

# INVOLVED IN DE NOVO 2-containing complex involved in RNA-directed DNA methylation in *Arabidopsis*

Israel Ausin<sup>a,b,1</sup>, Maxim V. C. Greenberg<sup>a,1</sup>, Dharendra K. Simanshu<sup>c,1</sup>, Christopher J. Hale<sup>a</sup>, Ajay A. Vashisht<sup>d</sup>, Stacey A. Simon<sup>e</sup>, Tzuu-fen Lee<sup>e</sup>, Suhua Feng<sup>a</sup>, Sophia D. Española<sup>a</sup>, Blake C. Meyers<sup>e</sup>, James A. Wohlschlegel<sup>d</sup>, Dinshaw J. Patel<sup>f</sup>, and Steven E. Jacobsen<sup>a,f,g,2</sup>

<sup>a</sup>Department of Molecular, Cell, and Developmental Biology, <sup>d</sup>Department of Biological Chemistry, David Geffen School of Medicine, <sup>f</sup>Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, and <sup>g</sup>Howard Hughes Medical Institute, University of California, Los Angeles, CA 90095; <sup>b</sup>Department of Plant Molecular Genetics, Campus Universidad Autónoma de Madrid, Madrid 28045, Spain; <sup>c</sup>Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065; and <sup>e</sup>Department of Plant and Soil Sciences and Delaware Biotechnology Institute, University of Delaware, Newark, DE 19711

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2011.

Contributed by Steven E. Jacobsen, April 20, 2012 (sent for review March 20, 2012)

At least three pathways control maintenance of DNA cytosine methylation in *Arabidopsis thaliana*. However, the RNA-directed DNA methylation (RdDM) pathway is solely responsible for establishment of this silencing mark. We previously described INVOLVED IN DE NOVO 2 (IDN2) as being an RNA-binding RdDM component that is required for DNA methylation establishment. In this study, we describe the discovery of two partially redundant proteins that are paralogous to IDN2 and that form a stable complex with IDN2 in vivo. Null mutations in both genes, termed IDN2-LIKE 1 and IDN2-LIKE 2 (IDNL1 and IDNL2), result in a phenotype that mirrors, but does not further enhance, the *idn2* mutant phenotype. Genetic analysis suggests that this complex acts in a step in the downstream portion of the RdDM pathway. We also have performed structural analysis showing that the IDN2 XS domain adopts an RNA recognition motif (RRM) fold. Finally, genome-wide DNA methylation and expression analysis confirms the placement of the IDN proteins in an RdDM pathway that affects DNA methylation and transcriptional control at many sites in the genome. Results from this study identify and describe two unique components of the RdDM machinery, adding to our understanding of DNA methylation control in the *Arabidopsis* genome.

genomics | mass spectrometry | siRNAs | epigenetics

DNA methylation is a stable epigenetic mark that is associated with the repression of genes and transposable elements. In *Arabidopsis thaliana*, maintenance of DNA methylation at silent loci is carried out by at least three methyltransferases: METHYLTRANSFERASE 1 (MET1), CHROMOMETHYLTRANSFERASE 3 (CMT3), and DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2). However, DRM2 is solely responsible for establishment of DNA methylation—or de novo methylation—of silent elements (1). DRM2 is guided to chromatin by small interfering RNAs (siRNAs) in a process known as RNA-directed DNA methylation (RdDM) (2). It has been proposed that RdDM also requires intergenic noncoding (IGN) transcripts that are synthesized by RNA Polymerase V (Pol V). These transcripts likely serve as platforms for the recruitment of siRNA-loaded ARGONAUTE 4 (AGO4) to methylated loci (3, 4).

Recently, we discovered the requirement of INVOLVED IN DE NOVO 2 (IDN2) for RdDM from a forward genetic screen (5). Alleles of this same gene were later reported from another screen for DNA methylation mutants (6). We previously demonstrated that IDN2 binds to double-stranded RNA through its XS domain, which is also observed in the XS-domain-containing protein SUPPRESSOR OF GENE SILENCING 3 (SGS3) (7). We also found that IDN2 was likely to act in a step downstream of initial siRNA biogenesis. The XS domain is conserved throughout the plant kingdom, and XS-domain-containing proteins are

involved in a wide range of processes such as viral defense (8) and stress response (9).

To gain a better understanding of the in vivo role of IDN2, we performed affinity purifications from complementing transgenic lines expressing epitope-tagged full-length IDN2 protein. We found that IDN2 forms a complex with two unique proteins from the same family as IDN2, which we termed IDN2-LIKE 1 (IDNL1) and IDN2-LIKE 2 (IDNL2). Through single-locus approaches as well as whole-genome bisulfite sequencing we find that IDNL1 and IDNL2 are also essential components of the RdDM pathway. In higher-order mutants, the methylation phenotype is not enhanced further than what is observed in the single *idn2-1* mutant, suggesting a nonredundant role between IDN2 and its paralogs in the IDN2 complex. Comparisons with known RdDM mutants in genome-wide methylation and expression analyses solidify the role of the IDN2 complex as a core component of RdDM machinery.

## Results

**Structural Analysis of IDN2 XS Domain.** We previously showed that IDN2 binds to double-stranded RNA with 5' overhangs in vitro through its XS domain (5). Bioinformatic analysis has suggested that the XS domain is likely to adopt a unique RNA-recognition motif (RRM) fold, which would be consistent with its in vitro activity (10). To gain further insights into the structure and mechanism of the XS domain we determined the structure of the IDN2-XS domain along with a small segment of adjacent coiled-coil region (120–292) by X-ray crystallography (Fig. 1). We found that the core structure of the XS domain superimposes closely over a known RRM domain. However, insertions in the XS domain form a few additional secondary structural elements: a  $\beta$ -strand ( $\beta_N$ ) at the N terminus, a longer loop having an antiparallel  $\beta$ -sheet (formed by  $\beta_{1A}$  and  $\beta_{1B}$ ) between  $\alpha_1$  and  $\beta_2$ , a longer loop having a small  $\alpha$ -helix ( $\alpha_3$ ) between  $\alpha_2$  and  $\beta_4$ , and two additional  $\alpha$ -helices ( $\alpha_3$  and  $\alpha_4$ ) at the C-terminal end of the XS domain (Fig. 1B and Fig. S1). Because RNA binding

Author contributions: I.A., M.V.C.G., D.K.S., and S.E.J. designed research; I.A., M.V.C.G., D.K.S., A.A.V., S.F., and S.D.E. performed research; B.C.M., J.A.W., and D.J.P. contributed new reagents/analytic tools; D.K.S., C.J.H., A.A.V., S.A.S., and T.-f.L. analyzed data; and I.A., M.V.C.G., and S.E.J. wrote the paper.

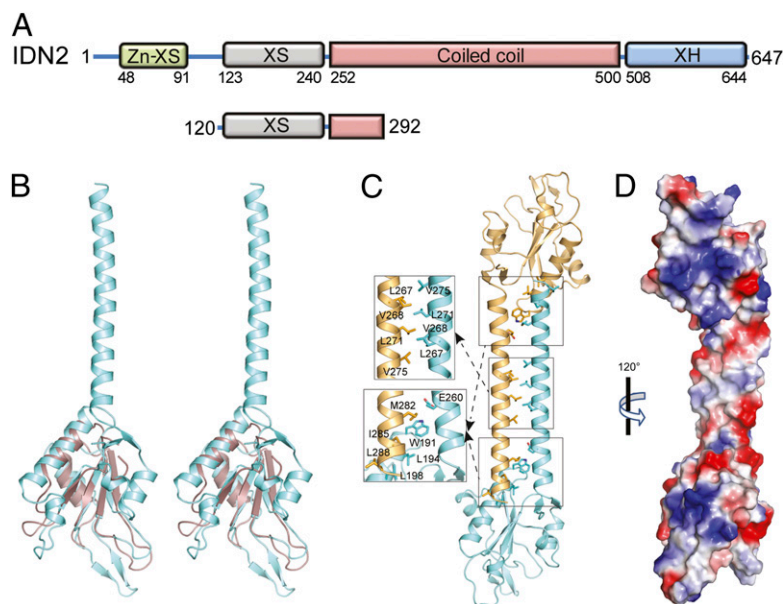
The authors declare no conflict of interest.

Data deposition: The atomic coordinates reported in this paper have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID code 4E8U) and the data reported in this paper deposited in the Gene Expression Omnibus (GEO) database, [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo) (accession nos. GSE36143, GSE37206, and GSE36129).

<sup>1</sup>I.A., M.V.C.G., and D.K.S. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. E-mail: [jacobsen@ucla.edu](mailto:jacobsen@ucla.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1206638109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1206638109/-DCSupplemental).



**Fig. 1.** *Arabidopsis* IDN2 domain architecture and crystal structure of IDN2-XS domain (120–292) of IDN2. (A) Domain architecture of *Arabidopsis* IDN2 (Upper) and XS domain and a small segment of coiled-coil region (Lower) used for structural studies. (B) Stereoview of structural superposition of the IDN2-XS domain with the RRM domain of FBP-interacting repressor [Protein Data Bank (PDB) ID code: 2QFJ]. The IDN2-XS domain is colored in cyan whereas the RRM domain is colored in brown. (C) Quaternary (dimeric) structure of the IDN2-XS domain. The two molecules are colored in cyan and light orange. The two subunits interact mainly via hydrophobic interactions formed by residues present in the  $\alpha$ -helix of the coiled-coil segment. A few hydrophobic residues present on the surface of the XS domain also interact with the residues present at the C-terminal end of the terminal helix. (D) Electrostatic surface representation of the dimeric IDN2-XS domain highlighting basic (blue) and acidic (red) regions on the IDN2-XS domain.

specificity by RRM fold proteins depends on the loops present between  $\alpha$ -helices and  $\beta$ -strands, it is likely that insertions in these regions in the XS domain result in its unique specificity toward 5'-overhang-containing dsRNA. This structural study provides unique empirical evidence that the XS domain has a unique RRM motif. Additionally, by examining the electrostatic surface of the XS domain, we observed an exposed positively charged basic patch (Fig. 1D). This patch likely interacts directly with negatively charged RNA molecules.

As seen in the crystal structure, IDN2 (120–292) dimerizes mainly via a coiled-coil segment (Fig. 1C and D). The fact that the coiled-coil region readily dimerizes suggests that IDN2 potentially exists *in vivo* in higher-order complexes with itself. However, as discussed below, the interaction is more likely between IDN2 and closely related homologs, which also contain a coiled-coil domain.

**IDN2 Forms a Complex with Paralog IDNL1 and IDNL2.** To determine the interacting partners of IDN2, we produced transgenic lines

expressing IDN2 fused to different epitope tags under the control of the IDN2 promoter region. These IDN2 epitope-tagged transgenic lines were able to complement the methylation defect of the *idn2-1* mutant at the *MEDEA-INTERGENIC SUBTELOMERIC REPEATS* (*MEA-ISR*), indicating normal protein functionality (Fig. S2). Upon establishing homozygous lines, we prepared protein extracts from the apical tissue of *BLRP::9xMyc::IDN2* complementing lines and performed affinity purification with streptavidin. Purified extracts were analyzed by multidimensional protein identification technology (MudPIT) (11). MudPIT analysis from two independent purifications revealed the presence of abundant peptides of the proteins At1g15910 and At4g00380, indicating that those two proteins and IDN2 could form a complex *in vivo* (Table 1). Much less abundant peptides from a few other proteins were also found in both replicas, although it is not known whether these are of any significance. At1g15910 and At4g00380 are in the same gene family as IDN2,

**Table 1.** Mass spectrometric analyses of IDN2 and IDNL1 affinity purifications

	Experiment I				Experiment II			
	Spectra	UniPepts	Coverage, %	NSAF	Spectra	UniPepts	Coverage, %	NSAF
IDN2 purification								
IDN2	651	91	68.0	16,261	398	52	53.0	26,405
AT1G15910 (IDNL1)	408	59	60.9	10,400	127	30	40.4	8,598
AT4G00380 (IDNL2)	163	39	49.4	4,148	82	24	32.3	5,543
AT5G24780	3	3	12.2	179	5	3	10.4	794
AT1G78300	7	6	27.0	436	3	3	17.8	497
AT4G16143	2	2	7.0	72	3	2	7.0	290
IDNL1 purification								
IDNL1	32	13	30.0	755	103	20	33.0	2,238
IDN2	6	3	6.2	138	87	22	31.4	1,852

NSAF, normalized spectral abundance factor. Only proteins appearing in two experiments are shown.

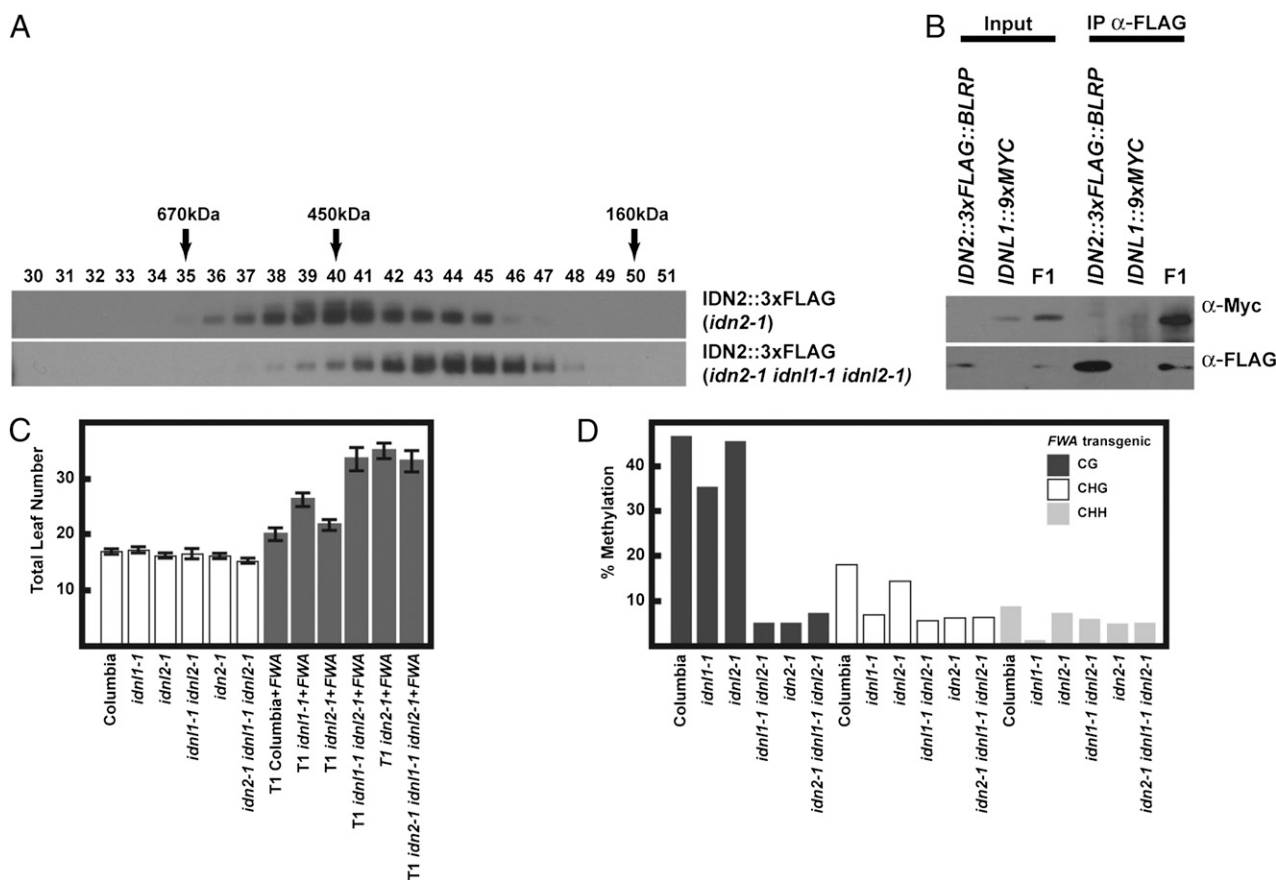
and share 92% amino acid identity with each other (Fig. S3A). The members of this gene family all contain the same domain architecture and organization as IDN2 (Fig. 1A). We termed At1g15910 and At4g00380 IDNL1 and IDNL2, respectively.

To confirm the interaction between IDN2 and IDNL1, we performed several assays including gel filtration, IDNL1 co-immunoprecipitation (co-IP), and IDNL1 affinity purifications (Fig. 2A and B and Table 1). For gel filtration assays we first generated a complementing transgenic line expressing IDN2::3xFLAG::BLRP in two different genetic backgrounds: the *idn2-1* mutant and the *idn2-1 idnl1-1 idnl2-1* triple mutant. After gel filtration and Western blotting, the elution profile revealed a significant delay in elution of the complex in the triple *idn2-1 idnl1-1 idnl2-1* mutant background compared with *idn2-1* (Fig. 2A). Given that both *idnl1-1* and *idnl2-1* are insertion mutants that do not generate a transcript (Fig. S3B and C), this delay can be explained by the absence of these two proteins in the complex.

To examine the IDN2 complex in further detail, we generated complementing transgenic lines carrying the IDNL1::9xMYC fusion under the control of IDNL1 promoter region (Fig. S2). This line was crossed to IDN2::3xFLAG::BLRP, and plants from the subsequent F<sub>1</sub> generation were analyzed by co-IP, confirming the in vivo interaction (Fig. 2B). Additionally, we affinity purified the transgenic IDNL1 using an anti-Myc antibody and performed MudPIT analysis. Consistent with both the co-IP and IDN2 MudPIT, we observed a substantial interaction with IDN2 (Table

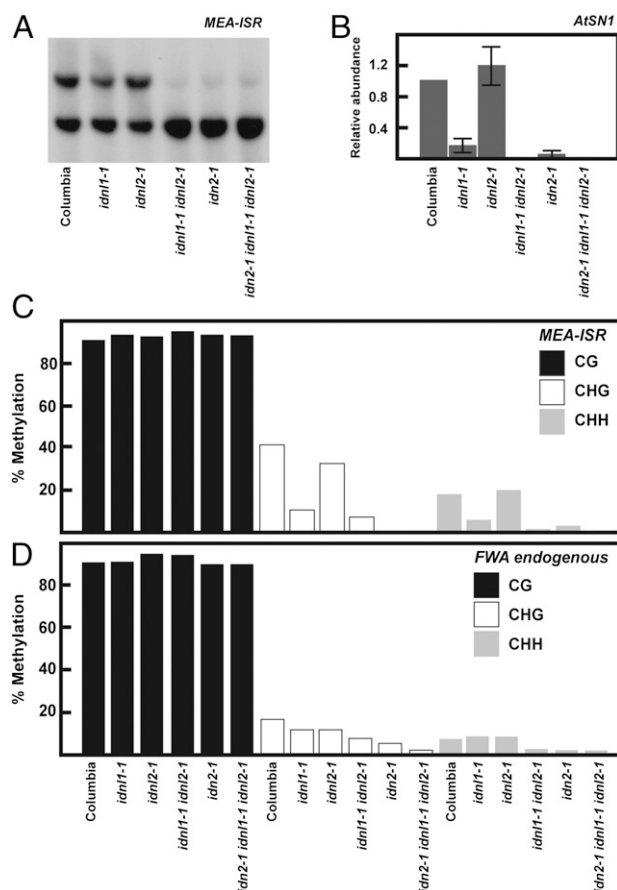
1). In fact, over two independent replicates, IDN2 was the only protein to be recovered from both. Taken together, these data confirm the IDNL1–IDN2 in vivo interaction.

**IDNL1 and IDNL2 Are Required for RdDM.** Given the fact that IDN2 affects de novo methylation (5) and both IDNL1 and -2 copurify with IDN2, it seemed likely that those two genes might be also required for de novo methylation. To assess this hypothesis, we used the well-studied gene *FLOWERING WAGENINGEN* (*FWA*). *FWA* is heritably silenced by methylation; however, unmethylated *fwa* epialleles exhibit ectopic expression that results in a dominant late-flowering phenotype (12). After *FWA* transformation, wild-type plants are able to methylate and silence *FWA* transgenes, whereas RdDM mutants fail to methylate *FWA* and thus flower late (1, 12). Using *FWA*, we transformed individual *idnl* mutants as well as the *idnl* double mutant and the *idn2-1 idnl1-1 idnl2-1* triple mutant. After *FWA* transformation, *idnl2-1* did not show any flowering defect, whereas *idnl1-1* displayed a slightly late flowering phenotype (Fig. 2C). This delay in flowering was correlated with a small decrease in methylation in all contexts; however, preestablished CG methylation remained unaffected at the *FWA* endogenous gene (Fig. 3D). Interestingly *idnl1-1 idnl2-1* double-mutant plants showed a late-flowering phenotype as strong as that of *idn2-1*. However, the *idn2-1 idnl1-1 idnl2-1* triple mutant did not show a defect any stronger than *idn2-1* (Fig. 2C and D). These data suggest partial redundancy



**Fig. 2.** IDN2–IDNLs in vivo interaction and de novo methylation phenotype of *idnl* mutants. (A) Gel filtration showing the elution profile of IDN2::3xFLAG fusion in two different mutant backgrounds. Arrows indicate the fraction in which size standards peak. (B) FLAG pull-down and coimmunoprecipitation assays confirming IDN2–IDNL1 interaction. Input lanes confirm expression of the epitope fusion proteins in the parental lines indicated above each lane. F<sub>1</sub> represents a cross between the two parental lines. (C) Flowering time measured as total number of leaves produced by wild-type (Columbia), *idnl* mutants, and *FWA*-transformed T<sub>1</sub> plants under long-day conditions. Error bars depict SE. (D) Methylation status of wild-type and *idnl* mutants at the transgenic copy of *FWA*.





**Fig. 3.** Maintenance methylation phenotype of the *idn* mutants. (A) Methylation-sensitive enzyme Southern hybridization assay at the *MEA-ISR* locus. *MspI* is blocked by methylation of the external C in the CCGG context. (B) *HaeIII* cutting assay at *AtSN1*. Genomic DNA is digested with *HaeIII* and then quantitative PCR is performed. Y-values are relative to uncut and then to Columbia. Errors bar represent SE from two independent experiments, each with two technical replicates. *HaeIII* is blocked by cytosine methylation in the GGCC context. (C) Methylation status of wild-type (Columbia) and *idn* mutants at endogenous *MEA-ISR*. (D) Methylation status of Columbia and *idn* mutants at endogenous *FWA*.

between IDNL1 and IDNL2, which is consistent with their high similarity at the amino acid level. The function of IDNL2 can be fully compensated by IDNL1, but IDNL2 seemingly cannot entirely replace the function of IDNL1.

All known genes affecting de novo methylation are involved in the maintenance of DRM2-mediated methylation at several loci (13, 14). Combining methylation-sensitive enzymes and PCR or Southern blot analysis, as well as bisulfite sequencing techniques, we examined the methylation status of known RdDM targets: *MEA-ISR*, *FWA* repeats, and the *AtSN1* transposon (Fig. 3A–D). As expected on the basis of its minor de novo methylation defect, *idn1-1* caused a slight reduction in non-CG methylation at all tested loci. Similarly, with the observed data for de novo methylation, *idn2-1* did not display any defect in methylation whereas *idn1-1 idn2-1* double mutants and *idn2-1 idn1-1 idn2-1* triple mutants showed a drastic reduction in non-CG methylation levels. Again, this severe reduction was comparable to that observed in the *idn2-1* mutant, reinforcing the hypothesis that IDNL1 and IDNL2 act redundantly, together with the required factor IDN2.

**IDN2 Complex Acts at a Downstream Step of RdDM.** To determine where in the pathway IDNL proteins are acting, we analyzed the abundance of siRNAs at several loci. IDN2 complex members

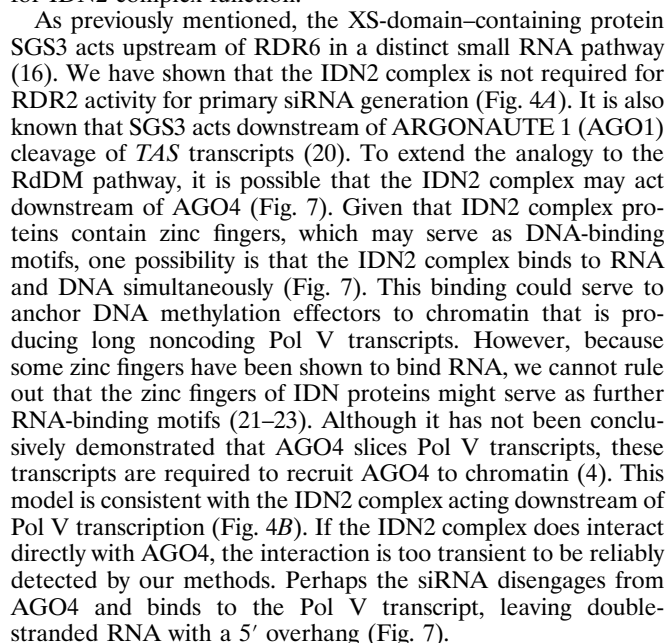
contain the double-stranded RNA-binding XS domain, shared with another *Arabidopsis* protein, SGS3 (15). SGS3 acts upstream of RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) in a small RNA pathway that is distinct from RdDM (16). However, in previous work we have shown that IDN2 is not required to generate 24-nt siRNAs associated with RdDM and thus does not act upstream of RDR2 (5). The *idn1-1 idn2-1* double mutant has a similar effect to *idn2-1*, in that siRNAs are reduced at some loci, but not eliminated, as is observed in mutants in *NUCLEAR RNA POLYMERASE D 1* (*NRPD1*)—the largest subunit of Pol IV (Fig. 4A). The siRNA levels were not further decreased in the triple-mutant background, providing strong confirmation that the IDN2 complex acts at a downstream step from siRNA biogenesis. Instead, the siRNA pattern is reminiscent of the *NUCLEAR RNA POLYMERASE E 1* (*NRPE1*) mutants—the largest subunit of Pol V—which affects siRNA levels at a subset of RdDM targets (17).

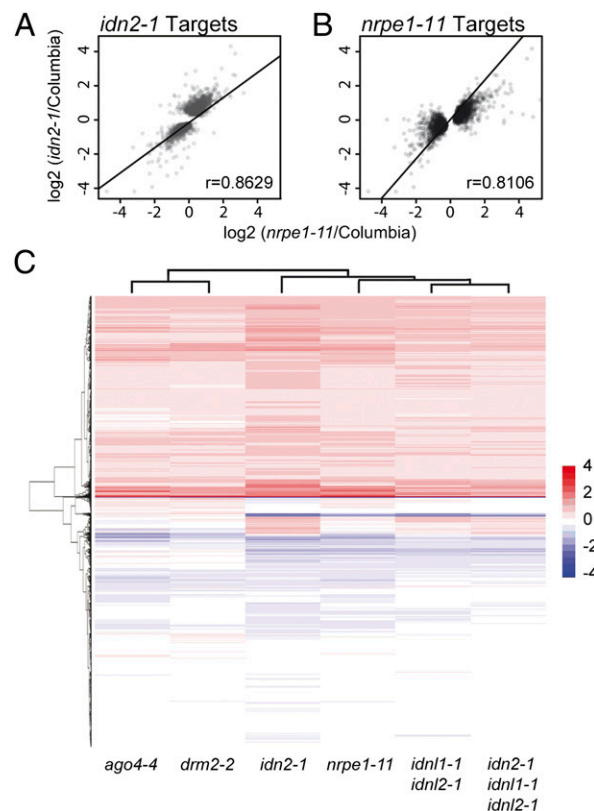
We next wanted to determine whether the IDN2 complex is required for the accumulation of Pol V transcripts. Pol V has been shown to be an active polymerase that transcribes intergenic noncoding (IGN) regions that are necessary for the recruitment of downstream RdDM components such as AGO4 (3, 4). Using quantitative reverse-transcriptase PCR (Q-RT-PCR), we tested the accumulation of IGN transcripts at two loci—*IGN5* and *MEA-ISR*—in *idn1* single mutants, the *idn1-1 idn2-1* double mutant, and the *idn2-1 idn1-1 idn2-1* triple mutant. Comparison with wild-type levels showed no significant differences, placing the IDN2 complex downstream of the production of IGN transcripts (Fig. 4B).

**Genome-Wide Methylation and Expression Analysis.** To gain a broader understanding of how the IDN2 complex affects the *Arabidopsis* epigenome, we performed shotgun bisulfite sequencing in wild-type Columbia, the various *idn* mutants, and *drm2-2* as a positive control (18). We defined differentially methylated regions (DMRs, Dataset S1) that showed reduced CHH context methylation in the *drm2-2* and *idn* libraries relative to Columbia and plotted the respective densities across the five chromosomes in the *Arabidopsis* genome (Fig. 5A). We found that the patterns of DMRs were markedly similar for all mutants tested, indicating that there are few, if any, loci affected by the IDN2 complex outside of those affected by DRM2. This last point is further supported by comparing the *drm2-2* and *idn2-1* DMRs (Fig. 5B). We observed a striking overlap between the regions identified in these datasets, which is consistent with the *idn2-1* methylation effect observed at the known RdDM target loci tested (Fig. 3). Moreover, when analyzing the methylation state of the “nonoverlapping regions,” we do in fact see a somewhat reduced DNA methylation state in *idn* mutants at *drm2-2* DMRs and vice versa (Fig. 5C). This result indicates that virtually all DRM2 targets are affected by loss of IDN2, even though some were excluded from the threshold used to call overlap of regions (Fig. 5B) due to the stringency cutoffs used to define DMRs.

An important aspect of the genome-wide bisulfite sequencing data comes from comparing the methylation of various *idn* mutants in DMRs defined in both *drm2-2* and *idn2-1* (Fig. 5C). In every instance, the patterns for *idn2-1*, the *idn1-1 idn2-1* double mutant, and the *idn2-1 idn1-1 idn2-1* triple mutant phenocopy one another. The genomic methylation data serve as strong additional evidence that the components of the IDN2 complex are likely to function together and are consistent with our methylation analysis at individual loci (Figs. 2B and C and 3).

We also performed whole-genome mRNA sequencing (mRNA-seq) in the *idn* class of mutants and various other RdDM mutants to better understand the regulatory role of the IDN2 complex (Fig. 6 and Fig. S4). Initially we performed the analysis in floral tissue and observed a direct correlation ( $P < 1 \times 10^{-15}$ )





**Fig. 5.** DMRs identified in *drm2-2* overlap with differentially methylated regions in *idn2* family mutants. (A) Density of DMRs across the genome. (B) Venn diagram showing overlap of *drm2-2* and *idn2-2* DMRs. (C) Boxplots representing methylation levels of each DMR class represented by the Venn diagram for Columbia and mutant genomes.

It is also unclear what purpose binding the Pol V transcripts by the IDN2 complex would serve with regard to DNA methylation. The IDN2 complex does not seem to be required for Pol V transcript degradation, because we did not observe an increased abundance of *IGN* transcripts in the mutant backgrounds (Fig. 4B). If the IDN2 complex serves to bind the junction formed by the hybridization of an AGO4 bound siRNA and a Pol V transcript, this action could serve to integrate the information from both the upstream siRNA generation part of the pathway (driven by the location of Pol IV transcription) and the downstream long noncoding RNA portion of the pathway (driven by the location of Pol V transcription). The IDN2 complex could then serve as a signaling molecule for the recruitment or activation of chromatin modification enzymes, ultimately culminating in the recruitment or activation of DRM2 to methylate DNA. Continued studies on the role of XS+XH-containing proteins will be an exciting area of research in understanding RNA-mediated transcriptional control.

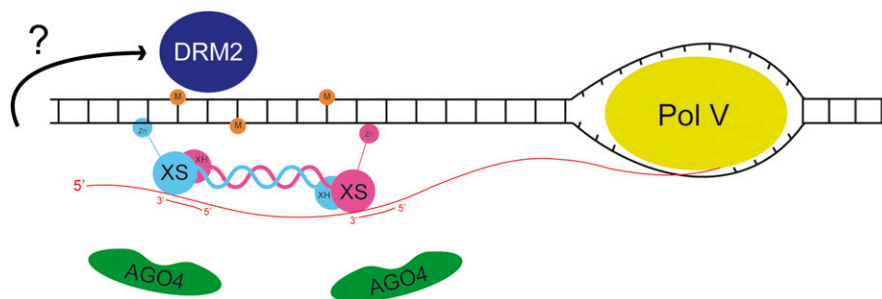
**Plant Materials.** All plants used in this study are the Columbia ecotype and grown under long-day conditions. The following mutant lines were used:

**Protein Expression and Purification.** The PCR-amplified cDNA fragment of *Arabidopsis* IDN2 (120–292) encoding the XS domain and a small segment of coiled-coil region was cloned in pET-Sumo vector (Invitrogen) and overexpressed in *Escherichia coli* BL21 (DE3) cells. The protein was purified from the soluble fraction by a Ni-affinity column, which was followed by overnight treatment with SUMO protease Ulp1 at 4 °C to cleave the His<sub>6</sub>-SUMO tag. Cleaved His<sub>6</sub>-SUMO tag was removed from the protein by a second round of Ni-affinity column chromatography. Protein was further purified by gel-filtration chromatography, using a HiLoad 16/60 Superdex-75 prep grade column (GE Healthcare). Purified proteins were concentrated to 15–20 mg/mL in 25 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 1 mM DTT; immediately frozen in liquid N<sub>2</sub>; and stored at –80 °C. L-selenomethionine (Se-Met)-labeled protein for ab initio phasing was produced by feedback inhibition of the methionine synthesis pathway.

**Crystallization and Structure Determination.** Crystals of IDN2 (120–292) were grown using the vapor-diffusion method by mixing the protein with an equal volume of reservoir solution containing 0.1 M BIS-Tris, pH 6.5, and 2.0 M ammonium sulfate. Small crystals appeared overnight at 20 °C and grew to full size within 1 wk. For data collection, crystals were flash frozen (100 K) in reservoir solution supplemented with 20% (vol/vol) ethylene glycol. Diffraction datasets were collected on a 24-ID-C beamline at the Advanced Photon Source (APS). Datasets were integrated and scaled using the HKL2000 suite (24).

Our attempt to solve the structure of IDN2 (120–292) using datasets collected on Se-Met-labeled protein-containing crystal was unsuccessful due to the presence of only one methionine (lack of sufficient phasing power).





**Fig. 7.** Model for the role of the IDN2 complex in RdDM. Pol V produces a transcript that recruits siRNA-loaded AGO4. The siRNA loaded in AGO4 hybridizes with the nascent Pol V transcript, and the AGO4 protein is released. The XS domain of IDN2 (magenta) is able to recognize this species of RNA. In this interpretation, IDNL1/IDNL2 (cyan) forms a complex with IDN2 in an antiparallel orientation along their respective coiled coils. It is possible that the XS domain of IDNL1/IDNL2 binds to a separate siRNA–Pol V transcript hybrid as well. The riboprotein complex is stabilized to chromatin by the IDN2 complex proteins binding to DNA via their zinc fingers. Through an unknown mechanism, the IDN2 complex and bound RNA promote DNA methylation by DRM2.

Phasing was finally carried out using an IDN2 construct having residues 120–270 with residues Leu198 and Thr247 mutated to methionine to improve the phasing power (Table S1). The structure of Se-Met–labeled IDN2 (120–270; L198M and T247M) was determined by the single-wavelength anomalous dispersion (SAD) method, using Phenix.Autosol (25). Phase improvement was carried out using density modification, producing a clearly interpretable electron density map from which an initial model was built manually using Coot (26). The structure of IDN2 (120–292) was solved by the molecular replacement program MOLREP (27), using the structure of the Se-Met–labeled IDN2 (120–270; L198M and T247M). The model was completed using several rounds of manual model building in Coot (26) and refinement using Phenix. Refine (25). The majority of the model has a clear and well-interpretable electron density map with the exception of a few solvent-exposed side chains, which were omitted in the final model. The geometry of the final model was checked using Procheck (28). The data collection and refinement statistics for IDN2 (120–292) are given in Table S1.

**Southern Blots.** Approximately 4  $\mu$ g of genomic DNA was separated on a 1% agarose gel and then transferred to Hybond N+ membranes. We blocked and washed the blot according to manufacturer's instructions (GE Healthcare). We probed the membranes using PCR products radiolabeled with [ $\alpha$ - $^{32}$ P]dCTP, using the Megaprime DNA Labeling System (GE Healthcare; RPN1606). Primers used for probe amplification are listed in Table S2.

**Bisulfite Sequencing and Analysis.** For sodium bisulfite sequencing, DNA was treated using the EZ DNA Methylation Gold kit (Zymo Research) by following the manufacturer's instructions. Amplified PCR fragments from each analyzed locus were cloned into pCR2.1-TOPO (Invitrogen) and sequenced. We analyzed 15–22 clone sequences per sample, using Lasergene SeqMan software. To distinguish the *FWA* transgene from the endogene, we destroyed a BglII restriction site in the transgenic copy in the region of PCR amplification. We then bisulfite treated genomic DNA of transgenic plants following a BglII digestion (37 °C, overnight), which prevented amplification of the endogenous gene. Additionally, the transgenic copy of *FWA* was derived from the Landsberg ecotype; thus we could distinguish between the transgene and the endogene on the basis of the existence of three single-nucleotide polymorphisms within the amplicon in case BglII digestion was not complete. Primers used for amplification are listed in Table S2.

**HaeIII Cutting Assay.** Analysis of asymmetric methylation at the *At5N1* locus was performed exactly as described in ref. 29. Primers used for amplification are listed in Table S2.

**Flowering Time.** We measured flowering time of plants as the total number of leaves (rosette and cauline leaves) developed by a plant at the time of flowering.

**Generation of Transgenic Plants.** Transgenic plants were generated as described in ref. 30.

**Gel Filtration.** We collected about 300 mg of young inflorescence tissue and homogenized it in IP buffer and then spun it in microfuge tubes for 5 min at 4 °C at 16,000  $\times$  g. Then we transferred the supernatant to a fresh tube and spun again. The supernatant was then filtered through a 0.2- $\mu$ m filter

and 500 mL was loaded onto a Superdex 200 (GE Healthcare; 17-5175-01). Two hundred fifty-milliliter fractions were collected and loaded on 4–12% (vol/vol) SDS/PAGE and then probed with anti-FLAG antibody, following standard Western blot procedures.

**Affinity Purification and Mass Spectrometric Analysis.** Approximately 10 g of flower tissue from transgenic 9xMyc-BLRP-IDN2, IDNL1-9xMyc, or Columbia (negative control) was ground in liquid nitrogen and resuspended in 50 mL of lysis buffer [LB: 50 mM Tris, pH 7.6, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 10% glycerol, 0.1% Nonidet P-40, 0.5 mM DTT, 1 mg/mL pepstatin, 1 mM PMSF, and 1 protease inhibitor mixture tablet (Roche; 14696200)]. Each supernatant was incubated at 4 °C for 2.5 h with 200 mL of Dynabeads MyOne Streptavidin C1 (Invitrogen) for 9xMyc-BLRP-IDN2 or Monoclonal 9E10 agarose beads (Covance; AFC-150P) for IDNL1-9xMyc; both incubations were also used for the negative control. The respective bead-bound complexes were then washed twice with 40 mL of LB and five times with 1 mL of LB. For each wash, the beads were rotated at 4 °C for 5 min. Proteins were then released from the Streptavidin beads by 3C cleavage or from the 9E10 agarose beads with two 10-min incubations with 400  $\mu$ L 8 M urea. Mass spectrometric analyses were conducted as described in ref. 11.

**Coimmunoprecipitation Analysis.** Coimmunoprecipitation and Western blotting were performed as described in ref. 31. IP was performed with M2 Flag agarose (50% slurry, Sigma; A2220). Western blotting was performed with ANTI-FLAG M2 Monoclonal Antibody-Peroxidase Conjugate (Sigma; A8592) and c-Myc 9E10 mouse monoclonal antibody (Santa Cruz Biotechnology; sc-40).

**RNA Analysis.** Pol V transcript assays and siRNA Northern blots were performed exactly as described in ref. 31. Sequences used for amplification and probing are listed in Table S2.

**Genome-Wide RNA Sequencing and Computational Analysis.** We prepared mRNA for Illumina sequencing from both floral and 3-wk-old leaf tissue. A summary of read counts for each library can be found in Dataset S3.

**Floral tissue.** Total RNA was isolated from mixed-stage inflorescence tissues of Columbia and the various *idn* mutants, using TRIzol reagent (Invitrogen). Total RNA (10  $\mu$ g) for each sample was used for purifying the poly(A)-containing mRNA molecules, RNA amplification, and synthesis of double-stranded cDNAs that were ligated to adapters, thus preparing the libraries for sequencing on the Illumina GAII. The mRNA-seq library preparation protocol was followed according to manufacturer's instructions. Libraries were sequenced on an Illumina GAII at the Delaware Biotechnology Institute (Newark, DE).

**Three-week-old leaf tissue.** Total RNA was prepared using a TRIzol (Invitrogen) extraction from 0.5 g of 3-wk-old plant aerial tissue. Four micrograms of total RNA was then used to prepare libraries for Illumina sequencing, following the Illumina TruSeq RNA Sample Prep guidelines. Multiplexed samples were sequenced at 50 bp length on an Illumina HiSeq 2000 instrument.

For data analysis, 50-bp sequences called by the Illumina pipeline were mapped to the *Arabidopsis* genome (TAIR8), using Bowtie (32). Only reads mapping uniquely to the genome with a maximum of two mismatches were used for further analysis. To quantify changes in gene expression, read counts over each *Arabidopsis* gene model were used to perform Fisher's exact tests between genotypes. False discovery rates (FDR) were estimated

by applying a Benjamini–Hochberg adjustment to resulting Fisher’s *P* values. All statistical analysis was conducted in the R environment.

The mRNA sequence data are available from the National Center for Biotechnology Information (NCBI)’s Gene Expression Omnibus (GEO) and are accessible via GEO Series accession nos. GSE37206 (floral) and GSE36129 (leaf).

**Shotgun Bisulfite Sequencing.** Genomic DNA was extracted from 1 g of 3-wk-old plant aerial tissue, using a DNeasy Plant Maxi Kit (Qiagen). Libraries for bisulfite sequencing were generated and sequenced as described in ref. 33, with the change that sequencing was carried out on an Illumina HiSeq 2000 instrument. A summary of library read counts can be found in [Dataset S3](#). Reads were subsequently mapped to the *Arabidopsis* genome (TAIR8), using the BSseeker wrapper (34) of the Bowtie aligner (32).

For data analysis, only cytosines with 5× coverage in all libraries compared were considered. DMRs were discovered using a sliding-window approach with 200-bp window sliding at 50-bp intervals. Fisher’s exact test was performed for methylated vs. unmethylated cytosines for each context, using the resultant windows, with FDRs estimated using a Benjamini–Hochberg adjustment of Fisher’s *P* values calculated in the R environment. Windows with a FDR < 0.05 were considered for further analysis and windows within 100 bp of each other were condensed to larger regions. Regions were then adjusted to extend to differentially methylated cytosines at each border. A cytosine was considered differentially methylated if it showed at least a twofold reduction in methylation percentage in the mutant. Finally regions

were also filtered to have at least 10 differentially methylated cytosines and have an average twofold reduction in methylation percentage per cytosine.

The bisulfite sequence data are available from the NCBI’s GEO and are accessible via GEO Series accession no. GSE36143.

**Protein Databank Accession Number.** Atomic coordinates and the structure factor for the crystal structure of *Arabidopsis* IDN2 (120–292) have been deposited in the Protein Data Bank under accession code 4E8U.

**ACKNOWLEDGMENTS.** We thank the staff of ID-24-C beamline at the Advanced Photon Source for their help with data collection; the University of California, Los Angeles, Broad Stem Cell Research Center BioSequencing Core Facility; M. Browne and M. Nguyen for their help preparing plant tissue; M. Akhavan for technical assistance; and members of the S.E.J. laboratory for supportive discussions. S.E.J. laboratory research was supported by US National Institutes of Health (NIH) Grant GM60398. I.A. was supported by a postdoctoral fellowship from the Ministerio de Educacion y Ciencia. M.V.C.G. was supported by US Public Health Service National Research Service Award GM07104 and a University of California, Los Angeles, Dissertation Year Fellowship. C.J.H. is a Howard Hughes Medical Institute fellow of the Damon Runyon Cancer Research Foundation. S.F. is a Special Fellow of the Leukemia and Lymphoma Society. Work in the B.C.M. laboratory was supported by National Science Foundation Award 0701745. J.A.W. laboratory research was supported by US NIH Grant GM089778 and funds from the Jonsson Cancer Center at University of California, Los Angeles. D.J.P. laboratory research was supported by the Abby Rockefeller Mauze Trust and the Maloris Foundation. S.E.J. is an investigator of the Howard Hughes Medical Institute.

- Cao X, Jacobsen SE (2002) Role of the Arabidopsis DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr Biol* 12:1138–1144.
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11:204–220.
- Wierzbicki AT, Haag JR, Pikaard CS (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135:635–648.
- Wierzbicki AT, Ream TS, Haag JR, Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* 41:630–634.
- 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.
- Zheng Z, et al. (2010) An SGS3-like protein functions in RNA-directed DNA methylation and transcriptional gene silencing in Arabidopsis. *Plant J* 62:92–99.
- Fukunaga R, Doudna JA (2009) dsRNA with 5′ overhangs contributes to endogenous and antiviral RNA silencing pathways in plants. *EMBO J* 28:545–555.
- Glick E, et al. (2008) Interaction with host SGS3 is required for suppression of RNA silencing by tomato yellow leaf curl virus V2 protein. *Proc Natl Acad Sci USA* 105:157–161.
- Qin Y, Ye H, Tang N, Xiong L (2009) Systematic identification of X1-homologous genes reveals a family involved in stress responses in rice. *Plant Mol Biol* 71:483–496.
- Zhang D, Trudeau VL (2008) The XS domain of a plant specific SGS3 protein adopts a unique RNA recognition motif (RRM) fold. *Cell Cycle* 7:2268–2270.
- Law JA, et al. (2010) A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in Arabidopsis. *Curr Biol* 20:951–956.
- Soppe WJ, et al. (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.
- Chan SW, et al. (2004) RNA silencing genes control *de novo* DNA methylation. *Science* 303:1336.
- Greenberg MV, et al. (2011) Identification of genes required for *de novo* DNA methylation in Arabidopsis. *Epigenetics* 6:344–354.
- Bateman A (2002) The SGS3 protein involved in PTGS finds a family. *BMC Bioinformatics* 3:21.
- Mourrain P, et al. (2000) Arabidopsis SGS2 and SGS3 genes are required for post-transcriptional gene silencing and natural virus resistance. *Cell* 101:533–542.
- Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE (2007) Role of RNA polymerase IV in plant small RNA metabolism. *Proc Natl Acad Sci USA* 104:4536–4541.
- Cokus SJ, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452:215–219.
- Xie M, Ren G, Costa-Nunes P, Pontes O, Yu B (2012) A subgroup of SGS3-like proteins act redundantly in RNA-directed DNA methylation. *Nucleic Acids Res*, 10.1093/nar/gks034.
- Yoshikawa M, Peragine A, Park MY, Poethig RS (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev* 19:2164–2175.
- Burdach J, O’Connell MR, Mackay JP, Crossley M (2012) Two-timing zinc finger transcription factors liaising with RNA. *Trends Biochem Sci* 37:199–205.
- Font J, Mackay JP (2010) Beyond DNA: Zinc finger domains as RNA-binding modules. *Methods Mol Biol* 649:479–491.
- Hall TM (2005) Multiple modes of RNA recognition by zinc finger proteins. *Curr Opin Struct Biol* 15:367–373.
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
- Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221.
- Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
- Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* 66:22–25.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) Procheck - a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291.
- Deleris A, et al. (2010) Involvement of a Jumonji-C domain-containing histone demethylase in DRM2-mediated maintenance of DNA methylation. *EMBO Rep* 11:950–955.
- Clough SJ, Bent AF (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of Arabidopsis thaliana. *Plant J* 16:735–743.
- Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE (2011) SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLoS Genet* 7:e1002195.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
- Feng S, Rubbi L, Jacobsen SE, Pellegrini M (2011) Determining DNA methylation profiles using sequencing. *Methods Mol Biol* 733:223–238.
- Chen PY, Cokus SJ, Pellegrini M (2010) BS Seeker: Precise mapping for bisulfite sequencing. *BMC Bioinformatics* 11:203.



# Supporting Information

Ausin et al. 10.1073/pnas.1206638109

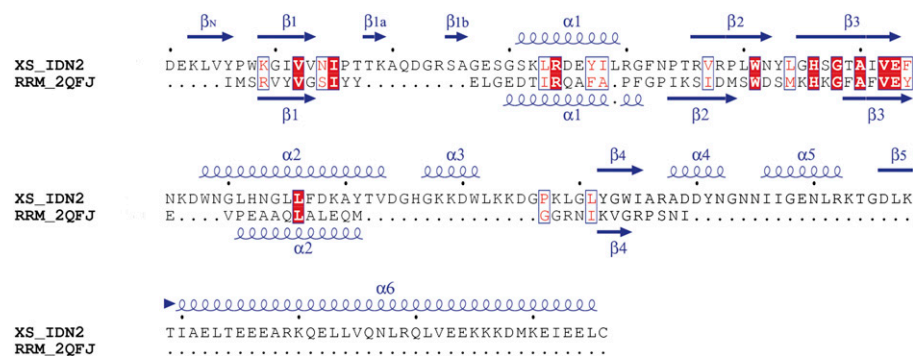


Fig. S1. IDN2 amino acid sequence alignments. (A) Structure-based sequence alignment of the IDN2-XS domain with the RRM domain of the FBP-interacting repressor. Secondary structural elements of the IDN2-XS domain and the RRM domain are shown above and below the aligned sequence, respectively. Additional secondary structural elements observed in the IDN2-XS domain are labeled  $\beta$ N,  $\beta$ 1a,  $\beta$ 1b,  $\alpha$ 3, and  $\alpha$ 4.

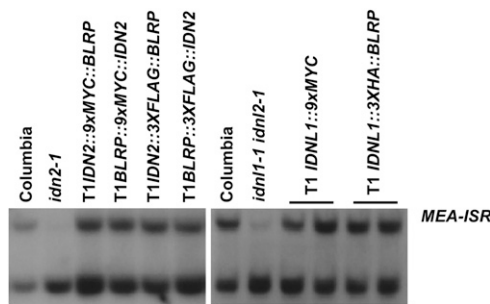


Fig. S2. *idn2-1* and *idn1-1* genetic complementation. Shown is the MEA-ISR Southern blot phenotype of wild-type, *idn2-1* and *idn1-1 idn2-1* mutants, and several individual transgenic lines expressing either the IDN2::tag or the IDN1::tag.







**Dataset S1.** Genomic locations of differentially methylated regions (DMRs) defined for this study

[Dataset S1 \(XLSX\)](#)

**Dataset S2.** Genomic locations and functional descriptions of genes defined as differentially methylated for this study

[Dataset S2 \(XLSX\)](#)

**Dataset S3.** Counts of total and mapping reads for Illumina libraries used in this study

[Dataset S3 \(XLSX\)](#)