LETTERS

Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases

Yannick Jacob[1]*, Hume Stroud[2]*, Chantal LeBlanc[1], Suhua Feng[3], Luting Zhuo[1], Elena Caro[2], Christiane Hassel[1], Crisanto Gutierrez[4], Scott D. Michaels[1] & Steven E. Jacobsen[2,3]

Multiple pathways prevent DNA replication from occurring more than once per cell cycle[1]. These pathways block re-replication by strictly controlling the activity of pre-replication complexes, which assemble at specific sites in the genome called origins. Here we show that mutations in the homologous histone 3 lysine 27 (H3K27) monomethyltransferases, *ARABIDOPSIS TRITHORAX-RELATED PROTEIN5* (*ATXR5*) and *ATXR6*, lead to re-replication of specific genomic locations. Most of these locations correspond to transposons and other repetitive and silent elements of the *Arabidopsis* genome. These sites also correspond to high levels of H3K27 monomethylation, and mutation of the catalytic SET domain is sufficient to cause the re-replication defect. Mutation of *ATXR5* and *ATXR6* also causes upregulation of transposon expression and has pleiotropic effects on plant development. These results uncover a novel pathway that prevents over-replication of heterochromatin in *Arabidopsis*.

We previously characterized two redundant histone methyltransferase genes, *ATXR5* and *ATXR6*, and demonstrated that the ATXR5 and ATXR6 proteins show H3K27 monomethylation (H3K27me1) activity *in vitro*, and that an *atxr5 atxr6* double mutant shows a reduction of H3K27me1 *in vivo*[2]. The *atxr5 atxr6* double mutant shows pleiotropic defects in plant development, including smaller misshapen leaves[2]. Overexpression of ATXR5 or ATXR6 also causes morphological defects and male sterility[3]. Furthermore, the double mutant displays reactivation of the expression of a variety of both DNA transposons and retrotransposons[2]. Notably, *atxr5 atxr6* did not disturb DNA methylation or histone H3K9 dimethylation (H3K9me2, a key repressive histone modification correlated with DNA methylation[4–6]), indicating that ATXR5 and ATXR6 act by means of a novel pathway to maintain gene silencing. Previous work also suggested that ATXR5 and ATXR6 show links with DNA replication. *ATXR5* and *ATXR6* expression is regulated by the cell cycle, with expression peaking just before DNA replication[3], and *ATXR6* expression is strongly co-regulated with CDT1, ORC2 and other DNA replication proteins[7]. In addition, ATXR5 and ATXR6 contain PCNA-interacting-protein (PIP) motifs and have been shown to interact with the two PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) proteins in *Arabidopsis* (AtPCNA1 and AtPCNA2)[3]. PCNA interacts with DNA polymerase and serves as a general loading platform for many proteins involved in diverse processes occurring at chromatin[8]. Here we show that *ATXR5* and *ATXR6* are critical factors that act in a novel pathway to suppress DNA re-replication, especially in heterochromatic regions of the *Arabidopsis* genome.

To determine whether ATXR5 and/or ATXR6 have a role in DNA replication, we analysed the DNA content of leaf nuclei by flow cytometry. Leaves are well suited for assessing DNA replication defects as they undergo both mitosis and endoreduplication (genome duplication without mitosis), which is responsible for the widespread polyploidy observed in mature leaf tissue[9]. Nuclei were extracted from mature rosette leaves of wild-type Columbia (Col), *atxr5*, *atxr6* and *atxr5 atxr6* plants, stained with propidium iodide, and analysed for DNA content. Col, *atxr5* and *atxr6* all showed well-resolved populations of 2C, 4C, 8C, 16C and 32C nuclei (Fig. 1a). Thus, *atxr5* and *atxr6* single mutations do not have an impact on DNA replication in leaves. In contrast, the 8C and 16C peaks for *atxr5 atxr6* were much broader and skewed to the right, indicating that many 8C and 16C nuclei have higher DNA contents than the corresponding wild-type nuclei (Fig. 1a). The distribution of nuclei between endoreduplication levels was similar between wild type and *atxr5 atxr6* double mutants (Fig. 1b), suggesting that the primary defect in *atxr5 atxr6* plants is in the fidelity of S-phase progression rather than the number of rounds of endoreduplication. This is in contrast to mutations previously reported to affect the level of endoreduplication, but not S-phase fidelity[10].

Our previous work has shown that ~65% of *atxr5 atxr6* nuclei show significant decondensation of constitutive heterochromatin (that is, chromocentres) (Fig. 1c)[2]. Consistent with the flow cytometry results showing that the DNA content phenotype of *atxr5 atxr6* mutants is observed most strongly in 8C and 16C nuclei, we found, by microscopic analysis of sorted nuclei, that the heterochromatic decondensation defect was also more extreme in these higher ploidy nuclei (Fig. 1c, d). As a control, we also analysed the DNA content of *decrease in dna methylation1* (*ddm1*) plants, which also show very strong chromocentre decondensation defects, as well as reduced DNA methylation and massively reactivated transposons[11–13]. The flow cytometry profile from extracted nuclei from *ddm1-2* leaves was similar to that of wild-type plants (Fig. 1a). These results show that the aberrant flow cytometry profiles observed in *atxr5 atxr6* mutants are not simply a result of chromatin decondensation defects and/or transposon derepression.

To test the hypothesis that there is indeed extra DNA in *atxr5 atxr6* mutants, we used an Illumina Genome Analyser II to sequence genomic DNA from sorted nuclei (2C, 4C, 8C and 16C) of both wild-type and *atxr5 atxr6* plants. A total of 84.9 million uniquely mapping 36-nucleotide reads were mapped to the *Arabidopsis thaliana* genome, allowing up to two mismatches. We examined the distribution of genomic DNA across all chromosomes by plotting the density of reads in non-overlapping 100-kilobase (kb) bins. For wild type, the ratio of 4C, 8C and 16C to 2C sequence reads was uniform across the genome, showing that the genome is uniformly endoreduplicated in wild-type *Arabidopsis* (Fig. 1e). However, in *atxr5 atxr6* mutants, we observed an enrichment of reads in the pericentromeric heterochromatin in 4C,

[1]Department of Biology, Indiana University, 915 East Third Street, Bloomington, Indiana 47405, USA. [2]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, California 90095, USA. [3]Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, California 90095, USA. [4]Centro de Biologia Molecular Severo Ochoa, Consejo Superior de Investigaciones Cientificas, Universidad Autonoma de Madrid, Nicolas Cabrera 1, Cantoblanco, Madrid 28049, Spain.
*These authors contributed equally to this work.

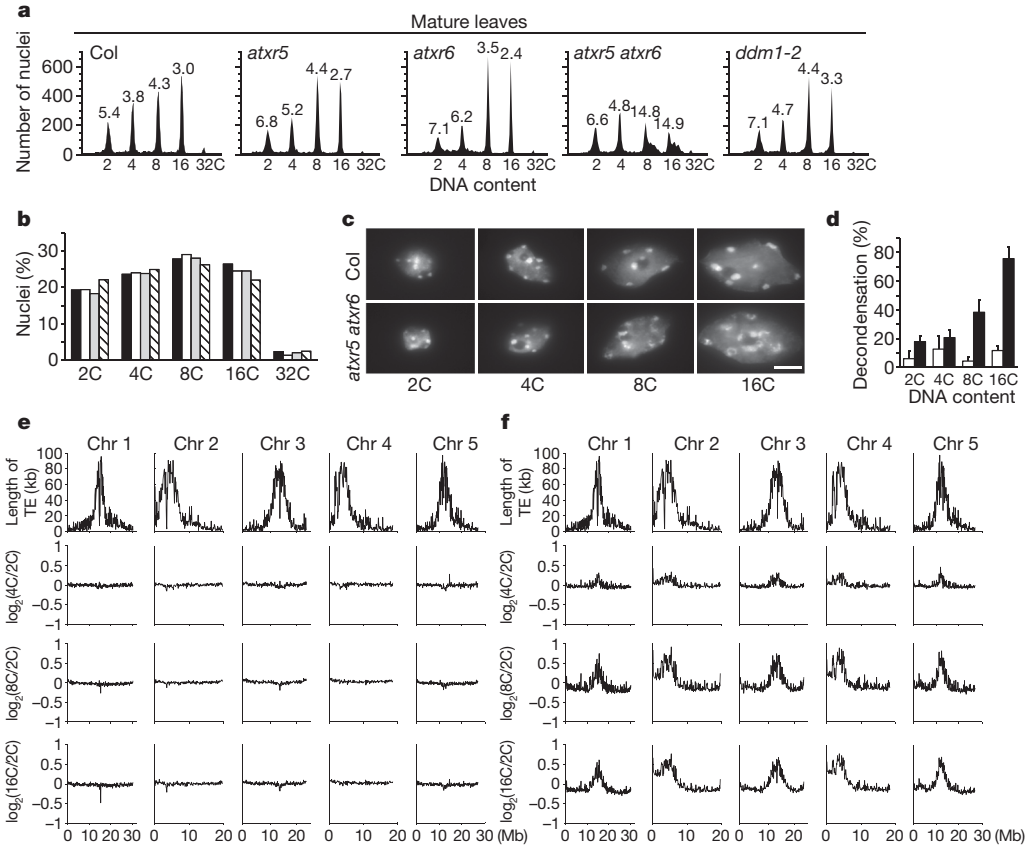

Figure 1 | Heterochromatic DNA is over-produced in *atxr5 atxr6* mutants. **a**, Flow cytometry profiles of Col, *atxr5*, *atxr6*, *atxr5 atxr6* and *ddm1-2* plants. Three-thousand gated events are plotted. The number above each peak (robust CV) indicates the number of fluorescence intensity units that enclose the central 68% of nuclei for that endoreduplication level. **b**, Quantification of nuclei at each ploidy level for samples in panel **a**; Col, black; *atxr5*, white; *atxr6*, grey; *atxr5 atxr6*, crosshatched. **c**, 4′,6-diamidino-2-phenylindole (DAPI) staining of sorted nuclei from Col and *atxr5 atxr6* leaves. Scale bar, 10 μm. **d**, Chromocentre decondensation occurs mainly in 8C and 16C nuclei. Thirty nuclei of each ploidy level from three biological replicates were analysed. White bars represent wild type and black bars represent *atxr5 atxr6*. Error bars indicate one standard deviation. **e**, DNA is replicated uniformly in wild-type nuclei during endoreduplication. The $\log_2$ ratios of genomic DNA Illumina reads from wild-type 4C versus 2C, 8C versus 2C and 16C versus 2C are plotted across the chromosomes in 100-kb-sliding windows. Plots of transposable element (TE) abundance (kb of transposon sequence per 100 kb genomic DNA) indicate pericentromeric regions. **f**, Similar analysis with *atxr5 atxr6* mutants showing an increased proportion of reads in pericentromeric heterochromatin in higher ploidy nuclei.

8C and 16C compared to 2C, indicating that heterochromatin is over-replicated in 4C, 8C and 16C nuclei compared to 2C nuclei (Fig. 1f), and that the over-replication is more severe in nuclei with higher ploidy levels.

A comparison of the distribution of reads from *atxr5 atxr6* mutants with wild type showed that even in the 2C nuclei, *atxr5 atxr6* mutants show over-replication of pericentromeric heterochromatin, although to a lower extent than in nuclei of higher ploidy levels (Supplementary Fig. 1a). Relative to wild type, *atxr5 atxr6* mutants showed a 2.9%, 11.0%, 29.1% and 28.4% increase in reads mapping to pericentromeric heterochromatin in 2C, 4C, 8C and 16C nuclei, respectively (Supplementary Fig. 1b). Sites of over-replication were well correlated at the different ploidy levels. For instance, the Pearson correlation between 8C and 16C *atxr5 atxr6* nuclei was 0.84, indicating that the same sites are over-replicating (Supplementary Fig. 1c). These results show that *atxr5 atxr6* mutants show over-replication of pericentromeric heterochromatin in both mitotic and endocycling cells, with progressively stronger defects observed in nuclei with higher ploidy levels.

To examine over-replication in *atxr5 atxr6* mutants at higher resolution, sequence reads were grouped and analysed in 200-base-pair non-overlapping bins. We found that over-replication of pericentromeric heterochromatin is the result of the over-replication of many densely spaced, but distinct, loci (Fig. 2a). We also observed localized over-replication in small regions of the euchromatic arms of chromosomes (Fig. 2b). These small regions of over-replication were highly enriched in transposons and other repeat elements. Using the BLOC algorithm (see Methods)[14], we identified 407 sites of over-replication in the arms of *atxr5 atxr6* chromosomes (Supplementary Table 1). The over-replicating regions were relatively small: 94% of regions were smaller than 25 kb, with a median size of 10.4 kb (Supplementary Fig. 2a). Most (80%) overlapped with previously defined H3K9me2 regions, a mark that is strongly correlated with DNA methylation and gene silencing[6] (Supplementary Fig. 2b). Thus, the regions that over-replicate in *atxr5 atxr6* mutants primarily consist of transposons and silent elements of the *Arabidopsis* genome. Over-replication was confirmed by performing quantitative polymerase chain reaction (qPCR) on defined sites (Supplementary Fig. 3). Elements that are transcriptionally reactivated in *atxr5 atxr6* mutants (*TSI*, *Ta3*, *CACTA*)[2] were found to be over-replicated, indicating a positive correlation between transposon reactivation and over-replication.

Re-replication is a well-known mechanism by which DNA is known to over-replicate and results when DNA replication is initiated from an origin multiple times during a single S phase[1]. Presumably because recently replicated chromatin is less compact, secondary replication forks move faster than primary forks, and collisions of the multiple forks result in successively smaller fragments of DNA reiteratively produced from the origin[15] (Fig. 2c). This model predicts that sequences in the centre of the origin will be the most highly over-replicated and that over-replication should drop off symmetrically on

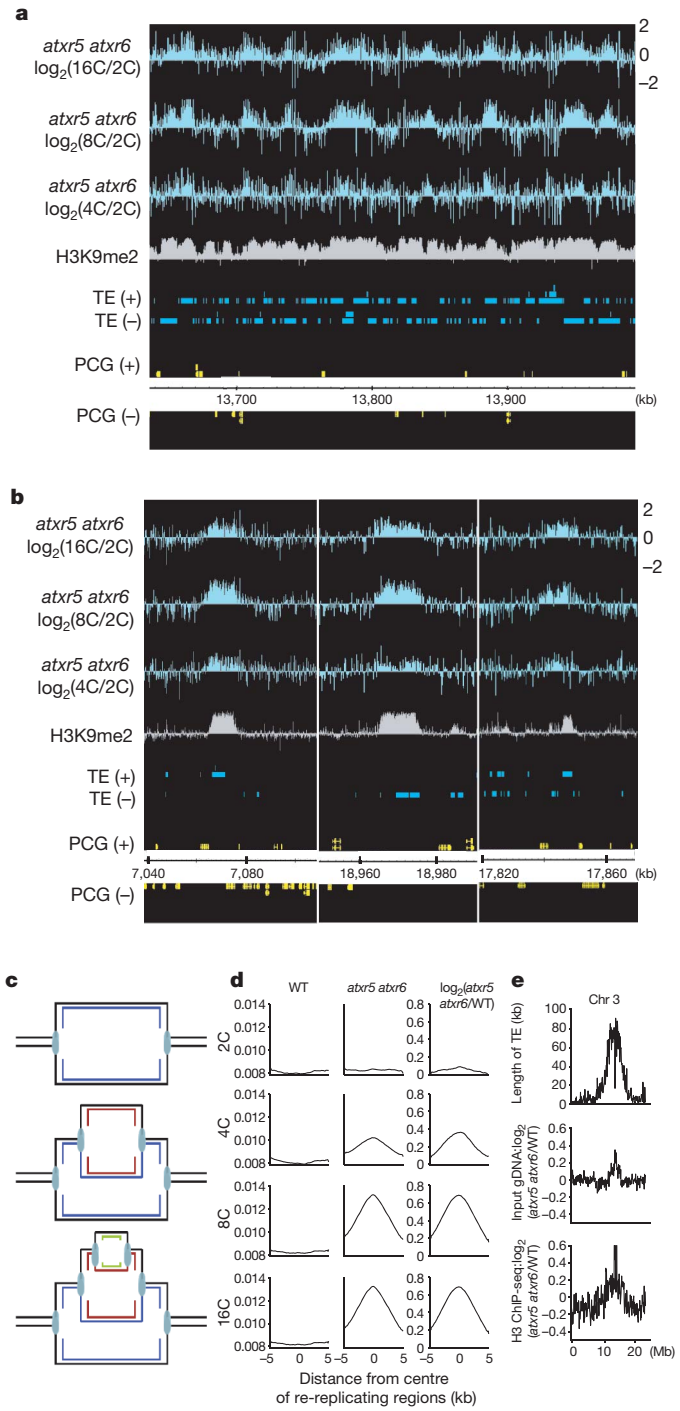

Figure 2 | Increased heterochromatic DNA in *atxr5 atxr6* mutants is consistent with re-replication of chromatin. **a**, Genome browser view of a region of pericentromeric heterochromatin. Pericentromeric heterochromatin contains densely spaced, ~10-kb over-replicating sites. Data are represented as $\log_2$ ratios (16C/2C, 8C/2C or 4C/2C) in 200-bp bins. H3K9me2 microarray data[6], TAIR8 protein-coding gene (PCG) and transposable element (TE) tracks are also shown on the plus (+) or minus (−) strand of the genome. **b**, Genome browser view of examples of over-replication in the arms of chromosomes. Three over-replicating regions are shown. **c**, Model for DNA re-replication (ref. 22). **d**, Distribution of Illumina reads in re-replicating regions. Plots of the average number of sequence reads ±5 kb relative to the centre of over-replicating regions in *atxr5 atxr6* mutants, wild type, or the *atxr5 atxr6* mutants/wild type $\log_2$ ratio (plotted in 100-bp bins). **e**, Histone content in re-replicating regions is higher in *atxr5 atxr6* mutants. Log_2 ratios of H3 ChIP-seq reads and input genomic DNA reads in *atxr5 atxr6* mutants relative to wild type, plotted over chromosome 3 in 100-kb sliding windows.

either side of the origin. To determine if over-replication in the *atxr5 atxr6* mutant is consistent with re-replication, plots of sequencing reads averaged over the over-replicated regions were generated (Fig. 2d). In contrast to wild type, in which sequencing reads were uniformly distributed, *atxr5 atxr6* mutants showed a bilaterally symmetrical distribution of reads, with the highest density of reads in the centre of the over-replicated regions (Fig. 2d). These results indicate that the extra DNA in *atxr5 atxr6* mutants is a result of repeated replication from defined sites.

We next examined whether chromatin or naked DNA is being re-replicated in *atxr5 atxr6* mutants. To test this, we performed chromatin-immunoprecipitation of unmodified histone H3 followed by Illumina sequencing (ChIP-seq) on wild type and *atxr5 atxr6* mutants. Compared with wild type, H3 ChIP-seq reads in *atxr5 atxr6* mutants were enriched in the pericentromeric heterochromatin to a similar extent as was the input genomic DNA (Fig. 2e and Supplementary Fig. 2c). This result indicates that chromatin (DNA and associated histones) is re-replicated in *atxr5 atxr6* mutants. Our data also indicate that the re-replicated DNA is properly methylated. If the re-replicated DNA was unmethylated, the per cent methylation in *atxr5 atxr6* mutants would be predicted to be lower than in wild type. However, we have previously shown that the per cent DNA methylation in *atxr5 atxr6* mutant leaves (where re-replication was observed) is the same as in wild type[2], which suggests that the re-replicated DNA is properly methylated. Furthermore, to determine whether re-replicated DNA was stably associated with the chromosomes, we performed qPCR on size-fractionated DNA (Supplementary Fig. 4). We found that the extra DNA could be detected in the high-molecular mass DNA fraction, indicating that at least part of the re-replicated DNA is stably associated with the chromosome. Being associated with chromosomes, rather than being extrachromosomal fragments, may help to explain the stability of re-replicating DNA fragments present in 3–4-week-old leaf cells.

Because ATXR5 and ATXR6 catalyse H3K27me1 (ref. 2), we wanted to examine whether the spatial distribution of H3K27me1 overlaps with re-replicating regions. Immunolocalization indicates that H3K27me1 is a heterochromatic mark enriched in chromocentres[2,16–18]. A detailed global map of H3K27me1, however, has not been reported. We therefore profiled H3K27me1 genome-wide using ChIP-seq. Consistent with the re-replication of pericentromeric heterochromatin in *atxr5 atxr6* (Fig. 1f), we found that H3K27me1 was strongly enriched in pericentromeric heterochromatin (Fig. 3a). We also observed H3K27me1 in the coding regions of protein-coding genes and found that the amount of H3K27me1 was anticorrelated with gene expression levels (Fig. 3b and Supplementary Fig. 5a). Together, these results support a role for H3K27me1 in gene silencing.

To gain additional evidence for a correlation between H3K27me1 and re-replication in *atxr5 atxr6* mutants, we examined the dispersed re-replicating regions in the arms of the chromosomes. We found that H3K27me1 ChIP-seq reads were significantly enriched in these regions compared to randomly selected control regions (permutation test, $P < 10^{-6}$) (Fig. 3c). In addition, plots of the ratio of H3K27me1 to H3 ChIP-seq reads averaged over these re-replicating regions showed strong enrichment of H3K27me1, confirming a positive correlation of H3K27me1 with sites that re-replicate in *atxr5 atxr6* mutants (Fig. 3d and Supplementary Fig. 5b).

Given that H3K27me1 levels correlate with the re-replicated regions of *atxr5 atxr6* mutants, an interesting question concerns the mechanism by which the spatial distribution of H3K27me1 is established. ATXR5 and ATXR6 both contain PHD domains, which have been shown in multiple species to mediate interactions with methylated or unmethylated forms of histone H3[19]. We performed *in vitro* binding assays with various H3 peptides using GST-tagged PHD domains of ATXR5 and ATXR6. The PHD domains of ATXR5 and ATXR6 bound strongly to an unmethylated peptide corresponding to amino acids 1–21 of H3 (Fig. 3e). This binding was unaffected by mono-, di-, or trimethylation at H3K9; however, binding was

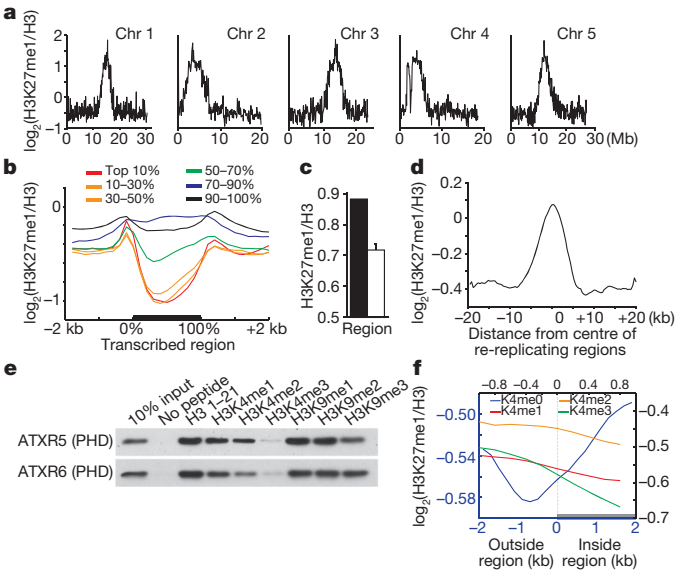

Figure 3 | Genome-wide mapping of H3K27me1 and anticorrelation with H3K4 methylation. **a**, H3K27me1 is enriched in heterochromatin. The $\log_2$ ratios of H3K27me1 reads to H3 ChIP-seq reads in wild type are plotted across the chromosomes (1 to 5) in 100-kb sliding windows. **b**, H3K27me1 is anticorrelated with gene expression level. H3K27me1 ChIP-seq reads normalized to H3 ChIP-seq reads averaged over TAIR8 protein-coding genes. The bodies of genes are scaled. Three-week-old wild-type plants were used for both ChIP-seq and RNA-seq. **c**, H3K27me1 is significantly enriched at sites of re-replication in the arms. Reads per base pair in re-replicating regions were calculated for both H3K27me1 and H3 ChIP-seq reads, and the ratio was calculated (black bar). Random regions with a similar distribution as re-replicating regions were generated 100,000 times and the same calculation was performed. The mean value obtained from random regions are shown (white bar) and the error bars represent the standard deviation. **d**, H3K27me1 is enriched in over re-replicating regions. The $\log_2$ ratio of H3K27me1 to H3 reads is plotted $\pm$20 kb relative to the centre of re-replicating regions of *atxr5 atxr6* mutants. Data were plotted in 400-bp bins and smoothed by taking the moving average over six bins. **e**, Pull-down assay using purified GST-tagged PHD domains of ATXR5 and ATXR6 and biotinylated H3 peptides with different methylated lysines. Interaction between the peptides and the GST–PHD domains was visualized by western blot using a GST antibody. **f**, Analysis of the relationship between H3K27me1 and H3K4 methylation. The $\log_2$ ratio of H3K27me1 to H3 is plotted over the boundaries of all H3K4me0/-me1/-me2/-me3 regions in the genome. Data are shown in 200-bp bins, and smoothed by taking the moving average over $\pm$2 bins. The scale for the plots over H3K4me0 is in blue, and the scale for the others is in black.

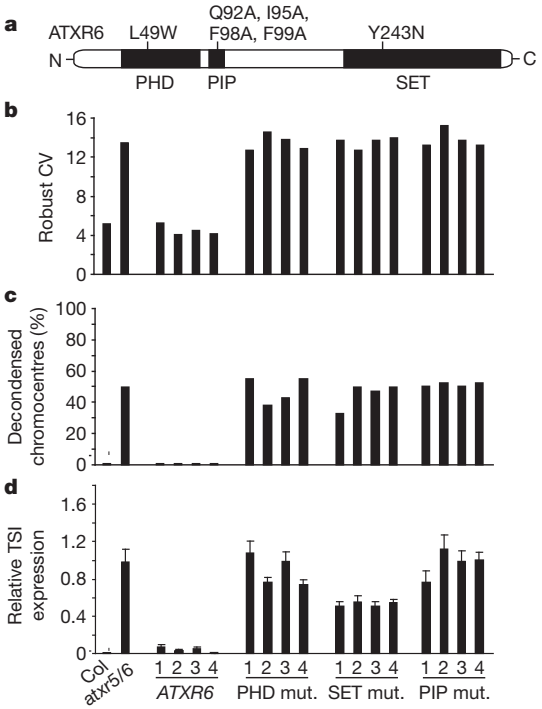

Figure 4 | Functional PHD and SET domains and the PIP motif are required for the regulation of DNA replication by *ATXR6*. **a**, Structure of ATXR6. The domains (below) and point mutations (above) made to generate *ATXR6* mutants are represented. **b**–**d**, Normal DNA replication as indicated by robust CV (**b**), chromatin condensation (**c**) and TSI gene silencing (**d**) is rescued in transgenic *atxr5 atxr6* plants expressing wild-type *ATXR6*, but not PHD, SET, or PIP mutants. All phenotypes were scored on the same four representative transgenic lines ($n > 20$) generated from each construct. Error bars indicate one standard deviation.

strongly reduced by increasing levels of H3K4 methylation. Thus, the PHD domains of ATXR5 and ATXR6 bound most strongly to H3 unmethylated at K4 (H3K4me0). Consistent with the hypothesis that binding of the ATXR5 and ATXR6 PHD domains to H3K4me0 chromatin is helping to guide H3K27 monomethylation activity, we observed a strong anticorrelation between H3K4 methylation[20] and H3K27me1 within genes and in the genome at large (Fig. 3f and Supplementary Fig. 5c).

Because loss of ATXR5 and ATXR6 leads to lower levels of H3K27me1 (ref. 2), it is possible that depletion of this mark is causing re-replication in *atxr5 atxr6* mutants. One prediction from this model is that the PHD and SET domains of ATXR5 and ATXR6 would be essential to prevent re-replication, as they are responsible for binding and methylating H3, respectively (Fig. 4a). To investigate this, we first created a genomic construct that expresses *ATXR6* under its own promoter and confirmed that it can rescue (>95% of T1 transformed plants analysed) the re-replication phenotype of *atxr5 atxr6* mutant plants (Fig. 4b). We then made PIP-, PHD- and SET-mutant *ATXR6* constructs by inserting point mutations designed to disrupt the activity of each functional element (Fig. 4a). Yeast-two-hybrid analysis and *in vitro* histone-peptide-binding and methyltransferase assays were used to confirm disruption of the PIP-motif, PHD-domain and SET-domain activities, respectively (Supplementary Fig. 6). Analysis of T1 plants transformed with each of the mutated *ATXR6* constructs showed that the re-replication phenotype was never rescued by constructs containing the mutated PIP motif, PHD domain or SET domain (Fig. 4b) ($n > 20$). These results show that the PIP motif, PHD domain and SET domain are all required for ATXR6 activity and indicate that depletion of H3K27me1 in the *atxr5 atxr6* double mutant is probably responsible for the re-replication phenotype. Consistent with this interpretation, we found that the restoration of H3K27me1 levels also required the wild-type PIP motif, PHD domain and SET domain (Supplementary Fig. 7). Furthermore, only the wild-type construct rescued chromatin decondensation and loss of gene silencing defect seen in *atxr5 atxr6* mutants (Fig. 4c, d). These results indicate that the three functional elements of ATXR6 contribute to the prevention of re-replication, chromatin decondensation and loss of gene silencing.

Our results indicate that *ATXR5* and *ATXR6* are components of a novel pathway required to suppress re-replication in *Arabidopsis*. Notably, most of the re-replicating sites in *atxr5 atxr6* mutants correspond to silent heterochromatin, which is composed mostly of transposon sequences. It is tempting to speculate that the ATXR5/ATXR6 system may have evolved to suppress excess DNA replication of transposon sequences that would otherwise result in transposon reactivation. Conversely, transposons are remarkable in requiring both the typical repressive modifications such as H3K9me2 and DNA methylation, as well as the novel ATXR5/ATXR6 H3K27me1 pathway for transcriptional suppression.

METHODS SUMMARY

FACS was used to generate flow cytometry profiles of leaf nuclei from Col and *atxr5 atxr6* and to sort nuclei based on endoreduplication level (2C, 4C, 8C and 16C). Genomic DNA isolated from sorted nuclei was sequenced using an Illumina Genome Analyser II. SeqMap[21] was used to map sequencing reads to the *Arabidopsis* genome. DNA from ChIP using H3 and H3K27me1 antibodies was sequenced and analysed in a similar fashion. *In vitro* binding assays used biotinylated H3 peptides (Millipore, Billerica).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 March; accepted 24 June 2010.
Published online 14 July 2010; corrected 19 August 2010 (see full-text HTML version for details).

1. Arias, E. E. & Walter, J. C. Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes Dev.* **21,** 497–518 (2007).
2. Jacob, Y. *et al.* ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nature Struct. Mol. Biol.* **16,** 763–768 (2009).
3. Raynaud, C. *et al.* Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in *Arabidopsis*. *Plant J.* **47,** 395–407 (2006).
4. Jackson, J. P., Lindroth, A. M., Cao, X. & Jacobsen, S. E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416,** 556–560 (2002).
5. Malagnac, F., Bartee, L. & Bender, J. An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J.* **21,** 6842–6852 (2002).
6. Bernatavichute, Y. V., Zhang, X., Cokus, S., Pellegrini, M. & Jacobsen, S. E. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis* thaliana. *PLoS ONE* **3,** e3156 (2008).
7. Obayashi, T., Hayashi, S., Saeki, M., Ohta, H. & Kinoshita, K. ATTED-II provides coexpressed gene networks for *Arabidopsis*. *Nucleic Acids Res.* **37,** D987–D991 (2009).
8. Moldovan, G. L., Pfander, B. & Jentsch, S. PCNA, the maestro of the replication fork. *Cell* **129,** 665–679 (2007).
9. Galbraith, D. W., Harkins, K. R. & Knapp, S. Systemic endopolyploidy in *Arabidopsis thaliana*. *Plant Physiol.* **96,** 985–989 (1991).
10. Caro, E., Desvoyes, B., Ramirez-Parra, E., Sanchez, M. P. & Gutierrez, C. Endoreduplication control during plant development. *SEB Exp. Biol. Ser.* **59,** 167–187 (2008).
11. Jeddeloh, J. A., Stokes, T. L. & Richards, E. J. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genet.* **22,** 94–97 (1999).
12. Fransz, P., ten Hoopen, R. & Tessadori, F. Composition and formation of heterochromatin in *Arabidopsis thaliana*. *Chromosome Res.* **14,** 71–82 (2006).
13. Soppe, W. J. *et al.* DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO J.* **21,** 6549–6559 (2002).
14. Pauler, F. M. *et al.* H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res.* **19,** 221–233 (2009).
15. Gomez, M. Controlled rereplication at DNA replication origins. *Cell Cycle* **7,** 1313–1314 (2008).
16. Fuchs, J., Demidov, D., Houben, A. & Schubert, I. Chromosomal histone modification patterns—from conservation to diversity. *Trends Plant Sci.* **11,** 199–208 (2006).
17. Lindroth, A. M. *et al.* Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with *CHROMOMETHYLASE3*. *EMBO J.* **23,** 4146–4155 (2004).
18. Mathieu, O., Probst, A. V. & Paszkowski, J. Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis*. *EMBO J.* **24,** 2783–2791 (2005).
19. Musselman, C. A. & Kutateladze, T. G. PHD fingers: epigenetic effectors and potential drug targets. *Mol. Interv.* **9,** 314–323 (2009).
20. Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M. & Jacobsen, S. E. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* **10,** R62 (2009).
21. Jiang, H. & Wong, W. H. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* **24,** 2395–2396 (2008).
22. Davidson, I. F., Li, A. & Blow, J. J. Deregulated replication licensing causes DNA fragmentation consistent with head-to-tail fork collision. *Mol. Cell* **24,** 433–443 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Lambert and D. Galbraith for assistance with flow cytometry; Y. Bernatavichute for assistance with ChIP experiments; and M. Pellegrini and S. Cokus for advice on data analyses. Y.J. was supported by a fellowship from Le Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT). S.F. is a Howard Hughes Medical Institute Fellow of the Life Science Research Foundation. Research in the Michaels' laboratory was supported by grants from the National Institutes of Health (GM075060), the Indiana METACyt Initiative of Indiana University, and the Lilly Endowment, Inc. C.G. was supported by grants from the Spanish Ministry of Science and Innovation (BFU2009-9783 and CSD2007-57B). S.E.J. is an investigator of the Howard Hughes Medical Institute.

Author Contributions S.D.M., S.E.J. and C.G. directed the research. Y.J., H.S., C.L., S.F., L.Z., E.C. and C.H. performed experiments. H.S. analysed data. H.S., Y.J., S.E.J. and S.D.M. prepared the manuscript.

Author Information Sequencing files have been deposited at GEO (accession codes GSE22411 and GSE21673). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.D.M. (michaels@indiana.edu) or S.E.J. (jacobsen@ucla.edu).

METHODS

Plant material. *atxr5* (SALK_130607) and *atxr6* (SAIL_240_H01) in the Columbia genetic background were obtained from the *Arabidopsis* Biological Resource Center. The *ddm1-2* seeds have been described previously[11,23]. Plants were grown under cool-white fluorescent light (~100 μmol $m^{-2} s^{-1}$) under long-day conditions (16 h of light followed by 8 h of darkness).

Flow cytometry. Plant tissue was chopped with a razor blade in 500 μl of Galbraith buffer (45 mM $MgCl_2$, 20 mM MOPS, 30 mM sodium citrate, 0.1% Triton X-100) containing 20 μg of RNase A. The lysate was filtered through a 40 μm cell strainer (BD Falcon) and propidium iodide was added to a final concentration of 20 μg ml^{-1}. For mature leaves, nuclei were extracted from rosette leaves 3 and 4 from 4-week-old plants; for young leaves, nuclei were extracted from rosette leaves 1 and 2 from 9-day-old plants. Flow cytometry profiles were obtained on a BD FACSCalibur (Becton Dickinson) based upon propidium iodide and 90° side angle scatter on a logarithmic scale due to the small size of the nuclei. Analysis was performed with CellQuest Pro software (Becton Dickinson). For nuclei sorting, 2.5 g of mature rosette leaves were collected from 4-week-old plants, chopped in 5 ml of Galbraith buffer containing 200 μg of RNase A, filtered and stained with propidium iodide. The procedure for sorting was based on parameters similar to the FACSCalibur on a BD FACS Aria II, using an 85 μm nozzle with sheath pressure at 35 p.s.i. Analysis was performed with FACSDiva version 6.1.1 (Becton Dickinson). For sequencing, 400,000 nuclei of each ploidy (2C, 4C, 8C and 16C) were isolated from Col and *atxr5 atxr6* leaves at a threshold rate of approximately 500 events s^{-1}, a flow rate of 2.5 and a variable sort rate dependent upon the peak sorted. Genomic DNA was extracted from the sorted nuclei using the PicoPure DNA Extraction Kit (Molecular Devices) according to the manufacturer's recommendations.

For analysis of sorted nuclei by microscopy, 0.25 g of leaves were fixed for 20 min in 4% formaldehyde in Tris buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 100 mM NaCl) and washed with Tris buffer 2 × 10 min. The leaves were then chopped in Galbraith buffer, filtered and stained with 4 μg ml^{-1} DAPI. Two-thousand nuclei of each ploidy were sorted directly onto microscope slides. Coverslips were then mounted using Vectashield mounting medium with DAPI (Vector Laboratories) and sealed with clear nail polish. Thirty nuclei for each ploidy type were analysed for three biological samples. Flow cytometry and nuclei sorting were performed at the Indiana University Flow Cytometry Core Facility.

cDNA synthesis, real-time PCR, protein expression and purification, and immunofluorescence. These procedures were performed as described previously[2]. Microscopy was performed at the Indiana University-Bloomington Light Microscopy Imaging Center.

Constructs. To make the *ATXR6* constructs used to transform *A. thaliana*, the promoter (293 bp upstream of start codon) and gene (exons and introns) were amplified by PCR and cloned into pENTR/D (Invitrogen). Point mutations in the PHD domain, SET domain and the PIP motif were made using pENTR/D-ATXR6 as template by site-directed mutagenesis. For the PHD-domain mutant, leucine 49 was replaced with tryptophan (L49W). This mutation is based on the structure of the PHD domain of the mammalian protein BHC80, which was shown to rely on an equivalent residue (M502) for preferentially binding H3K4me0 (ref. 24). To disrupt the function of the SET domain, we changed tyrosine 243 to asparagine (Y243N). This tyrosine is part of the conserved YXG motif of SET-domain proteins, which is responsible for orienting the ε-amino group for methyl transfer to occur on lysine[25]. An equivalent mutation (Y655N) in *Drosophila* Enhancer of zeste (E(z)) was shown to abolish H3K27 methylation without affecting folding of the protein[26]. Finally, glutamine 92, isoleucine 95, phenylalanine 98 and phenylalanine 99 of the conserved PIP motif (QTKIIDFF) of ATXR6 were replaced with alanine. The resulting ATXR6 constructs were subcloned first into pEG302 (ref. 27) using Gateway technology to acquire a C-terminal Flag-epitope sequence, then into pMDC30 (ref. 28) by restriction digest of the pEG302 vectors with SpeI and SbfI.

pGEX-6P was used to clone the PHD domains (amino acids 25–103, ATXR6; 57–133, ATXR5) and PHD-SET domains (25–349, ATXR6) for the *in vitro* binding and methyltransferase assays, respectively. The point mutations in the PHD (L49W) and SET (Y243N) domains were introduced in the pGEX vectors by site-directed mutagenesis.

For the yeast two-hybrid assay, the coding sequences of *ATXR6*, *ATXR6(pip)* and *AtPCNA1* (At1g07370) were cloned into pENTR/D, then subcloned into pDEST32 or pDEST22 using the Gateway system.

Histone peptide binding and histone methyltransferase assays. Biotinylated peptides were obtained from Millipore. The procedure for this assay has been described previously[29]. The binding buffer contained 250 mM NaCl. Histone methyltransferase assays were preformed as previously described[2].

Yeast two-hybrid assay. pDEST22-AtPCNA1 was co-transformed with each of the two pDEST32-ATXR6 vectors in *Saccharomyces cerevisiae* (strain MaV203, Invitrogen). As controls, pDEST22-AtPCNA1 and pDEST32-ATXR6/ATXR6(pip) were also used for co-transformation with empty pDEST32 and pDEST22 vectors, respectively. Five independent colonies from each transformation were selected and streaked on SC –Leu –Trp plates. The colonies were then replica plated on SC –Leu –Trp plates containing various concentrations (25, 50 or 100 mM) of 3-amino-1,2,4-triazole (3AT), and grown for 3–4 days at 30 °C.

Chromatin immunoprecipitation. ChIP was performed on crosslinked 3-week-old plants as previously described[6] using anti-H3 (Abcam 1791) and anti-H3K27me1 (Upstate 07-448) antibodies.

Quantitative PCR assays on genomic DNA. Fifty-thousand 16C nuclei were prepared from rosette leaves of 28-day-old plants (Col and *atxr5 atxr6*), as described above. For sorting, the same 16N gate size was used for both samples. Genomic DNA was extracted from the sorted nuclei using the PicoPure DNA Extraction Kit (Molecular Devices) according to the manufacturer's recommendations. qPCR was performed using Brilliant II SYBR green QPCR master mix (Agilent), also according to the manufacturer's recommendations.

Quantitative PCR assays on gel separated genomic DNA. Four 3-week-old leaves were ground and extracted in 400 μl extraction buffer (sorbitol 350 mM, Tris-HCl pH 7.5 100 mM, EDTA 5 mM) followed by lysis in 400 μl of lysis buffer (Tris-HCl pH 7.5 200 mM, EDTA 50 mM, NaCl 2 M, CTAB 2%) plus 27 μl of 10% sarkosyl. After a 30-min incubation at 65 °C, DNA was extracted with chloroform/isoamyl alcohol, precipitated and re-suspended in 100 μl of extraction buffer including RNase A (Qiagen). Around 10 μg of this DNA was run in a 1% low melting point agarose gel (Lonza) for both Col0 and *atxr5 atxr6* and the band for the genomic DNA was cut out of the gel (BIG) along with the rest of the lane below that band (SMALL, DNA <12 kb).

For the recovery of the BIG fraction of DNA from the gel we diluted it five times with TE (Tris pH 8 plus EDTA) and incubated at 65 °C until the agarose melted. We then performed serial extractions with phenol, phenol/chloroform and chloroform followed by isopropanol precipitation and resuspension in 50 μl of TE.

For the recovery of the SMALL fraction of DNA from the gel we used Qiagen gel extraction kit and followed the manufacturer's instructions. DNA was eluted in 50 μl of TE. qPCR was performed using the IQ -SYBR Green Supermix from Bio-Rad, a Stratagene MX3005P qPCR system.

Illumina library preparation. Illumina libraries for genomic DNA extracted from sorted nuclei and ChIP samples were made following the manufacturer's instructions. The libraries were sequenced using Illumina Genome Analyser II following manufacturer instructions, producing reads of 36 bp in length.

Illumina read alignment and analysis. Sequenced reads were based-called using the standard Illumina software. We used SeqMap (ref. 21) to align the reads to the *Arabidopsis thaliana* genome, allowing up to two mismatches. Identical reads were collapsed into single reads. We used two different mapping strategies for all of our data sets: (1) keeping reads that uniquely map to the genome and (2) remapping the reads that map to multiple locations in the genome, giving each read a weight. In (2), if a read mapped to x places in the genome, the read was given a score of $1/x$, hence uniquely mapping reads got a score of 1. Mapping strategy (2) was necessary because the majority of re-replicating regions were present in heterochromatic regions, and were used to define regions of re-replication. Because many heterochromatic regions are highly repetitive, we found that uniquely mapping the reads results in no signal in those regions.

Because 36-bp reads represented the ends of the DNA fragments in the library, for the analysis, we extended the read so that the data represent the actual DNA fragments of the libraries. The lengths of extensions were determined based on the distribution of the sizes of DNA fragments in the library. Each base pair of the read was given a score of 1 (or $1/x$). Therefore if a certain nucleotide in the genome had x uniquely mapping fragments overlapping, that nucleotide got a score of x. Data in (1) were normalized to total number of uniquely mapping reads in wild type, whereas in (2), data were normalized to the number of uniquely mapping reads plus the number of reads that map to multiple locations.

Re-replicating regions were defined by using BLOC[14]. Scores for *atxr5 atxr6* $\log_2$(16C/2C) were calculated in 60-bp bins, Z-score transformed, and then an average Z-score cutoff of 0.3 was applied for BLOC.

23. Vongs, A., Kakutani, T., Martienssen, R. A. & Richards, E. J. *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**, 1926–1928 (1993).
24. Lan, F. *et al.* Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448**, 718–722 (2007).
25. Dillon, S. C., Zhang, X., Trievel, R. C. & Cheng, X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* **6**, 227 (2005).

26. Joshi, P. *et al.* Dominant alleles identify SET domain residues required for histone methyltransferase of Polycomb repressive complex 2. *J. Biol. Chem.* **283,** 27757–27766 (2008).
27. Earley, K. W. *et al.* Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45,** 616–629 (2006).
28. Curtis, M. D. & Grossniklaus, U. A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta[w]. *Plant Physiol.* **133,** 462–469 (2003).
29. Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442,** 96 (2006).