarides, T.** (2011) Regulation of chromatin by histone modifications. *Cell Research*, **21**, 381-395.
- Bari, R. and Jones, J.D.G.** (2009) Role of plant hormones in plant defence responses. *Plant Molecular Biology*, **69**, 473-488.
- Barisic, D., Stadler, M.B., Iurlaro, M. and Schübeler, D.** (2019) Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature*, **569**, 136-140.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K.** (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823-837.

- Bäurle, I. and Dean, C.** (2006) The timing of developmental transitions in plants. *Cell*, **125**, 655-664.
- Becker, J.S., Nicetto, D. and Zaret, K.S.** (2016) H3K9me3-dependent heterochromatin: barrier to cell fate changes. *Trends in Genetics*, **32**, 29-41.
- Beischlag, T.V., Prefontaine, G.G. and Hankinson, O.** (2018) ChIP-re-ChIP: co-occupancy analysis by sequential chromatin immunoprecipitation. In *Chromatin Immunoprecipitation: Methods and Protocols*. New York, NY: Springer New York, pp. 103-112.
- Berger, N., Dubreucq, B., Roudier, F., Dubos, C. and Lepiniec, L.** (2011) Transcriptional regulation of *Arabidopsis* LEAFY COTYLEDON 2 involves RLE, a *cis*-element that regulates trimethylation of histone H3 at lysine 27. *The Plant Cell*, **23**, 4065-4078.
- Berry, S. and Dean, C.** (2015) Environmental perception and epigenetic memory: mechanistic insight through FLC. *The Plant Journal*, **83**, 133-148.
- Bezhani, S., Winter, C., Hershman, S., Wagner, J.D., Kennedy, J.F., Kwon, C.S., Pfluger, J., Su, Y. and Wagner, D.** (2007) Unique, shared, and redundant roles for the *Arabidopsis* SWI/SNF chromatin remodeling ATPases BRAHMA and SPLAYED. *The Plant Cell*, **19**, 403-416.
- Blackledge, N.P., Rose, N.R. and Klose, R.J.** (2015) Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nature Reviews. Molecular Cell Biology*, **16**, 643-649.
- Boss, P.K., Bastow, R.M., Mylne, J.S. and Dean, C.** (2004) Multiple pathways in the decision to flower: enabling, promoting, and resetting. *The Plant Cell*, **16**, S18-S31.
- Bouveret, R., Schönrock, N., Gruissem, W. and Hennig, L.** (2006) Regulation of flowering time by *Arabidopsis* *MSH1*. *Development*, **133**, 1693-1702.
- Brock, H.W. and Fisher, C.L.** (2005) Maintenance of gene expression patterns. *Developmental Dynamics*, **232**, 633-655.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S. and Zhang, Y.** (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, **298**, 1039-1043.
- Cao, R. and Zhang, Y.** (2004) The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Current Opinion in Genetics & Development*, **14**, 155-164.
- Carles, C.C. and Fletcher, J.C.** (2009) The SAND domain protein ULTRAPETALA 1 acts as a Trithorax group factor to regulate cell fate in plants. *Genes & Development*, **23**, 2723-2728.

- Carter, B., Bishop, B., Ho, K.K., Huang, R., Jia, W., Zhang, H., Pascuzzi, P.E., Deal, R.B. and Ogas, J.** (2018) The chromatin remodelers PKL and PIE1 act in an epigenetic pathway that determines H3K27me3 homeostasis in *Arabidopsis*. *The Plant Cell*, **30**, 1337-1352.
- Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.H., Sung, Z.R. and Goodrich, J.** (2004) Interaction of Polycomb group proteins controlling flowering in *Arabidopsis*. *Development*, **131**, 5263-5276.
- Chen, C., Li, C., Wang, Y., Renaud, J., Tian, G., Kambhampati, S., Saatian, B., Nguyen, V., Hannoufa, A., Marsolais, F., Yuan, Z.C., Yu, K., Austin, R.S., Liu, J., Kohalmi, S.E., Wu, K., Huang, S. and Cui, Y.** (2017) Cytosolic acetyl-CoA promotes histone acetylation predominantly at H3K27 in *Arabidopsis*. *Nature Plants*, **3**, 814-824.
- Chen, T.W., Li, H.P., Lee, C.C., Gan, R.C., Huang, P.J., Wu, T.H., Lee, C.Y., Chang, Y.F. and Tang, P.** (2014) ChIPseeker, a web-based analysis tool for ChIP data. *BMC Genomics*, **15**, 539.
- Chen, X.** (2004) A microRNA as a translational repressor of *APETALA 2* in *Arabidopsis* flower development. *Science*, **303**, 2022-2025.
- Chiang, G.C.K., Barua, D., Kramer, E.M., Amasino, R.M. and Donohue, K.** (2009) Major flowering time gene, *FLOWERING LOCUS C*, regulates seed germination in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of United States of America*, **106**, 11661-11666.
- Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebbersold, R. and Nogales, E.** (2012) Molecular architecture of human Polycomb repressive complex 2. *eLife*, **1**, e00005.
- Clapier, C.R. and Cairns, B.R.** (2009) The biology of chromatin remodeling complexes. *Annual Review of Biochemistry*, **78**, 273-304.
- Clarke, J.H. and Dean, C.** (1994) Mapping FRI, a locus controlling flowering time and vernalization response in *Arabidopsis thaliana*. *Molecular and General Genetics*, **242**, 81-89.
- Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C. and Coupland, G.** (2007) FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science*, **316**, 1030-1033.
- Cui, X., Lu, F., Qiu, Q., Zhou, B., Gu, L., Zhang, S., Kang, Y., Cui, X., Ma, X., Yao, Q., Ma, J., Zhang, X. and Cao, X.** (2016) REF6 recognizes a specific DNA sequence to demethylate H3K27me3 and regulate organ boundary formation in *Arabidopsis*. *Nature Genetics*, **48**, 694-699.

- Curtis, M.D. and Grossniklaus, U.** (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiology*, **133**, 462-469.
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K. and Scheible, W.R.** (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiology*, **139**, 5-17.
- da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., Kaneko, S., Helin, K., Reinberg, D., Stewart, A.F., Wutz, A., Margueron, R. and Heard, E.** (2014) Jarid2 is implicated in the initial Xist-induced targeting of PRC2 to the inactive X chromosome. *Molecular Cell*, **53**, 301-316.
- Davière, J. M. and Achard, P.** (2013) Gibberellin signaling in plants. *Development*, **140**, 1147-1151.
- Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C.R. and Pirrotta, V.** (2004) Polycomb silencing blocks transcription initiation. *Molecular Cell*, **13**, 887-893.
- Den Dulk-Ras, A. and Hooykaas, P.J.J.** (1995) Electroporation of *Agrobacterium tumefaciens*. In *Plant Cell Electroporation and Electrofusion Protocols* (Nickoloff, J.A. ed. Totowa, NJ: Springer New York, pp. 63-72.
- Deng, W., Buzas, D.M., Ying, H., Robertson, M., Taylor, J., Peacock, W.J., Dennis, E.S. and Helliwell, C.** (2013) *Arabidopsis* Polycomb repressive complex 2 binding sites contain putative GAGA factor binding motifs within coding regions of genes. *BMC Genomics*, **14**, 593.
- Derkacheva, M. and Hennig, L.** (2014) Variations on a theme: Polycomb group proteins in plants. *Journal of Experimental Botany*, **65**, 2769-2784.
- Dillon, S.C., Zhang, X., Trievel, R.C. and Cheng, X.** (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biology*, **6**, 227.
- Euskirchen, G., Auerbach, R.K. and Snyder, M.** (2012) SWI/SNF chromatin-remodeling factors: multiscale analyses and diverse functions. *The Journal of Biological Chemistry*, **287**, 30897-30905.
- Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H. and Karch, F.** (1994) The Trithorax-like gene encodes the *Drosophila* GAGA factor. *Nature*, **371**, 806-808.
- Farrona, S., Hurtado, L., Bowman, J.L. and Reyes, J.C.** (2004) The *Arabidopsis thaliana* SNF2 homolog AtBRM controls shoot development and flowering. *Development*, **131**, 4965-4975.



- Farrona, S., Hurtado, L., March-Diaz, R., Schmitz, R.J., Florencio, F.J., Turck, F., Amasino, R.M. and Reyes, J.C.** (2011a) Brahma is required for proper expression of the floral repressor FLC in *Arabidopsis*. *PLOS One*, **6**, e17997.
- Farrona, S., Thorpe, F.L., Engelhorn, J., Adrian, J., Dong, X., Sarid-Krebs, L., Goodrich, J. and Turck, F.** (2011b) Tissue-specific expression of *FLOWERING LOCUS T* in *Arabidopsis* is maintained independently of Polycomb group protein repression. *The Plant Cell*, **23**, 3204-3214.
- Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M.F., Kater, M.M. and Colombo, L.** (2003) MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *The Plant Cell*, **15**, 2603-2611.
- Ficz, G., Heintzmann, R. and Arndt-Jovin, D.J.** (2005) Polycomb group protein complexes exchange rapidly in living *Drosophila*. *Development*, **132**, 3963-3976.
- Flaus, A., Martin, D.M., Barton, G.J. and Owen-Hughes, T.** (2006) Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Research*, **34**, 2887-2905.
- Fornara, F., de Montaigu, A. and Coupland, G.** (2010) SnapShot: control of flowering in *Arabidopsis*. *Cell*, **141**, 550-550.
- Gan, E.S., Xu, Y. and Ito, T.** (2015) Dynamics of H3K27me3 methylation and demethylation in plant development. *Plant Signaling & Behavior*, **10**, e1027851.
- Gaspin, C., Rami, J.F. and Lescure, B.** (2010) Distribution of short interstitial telomere motifs in two plant genomes: putative origin and function. *BMC Plant Biology*, **10**, 283.
- Gendall, A.R., Levy, Y.Y., Wilson, A. and Dean, C.** (2001) The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. *Cell*, **107**, 525-535.
- Gendrel, A.V., Lippman, Z., Martienssen, R. and Colot, V.** (2005) Profiling histone modification patterns in plants using genomic tiling microarrays. *Nature Methods*, **2**, 213-218.
- Geraldo, N., Bäurle, I., Kidou, S.I., Hu, X. and Dean, C.** (2009) FRIGIDA delays flowering in *Arabidopsis* via a cotranscriptional mechanism involving direct interaction with the nuclear cap-binding complex. *Plant Physiology*, **150**, 1611-1618.
- Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P. and Bickmore, W.A.** (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell*, **118**, 555-566.

- Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M. and Coupland, G.** (1997) A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature*, **386**, 44-51.
- Han, P., Garcia-Ponce, B., Fonseca-Salazar, G., Alvarez-Buylla, E.R. and Yu, H.** (2008) *AGAMOUS-LIKE 17*, a novel flowering promoter, acts in a *FT*-independent photoperiod pathway. *The Plant Journal*, **55**, 253-265.
- Han, S.K., Wu, M.F., Cui, S. and Wagner, D.** (2015) Roles and activities of chromatin remodeling ATPases in plants. *The Plant Journal*, **83**, 62-77.
- Han, S.K., Sang, Y., Rodrigues, A., Biol, F., Wu, M.F., Rodriguez, P.L. and Wagner, D.** (2012) The SWI2/SNF2 chromatin remodeling ATPase BRAHMA represses abscisic acid responses in the absence of the stress stimulus in *Arabidopsis*. *The Plant Cell*, **24**, 4892-4906.
- Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M. and Helin, K.** (2008) A model for transmission of the H3K27me3 epigenetic mark. *Nature Cell Biology*, **10**, 1291-1300.
- Hargreaves, D.C. and Crabtree, G.R.** (2011) ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Research*, **21**, 396-420.
- Hartmann, U., Höhmann, S., Nettesheim, K., Wisman, E., Saedler, H. and Huijser, P.** (2000) Molecular cloning of *SVP*: a negative regulator of the floral transition in *Arabidopsis*. *The Plant Journal*, **21**, 351-360.
- Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., Antosiewicz-Bourget, J., Ye, Z., Espinoza, C., Agarwahl, S., Shen, L., Ruotti, V., Wang, W., Stewart, R., Thomson, J.A., Ecker, J.R. and Ren, B.** (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, **6**, 479-491.
- He, Y., Michaels, S.D. and Amasino, R.M.** (2003) Regulation of flowering time by histone acetylation in *Arabidopsis*. *Science*, **302**, 1751-1754.
- He, C., Chen, X., Huang, H. and Xu, L.** (2012) Reprogramming of H3K27me3 is critical for acquisition of pluripotency from cultured *Arabidopsis* tissues. *PLOS Genetics*, **8**, e1002911.
- Helliwell, C.A., Wood, C.C., Robertson, M., James Peacock, W. and Dennis, E.S.** (2006) The *Arabidopsis* FLC protein interacts directly *in vivo* with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. *The Plant Journal*, **46**, 183-192.
- Heo, J.B. and Sung, S.** (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*, **331**, 76-79.

- Herz, H.M., Garruss, A. and Shilatifard, A.** (2013) SET for life: biochemical activities and biological functions of SET domain-containing proteins. *Trends in Biochemical Sciences*, **38**, 621-639.
- Ho, L., Miller, E.L., Ronan, J.L., Ho, W.Q., Jothi, R. and Crabtree, G.R.** (2011) esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating Polycomb function. *Nature Cell Biology*, **13**, 903-913.
- Hodgson, J.W., Argiropoulos, B. and Brock, H.W.** (2001) Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid Polycomb group response element-dependent silencing. *Molecular and Cellular Biology*, **21**, 4528-4543.
- Hou, X., Zhou, J., Liu, C., Liu, L., Shen, L. and Yu, H.** (2014) Nuclear factor Y-mediated H3K27me3 demethylation of the SOC1 locus orchestrates flowering responses of *Arabidopsis*. *Nature Communications*, **5**, 4601.
- Howe, F.S., Fischl, H., Murray, S.C. and Mellor, J.** (2017) Is H3K4me3 instructive for transcription activation? *BioEssays*, **39**, e201600095.
- Hsieh, J.J.D., Cheng, E.H.Y. and Korsmeyer, S.J.** (2003) Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell*, **115**, 293-303.
- Huang, H., Sabari, B.R., Garcia, B.A., Allis, C.D. and Zhao, Y.** (2014) SnapShot: histone modifications. *Cell*, **159**, 458-458.
- Huijser, P. and Schmid, M.** (2011) The control of developmental phase transitions in plants. *Development*, **138**, 4117-4129.
- Hurtado, L., Farrona, S. and Reyes, J.C.** (2006) The putative SWI/SNF complex subunit BRAHMA activates flower homeotic genes in *Arabidopsis thaliana*. *Plant Molecular Biology*, **62**, 291-304.
- Jiang, D., Wang, Y., Wang, Y. and He, Y.** (2008) Repression of *FLOWERING LOCUS C* and *FLOWERING LOCUS T* by the *Arabidopsis* Polycomb repressive complex 2 components. *PLOS One*, **3**, e3404.
- Jiao, L. and Liu, X.** (2015) Structural basis of histone H3K27 trimethylation by an active Polycomb repressive complex 2. *Science*, **350**.
- Johnson, K.C.M., Xia, S., Feng, X. and Li, X.** (2015) The chromatin remodeler SPLAYED negatively regulates SNC1-mediated immunity. *Plant and Cell Physiology*, **56**, 1616-1623.
- Jones, P.A.** (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, **13**, 484-492.

- Kadoch, C. and Crabtree, G.R.** (2015) Mammalian SWI/SNF chromatin remodeling complexes and cancer: mechanistic insights gained from human genomics. *Science Advances*, **1**, e1500447.
- Kadoch, C., Copeland, R.A. and Keilhack, H.** (2016) PRC2 and SWI/SNF chromatin remodeling complexes in health and disease. *Biochemistry*, **55**, 1600-1614.
- Kageyama, T.** (2002) Pepsinogens, progastricsins, and prochymosins: structure, function, evolution, and development. *Cellular and Molecular Life Sciences CMLS*, **59**, 288-306.
- Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J. and Weigel, D.** (1999) Activation tagging of the floral inducer FT. *Science*, **286**, 1962-1965.
- Katsani, K.R., Hajibagheri, M.A. and Verrijzer, C.P.** (1999) Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. *The EMBO Journal*, **18**, 698-708.
- Ketel, C.S., Andersen, E.F., Vargas, M.L., Suh, J., Strome, S. and Simon, J.A.** (2005) Subunit contributions to histone methyltransferase activities of fly and worm Polycomb group complexes. *Molecular and Cellular Biology*, **25**, 6857-6868.
- Kidder, B.L., Palmer, S. and Knott, J.G.** (2009) SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells*, **27**, 317-328.
- Kim, D.H. and Sung, S.** (2014) Polycomb-mediated gene silencing in *Arabidopsis thaliana*. *Molecules and Cells*, **37**, 841-850.
- Kim, H.J., Hyun, Y., Park, J.Y., Park, M.J., Park, M.K., Kim, M.D., Kim, H.J., Lee, M.H., Moon, J., Lee, I. and Kim, J.** (2004) A genetic link between cold responses and flowering time through FVE in *Arabidopsis thaliana*. *Nature Genetics*, **36**, 167-171.
- Knizewski, L., Ginalski, K. and Jerzmanowski, A.** (2008) Snf2 proteins in plants: gene silencing and beyond. *Trends in Plant Science*, **13**, 557-565.
- Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T. and Wutz, A.** (2004) A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biology*, **2**, E171.
- Kouzarides, T.** (2007) Chromatin modifications and their function. *Cell*, **128**, 693-705.
- Krajewski, W.A., Nakamura, T., Mazo, A. and Canaani, E.** (2005) A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. *Molecular and Cellular Biology*, **25**, 1891-1899.

- Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., Adli, M., Kasif, S., Ptaszek, L.M., Cowan, C.A., Lander, E.S., Koseki, H. and Bernstein, B.E.** (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLOS Genetics*, **4**, e1000242.
- Kwon, C.S., Chen, C. and Wagner, D.** (2005) WUSCHEL is a primary target for transcriptional regulation by SPLAYED in dynamic control of stem cell fate in *Arabidopsis*. *Genes & Development*, **19**, 992-1003.
- Kwon, C.S., Hibara, K.I., Pfluger, J., Bezhani, S., Metha, H., Aida, M., Tasaka, M. and Wagner, D.** (2006) A role for chromatin remodeling in regulation of CUC gene expression in the *Arabidopsis* cotyledon boundary. *Development*, **133**, 3223-3230.
- Lafos, M., Kroll, P., Hohenstatt, M.L., Thorpe, F.L., Clarenz, O. and Schubert, D.** (2011) Dynamic regulation of H3K27 trimethylation during *Arabidopsis* differentiation. *PLOS Genetics*, **7**, e1002040.
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A. and Huala, E.** (2012) The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research*, **40**, D1202-1210.
- Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L.** (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.
- Lawrence, M., Daujat, S. and Schneider, R.** (2016) Lateral thinking: how histone modifications regulate gene expression. *Trends in Genetics*, **32**, 42-56.
- Laugesen, A., Højfeldt, J.W. and Helin, K.** (2019) Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. *Molecular Cell*, **74**, 8-18.
- Levy, Y.Y., Mesnage, S., Mylne, J.S., Gendall, A.R. and Dean, C.** (2002) Multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control. *Science*, **297**, 243-246.
- Lewis, E.B.** (1978) A gene complex controlling segmentation in *Drosophila*. *Nature*, **276**, 565-570.
- Li, B., Carey, M. and Workman, J.L.** (2007) The role of chromatin during transcription. *Cell*, **128**, 707-719.
- Li, C., Chen, C., Chen, H., Wang, S., Chen, X. and Cui, Y.** (2018a) Verification of DNA motifs in *Arabidopsis* using CRISPR/Cas9-mediated mutagenesis. *Plant Biotechnology Journal*, **16**, 1446-1451.

- Li, C., Chen, C., Gao, L., Yang, S., Nguyen, V., Shi, X., Siminovitch, K., Kohalmi, S.E., Huang, S., Wu, K., Chen, X. and Cui, Y.** (2015) The *Arabidopsis* SWI2/SNF2 chromatin remodeler BRAHMA regulates Polycomb function during vegetative development and directly activates the flowering repressor gene *SVP*. *PLOS Genetics*, **11**, e1004944.
- Li, C. and Cui, Y.** (2016) A DNA element that remembers winter. *Nature Genetics*, **48**, 1451.
- Li, C., Gu, L., Gao, L., Chen, C., Wei, C.Q., Qiu, Q., Chien, C.W., Wang, S., Jiang, L., Ai, L.F., Chen, C.Y., Yang, S., Nguyen, V., Qi, Y., Snyder, M.P., Burlingame, A.L., Kohalmi, S.E., Huang, S., Cao, X., Wang, Z.Y., Wu, K., Chen, X. and Cui, Y.** (2016) Concerted genomic targeting of H3K27 demethylase REF6 and chromatin-remodeling ATPase BRM in *Arabidopsis*. *Nature Genetics*, **48**, 687-693.
- Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E. and Yu, H.** (2008) A repressor complex governs the integration of flowering signals in *Arabidopsis*. *Developmental Cell*, **15**, 110-120.
- Li, Z., Fu, X., Wang, Y., Liu, R. and He, Y.** (2018b) Polycomb-mediated gene silencing by the BAH-EMF1 complex in plants. *Nature Genetics*, **50**, 1254-1261.
- Liu, C., Chen, H., Er, H.L., Soo, H.M., Kumar, P.P., Han, J.H., Liou, Y.C. and Yu, H.** (2008) Direct interaction of *AGL24* and *SOC1* integrates flowering signals in *Arabidopsis*. *Development*, **135**, 1481-1491.
- Liu, X., Yang, Y., Hu, Y., Zhou, L., Li, Y. and Hou, X.** (2018) Temporal-specific interaction of NF-YC and CURLY LEAF during the floral transition regulates flowering. *Plant Physiology*, **177**, 105-114.
- Liu, X.M., An, J., Han, H.J., Kim, S.H., Lim, C.O., Yun, D.J. and Chung, W.S.** (2014) ZAT11, a zinc finger transcription factor, is a negative regulator of nickel ion tolerance in *Arabidopsis*. *Plant Cell Reports*, **33**, 2015-2021.
- Lodha, M., Marco, C.F. and Timmermans, M.C.** (2013) The ASYMMETRIC LEAVES complex maintains repression of KNOX homeobox genes *via* direct recruitment of Polycomb repressive complex 2. *Genes & Development*, **27**, 596-601.
- Lopez-Vernaza, M., Yang, S., Müller, R., Thorpe, F., de Leau, E. and Goodrich, J.** (2012) Antagonistic roles of *SEPALLATA 3*, *FT* and *FLC* genes as targets of the Polycomb group gene *CURLY LEAF*. *PLOS One*, **7**, e30715.
- Lu, F., Cui, X., Zhang, S., Jenuwein, T. and Cao, X.** (2011) *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase. *Nature Genetics*, **43**, 715-719.

- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J.** (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251-260.
- Luo, C., Sidote, D.J., Zhang, Y., Kerstetter, R.A., Michael, T.P. and Lam, E.** (2013) Integrative analysis of chromatin states in *Arabidopsis* identified potential regulatory mechanisms for natural antisense transcript production. *The Plant Journal*, **73**, 77-90.
- Lynch, M.D., Smith, A.J., De Gobbi, M., Flenley, M., Hughes, J.R., Vernimmen, D., Ayyub, H., Sharpe, J.A., Sloane-Stanley, J.A., Sutherland, L., Meek, S., Burdon, T., Gibbons, R.J., Garrick, D. and Higgs, D.R.** (2012) An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *The EMBO Journal*, **31**, 317-329.
- Machanick, P. and Bailey, T.L.** (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, **27**, 1696-1697.
- Macknight, R., Bancroft, I., Page, T., Lister, C., Schmidt, R., Love, K., Westphal, L., Murphy, G., Sherson, S., Cobbett, C. and Dean, C.** (1997) *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell*, **89**, 737-745.
- Margueron, R. and Reinberg, D.** (2011) The Polycomb complex PRC2 and its mark in life. *Nature*, **469**, 343-349.
- Martens, J.A. and Winston, F.** (2003) Recent advances in understanding chromatin remodeling by SWI/SNF complexes. *Current Opinion in Genetics & Development*, **13**, 136-142.
- Mashtalir, N., D'Avino, A.R., Michel, B.C., Luo, J., Pan, J., Otto, J.E., Zullo, H.J., McKenzie, Z.M., Kubiak, R.L., St. Pierre, R., Valencia, A.M., Poynter, S.J., Cassel, S.H., Ranish, J.A. and Kadoch, C.** (2018) Modular organization and assembly of SWI/SNF family chromatin remodeling complexes. *Cell*, **175**, 1272-1288.
- Mateos, J.L., Madrigal, P., Tsuda, K., Rawat, V., Richter, R., Romera-Branchat, M., Fornara, F., Schneeberger, K., Krajewski, P. and Coupland, G.** (2015) Combinatorial activities of SHORT VEGETATIVE PHASE and FLOWERING LOCUS C define distinct modes of flowering regulation in *Arabidopsis*. *Genome Biology*, **16**, 31.
- Michaels, S.D. and Amasino, R.M.** (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant Cell*, **11**, 949-956.

- Michaels, S.D., Ditta, G., Gustafson-Brown, C., Pelaz, S., Yanofsky, M. and Amasino, R.M.** (2003) *AGL24* acts as a promoter of flowering in *Arabidopsis* and is positively regulated by vernalization. *The Plant Journal*, **33**, 867-874.
- Mlynárová, L., Nap, J.P. and Bisseling, T.** (2007) The SWI/SNF chromatin-remodeling gene *AtCHR12* mediates temporary growth arrest in *Arabidopsis thaliana* upon perceiving environmental stress. *The Plant Journal*, **51**, 874-885.
- Molitor, A., Latrasse, D., Zytnicki, M., Andrey, P., Houba-Herlin, N., Hachet, M., Battail, C., Del Prete, S., Alberti, A., Quesneville, H. and Gaudin, V.** (2016) The *Arabidopsis* hnRNP-Q Protein LIF2 and the PRC1 subunit LHP1 function in concert to regulate the transcription of stress-responsive genes. *The Plant Cell*, **28**, 2197-2211.
- Monribot-Villanueva, J., Zurita, M. and Vázquez, M.** (2017) Developmental transcriptional regulation by SUMOylation, an evolving field. *Genesis*, **55**, e23009.
- Mozgová, I. and Hennig, L.** (2015) The Polycomb group protein regulatory network. *Annual Review of Plant Biology*, **66**, 269-296.
- Mozgová, I., Muñoz-Viana, R. and Hennig, L.** (2017) PRC2 represses hormone-induced somatic embryogenesis in vegetative tissue of *Arabidopsis thaliana*. *PLoS genetics*, **13**, e1006562.
- Muller-Xing, R., Clarenz, O., Pokorný, L., Goodrich, J. and Schubert, D.** (2014) Polycomb-group proteins and *FLOWERING LOCUS T* maintain commitment to flowering in *Arabidopsis thaliana*. *The Plant Cell*, **26**, 2457-2471.
- Müller, J. and Kassis, J.A.** (2006) Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Current Opinion in Genetics & Development*, **16**, 476-484.
- Mutasa-Göttgens, E. and Hedden, P.** (2009) Gibberellin as a factor in floral regulatory networks. *Journal of Experimental Botany*, **60**, 1979-1989.
- Napp-Zinn, K.** (1987) Vernalization-environmental and genetic regulation. In JG Atherton, ed, *Manipulation of Flowering*. Butterworths, London, pp123-132.
- Neigeborn, L. and Carlson, M.** (1984) Genes affecting the regulation of *SUC2* gene expression by glucose repression in *Saccharomyces Cerevisiae*. *Genetics*, **108**, 845-858.
- Ng, D.W., Wang, T., Chandrasekharan, M.B., Aramayo, R., Kertbundit, S. and Hall, T.C.** (2007) Plant SET domain-containing proteins: structure, function and regulation. *Biochimica et Biophysica Acta*, **1769**, 316-329.



- Nicol, J.W., Helt, G.A., Blanchard, J.S.G., Raja, A. and Loraine, A.E.** (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics*, **25**, 2730-2731.
- Ojolo, S.P., Cao, S., Priyadarshani, S.V.G.N., Li, W., Yan, M., Aslam, M., Zhao, H. and Qin, Y.** (2018) Regulation of plant growth and development: a review from a chromatin remodeling perspective. *Frontiers in Plant Science*, **9**, 1232.
- Pan, J., McKenzie, Z.M., D'Avino, A.R., Mashtalir, N., Lareau, C.A., St. Pierre, R., Wang, L., Shilatifard, A. and Kadoch, C.** (2019) The ATPase module of mammalian SWI/SNF family complexes mediates subcomplex identity and catalytic activity-independent genomic targeting. *Nature Genetics*, **51**, 618-626.
- Paz Sanchez, M., Aceves-García, P., Petrone, E., Steckenborn, S., Vega-León, R., Álvarez-Buylla, E.R., Garay-Arroyo, A. and García-Ponce, B.** (2015) The impact of Polycomb group (PcG) and Trithorax group (TrxG) epigenetic factors in plant plasticity. *New Phytologist*, **208**, 684-694.
- Peirats-Llobet, M., Han, S.K., Gonzalez-Guzman, M., Jeong, C.W., Rodriguez, L., Belda-Palazon, B., Wagner, D. and Rodriguez, P.L.** (2016) A direct link between abscisic acid sensing and the chromatin-remodeling ATPase BRAHMA *via* core ABA signaling pathway components. *Molecular Plant*, **9**, 136-147.
- Pereira, C.F., Piccolo, F.M., Tsubouchi, T., Sauer, S., Ryan, N.K., Bruno, L., Landeira, D., Santos, J., Banito, A., Gil, J., Koseki, H., Merckenschlager, M. and Fisher, A.G.** (2010) ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. *Cell Stem Cell*, **6**, 547-556.
- Peters, A.H.F.M., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A.H.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., Martens, J.H.A. and Jenuwein, T.** (2003) Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Molecular Cell*, **12**, 1577-1589.
- Peterson, C.L. and Herskowitz, I.** (1992) Characterization of the yeast *SWI1*, *SWI2*, and *SWI3* genes, which encode a global activator of transcription. *Cell*, **68**, 573-583.
- Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B. and Zhang, Y.** (2003) Role of histone H3 lysine 27 methylation in X inactivation. *Science*, **300**, 131-135.
- Poethig, R.S.** (2003) Phase change and the regulation of developmental timing in plants. *Science*, **301**, 334-336.
- Pu, L. and Sung, Z.R.** (2015) PcG and TrxG in plants-friends or foes. *Trends in Genetics*, **31**, 252-262.

- Qian, S., Lv, X., Scheid, R.N., Lu, L., Yang, Z., Chen, W., Liu, R., Boersma, M.D., Denu, J.M., Zhong, X. and Du, J.** (2018) Dual recognition of H3K4me3 and H3K27me3 by a plant histone reader SHL. *Nature Communications*, **9**, 2425.
- Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F. and Manke, T.** (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, **44**, W160-165.
- Reyes, J.C.** (2014) The many faces of plant SWI/SNF complex. *Molecular Plant*, **7**, 454-458.
- Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V. and Helin, K.** (2014) Gene silencing triggers Polycomb repressive complex 2 recruitment to CpG islands genome wide. *Molecular Cell*, **55**, 347-360.
- Rounsley, S.D., Ditta, G.S. and Yanofsky, M.F.** (1995) Diverse roles for MADS box genes in *Arabidopsis* development. *The Plant Cell*, **7**, 1259-1269.
- Rubio-Somoza, I. and Weigel, D.** (2011) MicroRNA networks and developmental plasticity in plants. *Trends in Plant Science*, **16**, 258-264.
- Saha, A., Wittmeyer, J. and Cairns, B.R.** (2006) Chromatin remodelling: the industrial revolution of DNA around histones. *Nature Reviews Molecular Cell Biology*, **7**, 437-447.
- Salmon-Divon, M., Dvinge, H., Tammoja, K. and Bertone, P.** (2010) PeakAnalyzer: Genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics*, **11**, 415.
- Sang, Y., Silva-Ortega, C.O., Wu, S., Yamaguchi, N., Wu, M.F., Pfluger, J., Gillmor, C.S., Gallagher, K.L. and Wagner, D.** (2012) Mutations in two non-canonical *Arabidopsis* SWI2/SNF2 chromatin remodeling ATPases cause embryogenesis and stem cell maintenance defects. *The Plant Journal*, **72**, 1000-1014.
- Schatlowski, N., Stahl, Y., Hohenstatt, M.L., Goodrich, J. and Schubert, D.** (2010) The CURLY LEAF interacting protein BLISTER controls expression of Polycomb-group target genes and cellular differentiation of *Arabidopsis thaliana*. *The Plant Cell*, **22**, 2291-2305.
- Schommer, C., Palatnik, J.F., Aggarwal, P., Chételat, A., Cubas, P., Farmer, E.E., Nath, U. and Weigel, D.** (2008) Control of jasmonate biosynthesis and senescence by miR319 targets. *PLOS Biology*, **6**, e230.
- Schönrock, N., Bouveret, R., Leroy, O., Borghi, L., Köhler, C., Gruissem, W. and Hennig, L.** (2006) Polycomb-group proteins repress the floral activator *AGL19* in the *FLC*-independent vernalization pathway. *Genes & Development*, **20**, 1667-1678.

- Schuettengruber, B., Bourbon, H.M., Di Croce, L. and Cavalli, G.** (2017) Genome regulation by Polycomb and Trithorax: 70 years and counting. *Cell*, **171**, 34-57.
- Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.Y., Bourgon, R., Biggin, M. and Pirrotta, V.** (2006) Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nature Genetics*, **38**, 700-705.
- Schwartz, Y.B. and Pirrotta, V.** (2007) Polycomb silencing mechanisms and the management of genomic programmes. *Nature Reviews Genetics*, **8**, 9-22.
- Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasino, R.A. and Coupland, G.** (2006) The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. *Genes & Development*, **20**, 898-912.
- Seidah, N.G. and Chrétien, M.** (1997) Eukaryotic protein processing: endoproteolysis of precursor proteins. *Current Opinion in Biotechnology*, **8**, 602-607.
- Seo, S., Herr, A., Lim, J.W., Richardson, G.A., Richardson, H. and Kroll, K.L.** (2005) Geminin regulates neuronal differentiation by antagonizing Brg1 activity. *Genes & Development*, **19**, 1723-1734.
- Sheldon, C.C., Rouse, D.T., Finnegan, E.J., Peacock, W.J. and Dennis, E.S.** (2000) The molecular basis of vernalization: the central role of FLOWERING LOCUS C (FLC). *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 3753-3758.
- Shim, J.S., Kubota, A. and Imaizumi, T.** (2017) Circadian clock and photoperiodic flowering in *Arabidopsis*: CONSTANS is a hub for signal integration. *Plant Physiology*, **173**, 5-15.
- Sif, S.** (2004) ATP-dependent nucleosome remodeling complexes: Enzymes tailored to deal with chromatin. *Journal of Cellular Biochemistry*, **91**, 1087-1098.
- Simon, J., Chiang, A., Bender, W., Shimell, M.J. and O'Connor, M.** (1993) Elements of the *Drosophila* bithorax complex that mediate repression by Polycomb group products. *Developmental Biology*, **158**, 131-144.
- Simpson, G.G.** (2004) The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Current Opinion in Plant Biology*, **7**, 570-574.
- Simpson, G.G. and Dean, C.** (2002) *Arabidopsis*, the rosetta stone of flowering time? *Science*, **296**, 285-289.
- Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H.D. and Cordes, S.P.** (2009) A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell*, **138**, 885-897.

- Smaczniak, C., Immink, R.G.H., Angenent, G.C. and Kaufmann, K.** (2012) Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development*, **139**, 3081-3098.
- Springer, N.M., Napoli, C.A., Selinger, D.A., Pandey, R., Cone, K.C., Chandler, V.L., Kaeppeler, H.F. and Kaeppeler, S.M.** (2003) Comparative analysis of SET domain proteins in maize and *Arabidopsis* reveals multiple duplications preceding the divergence of monocots and dicots. *Plant Physiology*, **132**, 907-925.
- Srikanth, A. and Schmid, M.** (2011) Regulation of flowering time: all roads lead to Rome. *Cellular and Molecular Life Sciences*, **68**, 2013-2037.
- Stillman, B.** (2018) Histone modifications: insights into their influence on gene expression. *Cell*, **175**, 6-9.
- Su, Y., Kwon, C.S., Bezhani, S., Huvermann, B., Chen, C., Peragine, A., Kennedy, J.F. and Wagner, D.** (2006) The N-terminal ATPase AT-hook-containing region of the *Arabidopsis* chromatin-remodeling protein SPLAYED is sufficient for biological activity. *The Plant Journal*, **46**, 685-699.
- Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F. and Coupland, G.** (2001) CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature*, **410**, 1116-1120.
- Sun, B., Looi, L.S., Guo, S., He, Z., Gan, E.S., Huang, J., Xu, Y., Wee, W.Y. and Ito, T.** (2014) Timing mechanism dependent on cell division is invoked by Polycomb eviction in plant stem cells. *Science*, **343**, 505-513.
- Takada, S. and Goto, K.** (2003) TERMINAL FLOWER 2, an *Arabidopsis* homolog of HETEROCHROMATIN PROTEIN 1, counteracts the activation of *FLOWERING LOCUS T* by CONSTANS in the vascular tissues of leaves to regulate flowering time. *The Plant Cell*, **15**, 2856-2865.
- Tang, X., Bian, S., Tang, M., Lu, Q., Li, S., Liu, X., Tian, G., Nguyen, V., Tsang, E.W., Wang, A., Rothstein, S.J., Chen, X. and Cui, Y.** (2012a) MicroRNA-mediated repression of the seed maturation program during vegetative development in *Arabidopsis*. *PLOS Genetics*, **8**, e1003091.
- Tang, X., Hou, A., Babu, M., Nguyen, V., Hurtado, L., Lu, Q., Reyes, J.C., Wang, A., Keller, W.A., Harada, J.J., Tsang, E.W. and Cui, Y.** (2008) The *Arabidopsis* BRAHMA chromatin-remodeling ATPase is involved in repression of seed maturation genes in leaves. *Plant Physiology*, **147**, 1143-1157.
- Tang, X., Lim, M.H., Pelletier, J., Tang, M., Nguyen, V., Keller, W.A., Tsang, E.W.T., Wang, A., Rothstein, S.J., Harada, J.J. and Cui, Y.** (2012b) Synergistic repression of the embryonic programme by SET DOMAIN GROUP 8 and EMBRYONIC FLOWER 2 in *Arabidopsis* seedlings. *Journal of Experimental Botany*, **63**, 1391-1404.

- To, A., Valon, C., Savino, G., Guillemainot, J., Devic, M., Giraudat, J. and Parcy, F.** (2006) A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. *The Plant Cell*, **18**, 1642-1651.
- Torti, S. and Fornara, F.** (2012) *AGL24* acts in concert with *SOC1* and *FUL* during *Arabidopsis* floral transition. *Plant Signaling & Behavior*, **7**, 1251-1254.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L.** (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, **7**, 562-578.
- Tsai, C.C. and Fondell, J.D.** (2004) Nuclear receptor recruitment of histone-modifying enzymes to target gene promoters. In *Vitamins & Hormones*: Academic Press, pp. 93-122.
- Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G. and Colot, V.** (2007) *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLOS Genetics*, **3**, e86.
- van Kruijsbergen, I., Hontelez, S. and Veenstra, G.J.** (2015) Recruiting Polycomb to chromatin. *The International Journal of Biochemistry & Cell Biology*, **67**, 177-187.
- Veluchamy, A., Jégu, T., Ariel, F., Latrasse, D., Mariappan, K.G., Kim, S.K., Crespi, M., Hirt, H., Bergounioux, C., Raynaud, C. and Benhamed, M.** (2016) LHP1 regulates H3K27me3 spreading and shapes the three-dimensional conformation of the *Arabidopsis* genome. *PLOS One*, **11**, e0158936.
- Venkatesh, S. and Workman, J.L.** (2015) Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews Molecular Cell Biology*, **16**, 178-189.
- Verbsky, M.L. and Richards, E.J.** (2001) Chromatin remodeling in plants. *Current Opinion in Plant Biology*, **4**, 494-500.
- Verdin, E. and Ott, M.** (2014) 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nature Reviews Molecular Cell Biology*, **16**, 258-264.
- Wagner, D. and Meyerowitz, E.M.** (2002) SPLAYED, a novel SWI/SNF ATPase homolog, controls reproductive development in *Arabidopsis*. *Current Biology*, **12**, 85-94.
- Walley, J.W., Rowe, H.C., Xiao, Y., Chehab, E.W., Kliebenstein, D.J., Wagner, D. and Dehesh, K.** (2008) The chromatin remodeler SPLAYED regulates specific stress signaling pathways. *PLOS Pathogens*, **4**, e1000237.

- Wang, D., Tyson, M.D., Jackson, S.S. and Yadegari, R.** (2006) Partially redundant functions of two SET-domain Polycomb group proteins in controlling initiation of seed development in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 13244-13249.
- Wang, G.G., Allis, C.D. and Chi, P.** (2007) Chromatin remodeling and cancer, part II: ATP-dependent chromatin remodeling. *Trends in Molecular Medicine*, **13**, 373-380.
- Wang, H., Liu, C., Cheng, J., Liu, J., Zhang, L., He, C., Shen, W.H., Jin, H., Xu, L. and Zhang, Y.** (2016) *Arabidopsis* flower and embryo developmental genes are repressed in seedlings by different combinations of Polycomb group proteins in association with distinct sets of cis-regulatory elements. *PLOS Genetics*, **12**, e1005771.
- Wang, J.W., Czech, B. and Weigel, D.** (2009) miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell*, **138**, 738-749.
- Wang, Q. and Shen, W.H.** (2018) Chromatin modulation and gene regulation in plants: insight about PRC1 function. *Biochemical Society Transactions*, **46**, 957-966.
- Whitcomb, S.J., Basu, A., Allis, C.D. and Bernstein, E.** (2007) Polycomb group proteins: an evolutionary perspective. *Trends in Genetics*, **23**, 494-502.
- Whittaker, C. and Dean, C.** (2017) The *FLC* locus: a platform for discoveries in epigenetics and adaptation. *Annual Review of Cell and Developmental Biology*, **33**, 555-575.
- Wiles, E.T. and Selker, E.U.** (2017) H3K27 methylation: a promiscuous repressive chromatin mark. *Current Opinion in Genetics & Development*, **43**, 31-37.
- Woo, C.J., Kharchenko, P.V., Daheron, L., Park, P.J. and Kingston, R.E.** (2010) A region of the human HOXD cluster that confers Polycomb-group responsiveness. *Cell*, **140**, 99-110.
- Woodward, A.W. and Bartel, B.** (2005) Auxin: regulation, action, and interaction. *Annals of Botany*, **95**, 707-735.
- Wu, G., Park, M.Y., Conway, S.R., Wang, J.W., Weigel, D. and Poethig, R.S.** (2009) The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell*, **138**, 750-759.
- Wu, M.F., Yamaguchi, N., Xiao, J., Bargmann, B., Estelle, M., Sang, Y. and Wagner, D.** (2015) Auxin-regulated chromatin switch directs acquisition of flower primordium founder fate. *eLife*, **4**, e09269.

- Wu, M.F., Sang, Y., Bezhani, S., Yamaguchi, N., Han, S.K., Li, Z., Su, Y., Slewinski, T.L. and Wagner, D. (2012) SWI2/SNF2 chromatin remodeling ATPases overcome Polycomb repression and control floral organ identity with the LEAFY and SEPALLATA 3 transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 3576-3581.
- Xiao, J., Jin, R., Yu, X., Shen, M., Wagner, J.D., Pai, A., Song, C., Zhuang, M., Klasfeld, S., He, C., Santos, A.M., Helliwell, C., Pruneda-Paz, J.L., Kay, S.A., Lin, X., Cui, S., Garcia, M.F., Clarenz, O., Goodrich, J., Zhang, X., Austin, R.S., Bonasio, R. and Wagner, D. (2017) *Cis* and *trans* determinants of epigenetic silencing by Polycomb repressive complex 2 in *Arabidopsis*. *Nature Genetics*, **49**, 1546-1552.
- Xiao, J. and Wagner, D. (2015) Polycomb repression in the regulation of growth and development in *Arabidopsis*. *Current Opinion in Plant Biology*, **23**, 15-24.
- Xu, Q., Xiang, Y., Wang, Q., Wang, L., Brind'Amour, J., Bogutz, A.B., Zhang, Y., Zhang, B., Yu, G., Xia, W., Du, Z., Huang, C., Ma, J., Zheng, H., Li, Y., Liu, C., Walker, C.L., Jonasch, E., Lefebvre, L., Wu, M., Lorincz, M.C., Li, W., Li, L. and Xie, W. (2019) SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nature Genetics*, **51**, 844-856.
- Xu, Y., Guo, C., Zhou, B., Li, C., Wang, H., Zheng, B., Ding, H., Zhu, Z., Peragine, A., Cui, Y., Poethig, S. and Wu, G. (2016) Regulation of vegetative phase change by SWI2/SNF2 chromatin remodeling ATPase BRAHMA. *Plant Physiology*, **172**, 2416-2428.
- Yang, H., Berry, S., Olsson, T.S.G., Hartley, M., Howard, M. and Dean, C. (2017) Distinct phases of Polycomb silencing to hold epigenetic memory of cold in *Arabidopsis*. *Science*, **357**, 1142-1145.
- Yang, M., Wang, X., Huang, H., Ren, D., Su, Y., Zhu, P., Zhu, D., Fan, L., Chen, L., He, G. and Deng, X.W. (2016) Natural variation of H3K27me3 modification in two *Arabidopsis* accessions and their hybrid. *Journal of Integrative Plant Biology*, **58**, 466-474.
- Yang, S., Li, C., Zhao, L., Gao, S., Lu, J., Zhao, M., Chen, C.Y., Liu, X., Luo, M., Cui, Y., Yang, C. and Wu, K. (2015) The *Arabidopsis* SWI2/SNF2 chromatin remodeling ATPase BRAHMA targets directly to *PIN*s and is required for root stem cell niche maintenance. *The Plant Cell*, **27**, 1670-1680.
- Yang, Z., Qian, S., Scheid, R.N., Lu, L., Chen, X., Liu, R., Du, X., Lv, X., Boersma, M.D., Scalf, M., Smith, L.M., Denu, J.M., Du, J. and Zhong, X. (2018) EBS is a bivalent histone reader that regulates floral phase transition in *Arabidopsis*. *Nature Genetics*, **50**, 1247-1253.

- Yu, J., Li, Y., Ishizuka, T., Guenther, M.G. and Lazar, M.A.** (2003) A SANT motif in the SMRT corepressor interprets the histone code and promotes histone deacetylation. *The EMBO Journal*, **22**, 3403-3410.
- Yuan, W., Luo, X., Li, Z., Yang, W., Wang, Y., Liu, R., Du, J. and He, Y.** (2016) A *cis* cold memory element and a *trans* epigenome reader mediate Polycomb silencing of *FLC* by vernalization in *Arabidopsis*. *Nature Genetics*, **48**, 1527-1534.
- Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K. and Peng, W.** (2009) A clustering approach for identification of enriched domains from histone modification ChIP-seq data. *Bioinformatics*, **25**, 1952-1958.
- Zee, B.M., Levin, R.S., Xu, B., LeRoy, G., Wingreen, N.S. and Garcia, B.A.** (2010) *In vivo* residue-specific histone methylation dynamics. *The Journal of Biological Chemistry*, **285**, 3341-3350.
- Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J. and Jacobsen, S.E.** (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLOS Biology*, **5**, e129.
- Zhang, X., Henriques, R., Lin, S.S., Niu, Q.W. and Chua, N.H.** (2006) Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nature Protocols*, **1**, 641-646.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. and Liu, X.S.** (2008) Model-based analysis of ChIP-seq (MACS). *Genome Biology*, **9**, R137.
- Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.K., Shahab, A., Kuznetsov, V.A., Bourque, G., Oh, S., Ruan, Y., Ng, H.H. and Wei, C.L.** (2007) Whole-genome mapping of histone H3 lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell*, **1**, 286-298.
- Zhou, Y., Hartwig, B., James, G.V., Schneeberger, K. and Turck, F.** (2016) Complementary activities of TELOMERE REPEAT BINDING proteins and Polycomb group complexes in transcriptional regulation of target genes. *The Plant Cell*, **28**, 87-101.
- Zhou, Y., Romero-Campero, F.J., Gómez-Zambrano, Á., Turck, F. and Calonje, M.** (2017) H2A monoubiquitination in *Arabidopsis thaliana* is generally independent of LHP1 and PRC2 activity. *Genome Biology*, **18**, 69.
- Zhou, Y., Wang, Y., Krause, K., Yang, T., Dongus, J.A., Zhang, Y. and Turck, F.** (2018) Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in *Arabidopsis*. *Nature Genetics*, **50**, 638-644.



## Appendices

### Appendix A: Primers used in this thesis

Primer	Sequence (5' to 3')	Application
SWN-Pme1-F	AGCTTTGTTTAAACTCACTTATCTACTCGTGTTATGGGCAC	Cloning
SWN-Asc1-R	TTGGCGCGCCAATGAGATTGGTGCTTTCTGGCTCTACG	Cloning
CLF-Pme1-F	AGCTTTGTTTAAACTTGTCGATAGTCAAGGAATTGTACCTC	Cloning
CLF-Asc1-R	TTGGCGCGCCAAGCAAGCTTCTTGGGTCTACCAACAG	Cloning
SYD-Pme1-F	AGCTTTGTTTAAACATGATGAATCTTGGGAGGACC	Cloning
SYD-Asc1-R	TTGGCGCGCCACCAACAAAGTGGAGCTAGTGG	Cloning
clf-29-RP	GAGGCATTGACTTTGATTTGC	Genotyping
clf-29-LP	AAGAACTTGCTAGTTCCGCC	Genotyping
swn-4-RP	CGAGGAATTTTCTAATTCCGG	Genotyping
swn-4-LP	TGATTATTGCTCCGTTTCCAC	Genotyping
syd-5-RP	CTTCTCACGGTGAAGTCGTTT	Genotyping
syd-5-LP	CTGCGAAGTTAGCTGTTTTGG	Genotyping
agl17-3-RP	AAATGCAAACACAAACCGAAG	Genotyping
agl17-3-LP	AGGCAAGAATTGCACTCATTG	Genotyping
LBb1	GCGTGGACCGCTTGCTGCAACT	Genotyping
CLF-1-F	AAGCATTTACCTGGGGGTCG	Sequencing
CLF-1-R	GAGCCCGGGTATGATTCTGG	Sequencing
CLF-2-F	CCAGCAACGTTTCGTGTTTGT	Sequencing

---

CLF-2-R	ACCATTCCTGAACTGGCAGG	Sequencing
CLF-3-F	GAGGCCATTGACGACGAAGA	Sequencing
CLF-3-R	TGGAGAACTTCGTGGGTGTG	Sequencing
CLF-4-F	GCTGAGCGAGTTCCTCGTAA	Sequencing
CLF-4-R	TGCAGCTCTTTGGGCAACTA	Sequencing
CLF-5-F	AGCCCTGCCGTCAGTTTAAT	Sequencing
CLF-5-R	CACCCTGTGATCTCCAGCAA	Sequencing
CLF-R-4693	ACTTTCATCACTACCCGGCG	Sequencing
CLF-F-2675	GAGGTCTTCTATGGAACGAGGAGG	Sequencing
CLF-F-6144	GGAAGATATACGATCGCGAG	Sequencing
SWN-1-F	TGCAATGACGGGACACCTAC	Sequencing
SWN-1-R	TGAGCACGCCCTGAACTAAG	Sequencing
SWN-2-F	CTTAGTTCAGGGCGTGCTCA	Sequencing
SWN-2-R	GTTGCTATCGTCCGTCACCA	Sequencing
SWN-3-F	CTCCAGAGCCAGGCATGTCG	Sequencing
SWN-3-R	TGTCTTCGACCAACCACAGAC	Sequencing
SWN-4-F	TGCTGAGAGAATACCGCCATAC	Sequencing
SWN-4-R	TGTTTGCTGCAGGGTTTCCT	Sequencing
SWN-5-F	GGCGGTCAGAGAAGTACCAG	Sequencing
SWN-5-R	CAACAATGCCCAAAGTCGCA	Sequencing
SWN-6-F	TGGTGGCAAAAACCAGTCCT	Sequencing
SWN-6-R	CCTTTGCGTTGAGCATCGAG	Sequencing

---

---

SWN-F-985	GTACAGGTGATTAGTTTGTACTG	Sequencing
SWN-F-3973	ACTGTCGTCTGCATGGATGTT	Sequencing
SWN-F-5199	GATCTAGTACCCCAAATTTG	Sequencing
SWN-F-6790	GAAGGAGGACTAGGATTATAAC	Sequencing
SWN-R-7050	TATGCGGGAAGCGATGAGAC	Sequencing
F-SYD-1	GTGGCTTGTTGTAATGTTATG	Sequencing
F-SYD-2	CGAGCGGGAAGGTGACATTTG	Sequencing
F-SYD-3	GAATTCGGTCCCTACTGGTCTCC	Sequencing
F-SYD-4	CCCAAAATCTCTCTCTCTTTCCG	Sequencing
F-SYD-5	CATGAATTGTGGACCTCTGAG	Sequencing
F-SYD-6	ACACAATGCCCTATCAAGTC	Sequencing
F-SYD-7	GCGGTTGGAGTTAGCAATGAG	Sequencing
F-SYD-8	GTTGGGTTGTCCTCAGATGC	Sequencing
F-SYD-9	TACCTCGTTGGAATCATCCC	Sequencing
F-SYD-10	TGATCTGGGTGGCATTCTG	Sequencing
F-SYD-11	ATGTCACAGAAATGGATTATGGGTCG	Sequencing
F-SYD-12	CCATTAAGGAACACTTATTTGCTCAG	Sequencing
F-SYD-13	GCTTAAGAAATATTAGGAAGGAGATGTTTCGCC	Sequencing
F-SYD-14	CTGGCTTTAGGGAAAACTAGAGG	Sequencing
F-SYD-15	TGGGTAAGCTTTCTTCCATCTG	Sequencing
F-SYD-16	CACTGTACAACAATCACTTAAATGGC	Sequencing
F-SYD-17	GGCCTAAATTAAGCAAGATTCACTGGC	Sequencing

---

---

F-SYD-18	GAGGAGAATCTACTGATCATCAATCGAC	Sequencing
F-SYD-19	TGTTATTGTAAATGCAATGAAGCACTGGTTAGCGC	Sequencing
R-SYD-1	AGCATCAATCAGAATCGGTGAA	Sequencing
R-107-GFP	CCGTAGGTCAGGGTGGTGA	Sequencing
Ta3-F	GATTCTTACTGTAAAGAACATGGCATTGAGAGA	qChIP-PCR
Ta3-R	TCCAAATTCCTGAGGTGCTTGTAACC	qChIP-PCR
WOX1-F	CATCACCATTTCGATGACAAG	qChIP-PCR
WOX1-R	CGTAACCTCATTCTCAATTCACC	qChIP-PCR
EMF1-F	TTCTCTCTCCCTCCCCTTTC	qChIP-PCR
EMF1-R	AGCGGCAAGGGATAAAAACT	qChIP-PCR
MIR156C-F	TGTTTGGACCAGCCTAAGAAA	qChIP-PCR
MIR156C-R	GATTTGGTTCCCAATTGCAT	qChIP-PCR
LFY-F	ATTGGTTCAAGCACCACCTC	qChIP-PCR
LFY-R	TCAAGCTCCTCGTCCTTCAT	qChIP-PCR
SEP1-F	CTTTCCACGGTTGGAGAAGA	qChIP-PCR
SEP1-R	GAGGAAGAGTAGAGCTGAAGAGGA	qChIP-PCR
TCP13-F	CGCTACGACAAGACGTGAAA	qChIP-PCR
TCP13-R	CGCGTGAAACCCTAACAATC	qChIP-PCR
MYB5-F	GGGGATGAAGAGAGGACCAT	qChIP-PCR
MYB5-R	CGACGTAATTCCTCCACGTT	qChIP-PCR
WRKY23-F	ACTACCCGTCGTCACAAAGC	qChIP-PCR
WRKY23-R	CAAAGTCTTGATGCTGCTGAG	qChIP-PCR

---

---

FLC-F	CCTAATTTGATCCTCAGGTTTGGG	qChIP-PCR
FLC-R	CCGACGAAGAAAAAGTAGATAGGCA	qChIP-PCR
CLE9-F	ATCCTCCACCGCTTCTTCAA	qChIP-PCR
CLE9-R	TGAGGAGTCACGTGGTAAGG	qChIP-PCR
NSP2-F	GCTTCTCCCACTTCAACGTC	qChIP-PCR
NSP2-R	GACCCAAAAGCTCACGAAGG	qChIP-PCR
PDLP6-F	CACCACCGAGTCTCTACGAG	qChIP-PCR
PDLP6-R	TGAGGATAAGACGGTGGTGG	qChIP-PCR
SUC2-F	TCCGGCTGATCTAACTCACC	qChIP-PCR
SUC2-R	CACTACAACCAACGCAACAA	qChIP-PCR
GER3-F	GGCCTTTCCCAGAGTTTAGC	qChIP-PCR
GER3-R	TTTGCGCCGGATTTATCTCC	qChIP-PCR
OLE1-F	TCATCGGCAGAGACCAGTAC	qChIP-PCR
OLE1-R	GTAGTCAGATCCTCGTCCGG	qChIP-PCR
1G72100-F	CTTTACTAGCGGCGGAGGTA	qChIP-PCR
1G72100-R	ATGACTTCTCCCTCTGACGC	qChIP-PCR
2G15020-F	AGTCCGTTTCGGCTCTAAGTT	qChIP-PCR
2G15020-R	CGAGTCTGATTCACACGCAG	qChIP-PCR
4G08740-F	GCTCATCAACTCGTCATGCA	qChIP-PCR
4G08740-R	TGAGCTTCGTGTCCAAGGAT	qChIP-PCR
4G14315-F	TCGGTTTTGGCGATTGTTGT	qChIP-PCR
4G14315-R	TGGAAGAGATGGTGCAGTCA	qChIP-PCR

---

---

5G24130-F	AGCAGAGCCGAGTACATCAA	qChIP-PCR
5G24130-R	TGAGAATCCTGTCGCTGTGA	qChIP-PCR
GIK-1-F	GTATTTGTTTAAGTTCTAACTAAG	qChIP-PCR
GIK-1-R	GGGAAGAAAACGCGGAGATC	qChIP-PCR
GIK-2-F	GACTTAACCATTGGAGTAC	qChIP-PCR
GIK-2-R	GAAAGTATTAAGTAGATCTG	qChIP-PCR
GIK-3-F	ACTCTTAGGGCTCACATTC	qChIP-PCR
GIK-3-R	CTGACGGATGCTGACGTTAG	qChIP-PCR
GIK-4-F	CCTCTAGAAAAAAGAGCTTG	qChIP-PCR
GIK-4-R	CCACAATAGGCAAGTTCATTC	qChIP-PCR
2G18150-1-F	GCAAGTTAGCAGCGTAGGAT	qChIP-PCR
2G18150-1-R	CAGAGACTGTACAACCAATCCG	qChIP-PCR
2G18150-2-F	GGTGAGATCGAGACCTTGGT	qChIP-PCR
2G18150-2-R	CATTCCTGCACCCAACAACA	qChIP-PCR
2G18150-3-F	AACAACCTTCAAACCCGCGAG	qChIP-PCR
2G18150-3-R	GGTTGTGATGGATCGTTGCT	qChIP-PCR
2G18140-4-F	GGCTACAACCTGACCTCACGA	qChIP-PCR
2G18150-4-R	GAGAGTAACTATGGCGGCGA	qChIP-PCR
2G18140-1-F	TCTCACCCTCGAACCTGTC	qChIP-PCR
2G18140-1-R	GGGACTGTTGAATTCCGACC	qChIP-PCR
2G18140-2-F	GTCGAACAAGTTGTCGGGTT	qChIP-PCR
2G18140-2-R	GGAAGAAGAGATTCGGCAACT	qChIP-PCR

---

---

2G18140-3-F	ACAAGAGACAGTGTAGGGCA	qChIP-PCR
2G18140-3-R	CTCGCGGGTTTGAAGTTGTT	qChIP-PCR
2G18140-4-F	CACGATCTCCTCCGCTCTAG	qChIP-PCR
2G18140-4-R	GAGTAACTTTGGCGGCAACA	qChIP-PCR
AG-F	TCGGAGCTAGGAGGAGATTC	qChIP-PCR
AG-R	CGATTCGTTGTGTTCTCGAT	qChIP-PCR
FT-F	CCAGATGTTCCAAGTCCTAGCAACC	qChIP-PCR
FT-R	GGTGTGGGCTTTTTTGGGAGAC	qChIP-PCR
ACT2/7-F	CGTTTCGCTTTCCTTAGTGTTAGCT	qChIP-PCR
ACT2/7-R	AGCGAACGGATCTAGAGACTCACCTTG	qChIP-PCR
AGL17-P1-F	TTCGGTTTCAGTTTTGATTCAG	qChIP-PCR
AGL17-P1-R	AACACATGAGACAGAAGGAACAG	qChIP-PCR
AGL17-P2-F	ACGCCGAGGTCTGTCTCAT	qChIP-PCR
AGL17-P2-R	AGAAACCTGGAGCTGGCAAA	qChIP-PCR
AGL17-P3-F	CGACCATGATTGTTCTGTCTG	qChIP-PCR
AGL17-P3-R	GGACATGGCCAACTACAACC	qChIP-PCR
SMZ-RP	GAGATGCTTTCATCGCTTTTG	qChIP-PCR
SMZ-LP	TGCCTATAATCATCCACGACG	qChIP-PCR
LGO-RP	TTTATTTTCGTTTTCATGCC	qChIP-PCR
LGO-LP	AACATGTACCACCCATTGCTC	qChIP-PCR
scpl35-F	GCACTGAAAGTTATGCCGAGA	qChIP-PCR
scpl35-R	GGATAAGGAAGGTTGGTGACG	qChIP-PCR

---

---

AGP11-F	TGCTTTAGCTTTGGACACCTC	qChIP-PCR
AGP11-R	CAAAAGGGCAACTACGACAAA	qChIP-PCR
NACO49-F	GGGGGAGAAAGGAATGAAAT	qChIP-PCR
NACO49-R	CTTCTTCCTCGTTATCCACCA	qChIP-PCR
AGL15-F	CGAGAATGCGAATAGCAGAC	qChIP-PCR
AGL15-R	TACTCGAAGAGCTTGCCAGA	qChIP-PCR
4G18282-F	GCAACTGCAACTTCACCATC	qChIP-PCR
4G18282-R	GGAGAGAGAGCAAATGTTCCA	qChIP-PCR
SAG201-F	CCAACGTCTCAAGCAAAGTT	qChIP-PCR
SAG201-R	CGGGTCGGCTTTCTTATG	qChIP-PCR
RTFL16-F	ATCTTGCCAACATAGGAGCA	qChIP-PCR
RTFL16-R	AGAGTATCGAGCAGCAGAGGA	qChIP-PCR
RT-GAPDH-F	CTTGCACCAAGCAGCATGAA	qRT-PCR
RT-GAPDH-R	CCGATCCAGACACTGTACTTCCTT	qRT-PCR
RT-agl17-3-F	GCTGAAACTCTAAGGCAAGAATT	qRT-PCR
RT-agl17-3-R	ACTCCTTAACGCTCAAACCA	qRT-PCR

---



**Appendix B: Total read numbers of all ChIP-seq samples**

Samples	Total reads	Uniquely mapped
CLF rep1	13,000,000	6,700,000
CLF rep2	12,000,000	6,400,000
SWN rep1	42,000,000	9,500,000
SWN rep2	41,000,000	13,000,000
WT_K27 rep1	14,000,000	11,000,000
WT_K27 rep2	14,000,000	11,000,000
<i>clf-29_K27</i> rep1	15,000,000	11,000,000
<i>clf-29_K27</i> rep2	14,000,000	11,000,000
<i>swn-4_K27</i> rep1	16,000,000	11,000,000
<i>swn-4_K27</i> rep2	12,000,000	10,000,000
<i>clf-29 swn-4_K27</i> rep1	11,000,000	6,200,000
<i>clf-29 swn-4_K27</i> rep2	14,000,000	8,500,000
SYD seedling rep1	12,770,317	10,350,364
SYD seedling rep2	7,863,384	6,741,462
SYD seedling input	14,116,576	6,240,557
SYD inflorescence rep1	24,492,659	15,221,664
SYD inflorescence rep2	22,775,808	13,457,949
SYD inflorescence input	22,281,346	15,444,130

---

WT_H3K27me3 rep1	21,663,120	16,860,953
WT_H3K27me3 rep2	20,545,998	15,949,377
WT input	21,663,120	16,099,250
<i>syd-5</i> _H3K27me3 rep1	22,594,576	17,082,777
<i>syd-5</i> _H3K27me3 rep2	20,080,210	15,505,851
<i>syd-5</i> input	18,804,469	13,628,902

---

**Appendix C: Total read numbers of all RNA-seq samples**

Samples		Total	Mapped	Aligned pairs
WT rep1	Forward	24,247,121	23,569,958	21,545,227
	Reverse	24,247,121	21,898,473	
WT rep2	Forward	23,453,341	22,619,791	19,878,956
	Reverse	23,453,341	20,295,303	
WT rep3	Forward	17,662,886	17,062,223	15,649,123
	Reverse	17,662,886	15,937,038	
<i>clf-29</i> rep1	Forward	20,601,270	18,797,720	18,143,120
	Reverse	20,601,270	18,698,512	
<i>clf-29</i> rep2	Forward	18,695,240	18,090,129	16,798,857
	Reverse	18,695,240	17,106,009	
<i>clf-29</i> rep3	Forward	22,425,331	21,786,431	20,009,345
	Reverse	22,425,331	20,353,879	
<i>swn-4</i> rep1	Forward	18,630,733	18,008,285	16,637,370
	Reverse	18,630,733	16,952,099	
<i>swn-4</i> rep2	Forward	24,677,440	23,894,738	22,779,345
	Reverse	24,677,440	23,208,147	
<i>swn-4</i> rep3	Forward	24,154,138	23,493,065	21,191,193
	Reverse	24,154,138	21,545,710	

---

<i>clf-29 swn-4</i> rep1	Forward	23,896,240	22,776,814	21,859,458
	Reverse	23,896,240	22,445,589	
<i>clf-29 swn-4</i> rep2	Forward	23,065,073	22,008,176	21,205,347
	Reverse	23,065,073	21,817,292	
<i>clf-29 swn-4</i> rep3	Forward	25,944,577	24,735,884	23,818,219
	Reverse	25,944,577	24,492,594	
WT rep1 (for SYD study)	Forward	25,026,085	23,250,193	21,634,156
	Reverse	25,026,085	22,604,268	
WT rep2 (for SYD study)	Forward	21,994,219	20,189,031	18,532,898
	Reverse	21,994,219	19,475,792	
WT rep3 (for SYD study)	Forward	19,722,307	18,212,589	16,963,252
	Reverse	19,722,307	17,794,592	
<i>syd-5</i> rep1	Forward	21,323,898	19,787,956	18,247,069
	Reverse	21,323,898	19,066,122	
<i>syd-5</i> rep2	Forward	21,176,779	19,725,182	18,356,673
	Reverse	21,176,779	19,109,919	
<i>syd-5</i> rep3	Forward	20,847,567	19,288,466	17,962,490
	Reverse	20,847,567	18,826,812	

---

**Appendix D: List of genes regulated by BRM and PcG**

Genes occupied by CLF/SWN without H3K27me3

---

AT1G03780	AT1G72300	AT2G45790	AT3G61710	AT5G23230
AT1G07090	AT1G75490	AT3G02000	AT3G62450	AT5G36220
AT1G08300	AT1G75510	AT3G06590	AT4G01985	AT5G37540
AT1G09470	AT1G77630	AT3G07350	AT4G03295	AT5G38710
AT1G09490	AT1G78865	AT3G10282	AT4G12200	AT5G40030
AT1G10230	AT1G80450	AT3G12970	AT4G14170	AT5G41410
AT1G10657	AT1G80840	AT3G13510	AT4G14570	AT5G42630
AT1G11220	AT2G07042	AT3G13530	AT4G15160	AT5G42635
AT1G12970	AT2G16700	AT3G13810	AT4G17460	AT5G43020
AT1G13260	AT2G18340	AT3G13882	AT4G17820	AT5G43630
AT1G13380	AT2G18350	AT3G15300	AT4G18390	AT5G43760
AT1G14600	AT2G22170	AT3G18620	AT4G24990	AT5G45307
AT1G15010	AT2G22540	AT3G20720	AT4G25770	AT5G45310
AT1G18191	AT2G29660	AT3G22700	AT4G28190	AT5G51880
AT1G18400	AT2G31230	AT3G22886	AT4G28330	AT5G52630
AT1G23780	AT2G33509	AT3G25740	AT4G33040	AT5G52830
AT1G24140	AT2G35270	AT3G27700	AT4G34210	AT5G53110
AT1G25400	AT2G35700	AT3G28840	AT4G34215	AT5G55740

---

---

AT1G26920	AT2G36080	AT3G46060	AT4G34470	AT5G60230
AT1G30330	AT2G38304	AT3G47620	AT4G34760	AT5G60900
AT1G35140	AT2G39010	AT3G50740	AT4G36630	AT5G62020
AT1G44740	AT2G40330	AT3G54350	AT4G36870	AT5G63370
AT1G63030	AT2G41260	AT3G56400	AT4G39600	AT5G67060
AT1G65346	AT2G42040	AT3G56710	AT4G39850	AT5G67411
AT1G65347	AT2G42670	AT3G60390	AT5G02470	
AT1G66400	AT2G44910	AT3G60510	AT5G06580	
AT1G66840	AT2G45140	AT3G60640	AT5G13790	
AT1G72280	AT2G45780	AT3G61140	AT5G20740	

---

Genes occupied by CLF/SWN/BRM without H3K27me3

---

AT1G07090	AT1G78865	AT2G42040	AT3G50740	AT5G41410
AT1G10657	AT1G80450	AT2G45140	AT3G56400	AT5G43020
AT1G13260	AT1G80840	AT2G45780	AT3G56710	AT5G43630
AT1G14600	AT2G16700	AT3G06590	AT3G61140	AT5G43760
AT1G15010	AT2G18340	AT3G07350	AT4G17460	AT5G45307
AT1G18400	AT2G18350	AT3G10282	AT4G18390	AT5G45310
AT1G24140	AT2G22170	AT3G12970	AT4G28330	AT5G52830
AT1G25400	AT2G22540	AT3G13510	AT4G33040	AT5G53110

---

---

AT1G26920	AT2G29660	AT3G13530	AT4G34760	AT5G62020
-----------	-----------	-----------	-----------	-----------

AT1G30330	AT2G31230	AT3G13810	AT4G36870	AT5G67060
-----------	-----------	-----------	-----------	-----------

AT1G35140	AT2G36080	AT3G22886	AT5G06580	AT5G67411
-----------	-----------	-----------	-----------	-----------

AT1G66400	AT2G39010	AT3G25740	AT5G37540	
-----------	-----------	-----------	-----------	--

AT1G75490	AT2G40330	AT3G47620	AT5G38710	
-----------	-----------	-----------	-----------	--

---

## Curriculum Vitae

<b>Name:</b>	Jie Shu
<b>Post-secondary Education and Degrees:</b>	<p>Northwest Agriculture and Forestry University Yangling, Shaanxi, China 2010-2014 B.Eng.</p> <p>Western University London, Ontario, Canada 2014-present, Ph.D. Candidate</p>
<b>Honors and Awards:</b>	<p>Western Graduate Research Scholarship (WGRS) 2014-2019</p> <p>Chinese Government Scholarship for Abroad Doctoral Study 2014-2018</p> <p>Mitacs Globalink Graduate Fellowship 2014-2018</p> <p>Graduate Student Travel Award January 2016; April 2019 Department of Biology, Western University</p> <p>Mitacs Globalink &amp; Chinses Government Joint Scholarship 2013</p>
<b>Related Work Experience:</b>	<p><b>Teaching Assistant</b> Western University 2014-2019</p> <p><b>Research Assistant</b> York University, June 2013-September 2013</p>

### Selected Conference Presentations:

**Shu, J.**, Chen, C., Thapa, R.K., Bian, S., Nguyen, V., Yu, K., Yuan, Z.C., Liu, J., Kohalmi, S.E., Li, C., Cui, Y. Genome-wide occupancy of H3K27 methyltransferases CURLY LEAF and SWINGER in *Arabidopsis*. 30<sup>th</sup> International Conference on *Arabidopsis* Research. Wuhan, China, June 2019 [Poster]

Chen C., **Shu J.**, Thapa R.K., Kohalmi, S.E., Cui, Y. Plant ChIP-seq, assessment of published data and a visualizing platform for evaluation, comparison and sharing. Plant Biology. Montreal, Canada, July 2018 [Poster]



### **Publications:**

**Shu, J.**, Chen, C., Thapa, R.K., Bian, S., Nguyen, V., Yu, K., Yuan, Z.C., Liu, J., Kohalmi, S.E., Li, C. and Cui, Y. (2019) Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in *Arabidopsis* seedlings. *Plant Direct*, **3**, e00100. (**Ph.D. Thesis**)

Chen, C., **Shu, J.**, Li, C., Thapa, R.K., Nguyen, V., Yu, K., Yuan, Z.C., Kohalmi, S.E., Liu, J., Marsolais, F., Huang, S. and Cui, Y. (2019) RNA polymerase II independent recruitment of SPT6L at transcription start sites in *Arabidopsis*. *Nucleic Acids Research*, **47**, 6714-6725.