

## MOLECULAR BIOLOGY

ALBA protein complex reads genic R-loops to maintain genome stability in *Arabidopsis*Wei Yuan<sup>1,2\*</sup>, Jincong Zhou<sup>3\*</sup>, Jinjin Tong<sup>1,2\*</sup>, Wanqing Zhuo<sup>1</sup>, Lishuan Wang<sup>4</sup>, Yan Li<sup>3</sup>, Qianwen Sun<sup>3†</sup>, Weiqiang Qian<sup>1†</sup>

The R-loop, composed of a DNA-RNA hybrid and the displaced single-stranded DNA, regulates diverse cellular processes. However, how cellular R-loops are recognized remains poorly understood. Here, we report the discovery of the evolutionally conserved ALBA proteins (AtALBA1 and AtALBA2) functioning as the genic R-loop readers in *Arabidopsis*. While AtALBA1 binds to the DNA-RNA hybrid, AtALBA2 associates with single-stranded DNA in the R-loops in vitro. In vivo, these two proteins interact and colocalize in the nucleus, where they preferentially bind to genic regions with active epigenetic marks in an R-loop-dependent manner. Depletion of AtALBA1 or AtALBA2 results in hypersensitivity of plants to DNA damaging agents. The formation of DNA breaks in *alba* mutants originates from unprotected R-loops. Our results reveal that the AtALBA1 and AtALBA2 protein complex is the genic R-loop reader crucial for genome stability in *Arabidopsis*.

## INTRODUCTION

The R-loop is a naturally occurring chromatin structure composed of a DNA-RNA hybrid and a displaced single-stranded DNA (ssDNA). R-loops are prevalent in bacteria, yeast, animals, and plants and play crucial roles in the regulation of gene expression, chromatin structure, and DNA repair (1–4). In yeast, R-loop formation stimulates replication defects at transcribed regions (5). In mammals, R-loop formation facilitates transcriptional termination (6, 7), heterochromatin retention of histone lysine methyltransferases (8), and mitotic chromosome segregation (9). In plants, R-loops regulate gene expression and plant development (10–12).

However, R-loops pose a threat to genomic stability because the displaced ssDNA is susceptible to nucleotide changes and strand breakage (13, 14). R-loops are also structural barriers that impair DNA replication (15) and, in transcription-replication conflicts, can induce DNA damage and genome instability (16–18). Ribonuclease (RNase) H enzymes and RNA-DNA helicases dissolve R-loops and prevent DNA damage and genomic instability caused by persistent R-loop formation (10, 19–21). Replication protein A (RPA), an ssDNA binding protein, functions as a sensor of R-loops to recruit RNase H1 for removing R-loops and suppressing genomic instability in human cell lines (19). A number of proteins, such as Npl3 (22) in yeast and the THO-TREX complex (23) and BRCA2 (24) in human cells, prevent R-loop formation or stabilization and thereby protect genome stability. A recent genome-wide map of R-loop-induced DNA damage in yeast revealed that, even with R-loops, many genomic regions are not prone to DNA damage (25), suggesting that mechanisms, other than reducing R-loop levels, exist to protect DNA from being damaged.

Alba proteins are small, dimeric DNA/RNA binding proteins whose actions have been best characterized in archaea (26). Structural

and molecular investigations revealed that Alba dimers bind to DNA in a sequence-independent and cooperative manner (27, 28). At low protein-DNA ratio, Alba dimers interact with Alba dimers on the adjacent DNA duplex and bridge two DNA duplexes, while at high protein-DNA ratio, Alba dimers bind side by side to DNA duplexes and rigidify the DNA (29). The role of Alba proteins in shaping chromatin architecture resembles that of histones. Archaeal Alba proteins bind RNA with an affinity similar to the one detected for DNA (30) and may regulate RNA processing (31). Investigations in other organisms revealed that Alba proteins also regulate RNA stability (32) and protein translation by binding to RNA (33, 34). However, the functions of Alba proteins in plants and mammals are still unclear.

Here, we characterized the functions of two *Arabidopsis* ALBA proteins (AtALBA1 and AtALBA2). AtALBA1 and AtALBA2 have different nucleic acid-binding properties, but they colocalize and form heterodimers in the nucleus. On the basis of their activities, we found that, in vitro, they can bind R-loop structures. They preferentially bind to the genic regions with active epigenetic marks in an R-loop-dependent manner in vivo. Depletion of AtALBA1 or AtALBA2 results in hypersensitivity of plants to DNA damaging agents because R-loops targeted by AtALBA1 and AtALBA2 lost protection. Our findings suggest that AtALBA1 and AtALBA2 are R-loop readers that safeguard genome stability.

## RESULTS

## AtALBA1 and AtALBA2 bind different types of nucleic acids

According to phylogenetic analyses, the *Arabidopsis* genome encodes six Alba proteins belonging to two distinct subfamilies (31). Members of the Rpp20-like subfamily, including AtALBA1, AtALBA2, and AtALBA3, have only a conserved Alba domain, while the members of Mdp2-like subfamily, including AtALBA4, AtALBA5, and AtALBA6, have additional RGG (Arg-Gly-Gly) repeats, which are often found in proteins that regulate transcription, splicing, and translation (fig. S1).

To investigate the functions of ALBA proteins, we initially started to analyze the binding affinities of AtALBA1 and AtALBA2, two of the simplest proteins in the gene family, toward different forms of nucleic acids. For this purpose, we purified recombinant wild-type and K30E mutant forms of AtALBA1 and AtALBA2 (fig. S2A). K30

Copyright © 2019  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China. <sup>2</sup>Tsinghua-Peking Joint Center for Life Sciences and Center for Plant Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China. <sup>3</sup>Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China. <sup>4</sup>State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing 100193, China.

\*These authors contributed equally to this work.

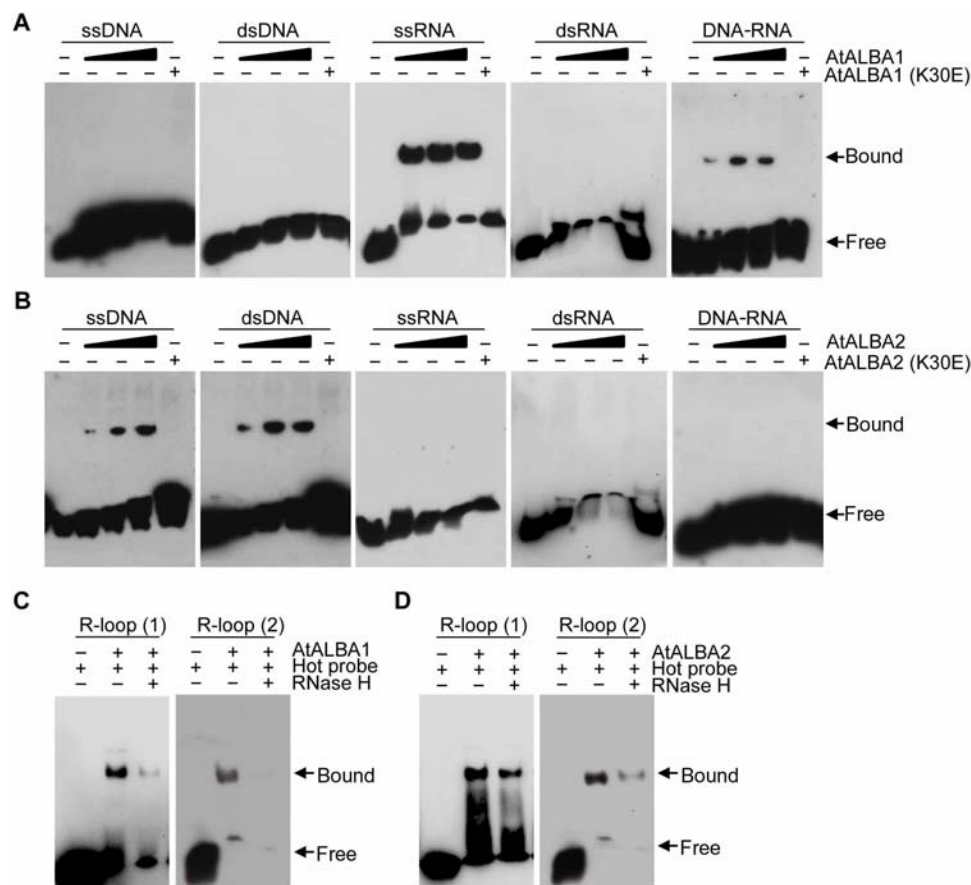
†Corresponding author. Email: sunqianwen@mail.tsinghua.edu.cn (Q.S.); wqian@pku.edu.cn (W.Q.)

corresponds in position to one of the critical DNA binding residues found in archaeal Alba proteins (K20 in ssoAlba1 and K11 in AfAlba2) and is conserved in AtALBA1, AtALBA2, and many Alba proteins in other species (fig. S2B) (35, 36). Purified AtNDX was also prepared as a positive control (12). We then performed electrophoretic mobility shift assays (EMSAs) using different substrates (fig. S2C). Our results revealed that wild-type AtALBA1-His bound to single-stranded RNA (ssRNA) and DNA-RNA hybrids (Fig. 1A). In contrast, wild-type AtALBA2-His bound to ssDNA and double-stranded DNA (dsDNA) (Fig. 1B). Consistent with previous results, AtNDX can bind ssDNA (fig. S3A). Because AtALBA1 and AtALBA2 bound to all sequences of nucleic acids we designed (fig. S3, B to D and table S1), their binding to nucleic acids was considered to be sequence independent. All of the observed bindings could be competed out with excess cold probe, and the binding of AtALBA1 to DNA-RNA hybrids was sensitive to RNase H digestion (fig. S3, B to D), indicative of the specificity of the bindings. The K30E mutation abolished the binding activities of AtALBA1-His and AtALBA2-His (Fig. 1, A and B), suggesting that the K30 residue is important for the binding of Alba

proteins to DNA, RNA, and DNA-RNA hybrids. To compare the relative affinities of AtALBA1 and AtALBA2 toward different types of nucleic acids, we quantified their affinities using the Agilent 2100 BioAnalyzer. Our results revealed that AtALBA1 and AtALBA2 have higher affinities toward DNA-RNA hybrids and dsDNA, respectively, in vitro (fig. S3, E and F).

### AtALBA1 and AtALBA2 colocalize and form heterodimers in the nucleus

We next investigated the subcellular localization of AtALBA1 and AtALBA2. We transiently expressed C-terminally green fluorescent protein (GFP)-tagged AtALBA1 and AtALBA2 (AtALBA1-GFP and AtALBA2-GFP) in *Arabidopsis* protoplasts. AtALBA1-GFP and AtALBA2-GFP were observed in both the cytoplasm and the nucleus (fig. S4A). These results were confirmed by subcellular fractionation experiments using transgenic plants (fig. S4B). Like Alba proteins in other species, AtALBA1 and AtALBA2 form homodimers and heterodimers, as determined by our split luciferase complementation and coimmunoprecipitation assays (fig. S4, C and D). To visualize



**Fig. 1. AtALBA1 and AtALBA2 bind R-loops in vitro.** (A) EMSA gel showing AtALBA1 binding to ssRNA and DNA-RNA hybrids. Different 5'-biotin-labeled substrates (5 nM) were incubated with increasing concentrations (25, 50, and 75 nM) of AtALBA1 wild-type protein (lanes 2 to 4) and 75 nM AtALBA1 (K30E) mutant protein (lane 5). (B) EMSA gel showing AtALBA2 binding to ssDNA and dsDNA. Different 5'-biotin-labeled substrates (5 nM) were incubated with increasing concentrations (25, 50, and 75 nM) of AtALBA2 wild-type protein (lanes 2 to 4) and 75 nM AtALBA2 (K30E) mutant protein (lane 5). (C) EMSA gel showing AtALBA1 binding to artificial R-loops. Artificial R-loop substrate (5 nM) with 5'-biotin-labeled DNA (1) or RNA (2) was incubated with 75 nM AtALBA1 wild-type protein. R-loop substrates were incubated with RNase H1 for 0 min and 10 min. (D) EMSA gel showing AtALBA2 binding to artificial R-loops. Artificial R-loop substrate (5 nM) with 5'-biotin-labeled DNA (1) or RNA (2) was incubated with 75 nM AtALBA2 wild-type protein. R-loop substrates were incubated with RNase H for 0 and 10 min. For EMSAs, at least three biological replicates were performed, and representative results are shown.

the nuclear localization patterns of homodimers and heterodimers formed from AtALBA1 and AtALBA2, we immunostained AtALBA1-Myc and AtALBA2-Flag in Col-0 and the F1 hybrid plants from the cross between *ALBA1-Myc* and *ALBA2-Flag* transgenic plants. AtALBA1 and AtALBA2 colocalized in approximately 92% of the transgenic nuclei, as shown by the yellow signals resulting from an overlap of the green and red signals (fig. S4E). No other signals, except the 4',6-diamidino-2-phenylindole (DAPI) signals, were detected in all wild-type nuclei (fig. S4E), suggesting the specificity of our staining. The colocalization of AtALBA1 and AtALBA2 are consistent with their heterodimerization.

### AtALBA1 and AtALBA2 bind R-loops in vitro

Because AtALBA1 and AtALBA2 interact and potentially heterodimerize in the nucleus and, based on our EMSA results, the heterodimers should be able to bind both DNA-RNA hybrids and the displaced ssDNA in R-loops, we hypothesized that AtALBA1 and AtALBA2 are R-loop-binding proteins. To test this hypothesis, we performed EMSAs using an artificial R-loop substrate (fig. S2C). Our results revealed that AtALBA1 and AtALBA2 bound artificial R-loops in a manner sensitive to RNase H treatment (Fig. 1, C and D). As expected, the positive control AtNDX also bound R-loops we designed (fig. S3A). Comparison of relative affinities toward R-loops using the Agilent 2100 BioAnalyzer revealed that the AtALBA1 and AtALBA2 heterodimer has a greater affinity toward R-loops than AtALBA1 or AtALBA2 alone (fig. S3G). Together, these results suggested that AtALBA1 and AtALBA2 can bind R-loops in vitro.

### AtALBA1 and AtALBA2 bind R-loops in vivo

To evaluate the possibility of AtALBA1 and AtALBA2 specifically recognizing R-loops in plants, we first performed chromatin immunoprecipitation (ChIP) combined with high-throughput sequencing (ChIP-seq) to identify genomic sites bound by AtALBA1. In total, 2146 binding peaks were consistently identified in two biological replicates of AtALBA1 ChIP-seq, and 2060 genes are associated with these peaks, accounting for approximately 4.63% of *Arabidopsis* genes (fig. S5A and table S2). Most of these peaks resided within genic regions, and AtALBA1 enrichment was observed across the gene body (Fig. 2, A and B). AtALBA1 was preferentially enriched on genes shorter than 2 kb (Fig. 2C). Analysis of the histone modification levels at peak regions revealed that AtALBA1 binding was highly coincident with histone modifications characteristic of actively transcribed genes, including H3K9Ac, H3K14Ac, H3K27Ac, H3K4me2, and H3K4me3. No correlation between AtALBA1 binding and repressive histone marks, such as H3K9me2, was found (Fig. 2D). Consistently, our immunostaining results showed that AtALBA1 and AtALBA2 are not enriched in repressive H3K9me1 domains (Fig. 2E). Further analysis of gene expression levels revealed that AtALBA1 peak-associated genes have significantly higher expression levels than non-AtALBA1-bound genes (fig. S5B). Our results indicated that AtALBA1 is more inclined to bind active genes.

To determine whether AtALBA2 binds to the same chromatin regions, we examined AtALBA1 and AtALBA2 enrichment at randomly selected genes by performing ChIP-quantitative polymerase chain reaction (qPCR). AtALBA2, like AtALBA1, was enriched at all examined AtALBA1-bound genes but not at non-AtALBA1-bound genes (fig. S5C). Our results suggest that AtALBA1 and AtALBA2 co-occupy a subset of chromatin regions. AtALBA1-FLAG and AtALBA2-MYC were not enriched on AtALBA1-bound genes when using AtALBA1-Flag

and AtALBA2-Myc transgenic plants in *alba1-1alba2-1* background, respectively, providing further evidence for heterodimerization of AtALBA1 and AtALBA2 at target loci (fig. S5C).

We then analyzed the presence or absence of R-loops in AtALBA1-bound genes using the available genome-wide R-loop data in *Arabidopsis* (11). We found a strong positive correlation between AtALBA1 binding and the presence of R-loops (Fig. 3A). Specifically, 75.5% of AtALBA1-bound genes were found to harbor R-loops (table S2). Genes harboring both sense and antisense R-loops (overlap R-loop) are significantly enriched in AtALBA1-bound genes (Fig. 3B). To further confirm that AtALBA1 and AtALBA2 specifically bind R-loop in vivo, we performed ChIP experiments after RNase H treatment. Our ChIP-qPCR results showed that the binding of AtALBA1 and AtALBA2 to randomly selected genes was sensitive to RNase H digestion (Fig. 3C). In contrast, the binding was not affected by RNase III treatment (fig. S5D). These results suggest that AtALBA1 and AtALBA2 can specifically recognize R-loops in vivo.

### R-loop levels are not affected in *alba* mutants

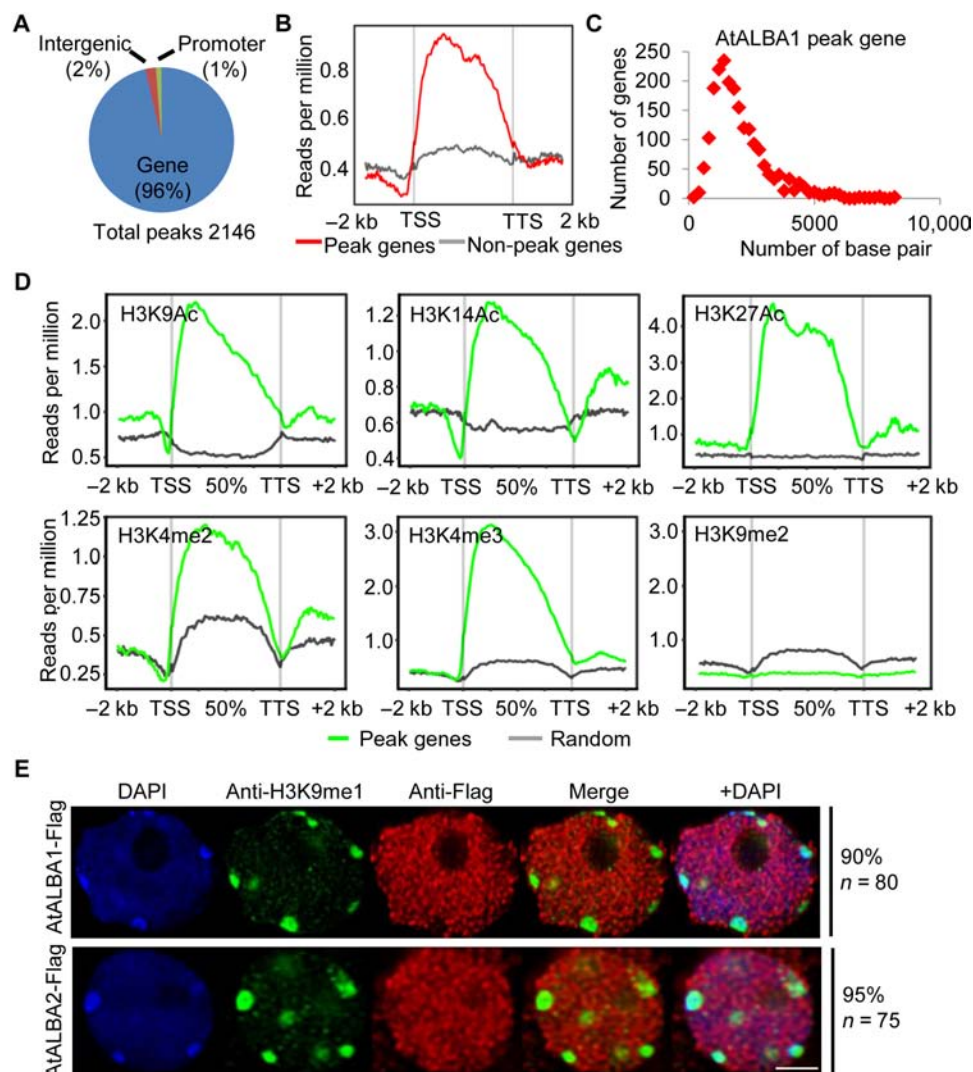
To study the functions of AtALBA1 and AtALBA2 in R-loop biology, we obtained transferred DNA (T-DNA) insertion mutants for *AtALBA1* and *AtALBA2* (fig. S6A). Reverse transcription (RT)-PCR experiments showed that *alba1-1* and *alba1-2* each had a complete loss of *AtALBA1* mRNA expression. A weak band corresponding to *AtALBA2* mRNA in *alba2-1* was detected, but it was shifted upward, which suggested that a nucleotide insertion event had occurred (fig. S6, B and C). Sanger sequencing confirmed a 27-nucleotide insertion within the T-DNA flanking sequence in the *AtALBA2* coding sequence (CDS) (fig. S6D), which caused a nine-amino acid insertion in the Alba domain of AtALBA2 (fig. S6E). The mutants for *AtALBA1* and *AtALBA2* did not exhibit obvious developmental phenotypes under normal growth conditions (fig. S6F).

Next, we tested whether the R-loop levels are affected in the *alba1-1alba2-1* double mutant. We immunostained nuclei isolated from Col-0 and *alba1-1alba2-1* plants using the R-loop antibody S9.6. Similar staining patterns were observed in the nuclei from each genotype (fig. S7A). To analyze genome-wide R-loop levels in *alba1-1alba2-1*, we performed single-strand DNA ligation-based library construction after DNA:RNA hybrid immunoprecipitation combined with next generation sequencing (ssDRIP-seq) (11). Our results revealed that the overall R-loop levels and R-loop levels at AtALBA1-bound loci in Col-0 and *alba1-2alba2-1* are comparable (fig. S7, B to E, and table S2). Together, these results suggest that AtALBA1 and AtALBA2 have minimal effects on R-loop stability. Because they bind to R-loops in vitro and in vivo, AtALBA1 and AtALBA2 may function as R-loop readers to recognize and bind to R-loops associated with genic regions within the *Arabidopsis* genome.

### AtALBA1 and AtALBA2 protect genic R-loops against DNA damage

R-loops are a source of genome instability (2). Protection of genic R-loop regions from being damaged is particularly important. Although AtALBA1 and AtALBA2 do not regulate R-loop levels, we next tested whether AtALBA1 and AtALBA2, as genic R-loop-binding proteins, can protect genic R-loops against DNA damage. Col-0 and *alba* single and double mutants were treated with or without the DNA alkylating agent methyl methanesulfonate (MMS). First, we detected  $\gamma$ H2AX foci by immunostaining using an anti- $\gamma$ H2AX antibody. In Col-0 and *alba* mutants without MMS treatment,  $\gamma$ H2AX foci could be barely detected (fig. S8A). The levels of  $\gamma$ H2AX foci were significantly increased in *alba* single and double mutants with MMS



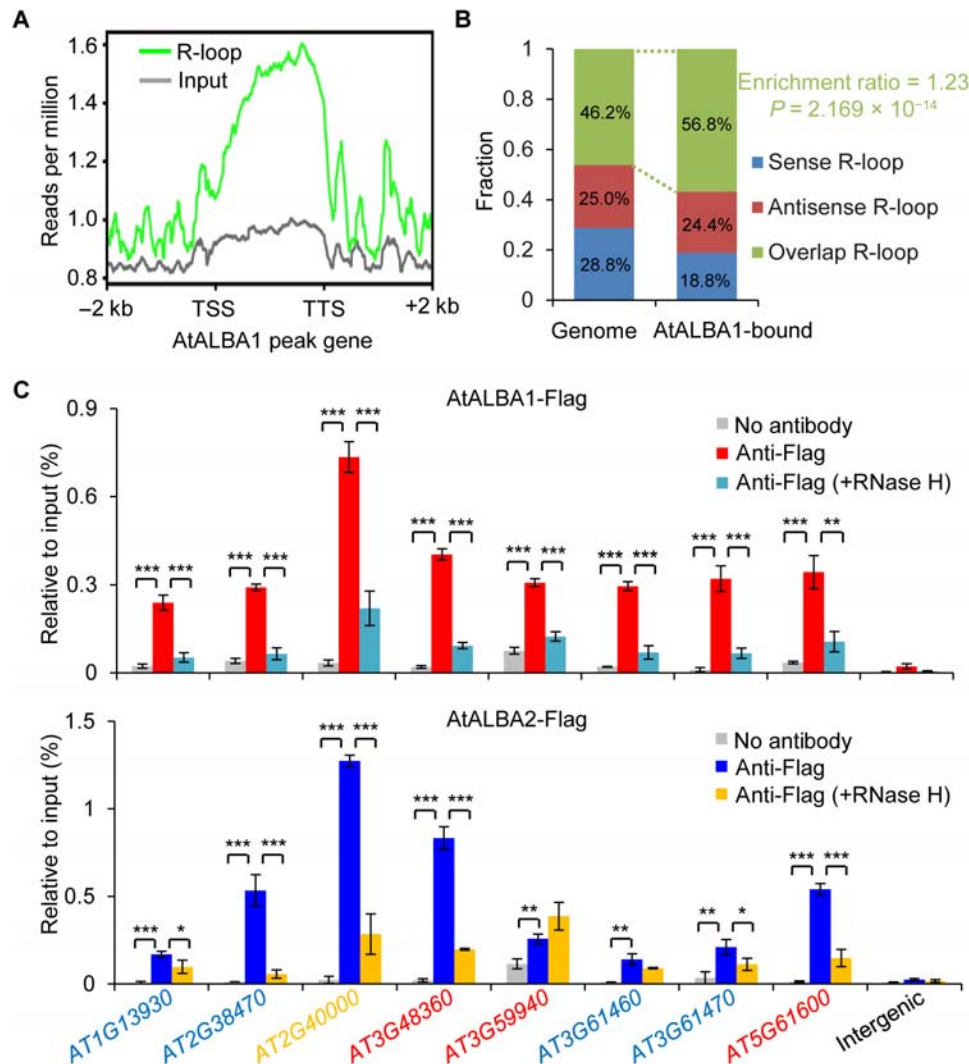


**Fig. 2. AtALBA1 preferentially binds gene body regions with active epigenetic marks in vivo.** (A) Total number and genomic distribution of AtALBA1 peaks identified by ChIP-seq. (B) Metagene plots of AtALBA1 ChIP-seq reads. TSS, transcription start site; -2 K and +2 K represent 2 kb upstream of the TSS and 2 kb downstream of the TTS, respectively. The y axis indicates AtALBA1 ChIP-seq read density. (C) Length distribution of AtALBA1-bound genes. The y axis indicates the number of genes. The x axis indicates the length of genes. (D) Metagene plots of histone modification levels on AtALBA1-bound genes. The y axis represents histone modification ChIP-seq read density. (E) The relationship between AtALBA1 and AtALBA2 binding and repressive histone modifications was determined by immunostaining. AtALBA1-Flag and AtALBA2-Flag in transgenic plants were stained with anti-Flag (red). H3K9me1 was stained with anti-H3K9me1 (green). DNA was stained with DAPI (blue). The frequency of nuclei displaying each interphase pattern is shown on the right. Scale bar, 2.5  $\mu$ m.

treatment (Fig. 4, A and B). The pattern of  $\gamma$ H2AX foci in the *alba* mutants resembles that induced by  $\gamma$ -irradiation (Fig. 4A and fig. S8B). Our results suggest that both AtALBA1 and AtALBA2 are required for the maintenance of genome stability. Second, we carried out RT-qPCR to detect the expression levels of *RAD51* and *BRCA1*, which are activated in response to DNA damage (37, 38). Our results revealed that the expression levels of *RAD51* and *BRCA1* were significantly increased in *alba* single and double mutants upon MMS treatment (fig. S8C), and this molecular phenotype can be complemented by *AtALBA1* or *AtALBA2* transgenes under the control of their native promoters (fig. S8D). Third, measurement of plant growth by measuring the fresh weight of plants revealed that *alba* single and double mutants were more sensitive than Col-0 plants to MMS (Fig. 4C). Notably, expression levels of *AtALBA1* and *AtALBA2*

were increased by MMS treatment (fig. S8E). The increase of *AtALBA1* and *AtALBA2* expression did not lead to altered localization pattern of *AtALBA1* and *AtALBA2* after MMS treatment (fig. S8F). To determine whether MMS treatment induces changes of R-loop levels, causing high sensitivity of *alba* mutants to DNA damage and induction of *AtALBA1* and *AtALBA2* expression, we analyzed R-loop levels in plants with and without MMS treatment (fig. S8G). The total R-loop levels remain unchanged with MMS treatment, although we could not exclude the possibility that the R-loop levels at specific loci change upon MMS treatment.

To demonstrate the direct role of AtALBA1 and AtALBA2 in the protection of genome stability, we next tested whether DNA damage in the *alba* mutants occurs at AtALBA1 and AtALBA2 target sites. We performed  $\gamma$ H2AX ChIP-seq using MMS-treated Col-0 and



**Fig. 3. AtALBA1 and AtALBA2 binding correlates with the presence of R-loops.** (A) Metagene plots of R-loop levels across AtALBA1-bound genes. The y axis indicates ssDRIP-seq read density. (B) Percentages of AtALBA1-bound genes overlapping with sense, antisense, and overlap (sense and antisense) R-loops. Enrichment ratio of AtALBA1-bound genes harboring overlap R-loops to all genes harboring overlap R-loops in the *Arabidopsis* genome was indicated. P value was calculated with R from Fisher's exact test. (C) Association of AtALBA1 and AtALBA2 with R-loops determined by ChIP-qPCR. Transgenic *AtALBA1-Flag/alba1-1* and *AtALBA2-Flag/alba2-1* plants were used. Expression of *AtALBA1-Flag* and *AtALBA2-Flag* was under the control of their respective native promoters. ChIP experiments were performed with anti-Flag antibody. The RNase H treatment was performed before cross-linking. Genes overlapping with sense, antisense, and overlap R-loops were represented by red, blue, and yellow colors, respectively. An intergenic region without R-loop formation is chosen as a negative control. Two biological replicates yielded very similar results. SEs were calculated from three technical replicates; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (two-tailed Student's *t* test).

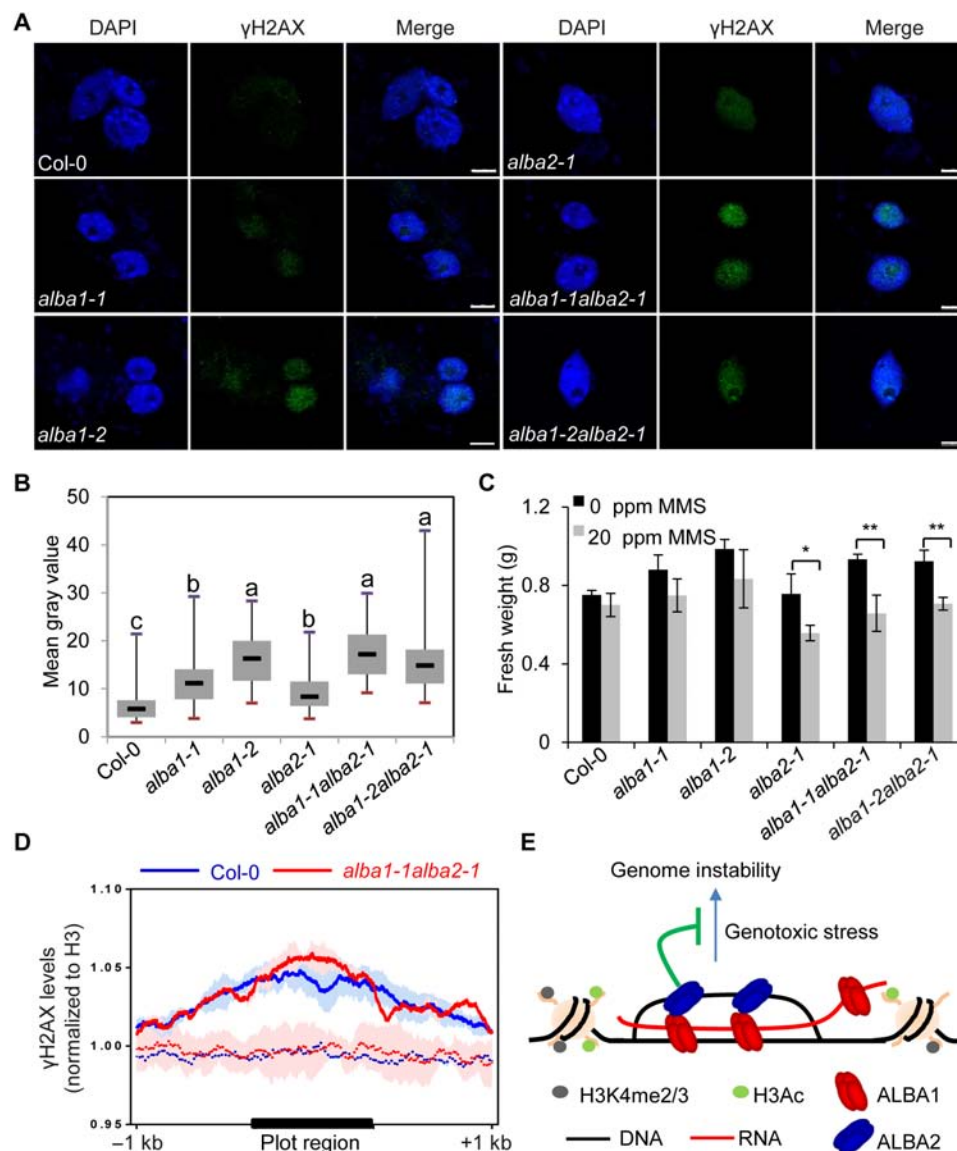
*alba1-1alba2-1* (fig. S8, H and I). Our results revealed that AtALBA1-bound regions were enriched with  $\gamma$ H2AX signals compared to randomly selected regions (Fig. 4D), and in *alba1-1alba2-1*, the  $\gamma$ H2AX signals were elevated compared to Col-0 in AtALBA1-bound regions (Fig. 4D). These results suggest that the AtALBA1-bound regions are particularly vulnerable to DNA damage and that AtALBA1 and AtALBA2 directly protect these regions from DNA damage.

## DISCUSSION

In this study, we find that AtALBA1 and AtALBA2 are the R-loop readers in *Arabidopsis*. They form heterodimers and bind a subset of R-loops at genic regions. Their binding protects genic R-loop regions from being damaged (Fig. 4E). Alba proteins in archaea and other

organisms have been found to regulate chromatin architecture, RNA metabolism, and protein translation (31). In plants, AtALBA1 and AtALBA2 have evolved to bind R-loops and maintain genome stability.

The unique characteristic of AtALBA1 and AtALBA2 is that they can bind DNA-RNA hybrid and ssDNA, respectively, and they can heterodimerize. This characteristic enables AtALBA1 and AtALBA2 to bind two parts of R-loops. Our EMSA and ChIP results demonstrate that they bind R-loops. AtALBA1 has a higher affinity to DNA-RNA hybrid than to ssRNA (fig. S3E). Thus, AtALBA1 preferentially recognizes R-loops. However, AtALBA2 has a lower affinity to ssDNA than to dsDNA (fig. S3F). To specifically bind R-loops, it may need to be recruited to R-loops by AtALBA1. In *alba1-1alba2-1* mutant background, AtALBA2 is not enriched on genes overlapping with R-loops (fig. S5C), suggesting that AtALBA1 and AtALBA2 bind



**Fig. 4. Depletion of AtALBA1 or AtALBA2 results in plant hypersensitivity to MMS.** (A) Representative microscopic images showing  $\gamma$ H2AX foci formation (green) in Col-0, *alba1-1*, *alba1-2*, *alba2-1*, *alba1-1alba2-1*, and *alba1-2alba2-1* plants treated with 50 ppm of MMS.  $\gamma$ H2AX foci were detected by immunostaining using an anti- $\gamma$ H2AX antibody. Nuclei were stained with DAPI (blue). Scale bars, 5  $\mu$ m. (B) Box plots showing the signal intensity of  $\gamma$ H2AX foci per nucleus for Col-0 plants and the indicated mutants. The  $\gamma$ H2AX signal intensity was analyzed by ImageJ software. Dark horizontal line, median; edges of boxes, 25th (bottom) and 75th (top) percentiles; whiskers, minimum and maximum gray values. The multiple comparison was calculated with Kruskal-Wallis. The  $\alpha$  parameter by default is 0.05. Post hoc test used the criterion Fisher's least significant difference. The adjustment methods include the Bonferroni correction and others. (C) Fresh weights of 10-day-old Col-0 seedlings and the indicated mutant seedlings grown on 1/2 MS medium supplemented with 0 or 20 ppm of MMS. The fresh weights of 120 seedlings were statistically analyzed. SEs were calculated from three biological replicates; \* $P$  < 0.05, \*\* $P$  < 0.01 (two-tailed Student's  $t$  test). (D) Metaplot of  $\gamma$ H2AX accumulation in AtALBA1-bound regions (solid lines) versus randomly selected regions (dash lines) in Col-0 and *alba1-1alba2-1* after MMS treatment. (E) A working model for the role of AtALBA1 and AtALBA2 in R-loop biology. AtALBA1 and AtALBA2 form a heterodimer or heteropolymer and bind R-loops at genic regions with active histone marks. By occupying R-loops, AtALBA1 and AtALBA2 protect R-loops from DNA damage and help maintain genome stability.

R-loops as heterodimers. This is different from all previously identified R-loop-associated factors, which target only one part of the R-loops. For instance, in *Arabidopsis*, the chloroplast-localized RNase H1 protein AtRNH1C cleaves the RNA strand of the DNA-RNA hybrid (10) and AtNDX binds the ssDNA of R-loop at the *COOLAIR* promoter (12). In human cells, many proteins interact with DNA-RNA hybrid parts of the R-loops (39).

AtALBA1 and AtALBA2 bind R-loops in a sequence-independent manner in vitro. We also could not find conserved DNA sequences for AtALBA1 to bind after bioinformatics analysis of our ChIP-seq data. However, AtALBA1 and AtALBA2 do not bind all R-loops in the *Arabidopsis* genome. About three-quarters of the 2060 AtALBA1-bound genes harbor R-loops. Thus, AtALBA1 binds approximately 1500 R-loops, corresponding to a small subset of R-loops in the *Arabidopsis* genome.



(~47,000 R-loops) (11). More than 90% of AtALBA1 binding resides within genic regions. AtALBA1 binding is preferentially associated with active epigenetic marks. In addition, genes harboring overlap R-loops are significantly enriched in AtALBA1-bound genes. However, the mechanisms through which AtALBA1 is recruited to R-loops with these features remain unclear. We propose that local chromatin environment may be important for determining the specificity of AtALBA1 targeting.

The functions of AtALBA1 and AtALBA2 in R-loop biology are also unique, as we found that the R-loop levels are not affected in the *alba* mutants. In previous studies, most, if not all, R-loop-associated factors regulate gene expression or genome stability through influencing R-loop levels. In *Arabidopsis*, AtNDX regulates *FLOWERING LOCUS C* expression and flowering by stabilizing the R-loop structure at the *COOLAIR* promoter (12). AtRNH1C, together with AtGyrases, maintains genome stability by restricting R-loop formation and releasing head-on transcription-replication conflicts in chloroplasts (10). In human cells, RNase H1, using RPA as the R-loop sensor, maintains genome stability by reducing R-loop levels (19). DXH9 helicase promotes transcriptional termination and prevents genome instability by suppressing R-loops (39).

Although AtALBA1 and AtALBA2 do not regulate R-loop levels, we found that AtALBA1 and AtALBA2 protect plant cells against DNA damage. Our  $\gamma$ H2AX ChIP-seq results further indicate that DNA damage in *alba1-1alba2-1* result from unprotected R-loops, suggesting that AtALBA1 and AtALBA2 directly prevent the occurrence of DNA damage on the R-loops they bind. Although R-loops are most enriched in promoters in human cells and plant cells (11, 40), genic R-loops are prevalent and, when these R-loops are not properly resolved, genomic instability (DNA double-strand breaks) can often be detected (11, 20, 41). Moreover, accumulation of R-loops in gene bodies causes asymmetric DNA mutagenesis (42). Thus, it is particularly important to resolve R-loops or protect R-loops in gene bodies. Because AtALBA1 and AtALBA2 recognize a subset of genic R-loops, they serve as specific caretakers of genic R-loops. Then, how do AtALBA1 and AtALBA2 execute their protective roles? In light of the previously documented roles of histones in protecting against spontaneous base mutations (43), oxidative DNA damage (44, 45), and radiation-induced DNA damage (46), we propose that, by occupying R-loops, AtALBA1 and AtALBA2 specifically protect R-loops from DNA damage (Fig. 4E). It is observed that *alba1-1* and *alba2-1* have some additive effects on  $\gamma$ H2AX accumulation (Fig. 4B). We reasoned that AtALBA1 and AtALBA2 may also form homodimers or heterodimerize with other AtALBA members to prevent the occurrence of DNA damage at different loci. In the future, it will be interesting to investigate the functions of AtALBA3 to AtALBA6 and their target specificity relative to AtALBA1 and AtALBA2.

## MATERIALS AND METHODS

### Plant materials and growth conditions

The T-DNA insertion lines SALK\_069210 (*alba1-1*), GK560\_B06 (*alba1-2*), and GK128\_D08 (*alba2-1*) were obtained from the Nottingham *Arabidopsis* Seed Center, UK. The genotypes of all homozygous mutants or double mutants were confirmed by PCR-based genotyping assays. After cold stratification for 2 days, sterilized seeds were grown on 1/2 Murashige-Skoog (MS) solid medium at 23°C under long-day condition (16 hours of light and 8 hours of darkness) for 14 days. The seedlings were then harvested for further

experiments or transplanted into soil and grown at 23°C with the same photoperiod.

For complementation of mutants, *AtALBA1* and *AtALBA2* genomic DNA with approximately 2-kb promoter regions was amplified from wild-type Col-0 genomic DNA by PCR and cloned into the binary vector *pCAMBIA1305* for plant transformation. *Agrobacterium tumefaciens* strain GV3101 carrying various *AtALBA1* or *AtALBA2* constructs was used to transform mutant plants via the standard floral dipping method. Primary transformants were selected on 1/2 MS plates containing hygromycin (25 mg/liter). T3 homozygous lines were used for further experiments. Refer to table S1 for detailed information regarding the primers used in this study.

### Transient expression of GFP-fusion constructs

To generate the GFP-fusion constructs, full-length *AtALBA1* and *AtALBA2* genomic DNA were PCR-amplified and cloned into the *Super1300-GFP* vector, which expresses the C-terminal GFP-tagged protein of interest under the control of a constitutive promoter. Transient expression assays were performed using mesophyll protoplasts from *Arabidopsis*. GFP signals were observed using a Leica TCS SP8 STED 3× confocal microscope.

### Nuclear-cytoplasmic fractionation

For nuclear-cytoplasmic fractionation, 14-day-old seedlings (0.5 g) were ground into fine powder in liquid N<sub>2</sub> using a cold mortar and pestle and then suspended in 1 ml of lysis buffer [20 mM Tris-HCl (pH 7.5), 20 mM KCl, 2 mM EDTA, 2.5 mM MgCl<sub>2</sub>, 25% glycerol, 250 mM sucrose, 5 mM dithiothreitol (DTT), and protease inhibitor cocktail]. After the homogenate was filtered through two layers of Miracloth, it was centrifuged at 1500g at 4°C for 10 min to pellet the nuclei. The supernatant was centrifuged at 10,000g at 4°C for 10 min and collected as the cytoplasmic fraction. The pellet was washed four times with 5 ml of nuclei resuspension buffer 1 (NRB1) [20 mM Tris-HCl (pH 7.5), 25% glycerol, 2.5 mM MgCl<sub>2</sub>, and 0.2% Triton X-100]. The pellet was resuspended in 500  $\mu$ l of NRB2 [20 mM Tris-HCl (pH 7.5), 0.25 M sucrose, 10 mM MgCl<sub>2</sub>, 0.5% Triton X-100, 5 mM  $\beta$ -mercaptoethanol, and protease inhibitor cocktail] and then carefully overlaid on top of 500  $\mu$ l of NRB3 [20 mM Tris-HCl (pH 7.5), 1.7 M sucrose, 10 mM MgCl<sub>2</sub>, 0.5% Triton X-100, 5 mM  $\beta$ -mercaptoethanol, and protease inhibitor cocktail]. Next, the sample was centrifuged at 16,000g for 45 min at 4°C. The final nuclear pellet was resuspended in 100  $\mu$ l of 2× protein loading buffer.

### Western blot

Proteins were separated using a 10% SDS-polyacrylamide gel electrophoresis and transferred onto polyvinylidene difluoride membranes. The membranes were blocked in TBST buffer [20 mM Tris-HCl (pH 7.5), 137 mM NaCl, and 0.1% Tween 20] with 5% nonfat milk for 1 hour and incubated with anti-Flag (F7425, Sigma), anti-Myc (05-724, Millipore), anti-Histone H3 (07-690, Millipore), or anti-tubulin antibodies (CW0098, CWBIO) overnight in TBST. After three washes with TBST, proteins were detected with a horseradish peroxidase chemiluminescence detection kit (CW0049, CWBIO).

### Split luciferase complementation assay

The full-length *AtALBA1* and *AtALBA2* CDSs were PCR-amplified and cloned into the *pCAMBIA1300-nLUC* vector or *pCAMBIA1300-cLUC* vector to generate an N-terminal or C-terminal luciferase-fusion construct, respectively. *A. tumefaciens* strain GV3101 carrying various

constructs was cultured in LB liquid medium with kanamycin (50 mg/liter) and rifampicin (50 mg/liter) at 28°C for 12 hours and resuspended in infiltration buffer [10 mM MES (pH 5.7), 10 mM MgCl<sub>2</sub>, and 150 μM acetosyringone] to reach an OD<sub>600</sub> (optical density at 600 nm) of 0.5. Equal amounts of the suspensions were mixed in different combinations, and the resulting mixtures were used to infiltrate *Nicotiana benthamiana* leaves. To prevent gene silencing, a construct encoding virus p19 protein was infiltrated at the same time at an OD<sub>600</sub> of 0.3. The infiltrated leaves were kept in the dark for 24 hours. Luciferase activity was detected with a luminescence imaging system (Princeton Instrument).

### Coimmunoprecipitation

F1 hybrids (14-day-old) from crosses between *AtALBA1-Myc* and *AtALBA2-Flag* transgenic plants and F1 hybrids from crosses between *AtALBA1-Flag* and *AtALBA1-Myc* transgenic plants and *AtALBA1-Flag* and *AtALBA2-Flag* transgenic plants were fast-frozen and ground in liquid N<sub>2</sub>. The resulting fine powder (1 g) was suspended in 2 ml of lysis buffer [50 mM tris-HCl (pH 8.0), 230 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, 0.2% NP-40, 0.5 mM DTT, 1 mM phenylmethylsulfonyl fluoride (PMSF), and protease inhibitor cocktail]. After centrifugation, the supernatant was incubated with anti-Myc agarose (A7470, Sigma) at 4°C for 3 hours. The beads were washed three times with 10 ml of washing buffer [50 mM tris-HCl (pH 7.5), 150 mM NaCl, and 5 mM EDTA]. The immunoprecipitates were subjected to Western blot analyses using anti-Flag (F1804, Sigma) and anti-Myc (05-724, Millipore) antibodies as primary antibodies.

### Immunolocalization

Immunofluorescence localization assays were performed as described by Martínez-Macías *et al.* (47). First, seedling tissue samples were used for nuclei preparation. The nuclei preparations were incubated at room temperature with different combinations of anti-Flag (F7425, Sigma), anti-Flag (F1804, Sigma), anti-Myc (05-724, Millipore), H3K9me1 (07-352, Millipore), S9.6 (from Q. Sun's laboratory, Tsinghua University), and anti-γH2AX (4418-APC-020, Trevigen) primary antibodies overnight, after which they were incubated with mouse Alexa594 (A23410, Abbkine)-conjugated or rabbit Alexa-488 (A23220, Abbkine)-conjugated secondary antibodies for 2 hours at 37°C. After washing with phosphate-buffered saline, DNA was counterstained using DAPI in Prolong Gold Antifade Mountant (Invitrogen). Nuclei were observed with a confocal microscope, Leica TCS SP8 STED 3× (Leica).

### Electrophoretic mobility shift assay

Full-length *AtALBA1* and *AtALBA2* CDSs were amplified and cloned into the pET28a expression vector for protein purification. The K30E mutation was introduced into the construct through site-directed mutagenesis with the QuikChange II XL Site-Directed Mutagenesis Kit according to the manufacturer's instructions (Agilent Technologies). The proteins were expressed in *Escherichia coli* DE3 (BL21) cells and purified by nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography. EMSA was performed as previously described (19). Oligonucleotide sequences used in this study are described in table S1. The indicated DNA or RNA oligonucleotides were synthesized and labeled with biotin at the 5' end. Next, the oligonucleotides were annealed to a complementary strand by heating them to 95°C for 5 min and cooling them slowly. The annealing created dsDNA, dsRNA, a DNA-RNA hybrid, and an R-loop structure. The oligonucleotides (5 nM) were incubated with *AtALBA1* or *AtALBA2* recombinant proteins at 25°C for 10 min in the binding buffer [20 mM tris-HCl (pH 7.6), 10 mM MgCl<sub>2</sub>, and

1 mM DTT]. The resulting protein-substrate complexes were resolved on 4% nondenaturing polyacrylamide gels at 80 V for 80 min using 1× TBE buffer (89 mM tris-HCl, 89 mM boric acid, and 2 mM disodium EDTA). After electrophoresis, the oligonucleotides in the gels were detected using a chemiluminescent biotin-labeled nucleic acid detection kit (D3308, Beyotime).

### ChIP assay and data analysis

For *AtALBA1* and *AtALBA2* ChIP assays, 14-day-old seedlings (2 g) were ground into powder in liquid N<sub>2</sub> and cross-linked in cold ChIP extraction buffer I [10 mM tris-HCl (pH 7.5), 10 mM MgCl<sub>2</sub>, and 400 mM sucrose] containing 1% formaldehyde at 4°C for 10 min. In some experiments, RNase H [M0297, New England Biolabs (NEB)] or RNase III (M0245, NEB) treatment was performed before cross-linking. The cross-linking reaction was quenched by adding glycine to a final concentration of 0.125 M. The homogenate was filtered through a cell strainer (431751, Falcon) and pelleted by centrifugation at 4000 rpm for 20 min at 4°C. The precipitates were washed several times with ChIP extraction buffer II [10 mM tris-HCl (pH 7.5), 10 mM MgCl<sub>2</sub>, 250 mM sucrose, and 1% Triton X-100] until they became white. The nuclei were suspended and incubated in 100 μl of nuclear lysis buffer [50 mM tris-HCl (pH 8.0), 10 mM EDTA, and 1% SDS] for 30 min at 4°C. After the addition of 200 μl of ChIP dilution buffer [16.7 mM tris-HCl (pH 8.0), 1.2 mM EDTA, 1.1% Triton X-100, and 167 mM NaCl], nuclei were sonicated for 24 cycles (UCD-200, Diagenode) to yield DNA fragments of 0.2 to 0.5 kb in length. After centrifugation, the chromatin supernatant was diluted with 700 μl of dilution buffer. For ChIP-seq, the sample was incubated with anti-Flag beads (M8823, Sigma) at 4°C overnight. For ChIP-qPCR, the sample was incubated with anti-Flag (F3165, Sigma) or anti-Myc (ab32, Abcam). After washing, elution, and the reversal of cross-linking, DNA was recovered by phenol/chloroform extraction and ethanol precipitation. For ChIP-seq, two biological replicates of the enriched DNA were subjected to library construction. An Illumina HiSeq 2000 instrument was used for high-throughput single-end sequencing of libraries. For ChIP-qPCR, three biological replicates of the enriched DNA were subjected to qPCR analyses.

For γH2AX ChIP-seq, the native ChIP method was applied. In detail, MMS-treated seedlings were harvested and ground to a fine powder in liquid nitrogen. The nuclei were extracted and washed with Honda buffer [0.44 M sucrose, 1.25% Ficoll, 2.5% Dextran T40, 20 mM Hepes (pH 7.4), 10 mM MgCl<sub>2</sub>, 0.5% Triton X-100, 5 mM DTT, and protease inhibitor cocktail]. Then, the nuclei were resuspended in 500 μl of MNase buffer [50 mM tris-HCl (pH 7.6), 5 mM CaCl<sub>2</sub>, 0.1 mM PMSF, and protease inhibitor cocktail] and incubated with RNase A at 37°C for 30 min. The sample volumes were adjusted to 1 ml with MNase buffer, and then, 4.5 μl of Micrococcal Nuclease (M0247S, NEB) was added. After incubation at 37°C for 30 min, the fragmentation process was stopped by the addition of EDTA (final concentration, 10 mM). Nucleosomes were released by the addition of 0.1% SDS and rotation of samples at 4°C for 3 hours. The samples were centrifuged, and supernatants were diluted with MNase dilution buffer (0.1% Triton X-100, 50 mM NaCl, 0.1 mM PMSF, and protease inhibitor cocktail). One day before the isolation of chromatin, protein Dynabeads G (Invitrogen) beads were incubated with γH2AX and H3 antibodies (ABclonal) in ChIP dilution buffer [1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM tris-HCl (pH 8.0), and 167 mM NaCl, protease inhibitor cocktail] at 4°C overnight. After the isolation of chromatin, protein Dynabeads G



beads conjugated with antibodies were washed twice with 1 ml of MNase dilution buffer and added into diluted chromatin. The immunoprecipitation step was performed at 4°C for 5 hours. Then, beads were washed twice with low-salt wash buffer [50 mM tris-HCl (pH 7.6), 10 mM EDTA, 50 mM NaCl, 0.1 mM PMSF, and protease inhibitor cocktail], washed twice with intermediate-salt wash buffer [50 mM tris-HCl (pH 7.6), 10 mM EDTA, 100 mM NaCl, 0.1 mM PMSF, and protease inhibitor cocktail], washed once with high-salt wash buffer [50 mM tris-HCl (pH 7.6), 10 mM EDTA, 150 mM NaCl, 0.1 mM PMSF, and protease inhibitor cocktail], and finally washed once with TE buffer [1 mM EDTA and 10 mM tris-HCl (pH 8.0)]. Immune complexes were eluted twice by the addition of 200  $\mu$ l of elution buffer (0.1% SDS and 0.1 M NaHCO<sub>3</sub>) at 65°C for 10 min. The samples were treated with 2  $\mu$ l of proteinase K (10 mg/ml) at 45°C for 1 hour. Final DNA was recovered by phenol/chloroform extraction and ethanol precipitation.

ChIP-seq reads were aligned to the *Arabidopsis* genome (TAIR10) using the Bowtie2 (v2.1.0) program (Illumina) with the default parameters. Duplicated reads and reads with low mapping quality were removed with SAMtools. Only perfectly and uniquely mapped reads were retained for further analysis. BigWig files of the alignments were generated using bam2wig and visualized using the Integrated Genome Browser. The number of reads from each biological replicate is summarized in fig. S8 and table S2. To determine the correlations between biological replicates, Pearson correlations were calculated from the normalized signal intensities using deepTools. MACS2 was used for peak calling with  $P = 1 \times 10^{-3}$ .

### Epigenetic feature analysis

Histone modification levels and gene expression levels at AtALBA1-bound loci were determined using previously published ChIP-seq (31, 48–51) and RNA sequencing (RNA-seq) datasets, respectively. ChIP-seq data of H3K9Ac and H3K14Ac (GSE89768), H3K27Ac (GSE80056), H3K4me2 and H3K4me3 (GSE73972), and H3K9me2 (SRA010097) for Col-0 and RNA-seq data (GSE80303) for Col-0 were downloaded from the Gene Expression Omnibus (GEO) database.

### ssDRIP-seq and data analysis

ssDRIP was performed as described previously (11) with some modifications. Briefly, 14-day-old seedlings (3 g) were ground into a fine powder in liquid N<sub>2</sub> and suspended in 30 ml of precooled Honda buffer [20 mM Hepes (pH 7.4), 0.44 M sucrose, 1.25% Ficoll, 2.5% Dextran T40, 10 mM MgCl<sub>2</sub>, and 0.5% Triton X-100]. After the homogenate was filtered through two layers of Miracloth, it was centrifuged at 2000g for 15 min at 4°C to pellet the nuclei. The nuclei were washed three times with NRB1 [20 mM tris-HCl (pH 7.5), 25% glycerol, 2.5 mM MgCl<sub>2</sub>, and 0.2% Triton X-100]. The pellet was resuspended in 2 ml of TE buffer [10 mM tris-HCl (pH 8.0) and 1 mM EDTA] supplemented with 0.5% SDS and proteinase K (0.33 mg/ml) and incubated at 37°C overnight with constant shaking at 400 rpm. DNA was recovered via phenol/chloroform extraction, and purified DNA was digested with Mse I, Mbo I, Dde I, and Nla III for fragmentation. For S9.6 immunoprecipitation, 4  $\mu$ g of fragmented DNA (the DNA concentration was measured with a Qubit dsDNA kit) was incubated with 10  $\mu$ g of S9.6 antibodies in DRIP-binding buffer [10 mM NaPO<sub>4</sub> (pH 7.0), 0.14 M NaCl, and 0.05% Triton X-100] at 4°C overnight. The DNA-antibody complexes were incubated with Protein G beads (Invitrogen) at 4°C for 4 hours with rotation. The Protein G beads were washed four times with DRIP-binding

buffer at room temperature. The beads were incubated with elution buffer [50 mM tris-HCl (pH 8.0) and 10 mM EDTA] at 55°C for 1 hour with rotation at 1000 rpm. Last, 200  $\mu$ g of proteinase K was added for protein digestion. DRIPed DNA was recovered via phenol/chloroform extraction and used for library construction as described previously (11).

ssDRIP data analysis was performed, as described previously (11). Trimmed reads were aligned to the *Arabidopsis* genome (TAIR10) using Bowtie 2 (v2.3.0) with the default settings. Reads with more than three mismatches and nonuniquely mapped reads were removed by SAMtools (v1.3.1). The set of mapped reads was divided into forward and reverse groups for sense/antisense R-loop analysis. The definition of sense/antisense R-loops was provided by Xu *et al.* (11). MACS2 was used to identify peaks for each sample. Binary Alignment Map (BAM) files were converted to normalized coverage files (BigWig) with 5-base pair bins using deepTools (v2.26.0). BigWig files were used for visualization and building of Metaplots with computeMatrix from deepTools.

### Slot blot hybridization analysis

Genomic DNA was purified as described in DRIP-seq. One hundred nanograms of genomic DNA from different backgrounds were treated with or without RNase H and then was slotted onto nitrocellulose membrane (Hybond N<sup>+</sup>, GE Amersham) and detected by an S9.6 antibody.

### Gene expression analysis

Total RNA was extracted from 14-day-old seedlings using the RNeasy Plant Mini Kit (Qiagen) and treated with RNase-free DNase (Qiagen). RT was performed using the PrimeScript II First-Strand Synthesis System (6210A, Takara). RNA transcript levels were determined by semi-quantitative RT-PCR or real-time PCR. Real-time PCR was performed using the Perfect Real-time Kit (Takara). *ACTIN2* was used as an internal control. The primers used for PCR are listed in table S1.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/5/eaav9040/DC1>

Fig. S1. Domain structure of ALBA proteins in *Arabidopsis*.

Fig. S2. Purification of AtALBA1 and AtALBA2 wild-type and mutant proteins and diagram of probes used in EMSAs.

Fig. S3. Characterization of the nucleic acid binding properties of AtALBA1 and AtALBA2.

Fig. S4. Subcellular localization and interaction of AtALBA1 and AtALBA2.

Fig. S5. Characterization of AtALBA1-bound loci.

Fig. S6. Characterization of the T-DNA insertion mutants for *AtALBA1* and *AtALBA2*.

Fig. S7. Detection of R-loop levels in Col-0 and *alba1-1alba2-1* by immunostaining and ssDRIP-seq.

Fig. S8. Molecular phenotypes of Col-0 and *alba1-1alba2-1* without and with MMS treatment.

Table S1. Primers and substrates used in this study.

Table S2. List of AtALBA1-bound loci.

### REFERENCES AND NOTES

1. K. Skourti-Stathaki, N. J. Proudfoot, A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* **28**, 1384–1396 (2014).
2. J. M. Santos-Pereira, A. Aguilera, R loops: New modulators of genome dynamics and function. *Nat. Rev. Genet.* **16**, 583–597 (2015).
3. F. Chédin, Nascent connections: R-loops and chromatin patterning. *Trends Genet.* **32**, 828–838 (2016).
4. J. Sollier, K. A. Cimprich, Breaking bad: R-loops and genome integrity. *Trends Cell Biol.* **25**, 514–522 (2015).
5. N. Taneja, M. Zofall, V. Balachandran, G. Thillainadesan, T. Sugiyama, D. Wheeler, M. Zhou, S. I. S. Grewal, SNF2 family protein Fft3 suppresses nucleosome turnover to promote epigenetic inheritance and proper replication. *Mol. Cell* **66**, 50–62.e6 (2017).

6. K. Skourti-Stathaki, K. Kamieniarz-Gdula, N. J. Proudfoot, R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439 (2014).
7. T. Nojima, M. Tellier, J. Foxwell, C. Ribeiro de Almeida, S. M. Tan-Wong, S. Dhir, G. Dujardin, A. Dhir, S. Murphy, N. J. Proudfoot, Deregulated expression of mammalian lncRNA through loss of SPT6 induces R-loop formation, replication stress, and cellular senescence. *Mol. Cell* **72**, 970–984.e7 (2018).
8. O. Velazquez Camacho, C. Galan, K. Swist-Rosowska, R. Ching, M. Gamalinda, F. Karabiber, I. De La Rosa-Velazquez, B. Engist, B. Koschorz, N. Shukeir, M. Onishi-Seebacher, S. van de Nobelen, T. Jenuwein, Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. *eLife* **6**, e25293 (2017).
9. L. Kabeche, H. D. Nguyen, R. Buisson, L. Zou, A mitosis-specific and R-loop-driven ATR pathway promotes faithful chromosome segregation. *Science* **359**, 108–114 (2018).
10. Z. Yang, Q. Hou, L. Cheng, W. Xu, Y. Hong, S. Li, Q. Sun, RNase H1 cooperates with DNA gyrase to restrict R-loops and maintain genome integrity in *Arabidopsis* chloroplasts. *Plant Cell* **29**, 2478–2497 (2017).
11. W. Xu, H. Xu, K. Li, Y. Fan, Y. Liu, X. Yang, Q. Sun, The R-loop is a common chromatin feature of the *Arabidopsis* genome. *Nat. Plants* **3**, 704–714 (2017).
12. Q. Sun, T. Csorba, K. Skourti-Stathaki, N. J. Proudfoot, C. Dean, R-loop stabilization represses antisense transcription at the *Arabidopsis* *FLC* locus. *Science* **340**, 619–621 (2013).
13. X. Li, J. L. Manley, Cotranscriptional processes and their influence on genome stability. *Genes Dev.* **20**, 1838–1847 (2006).
14. A. Aguilera, T. García-Muse, R loops: From transcription byproducts to threats to genome stability. *Mol. Cell* **46**, 115–124 (2012).
15. W. Gan, Z. Guan, J. Liu, T. Gui, K. Shen, J. L. Manley, X. Li, R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041–2056 (2011).
16. S. Hamperl, M. J. Bocek, J. C. Saldivar, T. Swigut, K. A. Cimprich, Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* **170**, 774–786.e19 (2017).
17. C. T. Stork, M. Bocek, M. P. Crossley, J. Sollier, L. A. Sanz, F. Chédin, T. Swigut, K. A. Cimprich, Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *eLife* **5**, e17548 (2016).
18. A. Helmrich, M. Ballarino, L. Tora, Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* **44**, 966–977 (2011).
19. H. D. Nguyen, T. Yadav, S. Giri, B. Saez, T. A. Graubert, L. Zou, Functions of replication protein A as a sensor of R loops and a regulator of RNaseH1. *Mol. Cell* **65**, 832–847.e4 (2017).
20. L. Wahba, J. D. Amon, D. Koshland, M. Vuica-Ross, RNase H and multiple RNA biogenesis factors cooperate to prevent RNA-DNA hybrids from generating genome instability. *Mol. Cell* **44**, 978–988 (2011).
21. H. E. Mischo, B. Gómez-González, P. Grzechnik, A. G. Rondón, W. Wei, L. Steinmetz, A. Aguilera, N. J. Proudfoot, Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell* **41**, 21–32 (2011).
22. J. M. Santos-Pereira, A. B. Herrero, M. L. García-Rubio, A. Marín, S. Moreno, A. Aguilera, The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes Dev.* **27**, 2445–2458 (2013).
23. I. Salas-Armenteros, C. Pérez-Calero, A. Bayona-Feliu, E. Tumini, R. Luna, A. Aguilera, Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. *EMBO J.* **36**, 3532–3547 (2017).
24. V. Bhatia, S. I. Barroso, M. L. García-Rubio, E. Tumini, E. Herrera-Moyano, A. Aguilera, BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365 (2014).
25. L. Costantino, D. Koshland, Genome-wide map of R-loop-induced damage reveals how a subset of R-loops contributes to genomic instability. *Mol. Cell* **71**, 487–497.e3 (2018).
26. M. Goyal, C. Banerjee, S. Nag, U. Bandyopadhyay, The Alba protein family: Structure and function. *Biochim. Biophys. Acta* **1864**, 570–583 (2016).
27. B. N. Wardleworth, R. J. M. Russell, S. D. Bell, G. L. Taylor, M. F. White, Structure of Alba: An archaeal chromatin protein modulated by acetylation. *EMBO J.* **21**, 4654–4662 (2002).
28. T. Tanaka, S. Padavattan, T. Kumarevel, Crystal structure of archaeal chromatin protein Alba2–double-stranded DNA complex from *Aeropyrum pernix* K1. *J. Biol. Chem.* **287**, 10394–10402 (2012).
29. N. Laurens, R. P. C. Driessen, I. Heller, D. Vorselen, M. C. Noom, F. J. H. Hol, M. F. White, R. T. Dame, G. J. L. Wuite, Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat. Commun.* **3**, 1328 (2012).
30. R. Guo, H. Xue, L. Huang, Ssh10b, a conserved thermophilic archaeal protein, binds RNA in vivo. *Mol. Microbiol.* **50**, 1605–1615 (2003).
31. L. Aravind, L. M. Iyer, V. Anantharaman, The two faces of Alba: The evolutionary connection between proteins participating in chromatin structure and RNA metabolism. *Genome Biol.* **4**, R64 (2003).
32. A. Dupé, C. Dumas, B. Papadopoulou, An Alba-domain protein contributes to the stage-regulated stability of amastin transcripts in *Leishmania*. *Mol. Microbiol.* **91**, 548–561 (2014).
33. J. Mani, A. Guttinger, B. Schimanski, M. Heller, A. Acosta-Serrano, P. Pescher, G. Spath, I. Roditi, Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery. *PLOS ONE* **6**, e22463 (2011).
34. M. Gissot, R. Walker, S. Delhay, T. D. Alayi, L. Huot, D. Hot, I. Callebaut, C. Schaeffer-Reiss, A. V. Dorsselaer, S. Tomavo, *Toxoplasma gondii* Alba proteins are involved in translational control of gene expression. *J. Mol. Biol.* **425**, 1287–1301 (2013).
35. S. D. Bell, C. H. Botting, B. N. Wardleworth, S. P. Jackson, M. F. White, The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation. *Science* **296**, 148–151 (2002).
36. K. Zhao, X. Chai, R. Marmorstein, Structure of a Sir2 substrate, Alba, reveals a mechanism for deacetylation-induced enhancement of DNA binding. *J. Biol. Chem.* **278**, 26071–26077 (2003).
37. M.-P. Doutriaux, F. Couteau, C. Bergounioux, C. White, Isolation and characterisation of the *RAD51* and *DMC1* homologs from *Arabidopsis thaliana*. *Mol. Gen. Genet.* **257**, 283–291 (1998).
38. S. Lafarge, M.-H. Montané, Characterization of *Arabidopsis thaliana* ortholog of the human breast cancer susceptibility gene 1: *AtBRCA1*, strongly induced by gamma rays. *Nucleic Acids Res.* **31**, 1148–1155 (2003).
39. A. Cristini, M. Groh, M. S. Kristiansen, N. Gromak, RNA/DNA hybrid interactome identifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. *Cell Rep.* **23**, 1891–1905 (2018).
40. L. Chen, J.-Y. Chen, X. Zhang, Y. Gu, R. Xiao, C. Shao, P. Tang, H. Qian, D. Luo, H. Li, Y. Zhou, D.-E. Zhang, X.-D. Fu, R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. *Mol. Cell* **68**, 745–757.e5 (2017).
41. E. Pefanis, J. Wang, G. Rothschild, J. Lim, J. Chao, R. Rabadan, A. N. Economides, U. Basu, Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* **514**, 389–393 (2014).
42. J. Lim, P. K. Giri, D. Kazadi, B. Laffleur, W. Zhang, V. Grinstein, E. Pefanis, L. M. Brown, E. Ladewig, O. Martin, Y. Chen, R. Rabadan, F. Boyer, G. Rothschild, M. Cogné, E. Pinaud, H. Deng, U. Basu, Nuclear proximity of Mtr4 to RNA exosome restricts DNA mutational asymmetry. *Cell* **169**, 523–537.e15 (2017).
43. X. Chen, Z. Chen, H. Chen, Z. Su, J. Yang, F. Lin, S. Shi, X. He, Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. *Science* **335**, 1235–1238 (2012).
44. M. Ljungman, P. C. Hanawalt, Efficient protection against oxidative DNA damage in chromatin. *Mol. Carcinog.* **5**, 264–269 (1992).
45. H. U. Enright, W. J. Miller, R. P. Hebbel, Nucleosomal histone protein protects DNA from iron-mediated damage. *Nucleic Acids Res.* **20**, 3341–3346 (1992).
46. H. Takata, T. Hanafusa, T. Mori, M. Shimura, Y. Iida, K. Ishikawa, K. Yoshikawa, Y. Yoshikawa, K. Maeshima, Chromatin compaction protects genomic DNA from radiation damage. *PLOS ONE* **8**, e75622 (2013).
47. M. I. Martínez-Macias, W. Qian, D. Miki, O. Pontes, Y. Liu, K. Tang, R. Liu, T. Morales-Ruiz, R. R. Ariza, T. Roldán-Arjona, J.-K. Zhu, A DNA 3'-phosphatase functions in active DNA demethylation in *Arabidopsis*. *Mol. Cell* **45**, 357–370 (2012).
48. L.-Q. Chen, J.-H. Luo, Z.-H. Cui, M. Xue, L. Wang, X.-Y. Zhang, W. P. Pawlowski, Y. He, ATX3, ATX4, and ATX5 encode putative H3K4 methyltransferases and are critical for plant development. *Plant Physiol.* **174**, 1795–1806 (2017).
49. X. Chen, L. Lu, K. S. Mayer, M. Scaff, S. Qian, A. Lomax, L. M. Smith, X. Zhong, POWERDRESS interacts with HISTONE DEACETYLASE 9 to promote aging in *Arabidopsis*. *eLife* **5**, e17214 (2016).
50. Y. J. Kim, R. Wang, L. Gao, D. Li, C. Xu, H. Mang, J. Jeon, X. Chen, X. Zhong, J. M. Kwak, B. Mo, L. Xiao, X. Chen, POWERDRESS and HDA9 interact and promote histone H3 deacetylation at specific genomic sites in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14858–14863 (2016).
51. C. Luo, D. J. Sidote, Y. Zhang, R. A. Kerstetter, T. P. Michael, E. Lam, Integrative analysis of chromatin states in *Arabidopsis* identified potential regulatory mechanisms for natural antisense transcript production. *Plant J.* **73**, 77–90 (2013).

**Acknowledgments:** We thank X. Wang for bioinformatics analysis, D. Li for technical support with ChIP assay, G. Li for technical support with protein purification, and C. Shan for image processing. **Funding:** This work was funded by grants from the Ministry of Science and Technology of China (grant no. 2016YFA0500800 to W.Q. and Q.S.) and the National Natural Science Foundation of China (grant nos. 31571326 and 31522005 to W.Q. and 91740105 and 31822028 to Q.S.) **Author contributions:** W.Q. and Q.S. conceptualized the study. W.Q. and W.Y. designed the experiments. W.Y. performed most of the experiments. J.T. performed the split luciferase complementation assay and coimmunoprecipitation assay. J.Z. performed

DRIP-seq and  $\gamma$ H2AX ChIP-seq and data analysis. L.W. performed a portion of the EMSAs, immunofluorescence experiments, and MMS treatment assays. W.Z. performed the ChIP assay. W.Y., J.Z., Y.L. Q.S., and W.Q. wrote the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. Sequence data is available under accession nos. SRP134706, GSE124943, and GSE121683.

Submitted 30 October 2018

Accepted 2 April 2019

Published 15 May 2019

10.1126/sciadv.aav9040

**Citation:** W. Yuan, J. Zhou, J. Tong, W. Zhuo, L. Wang, Y. Li, Q. Sun, W. Qian, ALBA protein complex reads genic R-loops to maintain genome stability in *Arabidopsis*. *Sci. Adv.* **5**, eaav9040 (2019).