nature genetics

A structural-maintenance-of-chromosomes hinge domain–containing protein is required for RNA-directed DNA methylation

Tatsuo Kanno[1,5], Etienne Bucher[1,4,5], Lucia Daxinger[1], Bruno Huettel[1], Gudrun Böhmdorfer[1], Wolfgang Gregor[2], David P Kreil[3], Marjori Matzke[1] & Antonius J M Matzke[1]

RNA-directed DNA methylation (RdDM) is a process in which dicer-generated small RNAs guide *de novo* cytosine methylation at the homologous DNA region[1,2]. To identify components of the RdDM machinery important for *Arabidopsis thaliana* development, we targeted an enhancer active in meristems for methylation, which resulted in silencing of a downstream *GFP* reporter gene. This silencing system also features secondary siRNAs, which trigger methylation that spreads beyond the targeted enhancer region. A screen for mutants defective in meristem silencing and enhancer methylation retrieved six *dms* complementation groups, which included the known factors DRD1 (ref. 3; a SNF2-like chromatin-remodeling protein) and Pol IVb subunits[4,5]. Additionally, we identified a previously unknown gene *DMS3* (At3g49250), encoding a protein similar to the hinge-domain region of structural maintenance of chromosomes (SMC) proteins. This finding implicates a putative chromosome architectural protein that can potentially link nucleic acids[6] in facilitating an RNAi-mediated epigenetic modification involving secondary siRNAs and spreading of DNA methylation.

In plants, small RNAs directed to promoter regions can trigger promoter methylation and transcriptional gene silencing[1]. In the two-component transgene system used here, this strategy has been modified to target an upstream enhancer that is active in meristems for methylation by a homologous hairpin RNA, which is encoded at an unlinked silencer locus (**Fig. 1a**). In the absence of the silencer locus, *GFP* expression can be observed in the root apical meristem of seedlings, whereas fluorescence is abolished in the presence of the silencer (**Fig. 1b**).

Through analysis of DNA methylation using bisulfite sequencing, we found that the target enhancer completely lacks cytosine methylation in the absence of the silencer locus (data not shown) but acquires heavy methylation of cytosines in all sequence contexts in the presence of the silencer (**Fig. 2**, T + S (target + silencer)). Unexpectedly, we also detected a high degree of methylation in the DNA sequence downstream of the targeted enhancer region in silenced plants (**Fig. 2**, T + S). This transcriptional silencing system therefore differs from several other systems in which methylation does not infiltrate significantly into nontargeted sequences[3,7].

The spreading of methylation beyond the originally targeted region suggests the presence of secondary siRNAs, which form adjacent to the target sites of primary siRNAs during RNAi in plants and *Caenorhabditis elegans*[8,9]. Secondary siRNAs depend on RNA-dependent RNA polymerase (RDR) for their biogenesis. Indeed, we detected 24-nt secondary siRNAs that are homologous to sequences downstream of the targeted enhancer, but only in silenced plants (**Fig. 3a**, lane T + S), which also contained the anticipated 21-, 22- and 24-nt primary siRNAs derived from the hairpin-RNA trigger (**Fig. 3b**, T + S). Of note, the appearance of secondary siRNAs was correlated with the substantial reduction of the expression of one or more transcripts (referred to hereafter as nascent RNAs) that were unexpectedly present in wild-type plants containing the target locus only (**Fig. 3**). Through sequence analysis, we found that the nascent RNAs include alternatively spliced transcripts, at least some of which likely initiate in the enhancer region and extend throughout the entire *GFP* coding sequence (**Supplementary Fig. 1** online). We did not anticipate the synthesis of these transcripts because a related viral enhancer did not show intrinsic promoter activity[10]. The data are consistent with a hypothetical model in which secondary siRNAs that induce methylation of the downstream region result from primary siRNA–induced turnover of an overlapping nascent RNA(s), perhaps through the sequential action of Argonaute (AGO), RDR and DICER-LIKE3 (DCL3) activities (**Fig. 4**). Whether secondary siRNAs and methylation spreading are necessary for silencing *GFP* expression, or whether they are simply by-products of this transgene system, is discussed below.

To obtain mutants defective in meristem silencing, we used ethyl methanesulfonate to mutagenize seeds of a plant doubly homozygous

[1]Gregor Mendel Institute for Molecular Plant Biology, Austrian Academy of Sciences, Dr. Bohrgasse 3, A-1030 Vienna, Austria. [2]Research Group of Molecular Pharmacology and Toxicology, Department of Science, University of Veterinary Medicine, A-1210 Vienna, Austria. [3]Department of Bioinformatics, Boku University Vienna, Muthgasse 18, A-1190 Vienna, Austria. [4]Present address: Laboratory of Plant Genetics, University of Geneva, CH-1211 Geneva 4, Switzerland. [5]These authors contributed equally to this work. Correspondence should be addressed to M.M. (marjori.matzke@gmi.oeaw.ac.at).

Received 5 September 2007; accepted 11 February 2008; published online 20 April 2008; doi:10.1038/ng.119

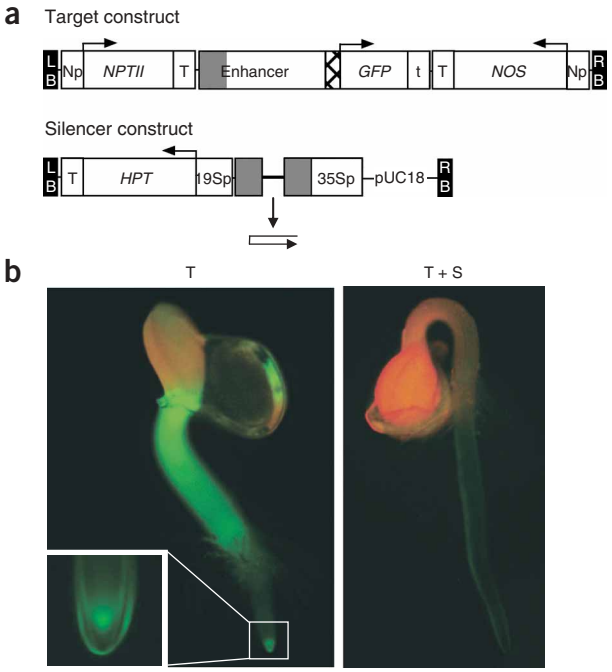

Figure 1 Transgene constructs and silenced phenotype. (**a**) The target DNA sequence is a ~300-bp segment (gray block) of a ~1.2-kb viral enhancer[24] positioned upstream of a minimal promoter (hatched region) and the coding sequence of enhanced green fluorescent protein (*GFP*). We targeted this region because it is essential for enhancer activity in *Arabidopsis*. The entire DNA sequence is shown in **Supplementary Figure 1**. In the silencer construct, the 35S promoter of cauliflower mosaic virus (35Sp) drives expression of an inverted DNA repeat of target enhancer sequences, producing a hairpin RNA trigger. Arrows indicate direction of transcription. LB, RB, left and right T-DNA borders, respectively; NOS, nopaline synthase; Np, NOS promoter; *NPTII*, neomycinphosphotransferase for selection of transformed plants on kanamycin-containing medium; T, NOS transcription terminator; t, 35S transcriptional terminator; 19Sp, 19S promoter from cauliflower mosaic virus; *HPT*, hygromycin phosphotransferase for selection of transformed plants on hygromycin-containing medium. (**b**) Germinating seedlings (seed coats still visible) showing GFP fluorescence in the root apical meristem in the unsilenced target line (T) and loss of fluorescence after introduction of the silencer construct (T + S). Strong GFP fluorescence in the hypocotyl is also observed in young seedlings of the target line. Cotyledons appear red because of chlorophyll autofluorescence at the excitation wavelength for GFP.

for both the target and silencer complexes and then screened seedlings of the M2 generation (the first generation when a recessive mutation can be homozygous) for recovery of green fluorescence in the root apical meristem. By screening approximately 125,000 M2 seedlings, we identified 130 mutants, which we then confirmed by PCR analysis to contain the silencer complex (data not shown). Intercrosses among approximately a quarter of these plants identified at least six *dms* complementation groups. We report here on *dms1*, *dms2*, *dms3* and *dms5*.

Through bisulfite sequencing, we found that in *dms1*, *dms2* and *dms5* mutants, all methylation is lost from the targeted enhancer as well as from the downstream region (data not shown). Very low levels of methylation remain in a *dms3* mutant, predominantly in the enhancer sequence targeted by primary siRNAs (**Fig. 2**, *dms3*). We observed primary siRNAs and one or more nascent transcripts in all four mutants; however, we did not detect secondary siRNAs, resulting in a unique RNA profile in the mutants (**Fig. 3**).

We mapped[11,12] the *dms1* mutation to a region of chromosome 2 that encodes DRD1 (At2g16390), a putative SNF2-like chromatin-remodeling protein identified in a previous forward screen for mutants defective in RdDM[3]. Through sequencing, we confirmed that *dms1* is indeed *drd1* (**Supplementary Fig. 2a** online). Given the identical methylation patterns of the targeted enhancer in *dms1*, *dms2* and *dms5* mutants, as well as the reduction of endogenous siRNAs in a *dms2* mutant (**Fig. 3c**,**d**, lane *dms2*), and the mapped location of the *dms5* mutation on the bottom arm of chromosome 2 (data not shown), we predicted that *DMS2* and *DMS5* would encode the Pol IVb subunits NRPD2a (At3g23780) and NRPD1b (At2g40030), which we recovered previously together with DRD1 in the same mutant screen[4]. Indeed, we sequenced *dms2* and *dms5* and found that they correspond to *nrpd2a* and *nrpd1b*, respectively (**Supplementary Fig. 2b**,**c**).

We mapped the *dms3* mutation to a region of chromosome 3 that did not contain any known silencing factors. Fine mapping defined a genetic interval that contained several possible candidate genes. On the basis of co-expression data with DRD1 (ref. 13), we sequenced one candidate, At3g49250, and found independent mutations in three alleles (**Supplementary Fig. 2d**). *DMS3* encodes a previously uncharacterized, 420 amino acid polypeptide that shows similarity to the hinge domain region of structural maintenance of chromosome

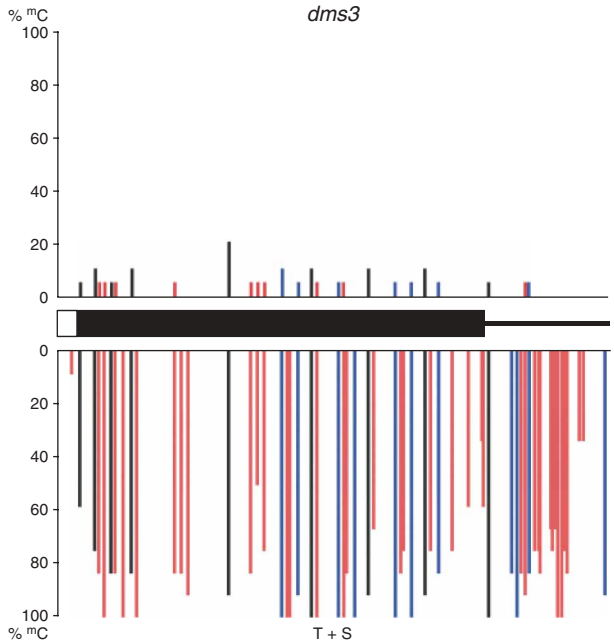

Figure 2 Bisulfite sequence analysis of DNA methylation. The ~300-bp targeted enhancer region and the sequence extending ~80 bp downstream are depicted by the bold and narrow black bars, respectively. Percent methylation of cytosines in different sequence contexts (CG, black; CNG, blue; CNN, red; where N is A, T or C) is indicated for wild-type plants (T + S) and a *dms3* mutant (*dms3-1*). The original data are shown in **Supplementary Figure 7** online. Analysis with methylation-sensitive restriction enzymes failed to detect methylation in the *GFP* coding region or the associated 35S transcription terminator (data not shown). Additional bisulfite sequencing in wild-type (T + S) plants suggested that methylation tapers off ~150 bp downstream of the targeted enhancer (data not shown).

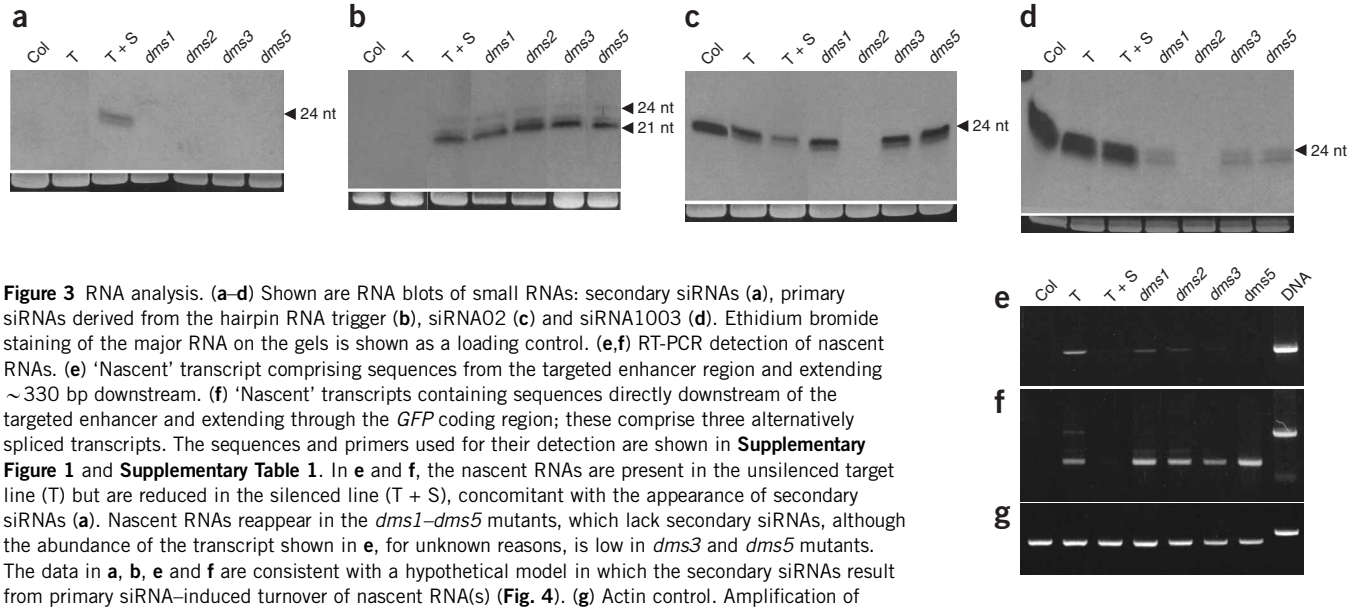

Figure 3 RNA analysis. (**a**–**d**) Shown are RNA blots of small RNAs: secondary siRNAs (**a**), primary siRNAs derived from the hairpin RNA trigger (**b**), siRNA02 (**c**) and siRNA1003 (**d**). Ethidium bromide staining of the major RNA on the gels is shown as a loading control. (**e**,**f**) RT-PCR detection of nascent RNAs. (**e**) 'Nascent' transcript comprising sequences from the targeted enhancer region and extending ~330 bp downstream. (**f**) 'Nascent' transcripts containing sequences directly downstream of the targeted enhancer and extending through the *GFP* coding region; these comprise three alternatively spliced transcripts. The sequences and primers used for their detection are shown in **Supplementary Figure 1** and **Supplementary Table 1**. In **e** and **f**, the nascent RNAs are present in the unsilenced target line (T) but are reduced in the silenced line (T + S), concomitant with the appearance of secondary siRNAs (**a**). Nascent RNAs reappear in the *dms1–dms5* mutants, which lack secondary siRNAs, although the abundance of the transcript shown in **e**, for unknown reasons, is low in *dms3* and *dms5* mutants. The data in **a**, **b**, **e** and **f** are consistent with a hypothetical model in which the secondary siRNAs result from primary siRNA–induced turnover of nascent RNA(s) (**Fig. 4**). (**g**) Actin control. Amplification of genomic DNA is shown in the last lanes. It is not known how to distinguish endogenous primary from endogenous secondary siRNAs, but reductions in endogenous RNA populations in *dms1*, *2*, *3* and *5* mutants (for example, siRNA1003 in **d**) might reflect loss of specifically secondary siRNAs. Col, nontransgenic, wild-type plants.

(SMC) proteins[14,15], including short coiled-coil regions on either side of the hinge (**Fig. 5**). SMC proteins, considered dynamic molecular linkers of the genome[6], are chromosomal ATPases that are highly conserved from bacteria to humans[16]. Authentic SMC proteins are much larger than DMS3, and they contain a central hinge surrounded by two long coiled-coil domains and N- and C-terminal ATP binding sites. In eukaryotes, six distinct SMC proteins form different heterodimers as part of larger complexes involved in sister chromatid cohesion (SMC1-SMC3), chromosome compaction and segregation (SMC2-SMC4) and DNA repair and checkpoint responses (SMC5-SMC6)[6,16].

DMS3 shows strong sequence similarity only to other plant genes, in particular, a known ortholog in *Medicago truncatula* and the unannotated ortholog Os01g0235200 in *Oryza sativa* (japonica cultivar-group) (**Supplementary Fig. 3** online). The protein encoded by one mutant allele, *dms3-1*, sustained glycine to glutamic acid conversion at position 339 (**Supplementary Fig. 2d**). This conserved glycine is at the edge of the hinge domain in a consensus sequence that is thought to impart flexibility to SMC proteins (**Fig. 5** and **Supplementary Fig. 3**)[15]. A previous study showed that the hinge region of an SMC1-SMC3 heterodimer together with a ~20 amino acid transition sequence into the coiled-coil domain is sufficient for dimerization and DNA binding[15]. Thus, DMS3 could potentially dimerize and bind DNA. Indeed, an initial yeast two-hybrid experiment suggests that DMS3 can form homodimers (**Supplementary Fig. 4** online).

To determine whether typical endogenous targets of RdDM are also affected in a *dms3* mutant, we analyzed methylation of 5S rDNA and 180-bp centromeric repeats[3,4]. The digestion patterns in *dms3* plants are indistinguishable from those observed in *drd1*, *nrpd1b* and *nrpd2a* mutants (**Supplementary Fig. 5a** online). Moreover, other targets of DRD1 and Pol IVb (comprising NRPD1b and NRPD2A), including a solo LTR that drives expression of the *IG/LINE* transcript[17], lose methylation (**Fig. 6a**) and/or are derepressed in a *dms3* mutant (**Fig. 6b** and **Supplementary Fig. 5b**).

Although we designed this silencing system to identify developmentally important genes needed for RdDM in meristems, *DMS3* expression is only moderately elevated in shoot and root apices compared to other cell and organ types (see URLs section below). In addition, similarly to the *drd1*, *nrpd2a* and *nrpd1b* mutants, the *dms3* mutants isolated in this study do not show overt morphological phenotypes when grown under standard conditions (data not shown).

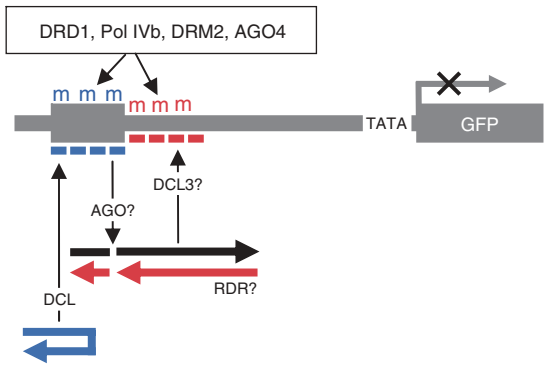

Figure 4 Hypothetical model for production of secondary siRNAs. Primary siRNAs (blue dashes) derived from dicer (DCL) processing of the hairpin RNA trigger induce primary RdDM (blue 'm') at the originally targeted enhancer sequence. Primary siRNAs are also postulated to guide Argonaute (AGO) cleavage of an overlapping nascent RNA(s) (black arrow), producing substrates for RNA-dependent RNA polymerase (RDR). Processing of the resulting double-stranded RNA (black and red arrows) by DCL3 produces 24-nt secondary siRNAs (red dashes), which then trigger secondary RdDM (red 'm') at the downstream region. Primary and secondary RdDM are both presumably catalyzed by the *de novo* methyltransferase DRM2 (ref. 2) and require AGO4 (ref.2), DRD1 and Pol IVb (refs. 3,4; this study). *GFP* expression is silenced, but whether this requires both primary and secondary RdDM is not known. The postulated AGO and RDR proteins required for secondary siRNA biogenesis remain unidentified.

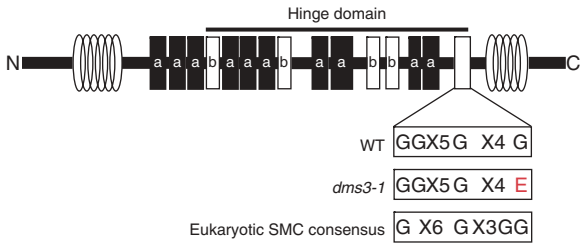

Figure 5 Domain structure of DMS3 (At3g49250). This protein has an interesting domain architecture, comprising only the SMC_hinge domain (pfam06470), indicated by the black bar, flanked by short coiled-coil regions (overlapping ovals). It is notably lacking the SMC_N domain (with an ATP binding site) of bonafide SMC proteins. There is evidence, however, suggesting that this domain structure forms active homodimers that can bind DNA[15]. DMS3 contains a conserved G pattern (box) that is similar to a consensus sequence in eukaryotic SMC proteins[15]. One mutated allele (*dms3-1*; **Supplementary Fig. 2**) contains a glycine (G) to glutamic acid (E) substitution in this region, which is thought to be essential for formation of an active dimer[15]. a, alpha helix; b, beta sheet.

These results, together with the finding that typical endogenous targets of RdDM are derepressed and/or lose methylation in a *dms3* mutant, suggest that DMS3 is a general component of the RdDM machinery. It is not yet known how DMS3 functions in this pathway, but given its relatively small size, we speculate that homodimers[14] of DMS3 bind and stabilize siRNA-DNA and/or siRNA-RNA complexes (**Supplementary Fig. 6a** online). Although primarily known for their roles during mitosis, SMC proteins are increasingly implicated in interphase functions, including diverse gene silencing phenomena[18,19]. Notably, SmcHD1, a mammalian SMC hinge domain–containing protein that shows sequence similarity to DMS3 in the hinge-domain region, is important for maintaining X-chromosome inactivation in mice (M. Blewitt and E. Whitelaw, Queensland Institute of Medical Research, personal communication).

It is unclear why DMS3 was recovered in this screen but not in a previous one that also identified DRD1, NRPD2a and NRPD1b[3,4]. A unique feature of the system described here is the nascent transcript(s) (**Fig. 3e**,**f**) that overlaps with the upstream enhancer region that is targeted by primary siRNAs (**Fig. 4** and **Supplementary Fig. 1**). The nascent transcript(s) and the overlapping primary siRNAs are apparently necessary but not sufficient for the generation of secondary siRNAs (see below) and hence for the spreading of methylation beyond the originally targeted enhancer region. Whether the secondary siRNAs and methylation spreading are essential for silencing the *GFP* gene is not known. A mutant defective in a gene product involved directly in secondary siRNA production would help to resolve this issue. Although secondary siRNAs are missing in mutants defective in DMS3, DRD1 and NRPD1b (**Fig. 3a**), these proteins are unlikely to function directly in secondary siRNA biogenesis, which is postulated to require currently unidentified AGO and RDR proteins, as well as DCL3 (**Fig. 4**). Moreover, the *dms3* mutant contains nearly wild-type levels of endogenous siRNA02 and siRNA1003 (**Fig. 3c**,**d**, lane *dms3*), indicating that DMS3 is not involved directly in siRNA production. A possible explanation for the lack of secondary siRNAs in *dms3*, *drd1* and *nrpd1b* mutants is the absence of primary siRNA–directed DNA methylation, which might be needed to attract an initiating component of the secondary siRNA–generating machinery. Considering previous proposals that Pol IVa transcribes methylated DNA[20,21], we hypothesize that primary methylation inhibits Pol II activity but attracts Pol IVa, which would take over synthesis of the nascent RNA(s) and in addition recruit AGO, RDR and DCL3 activities for secondary siRNA production (**Supplementary Fig. 6b**).

The identification of DMS3 adds a previously unknown protein to the RdDM machinery and expands the roles of SMC-related proteins to include an RNAi-mediated chromatin modification pathway. The transgene system we have developed might mimic natural genes that have noncoding transcripts derived from their promoter or upstream

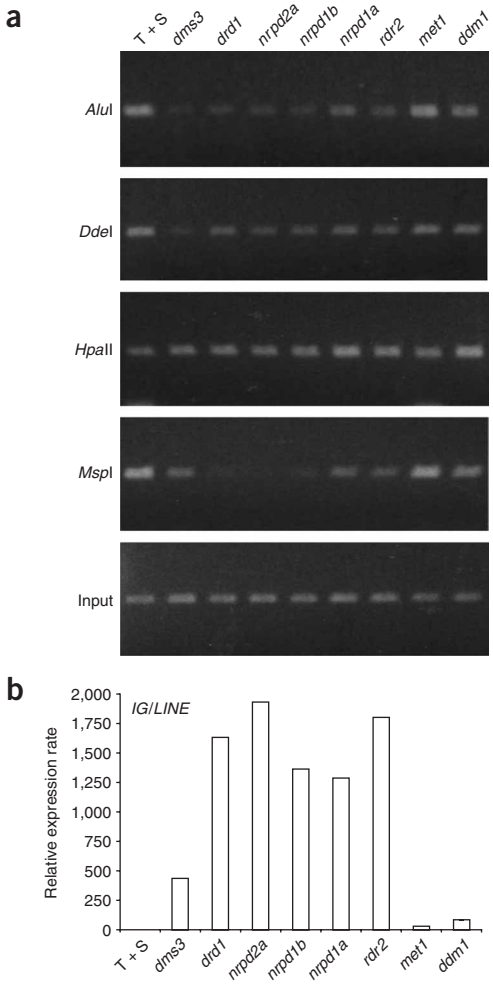

Figure 6 Loss of methylation and derepression of RdDM targets in a *dms3* mutant. (**a**) An intergenic solo LTR, which is a typical endogenous target of RdDM[17], shows less CNN and CNG methylation in a *dms3* mutant, similarly to *drd1*, *nrpd2a*, *nrpd1b*, *nrpd1a* and *rdr2* mutants[17]. This is indicated by reduced levels of the PCR amplification product following digestion with *Alu*I and *Dde*I (which report on CNN methylation) and *Msp*I (which reports on CNG methylation). CG methylation, assessed by digestion with *Hpa*II, is less affected. Mutations in the genes encoding MET1 and DDM1, which regulate primarily CG methylation, do not reduce methylation in the recognition site of the 376-bp solo LTR, which contains only two CG dinucleotides[17]. (**b**) Derepression of *IG/LINE* transcription occurs in a *dms3* mutant, and as reported previously, in *drd1*, *nrpd2a*, *nrpd1b*, *nrpd1a* and *rdr2* mutants[17], reflecting the loss of CNN and CNG methylation (**a**) in these mutants. Reactivation is minimal in *met1* and *ddm1* mutants. Transcription data are shown in a log scale in **Supplementary Figure 6**, together with data from other endogenous target sequences. T + S, wild-type plants containing target and silencer constructs.

regulatory regions[22,23]. Long noncoding RNAs might provide platforms for base-pairing interactions with primary siRNAs, which can trigger DNA methylation and/or repressive chromatin modifications at the targeted site as well as initiate the synthesis of secondary siRNAs that foster the spread of the silent state. The analysis of the other mutants obtained with our silencing system will likely identify additional factors required for RdDM, secondary siRNA biogenesis and spreading of DNA methylation.

METHODS

Plant material, transgene constructs and visualization of GFP. We carried out all analyses using wild-type *Arabidopsis thaliana* ecotype Columbia (Col-0). All mutants that we used are in the Col-0 background. We used the following alleles (note that *dms1*, *dms2* and *dms5* alleles correspond to newly identified alleles of *drd1*, *nrpd2a* and *nrpd1b*, respectively (**Supplementary Fig. 2**)): *dms1-1* (*drd1-7*), *dms2-1* (*nrpd2a-16*), *dms3-1* and *dms5-1* (*nrpd1b-10*) (**Fig. 3**); and *dms1-1* (*drd1-7*); *dms3-1*, *nrpd2a-4*, *nrpd1b-1*, *nrpd1a-4* (SALK_083051), *rdr2_1* (SAIL_1277H08), *met1-7* (SALK_076522) and *ddm1-10* (SALK_000590) (**Fig. 6**).

We made transgene constructs and introduced them into *Arabidopsis* according to standard procedures. We purchased a gene encoding enhanced GFP from Clontech. Details of the constructs are available from A.J.M. Matzke upon request. The 1.2-kb viral enhancer is derived from a *Nicotiana tomentosiformis* endogenous pararetrovirus[24]. This enhancer drives activity of *GUS* and *GFP* reporter genes in root and shoot meristems of transgenic *Arabidopsis* (this study; W. Gregor and A.J.M. Matzke, unpublished data).

We visualized GFP fluorescence in root apical meristems of approximately 10-day-old seedlings using a Leica stereo-fluorescence microscope MZFLIII. Seedlings were cultivated under sterile conditions on solid Murashige and Skoog medium in a 23 °C incubator with a light/dark cycle of 16 h light/8 h dark.

Mapping of mutations. Lehle Seeds carried out ethyl methanesulfonate (EMS) mutagenesis of seeds homozygous for the target and silencer complexes. We screened for mutants by selecting M2 seedlings that showed GPF activity in root meristems and double resistance to kanamycin and hygromycin. For genetic mapping and cloning of the *dms1*, *dms3* and *dms5* genes, we made F_2 mapping populations by crossing homozygous mutants of the M3 generation to ecotype Landsberg erecta, then selfing the resulting F_1 progeny to produce F_2 seedlings. We isolated genomic DNA from 30 F_2 plants that were GFP-positive and resistant to both kanamycin and hygromycin (indicating the presence of the target and silencer in a homozygous *dms* mutant background) and 30 F_2 plants that were GFP-negative and antibiotic resistant (that is, *DMS* wild-type plants). The initial approach involved hybridizing the pooled DNA from either mutant or wild-type plants to ATH1 microarrays[11], which usually narrowed the genetic interval down to ~2 Mb. We carried out further mapping on larger F_2 populations using cleaved amplified polymorphic sequence (CAPS) markers[12] and SNP markers from the Monsanto *Arabidopsis* polymorphism and Ler sequence collections (see URLs section below).

DNA sequencing. We isolated DNA from rosette leaves using a DNeasy Plant Maxi kit (Qiagen). We amplified the DNA fragment containing the gene encoding DMS3 (At3g49350) using the dms3-1 and dms3-2 primer combination (**Supplementary Table 1** online). After purifying the PCR product using a PCR purification kit (Qiagen) according to the manufacturer's instructions, we carried out the sequencing reaction using a Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) according to the manufacturer's instructions. Sequencing was carried out by VBC-Biotech.

DNA methylation. We made DNA blots to detect methylation of 180-bp centromeric and 5S rDNA repeats as described previously[4]. Bisulfite treatment of DNA (isolated as described above) was conducted using a EpiTect Bisulfite kit (Qiagen) according to the manufacturer's instructions with several modifications. Genomic DNA was pre-digested with *Hin*dIII and 500 ng of the DNA was treated in the reaction solution containing 35 µl of the DNA protect buffer. Following incubation for 2 min 95 C°, 8 cycles of these thermocycler conditions were performed: 1 min 95 C° and 2 h 75 C°. Primers are shown in **Supplementary Figure 1** and **Supplementary Table 1**. We analyzed methylation of the solo LTR using a PCR-based technique as detailed in a previous publication[17].

RNA analysis. Small RNAs were isolated from 21-day-old seedlings, rosette leaves or inflorescences using the mirVana miRNA isolation kit (Ambion) and analyzed by RNA blotting according to published procedures[4,17]. Primer information for specific probes is shown in **Supplementary Table 1**.

To detect nascent RNAs by RT-PCR, we isolated total RNA from seedlings using TRIZOL (Invitrogen). We synthesized cDNA using a First Strand cDNA Synthesis kit (Fermentas) according to the manufacturer's protocol using an oligonucleotide d(T) primer and 1 µg of total RNA. The primer sets are shown in **Supplementary Figure 1** and **Supplementary Table 1**.

Expression analysis of endogenous targets. We used real time PCR to analyze expression of *IG/LINE*, *IG5* and *IG5*** as described previously[17]. Primers used for *IG5*** (not previously reported) are listed in **Supplementary Table 1**.

DMS3 domain structure. For primary sequence analysis, we used visualization and query tools from the IMP Bioinformatics Large-Scale Sequence Annotation System (see URLs section below) and Pfam (see URLs section below)[25]. We applied algorithms including low complexity filtering by seg[26], detection of conserved domains by RPS-BLAST/CDD[27], detection of coiled-coils by Coils[28], and prediction of secondary structure by Predator[29].

URLs. Genevestigator, https://www.genevestigator.ethz.ch; Monsanto *Arabidopsis* Polymorphism and Ler Sequence Collections, http://www.arabidopsis.org/browse/Cereon/index.jsp; IMP Bioinformatics Large-Scale Sequence Annotation system, http://www.annotator.org; Pfam, http://www.sanger.ac.uk/Software/Pfam/.

Accession codes. GenBank: ABE81692.1, *Medicago truncatula* AT3g49250.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This work is supported by the European Science Foundation (ESF) under the EUROCORES Programme EuroDyna, through contract no. ERAS-CT-2003-980409 of the European Commission, DG Research, FP6, and the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF) (grant I26-B03), and the European Union (contract HPRN-CT-2002-00257). D.P.K. acknowledges support by the Vienna Science and Technology Fund (WWTF), Austrian Center of Biopharmaceutical Technology (ACBT), Austrian Research Center Seibersdorf and Baxter AG.

AUTHOR CONTRIBUTIONS

M.M. and A.J.M.M. designed and supervised the study. A.J.M.M. made the constructs and silenced *Arabidopsis* line. E.B. carried out the mutant screen, bisulfite sequencing and preliminary mapping of *dms1–dms5* mutations. T.K. identified and sequenced the dms1, dms2, dms3 and dms5 mutant alleles and analyzed nascent RNAs. L.D. analyzed small RNAs. B.H. analyzed reactivation of target sequences in the dms3 mutant. D.P.K. performed the bioinformatics analysis. M.M. wrote the paper with contributions by all coauthors.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Huettel, B. *et al.* RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. *Biochim. Biophys. Acta* **1769**, 358–374 (2007).
2. Chan, S.W.L., Henderson, I.R. & Jacobsen, S.E. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat. Rev. Genet.* **6**, 351–360 (2005).
3. Kanno, T. *et al.* Involvement of putative SNF2 chromatin remodelling protein DRD1 in RNA-directed DNA methylation. *Curr. Biol.* **14**, 801–805 (2004).
4. Kanno, T. *et al.* Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **37**, 761–765 (2005).
5. Pontier, D. *et al.* Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* **19**, 2030–2040 (2005).
6. Losada, A. & Hirano, T. Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes Dev.* **19**, 1269–1287 (2005).

7. Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J.M. & Matzke, M. RNA-directed DNA methylation in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **99**, 16499–16506 (2002).
8. Baulcombe, D.C. Amplified silencing. *Science* **315**, 199–200 (2007).
9. Vaistij, F.E., Jones, L. & Baulcombe, D.C. Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell* **14**, 857–867 (2002).
10. Mette, M.F. *et al.* Endogenous viral sequences and their potential contribution to heritable virus resistance in plants. *EMBO J.* **21**, 461–469 (2002).
11. Hazen, S.P. *et al.* Rapid array mapping of circadian clock and developmental mutations in *Arabidopsis*. *Plant Physiol.* **138**, 990–997 (2005).
12. Koneiczny, A. & Ausubel, F. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4**, 403–410 (1993).
13. Obayashi, T. *et al.* ATTED-II: a database of co-expressed genes and *cis* elements for identifying co-regulated gene groups in *Arabidopsis*. *Nucleic Acids Res.* **35**, D863–D869 (2006).
14. Hirano, M. & Hirano, T. Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *EMBO J.* **21**, 5733–5744 (2002).
15. Chiu, A., Revenkova, E. & Jessberger, R. DNA interaction and dimerization of eukaryotic SMC hinge domains. *J. Biol. Chem.* **279**, 26233–26242 (2004).
16. Hirano, T. At the heart of the chromosome: SMC proteins in action. *Nat. Rev. Mol. Cell Biol.* **7**, 311–322 (2006).
17. Huettel, B. *et al.* Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J.* **25**, 2828–2836 (2006).
18. Hagstrom, K.A. & Meyer, B.J. Condensin and cohesion: more than chromosome compactor and glue. *Nat. Rev. Genet.* **4**, 520–534 (2003).
19. Cobbe, N., Savvidou, E. & Heck, M.M. Diverse mitotic and interphase functions of condensins in *Drosophila*. *Genetics* **172**, 991–1008 (2006).
20. Herr, A.J., Jensen, M.B., Dalmay, T. & Baulcombe, D.C. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118–120 (2005).
21. Pontes, O. *et al.* The *Arabidopsis* chromatin modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**, 79–92 (2006).
22. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
23. Kapranov, P., Willingham, A.T. & Gingeras, T.R. Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* **8**, 413–423 (2007).
24. Gregor, W., Mette, M.F., Staginnus, C., Matzke, M. & Matzke, A.J.M. A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. *Plant Physiol.* **134**, 1191–1199 (2004).
25. Finn, R.D. *et al.* Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**, D247–D251 (2006).
26. Wootton, J.C. Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput. Chem.* **18**, 269–285 (1994).
27. Marchler-Bauer, A. *et al.* CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* **30**, 281–283 (2002).
28. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164 (1991).
29. Frishman, D. & Argos, P. Incorporation of long-distance interactions into a secondary structure prediction algorithm. *Protein Eng.* **9**, 133–142 (1996).