






# Genome-wide probing of histone H3 lysine 9 acetylation and dimethylation in Arabidopsis reveals correlation between multiple histone marks and gene expression

Junli Zhou  Xiangfeng Wang  Kun He   
Jean-Benoit F. Charron  Axel A. Elling   
Xing Wang Deng

Received: 9 March 2009 / Accepted: 20 December 2009  
Springer Science+Business Media B.V. 2010

**Abstract** Lysine residue 9 of histone H3 can either be acetylation (H3K9ac) and dimethylation (H3K9me2) in acetylated or mono-, di-, or tri-methylated. These epigenetic states have a diverse impact on regulating gene transcriptional activity and chromatin organization. H3K9ac is targeted 5,206 non-transposable element (non-TE) genes and 321 transposable elements (TEs), whereas H3K9me2 is targeted 2,281 TEs and 1,112 non-TE genes. H3K9ac was invariably correlated with transcriptional activation, whereas H3K9me2 has been reported to be mainly located at the 5' end of genes and peaked at the ATG position, while H3K9me2 tended to span the entire gene body. H3K9ac correlated with high gene expression, while H3K9me2 correlated with low expression. Analyses of H3K9ac and H3K9me2 with the available datasets of H3K27me3 and DNA methylation revealed a correlation between the occurrence of multiple epigenetic modifications and gene expression. Genes with H3K9ac alone were actively transcribed, while genes that were also modified by either H3K27me3 or DNA methylation showed a lower expression level, suggesting that a combination of repressive marks weakened the positive regulatory effect of H3K9ac. Furthermore, we observed a significant increase of the H3K9ac modification level of selected target genes in *hda19* (histone deacetylase 19) mutant seedlings, which indicated that HDA19 plays an important role in regulating the level of H3K9ac and thereby influencing the transcriptional activity in young seedlings.

Junli Zhou and Xiangfeng Wang contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-009-9594-7) contains supplementary material, which is available to authorized users.

J. Zhou · X. Wang · X. W. Deng  
National Institute of Biological Sciences, 7 Science Park Road,  
Zhongguancun Life Science Park, 102206 Beijing,  
People's Republic of China  
e-mail: Junli\_zhou@hotmail.com

X. Wang  
e-mail: Xiangfeng.Wang@yale.edu

J. Zhou · X. Wang · K. He · J.-B. F. Charron  
A. A. Elling · X. W. Deng (&)  
Department of Molecular, Cellular and Developmental Biology,  
Yale University, New Haven, CT 06520, USA  
e-mail: xingwang.deng@yale.edu

K. He  
e-mail: hekun78@hotmail.com

J.-B. F. Charron  
e-mail: jeanbenoit.frenettecharron@yale.edu

A. A. Elling  
e-mail: axel.elling@yale.edu

K. He · X. W. Deng  
Peking-Yale Joint Center of Plant Molecular Genetics  
and Agrobiotechnology, College of Life Sciences,  
Peking University, 100871 Beijing, People's Republic of China

**Keywords** Arabidopsis Epigenetics  
Histone modification H3K9ac H3K9me2

## Introduction

Nucleosomes are composed of 146 bp of DNA and a histone core, which is an octamer consisting of two copies each of histones H2A, H2B, H3 and H4. All four histones can be modified at specific residues, particularly their amino (N) terminal tails. Among these post-translational

modifications, acetylation and methylation are the two most commonly studied forms. Generally, acetylation of lysine residues of the N-terminal tails of histones H3 and H4 is associated with transcriptional activation, while mono-, di- and tri-methylation occurring at lysine or arginine residues is one of the most characterized epigenetic marks associated with active transcription and has been shown to influence numerous developmental and biological processes in higher plants (Benhamed et al. 2006; Ausio et al. 2004; Ng et al. 2006). However, a systematic identification of H3K9ac-associated regions in the entire genome is still lacking. Lysine residue 9 of histone H3 can either be acetylated or mono-, di-, or tri-methylated. These epigenetic states

Recently, a number of studies reported efforts on the global identification of DNA methylation and histone modifications in several model organisms. So far, about 20 modifications have been examined in yeast only (Kurdistani et al. 2004; Pokholok et al. 2005; Bernstein et al. 2002; Kurdistani et al. 2004; Wires et al. 2005; Pokholok et al. 2005; Rao et al. 2005; Sinha et al. 2006; Millar and Grunstein 2006; Liu et al. 2005). In higher eukaryotes, chromatin immunoprecipitation (ChIP)-Chip as well as ChIP-seq methods have been utilized to examine DNA methylation and histone modification patterns (Eckhardt et al. 2006; Schübeler et al. 2004; Bernstein et al. 2005; Kondo et al. 2004; Roh et al. 2004, 2005, 2006; Barski et al. 2007; Bernatavichute et al. 2008). These studies have greatly improved our understanding of the relationships between chromatin modifications/structure and gene expression. Furthermore, the availability of these mapping data allows us to progressively decipher the histone code by examining the combinatorial effects of these epigenetic marks (Liu et al. 2005; Agaloti et al. 2002; Kouzarides, Jenuwein and Allis 2001).

In *Arabidopsis* several repressive histone methylation

marks have been reported: H3K9me1/2, H3K27me1/2 and H4K20me1 preferentially associate with the heterochromatic chromocenters (Fuchs et al. 2006). Additionally

H3K9me2 was found to be localized in the heterochromatic knob region of chromosome 4, as well as in transposable elements (TEs) and related repeats (Lippman et al. 2004;

Turck et al. 2007), and H3K9me2 showed a higher than average level in pericentromeric/centromeric regions than in euchromatin regions (Bernatavichute et al. 2008). ChIP-chip analysis of H3K9me3 and H3K27me3 showed that these marks are mostly found in euchromatic regions of chromosome 4 indicating transcriptional repression of genes (Turck et al. 2007). Moreover, genome-wide approaches were used to map DNA methylation and H3K27me3 (Zhang et al. 2006, 2007; Zilberman et al. 2007), and it has been demonstrated that these two epigenetic marks play important roles in the regulation of gene expression and chromatin structure.

#### Materials and methods

#### Plant materials and growth conditions

*Arabidopsis* ecotype Columbia (Col-0) was the genetic background of all the lines used. T-DNA knockout lines for the HDA19 (At4g38130) gene were obtained from the Salk Institute Genome Analysis Laboratory (SALK\_139445.43.60.X), <http://signal.salk.edu> (Alonso et al. 2003). Surface-sterilized seeds were sown on sterile half strength Murashige and Skoog medium (Sigma-Aldrich). Seeds were stratified for 2 days at 4°C, then germinated and grown for 6 days at 22°C under continuous light as previously described (Feng et al. 2004). A 6-day-old seedling from three independent biological replicates was collected for further analysis.

Chromatin immunoprecipitation and probe labeling Wilcoxon-Signed-Rank-Test (Hollander and Wolf [1999](#)). We performed this test on paired data points (ChIP/Input) within a sliding window of  $\pm 1$  kb across each probe to determine whether hybridization intensities of the ChIP-3 and 10 $\mu$ L, respectively, of the following antibodies were enriched DNA samples were significantly greater than those of the total genomic DNA input samples. The algorithm of the Wilcoxon-signed-rank-test is described in Affymetrix's statistical algorithms description document. ChIP DNA and input DNA samples were amplified and labeled as previously described (Lee et al [2007](#)). An aliquot of untreated sonicated chromatin equal to the amount of ChIP-DNA was reverse cross-linked, amplified and used as total input DNA control. The values generated by the Wilcoxon-signed-rank-test were adjusted by the Benjamini and Hochberg false discovery rate procedure (Benjamini and Hochberg [1995](#)). A modified genomic region was defined by combining adjacent probes with a significance threshold of FDR-adjusted value less than 0.05, allowing a maximal gap of 200 bp and requiring a minimal run of two consecutive probes. We concluded that a gene is carrying a particular modification if we detected an overlap of at least 250 bp over this gene and a modified region. The mapping results of H3K9ac and H3K9me2 in this study are publicly available at <http://plantgenomics.biology.yale.edu> to characterize the relationships between epigenetic modifications and transcriptional activity we used a previously published expression dataset (Jiao et al [2005](#)). Raw and processed ChIP-Chip data were deposited in GEO (accession number GSEXXXX).

Chromatin immunoprecipitation (ChIP) was performed with 1.5 g of seedlings according to Bowler et al [2004](#). A 3 and 10 $\mu$ L, respectively, of the following antibodies were used: anti-H3K9ac (Upstate 07-352) and anti-H3K9me2 (Upstate 07-441). ChIP DNA and input DNA samples were amplified and labeled as previously described (Lee et al [2007](#)). An aliquot of untreated sonicated chromatin equal to the amount of ChIP-DNA was reverse cross-linked, amplified and used as total input DNA control.

Tiling microarray design The Arabidopsis genome assembly release 5.0 by TAIR was used to design the tiling-path microarray. Probes were selected using the NASA Oligonucleotide Probe Selection Algorithm (NOPSA; Stolc et al [2005](#)) and arrays were printed with the NimbleGen Maskless Array Synthesis system. A total of 390,580 36 bp probes, excluding highly repetitive regions, positioned on average every 250 bp along one strand of the genome were synthesized on a single array. This tiling array platform can be found in the NCBI gene expression omnibus (GEO) database with accession ID GPLXXXX.

Probe hybridization Microarray chips were prehybridized with prehybridization buffer (5 $\times$  SSC, 0.1% SDS, 1% BSA) at 42 for 1 h. Arrays were then washed twice with water and hybridized with labeled DNA in hybridization buffer (5 $\times$  SSC, 0.1% SDS, 0.5 mg/mL BSA, 0.5 mg/mL ssDNA) for 14 h at 50 C. After hybridization, the arrays were washed in 2 $\times$  SSC and 0.1% SDS, in 0.2 $\times$  SSC and 0.1% SDS and in 0.2 $\times$  SSC at room temperature for 10 min each and dried by centrifugation (1,000 rpm, 1 min). Microarrays were scanned with a GenePix 4200A scanner (Axon), and independent TIFF images for both Cy-3 and Cy-5 channels were acquired at a 5 $\mu$ m resolution.

Microarray data analysis First, microarray raw data were processed by the LOWESS normalization module in the LIMMA R package (Smyth et al. [2005](#)) to balance the hybridization intensities for biases between Cy3 and Cy5 fluorescence. As a second step, the Quantile normalization module in the NMP package (Wang et al [2006](#)) was used to remove the level of experimental variances across all the replicate arrays and samples. Identification of genomic regions enriched for H3K9ac or H3K9me2 modified histones was based on the

ChIP-PCR validation The ChIP-DNA and input DNA samples were resuspended in 30  $\mu$ L water, and 1 $\mu$ L was used for PCR amplification with specific primers as listed in additional Table 2. PCR conditions were as follows: 94 $^{\circ}$ C for 2 min, 28 $\times$  cycles at 94 $^{\circ}$ C for 1 min, 58 $^{\circ}$ C for 1 min, 72 $^{\circ}$ C for 1 min, followed by 72 $^{\circ}$ C for 10 min.

Quantitative real-time PCR ChIP-DNA samples and cDNA samples (synthesized from total RNA) were analyzed by quantitative real-time PCR. Primers were designed to amplify 100 $\times$ 300 bp DNA fragments (additional data Tables 3 and 4). Quantitative PCR was performed in 96-well plates using ABI PRISM 7500 fast real-time PCR system using a SYBR PreMix Ex Taq kit (Takara). Cycling conditions were as follows: 10 s at 95 $^{\circ}$ C, 40 cycles of 3 s at 95 $^{\circ}$ C, 34 s at 60 $^{\circ}$ C, 34 s at 72 $^{\circ}$ C, followed by a 60 $^{\circ}$ C dissociation protocol. The enrichment level of gene for ChIP was determined as previously described (Frank et al [2001](#)). For real time PCR analysis actin was used as internal control. The  $\Delta\Delta C_t$  values were calculated relative to reference PCR values (Livak and Schmittgen [2001](#)).

**Results**

2,280 TEs were found to be associated (Fig. 2a, additional data Files 1A, 1B, and 1C). Based on the genome annotation version 7 of Arabidopsis released by TAIR, our tiling array results identified 19.3% non-TE genes (5,206 out of 27,029) positive for H3K9ac and 4.1% non-TE genes (1,112 out of 27,029) positive for H3K9me2.

**Identification of genomic regions associated with H3K9ac and H3K9me2**

Genomic maps of H3K9ac and H3K9me2 were generated using chromatin immunoprecipitation (ChIP) and a custom oligonucleotide tiling array. The custom tiling array contained 390 K 36-mer probes, which covered 82% of the non-repetitive genome of Arabidopsis with an average resolution of 250 bp. Genomic regions associated with H3K9ac and H3K9me2 were identified by sliding a 1,000 bp window across the interrogated regions and testing whether hybridization signals of oligonucleotide sites (ATG). The gene territory was divided into bins of probes within the window were significantly higher when probed with H3K9ac and H3K9me2 ChIP samples than with input DNA. Probes with a significance threshold of  $P < 0.05$  along with adjacent probes within a maximum gap of 200 bp were selected as positive regions.

Using this methodology, we identified 5,853 H3K9ac regions and 4,009 H3K9me2 regions with median sizes of 656 and 506 bp, respectively. These modified regions covered 3.3 and 1.8% of the Arabidopsis nuclear genome. As expected, H3K9ac regions were highly enriched in the euchromatic regions resembling the distribution of genes while H3K9me2 regions were enriched in heterochromatin (Fig. 2b). It was evident that H3K9ac avoided TEs while H3K9me2 was enriched in TEs (Fig. 2a, b).

As expected, H3K9ac regions were highly enriched in the euchromatic regions resembling the distribution of genes while H3K9me2 regions were enriched in heterochromatin (Fig. 2b). It was evident that H3K9ac avoided TEs while H3K9me2 was enriched in TEs (Fig. 2a, b).

Two representative genomic regions of chromosome 2 and one region of chromosome 5 are shown in Fig. 1b to illustrate the distribution of the two modifications for individual genes.

To validate the accuracy of our ChIP-Chip method, we performed conventional ChIP-PCR on a random collection of 45 candidate genomic loci. These loci were either positive or negative for H3K9ac or H3K9me2 or positive for both. The genes targeted by each modification were classified into three groups of high, medium and low expression based on their normalized expression intensities. The genes in each category were aligned at their translation start sites (ATG) and their averaged modification level ratios were plotted. As shown in Fig. 3a, highly expressed genes possessed higher levels of H3K9ac near their translation start sites, while target genes that were expressed at significantly lower levels contained less H3K9ac modification. However, no significant correlation could be put forward between the presence of H3K9me2 and gene expression, since a uniform and low level of H3K9me2 was observed along gene territories of all three types of non-TE genes (Fig. 3b).

**Materials and methods**

5,100 expressed genes with known or predicted functions, 106 expressed genes with unknown functions, and 321 TEs were found to be associated with H3K9ac. This is in stark contrast with H3K9me2, with which 1,031 expressed genes with known or predictable functions, 81 expressed genes with unknown functions, and

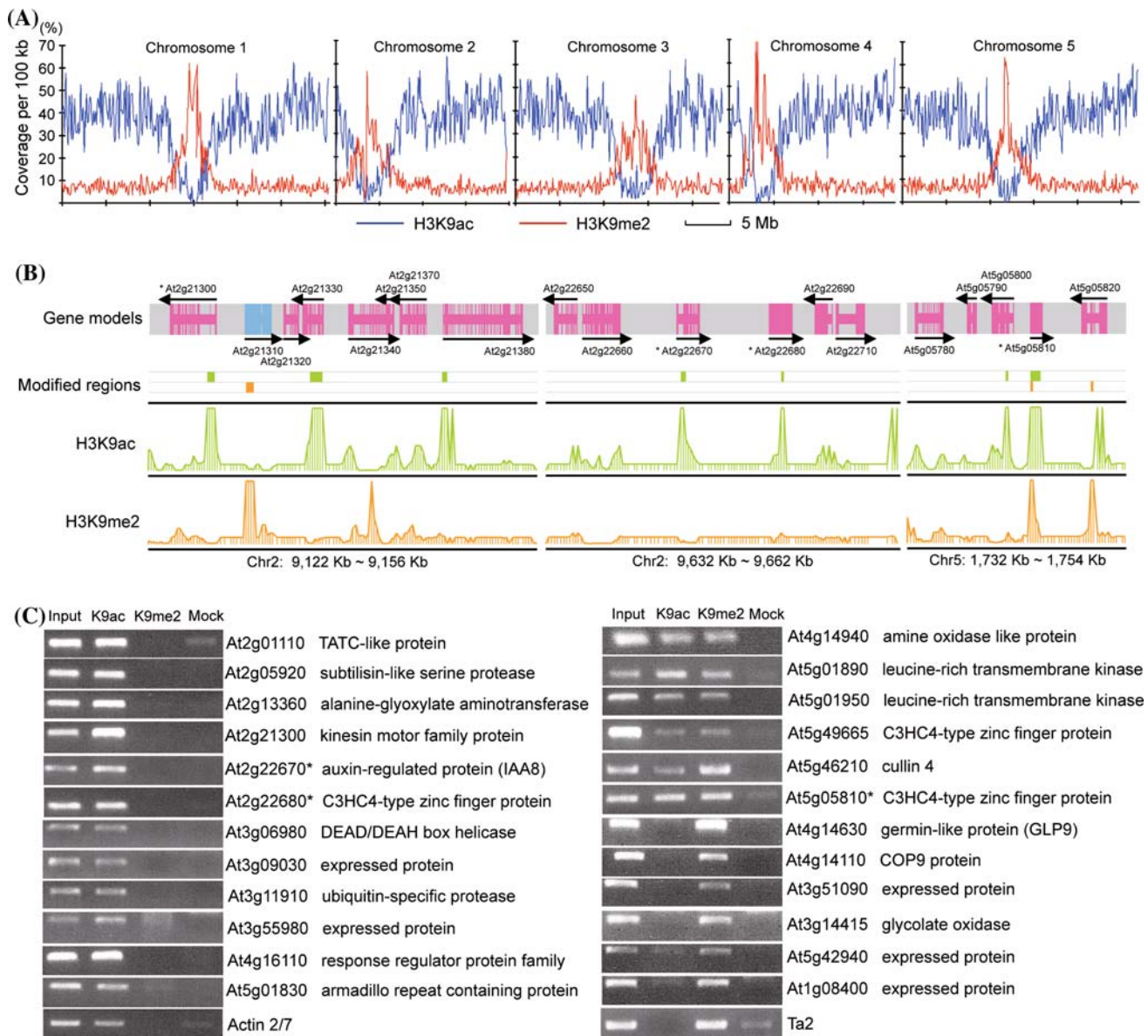


Fig. 1 Genome-wide landscapes of H3K9ac and H3K9me2 modifications across Arabidopsis chromosomes 2 and 5. The two lower tracks show the distribution of H3K9ac (green) and H3K9me2 (orange) signal. The top track shows gene models with TE-related genes in light blue and non-TE genes in pink. (Y-axis) Coverage per 100 kb region. (X-axis) Chromosomes 1-5. (B) Distribution of H3K9ac, H3K9me2, and annotated genes along three representative regions. (C) Conventional ChIP-PCR assays on randomly selected genes to validate the array analysis of target regions.

TE genes that were positive for H3K9ac (5,206 genes) or H3K9me2 (1,112 genes) into three groups: only promoter-modified, only body-modified, as well as both promoter and body-modified. This classification revealed that 64% of the H3K9ac target genes were modified in both body and promoter, 35.4% were only body-modified, and 1% was exclusively promoter-modified (Fig. 1a). In contrast, the H3K9me2 modification was either only in target gene promoters (46.4%) or only in gene bodies (48.8%), while it rarely occurred in both promoter and gene body of the target genes (4.8%) (Fig. 1a). Following this classification, the average expression intensities of each gene group along with a control group composed of unmodified genes were compared for H3K9ac and H3K9me2. This analysis clearly demonstrated that among H3K9ac target-genes, body-modified or both body and promoter-modified genes were expressed at relatively high levels. On the other hand, unmodified genes were expressed at much lower levels (Fig. 1b). Since the group of genes possessing H3K9ac solely in their promoters only contained 28 genes, we did not compare their expression with that of other gene groups (Fig. 1b). Moreover, genes

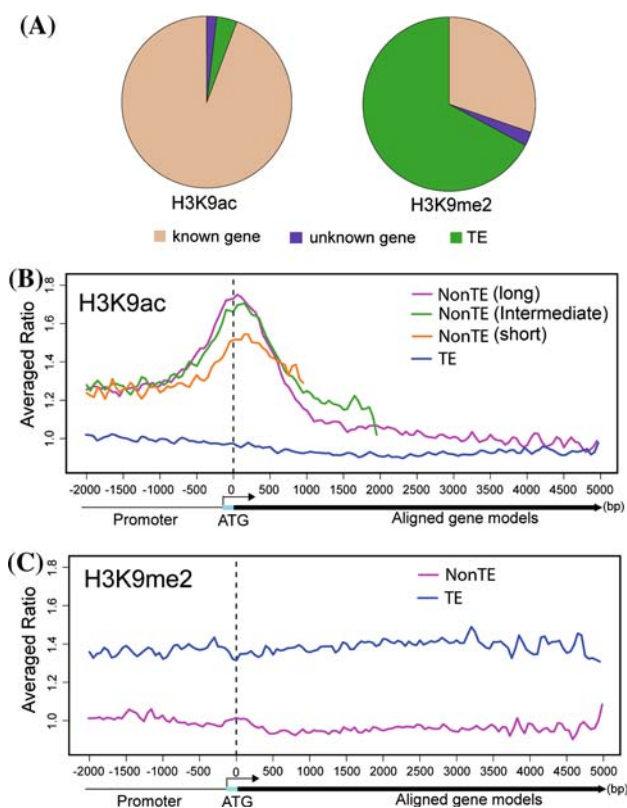


Fig. 2 Distribution of H3K9ac and H3K9me2 within Arabidopsis genes. a Frequencies of H3K9ac (left) and H3K9me2 (right) modified regions in known and unknown non-TE genes and TE-related genes. b Comparison of distributions of H3K9ac within short, intermediate, long non-TE genes and TE genes aligned at their translation start sites (ATG). In this study, a promoter region is defined as the 2,000 bp region upstream of the ATG. The x-axis shows the averaged ratios of the signals from the samples enriched for each modification to the signal from total genomic DNA. The average transcription start site of genes is shown by the 124 bp. c Comparison of the distribution of H3K9me2 within non-TE and TE genes aligned at their translation start sites (ATG). The Y axis shows the averaged ratios of the signals from the samples enriched for each modification to the signal from total genomic DNA.

lacking H3K9me2 had a higher average expression intensity compared to genes harboring this modification, suggesting that H3K9me2 correlates with gene suppression (Fig. 4b).

### Bivalent modification of H3K9ac and H3K9me2 in the same gene

We observed a bivalent modification of specific genes by both H3K9ac and H3K9me2, even though these two modifications were reported to have opposite effects on transcriptional activity (Pfluger and Wagner 2007). To further clarify if the occurrence of one modification affects the occurrence of the other, we compared their target gene sets. On the one hand, we observed that 40.1% (446 out of 1,114) of the non-TE genes targeted by H3K9me2 were H3K9ac-

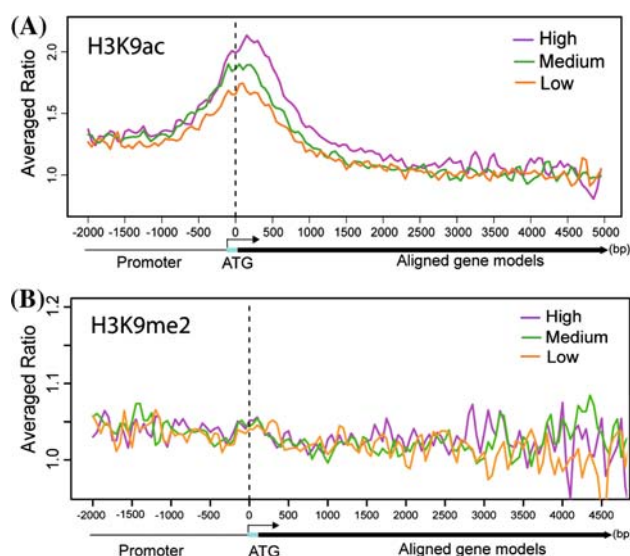


Fig. 3 Relationships between gene expression, H3K9ac and H3K9me2. a Comparison of distributions of H3K9ac within non-TE genes aligned at their translation start sites (ATG) and belonging to different expression intensity categories: low (1,000-3,000), medium (5,000-7,000) and high (9,000-11,000). The x-axis shows the averaged ratios of the signals from the samples enriched for each modification to the signal from total genomic DNA. b Comparison of the distribution of H3K9me2 within non-TE genes aligned at their translation start sites (ATG) and belonging to different expression intensity categories: low (1,000-3,000), medium (5,000-7,000) and high (9,000-11,000). The x-axis shows the averaged ratios of the signals from the samples enriched for each modification to the signal from total genomic DNA.

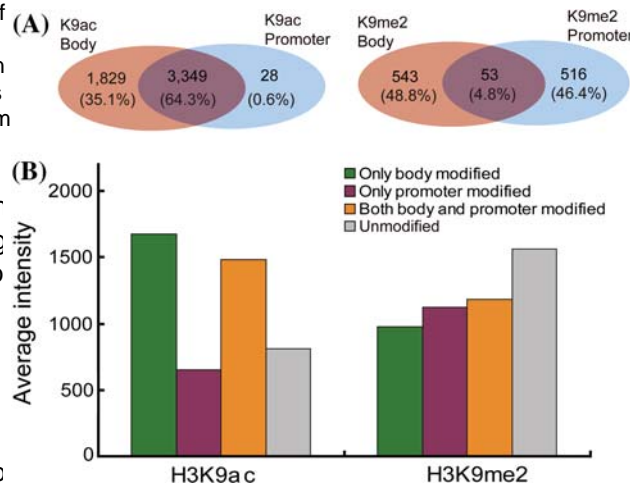
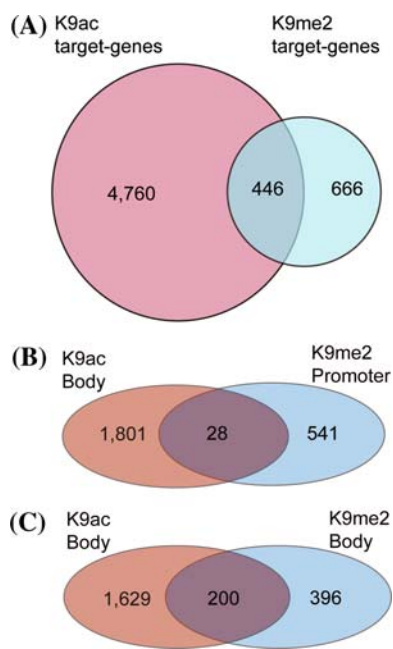


Fig. 4 H3K9ac and H3K9me2 modifications in gene promoter and body regions. a Numbers of “only promoter”, “only body”, “both promoter and body” H3K9ac-modified non-TE genes (left) and H3K9me2-modified non-TE genes (right). b Expression levels of “only promoter”, “only body”, “both promoter and body” modified, and “unmodified” genes containing H3K9ac, or H3K9me2. The Y axis shows the average expression level calculated from a previously published expression dataset (Jiao et al. 2015).



**Fig. 5** Relationship between H3K9ac and H3K9me2 modified genes. **(A)** Venn diagram comparing groups of non-TE genes containing H3K9ac and H3K9me2 modifications. **(B)** Venn diagram comparing groups of non-TE genes that are only body-modified by H3K9ac and promoter-modified by H3K9me2. **(C)** Venn diagram comparing groups of non-TE genes that are only body-modified by H3K9ac and body-modified by H3K9me2.

of the H3K9ac target-genes were also H3K9me2-modified (Fig. 5a). Considering the genic distributions of 19.3% for H3K9ac modified and 4.1% for H3K9me2 reported in this study, these results suggest there is a significant degree of overlap between these two epigenetic marks. However, this study does not provide evidence that these two modifications can coexist at the same nucleosome or histone, but only that they can occur at the same gene.

To further clarify the impact of a possible co-occurrence of these two marks we looked at their respective positions along the gene territories. This analysis revealed that H3K9me2 promoter-modified genes were unlikely to be only body-modified by H3K9ac (Fig. 5b). However, 34% (200 out of 596) of the genes that were H3K9me2 body-modified were also H3K9ac body-modified (Fig. 5c). This suggests that genes are more likely to have both H3K9me2 and H3K9ac body modifications, while genes that are H3K9me2 promoter-modified are less likely to be also H3K9ac body-modified.

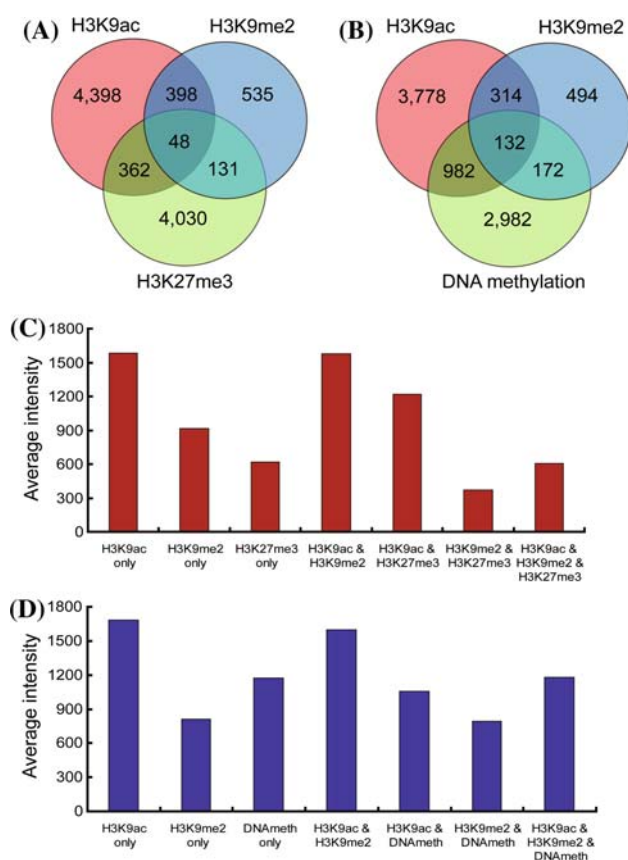
Correlations between combinations of multiple histone modifications and target gene expression status

To investigate a possible interplay among multiple histone modifications, we combined and analyzed the 4,517

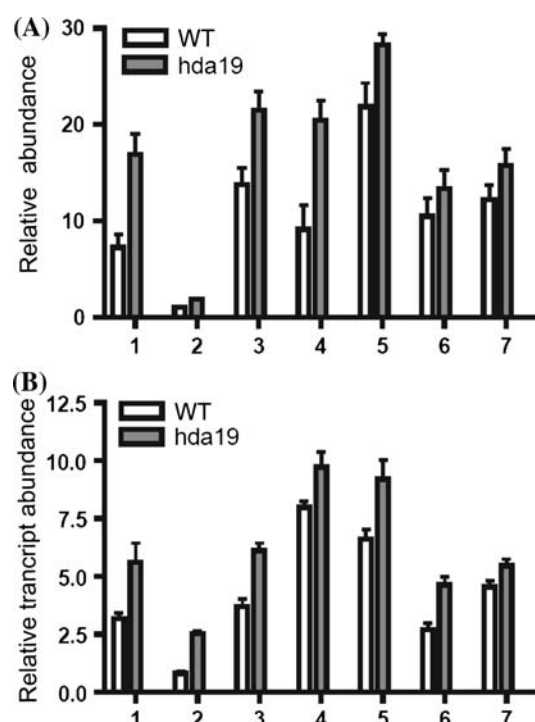
H3K27me3 modified genes identified in 10 to 14-day-old seedlings by Zhang et al (2007) and the 4,268 DNA methylation modified genes identified in the above-ground parts of 25-day-old plants by Zilberman et al (2007), together with our 5,206 H3K9ac modified genes and 1,112 H3K9me2 modified genes identified in 6-day-old seedlings. These previously published studies reported genic distributions of 20% for DNA methylation and 17% for H3K27me3. By comparing these data sets, we found that the fraction of H3K9ac target genes that also contained H3K27me3 was significantly lower (410 out of 5,206; 9.8%) than the average genic distribution (17%) reported for this modification by Zhang et al (2007), while the fraction of H3K9ac target genes that were also DNA methylated (1,114 out of 5,206; 21.4%) was found to be very similar to the value (20%) reported by Zilberman et al (2007) (Fig. 6a, b). On the other hand, the fraction of H3K9me2 target genes overlapping with genes targeted by DNA methylation (304 out of 1,112; 27.3%) was higher than the reported value, while the fraction of H3K9me2 also containing H3K27me3 (179 out of 1,112; 16.1%) was at the genome average (Fig. 6a, b). Collectively, these results indicate a negative correlation between the occurrence of H3K9ac and H3K27me3 and a positive correlation between the occurrence of H3K9me2 and DNA methylation.

It is conceivable that the observed correlations could reflect the potential transcriptional status of a given gene. Thus, we calculated the effects on gene expression of possible combinations of these four epigenetic modifications. Our results indicated that genes that were solely H3K9ac modified were highly expressed. However, lower expression levels were observed when H3K9ac target genes were also targeted by DNA methylation, H3K9me2, or H3K27me3 (Fig. 6c, d). In addition, genes that did not appear an H3K9ac modification mark, but which were possibly modified by a combination of either H3K9me2 and H3K27me3 or H3K9me2 and DNA methylation, were expressed at lower levels compared to genes modified by DNA methylation or by H3K27me3 alone (Fig. 6c, d). These results strongly suggest that H3K9me2, H3K27me3 and DNA methylation can repress the positive effect of H3K9ac on gene expression and that a combination of repressive modifications is probably even more effective.

To address whether decreasing HDA19 expression could influence the level of H3K9ac, we selected seven target genes belonging to different biological processes and performed real-time ChIP-PCR assays in wild type and hda19 seedlings. Our results indicated that decreasing HDA19 expression increased the H3K9ac level of the



**Fig. 6** Combinatorial effects of epigenetic modifications on transcription in Arabidopsis non-TE genes. Numbers of non-TE genes containing H3K9ac, H3K9me2 and H3K27me3 regions and various combinations of these modifications. Numbers of non-TE genes containing H3K9ac, H3K9me2 and DNA methylation regions and various combinations of these modifications. Expression of non-TE genes containing various combinations of H3K9ac and H3K9me2 and H3K27me3. The Y axis shows the average transcript abundance within each group. Expression dataset from Jiao et al. 2009.



**Fig. 7** Histone 3 lysine 9 acetylation state and expression analysis of seven genes in wild type and *hda19* mutant seedlings. **(A)** Histone 3 lysine 9 acetylation state of seven selected genes in Col-0 wild type and *hda19* mutant seedlings. Immunoprecipitates were obtained with an antibody specific to acetylated histone H3 lysine 9 and were analyzed by real-time PCR. Relative amounts of the PCR products were calculated and normalized with respect to the input chromatin. **(B)** Expression analysis of seven selected genes in wild type and *hda19* mutant seedlings. Total RNA was isolated from light-grown seedlings, reverse transcribed, and subjected to quantitative real-time PCR. Relative transcript abundance was calculated and normalized with respect to the actin transcript level. Error bars represent standard deviation (SD) values from at least three repetitions. The seven selected genes are: 1, *At1g65380 (LV2)*; 2, *At4g16280 (FCA)*; 3, *At5g63860 (JVR9)*; 4, *At2g22670 (AA8)*; 5, *At2g01570 (RGA1)*; 6, *At3g23150 (ETR2)*; 7, *At2g26990 (CSN2)*.

## Discussion

seven selected genes (Fig. 7a). To investigate whether the change of the H3K9ac level of these seven genes in *hda19* mutant background correlated with gene expression activity, we conducted an expression analysis of the seven genes under study in wild type and *hda19* mutant seedlings using real-time PCR. Our results showed that the expression levels of these seven selected genes were increased in *hda19* mutants compared to wild type (Fig. 7b), suggesting that HDA19 influences the H3K9ac levels and consequently the expression levels of these genes in Arabidopsis.

H3K9ac is restricted to euchromatin while H3K9me2 is found in both euchromatic and heterochromatic regions. Our work revealed that H3K9ac is widely distributed in the Arabidopsis genome, and shows that H3K9ac targets 5,206 non-TE genes. As expected, H3K9ac was restricted to euchromatin and was associated with active genes (Figs. 2a). This observation is consistent with the fact that histone acetylation has been genetically and biochemically linked to transcriptional activation (Kuo et al. 1998; Wu et al. 2000).

Conversely, the H3K9me2 level was significantly higher in heterochromatin than in euchromatin (Fig. 1a). This distribution pattern is in agreement with the proposed role of methylation system and H3K9 methylation (Lippman et al. 2004; Houben et al. 2003). A similar cooperation between H3K9me2 in heterochromatin formation and/or maintenance (Johnson et al. 2002; Houben et al. 2003; Bernatchi et al. 2008). To our surprise, 1,112 non-TE genes were found to be H3K9me2-modified, suggesting that this epigenetic mark is also an important gene silencing mechanism in plants that regulates a rather large number of genes.

H3K9ac modification correlates with high gene expression but with plant-specific features

A systematic analysis of the locations of lysine 9 acetylated sites revealed a striking correlation with the ends of annotated genes. These data are consistent with recent studies, which documented a similar distribution in yeast, human and mouse species in which the end was the most highly acetylated region within a gene (Kurdistanis et al. 2004; Pokholok et al. 2005; Sinha et al. 2006; Liu et al. 2005; Bernstein et al. 2005; Roh et al. 2004, 2005, 2006). However, in Arabidopsis the level of H3K9ac at the translation start site was correlated to gene length, since short genes presented a lower modification level (Fig. 6a) while longer genes presented higher H3K9ac levels. This is consistent with a previous study, which reported that highly expressed genes tend to have more and longer introns than genes expressed at a low level in both rice and Arabidopsis (Ren et al. 2006). Taken together, this could be a unique plant feature since these observations are in contrast with previous findings in yeast (Sinha et al. 2006) and suggest that even if the general correlation of H3K9ac with transcriptional activation is conserved during species evolution, the mechanism of action and the way this modification is established and maintained differs widely.

Combinatorial effect of multiple modifications on gene expression level

In this study, we found that 40.1% of the genes modified by H3K9me2 (446 out of 1,112) were also modified by H3K9ac (also named H3K9me1) (Fig. 5a) and that the expression levels of these genes were similar to those only bearing H3K9ac (Fig. 6c, d). Interestingly, genes with H3K9me2 but lacking H3K9ac are expressed at much lower levels. Hence, our findings support the presence of a balancing mechanism, in which the dual recruitment of a repressing and a stimulating modification at the H3 lysine residue 9 makes the transcription efficiency more stable and less sensitive to fluctuations. However, when this equilibrium is broken and the repressive mark becomes predominant, the transcriptional activity is repressed.

This balancing act is not the only example of a relationship between two epigenetic marks. It is well

established that chromatin compaction and gene silencing involve a functional relationship between the plant DNA-methylation system and H3K9 methylation (Lippman et al. 2004; Houben et al. 2003). A similar cooperation between methyl-H3K9 and DNA methylation in mammals was found (Sims et al. 2003), while in euchromatic regions or at the individual gene level, results have been less clear (Umlauf et al. 2004; Gartler et al. 2004; Zhao et al. 2005; Zhang et al. 2005; Yan et al. 2001). Our finding that a significant fraction of the H3K9me2 target genes was also DNA-methylated supports a cooperation model between these two repressive marks. This is further reinforced by

the fact that the fraction of H3K9ac target-genes that also contained DNA methylation was within reported values, suggesting that this cooperation is specific to the methylation of lysine residue 9.

Recently, H3K27me3 was reported as an important silencing mechanism independent of DNA methylation and RNA interference that targets 17% of the non-TE Arabidopsis genes (Zhang et al. 2007). Our analysis established that only 9.8% of the H3K9ac target genes were also targeted by H3K27me3 (Fig. 6a), while 16% of the H3K9me2 target genes were found to be H3K27me3-modified (Fig. 6b). This lower than average fraction of H3K9ac target genes also modified by H3K27me3 is not surprising since both modifications have opposite effects on gene transcription and are probably antagonistic to each other. A similar situation was reported for human cells, in which H3K27me3 regions did not overlap with chromatin regions presenting high levels of H3K9ac and H3K14ac (Roh et al. 2006).

HDA19 affects both H3K9ac and gene expression levels

Fifteen histone deacetylase (HDA) genes have been predicted to be encoded in the Arabidopsis genome and only a few were reported to be involved in a specific biological process (Verbsky and Richardson 2001). One of them, HDA19 (also named HD1) is involved in light signaling, growth, and development. Plants expressing a mutant allele of HDA19 exhibit a variety of developmental abnormalities. These abnormalities are believed to be the results of a perturbation of histone acetylation (Wu et al. 2000; Tian and Chen 2001; Tian et al. 2003). Our data showed that the H3K9ac level and expression activity of selected genes were affected in the HDA19 mutants at the seedling stage (Fig. 7), which suggests that HDA19 is mainly responsible for the deacetylation of lysine 9 of histone H3, therefore playing an important role in the regulation of gene expression.

We have mapped the distributions of H3K9ac and H3K9me2 in the Arabidopsis genome at a 250 bp

resolution. The data presented here will help in understanding how H3K9ac and H3K9me2 regulate the expression of individual genes and will facilitate a systematic insight into the overall structure and function of chromatin. The results presented here are consistent with previous findings in yeast, mouse, human and plants, and support a model in which H3K9ac enhances gene expression, and in which H3K9me2 is an important silencing system. Collectively, these results suggest that although certain epigenetic mechanisms might differ between plants and animals, the chromatin distributions and functions of H3K9ac and H3K9me2 seem to be widely conserved.

**Acknowledgments** This work was supported by a grant from the National 863 High-Tech Project of the National Ministry of Science and Technology, People's Republic of China (2003AA 2100701) and the National Institute of Biological Sciences, as well as by a grant from the NSF Plant Genome Program (DBI-0421675).

## References

- Agalioti T, Chen G, Thanos D (2002) Deciphering the transcriptional histone acetylation code for a human gene. *Cell* 111:381–392
- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P et al (2003) Genome-wide insertional mutagenesis in *Arabidopsis thaliana*. *Science* 301:653–657
- Austin I, Alonso-Blanco C, Jarillo JA, Ruiz-García L, Martínez-Zapater JM (2004) Regulation of flowering time by FVE, a retinoblastoma-associated protein. *Nat Genet* 36:162–166
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z et al (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837
- Bernamed M, Bertrand C, Servet C, Zhou DX (2006) *Arabidopsis* GCN5, HD1, and TAF1/HAF2 interact to regulate histone acetylation required for light-responsive gene expression. *Plant Cell* 18:2893–2903
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Statist Soc Ser B* 57:289–300
- Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine 9 methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS ONE* 3:e3156
- Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS et al (2002) Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci USA* 99:8695–8700
- Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ et al (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120:169–179
- Bowler C, Benvenuto G, Laßamme P, Molino D, Probst AV, Tariq M et al (2004) Chromatin techniques for plant cells. *Plant J* 39:776–789
- Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M et al (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 38:1378–1385
- Feng S, Shen Y, Sullivan JA, Rubio V, Xiong Y, Sun TP et al (2004) *Arabidopsis* CAND1, an unmodified CUL1-interacting protein, is involved in multiple developmental pathways controlled by ubiquitin/proteasome-mediated protein degradation. *Plant Cell* 16:1870–1882
- Frank SR, Schroeder M, Fernandez P, Taubert S, Amati B (2001) Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Genes Dev* 15:2069–2082
- Fuchs J, Demidov D, Houben A, Schubert I (2006) Chromosomal histone modification patterns: from conservation to diversity. *Trends Plant Sci* 11:199–208
- Gartler SM, Varadarajan KR, Luo P, Campbell TK, Traynor J, Francke U et al (2004) Normal histone modifications on the inactive X chromosome in ICF and Rett syndrome cells: implications for methyl-CpG binding proteins. *BMC Biol* 2:21
- Hollander M, Wolfe DA (1999) Nonparametric statistical methods. Wiley, New York, pp 228–235
- Houben A, Demidov D, Gernand D, Meister A, Leach CR, Schubert I (2003) Methylation of histone H3 in euchromatin of plant chromosomes depends on basic nuclear DNA content. *Plant J* 33:967–973
- Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293:1074–1080
- Jiao Y, Ma L, Strickland E, Deng XW (2005) Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and *Arabidopsis*. *Plant Cell* 17:3239–3256
- Johnson L, Cao X, Jacobsen S (2002) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* 12:1360–1367
- Kondo Y, Shen L, Yan PS, Huang TH, Issa JP (2004) Chromatin immunoprecipitation microarrays for identification of genes silenced by histone H3 lysine 9 methylation. *Proc Natl Acad Sci USA* 101:7398–7403
- Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128:693–705
- Kuo MH, Zhou J, Jambeck P, Churchill ME, Allis CD (1998) Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. *Genes Dev* 12:627–639
- Kurdistani SK, Tavazoie S, Grunstein M (2004) Mapping global histone acetylation patterns to gene expression. *Cell* 117:721–733
- Lee J, He K, Stolc V, Lee H, Figueroa P, Gao Y et al (2007) Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* 19:731–749
- Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR et al (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430:471–476
- Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N et al (2005) Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol* 3:e328
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔC<sub>T</sub></sup> method. *Methods* 25:402–408
- Millar CB, Grunstein M (2006) Genome-wide patterns of histone modifications in yeast. *Nat Rev Mol Cell Biol* 7:657–666
- Ng DW, Chandrasekharan MB, Hall TC (2006) Ordered histone modifications are associated with transcriptional poising and activation of the phaseolin promoter. *Plant Cell* 18:119–132
- Reuber J, Wagner D (2007) Histone modifications and dynamic regulation of genome accessibility in plants. *Curr Opin Plant Biol* 10:645–652
- Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI et al (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122:517–527
- Rao B, Shibata Y, Strahl BD, Lieb JD (2005) Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol Cell Biol* 25:9447–9459
- Ren XY, Vorst O, Fiers MW, Stiekema WJ, Nap JP (2006) In plants, highly expressed genes are the least compact. *Trends Genet* 22:528–532

- Roh TY, Ngau WC, Cui K, Landsman D, Zhao K (2004) High-resolution genome-wide mapping of histone modifications. *Nat Biotechnol* 22:1013-1016
- Roh TY, Cuddapah S, Zhao K (2005) Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev* 19:542-552
- Roh TY, Cuddapah S, Cui K, Zhao K (2006) The genomic landscape of histone modifications in human T cells. *Proc Natl Acad Sci USA* 103:15782-15787
- Schubeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F et al (2004) The histone modification patterns of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* 18:1263-1271
- Sims RJ III, Nishioka K, Reinberg D (2003) Histone lysine methylation: a signature for chromatin function. *Trends Genet* 19:629-639
- Sinha I, Wern M, Ekwall K (2006) Genome-wide patterns of histone modifications in *Saccharomyces cerevisiae*. *Chromosome Res* 14:95-105
- Smyth GK, Michaud J, Scott HS (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21:2067-2075
- Stolc V, Samanta MP, Tongprasit W, Sethi H, Liang S, Nelson DC et al (2005) Identification of transcribed sequences in *Arabidopsis thaliana* by using high-resolution genome tiling arrays. *Proc Natl Acad Sci USA* 102:4453-4458
- Tariq M, Saze H, Probst AV, Lichota J, Habu Y, Paszkowski J (2003) Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. *Proc Natl Acad Sci USA* 100:8823-8827
- Tian L, Chen ZJ (2001) Blocking histone deacetylation in *Arabidopsis* induces pleiotropic effects on plant gene regulation and development. *Proc Natl Acad Sci USA* 98:200-205
- Tian L, Wang J, Fong MP, Chen M, Cao H, Gelvin SB et al (2003) Genetic control of developmental changes induced by disruption of *Arabidopsis* histone deacetylase 1 (HD1) expression. *Genetics* 165:399-409
- Turck F, Roudier F, Farrona S, Martin-Magniette ML, Guillaume E, Buisson N et al (2007) *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet* 3:e86
- Umlauf D, Goto Y, Cao R, Cerqueira F, Wagschal A, Zhang Y et al (2004) Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nat Genet* 36:1296-1300
- Verbsky ML, Richards EJ (2001) Chromatin remodeling in plants. *Curr Opin Plant Biol* 4:494-500
- Wern M, Silverstein RA, Sinha I, Walfridsson J, Lee HM, Laurenson P et al (2005) Genomewide analysis of nucleosome density histone acetylation and HDAC function in *Saccharomyces cerevisiae*. *EMBO J* 24:2906-2918
- Wu K, Malik K, Tian L, Brown D, Miki B (2000) Functional analysis of a RPD3 histone deacetylase homolog in *Arabidopsis thaliana*. *Plant Mol Biol* 44:167-176
- Yan PS, Chen CM, Shi H, Rahmatpanah F, Wei SH, Caldwell CW et al (2001) Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res* 61:8375-8380
- Zhang Y, Fatima N, Dufau ML (2005) Coordinated changes in DNA methylation and histone modifications regulate silencing/derepression of luteinizing hormone receptor gene transcription. *Mol Cell Biol* 25:7929-7939
- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H et al (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126:1189-1201
- Zhang X, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, Goodrich J et al (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol* 5:e129
- Zhao W, Soejima H, Higashimoto K, Nakagawachi T, Urano T, Kudo S et al (2005) The essential role of histone H3 Lys9 di-methylation and MeCP2 binding in MGMT silencing with poor DNA methylation of the promoter CpG island. *J Biochem* 137:431-440
- Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299:716-719
- Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* 39:61-69