Histone variants and modifications in plant gene regulation
Roger B Deal1 and Steven Henikoff1,2

Genomes are packaged by complexing DNA with histone proteins, which provides an opportunity to regulate gene expression by dynamically impeding access of transcriptional regulatory proteins and RNA polymerases to DNA. The incorporation of histone variants into nucleosomes and addition of post-translational modifications to histones can alter the physical properties of nucleosomes and thereby serve as a mechanism for regulating DNA exposure. Chromatin-based gene regulation has profound effects on developmental processes including regulation of the vegetative to reproductive transition, as well as responses to pathogens and abiotic factors. Incorporation of the histone variant H2A.Z and methylation of histone H3 lysine residues 4 and 27 have emerged as key elements in the regulation of genes involved in each of these processes.

Introduction
The eukaryotic genome is tightly wrapped by histones to form nucleosomes, which must be densely packed to fit within the small volume of the nucleus. Despite these packaging constraints, nucleosomes must be able to allow access of sequence-specific binding proteins and enzymes such as DNA and RNA polymerases to DNA. Nucleosomes are composed of ~150 bp of DNA coiled around an octameric histone core particle containing two copies of each of the four core histone proteins (H2A, H2B, H3, and H4) [1,2]. The properties of a nucleosome can be altered in several ways, including replacement of canonical histones with specialized variant types, post-translational modification of histones, movement of the histone core relative to the underlying DNA sequence, and partial or complete removal of histones from the DNA. Chromatin-based gene regulation is effected by the interplay between sequence-specific DNA binding proteins, histone variants, histone modifying enzymes, chromatin-associated proteins, and ATP-dependent nucleosome remodelers, but how all of these components work together is still unclear [3]. An emerging view is that the composition of a nucleosome in terms of histone variants and histone post-translational modifications (PTMs) dictates its physical stability and propensity to be slid along the DNA or removed completely [4]. In this way, pathways directed at altering these nucleosome characteristics can be used to regulate exposure and occlusion of the DNA. This review will focus on recent advances in understanding histone variants and PTMs and their roles in regulating gene expression during plant development and in response to the environment.

Histone variants
In addition to the major histones, which are deposited during DNA replication, all eukaryotes have variant types of H2A and H3 that are incorporated into chromatin during interphase and can impart unique properties to the nucleosomes they occupy [5]. The H2A variants H2AX and H2A.Z are both found in plants, and multiple isoforms of each are present [6]. H2AX is phosphorylated on its unique C-terminal serine at sites of DNA damage and is involved in orchestrating the DNA repair pathway [7]. The role of this variant in repair pathways has been extensively studied in animals and it is presumed to play a similar role in plants [8], but has not been studied at a functional level.

The H2A.Z variant differs from H2A by many amino acid substitutions throughout the length of the protein, particularly in the C-terminal α-helical region [9]. This variant has been studied in yeast, animals, and plants, and has been implicated in many different genomic processes including transcriptional regulation, maintenance of genome integrity, and the formation of heterochromatin boundaries [10]. Genome-wide profiling of the location of H2A.Z by chromatin immunoprecipitation coupled to microarrays (ChIP-chip) has revealed that this variant is widespread throughout genomes and, as a rule, found in nucleosomes flanking the transcriptional start site (TSS) [11–13], where it appears to play a role in transcriptional regulation, at least partly, by helping to prevent DNA methylation [13,14]. H2A.Z is inserted into nucleosomes by the yeast Swr1 ATP-dependent nucleosome remodeling complex and by related complexes in animals and plants, which partially unwrap the nucleosome and replace an H2A/H2B dimer with an H2A.Z/H2B dimer [15]. Multiple H2A.Z variants and a SWR1-like deposition complex are found in plants, and these regulate
many genes involved in both development and environmental responses, as discussed below.

Two variants of H3 are found in all eukaryotes: CenH3 and H3.3. The CenH3 variant is incorporated at centromeres and is essential for chromosome segregation, as discussed in the review by X et al. (in this issue). The H3.3 variant differs from H3 at only 3–4 amino acids in both plants and animals [16] and is deposited into chromatin outside of DNA replication through different histone chaperones, including HirA and Daxx, depending on the genomic location [17–19]. Deposition of H3.3 into chromatin occurs predominantly within promoters, transcribed regions of expressed genes, and at gene regulatory elements, where nucleosomes are being rapidly disrupted and replaced [20,21].

In animals, H2A.Z and H3.3 show partially overlapping genome-wide distributions in which both variants are enriched near the TSS and in gene body nucleosomes at the 5′ end of expressed genes [20,22]. Nucleosomes containing H2A.Z but not H3.3 are relatively stable, whereas those that contain both variants are prone to disassembly in vivo [23]. These unstable double-variant nucleosomes are found near TSSs and so may modulate exposure of promoter DNA by promoting nucleosome turnover [24]. Multiple isoforms of H3.3 exist in plants, but none of these have yet been mapped genome-wide or studied functionally. However, the Arabidopsis homolog of the H3.3 chaperone HirA has been shown to mediate silencing of KNOX genes during leaf development [25], presumably through H3.3 deposition. Whether this is the result of a direct effect of H3.3 on the KNOX genes themselves is unknown.

**Histone post-translational modifications**

Biochemical studies of histone proteins have shown that they can be extensively modified post-translationally at their N-terminal tails through the addition of acetyl, methyl, phosphoryl, and ADP-ribose groups, as well as peptides such as SUMO and ubiquitin. Genome-wide mapping of these modified histones has revealed that certain groups of histone modifications tend to co-occur in a given region, and each PTM can be broadly categorized as being associated with actively transcribed genes, silenced genes, or transposons [26,27]. In particular, acetylation and methylation of H3 lysine residues play roles in promoting the expression or silencing, as well as switching between these states, of many genes in both plants and animals. With some exceptions, the types of histone modifications that occur, as well as their functions, are generally conserved between plants and animals [28].

Acetylation of histones alters the physical properties of nucleosomes directly by loosening association between histones and DNA, whereas other PTMs, including methylation, often create binding sites for other proteins that have specific effects on chromatin-based processes. In the case of modifications that are bound by specific effector proteins, these can either be involved in the repression of transcription by mechanisms such as compacting nucleosome arrays [29,30], or they can support transcription by recruiting chromatin remodeling complexes, modifying enzymes, or other complexes involved in elongation or splicing [31,32].

Two well-studied H3 methyl modifications are trimethylation of H3 lysine 4 (H3K4me3) and trimethylation of H3 lysine 27 (H3K27me3), which are correlated with transcriptional activation and silencing, respectively. H3K4me3 is enriched at the 5′ end of actively transcribed genes by the enzymatic action of trithorax group (trxG) protein complexes associated with the initiation form of RNA polymerase [33]. Similar distributions of this modification are found in the genomes of Arabidopsis, rice, and maize [34,35,36]. Nucleosomes carrying this modification are recognized by multiple protein complexes including histone acetyltransferases as well as the Chd1 and NURF complexes, which can remodel nucleosomes to support ongoing transcription [37]. Chd1 can also recruit the RNA splicing machinery to facilitate the coordination of transcription and splicing [38]. Thus, trimethylation of H3K4 facilitates transcription, and this effect can be reversed through the targeted action of specific H3K4 demethylase enzymes in order to silence a gene [39].

In contrast to H3K4me3, H3K27me3 is associated with stable silencing of developmentally important genes in both animals and plants. In animals, this modification is catalyzed by the Polycomb Repressive 2 (PRC2) complex and is found over regions many kilobases in size, often covering developmentally regulated genes. This modification is recognized by the Polycomb Repressive 1 (PRC1) complex, which acts to condense chromatin and thereby repress transcription [39,40]. In plants, this modification also appears to be deposited by one or more homologs of the PRC2 complex. Unlike in animals, H3K27me3 is generally restricted to promoters and transcribed regions of individual genes in plants, and is found at roughly 15–20% of genes in Arabidopsis [41,42] and 30–40% of genes in rice and maize [35,36]. Furthermore, H3K27me3 in Arabidopsis is bound by LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), which is related to animal HETEROCHROMATIN PROTEIN 1 (HP1) [43,44]. LHP1 and several plant homologs of animal PRC1 components appear to form a PRC1-like silencing complex [45,46]. In addition, the plant-specific protein EMBRYONIC FLOWER 1 (EMF1) also plays a PRC1-like role in silencing the floral homeotic gene AGAMOUS (AG) [46], suggesting that plants have other mechanisms for PcG silencing in addition to the PRC1 complex itself.
The vegetative to reproductive transition
In recent years, breakthroughs have been made in understanding the roles of H2A.Z and PTMs such as H3K4me3 and H3K27me3 in plant gene regulation. Perhaps the best understood example of the role of plant histone variants and PTMs in gene regulation is that of the Arabidopsis FLOWERING LOCUS C (FLC) gene, which controls the transition from vegetative growth to flowering. The FLC gene is expressed during the vegetative growth phase and acts to repress the transition to flowering by reducing the expression of genes that promote flowering; thus, FLC must be silenced for flowering to occur. The expression of FLC during vegetative growth and its silencing before flowering is an excellent example of the interplay between trxG activators and PcG silencing proteins, as well as the histone variant H2A.Z and other chromatin modifications in both stabilizing and switching between the active and silent states (Figure 1).

During vegetative growth the expression of FLC is promoted through the deposition of H3K4me3 by the Arabidopsis Paf1 complex and other trxG proteins [47–49].

Figure 1

Chromatin-level regulation of the FLC gene and the transition to flowering. A simplified depiction is shown of the main chromatin factors that promote activation and silencing of FLC to control the developmental phase of the plant. In the vegetative growth state, represented in the left panel, the FLC gene is expressed and the encoded protein represses genes that initiate flowering. Chromatin at the FLC locus is depicted at top, where brown circles represent nucleosomes and the TSS is shown as a black flag on the gene. FLC chromatin is made permissive for transcription by RNA polymerase (RNAP) through the action of the SWR1 complex that incorporates H2A.Z into nucleosomes around the TSS, histone acetyltransferase complexes (HATs) that add acetyl groups (Ac) to histones, and the PAF1 complex that trimethylates (me3) H3K4 as it travels with RNAP. The distribution and abundance of each chromatin mark are depicted below the locus diagram. In the active state H3K4me3 and H3 acetylation (H3Ac) are high at the 5' end of the gene whereas the levels of H3K27me3 and LHP1, the silencing protein that recognizes H3K27me3, are low throughout. H2A.Z is incorporated into nucleosomes around the TSS. FLC is silenced once the plant matures and environmental conditions are permissive, allowing expression of the flowering induction genes and production of an inflorescence. Silencing occurs through the combined action of histone deacetylases (HDACs) that remove acetyl groups from histones, histone demethylases (HDMs) that remove the methyl groups from H3K4, and the PRC2 complex that trimethylates H3K27. Once the H3K27me3 mark is imparted, it is recognized and bound by LHP1 that promotes stable silencing of the gene. H2A.Z continues to be incorporated into upstream nucleosomes by SWR1. When flowering and fertilization are complete, FLC must be reactivated in the embryo so that vegetative growth will commence upon germination. Incorporation of H2A.Z into the silent FLC chromatin appears to be required for this reactivation.
and incorporation of H2A.Z by the SWR1-like complex [50–52,53**]. Loss of Paf1 complex members, the trithorax group ATX proteins, subunits of the SWR1 complex, or H2A.Z itself all lead to reduced FLC expression and premature flowering. In addition to these components, other modifications to FLC chromatin, including H3 acetylation, H2B ubiquitination, and H3K36 methylation, are also involved in maintaining FLC expression and repressing flowering [54].

Once the plant has reached maturity and environmental conditions are optimal, FLC must be silenced so that flowering can occur. This switch from the active to silent state requires recruitment of H3K4 demethylase enzymes and a PRC2 complex that deposits H3K27me3, which leads to recruitment of LHP1 and silencing of the gene. In addition, arginine methylation of H3 and H4 as well as methylation of H3K9 is also important for stable silencing of FLC [54]. Interestingly, H2A.Z remains present and in fact increases in abundance in silenced FLC chromatin, indicating that it is one component that is necessary but insufficient for FLC expression [50]. In order for vegetative growth to resume in the next generation, FLC is reverted to an active state during embryo development, and the cycle continues [55*,56*].

Plant responses to the environment

Plants are constantly faced with environmental changes such as temperature, nutrient and water availability, as well as attack by pathogens and herbivores. In recent years a great deal of progress has been made in identifying many of the genes that are responsive to these varied environmental conditions, and also in understanding the role of chromatin alterations in regulating their expression. While some of the histone modifications discussed above are thought to be important for environmental responses, studies of plant responses to temperature, phosphate deficiency, and pathogens have revealed new and unexpected functions of H2A.Z.

In a genetic screen to identify regulators of the response to increased temperature in Arabidopsis, Kumar and Wigge [57**] found the ARP6 gene, which encodes an essential component of the SWR1 complex that is required for H2A.Z incorporation into chromatin. The authors found that the set of genes that are normally altered in expression in response to increased temperature were constitutively altered in arp6 mutants grown at normal temperature. Furthermore, they found that H2A.Z nucleosomes were lost from genes after a temperature increase in wild-type plants, suggesting that H2A.Z nucleosomes serve as temperature sensors and actually gate the response to increased temperature. In warmer temperatures some genes are upregulated and others are downregulated; thus, the role of H2A.Z in modulating transcription seems to depend on the gene context.

Downregulation of expression by H2A.Z has also been observed at phosphate starvation response (PSR) genes, which are induced in response to phosphate deficiency [58]. These genes are normally expressed at very low levels and many have H2A.Z-containing nucleosomes near their TSSs. Interestingly, it was observed that many of the PSR genes are highly induced in arp6 mutants, suggesting that H2A.Z serves to reduce expression of these genes. Similarly, gene expression profiling studies have indicated that many of the genes involved in the pathogen response known as systemic acquired resistance (SAR) are also constitutively upregulated in mutants in which the SWR1 complex is non-functional [53**].

Collectively, these examples show that H2A.Z is not simply required for induction or high-level transcription of genes, as it is for FLC, but in fact it seems to have the opposite effect at some genes. In the case of the warm temperature response, the physical basis of this effect appears to be the general thermal instability of H2A.Z nucleosomes, which would increase exposure of promoter DNA to transcriptional regulatory proteins thatdictate the specific response of each gene to increased temperature. Such a mechanism is also probably at play in the case of the PSR and SAR genes.

Conclusions and prospects for the future

Histone variants and PTMs are generally conserved in multicellular eukaryotes, but these chromatin modifications are diversified in different organisms in terms of the genes they regulate, and consequently the developmental processes they control. Recent progress in the chromatin field has shown that histone variants can have major effects on the physical stability of nucleosomes and therefore may regulate gene expression by modulating DNA exposure, similar to the direct effects of histone acetylation and deacetylation. Histone methylations probably create binding sites for effector proteins that can promote or repress transcription by mechanisms that remain to be fully elucidated. The effects on gene expression of other histone modifications, including phosphorylation and ubiquitylation, are poorly understood.

The transcriptionally active gene state generally correlates with H2A.Z, a number of different PTMs including acetylated H3, H3K4me3, as well as a lack of H3K27me3. Conversely, the silent state appears to require H3K27me3, histone deacetylation, and demethylation of H3K4. Each of these states is metastable at a given gene and can be interconverted in response to endogenous and environmental cues. It is this regulated interconversion between active and silent states that allows for precise timing of plant life cycle transitions, such as flowering, and appropriate responses to environmental conditions.

While genome-scale methods have provided clues as to possible roles of histone variants and PTMs in plant gene
regulation, important questions remain. In the coming years it will be necessary to gain a better understanding of the specific roles that many PTMs play in the activation and silencing of particular genes during development and in response to the environment. Another rich area of study will be the characterization of chromatin protein families. Many chromatin components including H2A.Z, H3.3, histone modifying enzymes, modification removal enzymes, as well as modification binding proteins are encoded by multi-gene families [65], providing the potential for functional specialization among family members. While in some cases there is evidence for redundancy with these gene families [53**59], it is clear that this is often not the case [60,61]. Thus, deeper functional study of such protein families will be important for understanding the full epigenomic repertoire of plants. Furthermore, intensified studies of rice and maize chromatin are needed in order to understand the commonalities and differences in chromatin-based gene regulation among divergent plant lineages.

In terms of methodology, it is important for plant researchers not only to continue to embrace new genomics technologies that provide genome-scale readouts, but also to perform gene expression and chromatin profiling studies of single cell types. Recently, several technologies have been developed that make possible the profiling of gene expression and chromatin in specific plant cell types [42**,62,63**]. As opposed to starting with whole tissue, these approaches have the advantage that the behavior of individual cell types can be monitored, rather than amalgamating signals from multiple cell types in a tissue. This issue cannot be ignored because individual cell types clearly have different gene expression profiles and chromatin features, and also respond to environmental stimuli in different ways [64]. Ultimately, studying the intricate relationships between chromatin alterations and gene expression in individual cell types will illuminate how the genome is reprogrammed during cell differentiation and will also lead to a better understanding of the unique structural and physiological characteristics of each cell type in the plant body.

Acknowledgements
We thank Paul Talbert for helpful suggestions on improving the manuscript. This work was supported by funding from the Howard Hughes Medical Institute to S.H. and NIH Postdoctoral Fellowship 1F32GM083449 to R.B.D.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:


This paper describes a comprehensive database of known and predicted chromatin proteins from many different organisms. This database is useful for informing functional studies as well as for comparative and phylogenetic studies.

6 Genome studies and molecular genetics


This paper shows that nucleosomes containing both H2A.Z and H3.3 are unstable compared to other nucleosomes and are enriched in active promoters.


The data presented in this paper show that Polycomb silencing occurs through chromatin compaction in vivo.


The authors report the distributions of H3 methylation to varying extents on K4 across the genome. This study begins to parse out the different roles of each histone modification state as well as the connections between Histone K4 methylation, H3K27 methylation, and DNA methylation.


This paper reports the first large-scale study of the genome-wide distributions and relationships between mRNAs, small RNAs, histone modifications, and DNA methylation in rice.


A genome-scale analysis of the distributions and relationships between mRNAs, small RNAs, histone modifications, and DNA methylation in maize this study reveals general principles of epigenome organization in maize and demonstrates the power of deep sequencing for profiling large and repetitive genomes.


A method for affinity isolation of nuclei from specific cell types is described in this paper. We show that this method can be used to determine genome-wide gene expression and chromatin profiles of individual cell types.


The authors use genetic and molecular analyses to identify a previously unrecognized PRC1-like complex in Arabidopsis that is involved in KNOX gene silencing. These results shed light on the mechanisms of Polycumb silencing in plants.


The plant-specific EMF1 protein is shown to effect Polycumb silencing in conjunction with the PRC2 complex. This indicates that there is at least one other mechanism for Polycumb silencing in plants besides the PRC1-like complex described in [*45*].


The authors show that the SWR1 complex and H2A.Z are responsible for the regulation of a class of genes specifically involved in pathogen response. The results also demonstrate that two of the three H2A.Z isoforms in Arabidopsis are functionally redundant.


This paper demonstrates that plant responses to increased temperature are modulated through the thermal instability properties of H2A.Z nucleosomes. The data show that H2A.Z has both positive and negative effects on transcription.


This paper describes a method for isolating nuclei from specific cell types by fluorescence-activated cell sorting. The authors demonstrate that these nuclei can be used for gene expression profiling in individual cell types.