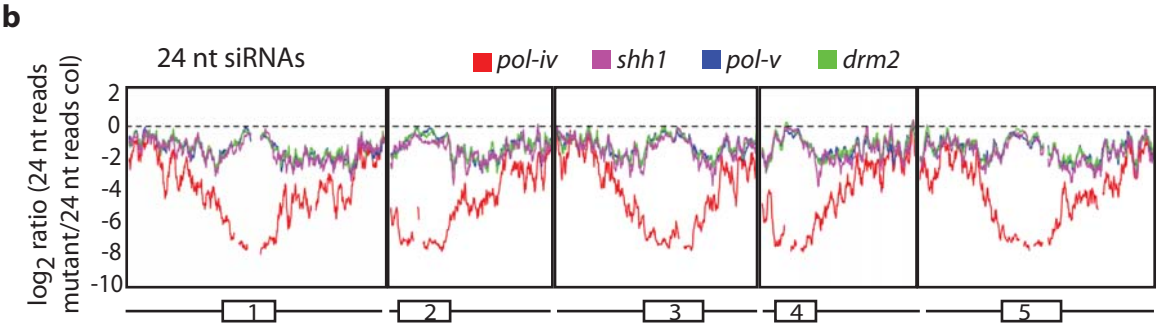
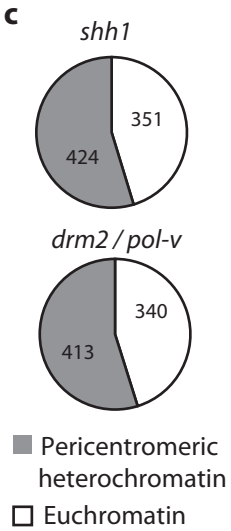
SUPPLEMENTARY INFORMATION

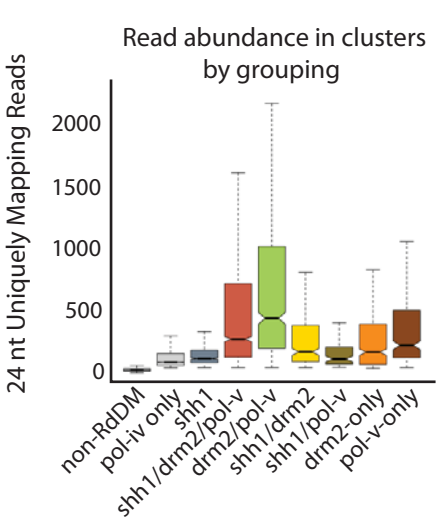
doi:10.1038/nature12178

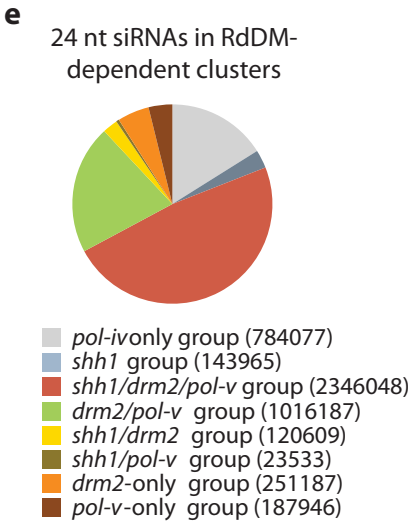
a

		Number of clusters	24 nt siRNA Reads
By Genotype	*pol-iv*	7,658	4,859,216
	shh1	3,394	2,634,155
	drm2	3,695	3,734,031
	pol-v	3,376	3,573,714
	Not affected in any genotype	4,871	163,984
By Subclass (Fig. 1b)	*pol-iv* only	2,883	784,077
	shh1	775	143,965
	shh1/drm2/pol-v	2,257	2,346,048
	drm2/pol-v	573	1,016,187

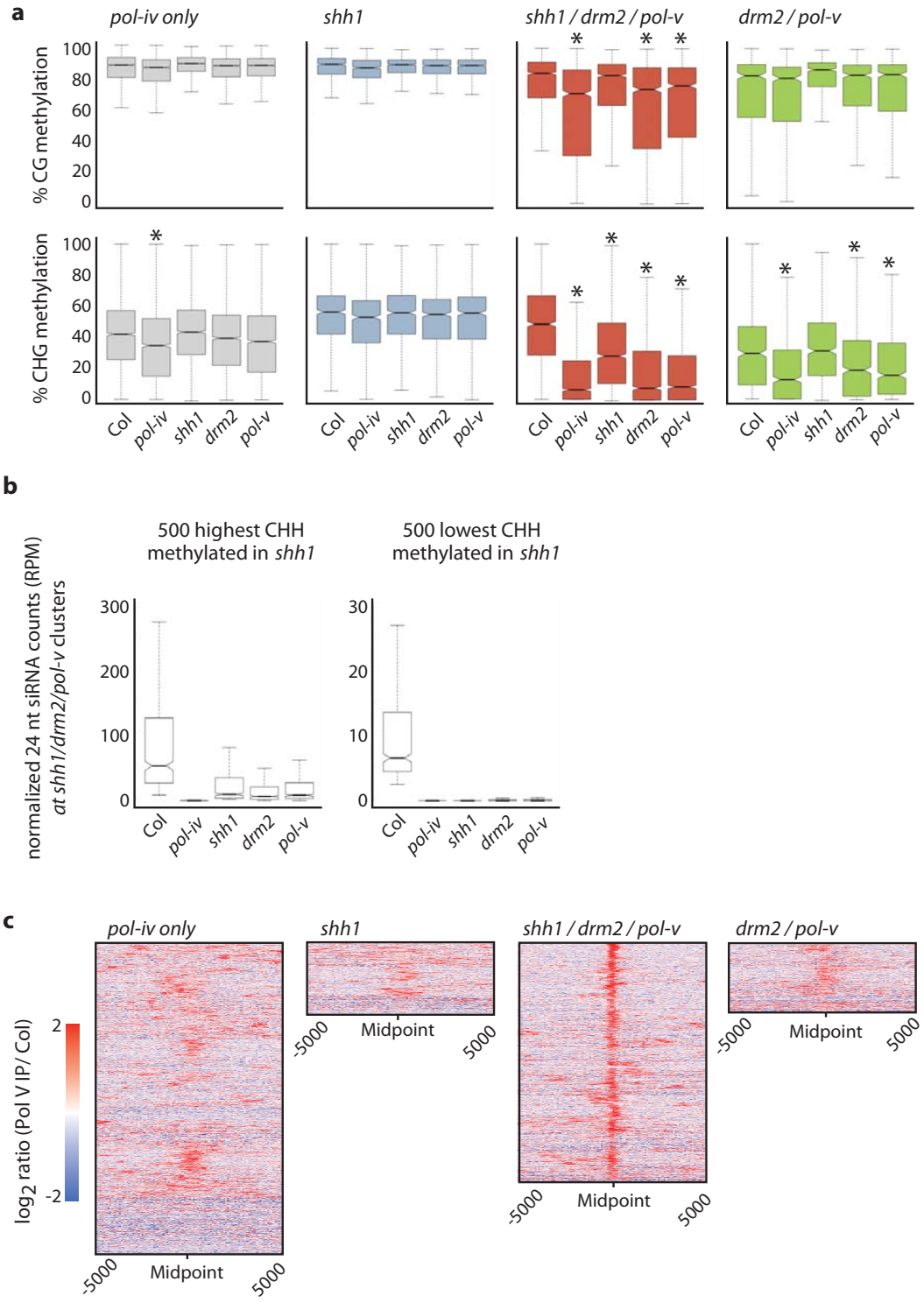




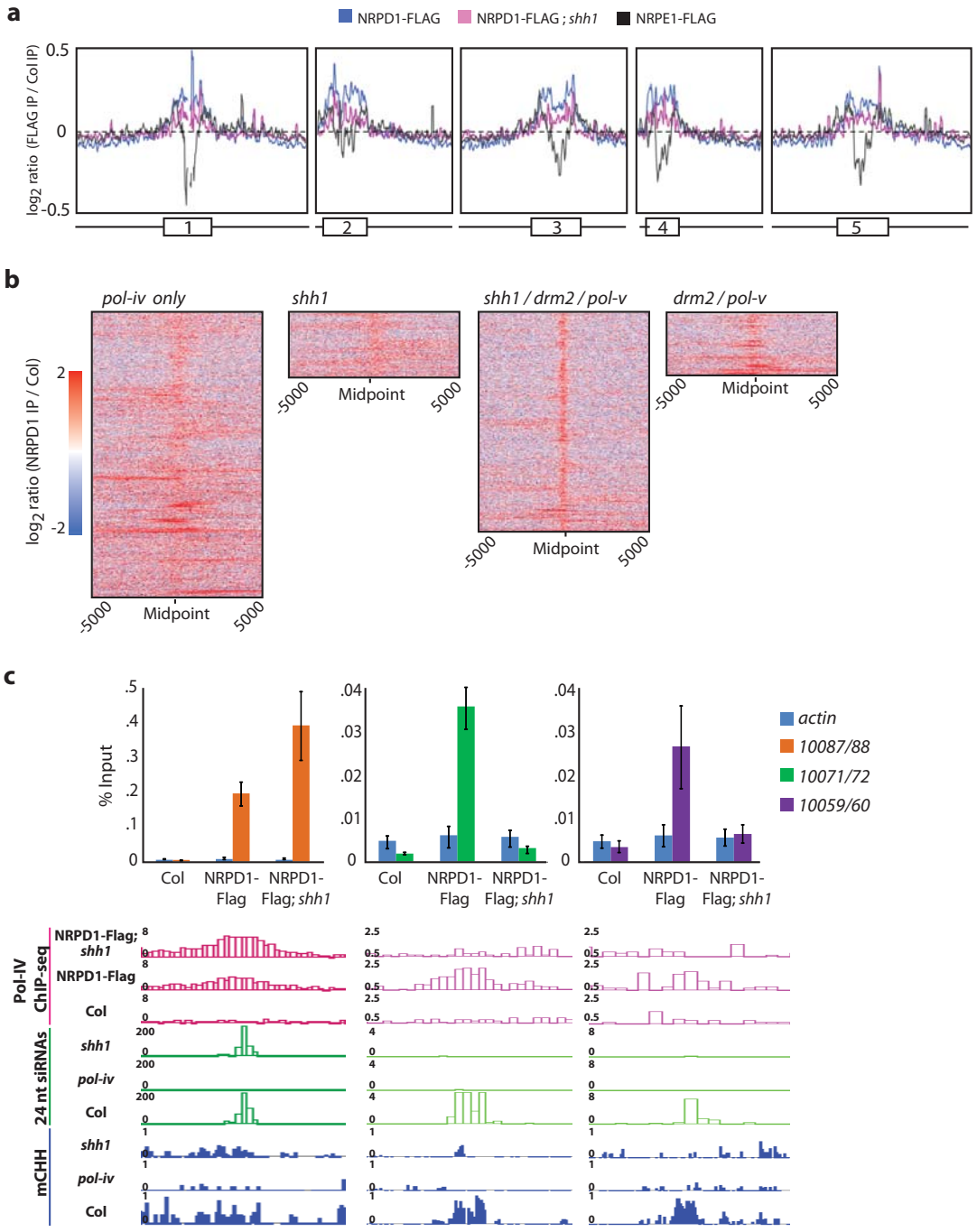




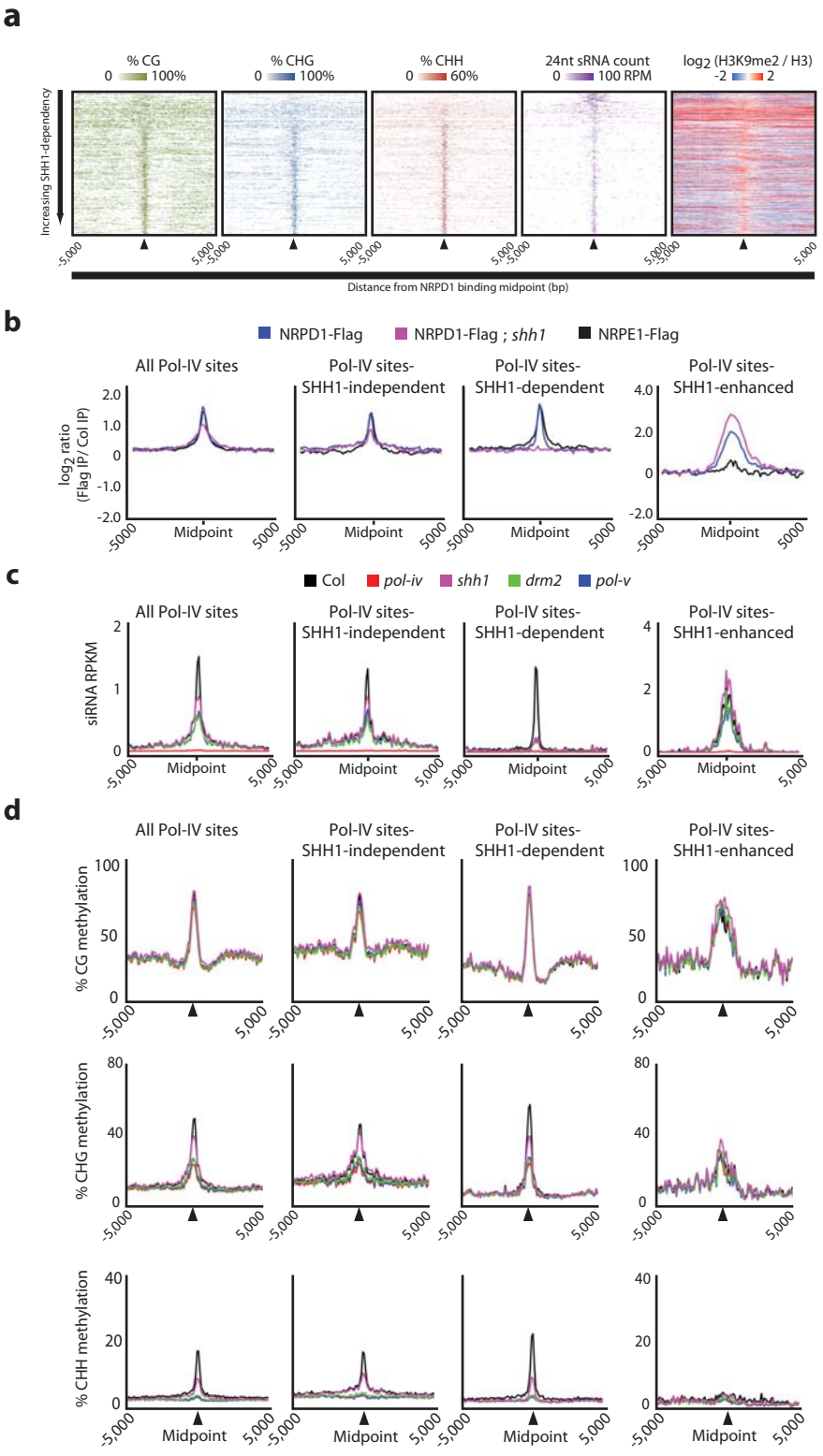
Supplementary Figure 1 Analysis of siRNA clusters, levels, and distribution in the Arabidopsis genome. **a**, Number of 24 nt siRNA clusters falling into the classes categorized by genotypes or siRNA subclasses defined in Fig. 1b. **b**, Rolling average of ten 50 kbp windows of 24 nt siRNA abundance relative to the wild-type (Col) pattern across each Arabidopsis chromosome for *pol-iv*, *pol-v*, *drm2*, and *shh1* mutants. Schematic representations of each chromosome are shown below, with the boxed regions corresponding to the pericentromeric chromatin. **c**, Pie charts analogous to those in Fig. 1c showing the distribution of affected siRNA clusters in the indicated subclasses. **d**, siRNA abundance in Col plants for clusters unaffected in RdDM mutants (non-RdDM), the cluster subclasses defined in Fig. 1b, and the clusters that fall outside these subclasses. **e**, A plurality of siRNAs within Pol-IV-dependent clusters fall within the *shh1/drm2/pol-v* siRNA subclass.



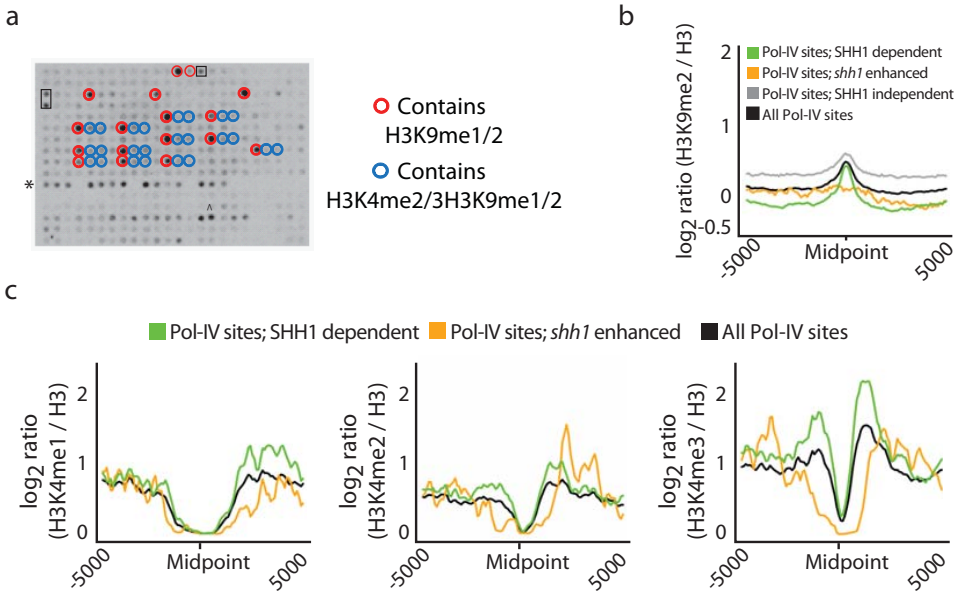
Supplementary Figure 2 DNA methylation levels at reduced siRNA clusters. **a**, Boxplots of CG and CHG DNA methylation profiles as assayed by BS-seq at the siRNA cluster subclasses defined in Fig. 1b. (* indicates significant reduction; P<1e-10 Mann-Whitney U test). **b**, Boxplots showing the siRNA levels (reads per million mapping reads) remaining in *shh1* mutants are higher at the clusters that retain the most CHH methylation (top 500 within the *shh1/drm2/pol-v* subclass) than those that retain the least CHH methylation (bottom 500 within the *shh1/drm2/pol-v* subclass). **c**, Heatmaps of Pol-V enrichment over affected siRNA clusters.



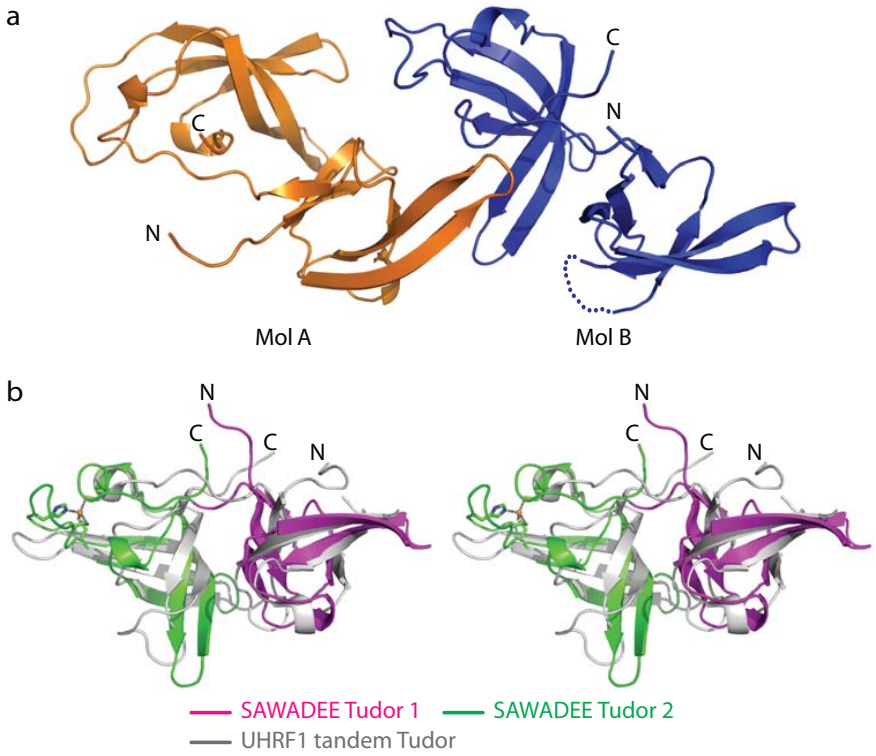
Supplementary Figure 3 Pol-IV ChIP-seq . **a**, Chromosomal views of NRPD1-FLAG IP enrichment as compared to an untagged control (Col) in wild type and shh1 mutants. Also shown is the previously published genome profile of NRPE1[10]. **b**, Heatmaps of Pol-IV enrichment over the defined siRNA clusters. **c**, Quantitative PCR validation of select Pol-IV peaks showing the standard error of two biological replicates and associated genome browser screen shots. mCHH tracks are at single bp resolution, for 24 nt siRNA and Pol-IV ChIP-seq tracks each bar represents 50 bps. actin: 5'AGCACGGATCGAATCACATA3' and 5'CTCGCTGCTTCTCGAATCTT3' 10087/88: 5'GAGATGATGTGCTTAGATTGTG3' and 5'GGTTTTGGTTAT GACGGC3' 10071/72: 5'GAATGGACTTCTTTCCGAAGG3' and 5'GGAGAAGATTTGGGGAAGGG3' 10059/60: 5'GGCTTCGATAGGAAGAATGCCC3' and 5'GTGAAACT GCCAGA TCCAAATTC3'



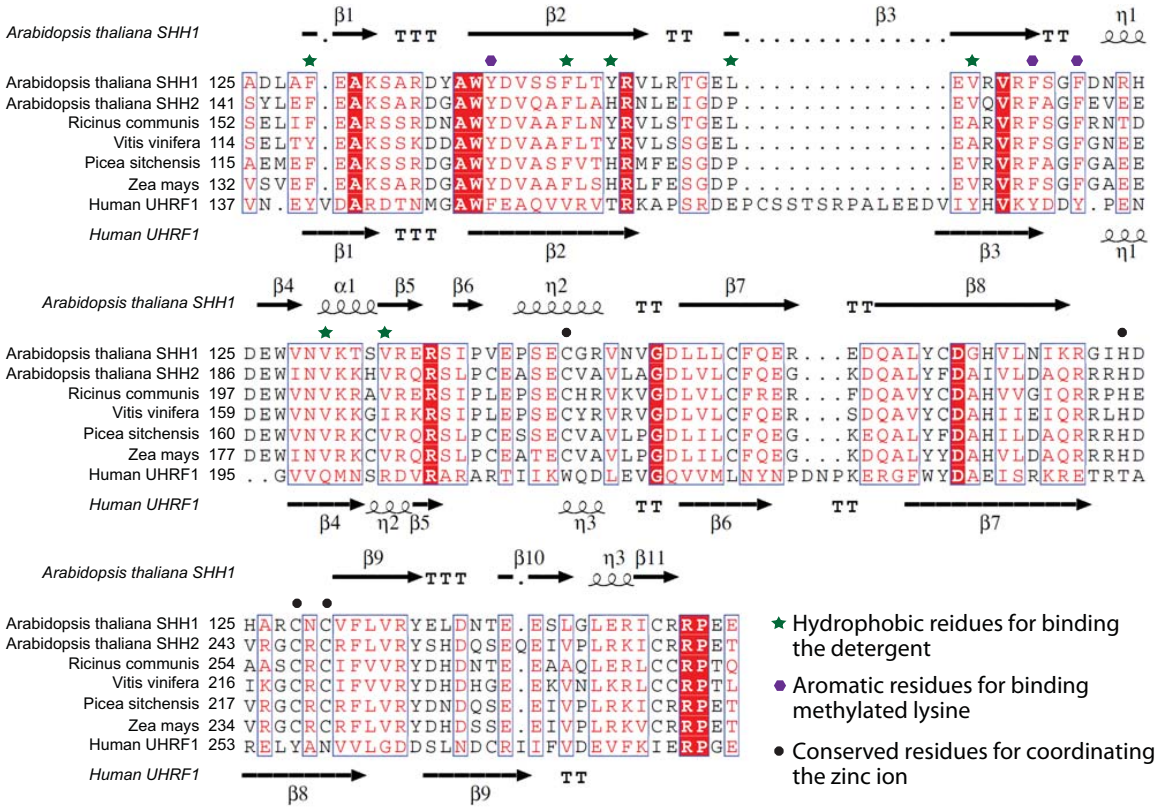
Supplementary Figure 4 Analysis of different Pol IV ChIP-seq peak classes. **a,** Heatmaps of epigenetic marks over the identified Pol IV peaks. **b,** Metaplots of polymerase enrichment **c**, siRNA abundance or **d**, % CG, CHG, and CHH methylation over Pol-IV peak classes.



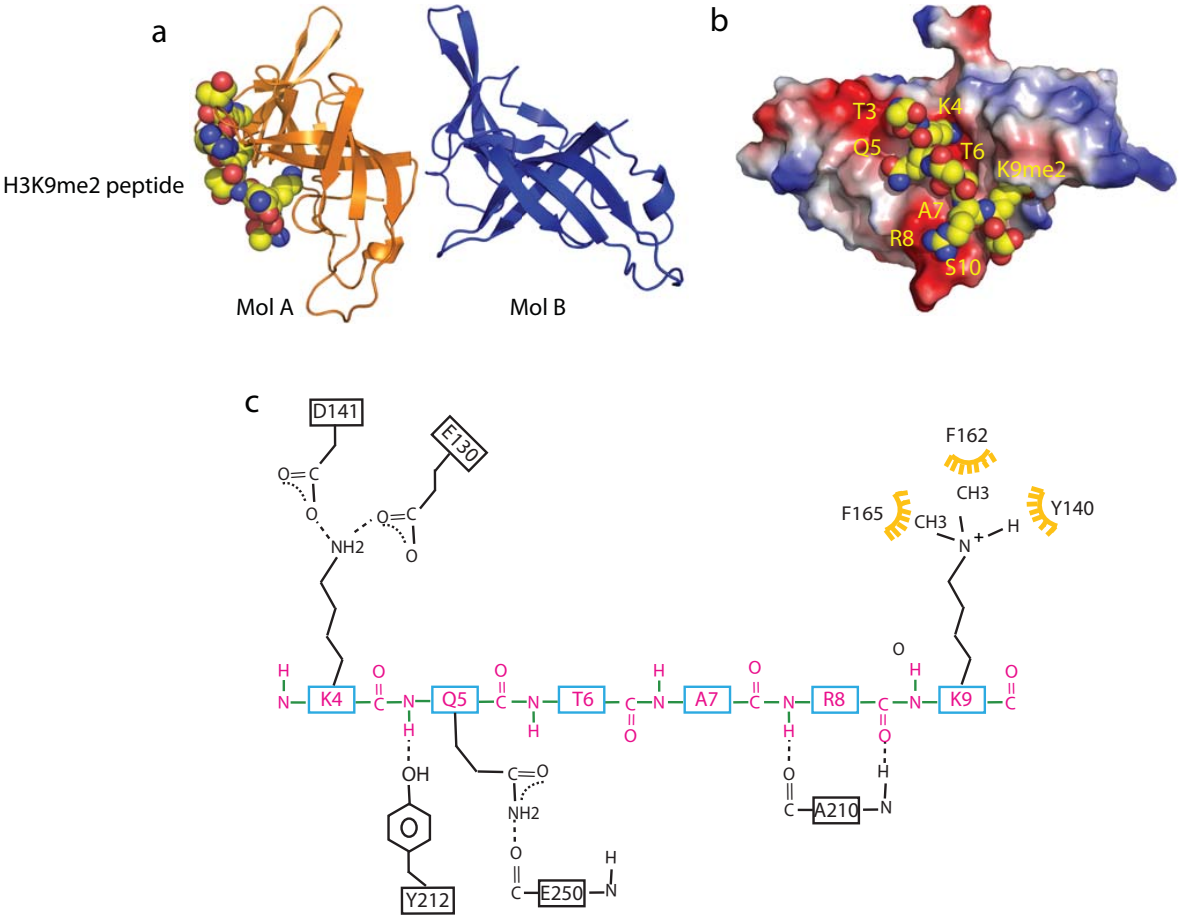
Supplementary Figure 5 Peptide array and Pol-IV peak analysis. a, Active Motif peptide array bound with a GST-SHH1 SAWADEE domain (125-258 aa) fusion protein and detected using a GST antibody. Strongly bound peptides that harbor H3K9me1 or H3K9me2 modifications either alone or in combination with other modifications are circled in red (O). Adjacent peptides that differ only in the addition of an H3K4me2 or H3K4me3 modification, and are not bound by the SHH1 SAWADEE domain, are circled in blue (O). Other lysine methylated peptides that were also bound by the SHH1 SAWADEE domain in this assay, but were bound very weakly or at undetectable levels by ITC, are indicated as follows: H3K4me3 modifications are marked by boxes, H3K27 methyl-modifications, which are present singly or in combination with other modifications in the row marked by an *, and peptides that contain H4K20me modifications are indicated by an ^. **b**, Metaplot analyses showing the enrichment of H3K9me2 over the indicated Pol-IV ChIP-seq peak classes. **c,** Metaplot analyses showing the enrichment of H3K4me1, me2, or me3 over the indicated Pol-IV ChIP-seq peak classes using previous published H3K4



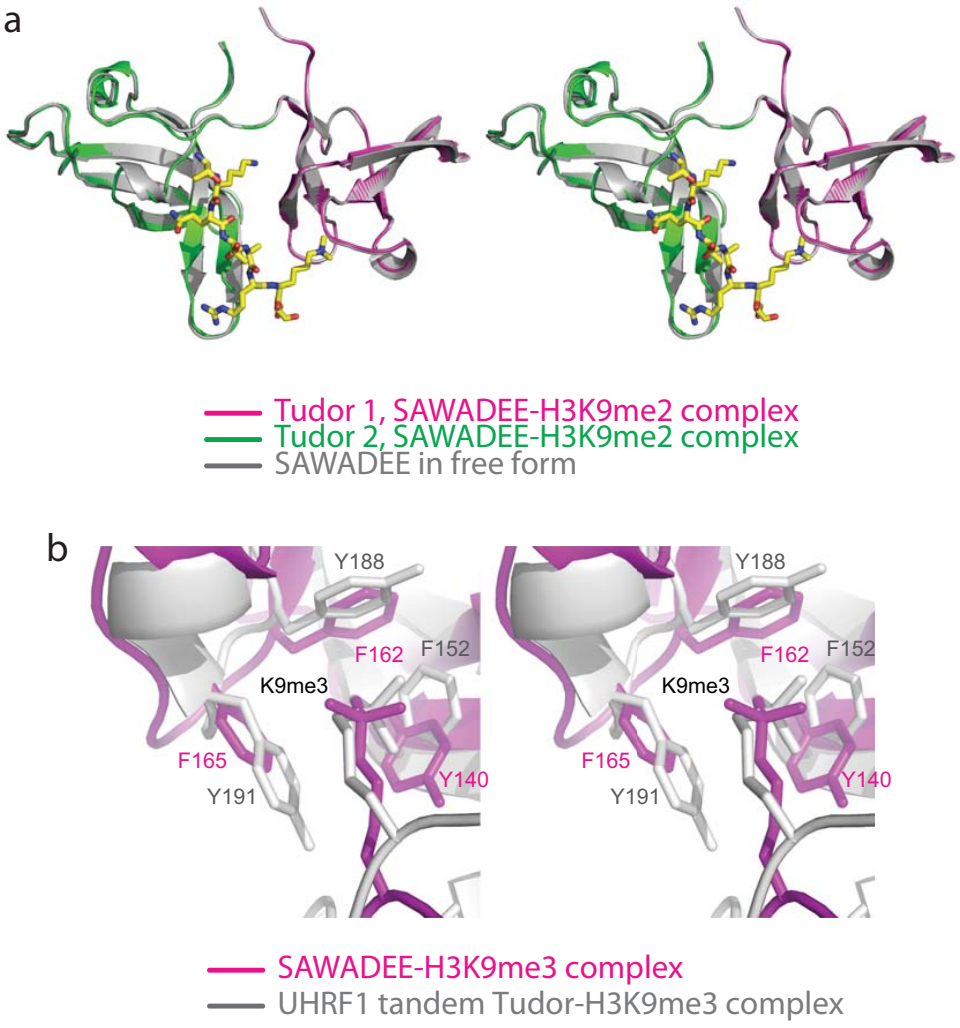
Supplementary Figure 6 Overall structure of the SHH1 SAWADEE domain in the free form. a, There are two molecules (Mol A and Mol B) in the asymmetric unit of the asymmetric dimer, which are shown in orange and blue, respectively. The dimer interface has several pairs of charge-charge interactions, which probably reflect contributions from crystal packing. The two monomers in the dimer adopt similar structures with an r.m.s.d. of 0.55 Å for 126 aligned C alpha atoms. **b**, Stereo view of superposition of the SHH1 SAWADEE domain (with Tudor 1 in magenta and Tudor 2 in green) and the UHRF1 tandem Tudor domain (in silver) reveals their structural similarity.



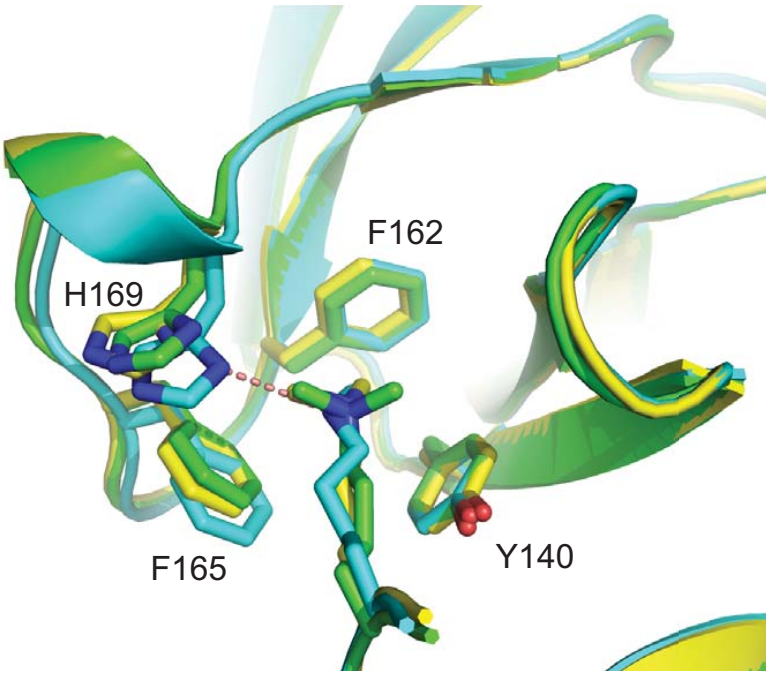
Supplementary Figure 7 Structure-based sequence alignment. Structure based sequence alignment of Arabidopsis thaliana SHH1 (Genbank ID: AAD39678.1), SHH2 (Genbank ID: AEE76089.1), and the SAWADEE domains from SHH1 orthologs in other species, including *Ricinus communis* (Genbank ID: EEF46394.1), *Vitis vinifera* (Genbank ID: CBI36072.3), *Picea sitchensis* (Genbank ID: ABR17173.1), and *Zea mays* (Genbank ID: ACL53048.1). The tandem Tudor domain of human UHRF1 is also included for comparison. The secondary structure of the SHH1 SAWADEE domain and the human UHRF1 tandem Tudor domain are presented on the top and bottom of the alignment, respectively. The partially conserved hydrophobic residues of the SAWADEE domain involved in binding the nonpolar tail of the detergent are marked by green stars. The conserved aromatic residues involved in recognition of methylated K9 are marked by purple hexagons. The conserved SAWADEE residues involved in coordinating the zinc ion are marked by black circles.



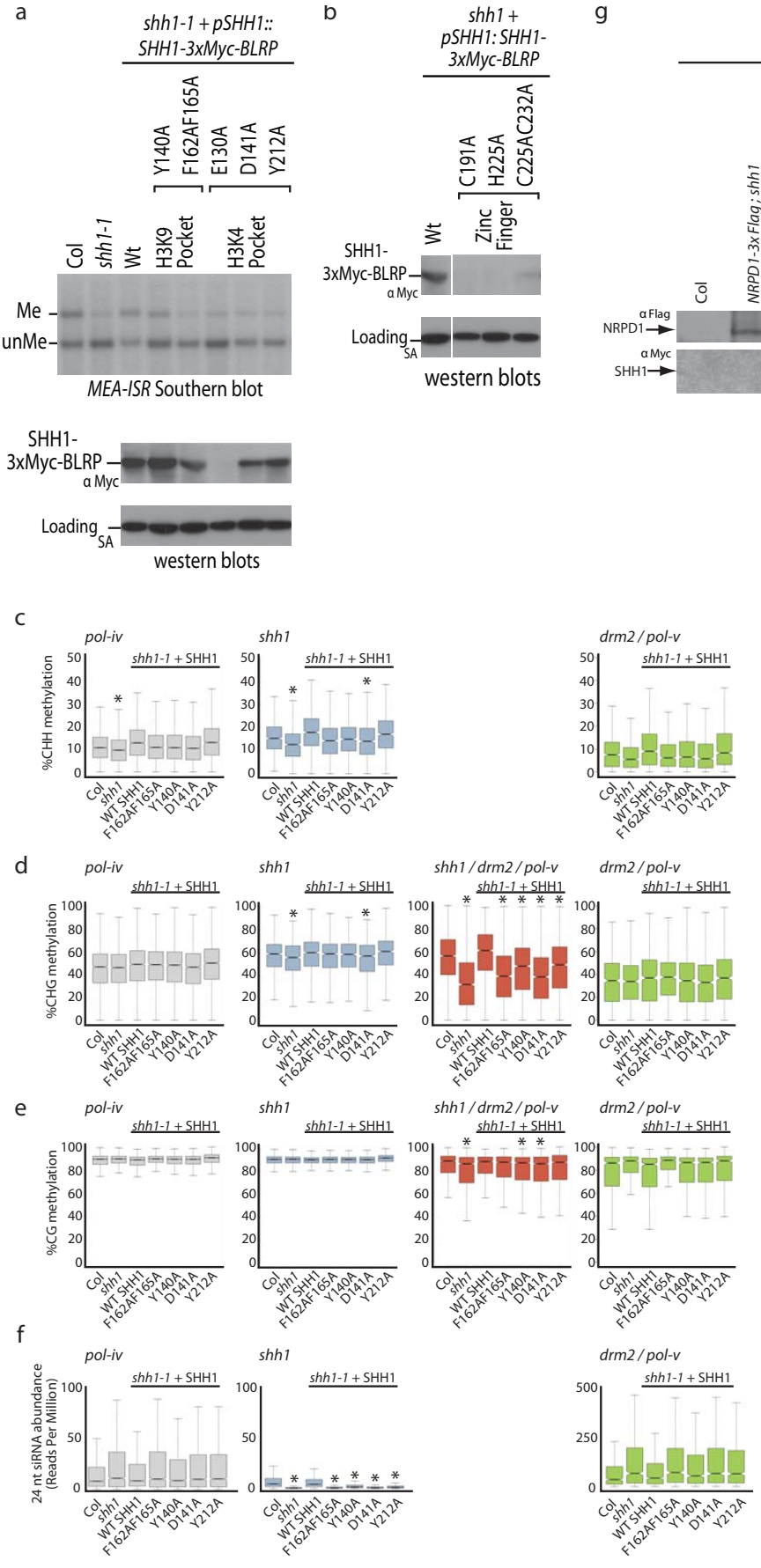
Supplementary Figure 8 Structure of the H3(1-15)K9me2-SHH1 SAWADEE complex. a, , In the crystal structure of the SHH1 H3(1-15)K9me2-SAWADEE peptide complex, there are two molecules (Mol A and Mol B) in the asymmetric unit that form an asymmetric dimer, which are shown in orange and blue, respectively. Only one SAWADEE molecule has the bound peptide, which is shown in a space-filling representation, while the peptide binding groove in the other molecule is occluded due to packing interactions. **b**, The overall structure of the H3(1-15)K9me2-SAWADEE complex. The SAWADEE domain is shown in an electrostatic surface representation and the peptide in a space-filling representation. **c**, Schematic representation of the intermolecular interactions between the SAWADEE domain and the bound H3(1-15)K9me2 peptide.



Supplementary Figure 9 Comparing the structure of SHH1 SAWADEE-histone peptide complexes with free form SAWADEE domain and UHRF1 tandem Tudor. a, Superposition of the H3K9me2 peptide bound SAWADEE domain (with Tudor1 in magenta and Tudor 2 in green) and the free form of the SAWADEE domain (in silver) reveals no significant overall conformational change upon peptide binding. **b**, A stereo view comparing the K9me3 binding pocket of the SHH1 SAWADEE domain and the UHRF1 tandem Tudor domain. The K9me3 inserts deeper into the binding pocket of the UHRF1 tandem Tudor, revealing the molecular basis underlining UHRF1's preference for higher methylated K9 containing peptides, while the SAWADEE domain does not distinguish different methylation states of H3K9me3 and H3K9me2.



Supplementary Figure 10 Structure basis for the non-discrimination of H3(1-15)K9me1/2/3 peptides by SAWADEE domain. The superposition of H3K9me3-SAWADEE (in green), H3K9me2-SAWADEE (in yellow), and H3K9me1-SAWADEE (in cyan) complexes positioned the methylated lysines within the same aromatic cage. The side chain of His169 undergoes a small but significant conformational change in order to hydrogen bond (highlighted with dash red line) with the monomethyllysine ammonium proton in the structure of the H3K9me1 complex which explains the SAWADEE domain exhibits almost equal affinity against H3K9me1/2/3 peptides.



Supplementary Figure 11. DNA methylation, siRNAs, and protein stability of SHH1 SAWADEE doamin point mutants a, Southern blot showing the methylation status of the MEA-ISR locus. Genomic DNA from plants of the indicated genetic backgrounds were digested with the methylation sensitive restriction enzyme MspI. The upper band (Me) corresponds to methylated DNA that is resistant to digestion by MspI and the lower band (unMe) corresponds to unmethylated DNA that is digested by MspI. The expression level of the Wt and mutant versions of the epitope tagged SHH1 protein are shown below in the anti-Myc (α Myc) western blot. An unidentified biotinylated background band detected by streptavidin (SA) was used as a loading control. BLRP; Biotin Ligase Recognition Peptide. **b,** western blot detecting the WT and various zinc binding pocket mutants (upper) or a nonspecific biotinylated protein as a loading control (lower). **c-e** Boxplots showing global methylation levels and **f** showing siRNA levels at the four subclasses of siRNA clusters in wildtype and *shh1* mutants as well as *shh1* mutants transformed with SHH1 constructs with either the wildtype SHH1 sequence or induced point mutations in the K9 (F162AF165A and Y140A) or K4 (D141A and Y212A) binding pockets. Missing panels are represented in the main figures. (* indicates significant reduction; P<1e-10 Mann-Whitney U test) **g,** Western blots showing the co-immunoprecipitation of the NRPD1 subunit of the Pol IV polymerase in an *shh1* null background(*NRPD1-3x Flag ; shh1)* and either wildtype or the indicated SHH1 SAWADEE domain point mutants in the F1 generation (*F1 NRPD1-3x Flag ; shh1 x SHH1 Wt-3xMyc ; shh1).* Non transgenic plants of the Col ecotype and the *NRPD1-3x Flag ; shh1* parental line were included as negative controls.

Supplementary Table 1. ITC-based Binding Parameters for Modified Histone Peptides Bound to SHH1 SAWADEE Domain

Protein	Peptide	N value	Kd (μM)	H (kcal/mol)	S(cal/mol/deg)
SAWADEE	H3(1-15)	0.80 ± 0.01	34.8 ± 0.9	-8.57 ± 0.11	-10.3
SAWADEE	H3(1-15)K9me3	1.06 ± 0.01	2.63 ± 0.2	-12.5 ± 0.2	-19.3
SAWADEE	H3(1-15)K9me2	0.86 ± 0.00	1.90 ± 0.06	-14.8 ± 0.0	-26.7
SAWADEE	H3(1-15)K9me1	1.23 ± 0.01	2.22 ± 0.20	-14.4 ± 0.2	-25.7
SAWADEE	H3(1-15)K4me1K9me1	0.99 ± 0.01	1.90 ± 0.16	-9.81 ± 0.10	-8.97
SAWADEE	H3(1-15)K4me1K9me2	1.02 ± 0.02	2.47 ± 0.50	-6.77 ± 0.18	1.43
SAWADEE	H3(1-15)K4me2K9me2	1.08 ± 0.02	7.81 ± 0.99	-8.90 ± 0.24	-8.49
SAWADEE	H3(1-15)K4me3K9me2	1.00 ± 0.02	11.0 ± 0.94	-8.70 ± 0.22	-8.45
SAWADEE	H3(1-15)K4AK9me2	1.18 ± 0.09	63.7 ± 15.3	-4.73 ± 0.61	2.26
SAWADEE	H3(1-15)K4me1	0.85 ± 0.03	26.9 ± 2.6	-2.37 ± 0.12	12.4
SAWADEE	H3(1-15)K4me2	0.98 ± 0.08	52.1 ± 7.8	-7.42 ± 0.84	-6.96
SAWADEE	H3(1-15)K4me3	-	NDB[a]	-	-
SAWADEE	H3(1-15)K4A	-	NDB	-	-
SAWADEE	H3(22-32)K27me1	0.93 ± 0.07	113 ± 9	-4.68 ± 0.42	1.29
SAWADEE	H3(22-32)K27me2	-	NDB	-	-
SAWADEE	H3(22-32)K27me3	-	NDB	-	-
SAWADEE	H3(31-41)K36me1	-	NDB	-	-
SAWADEE	H3(31-41)K36me2	-	NDB	-	-
SAWADEE	H3(31-41)K36me3	-	NDB	-	-
SAWADEE	H4(1-23)	-	NDB	-	-
SAWADEE	H4(1-23)K20me1	1.09 ± 0.08	72.5 ± 13.0	-6.85 ± 0.81	-5.59
SAWADEE	H4(1-23)K20me2	0.82 ± 0.02	80.6 ± 2.5	-5.30 ± 0.16	-0.26
SAWADEE	H4(1-23)K20me3	1.09 ± 0.15	147 ± 25	-6.63 ± 1.22	-6.21

Buffer: 100 mM NaCl, 10 mM HEPES, pH 7.5, 2 mM β-mercaptoethanol.
[a] NDB, no detectable binding. The standard deviation (s.d.) was derived from nonlinear fitting.

Supplementary Table 2. Diffraction data and structure refinement statistics of the *at*SHH1 SAWADEE domain in the free state

Data collection		
Crystal	Se-SAWADEE (L200M, L218M)[a]	SAWADEE
PDB code	4IUP	4IUQ
Beamline	APS-24ID-C	APS-24ID-C
Wavelength (Å)	0.9792 (Se-peak)	1.2819 (Zn-peak)
Space group	$P2_1$	$P2_12_12$
Cell dimensions		
a, b, c (Å)	53.3, 56.7, 59.3	54.7, 117.7, 53.4
β (°)	93.4	90
Resolution (Å)	30.0-1.9 (1.97-1.90)[b]	50.0-2.8 (2.90-2.80)[b]
R_{merge} (%)	12.6 (52.1)	10.5 (41.2)
Average I/σ(I)	12.2 (1.9)	14.3 (2.5)
Completeness (%)	100.0 (100.0)	99.8 (100.0)
Redundancy	3.7 (3.2)	3.1 (3.1)
Refinement and structure model		
Resolution (Å)	30.0-1.9 (1.97-1.90)[b]	50.0-2.8 (2.90-2.80)[b]
Unique reflections	54,961	16,076
R / R_{free} (%)	17.2 / 19.4	20.1 / 26.1
Number of atoms	2645	2220
Protein	2152	2185
Zn^{2+} ion	2	2
Ligand	48	33
Water	443	-
B factors (Å^2)	21.2	44.6
Protein	18.2	42.9
Zn^{2+} ion	13.0	59.6
Ligand	45.2	99.7
Water	32.8	-
RMS deviations		
Bond lengths (Å)	0.015	0.019
Bond angles (°)	1.187	1.692

[a] In order to solve the phase with Se-methionine substituted protein, Leu200 and Leu218 of the SAWADEE domain were mutated to methionine.
[b] Values in parentheses are for highest-resolution shell.

Supplementary Table 3. Diffraction data and structure refinement statistics of the *at*SHH1 SAWADEE domain in complexes with different modified histone H3 peptides

Data collection				
Crystal	SAWADEE-H3(1-15)K9me3	SAWADEE-H3(1-15)K9me2	SAWADEE-H3(1-15)K9me1	SAWADEE-H3(1-15)K4me1K9me1
PDB code	4IUR	4IUT	4IUU	4IUV
Beamline	NSLS-X29A	APS-24ID-E	APS-24ID-E	APS-24ID-E
Wavelength (Å)	1.0750	0.9792	0.9792	0.9792
Space group	$P2_1$	$P2_1$	$P2_12_12$	$P2_12_12$
Cell dimensions				
a, b, c (Å)	53.3, 56.2, 58.9	53.4, 56.7, 59.1	54.8, 118.1, 53.4	118.4, 54.5, 53.1
β (°)	93.0	93.2	90	90
Resolution (Å)	50.0-2.5 (2.59-2.50)[a]	50.0-2.7 (2.80-2.70)[a]	50.0-2.7 (2.80-2.70)[a]	50.0-2.8 (2.90-2.80)[a]
R_{merge} (%)	9.7 (54.1)	14.9 (56.0)	8.7 (41.0)	12.3 (51.4)
Average I/σ(I)	32.1 (4.7)	9.6 (1.9)	14.9 (2.0)	14.8(3.0)
Completeness (%)	99.9 (100.0)	99.3 (99.3)	99.5 (100.0)	100.0 (100.0)
Redundancy	6.0 (6.2)	3.5 (3.5)	3.8 (3.8)	5.9 (6.1)
Refinement and structure model				
Resolution (Å)	50.0-2.5 (2.59-2.50)[a]	50.0-2.7 (2.80-2.70)[a]	50.0-2.7 (2.80-2.70)[a]	50.0-2.8 (2.90-2.80)[a]
Unique reflections	12,197	9,744	10,043	8,999
R / R_{free} (%)	21.5 / 25.6	20.2 / 24.8	20.3 / 26.9	20.4 / 25.3
Number of atoms	2353	2379	2309	2313
Protein / peptide	2181 / 66	2190 / 65	2134 / 64	2152 / 72
Zn^{2+} ion	2	2	2	2
Ligand	44	44	55	44
Water	60	78	54	43
B factors (Å^2)	50.2	36.5	47.8	43.7
Protein / peptide	49.0 / 70.3	35.5 / 52.1	46.3 / 73.1	46.3 / 61.8
Zn^{2+} ion	49.1	30.8	52.6	41.3
Ligand	85.5	79.1	90.9	79.3
Water	46.3	27.9	34.9	36.9
RMS deviations				
Bond lengths (Å)	0.005	0.008	0.004	0.005
Bond angles (°)	0.946	1.538	0.734	1.002

[a] Values in parentheses are for highest-resolution shell.