Cell-type specific epi-genomic studies in *Arabidopsis thaliana*

Inaugural thesis presented to the Faculty of Mathematics and Natural Sciences of Heinrich Heine University Düsseldorf for the degree of

Doctor of Natural Sciences / Doctor of Philosophy in Natural Sciences

by

Yu Fu

Anyang, China

Düsseldorf Dec. 2019

From the Department/Institute of Biology
of Heinrich Heine University Düsseldorf

Printed by permission of the Faculty of Mathematics and Natural Sciences of Heinrich Heine University Düsseldorf
Examiners:
1. Prof. Dr Daniel Schubert
2. Prof. Dr Rüdiger Simon
Date of the oral defence:

Table of Contents

1. Introduction

1.1 Epigenetics

Since the term “epigenetics” was firstly introduced by Conrad H. Waddington in the 1940s to describe events during embryogenesis that controlling cell fate (Waddington, 1953), the definition of epigenetics has been evolved and clarified along with expanding understanding of the molecular regulatory network in development. The current widely accepted concept of epigenetics is “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” (Riggs et al., 1996; Riggs and Martienssen, 1996). In all eukaryotic organisms, besides DNA provide essential genetic information, epigenetic mechanisms act as an extra layer of guidance to regulate gene expression. It can explain how different cells have divergent identity and function while sharing the same DNA sequence, such as imprinting, paramutation, X chromosome inactivation, regulation of transposon activity, and so on.

In plants, epigenetic regulation plays a broad and crucial role in development. From the first discovery of non-Mendelian inheritance of *r1(Red1)* (Brink, 1956) and *b1(Boosters)* (Coe, 1959) expression in maize, to recent findings of epigenetic regulation on *FLOWERING LOCUS C* (*FLC*) during vernalization in *Arabidopsis thaliana* (Gendall et al., 2001; Lister et al., 2004) as well as imprinting mechanisms in seed development, studies have revealed various players encoding epigenetic regulations. There are DNA methylation, histone modifications, chromatin-remodeling, non-coding RNA, as well as prions that pass inheritable information through proteins (Halfmann and Lindquist, 2010).

1.1.1 Brief overview of chromatin structure

In eukaryotes, DNA is wrapped around a histone octamer to form the nucleosome which is the basic unit of chromatin. The nucleosomes are connected by linker histone H1 and packed into chromatin fiber which further compacted to form the higher level of structure, known as chromosomes.

From linear DNA to 3D structure, nucleosomes are not evenly distributed. The heterochromatin is organized in a compact form and the euchromatin is loosely packed. In turn, those divergent chromatin structures enable different accessibility to load transcription machinery along DNA. In fact, nucleosomes are not arranged in a static status but instead are dynamically regulated. Studies have shown that the euchromatin-located active promoters and enhancers usually have higher nucleosome turnover rate than inactive genomic features in the condensed heterochromatin (Deal et al., 2010; Hendzel et al., 2000; Schones et al., 2008). Although the detailed mechanisms underlying chromatin remodeling are still not well understood, it is widely accepted that histones, transcription factors (TFs), and chromatin remodelers act together to regulate the process (Kaplan et al., 2009; Valouev et al., 2011; Raveh-Sadka et al., 2012). For example, TFs couple with chromatin remodelers can mediate and maintain the depletion of nucleosomes in the nucleosome-free regions (NFR) which is upstream of the TSS (Transcription start site) in *Saccharomyces cerevisiae* (Ozonov and van Nimwegen, 2013); Histone variant H2A.Z which is enriched in regulatory regions (promoters and enhancers) influences chromatin accessibility in murine ESCs (Embryo stem cells) during differentiation and self-renewal (Dai et al., 2017).

1.1.2 Overview of epigenetic modifications

Besides the structural dynamics, chromatin can also be altered by adding or removing chemical modifications on DNA or histones. A histone octamer contains two copies of H2A, H2B, H3, and H4 which are abundant in lysine and arginine at the N-terminus. Those residues are subjected to various covalent modifications, including methylation, acetylation, ubiquitination, phosphorylation, and sumoylation, *etc*. Histone modifications are widely distributed along chromatin, influencing chromatin accessibility and affecting DNA replication and repair. Histone acetylation reduces the positive charge of the histone surface so that the affinity between histone and DNA is weakened (Robinson et al., 2008). In general, histone acetylation is associated with the active state of chromatin (Allegra et al., 1987). Histone phosphorylation also changes the charge of histones and affects chromatin structure. Although histone phosphorylation was less extensively studied, it is certainly associated with regulation of transcription (Lau and Cheung, 2011) and DNA repair (Su et al., 2017; Van Attikum et al., 2004). Histone ubiquitylation is a large modification that involves

attachment of ubiquitin to histone lysine via an isopeptide bond. This process is mediated by three classes of enzymes: E1 enzymes activate C-terminal glycine residue of ubiquitin, E2 enzymes conjugate ubiquitin, E3 enzymes in the last transfer the ubiquitin from E2 to target substrates (Hershko and Ciechanover, 1998). Histone ubiquitylations have been found associated with both gene silencing (H2A monoubiquitination, H2Aub1) (Blackledge et al., 2014) and transcription activation (H2B monoubiquitination, H2Bub1) (Kim et al., 2009). Similar to ubiquitylation, histone sumoylation adds small ubiquitin-like molecules to histone lysine through the sequential actions of E1, E2, and E3 enzymes. Based on current knowledge, sumoylation is associated with the repressive state of chromatin (Shiio and Eisenman, 2003). Histone methylations are more complex than other modifications, which are organized in various contexts. For example, the methylation on lysine can be mono-, di, tri-methylated, while arginine can be symmetrically mono-methylated or asymmetrically di-methylated. And those different methylations mark specific chromatin states. H3K36 and H3K4 methylations are associated with active transcription, while H3K9 and H3K27 methylations are associated with a repressive state of chromatin. However, “histone codes” are not a naive “on/off” binary setup, as studies have revealed that cross-talk between histone modifications complicate the “code book”. Certain histone modifications can recruit the molecular machinery of other modifications, such as the association between H2Bub1 and H3K4me (Lee et al., 2007a). Some histone modifications repel other modifications. One classic example is the active modification H3K4me3, which inhibit the binding of the molecular machinery of generating H3K27me3 (Schmitges et al., 2011; Schuettengruber et al., 2007). Studies in mammalian cells and tumor cells found that several protein complexes for establishing histone modifications may carry both methyltransferase and demethylase, to coordinate the establishment of active H3K4me while removing the repressive H3K27me3 (Cho et al., 2007; Kim et al., 2014; Wang et al., 2017). Meanwhile, H3K4me3 and H3K27me3 can bivalently target the same gene, subjected to poised chromatin state in mammalian ESCs (Bernstein et al., 2006). And similar bivalency state of H3K4me3 and H3K27me3 has also been identified in Arabidopsis seedlings, although it is not clear if such chromatin feature is present in a specific cell population or prevalent in all cell types (Sequeira-Mendes et al., 2014).

In addition, DNA can also be modified, including the methylation of DNA at the fifth position of cytosine (5mc) and the oxidized forms of 5mc. DNA methylation is widely found among eukaryote organisms. In mammalian cells, DNA methylation mostly occurs at CG sites (CpG sites), whereas methylated cytosines are found in CG, CHG, and CHH context (H can be A, T, C) in plants.

Together, histone and DNA modifications mark distinct chromatin states which can be pass on through cell divisions, known as the "epigenetic memory". Studies have revealed numbers of molecular mechanisms that underlie the inheritance of the "epigenetic memory". One well-known epigenetic regulatory system is coordinated by the Polycomb group (PcG) proteins (Grossniklaus and Paro, 2014) and Trithorax group (TrxG) proteins (Kingston and Tamkun, 2014). PcG and TrxG proteins mediate the cell fate specification through their antagonistic regulation on gene expression, whereas PcG complexes carried out the repressive state of the chromatin through H3K27me while TrxG complexes are involved in the establishment of active chromatin modifications H3K4 and H3K36 (Ringrose et al., 2003; Schuettengruber et al., 2007). In the following sections, I will review the current knowledge of PcG and TrxG regulatory network and respective histone modifications, including H3K4me and H3K27me3. In addition, as DNA methylation is the most prominent heterochromatin mark to maintain the silent state of heterochromatin, I will also briefly introduce the molecular mechanisms for establishing DNA methylation, mostly the RNA-directed DNA methylation pathway (RdDM).

1.1.3 Establishing the repressive chromatin state by PcG complexes

PcG proteins establish the silencing chromatin state via Polycomb repressive complex 2 (PRC2) and PRC1. The PRC2 catalyzes H3K27me3, while PRC1 recognizes H3K27me3 and recruits the transcriptional repression machinery. In addition, PRC2 and PRC1 are not always functioning together, as studies have shown that PRC2 and PRC1 can mediate transcription repression independently (Morey et al., 2012; Wu et al., 2013). PRC2 is conserved among mammalian and plant organisms. In Drosophila, PRC2 consists of four core subunits: enhancer of zeste (E[z]), suppressor of zeste 12 (Su[z]12), extra sex combs (Esc) and the nucleosome-remodeling factor Nurf55 (Lanzuolo and Orlando, 2012). In *Arabidopsis*, PRC2

subunits have several homologs: CURLY LEAF (CLF), MEDEA (MEA), FERTILIZATION INDEPENDENT SEED1 (FIS1) and SWINGER (SWN) are homologs of E(z) (Goodrich et al., 1997; Grossniklaus et al., 1998; Chanvivattana et al., 2004; Gendall et al., 2001); EMBRYONIC FLOWER 2 (EMF2), VERNALIZATION 2 (VRN2) and FIS2 are homologs of Su(z)12 (Yoshida et al., 2001; Gendall et al., 2001; Luo et al., 1999); FERTILISATION INDEPENDENT ENDOSPERM (FIE) is orthologs of Esc (Ohad et al., 1999); MULTIPLE SUPPRESSOR OF IRA 1-5 (MSI1-5) are homologs of Nurf55/p55 (Köhler et al., 2003). Various combinations of the subunits, as mentioned above, form several PRC2 complexes that catalyze H3K27me3 at different developmental stages and tissues. The EMF-PRC2 complex consists of EMF2, CLF/SWN, FIE, and MSI1, regulating various flowering genes during the phase transition from vegetative to reproductive, such as AGAMOUS (AG), APETALA3 (AP3) and FLOWERING LOCUS T (FT). Losing CLF activity causes early flowering, and narrow upwardly curled leave phenotype, while *emf2* mutants display even earlier flowering phenotype than *clf* (Chanvivattana et al., 2004). This indirectly reflects the redundant role of SWN and CLF as methyltransferase. The VRN-PRC2 complex (CLF/SWN, FIE, MSI1/EMF2, and VRN2/VRN5/VEL1/ VIN3) is involved in vernalization by maintaining the silent state of FLOWERING LOCUS C (FLC) (Lister et al., 2004). The expression of *FLC* is quantitatively decreased in accordance with the time course of cold, while the H3K27me3 is accumulated at first exon and intron of *FLC* after several weeks cold. And when the temperature goes back to warm, H3K27me3 spreads to the complete *FLC* locus to maintain the repressive state of *FLC* (Ietswaart et al., 2012). Studies have shown that the down-regulation of *FLC* during the cold is associated with the anti-sense non-coding *FLC* transcript *COOLAIR* and the sense non-coding transcript *COLDAIR*, in which the *COLDAIR* is associated with PRC2 component to drive H3K27me3 accumulation (Ietswaart et al., 2012). Nevertheless, the regulation of FLC is a complex network that also involves the TrxG complex, as well as autonomous pathway (Ietswaart et al., 2012). The FIS-PRC2 contains MEA, FIS2, FIE, and MSI1, which play a role in female gametophyte and seeds development (Ohad et al., 1999; Luo et al., 1999). *MEA* and *FIS2* are only expressed at female gametophyte and endosperm, while FIE and MSI1 are widely expressed in different tissues and developmental stages. The FIS-PRC2 is essential to prevent the proliferation of central cells before fertilization. The disrupted FIS-PRC2 complex in *Arabidopsis* results in the over-proliferation of endosperm (Kiyosue et al., 1999; Guitton et al., 2004).

PRC1 proteins not only have binding ability to H3K27me3 but also have ubiquitylation catalytic capacity to establish H2Aub1. In *Drosophila*, the PRC1 complex contains four subunits, including Polycomb (Pc) which shows binding affinity to H3K27me3, Posterior sex combs (Psc), Polyhomeotic (Ph) and dRING1/Sex combs extra. Psc and dRING1 are the core subunits to catalyze H2Aub1 (Kassis et al., 2017). In *Arabidopsis*, several homologs of PRC1 have been found in recent studies. The ring finger proteins RING1A/RING1B which are orthologs of dRING1, as well as BMI1A/BMI1B mediate the ubiquitination of histone (Xu and Shen, 2008; Bratzel et al., 2012, 2010). LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) which shows binding affinity to H3K27me3, likely plays a similar role as Pc in *Drosophila* (Zhang et al., 2007c). However, the H2Aub1 enrichment seems independent of LHP1 (Zhou et al., 2017), although RING1A/RING1B and BMI1A/BMI1B indeed bind to LHP1 (Xu and Shen, 2008; Bratzel et al., 2010). In addition, EMF1 is a plant specific protein, also plays a role in the PRC1 complex. EMF1 interacts with RING1B, BMI1A/BMI1B, and LHP1. Losing EMF1 results in a reduction of H2Aub1, and the mutant phenotype highly resembles the *emf2* (Yang et al., 2013; Wang et al., 2014). Furthermore, numbers of other protein factors are associated with the PRC1 complex, to "read" the repressive H3K27me3 in *Arabidopsis*. One well-studied example is ASYMMETRIC LEAVE1-2 (AS1-AS2). AS1-AS2 complex interacts with LHP1 and recruits the PRC2 complex to the cis-regulatory region of the class I *KNOTTED1-LIKE HOMEOBOX* (*KNOX*) genes to establish the repressive chromatin state during leaf differentiation (Lodha et al., 2013; Shen and Xu, 2009). The class I *KNOX* genes are important regulators to maintain the shoot apical meristem, and will be introduced in section 1.2. It worth to mention that the AS1-AS2 complex also involves in abaxial and adaxial cell fate determination by negatively regulating *AUXIN RESPONSIVE FACTORS 3/4* (*ARF3/4*) (Iwasaki et al., 2013), which will be introduced in section 1.2 as well.

As one of the major repressive marks on euchromatin, H3K27me3 is conserved among eukaryote organisms. It mostly marks broad genomic regions, covering genic regions as well as intergenic regions (Schwartz et al., 2006; Tolhuis et al., 2006). In *Drosophila*, a single H3K27me3 domain can spread hundreds of kilobases and cover multiple genes, whereas most of the H3K27me3 signal is enriched on a single gene in *Arabidopsis* (Zhang et al., 2007a). In principle, the repressive H3K27me3 is antagonistic to the active H3K4me3, studies based on mass spectrometry have shown that H3K27me3 does not colocalize on the

same H3 with H3K4me3 or H3K36me3 in mammalian cells (Schwämmle et al., 2014; Johnson et al., 2004). However, sequential chromatin immunoprecipitation experiments in mammalian ESCs and *Arabidopsis* seedling have found that H3K4me3 and H3K27me3 can colocalize on the same nucleosome, known as the bivalent state (Bernstein et al., 2006). Meanwhile, DNA methylation is also antagonistic to H3K27me3 in mammals and *Arabidopsis*. Losing DNA methylation leads to the accumulation of H3K27me3 at regions that were previously silenced by DNA methylation (Mathieu et al., 2005; Deleris et al., 2012; Reddington et al., 2013).

1.1.4 Establishing the active chromatin state through TrxG complexes

As the balanced counter partner of PcG proteins, TrxG proteins install the active state of the chromatin. One of the main functions of TrxG proteins is to establish H3K4me. In yeast, methyltransferase SET domain containing1 (Set1) together with other subunits Swd1/2/3, Bre2, Sdc1, Spp1 and Shg1 form Set1/COMPASS (Complex Proteins Associated with Set1), which catalyze methylation on H3K4 (Briggs et al., 2001; Miller et al., 2001; Wilm et al., 2001). Set1 is the core active unit of the enzyme complex. Lack of functional Set1 causes the complete loss of methyltransferase activity (Wilm et al., 2001). In mammalian cells, six SET1 homologs were identified, including SET1A/KMT2F, SET1B/KMT2G, MLL1/KMT2A, MLL2/KMT2B, MLL3/KMT2C and MLL4/KMT2D (Shilatifard, 2012), whereas two groups of TrxG proteins have been found in *Arabidopsis* (Alvarez-Venegas and Avramova, 2002; Springer et al., 2003; Baumbusch et al., 2001), including ARABIDOPSIS HOMOLOGUE OF TRITHORAX 1-5 (ATX1-5), ARABIDOPSIS TRITHORAX-RELATED 1-7 (ATXR1-7). Thereinto, loss of ATX1 causes mild reduction of H3K4me3 globally and strong reduction at some loci (Alvarez-Venegas and Avramova, 2005; Fromm and Avramova, 2014). ATXR7 (known as SDG25) also appears to influence H3K4me3 only on certain genomic loci (Tamada et al., 2009). On the other hand, ATXR3 (known as SDG2) and ATX3-5 likely function as major H3K4me3 methyltransferases, targeting numerous genes in *Arabidopsis* (Min et al., 2010; Chen et al., 2017).

In addition, respective homologs of subunits in COMPASS have also been found in *Arabidopsis*, including HUMAN WDR5 (WD40 REPEAT) HOMOLOG A (WDR5a) and WDR5b

which are homologs of Swd3, RBBP5 LIKE (RBL) which is homolog of Swd1, ARABIDOPSIS ASH2 RELATIVE (ASH2R) which is homolog of Bre2. Mutations of those subunits not only result in the reduction of H3K4me3, but also cause irregular development in *Arabidopsis*. The study has shown that ASH2R, RBL, WDR5a, together with ATX1 involved in flower transition and seeds development (Jiang et al., 2011). It worth to be noted that ATX1 can still recruit PoI II and TATA-binding protein to promoters to initiate transcription in *Arabidopsis* when ATX1's methyltransferase domain was mutated, indicating ATX1/COMPASS complex could regulate the initiation of transcription without the presence of H3K4me3 (Ding et al., 2012).

According to genome-wide profiles, H3K4me1/2/3 mark actively transcribed genomic features, but the precise distribution differs among H3K4me1, H3K4me2, and H3K4me3 (Li et al., 2007; Zhang et al., 2009). According to genomic studies in *Arabidopsis*, H3K4me2 and H3K4me3 mainly mark the promoter and TSS, whereas H3K4me3 appears slightly upstream of H3K4me2 (Zhang et al., 2009). H3K4me1 mostly covers the transcribed gene body while it is less enriched at the promoter regions (Zhang et al., 2009). Other than the localization on active genes, many studies have also discussed the distribution of H3K4me at the intergenic region. In mammalian cells, although H3K4me3 domains on intergenic regions have previously been assumed to be within unannotated promoter/gene or long non-coding RNAs (Mikkelsen et al., 2007; Guttman et al., 2009), further studies suggest that all three states of H3K4me can target enhancers (Koch and Andrau, 2011; Pekowska et al., 2011). H3K4me1 is generally enriched in enhancers but not necessarily associated with enhancer activity (Cui et al., 2009; Creyghton et al., 2010). Meanwhile, H3K4me3 may also target enhancers. A study in human T-cells have found that enhancers of genes involved in T-cell specification have gained enrichment of H3K4me3 during T-cell differentiation (Pekowska et al., 2011). Nevertheless, the association of enhancer activity with H3K4me3 has not been proved to be a prevalent principle to all enhancers, whereas H3K27ac, H3K18ac, and H3K9ac are classic marks for enhancer activity (Creyghton et al., 2010; Jin et al., 2011; Karmodiya et al., 2012; Rada-Iglesias et al., 2011). In fact, because H3K4me3 enrichment is generally low at enhancers than promoters, the ratio of H3K4me3/H3K4me1 was used in many studies to distinguish enhancers and promoters (Kim and Shiekhattar, 2015; Robertson et al., 2008).

1.1.5 Establishing DNA methylation via RNA-directed DNA methylation (RdDM)

As a heterochromatin mark, DNA methylation is essential to repress transposable element, 5S and 45S rRNA gene repeats. Meanwhile, DNA methylation can also be found on some protein coding genes. Although DNA methylation is a stable modification that transmits during cell divisions, it is dynamically regulated, either is added by DNA methyltransferases (DNMTs) or erased via DNA demethylases. In *Arabidopsis*, the *de novo* DNA methylation is carried out by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) which is guided by small interference RNA (siRNA), known as the RNA-directed DNA methylation (RdDM) pathway (Chan et al., 2004; Zilberman et al., 2003, 2004; Kim and Zilberman, 2014). The canonical RdDM is initiated with transcription of heterochromatic loci by RNA polymerase IV (Pol IV) and form dsRNA (26-45nt) by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) which physically interacts with Pol IV (Wierzbicki et al., 2008; Haag et al., 2012; Blevins et al., 2015; Zhai et al., 2015). Pol IV has 12 subunits, in which the largest subunit NUCLEAR RNA POLYMERASE D 1 (NRPD1) binds to NRPD2, forming the catalytic core (Ream et al., 2009). Pol IV-RDR2 produced dsRNA is cleaved by DICER-LIKE 3 (DCL3), which results in the generation of 24nt sRNAs (Zheng et al., 2007; Havecker et al., 2010; Nagano et al., 2014). Thus 24nt sRNAs load into ARGONAUTE4/6 (AGO4/6) and guide AGO4/6 to the target transcripts-chromatin scaffold, mediated by Pol V (Wierzbicki et al., 2008). The Pol V contains twelves subunits, of which the largest subunit NUCLEAR RNA POLYMERASE D1B (NRPE1) interacts with KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1). The KTF1 contains an AGO hook motif that can recruit AGO4/6 (Bies-etheve et al., 2009; He et al., 2009). The AGO4-sRNA recognize the matched Pol V transcripts, and together with RNA-DIRECTED DNA METHYLATION 1 (RDM1), recruit DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) to establish *de novo* DNA methylation (Gao et al., 2010). In addition, it was proposed that SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1) binds to H3K9me2 and SU(VAR)3-9 HOMOLOG 2/9 (SUVH2/9), which then binds to DNA methylation to facilitate the recruitment of Pol IV and Pol V to target loci (Kuhlmann et al., 2012; Law et al., 2013). Other than the canonical RdDM pathway, there are several non-canonical RdDM pathways have been reported, in which the establishment of DNA methylation can also be guided by 21-24nt sRNAs and mediated by alternative complexes, such as TAS-Pol II-RDR6-

DCL2/4 (Allen et al., 2005; Wu et al., 2012), TE-Pol II-RDR6-DCL2/3/4 (McCue et al., 2015; Marí-ordóñez et al., 2013).

The RdDM pathway is regulated non-cell autonomously. 23-24nt sRNAs are believed to move from shoot to roots to spread RdDM silencing signal (Melnyk et al., 2011; Molnar et al., 2010), whereas 21nt tasiRNA mobility has been shown in both short and long-distance transportation (Chitwood et al., 2009; Wu et al., 2012; Molnar et al., 2010). Studies in *Arabidopsis* have reported that the mobile sRNA can spread silencing signals between sperm cells and vegetative cells (Slotkin et al., 2009; Martínez et al., 2016). Based on current knowledge, cell to cell sRNAs movement is through plasmodesmata, of which permeability can be regulated by peroxidases (Liang et al., 2014). On the other hand, long-distance transportation of sRNA is established through the phloem, adopting the source-sink model (from shoot to root). Nevertheless, studies also demonstrated that the adverted long-distance movement of sRNA from root to shoot is possible in *Nicotiana benthamiana* (Bai et al., 2011; Liang et al., 2014).

In principle, besides sRNA, dsRNA can also be the mobile silencing signal. A study has found that dsRNA can be transported from neurons to germline to spread transgenerational silencing memories in *Caenorhabditis elegans* (Devanapally et al., 2015). However, the transmembrane protein SID-1 which enables systemic RNA silencing in *C.elegans* (Shih et al., 2009), does not have obvious orthologues in plant genomes.

1.2 Cell fate specification in the shoot apex

Plants are multicellular organisms, producing new tissues continuously after embryogenesis. The post-embryogenesis is initiated from groups of undifferentiated or less differentiated cells known as meristems that are able to self-renew and differentiate into various tissues, whereas SAMs (Shoot apical meristems) is responsible for the development of aerial tissues. The SAMs are dynamic but highly organized. In *Arabidopsis*, the SAMs are divided into three distinct zones according to cell division rate: central zone (CZ) which contains stem cells that have lower rate of cell division; peripheral zone (PZ) which produces cells for lateral organ development; rib zone (RZ), in which the differentiation of stems occurs (Meyerowitz, 1997)

(Figure 1.1). Besides, the SAM is also consisted by three specific cell layer to coordinate the onset of different cell types, including L1 cells which generate the epidermal cells, L2 cells which give rise to mesophyll cells, and cells in L3 which form other tissue in the center of leaf and stems (Meyerowitz, 1997) (Figure 1.1). It was previously believed that there were three stem cells in each cell layers in *Arabidopsis* (Irish and Sussex, 1992; Schnittger et al., 1996), but recent studies have found around 35 stem cells in CZ (Reddy and Meyerowitz, 2005; Yadav et al., 2009). The maintenance of stem cells is organized by the signal from organizing center cells (OC), which is marked by the expression of a homeodomain transcription factor *WUSCHEL* (*WUS*) (Laux et al., 1996). The WUS plays a role in the maintenance of SAMs by stimulating the *CLAVATA3 (CLV3)* expression, which specifically marks the stem cells in CZ (Laufs et al., 1998; Clark et al., 1995; Fletcher, 1999). In turn, the CLV3 can repress *WUS* expression, which is known as the negative feedback loop to WUS (Fletcher, 1999; Brand et al., 2000).

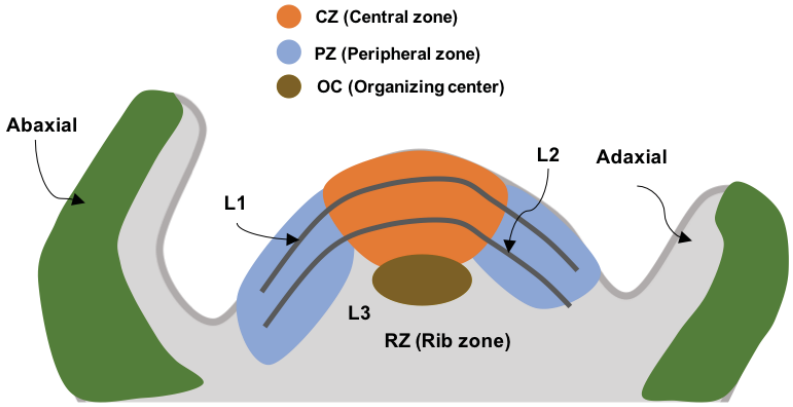


Figure 1.1 Organization of the Arabidopsis shoot apical meristem

Apart from the WUS-CLV3 regulatory network, *SHOOT MERISTEMLESS (STM)*, a gene in class I *KNOX* (knotted1-like homeobox) family, plays a role in meristem maintenance by suppressing differentiation (Scofield et al., 2008). *STM* is expressed in the whole meristem (Smith et al., 1992). Ectopic expression of *STM* causes excessive growth of SAM on the leaf, and loss of *STM* leads to the termination of shoot apex development (Scofield et al., 2008). Besides STM, other KNOX genes are also expressed in SAM, including *KNOTTED-LIKE FROM ARABIDOPSIS THALIANA1* (*KNAT1*, also known as *BP*), *KNAT2,* and *KNAT6*. KNAT1/BP and KNAT6 play redundant roles of STM to maintain SAM (Belles-boix et al., 2006; Byrne et al., 2002). Meanwhile KNAT2 cannot restore the development of SAM in *stm* mutant but mostly

influence carpel formation (Byrne et al., 2002). As mentioned previously, the expression of class I KNOX genes are mediated by PRC2-guided epigenetic machinery (Shen and Xu, 2009).

Once the meristematic genes, such as KNOX genes are repressed in leaf primordium founder cells, the leaf development begins. During the developmental course, the leaf primordium gains two distinct cell types: Adaxial and abaxial. The abaxial side is away from the meristem and later differentiates as the lower part of the leaf, while the adaxial side is toward the meristem and becomes the upper part of the leaf. The adaxial side of the leaf usually is dark green, having well-organized palisade mesophyll cells that efficiently capture sunlight, whereas the abaxial epidermal is enriched in stomata and arranged with spongy mesophyll cells that facilitate gas exchange. The establishment of abaxial and adaxial polarity in leaf primordium is regulated transcriptional as well as post-transcriptional. Various components have been extensively studied in the past.

CLASS III HOMEODOMAIN-LEUCINE ZIPPER (HD-ZIP III) family, including PHABULOSA (PHB), PHAVOLUTA (PHV), and REVOLUTA (REV) are the first gene family identified as determinants for adaxial cell (Mcconnell et al., 2001; Emery et al., 2003). The expression of HD-ZIP III genes is governed by abaxial expressing miR165/166, which mediate the degradation of HD-ZIP III mRNAs (Mcconnell et al., 2001; Juarez et al., 2004; Tang et al., 2003; Mallory et al., 2004). In addition, the different binding affinity between miR165/166 and AGO proteins indirectly regulate the expression of HD-ZIP III family genes. It is known that increased expression of miR165/166 in *ago10* is accompanied by the down-regulation of HD-ZIP III family genes (Liu et al., 2009). Meanwhile, studies have demonstrated that AGO10 sequesters miR165/166 from AGO1, in turn interrupting the cleavage of the HD-ZIP III family genes' transcripts by AGO1-miR165/166 (Zhou et al., 2015; Yu et al., 2017). In addition, post-translational regulation is also part of the network. LITTLE ZIPPERs (ZRPs) are believed to inhibit the activity of HD-ZIP III proteins by forming a non-functional HD-ZIP III/ZRP protein complex (Wenkel et al., 2007). Meanwhile, the expression of *ZRP1/2/3* is directly promoted by REV, forming a negative feedback loop (Brandt et al., 2013).

KANADI (KAN) family includes four members of GARP transcription factors, are recognized as abaxial determinants (Kerstetter et al., 2001; Eshed et al., 2004). Double or triple *kan*

mutants develop adaxialized organs that resemble the phenotype of plants with ectopic expression of HD-ZIP III (Eshed et al., 2004). *KAN* genes are antagonistically interact with HD-ZIP III according to mutant genetic studies (Kerstetter et al., 2001; Emery et al., 2003; Eshed et al., 2004). In addition, KAN1 was able to repress *AS2* expression by binding to the promoter of *AS2* (Wu et al., 2008), in turn AS2 can repress *KAN2* expression (Iwakawa et al., 2007; Lin et al., 2003; Wu et al., 2008).

YABBY (YAB) gene family are also known as abaxial determinants, likely to act downstream of KAN (Eshed et al., 2001). Among six members of YAB gene family in *Arabidopsis*, *FILAMENTOUS FLOWER (FIL/YAB1), YAB2, YAB3, and YAB5* are expressed in abaxial domain of lateral organs, while *CRABS CLAW (CRC)* and *INNER NO OUTER (INO)* are only detected in abaxial domain of carpels and the outer integument of ovules (Sawa et al., 1999; Siegfried et al., 1999; Villanueva et al., 1999). Mutations of *fil* and *yab3* result in narrow leaves and partially loss of abaxial cell fate, whereas *kan1 kan2 fil yab3* quadruple mutant shows loss of blade outgrowth, indicating potential role of YAB family in promoting lamina growth (Sawa et al., 1999; Siegfried et al., 1999; Eshed et al., 2004). Although the detailed role of the YAB family in the AB/AD polarity regulatory network is still unclear, it is believed that YAB proteins act together with LEUNIG (LUG) as a transcriptional repressor (Stahle et al., 2009).

AUXIN RESPONSE FACTORs ARF3/ETTIN and ARF4 are also well-studied regulators in AB/AD specification, whose mutant resembles *kan1 kan2* phenotype (Pekker et al., 2005). Genetic studies have shown that *ett/arf3* suppresses ectopic KAN1 activity, but *kan1 kan2* did not alter ARF3/4 expression (Kelley et al., 2012; Pekker et al., 2005), it is believed that ARF3/4 may function together with KAN proteins as a complex in leaf polarity determination. ARF3/4 abaxial localization is restricted by *TAS3*-derived *trans*-acting short interfering RNA (tasiARF) (Allen et al., 2005). tasiARF biogenesis is initiated by AGO7/miR390-mediated cleavage of *TAS3* transcript which recruits RNA DEPENDENT RNA POLYMERASE 6 (RDR6) and SUPPRESSOR OF GENE SILENCING 3 (SGS3) to generate dsRNA. Mediated by DICER-LIKE 4 (DCL4), dsRNAs are processed to 21nt tasiARF (Yoshikawa et al., 2005; Peragine et al., 2004; Vazquez et al., 2004; Gasciolli et al., 2005; Xie et al., 2005; Adenot et al., 2006; Montgomery et al., 2008). The adaxial restricted expression of *AGO7* and *TAS3* defined the strongest enrichment of tasiARF at the adaxial domain (Chitwood et al., 2009). However, mutants of

tasiARF biogenesis showed minor defects on leaf polarity (Peragine et al., 2004; Vazquez et al., 2004; Hunter et al., 2006). In addition, the AS1-AS2 complex could directly downregulate *ARF3* expression and repress miR166A and *YAB5* at the adaxial domain (Husbands et al., 2015), whereas physical binding of AS1 at *TAS3A* promoter may serve as a protection to the transcripts of *TAS3A* (Husbands et al., 2015).

In contrast, one of the activators among ARFs, *MONOPTEROS* (*MP*) (known as *ARF5*) is expressed in the low-auxin enriched adaxial domain, but only become active in the presence of auxin which is enriched in the abaxial domain, forming auxin signaling maxima in the middle domain (Qi et al., 2014). The auxin-dependent active MP protein promotes *PIN1* expression (Wenzel et al., 2007) and instructs PIN1 polarity directions, which further reinforce the auxin signaling gradient (Bhatia et al., 2016; Qi et al., 2014; Guan et al., 2017). Similarly, MP-mediated auxin signaling was also found in the specification of procambial cells (Wenzel et al., 2007). Meanwhile, KAN1 negatively affects the *PIN1* expression. Studies have shown that the defective vascular tissue phenotype in ectopic *KAN1*-expressing plants was restored by *PIN1* overexpression (Ilegems et al., 2010; Izhaki and Bowman, 2007).

1.3 Cell-type specific analysis

In order to address what governs the process of developing a particular cell type, one common strategy is to carry out the cell-type specific analysis. Years of studies have developed several methods to isolate specific cell types in plants, including Fluorescence-activated cell sorting (FACS), Laser capture microdissection (LCM) (Kerk et al., 2003; Torti et al., 2012), Translating ribosome affinity purification (TRAP), Isolation of nuclei tagged in specific cell types (INTACT), as well as recent adaptations of single-cell sequencing (Table 1.1). Some of the methods based on labeling specific cell types, such as FACS, TRAP, and INTACT, which require to generate respective transgenic report lines. LCM, on the other hand, uses staining methods for the selection of specific cell types, including fluorescence in situ hybridization and immunohistochemistry. Single-cell sequencing does not need additional labeling or staining technique, however, it is still a developing technique and only has been reported in root tissues in *Arabidopsis* (Shulse et al., 2019). FACS and INTACT have already been carried out in various tissues in *Arabidopsis* (Table 1.1), but require

adaptations upon different cell types. For example, FACS has been widely used in root and leaf tissues using protoplast strategy, however it is more difficult to apply in embryo tissues (Slane et al., 2014). Therefore, Slane et al. used FANS (fluorescent-activated sorting of nuclei) to isolate proembryo and suspensor in the embryo. Similarly, the INTACT method in various studies has been adopted with different adjustments upon different tissues (Deal and Henikoff, 2010; You et al., 2017; Moreno-romero et al., 2016).

Table 1.1 Selected studies of cell-type specific analysis in Arabidopsis

Tissue	**Method**	**Profiles**	**Sequencing platform**	**Citation**
Stem cell niche at SAM	FACS	Transcriptome of stem cells (CLV3), rib zone cells (WUS) and peripheral zone cells (FIL)	Affymetrix ATH1	(Yadav et al., 2009)
Stem cell niche at SAM	FACS	Transcriptome of epidermal cells, subepidermal cells, vasculature cells	Affymetrix ATH1	(Yadav et al., 2014)
Embryo	FACS	Transcriptome of proembryo and suspensor	Affymetrix ATH1	(Slane et al., 2014)
Stomata linage	FACS	Transcriptome, H3K4me3 and H3K27me3 on guard mother cells, young and mature guard cells	High-throughput sequencing	(Adrian et al., 2015; Lee et al., 2019)
SAM	Razor dissection	Transcriptome and H3K27me3 on SAM enriched tissue	Agilent 44k array and NimbleGen tiling microarray	(Lafos et al., 2011)
Root	INTACT	Transcriptome, H3K4me3 and H3K27me3 on root hair and non-root hair cells	NimbleGen whole-genome tiling microarrays	(Deal and Henikoff, 2010)
SAM	INTACT	Transcriptome, H3K4me3 and H3K27me3 in SAM during flowering course (0D, 1D, 2D and 3D)	High throughput sequencing	(You et al., 2017)
Endosperm	INTACT	H3K27me3 and DNA methylation profiles in endosperm	High throughput sequencing	(Moreno-romero et al., 2016)
Root	Microfluidic cell sorting	Various cell types of root	Single-cell RNA-seq	(Shulse et al., 2019)

Despite of technique differences, in the combination of genome-wide sequencing on expression and histone modification profiles, cell-type specific studies have clearly demonstrated that different cell types not only harbor distinct transcriptome profiles but are also dynamically regulated on post-transcription level in plants (Yadav et al., 2014; You et al., 2017; Lafos et al., 2011). The previous study in our lab had compared the H3K27me3 profile between hand-dissected meristematic tissue to differentiated leaf tissue, and identified hundreds of genes were differentially methylated of H3K27me3 during differentiation (Lafos et al., 2011). Together with other analyses on specific cell-types, suggesting that the single tissue/cell analysis is crucial for understanding the molecular mechanisms of cell fate determination.

1.4 Aim of the study

As a multicellular organism, plants have differentiated into different cell types with specialized functions. Studies have already shown that transcription and chromatin states are dynamically regulated upon different cell types (Lafos et al., 2011). However, the detailed mechanisms underlying the cell fate determination are largely unknown. In order to further dissect the cell fate developing trajectory during cell differentiation, we adopted the INTACT method (Deal and Henikoff, 2010) to isolate nuclei from stem cells and abaxial cells from *Arabidopsis* seedlings and generated respective transcriptome, H3K4me3, and H3K27me3 profiles. In this study, we firstly discussed technique adaptation toward INTACT methods and addressed the potential limitations of the technique. Then we described the overall profiles of transcription, H3K4me3, and H3K27me3 in different cell types. Although it did not generate convincing datasets for stem cells, we confirmed the accepted regulatory network of abaxial/adaxial cell fate determination at both transcription and post-transcription level. Furthermore, we also identified some other potential regulators in abaxial/adaxial cell fate determination, as well as the candidate mobile signals between abaxial and adaxial.

2. Materials and Methods

2.1 Materials

2.1.1 Media

LB, 1L (pH 7.0)

NaCl	10g
Yeast extract	5g
Bacto-Tryptone	10g
Agar	10g (1%)

YEB, 1L (pH 7.2)

Sucrose	5g
Yeast extract	1g
Peptone	5g
Beef extract	5g
MgCl	0.5g
Agar	10g (1%)

Plants growth media

1/2MS, 1L (pH 5.7)

MS	2.2g
MES	0.5g
Sucrose	5g
Agar	8g (0.8%)

2.1.2 Antibodies

	Company	Dilution
H3K4me3	Diagenode	0.5ug per IP
H3K27me3	Diagenode	0.5ug per IP

2.1.3 Oligonucleotides

Table 2.1 Oligonucleotides for cloning

FIL-promoter-F (AT2G45190)	TGAGATTTGTAATGGTTAATATTG
FIL-promoter-R (AT2G45190)	GGAGGACATAGACGACATAGA
UFO-promoter-F	TTTAGAAAGAGATGCTTCATTAA
UFO-promoter-R	GTTATTGATGAACACAGTTGAAT

Table 2.2 Oligonucleotides for qRT-PCR

CLV3-RT-F (AT2G27250)	ACTACCTGAGCACCCAGTCC
CLV3-RT-R (AT2G27250)	CTTGTACAGCTCGTCCATGC
GFP-RT-F	TGACCCTGAAGTTCATCTGC
GFP-RT-R	GAAGTCGTGCTGCTTCATGT
FIL-RT-F (AT2G45190)	AACCATCCTTGCGGTTAATG
FIL-RT-R (AT2G45190)	TTAACGGGTGGTGCTTTAGG
Rubisco-RT-F (AT5G38430)	CCACACCTTCATGCAGCTAA
Rubisco-RT-R (AT5G38430)	CACTGGTTTGAAGTCATCCG
N-black (AT4G34270)	GTGAAAACTGTTGGAGAGAAGCAA
N-black (AT4G34270)	TCAACTGGATACCCTTTCGCA

Table 2.3 Oligonucleotides for ChIP-qPCR

ACT7-F (AT5G09810)	CCAGGAATTGCTGACCGTAT
ACT7-R (AT5G09810)	GGTGCAACCACCTTGATCTT
ACT7-F-ATG (AT5G09810)	TAGTGAAAAATGGCCGATGG
ACT7-R-ATG (AT5G09810)	CCATTCCAGTTCCATTGTCA
FUS3-F (AT3G26790)	GTGGCAAGTGTTGATCATGG
FUS3-R (AT3G26790)	AGTTGGCACGTGGGAAATAG
CLV3-ATG-F (AT2G27250)	TTGGACTGTCCCCTTCTCAT
CLV3-ATG-R (AT2G27250)	GAGAGAAAGTGACTGAGTGAGAGAGA
CLV3-3'-F (AT2G27250)	GGTTTTACCATTTCGGGAGTC
CLV3-3'-R (AT2G27250)	ACCAAACGAAACAGATTGCAC
CLV3-exon1-F (AT2G27250)	TGGATTCGAAGAGTTTTCTGC
CLV3-exon1-R (AT2G27250)	CAGAAGCATCATGAAGGAACA
CLV3-exon3-F (AT2G27250)	GGTTTTACCATTTCGGGAGTC

CLV3-exon3-R (AT2G27250)	ACCAAACGAAACAGATTGCAC
FIL-exon1-F (AT2G45190)	ACCGGACCACTTCTCTCCTT
FIL-exon1-R (AT2G45190)	GGTTTGGCAAAAGTTGCATT
FIL-exon2-F (AT2G45190)	ACCAATCTCCTTTCGGTGAA
FIL-exon2-R (AT2G45190)	CTGGAGCTGGAGCTGGTTAG
FIL-exon6-F (AT2G45190)	CCACTTCCCCCACATACACT
FIL-exon6-R (AT2G45190)	GCATGTTGGTTTTCTTCACG
FIL-exon7-F (AT2G45190)	GAGGGAGAGGATAACATGGTGA
FIL-exon7-R (AT2G45190)	AACGTTAGCAGCTGCAGGA
STM-ATG-F (AT1G62360)	GGCTGACCCGATGTATCTTC
STM-ATG-R (AT1G62360)	CCTGAAACTGGGACATGGTT
STM-F (AT1G62360)	TAGGACACATCGGACCATCA
STM-R (AT1G62360)	GAGTGGTTCCAACAGCACTTC
PIN7-ATG-F (AT1G23080)	CCGGCGAACAACAATATGAT
PIN7-ATG-R (AT1G23080)	CGTGAGGACGGTGTAGAGGT

2.1.4 Constructs and Vectors

FIL::NTF, 6Kb upstream fragment of AT2G45190 (*FIL*) was cloned to TOPO vector and inserted to the pGreen NTF vector by Gateway system. UFO::NTF, 2.5Kb upstream fragment of AT1G30950 (*UFO*) was cloned and introduced to the pGreen NTF vector by Gateway system. All of the oligonucleotides used for fragment amplification were listed in Table2.1. All of the vectors used for INTACT were listed in Table 2.4 and 2.5.

Table 2.4 Entry vector

	Vector backbone	Backbone source	Construct	Oligos
FIL-pCR8GW-TOPO	pCR8/GW-TOPO	Invitrogen	FIL-promoter	FIL-promoter-F/R
UFO-pCR8GW-TOPO	pCR8/GW-TOPO	Invitrogen	UFO-promoter	UFO-promoter-F/R

Table 2.5 Destination vector

	Vector backbone	Backbone source

FIL::NTF	pGREEN-GW-NPT-NTF	Deal et al.,2010
UFO::NTF	pGREEN-GW-NPT-NTF	Deal et al.,2010

Table 2.6 Other vectors used in this study

	Created by
CLV3::NTF	Asif Arif
GATA23::NTF	Asif Arif
BBM::NTF	Asif Arif
CLV3::H2B-YFP	Helge Pallakies

2.2 Plant growth condition

All INTACT lines used in this study were in Columbia (Col) background. The INTACT NTF constructs were transformed into the ACT::BirA line (Deal and Henikoff, 2011). Then, INTACT lines were selected by Kanamycin for three generations. In T1, lines with INTACT construct were selected. In T2, lines segregated in the 3:1 ratio were selected for a single copy of the INTACT construct. In T3, lines with no segregation were selected for homozygous of the INTACT construct. All of the INTACT lines' expression patterns of GFP were screened under the confocal microscope.

Seeds sown on 1/2MS plate were surface sterilized with 70% ethanol and 0.05% TritonX-100 for 2min. Plants were grown in plate or soil, under Long day (LD) condition.

2.3 Methods

2.3.1 Isolation of Nuclei Tagged in specific Cell Types (INTACT)

The INTACT method was based on Deal et al. (Deal and Henikoff, 2011) with modifications. The streptavidin Dynabeads were pre-treated in 1%BR_NPB (1% blocking reagent from Roche, NPB buffer: 20mM MOPS pH 7; 40mM NaCl; 90mM KCl; 2mM EDTA; 0,5mM EGTA; 0,5mM spermidine; 0,2mM spermine; 1:200 protease inhibitor cocktail from Sigma) at least half an hour at room temperature with rotation. 5D old seedlings were collected. If the nuclei were isolated for the ChIP experiment (Chromatin immunoprecipitation), seedlings

were crosslinked in NPBf (NPB buffer with 1% paraformaldehyde and 0,1% Triton-X 100) for 15min on ice with the vacuum. The crosslink was stopped by adding glycine (final concentration: 0.125M) with a 5min vacuum on ice. If the nuclei were used for RNA extraction, the crosslinking procedure listed above was skipped. The collected tissue was grinded in liquid N2 and homogenized in 50ml NPB. Then the tissue suspension was filtered through the following filters: mirocloth (Millipore, 22uM); cell strainer (CORNING, 40uM); mirocloth (Millipore, 22uM). The filtered suspension was centrifuged at 1000g at 4 degree for 10 min. The nuclei pellet was resuspended in 1ml NPB buffer with 2ug/ml DAPI, and incubate on ice for 3 min. Then the nuclei suspension was centrifuged at 1000g at 4 degree for 5 min. The nuclei pellet was resuspended in 1.5ml ice-cold NPBb (NPB buffer with 1% BSA from Sigma). The pretreated streptavidin beads were washed by NPBb buffer on a magnetic stand (Invitrogen). The streptavidin beads slurry was resuspended in 20ul NPBb buffer and added into nuclei suspension. The nuclei and streptavidin beads mixture was incubated with rotation for 30 min at 4 degree, which then was diluted with 9ml NPBtb (NPB buffer with 0,1% TritonX-100 and 1% BSA). The beads-bound nuclei were captured using magnet stand (Invitrogen) and washed by NPBtb, which repeated 10 times. After the last wash, the nuclei pellet was resuspended in 1ml NPBtb buffer and 10ul purified beads and nuclei mixture were used to analyze the purity and yield under the fluorescence microscope. At last, the beads-bound nuclei slurry was captured by the magnet and resuspend in 1ml Trizol (Invitrogen) for RNA analysis or 200-300ul nuclei lysis buffer for ChIP analysis.

Another method used to evaluate the %contamination in INTACT was inspired by Jordi Moreno-Romero et al. (Moreno-romero et al., 2016). By mixing Ler-0 seedlings with INTACT lines which were in Col background, the Ler DNA in INTACT isolated nuclei and the starting material was quantified by qPCR. And the % of contamination was evaluated by Equation 1. Ta1-2 is a Ler specific retrotransposon, and GFP is only present in the INTACT line (Equation 1). Based on the ratio of INTACT line and *Ler*-0 in the whole nuclei extraction before the INTACT isolation ($Ct_{(GFP\_starting)}$ and $Ct_{(Ta1\text{-}2\_starting)}$), the amount of DNA (INTACT line) from unspecific binding of nuclei or chromatin in the INTACT isolated nuclei can be estimated by the quantity of *Ler* DNA in the INTACT isolated nuclei ($Ct_{(Ta1\text{-}2\_INTACT)}$). Thus, the %contamination is the ratio of the INTACT line DNA from unspecific binding to the total

INTACT line DNA in the INTACT isolated nuclei ($Ct_{(GFP\_INTACT)}$). In the equation, the $E_{(GFP)}$ and $E_{(Ta1\text{-}2)}$ represent the primer efficiency of respective gene locus. In principle, if there is no contamination, one would expect no Ct value from *Ler* specific locus *Ta1-2* in the INTACT isolated nuclei.

$$\%Contamination = \frac{\left(\frac{{E_{(GFP)}}^{-Ct_{(GFP\_starting)}}}{{E_{(Ta1-2)}}^{-Ct_{(Ta1-2\_starting)}}}\right) \times {E_{(Ta1-2)}}^{-Ct_{(Ta1-2\_INTACT)}}}{{E_{(GFP)}}^{-Ct_{(GFP\_INTACT)}}} \times 100\% \quad \text{(Equation 1)}$$

2.3.2 Extraction of DNA from INTACT isolated nuclei

The nuclei isolated from INTACT were resuspended in 500ul Nuclei Lysis Buffer (50mM Tris, pH 8; 10mM EDTA; 1% SDS; 0,2mM PEFABLOC; 1:200 protease inhibitor cocktail from Sigma), followed by 2min incubation with gentle votexing. 500ul phenol/chloroform/Isoamyl alcohol (Roth, phenol/chloroform/Isoamyl alcohol at a ratio of 25:24:1) was added and mixed to the nuclei suspension, followed by 5min centrifugation at 16000g. The upper aqueous phase was transferred to a new tube and then mixed with 0.7 volumes of isopropanol 1/10th volumes of 3M sodium acetate and 2ul glycogen. After 2 hours incubation at -70 degree, the DNA was precipitated, and pelleted after 30min centrifugation at 16000g at 4 degree. The DNA pellet was washed by 75% ethanol and dried in air at room temperature. The DNA pellet resolved in 20-50ul nuclease-free TE buffer (10mM Tris pH 8; 1mM EDTA)

2.3.3 Extraction of RNA from INTACT isolated nuclei

INTACT isolated nuclei were resuspended in 1ml Trizol (Ambion). After votexing and 5min incubation at room temperature, the nuclei-trizol sample was kept in -70 for at least 1 hour. Then the nuclei-trizol sample was thawed at room temperature. 200ul chloroform was added and vigorously shaken for 15 sec. The mixture was then incubated at room temperature for 2-3 min. Followed by 15min centrifugation at 12000g at 4 degree, two phases were separated. The aqueous phase was carefully transferred to a new eppendorf tube. And the aqueous phase was mixed with a half volume of Isopropanol and 2ul glycogen and incubated at -70 degree for at least 1 hour to precipitate RNA. Then the mixture was

centrifuged at 12000g at 4 degree for 20min. And pellet was washed by 1ml 75% cold Ethanol and followed by 15 min centrifugation at 7400g at 4 degree. In the end, the RNA pellet is dried in air for 5-10 min and resuspend in 10-20ul RNase free water.

2.3.4 cDNA and dscDNA synthesis

The cDNA and dscDNA synthesis were prepared by using the SMART cDNA synthesis kit (Clontech), following the manufacture's protocol.

2.3.5 Quantitative Real-time polymerase Chain reaction (qRT-PCR)

The quantitative real-time PCR was performed on dscDNA that synthesized from RNA or ChIP-DNA. The oligonucleotides used in qRT-PCR listed in table 2.2. The qRT-PCR reactions were prepared with KAPA SYBR FAST qPCR kit and run in Roche lightcycler 480. The reference gene At4g34270 was used in $\Delta\Delta$Ct calculation (Czechowski et al., 2005). The $\Delta\Delta$Ct calculation was previously described (Simon, 2003; Yuan et al., 2006).

2.3.6 Chromatin immunoprecipitation (ChIP) on INTACT isolated nuclei

The ChIP protocol was based on Schubert et al. (Schubert et al., 2005) with modifications. The protein A magnetic beads were conjugated with respective antibody by incubating at 4 degree with rotation. In general, each IP required 10ul beads with 0,5ug antibody. The nuclei isolated from INTACT were resuspended in 200-300ul nuclei lysis buffer (50mM Tris, pH 8; 10mM EDTA; 1% SDS; 0,2mM PEFABLOC; 1:200 protease inhibitor cocktail from Sigma). The nuclei suspension was sonicated using the bioruptor (Diagenode) with 20 cycles 30sec on and 30sec off at high power. The sonicated chromatin solution was centrifuged for 5min at 12000g at 4 degree and the supernatant of chromatin was transferred into a new tube. 10-20ul chromatin was taken as input and the remaining chromatin was diluted 10 times by ChIP dilution buffer (1,1% TritonX-100; 1,2mM EDTA; 16,7mM Tris, pH 8; 167mM NaCl; 0,2mM PEFABLOC). The antibody-coated protein A beads were added to diluted chromatin solution and incubated at 4 degree for 4hours to overnight. The beads-chromatin was washed in washing solution series twice (Low salt buffer: 150mM NaCl; 0,1% SDS; 1%

TritonX-100; 2mM EDTA; 20mM Tris pH 8. High salt buffer: 500mM NaCl; 0,1% SDS; 1% TritonX-100; 2mM EDTA; 20mM Tris pH 8. Lithiumchlorid salt buffer: 0,25M LiCl; 1% NP-40; 1% sodium deoxycholate; 1mM EDTA; 10mM Tris pH 8. TE buffer: 10mM Tris pH 8; 1mM EDTA). The immune complexes were eluted from beads after two times incubation of beads in 250ul elution buffer for 15min at 65 degree (1% SDS; 0,1M $NaHCO_3$). The combined 500ul eluted chromatin was incubated at 65 degree overnight with an additional 20ul 5M NaCl in the solution to reverse crosslink. The reverse crosslinked chromatin was treated by 2ul 10mg/ml proteinase K with 10ul 0,5M EDTA and 20ul 1M Tris (pH 6,5) at 45 degree for 1 hour to remove the protein. The DNA extraction was done by Phenol chloroform protocol listed above. Loci of interest were analyzed by qPCR on Input and IP. %input was calculated by IP signal versus Input signal with adjustment of dilution of starting chromatin.

2.3.7 Library preparation

The library was prepared by using the Microplex library preparation kit (Diagenode) following the manufacturer's protocol. The library size and quantity were analyzed by Agilent 2100 bioanalyzer using the high-sensitive chip. And the enrichment of some key genes was calculated using qPCR to validate the library. $(E_{(gene)} \text{ ^ } (-Ct_{(gene)}))*10^8$ was used to represent the signal of the tested locus. $E_{(gene)}$ is the primer efficiency of the tested locus.

2.3.8 Confocal microscope

The expression pattern of GFP in each INTACT line was analyzed using ZEISS LSM 510 or ZEISS LSM 780 with 40X objective, with the following filter and lasers setting: GFP: Emission 488nm Agron laser; PI: Emission 561nm. The images and projections were processed in the ZEN 2011 software.

2.3.9 Sequencing analysis

High throughput sequencing was conducted by BGI using Illumina Hiseq2000, pair-end with 100bp reads length. After clipped adapter sequences, all the low-quality reads were

removed, with the following criteria: Quality value<=20 is no less than 50%, or the rate of "N" in a read is no less than 10%.

ChIP-seq data

The filtered reads from the ChIP-seq experiment were mapped against the TAIR10 genomic sequence using Bowtie2, not allowing mismatches. The mapping output was shown in Table 2.7. The correlation of each replicates was assessed by calculating the Pearson correlation coefficient in R.

Table 2.7 Reads count of each ChIP-seq samples

	Clean reads	Mapped reads
FIL_Input	31,754,050	30,893,515
FIL_H3K4me3_I	39,954,132	38,631,650
FIL_H3K4me3_II	36,367,276	35,494,461
FIL_H3K4me3_III	34,668,750	31,465,358
FIL_H3K27me3_I	39,474,504	38,499,484
FIL_H3K27me3_II	32,660,974	31,641,952
FIL_H3K27me3_III	42,394,322	41,359,901
SC_Input	35,566,872	33,813,425
SC_H3K4me3_I	24,972,318	21,823,309
SC_H3K4me3_II	25,392,536	23,759,796
SC_H3K4me3_III	26,541,194	25,641,448
SC_H3K27me3_I	25,360,566	22,789,005
SC_H3K27me3_II	29,751,500	27,540,964
SC_H3K27me3_III	29,582,786	28,565,139
UN_Input	27,574,994	27,164,127
UN_H3K27me3_I	32,387,378	31,049,779
UN_H3K27me3_II	29,989,000	29,470,190
UN_H3K4me3_I	27,412,940	26,722,134
UN_H3K4me3_II	32,982,346	32,372,173

ChIP tag enriched domains were calculated by SICER algorism (Zang et al., 2009), using the following parameters: window size=300, gap size=300, duplicate reads=1, FDR=0.01.

Differential methylation analysis was conducted by SICER_df pipeline from the SICER package. Histone modification domains and differential methylated regions were annotated by genomic features (Genes, promoter, intergenic, or transposon elements) using TAIR10. Other bioinformatic analyses were using the following tools and algorisms: Bedtools (Quinlan and Hall, 2010); Seqplot in R (Stempor and Ahringer, 2016); ChIPpeakAnno in R (Zhu et al., 2010); DREME; ggplot2 in R (Wickham, 2016) and statistic tests in R.

RNA-seq data

The filtered reads from the RNA-seq experiment were trimmed to remove additional adapters that were ligated during dscDNA library preparation (5'-AAGCAGTGGTATCAACGCAGAGTAC-3'). The reads length smaller than 15 bp were filtered out. And the reads passed the selection were mapped against TAIR10 genomic sequence with Tophat2, which allows detecting the junction of sequences. The correlation of each replicate was assessed by calculating the Pearson correlation coefficient in R. Differential expression analysis was carried out by DEseq2 in R (Love et al., 2014). GO enrichment analysis was conducted on Gene Ontology Consortium (Carbon et al., 2019; Ashburner et al., 2000).

Table 2.8 Reads count of each RNA-seq samples

	Clean reads	Mapped reads
FIL_RNA_I	7,801,176	6,413,019
FIL_RNA_II	6,195,284	5,076,554
SC_RNA_I	3,775,854	2,975,134
SC_RNA_II	3,161,770	2,504,833
UN_RNA_I	6,846,760	5,579,838
UN_RNA_II	7,797,686	6,379,164

2.3.10 Optimization of SICER parameter for Peak-calling

To analyze H3K4me3 and H3K27me3 profiles in FIL labeled and non-FIL labeled cells, we firstly tested different parameter settings to optimize the peak calling performance in SICER. In general, the SICER algorism splits the genome into non-overlapping windows (w) and scores each window that passes for the negative logarithm of the probability of finding in a random background model. The windows which passed the threshold were further clustered as islands based on the setting of gap size(g). If g=0, the island only consists of continued windows. Window size influences the random background island score. The smaller window size will define a lower background score threshold. And both window size and gap size influence the resolution of islands (domains) (Zang et al., 2009).

To select the ideal window size, we run SICER on H3K4me3 and H3K27me3 ChIP-seq data from FIL-labeled cells (FIL), using following window size: 100bp (sequencing read length=100bp), 150bp (roughly equal to a nucleosome size), 200bp, 300bp, 500bp and 1000bp. The islands identified under different window settings were largely overlapping for both H3K4me3 and H3K27me3, despite SICER identified a few hundred more unique domains with window size=100bp (Figure 2.1 A and B). In fact, we found that different window sizes showed a larger impact on the coverage of islands (Figure 2.1 C-F). Under all window settings, the majority of the genes were covered by a single domain for both H3K4me3 and H3K27me3 (Figure 2.1 C and D). Nevertheless, we observed an increased number of genes covered by multiple islands in H3K4me3 under larger window size settings, which was caused by the expansion of the H3K4me3 domain from neighbor genes (Figure 2.1 C). In contrast, larger window size settings in H3K27me3 lead to fewer genes covered by multiple islands, which corresponded to its broad domain character and suggested the importance of the gap setting (Figure 2.1 D). Furthermore, comparing the average coverage of z-score normalized H3K4me3/H3K27me3 that calculated based on different window sizes, we found that the average coverage of the signals from w=100bp, 150bp, 200bp, and 300bp closely resembled the distribution of raw reads count (Figure 2.1 E and F). On the other hand, w=500bp and 1000bp showed an expanded domain with reduced enrichment (Figure 2.1 E and F). Together, we chose w=300 for both H3K4me3 and H3K27me3, which reflected actual reads distribution while using the least computer power.

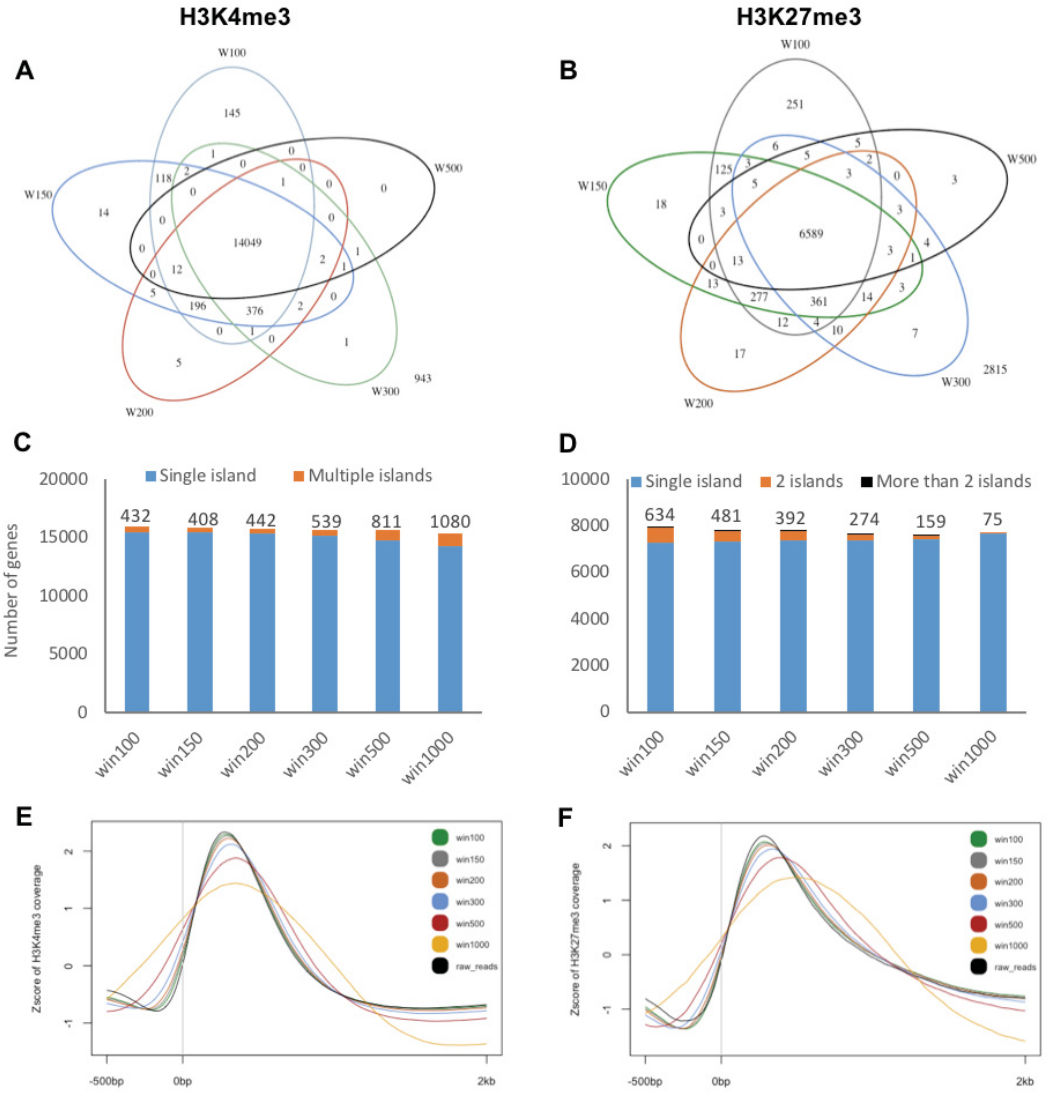


Figure 2.1. Comparison of peak calling performance upon different window sizes in SICER. (A) (C) and (E) Analysis output on ChIP-seq data of H3K4me3. (B) (D) and (F) Analysis output on ChIP-seq data of H3K27me3. (A and B) Venn diagram of islands (domains) that were detected under different window size settings, including 100bp, 150bp, 200bp, 300bp, 500bp, and 1000bp. (C and D) count of genes that were covered by a single island, 2 islands or multiple islands; in figure (C), the category of 2 islands and multiple islands were merged with multiple islands. (E and F) Z-score of average coverage of signals on target genes under different window size settings, 100bp (green), 150bp (gray), 200bp (orange), 300bp (blue), 500bp (red) and 1000bp (yellow), black line is the reference line using raw reads count; -500bp, 0bp, and 2Kb on the X axis represent the 500bp upstream of TSS, TSS and 2Kb downstream of TSS.

We then tested several gap sizes (g) on ChIP-seq data of H3K4me3 and H3K27me3 in FIL-labeled cells (FIL), including gap=0, 300bp, 600bp and 900bp (gap size should be in units of window size).

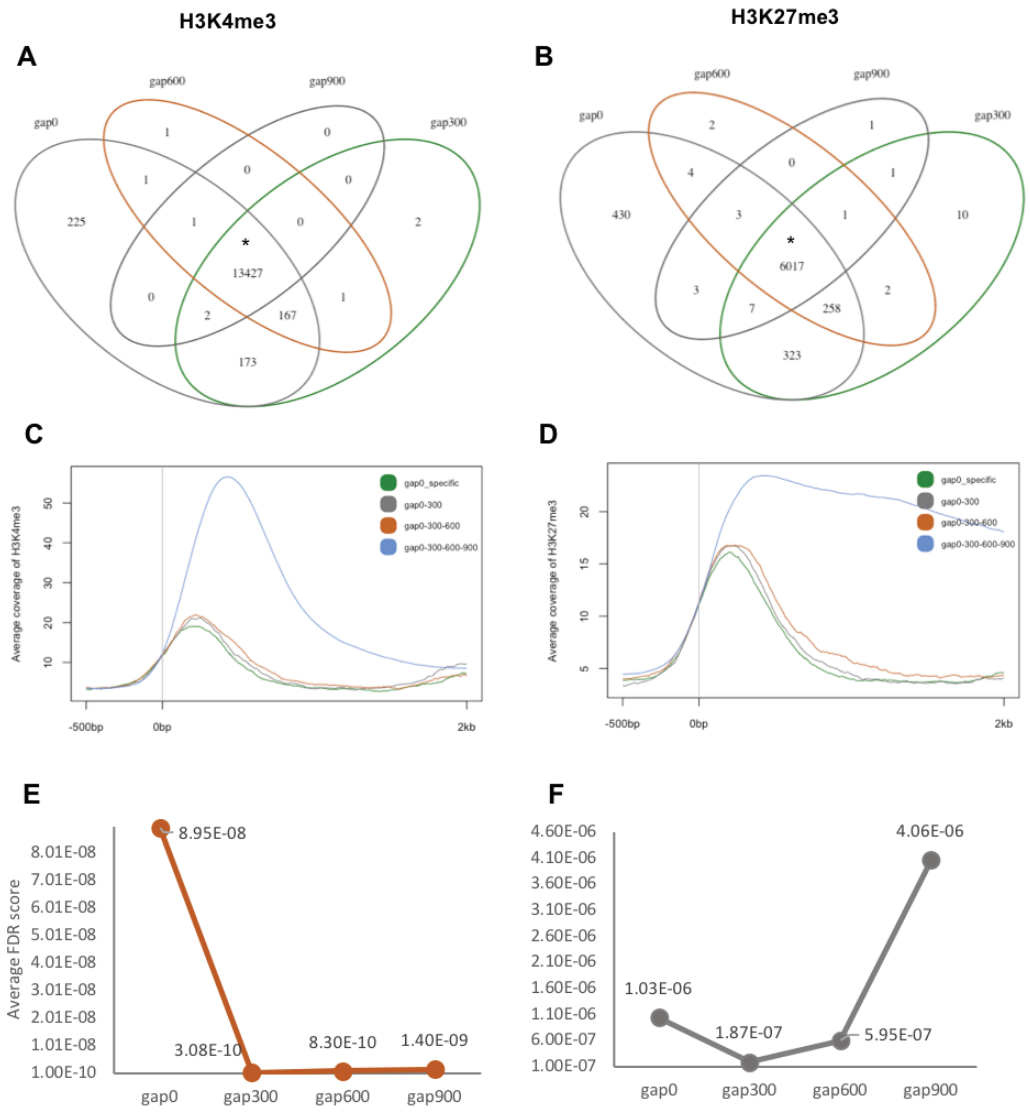


Figure 2.2. Comparison of peak calling performance upon different gap sizes in SICER. (A) (C) and (E) Analysis output on ChIP-seq data of H3K4me3. (B) (D) and (F) Analysis output on ChIP-seq data of H3K27me3. (A and B) Venn-diagram of islands (domain) that were identified under gap=0bp, 300bp, 600bp, and 900bp. (C and D) Average coverage of signals on domains from gap=0 (green line), overlapped islands of gap=0 and 300 (gray line), overlapped islands of gap=0 and 300 and 600 (orange), or overlapped domains in all four gap settings (blue line). (E and F) Average FDR of islands located on the overlapped islands, marked as * in figure (A) and (B). In gap parameter analysis, window=300bp, FDR threshold=0.01.

The islands identified under different gap sizes were mostly overlapped for both H3K4me3 and H3K27me3, despite SICER defined a few hundred more unique domains without a gap (Figure 2.2 A and B). In addition, those non-overlapped islands from different gap sizes were shown lower signal intensity compared to the overlapped islands (Figure 2.2 C and D), indicating a lower probability as a true domain. Therefore, to evaluate the confidence of islands identified among different gap sizes, we compare the FDR value given by SICER algorism. Although most of the islands were given FDR=0 in all gap settings, the sum of FDRs was the smallest for gap=300bp in both H3K4me3 and H3K27me3 (Figure 2.2 E and F). This suggested that islands (domains) of H3K4me3 and H3K27me3 had the highest probability as true domains under gap=300bp (1 window size).

2.3.11 Data analysis using published datasets

The following datasets from previous studies were used in our analysis: Microarray data of gene expression profile in CLV3-expression domain (Yadav et al., 2009) (GSE13596); transcriptome of INTACT isolated SAM dataset (You et al., 2017) (E-MTAB-5130); RDR2/NRPD1 dependent dsRNA(P4RNA) dataset were from Li et al. (Li et al., 2015) (GSE57215); Expression profile of *nrpd1* were from Yang et al. (Yang et al., 2017) (GSE98286); H3K4me3 profile in GL2-labeled cells and ADF8-labeled cells were from Deal et al. (Deal and Henikoff, 2010) (GSE19654); expression profile of leaf (GSE5630) and root tissue (GSE5631) from Schmid et al. (Schmid et al., 2005); expression and H3K4me3 profile from dehydration experiment (Dijk et al., 2010) (GSE11657); expression profiles in *morc6* were from Liu et al. (Liu et al., 2016) (GSM2125312).

3. Results

3.1 Isolation of specific cell types from *Arabidopsis*

Plants are multicellular organisms, in which different cell types are derived from undifferentiated stem cells. One of the key questions in developmental biology is to address the molecular mechanisms underlying cell fate establishment and maintenance. Cell-type specific profiling of transcription and chromatin states provide essential data to decipher respective machineries. However, isolation of single cell type can be challenging in plants, since its cells cannot be cultured. In turn, plant researchers developed several methods to enrich the target cell population from heterogeneous tissues. In this study, I tried two different approaches to isolate target cells residing in the aerial vegetative tissues.

One of the strategies we used to isolate specific cell types is FACS. This method is based on the flow cytometry technique to identify and sort a particular cell population by fluorescence, cell size, and internal complexity. An essential step in FACS application is the preparation of single-cell suspension, which can be the protoplast or the nucleus (Birnbaum et al., 2003; Carter et al., 2013; Yadav et al., 2009; Zhang et al., 2008). The protoplast is usually generated by enzymatically removing the cell wall (Birnbaum et al., 2003; Yadav et al., 2009), which is a stress shock to cells and results in an artifact in the downstream experiment. In addition, because some cell types are resistant to enzymatic treatment, the efficiency of protoplasting varies among different tissues (Slane et al., 2014). On the other hand, nuclei extraction is a relatively simple procedure that has been used in FACS analysis. It can avoid the artifact caused by enzymatic treatment in the protoplasting process. Therefore, we extracted nuclei instead of protoplasting for sorting, which was described as FANS (Slane et al., 2014). It should be noted that the RNA in nuclei are precursor mRNA at different stages, which is not the same as the mature RNA isolated from the cell.

Another method we adopted to isolate specific cell types was Isolation of Nuclei Tagged in specific Cell Types (INTACT) (Deal and Henikoff, 2011). The principle of this method is labeling the nuclei of interest by expressing a biotinylated nuclear envelope protein, and then enrich the nuclei of interest by the interaction between biotin and streptavidin. The

INTACT system requires two constructs. One is the fusion protein (NTF) driven by a cell-type specific promoter. The NTF consists of a nuclear envelope targeted protein, a green fluorescence protein (GFP), and a biotin ligase recognition sequence (BLRP), which is the substrate of E. coli biotin ligase BirA. Another construct is constitutively expressing BirA in all cell types, which is driven by the promoter of ACTIN2. Thus with the presence of both fusion protein and BirA, the nuclei of interest are labeled by biotin (Deal and Henikoff, 2011).

3.1.1 Application of Fluorescence-activated Nuclei sorting to enrich nuclei from meristematic tissues

To isolate stem cell in shoot apical meristem of *Arabidopsis*, we used report line which expressed HISTONE 2B-yellow fluorescent fusion protein in CLV3-expressing domain (CLV3::H2B-YFP) and wild type Col line (non-reporter line) to set up the isolation procedure and parameters in FACS machine (Figure3.1 A and B). In general, several variables are used in the FACS machine to separate different nuclei populations, including side scatter which mainly differentiates the complexity of nuclei, and wavelength of fluorescent emission which separate the fluorophore-labeled nuclei (Figure 3.1 E-J). In our case, the combination of propidium iodide wavelength (610/20-A filter) and side scatter were used to define nuclei population (gate P9 and P10) from background particles (gate P1) (Figure 3.1 G and H). And a 530/30-A filter was used to distinguish the YFP-positive nuclei (gate P9) from the unlabeled nuclei (gate P10) (Figure 3.1 E, F, I, and J). We firstly tested methods of preparing single-nuclei suspension in *Arabidopsis*: chopping by a razor blade or grinding with liquid nitrogen. For the chopping method, we can only work on a small scale of *Arabidopsis* tissue (30-50 inflorescences) each time, while grinding in liquid nitrogen can be applied on a larger amount of tissue (up to 5g). As a blank control, the wildtype nuclei extraction by chopping method showed zero soring event for YFP-positive nuclei (Figure 3.1 E). However, we observed plenty of YFP positive signals (P9 gate) in the wildtype sample when adopting the grinding method (Figure 3.1 F). A possible explanation is that the grinding process is a crude way to disrupt the cellular structure, might releasing components or metabolites that emit at GFP wavelength, which is known as auto-fluorescence (Figure 3.1 F). Thus, despite taking a longer time to prepare nuclei suspension, we chose the chopping method in the following FACS experiment to avoid false-positive events.

We then tested formaldehyde-treated nuclei in FACS sorting, because crosslink is an essential treatment for the downstream ChIP experiment (Data was shown in section 3.1.2). Comparing to the untreated nuclei, the formaldehyde-treated nuclei generated a diffused scatterplot that showed smear signals between different DNA ploidies (Figure 3.1 G and H). This is caused by the formaldehyde, which affects the binding of dye to DNA and results in an inaccurate measurement of DNA content. As shown in Figure 3.1 (H), without a proper illustration of DNA ploidy in the side scatterplot, it is not straightforward to define the true nuclei population (P1 gate in figure 3.1 G and H) among all particles in the sorting suspension.

Furthermore, we analyzed the sorted nuclei (in P9 gate) (Figure 3.1 I and J) under the fluorescence microscope and confirmed that those are indeed YFP positive nuclei (Figure 3.1 C and D). However, we can only recover less than a hundred of YFP positive nuclei from around 700 inflorescences (each inflorescence has approximately 5-6 CLV3-expressing domains). Altogether, besides inconvenience caused by full dependence on FACS facility, technical problems to set proper FACS parameters and low efficiency in FACS nuclei isolation prompted us to use another method for cell-type specific nuclei isolation.

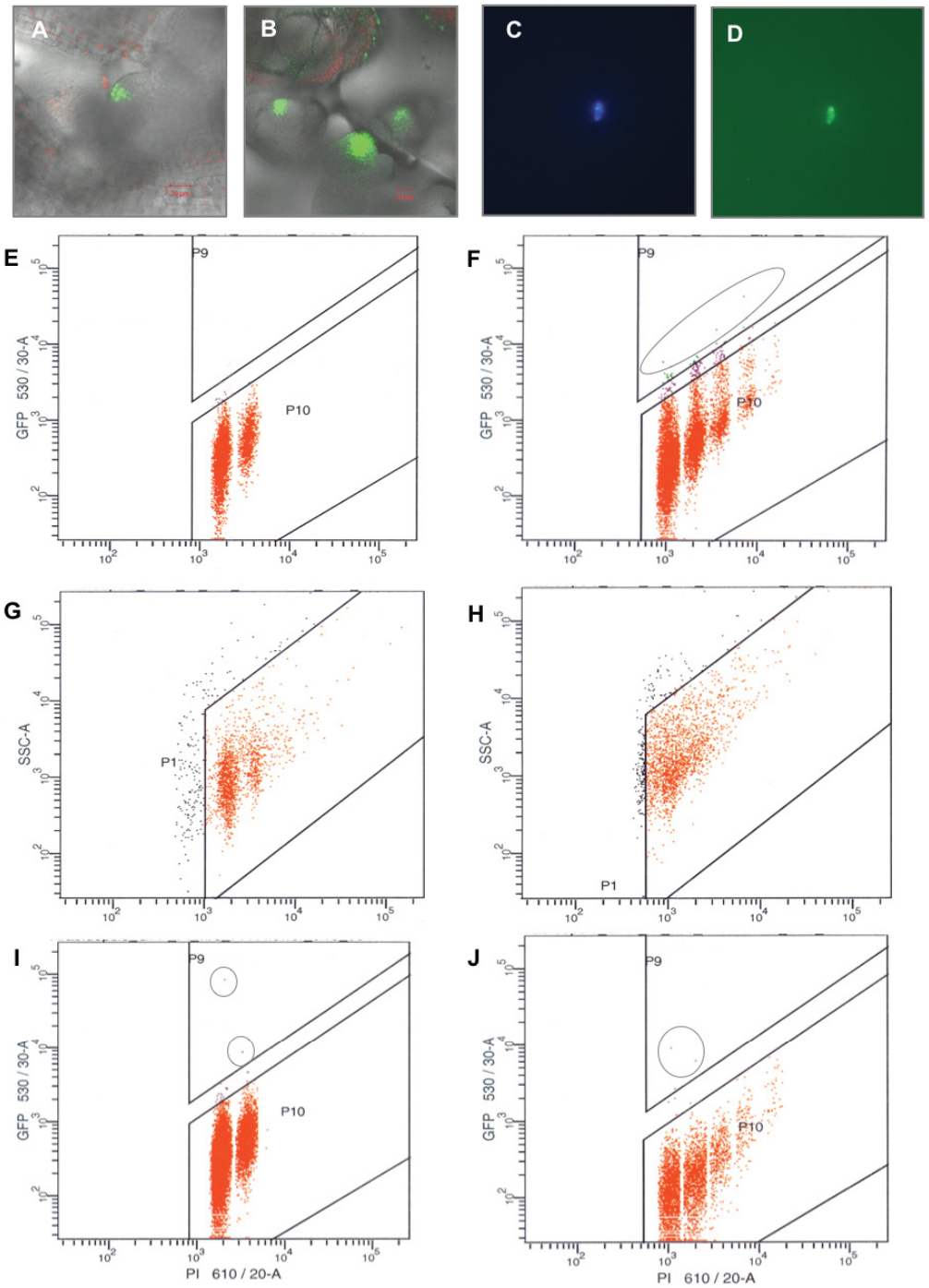


Figure 3.1. (A and B) Expression pattern of *CLV3*::*H2B-YFP* (green) reporter line in shoot apical meristem, (A) Side view, (B) Top view. (C and D) FACS-sorted YFP positive nuclei under fluorescence microscope, using the DAPI filter (blue) and the GFP filter (green) respectively. (E and F) Bivariate analysis of PI (610/20-A filter) fluorescence intensity versus GFP (530/30-A filter) fluorescence intensity on nuclei extraction of WT (Col-0) inflorescence; P10: the nuclei population that sorted as YFP-negative (red); P9: the nuclei population that sorted as YFP-positive (green). (E) nuclei extraction was prepared by the chopping method. (F) nuclei extraction was prepared by the grinding method; a circle marked the sorted YFP-positive particles. (G and H) PI (610/20-A filter) fluorescence intensity

(x axis) and side scatter (y axis) on nuclei extraction of *CLV3*::*H2B-YFP* inflorescence; P1: overall sorted nuclei (red), unsorted particles were in black. (G) inflorescences were not treated by crosslinker (formaldehyde). (H) inflorescences were treated with formaldehyde. (I and J) Bivariate analysis of PI (610/20-A filter) fluorescence intensity versus YFP (530/30-A filter) fluorescence intensity on nuclei extraction from non-crosslinked *CLV3*::*H2B-YFP* inflorescence; P10: the nuclei population that sorted as YFP-negative (red); P9: the nuclei population that sorted as YFP-positive (green).

3.1.2 Establishing the INTACT procedure

To establish the INTACT protocol in *Arabidopsis*, we generated several INTACT lines (Figure 3.2), including CLV3::NTF which labels stem cells in the shoot apical meristem (Figure 3.2 A, B and C), FIL::NTF line which labels abaxial cells of lateral organ (Figure 3.2 D and E), UFO::NTF line which labels the SAM excluding central meristem (Figure 3.2 F), GATA23::NTF, which labels the lateral root primordium (Figure 3.2 G), and BBM::NTF labels the QC of root meristem, although did not match the expression domain as described in previous study (Boutilier et al., 2002) (Figure 3.2 H). In this study, in order to analyze dynamic changes between stem cells and differentiated cells, we focused on domains marked by CLV3::NTF and FIL::NTF.

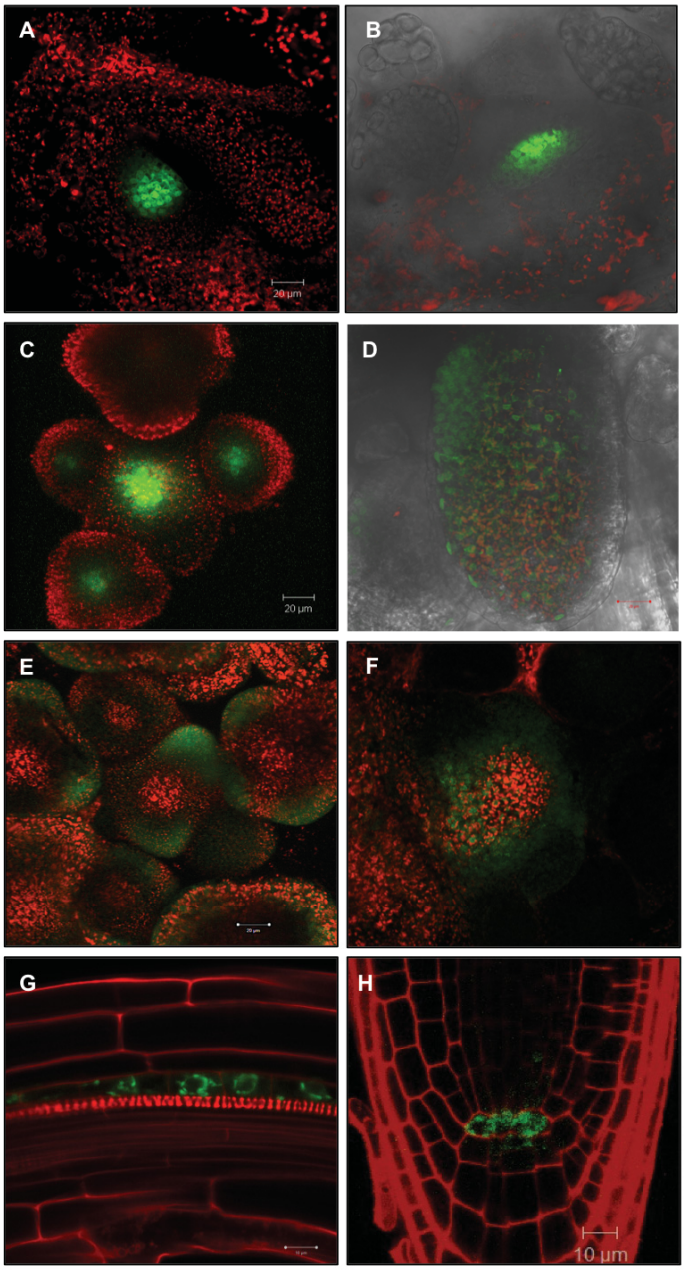


Figure 3.2: Expression pattern of INTACT lines. (A), (C), and (E-H) merged confocal images of 488nm (green) and 561nm (red) channels; (B and D) merged confocal images of brightfield (gray), 488nm (green), and 561nm (red) channels. (A and B) Expression pattern of CLV3::NTF line in vegetative SAM; (A) top view; (B) side view. (C) Expression pattern of CLV3::NTF line in reproductive SAM. (D and E) Expression pattern of FIL::NTF; (D) leaf primordium; (E) reproductive SAM. (F) Expression pattern of UFO::NTF line in the reproductive SAM, the GFP signal was depleted in the central zone of the meristem. (G) Expression pattern of GATA23::NTF line in lateral root primordium at initiation stage. (H) Expression pattern of BBM::NTF line marked nuclei envelope of QC in the root. Note: (A-F) the red signals are chlorophylls which emit auto-fluorescence; (G and H) the cell wall was stained by PI.

In the standard INTACT procedure (Deal and Henikoff, 2011), one can estimate the purity of the isolation by counting the nuclei with or without streptavidin beads under the fluorescence microscope (Figure 3.4 D and E). However, it does not consider the contamination introduced by unspecific binding of nuclei on streptavidin beads as well as the binding of random DNA or chromatin from disassociated nuclei. In fact, when we incubated streptavidin beads with pure DNA or chromatin, a trace of DNA or chromatin could still be found on streptavidin beads after standard washing steps in INTACT (Figure 3.3). Here, the DNA attached to streptavidin beads was evaluated by the relative quantification of *ACT7* between the starting DNA/chromatin solution and beads-attached DNA/chromatin.

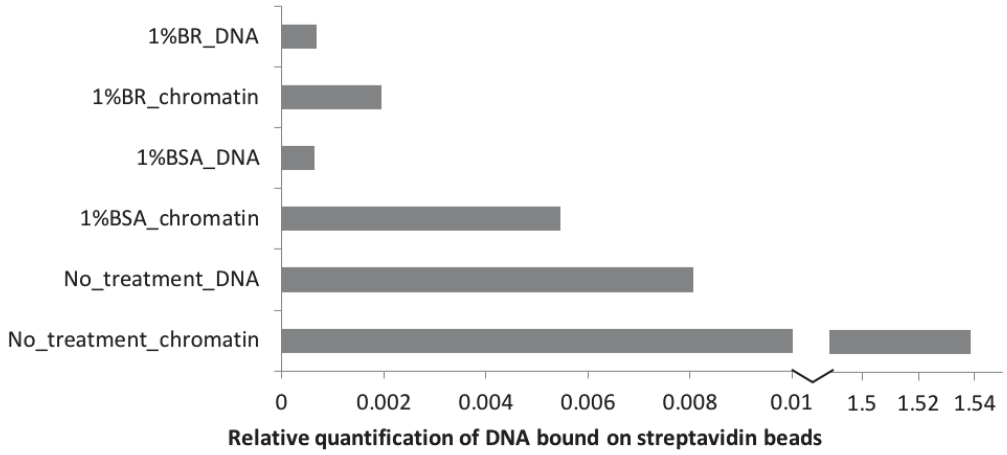


Figure 3.3 Quantification of random binding of DNA/chromatin on streptavidin beads. Y axis lists each treatment: 1%BR_DNA is the evaluation of random binding of DNA on 1% BR treated streptavidin beads; 1%BR_chromatin is the evaluation of random binding of chromatin on 1% BR treated streptavidin beads; 1%BSA_DNA is the evaluation of random binding of DNA on 1% BSA treated streptavidin beads; 1%BSA_chromatin is the evaluation of random binding of chromatin on 1% BSA treated streptavidin beads; No_treatment_DNA is the evaluation of random binding of DNA on streptavidin beads without any treatment; No_treatment_chromatin is the evaluation of random binding of chromatin on streptavidin beads without any treatment. X axis is the relative DNA quantification by qPCR analysis on *ACT7,* using the equation $((2 \wedge (Ct_{(starting\_DNA)}-Ct_{(Beads\_DNA)}))*100000)$.

In order to have a better estimation of contamination in INTACT, we adopted the second quantification method established by Moreno-Romero et al. (Moreno-romero et al., 2016). In their approach, taking advantage of endosperm (target cell population) which has a fixed 2:1 genome ratio of maternal to paternal, they crossed the INTACT line (Col) with *Ler* and

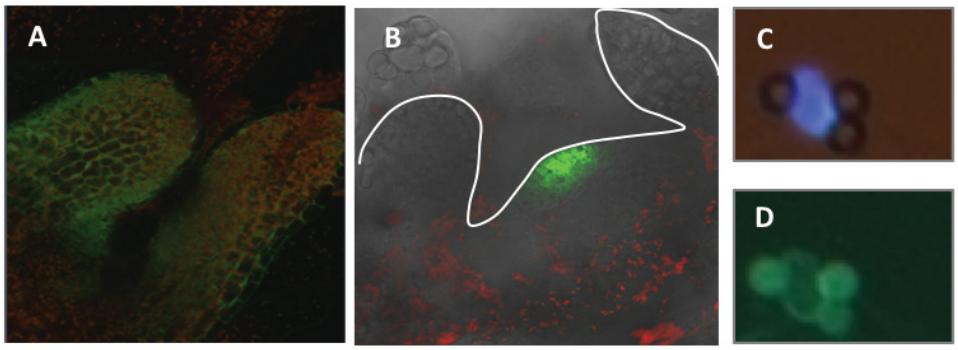
evaluated the ratio of *Ler* DNA in INTACT isolated endosperm. As a result, the percentage of contamination was calculated by using the reference of 2:1 or 1:2 of Col/*Ler*, depending on the crossing direction (Moreno-romero et al., 2016). Here, we mixed *Ler*-0 seedlings with INTACT lines which were in Col background before isolation of nuclei. By quantifying the percentage of *Ler* DNA in INTACT-isolated nuclei and the starting material, we estimated the percentage of contamination (Equation 1, section 2.3.1). Details of the calculation were explained in the materials and methods. Accordingly, we tested two contamination quantification methods on CLV3::NTF line (Figure 3.4 B). In general, when applying regular INTACT protocol on CLV3::NTF line (5g seedling), we could isolate around 80000 positive nuclei with 89-92% purity based on the counting strategy. However, when we applied the second contamination estimation strategy by mixing INTACT lines with *Ler* plants, we detected 51.67%-60.53% contamination (Figure 3.4 E). The large gap between two contamination quantifications further suggested that the counting method probably neglected unspecific bindings of nuclei or DNA on streptavidin beads. In the following part of the thesis, the percentage of contamination (% contamination) always refers to the second quantification method.

In order to avoid unspecific binding of nuclei and DNA on streptavidin beads, we tested commonly used blocking substrates to treat the streptavidin beads (Figure 3.3), including BSA (Sigma) and blocking reagent (BR) (Roche). When the streptavidin beads were treated by a blocking substrate, we observed a significant reduction of DNA or chromatin on streptavidin beads (Figure 3.3), whereas 1% BR showed better blocking performance on the random binding of chromatin (Figure 3.3). We then tested the 1% BR blocking treatment in the complete INTACT procedure on CLV3::NTF lines. We indeed observed around 10% reduction of contamination (Figure 3.4 E), although less striking differences when comparing to the in vitro experiment (Figure 3.3). On the other hand, the levels of contamination are largely dependent on the INTACT lines. We observed that the INTACT purified FIL labeled nuclei had only around 1.67%-7.09% contamination (Figure 3.4 F). One possible explanation of the different contamination rates between two INTACT lines is that FIL::NTF marks a larger cell population that already has a higher ratio of NTF-positive cells before the INTACT procedure (Figure 3.4 A and B).

In addition, it is worth to note that independent of the blocking strategy, a higher amount of DNA contamination was found if the streptavidin beads were incubated with chromatin (Figure 3.3). Meanwhile, the crosslinked samples had higher %contamination in the INTACT isolation than the untreated material (Figure 3.4 F). In the crosslinking procedure, the plant material was treated with 1% paraformaldehyde before INTACT procedure. It is known that the PFA treatment add the covalent bond between DNA and proteins (DNA-CH_2-protein), but the impact on nuclei envelopes is not clear by far.

Although the crosslink procedure generated higher contamination in INTACT, this step was crucial for the downstream ChIP experiment. When we compared ChIP performance upon different duration of crosslink, the 8min, 15min, and 30min crosslink showed differential enrichment of H3K27me3 between background (*ACT7* locus) and target region (*CLV3* locus). However, such enrichment was under detection by qPCR when the ChIP was done without crosslink or 4min crosslink (Figure 3.4 G). Therefore, we applied crosslink on INTACT material even it brought a higher contamination rate.

Altogether, counting nuclei cannot genuinely reflect the contamination rate in INTACT. Instead, we used a different *Arabidopsis* ecotype (*Ler*) as the background reference to measure the contamination in INTACT. In addition, contamination in INTACT can be reduced by blocking streptavidin beads with 1% BR. Most importantly, the contamination in INTACT is line dependent. The larger cell population of the target cell-types, the lower contamination rate will yield.



E CLV3::NTF line

	Streptavidin beads	Streptavidin beads & 1%BR
%Contamination	51.67%-60.53%	37.39%-54.75%

F FIL::NTF line

	No-crosslink	With-crosslinked
%Contamination	1.67%-7.09%	6.81%-14.36%

G Testing different crosslink time for ChIP on H3K27me3

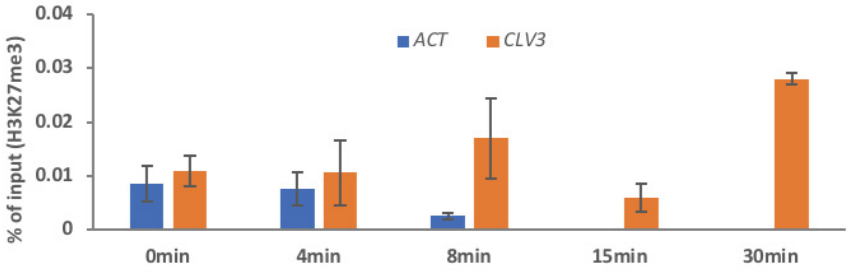


Figure 3.4: (A) Merged confocal images of 5LDs FIL::NTF line, green is the image of the 488nm channel, and red is the image under the 561nm channel. (B) Merged confocal images of 5LDs CLV3::NTF line, gray is the image of brightfield, green is the image of 488nm channel, and red is the image of 561nm channel. (C and D) Images of INTACT-isolated nuclei under the fluorescence microscope. (C) DAPI channel, Nucleus is in blue, and streptavidin beads are in brown. (D) GFP channel, besides the nucleus has green fluorescence, streptavidin beads also emit auto-fluorescence. (E) Quantification of % contamination on INTACT-isolated nuclei by un-treated or 1%BR treated streptavidin beads, using CLV3::NTF line. (F) % contamination of INTACT-isolated nuclei from crosslinked or non-crosslinked FIL::NTF line; In (E) and (F), the % contamination was calculated by using the Equation 1 in section 2.3.1 (Material and method). (G) Evaluation of H3K27me3 ChIP efficiency upon different crosslink duration, by ChIP-qPCR, Y axis is the relative quantification of H3K27me3 enrichment on *ACT7* and *CLV3*, by quantifying the immunoprecipitated DNA versus input DNA.

We then applied H3K4me3 and H3K27me3 ChIP on INTACT isolated nuclei, including FIL-labeled cells (FIL), cells which were not labeled by FIL (UN), CLV3-labeled cells (SC), cells which were not labeled by CLV3 (Negative_SC). qPCR was conducted on several commonly analyzed loci to evaluate the performance of ChIP, including *PIN7* which is highly expressed in seedlings (Žádníková et al., 2010; Sassi et al., 2012) and is used as positive control for H3K4me3, *FUS3* which is repressed during germination and is known as positive control for H3K27me3, *FIL* and *CLV3* are in accordance to INTACT isolated cell population (Figure 3.5 A and B). In general, we detected differential enrichment of H3K4me3 and H3K27me3 between *PIN7* and *FUS3* among all cell types (Figure 3.5 A and B), in which the enrichment is comparable to previous epi-genome studies in seedlings (Zhang et al., 2009, 2007b) (Figure 3.5 C). This indicated that ChIP experiment on INTACT-isolated nuclei were successful. In addition, comparing to other three tissues, FIL-labeled cells (FIL) showed increased H3K4me3 on the first exon of *FIL* and reduced H3K27me3 on both exon-1 and exon-7 of *FIL* (Figure 3.5 A and B). This observation fits in the expectation that FIL-expressing cells have increased active mark (H3K4me3) and decreased repressive mark (H3K27me3) on *FIL*. On the other hand, CLV3-labeled cells showed increased H3K4me3 at TSS of *CLV3* comparing to non-CLV3 labeled cells and FIL-labeled cells. But one should be cautious about this conclusion, since the H3K4me3 enrichment on *CLV3-ATG* was less than *FUS3* (negative control for H3K4me3), suggesting that respective enrichment maybe background noise (Figure 3.5 A). Therefore, it is not conclusive to link that increased enrichment of H3K4me3 at *CLV3* in CLV3-labeled cells with INTACT isolated CLV3-expressing cells. Altogether, we successfully conducted ChIP on nuclei isolated from INTACT, which showed sufficient resolution to identify differential enrichment of H3K4me3 and H3K27me3 on analyzed loci. Nevertheless, since we have only analyzed six loci by qPCR, this does not reflect the full picture of H3K4me3 and H3K27me3 profiles among different cell types.

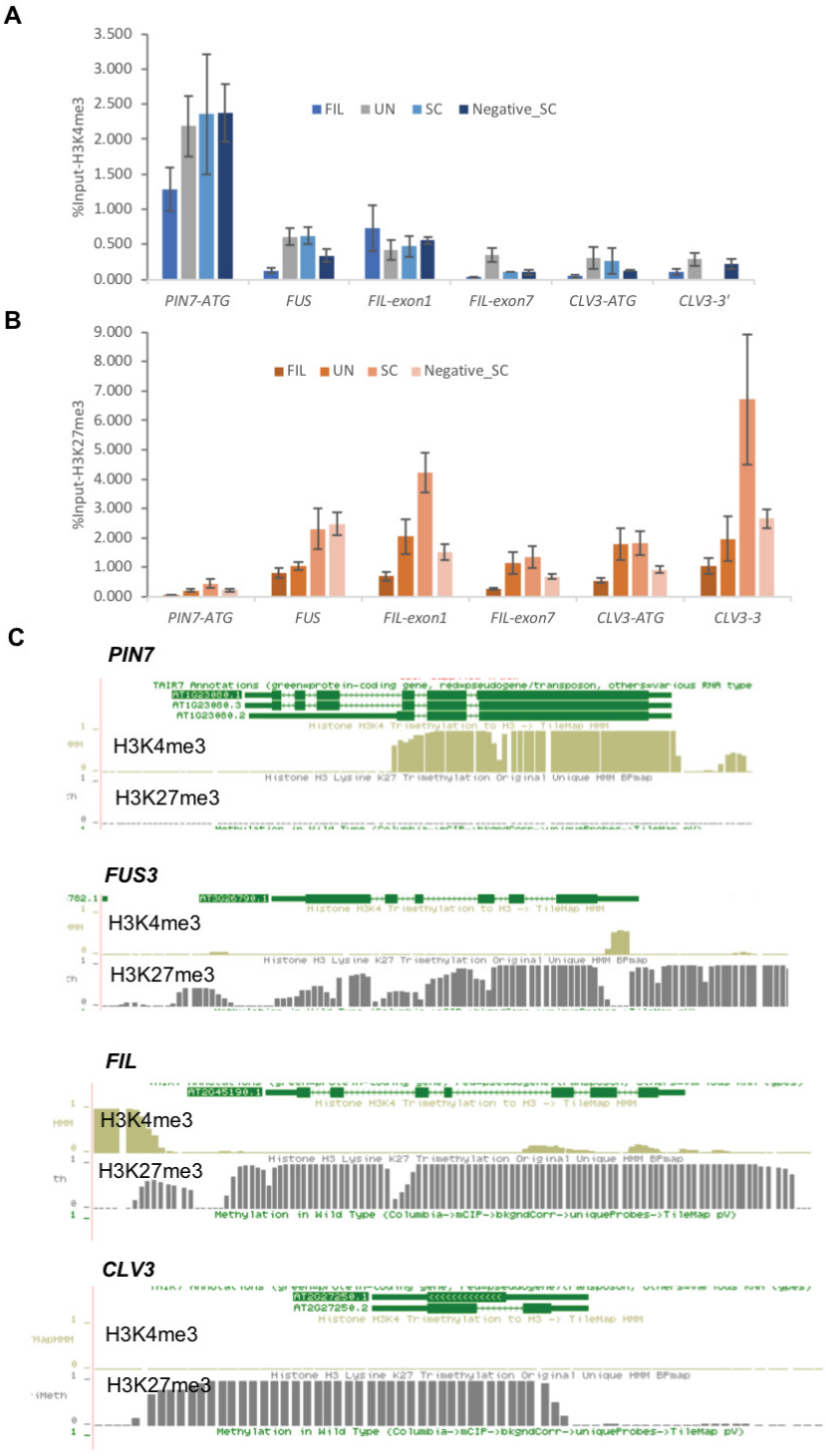


Figure 3.5. (A and B) ChIP-qPCR analysis of H3K4me3 and H3K27me3 on INTACT isolated FIL-labeled cells (FIL), CLV3-labeled cells (SC), non-FIL labeled cells (UN), non-CLV3 labeled cells (Negative_SC); Y axis is the relative quantification of H3K4me3 (A) and H3K27me3 (B) on *PIN7, FUS3, FIL,* and *CLV3*, by quantifying the immunoprecipitated DNA versus input DNA. (C) UCSC genome browser view of H3K4me3 (Zhang et al., 2009) and H3K27me3 (Zhang et al., 2007b) distribution on *PIN7*, *FUS3*, *FIL* and *CLV3*, tracks in yellow are H3K4me3, and tracks in gray are H3K27me3.

3.2. Profiles in CLV3-labeled cells do not reflect previous knowledge in meristematic tissues

We then conducted high-throughput sequencing on RNA and ChIP against H3K4me3 and H3K27me3 in CLV3-labeled cells (SC), FIL-labeled cells (FIL) and non-FIL labeled cells (UN) (Table 2.7 and 2.8). In the following sections, I will present the results from the genome-wide comparison of gene expression, H3K4me3, and H3K27me3 among CLV3-labeled cells (SC), FIL-labeled cells (FIL) and non-FIL labeled cells (UN).

Because of the contamination problem in CLV3-labeled cells (SC), we firstly examined transcriptome, H3K4me3, and H3K27me3 profiles on some key meristematic regulatory genes before a detailed genome-wide comparison. On the transcriptome data, we calculated covariance among expression profiles on genes from stem cell maintenance (GO:0019827) and abaxial specification (GO:0009798). As a reference, we also included the expression data from INTACT isolated shoot apical meristem(SAM) (You et al., 2017) in the comparison, which expect to show higher similarity to CLV3-labeled cells (SC) than FIL-labeled cells (FIL) or non-FIL labeled cells (UN). The PCA analysis on genes from stem cell maintenance (GO:0019827) showed that SAM (You et al., 2017) did not cluster with CLV3-labeled cells (SC) on PC2 which explained the largest variance of genes' expression between CLV3-labeled cells (SC) and FIL-labeled cells (FIL). On genes from abaxial specification (GO:0009798), PC2 separated FIL-labeled cells from CLV3-labeled cells (SC) and non-FIL labeled cells (UN). However, SAM (You et al., 2017) still did not cluster with CLV3-labeled cells (SC) on PC2, even though it showed smaller variance for CLV3-labeled cells (SC) comparing to FIL-labeled cells (FIL) and non-FIL labeled cells (UN). In addition, using FIL-labeled cells as the reference, we found that DEGs (differential expressed genes) of CLV3-labeled cells showed similar expression dynamic with non-FIL labeled cells (UN) (Figure 3.6 D, E, F, G). More than 60% of down-regulated DEGs in CLV3-labeled cells (SC) were also down-regulated in non-FIL labeled cells (UN) (Figure 3.6 D and E). Similar trends can also be found in up-regulated DEGs in CLV3-labeled cells (SC) (Figure 3.6 F and G). Furthermore, key meristematic regulatory genes, such as *CLV3*, *STM,* and *WUS* showed similar enrichment of H3K27me3 in all three cell types, whereas H3K4me3 is not enriched (Figure 3.6 G). In

contrast, H3K4me3 enrichment at *FIL* is different between FIL-labeled and non-FIL labeled cells (Figure 3.6 G).

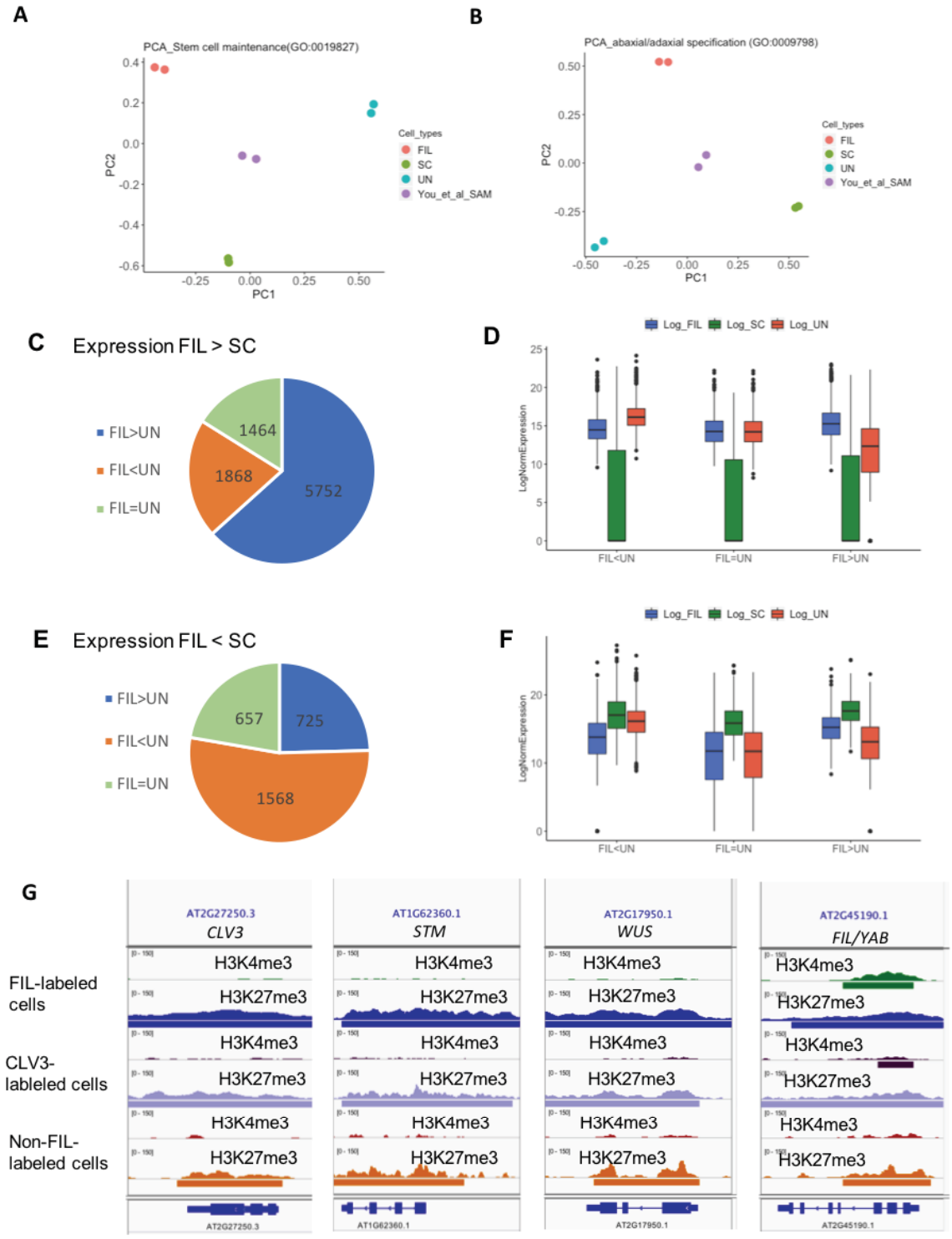


Figure 3.6. (A) PCA analysis of expression covariance of stem cell maintenance genes (GO:0019827) among FIL-labeled cells (FIL), CLV3-labeled cells (SC), non-FIL-labeled cells (UN) and shoot apical meristem from literature (You et al.,2017). (B) PCA analysis of expression covariance of genes

involved in abaxial cell fate determination (GO:0009798) among FIL-labeled cells (FIL), CLV3-labeled cells (SC), non-FIL-labeled cells (UN) and shoot apical meristem (You et al.,2017). (C-F) Analyses of DEGs between FIL-labeled cells (FIL) and CLV3-labeled cells (SC) versus DEGs between FIL-labeled cells (FIL) and non-FIL labeled cells (UN); FIL>UN represents gene expression in FIL is increased comparing to UN; FIL<UN represents gene expression in FIL is decreased comparing to UN; and FIL=UN represents expression in FIL is not significant different from UN; FIL>SC represents gene expression SC is decreased comparing to FIL; FIL<SC represents gene expression in SC is increased comparing to FIL. (C) Pie chart of the number of FIL<UN, FIL>UN and FIL=UN among down-regulated genes in SC (FIL>SC). (D) Boxplot of expression profile of genes in accordance to figure (C), among FIL-labeled cells (FIL), CLV3-laleled cells (SC), and non-FIL labeled cells (UN). (E) Pie chart of the number of FIL<UN, FIL>UN and FIL=UN among up-regulated genes in SC (FIL<SC). (F) Boxplot of expression profile of genes in accordance to figure (E), among FIL-labeled cells (FIL), CLV3-labeled cells (SC) and non-FIL labeled cells (UN). (G) iGV browser view of H3K4me3 and H3K27me3 enrichment on *CLV3, STM, WUS, and FIL* in FIL-labeled cells (FIL), CLV3-labeled cells (SC), non-FIL labeled cells (UN); Each gene model is presented in the bottom track.

In addition, following RNA-seq analysis pipeline, we found around 50% of genes were expressed in FIL-labeled cells (FIL) and non-FIL labeled cells (UN), resembling other transcriptome profiles of specific cell types from previous studies, either using FACS (Yadav et al., 2009) or INTACT (You et al., 2017) (Figure 3.6 A). However, less than 25% of genes were found expressed in CLV3-labeled cells (SC) (Figure 3.6 A). This introduced potential bias in reads count normalization in differential expression analysis among tissues (Figure 3.6 C). In the DEseq pipeline which we adopted to identify DEGs (differentially expressed gene), the size factor (normalization factor of reads count) is calculated by assuming the median ratio of reads count would fall on the non-differential expressed genes (Love et al., 2014). However, unlike FIL-labeled cells (FIL) and non-FIL labeled cells (UN) which showed even distribution of up and down-regulated genes in MA plot, we observed more genes were down-regulated in CLV3-labeled cells (SC) comparing to FIL-labeled cells (FIL) or non-FIL labeled cells (UN) (Figure 3.6 C). Together, those preliminary analyses among CLV3-labeled cells and FIL-labeled cells suggested that profiles of CLV3-labeled cells provided little and inconclusive information. Therefore, we had to give up the CLV3-labeled cells and focused on the profiles between FIL-labeled cells (FIL) and non-FIL labeled cells (UN) to explore the dynamic regulations of expression and H3K4me3/H3K27me3 enrichment.

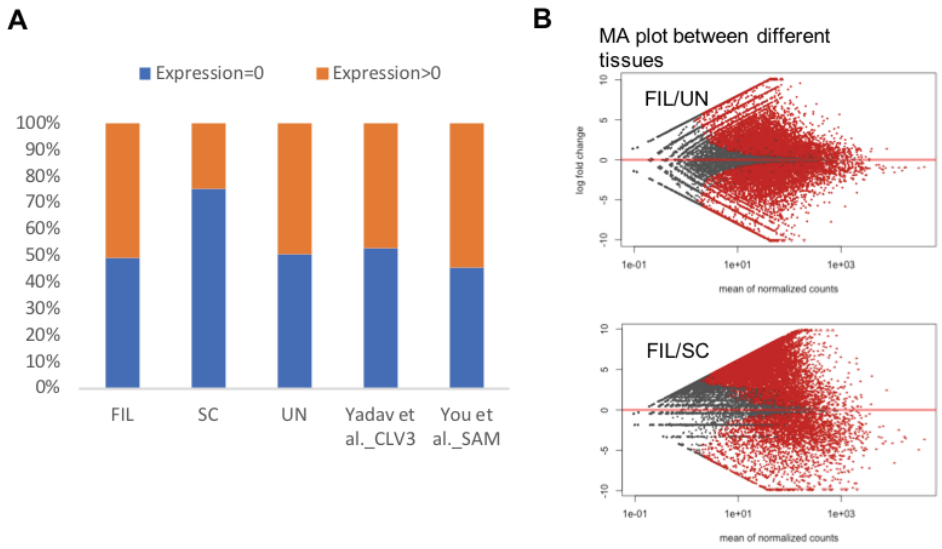


Figure 3.7. (A) Percentage of expressed and un-expressed genes (reads count=0) among transcriptome profile of FIL-labeled cell (FIL), CLV3 labeled cells (SC), non-FIL labeled cells (UN), FACS-sorted CLV3 labeled cells (Yadav et al., 2009) and shoot apical meristem (You et al.,2017). (B) MA plot of the expression fold change of FIL versus UN, FIL versus SC, calculated by DEseq2 in R; X axis is mean normalized reads count; Y axis log2 fold change; Red dots represent events with significant fold change.

3.3 Analysis of H3K4me3 and H3K27me3 profile in FIL-labeled cells (FIL) and non-FIL labeled cells (UN)

3.3.1 Overview of H3K4me3 and H3K27me3 profile in FIL-labeled cells (FIL) and non-FIL labeled cells (UN)

Overall, we observed more than 15294 H3K4me3 domains and more than 7552 H3K27me3 domains in FIL-labeled cells (FIL) and 13859 H3K4me3 domains and 6525 H3K27me3 domains in non-FIL labeled cells. By calculating K-means of signal intensity of H3K4me3 and H3K27me3 on *Arabidopsis* genes (TAIR10), we generated 8 clusters which illustrated various distribution of H3K4me3 and H3K27me3. In general, genes within cluster1-4 were targeted by H3K4me3, without H3K27me3 enrichment (Figure 3.8 A and B). Genes within cluster 5-7 were targeted by H3K27me3, without H3K4me3 enrichment (Figure 3.8 A and B). And genes within cluster 8 were enriched in neither modification (Figure 3.8 A and B). Similar to previous studies, H3K4me3 enrichment was associated with gene activation, and H3K27me3 was correlated with gene repression (Figure 3.8 C). Despite the difference in the precise distribution of H3K4me3 or H3K27me3 among clusters, genes from all four clusters (1-4) that enriched in H3K4me3 had shown a similar range of high-level expression, while genes from cluster 5-7 were mostly expressed at a low level or not expressed (Figure 3.8 C).

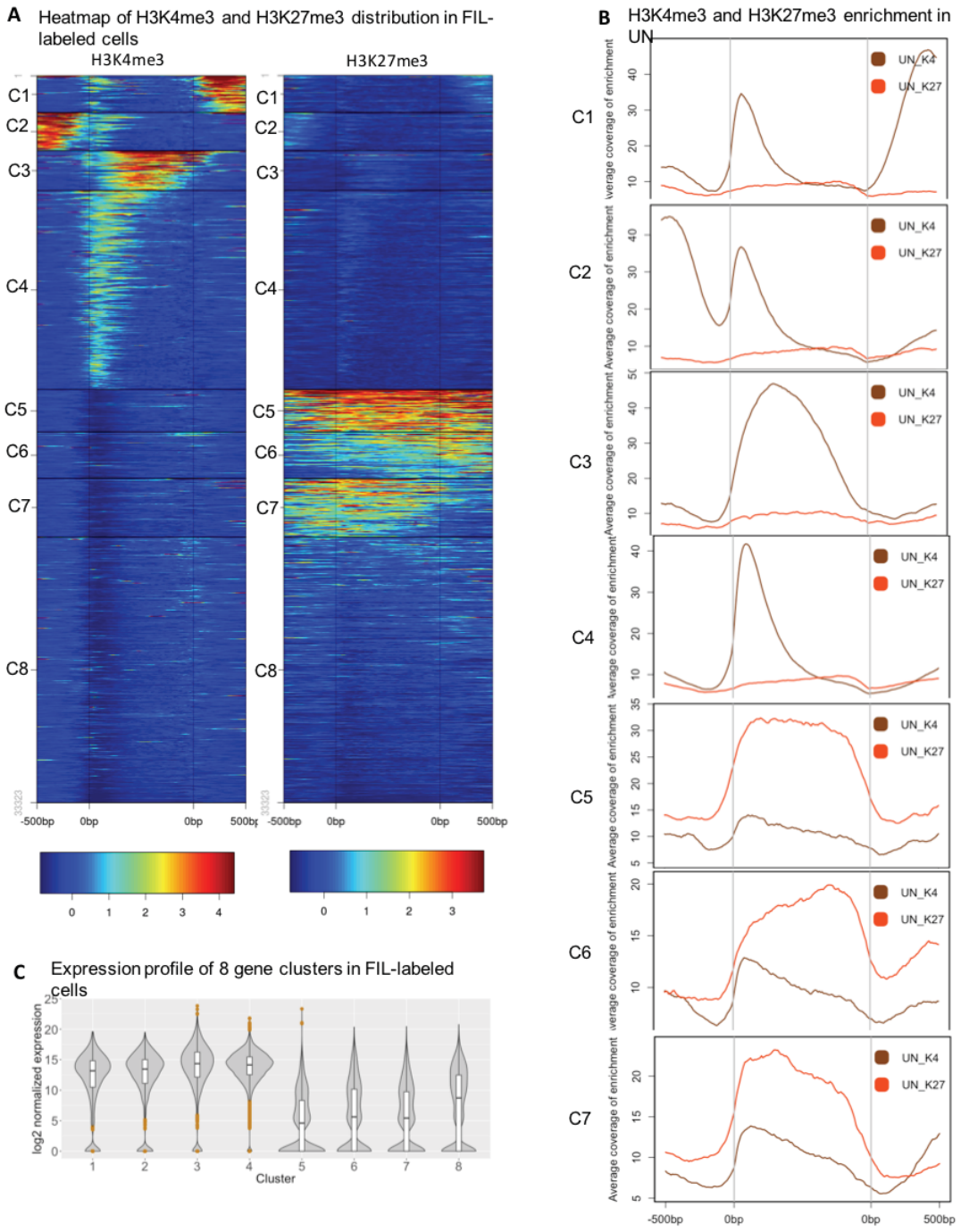


Figure 3.8. General description of H3K4me3 and H3K27me3 in FIL-labeled cells (FIL) and non-FIL labeled cells (UN). (A) Heat map of H3K4me3 and H3K27me3 distribution on coding genes of *Arabidopsis* in FIL-labeled cells, 8 clusters had been designed and calculated by K-mean of reads count (ChIP-seq of H3K4me3 and H3K27me3) on each gene. B. Average coverage of H3K4me3 and H3K27me3 on each gene cluster in non-FIL labeled cells, each gene cluster correspond to Figure (A), "UN_K4" in brown is average coverage of H3K4me3 in non-FIL labeled cells, "UN_K27" in orange is average coverage of H3K27me3 in non-FIL labeled cells. C. Violin box plot of gene expression on 8 clusters, each gene clusters correspond to Figure (A), Y axis is the log2 normalized reads count.

The associations between gene expression and H3K4me3/H3K27me3 were also reflected in the dynamic changes of H3K4me3 and H3K27me3 between FIL-labeled (FIL) and non-FIL labeled cells (UN). By the peaking-calling algorism SICER, we identified 1238 H3K4me3 hyper- differentially methylated regions (Hyper-DMRs), 383 H3K4me3 hypo- differentially methylated regions (Hypo-DMRs), 2529 H3K27me3 hyper-DMRs and 1589 hypo-DMRs in FIL-labeled cells (FIL) compared to non-FIL labeled cells (UN). Among genes that were hypermethylated by H3K4me3, there was a higher percentage of genes showed up-regulation (around 40%) than down-regulation (around 17%) in FIL-labeled cells (FIL) (Figure 3.9 A). In contrast, among the genes that were hypomethylated by H3K27me3, there was a higher percentage of genes that were up-regulated than down-regulated in FIL-labeled cells (Figure 3.9 A). The proportion of up and down-regulated genes in the H3K4me3 hyper-DMRs was significantly different from the proportion of up/down-regulated genes in the H3K4me3 hypo-DMRs (Table 3.1). Similar statistical significance was also found between H3K27me3 hyper-DMRs and H3K27me3 hypo-DMRs (Table 3.1).

However, the associations between expression and those two histone modifications were not universal to all H3K4me3 or H3K27me3 targets. Nearly 50% of genes that were differentially methylated by H3K4me3 showed no significant change in expression between FIL-labeled and non-FIL labeled cells (Figure 3.9 A). This was especially obvious for H3K4me3 hypo-methylated genes (Figure 3.9 A). And more than 75% of genes within H3K27me3 DMRs were not differentially expressed (Figure 3.9 A). The association between expression and histone modification became even weaker, if comparing the quantitative changes of expression and H3K4me3 enrichment (R^2=0.06) (Figure 3.9 C). The fold change of transcription showed larger variation, while H3K4me3 fold change was in a narrow range: 2-5 folds change (Figure 3.9 D). It needs to be noted that we calculated the fold change of H3K4me3 on in silico DMRs by assuming all DMRs start site were at the TSS of respective genes (Figure 3.9 B). In reality, the H3K4me3 is enriched at TSS, but the precise location of H3K4me3 domains varies among different experiments. Thus, it makes the quantitative evaluation between different experiments impossible if directly using the H3K4me3 domains (Figure 3.9 B).

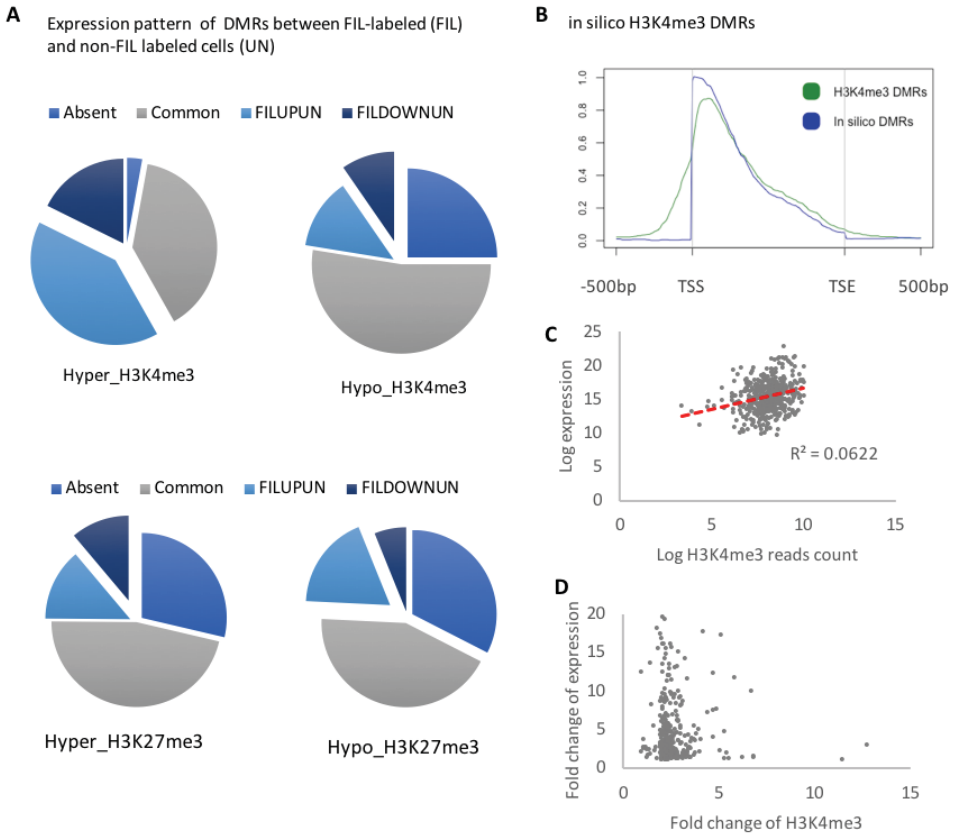


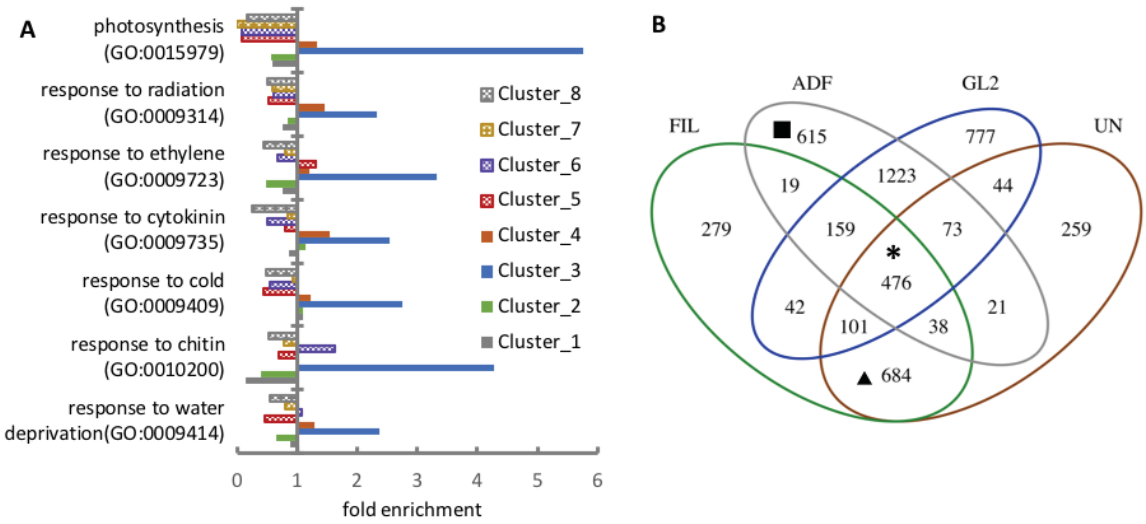
Figure 3.9. Analysis of correlation between dynamics of H3K4me3 or H3K27me3 and expression changes. (A) Pie chart of the number of genes in each group of DEGs among DMRs, including H3K4me3 hypermethylated domains (Hyper_H3K4me3), H3K4me3 hypomethylated domains (Hypo_H3K4me3), H3K27me3 hypermethylated domains (Hyper_H3K27me3), and H3K27me3 hypomethylated domains (Hypo_H3K27me3). The legend in figure (A) listed four DEG groups: Absent (expression in FIL=UN=0), Common (expression in FIL≈UN>0), FILUPUN (expression in FIL>UN), and FILDOWNUN (expression in FIL<UN). (B) Comparison of average localization of in-silico generated DMRs (blue) versus SICER identified DMRs (green); In X axis, -500bp is 500 upstream of TSS and 500bp is 500bp downstream of TES. (C) Dot plot of gene expression and H3K4me3 enrichment on respective in silico H3K4me3 DMRs, correlation co-efficiency R^2 were calculated by the R package. (D) Dot plot of the fold changes of gene expression versus changes of H3K4me3 enrichment of in silico DMRs.

Table 3.1. Fisher exact test on proportion of up-regulated and down-regulated genes among DMRs.

	Hyper_H3K4me3	**Hyper_H3K27me3**	**Hypo_H3K4me3**
Hypo_H3K4me3	3.36E-02*	4.18E-01	
Hypo_H3K27me3	1.02E-01	2.36E-05**	1.02E-02**
Hyper_H3K27me3	7.64E-08**		

3.3.2 Analysis of cluster3-type (broad distribution) of H3K4me3

In general, the majority of H3K4me3 domains were located at the TSS (Figure 3.8 A), which were comparable to the H3K4me3 profile in seedlings (Zhang et al., 2009). Nevertheless, by the K-means clustering method, we identified 1798 genes that were covered by the broad H3K4me3 domains in both FIL-labeled and non-FIL labeled cells (Figure 3.8 A and B). Those genes were grouped in cluster3 (Figure 3.8 A). Although overall genes from cluster3 had similar range of expression as clusters 1, 2, and 4 (Figure 3.8 A), we found photosynthesis and stress responsive genes were significantly enriched in cluster3 (Figure 3.10 A). A study in human neuron cells has found that the broad H3K4me3 domains might be associated with respective cell identity genes (Benayoun et al., 2014). To analyze if this hypothesis is also valid in *Arabidopsis*, we cross-compared broad H3K4me3 domains in distinct tissues: aerial tissues (FIL-labeled cells and non-FIL labeled cells) and root tissues (GL2-labeled and ADF8-labeled cells) (Deal and Henikoff, 2010). Considering photosynthesis is one of the most conserved functions in leaves, we would expect to observe genes which involved in root development or certain root-related functions (such as ion transportation, nutrient transportation) were targeted by broad H3K4me3, if the broad domain associated with cellular identity in plants. Between FIL-labeled and non-FIL labeled cells, broad H3K4me3 domains were largely overlapped (Figure 3.10 B). Even those unique broad domains in FIL-labeled cells or non-FIL labeled cells have shown the broad distribution of H3K4me3 in both FIL-labeled and non-FIL labeled cells, with differences on signal intensity (Figure 3.11). This is likely due to the nature of K-mean clustering, which does not cluster data by a definite criterion instead by data similarity. In the comparison with root tissues, despite we found few hundreds of distinct genes that was targeted by broad H3K4me3 domain (Figure 3.10 B and 3.11), no GO term related to root cell identity was found in the genes that was uniquely targeted by broad H3K4me3 domains in root tissues (GL2-labeled cells or ADF8-labeled cells) (Figure 3.10 C). Among those exclusive broad domains in each cell type, we only identified significant GO terms in ADF8-labeled cells, including translation (GO:0006412) and peptide biosynthetic process (GO:0043043). Together, the broad H3K4me3 distribution probably does not have the proposed association with cell identity as in human neural cells.



C GO analysis

	GO	Fold enrichment	p-value
FIL∩UN ▲	photosynthesis (GO:0015979)	8.65	5.93E-27
	cellular response to abiotic stimulus (GO:0071214)	2.13E-04	1.57E-02
	response to ethylene (GO:0009723)	3.39	1.04E-06
	response to cold (GO:0009409)	2.7	1.57E-05
	response to water deprivation (GO:0009414)	2.76	3.77E-05
ADF8 ■	translation (GO:0006412)	2.68	6.76E-03
	peptide biosynthetic process (GO:0043043)	2.65	8.01E-03
FIL∩UN∩ADF8 ∩GL2 *	response to chitin (GO:0010200)	5.37	5.55E-06
	response to hormone (GO:0009725)	2	6.74E-06

Symbols (* ■ ▲) mark the corresponding gene groups in Figure B

Figure 3.10. Analysis on the cluster3 type of broad H3K4me3 domains. (A) Fold enrichment of the significantly enriched GO terms in cluster3 genes; in comparison, fold enrichment of above mentioned GO terms on other clusters of genes were calculated in parallel. (B) Venn diagram of genes enriched in cluster3 type broad H3K4me3 among FIL-labeled (FIL), non-FIL labeled (UN), ADF8-labeled (ADF) and GL2-labeled cells (GL2), * marks the overlapped genes in all four tissues, ■ represents unique broad H3K4me3 domains in ADF8-labeled cells (ADF), ▲ represents the overlapped broad H3K4me3 domains of FIL-labeled (FIL) and non-FIL labeled cells (UN). (C) GO analysis on genes in accordance of Figure (B), including the overlapped broad H3K4me3 domains of (FIL) and (UN) (FIL∩UN), unique broad H3K4me3 domains in (ADF) and overlapped domains in all four tissues. Note that other groups of broad domains from figure (B) did not have any significant GO term.

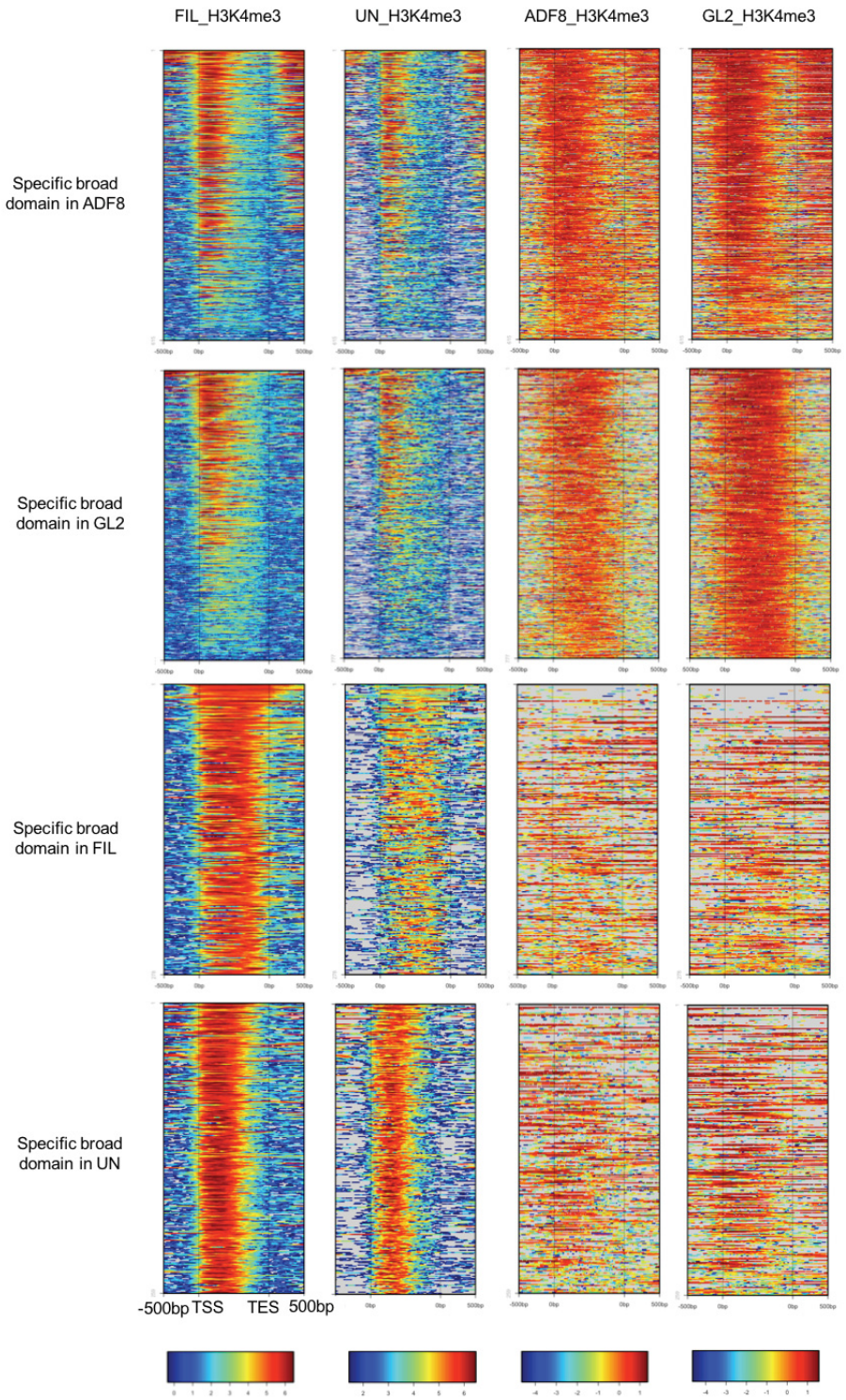


Figure 3.11. Heatmap of unique broad H3K4me3 domains of ADF8-labeled, GL2-labeled, FIL-labeled, and non-FIL labeled cells (UN) in all four cell types. From left to right, each row corresponds to H3K4me3 signals from FIL-labeled cells, non-FIL labeled cells, ADF8-labeled cells, and GL2-labeled cells. From top to bottom, each column corresponds to the unique broad H3K4me3 target genes that were used for plotting H3K4me3 signal, including ADF8-labeled, GL2-labeled, FIL-labeled, and non-FIL labeled cells (UN).

Besides photosynthesis, several GO terms for stress response were also found significantly enriched among genes with broad domains, such as response to water deprivation (GO:0009414), response to cold (GO:0009409) (Figure 3.10 A). The first question is whether the broad H3K4me3 domains have any biological function in stress response. Thus, we cross compared H3K4me3 broad domains in FIL-labeled cells to a published dehydration experiment by Dijk et al. (Dijk et al., 2010). Dijk et al.'s datasets include expression and H3K4me3 profiles of *Arabidopsis* seedlings under regular and drought conditions (Dijk et al., 2010). We analyzed the H3K4me3 distribution on the dehydration responsive genes that overlapped with cluster3 under drought and regular (water) conditions. In parallel, we also analyzed the dehydration responsive genes that were not overlapped with cluster3. As a result, we found that genes in cluster3 were widely covered by H3K4me3 under both drought and regular (water) conditions (Figure 3.12 A, B, D, and E). Although the down-regulated genes that were targeted by broad H3K4me3 domains showed reduced H3K4me3 intensity under drought stress, the breadth of H3K4me3 coverage was not affected (Figure 3.12 B and D). It worth to be noted that a group of up-regulated genes from other clusters showed expanded H3K4me3 coverage under drought stress, but the H3K4me3 peak still located at TSS (Figure 3.12 C). Thus, it is still not clear whether the broad H3K4me3 domain has any unique function in drought stress response. The cluster3 type of broad domains remain similar coverage under stress conditions. Meanwhile, some genes with expanded H3K4me3 domain are indeed up-regulated under dehydration stress, but those expanded domains do not resemble the cluster3 type domains.

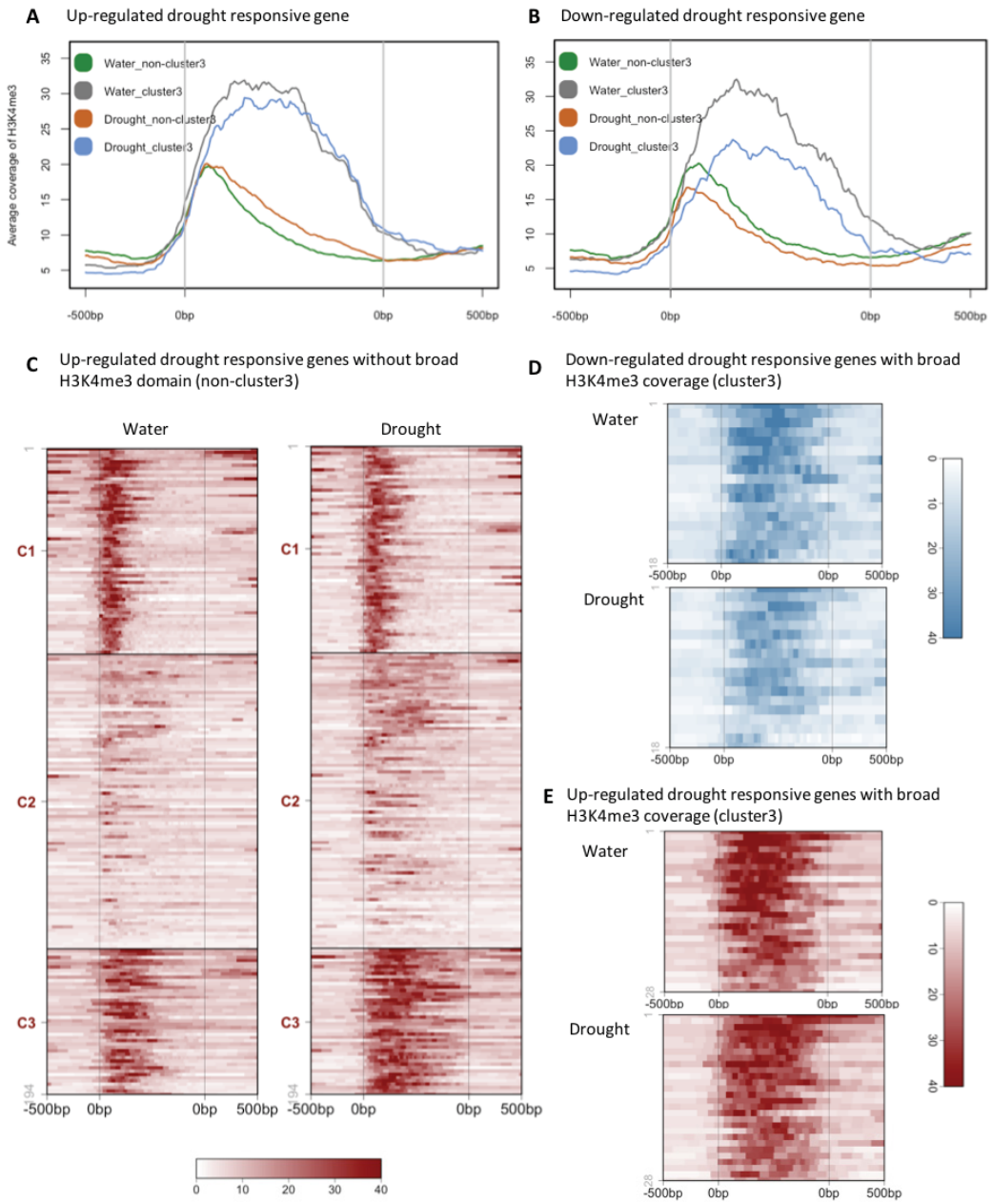


Figure 3.12. Analysis of broad H3K4me3 domains in dehydration response. (A) Average coverage of H3K4me3 on up-regulated drought responsive genes in cluster3 under drought condition (Blue) and regular condition (Gray), or up-regulated drought responsive genes in other clusters (non-cluster3) under drought condition (Orange) and regular condition (Green). (B) Average coverage of H3K4me3 on down-regulated drought responsive genes in cluster3 under drought condition (Blue) and regular condition (Gray), or down-regulated drought responsive genes in other clusters (non-cluster3) under drought condition (Orange) and regular condition (Green). (C) Heatmap of H3K4me3 distribution on up-regulated drought responsive genes in other clusters (non-cluster3) under drought condition (Drought) and regular conditions (Water). (D and E) Heatmap of H3K4me3 distribution on down-regulated (D) or up-regulated (E) drought responsive genes in cluster3 under drought condition (Drought) and regular condition (Water).

3.3.3 Analysis of H3K4me3 domains in the intergenic region

In general, H3K4me3 domains were located in genic regions, while H3K27me3 had relatively higher coverage on intergenic regions (Figure 3.13 A). Overall, around 2%-8% of reads from H3K4me3 were aligned to intergenic regions (excluding 1Kb-promoter) (Figure 3.13 A). To rule out the intergenic domains which were generated by the expansion of H3K4me3 domains from adjacent genes, we defined new H3K4me3 distribution groups in a hierarchy order of gene body, 1Kb-promoter, intergenic. Once a domain was classified in one group, it would be excluded from groups in the lower hierarchy. Based on the new classification criterion, we analyzed H3K4me3 distribution and respective expression profiles in FIL-labeled cells. The gene-body domains were enriched at TSS as expected (Figure 3.13 B). Meanwhile, those gene-body domains showed a small bump at the promoter region besides the peak at TSS. Between two peaks, it probably is the nucleosome-free region (NFR), in which the transcription factor would bind (Figure 3.13 B). In contrast, those 1KB-promoter located H3K4me3 domains showed a smooth coverage at the promoter and TSS, but the H3K4me3 signal intensity was reduced comparing to gene-body domains (Figure 3.13 B). For Intergenic located domains, an increased enrichment of H3K4me3 was shown at the upstream of 1Kb-promoter, while the H3K4me3 enrichment was reduced comparing to gene-body domains. Although few dozens of intergenic H3K4me3 domains were on unannotated gene/promoters, which were annotated as the pseudogene in the newer TAIR11 (Figure 3.14 D), the majority of intergenic domains were still annotated as an intergenic region by far. Among various H3K4me3 distribution, genes with H3K4me3 domain on gene body or 1Kb-promoter were highly expressed, whereas 1Kb-promoter H3K4me3 targets showed increased variation in expression (Figure 3.13 C). On the other hand, genes adjacent to the intergenic H3K4me3 domain showed the lowest expression level, with the largest variation among three types of H3K4me3 localization.

Meanwhile, we found few hundreds of gene-body and intergenic H3K4me3 domains were also enriched in H3K27me3 simultaneously, which we considered as potential bivalent domains. Here, we only found intergenic bivalent domains located at downstream regions of adjacent genes. As expected, those bivalently marked genes in FIL-labeled cells showed reduced expression comparing to H3K4me3 targets, in which bivalent gene-body domains

still had a higher level of expression than intergenic-bivalent domains (Figure 3.13 C). The expression profile comparison showed that the gene-body located H3K4me3 domain (Peak-valley-Peak) had the most significant association with gene activation, while the presence of H3K27me3 compromised such association (Figure 3.13 C). It should be noted that the bivalent marked genes have not been verified by sequential-ChIP, whether H3K4me3 and H3K27me3 are enriched in the same gene or even same nucleosome is not clear in this study.

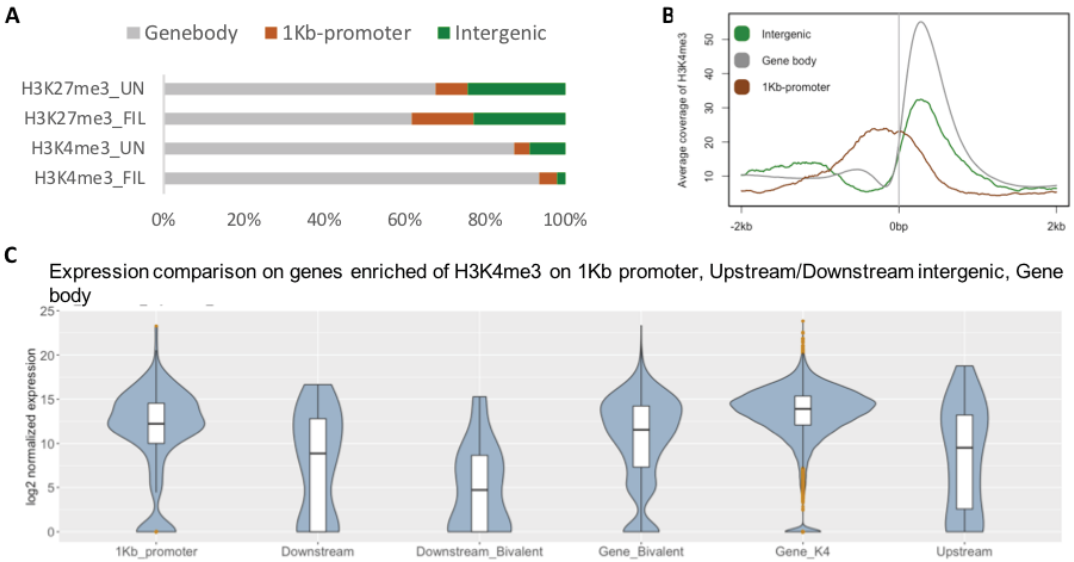


Figure 3.13. H3K4me3 distribution and expression comparison. (A) Reads count of H3K4me3 ChIP-seq and H3K27me3 ChIP-seq on different genome features, including gene-body, 1-KB promoter, and intergenic regions which exclude of the 1-KB promoter. (B) Average coverage of H3K4me3 on intergenic H3K4me3 domains, 1Kb-promoter H3K4me3 domains, and Gene-body H3K4me3 domains. Y axis is the average coverage of H3K4me3 signal; -2Kb, 0bp, 2Kb on X axis represent 2Kb upstream of TSS, TSS, and 2Kb downstream of TSS. (C) Violin box plot of gene expression within different H3K4me3 distributions, including H3K4me3 enriched at 1Kb-promoter of the gene (1Kb_promoter), H3K4me3 enriched at the downstream intergenic region of the gene (Downstream), both H3K4me3 and H3K27me3 enriched at the downstream intergenic regions of the gene (Downstream_Bivalent), both H3K4me3 and H3K27me3 enriched at gene-body (Genebody_Bivalent), H3K4me3 enriched at gene body (Gene_K4), and H3K4me3 enriched at upstream intergenic region of the gene (Upstream); Y axis is the Log2 normalized reads count.

In addition, we found that a higher percentage of intergenic H3K4me3 domains from both FIL-labeled (FIL) and non-FIL labeled cells (UN) were differentially methylated, comparing to other types of H3K4me3 domains (Figure 3.15 D). To better understand the characters of

intergenic H3K4me3 domains, we tested two hypotheses that were proposed in mammalian studies. It has been shown in the human genome study that H3K4me3 were enriched at bidirectional promoters which initiate transcription in both strands (Bornelöv et al., 2015). Thus, we asked if intergenic H3K4me3 domains were surrounded by bidirectional-expressing genes. Based on the strand information of *Arabidopsis* genes (TAIR10), an intergenic region was defined as bidirectional intergenic region if it sat at upstream of both adjacent genes (two adjacent genes are on opposite strand) (Figure 3.15 A). And unidirectional intergenic regions were located between two adjacent genes that were on the same strand (Figure 3.15 B). Using this definition, we counted the number of bidirectional and unidirectional intergenic regions that were targeted by H3K4me3 in FIL-labeled cells (FIL), as well as overall *Arabidopsis* intergenic regions (Figure 3.14 B). However, we found that a lower percentage of intergenic H3K4me3 domains were at bidirectional intergenic regions comparing to overall *Arabidopsis* intergenic regions (Figure 3.14 B).

On the other hand, a study in mammalian cells has shown that some active enhancers are enriched in H3K4me3 (Pekowska et al., 2011). Thus, to explore if those intergenic domains were associated with certain regulatory features, we analyzed the sequence composition and found that H3K4me3 targeted intergenic regions showed an increased percentage of GC content for both FIL-labeled (FIL) and non-FIL labeled cells (UN) (Figure 3.14 C). Interestingly, the CGGCGR motif was identified by DREME, using intergenic H3K4me3 domains against all *Arabidopsis* intergenic regions (E-value=$1.4e^{-18}$) (Figure 3.14 D). 205 domains out of 243 intergenic H3K4me3 domains in FIL-labeled cells contained the CGGCGR motif (Figure 3.14 E). Thereinto, 55 domains showed increased H3K4me3 enrichment in FIL-labeled cells (Figure 3.14 E and F, 3.15 E), while other CGGCGR motif-containing domains were also enriched in H3K4me3 in non-FIL labeled cells. (G)CGGCG or GCGGC(G) motifs are classified as the LBD motif, which can be recognized by LATERAL ORGAN BOUNDARIES (LOB) (Bell et al., 2012). Altogether, although it is still unclear what is the function of those intergenic H3K4me3 domains, it is certainly not the background or random enrichment of H3K4me3. Based on the motif analysis, those intergenic H3K4me3 domains in FIL-labeled cells may mark certain regulatory features.

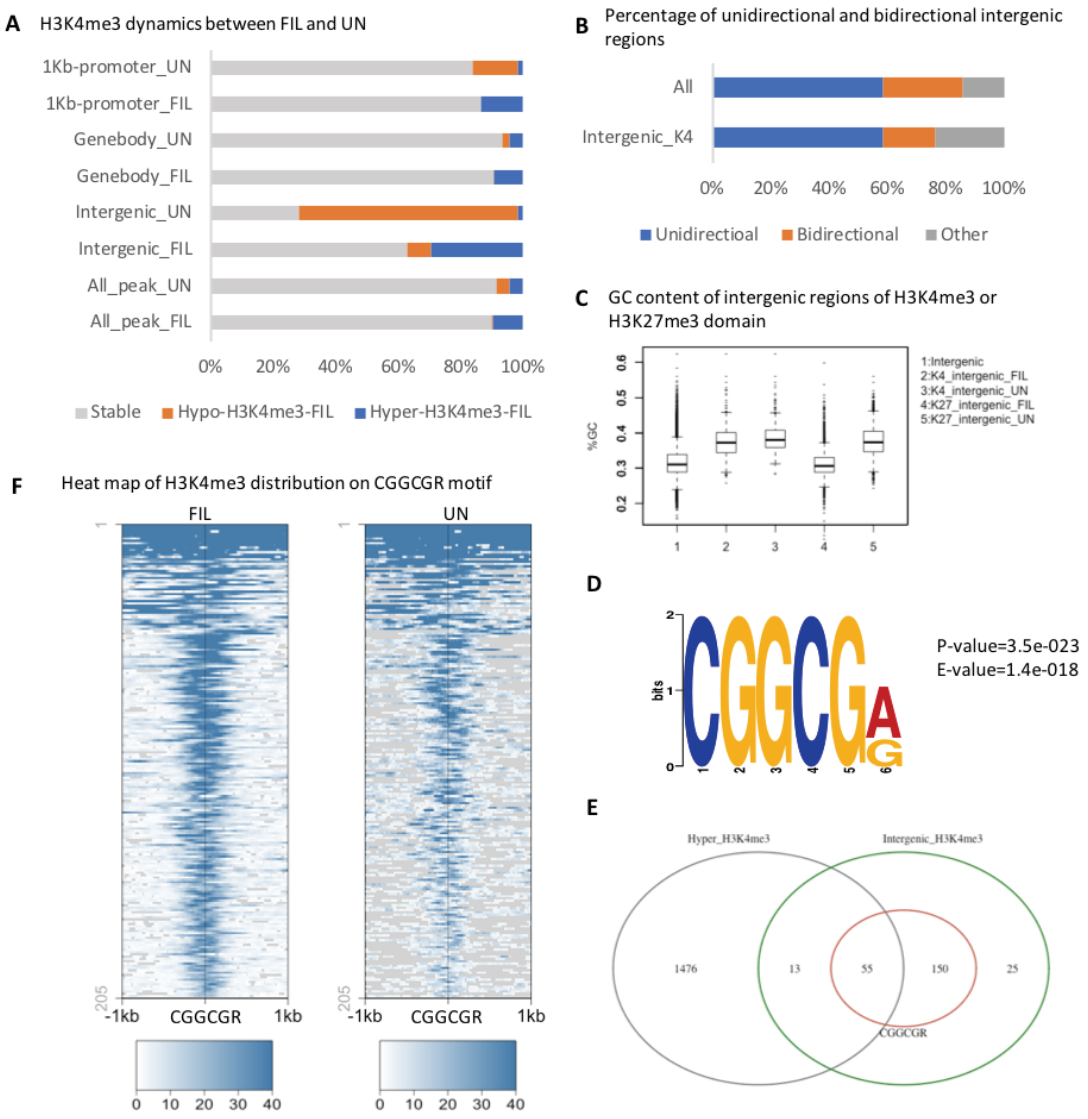


Figure 3.14. Analysis of H3K4me3 domains in intergenic regions. (A) Percentage of each group of H3K4me3-DMRs among gene-body H3K4me3 domains, 1Kb-promoter H3K4me3 domains and intergenic H3K4me3 domains in FIL-labeled cells (FIL) and non-FIL labeled cells (UN); H3K4me3 hypermethylated regions in (FIL) (hyper-H3K4me3-FIL) were in orange, H3K4me3 hypomethylated regions in (FIL) (hypo-H3K4me3-FIL) were in blue, and no significant changes of H3K4me3 between (FIL) and (UN) (stable) were in gray. (B) Percentage of unidirectional, bidirectional intergenic regions among intergenic H3K4me3 domains, examples of unidirectional, bidirectional intergenic regions were illustrated in figure 3.15. (C) Boxplot of %GC in H3K4me3 or H3K27me3 targeted intergenic regions; All intergenic regions in TAIR10 were used as the reference, plotted as "1" on the X axis. (D) The significantly enriched motif among intergenic H3K4me3 domains, identified by DREME algorism. (E) Venn diagram of intergenic H3K4me3 domains that contain CGGCGR motif, Hyper H3K4me3 DMRs, and all intergenic H3K4me3 domains. (F) Heatmap of H3K4me3 distribution on intergenic H3K4me3 domains that contain CGGCGR motif.

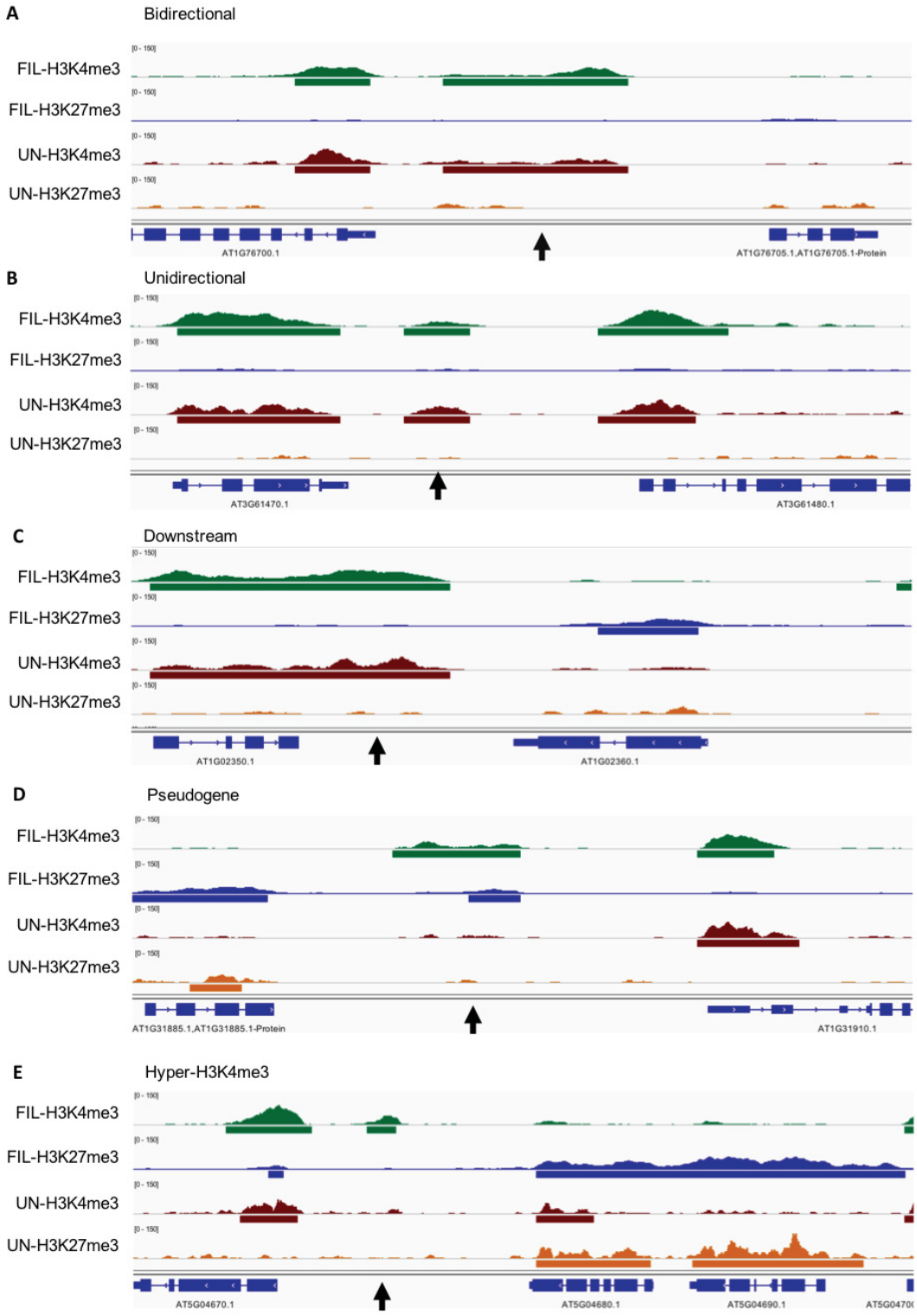


Figure 3.15. iGV browser view of examples of intergenic H3K4me3 domains. (A) H3K4me3 enriched at the bidirectional intergenic region. B. H3K4me3 enriched at the unidirectional region. C. H3K4me3 enriched at a shared downstream intergenic regions (D) Poor annotated intergenic region (hypothetic protein, pseudogene). (E) H3K4me3 hypermethylated intergenic region in FIL-labeled cells.

3.3.4 Analysis of the genome-wide distribution of H3K27me3

As mentioned previously in section 3.3.1, H3K27me3 hypermethylated genes were associated with down-regulation in FIL-labeled cells, while H3K27me3 hypomethylated genes tend to be up-regulated in FIL-labeled cells (Figure 3.9 A). Nevertheless, such association only presented in less than 25% of H3K27me3 DMRs (Figure 3.9 A). In this section, we scrutinized the impact of the size of H3K27me3 domains on gene expression in FIL-labeled and non-FIL labeled cells. H3K27me3 was mostly expanded along the gene body (Figure 3.8 A), which also showed a broader coverage on the intergenic region than H3K4me3 in both FIL-labeled and non-FIL labeled cells (Figure 3.13 A). Because the precise location of the H3K27me3 domains were variable between cell types, we defined five types of H3K27me3 distributions according to genomic features to make the comparison possible (Figure 3.16 A). The category includes type1 which has H3K27me3 on the gene body, the upstream intergenic and the downstream intergenic regions; type2 which has H3K27me3 on the gene body and the downstream intergenic region; type3 which has H3K27me3 on the gene body and the upstream intergenic region; type 4 which only has H3K27me3 on the gene body; type5 which only has H3K27me3 on the intergenic regions (Figure 3.16 A). Overall, broader coverage of H3K27me3 (type 1, 2, and 3) showed lower expression with less variation compare to type 4, which only showed enrichment on gene body (Figure 3.16 B and C). In addition, genes with type 5-H3K27me3 distribution, which only covered the upstream or downstream intergenic region, showed the highest expression profile among H3K27me3 targets. Furthermore, we found that the percentage of down-regulated genes was higher among genes had larger H3K27me3 domains in FIL-labeled than non-FIL labeled cells, comparing to the smaller H3K27me3 domains (Figure 3.16 D and E). Here, we compared the coverage of H3K27me3 size based on the five types of H3K27me3 domains. For example, if a gene was targeted by a type1 H3K27me3 domain in FIL-labeled cells while type2 in non-FIL labeled cells, we consider as the larger domain in FIL-labeled cells. As a result, we found that the percentage of down-regulation was higher among genes with a larger H3K27me3 domain in FIL-labeled. However, only around 20% of the H3K27me3 domains with different coverage showed differential expression, while the majority remain

silenced or low-level expressed regardless of the dynamics of domain size between FIL-labeled and non-FIL labeled cells (Figure 3.16 D and E).

A

	Type_1	Type_2	Type_3	Type_4	Type_5
Upstream intergenic	✓		✓		✓*
Downstream intergenic	✓	✓			✓*
Gene body	✓	✓	✓	✓	

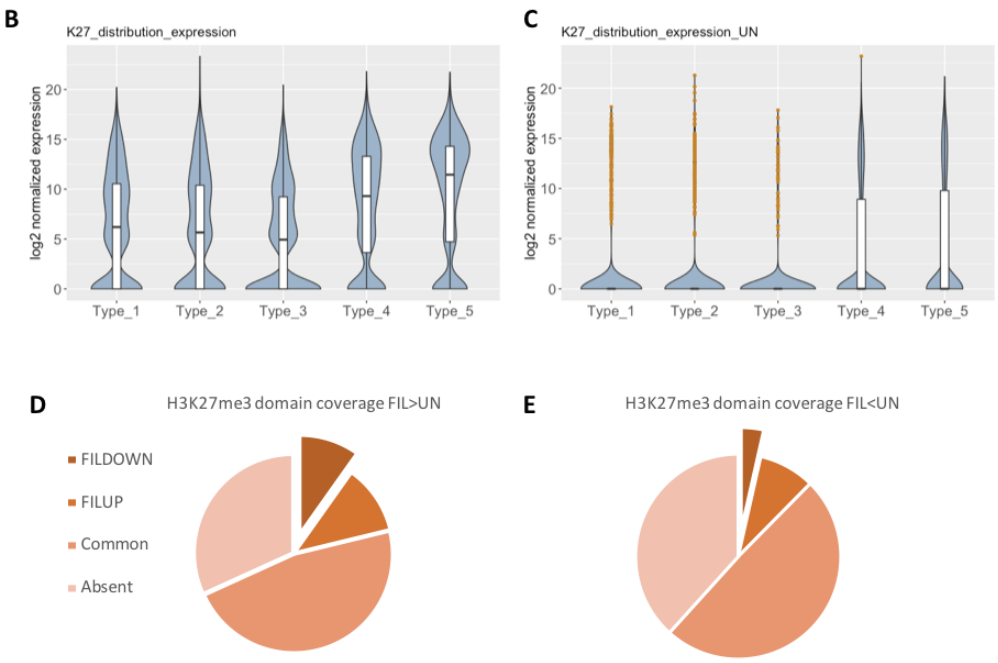


Figure 3.16 Analysis of different H3K27me3 distribution. (A) The definition of each type of H3K27me3 distribution, * means the H3K27me3 either enriched at downstream intergenic region or upstream intergenic region. B. Violin boxplot of gene expression among five types of H3K27me3 domains in FIL-labeled cells (FIL). (C) Violin boxplot of gene expression among five types of H3K27me3 domains in non-FIL labeled cells (UN). (B and C) five types of H3K27me3 domains refer to figure (A). (D) Pie chart of each expression groups among the expanded H3K27me3 domains in FIL-labeled cells (FIL) comparing to non-FIL labeled cells (UN). (E) Pie chart of each expression group among narrowed H3K27me3 domains in FIL-labeled cells (FIL) comparing to non-FIL labeled cells (UN). (D and E), Legend “FILDOWN” represents expression in FIL<UN, “FILUP” represents expression in FIL>UN, “Common” represents expression in FIL≈UN>0, “Absent” represents expression in FIL=UN=0.

Besides H3K27me3 enrichment on protein coding genes, we also found that 2828 TEs and 2679 TEs were targeted by H3K27me3 in FIL labeled cells (FIL) and non-FIL labeled cells

(UN). Among different transposon element families, DNA and LINE TE families were enriched in H3K27me3 targets in both cell types, especially DNA TE family showed more than three folds enrichment, whereas LTR/Gypsy was under-represented among H3K27me3 targets (Figure 3.17 A).

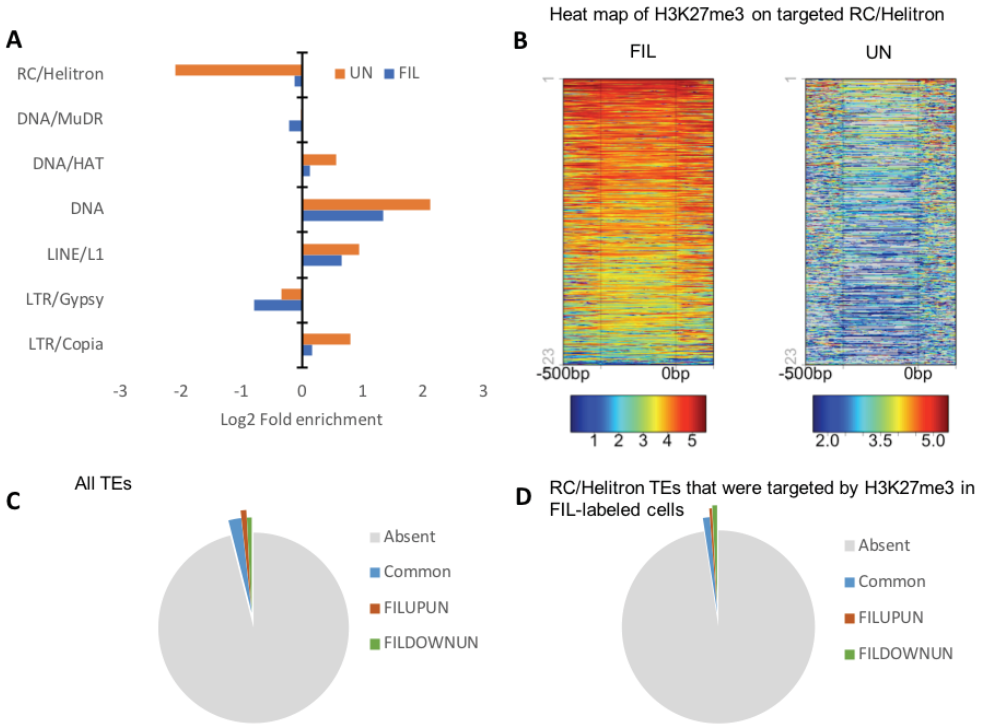


Figure 3.17. Analysis on H3K27me3 enriched transposon elements. (A) Fold enrichment of each TE family among H3K27me3 targets in FIL-labeled cells (FIL) and non-FIL labeled cells (UN), X axis is calculated by log2(frequency of each TE families among H3K27me3 targets/frequency of each TE families among all TE). (B) Heatmap of H3K27me3 distribution on the H3K27me3 targets of RC/Helitron in FIL labeled cells (FIL), and respective H3K27me3 distribution in non-FIL labeled cells. (C) Pie chart of each expression group of TEs (All *Arabidopsis* TEs). (D) Pie chart of each expression group of TEs among H3K27me3 targets in RC/Helitron TEs in FIL-labeled cells (FIL). (C and D) Legend "FILDOWNUN" represents expression in FIL<UN, "FILUPUN" represents expression in FIL>UN, "Common" represents expression in FIL≈UN>0, "Absent" represents expression in FIL=UN=0. Note, the TE families that have less than 300 TEs were not included in Figure (A) (SINE, DNA Marine, DNA/Pogo, DNA/Harbinger, DNA/En-Spm).

On the other hand, the RC/Helitron family as the most abundant TEs in *Arabidopsis* (41%, based on current annotation) showed decreased enrichment in non-FIL labeled cells. In total, 1123 RC/Helitron TEs were targeted by H3K27me3 in FIL-labeled cells, while depleted

of H3K27me3 in non-FIL labeled cells (Figure 3.17 B). However, the majority of TEs in RC/Helitron family remained in the silent state regardless of the differential enrichment of H3K27me3 (Figure 3.17 C and D). Therefore, it is not clear if the dynamic regulation of H3K27me3 on RC/Helitron has any biological function.

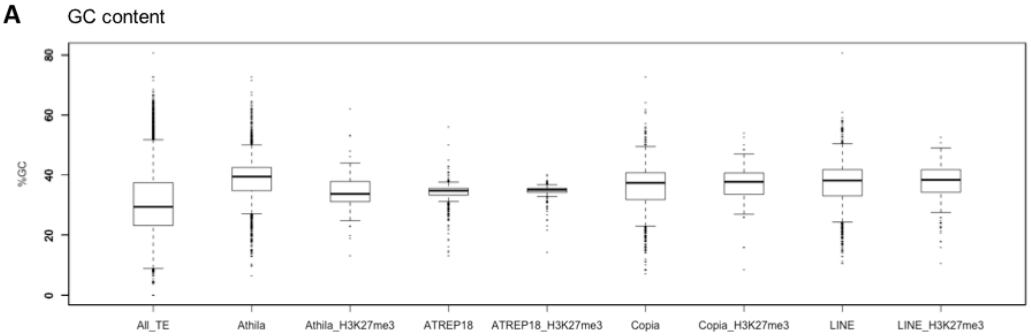
Table 3.2 Summary of fold enrichment of H3K27me3 targets in Athila super families

	Total number	FIL-labeled cells		Non-FIL labeled cells	
		No. Targets	Fold of enrichment	No. targets	Fold of enrichment
ATHILA	198	2	0.110996814	2	0.1432021
ATHILA0_I	138	2	0.159256298	1	0.10273194
ATHILA2	413	3	0.079820953	2	0.06865379
ATHILA3	243	0	0	6	0.35004957
ATHILA4	250	7	0.307683168	0	0
ATHILA4A	310	11	0.389921066	0	0
ATHILA4B_LTR	143	6	0.461063689	0	0
ATHILA4C	206	5	0.266715645	1	0.06882042
ATHILA5	131	5	0.419415442	2	0.21644285
ATHILA6A	247	4	0.177954406	3	0.17219037
ATHILA6B	134	2	0.164010218	2	0.21159714
ATHILA7A	31	2	0.708947392	0	0

In addition, in order to investigate if H3K27me3 targets share any sequence specificity, we further analyzed sequence composition among H3K27me3 enriched and depleted TE super families, in which ATREP18 belongs to DNA TE family, Copia belongs to LTR/Copia, LINE belongs to LINE/L1 and Athila is from LTR/Gypsy family. Interestingly, although Athila TEs were generally not targeted by H3K27me3 in both cell types, those Athila TEs that were targeted by H3K27me3 showed significantly decreased % of GC and dinucleotides of GC/CG/GG/CC (Figure 3.18 A and C). Meanwhile, other TEs family showed no significant change in GC composition in H3K27me3 targets (Figure 3.18 A and C). Previous studies have shown that Athila were silenced by DNA methylation (Slotkin et al., 2009; McCue et al., 2012; Brinkman et al., 2012; Jermann et al., 2014; Tanay et al., 2007; Walter et al., 2016), which explained the under-representation of Athila TEs among H3K27me3 targets, since the antagonistic relation between DNA methylation and H3K27me3. Meanwhile, one possible hypothesis for the decreased GC content in the H3K27me3-enriched Athila TEs is that

H3K27me3 may act as an additional silencing mark when Athila TEs have less cytosine for DNA methylation.

A



B

Table: Wilcox test on composition of GC content in H3K27me3 TE targets and background, in accordance of Figure A

	Athila	ATREP	Copia	LINE
ATHILA H3K27me3	7.54e-11**			
ATREP18 H3K27me3		0.127		
Copia H3K27me3			0.462	
LINE H3K27me3				0.5048

C

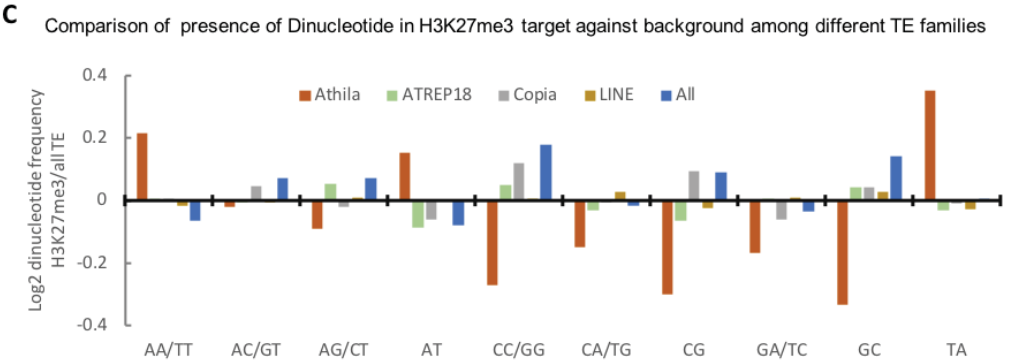


Figure 3.18. (A) Boxplot of %GC in all TE, Athila superfamily, H3K27me3 targets in Athila superfamily, ATREP18 super family, H3K27me3 targets in ATREP18 superfamily, Copia super family, H3K27me3 targets in Copia superfamily, LINE family, H3K27me3 targets in LINE family; Here, Copia family served as an additional reference which shows unbiased number of H3K27me3 targets. (B) Wilcox statistic test of %GC between H3K27me3 targets of TE families and overall respective TE families, those TE super families include Athila, ATREP18, Copia, and LINE; "*" marked the statistical significance. (C) Column plot of fold enrichment of dinucleotides among H3K27me3 targets of Athila (orange), ATREP18 (green), Copia (gray), LINE (yellow), and all TEs (blue); Y axis is calculated by log2(frequency of dinucleotides among H3K27me3 targets from each TE families/frequency of dinucleotides of respective TE families).

3.4 Integrated comparison of Transcriptome profiles, H3K4me3 and H3K27me3 profiles between FIL-labeled cells (FIL) and non-FIL labeled cells (UN)

3.4.1 Photosynthesis regulators were down-regulated in FIL-labeled cells (FIL)

In differential expression analysis, we found 7634 genes were up-regulated in FIL-labeled cells (FIL) comparing to non-FIL-labeled cells (UN), and 4153 genes were down-regulated in FIL labeled cells (FIL). As expected, many of the genes involved in abaxial and adaxial cell fate determination were differentially regulated, although none of the related GO terms were found in GO analysis (Table 3.3). Genes involved in abaxial and adaxial cell fate determination will be discussed in section 3.4.5. Based on GO analysis, we found photosynthesis-related genes were significantly enriched in down-regulated DEGs, but were expressed at a high level in both tissues. Most interestingly, photo-inhibition related genes were also significantly enriched among down-regulated genes (Figure 3.19 A), with more prominent expression fold changes comparing to photosynthesis related-genes (Figure 3.19 B and C). This corresponds to the photoinhibition gradient studies in sunflower and spinach (Oguchi et al., 2011), which have demonstrated that abaxial cells are more susceptible to high light intensity.

Table 3.3. Top 10 significantly enriched GO terms in down-regulated and up-regulated genes in FIL-labeled cells (FIL).

Expression group	GO ID	Fold of enrichment
Down-regulated (FILDOWNUN)	reductive pentose-phosphate cycle (GO:0019253)	4.77
	photosynthesis, dark reaction (GO:0019685)	4.43
	hexose biosynthetic process (GO:0019319)	4.24
	carbon fixation (GO:0015977)	4.14
	gluconeogenesis (GO:0006094)	4.08
	photosynthesis, light harvesting in photosystem I (GO:0009768)	4.02
	monosaccharide biosynthetic process (GO:0046364)	3.85
	glucose metabolic process (GO:0006006)	3.22
	protein-chromophore linkage (GO:0018298)	3.2
	photosynthesis, light harvesting (GO:0009765)	3.15
Up-regulated (FILUPUN)	nucleus organization (GO:0006997)	2.06
	protein modification by small protein removal (GO:0070646)	2.01
	DNA conformation change (GO:0071103)	1.97
	regulation of gene expression, epigenetic (GO:0040029)	1.92
	positive regulation of organelle organization (GO:0010638)	1.87
	regulation of chromosome organization (GO:0033044)	1.79
	gene silencing (GO:0016458)	1.78
	vegetative to reproductive phase transition of meristem (GO:0010228)	1.73
	mRNA processing (GO:0006397)	1.7
	macromolecule methylation (GO:0043414)	1.68

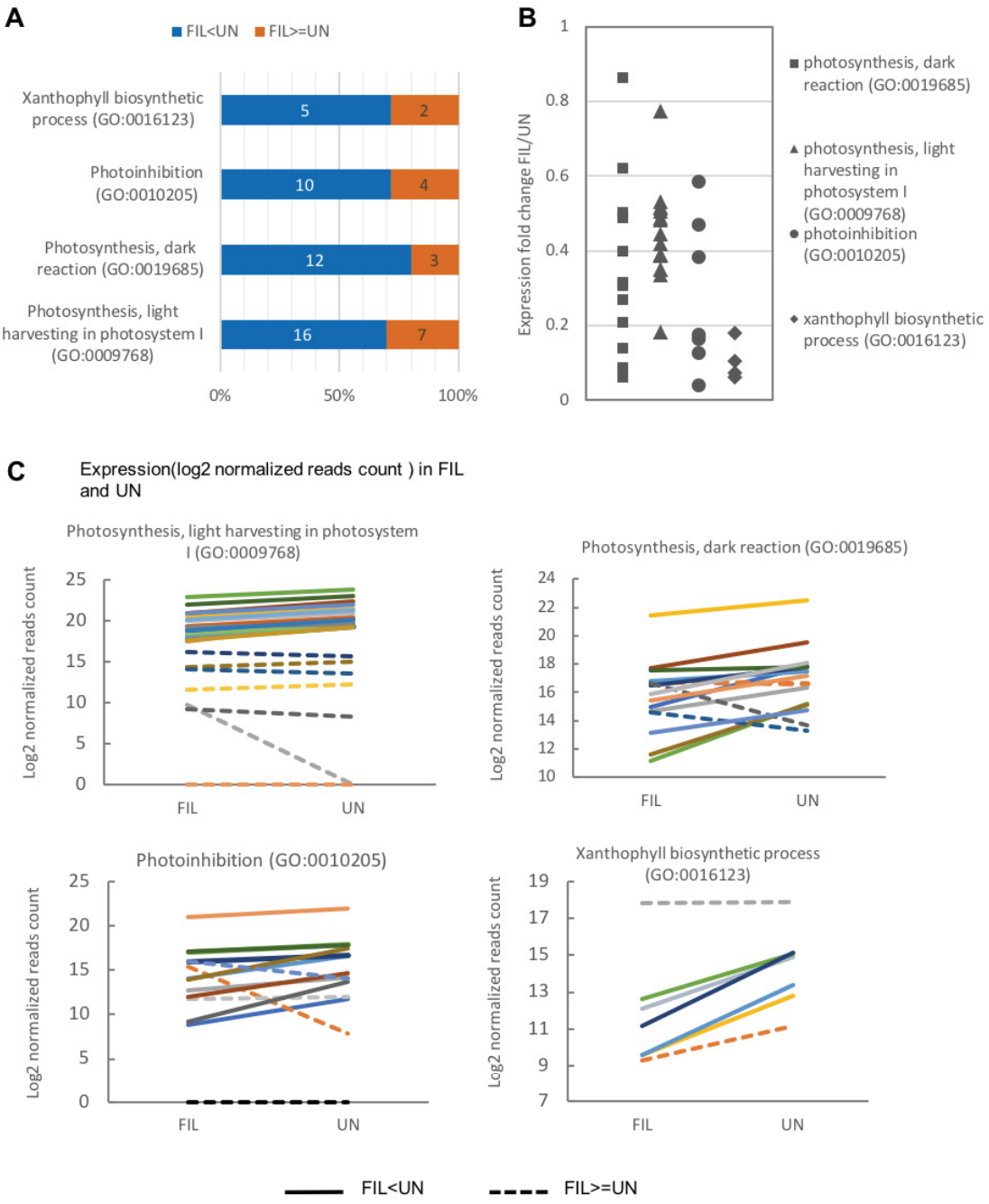


Figure 3.19. (A) Plot of the percentage of down-regulated genes (FIL<UN, in blue) and other genes (FIL>=UN in orange) among photosynthesis-related GO terms; The number of genes within each category were listed in respective columns. (B) Plot of the expression fold changes of down-regulated genes of each GO terms in accordance to figure (A). (C) Plot of expression pattern of genes of GO terms in accordance to figure (A), dash lines represent genes that were not down-regulated in FIL-labeled cells (FIL>=UN); Y axis is Log2 normalized read count.

3.4.2 Genes involved in dsRNA biogenesis were exclusively expressed in FIL-labeled cells (FIL)

In the GO analysis on up-regulated DEGs in FIL-labeled cells (FIL), several GO terms related to epigenetic gene regulation were found in this study (Table 3.3). Furthermore, we found that more than 50% of genes involved in the RdDM pathway showed significantly differential expression between FIL-labeled cell (FIL) and non-FIL labeled cell (UN). 18 genes showed increased expression in FIL labeled cells (FIL) (Table 3.4), 8 genes showed increased expression in non-FIL labeled cells (UN) (Table 3.5), and 13 genes maintained similar expression levels in both cell types (data not shown). Interestingly, NRPD1 (subunits of Pol IV) and RDR2 which involved in the biosynthesis of dsRNA (26-45nt) were found only expressed in FIL-labeled cells (FIL) (Table 3.4). In addition, RDR6 involved in the production of 21-22nt dsRNA was also found exclusively expressed in FIL-labeled cells (FIL) (Table 3.4). Meanwhile, subunits of Pol V which involved in the establishment of DNA methylation in the RdDM pathway, including *NRPE1*, *NRPD2,* and *NRPE5* kept relatively high-level expression in both tissues with slightly increased expression in FIL-labeled cells (FIL) (Table 3.4). Because of the key function of the RdDM pathway is to maintain the silent status of transposon element through the establishment of DNA methylation, we asked if TEs were differentially expressed between FIL-labeled cells and non-FIL labeled cells. By comparing the expression profiles of transposon element, we found that more than 90% of TEs remained silent in both tissues, despite the absence of *NRPD1* and *RDR2* expression in non-FIL labeled cells (UN). Furthermore, we cross compared expression profiles of *nrpd1* (Yang et al., 2017) with FIL-labeled cells (FIL) and non-FIL labeled cells (UN). However, only dozes of down-regulated genes and TEs in FIL-labeled cells (FIL) were overlapped with the derepressed loci of *nrpd1* (Yang et al., 2017) (Figure 3.20 A and B). Meanwhile, up-regulated genes and TEs in FIL-labeled cells also showed similar range of overlap with the derepressed loci of *nrpd1* (Figure 3.20 A and B). Thus, the absent of *NRPD1* expression in non-FIL labeled cells does not lead to the *nrpd1* mutant's expression profile.

Table 3.4. List of up-regulated genes involved in RdDM pathway in FIL-labeled cells

Proteins	Description	Leaf phenotype	FIL	UN
MORC6	A MORC-type ATPase	(Brabbs et al., 2013)	11.95	0.00
NRPD1	The unique, largest subunit of Pol IV	NA	9.56	0.00
HUA ENHANCER 1 (HEN1)	A small RNA methyltransferase	Small leave, upwardly curling	13.22	0.00
RDR2	An RNA-dependent RNA polymerase	Downwardly curling (*rdr1/2/6*, resemble *rdr6*)	12.02	0.00
HISTONE DEACETYLASE 6 (HDA6)	A histone deacetylase	NA	12.32	0.00
NEEDED FOR RDR2-INDEPENDENT DNA METHYLATION (NERD)	Involved in non-canonical RdDM	NA	20.83	19.12
AGO4	An Argonaute protein	Lacks trichomes	15.92	13.54
NRPE1	The unique, largest subunit of Pol V	NA	18.37	17.55
NRPD2/NRPE2	The shared second largest subunit of Pol IV and Pol V	NA	12.43	9.02
UBIQUITIN-SPECIFIC PROTEASE 26 (UBP26)	A histone H2B deubiquitinase	NA	17.39	14.57
NRPE5	A special fifth largest subunit of Pol V	NA	16.81	15.67
CHROMOMETHYLASE 2 (CMT2)	A DNA methyltransferase specific to CHH	NA	13.66	10.42
JMJ14	A histone demethylase	NA	14.88	13.77
NRPB1	The largest subunit of Pol II	NA	17.05	16.30
KTF1	Contains the AGO hook motif; involved in Pol V transcription	NA	15.91	15.02
SUVH4	A H3K9 methyltransferase	Curled leaves(*suvh4/drm2*)	12.73	7.87
IDN2 PARALOGUE 1 (IDP1)	Forms a complex with IDN2	NA	14.29	13.56
DRM2	A *de novo* DNA methyltransferase	Curled leaves(*drm2/cmt3 or suvh4/drm2*)	15.31	13.34

Table 3.5. List of downregulated genes involved in RdDM pathway in FIL-labeled cells

Proteins	Description	Leaf phenotype	FIL	UN
SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1)	Binds to methylated H3K9 and recruits Pol IV	NA	12.61	15.50
LYSINE-SPECIFIC HISTONE DEMETHYLASE 1 (LDL1)	A histone demethylase	NA	7.01	10.68
SUVH9	An SRA domain protein that binds to methylated DNA and recruits Pol V	NA	12.71	14.94
DEFECTIVE IN MERISTEM SILENCING 3 (DMS3)	A SMC solo hinge protein; part of the DDR complex; involved in the Pol V pathway	NA	11.01	14.73
AGO6	An Argonaute protein	NA	0.00	11.48
DICER-LIKE 3 (DCL3)	A Dicer endonuclease that produces 24-nucleotide siRNAs	Downwardly curling	0.00	16.72
IDP2	Forms a complex with IDN2	NA	0.00	11.38
NRPE9B	A special ninth largest subunit required for Pol V activity	NA	0.00	13.40

We then asked whether dsRNA was accumulated in FIL-labeled cells (FIL), since NRPD1 and RDR2 are mostly play a part in dsRNA biosynthesis (Eamens et al., 2008; Blevins et al., 2015). Due to the lack of dsRNA profile in this study, we compared transcriptome data from our study to a published dataset of RDR2/NRPD1 dependent dsRNA (P4RNA) (Li et al., 2015). There are in total 17605 P4RNA loci have been identified by dsRNA-seq on *dcl2 dcl3 dcl4* and *dcl2 dcl3 dcl4 nrpd1* (Li et al., 2015). The majority of P4RNA loci were not transcribed in FIL-labeled (FIL) and non-FIL labeled cells (UN), and the number of P4RNA loci that were hyper or hypo expressed showed no significant difference (Figure 3.20 C and E). Nevertheless, the total transcript reads count on all P4RNA loci were significantly increased in FIL-labeled cells (FIL) (Figure 3.20 D). Altogether, based on the exclusive expression of *NRPD1* and *RDR2* in FIL-labeled cells and the silent status of TE in non-FIL labeled cells, we speculate that dsRNA may be a mobile signal, moving from FIL-labeled cells to other cells. However, we lack of definitive evidences to demonstrate the mobility.

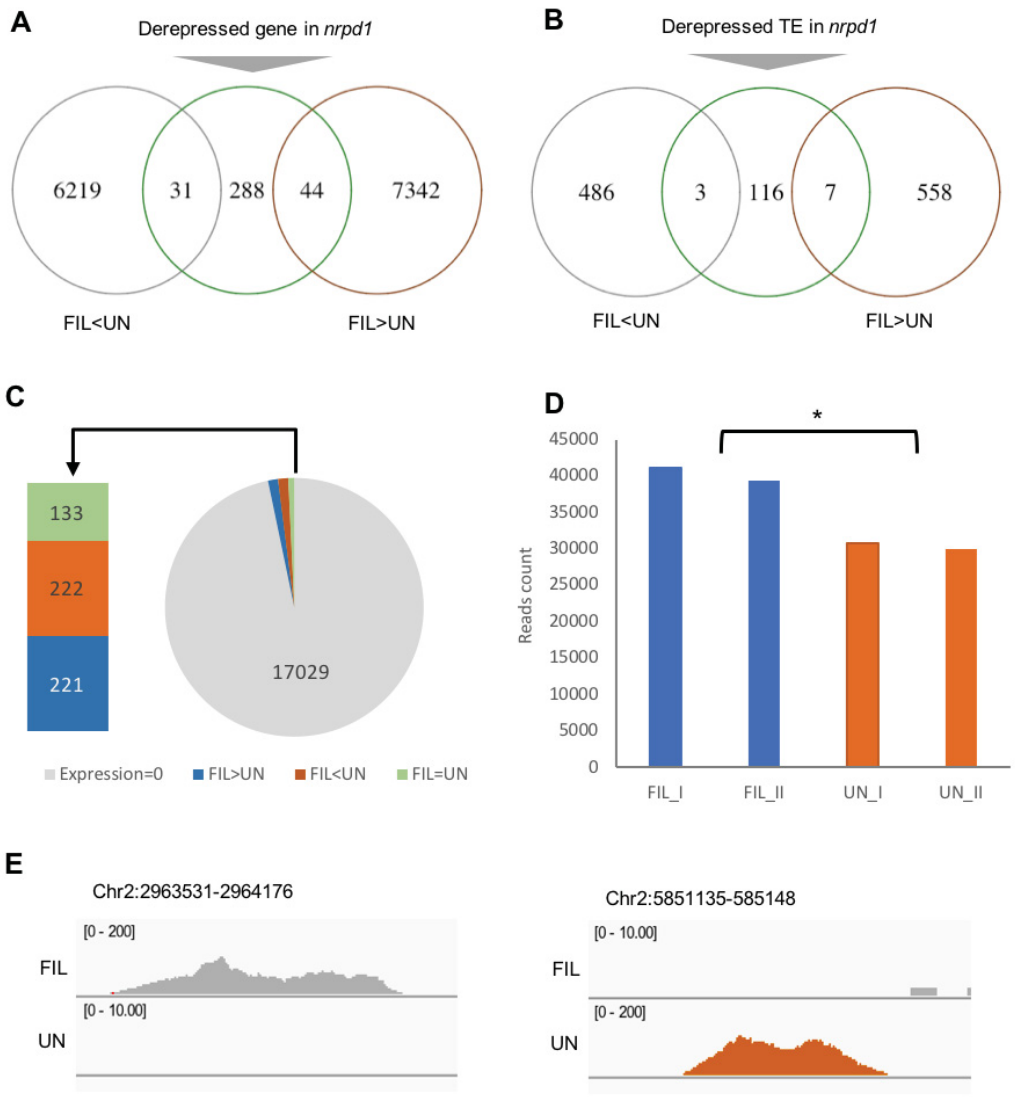


Figure 3.20. (A) Venn diagram of derepressed genes in nrpd1 (Yang et al., 2017), down-regulated genes, and up-regulated genes in FIL-labeled cells (FIL). (B) Venn diagram of derepressed TEs in nrpd1 (Yang et al., 2017), down-regulated TEs, and up-regulated TEs in FIL-labeled cells (FIL). (C) Pie chart of the number of each expression group on P4RNA loci, including expression=0, expression in FIL>UN, expression in FIL<UN, and expression in FIL=UN. (D) Total reads count of transcription on all P4RNA loci in FIL labeled cell (FIL) and non-FIL labeled cells (UN). (E) iGV browser of two examples of P4RNA loci that showed differential expression between FIL-labeled and non-FIL labeled cells. Note, P4RNA loci mentioned in figure (C-E) were identified by (Li et al., 2015).

3.4.3 *MORC6* was exclusively expressed in FIL-labeled cells

Among the up-regulated genes involved in RdDM pathway, *MORC6* was exclusively expressed in FIL-labeled cells (FIL) (Table 3.4, Figure 3.21 B). MORC6 is a member of the GHKL ATPase super family, containing an ATPase domain, a coiled-coil domain at C-terminal, and a motif of intraflagellar transport complex B(IFT20) which is known as a transporting molecule through microtubules (Figure 3.21 A). Studies have shown that MORC6 may be involved in RdDM by regulating chromatin condensation (Liu et al., 2016; Jing et al., 2016; Brabbs et al., 2013; Moissiard et al., 2012). However, by comparing our datasets to the published DNA-methylation dataset of *morc6* (Liu et al., 2016), we found that most of the MORC6-mediated DNA methylation target loci remain silenced in both FIL-labeled (FIL) and non-FIL labeled cells (UN) (Figure 3.21 C). This suggested that the absence of MORC6 expression did not lead to the de-repression of target loci in non-FIL labeled cells (UN) (Figure 3.21 C). Meanwhile, by comparing our datasets to the published transcriptome dataset of *morc6* (Liu et al., 2016), we found that upregulated coding genes in *morc6* were preferentially up-regulated in FIL-labeled (FIL) (*chi*-square p-value=4.10e-08) (Figure 3.21 D). But this observation contradicted to the exclusive expression of *MORC6* in FIL-labeled cells. Altogether, based on the exclusive expression of MORC6 in FIL-labeled cells and the presence of IFT20 motif in MORC6 protein, we propose that MORC6 may be a mobile protein that move from FIL-labeled cell to other tissues. Meanwhile, the exclusive expression MORC6 in of FIL-labeled cells do not correspond the observation that up-regulated genes in FIL-labeled cells were actually also up-regulated in *morc6*.

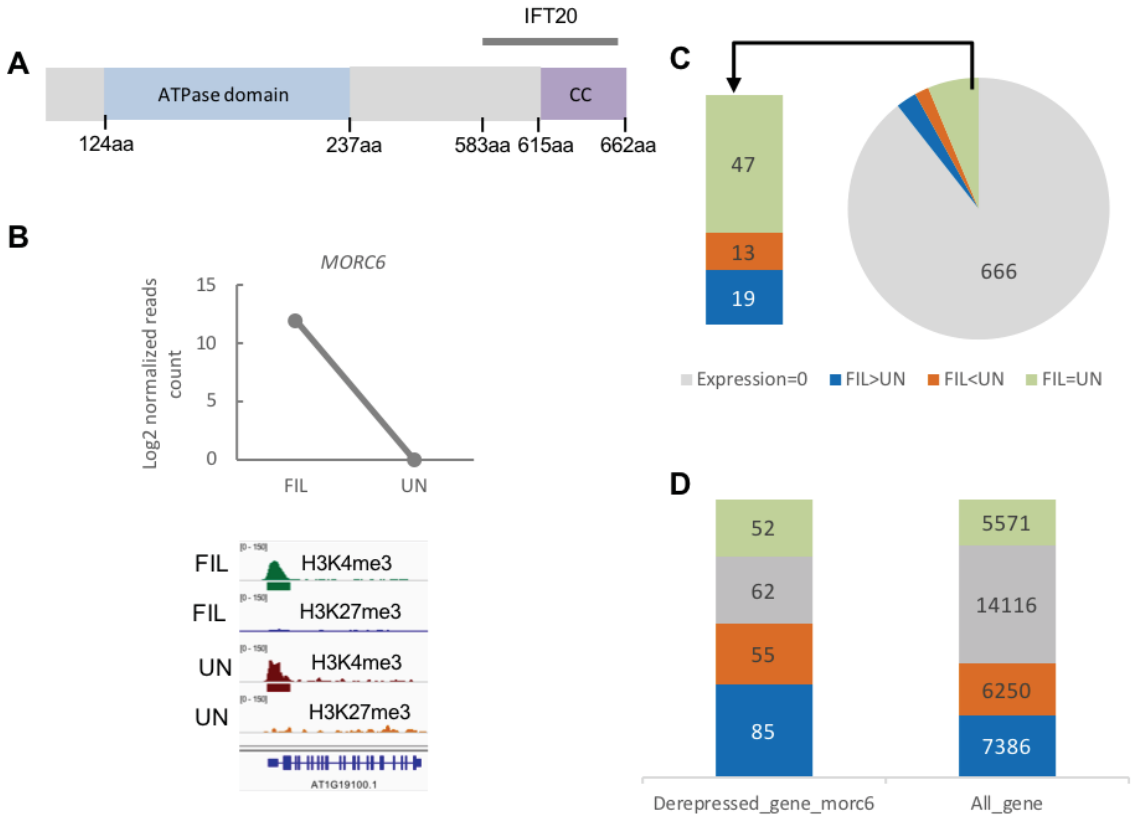


Figure 3.21. (A) Predicted protein domains in MORC6, CC is the abbreviation for the Coiled-Coil domain; Between 583aa and 658aa is an IFT20 motif (Intraflagellar transport complex B). (B) MORC6 expression, H3K4me3, and H3K27me3 status in FIL labeled cell (FIL) and non-FIL labeled cells (UN). (C) Pie chart of the number of MORC6 target genes in each expression group, including expression=0, expression in FIL>UN, expression in FIL<UN, and expression in FIL=UN; here the MORC6 targets refer to the DNA hypomethylated regions in *morc6* (Liu et al., 2016). (D) Stacked column plot of the percentage of FIL>UN, FIL<UN, FIL=UN, Expression=0 among upregulated genes in *morc6*; here the upregulated genes in *morc6* refer to (Liu et al., 2016); "All_gene" used the percentage of each expression category among all 33323 TAIR10 genes, as a reference.

3.4.4 Abaxial-adaxial cell determinants were dynamically regulated on transcription, H3K4me3, and H3K27me3

As expected, among genes that were differentially methylated between FIL-labeled (FIL) and non-FIL labeled cells (UN), several GO terms related to leaf development were enriched, including abaxial cell fate determination (GO:0010158) and xylem development (GO:0010089) (Table 3.6. Figure 3.22 and 3.23). Meanwhile, most of those differentially methylated genes also showed changes in expression (Figure 3.22 and 3.24). *YAB/KAN* gene family and *ARF3* are known as key determinants of abaxial cell fate (Siegfried et al., 1999; Kerstetter et al., 2001; Adenot et al., 2006; Sarojam et al., 2010; Kelley et al., 2012), was up-regulated in FIL-labeled cells (FIL) and also showed increased H3K4me3 enrichment. Interestingly, YAB and KAN gene families were also found targeted by H3K27me3, suggesting potential bivalency status (Figure 3.22). LUG, which forms a regulatory complex with YAB in leaves to promote abaxial cell identity (Stahle et al., 2009), also showed increased expression with increased H3K4me3 in FIL-labeled cell (Figure 3.22). *AS2*, which was known directly repressed by KAN (Matsumura et al., 2016; Iwasaki et al., 2013; Iwakawa et al., 2007; Bell et al., 2012; Lin et al., 2003), showed no expression in both tissues while targeted by increased H3K27me3 in FIL-labeled cells. Meanwhile, we found that several auxin-related genes were differentially expressed and methylated in FIL-labeled cells (Figure 3.22). Previous studies have shown that adaxial hyper expressed MP only becomes an active form with the presence of auxin which is enriched in the abaxial side. Moreover, *PIN1* expression is directly promoted by MP, and in turn reinforce the maximum auxin signaling at middle domain by directing auxin flow (Bhatia et al., 2016; Wenkel et al., 2007; Ilegems et al., 2010; Guan et al., 2017; Qi et al., 2014). In our study, *MP* showed down-regulation with increased enrichment of H3K27me3 at promoter region in FIL-labeled cells, while *PIN1*'s decreased expression was coupled with increased H3K27me3 coverage at gene body in FIL-labeled cells (Figure 3.22). In addition, the auxin influx carrier *AUX1* was also down-regulated in FIL-labeled cells (Figure 3.22). Together, it supported the network of auxin signaling in abaxial and adaxial cell fate determination.

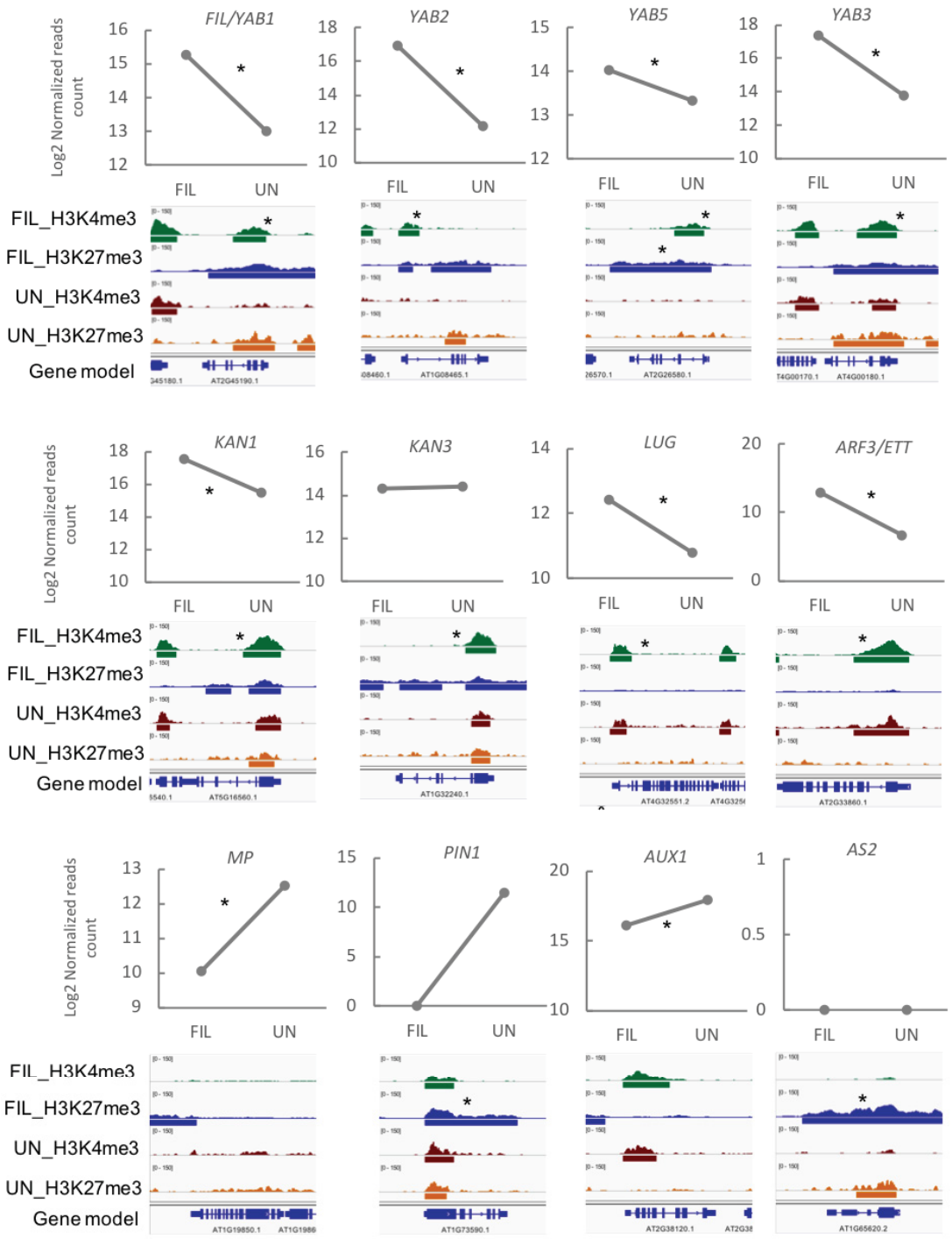


Figure 3.22. Gene expression and iGV browser view of H3K4me3 and H3K27me3 on genes involved in the abaxial cell fate determination pathway. * marked the gene expression or H3K4me3 or H3K27me3 is significantly different between FIL-labeled cells and non-FIL labeled cells.

Furthermore, several GOs related to xylem differentiation and secondary cell wall biogenesis were significantly enriched in H3K27me3 hyper-DMRs in FIL-labeled cells (FIL) (Table 3.6).

Table 3.6. Significantly enriched GO terms among genes showed hyper H3K27me3 or H3K4me3 in FIL-labeled cells (hypo-DMRs of H3K4me3 and H3K27me3 showed no significant GO terms)

DMRs	GO	Fold enrichment
Hyper_H3K27me3	regulation of secondary cell wall biogenesis (GO:2000652)	4.02
	xylem development (GO:0010089)	3.48
	cotyledon development (GO:0048825)	3.42
	maintenance of meristem identity (GO:0010074)	3.25
	regulation of cell wall organization or biogenesis (GO:1903338)	3.22
	brassinosteroid metabolic process (GO:0016131)	3.19
Hyper_H3K4me3	abaxial cell fate specification (GO:0010158)	14.2
	red light signaling pathway (GO:0010161)	9.46
	shade avoidance (GO:0009641)	7.1
	very long-chain fatty acid biosynthetic process (GO:0042761)	6.68
	photosynthesis, light harvesting in photosystem I (GO:0009768)	6.31
	hormone-mediated signaling pathway (GO:0009755)	2.13

However, most of those genes related to xylem development showed no significant difference in expression between FIL-labeled cells (FIL) and non-FIL labeled cells (UN). For example, *LBD18, LBD30, KNAT7* which play a part in xylem development (Endo et al., 2015; Soyano et al., 2008; Li et al., 2012), showed increased enrichment of H3K27me3 in FIL-labeled cells (Figure 3.23). Particularly, LBD30 and KNAT7 were exclusively enriched in H3K27me3 in FIL-labeled cells (Figure 3.23). The increased H3K27me3 in FIL-labeled cells can be explained by the fact that xylem is located on the adaxial side of the leaf. However, all three genes were not differentially expressed in both FIL-labeled and non-FIL labeled cells. One possible explanation could be that the mixed cell types in non-FIL labeled cells diluted

the expression signal. Meanwhile, it again suggested the weak association between the dynamics of H3K27me3 and expression.

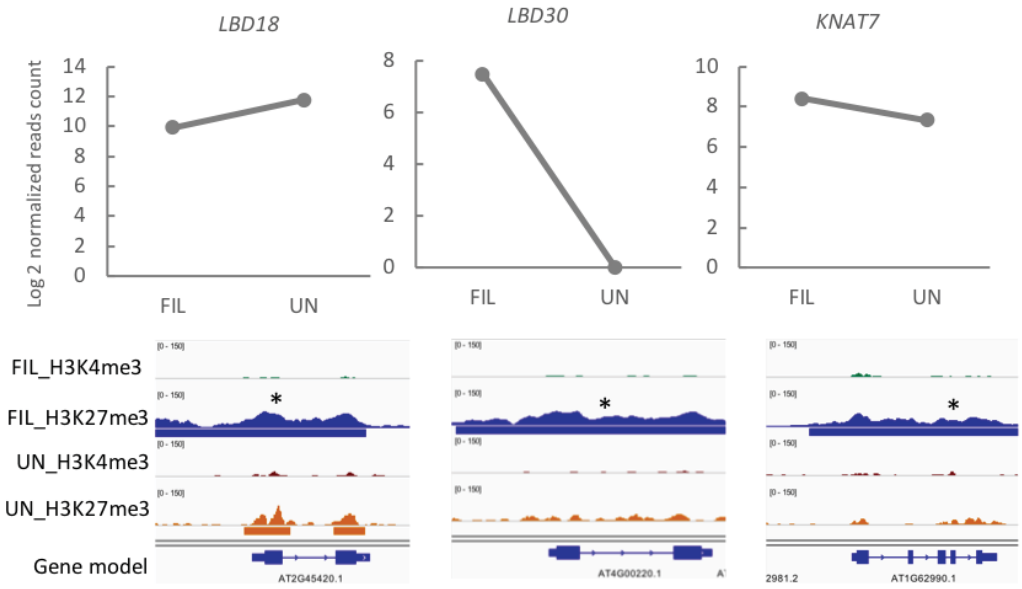


Figure 3.23. Gene expression and iGV browser view of H3K4me3 and H3K27me3 on genes involved in the xylem development. * marked the gene expression or H3K4me3 or H3K27me3 is significantly different between FIL-labeled cells and non-FIL labeled cells.

Also, we found post-transcriptional regulators that play a role in leaf polarity cell determination were differentially expressed without significant change of H3K4me3 or H3K27me3 between FIL-labeled (FIL) and non-FIL labeled cells (UN). Studies have shown that tasiARF is enriched in adaxial tissue and involved in guiding the expression pattern of *ARF3/4* on the leaf (Adenot et al., 2006; Allen et al., 2005). In agree with the tasiARF pathway, we found that *TAS3* and *SGS3* expression were significantly decreased, and *AGO7* was absent in FIL-labeled cells (FIL) (Figure 3.24). However, *RDR6* was exclusively expressed in FIL-labeled cells with slightly increased H3K4me3 enrichment (less than two folds) (Figure 3.24). This contradicts the doctrine of tasiARF-ARF3/4 regulation network and is also against the shared phenotype among *ago7-1, dcl4-2, rdr6, sgs3-1, tas3-1,* and 35S::*ARF3,* which showed down-ward curling leaf margin (Adenot et al., 2006). We lack data to verify whether the absence of *RDR6* expression in non-FIL labeled cells is caused by the dilution effect from mixed tissue or reflects the actual biology in abaxial tissue. If *RDR6* is only expressed in

abaxial tissue, one tempting explanation would be that the cleaved *TAS3* mRNA might move between abaxial and adaxial tissues.

In addition, players involved in the miR165/166-mediated post-transcriptional regulation in AB/AD cell fate determination were also found differentially expressed in FIL-labeled cells (Figure 3.24). It have reported that the expression of HD-ZIP III family genes are restricted at the adaxial side of the leaf, by miR165/166 which accumulates on the abaxial side of the leaf (Mcconnell and Barton, 1998; Mcconnell et al., 2001; Juarez et al., 2004). Although we still detected the expression of HD-ZIP III family genes in FIL-labeled cells, one possible explanation is that we were analyzing RNA from nuclei, which contains precursor-RNA at different stages. On the other hand, *REV* which is a member of HD-ZIP III family was indeed significantly down-regulated in FIL-labeled cells (Figure 3.24). *ZPR3* which is directly induced by REV (Brandt et al., 2013; Wenkel et al., 2007), was only expressed in non-FIL labeled cells (UN) (Figure 3.24).

Furthermore, *HEN1* and *HESO* were exclusively expressed in FIL-labeled cells (FIL) (Figure 3.24). Studies have shown that HEN1 plays a role in methylation of 3' terminal ribose to stabilize miRNA duplex (Jover-Gil et al., 2012) and HESO is associated with AGO1 to facilitate polyuridylation for mRNA degradation (Yu et al., 2005; Li et al., 2005; Tang et al., 2003). Therefore, HEN1 and HESO maybe two crucial players to maintain the expression pattern of miR165/166 and HD-ZIP III family genes. The abaxial expressing HEN1 can protect the stability of miR165/166, while the abaxial expressing HESO facilitates the cleavage of transcripts of HD-ZIP III genes in the miR165/166 directed PTGS pathway.

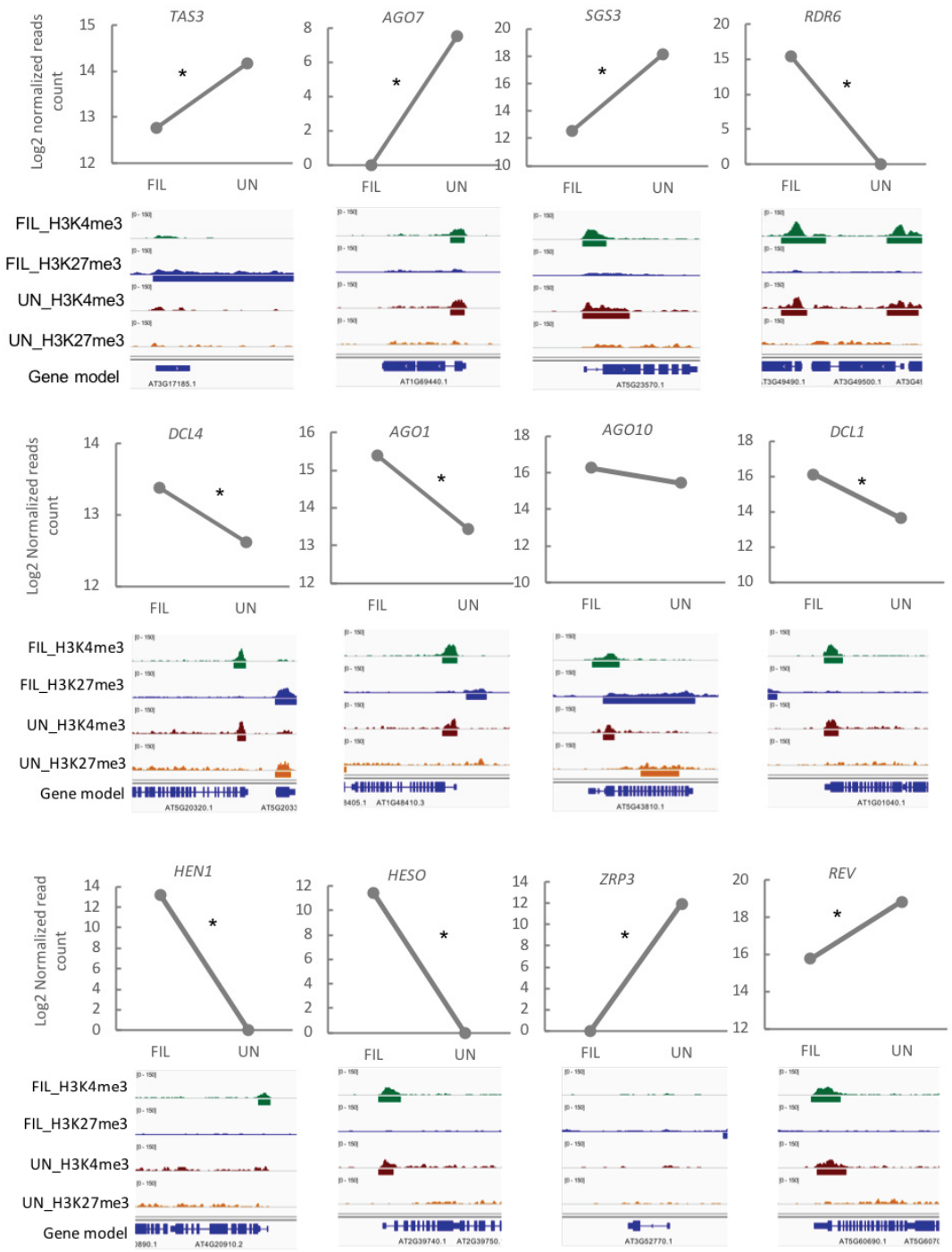


Figure 3.24. Gene expression and iGV browser view of H3K4me3 and H3K27me3 on genes involved in the abaxial cell fate determination pathway. * marked the gene expression or H3K4me3 or H3K27me3 is significantly different between FIL-labeled cells and non-FIL labeled cells.

4. Discussion

Plants develop distinct cell fates and tissue structures to adapt to various exogenous cues and to maximize their sensitivity towards a dynamically changing environment. However, the mechanisms in cell fate determination are still largely unknown. Cell-type specific analysis, together with high-throughput sequencing, provides high-resolution datasets to explore the regulatory network underlying cell fate determination. In this study, we generated cell-type specific transcriptome, and H3K4me3 and H3K27me3 epigenome profiles of CLV3-labeled cells and FIL-labeled cells (abaxial cells) in Arabidopsis shoot apex and investigated potential regulations in cell fate determination.

4.1 INTACT efficiency is variable on different tissues

Since the INTACT method has been reported (Deal and Henikoff, 2011), several studies have reported the isolation of different tissues, including root hair nuclei (GL2::NTF), non-root hair tissue (ARF8::NTF), endosperm tissue (Moreno-romero et al., 2016), the shoot apical meristem (You et al., 2017). In this study, we adopted INTACT to isolate nuclei from CLV3::NTF, which labeled a few dozens of nuclei in the vegetative shoot meristem, and FIL::NTF which labeled abaxial side of leaves, hence marking hundreds of cells in the vegetative shoot apex. Therefore, from root to vegetative tissue to reproductive tissue, INTACT can be used in various tissues in Arabidopsis. However, the yield is low, and a high contamination rate can be observed when working with a small target cell population. In our hands, INTACT isolated CLV3-labeled nuclei contained 40%-50% of contamination, whereas less than 10% of contamination was observed in FIL-labeled nuclei. As an essential quality control in INTACT, counting nuclei cannot quantify the contamination accurately, since the INTACT procedure also isolates unspecific bound nuclei. Hence, in this study, inspired by Jordi et al. (Moreno-romero et al., 2016), we quantified contamination by evaluating the enrichment of DNA from *Ler* after INTACT, in which both *Ler* plants and INTACT lines (Col) were mixed as starting material. But the disadvantage of this quantification method is that one has to bring the additional *Ler* in the sample. Therefore, this method can only be used for setup INTACT.

Besides the high contamination rate in CLV3-labeled cells, we have found that 80% of the genes showed no expression in the CLV3 profiles. This generated bias in differential expression analysis when conducting the reads counts normalization. And we have not found differential enrichment of H3K4me3 or H3K27me3 on meristematic markers genes, including *CLV3, STM, WUS*. Therefore, in this study, we focused on the profiles in FIL-labeled cells and non-FIL labeled cells.

4.2 Correlation between histone modifications and transcription

The histone modifications-H3K4me3 and H3K27me3 are associated with gene expression, marking active or repressive chromatin states, respectively. The association of these marks with their respective chromatin state can also be observed in this study: most of the genes enriched in H3K4me3 showed relatively high expression, whereas genes enriched in H3K27me3 were either not expressed or only at low levels (Figure 3.8 C). This correlation was also valid when comparing dynamic changes of expression and histone modifications between different cell types. Especially prominent for key regulators involved in abaxial cell specification, which had shown hyper-H3K4me3 coupled with up-regulation, or hyper-H3K27me3 with down-regulation in FIL-labeled cells (Figure 3.21). However, these associations were not consistent on all genes: Some genes were not expressed despite being enriched for H3K4me3, whereas some of the H3K27me3 target genes were highly expressed (Figure 3.8 C). And the correlation became minor when we quantified the changes in histone modification and transcription.

For H3K4me3, we found that around 40% of hyper-H3K4me3 target genes showed increased expression, but H3K4me3 fold changes are mainly in the range of 2-5 while transcriptional fold changes can be in the range of hundreds (Figure 3.9 A and D). These results are similar to the observation made by Howe et al.'s study (Howe et al., 2017). One potential explanation is that the quantity of histone modification and transcription are dependent on different factors. The enrichment of H3K4me3 is dependent on the target gene size and the number of cells that comprise the respective H3K4me3 state, while gene expression is dependent on the speed of transcription and degradation of transcripts. Therefore, the comparison of H3K4me3 enrichment on the same gene between FIL-labeled

and non-FIL labeled cells likely reflects the difference in the number of cells comprising H3K4me3 on this gene. The expression of the same gene in different cells can show striking differences. More than 50% of the genes showed an increased level of H3K4me3, but their expression showed no significant change or even decreased (Figure 3.9 A). Although one may argue that the mixed cell-types from non-FIL labeled cells can cause this result by diluting signals or adding noise to the comparison, it is also plausible that H3K4me3 itself is not instructive to transcription, but instead is rather a byproduct of transcription. Studies in yeast, mammals, and *Arabidopsis* have shown plenty of evidence that loss of H3K4me3 does not necessarily cause a reduction in transcription (Margaritis et al., 2012; Lenstra et al., 2011; Clouaire et al., 2012; Ding et al., 2011). In yeast, the *spp1* mutant has caused the reduction of H3K4me3 without alteration of gene expression (Margaritis et al., 2012; Lenstra et al., 2011). In *Arabidopsis*, one of the H3K4me3 methyltransferases, ATX1 shows independent dual functions: it can recruit Pol II at promoters to initiate transcription, and it can establish the H3K4me3 mark at +300bp region of an actively expressed gene (Ding et al., 2011). Blocking of the phosphorylated form of Pol II leads to reduced H3K4me3 and reduced recruitment of ATX1 at transcription region, but shows little impact on gene expression and enrichment of Pol II and ATX1 at the promoter (Ding et al., 2011). In fact, studies on dynamic regulation of meiotic genes in yeast have shown that the H3K4me3 mark is only “written” on target genes after the upregulation of transcription, at least 2hours later (Robine et al., 2009). Nevertheless, we have to admit that the relationship between transcription and H3K4me3 establishment is still under debate, and the current data is puzzling. For example, using the epigenome-editing technique, induced H3K4me3 is sufficient to reactivate transcription of the target gene, unless the region is on the DNA methylation state (Cano-rodriguez et al., 2016). Overall, considering our data together with the results from other studies, it is still too early to define the role of H3K4me3 in transcription activation, which probably is more complicated than instructive/causal. It may differ between a transient system and a developing organism, and vary among other contexts of epigenetic modifications.

For H3K27me3, our data showed an imperfect association with transcription on a quantitative level as well. More than 70% of genes showed no significant change in expression (silenced or low expression) upon differential enrichment of H3K27me3 between

FIL-labeled and non-FIL labeled cells (Figure 3.9 A). Although we have found that genes with expanded H3K27me3 coverage in their upstream or downstream region show a higher probability of being silenced than those only enriched in H3K27me3 in the gene body (Figure 3.15), this cannot be generalized for all targets. Weak quantitative correlation between H3K27me3 and transcription has also been shown in other studies. In mammalian cells, even though demethylation of H3K27me3 can lead to gene activation (Agger et al., 2007; Lee et al., 2007b), but there are genes can be activated with the presence of H3K27me3 (Kang et al., 2015; Prickaerts et al., 2012). In *Arabidopsis*, the H3K27me3 profile comparison among the shoot apical meristem from different stages has also shown that the correlation of H3K27me3 quantity and expression changes are not consistent with all genes during the transition of vegetative to reproductive (You et al., 2017). Taken together, although H3K27me3 is a key switch to control the transcription state, since H3K27me3 targets are mainly lowly or not expressed at all, its quantity is not a predictive bar to transcription.

In addition, we have found many genes that are differentially expressed, but were neither target of H3K4me3 nor H3K27me3, indicating the involvement of other components. For example, *MP* is dynamically regulated between FIL-labeled and non-FIL labeled cells (Figure 3.21), and is only targeted by H3K4me1 at a region between its 5th exon and 10th exon according to epigenome profiles in seedlings (Zhang et al., 2009, 2007b). It is not clear if H3K4me1 is differentially enriched between FIL-labeled and non-FIL labeled cells, and if H3K4me1 also contributes to the down-regulation of *MP*. Nevertheless, it is suggested that a full picture of epigenetic modification profiles may fulfill the imperfect correlation between transcription and H3K4me3/H3K27me3.

4.3 Distribution pattern of H3K4me3 may provide additional information

Among eukaryotic organisms, H3K4me3 is always known as a narrow domain mark, which is enriched at the transcriptional start site of target genes. With the explosion of epigenome data and detailed analysis on H3K4me3 profiles, studies have revealed various H3K4me3 distributions and proposed potential functions in gene regulation (Benayoun et al., 2014; Dincer et al., 2015; Chen et al., 2015; Pekowska et al., 2011; Pundhir et al., 2016). In this

study, we found that nearly 2000 genes were covered by expanded H3K4me3 domains which peaked in the middle of the gene-body (Figure 3.8 A). Among those broad H3K4me3 targets, photosynthesis, and stress responsive genes were significantly enriched (Figure 3.10 A). Several studies in mammalian and human cells have shown that broad H3K4me3 domains mark key cell identity gene in neuron cells (Dincer et al., 2015; Benayoun et al., 2014), as well as tumor suppressor genes in human T cells (Chen et al., 2015). However, we found that the association between cell identity and broad H3K4me3 domains is trivial when comparing FIL-labeled cells to published INTACT isolated root tissues (GL2-labeled and ADF8-labeled cells) (Deal and Henikoff, 2010) (Figure 3.10). None of the root related functional GO terms were found in the broad domain in GL2-labeled and ADF8-labeled cells (Figure 3.10 C). This challenged the link between cell identity and broad H3K4me3 domains. Other than that, studies suggested that broad domains may have an important role in transcriptional regulation. In human neurons, a study has shown that the broad H3K4me3 domain is preferably associated with transcriptional consistency than transcription levels. Although we lack single-cell epigenome and transcriptome data to test this hypothesis, the enrichment of photosynthesis-related genes in the broad H3K4me3 domain category is a good indication since photosynthesis genes are among the most conserved and consistently expressed genes in leaf tissues. In addition, a study in T cells has shown that the broad H3K4me3 domain is associated with increased H3K79me2, as well as an increased ratio of gene-body localized Pol II, suggesting that broad H3K4me3 might be associated with transcription elongation (Chen et al., 2015). We have observed that some of the dehydration-responsive genes comprise an expanded H3K4me3 domain under drought stress, while the peak is still at TSS (Figure 3.12). This suggests that the expanded H3K4me3 domains which were narrow previously are different from the broad domains in cluster3 which peak in the middle of the gene-body (Figure 3.12). Taken together, the broad H3K4me3 domain likely marks a different state of transcription compared to the narrow H3K4me3 domain. However, we need more data to verify this association, such as profiles of Pol II, H3K36me3, DNase sensitive sites, and DNA methylation profiles.

Furthermore, we also identified a few hundred intergenic H3K4me3 domains (Figure 3.13 and 3.14). In human cells, H3K4me3 is enriched at upstream of the TSS between the bidirectional gene pairs, marking bidirectional promoters (Bornelöv et al., 2015). However,

among the intergenic regions that are targeted by H3K4me3, we found a reduced percentage of intergenic regions between bidirectional genes (Figure 3.14 B). In addition, genes adjacent to intergenic H3K4me3 domains showed a lower expression level with larger variation compared to TSS enriched H3K4me3 (Figure 3.13 B and C), suggesting that those H3K4me3 marked intergenic regions probably are not cis-regulatory elements. Interestingly, intergenic H3K4me3 domains showed a higher percentage of GC content and were significantly enriched in an LBD motif, CGGCGR (Figure 3.14 D, E, and F). Considering that there was a higher ratio of DMRs in intergenic H3K4me3 domains than in the gene-body and promoter H3K4me3 domains, intergenic H3K4me3 domains may mark certain regulatory features and are dynamically regulated between different cell types. In fact, we identified 55 CGGCGR-containing intergenic regions that showed a hyper-methylation of H3K4me3 in FIL-labeled cells. In human cells, H3K4me3 is indeed associated with active enhancers, whereas the enhancer marker H3K4me1 does not provide information on enhancer activity (Pekowska et al., 2011). Nevertheless, to confirm that those intergenic H3K4me3 domains are indeed functional regulatory elements, we would need epigenome profiles of H3K4me1 and H3K27ac, which are enhancer markers, as well as 3C-seq to capture the genomic interactions.

Taken together, we have identified various H3K4me3 distributions, including TSS located narrow domains which show a “Peak-valley-peak” pattern, broad H3K4me3 domains which peaked in the middle of gene-body, as well as intergenic located domains. The detailed localization of H3K4me3 may provide additional information on a gene’s transcription status and marks specific regulatory features.

4.4 Dynamic regulation network in abaxial cell fate determination

In plants, the abaxial and adaxial of leaf are two distinct tissues, whose development is closely regulated on a transcriptional and a post-transcriptional level. In our study, we analyzed the transcriptome and H3K4me3/H3K27me3 profiles from FIL-labeled cells, which are enriched in abaxial cells, and non-FIL labeled cells, which are a mixture of other aerial tissues besides abaxial cells. Our data confirmed most of the key regulators in abaxial cell determination (Figure 3.22, 3.24 and 4.1). *YAB* and *KAN* gene families, which are key abaxial

determinants, were found up-regulated with hyper-methylation of H3K4me3 in FIL-labeled cells (Figure 3.22). Meanwhile, YAB and KAN gene families were also enriched in H3K27me3, especially YAB5 showed hyper-methylation of H3K27me3 in FIL-labeled cells (Figure 3.22), suggesting that the presence of H3K4me3 can overtake the repressive H3K27me3 regarding gene expression. Although we lack data to show that H3K4me3 and H3K27me3 were indeed enriched on the same genomic fragment, bivalent marks, i.e., H3K4me3 and H3K27me3 present at the same nucleosome, have been widely found in mammalian, human cells and *Arabidopsis* (Bernstein et al., 2006; Grandy et al., 2016; Cui et al., 2009). One possible explanation is that H3K27me3 is the chromatin memory passed on to differentiated abaxial cells from meristem cells, in which YAB and KAN genes were silenced (Stahle et al., 2009). It is worth mentioning that YAB5 was not targeted by H3K4me3 in non-FIL labeled cells while being expressed in both cell types (Figure 3.22), further suggesting that H3K4me3 is not instructive to transcription. ARF3/4 are other important abaxially expressed regulators, which are directly suppressed by AS2 and downregulated on the post-transcription level via *tasiARF* (Husbands et al., 2015). As expected, we detected upregulation of *ARF3* together with hyper-methylation of H3K4me3, while *AS2* is absent and hypermethylated by H3K27me3 in FIL-labeled cells (Figure 3.22). Furthermore, components involved in *tasiARF* biogenesis, such as *AGO7*, *TAS3* and *SGS3*, were down-regulated in FIL-labeled cells (Figure 3.23). However, RDR6, which is involved in the dsRNA formation during tasiRNA biogenesis, showed increased expression in FIL-labeled cells. The unexpected upregulation of *RDR6* in FIL-labeled cells is in contrast to the downward curling leaf phenotype in *rdr6*, which is a common phenotype among other tasi-mutants, such as *ago7-1, dcl4-2, sgs3-1, tas3-1,* and 35S::*ARF3* (Adenot et al., 2006). Despite the inverted RDR6 expression pattern, our data further confirmed an opposing gradient of *tasiARF* and ARF3 in abaxial and adaxial leaf tissues (Figure 4.3).

In contrast, the adaxial determinants HD-ZIP III genes (PHB, PHV, and REV) are restricted by KAN family genes and miR165/166 on the post-transcriptional level (Reinhart et al., 2013). However, we found expression of all HD-ZIP III genes in FIL-labeled cells, although *REV* was down-regulated. Because HD-ZIP III genes are regulated on the post-transcriptional level, this could be a possible explanation of why we could detect their expression. In fact, *ZPR3,* which is directly induced by REV (Brandt et al., 2013; Wenkel et al., 2007), showed exclusive

expression in non-FIL labeled cells, indicating that functional REV protein is absent in FIL-labeled cells (Figure 3.24). In addition, although we lack microRNA data to analyze miR165/166 expression profiles, the components involved in the miRNA silencing pathway showed distinct spatial expression patterns (Figure 3.24). HEN1 is a key regulator to stabilize the miRNA duplex against uridylation by catalyzing the methylation of 3'terminal ribose (Ren et al., 2014; Jover-Gil et al., 2012), and showed exclusive expression in FIL-labeled cells. Another regulator, HESO, is recruited by AGO1 to facilitate cleavage of target mRNA by adding polyuridyl tail (Ren et al., 2014; Zhao et al., 2012), and also showed exclusive expression in FIL-labeled cells. Therefore, it is likely that the presence of HEN1 at the abaxial side protects the miR165/166 from HESO, which can uridylate 3' terminal, while the adaxial side would not have efficient cleavage of HD-ZIP III genes without HESO to add a degradation signal to the cleaved mRNA. Together, the spatial expression of HEN1 and HESO probably are two important factors to restrict HD-ZIP III genes at the abaxial side by modulating the miR165/166 post-transcriptional silencing pathway.

Furthermore, several auxin related regulators were found to be differentially expressed between FIL-labeled and non-FIL labeled cells, corresponding to the auxin gradient model in abaxial/adaxial cell fate determination. MP/ARF5 is known to be down-regulated by the KAN family genes (Krogan and Berleth, 2012), which showed a hyperexpression in the adaxial side (Figure 3.22). However, MP protein can only become activated in the presence of auxin, which is depleted on the adaxial side of the leaf. Meanwhile, *PIN1* can be induced by MP(Wenzel et al., 2007), which also showed downregulation in FIL-labeled cells (Figure 3.22). Besides this, Aux1, the auxin influx carrier which plays an important role in xylem differentiation (Fàbregas et al., 2015), was also down-regulated in FIL-labeled cells (Figure 3.23). This dynamic regulation fits in the current auxin gradient model that argues that a maximum auxin signal is in the abaxial and adaxial boundary (Bhatia et al., 2016; Guan et al., 2017; Qi et al., 2014).

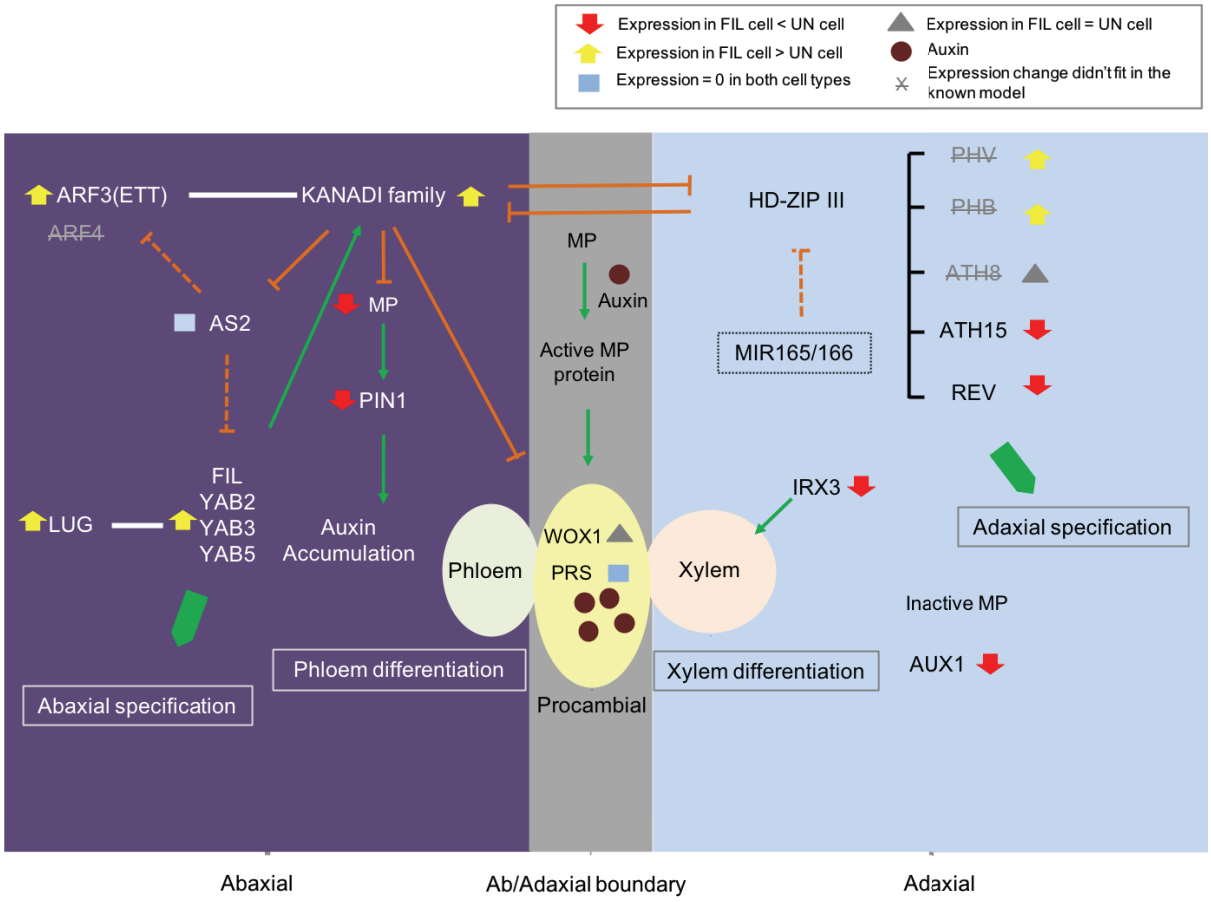


Figure 4.1 Dynamic regulation of key determinants involved in abaxial and adaxial specification.

4.5 dsRNA may be transported from the abaxial side of leaves to other tissues

In this study, when comparing the differential expressed genes between FIL-labeled cells and non-FIL labeled cells, we found that plenty of genes involved in RdDM pathway were differentially expressed, e.g. *NRPD1* (subunits of Pol IV), *RDR2* and *RDR6* were exclusively expressed in FIL-labeled cells (Table 3.4). NRPD1 and RDR2 are key components in the biosynthesis of dsRNA (26-45nt), whereas RDR6 plays a role in the production of 21-22nt dsRNA. However, cross comparing the derepressed loci of *nrpd1* (Yang et al., 2017) and DEGs in this study, we found that most of the NRPD1 targets were stably regulated in FIL-labeled and non-FIL labeled cells (Figure 3.19 A and B), suggesting that the exclusive expression pattern of *NRPD1*, *RDR2,* and *RDR6* in FIL-labeled cells has little impact on target gene expression. Meanwhile, by comparing transcriptome data from this study to the RDR2/NRPD1 dependent dsRNA (P4RNA) (Li et al., 2015), we found that total transcript reads count on all P4RNA was significantly increased in FIL-labeled cells, although the

majority of the P4RNA loci showed no expression in both tissues (Figure 3.19). Therefore, we speculate that dsRNA is spatially produced in FIL-labeled cells and is potentially transported to other cells to establish *de novo* DNA methylation. Actually, mobility of RNA is not only a concept, many studies have documented mobile RNA signaling to trigger gene silencing via cell to cell transport or long-distance transport in plants. Examples include tasiARF signaling in abaxial and adaxial leaf tissue, and the potential mobility of sRNA or sRNA precursors between vegetative and sperm cells in pollen (Chitwood et al., 2009; Schwab et al., 2009; Slotkin et al., 2009). In *C.elegans*, it was shown that dsRNA can be transported from neurons to the germline to spread transgenerational silencing (Devanapally et al., 2015). Although in plant there is no homologous protein to the transmembrane protein SID1, which is involved in the dsRNA transportation in *C.elegans*, data has shown that *nrpd1* and *rdr2* mutations affect cell-to-cell RNA silencing but not *ago4* and *dcl3* (Dunoyer et al., 2007; Smith et al., 2007). This strongly suggests that dsRNA may move from abaxial (phloem located on the abaxial side) to other tissues to establish RdDM mediated DNA methylation (Figure 4.3).

4.6 MORC6 may be a mobile protein, transmitting between tissues

In this study, when comparing the differential expressed genes between FIL-labeled cells and non-FIL labeled cells, we also found that MORC6 was also exclusively expressed in FIL-labeled cells (Figure 3.21, Table 3.4). Several studies have indicated that MORC6 is involved in RdDM and regulate chromatin condensation, although the mechanism is still unclear. However, among the published MORC6 target loci (Liu et al., 2016), we found the absent of MORC6 in non-FIL labeled cells did not alter the expression profile of respective genes (Figure 3.21 B and C). Nevertheless, a study has found that abaxial epidermis in *morc6* mutants are indeed less effective in de-repressing the previously silenced GFP, comparing to the adaxial epidermis (Brabbs et al., 2013). Considering that MORC6 can be purified from plasmodesmata (Fernandez-Calvino et al., 2011), we speculate that MORC6 may be a mobile protein that transported from abaxial to other tissues to affect RdDM directed gene silencing. By assuming MORC6 is a mobile protein, we can perfectly explain the *morc6* phenotype described in Brabb et al. (Brabbs et al., 2013) (Figure 4.2). In the dual transgene silencing systems, the WT shows no GFP signal while *morc6* EMS mutants partially

recovered GFP signals, in which *morc6-7* is a strong mutant that has SNP on ATPase domain and *morc6-5* is a weak allele which has SNP apart from ATPase motif (Brabbs et al., 2013). Because *morc6-7* still retains 3 out of 4 ATPase motifs, both mutants still comprise functional core ATPase domains to a certain degree. Thus, *morc6-5* and *morc6-7* probably will generate the less efficient form of MORC6 (Figure 4.2). In WT conditions, the WT version of MORC6 move from abaxial to adaxial tissues to establish transcriptional silencing (TGS) on *GFP*, showing no GFP signal in both abaxial and adaxial tissues (Figure 4.2). In *morc6-5* and *morc6-7* mutants, the it would require more of the defective version of *MORC6* to establish TGS on *GFP* in abaxial tissues. Thus, it would transport less *MORC6* to the adaxial side, leading to differential efficiency in establishing TGS in the adaxial and abaxial epidermal in *morc6* (Figure 4.2). Therefore, *morc6* plants have exhibited a slower and less effective establishment of TGS on GFP in adaxial side, showing stronger GFP signal (Brabbs et al., 2013). We still lack definitive evidences to prove the mobility of MORC6, therefore future study of its intraflagellar transport complex B motif and binding domain to plasmodesmata is necessary.

Altogether, based on our data and previous studies, we found several potential mobile signals in *Arabidopsis* (Figure 4.3). However, this is all based on transcriptome data. In the future, grafting experiments and the actual tracing of the mobility of respective molecules via the high-resolution microscopy are necessary to confirm our hypothesis.

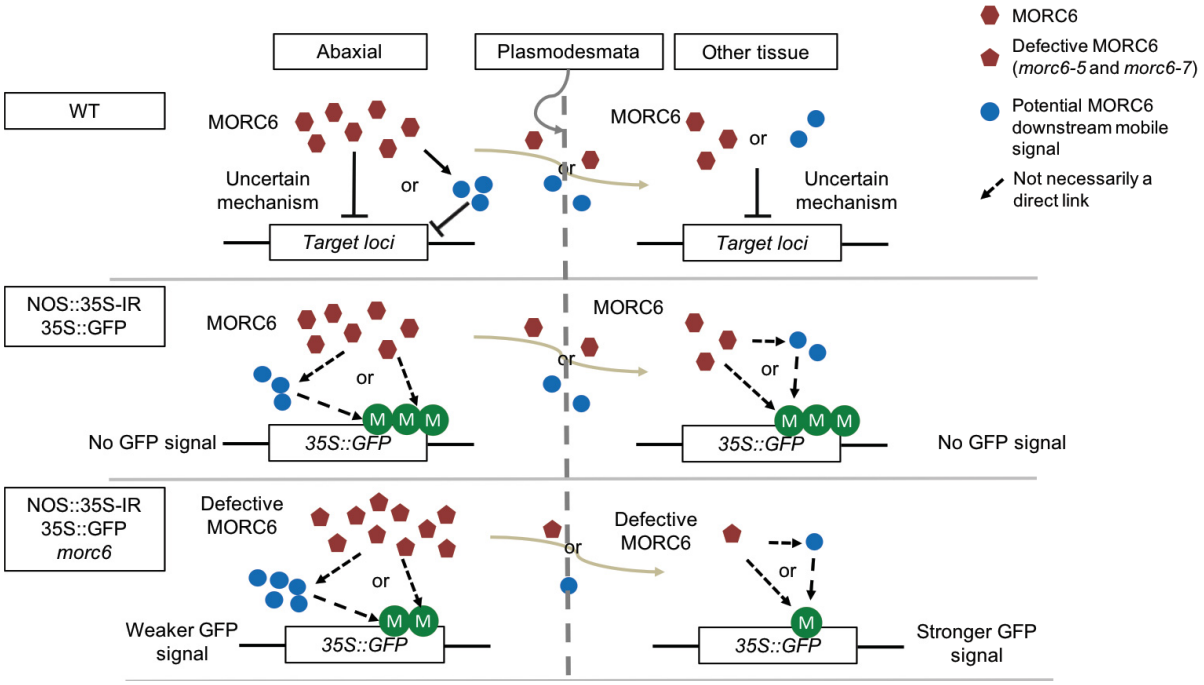


Figure 4.2 Potential model for MORC6 movement between abaxial cells and adaxial cells to spread TGS.

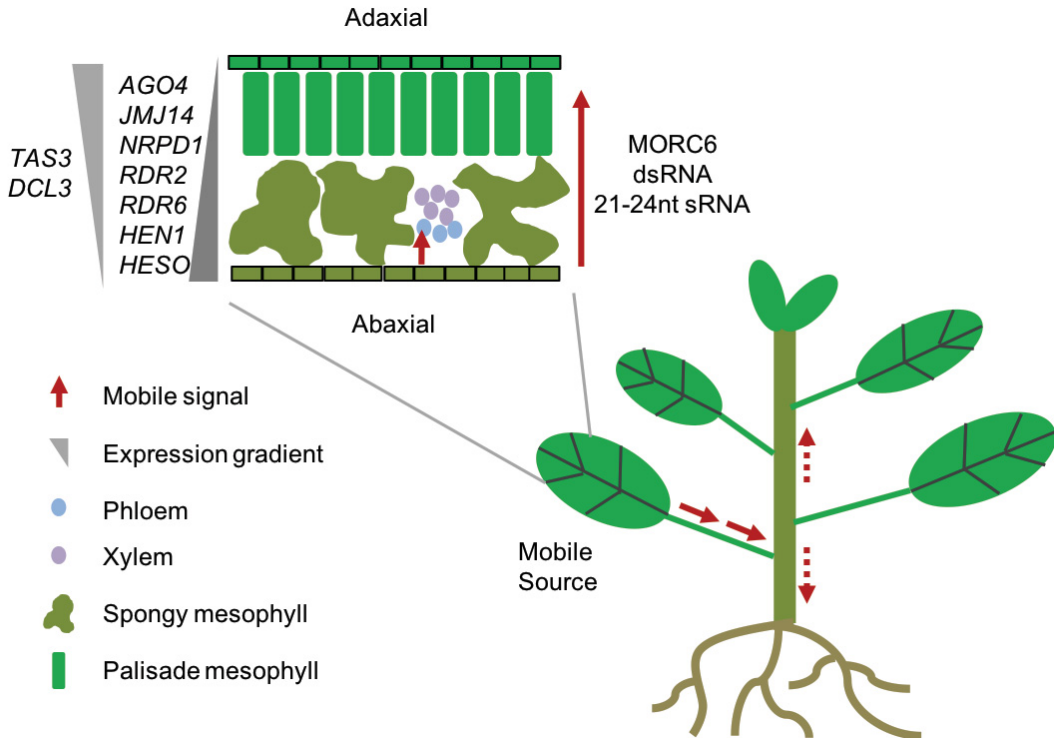


Figure 4.3 Possible mobile signals between abaxial cells and adaxial cells.

5. Summary

As multicellular organisms, plants harbor various cell types that are all derived from the undifferentiated stem cells in different stem cell niches. Distinct cell types acquire cell identity via transcription and post-transcription regulations. Therefore, profiling the transcriptome and epigenome of specific cell types is essential for characterizing the mechanisms of cell fate determination and cell-type specific responses.

In this study, we adopted the INTACT (Isolation of Nuclei Tagged in specific Cell Types) technique to isolate stem cell populations and abaxial cells of the *Arabidopsis* shoot apex. Together with high-throughput sequencing, we generated transcriptome and H3K4me3 and H3K27me3 genome-wide profiles in stem cells, abaxial cells, and other tissues apart from abaxial cells. Although the INTACT method was not suitable to elucidate stem cell transcriptomic and epigenomic patterns, we observed dynamic of transcription as well as H3K4me3 and H3K27me3 between abaxial cells and other non-abaxial tissues. Not only we confirmed the known regulatory network underlying the abaxial cell specification, but we also found other candidate genes that may restrict the spatial expression pattern of abaxial/adaxial determinants. In addition, we observed that the regulators of double-stranded RNA biogenesis were exclusively expressed in abaxial cells, suggesting that dsRNA might be generated in the abaxial side and move to other tissues. Apart from analyzing gene regulatory networks underlying abaxial cell fate determination, we also took advantage of the high level of homogeneous cells among INTACT-isolated abaxial cells and analyzed the detailed distribution of chromatin state (H3K4me3 and H3K27me3). Besides the enrichment of H3K4me3 at gene coding regions, we identified hundreds of intergenic regions were also enriched in H3K4me3, whereas an LBD (LOB DOMAIN) motif is significantly enriched. Furthermore, we scrutinized the broad H3K4me3 domains which were previously shown to be linked with certain cell fate identity in mammalian cells. Despite the enrichment of photosynthesis and response genes within broad H3K4me3 targets, we found no clear indication of the association of cell identity and broad H3K4me3 in this study.
Overall, this study generated high-resolution transcriptome and H3K4me3 and H3K27me3 profiles of abaxial cells and revealed dynamic regulations on abaxial cell identity.

6. References

Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouché, N., Gasciolli, V., and Vaucheret, H. (2006). DRB4-Dependent TAS3 trans-Acting siRNAs Control Leaf Morphology through AGO7. Curr. Biol. **16**: 927–932.

Adrian, J., Chang, J., Birnbaum, K.D., Bergmann, D.C., Adrian, J., Chang, J., Ballenger, C.E., Bargmann, B.O.R., and Alassimone, J. (2015). Transcriptome Dynamics of the Stomatal Lineage : Birth , Amplification , and Termination of a Self- Resource. Dev. Cell **33**: 107–118.

Agger, K., Cloos, P.A.C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). LETTERS UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature **449**: 731–735.

Allegra, P., Sterner, R., Clayton, D.F., and Allfrey, V.G. (1987). Affinity chromatographic purification of nucleosomes containing transcriptionally active DNA sequences. J. Mol. Biol. **196**: 379–388.

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell **121**: 207–21.

Alvarez-Venegas, R. and Avramova, Z. (2005). Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive Arabidopsis genes and in atx1 mutants. Nucleic Acids Res. **33**: 5199–5207.

Alvarez-Venegas, R. and Avramova, Z. (2002). SET-domain proteins of the Su(var)3-9, E(z) and Trithorax families. Gene **285**: 25–37.

Ashburner, M. et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. **25**: 25–29.

Van Attikum, H., Fritsch, O., Hohn, B., and Gasser, S.M. (2004). Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair. Cell **119**: 777–788.

Bai, S., Kasai, A., Yamada, K., Li, T., and Harada, T. (2011). A mobile signal transported over a long distance induces systemic transcriptional gene silencing in a grafted partner. J. Exp. Bot. **62**: 4561–4570.

Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R.,

Schulz, I., Reuter, G., and Aalen, R.B. (2001). The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res. **29**: 4319–4333.

Bell, E.M., Lin, W. -c., Husbands, A.Y., Yu, L., Jaganatha, V., Jablonska, B., Mangeon, A., Neff, M.M., Girke, T., and Springer, P.S. (2012). Arabidopsis LATERAL ORGAN BOUNDARIES negatively regulates brassinosteroid accumulation to limit growth in organ boundaries. Proc. Natl. Acad. Sci. **109**: 21146–21151.

Belles-boix, E., Hamant, O., Witiak, S.M., Morin, H., and Traas, J. (2006). KNAT6 : An Arabidopsis Homeobox Gene Involved in Meristem Activity and Organ Separation. Plant Cell **18**: 1900–1907.

Benayoun, B.A. et al. (2014). H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell **158**: 673–688.

Bernstein, B.E. et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell **125**: 315–26.

Bhatia, N., Bozorg, B., Larsson, A., Ohno, C., Jönsson, H., and Heisler, M.G. (2016). Auxin Acts through MONOPTEROS to Regulate Plant Cell Polarity and Pattern Phyllotaxis. Curr. Biol. **26**: 3202–3208.

Bies-etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., and Cooke, R. (2009). RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. EMBO Rep. **10**: 649–654.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science **302**: 1956–60.

Blackledge, N.P. et al. (2014). Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. Cell **157**: 1445–59.

Blevins, T., Podicheti, R., Mishra, V., Marasco, M., Wang, J., Rusch, D., Tang, H., and Pikaard, C.S. (2015). Identification of Pol IV and RDR2- dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. Elife **2**: 1–22.

Bornelöv, S., Komorowski, J., and Wadelius, C. (2015). Different distribution of histone modifications in genes with unidirectional and bidirectional transcription and a role of CTCF and cohesin in directing transcription. BMC Genomics **16**: 1–13.

Boutilier, K., Offringa, R., Sharma, V.K., Kieft, H., Ouellet, T., Zhang, L., Hattori, J., Liu, C.-

M., van Lammeren, A.A.M., Miki, B.L.A., Custers, J.B.M., and van Lookeren Campagne, M.M. (2002). Ectopic Expression of BABY BOOM Triggers a Conversion from Vegetative to Embryonic Growth. Plant Cell **14**: 1737 LP – 1749.

Brabbs, T.R., He, Z., Hogg, K., Kamenski, A., Li, Y., Paszkiewicz, K.H., Moore, K.A., O'Toole, P., Graham, I.A., and Jones, L. (2013). The stochastic silencing phenotype of Arabidopsis morc6 mutants reveals a role in efficient RNA -directed DNA methylation. Plant J. **75**: 836–846.

Brand, U., Fletcher, J.C., Hobe, M., Meyerowitz, E.M., and Simon, R. (2000). Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity. Science **289**: 617–619.

Brandt, R., Xie, Y., Musielak, T., Graeff, M., Stierhof, Y.D., Huang, H., Liu, C.M., and Wenkel, S. (2013). Control of stem cell homeostasis via interlocking microRNA and microProtein feedback loops. Mech. Dev. **130**: 25–33.

Bratzel, F., López-Torrejón, G., Koch, M., Del Pozo, J.C., and Calonje, M. (2010). Keeping cell identity in arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. Curr. Biol. **20**: 1853–1859.

Bratzel, F., Yang, C., Angelova, A., López-Torrejón, G., Koch, M., Del Pozo, J.C., and Calonje, M. (2012). Regulation of the new arabidopsis imprinted gene AtBMI1 requires the interplay of different epigenetic mechanisms. Mol. Plant **5**: 260–269.

Briggs, S.D., Strahl, B.D., Cheung, W.L., Bryk, M., Winston, F., Davie, J.K., and Dent, S.Y.R. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes Dev. **15**: 3286–3295.

Brink, R.A. (1956). A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. Genetics **41**: 872–89.

Brinkman, A.B., Gu, H., Bartels, S.J.J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., and Stunnenberg, H.G. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. **22**: 1128–1138.

Byrne, M.E., Simorowski, J., and Martienssen, R.A. (2002). ASYMMETRIC LEAVES1 reveals know gene redundancy in Arabidopsis. Development **129**: 1957–1965.

Cano-rodriguez, D., Gjaltema, R.A.F., Jilderda, L.J., Jellema, P., Dokter-fokkens, J., Ruiters, M.H.J., and Rots, M.G. (2016). Writing of H3K4Me3 overcomes epigenetic silencing in a

sustained but context-dependent manner. Nat. Commun. **7**: 1–11.

Carbon, S. et al. (2019). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. **47**: D330–D338.

Carter, A.D., Bonyadi, R., and Gifford, M.L. (2013). The use of fluorescence-activated cell sorting in studying plant development and environmental responses. Int. J. Dev. Biol. **57**: 545–552.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science **303**: 1336.

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**: 5263–76.

Chen, K. et al. (2015). Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor suppressor genes Kaifu. Nat. Genet. **47**: 1149–1157.

Chen, L.-Q., Luo, J.-H., Cui, Z.-H., Xue, M., Wang, L., Zhang, X.-Y., Pawlowski, W.P., and He, Y. (2017). ATX3 , ATX4 , and ATX5 Encode Putative H3K4 Methyltransferases and Are Critical for Plant Development . Plant Physiol. **174**: 1795–1806.

Chitwood, D.H., Nogueira, F.T.S., Howell, M.D., Montgomery, T.A., Carrington, J.C., and Timmermans, M.C.P. (2009). Pattern formation via small RNA mobility. Genes Dev. **23**: 549–554.

Cho, Y.W., Hong, T., Hong, S.H., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G.R., Copeland, T.D., Kalkum, M., and Ge, K. (2007). PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. J. Biol. Chem. **282**: 20395–20406.

Clark, S., Running, M., and Meyerowitz, E. (1995). CLAVATA3 is a specific regulator of shoot and floral meristem development affecting the same processes as CLAVATA1. Development **121**: 2057–2067.

Clouaire, T., Webb, S., Skene, P., Illingworth, R., Kerr, A., Andrews, R., Lee, J., Skalnik, D., and Bird, A. (2012). Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. Genes Dev.: 1714–1728.

Coe, E.H. (1959). a Regular and Continuing Conversion-Type Phenomenon At the B Locus in Maize. Proc. Natl. Acad. Sci. **45**: 828–832.

Creyghton, M.P. et al. (2010). Histone H3K27ac separates active from poised enhancers and

predicts developmental state. Proc. Natl. Acad. Sci. **107**: 21931–21936.

Cui, K., Zang, C., Roh, T.Y., Schones, D.E., Childs, R.W., Peng, W., and Zhao, K. (2009). Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation. Cell Stem Cell **4**: 80–93.

Czechowski, T., Stitt, M., Altmann, T., and Udvardi, M.K. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization. Plant Physiol. **139**: 5–17.

Dai, X., Bai, Y., Zhao, L., Dou, X., Liu, Y., Wang, L., Li, Y., Li, W., Hui, Y., Huang, X., Wang, Z., and Qin, Y. (2017). H2A.Z Represses Gene Expression by Modulating Promoter Nucleosome Structure and Enhancer Histone Modifications in Arabidopsis. Mol. Plant **10**: 1274–1292.

Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-Wide Kinetics of Nucleosome Turnover Determined by Metabolic Labeling of Histones. Science **328**: 1161–1165.

Deal, R.B. and Henikoff, S. (2010). A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. Dev. Cell **18**: 1030–1040.

Deal, R.B. and Henikoff, S. (2011). The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat. Protoc. **6**: 56–68.

Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., and Jacobsen, S.E. (2012). Loss of the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target Genes and Redistribution of H3K27 Trimethylation to Transposons in Arabidopsis thaliana. PLoS Genet. **8**: e1003062.

Devanapally, S., Ravikumar, S., and Jose, A.M. (2015). Double-stranded RNA made in C . elegans neurons can enter the germline and cause transgenerational gene silencing. Proc. Natl. Acad. Sci. **112**: 2133–2138.

Dijk, K. Van, Ding, Y., Malkaram, S., Riethoven, J.M., Liu, R., Yang, J., Laczko, P., Chen, H., Xia, Y., Ladunga, I., Avramova, Z., and Fromm, M. (2010). Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in Arabidopsis thaliana. BMC Plant Biol. **10**: 238.

Dincer, A., Gavin, D.P., Xu, K., Zhang, B., Dudley, J.T., Schadt, E.E., and Akbarian, S. (2015). Deciphering H3K4me3 broad domains associated with gene-regulatory networks and conserved epigenomic landscapes in the human brain. Transl Psychiatry **5**: e679.

Ding, Y., Avramova, Z., and Fromm, M. (2011). Two Distinct Roles of ARABIDOPSIS

HOMOLOG OF TRITHORAX1 (ATX1) at Promoters and within Transcribed Regions of ATX1-Regulated Genes. Plant Cell **23**: 350–363.

Ding, Y., Ndamukong, I., Xu, Z., Lapko, H., Fromm, M., and Avramova, Z. (2012). ATX1-Generated H3K4me3 Is Required for Efficient Elongation of Transcription , Not Initiation , at ATX1- Regulated Genes. PLoS Genet. **8**: e1003111.

Dunoyer, P., Himber, C., Ruiz-ferrer, V., Alioua, A., and Voinnet, O. (2007). Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways. Nat. Genet. **39**: 848–856.

Eamens, A., Vaistij, F.E., and Jones, L. (2008). NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in Arabidopsis. Plant J. **55**: 596–606.

Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial Patterning of Arabidopsis Shoots by Class III HD-ZIP and KANADI Genes. Curr. Biol. **13**: 1768–1774.

Endo, H., Yamaguchi, M., Tamura, T., Nakano, Y., Nishikubo, N., Yoneda, A., Kato, K., Kubo, M., Kajita, S., Katayama, Y., Ohtani, M., and Demura, T. (2015). Multiple Classes of Transcription Factors Regulate the Expression of VASCULAR-RELATED NAC-DOMAIN7 , a Master Switch of Xylem Vessel Differentiation. Plant Cell Physiol. **7**: 242–254.

Eshed, Y., Baum, S.F., Perea, J. V, and Bowman, J.L. (2001). Establishment of polarity in lateral organs of plants. Curr. Biol. **11**: 1251–1260.

Eshed, Y., Izhaki, A., Baum, S.F., Floyd, S.K., and Bowman, J.L. (2004). Asymmetric leaf development and blade expansion in Arabidopsis are mediated by KANADI and YABBY activities. Development **131**: 2997–3006.

Fàbregas, N., Formosa-Jordan, P., Confraria, A., Siligato, R., Alonso, J.M., Swarup, R., Bennett, M.J., Mähönen, A.P., Caño-Delgado, A.I., and Ibañes, M. (2015). Auxin Influx Carriers Control Vascular Patterning and Xylem Differentiation in Arabidopsis thaliana. PLoS Genet. **11**: 1–26.

Fernandez-Calvino, L., Faulkner, C., Walshaw, J., Saalbach, G., Bayer, E., Benitez-Alfonso, Y., and Maule, A. (2011). Arabidopsis plasmodesmal proteome. PLoS One **6**: e18880.

Fletcher, J.C. (1999). Signaling of Cell Fate Decisions by CLAVATA3 in Arabidopsis Shoot Meristems. Science **283**: 1911–1914.

Fromm, M. and Avramova, Z. (2014). ATX1/AtCOMPASS and the H3K4me3 marks: How do

they activate arabidopsis genes? Curr. Opin. Plant Biol. **21**: 75–82.

Gao, Z. et al. (2010). LETTERS An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. Nature **465**: 106–109.

Gasciolli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. (2005). Partially Redundant Functions of Arabidopsis DICER-like Enzymes and a Role for DCL4 in Producing trans - Acting siRNAs. Curr. Biol. **15**: 1494–1500.

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.

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.

Grandy, R.A., Whitfield, T.W., Wu, H., Fitzgerald, M.P., VanOudenhove, J.J., Zaidi, S.K., Montecino, M.A., Lian, J.B., van Wijnen, A.J., Stein, J.L., and Stein, G.S. (2016). Genome-Wide Studies Reveal that H3K4me3 Modification in Bivalent Genes Is Dynamically Regulated during the Pluripotent Cell Cycle and Stabilized upon Differentiation. Mol. Cell. Biol. **36**: 615–627.

Grossniklaus, U. and Paro, R. (2014). Transcriptional Silencing by Polycomb -Group Proteins. Cold Spring Harb Perspect Biol **6**: a019331.

Grossniklaus, U., Vielle-Calzada, J.P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal control of embryogenesis by MEDEA, a Polycomb group gene in Arabidopsis. Science **280**: 446–450.

Guan, C., Wu, B., Yu, T., Wang, Q., Krogan, N.T., Liu, X., and Jiao, Y. (2017). Spatial Auxin Signaling Controls Leaf Flattening in Arabidopsis. Curr. Biol. **27**: 2940–2950.

Guitton, A.E., Page, D.R., Chambrier, P., Lionnet, C., Faure, J.E., Grossniklaus, U., and Berger, F. (2004). Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in Arabidopsis thaliana. Development **131**: 2971–2981.

Guttman, M. et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature **458**: 223–227.

Haag, J.R., Ream, T.S., Marasco, M., Nicora, C.D., Norbeck, A.D., Pasa-tolic, L., and Pikaard, C.S. (2012). Short Article In Vitro Transcription Activities of Pol IV , Pol V , and RDR2 Reveal Coupling of Pol IV and RDR2 for dsRNA Synthesis in Plant RNA Silencing. Mol.

Cell **48**: 811–818.

Halfmann, R. and Lindquist, S. (2010). Epigenetics in the Extreme: Prions and the Inheritance of Environmentally Acquired Traits. Science **330**: 629–632.

Havecker, E.R., Wallbridge, L.M., Hardcastle, T.J., Bush, M.S., Kelly, K.A., Dunn, R.M., Schwach, F., Doonan, J.H., and Baulcombe, D.C. (2010). The Arabidopsis RNA-Directed DNA Methylation Argonautes Functionally Diverge Based on Their Expression and Interaction with Target Loci. Plant Cell **22**: 321–334.

He, X., Hsu, Y., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H., Wang, C., Jin, H., and Zhu, J. (2009). An Effector of RNA-Directed DNA Methylation in Arabidopsis Is an ARGONAUTE 4- and RNA-Binding Protein. Cell **137**: 498–508.

Hendzel, M.J., Lever, M.A., Th'ng, J.P.H., and Sun, X. (2000). Rapid exchange of histone H1.1 on chromatin in living human cells. Nature **408**: 873–876.

Hershko, A. and Ciechanover, A. (1998). the Ubiquitin System. Annu. Rev. Biochem. **67**: 425–479.

Howe, F.S., Fischl, H., Murray, S.C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? BioEssays **39**: 1–12.

Hunter, C., Willmann, M.R., Wu, G., Yoshikawa, M., Gutiérrez-nava, M.D.L., and Poethig, R.S. (2006). Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in Arabidopsis. Development **2981**: 2973–2981.

Husbands, A.Y., Benkovics, A.H., Nogueira, F.T.S., Lodha, M., and Timmermans, M.C.P. (2015). The ASYMMETRIC LEAVES Complex Employs Multiple Modes of Regulation to Affect Adaxial-Abaxial Patterning and Leaf Complexity. Plant Cell **27**: 3321–3335.

Ietswaart, R., Wu, Z., and Dean, C. (2012). Flowering time control: another window to the connection between antisense RNA and chromatin. Trends Genet. **28**: 445–53.

Ilegems, M., Douet, V., Meylan-Bettex, M., Uyttewaal, M., Brand, L., Bowman, J.L., and Stieger, P.A. (2010). Interplay of auxin, KANADI and Class III HD-ZIP transcription factors in vascular tissue formation. Development **137**: 975–984.

Irish, V.F. and Sussex, I.A.N.M. (1992). A fate map of the Arabidopsis embryonic shoot apical meristem. Development **753**: 745–753.

Iwakawa, H., Iwasaki, M., Kojima, S., Ueno, Y., Soma, T., Tanaka, H., Semiarti, E., Machida, Y., and Machida, C. (2007). Expression of the ASYMMETRIC LEAVES2 gene in the adaxial domain of Arabidopsis leaves represses cell proliferation in this domain and is

critical for the development of properly expanded leaves. Plant J. **51**: 173–184.

Iwasaki, M. et al. (2013). Dual regulation of ETTIN (ARF3) gene expression by AS1-AS2 , which maintains the DNA methylation level , is involved in stabilization of leaf adaxial-abaxial partitioning in Arabidopsis. Development **1969**: 1958–1969.

Izhaki, A. and Bowman, J.L. (2007). KANADI and Class III HD-Zip Gene Families Regulate Embryo Patterning and Modulate Auxin Flow during Embryogenesis in Arabidopsis. Plant Cell **19**: 495–508.

Jermann, P., Hoerner, L., Burger, L., and Schübeler, D. (2014). Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. Proc. Natl. Acad. Sci. **111**: E3415–E3421.

Jiang, D., Kong, N.C., Gu, X., Li, Z., and He, Y. (2011). Arabidopsis COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. PLoS Genet. **7**: e1001330.

Jin, Q., Yu, L.R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.E., Wang, C., Brindle, P.K., Dent, S.Y.R., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J. **30**: 249–262.

Jing, Y., Sun, H., Yuan, W., Wang, Y., Li, Q., Liu, Y., Li, Y., and Qian, W. (2016). SUVH2 and SUVH9 Couple Two Essential Steps for Transcriptional Gene Silencing in Arabidopsis. Mol. Plant **9**: 1156–1167.

Johnson, L., Mollah, S., Garcia, B.A., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Jacobsen, S.E. (2004). Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res. **32**: 6511–6518.

Jover-Gil, S., Candela, H., Robles, P., Aguilera, V., Barrero, J.M., Micol, J.L., and Ponce, M.R. (2012). The MicroRNA pathway genes AGO1, HEN1 and HYL1 participate in leaf proximal-distal, venation and stomatal patterning in arabidopsis. Plant Cell Physiol. **53**: 1322–1333.

Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A., and Timmermans, M.C.P. (2004). microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. Nature **428**: 84–88.

Kang, S.C., Kim, S.K., Chai, J.C., Kim, S.H., Won, K., Lee, Y.S., Jung, K.H., and Chai, Y.G. (2015). Transcriptomic Profiling and H3K27me3 Distribution Reveal Both Demethylase-Dependent and Independent Regulation of Developmental Gene Transcription in Cell

Differentiation. PLoS One **10**: e0135276.

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

Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., and Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics **13**: 424.

Kassis, J.A., Kennison, J.A., and Tamkun, J.W. (2017). Polycomb and trithorax group genes in drosophila. Genetics **206**: 1699–1725.

Kelley, D.R., Arreola, A., Gallagher, T.L., and Gasser, C.S. (2012). ETTIN (ARF3) physically interacts with KANADI proteins to form a functional complex essential for integument development and polarity determination in Arabidopsis. Development **139**: 1105–1109.

Kerk, N.M., Ceserani, T., Tausta, S.L., Sussex, I.M., and Nelson, T.M. (2003). Laser Capture Microdissection of Cells from Plant Tissues. Plant Physiol. **132**: 27–35.

Kerstetter, R.A., Bollman, K., Taylor, R.A., Bomblies, K., and Poethig, R.S. (2001). KANADI regulates organ polarity in Arabidopsis. Nature **411**: 706–709.

Kim, J.-H., Sharma, A., Dhar, S.S., Lee, S.-H., Gu, B., Chan, C.-H., Lin1, H.-K., and Lee, M.G. (2014). UTX and MLL4 Coordinately Regulate Transcriptional Programs for Cell Proliferation and Invasiveness in Breast Cancer Cells. Cancer Res. **74**: 1705–1717.

Kim, J., Guermah, M., McGinty, R.K., Lee, J.S., Tang, Z., Milne, T.A., Shilatifard, A., Muir, T.W., and Roeder, R.G. (2009). RAD6-Mediated Transcription-Coupled H2B Ubiquitylation Directly Stimulates H3K4 Methylation in Human Cells. Cell **137**: 459–471.

Kim, M.Y. and Zilberman, D. (2014). DNA methylation as a system of plant genomic immunity. Trends Plant Sci. **19**: 320–6.

Kim, T. and Shiekhattar, R. (2015). Review Architectural and Functional Commonalities between Enhancers and Promoters. Cell **162**: 948–959.

Kingston, R.E. and Tamkun, J.W. (2014). Transcriptional Regulation by Trithorax-Group Proteins. Cold Spring Harb Perspect Biol **6**: a019349.

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Control of fertilization-independent endosperm development by the MEDEA polycomb gene in

Arabidopsis. Proc. Natl. Acad. Sci. **96**: 4186–4191.

Koch, F. and Andrau, J.C. (2011). Initiating RNA polymerase II and TIPs as hallmarks of enhancer activity and tissue-specificity. Transcription, **2**: 263–268.

Köhler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., and Gruissem, W. (2003). Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. EMBO J. **22**: 4804–4814.

Krogan, N.T. and Berleth, T. (2012). A dominant mutation reveals asymmetry in MP/ARF5 function along the adaxial-abaxial axis of shoot lateral organs. Plant Signal. Behav. **7**: 940–943.

Kuhlmann, M., Florian, M., Atsn, Á., and Suvh, Á.S.Á. (2012). Developmentally non-redundant SET domain proteins SUVH2 and SUVH9 are required for transcriptional gene silencing in Arabidopsis thaliana. Plant Mol. Biol. **79**: 623–633.

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 Genet. **7**: e1002040.

Lanzuolo, C. and Orlando, V. (2012). Memories from the Polycomb Group Proteins. Annu. Rev. Genet. **46**: 561–589.

Lau, P.N.I. and Cheung, P. (2011). Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. Proc. Natl. Acad. Sci. **108**: 2801–2806.

Laufs, P., Grandjean, O., Jonak, C., Kiêu, K., and Traas, J. (1998). Cellular parameters of the shoot apical meristem in Arabidopsis. Plant Cell **10**: 1375–90.

Laux, T., Mayer, K.F.X., Berger, J., Jürgens, G., Genetik, L., and München, L. (1996). The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. Development **96**: 87–96.

Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M.S., Strahl, B.D., Patel, D.J., and Jacobsen, S.E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature **498**: 385.

Lee, J.S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R., and Shilatifard, A. (2007a). Histone Crosstalk between H2B Monoubiquitination and H3 Methylation Mediated by COMPASS. Cell **131**: 1084–1096.

Lee, L.R., Wengier, D.L., and Bergmann, D.C. (2019). Cell-type–specific transcriptome and

histone modification dynamics during cellular reprogramming in the Arabidopsis stomatal lineage. Proc. Natl. Acad. Sci. **116**: 21914–21924.

Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., Croce, L. Di, and Shiekhattar, R. (2007b). Demethylation of H3K27 Regulates Polycomb Recruitment and H2A Ubiquitination. Science **318**: 447–450.

Lenstra, T.L. et al. (2011). Resource The Specificity and Topology of Chromatin Interaction Pathways in Yeast. Mol. Cell **42**: 536–549.

Li, B., Carey, M., and Workman, J.L. (2007). The Role of Chromatin during Transcription. Cell **128**: 707–719.

Li, E., Bhargava, A., Qiang, W., Friedmann, M.C., Forneris, N., Savidge, R.A., Johnson, L.A., Mansfield, S.D., Ellis, B.E., Douglas, C.J., and Douglas, C.J. (2012). The Class II KNOX gene KNAT7 negatively regulates secondary wall formation in Arabidopsis and is functionally conserved in Populus. New Phytol. **194**: 102–115.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation Protects miRNAs and siRNAs from a 3-End Uridylation Activity in Arabidopsis. Curr. Biol. **15**: 1501–1507.

Li, S., Vandivier, L.E., Tu, B., Gao, L., Won, S.Y., Li, S., Zheng, B., Gregory, B.D., and Chen, X. (2015). Detection of pol IV/RDR2-dependent transcripts at the genomic scale in arabidopsis reveals features and regulation of sirna biogenesis. Genome Res. **25**: 235–245.

Liang, D., White, R.G., and Waterhouse, P.M. (2014). Mobile gene silencing in Arabidopsis is regulated by hydrogen peroxide. PeerJ **2**: e701.

Lin, W., Shuai, B., and Springer., P.S. (2003). The Arabidopsis LATERAL ORGAN BOUNDARIES–Domain Gene ASYMMETRIC LEAVES2 Functions in the Repression of KNOX Gene Expression and in Adaxial-Abaxial Patterning. Plant Cell **15**: 2241–2252.

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

Liu, Q., Yao, X., Pi, L., Wang, H., Cui, X., and Huang, H. (2009). The ARGONAUTE10 gene modulates shoot apical meristem maintenance and establishment of leaf polarity by repressing miR165/166 in Arabidopsis. Plant J. **58**: 27–40.

Liu, Z.W., Zhou, J.X., Huang, H.W., Li, Y.Q., Shao, C.R., Li, L., Cai, T., Chen, S., and He, X.J. (2016). Two Components of the RNA-Directed DNA Methylation Pathway Associate

with MORC6 and Silence Loci Targeted by MORC6 in Arabidopsis. PLoS Genet. **12**: 1–24.

Lodha, M., Marco, C.F., and Timmermans, M.C.P. (2013). The ASYMMETRIC LEAVES complex maintains repression of KNOX homeobox genes via direct recruitment of Polycomb-repressive complex2. Genes Dev. **27**: 596–601.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**: 1–21.

Luo, M., Bilodeau, P., Koltunow, A., Dennis, E.S., Peacock, W.J., and Chaudhury, A.M. (1999). Genes controlling fertilization-independent seed development in Arabidopsis thaliana. Proc. Natl. Acad. Sci. **96**: 296–301.

Mallory, A.C., Reinhart, B.J., Jones-rhoades, M.W., Zamore, P.D., Kathryn, M., and Bartel, D.P. (2004). MicroRNA control of PHABULOSA in leaf development : importance of pairing to the microRNA 5- region. EMBO J. **23**: 3356–3364.

Margaritis, T., Oreal, V., Brabers, N., Maestroni, L., Vitaliano-prunier, A., and Holstege, F.C.P. (2012). Two Distinct Repressive Mechanisms for Histone 3 Lysine 4 Methylation through Promoting 3-End Antisense Transcription. PLoS Genet. **8**: e1002952.

Marí-ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O. (2013). Reconstructing de novo silencing of an active plant retrotransposon. Nat. Genet. **45**: 1029–1039.

Martínez, G., Panda, K., Köhler, C., and Slotkin, R.K. (2016). Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. Nat. Plants **2**: 1–8.

Mathieu, O., Probst, A. V., and Paszkowski, J. (2005). Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. EMBO J. **24**: 2783–2791.

Matsumura, Y., Ohbayashi, I., Takahashi, H., Kojima, S., Ishibashi, N., Keta, S., Nakagawa, A., Hayashi, R., Sae, J., Echeverria, M., and Sugiyama, M. (2016). A genetic link between epigenetic repressor AS1-AS2 and a putative small subunit processome in leaf polarity establishment of Arabidopsis. Biol. Open **5**: 942–954.

Mcconnell, J.R. and Barton, M.K. (1998). Leaf polarity and meristem formation in Arabidopsis. Development **2942**: 2935–2942.

Mcconnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. Nature **411**: 709–713.

McCue, A.D., Nuthikattu, S., Reeder, S.H., and Slotkin, R.K. (2012). Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. PLoS Genet. **8**: e1002474.

McCue, A.D., Panda, K., Nuthikattu, S., Choudury, S.G., Thomas, E.N., and Slotkin, R.K. (2015). ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. EMBO J. **34**: 20–35.

Melnyk, C.W., Molnar, A., and Bassett, A. and D.C.B. (2011). Report Mobile 24 nt Small RNAs Direct Transcriptional Gene Silencing in the Root Meristems of Arabidopsis thaliana. Curr. Biol. **21**: 1678–1683.

Meyerowitz, E.M. (1997). Genetic Control of Cell Division Patterns in Developing Plants. Cell **88**: 299–308.

Mikkelsen, T.S. et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature **448**: 553–560.

Miller, T. et al. (2001). COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. Proc. Natl. Acad. Sci. **98**: 12902–12907.

Min, K.H. et al. (2010). SET DOMAIN GROUP2 is the major histone H3 lysine 4 trimethyltransferase in Arabidopsis. Proc. Natl. Acad. Sci. **107**: 18557–18562.

Moissiard, G. et al. (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science **336**: 1448–1451.

Molnar, A., Melnyk, C.W., Bassett, A., Hardcastle, T.J., Dunn, R., and Baulcombe, D.C. (2010). Small Silencing RNAs in Plants Are Mobile and Direct Epigenetic. Science **328**: 872–875.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008). Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans-Acting siRNA Formation. Cell **133**: 128–141.

Moreno-romero, J., Jiang, H., Santos-González, J., and Köhler, C. (2016). Parental epigenetic asymmetry of PRC2-mediated histone modifications in the Arabidopsis endosperm. EMBO J. **35**: 1298–1311.

Morey, L., Aloia, L., Cozzuto, L., Benitah, S.A., and Croce, L. Di (2012). RYBP and Cbx7 Define Specific Biological Functions of Polycomb Complexes in Mouse Embryonic Stem Cells. Cell Rep. **3**: 60–69.

Nagano, H., Fukudome, A., Hiraguri, A., Moriyama, H., and Fukuhara, T. (2014). Distinct substrate specificities of Arabidopsis DCL3 and DCL4. Nucleic Acids Res. **42**: 1845–1856.

Oguchi, R., Douwstra, P., Fujita, T., Chow, W.S., and Terashima, I. (2011). Intra-leaf gradients of photoinhibition induced by different color lights: Implications for the dual mechanisms of photoinhibition and for the application of conventional chlorophyll fluorometers. New Phytol. **191**: 146–159.

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell **11**: 407–415.

Ozonov, E.A. and van Nimwegen, E. (2013). Nucleosome Free Regions in Yeast Promoters Result from Competitive Binding of Transcription Factors That Interact with Chromatin Modifiers. PLoS Comput. Biol. **9**: e1003181.

Pekker, I., Alvarez, J.P., and Eshed, Y. (2005). Auxin Response Factors Mediate Arabidopsis Organ Asymmetry via Modulation of KANADI Activity. Plant Cell **17**: 2899–2910.

Pekowska, A., Zacarias-, J., Imbert, J., Ferrier, P., and Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. EMBO J. **30**: 4198–4210.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes Dev. **18**: 2368–2379.

Prickaerts, P., Niessen, H.E.C., Mouchel-vielh, E., Dahlmans, V.E.H., Akker, G.G.H. Van Den, Geijselaers, C., Adriaens, M.E., Spaapen, F., Takihara, Y., Rapp, U.R., Peronnet, F., and Voncken, J.W. (2012). MK3 controls Polycomb target gene expression via negative feedback on ERK. Epigenetics Chromatin **5**: 12.

Pundhir, S., Bagger, F.O., Lauridsen, F.B., Rapin, N., and Porse, B.T. (2016). Peak-valley-peak pattern of histone modifications delineates active regulatory elements and their directionality. Nucleic Acids Res. **44**: 4037–4051.

Qi, J., Wang, Y., Yu, T., Cunha, A., Wu, B., Vernoux, T., and Meyerowitz, E. (2014). Auxin depletion from leaf primordia contributes to organ patterning. Proc. Natl. Acad. Sci. **111**: 18769–18774.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841–842.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature **470**: 279–285.

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., and Segal, E. (2012). Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat. Genet. **44**: 743–750.

Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J., Hagen, G., Guilfoyle, T.J., and Pas, L. (2009). Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II. Mol. Cell **1**: 192–203.

Reddington, J.P. et al. (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biol. **14**: R25.

Reddy, G.V. and Meyerowitz, E.M. (2005). Stem-cell homeostasis and growth dynamics can be uncoupled in the Arabidopsis shoot apex. Science **310**: 663–7.

Reinhart, B.J., Liu, T., Newell, N.R., Magnani, E., Huang, T., Kerstetter, R., Michaels, S., and Barton, M.K. (2013). Establishing a Framework for the Ad/Abaxial Regulatory Network of Arabidopsis: Ascertaining Targets of Class III HOMEODOMAIN LEUCINE ZIPPER and KANADI Regulation. Plant Cell **25**: 3228–3249.

Ren, G., Xie, M., Zhang, S., Vinovskis, C., Chen, X., and Yu, B. (2014). Methylation protects microRNAs from an AGO1-associated activity that uridylates 5’ RNA fragments generated by AGO1 cleavage. Proc. Natl. Acad. Sci. **111**: 6365–70.

Riggs, A.D. and Martienssen, R.A. (1996). Overview of epigenetic mechanisms. In Epigenetic mechanisms of gene regulation, R.A. Russo VEA, Martienssen R, ed (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), pp. 29–45.

Riggs, A.D., Martienssen, R.A., and Russo, V.E.A. (1996). Introduction. In Epigenetic mechanisms of gene regulation, Russo VEA, ed (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), pp. 1–4.

Ringrose, L., Rehmsmeier, M., Dura, J.M., and Paro, R. (2003). Genome-wide prediction of polycomb/trithorax response elements in Drosophila melanogaster. Dev. Cell **5**: 759–771.

Robertson, A.G. et al. (2008). Genome-wide relationship between histone H3 lysine 4

mono- and tri-methylation and transcription factor binding. Genome Res. **18**: 1906–1917.

Robine, N., Lin, W., Bonfils, S., Nicolas, A., and Ge, V. (2009). Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. EMBO J. **28**: 99–111.

Robinson, P.J.J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R.G., and Rhodes, D. (2008). 30 nm Chromatin Fibre Decompaction Requires both H4-K16 Acetylation and Linker Histone Eviction. J. Mol. Biol. **381**: 816–825.

Sarojam, R., Sappl, P.G., Goldshmidt, A., Efroni, I., Floyd, S.K., Eshed, Y., and Bowman, J.L. (2010). Differentiating Arabidopsis Shoots from Leaves by Combined YABBY Activities. Plant Cell **22**: 2113–2130.

Sassi, M. et al. (2012). COP1 mediates the coordination of root and shoot growth by light through modulation of PIN1- and PIN2-dependent auxin transport in Arabidopsis. Development **3412**: 3402–3412.

Sawa, S., Watanabe, K., Goto, K., Kanaya, E., Morita, E.H., and Okada, K. (1999). Filamentous flower, a meristem and organ identity gene of Arabidopsis, encodes a protein with a zinc finger and HMG-related domains. Genes Dev. **13**: 1079–1088.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of Arabidopsis thaliana development. Nat. Genet. **37**: 501–506.

Schmitges, F.W. et al. (2011). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Mol. Cell **42**: 330–341.

Schnittger, A., Grini, P.E., Folkers, U., and Hu, M. (1996). Epidermal Fate Map of the Arabidopsis Shoot Meristem. Dev. Biol. **255**: 248–255.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell **132**: 887–898.

Schubert, D., Clarenz, O., and Goodrich, J. (2005). Epigenetic control of plant development by Polycomb-group proteins. Curr. Opin. Plant Biol. **8**: 553–61.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell **128**: 735–45.

Schwab, R., Maizel, A., Ruiz-ferrer, V., Garcia, D., and Bayer, M. (2009). Endogenous TasiRNAs Mediate Non-Cell Autonomous Effects on Gene Regulation in Arabidopsis

thaliana. PLoS One **4**: e5980.

Schwämmle, V., Aspalter, C.M., Sidoli, S., and Jensen, O.N. (2014). Large scale analysis of co-existing post-translational modifications in histone tails reveals global fine structure of cross-talk. Mol. Cell. Proteomics **13**: 1855–1865.

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. Nat. Genet. **38**: 700–705.

Scofield, S., Dewitte, W., Murray, J.A.H., Arabidopsis, T., and Knat, B.P. (2008). A model for Arabidopsis class - 1 KNOX gene function. Plant Signal. Behav. **3**: 257–259.

Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S.E., Bastolla, U., and Gutierrez, C. (2014). The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. Plant Cell **25**: 2351–2366.

Shen, W.-H. and Xu, L. (2009). Chromatin remodeling in stem cell maintenance in Arabidopsis thaliana. Mol. Plant **2**: 600–9.

Shih, J.D., Fitzgerald, M.C., Sutherlin, M., and Hunter, C.P. (2009). The SID-1 double-stranded RNA transporter is not selective for dsRNA length. RNA **15**: 384–390.

Shiio, Y. and Eisenman, R.N. (2003). Histone sumoylation is associated with transcriptional repression. Proc. Natl. Acad. Sci. **100**: 13225–13230.

Shilatifard, A. (2012). The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. Annu. Rev. Biochem. **81**: 65–95.

Shulse, C.N. et al. (2019). High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types. Cell Rep. **27**: 2241-2247.e4.

Siegfried, K.R., Eshed, Y., Baum, S.F., Otsuga, D., Drews, G.N., and Bowman, J.L. (1999). Members of the YABBY gene family specify abaxial cell fate in Arabidopsis. Development **126**: 4117–28.

Simon, P. (2003). Q-Gene: processing quantitative real-time RT-PCR data. Bioinformatics **19**: 1439–1440.

Slane, D., Kong, J., Berendzen, K.W., Kilian, J., Henschen, a., Kolb, M., Schmid, M., Harter, K., Mayer, U., De Smet, I., Bayer, M., and Jurgens, G. (2014). Cell type-specific transcriptome analysis in the early Arabidopsis thaliana embryo. Development **141**: 4831–4840.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J.D., Feijó, J. a, and Martienssen, R. a (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell **136**: 461–72.

Smith, L.G., Greene, B.E.N., Veit, B., and Hake, S. (1992). A dominant mutation in the maize homeobox gene , Knotted-1 , causes its ectopic expression in leaf cells with altered fates. Development **30**: 21–30.

Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S., and Baulcombe, D.C. (2007). An SNF2 Protein Associated with Nuclear RNA Silencing and the Spread of a Silencing Signal between Cells in Arabidopsis. Plant Cell **19**: 1507–1521.

Soyano, T., Thitamadee, S., Machida, Y., and Chua, N. (2008). ASYMMETRIC LEAVES2-LIKE19 / LATERAL ORGAN BOUNDARIES DOMAIN30 and ASL20 / LBD18 Regulate Tracheary Element Differentiation in Arabidopsis. Plant Cell **20**: 3359–3373.

Springer, N.M., Napoli, C.A., Selinger, D.A., Pandey, R., Cone, K.C., Chandler, V.L., Kaeppler, H.F., and Kaeppler, S.M. (2003). Comparative Analysis of SET Domain Proteins in Maize and Arabidopsis Reveals Multiple Duplications Preceding the Divergence of Monocots and Dicots. Plant Physiol. **132**: 907–925.

Stahle, M.I., Kuehlich, J., Staron, L., von Arnim, A.G., and Golz, J.F. (2009). YABBYs and the Transcriptional Corepressors LEUNIG and LEUNIG_HOMOLOG Maintain Leaf Polarity and Meristem Activity in Arabidopsis. Plant Cell **21**: 3105–3118.

Stempor, P. and Ahringer, J. (2016). SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. Wellcome Open Res. **1**: 14.

Su, Y., Wang, S., Zhang, F., Zheng, H., Liu, Y., Huang, T., and Ding, Y. (2017). Phosphorylation of Histone H2A at Serine 95: A Plant-Specific Mark Involved in Flowering Time Regulation and H2A.Z Deposition. Plant Cell **29**: 2197–2213.

Tamada, Y., Yun, J.-Y., Woo, S. chul, and Amasino, R.M. (2009). ARABIDOPSIS TRITHORAX-RELATED7 Is Required for Methylation of Lysine 4 of Histone H3 and for Transcriptional Activation of FLOWERING LOCUS C. Plant Cell **21**: 3257–3269.

Tanay, A., Donnell, A.H.O., Damelin, M., and Bestor, T.H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. Proc. Natl. Acad. Sci. **104**: 5521–5526.

Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev. **17**: 49–63.

Tolhuis, B., Muijrers, I., De Wit, E., Teunissen, H., Talhout, W., Van Steensel, B., and Van

Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. Nat. Genet. **38**: 694–699.

Torti, S., Fornara, F., Vincent, C., Andre, F., Schoof, H., and Coupland, G. (2012). Analysis of the Arabidopsis Shoot Meristem Transcriptome during Floral Transition Identifies Distinct Regulatory Patterns and a Leucine-Rich Repeat Protein That Promotes Flowering. Plant Cell **24**: 444–462.

Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. Nature **474**: 516–522.

Vazquez, F., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J., Bartel, D.P., and Cre, P. (2004). Endogenous trans -Acting siRNAs Regulate the Accumulation of Arabidopsis mRNAs. Mol. Cell **16**: 69–79.

Villanueva, J.M., Broadhvest, J., Hauser, B.A., Meister, R.J., Schneitz, K., and Gasser, C.S. (1999). INNER NO OUTER regulates abaxial – adaxial patterning in Arabidopsis ovules. Genes Dev. **13**: 3160–3169.

Waddington, C.H. (1953). Epigenetics and evolution. Symp Soc Exp Biol: 186–199.

Walter, M., Teissandier, A., Pérez-Palacios, R., and Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. Elife **5**: e11418.

Wang, S.P., Tang, Z., Chen, C.W., Shimada, M., Koche, R.P., Wang, L.H., Nakadai, T., Chramiec, A., Krivtsov, A. V., Armstrong, S.A., and Roeder, R.G. (2017). A UTX-MLL4-p300 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for Eliciting Transcription. Mol. Cell **67**: 308-321.e6.

Wang, Y., Gu, X., Yuan, W., Schmitz, R.J., and He, Y. (2014). Photoperiodic control of the floral transition through a distinct polycomb repressive complex. Dev. Cell **28**: 727–736.

Wenkel, S., Emery, J., Hou, B., Evans, M.M.S., and Barton, M.K. (2007). A Feedback Regulatory Module Formed by LITTLE ZIPPER and HD-ZIPIII Genes. Plant Cell **19**: 3379–3390.

Wenzel, C.L., Schuetz, M., Yu, Q., and Mattsson, J. (2007). Dynamics of MONOPTEROS and PIN-FORMED1 expression during leaf vein pattern formation in Arabidopsis thaliana. Plant J. **49**: 387–398.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. Cell **135**: 635–648.

Wilm, M., Roguev, A., Schaft, D., Pijnappel, W.W.M.P., Stewart, A.F., Shevchenko, A., and Aasland, R. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J. **20**: 7137–7148.

Wu, G., Lin, W., Huang, T., Poethig, R.S., Springer, P.S., and Kerstetter, R.A. (2008). KANADI1 regulates adaxial – abaxial polarity in Arabidopsis by directly repressing the transcription of ASYMMETRIC LEAVES2. Proc. Natl. Acad. Sci. **105**: 16392–16397.

Wu, L., Mao, L., and Qi, Y. (2012). Roles of DICER-LIKE and ARGONAUTE Proteins in TAS-Derived Small Interfering RNA-Triggered DNA Methylation. Plant Physiol. **160**: 990–999.

Wu, X., Johansen, J.V., and Helin, K. (2013). Article Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. Mol. Cell **49**: 1134–1146.

Xie, Z., Allen, E., Wilken, A., and Carrington, J.C. (2005). DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. Proc. Natl. Acad. Sci. **102**: 12984–12989.

Xu, L. and Shen, W.H. (2008). Polycomb Silencing of KNOX Genes Confines Shoot Stem Cell Niches in Arabidopsis. Curr. Biol. **18**: 1966–1971.

Yadav, R.K., Girke, T., Pasala, S., Xie, M., and Reddy, G.V. (2009). Gene expression map of the Arabidopsis shoot apical meristem stem cell niche. Proc. Natl. Acad. Sci. **106**: 4941–6.

Yadav, R.K., Tavakkoli, M., Xie, M., Girke, T., and Reddy, G.V. (2014). A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche. Development **141**: 2735–44.

Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M. (2013). VAL-and AtBMI1-Mediated H2Aub initiate the switch from embryonic to postgerminative growth in arabidopsis. Curr. Biol. **23**: 1324–1329.

Yang, R. et al. (2017). The developmental regulator PKL is required to maintain correct DNA methylation patterns at RNA-directed DNA methylation loci. Genome Biol. **18**: 1–18.

Yoshida, N., Yanai, Y., Chen, L., Kato, Y., Hiratsuka, J., Miwa, T., Sung, Z.R., and Takahashi, S. (2001). EMBRYONIC FLOWER2, a novel Polycomb group protein homolog, mediates shoot development and flowering in Arabidopsis. Plant Cell **13**: 2471–2481.

Yoshikawa, M., Peragine, A., Park, M.Y., and Poethig, R.S. (2005). A pathway for the biogenesis of trans -acting siRNAs in Arabidopsis. Genes Dev.: 2164–2175.

You, Y., Sawikowska, A., Neumann, M., Posé, D., Capovilla, G., Langenecker, T., Neher, R.A., Krajewski, P., and Schmid, M. (2017). Temporal dynamics of gene expression and histone marks at the Arabidopsis shoot meristem during flowering. Nat. Commun. **8**: 15120.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a Crucial Step in Plant microRNA Biogenesis. Science **307**: 932–936.

Yu, Y. et al. (2017). ARGONAUTE10 promotes the degradation of miR165/6 through the SDN1 and SDN2 exonucleases in Arabidopsis. PLoS Biol. **15**: e2001272.

Yuan, J.S., Reed, A., Chen, F., and Stewart, C.N. (2006). Statistical analysis of real-time PCR data. BMC Bioinformatics **7**: 85.

Žádníková, P., Petrášek, J., Marhavý, P., Raz, V., Vandenbussche, F., Ding, Z., Morita, M.T., Tasaka, M., Hejátko, J., and Straeten, D. Van Der (2010). Role of PIN-mediated auxin efflux in apical hook development of Arabidopsis thaliana. Development **617**: 607–617.

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**: 1952–8.

Zhai, J. et al. (2015). A One Precursor One siRNA Model for Pol IV- Dependent siRNA Biogenesis. Cell **163**: 445–455.

Zhang, C., Barthelson, R. a, Lambert, G.M., and Galbraith, D.W. (2008). Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. Plant Physiol. **147**: 30–40.

Zhang, X., Bernatavichute, Y. V, Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol. **10**: R62.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007a). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol. **5**: e129.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007b). Whole-genome analysis of histone H3 lysine 27 trimethylation

in Arabidopsis. PLoS Biol. **5**: 1026–1035.

Zhang, X., Germann, S., Blus, B.J., Khorasanizadeh, S., Gaudin, V., and Jacobsen, S.E. (2007c). The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. Nat. Struct. Mol. Biol. **14**: 869–871.

Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., and Theresa, T. (2012). HESO1, a nucleotidyl transferase in Arabidopsis, uridylates unmethylated miRNAs and siRNAs to trigger their degradation Yuanyuan. Curr. Biol. **22**: 689–694.

Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. EMBO J. **26**: 1691–1701.

Zhou, Y., Honda, M., Zhu, H., Zhang, Z., Guo, X., Li, T., Li, Z., Peng, X., Nakajima, K., Duan, L., and Zhang, X. (2015). Spatiotemporal sequestration of miR165/166 by arabidopsis argonaute10 promotes shoot apical meristem maintenance. Cell Rep. **10**: 1819–1827.

Zhou, Y., Romero-Campero, F.J., Gómez-ZambranoÁngeles, Turck, F., and Calonje, M. (2017). H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity. Genome Biol. **18**: 1–13.

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics **11**: 237.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation. Science **299**: 716–720.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of Arabidopsis ARGONAUTE4 in RNA-Directed DNA Methylation Triggered by Inverted Repeats. Curr. Biol. **14**: 1214–1220.

7. Appendix

7.1 Abbreviation

AB: Abaxial

AD: Adaxial

Bp: Base Pair

BR: Blocking Reagent

BSA: Bovine Serum Albumin

ChIP: Chromatin immunoprecipitation

Ct: Cycle Threshold

DAPI: 4',6-Diamidino-2-Phenylindole

DEG: Differentially Expressed Gene

DMRs: Differentially Methylated Regions

DNA: Deoxyribonucleic acid

DREME: Discriminative Regular Expression Motif Elicitation

dsDNA: double stranded Deoxyribonucleic acid

dsRNA: double stranded Ribonucleic acid

EDTA: Ethylenediaminetetraacetic acid

EGTA: Ethyleneglycotetraacetic acid

ESCs: Embryo Stem Cells

FACS: Fluorescence-activated cell sorting

FDR: False Discovery Rate

GFP: Green Fluorescent Protein

GO: Gene Ontology

INTACT: Isolation of Nuclei Tagged in specific Cell Types

mM: millimolar

MOPS: 3-(N-morpholino) propanesulfonic acid

NFR: Nucleosome-Free region

NPB: Nuclei Purification Buffer

NTF: Nuclei Tagged Fusion protein

PCR: Polymerase Chain Reaction

PFA: Paraformaldehyde

PI: Propidium Iodide

RdDM: RNA-directed DNA methylation

SAM: Shoot Apical Meristem

SICER: Spatial clustering approach for the identification of ChIP-enriched regions

TE: Transposon element

TES: Transcription End Site

TGS: Transcriptional Gene Silencing

TSS: Transcription Start Site

YFP: Yellow Fluorescent Protein

Acknowledgments

This research work was started in the Institute for Genetics at the Heinrich Heine University and finished in the AG Epigenetik der Pflanzen of Free University of Berlin under the supervision of Prof. Dr. Daniel Schubert. I would like to thank Daniel for the opportunity he gave me to do this PhD. I am particularly grateful to him for his excellent supervision, practical help and every kind of support I received during this project.

Special thanks to all the members of the labs, both in Düsseldorf and in Berlin, in particular to Dr. Marcel Lafos, Dr. Sara Farrona, Dr. Ines Trindade and Dr. Ralf Müller-Xing for sharing their knowledge and giving great suggestions to the project. I would like to thank Dr. Julia Kleinmanns and Dr. Pawel Mikulski for sharing ideas and support, in particular, I would like to thank Julia for the proofreading. I also would like to thank Jenny Sobial and Conny Klose for their technical support.

I wish to express my appreciation to Dr. Xinai Zhao, Dr. Feng Zhao, Dr. Yuanyuan Chen for reading my thesis.

Also, I sincerely thank my parents! Thanks for their love and countless support! I also want to thank my friends in China, especially Tiantian Liu.

Finally, I gratefully acknowledge Prof. Dr. Rüdiger Simon for the review of this dissertation