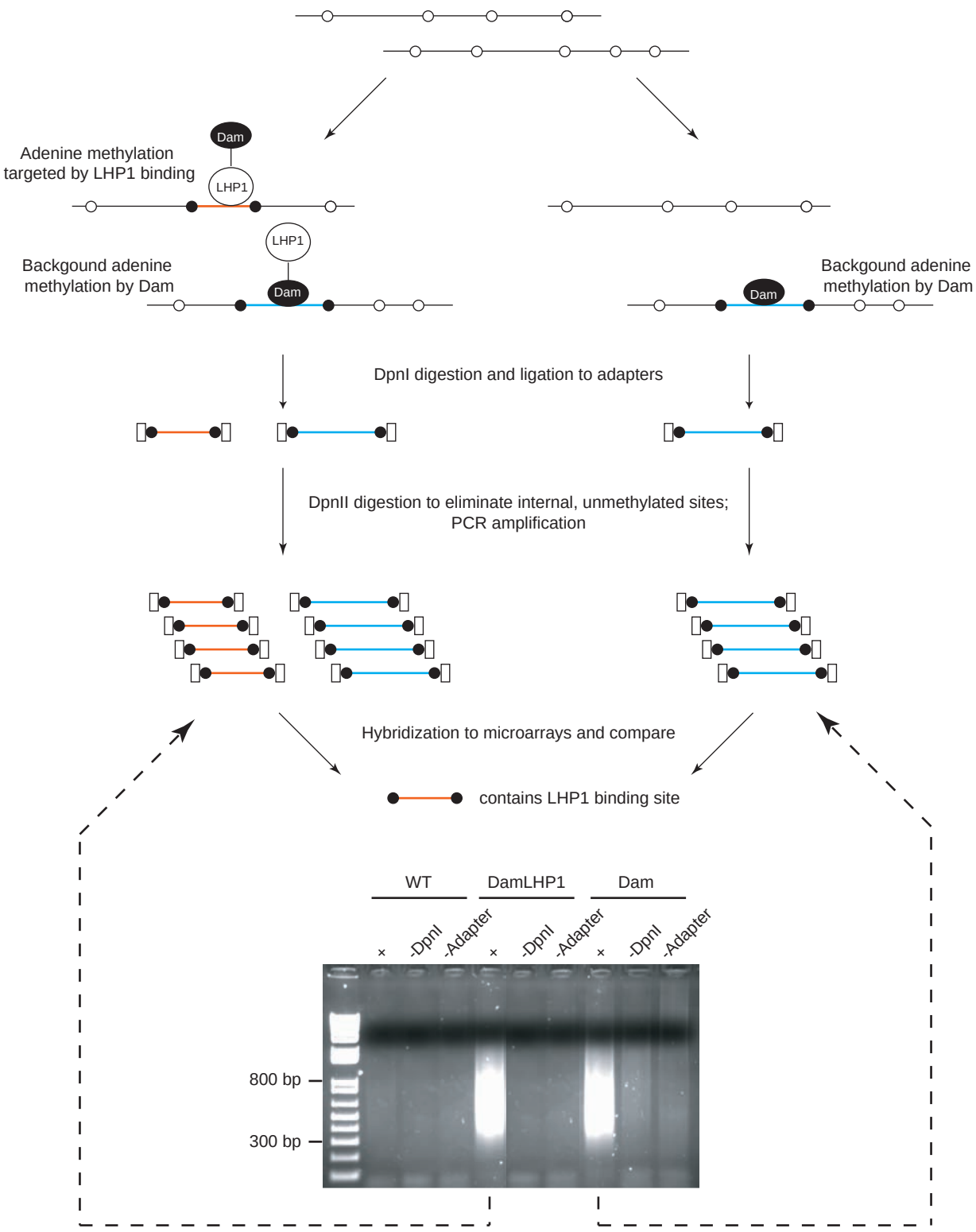
Supplementary Figures, Table, and Methods

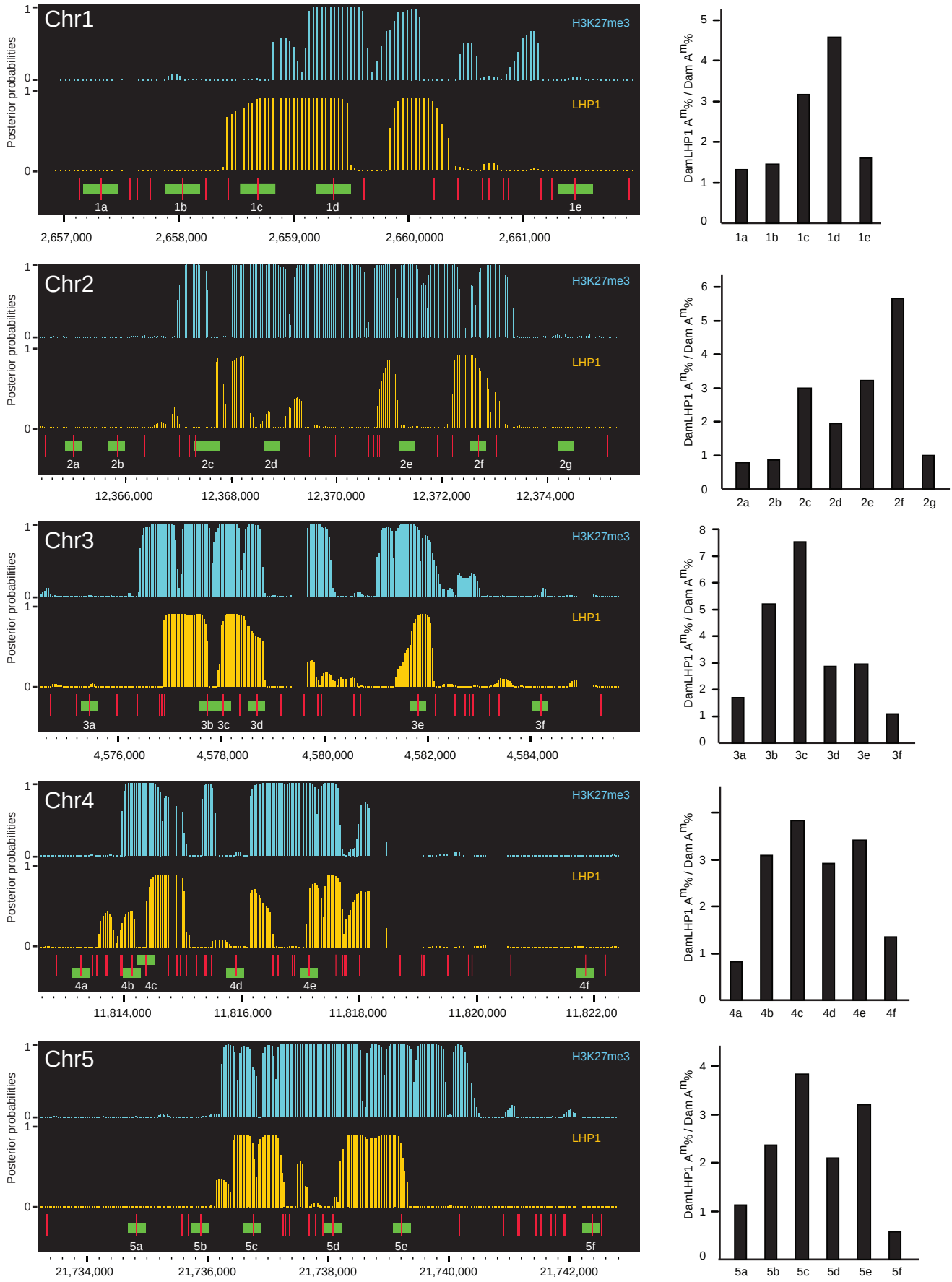
The *Arabidopsis* chromodomain-containing protein LHP1 colocalizes with histone H3 lysine27 trimethylation

Xiaoyu Zhang, Sophie Germann, Bartlomiej J. Blus,

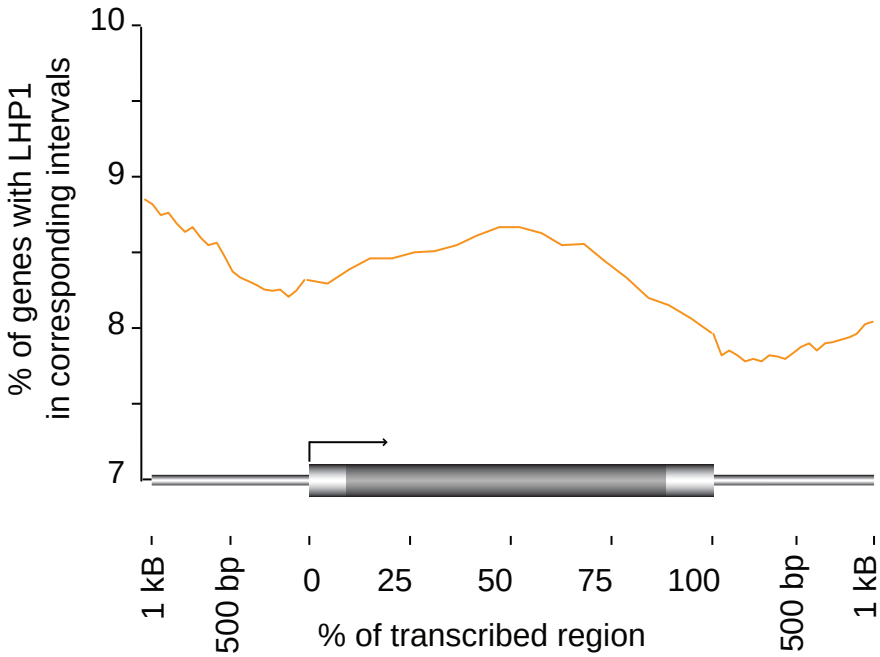
Sepideh Khorasanizadeh, Valerie Gaudin and Steven E. Jacobsen



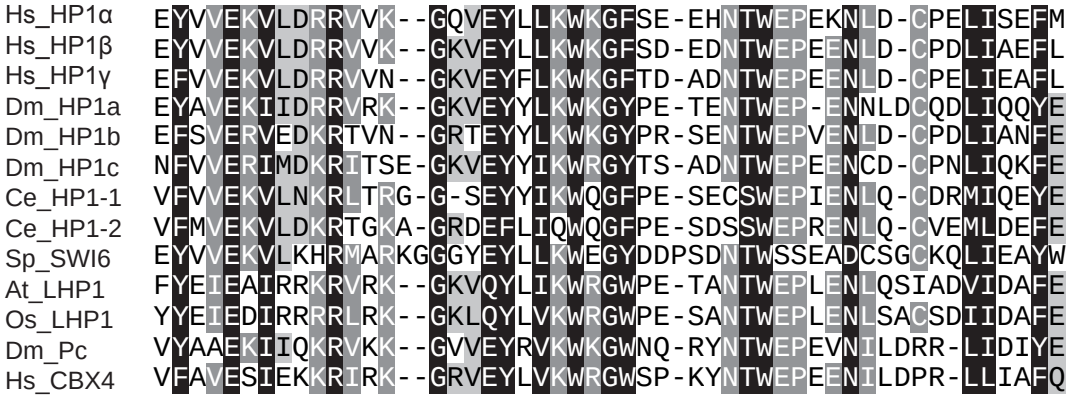
Supplementary Figure 1 The DamID-chip method. A schematic representation of the DamID-chip method is shown on top. Horizontal lines represent genomic DNA, open circles represent unmethylated GATC sites, filled circles represent adenine-methylated GATC sites, and open boxes represent adapters. The detailed procedures are described in **Supplemental Methods**. Methylation-specific PCR results are shown at the bottom. WT: DNA from wild type plants without transgene; DamLHP1: DNA from plants expressing the DamLHP1 fusion protein; Dam: DNA from plants expressing the Dam protein only. No PCR amplification occurred in control experiments either without DpnI digestion ("-DpnI") or without adapter ligation ("-Adapter").



Supplementary Figure 2 Validation of DamID-chip results. The experimental procedures are described in **Supplementary Methods**. Left panels: LHP1 localization results from DamID-chip (orange). H3K27me3 is shown in blue. Vertical red bars represent individual DpnI/DpnII sites and horizontal green bars represent regions validated. Right panels: validation results of corresponding regions. *y*-axis: ratio of the percentage of adenine methylation in DamLHP1 samples and Dam-only control samples.



Supplementary Figure 3 LHP1 binding sites are enriched in the transcribed regions of genes. Each of the 25,423 expressed *Arabidopsis* genes [thick horizontal bar, shown as the average length of *Arabidopsis* genes (2.5 kb)] was divided into 20 intervals (5% each interval), and the 1 kb regions upstream and downstream of each gene (thin horizontal bars) were divided into 50 bp intervals. The percentage of genes containing LHP1 binding sites in each interval was graphed (*y*-axis).



Supplementary Figure 4 The LHP1 CD is highly similar to members of both the HP1 and Pc families. The CLUSTALW multiple alignment of the CDs from LHP1, HP1 and Pc families includes the following sequences. Hs_HP1α: *H. sapiens* (gi:6912292); Hs_HP1β: *H. sapiens* (gi:48428808); Hs_HP1γ: *H. sapiens* (gi:5732187); Dm_HP1a: *D. melanogaster* (gi:17136528); Dm_HP1b: *D. melanogaster* (gi:24640713); Dm_HP1c: *D. melanogaster* (gi:21355721); Ce_HP1-1: *C. elegans* (gi:17568757); Ce_HP1-2: *C. elegans* (gi:71987888); Sp_SWI6: *S. pombe* (gi:510930); At_LHP1: *A. thaliana* (gi:110810410); Os_LHP1: *O. sativa* (gi:110810411); Dm_PC: *D. melanogaster* (gi: 129718); Hs_CBX4: *H. sapiens* (gi:15929016).

Supplementary Table 1 Primers used for real-time PCR validation and sites validated

Chr	DnpI/DpnII site	forward primer	reverse primer
1	2657326 - 2657329	GCTGTCATTGGCTCCAAAA	GAGATGCTAATTCCTGCACCA
1	2658037 - 2658040	GTTTTGAACCTGATGATGTATTGG	AGTCTGCCACACACCAAAAAC
1	2658694 - 2658697	GAAGTGCAACGCACCACTAA	CCTTTTGATGGTCACTCAGC
1	2659354 - 2659357	ATTACCATGCGTTTCCAAGC	ATATCAGCGCGAACGTCAAT
1	2661458 - 2661461	GACAAGTCCGGCAAAAACTC	CGATTACAATGACGCTTTGG
2	12365003 - 12365006	GGTTCATTGCACAGCAAAAG	GATTCCATCTCTGGGTGGAC
2	12365834 - 12365837	CATGAACCGCTGATAATCCA	CGTGCAGAAGAGTGGTGGTT
2	12367548 - 12367551	CAAGAATCATCCACCGTTCC	GAACTCTTAATCGTTCGTTGTCA
2	12368776 - 12368779	TGGGATTATGAAAAGGTTTGCT	TTCCCTAACGATTTTGTTGTATG
2	12371339 - 12371342	CCACCACCATGTTGAGACAG	GGCTGCGTGAGGTTGTAGTT
2	12372702 - 12372705	AGGAGCTTTGCCTTTTTCTTT	ACACATGTAAAAAGTAAGGGTTGTT
2	12374360 - 12374363	CGAACAGTTAAACCACGTCCT	TCAGCGATATCCCAAGAAGAA
3	4575447 - 4575450	TGGCCTTGAGGCTTTTAGTC	CACAAAGTGACATACCGAGCA
3	4577734 - 4577737	AGAGCAGCTTCTTGAGGAGAGA	ACATGGGTTTTGTTTCTTGAAT
3	4578051 - 4578054	TCACCTTTTGCTTGTTGCAG	AAAGTTTCTGCAGCCGTTTC
3	4578694 - 4578697	TGTTGCCCATGCAGACAC	TTCCTTTGTTGGCAATCCTT
3	4581831 - 4581834	CGAGCTCAAAATAAGCTGTGG	CAGACAAAGATAAACAGCTACGG
3	4584190 - 4584193	TTGGGCCTGCAGAGATAATAA	TGGTAGGGAATTGGATTGCT
4	11813272 - 11813275	CCAATTTGGACTATTTGACATCTCT	CAGTATGGTCTTTTGCAATTTGA
4	11814142 - 11814145	CGCATTGTTAATCTCCAGCA	GCTGCGTTTGTTTGATGTGT
4	11814372 - 11814375	CAAACGCAGCCATCTAACAC	GTCGTGGCTATGGTGTGGTT
4	11815907 - 11815910	TCAAGCCAATTCAAGTCCATT	CAACAAACCCTTGGTGGATG
4	11817148 - 11817151	AAGGCCTTGGTTGCTTTACC	GAAAGTTTGGTGGGTTTTGC
4	11821859 - 11821862	AGTCTTCAACGCTCCTTCGT	GAGTTGCGTCAGTTGGTTGA
5	21734835 - 21734838	CCATTATTTGATTTGGCTCGAA	ATCAGTTGCATTTGGCGTTT
5	21735897 - 21735900	CGTTTAGTAGGACAACGAAAACC	CTTAGGCGCACATACCCAAT
5	21736762 - 21736765	GAAAGTTGAAGTGCGAAACCA	TCCATAATTCTACTAAATCCCAAAA
5	21738076 - 21738079	ATGGTCAAGAGATGAACTTCCAC	TGTAAATGTTTGTGAACCAAGTGA
5	21739228 - 21739231	TATCAACCTTGCGTCCCAAC	TGTTTTAATTTTGGAGAAAGAATGG
5	21742368 - 21742371	TGAGAACATCTTCTACAAATGTAAGGT	TGGTTGAATGTTAGGAAATTGTTG

Supplementary Methods

Plant material

The *Arabidopsis thaliana* transgenic lines DamLHP1 x AlcR (DamLHP-1 x AlcR4 and DamLHP-2 x AlcR5) and Dam x AlcR (293Dam-3 x AlcR4 and Dam-1 x AlcR5) were described[1]. The expression level of the *DamLHP1* gene is approximately 1.5 fold lower than the endogenous *LHP1* gene as assayed by real-time RT-PCR (not shown). Plants were grown on soil, in a growth chamber under long-day (16 h light/8 h dark) conditions, at 23°C/15°C (day/night) and rosette leaves of five week old plants were harvested.

PCR amplification of Dam-methylated DNA

Genomic DNA extraction and quantification were performed as described[1]. A detailed protocol for selective PCR amplification of methylated DNA fragments can be found at http://www.nki.nl/nkidep/vansteensel[2], and a schematic representation of the procedure is shown in **Supplementary Figure 1** on line. Briefly, 2.5 µg of genomic DNA was digested for 16 h at 37°C with 10 units of DpnI (New England Biolabs, Beverly, MA, USA), which only cuts at adenine-methylate sites. After inactivation of DpnI at 80°C for 20 min, the DpnI-digested genomic DNA was ligated to 40 pmol of the double-stranded adaptor AdR (5'-CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGATCCTCG GCCG-3') for 2 h at 16°C with 5 units of T4 ligase (Roche Molecular Biochemicals, Mannheim, Germany). The enzyme was heat inactivated for 10 min at 65°C. To prevent amplification of DNA fragments bearing unmethylated

GATC sites, DNA was incubated for 1 h at 37°C with 10 units of the adenine methylation-sensitive enzyme DpnII (New England Biolabs, Beverly, MA, USA). PCR amplification was then performed with 0.5 μg of DpnII-digested DNA using the BD Advantage 2 PCR Enzyme System (reference 639201, Clontech Palo Alto, CA, USA) and the AdR_PCRprimer (5'-GGTCGCGGCCGAGGATC-3'). The PCR conditions were: 10 min at 68°C; one cycle of 1 min 94°C, 5 min 65°C and 15 min 68°C; 3 cycles of 1 min 94°C, 1 min 65°C and 10 min 68°C; and 17 cycles of 1 min 94°C, 1 min 65°C and 2 min 68°C. The PCR products were purified using the QIAquick PCR purification columns (QIAGEN) according to the manufacturer's recommendations.

Microarray experiments and data analysis

The Affymetrix whole-genome tiling microarrays used in this study cover ~97% of the *Arabidopsis* genome with one 25-nt oligo representing each 35-bp genomic region[3]. Microarray hybridization, washing, staining and scanning were performed as described[3]. 30 μg of adenine methylation-specific PCR products were used per replicate, and four biological replicates were performed for DamLHP1 x AlcR and Dam x AlcR plants. All microarray data were submitted to Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under the series accession number GSE8169. Raw microarray data from oligo probes that mapped to unique locations in the genome (representing ~90% of all probes) were quantile normalized and analyzed using Tilemap with the Hidden Markov model option as described[3,4]. DpnI restriction fragments containing putative LHP1 binding

sites were identified as those significantly preferentially amplified in the DamLHP1 samples compared to Dam only control samples (i.e., those overlapping with oligo probes yielding posterior probabilities of 0.5 or higher). In addition, all isolated single positive DpnI restriction fragments (with no other positive fragments located within 1 kb on either side) were excluded from further analysis as they might represent false-positives caused by spurious PCR amplification, and genomic regions that included multiple closely spaced positive fragments were summarized into LHP1 binding regions by allowing a maximal of 1-kb gaps. Under these conditions, no LHP1 binding region was detected in the chloroplast or mitochondrial genomes that were also represented on the microarrays, indicating a low false discovery rate. For each LHP1 binding region, a region of the same length was randomly selected within 5 kb upstream or downstream; these randomly selected regions resembled LHP1 binding regions in length and chromosomal distributions and were referred to as "control regions". The DNA methylation and siRNAs data used for comparison were previously published[3,5,6].

The identification of LHP1 binding regions by DamID-chip requires the presence of suitable DpnI sites in the region being tested. In addition, DpnI restriction fragments that are too short (< 300 bp) or too long may not be amplified and recovered efficiently during the ligation-based PCR step, resulting in a higher level of false negative rate in genomic regions that contain only short or long DpnI restriction fragments. Considering these limitations, we focused on H3K27me3 regions that were 1 kb or longer because it is more likely that these regions contain suitable DpnI sites. We thus re-analyzed the H3K27me3 localization using previously published

data[3,5,6], and derived 2,632 regions by requiring a minimal length of 1 kb and allowing a maximal gap of 1 kb.

A previous microarray study of gene expression in the *lhp1* mutant identified 87 up-regulated genes[7], of which 28 (32.2%) were found to be bound by LHP1 here. Considering 3,261 of the 25,407 (12.8%) expressed genes were found to be bound by LHP1, we concluded that LHP1 targets were >2.5 fold overrepresented in the up-regulated genes in *lhp1* ($p<10^{-5}$ according to a ypergeometric cumulative distribution function).

Validation of DamID-chip results

DamID-chip results were validated using the method described in Germann *et al*[1]. DNA samples from plants expressing DamLHP1 or Dam only were digested with DpnII, which is sensitive to adenine methylation (i.e., cutting is blocked by adenine methylation). For each DNA sample, real-time PCR was performed with primers flanking a selected DpnII restriction site, using either undigested genomic DNA or DpnII-digested DNA as template. The same amount of DNA was used for each reaction (15 ng), and the methylation level at a site was estimated as the ratio between the amplified quantity from the remaining uncut DNA in the DpnII-digested sample and the quantity amplified from the undigested sample. Real-time PCR reactions were performed using the iQ™ SYBR Green Supermix (BIO-RAD, Hercules, CA). The sites validated and the primers utilized are listed in **Supplementary Table 1**. The PCR parameters were: 1 cycle of 2 minutes at 95℃, 40 cycles of 15 seconds at 95℃, 30 seconds at 60℃ and 30 seconds at 72℃, 1 cycle of 1 minute at 95℃. The adenine methylation level at each site was then compared between

plants expressing DamLHP1 or Dam only to take into account chromatin accessibility differences at each site[1]. A similar level of adenine methylation in the two backgrounds indicates the lack of LHP1 binding, whereas a higher level of adenine methylation in the DamLHP1 background suggests that LHP1 is targeted to the vicinity of this site. This was done for 30 sites selected from the 5 chromosomes of *Arabidopsis* that were found to be either positive or negative for LHP1 binding by DamID-chip, and the results are shown in **Supplemental Figure 2**.

In vitro binding assays

The LHP1 chromodomain (residues 104-160) was amplified by PCR and cloned into the BamHI/NdeI sites of the pET11a vector (Novagen) to encode a fusion protein with N-terminal His-tag. The fusion protein was expressed in *E. coli* strain BL21(DE3) (Novagen), and purified by Ni^{2+}-affinity chromatography (QIAGEN). Protein concentration was determined by absorbance spectroscopy using predicted extinction coefficient ε_{280} = 19,630 $M^{-1}cm^{-1}$. Peptide concentrations were determined using absorbance spectroscopy (extinction coefficient for tyrosine, ε_{280} = 1,280 $M^{-1}cm^{-1}$; extinction coefficient for fluoresceinated peptides ε_{492} = 68,000 $M^{-1}cm^{-1}$). Fluorescence polarization binding assays were performed under conditions of 50 mM sodium phosphate, 25 mM NaCl (pH 7.8) at 15°C, and in the presence of 150 nM fluorescein-labeled peptide following a previously described protocol[8]. Data were obtained using a Beacon 2000 device. Titration binding curves were analyzed using KaleidaGraph (Synergy Software) as previously described[8]. The peptides used in the binding assays were H3K9me3

[ARTKQTARK(me3)STGGY], H3K9me2 [ARTKQTARK(me2)STGGY], H3K9me1 [ARTKQTARK(me1)STGGY], H3K27me3 [APRKQLATKAARK(me3)SAPSTY], and H3K9me3K27me3 [ARTKQTARK(me3)STGGKAPRKQLATKAARK(me3)SAPATGGVKKPHRY].

References Cited:

1. Germann, S., Juul-Jensen, T., Letarnec, B. & Gaudin, V. *Plant J* **48**, 153-63 (2006).
2. Greil, F. *et al*. *Genes Dev* **17**, 2825-38 (2003).
3. Zhang, X. *et al*. *Cell* **126**, 1189-201 (2006).
4. Ji, H. & Wong, W.H. *Bioinformatics* **21**, 3629-36 (2005).
5. Lu, C. *et al*. *Science* **309**, 1567-9 (2005).
6. Zhang, X. *et al*. *PLoS Biology* **5(5)**, e129 (2007).
7. Nakahigashi, K., Jasencakova, Z., Schubert, I. & Goto, K. *Plant Cell Physiol* **46**, 1747-56 (2005).
8. Jacobs, S.A., Fischle, W. & Khorasanizadeh, S. *Methods Enzymol* **376**, 131-48 (2004).