

MOLECULAR BIOLOGY

A DNA methylation reader complex that enhances gene transcription

C. Jake Harris^{1*}, Marion Scheibe^{2*}, Somsakul Pop Wongpalee^{1†}, Wanlu Liu¹, Evan M. Cornett³, Robert M. Vaughan³, Xueqin Li⁴, Wei Chen⁴, Yan Xue¹, Zhenhui Zhong^{1,5}, Linda Yen¹, William D. Barshop⁶, Shima Rayatpisheh^{6†}, Javier Gallego-Bartolome¹, Martin Groth^{1§}, Zonghua Wang^{5,7}, James A. Wohlschlegel⁶, Jiamu Du⁴, Scott B. Rothbart³, Falk Butter^{2||}, Steven E. Jacobsen^{1,8||}

DNA methylation generally functions as a repressive transcriptional signal, but it is also known to activate gene expression. In either case, the downstream factors remain largely unknown. By using comparative interactomics, we isolated proteins in *Arabidopsis thaliana* that associate with methylated DNA. Two SU(VAR)3-9 homologs, the transcriptional antisilencing factor SUVH1, and SUVH3, were among the methyl reader candidates. SUVH1 and SUVH3 bound methylated DNA in vitro, were associated with euchromatic methylation in vivo, and formed a complex with two DNAJ domain-containing homologs, DNAJ1 and DNAJ2. Ectopic recruitment of DNAJ1 enhanced gene transcription in plants, yeast, and mammals. Thus, the SUVH proteins bind to methylated DNA and recruit the DNAJ proteins to enhance proximal gene expression, thereby counteracting the repressive effects of transposon insertion near genes.

DNA methylation frequently marks transposable elements (TEs) in eukaryotic genomes (1–3). In plants, the RNA-directed DNA methylation (RdDM) pathway is responsible for the initial establishment of methylation in CG, CHG, and CHH contexts (4). TE insertions can exert a transcriptional effect on neighboring genes (5–8), and promoter methylation is typically associated with gene repression (9). However, exceptions exist where promoter methylation is required for gene expression (10–14). The downstream factors that perceive methylation to mediate these divergent transcriptional effects are still poorly characterized, and little is known of how methylation can stimulate gene transcription.

To identify proteins in *Arabidopsis thaliana* that recognize methylated DNA, we incubated

nuclear extract from floral bud tissue with either methylated or unmethylated biotinylated double-stranded DNA oligonucleotides, affinity purified the DNA, and subjected the associated proteins to high-resolution mass spectrometry followed by label-free comparative analysis (15) (fig. S1). We used DNA sequences that are naturally methylated in vivo and two distinct DNA sequences for each of the CG, CHG, and CHH methylation contexts (fig. S2). A total of 41 proteins were significantly methyl enriched in at least one pull-down assay, including many candidates with known or predicted methyl-binding activity involved in gene silencing and methylation control (fig. S3). By requiring that candidates be significantly enriched in both DNA sequences for each of CG, CHG, and CHH, we obtained a stringent list of 10 candidates (Fig. 1A). Of these, relatively little is known about the role of the highly related SUVH1 and SUVH3 proteins (16) or the DNAJ proteins.

Recently, SUVH1 was isolated from an anti-silencing screen and was shown to promote the expression of promoter methylated genes (17). As SUVH1 and SUVH3 contain a SET- and RING-associated (SRA) domain (18), they are predicted to bind methylated DNA directly. Using fluorescence polarization (FP) and microscale thermophoresis (MST), we confirmed an SRA-dependent methyl-binding preference for recombinant SUVH1 and SUVH3 proteins, respectively, in CG, CHG, and CHH contexts (Fig. 1B and fig. S4). Chromatin immunoprecipitation sequencing (ChIP-seq) of transgenic lines expressing FLAG-tagged SUVH1 or SUVH3 showed that their localization was essentially identical (fig. S5A) and that they colocalized with CHH methylation deposited by the RdDM pathway (Fig. 2A and fig. S5B). SUVH1 and SUVH3 displayed enrichment directly over NRPE1 sites (19) [the largest subunit of the

RdDM component RNA polymerase V (Pol V)] (Fig. 2B and fig. S5C) and showed preferential localization over short TEs and at the edges of long TEs (Fig. 2C and fig. S5D), which are hallmarks of RdDM localization (20, 21). There was a positive correlation between SUVH1 and SUVH3 enrichment and RdDM-deposited CHH methylation (mCHH) at both local and genome-wide scales (fig. S5, E to H). Using random forest regression, we observed that mCHH was the strongest predictor for SUVH1 binding in vivo (Fig. 2, D and E).

The nearly perfect colocalization of SUVH1 with RdDM sites predicts that RdDM pathway mutants might reduce SUVH1 occupancy. ChIP-seq of SUVH1 in *nrrpd1*, *nrrpe1*, or *drml2* RdDM mutant backgrounds (4) showed that SUVH1 enrichment was essentially eliminated (Fig. 2F and fig. S6). To exclude the possibility that interaction with RdDM proteins, rather than DNA methylation itself, was responsible for SUVH1 recruitment, we compared ChIP-seq results for an SRA domain amino acid change mutant [with tyrosine-277 mutated to alanine (Y277A)] that abrogated methyl binding, SUVH1^{Y277A} (Fig. 1B). Indeed, SUVH1^{Y277A} showed highly reduced recruitment and association with CHH methylation (Fig. 2G and fig. S7).

Whole-genome bisulfite sequencing (WGBS) revealed that SUVH1 ChIP-seq peaks were characterized by local CHH methylation maxima and that in *suvh1*, *suvh3*, and double mutant *suvh1 suvh3* plants, methylation levels were unperturbed (17) (fig. S8A). This indicated that SUVH1 and SUVH3 are not required for methylation maintenance and act strictly as methyl readers. RNA sequencing (RNA-seq) of *suvh1*, *suvh3*, and *suvh1 suvh3* confirmed many of the previously identified (17) promoter methylated genes that require SUVH1 for expression (fig. S8B) and showed reduced expression at genes proximal to RdDM sites (22) (fig. S8C).

SUVH1 and SUVH3 might enhance transcription by directly impacting chromatin (18), as both encode SET domains of the SU(VAR)3-9 family that typically methylate histone H3's lysine-9 (23). However, we were unable to detect histone methyltransferase (HMT) activity in vitro (fig. S9) or changes in dimethylation of histone 3 lysine-9 (H3K9me2) levels in *suvh1 suvh3* mutants in vivo (17) (fig. S10). Furthermore, SUVH1^{Y524F} and SUVH1^{Y638F} predicted HMT catalytic mutants (18), but not the SUVH1^{Y277A} methyl-binding mutant, were able to complement *suvh1*, indicating that HMT activity is nonessential for function in vivo (fig. S11). Chromatin accessibility, as profiled by ATAC-seq (a sequencing technique based on an assay for transposase-accessible chromatin), was also unchanged in *suvh1 suvh3* mutants (fig. S12).

Next, we assessed whether SUVH1 and SUVH3 might enhance transcription by acting as a recruitment platform (24). Immunoprecipitation followed by mass spectrometry (IP-MS) of SUVH1 and SUVH3 identified that each pulled down the other and also DNAJ1 and DNAJ2 (Figs. 1A and 3A and fig. S13). IP-MS of DNAJ1 and DNAJ2

¹Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, CA 90095, USA. ²Quantitative Proteomics, Institute of Molecular Biology, 55128 Mainz, Germany. ³Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI 49503, USA. ⁴National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Center for Plant Stress Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201602, China. ⁵State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, Fujian Agriculture and Forestry University, 350002 Fuzhou, China. ⁶Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. ⁷Institute of Oceanography, Minjiang University, 350108 Fuzhou, China. ⁸Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, CA, USA.

*These authors contributed equally to this work. †Present address: Department of Microbiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand 50200. ‡Present address: Genomic Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Dr., San Diego, CA 92121, USA. §Present address: Helmholtz Zentrum München, Institute of Biochemical Plant Pathology, 85764 Neuherberg, Germany. ||Corresponding author. Email: f.butter@imb-mainz.de (F.B.); jacobsen@ucla.edu (S.E.J.)

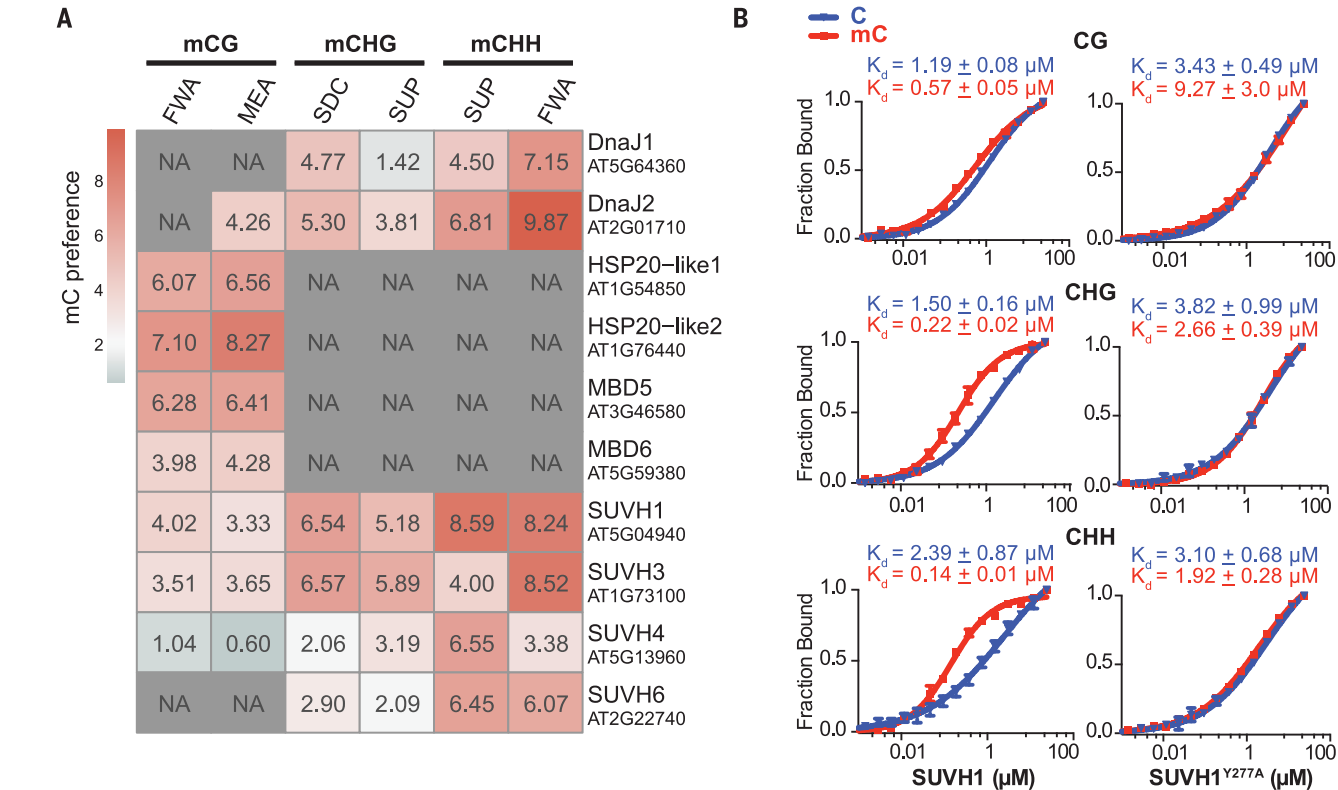


Fig. 1. Comparative interactomics identifies methyl reader proteins. (A) Heatmap of methyl-binding preferences for proteins identified as significantly enriched in two different underlying DNA sequences per methyl-cytosine (mC) context (mCG, mCHG, mCHH). NA, the protein was not detected. FWA, MEA, SDC, and SUP represent four in vivo-methylated loci. Probes are listed in fig. S2. (B) FP binding assays to

quantify the interaction of SUVH1 with methylated or unmethylated probes in CG, CHG, and CHH contexts (left) or an amino acid change version, SUVH1^{Y277A}, predicted to abrogate methyl binding (18) (right). Binding affinities are indicated by dissociation constants (K_d) values. Error bars represent SEM of technical replicates. The data are representative of two independent experiments.

showed that each of these pulled down the other and also SUVH1 and SUVH3 (Fig. 3A and fig. S13), indicating that SUVH1, SUVH3, DNAJ1, and DNAJ2 interact in vivo. We confirmed the interactions between SUVH1 and SUVH3 with DNAJ and DNAJ2 by coimmunoprecipitation in *Nicotiana benthamiana* and in yeast two-hybrid assays (figs. S14 and S15). To assess the strength of the interaction, we expressed all four proteins in the same bacterial cell and performed affinity purification of either SUVH1 or SUVH3, finding that both DNAJ1 and DNAJ2 remained associated even under 500mM NaCl conditions (fig. S16).

DNAJ1 and DNAJ2 lack any discernible methyl-binding domain, but they are robustly associated with SUVH1 and SUVH3, suggesting that SUVH1 and SUVH3 may be responsible for recruiting DNAJ1 and DNAJ2 to methylated DNA (Fig. 1A). We repeated a CHH context pulldown experiment with *suvh1 suvh3* and *dnaj1 dnaj2* double mutant plants. DNAJ1 and DNAJ2 were no longer associated with methyl-DNA in *suvh1 suvh3*, while SUVH1 and SUVH3 methyl-DNA binding was unaffected in *dnaj1 dnaj2* (fig. S17). Thus, SUVH1 and SUVH3 are required to recruit DNAJ1 and DNAJ2 to methylated DNA. We performed ChIP-seq of DNAJ1 and DNAJ2 and found a tight genome-wide correlation with SUVH1 and SUVH3 (Fig. 3, B and C, and fig. S18, A and B). As

with *suvh1 suvh3*, there was no effect on DNA methylation levels in *dnaj1 dnaj2* mutants, consistent with a downstream reader function (fig. S18C). To assess whether DNAJ1 and DNAJ2 are required for the transcriptional enhancement activity of SUVH1 and SUVH3, we performed RNA-seq on *dnaj1 dnaj2*, and double mutant *dnaj1 dnaj2* plants. The *dnaj1 dnaj2* transcriptome was strongly positively correlated with that of *suvh1 suvh3* (Fig. 3D and fig. S19), and RdDM proximal genes showed reduced expression in both *suvh1 suvh3* and *dnaj1 dnaj2* double mutants (fig. S20). *ROS1* is one of the few loci known to require methylation for expression (11, 12), and indeed we observed reduced expression of *ROS1* in both the *suvh1 suvh3* and *dnaj1 dnaj2* backgrounds, despite methylation levels being maintained (fig. S21). Furthermore, genes with promoters proximal to SUVH1 peaks generally showed reduced expression in both the *suvh1 suvh3* and *dnaj1 dnaj2* double mutants (Fig. 3E). Together, these data indicate that DNAJ1 and DNAJ2 interact with SUVH1 and SUVH3, are recruited to sites of RdDM, and promote the expression of proximal genes.

The yeast two-hybrid experiments revealed that binding domain (BD)-fused DNAJ1 induced expression of the reporter even when cotransformed with an unfused activation domain con-

struct (fig. S15). This suggested that DNAJ1 alone may be sufficient to stimulate expression of the reporter, which we confirmed in a yeast one-hybrid assay (fig. S22A). We fused DNAJ1 to a zinc finger protein (ZF108) (24) behind the *UBQ10* promoter and cotransformed it into *N. benthamiana* with a reporter construct containing either the ZF108 target site or a scrambled target site in the promoter region. Expression of the ZF108 target reporter was increased by approximately threefold above that of the scrambled promoter (fig. S22B). To assess whether DNAJ1 can function in a mammalian context (25), we transfected N2a cells and found that Gal4 DNA-binding domain (Gal4BD)-fused DNAJ1 was able to stimulate transcription of the reporter by 5- to 10-fold (fig. S22C).

Next, we generated stable transgenic *A. thaliana* lines using the *UBQ10::ZF108-DNAJ1* construct. The first-generation independent transgenic lines displayed severe morphological defects (fig. S23). RNA-seq and ChIP-seq (Fig. 4A) on these *UBQ10::ZF108-DNAJ1* lines showed that up- but not down-regulated genes were significantly enriched for overlap with ZF108-DNAJ1 ChIP-seq peaks (observed over expected = 2.26, hypergeometric test $P = 7.7 \times 10^{-71}$) (fig. S24). As controls, we generated *UBQ10::ZF108-YPET* and *UBQ10::DNAJ1* (without ZF108) transgenic plants and found no

Fig. 2. SUVH1 is recruited by RdDM-associated mCHH.

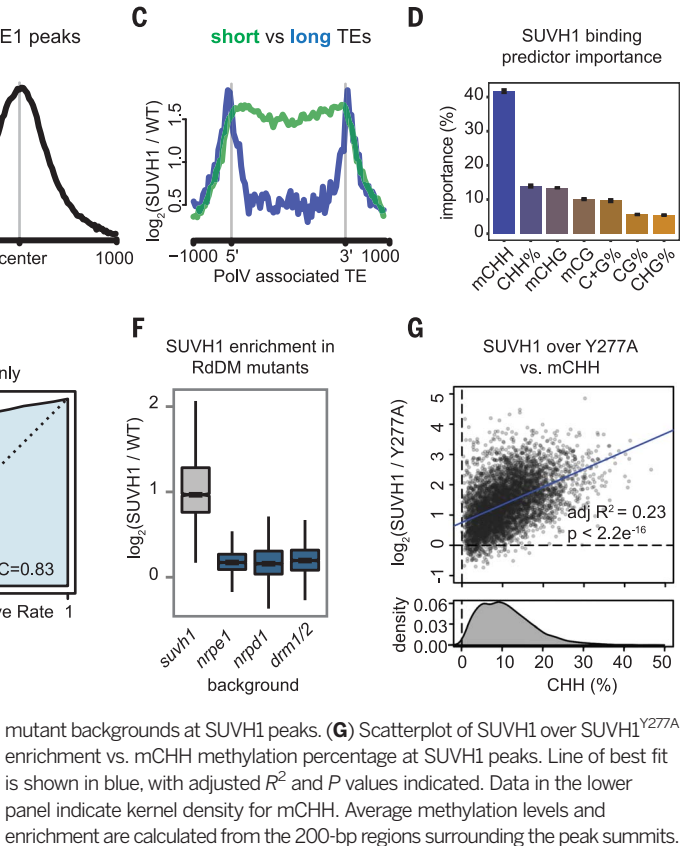
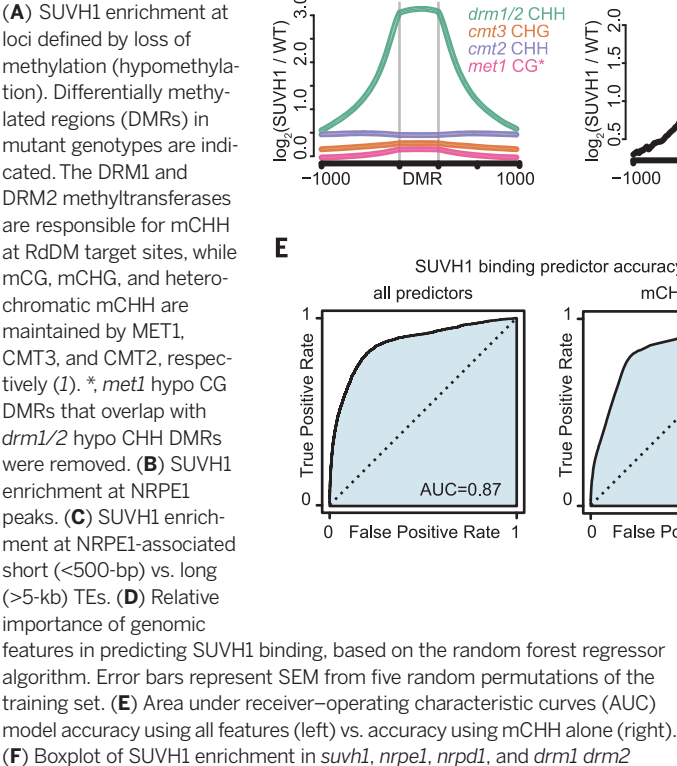
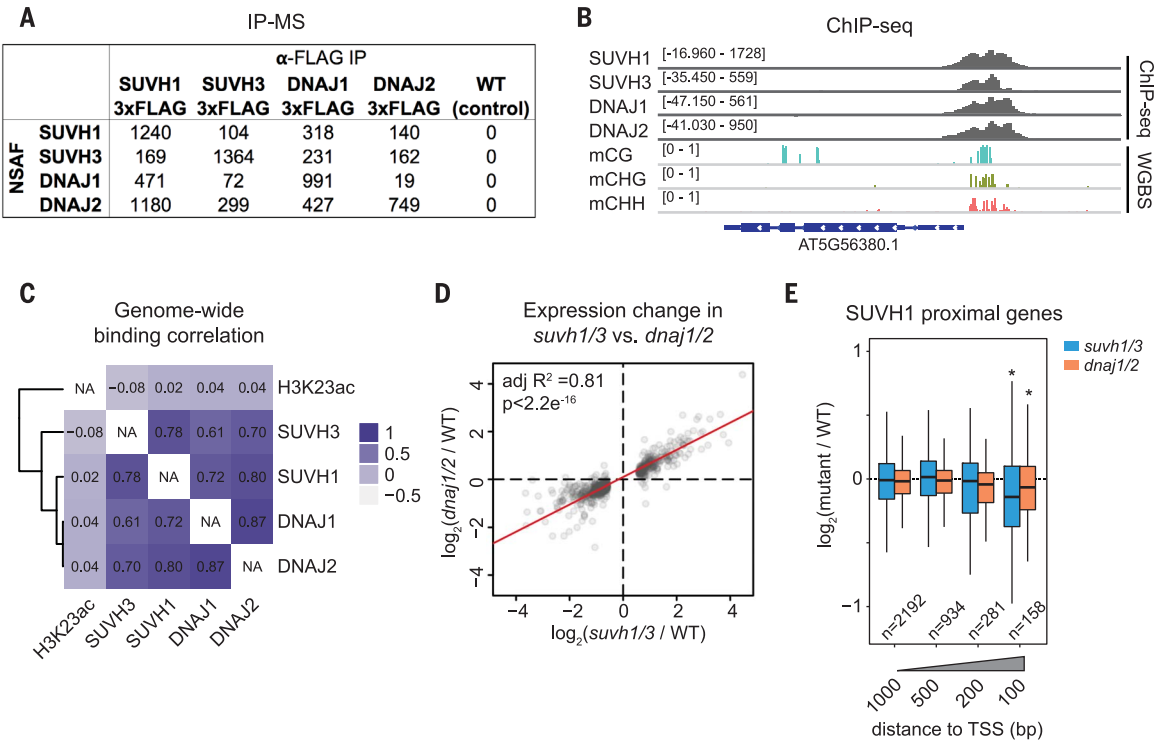
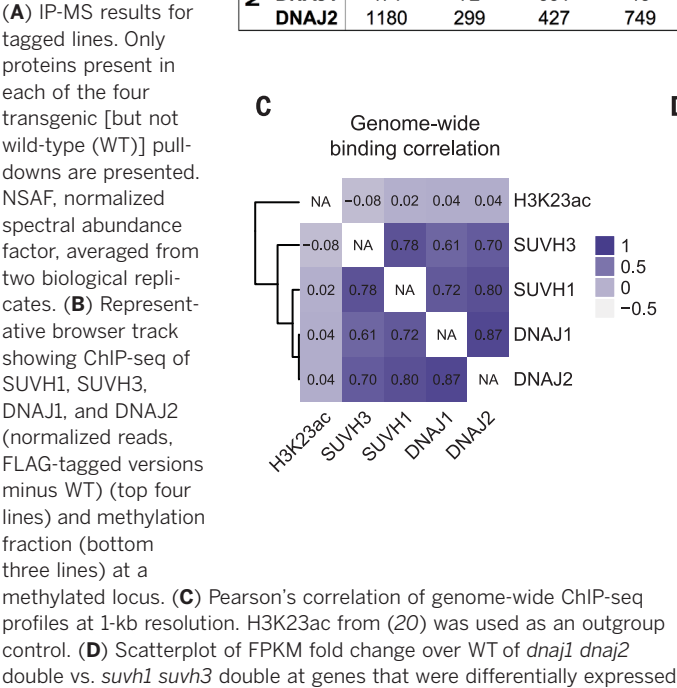


Fig. 3. SUVH1, SUVH3, DNAJ1, and DNAJ2 interact, colocalize, and are required for the expression of proximal genes.



in *suvh1 suvh3*. Line of best fit is shown in red, with adjusted R^2 and P values indicated. (E) Boxplot of expression change for genes proximal to SUVH1 binding sites. n , number of genes. * $P < 0.05$ (Mann-Whitney test).

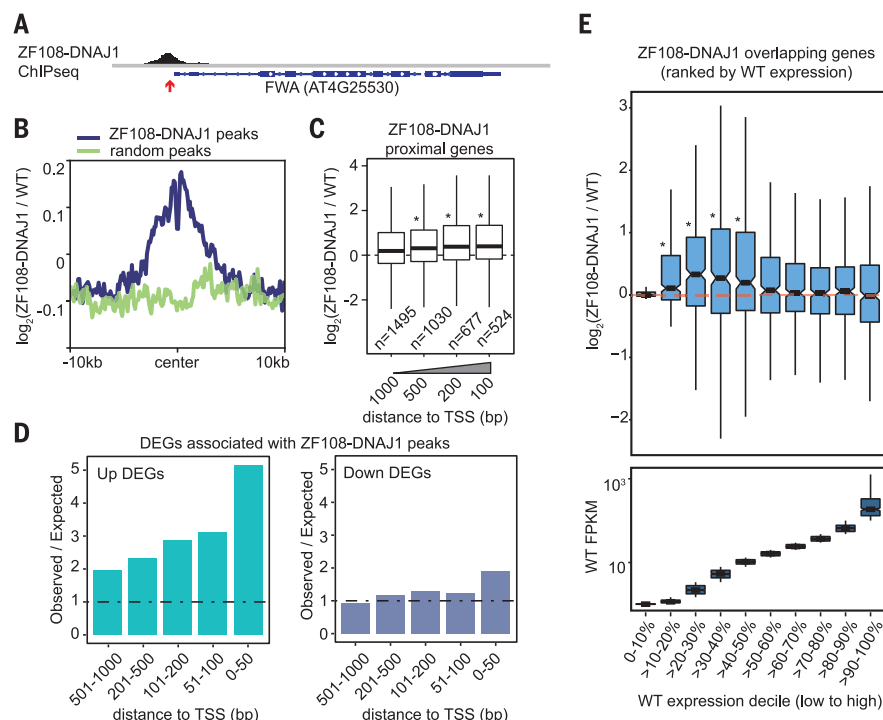


Fig. 4. ZF108-DNAJ1 transcriptionally activates mildly expressed proximal loci. (A) Browser track showing the ZF108-DNAJ1 ChIP-seq profile at FWA. The red arrow indicates the genomic location of the designed ZF108 target binding site. (B) Metaplot of expression change, centered on ZF108-DNAJ1 vs. random peaks. (C) Boxplot of expression changes for genes with promoters proximal to ZF108-DNAJ1 binding sites. n , number of genes. $*P < 0.05$ (Mann-Whitney test). (D) Observed over expected ratio for overlap of ZF108-DNAJ1 sites with up- or down-regulated ZF108-DNAJ1 gene promoters. (E) Boxplot of expression change for genes that overlap with ZF108-DNAJ1 peaks (upper panel), arranged by ascending WT expression decile (lower panel). Genes that lacked expression in both genotypes were removed. $*P < 0.05$ (Mann-Whitney test).

morphological defects or transcriptional changes associated with ZF108 peaks, indicating that neither ZF108 binding nor DNAJ1 overexpression was sufficient to cause the transcriptional defects observed (fig. S24). In addition, bulk levels of RNA were increased over ZF108-DNAJ1 peaks ($n = 4951$), and there was a clear promoter proximal effect on transcription (Fig. 4, B to D). In contrast, neither up- nor down-regulated gene sets showed an association with TEs or RdDM sites, indicating that ZF108-DNAJ1 acts primarily at ectopic locations driven by ZF108 binding (fig. S25). Together, these data showed that recruitment of DNAJ1 increases the expression of proximal neighboring genes.

Given that SUVH1, SUVH3, DNAJ1, and DNAJ2 are localized at RdDM sites, including many TE sequences, an interesting paradox is what prevents TEs themselves from being reactivated. FWA, the gene that ZF108 was designed to target (24), is stably silent in wild-type plants and experienced no transcriptional up-regulation in transgenic plants, despite clear localization of ZF108-DNAJ1 to FWA (Fig. 4A and fig. S26). We reasoned that the transcriptional enhancement effect of DNAJ1 may be limited to genes that are already expressed, as opposed to traditional transcriptional

activator proteins, such as VP16, that can activate transcription of stably silent genes (26). Parsing the ZF108-DNAJ1 overlapping genes into expression deciles revealed that only genes with moderate expression in the wild type, but not those in the lowest or higher expression deciles, experienced transcriptional assistance (Fig. 4E). This provides a simple explanation for the paradox, as only proximal expressed genes would be affected, leaving TEs silent.

We propose that SUVH1 and SUVH3 in complex with DNAJ1 and DNAJ2 evolved to counteract the repressive effect of TE insertion near genes (8, 27, 28), thereby facilitating access to the gene regulatory diversity provided by TE proliferation (29–31). This is consistent with SUVH1, SUVH3, DNAJ1, and DNAJ2 being recruited downstream of the RdDM pathway, which is known to target evolutionarily young TEs and to cause mild repression of genes near TEs (22). The complex of SUVH1, SUVH3, DNAJ1, and DNAJ2 also reveals a potential mechanism to explain examples of methylation-dependent gene expression (11–13). Overall, these findings shed light on how methylation can act to fine-tune gene expression by balancing both repressive and activating transcriptional effects.

REFERENCES AND NOTES

1. J. Du, L. M. Johnson, S. E. Jacobsen, D. J. Patel, *Nat. Rev. Mol. Cell Biol.* **16**, 519–532 (2015).
2. S. Feng et al., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8689–8694 (2010).
3. A. Zemach, I. E. McDaniel, P. Silva, D. Zilberman, *Science* **328**, 916–919 (2010).
4. M. A. Matzke, R. A. Mosher, *Nat. Rev. Genet.* **15**, 394–408 (2014).
5. A. E. Van't Hof et al., *Nature* **534**, 102–105 (2016).
6. T. Stuart et al., *eLife* **5**, e20777 (2016).
7. L. Quadrona et al., *eLife* **5**, e15716 (2016).
8. J. D. Hollister, B. S. Gaut, *Genome Res.* **19**, 1419–1428 (2009).
9. J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **11**, 204–220 (2010).
10. K. Bahar Halpern, T. Vana, M. D. Walker, *J. Biol. Chem.* **289**, 23882–23892 (2014).
11. B. P. Williams, D. Pignatta, S. Henikoff, M. Gehring, *PLOS Genet.* **11**, e1005142 (2015).
12. M. Lei et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3553–3557 (2015).
13. K. Shibuya, S. Fukushima, H. Takatsuki, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1660–1665 (2009).
14. J. Moreno-Romero, H. Jiang, J. Santos-González, C. Köhler, *EMBO J.* **35**, 1298–1311 (2016).
15. D. Kappei et al., *Nat. Commun.* **8**, 14334 (2017).
16. L. O. Baumbusch et al., *Nucleic Acids Res.* **29**, 4319–4333 (2001).
17. S. Li et al., *Nucleic Acids Res.* **44**, 608–620 (2016).
18. J. Du et al., *Mol. Cell* **55**, 495–504 (2014).
19. W. Liu et al., *Nat. Plants* **4**, 181–188 (2018).
20. H. Stroud et al., *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014).
21. A. Zemach et al., *Cell* **153**, 193–205 (2013).
22. X. Zhong et al., *Nat. Struct. Mol. Biol.* **19**, 870–875 (2012).
23. J. P. Jackson, A. M. Lindroth, X. Cao, S. E. Jacobsen, *Nature* **416**, 556–560 (2002).
24. L. M. Johnson et al., *Nature* **507**, 124–128 (2014).
25. T. Y. Erkina, A. M. Erkina, *Epigenetics Chromatin* **9**, 40 (2016).
26. L. G. Lowder et al., *Plant Physiol.* **169**, 971–985 (2015).
27. X. Wang, D. Weigel, L. M. Smith, *PLOS Genet.* **9**, e1003255 (2013).
28. D. Meng et al., *PLOS Genet.* **12**, e1006141 (2016).
29. K. Naito et al., *Nature* **461**, 1130–1134 (2009).
30. V. Sundaram, T. Wang, *BioEssays* **40**, 170155 (2018).
31. H. Ito et al., *Nature* **472**, 115–119 (2011).

ACKNOWLEDGMENTS

We thank S. Feng and M. Akhavan for the high-throughput sequencing performed at the UCLA Broad Stem Cell Research Center BioSequencing Core Facility. We thank J. A. Long for advice on nuclear isolation, M. F. Carey for providing plasmids for N2a transfections, and Y. Ma, J. Appell, J. Zhao, A. Thai, J. Nail, G. Anigol, R. Sahu, and A. Desouza for technical assistance. **Funding:** This work was supported by grants NIH R01 GM06398 (to S.E.J.), NIH R01 GM089778 (to J.A.W.), and NIH R35 GM124736 (to S.B.R.), by an EMBO Long-Term Fellowship (ALTF 1138-2014) (to C.J.H.), and by a Ruth L. Kirschstein National Research Service Award (GM007185) (to L.Y.). S.E.J. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** F.B. and S.E.J. conceived the study; C.J.H., M.S., J.A.W., J.D., S.B.R., F.B., and S.E.J. designed the research; C.J.H., S.P.W., Y.X., L.Y., J.G.B., and M.G. performed the experiments; M.S. performed the comparative interactomics; W.L. performed the random forest regression analysis; E.M.C. and R.M.V. performed the FP assays; X.L. and W.C. performed MST assays; W.D.B. and S.R. performed the mass spectrometry from immunoprecipitated samples; C.J.H., W.L., and Z.Z. performed bioinformatic analysis; C.J.H. and S.E.J. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The high-throughput sequencing data generated in this paper have been deposited in the Gene Expression Omnibus (GEO) database (GSE108414).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6419/1182/suppl/DC1
Materials and Methods
Figs. S1 to S26
References (32–60)

17 December 2017; resubmitted 7 September 2018
Accepted 31 October 2018
10.1126/science.aar7854