DNA hypermethylation and condensed chromatin correlate with chromosome-specific rRNA gene silencing in Arabidopsis

Gireesha Mohannath[1,*], Anastasia McKinlay[2,3,†], Ramya Enganti[2,3,†], Navinchandra Venkata Puppala[1,†], Gargi Prasad Saradadevi[1,†], Craig S. Pikaard[2,3,] and Todd Blevins[4]

[1]Department of Biological Sciences, Birla Institute of Technology and Science-Pilani, Hyderabad campus, Hyderabad, Telangana, 500078, India.

[2]Department of Biology and Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, IN, 47405, United States.

[3]Howard Hughes Medical Institute, Indiana University, Bloomington, IN, 47405, United States.

[4]Institut de Biologie Moléculaire des Plantes, CNRS, Université de Strasbourg, F-67084 Strasbourg, France.

* To whom correspondence should be addressed.

Tel: +91-40-66303697

+1- 812-855-1874

Fax: +91-40-66303998

Email: gireesha@hyderabad.bits-pilani.ac.in

† Joint second authors

ABSTRACT

In eukaryotes, hundreds of ribosomal RNA (rRNA) genes are clustered at chromosomal loci called nucleolus organizer regions (NORs). *Arabidopsis thaliana* has two NORs, one on chromosome 2 (*NOR2*) and the other on chromosome 4 (*NOR4*). Each NOR consists of ~ 400 rRNA gene copies. We recently showed that rRNA gene subtypes that map to *NOR2* are silenced during development, whereas those that map to *NOR4* are active. In several DNA methylation mutants of Arabidopsis, we show disruption of the *NOR2* gene silencing to varying degrees. Significantly, the highest disruption of *NOR2* gene silencing correlates with a maximum loss of cytosine methylation in the CHH context followed by the CG context, independent of RNA-directed DNA methylation (RdDM). Next, we show in Col-0 that *NOR2* genes are relatively hypermethylated and *NOR4* genes are hypomethylated using multiple methylation analysis of genomic DNA carried out with different types of methylation-sensitive restriction enzymes. We demonstrate similar differential methylation status between *NOR2* and *NOR4* genes in an introgression line named $Col^{Sf\text{-}NOR4}$, which carries *NOR2* from Col-0 and *NOR4* from ecotype Sf-2. Lastly, using Tn5 transposon-mediated transposition into native chromatin, we show that *NOR2* gene chromatin is in more condensed state than *NOR4* gene chromatin.

INTRODUCTION

In eukaryotes, hundreds to thousands of 45S ribosomal RNA (rRNA) genes are arranged as tandem repeats at chromosomal loci known as nucleolus organizer regions (NORs) (Fig. 1A). RNA polymerase I (Pol I) transcribes rRNA genes to produce 45S pre-rRNA transcripts that are then processed into 18S, 5.8S, and 25-28S rRNAs that comprise the catalytic core of ribosomes, which synthesize proteins in cells. (1-3). Although eukaryotic genomes have many rRNA genes, not all of them are transcribed all the time. Rather, the number of genes transcribed varies during development, dictated by the developmental stage-specific need for ribosomes and protein synthesis (4-6). rRNA 'dosage control' involves selective gene silencing rather than selective gene activation (7). Our recent findings in *Arabidopsis thaliana* (ecotype Col-0) indicate that such selective silencing is determined by rRNA gene chromosomal location and not rRNA gene sequence variation (8,9).

rRNA genes in *A. thaliana* are organized into two NORs; one NOR on chromosome 2 (*NOR2*) and the other NOR on chromosome 4 (*NOR4*). Each NOR consists of ~ 400 rRNA gene copies (10,11). The 18S, 5.8S, and 25S rRNA sequences are identical among 45S rRNA genes, but some variation exists in the internal transcribed spacer (ITS) and the external transcribed spacer (ETS) regions. Four distinct variants had been defined based on the variation in 3' ETS region named VAR1 through 4 (Fig. 1B-C) (12). Of these, VAR1, VAR2, and a subset of VAR3 (VAR3a) lack HindIII sites in the 3' ETS region, but other remaining variants (VAR3b,c and VAR4) have a HindIII site (8,9) (Fig.1C).

Earlier, we had shown that the rRNA gene variants that map primarily to *NOR2* (VAR1, VAR3a) are selectively silenced as a consequence of dosage control during development. In contrast, the variants that map primarily to *NOR4* (VAR2, VAR3c, and VAR4) remain active (Fig. 1D)(8). Silencing of many genes in eukaryotes involves cytosine methylation and chromatin condensation. However, relative role of cytosine methylation in all three contexts (CG, CHG and CHH) in selectively silencing *NOR2* genes is unknown. In *A. thaliana*, cytosine methylation is carried out by different DNA methyltransferases. DECREASE IN DNA METHYLATION 1 (DDM1) is a SWI2/SNF2 family chromatin remodeler which acts as a master regulator of DNA methylation (13). It promotes DNA methyltransferases' access to the H1 histone-containing heterochromatin to enable silencing of transposable elements (14). Loss of DDM1 function resulted in genome-wide demethylation at 70% of all methylated cytosines (15,16). The *ddm 1-2* mutation is known to reduce cytosine methylation in all three contexts (CG, CHG and CHH) (14). In plants, METHYLTRANSFERASE 1 (MET1) is involved in cytosine methylation in the CG context, CHROMOMETHYLASE 3 (CMT3) in the CHG context and DOMAINS REARRANGED

METHYLASE 2 (DRM2) and CMT2 in the CHH context (17). Cytosine methylation in the CHH context is facilitated by DDM1 and mediated by CMT2 independent of RNA-directed DNA methylation (RdDM) (14). *cmt2-3* is known to reduce most of cytosine methylation in the CHH context (14). An earlier study investigated effects of *drm2*, *cmt3* and *met1* mutants on rRNA gene silencing (18). Of these, only *met1* mutants disrupted *NOR2* gene silencing while others barely had any effect (18). In this study, we analyzed *ddm1* and *cmt2* mutants along with *drm2*, *cmt3*, and *met1* mutants for their relative effect on rRNA gene silencing. We show that disruption of *NOR2* gene silencing is the highest in *cmt3* mutants, followed by *ddm1* mutants.

Further, to determine if DNA methylation and chromatin condensation are involved in the selective silencing of rRNA genes, we analyzed the cytosine methylation and chromatin status of *NOR2*- and *NOR4*-associated genes using multiple molecular approaches. First, using methylation-sensitive restriction digestion analyses, we show hypermethylation of silent rRNA genes at *NOR2* and hypomethylation of the active rRNA genes at *NOR4*. We demonstrate similar results in an Arabidopsis introgression line named $Col^{Sf\text{-}NOR4}$ containing NORs from two different ecotypes (Col-0 and Sf-2) (8). Lastly, using Tn5 transposase-mediated transposition into native chromatin (19), we show that active *NOR4* genes exist in a more open chromatin than the silenced *NOR2* genes.

MATERIAL AND METHODS

Plant material and growth

A. thaliana ecotype Col-0 (stock no. CS22681), and mutants *met1-7* (stock no. SALK_076522), *cmt2-3* (stock no. SALK_012874C), *cmt3*, and *hda6-7* (stock no. CS66154) were obtained from the Arabidopsis Biological Resource Center (ABRC; http://abrc.osu.edu). *drm2-2b* (stock no. SAIL_70_E12) and $Col^{Sf\text{-}NOR4}$ were provided by Andrejez Wierzbicki and Scott Michaels, respectively, and *ddm1-2* and *met1-1* were provided by Eric Richards. Plants were grown in soil in a plant growth room under a 16-h light/8-h dark photoperiod cycle with 50–55% humidity and temperature of 23–25°C.

RNA isolation

Four-week-old leaf tissue of individual plants was frozen in liquid nitrogen and ground into a fine powder. RNeasy plant minikit (Qiagen) was used to extract total RNA from ground powder. Total RNA was then treated with Turbo DNA-free kit (Invitrogen) to eliminate contaminating DNA.

PCR and RT-PCR assays

For PCR of genomic DNA to amplify all rRNA variants, PCR amplification was conducted using 22-30 cycles (depending on the source of DNA) of 30 sec at 95°C, 30 sec at 55°C–60°C, and 60 sec at 72°C. For RT–PCR, the reverse transcription reaction was performed using 1 µg of total RNA using SuperScript III reverse transcriptase (Invitrogen). One microliter (~100 ng) of reverse transcription product was then used for PCR amplification of rRNA variable regions or an actin control using 30 cycles of 30 sec at 94°C, 30 sec at 55°C, and 60 sec at 72°C. In all the PCR or RT-PCR reactions, primers 5'allvar and 3'allvar were used to amplify rRNA variable regions (all the 4 variant types). PCR and RT–PCR products were resolved by electrophoresis on 2%–2.5% agarose gels in TBE buffer. Sequences of primers used in this study are given in Figure S1.

Generation of methylation heat maps

Whole genome bisulfite sequencing data (GSE39901) published by Stroud et al. (2013)(20), was used to assess the methylation state of rDNA in various methylation mutants. The sequencing data were analyzed using Methylpy (21). The reads were mapped to a converted 45S rDNA reference sequence allowing a single mismatch. The methylation state of all cytosines in CG, CHH and CHG contexts was then analyzed and represented as a fraction of methylated cytosines over the total cytosines in each context. Replicates for a given genotype were merged into a single file.

DNA isolation from leaf tissue and agarose gel pieces (gel elution).

Four-week-old individual plant leaf tissue was incubated for 15 min at 99°C in extraction buffer (200 mM Tris-HCl at pH 8.0, 250 mM NaCl, 25 mM EDTA, 0.5% SDS). Genomic DNA from the supernatant was recovered by ethanol precipitation and resuspended in 60 µL of Tris-HCl (pH 8.0). In some cases, DNA was isolated from leaf tissue of 4-wk-old plants by using Illustra Nucleon Phytopure extraction kit reagents (GE Healthcare). DNA was isolated from agarose gel pieces using QIAquick gel extraction kit (Qiagen).

Nuclei isolation

Nuclei were isolated using 10 g tissue from 3-week-old Col-0 plant seedlings as described previously (22) with a few modifications. Briefly, 3-week-old plants were harvested and washed in ice-cold water. Then plants were submerged in cold diethyl ether for 5 min and washed twice in ice-cold water. Homogenization of plant material was performed in 100 mL of nuclei isolation buffer (1M sucrose, 50 mM Tris-HCl pH 7.2, 5 mM $MgCl_2$, 5 mM KCl, 10

mM 2-mercaptoethnol, 0.2 mM PMSF) with a motorized homogenizer (Fisher Scientific Powergen 700). Homogenate was filtered through 1 layer of Miracloth and 1 layer of nylon mesh (pore size of 50 μm). After centrifugation of the filtered homogenate at 14,000 g for 15 min at 4°C the pellet was resuspended in 10 ml of nuclei isolation buffer and prepared for loading into discontinuous Percoll gradient (15%, 30%, 45%, and 60%) by adding one volume of "100% Percoll solution" (34.23 g sucrose, 5 mL 1 M Tris-HCl pH 7.2, 0.5 mL 1 M $MgCl_2$, 0.5 mL 1 M KCl, 34 μL 2-mercaptoethnol, 100 μL 2 mM PMSF and Percoll to 100 mL) to 19 volumes of nuclei. Gradients were centrifuged at 2,000 rpm for 10 min, then 8,000 rpm for 20 min. Nuclei were harvested from 30%/45% and 45%/60% boundaries and pooled. Nuclei were diluted with 5 volumes of nuclei isolation buffer, mixed by inversion and collected by centrifugation at 1,500 g for 10 min at 4°C. A second wash of nuclei was performed with 25 volumes of nuclei isolation buffer. Final wash was done with 50 mL of Nuclease digestion buffer (50 mM Tris-HCl, pH 8.0, 0.3 M sucrose, 5 mM MgCl2, 1.5 mM NaCl, 0.1 mM CaCl2, 5 mM beta-mercaptoethanol). The resulting pellet was resuspended in 2 mL of Nuclease digestion buffer.

Transposase-accessible chromatin assay

The assay was carried out based on the methodology described by Buenrostro et al. (19). Oligos (5'-Ez-Tn5 and 3'-Ez-Tn5) were annealed by heating a mixture of equal volumes of 100 mM oligos at 95 °C for 5 min, followed by turning the heating off and incubating for 10 min. Transposition reactions were carried out using 2 μL of purified nuclei, 20 μL of EZ-Tn5 transposase with varying amount of the oligos 5'-Ez-Tn5 and 3'-Ez-Tn5, according to the manufacturer's instructions (Epicentre). Incubation was performed at 37°C for 90 min. DNA was isolated using cetyl trimethyl ammonium bromide (CTAB) method. PCR1 and PCR2 were carried out as described above. The sequences of primers and oligos used in this study are listed in Figure S1.

RESULTS

Selective *NOR2* gene silencing is suppressed in DNA methylation mutants

To determine the relative impact of loss of cytosine methylation in three different contexts (CG, CHG and CHH) on selective *NOR2* gene silencing, we analysed rRNA variant expression in several DNA methylation mutants. We performed RT-PCR analysis using RNA isolated from 4-week-old leaf tissues of various mutants. VAR1 type rRNA genes constitute

~50% of total 45S rRNA genes, and most of them map to *NOR2*, and undergo selective silencing during development (8,12). Among the DNA methylation mutants, VAR1 expression levels was found to be the highest in *cmt2-3* and *ddm1-2* mutants, followed by *met1-7* (Fig. 2 and Fig. S2). We confirmed high VAR1 expression in these mutants by analysing multiple independent mutant plants using RT-PCR analysis (Fig. S2). However, *cmt3* and *drm2* mutants displayed low levels of VAR1 expression, if any (Fig. 2).

Using published whole-genome bisulfite sequencing data (20), we generated heatmaps of cytosine methylation for a 45S rRNA reference gene sequence (Fig. S3) using Methylpy (21) (Fig. 3). The cytosine methylation heatmaps confirm loss of most of the cytosine methylation in the CHH context in *cmt2-3* and *ddm1-2* mutants (Fig. 3), in which disruption of *NOR2* gene silencing was observed to the highest degree (Fig. 2). The *hda6-7* mutant was included as a positive control, as it is known to disrupt *NOR2* gene silencing to a high degree (8,23). Similarly, heatmaps confirm the loss of most of the cytosine methylation in the CG context in *ddm1-2* and *met1-7* mutants (Fig. 3).

***NOR2* genes are hypermethylated while *NOR4* genes are hypomethylated in Col-0**

We carried out methylation analysis of rRNA variants using HpaII-digested genomic DNA of Col-0 and *met1-7* mutants (Fig. 4). Because the rRNA variant-defining 3' ETS region lacks HpaII sites, HpaII digestion does not impact PCR amplification in this region. Therefore, we chose to separate HpaII-digested and HpaII-undigested fractions of genomic DNA by resolving them on an agarose gel. The fragments that remained at the top of the gel (>20 kb) constituted the HpaII-undigested fractions while those resolved at 0.6-1.0 kb range constituted the HpaII-digested fractions (Fig. 4A, left panel). Both the digested and the undigested fractions were gel-eluted, followed by PCR amplification of rRNA variants. Through this analysis, we were able to identify the rRNA variants that were enriched in HpaII-digested or undigested fractions. The PCR data revealed VAR2 genes from *NOR4* to be the most predominant in Col-0 among the HpaII-digested products, indicating relative hypomethylation of *NOR4* genes compared to *NOR2*-localized VAR1 genes (Fig. 4A, lanes 1 and 2, right panel). However, in *met1-7* mutants, VAR1 genes were the most predominant among the HpaII-digested products due to loss of cytosine methylation in all four rRNA variants (Fig. 4A, lanes 3 and 4, right panel,).

In another experiment, we subjected gDNA to digestion by McrBC, an enzyme which preferentially digests methylated DNA. McrBC digestion of gDNA followed by PCR

amplification of rRNA variants revealed that VAR1 was digested the most and VAR2 the least (Fig. 4B), consistent with hypermethylation of genes at *NOR2* and hypomethylation of genes at *NOR4*. HpaII digestion indicates cytosine methylation levels in the CG context only, but McrBc digestion indicates cytosine methylation levels in all three contexts (CG, CHG, and CHH).

Methylation maps of centromere-proximal regions of NORs show that ~ 75 kb of the *NOR2*-adjoining region is heavily methylated compared to the *NOR4*-adjoining region (Fig. S4), which has a protein-coding gene just 3 kb away (24). It is possible that the heavily methylated *NOR2*-adjoining region might act as a source for spreading methylation-mediated silencing into the *NOR2* genes, but this hypothesis remains to be tested.

***NOR2* genes are hypermethylated while *NOR4* genes are hypomethylated in Col$^{Sf\text{-}NOR4}$**

An introgression line named Col$^{Sf\text{-}NOR4}$ consists of *NOR2* from Col-0 (VAR1 and VAR3 genes) and *NOR4* from ecotype Sf-2 (8) (Fig. 5A). Sf-2 has only VAR1 type genes, but they can be differentiated from VAR1 genes of Col-0 by HindIII digestion. Sf-2 VAR1 genes have a HindIII site in their 3' ETS region, while Col-0 VAR1 genes lack it (8). In Col$^{Sf\text{-}NOR4}$, Sf-2 *NOR4* genes are active and Col-0 *NOR2* genes are selectively silenced (8). The methylation status of rRNA variants was determined using similar experiments as described above. In one experiment, genomic DNA of Col$^{Sf\text{-}NOR4}$ was subjected to HpaII digestion, followed by gel elutions of the undigested and the digested fractions. The gel-eluted DNA samples were then subjected to PCR amplification of rRNA variants followed by HindIII digestion. In control samples (no HpaII), HindIII digestion of PCR-amplified rRNA variants resulted in the expected products, which include digested Sf2-VAR1 genes, undigested Col-0 VAR1 genes, and partially digested Col-0 VAR3 genes (Fig. 5B, lanes 1 and 2, top right gel image, and Fig. 5C) (8). However, HindIII digestion of the PCR products obtained from the gel-purified HpaII-digested fraction as template, revealed that Col-0 VAR1 genes predominantly constitute the HpaII-undigested fraction (Fig. 5B, lanes 3 and 4, top right gel image, and Fig. 5C). We also observed a significant increase in the Sf-2 VAR1 genes in the HpaII-digested fraction (Fig. 5B, lanes 3 and 4, bottom right gel image, and Fig. 5C) compared to the undigested fraction (Fig. 5B, lanes 3 and 4, top right gel image, and Fig. 5C). These data indicate hypermethylation of Col-0 *NOR2* genes and hypomethylation of Sf-2 *NOR4* genes, despite the presence of the VAR1 genes on both the NORs in the Col$^{Sf\text{-}NOR4}$ introgression line.

In another experiment, gDNA of Col$^{Sf\text{-}NOR4}$ was first digested with McrBc, followed by PCR amplification of rRNA variants. These PCR products were then digested with HindIII and resolved using regular gel electrophoresis. HindIII-digested control PCR products (no McrBc digestion) behaved as expected (Fig. 5D&E, lanes 1 and 2). However, the McrBC digestion patterns of the rRNA variants (Fig. 5D&E) were the opposite of the experimental results involving HpaII digestion (compare to Fig. 5B&C). Predominantly, the Sf-2 VAR1 genes constituted the McrBc-undigested fraction, while the Col-0 VAR1 genes constituted the McrBc-digested fraction (Fig. 5D&E, lanes 3 and 4). We also observed complete digestion of the Col-0 VAR3 genes by McrBC (Fig. 5D, lane 3). These data suggest that *NOR4* genes are hypomethylated and *NOR2* genes are hypermethylated in Col$^{Sf\text{-}NOR4}$ despite the presence of VAR1 type of genes in both NORs.

***NOR2* genes are organized in more condensed chromatin than *NOR4* genes**

To investigate the chromatin condensation status of *NOR2* and *NOR4* genes, we carried out Tn5 transposase-mediated transposition using nuclei isolated from 3-week-old Col-0 plants. Tn5 is a prokaryotic hyperactive transposase that has been developed as a tool to insert short DNA sequences into open chromatin regions (19,25,26). A pair of 19 bp IS50 repeats that serve as ‘outside ends’ (OEs) sequences are sufficient for recognition by Tn5 transposase and transposition of OEs-containing sequences into open chromatin (27). We designed short DNA sequences consisting of OEs for Tn5 recognition and used them in a native chromatin transposition assay. Purified Col-0 nuclei were incubated with either no or 5 pmol oligos containing Tn5 recognition sites for transposition, followed by DNA extraction from the nuclei and PCR analysis of rRNA variants. The PCR data did not reveal differences between *NOR2* and *NOR4* genes (data not shown). We then tried an indirect approach to focus on the chromatin accessibility state of the promoter regions. In this approach, the DNA samples isolated from nuclei subjected to Tn5 transposition were subjected to two rounds of PCR amplification. The first round of PCR (PCR1) amplifies ~2.8 kb region that spans the 3’ ETS region, IGS and the promoter of the following gene repeat (Fig. 6B). Due to lack of variant-specific sequence variation in the promoter region, we exploited the proximity of the rRNA variant-defining 3’ ETS region of one rRNA gene to the promoter of the following rRNA gene. The occurrence of rRNA genes as tandem repeats (Fig. 1A) at NORs permitted this approach. The idea is that, if promoters of VAR1 genes (primarily present on *NOR2*) are more condensed than the promoters of VAR2 genes, then the 2.8 kb PCR products should be enriched for VAR1 genes. Because the 2.8 kb PCR products obtained during PCR1

(using primers 5'allvar and p34) can't discriminate between *NOR2* and *NOR4* genes, they were resolved on a 1% agarose gel followed by gel-purification of 2.8 kb amplicons. These amplicons were then subjected to the second round of PCR amplification (PCR2) with primers that either amplify only VAR1 and VAR2 genes (Fig. 6C) or primers that amplify all rRNA variants (Fig. 6D). The results showed severe depletion of *NOR4*-specific VAR2 genes in the reactions with 5 pmol of Tn5 oligos (Fig. 6C&D), indicating depletion of *NOR4*-specific fragments in the PCR1 amplicon pool. This depletion could be presumably due to higher transposition rates in the promoter regions of *NOR4*-specific rRNA variants compared to the *NOR2* variants. Collectively, these data indicate that *NOR4* genes are predominantly present in a more accessible open chromatin state than *NOR2* genes.

DISCUSSION

Our earlier findings revealed that during development, *NOR2* genes are silenced and *NOR4* genes remain transcriptionally active (8,9). In eukaryotes, gene silencing often involves DNA methylation and chromatin condensation. In this study, disrupted *NOR2* gene silencing was observed in most of the cytosine methylation mutants but was found to be the highest in *cmt2-3* and *ddm1-2,* followed *met1-7* mutants (Fig. 2). Notably, *ddm1-2* and *cmt2-3* mutants exhibit the highest loss of CHH methylation (Fig. 3), consistent with an earlier study in Arabidopsis that showed that most asymmetric cytosine methylation is mediated by CMT2 and DDM1, independent of the RNA-directed DNA methylation (RdDM) pathway (14). DRM2 is the methyltransferase primarily involved in RdDM. Although *ddm1-2* loses cytosine methylation in all contexts (CG, CHG, CHH) (15,16) (Fig. 3), *NOR2* gene silencing was not disrupted to the extent of *cmt2-3* (Fig. 2). One possible reason is that loss of CG and/or CHG methylation could dysregulate some of the positive regulators of *NOR2* gene silencing, limiting the disruption of *NOR2* gene silencing in *ddm1-2* mutants to some extent. Nevertheless, the loss of cytosine methylation in the CHH context appears to have the maximum effect on the disruption of *NOR2* gene silencing.

Next, to determine if DNA methylation and chromatin condensation are involved in selectively silencing *NOR2* genes, we analysed DNA methylation and chromatin condensation status of *NOR2* and *NOR4* genes using various approaches. To analyze DNA methylation status of *NOR2* and *NOR4* genes, we carried out rRNA variant analysis upon digestion of genomic DNA by different types of methylation-sensitive restriction enzymes (HpaII and McrBC) (Fig. 4). HpaII does not digest DNA if its recognition site (CCGG) has cytosine methylation either in the CG or CHG context. In contrast, McrBc preferentially

digests DNA if its recognition site ($R^mC(N_{40-3000})R^mC$) consists of methylated cytosines. From these experiments involving two methylation-sensitive restriction enzymes with contrasting mechanisms of detecting cytosine methylation, we find that *NOR2* genes are hypermethylated and *NOR4* genes are hypomethylated (Fig. 4). HpaII digestion reveals the methylation status of cytosines in the CG and CHG context, and McrBc digestion reveals the methylation status of cytosines in all three contexts (CG, CHG, CHH). Therefore, hypermethylation of cytosine on *NOR2* genes could likely have occurred in all three contexts.

We conducted similar experiments using an introgression line named Col-$^{Sf\text{-}NOR4}$ which consists of *NOR2* from Col-0 (VAR1 and VAR3 genes) and *NOR4* from ecotype Sf-2 (8) (Fig. 5). Interestingly, Sf-2 consists of only VAR1 type genes, but they can be differentiated from VAR1 genes of Col-0 by HindIII digestion. We have earlier shown that in Col$^{Sf\text{-}NOR4}$, Sf-2 *NOR4* genes are active and Col-0 *NOR2* genes are selectively silenced (8). Results from the experiments involving digestions with methylation-sensitive HpaII and McrBc showed hypermethylation of *NOR2* genes and hypomethylation of *NOR4* genes (Fig. 5), consistent with the results from the other experiments (Fig. 4). Based on these data, we conclude that *NOR2* genes are hypermethylated compared to *NOR4* genes, and DNA methylation is involved in selectively silencing *NOR2* genes.

Further, we examined the chromatin accessibility of *NOR2* and *NOR4* genes using a Tn5 transposase-mediated transposition into native chromatin. Tn5 is a prokaryotic hyperactive transposase that can insert short DNA sequences into open chromatin regions (19,25,26) where in ‘outside ends’ (OEs) sequences are sufficient for Tn5 recognition and transposition (27). The underlying principle is that the higher the frequency of insertion, the more open the chromatin will be. PCR analysis of DNA obtained from Tn5-treated nuclei did not reveal any difference in chromatin condensation states of *NOR2* and *NOR4* genes. However, an indirect approach involving two rounds of PCR reactions, as described above, revealed that *NOR2* genes are more condensed than *NOR4* genes (Fig. 6).

In conclusion, hypermethylation and chromatin condensation of *NOR2* genes correlates with selective *NOR2* gene silencing. However, despite striking similarities in rRNA gene sequences of *NOR2* and *NOR4* genes, why does *NOR2* genes undergo selective hypermethylation, condensation and silencing? Interestingly, the immediate centromere-proximal region of *NOR2* is composed of transposable elements and transposon remnants that stretch for ~ 60 kb (Fig. S4). The corresponding centromere-proximal regions of *NOR4* do not contain any stretch of transposons or their remnants (Fig. S4). In plants, transposons are silenced through methylation-mediated mechanisms. Could such silencing of the *NOR2*-

adjacent transposon-rich region spread to *NOR2* genes, resulting in their selective silencing? The question remains to be answered.

DATA AVAILABILITY

Whole genome bisulfite sequencing data published by Stroud et al. (2013)(20) with a GEO accession number GSE39901 can be found in NCBI database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39901).

AUTHOR CONTRIBUTIONS:

G.M. and C.S.P. designed research; G.M., A.M., R.E., G.P.S., and T.B. performed research; G.M., A.M., R.E., N.V.P, T.B., and C.S.P. analyzed data; and G.M., A.M., R.E., N.V.P, and C.S.P. wrote the paper.

ACKNOWLEDGEMENT

We thank the members of the Pikaard lab and the Mohannath lab for their critique of the manuscript.

FUNDING

G.M. is thankful to the Birla Institute of Science and Technology (BITS) Pilani, Hyderabad campus for the research grant (BITS/GAU/ACRG/2019/H0576) to G.M. and an Institute Fellowship to G.P.S., and to the Science and Engineering Research Board (SERB), Government of India, for the Ramanujan Fellowship Research Grant (SB-S2-RJN-062-2017) and CRG grant (CRG/2020/002855). Pikaard lab members A.M., R.E. were supported by funds to C.S.P. as an Investigator of the Howard Hughes Medical Institute.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Saradadevi, G.P., Priyadarshini, N., Bera, A. and Mohannath, G. (2020) Together we are on together we are off -a conserved rule for ribosomal RNA (rRNA) gene regulation? *Journal of Plant Biochemistry and Biotechnology*, **29**, 743-753.
2. Turowski, T.W. and Tollervey, D. (2015) Cotranscriptional events in eukaryotic ribosome synthesis. *Wiley interdisciplinary reviews. RNA*, **6**, 129-139.
3. Viktorovskaya, O.V. and Schneider, D.A. (2015) Functional divergence of eukaryotic RNA polymerases: unique properties of RNA polymerase I suit its cellular role. *Gene*, **556**, 19-26.
4. Grummt, I. (2013) The nucleolus—guardian of cellular homeostasis and genome integrity. *Chromosoma*, **122**, 487-497.
5. Hannan, K.M., Sanij, E., Rothblum, L.I., Hannan, R.D. and Pearson, R.B. (2013) Dysregulation of RNA polymerase I transcription during disease. *Biochimica et biophysica acta*, **1829**, 342-360.
6. Moss, T., Langlois, F., Gagnon-Kugler, T. and Stefanovsky, V. (2007) A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cell Mol Life Sci*, **64**, 29-49.
7. Chen, Z.J. and Pikaard, C.S. (1997) Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Genes Dev*, **11**, 2124-2136.
8. Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F. and Pikaard, C.S. (2016) Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in Arabidopsis. *Genes Dev*, **30**, 177-190.
9. Mohannath, G., Pontvianne, F. and Pikaard, C.S. (2016) Selective nucleolus organizer inactivation in Arabidopsis is a chromosome position-effect phenomenon. *Proc Natl Acad Sci U S A*, **113**, 13426-13431.
10. Copenhaver, G.P., Doelling, J.H., Gens, S. and Pikaard, C.S. (1995) Use of RFLPs larger than 100 kbp to map the position and internal organization of the nucleolus organizer region on chromosome 2 in Arabidopsis thaliana. *Plant J*, **7**, 273-286.
11. Copenhaver, G.P. and Pikaard, C.S. (1996) Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in Arabidopsis thaliana, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *Plant J*, **9**, 273-282.
12. Pontvianne, F., Abou-Ellail, M., Douet, J., Comella, P., Matia, I., Chandrasekhara, C., Debures, A., Blevins, T., Cooke, R., Medina, F.J. *et al.* (2010) Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in Arabidopsis thaliana. *PLoS genetics*, **6**, e1001225.
13. Brzeski, J. and Jerzmanowski, A. (2003) Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. *The Journal of biological chemistry*, **278**, 823-828.
14. Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L. and Zilberman, D. (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, **153**, 193-205.
15. Jeddeloh, J.A., Stokes, T.L. and Richards, E.J. (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature genetics*, **22**, 94-97.
16. Vongs, A., Kakutani, T., Martienssen, R.A. and Richards, E.J. (1993) Arabidopsis thaliana DNA methylation mutants. *Science*, **260**, 1926-1928.
17. Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature reviews. Genetics*, **11**, 204-220.
18. Pontvianne, F., Blevins, T., Chandrasekhara, C., Mozgova, I., Hassel, C., Pontes, O.M., Tucker, S., Mokros, P., Muchova, V., Fajkus, J. *et al.* (2013) Subnuclear partitioning of rRNA genes

441 between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Genes Dev*, **27**,
442 1545-1550.
443 19. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. and Greenleaf, W.J. (2013) Transposition
444 of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-
445 binding proteins and nucleosome position. *Nature methods*, **10**, 1213-1218.
446 20. Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V. and Jacobsen, S.E. (2013)
447 Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis
448 methylome. *Cell*, **152**, 352-364.
449 21. Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N.,
450 Nery, J.R., Urich, M.A., Chen, H. *et al.* (2015) Human body epigenome maps reveal
451 noncanonical DNA methylation variation. *Nature*, **523**, 212-216.
452 22. Gaudino, R.J. and Pikaard, C.S. (1997) Cytokinin induction of RNA polymerase I transcription
453 in Arabidopsis thaliana. *The Journal of biological chemistry*, **272**, 6799-6804.
454 23. Blevins, T., Pontvianne, F., Cocklin, R., Podicheti, R., Chandrasekhara, C., Yerneni, S., Braun,
455 C., Lee, B., Rusch, D., Mockaitis, K. *et al.* (2014) A two-step process for epigenetic inheritance
456 in Arabidopsis. *Molecular cell*, **54**, 30-42.
457 24. Lewis, M.S. and Pikaard, C.S. (2001) Restricted chromosomal silencing in nucleolar
458 dominance. *Proc Natl Acad Sci U S A*, **98**, 14536-14540.
459 25. Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B.,
460 MacKenzie, A.P., Caruccio, N.C., Zhang, X. *et al.* (2010) Rapid, low-input, low-bias
461 construction of shotgun fragment libraries by high-density in vitro transposition. *Genome*
462 *biology*, **11**, R119.
463 26. Goryshin, I.Y. and Reznikoff, W.S. (1998) Tn5 in vitro transposition. *The Journal of biological*
464 *chemistry*, **273**, 7367-7374.
465 27. Ahmed, A. (2009) Alternative mechanisms for tn5 transposition. *PLoS genetics*, **5**, e1000619.

466

467

468

469

470

471

472

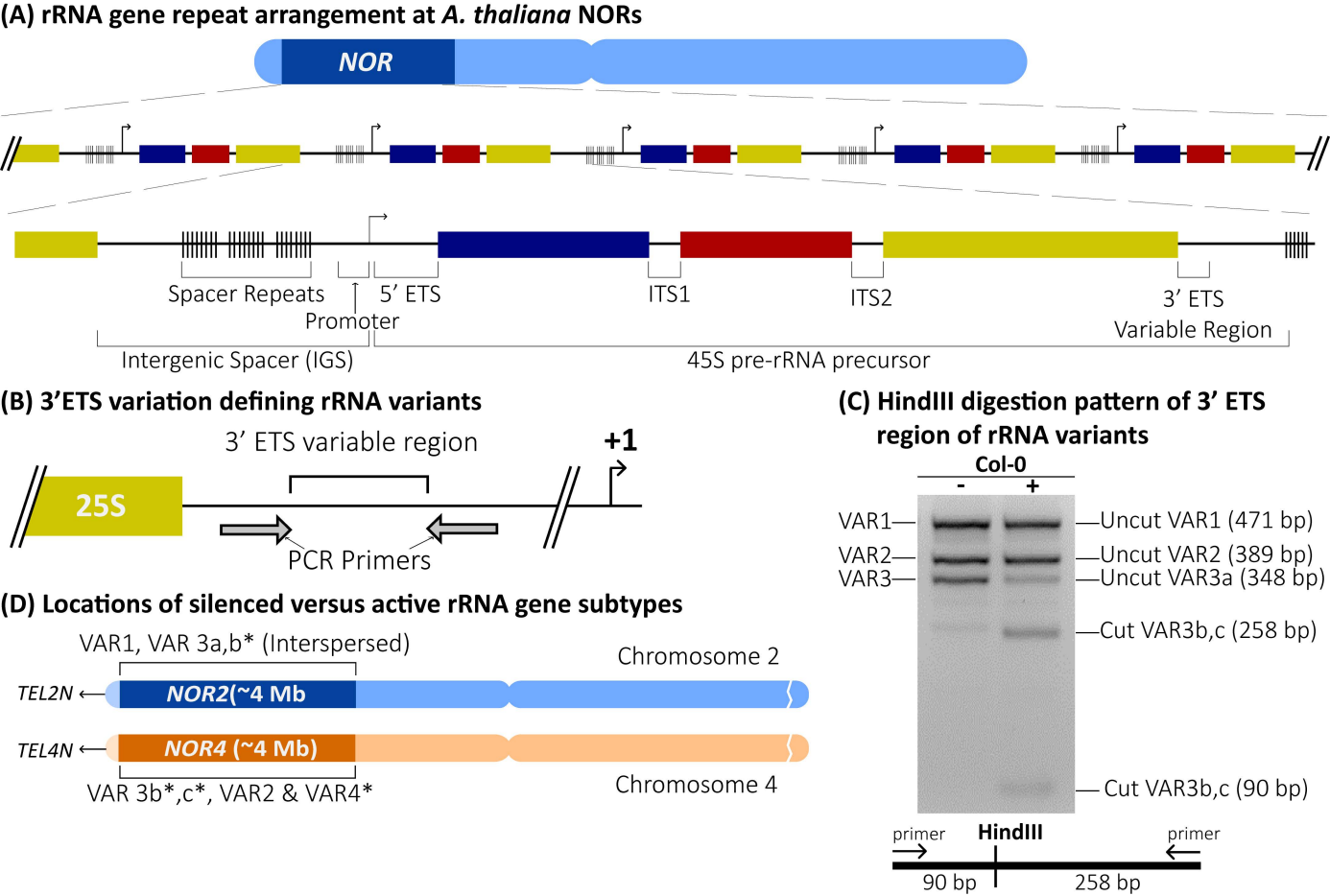


Figure 1. Ribosomal RNA (rRNA) gene organization, subtype variation, localization and their relative arrangement

rRNA subtype expression in DNA methylation mutants

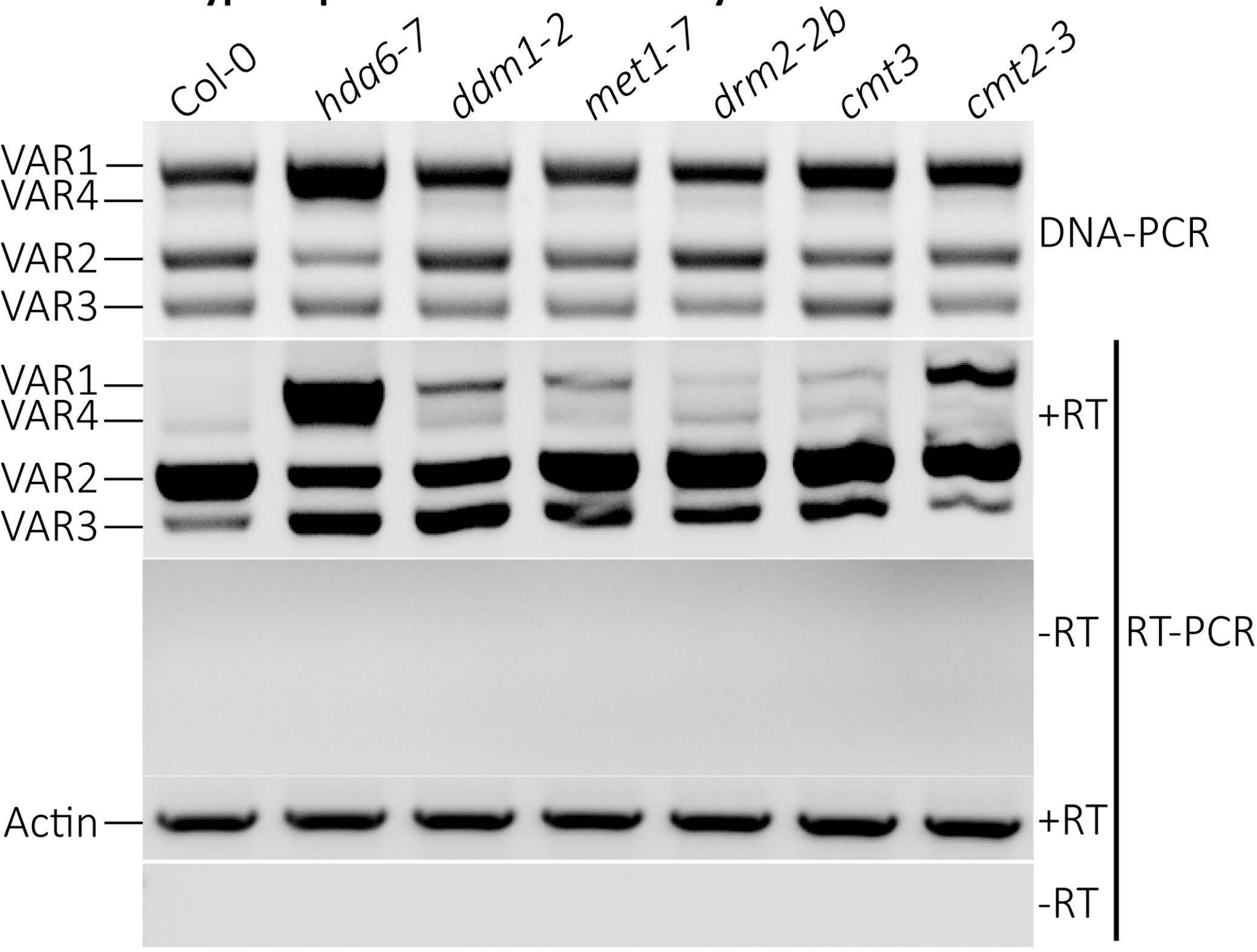


Figure 2. Disruption of *NOR2* gene silencing to varying degrees in DNA methylation mutants.

The gel images show DNA-PCR data for rRNA gene subtypes and RT-PCR expression data for rRNA gene subtypes and actin.

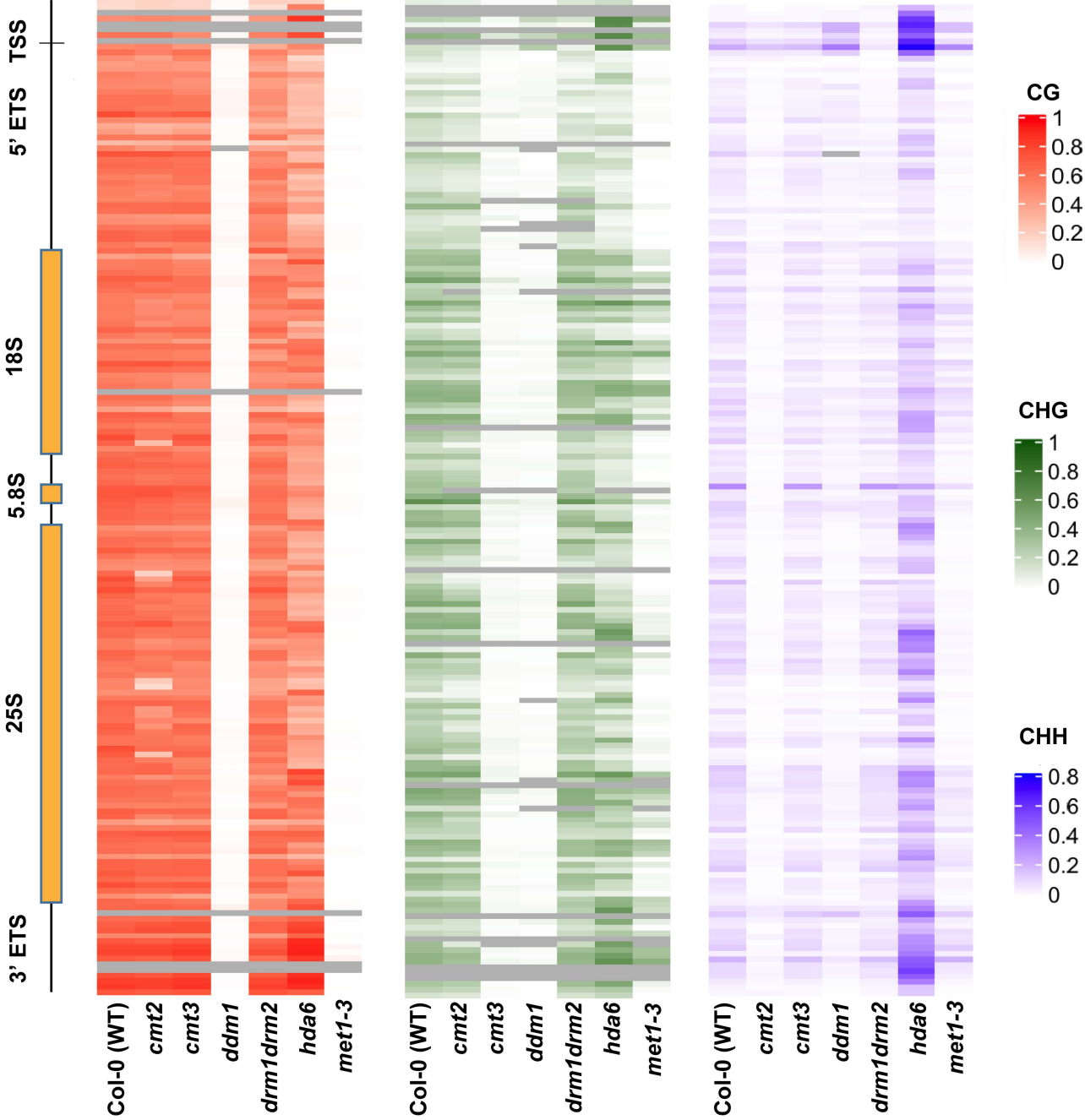


Figure 3: Methylation state of rDNA in CG, CHG and CHH contexts. Whole genome bisulfite sequencing data from various DNA methylation mutants was analyzed using Methylpy and the overall cytosine methylation state of the rDNA is shown as heatmaps in all three contexts. The Y-axis represents a single rDNA unit starting with a promoter and ending with 3' ETS. Each line on the heatmap represents the methylation state of cytosines in a given methylation context in 50 nucleotide bins. CG is shown in red, CHG in green and CHH in blue. Grey indicates that no methylation sites were present in that bin for a given context.

TSS= Transcription start site

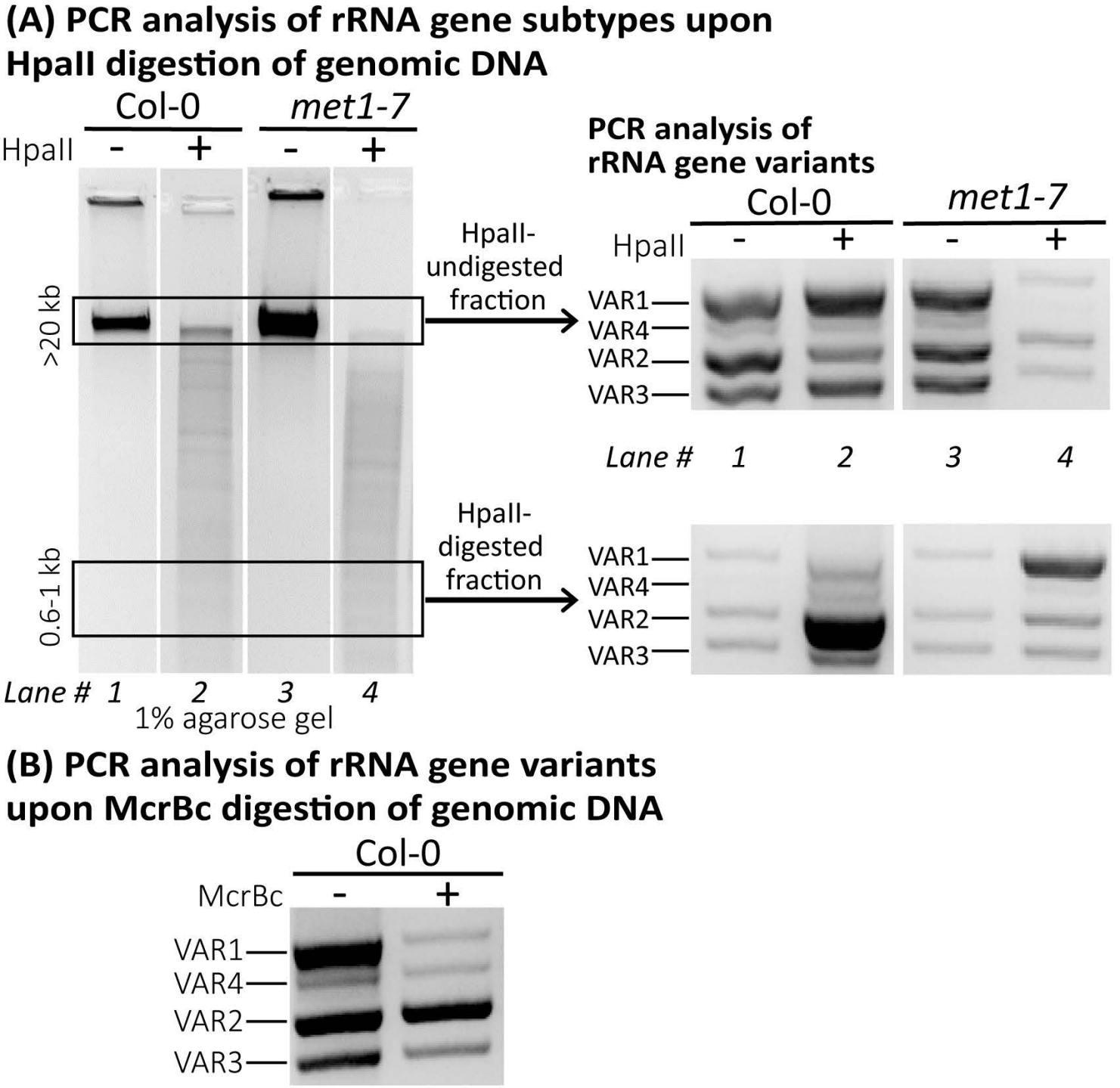
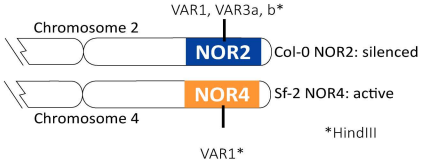


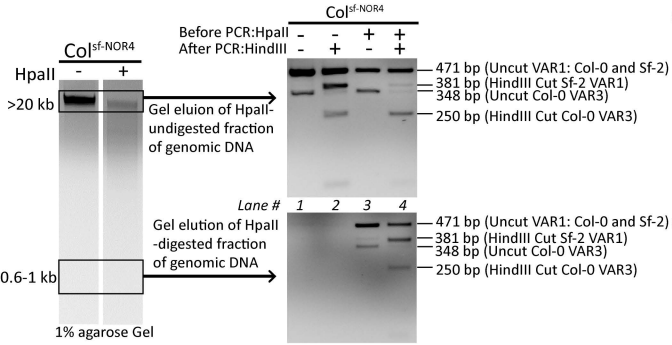
Figure 4. *NOR2* genes are hypermethylated and *NOR4* genes are hypomethylated.

(A) Gel images on the left show resolution of HpaII- or buffer-treated Col-0 and *met1-7* DNA samples and the gel images on the right show PCR amplification of rRNA gene subtypes. For these PCR reactions, gel-eluted DNA samples representing the HpaII-undigested (≥ 20 kb) and the HpaII-digested (0.6-1.0 kb) fractions were used. (B) The gel image shows rRNA gene subtype PCR products obtained from McrBc-digested or no enzyme-treated Col-0 DNA.

(A) Col$^{Sf\text{-}NOR4}$ line has Col-0 *NOR2* and Sf-2 *NOR4*



(B) HindIII CAPS analysis of rRNA gene subtypes amplified from gel-eluted fractions of HpaII-treated genomic DNA

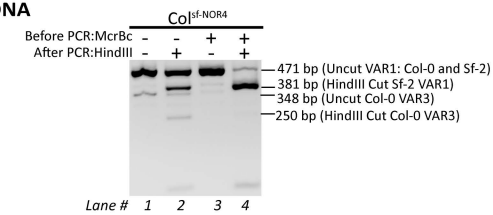


(C)

rRNA variant composition of PCR amplicons		Position of HindIII in PCR amplicons	Relative DNA methylation status	Lane 1 (no enzyme control)	Lane 2 (HindIII after PCR)	Lane 3 (HpaII before PCR & No HindIII)	Lane 4 (HpaII before PCR & HindIII after PCR)
VAR1	Sf-2 VAR1 (471 bp)	F HindIII R; 90 bp, 381 bp	hypomethylated	uncut	cut	cut by HpaII	cut by HpaII & HindIII
VAR1	Colo-0 VAR1 (471 bp)	F R	hypermethylated	uncut	uncut	uncut	uncut by HpaII or HindIII
VAR3	Col-0 VAR3a (348 bp)	F R	hypermethylated	uncut	uncut	uncut	uncut by HpaII or HindIII
VAR3	Col-0 VAR3b (348 bp)	F HindIII R; 90 bp, 258 bp	hypermethylated	uncut	cut	uncut	uncut by HpaII but cut by HindIII

F-forward primer; R-reverse primer

(D) HindIII CAPS analysis of rRNA gene subtypes amplified from McrBc-treated genomic DNA



(E)

rRNA variant composition of PCR amplicons		Position of HindIII in PCR amplicons	Relative DNA methylation status	Lane 1 (no enzyme control)	Lane 2 (HindIII after PCR)	Lane 3 (McrBc before PCR & No HindIII)	Lane 4 (HpaII before PCR & HindIII after PCR)
VAR1	Sf-2 VAR1 (471 bp)	F HindIII R; 90 bp, 381 bp	hypomethylated	uncut	cut	uncut	uncut by McrBc & cut by HindIII
VAR1	Colo-0 VAR1 (471 bp)	F R	hypermethylated	uncut	uncut	cut by McrBc	cut by McrBc & uncut by HindIII
VAR3	Col-0 VAR3a (348 bp)	F R	hypermethylated	uncut	uncut	cut by McrBc	cut by McrBc & uncut by HindIII
VAR3	Col-0 VAR3b (348 bp)	F HindIII R; 90 bp, 258 bp	hypermethylated	uncut	cut	cut by McrBc	cut by McrBc & HindIII

F-forward primer; R-reverse primer

Figure 5. Col-0 *NOR2* is hypermethylated and Sf-2 *NOR4* is hypomethylated in the Col$^{sf\text{-}NOR4}$ introgression line.
(A) Diagram showing rRNA subtype composition of the Arabidopsis introgression line called Col$^{Sf\text{-}NOR4}$ in which Sf-2 *NOR4* and adjacent sequences were introgressed into the Col-0 genetic background. (B) The agarose gel images on the left show resolution of HpaII- or buffer-treated Col$^{Sf\text{-}NOR4}$ gDNA samples, and the gel images on the right show PCR amplification of rRNA variants with or without HindIII digestion. For these PCR reactions, gel-eluted DNA samples representing the HpaII-undigested (≥ 20 kb) and the HpaII-digested (0.6-1.0 kb) fractions were used. (C) Summay of results shown in (B). (D) The Gel image shows PCR amplification of rRNA variants obtained from McrBc- or buffer-treated Col$^{Sf\text{-}NOR4}$ gDNA with or without HindIII digestion. (E) Summary of results shown in (D).

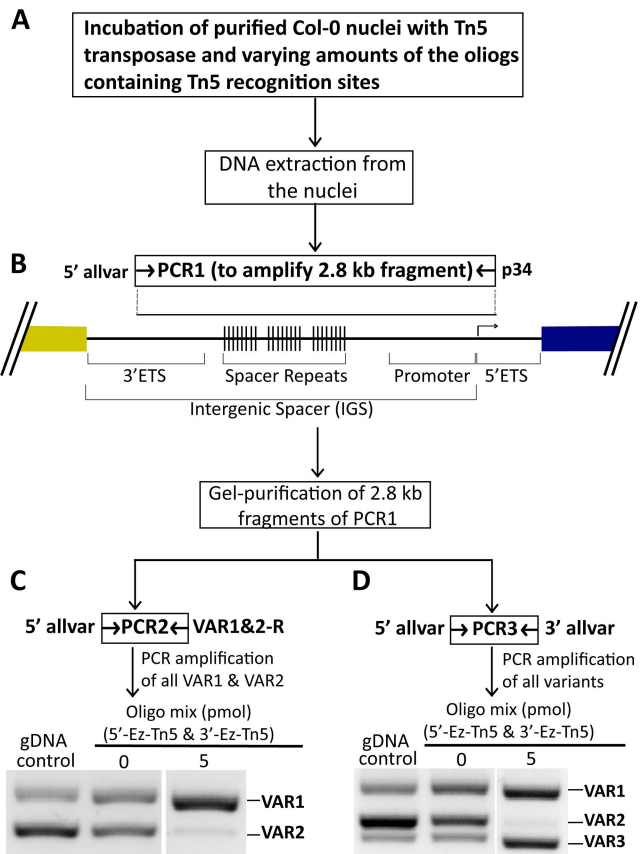


Figure 6. *NOR2* genes exist in a more condensed chromatin than *NOR4* genes.

(A) Tn5 Transposase-accessible chromatin assay (ATAC) was performed on nuclei isolated from wild-type Col-0 plants. (B) DNA isolated from the transposase-treated nuclei was used in PCR1 to amplify ~2.8 kb of rRNA gene promoter-proximal upstream regions involving the variable 3'ETS of the preceding gene and IGS. (C) The PCR1-amplified 2.8 kb fragments were gel-eluted followed by PCR2, products of which are visualized on a 2.8% agarose gel stained with ethidium bromide. Amplification of VAR1 and VAR2 genes from 3' ETS variable region from genomic DNA (gDNA) as a control is shown in the first lane. (D) The PCR1-amplified 2.8 kb gments were gel-eluted followed by PCR3, products of which are visualized on a 2.8% agarose gel stained with ethidium bromide. Here, PCR3 involves amplification of all rRNA variants. Amplification of all rRNA variants from 3' ETS variable region from genomic DNA (gDNA) as a control is shown in the first lane.