Supplementary discussion

Structural analysis of the interaction between RcATXR5 and the H3.1 peptide

The structure of RcATXR5 comprises an all α-helix nSET followed by the SET domain. In addition to these domains, the C-terminus of the SET domain (352-374) folds as a long loop (L3) that funnels through another loop linking β8 and α6 to form the pseudo-knot structure. Similar to other protein lysine methyltransferases, this region of the SET domain is in close proximity to both the peptide substrate and cofactor. An electrostatic representation of the histone H3.1 binding cleft reveals that the region of RcATXR5 interacting with the N-terminus of the peptide is negatively charged and likely forming an optimal environment for binding Arg-26 of histone H3.1. Conversely, the region binding the C-terminus of histone H3.1 is predominantly hydrophobic with weak negatively and positively charged pockets in proximity of the e-amine of Lys-36 and the backbone carbonyl groups of Gly-33/Gly-34, respectively. The extent of the interaction supports our initial HKM assays showing that deletion of residues succeeding Lys-18 and Gly-34 is detrimental to the enzymatic activity of RcATXR5 (fig. S2).

Inspection of the histone H3.1 binding cleft of RcATXR5 reveals that three loops (L1-L3), along with residues found in β6 and β8, make extensive hydrophobic contacts and hydrogen bond interactions with histone H3.1. The α3-α4 loop (L1) is found in close proximity of the C-terminal end of H3.1 and makes several interactions with the backbone and side chains of Ala-31 and Thr-32 while L2, found between β5-β6, interacts with Ala-25 and Arg-26 at the N-terminus of H3.1. Finally, the third loop (L3) is found at the N-terminal end of the post-SET domain of RcATXR5 and includes residues G363, Y364, E365 and E367. The L3 loop folds back on top of histone H3.1 shielding the peptide from the solvent in a conformation analogous to a safety belt. Residues forming the safety belt make extensive contacts with the central region of the H3.1 peptide (residues 26-30) and help in maintaining the peptide tightly bound to RcATXR5. In contrast to the loosely bound Lys-23, the carbonyl groups of Gly-33 and Gly-34 engages in direct hydrogen bonds with the amide carbonyl and the -amine of Q217 and K331, respectively. In vicinity of this region, Val-35 is tightly maintained in a hydrophobic pocket composed of M263, V346 and T348.

The divergent binding mode of ATXR5 for histone H3

Our analysis of the H3.1 binding cleft of RcATXR5 revealed several unique structural features compared to other SET domains. First, RcATXR5-bound H3.1 does not form extensive hydrogen bond networks reminiscent of a β-strand as seen for other SET domains bound to H3 peptides such as SET8 and DIM-5. Instead, the H3.1 peptide binds in an L-shaped conformation with a sharp bend at Thr-32. Second, RcATXR5 maintains the peptide with residues present in the nSET, SET and cSET domains. The implication of nSET in ATXR5 is in clear contrast to all other SET domain proteins that only use residues found in the SET and cSET regions. Third, residues found at the C-terminus of the SET domain (referred to as cSET) fold as a short loop while the cSET regions of

SET8 and vSET fold as α-helices. Finally, while most of the SET domains of protein lysine methyltransferases, such as SET7/9, SET8 and DIM-5, employ predominantly an α-helix (coined as iSET) to achieve binding and selectivity of the lysine substrate, ATXR5 ensconces the histone H3.1 peptide in a shallow peptide binding cleft using predominantly three loops (L1 to L3). Overall, these comparative analyses indicate that the SET domain of ATXR5 binds histone H3 using a novel binding mode.

Supplementary Figures

Figure S1: Alignment of plant ATXR5/6 proteins. Protein sequences were obtained from the National Center for Biotechnology Information and correspond to the following accession numbers: (*Arabidopsis thaliana* ATXR5: NP_001078559.1; *Arabidopsis thaliana* ATXR6: NP_197821.1; *Ricinus communis* ATXR5: XP_002517234.1; *Ricinus communis* ATXR6: XP_002519163.1; *Oryza sativa*: BAC05613.1; *Picea sitchensis*: ADE75939.1; *Selaginella moellendorffii*: XP_002970367.1; *Physcomitrella patens*_1: XP_001776886.1; *Physcomitrella patens*_2: XP_001784397.1). Identical residues are dark-shaded; similar residues are light-shaded. A red asterisk indicates the amino acids composing the selectivity pocket for A31 of histone H3.1. PIP box: PCNA-interacting protein motif. NLS: Nuclear localization signal.

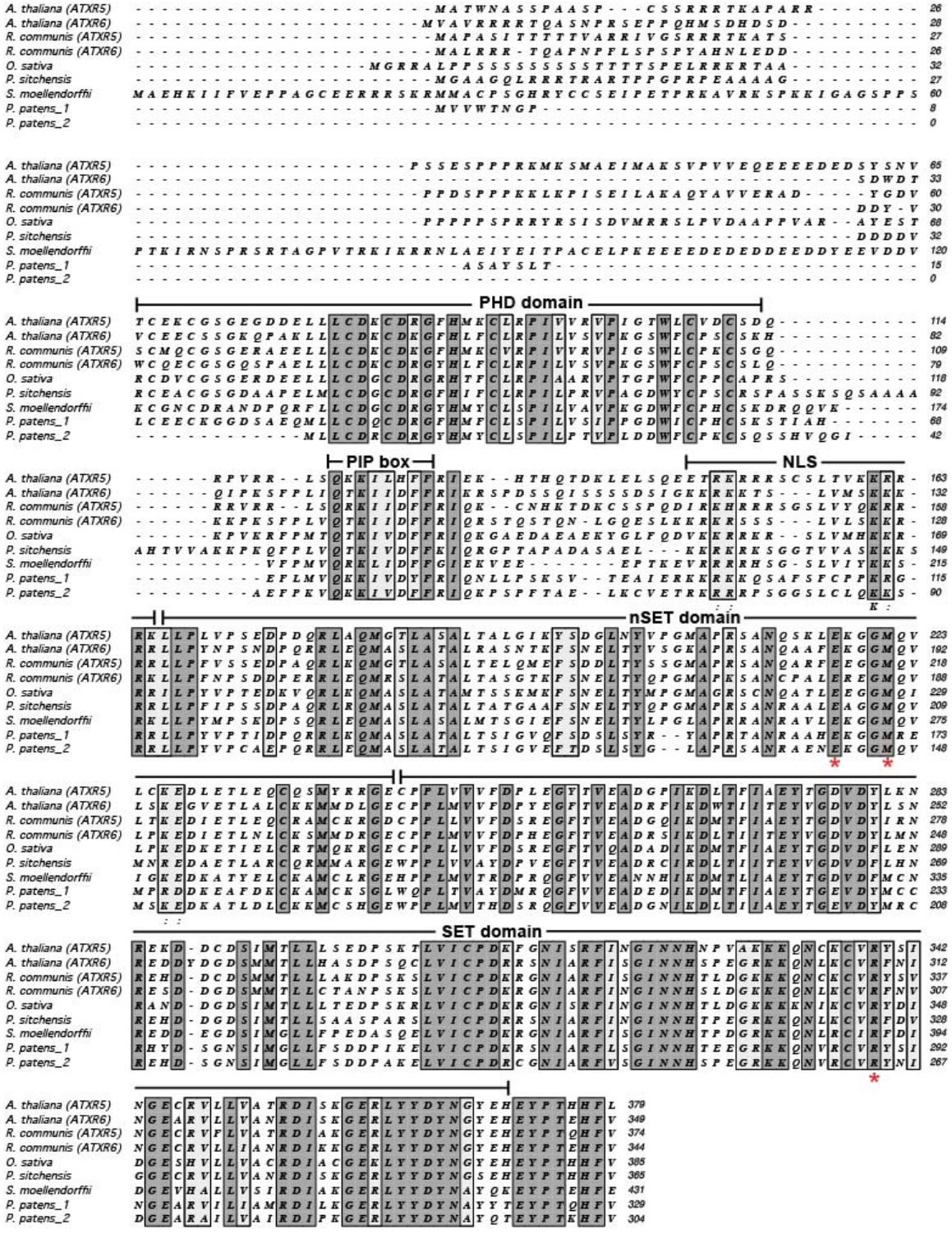

PHD domain
PIP box
NLS
nSET domain
SET domain

Figure S2: Defining the optimal substrate for ATXR5. HKM assays performed with peptides spanning different boundaries around the methylation site (K27) of RcATXR5. Values represent averages of two independent experiments performed in triplicate. Standard deviations are represented by the error bars.

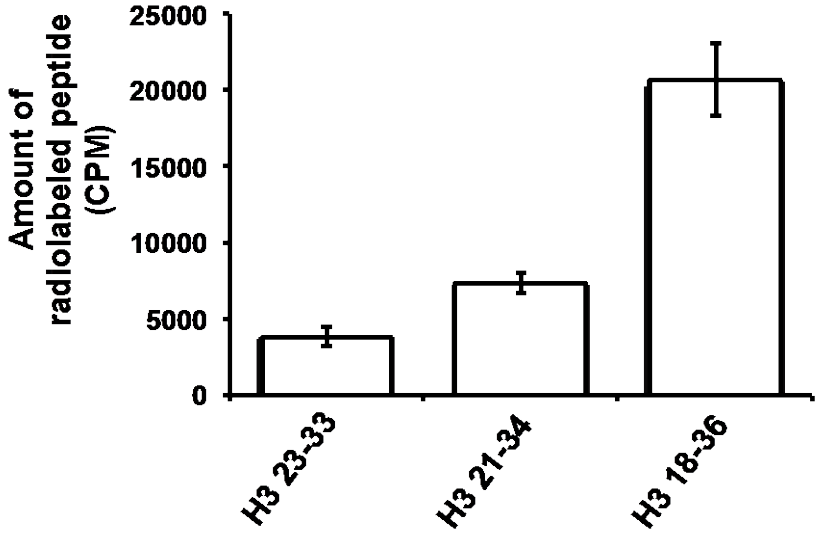


Figure S3: ATXR5 and ATXR6 bind histone H3.1 employing divergent structural determinants. Overall representation of ATXR5 (A), DIM-5 (B), vSET (C) and SET8 (D) and SET7/9 (E). In each complex, histone carbon atoms are highlighted in orange and the cofactor AdoHcy colored in magenta. Key structural elements such as the nSET (grey), iSET (beige), cSET (light blue) are also indicated.

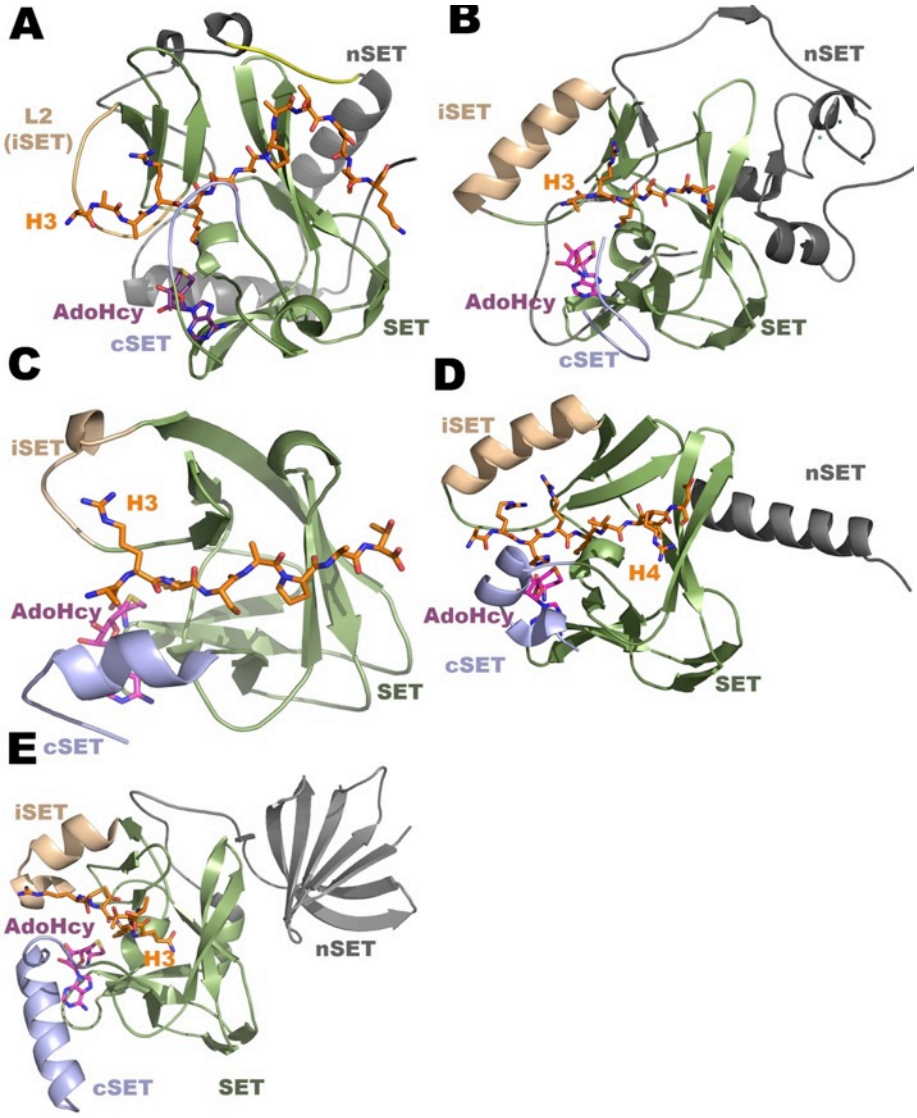


Figure S4: The selectivity pocket of ATXR5/6 is highly restrictive. Zoomed view of the selectivity pocket of RcATXR5. The histone H3.1 determinant Ala-31 was substituted for a threonine residue. Modeling of the threonine was performed such that it represents its most probable rotameric conformation (>98%). Clashes are highlighted as solid yellow lines. Modeling, detection of clashing atoms and the figures have been generated using the chimera software (*40*).

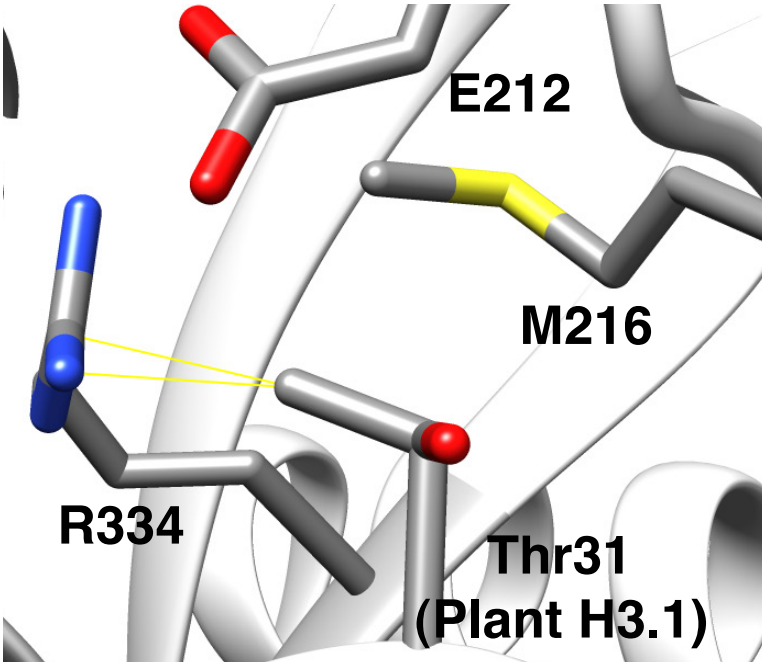


Figure S5: Relationship between H3.1 and H3K27me1 in Arabidopsis. A. Correlation between H3.1 occupancy (ChIP/Input) vs. H3K27me1 occupancy (ChIP/Input). Chromosomes were divided into 10 000 bp windows. Reads were normalized (RPM) and multiple mapping reads were weighted by the total number of positions to which they mapped. ρ: correlation coefficient for a Spearman's Rank correlation (non-parametric correlation) B. H3K27me1 occupancy (ChIP/Input) over significantly enriched regions of H3.1 occupancy (H3.1 peaks) and non-enriched regions (Background) as defined by SICER using following parameters: W=200, G=400, FDR=0.001. Reads were normalized (RPM) and multiple mapping reads were weighted by the total number of positions to which they mapped. Original data was obtained from previously published studies (*4*, *15*).

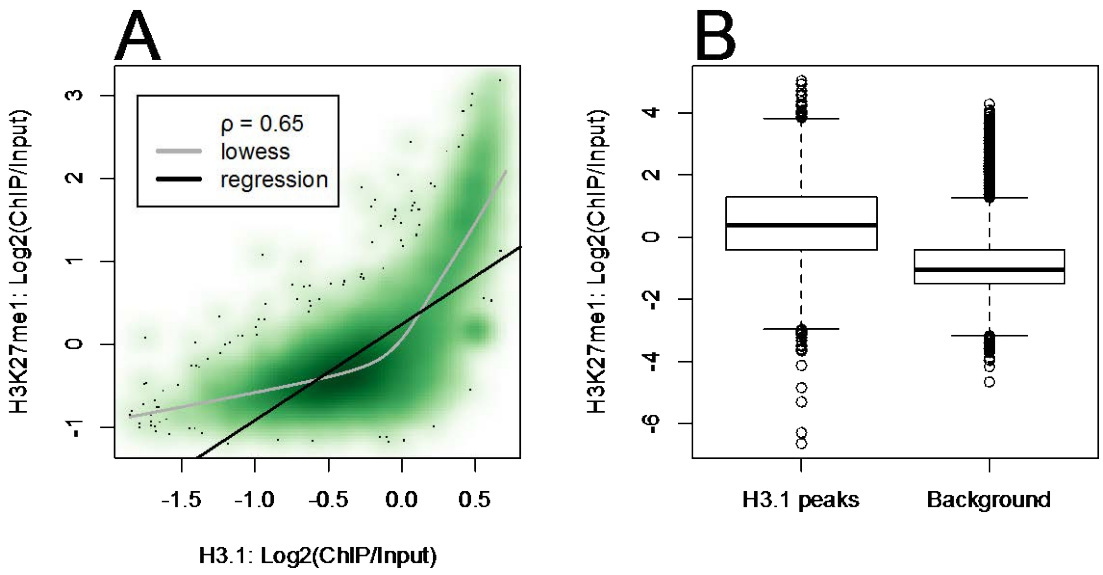


Figure S6: Distribution of H3.1 and H3.3 variants over a CACTA DNA transposon (*At4g03745*) and a COPIA-like retrotransposon (*At1g37110*). Significant islands of histone H3.1, but not H3.3, were identified (adjusted p value < 0.001).

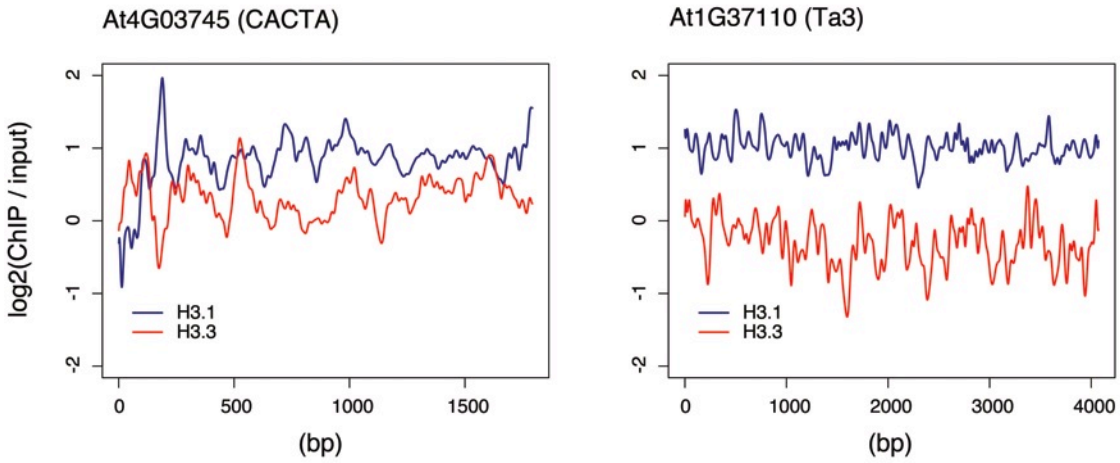


Figure S7: Arabidopsis PRC2 complexes have H3K27 monomethyltransferase activity. *In vitro* HKM assay using recombinant chromatin containing plant histone H3.1 and four different plant H3K27 methyltransferases. The products of the HKM assay were migrated on gel and detected using H3K27me1 and H3 antibodies.

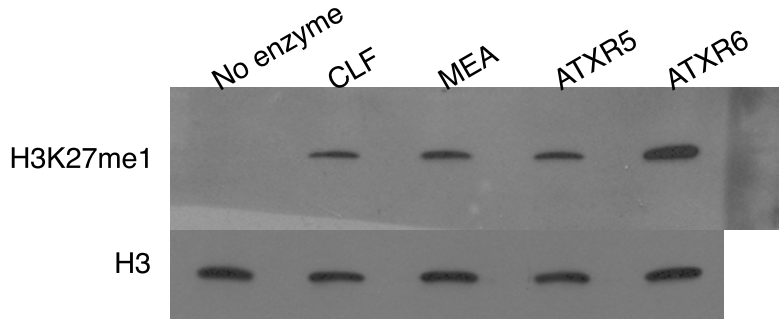


Figure S8: ChIP analysis of H3K27me1 in transgenic plants expressing WT H3.1 and H3.1 A31T. Results for H3K27me1 levels (upper panels) and H3 levels (lower panels) at the heterochromatic elements Ta3 (*At1g37110*) and CACTA *(At4g03745)* are presented. The results represent the average and standard deviation of three independent experiments.

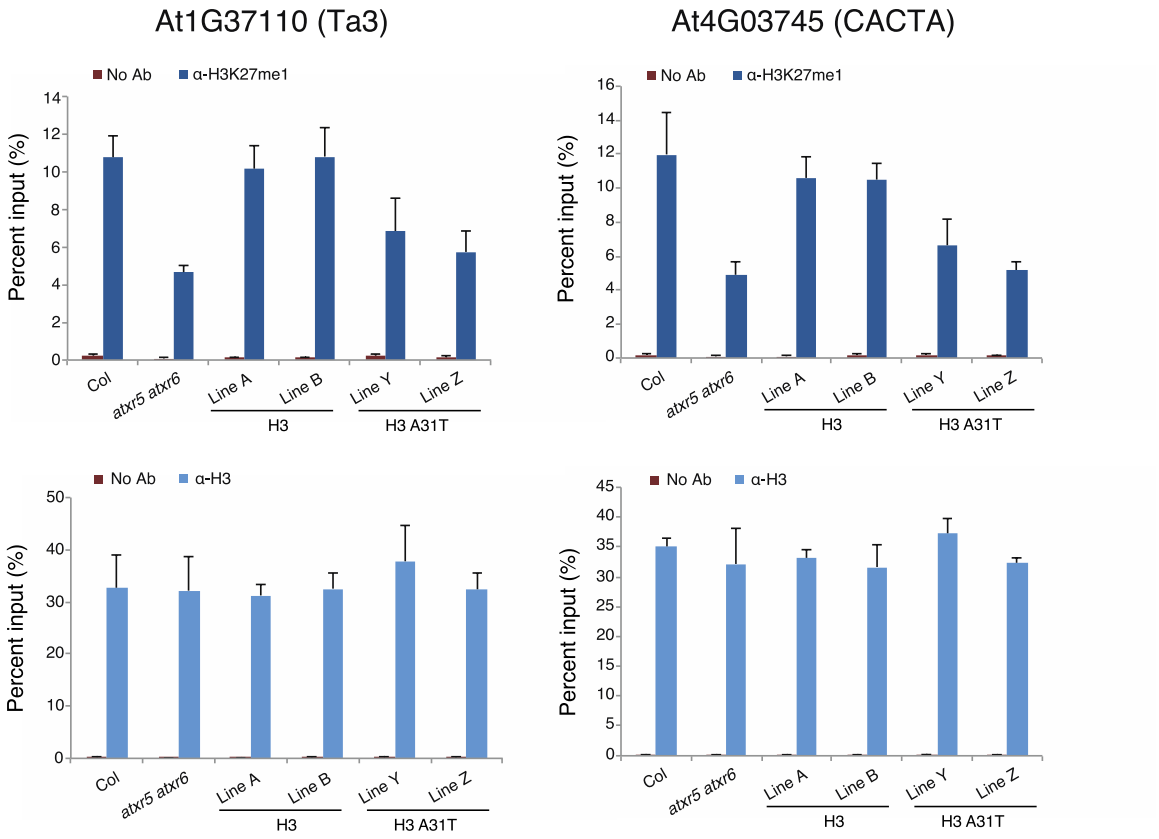


Figure S9: ChIP analysis of H3K27me1 in *fas-2* mutants. Results for H3K27me1 levels (upper panels) and H3 levels (lower panels) at the heterochromatic elements TSI, Ta3 (*At1g37110*) and CACTA *(At4g03745)* are presented. The results represent the average and standard deviation of three independent experiments.

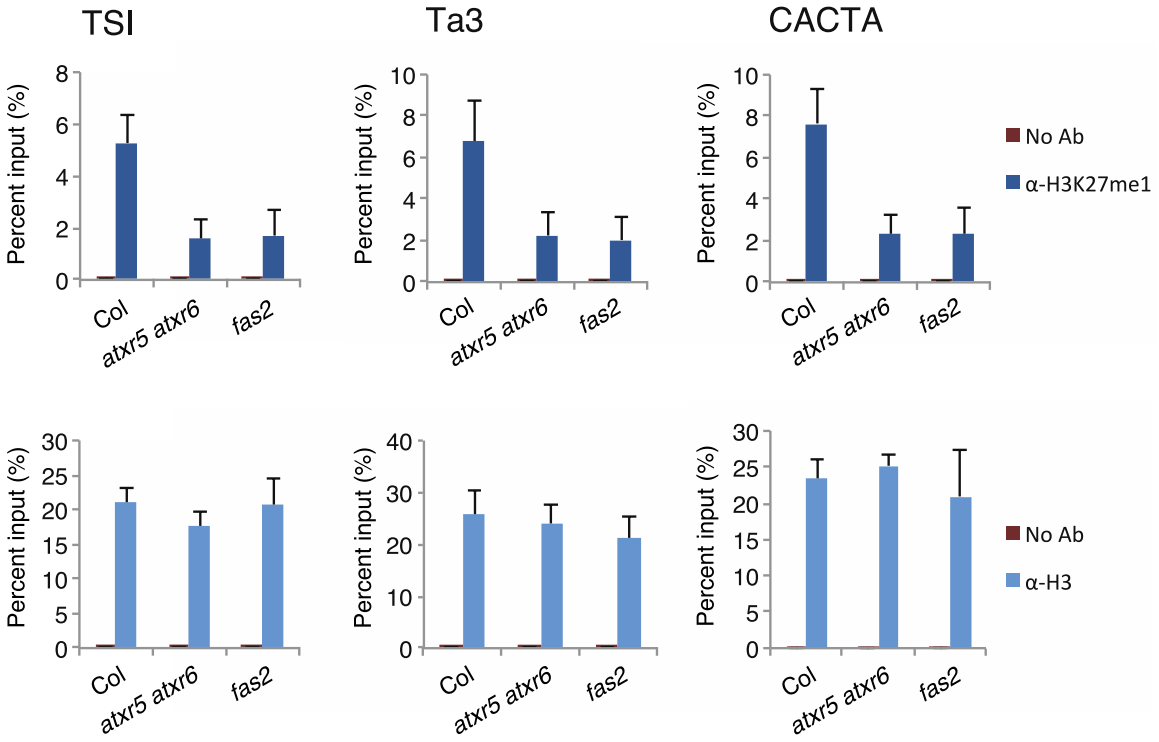


Figure S10: Binding affinity of H3K27me1 antibodies for H3.1 and H3.3. Quantitative analysis by Western blot of the binding affinity of different H3K27me1 antibodies for H3.1 and H3.3 peptides (a.a. 18-41). The binding activity for each antibody on histone H3.3 is relative to the activity of the antibody on histone H3.1.

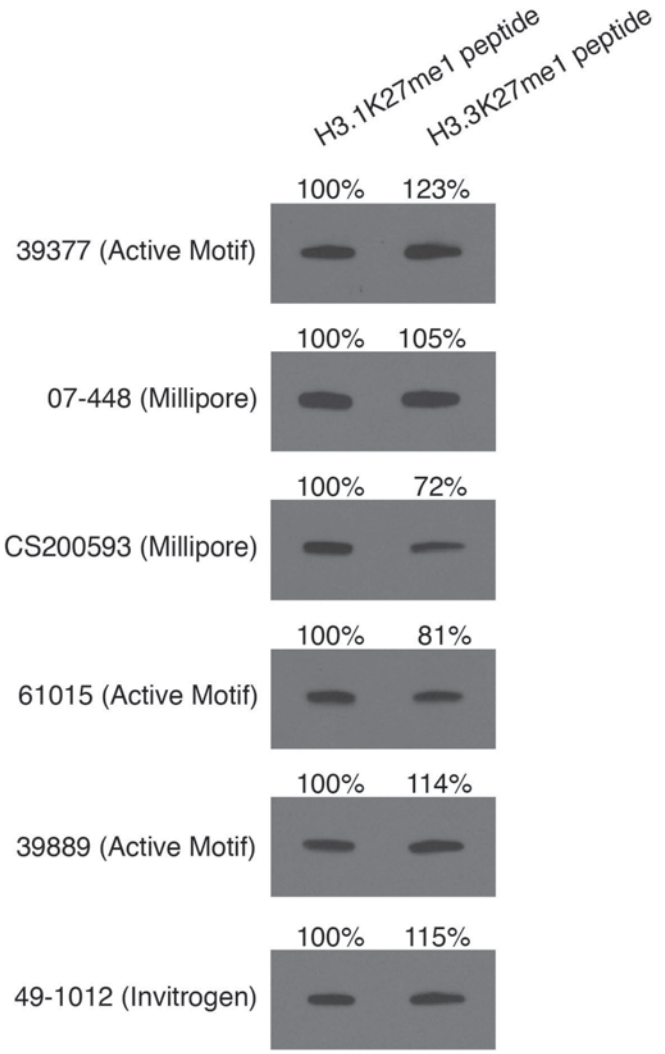


Figure S11: Flow cytometry profiles of leaf nuclei from Col, *atxr5 atxr6* and independent transgenic T4 (homozygous) lines expressing either wild-type H3.1 (Lines A-D) or H3.1 A31T (Lines W-Z). The numbers above the 8C and 16C peaks correspond to the robust CV values (PI units that enclose the central 68% of nuclei) for those peaks. High robust CV values at 8C and 16C peaks characterize heterochromatic over-replication in *atxr5 atxr6* mutants (*4*). PI: propidium iodide.

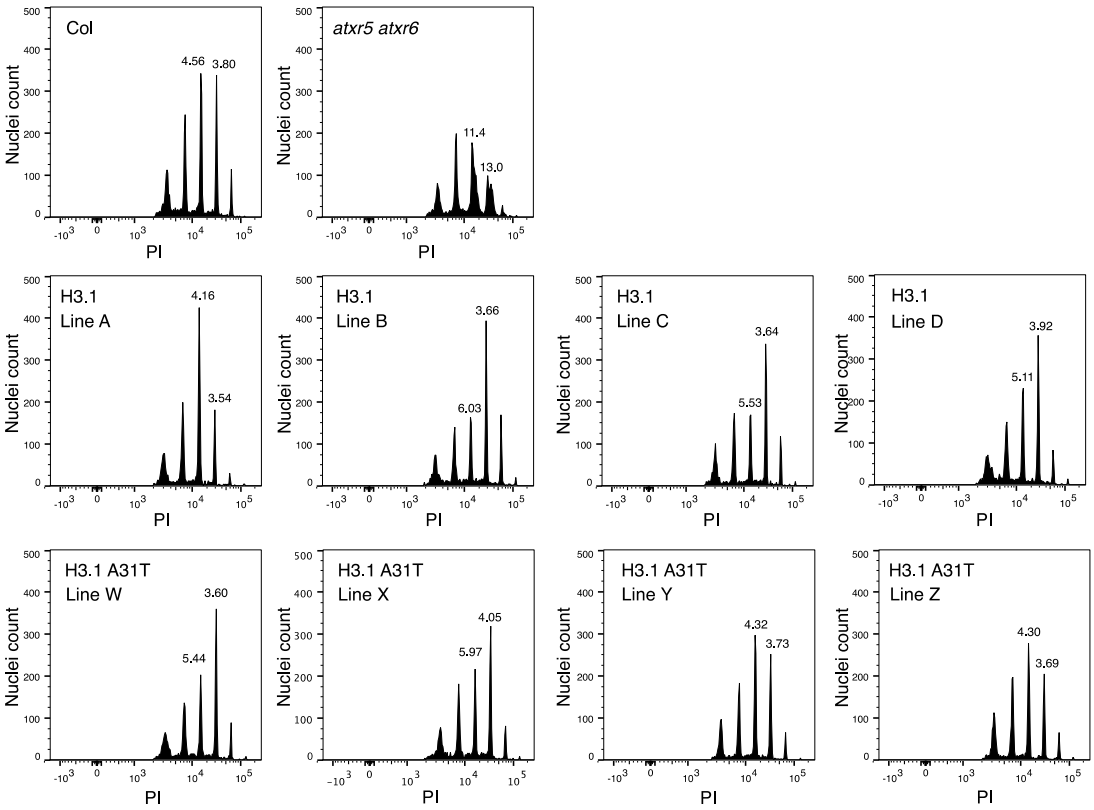
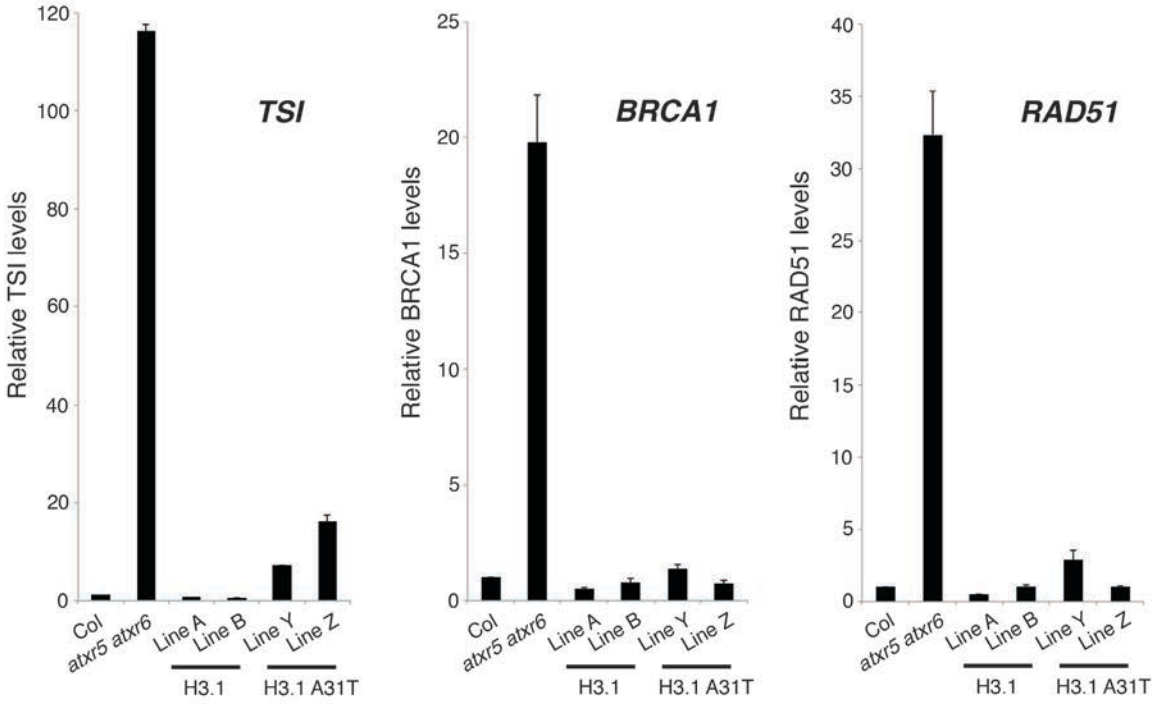


Figure S12: Expression analysis (RT-qPCR) of the DNA repair genes *BRCA1* and *RAD51* in the transgenic T4 (homozygous) lines expressing H3 and H3.1 A31T. *BRCA1* and *RAD51* are overexpressed in *atxr5 atxr6* and their overexpression was found to be dependent on heterochromatic over-replication (*5*). TSI expression is shown as a control to confirm the effect of the H3.1 A31T lines on heterochromatic transcription. The average and standard error of measurement for two independent experiments is presented.



Supplementary Table S1. Data collection, phasing and refinement statistics.

	□RcATXR5/H3.1/SAH
Data collection	
Space group	C2
Cell dimensions	
a, *b*, *c* (Å)	101.2, 87.3, 74.5
β□(□)	127.8
Resolution (Å)	27.1 – 2.10 (2.18-2.10)*
R_{sym}	6.0 (30.0)
I/ □I	12.1 (4.0)
Completeness (%)	98.3 (96.6)
Redundancy	4.1 (4.1)
Refinement	
Resolution (Å)	26.4 – 2.10
No. reflections	29410
R_{work}/R_{free}	20.6 / 25.2
No. atoms	
Protein	3418
H3.1/SAH	170/51
Water	249
B-factors (Å^2)	
Protein	21.2
H3.1/SAH	19.3/34.5
Water	48.2
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.15
Molprobity score	2.11

The data set has been collected on a single crystal
* Highest resolution shell is shown in parenthesis