Integrative epigenomic mapping defines four main chromatin states in Arabidopsis

François Roudier[1,*], Ikhlak Ahmed[1], Caroline Bérard[2], Alexis Sarazin[1], Tristan Mary-Huard[2], Sandra Cortijo[1], Daniel Bouyer[3], Erwann Caillieux[1], Evelyne Duvernois-Berthet[1], Liza Al-Shikhley[1], Laurène Giraut[4], Barbara Després[1], Stéphanie Drevensek[1], Frédy Barneche[1], Sandra Dèrozier[4], Véronique Brunaud[4], Sébastien Aubourg[4], Arp Schnittger[3], Chris Bowler[1], Marie-Laure Martin-Magniette[2,4], Stéphane Robin[2], Michel Caboche[4] and Vincent Colot[1,*]

[1]Institut de Biologie de l'Ecole Normale Supérieure, Centre National de la Recherche Scientifique (CNRS) UMR8197, Institut National de la Santé et de la Recherche Médicale (INSERM) U1024, Paris, France, [2]AgroParisTech/Institut National de la Recherhe Agronomique (INRA), UMR518 Mathématiques et Informatiques Appliquées, Paris, France, [3]Institut de Biologie Moléculaire des Plantes, CNRS UPR2357, Université de Strasbourg, Strasbourg, France and [4]Unité de Recherche en Génomique Végétale, UMR8114 INRA/CNRS/Université d'Evry Val d'Essonne (UEVE), Evry, France

Post-translational modification of histones and DNA methylation are important components of chromatin-level control of genome activity in eukaryotes. However, principles governing the combinatorial association of chromatin marks along the genome remain poorly understood. Here, we have generated epigenomic maps for eight histone modifications (H3K4me2 and 3, H3K27me1 and 2, H3K36me3, H3K56ac, H4K20me1 and H2Bub) in the model plant Arabidopsis and we have combined these maps with others, produced under identical conditions, for H3K9me2, H3K9me3, H3K27me3 and DNA methylation. Integrative analysis indicates that these 12 chromatin marks, which collectively cover ∼90% of the genome, are present at any given position in a very limited number of combinations. Moreover, we show that the distribution of the 12 marks along the genomic sequence defines four main chromatin states, which preferentially index active genes, repressed genes, silent repeat elements and intergenic regions. Given the compact nature of the Arabidopsis genome, these four indexing states typically translate into short chromatin domains interspersed with

*Corresponding authors. F Roudier or V Colot, Institut de Biologie de l'Ecole Normale Supérieure, CNRS UMR8197, INSERM U1024, 46 rue d'Ulm, 75230 Paris Cedex 05, France. Tel.: + 33 014 432 3538; Fax: + 33 014 432 3935; E-mails: roudier@biologie.ens.fr or colot@biologie.ens.fr

each other. This first combinatorial view of the Arabidopsis epigenome points to simple principles of organization as in metazoans and provides a framework for further studies of chromatin-based regulatory mechanisms in plants.

Introduction

Packaging of DNA into chromatin is pivotal for the regulation of genome activity in eukaryotes. The basic unit of chromatin is the nucleosome, which is composed of 147 bp of DNA wrapped around a protein octamer composed of two molecules each of the core histones H2A, H2B, H3 and H4. Covalent modifications of histones, DNA methylation, incorporation of histone variants, and other factors, such as chromatin-remodelling enzymes or small RNAs, all contribute to defining distinct chromatin states that modulate access to DNA (Berger, 2007; Kouzarides, 2007). In particular, different histone modifications are thought to act sequentially or in combination in order to confer distinct transcriptional outcomes (Strahl and Allis, 2000; Jenuwein and Allis, 2001; Berger, 2007; Lee *et al*, 2010a). More generally, it is now well established that the precise composition of chromatin along the genome, which defines the epigenome, participates in the selective readout of the genomic sequence.

Thanks in part to a compact, almost fully sequenced and well-annotated genome, the flowering plant *Arabidopsis thaliana* has become a model of choice for exploring the epigenomes of multicellular organisms and the contribution of chromatin to the regulation of genome activity during development or in response to the environment. Indeed, epigenomic profiling in Arabidopsis has begun to provide insights into the relationship between transcriptional activity and localization of chromatin marks or histone variants (Roudier *et al*, 2009; Feng and Jacobsen, 2011). For instance, H3K4me3 and H3K36me2 are detected at the 5′- and 3′-ends of actively transcribed genes, respectively (Oh *et al*, 2008; Zhang *et al*, 2009), while H3K27me3 broadly marks repressed genes (Turck *et al*, 2007; Zhang *et al*, 2007; Oh *et al*, 2008). In contrast, cytosine methylation (5mC) has a dual localization. It is present predominantly over silent transposable elements (TEs) and other repeats, where it is associated with H3K9me2 and

H3K27me1, but also in the body of $\sim$30% of genes, many of which are characterized by moderate expression levels (Lippman et al, 2004; Zhang et al, 2006; Zilberman et al, 2006; Turck et al, 2007; Vaughn et al, 2007; Bernatavichute et al, 2008; Cokus et al, 2008; Lister et al, 2008; Jacob et al, 2010). Furthermore, the variant histone H2A.Z, which is preferentially deposited near the 5'-end of genes and promotes transcriptional competence, antagonizes DNA methylation and vice versa (Zilberman et al, 2008). However, extensive combinatorial analyses of these and other chromatin marks have not been performed so far in Arabidopsis and meta-analysis of published data is complicated by the fact that biological materials and methodologies often differ between studies.

Here, we report the epigenomic profiles of eight histone modifications (H3K4me2, H3K4me3, H3K27me1, H3K27me2, H3K36me3, H3K56ac, H4K20me1 and H2Bub). Integrative analyses of these and other profiles, previously obtained under identical conditions for DNA methylation, H3K9me2, H3K9me3 and H3K27me3 (Turck et al, 2007; Vaughn et al, 2007), indicate a low combinatorial complexity of chromatin marks in Arabidopsis, as recently reported for metazoans (Wang et al, 2008; Hon et al, 2009; Ernst and Kellis, 2010; Gerstein et al, 2010; Roy et al, 2010; Kharchenko et al, 2011; Liu et al, 2011; Riddle et al, 2011; Zhou et al, 2011). Furthermore, our study identifies four main chromatin states in Arabidopsis, which have distinct indexing functions and which typically form short domains interspersed with each other. This first comprehensive view of the Arabidopsis epigenome suggests simple principles of organization, as recently proposed for Drosophila (Filion et al, 2010), and provides a resource to refine our understanding of the control of genome activity at the level of chromatin.

Results

Epigenomic profiling of 12 chromatin marks

Epigenomic maps were generated for eight histone modifications (H3K4me2, H3K4me3, H3K27me1, H3K27me2, H3K36me3, H3K56ac, H2Bub and H4K20me1) using chromatin extracted from young seedlings and immunoprecipitation followed by hybridization to a tiling microarray that covers the entire chromosome 4 of Arabidopsis at $\sim$900 bp resolution (Turck et al, 2007). Data previously obtained for 5mC (Vaughn et al, 2007), H3K9me2, H3K9me3 and H3K27me3 (Turck et al, 2007) using similar materials and methodologies were also considered. Epigenomic profiling was additionally performed for seven of these marks (H3K4me2, H3K4me3, H3K27me1, H3K27me3, H3K36me3, H2Bub and 5mC) using a tiling microarray covering the whole-genome sequence at 165 bp resolution. Chromosome 4 and whole-genome maps were also obtained for histone H3 to control for nucleosome occupancy. The 12 marks were chosen because they were shown in previous studies to be associated with distinct transcriptional activities or subnuclear localization in Arabidopsis. In addition, our selection was focussed to a large extent on histone lysine methylation, which exists in three forms (mono-, di- and trimethylation) and therefore has a versatile indexing potential (Sims and Reinberg, 2008).

Collectively, the 12 chromatin marks cover almost all of the regions that are detectably associated with histone H3, which amount to $\sim$90% of the total genome sequence (data not shown; Chodavarapu et al, 2010). The distribution of each chromatin modification was characterized in detail along chromosome 4. In agreement with previous reports (Lippman et al, 2004; Turck et al, 2007; Zhang et al, 2007, 2009; Oh et al, 2008; Tanurdzic et al, 2008), H3K4me2, H3K4me3, H3K9me3, H3K27me3 and H3K56ac are mostly found in euchromatin (Figure 1A; Supplementary Figure S1; Supplementary Table I), which reflects the fact that these different modifications are associated almost exclusively with genes (Figure 1B). H2Bub and H3K36me3, for which no epigenomic maps have been reported to date in plants, are also characterized by a predominant distribution over genes. In contrast, H4K20me1 is found in heterochromatin mainly and associates with TE and other repeat element sequences (Figure 1B), like H3K9me2 (Lippman et al, 2004; Bernatavichute et al, 2008). The present analysis reveals in addition that, like 5mC (Zhang et al, 2006; Zilberman et al, 2006), H3K27me1 and H3K27me2 are dual marks associated not only with TEs but also with a fraction of genes (Supplementary Tables II–IV).

Each chromatin mark defines domains of contiguous tiles and the number of these domains ranges from 306 for H3K9me2 to 1163 for H3K4me3. For H3K4me3, H3K36me3, H3K56ac, H3K9me3, H2Bub or H3K27me3, domains have similar median length between euchromatin and heterochromatin and mostly coincide with single transcription units (Supplementary Table II; Supplementary Figure S2). By contrast, H3K9me2, H4K20me1, H3K27me1, H3K27me2 and 5mC form small domains in euchromatin but large domains in heterochromatin, as a result of the dense clustering of TE and other repeat sequences in the latter (Supplementary Figure S2; Supplementary Table II).

Combinatorial analysis of chromatin marks

As a first step in exploring the combinatorial deposition patterns of chromatin marks, unbiased pairwise association analyses were carried out. A heat map generated from the calculated association values (Supplementary Table V) and organized by hierarchical clustering reveals two clear groups of correlated pairs that distinguish genes from TE sequences (Figure 1C). Next, co-occurrence of marks was registered over each of the $\sim$20 000 tiles of the chromosome 4 array. Of the $2^{12} = 4096$ combinations theoretically possible, only 665 were observed and among these, only 38 concerned at least 100 tiles (Supplementary Figure S3A). This indicates therefore a limited repertoire of chromatin signatures in Arabidopsis, as in other eukaryotes (Ernst and Kellis, 2010; Kharchenko et al, 2011; Liu et al, 2011). The four prevalent combinations of marks are H3K27me1 + 5mC + H3K9me2 + H4K20me1 + H3K27me2, H3K56Ac + H2Bub + H3K4me3 + H3K4me2 + H3K9me3 + H3K36me3, H3K27me3 + H3K27me2 + H3K4me2 and H3K27me3 + H3K27me2, which cover 10.9, 6.8, 4.7 and 4.6% of the tiling array, respectively. Whereas the first combination is almost exclusively associated with TE sequences, the other three are mainly present over genes (Supplementary Figure S3B). Furthermore, like H3K27me3 + H3K27me2, most of the remaining combinations represented by at least 100 tiles are subcombinations of the three prevalent ones (Supplementary Figure S3B and data not shown).
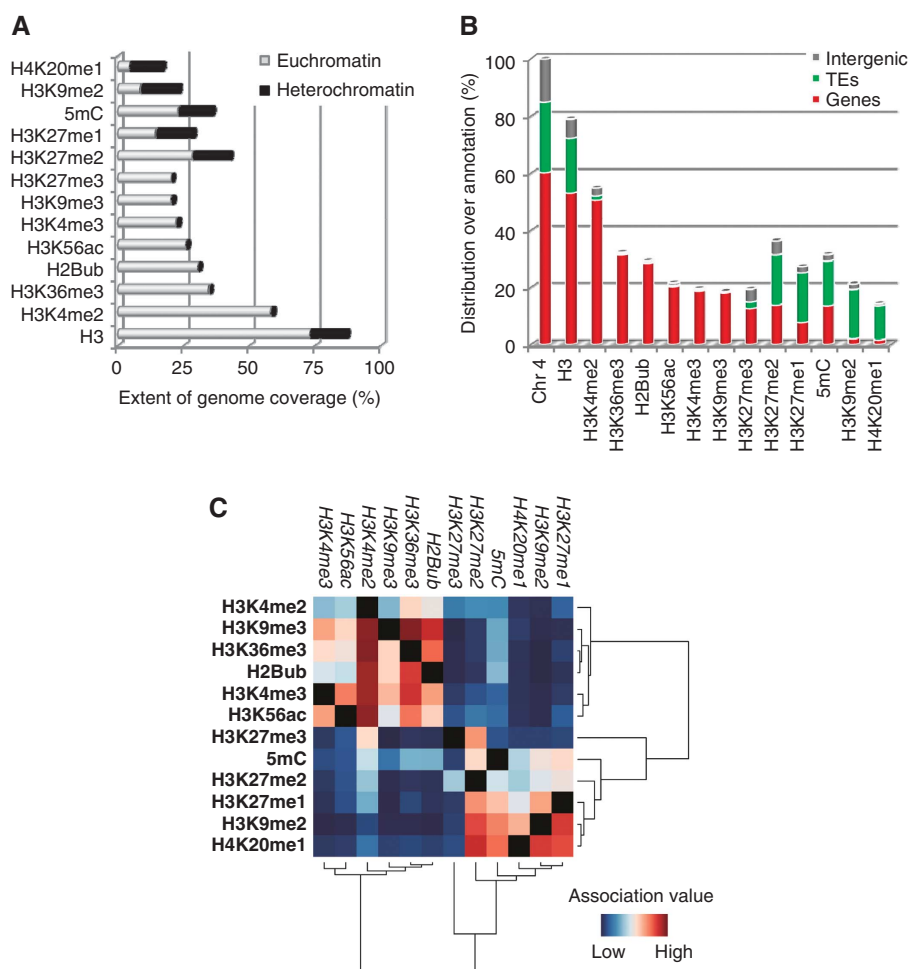
Figure 1 Genomic distribution of chromatin marks. (**A**) Relative coverage of chromatin marks in the euchromatin and heterochromatin of chromosome 4. Coordinates for heterochromatin are 1.61–2.36 Mb (knob) and 2.78–5.15 Mb (pericentromeric regions). (**B**) Chromosome-wide distribution of chromatin marks over annotated features. Tiles that overlap annotated genes or transposable elements (TAIR8) by at least 50 bp were assigned to the corresponding annotation and otherwise called 'intergenic'. (**C**) Pairwise association analysis of the 12 chromatin marks along chromosome 4. Mean association values were calculated for each pair of modifications over all marked tiles and are shown as a directional heat map organized by hierarchical clustering using Pearson's correlation distances.

To complement this tile-centric analysis and to identify the prevalent combinatorial patterns of the 12 chromatin marks, unsupervised *c*-means clustering was performed. The number of clusters (k) was varied from 2 to 11 and $k = 4$ was determined to be optimal in maximizing homogeneity within clusters and heterogeneity between them. The four chromatin states (CS1–CS4) defined by these four clusters are also identified by PCA analysis (data not shown), thus reinforcing their significance. Whereas CS1 regroups ∼90% of the tiles associated with H3K4me3, H3K36me3, H3K9me3 and H2Bub as well as the majority of H3K4me2- and H3K56ac-marked sequences, H3K27me3 and H3K27me2 are the most prevalent modifications in CS2 (Figure 2A). As expected from their composition, CS1 and CS2 are mainly associated with genes (Figure 2B) and have antagonistic indexing functions, being prevalent among active and repressed/lowly expressed genes, respectively (Figure 2C). CS3, which is associated predominantly with TE sequences (Figure 2B), regroups most of the tiles marked by H3K9me2, H4K20me1 and H3K27me1 as well as ∼50% of those marked by H3K27me2 and 5mC (Figure 2A). In contrast to the other

three chromatin states, CS4 is not particularly enriched in any chromatin mark (Figure 2A) and is found mainly outside of genes and TE sequences (Figure 2B). Nonetheless, CS4 also marks ∼10% of genes, most of which display low expression (Figure 2C). In keeping with the domain layout of individual marks, CS1–CS4 typically form small domains interspersed with each other, except in cytologically defined heterochromatin, where CS3 forms larger domains as a result of the clustering of TE sequences (Figure 2D; Supplementary Figure S4).

Chromatin signatures of genes

To investigate further the chromatin indexing of genes, pairwise analysis of chromatin modifications was carried out specifically over genic tiles, which revealed a tight association between H3K4me3 and H3K56ac, between H3K36me3, H3K9me3 and H2Bub and between H3K27me2 and H3K27me3 (Figure 3A). Next, average enrichment levels were calculated within and around genes for all marks except H3K9me2 and H4K20me1, which are almost exclusively associated with TE and other repeat sequences. As shown
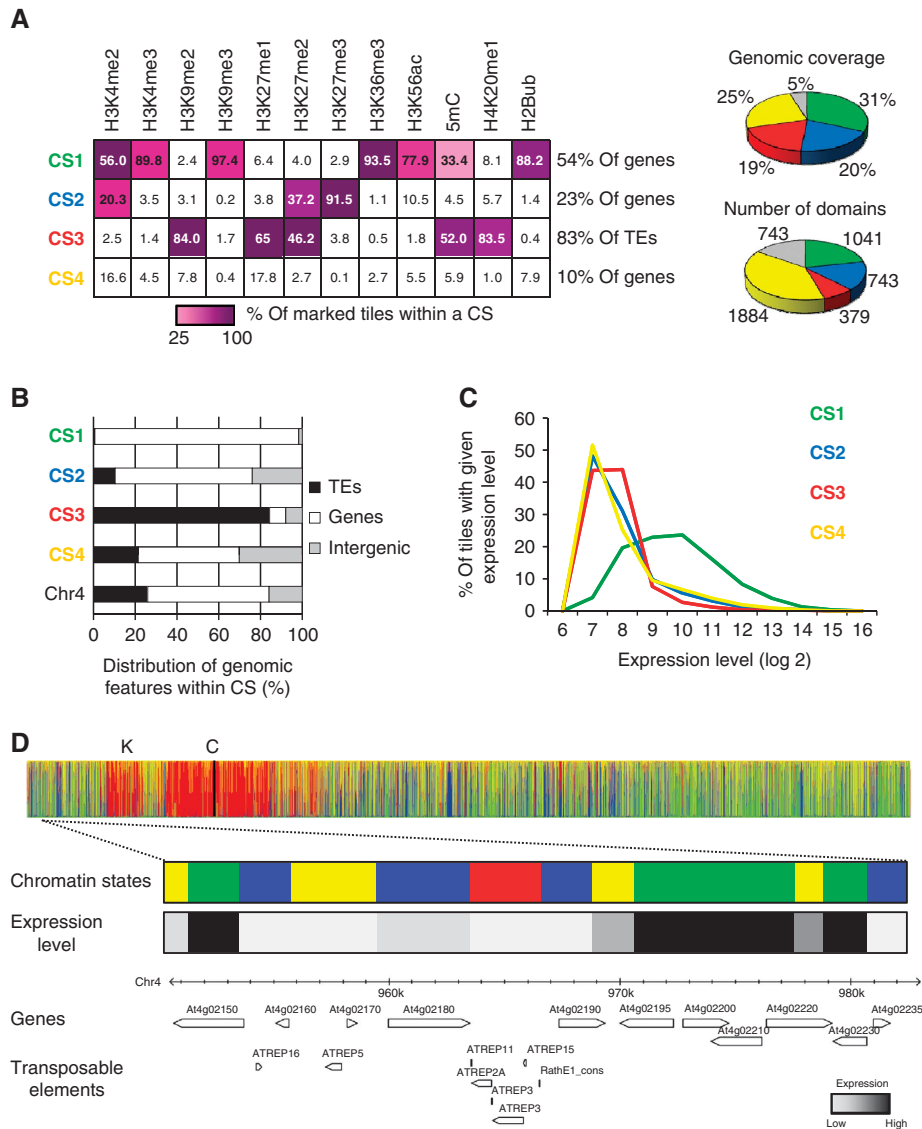
Figure 2 The Arabidopsis epigenome contains four predominant chromatin states. (**A**) The table on the left indicates the composition of the four predominant chromatin states (CS) identified by *c*-means clustering. The distribution of the 12 chromatin marks over the four CS is indicated as a heat map for values ranging from 25 % (light purple) to 100 % (dark purple). The degree of homogeneity of each CS is indicated by the percentage of tiles assigned to it that are associated with each of the 12 chromatin marks (numbers inside cells). Note that no single mark is present over > 20 % of the tiles assigned to CS4, in contrast to what is observed for CS1–CS3. The percentage of genes indexed by CS1, CS2 and CS4 and the percentage of TE annotations indexed by CS3 are also shown. Pie charts indicate the relative genomic coverage of the four CS and the number of domains that they each form. Grey colour corresponds to tiles that cannot be unambiguously assigned to any of the four CS (see Materials and methods). (**B**) Relative proportion of genomic features within each CS. Tiles that overlap annotated genes or transposable elements (TEs) (TAIR8) by at least 50 bp were assigned the corresponding annotation. All other tiles were considered as 'intergenic'. (**C**) Relationship between chromatin states and gene expression level. The percentage of tiles associated with a given CS is represented according to expression level. The dashed line represents the distribution of all annotated genes of chromosome 4. Expression data (Schmid *et al*, 2005) were obtained by averaging appropriate developmental stages. (**D**) Distribution of the four CS along chromosome 4. For each tile, membership to a given CS is colour coded. K: heterochromatic knob. The non-sequenced part of the centromere (C) is represented by the vertical black line. The high interspersion of chromatin states seen outside of heterochromatin is highlighted in a genome browser view of a 30-kb euchromatic region (positions 0.95–0.98 Mb).

in Figure 3B, values are highest within the transcribed region for the 10 chromatin modifications considered and are typically lowest upstream or downstream of it. However, distribution patterns vary substantially between marks, as previously established in several instances (Turck *et al*, 2007; Zhang *et al*, 2007; Jacob *et al*, 2010). H3K4me3, H3K56ac, H3K4me2, H3K36me3 and H3K9me3 all peak at the 5′-end of the transcribed region, but the first two marks more sharply than the other three (Figure 3B). In contrast, H2Bub as well as H3K27me1 are highest more centrally, 5mC is most enriched in the 3′-half of the transcribed region and both H3K27me2 and H2K27me3 show an even distribution across transcribed regions. Finally, H3K27me2 differs from all other marks including H3K27me3 in that it remains high in flanking regions, a difference which does not result from the presence of H3K27me2-marked TE sequences
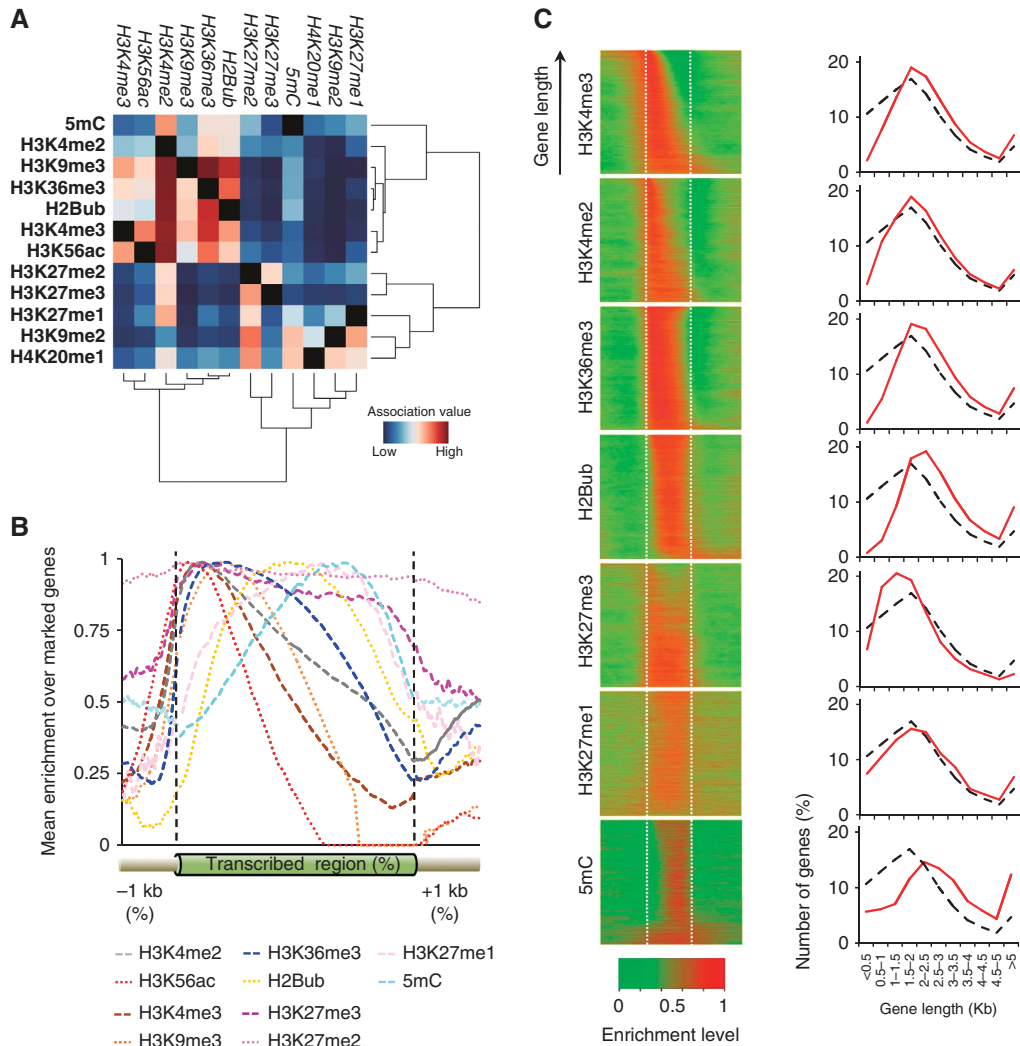
Figure 3 Distribution of chromatin marks over genes. (**A**) Pairwise association analysis of the 12 chromatin marks along chromosome 4. Mean association values were calculated for each pair of modifications over all marked genic tiles and are shown as a directional heat map organized by hierarchical clustering using Pearson's correlation distances metric. (**B**) Mean enrichment levels relative to histone H3 are plotted along marked genes (transcribed region, scaled to accommodate for different gene lengths, bin size of 1%) as well as up to 1 kb of upstream and downstream sequences (bin size of 10 bp). Maximum value for any given mark is arbitrarily set to 1. Data were obtained using the chromosome 4 tiling array. Note that values for H3K27me2 in upstream and downstream regions are significantly higher than for unmarked genes (>0.9 versus ~ 0.6, not shown). (**C**) Left panels: Enrichment levels relative to histone H3 for marked genes sorted by length. Each line represents a single gene as well as 1 kb of upstream and downstream sequences. Enrichment is indicated as a heat map, with maximal (red) and minimal (green) values set to 1 and 0, respectively. Right panels: Frequency distribution of marked (red line) and all genes (black dashed line) according to their length. Data were obtained using the whole-genome tiling array.

adjacent to genes nor from the lower signal to noise ratio measured for this mark (see legend of Figure 3A, data not shown). Using the genome-wide profiles obtained for seven chromatin modifications, we could show in addition that contrary to H3K27me3, which preferentially marks small genes as noted before (Luo and Lam, 2010), H2Bub, H3K36me3, 5mC and, to a lesser extent, H3K4me2 as well as H3K4me3 tend to be associated with longer genes (Figure 3C). Unlike these chromatin modifications, H3K27me1 does not exhibit preferential association in relation to gene length (Figure 3C).

It has been established that H3K4me3 and H3K56ac mark genes that are highly and broadly expressed (Oh *et al*, 2008; Tanurdzic *et al*, 2008; Zhang *et al*, 2009). Conversely, H3K27me3 is preferentially associated with genes that are

expressed at low levels or in a tissue-specific manner (Turck *et al*, 2007; Zhang *et al*, 2007; Oh *et al*, 2008; Jacob *et al*, 2010) and 5mC tends to mark moderately expressed genes (Zilberman *et al*, 2006; Vaughn *et al*, 2007). Our analysis confirms these results and indicates in addition that H2Bub, H3K36me3 and H3K9me3 tend to mark highly expressed genes, like H3K4me3 and H3K56ac (Figure 4A). On the other hand, H3K4me2 does not appear to index genes in relation to their expression level and H3K27me1 as well as H3K27me2 tend to be associated with genes that are expressed at low level or in a tissue-specific manner, like H3K27me3 (Figure 4A and B). However, H3K27me1 and H3K27me2/3 mark largely non-overlapping sets of genes with different ontologies (Figure 3A; Supplementary Tables III, IV, VI and VII), which suggests the existence of two
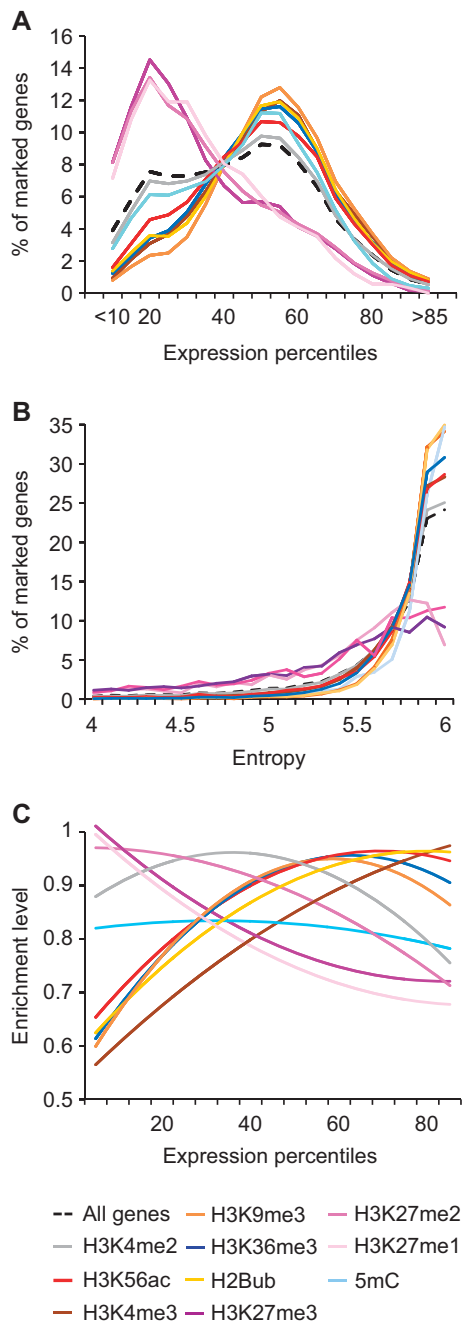
Figure 4 Chromatin indexing in relation to gene expression. (**A**) Distribution density of marked genes according to expression percentiles. Genes were binned according to their absolute expression values in whole seedlings. The dashed line indicates the distribution of all annotated genes on chromosome 4 across all expression percentiles. Expression data (Schmid *et al*, 2005) were obtained by averaging appropriate developmental stages. (**B**) Tissue specificity of marked genes as estimated by Shannon entropy calculation. Low entropy values indicate high tissue specificity. The fraction of marked genes associated with a given entropy value is plotted for each chromatin modification. (**C**) Relationship between gene expression and enrichment level for each chromatin modification. Maximum enrichment level is set to 1 in each case.

distinct gene repression systems associated with methylation of H3K27. For most chromatin marks, average enrichment levels correlate either positively or negatively with expression

levels (Figure 4C). Thus, values for H3K4me3, H3K56ac, H3K36me3, H2Bub and H3K9me3 increase gradually with gene expression, at least up to mid expression levels, whereas values for H3K27me1, H3K27me2 and H3K27me3 show an opposite trend. Whether these correlations reflect expression of genes in a variable number of cells, or true differential enrichment in relation to expression level, remains to be determined.

Collectively, our findings indicate that H3K4me3 and H3K27me3 are diagnostic of two antagonist chromatin states that are associated with most active and repressed genes, respectively. However, ∼13% (3433 out of 27 294) of genes marked by H3K4me3 or H3K27me3 in whole seedlings present both marks, in agreement with previous observations (Oh *et al*, 2008; Zhang *et al*, 2009). To explore this further, H3K4me3 and H3K27me3 were mapped genome-wide using chromatin extracted from roots and profiles were compared with those obtained for whole seedlings (this study) or aerial parts only (Oh *et al*, 2008). Out of the 3433 genes with both marks in whole seedlings, 284 genes (8.3%) are only marked by H3K4me3 in roots and by H3K27me3 in aerial parts or vice versa (Figure 5A; Supplementary Table VIII). Correspondingly, a majority of these genes show differential expression between roots and aerial parts (Figure 5B), which is in contrast to genes with persistent co-marking in both plant parts (Figure 5A and C). Thus, it can be concluded that co-marking in whole seedlings results for a number of genes from the mixing of cells with opposite chromatin indexing in the two plant parts. By extension, it is likely that persistent co-marking in one or the other plant parts (Figure 5A) reflects similar mixing of cells with distinct epigenomes, but this time within organs. Co-marking could nevertheless correspond to *bona fide* bivalent marking in some cases, as originally reported in mammals for key regulatory genes poised for activation (Wang *et al*, 2009) and as also described in Arabidopsis for a small number of genes encoding transcription factors (Jiang *et al*, 2008; Berr *et al*, 2010). In this respect, it is noteworthy that ontology analysis of the 224 genes with persistent co-marking in both roots and aerial parts (Figure 5A) indicates significant enrichment for terms associated with regulation of transcription (data not shown).

Discussion

A small number of prevalent chromatin states index the Arabidopsis genome

Using an integrative analysis of the distribution of 12 chromatin marks, we show that the Arabidopsis epigenome is organized around four predominant chromatin states with distinct biochemical, transcriptional and sequence properties. This representation refines the classical segmentation between cytologically defined heterochromatin and euchromatin. A first chromatin state (CS1) corresponds to transcriptionally active genes and is typically enriched in the trimethylated forms of H3K4 and H3K36. Two further states correspond to two distinct types of repressive chromatin. H3K27me3-marked repressive chromatin (CS2) is mainly associated with genes under PRC2-mediated repression (Turck *et al*, 2007; Zhang *et al*, 2007), while H3K9me2- and H4K20me1-marked repressive chromatin (CS3) corresponds to classical heterochromatin and is almost exclusively located
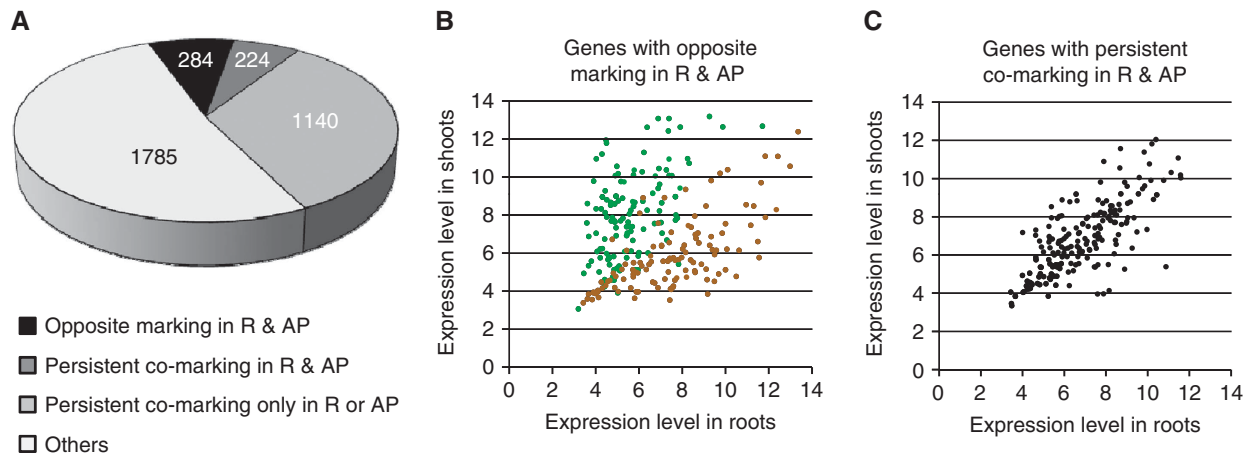
Figure 5 Analysis of genes co-marked with H3K27me3 and H3K4me3 in whole seedlings. (**A**) The 3433 genes co-marked in whole seedlings were split into different classes according to their marking in roots (R; this study) and aerial parts (AP; Oh *et al*, 2008). 'Others' indicate genes with other marking patterns in the two plant parts. This class, which is not expected based on the co-marking observed in whole seedlings, can be explained in part by the fact that the different data sets were not all generated using the same conditions and methodologies. (**B**) Expression analysis in roots and shoots (Schmid *et al*, 2005) for the 284 genes showing opposite marking in roots and aerial parts. Brown dots indicate genes with H3K4me3 in roots and H3K27me3 in aerial parts, green dots indicate genes with the opposite marking pattern. (**C**) Expression analysis in roots and shoots (Schmid *et al*, 2005) for the 224 genes showing persistent co-marking in roots and aerial parts.

over silent TEs (Lippman *et al*, 2004; Bernatavichute *et al*, 2008). A fourth chromatin state (CS4) is characterized by the absence of any prevalent mark and is associated with weakly expressed genes and intergenic regions.

This rather simple organization of Arabidopsis chromatin into four main states shows similarities with that recently reported for Drosophila cells. Indeed, based on the integration of epigenomic maps obtained for 53 chromatin proteins, it was concluded that the Drosophila epigenome is organized into a mosaic of five principal chromatin types that display distinct functional properties (Filion *et al*, 2010). Specifically, Arabidopsis CS2 and CS3 are similar to Drosophila 'BLUE' and 'GREEN' chromatin types, which correspond to repressive chromatin associated with the Polycomb pathway and classical heterochromatin, respectively. Furthermore, CS4, which has no prevalent chromatin mark and indexes some weakly expressed genes as well as intergenic regions is reminiscent of Drosophila 'BLACK' chromatin, which is relatively gene poor and constitutes a repressive environment distinct from heterochromatin. In contrast, transcriptionally active chromatin is represented by a single chromatin state in Arabidopsis (CS1) but by two distinct types in Drosophila that differ in several ways, including the enrichment of H3K36me3 in 'YELLOW' but not in 'RED' chromatin.

Other large-scale epigenomic studies have been performed in yeast (Liu *et al*, 2005), *C. elegans* (Gerstein *et al*, 2010; Liu *et al*, 2011), Drosophila (Kharchenko *et al*, 2011; Roy *et al*, 2010; Riddle *et al*, 2011) and human cells (Wang *et al*, 2008; Hon *et al*, 2009; Ernst and Kellis, 2010; Zhou *et al*, 2011), which all indicate a relatively low combinatorial complexity of chromatin marks. Furthermore, the two main repressive chromatin states defined in Arabidopsis (CS2 and CS3) have similar counterparts in metazoans, indicating that they are highly conserved between plants and animals. On the other hand, the single predominant chromatin state (CS1) that we have identified for transcriptionally active genes

in Arabidopsis has no obvious equivalent in these other organisms. Instead, several chromatin states have been associated with expressed genes in other organisms. This discrepancy likely results from the smaller size of genes and intergenic regions in Arabidopsis (∼2 kb each on average), as well as the relatively lower resolution of our data. Indeed, our analysis shows that distribution patterns vary substantially between chromatin marks associated with active genes (Figure 3B), which suggests that CS1 could be further refined into at least two additional chromatin signatures, specific to the promoter and transcribed region of these genes.

Although the number of chromatin states identified via this type of integrative approach may appear surprisingly low, such analyses aim to identify prevalent combinations of chromatin marks or chromatin proteins. Furthermore, the heterogeneity of the biological material used in many of these studies, including ours, likely hampered the detection of certain chromatin states such as those that are specific to rare cell types. Ultimately, only a knowledge of the epigenomes of individual cell types will enable a full understanding of the functional impact of chromatin-level regulation on genome activity.

Chromatin indexing of genes in Arabidopsis

Our work indicates that the Arabidopsis epigenome is mainly organized at the level of single transcription units and that the distribution of chromatin marks along genes is linked to the transcription process (Figures 2 and 3). For example, H3K4me3 peaks around the transcription start site of actively expressed genes, as observed in all other eukaryotes examined to date (Rando and Chang, 2009). Similarly, H3K56ac is specifically located at gene promoters and shows preferential marking of active genes, suggesting that, like in yeast, it could facilitate rapid transcriptional activation (Williams *et al*, 2008). In contrast to H3K4me3, H3K4me2 shows no particular association with highly expressed genes

or with specific parts of genes. Rather than being a constitutive mark of transcription, H3K4me2 may be implicated in fine tuning of tissue-specific expression, as recently reported in mammals (Pekowska *et al*, 2010).

The distribution of H3K36me3, H3K9me3 and H2Bub over the transcribed regions of expressed genes suggests that these modifications are linked with transcriptional elongation. In the case of H2Bub, this is in agreement with the distribution reported in mammals and yeast (Minsky *et al*, 2008; Schulze *et al*, 2009). For H3K9me3, enrichment over the coding region of expressed genes in Arabidopsis (this study; Caro *et al*, 2007; Turck *et al*, 2007; Charron *et al*, 2009) contrasts with the enrichment predominantly over heterochromatin in animals. However, association with the transcribed regions of some active genes has been reported in mammals (Vakoc *et al*, 2005, 2006; Squazzo *et al*, 2006). Whether H3K9me3 could serve different outcomes depending on genomic and or chromatin context and whether it has any role in transcription regulation in plants remains to be determined. Given the discrepancy between the low amounts of H3K9me3 reported in bulk histones (Jackson *et al*, 2004; Johnson *et al*, 2004) and its apparent abundance reported by ChIP-chip, it is also possible that the H3K9me3 antibody we used recognizes another modification in Arabidopsis, which would be H3K36me3 based on our epigenomic analysis. However, *in vitro* competition assays using an H3K36me3 peptide suggest that this is unlikely (Supplementary Figure S5).

H3K36me3 preferentially marks exons of transcribed genes in yeast, *C. elegans* and mammals (Kolasinska-Zwierz *et al*, 2009) and it was shown to be involved in the control of alternative splicing in mammals (Luco *et al*, 2010). In Arabidopsis, however, H3K36me3 peaks in the first half of the coding region, which is in contrast to the 3′-end enrichment reported in other organisms (Wang *et al*, 2009). This preferential enrichment at the 5′-end, which is not dependent on gene length, could indicate that the principles governing H3K36me3 deposition differ between plants and other eukaryotes. In fact, H3K36me3 distribution in Arabidopsis resembles that of H3K79me3 in mammals (Wang *et al*, 2009). As Arabidopsis lacks a clear homologue of the H3K79 methyltransferase Dot1 and has no H3K79me3 (Zhang *et al*, 2007), it is possible that H3K36me3 in plants serves a function equivalent to H3K79me3 in other eukaryotes. Furthermore, H3K36me2 could have a role similar to that attributed to H3K36me3 in other eukaryotes, as it peaks at the 3′-end of expressed genes in Arabidopsis (Oh *et al*, 2008).

Chromatin marks associated with transcription have been proposed to cross talk and serve as checkpoints in budding yeast and mammals (Suganuma and Workman, 2008; Weake and Workman, 2008; Lee *et al*, 2010a). A similar scenario could be envisioned in Arabidopsis based on the chromatin marks that predominate in CS1, whereby the RNA polymerase II-associated factor 1 complex would induce mono-ubiquitylation of H2B via the activity of the Rad6-Bre1 ubiquitin ligase homologues UBC1, 2 and 3 as well as HUB1 and 2, as shown at the *FLC* gene (Cao *et al*, 2008; Gu *et al*, 2009; Schmitz *et al*, 2009). H2Bub deposition would in turn help recruit COMPASS (COMplex Proteins ASsociated with Set1), thus mediating deposition of H3K4me3 and potentially H3K36me3 (in place of H3K79me3) as well as H3K36me2. Similarly to other eukaryotes, initiation

of another round of transcription would require the activity of the Ubp8 ubiquitin protease homologue, UBP26, which catalyses H2B deubiquitylation (Sridhar *et al*, 2007). Consistent with this, H3K36me3 but not H3K36me2 nor H3K4me3 is almost lost at the 5′-end of the gene *FLC* in *ubp26* mutant plants and this loss is associated with a reduction of *FLC* expression (Schmitz *et al*, 2009). The steady-state distribution pattern of H2Bub observed over expressed genes presumably results from targeted deubiquitylation of H2B at the 5′-end and probably 3′-end of the transcribed region, rather than from an increased ubiquitylation of H2B towards the middle of the gene.

Our epigenomic profiling of the three forms of H3K27 indicates that methylation of this lysine residue is generally associated with repressive chromatin and that its indexing function depends on the degree of modification (mono-, di- and tri-methylation). Thus, in agreement with previous studies, H3K27me3, which is the hallmark of CS2, is almost exclusively present over transcriptionally repressed genes (Turck *et al*, 2007; Zhang *et al*, 2007), while H3K27me1 is prevalent over silent TEs in pericentromeric regions, where it is thought to prevent over-replication (Jacob *et al*, 2009, 2010). Our analysis reveals in addition that H3K27me2 is enriched over H3K27me3-marked genes, as well as of over TE sequences. Although immunolocalization of H3K27me2 at chromocenters (Fuchs *et al*, 2006) was proposed to result from cross-reactivity of antibodies with H3K27me1 in Arabidopsis (Jacob *et al*, 2009), we did not observe extensive cross-reactivity of the H3K27me2 antibodies used in our study with H3K27me1 (Supplementary Figure S5). Moreover, while all forms of methylated H3K27 can be found over genes and are associated with transcriptional repression, little overlap is observed between the small group of genes marked by H3K27me1 and the much larger set of genes marked by H3K27me2/3, suggesting that these modifications define two repressive pathways with distinct gene targets (Supplementary Tables VI and VII). Whereas H3K27me3 deposition is catalysed by the evolutionarily conserved Polycomb Repressive Complexes 2 (Kohler and Hennig, 2010; Bouyer *et al*, 2011), H3K27me1 deposition over TE sequences is partly dependent on the activity of the two SET-domain proteins ATXR5 and ATXR6 (Jacob *et al*, 2009). Whether H3K27me1 deposition over genes requires the same or different histone methyltransferases and whether it is associated with the control of DNA replication remain to be determined. Irrespective of the mechanisms involved, it is noteworthy that whereas H3K27me1-marked TE sequences are also co-marked with H3K9me2 and 5mC, this is not the case for H3K27me1-marked genes.

Acetylation of H3K56 is another chromatin mark that has been linked with the replication process. In Arabidopsis cell cultures, early replicating sequences form broad domains of H3K56ac (Lee *et al*, 2010b). Our epigenomic profiling of H3K56ac reveals mostly short domains located at the 5′-end of expressed genes, which correspond to the replication-independent incorporation of acetylated H3K56. However, a few large domains ($\sim$20 kb) are also detected, which span several genes, intergenic regions and TEs. As our epigenomic maps have been derived from whole seedlings that comprise only a small proportion of mitotic cells, these large H3K56ac domains might correspond to sequences frequently used as endoreplication origins.

Although most Arabidopsis genes are associated with chromatin states CS1 or CS2, $\sim 10\%$ are instead associated with CS4, which is characterized by the absence of any prevalent chromatin mark among the 12 that were analysed in this work (Figure 3). Analysis of additional chromatin marks and proteins will be required to determine more precisely the nature of CS4 and notably the extent of its similarity to the repressive chromatin type BLACK of Drosophila (Filion *et al*, 2010).

To conclude, the first integrative view of the Arabidopsis epigenome provided here could be compared with a first sketch, which is progressively refined until a complete blueprint is produced. Importantly, key aspects of the Arabidopsis epigenome are already apparent in this first sketch, like the relative simplicity of designing principles, which appears to be shared with metazoans.

Materials and methods

Immunoprecipitation of chromatin and methylated DNA, labelling and microarray hybridization

All experiments were performed using wild-type *Arabidopsis thaliana* accession Columbia seedlings grown for 10 days either in liquid MS (whole seedlings) or on MS agar plates (roots and aerial parts) supplemented with 1% sucrose under long day conditions. ChIP and Me-DIP assays were carried out essentially as described (Lippman *et al*, 2005) using commercially available antibodies (Supplementary Table IX; Supplementary Figure S5). Specificity of the H3K27me2 and H3K9me3 antibodies was tested by peptide competition and western blotting analysis on nuclear extracts (Supplementary Figure S5) as described in Bouyer *et al* (2011) using H3K27me3, H3K27me2, H3K27me1, H3K9me3 and H3K36me3 peptides (Millipore, 12-565, 12-566, 12-567, 12-568 and Diagenode sp-058-050, respectively). Immunoprecipitated DNA (IP) and input DNA (INPUT) were amplified, differentially labelled and co-hybridized in dye-swap experiments as described (Lippman *et al*, 2004; Turck *et al*, 2007) for the chromosome 4 tiling array or according to the manufacturer's instructions for the Roche NimbleGen whole-genome tiling array. Two biological replicates were analysed (two dye-swaps). The chromosome 4 array contains 21 800 printed features, on average ~ 900 bp in size. The heterochromatic knob on the short arm and several megabases of pericentromeric heterochromatin are included and account for 16% of the 18.6 Mb covered by the array. Details of array design and production are described in Vaughn *et al* (2007). This platform has been deposited to GEO under accession number GPL10172. The whole-genome tiling array consists of 50–75 nt tiles, with 110 nt spacing on average, that are tiled across the entire genome sequence (TAIR7), without repeat masking. Tiles have a melting temperature of 74°C on average and 88% of them match a unique position in the genome. This custom design was either split into two arrays of 360 718 tiles each, with every other tile on each array (GEO accessions GPL10911 and GPL10918) or synthesized in triplicates of 711 320 tiles each on a single array (GEO accession GPL11005).

ChIP- and Me-DIP-chip data analysis

Hybridization data were normalized as described previously for the chromosome 4 array (Turck *et al*, 2007) or using an ANOVA model was applied to remove technical biases from data obtained using the whole-genome array. Data were averaged on the dye-swap to remove tile-specific dye bias. Normalized data were analysed using the ChIPmix method (Martin-Magniette *et al*, 2008), which was adapted to handle multiple biological replicates simultaneously. This method is based on a mixture model of regressions, the parameters of which are estimated using the EM algorithm. For each tile, a posterior probability is defined as the probability to be enriched given the log(Input) and log(IP) intensities, and is used to assign each tile into a normal or enriched class. A false-positive risk is determined by defining the probability of obtaining a posterior probability at least as extreme as the one that is actually observed when the tile is normal. False-positive risks are then adjusted by the Benjamini–Hochberg procedure and tiles for which the adjusted false-positive risk is <0.01 are declared enriched. Previously published data (Turck *et al*, 2007; Vaughn *et al*, 2007) were re-analysed using the same procedure. Neighbouring enriched tiles are joined into domains by requiring a minimal run of 1.6 kb or 400 bp and allowing a maximal gap of 800 or 200 bp for data obtained using the chromosome 4 or whole-genome arrays, respectively. Thus, 'singletons' are not considered for further analyses.

Computational analyses

General bioinformatics methods including positional, quantitative and class-based computations were conducted in Excel and using *ad hoc* scripts written in R, PERL or Python. Genes and transposable elements were annotated based on TAIR8 and other sequences are assumed to be intergenic. Gene Ontology analyses were done using the GOrilla (Eden *et al*, 2009) with an additional correction for multiple testing of the *P*-values. Pairwise association analysis, which is directional unlike correlation analysis, was calculated by scoring the frequency of co-occurrence of pairs of chromatin modifications among the 12 marks analysed on the chromosome 4 tiling array.

Whole seedlings transcriptome data were retrieved from Schmid *et al* (2005) and genes were binned into 20 expression percentiles according to their absolute expression values. Within each expression percentile, the number of genes marked by a given chromatin modification was calculated and represented as a percentage of all the genes marked by this modification. Shannon entropy for each set of marked genes was calculated as described (Zhang *et al*, 2006) using publicly available developmental expression series (Schmid *et al*, 2005), after filtering genes that showed no expression in any conditions.

Fuzzy *c*-means clustering using R MCLUST package was performed to classify tiles into principal chromatin states based on the 12 epigenomic maps. *c*-means clustering computes membership values for each tile towards all the clusters and all the membership values add up to 1. Each tile was assigned to one cluster only, based on a membership value equal or higher to 0.5. To identify the optimal number of clusters (*k*), cluster validity value, which is an estimate of homogeneity within the clusters and heterogeneity between them, was calculated for clusters from $k = 2–11$.

Data availability

Raw and processed data have been deposited to NCBI's Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under the super-series accession GSE24710 and to CATdb (http://urgv.evry.inra.fr/CATdb) (Samson *et al*, 2004; Gagnot *et al*, 2008). In addition, array data and genome annotation are displayed using a Generic Genome Browser, available for visualization at http://epigara.biologie.ens.fr/index.html.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature* **447:** 407–412

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. *PLoS One* **3:** e3156

Berr A, McCallum EJ, Menard 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

Bouyer D, Roudier F, Heese M, Ellen D, Andersen ED, Gey D, Nowack MK, Goodrich J, Renou J-P, Grini PE, Colot V, Schnittger A (2011) Polycomb Repressive Complex 2 controls the embryo to seedling phase transition. *PLoS Genet* **7:** e1002014

Cao Y, Dai Y, Cui S, Ma L (2008) Histone H2B monoubiquitination in the chromatin of FLOWERING LOCUS C regulates flowering time in Arabidopsis. *Plant Cell* **20:** 2586–2602

Caro E, Castellano MM, Gutierrez C (2007) A chromatin link that couples cell division to root epidermis patterning in Arabidopsis. *Nature* **447:** 213–217

Charron JB, He H, Elling AA, Deng XW (2009) Dynamic landscapes of four histone modifications during deetiolation in Arabidopsis. *Plant Cell* **21:** 3732–3748

Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Kramer U, Merchant SS, Zhang X, Jacobsen SE, Pellegrini M (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* **466:** 388–392

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452:** 215–219

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10:** 48

Ernst J, Kellis M (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28:** 817–825

Feng S, Jacobsen SE (2011) Epigenetic modifications in plants: an evolutionary perspective. *Curr Opin Plant Biol* (advance online publication; doi:10.1016/j.pbi.2010.12.002)

Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ, van Steensel B (2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143:** 212–224

Fuchs J, Demidov D, Houben A, Schubert I (2006) Chromosomal histone modification patterns–from conservation to diversity. *Trends Plant Sci* **11:** 199–208

Gagnot S, Tamby JP, Martin-Magniette ML, Bitton F, Taconnat L, Balzergue S, Aubourg S, Renou JP, Lecharny A, Brunaud V (2008) CATdb: a public access to Arabidopsis transcriptome data from the URGV-CATMA platform. *Nucleic Acids Res* **36:** D986–D990

Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, Alves P, Chateigner A, Perry M, Morris M, Auerbach RK, Feng X, Leng J, Vielle A, Niu W, Rhrissorrakrai K *et al* (2010) Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* **330:** 1775–1787

Gu X, Jiang D, Wang Y, Bachmair A, He Y (2009) Repression of the floral transition via histone H2B monoubiquitination. *Plant J* **57:** 522–533

Hon G, Wang W, Ren B (2009) Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput Biol* **5:** e1000566

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

Jacob Y, Feng SH, 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:** 763–796

Jacob Y, Stroud H, Leblanc C, Feng S, Zhuo L, Caro E, Hassel C, Gutierrez C, Michaels SD, Jacobsen SE (2010) Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* **466:** 987–991

Jenuwein T, Allis CD (2001) Translating the histone code. *Science* **293:** 1074–1080

Jiang D, Wang Y, He Y (2008) Repression of FLOWERING LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb repressive complex 2 components. *PLoS One* **3:** e3404

Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE (2004) Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res* **32:** 6511–6518

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, Linder-Basso D, Plachetka A, Shanower G, Tolstorukov MY, Luquette LJ, Xi R, Jung YL, Park RW, Bishop EP, Canfield TP *et al* (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471:** 480–485

Kohler C, Hennig L (2010) Regulation of cell identity by plant Polycomb and trithorax group proteins. *Curr Opin Genet Dev* **20:** 541–547

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* **41:** 376–381

Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128:** 693–705

Lee JS, Smith E, Shilatifard A (2010a) The language of histone crosstalk. *Cell* **142:** 682–685

Lee TJ, Pascuzzi PE, Settlage SB, Shultz RW, Tanurdzic M, Rabinowicz PD, Menges M, Zheng P, Main D, Murray JA, Sosinski B, Allen GC, Martienssen RA, Hanley-Bowdoin L, Vaughn MW, Thompson WF (2010b) Arabidopsis thaliana chromosome 4 replicates in two phases that correlate with chromatin state. *PLoS Genet* **6:** e1000982

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430:** 471–476

Lippman Z, Gendrel AV, Colot V, Martienssen R (2005) Profiling DNA methylation patterns using genomic tiling microarrays. *Nat Methods* **2:** 219–224

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133:** 523–536

Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ (2005) Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biol* **3:** e328

Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I, Cheung MS, Ercan S, Ikegami K, Jensen M, Kolasinska-Zwierz P, Rosenbaum H, Shin H, Taing S, Takasaki T, Iniguez AL, Desai A, Dernburg AF, Lieb JD, Ahringer J, Strome S *et al* (2011) Broad chromosomal domains of histone modification patterns in C. elegans. *Genome Res* **21:** 227–236

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T (2010) Regulation of alternative splicing by histone modifications. *Science* **327:** 996–1000

Luo C, Lam E (2010) ANCORP: a high-resolution approach that generates distinct chromatin state models from multiple genome-wide datasets. *Plant J* **63:** 339–351

Martin-Magniette ML, Mary-Huard T, Berard C, Robin S (2008) ChIPmix: mixture model of regressions for two-color ChIP-chip analysis. *Bioinformatics* **24:** i181–i186

Minsky N, Shema E, Field Y, Schuster M, Segal E, Oren M (2008) Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nat Cell Biol* **10:** 483–488

Oh S, Park S, van Nocker S (2008) Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. *PLoS Genet* **4:** e1000077

Pekowska A, Benoukraf T, Ferrier P, Spicuglia S (2010) A unique H3K4me2 profile marks tissue-specific gene regulation. *Genome Res* **20:** 1493–1502

Rando OJ, Chang HY (2009) Genome-wide views of chromatin structure. *Annu Rev Biochem* **78:** 245–271

Riddle NC, Minoda A, Kharchenko PV, Alekseyenko AA, Schwartz YB, Tolstorukov MY, Gorchakov AA, Jaffe JD, Kennedy C, Linder-Basso D, Peach SE, Shanower G, Zheng H, Kuroda MI, Pirrotta V, Park PJ, Elgin SC, Karpen GH (2011) Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res* **21:** 147–163

Roudier F, Teixeira FK, Colot V (2009) Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more. *Trends Genet* **25:** 511–517

Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, Lin MF, Washietl S, Arshinoff BI, Ay F, Meyer PE, Robine N, Washington NL, Di Stefano L, Berezikov E, Brown CD, Candeias R *et al* (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330:** 1787–1797

Samson F, Brunaud V, Duchene S, De Oliveira Y, Caboche M, Lecharny A, Aubourg S (2004) FLAGdb + + : a database for the functional analysis of the Arabidopsis genome. *Nucleic Acids Res* **32:** D347–D350

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. *Nat Genet* **37:** 501–506

Schmitz RJ, Tamada Y, Doyle MR, Zhang X, Amasino RM (2009) Histone H2B deubiquitination is required for transcriptional activation of FLOWERING LOCUS C and for proper control of flowering in Arabidopsis. *Plant Physiol* **149:** 1196–1204

Schulze JM, Jackson J, Nakanishi S, Gardner JM, Hentrich T, Haug J, Johnston M, Jaspersen SL, Kobor MS, Shilatifard A (2009) Linking cell cycle to histone modifications: SBF and H2B monoubiquitination machinery and cell-cycle regulation of H3K79 dimethylation. *Mol Cell* **35:** 626–641

Sims III RJ, Reinberg D (2008) Is there a code embedded in proteins that is based on post-translational modifications? *Nat Rev Mol Cell Biol* **9:** 815–820

Squazzo SL, O'Geen H, Komashko VM, Krig SR, Jin VX, Jang SW, Margueron R, Reinberg D, Green R, Farnham PJ (2006) Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome Res* **16:** 890–900

Sridhar VV, Kapoor A, Zhang K, Zhu J, Zhou T, Hasegawa PM, Bressan RA, Zhu JK (2007) Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. *Nature* **447:** 735–738

Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403:** 41–45

Suganuma T, Workman JL (2008) Crosstalk among histone modifications. *Cell* **135:** 604–607

Tanurdzic M, Vaughn MW, Jiang H, Lee TJ, Slotkin RK, Sosinski B, Thompson WF, Doerge RW, Martienssen RA (2008) Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biol* **6:** 2880–2895

Turck F, Roudier F, Farrona S, Martin-Magniette ML, Guillaume E, Buisine N, Gagnot S, Martienssen RA, Coupland G, Colot V (2007) Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet* **3:** e86

Vakoc CR, Mandat SA, Olenchock BA, Blobel GA (2005) Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol Cell* **19:** 381–391

Vakoc CR, Sachdeva MM, Wang H, Blobel GA (2006) Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol Cell Biol* **26:** 9185–9195

Vaughn MW, Tanurd Ic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, Colot V, Doerge RW, Martienssen RA (2007) Epigenetic natural variation in Arabidopsis thaliana. *PLoS Biol* **5:** e174

Wang Z, Schones DE, Zhao K (2009) Characterization of human epigenomes. *Curr Opin Genet Dev* **19:** 127–134

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40:** 897–903

Weake VM, Workman JL (2008) Histone ubiquitination: triggering gene activity. *Mol Cell* **29:** 653–663

Williams SK, Truong D, Tyler JK (2008) Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation. *Proc Natl Acad Sci USA* **105:** 9000–9005

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

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

Zhang XY, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen HM, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* **126:** 1189–1201

Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12:** 7–18

Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456:** 125–129

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2006) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39:** 61–69