

# 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*.

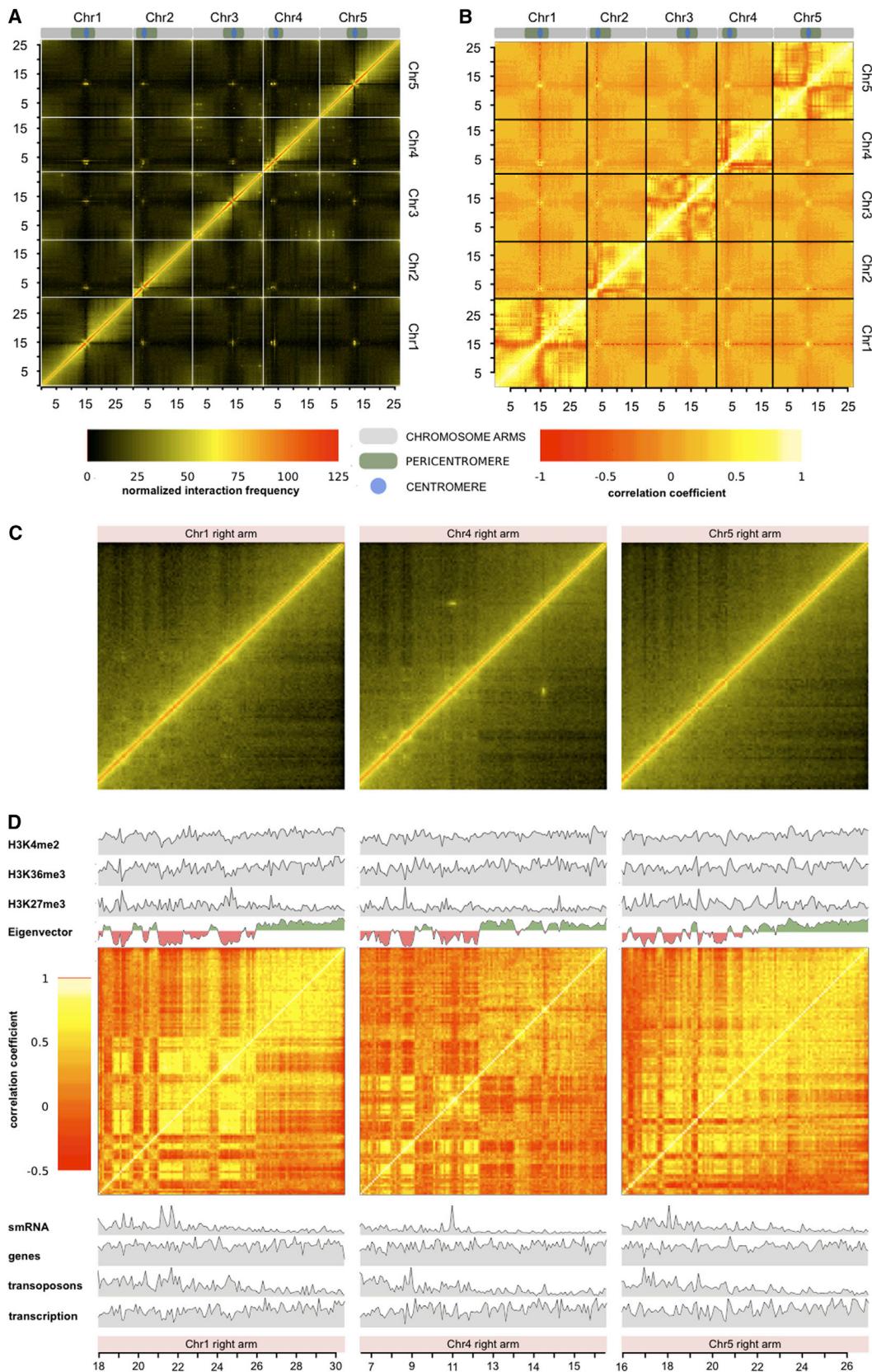
To obtain a numeric description of these SDs, we performed principal component analysis (PCA) on the correlation matrix of each individual chromosome (Chr). This led to a clear partitioning of the interactome into two SDs with either positive or negative Eigenvalues, with negative and positive Eigenvalues 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.

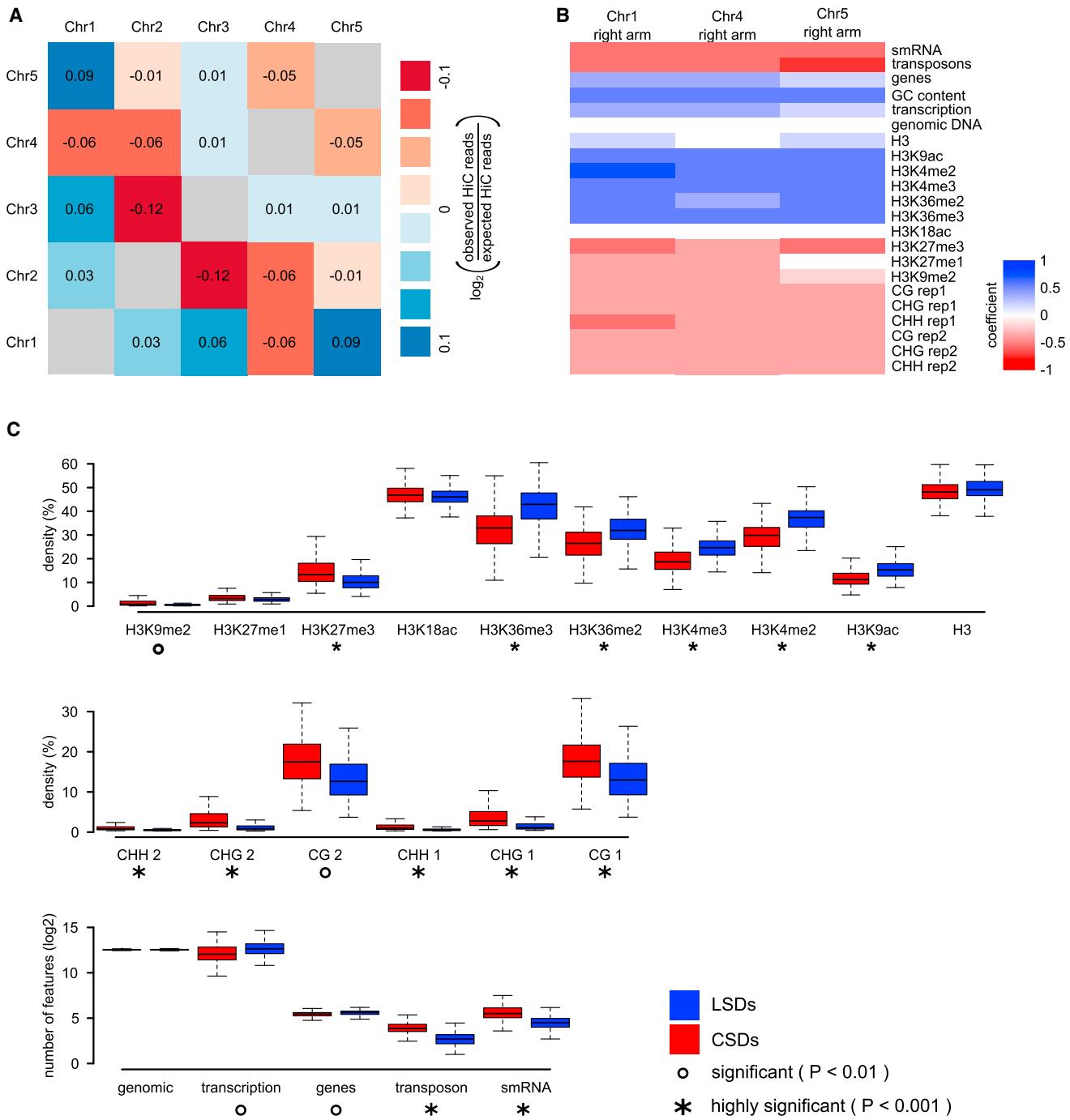
To investigate the impact of the *crwn1* and *crwn4* mutants, we calculated differences between all obtained Hi-C data according to a previously described method (Moissiard et al., 2012) (Figures 3A and S3). In short, we calculated the difference between all elements of two Hi-C matrices of interest. 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.

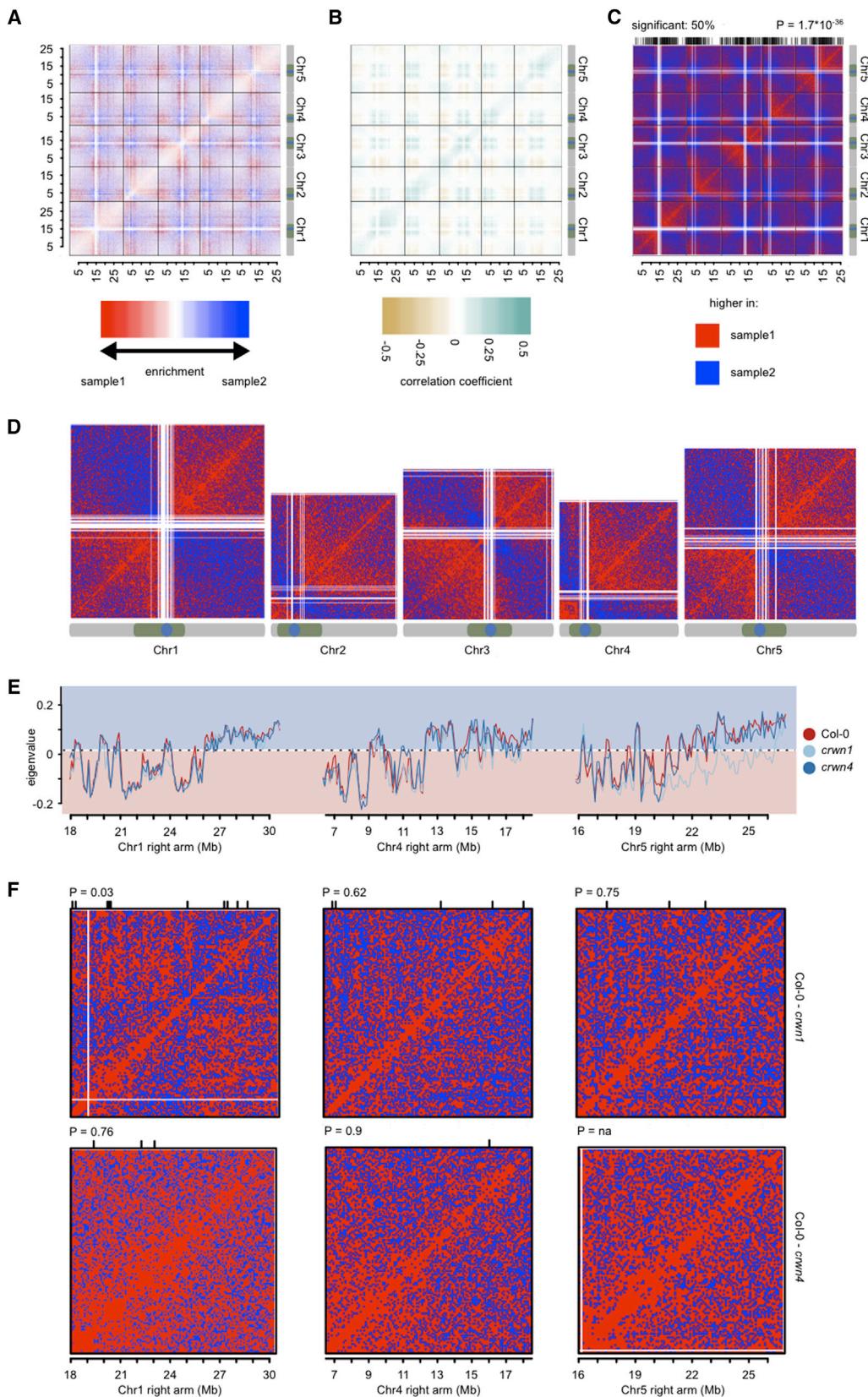
To reveal whether mutations affecting nuclear morphology such as *crwn1* and *crwn4* affect overall chromatin organization, we determined their genome-wide IDEs ( $IDE_{crwn1} = -0.834$ ,  $IDE_{crwn4} = -0.846$ ). These values are in agreement with the FGM of chromatin organization (Figure 4D). IDEs 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).

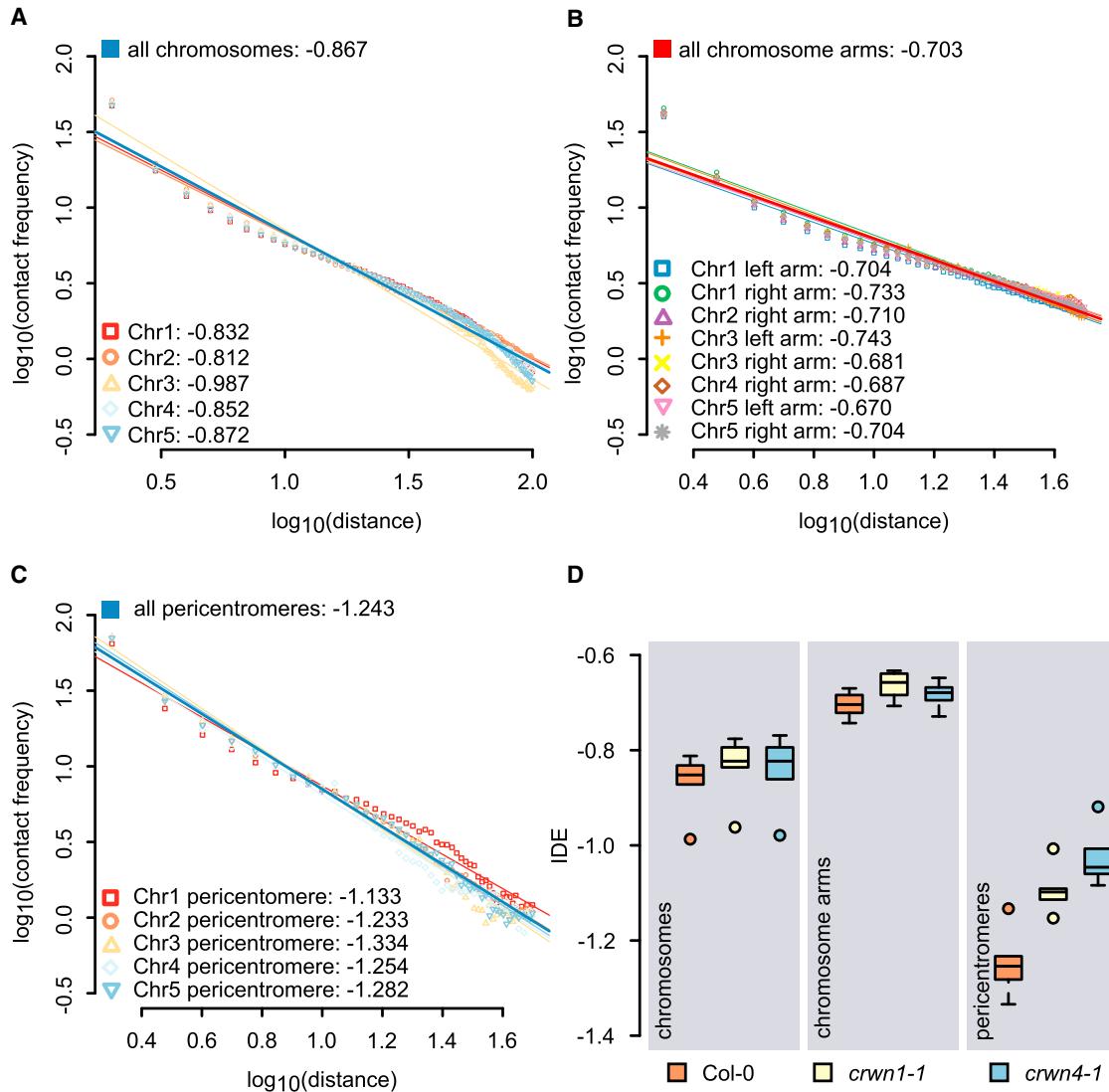
We hypothesize that *KEEs* are not epigenetically homogeneous as they are located in both PRs and CAs. If a genomic or epigenetic feature is characteristic for all *KEEs*, we postulate that the variance in density of that feature would be lower among *KEEs* than among randomly selected regions. However,

### 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.

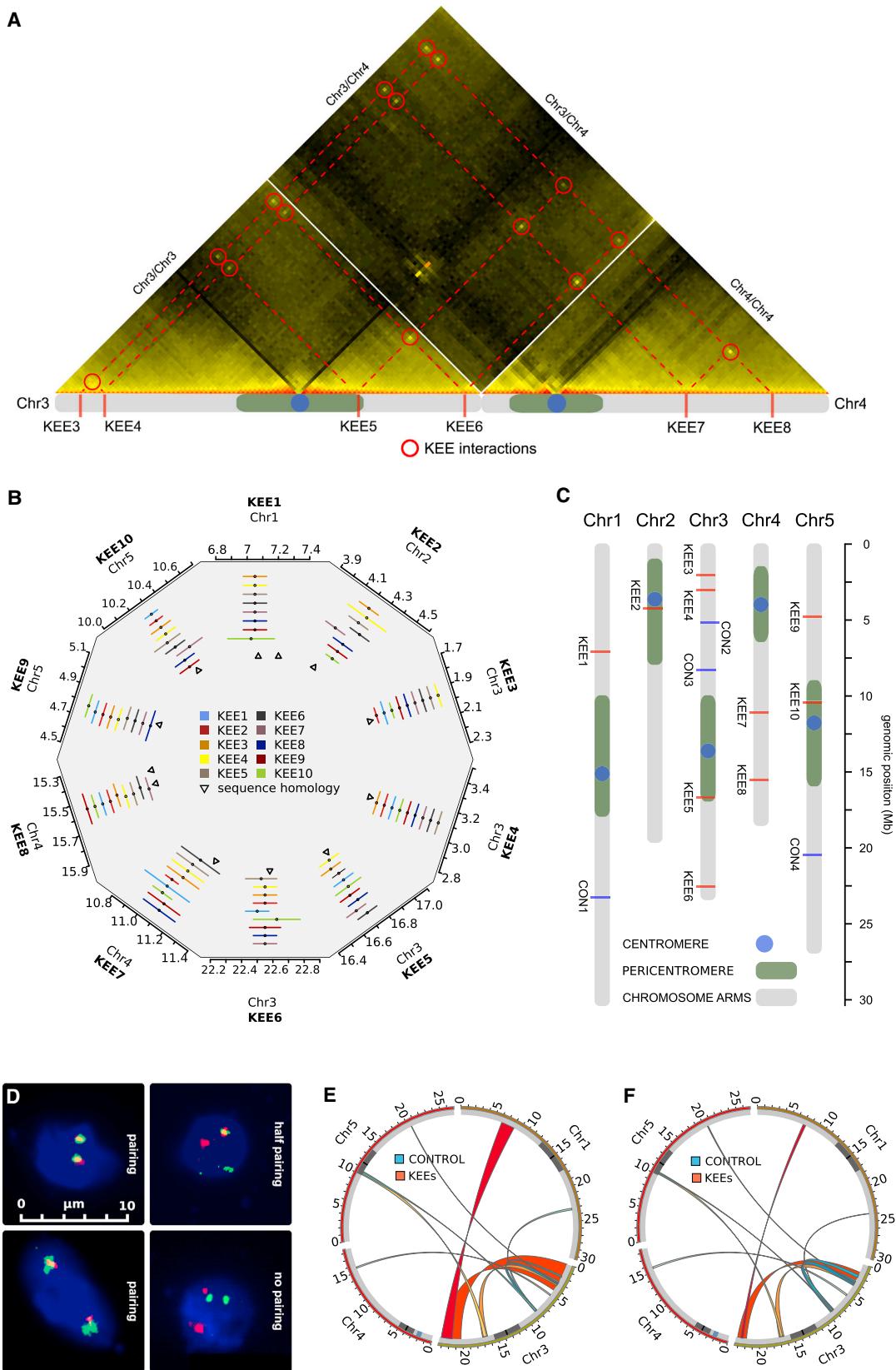
### ***Arabidopsis* Chromosomes Show a Simple Organization with Respect to Their Epigenetic Landscape and Interactome**

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.

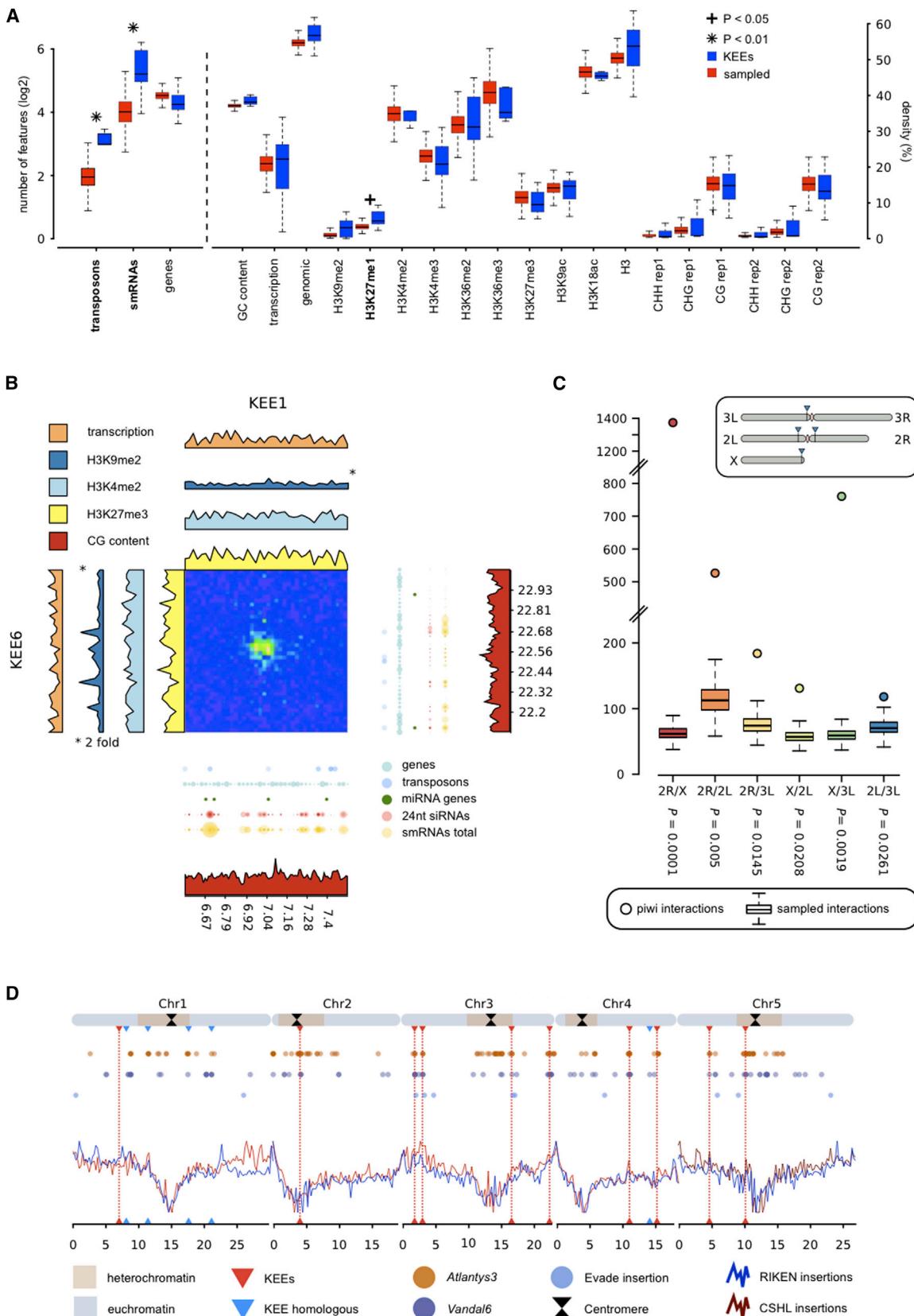
### IDEs Indicate a Distinct Chromatin Organization of CAs and PRs

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.

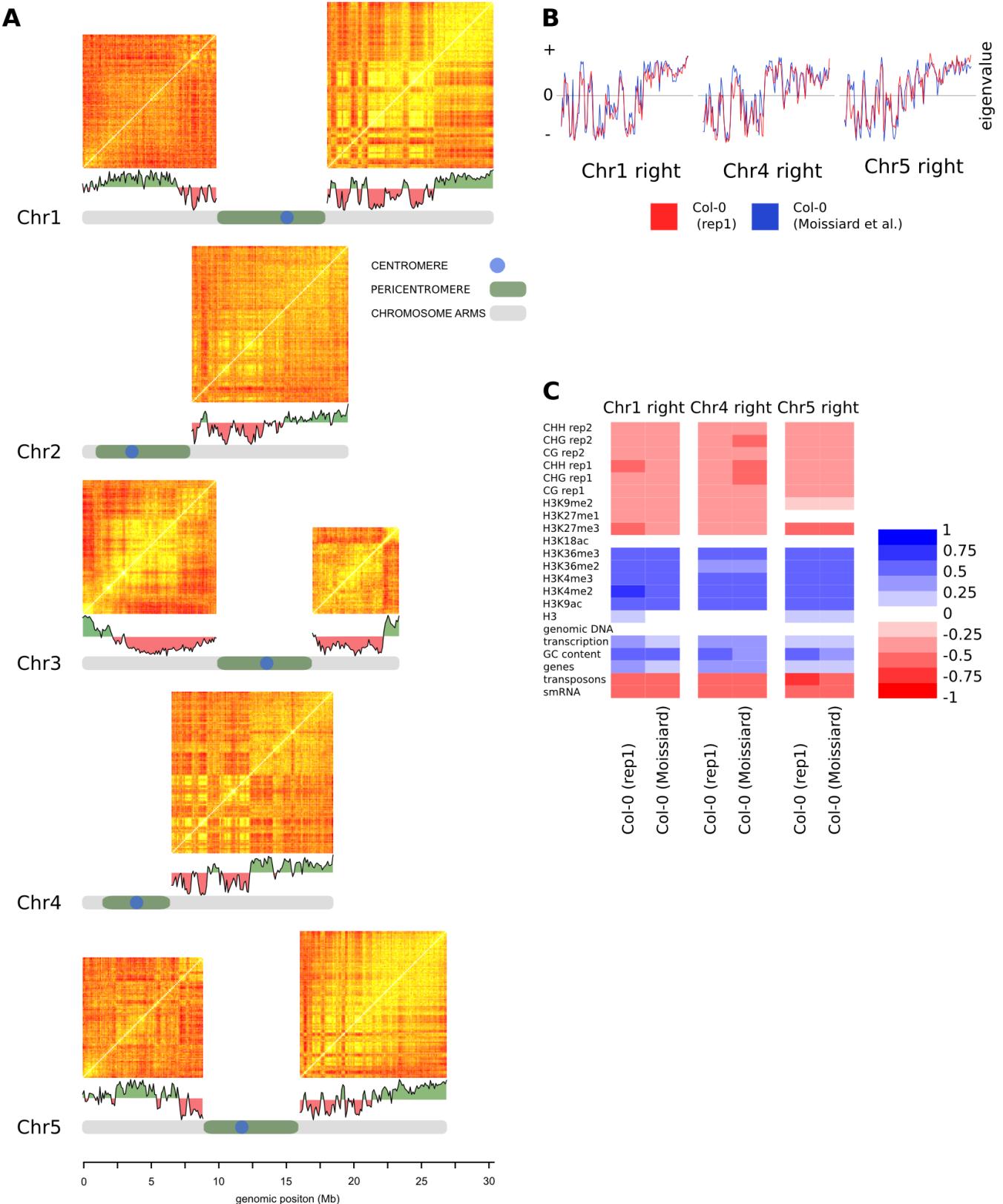
- Grob, S., Schmid, M.W., Luedtke, N.W., Wicker, T., and Grossniklaus, U. (2013). Characterization of chromosomal architecture in *Arabidopsis* by chromosome conformation capture. *Genome Biol.* 14, R129.
- Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664.
- Kleinboelting, 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.

**Molecular Cell, Volume 55**

**Supplemental Information**

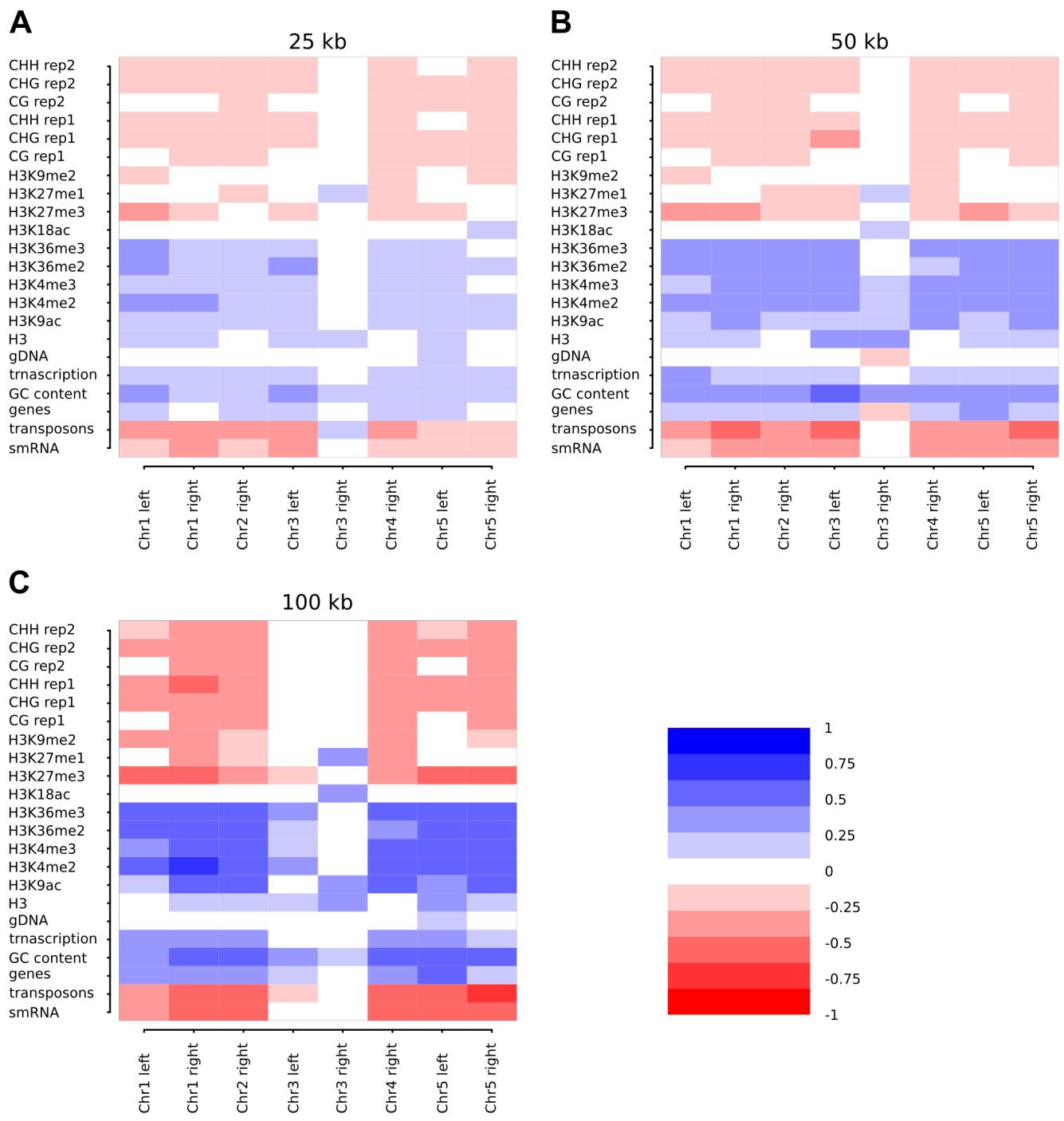
**Hi-C Analysis in *Arabidopsis* Identifies  
the *KNOT*, a Structure with Similarities  
to the *flamenco* Locus of *Drosophila***

Stefan Grob, Marc W. Schmid, and Ueli Grossniklaus



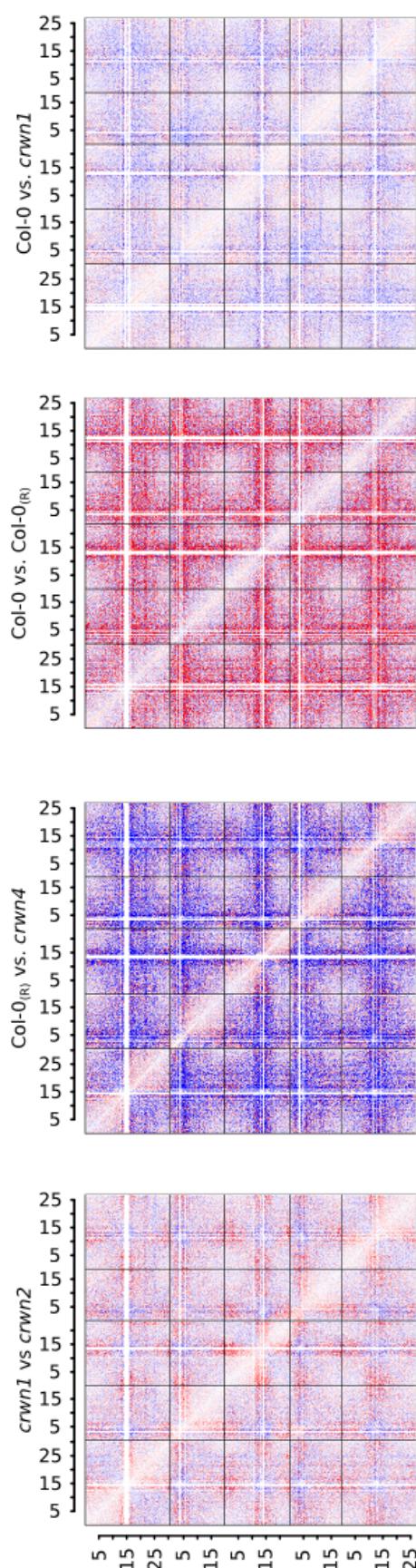
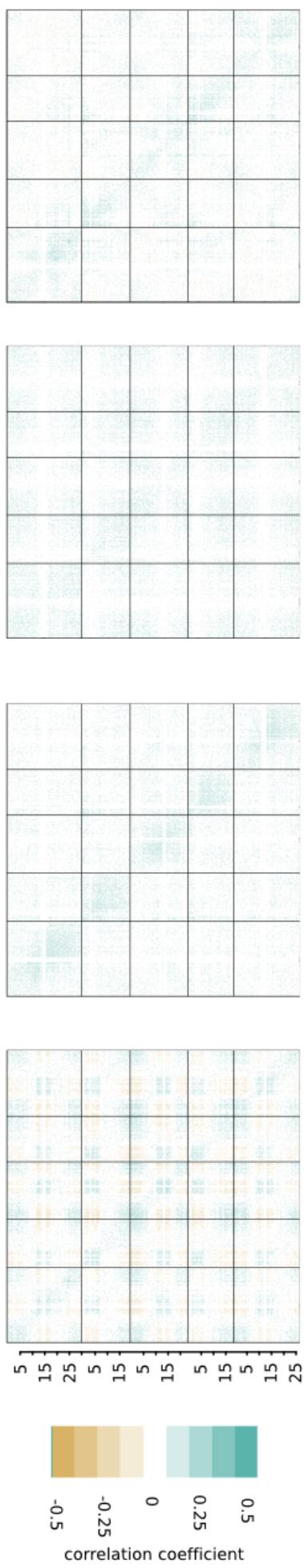
**Figure S1 Correlated HiC maps of 8 CAs and comparisons to previously published HiC data, related to Figure 1.**

(A) Shown are correlated HiC interaction matrices of single chromosome arms with their respective Eigenvectors. Red: negative Eigenvalues; Green positive Eigenvalues. Genomic bin size: 100 kb. (B) Comparisons between the Eigenvectors of the three chromosome arms shown in Figure 1D and the Eigenvectors calculated based on HiC interaction data by Moissiard et al. (2012). Red: Col-0 sample (this study), blue: Col-0 from Moissiard et al. (2012). The Eigenvectors highly correlate (Pearson's  $r = 0.94$ ,  $P = 2.2 \times 10^{-16}$ ) (C) Pearson's correlation analysis for regions with positive Eigenvalues (LSDs). The same epigenetic and genetic features correlate in both samples (Col-0 from our study and Col-0 from Moissiard et al. [2012]). Only features for which a  $P$ -value  $< 0.05$  was determined are colored. See also Figure 1.

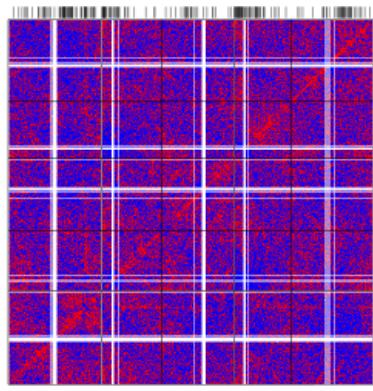


**Figure S2. Correlation coefficients for single chromosome arms, related to Figure 2.**

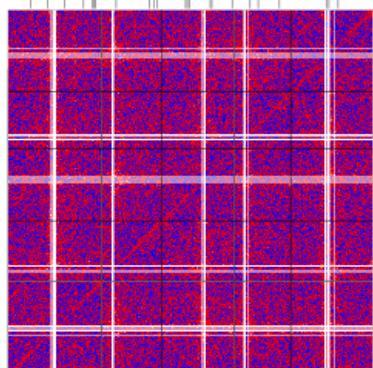
Color-coded correlation indices with a corresponding *P*-value of < 0.05, indicative for the correlation of an epigenetic or genomic feature and the Eigenvector of single chromosome arm. (A) genomic bin size: 25 kb; (B) genomic bin size: 50 kb; (C) genomic bin size: 100 kb. See also Figure 2.

**A****B****C**

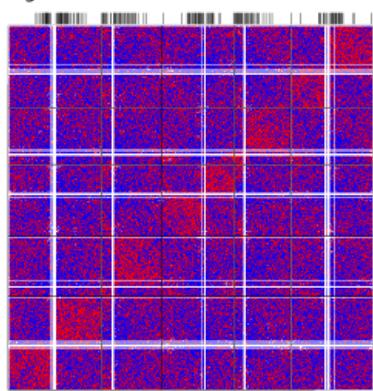
significant: 19%       $P = 2.2 \times 10^{-18}$



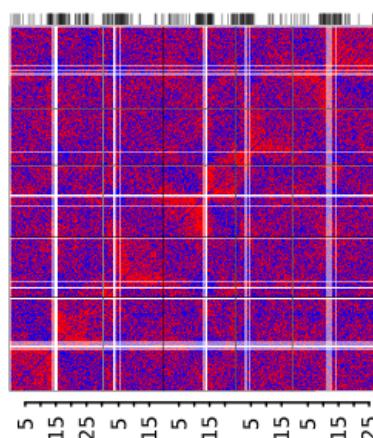
significant: 2%       $P = 0.76$



significant: 17%       $P = 2.3 \times 10^{-10}$



significant: 26%       $P = 6.0 \times 10^{-26}$



enrichment

sample1      enrichment      sample2

correlation coefficient

-0.5      0      0.25      0.5

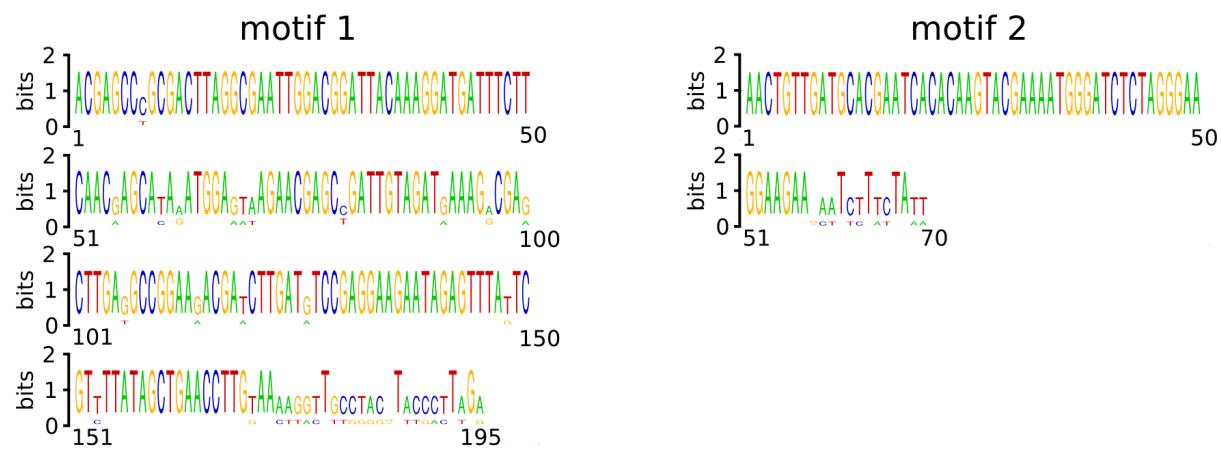
higher in:

	sample1
	sample2

**Figure S3. Additional comparison between two wild types to *crwn1* and *crwn4* mutants, related to Figure 3.**

- (A) Enrichment of interaction frequencies, obtained by calculating the relative difference between two interactomes. Col-0<sub>(R)</sub> represents an additional Col-0 WT replicate HiC interaction data set.
- (B) Pearson's correlation coefficients of differences between two interactomes.
- (C) Visualization of the SDM between two interactomes.

See also Figure 3

**A****B**

KEE ID	chrom	position <sup>1</sup>	position <sup>2</sup>	motif1	motif2	retrotransposons	transposons	identified by
KEE1	Chr1	7050000	7071324	no	no	ATLANTYS2(2)	TNAT1A (1)	BLAT*
KEE2	Chr2	4175000	4136555	yes	no	ATHILAO_I (7)		BLAST
KEE3	Chr3	1950000	1971581	yes	yes	ATLANTYS3(3)	VANDAL6 (5)	MEME
KEE4	Chr3	3100000	3121455	yes	yes	ATLANTYS3(3)	VANDAL6 (3)	MEME
KEE5	Chr3	16687500	16717396	yes	no	ATLANTYS3(3)	VANDAL6 (1)	BLAST
KEE6	Chr3	22525000	22580488	yes	yes	ATLANTYS3(2)	VANDAL6 (3)	MEME
KEE7	Chr4	11050000	11206537	yes	yes	ATLANTYS3(7)	VANDAL6 (2)	MEME
KEE8	Chr4	15537500	15441465	no	no	ATLANTYS3(5)	ATREP2 (1)	BLAT*
KEE9	Chr5	4762500	4800379	yes	yes	ATLANTYS3(3)	VANDAL6 (1)	MEME
KEE10	Chr5	10312500	10331725	no	no	ATLANTYS3(6)	VANDAL6 (3)	BLAT

**Figure S4: Occurrence of sequence motifs and (retro-) transposons in KEE regions, related to Figure 5.**

(A) Logo representation of motif1 and motif2 shared by most *KEEs*.

(B) Table summarizing location of *KEEs* and associated TEs. Positions of *KEEs* are either defined (1) by the center of the maximal overlap of interaction frequencies (center of bars in Figure 5B) or (2) by sequence homology among *KEEs* (see triangles in Figure 5B). Asterisk marks regions in which a motif or (retro-) transposon was found twice. Number in brackets indicates the number of TE copies found within  $\pm 150$  kb of the *KEE* position (defined by the overlap of *KEE* interactions as shown in Figure 5B).

See also Figure 5.

**Table S1: Coordinates of chromosome arms used for the analysis, related to Figure 1.**

See also Figure 1

<b>chromosome arm</b>	<b>start</b>	<b>end</b>
Chr1, left	0	10,000,000
Chr1, right	18,000,000	30,427,671
Chr2, left	0	1,000,000
Chr2, right	8,000,000	19,698,289
Chr3, left	0	10,000,000
Chr3, right	17,000,000	23,459,830
Chr4, left	0	1,500,000
Chr4, right	6,500,000	18,585,056
Chr5, left	0	9,000,000
Chr5, right	16,000,000	26,975,502

**Table S2: Pearson correlation coefficients between the principal component and epigenetic/genomic features and enrichment of epigenetic/genomic features in open chromatin compared to closed chromatin, related to Figure 2.**

Correlations on top, enrichments below. One / two / three asterisks mark correlations / enrichments with a *P*-value below 0.01/0.001/0.0001. Enrichments are log<sub>2</sub> transformed (positive value corresponds to enrichment in open chromatin). See also Figure 2.

Feature	Chr1 right arm	Chr4 right arm	Chr5 right arm	average
smRNA	***-0.67 ***-1.08	***-0.62 ***-1.14	***-0.66 ***-1.04	-0.65 -1.09
transposon	***-0.68 ***-1.06	***-0.66 ***-1.38	***-0.76 ***-1.29	-0.70 -1.24
genes	**0.34 **0.14	***0.38 **0.20	*0.30 0.12	0.34 0.15
GC content	***0.59 ***0.05	***0.52 ***0.05	***0.51 ***0.04	0.54 0.04
transcription	**0.33 ***0.61	***0.36 *0.62	*0.30 *0.53	0.33 0.59
genomic DNA	-0.01 -0.01	0.03 0.01	-0.09 0.00	-0.03 0.00
H3	*0.29 0.05	0.09 0.02	0.23 0.06	0.20 0.04
H3K9Ac	***0.55 ***0.40	***0.54 ***0.47	***0.56 ***0.35	0.55 0.40
H3K4me2	***0.75 ***0.37	***0.65 ***0.33	***0.64 ***0.32	0.68 0.34
H3K4me3	***0.63 ***0.37	***0.64 ***0.41	***0.54 ***0.30	0.61 0.36
H3K36me2	***0.58 ***0.33	***0.50 **0.25	***0.51 ***0.33	0.53 0.30
H3K36me3	***0.67	***0.62	***0.54	0.61

	***0.42	***0.40	***0.33	0.38
H3K18Ac	-0.07	-0.04	0.02	-0.03
	-0.04	-0.01	-0.02	-0.02
H3K27me3	***-0.53	***-0.49	***-0.53	-0.51
	***-0.53	**-0.49	***-0.52	-0.51
H3K27me1	***-0.35	**-0.32	-0.16	-0.28
	*-0.45	-0.42	-0.27	-0.38
H3K9me2	**-0.32	***-0.37	*-0.28	-0.33
	*-1.09	**-1.03	*-1.02	-1.05
GC methylation 1	***-0.35	***-0.40	***-0.43	-0.39
	**-0.35	**-0.41	**-0.34	-0.37
CHG methylation 1	***-0.50	***-0.49	***-0.44	-0.48
	***-1.10	***-1.11	***-0.99	-1.07
CHH methylation 1	***-0.50	***-0.47	***-0.41	-0.46
	***-0.87	***-0.88	***-0.85	-0.87
CG methylation 2	**-0.32	***-0.39	***-0.41	-0.37
	*-0.32	**-0.41	**-0.33	-0.35
CHG methylation 2	***-0.49	***-0.47	***-0.45	-0.47
	***-1.20	***-1.12	***-1.07	-1.13
CHH methylation 2	***-0.47	***-0.46	***-0.42	-0.45
	***-0.76	***-0.80	***-0.78	-0.78

**Table S3: BACs used for fluorescence *in situ* hybridization (FISH), related to Table 1.**

See also Table 1

	<b>chromosome</b>	<b>start</b>	<b>end</b>	<b>alias</b>
F15H21	Chr1	23079000	23343000	CON1
K7L4	Chr3	5120490	5185856	CON2
K14B15	Chr3	8241240	8324928	CON3
K6A12	Chr5	20405280	20479415	CON4
F5M15	Chr1	7065426	7164656	KEE1
F24P17	Chr3	1906274	1992295	KEE3
T22K18	Chr3	3047305	3143536	KEE4
F9K21	Chr3	16657512	16768491	KEE5
T27I15	Chr3	22502205	22614788	KEE6
F10M6	Chr4	15532305	15625657	KEE8
F21B23	Chr5	10317873	10389156	KEE10

**Table S4: Enrichment of epigenetic/genomic features and T-DNA/transposon insertions in KEE regions of variable size compared to random regions, related to Figure 6.**

Enrichments of epigenetic/genomic features were calculated using only euchromatic *KEEs*. Asterisks mark enrichments with *P*-values below 0.05 (one-sided). Enrichments are  $\log_2$  transformed (positive value corresponds to enrichment in *KEE* regions).

See also Figure 6.

feature	20 kb	50 kb	100 kb	150 kb	200 kb	300 kb
smRNA	*0.82	*1.82	*1.35	*0.99	*0.73	*0.47
transposon	-0.04	*1.03	*0.78	0.47	0.24	0.14
genes	-0.10	-0.16	-0.02	-0.08	-0.07	-0.06
GC content	*0.06	*0.05	*0.04	*0.04	0.02	0.01
transcription	-0.60	-0.16	0.00	*0.87	*0.71	*0.56
genomic DNA	-0.06	-0.02	-0.04	-0.05	-0.04	-0.01
H3	*0.25	0.07	*0.08	*0.08	0.04	0.02
H3K9Ac	-0.20	-0.08	-0.02	0.00	-0.04	0.01
H3K4me2	0.00	0.00	0.05	0.08	0.07	0.06
H3K4me3	0.04	-0.13	-0.05	0.00	-0.01	-0.01
H3K36me2	0.19	0.03	0.16	0.11	0.10	0.07
H3K36me3	-0.25	-0.18	0.01	0.03	0.02	0.01
H3K18Ac	-0.02	-0.05	-0.01	-0.01	0.00	0.00
H3K27me3	-0.84	-0.20	-0.18	-0.14	-0.11	-0.08
H3K27me1	*1.14	*0.91	*0.63	*0.66	*0.49	*0.38
H3K9me2	1.43	*1.43	0.89	0.66	0.32	0.48
GC methylation 1	-0.18	-0.09	0.02	0.02	-0.08	-0.13
CHG methylation 1	-1.35	0.53	0.07	0.23	-0.07	-0.07
CHH methylation 1	-1.16	0.59	0.09	0.22	-0.02	0.01
CG methylation 2	-0.21	-0.12	0.00	0.02	-0.07	-0.13
CHG methylation 2	-1.44	0.60	0.02	0.22	-0.07	-0.04
CHH methylation 2	-1.01	0.55	0.06	0.21	-0.03	0.01
CSHL	*0.94	1.58	1.40	1.38	1.30	1.09
FLAG	0.22	0.13	0.11	0.14	0.15	0.10
GABI	-0.33	-0.05	-0.03	0.00	-0.02	-0.02
RIKEN	*0.94	*1.20	*0.90	*0.83	*0.79	*0.64
SAIL	-0.13	-0.06	-0.05	0.04	0.16	*0.37
SALK	-0.26	-1.17	0.14	0.12	0.16	*0.44
WISC	*0.91	0.27	*0.43	*0.47	*0.45	0.37

## Supplemental Experimental Procedures

### Plant Material

The plant material for this study comprised several accessions from *Arabidopsis thaliana* (L) Heynh: Columbia-0 (Col-0) wild type and the two homozygous crowded nuclei mutants *crwn1-1* and *crwn4-1* (both donations from Eric Richards; Dittmer et al., 2007). Seedlings of all genotypes were grown on MS (4.3 g/l Murashige and Skoog salt (Carolina Biological Supply Company, Burlington, North Carolina, USA), 10 g/l sucrose (Applichem GmbH, Darmstadt, Germany), 7 g/l PHYTAGAR (Life Technologies Europe, Zug, Switzerland), pH5.6) culture plates under long-day condition (16 h and 22°C day / 8 h and 18° C night) for 14 days. For each HiC experiment, approximately 10 g of aerial tissue was collected and distributed to four conical 50 ml tubes.

### Fluorescence *in situ* Hybridization (FISH)

For the labeling of specific genomic regions, bacterial artificial chromosomes (BACs) were retrieved from the ABRC stock centre. After DNA extraction employing standard alkaline lysis protocol, the identities of BACs were confirmed by PCR (for detailed information of BACs used in this study, Table S4).

For each FISH experiment, a set of two BACs was labeled with either digoxigenin (DIG) or biotin, allowing for performing dual color FISH. For this, 500 - 1000 ng of BAC DNA was either labeled with DIG-nick translation mix or Biotin-nick translation mix (both Roche). The reactions were incubated for 2 h at 15°C and subsequently stopped by the addition of 1 µl of 0.5 M EDTA and heating up to 65°C for 10 min. The labeled BAC DNA was then purified using the QIAquick nucleotide removal kit (Qiagen, Hilden, Netherlands), followed by ethanol precipitation. The labeled BAC DNA was air-dried and resuspended in 10 µl of HB50 (50 % formamide, 50 mM sodium phosphate buffer pH7, 2x SSC (20x SSC: 3 M NaCl, 300 mM trisodium citrate, pH7)). After 15 min incubation at 42°C, 10µl of 20 %

dextran sulfate in HB50 was added and the DNA was denatured for 15 min at 75°C and stored on ice until the hybridization.

Young *Arabidopsis* rosette leaves were fixed in 4 % formaldehyde in TRIS buffer (10 mM TRIS-HCl, 10 mM EDTA, 100 mM NaCl, 0.1 % Triton X-100, pH 7.5) for 20 minutes under vacuum at RT. The rosette leaves were then washed 3 times in TRIS buffer and then homogenized in FISH nuclei isolation buffer (15 mM TRIS-HCl, 2 mM EDTA, 0.5 mM spermidin, 80 mM KCl, 20 mM NaCl, 15 mM 2-mercaptoethanol, 0.1 % Triton X-100). To remove residual cellular debris, the nuclei were filtered through a 30  $\mu$ m mesh. Subsequently, nuclei were flow sorted on a FACSAria Illu BL1 (BD Biosciences, San Jose, CA, USA) flow-sorting platform. Nuclei of a 2n:2c genomic content were subsequently placed within a drop of sucrose solution (100 mM TRIS-HCl, 50 mM KCl, 2 mM MgCl<sub>2</sub>, 0.05% Tween-20, 5% sucrose) on a microscopy glass slide. The air-dried microscopy slides holding the nuclei were then stored at -20°C.

After washing the slides twice in 2x SSC for 5 min, nuclei were fixed for 5 min in 1 % formaldehyde in PBS and subsequently rinsed in PBS for 5 min. To permeabilize the nuclear membrane, nuclei were incubated in pepsin (Sigma-Aldrich, Buchs, Switzerland) for 80 sec at 38°C. The nuclei were once more fixed in 1 % formaldehyde in PBS for 10 min, washed twice in PBS for 5 min each, and subsequently dehydrated in an ethanol gradient, stepping from 70 %, to 90 %, to 100 % ethanol. To prevent binding of labeled BAC DNA to endogenous RNAs, nuclei were treated with 100  $\mu$ g/ml in 2 x SSC RNase A (Roche) for 30 min at 37°C. To finally prepare the nuclei for hybridization, slides were washed 3 times in 2 x SSC for 5 min each and subsequently washed in PBS and dehydrated in ethanol. For hybridization, 20  $\mu$ l of labeled BAC DNA was applied to the slide: To denature the chromosomal and the labeled BAC DNA, the slides were placed on a heating block at 80°C for 2 minutes. Then, the nuclei were incubated in a moisture chamber for 18 h at 37°C.

After hybridization, the slides were rinsed at 42 °C 3 times in SF50 (50% formamide in 2X SSC, pH 7.0), twice in 2 x SSC, and once in 4T (4 x SSC, 0.05 % Tween-20) for 5 min each. Subsequently, the nuclei were incubated for 30 min at

37°C in 100  $\mu$ l of blocking solution (MB-1220; Vector Labs, Burlingame, CA, USA), which was directly applied on the nuclei. After washing twice for 5 min in 4T, the probe detection was performed.

For detection, 1000 x dilution in blocking solution of Texas Red Avidin DCS (A-2016; Vector Labs) was applied and the slides were incubated for 30 min at 37 °C and subsequently washed twice in twice in 4T and once in TNT (0.1 M TRIS, 0.15 M NaCl, 0.05% Tween-20) for 5 min each. Then, a 1:250 dilution in TNB of biotinylated Anti-Avidin D (BA-0300; Vector Labs) and mouse mouse anti-digoxigenin (Roche) was added and the nuclei were incubated for 30 min at 37°C. The slides were washed 3 times in TNT for 5 min each. Finally, Texas-Red (1:1000) and a 1:400 dilution in TNB (0.1 M TRIS, 0.15 M NaCl, 0.5 % blocking reagent (w/v; Boehringer-Ingelheim, Basel, Switzerland)) of goat anti-mouse conjugated with Alexa-488 (Life Technologies) was added and the slides were incubated for 30 min at 37°C. To remove excess of antibodies, the slides were washed in TNT 3 times (5 min each). Finally, the nuclei were dehydrated in an ethanol series and the DNA was stained with a small drop of Vectashield (H-1200; Vector Labs).

The FISH treated nuclei were analyzed using the epifluorescence microscope DM6000 (Leica, Wetzlar, Germany), equipped with a CCD camera (DFC350FXR2; Leica). The association rates were scored in two classes:

Pairing events (that is two dots completely overlap) were scored with value one, whereas close association (that is the two dots do not overlap, however are in very close proximity) was scored with 0.5. This yielded pairing rates within individual nuclei, ranging from 0 (all 4 signals can be detected separately), to 0.5 (one pair of signals are in proximity), to 1 (one pair full pairing or two pairs in close proximity), to 1.5 (one full pairing and one proximity event), and to 2 (two complete pairing events). To obtain final association rates, the sum of pairing rates was subsequently divided by 2 and subsequently divided by the total number of analyzed nuclei.

## HiC Sample Preparation

The HiC experiments for all genotypes were performed according to following protocol. The chromatin was cross-linked for 1 hour at room temperature (RT) in 15

ml freshly prepared nuclei isolation buffer (NIB: 20 mM Hepes (pH8), 250 mM sucrose, 1 mM MgCl<sub>2</sub>, 5 mM KCl, 40% (v/v) glycerol, 0.25% (v/v) Triton X-100, 0.1 mM phenylmethanesulfonylfluoride (PMSF), 0.1% (v/v) 2-mercaptoethanol) and 15 ml 4% formaldehyde solution. To quench the formaldehyde, 1.9 ml of 2 M glycine was subsequently added and incubated under vacuum for another 5 minutes at RT. Subsequently, frozen plant tissue from all four conical tubes was homogenized by grinding to a fine powder using mortar and pestle. Then, the homogenized plant material was equally distributed to two 50 ml conical tubes and resuspended in 10 ml NIB containing protease inhibitor (Complete Protease Inhibitor Tablets; Roche, Basel, Switzerland; two tablets in 150 ml NIB). We then filtered the suspension twice using Miracloth (Calbiochem/EMD Millipore, Darmstadt, Germany). For optimal recovery of nuclei, an additional 10 ml NIB was added to the left over material residing in the Miracloth. To collect the filtered nuclei, the filtrate was spun for 15 minutes at 4°C and 3000×g. The pellet was subsequently resuspended in 4 ml NIB and transferred to 2 fresh 1.5 ml reaction tubes. Then, the nuclei were washed 4 times in 1 ml NIB and recollected by 5 minutes centrifugation at 4°C and 1900×g (we used the same centrifugation conditions between each washing step). To remove traces of NIB for the subsequent restriction enzyme digestion, the nuclei were then washed twice with 1.2 × NEB buffer 4 (New England Biolabs, Ipswich, MA, USA) (10 × NEB buffer 4: 50 mM potassium acetate, 20 mM Tris acetate, 10 mM magnesium acetate, 1 mM dithiothreitol (DTT)) and finally resuspended in 500 ml 1.2 × NEB buffer 4. To permeabilize the nuclear membrane, the samples were incubated for 40 minutes at 65°C and 20 minutes at 37°C under constant shaking, with the addition of 5 µl of 20% SDS. Subsequently, to sequester the SDS, 50 µl of 20% Triton X-100 was added to the mixture followed by incubation for 1 hour at 37°C under constant shaking. For later analysis of digestion efficiency 60 µl of each tube was set aside as a pre-digestion control.

Subsequently, the extracted cross-linked chromatin was digested over night using a total of 400 U *Hind*III restriction enzyme (New England Biolabs), which were added in three steps. To facilitate digestion, the samples were diluted by adding 15

$\mu$ l 10x NEB buffer 4 and 115  $\mu$ l H<sub>2</sub>O. For later analysis of digestion efficiency 60  $\mu$ l of each tube was set aside as a post-digestion control.

For the later enrichment of HiC hybrid molecules, restriction fragment ends were labeled with biotinylated cytosine nucleotides as follows: 40  $\mu$ l of 0.4 mM biotin-14-dCTP (Life Technologies Europe, Zug, Switzerland), 1.6  $\mu$ l of each, 10 mM dATP, 10 mM dGTP, and 10 mM dTTP (Invitrogen/Life Technologies) and 60 U of Klenow polymerase (Large Klenow Fragment; New England Biolabs) were added to each but one tube and the mixture was incubated for 45 minutes at 37°C under constant shaking. The residing sample was set aside, for later analysis as a negative control for the efficiency of the fill-in reaction. The restriction and Klenow enzymes were inactivated by the addition of 20  $\mu$ l 20% SDS and 25 minutes incubation at 65°C under constant shaking.

To sequester the SDS, the samples were then incubated for 1 hour at 37°C under constant shaking in 745  $\mu$ l of 10x ligation buffer (0.5 M Tris-Cl, 0.1 M MgCl<sub>2</sub>, 0.1 M DTT, pH 7.5), 745  $\mu$ l of 10% Triton X-100, 80  $\mu$ l 10 mg/ml bovine serum albumin (BSA)), and 5.23 ml H<sub>2</sub>O. To obtain hybrid HiC fragments, blunt-ended restriction fragments were ligated for 5 h at 16°C with the addition of 80  $\mu$ l 100mM ATP (Roche, Basel, Switzerland) and 50 Weiss Units of T4 DNA ligase (Fermentas/Fisher Scientific, Wohlen, Switzerland). The non-filled-in negative control sample was ligated similarly, however as this sample was not treated with Klenow polymerase previously and therefore did not exhibit blunt fragment ends, less ligase was added (10 WU).

After ligation, the cross-linking was reversed by adding 50 ml of 10 mg/ml proteinase K (Qbiogene; MP Biomedicals, Santa Ana, CA, USA) and overnight incubation at 65°C under constant shaking. Next morning, an additional 50  $\mu$ l of proteinase K was added followed by 2h incubation at 65°C.

The DNA was extracted by adding 7 ml of Phenol and 7 ml of 24:1 Chloroform:Isoamylalcohol (v/v). A second purifying step was performed by addition of 7 ml of 24:1 Chloroform:Isoamylalcohol (v/v). Finally the hydrophilic phase was retained and mixed with 1.4 ml 3 M sodium acetate (NaOAc), 7 ml of H<sub>2</sub>O, and 30  $\mu$ l of glycogen. To precipitate the DNA, ice-cold 100% ethanol was added to a final

volume of 50 ml and the samples were then incubated at -80°C for 2 hours. After centrifugation, the DNA pellet was resuspended in 150  $\mu$ l of H<sub>2</sub>O with the addition of 1  $\mu$ l of 10mg/ml RNase A (Roche).

Subsequently, we analyzed both, the efficiency of the digestion and the fill-in reaction. For the digestion efficiency, we loaded 120 ng of DNA from each, pre-digestion control, post-digestion control, and the final HiC sample on a 1.5 agarose gel. The digestion efficiency was estimated by the appearance of a smear of DNA fragments with low molecular size.

The successful fill-in reaction, which was employed to label fragment ends with biotinylated cytosines created blunt-ended DNA fragments, whereas non-filled-in restriction fragment exhibited sticky ends. Upon ligation, two sticky ends theoretically produce the same restriction site (*Hind*III), which was initially used to digest the chromatin. Blunt-end ligation however, was expected to disrupt the *Hind*III restriction site (AAGCTT) and form a new *Nhe*I restriction site (GCTAGC). We amplified a specific genomic region in each sample and subsequently digested the PCR product with both, *Hind*III and *Nhe*I restriction enzymes. Samples, which exhibited low or no *Hind*III specific digestion products and high abundance of *Nhe*I digestion products were then classified as successfully labeled HiC templates. Samples, which exhibited both, satisfactory primary digestion and high efficiency end-labeling were pooled and subsequently used for the HiC sequencing library preparation.

The pooled HiC samples were then purified by adding 25:24:1 (v/v) phenol:chloroform:isoamylalcohol in equal volume to the pooled HiC sample. After an additional purifying step using 24:1 (v/v) chloroform:isoamylalcohol, the DNA was precipitated with 100 % ice-cold ethanol. To remove biotinylated cytosines from unligated fragment ends, the purified HiC samples were split into 2 and 1  $\mu$ l of 10 mg/ml BSA, 10  $\mu$ l of 10x NEB buffer 2 (New England Biolabs), 1  $\mu$ l 10mM dATP, 10 mM dGTP, 1.7  $\mu$ l T4 DNA polymerase (5.1 units; New England Biolabs), and 45.3  $\mu$ l H<sub>2</sub>O was added to 40  $\mu$ l of HiC sample each. The mixture was incubated for 2 h at 12°C and the reactions were stopped by the addition of 2  $\mu$ l of 0.5 M EDTA. Finally,

the HiC samples were purified once more with phenol:chloroform and subsequently precipitated with 100 % ethanol.

### HiC Library Preparation

The HiC samples were fragmented to a mean size of 300 bp by sonication using Covaris S2 sonication system (Covaris, Woburn, MA, USA) employing 5 cycles of 55 seconds, with intensity 5 and a cycle/burst ratio of 200. Subsequently, the fragment ends were repaired by the addition of 10  $\mu$ l resuspension buffer (RSB; Illumina, San Diego, USA) and 40  $\mu$ l End-Repair Mix (ERP) (Illumina) to 40  $\mu$ l of fragmented HiC sample. The mixture was then incubated for 30 minutes at 30°C. After standard purification using Agencourt AMPure beads (Beckman Coulter, Brea, CA, USA), biotin labeled HiC samples were specifically enriched with the use of Streptavidin C1 (Life Technologies) magnetic beads. For this, 60  $\mu$ l of Streptavidin beads were washed twice in 400  $\mu$ l Tween Wash Buffer (TWB; 5 mM Tris, 0.5 mM EDTA, 1M NaCl, 0.05% Tween-20). Between each washing step, the Streptavidin beads were recovered by placing the tubes on a magnetic stand. Subsequently, the beads were resuspended in 300  $\mu$ l of Binding Buffer (BB; 10 mM Tris, 1 mM EDTA, 2 M NaCl) and 300  $\mu$ l of the HiC sample was added. After 15 minutes incubation at RT under rotation, the supernatant was removed and the beads binding biotinylated HiC fragments were resuspended in 200  $\mu$ l of BB and 200  $\mu$ l of H<sub>2</sub>O. Then, the beads were washed once in 60  $\mu$ l of RSB and finally resuspended in 35  $\mu$ l of RSB. The fragment ends were then adenylated by adding 25  $\mu$ l of A-tailing Mix (ATL; Illumina) and 30 minutes incubation at 37°C. We then ligated 2.5  $\mu$ l of each Illumina paired-end sequencing adapter to the adenylated HiC fragment ends by addition of 5  $\mu$ l Ligation Mix (LIG; Illumina). The mixture was incubated 10 minutes at 30°C and the reaction was stopped by adding 10  $\mu$ l of Stop Ligation Mix (STL; Illumina). Finally, the bead-bound HiC samples were washed twice in 400  $\mu$ l TWB, once in 200  $\mu$ l BB, and once in 200  $\mu$ l RSB and were resuspended in 50  $\mu$ l RSB. Subsequently, the HiC libraries were amplified on bead by PCR (16 cycles) with adapter specific primers and the PCR products were purified applying the Agencourt AMPure beads standard protocol. The HiC libraries were then sequenced on a

Illumina Hi seq 2000 sequencing device (Illumina). Illumina sequencing of *HindIII* HiC fragments yielded in total 169,121,538 reads for Col-0, 190,929,344 reads for the Col-0 replicate, 219,474,805 reads for *crwn1-1*, and 233,011,638 reads for *crwn4-1* samples.

### HiC Sequencing Data Processing

To ensure high data recovery, it is important to consider the length of the sequenced ligation products, which was around 300 to 400 bp on average. Generally, a ligation product contains two parts of two distinct restriction fragments joined by a *HindIII* restriction site. To identify the interacting restriction fragments (and map them onto the genome), the ligation product is sequenced from both ends by paired-end sequencing. However, the *HindIII* restriction site separating the two restriction fragments can occur at any position within a ligation product. If the site is close to one of the ends, the corresponding read contains sequences from both restriction fragments, and therefore fails to align. Reads were thus trimmed to 30 bp and aligned to the Col-0 reference genome (TAIR10, Huala et al., 2001) using bowtie (version 0.12.7, Langmead et al. 2009) with the command line arguments –v 0 –m 1 –a (no mismatches and no multiple alignments). The aligned read-pairs can then be used to create an interaction matrix, in which each row (and column) corresponds to one fragment, and values represent the number of read-pairs aligned to the respective fragments. For further analysis, those matrices were binned into larger stretches along the genome (windows of size 10 kb, 25 kb, 50 kb, 100 kb, and 250 kb). These data were corrected for systematic biases using the approach from Jin et al., 2013, resulting in 16,291,506 (Col-0), 10,283,215 (Col-0 replicate), 44,744,537 (*crwn1-1*), and 36,508,909 (*crwn1-4*) read pairs in the final data sets. Matrices were then normalized as described in Zhang et al., 2012. We observed that some regions were highly variable between all samples. These regions were characterized by a high number of zeros and few non-zero interaction counts. It is thus likely that these differences did not reflect biologically significant differences. We therefore removed/ignored those bins with very low number of

interactions (i.e. the five percent of all rows/columns in the matrix with the highest number of zeros).

### Data on Epigenetic and Genomic Features

To add additional information, we used publicly available histone modification (Luo et al., 2012), cytosine methylation (Stroud et al., 2013), and transcriptional data (Luo et al., 2012). Data was obtained and processed as described previously in Grob et al., 2013. To control for sequencing biases, we used genomic DNA sequencing data (Jacob et al., 2010) and processed it as described for the transcriptional data (Grob et al., 2013). Regions associated with small RNAs were identified using data from (Kasschau et al., 2007; Gregory et al., 2008; Lister et al., 2008). siRNAs closer than 10 bp to each other were merged into a single target region (termed smRNA associated regions). Genomic features (i.e. genes and transposable elements (Huala et al., 2001), siRNA-associated regions, and T-DNA/transposon insertions (T-DNA: SALK, GABI, FLAG, WISC; transposon: CHSL, RIKEN, signal.salk.edu) were then mapped to the restriction fragments. If a feature did not span the entire restriction fragment, it was counted only half. For further analysis, values from individual restriction fragments were summarized across genomic regions with the size of choice (sum for count features, and average for density features, respectively). For comparison and statistical tests, the count data was log<sub>2</sub>-transformed.

### Calculation of the Interaction Frequency Decay Exponent

To estimate the genomic distance-dependent decay of the interaction probability (Interaction Decay Exponents, IDE), we used the 100 kb interaction matrices. For chromosome specific IDEs, the average interaction frequency between genomic bins sharing the same distance to each other was calculated for distances ranging from 100 kb to 10 Mb. Distance and average interaction frequencies were both log<sub>10</sub> transformed to fit a linear model. The resulting slope of the model corresponds to the IDE value. Calculation of IDEs of individual pericentromeres and chromosome arms was performed accordingly, however, due to their limited size, IDEs were calculated using a distance range of 100 kb to 5 Mb. The short arms of

chromosome 2 and 4 with euchromatic regions shorter than 2 Mb were entirely excluded.

### Determination of Chromosomal Neighborhoods

The enrichment of *trans*-interactions between a pair of chromosomes was calculated as described in Zhang et al., 2012. In short, the values are given as the  $\log_2$  ratio of the observed to the expected value.

### Identification of Structural Domains (SDs) in Chromatin

To identify interacting chromatin domains, we followed a strategy previously described by Lieberman and colleagues (Lieberman-Aiden et al., 2009). In brief, the approach relies on *intra*-chromosomal interactions and identifies chromatin domains in three steps: (i) distance-normalization, (ii) calculation of pair-wise Pearson correlation coefficients, and (iii) principal component analysis (PCA) on the correlation matrix. The first principal component can then be used to visualize differential behavior of genomic regions. Considering that the sign of the Eigenvector is arbitrary, it was set as such that positive Eigenvalues were associated with “loose structural domains” (LSD), i.e. a higher number of *trans*-interactions compared to “compacted structural domains” (CSDs). Using the whole chromosome, the first principal component generally separated the euchromatic chromosome arms from the heterochromatic pericentromeres. To identify distinct chromatin domains within the euchromatic arms, we therefore excluded the pericentromeres from the analysis and performed the PCA separately for each arm. Coordinates of chromosome arms were determined by visual inspection of transposon and gene density along the chromosomes (Table S1). The analysis was done on genomic bins with a size of 100 kb.

### Epigenetic Landscape and Chromatin Domains

To test whether epigenetic and genomic features are associated with either open or closed chromatin (i.e. the first Eigenvector of the PCA), we used two approaches. The first relied on a test for significance of correlation between the first Eigenvector of the PCA of a chromosome arm and the density/count of a given feature along

that chromosome arm using the built-in R ([www.r-project.org](http://www.r-project.org)) function `cor.test()`). In the second approach, genomic bins of a chromosome arm were split into two groups, according to the sign of the Eigenvalue of the first component of the PCA. We then performed two-sided Wilcoxon signed rank testing to determine whether a feature's density significantly differed between open and closed chromatin. Enrichment of a given feature was calculated as the ratio of the average density/count in the open chromatin over the closed chromatin.

### Identification of *KEE* Locations

To estimate the genomic location of *KEE* regions, we visualized each interaction-pair of the *KNOT* separately in high resolution using 10 kb genomic bins for *intra*-chromosomal interactions and 50 kb genomic bins for *inter*-chromosomal interactions. This resulted in estimated genomic regions of 150 kb to 300 kb. Subsequently, for each *KEE* region, all *KEE* partners were aligned against each other and the minimal overlap of all *KEE* partners was obtained by calculating mean of the maximal starting position and the minimal end position of all aligned *KEEs*. To allow for inaccuracy of the initial estimation, 150 kb to each side of the previously determined “core” *KEE* position was added.

To analyze sequence homology between *KEEs*, each *KEE* region (300 kb) was split into 500 bp fragments and aligned to all other *KEE* regions using BLAT with minimal identity threshold of 80% (Kent, 2002). Alignments longer than 60 bp were then used to calculate the number of *KEE* positions matching to a certain sequence within a given *KEE* region. For each *KEE*, the region with the highest coverage was then extracted for a more refined motif search using MEME (Bailey and Elkan, 1994). Motif search was limited to five motifs with a size between 50 to 300 bp. Nucleotide logos were generated using the publicly available Weblogo platform ([weblogo.berkeley.edu/logo.cgi](http://weblogo.berkeley.edu/logo.cgi)). To search for additional regions (termed “*KEE* homologous”), sharing sequence homology with the obtained sequence motifs, motif 1 and motif 2 were blasted against all available genomes using megablast ([blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)). Subsequently, regions with the highest scores were retrieved.

## Random Sampling Strategy for Analysis of *KEE* and *KEE* Homologous Regions

*KEE* and *KEE* homologous regions were identified as “point-positions” within the genome (i.e. residing at position X on chromosome Y) and not as genomic bins within the interaction matrix. To obtain empirical distributions of a certain characteristic C, we therefore sampled a set of “point-positions” within the genome (10'000 repetitions). Within each set, the randomly chosen regions reflected the numbers, as well as the locations, of the *KEE* (or *KEE* homologous) regions (i.e. each *KEE* (or *KEE* homologous) region was represented by a region randomly chosen from its own chromosome arm). Measures of C were then summarized within a window of a certain size (20 kb, 50 kb, 100 kb, 150 kb, 200 kb, 300 kb) centered at the *KEE*, *KEE* homologous, or sampled region (sum for count, and average for density data, respectively). For comparison and statistical tests, the count data was log<sub>2</sub>-transformed.

## Enrichment of Interaction Frequencies between *KEE* and *KEE* Homologous Regions

To test whether *KEE* (or *KEE* homologous) regions interact preferentially with each other, we compared the sum of interaction frequencies (SIF) between these regions to an empirical distribution of SIFs between sets of randomly chosen regions described above. Considering that *KEE* homologous regions were identified using the conserved sequences found within the *KEE* regions, *KEE* positions were chosen according to those conserved sequences as well. Significance of enrichment was then calculated as the fraction of SIFs within the empirical distribution, which was higher than the SIF between the *KEE* (or *KEE* homologous) regions (empirical P-value, one-sided). *KEE* regions interacted significantly more frequent between each other than randomly chosen regions. However, *KEE* homologous regions did not. These results were consistently observed for all window sizes tested.

## Enrichment of Epigenetic or Genomic Features in *KEE* Regions

To test the enrichment of epigenetic or genomic features at *KEE* positions, the density/count measures within the *KEE* regions ( $M_{KEE}$ ) were compared to those of

the randomly chosen regions described before ( $M_{\text{random}}$ ). For each feature, the empirical  $P$ -value was calculated as the fraction of randomly chosen regions with a higher density/count measure. Enrichment was given as the average  $M_{\text{KEE}}$  over the average  $M_{\text{random}}$ . Features with an enrichment below 1.5 or a  $P$ -value above 0.05 were discarded. From all “natural” features, only smRNA-associated regions and H3K27me1 density were consistently enriched in euchromatic *KEE* regions for all window sizes tested. Regarding T-DNA and transposon insertions, only the transposon insertion lines (CSHL and RIKEN) showed consistent significant enrichment (all windows, other features not in any).

### Epigenetic Variance among *KEE* Regions

To test whether *KEE* regions vary less among each other than expected, variation of density/count measures among *KEE* regions was compared to those of randomly chosen regions. None of the features exhibited significantly reduced variance among *KEE* regions consistently across all window sizes tested. In individual tests, only transposons ( $P_{50kb} = 0.025$ ) and GC density ( $P_{100kb} = 0.016$ ,  $P_{150kb} = 0.039$ ,  $P_{200kb} = 0.012$ ) showed slightly reduced variance among *KEE* regions.

### Occurrence of Natural Transposon Insertions in *KEE* Regions

Likewise, *EVADÉ* insertion events were mapped by blasting genotyping primer sequences obtained from (Marí-Ordóñez et al., 2013). Genomic positions of *ATLANTYS3* and *VANDAL6* retrotransposons were obtained from [www.arabidopsis.org](http://www.arabidopsis.org) (TAIR10, Huala et al., 2001).

### Difference Between HiC Data Sets

Three different approaches to analyze differences between HiC data sets were chosen. The first analysis was conducted according to previously published protocol (Moissiard et al., 2012). In short, the difference between two given HiC matrices was calculated, by calculating the difference between pairs of elements in the two HiC matrices sharing the same coordinates. Subsequently, to normalize for the interaction intensity of these elements, each element of the resulting difference

matrix was divided by the mean interaction frequency of the pair of elements, for which the difference was calculated.

To reveal domains of differences, Pearson's correlation coefficients were calculated for the difference matrices, obtaining correlated difference matrices. Thereby, each element of the correlation matrix represented the correlation coefficient of a given column and row of the difference matrix.

To analyze whether differences between two HiC interaction data sets A and B were stochastic, the difference between the two HiC data sets (100 kb bin size) was calculated without normalizing for absolute interaction frequency. Subsequently, a signed difference matrix (SDM) was generated, which contained three classes of elements, + (enriched interaction frequency in A), - (enriched interaction frequency in B), and 0 (no change between matrix A and B). As all HiC matrices were initially normalized for coverage, which yielded single interaction frequencies with many decimal places, the occurrence of zero difference between the two HiC interaction data sets was extremely low and mainly limited to interactions, which were absent in both data sets. Thus, for further statistical analysis, elements in the SDM with value zero were removed.

To analyze, whether positive or negative signs in the SDMs occur in clusters, we performed Wald-Wolfowitz runs test on each column of the signed difference matrix using the R function `runs.pvalue()` (package “randomizeBE”). Columns, for which a *P*-value < 0.01 was obtained, were then used for subsequent analysis. Thereby, we analyzed whether columns exhibiting significant *P*-values cluster along the genome. We therefore assigned all columns, which showed significant *P*-values with value +1, all other columns were assigned with value -1. Subsequently, another Wald-Wolfowitz runs test was performed on these values.

To validate our analytical pipeline, we compared the Col-0 WT HiC interactome presented in this study, with a previously generated Col-0 WT HiC interactome (which was not included in this study, due to generally low sequencing read number). The resulting SDM of 100 kb genomic bin size exhibited only 2 % significant columns. Furthermore, no significant clustering of the significant few columns was observed, suggesting that apparent differences between the two HiC

interactomes are not biologically relevant.

### Interaction Frequencies of *Drosophila* piRNA Clusters

To test for the enrichment of interaction frequencies of *Drosophila* piRNA clusters, we obtained pre-processed HiC interaction data described in a study by Sexton and colleagues (Sexton et al., 2012) from Gene Expression Omnibus (GSE34453). Positional information on piRNA clusters was obtained from a study by Brennecke and colleagues (Brennecke et al., 2007). Only piRNA clusters, for which unambiguous positional information was available, were included in the later analysis. These piRNA clusters were located on chromosome arms 2L (20,148,259 - 20,227,581), 2R (2,144,349 – 2,386,719), 3L (23,273,964 – 23,314,199), and on chromosome X (21,392,175 – 21,431,907). We then calculated SIFs between piRNA clusters, including four genomic bins on each side of the genomic bin harboring the piRNA cluster (bin size: 80 kb). The SIFs of each individual pair of interacting piRNA clusters was then compared to SIFs of 10,000 times randomly sampled genomic regions (including 4 genomic bins on each side of the sampled genomic bin). Genomic bins were exclusively sampled on the chromosome arms, harboring the piRNA clusters, which were tested. By comparing the SIFs from sampled genomic bins to the SIFs of interacting piRNA clusters, an empirical *P*-value was obtained, describing the fraction of randomly selected pairs of genomic bins exhibiting higher interaction frequencies than the pair of interacting piRNA clusters.

To obtain an empirical *P*-value, describing whether piRNA clusters generally interact more frequent than randomly selected genomic regions, we then compared the average SIF between all piRNA clusters to 10,000 average SIFs of sampled genomic regions. Thereby, each average was calculated across 6 (the number of possible combinations of individual pairs of interacting piRNA clusters) SIFs of previously sampled pairs of genomic regions.

## Supplemental References

- 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.
- 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.
- 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.
- Gregory, B.D., OMalley, R.C., Lister, R., Urich, M.A., Tonti-Filippini, J., Chen, H., Millar, A.H., and Ecker, J.R. (2008). A link between RNA metabolism and silencing affecting *Arabidopsis* development. Developmental Cell 14, 854–866.
- 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.
- Huala, E., Dickerman, A.W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W., Mueller, L.A., Bhattacharyya, D., Bhaya, D., Sobral, B.W., Beavis, W., Meinke, D.W., Town, C.D., Somerville, C., and Rhee, S.Y. (2001). The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res. 29, 102–105.
- Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503, 290–294.
- Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-wide profiling and analysis of *Arabidopsis* siRNAs. PLoS Biol. 5, e57.
- Kent, W.J. (2002) BLAT--The BLAST-Like Alignment Tool. Genome Research 12, 656–664.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25.

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., Richard Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., and Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* *326*, 289–293.

Lister, R., OMalley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* *133*, 523–536.

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.

Marí-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.

Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., Christopher J. Hale, C.J., Feng, W., Michaels, S.D., Frand, A.R., Pellegrini, M., Dekker, J., Kim, J.K., and Jacobsen, S.E. (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* *336*, 1448–1451.

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

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.

Stroud, H., Greenberg, M.V.C., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* *152*, 352–364.

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.