

# ATX3, ATX4, and ATX5 Encode Putative H3K4 Methyltransferases and Are Critical for Plant Development<sup>1</sup>[CC-BY]

Li-Qun Chen,<sup>a,2</sup> Jin-Hong Luo,<sup>b,2</sup> Zhen-Hai Cui,<sup>b,c,2</sup> Ming Xue,<sup>b</sup> Li Wang,<sup>d</sup> Xiao-Yu Zhang,<sup>e</sup> Wojciech P. Pawlowski,<sup>f,3</sup> and Yan He<sup>b,f,3</sup>

<sup>a</sup>State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing 100083, China

<sup>b</sup>National Maize Improvement Center of China, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100083, China

<sup>c</sup>College of Biological Science and Technology, Shenyang Agricultural University, Shenyang 110866, China

<sup>d</sup>Broad Institute of MIT and Harvard University, Cambridge, Massachusetts 02142

<sup>e</sup>Department of Plant Biology, University of Georgia, Athens, Georgia 30602

<sup>f</sup>School of Integrative Plant Science, Cornell University, Ithaca, New York 14853

ORCID IDs: 0000-0002-8179-6172 (L.-Q.C.); 0000-0003-3640-8502 (Y.H.).

Methylation of Lys residues in the tail of the H3 histone is a key regulator of chromatin state and gene expression, conferred by a large family of enzymes containing an evolutionarily conserved SET domain. One of the main types of SET domain proteins are those controlling H3K4 di- and trimethylation. The genome of *Arabidopsis* (*Arabidopsis thaliana*) encodes 12 such proteins, including five ARABIDOPSIS TRITHORAX (ATX) proteins and seven ATX-Related proteins. Here, we examined three until-now-unexplored ATX proteins, ATX3, ATX4, and ATX5. We found that they exhibit similar domain structures and expression patterns and are redundantly required for vegetative and reproductive development. Concurrent disruption of the *ATX3*, *ATX4*, and *ATX5* genes caused marked reduction in H3K4me2 and H3K4me3 levels genome-wide and resulted in thousands of genes expressed ectopically. Furthermore, *atx3/atx4/atx5* triple mutants resulted in exaggerated phenotypes when combined with the *atx2* mutant but not with *atx1*. Together, we conclude that *ATX3*, *ATX4*, and *ATX5* are redundantly required for H3K4 di- and trimethylation at thousands of sites located across the genome, and genomic features associated with targeted regions are different from the *ATXR3/SDG2*-controlled sites in *Arabidopsis*.

Changes in chromatin state involve several types of modification of an array of histone residues to allow fine-tuning of genes with diverse transcriptional profiles. A major regulator of chromatin state and controller of gene expression is methylation of Lys residues

in the tail of the H3 histone (Thorstensen et al., 2011). This modification is conferred by a large family of enzymes, histone Lys methyltransferases. Most of them contain an evolutionarily conserved SET (suppressor of variegation, enhancer of zeste and trithorax) domain, which is responsible for the methyl transferase activity (Tschiersch et al., 1994; Lei et al., 2012). The *Arabidopsis* (*Arabidopsis thaliana*) genome encodes a group of 49 such proteins, known as the SET DOMAIN GROUP (SDG; [www.chromdb.org](http://www.chromdb.org); Baumbusch et al., 2001; Ng et al., 2007; Berr et al., 2010).

SDG proteins are classified into five distinct classes, based on their activity and domain architecture (Springer et al., 2003; Zhao and Shen, 2004; Pontvianne et al., 2010). One of the main classes is class III, which contains proteins acting on Lys 4 (H3K4; Jackson et al., 2002; Alvarez-Venegas et al., 2003; Lindroth et al., 2008; Xu et al., 2008; Jacob et al., 2009). There are five class III SDG proteins in *Arabidopsis*, named ATX1–5, which are all homologs of Trithorax proteins found throughout eukaryotes (Veerappan et al., 2008; Schuettengruber et al., 2011). Only two of them have been characterized so far. ATX1 was found to be involved in H3K4 trimethylation and

<sup>1</sup> This work was supported by grants from the US National Science Foundation (IOS-1025881) and National Natural Science Foundation of China (31471172 and 31470279).

<sup>2</sup> These authors contributed equally to the article.

<sup>3</sup> Address correspondence to [yh352@cau.edu.cn](mailto:yh352@cau.edu.cn) or [wp45@cornell.edu](mailto:wp45@cornell.edu).

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors ([www.plantphysiol.org](http://www.plantphysiol.org)) is: Yan He ([yh352@cau.edu.cn](mailto:yh352@cau.edu.cn)).

Y.H., W.P.P. and X.-Y.Z. conceived the project; Y.H. and W.P.P. supervised the project; L.-Q.C., Y.H., Z.-H.C., M.X., and L.W. conducted experiments, and J.-H.L. performed bioinformatics and statistical analyses; manuscript was prepared by Y.H. and W.P.P. with input from J.-H.L. and L.-Q.C.; all authors read and approved the final manuscript.

[CC-BY] Article free via Creative Commons CC-BY 4.0 license.

[www.plantphysiol.org/cgi/doi/10.1104/pp.16.01944](http://www.plantphysiol.org/cgi/doi/10.1104/pp.16.01944)

required for root, leaf, and floral organ development (Baumbusch et al., 2001; Alvarez-Venegas et al., 2003; Alvarez-Venegas and Avramova, 2005; Fromm and Avramova, 2014; Napsucially-Mendivil et al., 2014), as well as for transcriptional regulation of several stress-response genes (Ding et al., 2011a, 2011b). ATX2 has sequence that is fairly similar to that of ATX1, but the protein shows diverged biochemical properties. It exhibits H3K4me2 activity, rather than H3K4me3 activity, and influences expression of a gene set that is largely different from that affected by ATX1 (Pien et al., 2008; Saleh et al., 2008).

In addition to the five ATX proteins, Arabidopsis has seven ATX-Related (ATXR) proteins, of which only two have been characterized (Baumbusch et al., 2001; Avramova, 2009). ATXR7/SDG25 is involved in flowering time regulation by activating expression of FLOWERING LOCUS C (FLC) (Berr et al., 2009; Tamada et al., 2009). ATXR3/SDG2, in contrast, has a much broader role, as it affects a large number of processes participating in both sporophyte and gametophyte development (Berr et al., 2010; Guo et al., 2010), which it does by controlling H3K4 trimethylation at thousands of loci genome-wide.

H3K4me2 and H3K4me3 are both marks of active chromatin. Interestingly, however, they display distinct genomic profiles (Zhang et al., 2009; Du et al., 2013). H3K4me3 marks tend to appear upstream of H3K4me2 marks within genes (Zhang et al., 2009) and are more narrowly spread compared to H3K4me2 marks (van Dijk et al., 2010). There is also indication that the two marks display different dynamics in response to external stimuli (van Dijk et al., 2010). These observations indicate that H3K4me2 and H3K4me3 are controlled by distinct cellular mechanisms and sets of genes. However, despite a large number of ATX and ATXR proteins identified, exact roles of specific genes have not been well defined, particularly for the genes involved in H3K4me2. To elucidate the roles of ATX proteins, we characterized the three so-far-unexplored ATX family members, ATX3, ATX4, and ATX5. We found that they are responsible for controlling a large number of genes critical for both vegetative and reproductive development. Furthermore, they affect H3K4me2 and H3K4me3 at thousands of sites throughout the genome. The former makes them the first enzymes with genome-wide H3K4me2 activity identified so far in Arabidopsis. ATX3/4/5 are similar to ATXR3/SDG2 in that they control genome-wide H3K4me3 profiles. However, the loci affected by ATX3/4/5 are largely distinct from those affected by ATXR3/SDG2, indicating that ATX3/4/5 act in a pathway separate from ATXR3/SDG2.

## RESULTS

### ATX3, ATX4, and ATX5 Exhibit Common Evolutionary Origin, Identical Domain Structure, and Similar Expression Patterns

The five Arabidopsis *TRX* genes have been shown to form two subfamilies, one containing *ATX1* and *ATX2*, and the other comprised of *ATX3*, *ATX4*, and *ATX5*

(Avramova, 2009). To examine the evolutionary origin of these genes, we conducted a phylogeny reconstruction using homologs of *ATX* genes from rice (*Oryza sativa*), maize (*Zea mays*), and *Physcomitrella patens* (Fig. 1A). Rice has one putative *ATX1/2* homolog (LOC\_Os09g04890) and two putative *ATX3/4/5* homologs (LOC\_Os01g46700 and LOC\_Os01g11952; Ng et al., 2007). The maize genome has one putative homolog of *ATX1/ATX2* (GRMZM2G013794) and three putative *ATX3/4/5* homologs (GRMZM2G372928, GRMZM2G085266, and GRMZM2G170412; Springer et al., 2003). In *Physcomitrella*, there are three *ATX1/2* homologs (XP\_001766115, XP\_001767466, and XP\_001780587) and a single *ATX3/4/5* homolog (XP\_001777592), which we identified using BLAST searches. Our phylogeny reconstruction showed that the two plant *ATX* subfamilies originated early in plant evolution, as the two groups are already present in *Physcomitrella* (Fig. 1A). However, the origins of the paralogs within each subfamily are more recent and occurred independently in different plant lineages.

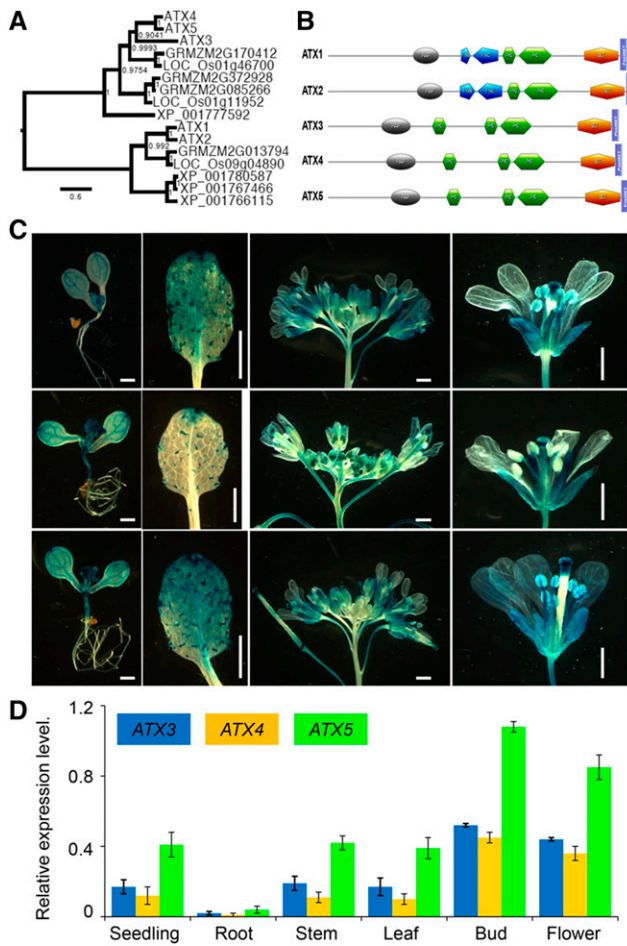
The main feature defining the two *ATX* subfamilies is the presence of domain associated with SET in trithorax (of either the F/Y-rich N terminus or F/Y-rich C terminus type) in the *ATX1/2* subfamily, whereas the *ATX3/4/5* subfamily lacks the domain associated with SET in trithorax but instead contains an additional plant homeodomain (PHD) finger domain (Fig. 1B). The latter structure and its amino acid sequence are conserved in all three Arabidopsis proteins (Fig. 1B; Supplemental Fig. S1), which suggests that they could exhibit similar biochemical activities.

To start investigating functional properties of *ATX3*, *ATX4*, and *ATX5*, we first examined the expression patterns of their corresponding genes. To do this, we generated transgenic plants carrying promoters of the three genes fused to the coding region of the *GUS* gene. The three transgenes exhibited similar patterns of expression, although *ATX5:GUS* showed stronger overall expression than *ATX3:GUS* and *ATX4:GUS* (Fig. 1C). All three transgenes were expressed in cotyledons, leaves, and hypocotyls, but not in roots (Fig. 1C). Strong expression was also observed in vascular tissues and trichomes (Fig. 1C). In reproductive tissues, expression was detected in sepals, pedals, anthers, filaments, styles, and stigmas, but not in mature pollen (Fig. 1C). To verify the expression patterns, we conducted real-time quantitative PCR (RT-qPCR). Consistent with the *GUS* results, we found transcripts of *ATX3*, *ATX4*, and *ATX5* in seedlings, stems, leaves, green buds, and open flowers. However, only low levels of transcripts were present in roots (Fig. 1D).

Taken together, our results indicated that Arabidopsis *ATX3*, *ATX4*, and *ATX5* are genealogically closely related and exhibit the same protein domain structure as well as similar tissue-specific expression patterns. These data suggested that the three genes may have similar functions.

### *ATX3*, *ATX4*, and *ATX5* Are Redundantly Critical for Plant Development

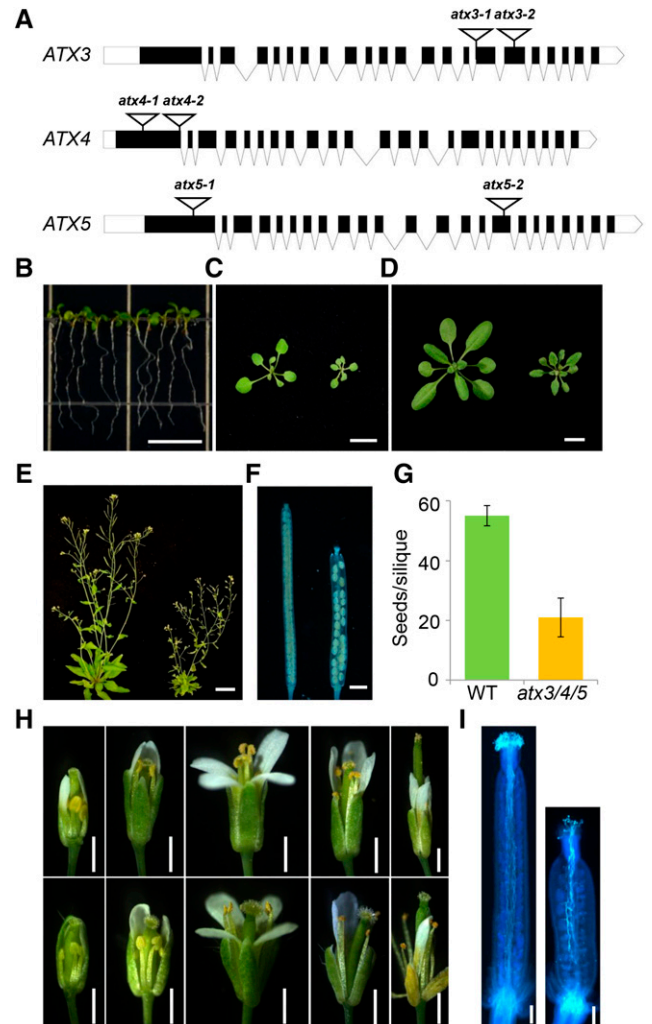
To examine whether *ATX3*, *ATX4*, and *ATX5* are indeed functionally redundant, we conducted mutant



**Figure 1.** Phylogenetic relationship, protein domain structures, and expression patterns of Arabidopsis *ATX3*, *ATX4*, and *ATX5* genes. **A**, Bayesian phylogeny reconstruction of ATX homologs in Arabidopsis, maize, rice, and *Physcomitrella patens*. Numbers next to branches indicate posterior probability values. The scale indicates number of substitutions per site. Protein sequences were aligned using ClustalX (Jeanmougin et al., 1998), and phylogeny reconstruction was conducted using MrBayes (Huelsenbeck et al., 2001). **B**, Domain organization of Arabidopsis TRX proteins. Protein structure was analyzed using NCBI-CD and SMART searches. Abbreviations: PWWP, Domain named after a conserved Pro-Trip-Trip-Pro motif; FYRN, F/Y-rich N terminus; FYRC, F/Y-rich C terminus; PHD, plant homeodomain. **C**, Analysis of spatial and temporal expression patterns of *ATX3* (top), *ATX4* (middle), and *ATX5* (bottom) using promoter-GUS fusion. Bars = 1 mm. **D**, RT-qPCR analysis of *ATX3*, *ATX4*, and *ATX5* expression. Expression levels relative to the expression level of the *ACTIN2* gene are presented with *ses*. Values are means of three independent experiments.

analyses. We obtained two T-DNA insertional mutants for each of the three *ATX* genes from various Arabidopsis mutant resources (see “Materials and Methods”). The T-DNA insertion sites were verified by sequencing their flanking regions (Fig. 2A). Using RT-PCR, we could not detect transcripts of the disrupted genes in any of the six mutants (Supplemental Fig. S2), suggesting that the mutations are null. However, none of the single mutants exhibited a discernible vegetative or reproductive mutant phenotype under normal growth conditions. We then

generated double mutants in different combinations: *atx3/atx4*, *atx3/atx5*, and *atx4/atx5*. However, we still could not find any apparent mutant phenotype. Consequently, we produced triple mutants, and these did display several strong phenotypes. The triple mutants were generated using all eight possible combinations of *atx* mutant alleles. However, since all combinations exhibited nearly identical phenotypes, we only focused on the *atx3-1/atx4-1/atx5-1* triple mutant in further phenotypic characterization.



**Figure 2.** Phenotypes of *atx3/4/5* triple mutant plants. **A**, Diagram of genomic structures of the *ATX3*, *ATX4*, and *ATX5* genes with T-DNA insertion sites marked with triangles. Bars indicate exons and lines represent introns. Scale bar = 1 kb. **B** to **E**, Phenotype comparison of 1-week-old (**B**), 2-week-old (**C**), 3-week-old (**D**), and 6-week-old (**E**) in wild-type (WT; left) and *atx3/4/5* mutant (right) plants. Bars = 1 cm in **B** to **D** and 2 cm in **E**. **F** and **G**, Phenotype comparison of siliques of wild-type and *atx3/4/5* mutant plants. Bars = 1 mm. **H**, Flower development at stages 12 to 16 (from left to right) in wild-type (top) and *atx3/4/5* mutant plants (bottom). Bars = 1 mm. **I**, Aniline blue staining of pollen tubes in pollinated gynoecium at stage 15 of flower development in wild-type (left) and *atx3/4/5* mutant plants (right). Bars = 200  $\mu$ m.

*atx3-1/atx4-1/atx5-1* triple mutant plants grew normally at early seedling stages of 1 to 7 d (Fig. 2B), but obvious abnormalities started appearing at 2 weeks of growth. In particular, leaves of 2-week-old seedlings were much smaller than those of wild-type plants (Fig. 2C). Furthermore, growth was drastically retarded throughout vegetative and reproductive stages, resulting in dwarf plants with small leaf rosettes (Fig. 2, D and E). Flowering time was not significantly affected in mutant plants. However, seed set was dramatically reduced: mutant plants exhibited an average of 21 seeds per silique, compared to 55 seeds per silique in wild type ( $n = 50$ ; Fig. 2, F and G).

To elucidate causes of the reduced seed set, we dissected flower development in the *atx3-1/atx4-1/atx5-1* mutant. We found that stamen development was significantly delayed starting at stage 12 (Fig. 2H). At that point, wild-type anthers matured and were yellowish, whereas anthers in the triple mutant remained green. Starting at stage 13, mutant stamens were also much shorter than wild-type stamens, which prevented pollen release from anther sacs onto stigma (Fig. 2H). However, aniline blue staining showed that the growth of pollen tubes that managed to land and germinated on stigma was normal (Fig. 2I). Following pollen release, anthers in the mutant were also slow to senesce (Fig. 2H). Moreover, mutant anthers were smaller than wild-type anthers and contained aborted pollen grains (Supplemental Fig. S3A), although most mature pollen grains developed normally (Supplemental Fig. S3B). The gynoeceium in mutant plants appeared normal (Supplemental Fig. S3C). These observations implicate defects in anther development as the cause for the reduced seed set in *atx3/4/5* mutant plants.

Taken together, our mutant analyses indicated that *ATX3*, *ATX4*, and *ATX5* are redundantly required for both vegetative as well as reproductive development.

### Disruption of *ATX3/4/5* Causes Decreases of H3K4 Di- and Trimethylation at Sites throughout the Genome

To study biochemical functions of *ATX3*, *ATX4*, and *ATX5*, we first examined the activity of the proteins in vitro by expressing the C-terminal, SET-domain-containing regions of the corresponding genes in *E. coli*. We strictly followed the methods previously used to produce a recombinant SDG2 protein (Guo et al., 2010). However, despite generating sufficient protein quantities, we were not able to detect HMT activity using recombinant mouse H3 histones as substrate (Supplemental Fig. S4). Based on this result, we concluded that the proteins may require additional factors for their HMTase activity, such as interacting proteins or posttranslational modifications.

We then investigated the activity of the three proteins in planta by using western blot analyses to compare the overall levels of several histone modifications, H3K4me1, H3K4me2, H3K4me3, H3K9me2, H3K27me3, and H3K36me3, in wild-type and *atx3/atx4/atx5* triple-mutant plants. We found that levels of H3K4me2 and H3K4me3 were consistently lower in the triple mutant compared to

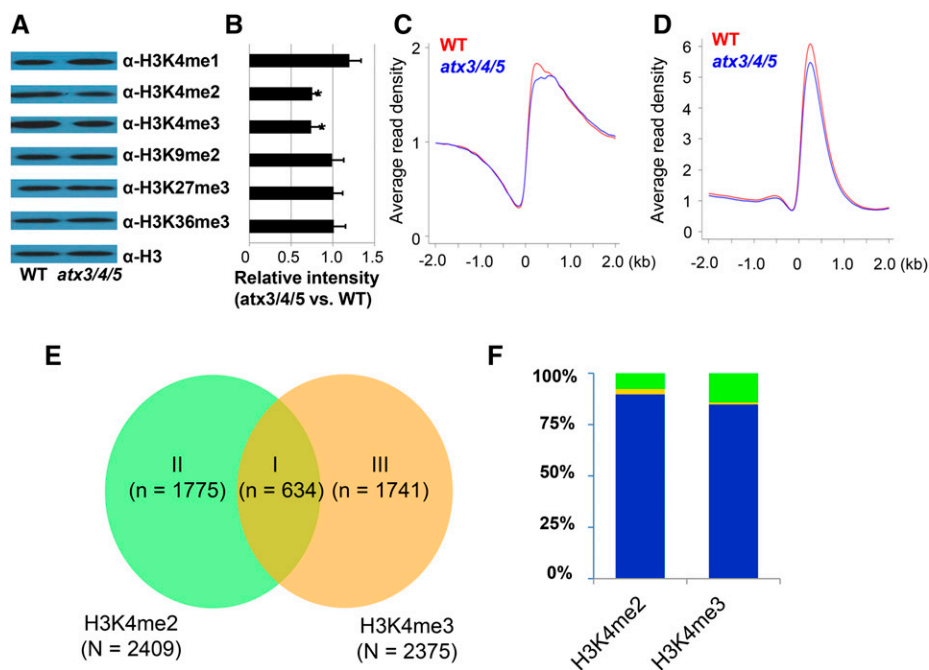
wild type (Fig. 3A). The quantification of blot using ImageJ software (<https://imagej.nih.gov/ij/>) further revealed that the reduction in both H3K4me2 and H3K4me3 was ~25% and statistically significant (Fig. 3B). In contrast, H3K4me1, H3K9me2, H3K27me3, and H3K36me3 were not affected. These data suggested that *ATX3/4/5* specifically catalyze H3K4 di- and trimethylation.

To learn more about the properties of *ATX3*, *ATX4*, and *ATX5*, we examined distribution of H3K4me2 and H3K4me3 marks in wild-type and *atx3/4/5* mutant plants using chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) (Supplemental Table S1). In wild-type plants, we identified 20,585 H3K4me2 sites (Supplemental Table S2) and 14,496 H3K4me3 sites (Supplemental Table S3). In *atx3/4/5* mutant plants, the number of H3K4me2 and H3K4me3 sites was substantially lower than in wild type (Supplemental Tables S4 and S5). As typical for H3K4me2 and H3K4me3 patterns (Zhang et al., 2009), we found the most enrichment occurring around gene transcription start sites (TSSs) in both wild-type and mutant plants (Fig. 3, C and D). Although the overall distribution of H3K4me2 or H3K4me3 tend to overlap in wild-type and *atx3/4/5* mutant plants, the intensity around peak summit around TSSs for both H3K4me2 and H3K4me3 was significantly decreased in mutant (Student's *t* test,  $P < 2.2e-16$ ) after conducting a simulation test ("Materials and Methods").

In the *atx3/4/5* mutant, we identified 2409 sites exhibiting decreased presence of H3K4me2 and 2375 sites with decreased H3K4me3, representing 11.7% and 16.4% of all wild-type H3K4me2 and H3K4me3 sites, respectively. These data demonstrated that *ATX3/4/5* control the landscape of both H3K4me2 and H3K4me3 throughout the genome (Fig. 3E; Supplemental Fig. S5).

The sites affected in the *atx3/4/5* mutant could be grouped into three subclasses. Class I ( $n = 634$ ) contains sites exhibiting a decrease in both H3K4me2 and H3K4me3 (Fig. 3E; Supplemental Fig. S6A; Supplemental Table S6). Class II ( $n = 1775$ ) are sites showing a decrease in H3K4me2 but not in H3K4me3 (Fig. 3E; Supplemental Fig. S6, B and C; Supplemental Table S7). Finally, class III ( $n = 1741$ ) contains sites showing no change in H3K4me2 but a decrease in H3K4me3 (Fig. 3E, Supplemental Fig. S6, D and E; Supplemental Table S8). The existence of class II and III sites is intriguing; as H3K4me2 is substrate for H3K4me3, changes in H3K4me3 should accompany changes in H3K4me2. A possible explanation of why these sites are present would be that most of them already lack H3K4me2 or H3K4me3 in wild type (Supplemental Fig. S6, B and D). However, as many as 26% of class II sites (Supplemental Fig. S6C) and 22% of class III sites (Supplemental Fig. S6E) exhibited both H3K4me2 and H3K4me3 in wild type. This finding suggests that there might be other HMTases acting on the same chromosomes sites independently from *ATX3/4/5*.

To investigate the mode of redundancy of *ATX3*, *ATX4*, and *ATX5* in regulating H3K4me2/3 levels at specific regions, we selected six loci from the list of



**Figure 3.** *ATX3/4/5*-affected H3K4me2/3 sites. A representative western blot analysis of H3K4me1, H3K4me2, H3K4me3, H3K9me2, H3K27me3, and H3K36me3 levels in wild-type (WT) and *atx3/4/5* mutant plants. For each antibody, the same amount of total protein was loaded. B, Relative intensity of protein levels in *atx3/4/5* versus wild-type. The extent of protein was determined by first normalizing the band intensity of specific antibodies relative to H3 antibody. Data were shown as means  $\pm$  se from three independent biological replicates. \*Significant at  $P \leq 0.01$ . C and D, Distribution of H3K4me2 (C) and H3K4me3 (D) around TSSs. Total numbers of ChIP-seq reads were normalized and quantified in 50-bp windows in regions 2 kb up- and downstream of TSS. E, Venn diagram showing the overlap of *ATX3/4/5*-affected H3K4me2 and H3K4me3 sites. F, Lack of correlation between a decrease in H3K4me2/3 levels and gene transcript levels. Genes showing a coordinated decrease in *ATX3/4/5*-affected H3K4me2/3 levels and transcript levels in the *atx3/4/5* triple mutant relative to wild type are represented by green bars. Genes whose transcript levels are up-regulated despite a decrease in H3K4me2/3 levels are orange. Genes showing no change in transcript levels while decrease in H3K4me2/3 levels are blue.

genomic sites affected in the *atx3/4/5* triple mutant (Supplemental Table S6) and conducted ChIP-qPCR to measure H3K4me2 and H3K4me3 levels in single and double mutants compared to the triple mutant. We found that only *atx5* showed an H3K4me2/3 decrease as a single mutation and in double mutant combinations, although the decrease was significantly smaller than that in the triple mutant (Supplemental Fig. S7). This effect was seen in four of the six loci. It is worth noting that the degrees of H3K4me2 and H3K4me3 reductions in the triple mutant in the ChIP-qPCR experiments were similar to those observed in ChIP-seq. Altogether, these results suggest that *ATX3*, *ATX4*, and *ATX5* act redundantly with *ATX5* having a larger role than *ATX3* and *ATX4*.

#### ATX3/4/5 Activity and Transcription

Distribution of H3K4me2 as well as H3K4me3 generally reflects the distribution of genes (Zhang et al., 2009). Of the H3K4me2 and H3K4me3 regions affected in the *atx3/4/5* triple mutant, 2542 and 2574 corresponded to genes, respectively. It is noted that some H3K4me2 and H3K4me3

sites may overlap two genes, which located in too close proximity to separate. Therefore, the number of genes appeared larger than the corresponding number of H3K4me2 or H3K4me3 sites. To investigate whether any functional groups of genes were disproportionately targeted by *ATX3/4/5*, we performed a gene ontology (GO) category analysis (Du et al., 2010). However, we found that most GO categories were equally affected, suggesting that *ATX3/4/5* targets do not exhibit a significant bias toward any functional gene class (Supplemental Fig. S8).

To investigate the effect of *ATX3/4/5*-mediated chromatin remodeling on gene transcription, we conducted an RNA-seq analysis in wild-type and *atx3/4/5* triple-mutant plants. We identified 1946 differentially expressed genes, of which 932 were up-regulated in the *atx3/4/5* mutant (Supplemental Table S9) and 1014 were down-regulated (Supplemental Table S10). Interestingly, decreases in H3K4me2/3 at individual gene loci were not always accompanied by corresponding reductions in transcript levels (Fig. 3F) and vice-versa (Supplemental Fig. S9). We confirmed these findings using H3K4me2 and H3K4me3 ChIP-qPCR and RT-qPCR on 15 genes selected to represent all

three classes of *ATX3/4/5* target sites (Supplemental Figure S10).

#### Functional Overlap between *ATX3/4/5* versus *ATX1* and *ATX2*

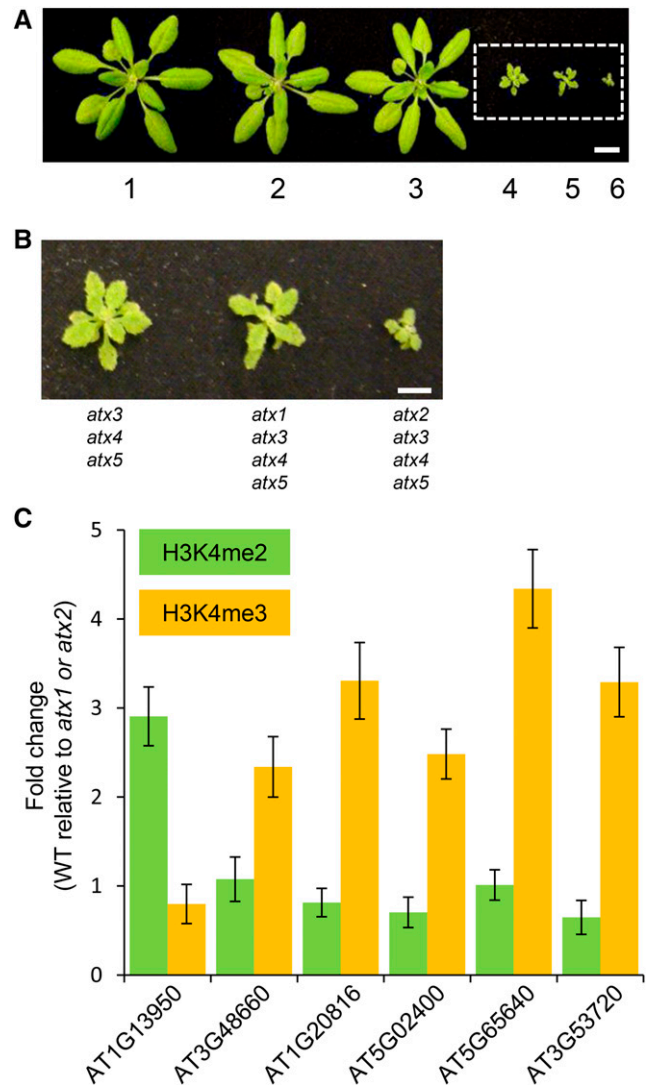
Whereas *ATX3/4/5* affects H3K4me2 as well as H3K4me3, *ATX1* was shown to only affect H3K4me3 and *ATX2* to only affect H3K4me2 (Pien et al., 2008; Saleh et al., 2008). To investigate the overlap between members of the two ATX subgroups, we constructed quadruple mutants, *atx1/3/4/5* and *atx2/3/4/5*. The *atx1/3/4/5* mutant showed a phenotype very similar to that of the *atx3/4/5* triple mutant (Fig. 4, A and B; Supplemental Fig. S11), suggesting that *ATX1* likely acts in the same genetic pathway as *ATX3/4/5*. In contrast, the *atx2/3/4/5* showed an obviously stronger phenotype compared to the *atx3/4/5* triple mutant, suggesting that *ATX2* may be in a pathway separate from *ATX3/4/5*.

We further examined these conclusions by investigating the effects of *atx1* and *atx2* mutations on H3K4 di- and trimethylation at six genomic sites that we previously found to be regulated by *ATX3/4/5* (Supplemental Table S6). ChIP-qPCR analyses showed that *atx1* resulted in a decrease of H3K4me3 at five out of the six sites (Fig. 4C). In contrast, *atx2* exhibited lower H3K4me2 at only one site. These data suggested that most genomic sites targeted by *ATX3/4/5* are also targets of *ATX1*, but not of *ATX2*. Consequently, *ATX1* and *ATX3/4/5* likely function in the same pathway whereas *ATX2* acts separately.

#### *ATX3/4/5*-Affected Genomic Sites Are Distinct from Those Controlled by *SDG2*

Similar to *ATX3/4/5*, *ATXR3/SDG2* also controls H3K4me3 at thousands of sites throughout the genome (Berr et al., 2010; Guo et al., 2010). To investigate whether *ATX3/4/5* act in the same way as *SDG2*, we examined H3K4me3 patterns in the *sdg2* mutant using ChIP-seq. We found 6723 regions exhibiting reduced levels of H3K4me3 in *sdg2* (Fig. 5A), which constituted 46.4% of all H3K4me3 sites in the genome (Supplemental Tables S11 and S12). This number is substantially larger than the number of H3K4me3 sites controlled by *ATX3/4/5*. Moreover, we also found that the *ATX3/4/5*-controlled H3K4me3 sites and the *SDG2*-controlled sites only partially overlapped (Fig. 5A; Supplemental Fig. S12, A–C).

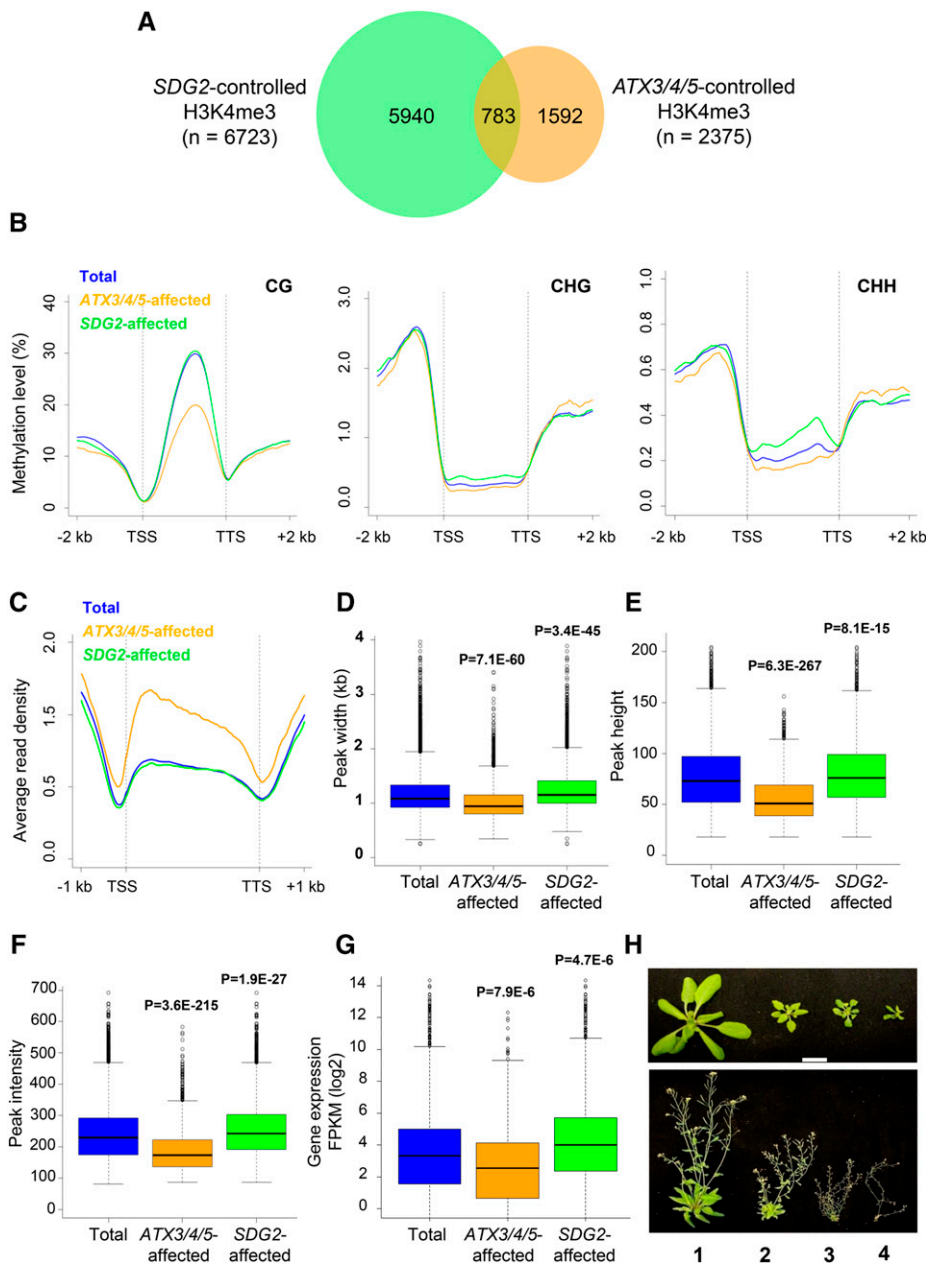
To examine what differentiates *ATX3/4/5* and *SDG2* target sites, we compared their characteristics. We found that H3K4me3 sites targeted by *ATX3/4/5* exhibited considerably low level of gene-body CG methylation (Fig. 5B) and high level of H3K27me3 mark (Fig. 5C). In addition, the H3K4me3 sites targeted by *ATX3/4/5* were remarkably smaller than the sites targeted by *SDG2*, exhibiting reduced peak width, height, and intensity (Fig. 5, D–F). Moreover, genes containing the *ATX3/4/5* target sites showed, on average, lower expression levels than genes overlapping *SDG2*-targeted H3K4me3 sites (Fig. 5G). These findings further indicate that *ATX3/4/5* and *SDG2* function in different ways. It is



**Figure 4.** Functional overlap between *ATX3/4/5* versus *ATX1* and *ATX2*. A, Comparison of phenotypes of 4-week-old plants. 1, wild type; 2, *atx1*; 3, *atx2*; 4, *atx3/atx4/atx5*; 5, *atx1/atx3/atx4/atx5*; 6, *atx2/atx3/atx4/atx5*. Bars = 2 cm. B, Zoom in of triple and quadruple mutants shown in the white box in A. Bars = 1 cm. C, Levels of H3K4me3 in *atx1* mutants (orange) and H3K4me2 in *atx2* mutants (green) at six representative genomic sites examined by ChIP-qPCR. y axis: Fold enrichment of H3K4me2/3 in wild type (WT) relative to *atx1* and *atx2* mutant. Enrichment was first normalized using the TA3 retrotransposon locus (AT1G37113).

noted that the same analysis was performed toward *ATX3/4/5*-targeted H3K4me2 sites, and the results showed a similar pattern as H3K4me3 with the exception that the peak height and intensity of H3K4me2 were higher for *ATX3/4/5*-targeted sites relative to the whole sites in the genome (Supplemental Fig. S13).

To further investigate the functional overlap and differences between *ATX3/4/5* and *SDG2*, we generated *atx3/atx4/atx5/sdg2* quadruple mutants. We were successful in producing *atx3/atx3 atx4/atx4 atx5/atx5 sdg2/+* triple homozygous/single heterozygous mutants. These analyses revealed mutant phenotypes that were more severe than



**Figure 5.** *ATX3/4/5* function in a different pathway from *SDG2*. **A**, Venn diagram showing that the majority of *ATX3/4/5*-affected H3K4me3 sites do not overlap with *SDG2*-affected H3K4me3 sites. **B**, Metagene plots of DNA methylation across gene bodies in three contexts (CG, CHG, and CHH) associated with *ATX3/4/5*- and *SDG2*-affected H3K4me3 sites. **C**, Metagene plots of H3K27me3 across gene bodies associated with *ATX3/4/5*- and *SDG2*-affected H3K4me3 sites. **D**, Boxplot of ChIP-seq peak width of *ATX3/4/5*- and *SDG2*-affected H3K4me3 sites. **E**, Boxplot of ChIP-seq peak height of *ATX3/4/5*- and *SDG2*-affected H3K4me3 sites. **F**, Boxplot of ChIP-seq peak intensity of *ATX3/4/5*- and *SDG2*-affected H3K4me3 sites. Peak intensity is defined as the number of reads in a peak normalized by peak width (reads/kb). **G**, Boxplot of expression levels of genes overlapping *ATX3/4/5*- and *SDG2*-affected H3K4me3 sites. **H**, Phenotype of *atx3/atx4/atx5/sgd2* quadruple mutants in 4-week-old (top) and 6-week-old (bottom) plants. 1, Wild type; 2, *atx3/atx4/atx5* triple mutant; 3, *sdg2*; 4, *atx3/atx4/atx5/sgd2/+*. Bars = 2 cm.

those of single *sdg2* or triple *atx3/4/5* mutants (Fig. 5H). However, we were not able to obtain homozygous quadruple mutants, despite examining a large population of segregating plants (>1000) and multiple combinations of parental genotypes (*atx3/+ atx4/+ atx5/+ sdg2/+*, *atx3/atx3 atx4/+ atx5/atx5 sdg2/+*, and *atx3/atx3 atx4/atx4 atx5/+ sdg2/+*), which further confirms that *ATX3/4/5* and *SDG2* function in separate and complementary genetic pathways.

## DISCUSSION

To elucidate the network controlling chromatin modification by SET domain proteins, we characterized three Arabidopsis H3K4 methyl transferases, *ATX3*, *ATX4*, and

*ATX5*. Our analyses revealed that the three proteins facilitate H3K4 di- as well as trimethylation at a large number of genomic sites. They all act in the same genetic pathway, which is required for both vegetative and reproductive development and is distinct from the previously identified pathway for genome-wide H3K4 trimethylation defined by *ATXR3/SDG2*.

### *ATX3/4/5* Regulate H3K4 Di- and Trimethylation at a Large Number of Genomic Sites

Prior to this study, nearly all H3K4 methyl transferases characterized in Arabidopsis were those controlling H3K4me3 (Berr et al., 2015; Song et al., 2015). The only enzyme described to catalyze H3K4me2 was

ATX2, which only controls a relatively small set of genes (Saleh et al., 2008; Ding et al., 2011a, 2011b). We postulate that ATX3/4/5 are major H3K4 di- and trimethylases acting across the genome based on three lines of evidence. First, the domain structure and sequence placed ATX3, ATX4, and ATX5 firmly within class III SDG proteins known to possess H3K4me2 and H3K4me3 activities (Pien et al., 2008; Saleh et al., 2008). Second, western blot and metagene analyses showed that H3K4me2 and H3K4me3 levels are substantially reduced in the *atx3/4/5* triple mutant. Third, investigations of H3K4me2 and H3K4me3 distribution revealed that levels of these two histone modifications were dramatically reduced at thousands of sites across the genome upon simultaneous disruption of ATX3, ATX4, and ATX5.

Five ATX proteins are all homologs of Trithorax protein, which is a large multifunctional protein including a SET domain, PHD fingers, and PWWP domains (Fig. 1B). The SET domain in Trithorax protein has been shown to have Lys methyltransferase activity with substrate specificity for histone H3K4 (Smith et al., 2004; Tie et al., 2014). However, the mechanisms of how the specificity of these Trithorax proteins conferred remain largely unknown.

Our data further show that the action of ATX3, ATX4, and ATX5 exerts a substantial effect on the transcriptome. Interestingly, not all affected transcripts contain ATX3/4/5-dependent H3K4me2/3 sites. However, a simple explanation for this phenomenon could be that the genes lacking these sites represent downstream effects of genes directly targeted by ATX3/4/5. Alternatively, it is likely that not all the changes in gene expression are necessarily accompanied with the alterations in H3K4me2/3 marks. Furthermore, it is also possible that the other chromatin states are more or less altered in the *atx3/4/5* mutant plants, leading to the secondary effects on gene expression.

### Functional Redundancies of ATX Genes

Our genetic experiments showed that ATX3, ATX4, and ATX5 act redundantly. Further analyses using ChIP-qPCR suggested that they operate in a dosage-dependent manner and their effects are additive, although ATX5 exhibits a stronger effect than ATX3 and ATX4. The stronger effect could be a result of ATX5 being expressed at a higher level than ATX3 and ATX4 (Fig. 1, C and D). We should note, however, that our results do not exclude a possibility that there are sites in the genome where the relationship of ATX3, ATX4, and ATX5 is different from these six sites that we examined.

We also examined the functional overlap between ATX3/4/5 and other proteins from different clades of the ATX family tree. Phenotypic analyses and ChIP-qPCR experiments indicated that ATX2 acts independently from ATX3/4/5 in regulating H3K4me2. In contrast, our evidence suggests that the ATX1 function partially overlaps with the function of ATX3/4/5 in controlling H3K4me3. It is possible that the ATX1 protein acts in the same complex as ATX3/4/5. Alternatively, ATX1 and ATX3/4/5 may target the same genomic sites. The latter possibility seems

more likely, based on the fact that the functions of ATX1 and ATX3/4/5 overlap only partially, as evidence particularly by the less severity of the *atx1* mutant phenotype compared to *atx3/4/5*.

Our biochemical and genetic studies also demonstrated that ATX3/4/5 act in a pathway different from the one defined by SDG2. First, ATX3/4/5 and SDG2 affect largely different sets of genomic target sites. Second, the ATX3/4/5-controlled H3K4me3 sites show distinct features compared to the sites targeted by SDG2, which also suggests that the activities of ATX3/4/5 and SDG2 are controlled by different mechanisms. In particular, we found that ATX3/4/5-targeted genic loci have fairly lower level of the gene-body CG methylation and higher level of H3K27me3 than SDG2 targets, raising the questions of why and how such characteristics are conferred. It has been known that, in general, the activity of gene transcription is positively correlated with the gene-body methylation and negatively correlated with H3K27me3 mark (Zhang et al., 2006; Kouzarides, 2007; Wang et al., 2015; Yang et al., 2016). Indeed, we found that the expression of genes for the ATX3/4/5-targeted loci is significantly lower than SDG2 targets. Therefore, it seems that ATX3/4/5 and SDG2 can target two classes of genic sites with differential gene transcription activity. The next intriguing question raised is how such association of is fulfilled. Recently, Song et al. (2015) have shown that the H3K4 methyltransferase complex can directly interact with specific transcription factor to regulate gene expression in Arabidopsis. Therefore, we speculate that ATX3/4/5 and SDG2 may interact with different transcription factors to facilitate the gene expression and generate the deposition of H3K4me3 mark at specific loci. In addition, two recent studies in mammals found that width of H3K4me3 sites is associated with specific cell identities and transcriptional profiles (Benayoun et al., 2014; Lei et al., 2015). Therefore, it is likely that ATX3/4/5 and SDG2 diverged to modulate transcriptional programs in different tissues or at different stages of development. SDG2 may play a larger role than ATX3/4/5 in facilitating transcription of genes necessary for leaf development, the stage at which tissue samples were collected for transcriptome analyses in this study. In other tissues, such as flowers, where ATX3/4/5 are more abundantly expressed than SDG2, ATX3/4/5 may be more critical. It would be very interesting to compare transcriptome profiles and assess the relative requirement for ATX3/4/5 versus SDG2 in several different tissue types. Finally, our *in vitro* experiments, while not completely conclusive, suggest that ATX3/4/5 may require different cofactors or posttranslational modifications than SDG2.

### CONCLUSION

We examined three until-now-uncharacterized SET-domain-containing Arabidopsis histone methyltransferases, ATX3, ATX4, and ATX5. The three proteins share ancient origin, predating the emergence of monocots and dicots, but have diverged from each other more



recently. They also have similar domain structure as well as expression patterns. They are redundantly required for plant vegetative and reproductive development, and their absence results in profound mutant defects. Unlike most other characterized SET domain proteins except SDG2, they act genome wide. Interestingly, the genomic sites affected by ATX3, ATX4, and ATX5 are largely different from those affected by SDG2. This observation suggests that ATX3, ATX4, and ATX5 define a novel major pathway for H3K4 di- and trimethylation that is different from the SDG2-controlled pathway. Our characterization of ATX3, ATX4, and ATX5 completes elucidation of ATX genes in Arabidopsis. Since the ATX gene family is widely conserved across the plant kingdom, this study should enhance the understanding of the occurrence, distribution, and function of H3K4me2 and H3K4me3 in plants.

## MATERIALS AND METHODS

### Plant Materials, Growth Conditions, and Genotyping

Mutants defective in ATX genes, *atx1-2* (Salk\_149002), *atx2-2* (Salk\_117262), *atx3-1* (CS413717, GK-128H01), *atx3-2* (CS412277, GK-143H01), *atx4-1* (SALK\_060156), *atx4-2* (SALK\_006002), *atx5-1* (CS912172, WiscDsLoxHs127\_10D), and *atx5-2* (CS831182, SAIL\_705\_H05) were obtained from the Arabidopsis Biological Resource Center (Columbus, OH). The mutants were backcrossed to wild-type Columbia (Col-0) plants. Plants were grown in 16 h day/8 h night at 22°C day/20°C night. Genomic DNA extraction and genotyping were conducted as described previously (He et al., 2009). PCR primers used for genotyping are listed in Supplemental Table S13.

### Reproductive Development Analyses

Identification of flower development stages was performed according to Smyth et al. (1990) and Alvarez-Buylla et al. (2010). To examine pollen germination, pollen tubes were stained with aniline blue as described by Wang et al. (2011). 4',6'-Diamidino-2-phenylindole staining was performed as described by McCormick (2004). Pollen viability staining was performed as described by Alexander (1969). Paraffin section was performed as described by Dou et al. (2011).

### Generating and Analyzing Transgenic Promoter-GUS Fusion Arabidopsis (*Arabidopsis thaliana*) Plants

Promoter regions of ATX3 (from -1296 to -1 bp), ATX4 (from -1916 to -20 bp), and ATX5 (from -2150 to -1 bp) were PCR-amplified from genomic DNA using primers ATX3-1L/1R, ATX4-1L/1R, and ATX5-1L/1R, respectively, and cloned into the pGEM-T vector (Promega). Resulting inserts were confirmed by sequencing and subcloned into pCAMBIA1305 using *Sall* and *Bam*HI restriction sites. *Agrobacterium tumefaciens* strain GV3101 was used to transform the construct into Arabidopsis plants by floral dipping (Clough and Bent, 1998). Transgenic plants were selected using 25 µg/mL hygromycin. Histochemical GUS assays were performed on twenty independent T2 transgenic plants for each construct as previously described (Jefferson et al., 1987).

### RNA Extraction and RT-PCR

RNA was extracted from 3-week-old rosette leaves using the RNeasy Plant Mini kit (Qiagen). cDNA was synthesized from 5 µg of total RNA using the ProtoScript First Strand cDNA Synthesis Kit (New England Biolabs) following manufacturer's instructions. RT-PCR reactions were performed as previously described (Zhang et al., 2015) using primers listed in Supplemental Table S13. Transcript levels were estimated using the comparative Ct (threshold cycle) method utilizing *ACTIN2* as an internal control for data normalization. Data shown in Figure 1D are averages of three independent experiments.

### Protein Expression and Western Blotting

C-terminal sequences of ATX3, ATX4, and ATX5 were cloned into the pET28a expression vector (Novagen). The constructs were expressed in *E. coli* strain BL21 (DE3). Recombinant proteins were purified with the Midi PrepEase kit (EMD Biosciences) using His tags. Western blotting experiments were performed as previously described (Jackson et al., 2004). The antibodies used were anti-H3 (Millipore, 05-499), anti-H3K4me1 (Millipore, 07-436), anti-H3K4me2 (Millipore, 07-030), anti-H3K4me3 (Millipore, 07-473), anti-H3K9me2 (Millipore, 07-441), anti-H3K27me3 (Millipore, 07-449), and anti-H3K36me3 (Millipore, 05-801).

### ChIP-seq Library Construction and Sequencing

ChIP was performed as previously described (Xue et al., 2015). Illumina libraries were constructed using the ChIP-seq DNA Sample Prep Kit (Illumina) according to manufacturer's protocol. Libraries were sequenced using the Single-end Cluster Generation and 100-cycle Sequencing Kits (Illumina) on the Illumina HiSeq2500 sequencing system according to manufacturer's instructions.

### Processing and Mapping Illumina Reads to the Genome

Base calling and read quality control were performed following the standard Illumina protocol. Reads passing quality control were aligned to the Arabidopsis genome (TAIR10; www.arabidopsis.org) using Bowtie v1.0.1 allowing for no more than two mismatches (Langmead et al., 2009). Only reads mapped uniquely to the genome were used for further analysis. SAMtools v1.2 was used to convert the mapped reads into BAM files (Li et al., 2009).

### Peak Calling

To identify peaks in ChIP-seq datasets, MACS (version 1.4.2) was used with a matching number of reads in treatment and control (Zhang et al., 2008). Peak calling was performed with following parameters: bandwidth = 150 nucleotides; m-fold = 5 to 50; *P* value cutoff = 1e-05. ATX3/4/5- or SDG2-dependent sites were identified after two filtering steps. First, H3K4me3 ChIP-seq read peak regions were called using the wild-type sample as treatment and the corresponding mutant sample as control. Next, ChIP-seq read peak regions were called using the wild-type sample as treatment and the corresponding input sample as control. Overlapping regions from the two steps were defined as ATX3/4/5- or SDG2-dependent sites. This procedure was devised to avoid the effects of occasional sequence coverage fluctuations, which affect between 0.2% and 5.8% of all reads (Supplemental Fig. S14; Supplemental Tables S6–S8 and S12).

### Simulation Test

Differences of H3K4me2 and H3K4me3 around TSSs between wild-type and *atx3/4/5* mutant were quantified using simulation method (King et al., 2000). Eighty percent of total H3K4me2/3 sites from wild-type and mutant was randomly shuffled to calculate the average read density of 50 bp up- and downstream of the peak summit 1000 times, and the significance was determined by the paired *t* test.

### RNA-seq Analyses

Polyadenylated mRNA was purified using the Dynabeads mRNA Purification kit (Invitrogen), and library construction was performed with the NEB Next mRNA Library Prep Master Mix Set (New England Biolabs) following manufacturer's instructions. Sequencing was using the Pair-end Cluster Generation kits with 125 cycles on the HiSeq2500 sequencing platform. Quality-controlled reads were aligned to the Arabidopsis genome using TopHat-2.0.10 (Trapnell et al., 2009, 2012). Gene expression levels were normalized into fragments per kilobase of transcript per million mapped reads using Cufflinks-2.2.1 with default settings (Trapnell et al., 2012). Genes exhibiting 2-fold change with FDR < 0.05 between wild type and mutant were considered as differentially expressed genes (Supplemental Tables S9 and S10).

### DNA Methylation Analysis

The BS-Seq datasets were downloaded from a published study and were analyzed according to the method as previously described (Tang et al., 2016).

In brief, the data used for the DNA methylation analysis were downloaded from the Gene Expression Omnibus (GEO; accession no. GSE83802). Clean reads were mapped to TAIR10 using BRAT-BW v2.0.1 with parameters -m 2 -i 0 -a 1000. Copy duplicates were removed and count mapped Cs and Ts at each cytosine of forward and reverse strands of the reference. For the mega gene plots, sliding window methods were used with the window size equal to 500 nucleotides and step 50 nucleotides, and methylation levels of CG, CHG, and CHH context were plotted with gene-coding regions plus 2 kb upstream of the TSS and downstream of the transcription termination site (TTS).

### H3K27me3 Analysis

The H3K27me3 datasets were downloaded from GEO with accession number GSE65329 (Cui et al., 2016) and were analyzed using the same method as described above for H3K4me2/3 data. Analyses span 1 kb upstream of the TSS and the gene body and 1 kb downstream of the TTS.

### Accession Numbers

Sequencing datasets are available for download from the NCBI GEO under series GSE73972. A summary of the datasets is provided in Supplemental Table S1.

### Supplemental Data

The following supplemental materials are available.

**Supplemental Figure S1.** Multiple sequence alignment of Arabidopsis ATX3, ATX4, and ATX5 proteins.

**Supplemental Figure S2.** RT-PCR analysis of *ATX3*, *ATX4*, and *ATX5* expression in homozygous mutant plants.

**Supplemental Figure S3.** Gametophyte phenotype of *atx3/4/5* triple-mutant plants.

**Supplemental Figure S4.** In vitro analyses of HMTase activities of ATX3, ATX4, and ATX5.

**Supplemental Figure S5.** IGV images of representative regions selected from all five chromosomes showing altered patterns of H3K4m2 and H3K4me3.

**Supplemental Figure S6.** IGV images of representative regions showing the overlap of *ATX3/4/5*-targeted H3K4me2 and H3K4me3 sites.

**Supplemental Figure S7.** Levels of H3K4me2/3 in six representative genes verified using ChIP-qPCR.

**Supplemental Figure S8.** GO categories of genes overlapping *ATX3/4/5*-dependent H3K4me2 and H3K4me3 sites.

**Supplemental Figure S9.** Lack of correlation between a decrease in gene transcript levels and H3K4me2/3 levels in the *atx3/4/5* triple mutant.

**Supplemental Figure S10.** Levels of H3K4me2/3 and gene transcript in 15 representative genes verified using ChIP-qPCR (top) and RT-qPCR (bottom).

**Supplemental Figure S11.** Quantification of the fresh weight of 4-week-old plants.

**Supplemental Figure S12.** Representative genomic regions affected by *ATX3/4/5* or *SDG2*.

**Supplemental Figure S13.** Genomic features associated with *ATX3/4/5*-targeted H3K4me2 sites.

**Supplemental Figure S14.** IGV images showing artifactual H3K4me2/3 peaks prior to second filtration of ChIP-seq data.

**Supplemental Table S1.** Summary of sequencing datasets used in this study.

**Supplemental Table S2.** Distribution of H3K4me2 sites in wild-type Arabidopsis.

**Supplemental Table S3.** Distribution of H3K4me3 sites in wild-type Arabidopsis.

**Supplemental Table S4.** Distribution of H3K4me2 sites in the *atx3/4/5* triple mutant.

**Supplemental Table S5.** Distribution of H3K4me3 sites in the *atx3/4/5* triple mutant.

**Supplemental Table S6.** List of ATX3/4/5-dependent H3K4me2/3 sites, class I.

**Supplemental Table S7.** List of ATX3/4/5-dependent H3K4me2/3 sites, class II.

**Supplemental Table S8.** List of ATX3/4/5-dependent H3K4me2/3 sites, class III.

**Supplemental Table S9.** List of genes up-regulated in the *atx3/4/5* triple mutant relative to wild type.

**Supplemental Table S10.** List of genes down-regulated genes in *atx3/atx4/atx5* triple mutant relative to wild type.

**Supplemental Table S11.** Distribution of H3K4me3 sites in the *sgd2* mutant.

**Supplemental Table S12.** List of SDG2-dependent H3K4me3 sites.

**Supplemental Table S13.** List of primers used in the study.

### ACKNOWLEDGMENTS

We thank members of our laboratories for helpful discussions and assistance during this project. We also thank Dr. Ye D (China Agricultural University) for his critical comments on the manuscript.

Received December 22, 2016; accepted May 23, 2017; published May 26, 2017.

### LITERATURE CITED

- Alexander MP (1969) Differential staining of aborted and nonaborted pollen. *Stain Technol* **44**: 117–122
- Alvarez-Buylla ER, Benítez M, Corvera-Poiré A, Chao Cador A, de Folter S, Gamboa de Buen A, Garay-Arroyo A, García-Ponce B, Jaimes-Miranda F, Pérez-Ruiz RV, et al (2010) Flower development. *Arabidopsis Book* **8**: e0127
- Alvarez-Venegas R, 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, Pien S, Saddler M, Witmer X, Grossniklaus U, Avramova Z (2003) ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes. *Curr Biol* **13**: 627–637
- Avramova Z (2009) Evolution and pleiotropy of TRITHORAX function in Arabidopsis. *Int J Dev Biol* **53**: 371–381
- Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R, Schulz I, Reuter G, Aalen RB (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
- Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, Devarajan K, Daugherty AC, Kundaje AB, Mancini E, et al (2014) H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell* **158**: 673–688
- Berr A, McCallum EJ, Ménard R, Meyer D, Fuchs J, Dong A, Shen WH (2010) Arabidopsis SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development. *Plant Cell* **22**: 3232–3248
- Berr A, Shafiq S, Pinon V, Dong A, Shen WH (2015) The *trxG* family histone methyltransferase SET DOMAIN GROUP 26 promotes flowering via a distinctive genetic pathway. *Plant J* **81**: 316–328
- Berr A, Xu L, Gao J, Cognat V, Steinmetz A, Dong A, Shen WH (2009) SET DOMAIN GROUP25 encodes a histone methyltransferase and is involved in FLOWERING LOCUS C activation and repression of flowering. *Plant Physiol* **151**: 1476–1485
- Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**: 735–743
- Cui X, Lu F, Qiu Q, Zhou B, Gu L, Zhang S, Kang Y, Cui X, Ma X, Yao Q, et al (2016) REF6 recognizes a specific DNA sequence to demethylate H3K27me3 and regulate organ boundary formation in Arabidopsis. *Nat Genet* **48**: 694–699

- Ding Y, Avramova Z, Fromm M (2011a) The Arabidopsis trithorax-like factor ATX1 functions in dehydration stress responses via ABA-dependent and ABA-independent pathways. *Plant J* **66**: 735–744
- Ding Y, Avramova Z, Fromm M (2011b) Two distinct roles of ARABIDOPSIS HOMOLOG OF TRITHORAX1 (ATX1) at promoters and within transcribed regions of ATX1-regulated genes. *Plant Cell* **23**: 350–363
- Dou XY, Yang KZ, Zhang Y, Wang W, Liu XL, Chen LQ, Zhang XQ, Ye D (2011) WBC27, an adenosine tri-phosphate-binding cassette protein, controls pollen wall formation and patterning in Arabidopsis. *J Integr Plant Biol* **53**: 74–88
- Du Z, Li H, Wei Q, Zhao X, Wang C, Zhu Q, Yi X, Xu W, Liu XS, Jin W, et al (2013) Genome-wide analysis of histone modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in *Oryza sativa* L. Japonica. *Mol Plant* **6**: 1463–1472
- Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res* **38**: W64–W70
- Fromm M, Avramova Z (2014) ATX1/AtCOMPASS and the H3K4me3 marks: How do they activate Arabidopsis genes? *Curr Opin Plant Biol* **21**: 75–82
- Guo L, Yu Y, Law JA, Zhang X (2010) SET DOMAIN GROUP2 is the major histone H3 lysine [corrected] 4 trimethyltransferase in Arabidopsis. *Proc Natl Acad Sci USA* **107**: 18557–18562
- He Y, Mawhinney TP, Preuss ML, Schroeder AC, Chen B, Abraham L, Jez JM, Chen S (2009) A redox-active isopropylmalate dehydrogenase functions in the biosynthesis of glucosinolates and leucine in Arabidopsis. *Plant J* **60**: 679–690
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755
- Jackson JP, Johnson L, Jasencakova Z, Zhang X, PerezBurgos L, Singh PB, Cheng X, Schubert I, Jenuwein T, Jacobsen SE (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana*. *Chromosoma* **112**: 308–315
- Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**: 556–560
- Jacob Y, Feng S, LeBlanc CA, Bernatavichute YV, Stroud H, Cokus S, Johnson LM, Pellegrini M, Jacobsen SE, Michaels SD (2009) ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat Struct Mol Biol* **16**: 763–768
- Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal x. *Trends in Biochemical Sciences* **23**: 403–405
- Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: Beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* **6**: 3901–3907
- King G, Tomz M, Wittenberg J (2000) Making the most of statistical analyses: Improving interpretation and presentation. *Am J Pol Sci* **44**: 347–361
- Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128**: 693–705
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25
- Lei L, Shi J, Chen J, Zhang M, Sun S, Xie S, Li X, Zeng B, Peng L, Hauck A, et al (2015) Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *Plant J* **84**: 1206–1218
- Lei L, Zhou SL, Ma H, Zhang LS (2012) Expansion and diversification of the SET domain gene family following whole-genome duplications in *Populus trichocarpa*. *BMC Evol Biol* **12**: 51
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079
- Lindroth AM, Park YJ, McLean CM, Dokshin GA, Persson JM, Herman H, Pasini D, Miró X, Donohoe ME, Lee JT, et al (2008) Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. *PLoS Genet* **4**: e1000145
- McCormick S (2004) Control of male gametophyte development. *Plant Cell* **16**(Suppl): S142–S153
- Napsucially-Mendivil S, Alvarez-Venegas R, Shishkova S, Dubrovsky JG (2014) Arabidopsis homolog of trithorax1 (ATX1) is required for cell production, patterning, and morphogenesis in root development. *J Exp Bot* **65**: 6373–6384
- Ng DW, Wang T, Chandrasekharan MB, Aramayo R, Kertbundit S, Hall TC (2007) Plant SET domain-containing proteins: Structure, function and regulation. *Biochim Biophys Acta* **1769**: 316–329
- Pien S, Fleury D, Mylne JS, Crevillen P, Inzé D, Avramova Z, Dean C, Grossniklaus U (2008) ARABIDOPSIS TRITHORAX1 dynamically regulates FLOWERING LOCUS C activation via histone 3 lysine 4 trimethylation. *Plant Cell* **20**: 580–588
- Pontvianne F, Blevins T, Pikaard CS (2010) Arabidopsis histone lysine methyltransferases. *Adv Bot Res* **53**: 1–22
- Saleh A, Alvarez-Venegas R, Yilmaz M, Le O, Hou G, Sadler M, Al-Abdallat A, Xia Y, Lu G, Ladunga I, et al (2008) The highly similar Arabidopsis homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions. *Plant Cell* **20**: 568–579
- Schuettengruber B, Martinez AM, Iovino N, Cavalli G (2011) Trithorax group proteins: Switching genes on and keeping them active. *Nat Rev Mol Cell Biol* **12**: 799–814
- Smith ST, Petruk S, Sedkov Y, Cho E, Tillib S, Canaani E, Mazo A (2004) Modulation of heat shock gene expression by the TAC1 chromatin-modifying complex. *Nat Cell Biol* **6**: 162–167
- Smyth DR, Bowman JL, Meyerowitz EM (1990) Early flower development in Arabidopsis. *Plant Cell* **2**: 755–767
- Song ZT, Sun L, Lu SJ, Tian Y, Ding Y, Liu JX (2015) Transcription factor interaction with COMPASS-like complex regulates histone H3K4 trimethylation for specific gene expression in plants. *Proc Natl Acad Sci USA* **112**: 2900–2905
- Springer NM, Napoli CA, Selinger DA, Pandey R, Cone KC, Chandler VL, Kaeppler HE, Kaeppler SM (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
- Tamada Y, Yun JY, Woo SC, Amasino RM (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
- Tang K, Lang Z, Zhang H, Zhu JK (2016) The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nat Plants* **2**: 16169
- Thorstensen T, Grini PE, Aalen RB (2011) SET domain proteins in plant development. *Biochim Biophys Acta* **1809**: 407–420
- Tie F, Banerjee R, Saiakhova AR, Howard B, Monteith KE, Scacheri PC, Cosgrove MS, Harte PJ (2014) Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. *Development* **141**: 1129–1139
- Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**: 562–578
- Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, Reuter G (1994) The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J* **13**: 3822–3831
- van Dijk K, Ding Y, Malkaram S, Riethoven JJ, Liu R, Yang J, Laczko P, Chen H, Xia Y, Ladunga I, et al (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
- Veerappan CS, Avramova Z, Moriyama EN (2008) Evolution of SET-domain protein families in the unicellular and multicellular Ascomycota fungi. *BMC Evol Biol* **8**: 190
- Wang H, Beyene G, Zhai J, Feng S, Fahlgren N, Taylor NJ, Bart R, Carrington JC, Jacobsen SE, Ausin I (2015) CG gene body DNA methylation changes and evolution of duplicated genes in cassava. *Proc Natl Acad Sci USA* **112**: 13729–13734
- Wang W, Wang L, Chen C, Xiong G, Tan XY, Yang KZ, Wang ZC, Zhou Y, Ye D, Chen LQ (2011) Arabidopsis CSLD1 and CSLD4 are required for cellulose deposition and normal growth of pollen tubes. *J Exp Bot* **62**: 5161–5177
- Xu L, Zhao Z, Dong A, Soubigou-Tacconat L, Renou JP, Steinmetz A, Shen WH (2008) Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. *Mol Cell Biol* **28**: 1348–1360
- Xue M, Long J, Jiang Q, Wang M, Chen S, Pang Q, He Y (2015) Distinct patterns of the histone marks associated with recruitment of the

- methionine chain-elongation pathway from leucine biosynthesis. *J Exp Bot* **66**: 805–812
- Yang M, Wang X, Huang H, Ren D, Su Y, Zhu P, Zhu D, Fan L, Chen L, He G, et al** (2016) Natural variation of H3K27me3 modification in two *Arabidopsis* accessions and their hybrid. *J Integr Plant Biol* **58**: 466–474
- Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE** (2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol* **10**: R62
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al** (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137
- Zhang C, Pang Q, Jiang L, Wang S, Yan X, Chen S, He Y** (2015) Dihydroxyacid dehydratase is important for gametophyte development and disruption causes increased susceptibility to salinity stress in *Arabidopsis*. *J Exp Bot* **66**: 879–888
- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al** (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**: 1189–1201
- Zhao Z, Shen WH** (2004) Plants contain a high number of proteins showing sequence similarity to the animal SUV39H family of histone methyltransferases. *Ann N Y Acad Sci* **1030**: 661–669