

# Hi-C Analysis in *Arabidopsis* Identifies the *KNOT*, a Structure with Similarities to the *flamenco* Locus of *Drosophila*

Stefan Grob,<sup>1</sup> Marc W. Schmid,<sup>1</sup> and Ueli Grossniklaus<sup>1,\*</sup>

<sup>1</sup>Institute of Plant Biology and Zürich-Basel Plant Science Center, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland

\*Correspondence: grossnik@botinst.uzh.ch

<http://dx.doi.org/10.1016/j.molcel.2014.07.009>

## SUMMARY

Chromosomes are folded, spatially organized, and regulated by epigenetic marks. How chromosomal architecture is connected to the epigenome is not well understood. We show that chromosomal architecture of *Arabidopsis* is tightly linked to the epigenetic state. Furthermore, we show how physical constraints, such as nuclear size, correlate with the folding principles of chromatin. We also describe a nuclear structure, termed *KNOT*, in which genomic regions of all five *Arabidopsis* chromosomes interact. These *KNOT ENGAGED ELEMENT* (KEE) regions represent heterochromatic islands within euchromatin. Similar to PIWI-interacting RNA clusters, such as *flamenco* in *Drosophila*, KEEs represent preferred landing sites for transposable elements, which may be part of a transposon defense mechanism in the *Arabidopsis* nucleus.

## INTRODUCTION

Eukaryotic nuclei represent highly complex structures and are involved in many cellular processes. The storage and reading of genetic information require elaborate packaging of chromosomes, which depends on two seemingly conflicting factors: condensation and accessibility of DNA.

Chromosomes are organized into distinct regions, referred to as chromosome territories (CTs). The two chromosome arms (CAs) of a CT form a tight interaction unit, clearly separated from each other (Grob et al., 2013; Schubert et al., 2012). In animals, CAs were initially subdivided into discrete chromatin domains that are distinguished by differential packaging densities and epigenetic state (Lieberman-Aiden et al., 2009). Less packaged domains are characterized by activating epigenetic marks, such as H3K4me3, whereas more densely packaged domains are enriched in the inactive epigenetic mark H3K27me3 (Sexton et al., 2012). Using higher resolution, our knowledge on mammalian chromatin organization could be refined by the finding of topological domains that are demarcated by an enrichment of the insulator protein CTCF (Dixon et al., 2012).

Interaction decay exponents (IDEs) describe the steepness of the slope with which chromatin interaction frequencies (IFs) ob-

tained in Hi-C experiments decay with distance from a given viewpoint. IDEs were used to predict polymer-folding principles in human nuclei, for which distinct models, the fractal globule model (FGM) and the equilibrium globule model (EGM), were proposed (Lieberman-Aiden et al., 2009). The EGM suggests a densely packed polymer with various knots, in which different regions of the polymer interlace. The FGM describes a polymer structure that exhibits globular substructures, reminiscent of beads on a string. As the FGM lacks knots, allowing for easy untangling of chromosomes, it is convenient to describe chromatin conformation. Both models differ in their theoretical IDEs: FGM and EGM yield IDEs of  $-1$  and  $-1.5$ , respectively. Several chromosome interaction studies reported IDEs supporting the FGM (Grob et al., 2013; Lieberman-Aiden et al., 2009; Sexton et al., 2012; Zhang et al., 2012). However, chromatin organization is unlikely uniform along a chromosome, being composed of constitutive heterochromatin in pericentromeric regions (PRs) and euchromatic CAs. Whether PRs and CAs exhibit different IDEs, reflecting a distinct chromatin organization, is not clear, but previous studies showed that IDEs can differ between chromatin states (Sexton et al., 2012).

In *Arabidopsis thaliana*, PRs and CAs clearly differ in appearance, with PRs being part of chromocenters, brightly DAPI-stained dots in interphase nuclei (Fransz et al., 2002). Thus, calculation of IDEs of different chromatin states promises more realistic insights into chromatin organization.

Nuclear architecture is expected to be influenced by extrinsic factors, including nuclear volume. CROWDED NUCLEI (CRWN1, CRWN2, CRWN3, and CRWN4) proteins control nuclear size and are localized to the nuclear periphery (Dittmer et al., 2007; Sakamoto and Takagi, 2013; Wang et al., 2013). In *crwn1* and *crwn4* mutants, nuclear size is up to 75% smaller. Additionally, *crwn4* mutants exhibit fewer and dispersed chromocenters, indicating a role in heterochromatin regulation. Although the effects of *crwn* mutants on nuclear morphology have been described, it remains unknown how these changes affect chromosomal architecture. Therefore, we analyzed chromosomal architecture by performing Hi-C experiments on nuclei of *crwn1* and *crwn4* mutant *Arabidopsis* seedlings.

To date, very few studies have been published assessing differences between wild-type (WT) and mutant Hi-C data sets. Thus, a gold standard on how to assess differences between Hi-C data sets is lacking. We propose a computational method to assess the significance of changes observed in different Hi-C data sets and report how *crwn1* and *crwn4* mutants affect

chromosomal architecture. Hi-C not only allows a description of the principles of chromatin organization but also identifies discrete chromosomal interactions, which might confer functional significance. We identified a structure consisting of an entanglement of ten chromosomal regions, the *KNOT*. As it shows certain similarities to the *flamenco* locus of *Drosophila*, which controls several transposable elements (TEs) by RNAi, we postulate a function of the *Arabidopsis KNOT* in TE regulation and processing.

## RESULTS

To gain insight into the chromosomal architecture of *Arabidopsis* nuclei, we performed Hi-C experiments on WT, *crwn1-1*, and *crwn4-1* seedlings of the Columbia-0 (Col-0) accession.

### Chromosomal Neighborhood

We sought to understand how CTs relate to each other and investigated the spatial distribution of chromosomes in the nucleus. We calculated the expected (Zhang et al., 2012) IFs for each pair of *trans*-interacting chromosomes and compared these values to the observed IFs between these pairs. The log-ratio between observed and expected Hi-C interactions was used to describe whether two given chromosomes interact more with each other than expected and hence are located in spatial proximity (Figure 2A). Deviations from expected IFs were low compared to a study in mice (Zhang et al., 2012), suggesting equal interactions between all five *Arabidopsis* chromosomes.

### Hi-C Interactions Form Defined Interaction Domains

The relationship between interactions of neighboring genomic bins allows insight into chromosomal architecture. As previously shown (Lieberman-Aiden et al., 2009; Sexton et al., 2012; Zhang et al., 2012), Hi-C interaction values are not independent of each other but correlate, forming domains of interacting regions (Figures 1A and 1C). Two Hi-C bins in close genomic proximity should share common interactors as they are physically connected. To better define structural domains (SDs), we calculated correlation coefficients of the distance-normalized interaction matrix. Visualization of the distance-corrected correlation matrix facilitated the observation of distinct SDs (Figure 1B). The major domains of chromatin organization were limited to euchromatin of CAs and heterochromatin found in PRs (Table S1 available online and Figure 5C). Yet, we could detect additional SDs within euchromatic CAs encompassing several megabases (Figures 1B–1D and S1).

As previously reported (Grob et al., 2013; Moissiard et al., 2012), we observed increased IFs and high correlation between the PRs of the *Arabidopsis* chromosomes, indicating clustering within the nucleus. Likewise, telomeric regions were observed to specifically interact among each other. Interactions between telomeres and PRs were depleted, suggesting differential compartmentalization (Figures 1A and 1B). Generally, we observed low IFs between euchromatic CAs and PRs, further supporting our previous observation (Grob et al., 2013) that heterochromatin and euchromatin represent distinct interactomes within the nucleus.

### Principal Component Analysis Reveals Distinct Chromatin States

By close inspection of the correlated Hi-C data, we observed discrete SDs, which appeared to highly interact among each other but exhibited rather low IFs with the rest of the genome. Thus, we termed them compacted structural domains (CSDs). In contrast, other SDs exhibited a loose state (loose structural domains [LSDs]), characterized by depleted IFs within them but enriched IFs with more distal regions both in *cis* and *trans*.

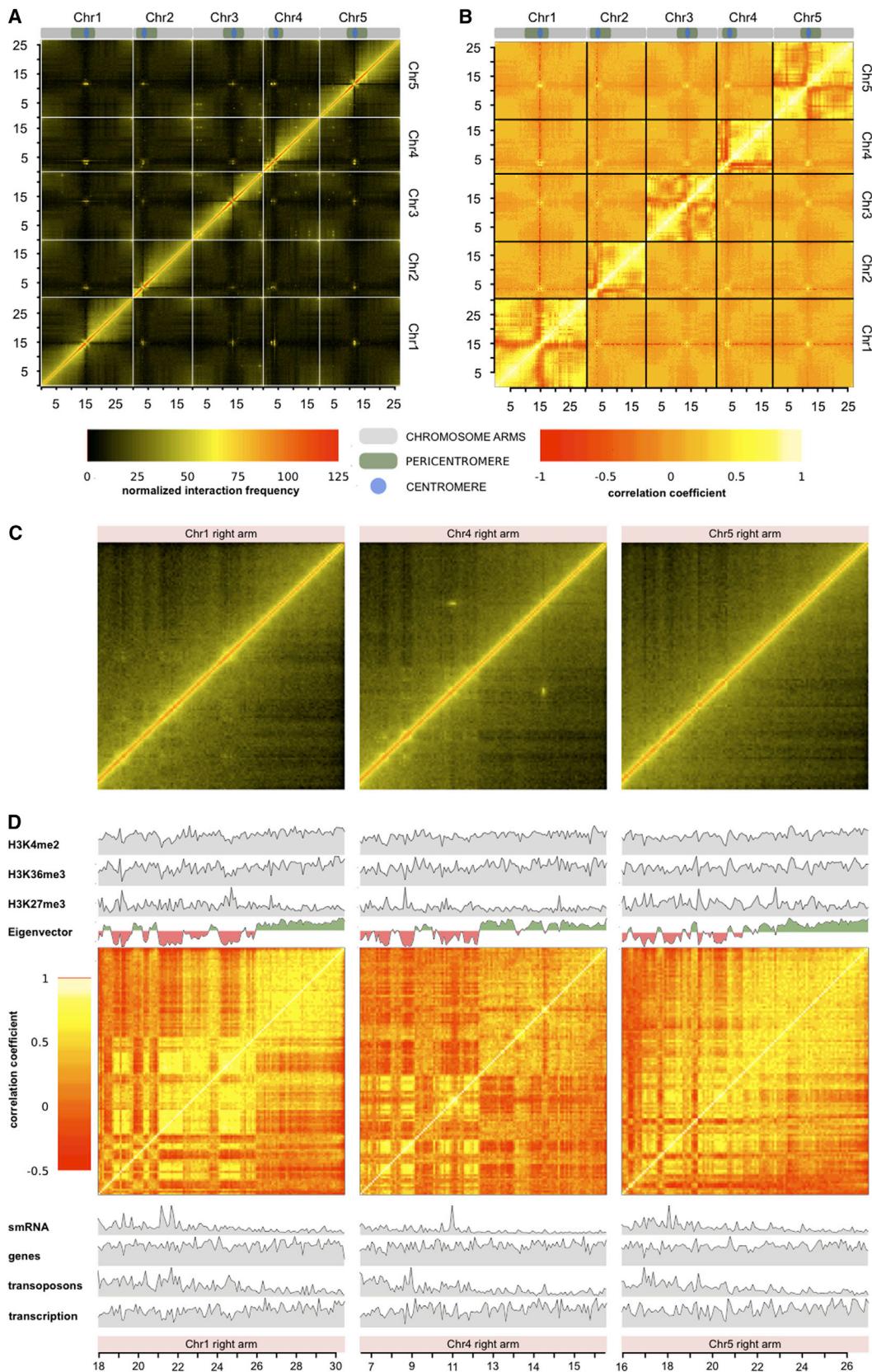
To obtain a numeric description of these SDs, we performed principal component analysis (PCA) on the correlation matrix of each individual chromosome (Chr). This led to a clear partitioning of the interactome into two SDs with either positive or negative Eigenvalues, with negative and positive Eigenvalues corresponding to CSDs and LSDs, respectively. The Eigenvalues can serve as a measure for domain structure, describing the accessibility—and therefore compaction state—of a given SD, and aid in accentuating the domain structure of chromatin (Figures 1C, 1D, and S1).

As expected, the first principal component (which describes the factor adding most to the variance of the data) was mainly dependent on the occurrence of constitutive heterochromatin or euchromatin, and it therefore hindered uncovering a detailed domain structure by PCA. To understand SD formation within euchromatin, we calculated correlations matrices and subsequently PCAs separately for each euchromatic CA, excluding heterochromatic PRs from analysis (Table S1). We found that the accentuation of discrete SDs varies between different CAs. The right arms of Chr1, Chr4, and Chr5 exhibited the clearest sequential arrangement of discrete SDs, whereas SDs on other CAs, although present, were less obvious (Figures 1B–1D and S1).

### LSDs and CSDs Correlate with Epigenetic Chromatin States

Previous reports suggested a correlation between interactome and epigenome (Grob et al., 2013; Lieberman-Aiden et al., 2009; Sexton et al., 2012). Thus, we speculated that specific epigenetic marks correlate with LSDs and CSDs in CAs. To test this hypothesis, we obtained publicly available data on epigenetic and genomic features (see *Supplemental Information*). We computed Pearson's correlation coefficients between each feature and the Eigenvector for all euchromatic CAs individually (Figures 1D, 2B, and S2; Table S2). For the robustness of these analyses, the detection of discrete SDs is crucial. Therefore, we focused specifically on the right arms of Chr1, Chr4, and Chr5, which exhibited the most readily recognizable SDs (Figures 1D and S1).

Generally, histone modifications associated with active euchromatin (Filion et al., 2010; Roudier et al., 2011) exhibited strong correlations with the Eigenvector and highly significant p values. Specifically, high correlations were observed for H3K36me3 and H3K4me2, whereas strong anticorrelation was found for the Polycomb-associated mark H3K27me3 (Figures 2B and S2; Table S2). Histone marks associated with constitutive heterochromatin (H3K27me1, H3K9me2) showed weak anticorrelations. Of genomic features tested, transcription rate



(legend on next page)

highly correlated, whereas the number of TEs highly anticorrelated (Figures 2B and S2; Table S2). In summary, correlation analysis revealed that active histone modifications and transcription rate positively correlated with LSDs, whereas CSDs highly correlated with inactive epigenetic marks and genomic features of inactive euchromatin, such as abundance of TEs and accumulation of associated small RNAs (smRNAs).

To quantify the difference in epigenetic landscape between the two SDs, we assigned each genomic bin to one of two groups, defined by positive or negative Eigenvalues. To test whether the groups significantly differed in epigenetic landscape, we individually performed Wilcoxon rank sum tests for each feature and each CA (Figure 2C; Table S2). H3K9ac, H3K4me2, H3K4me3, H3K36me2, and H3K36me3 were significantly ( $\alpha = 0.01$ ) higher in LSDs for all CAs analyzed. The enrichment of active marks in LSDs varied little, with an average enrichment of 1.2- to 1.3-fold compared to CSDs over all CAs analyzed. In contrast, we observed significant enrichment of H3K27me3 in CSDs (1.3-fold) (Figure 2C).

Despite showing a significant enrichment in CSDs for a subset of CAs, density levels of H3K9me2 and H3K27me1 were generally low, further suggesting that histone modifications characteristic of constitutive heterochromatin do not play a major role for SD formation in euchromatic CAs. Although previously described to colocalize with H3K27me3 (Luo et al., 2012), we did not observe significant differences in H3K18ac (Figure 2C).

In plants, cytosine methylation occurs in the CG, CHG, and CHH context (where H is any base but G). In CSDs, DNA methylation in the CG, CHG, and CHH context was enriched 1.3-, 2.1-, and 1.8-fold, respectively. We observed a significantly higher transcription rate (1.5-fold) in LSDs, while gene density appeared to be a minor factor, as it was only negligibly higher in LSDs (1.1-fold). In contrast, the number of loci associated with smRNAs (2.1-fold) and TEs (2.4-fold) was significantly enriched in CSDs (Figure 2C). We could exclude that sequencing and alignment artifacts perturbed our analyses, as both the density of H3 occupancy and genomic sequencing reads did not significantly differ between LSDs and CSDs (Figure 2C). Furthermore, the results were robust using various genomic bin sizes (25, 50, and 100 kb).

In summary, we could detect a clear correlation between the spatial organization of chromatin and the epigenetic landscape. Features that are predominantly associated with epigenetically inactive euchromatin were enriched in CSDs, whereas features characteristic for active euchromatin were observed at higher densities in LSDs. As we excluded regions of known constitutive heterochromatin (e.g., PRs), we did not observe a correlation between epigenetic marks associated with heterochromatin and either LSDs or CSDs.

#### Figure 1. Visualization of Hi-C Interactome

- (A) Visualization of WT Hi-C IFs; genomic bin size: 250 kb.
  - (B) Visualization of distance-normalized WT correlation matrix; genomic bin size: 250 kb.
  - (C) Magnified view on right arms of Chr1, Chr4, and Chr5; bin size: 100 kb.
  - (D) Visualization of correlative interactomes of the CAs in (C). Eigenvector for each CA representing the Eigenvalues of each 100 kb genomic bin is shown. Additional tracks are densities of epigenetic modifications or number of genomic features.
- See also Table S1 and Figure S1.

#### Arabidopsis Mutants Affecting Nuclear Size Affect the Interactome

We hypothesized that structural characteristics of nuclei could significantly influence chromosomal architecture. Nuclear size represents a likely factor affecting chromatin organization because it will limit the volume available to a CT. To investigate the effects of size constraints, we compared chromatin organization of WT nuclei with nuclei deficient for the structural components CRWN1 and CRWN4.

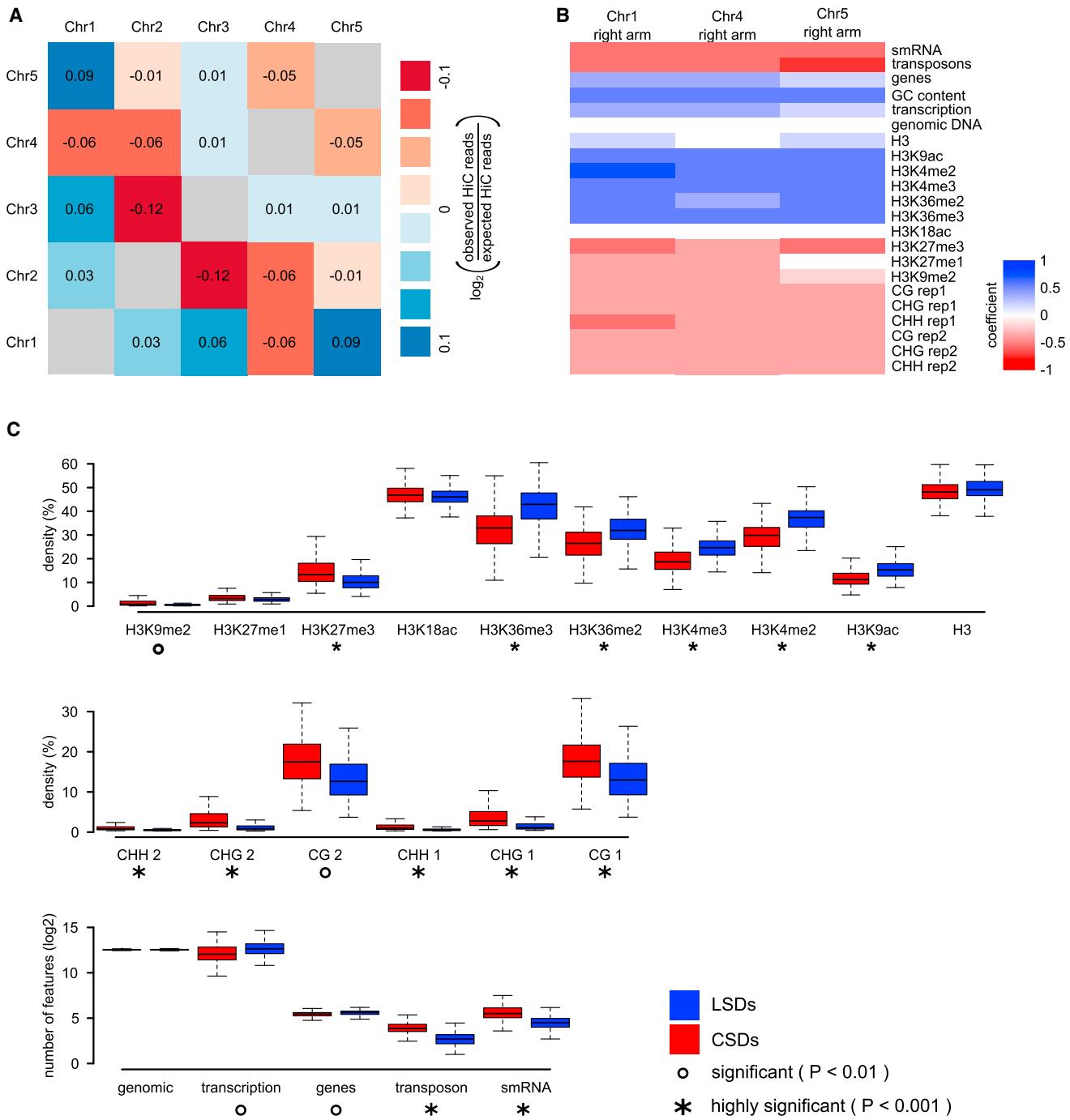
To investigate the impact of the *crwn1* and *crwn4* mutants, we calculated differences between all obtained Hi-C data according to a previously described method (Moissiard et al., 2012) (Figures 3A and S3). In short, we calculated the difference between all elements of two Hi-C matrices of interest. The resulting difference matrix was subsequently normalized according to the absolute IFs in the two Hi-C matrices of interest. By visual inspection of the plotted difference, we observed increased interchromosomal pericentromere interactions, increased interarm interactions, and slightly reduced intra-arm interactions in *crwn4* nuclei (Figures 3A and S3). The reduction of intra-arm interactions was most pronounced for interactions between PRs and more distal regions of the CAs. Complementarily, we observed increased interactions between the two halves of the PRs flanking the centromeres. In contrast, interactions within one-half of the PRs appeared to be depleted, and interactions of PRs and telomeres were reduced in *crwn4* nuclei.

Nuclei of *crwn1* showed similar changes in chromosomal architecture; however, differences to WT were less distinct and their overall magnitude was smaller (Figures 3A and S3). Generally, *crwn4* and *crwn1* nuclei exhibited enrichment in *trans*-interactions (both *trans*-arm and *trans*-chromosomal), suggesting higher genome-wide compaction in these mutants. These observations are consistent with previous studies (Dittmer et al., 2007; Sakamoto and Takagi, 2013), describing significantly smaller nuclei in *crwn* mutants, leading to space constraints and, thus, possibly higher *trans*-interactions among the chromosomes. Additionally, we observed increased IFs between the PRs of all five chromosomes (Figures 3 and S3).

#### Differences between *crwn1*, *crwn4*, and Col-0 Cluster in Defined Domains

As chromosomal architecture is partly influenced by stochastic factors, we expected that Hi-C data sets exhibit some variability not based on relevant biological differences. Therefore, we developed an analytical pipeline to reveal biologically significant changes between sets of Hi-C interactomes.

We made use of the axiom that regions in close genomic proximity, which are physically linked, correlate in their genome-wide interactomes. Thus, changes inflicted on the genome-wide interactome of a given genomic bin should be reflected by



**Figure 2. Chromosomal Neighborhood and Features Associated with Chromatin Organization**

(A) Log<sub>2</sub> ratio of observed to expected pairwise interchromosomal interactions.

(B) Pearson's correlation coefficients between the Eigenvector (on 100 kb genomic bins) and epigenetic and genomic features for the right arms of Chr1, Chr4, and Chr5.

(C) Distribution of epigenetic and genomic features in LSDs and CSDs.

See also Table S2 and Figure S2.

changes in interactomes of neighboring genomic bins. We calculated matrix-wise correlation coefficients to obtain matrices of correlated differences (Figures 3B and S3). The rep-

resentation of the correlation matrices showed that differences between Col-0 and the *crwn1* and *crwn4* mutants occurred in distinct domains.

To quantify this effect, we simplified the difference matrices, only considering whether a given interaction pair increases or decreases between two Hi-C data sets. This yielded a signed difference matrix (SDM) with the three possible elements: +, −, and 0 (for no difference) (Figures 3C, 3D, and S3). The Wald-Wolfovitz (WW) runs statistical test reveals whether the elements of a sequence are independent of each other. We expected that differences between two Hi-C data sets that arose from random noise in the data would be independent of each other for a given dimension of the matrix. Conversely, specific differences should occur in blocks of either positive or negative changes between the two Hi-C data sets. We calculated WW p values for each column in the SDM and counted the number of columns exhibiting a p value < 0.01; 50% of the genome-wide interactomes of genomic bins in the SDM of the pair *crwn4*-Col-0 exhibited significant p values. In comparison, 19% and 26% of the columns significantly differed in the *crwn1*-Col-0 and *crwn1*-*crwn4* SDMs, whereas only 2% significant differences were observed between two Col-0 replicates (Figure S3).

We then asked whether significant bins cluster along genomic positions. We expected significant columns to cluster if they contribute to changes that are based on biological differences between Hi-C data sets. Thus, we performed a second WW analysis, testing clustering of significant columns. This yielded extremely low p values for the pairs *crwn4*-Col-0, *crwn1*-Col-0, and *crwn1*-*crwn4*, but nonsignificant p values between two Col-0 replicates (Figures 3C and S3). In summary, alterations of chromosomal architecture associated with mutations in *crwn1* and *crwn4* clustered in defined domains, indicating a low contribution of stochastic variance to the observed differences.

### SD Organization of CAs Does Not Change in *crwn1* and *crwn4* Mutants

Mutations affecting structural components of *Arabidopsis* nuclei influence *trans*-interactions. Intuitively, such alterations were expected due to the reduced nuclear size of *crwn1* and *crwn4* mutants, but they could also affect organizational differences within mutant nuclei. To study *cis*-interactions, and thus potential changes in local domain structure, we analyzed single chromosomes in more detail. We applied the above-described strategy to reveal SDs. As for WT nuclei, we focused our analysis on the right arms of Chr1, Chr4, and Chr5.

Making use of the Eigenvectors of each CA, we sought to detect potential changes in domain organization between WT and mutant nuclei. We individually performed cross-wise Pearson's correlation analyses between the different Hi-C data sets for all the three CAs (Figure 3E). Despite the observed alterations in *trans*-interaction patterns for a subset of mutants, we did not detect significant changes in the domain organization of CAs. The domain structure of all genotypes analyzed highly correlated among each other with negligible p values (Figure 3F). Consistent with this observation, we did not detect significant changes in SD organization when performing WW tests on the three CAs. As the only exception, we observed a minor change on the right arm of Chr1 when comparing *crwn1* to both WT and *crwn4*. We found an accentuated boundary between two SDs; this boundary encompassed the *CRWN1* gene and, in the *crwn1-1* mutant,

the transfer DNA (T-DNA) insertion that caused the mutation (Figure 3F).

Hence, the SD organization of CAs appears to be a robust hallmark of chromosomal architecture, which is not significantly altered by mutations that affect nuclear size.

### Distance-Dependent Decay of Interactions

Using distance-dependent mean interaction values, we can describe how IFs are coupled to the genomic distance of a given interaction pair. Previously, the distance-dependent decay of interactions, measured by IDEs, has been used to characterize chromatin packaging, specifically whether chromatin organization follows the EGM or FGM (Lieberman-Aiden et al., 2009).

IFs were shown to decay in a power-law function with an exponent of −0.867 (Figure 4A), consistent with previously described IDEs in *Arabidopsis* (Grob et al., 2013) and other organisms (Lieberman-Aiden et al., 2009; Sexton et al., 2012; Zhang et al., 2012). The variation of single chromosome IDEs was low, suggesting that all chromosomes share a common organization. To analyze how the IDE relates to different chromatin states, we calculated IDEs separately for PRs and for CAs (Figures 4B and 4C). Whereas variation within CAs and PRs was small ( $sd_{CA} = 0.02$ ,  $sd_{PR} = 0.07$ ), we noticed clear differences in IDE values between them. The mean IDE of PRs was −1.243 (Figure 4C), whereas CAs exhibited a smaller mean IDE of −0.704 (Figure 4B). Different IDEs of heterochromatic and euchromatic regions indicate a fundamentally different chromatin organization.

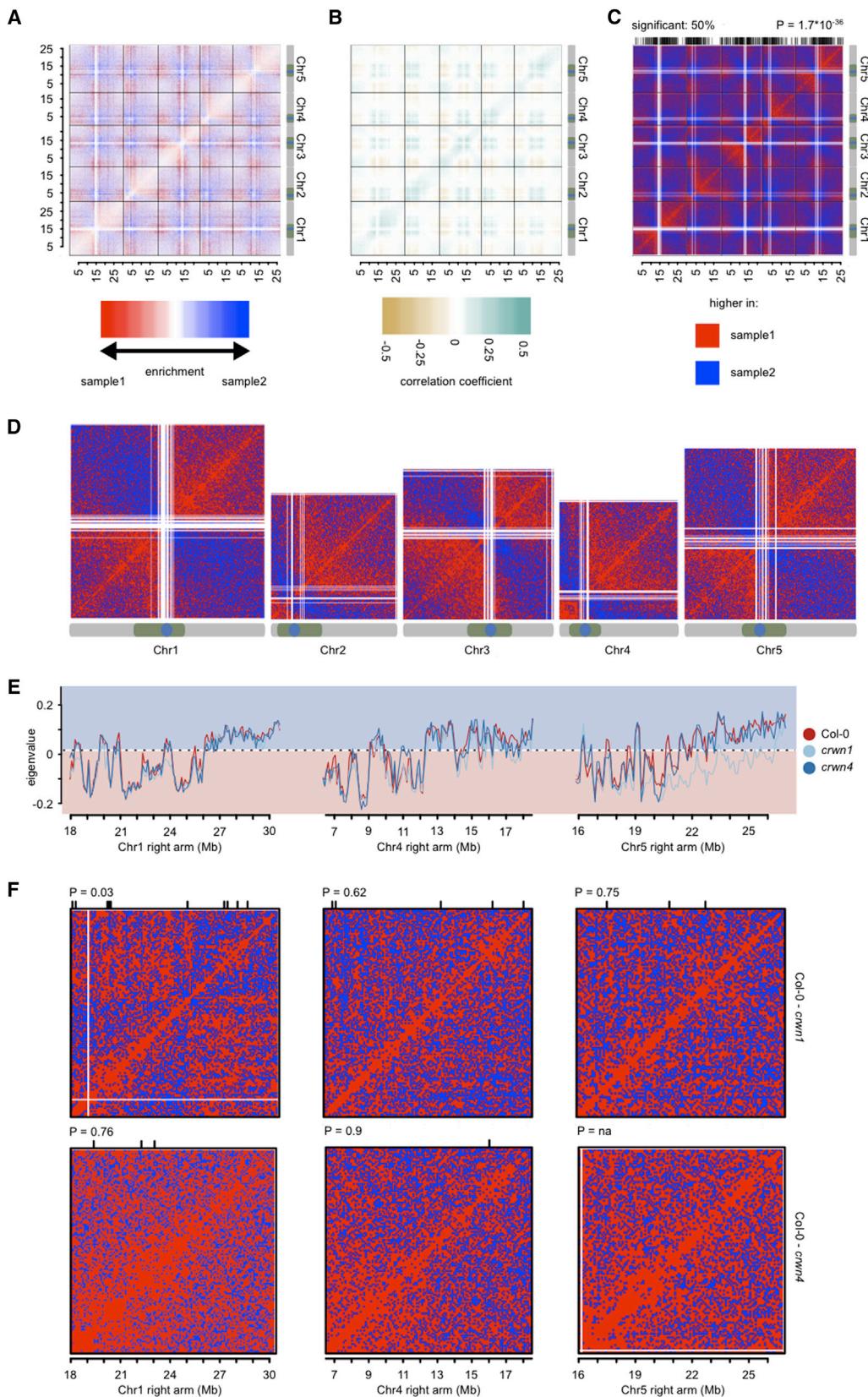
To reveal whether mutations affecting nuclear morphology such as *crwn1* and *crwn4* affect overall chromatin organization, we determined their genome-wide IDEs ( $IDE_{crwn1} = -0.834$ ,  $IDE_{crwn4} = -0.846$ ). These values are in agreement with the FGM of chromatin organization (Figure 4D). IDEs of PRs, however, exhibited clear differences between WT and mutant nuclei, implying differences in chromatin packaging. Pericentromeric IDEs of *crwn1* and *crwn4* were significantly higher than those of the WT ( $IDE_{crwn1} = -1.09$ ,  $IDE_{crwn4} = -1.02$ ; t test,  $p_{crwn1} = 0.006$ ,  $p_{crwn4} = 0.001$ ). This suggests an FGM of chromatin organization in PRs of mutant nuclei (Figure 4D).

In summary, Hi-C data sets differed considerably when their IDEs were calculated separately for PRs and CAs, indicating distinct packaging of these chromatin domains.

### Specific Chromosome Interactions Form the KNOT

Visualizing raw Hi-C data, we observed discrete dots, likely representing highly specific interactions (Figures 1A and 5A). These dots seemed to connect a unique set of ten genomic regions, which appeared to interact almost exclusively among each other with high frequency (Figures 1A–1C and 5A). We concluded that all these genomic regions form an interacting structure that, in reminiscence of the nondisentangleable Gordian Knot (Plutarch, 1727), we termed the *KNOT*. The *KNOT* consists of both long- and short-range intrachromosomal as well as interchromosomal interactions. We found regions involved in the *KNOT* to reside on all chromosomes and named them *KNOT ENGAGED ELEMENT1* (*KEE1*) to *KEE10* (Figures 5B and 5C).

To unravel the nature of the ten *KEEs*, we identified their exact genomic position. We visualized each interaction pair of the *KNOT* separately at high resolution and estimated the genomic



(legend on next page)

coordinates of regions comprising the high-frequency interaction. As we expected a selected *KEE* to interact with all other *KEEs* with a defined core region, we hypothesized that this core should be reflected by the overlap of all pairwise interactions of the other *KEEs* with the selected *KEE*. Thus, we calculated the minimal overlap of all highly interacting regions for each *KEE*. With only one exception, all estimated core *KEE* positions overlapped each other (Figures 5B and S4), indicating that all *KEEs* interact within the *KNOT* with the same core position.

### Fluorescence In Situ Hybridization Confirms the Existence of the *KNOT*

To independently confirm the robustness of the Hi-C data and the existence of the *KNOT*, we performed fluorescence in situ hybridization (FISH) on *Arabidopsis* seedling nuclei. We hybridized bacterial artificial chromosomes (BACs) to the chromatin of fixed leaf nuclei (Table S3). We selected BACs either encompassing *KEEs* or randomly chosen control regions. In each FISH experiment, we chose two distinctly labeled BACs in different combinations. These yielded nuclei in which either two *KEEs*, one *KEE* and one random region, or two random regions were labeled with different fluorescent markers (Figure 5D). Subsequently, association events between the two differentially labeled regions were analyzed by microscopy (Table 1; Figure 5F). As expected, we observed the highest association rates between regions located on the same chromosome, irrespective of whether the BACs encompassed *KEEs* or random regions.

However, we generally observed higher association rates between *KEEs* than between random regions. Strikingly, even *KEEs* separated by 20 Mb on different CAs showed higher association rates than a *KEE* and a random region located on the same CA and separated by only 6.1 Mb (Figure 5F). To analyze how the observed association rates relate to Hi-C interaction data, we performed *in silico* chromosome conformation capture (3C) experiments by calculating the sum of interactions between two regions (Figure 5E). Subsequently, by comparison of the Hi-C interaction values with the FISH association rates, we found the same interactions ranking high or low, respectively, in *in silico* 3C and FISH experiments (Figures 5E and 5F).

In summary, we could confirm the high IFs among *KEEs* by FISH and found comparable interaction and association rates, respectively, between FISH and Hi-C data.

### *KEEs* Share Common Sequence Motifs

To better understand specific interactions among *KEEs*, we searched for common characteristics, such as sequence similarity. We extracted regions with high similarity using cross-wise

alignments generated by the BLAST-like alignment tool (BLAT) (Kent, 2002), and we then refined the analysis with the motif search tool MEME (Bailey and Elkan, 1994). The highest similarity was detected for *KEE3*, *KEE4*, *KEE6*, *KEE7*, and *KEE9*, for which two motifs of 195 bp (motif1) and of 70 bp (motif2) were found (Figure S4).

To identify the genomic position of these motifs, we performed BLAST searches and found that motif1 corresponded to TEs of the *ATLANTYS3* (LTR/Gypsy superfamily) and motif2 to *VANDAL6* (DNA MutR superfamily) families. Although not identified in the MEME search, we found *KEE2* and *KEE5* to exhibit significant sequence similarity with one of the two motifs. For the remaining *KEEs*, searching the genome with the sequence obtained in the BLAT-alignment, we found *ATLANTYS2* and a *TNAT1A* family DNA transposon (*KEE1*), *ATREP3*, *ATREP2*, and *VANDAL8* (*KEE8*), and *ATLANTYS3* and *VANDAL6* (*KEE10*).

In addition to the *KEEs*, we detected several other genomic regions that share sequence similarity with the motifs. These regions harbored *ATLANTYS3* and *VANDAL6* (Figure S4). We tested for increased IFs between these regions sharing sequence similarity with the *KEEs*. While *KEEs* exhibited significantly higher IFs among each other than with randomly chosen genomic bins ( $p = 0.0004$ ), no enrichment of IFs was observed among regions sharing sequence homology to *KEEs* ( $p = 0.2931$ ).

In summary, *KEEs* exhibit high sequence similarity, mainly corresponding to *ATLANTYS3* and *VANDAL6*. However, sequence similarity among *KEEs* is unlikely the crucial factor for formation of the *KNOT* because other genomic regions with sequence similarity to *KEEs* showed similar TE compositions but did not interact at high frequency.

### *KEEs* Show a Specific Enrichment of Epigenetic and Genomic Features

As shown in this study, epigenetic features correlate with the interaction potential of a given region. To reveal common features, we analyzed the epigenetic landscape of *KEEs* (Figures 6A and 6B; Table S4). We observed a significant 2.7-fold enrichment of smRNAs associated with genomic regions surrounding the *KEEs* ( $p = 0.0022$ ). For all other tested epigenetic and genomic features, we could not detect a significant enrichment or depletion in *KEE* regions ( $\alpha = 0.05$ ; minimal enrichment or depletion: 1.5-fold).

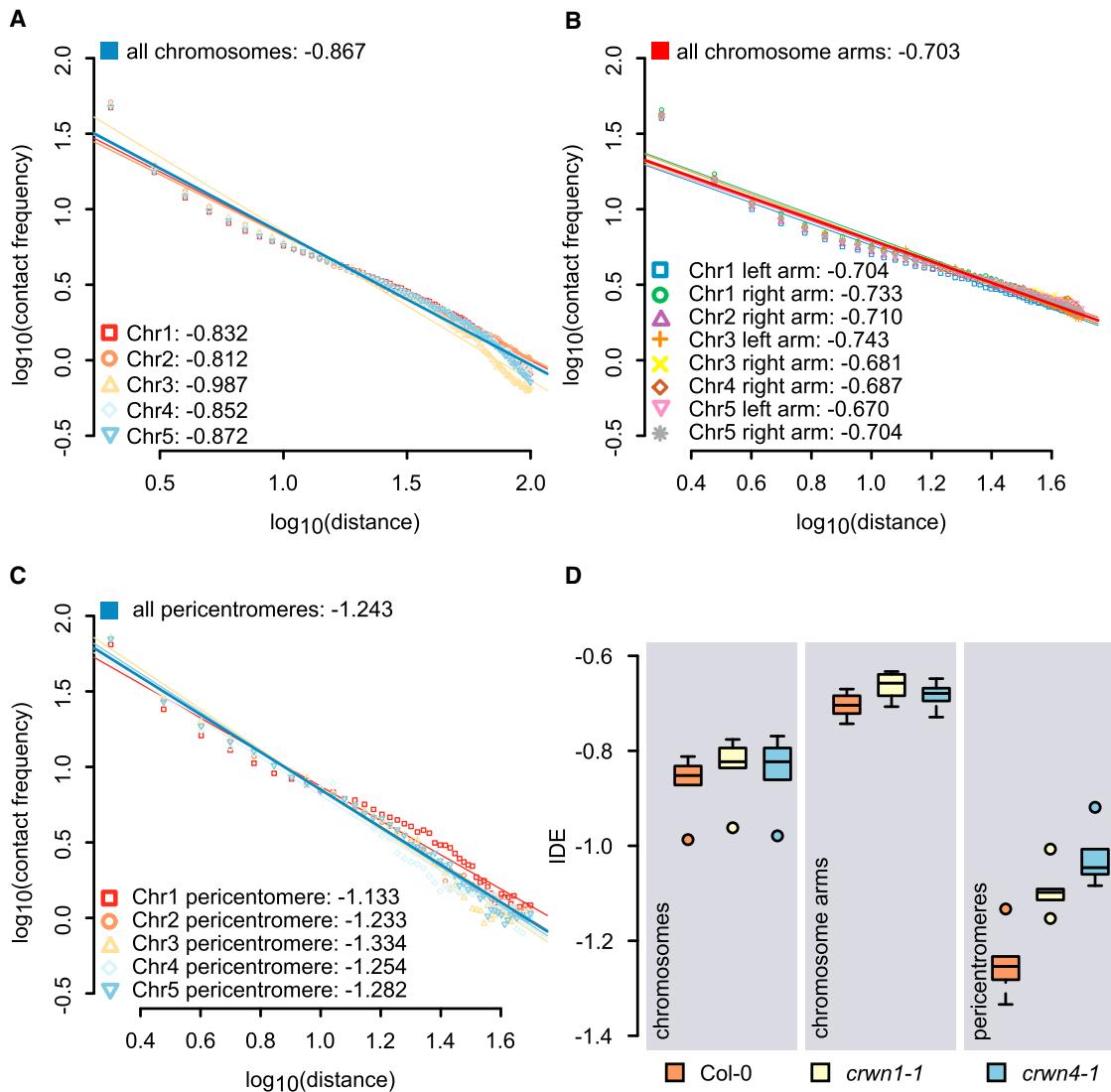
We hypothesize that *KEEs* are not epigenetically homogeneous as they are located in both PRs and CAs. If a genomic or epigenetic feature is characteristic for all *KEEs*, we postulate that the variance in density of that feature would be lower among *KEEs* than among randomly selected regions. However,

### Figure 3. Comparison of WT to *crwn* Mutants

- (A) Enrichment of IFs obtained by calculating the relative difference between Col-0 and *crwn4*.
- (B) Pearson's correlation coefficients of differences between Col-0 and *crwn4*.
- (C) Visualization of the SDM between Col-0 and *crwn4*.
- (D) SDMs between Col-0 and *crwn4* for individual chromosomes.
- (E) Comparison of the Eigenvectors of the right arms of Chr1, Chr4, and Chr5.
- (F) Visualization of SDMs of individual CAs.

The lines on top of the SDM plots (C, E) indicate the location of genomic bins exhibiting significant ( $\alpha < 0.01$ ) clustering of either positive or negative changes. (A)–(F) Genomic bin size: 100 kb.

See also Figure S3.

**Figure 4. IDEs**

- (A) IDEs along chromosomes.
- (B) IDEs along CAs.
- (C) IDEs along PRs.
- (D) Distribution of IDEs of the full genomes, CAs, and PRs for WT, *crwn1*, and *crwn4*.

In (A)–(C) dots represent average IFs between two regions of a given distance. The lines represent the fit of a linear model.

none of the investigated features varied significantly ( $\alpha = 0.05$ ) less than expected. Consequently, we refined the analysis by considering only euchromatic *KEEs* (*KEE1*, *KEE3*, *KEE4*, *KEE6*, *KEE7*, *KEE8*, and *KEE9*) to reveal significantly enriched features. As in the above-described analysis for all *KEEs*, we found that smRNAs associated with *KEE* regions of 50 kb exhibited a significant 3.5-fold enrichment ( $p < 0.0001$ ). In line with the observed enrichment of *VANDAL6* and *ATLANTYS3*, TEs were found two times more often in euchromatic *KEEs* than expected ( $p = 0.0033$ ). Additionally, the heterochromatic mark H3K27me1 was 1.9-fold enriched ( $p = 0.0119$ ) (Figures 6A and 6B; Table S4).

To confirm the robustness of these results, we repeated the analysis by testing for enrichment of a given feature within *KEE* regions of various size, i.e., 20, 50, 100, 150, 200, and 300 kb (Table S4). Whereas significant enrichments of smRNAs and H3K27me1 were observed in all window sizes tested, the enrichment of TEs was only significant in *KEE* regions of 50 and 100 kb. However, we additionally observed significantly increased transcription rates in *KEEs*, considering windows of 150, 200, and 300 kb.

Although rather heterogeneous concerning their epigenetic landscape, we conclude that *KEEs* in euchromatic CAs represent heterochromatic islands characterized by abundant

TEs, robust enrichment of smRNAs, and elevated levels of H3K27me1.

### **KEEs Are Preferred TE Insertions Sites**

The occurrence of TEs, as well as the enrichment of smRNAs, led to the question whether *KEEs* play a role in TE processing, e.g., *KEEs* may represent a preferred TE landing site. A large number of insertion lines, consisting of several thousand independent events, are available in *Arabidopsis*. The majority of these lines were generated by *Agrobacterium*-mediated insertion of T-DNAs (SALK [Alonso et al., 2003], SAIL [Sessions et al., 2002], GABI-Kat [Kleinboelting et al., 2012], and FLAG [Samson et al., 2002]). Insertion lines created at Cold Spring Harbor Laboratory (CSHL) (Sundaresan et al., 1995) and RIKEN (Kuromori et al., 2004) were generated by the activation of a *Dissociation* (*Ds*) transposon and represent a collection of individual TE insertions. Wisconsin *Ds*Lox T-DNA lines (WISC) (Woody et al., 2007) were generated by *Agrobacterium*-mediated T-DNA insertion, but the vector also contained a *Ds* element.

We gathered information about the insertion sites of all available insertion lines from the SIGNAL database and tested for a preferential insertion into *KEEs*. We compared insertion frequencies within *KEEs* with insertion frequencies of 10,000 random sets of genomic regions. From the seven tested insertion collections, the two *Ds* populations (CSHL, RIKEN) exhibited a significant enrichment of insertions within *KEEs* ( $p_{\text{CSHL}} = 0.0003$ ,  $p_{\text{RIKEN}} = 0.0008$ ) (Figure 6D). All other analyzed collections, which were generated by T-DNA transformation (SALK, SAIL, GABI, FLAG, WISC), did not show significantly enriched insertion frequencies (Table S4). We also analyzed insertion sites of the retrotransposon *EVADÉ* (Marí-Ordóñez et al., 2013), which was reactivated in backgrounds with reduced DNA methylation (Mirouze et al., 2009). From 11 new *EVADÉ* insertions, 4 inserted within 250 kb of a *KEE* (Figure 6D).

In *Drosophila*, several PIWI-interacting RNA (piRNA) clusters are involved in TE silencing (Brennecke et al., 2007; Malone et al., 2009). Thus, we asked whether *Drosophila* piRNA clusters exhibit a similar interaction pattern as *KEEs* in *Arabidopsis*. Indeed, by inspection of previously published *Drosophila* Hi-C data (Sexton et al., 2012), we found significantly ( $p < 0.0001$ ) enriched IFs between genomic regions harboring piRNA clusters (Brennecke et al., 2007) (Figure 6C).

In summary, we show that a *KNOT*-like structure is also formed by piRNA clusters in *Drosophila* and that *KEEs* are preferential insertion sites for TEs, suggesting a role in TE biology and thus genome integrity.

## **DISCUSSION**

### **There Is No Distinct Chromosomal Neighborhood for a Given Chromosome**

By calculating the deviation from the expected *trans*-IF between chromosomes, nuclear neighborhoods of CTs can be determined (Zhang et al., 2012). Compared to a study in mice (Zhang et al., 2012), the deviations from expected IFs in *Arabidopsis* nuclei are rather small. This suggests that any *Arabidopsis* chromosome has the same likelihood to stay in physical contact with any other, and that there is no preferential chromosome associ-

ation. This conclusion is in line with FISH studies showing that *Arabidopsis* chromosomes do not preferentially pair (Pecinka et al., 2004).

The small number of chromosomes in *Arabidopsis* can explain the absence of distinct chromosomal neighborhoods. The higher number of chromosomes in mouse nuclei increases the probability that a chromosome is located between another pair, thereby separating distinct CTs. Single-cell Hi-C suggested a discrete number of interchromosomal contacts in a single mouse nucleus (Nagano et al., 2013). However, these contacts vary between nuclei of the same cell type, which leads to a rather uniform distribution of interchromosomal contacts in ensemble Hi-C, indicating that the preference of interchromosomal interactions is stochastic.

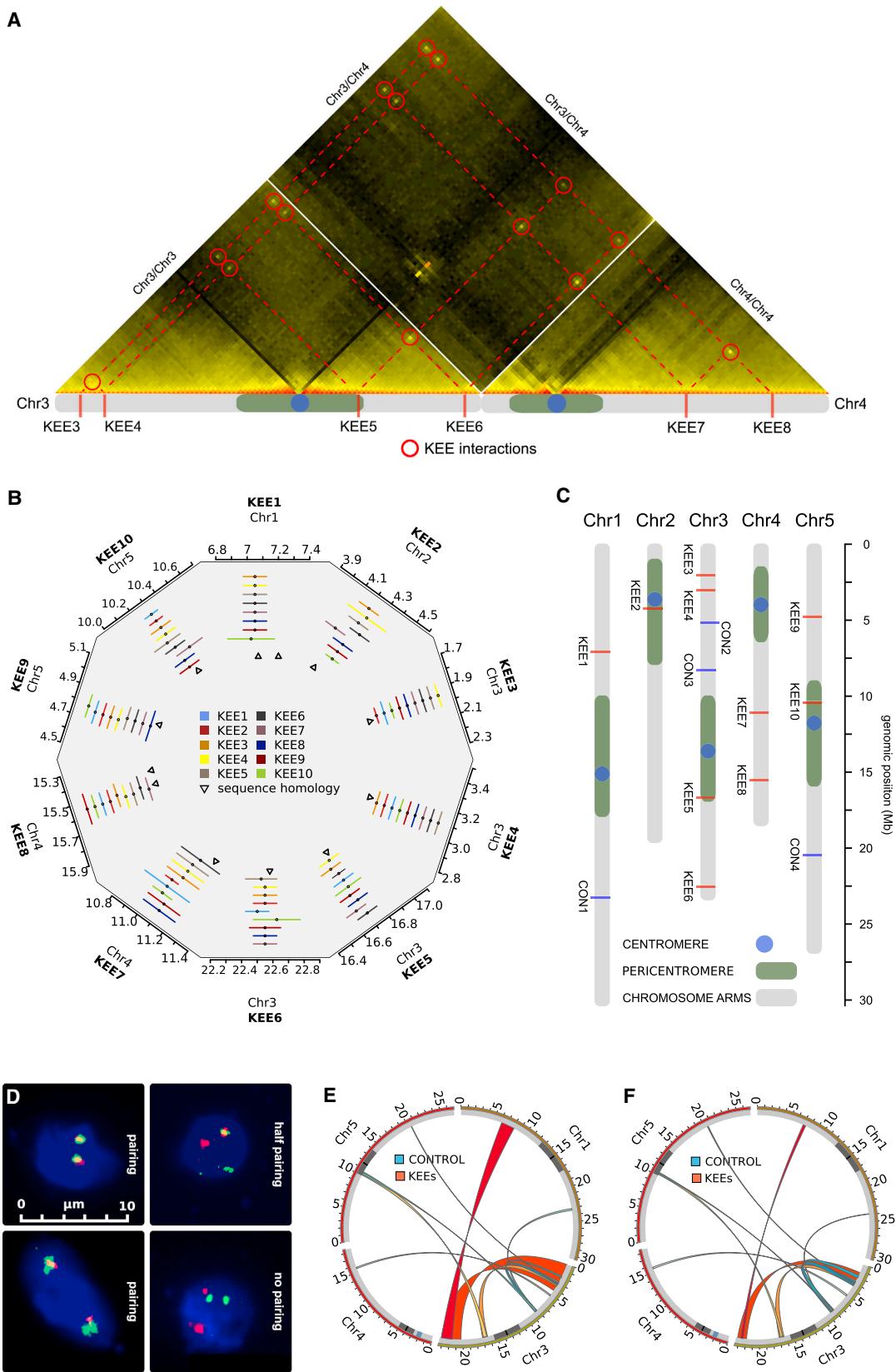
### ***Arabidopsis* Chromosomes Show a Simple Organization with Respect to Their Epigenetic Landscape and Interactome**

Our results show that the epigenetic landscape strongly correlates with chromosomal architecture. LSDs, characterized by low compaction and enriched IFs with more distal regions both in *cis* and *trans*, are associated with active epigenetic marks, whereas the more condensed CSDs are enriched in repressive epigenetic marks. The composition of these two epigenetic landscapes is reminiscent of active chromatin state (CS) 1 and repressive CS2 (Roudier et al., 2011).

LSDs and CSDs resemble A and B compartments described in human cells. Similar to LSDs and CSDs, regions of the A compartment are less densely packaged than genomic regions of the B compartment (Lieberman-Aiden et al., 2009). The two classes of SDs in our study were distinguished by their inherent interaction potential. Thus, a single LSD or CSD can be subdivided into consecutive SDs with a similar interaction potential. These subdomains could be compared to topologically associating domains (Bickmore and van Steensel, 2013). However, we generally observed SDs to alternate in LSDs or CSDs, which seem to act as boundaries for each other.

*Arabidopsis* chromosomes show a rather simple organization with regard to the occurrence of constitutive heterochromatin and euchromatin. In all chromosomes, except Chr4, constitutive heterochromatin is solely found within PRs, whereas euchromatin is associated with CAs. The only additional region of constitutive heterochromatin of significant size, the knob *hk4s*, is on the short arm of Chr4 (Fransz et al., 2000; Grob et al., 2013). The organization of CAs is surprisingly homogenous, as all CAs exhibit increasing active marks, and therefore increasing occurrence of LSDs, toward distal positions. This makes it difficult to distinguish distinct SDs for a number of CAs.

The simple chromatin organization in *Arabidopsis* contrasts that of mammalian nuclei, in which CAs are clearly divided into numerous consecutive SDs (Lieberman-Aiden et al., 2009; Zhang et al., 2012). However, *Drosophila* nuclei exhibit a rather simple chromatin organization similar to that of *Arabidopsis* (Sexton et al., 2012). As the most conspicuous difference between mammalian genomes and those of *Drosophila* and *Arabidopsis* is their size, we propose that the highly compact nature of these genomes explains the apparent absence of structurally complex CAs.



(legend on next page)

**Table 1. FISH Association Rates and Hi-C Interaction Scores**

Probe 1	Probe 2	FISH Association Rate (%)	Hi-C Interaction Score
KEE6	KEE1	20	87.43
CON3 <sup>a</sup>	CON1	3	5.44
CON3	KEE3	21	7.65
CON3	KEE4	31	8.36
KEE5	KEE4	35	34.96
KEE6	KEE3	66	92.39
KEE8	CON2	12	5.8
KEE5	KEE10	16	18.61
CON3	KEE10	9	4.11
CON4	KEE4	7	2

See also Table S3.

<sup>a</sup>CON, control BAC.

### Nuclear Morphology Affects *trans*-Chromosomal Interactions but Not Domain Structure in *Arabidopsis* Nuclei

CRWN proteins are important structural components of *Arabidopsis* nuclei, with *crwn1* and *crwn4* mutants affecting nuclear morphology (Dittmer et al., 2007; Sakamoto and Takagi, 2013). *crwn1* and *crwn4* nuclei exhibited increased *trans*-interactions compared to WT nuclei, suggesting higher chromosomal compaction. As the size of *crwn1* and *crwn4* nuclei is substantially smaller than that of WT, we suggest that increased *trans* IFs are the consequence of size constraints, within which CTs have to be organized.

As a hallmark of chromosomal architecture in *crwn4* and, to a lesser extent, in *crwn1* nuclei, we observed increased interactions between PRs. Similarly, a reduced number of chromocenters and a disruption of chromocenter organization were cytogenetically observed in *crwn4* mutants (Wang et al., 2013). We conclude that this reduced number of observable chromocenters does not depend on chromatin decondensation but relates to an increased frequency of PR pairing, leading to the merging of PR territories.

The increased nuclear compaction in *crwn4* and *crwn1* nuclei is most obvious in the general increase of *trans*-arm interactions. In contrast, local chromatin organization within individual CAs is not grossly affected. This is evident by the unchanged occurrence of CSDs and LSDs within individual CAs. We conclude that chromosomes are organized in a hierarchical manner, in which CAs appear to be a stable unit, largely unaffected by physical constraints of nuclear morphology. However, CTs appear to be influenced by nuclear morphology. With less space available,

two CA territories are forced into closer spatial proximity. Last, contacts between individual chromosomes appear to vary with nuclear size.

Variability in nuclear size and morphology is surprisingly high in *Arabidopsis*, which should influence *trans*-chromosomal interactions. However, much of this variation cannot easily be related to the transcriptional state of cells. Our results could provide a possible explanation for the lack of this relationship. The epigenetic landscape, and thus the transcriptional state of a cell, is mainly associated with the occurrence of SDs within CAs, which were shown to be largely independent of nuclear morphology.

### Stochastic Variability between Interactomes Has to Be Carefully Assessed to Draw Biologically Relevant Conclusions

Chromosomal architecture is prone to stochastic variation, which is unlikely caused by important biological processes (Nagano et al., 2013). Therefore, careful assessment of this variability is essential for a conclusive evaluation of comparisons between different Hi-C data sets. We suggest an analytical pipeline to quantify stochastic variability, making use of the axiom that neighboring genomic bins should exhibit correlative interaction profiles.

The inspection of plain difference matrices bears the risk of overestimating the observation of patterns within these matrices. Hi-C interaction matrices are often visualized in symmetrical plots that represent a mirror image of the actual interactome, representing each interaction twice. This visualization method pronounces apparent patterns within the matrix, which would probably not be perceived as a distinct structure in a non-symmetrical visualization of the matrix. Analyzing correlative differences between two given Hi-C interaction data sets aids in a better understanding of the biological relevance of changes in Hi-C interactomes. Even more powerful, as it allows a statistical investigation of changes, is the analysis of whether clustered changes occur in SDMs, providing an even deeper insight into alterations of chromatin organization between different Hi-C data sets. As a major advantage, this method not only reveals genomic locations that undergo significant changes, but also provides an overall estimate of the difference between two interactomes by the total number of significant changes observed between them.

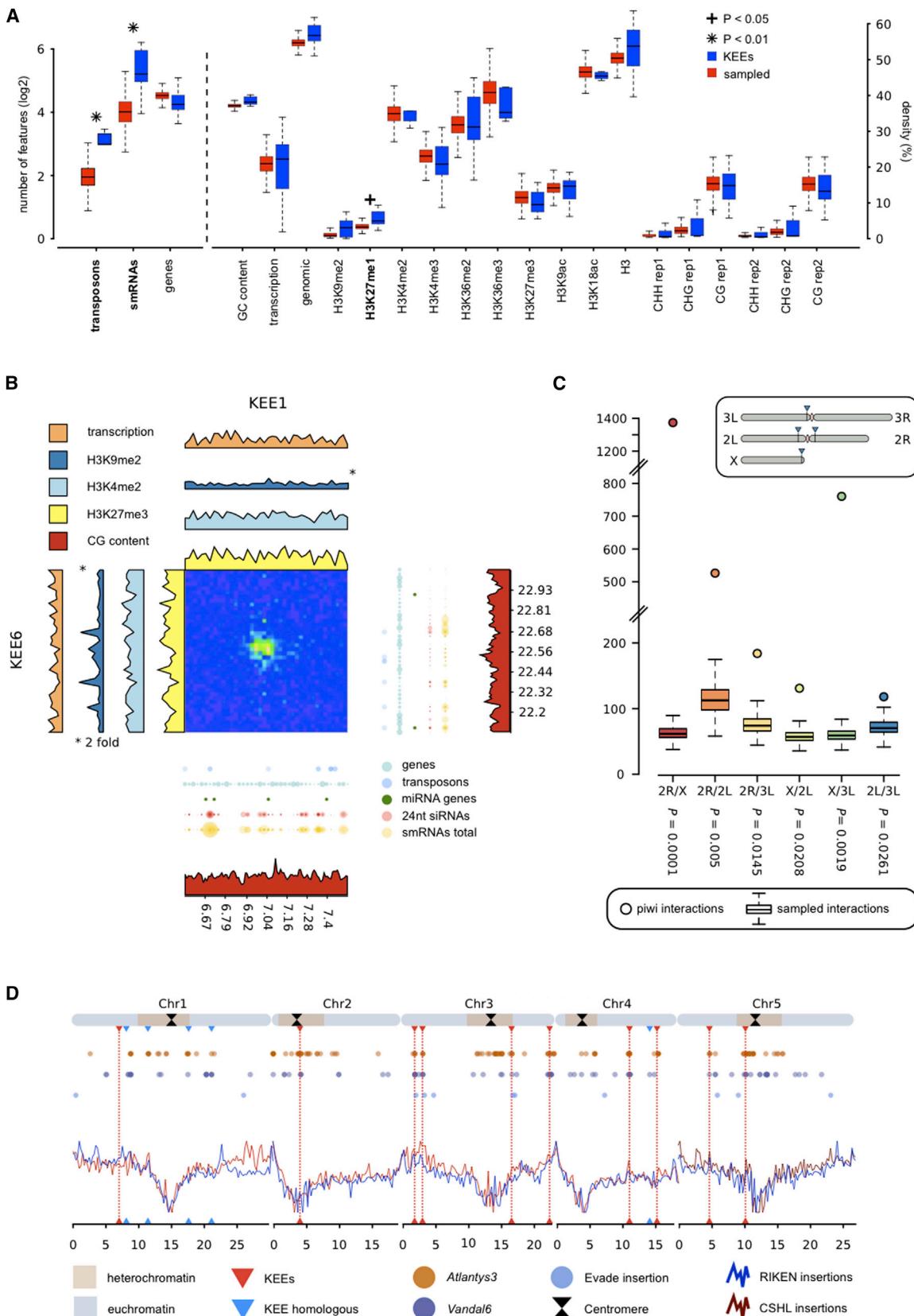
### IDEs Indicate a Distinct Chromatin Organization of CAs and PRs

Most reported IDEs are close to the theoretical IDE of the FGM (*Drosophila*,  $-0.85$  [Sexton et al., 2012]; mouse,  $-1.03$  [Zhang et al., 2012]; human,  $-1.08$  [Lieberman-Aiden et al., 2009]),

**Figure 5. Positioning of KEEs, Shared Sequence Motifs, and FISH Validation**

- (A) Close-up of the interactions between and within Chr3 and Chr4. Red circles indicate high-frequency interactions between KEE regions.
- (B) Estimated genomic intervals with the highest IF between a given KEE and all other KEEs (lines) and genomic positions of sequence homology among KEEs (triangles).
- (C) Overview of the genomic positions of the KEEs on the five *Arabidopsis* chromosomes.
- (D) Examples of FISH-analyzed nuclei. BACs are stained red and green, whereas DNA is stained blue.
- (E) Circos plot of a virtual 3C experiment between KEE and control regions.
- (F) Circos plot of FISH association rates.
- (E and F) Red, interactions between KEEs; blue, interactions between control regions and between control regions and KEEs.

See also Figure S4.



(legend on next page)

indicating that the fractal globule organization is a conserved hallmark. The genome-wide IDE calculated in the present study ( $-0.895$ ) further supports the FGM. By averaging IDEs of several circularized chromosome conformation capture (4C) experiments in *Arabidopsis*, we calculated an IDE of  $-0.73$  (Grob et al., 2013). This value differs considerably from the genome-wide IDE calculated in the present study. However, in our previous work, 4C viewpoints were exclusively chosen within CAs. When we compared the IDE obtained by 4C experiments with the mean IDE of CAs in the present Hi-C study, we observed only a small difference between the two values ( $-0.73$  and  $-0.703$ ).

Interestingly, IDEs of different chromatin states differed considerably. Whereas euchromatic CAs exhibited an IDE of  $-0.703$ , the average IDE of PRs was  $-1.243$ . The IDEs of PRs suggest a different chromatin organization, which more closely resembles the EGM. This is not surprising as heterochromatin can easily be distinguished from euchromatin by its appearance. Therefore, accessibility, which is facilitated in a fractal globule chromatin organization, may not be a feature of heterochromatin. A different chromatin organization, such as the equilibrium globule organization, could be favorable to fulfill the requirements for heterochromatin.

Previous observations in *Drosophila* suggested that active chromatin exhibits a different IDE than regions characterized by repressive epigenetic marks (Sexton et al., 2012). These IDEs are contrasting our results, as the IDE of heterochromatic PRs showed a higher IDE ( $-0.7$ ) than euchromatic CAs ( $-0.85$ ). However, the IDE of repressive epigenetic regions described in *Drosophila* cannot easily be compared to the IDE of constitutive heterochromatin of PRs described in our study. Sexton et al. (2012) pooled various chromatin states, namely, *Polycomb*-silenced chromatin, chromatin bound by Heterochromatin Protein 1, centromeric chromatin, and chromatin that was not enriched in any epigenetic mark (null state). In contrast, the heterochromatin of *Arabidopsis* PRs represents a homogeneous epigenetic state, likely explaining the different IDEs in the two studies.

In accordance with the unchanged SD organization of CAs in *crwn* mutants, the IDEs of CAs in *crwn4* and, to a lesser extent, in *crwn1* resembled IDEs of CAs in the WT. In contrast, the IDEs of PRs were indicative for the FGM and therefore significantly differed from the WT. It is unclear, whether this alteration in the organization of PRs is solely inflicted by reduced nuclear volume or by a function of CRWN4 in centromere organization. Since *crwn1* nuclei are at least as small as *crwn4* nuclei, but disrupted heterochromatic PRs have only been reported in *crwn4* (Wang et al., 2013), the different IDEs of the two mutants in Hi-C experiments support such a function.

In summary, *Arabidopsis* chromosomes are globally organized according to the FGM. However, the PRs are likely orga-

nized differently than euchromatic CAs, which can be explained by the fundamentally different roles the two chromatin states play in the nucleus.

### The KNOT Plays a Role as a Transposon Trap Similar to the *flamenco* Locus in *Drosophila*

As an unexpected, conspicuous feature of the interactome, we observed distinct high IFs between ten *KEEs*, resulting in a web of interactions that we termed *KNOT*. Although *KEE* regions varied among each other with respect to their epigenetic constitution, we observed significant enrichment of associated smRNAs in all *KEE* regions. As *KEEs* were found in fundamentally different chromatin states, such as constitutive heterochromatin of PRs and euchromatic CAs, we did not expect *KEEs* to represent an epigenetically uniform group. By solely considering *KEEs* embedded in euchromatin, we detected an enrichment of H3K27me1 and TEs, suggesting that *KEEs* are heterochromatic islands within euchromatin. However, *KEE* regions are not generally silenced, as actively transcribed genes were detected within them.

*Ds* transposons preferentially insert in the proximity of *KEEs*. Interestingly, preferential insertion appears to be limited to TEs as we did not observe enriched T-DNA transgene integration near *KEE* regions. The mechanism leading to preferential insertion of TEs within *KEEs* is not solely based on sequence identity of the TEs, as transgenes carrying a *Ds* transposon (WISC lines) were not found to be enriched.

Active TEs potentially harm genome integrity, as TE insertions can disrupt genes and important regulatory elements. Therefore, plants developed a sophisticated TE defense system that relies largely on the RNAi machinery, leading to either posttranscriptional gene silencing or RNA-directed DNA methylation (Castel and Martienssen, 2013). The observed enrichment of new *Ds* insertions and smRNAs, which are associated with *KEE* regions, leads us to propose that the *KNOT* may play a role in TE defense. The *KNOT* might act as a TE trap or relay station from which TEs are either excised or redirected to a TE safe house, such as the PRs.

In *Drosophila*, several piRNA clusters, including the *flamenco* locus, are involved in TE silencing (Brennecke et al., 2007; Malone et al., 2009). Interestingly, *Drosophila* piRNA clusters show similar chromatin interactions as *KEEs*, further supporting the involvement of the *KNOT* in TE defense. Furthermore, it was recently shown that the *flamenco* locus in *Drosophila* serves as a TE trap (Zanni et al., 2013). Based on these similarities, we hypothesize that the *KNOT* is a conserved nuclear structure and plays a similar role as piRNA clusters in *Drosophila*, and that there are nuclear structures analogous to the *KNOT* in other eukaryotes.

**Figure 6. The KNOT: Epigenetic and Genomic Features and TE Insertion Sites**

- (A) Distributions of epigenetic and genomic features in *KEEs* (blue) and sampled regions (red). Features that significantly differ in several bin sizes are marked bold.
  - (B) Interaction between *KEE1* and *KEE6* along 1 Mb. H3K9me2 tracks were 2-fold exaggerated for better visibility.
  - (C) Interactions among piRNA clusters. Dots represent IFs between piRNA clusters (spanning nine genomic bins of 80 kb each). Boxes indicate IFs of 10,000 randomly sampled regions, selected on chromosomes (ChrX) or CAs (2R, 2L, and 3L), which harbor the respective piRNA clusters. Inset: genomic positions of piRNA clusters in *Drosophila*.
  - (D) Distribution of natural TE insertion sites (dots) and TE insertion frequencies of RIKEN and CSHL lines (lines).
- See also Table S4.

## EXPERIMENTAL PROCEDURES

## Plant Material

Hi-C experiments were performed using 14-day-old *Arabidopsis thaliana* seedlings (Col-0 accession) grown on Murashige and Skoog culture medium. FISH experiments were performed on Col-0 leaf nuclei. Detailed information on plant materials and growth conditions can be found in the [Supplemental Experimental Procedures](#).

## FISH

Chromatin regions encompassing *KEEs* or control regions were hybridized with biotin- or digoxigenin-labeled BAC DNA probes ([Table S4](#)). Labeled probes were subsequently detected using secondary antibodies conjugated with fluorescent dyes (Texas Red [red] or Alexa 488 [green]). Pairing events (associations of green and red dots) were subsequently scored using fluorescence microscopy. A detailed protocol for FISH experiments can be found in the [Supplemental Experimental Procedures](#).

## Hi-C Experiments

*Arabidopsis* chromatin was crosslinked and digested using a six-cutter restriction enzyme (HindIII). Subsequently, the chromatin was religated in a large volume favoring intramolecular ligation events, yielding circular DNA molecules comprised of interacting restriction fragments. Identification and quantification of interacting partners were obtained by submitting the DNA to Illumina sequencing, providing genome-wide information on chromosome conformation. A more detailed protocol for Hi-C experiments can be found in the [Supplemental Experimental Procedures](#).

Sequencing reads were aligned to the *Arabidopsis* Col-0 reference genome (TAIR 10) without allowing mismatches and multiple alignments. For subsequent analyses, the mapped sequencing reads were pooled into genomic bins (10, 25, 50, 100, or 250 kb). We then generated matrices in which each element describes the interaction frequency of two genomic bins.

## Hi-C Data Analysis

For intrachromosomal interactions, Hi-C matrices were distance normalized by dividing the interaction frequency between two genomic bins by the average interaction frequency of all genomic bins that exhibited the same genomic distance. Subsequently, Pearson's correlation coefficients were calculated such that each element in the correlated Hi-C interaction matrix describes the correlation coefficient between two *in silico* 4C interactomes (i.e., rows and columns of the distance-normalized interaction matrix). To reveal LSDs and CSDs, respectively, PCA was performed on the correlated Hi-C interaction matrices of single chromosome arms. Genomic bins were then split into two groups according to the sign of the resulting Eigenvalue of each genomic bin (negative Eigenvalues, CSD; positive Eigenvalue, LSD).

To analyze the relationship of chromosome conformation and the epigenetic and genomic landscape, the density (e.g., for histone modifications) or the number (e.g., number of genes) of a given feature per genomic bin was calculated. Based on these values, enrichment or depletion of an epigenetic or genomic feature within LSDs was determined by performing Wilcoxon signed rank testing. Additionally, Pearson's correlation coefficients between the density or count values of a given feature and Eigenvalues of genomic bins was calculated.

SDMs were generated by calculating the sign of the difference between each element of two Hi-C interaction matrices. Subsequently, performing WW testing on each column revealed significant clustering of positive and negative signs, respectively, defining genomic bins that undergo significant architectural changes between the two Hi-C interaction data sets.

To analyze the epigenetic landscape of *KEE* regions and the interaction frequencies between piRNA clusters in *Drosophila*, a Monte-Carlo-based statistical approach was used.

A detailed description of all statistical analyses can be found in the [Supplemental Experimental Procedures](#).

## ACCESSION NUMBERS

Hi-C interaction data can be accessed under the Gene Expression Omnibus accession number GSE55960.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2014.07.009>.

## AUTHOR CONTRIBUTIONS

S.G. and U.G. conceived the study. S.G. performed the experiments. S.G. and M.W.S. performed the bioinformatic data analysis. S.G., M.W.S., and U.G. interpreted the data. S.G., M.W.S., and U.G. wrote the manuscript.

## ACKNOWLEDGMENTS

We are indebted to Erika Hughes and Eric J. Richards (Boyce Thomson Institute for Plant Research) for providing seeds of *crwn1* and *crwn4* mutants. We thank Keith Harshman from the Lausanne Genomic Technologies Facility (University of Lausanne) for hospitality during library construction, Eric J. Richards for useful comments on the manuscript, and Konstantinos Kritsas (University of Zürich) for advice on FISH. This work was supported by the University of Zürich, an iPhD project grant from SystemsX.ch, and an Advanced Grant of the European Research Council to U.G.

Received: April 4, 2014

Revised: May 15, 2014

Accepted: July 10, 2014

Published: August 14, 2014

## REFERENCES

- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657.
- Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
- Bickmore, W.A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. *Cell* 152, 1270–1284.
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089–1103.
- Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* 14, 100–112.
- Dittmer, T.A., Stacey, N.J., Sugimoto-Shirasu, K., and Richards, E.J. (2007). *LITTLE NUCLEI* genes affecting nuclear morphology in *Arabidopsis thaliana*. *Plant Cell* 19, 2793–2803.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.
- Filion, G.J., van Bemmelen, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., and van Steensel, B. (2010). Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* 143, 212–224.
- Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T., and Jones, G.H. (2000). Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell* 100, 367–376.
- Fransz, P., De Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I. (2002). Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. USA* 99, 14584–14589.

- Grob, S., Schmid, M.W., Luedtke, N.W., Wicker, T., and Grossniklaus, U. (2013). Characterization of chromosomal architecture in *Arabidopsis* by chromosome conformation capture. *Genome Biol.* 14, R129.
- Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664.
- Kleinboelting, N., Huep, G., Kloetgen, A., Viehoever, P., and Weisshaar, B. (2012). GABI-Kat SimpleSearch: new features of the *Arabidopsis thaliana* T-DNA mutant database. *Nucleic Acids Res.* 40, D1211–D1215.
- Kuromori, T., Hirayama, T., Kiyosue, Y., Takabe, H., Mizukado, S., Sakurai, T., Akiyama, K., Kamiya, A., Ito, T., and Shinozaki, K. (2004). A collection of 11 800 single-copy *Ds* transposon insertion lines in *Arabidopsis*. *Plant J.* 37, 897–905.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- Luo, C., Sidote, D.J., Zhang, Y., Kerstetter, R.A., Michael, T.P., and Lam, E. (2012). Integrative analysis of chromatin states in *Arabidopsis* identified potential regulatory mechanisms for natural antisense transcript production. *Plant J.* 73, 77–90.
- Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137, 522–535.
- Mari-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O. (2013). Reconstructing *de novo* silencing of an active plant retrotransposon. *Nat. Genet.* 45, 1029–1039.
- Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J., and Mathieu, O. (2009). Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* 461, 427–430.
- Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., et al. (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* 336, 1448–1451.
- Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64.
- Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M.A., Fuchs, J., and Schubert, I. (2004). Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. *Chromosoma* 113, 258–269.
- Plutarch. (1727). Alexander. In *Plutarch's Lives: Translated from the Greek* (London: J. Tonson).
- Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., et al. (2011). Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J.* 30, 1928–1938.
- Sakamoto, Y., and Takagi, S. (2013). LITTLE NUCLEI 1 and 4 regulate nuclear morphology in *Arabidopsis thaliana*. *Plant Cell Physiol.* 54, 622–633.
- Samson, F., Brunaud, V., Balzergue, S., Dubreucq, B., Lepiniec, L., Pelletier, G., Caboche, M., and Lecharny, A. (2002). FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants. *Nucleic Acids Res.* 30, 94–97.
- Schubert, V., Berr, A., and Meister, A. (2012). Interphase chromatin organization in *Arabidopsis* nuclei: constraints versus randomness. *Chromosoma* 121, 369–387.
- Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., et al. (2002). A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell* 14, 2985–2994.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472.
- Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D., Dean, C., Ma, H., and Martienssen, R. (1995). Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.* 9, 1797–1810.
- Wang, H., Dittmer, T.A., and Richards, E.J. (2013). *Arabidopsis* CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization. *BMC Plant Biol.* 13, 200.
- Woody, S.T., Austin-Phillips, S., Amasino, R.M., and Krysan, P.J. (2007). The WiscDsLox T-DNA collection: an *Arabidopsis* community resource generated by using an improved high-throughput T-DNA sequencing pipeline. *J. Plant Res.* 120, 157–165.
- Zanni, V., Eymery, A., Coiffet, M., Zytnicki, M., Luyten, I., Quesneville, H., Vaury, C., and Jensen, S. (2013). Distribution, evolution, and diversity of retrotransposons at the *flamenco* locus reflect the regulatory properties of piRNA clusters. *Proc. Natl. Acad. Sci. USA* 110, 19842–19847.
- Zhang, Y., McCord, R.P., Ho, Y.-J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148, 908–921.