



Complex interplay between exapted transposable element proteins and epigenetic mechanisms in the regulation of gene expression and silencing of DNA repeats in plants

Guillaume Moissiard

► To cite this version:

Guillaume Moissiard. Complex interplay between exapted transposable element proteins and epigenetic mechanisms in the regulation of gene expression and silencing of DNA repeats in plants. Life Sciences [q-bio]. UPVD, 2020. tel-03037768

HAL Id: tel-03037768

<https://univ-perp.hal.science/tel-03037768>

Submitted on 3 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITE DE PERPIGNAN

2020



Dossier présenté en vue d'obtenir

L'HABILITATION A DIRIGER DES RECHERCHES

Par

Guillaume Moissiard

Complex interplay between exapted transposable element proteins and epigenetic mechanisms in the regulation of gene expression and silencing of DNA repeats in plants

Soutenue le 05 mai 2020

Composition du jury

Dr. Vincent Colot	Rapporteur externe
Dr. Laurence Maréchal-Drouard	Rapporteur externe
Pr. Philippe Gallusci	Rapporteur externe
Dr. Aline Probst	Examineur externe
Dr. Antoine Martin	Examineur externe
Dr. Natacha Bies-Ethève	Examineur interne
Dr. Thierry Lagrange	Examineur interne

Laboratoire Génome et Développement des Plantes
CNRS/UPVD
UMR5096



“The major problem is chromatin... You can inherit something beyond the DNA sequence. That's where the real excitement of genetics is now”.

- James D. Watson (2003)

Remerciements.

Je remercie les membres de l'équipe MEAC ainsi que toutes les personnes du LGDP et en particulier, les responsables des différentes plateformes du laboratoire sans qui toute recherche ne serait pas possible. Je remercie aussi l'équipe administrative et les gestionnaires du LGDP pour leur soutien logistique et aide précieuse.

Je remercie aussi toutes les personnes qui m'ont soutenu pendant les périodes difficiles ; sans pour autant les nommer, elles se reconnaîtront... Je leur suis entièrement reconnaissant.

Finalement, je remercie tout particulièrement Dr. Laurence Maréchal-Drouard, Dr. Aline Probst, Dr. Natacha Bies-Ethève, Dr. Vincent Colot, Pr. Philippe Gallusci, Dr. Antoine Martin et Dr. Thierry Lagrange d'avoir accepté d'être membres du jury et pour le temps qu'ils ont consacré à la lecture de ce manuscrit.

UNIVERSITE DE PERPIGNAN

2020

Dossier présenté en vue d'obtenir

L'HABILITATION A DIRIGER DES RECHERCHES

Par

Guillaume Moissiard

Complex interplay between exapted transposable element proteins and epigenetic mechanisms in the regulation of gene expression and silencing of DNA repeats in plants

Soutenue le 05 mai 2020

Composition du jury

Dr. Vincent Colot	Rapporteur externe
Dr. Laurence Maréchal-Drouard	Rapporteur externe
Pr. Philippe Gallusci	Rapporteur externe
Dr. Aline Probst	Examineur externe
Dr. Antoine Martin	Examineur externe
Dr. Natacha Bies-Ethève	Examineur interne
Dr. Thierry Lagrange	Examineur interne

Laboratoire Génome et Développement des Plantes
CNRS/UPVD
UMR5096

Table of Contents

I. Curriculum vitae.....	1
II. Research management and supervision, scientific animation and popularization	2
III. List of publications, communications and collaborations	3
IV. Main introduction	5
1. The chromatin and subnuclear organization of the genome	5
2. A brief history of epigenetics	5
3. The nucleosome, first layer of chromatin organization	7
4. Post-translational modifications of histones	8
5. Histone lysine methyltransferases (HKMTs)	9
6. DNA methyltransferases (DNMTs)	10
7. Additional roles of DNA methylation	12
8. Epigenetic writers, erasers and readers	12
9. Other epigenetic players and the epigenetic “mille-feuille”	13
10. The complex interplay between epigenetic modifications and transcription factors	15
11. Transgenerational epigenetic inheritance and memory	18
12. Roles of TEs in the organization, evolution and innovation of eukaryotic genomes	18
V. Research studies	21
1. PhD thesis	21
2. Postdoctoral research	21
2.1. The role of JmjC JMJ14 in the DRM2-mediated maintenance of DNA methylation..	21
2.2. A forward genetic screen identifies new epigenetic players	21
2.3. The Microrchidia ATPases are required for gene silencing and heterochromatin compaction	22
2.4. MTHFD1 controls DNA methylation in <i>A. thaliana</i>	23
3. Current research projects	23
3.1. Characterization of the <i>ATCOPIA28::GFP</i> system	24
3.2. The <i>ATCOPIA28::GFP</i> forward genetic screen	25
3.3. The Plant Mobile Domain (PMD) protein MAINTENANCE OF MERISTEMS (MAIN) is required for the <i>ATCOPIA28::GFP</i> silencing	27

3.4. A brief introduction of the Plant Mobile Domain (PMD) proteins	27
3.5. MAIN and MAIL1 are required for TE silencing and the proper expression of several genes, including <i>MORC1</i> and <i>FLC</i>	29
3.6. Mode of action of MAIN/MAIL1/PP7L complex in the regulation of gene expression and TE silencing.....	32
3.6.1. Are PMD and PP7L proteins physically interacting with chromatin?	32
3.6.2. What is the role of PP7L in the MAIN/MAIL1 pathway?	33
3.6.3. Are MAIN, MAIL1 and PP7L transcriptional activators or repressors?	33
3.6.4. What is the structure of PMD?	34
3.7. Investigating the role of other PMD proteins in <i>A. thaliana</i> plant development	35
3.7.1. Preliminary results	35
3.7.2. What are the roles of MAIL2 and MAIL3?	36
3.7.3. What are the roles of PMD-B proteins?	37
3.8. The role of PMD proteins in <i>S. lycopersicum</i> (tomato) plant and fruit development	38
3.8.1. Scientific context	38
3.8.2. What is the expression pattern of <i>SIPMDs</i> at the tissue-specific level?	39
3.8.3. Are <i>SIPMDs</i> required for tomato plant development and ripening process?	40
3.8.4. Are <i>SIP7L</i> and <i>SIPMD</i> interacting together?	40
3.9. Evolutionary aspects of PMD in relation with TEs and other protein domains	40
3.9.1. Scientific context	40
3.9.2. Can PMD be beneficial for TEs?	41
3.9.3. Co-evolution of PMD and other protein domains.	43
4. Future studies and perspectives: beyond the PMD, role of ETE proteins in regulation of gene expression	43

VI. References	44
VII. Appendix (selected scientific production)	54
VIII. List of figures	55
IX. Abbreviations	56

I. Curriculum Vitae.

MOISSIARD GUILLAUME

Charge de recherche CNRS CRCN

Date of birth: December 10th, 1980 **Nationality:** French

Address: Laboratoire Genome et Developpement des Plantes

UMR5096 CNRS/UPVD Bâtiment T

58 Avenue Paul Alduy

66860 Perpignan

Tel: +33(0) 4 68 08 67 90

Email: guillaume.moissiard@univ-perp.fr

Diplomas and Training

10/2016 – Present: Group leader – CR1 researcher at LGDP, Perpignan, France.

2013 – 09/2016: Independent Senior Post-doc in the Voinnet lab at ETH, Zurich, Switzerland.

2007 – 2013: Post-doctoral fellow in the Steve Jacobsen lab at UCLA, Los Angeles, CA, USA.

2003 – 2007: PhD student in the Voinnet laboratory at IBMP – CNRS (Strasbourg, France).

2002 – 2003: Master of Sciences in Molecular Biology (II) at IBMP – CNRS (Strasbourg, France).

2001 – 2002: Master of Sciences in Molecular Biology (I) at University of Montpellier (France).

2000 – 2001: License of Sciences in Molecular Biology at University of Montpellier (France).

1998 – 2000: DEUG in Life Sciences at University of Aix-Marseille (Marseille, France).

Additional training

April 2002 - June 2002: Master training period. “Engineering chimerical HIV/SIV integrase to study the interaction with the Uracil-DNA-Glycosylase”. INSERM, Marseille Luminy (France) , Laboratory of Lentivirus (supervised by Joséphine Sire).

April 2001-May 2001: License training period. “Recombinant expression of sheep metallothionein in *E. coli* for phytoremediation”. CEA Cadarache, Saint-Paul-Lès-Durance (France). Laboratory of Ecophysiology of Photosynthesis (supervised by Sandrine Sauge-Merle).

University teaching responsibilities

2014 - 2015: ETH Zurich (Switzerland), “plant responses to biotic stress” and “Chromatin-based RNA silencing”.

2012: University of California Los Angeles (USA), Elaboration and supervision of a Journal Club class for UCLA students.

2003 - 2007: University of Medicine of Strasbourg (France) – *Monitorat*, “regulation of gene expression” and “Introduction to RNA silencing”.

Awards

2016/2019: Prime d’Encadrement Doctoral et de Recherche (PEDR)

2003/2007: MENRT grant allocated by the French Ministry of Research.

Funding

04-2019/12-2019: GDR EPIPLANT call 2019. 4500€ (Projet PMD)

01-2018/12-2018: LABEX TULIP Projet central. 6852€ (Projet PMD)

01-2017/12-2017: Gros projet BQR, UPVD. 15k€ (Projet *ATCOPIA28::GFP*)

II. Research management and supervision, scientific animation and popularization.

Supervision of PhD candidate

- **Melody Nicolau (UPVD Perpignan ED305, 10-2017 / present)**
Co-supervision with Frederic Pontvianne.
Title: Role of Plant Mobile Protein (PMD) in the regulation of gene expression and silencing of transposable elements.
Expected PhD defense: Fall 2020.

Supervision of Master and undergraduate students

- **Lorena Valdes (Master M1 student, UPVD Perpignan, 07-2019)**
Characterization of *pmd* mutants in *A. thaliana*.
- **Laurie Gimenez (IUT 1st year student, UPVD Perpignan, 07-2019)**
Characterization of *pmd* mutants in *A. thaliana*.
- **Faysal Yakhlef (Master M1 student, UPVD, Perpignan, 02/05-2017)**
Characterization of the *ddc16* mutant population.
- **Lea Garros (BTS student, Lycee Jean Mermoz Montpellier, 01/02-2017)**
Characterization of M2 mutant populations.
- **Benjamin Maier (Master M2 student, ETH Zurich, 10-2014/09-2015)**
Role of CROWDED NUCLEI (CRWN) proteins in the regulation of gene expression and chromatin organization.
- **Olivier Martel-Paradis (Master M2 student, ETH Zurich, 01-2014/06-2014)**
Role of Histone H1-like in the regulation of gene expression.
- **Huajun Zhou (Master M2 student, Jacobsen lab, UCLA, 09-2011/12-2011)**
Characterization of the M2 mutant population wt#162.
- **Supervision of several undergraduate students (Jacobsen lab, UCLA, 01-2008/06-2013)**
Dylan Husmann, Vanessa Trieu, Ashot Papikian, Caroline Kim, Haivy Nguyen, Zain Mehdi, Manan Damani, Omid Taghavi and Bunkuong Lim.

Supervision of research assistants

- **Julie Descombin (Ingenieure d'étude, UPVD Perpignan, 10-2016/present)**
Screen and characterization of *ATCOPIA28::GFP* mutant populations.
Technical support for other projects including the PMD project.

- **Benjamin Maier (Research assistant, ETH Zurich, 11-2015/01-2016)**
Role of CROWDED NUCLEI (CRWN) proteins in the regulation of gene expression and chromatin organization.
- **Dylan Husmann (Research assistant, Jacobsen lab, UCLA, 07-2012/06-2013).**
Role of MORC proteins in TE silencing.

Scientific animation and popularization

Participation in “Fete de la science” 2017-2019, attempt to set up a science club meeting with UPVD PhD applicants. Organization of LGDP seminars and Journal Club. “Ma these en bref” and interaction with high school students (during my PhD training).

III. List of publications, communications and collaborations.

Publications

Submitted

- Nicolau M., Picault N., Descombin J., Jami-Alahmadi Y., Feng S., Bucher E., Jacobsen S.E., Deragon J.M., Wohlschlegel J. and **Moissiard G.** The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*. *Submitted to PLOS Genetics, in revision.*

Published

- Groth M., **Moissiard G.**, Wirtz M., et al. MTHFD1 controls DNA methylation in *Arabidopsis*. *Nat Commun.* **2016**; 7: 11640.
- Harris, C.J., Husmann, D., Liu, W., Kasmi, F.E., Wang, H., Papikian, A., Pastor, W.A., **Moissiard, G.**, Vashisht, A.A., Dangl, J.L., et al. *Arabidopsis* AtMORC4 and AtMORC7 Form Nuclear Bodies and Repress a Large Number of Protein-Coding Genes. *PLoS Genet.* **2016**; 12: e1005998.
- **Moissiard, G.**, Bischof, S., Husmann, D., Pastor, W. A., Hale, C. J., Yen, L., Stroud, H., Papikian, A., Vashisht, A. A., Wohlschlegel, J. A. and Jacobsen, S. E. Transcriptional gene silencing by *Arabidopsis* microorchidia homologues involves the formation of heteromers. *Proc Natl Acad Sci U S A.* **2014**;111(20):7474-7479.
- Pastor, W. A., Stroud, H., Nee, K., Liu, W., Pezic, D., Manakov, S., Lee, S. A., **Moissiard, G.**, Zamudio, N., Bourc'his, D., Aravin, A. A., Clark, A. T., Jacobsen, S. E. MORC1 represses transposable elements in the mouse male germline. *Nat Commun.* **2014**; 5:5795. Erratum 2015; 6:7604.
- **Moissiard, G.**, Cokus, S. J., Cary, J., Feng, S., Billi, A. C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B. R., McCord, R. P., Hale, C. J., Feng, W., Michaels, S. D., Frand, A. R., Pellegrini, M., Dekker, J., Kim, J. K., Jacobsen, S. E. MORC family ATPases required for heterochromatin condensation and gene silencing. *Science.* **2012**;336: 1448-1451.
- Deleris, A., Greenberg, M. V., Ausin, I., Law, R. W., **Moissiard, G.**, Schubert, D., Jacobsen, S. E. Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. *EMBO Rep.* **2010**;11(12):950-955.

- Haas, G., Azevedo, J., **Moissiard, G.**, Geldreich, A., Himber, C., Bureau, M., Fukuhara, T., Keller, M., and Voinnet, O. Nuclear import of CaMV P6 is required for infection and suppression of the RNA silencing factor DRB4. *Embo J.* **2008**; 27, 2102-2112. Corrigendum 2015; 34, 20, 2591-2592.
- **Moissiard, G.**, Parizotto, E.A., Himber, C., and Voinnet, O. Transitivity in Arabidopsis can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins. *RNA* **2007**; 13, 1268-1278. Corrigendum 2016 Jan;22(1):162-5. doi: 10.1261/rna.055012.115. May 2016 22: 810; doi:10.1261/rna.056366.116
- **Moissiard, G.**, and Voinnet, O. Viral suppression of RNA silencing in plants. *Molecular plant pathology* **2004**; 5, 71-82. *Review*.
- Himber, C., Dunoyer, P., **Moissiard, G.**, Ritzenthaler, C., and Voinnet, O. Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *Embo J* **2004**; 22, 4523-4533.

Communications

- Seminar at Helmholtz Zentrum München, BIOP (invited by Martin Groth), 2019, Munich, Germany.
- Abcam Chromatin and Epigenetics: Inheritance and Design, 04-2019, Munich, Germany. (flash talk and poster).
- Seminar at GBF (invited by Julien Pirrello), 01-2019, Toulouse, France.
- International plant epi/genetics meeting 10-2018 Angers, France.
- Seminar at UCLA (invited by Steve Jacobsen), 2017, Los Angeles, USA

Collaborations

- **Christel Carles (LPCV, UMR CNRS 5168, Grenoble)**. Role of Plant Mobile Domain proteins in plant development. DNA-binding assays.
- **Julien Pirrello (GBF INRA/INP-ENSAT UMR990 Toulouse)**. Role of Plant Mobile Domain proteins in *Solanum lycopersicum* (tomato).
- **James Wohlschlegel (Department of Biological Chemistry, UCLA, Los Angeles, CA, USA)**. Proteomic analyses.
- **Steven E. Jacobsen (Department of Molecular, Cell and Developmental Biology, UCLA, Los Angeles, CA, USA)**. NGS analyses.
- **Etienne Bucher (Plant Breeding and Genetic Resources, Agroscope, Nyon, Switzerland)**. NGS mapping.
- **Jean-Marc Deragon (LGDP - CNRS/UPVD - UMR5096, Perpignan)**. Phylogenetic analyses.
- **Moaine EL BAIDOURI (LGDP - CNRS/UPVD - UMR5096, Perpignan)**. Evolution of TE/PMD.
- **Jiamu Du (SUSTech, Shenzhen, China)**. Crystal structure of PMD and PP7L.
- **Lionel Navaro (IBENS CNRS/INSERM UMR8197 Paris)**. *Pseudomonas syringae* infection (MAIL3 project).
- **Harald Keller (INRA1355-CNRS7254-Université Côte d'Azur, UMR Institut Sophia Agrobiotech, Sophia Antipolis)**. *Hyaloperonospora arabidopsidis* (Hpa) infection (MAIL3 project).

IV. Main introduction.

1. The chromatin and subnuclear organization of the genome.

In eukaryotes, the genetic material is precisely organized in the nucleus as chromatin, which is predominantly composed of DNA and protein molecules. Early microscopic observations of nuclei at interphase revealed the existence of two compaction states of chromatin¹. Constitutive heterochromatin, which appears as dense DNA foci called chromocenters (CCs), is composed of transcriptionally repressed DNA repeats and transposable elements (TEs) that are kept silenced because of their high mutagenic effect (Figure 1)^{2,3}. Constitutive heterochromatin tends to be more localized at the nuclear periphery or in close proximity to the nucleolus⁴. By contrast, euchromatin is more diffused within the nuclear interior, and is enriched in genic regions that are more permissive to transcription (Figure 1). Importantly, chromatin structure is not as dichotomic as originally thought, as up to nine chromatin states can prevail in the nucleus^{5,6}. Furthermore, facultative heterochromatin constitutes another form of chromatin that accumulates in the nuclear interior with very distinct epigenetic features (Figure 1 and see below).

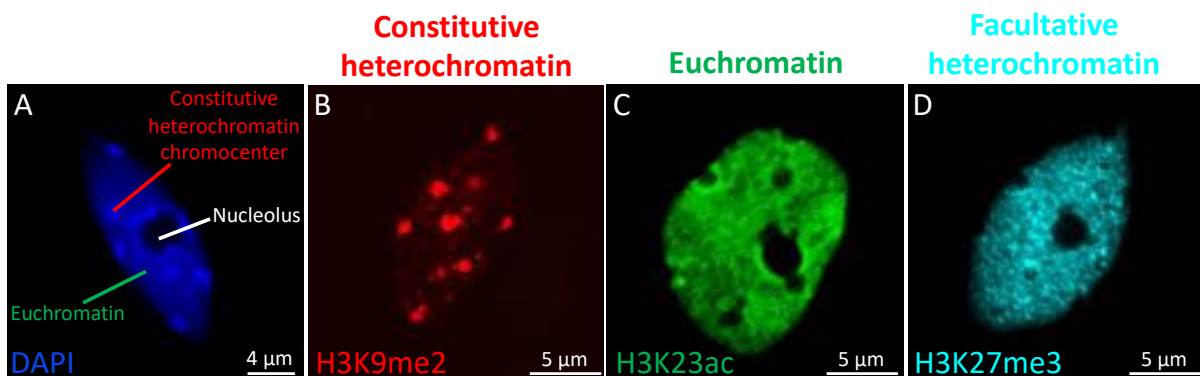


Figure 1. Chromatin states of *Arabidopsis thaliana* nuclei at interphase. A. Constitutive heterochromatin appears as dense DNA foci called chromocenters (CCs) preferentially located nearby nuclear periphery and nucleolus, while euchromatin is more relaxed and accumulates in the nuclear interior. B-C. Trimethylation of histone H3 at lysine 9 (H3K9me2) and histone acetylation (such as H3K23ac) are hallmarks of constitutive heterochromatin and transcribed euchromatin, respectively. Trimethylation of histone H3 at lysine 27 (H3K27me3) is a repressive epigenetic mark enriched in facultative heterochromatin. DAPI= 4',6-diamidino-2-phenylindole staining of DNA.

2. A brief history of epigenetics.

It is in the early forties that Conrad Hal Waddington used for the first time the term *epigenetics*. At that time, this term referred to *epigenesis* and *epigenotype*, which were defined by Waddington as

the complex developmental processes occurring between genotype and phenotype of an organism from its embryonic stage to adult life⁷. In 1957, Waddington also proposed for the first time the notion of “epigenetic landscape” as a metaphor in which any kind of developmental and environmental constraints that are facing a cell during its life impose to make crucial decisions that will impact cell fate and developmental outcome (Figure 2)⁸. The term epigenetic landscape will be latter reused by epigeneticists as a synonymous way of defining the epigenetic chromatin state at a specific genomic location.



Figure 2. The concept of “epigenetic landscape” as depicted by Conrad W. Waddington in 1957. In this visual metaphor, the cell, represented as a ball, is making decisions throughout its development. These decisions will impose specific cell trajectories within its epigenetic landscape, and will impact the outcomes and cellular fates. Figure extracted from reference⁸.

In the line of Waddington’s definition, the geneticist David L. Nanney wrote in 1958 an essay entitled “Epigenetic control systems”⁹. In this composition, Nanney wrote the following sentence:

“The existence of phenotypic differences between cells with the same genotype merely indicates that the expressed specificities are not determined entirely by the DNA present in the cell - that other devices, epigenetic systems, regulate the expression of the genetically determined potentialities.”

Based on microbiology studies, Nanney also suggested that “epigenetic systems” could be a way for micro-organisms to adapt to environmental conditions⁹.

In 1987, Robin Holliday meliorates the definition of epigenetics by connecting the processes of DNA methylation and gene expression during development together with the notions of reversibility and inheritability¹⁰. Finally, it is in 1994 that Holliday proposes the modern definition of epigenetics¹¹.

“The study of the changes in gene expression, which occur in organisms with differentiated cells, and the mitotic inheritance of given patterns of gene expression.”

Holliday also emphasized the notion of inheritability and reversibility:

“Nuclear inheritance which is not based on differences in DNA sequence.”

Nowadays, epigenetics can be defined as any reversible modifications of the chromatin such as DNA methylation or histone modifications that are inheritable, and alter gene expression without a change in the primary DNA sequence.

Since the term epigenetics was first coined by Waddington in 1942, this field has been extensively studied. Today, it is well established that epigenetics plays essential roles during the development of any organisms, but also in response to environmental stimuli and diseases¹². Epigenetic modifications can be inherited throughout the life of organism, as well as transgenerationally. Finally, epigenetics has spread beyond the fields of genetics and molecular biology, and epigenetic concepts have been embraced by various scientific disciplines such as among others, ecology, evolution, physiology or psychology¹³.

For a more thorough definition of epigenetics, and a more detailed history of this field, please refer to references^{8, 11, 12} and ¹³.

3. The nucleosome, first layer of chromatin organization.

The nucleosome is composed of 147-bp DNA helix wrapped around an octamer of core histone proteins (H2A, H2B, H3 and H4)¹⁴. This constitutes the so-called “beads-on-string” chromatin structure, which is the first layer of chromatin organization¹⁵. Importantly, this chromatin structure can only be observed under non-physiological conditions, and in the nucleus, the chromatin rapidly complexifies in a higher-order chromatin structure, the 30-nm chromatin fiber, through the activity of several factors including linker histone H1 that contacts linker DNA to bring adjacent nucleosomes together (Figure 3)¹⁶. In addition, epigenetic modifications such DNA methylation and post-translational modifications of histones, and specific chromatin remodelers contribute to the compaction state of chromatin¹⁵.

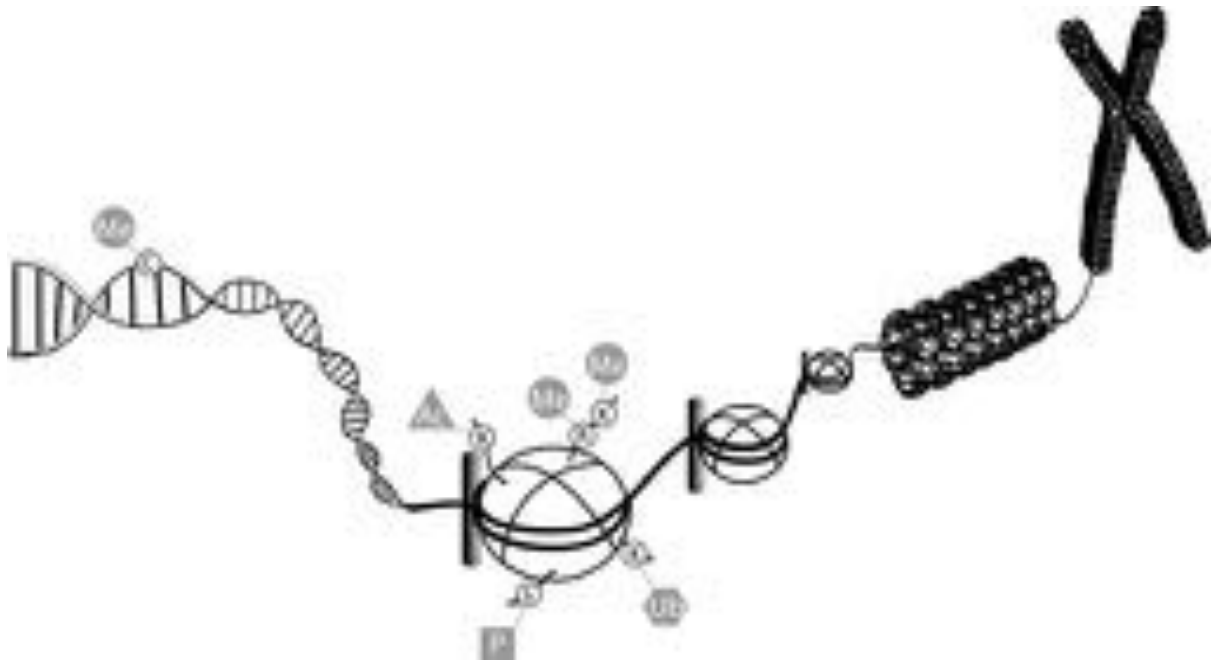


Figure 3. Artwork of chromatin organization from the double helix DNA molecule to the highly-condensed chromosome. Nucleosomes are depicted as “beads-on-string” associated with linker H1 histone (black cylinders). Compaction of nucleosomes form the 30-nm chromatin fiber and ultimately, a condensed chromosome, as observed during cell division. Methylation (Me) of DNA and post-translational modifications of histones such as Me, acetylation (Ac), phosphorylation (P) and ubiquitination (Ub) represent different epigenetic modifications of the chromatin. C = cytosine, K = lysine, R = arginine, S = serine.

4. Post-translational modifications of histones.

Post-translational modifications (PTMs) of histones are epigenetic marks involved in various cellular processes such as regulation of gene expression, cell division, DNA damage response, or TE and gene silencing. Several combinations of histone PTMs occur within the cell, and histones can be covalently modified by a myriad of chemicals (such as methyl, acetyl, phosphate, ubiquitin, small ubiquitin-like modifier [SUMO], ADP-ribose) at specific amino acids of their tails, which protrude from the histone core particles (Figure 3)¹⁷. In addition, histones can undergo citrullination (deimination of arginine into citrulline), or proline isomerization^{17,18}. Altogether, the different combinations of histone PTMs constitute the so-called histone code¹⁹. A brief description of enzymes specialized in histone PTMs is provided below.

Acetylation and deacetylation of histones at distinct lysine residues are mediated by various histone acetyl transferases (HATs) and histone deacetylases (HDACs), respectively^{17, 18}. Histone acetylation is an active mark of transcription, enriched in euchromatin (Figure 1C), while HDAC-mediated deacetylation of histones contributes to gene silencing^{17, 20}. In *A. thaliana*, HISTONE

DEACETYLASE 6 (HDA6) is required for transcriptional gene silencing²¹⁻²³. There are different classes of HATs and HDACs reviewed in reference¹⁸. Phosphorylation of histones at serine, threonine or tyrosine residues plays important role during cell division and chromosome condensation, but is also associated to transcription regulation^{17, 18}. Sequential events of H2A or H2B ubiquitination and deubiquitination, associated with complex crosstalk with other histone marks, have been linked to regulation of transcription elongation, nucleosome stability and gene silencing^{17, 18, 20, 24}. The Polycomb (PcG) repressive complex 1 (PRC1)-mediated H2A monoubiquitination is involved in gene silencing^{17, 18}. Conversely, in *A. thaliana*, ubiquitin-specific protease 26 (UBP26)-mediated H2B deubiquitination is required for DNA methylation and TE silencing²⁵. Histone methylation can occur at arginine and lysine residues¹⁷. Distinct protein arginine methyltransferases (PRMTs) can be specifically recruited at the promoters of genes, where they can exert a positive or negative influence on transcription²⁶. PRMTs have been involved in several cellular processes such as cell proliferation, transformation and anti-apoptotic function, all of which are related to tumorigenesis²⁷. Histone lysine methyltransferases (HKMTs) are more thoroughly described below.

5. Histone lysine methyltransferases (HKMTs).

SET DOMAIN GROUP (SDG) proteins carries a catalytic SET domain identified first in the *Drosophila* histone methyltransferases Suppressor of variegation 3-9 [su(var)3-9], Enhancer of zeste [E(Z)], and Tritheorax)²⁸. SDG proteins promoting the deposition of methyl groups on lysine residues of histones are called histone lysine methyltransferases (HKMTs)²⁸, while non-histone lysine methyltransferases (KMTs) transfer methyl groups on lysine residues of non-histone proteins²⁹.

In *Arabidopsis thaliana*, there are five classes of HKMTs.

- The class I HKMTs CURLY LEAF (CLF), MEDEA and SWINGER (SWN) are homologs of E(Z); essential components of the PcG pathway, they are involved in the deposition of trimethyl groups on H3 at lysine 27 (H3K27me3), a repressive epigenetic mark forming facultative heterochromatin, which is mostly composed of developmentally-regulated genes ([Figure 1D](#))^{30, 31}.
- Belonging to class II HKMTs, SDG4 and SDG8 are homologs of Set2, a HKMT that deposits methyl groups on H3 at lysine K36 (H3K36me3) during transcription elongation to repress cryptic transcription in yeast^{17, 20}.

- The class III HKMTs are also called ARABIDOPSIS TRITHORAX-LIKE (ATX) proteins, part of the COMPASS-related (AtCOMPASS) protein complex^{31,32}. Among them, SDG2 has been involved in the deposition of trimethyl groups on H3 at lysine 4 (H3K4me3), which is an active mark of transcription, enriched in euchromatin^{31,33}.
- The two class IV HKMTs ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 (ATXR5) and ATXR6 are required for the deposition of monomethyl group on H3 at lysine 27 (H3K27me1), a repressive mark concomitant to dimethylation of histone H3 lysine 9 (H3K9me2), and highly enriched in the constitutive heterochromatic CCs^{31,34,35}. ATXR5 and ATXR6 are plant-specific HKMTs, and enrichment of H3K27me1 in constitutive heterochromatin seems to be unique to the plant kingdom³⁴. In mouse embryonic stem cells (ESCs), E(Z)-mediated H3K27me1 is indeed enriched in transcriptionally active genes³⁶. In *A. thaliana*, ATXR5- and ATXR6-mediated H3K27me1 deposition has been involved in the control of heterochromatic DNA replication³⁷.
- Class V HKMTs are the most abundant HKMTs in *A. thaliana*³¹. SU(VAR)3-9 HOMOLOGOUS 4 (SUVH4, a.k.a. KRYPTONITE [KYP]), SUVH5 and SUVH6 catalyze the deposition of dimethyl groups on histone H3 at lysine 9 (H3K9me2), which is a hallmark of constitutive heterochromatin (Figure 1B)^{31,38}. In *A. thaliana*, H3K9me2 is predominantly found in the pericentromeric regions of chromosomes³⁹, where most of DNA repeats and TEs reside. SUVH4/SUVH5/SUVH6 are required for the maintenance of DNA methylation -mostly at CHG sites, through a self-reinforcing loop involving CHROMOMETHYLASE 3 (CMT3)³⁸. SUVH2 and SUVH9 have been involved in the recruitment of the plant-specific RNA polymerase V (RNA Pol V), which is involved in the RNA-directed DNA methylation (RdDM) pathway⁴⁰. SUVH1 and SUVH3 have been recently identified as transcriptional activators⁴¹ (and see below). SU(VAR)3-9 RELATED 2 (SUVR2) interacts with its close homolog SUVR1, as well as with the SNF2 chromatin remodeler-RING-HELICASE-LIKE 1 proteins FRG1 (a.k.a. CHR27) and FRG2 (a.k.a. CHR28) to allow efficient DNA methylation through the RdDM pathway^{42,43}. Finally, SUVR5 has been involved in the repression of stress-related genes through the deposition of H3K9me2 within their promoters, in a DNA methylation-independent manner⁴⁴.

6. DNA methyltransferases (DNMTs).

DNA methylation at cytosine residues (5-mC) is another important epigenetic mark that is mediated by DNA methyltransferases (DNMTs), and predominantly involved in gene silencing. In plants, DNA methylation occurs in three different cytosine contexts: CG, CHG and CHH (where H= A, T or C), involving specialized DNMTs⁴⁵. In *A. thaliana*, DNA methylation is mostly enriched in

pericentromeric regions of the chromosomes^{46, 47}, associated with H3K9me2 and H3K27me1. Nevertheless, DNA methylation can also be found in the body or promoters of expressed genes (see below).

DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) mediates de novo DNA methylation (a.k.a. INITIATION) in all sequence contexts through the RdDM pathway (Figure 4). RdDM involves short interfering (si)RNAs produced through the sequential activities of RNA-dependent RNA polymerase 2 (RDR2) and DICER-LIKE 3 (DCL3) proteins, and two specialized, plant-specific RNA polymerases RNA Pol IV and Pol V^{48, 49}. As mentioned above, the recruitment of RNA Pol V to the chromatin requires, among other factors, the activity of SUVH2 and SUVH9, which recognize methylated DNA through their SET- or RING-associated (SRA) domain^{38, 40}. The maintenance of CG methylation requires DNA METHYLTRANSFERASE 1 (MET1) and the 5-mC-binding proteins VIM1, VIM2 and VIM3⁴⁵. CMT3 and CMT2 are involved in the maintenance of CHG methylation, through a self-reinforcing loop involving H3K9me2 recognition by their chromodomain^{38, 45, 50}. Finally, CHH methylation is maintained by DRM2 through RdDM, and by CMT2 (Figure 4)^{38, 45}. DRM2 tends to preferentially target short TEs and edges of TEs, where H3K9me2 level is low⁵⁰. CMT2 targets long pericentromeric TEs with high level of H3K9me2⁵⁰. While DRM2 and MET1 are homologs of the mouse DNMT3A/B and DNMT1, respectively; CMTs are plant-specific DNMTs⁴⁵.

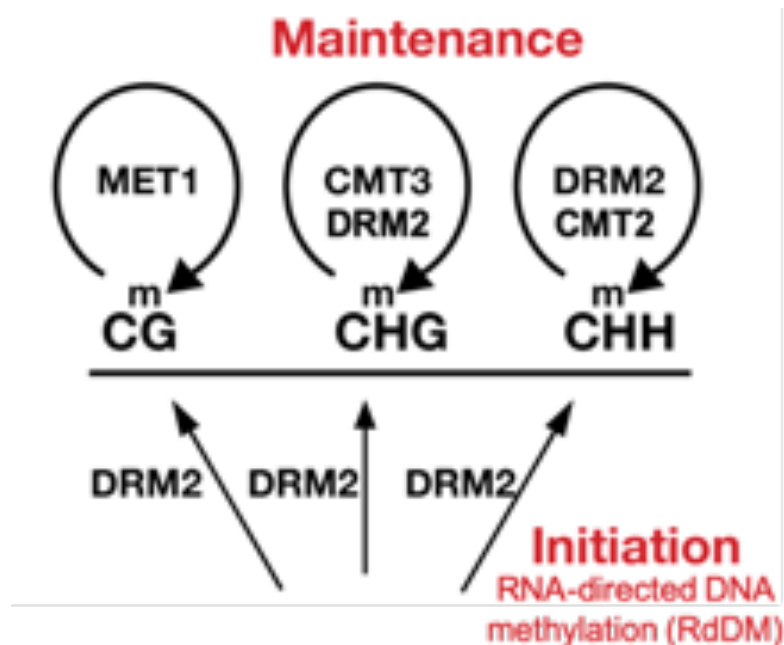


Figure 4. DNMTs involved in the initiation (de novo) and maintenance of DNA methylation in *A. thaliana*. DRM2 is required for de novo deposition of 5-mC in all cytosine sequence contexts through the RdDM pathway. After each cell division, mCG is exclusively maintained by MET1, while the maintenance of mCHG and mCHH requires DMR2, CMT3 and/or CMT2 activity. DRM1, not

represented here, is specifically expressed in egg cell, where it is involved in mCHH deposition⁵¹. This scheme has been modified from reference⁵².

7. Additional roles of DNA methylation.

DNA methylation plays an important role in the silencing of TEs. In *A. thaliana*, most TEs are heavily methylated in all cytosine contexts, which prevents their expression and transposition. Beside TEs, DNA methylation is also observed in the body of approximately one third of *A. thaliana* genes^{46, 47}. Gene body methylation (GbM) is evolutionary conserved among land plants, and almost exclusively found in the CG context⁵³. Genes displaying GbM tend to be constitutively expressed (housekeeping genes)⁵⁴. GbM genes would evolve slower than unmethylated genes, which suggests that GbM could, somehow, contribute to genome evolution pace. GbM is also found in animals, like for instance in several Hymenoptera species, such as *Apis mellifera* (honey bee)^{53, 55}.

Until very recently, the role of GbM in plants was still matter of debate. It had been proposed that GbM could regulate splicing, or inhibit the production of cryptic transcripts by repressing internal cryptic promoters^{53, 54}. A recent study has confirmed that GbM is indeed involved in the repression of intragenic antisense transcripts, acting in concert with the linker histone H1⁵⁶.

DNA methylation is primarily an epigenetic mark repressing transcription. However, several studies have reported that DNA methylation within the promoter of genes can enhance their expression. In *Petunia*, the ectopic expression of *pMAD3*, a class-C homeotic gene, is stimulated by the methylation of a specific CG site within its promoter⁵⁷. In human, CG methylation of *FoxA2* promoter is required for the expression of this transcription factor (TF), which is a master regulator of endoderm development⁵⁸. In *A. thaliana*, the expression of *REPRESSOR OF SILENCING 1 (ROS1)*, which encodes a DNA glycosylase/lyase involved in active DNA demethylation, requires DNA methylation of *ROS1* promoter^{59, 60, 61}. This phenomenon coined “methylstat” would act as an epigenetic “rheostat” (or “thermostat”) sensing DNA methylation level at *ROS1* promoter to control the genome-wide DNA methylation homeostasis⁵⁹. Finally, the *A. thaliana* SUVH1, SUVH3 and DNAJ domain-containing proteins DNAJ1 and DNAJ2 form a protein complex that reads DNA methylation within the promoter of genes to enhance their expression⁴¹. Thus, the SUVH1, SUVH3, DNAJ1 and DNAJ2 proteins can be considered as epigenetic readers⁴¹.

8. Epigenetic writers, erasers and readers.

As previously mentioned epigenetic modifications are reversible, and altogether, they define particular epigenetic landscapes recruiting specific protein factors. Among epigenetic writers, are DNMTs, HKMTs, HATs or PRMTs (Figure 5). Epigenetic marks are actively removed through the activity

of specialized enzymes called epigenetic erasers (Figure 5). In animals, the ten-eleven translocation methylcytosine dioxygenases (TETs) have been involved in active DNA demethylation⁶². In plants, 5-mC DNA glycosylase/lyases such as ROS1, TRANSCRIPTIONAL ACTIVATOR DEMETER (DME) and DME-LIKE (DML) proteins recognize and remove 5-mC⁵⁹. Jumonji C (JmjC) and histone lysine demethylases (HKDMs) catalyze the removal of methyl groups on lysine residues⁶². Some JmjC have also been involved in the demethylation of arginine residues²⁷. HDACs are another kind of epigenetic erasers, removing acetyl groups covalently bound to lysine residues of histones⁶². Epigenetic readers carry various protein domains, each specialized in the recognition of a specific epigenetic mark (Figure 5). There are hundreds of epigenetic readers. For instance, the SRA and methyl-CpG-binding (MBD) domains are motifs binding 5-mC, both in animals and plants⁶². Tudor domain, plant homeodomain (PHD), chromodomain recognize histone methylation⁶². Finally, bromodomain is an epigenetic reader binding acetylated histones⁶².

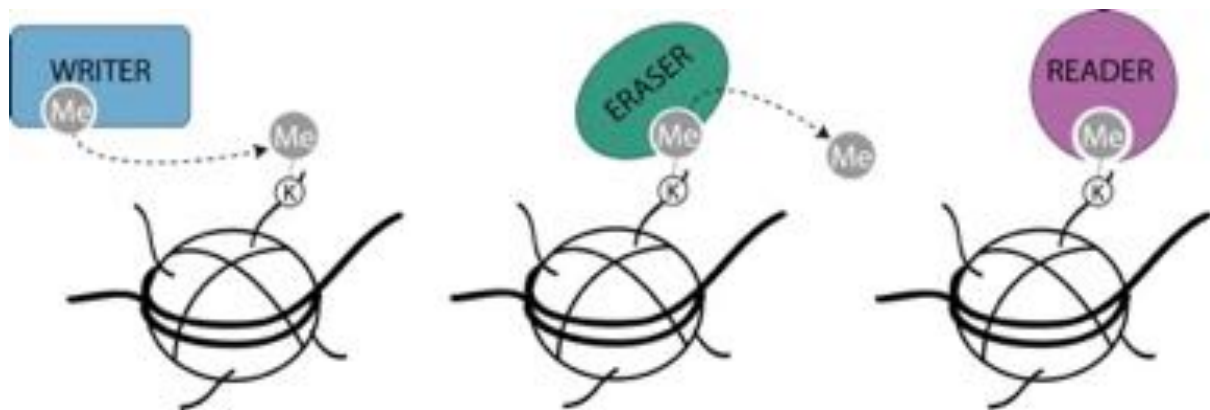


Figure 5. Epigenetic writers, erasers and readers. Schematic representation of writing, erasing and reading epigenetic marks, exemplified by histone lysine methylation, by distinct enzymatic activities and protein domains. Me= methylation, K=lysine.

9. Other epigenetic players and the epigenetic “mille-feuille”.

In addition to DNA methylation and histone modifications, other epigenetic players have been identified throughout the years. They are often involved in specific mechanisms and pathways contributing to gene silencing, regulation of gene expression and chromatin organization. Below is a non-exhaustive description.

Different histone variants have been linked to specific epigenetic states and chromatin-related processes⁶³. The centromeric H3 variant (called CENP-A in vertebrates, Cse4 in yeast and CENH3 in plants) is a marker of centromeric heterochromatin⁶³. H2AW and macroH2A variants are enriched at constitutive heterochromatin, and correlate with transcriptional silencing in plants and animals, respectively⁶³⁻⁶⁶. In *A. thaliana*, H3.1 and H3.3 variants are respectively found at heterochromatic and

actively-transcribed euchromatic regions of the genome^{66, 67}. Conversely, in animals, H3.3 is also enriched at heterochromatic and telomeric regions⁶³. H2A.Z is a multi-faceted histone variant, primarily found in nucleosomes located in gene promoters, and inversely correlated with DNA methylation^{63, 68}. In promoters, H2A.Z is involved in transcription initiation by recruiting RNA Pol II⁶³. H2A.Z is also enriched in the body of genes that are rather lowly expressed, and respond to environmental and developmental stimuli⁶⁹. Depending on the genomic location and organisms, H2A.Z plays antagonistic roles. In yeast, it has been involved in the protection of euchromatic regions from ectopic spreading of telomeric heterochromatin⁶³. In *D. melanogaster*, H2A.Z accumulates in pericentric heterochromatin, where it is required for heterochromatin assembly⁶³. Finally, the histone variant H2A.X plays an important role during DNA damage response (DDR). Upon DNA double-strand breaks (DSBs), H2A.X is phosphorylated (γ H2A.X), which initiates DDR⁶⁴.

Chromatin remodelers are protein complexes of various compositions playing essential role in most of chromatin-related processes such as regulation of transcription, gene and TE silencing, DNA replication, repair, and chromatin architecture⁷⁰⁻⁷². They are versatile tools used by the cell to reorganize chromatin through the regulation of nucleosome/DNA accessibility, or by performing nucleosome assembly, sliding, eviction or editing⁷⁰⁻⁷². There are 4 distinct families of ATP-dependent chromatin remodelers: chromodomain-helicase-DNA binding (CHD), inositol requiring 80 (INO80), switch/sucrose-non-fermenting (SWI/SNF) and imitation switch (ISWI)⁷⁰⁻⁷². In *A. thaliana*, DECREASE IN DNA METHYLATION 1 (DDM1) is a SWI/SNF chromatin remodeler that is required for the maintenance of DNA methylation, TE silencing and CC condensation⁷³⁻⁷⁷. It has been suggested that DDM1 would evict the linker histone H1 from constitutive heterochromatin to allow DNMTs to access and methylate DNA³². MORPHEUS' MOLECULE 1 (MOM1) is a plant-specific CHD3-like ATPase that is also required for transcriptional gene silencing (TGS) and repression of TEs⁷⁸. However, unlike DDM1, MOM1 is not required for the maintenance of DNA methylation and CC condensation^{75, 77}. MOM1 acts synergistically with RNA Pol V (RdDM pathway) to silence heterochromatic loci⁷⁹. Recently, the two PROTEIN INHIBITOR OF ACTIVATED STAT (PIAS)-type SUMO E3 ligase-like proteins, PIAL1 and PIAL2 were identified as MOM1 interactors that are required for TE silencing⁸⁰.

Noncoding RNAs (ncRNAs) are important regulators of gene expression and chromatin organization¹². Among ncRNAs, we distinguish small RNAs, such as microRNAs (miRNAs), short interfering RNAs (siRNAs) or PIWI-interacting RNAs (piRNAs) that play roles in transcriptional and post-transcriptional gene silencing (TGS and PTGS)^{12, 48}. In plants, both 24-nt and 21-nt siRNAs have been involved in the RdDM pathway^{48, 49}. More recently, circular RNAs (circRNAs) and tRNA-derived small RNAs (tsRNAs) were identified as new ncRNA molecules involved in gene and TE silencing^{81, 82}. Finally, long noncoding RNAs (lncRNAs) constitute another class of RNA molecules involved in the regulation

of gene expression, usually ranging from 200-nt to several kb long, and transcribed by RNA Pol II, RNA Pol IV or Pol V^{83, 84}.

Together with other processes (not mentioned here), the above-described epigenetic pathways participate in the subnuclear organization of chromatin, regulation of gene expression and TE silencing. They also contribute to other fundamental cellular processes such as DNA repair and cell division. Thus, altogether, they can be represented as an “epigenetic mille-feuille”, which is layered by all the epigenetic pathways that shape the eukaryotic genome (Figure 6)⁸⁵.

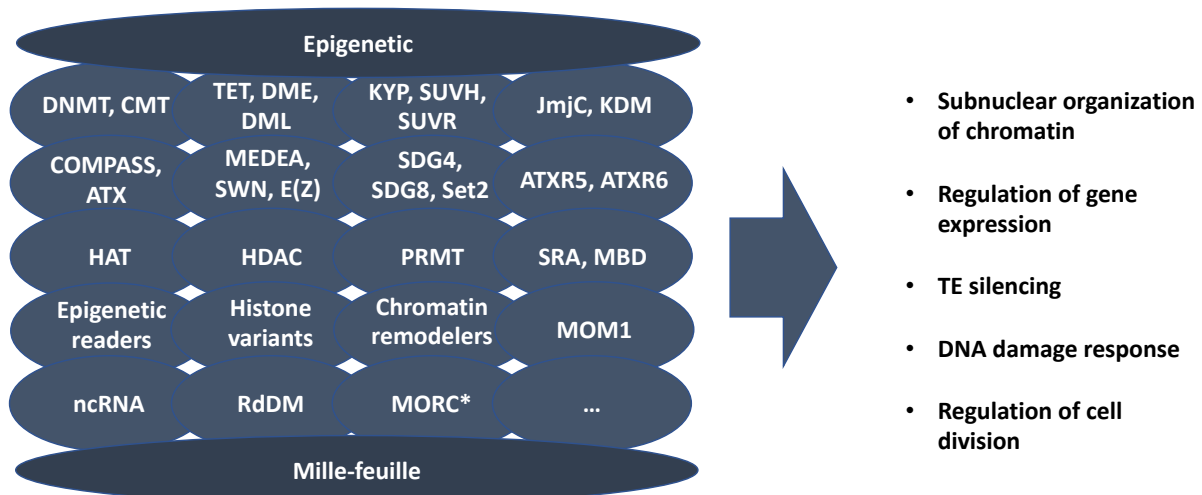


Figure 6. Intricate features of the “epigenetic mille-feuille”, composed of epigenetic pathways and players that have been described in this introduction. Please note that many epigenetic players were not described here. In addition, it is very likely that several epigenetic players are yet to be discovered. The notion of “epigenetic mille-feuille” was first proposed by Rigal and Mathieu in 2011⁸⁵.

* The Microrchidia (MORC) proteins will be described in Section V chapter 2.3.

10. The complex interplay between epigenetic modifications and transcription factors.

A precise and highly-coordinated regulation of transcription is essential among the cells of any given organism. The cellular mechanisms involved in the regulation of transcription are tremendously sophisticated, complex and various⁸⁶. Transcription factors (TFs) are major proteins regulating transcription that have been studied for decades⁸⁷. Beside their transcriptional activation domain and an optional signal-sensing domain, TFs carry a DNA-binding domain (DBD) that recognizes a specific DNA sequence, called cis-regulatory elements or modules (CREs or CRMs)^{86, 87}. There is a plethora of eukaryotic TFs, classified in five superclasses based on their DBD structure⁸⁸.

- **Superclass 1, Basic Domains**, including classes of Leucine zipper factors (bZIP), Helix-loop-helix factors (bHLH), NF-1, RF-X, bHSH.

- **Superclass 2, Zinc-coordinating DNA-binding domains**, including classes of Cys4 zinc finger (Znf) of nuclear receptor type, diverse Cys4 Znf, Cys2His2 Znf domain, Cys6 cysteine-zinc cluster, Znf of alternating composition.
- **Superclass 3, Helix-turn-helix**, including classes of Homeodomain, Paired box, Fork head / winged helix, Heat shock factors, Tryptophan clusters, TEA domain.
- **Superclass 4, beta-Scaffold Factors with Minor Groove Contacts**, including classes of RHD (Rel homology domain), STAT, p53, MADS box, beta-Barrel alpha-helix transcription factors, TATA-binding proteins, High mobility group (HMG), Heteromeric CCAAT factors, Grainyhead, Cold-shock domain factors, Runt.
- **Superclass 0: Other Transcription Factors** including classes of Copper fist proteins, HMGI(Y), Pocket domain, E1A-like factors, AP2/EREBP-related factors.

Some TFs are specific to metazoans, while others are also found in plants^{88, 89}. Conversely, some TFs are plant-specific such as, among others, APETALA 2 (AP2)/ethylene-responsive element binding factor (ERF), NAC [NAM (no apical meristem), ATAF1 and -2, and CUC2 (cup-shaped cotyledon)], WRKY, ABSCISIC ACID INSENSITIVE 3 (ABI3)/VIVIPAROUS1 (VP1); auxin response factor (ARF), and SQUAMOSA-promoter binding protein (SBP) proteins⁹⁰.

There are three different types of TFs.

- Pioneer factors that are essential TFs required for cell reprogramming and development⁹¹. Unlike other TFs, they have the capacity of binding heterochromatin to promote its decondensation, and subsequently allow the binding of specific and general TFs to allow gene expression⁹¹. Pioneer TFs are found in animals and plants. For instance, in animals, octamer-binding transcription factor 4 (OCT4) and SRY (sex determining region Y)-box 2 (SOX2) are essential factors for maintaining the self-renewal of ESCs, and reprogramming of somatic cells into pluripotent cells. In plants, LEAFY is required for floral meristem establishment; AP1 and SEPALLATA3 (SEP3) are essential for floral organ specificity by counteracting the PcG pathway to remove H3K27me3.
- Specific (or upstream) TFs are transcriptional activators or repressors acting through the recognition of a specific DNA sequence usually located upstream of the transcriptional start site (TSS)⁸⁶. They are often activated upon the perception of specific cell signals or environmental stimuli⁸⁶.
- General TFs are involved in the recruitment and formation of the RNA Pol II pre-initiation complex, through the recognition of specific DNA sequences, such as the TATA box, located nearby the TSS of any RNA Pol II-transcribed genes⁹². They act downstream of specific TFs.

In a well-established paradigm, the recruitment of a transcriptional activator TF at a specific gene promoter requires sequential events such as TF activation through various PTMs, chromatin accessibility and nucleosome eviction by diverse chromatin modifiers, removal of transcriptional repressors, and replacement of repressive epigenetic marks by activating marks. For instance, upon hypertrophic stimuli, the HAT p300 promotes Cys4 ZnF TF GATA4 and histones acetylation to allow the expression of cardiac genes in murine cardiomyocytes⁹³. Furthermore, histone acetylation allows the sequential recruitment and activation of RNA Pol II, as seen by single cell live imaging⁹⁴. In *A. thaliana*, the histone acetyltransferase GCN5 is required for the expression of heat stress-induced genes⁹⁴. Besides, the bZIP TFs bZIP28 and bZIP60 interact with the AtCOMPASS protein complex to promote H3K4 methylation, which will subsequently allow gene expression⁹⁵.

Conventionally, the binding of a TF to its CRE implies that the DNA motif is not methylated⁹⁶. Thus, DNA methylation or transcriptional repressors bound to 5-mC are usually sufficient to preclude most TFs from binding CREs (Figure 7A and B)⁹⁶. Nevertheless, DNA methylation can also positively influence TF binding. This has been specifically shown for several homeodomain TFs including the pioneer factor OCT4 and the RHD NF-AT TFs⁹⁷ (Figure 7C). Lastly, a TF can potentially recognize two distinct CREs based on their epigenetic signature (Figure 7D)⁹⁶.

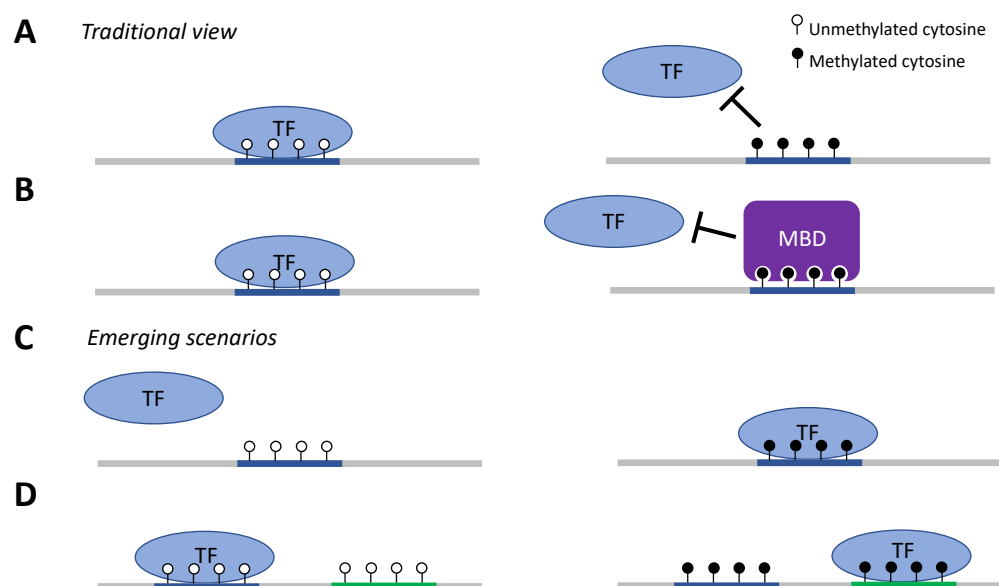


Figure 7. Traditional view and emerging scenarios describing TF/CRE interaction. A-B. Traditionally, TF/CRE interaction is only possible if DNA is not methylated (left panels). Either DNA methylation by itself (A, right panel) or the binding of a transcriptional repressor such as MBD protein (B, right panel) is sufficient to prevent TF binding. C-D. In these emerging scenarios, the TF can only bind methylated CRE (C, right panel), or shows distinct binding specificities for two different CREs that

are either unmethylated or methylated (D, left and right panels, respectively). Modified from reference⁹⁶.

11. Transgenerational epigenetic inheritance and memory.

Alterations of cell epigenome occur throughout the development of any organism. Several factors may promote epigenome modifications, such as for instance the genetic background, nutrition or exposure to particular environmental factors or chemicals¹². These epigenetic changes can be transgenerationally inherited throughout several generations, with progeny carrying an epigenetic memory of these alterations. This process is called transgenerational epigenetic inheritance (TEI), and if located within a gene or promoter, these epigenetic changes can create stable 'epialleles'¹². In animals, several studies have linked epigenetic modifications like H3K9 and DNA methylation, or ncRNAs such as siRNAs or piRNAs, as factors involved in transgenerational epigenetic inheritance (reviewed in ^{12, 98-101}). Originally described in plants, paramutation is another form of TEI involving interaction between two alleles, whereby the paramutagenic allele will convert the other one (called paramutated) into the same epigenetic state when both alleles are present in a heterozygote genotype¹⁰². Paramutation has also been described in animals, and like in plants, this phenomenon very often involves small RNAs¹⁰³. In *A. thaliana*, the epigenetic recombinant inbred lines (epiRILs) have been a powerful way to study TEI^{104, 105}. EpiRILs were initially generated by backcrossing either *met1* or *ddm1* mutants with wild type (WT) plants, and subsequently selecting and propagating independent inbred lines that were homozygote WT for each gene^{104, 105}. Loss of MET1 or DDM1 induces drastic changes in DNA methylation patterns (mostly CG hypomethylation, but also non-CG hypo- and CHG hypermethylation) that can be inherited in epiRILs, depending of the genomic location, and despite the restoration of enzymatic activities^{104, 105}. However, reestablishment of DNA methylation can occur at loci accumulating siRNAs, through the RdDM pathway¹⁰⁶. EpiRILs have also been a powerful tool to study TE transposition, and the role of TEs in TEI.

12. Roles of TEs in the organization, evolution and innovation of eukaryotic genomes.

TEs play important roles in the organization and evolution of eukaryotic genomes^{107, 108}. They contribute to genome architecture and dynamics, and modulate gene expression at the transcriptional and post-transcriptional level^{107, 109}. In specific genetic backgrounds (e.g. epigenetic mutants) or upon environmental stimuli, TEs can mobilize within the host genome. TE neoinsertions promote high mutation rates of the genome, which can sometimes be deleterious for the host, but also an important source of genetic variability, adaptation and innovation, through the modulation of gene expression and creation of genetic and phenotypic diversity^{107, 110, 111}. In plants, there are several examples of TE-driven modulation of gene expression that have led to phenotypic variations, with important

agronomic impacts¹⁰⁷. In mammals, a good example is the Eutherian-specific transposable element MER20, which controls the expression of a network of genes involved in placentation¹¹². Finally, in *A. thaliana*, a recent study has shown that retrotransposition of *ATCOPIA* TEs occurs predominantly in environmentally responsive genes displaying specific epigenetic signatures, suggesting that *ATCOPIA* retrotransposition can contribute to genetic diversity and genome adaptation in response to environmental changes¹¹³.

The process of TE gene domestication is another outstanding example of TE contribution to host genetic innovation¹⁰⁷. Also known as “gene exaptation” or “cooption”, this process can lead to the production of exapted TE (ETE) proteins involved in an entirely new function^{107, 114-116}. There are many documented *ETE* genes in the literature. For instance, in jawed vertebrates, the variable (V) diversity (D) joining (J), V(D)J recombination system, which is involved in immune response through the creation of an infinite variety of antibodies, requires the domesticated transposase RAG1 and recombination signal sequences (RSS) of TE origin¹¹⁴. Syncytins that are proteins involved in mammal placentation, derive from endogenous retroviruses (ERVs), which share similarities with retrotransposons¹¹⁴. In human, centromere binding protein B (CENP-B), which derives from a transposase of *Pogo*-like DNA transposons, is required for centromere formation and identity through the binding of a specific 17bp stretch of satellite DNA called the CENP-B box¹¹⁴. Telomerase reverse transcriptase (TERT), which is essential for telomere replication is likely to derive from *LINE*-like retrotransposons¹¹⁴. In prokaryotes, the Cas1 protein, involved in the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas system, show sequence similarities with transposases of *Casposons* TE¹¹⁴. Similarly, domestication of TE genes has greatly contributed to the diversification of plant protein repertoire, creating a panel of proteins that are involved in various cellular processes^{107, 114, 117}. The *hAT*-like transposase-derived Daysleeper is involved in DNA damage response and plant development¹¹⁸. Far-red impaired response 1 (FAR1) and FAR1-related sequences (FRS) proteins are TFs regulating the expression of genes involved in light perception and plant development¹¹⁹. The *PIF/Harbinger*-related transposase ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN1 (ALP1) antagonizes the PcG pathway, promoting the expression of PcG-target genes in the *clf* mutant¹²⁰. The Harbinger transposon-derived proteins (HDP1) and HDP2 are transposase- and SANT/Myb DNA binding-derived proteins, respectively, that form a histone acetyltransferase complex involved in active DNA demethylation¹²¹. Plant Mobile Domain (PMD) proteins are derived from *Ty3/Gypsy* TEs, and we and other labs have shown that they are involved in TE silencing, regulation of gene expression, genome stability and plant development (Nicolau et al., PLOS Genetics, in revision and¹²²⁻¹²⁴). Finally, *MULE* transposase-derived MUSTANG (MUG) proteins have been linked to plant development and fertility^{117, 125}.

V. Research studies.

1. PhD thesis.

I did my PhD at Institut de Biologie Moleculaire des Plantes (IBMP, Strasbourg) where I studied Post Transcriptional Gene Silencing (PTGS) as an antiviral defense mechanism, mostly in *A. thaliana*. I also described the effects of several viral proteins (called suppressors of RNA silencing) on transitivity, a mechanism developed by the plant to amplify the PTGS response through the activity of RNA-dependent RNA polymerases (RdRPs or RDRs), such as for instance RDR6¹²⁶. Finally, I initiated the characterization of a plant DNA virus-encoded suppressor of RNA silencing¹²⁷.

This part of my previous research will not be further described in this manuscript.

2. Postdoctoral research.

I joined Steve Jacobsen lab at UCLA, Los Angeles, USA in 2007, where I spent almost six years studying epigenetics using *A. thaliana* as a model organism. During this time, I had the opportunity to get involved in various projects, in an outstanding research environment.

Here, I will only describe the projects that lead to a publication.

2.1 The role of JmjC JMJ14 in the DRM2-mediated maintenance of DNA methylation.

This project was led by Angelique Deleris, another postdoctoral fellow. It was based on a reverse genetic approach aiming at identifying *jmj* mutants impaired in DNA methylation. Using bisulfite DNA sequencing (BS-seq) and chromatin-immunoprecipitation (ChIP) experiments, we showed that *jmj14* displayed DNA methylation defects at non-CG sites, and increase in H3K4me3 at DRM2 target sites. Further experiments suggested that JMJ14 was specifically involved in the DRM2-mediated maintenance of RdDM, most likely acting as a H3K4me3 HKDM¹²⁸.

2.2 A forward genetic screen identifies new epigenetic players.

Forward genetic is a powerful and unbiased approach aiming at identifying new components of a specific pathway. In the Jacobsen lab, my main project was to develop a forward genetic screen based on the *SDC::GFP* transgene, in which GFP expression is controlled by the promoter of *SUPPRESSOR OF drm1 drm2 cmt3 (SDC)* (Figure 8A). *SDC* promoter carries several tandem repeats, and in vegetative tissue, this gene is transcriptional repressed through the redundant activity of DRM2 and CMT3¹²⁹. In the endosperm of the seed, however, maternal copy of *SDC* is expressed, making *SDC* an imprinted gene^{130, 131}.

The *SDC::GFP* transgene was introduced in WT, *drm2*, *cmt3* single, and *drm2 cmt3 (dc)* double mutant backgrounds. As expected, the transgene was silenced in WT, *drm2* and *cmt3*, but strongly

upregulated in *dc* (Figure 8B). I performed an ethyl methanesulfonate (EMS) mutagenesis, and subsequently screened for mutant populations showing upregulation of GFP in WT, *drm2* or *cmt3* genetic backgrounds. Among others, the mutant populations WT #67, *cmt3* #49, *cmt3* #7 and WT #162 were identified (Figure 8C). The mutated genes responsible for the upregulation of the transgene were mapped using bulk segregant analyses coupled to whole genome resequencing. These mutants are described in sections 2.3 and 2.4.

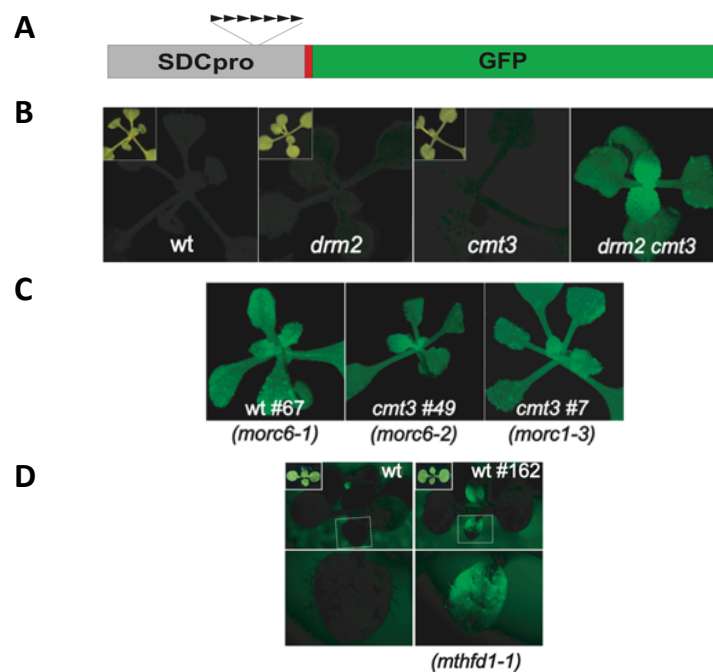


Figure 8. The *SDC::GFP* reporter gene allowed the identification of new epigenetic players. A. GFP is controlled by the *SDC* promoter, which carries seven tandem repeats. Red bar: nuclear localization signal. B. WT, *drm2* and *cmt3* show no GFP fluorescence under ultraviolet (UV). Conversely, GFP is upregulated in the *drm2 cmt3* double mutant. C-D. EMS-mutagenized populations WT #67, *cmt3* #49, *cmt3* #7 (C) and WT #162 (D) show strong GFP fluorescence. Insets: plants under white light. Figure 7 is a composite of figure panels that were published in references ¹³² and ¹³³.

2.3 The Microrchidia ATPases are required for gene silencing and heterochromatin compaction.

Mapping of mutant populations WT #67, *cmt3* #49 and *cmt3* #7 identified the mutated genes *MORC6* and *MORC1*, as responsible for GFP upregulation (Figure 8C).

The Microrchidia (MORC) proteins belong to the GHKL (gyrase, HSP90, histidine kinase, MutL) ATPases that are found in prokaryote and eukaryote organisms¹³⁴. They have been involved in various cellular processes such as TE and gene silencing, chromatin organization, DNA repair, plant immunity and cancer¹³⁴⁻¹³⁷. In *A. thaliana*, MORC1 and MORC6 (as well as MORC2 and MORC6) form heteromeric complexes, and accumulate as nuclear bodies localizing in the vicinity of CCs^{132, 138}. The *morc1*, *morc2* and *morc6* single, and higher order mutants thereof, show upregulation of several TEs and DNA-

methylated genes, and decondensation of constitutive heterochromatin^{132, 138-140}. In addition, The MORC1/6 pathway acts synergistically with MOM1 to repress TE¹³⁸. Conversely, MORC4 and MORC7 form homomeric complexes acting redundantly to regulate gene expression, and they are involved in plant response against the oomycete pathogen, *Hyaloperonospora arabidopsidis* (*Hpa*)¹⁴¹. Despite their role in the repression of TEs and DNA-methylated genes, the *morc* mutants show only modest changes in DNA methylation and H3K9me2 levels^{132, 139-141}. However, several reports showed that MORC1 and MORC6 interact with components of the RdDM pathway^{139, 142, 143}, and artificial targeting of MORC6 to a derepressed DNA-methylated gene can promote de novo DNA methylation and repression of the misregulated loci¹⁴⁴. Thus, it has been proposed that the MORC1/6 complex acts mostly downstream of DNA methylation, and would interact with components of the RdDM machinery to promote its own recruitment to heterochromatic loci that must remain condensed and silenced¹⁴⁴.

In mouse, the founding member Morc1 is required for male fertility, and *morc1* mutant exhibits small testis (hence the name microrchidia)¹⁴⁵. In the male germline, Morc1 is required for TE silencing, facilitating de novo DNA methylation¹⁴⁶. In nematode, the unique *Morc-1* gene is required for gene silencing, and recent biochemical studies showed that Morc-1 compacts chromatin through a DNA-loop trapping mechanism^{132, 147}. Considering that eukaryotic MORC proteins are highly conserved, it is likely that the *modus operandi* of other animal and plant MORCs is similar.

2.4 MTHFD1 controls DNA methylation in *A. thaliana*.

The *SDC::GFP* genetic screen identified the bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase (*MTHFD1*) gene as required for DNA and H3K9 methylation and, therefore, TE silencing (WT#162, [Figure 8C](#))¹³³. In short, *mthfd1* mutant is impaired in folate metabolism, which induces a misregulation of the methionine cycle and change in the methylation index [MI = S-Adenosylmethionine / S-Adenosylhomocystein], hence DNA and H3K9 methylation defect in this mutant¹³³.

3. Current research projects.

Since October 2016, I have been appointed as CRCN at LGDP where I co-lead, together with Frederic Pontvianne, the team Epigenetic Mechanism and Chromatin Architecture (MEAC). In brief, our team studies *i)* the mechanisms involved in the subnuclear organization of chromatin, *ii)* the epigenetic mechanisms regulating gene expression and repressing TEs, *iii)* the phenomena of TE gene domestication (“*ETE* genes”) and tandem duplication in direct orientation (TDDO) as sources of genetic innovation in plant.

3.1 Characterization of the *ATCOPIA28::GFP* system.

To identify new pathways involved in the silencing of TEs, I decided to engineer another GFP-based transgene, called *ATCOPIA28::GFP*, in which GFP expression is controlled by the 5' long terminal repeat (LTR) promoter region of a TE belonging to the *ATCOPIA28* retrotransposon family (Figure 9A). *ATCOPIA28::GFP* was first introduced in the *drm1 drm2 cmt3 (ddc)* triple mutant, and primary transformants showing strong GFP expression were selected. These plants were and further characterized to finally keep one independent line, in which segregation analyses confirmed that a single insertion event of *ATCOPIA28::GFP* had occurred. This line called "*ddc+++*" exhibits strong GFP fluorescence, which is transgenerationally maintained (Figure 9B and C). Backcrossing *ddc+++* to WT (F1 generation) promoted *ATCOPIA28::GFP* silencing, most likely through the RdDM pathway. In the F2 generation, WT plants carrying the transgene (*ATCOPIA28::GFP* WT) showed no GFP fluorescence (Figure 9D). By contrast, the F2 plants that were genotyped as *ddc* (called "*ddc+*") showed intermediate level of GFP fluorescence, in between *ATCOPIA28::GFP* WT and "*ddc+++*" (Figure 9E). F2 *ATCOPIA28::GFP* WT plant were ultimately crossed with untransformed *ddc* plants to generate F2' *ATCOPIA28::GFP ddc* plants (called "*ddc-*") that showed no GFP fluorescence (Figure 9F).

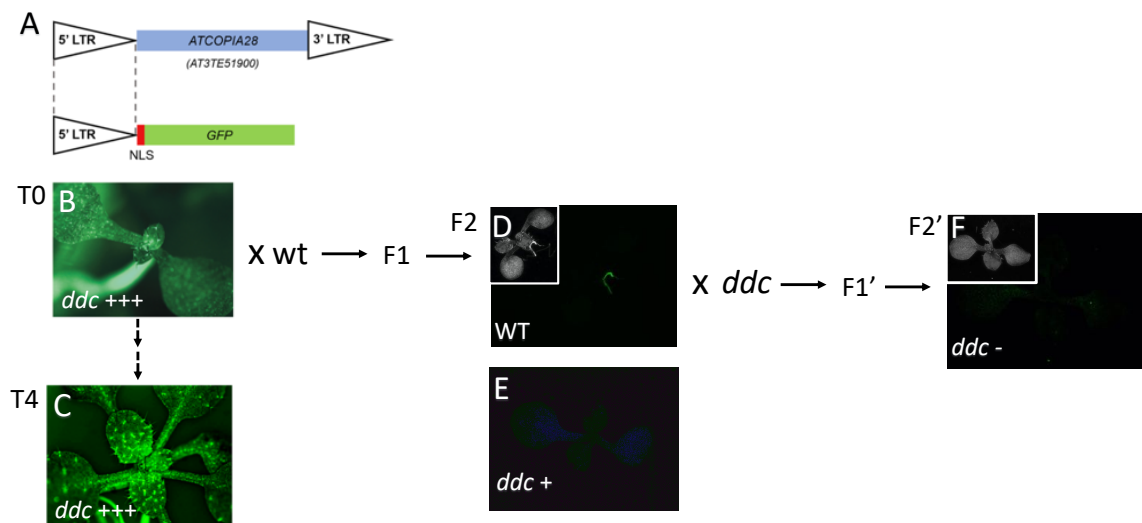


Figure 9. Transgenerational analyses of *ATCOPIA28::GFP* expression. A. *ddc* plants were transformed with *ATCOPIA28::GFP* to obtain the line *ddc+++* showing strong GFP fluorescence under UV light. B. GFP expression is transgenerationally maintained through several generations in the progeny of *ddc+++*. C-D. Backcrossing of *ddc+++* with WT Columbia plants allowed the identification of lines WT and *ddc+*, showing no and low GFP fluorescence, respectively. E. Reintroducing the *ddc* mutation by crossing the line WT with a *ddc* plant that does not carry the transgene led to the identification of line *ddc-*, which show no GFP fluorescence.

The expression level of *ATCOPIA28::GFP* in WT, *ddc +*, *ddc + + +* and *ddc -* anticorrelates with DNA methylation level at the 5'LTR (Figure 10A and B). In WT, the 5'LTR is highly methylated in the three cytosine sequence contexts, hence the silencing of the transgene. In *ddc + + +*, in which *DRM2* has remained mutated, there is no DNA methylation and strong GFP fluorescence. In *ddc +*, mCG shows an intermediate level, most likely because *DRM2* was not able of fully reestablish DNA methylation in the F1 parental line (Figure 9). By contrast, *ddc -* displays a WT level of mCG, probably because it derives from a cross using the WT line, in which *DRM2* had enough time to efficiently reestablish DNA methylation, and recruit the other DNMTs (Figure 10B). Hence, despite the lack of *DRM2* and *CMT3* activities in *ddc -*, the *MET1*-mediated mCG maintenance is sufficient to keep *ATCOPIA28::GFP* silenced. This makes *ATCOPIA28::GFP* different than *SDC::GFP*, whose silencing is impaired in *ddc*¹³².

In conclusion, through this genetic analysis, four *ATCOPIA28::GFP* epialleles have been generated. One epiallele is in WT; the three others are in *ddc* (Figure 10B).

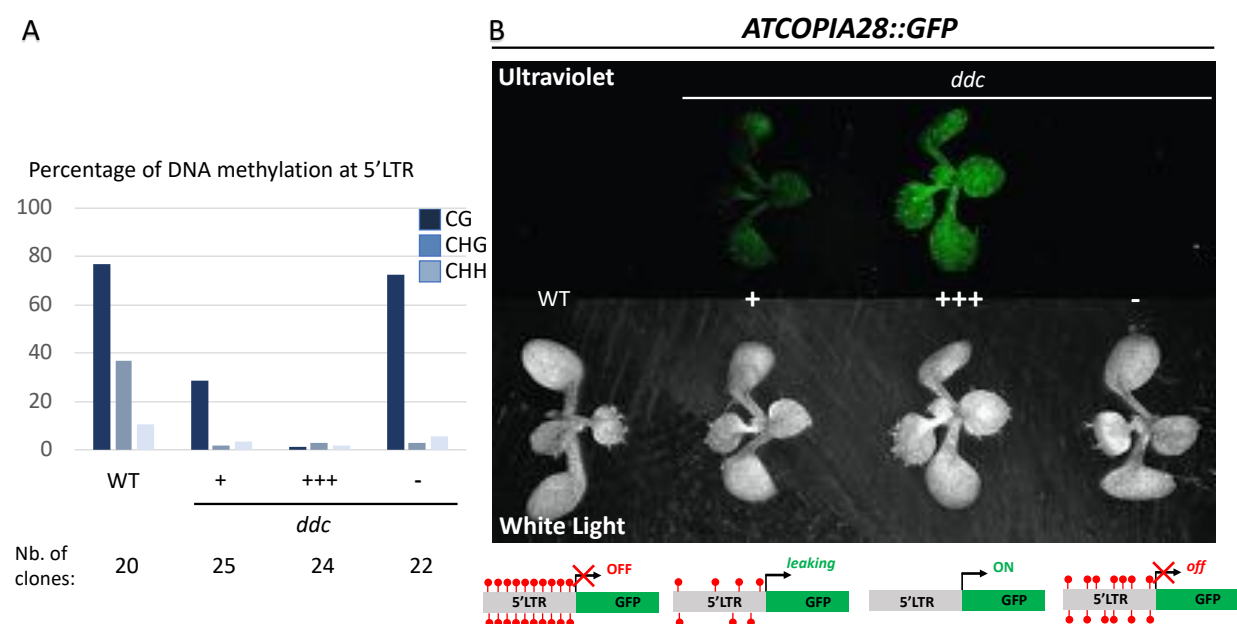


Figure 10. Characterization of four *ATCOPIA28::GFP* epialleles obtained in WT and *ddc* genetic backgrounds. A. DNA methylation analyses at the 5'LTR of *ATCOPIA28::GFP* in WT, *ddc +*, *ddc + + +* and *ddc -* assayed by bisulfite DNA sequencing. Nb of clones = number of clones that were sequenced for the analysis. B. Level of GFP fluorescence in the four *ATCOPIA28::GFP* lines, which anticorrelates with the DNA methylation status of the 5'LTR.

3.2 The *ATCOPIA28::GFP* forward genetic screen.

Following their characterization, I performed an EMS mutagenesis of WT and *ddc +* lines. The 5-mC-deficient *ddc +* line is considered as a sensitized genetic material, in which any additional mutation

in another epigenetic pathway converging towards the silencing of the transgene, would lead to a strong expression of the GFP. For each mutagenesis, approximately 1500 M2 mutant populations were generated. The WT screen is now finished. Nine M2 populations showing a moderate increase in GFP fluorescence were selected, and mutations responsible for the phenotype were mapped. Unfortunately, this screen did not allow us to identify new epigenetic players, as we only retrieved *MOM1* or *MET1* mutant alleles. The *ddc+* screen is still in progress, and so far, half of the M2 populations have been screened. Among the seventeen M2 *ddc+* populations that we mapped, the population *ddc #16* drew our attention. A summary of these screens is described in [Table 1](#).

Sequencing code	Population number	Candidate gene	Type of mutation
S1	WT#42	<i>MOM1</i>	nonsense
S2	WT#64	<i>MOM1</i> *	deletion
S3	WT#82	<i>MET1</i>	splicing defect
S4	WT#136	<i>MOM1</i> *	deletion
S5	<i>ddc</i> #16	<i>MAIN</i> *	missense
S6	<i>ddc</i> #18	<i>MORC6</i> *	missense
S7	<i>ddc</i> #28	<i>PIAL2</i>	nonsense
S8	<i>ddc</i> #72	<i>UBP26</i>	nonsense
S9	WT#909	<i>MOM1</i> *	nonsense
S10	WT#1049	<i>MOM1</i> *	nonsense
S11	WT#1421	<i>MOM1</i>	nonsense
S12	WT#1517	<i>MOM1</i>	nonsense
S13	WT#1568	<i>MOM1</i>	nonsense
S14	<i>ddc</i> #55	<i>MOM1</i>	nonsense
S15	<i>ddc</i> #100	<i>MORC6</i>	nonsense
S16	<i>ddc</i> #109	<i>UBP26</i>	missense
S17	<i>ddc</i> #117	<i>MOM1</i>	splicing defect
S18	<i>ddc</i> #124	unresolved	-
S19	<i>ddc</i> #194	tbd	-
S20	<i>ddc</i> #218	<i>HDA6</i>	nonsense
S21	<i>ddc</i> #259	<i>ATMS1</i> [§]	missense
S22	<i>ddc</i> #333	<i>PIAL2</i>	splicing defect
S23	<i>ddc</i> #344	<i>MORC6</i> *	missense
S24	<i>ddc</i> #370	unresolved	-
S25	<i>ddc</i> #431	<i>MORC6</i>	nonsense
S26	<i>ddc</i> #447	unresolved	-

Table 1. Summary of WT and *ddc* mutant populations that have been mapped. * confirmed by SANGER DNA sequencing. § most probable candidate. tbd: to be determined. *METHIONINE SYNTHASE1* (*ATMS1*) has been recently identified as required for DNA and H3K9 methylation at heterochromatic regions¹⁴⁸.

3.3 The Plant Mobile Domain (PMD) protein MAINTENANCE OF MERISTEMS (MAIN) is required for the *ATCOPIA28::GFP* silencing.

Bulk segregant analyses coupled to whole genome resequencing of *ddc #16* allowed the identification of an EMS mutation, creating the missense mutation (C230Y) within *MAINTENANCE OF MERISTEMS (MAIN)*. Genetic complementation analyses by crossing the EMS mutant *ddc #16* with the knock-out (KO) transferred DNA (T-DNA) insertion line *main-2* confirmed that *MAIN* was the mutated gene causing silencing defects. Thus, *ddc #16* was renamed *ddc main-3*.

This work is described in **Nicolau et al., PLOS Genetics, in revision, the plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*.**

In brief, the MAIN, DRM2 and CMT3 pathways act synergistically to silence TEs. MAIN and its close homolog MAIN-LIKE 1 (MAIL1) physically interact together to regulate a similar set of genes and silence TEs. The two PMD proteins interact also with a putative serine/threonine-specific phosphoprotein phosphatase (PPP) called PP7-LIKE (PP7L), and *main*, *mail1* and *pp7l* single and *mail1 pp7l* double mutants display similar developmental and molecular phenotypes. A substantial fraction of genes that are commonly downregulated in the *pmd* and *pp7l* mutants carry within their promoter a highly similar DNA motif (called the 'DOWN' motif), which suggests that transcriptional regulation of these loci could occur through the recognition of this DNA motif (see below section 3.6).

3.4 A brief introduction to the Plant Mobile Domain (PMD) proteins.

The PMD is a protein domain largely represented among the angiosperms, and phylogenetically spread over three clades¹²⁴. Early in silico analyses suggested that genic PMD versions derived from *Ty3/Gypsy* TEs upon TE gene domestication^{149, 150}. Later in the evolution, genic PMDs would have been captured by *Mutator-like elements (MULE)* *MuDR* TEs through the process of transduplication^{124, 149}. While genic PMDs belong to the PMD-B and -C clades, the *Gypsy*- and *MuDR*-associated PMDs belong to the PMD-A and -C clades, respectively¹²⁴. Comparative analyses of 33 angiosperm genomes showed that several plant species, such as rice or tomato for instance, carry both TE-associated and genic PMDs¹²⁴. Conversely, the genome of *A. thaliana* is deprived of TE-associated PMDs, and contains only genic versions. Protein domain analyses revealed that among genic PMDs, the standalone version is the most represented¹⁴⁹. Nevertheless, in many species, the PMD is fused to other protein domains such as, for instance, PPP, kinase or protease domains. The genome of *A. thaliana* encodes thirteen PMD proteins. The standalone PMD-C versions MAIN and MAIL1 have been related to genome stability, developmental processes, regulation of gene expression and TE silencing¹²²⁻¹²⁴ (and Nicolau et al., PLOS Genetics, in revision). They have two uncharacterized close homologs called MAIL2 and

MAIL3, and nine related PMD proteins named PMD5 to PMD13 of unknown function¹²⁴. MAIL3 is the unique *A. thaliana* protein harboring a PMD fused to a PPP domain, whose substrates remain unknown (Figure 11A). MAIN and MAILs proteins carry putative nuclear localization signals (NLS), and previous works showed that they are *bona fide* nuclear proteins¹²²⁻¹²⁴. We confirmed that MYC-epitope tagged MAIN and MAIL1 are localized in the nucleus (Figure 11B).

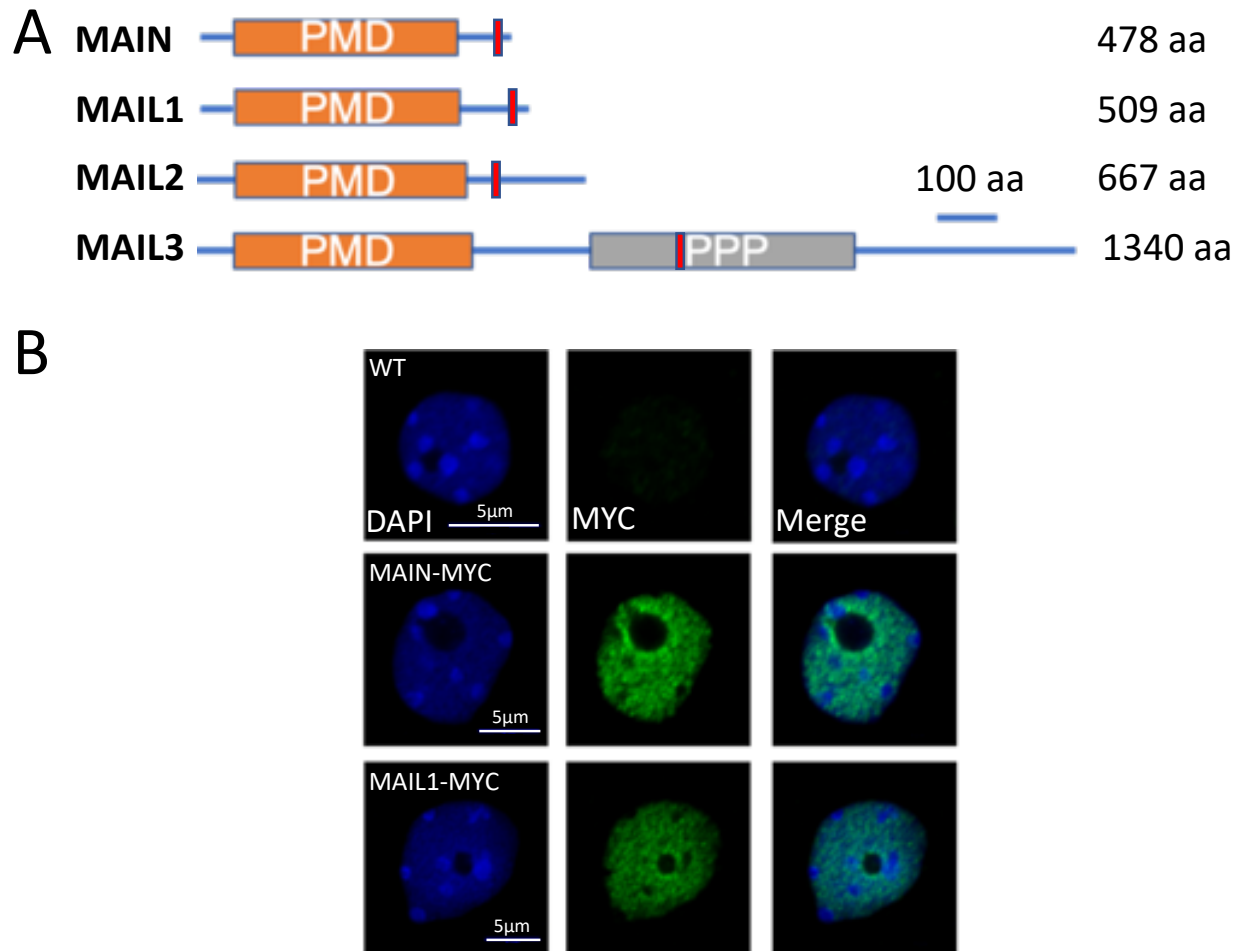


Figure 11. PMD-C protein organization and subnuclear localization. A. Protein domains of MAIN, MAIL1, MAIL2 and MAIL3 proteins with predicted nuclear localization signals (NLS) depicted by red rectangles. B. Subnuclear localization of MAIN- and MAIL1-MYC proteins assayed by immunofluorescence (IF) experiments using anti-MYC antibody. DAPI: = 4',6-diamidino-2-phenylindole staining of DNA. PMD: Plant Mobile Domain. PPP: serine/threonine-specific phosphoprotein phosphatase, aa: amino acids.

3.5 MAIN and MAIL1 are required for TE silencing and the proper expression of several genes, including *MORC1* and *FLOWERING LOCUS C (FLC)*.

Three independent RNA sequencing (RNA-seq) analyses have precisely defined the sets of TEs and genes that are misregulated in the *main-2* and *mail1-1* null mutants (Nicolau et al., PLOS Genetics, in revision). There are significant overlaps of misregulated loci between *main-2* and *mail1-1*, which is consistent with the similar developmental phenotypes of the two mutants (Figure 12A and B, and Nicolau et al., PLOS Genetics, in revision). We then identified subsets of genes and TEs that were commonly misregulated in *main-2*, *mail1-1* and the hypomorphic *main-3* mutant (Figure 12B). Remarkably, among the downregulated genes, we identified *MORC1*, and confirmed by quantitative reverse transcription PCR (RT-qPCR) that *MORC1*, but not *MORC6*, is indeed downregulated in *main-2*, *mail1-1* and *main-3* as well as in the *main-2 mail1-1* double mutants (Figure 12C-E). Thus, MAIN and MAIL1 are required for *MORC1* expression. We then compared the subsets of loci that were commonly misregulated in *main-2*, *mail1-1* and *main-3* with publicly available *morc1-4* RNA-seq data to determine the overlaps of misregulated loci between the *pmd* and *morc1* mutants¹³². We found that 39% of upregulated TEs and 25% of upregulated genes in the *main-2*, *mail1-1* and *main-3* overlaps were also upregulated in *morc1-4*. (Figure 12F). Conversely, no overlap was found between the downregulated genes in the *pmds* and *morc1* mutants (Figure 12F and G). Furthermore, all the genes commonly upregulated in the *pmds* and *morc1-4* were pericentromeric, DNA-methylated and transcriptionally-repressed genes in WT plants¹⁵¹.

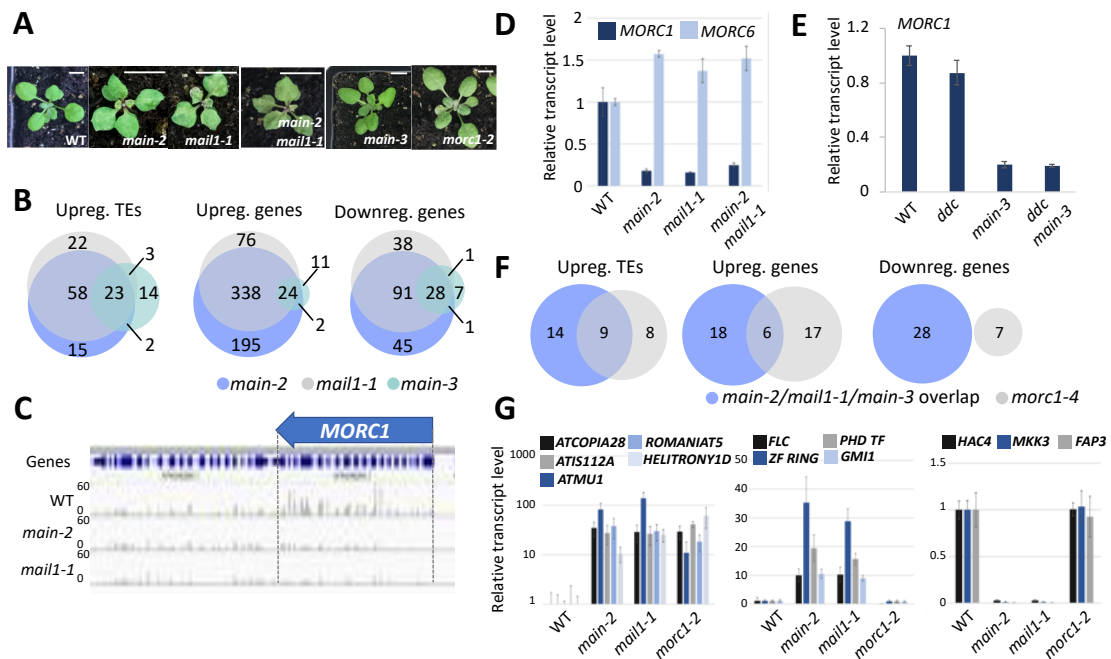


Figure 12. *MORC1* is downregulated in *main* and *mail1* mutants. A. Developmental phenotypes of *main-2*, *mail1-1*, *main-3*, *morc1-2* single and *main-2 mail1-1* double mutants in comparison to WT plant. bar = 1cm. B. Venn diagram analyses of misregulated loci in *main-2*, *mail1-1* and *main-3* (threshold: bar = 1cm).

$\log_2 \geq 2$ or $\log_2 \leq -2$; $\text{padj} < 0,01$). C. Snapshot of *MORC1* genomic location and mRNA accumulation from RNA-seq analyses in *main-2*, *mail1-1* and WT plants. D. Relative *MORC1* and *MORC6* transcript levels in *main-2*, *mail1-1* and *main-2 mail1-1* mutants over WT assayed by RT-qPCR and normalized to the housekeeping *RHIP1* gene. E. Same analyses as in D using *ddc*, *main-3* and *ddc main-3* mutants. F. Venn diagram analyses comparing the misregulated loci defined in the overlaps between *main-2*, *mail1-1* and *main-3* in panel B, with misregulated loci in the *morc1-4* mutant from publicly available *morc1-4* RNA-seq data¹³². G. RT-qPCR validation of overlapping TEs, and non-overlapping upregulated and downregulated genes described in F.

In conclusion, these results suggest that upregulation of TEs and DNA-methylated genes observed in the *main* and *mail1* mutants could be, at least for a subset of loci, an indirect effect of *MORC1* downregulation. To test this hypothesis and determine whether MAIN and MAIL1 regulate *MORC1* expression, we are currently developing several approaches:

- I) Chromatin-immunoprecipitation coupled to quantitative PCR (ChIP-qPCR) experiments using epitope-tagged MAIN/MAIL1 expressing lines to test if the two PMDs are enriched at *MORC1* genomic location (see also paragraph 3.6.1).
- II) Transformation of *main-2* and *mail1-1* plants with a *MORC1* transgene driven by a ubiquitous promoter (*UBIQUITIN10* promoter) to determine if *MORC1* overexpression can rescue some TE and gene silencing defects in the *main-2* and *mail1-1* mutants.
- III) RNA seq analyses of *pmd morc1* higher order mutants to determine the genetic interaction between the *pmd* and *morc1* mutations.

Considering that *pmds* and *morc1-2* mutants share a common set of upregulated TEs and genes, and *morc1-2* and WT plants are phenotypically undistinguishable (Figure 12A, F-G), it is unlikely that TE silencing defects observed in the *pmd* mutants account for their abnormal developmental phenotypes. Instead, it is likely that the *pmd* developmental phenotype is the aftermath of misregulation of genes essential for plant fitness. Among the genes that were commonly misregulated in *main-2* and *mail1-1* null mutants, we found enrichments for the gene ontology (GO) term ‘response to stress’ ($p\text{-value}=4.19\text{e-}14$, upregulated genes) and ‘disulfide oxidoreductase activity’ ($p\text{-value}=2.81\text{e-}3$, downregulated genes). Thus, these GO term enrichments could explain the abnormal developmental phenotype of *main-2* and *mail1-1* null mutants. Furthermore, we did not identify GO term enrichment in the lists of genes commonly misregulated in *main-2*, *mail1-1* and *main-3*. However, among these lists, we have identified, beside *MORC1*, several outstanding genes that are involved in important developmental and cellular processes. These genes are described in Figure 13A,

based on TAIR annotations, Araport ThaleMine and/or the following references¹⁵²⁻¹⁶¹. Among the genes that are upregulated *main-2* and *mail1-1*, we identified *FLC*, which is a MAD-box TF acting as floral repressor (Figure 13B)¹⁵². *FLC* is repressed by several pathways including PcG-mediated H3K27me3 deposition, the autonomous pathway and the lncRNA COOLAIR^{152, 162}. One hypothesis is that *FLC* upregulation leads to flowering delay of *main-2* and *mail1-1* mutants, and preliminary results suggest that H3K27me3 deposition at *FLC* is not impaired in the *main-2 mail1-1* double mutant (Figure 13C and D). To confirm this hypothesis, it will be essential to perform a proper flowering time assay in *main-2* and *mail1-1* plants. Importantly, flowering time assay using *pp7l-2* mutant showed that this mutant does not display a late flowering phenotype¹⁶⁵. Therefore, it is likely that *FLC* upregulation in *main-2*, *mail1-1* and *pp7l-2* plants is not causing a late flowering phenotype, but instead it is reflecting the growth delay of these mutants in comparison de WT plant. In the line of this second hypothesis, it is likely that *main-2*, *mail1-1* and *pp7l-2* plants show some plastochron defects.

Finally, the *main-2* and *mail1-1* mutant plants display strong developmental phenotypes: previous studies showed that the shoot apical meristem (SAM) is disorganized in the two *pmd* mutants, and that *main-2* develops fasciated stems^{122, 123}. Moreover, *main-2* flowers display impaired determinacy, leading to the production of extra carpels or flowers-in-flowers¹²³. To decipher the role of PMDs in SAM homeostasis and floral morphogenesis, we will analyze the effect of *pmd* mutations on the expression patterns of meristem and flower marker genes, using in situ hybridization (ish) and reporter-based transgenic lines. This part of the PMD project involves a collaboration with Christel Carles at the Plant and Cell Physiology Laboratory in CEA Grenoble.

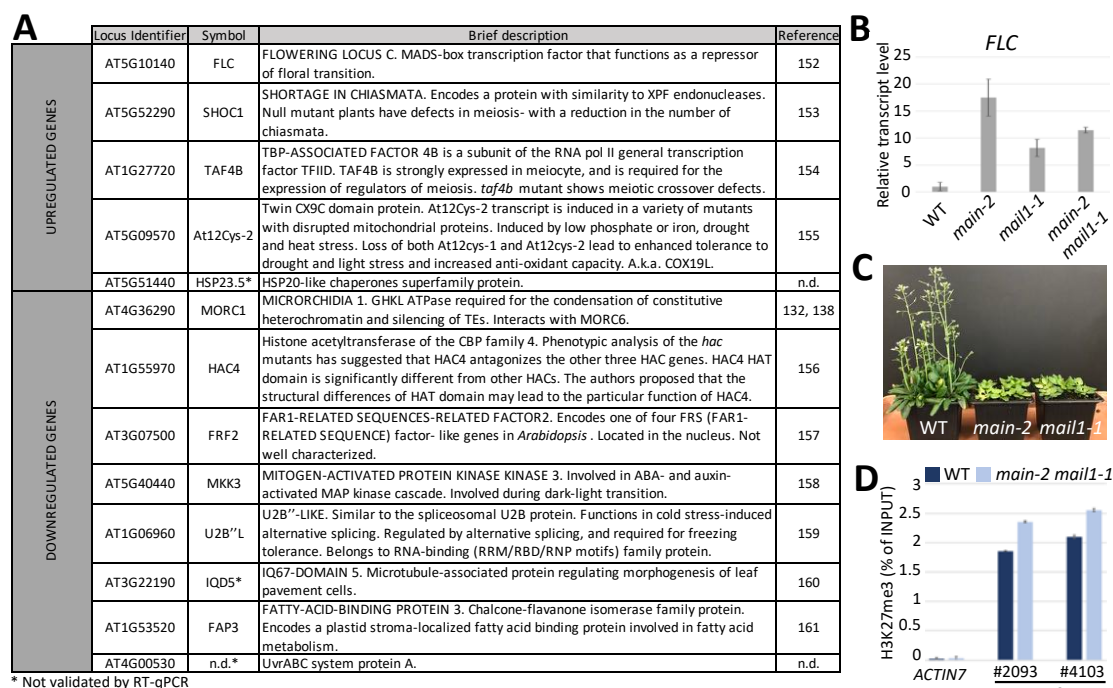


Figure 13. A. List of outstanding misregulated genes in the *main-2* and *mail1-1* mutants, including *FLC* and *MORC1*. These gene are also misregulated in *pp7l-2* and *mail1-1 pp7l-2*. B. Relative *FLC* transcript

level in *main-2*, *mail1-1* and *main-2 mail1-1* mutants over WT assayed by RT-qPCR and normalized to the housekeeping *RHIP1* gene. C. Growth and flowering delay of *main-2* and *mail1-1* mutants in comparison to WT plants. The *main-2 mail1-1* mutant shows same phenotype D. H3K27me3 levels at two *FLC* locations in *main-2 mail1-1* and WT plants, as defined in reference¹⁶². *ACTIN7* is used as negative control. Values are mean $\pm$ SEM from two independent samples, with data relative to INPUT.

3.6 Mode of action of MAIN/MAIL1/PP7L complex in the regulation of gene expression and TE silencing.

To decipher the molecular mechanisms involving the MAIN/MAIL1/PP7L protein complex in the regulation of gene expression and TE silencing, we are combining genetic, genomic, microscopic and biochemical approaches. We believe that altogether these approaches will allow us to address the following questions.

3.6.1 Are PMD and PP7L proteins physically interacting with chromatin?

Previous reports suggested that MAIN and MAIL1 could be putative TF-like proteins or nuclear factors involved in transcriptional processes^{123, 149}. Besides, we have found that a significant proportion of loci that are commonly misregulated in the *pmd* and *pp7l-2* single, and *mail1-1 pp7l-2* double mutants are downregulated genes carrying a specific DNA motif (called the 'DOWN' motif) within their promoters (Nicolau et al., PLOS Genetics, in revision). Thus, one hypothesis is that the MAIN/MAIL1/PP7L complex directly interacts with chromatin possibly through the recognition of the 'DOWN' motif to promote transcription. A second hypothesis is that the MAIN/MAIL1/PP7L complex is required for the activation of an unidentified TF, which would recognize the 'DOWN' motif to initiate transcription. In addition, we have identified genes and TEs that are upregulated in *pmd* and *pp7l-2* single, and in *mail1-1 pp7l-2* double mutants, which suggests that the MAIN/MAIL1/PP7L complex could act as transcriptional repressor at these genomic locations.

To test all these hypotheses, chromatin-immunoprecipitation coupled to DNA sequencing (ChIP-seq) experiments using FLAG- or MYC-tagged MAIN, MAIL1 and PP7L expressing lines will be carried out. ChIP-qPCR analyses will be also specifically performed at the genomic regions corresponding to misregulated loci. Particularly, these analyzes will allow us to determine if the MAIN/MAIL1/PP7L complex is enriched in the promoter regions carrying the 'DOWN' motif, such as the *MORC1* promoter. In parallel, we are planning to produce MAIN, MAIL1 and PP7L proteins using bacterial system and TnT-coupled wheat germ extract system to test if the proteins interact with DNA and/or chromatin. We will then test unbiasedly if the PMD and PP7L proteins bind a specific DNA motif, using complementary approaches based on systematic evolution of ligands by exponential enrichment

(SELEX) and DNA affinity purification sequencing (DAP-seq) experiments^{163, 164}. Electrophoretic mobility shift assay (EMSA) will ultimately validate these results.

3.6.2 What is the role of PP7L in the MAIN/MAIL1 pathway?

We and others have found that MAIN, MAIL1 and PP7L form a protein complex (Nicolau et al., PLOS Genetics, in revision) and ¹⁶⁵. We will perform gel filtration experiments to determine the molecular weight of this complex and its stoichiometric composition. We will also introduce epitope-tagged PMDs into the *pp7l-2* mutant to test if PP7L is required for MAIN/MAIL1 interaction and subnuclear localization using co-immunoprecipitation (co-IP) and immunofluorescence (IF) experiments, respectively. Site directed mutagenesis will generate epitope-tagged mutant versions of PP7L and PMDs (including MAIN C230Y) to test by co-IP if mutant versions can still interact together. This will allow to decipher the amino acids required for the interactions. Beside the PMDs, it is unknown whether PP7L interacts with other proteins. We will perform immunoprecipitation followed by mass spectrometry (IP-MS) analyses to identify PP7L putative interactors, and validate these results by co-IP experiments. In animals, several reports have shown that PPPs are involved in chromatin-related processes, by interacting with histones or DNA, or regulating TF activity¹⁶⁶. In plants, only a few studies have made such connections. For instance, a PPP protein belonging to the PP2A subfamily was involved in the dephosphorylation of histone H3 at serine 10¹⁶⁷. More recently, a study showed that PP4R3A, which is a regulatory subunit of the PP4-like PPP protein complex, interacts with chromatin to recruit RNA polymerase II and promote transcription of microRNA-encoding genes¹⁶⁸.

PP7L is a presumably inactive phosphatase, lacking essential amino acids to hydrolyze phosphate^{169, 170}, but potentially capable of fixing it. In spite of this, it is possible that, in complex with the PMDs, PP7L enzymatic activity is compensated. Using in vitro approaches, we will test if PP7L alone or in complex with PMDs binds phosphate, and displays phosphatase activity. Depending on the result, we will consider phosphoproteomic analyses. If PP7L can fix phosphate, we will test if it can bind phosphorylated histones at serine/threonine residues using histone peptide array. Indeed, beside a role in chromosome condensation and cell division, histone phosphorylation/dephosphorylation has been involved in transcription regulation¹⁶⁶.

3.6.3 Are MAIN, MAIL1 and PP7L transcriptional activators or repressors?

The ChIP-seq, DAP-seq and SELEX experiments described above should give new insights into the mode of action of MAIN/MAIL1/PP7L complex. Alternatively, to test if MAIN/MAIL1/PP7L positively regulate transcription, we will perform yeast one hybrid (Y1H) experiment using MAIN/MAIL1/PP7L fused to the Gal4 binding domain (Gal4BD) to test if together, they can activate the reporter gene¹⁷¹.

In *A. thaliana*, the floral repressor *FLOWERING WAGENINGEN* (*FWA*) is an imprinted gene, whose promoter carries SINE retrotransposon and related tandem repeats that are targeted by DNA and H3K9 methylation^{172, 173}. *FWA* silencing is required for flowering, hence the late flowering phenotype of plants overexpressing a hypomethylated *fwa* epiallele¹⁷². The use of zinc finger 108 (ZF108) has been a powerful way to target any protein of interest to *FWA* locus (either in WT or *fwa* epiallele mutant) to test for *FWA* activation or repression^{40, 41, 144, 174}. While targeting a TF-like proteins at *FWA* promotes its transcription and late flowering phenotype in WT background, targeting a transcriptional repressor rescue the late flowering phenotype of *fwa* plants by silencing *FWA*. To test if the MAIN/MAIL1/PP7L enhance or repress transcription, we will transform WT and *fwa* plants with constructs expressing MAIN/MAIL1/PP7L domains fused to ZF108. We will then analyze *FWA* expression level by RT-qPCR and score the flowering time of transgenic plants. In addition to *FWA*, ZF108 binds to other genomic locations, called "off target" regions¹⁴⁴. We will take the advantage of these "off target" regions to determine by RNA-seq analyses the genome-wide effects of MAIN/MAIL1/PP7L domains fused to ZF108.

Developing CRISPR-based approaches is an attractive alternative to address this fundamental question¹⁷⁵. Using the nuclease-deficient form of CRISPR-associated 9 (dCAS9) fused to the MAIN/MAIL1/PP7L domains in combination with specific guide RNAs (gRNAs) may allow to target them to specific genomic locations, and test if these protein domains act as transcriptional activator or repressor. In conclusion, these analyses will precise the role of MAIN, MAIL1 and PP7L on transcriptional regulation.

3.6.4 What is the structure of PMD?

The PMD is a plant-specific domain of unknown function whose three-dimensional (3D) structure remains elusive. Previous work proposed that PMD shared similarities with aminotransferase domain (InterPro IPR019557)¹⁴⁹, and I-TASSER protein structure prediction suggests that PMD is the most similar to proteins belonging to the "transferase" family such as the yeast fatty acid synthase (<https://www.rcsb.org/structure/2pff>)¹⁷⁶. Nevertheless, correlations between protein structure and molecular function of PMD remain difficult. As mentioned above, PP7L is presumably inactive, and has never been crystalized.

We have initiated a collaboration with the Jiamu Du laboratory at SUSTech China to crystalize the PMD and PP7L. Positioning the different conserved and mutated amino acids in the crystals will be essential to better understand the mode of action of these proteins. In particular, defining the position of Cys230, which is the amino acid that is mutated in the hypomorphic *main-3* mutant, would allow us to better understand the mode of action of PMD in chromatin-related processes. Finally, combining

the PMD with an oligonucleotide carrying the 'DOWN' motif could possibly stabilize the PMD, allowing its crystallization while in contact with DNA.

3.7 Investigating the role of other PMD-B and PMD-C proteins in *A. thaliana* plant development.

3.7.1 Preliminary results.

The *A. thaliana* genome encodes thirteen PMD proteins belonging to the PMD-B and PMD-C clades (Figure 14A)¹²⁴. RT-qPCR analyses confirmed previous results showing that the PMD-C *MAIN* and *MAIL1*, together with their close homologs *MAIL2* and *MAIL3* are expressed throughout the plant development (Figure 14B)¹²². *MAIL2* is a standalone PMD protein that seems to be essential for plant development, hence the dramatic phenotype of *mail2-1* mutant seedlings, failing to develop true leaves after germination. Conversely, the *mail3-2* mutant displays a WT phenotype, meaning that *MAIL3* is not as essential as *MAIN*, *MAIL1* or *MAIL2* for plant development in normal laboratory conditions. Also known as *longPP7*¹⁷⁰, *MAIL3* encodes the unique *A. thaliana* protein in which its PMD is fused to a PPP domain (Figure 11A, and Nicolau et al., PLOS Genetics, in revision). As mentioned previously, the *MAIL3* PPP domain is closely related to PP7L (Nicolau et al., PLOS Genetics, in revision)¹⁷⁰. To get more insights into the role of *MAIL2* and *MAIL3* in the regulation of gene expression, we performed RNA-seq experiments in *mail2-1* and *mail3-2* single mutants in comparison to their respective WT controls. Although these analyses are very preliminary and require validations, we found that most of the loci that were misregulated in these two *pmd* mutants were not affected in *main* and *mail1* mutants (*main-2/mail1-1/main-3* overlap, Figure 14C). Furthermore, these analyses revealed very different sets of misregulated genes, and different GO term enrichments between the *mail2-1* and *mail3-2* mutants (Figure 14D). Thus, *MAIL2* and *MAIL3* might be involved in different pathways regulating the expression of specific networks of genes.

Using available RNA-seq data of different WT *A. thaliana* tissues and growing conditions, we analyzed the expression patterns of *PMD-B* genes, and found tissue-specific expression patterns (Figure 14E). They seem to be all expressed in carpel, except *PMD8* and *PM10* which, instead, show a strong expression in pollen. In addition, *PMD11* and *PMD13* are expressed in roots, and overall, more than others (Figure 14E). Preliminary RT-qPCR analyses suggested that most of the *PMD-B* were expressed in flower and silique (Figure 14F).

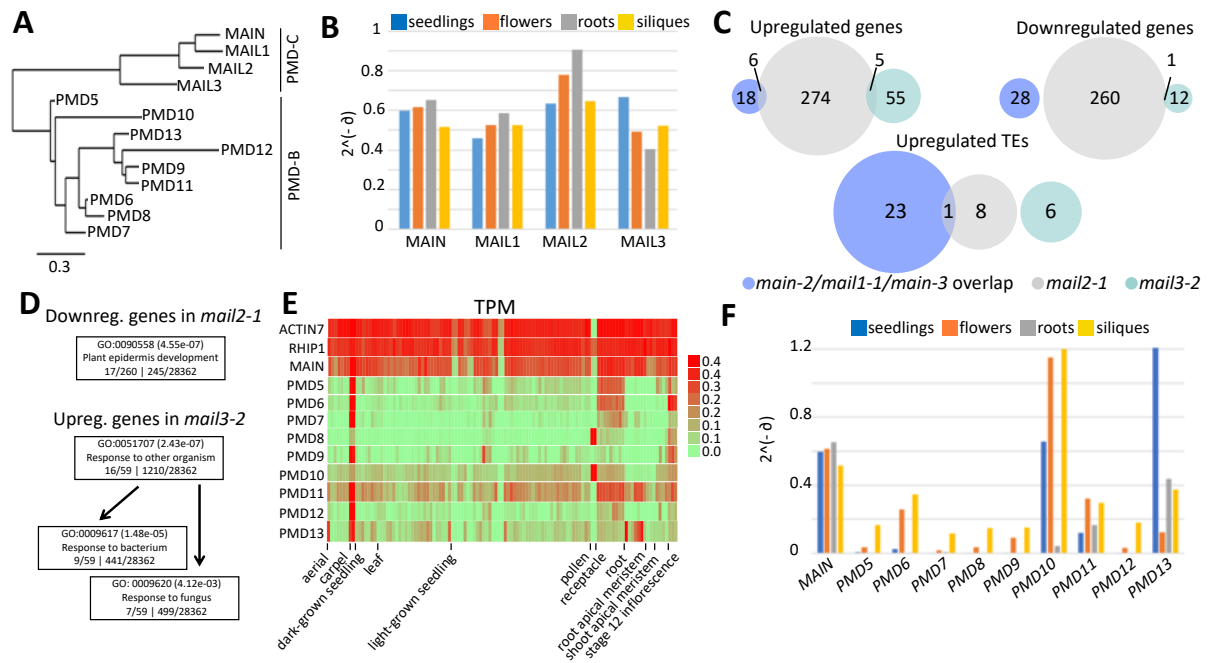


Figure 14. Preliminary studies of MAIL2, MAIL3 and PMD-B clade PMDs. **A.** Phylogenetic tree of *A. thaliana* PMDs **B.** RT-qPCR assaying MAIN and MAIL expression level in four different organs normalized to RHIP1 **C.** Venn diagrams analyses of misregulated loci in *main-2/mail1-1/main-3* overlaps (described in Figure 12B), *mail2-1* and *mail3-2*. **D.** GO term enrichment of down- and upregulated loci in *mail2-1* and *mail3-2*, respectively. **E.** Heat maps compiling RNA-seq data from different WT organs and growing conditions. TPM = Transcripts per Million. Color keys are log2 scale of TPM. MAIN and the two housekeeping genes ACTIN7 and RHIP1 are used for comparison. Source Araport ThaleMine. **F.** Same analyses as in B for PMD5 to PMD13, MAIN is shown as reference.

3.7.2 What are the roles of MAIL2 and MAIL3?

Preliminary RNA-seq and GO terms analyses of misregulated loci in *mail2-1* and *mail3-2* suggested that MAIL2 is required for the proper expression of genes essential for plant development, while MAIL3 may be involved in the repression of biotic stress-induced genes (Figure 14D). Importantly, they might regulate very different networks of genes in comparison to MAIN and MAIL1. Thus, these preliminary results suggest that MAIL2 and MAIL3 might be involved in distinct cellular pathways. Because *mail2-1* null mutant shows dramatic developmental phenotype, we will engineer a CRISPR-induced hypomorphic mutant, based on amino acid conservation in the PMD-C proteins. Phenotypic and molecular characterization of the obtained *mail2* hypomorphic mutant will allow us to better characterize the role of MAIL2 in plant development.

To determine the role of MAIL2 and MAIL3, we are using approaches similar to those described for MAIN and MAIL1. We are currently generating complementing lines expressing epitope-tagged genomic (g)MAIL2/MAIL3 constructs to carry out IP-MS to identify putative protein interactors. These

lines will be also used to perform ChIP experiments. MAIL3 PPP domain is presumably functional^{169, 170}. We will confirm its activity using in vitro phosphatase assay, and engineer a MAIL3 point mutant lacking the catalytic activity of the phosphatase domain to stabilize transient protein complexes and identify the substrates. As PP7L and MAIL3 PPP are phylogenetically related, we will test their genetic interactions by creating a *pp7l mail3* double mutant. Recently, it has been shown that TE silencing defects in *pp7l-3* can be partially complemented at four TE locations in *pp7l-3 mail3-2* double mutant¹⁶⁵. Further work will be needed to decipher the genetic interaction between PP7L and MAIL3 at the whole genome level, and to better understand the interplay between MAIL3 and the MAIN/MAIL1/PP7L complex.

Finally, as upregulated genes in *mail3-2* are enriched in the GO terms “response to bacterium” and “response to fungus” (Figure 14D), we have initiated collaborations with laboratories studying plant-microorganism interactions. Lionel Navarro laboratory at ENS Paris is currently testing if the *mail3-2* mutant is more resistant to *Pseudomonas syringae* infection. In parallel, Harald Keller at INRA Institut Sophia Agrobiotech is testing *mail3-2* resistance upon infection with the oomycete *H. arabidopsidis* (Hpa).

3.7.3 What are the roles of PMD-B proteins?

As a long-term project, we want to explore the role of PMD-B proteins. *A. thaliana* carries nine PMD-B genes. Five of them (i.e. *PMD6*, *PMD9*, *PMD11*, *PMD12* and *PMD13*) cluster in a ~30kb region on the long arm of chromosome 1. *PMD5* is located ~285kb downstream of this aforementioned PMD-B cluster. *PMD10*, *PMD7* and *PMD8* are located on the short arm of chromosome 1, the long arm of chromosome 4 and the short arm of chromosome 5, respectively (Figure 15).

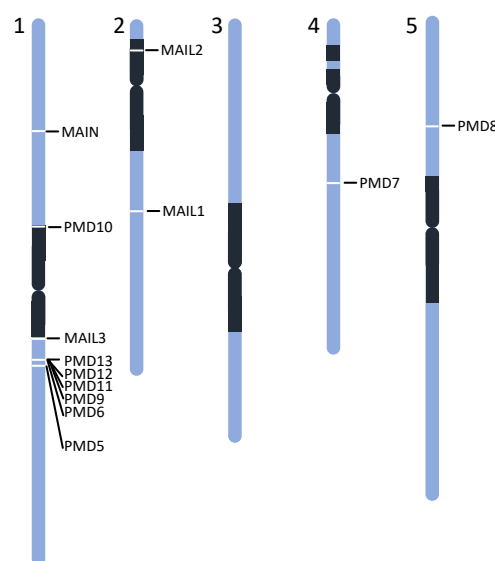


Figure 15. Chromosomal distribution of PMD genes within the *A. thaliana* genome. Dark boxes represent pericentromeric constitutive heterochromatin.

Preliminary results suggest that *PMD-B* genes show very distinctive expression pattern (Figure 14A, E and F). All of them except *PMD8* and *PMD10*, seem to be expressed in carpel (Figure 14E). By contrast, *PMD8* and *PMD10* might be expressed in pollen (Figure 14E). To confirm these observation, RT-qPCR analyses using WT pollen and carpel material will be carried out. In parallel, we will clone the β -glucuronidase (GUS) reporter gene under the control of *PMD-B* promoters, and transform WT plant to determine the tissue-specific expression *PMD-B* candidates. We will also possibly determine their subcellular (and possibly nuclear) localization by IF experiments coupled to confocal microscopy analyses.

In parallel, we are developing reverse genetic approach to generate higher order mutants, and screen for abnormal developmental phenotypes. Depending on the results, we will further characterize these mutants using appropriate experimental approaches (RNA-seq, IP-MS, co-IP, ChIP-seq, BS-seq...) In conclusion, we believe that this study will give insights into the role of *PMD-B* proteins during *A. thaliana* development. Considering that some *PMD-B* seem to be strongly expressed in pollen or carpel, it is tempting to speculate that they may play a regulatory function during flower development and reproduction.

3.8 The role of PMD proteins in *S. lycopersicum* (tomato) plant and fruit development.

3.8.1 Scientific context

As a climacteric fruit, tomato is characterized by a sharp increase in respiration marked by elevated CO₂ production, the so-called “climacteric crisis”, associated with a rise in autocatalytic production of the plant hormone ethylene. This volatile compound is responsible for the initiation and coordination of ripening process. TFs play important roles in fruit ripening as exemplified by the identification of ripening mutants such as ripening-inhibitor (*rin*), nonripening (*nor*) and Colorless nonripening (*Cnr*), in which *TF* genes are mutated^{177, 178}. The RIN, NOR and CNR TFs control fruit ripening by acting upstream of ethylene, and regulating the expression of ethylene biosynthesis/perception genes¹⁷⁹. DNA methylation and histone modifications have also been connected to ripening control. DEMETER-like DNA demethylase2 (SIDML2) is specifically expressed in fruit to demethylate promoters of genes, whose expression is required for this process¹⁸⁰ (for review^{181, 182}). Several genes related to ripening such *ACS2* or *RIN*, and *TDR4/FUL1* are associated with hyper-H3K27me3 marks in the *nor* and *cnr* mutants. This implies that the PcG pathway is disturbed in these mutants, and plays important role in fruit ripening¹⁷⁸, likely by fine-tuning gene expression. Accordingly, mutations in the PcG genes, such as *SIEZ1*, *SIEZ2*, *SIMS1* induce fruit ripening defects¹⁸¹. In conclusion, ethylene, specific TFs and epigenetic pathways orchestrate the proper expression of genes that are essential for tomato fruit ripening.

We have initiated a collaboration with Julien Pirrello at GBF INRA/INP-ENSAT Toulouse to investigate the role of PMD proteins in tomato. Considering that MAIN and MAILs are involved in regulation of gene expression and TE silencing, we want to decipher the role of their tomato counterparts (SIPMDs) during tomato plant development and fruit ripening. In addition, we want to determine if SIPMDs interact with the tomato homologs of *A. thaliana* PP7L/MAIL3 PPPs (SIPP7Ls), and if these latter are required for fruit development. Notably, SIPP2C1, a group A type 2C protein phosphatase, was recently involved in fruit ripening in tomato¹⁸³.

3.8.2 What is the expression pattern of *SIPMDs* at the tissue-specific level?

The tomato genome encodes seventeen *MAIN/MAIL* homologs (genic *SIPMDs*) of unknown function, and some of them show very distinctive expression patterns. In particular, *SIPMD2*, which is the *MAIN/MAIL1* closest homolog, is constitutively expressed throughout the plant development, with expression peaks in meristem and flower (Figure 16)¹⁸⁴. In addition, *SIPMD11* and *SIPMD14* are specifically expressed at major transition phases of plant development, such as anthesis and fruit ripening (Figure 16)¹⁸⁴. More precisely, spatiotemporal transcriptomic analyses of various fruit tissues revealed that the two genes are strongly expressed in placenta and locular tissue, where seeds are located¹⁸⁵. This implies that *SIPMD11* and *SIPMD14* might be regulated in a tissue-specific manner. To confirm *SIPMD2*, *SIPMD11* and *SIPMD14* expression patterns, we are planning to perform RT-qPCR analyses in different WT plant and fruit tissues. We will also transform WT tomato plants with constructs expressing GUS reporter gene under the control of each *SIPMD* native promoter.

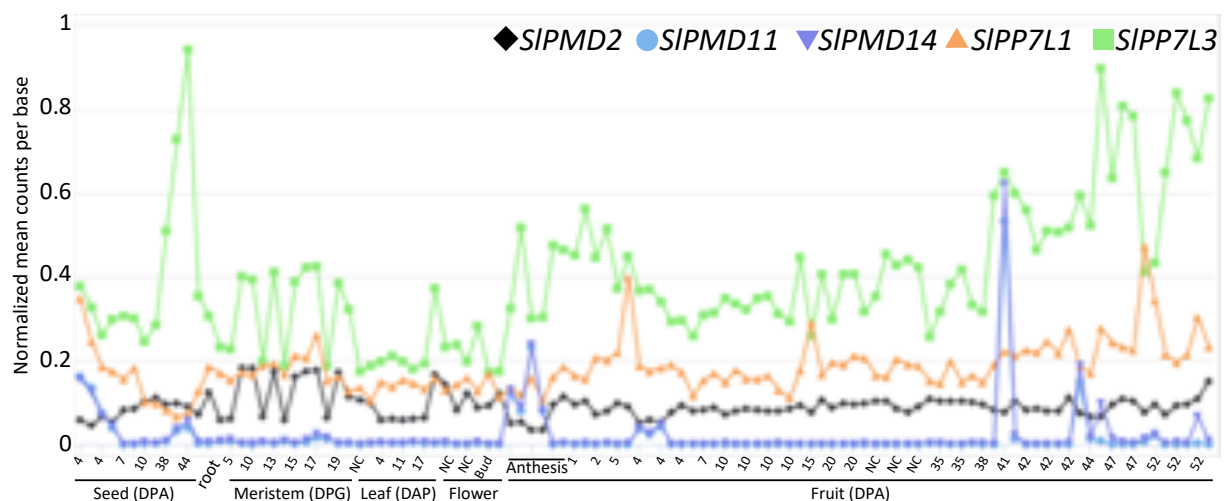


Figure 16. Expression pattern of three *SIPMDs* and two *SIPP7Ls* at different stages of tomato plant development. Fruit ripening starts at 42 DPA. DPA (day post anthesis), DPG (day post germination), DAP (day after planting). Source TomExpress¹⁸⁴.

3.8.3 Are SIPMDs required for tomato plant development and ripening process?

To tackle this fundamental question, the GBF lab is currently generating tomato plants misexpressing *SIPMD2*, *SIPMD11* and *SIPMD14* genes using CRISPR/Cas9 strategy. To overcome any potential redundancy, simultaneous *slpmd2/11/14* knockouts have been generated using the GoldenBraid technology combined to the CRISP-P tool to design efficient RNA guides^{186, 187}. Transgenic plants will be subsequently characterized to confirm *SIPMD* misexpression, and scored to determine any abnormal developmental phenotypes and/or fruit ripening defects. When relevant, RNA-seq analyses will be performed at different stages of plant development, and/or during the fruit formation and ripening to determine any transcriptional defects in the *slpmd* mutants. Depending on the results, we will consider epigenomic analyses of *slpmd* mutants.

3.8.4 Are SIPP7L and SIPMD interacting together?

We have identified two SIPP7L genes, called SIPP7L1 and SIPP7L3, that are close homologs of *A. thaliana* PP7L and MAIL3, and well expressed throughout the tomato plant and fruit development (Figure 16). We will first perform yeast two hybrid (Y2H) experiments to determine whether SIPMDs and SIPP7Ls interact together. Co-IP experiments using *N. benthamiana* and *A. thaliana* plants co-expressing the two putative interactors will be subsequently carried out to confirm the results. The validation of any SIPMD/SIPP7L interaction will initiate further work to generate *slpp7l* mutants and study their potential role in the SIPMD pathway.

3.9 Evolutionary aspects of PMD in relation with TEs and other protein domains.

3.9.1 Scientific context.

As previously mentioned, phylogenetic analyses of PMDs defined three clades¹²⁴. The PMD-A clade is widely spread among the angiosperms, and exclusively encoded by *Gypsy* TEs, while PMD-B and PMD-C clades are mostly genic versions that originated and diversified upon the domestication of a PMD-A variant by the plants¹²⁴. Among the PMD-C clade, which includes the *MAIN* and *MAIL* genes, some PMDs are associated to *MuDR* TEs, like in the monocot *Oryza sativa* (rice) and in *Amborella trichopoda*, the sister species of all angiosperms¹²⁴. It is believed that *MuDR* TEs acquired the PMD through transduplication events that occur during the angiosperms evolution¹²⁴. Thus, while the majority of angiosperm species, such as *S. lycopersicum*, carry mostly *Gypsy*-PMD-A and genic PMD-B/C versions, some other species, like *O. sativa*, also harbor numerous *MuDR*-PMD-C variants. Finally, *A. thaliana* is deprived of TE-associated PMD, and only genic PMD versions prevail¹²⁴.

In this part of the project, we want to investigate the relationship between PMD and TEs, and co-evolution features of PMD with other protein domains.

3.9.2 Can PMD be beneficial for TEs?

The widespread occurrence of PMD-TE association in angiosperms plants over a long evolutionary period, rises the questions of its potential benefits to TEs mobility and amplification. Here, we aim at determining whether PMD has been beneficial for TEs by increasing their ability to transpose and survive within host genomes.

In a preliminary study, we checked the presence of PMD domain in the entire genome of 331 angiosperm species. This analysis shows that PMDs copy number vary tremendously among species ranging from dozens to thousands of copies. For example, among the Brassicaceae family, we found that most of the species, including *A. thaliana*, carried 10 to 20 genic PMDs, except tetra-allopolyploid species such as *C. sativa* that carried twice as much because of whole genome duplication (WGD) (Figure 17A and B). However, in *Solanum* genus (e.g. tomato, potato, etc.), the number of PMDs can reach ~500 copies. A more in-depth investigation of *S. lycopersicum* genome reveals that in contrast to *A. thaliana*, the 427 identified PMDs were mostly Gypsy-PMDs (Figure 17A and B). Strikingly, we observed a huge variation in the PMD number across the closely related *Oryza* species: from 186 copies in *O. officinalis* to 668 copies in *O. sativa*, and up to 2163 copies in *O. punctata* (Figure 17A). Because *O. punctata* and *O. sativa* are diploid organisms with a genome size similar to other *Oryza* species¹⁸⁸, the difference in PMD number cannot be explained by WGD as for *Brassicaceae*. Instead, we hypothesized that it was due to the successful amplification of TEs carrying a PMD. Indeed, further investigations of these two species revealed that increase in PMD copy number are due to genomic amplification of mainly two different TE families carrying a PMD domain: The Gypsy *Osr30* family in *O. punctata* and the MuDR *Os16* family in *O. sativa* (Figure 17B). Remarkably, *Osr30* is the TE family with highest number of paralogs not only in *O. punctata* but in all the 11 sequenced *Oryza* species¹⁸⁹. Similarly, *Os16* is the most repeated MuDR TE family in *O. sativa*. The estimation of *Osr30* insertion age in *O. punctata* shows a recent TE burst (~1 Mya) with neo-insertions that occurred in the very recent past (Figure 17C), suggesting the presence of potentially functional and active *Osr30* copies. Similarly, the all-by-all comparison of hundreds of MuDR *Os16* paralogs revealed high sequence similarity (>98 %), indicating a recent burst in *O. sativa*. Moreover, nanopore RNA-seq analyses suggested that *Os16* was expressed in WT *O. sativa* (MANGO team, unpublished data).

In summary, considering that *i)* the unrelated TEs *Osr30* (Class I TE) and *Os16* (Class II TE) carry a PMD, *ii)* have undergone massive and recent burst *iii)* and appear to be expressed (at least *Os16*), it seems rational to assume that, to some extent, their PMD has contributed to their evolutionary success.

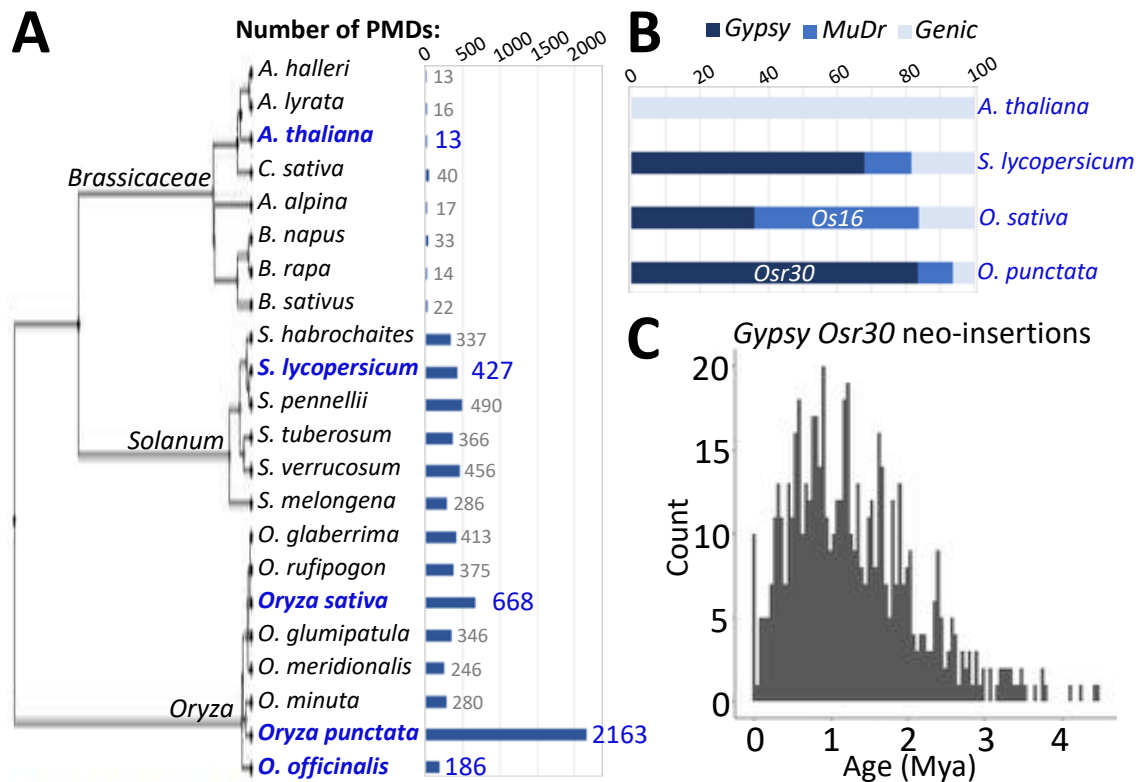


Figure 17. Evolution and distribution of PMDs among angiosperms. A. Total number of PMDs in several Brassicaceae, Solanum and Oryza species according to their phylogenetic distribution. B. Relative percentage of Gypsy-, MuDr-, and genic PMDs in four species. TEs of *Os16* and *Osr30* families carry PMD, and these families are the most repeated MuDr and Gypsy TE families in *O. sativa* and *O. punctata*, respectively. C. Counts of Gypsy *Osr30* neo-insertions in the recent history of *O. punctata*. Mya, millions of years.

To test this appealing hypothesis, we need first to confirm by RT-qPCR analyses that *Osr30* and *Os16* are active in the two *Oryza* species. Then, nanopore RNA-seq experiments will be carried out to identify full length RNA copies of *Osr30*, and mobilome experiments will identify extrachromosomal circular DNA (eccDNA) forms of active TEs in *O. punctata*¹⁹⁰. We will also transgenerationally score by qPCR the respective TE copy numbers in the genomes of the two *Oryza* species to determine if TEs are mobile. Thus, these experiments will confirm that *Osr30* and *Os16* are active and functionally capable of transposition. In addition, we will use WT *A. thaliana* as 'naïve' plant because its genome is deprived of TE-associated PMDs. First, WT plants will be transformed with *Osr30* and *Os16* TEs carrying a WT- or mutated (Δ)-PMD versions to test if PMDs contribute to TE fitness. Second, we will transform WT *A. thaliana* with standalone versions of *Osr30*- and *Os16*-PMDs to test, by RNA-seq analyses, if they impact gene expression of endogenous genes, possibly by outcompeting genic PMDs, such as MAIN or MAILs. In parallel, we will test if *main-2* plants overexpressing standalone *Osr30*- and *Os16*-PMDs

show a complementation of the mutant phenotype. Finally, we could use the ZF108 or dCas9/CRISPR system to transform WT *A. thaliana* plants with *Osr30*- and *Os16*-PMD fusion constructs and test if they can enhance or repress transcription.

3.9.3 Co-evolution of PMD and other protein domains.

Although most of genic PMDs are standalone versions, the PMD is often associated with other protein domains such as for instance PPP. Because the PPP protein PP7L interacts with MAIN and MAIL1, and MAIL3 is a close homolog of both MAIN/MAIL1 and PP7L through its PMD and PPP domains, respectively, we have decided to determine the distribution of PMD and PPP domains, and to retrace their evolutionary history among angiosperm species. This study, described in Nicolau et al., PLOS Genetics, in revision, suggests that the two protein domains have co-evolved to constitute a functional PMD/PP7 module. Importantly, this protein module could result from interactions *in trans*, as shown with MAIN, MAIL1 and PP7L, or *in cis* in proteins like MAIL3 (Figure 18).

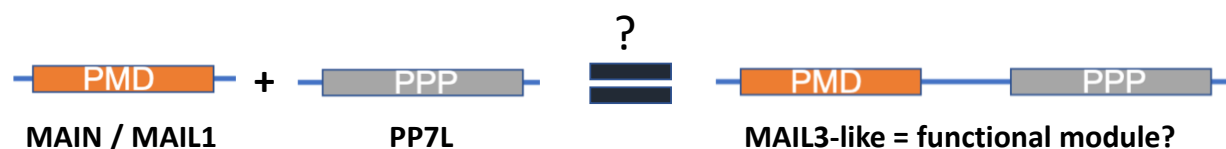


Figure 18. Hypothetical model in which the PMD and PPP domains constitute a functional protein module. Interaction *in trans* of MAIN/MAIL1 PMDs together with PP7L, or *in cis* between PMD and PPP domains of MAIL3 may form a functional protein module.

In addition to PPP, PMD can be associated to various protein domains, such as among others, kinase, protease, high mobility group (HMG), SWIM-type Znf or basic helix-loop-helix (bHLH) DBD domains (www.ebi.ac.uk/interpro/beta/entry/InterPro/IPR019557/domain_architecture/). In the future, it will be important to determine the biological function of proteins carrying a PMD in association with these DBDs. Altogether, these experiments will pave the way for a better understanding of the role of PMD in chromatin-related processes.

4. Future studies and perspectives: beyond the PMD, role of ETE proteins in regulation of gene expression.

Beside PMDs, MUGs or other TE-derived proteins, the *A. thaliana* genome carries many additional putative ETE genes encoding proteins of unknown function, and displaying the typical features of *bona fide* gene (well expressed, absence of siRNA, microsynteny between divergent genomes...) ¹⁹¹. Among those, are genes encoding transposases, Ribonuclease H (RNase H) or capsid-like gag proteins. In

conjunction with the study of PMD proteins, we will study the role of these ETEs during plant development, and in response to various stresses. Several studies have shown that TEs are highly responsive to abiotic and biotic stresses (such as salt, heat, cold, bacteria or viruses...) ¹⁰⁷. Considering that ETE proteins are derived from TEs, it is likely that they also respond to specific environmental stimuli. This assumption is supported by a recent study suggesting that ETE proteins play important roles in response to abiotic stresses ¹⁹². The potential role of PMDs in response to abiotic and/or biotic stresses will be also investigated.

In conclusion, our future line of research will be divided into two main parts. First, we will continue the *ATCOPIA28::GFP* forward genetic screen aiming at identifying new epigenetic factors. Second, we will study the role of ETE proteins during plant development and in response to environmental changes. We have already started this study through the characterization of PMD proteins, showing that the ETE proteins MAIN and MAIL1 are important factors required for TE silencing and regulation of gene expression. Further studies will give more insights into the role of ETE proteins in response to environmental changes, and their contribution to plant fitness.

VI. References.

1. Heitz E. Das heterochromatin der moose. *Jahrb Wiss Botanik* 1928; 69: 762-818.
2. Slotkin RK and Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 2007; 8: 272-285. DOI: 10.1038/nrg2072.
3. Grewal SI and Jia S. Heterochromatin revisited. *Nat Rev Genet* 2007; 8: 35-46.
4. Picart-Piccolo A, Picault N and Pontvianne F. Ribosomal RNA genes shape chromatin domains associating with the nucleolus. *Nucleus* 2019; 10: 67-72. 2019/03/15. DOI: 10.1080/19491034.2019.1591106.
5. Kharchenko PV, Alekseyenko AA, Schwartz YB, et al. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 2011; 471: 480-485. DOI: 10.1038/nature09725.
6. Sequeira-Mendes J, Araguez I, Peiro R, et al. The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. *Plant Cell* 2014; 26: 2351-2366. DOI: 10.1105/tpc.114.124578.
7. Waddington CH. The epigenotype. 1942. *Int J Epidemiol* 2012; 41: 10-13. 2011/12/22. DOI: 10.1093/ije/dyr184.
8. Goldberg AD, Allis CD and Bernstein E. Epigenetics: a landscape takes shape. *Cell* 2007; 128: 635-638. 2007/02/27. DOI: 10.1016/j.cell.2007.02.006.
9. Nanney DL. Epigenetic Control Systems. *Proc Natl Acad Sci U S A* 1958; 44: 712-717. 1958/07/15. DOI: 10.1073/pnas.44.7.712.
10. Holliday R. The inheritance of epigenetic defects. *Science* 1987; 238: 163-170. 1987/10/09.
11. Holliday R. Epigenetics: An Overview. *Dev Genet* 1994; 15: 453-457.
12. Cavalli G and Heard E. Advances in epigenetics link genetics to the environment and disease. *Nature* 2019; 571: 489-499. 2019/07/26. DOI: 10.1038/s41586-019-1411-0.
13. Deans C and Maggert KA. What do you mean, "epigenetic"? *Genetics* 2015; 199: 887-896. 2015/04/10. DOI: 10.1534/genetics.114.173492.

14. Luger K, Mader AW, Richmond RK, et al. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 1997; 389: 251-260.
15. Fransz P and de Jong H. From nucleosome to chromosome: a dynamic organization of genetic information. *Plant J* 2011; 66: 4-17. Review 2011/03/30. DOI: 10.1111/j.1365-3113X.2011.04526.x.
16. Song F, Chen P, Sun D, et al. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* 2014; 344: 376-380. 2014/04/26. DOI: 10.1126/science.1251413.
17. Kouzarides T. Chromatin modifications and their function. *Cell* 2007; 128: 693-705. DOI: 10.1016/j.cell.2007.02.005.
18. Bannister AJ and Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res* 2011; 21: 381-395. Review 2011/02/16. DOI: 10.1038/cr.2011.22.
19. Strahl BD and Allis CD. The language of covalent histone modifications. *Nature* 2000; 403: 41-45.
20. Li B, Carey M and Workman JL. The role of chromatin during transcription. *Cell* 2007; 128: 707-719.
21. Probst AV, Fagard M, Proux F, et al. Arabidopsis histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. *Plant Cell* 2004; 16: 1021-1034.
22. Aufsatz W, Mette MF, Van Der Winden J, et al. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *Embo J* 2002; 21: 6832-6841.
23. Hristova E, Fal K, Klemme L, et al. HISTONE DEACETYLASE6 Controls Gene Expression Patterning and DNA Methylation-Independent Euchromatic Silencing. *Plant Physiol* 2015; 168: 1298-1308. 2015/04/29. DOI: 10.1104/pp.15.00177.
24. Chandrasekharan MB, Huang F and Sun ZW. Histone H2B ubiquitination and beyond: Regulation of nucleosome stability, chromatin dynamics and the trans-histone H3 methylation. *Epigenetics* 2010; 5: 460-468. 2010/06/05. DOI: 10.4161/epi.5.6.12314.
25. Sridhar VV, Kapoor A, Zhang K, et al. Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. *Nature* 2007; 447: 735-738.
26. Jahan S and Davie JR. Protein arginine methyltransferases (PRMTs): role in chromatin organization. *Adv Biol Regul* 2015; 57: 173-184. 2014/09/30. DOI: 10.1016/j.jbior.2014.09.003.
27. Zhang J, Jing L, Li M, et al. Regulation of histone arginine methylation/demethylation by methylase and demethylase (Review). *Mol Med Rep* 2019; 19: 3963-3971. 2019/04/04. DOI: 10.3892/mmr.2019.10111.
28. Elgin SC and Reuter G. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb Perspect Biol* 2013; 5: a017780. DOI: 10.1101/cshperspect.a017780.
29. Carlson SM and Gozani O. Nonhistone Lysine Methylation in the Regulation of Cancer Pathways. *Cold Spring Harb Perspect Med* 2016; 6 2016/11/03. DOI: 10.1101/cshperspect.a026435.
30. Mozgova I and Hennig L. The polycomb group protein regulatory network. *Annu Rev Plant Biol* 2015; 66: 269-296. DOI: 10.1146/annurev-arplant-043014-115627.
31. Pontvianne F, Blevins T and Pikaard CS. Arabidopsis Histone Lysine Methyltransferases. *Adv Bot Res* 2010; 53: 1-22. 2010/08/13. DOI: 10.1016/S0065-2296(10)53001-5.
32. Zemach A, Kim MY, Hsieh PH, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 2013; 153: 193-205. 2013/04/02. DOI: 10.1016/j.cell.2013.02.033.
33. Guo L, Yu Y, Law JA, et al. SET DOMAIN GROUP2 is the major histone H3 lysine [corrected] 4 trimethyltransferase in Arabidopsis. *Proc Natl Acad Sci U S A* 2010; 107: 18557-18562. 2010/10/13. DOI: 10.1073/pnas.1010478107.
34. Jacob Y and Michaels SD. H3K27me1 is E(z) in animals, but not in plants. *Epigenetics* 2009; 4: 366-369. 2009/09/09. DOI: 10.4161/epi.4.6.9713.

35. Jacob Y, Feng S, LeBlanc CA, et al. ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat Struct Mol Biol* 2009; 16: 763-768.
36. Ferrari KJ, Scelfo A, Jammula S, et al. Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol Cell* 2014; 53: 49-62. 2013/12/03. DOI: 10.1016/j.molcel.2013.10.030.
37. Jacob Y, Stroud H, Leblanc C, et al. Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* 2010; 466: 987-991. 2010/07/16. DOI: nature09290 [pii] 10.1038/nature09290.
38. Du J, Johnson LM, Jacobsen SE, et al. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol* 2015; 16: 519-532. DOI: 10.1038/nrm4043.
39. Bernatavichute YV, Zhang X, Cokus S, et al. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS One* 2008; 3: e3156. 2008/09/09. DOI: 10.1371/journal.pone.0003156.
40. Johnson LM, Du J, Hale CJ, et al. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* 2014; 507: 124-128. 2014/01/28. DOI: 10.1038/nature12931.
41. Harris CJ, Scheibe M, Wongpalee SP, et al. A DNA methylation reader complex that enhances gene transcription. *Science* 2018; 362: 1182-1186. 2018/12/14. DOI: 10.1126/science.aar7854.
42. Groth M, Stroud H, Feng S, et al. SNF2 chromatin remodeler-family proteins FRG1 and -2 are required for RNA-directed DNA methylation. *Proc Natl Acad Sci U S A* 2014; 111: 17666-17671. 2014/11/27. DOI: 10.1073/pnas.1420515111.
43. Han YF, Dou K, Ma ZY, et al. SUV2 is involved in transcriptional gene silencing by associating with SNF2-related chromatin-remodeling proteins in *Arabidopsis*. *Cell Res* 2014; 24: 1445-1465. 2014/11/26. DOI: 10.1038/cr.2014.156.
44. Caro E, Stroud H, Greenberg MV, et al. The SET-domain protein SUV5 mediates H3K9me2 deposition and silencing at stimulus response genes in a DNA methylation-independent manner. *PLoS Genet* 2012; 8: e1002995. 2012/10/17. DOI: 10.1371/journal.pgen.1002995.
45. Law JA and Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 2010; 11: 204-220. 2010/02/10. DOI: nrg2719 [pii] 10.1038/nrg2719.
46. Cokus SJ, Feng S, Zhang X, et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 2008; 452: 215-219.
47. Lister R, O'Malley RC, Tonti-Filippini J, et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 2008; 133: 523-536.
48. Matzke MA and Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet* 2014; 15: 394-408. DOI: 10.1038/nrg3683.
49. Wendte JM and Pikaard CS. The RNAs of RNA-directed DNA methylation. *Biochim Biophys Acta* 2016. DOI: 10.1016/j.bbagr.2016.08.004.
50. Stroud H, Do T, Du J, et al. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat Struct Mol Biol* 2014; 21: 64-72. DOI: 10.1038/nsmb.2735.
51. Jullien PE, Susaki D, Yelagandula R, et al. DNA methylation dynamics during sexual reproduction in *Arabidopsis thaliana*. *Curr Biol* 2012; 22: 1825-1830. 2012/09/04. DOI: 10.1016/j.cub.2012.07.061.
52. Cao X, Aufsatz W, Zilberman D, et al. Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr Biol* 2003; 13: 2212-2217.
53. Zilberman D. An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol* 2017; 18: 87. 2017/05/11. DOI: 10.1186/s13059-017-1230-2.
54. Bewick AJ and Schmitz RJ. Gene body DNA methylation in plants. *Curr Opin Plant Biol* 2017; 36: 103-110. 2017/03/05. DOI: 10.1016/j.pbi.2016.12.007.
55. Feng S, Cokus SJ, Zhang X, et al. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A* 2010; 107: 8689-8694.

56. Choi J, Lyons DB, Kim MY, et al. DNA Methylation and Histone H1 Jointly Repress Transposable Elements and Aberrant Intragenic Transcripts. *Mol Cell* 2019; 2019/11/17. DOI: 10.1016/j.molcel.2019.10.011.
57. Shibuya K, Fukushima S and Takatsuji H. RNA-directed DNA methylation induces transcriptional activation in plants. *Proc Natl Acad Sci U S A* 2009; 106: 1660-1665. 2009/01/24. DOI: 10.1073/pnas.0809294106.
58. Bahar Halpern K, Vana T and Walker MD. Paradoxical role of DNA methylation in activation of FoxA2 gene expression during endoderm development. *J Biol Chem* 2014; 289: 23882-23892. 2014/07/13. DOI: 10.1074/jbc.M114.573469.
59. Zhang H, Lang Z and Zhu JK. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol* 2018; 19: 489-506. DOI: 10.1038/s41580-018-0016-z.
60. Williams BP, Pignatta D, Henikoff S, et al. Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS Genet* 2015; 11: e1005142. 2015/04/01. DOI: 10.1371/journal.pgen.1005142.
61. Lei M, Zhang H, Julian R, et al. Regulatory link between DNA methylation and active demethylation in Arabidopsis. *Proc Natl Acad Sci U S A* 2015; 112: 3553-3557. 2015/03/04. DOI: 10.1073/pnas.1502279112.
62. Biswas S and Rao CM. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. *Eur J Pharmacol* 2018; 837: 8-24. 2018/08/21. DOI: 10.1016/j.ejphar.2018.08.021.
63. Talbert PB and Henikoff S. Histone variants on the move: substrates for chromatin dynamics. *Nat Rev Mol Cell Biol* 2017; 18: 115-126. 2016/12/08. DOI: 10.1038/nrm.2016.148.
64. Lorkovic ZJ and Berger F. Heterochromatin and DNA damage repair: Use different histone variants and relax. *Nucleus* 2017; 8: 583-588. 2017/10/28. DOI: 10.1080/19491034.2017.1384893.
65. Yelagandula R, Stroud H, Holec S, et al. The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. *Cell* 2014; 158: 98-109. 2014/07/06. DOI: 10.1016/j.cell.2014.06.006.
66. Benoit M, Simon L, Desset S, et al. Replication-coupled histone H3.1 deposition determines nucleosome composition and heterochromatin dynamics during Arabidopsis seedling development. *New Phytol* 2019; 221: 385-398. 2018/06/14. DOI: 10.1111/nph.15248.
67. Stroud H, Otero S, Desvoyes B, et al. Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 2012; 109: 5370-5375. 2012/03/21. DOI: 10.1073/pnas.1203145109.
68. Zilberman D, Coleman-Derr D, Ballinger T, et al. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* 2008; 456: 125-129.
69. Coleman-Derr D and Zilberman D. Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genet* 2012; 8: e1002988. 2012/10/17. DOI: 10.1371/journal.pgen.1002988.
70. Clapier CR and Cairns BR. The biology of chromatin remodeling complexes. *Annu Rev Biochem* 2009; 78: 273-304. 2009/04/10. DOI: 10.1146/annurev.biochem.77.062706.153223.
71. Clapier CR, Iwasa J, Cairns BR, et al. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* 2017; 18: 407-422. 2017/05/18. DOI: 10.1038/nrm.2017.26.
72. Tyagi M, Imam N, Verma K, et al. Chromatin remodelers: We are the drivers!! *Nucleus* 2016; 7: 388-404. 2016/07/19. DOI: 10.1080/19491034.2016.1211217.
73. Vongs A, Kakutani T, Martienssen RA, et al. Arabidopsis thaliana DNA methylation mutants. *Science* 1993; 260: 1926-1928.
74. Jeddeloh JA, Stokes TL and Richards EJ. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat Genet* 1999; 22: 94-97. 1999/05/13. DOI: 10.1038/8803.
75. Probst AV, Fransz PF, Paszkowski J, et al. Two means of transcriptional reactivation within heterochromatin. *Plant J* 2003; 33: 743-749.

76. Soppe WJ, Jasencakova Z, Houben A, et al. DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *EMBO J* 2002; 21: 6549-6559.
77. Feng S, Cokus SJ, Schubert V, et al. Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. *Mol Cell* 2014; 55: 694-707. DOI: 10.1016/j.molcel.2014.07.008.
78. Amedeo P, Habu Y, Afsar K, et al. Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. *Nature* 2000; 405: 203-206.
79. Yokthongwattana C, Bucher E, Caikovski M, et al. MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. *Embo J*; 29: 340-351.
80. Han YF, Zhao QY, Dang LL, et al. The SUMO E3 Ligase-Like Proteins PIAL1 and PIAL2 Interact with MOM1 and Form a Novel Complex Required for Transcriptional Silencing. *Plant Cell* 2016; 28: 1215-1229. 2016/04/27. DOI: 10.1105/tpc.15.00997.
81. Li X, Yang L and Chen LL. The Biogenesis, Functions, and Challenges of Circular RNAs. *Mol Cell* 2018; 71: 428-442. 2018/07/31. DOI: 10.1016/j.molcel.2018.06.034.
82. Li S, Xu Z and Sheng J. tRNA-Derived Small RNA: A Novel Regulatory Small Non-Coding RNA. *Genes (Basel)* 2018; 9 2018/05/12. DOI: 10.3390/genes9050246.
83. Ransohoff JD, Wei Y and Khavari PA. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol* 2018; 19: 143-157. 2017/11/16. DOI: 10.1038/nrm.2017.104.
84. Wang HV and Chekanova JA. Long Noncoding RNAs in Plants. *Adv Exp Med Biol* 2017; 1008: 133-154. 2017/08/18. DOI: 10.1007/978-981-10-5203-3_5.
85. Rigal M and Mathieu O. A "mille-feuille" of silencing: epigenetic control of transposable elements. *Biochim Biophys Acta* 2011; 1809: 452-458. 2011/04/26. DOI: 10.1016/j.bbagr.2011.04.001.
86. Lelli KM, Slaterry M and Mann RS. Disentangling the many layers of eukaryotic transcriptional regulation. *Annu Rev Genet* 2012; 46: 43-68. 2012/09/01. DOI: 10.1146/annurev-genet-110711-155437.
87. Latchman DS. Transcription factors: an overview. *Int J Biochem Cell Biol* 1997; 29: 1305-1312. 1998/05/07. DOI: 10.1016/s1357-2725(97)00085-x.
88. Matys V, Kel-Margoulis OV, Fricke E, et al. TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 2006; 34: D108-110. 2005/12/31. DOI: 10.1093/nar/gkj143.
89. Jin J, Tian F, Yang DC, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res* 2017; 45: D1040-D1045. 2016/12/08. DOI: 10.1093/nar/gkw982.
90. Yamasaki K, Kigawa T, Seki M, et al. DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. *Trends Plant Sci* 2013; 18: 267-276. 2012/10/09. DOI: 10.1016/j.tplants.2012.09.001.
91. Lai X, Verhage L, Hugouvieux V, et al. Pioneer Factors in Animals and Plants-Colonizing Chromatin for Gene Regulation. *Molecules* 2018; 23 2018/08/02. DOI: 10.3390/molecules23081914.
92. Orphanides G, Lagrange T and Reinberg D. The general transcription factors of RNA polymerase II. *Genes Dev* 1996; 10: 2657-2683. 1996/11/01. DOI: 10.1101/gad.10.21.2657.
93. Bauer AJ and Martin KA. Coordinating Regulation of Gene Expression in Cardiovascular Disease: Interactions between Chromatin Modifiers and Transcription Factors. *Front Cardiovasc Med* 2017; 4: 19. 2017/04/22. DOI: 10.3389/fcvm.2017.00019.
94. Stasevich TJ, Hayashi-Takanaka Y, Sato Y, et al. Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature* 2014; 516: 272-275. 2014/09/26. DOI: 10.1038/nature13714.
95. Song ZT, Sun L, Lu SJ, et al. Transcription factor interaction with COMPASS-like complex regulates histone H3K4 trimethylation for specific gene expression in plants. *Proc Natl Acad Sci U S A* 2015; 112: 2900-2905. 2015/03/03. DOI: 10.1073/pnas.1419703112.

96. Zhu H, Wang G and Qian J. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet* 2016; 17: 551-565. 2016/08/02. DOI: 10.1038/nrg.2016.83.
97. Yin Y, Morgunova E, Jolma A, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 2017; 356 2017/05/06. DOI: 10.1126/science.aaj2239.
98. Skvortsova K, Iovino N and Bogdanovic O. Functions and mechanisms of epigenetic inheritance in animals. *Nat Rev Mol Cell Biol* 2018; 19: 774-790. 2018/11/15. DOI: 10.1038/s41580-018-0074-2.
99. Perez MF and Lehner B. Intergenerational and transgenerational epigenetic inheritance in animals. *Nat Cell Biol* 2019; 21: 143-151. 2019/01/04. DOI: 10.1038/s41556-018-0242-9.
100. Weiser NE and Kim JK. Multigenerational Regulation of the *Caenorhabditis elegans* Chromatin Landscape by Germline Small RNAs. *Annu Rev Genet* 2019 2019/06/01. DOI: 10.1146/annurev-genet-112618-043505.
101. Boskovic A and Rando OJ. Transgenerational Epigenetic Inheritance. *Annu Rev Genet* 2018; 52: 21-41. 2018/08/31. DOI: 10.1146/annurev-genet-120417-031404.
102. Chandler VL and Stam M. Chromatin conversations: mechanisms and implications of paramutation. *Nat Rev Genet* 2004; 5: 532-544. 2004/06/24. DOI: 10.1038/nrg1378.
103. Hollick JB. Paramutation and related phenomena in diverse species. *Nat Rev Genet* 2017; 18: 5-23. 2016/11/01. DOI: 10.1038/nrg.2016.115.
104. Johannes F, Porcher E, Teixeira FK, et al. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet* 2009; 5: e1000530. 2009/06/27. DOI: 10.1371/journal.pgen.1000530.
105. Reinders J, Wulff BB, Mirouze M, et al. Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev* 2009; 23: 939-950. 2009/04/25. DOI: 10.1101/gad.524609.
106. Teixeira FK, Heredia F, Sarazin A, et al. A role for RNAi in the selective correction of DNA methylation defects. *Science* 2009; 323: 1600-1604.
107. Lisch D. How important are transposons for plant evolution? *Nat Rev Genet* 2013; 14: 49-61. 2012/12/19. DOI: 10.1038/nrg3374.
108. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 2008; 9: 397-405. 2008/03/28. DOI: 10.1038/nrg2337.
109. Chuong EB, Elde NC and Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* 2017; 18: 71-86. 2016/11/22. DOI: 10.1038/nrg.2016.139.
110. O'Donnell KA and Burns KH. Mobilizing diversity: transposable element insertions in genetic variation and disease. *Mob DNA* 2010; 1: 21. 2010/09/04. DOI: 10.1186/1759-8753-1-21.
111. Rebollo R, Romanish MT and Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet* 2012; 46: 21-42. 2012/08/22. DOI: 10.1146/annurev-genet-110711-155621.
112. Lynch VJ, Leclerc RD, May G, et al. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* 2011; 43: 1154-1159. 2011/09/29. DOI: 10.1038/ng.917.
113. Quadrana L, Etcheverry M, Gilly A, et al. Transposition favors the generation of large effect mutations that may facilitate rapid adaptation. *Nat Commun* 2019; 10: 3421. 2019/08/02. DOI: 10.1038/s41467-019-11385-5.
114. Jangam D, Feschotte C and Betran E. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. *Trends Genet* 2017; 33: 817-831. 2017/08/29. DOI: 10.1016/j.tig.2017.07.011.
115. Joly-Lopez Z and Bureau TE. Exaptation of transposable element coding sequences. *Curr Opin Genet Dev* 2018; 49: 34-42. 2018/03/12. DOI: 10.1016/j.gde.2018.02.011.
116. Cosby RL, Chang NC and Feschotte C. Host-transposon interactions: conflict, cooperation, and cooption. *Genes Dev* 2019; 33: 1098-1116. 2019/09/05. DOI: 10.1101/gad.327312.119.

117. Joly-Lopez Z, Hoen DR, Blanchette M, et al. Phylogenetic and Genomic Analyses Resolve the Origin of Important Plant Genes Derived from Transposable Elements. *Mol Biol Evol* 2016; 33: 1937-1956. 2016/05/18. DOI: 10.1093/molbev/msw067.
118. Bundock P and Hooykaas P. An Arabidopsis hAT-like transposase is essential for plant development. *Nature* 2005; 436: 282-284. 2005/07/15. DOI: 10.1038/nature03667.
119. Lin R, Ding L, Casola C, et al. Transposase-derived transcription factors regulate light signaling in Arabidopsis. *Science* 2007; 318: 1302-1305. 2007/11/24. DOI: 10.1126/science.1146281.
120. Liang SC, Hartwig B, Perera P, et al. Kicking against the PRCs - A Domesticated Transposase Antagonises Silencing Mediated by Polycomb Group Proteins and Is an Accessory Component of Polycomb Repressive Complex 2. *PLoS Genet* 2015; 11: e1005660. 2015/12/08. DOI: 10.1371/journal.pgen.1005660.
121. Duan CG, Wang X, Xie S, et al. A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation. *Cell Res* 2017; 27: 226-240. 2016/12/10. DOI: 10.1038/cr.2016.147.
122. Uhlken C, Horvath B, Stadler R, et al. MAIN-LIKE1 is a crucial factor for correct cell division and differentiation in Arabidopsis thaliana. *Plant J* 2014; 78: 107-120. DOI: 10.1111/tpj.12455.
123. Wenig U, Meyer S, Stadler R, et al. Identification of MAIN, a factor involved in genome stability in the meristems of Arabidopsis thaliana. *Plant J* 2013; 75: 469-483. DOI: 10.1111/tpj.12215.
124. Ikeda Y, Pelissier T, Bourguet P, et al. Arabidopsis proteins with a transposon-related domain act in gene silencing. *Nat Commun* 2017; 8: 15122. DOI: 10.1038/ncomms15122.
125. Joly-Lopez Z, Forczek E, Hoen DR, et al. A gene family derived from transposable elements during early angiosperm evolution has reproductive fitness benefits in Arabidopsis thaliana. *PLoS Genet* 2012; 8: e1002931. 2012/09/13. DOI: 10.1371/journal.pgen.1002931.
126. Moissiard G, Parizotto EA, Himber C, et al. Corrigendum: Transitivity in Arabidopsis can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins. *RNA* 2016; 22: 810. DOI: 10.1261/rna.056366.116.
127. Haas G, Azevedo J, Moissiard G, et al. Nuclear import of CaMV P6 is required for infection and suppression of the RNA silencing factor DRB4. *EMBO J* 2015; 34: 2591-2592. DOI: 10.15252/embj.201570060.
128. Deleris A, Greenberg MV, Ausin I, et al. Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. *EMBO Rep* 2010; 11: 950-955. 2010/11/06. DOI: embor2010158 [pii] 10.1038/embor.2010.158.
129. Henderson IR and Jacobsen SE. Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev* 2008; 22: 1597-1606.
130. Hsieh TF, Shin J, Uzawa R, et al. Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc Natl Acad Sci U S A* 2011; 108: 1755-1762. 2011/01/25. DOI: 10.1073/pnas.1019273108.
131. Vu TM, Nakamura M, Calarco JP, et al. RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis. *Development* 2013; 140: 2953-2960. 2013/06/14. DOI: 10.1242/dev.092981.
132. Moissiard G, Cokus SJ, Cary J, et al. MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* 2012; 336: 1448-1451. Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't 2012/05/05. DOI: 10.1126/science.1221472.
133. Groth M, Moissiard G, Wirtz M, et al. MTHFD1 controls DNA methylation in Arabidopsis. *Nat Commun* 2016; 7: 11640. DOI: 10.1038/ncomms11640.
134. Iyer LM, Abhiman S and Aravind L. MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biology direct* 2008; 3: 8. Comparative Study Research Support, N.I.H., Intramural 2008/03/19. DOI: 10.1186/1745-6150-3-8.

135. Dong W, Vannozzi A, Chen F, et al. MORC Domain Definition and Evolutionary Analysis of the MORC Gene Family in Green Plants. *Genome Biol Evol* 2018; 10: 1730-1744. 2018/07/10. DOI: 10.1093/gbe/evy136.
136. Li DQ, Nair SS and Kumar R. The MORC family: new epigenetic regulators of transcription and DNA damage response. *Epigenetics* 2013; 8: 685-693. 2013/06/28. DOI: 10.4161/epi.24976.
137. Koch A, Kang HG, Steinbrenner J, et al. MORC Proteins: Novel Players in Plant and Animal Health. *Front Plant Sci* 2017; 8: 1720. 2017/11/03. DOI: 10.3389/fpls.2017.01720.
138. Moissiard G, Bischof S, Husmann D, et al. Transcriptional gene silencing by Arabidopsis microorchidia homologues involves the formation of heteromers. *Proc Natl Acad Sci U S A* 2014; 111: 7474-7479. 2014/05/07. DOI: 10.1073/pnas.1406611111.
139. Lorkovic ZJ, Naumann U, Matzke AJ, et al. Involvement of a GHKL ATPase in RNA-directed DNA methylation in Arabidopsis thaliana. *Curr Biol* 2012; 22: 933-938. DOI: 10.1016/j.cub.2012.03.061.
140. Brabbs TR, He Z, Hogg K, et al. The stochastic silencing phenotype of Arabidopsis morc6 mutants reveals a role in efficient RNA-directed DNA methylation. *Plant J* 2013; 75: 836-846. DOI: 10.1111/tpj.12246.
141. Harris CJ, Husmann D, Liu W, et al. Arabidopsis AtMORC4 and AtMORC7 Form Nuclear Bodies and Repress a Large Number of Protein-Coding Genes. *PLoS Genet* 2016; 12: e1005998. DOI: 10.1371/journal.pgen.1005998.
142. Liu ZW, Shao CR, Zhang CJ, et al. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLoS Genet* 2014; 10: e1003948. 2014/01/28. DOI: 10.1371/journal.pgen.1003948.
143. Liu ZW, Zhou JX, Huang HW, et al. Two Components of the RNA-Directed DNA Methylation Pathway Associate with MORC6 and Silence Loci Targeted by MORC6 in Arabidopsis. *PLoS Genet* 2016; 12: e1006026. 2016/05/14. DOI: 10.1371/journal.pgen.1006026.
144. Gallego-Bartolome J, Liu W, Kuo PH, et al. Co-targeting RNA Polymerases IV and V Promotes Efficient De Novo DNA Methylation in Arabidopsis. *Cell* 2019; 176: 1068-1082 e1019. 2019/02/12. DOI: 10.1016/j.cell.2019.01.029.
145. Inoue N, Hess KD, Moreadith RW, et al. New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum Mol Genet* 1999; 8: 1201-1207. Research Support, Non-U.S. Gov't
Research Support, U.S. Gov't, P.H.S. 1999/06/17.
146. Pastor WA, Stroud H, Nee K, et al. MORC1 represses transposable elements in the mouse male germline. *Nat Commun* 2014; 5: 5795. DOI: 10.1038/ncomms6795.
147. Kim H, Yen L, Wongpalee SP, et al. The Gene-Silencing Protein MORC-1 Topologically Entraps DNA and Forms Multimeric Assemblies to Cause DNA Compaction. *Mol Cell* 2019; 75: 700-710 e706. 2019/08/24. DOI: 10.1016/j.molcel.2019.07.032.
148. Yan X, Ma L, Pang H, et al. METHIONINE SYNTHASE1 Is Involved in Chromatin Silencing by Maintaining DNA and Histone Methylation. *Plant Physiol* 2019; 181: 249-261. 2019/07/25. DOI: 10.1104/pp.19.00528.
149. Babu MM, Iyer LM, Balaji S, et al. The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res* 2006; 34: 6505-6520. DOI: 10.1093/nar/gkl888.
150. Steinbauerova V, Neumann P, Novak P, et al. A widespread occurrence of extra open reading frames in plant Ty3/gypsy retrotransposons. *Genetica* 2011; 139: 1543-1555. DOI: 10.1007/s10709-012-9654-9.
151. Stroud H, Greenberg MV, Feng S, et al. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 2013; 152: 352-364. DOI: 10.1016/j.cell.2012.10.054.
152. Whittaker C and Dean C. The FLC Locus: A Platform for Discoveries in Epigenetics and Adaptation. *Annu Rev Cell Dev Biol* 2017; 33: 555-575. 2017/07/12. DOI: 10.1146/annurev-cellbio-100616-060546.

153. Macaisne N, Novatchkova M, Peirera L, et al. SHOC1, an XPF endonuclease-related protein, is essential for the formation of class I meiotic crossovers. *Curr Biol* 2008; 18: 1432-1437. 2008/09/25. DOI: 10.1016/j.cub.2008.08.041.
154. Lawrence EJ, Gao H, Tock AJ, et al. Natural Variation in TBP-ASSOCIATED FACTOR 4b Controls Meiotic Crossover and Germline Transcription in Arabidopsis. *Curr Biol* 2019; 29: 2676-2686 e2673. 2019/08/06. DOI: 10.1016/j.cub.2019.06.084.
155. Wang Y, Lyu W, Berkowitz O, et al. Inactivation of Mitochondrial Complex I Induces the Expression of a Twin Cysteine Protein that Targets and Affects Cytosolic, Chloroplastidic and Mitochondrial Function. *Mol Plant* 2016; 9: 696-710. 2016/02/02. DOI: 10.1016/j.molp.2016.01.009.
156. Li C, Xu J, Li J, et al. Involvement of Arabidopsis histone acetyltransferase HAC family genes in the ethylene signaling pathway. *Plant Cell Physiol* 2014; 55: 426-435. 2013/11/30. DOI: 10.1093/pcp/pct180.
157. Aguilar-Martinez JA, Uchida N, Townsley B, et al. Transcriptional, posttranscriptional, and posttranslational regulation of SHOOT MERISTEMLESS gene expression in Arabidopsis determines gene function in the shoot apex. *Plant Physiol* 2015; 167: 424-442. 2014/12/20. DOI: 10.1104/pp.114.248625.
158. Enders TA, Frick EM and Strader LC. An Arabidopsis kinase cascade influences auxin-responsive cell expansion. *Plant J* 2017; 92: 68-81. 2017/07/16. DOI: 10.1111/tpj.13635.
159. Calixto CPG, Guo W, James AB, et al. Rapid and Dynamic Alternative Splicing Impacts the Arabidopsis Cold Response Transcriptome. *Plant Cell* 2018; 30: 1424-1444. 2018/05/17. DOI: 10.1105/tpc.18.00177.
160. Mitra D, Klemm S, Kumari P, et al. Microtubule-associated protein IQ67 DOMAIN5 regulates morphogenesis of leaf pavement cells in Arabidopsis thaliana. *J Exp Bot* 2019; 70: 529-543. 2018/11/09. DOI: 10.1093/jxb/ery395.
161. Ngaki MN, Louie GV, Philippe RN, et al. Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature* 2012; 485: 530-533. 2012/05/25. DOI: 10.1038/nature11009.
162. Wu Z, Ietswaart R, Liu F, et al. Quantitative regulation of FLC via coordinated transcriptional initiation and elongation. *Proc Natl Acad Sci U S A* 2016; 113: 218-223. 2015/12/25. DOI: 10.1073/pnas.1518369112.
163. Franco-Zorrilla JM, Lopez-Vidriero I, Carrasco JL, et al. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci U S A* 2014; 111: 2367-2372. DOI: 10.1073/pnas.1316278111.
164. Bartlett A, O'Malley RC, Huang SC, et al. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc* 2017; 12: 1659-1672. 2017/07/21. DOI: 10.1038/nprot.2017.055.
165. de Luxan-Hernandez C, Lohmann J, Hellmeyer W, et al. PP7L is essential for MAIL1-mediated transposable element silencing and primary root growth. *Plant J* 2019 2019/12/19. DOI: 10.1111/tpj.14655.
166. Gil RS and Vagnarelli P. Protein phosphatases in chromatin structure and function. *Biochim Biophys Acta Mol Cell Res* 2019; 1866: 90-101. 2018/07/24. DOI: 10.1016/j.bbamcr.2018.07.016.
167. Biro J, Farkas I, Domoki M, et al. The histone phosphatase inhibitory property of plant nucleosome assembly protein-related proteins (NRPs). *Plant Physiol Biochem* 2012; 52: 162-168. 2012/01/31. DOI: 10.1016/j.plaphy.2011.12.010.
168. Wang S, Quan L, Li S, et al. The Protein Phosphatase 4 Complex Promotes Transcription and Processing of Primary microRNAs in Arabidopsis. *Plant Cell* 2019 2019/01/25. DOI: 10.1105/tpc.18.00556.
169. Goldberg J, Huang HB, Kwon YG, et al. Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature* 1995; 376: 745-753. 1995/08/31. DOI: 10.1038/376745a0.
170. Farkas I, Dombradi V, Miskei M, et al. Arabidopsis PPP family of serine/threonine phosphatases. *Trends Plant Sci* 2007; 12: 169-176. 2007/03/21. DOI: 10.1016/j.tplants.2007.03.003.

171. Reece-Hoyes JS and Marian Walhout AJ. Yeast one-hybrid assays: a historical and technical perspective. *Methods* 2012; 57: 441-447. 2012/08/14. DOI: 10.1016/j.ymeth.2012.07.027.
172. Soppe WJ, Jacobsen SE, Alonso-Blanco C, et al. The Late Flowering Phenotype of *fwa* Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene. *Mol Cell* 2000; 6: 791-802.
173. Lippman Z, Gendrel AV, Black M, et al. Role of transposable elements in heterochromatin and epigenetic control. *Nature* 2004; 430: 471-476.
174. Gallego-Bartolome J, Gardiner J, Liu W, et al. Targeted DNA demethylation of the Arabidopsis genome using the human TET1 catalytic domain. *Proc Natl Acad Sci U S A* 2018; 115: E2125-E2134. 2018/02/16. DOI: 10.1073/pnas.1716945115.
175. Kumlehn J, Pietralla J, Hensel G, et al. The CRISPR/Cas revolution continues: From efficient gene editing for crop breeding to plant synthetic biology. *J Integr Plant Biol* 2018; 60: 1127-1153. 2018/11/06. DOI: 10.1111/jipb.12734.
176. Zhang C, Freddolino PL and Zhang Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Res* 2017; 45: W291-W299. 2017/05/05. DOI: 10.1093/nar/gkx366.
177. Li S, Chen K and Grierson D. A critical evaluation of the role of ethylene and MADS transcription factors in the network controlling fleshy fruit ripening. *New Phytol* 2019; 221: 1724-1741. 2018/10/18. DOI: 10.1111/nph.15545.
178. Gao Y, Zhu N, Zhu X, et al. Diversity and redundancy of the ripening regulatory networks revealed by the fruitENCODE and the new CRISPR/Cas9 CNR and NOR mutants. *Hortic Res* 2019; 6: 39. 2019/02/19. DOI: 10.1038/s41438-019-0122-x.
179. Liu M, Pirrello J, Chervin C, et al. Ethylene Control of Fruit Ripening: Revisiting the Complex Network of Transcriptional Regulation. *Plant Physiol* 2015; 169: 2380-2390. DOI: 10.1104/pp.15.01361.
180. Liu R, How-Kit A, Stammitti L, et al. A DEMETER-like DNA demethylase governs tomato fruit ripening. *Proc Natl Acad Sci U S A* 2015; 112: 10804-10809. 2015/08/12. DOI: 10.1073/pnas.1503362112.
181. Giovannoni J, Nguyen C, Ampofo B, et al. The Epigenome and Transcriptional Dynamics of Fruit Ripening. *Annu Rev Plant Biol* 2017; 68: 61-84. 2017/02/23. DOI: 10.1146/annurev-arplant-042916-040906.
182. E. B, J. K, E. T, et al. Epigenetic Regulations of Fleshy Fruit Development and Ripening and Their Potential Applications to Breeding Strategies. *Advances in Botanical Research* 2018; 88: 327-360.
183. Zhang Y, Li Q, Jiang L, et al. Suppressing Type 2C Protein Phosphatases Alters Fruit Ripening and the Stress Response in Tomato. *Plant Cell Physiol* 2018; 59: 142-154. 2017/11/10. DOI: 10.1093/pcp/pcx169.
184. Zouine M, Maza E, Djari A, et al. TomExpress, a unified tomato RNA-Seq platform for visualization of expression data, clustering and correlation networks. *Plant J* 2017; 92: 727-735. DOI: 10.1111/tpj.13711.
185. Shinozaki Y, Nicolas P, Fernandez-Pozo N, et al. High-resolution spatiotemporal transcriptome mapping of tomato fruit development and ripening. *Nat Commun* 2018; 9: 364. DOI: 10.1038/s41467-017-02782-9.
186. Vazquez-Vilar M, Quijano-Rubio A, Fernandez-Del-Carmen A, et al. GB3.0: a platform for plant bio-design that connects functional DNA elements with associated biological data. *Nucleic Acids Res* 2017; 45: 2196-2209. 2017/01/06. DOI: 10.1093/nar/gkw1326.
187. Liu H, Ding Y, Zhou Y, et al. CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. *Mol Plant* 2017; 10: 530-532. 2017/01/17. DOI: 10.1016/j.molp.2017.01.003.
188. Copetti D, Zhang J, El Baidouri M, et al. RiTE database: a resource database for genus-wide rice genomics and evolutionary biology. *BMC Genomics* 2015; 16: 538. 2015/07/22. DOI: 10.1186/s12864-015-1762-3.

189. Stein JC, Yu Y, Copetti D, et al. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat Genet* 2018; 50: 285-296. 2018/01/24. DOI: 10.1038/s41588-018-0040-0.
190. Lanciano S, Carpentier MC, Llauro C, et al. Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants. *PLoS Genet* 2017; 13: e1006630. 2017/02/18. DOI: 10.1371/journal.pgen.1006630.
191. Hoen DR and Bureau TE. Discovery of novel genes derived from transposable elements using integrative genomic analysis. *Mol Biol Evol* 2015; 32: 1487-1506. 2015/02/26. DOI: 10.1093/molbev/msv042.
192. Joly-Lopez Z, Forczek E, Vello E, et al. Abiotic Stress Phenotypes Are Associated with Conserved Genes Derived from Transposable Elements. *Front Plant Sci* 2017; 8: 2027. 2017/12/19. DOI: 10.3389/fpls.2017.02027.

VII. Appendix.

1. Nicolau M., Picault N., Descombin J., Jami-AlahmadiY., Feng S., Bucher E., Jacobsen S.E., Deragon JM., Wohlschlegel J. and **Moissiard G.** The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*. *Submitted to PLOS Genetics, in revision*.
2. Groth M., **Moissiard G.**, Wirtz M., et al. MTHFD1 controls DNA methylation in *Arabidopsis*. *Nat Commun.* **2016**; 7: 11640.
3. Harris, C.J., Husmann, D., Liu, W., Kasmi, F.E., Wang, H., Papikian, A., Pastor, W.A., **Moissiard, G.**, Vashisht, A.A., Dangl, J.L., et al. *Arabidopsis* AtMORC4 and AtMORC7 Form Nuclear Bodies and Repress a Large Number of Protein-Coding Genes. *PLoS Genet.* **2016**; 12: e1005998.
4. **Moissiard, G.**, Bischof, S., Husmann, D., Pastor, W. A., Hale, C. J., Yen, L., Stroud, H., Papikian, A., Vashisht, A. A., Wohlschlegel, J. A. and Jacobsen, S. E. Transcriptional gene silencing by *Arabidopsis* microorchidia homologues involves the formation of heteromers. *Proc Natl Acad Sci U S A.* **2014**;111(20):7474-7479.
5. Pastor, W. A., Stroud, H., Nee, K., Liu, W., Pezic, D., Manakov, S., Lee, S. A., **Moissiard, G.**, Zamudio, N., Bourc'his, D., Aravin, A. A., Clark, A. T., Jacobsen, S. E. MORC1 represses transposable elements in the mouse male germline. *Nat Commun.* **2014**; 5:5795. Erratum 2015; 6:7604.
6. **Moissiard, G.**, Cokus, S. J., Cary, J., Feng, S., Billi, A. C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B. R., McCord, R. P., Hale, C. J., Feng, W., Michaels, S. D., Frand, A. R., Pellegrini, M., Dekker, J., Kim, J. K., Jacobsen, S. E. MORC family ATPases required for heterochromatin condensation and gene silencing. *Science.* **2012**;336: 1448-1451.
7. Deleris, A., Greenberg, M. V., Ausin, I., Law, R. W., **Moissiard, G.**, Schubert, D., Jacobsen, S. E. Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. *EMBO Rep.* **2010**;11(12):950-955.

The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*.

Short title: The PMD MAIN/MAIL1 and PP7L complex regulates gene expression and TE silencing.

Melody Nicolau^{1,2}, Nathalie Picault^{1,2}, Julie Descombin^{1,2}, Yasaman Jami-Alahmadi³, Suhua Feng⁴, Etienne Bucher⁵, Steven E. Jacobsen^{4,6}, Jean-Marc Deragon^{1,2}, James Wohlschlegel³ and Guillaume Moissiard^{1,2*}.

1 LGDP-UMR5096, CNRS, Perpignan, France.

2 LGDP-UMR5096, Université de Perpignan Via Domitia, France.

3 Department of Biological Chemistry, University of California at Los Angeles, Los Angeles, CA, USA.

4 Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, CA, USA.

5 Plant Breeding and Genetic Resources, Agroscope, Nyon, Switzerland.

6 Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, CA, USA.

* Corresponding author

E-mail: guillaume.moissiard@univ-perp.fr

ABSTRACT

Transposable elements (TEs) are DNA repeats that must remain silenced to ensure cell integrity. Several epigenetic pathways including DNA methylation and histone modifications are involved in the silencing of TEs, and in the regulation of gene expression. In *Arabidopsis thaliana*, the TE-derived plant mobile domain (PMD) proteins have been involved in TE silencing, genome stability, and control of developmental processes. Using a forward genetic screen, we found that the PMD protein MAINTENANCE OF MERISTEMS (MAIN) acts synergistically and redundantly with DNA methylation to silence TEs. We found that MAIN and its close homolog MAIN-LIKE 1 (MAIL1) interact together, as well as with the phosphoprotein phosphatase (PPP) PP7-like (PP7L). Remarkably, *main*, *mail1*, *pp7l* single and *mail1 pp7l* double mutants display similar developmental phenotypes, and share common subsets of upregulated TEs and misregulated genes. Finally, phylogenetic analyses of PMD and PP7-type PPP domains among the Eudicot lineage suggest neo-association processes between the two protein domains to potentially generate new protein function. We propose that, through this interaction, the PMD and PPP domains may constitute a functional protein module required for the proper expression of a common set of genes, and for silencing of TEs.

AUTHOR SUMMARY

The plant mobile domain (PMD) is a protein domain of unknown function that is widely spread in the angiosperm plants. Although most PMDs are associated with repeated DNA sequences called transposable elements (TEs), plants have domesticated the PMD to produce genic versions that play important roles within the cell. In *Arabidopsis thaliana*, MAINTENANCE OF MERISTEMS (MAIN) and MAIN-LIKE 1 (MAIL1) are genic PMDs that are involved in genome stability, developmental processes, and silencing of TEs. The mechanisms involving MAIN and MAIL1 in these cellular processes remain elusive. Here, we show that MAIN, MAIL1 and the phosphoprotein phosphatase (PPP) named PP7-like (PP7L) interact to form a protein complex that is required for the proper expression of genes, and the

silencing of TEs. Phylogenetic analyses revealed that PMD and PP7-type PPP domains are evolutionary connected, and several plant species express proteins carrying both PMD and PPP domains. We propose that interaction of PMD and PPP domains would create a functional protein module involved in mechanisms regulating gene expression and repressing TEs.

INTRODUCTION

In eukaryotes, DNA methylation and post-translational modifications of histones are epigenetic marks involved in chromatin organization, regulation of gene expression and silencing of DNA repeats such as transposable elements (TEs) [1-3]. Constitutive heterochromatin is highly condensed and enriched in silenced TEs that are targeted by DNA methylation and histone H3 lysine 9 dimethylation (H3K9me₂). Euchromatin is more relaxed and composed of genes that are more permissive to transcription, depending on the recruitment of transcription factors (TFs), cofactors and RNA polymerases [1, 4]. In plants, DNA methylation occurs in three different cytosine contexts: CG, CHG and CHH (where H = A, T or C), involving specialized DNA methyltransferases [5]. In *Arabidopsis thaliana*, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) and DRM1 mediate de novo DNA methylation in all sequence contexts through the RNA-directed DNA methylation (RdDM) pathway, which involves among other components, RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) and DICER-LIKE 3 (DCL3) for the production of short interfering (si)RNAs [6, 7]. The maintenance of CG methylation is specifically performed by METHYLTRANSFERASE 1 (MET1), while CHROMOMETHYLASE 2 (CMT2) and CMT3 are involved in the maintenance at CHG sites [8, 9]. CMT2 can also be involved in the deposition of CHH methylation at specific genomic location [10, 11]. Finally, DRM2 is mostly required for the maintenance of CHH methylation through the RdDM pathway [6, 7, 9]. Together with DNA methylation, additional pathways play important roles in TE silencing. The MICRORCHIDIA 1 (MORC1) and MORC6 ATPases interact together, and are required for heterochromatin condensation and repression of TEs, acting mostly downstream of DNA methylation and RdDM pathway [12-14].

More recently, the *A. thaliana* plant mobile domain (PMD) proteins MAINTENANCE OF

MERISTEM (MAIN) and MAIN-LIKE 1 (MAIL1) were identified as new factors required for TE silencing [15]. In addition, these two proteins have been involved in genome stability, and regulation of developmental processes such as cell division and differentiation [16, 17]. The PMD is a large protein domain of unknown function that is widely represented among the angiosperms, predominantly associated with TEs [15, 18]. It has been proposed that genic PMD versions, such as the MAIN and MAIL1 proteins derived from TEs after gene domestication [15, 18, 19]. Previous studies suggested that genic PMDs could act as cellular factors related to transcription, possibly acting as transcription factor (TF)-like, co-factor or repressor proteins regulating this cellular process [16, 18]. Nevertheless, the role of PMD proteins in the regulation of transcription remains elusive. Most of genic PMD proteins are standalone versions, however, in some cases, the PMD is fused to another protein domain, such as protease, kinase or metallo-phosphatase (MPP) domains. For instance in *A. thaliana*, the MAIL3 protein carries a PMD, which is fused to a putative serine/threonine-specific phosphoprotein phosphatase (PPP) domain phylogenetically related to the plant-specific protein phosphatase 7 (PP7) [20]. PP7 is a calmodulin-binding PPP that has been related to cryptochrome (CRY)-mediated blue-light signaling, and to the control of stomatal aperture [20-22]. PP7 is also involved in the perception of red/far red light by controlling the phytochrome pathway [23, 24]. In addition to PP7 and MAIL3 (also known as “long PP7”), the protein PP7-like (PP7L) belongs to the same phylogenetic clade [20]. PP7L was recently identified as a nuclear protein involved in chloroplast development and abiotic stress tolerance [25]. The *pp7l* mutant plants showed photosynthetic defects and strong developmental phenotype associated with misregulation of several genes [25].

In this study, we described a forward genetic screen based on a GFP reporter gene that allowed us to identify a mutant population in which *MAIN* is mutated, leading to GFP overexpression. We then deciphered the genetic interaction between the DRM2, CMT3 and MAIN, showing that these proteins are part of different epigenetic pathways that act redundantly or synergistically to repress TEs. Biochemical analyses indicated that MAIN and MAIL1 physically interact together. These analyses also identified PP7L as a robust interactor of MAIN and MAIL1 proteins. In addition, the

characterization of developmental and molecular phenotypes of *pmd* and *pp7l* single and double mutant plants strongly suggest that these proteins interact together to silence TEs, and regulate the expression of a common set of genes. Finally, phylogenetic analyses allowed us to determine the distribution of PMD and PP7/PP7L domains among the Eudicots. Based on these analyses, we have evidences of co-evolution linked to the neo-association of the PMD and PP7-type PPP domains on single proteins in several Eudicot species, suggesting a convergent evolution between these two protein domains.

RESULTS

Mutation in *MAIN* is responsible for TE silencing defects.

The *ATCOPIA28* retrotransposon *AT3TE51900* (hereafter called *ATCOPIA28*) is targeted by several epigenetic pathways such as DNA methylation and the MORC1/6 complex, which altogether contribute to its repression. We engineered a construct in which the 5' long terminal repeat (LTR) promoter region of *ATCOPIA28* controls GFP transcription (Fig 1A). While the *ATCOPIA28::GFP* transgene is fully silenced in wild type (WT) plants, it is weakly expressed in the DNA methylation-deficient *drm1 drm2 cmt3 (ddc)* triple mutant background (Fig 1B) [26]. We performed an ethyl methane sulfonate (EMS) mutagenesis using the *ATCOPIA28::GFP ddc* plants as sensitized genetic material, and screened for mutant populations showing GFP overexpression. Among, the selected populations, we retrieved two new mutant alleles of *MORC6* carrying missense mutations in either the GHKL or S5 domains of the protein (S1A-C Fig). We also identified the population *ddc #16* showing strong overexpression of GFP and misregulation of several endogenous TEs, including *ATCOPIA28* (Fig 1B-D). Mapping experiments based on whole genome resequencing and bulk segregant analysis indicated that *ddc #16* carries a missense point mutation (C230Y) in the gene *AT1G17930*, previously named *MAIN* (S1D and S1E Fig). Genetic complementation analyses by crossing the *ddc #16* EMS mutant with the knock-out (KO) transferred DNA (T-DNA) insertion line *main-2* generated F1 *ddc #16*

x *main-2* plants that did not express the GFP (S1F Fig). Transcriptional profiling analyses showed, however, that endogenous TEs, including *ATCOPIA28*, were upregulated in F1 *ddc #16* x *main-2* plants, but not in F1 control plants generated from the backcross of *ddc #16* with WT Columbia (Col) plants (S1G Fig). Self-fertilization of F1 *ddc #16* x *main-2* plants allowed us to retrieve several F2 *ddc #16* x *main-2* plants overexpressing the GFP (S1F Fig). Among these GFP positive F2 plants, we identified individuals that were either homozygote for the EMS mutation in the *MAIN* gene, or plants carrying both the EMS and T-DNA *main-2* mutant alleles (S1F Fig). Moreover, while all these plants were homozygote for the *drm2* mutation, half of them segregated the *cmt3* mutation. Thus, altogether, these analyses suggested that *ATCOPIA28::GFP* silencing is more DRM2- than CMT3-dependent. More importantly, they confirmed that *MAIN* was the mutated gene causing the upregulation of *ATCOPIA28::GFP* and several endogenous TEs. Therefore, *ddc #16* was renamed *ddc main-3*.

The MAIN, DRM2 and CMT3 pathways act synergistically to repress TEs and DNA-methylated genes.

To determine the genetic interaction of *ddc* and *main-3* mutations on TE silencing, we carried out two independent RNA sequencing (RNA-seq) experiments in the hypomorphic *main-3* single, *ddc* triple and *ddc main-3* quadruple mutant plants (Fig 2A and S2A Fig). As previously described, the *ddc* mutant showed upregulation of several TEs spread over the five chromosomes (Fig 2B-D and S2B Fig, and S1 Table) [11]. Loss of TE silencing was also observed to a milder degree in the *main-3* mutant, with the significant enrichment of pericentromeric TEs among the upregulates TEs (Fig 2B-D and S2B Fig, and S1 Table). The *ddc main-3* mutant showed an exacerbation of TE silencing defects, with a large number of pericentromeric TEs being specifically upregulated in this mutant background (Fig 2B-D and S2B Fig, and S1 Table). Comparative analyses revealed that upregulated TEs cluster into four distinct classes (Fig 2E and S2C Fig). Class I TEs are upregulated in *ddc*, *main-3* and *ddc main-3* mutants (Fig 2E and S2C-D Fig). Class II and class III TEs are targeted by the MAIN and DRM2/CMT3 pathways, respectively (Fig 2E and S2C-D Fig). However, the upregulation of class II and class III TEs is further enhanced in *ddc main-3*, which suggests that the MAIN and DRM2/CMT3 pathways can partially

compensate each other at these genomic locations (S2D Fig). Finally, the most abundant class IV TEs are only misregulated in *ddc main-3*, which implies that the MAIN and DRM2/CMT3 pathways act redundantly to silence these TEs (Fig 2E and S2C-D Fig).

Several genes were also misregulated in the three mutant backgrounds (S1 Table). Among these genes, a subset was commonly upregulated in *ddc*, *main-3* and *ddc main-3* (S2E Fig). Remarkably, genes that were upregulated in *ddc*, *main-3* or *ddc main-3* were significantly enriched in pericentromeric regions of chromosomes, where constitutive heterochromatin resides (S2F Fig). This is consistent with the fact that, among these upregulated genes, we identified a large proportion of genes that were DNA-methylated (in the three cytosine contexts) and targeted by H3K9me2 (S2F Fig). Conversely, we could only identify one gene commonly downregulated in *ddc*, *main-3* and *ddc main-3* (S2F Fig). Furthermore, downregulated genes in *ddc*, *main-3* or *ddc main-3* were rather enriched in chromosome arms, and most of them were not DNA-methylated genes (S2F Fig).

To further dissect the genetic interaction between the DRM2, CMT3 and MAIN pathways, we generated the *drm1 drm2 main-3* (*dd main-3*) and *cmt3 main-3* mutants (S2G Fig). We then analyzed the expression level of several TEs previously identified as misregulated in *ddc*, *main-3* and/or *ddc main-3*. The endogenous *ATCOPIA28* was the most expressed in *ddc main-3* and *dd main-3*, and to a lesser extent, in *cmt3 main-3* (Fig 2F). This is consistent with the fact that all the F2 *ddc* #16 x *main-2* plants overexpressing *ATCOPIA28::GFP* were *drm2* homozygote, although they segregated the *cmt3* mutation (S1F Fig). Further analyses showed that most of the tested TEs tend to be more expressed in *cmt3 main-3* than in *dd main-3*, with the exception of *ATIS112A* that was more upregulated in *dd main-3* than in *cmt3 main-3* (Fig 2G). In conclusion, these analyses showed complex genetic interactions between the DRM2, CMT3 and MAIN pathways, suggesting that MAIN and DNA methylation pathways act synergistically to repress TEs and DNA-methylated genes.

MAIN and MAIL1 are required for the proper expression of a common set of genes and TEs.

Beside a role of MAIN in TE and gene silencing, our transcriptomic analyses using the

hypomorphic *main-3* mutant suggested that MAIN would be required for the expression of several genes that are not controlled by the DRM2 and CMT3 pathway (S2E Fig). To further study the role of MAIN and MAIL1 in the regulation of gene expression and TE silencing, we performed two independent RNA-seq experiments in the *main-2* and *mail1-1* null mutants (RNA-seq Exp1 and Exp3), and combined these experiments with the reanalysis of previously published RNA-seq datasets (RNA-seq Exp2) [15]. Principal component analyses (PCA) showed that for each RNA-seq experiment, *main-2* and *mail1-1* mutant samples tend to cluster together, and away from the WT samples (S3A Fig). Analyzing these three RNA-seq experiments together allowed to identify large numbers of genes and TEs that were misregulated in the *main-2* and *mail1-1* null mutants (Fig3A and B, and S2 Table).

We then compared the transcriptomes of *main-2* and *mail1-1* mutants, together with the *main-3* mutant allele (Fig3A and B, S1 and S2 Tables). As expected by the fact that *main-2* and *mail1-1* are null mutants while *main-3* is a hypomorphic mutant allele, we identified greater numbers of misregulated loci in *main-2* and *mail1-1* in comparison to *main-3* (Fig3A and B). Fractions of these loci were specifically misregulated in each mutant background (Fig 3C and D). In addition, we identified subsets of genes and TEs that were only misregulated in *main-2* and *mail1-1* null mutants, but not in the hypomorphic *main-3* mutant (Fig 3C and D, and S3 Table). Finally, these analyses revealed subsets of loci that were commonly misregulated in the three mutant backgrounds (Fig 3C and D, S3B-D Fig and S3 Table).

The biggest overlaps between misregulated loci in *main-2*, *mail1-1* and *main-3* mutants were among the downregulated genes and upregulated TEs, whereas only a small proportion of genes commonly upregulated in *main-2* and *mail1-1* were also upregulated in *main-3* (Fig 3D). As observed in *main-3* (S2F Fig), upregulated TEs in *main-2* and *mail1-1* were enriched in pericentromeric regions, and genes that were downregulated in *main-2* and *mail1-1* were not targeted by DNA methylation, and mostly located in the chromosome arms (Fig 3E). However, unlike in *main-3*, the upregulated genes in *main-2* and *mail1-1* were not enriched in pericentromeric regions, and only small fractions of them were DNA-methylated genes (Fig 3E). This discrepancy can be explained by the fact that *main-2*

and *mail1-1* null mutations have a much greater impact on the misregulation of gene expression than the hypomorphic *main-3* mutant allele.

Finally, we compared the sets of misregulated loci in *main-2*, *mail1-1*, *ddc* and *ddc main-3* (S1 and S2 Tables). We found significant overlaps among upregulated genes and TEs between *main-2*, *mail1-1*, *ddc* and *ddc main-3* (S3E Fig). This suggests that MAIN, MAIL1, DRM2 and CMT3 cooperate to silence these subsets of genes and TEs. However, we could not find significant overlaps among downregulated genes between *main-2*, *mail1-1* and *ddc* (S3E Fig). Instead, a significant overlap was identified only by comparing the lists of downregulated genes in *main-2*, *mail1-1* and *ddc main-3*, three genetic backgrounds carrying a mutation in either *MAIN* or *MAIL1* (S3E Fig). Thus, this suggests that MAIN and MAIL1 are required for the expression of specific genes, in a DRM2- and CMT3-independent manner.

In conclusion, these comparative analyses allowed to precisely define the loci that were misregulated in *main-2* and *mail1-1* in comparison to *main-3*, *ddc* and *ddcmain-3* mutants. Among these loci, several TEs and DNA-methylated genes are commonly targeted by the MAIN, MAIL1, DRM2 and CMT3 pathways, which suggests that MAIN, MAIL1 and DNA methylation pathways cooperate to silence these TEs and DNA-methylated genes. Besides, several genes are downregulated in *main-2* and *mail1-1*, and subsets of these genes are also downregulated in *main-3*, and *ddcmain-3* but not in *ddc*. This suggests that the MAIN and MAIL1 act independently of DRM2 and CMT3 to ensure the expression of these genes. Finally, these results revealed important overlaps between the misregulated loci in *main-2* and *mail1-1* null mutants, which strongly suggests that the two proteins act in the same pathway to regulate the expression of common sets of loci.

Slight increase in non-CG methylation in the *main-2* mutant does not correlate with changes in gene expression and TE silencing defect.

Whole genome bisulfite sequencing (BS-seq) analyses showed that, at the chromosome scale, DNA methylation level is mostly unchanged in *main-2* in comparison to WT, with the exception of a

slight increase in CHG methylation in pericentromeric regions (Fig 4A). Subtle but statistically significant CHG hypermethylation was further confirmed in pericentromeric TEs and genes, which are mostly TE genes (Fig 4B and C). Slight CHG and CHH hypermethylation was also detected in TEs located in chromosome arms (Fig 4D). Conversely, genes located in chromosome arms did not show significant changes in DNA methylation level in *main-2* (Fig 4E). Identical results were obtained by analyzing the DNA methylation level at upregulated TEs and misregulated genes in *main-2* (Fig 4F-H). We then analyzed the DNA methylation level at genomic locations previously defined as differentially hypomethylated regions (hypo DMRs) at CHG and CHH sites in *cmt3* and *drm1drm2* (*dd*) mutants, respectively [26]. The *cmt3* and *dd* hypo DMRs are mostly located in TEs. As observed with pericentromeric genes and all TEs (Fig 4B-D), we found slight increases in CHG and CHH methylation at *cmt3* and *dd* hypo DMRs, respectively, in *main-2* (S4A and S4B Fig). Finally, DMR calling in *main-2* using stringent parameters only identified a few DMRs (S4C Fig). Thus, DNA methylation is mostly unaffected in *main-2*, with the exception of a slight increase in non-CG methylation at pericentromeric genes and all TEs. Moreover, this subtle non-CG hypermethylation does not correlated with changes in gene and TE expression observed in *main-2* because DNA methylation level in *main-2* is unchanged at these misregulated loci (Fig 4F-H).

MAIN, MAIL1 and the metallo-phosphatase PP7L physically interact together.

The *main-2* and *mail1-1* null mutants display similar molecular and developmental phenotypes (Fig 3 and Fig 5A). Thus, we hypothesized that MAIN and MAIL1 proteins may act in the same pathway, possibly by interacting together. To test this hypothesis, we generated transgenic lines expressing FLAG- and MYC-tagged genomic PMD versions driven by their endogenous promoters. We confirmed that epitope-tagged MAIN and MAIL1 proteins were produced at the expected sizes, and they could complement the respective developmental phenotypes of null mutant plants (Fig 5A and B). Importantly, they could also efficiently rescue the TE silencing and gene expression defects observed in *main-2* and *mail1-1* mutants, implying that epitope-tagged MAIN and MAIL1 are

functional proteins (Fig 5C-E). Using FLAG-tagged MAIN and MAIL1 expressing plants, immunoprecipitation followed by mass spectrometry (IP-MS) analyses were carried out to determine potential protein interactors. Mass spectrometry (MS) analyses indicated that MAIL1 was strongly immunoprecipitated with MAIN-FLAG and *vice versa* (Fig 5F). To validate IP-MS results, we crossed the MAIN-FLAG and MAIL1-MYC lines together. We then performed co-immunoprecipitation (co-IP) experiments using F1 hybrid plants co-expressing the two transgenes, and confirmed that MAIN and MAIL1 interact together (Fig 5G). MS analyses of MAIN-FLAG and MAIL1-FLAG IP also identified the metallo-phosphatase PP7L as putative interactor (Fig 5F). MAIN, MAIL1 and PP7L were the only three proteins reproducibly enriched across multiple replicates (Fig 5F). Co-IP experiments using plants co-expressing either PP7L-FLAG together with MAIN-MYC or MAIL1-MYC constructs confirmed the interaction between PP7L and each PMD protein (Fig 5H and I). Thus, the three proteins MAIN, MAIL1 and PP7L physically interact together.

The *main*, *mail1* and *pp7l* mutants display similar developmental and molecular phenotypes.

PP7L is a putative metallo-phosphatase that was recently identified as a nuclear protein required for photosynthesis [20, 25]. The *pp7l-2* null mutant displays abnormal developmental phenotype reminiscent of *main-2* and *mail1-1* mutant plants, and 3-week-old *mail1-1 pp7l-2* double mutant plants do not show exacerbation of this phenotype (Fig 6A). To determine the genetic interaction between PMD and PP7L, we compared the transcriptomes of *main-2*, *mail1-1*, *pp7l-2* single and *mail1-1 pp7l-2* double mutants (S5A Fig, and S2 and S4 Tables). We identified large numbers of misregulated loci in *pp7l-2* and *mail1-1 pp7l-2* (S5B-C Fig). As observed in *main-2* and *mail1-1*, TEs upregulated in *pp7l-2* and *mail1-1 pp7l-2* were enriched in pericentromeric regions, while up- and downregulated genes were mostly located in the chromosome arms, and not targeted by DNA-methylation (S5D Fig).

Comparative analyses revealed that significant proportions of loci were commonly misregulated in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2* mutants, which is consistent with the fact

280 that MAIN, MAIL1 and PP7L interact together to possibly regulate gene expression and silence TEs (Fig
281 6B-D and S5 Table). These analyses also identified loci that were specifically misregulated in *main-2*,
282 *mail1-1* or *pp7l-2*, which suggests that each protein is independently required for the proper
283 expression of subsets of loci (Fig 6B-C). Besides, these analyses revealed loci that were exclusively
284 misregulated in the *mail1-1 pp7l-2* double mutant, which implies that PP7L and MAIL1 may act
285 redundantly to ensure the proper expression of these loci (Fig 6B-C). Further analyses showed that,
286 among the loci that were misregulated in *mail1-1 pp7l-2*, upregulated genes were significantly more
287 expressed in the double mutant than in each single mutant, and upregulated TEs were significantly
288 differentially expressed only between *mail1-1 pp7l-2* and *pp7l-2* mutants (Fig 6E-F). Conversely, there
289 was no significant difference of expression between the double mutant and single mutants for the
290 downregulated genes (Fig 6G). Thus, these analyses suggest that combining the *pp7l-2* and *mail1-1*
291 mutations may lead to synergistic defects mostly at genes that are upregulated in the double mutant.

292 We then performed in silico analyses to identify enriched DNA motif within a 1kb promoter
293 region upstream of start codon of genes that were up- or downregulated in the different mutant
294 backgrounds. We could not detect any enrichment of a DNA motif among any lists of upregulated
295 genes (including overlapping lists). Likewise, we could not identify a DNA motif enriched in the lists of
296 downregulated genes in *pp7l-2* or *ddc*. However, we identified a discrete DNA motif (hereafter called
297 'DOWN' motif) that was partially enriched in the promoter of genes that were downregulated in *main-*
298 *2*, *mail1-1* and *mail1-1 pp7l-2* mutants (S5E Fig). The *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2* null
299 mutants display strong developmental phenotype, and large numbers of misregulated loci (Fig 6A).
300 Therefore, it is likely that some of the gene misregulation observed in these mutants might be due to
301 side effects of the mutations. To overcome this issue and refine our analysis, we investigated the
302 proportion of the 'DOWN' motif among downregulated genes in the hypomorphic *main-3* and *ddc*
303 *main-3* mutants, as well as in the different overlapping lists of genes commonly downregulated (S3
304 and S5 Tables). The 'DOWN' motif was strongly enriched among the downregulated genes in *main-3*,
305 and to a lesser extent in *ddc main-3* (S5E Fig). It was also significantly enriched in the overlapping lists

of commonly downregulated genes in *main-2*, *mail1-1* and *main-3* as well as in the *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2* overlap (S5E Fig). It was further enriched in the promoters of genes commonly downregulated in all the mutant backgrounds - except *ddc* - analyzed in this study: twenty-five out of twenty-six genes, 96% of enrichment (S5E Fig, S6 and S7 Tables). We analyzed the DNA methylation level of the 'DOWN' motif in the promoters of these twenty-five genes in WT and *main-2*, and found that this DNA motif was not targeted by DNA methylation. Besides, further analyses showed that only a small fraction of all *Arabidopsis* genes carried the 'DOWN' motif in their promoter (12,46%, S5E Fig). Finally, random test analyses based on twenty-six randomly picked genes strongly suggested that the enrichment of the 'DOWN' motif in the promoter of downregulated genes was substantial (S7 Table).

Thus, altogether, these analyses showed that MAIN, MAIL1 and PP7L are equally required for the repression of several genes and TEs. The three proteins are also required for the proper expression of a common set of genes that are downregulated in each single mutant as well as in *mail1-1 pp7l-2* double mutant, and significant fractions of these downregulated genes carry the 'DOWN' DNA motif in their promoter. Furthermore, the 'DOWN' DNA motif is strongly enriched among the genes that are always identified as downregulated in every mutant background carrying the *main-2*, *mail1-1*, *pp7l-2* or *main-3* mutant alleles. This suggests that transcriptional activation of this subset of loci equally requires MAIN, MAIL1 and PP7L activity, and possibly the recognition of the 'DOWN' DNA motif.

PP7L is not required for heterochromatin condensation.

WT *Arabidopsis* nuclei at interphase exhibit condensed DNA foci called chromocenters that are composed of constitutive heterochromatin, and are enriched in H3K9me2 [27]. In several epigenetic mutants, decondensation of constitutive heterochromatin correlates with disruption of chromocenters, and loss or diffusion of H3K9me2 in the nucleoplasm [27]. Thus, analyzing H3K9me2 subnuclear distribution by immunofluorescence (IF) experiments has been reproducibly used as a cytological approach to assay for heterochromatin decondensation [12, 27, 28]. A previous study

showed that subnuclear distributions of chromocenters and H3K9me2 were unchanged in *main-2* and *mail1-1* mutants [15]. However, fluorescent in situ hybridization (FISH) experiments using a DNA probe for the 106B pericentromeric repeats suggested that heterochromatin was decondensed in the *main-2* and *mail1-1* in comparison to WT plants [15]. We performed IF experiments to analyze the subnuclear distribution of H3K9me2 in the *pp7l-2* mutant. These analyses did not show any change in the condensation level of chromocenters in *pp7l-2* nuclei in comparison to WT (Fig 7). Instead, we observed that *pp7l-2* nuclei were proportionally more condensed than WT nuclei (Fig 7). This is likely due to the fact that *pp7l-2* mutant displays abnormal phenotype and growth delay in comparison to WT plants that are entering the floral transition stage, a developmental stage where partial chromocenter decondensation has been documented [29]. In conclusion, based on the H3K9me2 IF experiments, we can conclude that *pp7l-2* is not impaired in chromocenter condensation.

The PMD and PP7 domains have co-evolved among the Eudicots.

Among the Angiosperms, most of the genic PMDs, like MAIN and MAIL1, are standalone versions [18]. However, some genic PMDs can associate with other protein domains, such as for instance a PPP domain. In *A. thaliana*, the protein MAIL3, which carries a PMD fused to a PPP domain, is a close homolog of both MAIN/MAIL1 and PP7/PP7L through its PMD and PPP domains, respectively. Considering that the PMD proteins MAIN and MAIL1 interact with PP7L, and are required for the expression of similar set of loci, we decided to determine the distribution of related genic PMD and PPP domains, and to retrace their evolutionary history among plant species. The *A. thaliana* MAIN, MAIL1 and MAIL3 are all members of the PMD-C family that also includes MAIL2 [15]. Since our objective is to retrace the evolution of genic (and not TE-containing) PMD-C, we have decided to restraint our search to Eudicots. Indeed, Eudicot species contain mainly genic PMD-C, while other angiosperms may contain variable numbers of closely related genic and TE-associated PMD-C motifs that would be difficult to distinguish in our analysis. To retrace the evolution history of the genic PMD-C family, we used *A. thaliana* PMD-C genes to search and collect their relatives (paralogues and

orthologues) in 30 genomes representative of the Eudicot diversity (see S8 Table for a list of species and their corresponding codes used in Fig 8, and S9 Table for motif sequences).

In our phylogenetic analysis, the genic PMD-C family can be clearly separated in two major clades. The first clade is composed of orthologues of *A. thaliana* MAIL2, MAIL1 and MAIN, while the second one includes orthologues of *A. thaliana* MAIL3 (Fig 8A). MAIL2 orthologues were found in all species tested, forming a closely related group, which suggests that they are under strong purifying selection (see the very short branch lengths linking most MAIL2 genes in Fig 8A). In several species, additional MAIL2 paralogues were also detected. They were either imbedded in the major MAIL2 group, or forming independent and more divergent subgroups, like in the case of MAIL1 and MAIN that are Brassicaceae-specific MAIL2 paralogues. By comparison, MAIL3 orthologues were not found in all Eudicot species tested, and, except in Brassicaceae, MAIL3 genes appear to be under much weaker purifying selection compare to MAIL2 and MAIL2-like genes (see the longer branch lengths in the tree of Fig 8A). Brassicaceae MAIL3 genes contrast with other MAIL3, by forming a closely related group in the phylogenetic tree. This suggests a clear change in selection pressure, typical of a neofunctionalization event that could correlate with the acquisition of the PPP motif by these genes (Fig 8B and see below). Remarkably, another fusion event between PMD-C and PPP motifs occurred independently in grapevine, but this time involving a MAIL2 paralogue (VvMAIL2.2, Fig 8A).

We then used the PPP motif found in *A. thaliana* MAIL3, to collect orthologous genes and retrace the evolution history of this motif in the same Eudicot species used above. We confirmed that these genes can be clearly separated in two distinct clades: PP7 and PP7-like (PP7L) (Fig 8B). All tested species present one or several closely related PP7 paralogues. Although the Brassicaceae MAIL3 PPP motif belongs to the PP7 clade, it diverged significantly compared to other standalone PP7 paralogues (Fig 8B). Same observation was made regarding the PP7 domain of VvMAIL2.2. Thus, as described for the PMD of Brassicaceae MAIL3 and grapevine VvMAIL2.2, this suggests a fast-evolving period and neofunctionalization of the PP7 domain in these species, subsequently to the PMD-C/PP7 fusion. Conversely, PP7L orthologues were not found in all species tested and, accordingly, these genes are

under weaker purifying selection compare to genes belonging to the PP7 subfamily. In conclusion, phylogenetic analyses showed that, in at least Brassicaceae and grapevine, neo-association of PMD-C and PP7 domains have potentially create new protein functions that were maintained through evolution.

DISCUSSION

In *A. thaliana*, MAIN and MAIL1 are standalone PMD proteins that have been involved in genome integrity, regulation of cell division and differentiation, and silencing of TEs [15-17]. In this study, we show that TE silencing is widely impaired in the *ddc main-3* higher order mutant, which is both partially defective in DNA methylation and MAIN activity. We also identify the putative phosphatase protein PP7L as MAIN and MAIL1 protein interactor, and show that among the loci that are commonly misregulated in *pmd* and *pp7l* single and double mutants, a substantial fraction of downregulated genes carries the 'DOWN' DNA motif in their promoter. Finally, phylogenetic analyses among Eudicots suggest a mechanism of neofunctionalization between the PMD and PP7-type PPP, to potentially acquire a functional module that requires the two protein domains.

The PMD MAIN protein acts independently of DRM2- and CMT3 pathways to silence TEs and DNA-methylated genes.

Previous analyses showed that some TEs were synergistically upregulated in the *mail1 rdr2* double mutant plants, suggesting that MAIL1 acts independently of RdDM pathway [15]. In our whole genome transcriptomic analyses, we show that several TEs and DNA-methylated genes are upregulated in both *main-3* and *ddc* mutants, as well as in the *ddc main-3* quadruple mutant (Fig 2 and S2 Fig). We also identify TEs that are upregulated in either *ddc* or *main-3* mutants, but display stronger misregulation in the *ddc main-3* higher order mutant (Fig 2 and S2 Fig). Finally, we identify a large class of TEs that are only upregulated in *ddc main-3* (Fig 2 and S2 Fig). Altogether, these analyses reveal complex genetic interaction between the MAIN, DRM2 and CMT3 proteins to silence TE.

Previous work showed that DNA methylation is not impaired in *mail1-1* [15]. We found that DNA methylation is mostly unaffected in the *main-2* null mutant. However, we detected a mild but significant hypermethylation at non-CG sites in TEs and pericentromeric genes (Fig 4). One hypothesis is that CHG and CHH hypermethylation observed in *main-2* is a backup mechanism to compensate for MAIN loss of function, and to dampen TE silencing defects. Although further studies will be required to test this hypothesis, it is consistent with the fact that combining the *main-3* and *ddc* mutations leads to an exacerbation of TE silencing defects. Thus MAIN, DRM2 and CMT3 pathways cooperate to silence TE. Synergistic effects between different epigenetic pathways have already been described. For instance, it has been shown that MORPHEUS MOLECULE 1 (MOM1) and MORC1/MORC6 proteins, or MOM1 and the RdDM pathway act synergistically to efficiently silence TEs [13, 30]. Altogether, these observations contribute to the “mille-feuille” (i.e. “multiple layers”) model, in which different epigenetic pathways converge towards the silencing of TEs [31].

The putative phosphatase PP7L interacts with the PMD MAIN and MAIL1 protein to regulate a similar set of genes and TEs.

Recently, the putative phosphoprotein phosphatase PP7L was involved in the biogenesis of chloroplasts and plant response upon abiotic stress [25]. Here, we show that PP7L interact with MAIN and MAIL1, and *main-2*, *pp7l-2*, *mail1-1* single and *mail1-1 pp7l-2* double mutant plants display similar developmental and molecular phenotypes (Fig 5 and 6). We also show that, as described for *main-2* and *mail1-1* [15], the subnuclear distribution of chromocenters and H3K9me2 are unaltered in *pp7l-2* (Fig 7). The 106B pericentromeric repeats appeared decondensed in *main-2* and *mail1-1* mutants [15], future work will determine if similar phenotype is observed in *pp7l-2*. Although MAIN, MAIL1 and PP7L interact together, we cannot exclude that an additional protein is required for the interaction. In addition, PP7L may have additional partners independently of MAIN and MAIL1. Further biochemical studies such as IP-MS analyses using the FLAG-tagged PP7L line will contribute to addressing these points.

Transcriptomic analyses revealed complex genetic interaction between MAIN, MAIL1 and PP7L; the three proteins acting either independently or together to ensure the proper expression of genes, and to perform TE silencing. Moreover, transcriptome profiling of *mail1-1 pp7l-2* double mutant revealed that the two mutations may have synergistic effects, specifically at genes that are upregulated in the mutant. To further study the genetic interaction between the three proteins, it will be important to analyze the transcriptome of *main-2 mail1-1 pp7l-2* triple mutant. Altogether and considering that *i)* MAIN, DRM2 and CMT3 pathways cooperate to silence TEs, and *ii)* the *main-2* mutant show a slight increase in DNA methylation at CHG and CHH sites, we cannot rule out that MAIN is playing a dual role: regulating gene expression through its interaction with MAIL1 and PP7L, and involved in TE silencing through its genetic interaction with DNA methylation. In the future, it will be important to analyze DNA methylation in *pp7l-2*, but also in *pmd pp7l-2* higher order mutants. In parallel, studying the *ddc pp7l-2* mutant will allow to further decipher the genetic interaction between the PP7L and DNA methylation pathways.

A fraction of genes that are commonly downregulated in *main*, *mail1* and *pp7l* mutants carry the ‘DOWN’ motif in their promoters.

A substantial fraction of genes that are commonly downregulated in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2* carry the ‘DOWN’ motif in their promoter (S5E Fig and S7 Table). Furthermore, twenty-five out of twenty-six genes commonly downregulated in the all the mutant backgrounds analyzed in this study - except *ddc* - carry the ‘DOWN’ DNA motif in their promoter (S5E Fig and S7 Table). The ‘DOWN’ motif is also enriched in fractions of downregulated genes in *main-2*, *mail1-1*, *mail1-1 pp7l-2*, *main-3* and *ddc main-3*. However, it is not enriched among downregulated genes in *pp7l-2* mutant. One explanation for this discrepancy is that too many loci were identified as downregulated in *pp7l-2*, which created a dilution of the loci carrying the ‘DOWN’ motif in their promoter.

Based on our results, we hypothesize that the 'DOWN' motif may act as a putative cis-regulatory element (CRE) recognized by an unidentified TF, which would be required for the transcription of genes identified as downregulated in *pmd* and *pp7l* mutants. This unknown TF could be recruited or activated by the MAIN/MAIL1/PP7L protein complex. Another hypothesis is that the 'DOWN' motif is directly recognized by the MAIN/MAIL1/PP7L protein complex. Further study will be required to test if MAIN/MAIL1/PP7L protein complex interact with chromatin, and bind the 'DOWN' motif. In parallel, further biochemical analyses may allow to identify an uncharacterized putative TF as MAIN/MAIL1/PP7L protein interactor.

Altogether, these analyses suggest that MAIN, MAIL1 and PP7L are involved in three distinct activities. First, they are required for the silencing of TEs and DNA-methylated genes, cooperating with canonical epigenetic factors such as DRM2 and CMT3 to efficiently repress these loci. Second, they are required for the repression of subsets of genes that are not targeted by DNA methylation. For this category of loci, one hypothesis is that MAIN, MAIL1 and PP7L may act as transcriptional repressor. Third, MAIN, MAIL1 and PP7L are required for the transcriptional activation of several genes, and fractions of those genes carry the 'DOWN' motif in their promoter. In the future, it will be important to determine the molecular mechanisms that are involved in these three activities of MAIN, MAIL1 and PP7L.

The association of PMD-C and PP7/PP7L domains creates a functional protein module.

In this study, we identified PP7L has a protein partner of the two standalone PMDs MAIN and MAIL1, and showed that these proteins are required for the proper expression of a common set of genes, and for TE silencing. Besides, we showed that the Brassicaceae MAIL3 and the grapevine VvMAIL2.2 proteins carry a PMD fused to a PP7 domain. Based on these results, we hypothesize that depending on the configuration, the association of PMD-C and PP7/PP7L domains would create a functional protein module in trans or in cis. It is likely that the cis-association of PMD and PP7 found in the Brassicaceae MAIL3 proteins occurred in the common ancestors of this Eudicot lineage, possibly

through the process of gene duplication. Since then, the MAIL3 PMD/PP7 fusion was maintained under strong purifying selection, arguing for a neofunctionalization of the fusion protein. It is likely that a similar process happened in grapevine, and possibly, in closely related Vitaceae species. To some extent, the two distinct events that occurred in Brassicaceae and grapevine are reminiscent of convergent evolution processes leading to the production of a functional PMD/PP7 module.

The occurrence of PMD and PP7/PP7L protein fusion in several Brassicaceae and grapevine is reminiscent of the concept of Rosetta stone chimera proteins, which describes that two proteins interacting together in one organism can be found fused together in another species to facilitate enzymatic activity [32]. There are several examples of Rosetta stone proteins, described for instance with different subunits of DNA topoisomerase or RNA polymerase [32]. Here, we show that, at least in *A. thaliana*, the Rosetta stone chimera MAIL3 coexist with its close homologs MAIN/MAIL1 and PP7L that interact together. The fact that the PMD and PP7 domains are fused together in MAIL3 may be a strategy to optimize protein activity. Conversely, the enzymatic activity of the MAIN/MAIL1/PP7L protein complex could be further regulated by allowing, or not, the three proteins to interact together. Nevertheless, in both scenarios, it is likely that PMD and PP7/PP7L association creates a functional protein module, which might be specialized in distinct biological processes depending on its composition. Thus, we hypothesize that the MAIL3 and MAIN/MAIL1/PP7L protein complexes play different role in the plant. This is consistent with the fact that, unlike *main-2*, *mail1-1* and *pp7l-2* mutant, the *mail3-2* mutant does not show abnormal developmental phenotype [17]. Further studies will be required to describe the role of MAIL3 in the plants.

In conclusion, we show here that the two *A. thaliana* PMD MAIN and MAIL1 proteins interact with PP7L, and are involved in the regulation of a common set of genes and TEs. In addition, we show that distinct events of PMD-C and PP7 fusions have occurred among the Eudicots (among several Brassicaceae species and in grapevine), suggesting some convergent evolution processes and a potential neofunctionalization of PMD/PP7 module in cis. The biological significance of PMD/PP7 fusion proteins will be investigated in the future by studying the role of MAIL3 in *A. thaliana*. In

addition, it will be important to determine whether the PMD proteins play important roles in other plant species with agronomic value.

MATERIALS AND METHODS

Plant material and growing conditions. All the plant material is in the Columbia (Col) ecotype. Col= Non-transgenic WT Columbia ecotype. The *drm1-2* (SALK_031705), *drm2-2* (SALK_150863), *cmt3-11* (SALK_148381), *ddc* triple, *main-2* (GK-728H05), *mail1-1* (GK-840E05) and *pp7l-2* (SALK_003071) null mutant lines were previously described [15-17, 25, 26], and obtained from The Nottingham Arabidopsis Stock Centre. The *mail1-1 pp7l-2* double mutant was obtained by crossing the respective single mutants. T-DNA insertions were confirmed by PCR-based genotyping and RT-qPCR analyses. The *ATCOPIA28::GFP* WT line (WT) carries the transgene in WT Col ecotype. The *ATCOPIA28::GFP ddc* line (*ddc*) carries the transgene in *ddc*. The *ATCOPIA28::GFP ddc main-3* line (*ddc main-3*= *ddc* #16) carries the transgene in the *ddc main-3* background. The *ATCOPIA28::GFP main-3* line (*main-3*) was obtained by backcrossing *ddc main-3* with WT, F1 plants were self-fertilized, and F2 plants were screened by PCR-based genotyping to identify plants homozygote for the *main-3* mutation and WT for *DRM2* and *CMT3*. The *main-3* mutant allele was scored by derived cleaved amplified polymorphic sequences (dCAPS) using the restriction enzyme FokI. Primer sequences are described in S10 Table. All the WT Col and T-DNA mutant plants were grown on soil under a 16h-light/8h-dark cycle. When experiments required to screen for GFP expression under UV light, plants carrying the *ATCOPIA28::GFP* transgene were first grown on Murashige and Skoog (MS) plates under continuous light, 10-day old plants were then screened for GFP expression under UV light, and subsequently transferred onto soil. For *in vitro* plant culture, seeds were surface-sterilized and sowed on solid MS medium containing 0.5% sucrose (w/v).

538 **Cloning of ATCOPIA28::GFP.** The pCambia3300-NLS-GFP-T35S vector was previously described [12].
539 The 5'LTR promoter corresponding to a region of ~1 kb upstream of *ATCOPIA28* (*AT3TE51900*) was
540 PCR amplified from WT genomic DNA, and cloned into pCR2.1 TOPO vector (Invitrogen). Quikchange
541 site-directed mutagenesis (Stratagene) was performed according to Manufacturer's instruction to
542 create a polymorphism site (MfeI→NdeI) within the 5'LTR promoter, which was subsequently
543 mobilized into pCambia3300 upstream of NLS-GFP-T35S sequence. *ddc* triple mutant plants were
544 transformed with the *ATCOPIA28::GFP* construct using the *Agrobacterium*-mediated floral dip
545 method [33]. Transgenic plants showing GFP fluorescence were backcrossed with a WT plant to
546 promote the silencing of *ATCOPIA28::GFP* in the F1 generation. F1 plants were self-crossed and their
547 F2 progenies were screened for GFP fluorescence, and PCR-based genotyped to obtain
548 *ATCOPIA28::GFP* WT and *ATCOPIA28::GFP ddc* plants. Primer sequences used for *ATCOPIA28::GFP*
549 cloning and PCR genotyping are described in S10 Table.

550

551 **EMS mutagenesis, GFP screening and mapping analyses.** Five thousand seeds of *ATCOPIA28::GFP ddc*
552 were mutagenized in 0.26% EMS solution for 12 hours with rotation. Seeds were subsequently washed
553 with water and sown on soil. Fifteen hundred M2 populations were collected, and subsequently
554 screened for GFP fluorescence under UV light using a SMZ18 Nikon Fluorescence Stereomicroscope
555 coupled with the C-HGFI intensilight fluorescence filter. Pictures were taken using the DS Qi1MC digital
556 camera kit. Mapping and identification of the EMS mutation responsible for the phenotype were
557 performed by bulk segregant analysis coupled with deep genome re-sequencing as previously
558 described [12], with the following differences. Reads were mapped against the reference genome
559 (*Arabidopsis* TAIR10) and single nucleotide polymorphisms called in Geneious (Biomatters). Using R,
560 single nucleotide polymorphisms were filtered for EMS mutations (G:C→A:T) and zygosity called
561 based on the variant frequency provided by Geneious (≥80% homozygous mutation, ≥45%, and ≤55%
562 heterozygous mutation). Plots were then created by calculating the ratio of the number of
563 homozygous and heterozygous and mutations in a 500-kb window as previously described [34].

564

565 **Cloning of epitope-tagged versions of PMD and PP7L proteins.** *MAIN*, *MAIL1* and *PP7L* genomic
566 regions were PCR amplified and FLAG or Myc epitopes were added to the C-terminus of each protein
567 as previously described [12]. Each time, the amplified region includes a ~1Kb promoter sequence
568 upstream of the respective transcriptional start site. For the *MAIN* promoter, a MluI site was modified
569 to allow LR reaction without changing the sequence integrity of the gene. *main-2*, *mail1-1* and *pp7l-2*
570 mutant plants were transformed with the *MAIN-FLAG*, *MAIN-MYC*, *MAIL1-MYC* and *PP7L-*
571 *FLAG* constructs using the *Agrobacterium*-mediated floral dip method [33]. Primer sequences are
572 described in S10 Table.

573

574 **IP and MS analysis.** Ten grams of 3-week-old seedling tissue were ground in liquid nitrogen and
575 resuspended in 50mL ice-cold IP buffer [50mM Tris HCl pH 7.6, 150mM NaCl, 5mM MgCl₂, 0.1%
576 Nonidet P-40, 10% glycerol (v/v), 0.5mM DTT, 1x Protease Inhibitor Mixture (Roche)] and centrifuged
577 2 times for 15 min at 4°C at 15 350g. 400μL of M2 magnetic FLAG-beads (Sigma, M8823) were added
578 to the supernatants, and incubated for 90min rotating at 4°C. M2 magnetic FLAG-beads were washed
579 seven times in ice-cold IP buffer for 5 min rotating at 4°C, and immunoprecipitated proteins were
580 eluted 3 times with 150μL 3x-FLAG peptides (Sigma, F4799) for 25 min each at 25°C. The eluted protein
581 complexes were precipitated by trichloroacetic acid and subjected to MS analyses as previously
582 described [13]. Peptide and protein-level false discovery rates were calculated by the DTASelect
583 algorithm using the decoy database approach. Based on a peptide PSM level p-value filter of less than
584 0.01 and a requirement for at least two peptides per protein, the protein-level false discovery rate
585 was less than 1% for all proteins detected.

586

587 **Co-IP and immunoblotting.** 0.5 g of 3-week-old seedling tissue were ground in liquid nitrogen,
588 resuspended in 1.5mL ice-cold IP buffer [50mM Tris pH 7.6, 150mM NaCl, 5mM MgCl₂, 0.1% Nonidet
589 P-40, 10% glycerol, 0.5 mM DTT, 1x Protease Inhibitor Mixture (Roche)], and centrifuged 2 times for

590 15 min at 4°C, 16 000g. 50µL M2 magnetic FLAG-beads (Sigma, M8823) were added to the
591 supernatants and incubated for 2 hour rotating at 4°C. Beads were washed 3 times in ice-cold IP buffer
592 for 10 min rotating at 4°C. Immunoprecipitated proteins were denatured in Laemmli buffer for 5min
593 at 95°C. 10µL of input and bead elution were run on 10% SDS-PAGE gels, and proteins were detected
594 by western blotting using either Anti-FLAG M2 monoclonal antibody-peroxidase conjugate (Sigma,
595 A8592) at a dilution of 1:10000, or c-Myc rat monoclonal antibody (Chromotek, 9E1-100) at a dilution
596 of 1:1000 followed by goat anti-rat IgG horseradish peroxidase (Abcam, ab205720) used at a dilution
597 of 1:20000 as secondary antibody. Western blots were developed using Substrat HRP Immobilon
598 Western (Merck Millipore, WBKLS0500).

599

600 **RNA extraction.** Total RNA was extracted from aerial parts of 3-week-old seedlings grown on soil using
601 either RNeasy Plant Mini Kit (Qiagen, 74904) or Monarch Total RNA Miniprep Kit (NEB, T2010)
602 according to the manufacturer's protocols.

603

604 **RNA sequencing.** RNA-seq libraries were generated from 1µg of input RNA using NEBNext Ultra II
605 Directional RNA Library Prep Kit for Illumina (NEB, E7490) according to the manufacturer's protocols.
606 Libraries were sequenced on an Illumina HiSeq 4000 or NextSeq 550 machines. Reads were trimmed
607 using Trimmomatic [35], and mapped to the *A. thaliana* genome (*Arabidopsis* TAIR10 genome) using
608 HISAT2 [36]. The sequence alignment files were sorted by name and indexed using SAMtools [37].
609 Files were converted to BAM files and number of reads mapped onto a gene calculated using HTSeq-
610 count [38]. Differentially expressed genes were obtained with DESeq2 [39], using a log2 fold-change
611 ≥ 2 (up-regulated genes) or ≤ -2 (down-regulated genes) with an adjusted *p*-value of 0,01. Batch effects
612 were modeled within the DESeq2 study design. For PCA, we removed the batch effect using limma's
613 'removeBatchEffect' function [40]. Heat map visualizations were realized using the heatmap2 function
614 from the R gplots package. Boxplots were realized using boxplot function from R. Re-analyses of

previously published RNA-seq datasets from *main-2* and *mail1-1* (PRJEB15202) [15] were performed as described above.

RT-qPCR. 1 µg of input RNA was converted to cDNA using GoScript Reverse Transcriptase (Promega A501C) according to the manufacturer's protocol. The final reaction was diluted 6 times with RNase free water. RT-qPCR experiments were performed with 4µL of cDNA combined to the Takyon No Rox SYBR MasterMix (Eurogentec, UF-NSMT-B0701), using a LightCycler 480 instrument (Roche). Amplification conditions were as follows: 95°C 5 min; 45 cycles, 95°C 15s, 60°C 15s, 72°C 30s; melting curves. RT-qPCR analyses used the $2^{-\Delta\Delta Ct}$ method. For each analysis, ΔCt was first calculated based on the housekeeping *RHIP1* gene Ct value [41]. $\Delta\Delta Ct$ were then obtained by subtracting the wt ΔCt from the ΔCt of each sample. Values were represented on bar charts relative to WT. Three technical replicates were performed per biological replicate, and 3 biological replicates were used in all experiments, unless otherwise stated. Primer sequences are described in S10 Table.

DNA motif detection. The motifs for enhancer sequences (1kb upstream the TSS) were discovered using MEME (Multiple Em for Motif Elicitation). MEME represents motifs as position-dependent letter-probability matrices which describe the probability of each possible letter at each position in the pattern [42].

Bisulfite sequencing. Genomic DNA was extracted from aerial parts of 3-week-old seedlings using Quick-DNA Plant/Seed Miniprep Kit (Zymo research, D6020) according to the manufacturer's protocol. Whole genome bisulfite sequencing (WGBS) library was prepared from 50 ng genomic DNA using NuGen Ovation Ultralow Methyl-Seq kit. Bisulfite treatment was carried out by Qiagen Epitect bisulfite kit. WGBS libraries were sequenced on an Illumina HiSeq 4000 machine. The raw reads (single end) were trimmed using Trimmomatic in order to remove adapter sequences [35]. The remaining sequences were aligned against the *A. thaliana* genome TAIR10 version using Bismark [43]. Duplicated

reads were collapsed into one read. For metaplot and boxplot visualization, we used ViewBS [44]. Boxplots were realized using boxplot function from R. DMRs (differentially methylated regions) were defined comparing methylation in wildtype with the *main-2* mutant analyzed using the R package “DMRcaller” [45]. We used “noise filter” method to compute CpG, CpHpG and CpHpH DMRs. We selected bins where the p-value was less than 0.01, the difference in methylation level was at least 40% in the CG context, 20% in the CHG context or 10% in the CHH context, with at least four cytosines; each cytosine had on average at least four reads.

Sequence selection, multiple sequences alignments and phylogenetic reconstruction.

Blast searches (blastp) were performed starting from known *A. thaliana* PMD-C and PP7/PP7L motifs on the thirty species representing the diversity of the Eudicot lineages. When necessary tblastn searches were also used to obtain complete protein sequences. To build the phylogenetic trees, PMD-C or PP7/PP7L motifs were aligned using the multiple sequence comparison by log-expectation (MUSCLE v3.7) software [46]. Trees were reconstructed using the fast-maximum likelihood tree estimation program PHYML [47] using the LG amino acids replacement matrix [48]. Statistical support for the major clusters were obtained using the approximate likelihood-ratio test (aLRT) [49].

Immunofluorescence and DAPI-staining. Leaves from 3-week-old plants, were fixed for 20 min rotating at 4°C in 2% formaldehyde in Tris buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 100 mM NaCl), washed two times for 10 min rotating at 4°C in cold Tris buffer and subsequently chopped in LB01 buffer (15 mM Tris-HCl pH 7.5, 2 mM EDTA, 0.5 mM spermine, 80 mM KCl, 20mM NaCl and 0.1% Triton- X-100). Nuclei were filtered through a 30 µm cell strainer cap (Sysmex, 04-0042-2316) and 5µl of the nuclei solution was diluted in 10 µl of sorting buffer (100mM Tris-HCl pH 7.5, 50 mM KCl, 2 mM MgCl₂, 0.05% Tween-20 and 5% sucrose). 20µl of the nuclei dilution were spread onto a polylysine slide and air-dried for 40 min. Slides were post-fixed in 2% formaldehyde in 1X PBS for 5 min and washed 2 times with water. Slides were incubated 15 min in 1X PBS, 0.5% Triton X-100 at RT and

washed 3 times with 1X PBS for 5 min. For detection, slides were incubated over night with a mouse anti-H3K9me2 monoclonal antibody (Abcam, Ab 1220) at 1:500 in 3% BSA, 0.05% Tween in 1X PBS at 4°C in a moist chamber. After 3 washes in 1X PBS for 5 min, slides were incubated 2h with a goat anti-mouse antibody coupled to Alexa fluor 568 (Invitrogen, A11004) at 1:1000 in 3% BSA, 0.05% Tween in 1X PBS in a moist chamber. Slides were washed 1 time 5 min with 1X PBS, 1 time 10 min with 1X PBS, 1µg/mL DAPI, and 1 time 5 min with 1X PBS. DNA was counterstained with 1µg/mL DAPI in Vectashield mounting medium (Vector Laboratories). Observation and imaging were performed using a LSM 700 epifluorescence microscope (Zeiss).

Data availability. Nucleotide sequencing data generated in this study have been deposited in European Nucleotide Archive (ENA) under the accession number PRJEB33240 (<http://www.ebi.ac.uk/ena/data/view/PRJEB33240>). The proteomics data have been deposited to the MassIVE data repository (<https://massive.ucsd.edu>) with the dataset identifier MSV000084089. All other data and material are available within the manuscript and its supplementary files, or from the corresponding author upon request.

ACKNOWLEDGMENTS

The authors want to thank Thierry Lagrange, Frederic Pontvianne and other team members for fruitful discussions, and all the LGDP platform members for their outstanding technical assistance and plant care.

REFERENCES

1. Grewal SI, Jia S. Heterochromatin revisited. *Nat Rev Genet.* 2007;8(1):35-46.
2. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007;8(4):272-85.
3. Deniz O, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications. *Nat Rev Genet.* 2019;20(7):417-31.
4. Dergai O, Hernandez N. How to Recruit the Correct RNA Polymerase? Lessons from snRNA Genes. *Trends Genet.* 2019;35(6):457-69.
5. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet.* 2010;11(3):204-20.

6. Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet.* 2014;15(6):394-408.
7. Wendte JM, Pikaard CS. The RNAs of RNA-directed DNA methylation. *Biochim Biophys Acta.* 2016.
8. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol.* 2015;16(9):519-32.
9. Zhang H, Lang Z, Zhu JK. Dynamics and function of DNA methylation in plants. *Nat Rev Mol Cell Biol.* 2018;19(8):489-506.
10. Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell.* 2013;153(1):193-205.
11. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol.* 2014;21(1):64-72.
12. Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, Stroud H, et al. MORC family ATPases required for heterochromatin condensation and gene silencing. *Science.* 2012;336(6087):1448-51.
13. Moissiard G, Bischof S, Husmann D, Pastor WA, Hale CJ, Yen L, et al. Transcriptional gene silencing by Arabidopsis microorchidia homologues involves the formation of heteromers. *Proc Natl Acad Sci U S A.* 2014;111(20):7474-9.
14. Lorkovic ZJ, Naumann U, Matzke AJ, Matzke M. Involvement of a GHKL ATPase in RNA-directed DNA methylation in Arabidopsis thaliana. *Curr Biol.* 2012;22(10):933-8.
15. Ikeda Y, Pelissier T, Bourguet P, Becker C, Pouch-Pelissier MN, Pogorelnik R, et al. Arabidopsis proteins with a transposon-related domain act in gene silencing. *Nat Commun.* 2017;8:15122.
16. Wenig U, Meyer S, Stadler R, Fischer S, Werner D, Lauter A, et al. Identification of MAIN, a factor involved in genome stability in the meristems of Arabidopsis thaliana. *Plant J.* 2013;75(3):469-83.
17. Uhlken C, Horvath B, Stadler R, Sauer N, Weingartner M. MAIN-LIKE1 is a crucial factor for correct cell division and differentiation in Arabidopsis thaliana. *Plant J.* 2014;78(1):107-20.
18. Babu MM, Iyer LM, Balaji S, Aravind L. The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* 2006;34(22):6505-20.
19. Steinbauerova V, Neumann P, Novak P, Macas J. A widespread occurrence of extra open reading frames in plant Ty3/gypsy retrotransposons. *Genetica.* 2011;139(11-12):1543-55.
20. Farkas I, Dombradi V, Miskei M, Szabados L, Koncz C. Arabidopsis PPP family of serine/threonine phosphatases. *Trends Plant Sci.* 2007;12(4):169-76.
21. Sun X, Kang X, Ni M. Hypersensitive to red and blue 1 and its modification by protein phosphatase 7 are implicated in the control of Arabidopsis stomatal aperture. *PLoS Genet.* 2012;8(5):e1002674.
22. Liu HT, Li GL, Chang H, Sun DY, Zhou RG, Li B. Calmodulin-binding protein phosphatase PP7 is involved in thermotolerance in Arabidopsis. *Plant Cell Environ.* 2007;30(2):156-64.
23. Genoud T, Santa Cruz MT, Kulisic T, Sparla F, Fankhauser C, Metraux JP. The protein phosphatase 7 regulates phytochrome signaling in Arabidopsis. *PLoS One.* 2008;3(7):e2699.
24. Uhrig RG, Labandera AM, Moorhead GB. Arabidopsis PPP family of serine/threonine protein phosphatases: many targets but few engines. *Trends Plant Sci.* 2013;18(9):505-13.
25. Xu D, Marino G, Klingl A, Enderle B, Monte E, Kurth J, et al. Extrachloroplastic PP7L Functions in Chloroplast Development and Abiotic Stress Tolerance. *Plant Physiol.* 2019.
26. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell.* 2013;152(1-2):352-64.
27. Fransz P, de Jong H. From nucleosome to chromosome: a dynamic organization of genetic information. *Plant J.* 2011;66(1):4-17.

28. Zhao S, Cheng L, Gao Y, Zhang B, Zheng X, Wang L, et al. Plant HP1 protein ADCP1 links multivalent H3K9 methylation readout to heterochromatin formation. *Cell Res.* 2019;29(1):54-66.
29. Tessadori F, Schulkes RK, van Driel R, Fransz P. Light-regulated large-scale reorganization of chromatin during the floral transition in *Arabidopsis*. *Plant J.* 2007;50(5):848-57.
30. Yokthongwattana C, Bucher E, Caikovski M, Vaillant I, Nicolet J, Mittelsten Scheid O, et al. MOM1 and Pol-IV/V interactions regulate the intensity and specificity of transcriptional gene silencing. *Embo J.* 2009;28(2):340-51.
31. Rigal M, Mathieu O. A "mille-feuille" of silencing: epigenetic control of transposable elements. *Biochim Biophys Acta.* 2011;1809(8):452-8.
32. Galperin MY, Koonin EV. Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol.* 2000;18(6):609-13.
33. Clough SJ, Bent AF. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 1998;16(6):735-43.
34. Hristova E, Fal K, Klemme L, Windels D, Bucher E. HISTONE DEACETYLASE6 Controls Gene Expression Patterning and DNA Methylation-Independent Euchromatic Silencing. *Plant Physiol.* 2015;168(4):1298-308.
35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-20.
36. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357-60.
37. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
38. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31(2):166-9.
39. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.
41. Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* 2005;139(1):5-17.
42. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37(Web Server issue):W202-8.
43. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics.* 2011;27(11):1571-2.
44. Huang X, Zhang S, Li K, Thimmapuram J, Xie S, Wren J. ViewBS: a powerful toolkit for visualization of high-throughput bisulfite sequencing data. *Bioinformatics.* 2018;34(4):708-9.
45. Catoni M, Tsang JM, Greco AP, Zabet NR. DMRcaller: a versatile R/Bioconductor package for detection and visualization of differentially methylated regions in CpG and non-CpG contexts. *Nucleic Acids Res.* 2018;46(19):e114.
46. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792-7.
47. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 2003;52(5):696-704.
48. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol.* 2008;25(7):1307-20.
49. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol.* 2006;55(4):539-52.
50. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS ONE.* 2008;3(9):e3156.

51. Vergara Z, Gutierrez C. Emerging roles of chromatin in the maintenance of genome organization and function in plants. *Genome Biol.* 2017;18(1):96.

FIGURE CAPTIONS

Fig 1. The *ddc* #16 EMS population shows overexpression of *ATCOPIA28::GFP* and upregulation of endogenous TEs.

(A) Schematic representation of the *ATCOPIA28::GFP* transgene. The 5' long terminal repeat (LTR) promoter region of an *ATCOPIA28* LTR-retrotransposon (*AT3TE51900*) is used to control the expression of GFP. The construct carries a Nuclear Localization Signal (NLS) to target the GFP in the nucleus. (B) WT and *drm1 drm2 cmt3* (*ddc*) triple mutant plants carrying the *ATCOPIA28::GFP* transgene showed no and weak GFP fluorescence under UV light, respectively. By comparison, the *ddc* #16 EMS mutant showed strong GFP fluorescence. Insets show plants under white light. (C) Western blot using anti-GFP antibody confirmed *ATCOPIA28::GFP* overexpression in *ddc* #16. Coomassie staining of the large Rubisco subunit (rbcl) is used as a loading control. KDa: kilodalton. (D) Relative expression analyses of *ATCOPIA28::GFP* (*GFP*) and three endogenous TEs in *ddc* and *ddc* #16 assayed by Real-Time quantitative PCR (RT-qPCR). RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the mutants are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. Screening of EMS mutant populations was done on MS plates to allow for visualization of GFP-positive individuals under UV light.

Fig 2. MAIN, DRM2 and CMT3 act synergistically to repress TEs.

(A) Representative pictures showing the developmental phenotype of 3-week-old *ddc*, *main-3* and *ddc main-3* mutants in comparison to WT plant. (B) Number of upregulated TEs in *ddc*, *main-3* and *ddc*

main-3, and classified by TE superfamily. (C) Chromosomal distributions of misregulated loci in *ddc*, *main-3* and *ddc main-3* over WT. Chromosome arms are depicted in light grey, pericentromeric regions in dark grey as defined in [50]. Upregulated genes and TEs are represented in blue and red, respectively; downregulated genes are represented in green. (D) Fraction of upregulated TEs in *ddc*, *main-3* and *ddc main-3* located in chromosome arms or in pericentromeric regions as defined in [50]. Asterisks indicate statistically significant enrichments of TEs in pericentromeric regions in comparison to the genomic distribution of all *A. thaliana* TEs (Chi-Square test, *: p-value ≤ 0.05, **: p-value ≤ 0.01 n.s: not significant). (E) Heatmap showing upregulated TEs in *ddc*, *main-3* and *ddc main-3* mutants in comparison to WT plants. (F-G) Relative expression analyses of *ATCOPIA28* (F) and several endogenous TEs (G) in *ddc*, *main-3*, *ddc main-3*, *cmt3 main-3* and *drm1 drm2 (dd) main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the mutants are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. RNA-seq threshold: $\log_2 \geq 2$, or $\log_2 \leq -2$; p-adj < 0.01.

Fig 3. MAIN and MAIL1 are required for the proper expression of similar genes, and for TE silencing.

(A-B) Number of misregulated genes (A) and upregulated TEs (B) in *main-2*, *mail1-1* and *main-3* mutants in comparison to WT Col plants. TEs are classified by superfamily. (C) Heatmap showing misregulated loci in *main-2*, *mail1-1* and *main-3* in comparison to Col and WT controls, respectively. Asterisks represents loci that are commonly misregulated in the three mutant backgrounds. (D) Venn diagrams analyses representing the overlaps between misregulated loci in *main-2*, *mail1-1* and *main-3*. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value < 2.2.10e-16). (E) Fraction of misregulated loci in *main-2* and *mail1-1* located in chromosome arms or in pericentromeric regions as defined in [50]. Asterisks indicate statistically significant enrichments of downregulated genes and upregulated genes and TEs in chromosome arms and pericentromeric regions, respectively, in comparison to the genomic distributions of all *A. thaliana* genes and TEs (Chi-

Square test, *: p-value $\leq$ 0.05, **: p-value $\leq$ 0.01, n.s: not significant). Percentages of genes targeted by DNA methylation and H3K9me2 were calculated based on enrichment in heterochromatin states 8 and 9 as defined in [51]. RNA-seq threshold: log₂ $\geq$ 2, or log₂ $\leq$ -2 ; p-adj< 0.01.

Fig 4. The *main-2* mutation has a slight effect on non-CG DNA methylation levels.

(A) Genome-wide DNA methylation levels along the five *Arabidopsis* chromosomes in *main-2* versus WT Col plants. Chromosome arms are depicted in light grey, pericentromeric regions in dark grey as defined in [50]. Mb: megabase. (B-H) Boxplot analyses in two *main-2* and WT Col biological replicates showing the DNA methylation levels of all pericentromeric TEs (B) and genes (C), all chromosome arms TEs (D) and genes (E), TEs that are upregulated in *main-2* (F), and genes that are upregulated (G) and downregulated (H) in *main-2*. p-values were calculated using a Wilcoxon test. ***: p-value < 2.10e-16.

Fig 5. MAIN, MAIL1 and PP7L physically interact together.

(A) Representative pictures of 3-week-old *main-2* and *mail1-1* mutants, and epitope-tagged complementing lines in comparison to WT Col plants. (B) Western blots using anti-FLAG and anti-Myc antibodies showing the accumulation of epitope-tagged PMD proteins at the expected sizes in the different complementing lines. Coomassie staining of the large Rubisco subunit (rbcl) is used as a loading control. KDa: kilodalton. (C-E) Relative expression analyses of upregulated TEs (C), upregulated genes (D) and downregulated genes (E) in the different complementing lines assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the complementing lines and mutants are represented relative to WT Col. Error bars indicate standard deviation based on three independent biological replicates. (F) FLAG-tagged MAIN and MAIL1 proteins were immunoprecipitated and putative interacting proteins were identified by mass spectrometry. Numbers of identified spectra, peptides and the normalized spectral abundance factor (NSAF₅) are

shown for two independent experiments, including three *main-2* and two *mail1-1* replicates. WT replicates are used as a negative control. Only proteins reproducibly enriched in all the FLAG-MAIN and FLAG-MAIL1 IP, and depleted in WT controls across multiple replicates are described in the table. (G) MAIL1-MYC was co-immunoprecipitated with MAIN-FLAG in F1 plants obtained by crossing MAIL1-MYC and MAIN-FLAG lines together. Parental MAIL1-MYC and MAIN-FLAG lines were used as negative controls. (H) The MAIN-MYC line was supertransformed with the PP7L-FLAG construct, and MAIN-MYC was co-immunoprecipitated with PP7L-FLAG. Plants expressing only MAIN-MYC or PP7L-FLAG were used as negative controls. (I) Same as H but using MAIL1-MYC plants supertransformed with the PP7L-FLAG construct. Epitope-tagged proteins were detected by Western blotting. Arrowheads indicates expected bands. Asterisks indicates non-specific hybridization. Co-exp: plants co-expressing PP7L-FLAG and MAIN-MYC (H) or PP7L-FLAG and MAIL1-MYC (I).

Fig 6. *main-2*, *mail1-1*, *pp7l-2* single and *mail1-1 pp7l-2* double mutants display similar developmental and molecular phenotypes.

(A) Representative pictures of 3-week-old *main-2*, *mail1-1*, *pp7l-2* single and *mail1-1 pp7l-2* double mutants in comparison to WT Col plant. (B) Heatmap showing misregulated loci in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2* mutants in comparison to WT Col plants using the datasets of RNA-seq Exp1, Exp2 and Exp3 (S2 and S4 Tables). One asterisk defines the loci that are commonly misregulated in all mutant backgrounds. Two asterisks define the loci that are misregulated in the *mail1-1 pp7l-2* double mutant. (C) Venn diagrams analyses representing the overlaps between misregulated loci in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2*. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <2.2.10e-16). (D) Relative expression analyses of upregulated TEs, genes and downregulated genes in the different genotypes assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different mutants are represented relative to WT Col. Error bars indicate standard deviation based on three independent

biological replicates. (E-G) Boxplots analyses showing average RPKM values of upregulated TEs (E), upregulated genes (F) and downregulated genes (G) in *mail1-1 pp7l-2* in the indicated genotypes of RNA-seq Exp3. These analyses are based on the misregulated loci datasets defined by ** in panel B. P-values were calculated using a Wilcoxon test, and only significant p-values are shown. *: p-value < 1.10e-3; **: p-value < 3.10-6; ***: p-value < 2.10e-16.

Fig 7. Constitutive heterochromatin appears unaltered in *pp7l-2* mutant.

Proportion of nuclei showing condensed, partially decondensed (intermediate), or decondensed chromocenters in the *pp7l-2* mutant in comparison to WT control (Col) based on H3K9me2 immunostaining of nuclei. Representative pictures of nuclei displaying condensed, partially decondensed or decondensed chromocenters. DAPI: DNA stained with 4',6-diamidino-2-phenylindole.

Fig 8. Evolutionary history of PMD-C and PP7 proteins in plants.

(A) An alignment of the PMD-C motifs from 30 representative Eudicot species was used to construct a phylogenetic tree. The two major clades (MAIL2/MAIL2-like and MAIL3) are indicated. The species codes are given in S11 Table, and corresponding protein sequences in S12 Table). In red are genes presenting a fusion between a PMD-C and a PP7 motif. Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated. Scale bar indicates one substitution/site. The tree was rooted using the *Amborella trichopoda* PMD-C motif (Atr1PMDC). (B) Phylogenetic tree constructed using an alignment of the PP7 motif from the same species as in (A). The two major clades (PP7 and PP7L) are indicated. In red are genes presenting a fusion between a PP7 and a PMD-C motif. Statistical supports of key nodes calculated with the approximate likelihood-ratio test are indicated. Scale bar indicates one substitution/site. The tree was rooted using the *A. thaliana* PP5 motif (AtPP5).

SUPPORTING INFORMATION CAPTIONS

S1 Fig. *MAIN* is the mutated gene responsible for *ATCOPIA28::GFP* and TE overexpression in the *ddc* #16 mutant.

(A) Representative pictures of *ddc* #18 (*ddc morc6-8*) and *ddc* #344 (*ddc morc6-9*) mutants in comparison to *ATCOPIA28::GFP* WT and *ddc* control plants under UV light. Insets show plants under white light. (B) Enrichment in homozygote/heterozygote ratio of EMS over WT single nucleotide polymorphisms (SNPs), defining the linkage intervals for the populations *ddc* #18 and *ddc* #344. Mb: megabase. Gray-shaded rectangles delimit the mapping intervals. (C) Location of the point mutations corresponding to the *morc6-8* and *morc6-9* alleles within the *MORC6* genomic sequence. Nucleotide and corresponding amino acid changes are indicated above the gene. Positions of the mutations are indicated relative to the transcription start site (+1). Grey boxes represent 5' and 3' UTR, blue boxes and lines represent exons and introns, respectively. (D) Enrichment in homozygote/heterozygote ratio of EMS over WT single nucleotide polymorphisms (SNPs), defining the linkage intervals for the population *ddc* #16. Gray-shaded rectangle delimits the mapping interval. (E) Location of the point mutation corresponding to the *main-3* mutant allele within the *MAIN* genomic sequence. (F) Genetic complementation analyses using the KO T-DNA insertion line *main-2*. *ddc* #16 plants were crossed with *main-2* plants. F1 plants were self-crossed, and F2 plants were screened under UV light to select GFP-overexpressing plants. Western blotting using anti-GFP antibodies confirmed GFP overexpression in selected F2 plants. Coomassie staining of the large Rubisco subunit (rbcl) is used as a loading control. KDa: kilodalton. Among the selected F2 plants, the presence of *main-3* EMS and *main-2* T-DNA mutant alleles were determined by dCAPS-PCR and PCR analyses, respectively. *DRM2* and *CMT3* genotyping were determined by PCR analyses. WT: Wild type, Ho: Homozygote mutant. He: Heterozygote. (G) Relative expression analyses of several TEs in the indicated genotypes assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels

in the different genotypes are represented relative to WT. Error bars indicate standard deviation based on two independent biological replicates. Screening of EMS mutant populations was done on MS plates to allow for visualization of GFP-positive individuals under UV light.

S2 Fig. Combining the *drm2*, *cmt3* and *main-3* mutations exacerbate TE silencing defects.

(A) Principal component analysis (PCA) performed after batch correction for first two components of the sixteen samples described in RNA-seq EMS Exp1 and Exp2. (B) Relative expression analyses of *ATCOPIA28* and *HELITRONY1D* (*AT5TE35950*) in *ddc*, *main-3* and *ddc main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. (C) Venn diagrams analysis showing the overlaps between reproducibly upregulated TEs in *ddc*, *main-3* and *ddc main-3*. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <2.2.10e-16). (D) Same as panel B for TEs defined as class I-IV TEs. Frames of RT-qPCR graphs are using the same color code as shown in panel C. (E) Venn diagrams analyses defining the overlaps between up- and downregulated genes in the different genotypes. Fisher's exact test statistically confirmed the significance of Venn diagram overlaps (p-value <2.2.10e-16). (F) Fraction of misregulated genes in *ddc*, *main-3* and *ddc main-3* located in chromosome arms or in pericentromeric regions as defined in [50]. Asterisks indicate statistically significant enrichments of misregulated genes in chromosome arms or pericentromeric regions in comparison to the genomic distributions of all *A. thaliana* genes (Chi-Square test, **: p-value ≤ 0.01). Percentages of genes targeted by DNA methylation and H3K9me2 were calculated based on enrichment in heterochromatin states 8 and 9 as defined in [51]. (G) Relative expression analyses of *DRM2* and *CMT3* in *ddc*, *main-3*, *ddc main-3*, *cmt3 main-3* and *dd main-3* assayed by RT-qPCR. RT-qPCR analyses were normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are represented relative to WT. Error bars indicate standard deviation based on three independent biological replicates. Screening of EMS mutant populations was done on MS plates

973 to allow for visualization of GFP-positive individuals under UV light.

974 **S3 Fig. Identification of reproducibly misregulated loci in *main-2*, *mail1-1* and *main-3*.**

975 (A) Principal component analysis (PCA) performed after batch correction for first two components of
976 the twenty-four *main-2*, *mail1-1* and WT Col samples described in RNA-seq Exp1, Exp2 and Exp3. (B-
977 D) Relative expression analyses of several upregulated TEs (B), upregulated genes (C), and
978 downregulated genes (D) in *main-2*, *mail1-1* and *main-3* assayed by RT-qPCR. RT-qPCR analyses were
979 normalized using the housekeeping *RHIP1* gene, and transcript levels in the different genotypes are
980 represented relative to respective WT controls. Error bars indicate standard deviation based on three
981 independent biological replicates. (E) Venn diagrams analyses representing the overlaps between
982 misregulated loci in *main-2*, *mail1-1*, *ddc* and *ddc main-3*. Fisher's exact test statistically confirmed the
983 significance of Venn diagram overlaps (p-value <0.005).

984 **S4 Fig. DNA methylation analyses in the *main-2* mutant**

985 (A-B) Boxplot analyses in two *main-2* and WT Col biological replicates showing the DNA methylation
986 levels at genomic sites previously defined as hypo CHG differentially methylated regions (DMR) in
987 *cmt3* (A) and hypo CHH DMR in *drm1 drm2* (B) based on [26]. p-values were calculated using a
988 Wilcoxon test. *: p-value <5.10e-7, **: p-value <5.10e-10, ***: p-value < 2.10e-16.

989 **S5 Fig. MAIN, MAIL1 and PP7L are required for the proper expression of similar loci, and commonly**
990 **downregulated genes carry the 'DOWN' DNA motif in their promoter.**

991 (A) Principal component analysis (PCA) performed after batch correction for first two components of
992 the thirty-two samples described in RNA-seq Exp1, Exp2 and Exp3. (B) Number of misregulated genes
993 in the different genotypes in comparison to WT Col plants from RNA-seq Exp3 (four biological
994 replicates, S3 and S6 Tables). (C) Number of upregulated TEs in *pp7l-2* and *mail1-1 pp7l-2*, and
995 classified by TE superfamily. (D) Fraction of misregulated loci in *pp7l-2* and *mail1-1 pp7l-2* located in
996 chromosome arms or in pericentromeric regions as defined in [50]. Asterisks indicate statistically

significant enrichments of downregulated genes, upregulated genes and TEs in chromosome arms and pericentromeric regions, respectively, in comparison to the genomic distributions of all *A. thaliana* genes and TEs (Chi-Square test, *: p-value $\leq$ 0.05, **: p-value $\leq$ 0.01, n.s: not significant). Percentages of genes targeted by DNA methylation and H3K9me2 were calculated based on enrichment in heterochromatin states 8 and 9 as defined in [51]. (E) Identification and proportions of the 'DOWN' DNA motif among the promoters of downregulated genes and all *Arabidopsis* genes using the MEME software. Promoter regions are defined as 1kb upstream of ATG. The list of all *Arabidopsis* genes used to determine genomic distributions is based on the TAIR file: TAIR10_upstream_1000_translation_start_20101028. RNA-seq threshold: log2 $\geq$ 2, or log2 $\leq$ -2 ; p-adj< 0.01.

S6 Fig. Full size images of panels described in Fig 5G-I.

S1 Table. Lists of differentially expressed loci in *ddc*, *main-3* and *ddc main-3*.

S2 Table. Lists of differentially expressed loci in *main-2* and *mail1-1*.

S3 Table. Lists of loci commonly misregulated in *main-2*, *mail1-1* and *main-3*.

S4 Table. Lists of differentially expressed loci in *pp7l-2* and *mail1-1 pp7l-2*.

S5 Table. Lists of loci commonly misregulated in *main-2*, *mail1-1*, *pp7l-2* and *mail1-1 pp7l-2*.

S6 Table. Lists of loci commonly misregulated in all mutant backgrounds (except *ddc*) analyzed in this study.

1022 **S7 Table. Lists of commonly downregulated genes displaying the “DOWN” motif in their promoter**
1023 **and random test analyses.**

1024

1025 **S8 Table. List of species used to construct the two trees of figure 8, their codes and the**
1026 **presence/absence of the different PMD-C and PP7 motifs.**

1027

1028 **S9 Table. (A) PMD-C and (B) PP7/PP7L motifs used to construct the two phylogenetic trees of**
1029 **Figure 8.**

1030

1031 **S10 Table. List of primers used in this study.**

1032

1033 **S11 Table. Next Generation Sequencing (NGS) mapping and coverage statistics.**

1034

Fig1

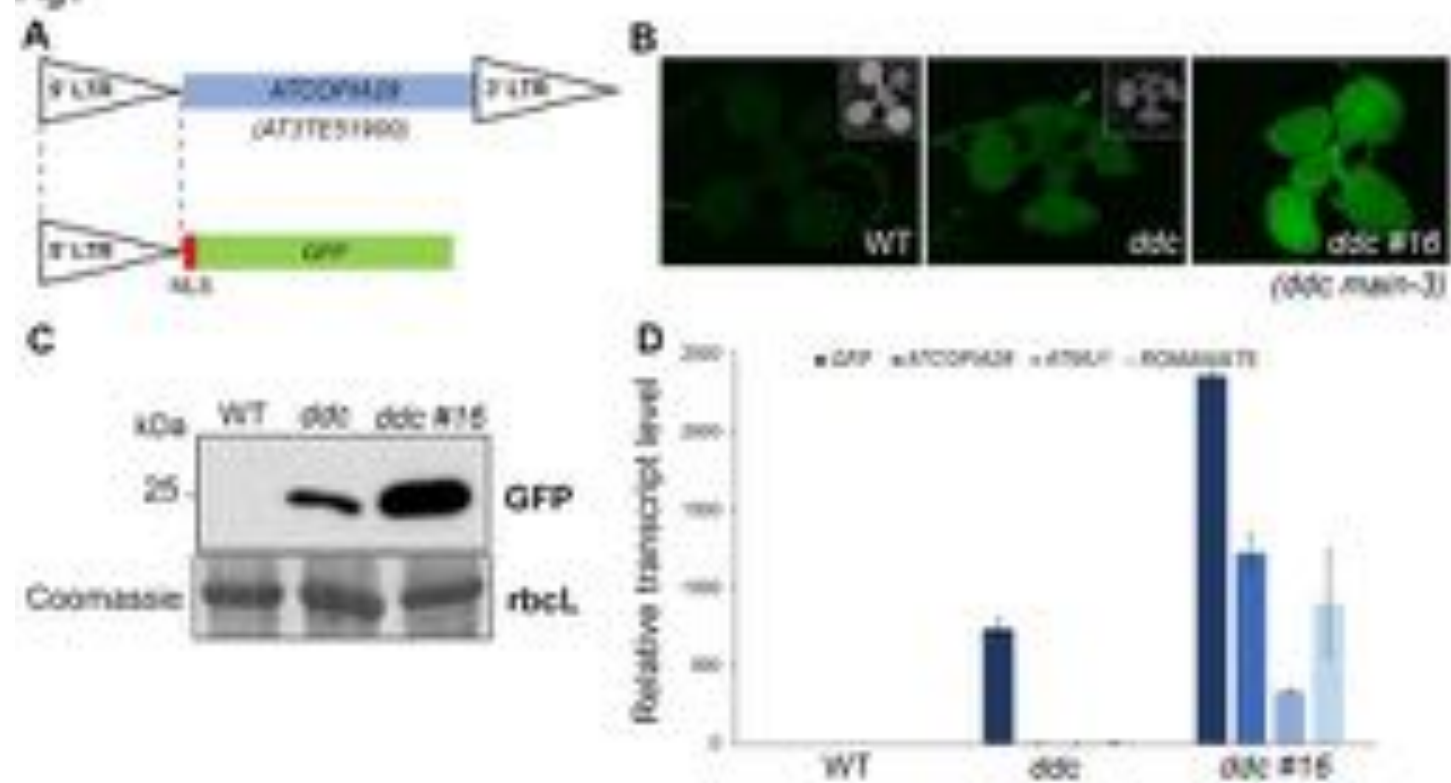


Fig 2

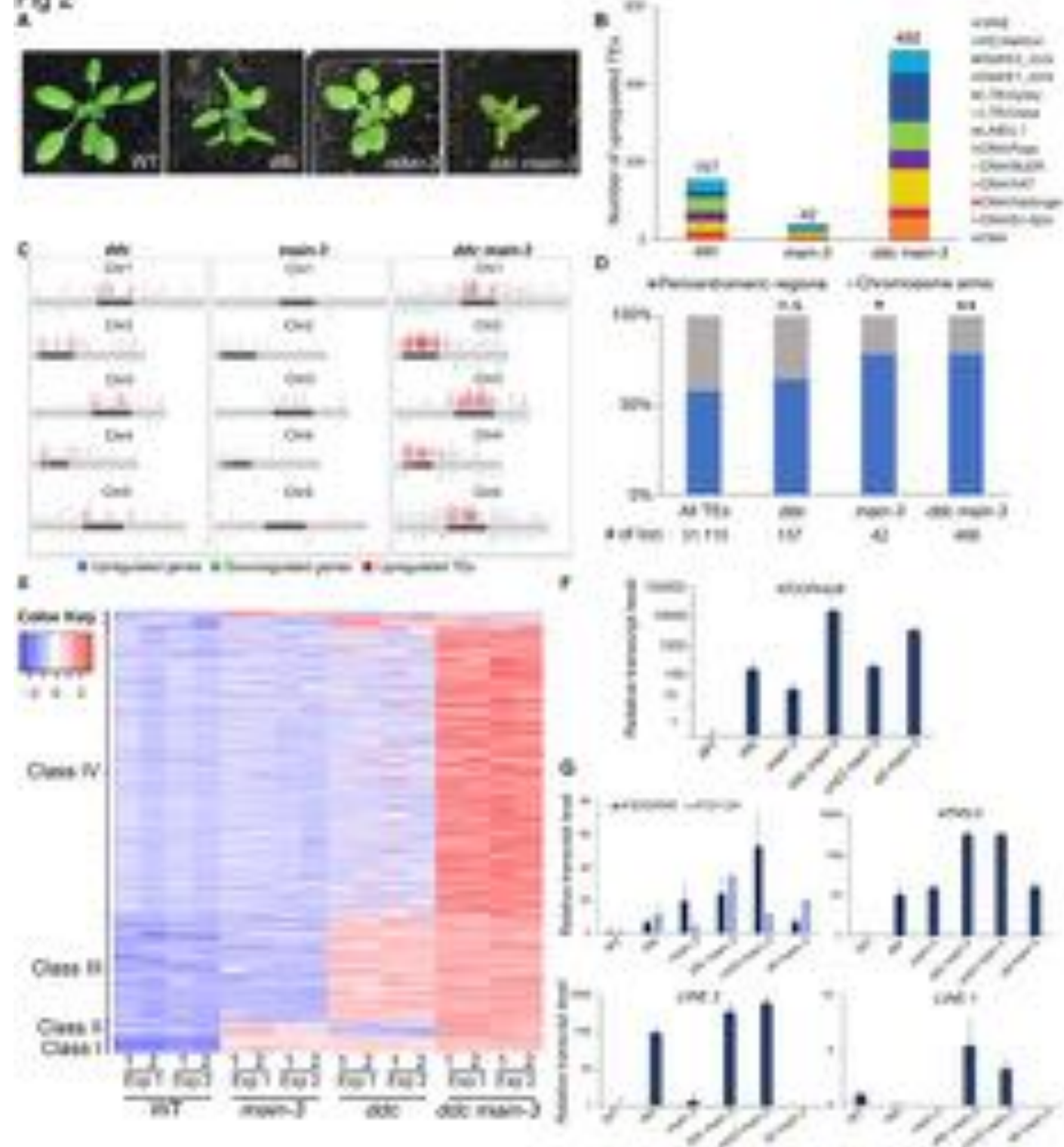


Fig 4

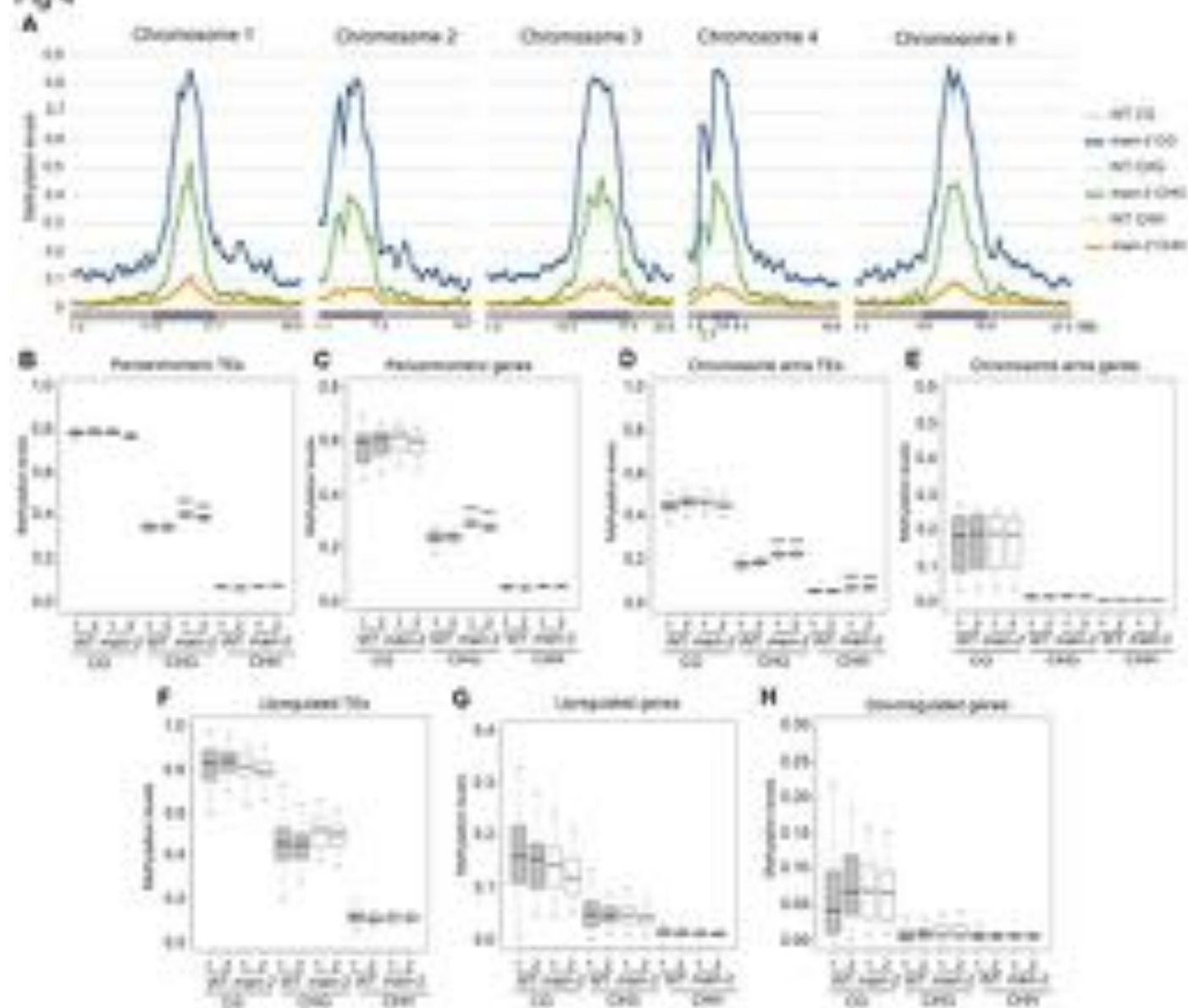


Fig 5

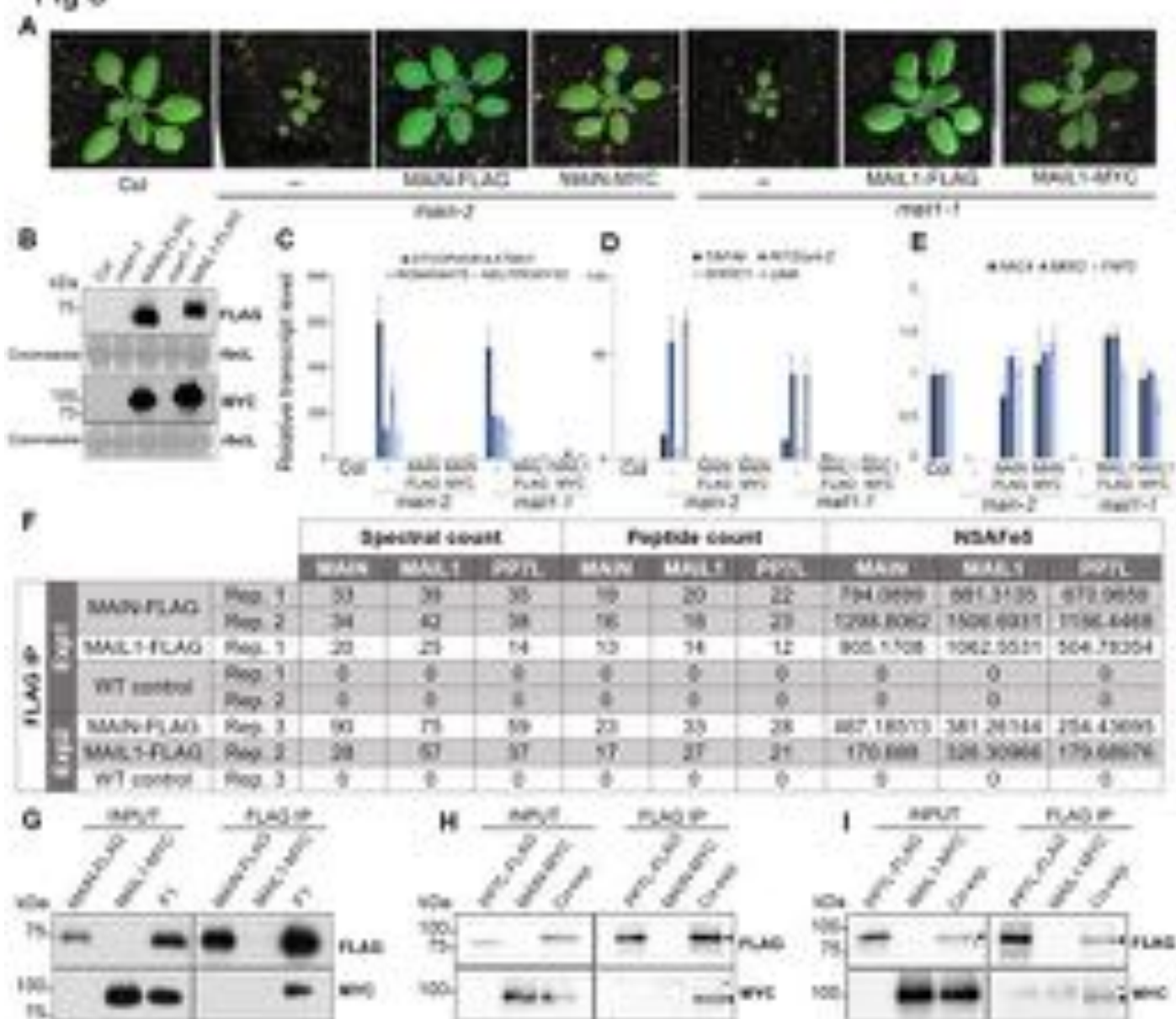


Fig 6

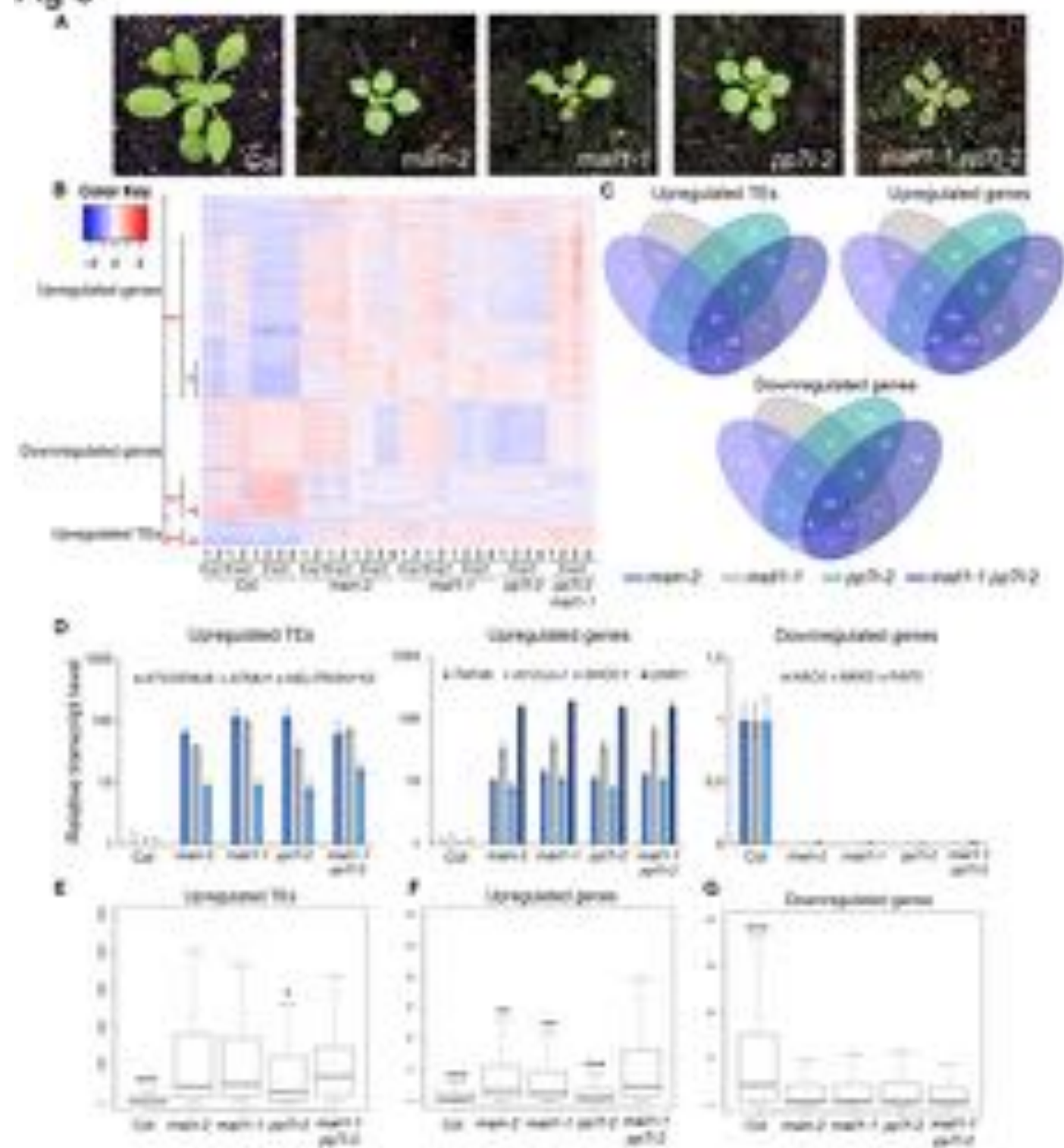


Fig 7

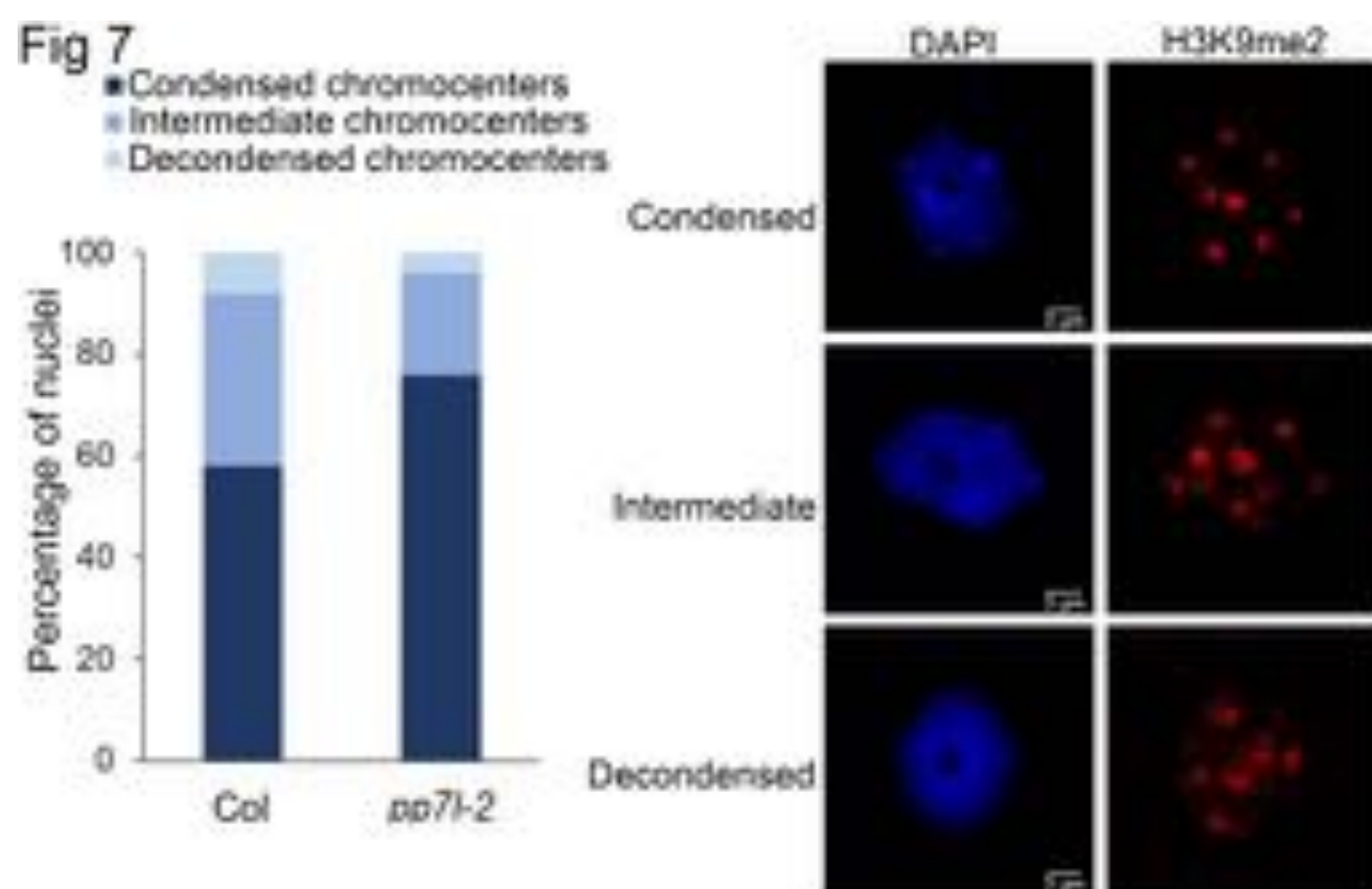
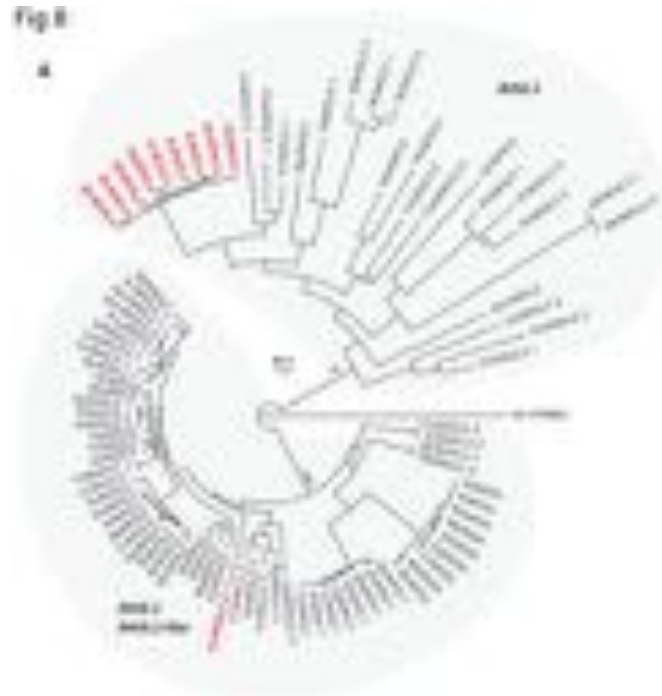


Fig 8

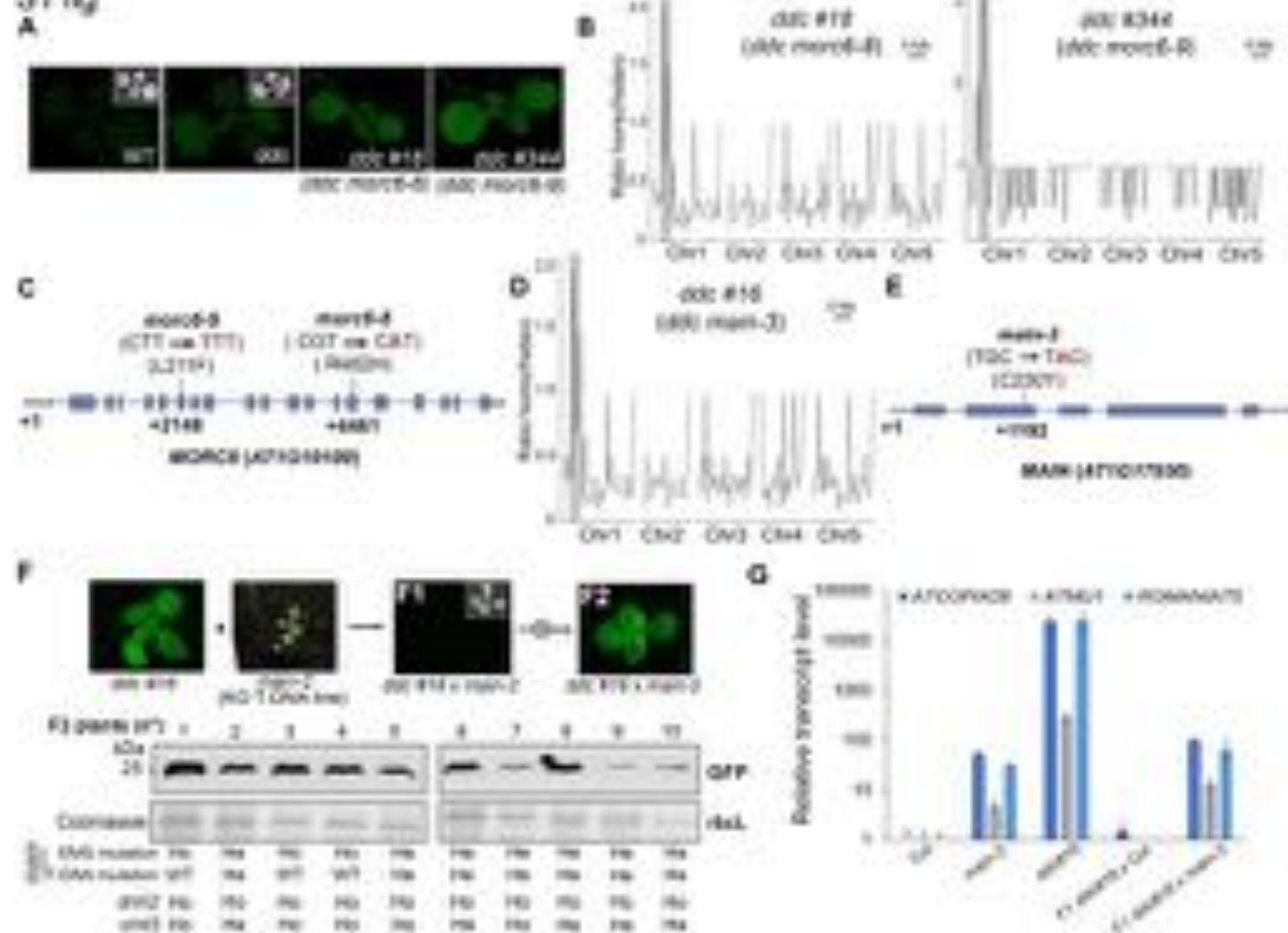
a



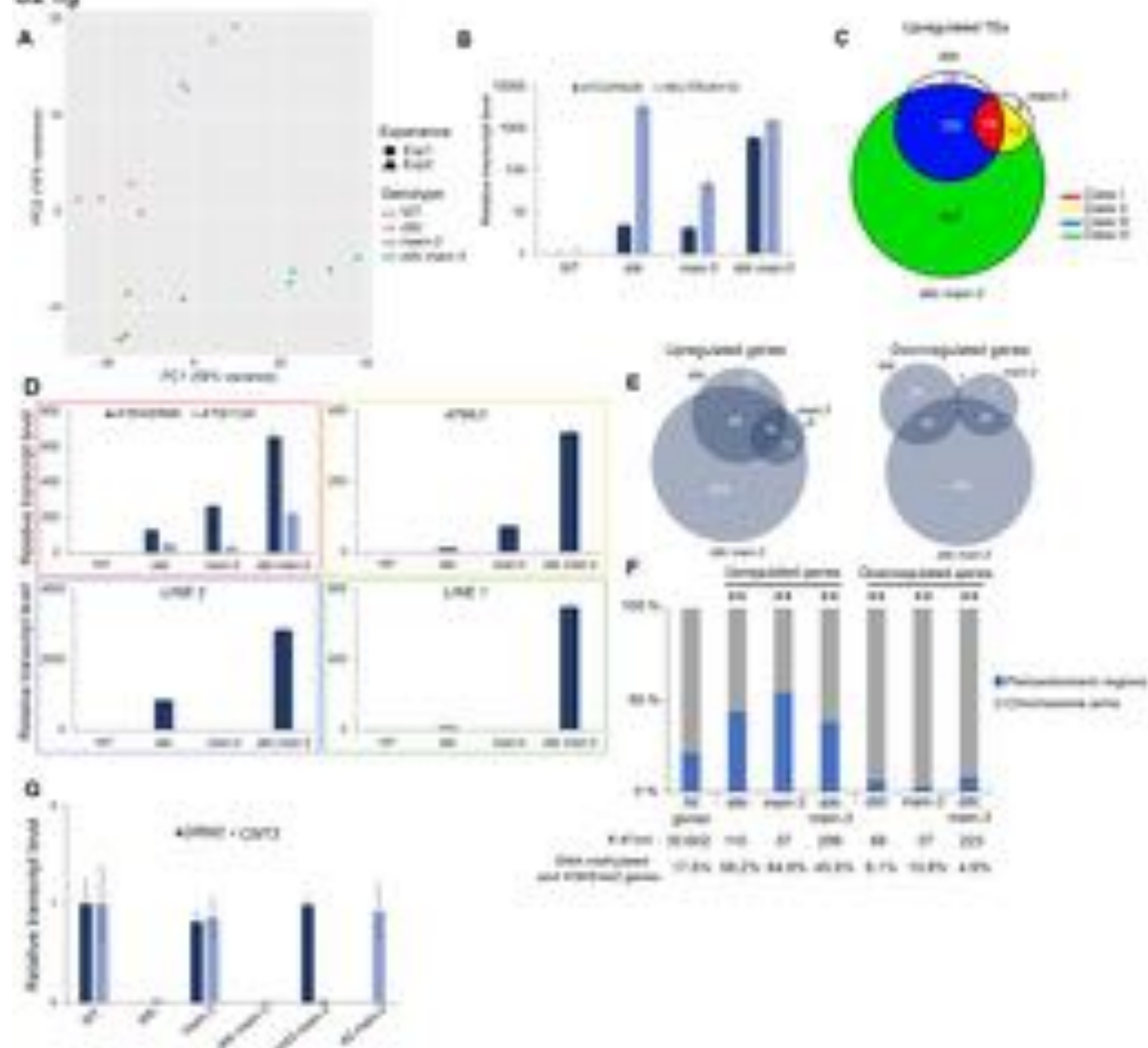
b



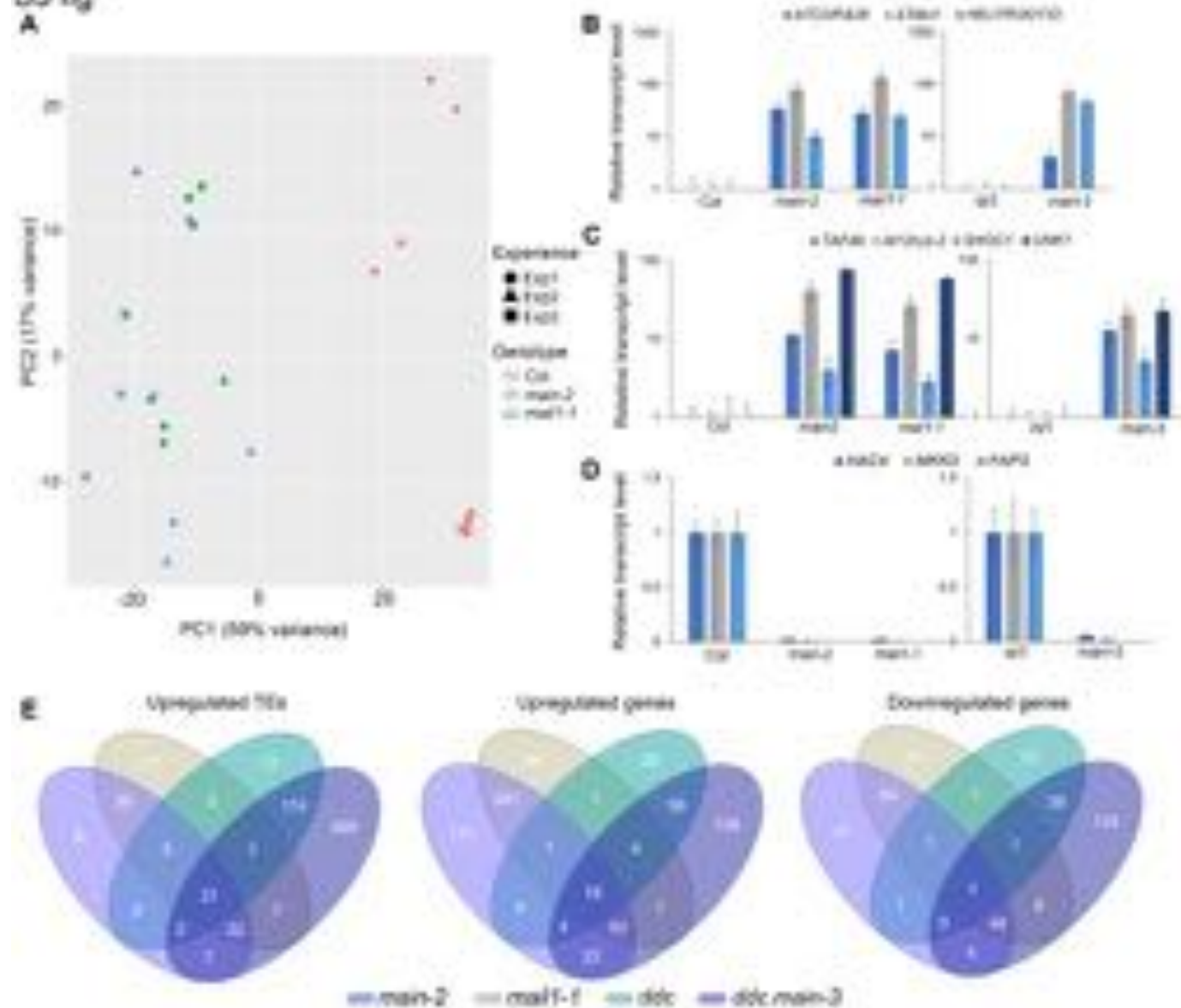
S1 fig



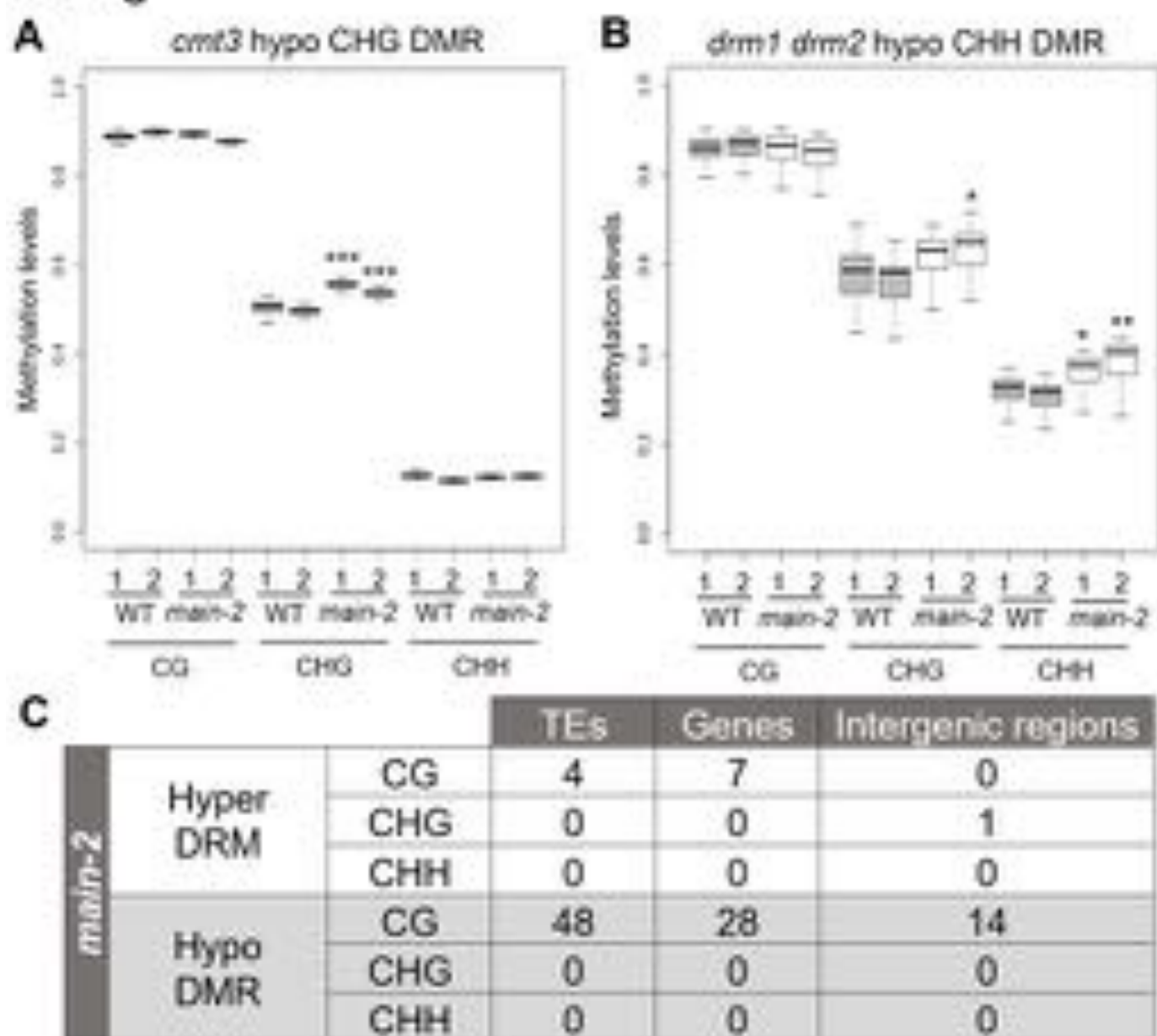
S2 fig



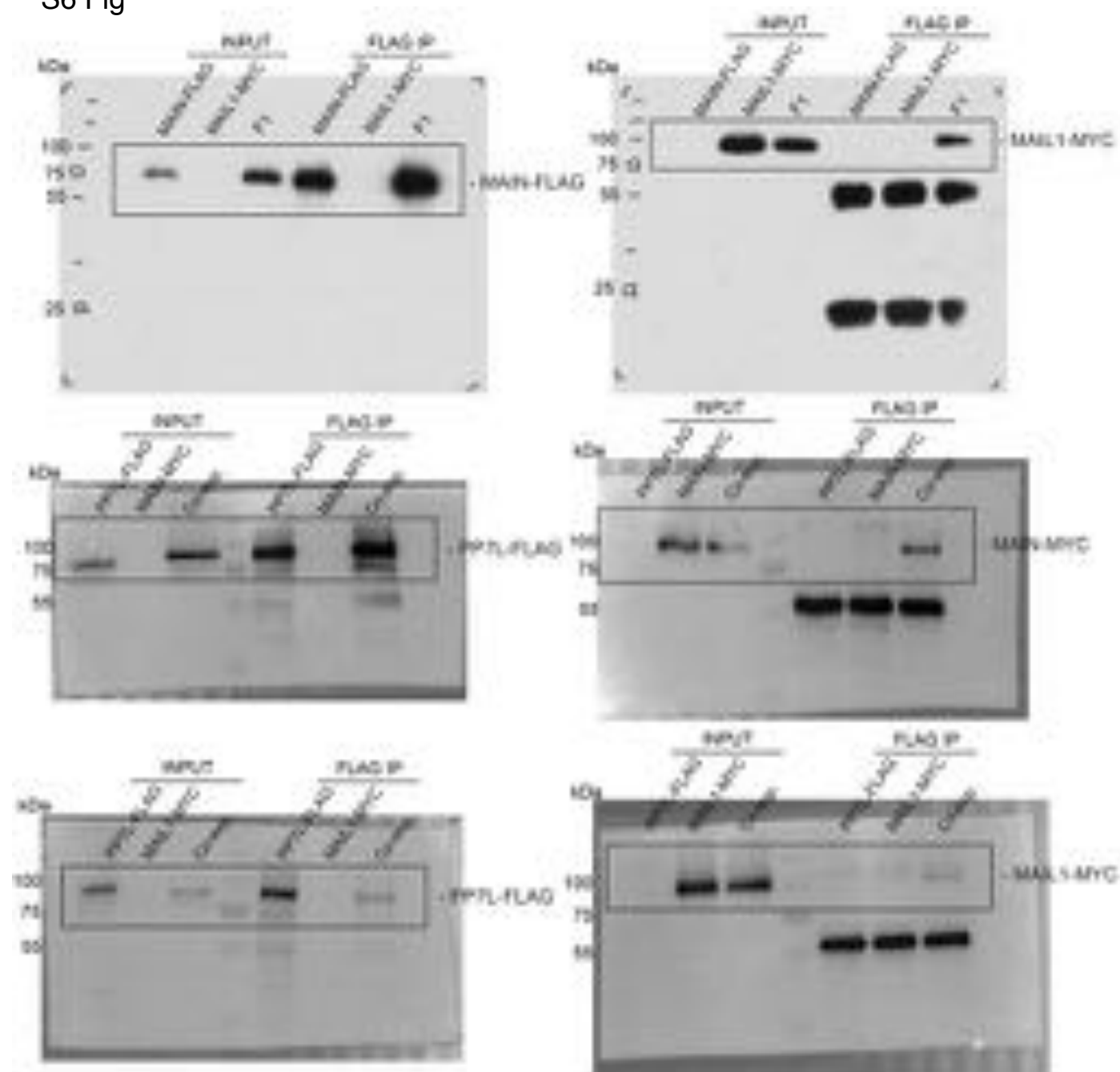
S3 fig



S4 fig



S6 Fig



ARTICLE

Received 26 Oct 2015 | Accepted 14 Apr 2016 | Published 13 Jun 2016

DOI: 10.1038/ncomms11640

OPEN

MTHFD1 controls DNA methylation in *Arabidopsis*

Martin Groth¹, Guillaume Moissiard¹, Markus Wirtz², Haifeng Wang³, Carolina Garcia-Salinas⁴, Perla A. Ramos-Parra⁴, Sylvain Bischof¹, Suhua Feng^{1,5,6}, Shawn J. Cokus¹, Amala John¹, Danielle C. Smith¹, Jixian Zhai¹, Christopher J. Hale¹, Jeff A. Long¹, Ruediger Hell², Rocío I. Díaz de la Garza⁴ & Steven E. Jacobsen^{1,5,6}

DNA methylation is an epigenetic mechanism that has important functions in transcriptional silencing and is associated with repressive histone methylation (H3K9me). To further investigate silencing mechanisms, we screened a mutagenized *Arabidopsis thaliana* population for expression of *SDCpro-GFP*, redundantly controlled by DNA methyltransferases DRM2 and CMT3. Here, we identify the hypomorphic mutant *mthfd1-1*, carrying a mutation (R175Q) in the cytoplasmic bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase (MTHFD1). Decreased levels of oxidized tetrahydrofolates in *mthfd1-1* and lethality of loss-of-function demonstrate the essential enzymatic role of MTHFD1 in *Arabidopsis*. Accumulation of homocysteine and S-adenosylhomocysteine, genome-wide DNA hypomethylation, loss of H3K9me and transposon derepression indicate that S-adenosylmethionine-dependent transmethylation is inhibited in *mthfd1-1*. Comparative analysis of DNA methylation revealed that the CMT3 and CMT2 pathways involving positive feedback with H3K9me are mostly affected. Our work highlights the sensitivity of epigenetic networks to one-carbon metabolism due to their common S-adenosylmethionine-dependent transmethylation and has implications for human MTHFD1-associated diseases.

¹Department of Molecular, Cell, and Developmental Biology, University of California Los Angeles, Los Angeles, California 90095, USA. ²Centre for Organismal Studies, University of Heidelberg, Heidelberg 69120, Germany. ³Basic Forestry and Proteomics Research Center, Haixia Institute of Science and Technology (HIST), Fujian Province Key Laboratory of Plant Virology, Institute of Plant Virology, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China. ⁴Tecnologico de Monterrey, Campus Monterrey, Ave. Eugenio Garza Sada 2501, Monterrey 64849, México. ⁵Eli & Edythe Broad Center of Regenerative Medicine & Stem Cell Research, University of California Los Angeles, Los Angeles, California 90095, USA. ⁶Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, California 90095, USA. Correspondence and requests for materials should be addressed to M.G. (email: grothmn@gmail.com) or to S.E.J. (email: jacobsen@ucla.edu).

DNA methylation serves as a defense mechanism against transposable elements (TEs) and other types of repetitive DNA that can harm the genome of the organism they inhabit. DNA methylation promotes the packaging of DNA into so-called heterochromatin as to enforce a silent state, for example, by making the DNA inaccessible to transcription activators. Because of its transgenerational stability, DNA methylation is the prime example of an epigenetic mechanism. The stability is provided by feedback loops within, as well as crosstalk between, different methylation pathways, as established in *Arabidopsis thaliana* (*Arabidopsis*), where DNA methylation occurs at CG, CHG and CHH (H = A, T or C)¹. In contrast to CHG and CHH methylation, which are exclusively involved in heterochromatin formation and transcriptional gene silencing (TGS), CG methylation also occurs over gene bodies.

DNA methylation is generated by the activity of DNA methyltransferases (DNMTs), which enzymatically transfer a methyl group from S-adenosyl methionine (SAM) to cytosine. In *Arabidopsis*, DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2) catalyses *de novo* methylation in all sequence contexts². Once established, DNA methylation at symmetric CG and CHG sequences is maintained by DNA METHYLTRANSFERASE1 (MET1) and CHROMOMETHYLASE3 (CMT3), respectively. Maintenance of CG methylation is based on the recognition of hemimethylated signatures after semi-conservative DNA replication. Analogous to the recruitment of the mammalian maintenance methylase DNMT1 through UHRF1, members of the VARIANT IN METHYLATION (VIM) family bind to hemimethylated DNA with their SET and RING-associated (SRA) domains and are required for CG methylation by MET1 (refs 1,3). In contrast, CHG methylation is maintained by a reinforcing loop between non-CG methylation and methylation of lysine 9 of histone H3 (H3K9me), which involves the histone methyltransferases KRYPTONITE/SUVH4, 5 and 6. These preferentially bind methylated non-CG sequences via their SRA domains and modify the wrapped nucleosome with H3K9me (ref. 4). In turn, CMT3 binds H3K9me through its chromo and BAH domains and catalyses the remethylation of CHG sites during replication⁵. Similarly, maintenance of CHH methylation by CMT2 also depends on SUVH4/5/6-mediated H3K9me (ref. 6).

While CMT2 and 3 mostly target transposons in the pericentromeric heterochromatin, DRM2 is mainly required for maintenance of CHH methylation and TGS in the chromosomal arms⁷. Targeting of DRM2 is mediated by the concerted action of short transcripts that are processed into 24 nt small-interfering RNAs and complementary long noncoding transcripts produced by the plant-specific RNA polymerase complexes Pol IV and Pol V, respectively. In the canonical RNA-directed DNA methylation pathway (RdDM), 24 nt RNAs are incorporated into ARGONAUTE4 (AGO4) in order to match the RNA-induced silencing complexes with Pol V transcripts⁸. Subsequently, DRM2 is recruited to the target CHH sites by direct interaction with AGO4 (ref. 9). Recruitment of Pol IV to heterochromatin is also dependent on H3K9me interaction via SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1) (ref. 10), whereas Pol V is recruited via the non-catalytic SUV39 homologues SUVH2 and 9 that bind to methylated DNA via their SRA domains¹¹.

TGS is often reinforced by the synergistic action of different DNA methylation pathways. This is exemplified by the *SUPPRESSOR OF DRM2 CMT3* (*SDC*) locus, which is redundantly silenced by the CMT3- and DRM2-mediated methylation of tandem repeats in the promoter region¹². Hence, *SDC* is ectopically expressed in *drm2 cmt3* double mutants, but repressed during most of development in the single mutants, making *SDC* a

powerful genetic marker of simultaneous impairment of CHG and CHH methylation pathways. We generated stable transgenic lines carrying an *SDC_{pro}-GFP* fusion construct in wild-type (WT) and *cmt3* genetic background and screened M2 populations for EMS-mutants that express GFP. The identification of *microrchidia 1* (*morc1*) and *morc6* mutants from this screen was published previously¹³. Here, we identified a mutant from the WT background that carries a missense mutation in the *Arabidopsis* METHYLENETETRAHYDROFOLATE DEHYDROGENASE/METHENYLTETRAHYDROFOLATE CYCLOHYDROLASE 1 (*MTHFD1*) gene. The mutation disrupts folate metabolism and leads to accumulation of homocysteine (Hcy), a hallmark of an impaired methionine (Met) cycle, whose main function is to produce SAM for transmethylation reactions and recycle the byproduct S-adenosyl homocysteine (SAH)¹⁴. Genome-wide loss of CHG and CHH methylation, reduced H3K9me and derepression of TEs in the *mthfd1-1* mutant indicate that the Met cycle constitutes an ‘Achilles heel’ of the feedback mechanisms between DNA and histone methylation.

Results

The R175Q mutation in MTHFD1 leads to *SDC_{pro}-GFP* expression. Mutant #162 was identified by screening M2 seedlings of an EMS-mutagenized population of *Arabidopsis* that carried an *SDC_{pro}-GFP* insertion event in WT background (herein after referred to as WT; Col refers to non-transgenic WT) for individuals that showed GFP fluorescence (Fig. 1a). Using deep sequencing of bulked GFP-positive F2 progeny of mutant #162 crossed with a WT plant of ecotype Landsberg *erecta* (*Ler*), we confined the target region containing the causative mutation to the north end of chromosome 3 (Supplementary Fig. 1). To identify the causative mutation, mutant #162 was crossed to WT and the co-segregation of candidate EMS mutations in GFP-positive F2 progeny was analysed using dCAPS markers (Supplementary Fig. 2). A guanine to adenine transition in a gene (At3g12290) encoding a putative methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase (*MTHFD1*) showed 100% co-segregation in 98 GFP-positive F2 individuals (Supplementary Fig. 2). The mutant phenotype segregated as a recessive monogenic trait (103 GFP-positive versus 368 GFP-negative, χ^2 *P* value (3:1) = 0.12). The mutation, herein after named *mthfd1-1*, leads to a predicted substitution of a conserved arginine by glutamine at residue 175 (R175Q) (Fig. 1b). To confirm that *mthfd1-1* caused the expression of *SDC_{pro}-GFP*, a #162 M3 mutant (*mthfd1-1/mthfd1-1*) was crossed with a heterozygous plant containing a transfer DNA (T-DNA) insertion allele, *MTHFD1/mthfd1-2* (Fig. 1b). GFP expression in F1 progeny co-segregated with the *mthfd1-2* allele (Fig. 1a,c), confirming that *mthfd1-1* caused the expression of *SDC_{pro}-GFP* in #162.

GFP-positive plants originating from mutant #162 had pale leaves, reduced seed set, were smaller and developed more slowly compared with WT but did not display other morphological defects (Fig. 1d and Supplementary Fig. 3a). In contrast, *mthfd1-2* homozygous mutants showed severe developmental defects, including dwarfism, pale, shortened leaves, reduced apical dominance, delayed flowering, prolonged vegetative phase and infertility (Fig. 1d and Supplementary Fig. 3a,b). Moreover, on average we only retrieved one viable homozygous *mthfd1-2* mutant out of 18 seeds from a heterozygous parent, indicating that more than 75% of homozygous *mthfd1-2* mutants died prematurely (Supplementary Table 1). Viability seemed to be affected during or after germination, because siliques from WT and heterozygous *MTHFD1/mthfd1-2* did not show differences in ovule and seed development (Supplementary Fig. 3a). We did not

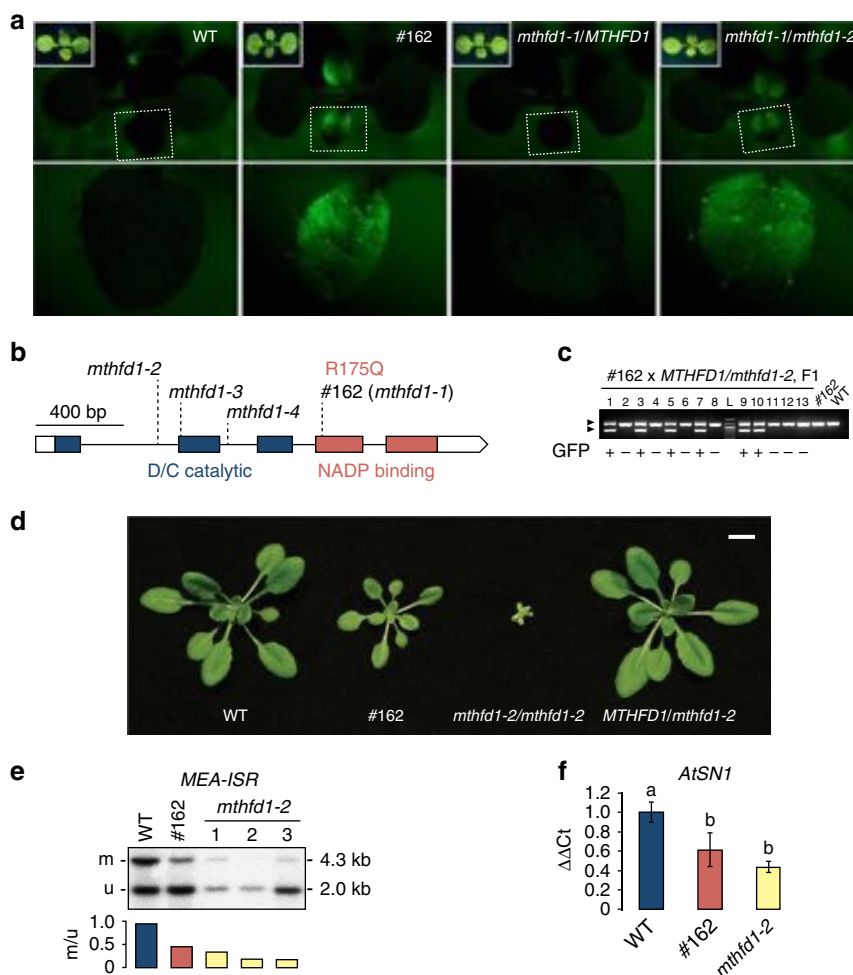


Figure 1 | SDCpro-GFP expression and DNA demethylation caused by R175Q mutation in *MTHFD1*. (a) GFP fluorescence micrographs of WT, #162 M2, *MTHFD1/mthfd1-2* F1 and #162/*mthfd1-2* F1 seedlings. F1 are progeny of #162 M2 x *MTHFD1/mthfd1-2*. Dashed boxes indicate magnified areas shown in lower panels. Inlets show bright-field images. (b) Gene structure, positions of mutations and conserved domains of *MTHFD1*. The EMS mutation in #162 lead to a R175Q substitution of a conserved residue required for NADP binding²⁸. (c) PCR-based genotype analysis of 13 F1 seedlings and two control samples. Arrowheads mark bands corresponding to WT/*mthfd1-1* (upper) and *mthfd1-2* (lower). The *mthfd1-2* allele co-segregates with GFP fluorescence in F1 (+: present, -: absent). L, ladder. (d) Habit of different genotype plants 20 days after germination. Scale bar, 10 mm. (e) DNA blot analysis of non-CG methylation at the *MEA-ISR* locus. Genomic DNA was digested with methylation-sensitive *MspI*; upper and lower bands correspond to methylated (m) and unmethylated (u) fragments, respectively. Ratios of band intensities for each lane are shown under the gel image. (f) Levels of non-CG methylation at the *AtSN1* locus by quantitative chop PCR analysis of genomic DNA after digestion with methylation-sensitive *HaeIII* relative to undigested DNA. Mean values $\pm$ s.d. ($n = 3$). Different letters above bars indicate significant differences between pairwise comparisons by Student's *t*-test ($P < 0.05$).

retrieve any homozygous mutants from the T-DNA alleles *mthfd1-3* or *mthfd1-4* (Fig. 1b). These results led to the conclusion that *MTHFD1* is an essential gene in *Arabidopsis* and that the T-DNA insertion in the first intron in *mthfd1-2* does not completely abolish the function of *MTHFD1*. Moreover, the results indicate that R175Q in *mthfd1-1* partially impairs gene function, causing a hypomorphic phenotype that does not affect viability.

To test if DNA methylation is altered in *mthfd1* mutants, we analysed the *MEDEA INTERGENIC SUBTELOMERIC REPEAT (MEA-ISR)* locus using a well-established Southern blot assay¹⁵. In WT, approximately half of the *MEA-ISR* alleles are methylated and restriction digest with methylation-sensitive *MspI* produces two fragments of similar abundance. Both #162 and *mthfd1-2* mutants showed a reduction of the methylated band, and the loss of methylation was stronger in *mthfd1-2* than in #162 (Fig. 1e). We also analysed DNA methylation at *Arabidopsis thaliana* *SHORT INTERSPERSED ELEMENT 1 (AtSN1)* by quantitative PCR following methylation-sensitive restriction digestion and

observed similar reductions in DNA methylation (Fig. 1f). Furthermore, bisulfite (BS)-PCR analysis of the levels of methylation at CG, CHG and CHH sites in the tandem repeat region of the transgenic *SDC* promoter showed a decrease in CHG methylation of the transgenic *SDC* in #162 mutants compared with the WT reference (Supplementary Fig. 4). In summary, the DNA methylation assays confirmed that *MTHFD1* is required for DNA methylation at different loci in different sequence contexts, and the R175Q amino-acid substitution leads to reduced DNA methylation. Because of the limiting amounts of tissue available from *mthfd1-2* mutants, only #162 mutants (referred to as *mthfd1-1* herein after) were analysed subsequently.

***MTHFD1* is required for epigenetic silencing.** To get a general view of the DNA methylation defects in *mthfd1-1*, we analysed genome-wide DNA methylation at single-nucleotide resolution by BS-seq. The average global DNA methylation was reduced by ~40% relative to WT (Fig. 2a). The strongest effect was observed

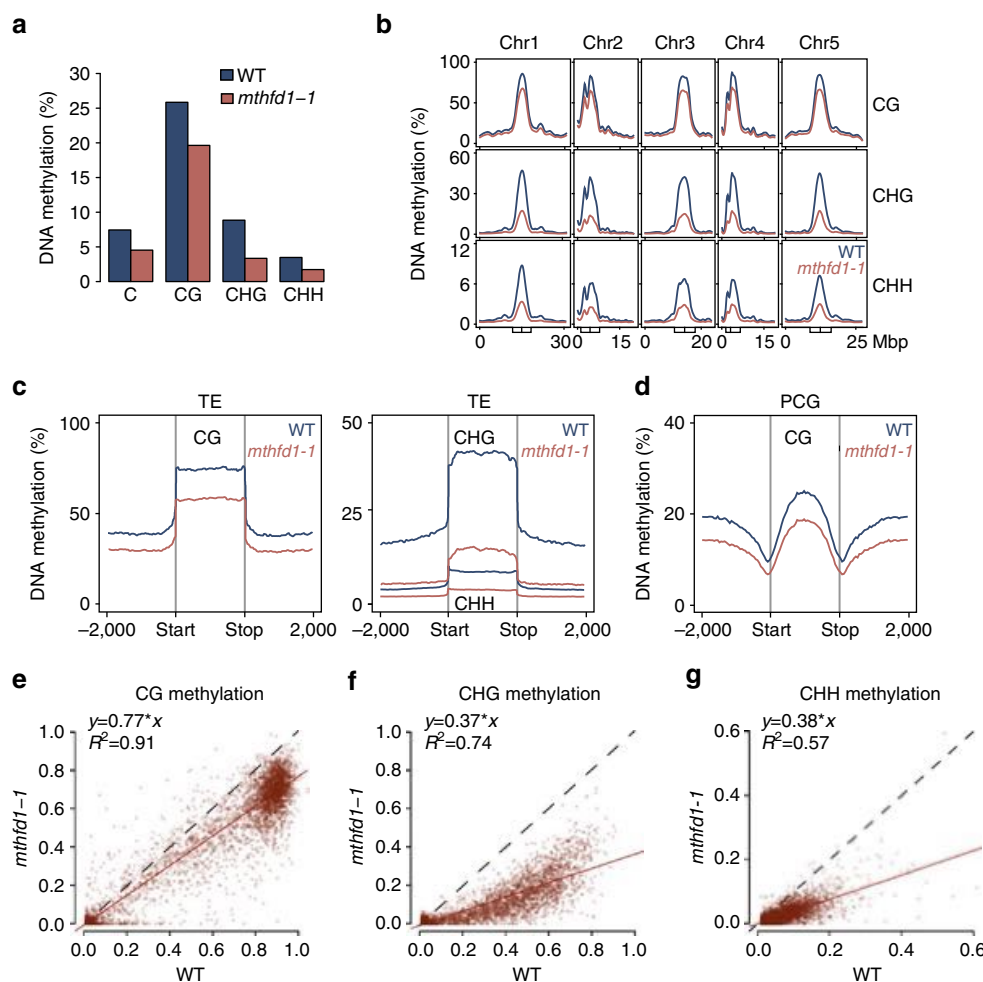


Figure 2 | DNA methylation is globally decreased in *mthfd1-1* mutants. (a) Average genome-wide DNA methylation for all Cs and in individual sequence contexts (H = C, A or T). (b) Chromosomal distribution of fractional DNA methylation in individual sequence contexts. Boxes and vertical lines inside boxes mark pericentromeric regions and centromeres, respectively. (c,d) Average distribution of DNA methylation over TEs (c) and PCGs (d) and the flanking 2,000 bp in individual sequence contexts. (e–g) Comparison of DNA methylation levels in 5,000 random 100 bp bins with WT methylation levels > 0.01 in CG (e), CHG (f) and CHH (g) contexts. Red line: linear regression between *mthfd1-1* and WT levels; corresponding coefficients are shown in top left corners. Dashed: identity line.

in the CHG context, which lost 62% of DNA methylation relative to WT, followed by CHH and CG methylation with 50% and 24% decreases, respectively (Fig. 2a). DNA methylation in all sequence contexts was mostly decreased over the TE-rich pericentromeric regions, which contain most of the DNA methylation along the chromosomes (Fig. 2b). Accordingly, average methylation levels over TEs were strongly decreased in *mthfd1-1* compared with WT, especially in CHG and CHH contexts (Fig. 2c). DNA methylation was also moderately decreased over protein-coding genes (PCGs) (Fig. 2d), indicating that MTHFD1 is not only required for repressive DNA methylation at TEs but also for efficient gene body CG methylation, although to a lesser degree, compared with non-CG methylation. The different effects on DNA methylation in the different sequence contexts was also apparent when comparing WT and *mthfd1-1* CG, CHG and CHH methylation levels in randomly selected 100 bp windows of the genome with methylation thresholds > 1% (in order to exclude unmethylated bins during randomization) (Fig. 2e–g). This comparison additionally shows a positive correlation of DNA methylation levels between WT and *mthfd1-1* in all three sequence contexts, which indicates that DNA methylation was decreased uniformly across the genome (Fig. 2e–g), as opposed to the DNA methylation patterns in loss-of-function DNMT

mutants, which show either nearly complete loss across the sample pool (Supplementary Fig. 5a,b), or at a subset of regions (Supplementary Fig. 5c).

To define the effect of the *mthfd1-1* mutation on the different DNA methylation pathways, we calculated differentially methylated regions (DMRs) that had decreased DNA methylation in *mthfd1-1* or the DNMT mutants compared with the WT reference (hypo-DMRs). The comparison of hypo-DMRs clearly showed that CMT3-dependent CHG methylation was the most affected, followed by CMT2- and DRM1,2-dependent CHH methylation (DRM1 is a lowly expressed paralog of DRM2 (ref. 16)), and finally MET1-dependent CG methylation that was the least affected in *mthfd1-1* (Fig. 3a–c). It is noteworthy that *mthfd1-1* DNA methylation levels were also decreased in regions that were only defined as DMRs in the DNMT mutants, but not in *mthfd1-1* (Fig. 3d–f). Moreover, all subsets of *mthfd1-1* hypo-DMRs showed residual DNA methylation, which supports a uniform genome-wide decrease in *mthfd1-1* (Fig. 3d–f). Heat maps of hierarchically clustered CHG and CHH hypo-DMRs further illustrated that DNA methylation levels in *mthfd1-1* are evenly decreased at moderate degrees and thus generally proportional to WT levels (Fig. 3g). The equal ratio of TE- versus PCG-overlapping CG hypo-DMRs in *met1* and *mthfd1-1* is also

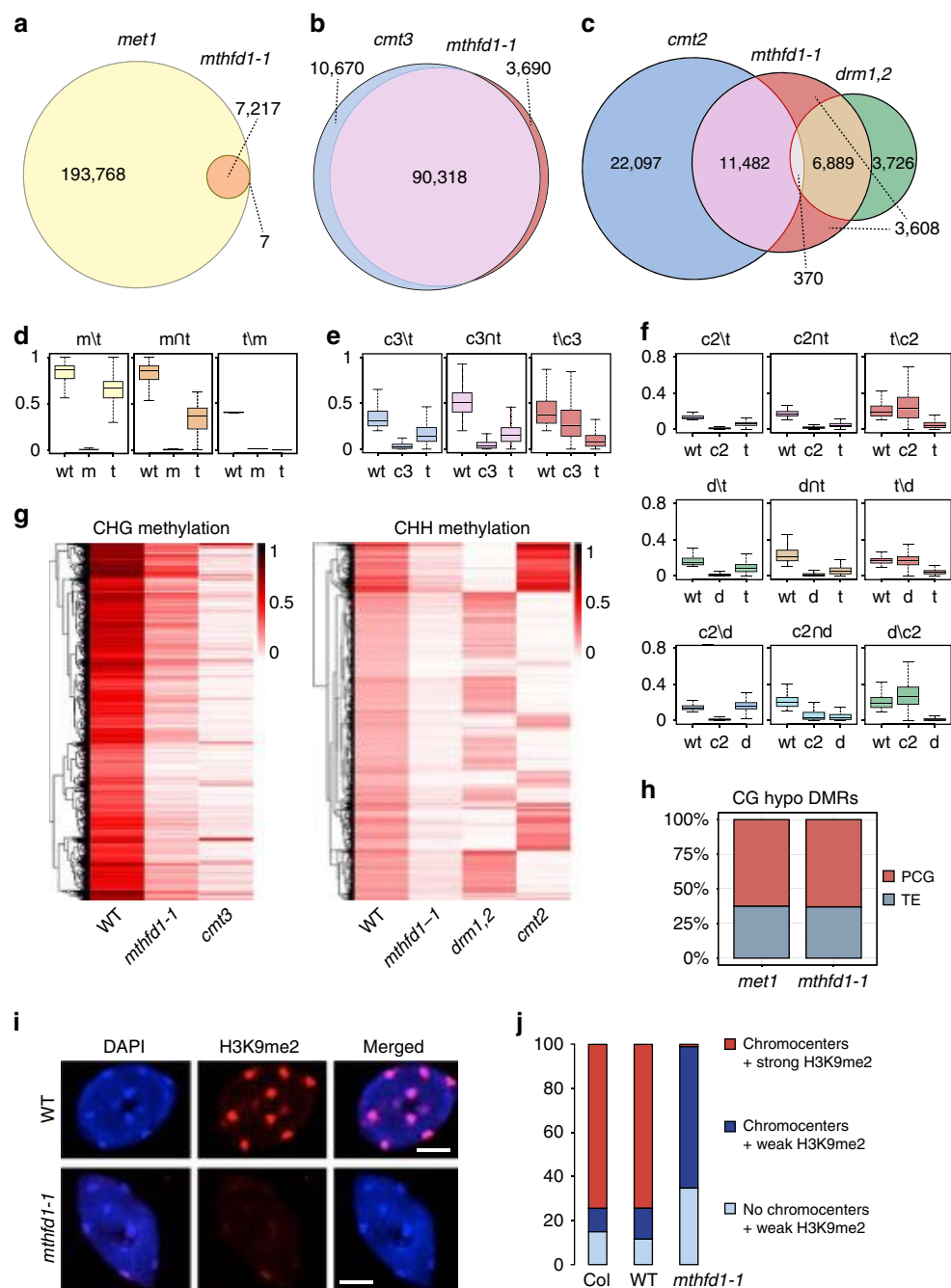


Figure 3 | *mthfd1-1* mostly interferes with non-CG and H3K9 methylation. (a–c) Overlap between hypo-DMRs of different mutants in CG (a), CHG (b) and CHH (c) contexts. (d–f) CG (d), CHG (e) and CHH (f) methylation levels in DMR fractions corresponding to (a–c), respectively. w = wild-type, m = *met1*, t = *mthfd1-1*, c3 = *cmt3*, c2 = *cmt2*, d = *drm1,2*. Box plot (herein and after): horizontal line, median; edges of boxes, 25th (bottom) and 75th (top) percentiles; error bars, minimum and maximum points within $1.5 \times$ interquartile range. (g) Heat map of DNA methylation levels in *mthfd1-1* hypo-DMRs (rows) clustered by methylation levels. (h) Overlap of *met1* or *mthfd1-1* CG hypo-DMRs with PCGs or TEs. (i) Fluorescence micrographs of representative nuclei from WT and *mthfd1-1*. DNA was stained with DAPI and H3K9me2 was immunostained using Alexa Fluor 647 as secondary antibody. Scale bar, 5 μ m. (j) Number of nuclei classified by DAPI staining and H3K9me2 immunofluorescence.

in agreement with a uniform decrease in genomic DNA methylation in *mthfd1-1* (Fig. 3h). Accordingly, the small subset of *met1* CG hypo-DMRs shared by *mthfd1-1* does not seem to represent site-specific MTHFD1 function, but is more likely due to stringent criteria for DMR calling, which were not met by the majority of sites in *mthfd1-1* despite reduced CG methylation levels (Fig. 3a,d). In the CHH context, RdDM and the CMT2 pathway were both affected by *mthfd1-1* (Fig. 3c,g). The chromosomal distribution of *mthfd1-1* CHH hypo-DMRs, with

high densities inside and at the peripheries of the pericentromeric regions, and lower densities in the chromosomal arms, reflects the overlaps with the alternative CHH methylation pathways (Supplementary Fig. 5d). In summary, the DNA methylation analysis of *mthfd1-1* revealed a rather uniform genome-wide decrease, which is most pronounced in the CHG and CHH context.

DNA methylation by CMT3, CMT2 and RdDM is functionally linked to histone methylation, because CMT2 and CMT3 directly

bind H3K9me in a feedback loop with the histone methyltransferases KYP/SUVH4, SUVH5 and SUVH6 (refs 5,6), and Pol IV is recruited to chromatin by K3K9me-binding SHH1 (ref. 10). To test if H3K9me is also affected in *mtfhd1-1*, we analysed H3K9 dimethylation (H3K9me₂) by immunofluorescence. Nuclei of *mtfhd1-1* mutants showed a strong decrease in H3K9me₂, but the majority of the nuclei still contained DNA-dense chromocenters visualized by DAPI staining (Fig. 3i,j). Therefore, the strong decrease in CHG and CHH methylation in contrast to the small decrease in CG methylation is likely explained by the combined effect of impaired DNA and H3K9 methylation.

To test if loss of DNA methylation in *mtfhd1-1* led to transcriptional derepression, we analysed the expression levels of different retrotransposons that were previously identified as upregulated in *drm1,2 cmt3* triple mutants¹³. Quantitative reverse transcription (RT)-PCR showed that these TEs are also strongly induced in *mtfhd1-1* (Fig. 4a and Supplementary Data 1). Transcriptome analysis of *mtfhd1-1* and WT by RNA-seq confirmed that average transcript levels over CHH and CHG hypo-DMRs are higher in *mtfhd1-1* than in WT (Fig. 4b). Many transcripts from TEs that were silenced in WT were highly abundant in *mtfhd1-1*, whereas transcriptional differences of PCGs were more even and showed a slight tendency towards higher transcript levels in WT (Fig. 4c–e). Correspondingly, pericentromeric regions showed many differentially upregulated TEs and—to a lesser degree—PCGs in *mtfhd1-1*, whereas chromosome arms contained approximately equal distributions of up- and downregulated PCGs (Fig. 4d,e). The differentially upregulated TEs in *mtfhd1-1* belonged to class I, as well as class II transposons. Among the differentially upregulated TEs, members of the LTR/Gypsy family were overrepresented, and members of the RC/Helitron family were under-represented compared with the genomic distribution (Supplementary Fig. 6). In summary, the transcriptional analyses have shown that the loss of DNA methylation in *mtfhd1-1* led to derepression of transposons and genes (Fig. 4g), predominantly in the pericentromeric region (Fig. 4d). The observed transcriptional changes in the chromosome arms seem to be mainly a consequence of pleiotropic effects of impaired MTHFD1 function. This is supported by an analysis of GO terms annotated to genes that are significantly downregulated in *mtfhd1-1* compared with WT (Fig. 4f and Supplementary Data 2). The 10 statistically most significantly enriched biological processes indicate that MTHFD1 serves important functions in sugar metabolism, isoprenoid synthesis, redox homeostasis and photosynthesis. Since *mtfhd1-1*-downregulated genes did not show a significant loss of DNA methylation, the overall effects on transcript abundance caused by decreased gene body methylation in *mtfhd1-1* are likely to be negligible (Fig. 4g).

***mtfhd1-1* mutants show accumulation of S-adenosylhomocysteine.**

Arabidopsis MTHFD1 contains two highly conserved protein domains, a catalytic domain in the N-terminal half, and a NAD(P⁺)-binding domain of the Rossmann fold superfamily in the C-terminal half (Fig. 1b). Therefore MTHFD1 is probably required for the interconversion of tetrahydrofolate (THF) species in one-carbon metabolism of *Arabidopsis*. Members of the bifunctional enzyme family catalyse the reversible interconversion of 5,10-methylenetetrahydrofolate (5,10-CH₂-THF) to 5,10-methenyltetrahydrofolate (5,10-CH=THF) (NADP⁺-dependent dehydrogenase activity) and further to 10-formyltetrahydrofolate (10-CHO-THF) (cyclohydrolase activity) (Fig. 5) (ref. 17). These enzymatic activities have previously been detected in plant extracts¹⁸. The *Arabidopsis* genome encodes four homologues, the mitochondrial MTHFD2/FOLD1 (AT2G38660), FOLD3

(At4g00600) and FOLD4 (At4g00620), which are putatively plastidic, and MTHFD1/FOLD2, which lacks an N-terminal targeting peptide and is presumably localized in the cytoplasm (Supplementary Fig. 7) (refs 19–21). We confirmed the subcellular localization of the latter two homologues by expression and *in vivo* imaging of the full-length fusion proteins MTHFD1-YPET-3xFLAG, MTHFD1_R175Q-YPET-3xFLAG and FOLD4-YPET-3xFLAG in *Nicotiana benthamiana* (Fig. 6a,b and Supplementary Fig. 8), showing that MTHFD1 is predominantly in the cytoplasm. The Met cycle is exclusively localized in the cytoplasm and is required for the synthesis of SAM, which serves as methyl donor for many transmethylation reactions, including those catalysed by histone and DNMTs (Fig. 5). During transmethylation SAM is converted to SAH, which is further processed into adenosine and Hcy by the SAH hydrolase (SAHH) (Fig. 5). Hcy is recycled to Met by 5-CH₃-THF-dependent transmethylation activity of methionine synthase and can serve for a new round of SAM synthesis (Fig. 5) (refs 14,22). To test if the DNA methylation defects in *mtfhd1-1* are caused by changes in the Met cycle, we analysed the levels of SAM, SAH and Hcy, as well as cysteine in *mtfhd1-1*, WT and Col leaves. SAM and SAH were both significantly increased in *mtfhd1-1*, but the stronger increase in SAH levels led to an overall decrease of the methylation index (MI = SAM/SAH) (Fig. 7a–c). MI is an important measure of the organismal methylation status, because SAH is a strong competitive inhibitor of SAM-dependent transmethylation²³. Because of the low intracellular concentration and high affinity for methyltransferases, it has been suggested that even small changes in the MI can lead to a reduction in transmethylation activity²⁴. Therefore, it is likely that the decreased MI in *mtfhd1-1* leads to decreased activities of DNA and histone methyltransferases, as reflected by the observed DNA and histone methylation defects. This is further supported by a 12-fold increase in Hcy (and an associated increase in cysteine levels) in *mtfhd1-1* (Fig. 7d), because Hcy accumulation leads to inhibition of SAH hydrolysis and consequently to a lower MI^{22,25}.

Regulation of folate homeostasis by MTHFD1. Analysis of folate metabolites in leaves of *mtfhd1-1* and WT showed that total folate content did not differ significantly, but *mtfhd1-1* had an 8.8-fold increase in THF + 5,10-CH₂-THF (these two compounds cannot be distinguished by the analysis). On the other hand, the levels of 5,10-CH=THF, which also include the 10-CHO-THF pool, were reduced by 30% and 5-CHO-THF levels were also ~33% lower (Fig. 7e). Unexpectedly, there was no significant difference in 5-CH₃-THF contents, which is the product of 5,10-CH₂-THF reductase that serves as co-substrate for the methylation of Hcy to Met and constitutes the most abundant active THF species²⁶. The low levels of the oxidized folates, 10-CHO-THF + 5,10-CH=THF and 5-CHO-THF, suggest an impairment in the dehydrogenase and cyclohydrolase activities towards the formation of 10-CHO-THF. On the basis of the crystal structure and site-directed mutagenesis of human MTHFD1, the conserved arginine that is mutated in *mtfhd1-1* (R175Q), forms a hydrogen bond with NADP⁺ and is required for dehydrogenase, but not cyclohydrolase activity^{27,28}. Therefore, we anticipate that the R175Q mutation has a similar effect in *mtfhd1-1* mutants. Accordingly, the accumulation of the reduced forms THF + 5,10-CH₂-THF in *mtfhd1-1* might reflect reduced conversion of 5,10-CH₂-THF to 5,10-CH=THF, which is in agreement with the homology-based prediction that the R175Q mutation affects NADP⁺ binding and dehydrogenase activity^{27,28}. The decreased pool of oxidized THFs further suggests that one-carbon supply from formate via FTHFS and reverse cyclohydrolase by MTHFD1 is not sufficient to

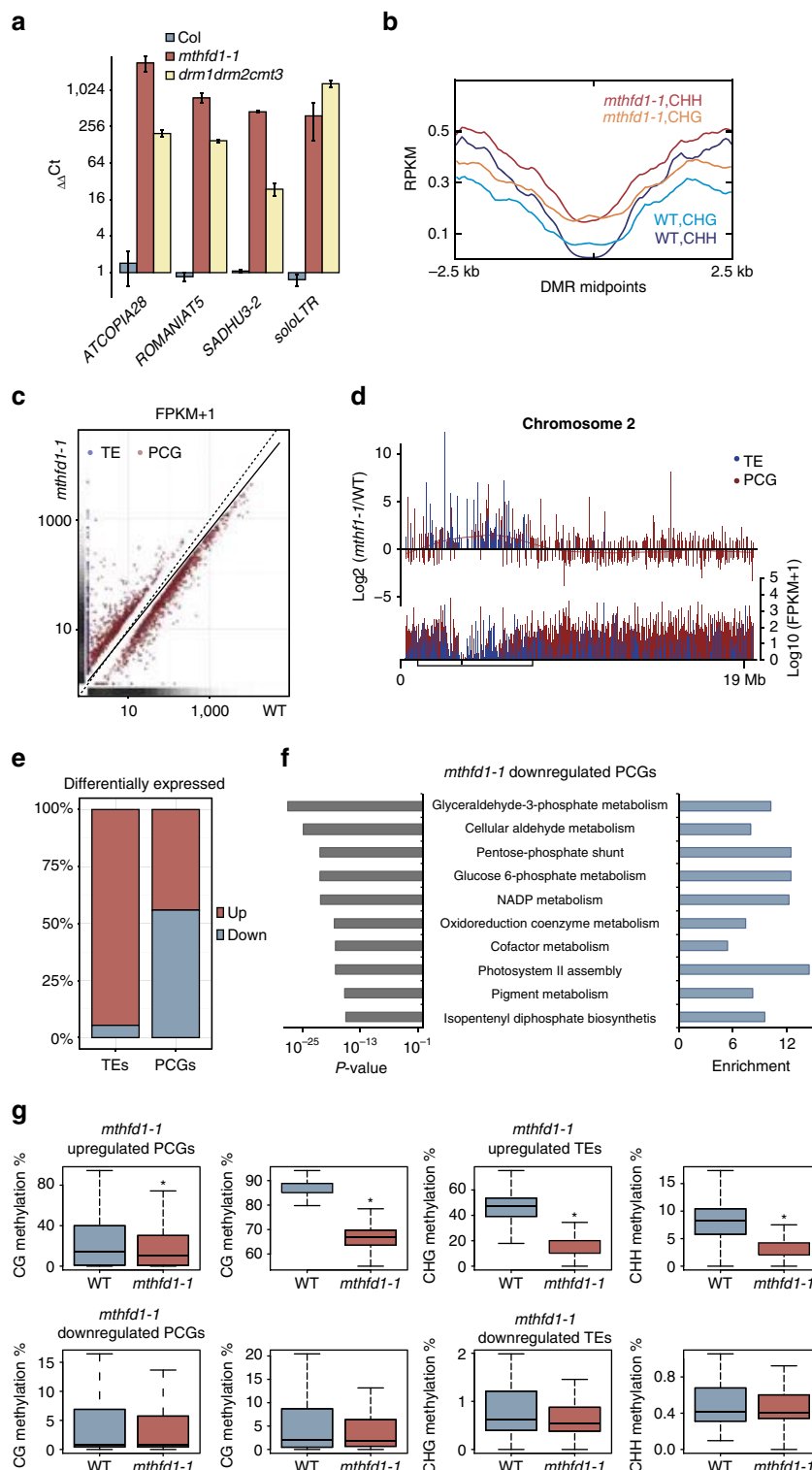


Figure 4 | *mthfd1-1* mutants show loss of TE silencing and pleiotropic transcriptional deregulation of PCGs. (a) Quantitative RT-PCR analysis of four exemplary TEs. Transcript levels are normalized to *ACTIN7* and relative to WT, and mean values $\pm$ s.e.m. ($n=3$) are shown. (b) Average distribution of normalized RNA-seq reads (RPKM) over *mthfd1-1* hypo-DMRs in CHG or CHH context. X axis indicates distance from the DMR midpoints. (c) Scatter plot showing normalized transcript levels (FPKM + 1) of differentially expressed PCGs and TEs in *mthfd1-1* versus WT. Solid line: linear regression through all TEs and PCGs. Dashed: identity line. Marginal density plots: distribution of all TEs and PCGs in WT (x axis) and *mthfd1-1* (y-axis). (d) Distribution of average normalized transcript levels (FPKM + 1) of all TEs and PCGs from *mthfd1-1* and WT (lower panel) and fold change in normalized transcript levels of differentially expressed PCGs and TEs (upper panel) along chromosome 2. Box and vertical line inside box mark pericentromeric region and centromere, respectively. (e) Fraction of TEs and PCGs significantly up- or downregulated in *mthfd1-1*. (f) Enrichment score and statistical significance of GO processes annotated to *mthfd1-1*-downregulated PCGs. (g) Methylation levels over *mthfd1-1* differentially expressed PCGs and TEs in different sequence contexts.

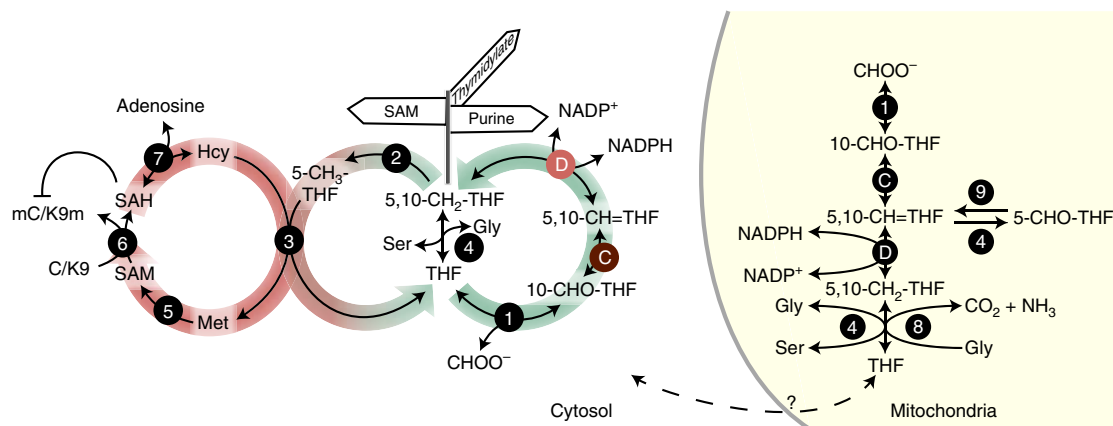


Figure 5 | Schematic representation of plant SAM and folate metabolism in the cytosol and mitochondria. One-carbon enters the cytoplasmic folate cycle (green) either through formyltetrahydrofolate synthetase (1) or SHMT (4); MTHFD1 reversibly interconverts 10-CHO-THF to 5,10-CH₂-THF by cyclohydrolase (C) and NADP⁺-dependent dehydrogenase (D) activity. 5,10-CH₂-THF serves for thymidylate synthesis or is converted by methylenetetrahydrofolate reductase (2) to 5-CH₃-THF, which enters the Met cycle (red) and serves for Hcy remethylation to Met by methionine synthase (3) (ref. 17). SAM synthetase (5) converts Met to SAM, which is further converted to SAH (6) during methylation of cytosines, H3K9 and so on. SAH is a competitive inhibitor of methyltransferases (6) and is recycled to Hcy by SAH hydrolase (7) (ref. 23). In mitochondria, one-carbon is transferred to THF by the glycine decarboxylase complex (8), but surplus of Gly due to photorespiration leads to consumption of one-carbon by SHMT during serine production³⁰. 5-CHO-THF, a byproduct of SHMT, is metabolized by mitochondrial 5-formyltetrahydrofolate cycloligase in order to re-enter the folate cycle (9) (ref. 59). Shuttle of THF between mitochondria and the cytosol has been described in other organisms, but remains uncharacterized in plants.

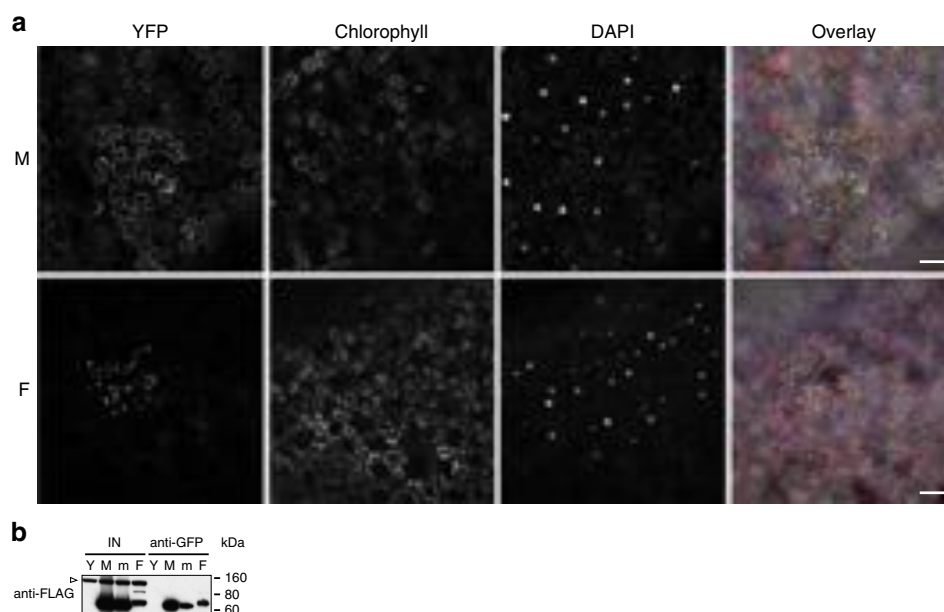


Figure 6 | Cytoplasmic localization of MTHFD1-YPET-3xFLAG. (a) Confocal micrographs of MTHFD1-YPET-3xFLAG (M) and FOLD4-YPET-3xFLAG (F) transiently expressed in *N. benthamiana*. Excitation (λ , nm)/filters (λ , nm): YFP = 514/519–559, chlorophyll = 488/630–730, DAPI = 405/409–530, and fluorescence overlay with bright field. Scale bars, 50 μ m. (b) Western blot using anti-FLAG antibody against anti-GFP-immunopurified extracts from *N. benthamiana* (IN) transiently expressing free YFP (Y), MTHFD1-YPET-3xFLAG (M), MTHFD1_R175Q-YPET-3xFLAG (m), or FOLD4-YPET-3xFLAG (F). Arrowhead indicates unspecific binding of anti-FLAG (shown as loading reference).

compensate for reduced MTHFD1 dehydrogenase activity (Fig. 5). This is in accordance with metabolic analyses showing relatively low one-carbon flow from formate to serine (Ser)²⁹.

Ser can serve as a one-carbon source through the reversible enzymatic activity of serine hydroxymethyltransferase (SHMT), which converts Ser and THF to 5,10-CH₂-THF and glycine (Gly) (Fig. 5). In addition, the Gly decarboxylase complex converts Gly and THF to 5,10-CH₂-THF, carbon dioxide and ammonia during photorespiration, which in turn can lead to Ser synthesis by

SHMT in the mitochondria³⁰ (Fig. 5). We analysed amino-acid levels in rosette leaves and found a threefold increase of Gly levels in *methfd1-1* compared with WT and Col controls, whereas Ser levels were only slightly increased in *methfd1-1* (Fig. 7f). In accordance with our folate analysis, the lower Ser/Gly ratio in *methfd1-1* might be due to increased SHMT activity towards 5,10-CH₂-THF formation¹⁹. Furthermore, *methfd1-1* mutants showed a 5.4-fold and 1.9-fold increase in proline and Met levels, respectively (Supplementary Fig. 9). The increase in Met might be due to increased *de novo* synthesis (which occurs

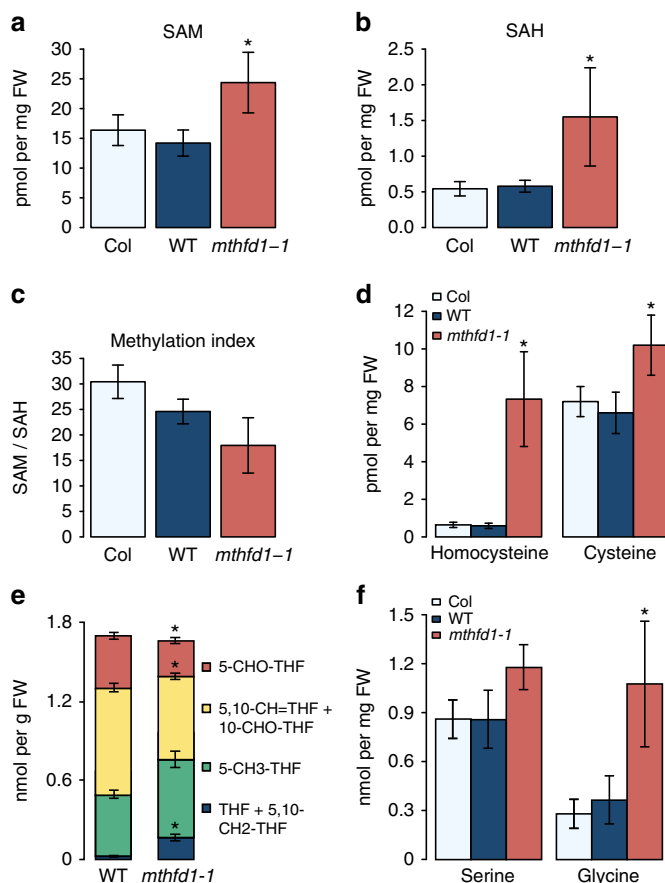


Figure 7 | *mthfd1-1* mutants show impaired one-carbon cycle.

(a–f) Steady-state levels of SAM (a), SAH (b), Methylation Index (MI) (c), selected thiols (d), folates (e) and selected amino acids (f) in leaves of Col, WT and the *mthfd1-1* mutant. Data represent means $\pm$ SD. Asterisks indicate significant differences determined by Student's *t*-test ($P < 0.05$, $n \geq 3$).

exclusively in plastids) in response to impaired Met recycling, whereas proline accumulation is typically observed as a stress response³¹.

We sought to manipulate one-carbon metabolism in *mthfd1-1* by exogenous application of metabolites involved in the folate and Met cycle and monitor root growth and DNA methylation effects. Previously it has been shown that growth phenotypes, DNA methylation and epigenetic silencing defects caused by decreased activated THF pools because of impaired plastidic folylpolyglutamate synthetase (FPGS1), or chemical inhibition of folate synthesis by sulfamethazine (SMZ) were complemented by exogenous application of 5-CHO-THF or Met^{32–35}. Application of 5-CHO-THF, as well as 5-CH3-THF, tests for defects in *mthfd1-1* caused by reduced folate availability before the flow of one-carbon into the Met cycle, whereas application of Met tests for defects in the Met cycle (Fig. 5), and SMZ was used as a control. In contrast to previously described complementation of *fpgs1* mutants, 5-CHO-THF strongly inhibited root growth of *mthfd1-1* seedlings without showing an inhibitory effect on WT seedlings (Fig. 8a). Although analysis of global DNA methylation levels revealed that these are largely independent of the observed root growth responses, average *mthfd1-1* CHG methylation over the previously defined CHG hypo-DMRs significantly increased upon Met application. On the other hand, 5-CHO-THF, as well as 5-CH3-THF, did not rescue the DNA methylation defects in *mthfd1-1* (Fig. 8b). These results indicate that, in contrast to the *fpgs1* mutants and SMZ inhibition^{34,35}, the DNA methylation

defects in *mthfd1-1* are probably not the mere result of diminished folate pools, but rather point towards an inhibition of methionine synthase.

Discussion

Methylation patterns in the *Arabidopsis* genome are remarkably stable not only from one generation to the next but also at evolutionary timescales^{36–39}. Comparative genomics and genome-wide association studies have linked DNA methylation and phenotypic variation in Brassicaceae to genetic polymorphisms in the DNA methylation machinery^{40,41}, and consequently support an adaptive role of spontaneous epigenetic changes. For example, two independent studies have revealed that different alleles of *CMT2* and the concomitant differences in CHH methylation are associated with climate adaptation^{40,42}. Here we have identified an EMS-induced polymorphism in the essential folate metabolic enzyme MTHFD1 from *Arabidopsis*, which causes a strong, genome-wide decrease in DNA methylation. This finding highlights that DNA methylation patterns in *Arabidopsis* not only depend on the pathways and catalytic activities of the DNMTs but also on the metabolic network that regulates the availability of the methyl donor SAM and the adequate functioning of the activated methyl cycle. It is therefore conceivable that regulatory mechanisms have evolved, which connect nutritional changes to epigenetic gene regulation by DNA and histone methylation. Although direct examples in plants are still lacking, it has been shown that a folate-rich diet in mice leads to changes in coat colour of the offspring that is caused by altered expression of the *agouti* gene due to increased DNA methylation of a transposon in the *agouti* locus (*A^y*)⁴³. This finding illustrates an example of how such regulatory mechanisms could work.

The EMS-allele *mthfd1-1* was identified through a genetic screen for mutants that simultaneously affect CHG and CHH methylation. Correspondingly, our genome-wide BS-seq analysis of *mthfd1-1* mutants revealed extensive hypomethylation in CHG and CHH sequence contexts. In contrast, loss of CG methylation was comparatively low. We therefore reason that the feedback regulation between CHG/CHH and H3K9 methylation is particularly prone to changes in one-carbon metabolism, because transmethylation by DNA and histone methyltransferases are both SAM-dependent. Accordingly, *mthfd1-1*, as well as previous analyses of *fpgs1* mutants and plants treated with SMZ^{34,35}, showed reduced H3K9me2. Because of the mechanistic interdependence of non-CG and H3K9 methylation⁵, it is difficult to tell whether histone or DNA methylation is more directly affected by impaired one-carbon metabolism. The predominant effect on CHG methylation might suggest that loss of H3K9me2 is the primary defect. However, the fact that CG methylation is also decreased in *mthfd1-1*, a type of methylation that is not linked with H3K9 methylation, suggests a general inhibition of the enzymatic activity of different methyltransferases, including MET1.

Our transcriptome analysis demonstrated that loss of DNA methylation in *mthfd1-1* mutants leads to derepression of TEs and a generally higher abundance of transcripts from the pericentromeric heterochromatin compared with WT. With respect to the almost equal numbers of up- and downregulated PCGs, loss of gene body methylation can only account for some of the observed changes, whereas the majority of differential gene expression is probably caused by pleiotropic effects of impaired MTHFD1 function. Secondary to its role in one-carbon metabolism, a shortfall of NADP⁺ conversion by MTHFD1 is expected to severely disturb the redox state⁴⁴. Accordingly, GO term analysis revealed a significant enrichment of genes involved in cell redox homeostasis, oxidative pentose phosphate pathway and glycolysis (Fig. 4f and Supplementary Data 2) (ref. 45).

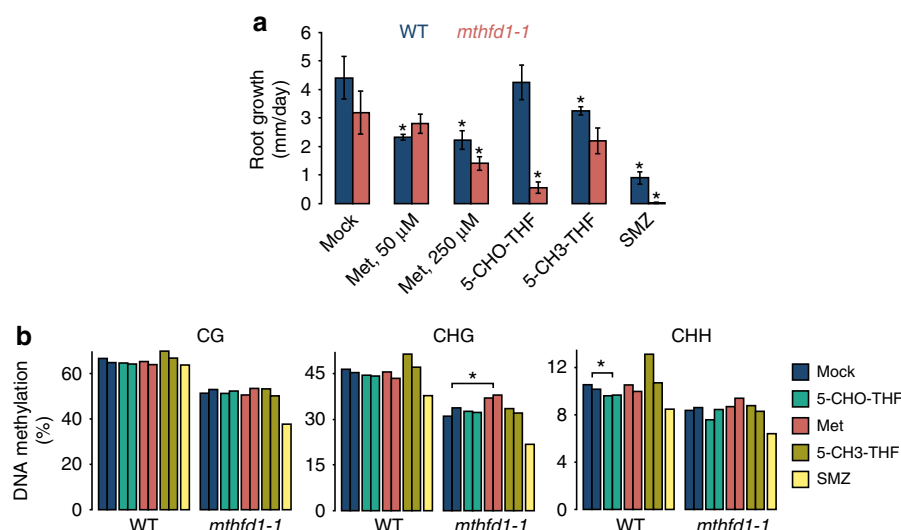


Figure 8 | *mthfd1-1* mutants are hypersensitive to exogenous 5-CHO-THF but tolerant to exogenous methionine. (a) Root growth of seedlings on solid media containing mock, Met, 0.5 mM 5-CHO-THF, 0.5 mM 5-CH3-THF or 5 μ M SMZ. Mean values $\pm$ s.d. ($n=3$) are shown. (b) Average CG, CHG and CHH methylation levels at previously defined *mthfd1-1* CG, CHG and CHH hypo-DMRs, respectively, in seedlings grown for 14 days on solid media containing mock, 0.1 mM Met, 0.5 mM 5-CHO-THF, 0.5 mM 5-CH3-THF or 5 μ M SMZ. Two biological replicates are shown, except for SMZ. * indicates significant difference between mock and chemical treatment ($P<0.05$, Student's *t*-test).

MTHFD1 related proteins in different species have mono-, bi- or trifunctional enzymatic activity. Yeast and mammalian cytosolic homologues, known as C1-THF synthases, are trifunctional and reversibly catalyse the stepwise oxidation from 5,10-CH₂-THF to 10-CHO-THF, which serve for thymidylate/pantothenate and *de novo* purine/*N*-formylmethionine synthesis, respectively, and the conversion of 10-CHO-THF to THF and formate (reverse FTHFS activity) (Fig. 5) (refs 14,46,47). Bifunctional forms, which are found in certain bacteria and in plants, lack the FTHFS activity^{18,48,49}. Moreover, methylenetetrahydrofolate reductase converts 5,10-CH₂-THF to 5-CH₃-THF and thereby directs activated methyl towards SAM (Fig. 5). As such, the reversible enzymatic activity of MTHFD1 channels one-carbon into different pathways and acts as a crucial regulatory hub (Fig. 5). Correspondingly, functional mutations, such as in *mthfd1-2*, have severe pleiotropic effects and are mostly lethal. This is in contrast to the subtle morphological defects generally observed in epigenetic *Arabidopsis* mutants, for example, *drm1 drm2 cmt3* triple knockout mutants⁵⁰, and denotes that inhibition of MTHFD1 leads to pleiotropic morphological defects that are independent of its impact on DNA and histone methylation. The essential nature of MTHFD1 further indicates that the additional three *Arabidopsis* MTHFD homologues have plastid- and mitochondrion-specific functions that cannot compensate for a loss of cytoplasmic MTHFD1 function.

Because of its role in nucleotide biosynthesis and DNA methylation, folate metabolism is of central relevance in cancer research, as exemplified by the therapeutic use of antifolates⁵¹. Polymorphisms in human MTHFD1 C1-THF synthase have been associated with cancers, as well as neural tube defects and other illnesses⁵¹. Interestingly, one polymorphism (R173C) resides in the same conserved residue that is mutated in the EMS-allele *mthfd1-1* and was linked to severe combined immunodeficiency, megaloblastic anaemia and altered Met metabolism, including Hcy accumulation⁵². Analyses of fibroblasts harbouring this mutation showed signs of DNA damage and uracil misincorporation into DNA due to impaired *de novo* thymidylate synthesis⁵². Interestingly, the mutation had the strongest impact on one-carbon flow towards the Met cycle⁵². Impaired dehydrogenase activity was partially compensated by

increased SHMT activity, as well as increased salvage thymidylate synthesis, whereas *de novo* purine synthesis was not affected⁵². The study did not include DNA methylation analyses, but given the conserved function of MTHFD1 and shared effects on Hcy remethylation, our results predict that DNA methylation is affected by the MTHFD1 R173C mutation and might be involved in certain types of severe combined immunodeficiency and megaloblastic anaemia. In reverse, analogous redirection of one-carbon flow towards nucleotide synthesis at the expense of Hcy remethylation is a possible explanation of the defects observed in *mthfd1-1* and was also suggested to occur upon methotrexate-induced THF depletion in *Arabidopsis*, based on folate measurements and transcriptional analyses⁵³.

Hcy accumulation and decreased MI, as observed in *mthfd1-1* and previous studies^{32–35}, are hallmarks of impaired Hcy remethylation due to impaired folate metabolism. Increased Hcy levels lead to decreased SAH hydrolase activity and accumulation of SAH, which competitively inhibits SAM-dependent transmethylation, including DNA and histone methylation^{54,55}.

Decreased flux of one-carbon towards the remethylation of Hcy should intuitively lead to decreased levels of Met, yet we observed increased cellular Met in *mthfd1-1*. However, primary metabolites, in particular the sulfur amino acids Cys and Met, are often controlled by multiple layers of regulatory circuits, as exemplified by the *sir1-1* mutant, which also show an increased Met steady level despite a 20-fold decreased flux of sulfur through the assimilatory sulfate reduction pathway⁵⁶. This increase was the result of decreased flux into Met sinks, due to an attenuation of translation and growth in *sir1-1* (ref. 56). To that effect, it is conceivable that Met levels were increased in the *mthfd1* mutant because histone and DNA methylation (two major one-carbon sinks) and growth were decreased. Moreover, increased Met levels might also be due to increased Met *de novo* synthesis. Accordingly, the transcriptome analysis of *mthfd1-1* shows a 4.3-fold increase in transcripts corresponding to *MRU1* (At5g35490), which is also upregulated in the Met over-accumulating mutant *mtol-1* (ref. 57).

In the cases of chemically inhibited THF synthesis or impaired THF polyglutamylation, exogenous application of 5-CHO-THF,

which is readily assimilated and metabolized to active THF forms in *Arabidopsis*^{53,58}, successfully reversed the feedback inhibition of transmethylation and TGS^{34,35}. Interestingly, 5-CHO-THF feeding to *mtfhd1-1* mutants did not complement the DNA methylation defect and had a strong adverse effect on root growth. This hypersensitivity could be attributable to an inhibitory effect of 5-CHO-THF on SHMT⁵⁹. As in R173C fibroblasts⁵², it is likely that the supply of 5,10-CH₂-THF in *mtfhd1-1* mutants depends on cytosolic SHMT. Although cell compartmentalization demands a cautious interpretation of the metabolic profiles, the observed decrease in 5-CHO-THF levels by 33% in *mtfhd1-1* versus WT might have led to an increase in SHMT activity. Since the reaction equilibrium catalysed by SHMT favours Gly production⁶⁰, increased SHMT activity might have contributed to the threefold increase in steady-state Gly levels observed in *mtfhd1-1* versus WT. An inhibition of SHMT by exogenously applied 5-CHO-THF would accordingly cut off the cytosolic 5,10-CH₂-THF supply and explain the enhanced root growth defect in *mtfhd1-1* mutants. It is noteworthy that we did not observe an enhanced DNA methylation defect upon 5-CHO-THF feeding, which suggests that even under normal growth conditions SHMT-dependent 5,10-CH₂-THF production is unable to perpetuate the Met cycle in *mtfhd1-1*. This is further supported by the lack of phenotypic rescue of *mtfhd1-1* by exogenous 5-CH₃-THF. On the other hand, exogenous Met partially restored global CHG methylation, which together with the folate quantifications and feeding experiments suggests that transmethylation in *mtfhd1-1* is impaired due to an inhibition of Hcy remethylation, as opposed to limited availability of folate intermediates^{34,35} or inhibition of SAHH⁵⁴. As such, the described DNA methylation and gene regulatory defects in *mtfhd1-1* highlight a central regulatory role of MTHFD1 in one-carbon distribution towards different cell physiological processes.

Methods

Plant material. All plants used in this study were of the Columbia-0 ecotype. T-DNA insertion mutants *mtfhd1-2* (WiscDsLox244C04), *mtfhd1-3* (SALK_015165) and *mtfhd1-4* (SALK_039538) were obtained from the *Arabidopsis* Biological Research Center (Ohio State University). Genotypes were analysed by PCR using primers listed in Supplementary Table 2. The triple mutant *drm1 drm2 cmt3* and the WT transgenic line carrying the *SDC_{pro}-GFP* fusion construct were published previously^{4,13}. *mtfhd1-1* mutants have been backcrossed with WT plants carrying *SDC_{pro}-GFP*. Plants were grown in the greenhouse at long day light cycles, unless stated differently.

Genetics screening and mapping analyses. WT seeds (2,000) were suspended in 0.3% EMS solution for 13 h with rotation, washed with water and planted on soil. Approximately 1,000 independent M2 populations were collected and screened for GFP fluorescence using a Leica MZ16F Fluorescence Stereomicroscope equipped with a GFP Plus filter. Pictures were taken with a DFC300 FX digital camera. For mapping and identification of EMS mutations, mutant #162 was crossed with WT *Ler* and 10-days-old F2 seedlings grown on media containing 1 × Murashige and Skoog basal salt mixture (MP) and 20 μg ml⁻¹ glufosinate ammonium (Sigma) were analysed for GFP expression. Genomic DNA was isolated from pooled tissue of 50 GFP-positive F2 mutants and analysed by whole-genome re-sequencing for co-segregating single-nucleotide polymorphisms between Col and *Ler*¹³. Primer sequences of CAPS markers for co-segregation analyses are shown in Supplementary Table 2.

Local DNA methylation analyses. Genomic DNA was isolated from aerial tissue of 4–5-weeks-old plants. The *MEA-ISR* probe for DNA blot analysis was amplified using primers JP980 and JP981 (Supplementary Table 2) (ref. 15). Vertically uncropped images of all blots and gels shown in this study are provided in Supplementary Fig. 10. Chop-PCR analysis of *AtSN1* was performed by real-time PCR using primers JP6349 and JP6350 (Supplementary Table 2) (ref. 61). For DNA methylation analysis of the transgenic *SDC* promoter, DNA was BS converted using EZ DNA Methylation Gold kit (Zymo Research) and PCR amplified using primers listed in Supplementary Table 2. PCR fragments were cloned into pCR2.1-TOPO (Thermo Fisher Scientific), and 20 clones per genotype were sequenced.

Whole-genome bisulfite sequencing. Genomic DNA was extracted from rosette leaves of 3-weeks-old plants using DNeasy Plant Mini Kit (Qiagen) and fragmented into 200 bp average size with a Covaris S2 sonicator. Next, fragmented DNA was end repaired, adenylated and ligated with TruSeq DNA LT adapters (Illumina) using NEBNext DNA library prep reagent set (NEB). Subsequently, BS conversion was performed with CpGenome DNA modification kit (Millipore). Libraries were amplified using PCR primer cocktail (Illumina) and Pfu Turbo Cx hotstart DNA polymerase (Agilent). Sequencing was performed on a HiSeq 2000 platform at 50 bp length. Identical reads were removed and unique reads were aligned to the *Arabidopsis* reference genome (TAIR10) using BSMAP 2.87 (ref. 62). Read statistics are listed in Supplementary Table 3. Data for mutants other than the *mtfhd1-1* were obtained from GSE39901 (ref. 63). Methylation levels were calculated as #C/(#C + #T). DMRs were defined by dividing the genome into 100 bp bins and comparing mutants and WT by the number of methylated and unmethylated Cs with at least four Cs covered using Fisher's exact test and cutoffs of Benjamini-Hochberg corrected false discovery rate < 0.01. Moreover, absolute methylation difference of each bin had to be at least 0.4, 0.2 and 0.1 for CG, CHG and CHH, respectively. Heat maps of DMRs were generated by 'pheatmap' package in R software and clusters were grouped by the complete linkage method with Euclidean distance measurement. Venn diagrams were generated by calculating overlaps of 100 bp DMRs.

RNA analyses. Total RNA was isolated with TRIzol (Thermo Fisher Scientific) from 0.1 g of rosette leaves from 3-weeks-old plants. For real-time RT-PCR analysis, 2 μg of DNase I-treated total RNA were reverse-transcribed with SuperScript III (Thermo Fisher Scientific) and cDNA was amplified at target loci (primers listed in Supplementary Table 2) using iQ SYBR Green Supermix (Bio-Rad) and a Mx3005P qPCR system (Agilent Technologies).

For RNA-seq analysis, unstranded libraries from poly-A-tailed RNA were generated according to the manufacturer's instructions (Illumina TruSeq) and sequenced with the HiSeq 2,000 platform at 50 bp length. Reads were mapped to the TAIR10 genome with TopHat2 (ref. 64) using default settings, except that intron length was set to 40–5,000. Read statistics are listed in Supplementary Table 3. Fragments per kilobase of exon per million fragments mapped (FPKM) values and differential gene expression were analysed with Cufflinks⁶⁵ using default settings, except that maximum intron length was set to 5,000 and the -u option was used. The reference annotation for Cufflinks analysis was downloaded from TAIR and combined genes, including pseudogenes and TE genes, and TEs. GO term enrichment in genes that were significantly down regulated in *mtfhd1-1* compared with WT by at least twofold was analysed with GOrilla⁶⁶, using all *Arabidopsis* PCGs as background list.

Immunofluorescence analysis. Nuclei from rosette leaves of 3-weeks-old plants were immunostained with anti-H3K9me2 primary (Abcam ab1220, 5 μg ml⁻¹) and Alexa Fluor 647-conjugated anti-mouse IgG secondary (Thermo Fisher Scientific A-31571, 10 μg ml⁻¹) antibodies, and counterstained with DAPI (1 μg ml⁻¹) (ref. 13). Stained nuclei were imaged with a LSM 710 confocal microscope (Zeiss), with a C-Apochromat × 40/1.2 W Corr M27 objective and detection at λ (nm) = 410–504 (DAPI) and λ (nm) = 653–680 (Alexa Fluor 647).

Subcellular localization. To generate C-terminally tagged translational fusion proteins MTHFD1-YPET-3xFLAG, MTHFD1_R175Q-YPET-3xFLAG and FOLD4-YPET-3xFLAG, genomic DNA from Col and *mtfhd1-1* was amplified with primer pairs JP14184/5 and JP14190/1 (Supplementary Table 2), spanning the entire ORF (excluding Stop) and 1147 and 866 bp 5' of the ORF of *MTHFD1* and *FOLD4*, respectively. The amplified products were digested with XhoI & SpeI or SalI & SpeI and ligated with the plasmid pBJ36 (ref. 67), which has been linearized with XhoI & XbaI and contained an insertion of *YPET-3xFLAG* on the 3'-side of the XbaI site. NotI fragments from the resulting plasmids were inserted into the NotI site of the binary vector pMLBART⁶⁷. Overnight cultures of transformed *Agrobacterium tumefaciens* strain ASE were adjusted to OD₆₀₀ = 0.3 and coinfiltrated with p19 into *N. benthamiana* leaves⁶⁸. Leave discs were imaged 4 days after infiltration with a LSM 710 confocal microscope (Zeiss), using a Plan-Apochromat × 20/0.8 M27 objective and sequential scanning at excitation/detection λ (nm) = 514/519–559 (YFP), 488/630–730 (chlorophyll) and 405/409–530 (DAPI).

Immunopurification and western blot analysis. Agro-infiltrated *N. benthamiana* leaves (0.5 g) were ground in liquid nitrogen, and ground tissue was resuspended in 3 ml of IP buffer (50 mM Tris pH7.6, 150 mM NaCl, 5 mM MgCl₂, 5% (vol/vol) glycerol, 1% Tergitol (Type NP-40, Sigma), 2.8 mM β-mercaptoethanol, 1 μg ml⁻¹ pepstatin, 1 mM PMSF and 1 × protease inhibitor mixture tablet (Roche, 14696200)). Cleared lysates were incubated with 4 μl of anti-GFP antibody (A-11122, Molecular Probes), followed by 50 μl of Dynabeads Protein G (Thermo Fisher Scientific) at 4 °C for 1 h each. Western blotting was performed with anti-FLAG M2-Peroxidase (horseradish peroxidase) antibody (A8592-1MG, Sigma, 1:7,500 dilution).

Quantification of metabolites. Thiols, amino acids and adenosine nucleotides were extracted with 0.5 ml of 0.1 M hydrochloric acid from 0.1 g of in liquid nitrogen grinded rosette leaves from 4-weeks-old plants ($n=6$) grown at long day light and 21 °C. Amino acids and thiol were labelled with AccQ-Tag (Waters) and monobromobimane (Callbiochem), respectively, and quantified after separation by reverse phase chromatography⁶⁹. SAM and SAH were converted by chloroacetaldehyde treatment to their fluorescent etheno-derivates and quantified according to Burstenbinder *et al.*⁷⁰ after separation on a Gemini-NX C18 column (150 × 3 mm, 5 µm, 110 Å, Phenomenex, Germany) connected to a Waters 600 HPLC system with a flow rate of 1 ml min⁻¹ using the following gradient: 5 min 100% buffer A (50 mM tri-sodium phosphate deodecahydrate, 10 mM sodium 1-heptane sulfonate, 4% acetonitrile, pH 3.2); linear gradient for 15 min to 15% buffer B (pure acetonitrile); 7 min linear gradient to 90% buffer B; and 3 min 90% buffer B followed by re-equilibration of the column in 100% buffer A for 20 min.

Folate analysis. *Arabidopsis* rosette leaves (~0.15 g) were pulverized in a mortar with addition of liquid N₂ and homogenized with 10 ml folate extraction buffer (50 mM HEPES, 50 mM CHES, 10 mM β-mercaptoethanol, 2% Na-ascorbate (p/v), pH 7.9). The extracts were deglutamylated with a recombinant conjugase from *Arabidopsis* (100 µg AtGGH2 g⁻¹ sample) for 1 h at 37 °C. Foliates were purified by affinity chromatography using folate-binding columns. Purified folates were separated by liquid chromatography (Agilent Technologies, Santa Clara, CA, USA) using a Prodigy ODS(2) column (150 × 3.2 mm; 5 µm particle size) (Phenomenex, Torrance, CA, USA) with a 33 min nonlinear gradient of phase A (28 mM K₂HPO₄, 59 mM H₃PO₄) and phase B (75% phase A, 25% acetonitrile): 10% B (0–2 min); 10–20% B (2–4 min); 20–47% B (4–20 min); 47–80% B (20–25 min); 100% B (25–30 min); and 10% B (30–33 min) with a 1 ml min⁻¹ flow. Folate derivatives were detected by a four-channel electrochemical detector (CoulArray Model 5600A, ESA, Massachusetts, USA) with potentials set at 100, 200, 300 and 400 mV. THF, 5-methyl-THF, 5,10-methenyl-THF and 5-formyl-THF were quantified using calibration curves made with standards obtained from Schircks (Schircks Laboratories, Buechstrasse, Jona Switzerland). Because of the acidic pH of the mobile phase, in these analyses, THF represents THF + 5,10-methylene-THF and 5,10-CH = THF comprises 5,10-CH = THF + 10-CHO-THF³³.

Root growth assays and global DNA methylation analyses. Seeds were germinated on Phyto agar (RPI Corp.) containing 1 × Murashige and Skoog Basalt Salt Mixture (MP) and 500 µM (6R,S)-5-CHO-5,6,7,8-THF calcium salt (Schircks Laboratories); 50, 100 or 250 µM L-methionine (SIGMA); 5 µM SMZ (SIGMA); 500 µM (6R,S)-5-CH₃-5,6,7,8-THF calcium salt (Schircks Laboratories), which has been re-applied directly to the roots every 3 days due to its instability; or mock. Seedlings were grown vertically at 16 h light/ 8 h dark cycles and 22 °C. Root lengths were measured at 6, 8, 10, 12 and 14 days after germination and growth rates were calculated by linear regression from at least 10 seedlings per genotype per replicate. For measurement of global DNA methylation, seedlings were pooled per genotype and isolated genomic DNA was analysed by BS-seq as described above, except that EZ DNA Methylation-Lightning kit (Zymo Research) was used for BS conversion. DNA methylation levels were calculated as $\%C/(\%C + \%T)$ at CG, CHG and CHH sites and averaged over previously defined *mtf1d1-1* DMRs. Read statistics are listed in Supplementary Table 3.

Data availability statement. Primary high-throughput sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) with the accession code GSE77966 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77966>).

Secondary high-throughput sequencing data that support the findings of this study are available in the Gene Expression Omnibus (GEO) with the accession code GSE39901 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39901>).

The authors declare that all other relevant data supporting the findings of this study and computer code are available within the article and its Supplementary Information files or on request.

References

- Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
- Cao, X. & Jacobsen, S. E. Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr. Biol.* **12**, 1138–1144 (2002).
- Woo, H. R., Dittmer, T. A. & Richards, E. J. Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. *PLoS Genet.* **4**, e1000156 (2008).
- Johnson, L. M. *et al.* The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr. Biol.* **17**, 379–384 (2007).
- Du, J. *et al.* Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**, 167–180 (2012).
- Stroud, H. *et al.* Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014).
- Zemach, A. *et al.* The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
- Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* **15**, 394–408 (2014).
- Zhong, X. *et al.* Molecular mechanism of action of plant DRM *de novo* DNA methyltransferases. *Cell* **157**, 1050–1060 (2014).
- Law, J. A. *et al.* Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**, 385–389 (2013).
- Johnson, L. M. *et al.* SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**, 124–128 (2014).
- Henderson, I. R. & Jacobsen, S. E. Tandem repeats upstream of the *Arabidopsis* endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev.* **22**, 1597–1606 (2008).
- Moissiard, G. *et al.* MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* **336**, 1448–1451 (2012).
- Hanson, A. D. & Roje, S. One-carbon metabolism in higher plants. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **52**, 119–137 (2001).
- Cao, X. & Jacobsen, S. E. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc. Natl Acad. Sci. USA* **99**, 16491–16498 (2002).
- Cao, X. *et al.* Conserved plant genes with similarity to mammalian *de novo* DNA methyltransferases. *Proc. Natl Acad. Sci. USA* **97**, 4979–4984 (2000).
- Cossins, E. A. & Chen, L. Foliates and one-carbon metabolism in plants and fungi. *Phytochemistry* **45**, 437–452 (1997).
- Kirk, C. D., Chen, L., Imeson, H. C. & Cossins, E. A. A 5,10-methylenetetrahydrofolate dehydrogenase: 5,10-methylenetetrahydrofolate cyclohydrolase protein from *Pisum sativum*. *Phytochemistry* **39**, 1309–1317 (1995).
- Collakova, E. *et al.* *Arabidopsis* 10-formyl tetrahydrofolate deformylases are essential for photorespiration. *Plant Cell* **20**, 1818–1832 (2008).
- Ito, J. *et al.* Analysis of the *Arabidopsis* cytosolic proteome highlights subcellular partitioning of central plant metabolism. *J. Proteome. Res.* **10**, 1571–1582 (2011).
- Zybailov, B. *et al.* Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE* **3**, e1994 (2008).
- Sauter, M., Moffatt, B., Saechao, M. C., Hell, R. & Wirtz, M. Methionine salvage and S-adenosylmethionine: essential links between sulfur, ethylene and polyamine biosynthesis. *Biochem. J.* **451**, 145–154 (2013).
- De La Haba, G. & Cantoni, G. L. The enzymatic synthesis of S-adenosyl-L-homocysteine from adenosine and homocysteine. *J. Biol. Chem.* **234**, 603–608 (1959).
- Moffatt, B. A. *et al.* Adenosine kinase deficiency is associated with developmental abnormalities and reduced transmethylation. *Plant. Physiol.* **128**, 812–821 (2002).
- Guranowski, A. & Pawelkiewicz, J. Adenosylhomocysteinase from yellow lupin seeds. Purification and properties. *Eur. J. Biochem.* **80**, 517–523 (1977).
- Roje, S. *et al.* Isolation, characterization, and functional expression of cDNAs encoding NADH-dependent methylenetetrahydrofolate reductase from higher plants. *J. Biol. Chem.* **274**, 36089–36096 (1999).
- Allaire, M., Li, Y., MacKenzie, R. E. & Cygler, M. The 3-D structure of a folate-dependent dehydrogenase/cyclohydrolase bifunctional enzyme at 1.5 Å resolution. *Structure* **6**, 173–182 (1998).
- Pawelek, P. D., Allaire, M., Cygler, M. & MacKenzie, R. E. Channeling efficiency in the bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase domain: the effects of site-directed mutagenesis of NADP binding residues. *Biochim. Biophys. Acta* **1479**, 59–68 (2000).
- Li, R., Moore, M. & King, J. Investigating the regulation of one-carbon metabolism in *Arabidopsis thaliana*. *Plant Cell Physiol.* **44**, 233–241 (2003).
- Prabhu, V., Chatson, K. B., Abrams, G. D. & King, J. ¹³C nuclear magnetic resonance detection of interactions of serine hydroxymethyltransferase with C1-tetrahydrofolate synthase and glycine decarboxylase complex activities in *Arabidopsis*. *Plant Physiol.* **112**, 207–216 (1996).
- Hare, P. D. & Cress, W. A. Metabolic implications of stress-induced proline accumulation in plants. *Plant Growth Regul.* **21**, 79–102 (1997).
- Mehrshahi, P. *et al.* Functional analysis of folate polyglutamylation and its essential role in plant metabolism and development. *Plant J.* **64**, 267–279 (2010).
- Srivastava, A. C. *et al.* The folylpolyglutamate synthetase plastidial isoform is required for postembryonic root development in *Arabidopsis*. *Plant. Physiol.* **155**, 1237–1251 (2011).
- Zhang, H. *et al.* Sulfamethazine suppresses epigenetic silencing in *Arabidopsis* by impairing folate synthesis. *Plant Cell* **24**, 1230–1241 (2012).
- Zhou, H. R. *et al.* Folate polyglutamylation is involved in chromatin silencing by maintaining global DNA methylation and histone H3K9 dimethylation in *Arabidopsis*. *Plant Cell* **25**, 2545–2559 (2013).

36. Becker, C. *et al.* Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* **480**, 245–249 (2011).
37. Hagmann, J. *et al.* Century-scale methylome stability in a recently diverged *Arabidopsis thaliana* lineage. *PLoS Genet.* **11**, e1004920 (2015).
38. Schmitz, R. J. *et al.* Transgenerational epigenetic instability is a source of novel methylation variants. *Science* **334**, 369–373 (2011).
39. van der Graaf, A. *et al.* Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proc. Natl Acad. Sci. USA* **112**, 6676–6681 (2015).
40. Dubin, M. J. *et al.* DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *Elife* **4**, e05255 (2015).
41. Willing, E.-M. *et al.* Genome expansion of *Arabidopsis alpine* linked with retrotransposition and reduced symmetric DNA methylation. *Nat. Plants* **1**, 14023 (2015).
42. Shen, X. *et al.* Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. *PLoS Genet* **10**, e1004842 (2014).
43. Waterland, R. A. & Jirtle, R. L. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell. Biol.* **23**, 5293–5300 (2003).
44. Fan, J. *et al.* Quantitative flux analysis reveals folate-dependent NADPH production. *Nature* **510**, 298–302 (2014).
45. Van Wilder, V. *et al.* C1 metabolism and chlorophyll synthesis: the Mg-protoporphyrin IX methyltransferase activity is dependent on the folate status. *New Phytol.* **182**, 137–145 (2009).
46. Hum, D. W., Bell, A. W., Rozen, R. & MacKenzie, R. E. Primary structure of a human trifunctional enzyme. Isolation of a cDNA encoding methylenetetrahydrofolate dehydrogenase-methylenetetrahydrofolate cyclohydrolase-formyltetrahydrofolate synthetase. *J. Biol. Chem.* **263**, 15946–15950 (1988).
47. Paukert, J. L., Williams, G. R. & Rabinowitz, J. C. Formyl-methylenetetrahydrofolate synthetase (combined); correlation of enzymic activities with limited proteolytic degradation of the protein from yeast. *Biochem. Biophys. Res. Commun.* **77**, 147–154 (1977).
48. Dev, I. K. & Harvey, R. J. A complex of N5,N10-methylenetetrahydrofolate dehydrogenase and N5,N10-methylenetetrahydrofolate cyclohydrolase in *Escherichia coli*. Purification, subunit structure, and allosteric inhibition by N10-formyltetrahydrofolate. *J. Biol. Chem.* **253**, 4245–4253 (1978).
49. Ljungdahl, L. G., O'Brien, W. E., Moore, M. R. & Liu, M. T. Methylenetetrahydrofolate dehydrogenase from *Clostridium formicoaceticum* and methylenetetrahydrofolate dehydrogenase, methylenetetrahydrofolate cyclohydrolase (combined) from *Clostridium thermoaceticum*. *Methods Enzymol.* **66**, 599–609 (1980).
50. Chan, S. W. *et al.* RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in *Arabidopsis*. *PLoS Genet.* **2**, e83 (2006).
51. Krajcinovic, M. MTHFD1 gene: role in disease susceptibility and pharmacogenetics. *Pharmacogenomics* **9**, 829–832 (2008).
52. Field, M. S., Kamynina, E., Watkins, D., Rosenblatt, D. S. & Stover, P. J. Human mutations in methylenetetrahydrofolate dehydrogenase 1 impair nuclear *de novo* thymidylate biosynthesis. *Proc. Natl Acad. Sci. USA* **112**, 400–405 (2015).
53. Loizeau, K. *et al.* A genome-wide and metabolic analysis determined the adaptive response of *Arabidopsis* cells to folate depletion induced by methotrexate. *Plant Physiol.* **148**, 2083–2095 (2008).
54. Rocha, P. S. *et al.* The *Arabidopsis* HOMOLOG-DEPENDENT GENE SILENCING1 gene codes for an S-adenosyl-L-homocysteine hydrolase required for DNA methylation-dependent gene silencing. *Plant Cell* **17**, 404–417 (2005).
55. Tehlivets, O., Malanovic, N., Visram, M., Pavkov-Keller, T. & Keller, W. S-adenosyl-L-homocysteine hydrolase and methylation disorders: yeast as a model system. *Biochim. Biophys. Acta* **1832**, 204–215 (2013).
56. Khan, M. S. *et al.* Sulfite reductase defines a newly discovered bottleneck for assimilatory sulfate reduction and is essential for growth and development in *Arabidopsis thaliana*. *Plant Cell* **22**, 1216–1231 (2010).
57. Goto, D. B. & Naito, S. AtMRD1 and AtMRU1, two novel genes with altered mRNA levels in the methionine over-accumulating mto1-1 mutant of *Arabidopsis thaliana*. *Plant. Cell. Physiol.* **43**, 923–931 (2002).
58. Goyer, A. *et al.* 5-Formyltetrahydrofolate is an inhibitory but well tolerated metabolite in *Arabidopsis* leaves. *J. Biol. Chem.* **280**, 26137–26142 (2005).
59. Roje, S., Janave, M. T., Ziemak, M. J. & Hanson, A. D. Cloning and characterization of mitochondrial 5-formyltetrahydrofolate cycloligase from higher plants. *J. Biol. Chem.* **277**, 42748–42754 (2002).
60. Besson, V., Rebeille, F., Neuburger, M., Douce, R. & Cossins, E. A. Effects of tetrahydrofolate polyglutamates on the kinetic parameters of serine hydroxymethyltransferase and glycine decarboxylase from pea leaf mitochondria. *Biochem. J.* **292**, 425–430 (1993).
61. Deleris, A. *et al.* Involvement of a Jumoni-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. *EMBO Rep.* **11**, 950–955 (2010).
62. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinformatics* **10**, 232 (2009).
63. Stroud, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V. & Jacobsen, S. E. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**, 352–364 (2013).
64. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
65. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
66. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
67. Eshed, Y., Baum, S. F., Perea, J. V. & Bowman, J. L. Establishment of polarity in lateral organs of plants. *Curr. Biol.* **11**, 1251–1260 (2001).
68. Jarsch, I. K. & Ott, T. Quantitative image analysis of membrane microdomains labelled by fluorescently tagged proteins in *Arabidopsis thaliana* and *Nicotiana benthamiana*. *Bio-protocol* **5**, e1497 (2015).
69. Wirtz, M., Droux, M. & Hell, R. O-acetylserine (thiol) lyase: an enigmatic enzyme of plant cysteine biosynthesis revisited in *Arabidopsis thaliana*. *J. Exp. Bot.* **55**, 1785–1798 (2004).
70. Burstenbinder, K., Rzewuski, G., Wirtz, M., Hell, R. & Sauter, M. The role of methionine recycling for ethylene synthesis in *Arabidopsis*. *Plant J.* **49**, 238–249 (2007).

Acknowledgements

We thank Mahnaz Akhavan for her support in high-throughput sequencing at the University of California, Los Angeles (UCLA) Broad Stem Cell Research Center BioSequencing Core Facility. Work in the Jacobsen laboratory was supported by NIH grant GM60398. M.G. was supported by an EMBO ALTF 986-2011 Long-term Fellowship. Work in the Díaz de la Garza laboratory was supported by CONACyT grant CB243058. C.G.-S. was supported by CONACyT doctoral fellowship 372615. S.F. was a Special Fellow of the Leukemia & Lymphoma Society. S.B. is supported by a postdoctoral fellowship of the Swiss National Science Foundation. J.A.L. is supported by NIH grant GM072764. The work in the Hell/Wirtz laboratory was supported by the German research foundation (grants HE 1848/16-1 and WI 3560/2-1) and the Metabolomics Core Technology Platform of the Excellence cluster 'CellNetworks' (University of Heidelberg). S.E.J. is an Investigator of the Howard Hughes Medical Institute.

Author contributions

G.M. performed genetic screening. M.G. performed co-segregation analyses, complementation crosses and analysed *mthfd1* T-DNA insertion mutants. M.G., G.M., A.J. and S.F. generated sequencing libraries. H.W., M.G., S.J.C., J.Z. and C.J.H. performed bioinformatics analyses. M.G. performed immunofluorescence analysis. J.L., S.B. and M.G. performed the subcellular localization analysis. M.W. analysed thiols, amino acids and adenosine nucleotides. C.G.-S., P.A.R.-P. and R.I.D.I.G. analysed folates. A.J. and D.C.S. performed growth assays. M.G. and S.E.J. wrote the manuscript. R.H., R.I.D.I.G. and S.E.J. coordinated research.

Additional information

Accession codes: All generated high-throughput sequencing data are available at NCBI's Gene Expression Omnibus (GEO) and are accessible via GEO Series accession number GSE77966.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no conflict of interest.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Groth, M. *et al.* MTHFD1 controls DNA methylation in *Arabidopsis*. *Nat. Commun.* **7**:11640 doi: 10.1038/ncomms11640 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

RESEARCH ARTICLE

Arabidopsis AtMORC4 and AtMORC7 Form Nuclear Bodies and Repress a Large Number of Protein-Coding Genes

C. Jake Harris¹, Dylan Husmann¹^{¶a}, Wanlu Liu¹, Farid El Kasmi², Haifeng Wang^{3,4}, Ashot Papikian¹, William A. Pastor¹, Guillaume Moissiard¹^{¶b}, Ajay A. Vashisht⁵, Jeffery L. Dangl^{2,6}, James A. Wohlschlegel⁵, Steven E. Jacobsen^{1,7}^{*}

1 Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, California, United States of America, **2** Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **3** Basic Forestry and Proteomics Research Center, Haixia Institute of Science and Technology (HIST), Fujian Agriculture and Forestry University, Fuzhou, Fujian, China, **4** Fujian Province Key Laboratory of Plant Virology, Institute of Plant Virology, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China, **5** Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, United States of America, **6** Howard Hughes Medical Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **7** Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, California, United States of America

 These authors contributed equally to this work.

^{¶a} Current address: Department of Biology, Stanford University, Stanford, California, United States of America

^{¶b} Current address: Swiss Federal Institute of Technology (ETH-Z), Department of Biology, Zurich, Switzerland

* jacobsen@ucla.edu



OPEN ACCESS

Citation: Harris CJ, Husmann D, Liu W, Kasmi FE, Wang H, Papikian A, et al. (2016) Arabidopsis AtMORC4 and AtMORC7 Form Nuclear Bodies and Repress a Large Number of Protein-Coding Genes. PLoS Genet 12(5): e1005998. doi:10.1371/journal.pgen.1005998

Editor: Nick Gilbert, Edinburgh Cancer Centre, UNITED KINGDOM

Received: November 12, 2015

Accepted: March 30, 2016

Published: May 12, 2016

Copyright: © 2016 Harris et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database (accession number GSE78836).

Funding: CJH is supported by an EMBO Long-Term Fellowship (ALTF 1138-2014). WL is supported by a Philip J. Whitcome Fellowship from the UCLA Molecular Biology Institute and a scholarship from the Chinese Scholarship Council. FEK is supported by a DFG Postdoctoral Research Fellowship (EL 734/1-1). JAW is supported by an NIH grant (GM089778). SEJ and JLD are investigators of the Howard Hughes

Abstract

The MORC family of GHKL ATPases are an enigmatic class of proteins with diverse chromatin related functions. In Arabidopsis, AtMORC1, AtMORC2, and AtMORC6 act together in heterodimeric complexes to mediate transcriptional silencing of methylated DNA elements. Here, we studied Arabidopsis *AtMORC4* and *AtMORC7*. We found that, in contrast to AtMORC1,2,6, they act to suppress a wide set of non-methylated protein-coding genes that are enriched for those involved in pathogen response. Furthermore, *atmorc4 atmorc7* double mutants show a pathogen response phenotype. We found that AtMORC4 and AtMORC7 form homomeric complexes *in vivo* and are concentrated in discrete nuclear bodies adjacent to chromocenters. Analysis of an *atmorc1,2,4,5,6,7* hexuple mutant demonstrates that transcriptional de-repression is largely uncoupled from changes in DNA methylation in plants devoid of MORC function. However, we also uncover a requirement for MORC in both DNA methylation and silencing at a small but distinct subset of RNA-directed DNA methylation target loci. These regions are characterized by poised transcriptional potential and a low density of sites for symmetric cytosine methylation. These results provide insight into the biological function of MORC proteins in higher eukaryotes.

Medical Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Author Summary

Keeping selfish genetic elements—such as transposons—silent, while maintaining access to genes, is a fundamental challenge for eukaryotes. Different pathways frequently converge in order to identify transposons and maintain their repression, and in *Arabidopsis thaliana*, transposons are marked with DNA methylation. Previous studies of the Arabidopsis MORC proteins, which represent a highly conserved protein family, showed that AtMORC1, AtMORC2, and AtMORC6 are required for repression of methylated target transposons. Here, we describe the Arabidopsis genes *AtMORC4* and *AtMORC7*, which, instead of targeting methylated elements, appear to act redundantly to repress a large set of protein-coding genes and are required to mount a full defense against pathogen challenge. These proteins localize throughout the nucleus and form punctate bodies at the boundaries of highly compacted chromatin. By knocking out all functional copies of MORC genes in *Arabidopsis*, we find that major changes in transcription are not generally associated with the loss of DNA methylation. However, MORC may be recruited to assist in silencing of methylated regions that are unusually susceptible to transcriptional re-activation. This indicates that MORC and DNA methylation are convergently required to maintain repression at transposon targets.

Introduction

Maintaining regulatory access to genes while repressing the expression of potentially deleterious transposable elements is a fundamental challenge for living organisms. Eukaryotes achieve this in part by parsing their genomes into functional units characterized by distinct chromatin features [1,2]. The most stable chromatin mark is cytosine DNA methylation [3]. In plants, DNA methylation is often associated with transcriptionally silent regions [4,5] and occurs primarily in three sequence contexts, CG, CHG and CHH (where H is defined by any base except G). Methylation at the symmetrical CG and CHG sites is maintained by the action of MET1—the homologue of mammalian DNMT1—and CMT3, respectively [6]. Asymmetric CHH methylation must be continuously re-established. In pericentromeric heterochromatin, this is mostly mediated by CMT2 [7,8]; while in small patches of heterochromatin in the otherwise euchromatic arms, CHH methylation is mostly maintained by the action of DRM2 in the RNA-directed DNA methylation (RdDM) pathway [9–11].

RdDM primarily targets transposable elements through the combined action of two plant specific RNA polymerases [12,13]. During RdDM, Polymerase IV (Pol IV) is in part recruited by SHH1 [14] to generate short transcripts [15–17], which are made double-stranded by the action of RDR2 and diced into 24nt small RNAs by DCL3. Polymerase V (Pol V) is targeted to methylated sites via SUVH2/9 [18,19] and generates scaffold transcripts to recruit 24nt small RNA directed complexes [20,21], which then recruit the *de novo* methyltransferase DRM2 to induce DNA methylation in all sequence contexts [10]. The RdDM pathway results in a robust self-reinforcing loop; however, a potential role for 21nt small RNAs and RDR6 during the early stages of methylation establishment has recently emerged [22–24].

To identify novel factors involved in transcriptional gene silencing, forward genetic screens from three independent laboratories isolated alleles of *AtMORC6* [NP_173344; *AT1G19100*; *CRH6*; *Defective in Meristem Silencing 11* (*DMS11*)] [25–27]. MORC proteins are members of the GHKL ATPase superfamily [28,29] and by evolutionary comparison with prokaryotes are predicted to play a role DNA superstructure manipulations in response to epigenetic signals [30]. While the involvement of *AtMORC6* in transcriptional repression is established, the

extent to which it contributes to DNA methylation at target loci has varied between reports [25–27]. For instance, a 2012 study [25] found little evidence for methylation changes at either the de-repressed reporter construct or genome wide, while Lorković et al., 2012 [26] and Brabbs et al., 2013 [27] both observed minor reductions in DNA methylation at their reporter loci. It therefore remains uncertain whether transcriptional activation is associated with loss of DNA methylation in *atmorc* mutants and to what extent AtMORC proteins are involved in the RdDM pathway.

Another member of the *A. thaliana* MORC family, *AtMORC1* [NP_568000; AT4G36290; *Compromised Recognition of Turnip Crinkle Virus 1 (CRT1)*], is involved in plant defense and was isolated as a mutant that is hyper-sensitive to Turnip Crinkle Virus [31]. Interestingly, *AtMORC1* was also identified in the same transcriptional repression screen that isolated *AtMORC6* [25]. Recent studies have implicated changes in DNA methylation and transcriptional responses to pathogen infection [32–34]. Yet it is unclear how AtMORC1 might function in both plant defense and transcriptional repression at RdDM targets. AtMORC1 and its very close homolog AtMORC2 act in mutually exclusive heteromeric complexes with AtMORC6, and an *atmorc1 atmorc2 atmorc6* triple mutant resembles that of *atmorc6* with regard to transcriptional profile and methylation state [35].

As there are seven members of the MORC family in Arabidopsis, we sought to characterize the remaining *AtMORC* genes in order to help elucidate MORC function. We found that the highly related AtMORC4 [NP_199891; AT5G50780; *CRH4*] and AtMORC7 [NP_194227; AT4G24970; *CRH3*] proteins act partially redundantly to transcriptionally repress a large regulon and also play a role in plant defense. Both AtMORC4 and AtMORC7 were found to form stable homomers, but do not interact with each other, suggesting that they act in parallel to control gene silencing. We also found that AtMORC4 and AtMORC7, like AtMORC1 and AtMORC6 [25], form nuclear bodies that are adjacent to chromocenters. Finally, by generating a compound mutant devoid of all MORC function, we demonstrate that transcriptional de-repression can be largely uncoupled from changes in DNA methylation. However, a small but distinct subset of RdDM loci that are poised for transcriptional reactivation exhibit MORC-dependent methylation changes and reduced symmetric methylation potential.

Results and Discussion

AtMORC4 and AtMORC7 act semi-redundantly at a common set of loci

AtMORC4 and *AtMORC7* are highly related to one another (Fig 1A and 1B) [35]. We obtained T-DNA knockout lines for these genes (*atmorc4-1* and *atmorc7-1*) (S1A Fig). RT-PCR at targets known to be de-repressed in the *atmorc6* background [25,35] showed little change in transcript levels in the homozygous knockouts. However, when we crossed the lines to create an *atmorc4-1 atmorc7-1* double knockout, we observed de-repression at several of the candidate loci, suggesting that AtMORC4 and AtMORC7 act redundantly (S1B Fig). To determine the extent of redundancy between AtMORC4 and AtMORC7, we performed mRNA-Sequencing (RNA-seq) on leaves from individual plants of Col-0, *atmorc4-1*, *atmorc7-1*, and *atmorc4-1 atmorc7-1* backgrounds (hereafter referred to as wild-type (wt), *atmorc4*, *atmorc7* and *atmorc4/7*, respectively). We found that AtMORC4 and AtMORC7 affect a highly overlapping gene set with AtMORC7 playing a more dominant role (Fig 1C–1E). In *atmorc7*, 348 annotated loci were differentially expressed (FDR < 0.05) with 84% being up-regulated. In *atmorc4*, the 33 differentially expressed loci (30 up, 3 down) were largely a subset of those altered in *atmorc7*, with 29 of the 30 up-regulated loci also up-regulated in *atmorc7*. In the *atmorc4/7* double knockout, 50% more loci were differentially expressed than in the individual knockouts combined, suggesting a significant level of redundancy between AtMORC4 and AtMORC7.

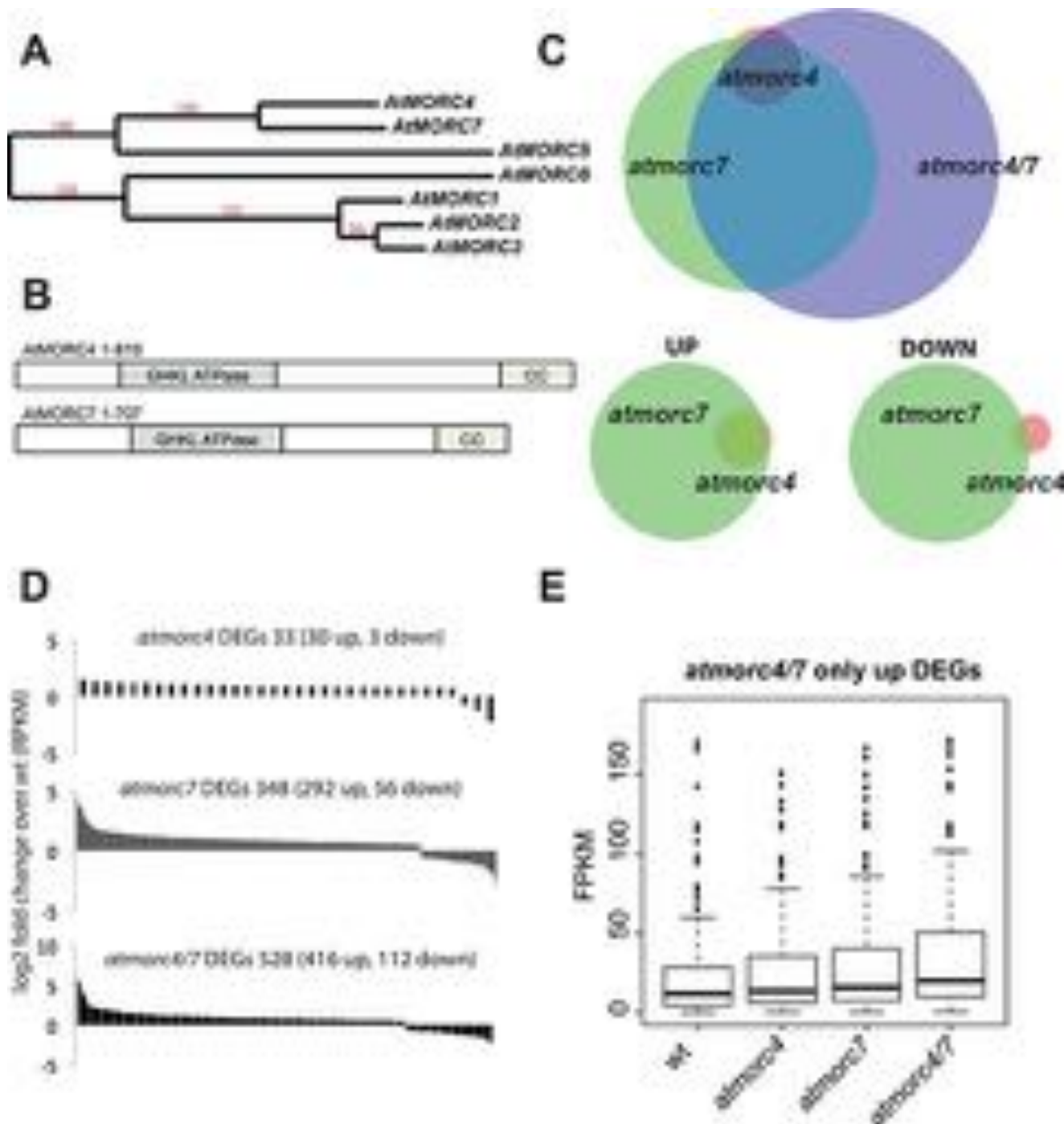


Fig 1. AtMORC4 and AtMORC7 act in a partially redundant manner to repress a highly overlapping gene set. (A) Phylogenetic reconstruction of *Arabidopsis thaliana* AtMORC genes (genomic sequence). Red numbers indicate branch support values in percentage (<http://www.phylogeny.fr>). (B) Schematic representation of AtMORC4 and AtMORC7 domains, drawn approximately to scale (CC = coiled coil). (C) Upper: Overlap of differentially expressed genes (DEGs—includes both genes and transposons—FDR<0.05) in the mutants indicated. Lower: overlap of *atmorc4* and *atmorc7* in either upregulated (UP) or downregulated (DOWN) loci. There is greater overlap for the upregulated loci. Within each overlap, circle size and overlap is proportional to number of DEGs therein. (D) log₂ fold change for individual DEGs in each of the mutants indicated (ranked highest to lowest). Most are upregulated. (E) FPKM (fragments per kilobase per million reads) boxplots for upregulated DEGs only present in *atmorc4/7*, showing that the *atmorc4* and *atmorc7* single mutants also show a similar trend at these loci.

doi:10.1371/journal.pgen.1005998.g001

Taken together, the results suggest that AtMORC4 and AtMORC7 act in a partially redundant manner, with AtMORC7 having a stronger effect than AtMORC4, to mainly repress a highly overlapping gene set.

AtMORC4 and AtMORC7 form homomeric complexes *in vivo*

We have previously shown that AtMORC6 forms mutually exclusive heteromeric complexes with either AtMORC1 or AtMORC2 [35]. To assess whether AtMORC4 and AtMORC7 form heteromeric complexes, we generated endogenous promoter driven MYC or FLAG tagged lines for both AtMORC4 and AtMORC7 in their respective T-DNA backgrounds. By co-immunoprecipitation, we detected a homotypic association of AtMORC4 and AtMORC7 but did not detect heteromers (Fig 2A–2C). These results were confirmed by mass spectrometry of the immunoprecipitated samples (IP-MS), showing that the AtMORC4 and AtMORC7 precipitates do not contain peptides from AtMORCs other than themselves (Fig 2D). Together, this indicates that AtMORC4 and AtMORC7 form homomeric complexes *in vivo*, consistent with the genetic redundancy observed between them (see Fig 1, S1 Fig).

Transcriptome comparison between *AtMORC* knockouts

To directly compare the phenotypes of the *atmorc4* and *atmorc7* mutants with the previously characterized *atmorc6-3* (hereafter referred to as *atmorc6*), we performed a second round of RNA-seq analysis. We also sought to generate a genetically MORC-less plant to obtain an unobscured view of MORC function. For this, we created a higher order knockout plant

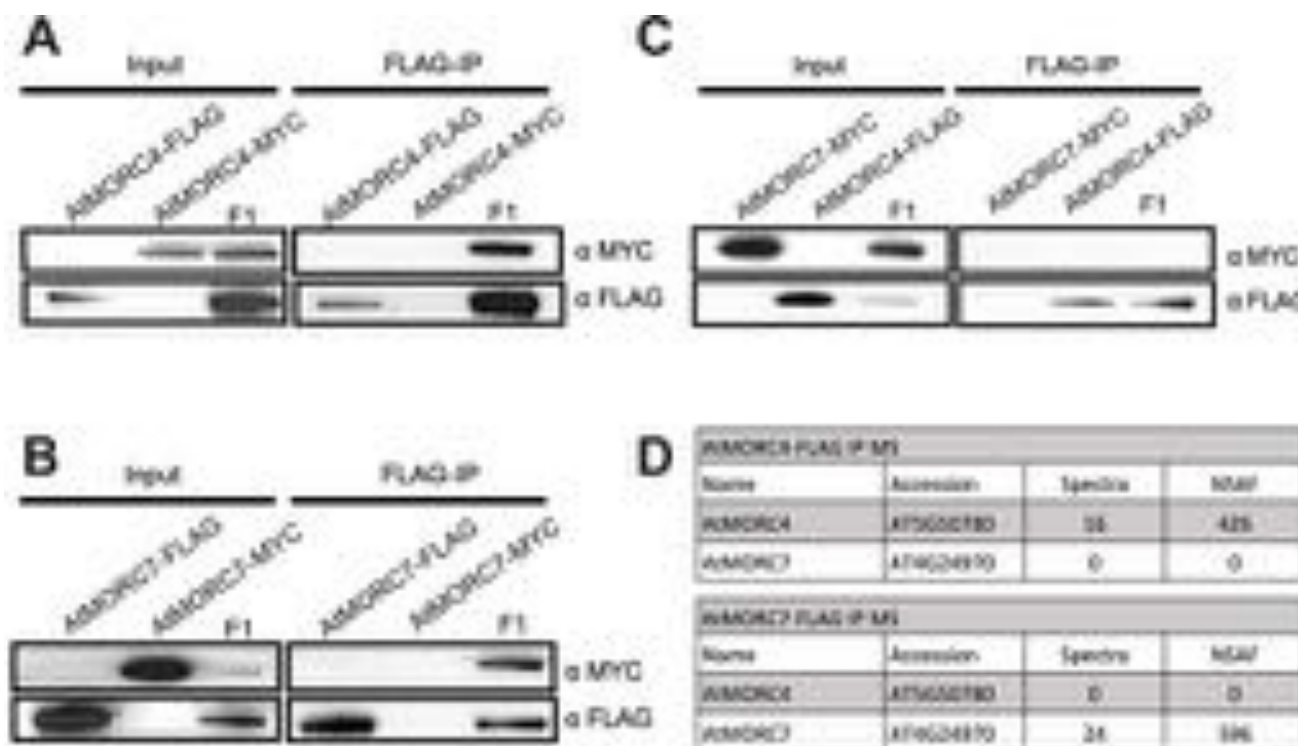


Fig 2. AtMORC4 and AtMORC7 form homomeric complexes *in vivo*. (A) Co-immunoprecipitation of AtMORC4-MYC with AtMORC4-FLAG in F1 plants (B) Co-immunoprecipitation of AtMORC7-MYC with AtMORC7-FLAG in F1 plants. (C) No interaction by co-immunoprecipitation between AtMORC7-MYC and AtMORC4-FLAG in F1 plants. (D) Table from immunoprecipitation followed by mass spectrometry (IP-MS) of FLAG tagged AtMORC4 and AtMORC7 plants showing peptides from themselves but not each other. NSAF = normalized spectral abundance factor.

doi:10.1371/journal.pgen.1005998.g002

containing T-DNA inserts in six out of the seven *MORC* genes in Arabidopsis, *atmorc1-2*, *atmorc2-1*, *atmorc4-1*, *atmorc5-1*, *atmorc6-3*, and *atmorc7-1* (*atmorc1/2/4/5/6/7*). While a previous study reported embryonic lethality for a T-DNA insertion in *AtMORC3* [NP_195350; AT4G36270; CRH2] [36], it is likely that this is an indirect effect caused by an unknown linked mutation in the SALK line (SALK_000009), as we found evidence suggesting that *AtMORC3* is in fact a pseudogene (S2 Fig). We found a premature stop codon in exon three in Col-0 (causing either an un-translated or truncated protein). Additionally, an independent homozygous T-DNA allele (SALK_043244) with an exonic insertion exhibited no discernable phenotype. Given that *AtMORC3* is non-functional in Col-0, the *atmorc1/2/4/5/6/7* line effectively lacks any functional AtMORC protein.

RNA-seq on individual plants (3 replicates each) from *atmorc6*, *atmorc4/7*, *atmorc4/6/7*, and *atmorc1/2/4/5/6/7* revealed 39, 815, 1188, and 1519 differentially expressed genes (FDR < 0.05) relative to wt, respectively, with a variety of interesting features (Fig 3). Twenty times more loci were differentially expressed in *atmorc4/7* as compared to *atmorc6*, suggesting that AtMORC4 and AtMORC7 play a more central role in gene expression (Fig 3A). As the majority of these *atmorc4/7* differentially expressed genes were up-regulated (87%), this is consistent with a repressive role and direct regulation at these targets. However, we cannot exclude the possibility of indirect effects. The difference between *atmorc6* and *atmorc4/7* is also clearly apparent from a heatmap over the union set of differentially expressed loci, which shows that *atmorc6* is most similar to wt (Fig 3B). In *atmorc6*, transposable elements (TEs) constitute 29% (11 total) of the differentially expressed loci while in *atmorc4/7*, only 1% (9 total) were misregulated, suggesting that AtMORC6 is preferentially involved in TE repression while AtMORC4 and AtMORC7 are primarily responsible for the repression of protein-coding genes.

Comparing *atmorc4/7* to *atmorc6* revealed that while there was a generally positive correlation, many loci are specifically affected in either *atmorc6* or *atmorc4/7* (Fig 3C). One example is *ZF1*, which encodes a stimulus response zinc finger protein characteristic of the types of genes up-regulated in *atmorc4/7* (see below) and is up-regulated only in *atmorc4/7*. On the other hand, the gene *SDC* [37] was much more highly up-regulated in *atmorc6* than it was in *atmorc4/7*, consistent with the use of its promoter in the forward genetic screen that resulted in isolation of *atmorc6* [25]. A similar plot comparing *atmorc4/6/7* versus *atmorc1/2/4/5/6/7* showed an extremely close correlation (Fig 3D and see S3 Fig). This demonstrates that AtMORC1, AtMORC2, and AtMORC5 [NP_196817; At5G13130; CRH5] do not have a significant impact on the transcriptome, consistent with the previous report indicating that *atmorc1/2* is equivalent to that of *atmorc6* and that the expression of *AtMORC5* is pollen specific [35].

AtMORC4 and AtMORC7 play a role in plant defense

We performed GO term analysis on the genes misregulated in *atmorc4/7*, which revealed a striking enrichment for immune response genes, especially ‘response to chitin’ (p value = $2.3e^{-47}$) (S4 Fig). Interestingly, we had previously noted ‘response to chitin’, albeit with lower significance, (p < $6e^{-4}$), for genes misregulated in *atmorc6* [35]. Chitin is a component of the fungal cell wall and acts as a basal defense response elicitor [38]. In addition, *AtMORC7* appears in an RNA co-expression network with multiple disease resistance genes, including *LURP1* [39], *PUB12* [40], *ACD6* [41], *SDE5* [42] and three NB-LRR type proteins [43] (Fig 4A).

Since *LURP1* mutants are compromised in defense against the Emwa1 isolate of the oomycete pathogen, *Hyaloperonospora arabidopsidis* (*Hpa*) [39] and *atmorc1* was also identified as showing enhanced susceptibility to this pathogen [44], we challenged *atmorc1*, *atmorc6*, *atmorc4*, *atmorc7*, *atmorc4/7*, *atmorc4/6/7* and *atmorc1/2/4/5/6/7* with Emwa1 *Hpa*. We observed significantly increased susceptibility in *atmorc1*, *atmorc6*, *atmorc4/7*, *atmorc4/6/7*

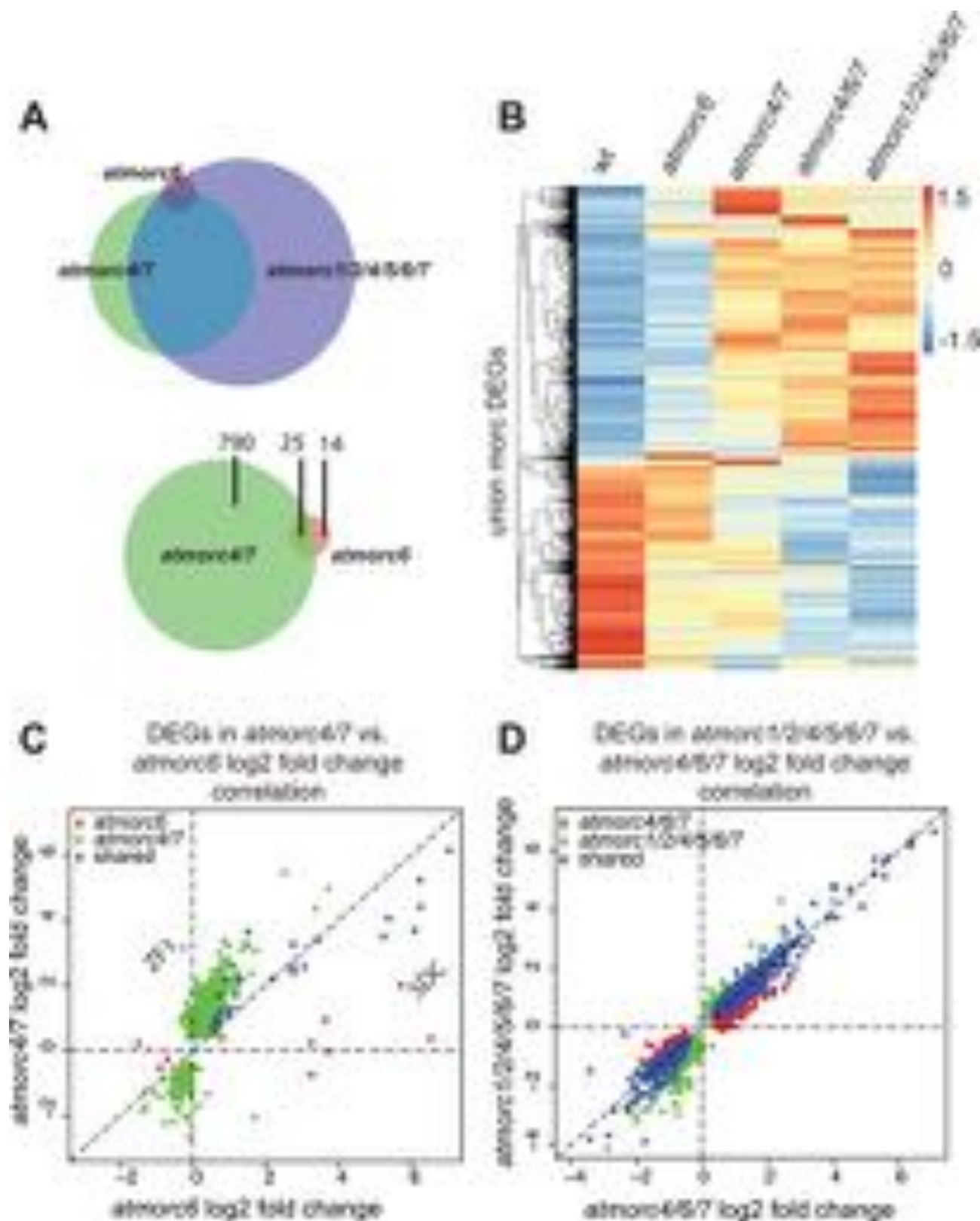


Fig 3. AtMORC4 and AtMORC7 target a wide gene set. (A) Upper: overlap of DEGs in the genotypes indicated with circle size and overlap proportional to number of DEGs therein. Lower: overlap between *atmorc4/7* and *atmorc6* DEGs, with number of DEGs indicated. (B) Heatmap over the union set of DEGs

(FDR<0.05) in the different genotypes. Each row is normalized by z-score (red = relatively higher, blue = relatively lower expression in that genotype). **(C)** Correlation between *atmorc4/7* and *atmorc6* DEGs. *ZF1* and *SDC* are indicated as examples of loci specifically upregulated in *atmorc4/7* or *atmorc6*, respectively. **(D)** Correlation between *atmorc4/6/7* and *atmorc1/2/4/5/6/7* DEGs.

doi:10.1371/journal.pgen.1005998.g003

and *atmorc1/2/4/5/6/7* as compared to wt (Fig 4B). The individual *atmorc4* and *atmorc7* mutants did not show a difference from wild type, providing further support for the functional redundancy between AtMORC4 and AtMORC7. As we did not observe an additive increase in susceptibility in the higher order *atmorc* mutants, we reasoned that this might reflect non-additive changes in the transcriptome. Indeed, the *atmorc4/6/7* and *atmorc1/2/4/5/6/7* plants showed no further increase in expression of the ‘response to chitin’ (GO:0010200) gene set than did *atmorc4/7* (S5 Fig). While the mis-expression of specific genes in this set may contribute to pathogen susceptibility, it also remains possible that AtMORC proteins play a more direct role in defense [31,36,45]. Together, these results suggest that—in addition to AtMORC1—AtMORC6, AtMORC4, and AtMORC7 act as positive regulators of defense in *A. thaliana* against the oomycete *Hpa*.

Chromocenter adjacent enrichment of AtMORC4 and AtMORC7 in the nucleus

In *Arabidopsis*, interphase chromosomes are organized into distinct chromosomal territories, with euchromatic arms looping out from condensed heterochromatic chromocenters [46–48]. These chromocenters constitute repeat and transposon-rich pericentromeric heterochromatin and are readily visible by light microscopy as intensely DAPI stained nuclear foci. AtMORC1 and AtMORC6 form punctate bodies adjacent to chromocenters and in *atmorc6* mutants, pericentromeric regions are decondensed, suggesting that AtMORC6 plays a role in higher order chromatin compaction at the interface of these transposon-rich regions [25,48]. Because AtMORC4 and AtMORC7 were found to target both genes and transposons, we determined their localization in the nucleus. Using pAtMORC4::AtMORC4-MYC and pAtMORC7::AtMORC7-MYC lines, we observed chromocenter adjacent bodies formed by both AtMORC4 and AtMORC7 (Fig 5A and 5B and S1 and S2 Videos). AtMORC7 bodies were generally more intensely stained than AtMORC4 bodies. Consistent with the effects of *atmorc4/7* mutation on euchromatic gene expression, AtMORC4 and AtMORC7 were also uniformly distributed throughout the nucleoplasm whereas AtMORC1 and AtMORC6 tended to appear as punctate nuclear foci (see Fig 5C and 5D and previously observed [25]). AtMORC4 and AtMORC7 staining was specifically excluded from chromocenters, but was frequently enriched along chromocenter boundaries, forming multiple foci or forming rings around chromocenters (Fig 5). The function of these nuclear bodies is currently unknown.

The contribution of MORC to DNA methylation patterning

We utilized the *atmorc1/2/4/5/6/7* hexuple mutant to determine the contribution of AtMORCs to DNA methylation patterning. We performed whole-genome bisulfite sequencing (BS-seq), to examine DNA methylation at single cytosine resolution, in *atmorc1/2/4/5/6/7* as well as *atmorc4/7* and wt (2 biological replicates each). We also included the previously published BS-seq dataset for *atmorc6* [35] in our analysis. Global levels of methylation over the chromosomes were unaltered in any AtMORC knockout background in all three sequence-contexts (S6A Fig). Focusing specifically on loci that were de-repressed in *atmorc1/2/4/5/6/7*, we observed very little overall change in methylation upstream, downstream or throughout the gene body at these loci (S6B Fig). These results suggest that the most significant changes in transcription

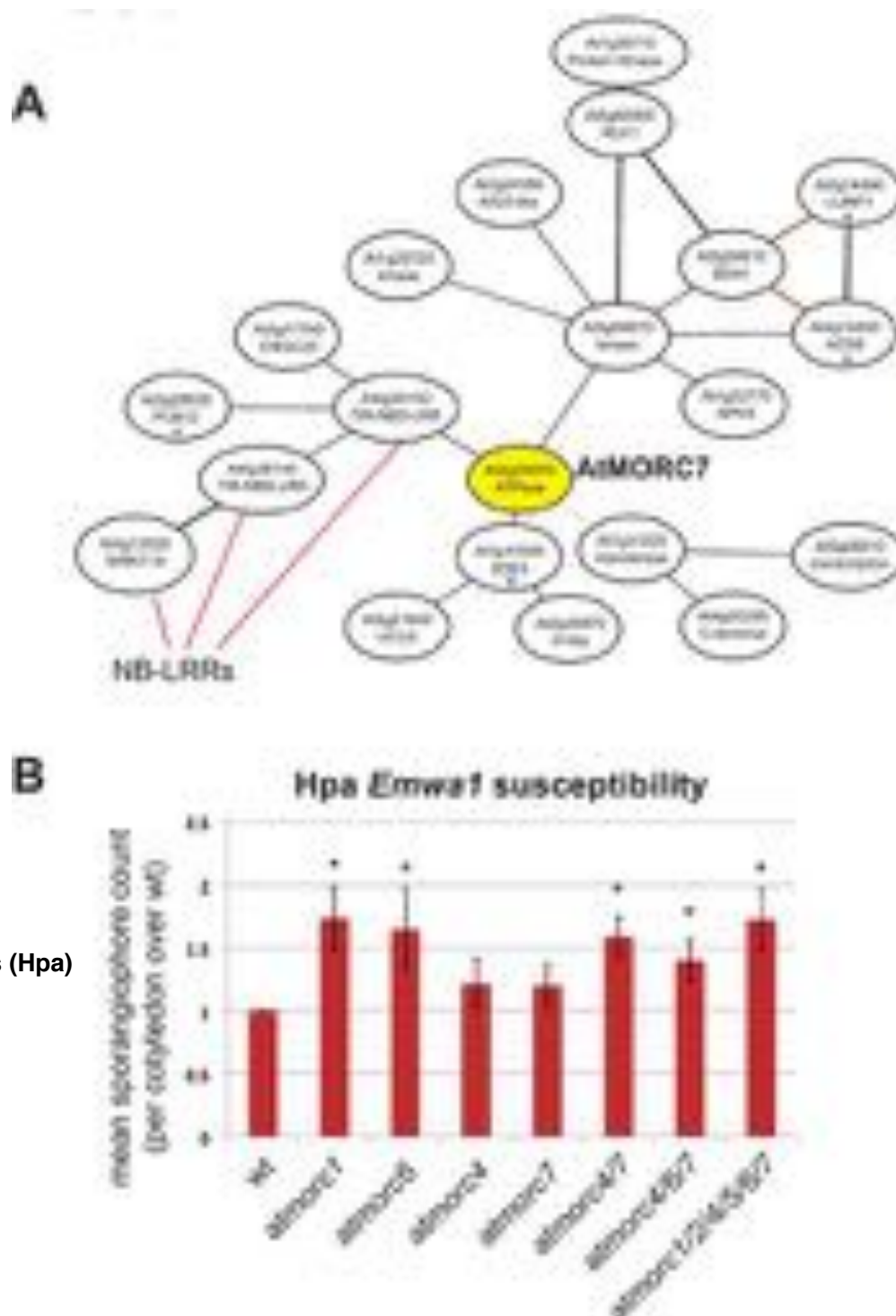


Fig 4. AtMORC4 and AtMORC7 act redundantly in pathogen defense. (A) ATTED-II microarray co-expression network for AtMORC7 [http://atted.jp]. AtMORC7, shown in yellow, is co-expressed with multiple immunity related genes. Blue asterisk indicates genes with established roles in pathogen defense [39–42] and NB-LRRs are classic resistance genes [43]. (B) Mean Emwa1 *Hpa* sporangiospore count per cotyledon over wt, (4–5 days post inoculation of 10 day old seedlings, >100 cotyledons scored per genotype). Data from three individual replicates of the experiment. Error bars represent SEM. * indicates significant difference from wt (p-value < 0.05).

doi:10.1371/journal.pgen.1005998.g004

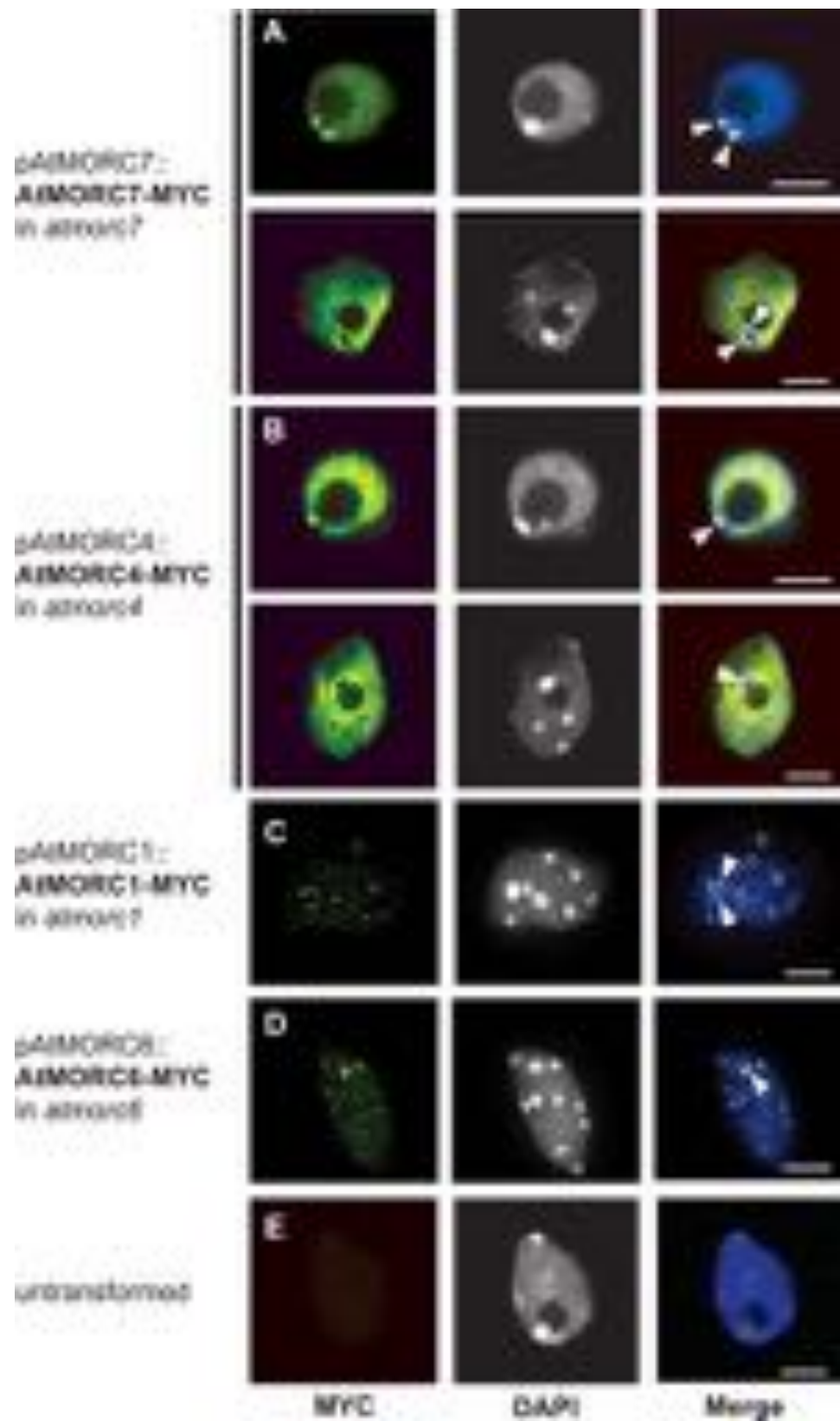


Fig 5. Chromocenter adjacent enrichment of AtMORC4 and AtMORC7 in the nucleus. (A-D)

Representative examples of body forming AtMORC7-MYC, AtMORC4-MYC, AtMORC1-MYC, and AtMORC6-MYC nuclei, respectively. **(E)** Untransformed wt nucleus subjected to the same antibody staining and imaging procedure. Left panels = anti-MYC channel; middle panels = DAPI channel (gray scaled). DAPI stains DNA, defining the position of dense chromocenters as high intensity white foci; right panels = merged channels (DAPI in blue, MYC in green). White triangles indicate examples of chromocenter adjacent AtMORC localization. Scale bars = 5 μ M.

doi:10.1371/journal.pgen.1005998.g005

resulting from the loss of AtMORCs are not generally accompanied by losses in DNA methylation.

Next we examined the potential contribution of AtMORC to the different DNA methylation pathways. MET1 maintains CG methylation throughout the genome, CMT3 maintains the majority of CHG methylation, DRM2 maintains CHH methylation at RdDM sites, and CMT2 maintains CHH methylation in pericentromeric heterochromatin [3,7,8,10]. Using previously defined loci whose methylation is dependent upon these methyltransferases [8,49], we examined methylation levels in the *AtMORC* mutants. Again we found essentially no reduction in methylation in the *AtMORC* knockouts, suggesting that AtMORCs do not play a significant role in any of the major DNA methylation pathways in Arabidopsis (S7A Fig). We also tested whether AtMORCs might act downstream of DNA methylation from any of these specific methyltransferase pathways by plotting RNA-seq reads over differentially methylated regions (DMRs) defined as changing in the different methyltransferase mutant backgrounds; however, we did not observe any consistent changes in bulk levels of RNA in the *AtMORC* knockouts at these collections of methylated loci (S7B Fig).

Since AtMORC6 has been implicated in transcriptional silencing at RdDM loci, reportedly interacting with members of the RdDM pathway [19,26], we examined whether there might be more localized changes in DNA methylation by parsing the genome into 100bp windows and searching for DMRs. We found 519 *atmorc1/2/4/5/6/7* hypomethylated CHH DMRs, 54% of which overlapped with *drm1/2* hypomethylated CHH DMRs (Fig 6A, S8A Fig). In addition, the remaining 46% of hypomethylated CHH DMRs that were called as being specific to *atmorc1/2/4/5/6/7* in fact showed dramatically reduced methylation in *drm1/2* (Fig 6B, right panel), suggesting that even though these DMRs did not make the stringent cutoff required to be a DMR, the majority of *atmorc1/2/4/5/6/7* hypomethylated DMRs correspond to sites of RNA directed DNA methylation. In contrast, only 2% of *atmorc1/2/4/5/6/7* hypomethylated DMRs exclusively overlapped with *cmt2* hypomethylated CHH DMRs (S8A Fig). We also checked whether these *atmorc1/2/4/5/6/7* hypomethylated CHH DMRs might be the result of spontaneous epi-allelic variation by comparison with a previously defined set of DMRs that are known to change states in the wild type [50], but found only a 3% overlap (S8B Fig). Together, these data suggest that AtMORCs are required for CHH methylation at a small subset of *drm1/2*-RdDM loci.

Comparing *atmorc6* with *atmorc4/7* at *atmorc1/2/4/5/6/7* hypo CHH DMRs, we found that *atmorc6* more strongly resembles that of *atmorc1/2/4/5/6/7* (S9 Fig). Interestingly, *atmorc4/7* and *atmorc6* do not appear to affect mutually exclusive regions, suggesting that AtMORC4/7 and AtMORC6 are required at overlapping target loci (S9A Fig). However, *atmorc4/7* generally showed less severe CHH methylation loss than *atmorc6* (S9A and S9B Fig), which is consistent with AtMORC4 and AtMORC7 being primarily involved in repression of protein-coding genes, and AtMORC6 being predominantly involved in repression of methylated elements.

Since the AtMORCs appear to be transcriptional repressors, we plotted RNA-seq data over the *atmorc1/2/4/5/6/7* hypomethylated CHH DMRs. We observed a clear increase in bulk levels of RNA over these sites in the *atmorc1/2/4/5/6/7* knockout (S10A Fig). While this result might

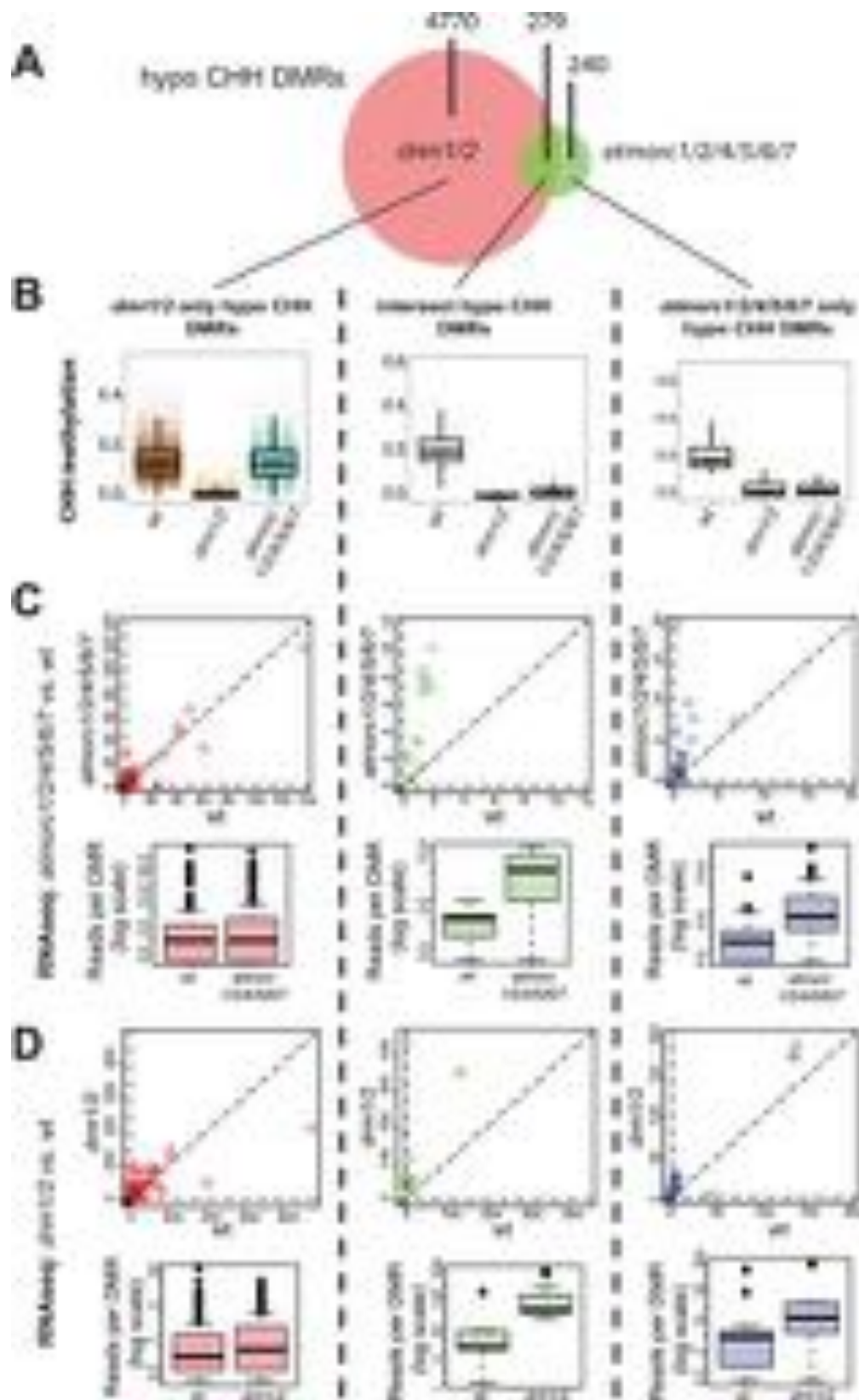


Fig 6. *atmorc* but not *drm1/2* specific hypomethylated CHH DMRs are associated with transcriptional de-repression. (A) Overlap between *atmorc1/2/4/5/6/7* hypo-CHH DMRs and *drm1/2* hypo-CHH DMRs. (B) Boxplot for CHH methylation levels in wt, *drm1/2*, and *atmorc1/2/4/5/6/7* at the hypo CHH DMR regions indicated. Note that although 241 loci were defined as '*atmorc1/2/4/5/6/7* only' in (A), they still lose significant of CHH methylation in *drm1/2*, indicating that these regions are still likely targets of RdDM. (C) Upper: Scatter plot showing RNA-seq reads over DMR regions indicated from *atmorc1/2/4/5/6/7* vs. wt (average from three replicates each). Each dot represents a single DMR. Lower: Boxplots using the same RNA-seq data as above. (D) Same as in (C) except using RNA-seq data from *drm1/2* vs. wt from (data from GEO:GSE51304) [8] (average from two replicates each). In (C) and (D) only DMRs with transcripts detectable in both genotypes were included.

doi:10.1371/journal.pgen.1005998.g006

seem intuitive, this was not the case for *drm1/2* hypomethylated CHH DMRs, where loss of *DRM1/2* did not result in significant transcriptional re-activation (S10B Fig and [8]). To determine whether the overall change in transcription seen in *atmorc1/2/4/5/6/7* knockout is caused by a small number of jackpot sites or is the result of many DMRs becoming transcriptionally reactivated at a moderate level, we plotted RNA-seq reads from individual DMRs (Fig 6C and 6D). We found that *atmorc1/2/4/5/6/7* hypomethylated CHH DMRs were frequently characterized by transcriptional de-repression, while *drm1/2* exclusive hypomethylated CHH sites were not. Interestingly, the *atmorc1/2/4/5/6/7* defined hypomethylated CHH sites were also transcriptionally reactivated in the *drm1/2* background (Fig 6D). Thus this set of sites is susceptible to transcriptional depression when CHH methylation is lost, either by loss of RdDM or by loss of MORC function.

In order to determine if the 519 *atmorc1/2/4/5/6/7* hypomethylated DMR regions might have unique qualities that distinguish them from other sites that do not lose CHH methylation, we analyzed their DNA sequence composition. Interestingly, when we calculated CG, CHG, and CHH density, we found that the *atmorc1/2/4/5/6/7* defined subset had significantly fewer CG and CHG sites as compared to the rest of the RdDM loci and compared to the genome average (Fig 7). An attractive hypothesis therefore is that a low density of symmetric methylation (due to a low density of methylatable sites) may not be sufficient to maintain silencing once asymmetric CHH methylation is lost, which would explain why these particular regions become reactivated in *drm1/2*. Since AtMORCs are not generally required for CHH methylation maintenance, it would then seem likely that AtMORCs primary role would be to help maintain transcriptional repression at these regions of diffuse symmetric methylation and poised transcriptional potential. The transcriptional reactivation of these sites in *atmorc* may then secondarily lead to loss of CHH methylation at these loci, and it is indeed known that positive epigenetic marks associated with transcription can lead to a loss of RdDM function [14,51,52]. In addition, symmetric CG methylation plays a role in the stable association of Pol V to chromatin, and thus perpetuates RdDM and CHH methylation [18]. Thus we hypothesize that this unique set of 519 *atmorc1/2/4/5/6/7* hypomethylated DMR regions experience a loss of methylation because they are both depleted in symmetric methylation and because they become transcriptionally reactivated in *atmorc* mutants.

Conclusion

In this study, we established a role for the previously uncharacterized *AtMORC4* and *AtMORC7* genes in widespread repression of protein-coding genes and in pathogen defense. We found that these proteins act partially redundantly, forming mutually exclusive homomeric complexes, which explains why they have not previously been identified in forward genetic screens. In addition, AtMORC4 and AtMORC7 formed bodies adjacent to chromocenters while also showing localization throughout the nucleoplasm. By analysing a compound mutant devoid of all MORC function, we showed that AtMORC is not a key component in the

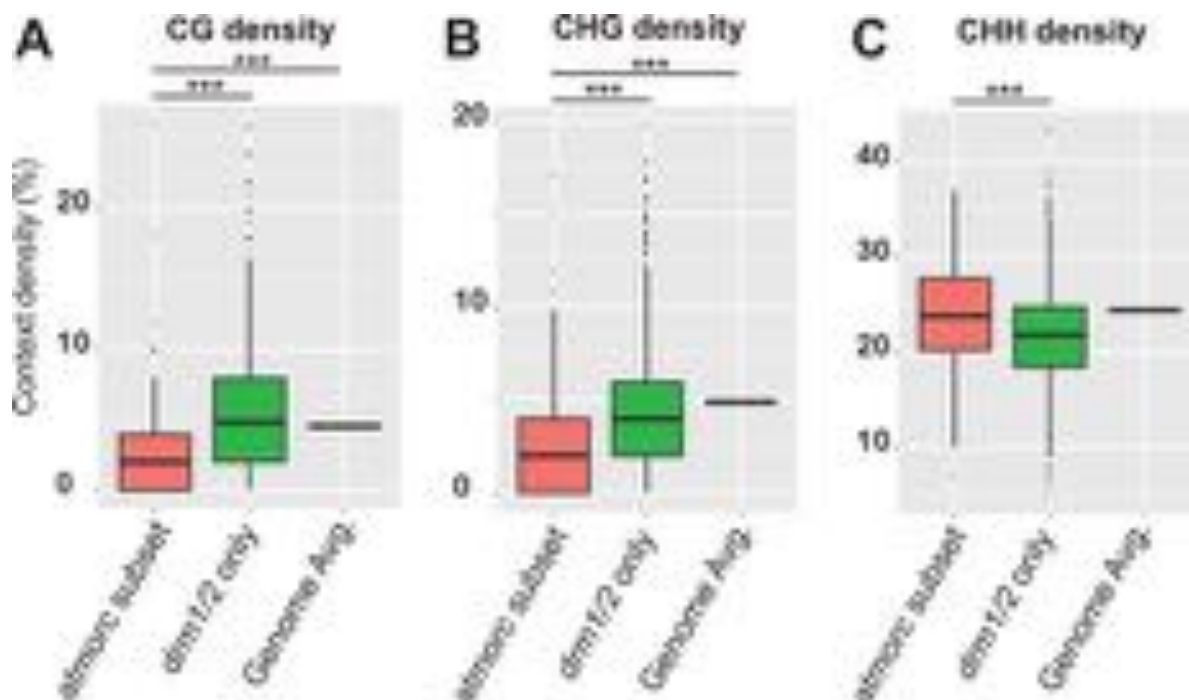


Fig 7. *atMORC* defined transcriptionally activatable subset of RdDM loci are characterized by reduced symmetric CG and CHG site density. In Fig 6 we showed that *atMORC* hypo CHH DMRs defined a subset of RdDM loci that become transcriptionally reactivated when CHH methylation is lost. Here we calculated density per base pair of CG (A), CHG (B), and CHH (C) sites at this subset of RdDM loci, termed '*atMORC* subset' (defined as the intersect between *atMORC*1/2/4/5/6/7 and *drm1/2* hypo CHH DMRs, $n = 279$, see Fig 6A), and compare it to the rest of RdDM loci, termed '*drm1/2* only' ($n = 4770$, see Fig 6A), and the genome average 'Genome Avg.'. While asymmetric CHH density is relatively high at the '*atMORC* subset', the density of sites for symmetric CG and CHG methylation are depleted by approximately half as compared to the '*drm1/2* only' loci and the genome average. Counts of CG, CHG, and CHH reflect presence on either strand, ie 2% CG indicates two CpG sites—one on each strand—for every 100bps. *** indicates statistically significant difference, $p < 0.001$.

doi:10.1371/journal.pgen.1005998.g007

maintenance of any of the major DNA methylation pathways and that major changes in transcription were not generally accompanied by loss of DNA methylation. However, at a small subset of RdDM targets (approximately 5%), AtMORC was required for both methylation and silencing, suggesting that these methylation losses are likely an indirect consequence of the loss of gene silencing. These findings reconcile our laboratory's previous reports of methylation-independent silencing [25] with that of other laboratories reporting hypomethylation at specific de-repressed reporter loci in *atMORC6* mutant backgrounds [26,27].

We recently reported that mouse MORC1 is required for DNA methylation and silencing at a specific subset of transposon promoters that are normally methylated at a developmentally late stage during the wave of global *de novo* methylation in the male germ line [53]. As in *Arabidopsis*, there were no genome wide changes in DNA methylation in the mouse *morc1* mutant, but specific methylation defects at a class of transposons that failed to establish silencing. These commonalities suggest that *Arabidopsis* MORCs may act similarly to mammalian MORC1, to maintain silencing at loci that are poised for transcriptional de-repression, with DNA hypomethylation as a secondary effect.

Nuclear localization of AtMORC4 and AtMORC7 broadly reflected that of their euchromatic gene and pericentromeric transposon targets, with both chromocenter adjacent enrichment and distribution throughout the nucleus. Since we previously reported that AtMORC6 and AtMORC1 form chromocenter adjacent bodies [25] (and see Fig 5), this appears to be a general feature of *Arabidopsis* MORC proteins, although the function of these bodies is at

present completely unknown. In the future, it will be important to determine the precise molecular mechanisms by which MORC proteins interact with chromatin and regulate gene expression.

Materials and Methods

Plant materials and growth

Wild-type and all mutant lines are from the ecotype Columbia (Col-0) and were grown under either continuous light ([S1 Fig](#), [Fig 2](#)) or long days (16 hour light—all other experiment). The T-DNA lines used in this study were: *atmorc1-2* (gene AT4G36290) SAIL_893_B06 (aka *crt1-2*), *atmorc2-1* (gene AT4G36280) SALK_072774C (aka *crh1-1*), *atmorc3-2* (gene AT4G36270) SALK_043244, *atmorc4-1* (gene AT5G50780) GK-249F08 (aka *crh4-2*), *atmorc5-1* (gene AT5G13130) SALK_049050C (aka *crh5-2*), *atmorc6-3* (gene AT1G19100) GABI_599B06 (aka *crh6-5*), and *atmorc7-1* (gene AT4G24970) SALK_051729 (aka *crh3-1*). T-DNAs were confirmed by PCR based genotyping. Primer sequences are described in [S1 Table](#).

Plasmid construction and transgenic plants

The pAtMORC4::AtMORC4-MYC, pAtMORC4::AtMORC4-FLAG, pAtMORC7::AtMORC7-MYC, and pAtMORC7::AtMORC7-MYC constructs were generated by the same method described in [\[35\]](#). Briefly, the AtMORC4 and AtMORC7 genomic regions, including ~1 kb upstream from the transcriptional start sites, were PCR amplified and cloned into a pENTR/D-TOPO vector (#K2400-20, Thermo Fisher). The cloned genomic regions were then transferred into a pEG302 based binary destination vector that included a MYC or FLAG epitope tag at the C-terminus via a Gateway LR Clonase II reaction (#11791–100, Thermo Fisher). *Agrobacterium tumefaciens* AGLO strain carrying these constructs were used to transform *A. thaliana* plants in their respective mutant backgrounds using the floral dip method [\[54\]](#).

BS-seq libraries

2–3 leaves from individual 3-week old plants were used to make individual BS-seq libraries based on methods described by [\[49\]](#). Briefly, genomic DNA was extracted using DNeasy Plant Mini kit (#69106) and 500ng was sheared using the Covaris S2 instrument. Libraries were generated using the Kapa Hyper Prep Kit (#KK8502) with bisulfite conversion using the EZ DNA Methylation Lightning Kit (#D5030). Libraries were sequenced on a HiSeq 2000 (Illumina).

RNA-seq libraries and RT-PCRs

RNA was extracted from 2–3 leaves of 3-week old plants using Trizol reagent and DNase treated using TURBO DNA-free kit (#AM1907). For RNA-seq, 1–2.5 µg of RNA starting material per library was first rRNA depleted using Epicentre RiboZero (#MRZPL1224) prior to library generation using Epicentre ScriptSeqv2 (#SSV21124). Libraries were sequenced on a HiSeq 2000 (Illumina). For RT-PCRs, cDNA was generated using SuperScript III (#18080–044, ThermoFisher) with random hexamer priming. The samples were digested with RNase H in accordance with manufacturer's protocol. RT-PCR was then performed with iQ SYBR Green Mastermix (BioRad) using an Agilent Technologies Mx3005p qPCR System (Stratagene).

Hpa assay

Hyaloperonospora arabidopsidis (Hpa) isolate Emwa1 was propagated on the susceptible *Arabidopsis* ecotype Ws. Conidiospores of Hpa strain Emwa1 were resuspended in autoclaved RO-water at a concentration of 3×10^4 spores/mL and spray-inoculated onto 10-day old seedlings.

Inoculated plants were covered with a lid to increase humidity and grown at 19°C under a 9-hour light period. Sporangioophores per cotyledon were counted 4 to 5 days post inoculation using a Leica M205 FA stereoscope. The experiments were repeated 3 times and the sporangioophores on approximately 100 cotyledons per genotype were counted in each experiment.

Co-Immunoprecipitation (Co-IP) and Immunoprecipitation Mass spectrometry (IP-MS)

Co-IP and IP-MS on pAtMORC4::AtMORC4-MYC/FLAG and pAtMORC7::AtMORC7-MYC/FLAG lines were performed as previously described [35]. For IP-MS, M2 magnetic FLAG-beads (SIGMA, M8823) were added to the supernatant and immunoprecipitated proteins were eluted using 3×FLAG peptides (SIGMA, F4799). The MS was performed as described by [55]. For the Co-IPs, we added 100 µL M2 magnetic FLAG-beads (SIGMA, M8823) to the supernatant for pulldown. For the western blots, we used HRP-coupled FLAG-specific antibody (SIGMA, A8592) and MYC-specific antibodies (Pierce, MA1-980).

Nuclear immunofluorescence

Nuclear immunofluorescence experiments for AtMORC4/7-MYC tagged lines were performed based on the method described in [25]. Leaves from three-week old plants were fixed in 4% paraformaldehyde in TRIS buffer (10 mM TRIS pH 7.5, 10 mM EDTA, and 100 mM NaCl) for 20 minutes and washed twice in TRIS buffer. Leaves were chopped in 200–400 microliters lysis buffer (15 mM TRIS pH 7.5, 2 mM EDTA, 0.5 mM spermine, 80 mM KCl, 20 mM NaCl, and 0.1% Triton X-100) and filtered through a 3 µm cell strainer (Corning, #352235). 5 µL of nuclei suspension was added to 12 µL of sorting buffer (100mM TRIS pH 7.5, 50mM KCl, 2mM MgCl₂, 0.05% Tween-20, and 20.5% sucrose) and air dried on chloroform dipped microscope slides for two hours and then post-fixed in 4% paraformaldehyde in PBS for 20 minutes. Slides were washed three times in PBS and incubated in blocking buffer (3% BSA, and 10% horse serum in PBS) for 30 minutes at 37°C. Nuclei were incubated at 4°C overnight in mouse monoclonal antibody against c-Myc (9E10, Abcam ab32; 1:200). Slides were washed in PBS and incubated with goat anti-mouse FITC antibody (Abcam, ab7064; 1:200) for 90 minutes at room temperature. Following PBS washes, nuclei were counterstained and mounted in Vectashield mounting media with DAPI (Vector, H-1200). Nuclei were analyzed with a Zeiss LSM 710 Confocal microscope at 63X or 100X magnification using Zen software.

Bioinformatics

For RNA-seq analysis, reads were aligned with TopHat, including the fr-secondstrand parameter. Cufflinks was used to generate count data using annotation from TAIR10 that was fed into the DESeq2 package in R for differential expression analysis. For BS-seq, reads were aligned using BSMAP with methylation levels calculated and DMRs defined as previously described [49]. For the *atmorc* DMRs, each biological replicated (two per mutant) was compared against two wild type biological replicates from the same experiment, requiring that the DMR be identified in all four mutant vs. wt comparisons to be considered a ‘true’ DMR. The *dmr1/2*, *cmt2*, *cmt3*, and *met1* DMRs were previously defined [49], using a single mutant biological replicate compared against three biological wild type replicates.

Data deposition

The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database (accession number GSE78836).

Supporting Information

S1 Fig. *atmorc4/7* double mutant shows de-repression at AtMORC6 transposon targets.

(A) RT-PCR on cDNA derived from *atmorc4-1/atmorc7-1* double mutant compared to wt showing no detectable wild type transcript in these T-DNA mutants. Primers were designed to span the T-DNA region in *atmorc4-1* (upper) and *atmorc7-1* (middle) (S1 Table). UBQ10 (lower) was amplified as a loading control (S1 Table). (B) RT-PCR at AtMORC6 targets indicated using the genotypes indicated. Error bars indicate standard error of the mean (SEM). (PDF)

S2 Fig. *AtMORC3* is likely to be a pseudogene. (A) TAIR predicted gene structure for *AtMORC1*, *AtMORC2*, and *AtMORC3*. Boxes = exons, light blue = UTR, and dark blue = CDS. *AtMORC1*, *AtMORC2*, and *AtMORC3* are highly related to one another, (see Fig 1A, and (B) below), encode the same number of exons, and lie directly adjacent to one another on *A. thaliana* chromosome four, indicating that they likely arose from a tandem duplication event. In the predicted 5' UTR of *AtMORC3*, there is an ATG start codon. However, a G to A mutation causes a W to Stop codon in exon three. BLAST of this *in silico* translated region identifies all other AtMORC proteins. However, because this ORF is predicted to be too small, TAIR finds the next in-frame ATG in exon 5, annotating this to be the translational start. If this protein were made, it would be N-terminally truncated, missing half of the GHKL ATPase including two out of the four motifs thought to be essential for ATP binding [28,29]. (B) Phylogenetic reconstruction of *AtMORC* genes in *Arabidopsis thaliana* and close relatives, *Capsella rubella* and *Arabidopsis lyrata*. The tandem arrangement of *AtMORC1*, *AtMORC2*, and *AtMORC3*, and the premature stop codon identified in *AtMORC3* is consistent with the pseudogenisation of a redundant paralogue. Therefore, we checked whether *AtMORC1*, *AtMORC2*, and *AtMORC3* are also present in *A. thaliana* sister species. We found that while the closely related *A. lyrata* encodes a single copy of each of *A. thaliana*'s *AtMORC* genes, the slightly more distantly related *C. rubella* does not encode a copy of either *AtMORC2* or *AtMORC3* (and encodes two copies of *AtMORC4*). Therefore *C. rubella* has either lost its versions of *AtMORC2*/*AtMORC3* or the tandem duplication of *AtMORC1* occurred after the divergence of *A. thaliana* and *A. lyrata* from *C. rubella*. In either scenario, it suggests that *AtMORC2* and *AtMORC3* are likely non-essential and may act redundantly with *AtMORC1*. In support of this hypothesis, we have already shown that *AtMORC2* is redundant with *AtMORC1* [35]. (C) Positions of the SALK_000009 and SALK_043244 insertions in *AtMORC3*. (D) Sequence of SALK_043244 T-DNA homozygous insert in *AtMORC3*. As the SALK_000009 line, which has a T-DNA insert in the 5' UTR of *AtMORC3*, was found to be embryonic lethal [36], we took an independent *AtMORC3* T-DNA line to homozygosity and sequence confirmed the presence of the insert in exon 11, finding that this line displays no discernable phenotype. Together with the premature stop codon in exon 3, it is likely that *AtMORC3* is a non-functional pseudogene in Columbia-0. (PDF)

S3 Fig. Comparison of RNA-seq in *atmorc4/6/7* vs. *atmorc1/2/4/5/6/7*. (A) Overlap between *atmorc4/6/7* and *atmorc1/2/4/5/6/7* upregulated DEGs. (B) Boxplot showing the FPKM (fragments per kilobase per million reads) for the 241 genes in *atmorc1/2/4/5/6/7* that did not overlap with *atmorc4/6/7* (purple section in (A)). This shows that while these genes did not make the significance cutoff required to be called DEGs in *atmorc4/6/7*, they still show the same trend for upregulation, indicating that the addition of *atmorc1*, 2 and 5 has very little additional impact on the transcriptome (also see Fig 3D). (PDF)

S4 Fig. DEGs in *atmorc4/7* are highly enriched for pathogen defense. (A) Top ten listed GO term categories from *atmorc4/7* misregulated genes (FDR<0.05) [<http://bioinfo.cau.edu.cn/agriGO>] identified RNA-seq round 2 (see Fig 3). (B) Top ten listed GO term categories from *atmorc4/7* misregulated genes (FDR<0.05) [<http://bioinfo.cau.edu.cn/agriGO>] identified RNA-seq round 1 (see Fig 1).

(PDF)

S5 Fig. No additive transcriptional effect at ‘response to chitin’ genes in higher-order *atmorc* knockouts. Boxplot showing FPKMs at the ‘response to chitin’ gene set (GO:0010200) in the genotypes indicated.

(PDF)

S6 Fig. Negligible DNA methylation changes genome wide and at AtMORC targets in *AtMORC* knockouts. (A) Genome wide profiles of CG, CHG, and CHH context methylation in the wt, *atmorc4/7*, *atmorc6*, and *atmorc1/2/4/5/6/7* backgrounds. Average of two biological replicates of each genotype, except *atmorc6* (data obtained from GSE54677) [35]. (B) Metaplot of methylation levels in wt, *atmorc4/7* and *atmorc1/2/4/5/6/7* over DEGs (>2 fold change, FDR<0.05) in *atmorc1/2/4/5/6/7* background, in CG, CHG and CHH contexts. TSS = transcriptional start site, TTS = transcriptional termination site.

(PDF)

S7 Fig. Loss of AtMORC does not significantly impact any of the major DNA methylation pathways and does not act downstream of DNA methylation. (A) Boxplots for methylation levels at *drm1/2* CHH, *cmt2* CHH, *cmt3* CHG, and *met1* CG defined hypomethylated DMRs [8,49] in the wt, *atmorc4/7*, *atmorc6*, *atmorc1/2/4/5/6/7*, and control methyltransferase mutant backgrounds indicated. (B) RNA-seq from wt and *atmorc1/2/4/5/6/7* (black and green, respectively, three replicates each, see Fig 3) over methylated loci defined by *drm1/2* CHH, *cmt2* CHH, *cmt3* CHG, and *met1* CG hypo DMRs (as in (A)).

(PDF)

S8 Fig. *atmorc1/2/4/5/6/7* hypo CHH DMRs overlap with RdDM sites. (A) Overlap of *atmorc1/2/4/5/6/7* defined hypo CHH DMRs with previously defined *drm1/2* and *cmt2* hypo CHH DMRs [8,49]. (B) Overlap of *atmorc1/2/4/5/6/7* hypo CHH DMRs with CHH loci prone to spontaneous epiallelic variation [50].

(PDF)

S9 Fig. Comparison of *atmorc6* with *atmorc4/7* at *atmorc1/2/4/5/6/7* hypo CHH DMRs. (A) Heatmap showing CHH methylation levels at all *atmorc1/2/4/5/6/7* hypo CHH DMRs in the genotypes indicated. *atmorc4/7* and *atmorc6* appear to affect many similar targets. Scale 0–0.6 indicates CHH methylation level. (B) Boxplot for methylation levels at same *atmorc1/2/4/5/6/7* hypo CHH DMRs as in (A). *drm1/2* is used as a control in (A) and (B), and demonstrates that *atmorc* hypo CHH DMRs are primarily RdDM target loci.

(PDF)

S10 Fig. *atmorc1/2/4/5/6/7* hypo CHH DMRs show evidence for transcriptional de-repression. (A) RNA-seq metaplot of wt vs. *atmorc1/2/4/5/6/7* (black and green, respectively, three replicates each, see Fig 3) over *atmorc1/2/4/5/6/7* defined hypo CHH DMRs. (B) RNA-seq metaplot of wt vs. *drm1/2* (black and red, respectively, two replicates each) over *drm1/2* hypo CHH DMRs (data from GEO:GSE51304) [8].

(PDF)

S1 Video. AtMORC7-MYC rotate. z-stack at 0.83 μ M intervals through the AtMORC7-MYC expressing nucleus depicted in Fig 5A was rendered in 3D with interpolation and rotated 360 degrees about the y-axis. Blue channel = DAPI staining; green channel = anti-MYC staining. (AVI)

S2 Video. AtMORC7-MYC stack. z-stack at 0.83 μ M intervals through the AtMORC7-MYC expressing nucleus depicted in Fig 5A. z-stack slices from the furthest to closest depth are shown in sequence (5 frames per second), illustrating the presence of AtMORC7-MYC bodies first at one chromocenter (upper middle of nucleolus) and then more prominently at another (middle left, between nucleolus and nuclear periphery). Blue channel = DAPI staining; green channel = anti MYC staining. Scale bar = 2 μ M (AVI)

S1 Table. Primers used in this study. List of relevant primers used in the study. (PDF)

Acknowledgments

We thank Jeffrey A. Long for providing access to and support for the Zeiss LSM 710 confocal microscope, and Suhua Feng, Mahnaz Akhavan, Yifan Ma, Jett Appel, Jamie Ho, Danielle Smith, Ryan Narbutas, and Lisa Wünsch for technical assistance. High-throughput sequencing was performed at the UCLA Broad Stem Cell Research Center BioSequencing Core Facility.

Author Contributions

Conceived and designed the experiments: CJH DH FEK JLD SEJ. Performed the experiments: CJH DH FEK AP WAP GM AAV. Analyzed the data: WL CJH HW. Contributed reagents/materials/analysis tools: AAV JAW. Wrote the paper: CJH SEJ.

References

1. Roudier F, Ahmed I, Berard C, Sarazin A, Mary-Huard T, Cortijo S, et al. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *Embo J*. 2011; 30(10):1928–38. doi: [10.1038/emboj.2011.103](https://doi.org/10.1038/emboj.2011.103) PMID: [21487388](https://pubmed.ncbi.nlm.nih.gov/21487388/)
2. Kharchenko P V, Alekseyenko A a, Schwartz YB, Minoda A, Riddle NC, Ernst J, et al. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* [Internet]. Nature Publishing Group; 2011; 471(7339):480–5. Available from: <http://dx.doi.org/10.1038/nature09725> doi: [10.1038/nature09725](https://doi.org/10.1038/nature09725) PMID: [21179089](https://pubmed.ncbi.nlm.nih.gov/21179089/)
3. Law J a, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* [Internet]. Nature Publishing Group; 2010 Mar [cited 2013 Jan 30]; 11(3):204–20. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3034103&tool=pmcentrez&rendertype=abstract> doi: [10.1038/nrg2719](https://doi.org/10.1038/nrg2719) PMID: [20142834](https://pubmed.ncbi.nlm.nih.gov/20142834/)
4. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*. 2008; 452(7184):215–9. doi: [10.1038/nature06745](https://doi.org/10.1038/nature06745) PMID: [18278030](https://pubmed.ncbi.nlm.nih.gov/18278030/)
5. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar a H, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* [Internet]. 2008 May 2 [cited 2014 May 23]; 133(3):523–36. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2723732&tool=pmcentrez&rendertype=abstract> doi: [10.1016/j.cell.2008.03.029](https://doi.org/10.1016/j.cell.2008.03.029) PMID: [18423832](https://pubmed.ncbi.nlm.nih.gov/18423832/)
6. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol* [Internet]. Nature Publishing Group; 2015; 16(9):519–32. Available from: <http://www.nature.com/doi/10.1038/nrm4043> doi: [10.1038/nrm4043](https://doi.org/10.1038/nrm4043) PMID: [26296162](https://pubmed.ncbi.nlm.nih.gov/26296162/)
7. Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* [Internet]. Elsevier Inc.; 2013 Mar 28 [cited 2013 May 21]; 153(1):193–205. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23540698> doi: [10.1016/j.cell.2013.02.033](https://doi.org/10.1016/j.cell.2013.02.033) PMID: [23540698](https://pubmed.ncbi.nlm.nih.gov/23540698/)

8. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* [Internet]. Nature Publishing Group; 2014 Jan [cited 2014 Apr 29]; 21(1):64–72. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24336224> doi: [10.1038/nsmb.2735](https://doi.org/10.1038/nsmb.2735) PMID: [24336224](https://pubmed.ncbi.nlm.nih.gov/24336224/)
9. Cao X, Jacobsen SE. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* [Internet]. 2002 Jul 9; 12(13):1138–44. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12121623> PMID: [12121623](https://pubmed.ncbi.nlm.nih.gov/12121623/)
10. Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, et al. Molecular Mechanism of Action of Plant DRM De Novo DNA Methyltransferases. *Cell* [Internet]. Elsevier Inc.; 2014 May [cited 2014 May 22]; 157(5):1050–60. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0092867414004905> doi: [10.1016/j.cell.2014.03.056](https://doi.org/10.1016/j.cell.2014.03.056) PMID: [24855943](https://pubmed.ncbi.nlm.nih.gov/24855943/)
11. Matzke M a., Mosher R a. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet* [Internet]. Nature Publishing Group; 2014 May 8 [cited 2014 May 9]; 15(6):394–408. Available from: <http://www.nature.com/doi/10.1038/nrg3683> doi: [10.1038/nrg3683](https://doi.org/10.1038/nrg3683) PMID: [24805120](https://pubmed.ncbi.nlm.nih.gov/24805120/)
12. Matzke M a, Kanno T, Matzke AJM. RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annu Rev Plant Biol* [Internet]. 2014 Dec 10 [cited 2014 Dec 11]; (December 2014):1–25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25494460>
13. Bond DM, Baulcombe DC. Small RNAs and heritable epigenetic variation in plants. *Trends Cell Biol* [Internet]. Elsevier Ltd; 2013 Sep [cited 2013 Sep 6]; 1–8. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0962892413001347>
14. Law J a, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* [Internet]. Nature Publishing Group; 2013 May 1 [cited 2013 May 25]; 498(7454):385–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23636332> doi: [10.1038/nature12178](https://doi.org/10.1038/nature12178) PMID: [23636332](https://pubmed.ncbi.nlm.nih.gov/23636332/)
15. Zhai J, Bischof S, Wang H, Feng S, Lee T, Teng C, et al. A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis. *Cell* [Internet]. Elsevier Inc.; 2015; 163(2):445–55. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0092867415011940> doi: [10.1016/j.cell.2015.09.032](https://doi.org/10.1016/j.cell.2015.09.032) PMID: [26451488](https://pubmed.ncbi.nlm.nih.gov/26451488/)
16. Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, et al. Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *Elife*. 2015;(October).
17. Li S, Vandivier LE, Tu B, Gao L, Won SY, Zheng B, et al. Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis. *Genome Res* [Internet]. 2014 Nov 20 [cited 2014 Nov 21]; Available from: <http://genome.cshlp.org/cgi/doi/10.1101/gr.182238.114>
18. Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, et al. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* [Internet]. Nature Publishing Group; 2014 Jan 22 [cited 2014 Jan 22]; Available from: <http://www.nature.com/doi/10.1038/nature12931>
19. Liu Z-W, Shao C-R, Zhang C-J, Zhou J-X, Zhang S-W, Li L, et al. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLoS Genet* [Internet]. 2014 Jan [cited 2014 Dec 2]; 10(1):e1003948. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3898904&tool=pmcentrez&rendertype=abstract> doi: [10.1371/journal.pgen.1003948](https://doi.org/10.1371/journal.pgen.1003948) PMID: [24465213](https://pubmed.ncbi.nlm.nih.gov/24465213/)
20. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* [Internet]. 2009; 41(5):630–4. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2674513&tool=pmcentrez&rendertype=abstract> doi: [10.1038/ng.365](https://doi.org/10.1038/ng.365) PMID: [19377477](https://pubmed.ncbi.nlm.nih.gov/19377477/)
21. Böhmendorfer G, Rowley MJ, Kuciński J, Zhu Y, Amies I, Wierzbicki AT. RNA-directed DNA methylation requires stepwise binding of silencing factors to long non-coding RNA. *Plant J* [Internet]. 2014 May 26 [cited 2014 Jul 10]; 181–91. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24862207>
22. Bond DM, Baulcombe DC. Epigenetic transitions leading to heritable, RNA-mediated de novo silencing in Arabidopsis thaliana. *Proc Natl Acad Sci* [Internet]. 2015; 112(3):917–22. Available from: <http://www.pnas.org/lookup/doi/10.1073/pnas.1413053112> doi: [10.1073/pnas.1413053112](https://doi.org/10.1073/pnas.1413053112) PMID: [25561534](https://pubmed.ncbi.nlm.nih.gov/25561534/)
23. McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK. ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *Embo J*. 2014;1–16.
24. Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, et al. The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21–22 nucleotide small interfering

- RNAs. *Plant Physiol* [Internet]. 2013 May [cited 2014 May 26]; 162(1):116–31. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3641197&tool=pmcentrez&rendertype=abstract> doi: [10.1104/pp.113.216481](https://doi.org/10.1104/pp.113.216481) PMID: [23542151](https://pubmed.ncbi.nlm.nih.gov/23542151/)
25. Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, Stroud H, et al. MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* [Internet]. 2012 Jun 15; 336(6087):1448–51. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3376212&tool=pmcentrez&rendertype=abstract> doi: [10.1126/science.1221472](https://doi.org/10.1126/science.1221472) PMID: [22555433](https://pubmed.ncbi.nlm.nih.gov/22555433/)
26. Lorković ZJ, Naumann U, Matzke AJM, Matzke M. Involvement of a GHKL ATPase in RNA-directed DNA methylation in *Arabidopsis thaliana*. *Curr Biol* [Internet]. 2012 May 22 [cited 2014 Aug 27]; 22(10):933–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22560611> doi: [10.1016/j.cub.2012.03.061](https://doi.org/10.1016/j.cub.2012.03.061) PMID: [22560611](https://pubmed.ncbi.nlm.nih.gov/22560611/)
27. Brabbs TR, He Z, Hogg K, Kamenski A, Li Y, Paszkiewicz KH, et al. The stochastic silencing phenotype of *Arabidopsis morc6* mutants reveals a role in efficient RNA-directed DNA methylation. *Plant J* [Internet]. 2013 Sep [cited 2014 Nov 13]; 75(5):836–46. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23675613> doi: [10.1111/tpj.12246](https://doi.org/10.1111/tpj.12246) PMID: [23675613](https://pubmed.ncbi.nlm.nih.gov/23675613/)
28. Dutta R, Inouye M. GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci*. 2000; 25(1):24–8. PMID: [10637609](https://pubmed.ncbi.nlm.nih.gov/10637609/)
29. Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A, Forterre P. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature*. 1997. p. 414–7.
30. Iyer LM, Abhiman S, Aravind L. MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol Direct* [Internet]. 2008 Jan [cited 2014 Sep 1]; 3:8. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2292703&tool=pmcentrez&rendertype=abstract> doi: [10.1186/1745-6150-3-8](https://doi.org/10.1186/1745-6150-3-8) PMID: [18346280](https://pubmed.ncbi.nlm.nih.gov/18346280/)
31. Kang H-G, Kuhl JC, Kachroo P, Klessig DF. CRT1, an *Arabidopsis* ATPase that interacts with diverse resistance proteins and modulates disease resistance to turnip crinkle virus. *Cell Host Microbe* [Internet]. 2008 Jan 17 [cited 2013 Feb 9]; 3(1):48–57. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18191794> doi: [10.1016/j.chom.2007.11.006](https://doi.org/10.1016/j.chom.2007.11.006) PMID: [18191794](https://pubmed.ncbi.nlm.nih.gov/18191794/)
32. Luna E, Bruce TJ a, Roberts MR, Flors V, Ton J. Next-generation systemic acquired resistance. *Plant Physiol* [Internet]. 2012 Mar [cited 2013 Jan 30]; 158(2):844–53. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3271772&tool=pmcentrez&rendertype=abstract> doi: [10.1104/pp.111.187468](https://doi.org/10.1104/pp.111.187468) PMID: [22147520](https://pubmed.ncbi.nlm.nih.gov/22147520/)
33. Downen RH, Pelizzola M, Schmitz RJ, Lister R, Downen JM, Nery JR, et al. Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci U S A* [Internet]. 2012 Aug 7 [cited 2014 Jul 11]; 109(32):E2183–91. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3420206&tool=pmcentrez&rendertype=abstract> doi: [10.1073/pnas.1209329109](https://doi.org/10.1073/pnas.1209329109) PMID: [22733782](https://pubmed.ncbi.nlm.nih.gov/22733782/)
34. Yu A, Lepère G, Jay F, Wang J, Bapaume L, Wang Y, et al. Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. *Proc Natl Acad Sci U S A* [Internet]. 2013 Feb 5 [cited 2014 Oct 15]; 110(6):2389–94. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3568381&tool=pmcentrez&rendertype=abstract> doi: [10.1073/pnas.1211757110](https://doi.org/10.1073/pnas.1211757110) PMID: [23335630](https://pubmed.ncbi.nlm.nih.gov/23335630/)
35. Moissiard G, Bischof S, Husmann D, Pastor W a, Hale CJ, Yen L, et al. Transcriptional gene silencing by *Arabidopsis* microRNA homologues involves the formation of heteromers. *Proc Natl Acad Sci U S A* [Internet]. 2014 May 20 [cited 2014 Sep 17]; 111(20):7474–9. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4034193&tool=pmcentrez&rendertype=abstract> doi: [10.1073/pnas.1406611111](https://doi.org/10.1073/pnas.1406611111) PMID: [24799676](https://pubmed.ncbi.nlm.nih.gov/24799676/)
36. Kang H-G, Oh C-S, Sato M, Katagiri F, Glazebrook J, Takahashi H, et al. Endosome-associated CRT1 functions early in resistance gene-mediated defense signaling in *Arabidopsis* and tobacco. *Plant Cell* [Internet]. 2010 Mar [cited 2011 Sep 7]; 22(3):918–36. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2861469&tool=pmcentrez&rendertype=abstract> doi: [10.1105/tpc.109.071662](https://doi.org/10.1105/tpc.109.071662) PMID: [20332379](https://pubmed.ncbi.nlm.nih.gov/20332379/)
37. Henderson IR, Jacobsen SE. Tandem repeats upstream of the *Arabidopsis* endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev* [Internet]. 2008 Jun 15 [cited 2014 Oct 22]; 22(12):1597–606. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2428058&tool=pmcentrez&rendertype=abstract> doi: [10.1101/gad.1667808](https://doi.org/10.1101/gad.1667808) PMID: [18559476](https://pubmed.ncbi.nlm.nih.gov/18559476/)
38. Boller T, Felix G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu Rev Plant Biol* [Internet]. 2009 Jan [cited 2013 Aug 8]; 60:379–406. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19400727> doi: [10.1146/annurev.arplant.57.032905.105346](https://doi.org/10.1146/annurev.arplant.57.032905.105346) PMID: [19400727](https://pubmed.ncbi.nlm.nih.gov/19400727/)
39. Knoth C, Eulgem T. The oomycete response gene LURP1 is required for defense against *Hyaloperonospora parasitica* in *Arabidopsis thaliana*. *Plant J*. 2008; 55(1):53–64. doi: [10.1111/j.1365-3113X.2008.03486.x](https://doi.org/10.1111/j.1365-3113X.2008.03486.x) PMID: [18346188](https://pubmed.ncbi.nlm.nih.gov/18346188/)

40. Lu D, Lin W, Gao X, Wu S, Cheng C, Avila J, et al. Direct ubiquitination of pattern recognition receptor FLS2 attenuates plant innate immunity. *Science* [Internet]. 2011 Jun 17 [cited 2011 Jul 24]; 332(6036):1439–42. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21680842> doi: [10.1126/science.1204903](https://doi.org/10.1126/science.1204903) PMID: [21680842](https://pubmed.ncbi.nlm.nih.gov/21680842/)
41. Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, et al. Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* [Internet]. Nature Publishing Group; 2010 Jun 3 [cited 2013 Sep 19]; 465(7298):632–6. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3055268&tool=pmcentrez&rendertype=abstract> doi: [10.1038/nature09083](https://doi.org/10.1038/nature09083) PMID: [20520716](https://pubmed.ncbi.nlm.nih.gov/20520716/)
42. Hernandez-Pinzon I, Yelina NE, Schwach F, Studholme DJ, Baulcombe D, Dalmay T. SDE5, the putative homologue of a human mRNA export factor, is required for transgene silencing and accumulation of trans-acting endogenous siRNA. *Plant J*. 2007; 50(1):140–8. PMID: [17397509](https://pubmed.ncbi.nlm.nih.gov/17397509/)
43. Eitas TK, Dangl JL. NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr Opin Plant Biol* [Internet]. Elsevier Ltd; 2010 Aug [cited 2013 Feb 9]; 13(4):472–7. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2910844&tool=pmcentrez&rendertype=abstract> doi: [10.1016/j.pbi.2010.04.007](https://doi.org/10.1016/j.pbi.2010.04.007) PMID: [20483655](https://pubmed.ncbi.nlm.nih.gov/20483655/)
44. Wang W, Barnaby JY, Tada Y, Li H, Tör M, Caldelari D, et al. Timing of plant immune responses by a central circadian regulator. *Nature* [Internet]. 2011 Feb 3 [cited 2011 Jul 21]; 470(7332):110–4. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21293378> doi: [10.1038/nature09766](https://doi.org/10.1038/nature09766) PMID: [21293378](https://pubmed.ncbi.nlm.nih.gov/21293378/)
45. Kang H-G, Hyong WC, von Einem S, Manosalva P, Ehlers K, Liu P-P, et al. CRT1 is a nuclear-translocated MORC endonuclease that participates in multiple levels of plant immunity. *Nat Commun* [Internet]. 2012 Jan [cited 2013 Feb 9]; 3:1297. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23250427> doi: [10.1038/ncomms2279](https://doi.org/10.1038/ncomms2279) PMID: [23250427](https://pubmed.ncbi.nlm.nih.gov/23250427/)
46. Franz P, De Jong JH, Lysak M, Castiglione MR, Schubert I. Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc Natl Acad Sci U S A* [Internet]. 2002; 99(22):14584–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12384572> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=137926&tool=pmcentrez&rendertype=abstract> PMID: [12384572](https://pubmed.ncbi.nlm.nih.gov/12384572/)
47. Schubert I, Shaw P. Organization and dynamics of plant interphase chromosomes. *Trends Plant Sci* [Internet]. Elsevier Ltd; 2011; 16(5):273–81. Available from: <http://dx.doi.org/10.1016/j.tplants.2011.02.002> doi: [10.1016/j.tplants.2011.02.002](https://doi.org/10.1016/j.tplants.2011.02.002) PMID: [21393049](https://pubmed.ncbi.nlm.nih.gov/21393049/)
48. Feng S, Cokus SJ, Schubert V, Zhai J, Pellegrini M, Jacobsen SE. Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in *Arabidopsis*. *Mol Cell* [Internet]. Elsevier Inc.; 2014 Aug [cited 2014 Aug 18]; 55(5):694–707. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1097276514006017> doi: [10.1016/j.molcel.2014.07.008](https://doi.org/10.1016/j.molcel.2014.07.008) PMID: [25132175](https://pubmed.ncbi.nlm.nih.gov/25132175/)
49. Stroud H, Greenberg MVC, Feng S, Bernatavichute Y V, Jacobsen SE. Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the *Arabidopsis* Methylome. *Cell* [Internet]. Elsevier Inc.; 2012; 152(1–2):352–64. Available from: <http://dx.doi.org/10.1016/j.cell.2012.10.054>
50. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, et al. Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. *Science* (80-). 2011; 334(October):369–73. doi: [10.1126/science.1212959](https://doi.org/10.1126/science.1212959) PMID: [21921155](https://pubmed.ncbi.nlm.nih.gov/21921155/)
51. Greenberg MVC, Deleris A, Hale CJ, Liu A, Feng S, Jacobsen SE. Interplay between active chromatin marks and RNA-directed DNA methylation in *Arabidopsis thaliana*. *PLoS Genet* [Internet]. 2013 Nov [cited 2014 Oct 26]; 9(11):e1003946. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3820799&tool=pmcentrez&rendertype=abstract> doi: [10.1371/journal.pgen.1003946](https://doi.org/10.1371/journal.pgen.1003946) PMID: [24244201](https://pubmed.ncbi.nlm.nih.gov/24244201/)
52. Li Q, Gent JI, Zynda G, Song J, Makarevitch I, Hirsch CD, et al. RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. *Proc Natl Acad Sci U S A* [Internet]. 2015; 1514680112 –. Available from: <http://www.pnas.org/content/early/2015/11/05/1514680112.short?rss=1>
53. Pastor W a, Stroud H, Nee K, Liu W, Pezic D, Manakov S, et al. MORC1 represses transposable elements in the mouse male germline. *Nat Commun*. 2014;
54. Clough SJ, Bent AF. Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J*. 1998; 16(June 1998):735–43. PMID: [10069079](https://pubmed.ncbi.nlm.nih.gov/10069079/)
55. Du J, Zhong X, Bernatavichute Y V, Stroud H, Feng S, Caro E, et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* [Internet]. Elsevier Inc.; 2012 Sep 28 [cited 2013 Feb 6]; 151(1):167–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23021223> doi: [10.1016/j.cell.2012.07.034](https://doi.org/10.1016/j.cell.2012.07.034) PMID: [23021223](https://pubmed.ncbi.nlm.nih.gov/23021223/)

Transcriptional gene silencing by *Arabidopsis* microorchidia homologues involves the formation of heteromers

Guillaume Moissiard^{a,1,2}, Sylvain Bischof^{a,2}, Dylan Husmann^a, William A. Pastor^a, Christopher J. Hale^a, Linda Yen^a, Hume Stroud^{a,3}, Ashot Papikian^a, Ajay A. Vashisht^b, James A. Wohlschlegel^b, and Steven E. Jacobsen^{a,c,4}

^aDepartment of Molecular, Cell and Developmental Biology and ^cHoward Hughes Medical Institute, University of California, Los Angeles, CA 90095; and ^bDepartment of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, CA 90095

Contributed by Steven E. Jacobsen, April 14, 2014 (sent for review January 31, 2014)

Epigenetic gene silencing is of central importance to maintain genome integrity and is mediated by an elaborate interplay between DNA methylation, histone posttranslational modifications, and chromatin remodeling complexes. DNA methylation and repressive histone marks usually correlate with transcriptionally silent heterochromatin, however there are exceptions to this relationship. In *Arabidopsis*, mutation of *Morpheus Molecule 1* (*MOM1*) causes transcriptional derepression of heterochromatin independently of changes in DNA methylation. More recently, two *Arabidopsis* homologues of mouse microorchidia (*MORC*) genes have also been implicated in gene silencing and heterochromatin condensation without altering genome-wide DNA methylation patterns. In this study, we show that *Arabidopsis* microorchidia (*AtMORC6*) physically interacts with *AtMORC1* and with its close homologue, *AtMORC2*, in two mutually exclusive protein complexes. RNA-sequencing analyses of high-order mutants indicate that *AtMORC1* and *AtMORC2* act redundantly to repress a common set of loci. We also examined genetic interactions between *AtMORC6* and *MOM1* pathways. Although *AtMORC6* and *MOM1* control the silencing of a very similar set of genomic loci, we observed synergistic transcriptional regulation in the *mom1/atmorc6* double mutant, suggesting that these epigenetic regulators act mainly by different silencing mechanisms.

epigenetics | plant biology

DNA methylation and histone posttranslational modifications are essential for silencing of transposable elements (TEs) and other repeat sequences. In the plant model organism *Arabidopsis thaliana*, DNA methylation sites are found in three different cytosine contexts: CG, CHG, and CHH (in which H is A, T, or C) (1). Symmetric CG and CHG methylations are mediated by DNA Methyltransferase 1 (*MET1*) and Chromomethylase 3 (*CMT3*), respectively (2, 3). Asymmetric CHH methylation is maintained at nonoverlapping sites by *CMT2* and *Domains Rearranged Methyltransferase 2* (*DRM2*) (4, 5). In the RNA-directed DNA methylation (RdDM) pathway, de novo methylation of CHH sites is established by *DRM2* and involves 24-nucleotide small interfering RNAs and long noncoding RNAs (6–11). Genome-wide studies revealed that DNA methylation and repressive histone modifications such as dimethylation of histone 3 lysine 9 (H3K9me2) correlate with transcriptionally silent chromatin (12–16). Furthermore, transcriptional derepression of silenced methylated loci is accompanied by loss of DNA methylation. A prominent exception to this interdependence is the *Morpheus Molecule 1* (*MOM1*).

MOM1 is unique to the plant kingdom and was identified in a random transfer-DNA (T-DNA) insertion screen reporting the derepression of a silenced transgene (17). The *mom1* mutant shows a loss of transcriptional gene silencing at loci located predominantly in the pericentromeric regions of the chromosomes (18). Interestingly, these transcriptional gene-silencing defects occur without major changes in DNA methylation or histone marks (17–21). RNA Polymerase IV and V (*PolIV* and

PolV), which are key components of the RdDM pathway, were identified as enhancers of the *mom1* phenotype (18). To date, the extent to which *MOM1* is implicated in RdDM as well as its molecular mechanism of action remain poorly understood. Because *MOM1* shows partial sequence similarities to chromodomain–helicase–DNA binding proteins, it has been proposed that *MOM1* is involved in heterochromatin compaction (17, 22). However, the *mom1* mutant does not show any heterochromatin decondensation (20, 23).

Recently, members of the *Arabidopsis* microorchidia (*AtMORC*) ATPase family have also been shown to be involved in transposon repression and gene silencing (24–26). The *MORC1* gene was originally described in mice, where it was found to be essential for male primordial germ cell development (27, 28). The *Arabidopsis* genome contains seven *MORC* homologs, which were termed *AtMORC1* [NP_568000; AT4G36290; *Compromized Recognition of Turnip Crinkle Virus 1* (*CRT1*)], *AtMORC2* [NP_195351; AT4G36280; *CRT1-Homolog 1* (*CRH1*)], *AtMORC3* (NP_195350; AT4G36270; *CRH2*), *AtMORC4* (NP_199891; AT5G50780; *CRH4*), *AtMORC5* (NP_196817; AT5G13130; *CRH5*), *AtMORC6* [NP_173344; AT1G19100;

Significance

Members of the *Arabidopsis* microorchidia (*AtMORC*) ATPase family are involved in gene silencing and heterochromatin condensation without altering genome-wide DNA methylation patterns. Here, we examine the functional relationship between several family members and show that *AtMORC6* interacts in two mutually exclusive protein complexes with *AtMORC1* and its closest homologue, *AtMORC2*. Consistently, RNA sequencing of high-order mutants indicates that *AtMORC1* and *AtMORC2* act redundantly in gene silencing. We also examine the genetic interactions between *AtMORC6* and the transcriptional repressor *Morpheus Molecule 1* (*MOM1*). We observe a synergistic transcriptional regulation in the *mom1/atmorc6* double mutant, indicating that these epigenetic regulators act mainly in different silencing pathways, both independently of DNA methylation.

Author contributions: G.M., S.B., D.H., W.A.P., L.Y., A.P., and S.E.J. designed research; G.M., S.B., D.H., W.A.P., L.Y., and A.P. performed research; C.J.H., H.S., A.A.V., and J.A.W. analyzed data; and G.M., S.B., and S.E.J. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE54677).

¹Present address: Department of Biology, ETH Zurich, 8092 Zurich, Switzerland.

²G.M. and S.B. contributed equally to this work.

³Present address: Department of Neurobiology, Harvard Medical School, Boston, MA 02115.

⁴To whom correspondence should be addressed. E-mail: jacobsen@ucla.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1406611111/-DCSupplemental.

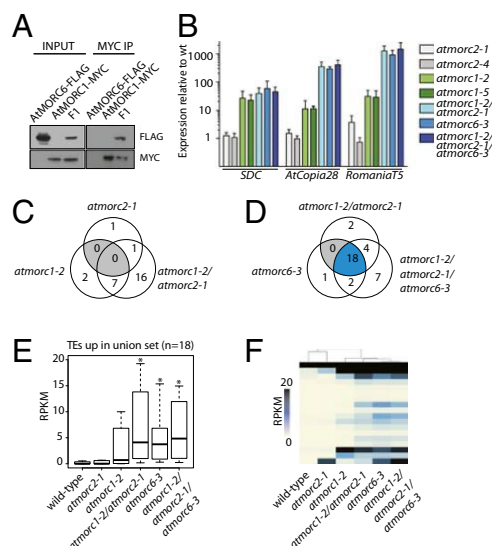


Fig. 1. Redundancy of *AtMORC1* and *AtMORC2* in transposon silencing. (A) *AtMORC1* physically interacts with *AtMORC6*. *AtMORC6*-FLAG was coimmunoprecipitated with *AtMORC1*-MYC in *F1* plants expressing both epitope-tagged proteins. Epitope-tagged proteins were detected by Western blotting. (B) RT-PCR assessing endogenous expression of *SDC*, *AtCopia28*, and *RomaniaT5*. Three biological replicates were performed for each tested genotype. Two individual alleles were used for *atmorc1* and *atmorc2*. (C and D) Venn diagrams of overlap between TEs up-regulated (fourfold increase; FDR, 0.05; Fisher's exact test) in each genotype. Gray regions represent categories with no TEs counted. Blue shading represents the union set of TEs up-regulated in *atmorc* mutants. (E) Boxplot and (F) heatmap of average reads per kilo base per million (RPKM) values between two biological replicates for TEs in a union set for different genotypes. An asterisk indicates a significant increase relative to wild-type samples ($P < 1e-3$, Mann-Whitney *U* test).

FLAG-*AtMORC2* was principally present in the elution fractions around 200–300 KDa, corresponding to similar elution fractions as *AtMORC6*-FLAG (Fig. S2). In summary, our biochemical analyses indicate that *AtMORC6* physically interacts with *AtMORC1* and *AtMORC2* in the form of two mutually exclusive heteromers.

AtMORC6 was shown to interact *in vitro* with DMS3 when both proteins were coexpressed in *Escherichia coli*, providing a physical link to the RdDM pathway (25). DMS3 is a structural maintenance of chromosomes hinge domain-containing protein that lacks an ATPase domain (36). Based on the stimulation of *AtMORC6* ATPase activity by *in vitro* interaction with DMS3, it was proposed that *AtMORC6* and DMS3 cooperate to promote transcriptional repression. DMS3 has also been shown to interact with additional components of the DRD1-DMS3-RDM1 (DDR) complex including Defective in RNA-Directed DNA Methylation 1 (DRD1) or RDM1 as well as with the largest subunit of PolV (37). Furthermore, genome-wide association of PolV to chromatin and thus the production of PolV-dependent transcripts and subsequent DNA methylation are dependent on all members of the DDR complex (37, 38). However, we did not detect DMS3 or other components of the DDR complex in our IP-MS experiments. Also, previous IP-MS experiments using FLAG-tagged DRD1 and DMS3 proteins as bait did not immunoprecipitate *AtMORC6* (37). Nevertheless, we cannot rule out that the interactions between components of the DDR complex and *AtMORC6* are weak or ephemeral and could not be detected under our IP conditions.

A recent study found that *AtMORC6* was immunoprecipitated in flowers in very small amounts with SUVH9, an SRA- (SET [suppressor of variegation 3–9 [Su(var)3–9], enhancer of zeste [E(z)], and trithorax (Trx)] and RING [really interesting new gene] associated)- and SET-domain-containing protein (35). SUVH9 and its closest homolog, SUVH2, were shown to bind methylated DNA and recruit PolV to chromatin through an interaction with

the DDR complex (11, 35, 39). Yeast two-hybrid assays further indicated that the interactions between *AtMORC* proteins and SUVH proteins were direct (35). These data, together with the slight changes observed in DNA methylation of certain RdDM target loci (24, 25, 40), suggest that *AtMORC* proteins modulate RdDM through interactions with the DDR complex and SUVH proteins. Nevertheless, the mild changes of small RNAs and DNA methylation genome-wide in *atmorc* mutants (26) suggest that *AtMORCs* are unlikely to be canonical RdDM factors. It is also plausible that *AtMORCs* contribute to processing of target loci transcripts, thus leading to posttranslational silencing. Future experiments are needed to clarify the precise function in gene silencing and degree of involvement of *AtMORCs* in the RdDM pathway.

***AtMORC2* Acts Redundantly with *AtMORC1* to Achieve Gene Silencing.**

To further study the role of *AtMORC2* in gene silencing and its functional relationship with *AtMORC1* and *AtMORC6*, we generated high-order mutants and performed transcriptional profiling analyses. Real-time PCR (RT-PCR) from RNA extracted from leaf tissue indicated that *SDC* was derepressed in *atmorc1* but not *atmorc2* (Fig. 1B), consistent with the fact that *AtMORC2* was not identified in the genetic screens that identified *AtMORC1* and *AtMORC6* (24–26, 31). RT-PCR also showed an increased derepression of two transposons, *AtCopia28* and *RomaniaT5*, in the *atmorc1/atmorc2* double mutant compared with *atmorc1* and *atmorc2* single mutants (Fig. 1B), indicating that *AtMORC1* and *AtMORC2* act redundantly in transposon silencing. Further genome-wide characterization of the transcriptome by RNA-seq indicated that only two transposons were significantly up-regulated in *atmorc2* compared with wild type [using a very stringent cutoff of fold change ≥ 4 ; false discovery rate (FDR) < 0.05], whereas nine TEs were up-regulated in *atmorc1* (Fig. 1C). Transcriptional derepression of protein-coding genes was also more pronounced in *atmorc1* compared with *atmorc2* (Fig. 2A). Publicly available microarray data indicate that expression of *AtMORC1* is higher than *AtMORC2* in most tissues and developmental stages (Fig. S3A), providing a plausible explanation for the stronger silencing defects observed in *atmorc1* compared with *atmorc2*. Interestingly, combined deletion of *AtMORC1* and *AtMORC2* led to significantly higher transcription of TEs and protein-coding genes compared with both single mutants (Fig. 1C, E, and F and Fig. 2A, C, and D), confirming that *AtMORC1* and *AtMORC2* are functionally redundant. In addition, the overexpression of FLAG-*AtMORC2* succeeded in complementing transcriptional derepression in the *atmorc1/atmorc2* double mutant (Fig. S1D).

The observed redundancy between *AtMORC1* and *AtMORC2* and their physical interaction with *AtMORC6* in two mutually exclusive heteromers predict that a loss of *AtMORC6* should be phenotypically comparable to the combined loss of *AtMORC1* and *AtMORC2*. To test this hypothesis, we compared the transcriptomes of *atmorc1/atmorc2* with the *atmorc6* single mutant. RNA-seq revealed a high overlap of transcriptional derepression between *atmorc1/atmorc2* and *atmorc6* (Fig. 1D–F and Fig. 2B–D), supporting the notion that *AtMORC6* function is epistatic to both *AtMORC1* and *AtMORC2* combined. Derepressed transposons were not restricted to a specific family in any of the mutant backgrounds analyzed (Fig. S3B). Finally, the observed transcriptional derepression did not significantly increase in a triple mutant lacking *AtMORC1*, *AtMORC2*, and *AtMORC6* (Fig. 1D–F and Fig. 2B–D). These results are consistent with the model that *AtMORC6* interacts exclusively with either *AtMORC1* or *AtMORC2* to achieve gene silencing and that *AtMORC1* is functionally redundant with *AtMORC2*.

It appeared that up-regulated genes were preferentially localized in H3K9me2-enriched heterochromatin (12) even though they are protein-coding (Fig. 2E). This is in agreement with the previous observations that *AtMORC1* and *AtMORC6* are mainly involved in silencing and compaction of heterochromatin (26). Gene ontology term analysis using AmiGO (41) of all up-regulated protein-coding genes indicated enrichments ($P < 6e-4$) in response

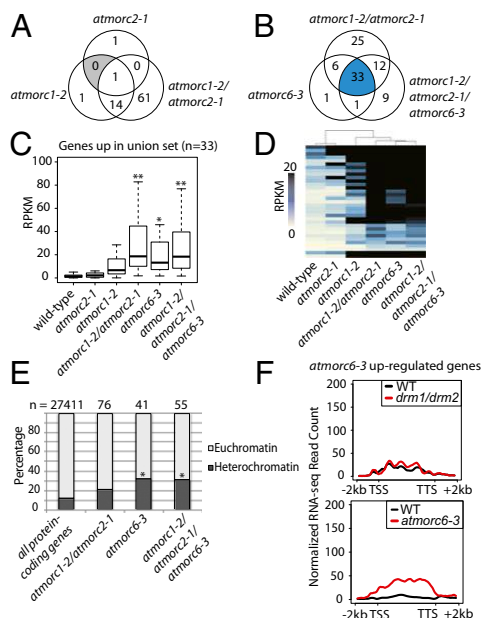


Fig. 2. Redundancy of *AtMORC1* and *AtMORC2* in gene silencing. (A and B) Venn diagrams showing relationships between sets of protein-coding genes called up-regulated (fourfold increase in expression; FDR < 0.05) for different genotypes. Gray regions represent categories with no gene counted. Blue shading represents the union set of genes up-regulated in *atmorc* mutants. (C) Boxplot and (D) heatmap of average RPKM values for protein-coding genes in a union set for different genotypes. An asterisk indicates a significant increase relative to wild-type samples ($P < 1e-8$, Mann-Whitney U test). Two asterisks represent a significant increase relative to wild-type samples and the *atmorc1* single mutant ($P < 1e-2$, Mann-Whitney U test). (E) Overrepresentation in H3K9me2-enriched heterochromatin of protein-coding genes significantly up-regulated in *atmorc1-2/atmorc2-1*, *atmorc6-3*, or *atmorc1-2/atmorc2-1/atmorc6-3* mutants. An asterisk indicates a significant increase relative to all protein-coding genes ($P < 1e-3$, Fisher's exact test). (F) Metagene analysis of RNA-seq reads over protein-coding genes called up-regulated in *atmorc1-2/atmorc2-1*, *atmorc6-3*, or *atmorc1-2/atmorc2-1/atmorc6-3* mutants. Reads are derived from previously published RNA-seq libraries for two replicates of the *drm1/drm2* double mutant and the corresponding wild type (WT).

to chitin and in response to organonitrogen compounds in *atmorc1/atmorc2* and in *atmorc1/atmorc2/atmorc6*. It is interesting to note that chitin has been recognized as a general elicitor of plant defense responses (42), which is in agreement with the reported implication of *AtMORC1* in plant immunity (31). To assess if protein-coding genes up-regulated in *atmorc6* were also targets of the RdDM machinery, we looked at their expression in a mutant lacking the methyltransferases *DRM1* and *DRM2* that is thus defective in RdDM (4). These were not significantly up-regulated in *drm1/drm2* (Fig. 2F), indicating that *AtMORCs* are unlikely to be canonical RdDM factors.

Our combined genetics and RNA-seq data show that the simultaneous absence of *AtMORC1* and *AtMORC2* in *atmorc1/atmorc2* cannot be functionally compensated by the presence of *AtMORC6* alone (Figs. 1 and 2). Also, the loss of *AtMORC6* in *atmorc6* cannot be compensated by the presence of *AtMORC1* and *AtMORC2* (Figs. 1 and 2). Furthermore, the *atmorc1/atmorc2/atmorc6* triple mutant does not have a stronger phenotype than the *atmorc1/atmorc2* double mutant (Fig. 1B and D–F and Fig. 2B–D). Together with the observation that *AtMORC1* and *AtMORC2* did not interact, these results lead to the conclusion that *AtMORCs* function as heteromers and not as homomers.

AtMORC6 and MOM1 Act Synergistically to Silence a Common Set of Transposons. *AtMORC1* and *AtMORC6* were identified in a forward genetic screen reporting the derepression of an *SDC::GFP*

transgene in wild type or in the *cmt3* mutant background (26). Further screening of ethyl methanesulfonate (EMS) mutagenized seeds followed by deep genome resequencing identified two new alleles of *AtMORC6* in the *cmt3* background. In the first line, *cmt3* 262, glycine 212 was mutated to glutamic acid, and in *cmt3* 379, a guanine (chr1:6599258) was mutated to adenine in the splice site before exon 14. Interestingly, we also identified three loss-of-function alleles of the *MOM1* gene in the same genetic screen. The EMS mutations in these new *mom1* alleles were a stop codon introduced at amino acid 603 (line 337 in a wild-type background), a stop codon introduced at amino acid 586 (*cmt3* 265), and a substitution of Leucine 656 to Phenylalanine (*cmt3* 113).

MOM1 is unique to the plant kingdom and has no homologs in the *Arabidopsis* genome. Previous studies showed that DNA methylation in *mom1* mutants was similar to the wild-type level (17–19, 21). This observation was recently confirmed by genome-wide bisulfite-sequencing (BS-seq) analyses (43). RNA-seq analyses showed that 52 TEs were significantly up-regulated in *mom1* using similarly stringent cutoffs as for *atmorc* mutants (Fig. 3A), and we found that the DNA methylation levels of these TEs also remained unchanged in *mom1* compared with wild type (Fig. 3D). Nineteen transposons were significantly derepressed in *atmorc6* in this experiment, and most of these were also derepressed in *mom1* (Fig. 3A). The numbers of TEs significantly up-regulated in *atmorc6* slightly vary between the two RNA-seq experiments performed (Figs. 1D and 3A) because both experiments were done independently. As shown previously, DNA methylation was not significantly changed in TEs up-regulated in *atmorc6* (26) (Fig. 3D). These data indicate that overall transcriptional derepression is higher in *mom1* compared with *atmorc6* and that *MOM1* and *AtMORC6* mediate the silencing of a subset of common targets as well as of a number of independent loci.

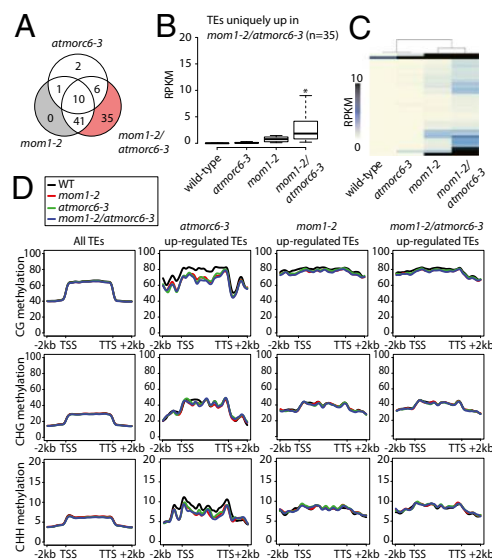


Fig. 3. Synergy of *AtMORC6* and *MOM1* in transposon silencing. (A) Venn diagram showing relationships between sets of TEs called up-regulated (fourfold increase in expression; FDR < 0.05) for different genotypes. Grayed regions highlight sets with no elements, and red shading highlights TEs uniquely called up-regulated in the higher order mutant. (B) Boxplot and (C) heatmap of average RPKM values between two biological replicates for TEs uniquely called up-regulated in the *mom1/atmorc6* mutant background for different genotypes. An asterisk indicates a significant increase relative to all other genotypes ($P < 1e-8$, Mann-Whitney U test). (D) Metagene analysis of DNA methylation levels across all *Arabidopsis* TEs for the *atmorc6-3*, *mom1-2*, *mom1-2/atmorc6-3*, and wild-type genotypes. Also shown are the methylation levels at TEs up-regulated in mutant genotypes.

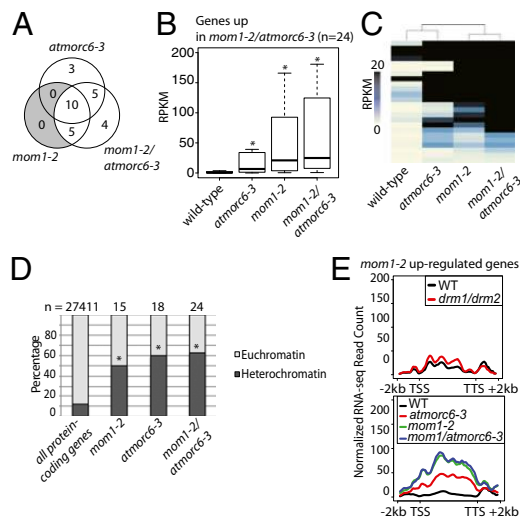


Fig. 4. Synergy of *AtMORC6* and *MOM1* in gene silencing. (A) Venn diagram showing relationships between sets of protein-coding genes called up-regulated (fourfold increase in expression; FDR < 0.05) for different genotypes. Grayed regions highlight sets with no elements. (B) Boxplot and (C) heatmap of average RPKM values for protein-coding genes uniquely called up-regulated in the *mom1/atmorc6* mutant background. An asterisk represents a significant increase relative to wild-type samples ($P < 1e-2$, Mann-Whitney U test). (D) Overrepresentation in H3K9me2-enriched heterochromatin of protein-coding genes significantly up-regulated in *atmorc6-3*, *mom1-2*, or *mom1-2/atmorc6-3* mutants. An asterisk indicates a significant increase relative to all protein-coding genes ($P < 1e-3$, Fisher's exact test). (E) Metagene analysis of RNA-seq reads over protein-coding genes called up-regulated in *atmorc6-3*, *mom1-2*, or *mom1-2/atmorc6-3* mutants. Reads are derived from previously published RNA-seq libraries for two replicates of the *drm1/drm2* double mutant and the corresponding wild type (WT).

To further understand the relationship between *MOM1*- and *AtMORC6*-mediated transcriptional silencing, we generated a double mutant lacking *MOM1* and *AtMORC6*. RNA-seq analyses in *mom1/atmorc6* showed a significant increase in derepression of TEs and to a smaller extent of protein-coding genes compared with both single mutants (Fig. 3 A–C and Fig. 4 A–C). RT-PCR analyses corroborated the synergistic derepression of *SDC* and *RomanistT5* (Fig. S44). Overexpressed TEs in all three genotypes profiled by RNA-seq are predominantly located in the pericentromeric heterochromatin and belong to diverse families, consistent with previous reports (18, 26) (Fig. S4 B and C). Genome-wide BS-seq analysis showed that DNA methylation was unchanged in TEs up-regulated in *mom1/atmorc6* (Fig. 3D). Similar to *AtMORC6* target loci, protein-coding genes significantly up-regulated in *mom1* were preferentially located in heterochromatin (Fig. 4D). Furthermore, transcription of these was not affected in the *drm1/drm2* mutant, suggesting a limited role of *MOM1* in RdDM (Fig. 4E). Altogether, these results indicate that *AtMORC6* and *MOM1* act synergistically to silence a largely common set of heterochromatic DNA elements through two independent pathways.

Conclusion

In this study, we combined biochemistry, genetics, and genomics to understand further the mode of action of the recently discovered *Arabidopsis* MORC homologs. We found that *AtMORC6*-mediated transcriptional silencing requires the formation of mutually exclusive heteromers with *AtMORC1* and its closest homolog, *AtMORC2*. Further biochemical studies involving domain deletions or point mutations should uncover the molecular mechanisms of the *AtMORC* proteins and the implication of heteromerization for ATPase activity. It is interesting to note the similarities between *AtMORCs* and the structural maintenance of chromosome proteins cohesin and condensin (44). These three

protein families are ATPases that function in vivo as heteromers and modulate chromatin superstructure to regulate proper expression and maintenance of genomic integrity.

Genetic and RNA-seq analyses showed that *AtMORC6* acts synergistically with the putative chromatin remodeler *MOM1* to silence a common set of heterochromatin-localized loci. The synergistic effect observed in the *mom1/atmorc6* double mutant suggests that *AtMORC6* and *MOM1* act in two convergent pathways that are both required for the proper silencing of pericentromeric heterochromatin. It has been previously shown that *AtMORC6* and *AtMORC1* accumulate in the nucleus as discrete nuclear bodies that localize in the vicinity of the heterochromatic chromocenters (26). It will be interesting to determine in the future whether *MOM1* accumulates in a similar fashion in the nucleus to form distinct nuclear bodies. The identification of *MOM1* interactors will also be crucial to understanding its mode of action.

Materials and Methods

Plant Material and Growing Conditions. Wild-type and all mutant lines are from the ecotype Columbia and were grown under continuous light. Plant lines used include *atmorc1-2* (SAIL_893_B06; crt1-2), *atmorc1-4* (SAIL_1239_C08), *atmorc1-5* (SAIL_131_H11; crt1-5), *atmorc2-1* (SALK_072774C; crh1-1), *atmorc2-4* (SALK_021267C; crh1-4), *atmorc6-3* (GABI_599B06), *cmt3-11* (SALK_148381), and *mom1-2* (SAIL_610_G01). EMS mutagenized *atmorc6-1* and *cmt3/mor1-3* lines and complementing *AtMORC1-MYC* and *AtMORC6-MYC* lines are described in ref. 26. T-DNA insertions were confirmed by PCR-based genotyping. Primer sequences are described in Table S1.

Cloning of pAtMORC1::AtMORC1-FLAG, pAtMORC2::FLAG-AtMORC2, and pAtMORC16::AtMORC6-FLAG. Cloning was done according to ref. 26. Briefly, *AtMORC1* and *AtMORC6* genomic regions were PCR amplified and the FLAG epitope was added to the C terminus of *AtMORC1* and *AtMORC6* and at the N terminus of *AtMORC2*. The amplified region includes a ~1 Kb promoter sequence upstream of the respective transcriptional start site.

IP and MS Analysis. Ten grams of 2-wk-old seedling tissue of each epitope-tagged line were ground in liquid nitrogen and resuspended in 45 mL ice-cold IP buffer [50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM MgCl₂, 0.1% Nonidet P-40, 10% (vol/vol) glycerol, 1× Protease Inhibitor Mixture (Roche)] and centrifuged for 10 min at 4 °C at 16,000 × *g*. We added 200 μL M2 magnetic FLAG-beads (SIGMA, M8823) to the supernatants and incubated it for 60 min rotating at 4 °C. M2 magnetic FLAG-beads were washed five times in ice-cold IP buffer for 5 min rotating at 4 °C, and immunoprecipitated proteins were eluted three times with 100 μL 3×-FLAG peptides (SIGMA, F4799) for 15 min at 25 °C. The eluted protein complexes were precipitated by trichloroacetic acid and subjected to MS analyses as previously described (14).

Co-IP and Immunoblotting. We ground 1.5 g of 2-wk-old seedling tissue of each epitope-tagged line in liquid nitrogen, resuspended it in 12 mL ice-cold IP buffer [50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM MgCl₂, 0.1% Nonidet P-40, 10% (vol/vol) glycerol, 1× Protease Inhibitor Mixture (Roche)], and centrifuged it for 10 min at 4 °C at 16,000 × *g*. We added 100 μL M2 magnetic FLAG-beads (SIGMA, M8823) or 150 μL MYC-conjugated agarose beads (COVANCE, AFC-150P-1000) to the supernatants and incubated it for 60 min rotating at 4 °C. Beads were washed five times in ice-cold IP buffer for 5 min rotating at 4 °C, and immunoprecipitated proteins were eluted in 1× Laemmli buffer for 15 min at 80 °C.

Western blots were performed as previously described (26) with GFP-specific antibody (Invitrogen, AA1122), HRP-coupled FLAG-specific antibody (SIGMA, A8592), and MYC-specific antibody (Pierce, MA1-980).

Gel Filtration. Gel filtration experiments were performed according to ref. 37. Briefly, 0.5 g of 2-wk-old seedling tissue of each epitope-tagged line were ground in liquid nitrogen and resuspended in 1 mL of ice-cold IP buffer [50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Nonidet P-40, 10% (vol/vol) glycerol, 1× Protease Inhibitor Mixture (Roche)] and centrifuged for 10 min at 4 °C at 16,000 × *g*. The supernatants were centrifuged again for 10 min at 4 °C at 16,000 × *g*. The supernatants were then centrifuged through a 0.2 μm filter (Millipore), 500 μL were loaded onto a Superdex 200 10/300GL column (GE Healthcare, 17-5175-01) column, and 250 μL fractions were collected. We ran 20 μL of every collected fraction on a 4–12% SDS/PAGE. Before use, the column was equilibrated and calibrated with gel filtration standards (Biorad, 151-1901).

RNA Extraction. We froze 100 mg of 20-d-old leaf tissue in liquid nitrogen. The frozen leaves were then added to a mortar containing liquid nitrogen. Immediately after the liquid nitrogen boiled off, the leaf tissue was crushed to powder using a pestle. We immediately added 1.2 mL of TRIzol Reagent (Life Technologies 15596) to the cold powder, and then it was pulverized further until a clear, dark brown solution was visible. The solution was transferred to a chilled Eppendorf tube, and 400 μ L of chloroform was added. The tube was vortexed for 5 s at maximum power, then spun in a centrifuge at $16,000 \times g$ (4 $^{\circ}$ C) for 10 min to separate the aqueous and organic phases. We collected 700 μ L of the aqueous (top) phase. To precipitate the RNA, 700 μ L of isopropanol was added to the aqueous material, the solution was vortexed for 5 s at maximum power, and then it was centrifuged for 10 min at $16,000 \times g$ (4 $^{\circ}$ C). The supernatant was removed, and 500 μ L of room temperature 80% (vol/vol) ethanol was added to the pellet, which was then spun for 5 min at $16,000 \times g$ (4 $^{\circ}$ C). The supernatant was removed and the pellet was air-dried for 5 to 10 min. The pelleted RNA was resuspended in 100 μ L water and then purified using the Qiagen RNeasy Mini (Qiagen 74104) "RNA Cleanup Protocol" according to manufacturer's instructions. RNA was quantified using Nanodrop.

RT-PCR. We treated 1 μ g of input RNA with DNase I (Life Technologies, 18068) according to the manufacturer's protocol. Of the 11 μ L final reaction volume, 3 μ L was set aside as a negative control for RT-PCR, whereas 8 μ L was converted to cDNA using SuperScript III (Life Technologies 18080). We used 5% of cDNA for each RT-PCR. RT-PCR was performed using IQ SYBR Green Supermix (Bio-Rad 170-8880), with 375 nM final primer concentration using a Stratagene Mx3005p instrument. Amplification conditions were as follows: 95 $^{\circ}$ C 10:00; 40

cycles, 95 $^{\circ}$ C, 30 s, 55 $^{\circ}$ C 1:00, 72 $^{\circ}$ C 1:00; melting curve. At least two technical replicates were performed per biological replicate, and three biological replicates were used in all experiments. Relative abundance of transcripts was calculated using the difference of squares method. Primer sequences are described in Table S1.

BS-Seq, RNA-Seq, and Accession Codes. BS-seq was done according to ref. 26. RNA-seq libraries were generated using 2 μ g of input RNA using TruSeq RNA Sample Preparation Kit v2 (Illumina RS-122-2001) according to the manufacturer's protocols. Sequencing data were deposited into Gene Expression Omnibus under accession no. GSE54677.

ACKNOWLEDGMENTS. We are grateful to Daniel F. Klessig (Boyce Thompson Institute for Plant Research, Ithaca, NY) for the *atmorc1/atmorc2* double knockout line. We are grateful to Michael F. Carey for his advice and for providing access to equipment. High-throughput sequencing was performed in the University of California, Los Angeles (UCLA) Broad Stem Cell Research Center BioSequencing Core Facility, and we are especially grateful to Mahnaz Akhavan for her support. We thank Bhumi Parekh, Colin Shew, Beatrice Sun, Lillian Tao, and Vanessa Trieu for technical assistance. This work was supported by National Institutes of Health (NIH) Grant GM60398 (to S.E.J.). C.J.H. is supported by the Damon Runyon postdoctoral fellowship, W.A.P. is supported by the Jane Coffin Childs Memorial Fund for Medical Research, H.S. is supported by a UCLA Dissertation Year Fellowship, L.Y. is supported by Ruth L. Kirschstein National Research Service Award GM007185, S.B. is supported by a postdoctoral fellowship of the Swiss National Science Foundation, and J.A.W. is supported by NIH Grant GM089778. S.E.J. is an investigator of the Howard Hughes Medical Institute.

- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11(3):204–220.
- Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416(6880):556–560.
- Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL (1996) Demethylation-induced developmental pleiotropy in Arabidopsis. *Science* 273(5275):654–657.
- Stroud H, et al. (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* 21(1):64–72.
- Zemach A, et al. (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153(1):193–205.
- Cao X, Jacobsen SE (2002) Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* 12(13):1138–1144.
- Chan SW, et al. (2004) RNA silencing genes control de novo DNA methylation. *Science* 303(5662):1336.
- Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* 308(5718):118–120.
- Law JA, et al. (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* 498(7454):385–389.
- Onodera Y, et al. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120(5):613–622.
- Johnson LM, et al. (2014) SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* 507(7490):124–128.
- Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. *PLoS ONE* 3(9):e3156.
- Cokus SJ, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452(7184):215–219.
- Du J, et al. (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* 151(1):167–180.
- Johnson L, Cao X, Jacobsen S (2002) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* 12(16):1360–1367.
- Tariq M, et al. (2003) Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. *Proc Natl Acad Sci USA* 100(15):8823–8827.
- Amedeo P, Habu Y, Afsar K, Mittelsten Scheid O, Paszkowski J (2000) Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. *Nature* 405(6783):203–206.
- Yokthongwattana C, et al. (2010) MOM1 and Pol-IV/IV interactions regulate the intensity and specificity of transcriptional gene silencing. *EMBO Journal* 29:340–351.
- Habu Y, et al. (2006) Epigenetic regulation of transcription in intermediate heterochromatin. *EMBO Rep* 7(12):1279–1284.
- Probst AV, Franz PF, Paszkowski J, Mittelsten Scheid O (2003) Two means of transcriptional reactivation within heterochromatin. *Plant J* 33(4):743–749.
- Vaillant I, Schubert I, Tourmente S, Mathieu O (2006) MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. *EMBO Rep* 7(12):1273–1278.
- Caikovski M, et al. (2008) Divergent evolution of CHD3 proteins resulted in MOM1 refining epigenetic control in vascular plants. *PLoS Genet* 4(8):e1000165.
- Mittelsten Scheid O, Probst AV, Afsar K, Paszkowski J (2002) Two regulatory levels of transcriptional gene silencing in Arabidopsis. *Proc Natl Acad Sci USA* 99(21):13659–13662.
- Brabbs TR, et al. (2013) The stochastic silencing phenotype of Arabidopsis morc6 mutants reveals a role in efficient RNA-directed DNA methylation. *Plant J* 75(5):836–846.
- Lorković ZJ, Naumann U, Matzke AJ, Matzke M (2012) Involvement of a GHKL ATPase in RNA-directed DNA methylation in Arabidopsis thaliana. *Curr Biol* 22(10):933–938.
- Moissiard G, et al. (2012) MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* 336(6087):1448–1451.
- Inoue N, et al. (1999) New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum Mol Genet* 8(7):1201–1207.
- Watson ML, et al. (1998) Identification of morc (microorchidia), a mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse. *Proc Natl Acad Sci USA* 95(24):14361–14366.
- Langen G, et al. (2014) The compromised recognition of turnip crinkle virus1 subfamily of microorchidia ATPases regulates disease resistance in barley to biotrophic and necrotrophic pathogens. *Plant Physiol* 164(2):866–878.
- Kang HG, Klessig DF (2008) The involvement of the Arabidopsis CRT1 ATPase family in disease resistance protein-mediated signaling. *Plant Signal Behav* 3(9):689–690.
- Kang HG, Kuhl JC, Kachroo P, Klessig DF (2008) CRT1, an Arabidopsis ATPase that interacts with diverse resistance proteins and modulates disease resistance to turnip crinkle virus. *Cell Host Microbe* 3(1):48–57.
- Kang HG, et al. (2010) Endosome-associated CRT1 functions early in resistance gene-mediated defense signaling in Arabidopsis and tobacco. *Plant Cell* 22(3):918–936.
- Kang HG, et al. (2012) CRT1 is a nuclear-translocated MORC endonuclease that participates in multiple levels of plant immunity. *Nat Commun* 3:1297.
- Dutta R, Inouye M (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci* 25(1):24–28.
- Liu ZW, et al. (2014) The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLoS Genet* 10(1):e1003948.
- Kanno T, et al. (2008) A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat Genet* 40(5):670–675.
- Law JA, et al. (2010) A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in Arabidopsis. *Curr Biol* 20(10):951–956.
- Zhong X, et al. (2012) DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat Struct Mol Biol* 19(9):870–875.
- Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE (2008) SRA-domain proteins required for DRM2-mediated de novo DNA methylation. *PLoS Genet* 4(11):e1000280.
- Lorković ZJ (2012) MORC proteins and epigenetic regulation. *Plant Signal Behav* 7(12):1561–1565.
- Carson S, et al.; AmiGO Hub; Web Presence Working Group (2009) AmiGO: Online access to ontology and annotation data. *Bioinformatics* 25(2):288–289.
- Boller T (1995) Chemoperception of microbial signals in plant-cells. *Annu Rev Plant Physiol* 46:189–214.
- Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152(1–2):352–364.
- Wood AJ, Severson AF, Meyer BJ (2010) Condensin and cohesin complexity: The expanding repertoire of functions. *Nat Rev Genet* 11(6):391–404.

Supporting Information

Moissiard et al. 10.1073/pnas.1406611111

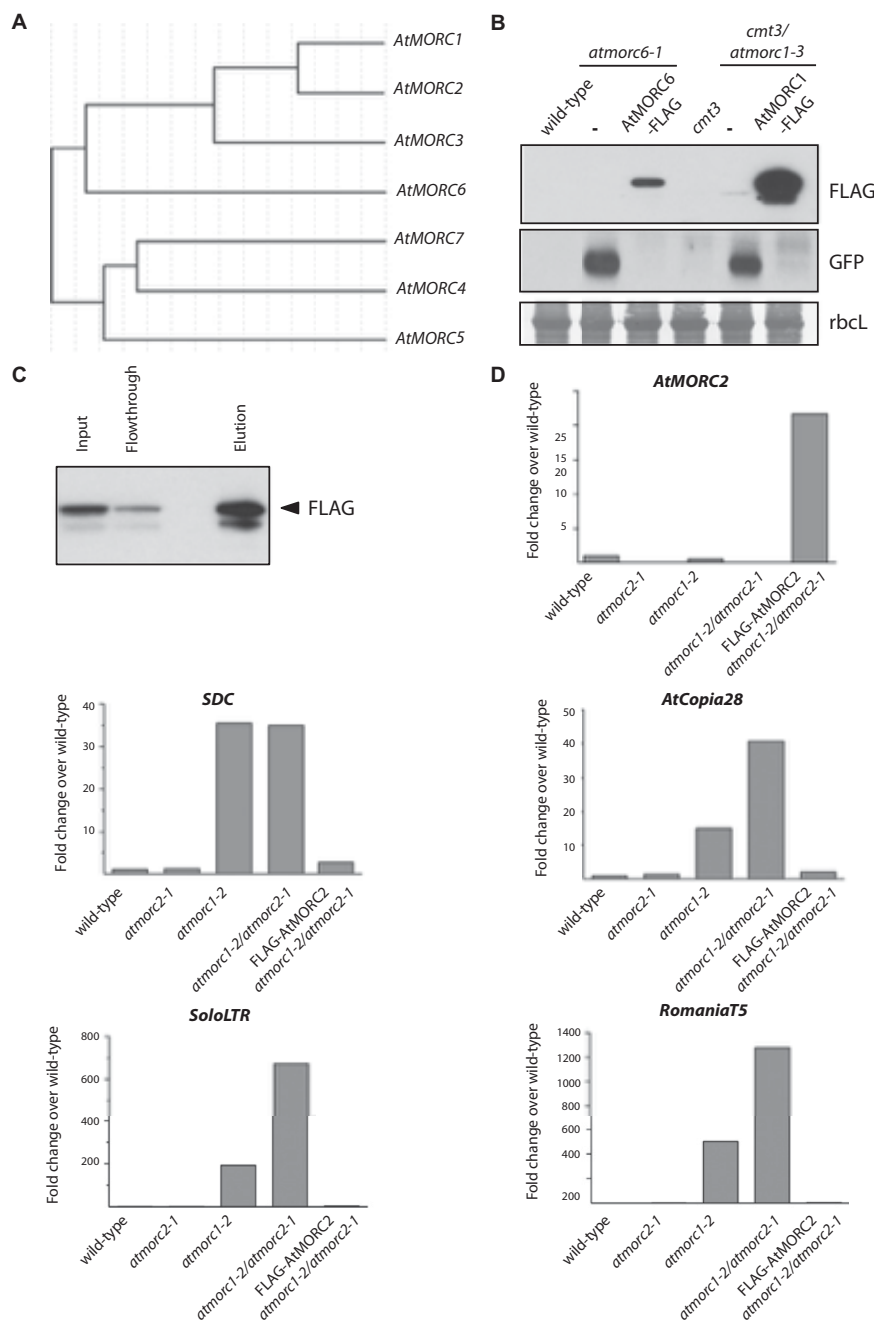
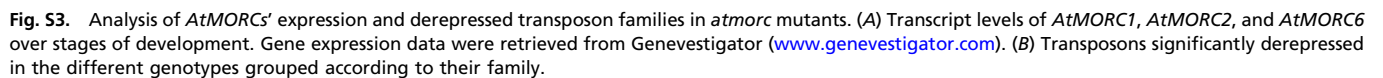
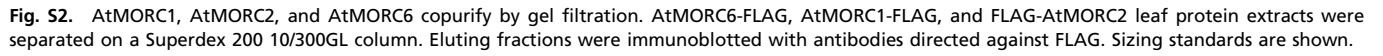


Fig. S1. Phylogenetic analysis and epitope-tagging of *Arabidopsis* Microrchidia (AtMORC)1, AtMORC2, and AtMORC6. (A) Phylogenetic analysis of the AtMORC gene family in *Arabidopsis thaliana*. The alignment was made with GeneBee using the default parameters. (B) AtMORC1 and AtMORC6 transgenic lines. FLAG epitope-tagged AtMORC1 and AtMORC6 were expressed under their respective endogenous promoter in their respective mutant background. Protein expression and complementation of the *SDC::GFP* silencing defects were probed by Western blotting. The large subunit of rubisco (*rbcl*) was used as the loading control. (C) Transgenic AtMORC2 line. FLAG epitope-tagged AtMORC2 was expressed under its respective endogenous promoter in the *atmorc1/atmorc2* double mutant background. FLAG-AtMORC2 is enriched in the elution fraction after immunoprecipitation. (D) Complementation of transcriptional derepression by expression of FLAG-AtMORC2 in *atmorc1/atmorc2*. RT-PCR shows increased levels of FLAG-AtMORC2 transcripts compared with wild-type and *atmorc1-2* despite being expressed under its respective endogenous promoter. Derepression of *suppressor of drm2 cmt3* (*SDC::GFP*), *AtCopia28*, *Solo long terminal repeat* (LTR), and *RomaniaT5* is suppressed by overexpression of FLAG-AtMORC2.



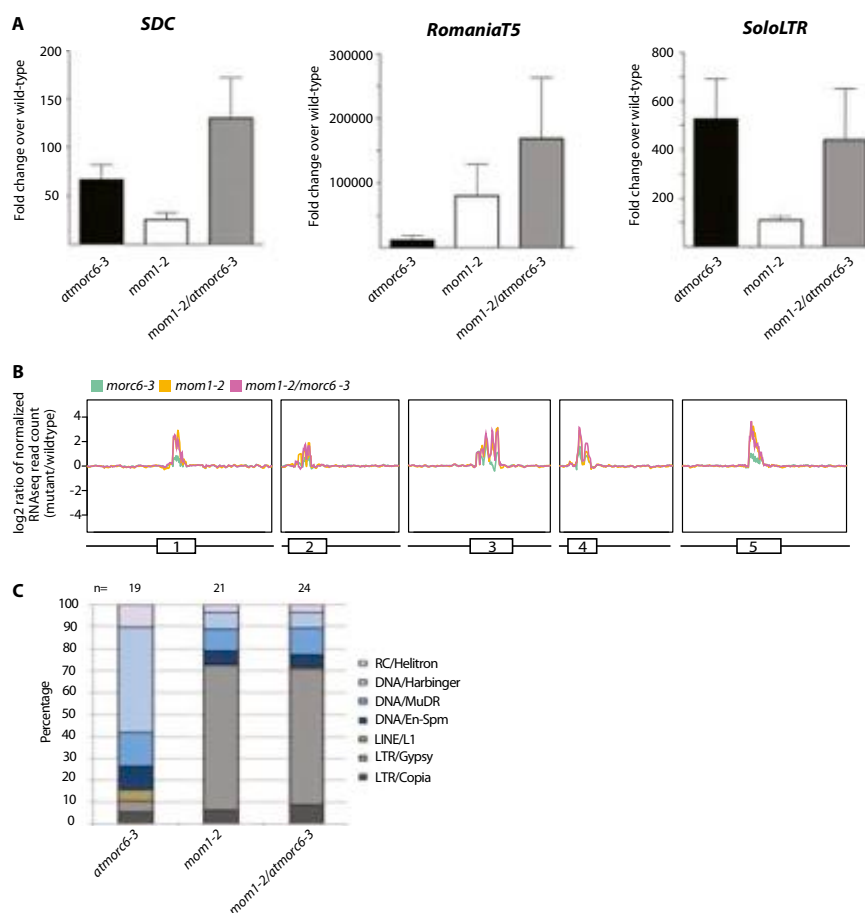


Fig. S4. Synergy of *AtMORC6* and *Morpheus Molecule 1 (MOM1)* in transposon silencing. (A) RT-PCR assessing endogenous expression of *SDC*, *RomaniaT5*, and *SoloLTR*. Two biological replicates were performed for each tested genotype. (B) Chromosomal view of percent change in read abundance between higher order mutants and other genotypes. (C) Transposons significantly derepressed in the different genotypes grouped according to their family.

Table S1. Primers used in this study

Genotyping			
Genotype	Insertion line	Left border primer	Right border primer
<i>atmorc1-2</i>	SAIL_893_B06	TTGCAGTTTGGAAACAAAATC	TGAGTTTTGACGACGATGATG
<i>atmorc1-4</i>	SAIL_1239_C08	CGTATCTCAGCCGCTAACTTG	AAGCAGCTGCAGTGGATTATG
<i>atmorc1-5</i>	SAIL_131_H11	CGTATCTCAGCCGCTAACTTG	AAGCAGCTGCAGTGGATTATG
<i>atmorc2-1</i>	SALK_072774C	CTACTCAGAGCGTTGGCATTCT	GTTGTAGCTGTATGGGGCTTG
<i>atmorc2-4</i>	SALK_021267C	TTTCGTCACTATTGCTTTTCC	GGTTGACTCTTCCACTGCTTG
<i>atmorc6-3</i>	GABI_599B06	GGAAAGCTGGAAGCTATAATGATG	GATGACATCTGCCCAAGTCTC
<i>cmt3-11</i>	SALK_148381	TAACGGAAGGATGCCAGATT	CAAGAAATGGGCTGTTGACAT
<i>mom1-2</i>	SAIL_610_G01	ACAATGCAGGAGCAAACACTC	GGAAAGGAGATACTTACCCGG
LB3	SAIL T-DNA	TAGCATCTGAATTTCTAAACCAATCTCGATACAC	—
LBA1	SALK T-DNA	TGGTTCACGTAGTGGGCCATCG	—
LB O8409	GABI-KAT	ATATTGACCATCATACTCATTGC	—
RT-PCR			
Locus		Left border primer	Right border primer
Actin 7		TCGTGGTGGTGAGTTTGTTAC	CAGCATCATCACAAAGCATCC
SDC		AATGTAAGTTGTAAACCAATTTGAACGTGACC	CAGGCATCCGTAGAACTCATGAGC
AtCopia28		AGTCCTTTTGTTGCTGAACA	CCGGATGTAGCAACATTCACT
RomaniaT5		GTATCCTTTGGCCCGGTATT	GCCTCTTCGAAATGCCATAA
SoloLTR		AACTAACGTCATTACATACACATCTTG	AATTAGGATCTTGTTGCCAGCTA

ARTICLE

Received 12 Sep 2014 | Accepted 7 Nov 2014 | Published 12 Dec 2014

DOI: 10.1038/ncomms6795

OPEN

MORC1 represses transposable elements in the mouse male germline

William A. Pastor^{1,*}, Hume Stroud^{1,*†}, Kevin Nee¹, Wanlu Liu¹, Dubravka Pezic², Sergei Manakov², Serena A. Lee¹, Guillaume Moissiard¹, Natasha Zamudio³, Déborah Bourc'his³, Alexei A. Aravin², Amander T. Clark^{1,4} & Steven E. Jacobsen^{1,4,5}

The Microrchidia (*Morc*) family of GHKL ATPases are present in a wide variety of prokaryotic and eukaryotic organisms but are of largely unknown function. Genetic screens in *Arabidopsis thaliana* have identified *Morc* genes as important repressors of transposons and other DNA-methylated and silent genes. MORC1-deficient mice were previously found to display male-specific germ cell loss and infertility. Here we show that MORC1 is responsible for transposon repression in the male germline in a pattern that is similar to that observed for germ cells deficient for the DNA methyltransferase homologue DNMT3L. *Morc1* mutants show highly localized defects in the establishment of DNA methylation at specific classes of transposons, and this is associated with failed transposon silencing at these sites. Our results identify MORC1 as an important new regulator of the epigenetic landscape of male germ cells during the period of global *de novo* methylation.

¹Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, 4028 Terasaki Life Sciences Building, 610 Charles E. Young Drive East, Los Angeles, California 90095, USA. ²Division of Biology and Biochemical Engineering, California Institute of Technology, 1200 E California Boulevard, Pasadena, California 91125, USA. ³Unité de génétique et biologie du développement, Institut Curie, CNRS UMR3215, INSERM U934, Paris 75005, France. ⁴Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, 615 Charles E. Young Drive South, Los Angeles, California 90095, USA. ⁵Howard Hughes Medical Institute, University of California Los Angeles, 675 Charles E. Young Drive South, Los Angeles, California 90095, USA. * These authors contributed equally to this work. † Present address: Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115 USA. Correspondence and requests for materials should be addressed to A.T.C. (email: clarka@ucla.edu) or to S.E.J. (email: jacobsen@ucla.edu).

Two *Morc* genes in *A. thaliana*, *AtMorc1* and *AtMorc6*, were identified in forward genetic screens for novel transcriptional repressors^{1,2}. *AtMORC1* and *AtMORC6* are required for silencing of a variety of transposons and are essential for higher-order chromatin compaction. The single *Morc* gene in *Caenorhabditis elegans* was also shown to be required for silencing of a repetitive transgene locus¹. The founding member of the *Morc* gene family is mammalian *Morc1*. *MORC1* is highly expressed in the blastocyst and male germline but is not expressed in most differentiated cells³. Mice deficient for *MORC1* are normal, except that homozygous mutant males are infertile with small testicles (hence the name microrchidia)^{4,5}. Male germ cells in the *Morc1* mutant do not undergo successful chromosomal pairing during the zygotene stage of meiosis and instead undergo apoptosis, with no germ cells surviving to complete prophase I.

During germ cell development, most DNA methylation is lost between E8.5 and E13.5. Then, between E13.5 and birth (~E19), the genome undergoes global *de novo* methylation^{6–8}. Failure to establish DNA methylation at this time causes transposon upregulation and meiotic failure. Indeed, the meiotic block in the *Morc1* mutant is similar to that observed for mice that have defects in DNA methylation and transposon repression, including mice deficient for DNA methyltransferases^{9–11} or the pre-meiotic Piwi-interacting RNA (piRNA) pathway^{12,13}. Therefore, we hypothesized that *MORC1* might be a critical factor for transposon silencing and DNA methylation in the mouse germline. Here we demonstrate that *MORC1*-deficient male germ cells undergo transposon derepression starting in late embryogenesis and continuing through the onset of meiosis. We also demonstrate that this phenotype is associated with failed locus-specific *de novo* methylation targeted specifically towards late-methylating transposon sequences.

Results

***MORC1* represses transposons in the male germline.** To further characterize *MORC1* we used a previously described FVB/N *Morc1* mutant (*Morc1^{tg}*) mouse strain in which a tyrosinase gene was integrated into the *Morc1* locus⁵. Transgene insertion resulted in loss of exons 2–4, eliminating a large region of the GHKL ATPase domain including binding residues predicted to be critical for catalysis and ATP binding¹⁴ (Fig. 1a). Consistent with previous reports, we found that *Morc1^{tg/tg}* mice have a spermatogenesis defect with a complete absence of post-meiotic spermatids and spermatozoa (Supplementary Fig. 1a–c).

Quantitative reverse transcription-PCR (qRT-PCR) from wild type (WT), embryonic whole testis indicates that *Morc1* messenger RNA becomes detectable at E14.5 and peaks at E16.5 (Supplementary Fig. 2a), which intriguingly is a period of rapid transposon methylation in the male germline. We generated an antibody against the coiled-coil domain of mouse *MORC1* (Supplementary Fig. 2b) and found that *MORC1* was localized to the nucleus of male germ cells at E16.5 (Fig. 1b). Conversely, *MORC1* protein was undetectable in *Morc1^{tg/tg}* mutant germ cells (Fig. 1b). To test for transposon derepression in *Morc1^{tg/tg}* germ cells, we performed immunostaining for LINE1 ORF1p and intracisternal particle A (IAP) at postnatal day 14.5 (P14.5). We found that both of these transposon classes were ectopically expressed during early postnatal development, with particular enrichment of LINE1 ORF1p in the meiotic cells towards the centre of the tubule (Fig. 1c,d). LINE1 ORF1p was also derepressed at E16.5 and E18.5, showing that transposon derepression in *Morc1^{tg/tg}* arises well before the apparent meiotic defect (Supplementary Fig. 2c).

To identify which genes and specific transposons are targets of *MORC1*-mediated repression, we performed mRNA-Sequencing

(RNA-Seq) of whole testes from *Morc1^{tg/tg}* and *Morc1^{tg/+}* heterozygous controls at P10.5 and P14.5. Given that germ cells make up only a small percentage of the testis during the embryonic early and postnatal period, we also performed RNA-Seq on ribosomal RNA-depleted total RNA from sorted germ cells at E16.5, E18.5, P2.5 and P10.5. To purify germ cells, we crossed the *Morc1^{tg}* allele into the *Oct4-IRES-eGfp* reporter strain, which exhibits a distinct eGFP⁺ population from e9.5 to P2.5 (Supplementary Fig. 3a)^{15,16}. To isolate germ cells at P10.5, we gated the side scatter (SSC) low (SSC^{Lo}), epithelial cell adhesion molecule (EPCAM) high (EPCAM^{Hi}) and major histocompatibility complex negative (MHC1[−]) population of testicular cells (Supplementary Fig. 3b). A list of sequencing samples and the experiments to which they correspond is included in Supplementary Data 1. Consistent with RT-PCR and published data^{5,17}, *Morc1* mRNA was confirmed to show high expression at E16.5 and lower expression at later time points in the sorted germ cells (Fig. 2a).

The RNA-Seq data showed a broad transposon derepression defect in *Morc1^{tg/tg}* starting at E16.5 (Fig. 2c–e and Supplementary Data 2). Quantitative RT-PCR analysis of various transposable element classes gave similar results as RNA-Seq data (Supplementary Fig. 4a). In addition, RNA-Seq on sorted *Morc1^{tg/tg}* and control germ cells at P10.5 resembled the pattern of transposon derepression observed in whole testes (Supplementary Fig. 4a). Heterozygous *Morc1^{tg/+}* mice showed no marked increase in transposon expression relative to WT *Morc1^{+/+}* mice, confirming their validity as littermate controls (Supplementary Fig. 4b).

Different transposon classes showed different patterns of derepression in *Morc1^{tg/tg}*. Some classes (RLTR4, RLTR6, MuRRS and Etn) were upregulated during embryogenesis but silenced even in the knockout at later time points (Fig. 2b). Other transposons (MMERVK10C, GLN and some IAP species) were most highly upregulated at postnatal time points (Fig. 2c,d). LINEs were upregulated both in late embryogenesis and again at P14.5 after the onset of meiosis (Fig. 2e), which was confirmed by immunofluorescence (Fig. 1d and Supplementary Fig. 2c). These fluctuations in transposon upregulation may reflect differences in the inherent transcriptional programmes of certain transposon classes, as well as varied effectiveness of other, partially redundant transposon repression pathways at different times. In the aggregate, however, these results indicate that *MORC1* constitutes a new participant in transposon repression in the mammalian germline, acting on many different elements. Notably, *MORC1* silences many transposon classes well after it is downregulated, consistent with it acting through an epigenetic mark such as DNA methylation.

piRNA biogenesis occurs normally in the *Morc1* mutant.

During this period in germline development (E14.5 to birth), germ cells undergo mitotic arrest and global nuclear reprogramming that most notably involves genome-wide *de novo* DNA methylation mediated by the DNMT3A/DNMT3L complex^{11,18}. The pre-meiotic piRNA pathway, involving the nuclear PIWI protein MIWI2, is also active during this period in promoting transposon silencing. To evaluate whether *MORC1* acts on the same transposon classes as DNMT3L or MIWI2, we performed RNA-Seq on whole testes from *Dnmt3l^{−/−}* (ref. 10) and *Miwi2^{−/−}* (ref. 12) mice and their respective controls at P10.5, and compared this with the *Morc1^{tg/tg}* P10.5 whole testis data set. *Morc1^{tg/tg}* and *Dnmt3l^{−/−}* exhibited derepression of an overlapping set of transposons, primarily long terminal repeat (LTR) retrotransposons, while the *Miwi2^{−/−}* mutant testes had a milder phenotype and showed derepression of specific LINE elements and the IAP-Ey class of retrotransposons, which was not affected in *Morc1^{tg/tg}* or *Dnmt3l^{−/−}* mutant testes (Fig. 2f–h).

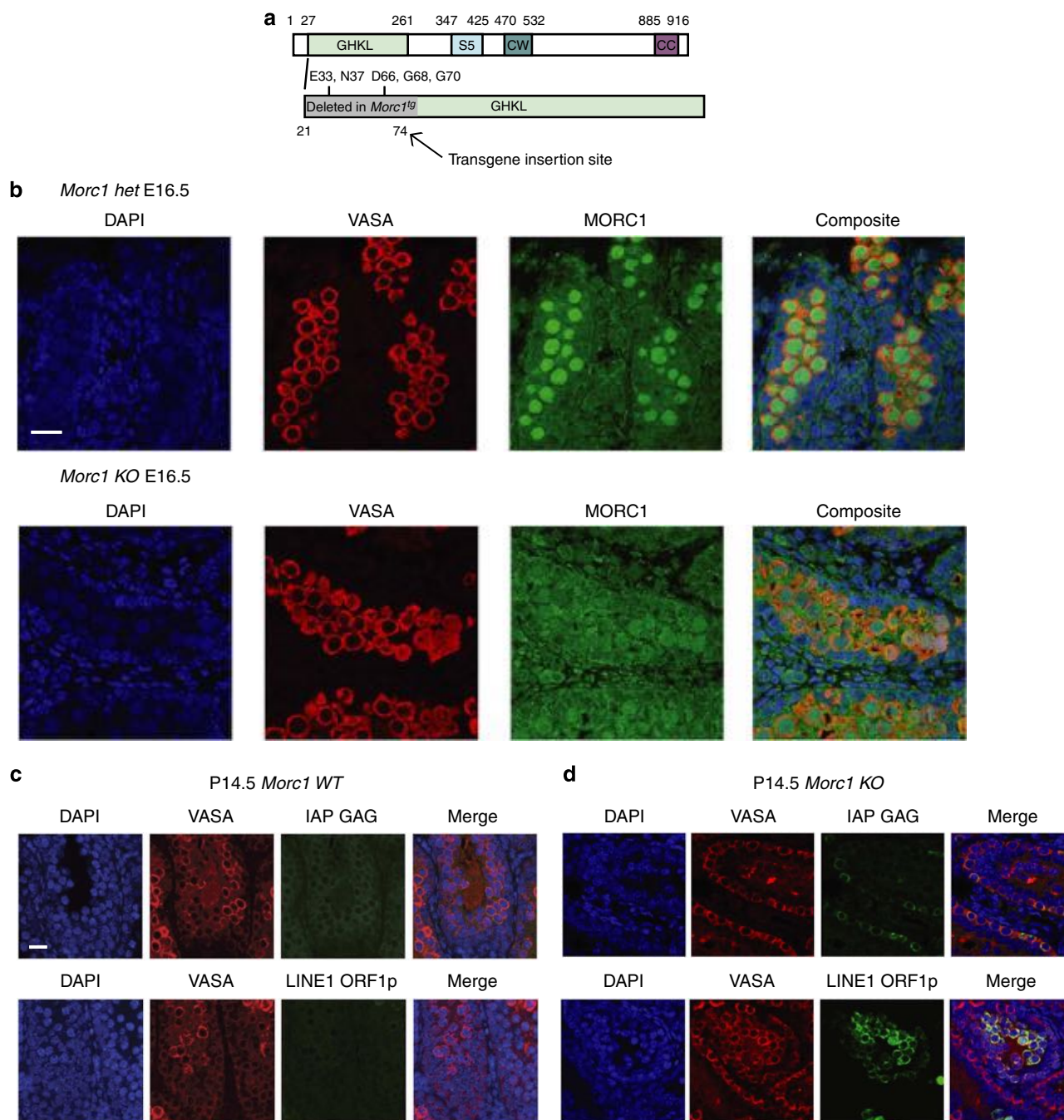


Figure 1 | MORC1 is a nuclear protein essential for transposon repression. (a) Domain structure of *Morc1* gene and disruption in *Morc1^{tg}* allele. Deleted residues predicted to be critical for catalytic activity or ATP binding are denoted. (b) Detection of MORC1 by immunofluorescence (IF) in E16.5 testes. MORC1 is present as a germ cell-specific nuclear protein in the *Morc1^{tg/+}* (het) control but is absent from the *Morc1^{tg/tg}* (knockout (KO)). Aberrant expression of IAP GAG (c) and LINE ORF1p (d) in *Morc1^{tg/tg}* as detected by IF at P14.5. Note that IAP is overexpressed in most germ cells, whereas LINE is primarily present in the more differentiated cells deeper into the tubule. Scale bars, 20 μ m (b–d).

The lack of overlap between *Morc1* and *Miwi2* predicts that piRNAs would be unperturbed in *Morc1^{tg/tg}* mice. To test this, we performed small RNA sequencing of the testis at E16.5 to examine piRNA production. Our data revealed that the ratio of piRNA/microRNA and the generation of antisense piRNAs were unaltered in *Morc1^{tg/tg}* (Table 1), indicating that the piRNA pathway remains largely intact in *Morc1^{tg/tg}* testis at E16.5, and that transposon derepression in *Morc1^{tg/tg}* is most likely to be independent of the piRNA pathway. In fact, at P10.5 we observed an increase in the fraction of piRNAs derived from LTR

retrotransposons, especially of the IAP family, in the MORC1-deficient testis (Supplementary Fig. 5a–c), similar to that observed in *Dnmt3l^{-/-}*¹⁹. These LTR transposon-derived piRNAs corresponded to primary sense piRNAs (Supplementary Fig. 5d,e), suggesting that they are probably more abundant simply because the underlying mRNA species are derepressed in *Morc1^{tg/tg}*, and some fraction are converted to piRNAs (Supplementary Fig. 5f). Hence, our results indicate that MORC1 acts in a transposon-silencing pathway independent of piRNA production.

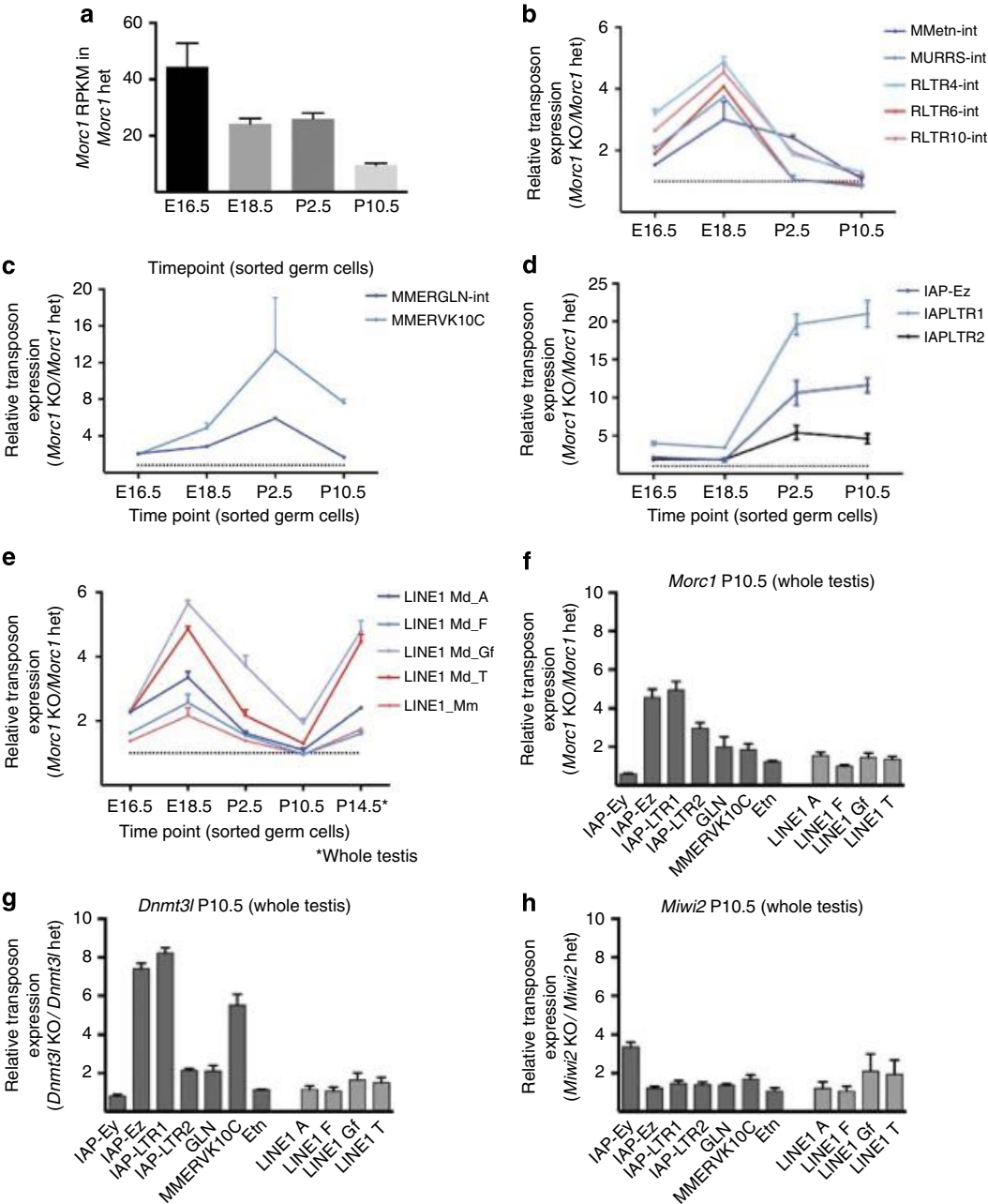


Figure 2 | *Morc1*^{tg/tg} shows transposon upregulation resembling *Dnmt3l*^{-/-}. (a) Expression of *Morc1* mRNA over development in *Morc1*^{tg/+}, as measured by RNA-seq. (b–e) Overexpression of transposon species over the course of mammalian development, represented as a ratio of expression in the *Morc1*^{tg/tg} and the *Morc1*^{tg/+} control. Some LTR transposons show upregulation selectively in late embryogenesis (b), while others are overexpressed postnatally (c,d), and LINE elements are overexpressed both during late embryogenesis and again at the onset of meiosis (e). Overexpression of transposons in MORC1- (f), DNMT3L- (g) and MIWI2- (h) deficient whole testis. For a–e, the dotted line indicates a fold change of one. For a–h. Two to four replicates per genotype were analysed; all data are RNA-seq from sorted germ cells or whole testis as indicated. Mean + s.e. plotted.

Table 1 piRNA abundance and characteristics in E16.5 <i>Morc1</i> ^{tg/+} and <i>Morc1</i> ^{tg/tg} testes.		
	<i>Morc1</i> het	<i>Morc1</i> KO
Putative piRNA/miRNA	0.55	0.59
Sense/antisense	1.34	1.34
Primary/secondary	4.43	3.91

miRNA, micro RNA; piRNA, Piwi-interacting RNA; smRNA, small RNA. Ratios of putative piRNA/miRNA, sense piRNA/antisense piRNA and primary piRNA/secondary piRNA populations are indicated for smRNA obtained from pooled E16.5 *Morc1*^{tg/+} and *Morc1*^{tg/tg} testes. No substantial defect in piRNA biogenesis is observed in MORC1-deficient testis.

Hypomethylation of transposable elements in *Morc1*^{tg/tg}. Because of the resemblance between transposons derepressed in *Morc1*^{tg/tg} and *Dnmt3l*^{-/-} testes (Fig. 2f,g), we sought to examine whether *Morc1* might affect global DNA methylation levels. To address this, we performed whole genome bisulfite sequencing at E16.5, P2.5 and P10.5 on sorted *Morc1*^{tg/tg} and control germ cells isolated as above. At E16.5, the germline is undergoing *de novo* DNA methylation and by P2.5 *de novo* methylation is largely complete. Between roughly P2.5 and P10.5, germline cells re-enter the cell cycle and either initiate the first wave of spermatogenesis to generate meiotic cells or localize to the basement membrane and generate the long-term

self-renewing spermatogonial stem cell population^{20,21}. In contrast to *Dnmt3l*^{-/-} mutant germ cells that show a dramatic global reduction in DNA methylation²², we found no change in global levels of methylation at any time point in *Morc1*^{tg/tg}-sorted germ cells (Fig. 3a). Thus, despite the similar morphological phenotypes and transposon expression defects of *Morc1*^{tg/tg} and *Dnmt3l*^{-/-} mice, MORC1 does not act by controlling *de novo* or maintenance methylation at a genome-wide level.

In mammals, DNA methylation is very dynamic and promoter DNA methylation frequently correlates with gene repression. To determine whether there may be localized defects in DNA methylation in *Morc1*^{tg/tg}, and whether these are associated with derepressed transposons identified by RNA-Seq, we calculated statistically significant differentially methylated regions (DMRs) in the *Morc1*^{tg/tg} germ cells relative to the *Morc1*^{tg/+} control. At E16.5, we found very few DMRs (Fig. 3b). However, at P2.5 we identified 6,309 hypomethylated regions (Supplementary Data 3) but only 145 hypermethylated regions (Supplementary Data 4), indicating that *Morc1*^{tg/tg} germ cells have locus and stage-specific DNA methylation defects (Fig. 3b). In addition, the overwhelming majority of regions identified as hypomethylated at P2.5 remain hypomethylated at P10.5 (Fig. 3c) and only a few regions lost methylation between P2.5 and P10.5 (Fig. 3d).

The hypomethylated DMRs in *Morc1*^{tg/tg} germ cells were highly enriched for LINE and LTR transposons rather than protein-coding genes compared with control regions (see Supplementary Methods), consistent with the transposon expression defects observed in *Morc1*^{tg/tg} germ cells (Fig. 4a). Indeed,

93.9% of hypomethylated DMRs contained an LTR or LINE, compared with 40.6% of control DMRs. The hypomethylated DMRs were strongly concentrated in the categories of transposons that showed evidence of derepression (Fig. 4b–d) during some stage of development before meiosis.

A partial exception to this trend were IAP elements. Hypomethylated DMRs were strongly enriched for IAP elements and corresponding LTRs (Supplementary Fig. 6a), but there was a poor correspondence between the extent to which a subcategory of IAPs was upregulated and the frequency of overlap with DMRs (Fig. 2d and Supplementary Fig. 6b). This is probably because certain highly similar repetitive elements such as LTR1 give very few uniquely mapping reads and are therefore missing from the data set. To overcome this, we also mapped the BS-seq data to RepBase consensus sequences for relevant transposons. We confirmed hypomethylation of the upregulated IAPLTR1 class (Supplementary Fig. 6c). Mapping to repeat consensus sequences also confirmed hypomethylation of LINE and LTR classes, which frequently overlap with DMRs (Fig. 4c–e).

Only 20 protein-coding genes contained an annotated transcription start site (TSS) within 1 kb of a hypomethylated DMR (Supplementary Data 5) and only 3 contained a TSS within a DMR. Interestingly, all three of these genes (*Nebulin*, *Tmc2* and *Cdkl4*) contain an RLTR10A transposable element immediately upstream of the TSS and all three genes showed a statistically significant increase in expression (Supplementary Fig. 7 and Supplementary Data 5). Thus, at a very few loci, MORC1 regulates genic expression, probably as a byproduct of its transposon repression activity in the local neighbourhood.

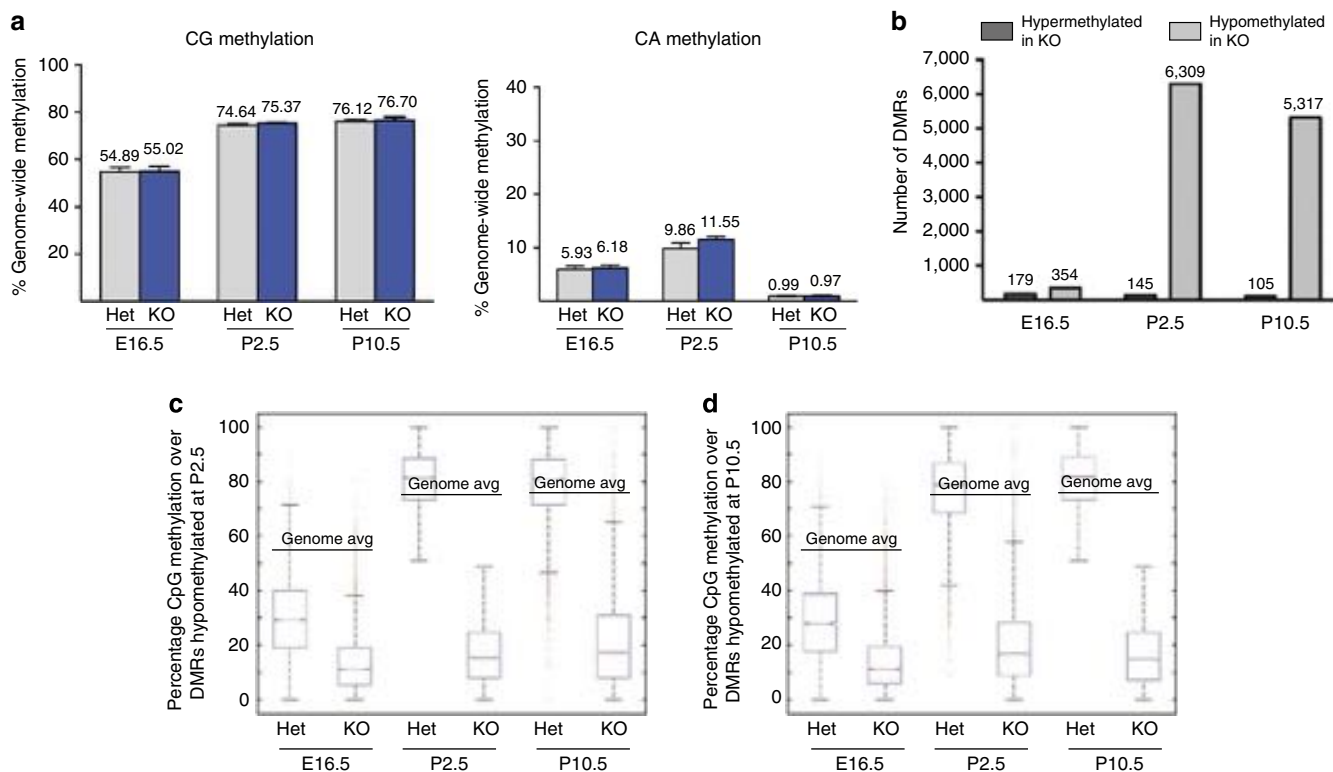


Figure 3 | MORC1 regulates DNA methylation in a stage- and locus-specific manner. (a) Total CG and CA methylation levels in *Morc1*^{tg/+} and *Morc1*^{tg/tg} sorted germ cells. Three to four replicates per genotype and time point were analysed; mean and s.e. are indicated. (b) Number of hypermethylated and hypomethylated DMRs, calculated using pooled BS-seq data at each time point. (c) Boxplots of DNA methylation levels at E16.5, P2.5 and P10.5 are shown for the set of regions that were computed as hypomethylated at P2.5 in *Morc1*^{tg/tg} germ cells. (d) Box plots of levels of DNA methylation levels in germ cells at E16.5, P2.5 and P10.5 are shown for the set of regions that were computed as hypomethylated at P10.5 in *Morc1*^{tg/tg}. For c,d, each DMR constitutes one point in each box plot. Red lines, median; edges of boxes, 75th (top) and 25th (bottom) percentiles; whiskers, minimum and maximum points within 1.5 × the interquartile range; red dots, outliers.

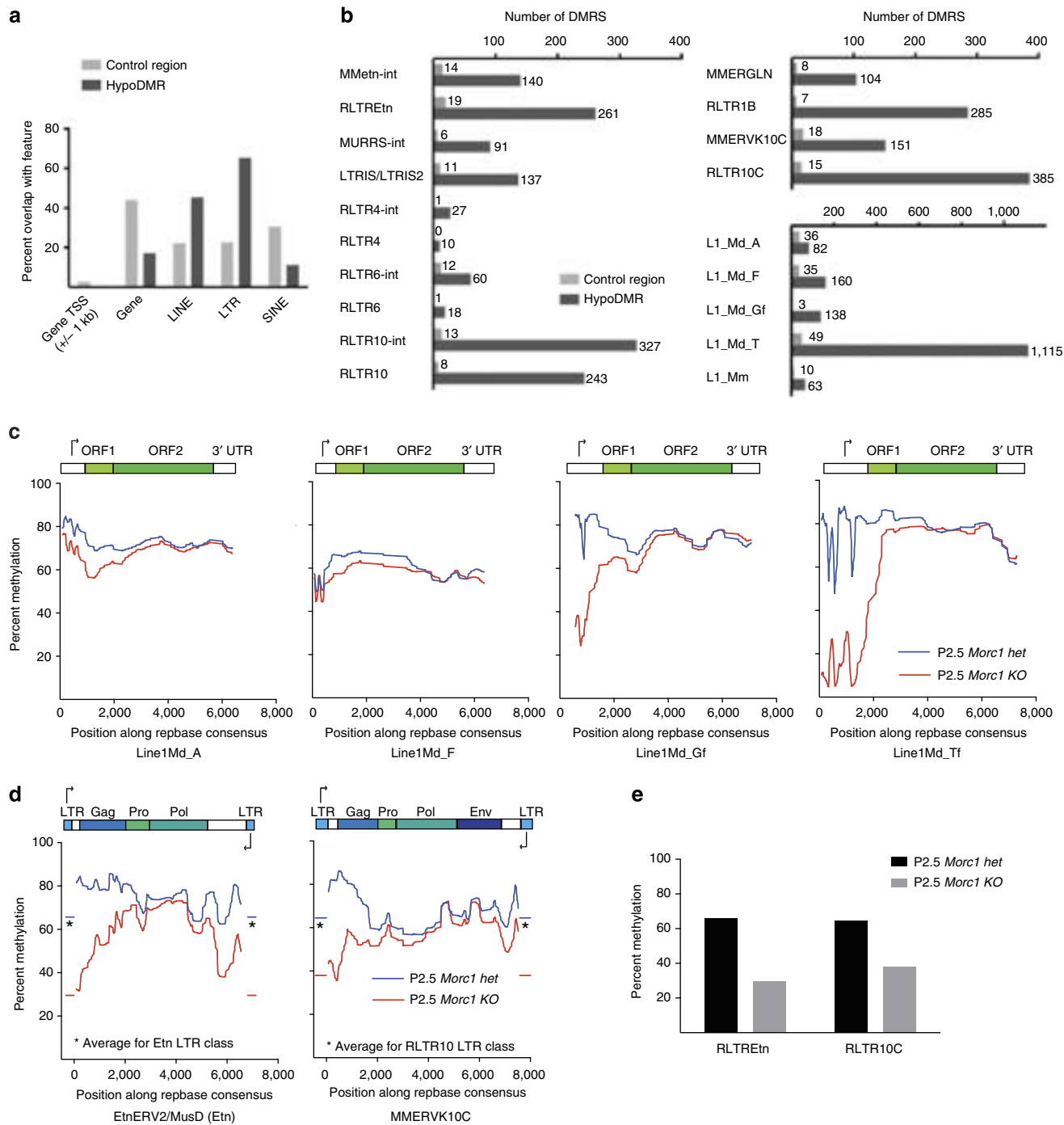


Figure 4 | Hypomethylated regions in *Morc1^{tg/tg}* correspond to upregulated transposon classes. (a) The overlap of hypomethylated DMRs and control regions (randomly selected regions whose methylation is unaffected by loss of MORC1) with genes and transposon classes is indicated. A DMR or control is counted as overlapping with a feature if there was at least one basepair overlap. (b) Overlap of hypomethylated DMRs with transposon classes upregulated in *Morc1^{tg/tg}*. Metaplot of methylation over LINE (c) and LTR (d) retrotransposons. BS-seq data were mapped to annotated RepBase consensus sequences for each transposon class and methylation is plotted at each CG in the annotated RepBase consensus sequence relative to the consensus sequence. The first 50 bases from each element, which often have low read coverage, are omitted. As it was usually not possible to determine the orientation of LTR-derived reads relative to LTR transposons, the average methylation level for the relevant LTR species is shown. (e) Global hypomethylation of LTR species corresponding to upregulated transposon classes. Again, BS-seq data were mapped to RepBase consensus sequence for these LTRs.

Considering MORC1's role as local modulator of DNA methylation, we examined changes in methylation in the three well-characterized paternally methylated imprinted loci²³. Methylation occurred normally at two of the three loci (*H19* and *Dlk1-Gtl2*), but the imprinting control region of *Rasgrf1* showed increased transcription and hypomethylation in the *Morc1* mutant (Supplementary Fig. 8). Interestingly, this is a transposon-rich area, which has previously been demonstrated as a target of the piRNA pathway²⁴ (see Discussion below).

Of the very few hypermethylated loci observed in the *Morc1* mutant, most were not conserved across time points and are probably a consequence of biological or statistical noise. However, 15 hypermethylated DMRs were reproducible between P2.5 and P10.5. Nine of these 15 were embedded in 2 transcripts upregulated in *Morc1^{tg/tg}*: 6 DMRs contained within the body of the *Cdk14* gene described above and 3 DMRs in an unannotated transcript probably originating from a hypomethylated IAPLTR1 element (Supplementary Fig. 9). These are probably examples of transcriptional run-through from a nearby promoter causing methylation of a locus, a phenomenon that has been described for some imprinted loci²⁵.

DMRs are sites of transposon transcriptional initiation. The highly localized affect of MORC1 on the germline epigenome suggests that MORC1 may function at the transcriptional start sites of transposons to facilitate their silencing and methylation. In support of this, we discovered that hypomethylation in *Morc1^{tg/tg}* mutant germline cells was concentrated at the 5'-ends of LINE elements coincident with the location of transcriptional initiation (Fig. 4c)²⁶. Furthermore, LTR transposons, which are typically flanked by LTRs that serve promoter and enhancer functions²⁷, showed hypomethylation on both ends in *Morc1^{tg/tg}* (Fig. 4d), and the LTRs themselves are heavily hypomethylated (Fig. 4d,e).

We also noted that hypomethylated DMRs in *Morc1^{tg/tg}* germ cells were late targets for *de novo* methylation during the course of epigenetic reprogramming, since in control *Morc1^{tg/+}* cells these genomic regions were also hypomethylated relative to the genome average at E16.5 (Fig. 3c,d and Supplementary Fig. 10). This suggests that these loci are somewhat resistant to *de novo* methylation. Consistent with this possibility, we also discovered that these *Morc1* affected genomic regions have increased H3K4me3 relative to control regions of the genome in WT cells E13.5 (Supplementary Fig. 11a)²⁸. To determine whether these loci are enriched in H3K4me3 during the dynamic *de novo* methylation of the germline genome, we performed chromatin immunoprecipitation (ChIP) analysis for H3K4me3 at E16.5 and confirmed that these regions still exhibited higher H3K4me3 (Fig. 5a). In contrast, other chromatin marks such as H2B10ac, K3K27ac and H3K27me3 showed no correlation with *Morc1^{tg/tg}* DMRs (Supplementary Fig. 11b). It is well established that H3K4 methylation antagonizes *de novo* DNA methylation by blocking DNMT3A/3L binding to histone H3 (ref. 29). Thus, the presence of H3K4me3 could potentially explain why these loci methylate with slow kinetics and require an additional factor (MORC1) for eventual silencing and methylation. It is also well established that H3K4me3 is a mark of transcriptional start sites, consistent with the idea that many DMRs are TSSs for transposons that are active in the embryonic germline.

At E16.5, we found that RNA transcripts were significantly elevated at *Morc1^{tg/tg}*-hypomethylated DMRs relative to the surrounding areas, even in control *Morc1^{tg/+}* germ cells (Fig. 5b). However, by P10.5, this RNA expression was severely repressed in control *Morc1^{tg/+}* germ cells with modest but increased expression in *Morc1^{tg/tg}* (Fig. 5b). This data is consistent with a model in which the hypomethylated DMRs correspond to TSSs that are normally methylated and suppressed during development.

To further confirm that these DMRs correspond to transcriptional start sites, we employed ATAC-seq, which can be used to identify areas of open chromatin that are a signature of promoter and enhancer sites³⁰. We confirmed that ATAC-seq can be accurately adopted for small sample sizes, that reads cluster near

transcriptional start sites in E16.5 germ cells, and that ATAC-seq read density at the TSS correlates with gene expression (Supplementary Fig. 12). We found that ATAC-seq peaks overlapped tightly with DMRs (Fig. 5c,d) and most DMRs showed ATAC-seq reads substantially elevated over background (Fig. 5e). In contrast, reads from the naked DNA control were not enriched over DMRs (Fig. 5d). At P10.5, a more limited subset of DMRs exhibited elevated ATAC-seq reads (Fig. 5e), consistent with the observation that transcription is retained only at some DMRs in postnatal germ cells (Supplementary Fig. 13a,b). Other sites lose ATAC peaks and transcription (Supplementary Fig. 13c), either because relevant transcription factors are absent or because other mechanisms of transposon silencing are effective. Importantly, at E16.5, where we observe expression from DMR regions in both *Morc1^{tg/+}* and *Morc1^{tg/tg}* cells, we also observed a high ATAC-seq signal in both *Morc1^{tg/+}* and *Morc1^{tg/tg}* cells (Fig. 5d). In contrast, at P10.5, where DMRs are silenced in heterozygotes but remain expressed in *Morc1^{tg/tg}*, we only observed high ATAC-seq signal in the *Morc1^{tg/tg}* cells (Fig. 5d). These results support the view that DMRs in *Morc1^{tg/tg}* cells correspond to promoters of transposons that fail to silence properly, leading to an inappropriately open chromatin state, and ectopic transposon expression, which is retained at P10.5, even after MORC1 expression has ceased.

Discussion

The results of this study identify MORC1 as a critical regulator of transposon repression in the male germline. MORC1 does not act as a global regulator of DNA methylation. Instead, MORC1 functions to facilitate DNA methylation of a variety of transposons in the germline with very little effect on the expression or methylation of protein-coding genes. The observation that *Morc* homologues are required for gene silencing in *Arabidopsis*, *C. elegans* and now mammals suggests that the *Morc* family of proteins constitute conserved epigenetic regulators that probably function in a wide variety of eukaryotic organisms and developmental contexts.

The *Morc1^{tg/tg}* phenotype of transposon derepression and a block in meiosis prophase I superficially resembles the phenotype observed in mice deficient for proteins involved in the pre-pachytene piRNA pathway, including *Mili*^{31,32}, *Miwi2* (ref. 12), *MitoPLD*³³, *Mov10L1* (refs 34,35), *Mae*^{36,37}, *Tdrkh*³⁸, *Tdrd9* (ref. 13) and *MVH*^{39,40}. What distinguishes *Morc1^{tg/tg}* from these characterized pre-pachytene piRNA mutants, however, is the apparently normal piRNA biogenesis in *Morc1^{tg/tg}* (Table 1). We do note similarities in the pattern of hypomethylation in *Morc1^{tg/tg}* and *Mili*^{-/-} mutant germ cells, including the *Rasgrf1* imprinting control region²⁴, as well as many of the same transposon families⁴¹. The dissimilarity in transposon repression observed in *Miwi2*^{-/-} and *Morc1^{tg/tg}* germ cells (Fig. 2f,h) suggests that MORC1's role in the nucleus is independent from the nuclear piRNA pathway mediated by MIWI2. It is possible that MORC1 participates downstream of the nuclear piRNA pathway during embryogenesis and has a separate, piRNA-independent silencing role during the postnatal stages. This could cause *Morc1^{tg/tg}* to have a broader transposon derepression phenotype than *Miwi2*^{-/-}. Alternatively, there may exist a MILI-dependent, MIWI2-independent mechanism for promoting methylation of target loci.

TEX19.1 has also been implicated in transposon repression in the male germline and has no known link to the piRNA pathway⁴². However, TEX19.1 is cytoplasmic^{42,43}, shows dysregulation only of MMERV10C elements^{42,44} and *Tex19.1*^{-/-} has an incomplete infertility defect^{42,43}. Thus, the *Morc1^{tg/tg}* and *Tex19.1*^{-/-} defects are fairly dissimilar and there is no evidence that they participate in the same pathway.

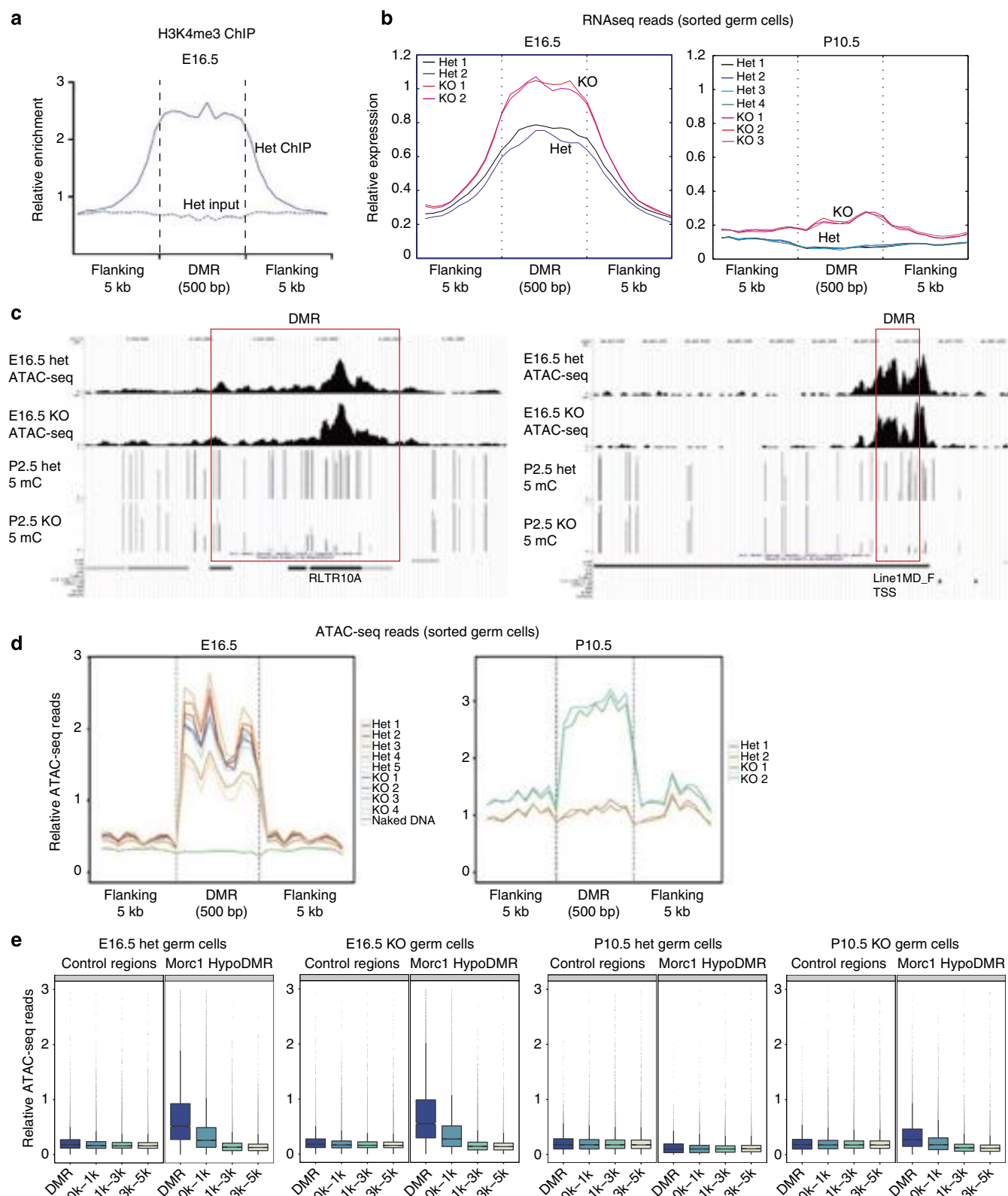


Figure 5 | Hypomethylated regions in *Morc1^{tg/tg}* correspond to TSS of transposons active in late embryogenesis. (a) H3K4me3 abundance at E16.5 is calculated within regions identified as hypomethylated DMRs (Hypo DMR) in P2.5 *Morc1^{tg/tg}* germ cells. (b) Average distributions of uniquely mapping E16.5 RNA-Seq reads (left) and P10.5 RNA-seq reads (right) from individual replicates are plotted over regions identified as hypomethylated at P2.5. (c) Pooled E16.5 ATAC-seq reads are plotted relative to methylation distribution at two loci with hypomethylated DMRs. Each CG is represented as by a bar, with the height of the bar indicating the frequency with which the CG is methylated. A dot at a position indicates no methylation. At least one read must map to the CG for a bar to appear. (d) ATAC-seq reads from individual replicates at E16.5 (left) and P10.5 (right) are plotted relative to DMRs. (e) ATAC-seq read abundance at DMRs and adjacent regions is represented as a boxplot, with each DMR constituting one point in the boxplot. For a–e, DMRs refer to regions hypomethylated in *Morc1^{tg/tg}* germ cells at P2.5.

Although we have revealed a critical role for MORC1 in transposon silencing, the actual mechanism by which MORC1 promotes DNA methylation in the male germline is unknown. Our study suggests at least three potential routes by which MORC1 represses transposons and facilitates DNA methylation. One possibility is that MORC1 directly silences transcription, perhaps using its ATPase activity to compact chromatin, thereby reducing H3K4 methylation levels at target sites. This silencing would allow for normal *de novo* methylation by DNMT3L. A second possibility is that MORC1 could recruit an H3K4 demethylase, which would similarly promote DNA methylation. Either mechanism agrees with our observation that MORC1-hypomethylated DMRs originate from loci with increased H3K4me3 at E13.5. A third non-mutually exclusive possibility is that MORC1 directly recruits the DNA methylation machinery to target loci, mediating methylation and silencing.

In conclusion, a robust genome defense system in the male germline is critical to safeguard genome integrity. We have identified a new participant that acts by facilitating DNA methylation of specific repetitive elements classes.

Methods

Mice. FVB/N-*Morc1*^{tg/+} (Tyr)^{1Az/J} mice (*Morc1*^{tg}) were recovered from cryopreservation at the Jackson Laboratory and maintained by intercrossing brothers and sisters in the FVB/N background. Male *Morc1*^{tg/tg} mice were viable healthy but infertile, whereas female *Morc1*^{tg/tg} mice were viable healthy and fertile. For PCR genotyping, the WT allele was detected as a 347-bp band with the following primers; forward: 5'-ATGCAACTTGAGGGGAAACA-3' and reverse: 5'-GCAGGAGTTATGCGATGTCA-3', and the mutant allele was detected as a 244-bp band with the following primers; forward: 5'-AGTTAGCCGTTATTAGTGGAGAGG-3' and reverse: 5'-AGAAAGCCTGCCTCAAAACA-3'. PCR conditions involved ten cycles of 94 °C, 65 °C and 68 °C, followed by 28 cycles of 94 °C, 50 °C and 72 °C. For sorting germ cells from E16.5–P2.5, *Morc1*^{tg/tg} females were crossed into the *Oct4*-IRES-*Gfp* mixed background. For embryonic staging, timed pregnancies were established and the day a vaginal plug was identified was called embryonic day 0.5 (E0.5). For postnatal time points, the day a litter was first observed was referred to a postnatal day 0.5 (P0.5).

All animal experiments were approved by The UCLA Institutional Animal Care and Use Committee, also known as the Chancellor's Animal Research Committee.

Antibodies. Murine Morc1 coiled-coil domain (amino acids 788–950), expressed in and purified from bacteria, was provided by Jiamu Du and Dinshaw Patel (Sloan Kettering). Anti-Morc1 antibody was raised in rabbit in collaboration with Rockland Immunochemicals.

Anti-LINE Orf1p antibody was provided by Alex Bortvin (Carnegie Institution for Science) and anti-IAP Gaga antibody was provided by Bryan Cullen (Duke).

Immunofluorescence. Whole testes were fixed with 4% paraformaldehyde, immobilized in paraffin and sectioned. After removal of paraffin, sections were stained at the following antibody concentrations: anti-LINE Orf1p (1:300), anti-IAP Gag (1:300), anti-MORC1 (1:100), anti-VASA (1:100, R&D Systems AF2030), stained with fluorescent secondary antibody and mounted with DAPI (4',6'-diamidino-2-phenylindole). Slides were imaged by Confocal microscopy.

Embryonic germ cell purification. Collection of embryonic testes were performed following institutional approval for appropriate care and use of laboratory animals. Pregnant females were euthanized using CO₂ and the embryos removed from the womb and stored on a 10-cm dish filled with chilled 1 × PBS. Testicles were removed from the embryos, placed in an individual 15-ml falcon tube with 3 ml of 0.25% Trypsin, with 3 µl of DNase I 1 Unit per 1 µl (Life Technologies). Testes were incubated for 15 min at 37 °C. After incubation, the cells were agitated into suspension gently by pipetting. The trypsin was then quenched using 5 ml DMEM/10% fetal bovine serum (Life Technologies). The cells were centrifuged at 278g for 5 min and resuspended in 500 µl FACS buffer (1 × PBS 1% BSA). 7-Aminoactinomycin D was added at a 1:50 dilution (BD Biosciences) and the cells strained through BD FACS tubes (Corning) before analysis. Green fluorescent protein-positive cells were sorted into Buffer RLT (Qiagen) or ATL (Qiagen) for RNA or DNA extraction, respectively.

Postnatal germ cell purification. Pups were euthanized using isoflurane. The testes were removed using tweezers, placed in a 1.5-ml centrifuge tube and chilled on ice. When all testes had been removed, each pair was placed in 1 ml of type IV collagenase (Invitrogen) in an ultra-low-attachment six-well plate (Corning).

All extraneous tissue and the tunica were removed and the seminiferous tubules were teased apart. The samples were then incubated at 37 °C for 15 min and centrifuged for 5 min at 278g. Testes were then resuspended in 500 µl of 0.25% Trypsin (Life Technologies) and incubated for 5 min at 37 °C. After the incubation period, the testes were agitated gently into suspension by pipetting. Five hundred microlitres of DMEM/10% fetal bovine serum was added and the samples were centrifuged for 5 min at 200g.

For the P2.5 timepoints, green fluorescent protein-positive cells were sorted as with embryonic time points. To sort germ cells at P10.5, the cells were washed with 1 ml FACS buffer and then resuspended in 500 µl FACS buffer. Cells were then incubated with 1:160 EPCAM PE (Biolegend 118205) and 1:250 µl H2-K^d 647 (Biolegend 115106) on ice for 20 min in the dark, then centrifuged 5 min at 200g and resuspended in 500 µl FACS buffer. DAPI was added (1:1,000, Life Technologies) and the cells were strained through BD FACS tubes (Corning) before analysis. SSC^{lo} EPCAM^{hi} H2-K^d cells were sorted into Buffer RLT or ATL for RNA or DNA extraction, respectively.

qRT-PCR of Morc1. For embryonic samples, gonads were pooled from approximately five to seven mice per time point. RNA was extracted by the TRIzol method and DNase-treated (Qiagen) before complementary DNA conversion (Superscript III, Life Technologies). Quantitative amplification of cDNA was performed in triplicate using SYBR Green quantitation (PCR primers listed below) on a 7900 HT Fast Real Time PCR System (Applied Biosystems).

Rrm2, F: 5'-CCGAGCTGGAAAGTAAAGCG-3'

R: 5'-ATGGGAAAGACAACGAAGCG-3'

Morc Exon 7, F: 5'-GACCCGACAGAGTCTTCTCA-3'

R: 5'-TGTCGATCAATTCAGCTTC-3'.

RNA preparation. RNA for was extracted from whole testes or cells using the RNeasy Micro Kit (Qiagen 74004). The material was quantified using a Nanodrop ND-1000 (Nanodrop) for RNA from whole testis or the Qubit RNA High Sensitivity Assay (Life Technologies) for RNA from sorted germ cells. RNA quality for material from whole testis was assessed by gel electrophoresis and visualization of the 28S and 18S rRNA bands.

DNA preparation. DNA for bisulfite sequencing was extracted using the QiaAMP DNA Micro kit (Qiagen) and quantified using the Qubit dsDNA High Sensitivity Kit (Life Technologies).

qRT-PCR of retrotransposons. qRT-PCR for retrotransposons was conducted using published primer sets⁴⁵. One microgram total RNA was treated with DNase I Amplification Grade (Life Technologies) and converted to cDNA using SuperScript II Reverse transcriptase and random hexamers as primer (Life Technologies). The samples were digested with RNase H in accordance with manufacturer's protocol. RT-PCR was then performed using iQ SYBR Green Mastermix (BioRad) with 750 nM concentration of each primer. The samples was amplified (PCR programme: 95 °C 10:00, 50x (95 °C 30s, 55 °C 30s, 72 °C 30s)) with detection of PCR product after each elongation step and determination of melting temperature after the completion of PCR. The reaction was performed using an Agilent Technologies Mx3005p qPCR System (Stratagene). Upregulation of transposon transcript in the mutant is estimated using difference of squares with glyceraldehyde 3-phosphate dehydrogenase as a control.

RNA-seq library preparation. RNA from whole testes was processed for sequencing using a TruSeq RNA Sample Preparation Kit v2 (Illumina) with 250 ng–2 µg total RNA as starting material. Mutant and controls were always matched for starting RNA content. RNA from sorted germ cell was processed using the Ovation Human FFPE RNA-Seq Multiple kit (Nugen) using custom primers for depletion of murine rRNA provided by the manufacturer, using 10 ng of total RNA. Each library was prepared using RNA from one individual mouse.

Small-RNA isolation and library preparation. Total RNA was isolated from embryonic testes using Ribozol. Thirty micrograms of total RNA was loaded on 12% urea-polyacrylamide (PAA) gel. The 19–30 nt fraction was excised and snap-frozen in liquid nitrogen in 400 µl 0.4 M NaCl. RNA was eluted from the gel overnight at 16 °C while shaking at 1,000 r.p.m. and then precipitated with 3 vol absolute ethanol. Pre-adenylated 3'-linker (5'-Phos/TGGAATTCTCGGGTGCC AAGGAATC/3'-ddC; 5'-DNA adenylation kit, NEB) was ligated to RNA overnight at 4 °C using Truncated RNA Ligase 2 (NEB). Ligation reactions were loaded onto 10% urea-PAGE, the 45–56 nt fraction was excised and nucleic acids extracted as above. 5'-Linker (5'-rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrGrArCrGrArUrCr-3') was ligated to the samples using RNA Ligase 1 (NEB) overnight at 4 °C. Ligation reactions were loaded on 10% Urea-PAA gel, 72–83 nt fraction was excised and nucleic acids extracted as above. Extracted samples were reverse transcribed (primer sequence: 5'-GGAGTTCCTTGGCACCCGAGA-3') and library amplified by PCR using standard Illumina primers. Final libraries were excised from the agarose gel and sequenced.

Bisulfite library preparation. Libraries were prepared using the Ovation Ultralow Methyl-Seq Library System (Nugen). Five to 25 ng DNA was used as starting material. Matched mutant and control samples always contained identical quantities of DNA. Unmethylated Lambda phage DNA (NEB) was spiked in at 0.5% input DNA quantity to determine conversion efficiency, which was consistently > 98%. Each library was prepared using DNA from one individual mouse.

ChIP sequencing. The ChIP sequencing (ChIP-seq) protocol was adapted from published sources²⁸. FACS-sorted cells from an individual mouse were diluted to 292 μ l with 1 $\times$ PBS at room temperature. Formaldehyde (Sigma) was added to a final concentration of 1% and the sample was incubated for 10 min at room temperature with rocking. One molar glycine was then added to yield a final concentration of 0.14 M and the samples were quenched 30 min with rocking. Cells were then spun at 425g for 10 min at room temperature. The cell pellet was flash frozen.

After thawing, the cells were resuspended in 200 μ l lysis buffer (50 mM Tris-Cl pH 8.0, 20 mM EDTA pH 8.0, 1% SDS, $1 \times$ Complete Protease Inhibitor (Roche)) and incubated on ice for 10 min. Samples were then subjected to a 9-min disruption using a Bioruptor on 'High' setting, with 30 s/30 s off disruption (hence, 4.5 min of disruption in total). Samples were spun at 14,000g for 10 min, to remove insoluble material. The soluble sample was diluted to 500 μ l with dilution buffer (16.7 mM Tris pH 8, 0.01% SDS, 1.1% TritonX-100, 1.2 mM EDTA, 167 mM NaCl) and 10% of material was saved as input. Sample was precleared with 30 μ l Protein A Dynabeads (Life Technologies) and preincubated for 1 h. The cleared material was incubated with 1 μ l anti-H3K4me3 antibody (Millipore 04-745) overnight.

The samples were incubated with 30 μ l Protein A Dynabeads and the precipitated material was recovered with a magnet. The beads were washed 2 $\times$ for 4 min with Buffer A (50 mM HEPES pH 7.9, 1% Triton X-100, 0.1% Deoxycholate, 1 mM EDTA, 140 mM NaCl), 2 $\times$ for 4 min with Buffer B (50 mM HEPES pH 7.9, 0.1% SDS, 1% Triton X-100, 0.1% Deoxycholate, 1 mM EDTA, 500 mM NaCl) and 2 $\times$ for 4 min with 10 mM Tris/1 mM EDTA. Bound material was eluted with 100 μ l elution buffer (50 mM Tris pH 8.0, 1 mM EDTA, 1% SDS) at 65 $^{\circ}$ C for 10 min and then eluted a second time with 150 μ l elution buffer.

The input samples were thawed and diluted with 200 μ l buffer. Cross-linking of ChIP and input samples was reversed by incubating 16 h at 65 $^{\circ}$ C. Samples were cooled and treated with 1.5 μ l of 10 mg ml⁻¹ RNaseA (PureLink RNaseA A, Invitrogen 12091-021) for 30 min at 37 $^{\circ}$ C. One hundred micrograms of Proteinase K was then added and the samples treated for 2 h at 56 $^{\circ}$ C. The samples were then purified using a Qiagen Minelute kit.

Samples were amplified by a SeqPlex DNA Amplification kit (Sigma) and then converted to libraries using an Ovation Rapid Library kit.

ATAC-seq library construction. Libraries were generated using a method adapted from published protocol³⁰. Briefly, FACS-collected cells from individual mice were spun at 500g for 5 min at 4 °C. Cells were resuspended in 50 μ l lysis buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP40, 1 $\times$ Complete Protease Inhibitor (Roche)) and spun at 500g for 10 min. at 4 °C to collect nuclei. The nuclei were resuspended in 50 μ l Transposase reaction (25 μ l 2 $\times$ Tagmentation buffer, 22.5 μ l water, 2.5 μ l Tn5 Transposase enzyme) and reacted for 30 min at 37 °C on a PCR machine. The material was purified using a Qiagen MinElute protocol, eluting with 14 μ l EB (Qiagen).

To amplify ATAC-seq libraries from the treated material, we amplified using the Ad1 primer below and a different Ad2 primer for each sample, which functions as a barcode

Ad1: 5'-AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCG
TCAGATGTG-3'

Ad2.1_TAAGGCGA: 5'-CAAGCAGAAGACGGCATACGAGATTCGCCTT
AGTCTCGTGGGCTCGGAGATGT-3'

Ad2.2_CGTACTAG: 5'-CAAGCAGAAGACGGCATACGAGATCTAGTACG
GTCTCGTGGGCTCGGAGATGT-3'

Ad2.3_AGGCAGAA: 5'-CAAGCAGAAGACGGCATACGAGATTTCTGCCT
GTCTCGTGGGCTCGGAGATGT-3'

Ad2.4_TCCTGAGC: 5'-CAAGCAGAAGACGGCATACGAGATGCTCAGG
AGTCTCGTGGGCTCGGAGATGT-3'

Ad2.5_GGACTCCT: 5'-CAAGCAGAAGACGGCATACGAGATAGGAGTCC
GTCTCGTGGGCTCGGAGATGT-3'

Ad2.6_TAGGCATG: 5'-CAAGCAGAAGACGGCATACTGAGATCATGCC
TAGTCTCGTGGGCTCGGAGATGT-3'

Ad2.7_CTCTCTAC: 5'-CAAGCAGAAGACGGCATACGAGATGTAGAGAG
GTCTCGTGGGCTCGGAGATGT-3'

Ad2.8_CAGAGAGG: 5'-CAAGCAGAAGACGGCATACGAGATCCTCTCTG
GTCTCGTGGGCTCGGAGATGT-3'

Ad2.9_GCTACGCT: 5'-CAAGCAGAAGACGGCATACGAGATAGCGTAGC
GTCTCGTGGGCTCGGAGATGT-3'

Ad2.10_CGAGGCTG: 5'-CAAGCAGAAGACGGCATACGAGATCAGCCTC
GGTCTCGTGGGCTCGGAGATGT-3'

The eluted material was amplified in 50 μ l volume using 1.25 μ M primer concentration and a $1 \times$ concentration NEBNext High-Fidelity Master-Mix (NEB) (programme: 72 °C 5:00, 98 °C 30 s, $5 \times$ (98 °C 10 s, 63 °C 30 s, 72 °C 1 min), 4 °C hold). After these five cycles of amplification, the tube was kept on ice.

A 5- μ l aliquot was then removed and used to perform a 15- μ l side reaction with identical concentrations of primer and enzyme as above, except that 0.6 $\times$ SYBR Green (Invitrogen S-7563) is included to monitor amplification. This side reaction was amplified on a Stratagene Mx3005p qPCR (Agilent) system with the following amplification conditions (98 $^{\circ}$ C 30 s, 20 $\times$ (98 $^{\circ}$ C 10 s, 63 $^{\circ}$ C 30 s, 72 $^{\circ}$ C 1:100)). The number of additional cycles 'N' required to reach one-fourth maximum fluorescence was observed. The purpose of this side reaction was to minimize the number of PCR cycles required used to generate the libraries, as length and GC bias increases with more amplification. The remaining 45 μ l of the reaction was then further amplified (98 $^{\circ}$ C 30 s, N $\times$ (98 $^{\circ}$ C 10 s, 63 $^{\circ}$ C 30 s, 72 $^{\circ}$ C 1:00), 4 $^{\circ}$ C) and the libraries were purified by a Qiagen MinElute kit eluting with 20 μ l volume. Libraries were visualized by running on a 5% TBE gel and imaged by incubating for 20 min in 1 $\times$ SYBR Green/1 $\times$ TBE. Libraries quantified using the KAPA Library Quantification Kit (Kapa Biosystems).

RNA-seq analysis. For all analyses, reads were trimmed to 50 bp and those mapping to ribosomal RNA (GenBank identifiers: 18S NR_003278.3, 28S NR_003279.1, 5S D14832.1, 5.8S K01367.1) by up to three mismatches were discarded.

Analysis on repeat families. Reads were then mapped to the mm9 genome allowing no mismatches and keeping reads that map up to 10,000 sites in the genome using Bowtie⁴⁶. Each mapping read was assigned a score of $1/n$, where n is the number of sites in the genome the read mapped to. Repeats were obtained from RepeatMasker. Expression values for each repeat family was calculated by adding the scores contained within the repeat body, divided by the total million reads mapped and average length (kb) of repeats within the family.

Analysis on individual genes and repeats. Reads were then mapped to known mm9 gene and repeat annotations by allowing up to two mismatches and only retaining reads that mapped to one location. When reads did not map to the annotated genes and repeats, the reads were mapped to the mm9 genome. Number of reads mapping to genes and repeats were determined by using HTSeq (doi: 10.1101/002824) using default parameters. Expression values were calculated as reads per kilobase of exons per million mapping reads. Differential gene and repeat expression was determined by using DESeq⁴⁷, by using default parameters.

Whole-genome bisulfite sequencing. Reads were split into 50 bp reads before mapping. Reads were mapped to the mm9 genome as well as the lambda genome using BS seeker2 (ref. 48) using default parameters. Methylation levels were determined by $\#C/(\#C + \#T)$. For identifying DMRs, the genome was tiled into 500 bp bins and CG methylation levels in knockout and control were compared within each bin. Bins that had a methylation level difference of 50% as well as a false discovery rate < 0.05 calculated by Fisher's exact test corrected by the Benjamini-Hochberg procedure were selected. Finally, DMRs containing at least four cytosines in CG contexts, each covered by at least four reads were retained. Control regions were defined completely randomly, except that control regions have (1) exact same coverage of cytosines in CG contexts as the *Morcl*^{+/-g} data within DMRs; (2) WT CG methylation levels are similar as the *Morcl*^{+/-g} data within DMRs ($< 5\%$). (3) same number of regions per chromosome as DMRs. We defined genes as associated with DMRs when the TSS of an Ensembl transcript model was within 1 kb of a DMR.

To align to Repeat consensus sequences, the RepBase consensus sequences for 30 repetitive elements (B1_SINE, ERVB7 1-1 MM (EtnERV2/MusD), IAP-d, IAPEY3_I, IAPEY_I, IAPEY_LTR, IAPEY3_LTR, IAPEZI, IAPLTR1_MM, IAPLTR2_MM, IAPLTR3, IAPLTR3_I, IAPLTR4, IAPLTR4-I, L1MdA_I, L1MdF_I, L1MdGf_I, L1MdTf_I, First 234 bases of GSAT_MM (Major_satellite), MMEGLN_I, MMEGLN_LTR, MERVL, MERVL_LTR, First 120 bases of SATMIN (Minor_satellite), MMERVK10C, RLTR10C, RLTR27_MM, RLTR6_MM, RLTR6I_MM, RLTR6TN_MM and RSINE1) were combined into a microgenome. Then, whole-genome bisulfite sequencing reads were mapped to the microgenome using BSMAP⁴⁹, accepting uniquely mapping reads only [- w 1], mapping to two forward possible strands [- n 0] and allowing 2 mismatches [- v 2]. Methylation levels were determined by #C/(#C + #T). The methylation levels at each CG site was calculated.

Small RNA sequencing. Sequence adaptors were removed using a custom-designed dynamic programming algorithm that recognizes both exact and inexact matches, and the trimmed reads were aligned to the mm9 genome following a custom suffix array-based procedure⁵⁰. Reads with lengths > 24 nt were considered for piRNA analysis. Based on alignment coordinates, the reads were annotated as derived from exons, introns, transposons and other repeats according to the genome annotation obtained via UCSC Genome Bioinformatics⁵¹. Reads that had multiple valid alignments were annotated based on ten alignments selected at random, and the majority annotation was assigned as the final annotation. In case of ties, annotation was picked based on a fixed hierarchy principle⁵⁰. Sense or antisense annotation was assigned to piRNA reads with respect to the strandedness of an underlying genomic feature. If a piRNA read contained U in position 1, such piRNA was considered as primary, while the presence of A in position 10 defined secondary piRNAs.

ChIP-seq analysis. Previously published ChIP-seq data²⁸ was obtained from GSE38165 in Gene Expression Omnibus. Reads were mapped to the mm9 genome using Bowtie by allowing up to two mismatches and only retaining reads that mapped to one location in the genome. Reads mapping to the same location were collapsed into one read. For all analyses, the data were normalized to total number of mapping reads in the library

ATAC-seq analysis. Data were collected using 50 bp paired end sequencing on a HiSeq. In keeping with established methodologies³⁰, reads were aligned to mm9 using Bowtie⁵² with the parameters -X2000 and -m1. The -X2000 parameter allows the fragments <2 kb to align and only unique aligning reads were collected (-m1). Duplicated reads were removed with samtools (rmdup function)⁵³. Previous results show that for Tn5 transposase, the transposon binds as a dimer and insert two adaptors separated by 9 bp⁵⁴. Thus, all reads aligned to the positive strands were offset by +4 bp and all reads aligned to the negative strands were offset by -5 bp.

References

- Moissiard, G. *et al.* MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* **336**, 1448–1451 (2012).
- Lorkovic, Z. J., Naumann, U., Matzke, A. J. & Matzke, M. Involvement of a GHKL ATPase in RNA-directed DNA methylation in *Arabidopsis thaliana*. *Curr. Biol.* **22**, 933–938 (2012).
- Hruz, T. *et al.* Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinformatics* **2008**, 420747 (2008).
- Watson, M. L. *et al.* Identification of morc (microorchidia), a mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse. *Proc. Natl Acad. Sci. USA* **95**, 14361–14366 (1998).
- Inoue, N. *et al.* New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum. Mol. Genet.* **8**, 1201–1207 (1999).
- Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).
- Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862 (2012).
- Kobayashi, H. *et al.* High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.* **23**, 616–627 (2013).
- Kaneda, M. *et al.* Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900–903 (2004).
- Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99 (2004).
- Webster, K. E. *et al.* Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc. Natl Acad. Sci. USA* **102**, 4068–4073 (2005).
- Carmell, M. A. *et al.* MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* **12**, 503–514 (2007).
- Shoji, M. *et al.* The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev. Cell* **17**, 775–787 (2009).
- Iyer, L. M., Abhiman, S. & Aravind, L. MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol. Direct* **3**, 8 (2008).
- Wernig, M. *et al.* In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
- Vincent, J. J. *et al.* Single cell analysis facilitates staging of Blimp1-dependent primordial germ cells derived from mouse embryonic stem cells. *PLoS ONE* **6**, e28960 (2011).
- Small, C. L., Shima, J. E., Uzumcu, M., Skinner, M. K. & Griswold, M. D. Profiling gene expression during the differentiation and development of the murine embryonic gonad. *Biol. Reprod.* **72**, 492–501 (2005).
- Hata, K., Kusumi, M., Yokomine, T., Li, E. & Sasaki, H. Meiotic and epigenetic aberrations in Dnmt3L-deficient male germ cells. *Mol. Reprod. Dev.* **73**, 116–122 (2006).
- Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* **31**, 785–799 (2008).
- Bellve, A. R., Millette, C. F., Bhatnagar, Y. M. & O'Brien, D. A. Dissociation of the mouse testis and characterization of isolated spermatogenic cells. *J. Histochem. Cytochem.* **25**, 480–494 (1977).
- Oatley, J. M. & Brinster, R. L. Spermatogonial stem cells. *Methods Enzymol.* **419**, 259–282 (2006).
- Niles, K. M., Chan, D., La Salle, S., Oakes, C. C. & Trasler, J. M. Critical period of nonpromoter DNA methylation acquisition during prenatal male germ cell development. *PLoS ONE* **6**, e24156 (2011).
- Tomizawa, S. *et al.* Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* **138**, 811–820 (2011).
- Watanabe, T. *et al.* Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgr1 locus. *Science* **332**, 848–852 (2011).
- Chotalia, M. *et al.* Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev.* **23**, 105–117 (2009).
- Sookdeo, A., Hepp, C. M., McClure, M. A. & Boissinot, S. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mobile DNA* **4**, 3 (2013).
- Leung, D. C. & Lorincz, M. C. Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem. Sci.* **37**, 127–133 (2012).
- Ng, J. H. *et al.* In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures. *Dev. Cell* **24**, 324–333 (2013).
- Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- Kuramochi-Miyagawa, S. *et al.* Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**, 839–849 (2004).
- Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744–747 (2007).
- Watanabe, T. *et al.* MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Dev. Cell* **20**, 364–375 (2011).
- Frost, R. J. *et al.* MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. *Proc. Natl Acad. Sci. USA* **107**, 11847–11852 (2010).
- Zheng, K. *et al.* Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc. Natl Acad. Sci. USA* **107**, 11841–11846 (2010).
- Soper, S. F. *et al.* Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev. Cell* **15**, 285–297 (2008).
- Aravin, A. A. *et al.* Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genet.* **5**, e1000764 (2009).
- Saxe, J. P., Chen, M., Zhao, H. & Lin, H. Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. *EMBO J.* **32**, 1869–1885 (2013).
- Tanaka, S. S. *et al.* The mouse homolog of *Drosophila* vasa is required for the development of male germ cells. *Genes Dev.* **14**, 841–853 (2000).
- Kuramochi-Miyagawa, S. *et al.* MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev.* **24**, 887–892 (2010).
- Molaro, A. *et al.* Two waves of de novo methylation during mouse germ cell development. *Genes Dev.* **28**, 1544–1549 (2014).
- Ollinger, R. *et al.* Deletion of the pluripotency-associated Tex19.1 gene causes activation of endogenous retroviruses and defective spermatogenesis in mice. *PLoS Genet.* **4**, e1000199 (2008).
- Tarabay, Y. *et al.* The mammalian-specific Tex19.1 gene plays an essential role in spermatogenesis and placenta-supported development. *Hum. Reprod.* **28**, 2201–2214 (2013).
- Reichmann, J. *et al.* Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. *PLoS Comput. Biol.* **8**, e1002486 (2012).
- Karimi, M. M. *et al.* DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**, 676–687 (2011).
- Langmead, B., Schatz, M. C., Lin, J., Pop, M. & Salzberg, S. L. Searching for SNPs with cloud computing. *Genome Biol.* **10**, R134 (2009).
- Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
- Guo, W. *et al.* BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**, 774 (2013).
- Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10**, 232 (2009).
- Olson, A. J., Brennecke, J., Aravin, A. A., Hannon, G. J. & Sachidanandam, R. Analysis of large-scale sequencing of small RNAs. *Pac. Symp. Biocomput.* **13**, 126–136 (2008).
- Karolchik, D. *et al.* The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.* **42**, D764–D770 (2014).
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11**, R119 (2010).

Acknowledgements

We acknowledge the UCLA BSCRC Flow Cytometry core for flow and FACS assistance, and the UCLA BSCRC High Throughput Sequencing Core. Bryan Cullen provided anti-IAP Gag antibody and Alexander Bortvin provided anti-LINE ORF1p antibody. Jiamu Du and Dinshaw Patel provided recombinant MORC1 coiled-coil domain to raise antibody.

W.A.P. is supported by the Jane Coffin Childs Memorial Fund for Medical Research. H.S. is an HHMI Fellow of the Damon Runyon Cancer Research Foundation (DRG-2194-14). W.L. is supported by a grant from the Chinese Scholar Council. This work was supported by grants from the NIH (R01 HD058047 and DP2 OD007371A), a Research Award from the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research and the Nogradi Fund, the Searle Scholar and the Packard Fellowship Awards, and an ANR grant ("TranspoFertil"). S.E.J. is an investigator of the Howard Hughes Medical Institute.

Author contributions

W.A.P., K.N., D.P., N.Z. and S.A.L. managed mice, performed dissections and purified DNA and RNA from materials. W.A.P., D.P. and G.M. generated sequencing libraries. H.S., W.L. and S.M. performed bioinformatics analysis. W.A.P. and S.A.L. performed immunofluorescent staining of target tissue. W.A.P., A.T.C. and S.E.J. wrote the manuscript. D.B., A.A.A., A.T.C. and S.E.J. coordinated research.

Additional information

Accession codes: Raw sequencing data, including RNA-seq, whole-genome bisulfite-seq, small RNA-seq, ATAC-seq and ChIP-seq, generated for this study have been deposited in the GEO database under accession number GSE63048.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Pastor, W. A. *et al.* MORC1 represses transposable elements in the mouse male germline. *Nat. Commun.* 5:5795 doi: 10.1038/ncomms6795 (2014).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Erratum: MORC1 represses transposable elements in the mouse male germline

William A. Pastor, Hume Stroud, Kevin Nee, Wanlu Liu, Dubravka Pezic, Sergei Manakov, Serena A. Lee, Guillaume Moissiard, Natasha Zamudio, Déborah Bourc'his, Alexei A. Aravin, Amander T. Clark & Steven E. Jacobsen

Nature Communications 5:5795 doi: 10.1038/ncomms6795 (2014); Published 12 Dec 2014; Updated 10 Jul 2015

This Article contains errors in Fig. 2 that were introduced during the production process. The bars of the graph in panel f were inadvertently switched with those of the graph in panel g. The correct version of the figure appears below.

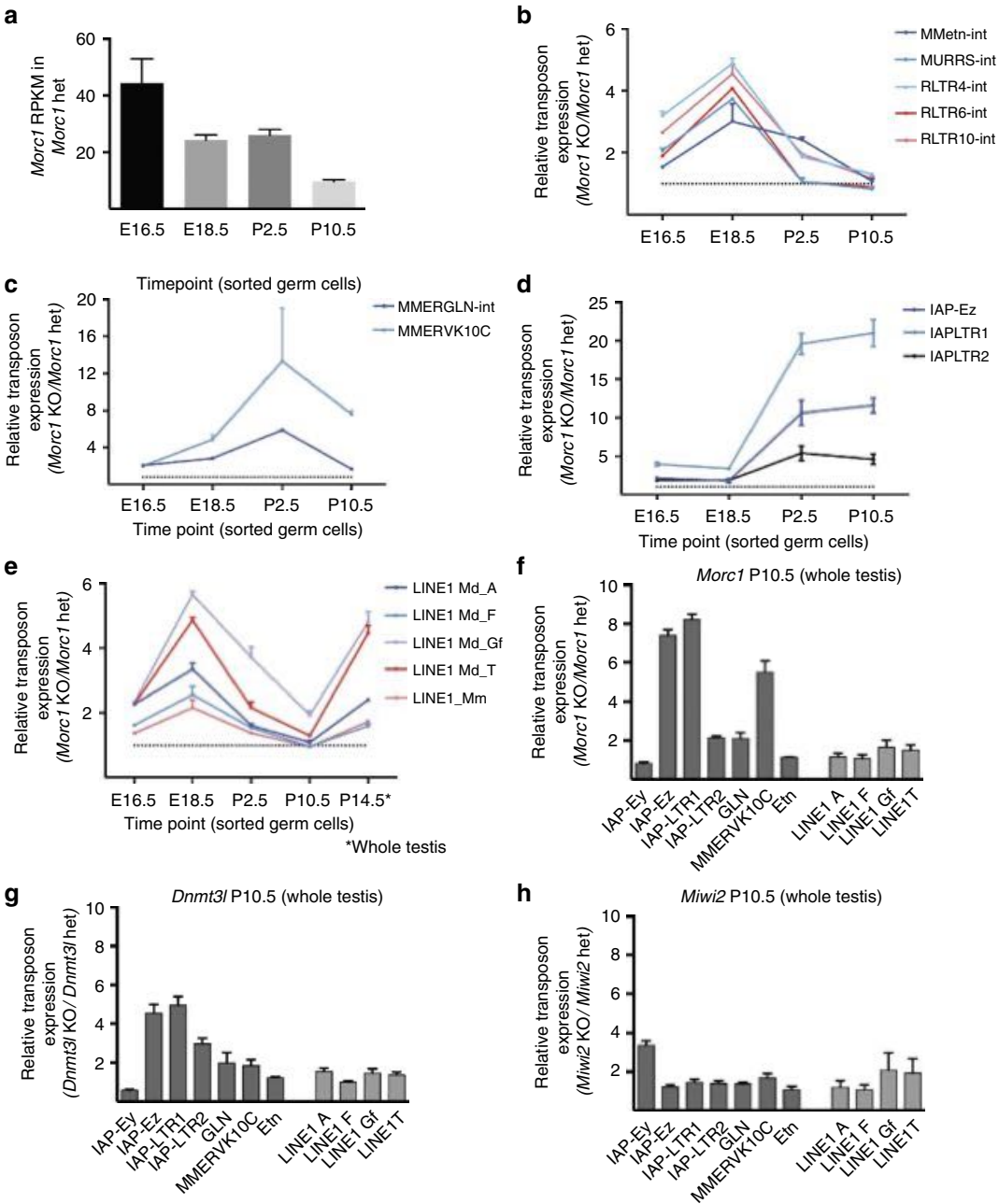


Figure 2

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of October 21, 2014):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/336/6087/1448.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2012/05/02/science.1221472.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/336/6087/1448.full.html#related>

This article **cites 44 articles**, 14 of which can be accessed free:

<http://www.sciencemag.org/content/336/6087/1448.full.html#ref-list-1>

This article has been **cited by** 13 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/336/6087/1448.full.html#related-urls>

This article appears in the following **subject collections**:

Molecular Biology

http://www.sciencemag.org/cgi/collection/molec_biol

7. Y. Onodera *et al.*, *Cell* **120**, 613 (2005).
8. J. Penterman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6752 (2007).
9. R. Lister *et al.*, *Cell* **133**, 523 (2008).
10. X. Zhang, Y. V. Bernatavichute, S. Cokus, M. Pellegrini, S. E. Jacobsen, *Genome Biol.* **10**, R62 (2009).
11. A. Kirmizis *et al.*, *Nature* **449**, 928 (2007).

Acknowledgments: This work was supported by National Institutes of Health grants R01GM070795 and R01GM059138 to J.-K.Z. and by the Chinese Academy of Sciences (CAS), China. We thank B. Stevenson for technical assistance and X. Zhang for the gift of glutathione S-transferase H3 (amino acids 1 to 57) construct. Sequence data are available in the Gene Expression Omnibus database (GEO accession GSE33071).

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6087/1445/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 to S7
References (12–20)

20 January 2012; accepted 26 April 2012
10.1126/science.1219416

MORC Family ATPases Required for Heterochromatin Condensation and Gene Silencing

Guillaume Moissiard,¹ Shawn J. Cokus,¹ Joshua Cary,¹ Suhua Feng,¹ Allison C. Billi,² Hume Stroud,¹ Dylan Husmann,¹ Ye Zhan,³ Bryan R. Lajoie,³ Rachel Patton McCord,³ Christopher J. Hale,¹ Wei Feng,⁴ Scott D. Michaels,⁴ Alison R. Frand,⁵ Matteo Pellegrini,^{1,6} Job Dekker,³ John K. Kim,² Steven E. Jacobsen^{1,5,6,7*}

Transposable elements (TEs) and DNA repeats are commonly targeted by DNA and histone methylation to achieve epigenetic gene silencing. We isolated mutations in two *Arabidopsis* genes, *AtMORC1* and *AtMORC6*, which cause derepression of DNA-methylated genes and TEs but no losses of DNA or histone methylation. *AtMORC1* and *AtMORC6* are members of the conserved *Microrchidia* (MORC) adenosine triphosphatase (ATPase) family, which are predicted to catalyze alterations in chromosome superstructure. The *atmorc1* and *atmorc6* mutants show decondensation of pericentromeric heterochromatin, increased interaction of pericentromeric regions with the rest of the genome, and transcriptional defects that are largely restricted to loci residing in pericentromeric regions. Knockdown of the single MORC homolog in *Caenorhabditis elegans* also impairs transgene silencing. We propose that the MORC ATPases are conserved regulators of gene silencing in eukaryotes.

Gene silencing in the *Arabidopsis* genome is highly correlated with DNA methylation, which is found in three different

cytosine contexts. Methylation of symmetric CG and CHG sites (in which H is A, T, or C) are mediated by DNA METHYLTRANSFERASE1

(MET1) and CHROMOMETHYLASE3 (CMT3), respectively, whereas CHH methylation is mainly catalyzed by DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2) (1). Silent loci are also enriched in the repressive histone H3 lysine 9 dimethylation mark (H3K9me2) (2, 3).

Suppressor of drm2 cmt3 (*SDC*) is a gene whose repression in most tissues depends on the redundant activities of DRM2 and CMT3 (4, 5). Hence, a loss of *SDC* silencing is observed in the *drm2 cmt3* double mutant but not in *drm2* or *cmt3*

¹Department of Molecular, Cell, and Developmental Biology, University of California at Los Angeles, Terasaki Life Sciences Building, 610 Charles Young Drive East, Los Angeles, CA 90095–723905, USA. ²Life Sciences Institute and Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA. ³Lazare Research Building 570N, Program in Systems Biology and Gene Function and Expression, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA 01605, USA. ⁴Department of Biology, Indiana University, Bloomington, IN 47405, USA. ⁵Department of Biological Chemistry, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA. ⁶Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, CA 90095, USA. ⁷Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, CA 90095, USA.

*To whom correspondence should be addressed. E-mail: jacobsen@ucla.edu

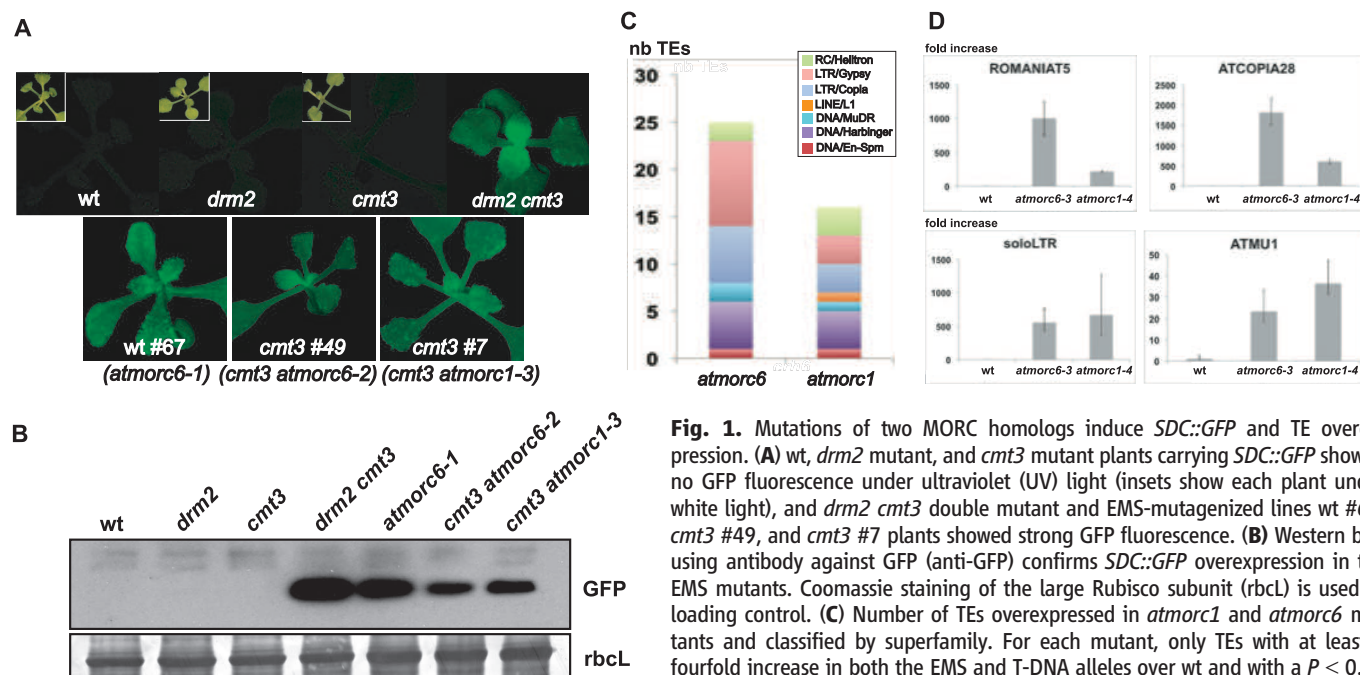


Fig. 1. Mutations of two MORC homologs induce *SDC::GFP* and TE overexpression. (A) wt, *drm2* mutant, and *cmt3* mutant plants carrying *SDC::GFP* showed no GFP fluorescence under ultraviolet (UV) light (insets show each plant under white light), and *drm2 cmt3* double mutant and EMS-mutagenized lines wt #67, *cmt3* #49, and *cmt3* #7 plants showed strong GFP fluorescence. (B) Western blot using antibody against GFP (anti-GFP) confirms *SDC::GFP* overexpression in the EMS mutants. Coomassie staining of the large Rubisco subunit (rbcl) is used as loading control. (C) Number of TEs overexpressed in *atmorc1* and *atmorc6* mutants and classified by superfamily. For each mutant, only TEs with at least a fourfold increase in both the EMS and T-DNA alleles over wt and with a $P \leq 0.05$ are represented. (D) Relative fold increase of four TE transcripts in *atmorc1-4* and

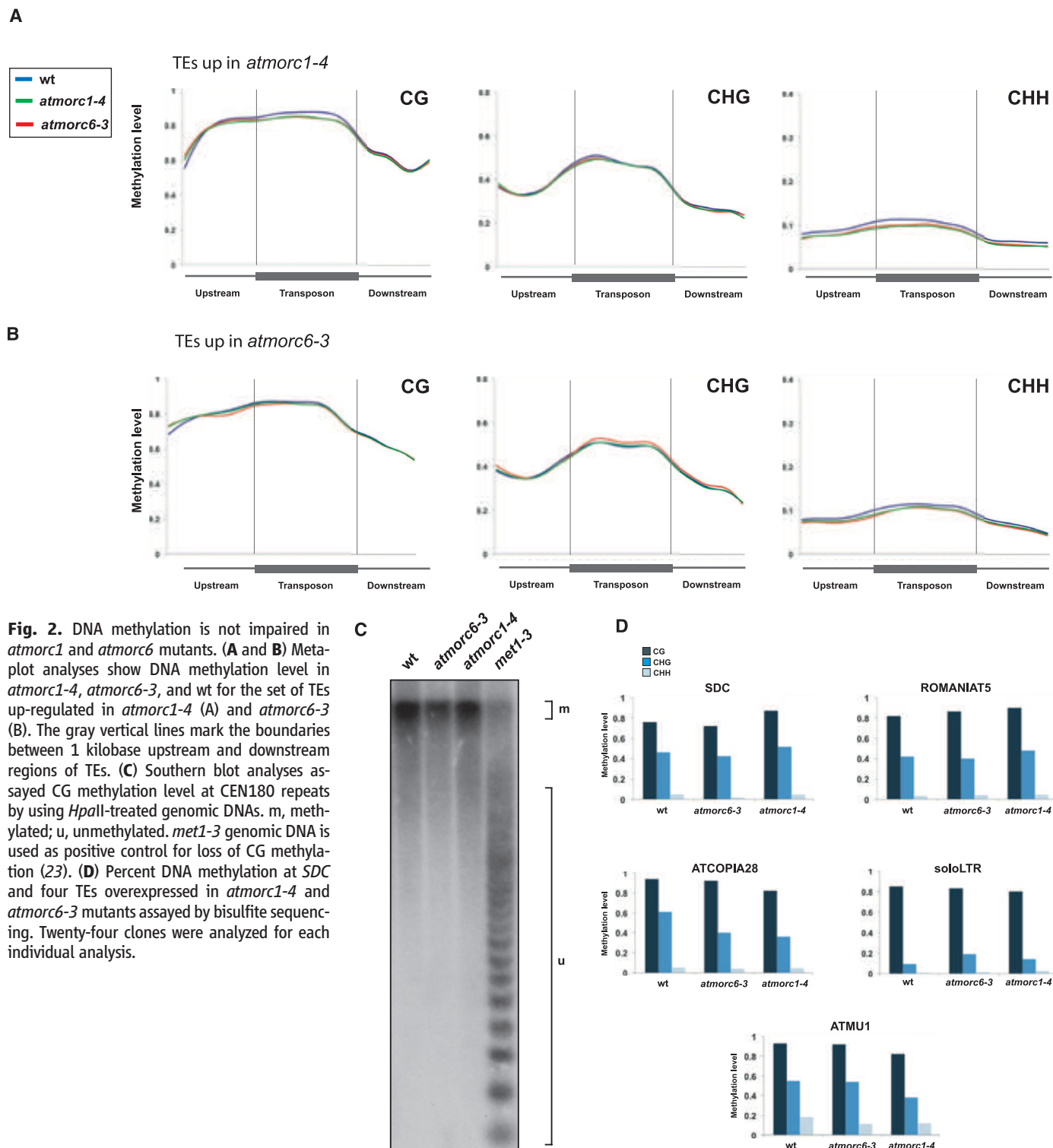
atmorc6-3 over wt assayed by real-time quantitative polymerase chain reaction (RT-qPCR) and normalized to *ACTIN7*. Errors bars indicate standard deviation based on three independent biological replicates.

single mutants. The *SDC* promoter carries seven tandem repeats, which recruit the DNA methylation machinery and cause transcriptional gene silencing. We engineered a green fluorescent protein (GFP)-based sensor construct controlled by the *SDC* promoter (fig. S1A). The *SDC::GFP* transgene behaves similarly to endogenous *SDC*, and GFP fluorescence is not detectable in wild-

type, *drm2*, or *cmt3* plants but is highly expressed in *drm2 cmt3* double mutant (Fig. 1A).

We carried out ethyl methanesulfonate (EMS) mutagenesis screens in wild-type (wt) or *cmt3* backgrounds for mutants showing *SDC::GFP* overexpression and identified the wt #67, *cmt3* #7, and *cmt3* #49 mutants (Fig. 1, A and B). Mapping experiments using bulk segregant anal-

ysis coupled to deep genome resequencing indicated that *cmt3* #7 contained a mutation in *At4g36290* (*AtMORC1*), previously also named *COMPROMISED RECOGNITION OF TCV-1* (*CRT1*) (6, 7), whereas wt #67 and *cmt3* #49 both contained mutations in *At1g19100* (*AtMORC6*) (7) (figs. S1B, S2, and S3A). An *atmorc1* allele was previously reported to show reduced resistance to



the turnip crinkle virus (TCV) (6, 7), suggesting that AtMORC1 is involved in viral resistance in addition to its role in gene silencing described in this study, whereas mutations in *AtMORC6* have not been described. To ensure that *atmorc1* and *atmorc6* mutations were those responsible for the loss of *SDC* silencing, we isolated knock-out transferred DNA (T-DNA) insertion lines *atmorc1-4* and *atmorc6-3* and confirmed *SDC* overexpression in these two mutant alleles (fig. S3, B to D). Genetic complementation crosses between the recessive EMS and T-DNA mutants confirmed *AtMORC1* and *AtMORC6* as the mutated genes responsible for *SDC::GFP* activation in the three EMS lines (fig. S3E). Therefore, #7, #67, and #49 were renamed *atmorc1-3*, *atmorc6-1*, and *atmorc6-2*, respectively.

By using RNA sequencing (RNA-seq) (8), we found that the majority of RNAs significantly affected in the *atmorc1* and *atmorc6* mutants showed up-regulation, and many of these were transposable elements (TEs) belonging to various transposon superfamilies, including, among others, the LTR/Gypsy, LTR/Copia, DNA/MuDR, and DNA/Harbinger families (Fig. 1, C and D; fig. S4A; table S1). The expression defects in the *atmorc1* and *atmorc6* mutants were very similar, with all but two of the transposons up-regulated in *atmorc1* also up-regulated in *atmorc6* (fig. S4B). Protein-coding genes overexpressed in the *atmorc1* and *atmorc6* EMS and T-DNA mutants included endogenous *SDC* (table S2). There was a high degree of overlap between the genes up-regulated in *atmorc1* and *atmorc6* (fig. S4C), most of them corresponding to DNA-methylated and silenced loci (fig. S4, D and E). We also performed RNA-seq in the *atmorc1 atmorc6* double mutant and found a very similar set of genes and transposons up-regulated, with only a few genes up-regulated in the double mutant that were not up-regulated in each of the single mutants (table S3), suggesting that AtMORC1 and AtMORC6 may act together to enforce gene silencing.

Whole-genome bisulfite sequencing (BS-seq) (9) revealed that DNA methylation levels in all sequence contexts were unaltered in *atmorc1* or *atmorc6* relative to wild type at TEs up-regulated in *atmorc1* or *atmorc6* (Fig. 2, A and B), nor were there any bulk alterations in protein-coding genes or TEs in the genome (fig. S5, A and B). In addition, analyses at the pericentromeric satellite CEN180 repeats and five loci up-regulated in *atmorc1* and *atmorc6* showed that the DNA methylation patterns in *atmorc1-4* and *atmorc6-3* were similar to those of wild type (Fig. 2, C and D). Chromatin immunoprecipitation sequencing (ChIP-seq) analyses of H3K9me2 also did not reveal any changes in the *atmorc1* or *atmorc6* mutants at *SDC* or other up-regulated locations (fig. S6, A and B). Lastly, small RNA sequencing analyses showed that elements up-regulated in *atmorc1* and *atmorc6* mutants were enriched in small interfering RNAs (siRNAs), but these siRNA levels did not change in the mutants (fig. S7). Thus, *AtMORC1* and *AtMORC6* are not required

to maintain DNA methylation, H3K9me2, or siRNAs, suggesting that AtMORC1 and AtMORC6 are likely to either act downstream of DNA methylation or enforce silencing by a novel mechanism.

AtMORC1 and *AtMORC6* are homologs of mouse *Microrchidia1* (*MORC1*) (10, 11) and contain gyrase, Hsp90, histidine kinase, and MutL (GHKL) and S5 domains, together comprising an adenosine triphosphatase (ATPase) module (6) in addition to a putative C-terminal coiled-coil domain (fig. S1B). The EMS mutations found

in *atmorc1-3*, *atmorc6-1*, and *atmorc6-2* alleles all introduced premature stop codons within the GHKL domain (fig. S1B).

Because of the similarity of AtMORC1 and AtMORC6 to ATPases involved in manipulating chromatin superstructure (12), these proteins may affect gene silencing through higher-order compaction of methylated and silent chromatin. In wild-type nuclei, pericentromeric heterochromatin forms densely staining nuclear bodies called chromocenters that localize to the nuclear periphery

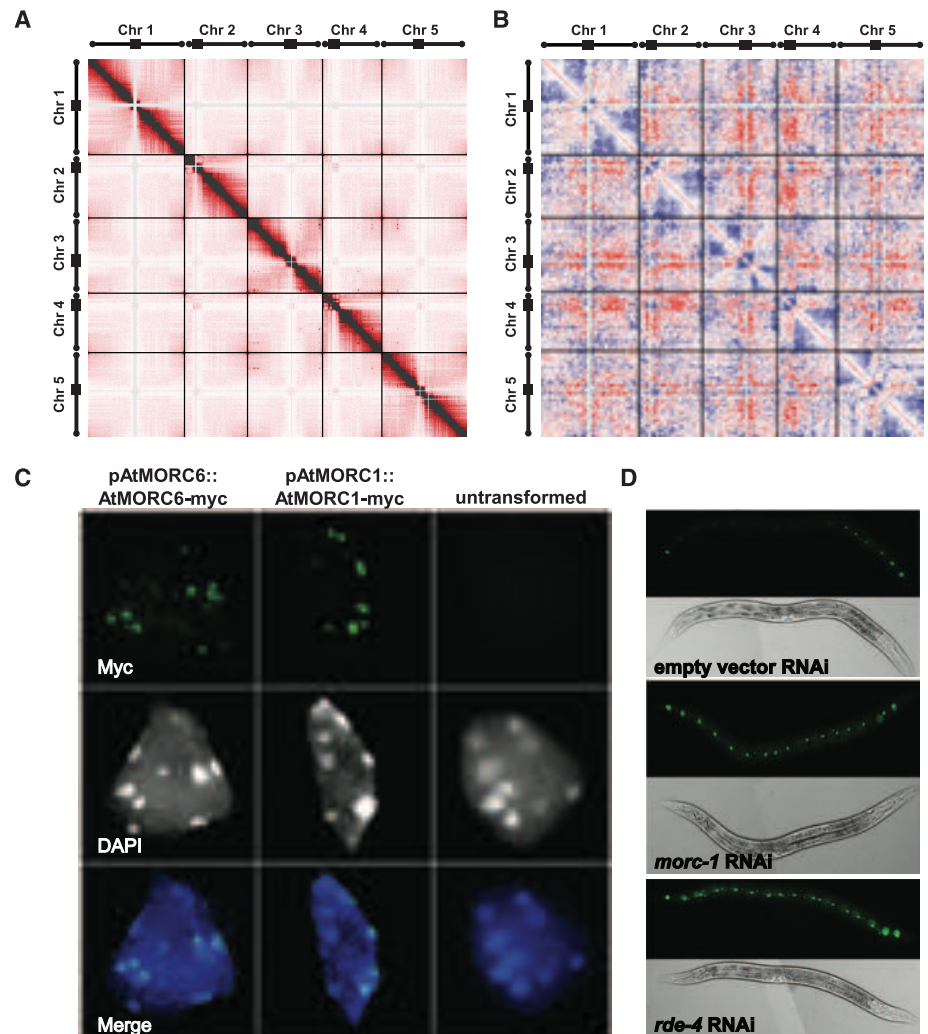


Fig. 3. AtMORC1 and AtMORC6 are required for maintenance of chromatin architecture and form nuclear bodies near chromocenters, and *morc-1* is involved in gene silencing in *C. elegans*. (A) Interaction matrix of the wt *Arabidopsis* genome from Hi-C analysis. Positions along the five chromosomes are shown from left to right and top to bottom, and each pixel represents interactions from uniquely mapping paired end reads in 200-kilobase bins. Black bars and circles mark the positions of the pericentromeric and telomeric regions, respectively. Light gray regions represent areas masked out because of problematic mapping. Black bars show separation between chromosomes. (B) Difference plot shows enrichment of Hi-C interactions in *atmorc6-1* in red and interactions depleted in *atmorc6-1* in blue. (C) Anti-Myc immunostaining showing localization of pAtMORC6::AtMORC6-Myc and pAtMORC1::AtMORC1-Myc in nuclear bodies adjacent to chromocenters. AtMORC1 and AtMORC6 showed 2.0 ± 1.0 (average $\pm$ standard deviation) and 2.5 ± 1.2 bodies per chromocenter, respectively. DAPI (4',6-diamidino-2-phenylindole) staining shows chromocenter location. Bottom images are merges. (D) A silenced seam cell-specific GFP transgene in the *eri-1* (*mg366*) sensitized background is overexpressed in worms fed with bacteria expressing double-stranded RNA targeting *morc-1* or *rde-4* but not in worms fed with bacteria expressing a control empty vector. Results are representative of five independent replicates.

(13). We observed decondensation of chromocenters in the *atmorc1* and *atmorc6* mutants (as well as in *atmorc1 atmorc6* double mutant) (figs. S8 to S11) and found that loci transcriptionally derepressed in the mutants mostly localized to pericentromeric heterochromatin (fig. S12 and tables S1 and S3). To directly examine whole-genome chromatin interactions, we performed Hi-C analyses in wild type and *atmorc6-1* (14). Consistent with previous cytological studies (13), the wild-type genome showed interactions between telomeres as well as between euchromatic regions on the same chromosome arm (Fig. 3A). In contrast, pericentromeric heterochromatin regions interacted very weakly with the rest of the genome, consistent with their compaction in chromocenters (Fig. 3A). Although *atmorc6-1* showed a roughly similar chromatin architecture (fig. S13), plotting the differences between mutant and wild type showed that *atmorc6-1* shows an increase in interactions between the pericentromeric regions of all chromosomes with the euchromatic arms of all chromosomes and a corresponding depletion of interactions of euchromatic arms with themselves. Because the analysis reports relative changes with the sum of differences set to zero, the most likely interpretation of these findings is that pericentromeric regions interact more strongly with the euchromatic arms in *atmorc6-1*, although we cannot exclude that the mutant also has effects on the euchromatic arms (Fig. 3B). This interpretation is consistent with the cytological observations showing that chromocenters expand out into a larger area of the nucleus in the mutants (fig. S8). We also found, by using complementing myc-tagged transgenes, that AtMORC1 and AtMORC6 proteins formed small nuclear bodies that were usually adjacent to but not within chromocenters (Fig. 3C and figs. S14 and S15). These results are all consistent with a model in which AtMORC1 and AtMORC6 enforce compaction and gene silencing of pericentromeric heterochromatin, although it is also possible that changes in chromatin and gene expression in the mutant secondarily lead to the observed changes in chromatin compaction. Mutation of the plant-specific *MOM1* gene has also been shown to affect gene silencing but not DNA methylation in *Arabidopsis*; however, *mom1* mutants do not show chromocenter decondensation and therefore are likely to act via a different mechanism (15, 16).

A single MORC homolog, *more-1*, is present in the worm *Caenorhabditis elegans*, which is devoid of DNA methylation (17). To test whether the *C. elegans more-1* (ZC155.3) is involved in gene silencing, we performed RNA interference (RNAi)-mediated knockdown of *more-1* in the *eri-1* sensitized background, in which a GFP transgene is silenced in most of the worm seam cells (Fig. 3D) (18). *more-1*-depleted worms showed GFP reactivation similar to worms depleted of *rde-4*, an essential component of gene silencing in *C. elegans* (Fig. 3D) (19). These results suggest that MORCs may play an ancient and conserved role in gene silencing. In addition, the

observation that *more-1* is required for gene silencing in *C. elegans* reinforces our view that MORCs in *Arabidopsis* are enforcing silencing by a mechanism that may not be directly linked with DNA methylation. It is interesting to note that the phenotype of the *Morc1*-knockout mouse resembles *Miw12*- and *Dnmt3L*-knockout mouse phenotypes, showing male-specific meiotic defects during spermatogenesis (10, 20–22). *Miw12* and *Dnmt3L* are both required for TE silencing, and it is possible that *Morc1* might be involved in transposon silencing in mammals as well. We propose that MORC family ATPases act to regulate chromatin architecture and gene silencing in a wide variety of eukaryotes.

References and Notes

1. J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **11**, 204 (2010).
2. J. P. Jackson, A. M. Lindroth, X. Cao, S. E. Jacobsen, *Nature* **416**, 556 (2002).
3. F. Malagnac, L. Bartee, J. Bender, *EMBO J.* **21**, 6842 (2002).
4. X. Zhang et al., *Cell* **126**, 1189 (2006).
5. I. R. Henderson, S. E. Jacobsen, *Genes Dev.* **22**, 1597 (2008).
6. H. G. Kang, J. C. Kuhl, P. Kachroo, D. F. Klessig, *Cell Host Microbe* **3**, 48 (2008).
7. H. G. Kang et al., *Plant Cell* **22**, 918 (2010).
8. Z. Wang, M. Gerstein, M. Snyder, *Nat. Rev. Genet.* **10**, 57 (2009).
9. S. J. Cokus et al., *Nature* **452**, 215 (2008).
10. M. L. Watson et al., *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14361 (1998).
11. N. Inoue et al., *Hum. Mol. Genet.* **8**, 1201 (1999).
12. L. M. Iyer, S. Abhiman, L. Aravind, *Biol. Direct* **3**, 8 (2008).
13. P. Fransz, J. H. De Jong, M. Lysak, M. R. Castiglione, I. Schubert, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14584 (2002).

14. E. Lieberman-Aiden et al., *Science* **326**, 289 (2009).
15. P. Amedeo, Y. Habu, K. Afsar, O. Mittelsten Scheid, J. Paszkowski, *Nature* **405**, 203 (2000).
16. A. V. Probst, P. F. Fransz, J. Paszkowski, O. Mittelsten Scheid, *Plant J.* **33**, 743 (2003).
17. V. J. Simpson, T. E. Johnson, R. F. Hammen, *Nucleic Acids Res.* **14**, 6711 (1986).
18. S. Kennedy, D. Wang, G. Ruvkun, *Nature* **427**, 645 (2004).
19. H. Tabara et al., *Cell* **99**, 123 (1999).
20. M. A. Carmell et al., *Dev. Cell* **12**, 503 (2007).
21. D. Bourc'his, T. H. Bestor, *Nature* **431**, 96 (2004).
22. A. A. Aravin et al., *Mol. Cell* **31**, 785 (2008).
23. M. W. Kankel et al., *Genetics* **163**, 1109 (2003).

Acknowledgments: We thank M. Akhavan for sequencing, L. Goddard and L. Iruela-Arispe for assistance with confocal microscopy, and P. Fransz and I. Schubert for helpful discussions. S.F. is a Special Fellow of the Leukemia and Lymphoma Society. J.C. is supported by the Ruth L. Kirschstein National Research Service Award GM007185. R.P.M. is supported by the National Institute of General Medical Sciences (grant F32GM100617), and J.D. is supported by a W. M. Keck Foundation Distinguished Young Scholar in Medical Research Award. Research in the Jacobsen, Kim, Michaels, and Dekker laboratories was supported by NIH grants GM60398, GM088565, GM075060, and HG003143, respectively. Sequencing files have been deposited at Gene Expression Omnibus (GEO) (accession code GSE37644). The authors declare no competing financial interests. S.E.J. is an investigator of the Howard Hughes Medical Institute. Correspondence and requests for materials should be addressed to S.E.J.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1221472/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S4
References (24–44)

5 March 2012; accepted 24 April 2012
Published online 3 May 2012;
10.1126/science.1221472

The Structures of COPI-Coated Vesicles Reveal Alternate Coatomer Conformations and Interactions

Marco Faini,¹ Simone Prinz,¹ Rainer Beck,² Martin Schorb,¹ James D. Riches,^{1*} Kirsten Bacia,³ Britta Brügger,² Felix T. Wieland,^{2†} John A. G. Briggs^{1,4†}

Transport between compartments of eukaryotic cells is mediated by coated vesicles. The archetypal protein coats COPI, COPII, and clathrin are conserved from yeast to human. Structural studies of COPII and clathrin coats assembled in vitro without membranes suggest that coat components assemble regular cages with the same set of interactions between components. Detailed three-dimensional structures of coated membrane vesicles have not been obtained. Here, we solved the structures of individual COPI-coated membrane vesicles by cryoelectron tomography and subtomogram averaging of in vitro reconstituted budding reactions. The coat protein complex, coatomer, was observed to adopt alternative conformations to change the number of other coatomers with which it interacts and to form vesicles with variable sizes and shapes. This represents a fundamentally different basis for vesicle coat assembly.

Cellular transport vesicles are formed by conserved protein coats (1–3). Detailed structural information about vesicle coats assembled on a membrane bilayer has remained elusive. The clearest insights into the architecture of vesicle coats have been obtained by applying

electron microscopy (EM) to coat protein complex COPII and clathrin protein cages, assembled in vitro from outer coat protein components in the absence of membranes (1, 4, 5). The cages have point group symmetries and discrete size distributions (6), whereas in vivo formed clathrin-

scientific report

Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation

Angelique Deleris^{1*}, Maxim V.C. Greenberg^{1*}, Israel Ausin¹, Rona W.Y. Law¹, Guillaume Moissiard¹, Daniel Schubert² & Steven E. Jacobsen^{1,3+}

¹Department of Molecular, Cell, and Developmental Biology, University of California at Los Angeles, Los Angeles, California, USA,

²Institute of Genetics, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany and ³Howard Hughes Medical Institute,

University of California at Los Angeles, Los Angeles, California, USA

 This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial Share Alike 3.0 Unported License, which allows readers to alter, transform, or build upon the article and then distribute the resulting work under the same or similar licence to this one. The work must be attributed back to the original author and commercial use is not permitted without specific permission.

Histone demethylases—both lysine-specific demethylase 1 (LSD1) and Jumonji-C (JmjC) domain-containing proteins—are broadly implicated in the regulation of chromatin-dependent processes. In *Arabidopsis thaliana*, histone marks directly affect DNA methylation, and mutations in LSD1 homologues show reduced DNA methylation at some loci. We screened transfer DNA mutations in genes encoding JmjC domains for defects in DNA methylation. Mutations in *jmj14* result in reduced DNA methylation in non-CG contexts at targets of DRM2 (domains rearranged methyltransferase 2)-mediated RNA-directed DNA methylation (RdDM), which is associated with an increase in H3K4m3. Unlike other components of RdDM, JMJ14 is not required for *de novo* methylation of a transgene, suggesting that JMJ14 is specifically involved in the maintenance phase of DRM2-mediated RdDM.

Keywords: DNA methylation; epigenetics; Jumonji-C; histone demethylase; *Arabidopsis*

EMBO reports (2010) 11, 950–955. doi:10.1038/embor.2010.158

INTRODUCTION

Cytosine DNA methylation is an epigenetic modification that is conserved in all kingdoms of eukaryotes and is largely associated with heterochromatic regions undergoing transcriptional gene silencing. In the model plant *Arabidopsis thaliana*, at least three methylation pathways exist and each is associated with a specific methyltransferase. Methyltransferase 1 (MET1) is a homologue of mammalian DNA methyltransferase 1 (DNMT1) and maintains methylation in the CG dinucleotide context. Chromomethylase 3 (CMT3) is a plant-specific methyltransferase that preferentially deposits the methyl mark in CHG contexts (where H is adenine, thymine or cytosine). Finally, the mammalian DNMT3 homologue DRM2 (domains rearranged methyltransferase 2) performs *de novo* DNA methylation, and maintains CHH or asymmetrical methylation through a small interfering RNA (siRNA)-driven signal in a process known as RNA-directed DNA methylation (RdDM; Law & Jacobsen, 2010). At some loci, CMT3 and DRM2 act redundantly to control the maintenance of both CHG and CHH methylation, but DRM2 alone is responsible for *de novo* DNA methylation (Cao & Jacobsen, 2002a; Chan *et al*, 2004).

Methylation patterns are correlated with specific histone modification signatures. For example, genome-wide studies in *Arabidopsis* have shown that histone 3 Lys9 dimethylation (H3K9m2) is a histone mark that often occurs with CHG methylation and endogenous clusters of siRNAs (Bernatavichute *et al*, 2008). H3K9m2 directed by the Kryptonite (KYP), SU (VAR) 3–9 homologue (SUVH) 5 and 6 histone methyltransferases is required for the maintenance of CHG DNA methylation (Jackson *et al*, 2002; Malagnac *et al*, 2002; Ebbs & Bender, 2006), probably through direct targeting of CMT3 (Lindroth *et al*, 2004). Conversely, histone 3 Lys4 mono/di/trimethylation (H3K4m1/2/3) is strongly negatively correlated with DNA methylation at nongenic silent loci (Zhang *et al*, 2009).

¹Department of Molecular, Cell, and Developmental Biology, University of California at Los Angeles, Terasaki Life Sciences Building, 610 Charles Young Drive East, Los Angeles, California 90095-723905, USA

²Institute for Genetics, Geb. 26.03, U1, R11, Heinrich-Heine-University Düsseldorf, Universitätsstrasse 1, 40255 Düsseldorf, Germany

³Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, California, USA

*These authors contributed equally to this work

+Corresponding author. Tel: +1 310 825 0182; Fax: +1 310 206 3987;

E-mail: jacobsen@ucla.edu

Received 5 May 2010; revised 5 September 2010; accepted 6 September 2010; published online 5 November 2010

The discovery in mammals of two classes of enzyme that are able to demethylate histones—lysine-specific demethylase 1 (LSD1; Shi *et al*, 2004) and Jumonji-C (JmjC) domain-containing proteins (Klose *et al*, 2006)—revealed that active removal of methyl marks from histones is necessary for proper epigenetic regulation. Two plant homologues of the mammalian histone demethylase LSD1—LSD1-LIKE 1 (LDL1) and 2 (LDL2)—are required for H3K4 demethylation at the *FLC* and *FWA* loci (Jiang *et al*, 2007). Although *FLC* is not a DNA-methylated gene, *FWA* transcription is controlled by DNA methylation at the tandem repeats in its 5'-untranslated region (5'-UTR), and *FWA* hypomethylation results in ectopic expression and a late-flowering phenotype (Soppe *et al*, 2000). Interestingly, *ldl1 ldl2* double mutants flower late, and molecular analysis showed hypomethylation at *FWA*. These data suggest that persistent H3K4 demethylation is required to maintain DNA methylation at some loci in the genome. To gain further insight into the relationship between active histone demethylation and DNA methylation at silent loci, we compiled a collection of homozygous transfer DNA insertion mutants in genes containing JmjC domains in *Arabidopsis*. We show that JM14 is required to maintain full levels of non-CG methylation at sites controlled by DRM2. We also found that the loss of non-CG methylation in *jmj14* mutants corresponded with increases in H3K4m3 marks, suggesting that JM14 targets DNA-methylated loci. Interestingly, *jmj14* mutants had no effect on DRM2-mediated establishment of methylation of an incoming *FWA* transgene, which is in contrast to all other mutants that were tested in the DRM2 pathway (Chan *et al*, 2004; Johnson *et al*, 2008; Ausin *et al*, 2009; Law & Jacobsen, 2010). These results suggest that establishment and maintenance of methylation mediated by DRM2 can be differentially regulated, and that JM14 has a specific role in the maintenance of RdDM.

RESULTS

jmj14 mutations affect non-CG maintenance methylation

Arabidopsis contains 21 genes with domains homologous to JmjC histone demethylases (Lu *et al*, 2008; Hong *et al*, 2009). To examine potential effects on DNA methylation, we analysed 17 JmjC mutants for which null alleles were available, at the *medea*-intergenic subtelomeric repeat (*MEA-ISR*) locus by using Southern blotting (supplementary Table S1 online). The *MEA-ISR* is a set of seven tandem repeats downstream from the *medea* (*MEA*) gene. Both MET1 (CG methylation) and DRM2 (CHG and CHH methylations) maintain DNA methylation at *MEA-ISR*, and hypomethylation phenotypes can be observed after digestion with the methylation-sensitive enzyme *MspI* (Cao & Jacobsen, 2002a). By Southern blot analysis, we were able to observe a consistent reduction of *MEA-ISR* methylation in two null alleles of *jmj14* (Fig 1A). JM14—also referred to as JM14 and putative lysine demethylase 7B (PKDM7B)—is the protein encoded by At4g20400 (Lu *et al*, 2008). To confirm the *jmj14* methylation defect, we performed bisulphite sequencing at the *MEA-ISR* locus (Fig 1B). Data from this analysis showed a reduction in non-CG methylation, but CG methylation was unchanged compared with the wild-type control. This indicates that the *jmj14* mutation interacts with the DRM2 pathway, but not the MET1 pathway.

To confirm the genetic interaction of JM14 with the DRM2 pathway, we examined the effect of the mutation on other RdDM targets. Analysis of the methylation state of the 5'UTR of *FWA* was

performed by using bisulphite sequencing. *FWA*, similarly to *MEA-ISR*, is mainly targeted by MET1 and DRM2 (Cao & Jacobsen, 2002a). Similarly to the bisulphite data at *MEA-ISR*, we observed a reduction in non-CG methylation but no effect at CG sites at *FWA* (Fig 1C). Finally, to examine DRM2-dependent methylation at the transposable element *AtSN1*, DNA from both wild type and *jmj14* mutants was digested with the restriction endonuclease *HaeIII* that cleaves GGCC sequences, but not GGmCC. Digested DNA was analysed by real-time quantitative PCR using primers that amplify a region spanning three asymmetrically methylated restriction sites (Fig 1D). Relative quantification of uncut DNA in the digested samples showed a significant decrease in CHH methylation in *jmj14* mutants compared with wild type, although not to the same extent as in *drm2*. To examine whether the *jmj14* mutant defects were specific to the DRM2 pathway, we also analysed the methylation state of *Ta3*—a single-copy transposable element that is methylated by CMT3 but not DRM2 (Cao & Jacobsen, 2002a). We observed no effect on methylation in any context for *jmj14* compared with the wild-type control (Fig 1E). This indicates that JM14 acts primarily in the DRM2 pathway.

jmj14 affects chromatin at RdDM target loci

To examine the localization of JM14, we created a carboxy-terminal epitope-tagged (9 × Myc) *JMJ14* transgene driven by the endogenous *JMJ14* promoter and showed that this transgene fully complements the early-flowering phenotype (Jeong *et al*, 2009) of the *jmj14* mutant (Fig 2A,B). Immunostaining for the Myc epitope revealed strong nuclear staining, consistent with the function of JM14 as a histone demethylase. Interestingly, we observed a specific pattern in which staining was uniformly present throughout the nucleoplasm but not in the nucleolus and the chromocentres (areas of dense heterochromatin that are highly enriched for H3K9m2; Fig 2C). This pattern is similar to that found for DRM2 (Li *et al*, 2006), consistent with the hypothesis that JM14 acts in the DRM2 pathway.

Phylogenetic analyses have shown that the JM14 sequence is closest to human lysine demethylase 5/Jumonji/Arid-domain containing protein 1 family histone demethylases (Lu *et al*, 2008) that are able to specifically demethylate H3K4m1, H3K4m2 and H3K4m3 (Christensen *et al*, 2007; Iwase *et al*, 2007; Lee *et al*, 2007; Seward *et al*, 2007). A recombinant JM14 was shown to efficiently demethylate H3K4m3 *in vitro* and to a lesser extent H3K4m2 and H3K4m1 (Jeong *et al*, 2009; Lu *et al*, 2010; Yang *et al*, 2010). This H3K4 demethylase activity was confirmed by an *in vivo* assay in *Nicotiana benthamiana* in which overexpression of *JMJ14* correlated with a strong reduction in H3K4m3 and H3K4m2 marks (Lu *et al*, 2010). Finally, in *Arabidopsis*, JM14 was shown to demethylate H3K4m3 and H3K4m2 at two loci involved in floral transition and not controlled by DNA methylation (Jeong *et al*, 2009; Yang *et al*, 2010).

This suggests that the defect in DNA methylation at non-CG sites was caused by an increase in H3K4 methylation in *jmj14* mutants. To confirm this hypothesis, we used chromatin immunoprecipitation (ChIP) analysis to assess the levels of H3K4m2 and H3K4m3 at silent loci analysed for DNA methylation in wild type and *jmj14*. We observed a consistent increase in H3K4m3 marks at *AtSN1*, *FWA* and *MEA-ISR* (Fig 3). The extent of this increase was similar to that which has been found in *jmj14* mutants at the floral transition loci *flowering locus T* (*FT*) and *twin sister of FT*

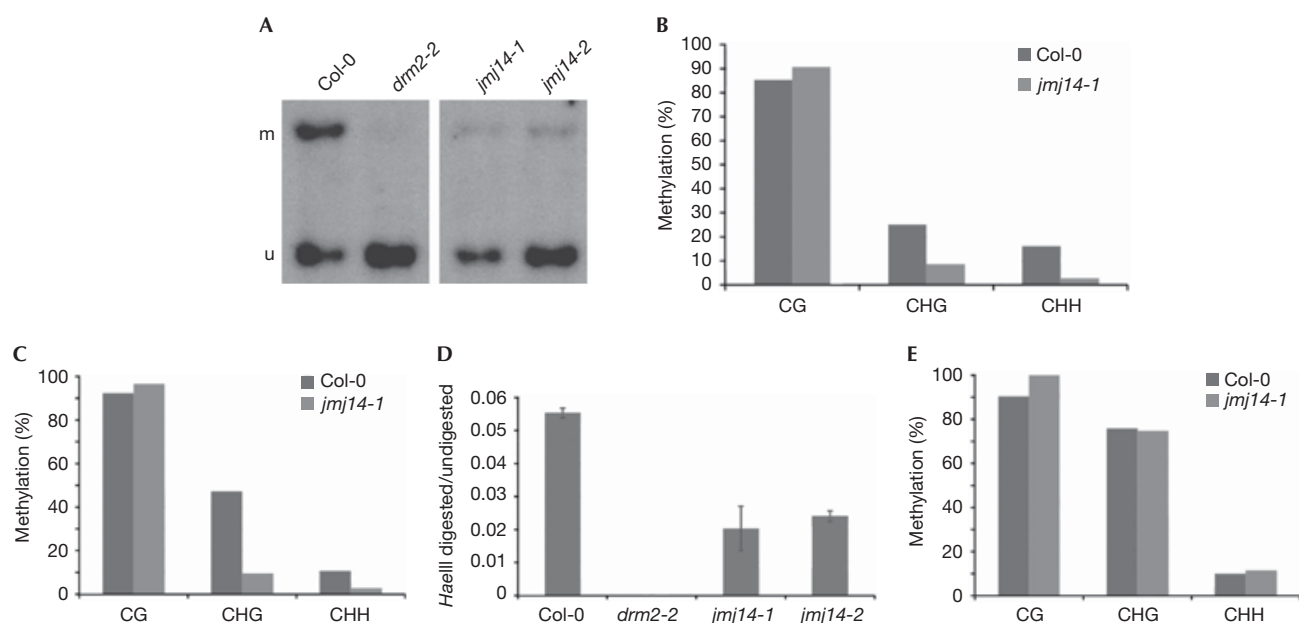


Fig 1 | DNA methylation analysis of *Jumonji 14* mutants. (A) *MEA-ISR* Southern blot. Genomic DNA was digested with the non-CG methylation-sensitive restriction endonuclease *MspI*, and probed for *MEA-ISR*. The high-molecular-weight band (m) represents methylated DNA and the low-molecular-weight band (u) represents unmethylated DNA. Two alleles of *jmj14* show a methylation phenotype intermediate between wild type and the *drm2* mutant. (B) *MEA-ISR* bisulphite sequencing. Genomic DNA was treated with sodium bisulphite and amplified with primers specific for *MEA-ISR*. Sequencing shows an effect at non-CG sites compared with wild type, but not in the CG context. (C) *FWA* endogene bisulphite sequencing. The *FWA* locus has a similar pattern to *MEA-ISR* in the *jmj14-1* mutant. (D) *AtSN1* *HaeIII* Chop-qPCR. Genomic DNA was digested with non-CG methylation-sensitive restriction endonuclease *HaeIII*. Digested DNA was quantified by using real-time qPCR with primers specific for a region of *AtSN1* spanning three restriction sites, and the signal was normalized to an undigested control. Two *jmj14* alleles had significantly more digestion compared with the wild-type control, thus there was less methylation. (E) *Ta3* bisulphite sequencing. The methylation state of *Ta3* shows no discernible defect in the *jmj14* mutant compared with wild type. *jmj14*, *Jumonji 14*; *MEA-ISR*, *medea*-intergenic subtelomeric repeats; qPCR, quantitative PCR.

(*TSF*; Jeong *et al*, 2009; Yang *et al*, 2010). We also saw a small but significant increase in H3K4m2 marks at the *FWA* locus, but not at *AtSN1* or *MEA-ISR* (Fig 3). The minor effects on H3K4m2 might be due to the redundant activity of other demethylases, such as LDL1 and LDL2 (Jiang *et al*, 2007). Overall, these results show that JM14 might directly target silent chromatin, and suggest that the active removal of H3K4 methyl marks at silent loci might be necessary for DRM2 to maintain proper DNA methylation patterns.

jmj14 does not affect *de novo* DNA methylation

All components of the RdDM machinery that have been tested thus far have been shown to be required both for DRM2-dependent non-CG maintenance DNA methylation at *MEA-ISR* and other loci, and for establishment of methylation in all sequence contexts on previously unmethylated sequences—or *de novo* methylation—of an incoming transgene (Chan *et al*, 2004; Johnson *et al*, 2008; Ausin *et al*, 2009; Law & Jacobsen, 2010). When *FWA* is introduced into wild-type plants, siRNAs are able to target the repeats in the 5'UTR and the incoming transgene becomes methylated, and thus silenced. However, in RdDM mutants, the transgene remains unmethylated in all sequence contexts and is expressed (Cao & Jacobsen, 2002b; Chan *et al*, 2004). As we had observed non-CG maintenance methylation phenotypes at known RdDM targets in *jmj14*, we used the *FWA* transgene system to test for a function of JM14 in *de novo*

methylation. Ectopic *FWA* expression leads to a late-flowering phenotype that gives a quantitative readout of the methylation establishment phenotype.

The *jmj14* mutant flowers earlier than the wild-type plants, which has previously been shown to be due to de-repression of *FT* (Fig 4A; Jeong *et al*, 2009; Lu *et al*, 2010; Yang *et al*, 2010). Surprisingly, *FWA*-transformed *jmj14* continued to flower earlier than wild-type control plants (Fig 4A). We note that other mutants with weak RdDM phenotypes—such as *dicer-like 3* (*dcl3*) which shows only partial losses of *MEA-ISR* methylation (equivalent to those of *jmj14*)—do show substantial effects on *FWA de novo* DNA methylation establishment, and thus flower later (Henderson *et al*, 2006). These results suggest that the *jmj14* mutation does not affect *FWA de novo* DNA methylation.

To confirm these findings, we analysed the methylation state of the newly introduced *FWA* transgene by using bisulphite sequencing (Fig 4B). We observed in the *FWA* transgene that CG methylation levels of the *jmj14* mutant were comparable with those of wild type; however, there was a significant decrease in non-CG methylation. By contrast, the *dcl3* mutant shows substantially less *de novo* methylation than wild type in all three sequence contexts, even though it exhibited a similar non-CG maintenance phenotype (Henderson *et al*, 2006). These results show that the CG DNA methylation that is primarily responsible for silencing *FWA* is fully established in *jmj14*. Once CG methylation is established, it is maintained by the MET1 pathway

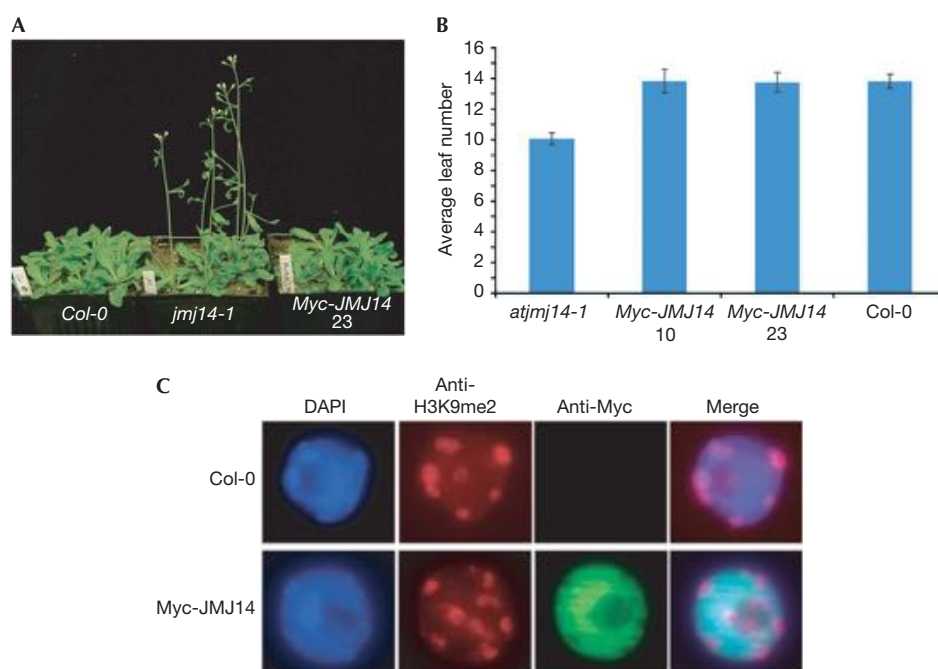


Fig 2 | Analysis of complementing Myc-tagged lines. (A) Myc-tagged JMJ14 constructs complement the early-flowering phenotype observed in the *jmj14-1* mutant background. (B) Flowering-time assay. Quantification of complementation for tagged JMJ14 lines. Note: Line 10 was used for immunofluorescence assay. (C) Immunolocalization of epitope-tagged JMJ14. A transgenic line expressing Myc-tagged complementing JMJ14 under its endogenous promoter was analysed by using fluorescent microscopy. JMJ14 is localized in the nucleus, but is depleted from the chromocentres (marked by histone 3 Lys 9 dimethylation enrichment and dense DAPI staining). DAPI, 4',6-diamidino-2-phenylindole; JMJ14, Jumonji 14.

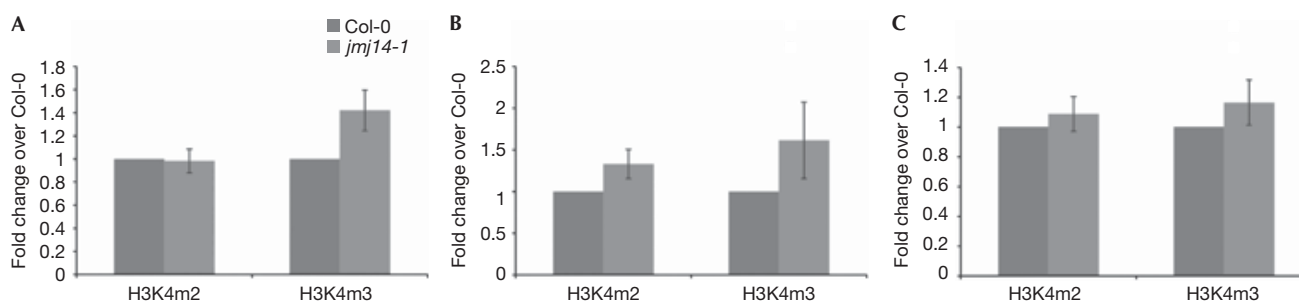


Fig 3 | Analysis of H3K4m2 and H3K4m3 state at RdDM targets by using chromatin immunoprecipitation. The immunoprecipitated DNA corresponding to (A) *AtSN1*, (B) *FWA* and (C) *MEA-ISR* was quantified by real-time PCR and normalized to an internal control—we used an intergenic region upstream from the isocitrate dehydrogenase gene—intergenic region (*ICDH-IGR*) and unlikely to be targeted by JMJ14. The fold enrichment in *jmj14-1* over wild type is shown at each locus (the wild type values were set to one). The values are the average ratio obtained from three independent ChIP experiments $\pm$ s.e. ChIP, chromatin immunoprecipitation; H3K4m2/m3, histone 3 Lys 4 dimethylation/trimethylation; ICDH, isocitrate dehydrogenase; JMJ14, Jumonji 14; RdDM, RNA-directed DNA methylation.

independently of DRM2, whereas DRM2 maintains non-CG marks. Consistent with a function in DRM2-mediated maintenance of non-CG methylation, and similarly to the *FWA* endogene (Fig 1C), we observed that maintenance of CHG and CHH methylation at the *FWA* transgene was reduced in the *jmj14* mutant (Fig 4B).

DISCUSSION

JMJ14 is required for the maintenance of DRM2-mediated non-CG DNA methylation. Consistent with our findings, a recent study

described the identification of JMJ14 through a forward-genetic screen for mutants impaired in hairpin-induced transcriptional silencing of the *phytoene desaturase* endogene (Searle et al, 2010).

We observed a moderate but consistent increase in H3K4m3 levels at RdDM targets analysed in *jmj14*, suggesting that active demethylation of H3K4 is required for proper DRM2-pathway function, perhaps due to competition between the active H3K4 methylation mark and repressive marks such as DNA methylation (Fig 5). The fact that two enzyme families—JmjC domain and LSD-like (Jiang et al, 2007)—have functions in the demethylation

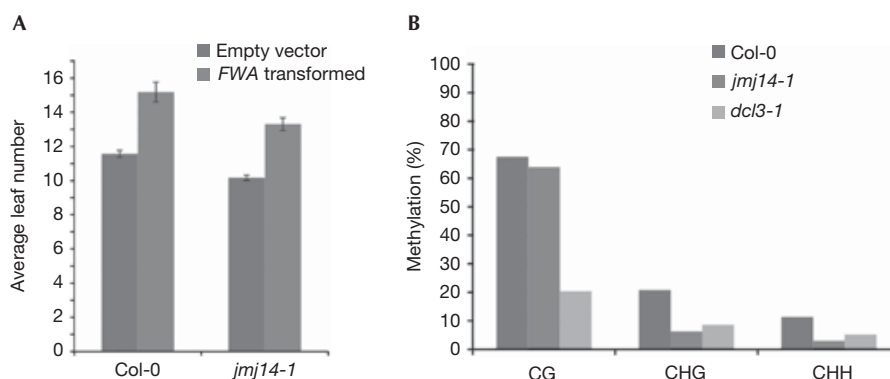


Fig 4 | The *de novo* DNA methylation analysis. (A) *FWA* flowering-time assay. Total leaf number on flowering was assessed for wild-type Col-0 and *jmj14-1* for both *FWA* and empty-vector transformants. (B) *FWA* transgene bisulphite sequencing. *jmj14-1* transformants had a minimal effect on CG methylation compared with *dcl3-1*. The effect on non-CG might be due to a maintenance defect after the initial methylation has been established. *jmj14-1*, *Jumonji 14-1*.

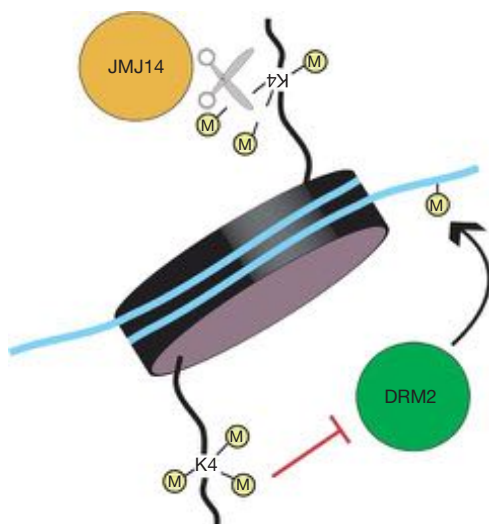


Fig 5 | Model for the role of Jumonji 14 in DRM2-mediated maintenance methylation. It is proposed that histone 3 Lys 4 methylation inhibits DRM2 pathway components. Active demethylation of the residue is needed for complete DRM2 maintenance activity. DRM2, domains rearranged methyltransferase 2; JM14, Jumonji 14.

of H3K4 methyl marks at silent loci/RdDM targets underlies the importance of removing those marks for the maintenance of proper DNA methylation patterns.

Interestingly, *jmj14* mutants showed no effect on DRM2-mediated *de novo* methylation of an incoming *FWA* transgene. This is in contrast to all other mutants tested in the DRM2 pathway: *nrpd1*, *nrpe1*, *dcl3*, *rdr2*, *ago4*, *drd1*, *suvh2*, *dms3* and *idn2* (Chan et al, 2004; Johnson et al, 2008; Ausin et al, 2009; Law & Jacobsen, 2010). This indicates that JM14 is required to maintain non-CG methylation patterns, but is not involved in the initial targeting of DNA methylation. This is an interesting finding as it implies that the maintenance activity of DRM2 can be mechanistically distinguished from its *de novo* methylation establishment activity, suggesting that during the maintenance

phase there is another level of regulation of DRM2 activity by histones. The relationship between DRM2 activity and H3K4 methylation status is also interesting in the light of activity mechanisms of the mammalian DRM2 homologue DNMT3A. DNMT3A is in part recruited to silent loci through interaction with a related protein (DNMT3L) that can bind to H3 specifically when Lys 4 is unmethylated (Jia et al, 2007; Ooi et al, 2007). Future analyses might determine how H3K4 methyl marks antagonize the DRM2 pathway in *Arabidopsis*.

METHODS

Plant materials. We used the following *Arabidopsis* strains: wild-type Col-0 and the recessive alleles *dcl3-1* and *drm2-2* in the Col-0 background. The list of alleles of JmjC mutants tested is presented in supplementary Table S1 online.

Southern blotting and bisulphite analysis. See the supplementary information online for details.

HaeIII Chop-qPCR. DNA from young flowers was extracted using a standard Cetyl trimethyl ammonium bromide protocol. A total of 200 ng of genomic DNA was digested overnight at 37 °C with *HaeIII* side-by-side with samples containing buffer and no enzyme (undigested). Quantitative real-time PCR validation of uncut DNA after *HaeIII* digestion was performed using the Bio-Rad Synergy Brands Green SuperMix on a MX3000 Stratagene cycler. The PCR parameters are as follows: one cycle of 10 min at 95 °C, 40 cycles of 30 s at 95 °C, 1 min at 55 °C and 1 min at 72 °C. PCR primers sequences are listed in supplementary Table S2 online.

FWA transformation. See the supplementary information online for details.

Flowering-time analysis. We measured flowering time as the total number of leaves (rosette and cauline leaves) developed by a plant.

Generation of epitope-tagged complementing lines. Epitope-tagged protein constructs were made by cloning 1.6 kb of genomic DNA upstream from the *JM14* open reading frame and including the entire open reading frame into pENTR. A 9 × Myc epitope tag was introduced at the C-terminus. The tagged construct was then recombined into a modified pDEST vector and introduced into *Agrobacterium* strain AGL1.

Protein immunofluorescence analysis. We prepared nuclei for immunofluorescent imaging as described in Li *et al*, 2006. See supplementary information online for more details.

ChIP. The ChIP experiments were performed as previously described (Bernatavichute *et al*, 2008; Johnson *et al*, 2008; Zhang *et al*, 2009). See supplementary information online for more details.

Supplementary information is available at EMBO reports online (<http://www.emboreports.org>).

ACKNOWLEDGEMENTS

We thank M. Mobilia for help with fluorescent microscopy and A. Hooper for genotyping. Jacobsen lab research was funded by United States National Institutes of Health grant GM60398. A.D. was funded by a European Molecular Biology Organization fellowship ALTF 1161-2007 and Human Frontier Science Program fellowship LT 0056-2008L. M.V.C.G. was funded by United States Public Health Service National Research Service Award GM07104. I.A. was funded by the Ministerio de Educacion y Ciencia. R.W.Y.L. was funded by a Van Trees Scholarship under the University of California at Los Angeles Undergraduate Research Scholars Program. S.E.J. is an investigator at the Howard Hughes Medical Institute.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- Ausin I, Mockler TC, Chory J, Jacobsen SE (2009) IDN1 and IDN2 are required for *de novo* DNA methylation in *Arabidopsis thaliana*. *Nat Struct Mol Biol* **16**: 1325–1327
- Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS ONE* **3**: e3156
- Cao X, Jacobsen SE (2002a) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc Natl Acad Sci USA* **99** (Suppl 4): 16491–16498
- Cao X, Jacobsen SE (2002b) Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr Biol* **12**: 1138–1144
- Chan SW, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE (2004) RNA silencing genes control *de novo* DNA methylation. *Science* **303**: 1336
- Christensen J, Agger K, Cloos PA, Pasini D, Rose S, Sennels L, Rappilber J, Hansen KH, Salcini AE, Helin K (2007) RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. *Cell* **128**: 1063–1076
- Ebbs ML, Bender J (2006) Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. *Plant Cell* **18**: 1166–1176
- Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE (2006) Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat Genet* **38**: 721–725
- Hong EH, Jeong YM, Ryu JY, Amasino RM, Noh B, Noh YS (2009) Temporal and spatial expression patterns of nine *Arabidopsis* genes encoding Jumonji C-domain proteins. *Mol Cell* **27**: 481–490
- Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L, Huarte M, Qi HH, Whetstone JR, Bonni A, Roberts TM, Shi Y (2007) The X-linked mental retardation gene *SMCX/JARID1C* defines a family of histone H3 lysine 4 demethylases. *Cell* **128**: 1077–1088
- Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**: 556–560
- Jeong JH, Song HR, Ko JH, Jeong YM, Kwon YE, Seol JH, Amasino RM, Noh B, Noh YS (2009) Repression of FLOWERING LOCUS T chromatin by functionally redundant histone H3 lysine 4 demethylases in *Arabidopsis*. *PLoS ONE* **4**: e8033
- Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. *Nature* **449**: 248–251
- Jiang D, Yang W, He Y, Amasino RM (2007) *Arabidopsis* relatives of the human lysine-specific demethylase1 repress the expression of FWA and FLOWERING LOCUS C and thus promote the floral transition. *Plant Cell* **19**: 2975–2987
- Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE (2008) SRA-domain proteins required for DRM2-mediated *de novo* DNA methylation. *PLoS Genet* **4**: e1000280
- Klose RJ, Kallin EM, Zhang Y (2006) JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* **7**: 715–727
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**: 204–220
- Lee MG, Norman J, Shilatifard A, Shiekhattar R (2007) Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBL, a polycomb-like protein. *Cell* **128**: 877–887
- Li CF, Pontes O, El-Shami M, Henderson IR, Bernatavichute YV, Chan SW, Lagrange T, Pikaard CS, Jacobsen SE (2006) An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126**: 93–106
- Lindroth AM *et al* (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J* **23**: 4286–4296
- Lu F, Li G, Cui X, Liu C, Wang XJ, Cao X (2008) Comparative analysis of JmjC domain-containing proteins reveals the potential histone demethylases in *Arabidopsis* and rice. *J Integr Plant Biol* **50**: 886–896
- Lu F, Cui X, Zhang S, Liu C, Cao X (2010) JM14 is an H3K4 demethylase regulating flowering time in *Arabidopsis*. *Cell Res* **20**: 387–390
- Malagnac F, Bartee L, Bender J (2002) An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J* **21**: 6842–6852
- Ooi SK *et al* (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448**: 714–717
- Searle IR, Pontes O, Melnyk CW, Smith LM, Baulcombe DC (2010) JM14, a JmjC domain protein, is required for RNA silencing and cell-to-cell movement of an RNA silencing signal in *Arabidopsis*. *Genes Dev* **24**: 986–991
- Seward DJ, Cubberley G, Kim S, Schonewald M, Zhang L, Tripet B, Bentley DL (2007) Demethylation of trimethylated histone H3 Lys4 *in vivo* by JARID1 JmjC proteins. *Nat Struct Mol Biol* **14**: 240–242
- Shi Y, Lan F, Matson C, Mulligan P, Whetstone JR, Cole PA, Casero RA (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**: 941–953
- Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJ (2000) The late flowering phenotype of FWA mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell* **6**: 791–802
- Yang W, Jiang D, Jiang J, He Y (2010) A plant-specific histone H3 lysine 4 demethylase represses the floral transition in *Arabidopsis*. *Plant J* **62**: 663–673
- Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE (2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol* **10**: R62



EMBO reports is published by Nature Publishing Group on behalf of European Molecular Biology Organization. This article is licensed under a Creative Commons Attribution Noncommercial Share Alike 3.0 Unported License [<http://creativecommons.org/licenses/by-nc-sa/3.0/>]

VIII. List of figures.

Figure 1. Chromatin states of *Arabidopsis thaliana* nuclei at interphase. (page 5)

Figure 2. The concept of “epigenetic landscape” as depicted by Conrad W. Waddington in 1957. (page 6)

Figure 3. Artwork of chromatin organization from the double helix DNA molecule to the highly-condensed chromosome. (page 8)

Figure 4. DNMTs involved in the initiation (de novo) and maintenance of DNA methylation in *A. thaliana*. (page 11)

Figure 5. Epigenetic writers, erasers and readers. (page 13)

Figure 6. Intricate features of the “epigenetic mille-feuille”, composed of epigenetic pathways and players that have been described in this introduction. (page 15)

Figure 7. Traditional view and emerging scenarios describing TF/CRE interaction. (page 17)

Figure 8. The *SDC::GFP* reporter gene allowed the identification of new epigenetic players. (page 22)

Figure 9. Transgenerational analyses of *ATCOPIA28::GFP* expression. (page 24)

Figure 10. Characterization of four *ATCOPIA28::GFP* epialleles obtained in WT and *ddc* genetic backgrounds. (page 25)

Table 1. Summary of WT and *ddc* mutant populations that have been mapped. (page 26)

Figure 11. PMD-C protein organization and subnuclear localization. (page 28)

Figure 12. *MORC1* is downregulated in *main* and *mail1* mutants. (page 29)

Figure 13. A. List of outstanding misregulated genes in the *main-2* and *mail1-1* mutants, including *FLC* and *MORC1*. (page 31)

Figure 14. Preliminary studies of MAIL2, MAIL3 and PMD-B clade PMDs. (page 36)

Figure 15. Chromosomal distribution of *PMD* genes within the *A. thaliana* genome. (page 37)

Figure 16. Expression pattern of three *SIPMDs* and two *SIPP7Ls* at different stages of tomato plant development. (page 39)

Figure 17. Evolution and distribution of PMDs among angiosperms. (page 42)

Figure 18. Hypothetical model in which the PMD and PPP domains constitute a functional protein module. (page 43)

IX. Abbreviations.

5-mC	DNA cytosine methylation
ABI	ABSCISIC ACID INSENSITIVE
ALP1	ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN1
AP	APETALA
ARF	Auxin response factor
ATX	ARABIDOPSIS TRITHORAX-LIKE
ATXR	ARABIDOPSIS TRITHORAX-RELATED PROTEIN
bHLH	Helix-loop-helix factor
BS-seq	Bisulfite DNA sequencing
bZIP	Leucine zipper factor
Cas	CRISPR-associated
CC	Chromocenter
CENP	Centromere binding protein
CHD	Chromodomain-helicase-DNA binding
ChIP-qPCR	Chromatin-immunoprecipitation coupled to quantitative PCR
ChIP-seq	chromatin-immunoprecipitation coupled to DNasequencing
circRNA	circular RNA
CLF	CURLY LEAF
CMT	CHROMOMETHYLASE
Cnr	Colorless nonripening
Co-IP	Co-immunoprecipitation
CRE or CRM	cis-regulatory element or module
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAP-seq	DNA affinity purification sequencing
DBD	DNA-binding domain
dCas	nuclease-deficient CRISPR-associated
DCL	DICER-LIKE
DDM1	DECREASE IN DNA METHYLATION 1
DDR	DNA damage response
DME	TRANSCRIPTIONAL ACTIVATOR DEMETER
DML	DEMETER-LIKE
DNMT	DNA methyltransferase
DRM	DOMAINS REARRANGED METHYLTRANSFERASE
eccDNA	extrachromosomal circular DNA
EMS	Ethyl methanesulfonate
EMSA	Electrophoretic mobility shift assay
EpiRL	Epigenetic recombinant inbred line
ERF	Ethylene-responsive element binding factor
ERV	Endogenous retroviruse
ESC	Embryonic stem cell
ETE protein	Exapted TE protein

FAR1	Far-red impaired response 1
FLC	FLOWERING LOCUS C
FRS	FAR1-related sequences
FWA	FLOWERING WAGENINGEN
Gal4BD	Gal4 binding domain
GbM	Gene body methylation
GHKL	Gyrase, HSP90, histidine kinase, MutL
GO	Gene Ontology
gRNA	guide RNA
GUS	Beta-glucuronidase
HAT	Histone acetyl transferase
HDAC, HDA	Histone deacetylases
HDP	Harbinger transposon-derived protein
HKDM	Histone lysine demethylase
HKMT	Histone lysine methyltransferase
HMG	High mobility group
INO80	Inositol requiring 80
IP-MS	Immunoprecipitation followed by mass spectrometry
ish	in situ hybridization
ISWI	Imitation switch
JmjC	Jumonji C
KMT	Lysine methyltransferase
KYP	KRYPTONITE (a.k.a SUVH4)
LINE	Long interspersed nuclear elements
lncRNA	long noncoding RNA
LTR	Long terminal repeat
MAIL	MAIN-LIKE
MAIN	MAINTENANCE OF MERISTEMS
MET1	DNA METHYLTRANSFERASE 1
miRNA	microRNA
MOM1	MORPHEUS' MOLECULE 1
MORC	Microrchidia
MTHFD1	Methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase
MUG	MULE transposase-derived MUSTANG
MULE	Mutator-like element
NAC	NAM (no apical meristem), ATAF1 and -2, and CUC2 (cup-shaped cotyledon)
ncRNA	noncoding RNA
NLS	Nuclear localization signal
nor	nonripening

OCT4	Octamer-binding transcription factor 4
PcG	Polycomb
PHD	Plant homeodomain
PIAS	PROTEIN INHIBITOR OF ACTIVATED STAT
piRNA	PIWI-interacting RNA
PMD	Plant Mobile Domain
PP7L	PP7-LIKE
PPP	Phosphoprotein phosphatase
PRC	Polycomb repressive complex
PRMT	Protein arginine methyltransferase
PTGS	Post-transcriptional gene silencing
PTM	Post-translational modification
RdDM	RNA-directed DNA methylation
RDR or RdRP	RNA-dependent RNA polymerase
RHD	Rel homology domain
rin	ripening-inhibitor
RING	Really Interesting New Gene
RNA Pol	RNA Polymerase
RNA-seq	RNA sequencing
ROS1	REPRESSOR OF SILENCING 1
RT-qPCR	Quantitative reverse transcription PCR
SAM	Shoot apical meristem
SBP	SQUAMOSA-promoter binding protein
SDC	SUPPRESSOR OF drm1 drm2 cmt3
SDG	SET DOMAIN GROUP
SELEX	Systematic evolution of ligands by exponential enrichment
SEP3	SEPALLATA 3
siRNA	ihort interfering RNA
SIDML2	DEMETER-like DNA demethylase2
SIPMD	<i>S. lycopersicum</i> PMD
SIPP7L	<i>S. lycopersicum</i> PP7L
SOX2	Sex determining region Y-box 2
SRA	SET- or RING-associated
SUVH	SU(VAR)3-9 HOMOLOGOUS
SUVR	SU(VAR)3-9 RELATED
SWI/SNF	Switch/sucrose-non-fermenting
SWN	SWINGER
TE	Transposable element
TERT	Telomerase reverse transcriptase
TET1	Ten-eleven translocation methylcytosine dioxygenase 1
TF	Transcription factor
TGS	Transcriptional gene silencing

tsRNA	tRNA-derived small RNA
TSS	Transcriptional start site
UBP26	Ubiquitin-specific protease 26
V(D)J	Variable diversity joining recombination system
VP	VIVIPAROUS
WGD	Whole genome duplication
Y1H	Yeast one hybrid
Y2H	yeast two hybrid
Znf	Zinc finger

Résumé. Les mécanismes épigénétiques tels que la méthylation de l'ADN ou les modifications des protéines histones jouent un rôle essentiel au sein de la cellule. Ils sont impliqués dans de nombreux processus, incluant la régulation de l'expression des gènes et la répression des éléments d'ADN répétés tels que les éléments transposables (TEs). Les TEs sont des éléments génétiques dit égoïstes et hautement mutagènes qui doivent être réprimés afin de maintenir l'intégrité de la cellule. Cependant, les TEs peuvent aussi avoir un effet positif sur le génome de l'hôte, contribuant à son dynamisme et évolution. De plus, les TEs sont une importante source d'innovation génétique, comme par exemple dans le processus de domestication de gènes de TEs, aussi connu sous le nom d'« exaptation de gènes ». Plusieurs gènes exaptés de TEs (gènes ETE) ont été décrits chez les eucaryotes et les procaryotes, jouant des rôles essentiels dans de nombreux processus cellulaires fondamentaux. Chez *Arabidopsis thaliana*, des gènes ETE ont été impliqués dans la régulation de l'expression des gènes, agissant de manière antagoniste ou coopérative avec les voies épigénétiques. Cependant, plusieurs gènes ETE restent non étudiés.

Au sein de l'équipe « Mécanismes Épigénétiques et Architecture de la Chromatine », nous combinons des approches de génétique directe et inverse afin d'identifier de nouveaux acteurs épigénétiques et d'étudier le rôle des gènes ETE dans les mécanismes chromatinien. Nous avons récemment identifié le gène « Plant Mobile Domain » (PMD) « *MAINTENANCE OF MERISTEMS* » (*MAIN*) comme étant requis pour l'expression de gènes et la répression de TEs. Le PMD est un domaine protéique de fonction inconnue qui est exclusivement trouvé chez les angiospermes, principalement associé aux TEs. Il a été suggéré qu'au cours de l'évolution, le PMD aurait été domestiqué par les plantes à partir de TEs pour générer des versions géniques (ou ETE) de PMD, tel que *MAIN*. *MAIN* et son plus proche homologue « *MAIN-LIKE 1* » (*MAIL1*) ont été impliqués dans les processus de stabilité génomique, développementaux et de répression de TEs. Notre équipe a récemment trouvé que *MAIN* et *MAIL1* interagissent ensemble, ainsi qu'avec la phosphoprotéine phosphatase (PPP) « *PP7-LIKE* » (*PP7L*), et que les trois protéines sont similairement requises pour la bonne expression de plusieurs gènes et la répression de TEs. De plus, des analyses phylogénétiques des protéines PMD et PPP de type PP7 au sein des Eudicotylédones suggèrent que ces deux domaines protéiques pourraient constituer un module protéique fonctionnel interagissant en *cis* ou en *trans*.

En se basant sur ces résultats, nous développons des approches complémentaires permettant de comprendre le rôle des protéines PMD au cours du développement. En se focalisant d'abord sur le complexe protéique *MAIN/MAIL1/PP7L*, nous voulons décortiquer les mécanismes impliquant ces protéines dans la régulation de l'expression de gènes et la répression des TEs. Nous étudierons aussi le rôle d'autres protéines PMD d'*Arabidopsis* durant le développement. En particulier, nous voulons comprendre le rôle de *MAIN-LIKE 2* (*MAIL2*), le deuxième plus proche homologue de *MAIN*, dans la régulation de l'expression génique. Ceci est d'autant plus pertinent du fait que l'expression de *MAIL2* est essentielle au bon développement de la plante. Les processus cellulaires impliquant les protéines PMD sont largement inconnus. De plus, ces protéines ont seulement été étudiées chez *Arabidopsis*. Par conséquent, nous avons décidé d'étudier le rôle des PMD durant le processus de maturation du fruit chez l'espèce d'intérêt agronomique *Solanum lycopersicum* (tomate). En parallèle, nous voulons étudier la signification biologique de l'association PMD/TE, et déterminer si le PMD peut être bénéfique pour l'adaptation des TEs.

En conclusion, ce projet de recherche apportera une meilleure compréhension des mécanismes impliquant les protéines PMD et ETE dans les processus chromatinien, afin de décortiquer les interactions complexes entre mécanismes épigénétiques et protéines ETE, mais aussi entre la cellule hôte et les TEs.

Abstract. Epigenetic mechanisms such DNA methylation and histone protein modifications play essential roles in the cell. They are involved in several cellular processes, including regulation of gene expression and silencing of DNA repeats such as transposable elements (TEs). TEs are considered as selfish and highly mutagenic genetic elements that must remain silenced to maintain cell integrity. However, TEs can also positively impact the host genome, contributing its dynamics and evolution. Besides, TEs are an important source of genetic innovation, as exemplified through the process of TE gene domestication, also known as ‘gene exaptation’. Several exapted TE (ETE) genes have been described in eukaryotes as well as in prokaryotes, playing essential roles in fundamental cellular processes. In *Arabidopsis thaliana*, ETE genes have been involved in the regulation of gene expression by either antagonizing or cooperating with epigenetic pathways. However, several ETE genes remain uncharacterized.

In the team ‘Epigenetic Mechanisms and Chromatin Architecture’, we combine forward and reverse genetic approaches to identify new epigenetic players and to study the role of ETE genes in chromatin-related processes. We recently identified the Plant Mobile Domain (PMD) gene *MAINTENANCE OF MERISTEMS (MAIN)* as required for the proper expression of genes and TE silencing. The PMD is a protein domain of unknown function that is exclusively found in the angiosperms, predominantly associated with TEs. It has been suggested that during evolution, the PMD would have been domesticated by the plants from TEs to generate genic (or ETE) PMD versions, such as *MAIN*. *MAIN* and its closest homolog *MAIN-LIKE 1 (MAIL1)* have been involved in genome stability, developmental processes, and TE silencing. In our team, we have recently found that *MAIN* and *MAIL1* proteins interact together, as well as with the phosphoprotein phosphatase (PPP) *PP7-LIKE (PP7L)*, and the three proteins are required for the proper expression of a common set of genes and TE silencing. In addition, phylogenetic analyses of PMD and PP7-type PPP proteins among the Eudicot lineage suggest that these two protein domains may constitute a functional protein module through trans or cis interactions.

Based on these results, we are developing complementary approaches to study the role of PMD proteins during plant development. Focusing first on the *MAIN/MAIL1/PP7L* protein complex, we want to decipher the mechanisms involving these proteins in the regulation of gene expression and TE silencing. We are also planning to study the role of other *Arabidopsis* PMD proteins during plant development. Particularly, we are interested in determining the role of *MAIN-LIKE 2 (MAIL2)*, which is *MAIN* second closest homolog, in the regulation of gene expression. This is especially relevant considering that *MAIL2* expression is essential for plant development. The cellular processes involving genic PMD proteins remain largely unknown. Moreover, these proteins have only been studied in *A. thaliana*. Therefore, we decided to study the role of genic PMD proteins during fruit ripening process in the model fruit-bearing crop *Solanum lycopersicum* (tomato). Besides, we want to investigate the biological significance of PMD/TE association, and determine whether the PMD could be beneficial for TE fitness.

Altogether, this research project will give more insights into the mechanisms involving PMD and other ETE proteins in chromatin-related processes, to eventually decipher the complex interplay between epigenetics mechanisms and ETE proteins, as well as between host cell and TEs.